UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
CURSO DE CIÊNCIA DA COMPUTAÇÃO

CAROLINA LANGE MELLO

Analysis and Visualization of Data about the Informatics Institute's Alumni

Work presented in partial fulfillment of the requirements for the degree of Bachelor in Computer Science

Advisor: Prof. Dr. Carla Maria Dal Sasso Freitas

Porto Alegre
May 2022

AGRADECIMENTOS

Primeiramente agradeço aos meus pais, Mara e Nei, por todo apoio e incentivo durante minha jornada acadêmica. Também quero agradecer aos meus amigos e colegas que me ajudaram nos momentos de dificuldade, sendo apenas fazendo companhia ou me ajudando com dúvidas técnicas.

Agradeço, também, ao Instituto de Informática da UFRGS, pela disponibilização dos dados necessários para este trabalho.

Por fim, agradeço minha orientadora Carla Freitas, uma profissional incrível que teve muita paciência e esteve sempre disponível para me ajudar.

ABSTRACT

Monitoring former students of undergraduate and graduate courses is one of the ways to measure the quality of teaching and research activities developed at the Instituto de Informática. This follow-up also allows for devising strategies to motivate students to enter and remain in the course. In particular, there is a worldwide movement for the expansion of women in the STEM careers, where Computing is inserted. Women still count for a low percentage in companies and educational institutions. The present work aims to collect, analyze and visualize data from the alumni of the Instituto de Informática. Data collection is done using an existing form on the Instituto de Informática website. Visualizations developed as part of a web application summarize the collected data, not revealing sensitive data in line with the LGPD, and provide an overview of graduates at various levels, even if incomplete, since the collection depends on the form completion by the alumni themselves.

Keywords: Visualization. Alumni. Institute of Informatics.

RESUMO

O acompanhamento de egressos dos cursos de graduação e pós-graduação é uma das formas de medir a qualidade do ensino e das atividades de pesquisa desenvolvidas no Instituto de Informática. Esse acompanhamento permite, também, traçar estratégias para motivar estudantes a ingressarem e se manterem no curso. Em particular, existe um movimento mundial para a ampliação de mulheres nas carreiras de Exatas, onde a Computação está inserida. Elas são ainda um percentual muito baixo nas empresas e nas instituições de ensino. O presente trabalho tem como objetivo coletar, analisar e visualizar dados de egressos do Instituto de Informática. A coleta de dados é feita de um formulário existente no site do Instituto de Informática. As visualizações desenvolvidas como parte de uma aplicação web sumarizam os dados coletados não revelando dados sensíveis em consonância com a LGPD e proporcionam uma visão geral dos egressos em vários níveis, mesmo que incompleta em função da coleta depender do preenchimento pelos próprios ex-alunos.

Palavras-chave: Visualização, Egressos, Instituto de Informática.

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

API   Application Programming Interface

CSV   Comma-separated values

HTML   Hypertext Markup Language

JSON   JavaScript Object Notation

LGPD   Lei Geral de Proteção a Dados Pessoais

SAGA   STEM and Gender Advancemen

SPA   Single Page Application

STEM   Science, Technology, Engineering, and Mathematics

UIS   UNESCO Institute for Statistics

W3C   Worl Wide Web Consortium

XML   Extensible Markup Language

CONTENTS

4

1 INTRODUCTION

The success of an undergraduate or graduate program can be measured by the success of its students in any contest or selection process they participate in and also by the successful career of its former students. Recently, CAPES, the federal funding agency that coordinates an assessment of all graduate programs in Brazil, has asked every program to list their most successful former students. Therefore, keeping track of these former students is important. One can also think that successful former students may become partners in projects or provide financial aid for the institution or students.

Another important reason for keeping track and maintaining close contact with former students is to trace strategies for motivating new students to enter and stay in the institution as long as possible, i.e., seeking MBA, MSc, or Ph.D. degrees after graduating. Regarding motivating new students, we have a special interest in inspiring girls to pursue a career in Computer Science. According to data compiled on women in technology from Women in Digital Scoreboard and The State of European Tech, only 17% of ICT specialists and 34% of STEM graduates are women. (WOMENTECH, 2022) Thus, keeping track of our former female students is especially important.

A few years ago, the Institute of Informatics started to collect data from its former students through a campaign using social media. A web page with a form is available[1], and such a campaign is repeated from time to time. However, there is no data curation, and one lacks ways of inspecting the collected data.

1.1 Objective and scope of this work

Motivated by such context and issues, this project aims to analyze and visualize data about former students of the Institute of Informatics in general and also focusing on women, improving the data about the history of the Institute regarding relevant information about its alumni.

We developed an application to summarize the data collected through the above mentioned form, aimed at providing visitors to the Institute's website with ways to analize such data. Web visitors might be interested in the alumni per course, in the number of women who have studied in the Institute, or general information about former students of all levels (graduated, MSc and Ph.D.).

---

[1]https://www.inf.ufrgs.br/site/pessoas/alumni-ex-alunos/

1.2 Structure of this work

This work is organized as follows. Next chapter (Chapter 2) briefly summarizes how some communities and institutions display data about their members or former members. Then, in Chapter 3, we describe our target dataset and technologies we used to build the application. In Chapters 4 and 5 we describe the application and the use cases to demonstrate its usefulness, respectively. Finally, in Chapter 6, we summarize our work and draw comments about future work.

2 RELATED WORK

In this chapter, we briefly describe works that served as inspiration for this project. The chapter presents two projects that inspired ours: the Stackoverflow Survey, that show information about developers, and the UFRGS' Data Panel, that presents general information about UFRGS (current) students. We also comment articles and reports that focus on women in STEM careers, since such a theme was the main motivation for our work.

2.1 Institutional data panels

In this section, we will analyze the data panels that served as inspiration for the project. We will highlight each of their particularities and provide a comparison between them.

2.1.1 Stackoverflow Survey

Every year since 2011, Stackoverflow[1] provides a survey with information about the developers that use the platform. They provide this data both in charts and in CSV format, so this information can be viewed and analyzed by others.

Until 2020, the data provided in the survey focused on users that posted technical issues, such as type of developer, technologies used, and experience. In 2020, the survey added information about a point that turned to be very important in the developer community: diversity. Data such as experience by gender and sexual orientation shows that certain groups, such as women, are still the minority even nowadays, as shown in Figure 2.1.

2.1.2 UFRGS' Data Panel

In 2017, in a project at the university's Data Processing Center, Bonatto (2018) developed the first version of the Data Panel ("Painel de Dados"), which aimed at creating a data visualization panel where interaction and data exploration would be easy and intuitive.

---

[1]https://insights.stackoverflow.com/survey/2020#overview

Figure 2.1 – Example of Stackoverflow infographics

Nowadays, the Data Panel has evolved. It is a web page that presents institutional information about UFRGS, such as undergraduate students and courses ("Graduação"), graduate students and programs ("Pós-Graduação) and professors and staff ("Gestão de Pessoas") among other data (Fig. 2.2). The graphics shown on the page are easy to understand and interactive, allowing the user to filter data according to their interest.

Having this amount of data on a single page makes it easy and practical to analyze students in general: we are able to measure the number of students enrolled, average age, how many earned master's and doctoral degrees. Moreover, one can filter (using the options accessible through the top tool bar menu). However, we cannot obtain information about former students, which is the main idea of this work.

2.1.3 Comparison between the data panels

Both data pannels have unique and interesting visualizations, despite the difference in their purposes. The Stackoverflow Survey has a similar focus to our project, it focuses on diversity and provides insight on the gap that exists between representation of men and women in the Tech Industry. Meanwhile, the UFRGS Data Panel focuses on design and data visualization, which gave us insight on ways to interact with data, how to display and organize it.

We decided to combine the strengths of both data pannels, to combine the insights

Figure 2.2 – UFRGS Data Pannel
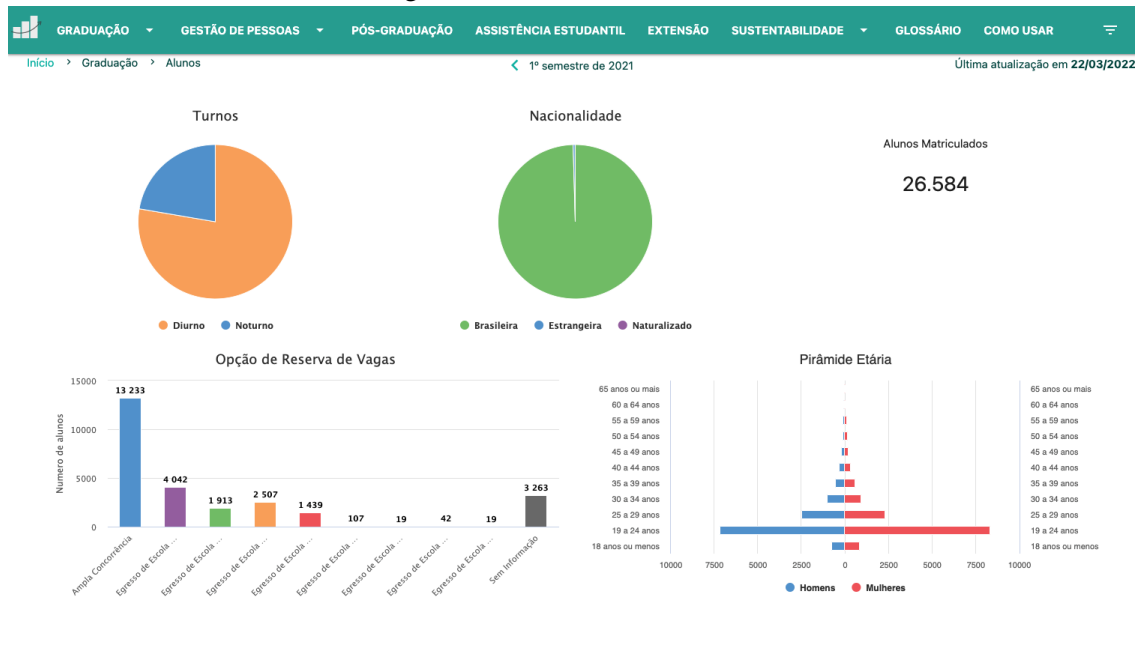
and the analysis from the Stackoverflow pannel with the visualization and interactivity of the UFRGS data pannel.

