Gender in 30 Years of IEEE Visualization

Project Repository link: https://github.com/Joel-kiran/DataVisualization.git

Group Members:

1. Joel Kiran Kumar - joelkik@clemson.edu, CID - C17113752
2. Pooja Yadav - poojay@clemson.edu, CID - C37076363
3. Chaitanya Gawade - cgawade@clemson.edu, CID - C21147285

Background and Motivation:

We chose this project idea to study and showcase the diversity among the authors, committee members, and award winners at the IEEE Visualization (VIS) conference over the last 30 years. Our primary focus would be on the female gender and to visualize their trends in career growth, number of awards, number of publications, collaboration, carrier age, dropout rate, etc., compared to males. We would try to answer these questions by representing them with appropriate visuals.

Project Objectives:

While going through the research paper, we encountered some primary questions that we will try to answer with the help of visualization; they are as follows:

1. How is male/female gender represented among VIS authors, committees, and award winners, and how has their representation changed over the years?
2. What are gender-related collaboration and publication patterns in the VIS community, and how have they evolved?
3. How do different gender collaborations contribute to the publication count?
4. How is credit shared among female authors and analyzing co-authorship relationships in the VIS community?
5. What is the percentage of female authors by position each year?
6. What is the percentage of awardees per gender compared to the overall proportion of female/male authors?

By the end of this project, we would know how the growth of the women have been, how many awards they received, and how many senior positions they hold; it would also help answer how old the women are compared to males in different segments over the decades. We would know the trend in female percentages over the decades in various features of the conference.

Data and Data Processing:

In this project, we would be utilizing the dataset from the website - https://osf.io/ydfj4/, and for additional information, we would also refer to this website - https://nyu.databrary.org/volume/1301. The dataset contains meta-data such as Year, DOI, Author name, Author gender, Career age, etc. The dataset is primarily clean for the data processing part, so we will combine appropriate files listed in the dataset and use the other files as-is.

Must-have Features:

The essential must-have features for our visualization would be gender, year, total authors, female authors, male authors, conference, career age, dropout rate, number of female publications, number of male publications, percentage of females in first, last and middle positions by year, publication paper count by female, male collaborations, percentage of female in OC and PC over the years, average carrier age of OC and PC members and percentage of female awardees. We think these are critical features that will make our visualization look more expressive and better convey the information.

Optional Features:

Our primary focus would be to ensure that we have implemented the must-have features first, but we won't stop there itself. We will keep enhancing our visualization and will add more features to it. The optional features for our visualization would be a percentage of female co-authors by publication, average carrier age of all, new authors over the years, average carrier age among awardees, average publications counts among awardees.

Project Schedule:

From Date	To Date	Tasks
10-18-2021	10-31-2021	We plan to implement the charts as mentioned above.
11-01-2021	11-07-2021	To decide on the final layout for the website prototype and make corrections in the previous task, if any.
11-08-2021	11-14-2021	We will work on the Project Review comments.
11-15-2021	11-29-2021	We will work on polishing the website and fixing

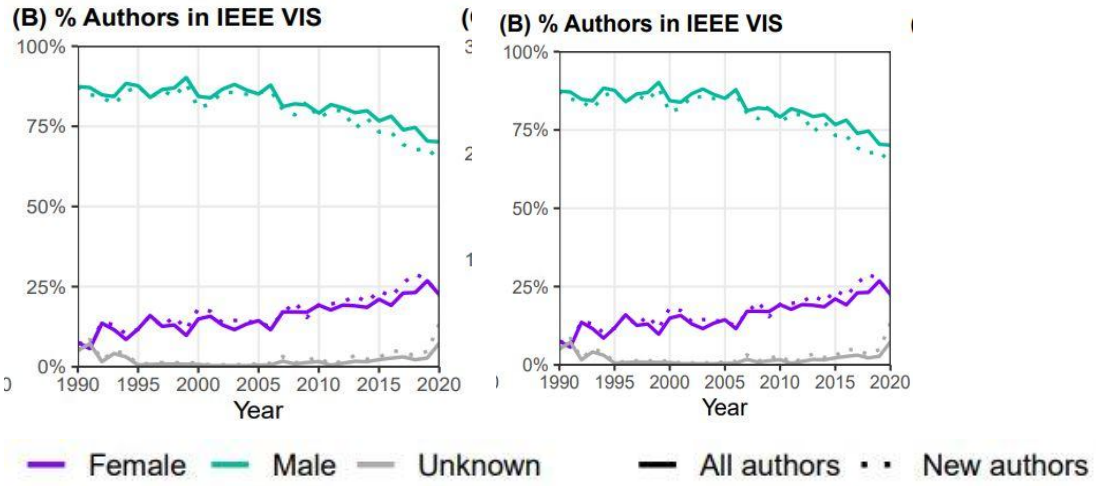
		the issues.
11-29-2021	12-04-2021	We will work on the Final Project Report and Project screencast.

