BlueBikes Visualization Tool

Kevin Johnson

4/20/2020

Introduction

As an avid cyclist myself, I was originally motivated to see what the City of Boston can do to encourage people to use cycling as their form of transportation over alternatives. Cycling usage can be seen as a self-reinforcing cycle, where more cyclists lead to increased investment in cycling infrastructure, which drives more people to chose to cycle. By identifying features of an area which drive people to chose to cycle as their mode of transportation, the city can work to highlight and improve these features.

Which of the features are associated with increased ridership will likely not be novel. What I hope to bring to the table is a discussion on the relative priorities of these features. For example, when other variables are controlled, is `percent tree cover` or `percent residential zoning` a greater influencer on ridership.

To answer these questions, I will aggregate data from a handful of data-sources by census tract. Then I will run a regression analysis to identify the correlation between ridership and other factors. Once I have the regression analysis, I will identify the trends and generate some comparison maps and scatterplots to visually explore the trends.

TODO Research questions and hypothesises

Methods

Datasets

Census Tracts

The fundamental dataset that I will be using is the 2010 Census Tracts. This data is provided by the Analytics Team at Boston Maps. This dataset defines the geography of the different census tracts within the City of Boston. On it's own, this data does not provide me much that I need, but by aggregating different data sources into each census tract I will be able to develop an analysis.

The geometry of each census tract has been used to correlate the following datasets with a tract `id`. This is the key with which I will do my aggregation on.

Bluebikes Usage Data

Critical to any analysis about cycling in Boston, the BlueBikes usage data has detailed information about every trip that BlueBikes users have taken. This includes information about the starting and ending station, duration, start and end time, rider demographics and other data. We will be using this data in aggregate over the course of a month, so the privacy of individual users will be respected.

Bluebikes Station Data

Valuable for correlating the usage data with census tracts, the station data is provided by the BlueBikes company and includes the latitude and longitude for each station. I have done additional processing on this data to identify the census tract containing each station.

Boston Zoning Data

One feature of census tracts that I was interested in correlating with BlueBikes usage was the type of buildings in the tract. An unanswered question was whether Residential

This data was harder to aquire, as it was not already in an easily accessible form such as JSON or CSV. Instead the City of Boston was hosting an ArcGIS server which included the zoning of each building. I wrote a script which manually downloaded each building and zoning region from the ArcGIS server and wrote it out to a CSV file for processing in R.

Tree Canopy Data

The tree canopy data provided by BARI has information on the tree cover by census tract. I downloaded the data in GeoJSON format, and then t

Aggregation

Data Aquisition

Analysis

Results

Discussion

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.