Project Proposal

CIS 602-02 Data visualization

Duc Tu Luong - 01649122

2017

# Table of Content

# Datasets

In this project, I will analyze about Uber trips data in New York city compared to other FHV (for-hire vehicles) companies. The data contains over 4.5 million Uber pickups in New York City from April to September 2014, and 14.3 million more Uber pickups from January to June 2015. Data also contains 10 other for-hire vehicle (FHV) companies including Lyft company.

*Datasets URLs*

* **Uber TLC FOIL Response:** <https://github.com/fivethirtyeight/uber-tlc-foil-response>
* **NYC Taxi & Limousine Commission (TLC):** <http://www.nyc.gov/html/tlc/html/home/home.shtml>
* **Unified New York City Taxi and Uber data:** <https://github.com/toddwschneider/nyc-taxi-data>
* **New York map data:** <http://data.beta.nyc//dataset/3bf5fb73-edb5-4b05-bb29-7c95f4a727fc/resource/6df127b1-6d04-4bb7-b983-07402a2c3f90/download/f4129d9aa6dd4281bc98d0f701629b76nyczipcodetabulationareas.geo.json>

*Datasets types and semantics*

Data contains “Date/Time”, “Latitude”, “Longitude”, and base code for FHV based on TLC base company (<http://www.nyc.gov/html/tlc/html/industry/base_and_business.shtml)> of each pickup by Uber and other 10 FHV companies in New York City from April to September 2014, and from January to June 2015. The datasets are obtained from NYC TLC Company by submitting a Freedom of Information Law request on July 20, 2015 from FiveThirtyEight (<https://github.com/fivethirtyeight)>.

Due to passenger privacy concern recently, the latest data was 2 years ago, in 2015. Moreover, because the size of data is tremendous, I will try to visualize data as much as possible, to find out the success of Uber when it takes millions of rides from traditional taxis in New York, the place taxis are everywhere.

# Tasks List

Uber trending

* How many rides Uber have more each month/year?
* How many rides traditional taxis have less each month/year?

Density

* Which location has the highest/lowest density of Uber rides?
* Which day in week has the highest/lowest density of Uber rides?

Difference

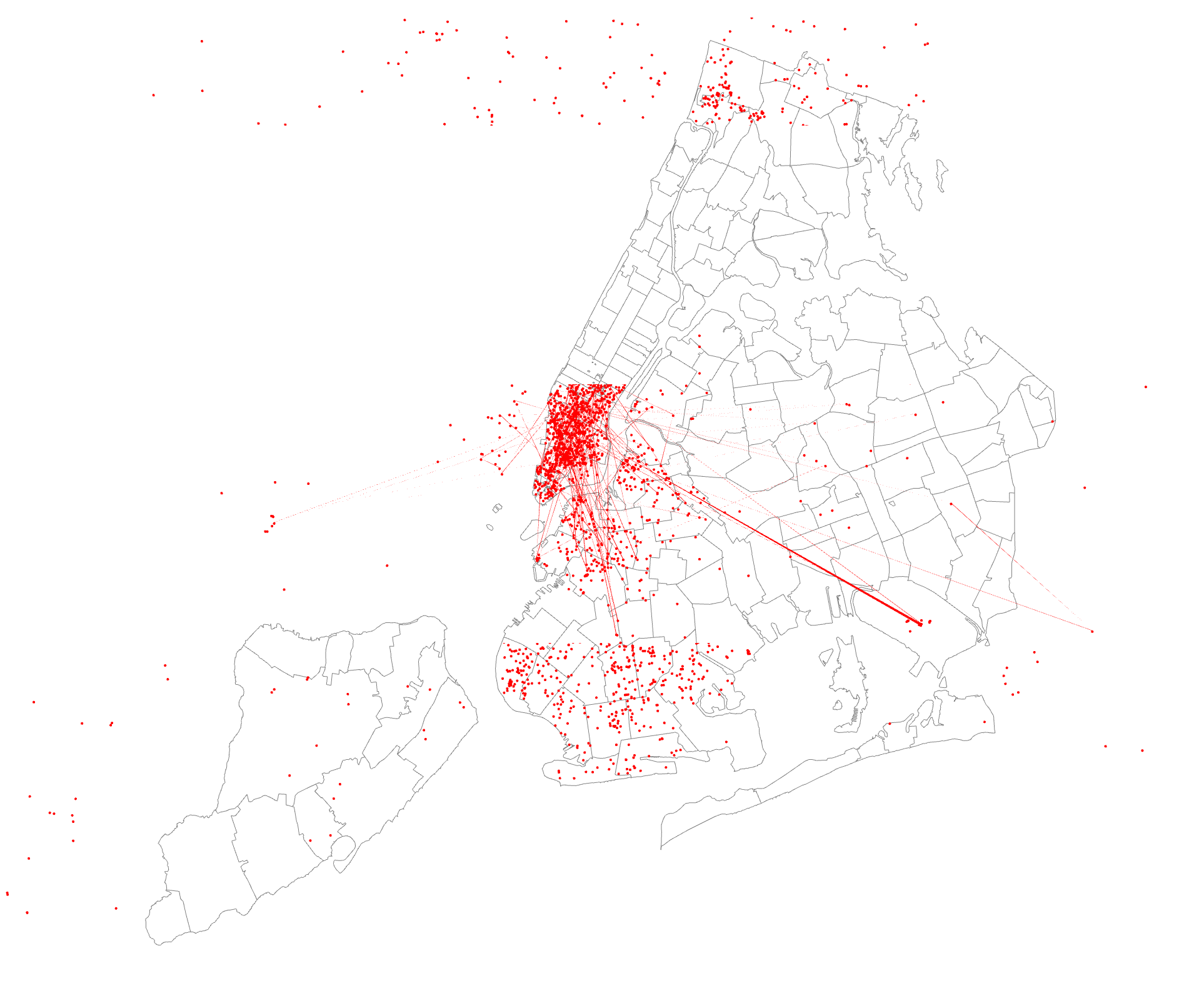
* Difference between taxis and Uber?
* Difference between weekdays and weekends ride density?

Other ideas

* Any change in Uber policy/pricing affects number of rides?
* Other factors such as weather, events, holidays affect Uber rides?

# Visualization Sketch

**Uber pickup locations density plot**



# Ideas and Requirements