Visualization Design:

LAYOUT 1

We plan to place the below figure as it appears; figure1 would be on the first row of the website, figure2 would be on the second row, figure 3 on the third row, and so on.

Figure 1



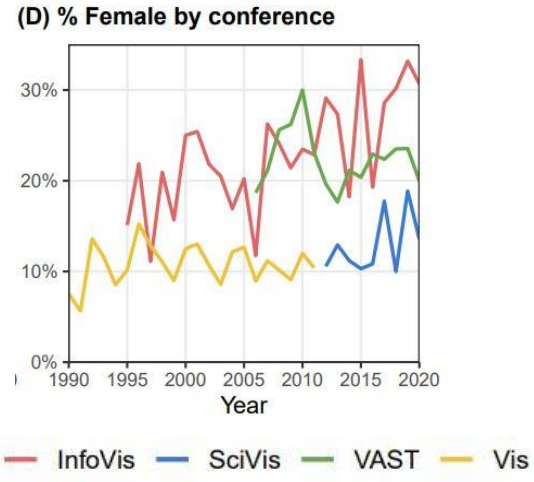


Figure 2

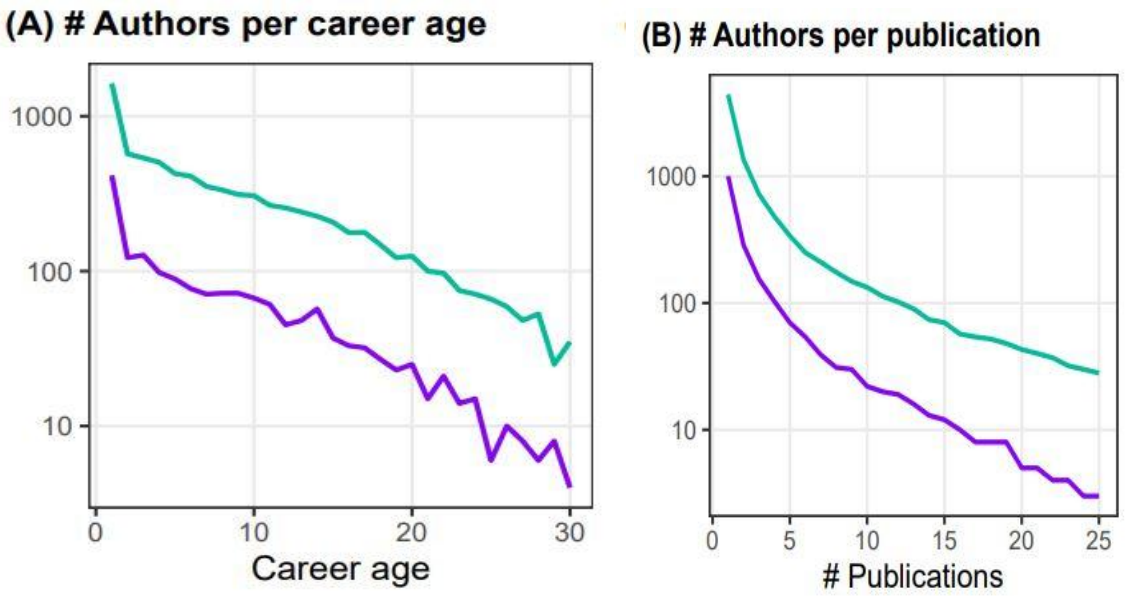


Figure 3

