PUCRS

ESCOLA POLITÉCNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
DOUTORADO EM CIÊNCIA DA COMPUTAÇÃO

KARINA KOHL SILVEIRA

AN EMPIRICAL STUDY ABOUT THE IMPACT OF GENDER DIVERSITY IN SOFTWARE DEVELOPMENT TEAMS

Porto Alegre
2022

PÓS-GRADUAÇÃO - *STRICTO SENSU*

Pontifícia Universidade Católica
do Rio Grande do Sul

PONTIFICAL CATHOLIC UNIVERSITY OF RIO GRANDE DO SUL
SCHOOL OF TECHNOLOGY
COMPUTER SCIENCE GRADUATE PROGRAM

AN EMPIRICAL STUDY ABOUT THE IMPACT OF GENDER DIVERSITY IN SOFTWARE DEVELOPMENT TEAMS

KARINA KOHL SILVEIRA

Doctoral Thesis submitted to the Pontifical Catholic University of Rio Grande do Sul in partial fulfillment of the requirements for the degree of Ph. D. in Computer Science.

Advisor: Prof. Rafael Prikladnicki

Porto Alegre
2022

Ficha Catalográfica

S587a Silveira, Karina Kohl

An Empirical Study About the Impact of Gender Diversity in Software Development Teams / Karina Kohl Silveira. – 2022.
170 p.
Tese (Doutorado) – Programa de Pós-Graduação em Ciência da Computação, PUCRS.

Orientador: Prof. Dr. Rafael Prikladnicki.

1. Software Engineering. 2. Diversity. 3. Gender. 4. Empirical Research. 5. Quantitative. I. Prikladnicki, Rafael. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da PUCRS
com os dados fornecidos pelo(a) autor(a).
Bibliotecária responsável: Loiva Duarte Novak CRB-10/2079

KARINA KOHL SILVEIRA

AN EMPIRICAL STUDY ABOUT THE IMPACT OF GENDER DIVERSITY IN SOFTWARE DEVELOPMENT TEAMS

This Doctoral Thesis has been submitted in partial fulfillment of the requirements for the degree of Ph. D. in Computer Science, of the Computer Science Graduate Program, School of Technology of the Pontifical Catholic University of Rio Grande do Sul

Sanctioned on March 9th, 2022.

COMMITTEE MEMBERS:

Prof. Dr. Igor Fabio Steinmacher (PPGCC-CM/UTFPR)

Prof. Dr. Marcelo Soares Pimenta (PPGC/UFRGS)

Prof[a]. Dr[a]. Milene Selbach Silveira (PPGCC/PUCRS)

Prof. Rafael Prikladnicki (PPGCC/PUCRS - Advisor)

““*...her wings are cut and then she is blamed for not knowing how to fly.*””
(Simone de Beauvoir, The Second Sex.)

ACKNOWLEDGMENTS

First of all, I would like to thank my mother, Noeli; the best friend one can have. The most strong, amazing, and sweet woman I've ever known. The one who taught me to be an independent woman and always told me education and knowledge would be the path to that. The best people manager I know. Love you like crazy, Dona Noeli.

Also, I cannot forget the friends along the journey! People that I met in different moments of life: the "german," Mariana Vicilli (sorry for all the "*I can't do it this weekend! I have a paper to write*"); the wonderful trio that PUCRS brought me in 2018 (in alphabetical order to avoid issues): Aniele Lisboa, Daniela Kuinchtner and Cassio Trindade; Olimar Borges, for the countless messages like "hey, how's the thesis going?"; the amazing Luciane Fortes and Amanda Nery, for the friendship, support and for being such power of nature. To all the colleagues, software engineers, product managers, engineering managers, directors from different places, I interacted through these last four years.

To Globo and HPE for the financial support.

To the amazing Professors, who directly or indirectly contributed to my development as a researcher and to the development of this thesis: Professor Marcelo Pimenta (UFRGS), always present in different points of my journey; Professor Milene Silveira (PUCRS), for being a reference as a researcher and a warm and welcoming person; Professor Tayana Conte (UFAM), also for being a reference as a researcher and a power of nature; Professor Sabrina Marckzak (PUCRS); Professor Bogdan Vasilescu (CMU) and Professor Alexander Serebrenik (TUE), my former references in diversity research; Professors Soraia Musse, Isabel Manssour and Renata Vieira, for sharing with me an award-winning paper; Professors Eduardo Pellanda and André Pase (PUCRS), that gave me perspectives on how technology research is done in communication; Jorge Kotick Audy, the best networker I know, for being the one who introduced me to Rafael Prikladnicki. Last but not least, thanks to my advisor, Professor Rafael Prikladnicki: thanks for the autonomy you gave me and for being the advisor that I needed, not the one I (maybe) wanted.

AN EMPIRICAL STUDY ABOUT THE IMPACT OF GENDER DIVERSITY IN SOFTWARE DEVELOPMENT TEAMS

ABSTRACT

Diversity is a subject that different areas of knowledge in society have widely discussed. Studies say that diverse teams deliver better results but also show that there are gender biases that impact hiring decisions or that women in Open Source are sometimes less likely to have their code accepted. Discussions and research on diversity in Software Engineering have also been presenting results to make the area more equalitarian. In the context of empirical research on human and social aspects of Software Engineering, using a constructivist approach, this research aims to understand and describe which factors in software development teams can be more impacted by gender diversity. We performed two systematic mapping reviews, a case study with 14 software development teams, and a survey. We collected data from practitioners in the Software Engineering industry in multiple ways, allowing us to perform quantitative and qualitative analysis. Thus, this research aims to contribute to the scientific community by describing the behavior and performance of software development teams related to gender diversity based on data collected in the field.

Keywords: Software Engineering, Diversity, Gender, Empirical Research, Quantitative, Qualitative.

UM ESTUDO EMPÍRICO SOBRE O IMPACTO DA DIVERSIDADE DE GÊNERO EM TIMES DE DESENVOLVIMENTO DE SOFTWARE

RESUMO

A diversidade é um assunto amplamente discutido em diferentes áreas do conhecimento da sociedade. Estudos dizem que a diversidade torna as equipes melhores e oferece melhores resultados, mas também mostram que há preconceitos de gênero que afetam as decisões de contratação ou que as mulheres em Open Source, às vezes, têm menos probabilidade de ter seu código aceito. Discussões e pesquisas sobre diversidade na Engenharia de Software também têm apresentado resultados para tornar a área mais igualitária. No contexto da pesquisa empírica sobre os aspectos humanos e sociais da Engenharia de Software, utilizando uma abordagem construtivista, deseja-se compreender e descrever quais fatores nas equipes de desenvolvimento de software podem ser mais impactados pela diversidade de gênero. Realizamos duas revisões sistemáticas de mapeamento, um estudo de caso com 14 equipes de desenvolvimento de software e um *Survey*. A coleta de dados de profissionais da indústria de Engenharia de Software foi realizada utilizando diversas abordagens, permitindo realizar análises quantitativas e qualitativas. Dessa forma, esta pesquisa visa contribuir com a comunidade científica ao descrever o comportamento e desempenho das equipes de desenvolvimento de software em relação à diversidade de gênero, tendo como base, dados coletados em campo.

Palavras-Chave: Engenharia de Software, Diversidade, Gênero, Pesquisa Empírica, Quantitativo, Qualitativo.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1. INTRODUCTION

1.1 Motivation

There has been an increasing awareness that software engineering is multidisciplinary, a "social activity," and "essentially a human activity," and that Software Engineering (SE) researchers must make observations in the "real world." Consequently, the Software Engineering field has now widely embraced these alternative approaches to study human aspects [125].

Software engineering involves real people in real environments. People create software, people maintain software, people evolve software. Accordingly, to truly understand software engineering, it is imperative to study people - software practitioners as they solve real software engineering problems in real environments. This means conducting studies in field settings[93].

Storey et al. [126] say Software Engineering is at the forefront of innovation and research and involves the consideration of both human and technical aspects. The origins of Software Engineering come from the 1950s and 1960s when the field emerged as a sub-discipline of computer science and engineering. As such, it was highly technical and focused on solving technical, logical, and mathematical problems and solutions. However, seminal works have gradually drawn attention to the importance of developers and social factors in Software Engineering [148, 121, 22, 33], as well as the need to understand software development processes. Nowadays, software development is widely recognized as a socio-technical endeavor [151]: many researchers consider both technical and human aspects of software development in their efforts to understand software engineering practices and improve tools, and special-interest communities with a focus on human aspects (such as the CHASE, International Conference on Cooperative and Human Aspects of Software Engineering, community) have been formed.

Storey et al. [126] developed a strategy to understand how human aspects are studied (or not). When applying the strategy to a cohort of papers from the Software Engineering community they found that, at the community level, the papers analyzed strongly favour data strategies over strategies that directly study human and social aspects, and most research contributions consist of the design or evaluation of technical solutions. So, they propose that the community diversify the use of research strategies to have a deeper understanding of human and social aspects of software development practice while balancing the design and evaluation of innovations on the technical side.

Software development teams are formed by different people and lately, it is being discussed that we have under-represented groups such as gender, ethnicity, culture, and

others [28, 146, 80, 79, 27, 50, 26, 139, 140, 138]. Diversity is being intensively discussed in different knowledge areas of society and the discussions about it in Software Engineering are increasing as well.

Diversity is highly valued in modern societies. Social cohesion, tolerance, and integration are linked to tangible benefits including economic vibrancy and innovativeness. Diversity is a complex issue, as groups can be diverse in terms of various attributes, such as ethnicity, gender, age, and socio-economic background [4]. Furthermore, diversity can be a divisive topic that is clouded by emotion, partisan loyalties, and political correctness, all of which can hinder impartial discussions [54].

Page [106] says we cannot say if diversity is good or bad unless we first know what diversity is. He defines two kinds of diversity. Cognitive diversity is the differences in how we interpret, reason and solve. Identity diversity is defined as the differences in race, gender, age, physical, capabilities, and sexual orientation. Page [105] says diversity leads to better outcomes and identity diversity produces better outcomes indirectly. Identity diverse groups do perform better than homogeneous groups when the tasks are primary problem solving, when their identities translate into relevant tools, when they have little or no preference diversity, and when their members get along with one another. These features translate into high benefits of diversity and low costs. However, the connections between identity diversity and cognitive diversity in mathematics, for example, do not seem obvious. Can gender, race, ethnicity, or physical abilities influence the representations and analytical tools that a mathematician applies? Yes. In mathematics research, identity is less pertinent than academic training, however, it is possible that a person's identity can influence how they represent a mathematical problem as well as how they choose to solve it [105].

Also, in his studies, Page [105] found that a diverse group of problem solvers - groups of people with diverse tools - consistently outperformed groups of the best and the brightest. Forming two groups, one random (and therefore diverse) and one consisting of the best individual performers, the first groups almost always did better. So, he says that in his model, *diversity trumps ability*.

Gender diversity often refers to an equitable or fair representation of people of different genders [129]. It most commonly refers to an equal ratio of men and women but may also include people of non-binary genders. Non-binary is a spectrum of gender identities that are not exclusively masculine or feminine — identities outside the gender binary [137].

The lack of an obvious logical connection between identity diversity and cognitive diversity pertinent in areas of *STEM - Science, Technology, Engineering, and Mathematics* does not mean that those areas do not need to be inclusive [106]. On the contrary, because the mathematical community faces complex problems, it needs to be cognitively diverse. Years ago, it was reckoned that the under-representation of women and some racial groups in math and science was generally lacking in interest. "*Women don't want to become physical*." Recently, some have attributed the low numbers to a lack of cognitive abilities. Current

thoughts lead to the effects of limited opportunities and exposure, the absence of role models, and the effects of non-inclusive behaviors and discrimination. Conflicting ideological positions between "*diverse teams do better*" or that "*we should hire for skills*" get nowhere. It is necessary to understand the conditions under which the benefits of diversity take place [106].

1.2 Research Questions

In the context of empirical research on human and social aspects of software engineering, there is an opportunity to identify which factors benefit most from gender diversity in software development teams. This research aims to understand and describe factors in software development teams which can be more impacted by gender diversity, such as Pull Request Sizes and Pull request Lead Time, through a constructivist approach. To achieve that, this thesis intends to answer the following research question:

What are the impacts of gender diversity on the performance and results of software development teams?

To help to answer this research question, the following sub-questions were defined:

- **RQ1** *What are the effects of gender diversity on software development teams when we analyzed pull based metrics?*
- **RQ2***What are the perceived benefits of gender diversity on software development teams reported by individuals?*
- **RQ3** *What are the perceived difficulties of gender diversity on software development teams reported by individuals?*

1.3 Goals

The main objective of this research is, through the use of empirical strategies, to understand and describe factors in software development teams which can be more impacted by gender diversity.

To reach this goal, we identified the following specific goals:

- Deepen studies on human and social aspects in software engineering;
- To conduct a systematic literature mapping to understand the state of the art in software engineering diversity studies;

- To run a case study to observe in the field the impacts of gender diversity on software development teams;
- To run a survey to observe in the field the impacts of gender diversity on software development teams;
- To document and to report the results to the scientific community.

1.4 Volume Organization

The remainder of this document is organized as follows: Chapter 2 presents the background on fundamental concepts. Chapter 3 describes the proposed research methodology for this work. Chapters 4, 5, and 6 present the studies undertaken to answer our research questions. Chapter 7 discusses the results. Finally, Chapter 8 shows our conclusion and final thoughts.

2. THEORETICAL BACKGROUND

This chapter presents the background on the core concepts related to this research: Software Engineering, Empirical Studies in Software Engineering, Social and Human Aspects in Software Engineering, Diversity, Gender Diversity, Pull-based Software Development, Pull request metrics, Repositories Activities, and Empathy.

2.1 Software Engineering

The IEEE Standard Glossary of Software Engineering Terminology [114], defines Software Engineering as the application and studies for systematic, disciplined, quantifiable approach to the development, operation, and maintenance of software; that is, the application of engineering to software.

The discipline aims to enable the successful production of software, where the criteria for success can include such quality characteristics as accuracy, appropriateness, functionality, reliability, usability, efficiency, maintainability, and portability, as well as timeliness, cost-effectiveness, customer satisfaction, or even political expedience [39].

Wohlin et al. [154] say Software Engineering is a cross-disciplinary subject. It stretches from technical issues such as databases and operating systems, through language issues, for example, syntax and semantics, to social issues and psychology. Software development is human-intensive. It is a discipline based on creativity and the ingenuity of the people working in the field. Nevertheless, we should, when studying and doing research in Software Engineering, aim at treating it as a scientific discipline. This implies using scientific methods for doing research and when making decisions regarding changes in the way we develop software [154].

2.1.1 Social and Human Aspects in Software Engineering

Modern software engineering involves both human and technical aspects, the importance of which is widely accepted by practitioners and researchers alike. At a community level, software engineering researchers may be expected to choose a balance of research strategies that capture both social and technical characteristics of software development [127].

Easterbrook et al. [42] say Software Engineering is a multi-disciplinary field, crossing many social and technological boundaries. To understand how software engineers construct and maintain complex, evolving software systems, we need to investigate not just the

tools and processes they use, but also the social and cognitive processes surrounding them. This requires the study of human activities. We need to understand how individual software engineers develop software, as well as how teams and organizations coordinate their efforts [42].

Nowadays, we recognize software engineering as a socio-technical endeavor [151]. Social aspects are becoming an increasingly critical part of the software engineering practice and research landscape [46]. What is more, while we may expect that many of our contributions are purely technical, somewhere, at some time, a software developer may be affected by our work. It is crucial to account for the social aspects of software engineering in our research, and we know that to capture them, we need appropriate driving research questions and methods as well as a focus on relevant stakeholders [119].

Storey et al. [128] say social aspects can be approached methodologically by inferring behavior from analyzing trace data of developers' past activities (e.g., code commits, code review comments, posted questions and answers on developer forums, etc.). But the analysis of trace data alone is fraught with threats to validity as it shows an incomplete picture of human behavior, intent, and social interactions in software engineering. Furthermore, trace data alone cannot be used to predict how a new solution may perturb an existing process in industry settings, although relying on trace data can bring early insights about the feasibility of a solution design. To appropriately capture and account for social aspects in software engineering research, we need to use dedicated methods that directly involve human participants in our empirical studies [128].

Empirical studies can be used as a research strategy when the research objective is related to the human and social aspects of Software Engineering [113]. Empirical research is research that is based on observation and measurement of phenomena, as directly experienced by the researcher. The data thus gathered may be compared against a theory or hypothesis, but the results are still based on real life experience. The data gathered is all primary data, although secondary data from a literature review may form the theoretical background [113].

2.1.2 The Software Team

DeMarco and Lister [41] say there is a tendency to use the word team consistently and vaguely in the business area, calling any group of professionals assigned to work together a "team." However, many of them do not resemble teams. There is no standard definition of success or an identifiable team spirit. What is missing is a phenomenon called consistency. A consistent team is a group of people so cohesive that the whole is greater than the sum of the parts. When a team starts to be consistent, the probability of success

increases, the team can become unbeatable, and certainly won't need to be motivated. It acquires speed and momentum [41].

DeMarco and Lister [111] argue that consistent team members are significantly more productive and motivated than average. They share a common goal, a common culture, and, in many cases, a sense of belonging to an elite team that makes them unique [111].

There is no infallible method for creating a consistent team [111]. However, there are attributes commonly found in effective software teams. It is suggested that an effective software team establish *a sense of purpose*, for example, all team members agree that their goal is to develop software that will transform a product category and, as a consequence, transform your company into an industry leader, they have a strong sense of purpose. An effective team must also embody a *sense of involvement* that allows each member to feel that their strengths and contributions are valuable. There is also the *sense of trust*. The software engineers on the team must trust the skills and competence of their colleagues and managers. The team should encourage a *sense of improvement*, periodically reflecting on their approach to software engineering and looking for ways to improve their work [111].

However, not all teams become consistent. Jackman [81] says many teams suffer from so-called "team toxicity," defined by five factors that promote a potentially toxic team environment: (1) a frenetic work atmosphere; (2) a high degree of frustration that causes friction among team members; (3) a fragmented or inefficiently coordinated software process; (4) a vague definition of roles within the team and (5) continual and repeated exposure to failure.

The most effective software teams are diverse in the sense that they combine a variety of different strengths. Highly skilled technicians are complemented by members who may have a less technical background but better understand the needs of those involved [111].

Pressman [111] say software teams often expend effort on the different characteristics of their members. Some are extroverts, others introverts. Some gather information intuitively, distilling broad concepts out of disparate facts. Others process information linearly, collecting and organizing minimal details of the data provided. Some feel comfortable making decisions only when a logical and orderly argument is presented. Others are intuitive, used to making decisions based on insight. Some developers want a detailed schedule filled with organized tasks that allow them to get to the closure of some element of the project. Still, others prefer a more spontaneous environment, in which results and open-ended questions will suffice. Some work hard to complete the steps well ahead of the set date, thus avoiding stress as the deadline approaches, while others are spurred on by the rush to make it to the deadline. It is important to note that recognizing human strengths is the first step towards creating consistent teams [111].

2.1.3 Software Engineering Metrics

Software metrics are quantitative measures that allow verifying the effectiveness of the software process [111]. Basic quality and productivity data is collected, which is analyzed against past averages and evaluated to determine whether the quality and productivity improvements have occurred [111]. Metrics are also used to pinpoint problem areas so that fixes can be developed and the software process improved [111].

The analysis of metrics is essential so that assessments are not subjective. With measurement, we can detect trends (both good and bad), the estimates can be better, and significant improvements can be obtained over time [111].

We can apply measurements to the software process to continually improve it [111]. We can use them during a software project to support estimation, quality control, productivity, and project control. Software engineers can also use them to assess the quality of artifacts and assist in tactical decision-making as the project progresses [111].

In the context of the software development process and projects, a software team is primarily concerned with productivity and quality metrics [111]. Software development output is measured as a function of effort and time applied, and fitness-for-use measures of the artifacts produced. For planning and estimating purposes, the interest is historical. What was the productivity of software development in past projects? What was the quality of the software produced? How can productivity and quality data from the past be extrapolated to the present? How can this help make more accurate plans and estimates? [111].

2.1.4 Pull-Based Software Development

The case study we will present in Chapter 5 rely on quantitative analysis of data extracted from software development teams code repositories. Those data are based mainly on Pull Requests submitted by the different team members to the code repositories of the team. Considering this, it is essential to contextualize pull-based software development, pull requests, and how they unfold for this research.

Software Engineering methodologies rely on version control systems such as git to store source code artifacts and manage changes to the codebase [104]. The pull-based development model [12] is widely used in distributed software teams to integrate incoming changes into a project's codebase [60].

The pull-based software development model, exemplified and popularized by GitHub [1], decouples a software development task from the decision to incorporate its results in the

[1] https://www.github.com

codebase: when the software development task is completed, its author submits a pull request [49]. A pull request is the ability to propose changes to a remote codebase, which is often the one considered to be the central repository. A pull request is based on those changes, or delta, between a locally cloned codebase and the target central repository. As its name indicates, a pull request is not an immediate change in the codebase but a request for it [104].

Pull-requests are used in many scenarios beyond basic patch submission, e.g., conducting code reviews, discussing new features [60]. Pull requests include chunks of source code, history of changes, log messages around a proposed change of the mainstream codebase, and much discussion on whether to integrate such changes or not [104]. When team members open a pull request, they propose changes and request that someone review and pull in the contribution and merge them into their branch [57]. The decision to merge the pull request is ultimately taken by people in the roles of core developers (those who directly perform changes in the codebase) or integrators (those who are responsible for integrating the changes into the codebase) [104].

Pull Request Metrics

Pull request metrics are important for software development. They can tell a lot about the efficiency of the development process and the teams' speed and capacity.

Pull Requests are part of the development workflow. It consists of comparing the changes of a branch with the repository's base branch. Pull Requests provide useful and actionable metrics[94].

We opted for using Pull Request Size and Pull Request Lead Time as the pull request metrics in this research, once they could be easily extracted from the repository data we had access. Their definitions follows:

- **Pull Request Size (PRS)** is the number of lines involved or the number of added/deleted lines [71]. The highest the number, the bigger is the pull request. Large pull requests carry more risk when deploying to production and are more challenging to review, merge, and release. Deploying pull requests of a reasonable size enables the team to review and ship new features at a faster cadence and with greater confidence [56, 68].

- **Pull Request Lead Time (PRLT)** metric gives an idea of how many times (usually in days) pull requests take to be merged or closed [68].It is the time between the first commit on a branch and the merge action of a pull request on that branch. The timestamp of the first commit on a branch is subtracted from the timestamp on the merge action of the pull request [56]. It differs from Time to Merge, which is how much time it takes for the first commit of a branch to reach master [56, 68] .

There are some others pull request metrics as Pull Request Flow Ratio and Pull Requests Discussions [68]. We decided for Pull Request Size and Pull Request Lead Time once were the ones we had enough information in our dataset to measured.

2.2 Diversity

Page [106] defines cognitive diversity as the differences in how we interpret, reason, and solve problems. He defines identity diversity as the differences in race, gender, age, physical, capabilities, and sexual orientation. Hewlett et al. [75] distinguishes between acquired and inherent diversity. Acquired diversity consists of experiences, along with learned behaviors and traits. We choose some acquired differences. Others we obtain by chance. Inherent diversity consists of immutable attributes: race, age, physical qualities, gender, ethnicity, and sexual orientation. Jehn et al. [82] distinguishes between social category diversity, informational diversity, and value diversity. Social category diversity or identity diversity refers to differences in age, race, gender, ethnicity, physical qualities, sexual orientation, and religion. This differs from inherent diversity in that some types of social diversity, notably religion, can be acquired. Informational diversity refers to differences in knowledge and perspectives. Value diversity corresponds to differences in principles and standards.

Page [105] says diversity leads to better outcomes and that identity diversity produces better outcomes indirectly. Any claim that identity diversity creates collective benefits requires two links. The first link connects identity diversity to cognitive diversity. The second link connects these diverse talents to relevant problems. We can take the connection between identity and cognition too far. Identity diverse people can think alike, and people belonging to the same race, age, gender, religion, and social class can also think differently. For the second link, identity difference can contribute to better outcomes only if the task is appropriate, and it happens in problem solving and prediction. Page [105] also says in many cases, identity diverse groups do perform better than homogeneous groups, and those situations are far from random. Identity diverse groups perform better when the tasks are primary problem solving, when their identities translate into relevant tools, when they have little or no preference diversity, and when their members get along with one another. These features translate into high benefits of diversity and low costs.

2.2.1 Diversity Impact on Teams

Page [106] reviews 50 years of empirical studies on team performance, where the direct effects of diversity are tested. We list some of them below:

Empirical evidence for the benefits of cognitive diversity and identity diversity takes many forms. They include correlational data, controlled experiments, and case studies. In some domains, prediction, in particular, the evidence for the significance of the diversity bonus is unequivocal. As expected from the theory, the direct evidence for identity diversity bonuses will be more mixed. It will exist in some cases but not in others. However, we should not view empirical evidence as a final arbiter. Data reveals the world as it is, not as it could be [106].

For more than two decades, organizational scholars have noticed that the predominance of teamwork has increased. An example from the creative area: teams of three or more composers wrote most of the top 100 *hits* on the *Billboard*. Teams perform better. When teams compete with individuals, they usually win as they can extract more extensive cognitive repertoires. A team has more information, ideas, knowledge, and ways of thinking than a single individual. A team can access more perspectives and more tools. This abundance of cognitive tools allows teams to produce more ideas and find improvements in the ideas found [106].

A study by McKinsey & Company [78], analyzes the senior management teams of 366 companies in the United States, United Kingdom, Canada, and Latin America and finds a positive linear relationship between diversity and financial performance. Top quartile companies for gender diversity outperform those in the bottom quartile by 15%. Companies in the top quartile for ethnic diversity outperform those in the bottom quartile by 35% [106].

The literature that analyzes racial and cultural diversity and the economic performance of cities and regions also shows a correlation. Racial diversity significantly improves performance in advertising, finance, entertainment, legal services, healthcare, hotels, bars and restaurants, and computer manufacturing. Industry-level analysis suggests that racial diversity improves performance in problem-solving, creative thinking, and understanding customers. As one would expect from the logic of diversity bonuses, increasing racial diversity does not increase performance in industries that involve physical labor (Page [106] apud Sparber, 2009).

2.2.2 Gender Diversity

Gender diversity often refers to an equitable or fair representation of people of different genders [129]. It most commonly refers to an equal ratio of men and women but may also include people of non-binary genders. Non-binary is a spectrum of gender identities that are not exclusively masculine or feminine — identities outside the gender binary [137].

Page [105] says that identity differences lead to experiential differences that in turn create tool differences. We can see this in the context of gender differences. Most people treat and react to men and women differently. Because we treat men and women differently,

we provide them with different experiences. As a result, they learn to think about situations differently. Men and women may differ in the tools they choose to acquire, yet this does not in any way imply that they differ in the perspectives, heuristics, interpretations, and predictive models they *could* acquire [105].

Mendez et al. [95] say social diversity has a positive effect on productivity, teamwork, and quality of contributions. Also, their research shows that gender diversity positively affects productivity in Open Source Software (OSS) communities. Kohl and Prikladnicki [122] mapped how the Software Engineering community have studied the subject from 2001 to 2018, reaching 221 studies, most of them, 129, focusing on diversity and agile methodologies. In Chapter 4 we present a recent Systematic Mapping, considering the period of January 2010 to March 2020, with publications in Software Engineering focusing mainly on gender issues.

Women in Computing

Women had a strong presence in computer science programs in the United States and Brazil until the mid-1980s when the trend reversed, and the area became occupied mainly by men. This situation gradually began to attract some higher education institutions, such as Carnegie Mellon University, in Pennsylvania, in the United States, which was one of the first to try to understand and reverse this imbalance in its computer science programs. The movement, over time, spread to other institutions, including Brazil [5].

The need to have more women in computer science is not just a matter of gender equity but also an economic one. The number of computer courses grew 586% in the last 24 years in Brazil [5]. The percentage of women enrolled in these courses decreased from 34.8% to 15.5%, according to data from the National Institute for Educational Studies and Research (Inep) of the Ministry of Education (MEC). The scenario is similar in the United States. The estimation for 2020 was approximately 1.4 million vacancies in the information technology area and a labor deficit of around 1 million professionals, according to the organization *Code.org* [5].

Women were once the majority in computer science courses at some of the main universities in Brazil. In 1974, 14 of the 20 students who graduated from the first class of the bachelor's degree in computer science at USP's Institute of Mathematics and Statistics (IME) were women. In 2016, however, of the 41 students who completed the course, only six were women. Researcher Claudia Bauzer Medeiros, from the Institute of Computing at the State University of Campinas (IC-Unicamp), says that the case of women in this field clashes with other STEM areas (*Science, Technology, Engineering, and Mathematics* - Science, Technology, Engineering, and Mathematics) because they were already the majority in the course [5]. This phenomenon follows the same movement observed in the United States. According to data from the National Center for Education Statistics in that country, women

represented nearly 37% of all undergraduate computer science students between 1984 and 1985. At the same time, data from the *American Bar Association*, *American Association of Medical Colleges* and the *National Science Foundation* point to growing participation of women in higher education courses in the fields of law, physical sciences, medicine, and computer science between the 1960's and 1980's [5].

However, from 1985 onwards, while the incidence of women in other courses continued to increase, exceeding 40% in 2015, in computer science, the movement turned to a fall, going from about 35% to less than 20% in 2015. The primary explanation for the phenomenon, both in Brazil and in the United States, would be in the 1980s, with the popularization of personal computers [5].

Then huge calculating machines, after the end of World War II (1939-1945) computers were mainly used in activities associated with the secretarial function, such as data processing and electronic tabulation. The predominance of women was evident. When calculating the payroll of companies, for example, they used to write codes that would later be transformed into punched cards to be read by machines [5].

In the mid-1980's, with the arrival of the first personal computers, this scenario changed. A 1985 report produced by the National Center for Education Statistics found that boys in the United States were much more likely to use these machines at home than girls, possibly because the manufacturers' marketing was primarily aimed at them. This may have contributed to the men's learning and interest in programming. Over time, the notion that the programming activity was male became the standard narrative, contributing to the forgetting of the names of female characters with essential contributions to the field [5].

Marie Hicks [76] says women were the computer industry's largest technically trained workforce during World War II and into the mid-1960s. They operated room-sized electromechanical computers that cracked codes, worked out military logistics, and did ballistic calculations. They later went to work for civil service departments, operating the computers needed by the government to collect data and process it correctly, which was seen as "unskilled and highly feminized work." Management considered women ideal for the computer industry because they didn't think they needed any career. The expectation was that a woman's career would be short because of marriage and children, which meant a workforce that would not be frustrated or demand higher promotions and salaries [76].

In the 1970's, there was a change in mindset, and women were no longer welcome in the workplace: government and industry had understood the capabilities of computers and wanted to integrate their use at the managerial level [76]. Women would not be placed in charge of computers as they were seen as low-level workers. Women were systematically eliminated and replaced by men who were paid more and had better jobs [76].

3. RESEARCH METHODOLOGY

In this chapter, we discuss aspects related to the research methodology followed in this thesis. In Section 3.1 we explain the research methods used in this work. Section 3.2 depicts the research design.

3.1 Methodological Background

We used the following research methods for this thesis: systematic literature mapping study, case study, and survey. We explain each one as follows.

3.1.1 Systematic Literature Mapping Study

Kitchenham and Stuart [87] define a systematic literature mapping study (also referred to as a scoping study) as a broad review of primary studies in a specific topic area that aims to identify what evidence is available on the topic. Petersen et al. [107] say systematic mapping study provides a structure of the type of research reports and results that have been published by categorizing them and often gives a visual summary, the map, of its results. It often requires less effort (than a systematic literature review) while providing a more coarse-grained overview [107].

The main goal of systematic mapping studies is to provide an overview of a research area and identify the quantity and type of research and results available within it. Often one wants to map the frequencies of publication over time to see trends. A secondary goal can be to identify the forums in which research in the area has been published [107].

The Systematic Mapping Process

Petersen et al. [107] say the essential process steps for a systematic mapping study are the definition of research questions, searching for relevant papers, screening of papers, keywording of abstracts, and data extraction and mapping (Figure 3.1). Each process step has an outcome, the final outcome of the process being the systematic map.

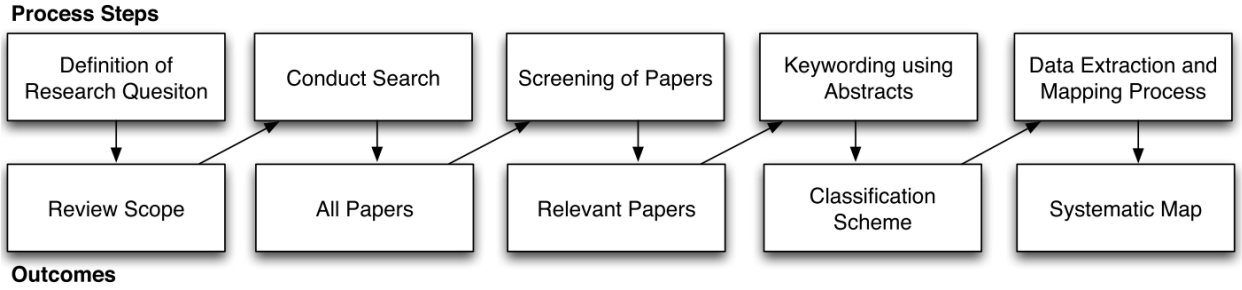


Figure 3.1: Systematic mapping Process, by Petersen et al. [107]

The primary studies are identified by using search strings on scientific databases or browsing manually through relevant conference proceedings, or journal publications [107]. Inclusion and exclusion criteria are used to exclude studies that are not relevant to answer the research questions [107].

Keywording of abstracts (Classification Scheme) [107] is a way to reduce the time needed in developing the classification scheme and ensuring that the scheme takes the existing studies into account. Keywording is done in two steps. First, the reviewers read abstracts and look for keywords and concepts that reflect the paper's contribution. While doing so, the reviewer also identifies the context of the research. When this is done, the set of keywords from different papers are combined together to develop a high-level understanding of the nature and contribution of the research. This helps the reviewers come up with a set of representative categories of the underlying population. When abstracts are of too poor quality to allow meaningful keywords to be chosen, reviewers can choose to study also the introduction or conclusion sections of the paper. When a final set of keywords has been chosen, they can be clustered and used to form the categories for the map [107].

In the context of this research, two systematic literature mapping studies were performed. The first one, at the beginning of the process, focusing on diversity in a broader way and aiming to understand the state of studies related to diversity in software engineering. The study was published in CHASE 2019 [90]. A second one in 2020, narrowing the focus to gender diversity and aiming to see how the studies evolve in. These two systematic mapping studies support addressing all the research questions of this thesis.

3.1.2 Case Study

A case study is an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident [155]. Its designs can be single-case or multiple-case studies, and they may entail a single unit or multiple units of analysis[155].

Yin [155] says research questions focusing on "*what*" questions, can be exploratory questions. This type of question is a justifiable rationale for conducting an exploratory study,

the goal being to develop pertinent hypotheses and propositions for further inquiry. As an exploratory study, an exploratory case study can be used as a research strategy.

Runeson et al. [116] say case studies are, by definition, conducted in real-world settings and thus have a high degree of realism, mostly at the expense of the level of control. The data collected in an empirical study may be quantitative or qualitative. Quantitative data involves numbers and classes, while qualitative data involves words, descriptions, pictures, diagrams, etc. Quantitative data is analyzed using statistics, while qualitative data is analyzed using categorization and sorting. Case studies tend mostly to be based on qualitative data, as these provide a richer and deeper description [116]. However, Yin [155] say case studies can include and even be limited to quantitative evidence. In fact, the contrast between quantitative and qualitative evidence does not distinguish the various research strategies. Note that, as analogous examples, some experiments and some survey questions rely on qualitative and not quantitative evidence. Likewise, historical research can include enormous amounts of quantitative evidence.

