Data Visualization, Spring 2024

CFLens: See Beyond the Code

Priet Ukani (20221111039), Kushal Shah (2022101059), and Garvit Gupta (2022101113)

April 23, 2024

1 Brief Description of Data

Our Dataset is the Codeforces data which is publically available through Codeforces API and various other online sources. The Codeforces API provides a large number of attributes in the data. Currently the data which we are visualising consists of the following things, while more data will be utilisied based on the further plots.

Ratings of Users at different time intervals All user ratings are included in this data at certain predetermined intervals (one data point per month in this example). Each data point consists of Username, date, rating value, and the user's country. The data is portrayed using a bar chart race, wherein the top coders' ratings are displayed as they vary over time.

Number of Problems by Tags The number of problems under each tag on Codeforces is the data we consider here, each data entry has two values Tag Name and Number of Problems, We plot a simple donut-chart to visualise which Problem Tags (Types of Problems) occur more frequently.

Rating Statistics of various Organisations from Different Countries This data includes Max Rating, Average Rating, Number of user of a Organisation from various countries across the globe. This data is visualised with the help of the Choropleth Map where a user can click on a specific country to view that country's top 10 Organisations Average and Max Rating.

Country Rating Statistics The overall average and maximum ratings for every nation are included in this data. A visualisation of this is provided by a choropleth map.

Time distribution of solves in a Contest The time of solve of a particular question for different users in a contest is included in this data. This data is visualised using a radar chart. Here the time spent on that problem by a user is shown in percentages of the total time of the contest. This is done for multiple users and contest which are taken input from the user.

Relative Ratings of Friends in the period of a Contest The relative ratings of friends in a contest are included in this data. This data is visualised using a bump chart. Here the relative rating of friends is shown over the period of the contest with intervals of 15 minutes. This is done for multiple users and a particular contest.

Daily submission data of a user The daily submission data of a user is included in this data. This data is visualised using a Heatmap of a calender. Here the user can see the number of submissions made on a particular day of the year. This is done for a single user. Hovering on the day will show the number of submissions made on that day along with the verdict distribution.

Contest Submission Statistics The submission statistics of a contest are included in this data. This data is visualised using a funnel chart which shows the relation between number of users who solved 1 problem, 2 problems, 3 problems and so on. This is done for a single contest.

Submission Language Data The submission language data of a user is included in this data. This data is visualised using a radial bar chart. Here the user can see the number of submissions made in a particular language along with the verdicts of the submissions. This can give an idea of the language used by the user for submissions.

2 Most Important Parts of Data which needs to be Visualised

- The time stamp of the data and the users' current rating values at that particular moment are the most significant aspects of the visualised data for the first visualisation, since it's most exciting and inspiring for spectators to look at leaderboards for top coders, the top 10 coders at that time are shown.
- The Number of Problems for a Tag is essential for the second visualisation. Additionally, since for the tags with very low number of problems, such questions are not relevant as they don't occur frequently, so the first 10-15 rows of this data are important.
- The Max Rating and Average Rating of a Country/Organisation are the most crucial data points for the thrid visualisation.

3 Visualisations

Currently we have worked upon the following visualisations:

3.1 Race Bar Chart

- This chart used a horizontal bar chart to show the competition between the topmost coders on codeforces and their rating changes over the years. The horizontal axis represents the rating and different bars represent users. Right from the start Tourist from Belarus has been at the top, while Indians have struggled to get into the top standings despite being in large numbers on codeforces today.
- Since we wanted to show the standings change live right from the start of the Codeforces uptil today, we chose a bar chart race as the most suitable chart which keeps on changing as time changes. The graph is shown in colour to make it aesthetic and eye pleasing as shown in Figure 1. Also the colour does not signify anything, so for a person who can not distinguish between two shades can also easily visualise the data as shown in Figure 2.