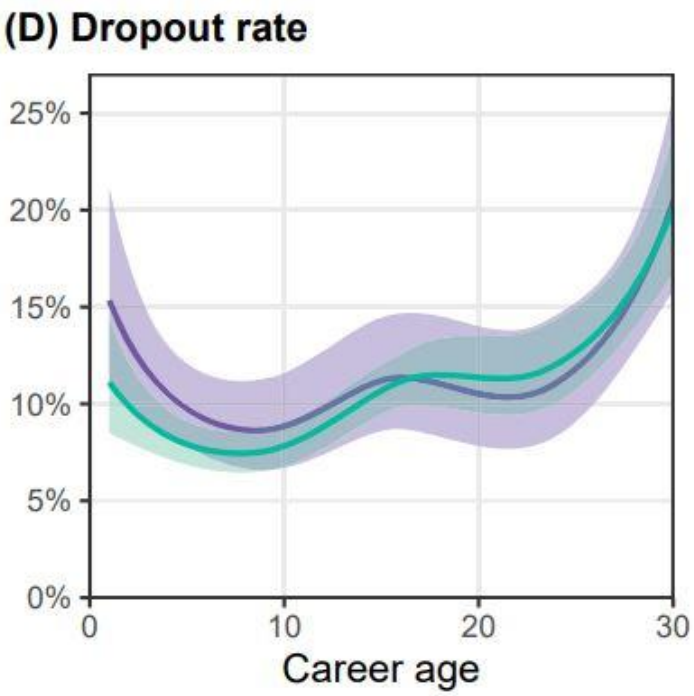


Figure 4

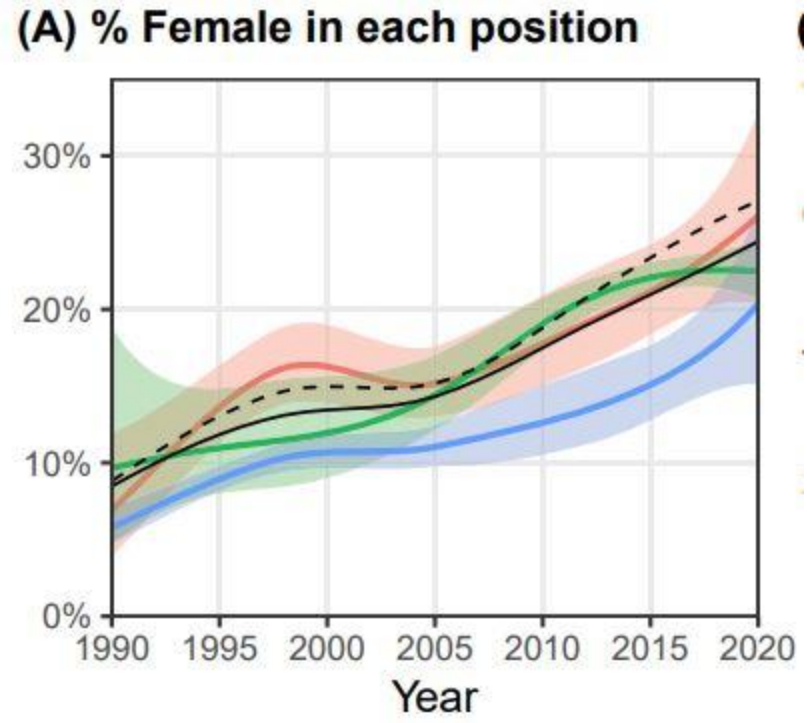


Figure 5

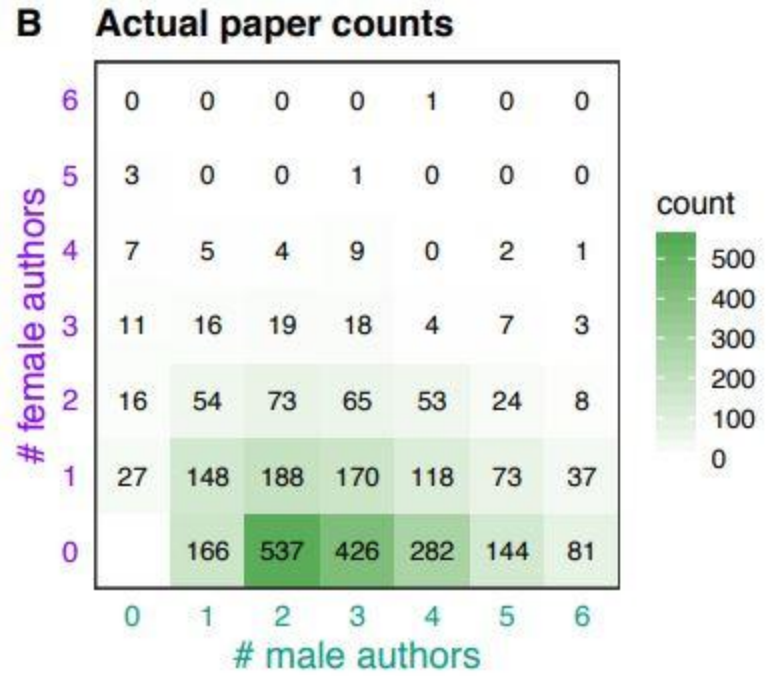


Figure 6

