Olympic Data Visualization

By:

Omkar Chekuri

Olympics

- The Olympics are the leading international sports event with more that 200 countries participating.
- The first Olympics was help in Athens, Greece in the year of 1896.
- Over the 120 years time period, the Olympics has faced a lot of challenges such as sex discrimination , politics etc.

Motivation

- My basic motivation for the project was to see trends in the Olympics data set.
- I picked the Olympics data set as it is an accumulation of all kinds of sports and is interesting for our team.
- In particular, I wanted to see the trends in Olympics over 120 years.

The Data Set

- The Olympic data set was taken from Kaggle.
- There are 15 columns and 271116 rows in our data set giving us information regarding things like Team, Medal, Year, Sex etc.
- Figure 1 shows a sample of our dataset.

```
Observations: 271,116
Variables: 15
$ ID      <int> 1, 2, 3, 4, 5, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 7, 7, 8, 8, 9, 10, 11,…
$ Name    <fct> A Dijiang, A Lamusi, Gunnar Nielsen Aaby, Edgar Lindenau Aabye, Christine Jacoba Aaftink, Chri…
$ Sex     <fct> M, M, M, M, F, F, F, F, F, F, M, M, M, M, M, M, M, M, M, M, M, M, M, M, F, F, M, M, M, M…
$ Age     <int> 24, 23, 24, 34, 21, 21, 25, 25, 27, 27, 31, 31, 31, 31, 33, 33, 33, 33, 31, 31, 31, 31, 33, 33…
$ Height  <int> 180, 170, NA, NA, 185, 185, 185, 185, 185, 185, 188, 188, 188, 188, 188, 188, 188, 188, 183, 1…
$ Weight  <dbl> 80.0, 60.0, NA, NA, 82.0, 82.0, 82.0, 82.0, 82.0, 82.0, 75.0, 75.0, 75.0, 75.0, 75.0, 75.0, 75…
$ Team    <fct> China, China, Denmark, Denmark/Sweden, Netherlands, Netherlands, Netherlands, Netherlands, Net…
$ NOC     <fct> CHN, CHN, DEN, DEN, NED, NED, NED, NED, NED, NED, USA, USA, USA, USA, USA, USA, USA, USA, USA,…
$ Games   <fct> 1992 Summer, 2012 Summer, 1920 Summer, 1900 Summer, 1988 Winter, 1988 Winter, 1992 Winter, 199…
$ Year    <int> 1992, 2012, 1920, 1900, 1988, 1988, 1992, 1992, 1994, 1994, 1992, 1992, 1992, 1992, 1994, 1994…
$ Season  <fct> Summer, Summer, Summer, Summer, Winter, Winter, Winter, Winter, Winter, Winter, Winter, Winter…
$ City    <fct> Barcelona, London, Antwerpen, Paris, Calgary, Calgary, Albertville, Albertville, Lillehammer, …
$ Sport   <fct> Basketball, Judo, Football, Tug-Of-War, Speed Skating, Speed Skating, Speed Skating, Speed Ska…
$ Event   <fct> "Basketball Men's Basketball", "Judo Men's Extra-Lightweight", "Football Men's Football", "Tug…
$ Medal   <fct> NA, NA, NA, Gold, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
```

Figure 1: Sample of the original dataset

Questions I would like to be answered…

The questions I would like to answer through our visualizations are:

Q1) What are the changes in the trends of each country winning the Olympics over the years?

Q2) Is weight and age effecting a person's sport?

Q3) Is there a change in the trend of men and women winning over the years?

Q4) What is the change in the trend of winning over time?

Visualizations we chose..

Four visualizations were selected to answer our four questions:

1) Bar Chart for Q4.
2) Map for Q1.
3) Line Graph for Q3.
4) Heatmap for Q2.

System design

- Initial design of the system includes 4 views:
- The first view is bar chart to visualize Number of medals vs year.
- Second view is map view to visualize the number of medals won by different countries over the years.
- Third view is a heat map view to visualize the average age or weight of a person who won different medals.
- Fourth view is a line chart to visualize trends in participation of men and women in Olympics over the years.
- Figure 2 shown an initial system design drawn on an ipad.

