CS6630 PROJECT(MILESTONE)

VISUALIZATION FOR TAXI RIDES IN NEW YORK CITY

Zheng Wang, u1208847, u1208847@utah.edu

Yishuai Du, u0884588, u0884588@utah.edu

CONTENTS

1. 1 BASIC INFORMATION

- Project Title: ***Visualization for Taxi Rides in New York City***

- *Group members:* **Zheng Wang, u1208847, u1208847@utah.ed**
  Yishuai Du, u0884588, u0884588@utah.edu

- Github: https://github.com/GregDobby/CS6630VisProj

1.2 BACKGROUND AND MOTIVATION

- Dating back to 1605, the horse-drawn hackney carriages can be the first form of taxi. With the development of technologies, horse-powered taxis were replaced by electric-powered taxis which were also quickly superseded by gas-powered taxis in 1899 in Paris. The advent of the first gas-powered taxi in New York City in 1907 laid the foundation of the largest taxi market in the United States. Since then, the taxi industry has become a critical component of the transportation infrastructure in large urban areas. Over 800 million consumers are transported annually in the United States, generating $23 billion revenues.
- However, the attention to taxi market regulation was scant until the first momentum of taxi industry in late 1920s, during which price war between taxi companies built threats to the society stability. Unfortunately, similar situation seems to happen again these days. Ride-sharing companies, namely Uber and Lyft, offer consumers rides with a lower price and less waiting time, while the incumbent and established taxi companies remain basically the same. And the established companies are doing everything possible to get regulators to help them by stopping new entry into their market. The increasingly tense situation requires regulation on ride-sharing cars as well as, more importantly, exploring more efficient regulation in taxi industry.

1.2 BACKGROUND AND MOTIVATION

- New entry into taxi market is only an external trigger to alteration of old-fashioned regulation. In fact, the demand for a more efficient regulation has always existed as a consequence of inherent characteristics of taxi industry: highly fragmented service, fixed two-part tariff fare regulation and spatial misallocation. A large number of small taxi companies, sometimes of a single owner-operator, result in highly fragmented taxi service. Many non-owner taxi drivers who lease cars from taxi companies has to pay a fixed leasing cost and their own gas and insurance. Since the fixed fare rate regulated by local municipalities, drivers have to increase their chances to get more consumers to increase their income, which indirectly causes spatial misallocation: densely populated area like downtown tends to be over-supplied and sparsely populated area like suburban district tends to be undersupplied.
- Our project tends to provide a visualization of taxi rides in New York City as a useful tool to analyze the relationship between supply and demand for a specific period in taxi market and help people adjust the regulation. More specifically, the distribution of people in New York City can vary according time in one day, date(weekday, weekend and festivals), weather and some other factors, so the demand can always change. By predicting the demand and adjusting the supply and change fare regulation, it can both reduce consumers' waiting time and increase the total revenues, which is a win-win situation.

1.3 PROJECT OBJECTIVE

- As the largest taxi industry, there are lots of taxis compete for consumers by driving to different locations around the New York city. Due to price regulation and search frictions, leaving empty taxis in some areas, and leaving excess demand in other areas as a big problem. The spatial mismatch and the waiting time of taxi make reduce lots of daily income. So in our project, we will use a statistical model of search and matching between taxis and consumers under regulation to optimize this problem.

2.1DATA COLLECTION

Data collection are as follows:

TLC Trip Record Data in New York City:

http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

2014-2015 Uber Rides in New York City:

https://www.kaggle.com/fivethirtyeight/uber-pickups-in-new-york-city

NYC Zoning GIS Data:

https://data.cityofnewyork.us/City-Government/Zoning-GIS-Data-Shapefile/kdig-pewd

Taxi Zone Lookup Table

http://www.nyc.gov/html/exit-page.html?url=https://s3.amazonaws.com/nyc-tlc/misc/taxi+_zone_lookup.csv

Taxi Zone Shapefile

http://www.nyc.gov/html/exit-page.html?url=https://s3.amazonaws.com/nyc-tlc/misc/taxi_zones.zip

2.2 DATA CLEANING UP

- Since there two columns (pick up location and drop off location) were differences after 2015. Before 2015, all locations are longitude and latitude, and after 2015 all locations are index as a specific zone. So we transform all longitude and latitude to index as a specific zone.