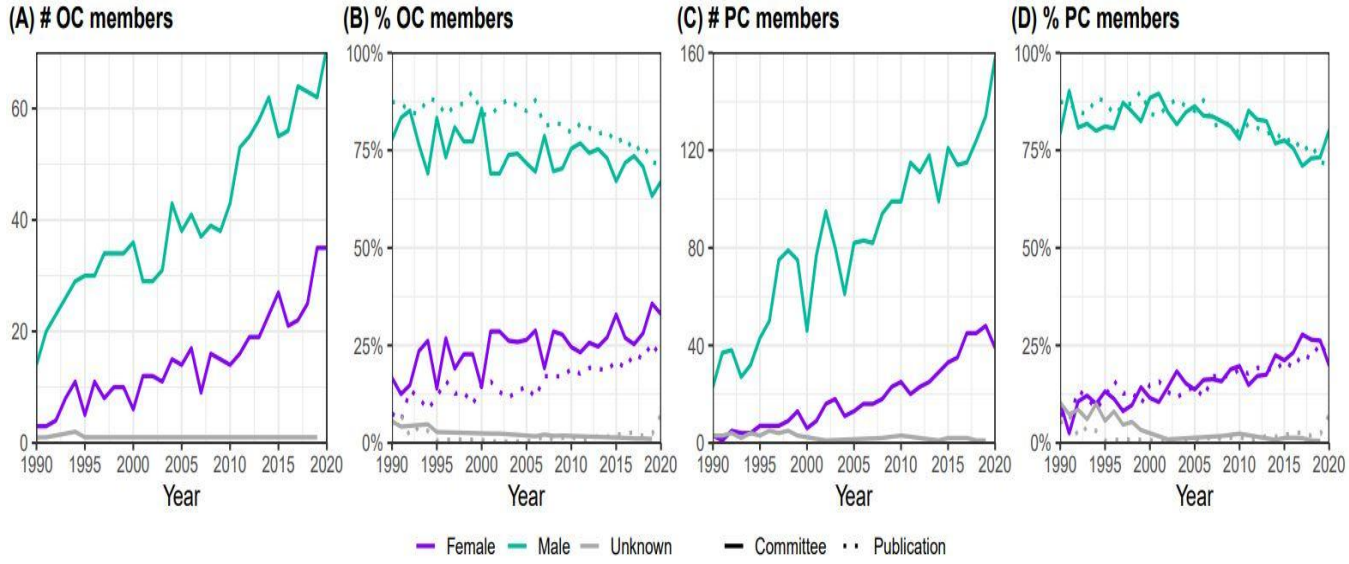


Figure 7

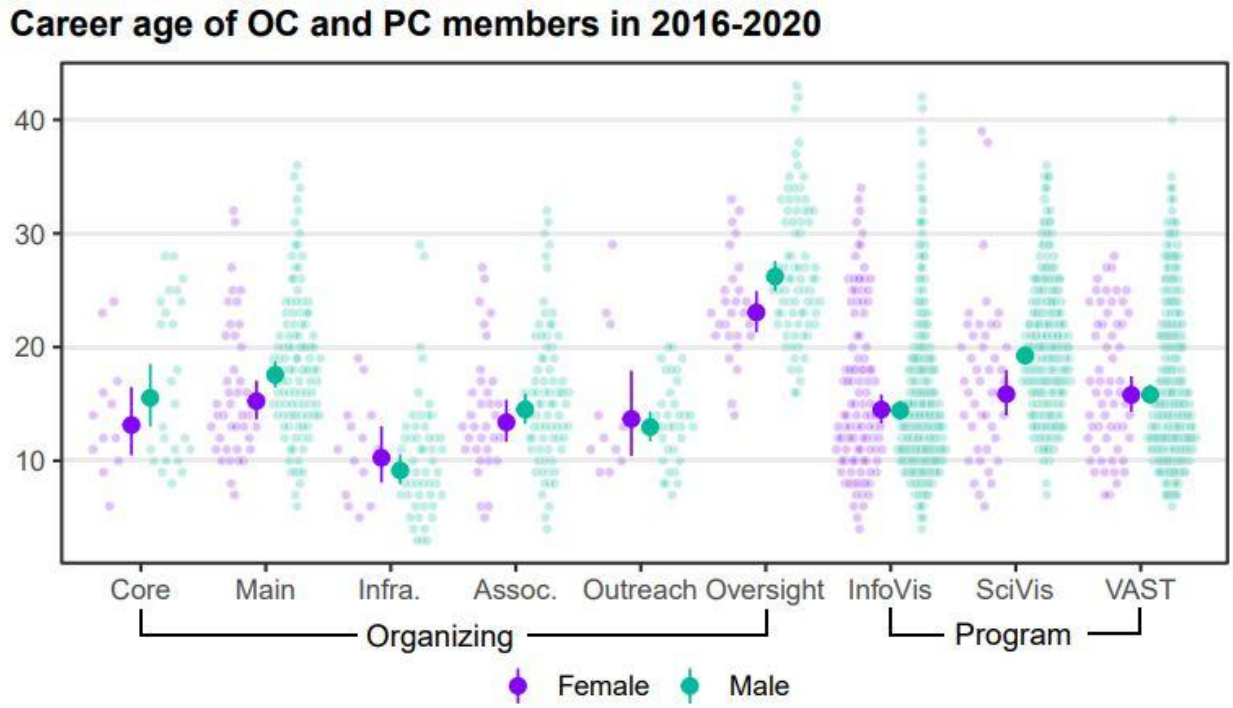
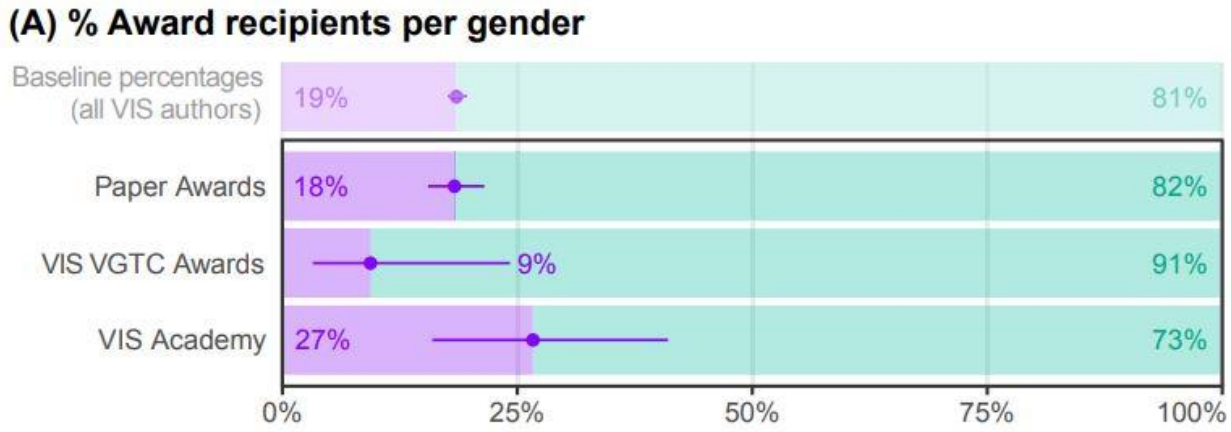
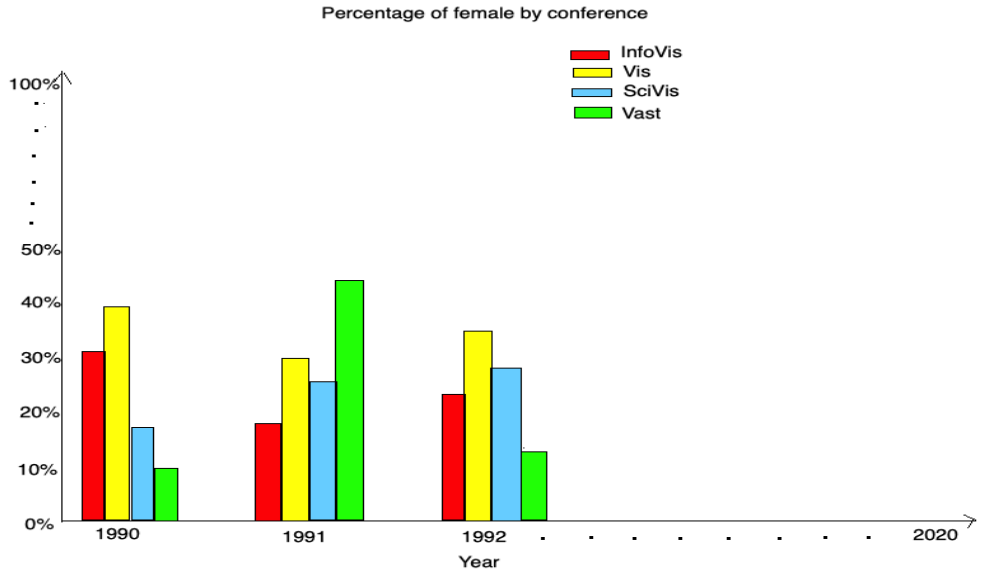