2.2 Women in STEM careers

In this section we comment on the articles focusing on women in STEM careers, which served as motivation for the project.

2.2.1 Analysis of women participation in the Computer Science course at UFCG

Despite having been quite important for the development of Computer Science in the early days of the area, the presence of women in companies and universities has declined over the years. Slowly, a global movement has started to invert this trend. One of the ways is encouraging girls to pursue degrees in Computer Science (CS) and related fields.

Despite these efforts, women still have low participation in undergraduate courses in CS. In a recent article, Miranda et al. (2021) seek to understand the motivation for the low participation of women in the area of technology, focusing on the Computer Science

course at the Federal University of Campina Grande (UFCG). By analyzing data provided by the institution, they also investigate whether gender would be a determining factor for success in the course.

The article goes deep into these issues and makes comparisons regarding the dropout of women and men from the course, as well as showing a comparison of the students' grades by gender. The results presented show that gender is not a determining factor for success in graduation. In terms of course dropout, success and performance, both women and men have a very close percentage, which serves as motivation and evidence that the area of technology is for both men and women.

2.2.2 UNESCO Institute for Statistics (UIS) and STEM

The UIS (UNESCO Institute for Statistics) is the official statistical agency of UNESCO and the source of the internationally-comparable data on education, science, culture and communication. It provides free access to data from all UNESCO countries and regional groupings, from 1970 to the most recent year available (UIS.STAT, 2022).

Less than 30% of the world's researchers are women, according to UIS data. The report also shows if these women work in public, private or academic sectors and their research fields. Based on this information, the UIS started developing a series of new indicators of what shapes women's decisions when pursuing careers of science, technology, engineering and mathematics (STEM) (UNESCO, 2022).

At the time the ideia was to use the data to improve the situation of women and reducing the gender gap in STEM fields through a new project: SAGA (STEM and Gender Advancement). [2]

Encouraged by the UNESCO initiative to provide such data and also by their analyses regarding diversity, we thought it would be interesting to provide similar data to INF's web page visitors. Next section describe the dataset and the tools we usedto build the INF's Alumni Data Panel.

---

[2]http://www.unesco.org/new/en/natural-sciences/priority-areas/gender-and-science/improving-measurement-of-gender-equality-in-stem/stem-and-gender-advancement-saga/

3 MATERIAL AND METHODS

This chapter presents an overview of the data collected and the technologies used for developing our application. First, in section 3.1 we describe our development flow and method; in section 3.2, we describe the dataset collected from the form available at the INF's web page; then, in section 3.3, we briefly introduce the technologies we used in this work.

3.1 Development Method

To create a Web Application with graphics regarding the alumni, the project development was divided into two initial parts: the acquisition of data and the beginning of the development of the front-end of the application, as shown in Figure 3.1.

Figure 3.1 – Application development method and flow
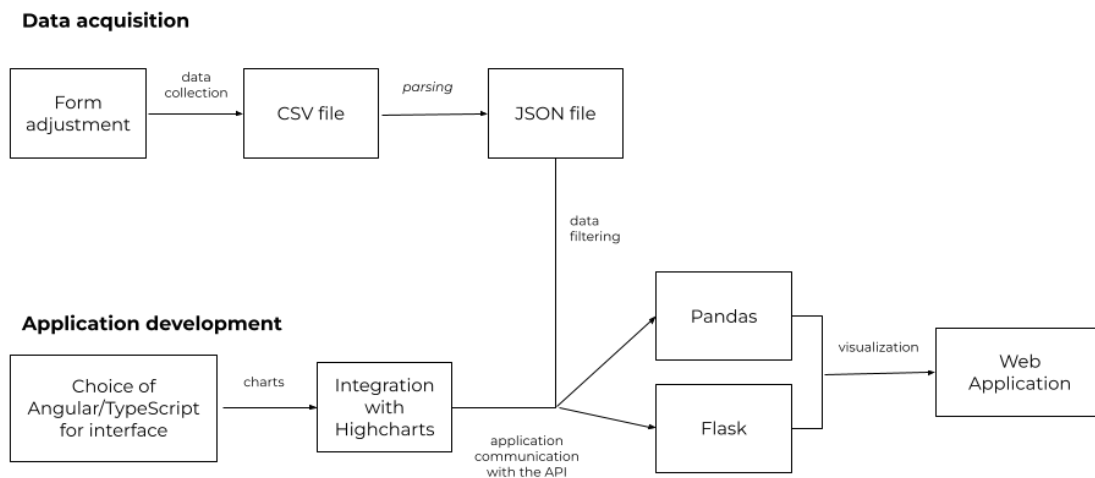
The alumni data was collected from a form, which needed some adjustments – section 3.2 provides more details. After adjusting this form and collecting the data in a CSV file, we converted this file into a JSON with information removed and filtered, ready to be used. Meanwhile, we chose Angular and Typescript as the technologies that would be used in the front-end and did the initial setup, integrating Angular with the Highcharts library to create the graphics.

With these two parts ready, the next step was to create an API with Flask to communicate with the front-end. Pandas was also used to filter the data and pass it in the

correct format to the interface, and so we were able to create the visualizations in the web application.

3.2 INF's Alumni dataset

In 2019, for the 30th anniversary of INF, a form was made available with the purpose of "reconnecting" with the alumni. As of today (May 2022), the form has collected data from around 800 former students.

With the idea of building visualizations with alumni's data, the first step was to analyze what was requested in the form and what information we would need for our project, which included investigating how many women have studied at INF. Originally the form asked for various informations:

- Full name
- Personal e-mail
- Country of residence
- City of residence
- Address
- Linkedin
- First connection with INF and year of completion
- Second connection with INF and year of completion
- Third connection with INF and year of completion
- Institution/Company name
- Place of operation
- City if Brazil/Country if abroad
- Do you still work in the area of Computing/informatics?
- Sector of activity
- Role in the institution/company
- Montlhy income

Most of them would be utilized for the project. However, some changes had to be made. The most important was to add a field that asked about gender since the focus of this work is to visualize the former students in general, and to understand the lack of gender diversity within the Institute. So, we asked for the addition of a question "Com qual

gênero você se identifica", with four possible answers: "Masculino", "Feminino", "Prefiro não informar", and "Outro", the last answer being a field to be filled in. Regarding the connections of the alumni, some changes were also made; instead of asking about the first, second, and third connections, the questions were modified to most recent connection, older conection, and if there was a third connection with the Institute.

The current form can be seen in the Appendix.

Of course, the modifications did not reflect on the answers that already existed in the form. So, the gender information of all the alumni on the dataset was inserted by hand, based on their names. We understand this may generate inaccurate data since the gender was filled considering only the graduate's name, limiting the answers to "Male" and "Female". Gender identity is broader than these two options, but unfortunately this was the best alternative we had due to the lack of this data in the form and its importance for this project.

Another matter to consider was the lack of consent from the former students that the data could be used with a purpose other than just reconnecting with them, so a policy and privacy field was added to the form. This field was added with LGPD in mind.

LGPD is law No. 13.709, approved in August 2018 and effective as of August 2020. It is a law that aims at creating a safe scenario for personal data of every citizen who is in Brazil (GOV.BR, sd). According to its 4th article, this law does not apply to academic research regarding personal data (GOV.BR, 2018), but we choose to add the policy and privacy field to the form, because an essential element of the LGPD is consent.

It is important to know the difference between personal and sensitive data, especially in the case of academic research, because we can use personal data without consent, but not sensitive data. Personal data is any information that makes it possible to identify an individual, directly or indirectly, for example: name, identity, address, photo, date and place of birth, IP adress, bank card, cookies, income, and others. [1]. Sensitive data are those that reveal racial or ethnic origin, political opinion, religious or philosophical beliefs, trade union membership, genetics, biometrics, and a person's health or sex life [2]. In this project, we do not use any sensitive data, only personal data.

---

[1]https://www.serpro.gov.br/lgpd/menu/protecao-de-dados/dados-pessoais-lgpd
[2]https://www.serpro.gov.br/lgpd/menu/protecao-de-dados/dados-sensiveis-lgpd

3.3 Technologies

Our application is web-based, and we used different technologies in its development. We analyzed some alternatives, and the ones we selected are presented herein.

