Mobility change in NYC with MTA turnstile data

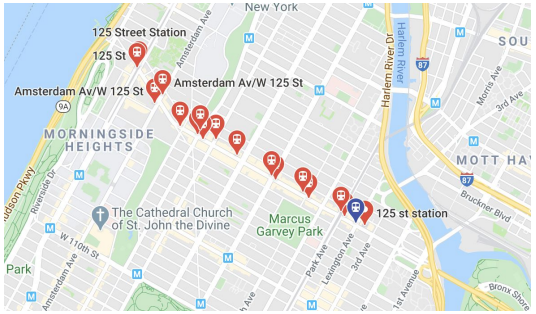
Problems & Approaches

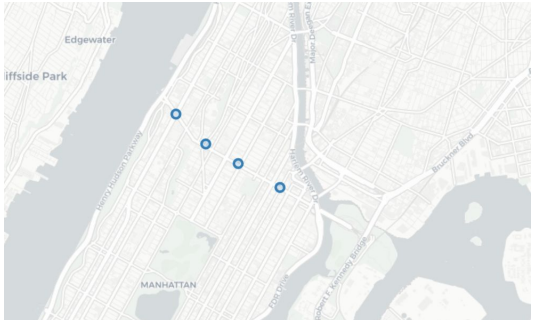
- How does mobility change over time? How does stat-at-home policies affect the trends?
  - compare with last year and previous month
- Does mobility change show different patterns in different areas?
  - look at aggregations of the data over zip codes or neighborhoods
  - the flow to and from work
  - see if patterns change in key bridges and tunnels
- How does subway ridership correlate to other matrices such as confirmed cases?
- How do people feel about the mobility change because of Covid-19?

Data Cleaning

MTA Turnstile Data http://web.mta.info/developers/turnstile.html

- Accumulative counters
  - Negative counters
  - Resets
  - Jumping counters, …
- Stations
  - Matching station and location data
  - Station definition: station name + lines



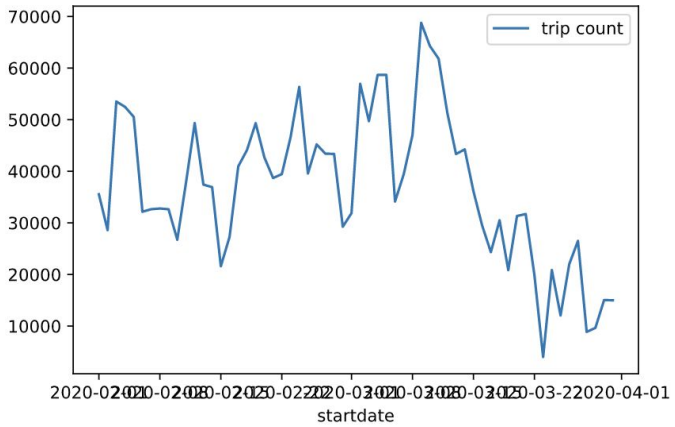


Other traffic datasets

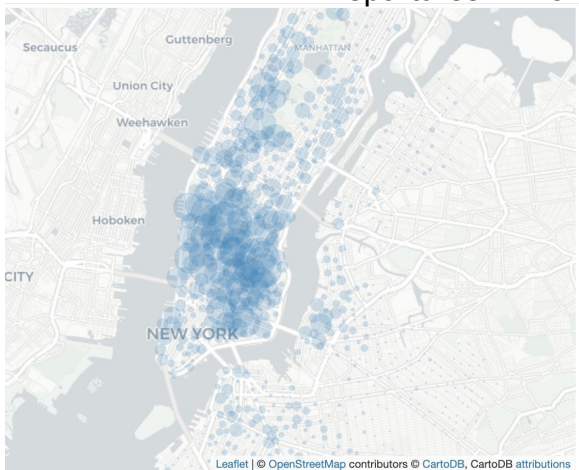
CitiBike Trip History in February and March