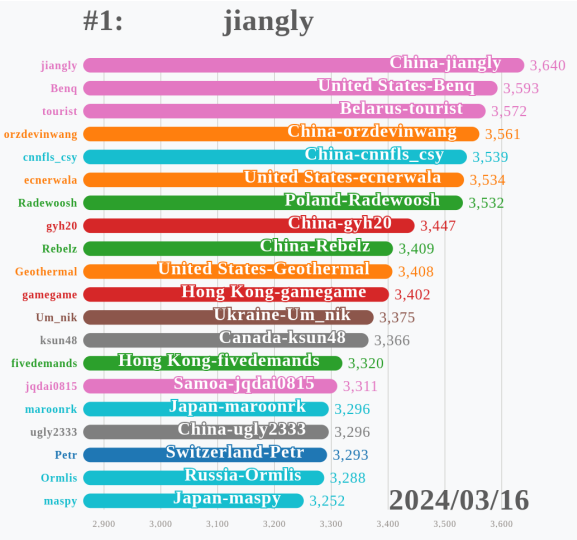

Figure 1: Bar Chart Race Visualisation Snapshot in Colour

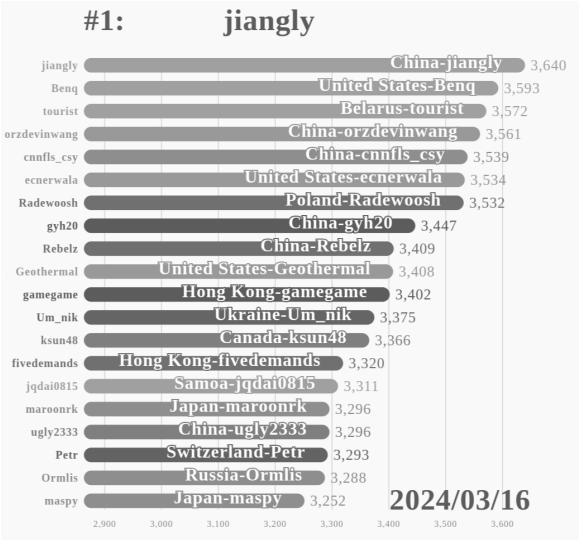

Figure 2: Bar Chart Race Visualisation Snapshot in Black and White

3.2 Donut Chart

- This chart shows the distribution of problems across tags on Codeforces. This can be useful to the users to find out which tags to practice more based on the occurance in problems. The graph is made interactive with tooltips showing the number of problems of each tag with the tagname.
- Two Buttons below enable the donut chart to either show all tags or only show the top 10 tags which enable better viewing.

