

Gender in 30 Years of IEEE Visualization

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

Group Members:

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

Overview and Motivation:

We choose to study this project as it analyzes the trend in gender over the 30 years of IEEE visualization conference. This trend is analyzed taking into consideration many factors like paper publication, career age, percentage in different child conferences, collaboration between the genders. We would try to answer the questions by representing them with appropriate visualizations.

Related work and Questions:

The research papers that we studied in the IEEE VIS mostly had one common factor that is the scarcity of female authors in all the published papers. This encouraged us to pick a paper that analyzes the trend in IEEE VIS conferences over 30 years.

We encountered some primary questions that we will try to answer with the help of visualization; they are as follows:

1. How are male/female gender represented among the VIS community 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. What is the percentage of female authors by position each year?

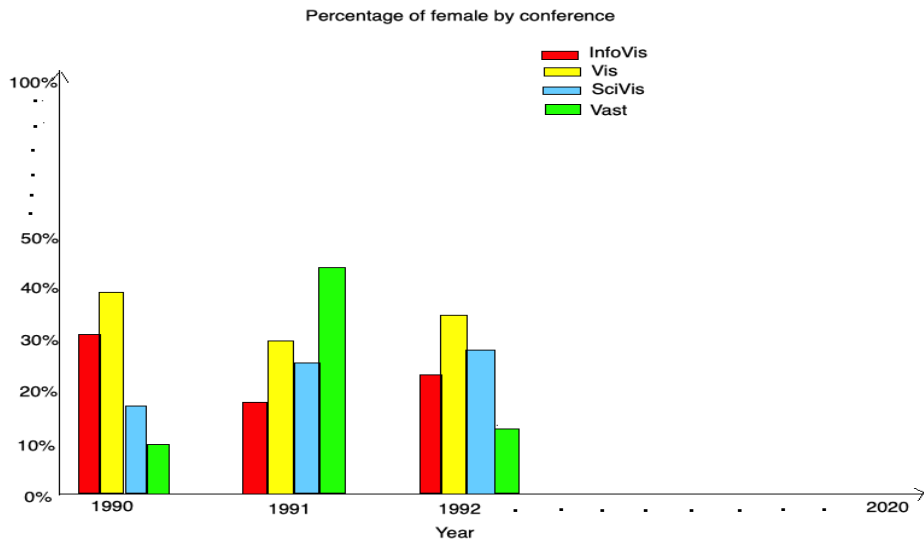
By the end of this project, we would know how the growth of the women have been, how many awards they received, and how many senior positions they hold; it would also help answer how old the women are compared to males in different segments over the decades. We would know the trend in female percentages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

Data Source:

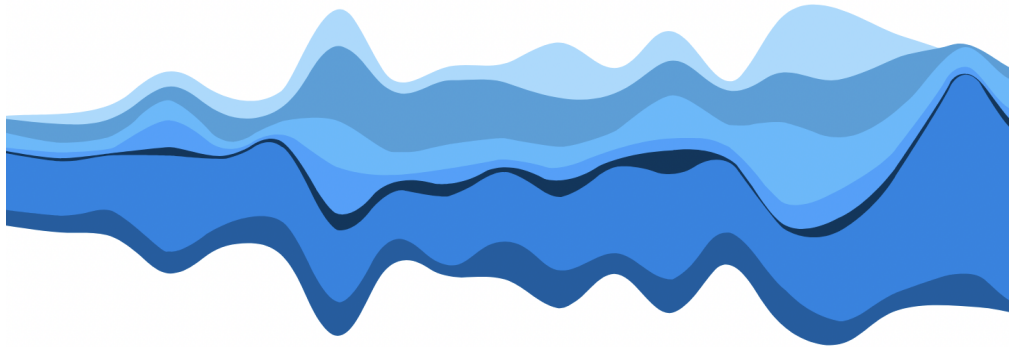
In this project, we would be utilizing the dataset from the website - <https://osf.io/ydfj4/>, and for additional information, we would also refer to this website - <https://nyu.databrary.org/volume/1301>. The dataset contains meta-data such as Year, DOI, Author name, Author gender, Career age, etc. 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. There was no data related to author gender, hence we had to scrap the data to find the authors gender. To represent the data in the bubble chart without crowding it we mined the data and retained the relevant information.