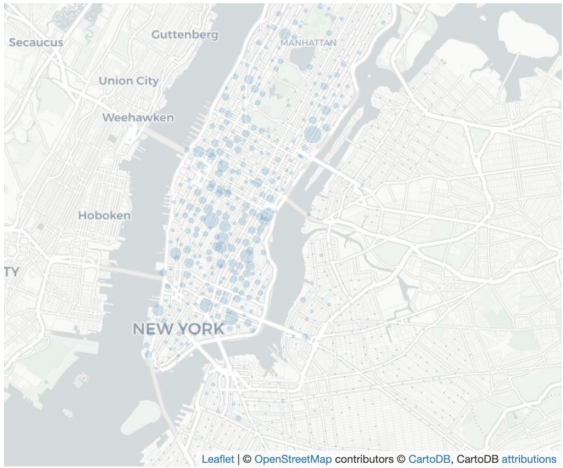
https://www.citibikenyc.com/system-data

- Number of trip has dropped
- Hard to define normal pattern



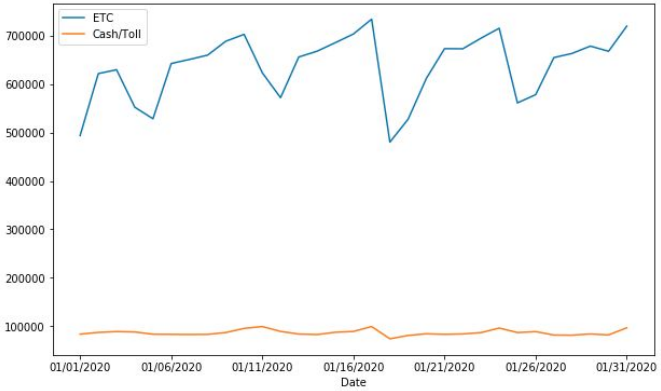
Departures in March 1 and. March 25

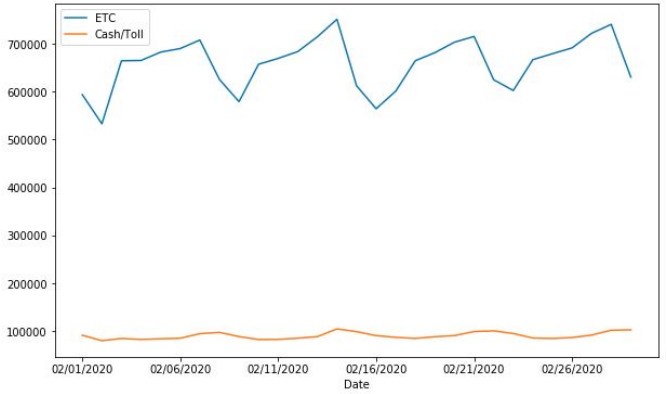


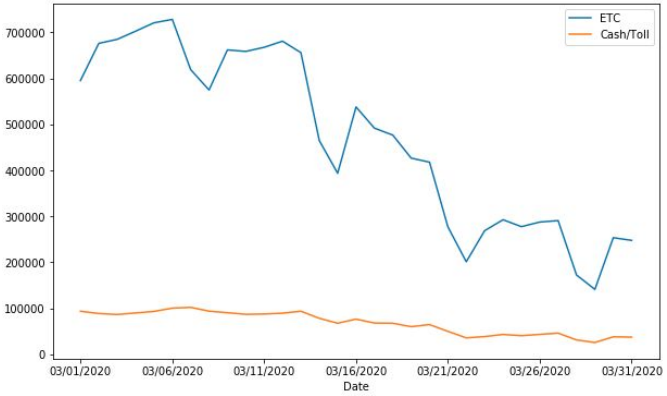


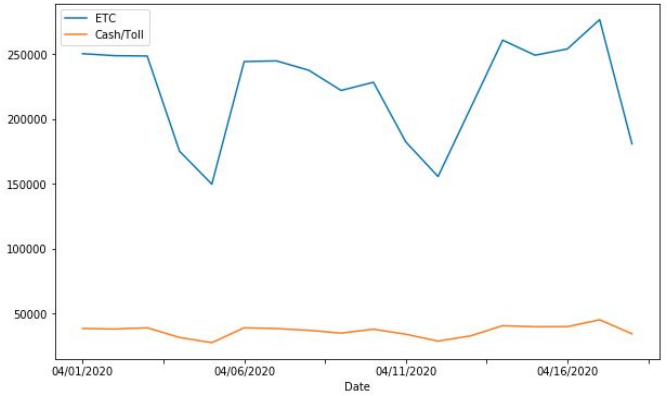
Bridges and Tunnels Traffic

- Compare NYC bridges and tunnels Traffic between Jan 2020 to Apr 2020
- Around mid-March the traffic starts to plummet which is consistent with covid outbreak
- Maintains at a low level in Apr