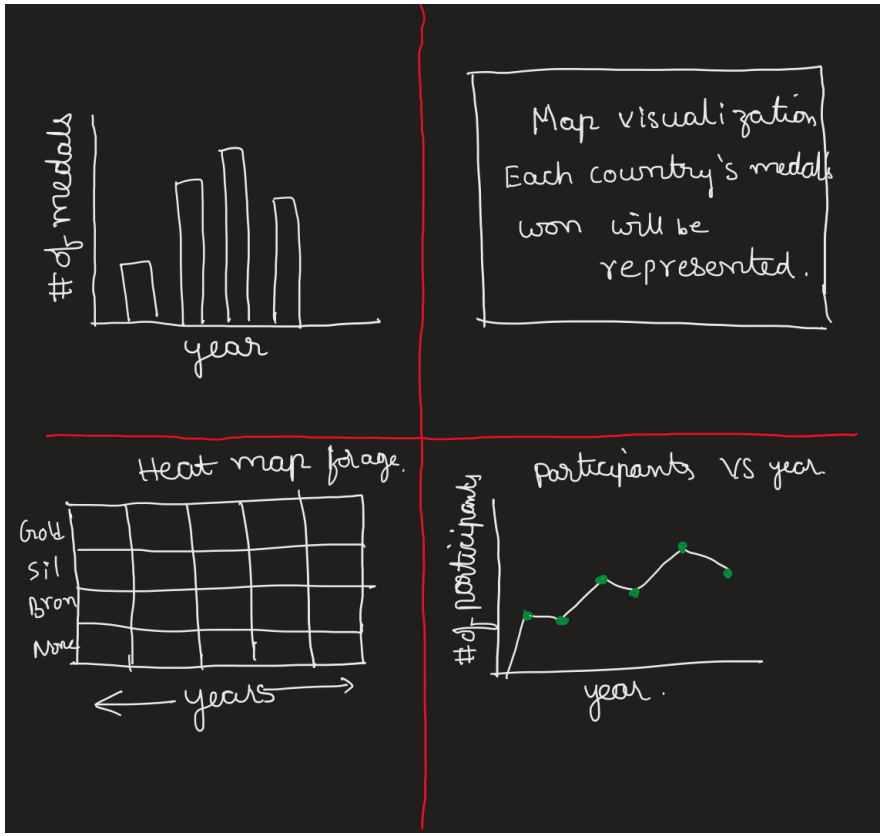


Figure 2: Initial system design

Tools we used…

The following tools were used in the development of the Olympic visualization:

1) R Studio: We used R Studio for data processing and generation of the required structured data.

2) D3js: We used D3 for the development of the visualizations.

3) React JS: We used React JS for the purpose of achieving coordination and interaction between our visualizations.

Phases of development of the system…

- Phase 1: D3js was used for basic visualizations like bar chart.
- Phase 2: R scripts were used to generate the required data files suitable for visualizations.
- Phase 3: Visualizations were developed using D3js and used React JS to introduce the concept of coordination and interaction between these views.

Phase 1

Development:

- In the first phase D3js was used for making simple bar chart visualization.
- The visualization is as shown in the figure 3.

Challenge using d3:

It was difficult to query the data using d3.js as the queries were complex and we were not getting expected results. Therefore, we moved to phase 2 of development.

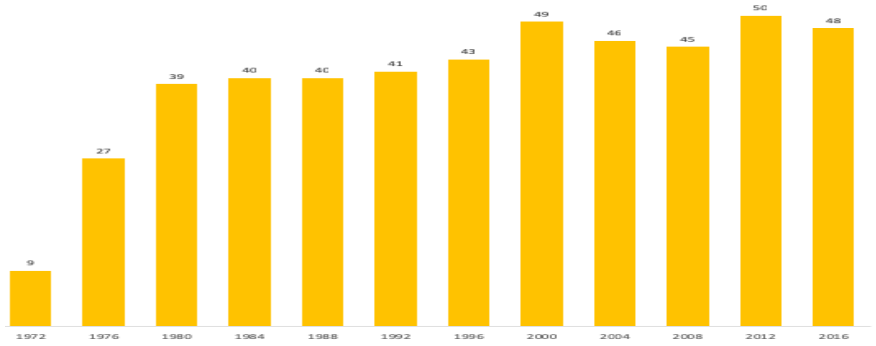


Figure 3: Basic bar chart in phase 1

Phase 2

Development:

- To overcome the challenges faced in Phase 1 , we used R Studio to query data.
- By making use of inbuilt packages in R to obtain the required structure for the visualizations.
- For missing data in map, some coordinates were missing. So, we had to manually add the coordinated.

Challenge faced in phase 2:

It was difficult for us to obtain coordination between views in d3js. Therefore, I proceeded to phase 3 of our project.

Phase 3

Development:

- To achieve interaction between data sets, I used a combination of React and D3 for our visualizations.
- In this phase, I took all the data files and built the visualizations using D3js at the beginning of phase 3.
- I then used React to achieve interaction and coordinated views.