3.3.1 Angular and Typescript

Angular is an open-source *framework* created by Google for developing efficient and sophisticated web applications (ANGULAR, 2010). It is built on TypeScript[3], an extension to JavaScript that adds static types and type-checking, which prevents common errors like typos, forgetting to handle null and undefined variables (MICROSOFT, 2021).

The framework provides resources for the development of Single Page Applications (SPA). A SPA is an application where all the functions exist in a single HTML page, so the code is obtained with just one page load. Moreover, the browser needs to render only the parts that matter to the user, without the need to load the page again in case of any update. According to the Angular documentation, this pattern can significantly improve the user experience[4].

We chose to use Angular and Typescript in the project for their ease of learning in a short period of time.

3.3.2 Highcharts

Highcharts is a JavaScript library based on SVG (Scalable Vector Graphics) used for generating charts. It allows the user to create responsive and interactive graphics easily and publish them on a web page (HIGHCHARTS, 2022).

SVG was developed by World Wide Web Consortium (W3C) and is a XML language to describe two-dimensional drawings and graphics in vector form with support for interactivity and animation. For being a vector-graphic, SVG images can be scaled in size without loss of quality (W3C, 2010).

---

[3]https://www.typescriptlang.org/docs/handbook/angular.html
[4]https://angular.io/guide/router-tutorial

3.3.3 Python, Pandas and Flask framework

Python is an object-oriented, high-level and open-source programming language created by Guido van Rossum that can be used for a variety of applications [5]. It is easy to learn and lets the developer to work more quickly, because of the support to integration with other systems (PYTHON, 2001). It also provides good support for filtering data and internet protocols, and that was the reason why we chose it as the programing language for the back-end of our application.

Flask is a web framework, a Python module that supports the development of web applications. It is defined as a microframework, which means that Flask aims to keep the core simple but extensible (PYTHONBASICS, 2010).

As it is said in the Flask documentation, the framework do not make decisions for the developer, such as what database to use, since it does not include a database abstraction layer or form validation. Instead, there is a variety of extensions that Flask supports to add the needed funcionalities to the application as if it was implemented in Flask itself (FLASK, 2010). The developer is also free to make design decisions that fits the projects' needs. For example, in the case of this project, instead of using an advanced pattern in any database tool, we choose to read a local JSON file, manipulate it with Pandas and send the manipulaged JSON object to our Angular application with the Flask web framework.

Pandas is a Python package mostly used for data analysis and manipulation tools (PANDAS, 2008). In this project we chose Pandas to manipulate and filter the data, since it is very flexible and easy to use.

---

[5]https://opensource.com/resources/python

4 VISUALIZING DATA ABOUT THE INFORMATICS INSTITUTE'S ALUMNI

In this chapter, we describe the web application "Painel de Dados de Egressos do INF"(INF's Alumni Data Panel) that allows the visualization of data about INF graduates. The panel allows data to be analyzed, both as an overview and by course, focusing on gender.

4.1 Architecture

The application was structured according to the three-tier architecture, which is a modular client-server architecture divided into: the presentation or user interface tier, the application tier, and the data tier. Separating an application into three tiers allows to update the technology of one tier without impacting the others, improving scalability and reliability. It also provides an ease of maintenance of the code base (INSIGHTSOFT-WARE, 2021). Figure 4.1 ilustrates the three-tier architecture of our project.

Figure 4.1 – Application Architecture
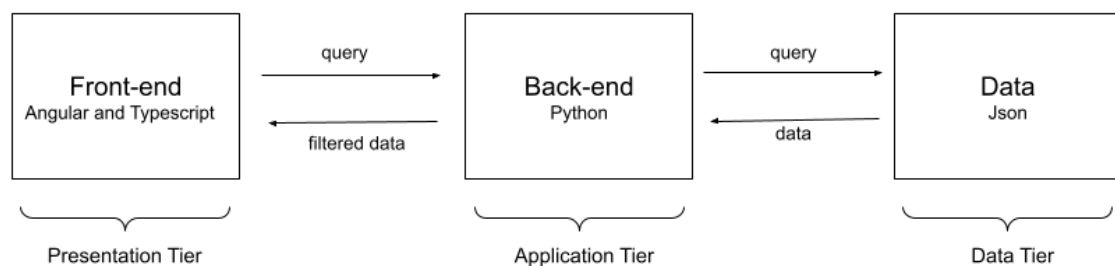
The presentation tier is the user interface, the *front-end*, where the user interacts with the application. Its main purpose is to display and collect user input. In this project, the presentation tier is the "Painel de Dados de Egressos do INF", which displays the data collected through the form mentioned before.

The application tier, or *back-end*, contains the business logic that supports the application's core functions. In this work, this layer processes the data and performs the filtering logic of this data.

The data tier, that may also be referred to as storage tier, consists of the database and a program for managing access to a database. In this project, we did not make exactly a data tier, since we did not use a database. Considering that the data is taken from a form

that is constantly being filled in, and we had to make modifications by hand, we decided to make this layer simpler: we have a JSON file as our "data layer", which was generated from the CSV file by a parser. The parser is further explained in section 4.1.1.

4.1.1 Parser

The data obtained from the form is in CSV format. In order to manipulate the data and for it to be read by the Angular application correctly, we implemented a parser in Python. This parser reads the CSV and filters the data according to what is expected by the application, returning a JSON file.

Some of the information in the original CSV, for example, personal e-mail and address, were deleted as they were not needed for the application. Others were renamed to simplify because the names were too long. The JSON file is shown in Figure 4.2 .

Figure 4.2 – Parsed Data

```json
{
    "name":"Carolina Lange Mello",
    "gender":"Feminino",
    "country":"Brasil",
    "city":"Porto Alegre",
    "connection1":"Graduação – Ciência da Computação",
    "year1":"Em andamento",
    "connection2":null,
    "year2":null,
    "connection3":null,
    "year3":null,
    "company":"Globo",
    "placeOfCompany":"Brasil",
    "cityOrCountry":"Porto Alegre",
    "active":"Sim",
    "sector":"Indústria",
    "function":"Desenvolvedor"
}
```

4.2 Visualizations

The initial objective of this work was to provide visualizations that would allow an overview of the alumni showing their distribution in all categories as well as by gender.

However, as we were designing the visualizations, the idea of also separating them by course came up. This separation is interesting considering that each course may have its particularities and specific information that users might like to see. So, we separated the data panel into two parts: Overview and Course View. In both the Overview and Course View we have charts showing data separated by gender.

Each of our application's views needs different alumni dataframes, with distinct information for each view. Thus, the front-end accesses the API on the back-end, which is responsible for fetching the data in the JSON file, and organizing and filtering them according to the needs of each view, through an endpoint. This endpoint returns, for each chart, the data ready to be used by the Highcharts library, responsible for generating the charts. The overview and course view endpoints are different.

Next sections give details about the two views.

4.2.1 Overview

In the Overview we have the visualization of the data of all the alumni who answered the form. In Figure 4.3, we show the home page of the INF Alumni Data Panel (Painel de Dados de Egressos do INF).

Figure 4.3 – Part of the Overview view
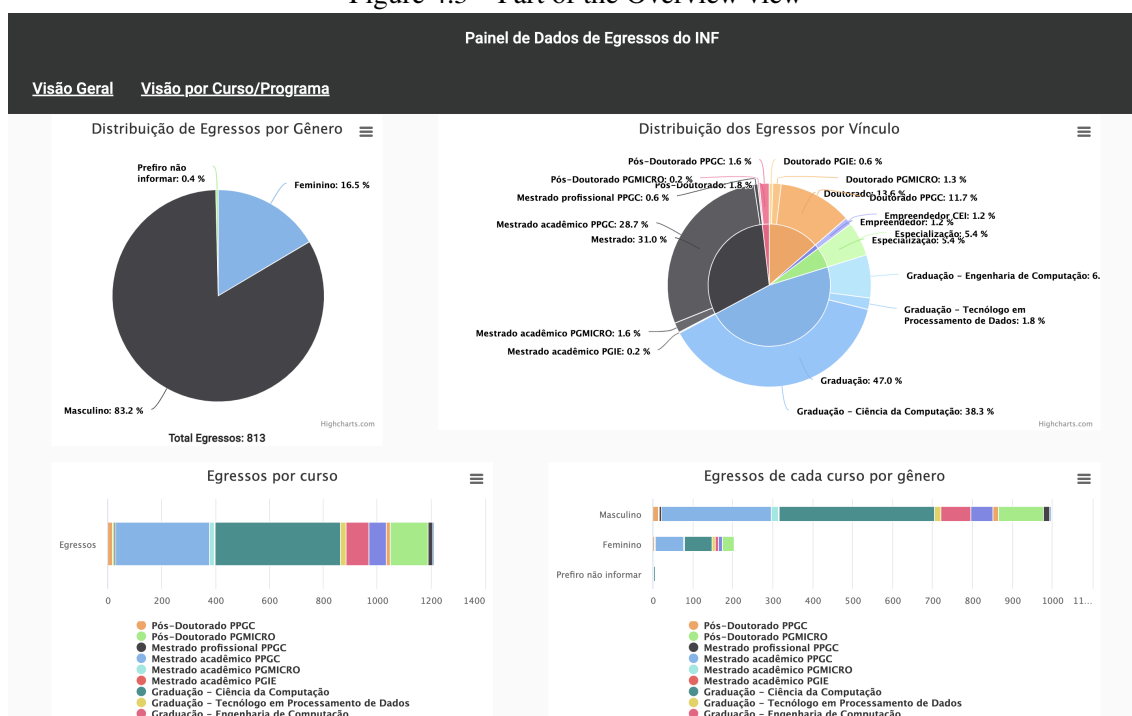
The process of defining each graphic was done thinking how to show information that we deem necessary among the data obtained through the form, focusing mainly in the distribution of gender and also all the courses. In the case of distributing the charts on the page, we prioritized placing them close to those that had similar themes, such as the two sector graphics that are side by side. We also considered the sizes of the charts, some needed more space and we prefer not to leave them crammed with others of the same theme.

