Gentrification Detection

Yuting Gui   Chaoyue Liu   Yining Gao

New York University   Superviser: Juliana Freire

Introduction

Gentrification is the influx of middle or upper-class people into a lower income neighborhood, resulting in the renewal of the area.

The main objective of this project is to :

- Quantify measurement of gentrification using Property Construction Data & US Census Data.
- Detect some correlation between Gentrification and Taxi Trips.

Data Preparation

Data Source:

- *NYC Property Construction Activity Data (1990 - 2012)* , Size: 1.04G
- *LTDB: American Community Survey and Census Data (2000 & 2010)*, which includes population, poverty ratio, rental price, etc. Each record has a location based tract ID.
- *NYC Yellow Cab Trips Data (2009 - 2013).* Size: 10-20 GB for each year.

Data Processing:

- *Spatial Index Assignment:* LTDB provides polygon shape information associated with each Tract ID. We assign each record in other two datasets with a Tract ID in Map Script.
- *Map Reduce Based Aggregation:* For each year's trip data, Map produces a {Track ID, 1} pair . Reduce counts the number of pairs.
- *Data Transformation* For trips data, we replaced trips amount with trips ratio. For other Census/Construction Data, we calculate the growth percentage. We then normalized all data into [0,1] and projected them into 11 bins.

