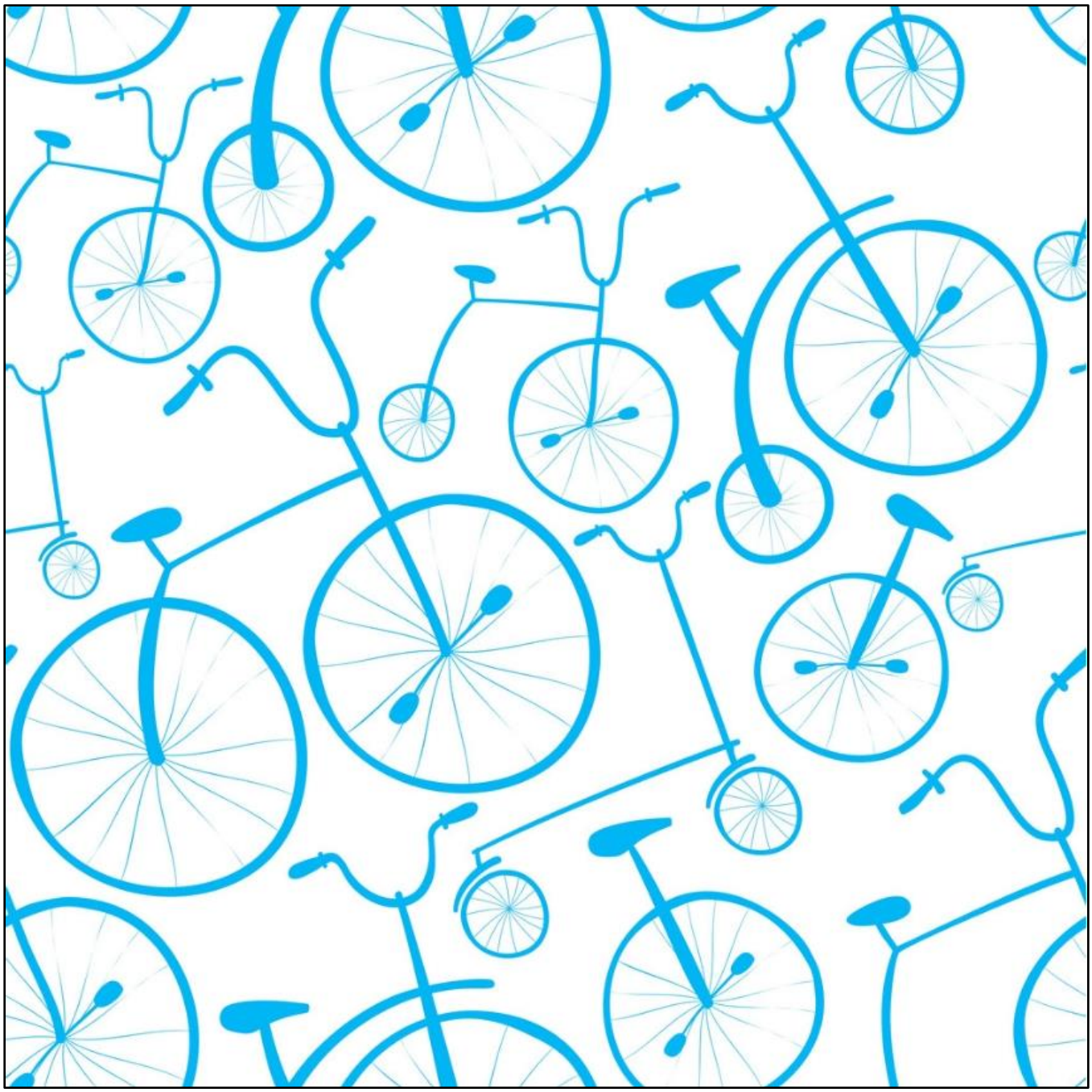


Finding Optimal Placement for New BLUEbikes Stations

CS504 project | Yufeng Chen | yufeng72@bu.edu

Project Propose

Boston is a large growing city, and public transportations is an important aspect of its city planning. In 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 like me ride their bikes every day 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). For example, Blue bikes only has one bike station with 10 docks at BU campus, and sometimes when I ride to the campus, I have to look for another bike station nearby because that one is already full. I was actually late twice for CS504 class this semester because of this! That makes me thinking: if Bluebikes can learn more about their bike-using situation and place their bike stations more properly, it will be very nice for their users. Thus 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:

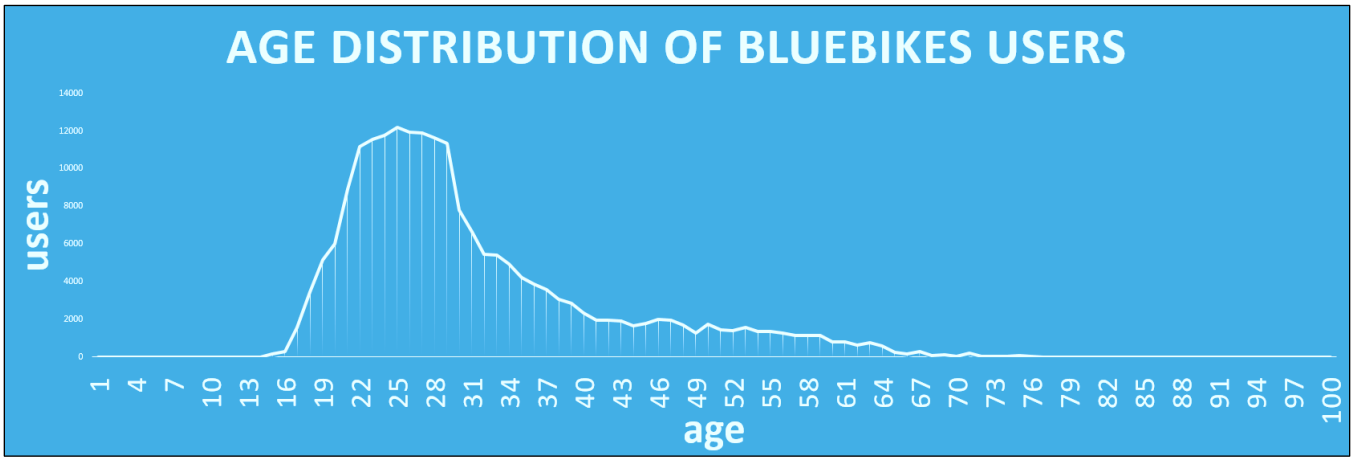
Subway stop locations	(.json file)
Bus stop locations	(.csv file)
College and university locations	(.csv file via BostonMaps Open Data API)
Bluebikes station locations	(.csv file via Bluebikes API)
Bluebikes trip data 2018.9	(.csv file)

Data Processing

After retrieving datasets and saving them into MongoDB, it's time to get some useful information from the data. I implemented three transformations using selection, projection and combination. The first transformation outputs bus stops, colleges and universities with valid latitude and longitude in Boston to let me know possible places for placing bike stations. The second one calculates distances between every Bluebikes station and every college or university in Boston, then count the number of the Bluebikes stations near these colleges and universities. The last transformation analyses Bluebikes' trip data and finds out which area or places are more popular among Bluebikes users.

Solving the problem

Now it's time to actually solve the bike station placement problem. First I analyzed Bluebikes' trip data to get the distribution of its users. The result are provided below:



As you can see, young people from age 19 to 28 are the main users of Bluebikes. With the fact that Boston has the largest student population and proportion in the world, it's easy to tell that the majority of these young people are college students. So I decided to mainly use college and university data to determine the placement of new bike stations.

First I implemented K-Means algorithm on my dataset to get several optimal locations. However, there is one problem: K-Means only care about the distance, but the distance is not the only factor that matters. The population density and the trip density are more important. Thus I decided to set a congestion score calculated by these factors to determine if there are enough bike stations in one area. The placement with the smallest average congestion score is the best placement.