There are several charts being shown, and they will be explained herein. The *Pie Chart*, known as being published for the first time by William Playfair (YAU, 2011), is the first chart that can be seen in the page. It is used to illustrate the quantity of graduates of each gender. As was explained in the section 3.2, most of the gender information was included by hand as the gender field was added after several graduates had already filled out the form, so the gender in the Pie Chart are: "Feminino", "Masculino" and "Prefiro não informar", as shown in Figure 4.4.
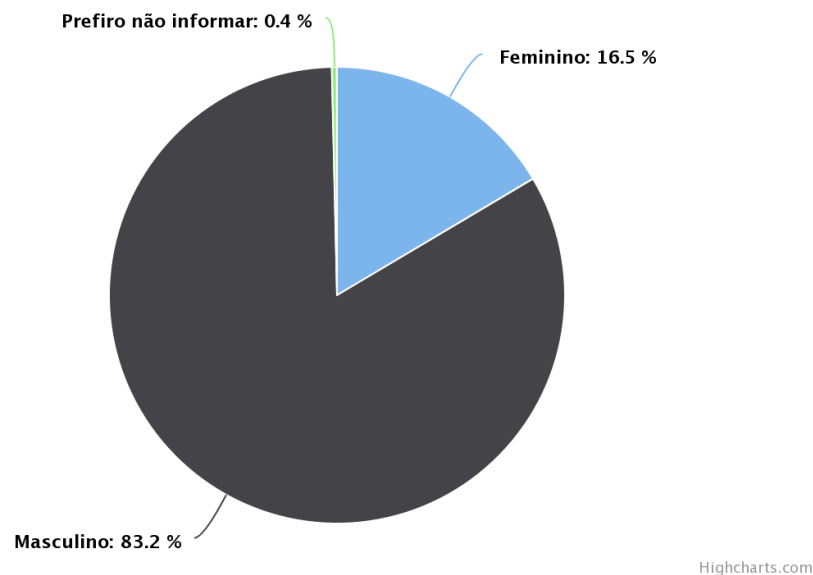
Figure 4.4 – Pie Chart of gender distribution

Figure 4.5 shows the genre data being queried from the JSON file. In line 1, the data from the JSON file is read to a dataframe, which is then grouped by gender in line 3. The result is a dataframe with two columns. One of the columns is called *gender*, and the rows of this column are all the genders of the graduates (for example, Male and Female). The other column is called *quantity*, and each of the rows has the number of graduates of each gender. This query is performed for both the Overview and Course view.

Figure 4.5 – Code for separating data by gender

```python
1  df = pd.read_json(file)
2  df = df["gender"]
3  df = df.groupby(by=["gender"]).count()
```

To the right of the Pie Chart there is a *Donut Chart* that shows the number of graduates per connection. The chart has two parts: the middle that represents the type of connections the graduates had (for example, Graduate, MSc, Ph.D., etc), and the part around the chart that shows the courses of each connection and their percentage.

Figure 4.6 – Donut Chart showing the distribution of alumni per category

Distribuição dos Egressos por Vínculo
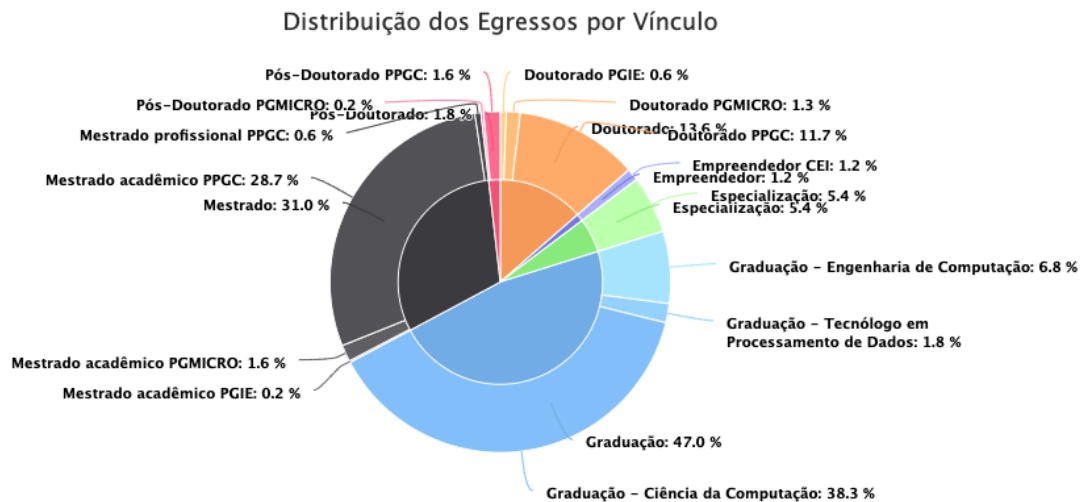
Selecting the information needed for this chart was more complicated. As demonstrated in Figure 4.2, we have three possible connections with each alumni (*connection1*, *connection2*, *connection3*) that refer to possible courses taken by them. In this chart, we also need the information of the alumni's connection, which we call *typeConnection* in the code.

In Figure 4.7, in lines 3 to 5, the data needed from connections and typeConnections is selected, forming three dataframes of connections. In line 8, the three dataframes are concatenated, with all information distributed in two columns: connection (that contains all information of connection1, connection2 and connection3) and typeConnection (that contains all information of typeConnection1, typeConnection2 and typeConnection3). Finally, in line 9, the data is grouped by the two variables, and the result is a dataframe with three columns: *connection*, *typeConnection*, and *quantity*.

The Stacked Bar Chart in Figure 4.8 has very similar information to that shown in

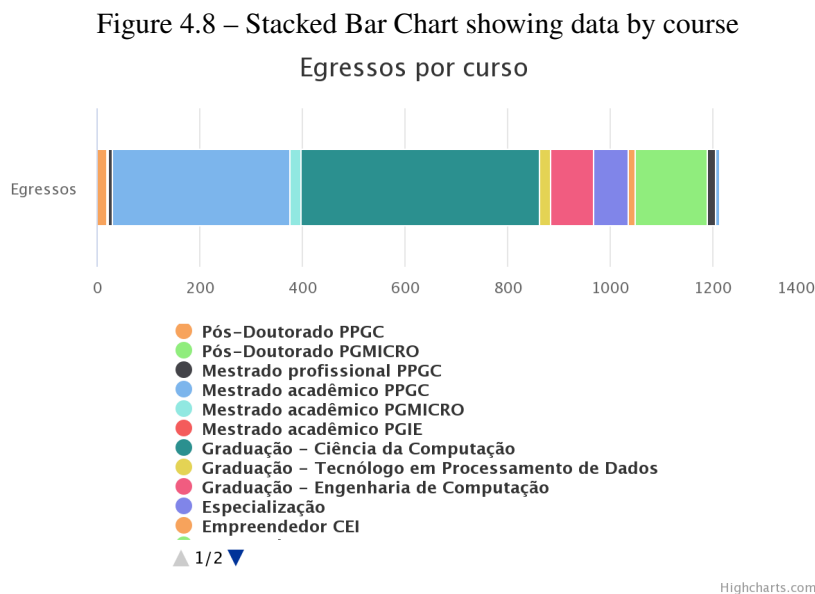Figure 4.7 – Code for separating data by connection type

```python
df = pd.read_json(file)

connection1 = df[["connection1", "typeConnection1"]]
connection2 = df[["connection2", "typeConnection2"]]
connection3 = df[["connection3", "typeConnection3"]]

connectionsArray = [connection1, connection2, connection3]
connections = pd.concat(connectionsArray).reset_index()
connections = connections.groupby(by=["connection", "typeConnection"])\
                .sum()\
                .reset_index()
```

Figure 4.6. It has a different format and presents only one category, the alumni, and the courses as the sub-categories. The bar shows the amount of courses taken by the graduates more clearly and it is important to emphasize that each graduate may have taken more than one course.

Figure 4.8 – Stacked Bar Chart showing data by course

The selection of the information needed to build the chart in Figure 4.9 is also very similar to the one shown in Figure 4.7, but simpler. We only have to search for the connections of the alumni, which refer to the courses taken by them, and group these connections. The result is a dataframe with two columns: *connection*, with all the courses, and *quantity*, which is the number of graduates by course.

Another Stacked Bar Chart (Fig. 4.10) presents the amount of graduates of each course by gender. Manipulating the data was also very similar to what is shown in Fig-

Figure 4.9 – Code for separating data by connection

```
1  df = pd.read_json(file)
2
3  connection1 = df[["connection1"]]
4  connection2 = df[["connection2"]]
5  connection3 = df[["connection3"]]
6
7  connectionsArray = [connection1, connection2, connection3]
8  connections = pd.concat(connectionsArray).reset_index()
9  connections = connections.groupby(["connection"]).sum().reset_index()
```

ure 4.7, but adding the field gender when doing the *groupby* on line 9.