Subway Tweets Analysis

- Positive and neutral tweets are the majority
- #Aline, #Fline, etc are popular hashtags
- Incident reports tends to have more retweets
- No pattern to people's emotions over time

Turnstile data Analysis

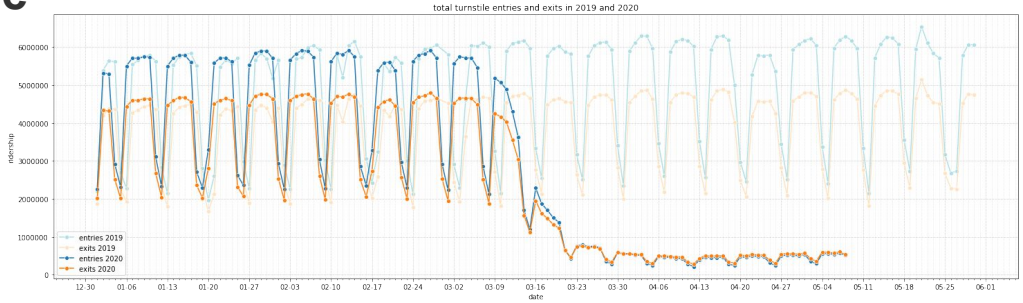
1. Overall ridership change over time
2. Ridership change by station and location
3. Ridership change by time-slot
4. Ridership by spatial distribution

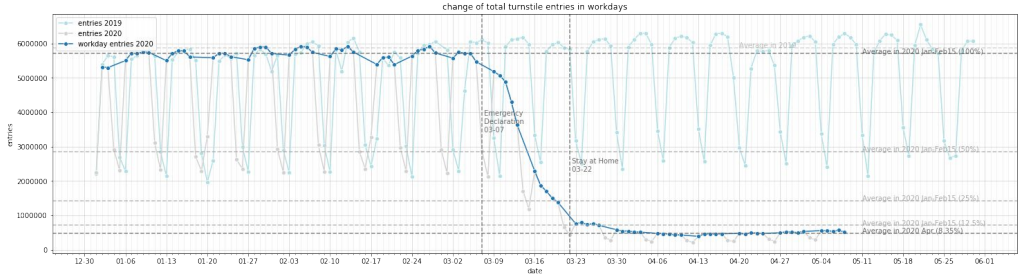
1. Overall ridership change over time

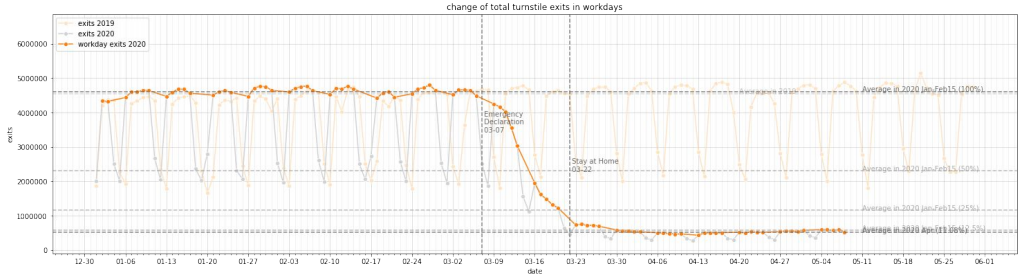
Normal days

- Entries always higher than exits
- Workday much higher than weekends and holiday

