Interactive Visualization of MIT European Career Fair Statistics

Mustafa Doga Dogan
MIT CSAIL
doga@mit.edu

ABSTRACT

In this project, we develop an interactive webpage that visualizes the statistics of the MIT European Career Fair. It includes numerous visualizations about both the student candidates and employers. We hope that this tool will allow both parties to inform themselves better about the fair to achieve their goals, which are to find jobs in Europe and to attract best talents from top US schools, respectively.

Author Keywords

Interactive data visualization; career fairs.

INTRODUCTION

The MIT European Career Fair (ECF) is an annual recruiting event in the Greater Boston Area, organized by members of the MIT European Club [8] at the Massachusetts Institute of Technology (MIT). Its mission is to connect top-tier US students and young professionals with European companies, universities, and non-profit organizations and help them learn about career opportunities in various fields, including robotics, pharmaceuticals, finance, aviation, and artificial intelligence [9].

Organized since 1997, the ECF is the largest of its kind in the United States and has a proven track record providing domestic, foreign, and multinational companies access to some of the highest caliber applicants worldwide. Attending candidates are students and recent alumni of and renowned European and American universities. Company participants are diverse and range from renowned multinationals to smaller, innovative firms and startups.

Visualizing the statistics of the fair is important for both MIT students and attending companies.

For MIT and its students: Every year, the European Club donates most of the ECF revenue to MIT for the funding of MIT International Science and Technology Initiatives (MISTI) programs in Europe (ca. $40,000). For more information, please refer the MIT News articles that were published in previous years [2, 3].

For employers: The main reason why companies come all the way from Europe to Boston is to meet and recruit top students in US institutions. Particularly, they are most interested in MIT and Harvard students. After each year's fair, they follow up and request statics about the attendees, including their schools, degrees, and fields of interest. Knowing and conveying this information in a clear convincing way ensures they keep coming to ECF every year and we keep donating our revenue to MISTI programs.

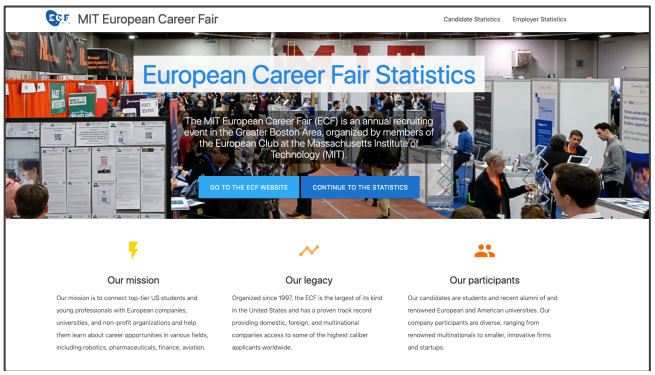

Figure 1: This project involves the design and development of a Web-based visualization of MIT European Career Fair statistics.

For this reason, I have built a Web-based interactive visualization (Figure 1) that tells the story of the ECF by providing insights on attending candidates and employers.

RELATED WORK

In previous years, ECF organizers used different ways to visualize the statistics, which they later sent to companies that have attended the fair or consider attending it in the future.

To better understand previous fair organizers' strategy, I provide some of these visualizations in this section, which I found as I was digging the archives of the European Club. In these promotional materials, I saw that most of the charts used were pie charts, such as the one in Figure 2 from 2008.

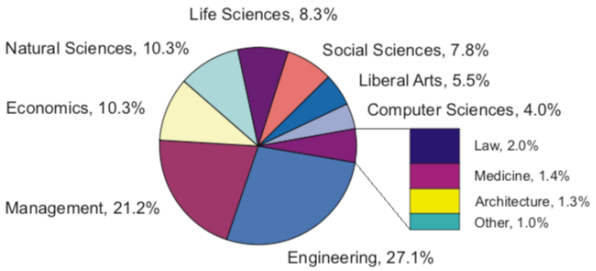

Figure 2: In 2008, a simple pie chart, alongside stacked bars to represent smaller data elements, was used to visualize the fields of study of attending candidates.