Exploratory Data Analysis:

We planned to use a grouped bar graph to compare the data of different child conferences that make up the IEEE VIS conference. The grouped-bar graph would give a better representation for comparison of data among the four conferences over the years gender wise.



We would use a streamgraph which would give a better sense to compare the variations over the years for different categories compared to grouped charts.



A line chart is better for visualizing the trend only for different categories. In our case the categories were the gender and trend was to show the evolution in number of authors over the years. The other data to display would be the count of authors versus the career age which shows the trend.

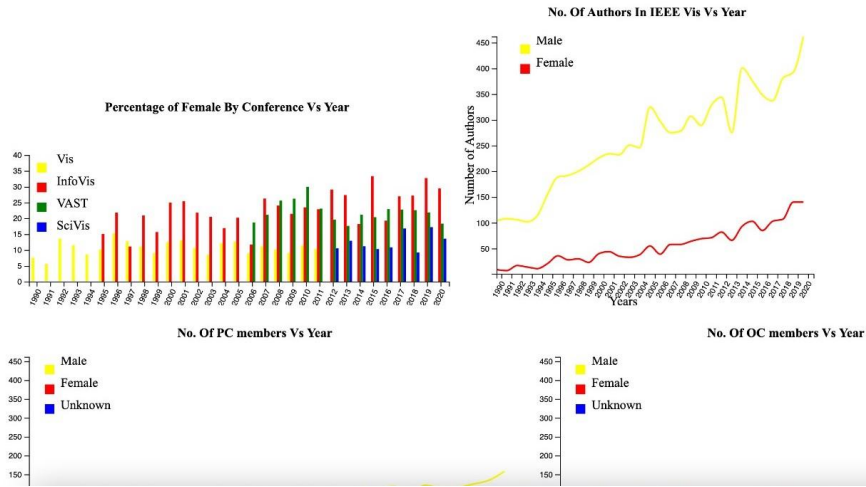
Design Evolution:

After our Initial visualizations, we started considering other visualization designs to see if it suits the data. We included line graph, bubble chart and Heatmap graph along with the initial grouped bar graph. While designing the visualization, we followed the principles of graphical excellence. Our visualizations are able to communicate ideas and data with clarity, precision and efficiency.

The first phase of our design was to show the graphs on different pages which could be navigated by the radio buttons. The categories are distributed by years, age and collaboration.

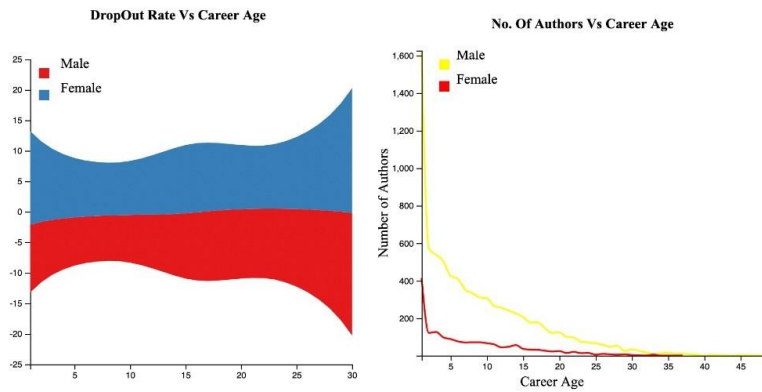
Gender in 30 Years Of IEEE Visualization

