*Authors:*
Alami-Idrissi Ali - Benlalah Ali - Fikrat Zakaria

PROCESS BOOK

Overview, motivation, target audience:
One joy of joining EPFL is receiving a CAMIPRO card which gives you the feeling that you're officially part of the student's community. When we first got it, we were impressed that some statistics about the school were printed on the card's back. We really liked the idea but still wished to get to know more about the school and have the control on what data to see.
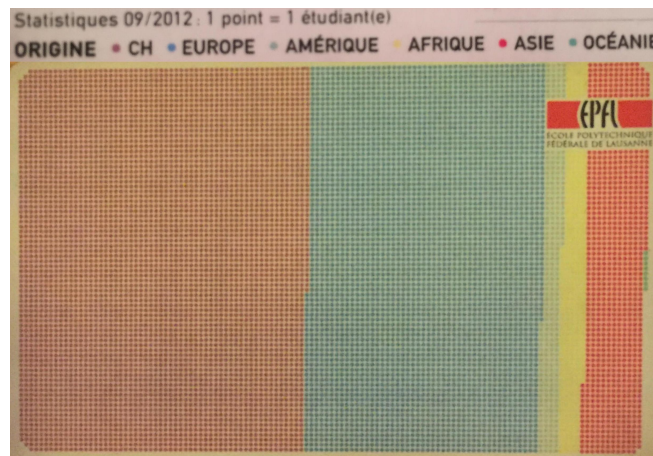
Our visualisation is an attempt to recreate a complex version of the old CAMIPRO by showing different statistics about EPFL students. It targets EPFL people who are interested in getting insights about their school and future EPFL members who want to have an idea about its diversity and students' distribution.

Related work and inspiration:
Our inspiration was mainly coming from two parts:
The first one concerning the idea that we have developed by having a look at the data shown on the CAMIPRO of nationalities at EPFL. The second is about the design from this visualisation https://www.nationalgeographic.com/what-the-world-eats/ which shows the evolution of the world's food consumption within time through dynamic donut charts. We found that this visualisation is really expressive and gives the user a great experience while playing with the data.

What are we trying to show with our visualization?
The main purpose of our visualisation is to show the internalisation of EPFL, as well as the evolution of student's number. Our graphical interface is also designed in such way to highlight the sections with most students throughout time.

Where the data comes from, and what are the processing steps?
First things first, we had to gather data about students. This was done by extracting it from the is-academia website which provides public Excel files featuring information about enrolled students. For our convenience, we choose to collect data from 2007 to 2017.
The structure of excel files is as follows:
Every file contains headers that separate students enrolled in different programs.
Every entry in the excel files contains the gender, the name, the SCIPER and the nationality of a student along with other information not used in our project.

Below is an example of a row without the non-used columns:

AR Exchange, 2014-2015, Autumn semester (28 ét.)						
Civilité	**Nom Prénom**	**Statut**	**Type Echange**	**Ecole Echange**	**No Sciper**	**Nationalité**
Mister Ano Nyme		Present	Erasmus	Universitat Politècnica de Catalunya, Barcelona	12345	Spanish

Now that we have collected the data, a processing was made to clean it in order to facilitate the analysis.

For each student entry (a row in the Excel file) we associate its corresponding header (e.g. the entry of a student enrolled in Bachelor 1 in Computer Science in 2017-2018 will have a new column containing "Computer Science 2017-2018, Bachelor semester 1" which is the header associated to the row in question)

We then parsed the header associated to each entry to extract the following information:
- The academic year
- The student's section
- The student's academic degree which determines if the student is a master, bachelor, exchange, CMS or a student doing a "HES pathway"

After that we cleaned and normalised multiple fields:
- We kept only one citizenship for students having multiple ones
- Stripped non-printable characters from our cells
- Generated unique SCIPER numbers for students without one

Since the collected excel files lists enrolled student at EPFL per semester, we had to keep only one entry per student for each year. Finally, we translated citizenship denominations to English and created a dictionary which maps citizenships to countries codes. The latter will be used for building to render the map.