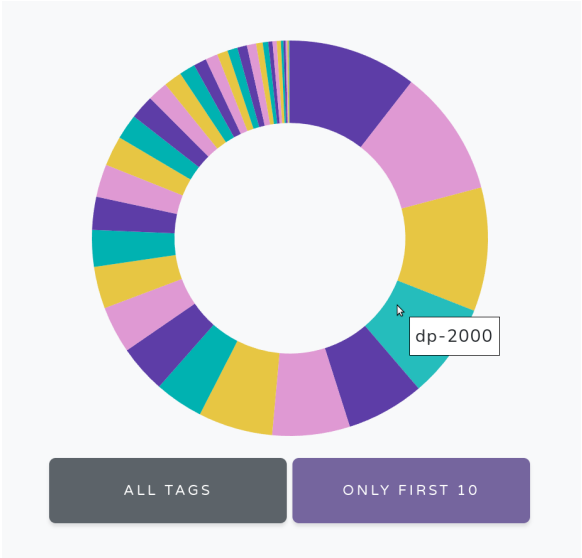

Figure 3: Donut Chart with All Tags

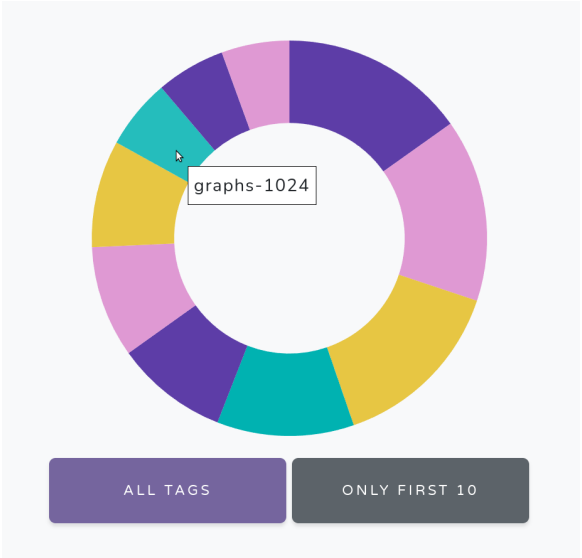

Figure 4: Donut Chart with only top 10 Tags

3.3 Choropleth Map

- The data used for the construction of World Map visualisation is extracted from codeforces website using it's public API and web scraping. All the personal data including the country, institution, rating, username is collected and visualised using the following plot.
- By the use of world map we see the distribution of the sport of the competitive programming on Codeforces with useful metrics like average rating and maximum ratings of countries. This gives a brief about the enthusiasm and spread of competitive programming culture over the world. The world map is made interactive with tooltips on hover to show the values. Also there is a dropdown to select the datatype to show whether average or maximum rating. We have shown the density with the help of sequential shades of green.
- With the help of this one can view and compare various top institutions of a particular country. This can be extremely useful for new college joining students to judge the competitive coding cultures of various top colleges.

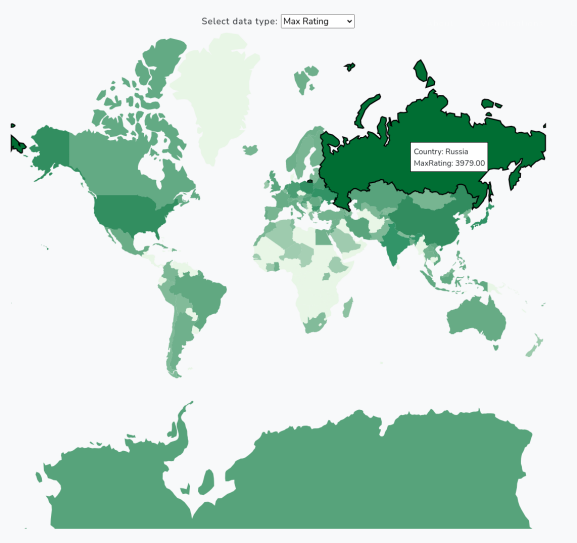

Figure 5: World Map with hover effects of Maximum Rating

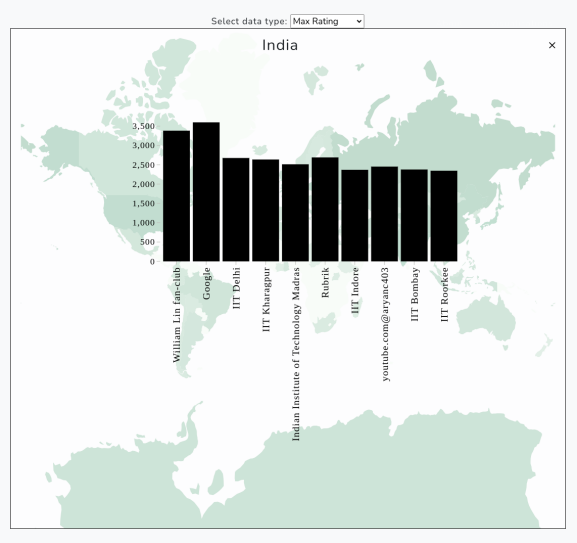

Figure 6: Pop out of Indian institutions on clicking on India

3.4 Line Chart

- User enters 2 users to compare and the line chart shows the ratings of the 2 users over time. By default 2 users are loaded without user input.
- The line chart is interactive and the user can hover over the points to see the exact rating at that time. Legends are provided to show which line represents which user.

3.4.1 Why our visualisation does what it is supposed to do

The aim was to provide comparative analysis over time for 2 users ratings. The best visualisation to do this is the Line chart. The line chart is perfect in showing progressive data where easy comparision can be done between the lines.

3.4.2 How it works

- User enters 2 users to compare and the line chart shows the ratings of the 2 users over time. By default 2 users are loaded without user input.
- The data is live fetched from the Codeforces API and the line chart is updated with the new data.
- The x-axis represents the time and the y-axis represents the rating. The line chart is interactive and the user can hover over the points to see the exact rating at that time. Legends are provided to show which line represents which user.