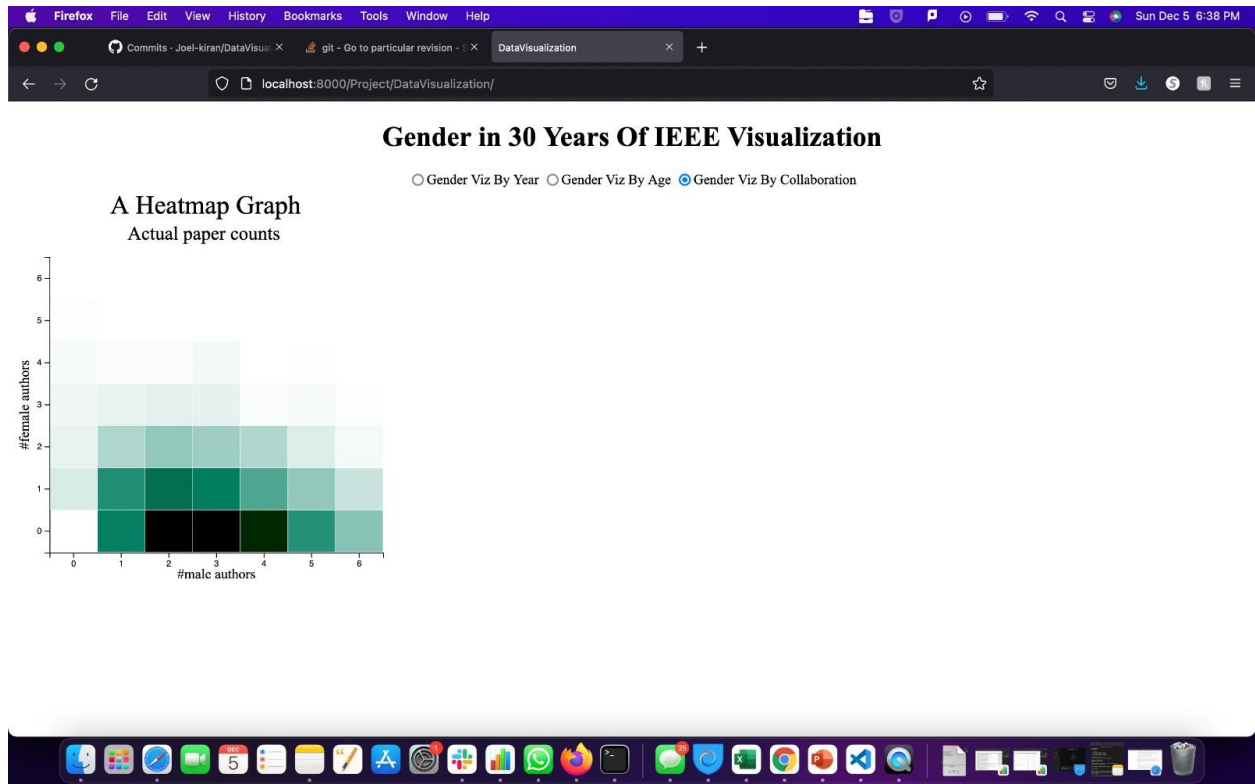
Gender Viz By Year Gender Viz By Age Gender Viz By Collaboration



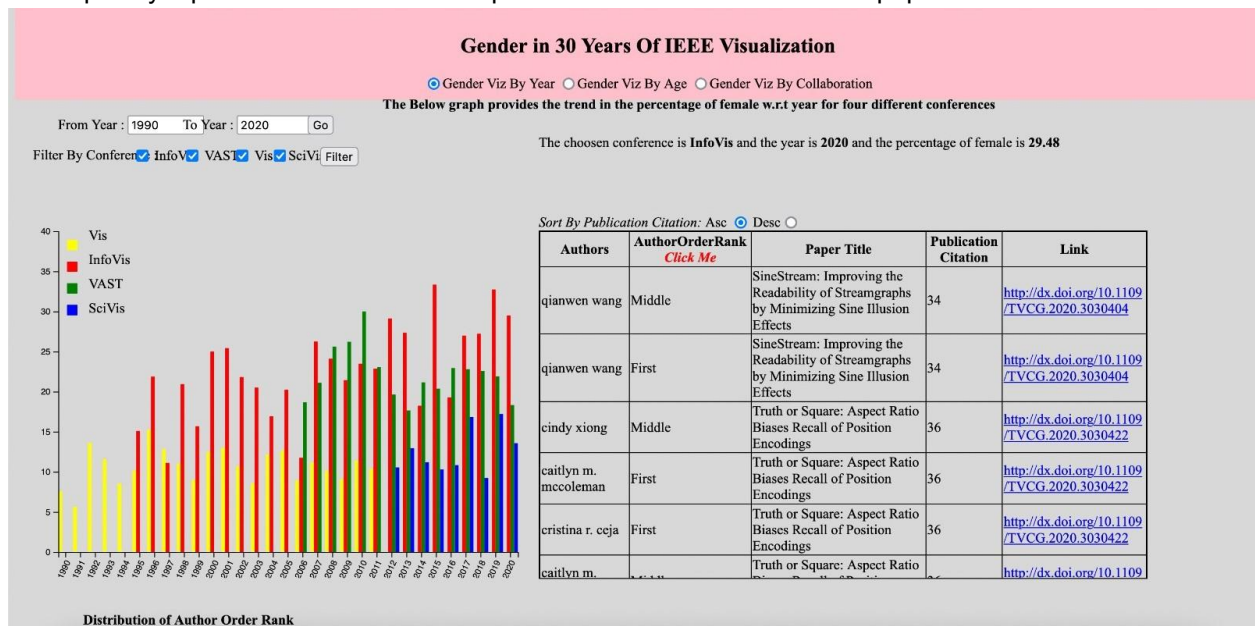
Gender in 30 Years Of IEEE Visualization