Challenges:

- Due to time contraints, we limited the possible set of interactions that could be performed on the visualization system.

Visualizations – Bar Chart

- The bar chart shows us the variation of the number of medals won in each year.
- Our X axis shows the years and the Y axis has the count of the medals.

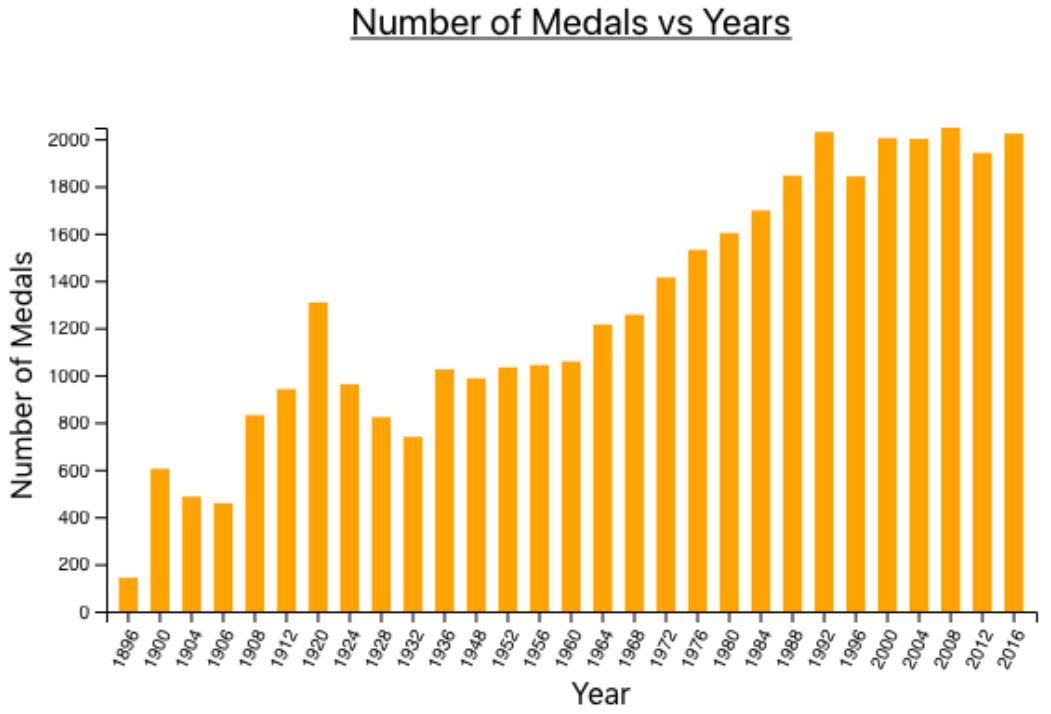


Figure 4: Bar chart visualization

Visualizations - Map

- In this visualization, by using encodings regarding color and size information is provided such as count of medals by each country is shown using red and the size to define the count.
- The data on the map changes interactively as each bar for every year is clicked and the corresponding data is shown on the map.
- The countries are shown in the map by using their general location on the map.