Figure 4.10 – Stacked Bar Chart by gender
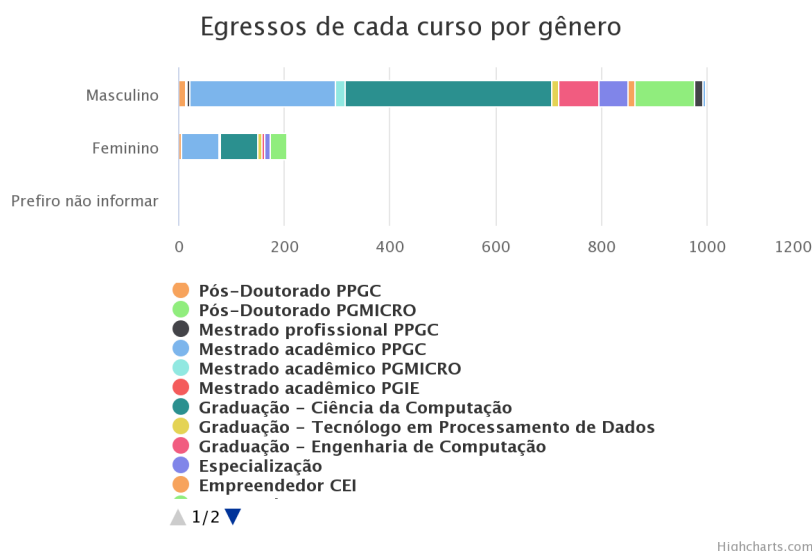
On the Overview page, we also have three Basic Column charts, which allow comparisons. The first Basic Column Chart (Fig. 4.11) not only shows the number of graduates per country they live in, but it also shows the number of graduates of each gender per country.

The second Basic Column Chart (Fig. 4.12) shows the alumni per year and separated by gender. The third Basic Column Chart (Fig. 4.13) also shows the alumni per year, but separated by connection type. Selecting the necessary information to make these three charts is similar to the others already shown.

The last two charts of the Overview view are both focusing on the economic sectors in which the graduates are working on or worked on. Figure 4.14 shows a Semi-donut Chart that presents all sectors, where the size of each sector corresponds to the number of alumni working in that sector.

Figure 4.11 – Basic Column Chart by country and by gender
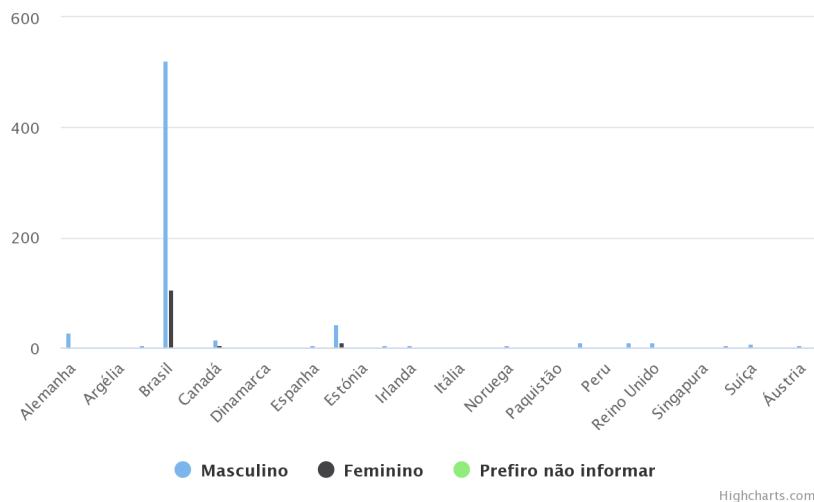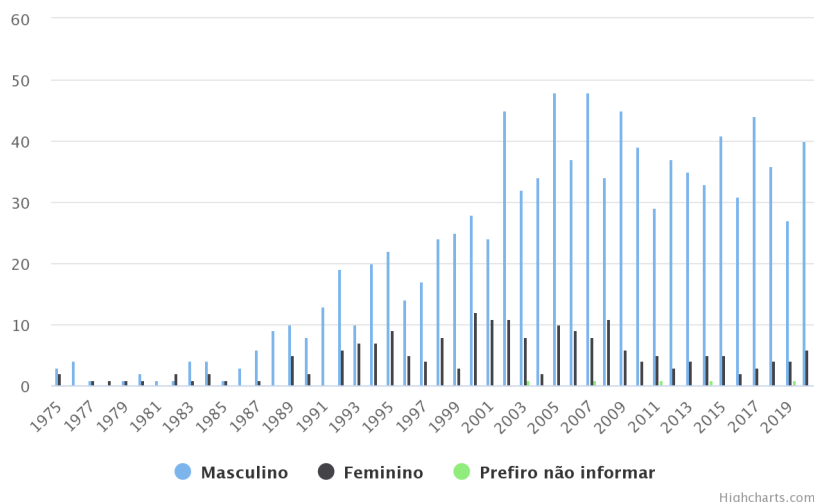

Egressos por país de destino por gênero

Figure 4.12 – Basic Column Chart by year and by gender

Egressos por ano por gênero

As we can see in the Semi-Donut Chart, there are six options: "Academia", "Empreendedor", "Governo", "Indústria", "Outro", and "Serviços". In the form, however, the sector is a field with options that the graduates can choose from (more than one), and a field (for "Outro"), where they can write whatever they want. So there are many different sector answers in this form field, and it was very difficult to visualize them in the chart. For this reason we decided to divide it into six categories, and the "Outro" encompasses all the responses other than the five options. This is demonstrated in line 12 of Figure 4.15.

As already said, the Sector is a field in the form with options and the graduates could choose one or more, that is, an alumni may have worked in more than one sector, so we had to split this data as shown in line 10 of Figure 4.15.
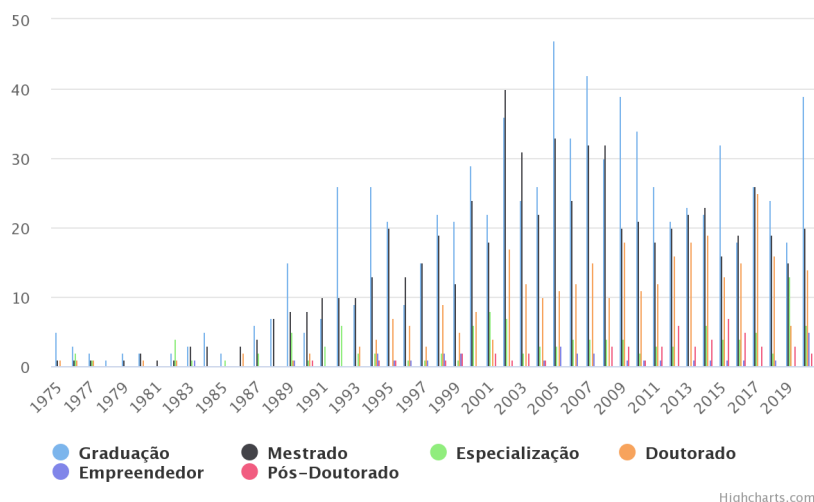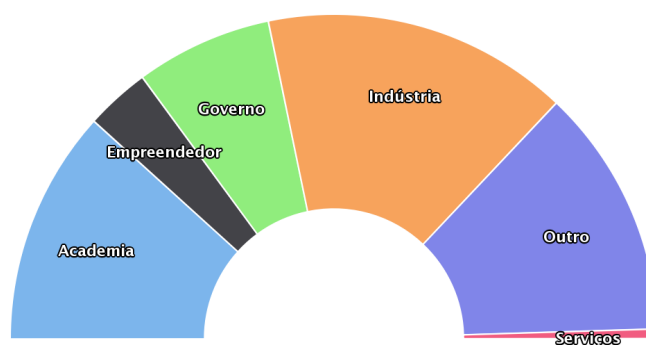
Figure 4.13 – Basic Column Chart by connection

Figure 4.14 – Semi-donut Chart by sectors

The last chart (Fig. 4.16) is a Stacked Bar Chart that let us visualize the sectors by gender. In this chart, our categories are genders and the sub-categories are the sectors.

