

Mobility change in NYC with MTA turnstile data

Problems & Approaches

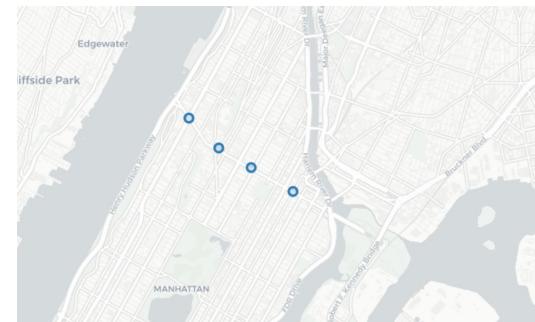
- How does mobility change over time? How does stat-at-home policies affect the trends?
 - compare with last year and previous month
- Does mobility change show different patterns in different areas?
 - look at aggregations of the data over zip codes or neighborhoods
 - the flow to and from work
 - see 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>

Readings of cumulative entries and exits count of subway turnstiles every 4 hour

- Accumulative counters
 - Negative counters
 - Resets
 - Jumping counters, ...
- Stations
 - Matching station and location data
 - Matching station name
 - Data from individual project
 - Manual repair
 - Station definition: station name + lines
 - Control area, remote unit, division...

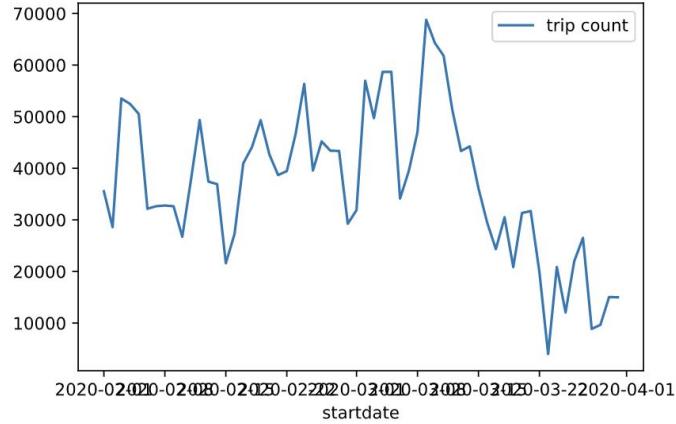


Other traffic datasets

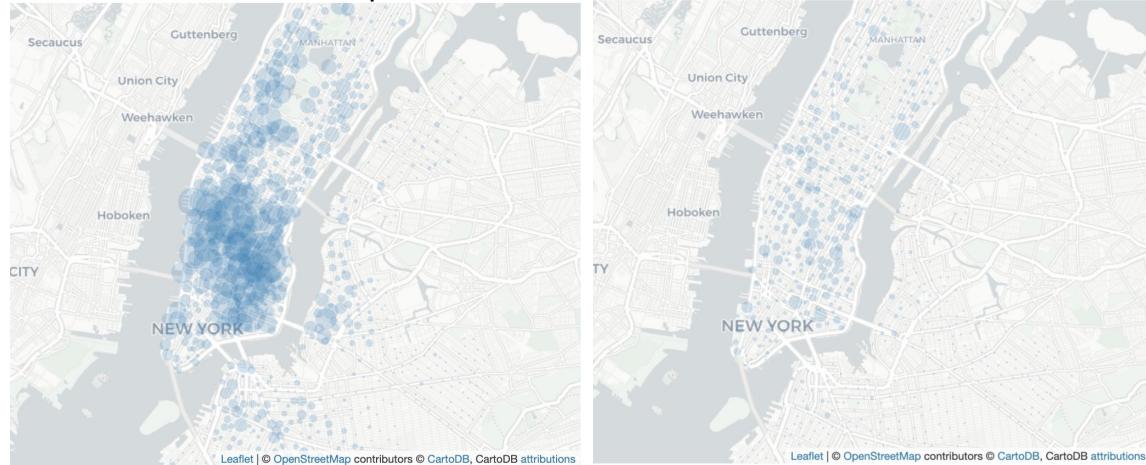
CitiBike Trip History in February and March

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

- Number of trip has dropped
- Most stations in Manhattan
- Hard to define normal pattern