Can take 2020 Jan-Feb15 as normal pattern





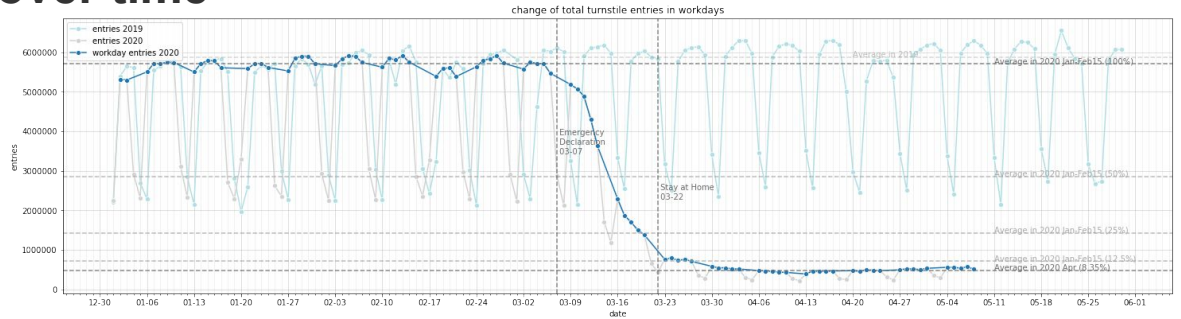


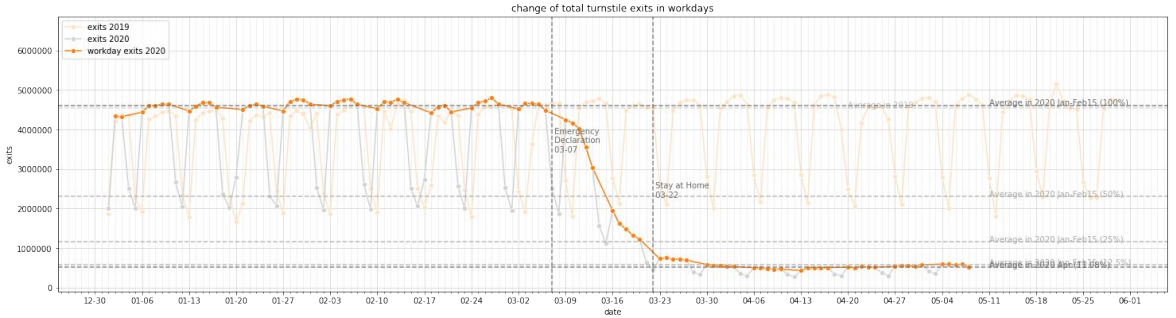
1. Overall ridership change over time

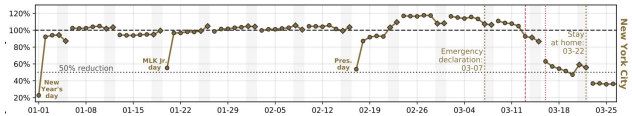
Quantify the change

- Entries dropped to 8.35%
- Exits dropped to 11.8%

Compare with other metrics







percent of daily commute
Brennan Klein et al.
[Assessing changes in commuting and individual mobility in major metropolitan areas in the United States during the COVID-19 outbreak]

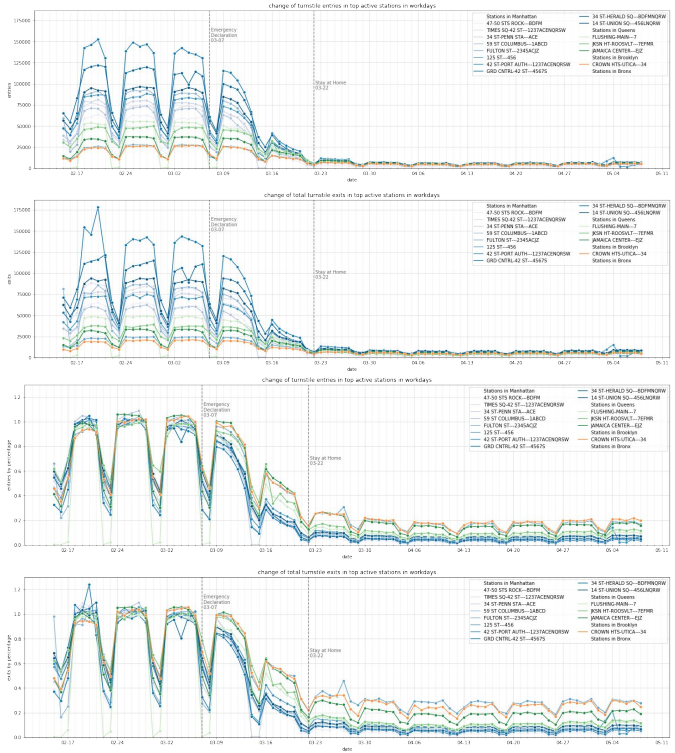
2. Ridership change by station and location

New top 10 stations in April 2020

- JAMAICA CENTER(EJZ) Queens
- CROWN HTS-UTICA(34) Brooklyn
- JKSN HT-ROOSVLT(7EFMR) Queens
- 125 ST(456) Manhattan
- FLUSING_MAIN(7) Queen

By percentage of change, Manhattan stations dropped the most, but..