- We also remove some columns we do not use that.

- [pickup_datetime, pickup_location, dropoff_datetime, dropoff_location, trip_distance, fare_amount , mta_tax, tolls_amount, tip_amount , extra total_amount]

2.3 DATA INTEGRATION

- TLC Trip Record Data in New York City provides us with information about taxi rides in New York City. More specifically, it provides pick up time, drop off time, transportation fare, pick up location and destination. However, to do a useful visualization, we also need demand information such as consumers' waiting time and supply information such as taxis' waiting time and cruising area. We will use estimated the data according to information provided.

- (Uber…)

2.3 DATA INTEGRATION

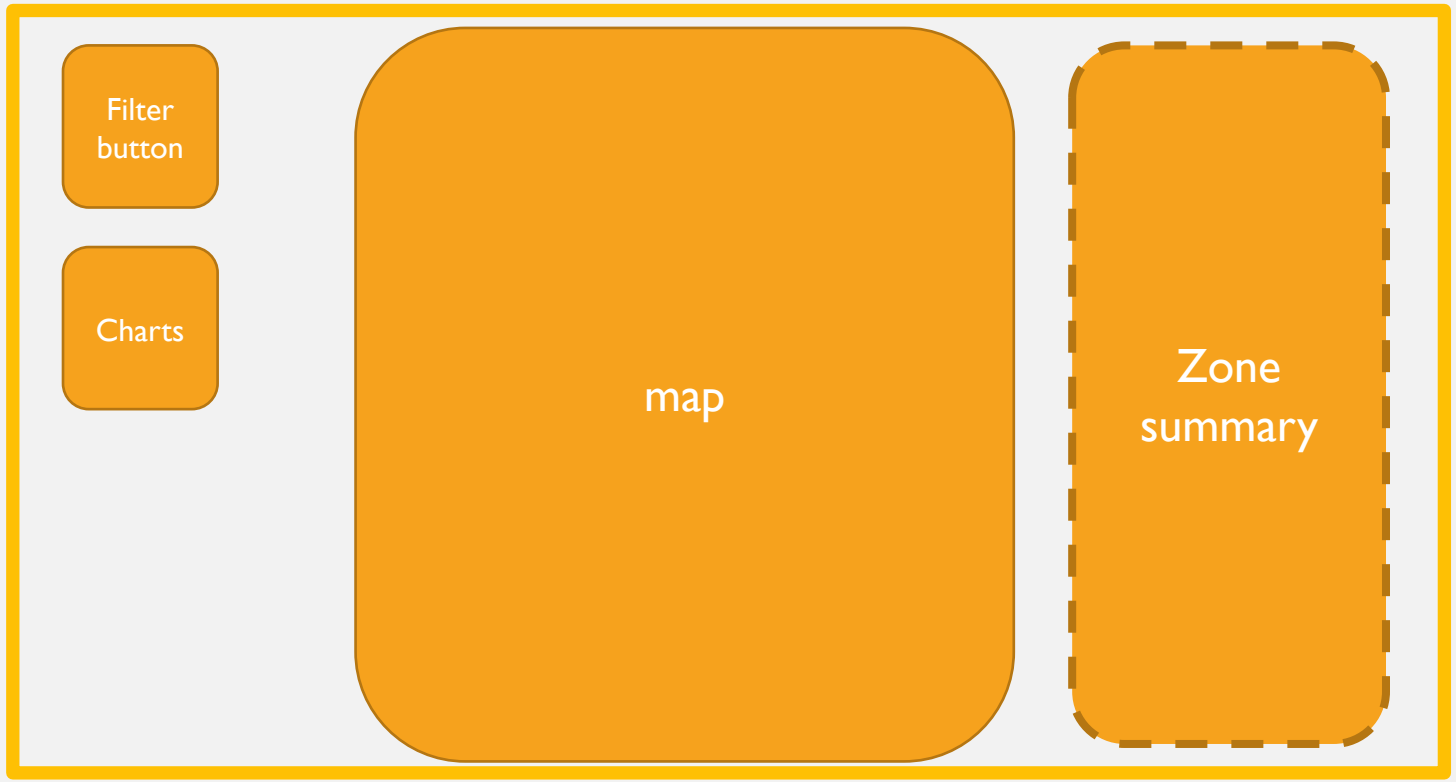
- **Map data:**
  - The geoJson of New York City without any modification.
- **Taxi Market:**
  - Estimate the number of active taxis by counting different taxis at different times in a day, reflecting the development of taxi market
  - Avenues in days, months and years
- **Trip information**
  - Calculate average trip time and total fare from a certain zone to another at different times, providing useful information to consumers
- **Demand & Supply by math algorithm**
  - Estimate taxis' searching time and consumers' waiting time[reference], reflecting spatial misallocation

