Finding Optimal Placement for New BLUEbikes Stations

CS504 project | Yufeng Chen | yufeng72@bu.edu

Project Purpose

Boston is a large growing city, and public transportations is an important aspect of its city planning. Recent years bike sharing systems like Bluebikes become more and more popular, and now it has over 2500 bikes and 260 stations around Boston. People ride their bikes everyday and it's really convenient. However, the placement of their bike stations are not very reasonable, and that brings bad experiences to its users (for example: college students like me). The purpose of this project is to find out which places are good choices for placing new Bluebikes stations or replacing old ones.

Datasets and Data Portal

I used five datasets as data sources of this project, they are:

Subway stop locations (.json file)

Bus stop locations (.csv file)

*College and university locations* (via BostonMaps Open Data API)

Bluebikes station locations (via Bluebikes API)

Bluebikes trip data 2018.9 (.csv file)

Methodology

To get some useful information from data I retrieve, I implemented several transformations to process them. First I looked for bus stops, colleges and universities with latitude and longitude in Boston to find possible places for placing bike stations. The second step is to count the number of the Bluebikes stations near all colleges and universities in Boston. Then I analyzed the Bluebikes trip data of one month (2018.9) to see what places are more popular and what is the age distribution of Bluebikes users. To select locations for new stations, I tried several methods like K-Means, and finally decided to set a congestion score to determine optimal locations.