Gender Viz By Year Gender Viz By Age Gender Viz By Collaboration

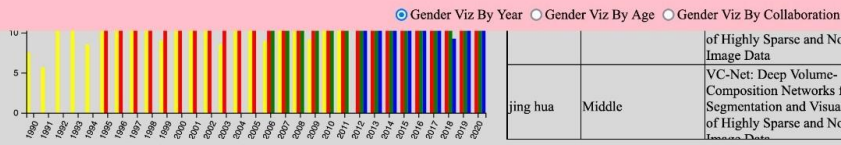




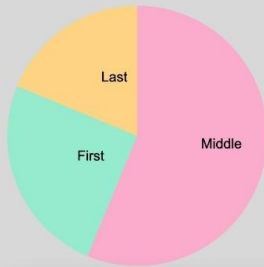
We fixed the background color to grey so that all colors can be seen well on the graph. We added some interactions for the bar graph as filters and hover functions. A table showing the data for all papers and consequently a pie chart would show the position of female authors in the paper.



Gender in 30 Years Of IEEE Visualization



Distribution of Author Order Rank

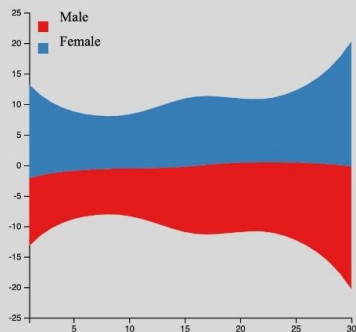


		of Highly Sparse and Noisy Image Data		https://doi.org/10.1109/TVCG.2020.3030374
jing hua	Middle	VC-Net: Deep Volume-Composition Networks for Segmentation and Visualization of Highly Sparse and Noisy Images Data	47	http://dx.doi.org/10.1109/TVCG.2020.3030374