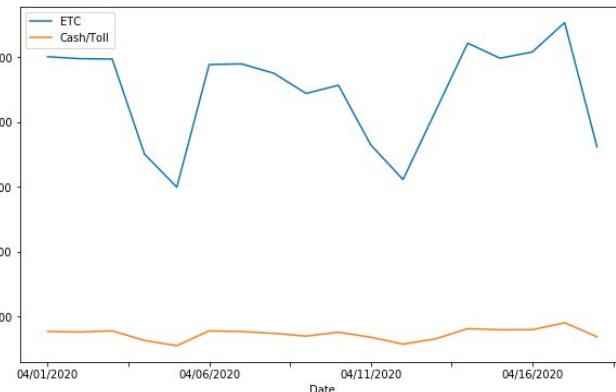
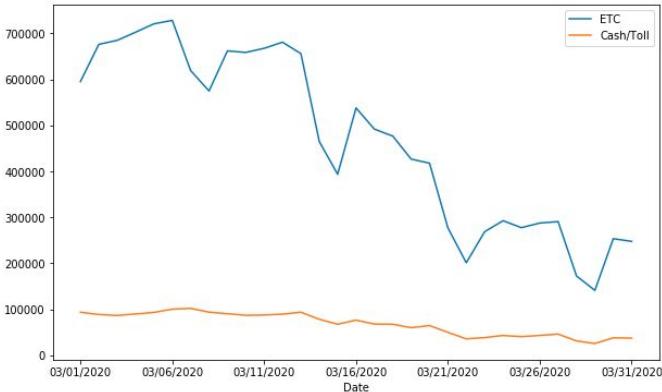
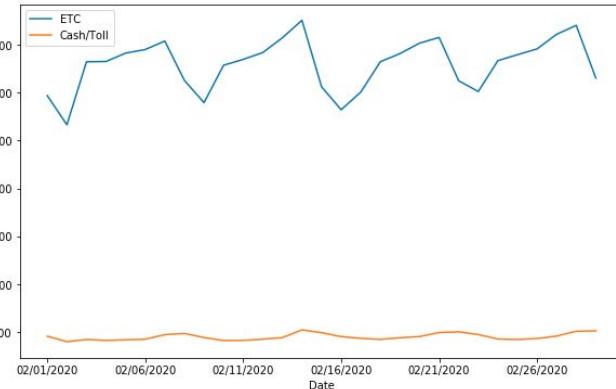
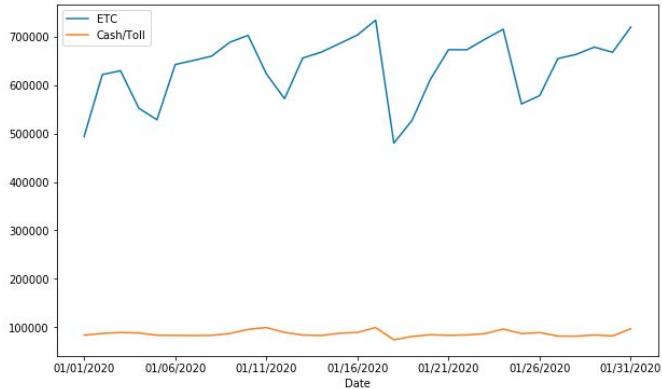