Statistical Analysis

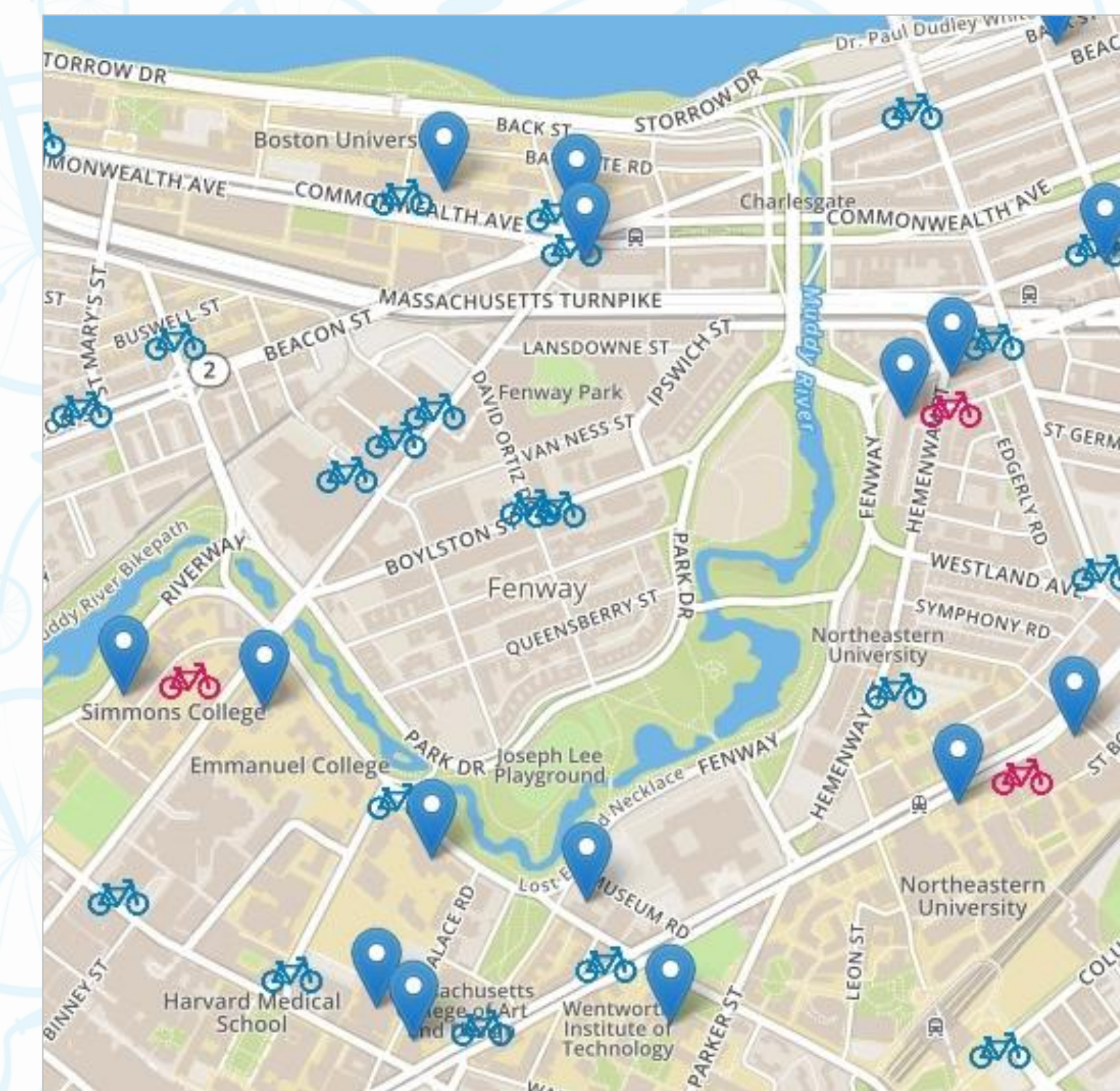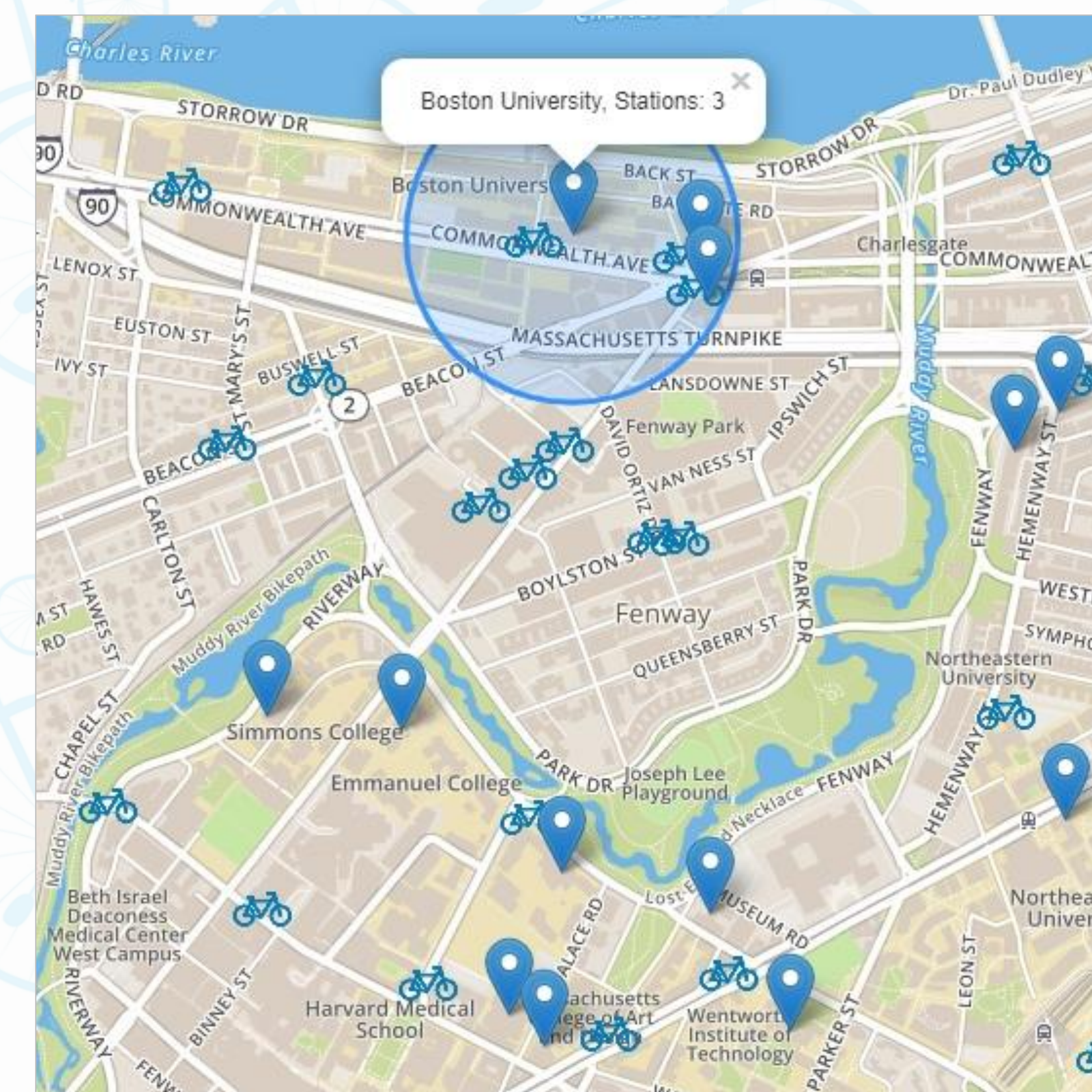**AGE DISTRIBUTION OF BLUEBIKES USERS**
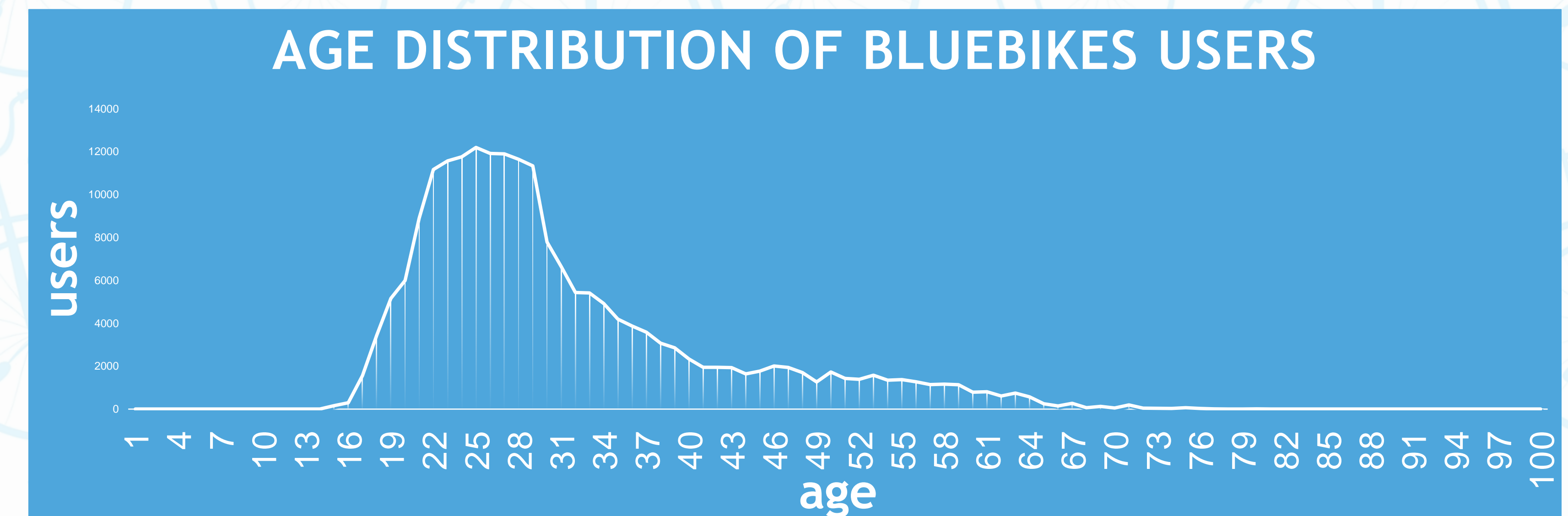
Correlation Coefficient (age and member):   -0.3216
Correlation Coefficient (trips and stations):   0.7389
Correlation Coefficient (trips and new stations): 0.9026

Conclusion and Future Work

The methods used in this project calculate the lowest average congestion score given the number of new Bluebikes stations and thus can tell us the optimal placement for these stations. To make it more practical, we can consider using real distance (the distance of the actual route between two places) instead of straight line distance. That requires more data like streets and crossroads information, and I'll keep working on it.