Gender in 30 Years Of IEEE Visualization

Gender Viz By Year Gender Viz By Age Gender Viz By Collaboration

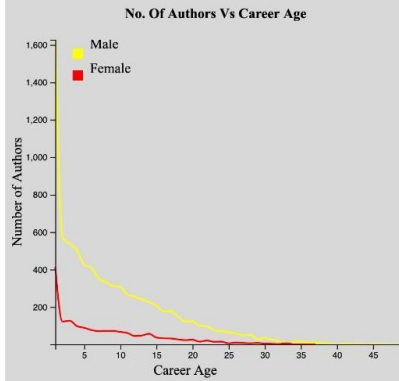
DropOut Rate Vs Career Age



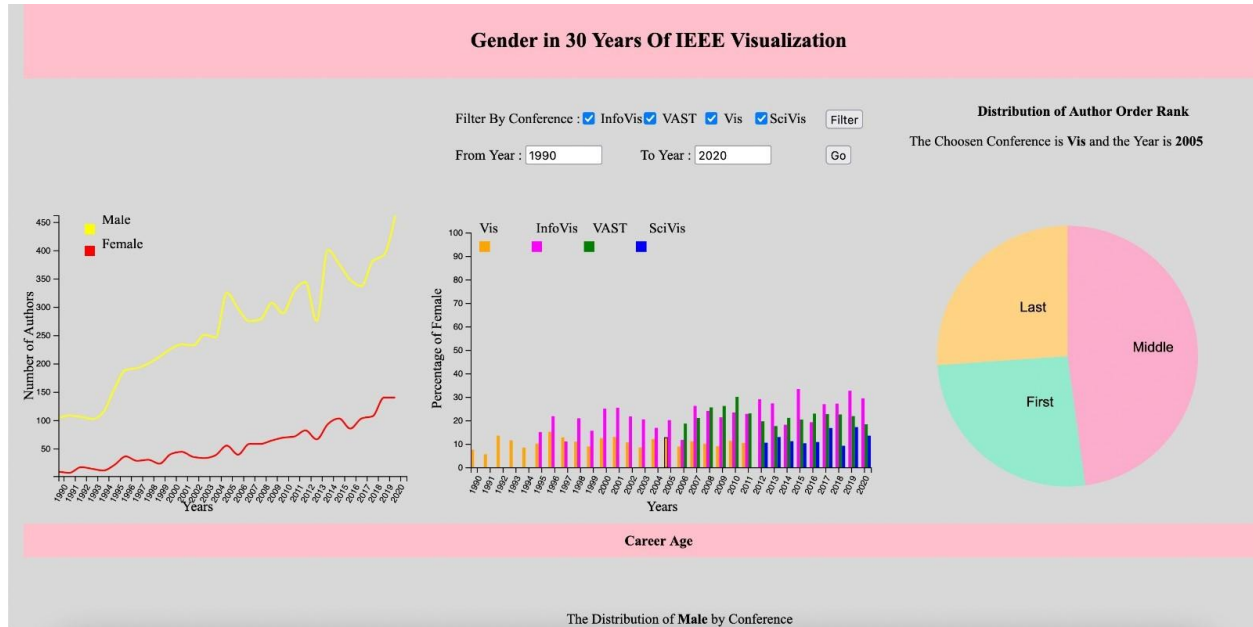
This graph provides an overview of the trend in dropout rate (aka quitting job) with respect to career age for Male and Female authors

The below graphs provides an overview of the trend w.r.t to number of authors and career age, it helps us understand how many authors of a particular career age are present in the organisation. **click on the male/female line in the graph, to explore more details**

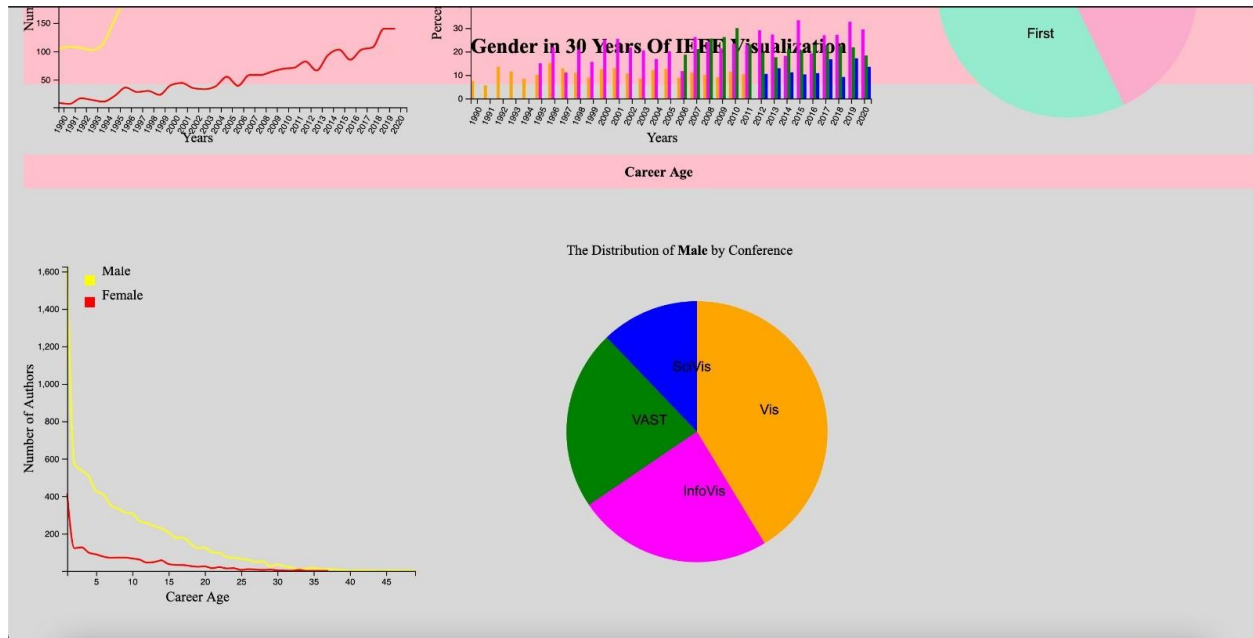
The below graphs provides an overview of the trend w.r.t to number of authors and career age, it helps us understand how many authors of a particular career age are present in the organisation. **click on the male/female line in the graph, to explore more details**



In the next phase, we fixed the colors so that it would be consistent throughout the visualization. We removed the table that showed the data of all the papers and were looking for a different way to visualize it. The major change in this phase was to put all the graphs on the same page so the user does not have to navigate to different pages for different categories of data. In this we removed the stream graph depicted in the previous phase as it wasn't conveying much data required to show the trends as a whole.



