

stackoverflow as a Social Network

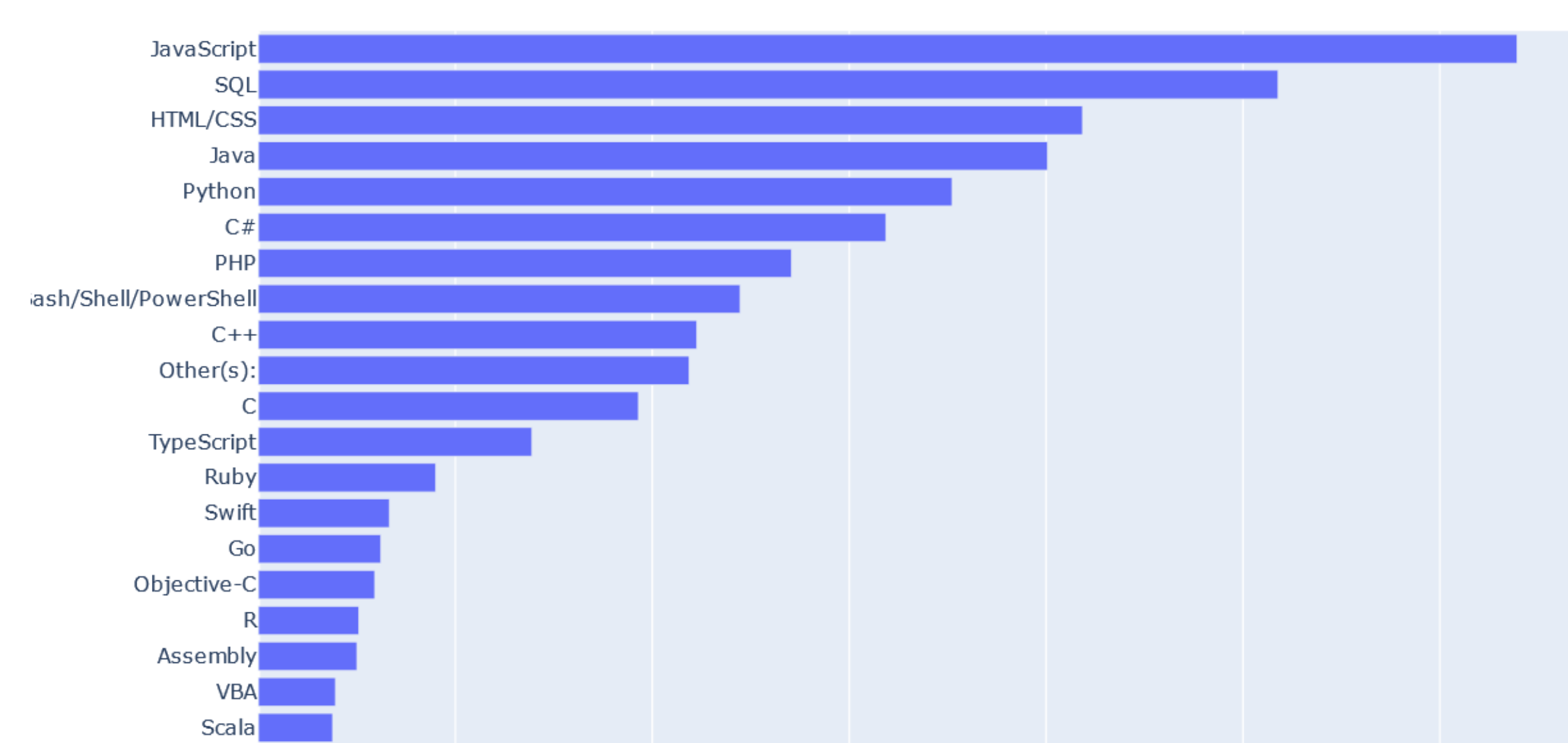
This project aims to study the characteristics of the user interaction on Stack Overflow (SO). How and when are users from all over the world communicating on SO? Are there any noticeable programmer communities on SO?

