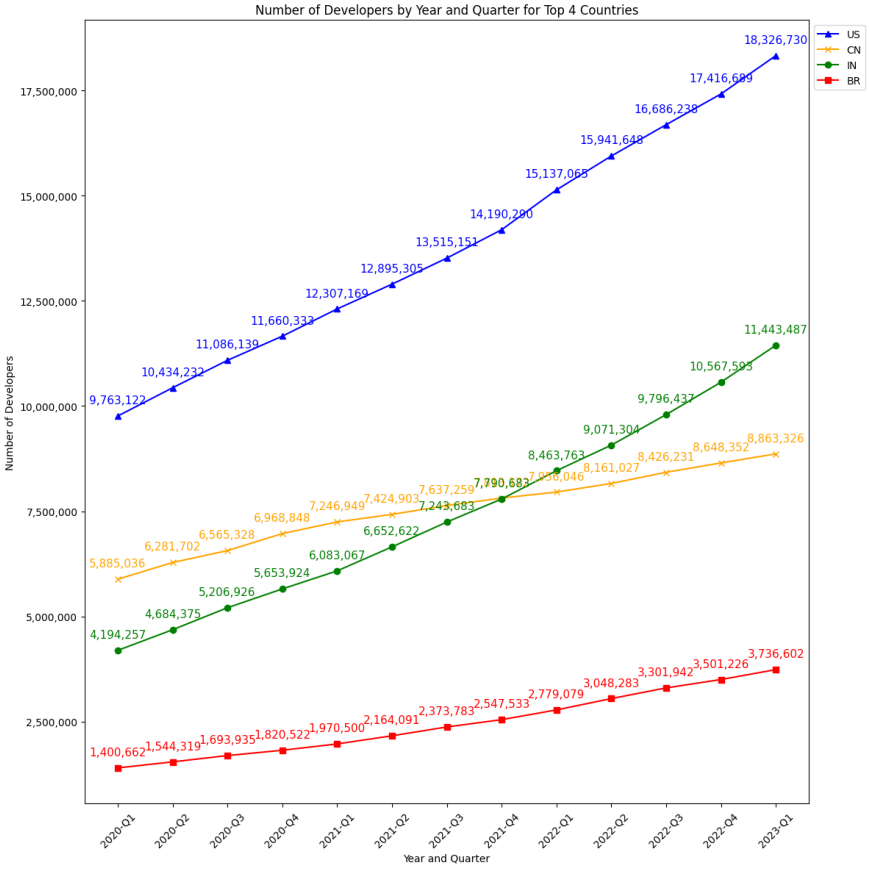
INFSCI 2415 Information Visualization Final Report

GitHub Developer Characteristics Study

Figure I: Number of Developers by Years and Quarter for Top 4 Countries



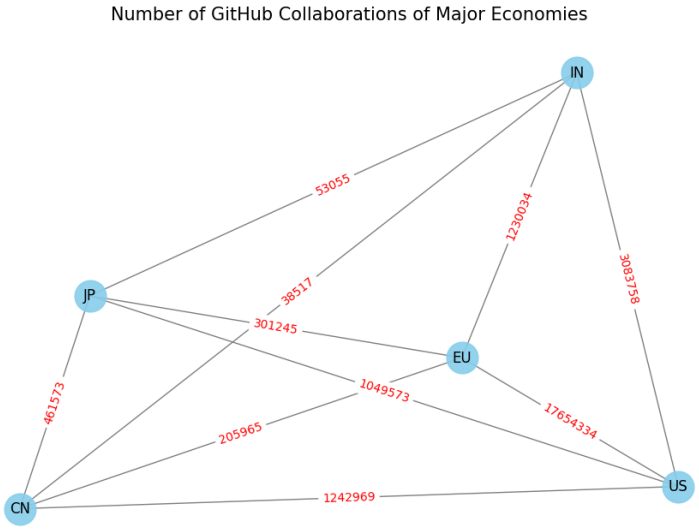
Legend Explained:

- The blue triangular line represents the United States
- The Orange "x" line represents China.
- The green dot line represents India.
- The red square line represents Brazil.
- The colored values represent the number of GitHub developers for the corresponding country in different years and quarters.

Findings text introducing highlights of the produced figure in bulletin points:

- This is a line plot that shows the number of GitHub developers in different countries from the first quarter of 2021 through the first quarter of 2023.
- From the plot, the US has always had the most GitHub developers in these years. It's in the top four countries with the most GitHub developers in the world, along with China, India, and Brazil.
- 2021 Q4 was a turning point, after which India started surpassing China in terms of the number of GitHub developers.
- All the top four countries have maintained their growth trends in the number of GitHub developers.