We added an interaction in which when we hover over the line graph of the career age it shows which conferences the authors were most active in their career ages.



In the last and the final phase we depicted the data of all the papers in the form of a bubble graph with forces which became the unique feature of our visualization. We also changed the interaction of the bar graph. Now when we hovered over the bars it would display the top cited authors in that year and that conference.

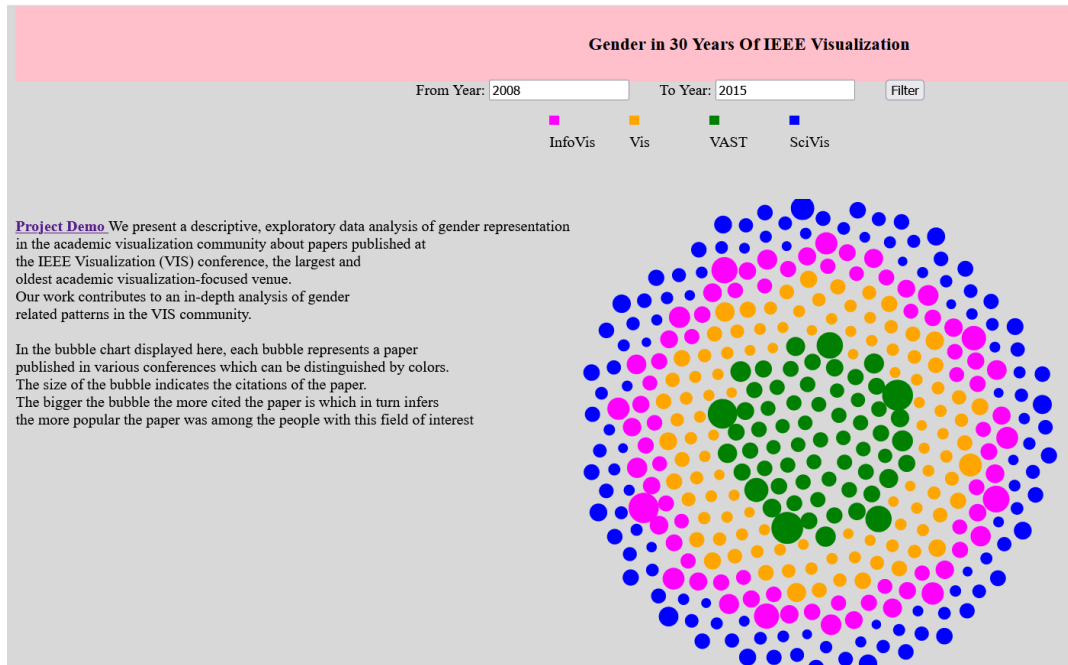
The line chart and the bar chart were now linked to each other and clicking on the line of a particular chart would show the distribution of that gender in the bar chart.

Implementation:

Our visualization was completed and tested mostly on **Firefox**.

