Data – Visualisation

Project Final Submission

Team Number – 1

Team Name – Pattern Pros

Team Members – Aryan Bansal (2021111018)

Chirag Jain (2021101100)

Pratham Priyank Thakkar (2021101077)

- Brief introduction of the project:
    - As a part of the data visualisation course, we have made a Codeforces Statistics Visualizer. Codeforces is a competitive programming platform.
    - In this project we have made a website that aims to help members of the Codeforces community to understand the Codeforces platform and the sport of competitive programming in an easy and better way. The project uses data collected from the Codeforces API to create several visualizations that help competitive programmers improve their game in the field of competitive programming.
    - The website includes several interactive charts and graphs that allow users to explore different aspects of the Codeforces platform, including the distribution of competitive programmers across the globe, the frequency of different tags asked in different problem levels, the types of questions solved by users in each rating range, and many more.
    - The visualizations are designed with several data visualization principles in mind, including abstraction, tasks, marks, tables, colour, story, and interaction, to help users quickly and easily understand the data and draw meaningful insights from it. The website is built using HTML, CSS, JavaScript, and D3.js to provide an intuitive and user-friendly experience.
    - Overall, using all the data visualisation principles we learned in this course, we have tried to make this website an excellent resource for competitive programmers looking to improve their skills and better understand the Codeforces platform.

- Data Visualisation Principles used:
    - Abstraction: The project uses various techniques to abstract complex data sets into meaningful insights, such as using interactive charts to show the distribution of competitive programmers across the globe and the frequency of different tags used in various problems. We also performed abstraction at task level to decide the relevant tasks and important {action, target} pairs like {finding trend on tags as problem difficulty changes, deciding to use Multi-Line chart for it}.
    - Tasks: The project focuses on tasks that are important for competitive programmers, such as understanding the distribution of tags across different ratings, analysing the time taken to solve a particular problem, and exploring the most popular programming languages used on the platform.
    - Marks: The project uses different types of visual marks such as bars, lines, scatterplots, and pie charts to represent data in an understandable manner.
    - Tables: The project uses tables in the form of bar charts to show the distribution of users in different rating ranges and the time taken to solve a particular problem.
    - Colour: The project uses colour to highlight important insights and patterns, such as showing the density of competitive programmers across the globe using a world map with colour showing the density of programmers across the world. We have considered different types of colour palettes like sequential, categorical, divergent, highlight while deciding colour scheme for any graph.
    - Story: The project presents a coherent story that follows the flow of data analysis, starting with the distribution of competitive programmers across the globe, analysing the frequency of tags used in different problems, understanding the distribution of tags across different ratings, analysing the time taken to solve a particular problem, exploring the most popular programming languages, and showing the increase in the awareness of competitive programming in India.
    - Interaction: The project makes use of interaction to allow users to explore the data more deeply, such as using play and pause buttons to individually analyse each and every tag in the frequency of

different tags vs. the rating of problems chart, collapsible tree of D3js to show recommended tags for users of different rating ranges, and a bar chart race of D3js to show the increase in the awareness of competitive programming in India.

- o Data to Ink Ratio: We have also tried to make the website such that the data to ink ratio is maximized.