A technique that was used in the following years was making these pie charts three-dimensional (Figure 3), which I do not believe is a more effective way to convey information about the fair.

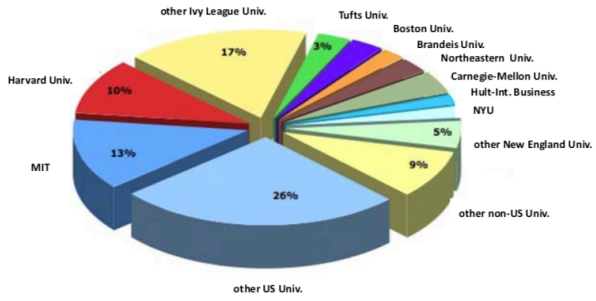

Figure 3: In 2011, a 3D pie chart was used to visualize the universities of candidates.

A more fundamental visual encoding mistake was made in 2017 graphs as the organizers chose to use different shades of the same color in pie charts, making it almost impossible to distinguish different categories (Figure 4).

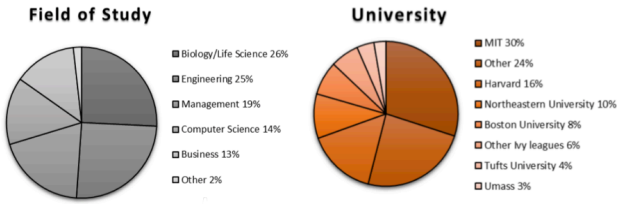

Figure 4: In 2017, the use of color gradient for non-numerical data made it much harder to read the charts.

After having identified these visualization issues in previous years, I tried to avoid similar design mistakes in my own design and development process.

METHODS

The method of developing this tool involved multiple steps regarding the data and its presentation: data exploration and wrangling, followed by the design and development of the Web-based visualization.

Data Exploration

I am one of the organizers of the fair, so I had access to the registration data. The raw dataset that came from the online fair registration platform was relatively large (e.g., 2090 columns and 2961 rows for the 2020 fair) and needed a lot of processing and manipulation. To understand the data better, I explored it in *Tableau* and *Microsoft Excel* first.

One of the first things I did was to rank top attending universities in a bar chart in *Tableau* (Figure 5), which showed some inconsistencies caused by user input during the fair registration process. For instance, equivalent names (e.g., abbreviations) were used for certain school names (e.g., MIT or NYU vs. the full names). Furthermore, many users left that field empty, causing *N/A* to be the top bar.

Continuing the exploration in *Excel*, I noticed that these initial charts had even more flaws because of the way the raw CSV file was formed. For example, for the "university" and "degree" columns, multiple values were present for each candidate (i.e., in case they have marked themself as having multiple degrees from multiple schools). Some examples are provided in Figure 6.

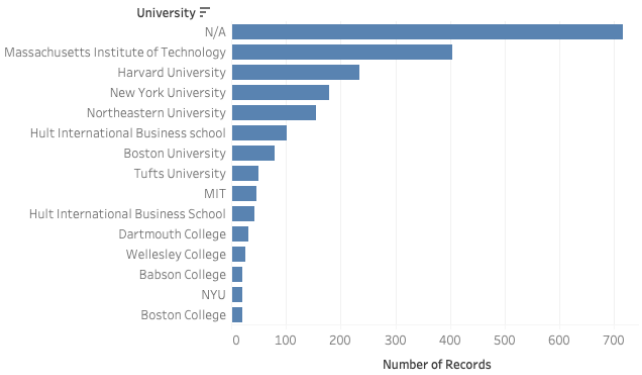

Figure 5: Data exploration in Tableau. The raw *University* data was ranked in a bar chart.

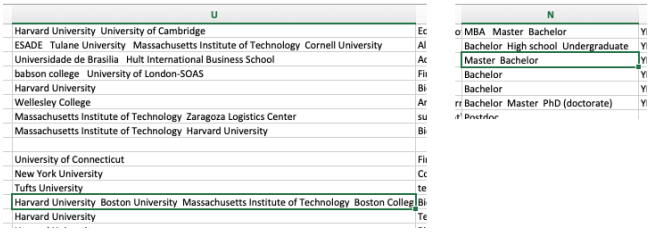

Figure 6: Data exploration in Excel. University (left) and Degree (right) columns have multiple values in many rows.