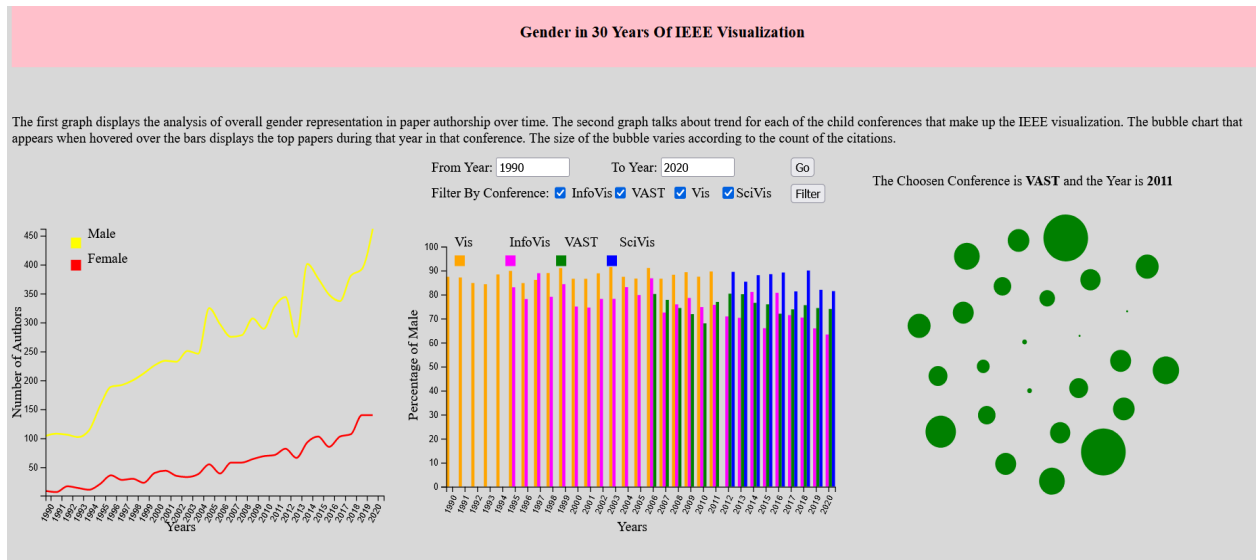
Visualization 1: Bubble graph with forces

- The bubble chart with forces is the most important visualization in the website as it visualizes the main dataset of the project. All the other charts are a derivation of this dataset to analyze the trend.
- Each bubble in the bubble chart represents a paper published in the conferences(InfoVis, Vis, VAST, SciVis) which can be distinguished by colors.
- When we hover over the bubble, we can see the title of the paper and the citation count. The size of the bubble varies according to the citation count.
- The filter above the graph allows the user to filter the data w.r.t years.



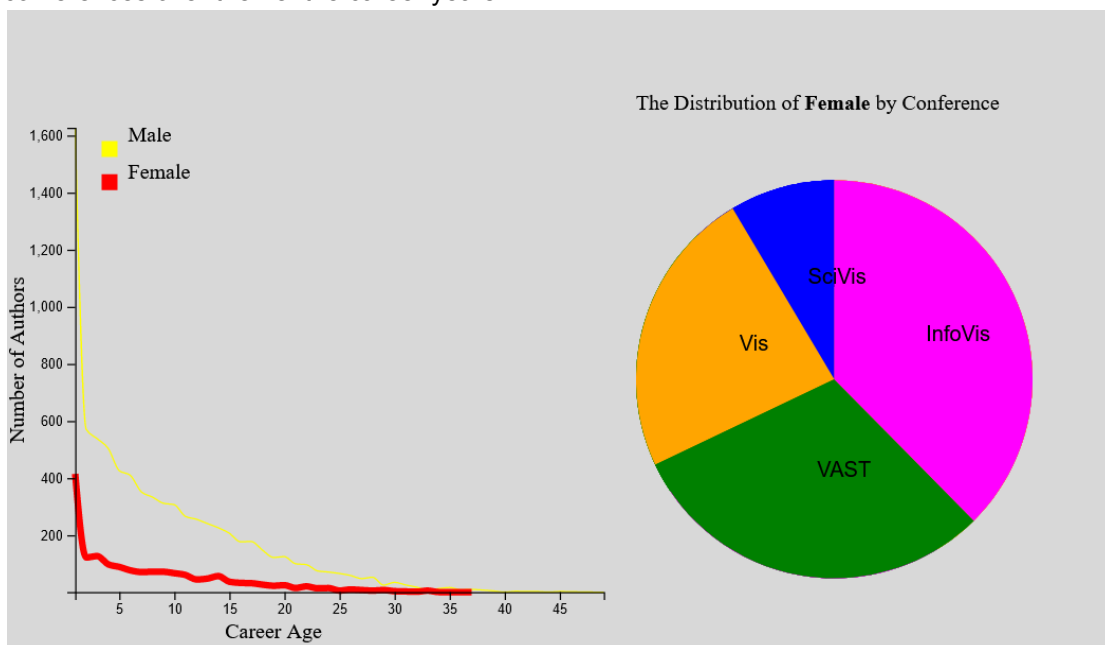
Visualization 2: Line graph & Grouped-bar graph