- Charts used in the project:
  1. World Map:
     - For the visualisation of distribution of competitive programmers across the globe we have made an interactive world map.
     - Using the codeforces API we can get the personal data for each of the users on the platform, for example: his codeforces handle, his first name, last name, country and so on. So, as a starting visualisation to give the user a brief idea of distribution of competitive programmers across the globe, after **abstracting** the obtained raw data from the api, we made an **interactive** world map.
     - Continuing further to show the density of competitive programmer's we have used an **ordered/sequential** colour palette (shades of green). To increase interactivity, we have also added a tooltip to show different categories of competitive programmers in that country.
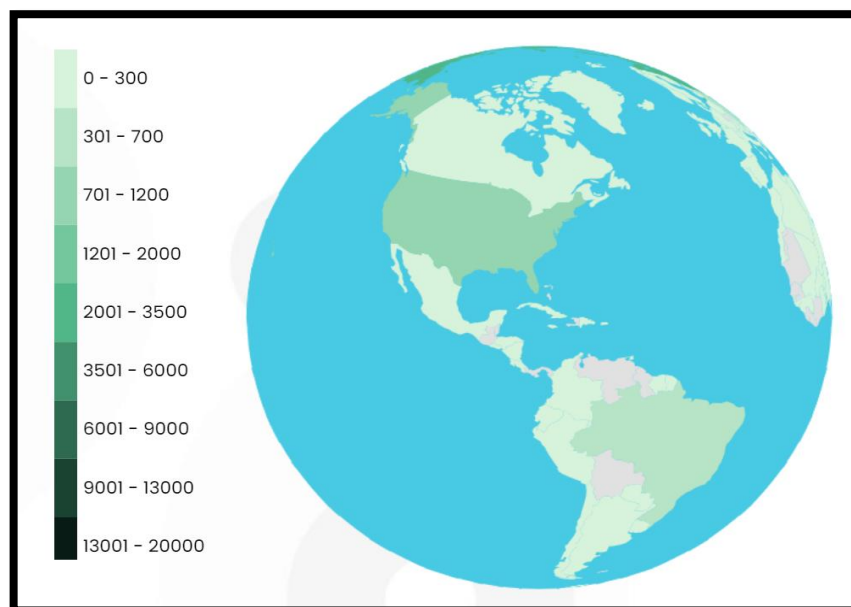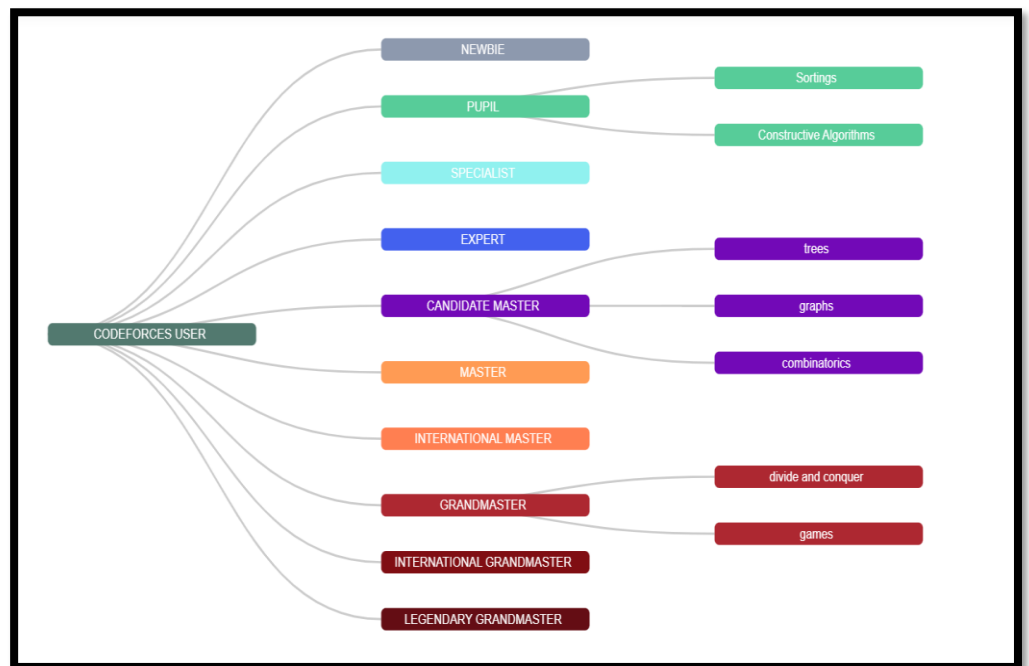
2. Multi – Line Chart:
   - For the visualisation of distribution of different tags across problem level (A, B, C, D, E, F) and problem difficulty we have used a multiline chart.
   - Since, the number of problem tags on the codeforces platform are around 30 - 40, hence showing the trend of frequency of tags for each of the problem level and difficulty was not possible using bars. Hence, rather to show the trend is an easier fashion and allow easy across tag comparisons we used a Multi-Line Chart, with a categorical colour palette.