4.2.2 Course View

The Course View is arranged as eight tabs, one for each of the courses offered by INF (Figure 4.17). The purpose of this arrangement is to allow the analisys of the data filtered by course, which migh be interesting for professors and coordinators of these courses, and also because a course may be better represented by some graphs that do not
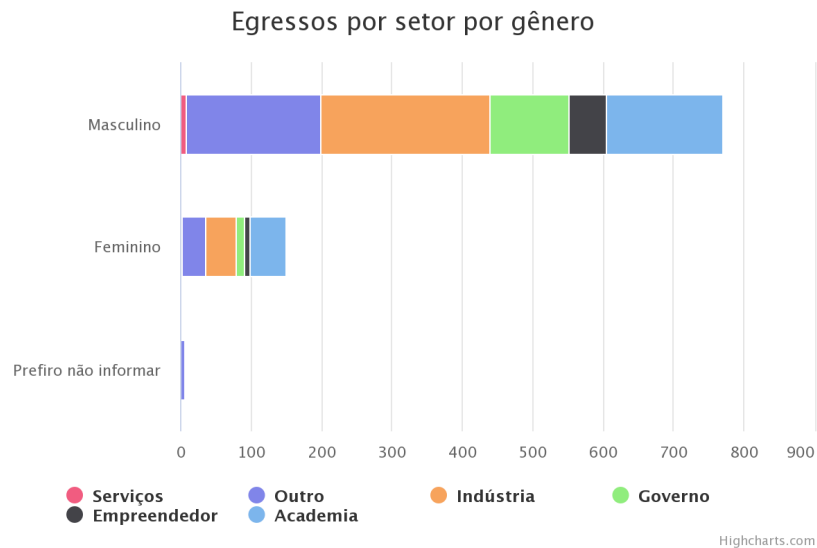
Figure 4.15 – Code for separating data by sector

```python
df = pd.read_json(file)
sectorNames = ['Academia',
               'Indústria',
               'Empreendedor',
               'Governo',
               'ONG',
               'Serviços']

df = df[["sector"]]
df['sector'] = df['sector'].str.split(";")

df.loc[~df["sector"].isin(sectorNames), "sector"] = "Outro"
df = df.groupby(by=["sector"]).count()
```

Figure 4.16 – Stacked Bar Chart by sector by gender

make sense to the others.

This page has the same charts already described in Section 4.2.1, but the API route accepts GET and POST methods, so an argument is passed with the tab's course in the URL when a request is made to the API.

Comparing Figures 4.17 and 4.18, some differences can be noticed despite both being course tabs. This is because Computer Science, like Computer Engineering, is an Undergraduate course while PPGC (and also PPGMICRO and PGIE) are graduate programs. So, some charts with the courses division of the PPGC courses were added to the corresponding tab.

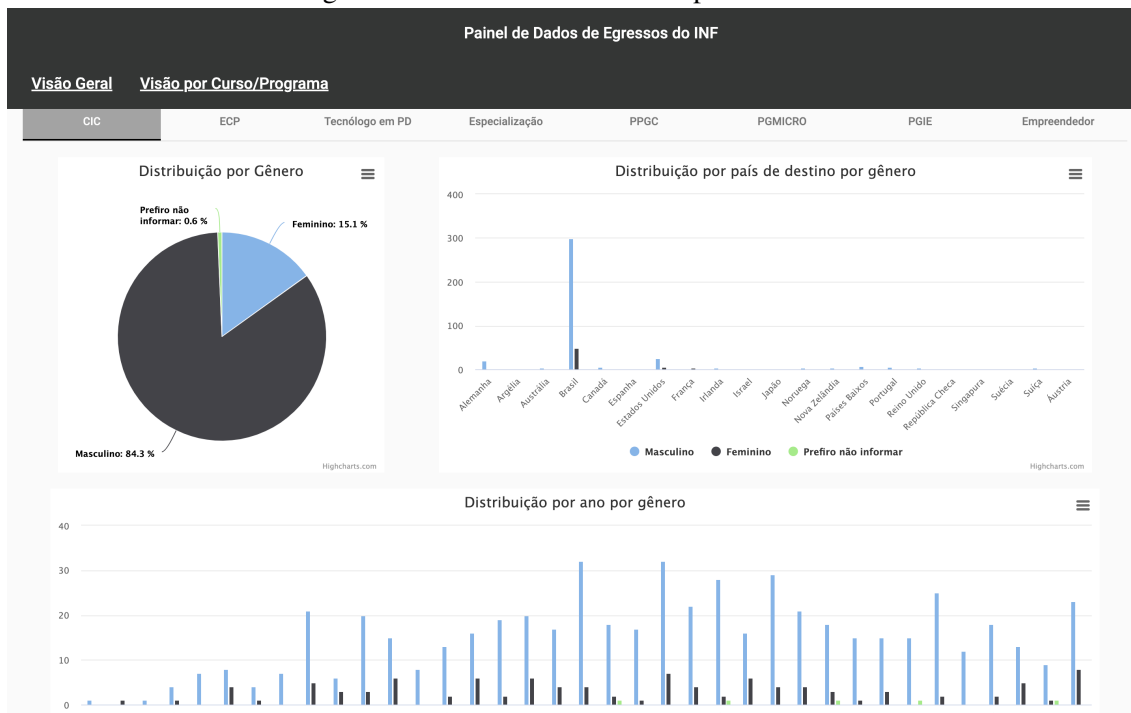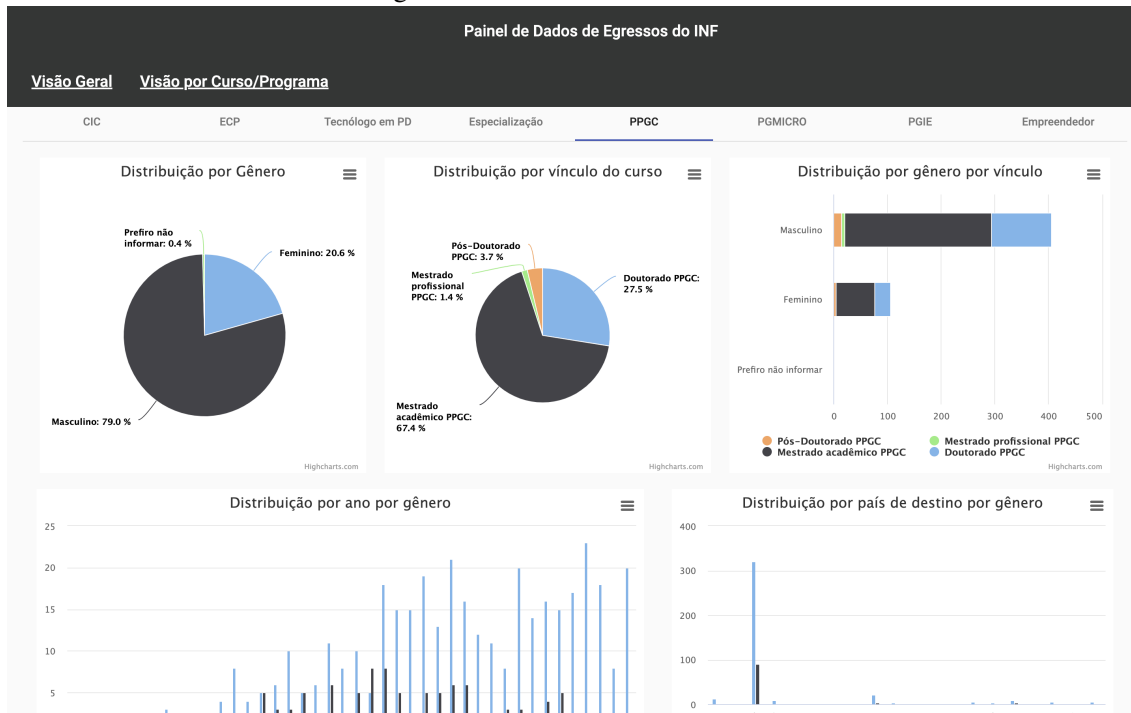Figure 4.17 – Course View - Computer Science

Figure 4.18 – Course View - PPGC

5 USE CASES AND ANALYSIS

In this chapter we will show use cases through a series of visualizations shown in Figures 5.1, 5.2, 5.3, 5.4 and 5.5, explaining how to interpret the information in each chart.

Figure 5.1 – Pie Chart presenting the gender distribution with two possible interactions: a) selecting a sector triggers the display of a tooltip with the name of that sector, percentage and the total; b) clicking on the sector highlights it.
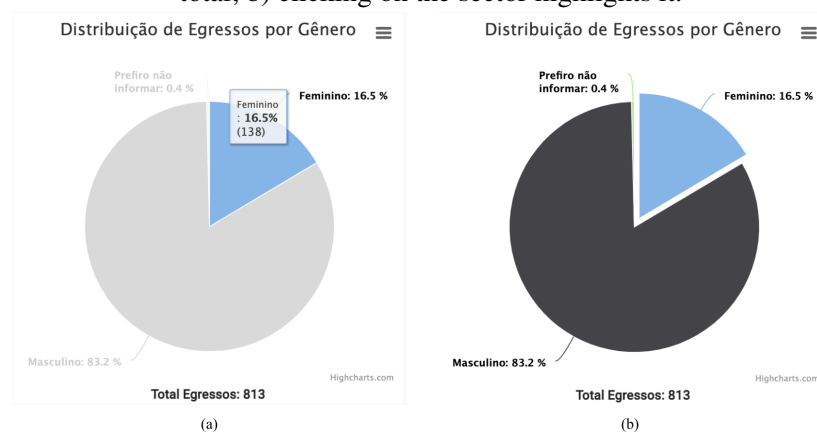