Case Study Research Process

Runeson et al. [116] say when conducting a case study, there are five major process steps to be walked through:

1. Case study design: objectives are defined, and the case study is planned.
2. Preparation for data collection: procedures and protocols for data collection are defined
3. Collecting evidence: execution with data collection on the studied case.
4. Analysis of collected data
5. Reporting

This process is almost the same for any kind of empirical study. However, as case study methodology is a flexible design strategy, there is a significant amount of iteration over the steps. The data collection and analysis may be conducted incrementally. If insufficient data is collected for the analysis, more data collection may be planned, etc. [116].

3.1.3 Survey

Survey research is the process of conducting research using surveys that are sent to respondents by the researchers [10]. The data collected from this process is then analyzed in order to draw conclusions. Surveys are also favored for "what" questions [155].

Kitchenham et al. [108] say the survey is not just the instrument (the questionnaire or checklist) for gathering information. It is a comprehensive system for collecting information to describe, compare or explain knowledge, attitudes, and behavior.

Thus, the survey instrument is part of a larger survey process with clearly defined activities [108]:

1. Setting specific, measurable objectives
2. Planning and scheduling the survey
3. Ensuring that appropriate resources are available
4. Designing the survey
5. Preparing the data collection instrument
6. Validating the instrument
7. Selecting participants
8. Administering and scoring the instrument
9. Analyzing the data
10. Reporting the results.

3.2 Research Design

This section describes the research design of this thesis. First, an overview of the research design is presented, based on the model proposed by Wohlin and Aurum [153].

The process of preparing research involves several decision points. Wohlin e Aurum [153] illustrate some important decision points when conducting empirical research in software engineering. These decision points are grouped into three phases: strategy, tactical, and operational.

- The **strategy phase** involves a plan that gives direction to the researcher for the tactical and operational phase of the research. This phase enables the researcher to conduct the research systematically and to position it in relation to different general approaches to research [153].
- The **tactical phase** involves decisions on how to operationalize the research activities in terms of how to approach the research questions more specifically. The decision

points are the research process and research methodology. The tactical decisions enable the research to achieve the research goal. The tactical phase focuses on selecting the actual process and methodology to use to achieve the research goal [153].

- The **operational phase** involves decisions on actions that will be taken when implementing the research, including data collection methods and data analysis techniques. Thus, actually planning the details and collecting the data to be able to respond to the stated research questions. The operational phase is focused on actually carrying out the empirical research by collecting and analyzing the data [153].

Figure 3.2 presents the research design diagram to conduct the study proposed in this thesis, based on the model proposed by Wohlin and Aurum [153]. Figure 3.3 presents the research roadmap for each phase and decision point through time.

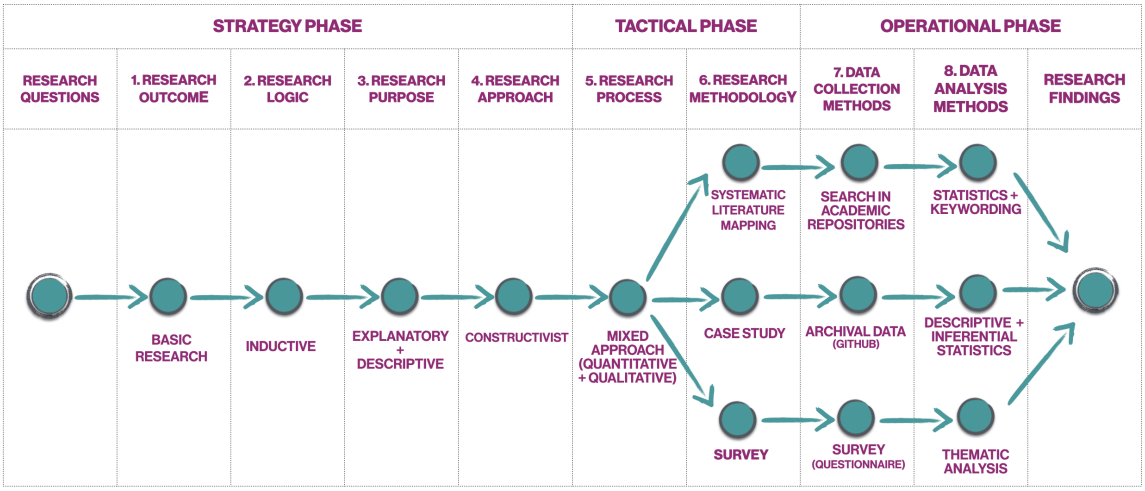


Figure 3.2: Research Design. Source: The Author.

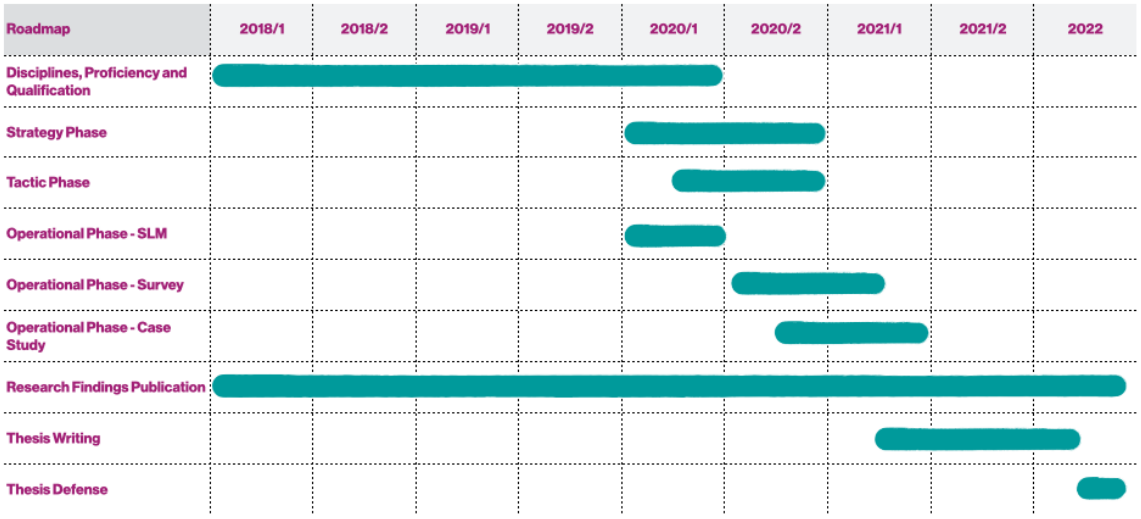


Figure 3.3: Research Roadmap. Source: The Author.

The research design for this thesis consists of quantitative and qualitative studies, using a combination of the following methodologies: systematic literature map, case study, and survey. Data collection happened through collecting data from code versioning repositories of the teams and data analysis through descriptive and inferential statistics. The survey was used as a methodology and data collection strategy, and for data analysis, we used thematic analysis. The three phases of the research design and respective decision points are described in the next Sections.

3.2.1 Research Strategy Phase

The strategy phase enables the researcher to conduct the research systematically and to position it in relation to different general approaches to research. The research strategy involves decisions on research outcome, research logic, research purpose, and research approach. The strategy phase sets the stage for the research [153]. The decision points related to the strategy phase follow.

Research Question:

Formulation of the research question(s) is critical. A research question determines or strongly influences the rest of the process in the research, including research methodology, data collection methods, and data analysis methods [153]. However, in practice, the research question(s) may evolve during the research, and the researcher may need to adjust the research question(s) several times to fit with the results of their findings [42, 116].

A research question may be related to a set of hypotheses, concepts, or relationships between concepts or two phenomena that require clarification [153]. Easterbrook et al. [42] argue that in an early stage of the research, the research question(s) tends to be explorative, but once the researcher has a clear idea about the problem, the research question(s) tends to search for the patterns, the relationships between the two phenomena or search for a causal effect between the two phenomena. The research questions for this thesis were presented in Section 1.

Research outcome (Decision Point 1):

The outcome of the research can be classified as basic or applied research. *Basic research* is applied to a problem where the emphasis is the understanding of the problem rather than providing a solution to a problem; hence the main contribution is the knowledge generated from the research. This type of research tends to be less specific, and the outcome of this type of research is knowledge [153]. *Applied research* is the type of research

where the researcher provides a solution to a specific problem by applying knowledge with the aim of improving existing practice or application. Applied research cannot stand alone; it relies on basic research because it applies the scientific knowledge from basic knowledge in an existing practice [153].

The outcome of this thesis is a **basic research** once the main goal is to understand the impact of gender diversity in software development teams rather than providing a solution to a problem. The main contribution is the knowledge generated from the research.

Research Logic (Decision Point 2):

Research logic refers to in which direction the research proceeds in terms of whether it moves from general to specific or vice versa. There are two common ways of reasoning in empirical software engineering research: deductive versus inductive research [153].

Deductive research works from the more general to the more specific. It allows researchers to establish hypotheses by using theory. The researcher collects data to confirm or reject the hypothesis. Deductive research tends to lend itself to quantitative research as it aims to test a theory [153]. *Inductive research* is based on inductive arguments, and it moves from the specific to the general. The researcher infers theoretical concepts and patterns from observed data. The researcher begins with specific observations, detects theoretical patterns, and develops some general conclusions or theories. Inductive reasoning works from specific observation to a more general conclusion, which may lend itself to both qualitative and quantitative research[153].

This thesis follows an **inductive research** approach. The study intends to observe software development teams through surveys and historical data from repositories. From the data collection and analysis, we try to understand processes, people, and software development environment (specific observation), aiming to reach a more general conclusion.

Research Purpose (Decision 3):

The research purpose can be classified as exploratory, descriptive, explanatory, and evaluation research methods and data analysis methods [32].

Exploratory research is applied when there is not much information available in the topic area, and the researcher aims to gather some insights about the problem. The aim is to explore the problem area and provide background information that can be used for the descriptive or explanatory research [153]. *Descriptive research* is, as its name suggests, applied to describe a phenomenon or characteristics of a problem. It is more focused than exploratory research and goes further than exploratory research. The research questions aim to describe the phenomena [32]. *Explanatory research* is applied when examining the nature of certain relationships between the elements of a problem. The research questions

aim to explain the phenomena [32]. *Evaluation research* aims to determine the impact of methods, tools, or frameworks that may encompass the other three research purposes: exploratory, descriptive, and explanatory research [44].

For this thesis, the decision point regarding research purpose is a combination of **descriptive and explanatory research,** aiming to describe and explain the relationship between software development teams and gender diversity.

Research Approach (Decision Point 4):

Wohlin e Aurum [153] point to three kinds of approaches in their research decision-making structure: positivist, interpretivist (constructivist), and critical.

Positivist research advocates an objective approach and believes that research is reliable if it can be repeated and another researcher would reach a similar conclusion [88]. It tends to use quantitative methods. It tries to measure the world through empirical data, formal propositions, and quantifiable measures of variables, hypotheses testing, and the drawing of inferences about a phenomenon from a sample population [103]. Common methods are controlled experiments, surveys, and archival data analysis. [153]. *Interpretivist research* aims, also known as *construtivist*, aims to understand the human activities in a specific situation from the participants' perspective; hence it emphasizes the context [88]. It rejects the possibility of "objective" research and believes that research can be subjective. It assumes the behavior is influenced by the meanings people attach to events [103]. It aims to understand the deeper structure of a phenomenon within cultural and contextual situations where the phenomenon is studied in its natural setting and from the participant's perspective without including the researcher's prior understanding of the situation. It assumes that the validity of research can be gained by gathering qualitative data that is rich and in-depth [103]. It tends to use qualitative methods, e.g., interviews or ethnographies. An interpretive case study or a survey may also fall in the exploratory and descriptive categories as a research purpose [42]. *Critical research* aims to critically evaluate the existing system based on the assumptions that social and cultural variables impact the existing system and that the interconnections cannot be ignored. In critical research, knowledge is considered subjective, depending on whose perspective the researcher takes and whose eyes view the problem [21]. Critical research aims to reveal contradictions and conflicts within the existing system, while positivist and interpretivist research aim to predict or explain the current situation. Critical research often involves long-term historical studies of organization processes and structure. It tends to use qualitative methods and is likely to be a longitudinal study[103].

This thesis use an **interpretivist** approach, once its objective is to observe human aspects, more specifically, the impact of gender diversity, within software development teams, in a real environment.

3.2.2 Research Tactical Phase

The tactical phase involves decisions on how to operationalize the research activities in terms of how to approach the research question more specifically. The focus is to select the process and methodology to achieve the research goal [153].

Research Process (Decision Point 5):

In general, there are two widely recognized research processes called quantitative research and qualitative research. An alternative option is the combination of both qualitative and quantitative research, denoted as mixed research [36].

Qualitative research is a matter of inquiry that aims to study social and cultural phenomena. It is conducted when a researcher aims to understand the perspectives of their research subjects. The main idea is that by gaining access to the perspectives of insiders, researchers can also gain access to new ways of seeing the world [70]. Qualitative data refers to verbal descriptions by reflecting the world as seen by participants. Qualitative research involves the use of qualitative data collection such as interviews, written documents, and participant observation to understand and explain social phenomena. Qualitative methods are well suited for building theory, writing detailed descriptions, explaining relationships, and describing groups of norms, e.g., standards, models, and frameworks [153]. *Quantitative research* involves studies that refer to collecting quantitative data directly or cases where qualitative data is quantified to allow, for example, for statistical analysis [88]. The quantification of qualitative data is one form of a mixed research process. The objective is to describe the characteristics of the population and, in many cases, predict causal relationships. Quantitative research emphasizes using metrics, measuring with numbers, and analyzing data by using statistical techniques. A *mixed research* approach involves studies collecting both qualitative and quantitative data. Mixing may involve not only the type of research process, research methods, and data analysis methods [36]

Considering the ways data was collected and analyzed, this thesis uses a **mixed approach** in the research process. Further details will be given in Section 3.2.3 on the Operational Phase.

Research Methodology (Decision Point 6):

An essential part of research is the decision on research methodologies, which encompasses the combination of research methods, processes, and frameworks. For this work, we used a combination of three different methodologies: **systematic literature map, case study, and survey.**

3.2.3 Research Operational Phase

The operational phase research refers to the process of using data collection and data analysis when investigating a research question [153].

Data Collection Methods (Decision Point 7):

The data collection method depends on the research question [14]. Data collection methods may involve qualitative or quantitative data.

Lethbridge et al. [93] created a taxonomy for the data collection techniques where each technique is categorized according to the degree of human contact it requires. The first-degree contact requires direct access to a participant population. Second-degree contact requires access to participants' environment as they work, but without requiring either direct access to participants or for participants and researchers to interact. Finally, third-degree contact requires access only to work artifacts, such as source code or documentation [93].

Cost is a function of the effort required to collect the data, the record-keeping technique used, the amount of data produced, and the effort required to analyze the data [93]. In general, lower degree techniques are more expensive to use because they require more time and effort from researchers and study participants. Methods that produce more data require more time to analyze that data. Computer-based records are easier to analyze because software can be brought to bear on the data, more so than other data sources, such as videotapes. Humans tend not to be reliable reporters, as they often do not remember past events with a high degree of accuracy. Records of activities, such as tapes, work products, and repositories tend to be more reliable. However, care must be taken when interpreting these data sources as they may not be consistent, internally or with each other. Despite their drawbacks, first-degree techniques are invaluable because of their flexibility and the phenomenon they can be used to study. Existing logs and repositories are easy to use, but the data available is highly constrained. Software engineers, on the other hand, can be asked about a much wider range of topics [93]. Second-degree techniques lie somewhere in between. However, all of these techniques can still be used for exploratory questions. Finally, some contact with software engineers is necessary to find out what they think or feel. More removed techniques can only tell you what they were doing. However, this is not a problem if actions or work practice is the main interest of the study. Some inferences can be made regarding cognition from behavior, but they need to be confirmed with direct inquiries. In summary, each category of technique has its drawbacks, so it is necessary to use the appropriate combination to provide the data necessary to provide evidence to answer the research questions. First and second-degree techniques are more complex than third-degree techniques, but in many situations, this is an argument in their favor [93].

In this thesis, data were collected from different sources, and **first-degree techniques were used for the survey**, and **third-degree techniques were used for the case study.**

Data Analysis Methods (Decision Point 8):

Through the research methods, a lot of data may have been collected in qualitative or quantitative form. The data provide insight and evidence into the phenomenon studied. Once the data are collected, the researcher needs to analyze the data by using qualitative or quantitative data analysis techniques [153].

Thematic analysis is widely used as a qualitative data analysis technique in empirical software engineering research as it provides a deeper understanding of the data content. Braun and Clarke [19] describe thematic analysis as a method for identifying, analyzing, and reporting themes within data. The authors describe six phases of the thematic analysis process: familiarizing yourself with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the report. Thematic analysis generally involves open coding, where the codes are used to organize themes.

Statistics is used for analyzing quantitative data. Data can be analyzed using descriptive or inferential analysis [153]. The descriptive analysis involves summarizing data by describing and aggregating data and presenting associations between the constructs. Mean, median, mode, average, deviation, and variance are examples of methods used in the descriptive analysis as well as different types of plots[153].

The inferential analysis involves, for example, statistical testing of hypotheses, regression analysis and estimation using data mining techniques. Hypothesis testing is used to make inferences about a population. Regression analysis refers to methods that help to understand how changes in one variable affect another variable. Data mining is an automatic or semi-automatic approach to discern interesting data patterns using different statistical techniques such as cluster analysis [153].

In this thesis, the data collected through the **survey were analyzed qualitatively through thematic analysis**. To describe the Software Development Teams of the case study, we will present the data distribution and measures such as **mean, median, standard deviation, minimum, maximum**. Different cutouts for the number of team members, number of women in teams, number of men in teams. To explore the relations between gender diversity and pull-based metrics of the Software Development Teams we will use **regression analysis** and **hypothesis testing** on the regression coefficients of the simple linear regressions.

3.3 Ethics Committee

This research is with human beings, and the approval of ethics committees is a step not covered in the research design model by Wohlin and Aurum [153]. However, the approval of the research protocol in the Ethics Committee of the University is mandatory and we see it fitting as a step in the tactical phase.

The entire research project was submitted to the SIPESQ, the Research System of PUCRS, under the number 9926, and received approval in June 2020. After that, it was submitted to CEP, the Research Ethics Committee from PUCRS, under the CAAE number 33491720.0.0000.5336, receiving approval in June 2020, under number 4.110.771.

Appendix B presents the suggested research protocol where the steps for conducting the case study on the influence of gender diversity on the results of software development teams and the factors that can be established are established. Appendix C presents the survey protocol.

4. SYSTEMATIC LITERATURE MAPPING

In March of 2020 we ran a systematic literature mapping study on Gender Diversity, considering the period from January 2010 to March 2020. The objectives were to nail down which subjects related to software engineering are studied from the perspective of gender diversity and also if similar studies were conducted and could complement our empirical results. The Systematic Literature Mapping was conducted using the guidelines provided by Petersen et al. [107].

4.1 Research Questions

The goal of this systematic literature mapping study is to determine what is reported about gender studies in Software Engineering literature, focusing mainly on women. It leads us to the following research questions (RQs):

- *RQ1. Which is the frequency of publications in Software Engineering literature about Gender and Women?*
- *RQ2. Which venues are used to publish about Gender and Women in Software Engineering?*
- *RQ3: Which research topics have been most/less studied?*

4.2 Data Sources and Search Strategy

The research started with defining a proper scope, which was initially set to cover Software Engineering and Gender Diversity. It led to setting the preliminary mapping questions and identifying the keywords. The initial keywords were searched in well-known databases such as ACM Portal and IEEE Xplore. Based on the search results, the research scope, research questions, and keywords were refined, search strings were reformulated, and searches were re-conducted. Moreover, the list of databases was expanded to collect as many relevant papers as possible and to run the final string we included Scopus as well. In parallel, a list of control papers was generated, which was used as a validation list to ensure the reliability and relevancy of the searches and to evaluate the search strings.

4.3 Search String

The search string was formulated by combining variations for "Software Engineer" and "Software Engineering," gender, and also variations for "woman" and "gender." The variations were included after partially testing the strings and notice that we retrieve different results considering singular and plural. The final string is:

(woman OR women OR gender OR female) AND ("software engineering" OR "software engineer" OR "software development" OR "software developer")

Furthermore, some limitations were applied to the searches. The publication year was set to be between January 2010 to March 2020. This study was conducted in March 2020, so we have at least ten years of publications in the area.

4.4 Study Selection

We defined inclusion and exclusion criteria to perform the filtering in the results obtained by running the search string in the chosen databases as the following:

- Inclusion Criteria
  - Terms fulfill the search string or by similarity to the subject;
  - Book, Academic journal, conference and workshop papers;
  - Papers written in English;
- Exclusion Criteria
  - Keywords and Abstracts which do not focus on software engineering or related areas;
  - Papers which do not focus on software engineering or related areas;
  - Papers where the word diversity is not related to Gender Diversity;
  - Proceedings; Courses; Standards; Panels;
  - Format is not pdf.

4.4.1 Control Papers

We generated a list of control papers that were used as a validation list to ensure the reliability and relevancy of the searches and to evaluate the search strings. Researchers in the area previously indicated those papers:

- Hill et al., "GenderMag experiences in the field: The whole, the parts, and the workload," [77].
- Mendez et al., 2018, "Gender in Open Source Software: What the Tools Tell," [96].
- Wang and Redmiles, 2019, "Implicit gender biases in professional software development: an empirical study," [146].
- Carver and Serebrenik, 2019, "Gender in Software Engineering," [26].

4.5 Results

The result selection began from 921 papers returned by the databases after running the search string (IEEE Explorer: 338; ACM: 404; Scopus: 179). After applying the inclusion and exclusion criteria over the titles, abstracts, and keywords, the number of papers diminished to 223 (IEEE Explorer: 66; ACM: 104; Scopus: 53). In the second round, we eliminated the duplicated ones and also read the full papers (IEEE Explorer: 64; ACM: 7; Scopus: 55). Appendix A presents the final list with 126 papers.

4.5.1 Research Questions Answers

In this Section, we answer the research questions presented prior to this study.

RQ1. Which is the frequency of publications in Software Engineering literature about Gender and Women?

Figure 4.1 shows the frequency distribution per year from January 2010 to March 2020. The distribution indicates increasing interest in the subject, and 2018 and 2019 are the most prolific ones with 25 and 46 published studies, respectively. A highlight for 2018 is the occurrence of the First Workshop on Gender Equality in Software Engineering during the International Conference on Software Engineering (ICSE), contributing with seven papers in

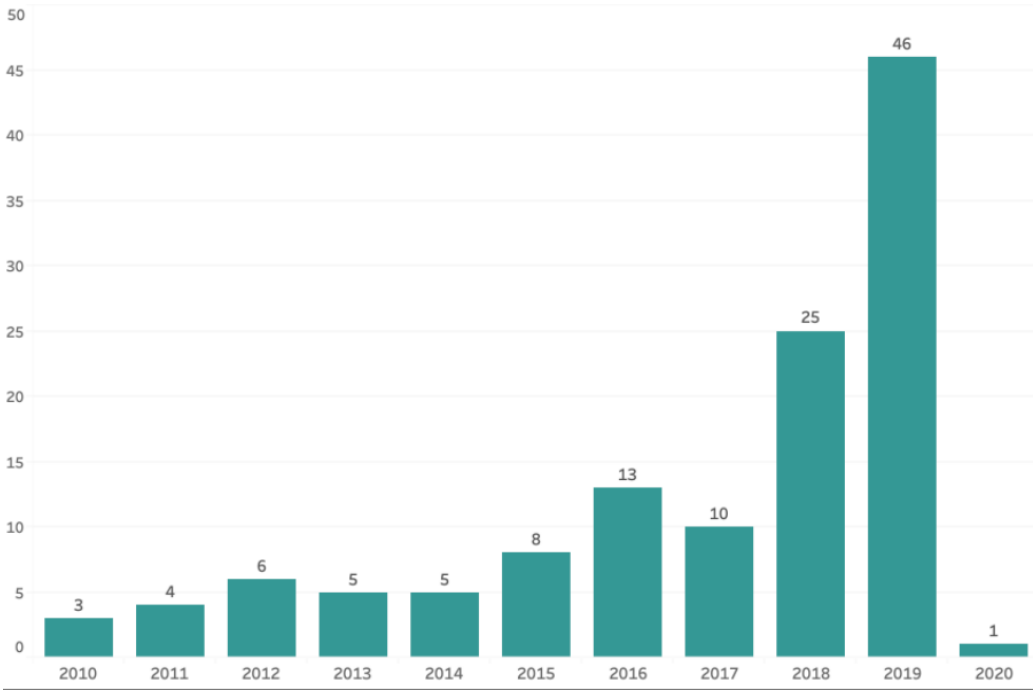


Figure 4.1: Papers per year - From January 2010 to March 2020

this systematic mappings. The same workshop contributed with ten papers in 2019. The International Conference on Software Engineering (ICSE) itself contributed with five published papers in 2018 and nine in 2019, the most significant number of contributions.

A hypothesis for the increasing interest in the research about the subject it is to follow the social trend. In the last years, society began to talk intensively about diversity, mainly in STEM (Science, Technology, Engineering, Mathematics) areas. Companies are publishing annual diversity reports. Marissa Mayer took the CEO position at Yahoo! in 2012 and Sheryl Sandberg, the COO position at Facebook in 2012, among other events.

RQ2. Which venues are used to publish about Gender and Women in Software Engineering?

We found out 65 publications about the subject in conferences, 20 in workshops, 10 in Symposiums. We also have 18 papers in journals, nine in magazines, three in lecture notes, and one book chapter. Table 4.1 shows the venues most target in which more than two papers about Gender and Women in Software Engineering have been published from January 2010 to March 2020.

Table 4.1: Conferences and Journals in which more than two papers have been published in the period of January 2010-March 2020

Type	Acronym	Conference/Journal Name	Published Studies
Conferences	**ICSE**	International Conference on Software Engineering	17
	ECSA	European Conference on Software Architecture	6
	FIE	Frontiers in Education	4
	GenderIT	Conference on GenderIT	2
	CHI	ACM Conference on Human Factors in Computing Systems	2
	CIbSE	Ibero-American Conference on Software Engineering	2
	EDUCON	IEEE Global Engineering Education Conference	2
	HICSS	Hawaii International Conference on System Sciences	2
	ICE/ITMC	IEEE International Conference on Engineering, Technology and Innovation	2
	IFIP AICT	Advances in Information and Communication Technology	2
	MSR	Mining Software Repositories	2
	SIGPLAN	ACM SIGPLAN International Conference on Systems, Programming, Languages, and Applications: Software for Humanity	2
Workshops	**GE**	Workshop on Gender Equality in Software Engineering (ICSE Workshop)	17
	CHASE	International Workshop on Cooperative and Human Aspects of Software Engineering (ICSE Workshop)	2
Symposiums	**VL/HCC**	IEEE Symposium on Visual Languages and Human-Centric Computing	4
	ESEM	Empirical Software Engineering and Measurement	2
Journals		Interacting with Computers	3
		Journal of Computing Sciences in Colleges	2
Magazines		IEEE Software	6
Books	**Book Chapters**		1

RQ3: Which research topics have been most/less studied?

We were also interested in which research topics have been the most or less studied in the published papers. Table 4.2 shows the topics which were extracted from the keywords, abstracts, and full text of the papers. Open Source Software (OSS) is on the top of the list with 28 studies. They go from the bias of the inclusion to the characterization of women's participation in OSS communities. Career and education also topics with a high number of papers, with topics fluctuating from career change and bias on the career evolution and education to construct inclusion since formation.

4.6 DISCUSSION

When formulating the search string, we considered not only specific words that refer to women (woman/women and female) but also the word gender. The idea was to comprehend how much research on gender in software engineering is relying only on women/female studies or if the studies are also going beyond the binary gender model.

From the 126 papers selected, three papers use word female, nine use the word women and seven, gender. Four mentions diversity more broadly, and one links the word

Table 4.2: Topics observed in the published papers

Topic	Published Studies	Studies
Open Source Software	28	P[9], P[12], P[14], P[19], P[26], P[29], P[30], P[33], P[35], P[36], P[37], P[46], P[48], P[51], P[54], P[55], P[59], P[67], P[92], P[96], P[98], P[102], P[107], P[108], P[113], P[117], P[120], P[126]
Career	22	P[7], P[8], P[15], P[16], P[22], P[23], P[31], P[32], P[41], P[42], P[44], P[45], P[47], P[57], P[63], P[64], P[70], P[80], P[83], P[87], P[91], P[110]
Education	20	P[4], P[5], P[6], P[18], P[20], P[24], P[49], P[50], P[52], P[60], P[69], P[71], P[73], P[85], P[88], P[100], P[101], P[104], P[111], P[123]
Team Composition	18	P[1], P[2], P[25], P[34], P[38], P[56], P[68], P[82], P[84], P[86], P[93], P[97], P[103], P[105], P[109], P[116], P[121], P[122]
Process	9	P[3], P[11], P[17], P[72], P[75], P[115], P[119], P[124], P[125]
Social Development	8	P[21], P[28], P[39], P[74], P[77], P[81], P[94], P[95]
Human-Computer Interaction	6	P[53], P[76], P[78], P[89], P[90], P[99]
Research	4	P[27], P[62], P[65], P[118]
Cultural	2	P[40], P[112]
Intersectionality	2	P[66], P[79]
Skills	2	P[13], P[106]
Social/Historical	2	P[43], P[114]
Emotions	1	P[61]
Performance	1	P[58]
Gender Prediction	1	P[10]

diversity to women only. However, 99 papers used the words gender, woman/women/female, indistinctly considering binary gender studies. It means around 79% of the final list.

Few studies from the selected list bring brought the discussion to the light. Izquierdo et al. P[36] mention the challenges of measuring gender diversity once gender is a complex social construct; most studies reduce gender to binary and employ heuristics to determine whether the individual is a woman or a man. Risks implied by this strategy are related to the inherent imprecision of the heuristics to determine gender and the marginalization of non-binary individuals. Kruger et al. P[10] also discuss the possible harm that automatic gender identification systems might entail due to their inaccuracy and also given that they are assuming a binary gender model. In a broader societal context, such systems have the potential of furthering the erasure of non-binary people.

Draude et al. P[65] say that gender studies refrain from common, simplified, often binary, heteronormative, reifications of gender. Also say gender itself should be understood in its relation to other social categories, such as race, ethnicity, sexuality, class, disability, considering the concept of intersectionality. Remedios et al. [115] say that intersectionality theory states that social identities are interdependent and that the meaning of identities derives from their relationship to one another. Only two papers from the mapping mention intersectionality.

Few papers mention non-binary genders, often as a threat to validity, that a simplified assumption that gender is binary was made for the paper. Wang P[74] made the simplifying assumption that gender is binary and mentioned that as the limitations and drawbacks of the study. Lin et al. P[95] cite the simplification assumption of the gender binary between the threats to validity. Ford P[81] used a tool that reports the gender of the user is male, female, unisex, or undetermined, and in their work, they reported females as women

and males as men. Vasilescu P[98] says that a team can reach its maximal gender diversity by having equally many female and male members when assuming a simplified, binary gender. Vandana P[46] briefly mentions non-binary data. Kohl et al. P[13] differentiate non-binary in the data presented in the paper as well.

Studies that talks about transgender software engineers are also rare. Ford et al. P[31] present a position paper that identified three themes that resonate across the trans experience and intersect with the advantages of working in software development remotely: identity disclosure, high-impact technical work, and the autonomy to disengage and re-engage. Singh P[48] talks about women-only spaces in the open-source software community. Spaces dedicated to providing a safe environment for women, free from judgment, abuse, hostility, and discrimination also explicitly welcome minorities and volunteer to be safe spaces for them as well (the paper mentions explicitly trans and ethnic or racial minorities).

4.7 CONCLUSION

The importance of considering the social and human aspects of Software Engineering is increasing, and discussions about gender diversity are gaining attention also in Software Engineering structures. We presented a systematic mapping of the literature from January 2010 to March 2020 about gender and women in Software Engineering. We identified a list of 126 qualified papers and evaluated them concerning how gender and women studies pervade Software Engineering. We found out that 79% of the published studies used the words gender, woman/women/female, indistinctly considering binary gender studies. However, few studies from the selected list bring the non-binary gender discussion to the light. There are challenges once gender is a social construct, and most studies reduce gender to binary (whether the individual is a woman or a man), risking a potential omission of non-binary people. Studies mentioning explicitly transgender individuals are also uncommon. A list of the topics found in the papers was provided as well.

5. CASE STUDY

The importance of collecting empirical evidence in studies involving human and social aspects in software engineering is known. The influence of gender diversity on the daily activities and deliveries of software development teams fits into this category of studies. What metrics, factors, and perceptions benefit or impact gender diversity in software development teams? This case study focuses on the following research question:

- **RQ1** What are the effects of gender diversity on software development teams when we analyzed pull-based metrics?

We study 14 software development teams of Brazilian's technology company, analyzing quantitative data from projects from their GitHub repositories using descriptive statistics and regressions. As mentioned in Chapter 2 w opted for using Pull Request Size and Pull Request Lead Time as the pull request metrics in this research, once they could be easily extracted from the repository data we had access. We were interested in the effects of gender diversity in the Pull Request Size (PRS) and Pull Request Lead Time (PRLT). We wanted to understand if they increase, decrease, or have no impact by a more or less gender-diverse team.

5.1 Data Collection

We selected 14 software development teams from a technology company from Brazil, representing around 10% of the teams with repositories in the company's GitHub. The selection of these teams took into consideration (summarized in Table 5.1):

- The number of members represented almost 88% of active developers of the company;
- The number of repositories touched by these teams, which represented almost 52% of active repositories;
- The activity in their repositories during the year 2020.

Table 5.1: Team Selection Filtering

	Total Number	Final Number	Covered
People	435	381	87.59%
Teams	142	14	9.86%
Repositories	238	123	51.68%

We analyzed data using descriptive statistical analysis and regression analysis for estimating the relationships between a dependent variable and one or more independent variables. We used Python scripts and GitHub APIs to extract data from the team's repositories and assembly a longitudinal dataset. We retrieved data from January 1_{st} 2020 to December 31_{st} 2020, so we have one year of data for analysis. We followed the steps below:

- Retrieve the identifiers of the selected teams from GitHub.
- For each team, we ran Python scripts which connected to GitHub APIs to retrieve:
  - The teams' members;
  - The repositories the teams used to work with;
  - The Pull Requests in these repositories in the period we considered;
  - The files modified in these Pull Requests

We stored the data from each team in .csv files for processing and filtering, which we will explain in the next Section.

5.1.1 **Preprocessing**

Anonymization

After retrieving the data, we anonymized the data. For the user login, we changed the user identifier using concatenation in the following format: *Gender_TeamNumber_UserId*, where:

- Gender: M (man), W (woman)
- Team number: from 01 to 14
- User Id: from 01 to *n*, where *n* is the number of team members

For the team name data, we changed the team identifiers using concatenation in the following format: *T_TeamNumber*, where the team number goes from 01 to 14.

Teams' Data

We included the total number of team members (#TeamMembers), the total number of men and women (#Men and #Women), and we calculated the *Blau Index* of each team, as follows in the following subsection.

Gender Diversity - Blau Index

We measure teams' gender diversity using the *Blau index* [17], which is defined as $1 - \sum_i^n p_i^2$ where p_i are the fraction of men and women team members. The Blau index (for categorical variables) is frequently used to capture how diverse groups are, i.e., the higher the measures, the more diverse team members are concerning a given attribute. For example, a team consisting only of same-gender members is not at all gender diverse; in contrast, a team can reach its maximal gender diversity by having equally many women and men members (assuming a simplified, binary gender), regardless of team size. Using the Blau index, gender uniformity is encoded as 0, while maximal gender diversity is encoded as 0.5 [141].