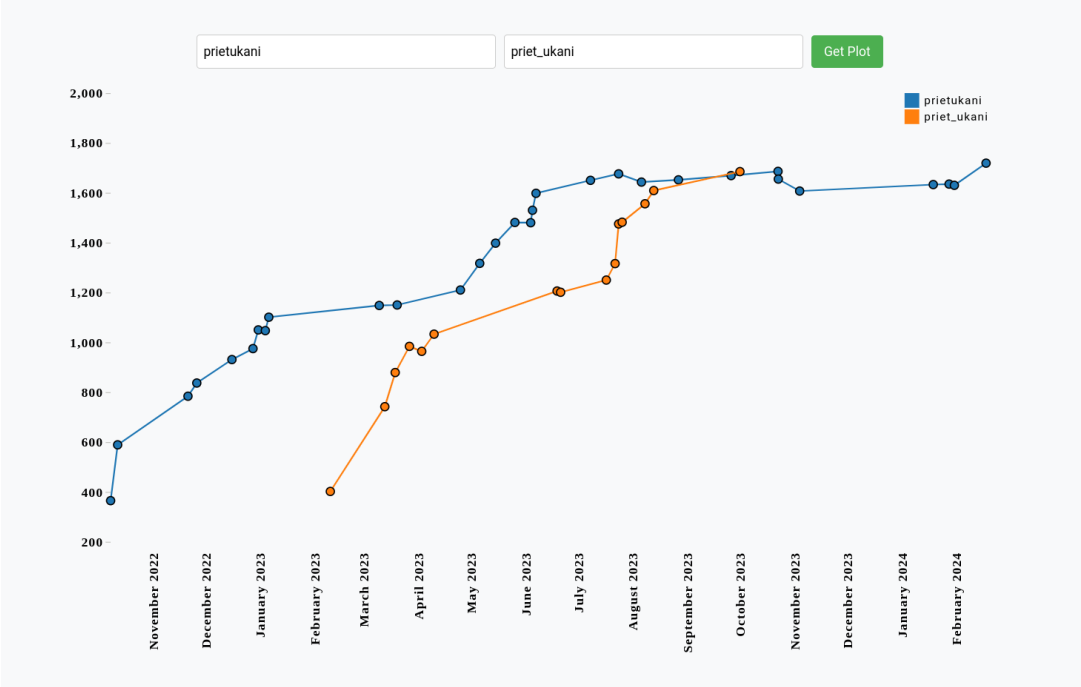

Figure 7: Line Chart showing comparision of 2 users ratings over time

3.5 Radar Chart

3.5.1 Why our visualisation does what it is supposed to do

The aim was to show the time distribution of solves in a contest. The best visualisation to do this is the Radar chart. The radar chart can provide a comparision of the time spent on each problem by a user in a contest.

3.5.2 How it works

- User enters the contest id and the usernames of the users to compare. The radar chart shows the time distribution of the solves of the users in the contest.
- The radar chart is shown in a circular form where the time spent on each problem is shown in percentages of the total time of the contest.
- The data is live fetched from the Codeforces API and the radar chart is updated with the new data.