Figure 5.2 – Stacked Bar Chart presenting the gender distribution of each course with two possible interactions: a) selecting a sector in the chart displays a tooltip with the name of that sector and the quantity of graduates of that gender; b) clicking on some courses, for example, the undergraduate ones, removes them temporarily from the chart (until we click on them again) to view the data without those courses.
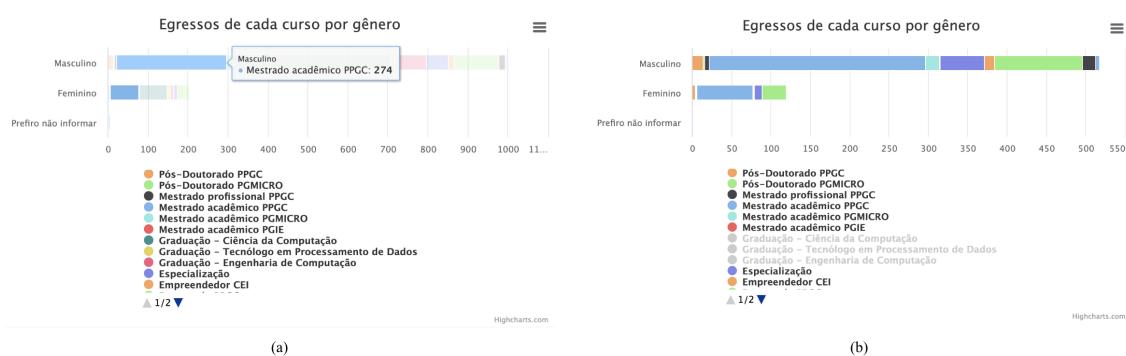
Analyzing these images, we can already take some interesting conclusions about the graduates. For example, in Figure 5.4 we can see that the undergraduate courses contain the most significant percentage of graduates. We can also compare the courses which have a more considerable quantity than the others, in the graphic, it is Computer

Figure 5.3 – Semi-Donut Chart presenting the sectors distribution. Upon interaction, a tooltip shows the name of the sector, the percentage and total of alumni that works or have worked in the selected sector .
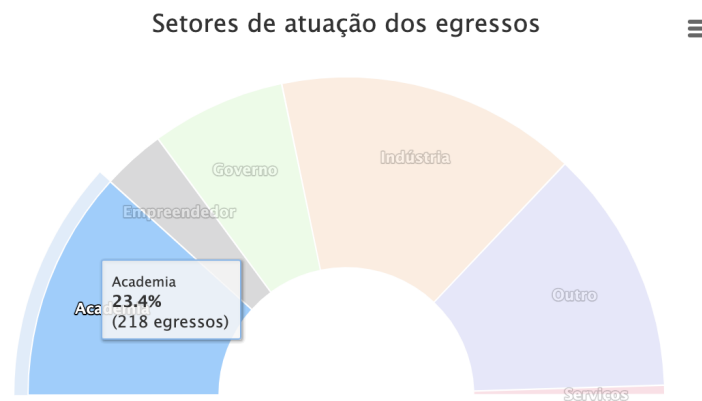
Science.

Another interesting, quite important comparison is the number of graduates of each gender. Looking at Figures 5.1 and 4.10 one can observe the difference between the number of men and women, since we have a much larger number of male graduates. This difference is not only present in general, but also considering each course and none of the courses have more women than men, not even the same quantity of both, which was already expected. It can also be analyzed in the charts of Figure 5.5 that the number of women had some peaks in periods between 2001 and 2009, but has been low in the recent years.

Switching between tabs in the Course View also allows us to compare their charts, especially those that make their differences more apparent. The Gender Pie Chart, being one of the first charts on the page, often draws attention to the number of women in a course that has very few women compared to the number of men (Computer Engineering with 8.4%, for example) and a course that has more (PPGC with 20.6%).

Of course, the graphics are not entirely accurate because we only have data from those who answered the survey. There are undoubtedly other graduates who did not respond. For example, PPGC has more than 2,000 alumni (1,700 MSc and 390 Ph.D.), and our dataset has around 800 records, including CS and CE former students. In addition, when placing the gender data by hand based only on names, some misunderstandings may have occurred. In any case, the difference between the number of men and women is very large and would hardly change.
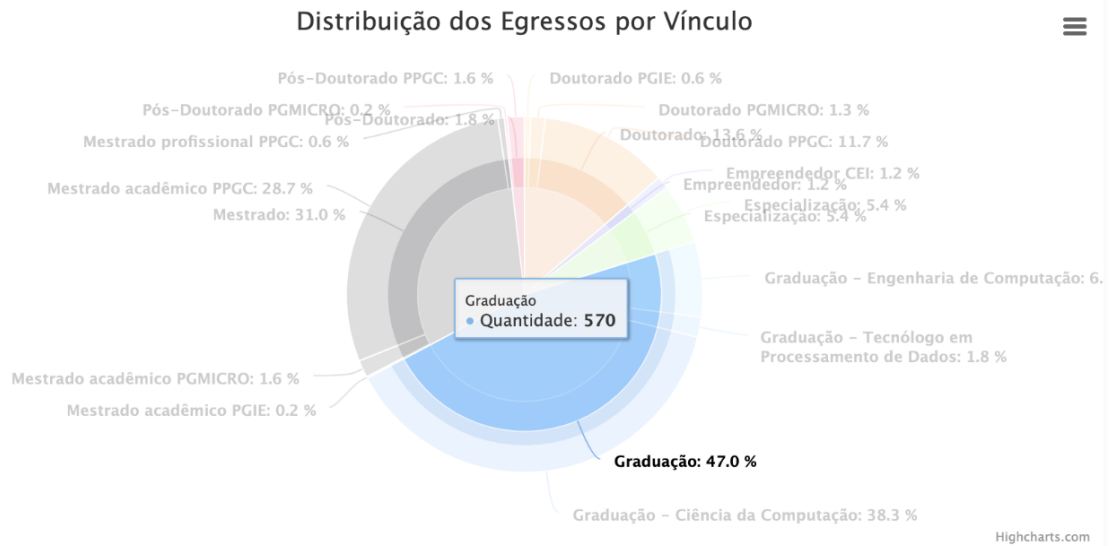
Figure 5.4 – Donut Chart presenting the distribution of alumni per type of connection, with two possible interactions: a) selecting a sector in the middle of the chart displays a tooltip with the name of that connection type and the total of graduates; b) selecting a sector on the outside part of the chart displays a tooltip with the name of the course and the total of graduates.

(a)
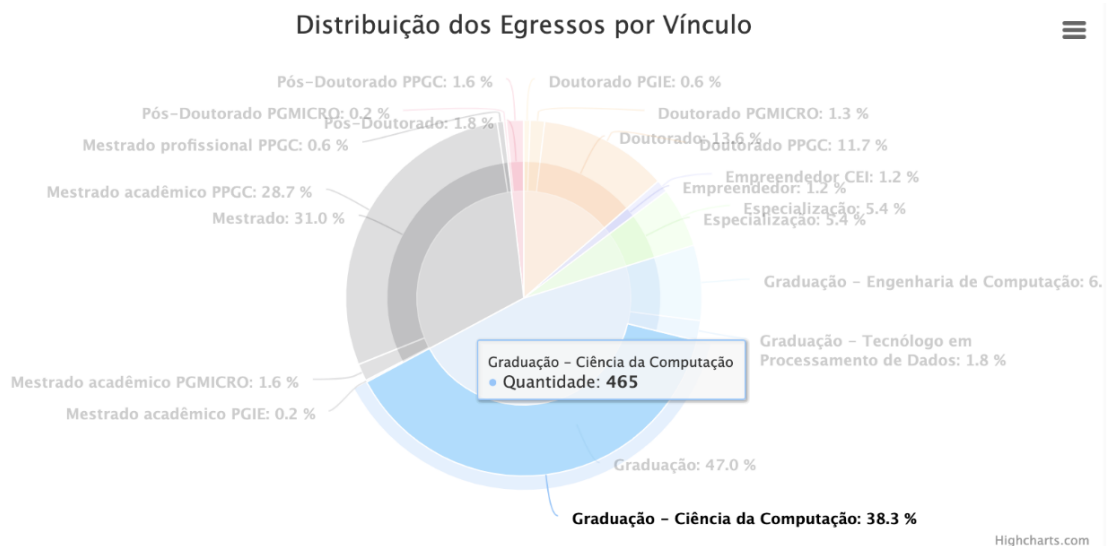
(b)

Figure 5.5 – Basic Column Chart presenting the gender distribution per year with three possible interactions: a) selecting a column of the chart displays a tooltip with the year of that column and the genders with their total of graduates on that year; b) selecting one of the categories make the others more transparent, and the selected category is highlighted; c) clicking on one of the categories, for example Masculine, removes them from the chart temporarily (until we click on them again).
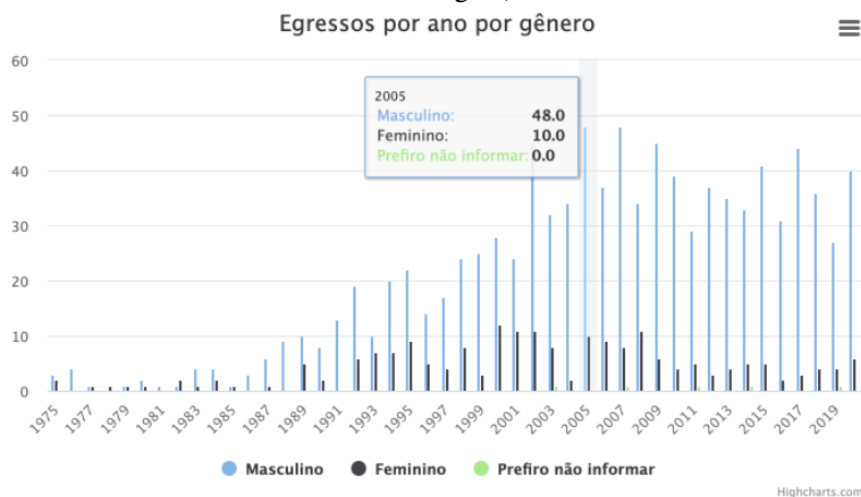
(a)

(b)

(c)

6 CONCLUSIONS

This project aimed to provide an overview and views by course of the Institute's alumni. In this first version, visualizations were developed to bring quantitative data about the graduates, focusing mainly on the issue of gender distribution. Then, the project evolved to provide more information based on what is available from a survey form developed by the Institute.