We used three Datasets:

1. Stack Overflow Annual Developer Survey
2. Stack Overflow Data from Google BigQuery
3. World Bank Open Data / UN Data / Kaggle World Bank

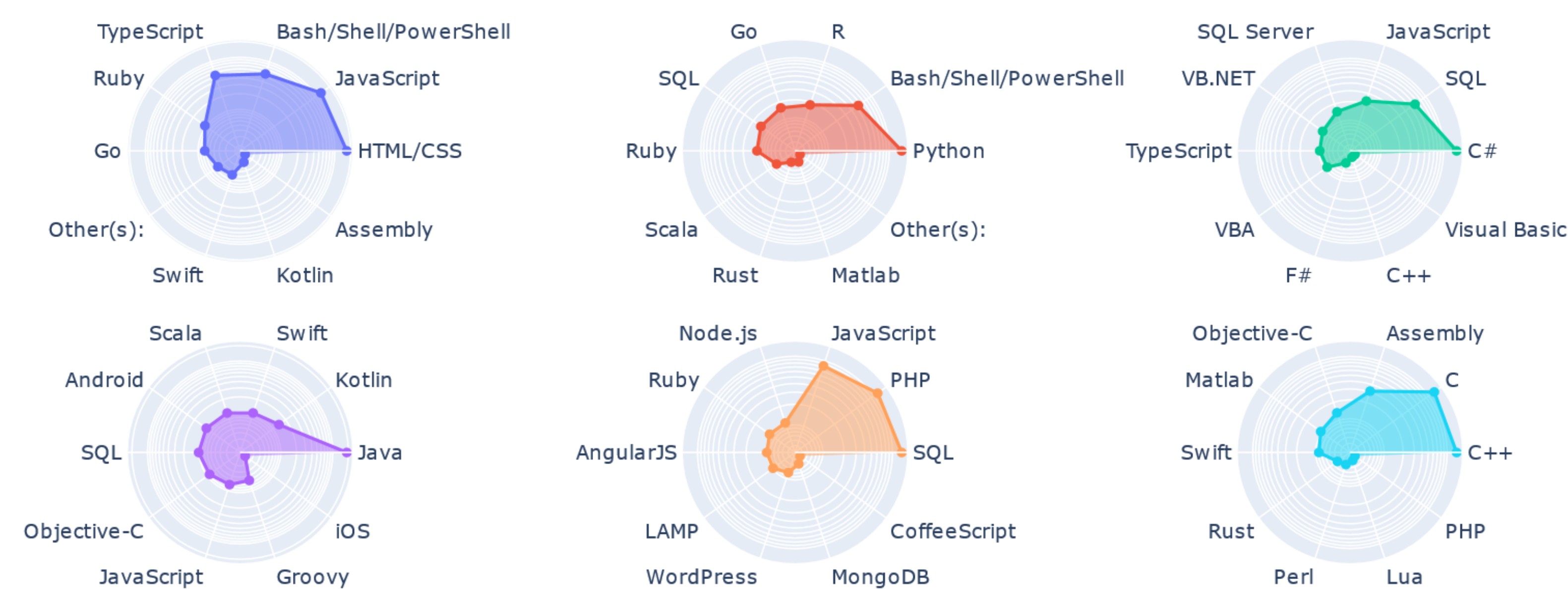
We categorized our most interesting findings into three sections, **User**, **Time** and **Country** related.