- Regular ridership in Manhattan stations are way higher
- Top April station in Manhattan (14 ST-UNION SQ)
- Outlier: 125 ST (456) dropped to about 30%



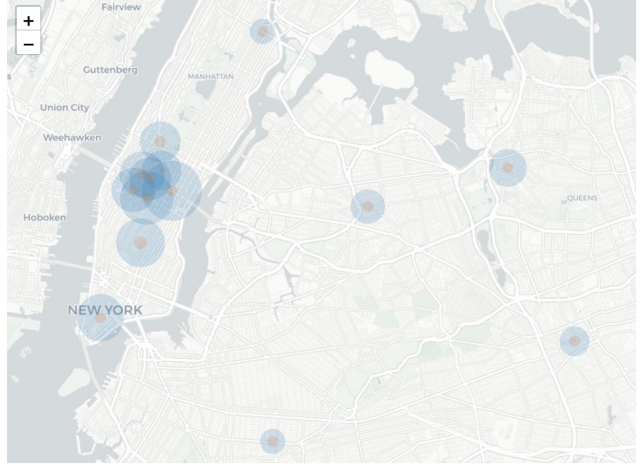
2. Ridership change by station and location

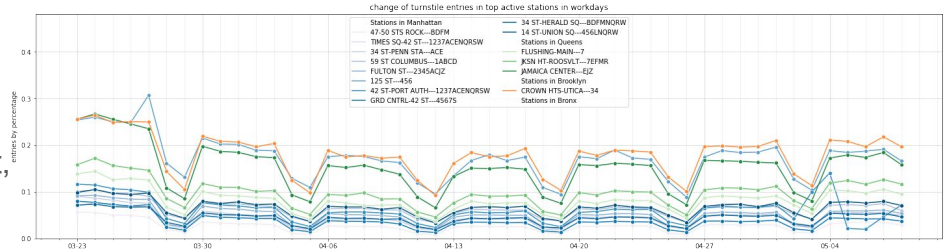
New top 10 stations in April 2020

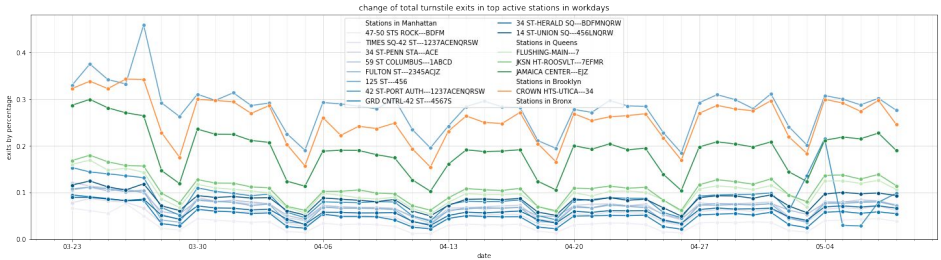
- JAMAICA CENTER(EJZ) Queens
- CROWN HTS-UTICA(34) Brooklyn
- JKSN HT-ROOSVLT(7EFMR) Queens
- 125 ST(456) Manhattan
- FLUSING_MAIN(7) Queen

By percentage of change,Manhattan stations dropped the most, but..