Analysis and Conclusion

Spearman Correlation Coefficient: Spearman Correlation statistically measures the strength of association between two ranked variables. Statistically, the correlation between trips data and the economic indicator listed is not very strong (< 0.5).
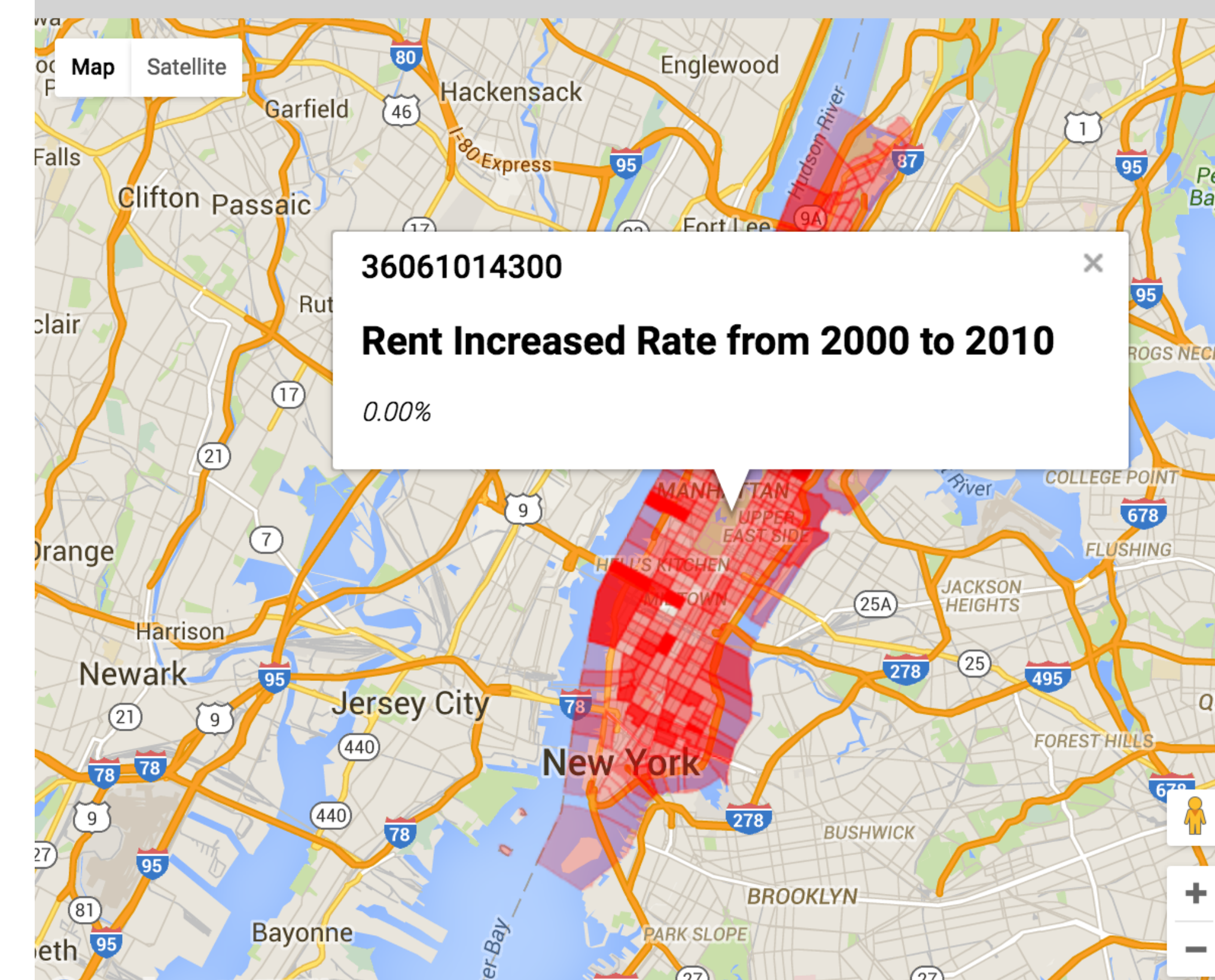