Pull Request Lead Time (PRLT) Preprocessing

Using Python scripts for the data of each team, we followed the next steps:

1. We got every Pull Requests opened after January 1_{st} 2020 and merged until December 31_{st} 2020;
2. We calculated the lead time for each Pull Request using the timestamp of the first commit on a branch subtracted from the timestamp on the merge action of the pull request: pullMergedAt - pullCreatedAt;
3. We cast the result to days, for simplicity;
4. We used Pandas, a library for data manipulation and analysis to calculate the statistics for each individual of the team: Min, Max, Mean and Standard Deviation;
5. Then we calculated the statistics for the entire team;
6. Last step, we grouped the Pull Requests of each team by gender (Women and Men) and calculated the same statistics.

Pull Request Size (PRS) Preprocessing

Using Python scripts for the data of each team, we followed the next steps:

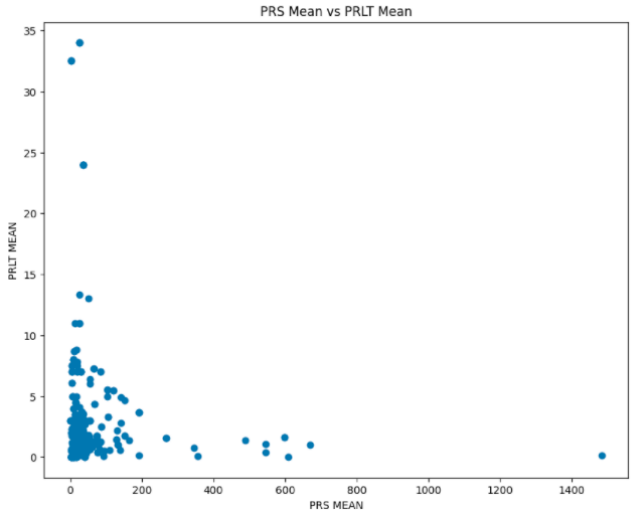
1. Once we could have multiple files modified in each pull request, we group them by the Pull Request ID, and then we summed the number of changes (additions + deletions) in each file, having the Pull Request Size.
2. We used Pandas, to calculate the PRS statistics for each individual of the team: Min, Max, Mean and Standard Deviation;

3. Then we calculated the PRS statistics for the entire team: Min, Max, Mean and Standard Deviation;

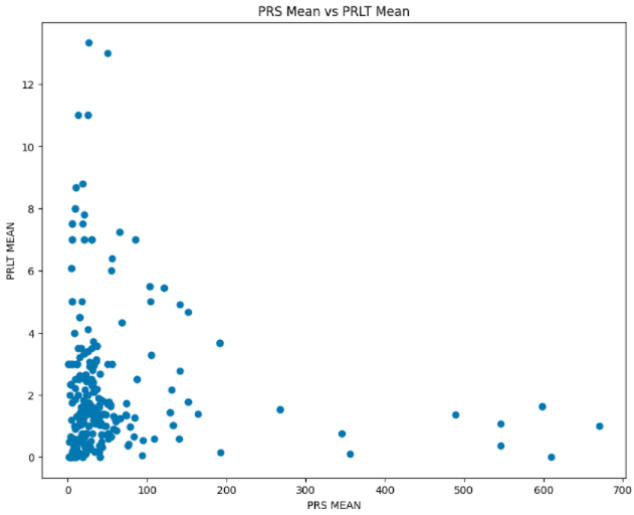
4. Last step, we grouped the PRS of each team by gender (Women and Men) and calculated the same statistics.

Outliers Removal

We used a visual approach to identify outliers. We plotted a scatter plot diagram showing us the relation between the number of pull requests and the number of days to merge and remove the outliers. Figure 5.3 shows the scatter plot before and after the removal of the outliers. We removed entries where the Pull Request Size was higher than 800 and Pull Request Lead Time higher than 15 days, the limits we visually identified were too different from most of the entries.



(a) With outliars



(b) No outliars

Figure 5.1: Dataset with and without outliars

Dataset Consolidation

We consolidated the data in two datasets and made it available in a GitHub repository [1]:

1. The first dataset has the results of the descriptive statistical analysis for each individual of the Software Development teams considered in the study. For each individual, we have the following data;
   (a) TeamId, Gender, UserId, #TeamMembers, #Men, #Women, BlauIndex, PRSMin, PRSMax, PRSMean and PRSStd, PRLTMin, PRLTMax, PRLTMean and PRLTStd
2. The second dataset is the data consolidation of each team, where we calculated the descriptive statistics over the data of the individuals of the teams;
   (a) TeamId, Gender, UserId, #TeamMembers, #Men, #Women, BlauIndex, PRSMin, PRSMax, PRSMean and PRSStd, PRLTMin, PRLTMax, PRLTMean and PRLTStd

5.2 Data Analysis

This section describes the variables we observed to test our hypotheses and answer our research question. Also, we describe the statistical modeling we used to analyze the data. We present the result in Section 5.3.

5.2.1 Measures

Dependent Variables

The *dependent variables (or response variables)* result from the experiment where the independent variable is manipulated. It is a factor whose variation is explained by other factors.

For this study, we measured as dependent variables the **Pull Requests Size (PRS)** and the **Pull Request Lead Time (PRLT).** The intention was to understand the effect of gender diversity of the teams on these variables.

Independent Variables and Control Variables

The *independent variable (or explanatory variables)* is the factor manipulated in an experiment by a researcher. It is used to determine the changes caused in the response

[1] https://github.com/karinakohl/ThesisData

variable. The *control variable* is anything held constant or limited in a research study. It is a variable that is not of interest to the study aims, but it is controlled because it could influence the outcomes.

To answer the research question, we considered **Gender Diversity** as the independent variable and the **Team Size** as a control variable. We measure teams' gender diversity using the *Blau index* [17], as mentioned in Section 5.1.1. The Team Size means the number of contributors in a team during 2020.

5.2.2 **Hypotheses**

From the literature, we know large pull requests carry more risk when deploying to production and are more challenging to review, merge, and release. Deploying pull requests of a reasonable size enables the team to review and ship new features at a faster cadence and with greater confidence [56].

The Pull Request Size can impact the time to merge. Depending on how many lines of code change, it requires more or less effort to review. However, developers tend to merge long pull requests faster. People get lazy to perform thorough reviews when too many things are going on. So, they immediately approve changes. There are two problems here: the time to merge goes up, and the quality goes down [68].

To answer our research question, we want to understand if there are relationships between the gender diversity in software development teams and the Pull Requests Size and Pull Requests Lead Time. So, we posit some null hypotheses (reported in Table 5.2) to test them via linear regression analysis. For simple linear regression, the chief null hypothesis is $H_0 : b_1 = 0$, and the corresponding alternative hypothesis is $H_1 : b_1 \neq 0$ [120], where b_1 is the slope of the independent variable, if positive this indicates that as X (independent variable) increases Y (dependent variable) also tends to increase, if negative, suggests that as X increases Y tends to decline. In Section 5.2.3 we will describe our statistical modeling.

Table 5.2: Hypotheses - Gender Diversity

Null (H_0)	**Alternate (H_1)**
$H_{1,0}$: Gender Diversity in a software development team does not impact the Mean Pull Request Size of the team.	$H_{1,1}$: Gender Diversity in a software development team impacts, increasing or decreasing, the Mean Pull Request Size of the team.
$H_{2,0}$: Gender Diversity in a software development team does not impact the Mean Pull Request Lead Time of the team.	$H_{2,1}$: Gender Diversity in a software development team impacts, increasing or decreasing, the Mean Pull Request Lead Time of the team.

5.2.3 Statistical Modeling

The Pythons scripts used for the descriptive statistical analysis and the regression analysis are available at GitHub repository[2] of this thesis.

Descriptive Statistical Analysis

Descriptive statistics are brief descriptive coefficients that summarize a given data set, representing the entire population or a sample of a population. We calculated the Minimum, Maximum, Means, and Standard Deviation for the Pull Request Size and Pull Request Lead Time of the teams. We used these results as inputs for the regression analysis, which we describe in the next Section.

Regression Analysis

Regression analysis is the study of how a response variable depends on one or more predictors [149]. Regression analysis is a well-known statistical learning helpful technique to infer the relationship between a dependent variable Y and independent variables X [6]. In regression analysis, you have your dependent variable — the main factor that we are trying to understand or predict, and then the independent variables — the factors you suspect have an impact on your dependent variable [55].

The Linear Regression Model is one of the oldest and more studied topics in statistics and is the type of regression most used in applications. Despite the fact that linear models are simple and easy to handle mathematically, they often provide an adequate and interpretable estimate of the relationship between X and Y [6]. A linear relationship can be positive (independent variable goes up, dependent variable goes up) or negative (independent variable goes up, dependent variable goes down). Additionally, it is not needed to assume normal distributions to perform regression.

To model the relationship between gender diversity in Pull Request Size (PRS) and Pull Request Lead Time (PRLT), we used Simple Linear Regression (SLR). In an SLR model, we build a model based on data, and we do not need the relationship between X and Y to be exactly linear. SLR models also include the errors in the data (also known as residuals). For modeling, we used the method of Ordinary Least Squares (OLS) supported by Statsmodel and scikit-learn, both scientific libraries for Python language. We segmented our data considering three sizes for the teams: small teams ($\leq$10 members), medium-sized teams (11 to 30 members), and a third for large teams ($\geq$ 30). For regression analysis of diverse teams, Vasilescu et al. [140] segmented team's data this same way, intending to

[2]https://github.com/karinakohl/ThesisData

avoid instability in the models and understanding that they do not expect that projects will respond to different team sizes in the same way, e.g., some types of projects may benefit significantly from the addition of a new team member, while others less so. We follow the same team size segmentation in this work. Table 5.3 shows the number of teams and members in these groups.

Table 5.3: Segmentation of Teams for Regression Analysis - Gender Diversity versus PRS and PRLT

	Size	**# of Teams**	**# of Members**
All Teams		14	381
Small Teams	$\leq$ 10 members	5	24
Medium Teams	10 $<$ members $\leq$ 30	5	77
Large Teams	$>$ 30 members	4	280

Our independent variable X is the Gender Diversity of the team, which is represented in our research by the Blau index. The index encoded gender uniformity as 0 (all men or all women) and maximal gender diversity as 0.5 (same number of men and women). Any number between 0 and 0.5 could represent more men or more women. In our data, we do not have any team with a higher number of women than men, so, as near the Blau Index is from 0.5, more women we have in a team, but never more women than men.

5.3 Results

5.3.1 Descriptive Statistical Analysis

Table 5.4 presents the descriptive statistical results for the teams analyzed. We used the results of the descriptive statistics as input for the regression analysis (mainly Team Size, Blau Index, PRS Mean, and PRLT Mean).

Table 5.4: Teams' Descriptive Statistics

TEAMID	TEAM SIZE	MEN	WOMEN	Blau Index	PRS Min	PRS Max	PRS Mean	PRS Std	PRLT Min	PRLT Max	PRLT Mean	PRLT Std
T01	4	3	1	0.3750	0	1123	20.5514	42.4750	0	14	0.5602	1.1480
T02	55	52	3	0.1031	0	42610	56.8284	687.3365	0	84	1.3438	3.7911
T03	4	3	1	0.3750	0	143	20.0000	27.9874	0	24	6.0000	8.2327
T04	9	8	1	0.1975	0	443	17.4964	38.0695	0	64	1.8163	7.2458
T05	4	4	0	0.0000	0	283	14.9719	32.1827	0	21	1.8235	4.5758
T06	13	10	4	0.3136	0	5098	30.9224	162.4681	0	32	3.2387	4.9979
T07	3	2	1	0.4444	0	1870	22.6939	63.0717	0	18	2.5248	3.4106
T08	44	37	7	0.2676	0	3184	21.9352	55.2879	0	69	0.9837	4.2717
T09	16	12	4	0.3750	0	15020	78.0544	437.8195	0	24	1.2601	3.0465
T10	23	20	3	0.2268	0	15020	62.4692	373.8955	0	25	1.3231	3.4339
T11	24	20	4	0.2778	0	42610	352.4524	2344.6089	0	20	0.8082	2.2727
T12	59	52	4	0.2186	0	36102	53.0490	585.0790	0	84	1.4517	3.9408
T13	62	54	8	0.2248	0	36102	48.3762	575.5634	0	69	1.4384	3.9043
T14	63	53	10	0.2671	0	42610	47.5278	604.1441	0	84	1.5075	4.0248

5.3.2 **Regression Analysis**

To analyze the results of our Linear Regression Model, we considered the following:

- $P - values$ and coefficients in regression analysis work together to tell you which relationships in your model are statistically significant and the nature of those relationships. The $p - values$ for the coefficients indicate whether these relationships are statistically significant [52]. Typical values for $p - values$ are 0.1 (10%), 0.05 (5%), and 0.01 (1%). These values correspond to the probability of observing such an extreme value by chance [92]. Fisher [48] says if $p - value$ is between 0.1 and 0.9, there is certainly no reason to suspect the hypothesis tested. If it is below 0.02, it is strongly indicated that the hypothesis fails to account for the whole of the facts. We shall not often be astray if we draw a conventional line at 0.05. Considering the first affirmation, in this research, we opted for using $p - value < 0.1$.

- The *coefficients*, intercept (b_0) and slope (b_1), describe the mathematical relationship between each independent variable and the dependent variable (positive or negative). The slope indicates the steepness of a line, and the intercept indicates the location where it intersects an axis. In the equation: $y = b_0 + b_1 x$. The greater the magnitude of the slope, the steeper the line and the greater the rate of change.

- $R - Squared$ is a goodness-of-fit measure for linear regression models. This statistic indicates the percentage of the variance in the dependent variable that the independent variables explain collectively. $R - Squared$ measures the strength of the relationship between the model and the dependent variable on a 0 – 100% scale [52].

5.3.3 Gender Diversity and Pull Request Size (PRS)

To understand the relationship between gender diversity and PRS, first, we ran the Linear Regression Model for the entire dataset, considering all the teams without considering their sizes. We did not observe statistical significance to reject the null hypotheses. A low $p - value$ (< 0.1) for the coefficient related to the gender diversity index (b_1) would indicate we could reject the null hypothesis, meaning the gender diversity impacts the PRS. The $p - value$ = 0.468, and when a $p - value$ is greater than the significance level, it is not possible to conclude there is a statistically significant association between the dependent and the independent variable. More than that, the $R - Squared$, the value which tells how much variation is explained by the model or how well the regression model fits the observed data, is very low (0.001), meaning that the model explains only 0.1% of the variation. So, the model does not explain much variation of the data, and it is not significant (worst scenario).

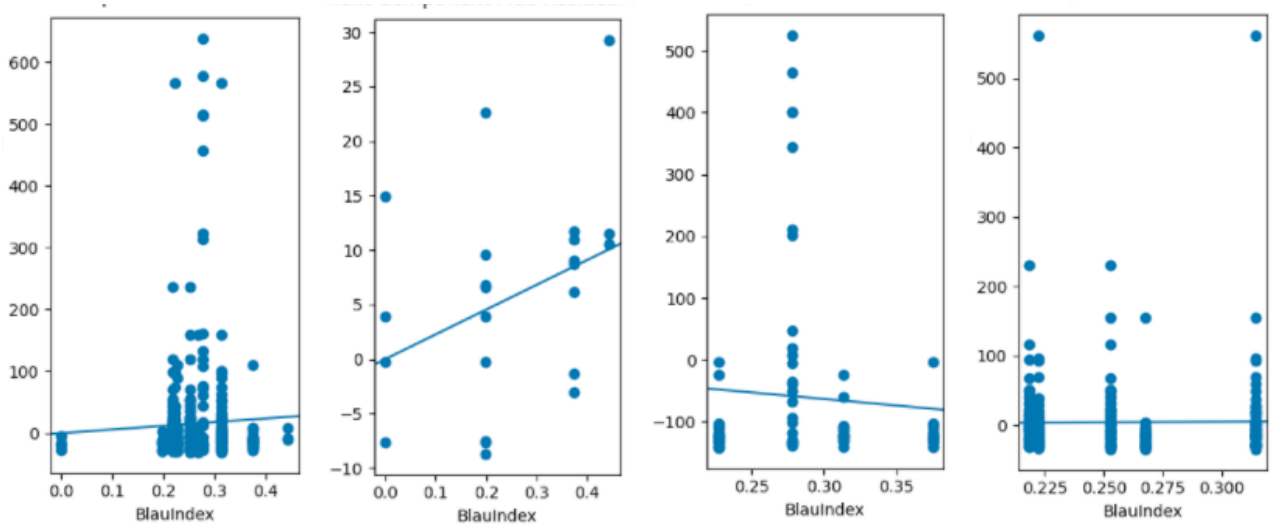
Then, we segmented the data, splitting it into three groups: small teams ($\leq$ 10 members), medium-sized teams (11 to 30 members), and large teams ($\geq$ 30 members). Table 5.5 shows the results for linear regression comparing the independent variable of gender diversity with the Pull Request Size for all the scenarios.

Table 5.5: Linear Regression - Diversity Index vs Pull Request Size (*$p - value < 0.1$)

	ALL TEAMS	SMALL TEAMS	MEDIUM TEAMS	LARGE TEAMS
R-Squared	0.001	0.126	0.006	0
Adj. R-Squared	-0.001	0.085	-0.007	-0.004
b0 (const) intercept	31.9583	11.6962	144.5924	36.1284
Std Err b0	21.83	3.771	93.041	26.929
p-value b0	0.144	0.005	0.124	0.181
Coeff (b1) slope	59.1615	22.6907	-211.4858	14.9276
Std Err b0	81.508	13.023	316.082	104.41
p-value (Blau)	0.468	0.096*	0.506	0.886

When running the model for **small teams**, we had statistical significance with a $p - value$ = 0.096. Then, for this segment, we rejected the null hypothesis $H_{1,0}$: *Gender Diversity in a software development team does not impact Mean Pull Request Size of the team*. It means that changes in the independent variable (gender diversity) are associated with changes in the dependent variable (PRS) at the population level. Evaluating the $R - Squared$, we can say the model explains 12.6% of data variance. We have a positive slope (b_1 = 22.6907), meaning when the diversity of the team increases, the Mean PRS also increases. Figure 5.2b shows the scatter plot that graphs the linear relationship between gender diversity and the mean PRS for small teams.

When running the model for **medium teams** and **large teams**, we did not observe statistical significance, considering $p - value < 0.1$. Then, for these two segments, we did not reject the null hypothesis $H_{1,0}$: *Gender Diversity in a software development team*



(a) All Teams (b) Small Teams (c) Medium Teams (d) Large Teams

Figure 5.2: Single Linear Regression - Gender Diversity Index vs PRS

does not impact the Mean Pull Request Size of the team. A $p-value$ that is greater than the significance level indicates insufficient evidence in the sample to conclude that if the diversity of the team increases or decreases, the Mean PRS also increases or decreases. Figure 5.2 shows the scatter plots that graph the linear relationship between gender diversity and the mean PRS for all the segments.

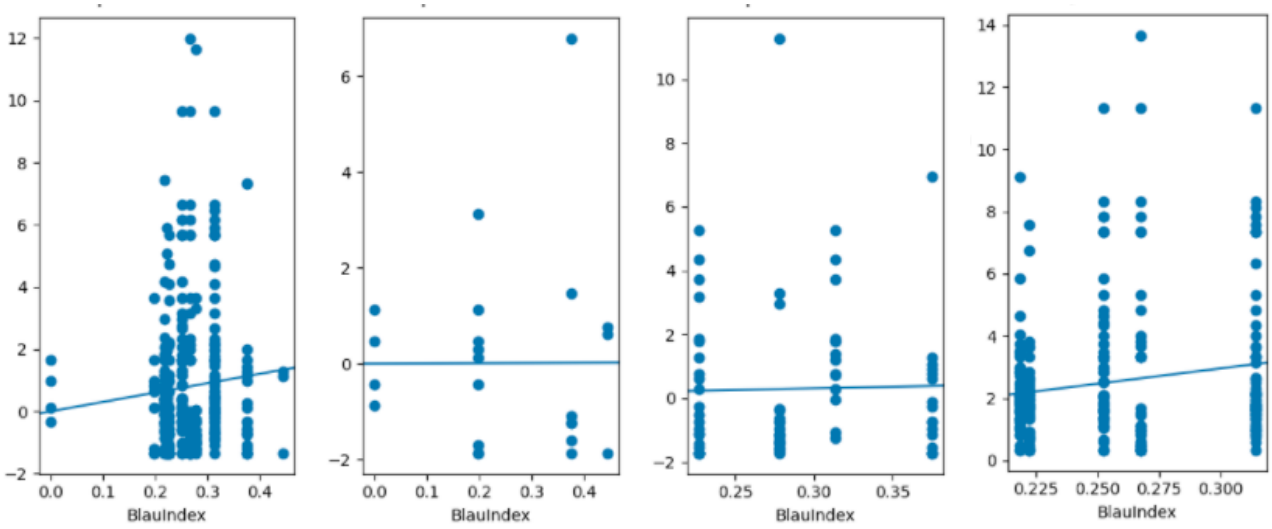
5.3.4 **Gender Diversity and Pull Request Lead Time (PRLT)**

As we previously did for gender diversity and Mean SPR, to understand the relationship between gender diversity and PRLT, we first ran the Linear Regression Model for the entire dataset, considering all the teams without considering their sizes. We did not have statistical significance to reject the null hypotheses. A low $p-value$ (< 0.1) for the coefficient related to the gender diversity index (b_1) would indicate we could reject the null hypothesis, meaning the gender diversity impacts the PRLT. The $p-value = 0.152$, and when a $p-value$ is greater than the significance level, it is not possible to conclude there is a statistically significant association between the dependent and the independent variable. More than that, the $R-Squared$, the value which tells how much variation is explained by the model or how well the regression model fits the observed data, is very low (0.006), meaning that the model explains only 0.6% of the variation. So, the model does not explain much variation of the data and is not significant (worst scenario).

So, again, we segmented the data, splitting the data into three groups: small teams ($\leq$ 10 members), medium-sized teams (11 to 30 members), and large teams ($\geq$ 30 members). Table 5.6 shows the results for linear regression comparing the independent variable of gender diversity with the Pull Request Lead Time for all the scenarios.

Table 5.6: Linear Regression - Diversity Index vs Pull Request Lead Time (PRLT) (*$p - value < 0.1$)

	All Teams	SMALL TEAMS	MEDIUM TEAMS	LARGE TEAMS
R-Squared	0.006	0	0.001	0.025
Adj. R-Squared	0.003	-0.048	-0.013	0.022
b0 (const) intercept	1.3419	1.8733	1.7238	-0.3319
Std Err b0	0.563	0.836	1.498	0.965
p-value b0	0.018	0.036	0.254	0.731
Coeff (b1) slope	3.0148	0.0413	1.0236	9.8512
Std Err b0	2.101	2.886	5.09	3.741
p-value (Blau)	0.152	0.989	0.841	0.009*



(a) All Teams (b) Small Teams (c) Medium Teams (d) Large Teams

Figure 5.3: Single Linear Regression - Gender Diversity Index vs. PRLT

When running the model for **large teams**, we had statistical significance with a $p - value = 0.009$. Then, for this segment, we rejected the null hypothesis $H_{2,0}$: *Gender Diversity in a software development team does not impact the Mean Pull Request Lead Time of the team.* It means that changes in the independent variable (gender diversity) are associated with changes in the dependent variable (PRLT) at the population level. Evaluating the $R - Squared$, the model explains 2.5% of data variance, meaning the model does not explain much data variation. Still, it is significant (better than not having a model). The analysis showed a positive slope ($b_1 = 9.8512$), meaning when the diversity of the team increases, the Mean PRLT also increases. Figure 5.3d shows the scatter plot that graphs the linear relationship between gender diversity and the mean PRLT.

When running the model for **small teams** and **medium teams**, we did not observe statistical significance, considering $p - value < 0.1$. Then, for these two segments, we did not reject the null hypothesis $H_{2,0}$: *Gender Diversity in a software development team does not impact the Mean Pull Request Lead Time of the team.* A $p - value$ that is greater than

the significance level indicates insufficient evidence in the sample to conclude that if the diversity of the team increases or decreases, the Mean PRLT also increases or decreases.

5.3.5 **Additional Analysis - Team Size and PRS and PRLT**

When observing the regression analysis results for gender diversity and PRS and PRLT, we hypothesized what other factors combined with gender diversity could be a complimentary analysis of the impact on PRS and PRLT. Different factors could be considered: age, social and educational background, tenure, etc. We choose to analyze the effect of team size once we already have the information in our dataset and we were familiar with the literature about it from other researchers as in [140, 141, 139]

We adapt our models considering team size as the independent variable and gender diversity as a control variable to perform this additional analysis. We also used Simple Linear Regression to model the relationship between the size of the software development teams and the PRS and PRLT. As we did previously, we segmented the analysis by splitting the data into three groups, one for small teams ($\leq$10 members), the second for medium-sized teams (11 to 30 members), and a third for large teams ($\geq$ 30). The hypotheses are presented in Table 5.7.

Table 5.7: Hypotheses - Team Size

Null (H_0)	**Alternate (H_1)**
$H_{3,0}$: The size of a software development team does not impact the Mean Pull Request Size of the team.	$H_{3,1}$: The size of a software development team impacts, increasing or decreasing, the Mean Pull Request Size of the team.
$H_{4,0}$: The size of a software development team does not impact the Mean Pull Request Lead Time of the team.	$H_{4,1}$: The size of a software development team impacts, increasing or decreasing, the Mean Pull Request Lead Time of the team.

Team Size and Pull Request Size (PRS)

To understand the relationship between team size and PRS, first, we ran the Linear Regression Model for the entire dataset, considering all the teams without considering their sizes. We did not have statistical significance to reject the null hypotheses. A low $p - value$ (< 0.1) for the coefficient related to the team size (b_1) would indicate we could reject the null hypothesis, meaning the team size impacts the PRS. The $p - value$ = 0.262, and when a $p - value$ is greater than the significance level, it is not possible to conclude there is a statistically significant association between the dependent and the independent variable. More than that, the $R - Squared$, the value which tells how much variation is explained by the model or how well the regression model fits the observed data, is very low (0.003), meaning that the model explains only 0.3% of the variation. So, the model does not explain much variation of the data, and it is not significant (worst scenario).

Then, we segmented the data, as we did for the linear regression of gender diversity, splitting the data into three groups: small teams ($\leq$ 10 members), medium-sized teams (11 to 30 members), and large teams ($\geq$ 30 members). Table 5.6 shows the results for linear regression comparing the independent variable of team size with the Pull Request Size for all the scenarios.

For **small teams**, we did not observe statistical significance, considering $p-value <$ 0.1. Then, for these segments, we cannot reject the null hypothesis $H_{3,0}$: *The size of a software development team does not impact the Mean Pull Request Size of the team*. A $p-value$ that is greater than the significance level indicates insufficient evidence in the sample to conclude that if the team size increases or decreases, the Mean PRS also increases or decreases. Figure 5.2 shows the scatter plots that graph the linear relationship between Team Size and the Mean PRS for all the segments.

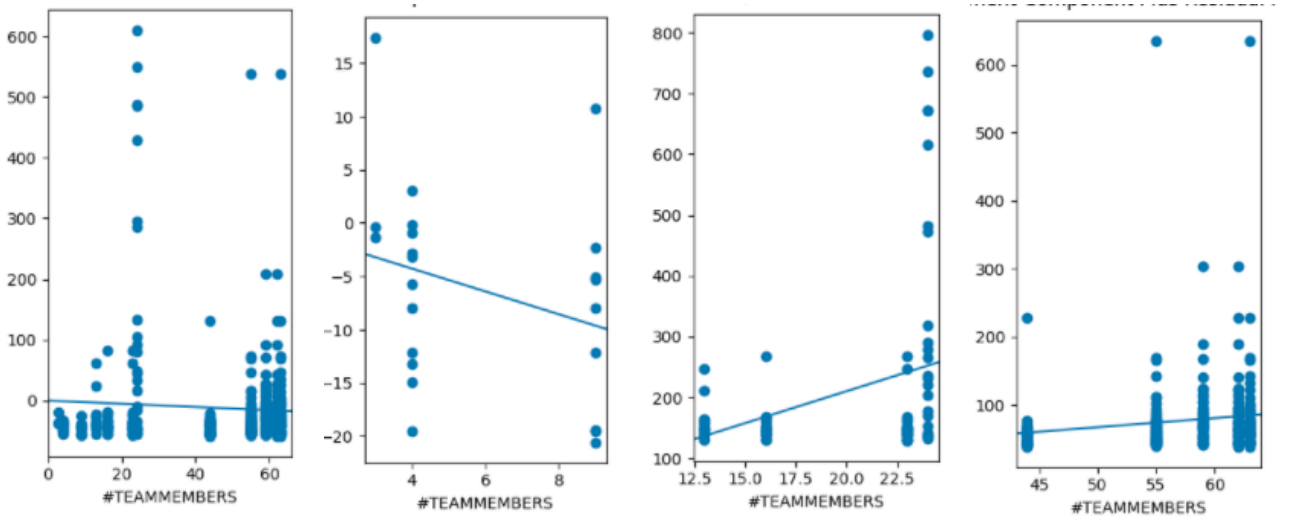
When we ran the model for **medium teams** and **large teams**, we had statistical significance with a $p - value$ = 0.005 and $p - value$ = 0.023, respectively. So, for these segments, we rejected the null hypothesis $H_{3,0}$: *The size of a software development team does not impact the Mean Pull Request Size of the team*. It means that changes in the independent variable (Team Size) are associated with changes in the dependent variable (PRS) at the population level. Evaluating the $R - Squared$, the model explains 10.3% of the data variance for medium teams and 1.9% for large teams. We also have had positive slopes (b_1 = 10.52 for medium and b_1 = 1.34 for large), meaning when the size of the team increases, the Mean PRS also increases. Figures 5.4c and 5.4d show the scatter plot that graphs the linear relationship between Team Size and the Mean PRS.

Table 5.8: Linear Regression - Team Size vs Pull Request Size (PRS) (*$p - value <$ 0.1)

	All Teams	**SMALL TEAMS**	**MEDIUM TEAMS**	**LARGE TEAMS**
R-Squared	0.003	0.085	0.103	0.019
Adj. R-Squared	0.001	0.042	0.091	0.016
b0 (const) intercept	59.7596	23.5896	-126.7859	-37.0129
Std Err b0	11.833	4.864	73.805	33.756
p-value b0	0	0	0.09	0.274
Coeff (b1) slope	-0.2641	-1.0701	10.5153	1.3387
Std Err b0	0.235	0.764	3.604	0.584
p-value (Blau)	0.262	0.176	0.005*	0.023*

Team Size and Pull Request Lead Time (PRLT)

To understand the relationship between team size and PRLT, first, we ran the Linear Regression Model for the entire dataset, considering all the teams without considering their sizes. We did not have statistical significance to reject the null hypotheses. A low $p - value$ ($<$ 0.1) for the coefficient related to the team size (b_1) would indicate we could reject the null hypothesis, meaning the gender diversity impacts the PRS. The $p - value$ = 0.403, and



(a) All Teams (b) Small Teams (c) Medium Teams (d) Large Teams

Figure 5.4: Single Linear Regression - Team Size vs PRS

when a $p - value$ is greater than the significance level, it is not possible to conclude there is a statistically significant association between the dependent and the independent variable. More than that, the $R - Squared$, the value which tells how much variation is explained by the model or how well the regression model fits the observed data, is very low (0.002), meaning that the model explains only 0.2% of the variation. So, the model does not explain much variation of the data, and it is not significant (worst scenario).

In this case, we also segmented the data, as we did previously, splitting the data into three groups: small teams ($\leq$ 10 members), medium-sized teams (11 to 30 members), and large teams ($\geq$ 30 members). Table 5.6 shows the results for linear regression comparing the independent variable of Team Size with the Pull Request Lead Time for all the scenarios.

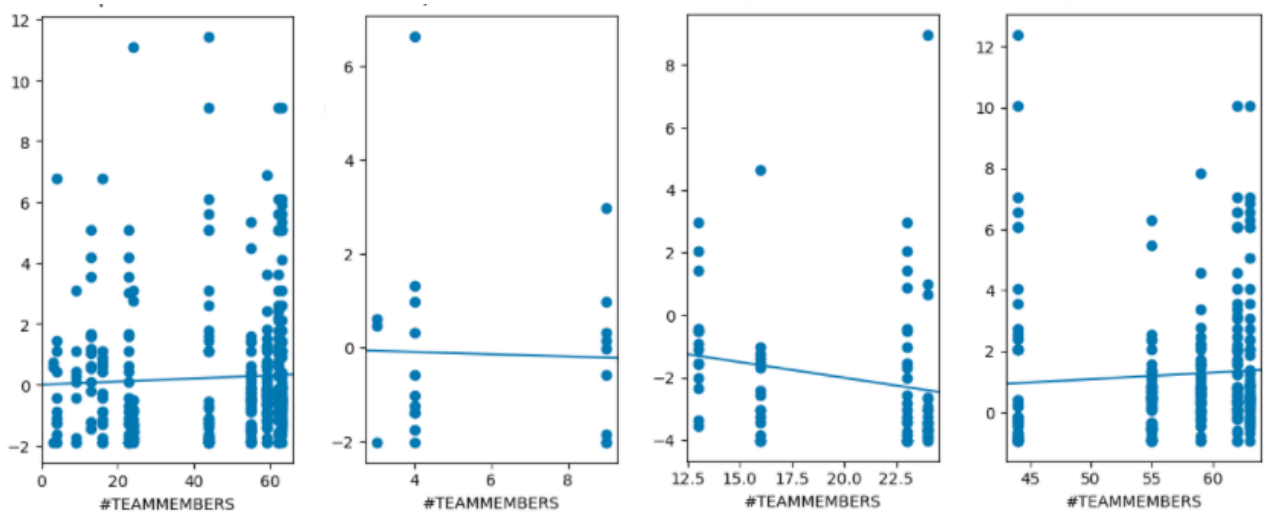
When we ran the model for **medium teams**, we had statistical significance with a $p - value$ = 0.099. So, for this segment, we rejected the null hypothesis $H_{4,0}$: *The size of a software development team does not impact the Mean Pull Request lead Time of the team*. It means that changes in the independent variable (Team Size) are associated with changes in the dependent variable (PRLT) at the population level. However, evaluating the $R - Squared$, the model explains only 0.4% of the data variance. We have a negative slope ($b_1 = -0.1001$), meaning when the size of the team increases, the Mean PRLT decreases. Figure 5.5c shows the scatter plot that graphs the linear relationship between team Size and the Mean PRLT. Figure 5.5d shows the scatter plot that graphs the linear relationship between team Size and the Mean PRLT.

When running the model for **small teams** and **large teams**, we did not observe statistical significance, considering $p - value < 0.1$. Then, for these two segments, we did not reject the null hypothesis $H_{4,0}$: *The size of a software development team does not impact the Mean Pull Request lead Time of the team*. A $p - value$ that is greater than the significance level indicates insufficient evidence in the sample to conclude that if the team

size increases or decreases, the Mean PRLT also increases or decreases. Figure 5.5 shows the scatter plots that graph the linear relationship between Team Size and the Mean PRLT for all the segments.

Table 5.9: Linear Regression - Team Size vs Pull Request Lead Time (PRLT) (*$p - value <$ 0.1)

	All Teams	SMALL TEAMS	MEDIUM TEAMS	LARGE TEAMS
R-Squared	0.002	0.001	0.036	0.004
Adj. R-Squared	-0.001	-0.047	0.023	0
b0 (const) intercept	1.895	2.0242	4.0209	0.9371
Std Err b0	0.306	1.053	1.229	1.234
p-value b0	0	0.068	0.002	0.448
Coeff (b1) slope	0.0051	-0.0241	-0.1001	0.0217
Std Err b0	0.006	0.165	0.06	0.021
p-value (Blau)	0.403	0.885	0.099*	0.31



(a) All Teams (b) Small Teams (c) Medium Teams (d) Large Teams

Figure 5.5: Single Linear Regression - Team Size vs PRLT

5.4 Discussion

From the 16 combinations analyzed (Gender Diversity vs. PRS and PRLT for different segments of Team Sizes and Team Size vs. PRS and PRLT also for different segments of Team Sizes), we observed statistical significance in four of them. We rejected the null hypotheses H_0 for those cases, meaning the changes in the independent variable were associated with changes in the dependent variable.