- Regular ridership in Manhattan stations are way higher
- Top April station in Manhattan (14 ST-UNION SQ)
- Outlier: 125 ST (456) dropped to about 30%







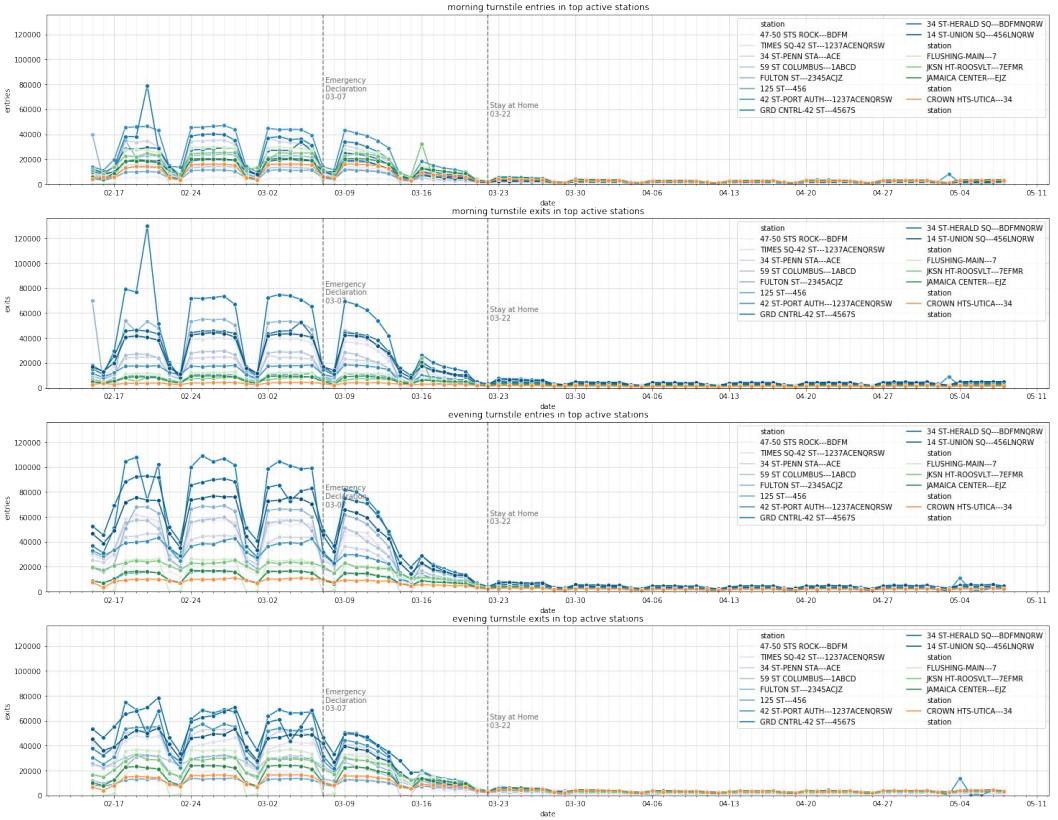
3. Ridership change by time-slot

4 hour time interval

- Morning: before 12 p.m.
- Evening: after 12 p.m.

The flow of commuters

- From elsewhere to big stations in Manhattan.
- After the outbreak, Manhattan stations are still the top target.
- But number of commuters in Manhattan may dropped the most.