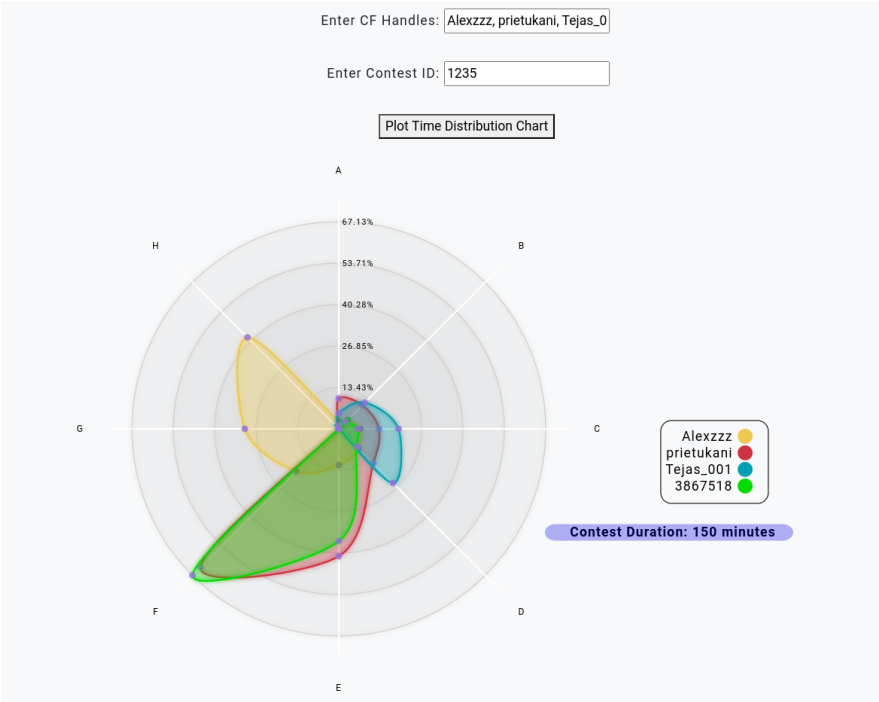

Figure 8: Radar Chart showing time distribution of solves in a contest

- The radar chart is interactive and the user can hover over the points to see the exact percentage of time spent on that problem. Legends are provided to show which line represents which user. On Hovering the part of that user is highlighted to make the display more interactive.

3.6 Bump Chart

3.6.1 Why our visualisation does what it is supposed to do

The aim was to show the relative ratings of friends in a contest. The best visualisation to do this is the Bump chart. The bump chart can provide the change of rankings after some fixed time and shows the progress of the users in the contest.

3.6.2 How it works

- The bump chart shows the data of 9 users in a contest. Here the data is fixed for the contest and is fetched from the Codeforces API.
- The bump chart is shown in a linear form where the relative rating of the users is shown over the period of the contest with intervals of 15 minutes.
- The bump chart is interactive and the user can hover over the points to see the exact relative rating of the user at that time. Legends are provided to show which line represents which user. On Hovering the part of that user is highlighted to make the display more interactive.

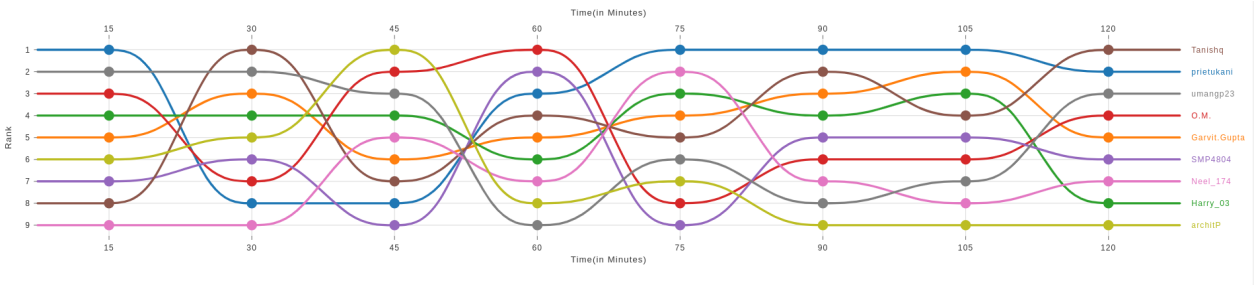

Figure 9: Bump Chart showing relative ratings of friends in a contest

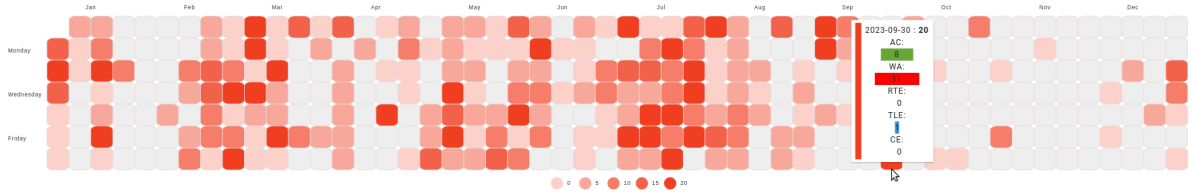

Figure 10: Heatmap showing daily submission data of a user

3.7 Heatmap

3.7.1 Why our visualisation does what it is supposed to do

The aim was to show the daily submission data of a user. The best visualisation to do this is the Heatmap. The heatmap can provide the number of submissions made on a particular day of the year.