There is insufficient evidence in the samples to conclude that changes in the independent variable are associated with the dependent variable's changes. For the 12 others,

we did not observe statistical significance and could not reject the null hypotheses H_0. Table 5.10 shows a summary of the main findings from the regression analysis.

Table 5.10: Summary of Main Findings from Regression Analysis

Independent Variable		**Dependent Variable**		**Impacts**
Gender Diversity	↑	PRS	↑	Small Teams
Gender Diversity	↑	PRLT	↑	Large Teams
Team Size	↑	PRS	↑	Medium and Large Teams
Team Size	↑	PRLT	↓	Medium Teams

From the regression analysis results, we discuss the findings aiming to answer our research question in the following Section.

5.4.1 **RQ1. What are the effects of gender diversity on software development teams when we analyzed pull-based metrics?**

For small teams ($\leq$ 10 members), we rejected the null hypothesis that gender diversity does not impact Pull Request Size, meaning that changes in the gender diversity index of the teams have a relationship with the Pull Request Size (PRS). Figure 5.6 plots the teams considered in the segment of small teams, their diversity index (Blau Index), and the Mean PRS by gender for each team. Once the slope is positive, we can say that as gender diversity increases, the PRS also increases.

Team 05 does not have any women, so its diversity index is zero. Team 07 is the team with the highest diversity index (the most significant number of women in a team in this segment), and we can see the Mean PRS for women is higher than the Mean PRS of men. However, Team 04 has the smallest diversity index and considers the Mean PRS for women is higher than the Mean PRS of men. Team 01 and Team 03 have the same diversity index, and the Mean PRS for men is higher than the Mean PRS of women.

So, from the regression analysis, we see a tendency to increase the diversity of a team and increase the PRSs of the teams. Looking at the teams' data nearer, from a simplistic point of view, we can say that a minimal number of women for a small team is also not good for the Pull Request Size.

For large teams ($\geq$ 30 members), we rejected the null hypothesis that gender diversity does not impact Pull Request Lead Time, meaning that changes in the gender diversity index of the teams have a relationship with the Pull Request Lead time (PRLT). The slope is positive, so as gender diversity increases, the PRLT also increases.

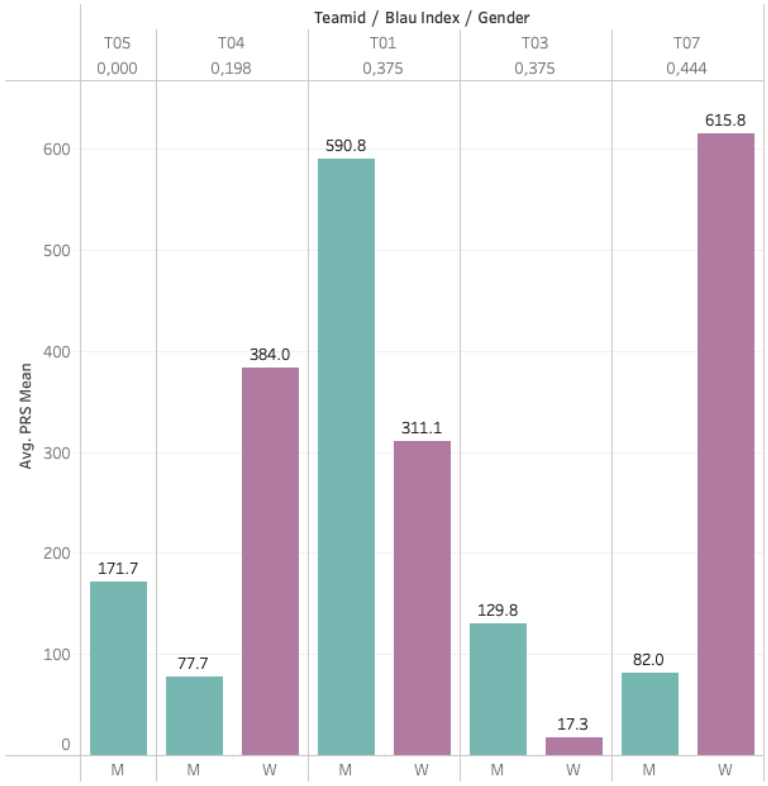


Figure 5.6: Gender Diversity and PRS Small Teams

Figure 5.7 plots the teams considered in the segment of large teams, their diversity index (Blau Index), and the Mean PRLT by gender for each team. Visual analysis of this segment's data shows that only for Team 12 are the women's Mean PRLT higher than men's Mean PRLT. This team is also the one with the smallest diversity index. For the other four teams, men's Mean PRLTs are higher than women's Mean PRLT. So, we can say that a more diverse team, in a large segment, increases the team's Mean PRLT, and this increase comes from the men's mean Mean PRLT.

We also analyzed the influence of the team size on Pull Request Size and Pull Request Lead Time. For medium ($\leq$ 10 members) and large teams (size between 11 and 30 members), we rejected the null hypotheses for Pull Request Size (PRS), meaning that changes in the team size have a relationship with the PRS. Once the slope is positive, we can say that as team size increase, the PRS also increase. Figure 5.8 plots the teams considered in the segments of medium and large teams, their sizes, and the Mean PRS for each team. For this case, the regression plots in Figure 5.4 give more precise visualization of the results.

For large teams (between 11 and 30 members), we rejected the null hypothesis that team size does not impact Pull Request Lead Time (PRLT), meaning that changes in the team size have a relationship with the PRLT. The slope is negative, so as the team size increases, the team's mean PRLT decreases. Figure 5.9 plots the teams considered in the segment of medium teams. In the left axis, we have the Mean PRLT of the teams and in the

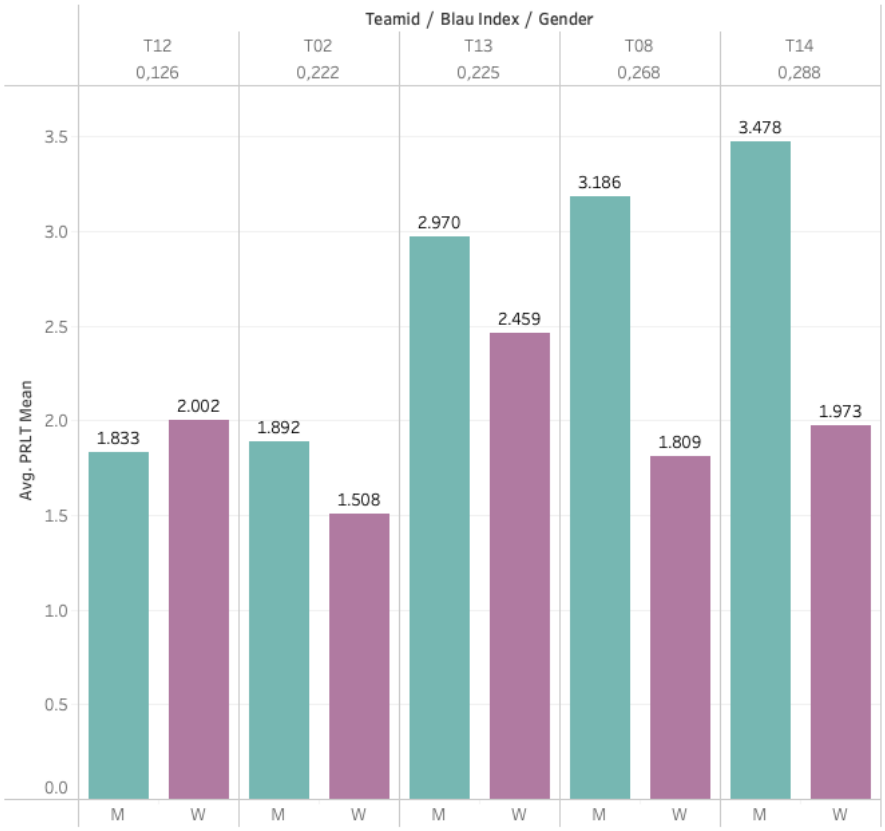
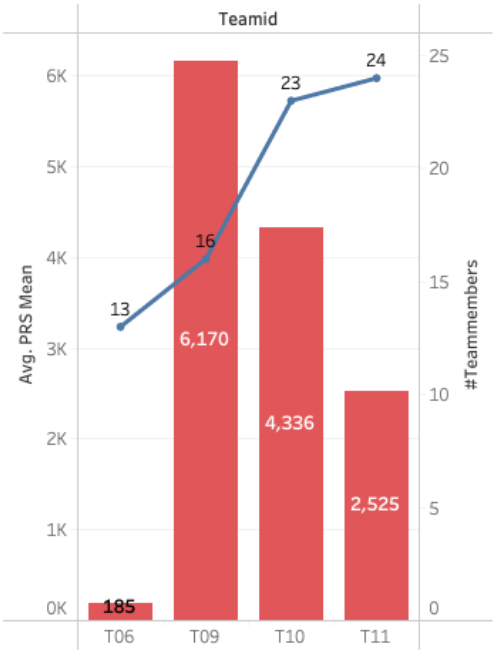
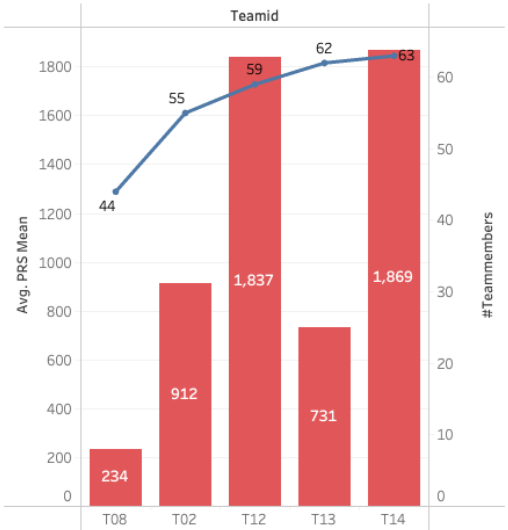


Figure 5.7: Gender Diversity and PRLT Large Teams



(a) Medium Teams



(b) Large Teams

Figure 5.8: Team Size and PRSs Medium and Large Teams

right axis, the team size, corroborating the regression analysis findings. As larger the team, the workload to review and approve pull requests is diluted between the team members.

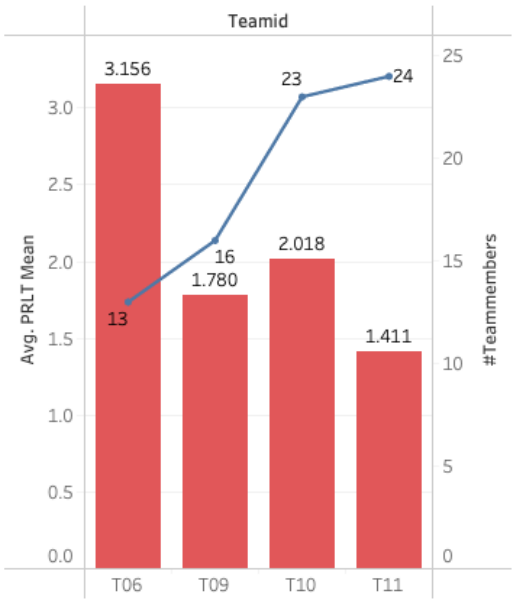


Figure 5.9: Team Size and PRLTs Large Teams

5.4.2 Conclusion

We were interested to understand the effects of gender diversity in the Pull Request Size (PRS) and Pull Request Lead Time (PRLT) of software development teams. We would like to understand if they increase, decrease, or have no impact by a more or less gender-diverse team. Also, we wanted to know if the team size factor, more than gender diversity, could have a different impact. To do that, we analyzed quantitative data from GitHub repositories of 14 software development teams from a technology company from Brazil.

From our analysis, we highlight the following findings that support answering our research questions:

- Increasing gender diversity in small teams, also increase their Mean PRS;
- Increasing gender diversity in large teams, increase their Mean PRLT. Using visual analysis, we observed that the men's Mean PRLT are more impacted;
- Increasing the team size in medium and large teams, increase their Mean PRS;
- Increasing the team size for medium teams helps decrease their Mean PRLT, probably diluting the workload through the team members.

These findings suggest that analyzing gender without considering other factors can be insufficient and lead to wrong conclusions about the impact of gender diversity in software development teams. For this thesis, we considered only team size as a comparison factor, once it was the one we could extract from our data. Other factors that could extend this study

are age, tenure, race/ethnicity, educational background, etc. Additionally, it is important to reinforce that the data of the 14 teams studied are from the same company. The study's replication to other companies can bring different and complementary results.

Some other studies bring insights that complement our findings. Vasilescu et al. [140] studied impacts of gender and tenure in OSS projects in GitHub. They expected that given a small overall number of women involved in OSS projects, the impact of gender diversity to be more visible for smaller teams. They found that gender and tenure diversity have significant, positive effects on productivity, across different team size segments, when controlled for other effects and, commit-tenure diversity has no significant productivity effect for smaller teams. Terrell et al.[132] performed a large-scale study on gender bias, where they compared acceptance rates of contributions (Pull Requests) from men versus women in the open-source software (OSS) community. Their results show that women's contributions in OSS tend to be accepted more often than men's. However, for contributors who are outsiders to a project whose gender is identifiable, men's acceptance rates are higher, suggesting that although women on GitHub may be more competent overall, bias against them exists nonetheless. Vasilescu et al. [138] performed a quantitative study about women's online community participation, e.g., StackOverflow. One finding was women use neuter names or "male profiles" to cope and be accepted by the predominantly male audiences. This is possible in online participation but not in industry teams. Antin e al. [7] research focused on differences in gender contribution in other kinds of virtual collaborative environments, mainly Wikipedia. They followed the activity of 437 contributors with self-identified genders on Wikipedia and found that men made more frequent contributions of the most active users while women made larger contributions.

6. SURVEY

The survey conducted in this research aimed to collect qualitative evidence to support answering the following research questions:

- **RQ2** What are the perceived benefits of gender diversity on software development teams reported by individuals?
- **RQ3** What are the perceived difficulties of gender diversity on software development teams reported by individuals?

6.1 Survey on Gender Diversity in Software Engineering

To study how people in software development teams feel about gender diversity in Software Engineering, we applied a qualitative, unsupervised, largely open-ended survey, using a web self-administered questionnaire. The approach was descriptive design, where our survey is case-control, which means that it is a retrospective: we asked participants about their previous circumstances to help explain a current phenomenon [108].

When we administer a survey, it is not usually cost-effective (and sometimes not even possible) to survey the entire population [108]. Instead, we survey a subset of the population, called a sample, hoping that the responses of the smaller group represent what would have been the responses of the entire group [108]. Our target population consists of people who identify themselves as part of a Software Development Team worldwide. We shared the web link for the survey in social networks to reach professionals of software development (e.g., LinkedIn and Twitter). So, we opted for the non-probabilistic sampling method called convenience sampling [108]. This kind of sampling involves obtaining responses from those people who are available and willing to take part, and we were aware that the main problem with this approach is that the people who are willing to participate may differ in important ways from those who are not willing [108].

The survey questions for this work were formulated based on the Empathy Map Canvas, so first, we would like to introduce the concept of Empathy. Decety and Cowell [40] say empathy is the ability to share in and understand others' experiences vicariously. Henschel et al. [73] say empathy corresponds to the ability to understand others' minds, feel their emotions outside our own, and respond with kindness, concern, and care to their emotions. Cameron et al. [25] suggest people may set the limits of empathy based on how hard they want to work. A study from Weisz and Zaki [150] suggests that people want to empathize with those most relevant to them. This tendency goes beyond group membership;

people are motivated to empathize with those who look like them, are kind to them, and those close to them.

In Human-Computer Interaction (HCI) discipline, empathy appears in different works, as in Bennett and Rosner [15] that studied empathy around disability. Gonzales et al. [58] used the empathy map as a tool to analyze Human-Computer Interaction in the elderly. Ferreira et al. [47] used personas to elicit requirements and Empathy Map to enhance the user experience. Graziotin et al. [66] say in the software engineering research arena, very few studies have been conducted into how developers feel or the consequences of their feelings.

Once we aimed to understand and describe how gender diversity impacts software development teams, we considered it essential to connect with the participants' experiences and emotions. So, we understood the Empathy Map Canvas could provide us a way to connect deeply with software engineering practitioners.

Dave Gray created the Empathy Map Canvas technique in 2009 as a tool that helps teams develop deep, shared understanding and empathy for other people. People use it to improve customer experience, navigate organizational politics, design better work environments, and host other things. The original goal of the Empathy Map is to gain a deeper level of understanding of a stakeholder in your business ecosystem, which may be a client, prospect, partner, etc., within a given context, such as a buying decision or an experience using a product or service [62].

Once we used the Empathy Map as our reference, our questions were open, and the respondents could frame their replies. A list of predefined choices was applied only for the demographic questions. Appendix C presents the protocol, the Empathy Map Canvas, and the full list of questions applied in the survey. The survey was built and distributed using Qualtrics[1].

Table 6.1 shows the response rate: the proportion of participants who responded compared to the number who were approached. Participation was voluntary and confidential, and it was estimated to take about 30 minutes. Using the mechanisms of the social networks, we could estimate how many people have access to the post/link for the survey, helping to estimate the response rate. The survey link reached around 1280 people, and 149 answered totally or partially. We had respondents who answered the survey partially. We evaluate that 60 respondents answered less than 20% of the questions, mostly only the demographics. We could not use these answer1 to extract information. In the end, we used answers from 88 respondents, corresponding to 6.88% of the people impacted by the post/link to the survey.

[1] https://https://www.qualtrics.com/

Table 6.1: Survey Response Rate

Linkedin + Twitter (How many people saw the link for the survey)	**Number of Respondents**	**Completude (Number of questions answered)**		**Discard**	**Final**	
		60-100%	20-59%	<20%		
1280	149	37	52	61	88	6.88%

6.2 Demographics

In this section, we present the demographics related to respondents. Table 6.2 presents the self-declared gender of respondents. We had 97.75% of the respondents self-declared as man or woman. Only one self-declared as Non-Binary. One respondent self-declared as Bisexual, which is a sexual orientation, not gender. Once it was self-declared and we did not have a way to identify gender, we considered it important to keep the information.

Table 6.2: Gender of Respondents (Self-Declared)

Gender	**Number of Respondents**	**%**
Man	44	49.44%
Woman	42	47.19%
Non Binary	1	1.12%
Bisexual	1	1.12%

Table 6.3 shows the age of respondents. On average, the respondents are 35 years old. The oldest one is 61, and the youngest, 23 years old.

Table 6.3: Age of Respondents (Self-Declared)

Gender	**Avg. Age**
Man	36.09
Woman	35.15
Non Binary	31
Bisexual	24
Average	**35.58**

Table 6.4 presents how respondents self-declared themselves in terms of race or ethnicity. In Merriam-Webster [147] we have the following definitions for race and ethnicity:

> "*The term race is understood today as primarily a sociological designation that identifies a group sharing some outward physical characteristics and some commonalities of culture and history, while ethnicity is a word for something you acquire based on where your family is from and the group which you share cultural,*

traditional, and familial bonds and experiences with. The end result: people may have racial similarity but ethnic dissimilarity.."

One respondent used the free text space to, instead of answer their race or ethnicity, to express their opinion on the subject: "*Don't be racist, skin color doesn't affect engineer skill.*" Their responses to the survey were meaningful, so we kept the answers, and we decided to give visibility on how this demographic question was answered.

Table 6.4: Race/Ethnicity (Self-Declared)

Race/Ethnicity	**Number of Respondents**	**%**
White	65	73.86%
Brown	4	4.55%
Black	5	5.68%
Not Informed	4	4.55%
Greek	2	2.27%
Latin	2	2.27%
Brazilian	1	1.14%
Chinese	1	1.14%
Brown	1	1.14%
Swedish	1	1.14%
Indian/Asian	1	1.14%
"Don't be racist, skin color doesn't affect engineer skill"	1	1.14%

6.2.1 **Professional Demographics**

This Section aims to be more specific on professional demographics. We had 79 answers for this question, and, on average, respondents work in technology for around 13 years (Max: 30 years and Min: 0.3 years). Table 6.5 presents the grouping of years in technology of the respondents.

Table 6.5: Years in Technology

Years in Technology	**Number of Respondents**
<10	36
11 - 20	31
21-30	12
Total	**79**

Table 6.6 shows the list of roles of respondents. Almost 51% of the respondents identified themselves as Developers. Table 6.7 presents the size of the companies the respondents' current work in terms of the number of employees. Table 6.8 shows how the teams are geographically distributed.

Table 6.6: Role of the Respondents

Role	Number of Respondents
Developer	40
Scrum Master	7
Product Owner / Product Manager	6
Project Leader / Project Manager	5
Engineering Manager	4
DevOps	3
Business Analyst	1
Data Analyst	1
Data Scientist	1
Director	1
Planning	1
Professor	1
QA	1
Research Engineer	1
Software Architect	1
Software Engineer	1
System's Analyst	1
UX	1
Team Lead	1
Trainee	1
Service Manager	1
Total	**80**

Table 6.7: Size of Company

Size of Company (number of employees)	Number of Respondents
1 - 10	6
11 - 50	13
51 - 250	9
251 - 1000	14
1001 - 2000	8
more than 2000	35
Total	**85**

Table 6.8: Teams' Distribution

Teams' Distribution	Number of Respondents
Distributed in National Territory	35
Globally Distributed	22
Locally	31
Total	**88**

6.3 Thematic Analysis

To analyze the survey data, we used *Thematic Analysis*, a technique used to gain a deeper understanding of the data content. Braun and Clarke [19] describe thematic analysis as a method to identify, analyze, and report on issues related to data. The thematic analysis process of this work followed the steps mentioned by Runeson et al. [117]:

1. Get the initial set of data. For this research: completed questionnaires;
2. Have the material studied in detail;
3. Formulate a set of codes of interest for the research, based on the research questions. Rely on another researcher and referenced literature to formulate the codes;
4. Read all texts and mark where codes fit into the content. Rephrase some codes if necessary: split the codes and create new ones if necessary. In this part, the process is iterative and, if new codes are formulated, the researcher will need to go back and re-code the material;
5. Use coded material to draw conclusions;
6. Compare text for different codes;
7. Compare different codes;
8. The process is iterative: there is the possibility to go back and adapt and change codes, sections, notes, and so on. It is also possible to go back and interview respondents again and identify new respondents if necessary.

To support this work, we used a software tool for qualitative and mixed methods research called MAXQDA[2]. First, we export the respondents' answers from Qualtrics in .csv format and import them in MAXQDA. The tool supports the navigation through the respondents' answers, allowing to mark up the text and code it, automatically saving the codes and allowing to reuse them. Through the process, we created 429 codes for 709 segments of the answers of the 88 respondents. Using the creative coding functionality of MAXQDA, which allows us to group the codes into themes, we grouped the codes into 18 themes. From these 18 themes, we identify that for five we could split them in sub-themes, better supporting to answer our research questions. Table 6.9 presents a summary of the themes and sub-themes. The entire list of themes, sub-themes together with the codes and frequencies can be found in Appendix D.

[2]https://https://www.maxqda.com/

Table 6.9: Summary of Themes and Sub-Themes

Theme	Sub-Themes
Benefits of Diversity	
Diversity as a whole	
Companies trying to be Diverse and Inclusive	
Men	"Bro" Culture
	Desconstruct Prejudice
	Men Characteristics
Women	Small number of women
	Women are disrespected
	Women Supporting Women
	Women Characteristics
No Diversity	Far Away from Diversity
	No Equity
	All Men
	Old Patterns
Leadership and Management Roles	
Sexism and Prejudice	
Professional Insecurities	
Companies do not support Diversity and Inclusion	
Missing Affirmative Actions and Initiatives	
Family and Personal Life	
Intersecctionality	
Meritocracy and Elitism	
Unawareness	
Sentiment and Emotions	Anger
	Anticipation
	Trust
	Fear
	Happyness/Joy
	Sadness
Overload	Causes
	Consequences
Financial Concerns	

6.4 Results

This section combined the themes we defined during the thematic analysis and the results of related work published by other researchers in the area to answer our research questions.

6.4.1 RQ2. What are the perceived benefits of gender diversity on software development teams reported by individuals?

Diversity is good beyond ethical reasons; it is recognized as valuable and, a lot of studies have been done about it [105]. Large technology companies have been creating annual reports of their efforts to have a more diverse workforce [59, 152, 97], increasing minority numbers through recruiting, working to minimize unconscious bias, and also investing in programs to increase representativeness [89].

From the answers of the survey, we extracted five themes respondents perceived as benefits of gender diversity, on Software Development Teams: *Benefits of Diversity, Diversity as a Whole, Companies trying to be Diverse and Inclusive, Men (Deconstruct Prejudice), Women (Women Characteristics and Women Supporting Women.*

In summary, for companies, the main gains and benefits are improving the hiring process, being more gender-inclusive, solid cultural codes, and encouragement of gender-neutral workplaces. As a result, companies reach better ideas sharing, better decision-making, and innovation. We saw evidence that men are touched by the subject, trying to deconstruct their prejudice and misconceptions about women in technology. Also, women try to support and inspire other women to remain in technology or enter the field. Characteristics as empathy, flexibility, and collaboration related to women.

Benefits of Diversity

Respondents see gender-diverse environments as more prone to innovation, creativity, and better decision-making.

> *"That having a healthy gender mix results in **better team working, better decisions and a better product**. A lot of people are tired of the "bro" culture around startups, and having a healthy gender mix ensures you don't end up with a laddish / pizza and beer feel to the team."* [Man]

> *"I talk about a health mix. Culture. **Better decision making**. Healthy working culture. Respectful language."* [Man]

> *"A team with more diversity tends to be even **more creative** and provide **better sharing of ideas, opinions, favoring innovation**."* [Man]

Nonaka and Takeuchi [101] emphasize that a team made of members with different backgrounds, perspectives, and motivations is critical for organizational knowledge creation to take place. Knowledge creation is the basis of repeatable innovation in companies. *"An*

organization's internal diversity must match the variety and complexity of the environment in order to deal with the challenges posed" [131].

Judy [85] says it is possible to create a workplace more tolerant and inviting to women if we recognize gender disparity as an impediment to achieving Agile. Agile values of collaboration, craft, quality, and delivery are a framework for addressing these factors.

Pretorius et al. [112] found that cognitive style was unrelated to feature novelty on its own. Neither a more intuitive nor a more rational dispositional style per se led participants to design a software feature of higher novelty. Gender, in contrast, was positively associated with feature novelty. They found that the female practitioners in their experiment produced more novel software features than the male practitioners did. Cognitive style and gender took together are also positively related to feature novelty. Female practitioners with a higher intuitive preference designed significantly more novel software features. Additionally, we found that female practitioners produced the most novel features when they had a preference for both intuition and rationality [112] .

Diversity as a whole

Respondents believe diversity can create a more democratic workplace. They try to give visibility to unconscious bias (the social stereotypes about certain groups of people that individuals form outside their own conscious awareness [98]) and have a neutral workplace.

> *"I can actively work and encourage a **gender neutral workplace**. I can make sure team activities are not exclusive in nature."* [Man]

> *"It's very much a positive focus on **diverse hiring**. Lots of internal **training to avoid unconscious bias**, etc."* [Man]

> *"I believe that greater diversity would make the environment **richer in ideas, dynamic and democratic**."* [Man]

Judy [85] says in a performing team, each member relates to the other as equals. A principled Agile team will not tolerate a hostile environment towards a teammate or the business people upon which it depends for work. Through the sometimes conflict-ridden process of team building, each member will come to tacitly understand and participate willingly in the unwritten norms of behavior that define the team's identity. The team will inspect its behavior and continuously improve the social skills required to communicate. It is a requirement for trust and the kind of collaboration that leads to cohesion and self-direction.

Companies trying to be Diverse and Inclusive

Some respondents work in companies that support gender diversity. They mentioned the culture codes of the companies as important tools to address ambiguous situations.

> *"The company I work **respects and accept the diversity of genders**."* [Woman]

> *"Our company has a **beautiful culture code** that encourages us to be the best person every day. If someone feels bad about a situation, it should be addressed to an anonymous channel, and HR will take care of it."* [Woman]

> *"**The company where I work seeks diversity and women in leadership**. Motherhood here is not seen as an issue, and there is an option to work part-time after the end of the maternity leave, which lasts from 15 to 18 months."* [Man]

A contested and consequential space where gender is legally not supposed to appear but does appear in the actions of recruiters and hiring managers is the technical job market. Especially for the younger women, gender is more visible as a factor in hiring than for the rest of the participants. Pozos and Friend [110], in their research, had respondents saying women have an easier time being hired because of their gender now that companies in Silicon Valley are feeling pressure to diversify their employee base. Recruiters can be both barriers and catalysts to finding tech jobs. However, there are also ambiguous situations, for example, when the participants are being recruited to change jobs. Gatekeeping happens implicitly, as for example, a woman was immediately directed to retail jobs at a high tech company by a male recruiter as she approached his booth at a career fair, and explicitly, for example, when another woman was offered an interview by a female recruiter just because she was a woman [110].

Tourani et al. [133] say in Open Source, a mixture of people with different cultures, personalities, and interests may increase the risk of offensive behaviors happening. Therefore, it seems essential for open source communities to protect their members from these kinds of unacceptable, destroying behaviors and provide a welcoming, safe, friendly, and inclusive environment where people can collaborate effectively towards presenting successful products. Common means for projects to achieve such an environment is the "code of conduct" concept. Such a code of conduct establishes ground rules for communications between participants, outlines enforcement mechanisms for violations, and tries to codify the spirit of a community. Anyone can contribute comfortably regardless of, e.g. gender, ethnicity, or sexual orientation.

Tourani et al. [133] say most codes of conduct aim to protect members from harassment, thus it seems important to understand what workplace harassment means. According to the Oxford dictionary, harassment is "Aggressive pressure or intimidation." As

such, workplace harassment is any offensive, belittling, or threatening behavior toward an individual worker or group of workers. It results in an unpleasant, humiliating, or intimidating environment employees feel uncomfortable in and consequently damages the effective work and productivity of employees [133].

Men

Respondents have different views about the role of men in supporting gender diversity. This theme was split in three sub-themes: *"Bro" Culture*, *Deconstruct Prejudice* and *Men Characteristics*. In this Section, to answer the perceived benefits of gender diversity in software development teams, the sub-theme we are interested is Deconstruct Prejudice.

Deconstruct Prejudice

Some men reported they are not only empathetic with the subject but they are also to identify their gaps to support better work environments. Some quotes follow:

> *"I am always willing to learn more and more. In the end, I'm a cisgender man, and with that, everything I say or think is projected through the search for information. I'm not in my place of talking about gender diversity, so* ***I always try to put myself in a learning position, deconstruction of my prejudices and construction of a space for diversity.****"* [Man]

> *"My workmates are mostly men. The younger ones are very conscious."* [Woman]

> *"I review my posture year after year because* ***I believe that my speech must be constantly criticized****. I do not want to incur in mistakes that I've already made."* [Man]

Viana et al. [142] analyze the stereotypes attributed to "egalitarian men," men who support gender equality in relation to domestic and family responsibilities as well as inclusion in the workforce. They found out the egalitarian man is perceived as fragile, sensitive, incompetent, and feminine. On the other hand, he is also seen as more competent and social than egalitarian women and traditional men [142].

Viana et al. [142] also hypothesised why are men so resistant to change when it comes to the sexual division of roles. Firstly, it could be the perception that more egalitarian relations between men and women may lead to the loss of certain privileges linked to the vision of a patriarchal society, in which it would be up to men to control economic, legal, and political institutions and, to women, to take care of the house and children and to satisfy their husband's sexuality. This division of roles endowed men with a structural power that gave them the primacy of the dominant group and made the family a privileged locus for the reproduction of patriarchal values concerning male superiority and female inferiority [142].

Women

This theme was split in four sub-themes: *Small number of women*, *Women are disrespected*, *Women supporting women* and *Women characteristics*. In this Section, to answer the perceived benefits of gender diversity in software development teams, the sub-themes we are interested are Women characteristics and Women supporting women.

Women Characteristics

Respondents identified some characteristics that are more prominent in women, as empathy, flexibility and collaboration.

"Empathetic, flexible, collaborative." [Woman]

"Work in an organized way." [Woman]

"Will to win." [Woman]

Curtis [37] proposed a predominantly theoretical model of influencing factors on programmers' performance containing intellectual aptitudes, knowledge base, cognitive style, motivational structure, personal characteristics, and behavioral characteristics. Turley and Bieman [134, 135] identified specific competencies of knowledge, personality, and attitudes as significant factors influencing software engineers' performance. Darley and Smith [38] say females pay more attention to details and disparate, multiple cues for information processing in simple and complex tasks.

Females are often motivated by their mentors, and their interest in the task [1]. Competitive behavior is rather demotivating [61].

Investigations of humans' characteristics contribution to task performance in software engineering often examine cognitive aspects. The influence of prior knowledge and or training in modeling skills is focused in several studies but did rarely consider personal characteristics as additional influencing factors on performance [145, 61].

Of course, software engineering requires problem-solving skills and analogical reasoning. Along with the cognitive-oriented studies in programming behavior in software engineering, studies rarely consider different personal factors and their impact on performance in software engineering tasks [61].

Nguyen-Duc et al. [99] another style category describes gender difference in terms of transformational and transactional leadership [13]. Transformational leaders are characterized as inspiring, motivating, being attentive to, and intellectually challenging their followers, where transactional leaders are described as contractual, corrective, and critical in their interactions with employees [13]. Female leaders were found to be more transformational than male leaders and also engaged in more of the contingent reward behaviors that are a component of transactional leadership [69].

Women Supporting Women

Respondents emphasized the importance of women supporting other women.

"My company supports diversity, ***I usually mentor women and people in social vulnerability****."* [Man]

"I'm quite introverted, and I don't have too much courage to speak, but I usually write a lot, and ***several people have already felt motivated by my texts and publications****."* [Woman]

*"****I always try to share and praise the women I admire****, and who motivated me to continue in the software development area."* [Woman]

*"****I try to indicate women [to work positions]****, I work in affinity groups for the inclusion of women, I give feedback when a colleague is sexist, I try to support other women."* [Woman]

"There are some very strong women, who are my ***inspiration****."* [Woman]

The relative absence of peers, mentors, and role models isolates women [85]. Research indicates that the social aspect of Agile practice, particularly routine face-to-face meetings and pair programming, reduces women developers' sense of isolation and raises their satisfaction and confidence [16, 85]. It also reduces feelings of internal competition and builds trust [72, 85]. Studies have shown that mentors can help women to motivate them and improve their self-confidence to achieve their goals Vidal et al. [143]

A recent report of UNESCO [136] highlights gender inequalities in the teaching of science, technology, engineering, and mathematics (STEM). The report also presents some kinds of interventions that help increase women's engagement with STEM education: supportive learning environments, mentoring, and role models [136].

6.4.2 RQ3. What are the perceived difficulties of gender diversity on software development teams reported by individuals?

From the answers of the survey, we extract 11 themes where respondents highlighted perceived difficulties about gender diversity on Software Development Teams: *No Diversity, Leadership and Management Roles, Sexism and Prejudice, Men, Professional Insecurities, Companies do not support Diversity and Inclusion, Missing Affirmative Actions and Initiatives, Intersectionality, Meritocracy and Elitism, Unawareness*.

In summary, respondents do not see diversity in their companies, but the same pattern repeating itself in the teams and the leadership: white, men, cisgender. More than

that, the pattern leads to male protectionism and privileges. Women suffer from sexism and prejudice, are frequently disrespected, and face difficulties thriving in their careers. Other dimensions of diversity pervade the answers as racism and ageism. Social vulnerability comes as a dimension related to meritocracy and elitism; once, at least in Brazil, equipment and training in Software Engineering are usually expensive.