Interactive Visualization We deployed the Google Map API, and use the KML file to display analysis. Map on the left displays US Census/Construction Activity data. Map on the right displays taxi data amount for each region.
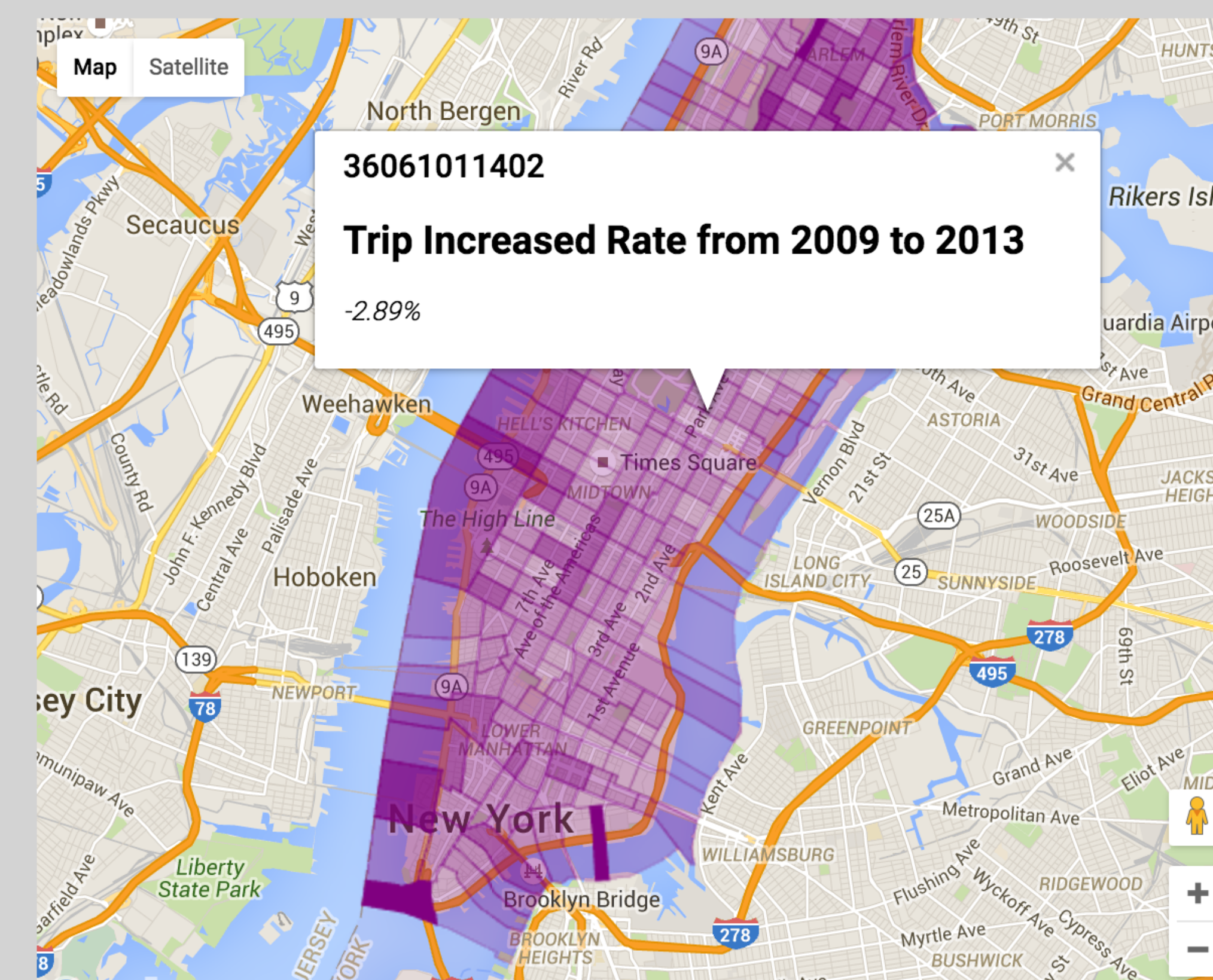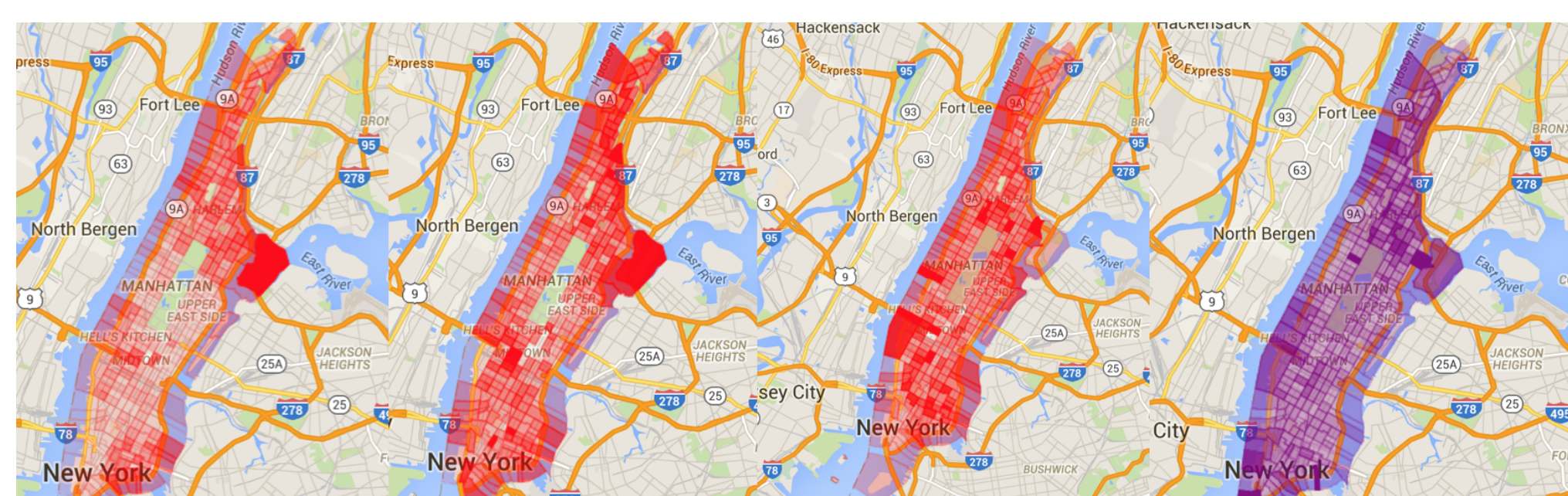