Data Wrangling

I wrote a Python script to wrangle the data and fix the aforementioned issues. I used *Pandas*, a data analysis library [4]. The code can be found as a Jupyter notebook in the GitHub repository.

Essentially, the script has a priority list (e.g., "Postdoc" comes after "PhD") for processing relevant fields (highest degree, university name, field of study). I created these rankings based on the interests of employers. For example, if a student has multiple affiliations and one of them is MIT, they are listed under MIT. If their fields of study include data science and industrial engineering, they are listed under the former. The script also combines equivalent names (e.g., "MIT" and "Massachusetts Institute of Technology"). Finally, it creates a crosstab sheet showing the number of attendees in a separate column for each subgroup (e.g., PhD students at MIT studying computer science: XX many attendees).

Web design and development

For the Web-based visualization, I used the *D3.js* [1] and *D3plus* [6] JavaScript libraries, as well as the *Materialize* CSS library [10], which helped me create a responsive design.

My previous class assignment also used this dataset to visualize a smaller subset of it. For the final project, I consulted other co-organizers of the fair and asked for their feedback

to extend the visualization in the most useful way. For this reason, the new features and sections are based on the feedback I have received from them.

Also, based on the feedback from the course staff, I intended to have a narrative visualization that tells a story about the ECF. More specifically, I decided to use the *Martini Glass* structure [5], which begins with an author-driven approach that first provides observations and written articles to introduce the visualization. Once the intended narrative is complete, the visualization opens up to a "reader-driven stage," where users are "free to interactively explore the data."

Following this logic, I first wanted to clarify on the webpage itself why this visualization is important to both MIT and the employers (for reasons explained in the introduction section), so I inserted a short section at the top, which also includes links to previously published MIT News articles about the fair and its contribution to MISTI (Figure 7). Since I knew the information that students and employers look for are different, I also created two distinct major sections: *Candidate statistics* and *Employer statistics*. These sections include various visualization techniques and interaction types, which I will explain in the following results section.

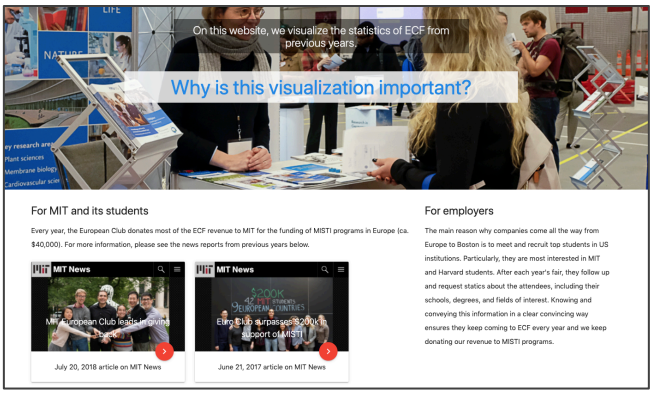

Figure 7: Before providing the charts, a section has been dedicated to explaining the importance of this visualization.

Lastly, for the page background, I added large pictures from the fair itself based on my co-organizers feedback as they thought this would made the story more tangible. I used the parallax scrolling effect (page elements scrolling at different speeds) for a visually appealing viewing experience, mostly inspired by Wonyoung So's *Cartographers of North Korea* [7].

RESULTS

In this section, we show several visualizations from the webpage and some of the insights that we got from them.

Some of the initial charts about the candidates show statistics about their schools and fields of study (Figure 8). We see that the majority of the candidates are MIT and Harvard students and graduates. Furthermore, the tree map about shows a comparative view of different majors over the years. We can observe recent trends in this chart, for instance, the block representing data science goes from less than 1% to >7% in three years. These figures have temporal data; the year can be adjusted from the bottom selection.