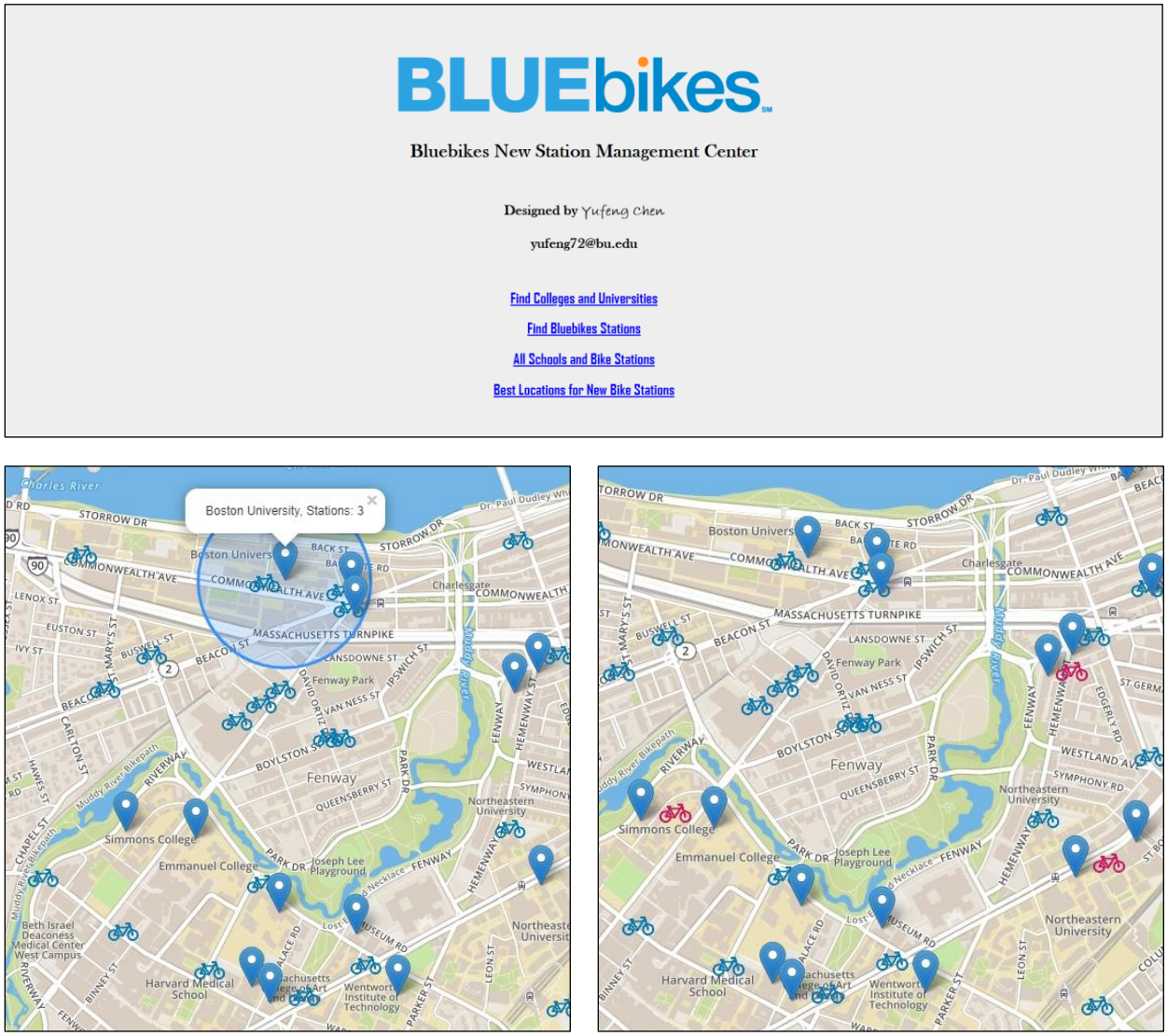
Then I treated the problem as a constraint satisfaction and optimization problem to solve it. The first satisfaction question is: can we achieve a congestion score less than 600 for all colleges and universities with 20 new bike stations? The answer is no, and the smallest number to satisfy the constraint is 55. The second

question is: what is the best placement (lowest average congestion score) given the number of new bike stations? I got the top 10 choices for this question and saved them into the database.

After solving the problems, I did some extra statistical analysis. For the old placement, the correlation coefficient of trip density and bike station number in an area is 0.7389, and for my new placement the correlation coefficient comes to 0.9026, which means my placement is more reasonable.

Visualization

I add some interactive web-based visualizations as new features to my project. These visualizations can give you an intuitive explanation of the problems and my solutions. Here are some screenshots:



Homepage (top) / Check stations around the school (bottom left) / Optimal placement (bottom right)

Conclusion and Future Work

The methods used in this project calculate the lowest average congestion score, and thus can tell us the optimal placement for the Bluebikes stations. However, there are still space for improvement, like using real distance (the distance of the actual route between two places) instead of straight line distance to make the solution more practical. That requires more data like streets and crossroads information, and I'll keep working on it.

Running My Code

To run functions except visualizations, type:

```
python execute.py yufeng72
```

I implemented trial mode for main functions. To run in trial mode:

```
python execute.py yufeng72 --trial
```

To run visualizations, first run execute.py and copy auth.json and config.json under /web. Then go to /web and type:

```
python app.py
```