Departures in March 1 and March 25



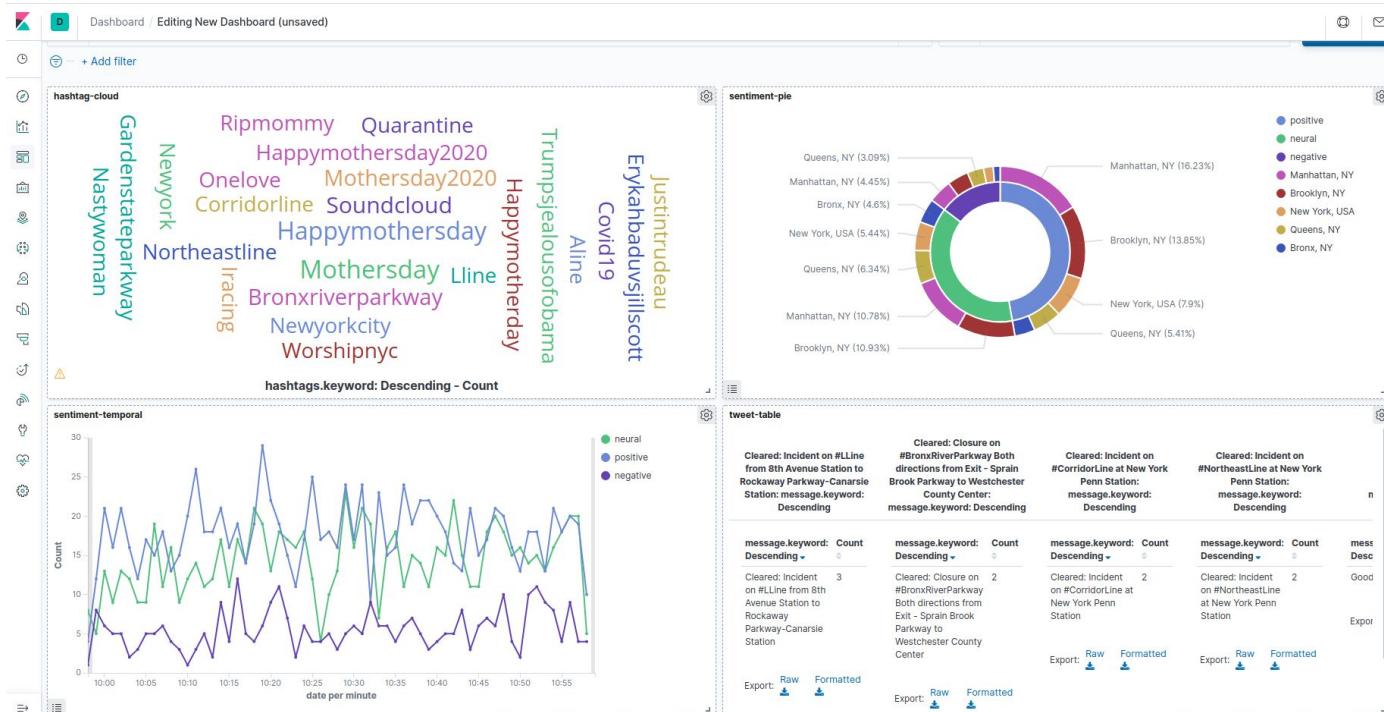
Bridges and Tunnels Traffic

- 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

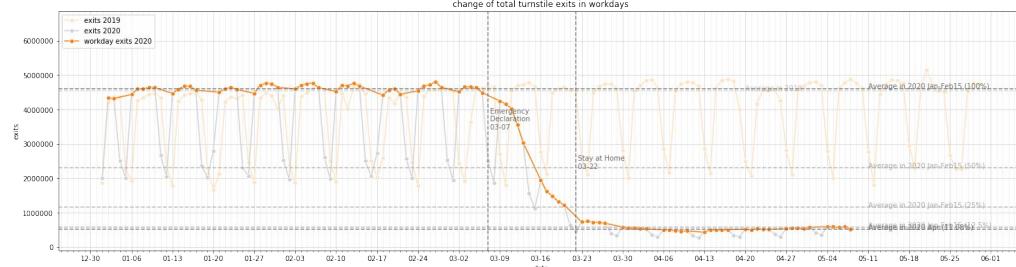
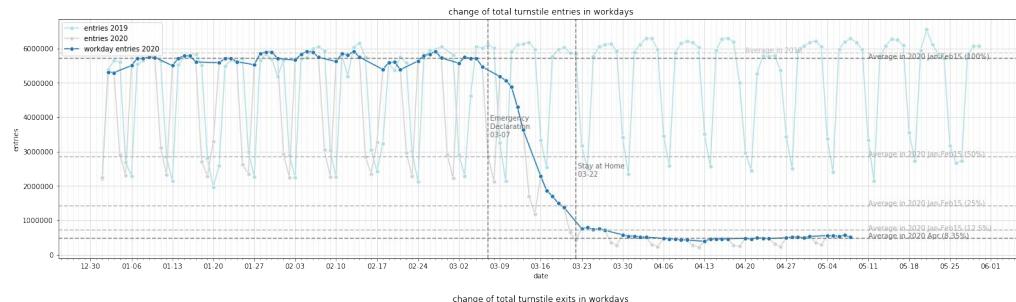
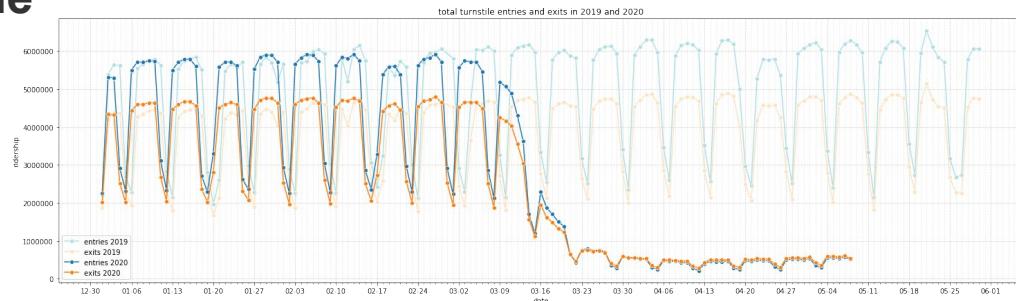
1. Overall ridership change over time
2. Ridership change by station and location
3. Ridership change by time-slot
4. Ridership by spatial distribution