Data anonymization:
In order to respect students' confidentiality, we removed the name and SCIPER column

What visualization have we used to gain insights on the data?
Since our idea is to build a dashboard, we simply used the latter to explore our dataset. Besides, our main contribution is to provide the reader with meaningful insights regarding EPFL students' growth in the past 10 years, which constitutes the "Discover" feature. (see below for further explanations)

What are the different visualizations we considered?
Our main goal is to show for each academic year between 2007-2017:
- The total number of students at EPFL
- The gender ratio
- The number of students per section
- The number of bachelor, master, exchange, CMS student as well as the number of students doing the "HES pathway"
- The number of students per country

We wanted to allow the user to filter data per gender, nationality, section and student type. Hence our main challenge was to represent our multidimensional data in a compact and elegant way. Our first attempt was to represent the gender using a bar chart, nevertheless, we ended up representing this attribute using dynamic SVG icons which were filled in function of the gender ratio.
To represent the academic degree attribute, we first used a donut chart. However, it didn't fit well our data since some categories had a small ratio and were barely visible, making the comparison between different parts of the chart tedious. Subsequently, we decided to choose a unique stacked bar were each chunk represents a category. The result was a neat plot leaving more space between features for a better visual experience.

The most challenging feature to represent was the distribution of students per major as it contains around twenty categories. Our first try was to represent every major on its own as a gauge chart with a filling level proportional to the major's weight (i.e. % students in the major).

However, this was not a viable solution since we ended up with multiple gauges driving away the user's attention.
A solution we tried was grouping majors by faculties and representing them as a multi-layered donut chart. The first layer contained faculties ratios and the second showed major's ratios. Nevertheless, the small ratios visibility issue rises again.
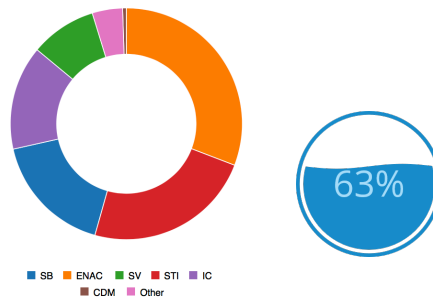
Our final solution consisted of using multi-layered horizontal bar chart which solved the aforementioned problem by leaving more space between each category.

**Did we deviate from our initial proposal?**
Fortunately, no :).

Description of the interactive visualization implemented:
Our data is composed of multiple independent components each responsible for representing a variable of our data.

- We use a bar chart to represent the number of students per year. Whenever a component (year, gender, section etc.) is selected, the data is filtered according to that choice.
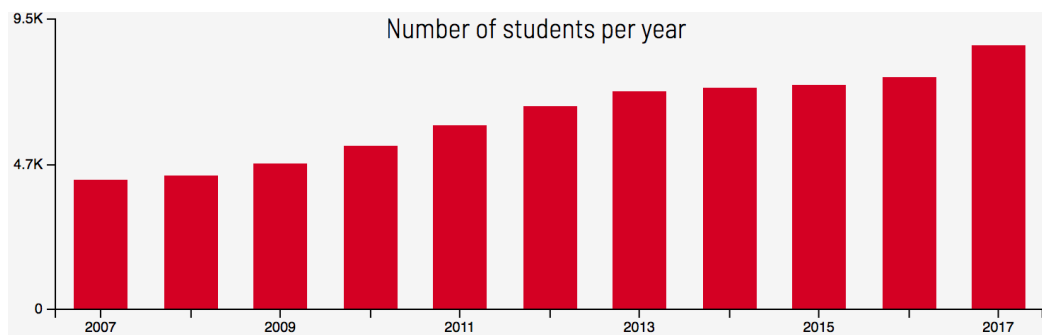

- A multi-layered horizontal bar chart displaying the number of students per faculty and major.
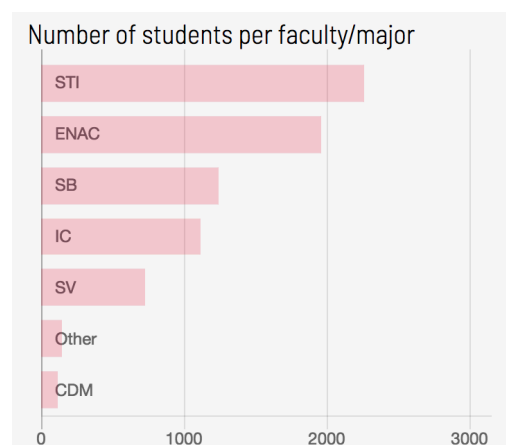
- SVG icons representing male/female are filled depending on the gender ratio. The exact number and percent are displayed near each icon. When clicked after selection, filtering by gender is removed.