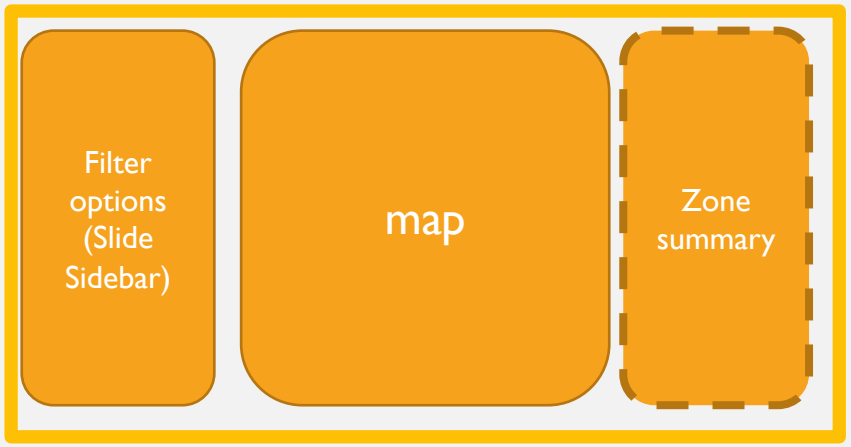
3.1 DESCRIPTION

-taxi_rides	
|-data	data has geo data, clean-up data
|-src	
|- css	css is to store all color schemes
|-js	js is to store all coding
|- index.html	index.html is the final display
|-py	py is to store coding for clean-up dataset

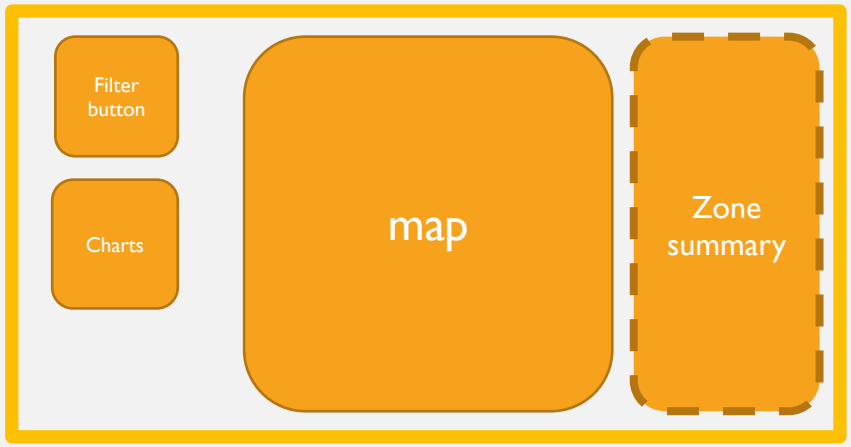
3.2 PROTOTYPE

The final look should be like this:

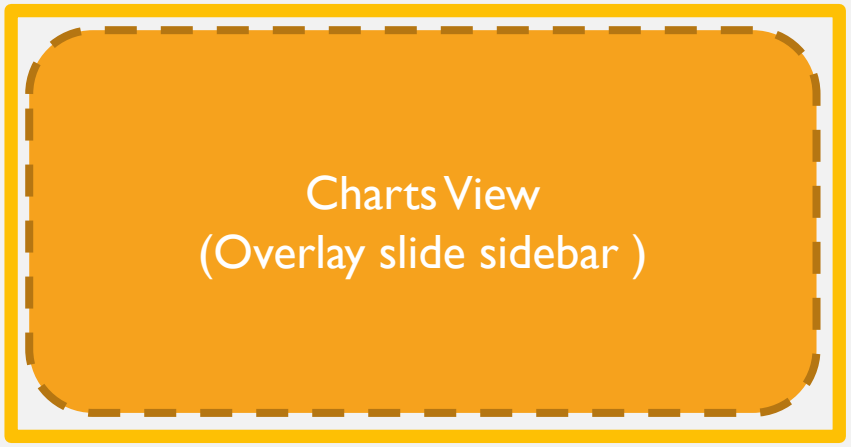




Select Filter options first

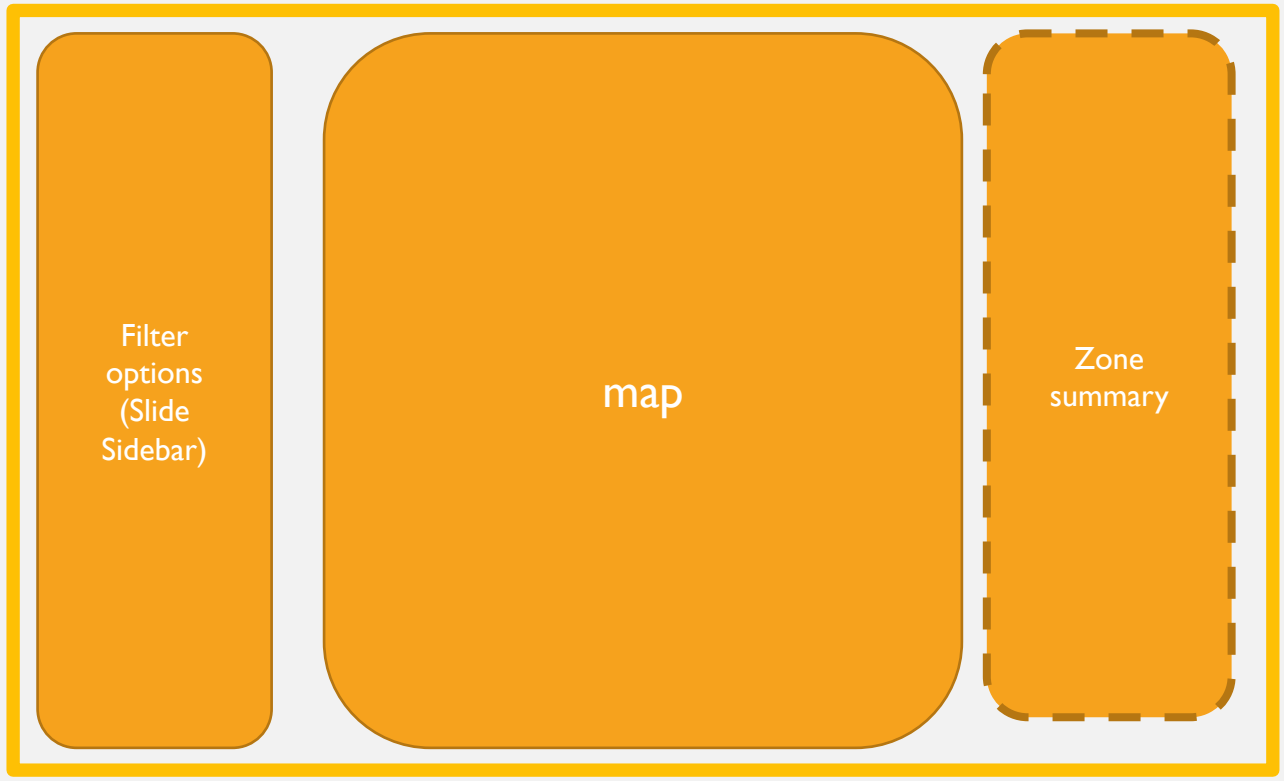