- The line graph shows the general trend of representation of male and female authors in the conferences over the years. We can see the percentage of female employment has increased and has been growing steadily in the last few years.
- The grouped-bar graph besides the line graph is linked to it. When we click on the male author line or the female author line, the grouped-bar graph changes accordingly. The grouped-bar graph shows the percentage of male and female authors in these child conferences that make up the IEEE visualization conference.
- We have provided filters for the years so that we can choose to see conferences in that particular year range. We also have filters for conferences so that we can focus on one or two conferences' data which we wish to get displayed. For example: If we select only InfoVis and deselect other conferences, then the grouped-bar graph will show us the data of InfoVis conferences only for each gender author.
- When we hover over the bars, the top cited papers that were published in that conference are shown alongside in the form of bubble charts. The higher citation paper has a bigger bubble compared to the paper who has lower citation. From this we can infer which author's paper has been cited and how many times it has been cited.



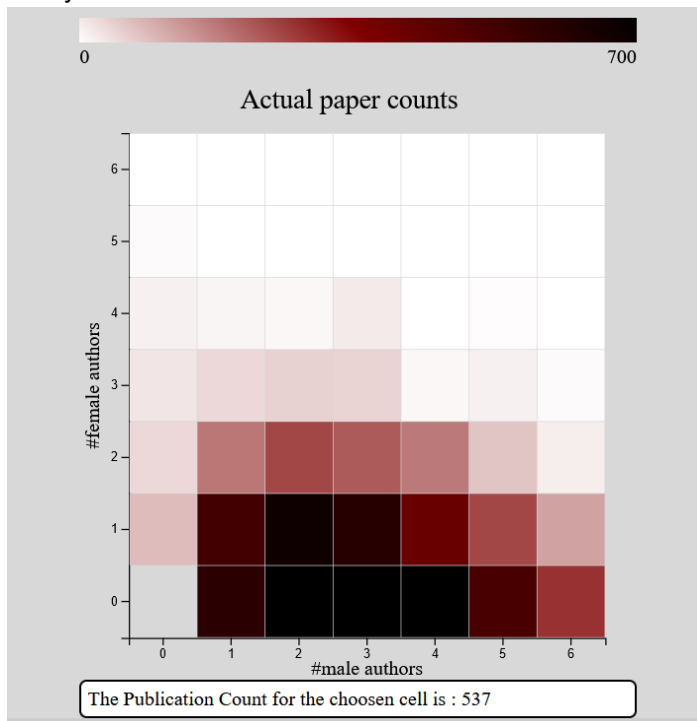
Visualization 3: Line graph

- The graph shows the number of authors versus their career age. Career age means the seniority of the author in the IEEE VIS community.
- The pie chart is associated with the line graph and it displays the distribution of authors by conference over their entire career age.
- The pie chart changes accordingly when we click on the male or female author line in the line graph.
- The intent of this graph and chart is to observe the involvement of each gender in various conferences over their entire career years.



Visualization 4: Heatmap graph

- Graph represents the collaboration between male and female authors for the research paper publication.
- Each cell represents the number of papers published either by authors on their own or published in collaboration with other authors. Initially you won't see the number of papers but when you hover over each cell, a box below the graph will show the publication count.
- One thing we can observe is that publication count between male authors is much higher compared to female authors.
- The intent here is to see if a gender likes to collaborate with the same gender for collaboration work or if they prefer to work with the opposite gender.
- One more thing is that, this particular visualization can be used as reference in near future to encourage and promote more female authors in IEEE conference since as of now the publication done by individual female authors and publication done in collaboration between female authors is way too low.



Evaluation:

By using the visualizations, we realized that the employment of women have increased over the past thirty years and they seem to be occupying the higher positions in the conference. We also inferred that most papers are published when authors collaborate with the same gender.

Even though the number of women in all of the data seems much less than the count of men, there has been an incline in the number of numbers if we look at the early years which seems as a positive result.

We hope this trend continues to grow so that someday the numbers are almost equal in all the fields

Our visualizations work well and conveniently. We were able to answer questions proposed initially. We plan to improve it further once we get more data to discover patterns, without which we can't proceed further.