Figure 1: Snapshots Of Webpage

Figure 2: Different Economics Indicators VS Trips Data (Left -> Right: Poverty Proportion, Unemployment Proportion , Rent Increase and Trips Amount)

- For most regions, the more trips drop off amount, the higher Poverty & Unemployment ratio, vice versa. E.g. Upper East Side.
- This phenomenon indicates Gentrification, as taxi influx might indicate that middle/upper class people are moving into those lower income areas.
- Weekend Places/Tourist Attractions might also explain some trips amount, E.g. Wards Island.

Error Analysis

Some areas have low Poverty & Unemployment Ratio but high trip drop off amount. (e.g. Statue of Liberty)

We believe tourist attraction is a main reason that causes inaccuracy in Gentrification Detection using trips data.
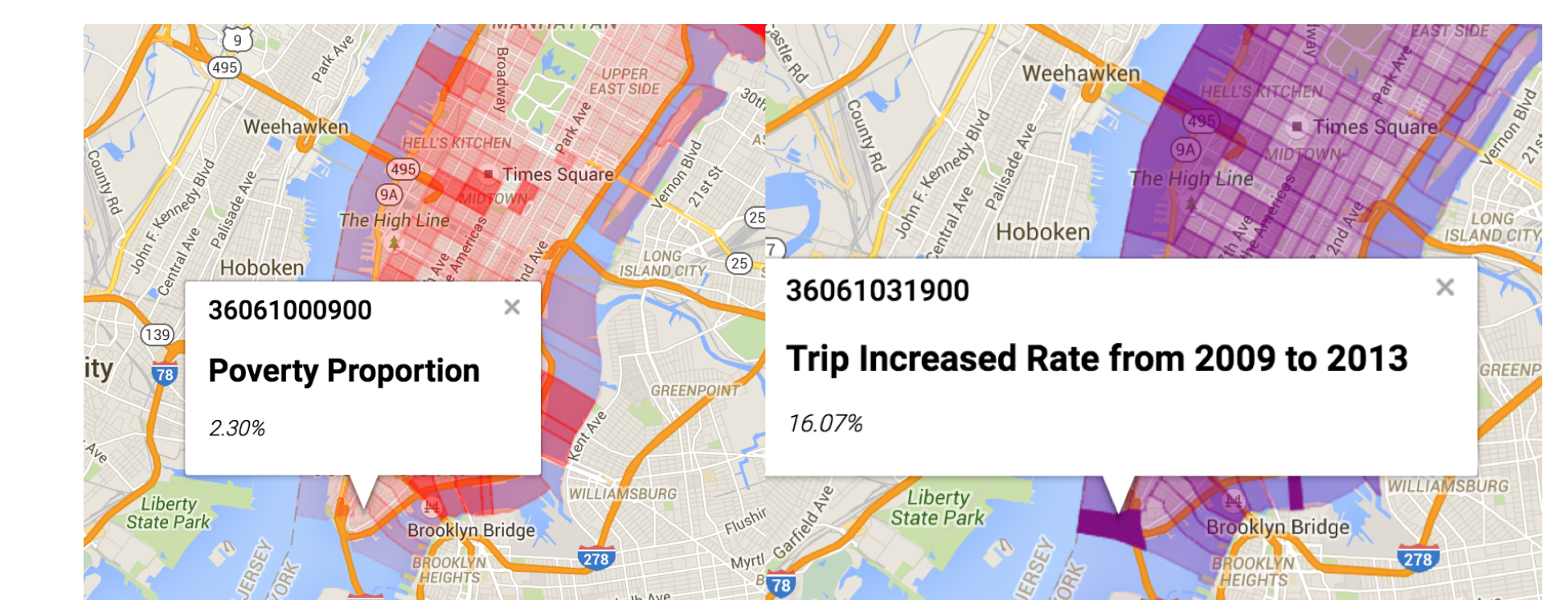

Figure 3: High Drop-off count with Low Poverty Rate

Also, after doing some research on Uber data, we believe Uber distorts the trips data distribution, which further decrease the correlation between Gentrification and Trips data.

Future Work

- How will Uber data affects our analysis? The popularity of Uber greatly changes trips distribution. We believe a complete analysis on Gentrification should both utilize Uber and Trips data.

- Is Manhattan data enough for Gentrification Detection ? It?s very possible that people from Manhattan are moving to other lower income neighborhood, as the income level in Manhattan is relatively high. To better detect this trend, we need to include other parts of the city.