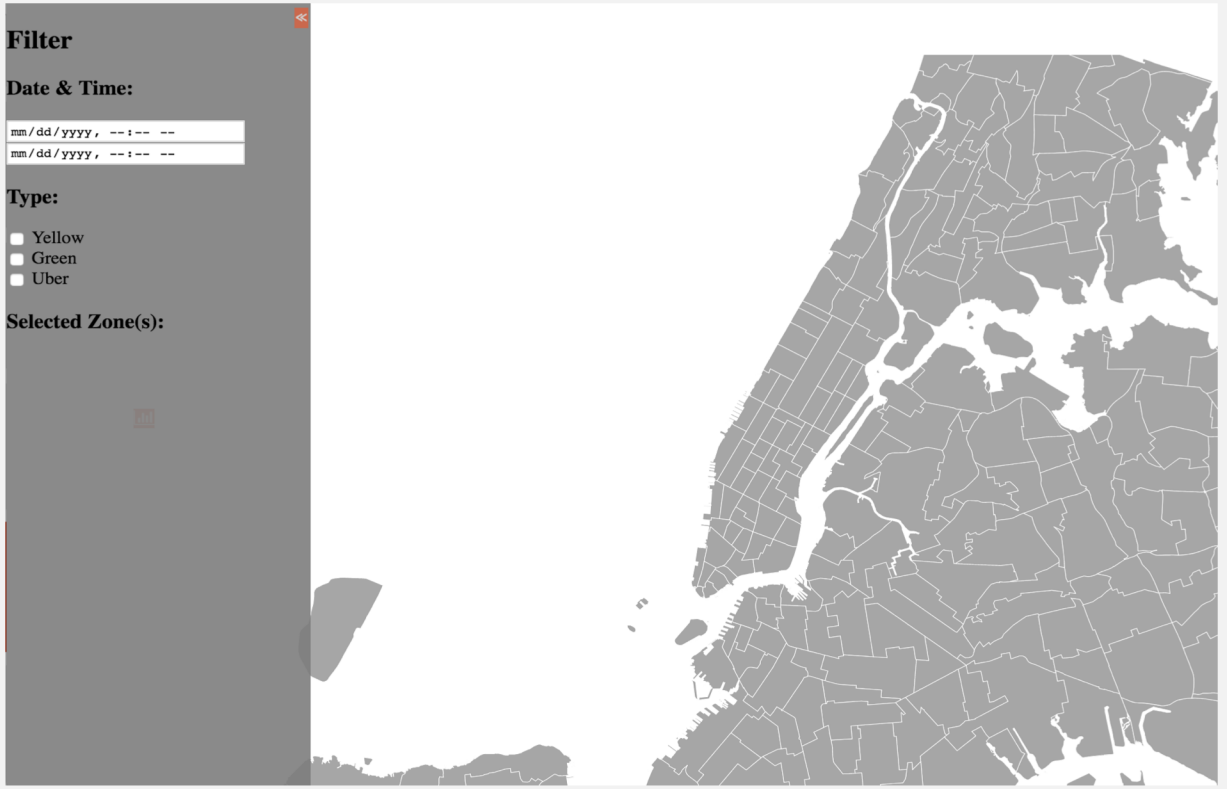
If you only want to know **one** zone summary, just have a mouseover on map. The summary will show on Zone summary.