   - Also, to make the chart interactive we have added tooltip which includes the crucial summary of that tag, and on hover focus feature, to help the user focus on tag of his choice. Further we have also provided the user with a play and pause button using which the user can get the summary of each of the tag in a sequential fashion, which can help the user to know about all the tags and it's relevancy to his level in around 4-5 minutes easily in a video type interactive fashion.
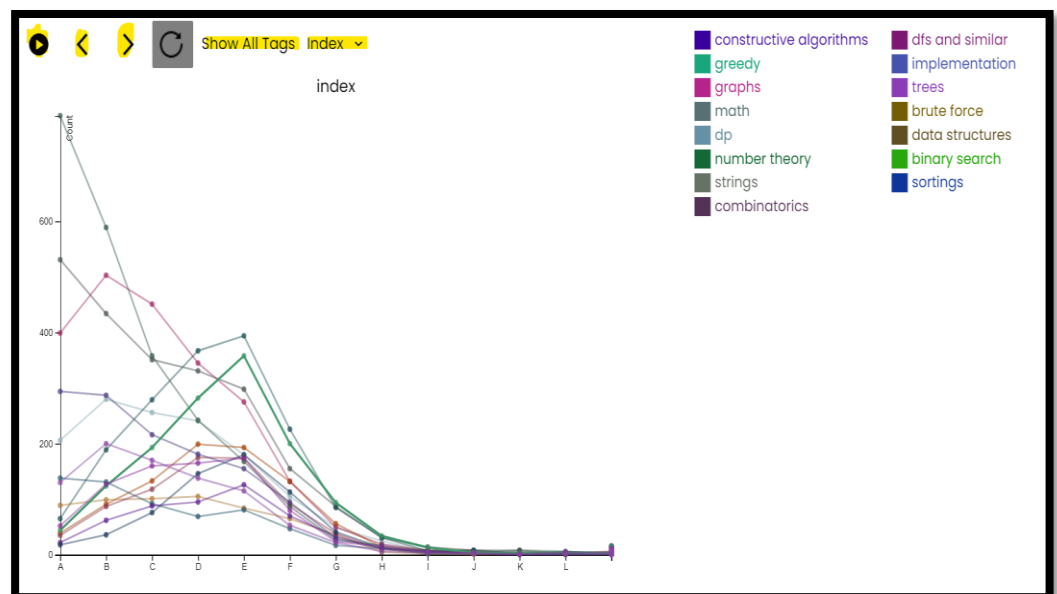
3. Collapsible Tree:
   - After going through the trends of each of the tags, one of the most important thing the user would like is a overall summary of what he should practice depending on his level and to achieve this we have made an interactive collapsible tree.
   - To show any kind of decision making process visually, the trees (like decision trees) are one best and hence to achieve our task of showing relevant tag for users of different rating range, so that they can decide what to focus on, we built a collapsible tree, using appropriate categorial colouring.
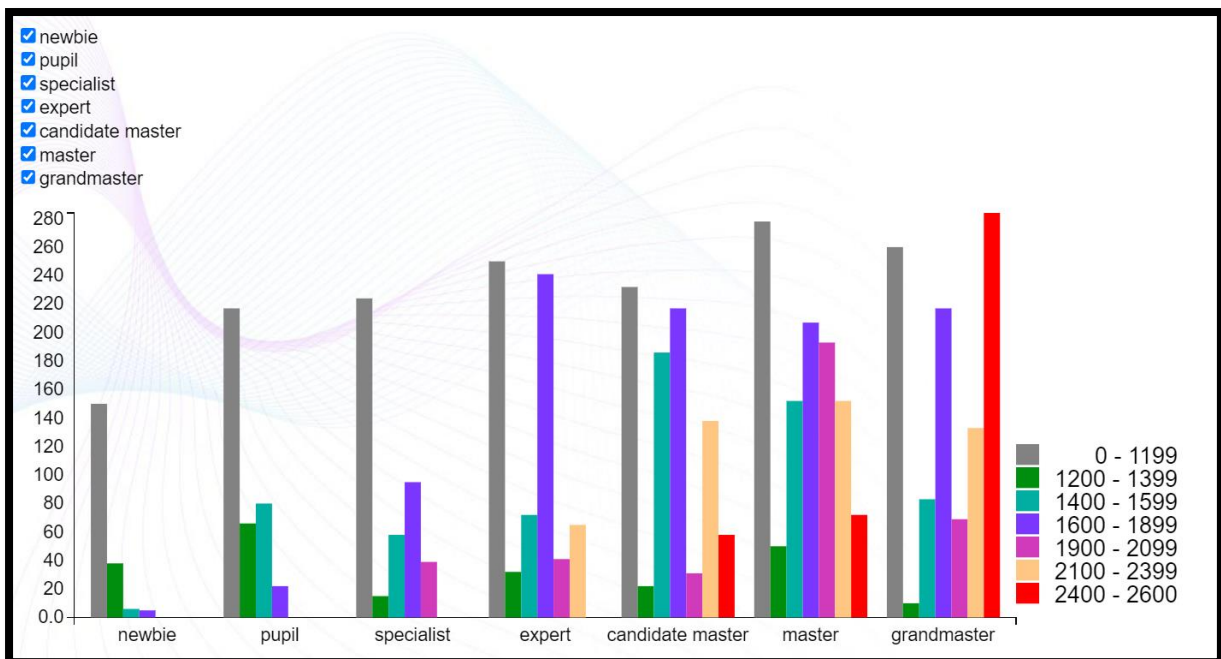
4. Horizontal Bar Chart:
   - Continuing the Summary of the Multiline Chart, we have made a horizontal bar chart, to show the frequency of various tags. The goal of this chart is to provide a final summary about problem tags.
   - We used Bar chart here, since tags are a categorical attribute and we made it horizontal since we wanted to ==show it like a ranking== and the number of categories were large and hence using a vertical chart can cause occlusion and a lesser readability if we displayed tags vertically. We used appropriate categorial colouring for this graph.
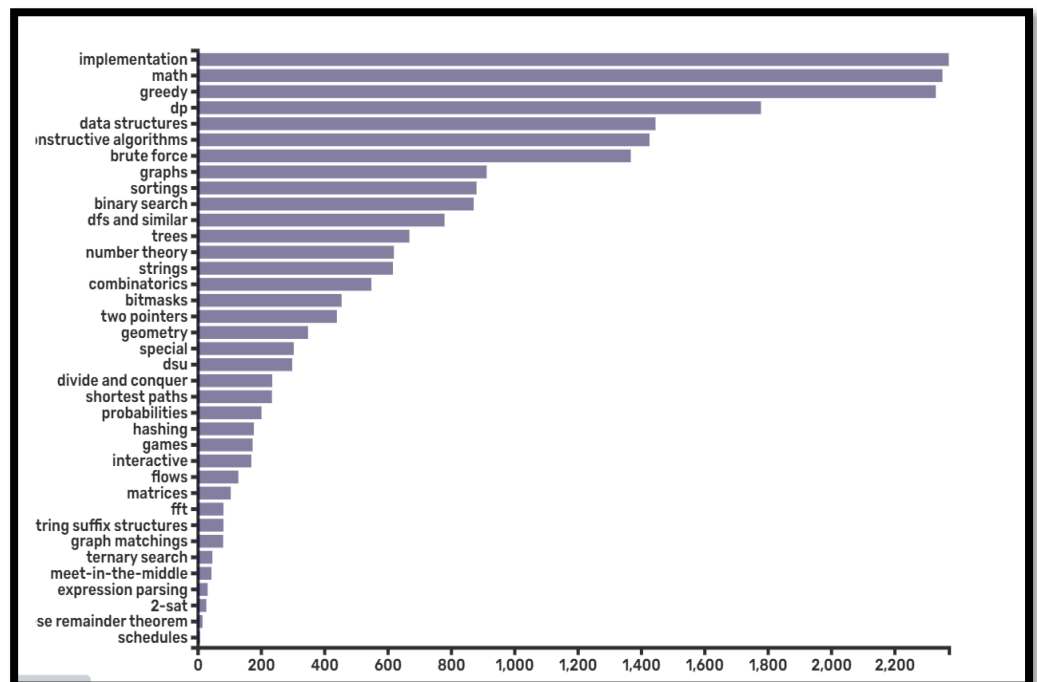
5. Grouped Bar Chart
   - The purpose of this chart is to help users understand the distribution of the problems solved by other users in each rating range with respect to problem rating.
   - Since, we had to represent 2 categorial attributes and 1 quantitative attribute, hence we choose a grouped bar chart, with each user rating range on X – axis, and for each user rating range each problem rating as a separate bar representing the number of problems solved in that rating range. We have used a categorial colour palette, with appropriate colour for each category. The chart has been made interactive with check boxes to allow user to compare a particular problem rating range across different user level's.