Figure 8

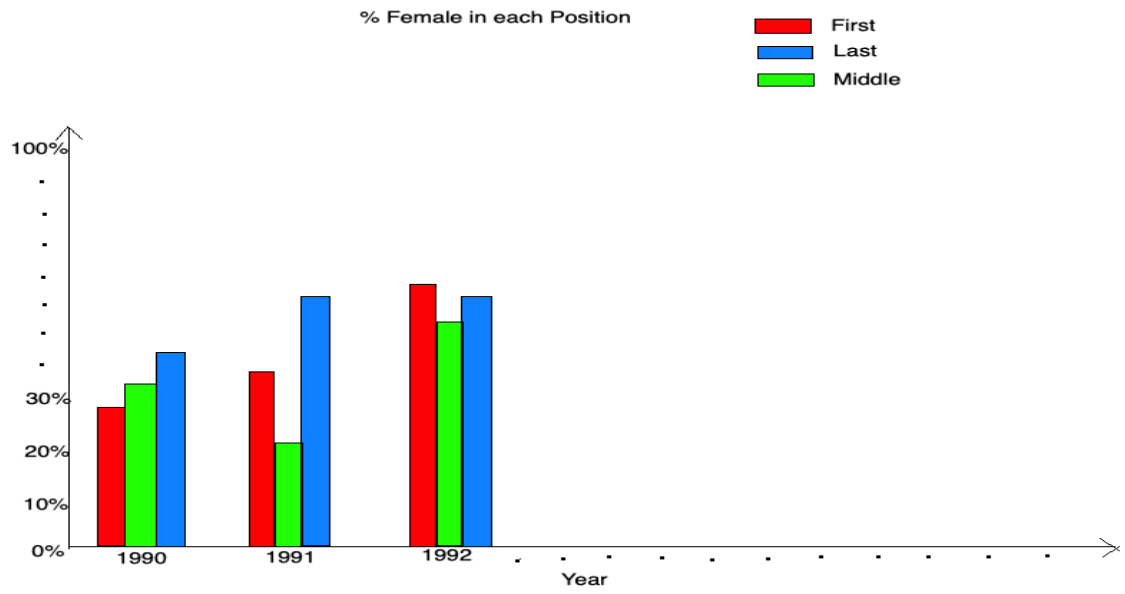


Layout 2

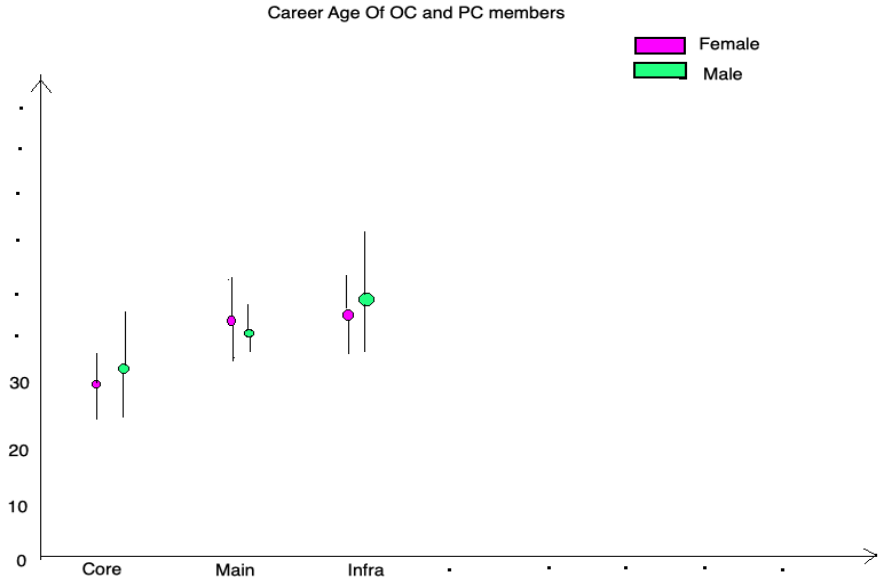
We plan to change figure 1d, 4a and 7



We plan to use a grouped bar graph instead of line graph in 1d because this gives a better representation to compare the four conferences for the specific year and explains the trend over the years.

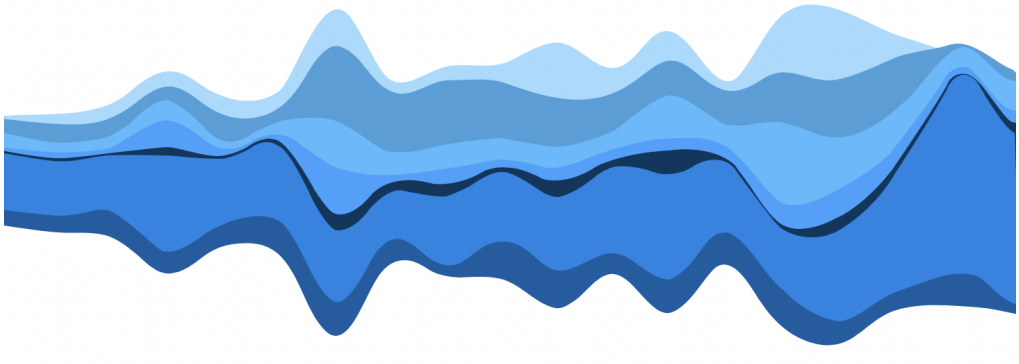


We plan to use group bar graph figure instead of 4a because this gives a better representation to compare first, middle and last positions held by females over the specific year, and it also explains the trend over the years.



We plan to use this mean graph for figure 9 because it gives an idea of the average age of males and females in the core, central, infra, and other committees. We don't care about the different points other than the average.

Layout 3



We plan to use this figure for 1d, 4a instead of a grouped bar chart as discussed in layout2 because this would give a better sense to compare the variations over the years for different categories compared to grouped charts.