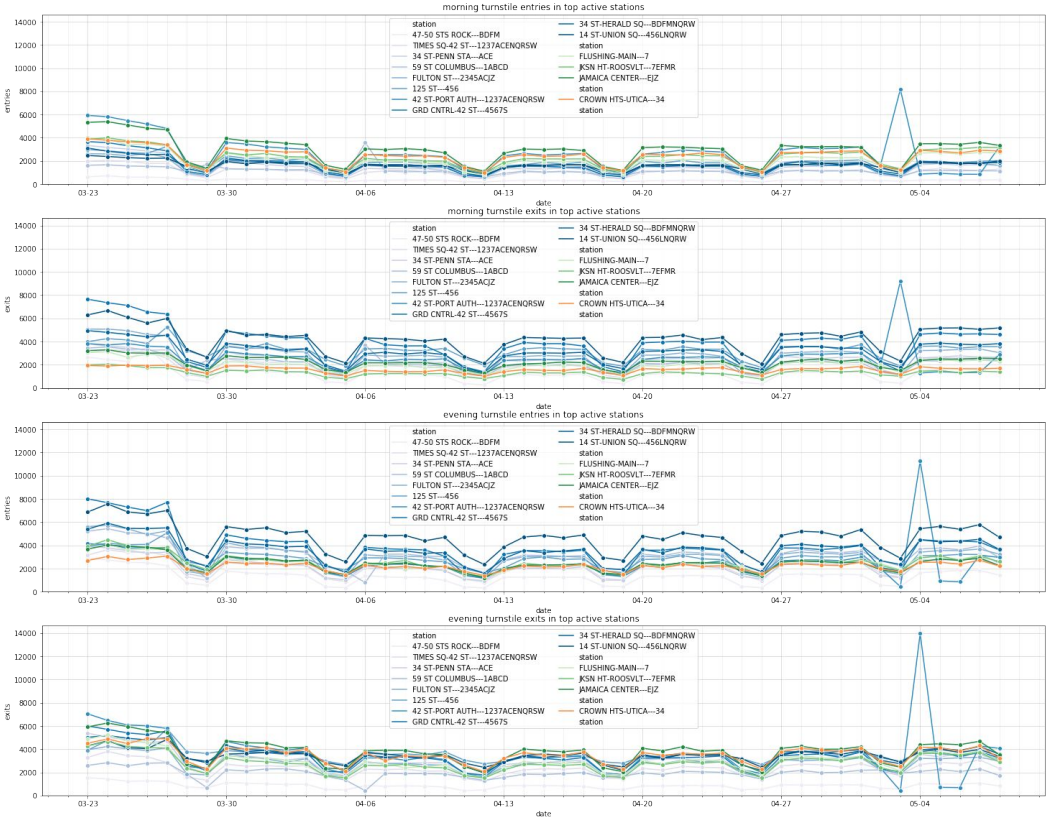
3. Ridership change by time-slot

4 hour time interval

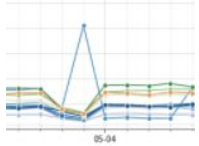
- Morning: before 12 p.m.
- Evening: after 12 p.m.

The flow of commuters

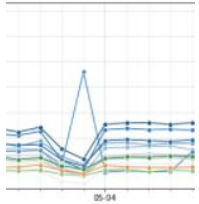
- From elsewhere to big stations in Manhattan.
- After the outbreak, Manhattan stations are still the top target.
- But number of commuters in Manhattan may dropped the most.



Brief discussion on previous sections

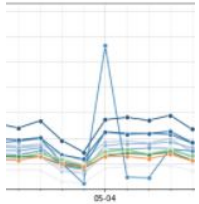


1. Overall ridership change over time
2. Ridership change by station and location
3. Ridership change by time-slot



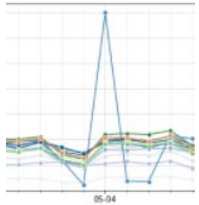
Pros:

- Better data cleaning, comparable with other quantified studies
- Flow to and from work



Limitations:

- Data quality: How to explain outliers?
- Limitation of turnstile dataset
- Static plots: Use more interactive demonstration in the future?



4. Ridership by Spatial Distribution - Methodology

Spatial Aggregation: Quadtree Spatial Index

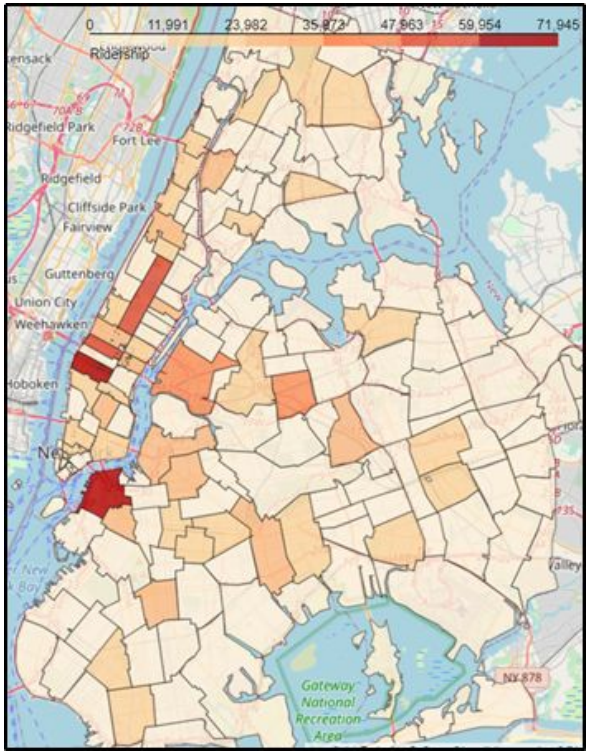
To generate the distribution of ridership based on zip codes, apply spatial index to accelerate aggregation

