Patterns of Life in NYC with Citibike

Objectives

1. **Does Citibike increase economic activity?**

   Citibike allows residents and visitors of New York City to explore areas of the city that may not be as easily reached by NYC metro alone. I will investigate whether use of Citibike corresponds to changes in an area's economic activity. As patrons move around NYC via Citibike, do residents experience increases in taxable income?

2. **Does Citibike increase exploration of the city?**]

   How are New Yorkers moving around their city via citibike? Do users typically take short trips in one area, or do the bikes move to vastly different areas of the city. I will be tracking the movement of individual bikes and trying to draw conclusions about their movement.

3. **Can we make any predictions about economic growth based on Citibike data?**

   Given the previous conclusions, are we able to look at changes in Citibike patterns to predict what areas will experience growth? Does the movement of people on bikes contribute to gentrification of neighborhoods?

Description

The Citibike data is extremely large. Consider how many people are in New York City. Every ride is logged with *date, time, departing station, returning station, duration, bike id, rider sex, and rider birth year*. The station data is provided as WGS84 latitude, and longitude. The NYC open data platform provides GIS polygons providing zip code boundaries. There are 139,906 points that need to be transformed from the NAD83/Long Island, NY coordinate reference system, to WGS84. This is a computationally moderate mathematical process that will benefit from parallelization. I intend to utilize CUDA to speed up the transformation. There are two opportunities for parallelization. Clearly each point can be computed in parallel, but within each point the X and the Y can be computed independently.

Observing how bikes move around NYC is analogous to a large graph problem. There are 187 stations, and over 10,000 bicycles. I believe that the data can be viewed as a social network with stations as nodes and bicycle trips as directed edges. Processing this graph can help us draw conclusions about increasing connectivity of the city. There are over 1 million trips ***per quarter***.