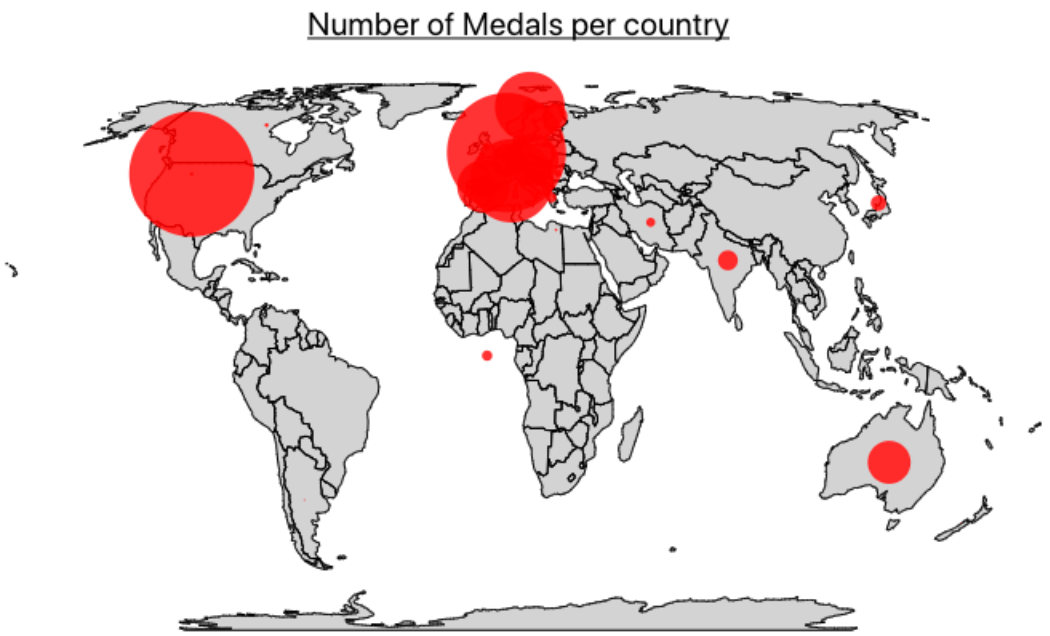


Figure 5: Map visualization

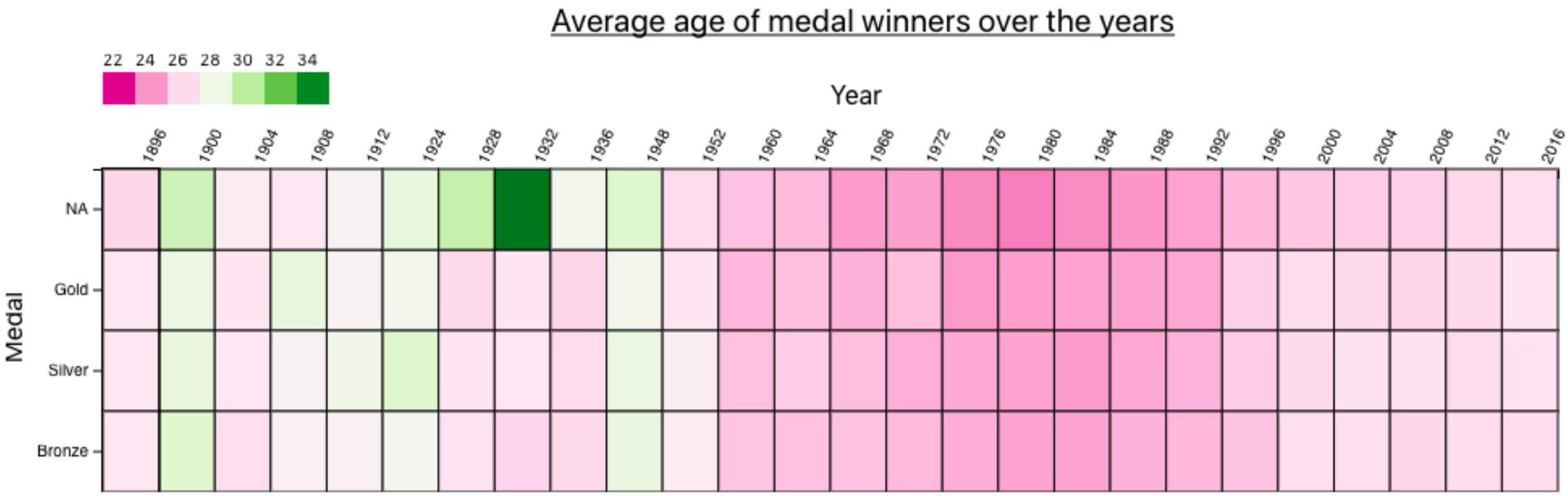


Figure 6: Heat map visualization

Visualizations - Heatmap

- This visualization gives us an idea of the variation in average age and average weight for every year.
- We used two different types of color encodings to specify average weight and average age.
- Interaction in the form of a drop down is also provided to select either the age or weight.

Visualizations – Line Graph

- In this visualization, a line graph is drawn to understand the variation between count of medals won by women and men every year.
- A drop down is provided to pick either women or men and the line chart is generated for that data.

