

PS5 McGuire

Joel McGuire

February 2018

1 WalkScore Scraped

The data I scraped was from walkscore.com which finds the distance or average distance it takes for one to walk to a basket of services normally necessary to live a happy and healthy life. I'm more interested in the premise than the data itself. It was too difficult to run a crawler. The data itself is already in a ranked and tabular format that conveys considerable information. The most walkable large cities are located on the east or west coast or not in the USA. Indexes are positively correlated. I may use it later because I'm interested in seeing the relationship between walkability and the trust in a community, because trust seems a metric that binds a lot of other socially desirable measurements together and its a channel that generally connects well-being and prosperity. The only tutorial I used was the one presented in the notes.

2 Twitter API

I used rtweet for accessing the twitter API and ,dplyr, tidyr, and tidytext for cleaning the tweets and formatting it into a table. The only thing I found to be noteworthy was how filtering for english and geolocation cuts down the size of the data. In my most previous run I went from 2700 randomly sampled tweets to 9 tweets that were geolocated and in english.