The current version of the application not only allows us to visualize the collected data, but we can also make several comparisons between the data shown and between the generated graphs.

6.1 Limitations

The biggest limitation of this project was the fact that the form used was not made with this project in mind, so necessary information was not included in the first version, for example, the gender of each alumni. Not only we had to put the gender information by hand, in our prototype, but we had some issues with the field that collects data about the economic sectors that the alumni are or have been working in. The very fact that the information came from a form that had to be filled in voluntarily is also a limitation, as we certainly did not get the data of all the students who had already been to the Informatics Institute.

6.2 Future Work

There are several possiblities for future work. The first one is to create a database, for example with MongoDB, instead of using a JSON file. It would be interesting and would make it a lot easier to do some automation regarding downloading the form data and updating the database from time to time, considering that the form already meets all the necessary information for analytical purposes.

Considering that a tab was created for each course in the application, the next step would be to talk to professors and coordinators to understand the specifics of each course, and then create views and filters that make sense for each one. This would allow each course unique visualizations.

In the future, there could also be more focus on diversity, not only on the issue of gender distribution but also on racial aspects, sexual orientation, and people with disabilities, among others.

Currently the web application runs only locally as a prototype. So, future work also is to make the application available in a URL at the Institute site. Finally, the application is not responsive; future work is also needed to deploy it for mobile devices.

REFERENCES

ANGULAR. **Introduction to the Angular Docs**. 2010. Disponível em <https://angular.io/docs>.

BONATTO, M. A. **Visualização de dados institucionais da UFRGS**. 2018. Trabalho de Conclusão (Bacharelado em Ciência da Computação), Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018. Disponível em: <http://hdl.handle.net/10183/175106>.

FLASK. **Flask**. 2010. Disponível em <https://flask.palletsprojects.com/en/2.0.x/foreword/>.

GOV.BR. **LEI Nº 13.709, DE 14 DE AGOSTO DE 2018**. 2018. Disponível em <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm>.

GOV.BR. **O que muda com a LGPD**. sd. Disponível em <https://www.serpro.gov.br/lgpd/menu/a-lgpd/o-que-muda-com-a-lgpd>.

HIGHCHARTS. **Highcharts**. 2022. Disponível em <https://www.highcharts.com/products/highcharts/>.

INSIGHTSOFTWARE. **5 Benefits of a 3 tier architecture**. 2021. Disponível em <https://insightsoftware.com/blog/5-benefits-of-a-3-tier-architecture/>.

MICROSOFT. **Announcing TypeScript 4.2**. 2021. Disponível em <https://devblogs.microsoft.com/typescript/announcing-typescript-4-2/>.

MIRANDA, B. A. de et al. **Análise da participação feminina no curso de Ciência da Computação da UFCF**. 2021. Artigo, Departamento de Sistemas e Computaçã -Universidade Federal de Campina Grande (UFCG), Campina Grande, 2021. Disponível em: <https://sol.sbc.org.br/index.php/wit/article/view/15837/15678>.

PANDAS. **Pandas Documentation**. 2008. Disponível em <https://pandas.pydata.org/docs/>.

PYTHON. **Python**. 2001. Disponível em <https://www.python.org/>.

PYTHONBASICS. **What is Flask Python**. 2010. Disponível em <https://pythonbasics.org/what-is-flask-python/>.

UIS.STAT. **Data UIS**. 2022. Disponível em <http://data.uis.unesco.org/>. Access in: 2022.

UNESCO. **Women in Science**. 2022. Disponível em <http://uis.unesco.org/en/topic/women-science>.

W3C. **Scalable Vector Graphics (SVG)**. 2010. Disponível em <https://www.w3.org/Graphics/SVG/>.

WOMENTECH. **Women in Technology Statistics: Where are We?** 2022. Disponível em <https://www.womentech.net/en-br/women-technology-statistics>. Access in: 2022.

YAU, N. **Visualize this: the FlowingData guide to design, visualization, and statistics**. [S.l.]: John Wiley & Sons, 2011.

APPENDIX A — ALUMNI FORM

INF-UFRGS: Mantenha a Conexão!

Cada aluno que já fez parte do Instituto de Informática (INF) contribuiu para a construção dessa instituição que cada vez mais fortalece sua vocação de excelência e qualidade na formação de recursos humanos, tanto na Graduação quanto na Pós-Graduação. Sabemos que nossos ex-alunos têm trajetórias de sucesso tanto no Brasil quanto no exterior, na Academia ou Indústria. Gostaríamos de estimular você, ex-aluno do INF, a manter contato conosco respondendo à pesquisa que elaboramos por meio da campanha "Mantenha a Conexão" (estimamos 5 min para responder).

Com esses dados, queremos criar e manter uma rede de contatos para nos comunicarmos daqui para a frente. Entendemos que manter sua conexão com o INF é de mútuo benefício: podemos divulgar oportunidades entre os alunos, divulgar palestras de interesse, oferecer atualizações profissionais e muitas outras possibilidades. Caso você permita, enviaremos atualizações por email.

Faça login no Google para salvar o que você já preencheu. Saiba mais

*Obrigatório

Aviso de privacidade *

☐ Ao avançar para a próxima etapa do formulário, você aceita que seus dados sejam utilizados em estatísticas pelo Instituto de Informática da UFRGS. Ressaltamos que, para a sua segurança, informações consideradas sensíveis nos termos da LGPD (Lei Geral de Proteção de Dados) não serão divulgadas.

Próxima   Limpar formulário

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. Denunciar abuso - Termos de Serviço - Política de Privacidade

Google Formulários

INF-UFRGS: Mantenha a Conexão!

Faça login no Google para salvar o que você já preencheu. Saiba mais

*Obrigatório

Informações Pessoais - Parte 1 de 2

Nome completo *

Sua resposta

E-mail pessoal *

Sua resposta

☐ Aceito receber notícias e informações em meu e-mail sobre o INF.

Com qual gênero você se identifica? *

○ Feminino

○ Masculino

○ Prefiro não informar

○ Outro:

País de residência *

Escolher ▾

Cidade de residência *

Sua resposta

Endereço residencial

Sua resposta

LinkedIn

Sua resposta

Qual foi o seu vínculo mais recente como egresso do INF? *

Escolher ▾

Ano de conclusão *

Escolher ▾

Escolher   ▼

Qual foi o seu vínculo mais antigo como egresso do INF?

Selecione caso tenha tido mais de um vínculo com o Instituto.

Escolher   ▼

Ano de conclusão

Escolher   ▼

Você teve um terceiro vínculo como egresso do INF?

Selecione caso tenha tido mais de dois vínculos com o Instituto.

Escolher   ▼

Ano de conclusão

Escolher   ▼

Voltar   Próxima   Limpar formulário

Nunca envie senhas pelo Formulários Google.

Google Formulários

INF-UFRGS: Mantenha a Conexão!

Faça login no Google para salvar o que você já preencheu. Saiba mais

*Obrigatório

Informações Profissionais - Parte 2 de 2

Nome da instituição/empresa de atuação *

Sua resposta

Local de atuação *

◯ Brasil

◯ Exterior

Cidade se Brasil / País se Exterior

Responda com base na pergunta anterior.

Sua resposta

Continua atuando na área de Computação? *

◯ Sim

◯ Não

Setor de atuação *

- ☐ Academia
- ☐ Indústria
- ☐ Empreendedor
- ☐ Governo
- ☐ ONG
- ☐ Outro:

Função na instituição/empresa *

Sua resposta

Renda bruta mensal

Órgãos como CAPES, Editora Abril (Guia do Estudante) e empresas que consolidam rankings de Instituições de Ensino Superior solicitam ao Instituto de Informática informações estatísticas sobre seus egressos como maneira de monitorar a importância e a qualidade das atividades sendo desenvolvidas. Se você estiver disposto a ajudar nesse sentido, pedimos que responda a pergunta opcional abaixo. Essas informações permanecerão sigilosas e, como já mencionado, serão utilizadas apenas para fins estatísticos.

- ○ Acima de 20 SM (R$ 19.960,01 ou mais)
- ○ De 10 a 20 SM (R$ 9.980,01 a R$ 19.960,00)
- ○ De 4 a 10 SM (R$ 3.992,01 a R$ 9.980,00)
- ○ De 2 a 4 SM (R$ 1.996,01 a R$ 3.992,00)
- ○ Até 2 SM (até R$ 1.996,00)
- ○ Sem renda

INF-UFRGS: Mantenha a Conexão!

◯ Sem renda

| Voltar | Enviar | Limpar formulário |

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. Denunciar abuso - Termos de Serviço - Política de Privacidade

Google Formulários