6. Horizontal Normalized Stacked Bar Chart
   - This visualization shows the average time taken by a user of a particular rating range to solve a particular problem in a contest. It helps the users to understand how much time they should take to solve a problem to reach a particular rating range.
   - Here, since the aim was to allow the user to do even <mark>part to whole comparison of ratio of total time spent of a particular problem by</mark> people of different rating ranges we used stacked bar chart. We normalized it since same total time is available

to everyone in a contest and since the number of categories were large to avoid occlusion we made it horizontal instead of vertical.

- Here also we have used sequential colour palette, with slightly increasing colour saturation, we choose this colour palette because the problem level were an ordered attribute and to we have added tooltips to make the graphs interactive.
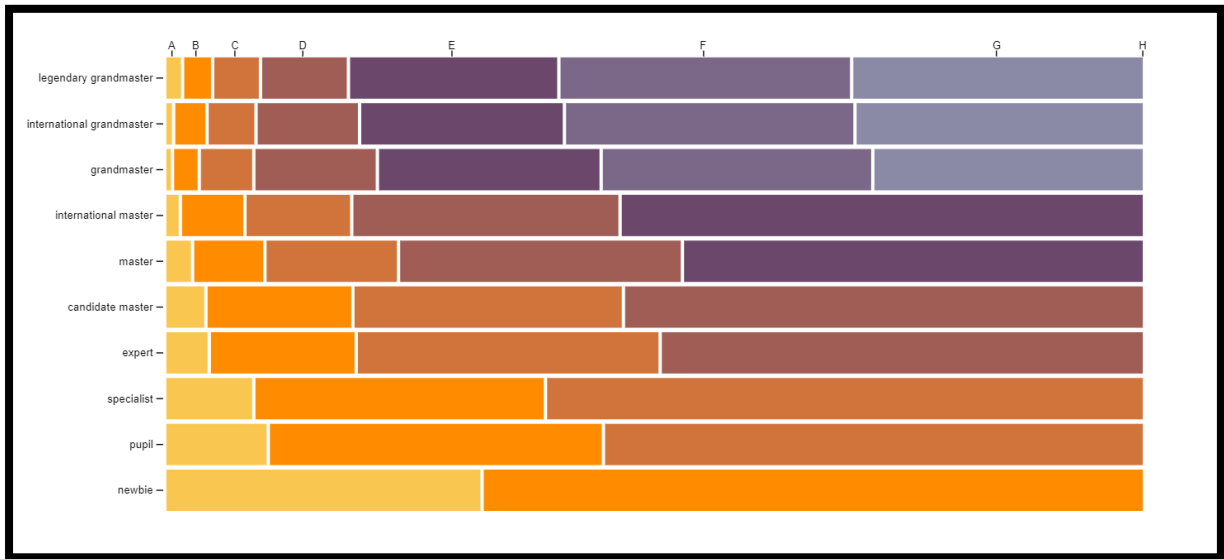
7. Scatter Plot

- This chart helps users understand the time it takes for users to reach a particular rating range. The x-axis represents the user's current rating, and the y-axis represents the time taken in days to reach that rating. The chart shows that users who have achieved higher ratings take more time to get there than those with lower ratings, and it highlights the importance of patience in the field of competitive programming.

- We made a scatter plot here since we wanted to show the relationship between 2 quantitative attributes namely time

taken and user rating. Since we wanted to show the general trend along with the outlier's which can be due to some exceptional people or people with previous experience. Here we have used categorical colour palette, with different colour for each rating category.
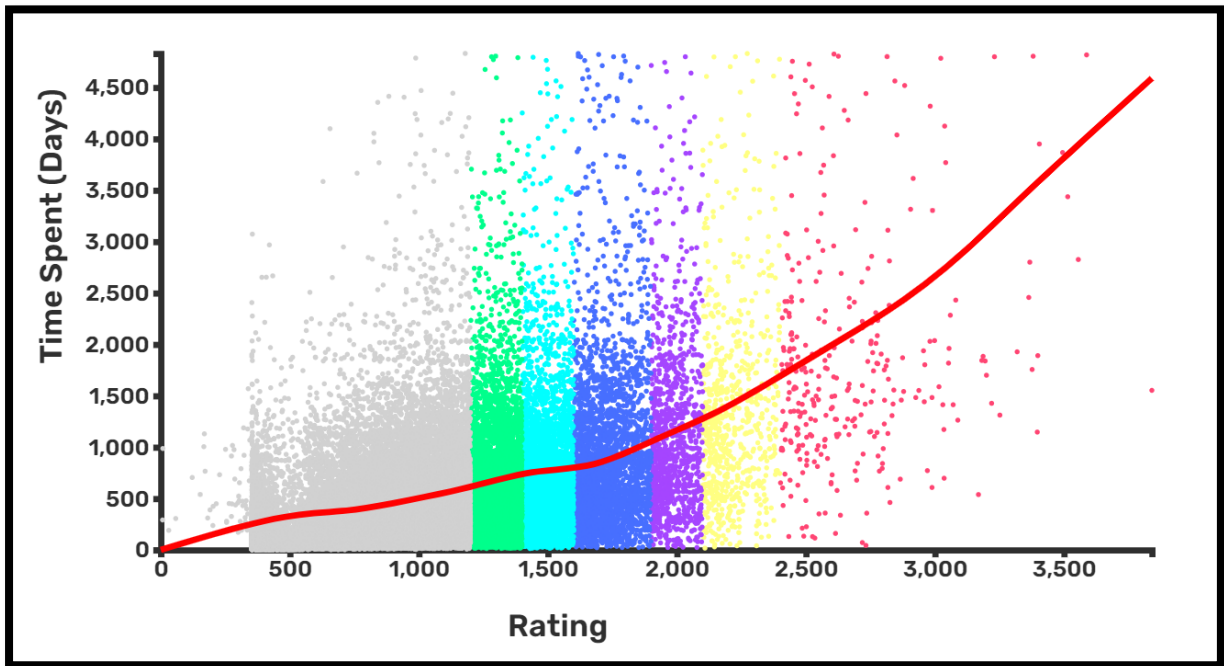