No Diversity

For the No Diversity theme, we identify four sub-themes: *Far Away from Diversity*, *No Equity*, *All Men* and *Old Patterns*. All of them are complementary when answering the perceived difficulties about gender diversity in software development teams. Respondents from the survey mentioned few women in the teams (it is common to be only one) and no other gender than man. The same pattern repeats itself in leadership positions, when there are, most of the time, white men, cisgender (a person whose gender identity matches their sex assigned at birth.) These findings are aligned with the previous work of other researchers. The stereotype of a computer scientist is a middle-class white man who is often geeky and anti-social [45, 110]. While stereotypes are not necessarily representative of the general population, they do impact the perception of who belongs in the field and can act as exclusionary forces for people who do not fit the stereotype [30, 124, 110].

> *"In the teams I've worked, **I was the only woman**. I have always felt "a fish out of water" in every company because women are a minority or practically none."* [Woman]

> *"I say I'm out of patience. I'm tired of walking into meetings and seeing the **same pattern (white cisgender men).**"* [Woman]

Leadership and Management Roles

Respondents mentioned different characteristics of women that can be considered leadership ones: collaboration, conciliation, determination, empathy, flexibility, organized way of working and, will to win. However, respondents also mentioned the difficulties for women to foster their careers to leadership positions.

> *"I see in the company I work, an exact sample of the market: **women are occupying "authorized spaces."** It seems that I have been given the right to be where I am and that it is the part I deserve."* [Woman]

> *"With a lot of effort, we managed to advance some steps in terms of representation. However, **in more strategic positions, the participation of women is almost nil.**"* [Woman]

Jetter et al. [83] mention that leadership style, the "*manner and approach of providing direction, implementing plans, and motivating people*," has a significant impact on team performance and the achievement of organizational goals. Lead to highly functional teams and fostered team innovation. Gender is a factor that has been investigated in leadership style studies. It is sometimes claimed that women lead differently than men and are more collaboration-oriented, communicative, and less aggressive than their male counterparts.

Jetter and Walker [84] say it has almost become a stylized fact that, on average, women are more likely to avoid competition, under-perform in competitive environments, and exhibit higher risk aversion than men. Persistent social phenomena, such as the gender wage gap or the under-representation of women in highly competitive occupations and job positions, have been linked to such observations. One prominent hypothesis to explain this phenomenon relates to the idea that the gender of one's opposition could influence competitive behavior. More generally, people may behave differently when competing against adversaries from the opposite sex. If true, this would imply wide-ranging consequences in a number of settings. For instance, numerous work environments are characterized by persistent under-representation of one gender. Women are especially under-represented in jobs generally associated with high-pressure environments and large stakes, such as financial management (the share of females at Wall Street remains at approximately 10 percent) or CEO positions in the US (2.5 percent). Other areas with low female employee shares include IT- and math-related occupations, where women usually occupy less than 20 percent of positions [84] .

Eagly and Carli [69] say any female advantage in leadership style might be offset by disadvantage that flows from prejudice and discrimination directed against women as leaders. Prejudice consists of an unfair evaluation of a group of people based on stereotypical judgments of the group rather than its members' behavior or qualifications. When people hold stereotypes about a group, they expect it to possess characteristics and exhibit behavior consistent with those stereotypes. They also say, consistent with role incongruity theory, stereotype research reveals that people do consider men to be more agentic than women and women to be more communal than men. Also, the communal qualities that people associate with women, such as warmth and selflessness, diverge from the agentic qualities, such as assertiveness and instrumentality, that people perceive as characteristics of successful leaders. In contrast, the predominantly agentic qualities that people associate with men are similar to the qualities perceived to be needed for success in high-status occupations, including most managerial occupations [69].

Sexism and Prejudice, Men and, Professional Insecurities

In our survey, respondents reported male protectionism and privileges preventing women from advancing in their careers, so, we did not considered these themes as sub-

themes on one another, however, together, they lead to a chain of difficulties for women, as we list below:

- Women suffer from **sexism and prejudice**, that can be veiled and also it can lead to harassment.

  > *"Some time ago, my wife reviewed a code, and even though she was a senior,* ***their colleague said that he did not accept her review because she was a woman****. She complained, and the company did nothing. There was also the case that his manager called her to dinner at a one-on-one meeting."* [Man]

- There are two sub-themes from the theme **Men**, related to attitudes observed in men: the *Men characteristics* and the *"Bro" culture.* Both are related with what the respondents called the male protectionism.

  > *"A network of* ***male protectionism****, which prevents women from advancing in strategic positions (or advancing with great difficulty, even requiring a certain masculinization to do so)."* [Woman]

  > *"In previous companies I worked, it was much more difficult for women to have the same respect as men, or even the opinion to have the same weight as a man.* ***If they had a man and a woman in the same position, men preferred to hear a man's answer****, even if the woman was thinking the same thing or even has more experience."* [Woman]

- Women goes through **Professional Insecurities**. They feel they need to prove they are as capable as men, and it generates the feeling of never being competent enough, and their opinions are worthless and disrespected.

  > *"I often feel that* ***I need to go beyond my peers*** *concerning training like I'm never competent enough."* [Woman]

  > *"Get myself over the* ***mistreatment and disrespect*** *to keep doing the job."* [Woman]

Some other researchers identified what they call "*brogrammer culture*," a term that acts as shorthand for pointing to sexism in the tech industry. The term, "*brogrammer*," began as a satirical term to refer to a man who can code and succeeds with the behaviors of a stereotypical "frat-boy" and ambition to become rich fast [91, 110]. By definition, women are excluded from this group and often objectified and pushed out of workplaces because of the fraternity-like environments created as a result of *brogrammers* being in the space. Women who succeed in these environments may also face difficulties due to their gender, as

being better than men can be seen as threatening. In addition, how women perceive being gendered in male-dominated spaces indicates how they relate their experience to gender. Women's beliefs about the reason for the lack of women at high levels in companies, for example, influence whether they are motivated to fight structural barriers for other women or if they reify glass ceilings [29, 110]. Within *brogrammer* culture, there are often so few women that those who are there become accustomed to the environment [110].

Pozos and Friend [110] say the stereotype of a person passionate about computing implies that one has been programming and hacking for years before college. For several of the women in the study, coming "late" to Computer Science or not majoring in Computer Science is perceived as more of a barrier than gender because it is harder to be part of the Computer Science crowd without knowing the jargon, habits, etc. even when they have been involved with STEM fields before finding computing [110].

Buhnova et al. [23] shared the experience with a project assisting women in their 20s and 30s in changing careers towards technology. They identified four main barriers that women face when considering their transition into tech and computing, very much aligned with the findings of our survey: 1) They think it is not interesting (because they have a history of very low exposure to tech and computing); 2) They think they would not be good at it (Even if these women become equally or better skilled in tech than their male counterparts, they tend to keep higher levels of tech anxiety and lower levels of confidence in their tech skills); 3) Who they would be working with (the perception of tech working environment is not very good); 4) Missing guidance [23].

Companies do not support Diversity and Inclusion

From our survey, respondents reported that companies seem to have difficulty supporting diversity and inclusion.

> *"I see more of an incentive to diversity as a form of marketing positioning than as a structured and serious policy."* [Man]

Missing Affirmative Actions and Initiatives

From our survey, affirmative initiatives for diversity are scarce: too much talking and no action. More than that, the perception is that companies are hidden behind a facade of employers' branding initiatives that happen only from the door out.

> *"I see initiatives to promote greater diversity, but they are still shy in my point of view."* [Man]

> *"**Too much talk and little action**, non-recognition of women in development teams, the lack of looking at different aspects of diversity."* [Woman]

Simmonds et al. [123] observe that affirmative action programs to boost female enrollment in programs can have positive effects for science, technology, engineering, and mathematics undergrad programs; however, the initiatives yield weaker results for computer science/ software engineering majors.

Niederle et al. [100] talk about quotas as affirmative action in competition. They say quotas distort the objective probability of women winning the competition, but they also improve the confidence of women in their ability to succeed while reducing men's overconfidence.

Villeval [144] says affirmative action may thus ensure fairness if its main effect is to motivate talented but shy women to enter more frequently into competitive schemes. Still, men who think that affirmative action creates unfair competition may be less willing to cooperate with women after competing under affirmative action.

Intersecctionality

Intersectionality is a subject that permeates diversity. Intersectionality is an analytical framework for understanding how aspects of a person's social and political identities combine to create different modes of discrimination and privilege [35]. The term was conceptualized and coined by Kimberlé Williams Crenshaw in 1989 [35]. For some women, in particular women of color who are doubly and sometimes triply excluded because of their skin color and language practices, the discourse on gender may itself be harmful [34, 110].

Intersectionality is a complex and relevant approach that is little known in the scope of the Software Engineering research field. Respondents brought to the light they see the number of opportunities increasing for women, however, the initiatives seem to be focused on white women. More than that, they mentioned they have few black and transgenders workmates. Another face of diversity mentioned was ageism, the process of systematic stereotyping or discrimination against people because they are old, just as racism and sexism accomplish with skin color and gender [24].

> *"In the previous company,* ***I was the only woman*** *on the team I was on,* ***there was only one black man****, and diversity was not encouraged, not everyone had a voice in the company."* [Woman]

> *"Most black people have a history of more sacrifices to be able to gain good positions in the market."* [Man]

> *"Lack of opportunities due to* ***ageism****."* [Man]

> *"White men. Few girls and black people.* ***No transgender people."*** [Man]

Gordón and Palacios [130] say the diversity crisis is not limited to women, it is about social identities that go beyond gender and race, but it is mainly about power. There are cases in which software workers belong to two or more underrepresented groups. It is clear that systems of privilege and oppression often converge for underrepresented groups, i.e., there are organizational power dynamics that have historically privileged some groups and marginalized others in the Software Engineering field [8, 130].

Gordón and Palacios [130] say an intersectional approach invites Software Engineering researchers to read data in different ways and ask other questions that increasingly demonstrate the flaws of a race-only or gender-only approach.

Meritocracy and Elitism

From our answers, meritocracy and elitism are also understood as problems related to diversity.

Meritocracy is a social system in which advancement in society is based on an individual's capabilities and merits rather than based on family, wealth, or social background [86]. Nowadays, meritocracy is a subject with a lot of attention, however, the "*if you want you can*" speech opposes the lack of financial resources some individuals have to enter the software engineering field. The area is defined as elitist once equipment and training are usually expensive. If an individual does not have the means to be trained as a software developer, the discourse of meritocracy does not apply.

> *"**I see a lot that plasticized discourse of meritocracy**, defending that it is enough for a person to want what he can do."* [Man]

> *"**It is an elitist area. And the elite has one color and one gender**. The market would easily accept black or trans people with the appropriate technical skills, but why does the market have almost no such people? Because these people are marginalized, they are on the periphery, they are without access to our world. Equipment, courses, internet, everything is expensive, so only the same group, the same elite as ever, qualify well. Companies are more open to diversity, but to what extent will this bring a truly diversified team?"* [Man]

However, given the structural mechanisms that have played a critical role in women's exclusion from computing, the educational implications of this history are more challenging than current efforts imply [110]. As Hicks [76] cautions readers:

> "Despite the rhetoric of meritocracy, patterns like these will not be undone by the individual career choices of workers, especially if they belong to groups that lack the power to participate in the structures of dominance and control that created institutionalized discrimination in a given organization or industry in the first place." (p. 238)

Pozos and Friend [110] say it is not enough for educators to support the development of women's interest in computing. There is also a need to address entrenched social values and stereotypes at a systems change level. It will require rethinking how technical expertise is assessed (e.g., challenging the effectiveness of whiteboard technical interviews) and expressed and bridging the gap between entry-level and leadership positions [110].

Unawareness

The survey respondents reported that diversity does not matter or should not be considered when talking about software development teams. Accordingly to respondents, people should be treated equally, no matter their gender, and what must be considered is their ability, not their identity. Also, comments that the career is not for women.

> *"I've heard several people, mostly men (professors, college friends, lecturers), saying that **[diversity] it's not important, that this profession is not for women**. There are still a lot of people who are resistant to this issue."* [Woman]

> *"Some **people are just trying to take advantage of this**, focused on political issues and does not care about the real issues."* [Man]

> *"Do not classify humans in genders when you are focused on the team."* [Man]

> *"For me, **I don't think the gender of the person makes a difference**."* [Woman]

> *"There's lots of discussion about women in tech, but I still see very few applications from women for technical roles."* [Man]

> *"I see that technology is predominant male, but as other areas like psychology are predominant female, I don't see any problem with that."* [Man]

Alba [3] says gender equality does not mean pretending that "male" and "female" do not exist. Gender equality also does not mean that males and females must always be treated the same. In some cases, what is required is not equal treatment but equitable treatment. Equity means recognizing that differences in ability mean that fairness often requires treating people differently so that they can achieve the same outcome. At times equity is necessary to achieve gender equality, but there are many instances where this is not the case [3].

Alba [3] also says, most of the time, women and girls are at no inherent disadvantage due to a lack of ability that warrants differential treatment. Gender equality can often be achieved by holding everyone to the same standard. The problem is the irrational gender bias that women and girls are routinely subjected to [3].

Bias is a person or entity's inclination towards or prejudices against one thing, person, or group compared with another, usually in a way that's considered to be unfair [31]. A person or entity's bias primarily based on another person's gender is 'gender bias' [18]. There are primarily two types of biases:

1. Conscious bias (or explicit bias): Conscious bias is to be aware, intentional, and responsive. It is a tendency, trend, inclination, feeling, or opinion, which is particularly preconceived and certainly unreasonable [18];

2. Unconscious bias (or implicit bias): Unconscious biases are social stereotypes about certain groups that individuals form outside their own conscious awareness. Everyone holds unconscious beliefs about various social and identity groups, and these biases stem from one's tendency to categorize social worlds. Unconscious bias is far more prevalent than conscious prejudice and often incompatible with one's conscious values. Certain scenarios can activate unconscious attitudes and beliefs. There is evidence of widespread prejudice against women and girls from decades of psychological research [3, 18].

Bosu and Sultana[18] identified examples of unconscious biases that women often encounter in OSS projects:

- In our society, there is a trend of a belief that women have the instinct to do better or worse at certain tasks or have different interests.

- People often strain to argue that women are 'biologically' and 'culturally' good and trained at certain tasks such as communication, visual design, documentation. Therefore, women in some OSS projects are often forced to do certain activities like documentation, tutorial, or UI work even though they do not feel interested in them.

- Due to the lack of women in OSS, the community often behaves as an all-male environment. For example, desktop wallpapers, advertisements, and conference presentations are often decorated with some graphics and texts that are uncomfortable to the women counterparts.

6.4.3 **Other Findings**

Besides gender issues, some other subjects appeared and they permeate diversity issues. We believe the Empathy Map questions supported these subjects to appear once they connect with what people think, feel, say, and do. We can develop rich insights which help to identify genuine issues, problems, or concerns that the respondents may have. The subjects have touch points with diversity and gender issues. The themes that support this

section are: *Sentiment and Emotions, Overload, Financial Concerns, Family and Personal Life, What People Want, Needs.*

Sentiment and Emotions

The first theme we approach in this Section is Sentiment and Emotions. During the coding process, we observed the theme pervaded the respondents' answers.

Sentiment and Emotions have gained attention from both the research community and industry. This growing attention to emotion and sentiment can be attributed to a deepening awareness of the affective impact states have on work performance, and team collaboration [102].

Ekman et al. [43] obtained evidence for six basic emotions: happiness, surprise, fear, sadness, anger, and disgust combined with contempt. Plutchick [109] lists eight basic emotions, each primary emotion with a polar opposite so that: joy as opposite of sadness, fear as the opposite of anger, anticipation as the opposite of surprise, and disgust as the opposite of trust. When coding and grouping the codes in themes for this work, we use the combination of the list of emotions from both authors, Ekman and Plutchik [43, 109] as sub-themes.

Happiness/Joy: Feelings of contentment;
Respondents reported happiness, satisfaction, and joy as thoughts and feelings that motivate them. Also, happiness and joy were mentioned with other aspects of respondents' lives, such as family and financial stability.

"Happiness, health, and money." [Man]

"Happiness, social well-being and family well-being." [Man]

Trust: Feelings of peace and connection;
Trust was reported as related to trust workmates and management, impacting financial and career growth.

"Be trusted by the team and management. ***To have a good salary and be promoted.***" [Woman]

Fear: Negative feelings of fright and anxiety;
Fear was reported mainly related ,with the possibility of losing a job or being in the wrong place.

*"**Fear of being fired**."* [Woman]

"Fear of losing my job." [Man]

Sadness: Negative feelings of loss or disappointment;

Respondents reported sadness for the small representation of women in technology and its struggle.

> *"**Sometimes discouragement, sometimes sadness.** But sometimes, even the small representation gives me hope to continue opening the way for more women in technology."* [Woman]

> *"A lot of questioning and nothing is simple to solve**. I feel tired** because I think a lot before proposing something because I want to make a difference. As I said at the beginning, I feel that I need to take a stand and always make contributions."* [Woman]

Anger: Feelings of frustration;

Respondents reported feeling frustrated when they realized that diversity is not important for some people.

> *"**I get frustrated** because for many people diversity is not important, making it difficult for other women to enter this area (because they act as if that is not their place, not only making women not want to enter this area, but also giving up)."* [Woman]

Anticipation: Feelings centered around waiting for an event;

We considered stress and mental strain as evidence of anticipation. Stress is our body's response to pressure [51]. Strain is the state of a system on which excessive demands are made, as muscular strain and psychological strain, usually due to an emotional overload [9].

> *"**Mental strain** from performing tasks from different contexts."* [Woman]

> *"**Stress**. Too much to do and too little time and/or energy."* [Woman]

Surprise: Positive feelings of shock;

We did not identify feelings of surprise coming from the respondents from the survey.

Disgust: It relates to feelings of avoidance;

We did not identify feelings of disgust coming from the respondents from the survey.

Focusing on the negative may already be intuitive to many developers. It is a common occurrence that developers share horror stories about their working experience [66]. Graziotin et al. [65] link happiness, software quality, and developer productivity. It is an attractive and important endeavor to understand happiness and unhappiness in all its components – positive and negative emotions and moods. Scholars in industrial and organizational psychology have suggested that understanding happiness and unhappiness could

lead to cost-effective ways of enhancing working conditions, job performance and limiting the occurrence of psychological disorders [65]. Graziotion et al. [64] say unhappiness's highest impact is experienced on programming productivity as expressed by cognitive performance, including creativity and flow and process-related performance. They found several job-related adverse effects and even indications of mental disorders: work withdrawal, stress, anxiety, burnout, and depression.

Novielli and Serebrenik [102] say awareness of mood during work on a project and how affective states are reflected in our communication style might help developers work more effectively in teams, thus improving the outcome of collaborative development. Also, scaling up to the organizational level, information about the developers' emotions can be used to assess the effectiveness of a software development methodology. For example, a company that implements agile development might interpret negative emotions, such as stress and frustration, as a sign that agile principles are not being applied correctly or that developers are not aligned with such principles, thus increasing the risk of developer burnout and undesired turnover [102].

From this theme, we can observe happiness and trust pointing to well-being, remuneration, and career growth. Fear and anticipation appear related to the fear of being fired, which could lead to being productive and working on too many tasks and small deadlines. If evidence from different researches on the subject suggest positive mood and emotions pay off in terms of performance [67], from our survey and previous researches, we can assume that a wealthy life supported by a diverse and psychologically safe work environment could bring benefits for both, employees and companies.

Overload

Work overload was also a theme that permeated the answers from the respondents of the survey. We identified two themes: causes and consequences of overloading. Respondents reported things as interruptions, lack of focus and too many tasks as causes for overload. The consequences reported are mainly related to low productivity, low quality and delayed deliveries. Some quotes, follow:

> *"**Work overload, excessive interruptions,** few colleagues to count on for certain tasks, delays, poor quality of deliveries."* [Woman]

> *"**Overbooked**. A lot of priorities are to be completed in a short time. A lot of obstacles that avoid or delay the project."* [Woman]

> *"**Too much to do and too little time and/or energy**. Sometimes feels like a lonely uphill battle."* [Woman]

Zika et al. [156] say that their study indicates that there are associations between a high level of project overload and (a) high levels of psychological stress reactions, (b) decreased competence development, and (c) deviations from time schedules.

Claes et al. Claes2018a worked empirically in an approach to detect stress and overload from the point of view of abnormal working hours. They say poor working patterns can reduce individual health, well-being, and productivity.

Fujigaki [53] examined the mental health of software developers and found a statistically significant correlation between job events and an increase in depressive symptoms up to a week after the job event. The study did not differentiate between different job events, but instead, they included: time pressure of a deadline, work-overload, amount of work increase, responsibility increase, and trouble with clients.

As mentioned previously in the theme "Sentiment and Emotions," affective states have been shown to alter the productivity of software developers. Effects of unhappiness while developing software include low cognitive performance, low motivation, work withdrawal, low productivity, and low code quality [64].

Judy [85] says traditional, and startup software cultures celebrate long, hard hours crashing for a deadline. The modern global distribution of business requires communication and coordination with teams in vastly different time zones. While women can feel fulfilled and challenged by extreme work schedules, they lead to attrition from the industry [74] though not necessarily any more than for their male counterparts.

Financial Concerns

Respondents reported different concerns regarding finances, but we did not group them in sub-themes once the codes are highly related. They highlight the importance of earning what is fair. For women, fair is to earn the same as their male peers. They also express preoccupation with the financial stability to provide for their families and to have a sustainable retirement. Fear of unemployment was also mentioned.

> *"**I want piece of mind**, a sustainable retirement and traveling."* [Man]

> *"Will I get the same salary a man gets? **Will I receive the same respect and success as a man?**"* [Woman]

> *"**Earning well to help my parents**. So my father can stop working."* [Woman]

> *"**I fear losing my income** and ending up in a difficult situation."* [Man]

> *"I need a fair salary raise like my male colleagues."* [Woman]

> *"I want my own home and financial security."* [Woman]

Family and Personal Life

For this theme we did not split themes in sub-themes, once all the codes are highly related. From the respondents' answers, work-life balance and the importance of family appeared. Respondents want quality time with their families, and also they want to have a wealthy life to support them.

> *"Living away from home showed me how valuable it is to be with one's family. With the pandemic (Covid-19), I was able to go home and experience it again every day. Being in the family is what motivates me. My parents motivate me."* [Woman]

> *"I want to be able to* ***help my parents.****"* [Woman]

> *"Happiness and life quality."* [Man]

> *"I dream of giving a better life to my family."* [Man]

> *"****Work life balance.****"* [Woman]

Aboobaker et al. [2] say various workplace factors like total working hours, over time requisites, inflexible work schedules, unaccommodating peers, and an uncongenial organizational culture create conflict between the work and family roles. Conflict between work and family is associated with increased occupational stress and burnout, intention to quit the organization, lower health, and job performance, low job satisfaction, and performance, high absenteeism rates, reduced career commitment, increased psychological distress, increased parental conflict, and marital distress, increase in child behavior problems and poor parenting styles and lower satisfaction with parenting. Conflict between work and family is a significant factor that contributes to turnover intention [2].

Judy [85] says the industry needs to make a specific commitment to being more supportive of women workers, and one of these commitments is to support their employees' full range of interests and better balance work and family life. Machado et al. [118] say for women, motherhood represents a special and tense period to reconcile professional life with family responsibilities.

> *"I want low stress levels."* [Man]

> *"I talk about a health mix. Culture. Better decision-making. Healthy working culture. Respectful language. Understanding that people don't need to be open source contributors or work 18 hour days to be good developers. "* [Man]

6.5 Discussion

We conducted this survey to qualitatively understand the perceived benefits and difficulties of gender diversity in software development teams. The answers we had aligned with different related work performed by researchers from the Software Engineering community. We also found out studies from other knowledge fields, such as psychology, that studied gender issues in technology and corroborated our findings.

From a beneficial point of view, we have that gender-diverse workplace are prone to have better ideas sharing, better decision making, creativity, and innovation. Characteristics linked to women were highlighted as empathy, flexibility, and collaboration.

To achieve the benefits mentioned, respondents reported that some companies worked to improve the hiring process to be more gender-inclusive. To support and guarantee inclusion and safety, solid cultural codes were created.

There is mutual support from women to women. Women support and inspire each other to remain in technology or enter the field. More than that, some men reported being touched by the issue and diligently deconstructing their prejudice and misconceptions about women in technology.

However, there are also difficulties from the point of view of gender diversity in software development teams. There are still cases the respondents do not see diversity. It is common to see only one woman in teams or just a few. More than that, no other gender than men and women, so the white, cisgender man is the pattern most of the time. The same pattern repeats itself in leadership positions.

This reported pattern leads to an issue: male protectionism and privileges. Due to that, women report being frequently disrespected and face difficulties in thriving in their careers. Sexism and prejudice used to happen together.

But there are other dimensions of diversity that pervaded the answers, and intersectionality is mentioned. Intersectionality aims to study how aspects of a person's social and political identities combine to create different modes of discrimination and privilege. So, when racism is mentioned, we can exemplify when respondents brought to the light they see the number of opportunities increasing for women, however, the initiatives seem to be focused on white women, excluding the intersection of race and gender of black women. Ageism, the discrimination against people because they are old, was also mentioned. Some people feel they lost, or even do not have opportunities, due to their age.

A little less explored as a dimension of diversity, social vulnerability is related to meritocracy and elitism. The "*if you want you can*" speech opposes the lack of financial resources some individuals have to enter the software engineering field. In Brazil, equipment

and training in Software Engineering are usually expensive, so the entry point to the field is difficult for people in social vulnerability.

We also observed a lack of awareness about the need to talk about the subject. The survey respondents reported that diversity does not matter or should not be considered when talking about software development teams. Accordingly to respondents, people should be treated equally, no matter their gender, and what must be considered is their ability, not their identity. It would be the ideal scenario, however, due to unconscious bias that women and girls (and other dimensions of diversity) are routinely subjected to and the different other points we already mentioned in this work, we see it as important to keep an active awareness about the subject. The unconscious bias leads to people arguing that women are 'biologically' and 'culturally' good and trained at certain tasks such as communication, visual design, and documentation even though they do not feel interested in them.

We believe the Empathy Map questions were positive once additional and complementary subjects appeared, besides gender diversity, once it aims to connect with what people think, feel, say, and do. We retrieved rich insights which help to identify genuine issues, problems, or concerns that the respondents may have. The subjects have touch points with gender issues. Themes such as *Sentiment and Emotions, Overload, Financial Concerns, Family and Personal Life* appeared consistently through the answers.

7. DISCUSSION

In the context of empirical research on human and social aspects of software engineering, we understood there was an opportunity to identify factors that could benefit most from gender diversity in development teams. Through a case study based on a longitudinal quantitative analysis and a qualitative survey, we intended to answer the following research question:

What are the impacts of gender diversity on the performance and results of software development teams?

First of all, we ran a systematic literature review. We considered the period from January 2010 to March 2020, focusing in gender and women in Software Engineering. We identified a list of 126 qualified papers and evaluated them concerning how gender and women studies pervade Software Engineering. We found out that 79% of the published studies used the words gender, woman/women/female, indistinctly considering binary gender studies. However, few studies from the selected list bring the non-binary gender discussion to the light. There are challenges once gender is a social construct, and most studies reduce gender to binary (whether the individual is a woman or a man), risking a potential omission of non-binary people. Studies mentioning explicitly transgender individuals are also uncommon. These mapping was responsible the foundation of this research, once we would like to understand how the scientific community was approaching the subject.

From the systematic literature review, more than to identify the foundations for this thesis, we see additional opportunities to evolve gender research in Software Engineering in different fronts in future work. The expansion of gender studies to non-binaries appears to be a prolific area, as are the studies for transgender individuals inclusion. Additionally, to focus on intersectional studies and the ones related to cultural differences, skills, social/historical, emotions, performance, and gender prediction (on the dimension of non-binary).

Then, to run the case study we presented in Chapter 5, we relied on quantitative analysis of data extracted from software development teams code repositories. We would like to understand the impacts on the software assets delivered by the teams when they are more or less diverse. The pull-based software development model, exemplified and popularized by GitHub, decouples a software development task from the decision to incorporate its results in the codebase: when the software development task is completed, its author submits a pull request [49]. A pull request is the ability to propose changes to a remote codebase, which is often the one considered to be the central repository. A pull request is based on those changes, or delta, between a locally cloned codebase and the target central repository. As its name indicates, a pull request is not an immediate change in the codebase but a request for it [104].

So, we studied 14 software development teams from a technology company from Brazil, analyzing quantitative data from projects of their GitHub repositories. We analyzed data using descriptive statistics and regressions. For these teams, we identified that increasing gender diversity in small teams, also increase their Mean PRS and increasing diversity in large teams, increase their Mean PRLT. However, we also observed that the men's Mean PRLT were the most impacted. For this last specific finding, we did not have enough data to extend the study and understand why it happens. Here, we think that a more granular approach on data, extracting information as who were the team members who reviewed the code, how much time each member who reviewed the code took, what were the comments and how these comments impacted on the continuity of the reviewing. All this points could help to explain what kind of behavior leaded to this relationship. More than that, the analysis per gender would be also necessary. Some hypotheses we could bring are: Do women take more time reviewing code? If yes, why? Do they have more difficulties or they are more careful? From grey literature, we know the pull request size can impact the time to merge [68]. Depending on how many lines of code change, it requires more or less effort to review. Ironically, developers tend to merge long pull requests faster. People get lazy to perform through reviews when too many things are going on. So, they immediately approve changes. There are two problems here: your time to merge goes up, and your quality goes down [68]. So, we believe this finding has potential to support future work.

Besides gender, we also performed regression analysis to understand if the team size would impact more than gender diversity when talking about Pull Request Size of the teams or their Pull Request Lead Time. The idea was to ventilate if the size of a team could be more relevant than how much a team is gender diverse. In this case, we observed that increasing the team size in medium and large teams, increase their Mean PRS. Also, increasing the team size for medium teams helps decrease their Mean PRLT, probably diluting the workload through the team members.

These findings suggest that analyzing gender without considering other factors can be insufficient and can lead to wrong conclusions about the impact of gender diversity in software development teams. For this thesis we considered only team size as a comparison factor, once it was the one we could extract from our data. Other factors that could be used to extend this study are age, tenure, race/ethnicity, educational background, etc. Additionally, is is important to reinforce that the data of the 14 teams studied are from the same company. To replicate the study to other companies can bring different and complementary results.

From the qualitative point of view, the Survey results brought some perspectives about the kind of gains of having more gender diversity in teams. In summary, for companies, the main gains and benefits are the improvement of hiring process, to be more gender inclusive, solid cultural codes and encouragement of gender neutral workplaces. As results, companies reach better ideas sharing, better decision making and innovation. We saw evidences there are men, who are touched by the subject, trying to deconstruct their

prejudice and misconceptions about women in technology. Also, women try to support and inspire other women to remain in technology or to enter the field. respondents highlighted the importance of this kind of support.

When talking about perceived difficulties of gender diversity on software development teams, the respondents said, they do not see diversity in their companies, but they see the same pattern repeating itself in the teams and the leadership: white, men, cisgender. More than that, this pattern leads to male protectionism and privileges. Women suffer from sexism and prejudice, are frequently disrespected and, face difficulties to thrive in their careers. Other dimensions of diversity pervade the answers, as racism and ageism. Social vulnerability comes as a dimension related to meritocracy and elitism, once, at least in Brazil, equipment and training in Software Engineering are usually expensive.

Also, in our survey, respondents reported male protectionism and privileges preventing women to advance in their careers, which leads to a chain of difficulties for women: Women suffer from sexism and prejudice, that can be veiled and also it can lead to harassment. Women goes through professional insecurities. They feel they need to prove they are as capable as men and it generates the feeling of never being competent enough and their opinions' are worthless and disrespected.

Companies have a hard time supporting diversity and inclusion. Affirmative initiatives are scarce: too much talking and no action. More than that, the perception is that companies are hidden behind a facade of employers branding initiatives that happen only from the door out.

To conduct the Survey using the empathy map as a tool, opened space to other faces of diversity and also other kind of problems be mentioned as ageism, the process of systematic stereotyping or discrimination against people because they are old, just as racism and sexism accomplish with skin color and gender [24].

Meritocracy and elitism are also understood as problems. The "*if you want you can*" speech opposes to the lack of financial resources some individuals have to enter software engineering field. The area is defined as elitist once equipment and training are usually expensive. If an individual does not have the means to be trained as a software developer, the discourse of meritocracy does not apply. So we have four other aspects of diversity that are perceived as difficulties besides gender: Intersectionality, racism, ageism and social vulnerability.

Happiness and trust pointed to well being, remuneration, and career growth. Fear and anticipation appear related to the fear of being fired, what could lead to the need of being productive, working on too much tasks and small deadlines. If evidence from different researches on the subject suggest positive mood and emotions pay off in terms of performance [67], from our survey and previous researches, we can assume that a wealth life supported by a diverse and psychological safe work environment, could bring benefits for both, employees and companies. Regarding finances and remuneration, respondents

highlighted the importance of earning what is fair. For women, fair is to earn the same as they male peers. They also express preoccupation with financial stability to provide for their families and to have a sustainable retirement. Fear of unemployment was also mentioned.

Work overload was also a theme that pervaded the answers from the respondents of the survey. They reported delayed deliveries, lack of focus, low productivity, work overload, no quality in deliveries, etc. Work-life balance and the importance of family were mentioned as important. Respondents want quality time with theirs families and also they want to have a wealth life to support them.

Storey et al. [127] say modern software engineering involves both human and technical aspects and software engineering researchers may be expected to choose a balance of research strategies that capture both social and technical characteristics of software development. Storey et al. [128] also say social aspects can be approached methodologically by inferring behavior from analyzing trace data of developers' past activities (e.g., code commits, code review comments, posted questions and answers on developer forums, etc.). But the analysis of trace data alone is fraught with threats to validity as it shows an incomplete picture of human behavior, intent, and social interactions in software engineering. Furthermore, trace data alone cannot be used to predict how a new solution may perturb an existing process in industry settings, although relying on trace data can bring early insights about the feasibility of a solution design. To appropriately capture and account for social aspects in software engineering research, we need to use dedicated methods that directly involve human participants in our empirical studies.

Our findings are aligned with what is proposed by Storey et al. [127, 128]. Studies regarding human and social aspects in Software Engineering need qualitative and quantitative data triagularization to achieve more accurate results. Software Engineering is complex and changes constantly, so, the importance of having different sources of data and also, the importance of considering results of different studies from different researchers, so we can have a plural result.

Our findings suggest that, when performing gender diverse quantitative studies, is important considering different control variables to have a more accurate idea of the complex social environment where the software development teams are inserted. Also, we reinforce the importance of the qualitative studies to highlight the gender bias exists as the perpetuation of the idea the Software Engineering is a men's field.

More than ever, the subject is being intensively discussed by society and Software Engineering researchers. To give visibility on these discussions is important so we can put the results in perspective and to reach a field less biased and more inclusive.

8. CONCLUSION

Diversity is being intensively discussed in different knowledge areas of society and it is a complex issue, as groups can be diverse in terms of various attributes, such as ethnicity, gender, age, and socio-economic background. The discussions about diversity in Software Engineering increased, in the last years, as well. Software engineering involves real people in real environments, creating and maintaining software, so it is imperative to study how software practitioners solve problems in real environments and how they interact.

8.1 Contributions

Through the development of this work we collected evidences that the subject is relevant to the field. In a daily basis, women use to report their difficulties working with Software Engineering, in different roles. However, it is frequently see as anecdotal or related to a few.

The quantitative study brought information mainly for small and large teams. For small teams we observed that the increasing in gender diversity also increase the Pull request Size of the teams. For large teams, we observed increasing the Pull Request Lead Time. However, in this case, when we split the observation, we see that this increase happens mainly for men. It opens opportunity to extends the study to understand this factor. The findings suggested that analyzing gender without considering other factors can be insufficient and can lead to wrong conclusions about the impact of gender diversity in software development teams. We considered team size as a comparison factor, once it was the one we could extract from our data. Other factors that could be used to extend this study are age, tenure, race/ethnicity, educational background, etc.