3.7.2 How it works

- The heatmap shows the data of a single user. Here the data is fixed for the user and is fetched from the Codeforces API.
- The heatmap is shown in a calender form where the user can see the number of submissions made on a particular day of the year.
- The heatmap is interactive and the user can hover over the days to see the number of submissions made on that day along with the verdict distribution.
- The legend is shown to show the color mapping of the number of submissions.
- Tooltip also shows a small bar chart of the verdict distribution on hovering over a day.

3.8 Funnel Chart

3.8.1 Why our visualisation does what it is supposed to do

The aim was to show the submission statistics of a contest. The best visualisation to do this is the Funnel chart. The funnel chart can provide the relation between the number of users who solved 1 problem, 2 problems, 3 problems and so on. This looks like a funnel as the number of users decreases as the number of problems solved increases.

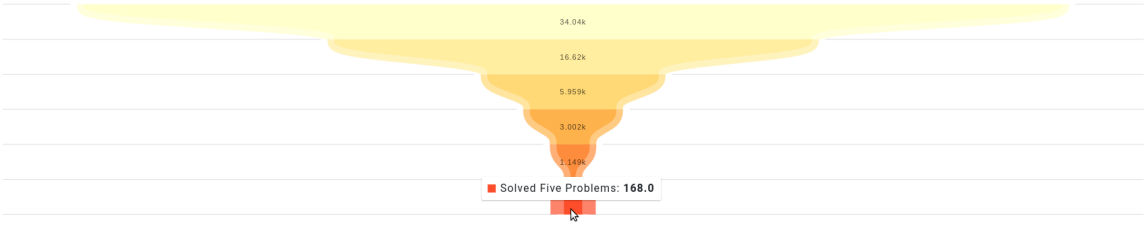

Figure 11: Funnel Chart showing submission statistics of a contest

3.8.2 How it works

- The funnel chart shows the data of a single contest. Here the data is fixed for the contest and is fetched from the Codeforces API.
- The funnel chart is shown in a funnel form where the relation between the number of users who solved 1 problem, 2 problems, 3 problems and so on is shown.
- The funnel chart is interactive and the user can hover over the parts to see the number of users who solved that many problems.
- Color gradient is used to show progressive more number of solves.

3.9 Radial Bar Chart

3.9.1 Why our visualisation does what it is supposed to do

The aim was to show the submission language data of a user. The best visualisation to do this is the Radial Bar chart. The radial bar chart can provide the number of submissions made in a particular language along with the subparts of the verdicts of the submissions. Stacked bar chart can be used to show the subparts of the verdicts. But the radial bar chart is more aesthetic and can show the data in a visulaisable form.

3.9.2 How it works

- The radial bar chart shows the data of all submissions of a particular contest. Here the data is fixed for the user and is fetched from the Codeforces API.
- The radial bar chart is shown in a circular form where each bar shows the language and the subparts shows the verdicts of the submissions.
- The radial bar chart is interactive and the user can hover over the parts to see the tooltip which shows number of submissions made in that language with some particular verdict.
- The legend is shown to show the color mapping of the verdicts.

4 Why our visualisation does what it is supposed to do

4.1 Choropleth Map

- The aim was to show the popularitity and distribution of codeforces over the world. The best visualisation to do this is the Choropleth map. Also an added functionality of popout which shows the comparision of various organisations of that country is best fit to be shown with the help of bar graph.