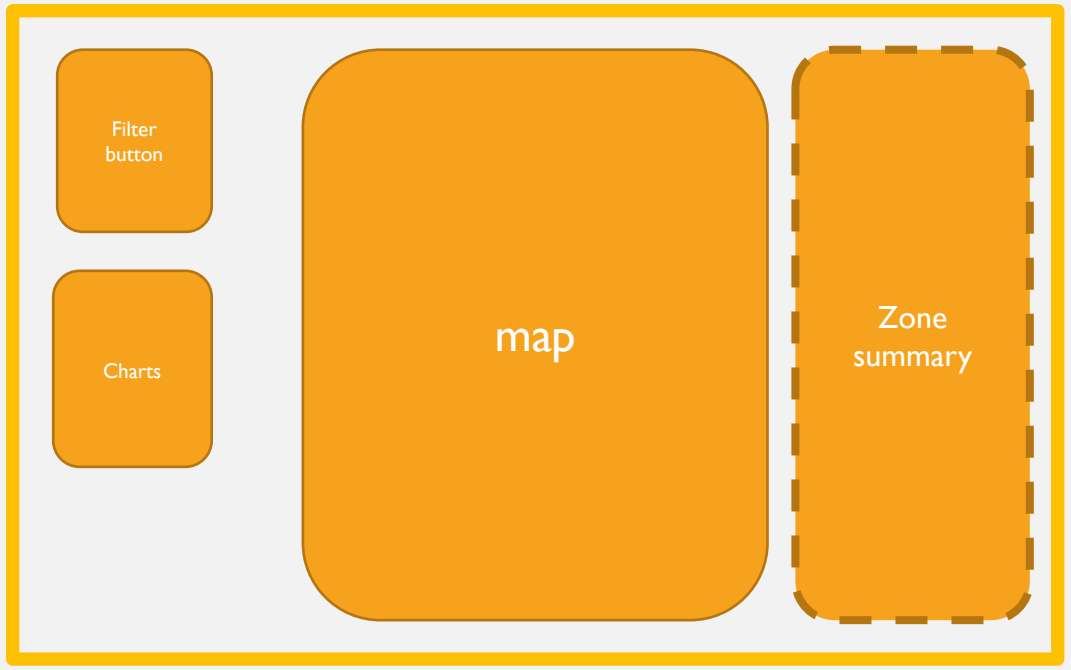


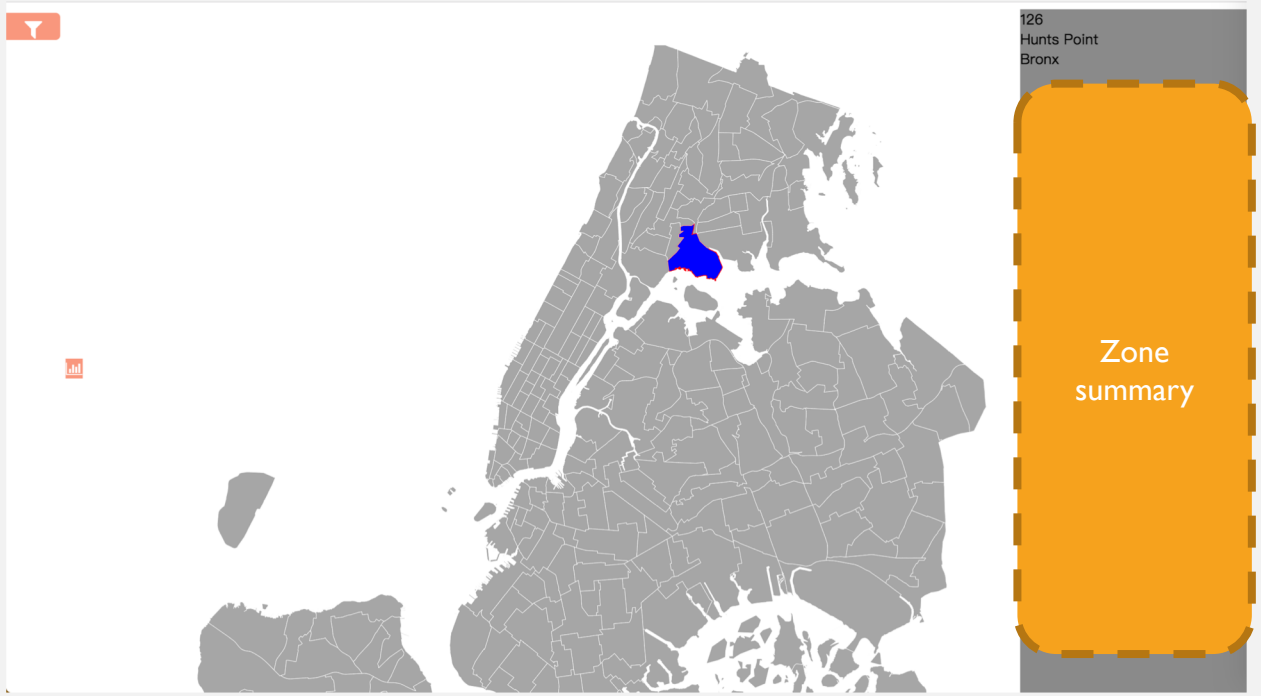
If you only want to know **Two or more** zones summary, click all of them on map. Then click Charts, all summaries will show on Chars View.