Figure II: Number of GitHub Collaborations of Major Economies



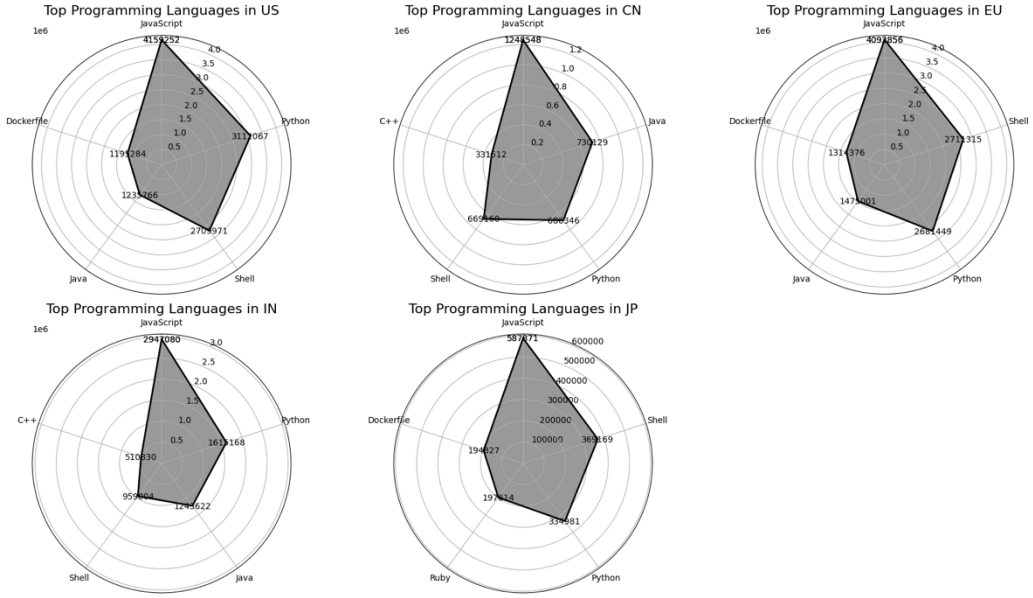
Legend Explained:

- The five dots represent the United States (US), China (CN), the European Union (EU), India (IN) and Japan (JP).
- The connecting lines between each of the two points represent the cooperative relationship between the two economies.
- The numbers on the connecting lines represent the total number of collaborations between the two economies.

Findings text introducing highlights of the produced figure in bulletin points:

- This is a network plot that shows the number of GitHub collaborations of five major economies from the first quarter of 2021 through the first quarter of 2023.
- From the plot, there is a great deal of collaboration among developers in all five economies. Among them, the United States has the most cooperation with other countries. Its cooperation with the European Union is the most, reaching 17,654,334 instances over three years.

Figure III: Top Programming Languages in Major Economies



Legend Explained:

- Each radar chart represents the United States (US), China (CN), the European Union (EU), India (IN), and Japan (JP).
- The labels on the radar chart show the number of GitHub commits for popular programming languages.

Findings text introducing highlights of the produced figure in bulletin points:

- These radar plots show the number of the five most popular programming languages in five major economies from the first quarter of 2021 through the first quarter of 2023.
- From the plot, JavaScript is the most popular programming language in the world. In all the five economies, JavaScript is ranked #1 among all the programming languages.
- In addition to Javascript, Python and Shell are also among the top five most popular languages in five countries. Java did not become one of the top five favorite languages for Japanese developers.

Data and method text describing the data and method used in this process.

- I used three separate data files, which counted the number of GitHub developers in different regions (developers.csv), the number of collaborations (economy_collaborators.csv), and the programming languages usage (languages.csv).
- I used Pandas, Matplotlib, Numpy, and NetworkX to create all the plots in Jupyter Notebook.
- For the first plot, I mainly used matplotlib.pyplot() to create a line plot.
- For the second plot, NetworkX, a powerful network analysis tool, was utilized to construct and visualize the network graph. I used Pandas to process the data to aggregate the number of collaborations and used NetworkX to create the plot.
- For the third plot, I processed the data using the Pandas library to filter out the five most used programming languages in each economy. The Numpy library was then applied to calculate the angle of each axis of the radar plot, ensuring that the data was evenly distributed across the plot. I ended up creating the radar plots using Matplotlib.

Significance statement on why the presented figure is important.

As an information science student, it is important to understand the preferences of other developers around the world. The three plots above show the characteristics of GitHub developers from three perspectives and allow us to clearly determine how many developers are in a specific country, the growth trends of the numbers of developers, their preferred programming languages, and partnerships between countries.

Data Source: https://www.kaggle.com/datasets/konradb/github-innovation-graph

GitHub Link: https://github.com/BarclayFu/INFSCI2415