1 Introduction

The purpose of this visualisation is to bring awareness to sexual misconduct that often occurs in academia. After coming across the "Sexual Harassment in the Academy" dataset, I came to the decision that the data collected needs to be communicated more effectively.

The main focus is on a specific subset of the data - the event itself, the rank of the target, the rank of the perpetrator, the academic field that the victim and/or perpetrator belongs/belonged to. The goal is to unveil which academic fields appear to have a higher frequency of sexual misconduct, which academic rank appears to have higher incidences of sexual harassment and humanizing the data by allowing our audience to read these personal stories of abuse.

2 Description

Data –

The anonymous crowdsourced survey, "Sexual Harassment in the Academy" [4] was made by Dr. Karen Kelsky, a former anthropology professor. The motive behind it was to provide the victims of sexual harassment/assault a space to report their experience without the fear of consequences.

The dataset [5] contains information about the nature of abuse, rank of the victim and the perpetrator, academic field of the victim/perpetrator, institutions' responses, punishments given (if any), and effects of the abuse on victim's mental health, career and life in general. Since the data was nominal and open to different kinds of interpretations, it had inconsistencies and the data needed to be cleaned and converted to JSON.

In order to organise data properly, ranks were given to each quantitative value of the academic system. "Not disclosed" was given a rank of 0, and students, as they rose through the academic system (K-12, Undergraduate, Graduate, Masters, and PhD), were each given a value of 1. All technicians, lecturers, and instructors, along with any adjunct or visiting professors fall under the category of faculty because I feel that they hold temporary status at an institution. On the other hand, those who were just starting their academic teaching career are assistant professors and those who are tenure-tracked are associate professors. Administrative includes all deans, chairs, and department heads who might hold more power in the office over other tenured professors.

After the additional cleaning and making sure that each instance was consistently labelled, all fields with 20 or fewer total instances were excluded, given the small sample size.

Tasks –

The overall communication goal is to show the extent of sexual misconduct taking place in academia as a result of hierarchal imbalance of power. The user should be able to discover which academic fields appear to have a higher frequency of sexual misconduct and understand the result of imbalance of power in a majority of the accounts. Looking at the individual accounts, the user discovers the prevalence of sexual harassment and lack of support for its victims in academia.

Approach –

There is already a functioning solution created by Dr. David R. Karger, an MIT professor. Dr. Karger created a very basic tool [6] which allows users to see the dataset in a somewhat readable format. However, it is primarily text-based.

With my solution, I also wanted to communicate a similar message but wished to make it better so it is both visually pleasing and impactful. For this purpose, I went with D3.js. By utilizing a stacked bar chart to represent the data and by encoding each data point as a dot on our visualization, my intention was to immediately communicate to the audience and to create understanding of the prevalence and sheer amount of sexual harassment in academia. Additionally, rather than having the audience try to compare and track the incidents by rank and discipline, using our filtering and sorting functionalities, I

also allow the audience to quickly perceive the differences. I allow my audience to select an individual data point and to read the victim's actual account/narrative. I feel that this aspect of the solution is what is most impactful and striking to the audience as it humanizes the data and frames it in a sobering context.

The solution allows users to switch between three different views, to filter by target or perpetrator role, and to select the individual points to read details about each event through elaboration. Interactivity was chosen to present the different views instead of a method such as small multiples because the points are sorted differently in each view. To make this difference in sorting order clear to user(s), staged transitions were used to first change the colour hue of the points (according to the scale of the selected view), and then change the position of the points.

Dynamic baseline approach was chosen for filtering by target or perpetrator status, moving the points from the selected category towards the y-axis to allow for better comparison of the filtered value between academic fields. Staged transitions were also used here, first reducing the opacity of the points not in the selected category, and then changing the position of the points.

Finally, I chose to use interactivity to allow users to obtain details on demand for each of the incidents through elaboration. When a user selects a point from the plot, they are shown a table containing information about that particular incident, including the target status, the perpetrator status, the description of the incident, the institution where it took place, the institutional response to the incident, and the impact of the event on the target's life, career, and mental health. By allowing users to read about each incident, the qualitative nature of the survey responses is being preserved, which I believe helps to (1) humanize the data, and (2) elicit a stronger response from the user(s).

Limitations –

As the dataset is voluntary crowdsourced, the data is somewhat biased as it represents only the incidents that victims have chosen to report. For example, while English appears to be the discipline with the most incidents of sexual harassment, it's a possible that other disciplines, like STEM, have an equivalent number of cases of sexual misconduct, but cultural norms within STEM fields may prevent victims from reporting, even anonymously. Furthermore, while the gender of the perpetrator is given, the survey doesn't ask for the gender of the victim; it is normatively implied that a majority of the victims are women.

References

[1]Knuth. (n.d.). *Computers and Typefaces.* Retrieved from http://www-cs-faculty.stanford.edu/̃uno/abcde.html

[2] Munzner, T. (2014). *Visualization Analysis and Design.* AK Peters / CRC Press.

[3] *Overleaf.* (n.d.). Retrieved from http://www.overleaf.com

[4] https://docs.google.com/forms/d/e/1FAIpQLSeqWdpDxVRc-i8OiiClJPluIpjMlM41aUlU2E0rrQ4br_rQmA/viewform

[5] https://docs.google.com/spreadsheets/d/1S9KShDLvU7C-KkgEevYTHXr3F6InTenrBsS9yk-8C5M/edit#gid=1530077352

[6] http://people.csail.mit.edu/karger/Exhibit/Harass/

[7] https://github.com/d3/d3/wiki