Programming Language Distribution



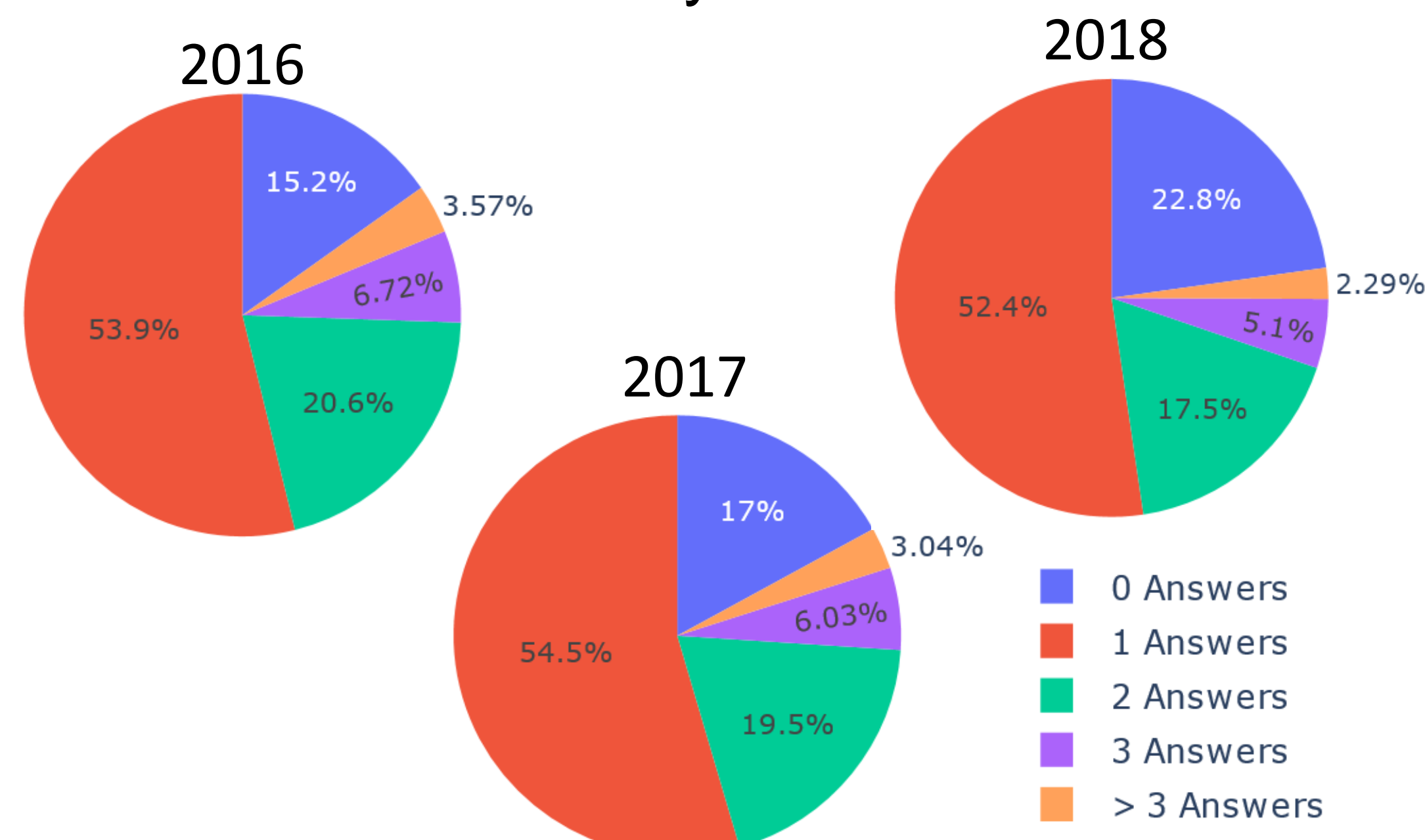
JavaScript and SQL are the most widely used languages. A handful of other languages, like Java and Python, also show up as "large" languages.