8. Calendar Heatmap
   ▪ This chart uses a calendar heatmap to show the distribution of contests on Codeforces across months. The heatmap is a grid where each square represents a day, and the colour of the square represents the type of contest that occurred on that day. This chart helps users understand the frequency and seasonality of the contests, which can help them plan and prepare for upcoming contests.
   ▪ Here also we have used categorical colour palette with different colour for each type of the contests. Since, there

was no ordered possible among contests we used categorical colour palette and we have also given tooltip to provide interactivity to the user, to help him plan his schedule so that he can participate in maximum contests possible.
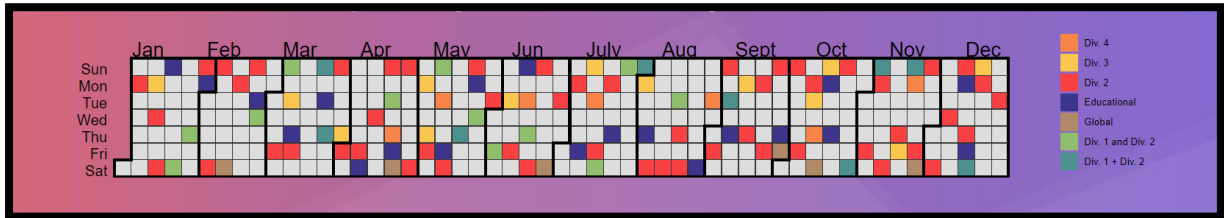