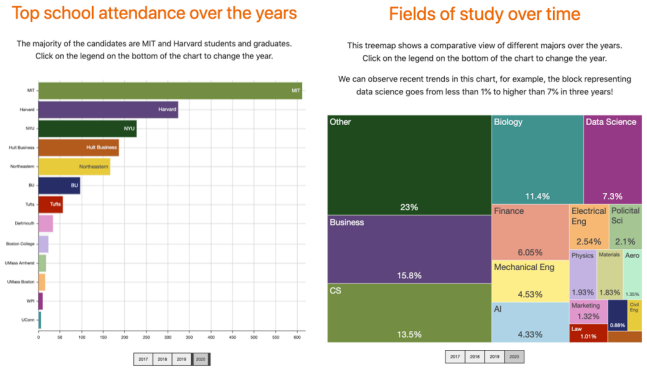

Figure 8: Top school attendance (left) and study fields of candidates (right) over the years.

One of the most sophisticated graphs shows a breakdown of different study fields for the highest attending universities. These two categories of information are those that are most sought-after by the attending organizations. For this purpose, a stacked bar chart was the perfect means (Figure 9). The different colors used make it easier for the viewer to distinguish different majors. Hovering over the bar sections shows the number of attendees in that specific cross-section of the data. By clicking on individual titles in the legend, they can also focus on specific majors and compare the number of attendees across universities. When clicked, the axes are rescaled in an animation to adapt to the new range of the sub-data.

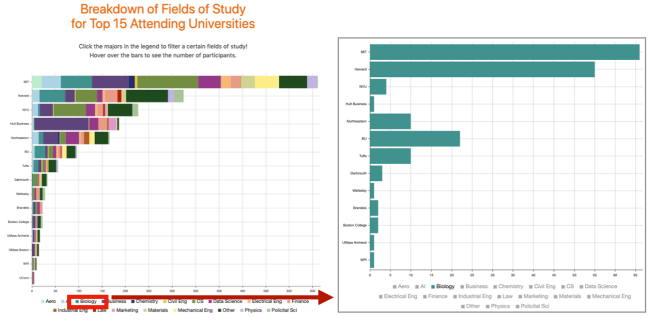

Figure 9: A stacked bar chart depicting the breakdown of fields of study for universities (left) turns into a major-specific view (right) by clicking one from the legend.

As for the employer statistics, the page has a map of European countries. Basically, it is a heatmap showing the number of attending organizations coming from each country. We see that the top attending countries are Germany and France. Clicking a country shows some of the companies from that country. Figure 10 shows an example for France.

DISCUSSION

The webpage was received very well by the ECF organization team. They were even more content after implementing

the changes that they had requested. Some of the co-organizers have noted that such a Web-based presentation of the statistics is much more interactive than a static PDF document.

Figure 10: Users can click on a country to view the respective organizations attending.

FUTURE WORK

We are planning to add even more statistical charts about the employers and make this visualization public (not just on GitHub pages) by publishing it on the ECF website. We hope that our tool will be useful to prospective candidates and employers in the future.

ACKNOWLEDGMENTS

I would like to thank Rishabh Chandra for introducing me to the Pandas library during office hours, as well as Jumana Almahmoud for giving me suggestions for this assignment.

REFERENCES

[1] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D^3 Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (December 2011), 2301–2309. https://doi.org/10.1109/TVCG.2011.185

[2] Lily Keyes. MIT European Club leads in giving back. *MIT News*. http://news.mit.edu/2018/mit-european-club-leads-in-giving-back-0720

[3] Caroline Knox. A unique partnership. *MIT News*. http://news.mit.edu/2017/mit-european-club-raises-200k-for-misti-fellowships-in-unique-partnership-0621

[4] Pandas. Python Data Analysis Library. https://pandas.pydata.org/

[5] E Segel and J Heer. 2010. Narrative Visualization: Telling Stories with Data. *IEEE Trans. Visual. Comput. Graphics* 16, 6 (November 2010), 1139–1148. https://doi.org/10.1109/TVCG.2010.179

[6] Alex Simoes. D3plus. https://d3plus.org/

[7] Wonyoung So. Cartographers of North Korea. https://cartographers-nk.wonyoung.so/

[8] MIT European Club. *euro-club*. https://www.euroclub.mit.edu

[9] MIT European Career Fair. https://euro-career.mit.edu/#!/

[10] Materialize. https://materializecss.com/