Six Major Types of Programmer on SO



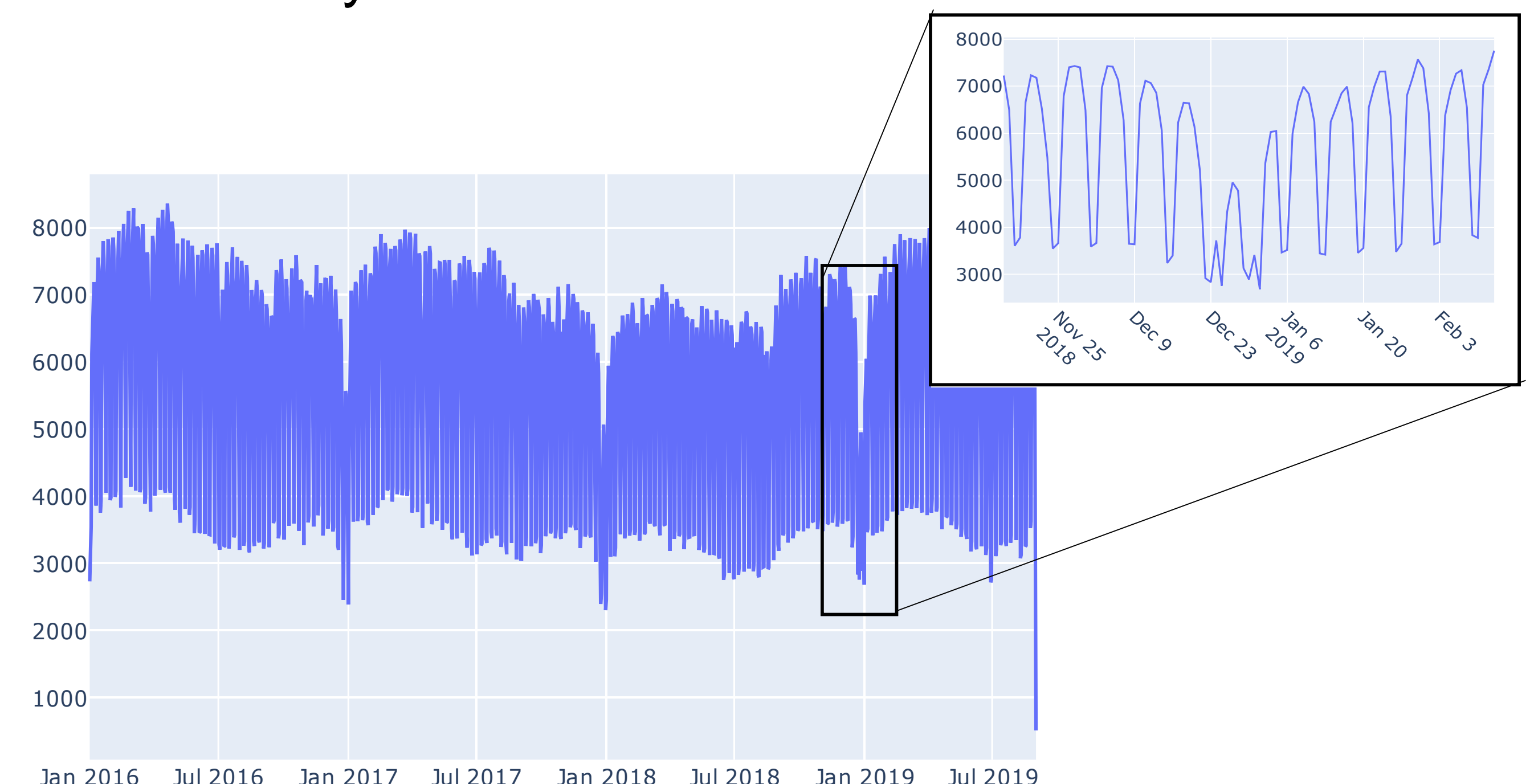
Using non-negative matrix factorization on the user language used data, we discovered 6 programming "communities" or programming archetypes, which intuitively also correspond to different types of programming.

Answers Received by Questions Over Time



1. Noticably 15% to 20% of unanswered questions.
2. Majority of the questions don't receive more than one answer.
3. More questions are being ignored over the years.