Quadtree

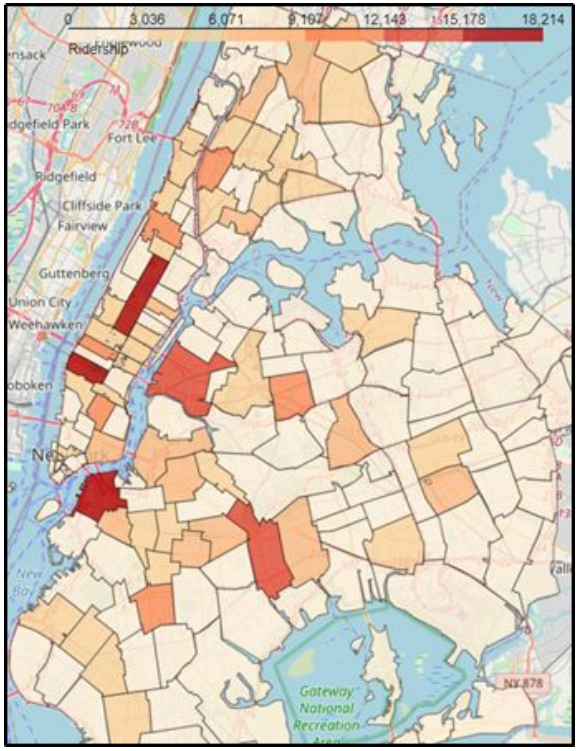
- Represents 2D planes
- Each internal node has 4 children
- Each children represents a quadrant of sub-plane
- Iterate through an array row-by-row
- If all 4 sub-regions exhibit desired property, merge into single node

4. Ridership by Spatial Distribution - Overview

02 / 19 Ridership Map



03 / 18 Ridership Map



04 / 22 Ridership Map

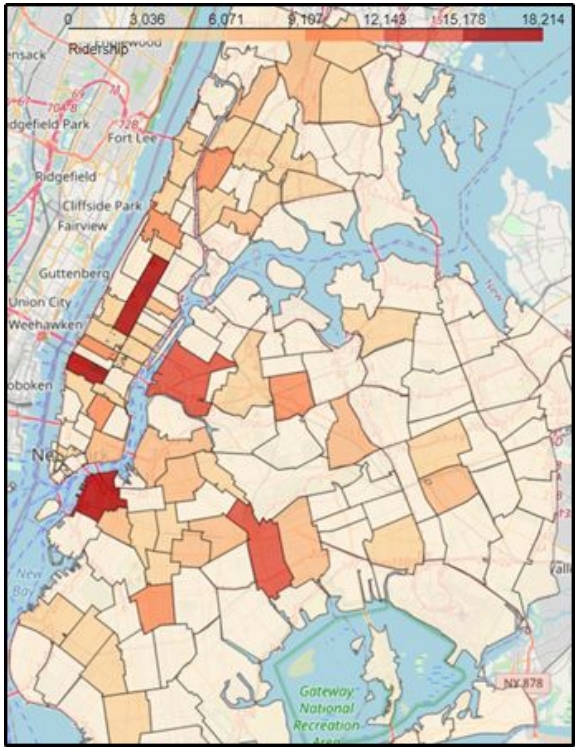
4. Ridership by Spatial Distribution - Founds

Similar Features

- Depend on the layout of subway lines
- Highest ridership in midtown Manhattan
- Higher ridership in regions surrounding Manhattan

Trends during epidemics

- Huge decrease of ridership, especially in Manhattan
- Ridership in different regions become more recognizable
- Increased proportion of ridership in inland regions



04 / 22 Ridership Map

4. Ridership by Spatial Distribution - Possible Influential Factors

Possible Factors

- Population
- Positive Rate
- Regional Income

Conducting Correlation Matrix

Positive ON

- Populations in Each Region

Slightly Negative ON

- Positive rate
- Regional Income
- Positive Cases Per Thousand People