From the qualitative study, we had different evidences of how the subject is ignored for some people and how it impacts career and emotions of women and people in other dimensions of diversity. Unawareness and bias appeared in the answers, bringing evidences that, yes, there are individuals that minimize the importance of considering diversity. Also, the answers brought how it impacts women in a daily basis, when people reported difficulties to thrive in careers due to sexism, racism, ageism and social vulnerabilities. Intersectionality, meritocracy and elitism appeared as related topics, what makes the need of bringing these subjects to the table more efficiently in research.

But the qualitative study also gave us evidences about the gains and benefits of gender diversity. Some companies are diligently working on improve their hiring process, to be more gender inclusive, solid cultural codes and encouragement of gender neutral workplaces. They see they can reach reach better ideas sharing, better decision making and

innovation. Also, there are men, trying to deconstruct their prejudice and misconceptions about women in technology. Also, women try to support and inspire other women to remain in technology or to enter the field.

8.2 Threats to validity

The validity of a study denotes the trustworthiness of the results and to what extent the results are true and not biased by the researchers' subjective point of view [154]. There are different ways to classify aspects of validity and threats to validity in the literature [154]. For this research, we used a classification scheme that is suggested by Wohlin et al. [154], and Yin [155], as follows:

8.2.1 Construct Validity

This aspect of validity reflects what extent the operational measures that are studied represent what the researcher has in mind and what is investigated according to the research questions [154].

We see potential threats to construct validity in this research, coming mainly from the Survey Chapter 6. If the constructs discussed in the survey questions were not interpreted in the same way by the researcher and the respondents, there is a threat to construct validity.

8.2.2 Internal Validity

This aspect of validity is of concern when causal relations are examined. When the researcher investigates whether one factor affects an investigated factor, there is a risk that the investigated factor is also affected by a third factor. If the researcher is not aware of the third-factor andor does not know to what extent it affects the investigated factor, there is a threat to internal validity [154].

Regarding internal validity, we can mention *Researcher bias:* refers to the potential bias that the author of the study may have when interpreting or synthesizing the extracted results of the survey Chapter 6. This type of bias can occur in relation to a topic or because only one author worked on data synthesis.

8.2.3 External Validity

This aspect of validity is concerned with the extent to which it is possible to generalize the findings and the extent to which the findings are of interest to other people outside the investigated case. During analysis of external validity, the researcher tries to analyze to what extent the findings are of relevance for other cases. In case studies, there is no population from which a statistically representative sample has been drawn. However, for case studies, the intention is to enable analytical generalization where the results are extended to cases that have common characteristics and hence for which the findings are relevant, i.e., defining a theory [154].

Concerning external validity, the results presented for the Case Study in Chapter 5 are valid for the software development teams of the company we analyzed. Diversity culture and development process may vary across companies, bringing different results. Same rational can be applied to the results from the Survey in Chapter 6. However, this threat is also an opportunity to extend the study and invite researchers to replicate it in other technology companies,

8.2.4 Reliability

This aspect is concerned with to what extent the data and the analysis are dependent on the specific researchers. Hypothetically, if another researcher, later on, conducted the same study, the result should be the same. Threats to this aspect of validity are, for example, if it is not clear how to code collected data or if questionnaires or interview questions are unclear [154].

For this research, we understand that reliability threats are mitigated once we provide a protocol for the case study (B) and a descriptive data collection and analysis Sections in 5. Same happens for the survey, where we provide the protocol and question in C.

8.3 Congresses and Conferences

Through the journey to deliver this thesis, this research resulted in scientific publications. Additionally, the PhD. Candidate worked as reviewer for different conferences, workshops and journals, and also as co-chair in a workshop. In the following subsections, we present these results.

8.3.1 Publications

Eight publications , 24 citations (January 2022)

- *K. Kohl* and R. Prikladnicki, "Perceptions on Diversity in Brazilian Agile Software Development Teams: A Survey," 2018 IEEE/ACM 1st International Workshop on Gender Equality in Software Engineering (GE), Gothenburg, 2018, pp. 37-40.
  (QUALIS B3) (Cited by 9)

- *Silveira, K.;* Musse, S.; Manssour, I.; Vieira, R. and Prikladnicki, R. (2019)."Reinforcing Diversity Company Policies: Insights from StackOverflow Developers Survey." In Proceedings of the 21st International Conference on Enterprise Information Systems - Volume 2: ICEIS, ISBN 978-989-758-372-8, pages 119-129. (Best Student Paper, Area: Information Systems Analysis and Specification)
  (QUALIS A3) (Cited by 1) * Best student paper

- *K. Kohl Silveira* and R. Prikladnicki, "A Systematic Mapping Study of Diversity in Software Engineering: A Perspective from the Agile Methodologies," 2019 IEEE/ACM 12th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE), Montreal, QC, Canada, 2019, pp. 7-10.
  (QUALIS A4) (Cited by 8)

- *K. Kohl Silveira*, S. Musse, I. H. Manssour, R. Vieira and R. Prikladnicki, "Confidence in Programming Skills: Gender Insights From StackOverflow Developers Survey," 2019 IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion), Montreal, QC, Canada, 2019, pp. 234-235.
  (QUALIS A1) (Cited by 5)

- *K. Kohl Silveira*, "Future of Agile is Diversity," 2019 Agile Conference, Session Track: The Future of Agile Software Development (IEEE Software), Washington DC, USA, 2019.
  (QUALIS A3)

- *K. Kohl Silveira*, B. Vasilescu and R. Prikladnicki, "Multitasking Across Industry Projects: A Replication Study," 2020 IEEE/ACM 13th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE), Seoul, South Korea 2020.
  (QUALIS A4)

- Ivica Crnkovic, *Karina Kohl Silveira*, and Sara Sprenkle. 2020. Summary of the 2nd Workshop on Gender Equality in Software Engineering (GE 2019). SIGSOFT Softw. Eng. Notes 45, 3 (July 2020), 25–27.
  (QUALIS -)

- *Kohl, K.* and Prikladnicki, R. (2021). Challenges Women in Software Engineering Leadership Roles Face: A Qualitative Study. In Proceedings of the 23rd International Conference on Enterprise Information Systems - Volume 2: ICEIS, ISBN 978-989-758-509-8; ISSN 2184-4992, pages 205-212.
  (QUALIS A3)(Cited by 1)

8.3.2 Reviewer

- Reviewer Agile Conference 2019, Session Track: The Future of Agile Software Development (IEEE Software), Washington DC, USA, 2019
- Reviewer IEEE Software, SWSI-2019-06-0114. SWSI: Software Engineering in Society.
- Reviewer IEEE Software, SWSI-2020-08-0222. SWSI: The Diversity Crisis in Software Development
- Reviewer SBES 2020 mentored by Prof. Dr. Sabrina Marczak.
- Reviewer for EMSE - Empirical Software Engineering, in 2021.
- Program Committee, 3_{rd} Workshop on Gender Equality, Diversity, and Inclusion in Software Engineering (GE 2022) collocated with ICSE 2022.

8.3.3 Chair

- Co-Chair 2019 IEEE/ACM 2nd International Workshop on Gender Equality in Software Engineering (GE), ICSE Workshop, together with Prof. Dr. Ivica Crnkovic, Chalmers University of Technology (Gotemburgo, Suécia), and Prof. Dra. Sara Sprenkle, Washington and Lee University (Virginia, EUA)

8.4 Future Work

We observed that increasing diversity in large teams, increase their Mean PRLT. However, we also observed that the men's Mean PRLT were the most impacted. For this last specific finding, we did not have enough data to extend the study and understand why it happens. Here, we think that a more granular approach on data, extracting information as who were the team members who reviewed the code, how much time each member who

reviewed the code took, what were the comments and how these comments impacted on the continuity of the reviewing. All this points could help to explain what kind of behavior leaded to this relationship. More than that, the analysis per gender would be also necessary.

Also, previous studies from other researchers bring some possibilities to extend this work. For example, for small teams we found out the Pull Request Size tend to increase. Antin et al. [7] mention that in Wikipedia contributions, men do frequent contributions and women larger contributions. Is it a behavior that maybe can be tested in Pull Requests of Software Development Teams? Vasilescu et al. [138] and Terrell et al. [132] have studies which women that use male profiles or the gender is not identifiable, have better acceptance. Blind reviews for industry teams can support a more balanced PRLT?

REFERENCES

[1] A., K.; Mejlgaard, N.; Haase, S.; Holgaard, J. E. “Motivational factors, gender and engineering education”, *European Journal of Engineering Education*, May 2013, pp. 340–358.

[2] Aboobaker, N.; Edward, M.; Pramatha, K. P. “Work-family Conflict, Family-work Conflict and Intention to Leave the Organization: Evidences Across Five Industry Sectors in India”, *Global Business Review*, Apr 2017, pp. 524–536.

[3] Alba, B. “To achieve gender equality, we must first tackle our unconscious biases”. Source: https://theconversation.com/to-achieve-gender-equality-we-must-first-tackle-our-unconscious-biases-92848, Sep 2021.

[4] AlShebli, B. K.; Rahwan, T.; Woon, W. L. “The preeminence of ethnic diversity in scientific collaboration”, *Nature Communications*, Dec 2018, pp. 5163.

[5] Andrade, R. O. “A retomada do espaço da mulher na computação”. Source: https://revistapesquisa.fapesp.br/2019/05/10/a-retomada-do-espaco-da-mulher-na-computacao/, Jan 2020.

[6] Angelini, C. “Regression analysis”. In: *Encyclopedia of Bioinformatics and Computational Biology*, Ranganathan, S.; Gribskov, M.; Nakai, K.; Schönbach, K. (Editors), Oxford: Academic Press, 2019, pp. 722–730.

[7] Antin, J.; Yee, R.; Cheshire, C.; Nov, O. “Gender differences in wikipedia editing”. In: Proceedings of the International Symposium on Wikis and Open Collaboration, 2011, pp. 11–14.

[8] Aspray, W. “Women and Underrepresented Minorities in Computing: A Historical and Social Study.” Springer, 2016, 271p.

[9] Association, A. P. “Strain”. Source: https://dictionary.apa.org/strain, Sep 2021.

[10] Babbie, R. “Survey Research Methods”. Wadsworth Publishing Company, 1990, 395p.

[11] Bailenson, J. N. “Nonverbal overload: A theoretical argument for the causes of zoom fatigue”, *Technology, Mind, and Behavior*, Feb 2021, pp. 1–6.

[12] Barr, E. T.; Bird, C.; Rigby, P. C.; Hindle, A.; German, D. M.; Devanbu, P. “Cohesive and isolated development with branches”, *Lecture Notes in Computer Science*, Apr 2012, pp. 316–331.

[13] Bass, B. M. “From transactional to transformational leadership: Learning to share the vision”, *Organizational Dynamics*, Jan 1990, pp. 19–31.

[14] Benbasat, I.; Goldstein, D. K.; Mead, M. “The case research strategy in studies of information systems”, *Management Information Systems Quarterly*, Sep 1987, pp. 369–386.

[15] Bennett, C. L.; Rosner, D. K. “The promise of empathy: Design, disability, and knowing the "other"”. In: Proceedings of the Conference on Human Factors in Computing Systems, 2019, pp. 1–13.

[16] Berenson, S. B.; Slaten, K. M.; Williams, L.; Ho, C.-W. “Voices of women in a software engineering course: Reflections on collaboration”, *Journal on Educational Resources in Computing*, Mar 2004, pp. 3–es.

[17] Blau, P. M. “Inequality and Heterogeneity - A Primitive Theory Of Social Structure”. Free Press New York, 1977, 308p.

[18] Bosu, A.; Sultana, K. Z. “Diversity and inclusion in open source software (oss) projects: Where do we stand?” In: Proceedings of the ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, 2019, pp. 1–11.

[19] Braun, V.; Clarke, V. “Using thematic analysis in psychology”, *Qualitative Research in Psychology*, Jul 2006, pp. 77–101.

[20] Brereton, P.; Kitchenham, B.; Budgen, D.; Li, Z. “Using a protocol template for case study planning”. In: Proceedings of the International Conference on Evaluation and Assessment in Software Engineering, 2008, pp. 41–48.

[21] Brooke, C. “What does it mean to be ‘critical’ in is research?”, *Journal of Information Technology*, Dec 2002, pp. 49–57.

[22] Brooks, F. P. “The Mythical Man-Month (Anniversary Ed.)”. Addison-Wesley Longman Publishing Co., Inc., 1995, 336p.

[23] Buhnova, B.; Jurystova, L.; Prikrylova, D. “Assisting women in career change towards software engineering: Experience from czechitas ngo”. In: Proceedings of the European Conference on Software Architecture, 2019, pp. 88–93.

[24] Butler, R. “Age-ism: another form of biogtry”, *Gerontologist*, Dec 1969, pp. 243–246.

[25] Cameron, C. D.; Hutcherson, C. A.; Ferguson, A. M.; Scheffer, J. A.; Hadjiandreou, E.and Inzlicht, M. “Empathy is hard work: People choose to avoid empathy because of its cognitive costs”, *Journal of Experimental Psychology: General*, Jun 2019, pp. 962–976.

[26] Carver, J. C.; Serebrenik, A. “Gender in software engineering”, *IEEE Software*, Nov 2019, pp. 76–78.

[27] Catolino, G.; Palomba, F.; Tamburri, D. A.; Serebrenik, A.; Ferrucci, F. “Gender diversity and women in software teams: How do they affect community smells?” In: Proceedings of the IEEE/ACM International Conference on Software Engineering: Software Engineering in Society, 2019, pp. 11–20.

[28] Catolino, G.; Palomba, F.; Tamburri, D. A.; Serebrenik, A.; Ferrucci, F. “Gender diversity and community smells: Insights from the trenches”, *IEEE Software*, Jan 2020, pp. 10–16.

[29] Cech, E. A.; Blair-Loy, M. “Perceiving Glass Ceilings? Meritocratic versus Structural Explanations of Gender Inequality among Women in Science and Technology”, *Social Problems*, Jul 2014, pp. 371–397.

[30] Cheryan, S.; Ziegler, S. A.; Montoya, A. K.; Jiang, L. “Why are some stem fields more gender balanced than others?”, *Psychological Bulletin*, Jan 2017, pp. 1–35.

[31] Coffman, E.; Galambos, J.; Martello, S.; Vigo, D. “Oxford english dictionary.” Source: https://www.oxfordlearnersdictionaries.com/, Sep 2021.

[32] Collis, J.; Hussey, R. “Business Research: A Practical Guide for Undergraduate and Postgraduate Students”. Palgrave Macmillan, 2013, 376p.

[33] Conradi, R.; Wang, A. I. “Empirical Methods and Studies in Software Engineering - Experiences from ESERNET”. Springer-Verlag Berlin Heidelberg, 2003, 284p.

[34] Convertino, C. “Nuancing the discourse of underrepresentation: a feminist post-structural analysis of gender inequality in computer science education in the us”, *Gender and Education*, Jun 2020, pp. 594–607.

[35] Crenshaw, K. “Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics.”, *University of Chicago Legal Forum*, Dec 1989, pp. 1–31.

[36] Creswell, J. “Research Design: Qualitative, Quantitative, and Mixed Methods Approaches”. SAGE Publications, 2014, 304p.

[37] Curtis, B. “Chapter 4 - five paradigms in the psychology of programming”. In: *Handbook of Human-Computer Interaction*, Helander, M. (Editor), Amsterdam: North-Holland, 1988, pp. 87–105.

[38] Darley, W. K.; Smith, R. E. “Gender differences in information processing strategies: An empirical test of the selectivity model in advertising response”, *Journal of Advertising*, Apr 1995, pp. 41–56.

[39] Dawson, R.; Bones, P.; Oates, B. J.; Brereton, P.; Azuma, M.; Jackson, M. L. “Empirical methodologies in software engineering”. In: Proceedings of the International Workshop on Software Technology and Engineering Practice, 2003, pp. 52–58.

[40] Decety, J.; Cowell, J. M. “Friends or foes: Is empathy necessary for moral behavior?”, *Perspectives on Psychological Science*, Sep 2014, pp. 525–537.

[41] DeMarco, T.; Lister, T. R. “Peopleware: Productive Projects and Teams”. Dorset House Publishing Company, 1987, 272p.

[42] Easterbrook, S.; Singer, J.; Storey, M.-A. “Selecting Empirical Methods for Software Engineering Research”. Springer-Verlag London, 2008, chap. 11, pp. 285–286.

[43] Ekman, P.; Friesen, W.; Ellsworth, P.; Goldstein, A.; Krasner, L. “Emotion in the Human Face: Guidelines for Research and an Integration of Findings”. Elsevier Science, 1972, 191p.

[44] Engel, R.; Schutt, R. “The Practice of Research in Social Work”. SAGE, 2009, 496p.

[45] Ensmenger, N. L. “The Computer Boys Take Over: Computers, Programmers, and the Politics of Technical Expertise”. The MIT Press, 2010, 336p.

[46] Feldt, R.; Torkar, R.; Angelis, L.; Samuelsson, M. “Towards individualized software engineering: Empirical studies should collect psychometrics”. In: Proceedings of the International Workshop on Cooperative and Human Aspects of Software Engineering, 2008, pp. 49–52.

[47] Ferreira, B.; Conte, T.; Diniz Junqueira Barbosa, S. “Eliciting requirements using personas and empathy map to enhance the user experience”. In: Proceedings of the Brazilian Symposium on Software Engineering, 2015, pp. 80–89.

[48] Fisher, R. “Statistical methods for research workers (5th ed.)”. Oliver and Boyd: Edinburgh, 1934, 378p.

[49] Ford, D.; Behroozi, M.; Serebrenik, A.; Parnin, C. “Beyond the Code Itself: How Programmers Really Look at Pull Requests”. In: Proceedings of the IEEE/ACM International Conference on Software Engineering: Software Engineering in Society, 2019, pp. 51–60.

[50] Ford, D.; Milewicz, R.; Serebrenik, A. “How remote work can foster a more inclusive environment for transgender developers”. In: Proceedings of the IEEE/ACM International Workshop on Gender Equality in Software Engineering, 2019, pp. 9–12.

[51] Foundation, M. H. “Stress”. Source: https://www.mentalhealth.org.uk/a-to-z/s/stress, Sep 2021.

[52] Frost, J. “How to interpret p-values and coefficients in regression analysis”. Source: https://statisticsbyjim.com/regression/interpret-coefficients-p-values-regression/, Nov 2021.

[53] Fujigaki, Y. “Time series investigation of job-events and depression in computer software engineers”, *Industrial Health*, Jan 1996, pp. 71–79.

[54] Galinsky, A. D.; Todd, A. R.; Homan, A. C.; Phillips, K. W.; Apfelbaum, E. P.; Sasaki, S.; Richeson, J. A.; Olayon, J. B.; Maddux, W. W. “Maximizing the gains and minimizing the pains of diversity: A policy perspective”, *Perspectives on Psychological Science*, Nov 2015, pp. 742–748.

[55] Gallo, A. “A refresher on regression analysis”. Source: https://hbr.org/2015/11/a-refresher-on-regression-analysis, Nov 2021.

[56] GitHub. “Github docx. metrics available with github insights.” Source: https://docs.github.com/en/enterprise-server@2.19/insights/exploring-your-usage-of-github-enterprise/metrics-available-with-github-insights/en/free-pro-team@latest/github/collaborating-with-issues-and-pull-requests/about-pull-requests, Nov 2020.

[57] GitHub. “Github guides. hello world.” Source: https://guides.github.com/activities/hello-world/, Nov 2020.

[58] González-Bañales, D. L.; Ortíz, L. E. S. “Empathy map as a tool to analyze human-computer interaction in the elderly”. In: Proceedings of the Latin American Conference on Human-Computer Interaction, 2017, pp. 1–3.

[59] Google. “Google diversity”. Source: https://diversity.google/, Sep 2021.

[60] Gousios, G.; Zaidman, A.; Storey, M. A.; Van Deursen, A. “Work practices and challenges in pull-based development: The integrator’s perspective”, *Proceedings of the International Conference on Software Engineering*, Aug 2015, pp. 358–368.

[61] Gramß, D.; Frank, T.; Rehberger, S.; Vogel-Heuser, B. “Female characteristics and requirements in software engineering in mechanical engineering”. In: Proceedings of the International Conference on Interactive Collaborative Learning, 2014, pp. 272–279.

[62] Gray, D. “Empathy map”. Source: https://gamestorming.com/empathy-mapping/, Jun 2021.

[63] Gray, D.; Brown, S.; Macanufo, J. “Gamestorming: A Playbook for Innovators, Rulebreakers, and Changemakers”. O’Reilly Media, Inc., 2010, 266p.

[64] Graziotin, D.; Fagerholm, F.; Wang, X.; Abrahamsson, P. “Consequences of unhappiness while developing software”. In: Proceedings of the International Workshop on Emotion Awareness in Software Engineering, 2017, pp. 42–47.

[65] Graziotin, D.; Fagerholm, F.; Wang, X.; Abrahamsson, P. “What happens when software developers are (un)happy”, *Journal of Systems and Software*, Jun 2018, pp. 32–47.

[66] Graziotin, D.; Wang, X.; Abrahamsson, P. “Happy software developers solve problems better: Psychological measurements in empirical software engineering”, *PeerJ*, Mar 2014, pp. 1–23.

[67] Graziotin, D.; Wang, X.; Abrahamsson, P. “Software developers, moods, emotions, and performance”, *IEEE Software*, Jul 2014, pp. 24–27.

[68] Guimarães, G. “5 metrics engineering managers can extract from pull requests”. Source: https://sourcelevel.io/blog/5-metrics-engineering-managers-can-extract-from-pull-requests, Nov 2021.

[69] H., E. A.; Carli, L. L. “The female leadership advantage: An evaluation of the evidence”, *The Leadership Quarterly*, Dec 2003, pp. 807–834.

[70] Hannah, D. R.; Lautsch, B. A. “Counting in qualitative research: Why to conduct it, when to avoid it, and when to closet it”, *Journal of Management Inquiry*, Sep 2011, pp. 14–22.

[71] Hattori, L. P.; Lanza, M. “On the nature of commits”. In: Proceedings of the IEEE/ACM International Conference on Automated Software Engineering, 2008, pp. III–63–III–71.

[72] Hazzan, O.; Dubinsky., Y. “Empower Gender Diversity with Agile Software Development”. Ed. Hershey, 2006, pp. 249–256.

[73] Henschel, S.; Nandrino, J.-L.; Doba, K. “Emotion regulation and empathic abilities in young adults: The role of attachment styles”, *Personality and Individual Differences*, Apr 2020, pp. 109763.

[74] Hewlett, S.; Luce, C. “The Athena Factor: Reversing the Brain Drain in Science, Engineering, and Technology”. Harvard Business Review, 2008, 108p.

[75] Hewlett, S. A.; Marshall, M.; Sherbin, L. “How diversity can drive innovation”. Source: https://hbr.org/2013/12/how-diversity-can-drive-innovation, Jan 2022.

[76] Hicks, M. “Programmed Inequality – How Britain Discarded Women Technologists and Lost Its Edge in Computing”. MIT Press, 2017, 352p.

[77] Hill, C.; Ernst, S.; Oleson, A.; Horvath, A.; Burnett, M. “Gendermag experiences in the field: The whole, the parts, and the workload”. In: Proceedings of the IEEE Symposium on Visual Languages and Human-Centric Computing, 2016, pp. 199–207.

[78] Hunt, V.; Layton, D.; Prince, S. “Why diversity matters”. Source: https://www.mckinsey.com/business-functions/organization/our-insights/why-diversity-matters, Jan 2020.

[79] Imtiaz, N.; Middleton, J.; Chakraborty, J.; Robson, N.; Bai, G.; Murphy-Hill, E. “Investigating the effects of gender bias on github”. In: Proceedings of the International Conference on Software Engineering, 2019, pp. 700–711.

[80] Izquierdo, D.; Huesman, N.; Serebrenik, A.; Robles, G. “Openstack gender diversity report”, *IEEE Software*, Jan 2019, pp. 28–33.

[81] Jackman, M. “Homeopathic remedies for team toxicity”, *IEEE Software*, Jul 1998, pp. 43–45.

[82] Jehn, K. A.; Northcraft, G. B.; Neale, M. A. “Why differences make a difference: A field study of diversity, conflict and performance in workgroups”, *Administrative Science Quarterly*, Dec 1999, pp. 741 – 763.

[83] Jetter, A. J.; Loanzon, E.; Jahromi, S.; Nour, A. H.; Pakdeekasem, P. “An exploratory study on the leadership style preferences of male and female managers: Implications on team performance”. In: Proceedings of PICMET: Technology Management in the IT-Driven Services, 2013, pp. 1161–1181.

[84] Jetter, M.; Walker, J. K. “The gender of opponents: Explaining gender differences in performance and risk-taking?”, *European Economic Review*, Oct 2018, pp. 238–256.

[85] Judy, K. H. “Agile values, innovation and the shortage of women software developers”. In: Proceedings of the Hawaii International Conference on System Sciences, 2012, pp. 5279–5288.

[86] Kim, C. H.; Choi, Y. B. “How meritocracy is defined today?: Contemporary aspects of meritocracy”, *Economics and Sociology*, Dec 2017, pp. 112–121.

[87] Kitchenham, B.; Charters, S. “Guidelines for performing Systematic Literature reviews in Software Engineering Version 2.3”, *Engineering*, Jul 2007, pp. 1051.

[88] Klein, H. K.; Myers, M. D. “A set of principles for conducting and evaluating interpretive field studies in information systems”, *Management Information Systems Quarterly*, Mar 1999, pp. 67–93.

[89] Kohl, K.; Prikladnicki, R. “Perceptions on diversity in brazilian agile software development teams: A survey”. In: Proceedings of the IEEE/ACM International Workshop on Gender Equality in Software Engineering, 2018, pp. 37–40.

[90] Kohl, K.; Prikladnicki, R. “A systematic mapping study of diversity in software engineering: A perspective from the agile methodologies”. In: Proceedings of the IEEE/ACM International Workshop on Cooperative and Human Aspects of Software Engineering, 2019, pp. 7–10.

[91] Kumar, D. “Disrupting the cultural capital of brogrammers”, *ACM Inroads*, Sep 2014, pp. 28–29.

[92] Lacey, M. “Tests of significance”. Source: http://www.stat.yale.edu/Courses/1997-98/101/sigtest.htm, Nov 2021.

[93] Lethbridge, T. C.; Sim, S. E.; Singer, J. “Studying software engineers: Data collection techniques for software field studies”. Springer, 2005, pp. 311–341.

[94] Level, S. “Pull requests — checklists, metrics and best practices, a definitive guide”. Source: https://sourcelevel.io/pull-requests-checklists-metrics-and-best-practices-a-definitive-guide, Aug 2021.

[95] Mendez, C.; Padala, H. S.; Steine-Hanson, Z.; Hilderbrand, C.; Horvath, A.; Hill, C.; Simpson, L.; Patil, N.; Sarma, A.; Burnett, M. “Open source barriers to entry, revisited: A sociotechnical perspective”. In: Proceedings of the International Conference on Software Engineering, 2018, pp. 1004–1015.

[96] Mendez, C.; Sarma, A.; Burnett, M. “Gender in open source software: What the tools tell”. In: Proceedings of the IEEE/ACM International Workshop on Gender Equality in Software Engineering, 2018, pp. 21–24.

[97] Microsoft. “Microsoft diversity and inclusion”. Source: https://www.microsoft.com/en-us/diversity/default.aspx, Sep 2021.

[98] Navarro, R. “Unconscious bias.” Source: https://diversity.ucsf.edu/resources/unconscious-bias, Dec 2021.

[99] Nguyen-Duc, A.; Khodambashi, S.; Gulla, J. A.; Krogstie, J.; Abrahamsson, P. “Female Leadership in Software Projects—A Preliminary Result on Leadership Style and Project Context Factors”. Cham: Springer International Publishing, 2018, chap. 11, pp. 149–163.

[100] Niederle, M.; Segal, C.; Vesterlund, L. “How costly is diversity? Affirmative action in light of gender differences in competitiveness”, *Management Science*, Apr 2013, pp. 1–16.

[101] Nonaka, I.; Takeuchi, H. “The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation”. Oxford University Press, 1995, 68p.

[102] Novielli, N.; Serebrenik, A. “Sentiment and Emotion in Software Engineering”, *IEEE Software*, Aug 2019, pp. 6–23.

[103] Orlikowski, W. J.; Baroudi, J. J. “Studying information technology in organizations: Research approaches and assumptions”, *Information Systems Research*, Mar 1991, pp. 1–28.

[104] Ortu, M.; Destefanis, G.; Graziotin, D.; Marchesi, M.; Tonelli, R. “How do you Propose Your Code Changes? Empirical Analysis of Affect Metrics of Pull Requests on GitHub”, *IEEE Access*, Jun 2020, pp. 110897–110907.

[105] Page, S. E. “The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies”. Princeton University Press, 2007, 424p.

[106] Page, S. E. “The Diversity Bonus: How Great Teams Pay Off in the Knowledge Economy (Our Compelling Interests)”. Princeton University Press, 2017, 328p.

[107] Petersen, K.; Feldt, R.; Mujtaba, S.; Mattsson, M. “Systematic mapping studies in software engineering”. In: Proceedings of the International Conference on Evaluation and Assessment in Software Engineering, 2008, pp. 68–77.

[108] Pfleeger, S. L.; Kitchenham, B. A. “Principles of survey research: Part 1: Turning lemons into lemonade”, *Software Engineering Notes*, Nov 2001, pp. 16–18.

[109] Plutchik, R. “The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice”, *American Scientist*, Jul 2001, pp. 344–350.

[110] Pozos, R. K.; Friend, M. “"you sound like a good program manager": An analysis of gender in women’s computing life histories”. In: Proceedings of the ACM Technical Symposium on Computer Science Education, 2021, pp. 692–698.

[111] Pressman, R.; Maxim, B. “Engenharia de Software - 8ª Edicao”. McGraw Hill Brasil, 2016, 968p.

[112] Pretorius, C.; Razavian, M.; Eling, K.; Langerak, F. “Combined intuition and rationality increases software feature novelty for female software designers”, *IEEE Software*, Mar 2021, pp. 64–69.

[113] Publishing, E. “How to conduct empirical research”. Source: https://www.emeraldgrouppublishing.com/research/guides/methods/empirical.htm?part=1, Jan 2020.

[114] Radatz, J. “Ieee standard glossary of software engineering terminology”, *IEEE Standard Glossary of Software Engineering Terminology*, Dec 1990, pp. 1–84.

[115] Remedios, J. D.; Akhtar, M. “9 - intersectional approaches to the study of confronting prejudice”. In: *Confronting Prejudice and Discrimination*, Mallett, R. K.; Monteith, M. J. (Editors), Academic Press, 2019, pp. 179–200.

[116] Runeson, P.; Höst, M. “Guidelines for conducting and reporting case study research in software engineering”, *Empirical Software Engineering*, Apr 2009, pp. 131–164.

[117] Runeson, P.; Host, M.; Rainer, A.; Regnell, B. “Case Study Research in Software Engineering: Guidelines and Examples”. Wiley Publishing, 2012, 237p.

[118] Santos Machado, L.; Perlin, M.; Colla Soletti, R.; Kmetzch Rosa e Silva, L.; Doerderlein Schwartz, I. V.; Seixas, A.; Klein Ricachenevsky, F.; Tamajusuku Neis, A.; Staniscuaski, F. “Parent in science: The impact of parenthood on the scientific career in brazil”. In: Proceedings of the IEEE/ACM International Workshop on Gender Equality in Software Engineering, 2019, pp. 37–40.

[119] Seaman, C. “Qualitative methods in empirical studies of software engineering”, *IEEE Transactions on Software Engineering*, Jul 1999, pp. 557–572.

[120] Seltman, H. “Experimental Design and Analysis”. CMU, 2018, 428p.

[121] Shneiderman, B. “Software psychology: human factors in computer and information systems”. Little, Brown, 1980, 320p.

[122] Silveira, K. K.; Prikladnicki, R. “A systematic mapping study of diversity in software engineering: A perspective from the agile methodologies”. In: Proceedings of the International Workshop on Cooperative and Human Aspects of Software Engineering, 2019, pp. 7–10.

[123] Simmonds, J.; Bastarrica, M. C.; Hitschfeld-Kahler, N. “Impact of affirmative action on female computer science/software engineering undergraduate enrollment”, *IEEE Software*, Feb 2021, pp. 32–37.

[124] Steele, C. M. “A threat in the air. how stereotypes shape intellectual identity and performance.”, *The American psychologist*, Jun 1997, pp. 613–629.

[125] Stol, K.-J.; Fitzgerald, B. “The abc of software engineering research”, *ACM Transactions on Software Engineering Methodology*, Jul 2018, pp. 1–51.

[126] Storey, M.; Williams, C.; Ernst, N. A.; Zagalsky, A.; Kalliamvakou, E. “Methodology matters: How we study socio-technical aspects in software engineering”, *ACM Transactions on Software Engineering and Methodology*, Aug 2019, pp. 111:1–111:22.

[127] Storey, M.-A. “Publish or perish: Questioning the impact of our research on the software developer”. In: Proceedings of the IEEE/ACM International Conference on Software Engineering: Companion Proceedings, 2019, pp. 2–2.

[128] Storey, M. A.; Ernst, N. A.; Williams, C.; Kalliamvakou, E. “The who, what, how of software engineering research: a socio-technical framework”, *Empirical Software Engineering*, Sep 2020, pp. 4097–4129.

[129] Sytsma, S. E. “Ethics and intersex.” Springer e-books. Dordrecht: Springer, 2006, 351p.

[130] Sánchez-Gordón, M.; Colomo-Palacios, R. “A framework for intersectional perspectives in software engineering”. In: Proceedings of the IEEE/ACM International Workshop on Cooperative and Human Aspects of Software Engineering, 2021, pp. 121–122.

[131] Takeuchi, H.; Nonaka, I. “Hitotsubashi on Knowledge Management”. Wiley, 2004, 250p.

[132] Terrell, J.; Kofink, A.; Middleton, J.; Rainear, C.; Murphy-Hill, E.; Parnin, C.; Stallings, J. “Gender differences and bias in open source: Pull request acceptance of women versus men”, *PeerJ*, May 2017, pp. 1–30.

[133] Tourani, P.; Adams, B.; Serebrenik, A. “Code of conduct in open source projects”. In: Proceedings of the IEEE International Conference on Software Analysis, Evolution and Reengineering, 2017, pp. 24–33.

[134] Turley, R. T.; Bieman, J. M. “Identifying essential competencies of software engineers”. In: Proceedings of the Annual ACM Computer Science Conference on Scaling up, 1994, pp. 271–278.

[135] Turley, R. T.; Bieman, J. M. “Competencies of exceptional and nonexceptional software engineers”, *Journal of Systems and Software*, Jan 1995, pp. 19–38.

[136] Unesco. “Cracking the code: girls’ and women’s education in science, technology, engineering and mathematics (STEM)”. UNESCO Publishing, 2017, 85p.

[137] Usher, R. “North American Lexicon of Transgender Terms.” GLB Publishers., 2006, 64p.

[138] Vasilescu, B.; Capiluppi, A.; Serebrenik, A. “Gender, representation and online participation: A quantitative study”, *Interacting with Computers*, Sep 2013, pp. 488–511.

[139] Vasilescu, B.; Filkov, V.; Serebrenik, A. “Perceptions of diversity on GitHub: A user survey”. In: Proceedings of the International Workshop on Cooperative and Human Aspects of Software Engineering, 2015, pp. 50–56.

[140] Vasilescu, B.; Posnett, D.; Ray, B.; van den Brand, M. G. J.; Serebrenik, A.; Devanbu, P.; Filkov, V. “Gender and tenure diversity in GitHub teams”. In: Proceedings of the Conference on Human Factors in Computing Systems, 2015, pp. 3789–3798.