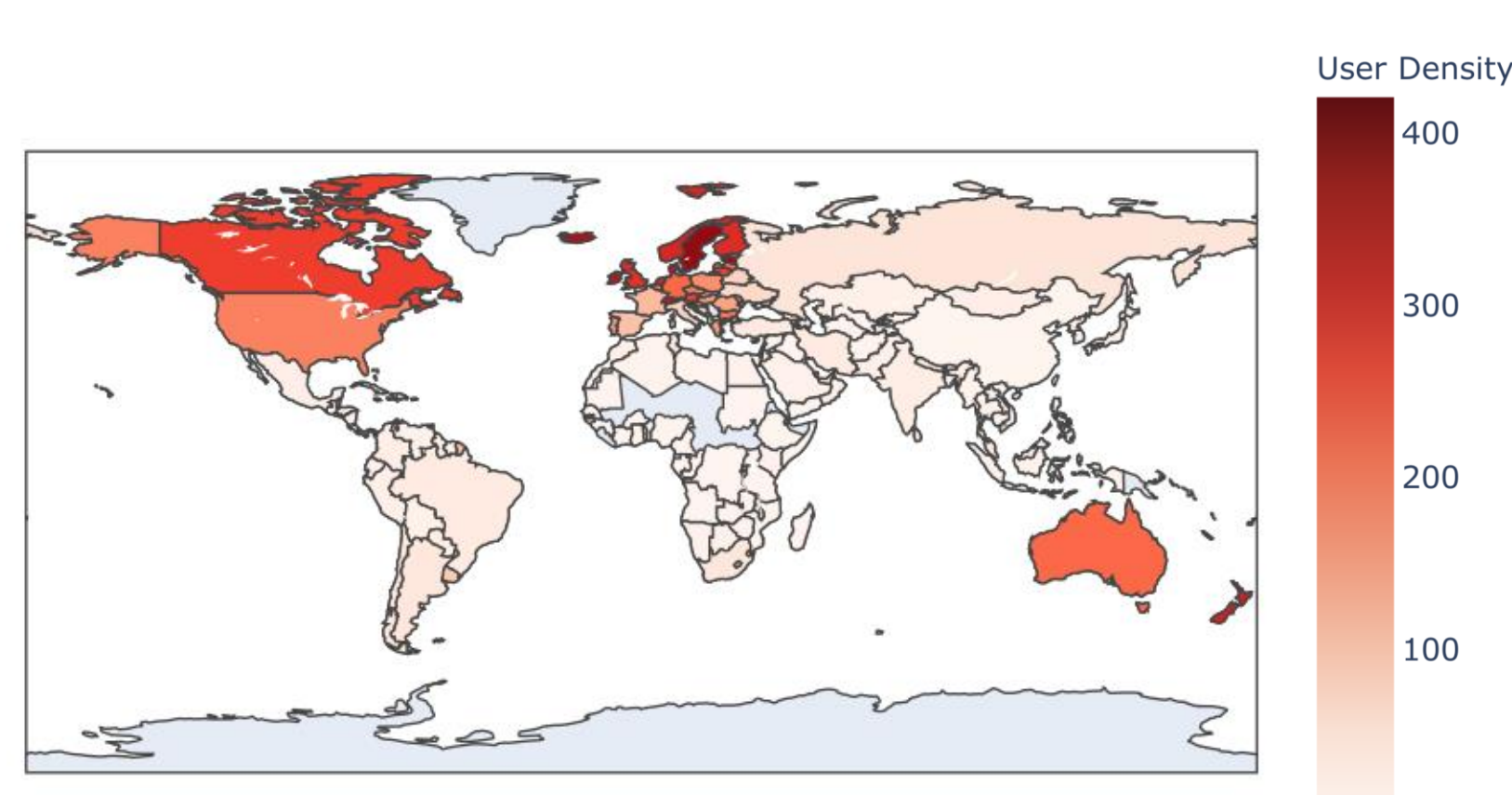
Daily Question Pattern on SO



1. A pattern of usage rate shows up where the amount of questions decreases to almost half during the weekends compared to working days.
2. The count drops even lower during Christmas and New Year's every year.