9. Pie Chart
   - This chart uses a pie chart to show the distribution of top 5 programming languages used on the Codeforces platform. The chart shows the percentage of users who use each language, and it helps beginners understand which language is most popular and widely used on the platform.
   - Here, since we wanted to show only top 5 languages, with special focus on the most used one, so that the new users can easily decide which language to pick we decided to make a pie chart.
   - Here, we have followed highlight colouring strategy, where we have coloured the most popular language as blue and provided shades of grey (colour luminance) to rest of the languages. This will help the user to easily see the most popular language and take an easy decision on which language to choose, also to provide interactivity we have added tooltip.
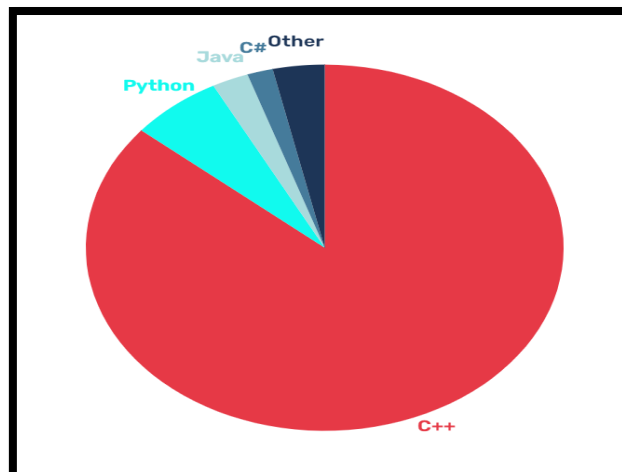
10. Bar Chart Race
- This chart uses a bar chart race to show the increase in the awareness of competitive programming in India over time. The x-axis represents the number of users, and the y-axis represents the timeline. The chart shows how the number of users interested in competitive programming has increased over time in India, and it highlights the impact of COVID and online education on the popularity of the sport. This chart can be useful for users to understand the growth of the competitive programming community in India and to track their own progress relative to others.
- Here, since we wanted to show over time change in number of competitive programmers across nations we used a bar chart race, which keeps on changing as the time changes. Here, also we have used highlight colouring strategy and provided blue colour to India and grey to rest of the nations, to help the user to get a direct track of the growth of the sport in our country India.
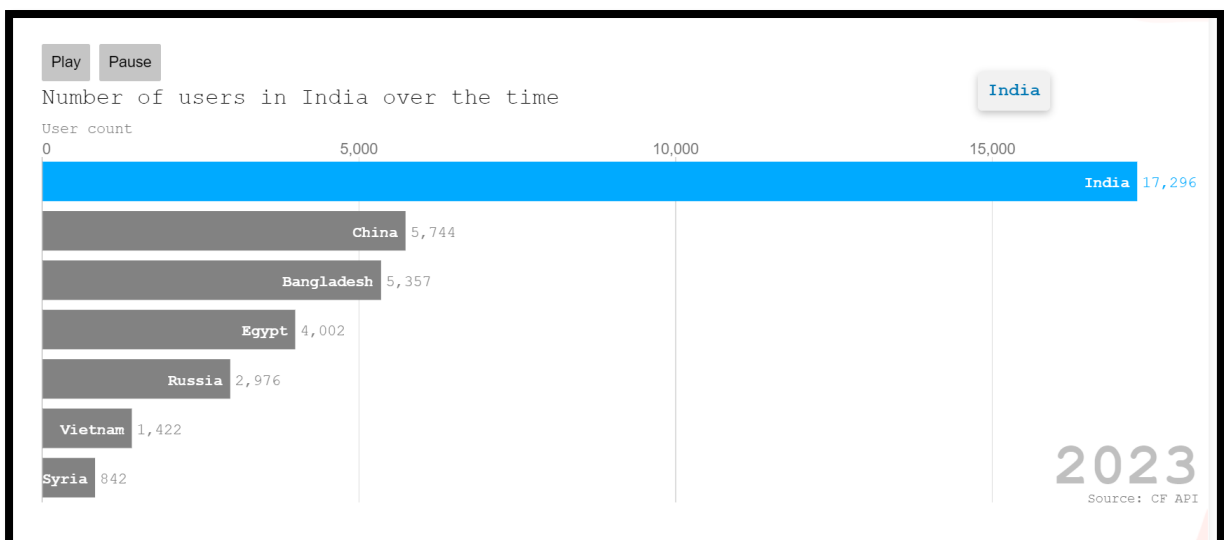
- Conclusion:
  - Overall, this data visualization project on Codeforces is useful for competitive programmers to gain insights into the Codeforces platform and the sport of competitive programming in an easy and better way. The project allows users to explore data interactively and gain a deeper understanding of various aspects of the platform, such as the distribution of programmers across

the world, the frequency of tags used in different problems, and the most popular programming languages used on the platform. The project also provides recommendations on which tags to focus on based on the user's rating range, and it allows users to understand the importance of patience in the field of competitive programming by analyzing the time taken to reach a particular rating range. Overall, this project is a valuable resource for anyone interested in competitive programming and can be used to enhance their skills and knowledge of the platform.

- Link to the video:

  https://www.youtube.com/watch?v=LcvWcN2BSck

- Link to the website:

  https://pattern-pros.netlify.app/

- Link to the code:

  https://github.com/chir263/DataVisualizationProject

- Signed Statement:
  - The contribution to the project goes as follows and we all agree to it:
    - Chirag Jain: Globe Chart, Multiline Chart, Bar Chart Race and getting data from codeforces API.
    - Pratham Priyank Thakkar: Grouped Bar Chart, Normalized Horizontal Bar Chart, Collapsible Tree, Calendar Heatmap and getting data from codeforces API.
    - Aryan Bansal: Scatter Plot, Pie Chart, Horizontal Chart, and a customized bar chart (which we gave in mid submission but later replaced with scatter plot as we found scatter plot more suitable for our visualisation).
    - We combinedly worked on the website.