[141] Vasilescu, B.; Serebrenik, A.; Filkov, V. “A Data Set for Social Diversity Studies of GitHub Teams”. In: Proceedings of the Working Conference on Mining Software Repositories, 2015, pp. 514–517.

[142] Viana, A. H.; Rosas Torres, A. R.; Álvaro Estramiana, J. L. “Egalitarian men: stereotypes and discrimination in the labor market”, *Acta Colombiana de Psicología*, May 2020, pp. 111–128.

[143] Vidal, E.; Castro, E.; Montoya, S.; Payihuanca, K. “Closing the gender gap in engineering: Students role model program”. In: Proceedings of the International Convention on Information, Communication and Electronic Technology, 2020, pp. 1493–1496.

[144] Villeval, M. C. “Ready, steady, compete”, *Science*, Feb 2012, pp. 544–545.

[145] Vogel-Heuser, B.; Obermeier, M.; Braun, S.; Sommer, K.; Jobst, F.; Schweizer, K. “Evaluation of a uml-based versus an iec 61131-3-based software engineering approach for teaching plc programming”, *IEEE Transactions on Education*, Nov 2013, pp. 329–335.

[146] Wang, Y.; Redmiles, D. “Implicit gender biases in professional software development: An empirical study”. In: Proceedings of the International Conference on Software Engineering: Software Engineering in Society, 2019, pp. 1–10.

[147] Webster, M. “The difference between race and ethnicity how they differ and overlap”. Source: https: //www.merriam-webster.com/words-at-play/difference-between-race-and-ethnicity, Sep 2021.

[148] Weinberg, G. M. “The Psychology of Computer Programming”. Van Nostrand Reinhold, 1971, 292p.

[149] Weisberg, S. “Linear hypothesis: Regression (basics)”. In: *International Encyclopedia of the Social Behavioral Sciences*, Smelser, N. J.; Baltes, P. (Editors), Oxford: Pergamon, 2001, pp. 8884–8888.

[150] Weisz, E.; Zaki, J. “Motivated empathy: a social neuroscience perspective”, *Current opinion in psychology*, Dec 2018, pp. 67—71.

[151] Whitworth, B. “The Social Requirements of Technical Systems”. IGI Global, 2011, chap. 97, pp. 1461–1481.

[152] Williams, M. “Facebook diversity update: Increasing representation in our workforce and supporting minority-owned businesses”. Source: https://about.fb.com/news/2021/07/facebook-diversity-report-2021/, Sep 2021.

[153] Wohlin, C.; Aurum, A. “Towards a decision-making structure for selecting a research design in empirical software engineering”, *Empirical Software Engineering*, Dec 2015, pp. 1427–1455.

[154] Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B., Wesslén, A. “Experimentation in Software Engineering”. Springer-Verlag Berlin Heidelberg, 2012, 236p.

[155] Yin, R. “Case Study Research: Design and Methods”. SAGE Publications, 2003, 181p.

[156] Zika-Viktorsson, A.; Sundström, P.; Engwall, M. “Project overload: An exploratory study of work and management in multi-project settings”, *International Journal of Project Management*, Jul 2006, pp. 385–394.

APPENDIX A – SYSTEMATIC LITERATURE MAPPING - LIST OF PAPERS

P[1] Catolino, G.; Palomba, F.; Tamburri, D. A. ; Serebrenik, A.; Ferrucci, F. "Gender Diversity and Community Smells: Insights From the Trenches", *IEEE Software*, vol. 37, Jan 2020, pp. 10-16.

P[2] Gilal A.R., Jaafar J., Omar M., Basri S., Aziz I.D.A., "A Set of Rules for Constructing Gender-Based Personality Types' Composition for Software Programmer", *Lecture Notes in Electrical Engineering*, vol. 520, Aug 2019, pp. 1-12.

P[3] Kohl, K.; Prikladnicki, R. "A systematic mapping study of diversity in software engineering: A perspective from the agile methodologies". In: IEEE/ACM 12th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE), 2019, pp. 7–10.

P[4] Grass, B.E., Coto, M., Collazos, C. (2019). "Academic Emotions in Programming Learning: Women's Impact on the Software Sector", *Communications in Computer and Information Science*, vol 847, Dec 2018, pp. 1-10.

P[5] Garcia-Holgado, A.; Vázquez-Ingelmo, A. ; Verdugo-Castro, S. ; González, C. ; Gómez, M. C. S.; Garcia-Peñalvo, F. J. "Actions to Promote Diversity in Engineering Studies: a Case Study in a Computer Science Degree", In: IEEE Global Engineering Education Conference (EDUCON), 2019, pp. 793-800.

P[6] Nguyen-Duc, A. ; Jaccheri, L. ; Abrahamsson, P. "An Empirical Study on Female Participation in Software Project Courses," In: IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion), 2019, pp. 240-241.

P[7] Buhnova B., Jurystova L., Prikrylova D. "Assisting women in career change towards software engineering: experience from Czechitas NGO". In: Proceedings of the 13th European Conference on Software Architecture (ECSA '19), 2019, pp. 88–93.

P[8] , Canedo E.D., Tives H.A., Marioti M.B., Fagundes F., de Cerqueira J.A.S. "Barriers Faced by Women in Software Development Projects", *Information*, Sep 2019, pp. 1-20.

P[9] Ford, D. ; Behroozi, M. ; Serebrenik, A. ; Parnin, C. "Beyond the Code Itself: How Programmers Really Look at Pull Requests". In: IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS), 2019, pp. 51-60.

P[10] Krüger S.; Hermann B. "Can an Online Service Predict Gender? On the State-of-the-Art in Gender Identification from Texts". In: IEEE/ACM 2nd International Workshop on Gender Equality in Software Engineering (GE), 2019, pp. 13-16.

P[11] Gilal R., Omar M., Gilal A.R., Md Rejab M., Waqas A., Sharif K.I.M. “Can Time Pressure and Personality Make any Sense together in Software Engineering?” *International Journal of Innovative Technology and Exploring Engineering*, Nov 2019, pp. 1071-1075.

P[12] Wurzelova, P. ; Palomba, F. ; Bacchelli, A. "Characterizing Women (Not) Contributing to Open-Source". In: IEEE/ACM 2nd International Workshop on Gender Equality in Software Engineering (GE), 2019, pp. 5-8.

P[13] Silveira, K.K.; Musse, S.; Manssour, I. H.; Vieira, R.; Prikladnicki, R. "Confidence in Programming Skills: Gender Insights From StackOverflow Developers Survey," In: IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion), 2019, pp. 234-235.

P[14] Bosu, A.; Sultana, K. Z. "Diversity and Inclusion in Open Source Software (OSS) Projects: Where Do We Stand?" In: ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), 2019, pp. 1-11.

P[15] Brockmann P.; Schuhbauer H.; Hinze A., "Diversity as an advantage: An analysis of career competencies for it students". In: 16th International Conference on Cognition and Exploratory Learning in Digital Age, 2019, pp. 1-8.

P[16] Murphy A.; Kelly B., Bergmann K.; Khaletskyy K.; O’Connor R.V.; Clarke P.M. "Examining Unequal Gender Distribution in Software Engineering". In: 26th European and Asian Conference on Systems, Software and Services Process Improvement , Sep 2019, pp. 1-14.

P[17] Paul, R. ; Bosu, A. ; Sultana, K. Z. . "Expressions of Sentiments during Code Reviews: Male vs. Female". In: Proceedings of the 26th IEEE International Conference on Software Analysis, Evolution and Reengineering, 2019, pp. 1-12.

P[18] Seibel, S.; Veilleux, N. "Factors influencing women entering the software development field through coding bootcamps vs. computer science bachelor’s degrees". *Journal of Computing Sciences in Colleges*, v. 34, Apr 2019, pp. 84–96.

P[19] Lee, A. ; Carver, J. C. "FLOSS Participants’ Perceptions About Gender and Inclusiveness: A Survey". In: Proceedings of the 41st International Conference on Software Engineering (ICSE), 2019, pp. 677–687.

P[20] Eiband, B.J.; Bergande, B.; Schedel, A.; Brune, P. "Game of Codes: Towards Learning Java by an Educational Mobile Game Adapted to Female Programming Novices". In: Proceedings of the Evaluation and Assessment on Software Engineering (EASE), 2019, pp. 362–363.

P[21] May, A.; Wachs, J.;Hannák, A. "Gender differences in participation and reward on Stack Overflow", *Empirical Software Engineering*, Feb 2019, pp. 1-23.

P[22] Bastarrica, M. C. ; Simmonds, J. "Gender Differences in Self and Peer Assessment in a Software Engineering Capstone Course". In: IEEE/ACM 2nd International Workshop on Gender Equality in Software Engineering (GE), 2019, pp. 29-32.

P[23] Bano, M. ; Zowghi, D. "Gender Disparity in the Governance of Software Engineering Conferences". In: IEEE/ACM 2nd International Workshop on Gender Equality in Software Engineering (GE), 2019, pp. 21-24.

P[24] Colomo-Palacios R.; Ben Yahia N.; Larrucea X. "Gender diversity among computing students: Reflections from Norway, Spain and Tunisia". In: Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'), 2019, pp. 196–200.

P[25] Catolino, G. ; Palomba, F.; Tamburri, D. A.; Serebrenik, A.; Ferrucci, F. :Gender Diversity and Women in Software Teams: How Do They Affect Community Smells?" In: Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS), 2019, pp. 11–20.

P[26] Widder, D. G. "Gender in Open Source Communities: Different Migration Patterns and Forms of Work". In: IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), 2019, pp. 241-242.

P[27] Carver, J. C. ; Serebrenik, A. "Gender in Software Engineering", *IEEE Software*, vol. 36, Nov 2019, pp. 76-78.

P[28] Wang, Y. "Gender Reputation Differences on Online Programming QA Communities", *Journal of Computer Information Systems*, Feb 2019, pp. 1-11.

P[29] Vedres B.; Vasarhelyi O. "Gendered behavior as a disadvantage in open source software development", *EPJ Data Science*, Jul 2019, pp. 1-18.

P[30] Qiu, H. S.; Nolte, A.; Brown, A.; Serebrenik, A.; Vasilescu, B. "Going Farther Together: The Impact of Social Capital on Sustained Participation in Open Source". In: IEEE/ACM 41st International Conference on Software Engineering (ICSE), 2019, pp. 688-699.

P[31] Ford, D.; Milewicz, R.; Serebrenik, A. "How Remote Work Can Foster a More Inclusive Environment for Transgender Developers". In: IEEE/ACM 2nd International Workshop on Gender Equality in Software Engineering (GE), 2019, pp. 9-12.

P[32] Wang, Y. ; Redmiles, D. "Implicit Gender Biases in Professional Software Development: An Empirical Study". In: EEE/ACM 41st International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS), 2019, pp. 1-10.

P[33] Imtiaz, N.; Middleton, J.; Chakraborty, J. ; Robson, N.; Bai,G.; Murphy-Hill, E. "Investigating the Effects of Gender Bias on GitHub". In: IEEE/ACM 41st International Conference on Software Engineering (ICSE), 2019, pp. 700-711.

P[34] Heels, L.; Devlin M. "Investigating the role choice of female students in a software engineering team project". In: Proceedings of the 3rd Conference on Computing Education Practice(CEP), 2019, pp. 1-4.

Kofink A., "Contributions of the under-appreciated: Gender bias in an open-source ecology". In: Companion Proceedings of ACM SIGPLAN International Conference on Systems, Programming, Languages and Applications: Software for Humanity, 2015, pp. 83-84.

P[36] Izquierdo, D.; Huesman, N.; Serebrenik, A.; Robles, G. "OpenStack Gender Diversity Report", *IEEE Software,* Jan 2019, pp. 28-33.

P[37] Santos Machado,L.; Perlin,M.; Colla Soletti, R.; Kmetzch Rosa e Silva, L.; Doerderlein Schwartz, I. V. ; Seixas, A.; Klein Ricachenevsky, F.; Tamajusuku Neis, A.; Staniscuaski, F. "Parent in Science: The Impact of Parenthood on the Scientific Career in Brazil". In: IEEE/ACM 2nd International Workshop on Gender Equality in Software Engineering (GE), 2019, pp. 37-40.

P[38] Blincoe, K.; Springer, O. ; Wrobel, M. R. "Perceptions of Gender Diversity's Impact on Mood in Software Development Teams". *IEEE Software*, Sep 2019, pp. 51-56.

P[39] Silveira K.K.; Musse S.; Manssour I.; Vieira R.; Prikladnicki R. "Reinforcing diversity company policies: Insights from Stackoverflow developers survey". In: Proceedings of the 21st International Conference on Enterprise Information Systems, 2019, pp. 119-129.

P[40] Alharthi A.D.; Alsanoosy T.; Spichkova M.; Hamilton M.; "Social Position and Gender Perspectives of eLearning Systems: A Study of Social Sustainability". In: Lecture Notes in Information Systems and Organisation, Apr 2019, pp. 169-185.

P[41] Hyrynsalmi, S. "Software Engineering Studies Attractiveness for the Highly Educated Women Planning to Change Career in Finland". In: IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion), 2019, pp. 304-305.

P[42] Hyrynsalmi S.; Sutinen E. "The role of women software communities in attracting more women to the software industry". In: IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), 2019, pp. 1-7.

P[43] Patitsas, E. "The Social Closure of Undergraduate Computing: Lessons for the Contemporary Enrolment Boom". In: IEEE/ACM 2nd International Workshop on Gender Equality in Software Engineering (GE), 2019, pp. 33-36.

P[44] Hyrynsalmi, S. M. "The Underrepresentation of Women in the Software Industry: Thoughts from Career-Changing Women". In: Proceedings of the 2nd International Workshop on Gender Equality in Software Engineering (GE), 2019, pp. 1-4.

P[45] Hyrynsalmi, S. "What motivates adult age women to make a career change to the software industry?" In: IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), 2019, pp. 1-8.

P[46] Vandana, S. "Women Participation in Open Source Software Communities". In: Proceedings of the 13th European Conference on Software Architecture, 2019, pp. 94-99.

P[47] Buhnova, B. ; Prikrylova, D. "Women Want to Learn Tech: Lessons from the Czechitas Education Project". In: IEEE/ACM 2nd International Workshop on Gender Equality in Software Engineering (GE), 2019, pp. 25-28.

P[48] Singh, V. "Women-Only Spaces of Open Source". In: IEEE/ACM 2nd International Workshop on Gender Equality in Software Engineering (GE), 2019, pp. 17-20.

P[49] Janzen, D. S.; Bahrami, S.; Silva, B. C.; Falessi, D. "A Reflection on Diversity and Inclusivity Efforts in a Software Engineering Program". In: IEEE Frontiers in Education Conference (FIE), 2018, pp. 1-9.

P[50] Bastarrica, M. C.; Hitschfeld, N. ; Marques Samary, M. ; Simmonds, J. "Affirmative Action for Attracting Women to STEM in Chile". In: IEEE/ACM 1st International Workshop on Gender Equality in Software Engineering (GE), 2018, pp. 45-48.

Terrell J., Kofink A., Middleton J., Rainear C., Murphy-Hill E., Parnin C., Stallings J. "Gender differences and bias in open source: Pull request acceptance of women versus men". *PeerJ Computer Science*, May 2017, pp. 1-30.

P[52] Borsotti V., "Barriers to gender diversity in software development education: Actionable insights from a danish case study". In:Proceedings of the 40th International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET), 2018, pp. 146-152.

P[53] Shekhar A., Marsden N., "Cognitive walkthrough of a learning management system with gendered personas". In: Proceedings of the 4th Conference on Gender & IT (GenderIT), 2018. pp. 191-198.

P[54] Wang, Z.; Wang, Y. ; Redmiles, D. "Competence-Confidence Gap: A Threat to Female Developers' Contribution on GitHub".In: IEEE/ACM 40th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS), 2018, pp. 81-90.

P[55] Robson, N. "Diversity and Decorum in Open Source Communities". In: Proceedings of 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE), 2018, pp. 986–987.

P[56] Nguyen-Duc A.; Khodambashi S.; Gulla J.A.' Krogstie J.; Abrahamsson P. "Female leadership in software projects—A preliminary result on leadership style and project context factors, Studies in Computational Intelligence". *Studies in Computational Intelligence*, 2018, pp. 149-163.

P[57] Reeves, J. "Gender Equality in Software Engineering". In: IEEE/ACM 1st International Workshop on Gender Equality in Software Engineering (GE), 2018, pp. 33-36.

P[58] Raura G.; Fonseca C E.R.; Castro J.W.; Gualotuña T.; Rebeca Mejía C.; Mónica Santillán T.; Pons C., Dieste O. "Gender gap in computing: A preliminary empirical study". In: 21st Conferencia Iberoamericana en Software Engineering (CIbSE), 2018, pp. 57-70.

P[59] Mendez, C.; Sarma, A.; Burnett, M. "Gender in Open Source Software: What the Tools Tell". In: IEEE/ACM 1st International Workshop on Gender Equality in Software Engineering (GE), 2018, pp. 21-24.

P[60] Aller C.F.; Navarro S.R. "Gender in software engineering degrees". In: Proceedings of the 12th European Conference on Software Architecture: Companion Proceedings (ECSA), 2018, pp. 1–4.

P[61] Carver J., Capilla R., Penzenstadler B., Serebrenik A., Valdezate A., Gender, "Sentiment and Emotions, and Safety-Critical Systems". *IEEE Software*, Nov 2018, pp. 16-19.

P[62] Clarke L.A.; Pollock L.; Stout J.G.; Ellis C.; Camp T.; Bizot B.; McKinley K.S. "Improving diversity in computing research: An overview of CRA-W activities". In: IEEE/ACM 1st International Workshop on Gender Equality in Software Engineering (GE), 2018, pp. 41-44.

P[63] Garcia-Holgado A.; Mena J.; Garcia-Penalvo F.J.; Gonzalez C. "Inclusion of gender perspective in Computer Engineering careers: Elaboration of a questionnaire to assess the gender gap in tertiary education". In: IEEE Global Engineering Education Conference (EDUCON), 2018, pp. 1547-1554.

P[64] Bennaceur, A. ; Cano, A. ; Georgieva, L. ; Kiran, M.; Salama, M. ; Yadav, P. "Issues in Gender Diversity and Equality in the UK". In: IEEE/ACM 1st International Workshop on Gender Equality in Software Engineering (GE), 2018, pp. 5-9.

P[65] Draude C.; Maab S. "Making IT work integrating gender research in computing through a process model". In: Proceedings of the 4th Conference on Gender IT (GenderIT), 2018, pp. 43–50.

P[66] Gren, L. "On Gender, Ethnicity, and Culture in Empirical Software Engineering Research". In: IEEE/ACM 11th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE), 2018, pp. 77-78.

P[67] Mendez, C.; Padala, H. S.; Steine-Hanson, Z.; Hildebrand, C.; Horvath, A.; Hill, C.; Simpson, L.; Patil, N.; Sarma, A.; Burnett, M. "Open Source Barriers to Entry, Revisited: A Sociotechnical Perspective". In: IEEE/ACM 40th International Conference on Software Engineering (ICSE), 2018, pp. 1004-1015.

P[68] Kohl, K.; Prikladnicki, R. "Perceptions on Diversity in Brazilian Agile Software Development Teams: A Survey". In: IEEE/ACM 1st International Workshop on Gender Equality in Software Engineering (GE), 2018, pp. 37-40.

P[69] Borsotti, V. "Barriers to Gender Diversity in Software Development Education: Actionable Insights from a Danish Case Study". In: Proceedings of the 40th International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET), 2018, pp. 146–152.

P[70] Jász, J. ; Beszédes, Á. "Software Testing Conferences and Women". In: IEEE/ACM 1st International Workshop on Gender Equality in Software Engineering (GE), 2018, pp. 17-20.

P[71] Castro L.M. "Teaching the next generation of software architects: A gender-focused survey on worldwide curricula". In: Proceedings of the 12th European Conference on Software Architecture: Companion Proceedings (ECSA), 2018, pp. 1–4.

P[72] Ahmar Y.E.; Pallec X.L.; Gérard S. "The visual variables in UML: How are they used by women?" In: Proceedings of the 12th European Conference on Software Architecture: Companion Proceedings (ECSA), 2018, pp. 1–5.

P[73] Ribaupierre, H.; Jones, K.; Loizides, F.; Cherdantseva, Y. "Towards Gender Equality in Software Engineering: The NSA Approach". In: IEEE/ACM 1st International Workshop on Gender Equality in Software Engineering (GE), 2018, pp. 10-13.

P[74] Wang, Y. "Understanding the Reputation Differences between Women and Men on Stack Overflow". In: 25th Asia-Pacific Software Engineering Conference (APSEC), 2018, pp. 436-444.

P[75] Gómez O.S.; Solari M.; Pardo C.J.; Ledezma A.C. "A controlled experiment on productivity of pair programming gender combinations: Preliminary results". In: , In: XX Conferencia Iberoamericana en Software Engineering (CIbSE), 2017, pp. 679-692.

P[76] Burnett, M.; Counts, R.; Lawrence, R.; Hanson, H. "Gender HCI and microsoft: Highlights from a longitudinal study". IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), 2017, pp. 139-143.

P[77] Savannah, M. "How Are Programming Questions from Women Received on Stack Overflow? A Case Study of Peer Parity". In: Proceedings Companion of ACM SIGPLAN International Conference on Systems, Programming, Languages, and Applications: Software for Humanity (SPLASH Companion), 2017, pp. 39–41.

P[78] Søndergaard M.L.J. "Intimate design: Designing intimacy as a critical-feminist practice". In: Proceedings of Conference Extended Abstracts on Human Factors in Computing Systems (CHI), 2017, pp. 320–325.

P[79] Agarwal, S.; Mittal, N.; Sureka, A. "Minority Ethnic Groups in Computer Science Research: What is the Bibliography Data Telling Us?", *ACM SIGCAS Computers and Society*, vol. 47, Jun 2017, pp. 5–15.

P[80] Spichkova M.; Schmidt H.; Trubiani C. "Role of women in software architecture: An attempt at a systematic literature review". In: Proceedings of the 11th European Conference on Software Architecture: Companion Proceedings (ECSA), 2017, pp. 31–34.

P[81] Ford D.; Harkins A.; Parnin C. "Someone like me: How does peer parity influence participation of women on stack overflow?" In: IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), 2017, pp. 239-243.

P[82] Andrejczuk E.; Roig C.; Rodrfguez-Aguilar J.A.; Sierra C. "Synergistic team composition". In: Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), 2017, pp. 1463-1465.

P[83] James, T.; Galster, M.; Blincoe, K.; Miller, G. "What is the perception of female and male software professionals on performance, team dynamics and job satisfaction? Insights from the trenches". In: IEEE/ACM 39th International Conference on Software Engineering: Software Engineering in Practice Track (ICSE-SEIP), 2017, pp. 13-22.

P[84] Gilal A.R.; Jaafar J.; Omar M.; Basri S.; Waqas A. "A rule-based model for software development team composition: Team leader role with personality types and gender classification". *Information and Software Technology*, Jun 2016, pp. 105-113.

P[85] Brinkman B.; Diekman A. "Applying the communal goal congruity perspective to enhance diversity and inclusion in undergraduate computing degrees". In: Proceedings of the 47th ACM Technical Symposium on Computing Science Education (SIGCSE), 2016, pp. 102–107.

P[86] Gilal A.R.; Jaafar J.; Omar M.; Basri S.; Aziz I.A. "Balancing the personality of programmer: Software development team composition", *Malaysian Journal of Computer Science*, Jun 2016, pp. 145–155.

P[87] Razavian, M.; Lago, P. "Feminine Expertise in Architecting Teams", *IEEE Software*, vol. 33, Jul 2016, pp. 64-71.

P[88] Hamilton, M.; Luxton-Reilly, A.; Augar, N.; Chiprianov, V.; Gutierrez, E.; Duarte, E.; Hu, H. H.; Ittyipe, S.; Pearce, J. L.; Oudshoorn, M.; Wong, E. "Gender Equity in Computing: International Faculty Perceptions and Current Practices". In: International Faculty Perceptions and Current Practices. In Proceedings of ITiCSE Working Group Reports (ITiCSE), 2016, pp. 81–102.

P[89] Hill, C.; Ernst, S.; Oleson, A.; Horvath, A.; Burnett, M. "GenderMag experiences in the field: The whole, the parts, and the workload". In: IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), 2016, pp. 199-207.

P[90] Burnett M.; Stumpf S.; Macbeth J.; Makri S.; Beckwith L.; Kwan I.; Peters A.; Jernigan W. "GenderMag: A method for evaluating software's gender inclusiveness", *Interacting with Computers*, Oct 2016, pp. 760-787.

P[91] Sudbery C. "How XP can improve the experiences of female software developers", *Lecture Notes in Business Information Processing*, May 2016, pp. 261-269.

P[92] Parra, E.; Haiduc, S.; James, R. "Making a Difference: An Overview of Humanitarian Free Open Source Systems". In: IEEE/ACM 38th International Conference on Software Engineering Companion (ICSE-C), 2016, pp. 731-733.

P[93] Gilal A.R.; Jaafar J.; Basri S.; Omar M.; Tunio M.Z. "Making programmer suitable for team-leader: Software team composition based on personality types". In: International Symposium on Mathematical Sciences and Computing Research (iSMSC), 2015, pp. 78-82.

P[94] Ford D.; Smith J.; Guo P.J.; Parnin C. "Paradise unplugged: Identifying barriers for female participation on stack overflow". In: In Proceedings of 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering (FSE), 2016, pp. 846–857.

P[95] Lin, B.; Serebrenik, A. "Recognizing Gender of Stack Overflow Users". In: IEEE/ACM 13th Working Conference on Mining Software Repositories (MSR), 2016, pp. 425-429.

P[96] Robles G.; Reina L.A.; González-Barahona J.M.; Domínguez S.D. "Women in free/libre/open source software: The situation in the 2010s", *IFIP Advances in Information and Communication Technology*, May 2016, pp. 163–173.

P[97] Choi K.S. "A comparative analysis of different gender pair combinations in pair programming". *Behaviour and Information Technology*, Jun 2015, pp. 825-837.

P[98] Vasilescu, B.; Serebrenik, A.; Filkov, V. "A Data Set for Social Diversity Studies of GitHub Teams". In: In Proceedings of the 12th Working Conference on Mining Software Repositories (MSR), 2015, pp. 514–517.

P[90] Kotamraju N.P. "Playing stupid, caring for users, and putting on a good show: Feminist acts in usability study work", *Interacting with Computers*, Apr 2011, pp. 439-446.

P[100] Gramß, D.; Vogel-Heuser, B. "Contribution of personal factors for a better understanding of the gender effects of freshmen in mechanical engineering". In: IEEE International Conference on Industrial Technology (ICIT), 2015, pp. 3258-3263.

P[101] Lyon, L. A.; Jameson, K. "From clicks to code: Resources women use to learn to code in apex". In: IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), 2015, pp. 303-304.

P[102] Vasilescu B.; Posnett D.; Ray B.; Van Den Brand M.G.J.; Serebrenik A.; Devanbu P.; Filkov V. "Gender and tenure diversity in github teams". In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI), 2015, pp. 3789–3798.

P[103] Weilemann E.; Brune P. "Less distress with a scrum mistress? on the impact of females in agile software development teams". In: Proceedings of 24th Australasian Software Engineering Conference (ASWEC), 2015, pp. 3–7.

P[104] Marques, M. "Software engineering education — Does gender matter in project results? — A Chilean case study". In: IEEE Frontiers in Education Conference (FIE), 2015, pp. 1-8.

P[105] Williams G. "Are you sure your software is gender-neutral?" *Interactions*, Jan 2014, pp. 36-39.

P[106] Gramß, D.; Frank, T. ; Rehberger, S. ; Vogel-Heuser, B. "Female characteristics and requirements in software engineering in mechanical engineering". In: International Conference on Interactive Collaborative Learning (ICL), 2014, pp. 272-279.

P[107] Vasilescu B.; Capiluppi A.; Serebrenik A.; "Gender, representation and online participation: A quantitative study", *Interacting with Computers*, Sep 2014, pp. 488-511.

P[108] Vasilescu B. "Human aspects, gamification, and social media in collaborative software engineering". In: Companion Proceedings of the 36th International Conference on Software Engineering (ICSE Companion), 2014. pp. 646–649.

P[109] Gila, A. R.; Jaafa, J.; Omar, M.; Tunio, M. Z. "Impact of personality and gender diversity on software development teams' performance". In: International Conference on Computer, Communications, and Control Technology (I4CT), 2014, pp. 261-265.

P[110] Rajagopalan, S.; Rajamani, L. "A Fuzzy Logic Rule Based Forecasting Model: Work-Life Balance in IT among Software vs. Services Industry on the View of Women Software Engineer". In: International Conference on Machine Intelligence and Research Advancement, 2013, pp. 241-246.

P[111] Li, Z.; Plaue, C.; Kraemer, E. "A spirit of camaraderie: The impact of pair programming on retention". In: 26th International Conference on Software Engineering Education and Training (CSEE&T), 2013, pp. 209-218

P[112] Zeid, A.; El-Bahey,R. "Establishing a global software development course: A cultural perspective". In: IEEE Frontiers in Education Conference (FIE), 2013, pp. 1695-170.

P[113] Moon, E. "Gendered Patterns of Politeness in Free/Libre Open Source Software Development". In: 46th Hawaii International Conference on System Sciences, 2013, pp. 3168-3177

P[114] Hemmendinger, D. "The computer boys take over: computers, programmers, and the politics of technical expertise". MIT Press, 2013, 336p.

P[115] Judy, K. H. "Agile Values, Innovation and the Shortage of Women Software Developers". In: 5th Hawaii International Conference on System Sciences, 2012, pp. 5279-5288.

P[116] Fernández-Sanz, L. ; Misra, S. "Analysis of cultural and gender influences on teamwork performance for software requirements analysis in multinational environments", *IET Journals*, Jun 2012, pp. 167-175.

P[117] Kuechler V.; Gilbertson C.; Jensen C. "Gender differences in early free and open source software joining process". In: IFIP Advances in Information and Communication Technology, 2012, pp. 78-93.

P[118] Vela B.; Cáceres P.; Cavero J.M. "Participation of women in software engineering publications", *Scientometrics*, May 2012, pp. 661-679.

P[119] Sharafi, Z.; Soh, Z.; Guéhéneuc, Y.; Antoniol, G. "Women and men — Different but equal: On the impact of identifier style on source code reading". In: 20th IEEE International Conference on Program Comprehension (ICPC), 2012, pp. 27-36.

P[120] Mahmod M.; Dahalin Z.M. "Women in open source software innovation process: Where are they?" *Journal of Information and Communication Technology*, Mar 2012, pp. 113–129.

P[121] Narasimhan, V. L. "A subjective perspective on genderization issues in software development life cycle". In: International Conference on Recent Trends in Information Technology (ICRTIT), 2011, pp. 1335-1340.

P[122] Sahin Y.G. "A team building model for software engineering courses term projects", *Computers and Education*, Apr 2011, pp. 916-922.

P[123] Zeid, A.; El-Bahey, R. "Impact of introducing single-gender classrooms in higher education on student achievement levels: A case study in software engineering courses in the GCC region". In: Frontiers in Education Conference (FIE), 2011, pp. T2H-1-T2H-6.

P[124] Dou W.; He W. "Compatibility and requirements analysis of distributed pair programming". In: Second International Workshop on Education Technology and Computer Science, 2010, pp. 467-470.

P[125] Burnett M.; Fleming S.D.; Iqbal S.; Venolia G.; Rajaram V.; Farooq U.; Grigoreanu V.; Czerwinski M. "Gender differences and programming environments: Across programming populations". In: Proceedings of ACM-IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), 2010, pp. 1–10.

P[126] Mahmod, M.; Yusof, S. A. M.; Dahalin, Z. M. "Women contributions to open source software innovation: A social constructivist perspective". In: International Symposium on Information Technology, 2010, pp. 1433-1438.

APPENDIX B – CASE STUDY PROTOCOL

This protocol is based on the model proposed by Brereton et al. [20] except for the Ethics Considerations section, added by the author, which is mentioned by Runenson et al. [116, 117]

The protocol established the steps to conduct the case study that partially answers the research questions proposed in this thesis. Broadly, the goal is to understand the effect of gender diversity on software development teams from the point of view of Pull Request metrics as Pull Request Size and Pull Request Time to Merge, as use the semantics of commit messages to extract qualitative information from the teams.

B.1 Research Question

As already mentioned, the protocol established the steps to conduct the case study that partially answers the research questions proposed in this thesis. The main research question is:

What are the effects of gender diversity on the performance and results of software development teams?

To help to answer this research question, four sub-questions were defined and presented in Chapter 1. This case study helps to answer the **RQ1** and **RQ1.1** sub-questions, which follows:

- **RQ1** *What are the effects of gender diversity on software development teams when we analyzed pull based metrics as Pull Request Size and Pull Request Time to Merge?*
  - **RQ1.1** *Is there any other factor than gender diversity that has effect in Pull Request Size and Pull Request Time to Merge?*

B.2 Case Study Design

The importance of collecting empirical evidence in studies involving human and social aspects in software engineering is known. The influence of gender diversity on the daily activities and deliveries of software development teams fits into this category of studies. What metrics, factors, and perceptions bring benefits or impact gender diversity in software development teams?

B.3 Unit Analysis

According to Yin [155], we can classify this study as a ***single embedded case study***, that is, multiple units of analysis and a single case. Each software development team participating in the research is considered as a ***unit of analysis***, and then, it is possible to compare them. All teams belong to the same technology company, characterizing the single embedded case study.

As a software development team, we are considering the group of individuals playing different roles and functions within a team, as software engineers, testers, DevOps, DevSecOps, *product owners, scrum masters*, business analysts, *agile coaches*, people managers, UX *Designers*, etc.).

Factor commonly used in software development we considered for collection: pull requests.

Software Development Team is the group of software engineers, in their different roles, responsible for designing and building the software products required by stakeholders. We considered the size of the team and the roles and responsibilities of each member.

Pull Request Size is the average of the total lines of code changed (added + removed). The highest the number, the bigger is the pull request. Large pull requests carry more risk when deploying to production and are more challenging to review, merge, and release. Deploying pull requests of a reasonable size enables the team to review and ship new features at a faster cadence and with greater confidence [56].

Pull Request Lead Time is the time between the first commit on a branch and the merge action of a pull request on that branch. The timestamp of the first commit on a branch is subtracted from the timestamp on the merge action of the pull request [56].

B.4 Replication Strategy

In the case of other researchers got interested in replicating the study they must run this protocol for different Software Development Team of a technology company. It is essential that researchers interested in replicating the experiment check the applicable rules related to their respective ethics committees.

B.5 Data Collection

Data collection occurred during the year 2020. We used third-degree techniques [93], extracting data from code repositories (GitHub).

B.6 Data Collection Plan

The second semester of 2020 was used to create Python scripts to retrieve data from code repositories. In January and February of 2021, we ran the scripts to collect data from 14 software development teams with different numbers of members and different gender distributions.

B.7 Data Storage

After anonymization, organization, cleaning, and completion, the data will be stored and shared in GitHub repositories with the scientific community.

B.8 Data Analysis

Runeson et al. [117] say the amount and types of data being collected can influence the types of analyzes that will be performed. During protocol development, it may be premature to know which specific analytical techniques will be used. However, the researcher must be able to anticipate the types of techniques. The nature of the case studies, including the volume and variety of data collected during the case studies, means that researchers will need to design their study to handle large, complex, and diverse datasets. For the analysis we identified the need to use descriptive statistics and linear regression analysis.

Descriptive statistics, such as mean values, standard deviations, histograms, and scatter plots, are used to understand the data collected. Regression analysis is a well-known statistical learning technique useful to infer the relationship between a dependent variable Y and independent variables X [6]. In regression analysis you have your dependent variable — the main factor that you are trying to understand or predict, and then you have your independent variables — the factors you suspect have an impact on your dependent variable [55]. Hypothesis testing is conducted to determine if there is a significant effect of one

or several variables (independent variables) on one or several other variables (dependent variables) [154].