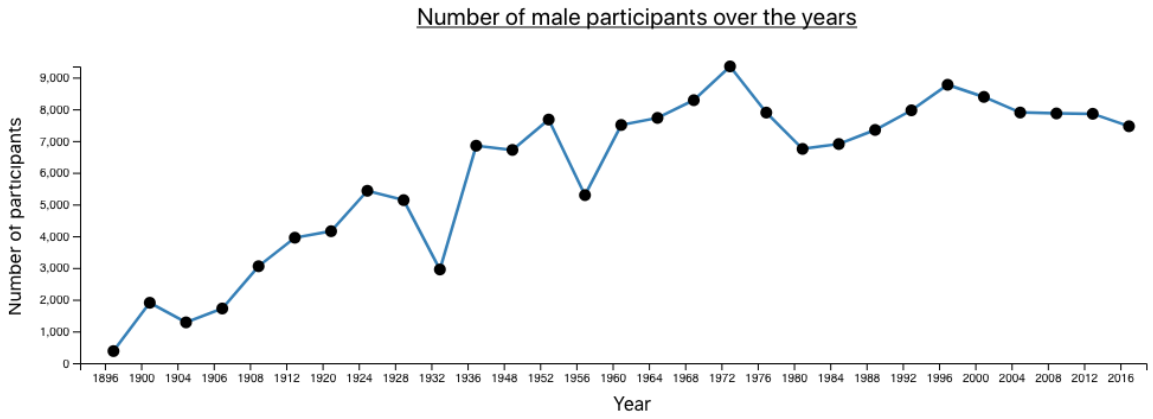
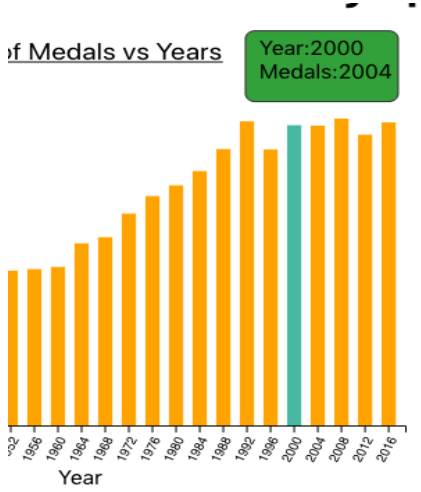


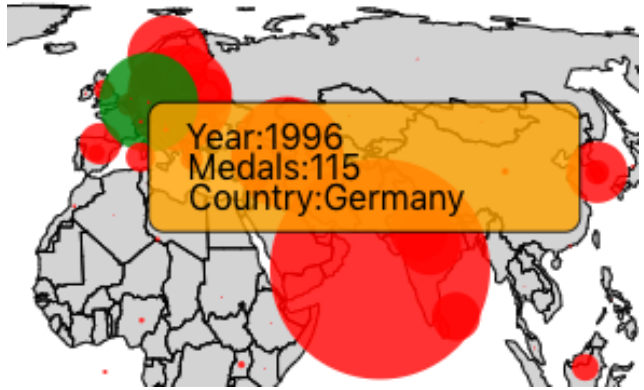
Figure 7: Line chart

Tooltip for details-on-demand

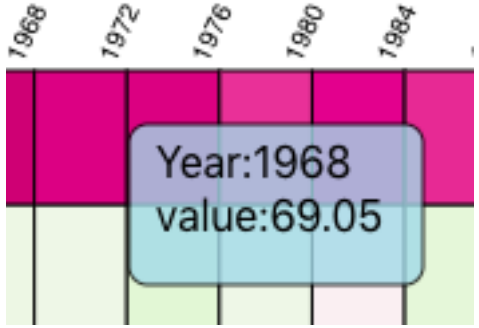
- A tooltip is added for all the visualizations.
- Tooltips for different visualizations are as shown in figure 8.



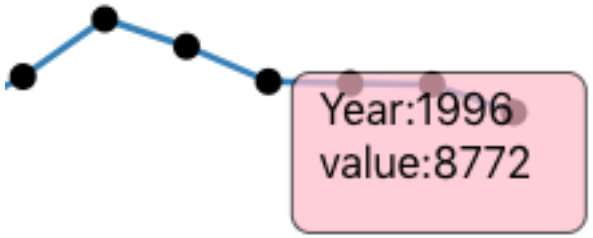
Tooltip for bar chart



Tooltip for map visualization



Tooltip for Heat map



Tooltip for Line chart

Figure 8: Tooltips

Selection Menu

- Selection menu was implemented to choose between different option for a Heat map and Line chart.
- The selection box can be seen in figure 9.

Figure 9: Options for the user.

Usability Study

- A usability study was conducted by taking 5 participants- 2 from Computer Science background, 1 from Electronics and Communications background, 1 from Industrial background and 2 from Data science background.
- The visualization was presented in front of them and they were asked to rate the visualization in terms of understandability, ease of use , usability and aesthetically on a scale of 1 to 5.
- A few simple tasks were also given to the users and in specific the usability was rated based on how well the users were able to perform these tasks.

Results

- Participants from CS and DSA background were able to navigate through the system and get the results easily.
- Participant from Industrial background needed help to understand the system.
- We noticed that participants with DSA background was able to perform the tasks easily.
- Most of the users used the tooltip for the Heatmap instead of looking at the legend to see the values.

Future Work

- This project can be extended in the future to add more interaction for the user like zoom onto a visualization.
- We can give more options to the user for heatmap to view event wise average age of winners.
- We can add more visualizations to see other trends in the data.

Thank you!