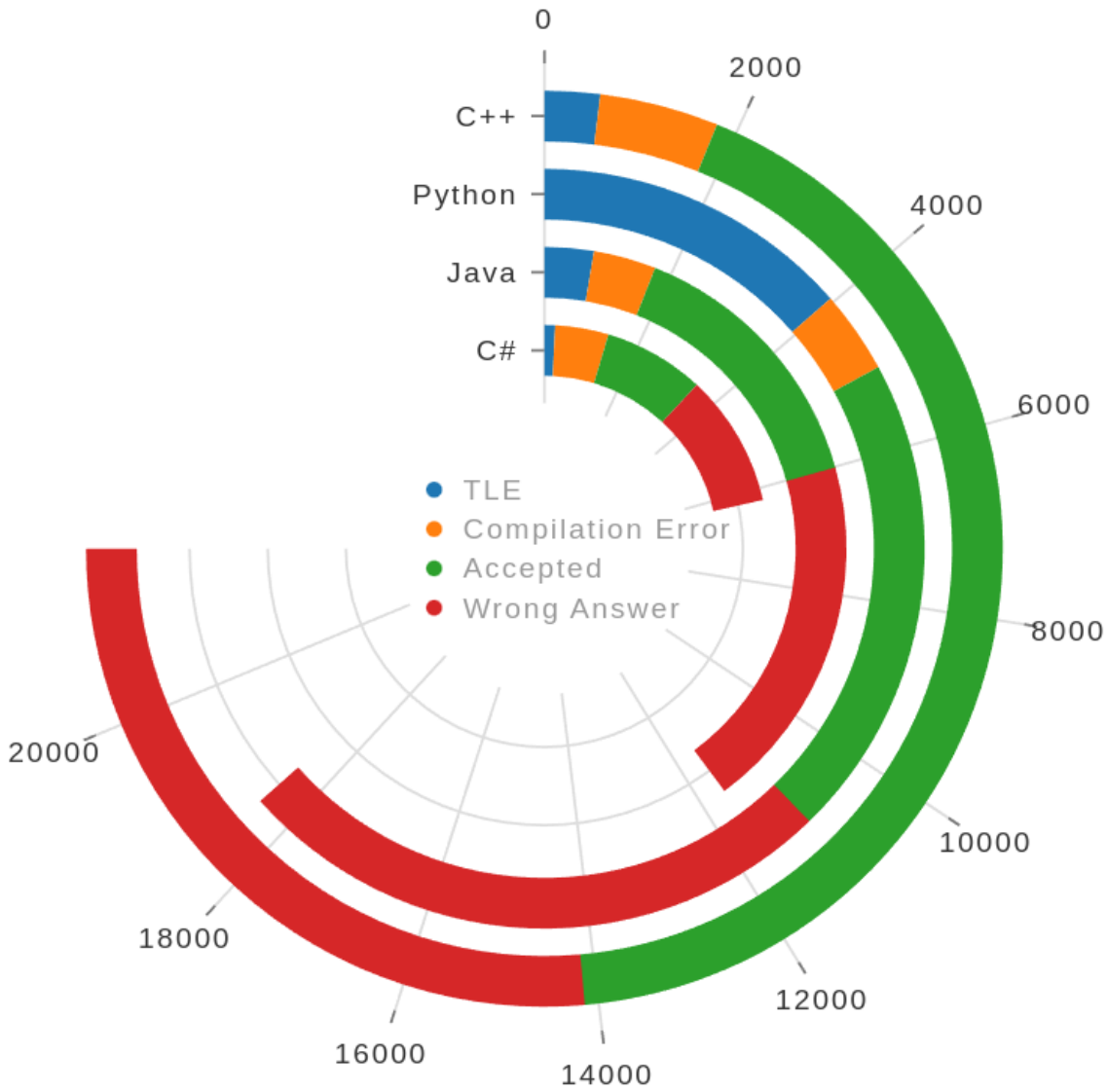

Figure 12: Radial Bar Chart showing submission language data of a user

- By hovering over a country and adjusting the filter between Average and Max rating, users can discern the popularity and expertise levels of users from that country.
- Analyzing the statistics of top organizations within a country allows us to gauge the level of interest and popularity associated with that country.

4.2 Donut Chart

- The aim was to show the comparative analysis of number of occurance of tags where the user can find out which tag occurs more than others. Donut chart best fits this use case.
- Users can observe the most important tags by number of problems and can focus on practicing them.

4.3 Bar Chart Race

- The aim of visualizing a bar chart race plot showcasing the top coders over the years on platforms like Codeforces is to inspire and motivate newbie coders by providing them with a visual representation of the progression, competitiveness, trends, and sense of community within the coding community.
- It serves as a medium for spectators to view the history of leaderboards, it represents players competing and their ups and downs, it also provides a sense of competition for experienced coders observe the rise and falls of the best and to aim to compete with them.