B.9 Tools to Support Quantitative Analysis

To support the quantitative analysis in this work, we used Python scripts and scientific libraries like Pandas, Numpy, Scipy, Statsmodel, and scikit-learn.

B.10 Validity

The validity of a study denotes the trustworthiness of the results and to what extent the results are true and not biased by the researchers' subjective point of view [154].

There are different ways to classify aspects of validity and threats to validity in the literature [154]. For this research, we used a classification scheme that is suggested by Wohlin et al. [154], and Yin [155], as follows:

- **Construct Validity**: This aspect of validity reflects what extent the operational measures that are studied represent what the researcher has in mind and what is investigated according to the research questions. If, for example, the constructs discussed in the interview questions are not interpreted in the same way by the researcher and the interviewed persons, there is a threat to construct validity [154].
- **Internal Validity**: This aspect of validity is of concern when causal relations are examined. When the researcher investigates whether one factor affects an investigated factor, there is a risk that the investigated factor is also affected by a third factor. If the researcher is not aware of the third-factor andor does not know to what extent it affects the investigated factor, there is a threat to internal validity [154].
- **External Validity**: This aspect of validity is concerned with the extent to which it is possible to generalize the findings and the extent to which the findings are of interest to other people outside the investigated case. During analysis of external validity, the researcher tries to analyze to what extent the findings are of relevance for other cases. In case studies, there is no population from which a statistically representative sample has been drawn. However, for case studies, the intention is to enable analytical generalization where the results are extended to cases that have common characteristics and hence for which the findings are relevant, i.e., defining a theory [154].
- **Reliability**: This aspect is concerned with to what extent the data and the analysis are dependent on the specific researchers. Hypothetically, if another researcher, later on,

conducted the same study, the result should be the same. Threats to this aspect of validity are, for example, if it is not clear how to code collected data or if questionnaires or interview questions are unclear [154].

B.10.1 Internal Validity

This aspect of validity is of concern when causal relations are examined. When the researcher investigates whether one factor affects an investigated factor, there is a risk that the investigated factor is also affected by a third factor. If the researcher is not aware of the third-factor andor does not know to what extent it affects the investigated factor, there is a threat to internal validity [154].

Regarding internal validity, possible threats that we can anticipate for this work:

- *Researcher bias:* refers to the potential bias that the author of the study may have when interpreting or synthesizing the extracted results. This type of bias can occur in relation to a topic or because only one author worked on data synthesis. In this case, to reduce the risk of the threat, the author will send the research to other researchers for *feedback*.

- *Change in teams:* an individual leaves the development team or a new one joins. This can influence how study participants feel and act.

B.10.2 External Validity

This aspect of validity is concerned with the extent to which it is possible to generalize the findings and the extent to which the findings are of interest to other people outside the investigated case. During analysis of external validity, the researcher tries to analyze to what extent the findings are of relevance for other cases. In case studies, there is no population from which a statistically representative sample has been drawn. However, for case studies, the intention is to enable analytical generalization where the results are extended to cases that have common characteristics and hence for which the findings are relevant, i.e., defining a theory [154].

Concerning external validity, a possible threat that we can anticipate for this work is there are not enough software development teams for analysis. Since the subject is sensitive and diversity culture and policies may vary across companies, the survey can generate better analytical generalization with the participation of different teams from different companies;

B.10.3 Reliability

This aspect is concerned with to what extent the data and the analysis are dependent on the specific researchers. Hypothetically, if another researcher, later on, conducted the same study, the result should be the same. Threats to this aspect of validity are, for example, if it is not clear how to code collected data or if questionnaires or interview questions are unclear [154].

B.11 Reports

An empirical study cannot be distinguished from its reports. The report communicates the results of the study, but it is also the main source of information to assess the quality of the study [154].

A variety of reports should be produced during and after the case study, including:

- A series of interim reports for research colleagues and doctoral advisors;
- Reports for teams and companies that support the survey
- Submission of papers to conferences and academic journals reporting intermediate results of the case study;
- Submission of papers to conferences and academic journals after completion of the case study;
- The doctoral thesis.

B.12 Schedule

The schedule for the execution of this research protocol is presented together with the schedule for this thesis in Chapter 3.

APPENDIX C – SURVEY PROTOCOL

Kitchenham et al. [108] say the survey is not just the instrument (the questionnaire or checklist) for gathering information. It is a comprehensive system for collecting data to describe, compare or explain knowledge, attitudes, and behavior.

This protocol established the steps to conduct a survey study that partially answers the research questions proposed in this thesis and it is based on the model proposed by Brereton et al. [20] except for the Ethics Considerations section, added by the author, which is mentioned by Runenson et al. [116, 117].

Kitchenham[108] say the survey instrument is part of a larger survey process with clearly defined activities we describe below:

C.1 Setting specific, measurable objectives

As already mentioned, the protocol established the steps to conduct the survey that partially answers the research questions proposed in this thesis. The main research question is:

What are the effects of gender diversity on the performance and results of software development teams?

To help to answer this research question, four sub-questions were defined and presented in Chapter 1. This survey helps to answer the **RQ2 and RQ3** sub-questions, which follows:

- **RQ2** What are the perceived benefits of gender diversity on software development teams reported by individuals?
- **RQ3** What are the perceived difficulties of gender diversity on software development teams reported by individuals?

C.2 Planning and scheduling the survey

The survey was planned, administered, and analyzed following the schedule presented in Chapter 3, Figure 3.2.

C.3 Ensuring that appropriate resources are available

To support the execution of the Survey process, we used Qualtrics [1], which is a use web-based survey tool to conduct survey research, evaluations, and other data collection activities. To support the qualitative analysis of this work, we used the MAXQDA tool [2].

C.4 Designing the survey

To design the survey, we used the empathy map canvas technique to inspire the questions. The original idea was to use the empathy map in interviews, focus groups, and surveys. However, in 2020, the Covid-19 pandemic forced a drastic increase in the number of videoconference meetings, which impacted our process of interviewing people and doing focus groups. Bailenson [11] says videoconferencing was a critical tool that allowed schools and many businesses to continue working during shelter-in-place, and Zoom, in particular, helped hundreds of millions of people by making video conferencing free and easy to use. However, something about being on videoconference all day seems particularly exhausting, and the term "Zoom Fatigue" caught on quickly. Bailenson [11] says outlining nonverbal overload as a possible explanation for Zoom Fatigue in both work and social life. So, we opted for using only the survey through the web where respondents could answering asynchronously. The idea was to minimize the fatigue and exhaustion of people who were willing to support the research.

The empathy map technique was created by Dave Gray in 2009 as part of a user-centric design toolkit called *Gamestorming* [63]. It's a method for understanding the audience, including users, customers, and other participants in any business ecosystem. It helps teams develop a deep and shared understanding and empathy for others. The empathy map allows exploring the external observable world and the user's internal mindset: what the user is doing, seeing, hearing, thinking, and feeling (including pain and gains).

C.4.1 Informed Consent Form

Before answering the survey, respondents needed to agree with the following Informed Consent Form.

[1] https://pucrs.qualtrics.com/
[2] https://www.maxqda.com/

AN EMPIRICAL STUDY ON THE IMPACT OF GENDER DIVERSITY IN SOFTWARE DEVELOPMENT TIMES

Please read and, if deemed appropriate, express your consent to participate in this survey before starting, selecting the ACCEPT option presented at the end of this term.

Please, if you have any questions before, during or after your participation, send an email to karina.kohl@edu.pucrs.br and / or rafael.prikladnicki@pucrs.br

CONSENT TO PARTICIPATE IN RESEARCH

You are being invited to participate in a research developed by Karina Kohl Silveira (doctoral student) and Rafael Prikladnicki (professor advisor), “An Empirical Study on the Impact of Gender Diversity in Software Development Teams”, from the School of Technology at PUCRS - Brazil.

The objective is to observe which are the factors related to software engineering that most benefit from gender diversity in software development teams.

You should read the information below and if you do not understand or have any questions you can contact us by e-mail before deciding whether or not to participate.

This survey is voluntary and you have the right to stop participating at any time for any reason.

It will take you about 30 minutes (estimated time to complete the survey).

You will not receive any reward for this.

No identifying information about you will be included in publications that may result from this research.

The collection of information for this project will be completed in December 2021 and all information obtained will be stored securely for a period of one year after this date or until the completion of the written work.

By continuing this survey, you agree that any questions you had were adequately answered and that you agree to participate in this study. I understood the procedures described above.

C.4.2 Questions

Below, we list the demographic questions and the empathy map questions as defined by Dave Gray [62], as we included in the survey:

Demographic Questions

- Name (Optional)
- Email (Optional)

- Education (Desirable)
- Age (Desirable)
- Gender (Required)
- Race / Ethnicity (Desirable. By answering this question, you help to consider aspects of Intersectionality in the study)
- Name of the company where you work (Optional)
- How big is the company? (may be an approximate size) (Desirable)
  - 1 - 10 employees
  - 11 - 50 employees
  - 51 - 250 employees
  - 251 - 1000 employees
  - 1001 - 2000 employees
  - more than 2000 employees
- Briefly describe the company's main focus: (Desirable)
- How is your team organized? (Required)
  - Locally
  - Distributed in the national territory
  - Globally distributed
- Is the team located in the same physical space? (Required)
  - Yes
  - No
  - Normally yes, not during the Covid-19 pandemic

Professional Experience

- How long do you work with technology? (Required)
- How long do you work at the company you are currently in? (Desirable)
- How long have you been part of your team? (Required)
- What role do you play? (Required)
  - Business Analyst

  - Project Leader / Project Manager
  - Software architect
  - Developer
  - DevOps
  - UX
  - Product Owner / Product Manager
  - Scrum Master
  - Other. Which one?

- How do you evaluate your experience in this role? (Desirable)
  - Beginner (up to 1 year experience)
  - Experienced (1 - 3 years experience)
  - Specialist (over 3 years of experience)

About your team

- How many people in your team currently? (Required)
- What methodology is used? (Desirable)
  - Waterfall
  - scrum
  - Kanban
  - SAFe
  - Other. Which one?

Empathy Map Questions

The next questions are related to the empathy map on gender diversity in software development teams. Answer the questions thinking about you and the gender you identify with. If in any question you find it important to score something about your race / ethnicity, the study will become enriched in terms of intersectionality.

- *What do you need to do in your daily activities?*:
  - What types of tasks do you want to do or need to do in the context of your team? (Mandatory)
  - What types of tasks do you want to do or need to do in the context of your team? (Mandatory)

  - What kind of decisions do you need to make in your daily activities? (Mandatory)
  - How do you know that you have succeeded in the tasks you performed? (Mandatory)

- *What do you see?*:
  - What do you see in the job market in terms of gender diversity in software development teams? (Mandatory)
  - What do you see in your immediate environment when you observe gender diversity in software development team? (Mandatory)
  - What do you see others saying or doing about gender diversity in software development teams? (Mandatory)
  - What do you usually see and read about gender diversity in software development teams? (Mandatory)
  - Who's around you? (Mandatory)
  - Who are your friends / colleagues? (Mandatory)
  - What kind of tasks are you exposed to on a daily basis? (Mandatory)
  - What kind of problems do you have to perform these tasks? (Mandatory)

- *What do you say?*:
  - What do you usually say about gender diversity in software development teams? (Mandatory)
  - What do you imagine others saying about gender diversity in software development teams? (Mandatory)

- *What do you do?*:
  - What is your observed behavior about gender diversity in software development teams? Do you actively do something about it that other people notices? (Mandatory)
  - What is your attitude towards gender diversity in software development teams)? (Mandatory)
  - What can you be saying to other people through your behavior towards gender diversity in software development teams? (Mandatory)

- *What do you LISTEN to?*:
  - What do you hear others say about gender diversity in software development teams? (Mandatory)

  - What do you hear friends say about gender diversity in software development teams? (Mandatory)
  - What do you hear your colleagues say about gender diversity in software development teams? (Mandatory)
  - What do you hear other people talking about gender diversity in software development teams? (Mandatory)

- *What do you THINK and FEEL?*:
  - What are your biggest frustrations related to gender diversity in software development teams)? (Mandatory)
  - What obstacles there are between you and what you want / need to achieve (professionally and within your team)? (Mandatory)
  - What risks do you fear to face ? (Mandatory)
  - What do you really want or need to get? (Mandatory)
  - How do you measure your success ? (Mandatory)
  - Think of some strategies to achieve your goals. Can you share some of them?
  - What are other thoughts and feelings that motivate your behaviors (in general)? (Mandatory)
  - What is really important to you (and which you may not say publicly)? (Mandatory)
  - Imagine your emotions, what motivates you? (Mandatory)
  - What can keep you awake at night? (Mandatory)

Closing Questions

Any comments on the topic that you would like to report? Would you like to collaborate with data for the quantitative analysis of this research? We are looking for data from repositories (github, gitlab, etc.) that contain basic information about projects, repositories, commits and issues. All data will be anonymized. If you want / can collaborate, please leave a contact below. Thank you!

C.5 Preparing the data collection instrument

As mentioned, to support the execution of the Survey process, we used Qualtrics. We included the questions in the tool in Portuguese and English to reach not only the Brazilian community of software development but also people from all around the globe.

C.6 Validating the instrument

To validate the survey, we ran the first round with three respondents to give feedback about the questions. Once the report was positive, we went through the next step.

C.7 Selecting participants and Administering and scoring the instrument

Our target population consists of people who identify themselves as part of a software development team worldwide. We shared the web link for the survey in social networks to reach software development professionals (e.g., LinkedIn). So, we opted for the Non-Probabilistic Sampling Method called Convenience sampling. This kind of sampling involves obtaining responses from those people who are available and willing to take part, and we were aware that the main problem with this approach is that the people who are willing to participate may differ in important ways from those who are not willing [108].

C.8 Analyzing the data

Thematic analysis is the technique used to gain a deeper understanding of the data content. Braun and Clarke [19] describe thematic analysis as a method to identify, analyze, and report on issues related to data. The authors describe six phases of the thematic analysis process: familiarization with the data, generating initial codes, researching themes, reviewing themes, defining and naming themes, and producing the report. Thematic analysis usually involves open-source, where codes are used to organize the themes.

The thematic analysis process of this work will follow the steps mentioned by Runeson et al. [117] and summarized below:

1. Get the initial set of data. For this research: completed questionnaires;
2. Have the material studied in detail;
3. Formulate a set of codes of interest for the research, based on the research questions. Rely on another researcher and referenced literature to formulate the codes;
4. Read all texts and mark where codes fit into the content. Rephrase some codes if necessary: split the codes and create new ones if necessary. In this part, the process is iterative and, if new codes are formulated, the researcher will need to go back and re-code the material;

5. Use coded material to draw conclusions;
6. Compare text for different codes;
7. Compare different codes;
8. The process is iterative: there is the possibility to go back and adapt and change codes, sections, notes, and so on. It is also possible to go back and interview respondents again and identify new respondents if necessary.

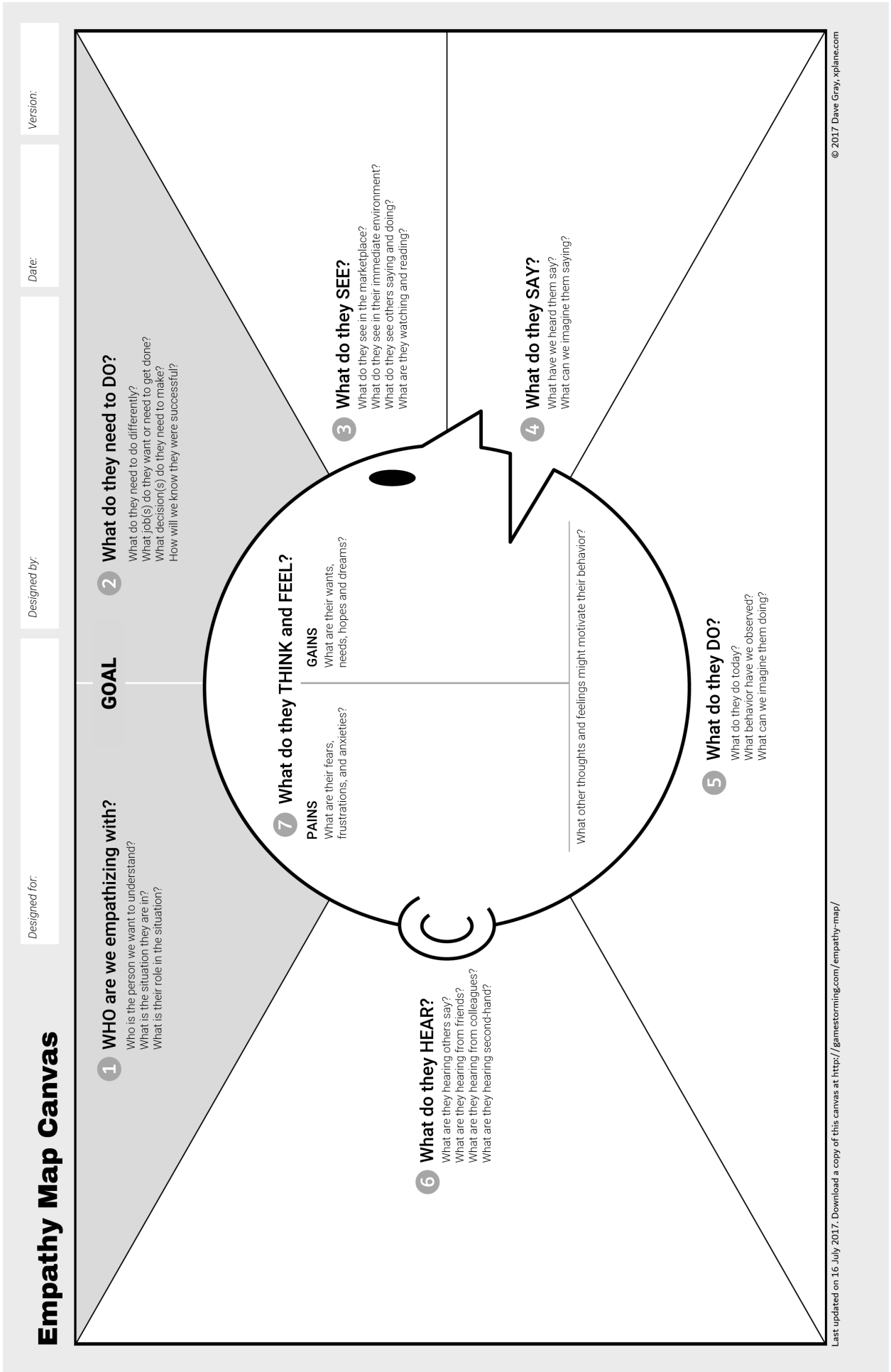


Figure C.1: Empathy Map Canvas, by Dave Grey [62]

APPENDIX D – SURVEY - THEMATIC ANALYSIS DATA

Table D.1: Missing Affirmative Actions and Initiatives

Theme	Sub-Themes	Codes	Frequency
Missing Affirmative Actions and Initiatives		Diversity awareness campaign in company	1
		Diversity is not discussed	4
		Diversity need prioritization	1
		Do not change is risky	1
		Few diversity initiatives	4
		Focus on diverse hiring	3
		I work in a company that is not opened to change and innovation	2
		It is important to think about gender diversity	2
		More initiatives do increase diversity	5
		Need for affirmative action about all kinds of diversity	2
		Need for affirmative action about gender diversity	2
		Need to increase opportunities	3
		Prioritize Diversity	0
		Remote difficulted interaction	1
		Slow process of diversity awareness	2
		Social affirmation policies	1
		Teach about gender neutrality	1
		Teams do not talk about diversity	3
		We should talk about diversity	2
		Working remotely the subject is not visible enough	1

Table D.2: Benefits of Diversity

Theme	Sub-Themes	Codes	Frequency
Benefits of Diversity		Mixed gender teams results better decisions	3
		More diversity more different perspectives	4
		More diversity, more criativity	1
		More diversity, more innovation	1
		New ideas are not accepted easily in a non diverse environment	1

Table D.3: Companies do not support D&I

Theme	Sub-Themes	Codes	Frequency
Companies do not support D&I		Diversity only on the facade	4
		Company does not let the learning curve to happen	1
		"It's working this way"	1
		Companies are imature to deal with diversity	3
		Companies do not support diversity	2
		Companies don't engage in diversity to not disturb the majority	1
		Companies hiding in fake strategies of employer branding	2
		Companies lack intention to talk about diversity	3
		Companies not positioning about diversity	2
		Diversity as marketing	1
		Diversity exploited to reduce the cost of salaries	1
		Diversity is in direct conflict with capitalism	1
		HR not concerned about diversity and inclusion	1
		I need authorization from company to talk about diversity	1
		It lack intentionality in the hiring proccess	2
		Large companies are doing good work on hiring women	1
		Lot of talking but no action (about diversity)	14
		No headcount to hire for diversity	1
		People are trying to take advantage (from diversity)	1
		Some companies strive to have a fairer environment	1
		The company did not encourage diversity	1
		Workplaces don't care about gender	1
		Diversity as a hidden agenda to reduce costs	1

Table D.4: Companies trying to be Diverse and Inclusive

Theme	Sub-Themes	Codes	Frequency
Companies trying to be Diverse and Inclusive		The company I work supports diversity	3
		Code Culture	2
		Companies say they do the possible	1
		Companies try to create a diverse environment	3
		Companies want to hire women	5
		Create new rules for hiring	1
		I speak my mind about hiring proccess	1
		In the company I work, all the voices are heard	1
		Maternity leave is not wrong where I work	1
		Openess to dialog	1
		Strive to be an inclusive company	1
		There is equity in the company I work	1
		Treat candidates equally	1
		Use blind screening to avoid bias	1

Table D.5: Diversity as a whole

Theme	Sub-Themes	Codes	Frequency
Diversity as a whole		Encourage gender neutral workplace	1
		Different backgrounds	1
		Gender neutral interview feedback	1
		Avoid unconscious bias	2
		Background diversity is important	1
		Care about gender diversity	2
		Cultural Diversity	1
		Differences in thinking	1
		Diverse people don't apply to jobs in tech	1
		Diversity being encouraged	2
		Diversity makes environment dynamic and democratic	2
		Diversity makes us strong	1
		Gender diversity matters	1
		I support diversity	5
		I support heterogeneous teams	1
		I talk assertively about gender diversity	1
		If we stop talking about diversity we come back to old patterns	1
		Only the ones that identify themselves as diverse seems to care	2
		Treat everyone as an equal	9

Table D.6: Sentiments and Emotions

Theme	Sub-Themes	Codes	Frequency
Sentiment and Emotions	**Anger**	Frustration	1
	Anticipation	Stress	2
		Mental strain by performing tasks in different contexts	1
	Trust	Trust	1
	Fear	Fear of being fired	1
		Fear of losing my job	1
		Fear of not beign in the right place	1
	Happyness Joy	Happyness	6
		Joy	1
		Satisfaction	1
	Sadness	Sadness	5
		Discouragment	1
		I feel sad for being privileged	1
		Tiredness	1

Table D.7: Family and Personal Life

Theme	Sub-Themes	Codes	Frequency
Family and Personal Life		Family in peace	1
		Family is valuable	3
		Family time	2
		Help kids with homework	1
		I take care of home activities	1
		I want to help my parents	1
		Life quality	4
		My personal friends are engaged in diversity subject	2
		People in my life are diverse	1
		Stability for family	1
		Supportive friends	1
		Take care of my family	3
		To see people I care safe and healthy	1
		Tranquility	3
		Wellness of my kids and family	5
		Work life balance	7

Table D.8: Financial Concerns

Theme	Sub-Themes	Codes	Frequency
Financial Concerns		sustainable retirement	1
		Equal opportunities	2
		Equal payment	3
		Equity	7
		Fear unemployment	1
		Financial Stability	4
		Good earnings so my father can stop working	1
		Good earnings	2
		I fear lose my income	1
		I get a decent salary	1
		Money	1
		My obstacles to enter SE were financial	1
		My workmates talk about diversity	1
		Pay my bills	1
		Pay women what they owned	3
		Promotion	1
		Salary increase	2
		Save money	1
		The salary is the higher I ever had	1
		To have my own home	3
		Women are paid less	2

Table D.9: Intersecctionality

Theme	Sub-Themes	Codes	Frequency
Intersecctionality		Age diversity is important	2
		Ageism	2
		Bad about gender diversity, worst about race	1
		Black people need to put more effort to succeed	1
		Company focus only on women diversity	1
		Computer science undergrad courses with no women	1
		Development initiatives focusing in white women	2
		Few transgenders	1
		Few women and black people	1
		I don't see black people in leadership roles	1
		I read about anti racism	1
		I've never had black professors or workmates	1
		I've never worked with a trangender person	1
		Most women developers are white	1
		My family and friends are majoritary white women	1
		My team doesn't have black people	1
		My team had only one black man	1
		No transgender people	1
		Only gender diversity	1
		People who talk about diversity are white heterosexual people	1
		We have people from all ethnicities and cultures	1
		We need to look all kinds of diversity	1

Table D.10: Leadership and Management Roles

Theme	Sub-Themes	Codes	Frequency
Leadership and Management Roles		High number of Women Professors in College	1
		Assertive positioning	1
		Brought the subject to the leadership	1
		Few women managers	1
		I am the only woman manager in my area	1
		I do not see high level management caring about diversity	1
		Leadership is majority men, it is an barrier for women	1
		Leadership is majority white	1
		Leadership is not concerned about diversity	1
		Little to no diversity in management does not help	1
		Managers unprepared for diversity	2
		No women in strategic roles	1
		People Management almost 50/50 but majoritary men	2
		Project management majoritary women	2
		To be Inspirational leadership in D&I	1
		We have women in leadership	1
		Women in leadership supporting hard time for learning	1
		Women in middle management role	1

Table D.11: Men

Theme	Sub-Themes	Codes	Frequency
Men	"Bro" Culture	"Bro" culture	1
		A network of male protectionism	1
		Companies always hire the most talkative man	1
		Embarrassed about a fully white men team	2
		I do not hear men talking about diversity	1
		Men are more heard than women	1
		Men have doubts about the ability of women in IT	1
		Men in strategical roles	1
		Neutrality in controversial subjects	1
		People are tired of "bro" culture	1
		Policies privileged rise of men	1
		Silence from men (about diversity)	1
		Tech is still mostly a boys club	1
		Toxic masculinity	1
		Typical boys probably hate diversity	1
		Undergrad students majoritary men	2
		We still have few men allies	1
		Women give feedback to men about sexism	2
	Desconstruct Prejudice	As a men, I put myself in a learn position	1
		I am a diversity ally	3
		I desconstruct the wrong ideas I have about diversity	1
		I review my actions every year	1
		I search different perspectives to learn from	1
		I try to desconstruct my prejudice	1
		Live with diverse people to bring awareness and desconstruction	1
		Look around to see if the environment is diverse	1
		People trying to get better as people	2
	Men Characteristics	Reinvent myself	1
		Young men are more conscious about diversity	1
		Innefective communication	1
		No communication	1
		No strategy and assertivity	1

Table D.12: Meritocracy and Elitism

Theme	Sub-Themes	Codes	Frequency
Meritocracy and Elitism		I hear people saying that everybody is equal	1
		I heard about meritocracy: if you want, you can	1
		Meritocracy	1
		Software Engineering is elitist	1

Table D.13: No Diversity

Theme	Sub-Themes	Codes	Frequency
No Diversity	**Far Away from Diversity**	I wish better gender balance	1
		Companies need diversity	1
		Far away fromn having diverse teams	1
		No diversity in IT	7
		No diversity in my team	7
		No diversity where I work	6
		People do not have knowledge about diversity	1
		People do not talk about diversity	3
		Women are minority	4
	No Equity	Far away to reach equity	3
		People complain there is not equity	1
		We won't reach equity if we are not able to hire	1
	All Men	All white hetero men	1
		Company does not care about trans and non-binary people	1
		Developers majoritary are men	7
		Fully white men team	1
		IT is composed by white, cis, men	1
		IT is predominant male	1
		More men thean women in SW teams	5
		Most men around	7
		Most white men	2
		Most white people around	6
		Not be a whote man make diffcult to grow in carreer	1
		White men around	3
	Old Patterns	Do not accept prejudice	1
		From hiring process most CVs are from men	1
		Junior developers documenting the code	1
		No patience for the same old behavior	1
		Software engineers are a pattern (white, cis, men)	2
		Tired of the same old patter	1

Table D.14: Overload

Theme	Sub-Themes	Codes	Frequency
Overload	**Causes**	I am a "do it all" in my team	1
		Interruptions	1
		Lack of Focus	1
		It is not easy to do what I do	1
		Overbooked	1
		Senior developers overloaded	1
		To many thing to solve	1
		Too much to do and too little time and/or energy	1
		Work overload	1
	Consequences	Delayed deliveries	1
		No quality on the deliveries	1
		Low Productivity	1

Table D.15: Professional Insecurities

Theme	Sub-Themes	Codes	Frequency
Professional Insecurities		As a woman I need to prove my capacity all the time	2
		As a woman I need to prove myself much more	2
		As a woman, I do not have space for insecurities	1
		Fight against my insecurities	1
		Fight againt Impostor Syndrom	1
		I am afraid of not being good enough	1
		I asked myself a lot	1
		I do not have the courage to talk, so I write	1
		I feel I am never good enough	1
		I feel I need to go beyond my peers	1
		I feel I was authorized (by men) to be where I am	1
		I feel insecure about my knowledge	2
		I feel that I am occupying what men think I deserved	1
		Lack of Opportunities	1
		Lack of knowledge to go ahead	1
		Lack of recognition	3
		My potential is not used	1
		No perspectives	1
		No positive feedbacks about my work	1
		No support from workmates	1
		No women participation	1
		Professional frustration	2
		Questions about decision to pursue software engineering path	1
		Recognition	10
		Research incentive	1
		Stagnation	1
		The daily struggle about being the only woman in a team	2
		Thoughts that I am not doing enough	1
		We put ourselves unpleasant situations to little progress	1
		What if nothing works before I have to retire?	1
		Women do not succeed to strategical positions	3
		Women have no space to grow	1
		Women in begginer levels	1
		It seems I must be available all the time	1
		Lack of rewards	1

Table D.16: Sexism and Prejudice

Theme	Sub-Themes	Codes	Frequency
Sexism and Prejudice		Desconstruct sexism	1
		Humiliation	1
		Prejudice	7
		Prejudiced jokes	4
		Stereotypes	2
		Structural Racism	1
		Structural sexism	9
		Team is sexist	1

Table D.17: Unawareness

Theme	Sub-Themes	Codes	Frequency
Unawareness		My workmates don't care about diversity	3
		People think the subject is not important	6
		"I look after the dog when my wife needs to go out "	1
		Concerns	1
		Conflict	1
		Difficult to find skilled women	2
		Diversity is to focus on political issues and not real issues	1
		Do not classify humans in genders, focus on the team	2
		Do not discriminate anyone	1
		Gender does not matter, Knowledge matters	4
		Girls have advantages when applying to jobs	1
		Hire women because of gender and not skills is demeaning	1
		I am accepted as anyone else in the team	1
		I am good for my skills and not my gender	1
		I do not have any frustrations about diversity	1
		I do not have any obstacles	1
		I do not have diversity problems where i work	3
		I do not have friends at work, only workmates, men and women	1
		I do not make distinctions regarding gender in my team	1
		I don't do anything to support diversity	1
		I don't talk about diversity	1
		I talk about people, not about gender	1
		IT people are focused on solving problems	1
		Ignorance	1
		It is a constant struggle	1
		It is difficult to raise awareness	1
		Males and females working together where I work	1
		Old ladies doing project and people management	1
		People say they are open and don't have any prejudice	1
		Problems are part of life, we need to work to solve them	1
		Promote women in tech in unethical	1
		Respect all people	7
		Teams say that diversity is whining	2
		To clean the mess of other developers because they didn't listen to me	1
		We are all equals	1
		We are all people	1
		We can't control how new hires act in terms of Diversity and Inclusion	1
		Women don't apply to it roles	5
		Women don't have technical background	1
		Women don't want tech	3
		Women dropped out at the last minute in the last hiring round.	1
		Psychology are predominant female	1
		I don't see problems with that	

Table D.18: Women

Theme	Sub-Themes	Codes	Frequency
Women	Women are disrespected	I was the only women in the team	1
		Few Women developers	3
		I do not see senior women developerts	1
		I see the number of women in IT increasing	1
		More women should join	1
		Quality assurance majoritary women	3
		We need more women in technology	2
		Women are as good as any men in SE	3
		A large part of issue is harassment in SE	1
		Abuse of authority	1
		As a woman, I've been through some bad situations in SE	1
		Bad-mouthing	1
		Feelings of no belonging	1
		Fight and not succeed	1
		Get over the disrespect to keep doing the job	1
		Get over the mistreatment to keep doing the job	1
		Harassment	1
		I alway need to prove myself	1
		I asked for permission to talk about technical subjects	1
		I impose myself at work	1
		Injustice	2
		It seems I always need to have a position about some subject	1
		Jokes about women being better on delivering than men	1
		Lack of acceptance	1
		Less senior men did not accept code review from women	1
		Manager ask a women direct report for a date	1
		Mansplaining	1
		Mansplainning	1
		Manterupting	1
		Men saying how I should work	1
		No career projection	1
		No recognition of women in software development teams	2
		People blame women for the lake of interest in IT	1
		People make jokes about diversity	1
		People make women believe SE is not for them	2
		People say I am exagerating	2
		Power abuse from senior workmates	1
		Reprimand for try to fight sexism	2
		Retaliation	1
		Shame until find my place	1
		Software development teams are toxic for women	1
		Some cultures are hard for women	1
		Some people say SE is not for women	1
		Some women gave up	1
		To be silenced by workmates	1
		Women are dramatic and talk to much	1
		Women are hired only when better than men	1
		Women are not respected as men are	1
		Women are not skilled to be hired	2
		Women are put in uncomfortable positions	1
		Women are seen as technically weaker	1
		Women are silenced	1
		Women deal daily with sexism	2
		Women face difficult situations in their teams	1
		Women need to prove themselves more than men	2
		Women occupy "authorized" spaces	1
		Women reporting harassment	1
		Women reporting prejudice from peers and managers	1
		Women's opinion are less valued than men's ones	1
		Workmates say I need to have hard times do learn	1

Table D.19: Women

Theme	Sub-Themes	Codes	Frequency
Women	Women Supporting Women	Be inspired by other women	1
		Followup through online meetups	2
		Shadowing	1
		Hope to open paths for women	1
		I heard women talking about diversity	2
		I mentor women	1
		I motivate women through my writing	1
		I praise women that I admire	1
		I praise women who motivate me to continue in IT	1
		I recommend women for positions	3
		I reinforce the importance of women in software development	1
		I support a women in tech group	3
		I support other women	1
		I try to contribute to empower women in tech	2
		I want to build the next generation of women in leadership	1
		I want to hire more women and black people	1
		I want to show women can reach better positions	1
		Initiatives to women development	2
		Make women comfortable in technology	1
		Search more women to IT	1
		See me and other women succeding in carrers	1
		Strong women are my inspiration	1
		Welcome other women	1
		Women searching for other women to hire	4
		Women supporting each other	7
		Women's Community are the best part	1
		Collaborative	1
	Women Characteristics	Conciliator	1
		Determination	1
		Empathetic	1
		Flexible	1
		Organized way of working	1
		Will to win	1

Pontifícia Universidade Católica do Rio Grande do Sul
Pró-Reitoria de Graduação
Av. Ipiranga, 6681 - Prédio 1 - 3º. andar
Porto Alegre - RS - Brasil
Fone: (51) 3320-3500 - Fax: (51) 3339-1564
E-mail: prograd@pucrs.br
Site: www.pucrs.br