5.5K (73%)

2.1K (27%)
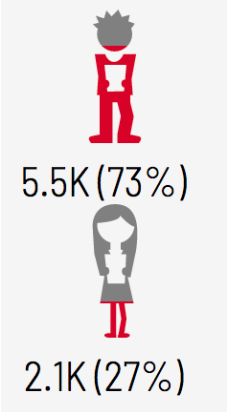
- A unique stacked bar classifying students per academic degree (Bachelor/Master/Exchange/CMS), displaying the number in each when hovering over. When clicked after selection, filtering is removed

Proportion of bachelor and master students

● Bachelor ● Master ● Exchange ● CMS
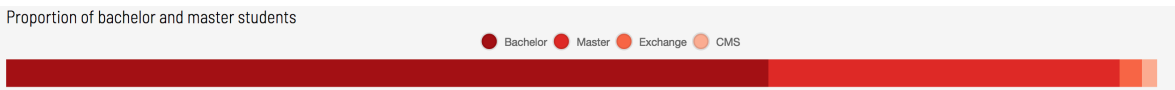
- A chloropleth map showing the number of student per country. When hovering over a country, the exact number of students from it is displayed.

Number of students per country

+
—

Country
Hover over a country

1–6
6–28
28–142
142–738
738–3845

Leaflet
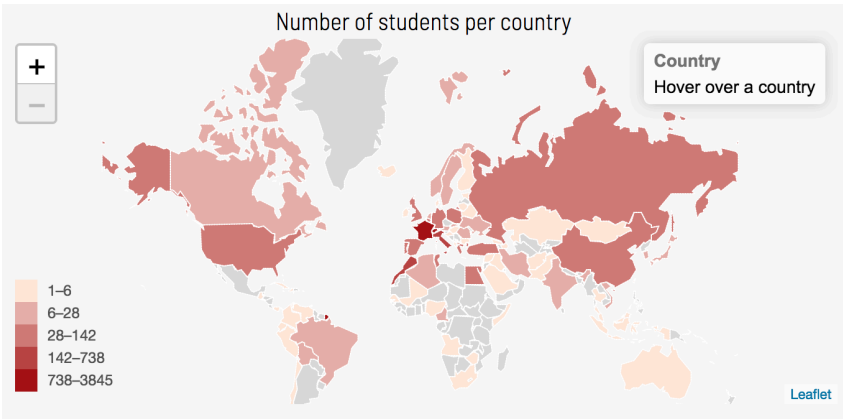
- A title giving the summary of selected filters.

Swiss students majoring in IC

- The total number of students satisfying the current selection is shown below the title.

8.6K

- Two buttons allowing to reset the currently selected filters and to discover meaningful insights about the data.

[Discover] [Reset]

What did we learn about the data by using our visualization and what are the further improvements to be made?

By using our visualization, we learned and got insights about the students we didn't had the ability to see using the raw is-academia data. Also, it is done in a fun gamified way which, from many users' feedbacks, makes you stick with it and want to know more and more about the data.

As students in EPFL for more than 3 years, we discovered interesting facts such as the stagnation of the Swiss student's number since 2013 with the increase of student's number from other nationalities, which really shows how international EPFL is. Another fun fact is that females are less attracted by Engineering majors compared to males.

Our visualisation works well, but due to lack of time we couldn't include the following feature to make the interaction more user friendly:
- Add a panel showing the currently selected filters
- Allow selection of multiple countries at the same time
- Allow selection of multiple academic degrees at the same time

Peer assessment: (Preparation, contribution, respect for others' ideas, flexibility)
Ali Alami-Idrissi: Yes, yes, yes, yes
Ali Benlalah: Yes, yes, yes, yes
Zakaria Fikrat: Yes, yes, yes, yes