4.1 OVERVIEW

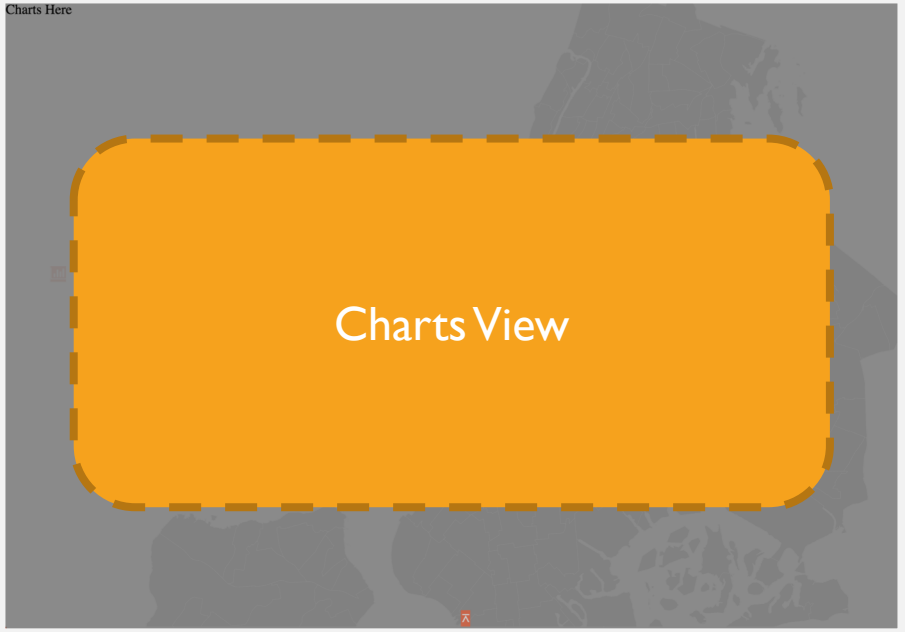

Filter options (Slide Sidebar)
map
Zone summary


Filter
Date & Time:
mm/dd/yyyy, --:-- --
mm/dd/yyyy, --:-- --
Type:
Yellow
Green
Uber
Selected Zone(s):


Filter button
Charts
map
Zone summary


126
Hunts Point
Bronx
Zone summary

Charts View
(Overlay slide sidebar)



RELATED WORK

- *Spatial Equilibrium, Search Frictions and Dynamic Efficiency in the Taxi Industry* **[12/2017]**

EVALUATION

- We hope everyone can have fun with this project. Our project combines a math algorithm and visualization. We want to use these two technology help government , drivers and costumers have an optimal way to deal with taxi problem in the New York, such as how to solve the taxi of demand and supply in different zone, how to assign work for each taxi, and so on. The process of data is hard time, we have 200gb data for that. We need to clean up and transform them to have a same form. And also need to do lots of statically computing. Finally, we hope more people can seek taxi easier in the New York city future.