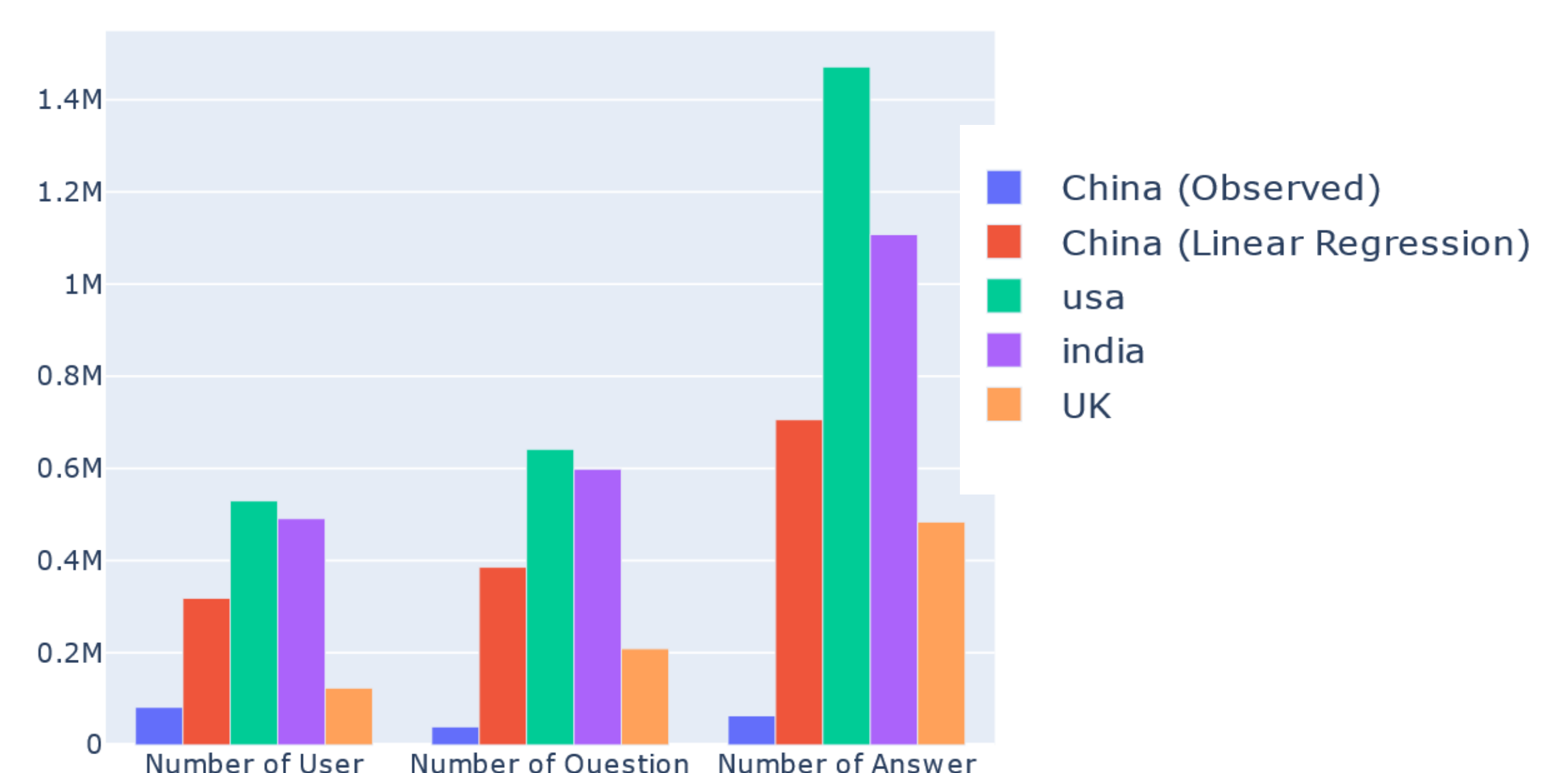
User Density by Country

2016-2019 StackOverflow User Density



1. Sweden is the top country in terms of amount of users per capita.
2. The user density of China is abnormally low, which is probably due to Stack Overflow being partly blocked in China.

Potential of China for SO



A linear regression model is trained to predict the user activity for a country with features such as population, internet usage rates and gdp per capita, using data from countries other than China. China could become the country with the third largest amount of Stack Overflow users under the model.