1. Overall ridership change over time

Normal days 2019

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

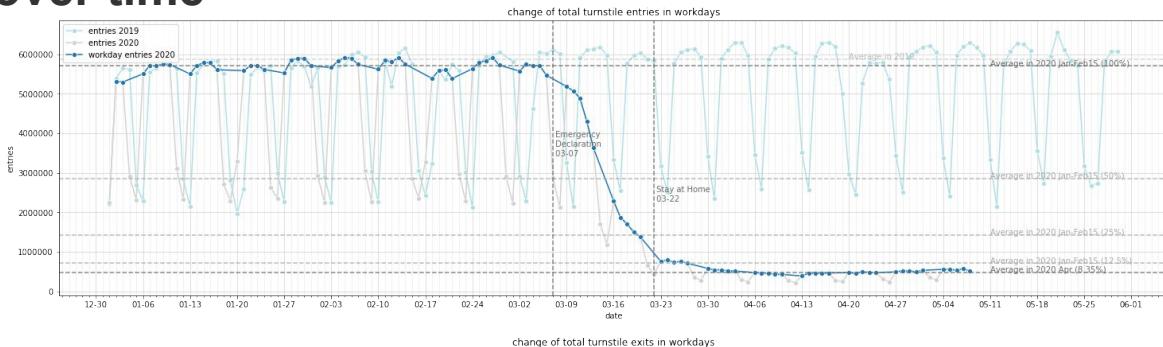
Can take 2020 Jan-Feb15 as normal pattern



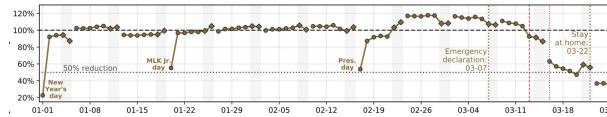
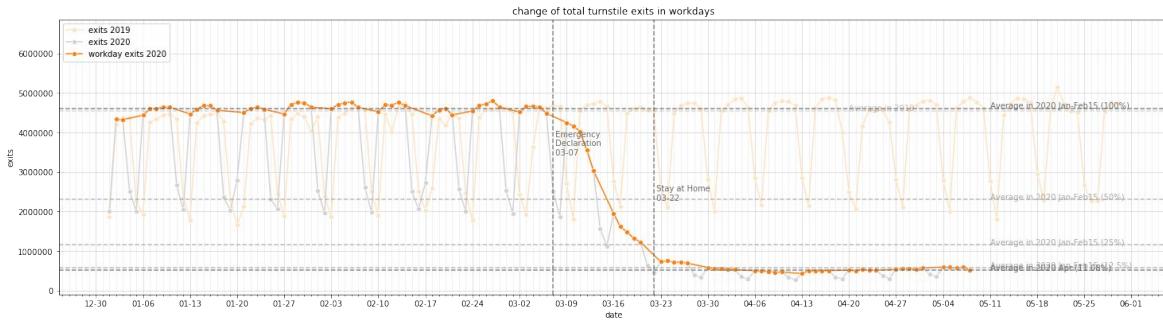
1. Overall ridership change over time

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

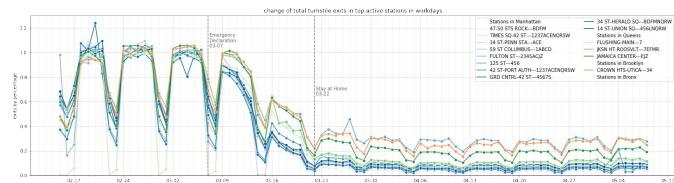
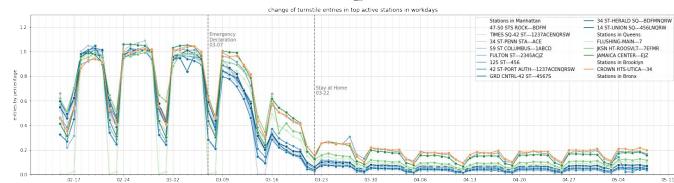
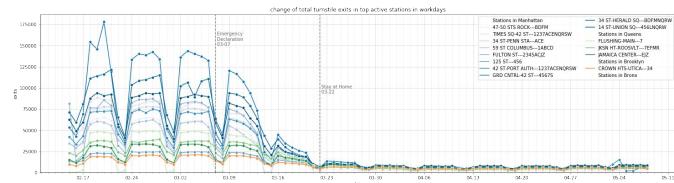
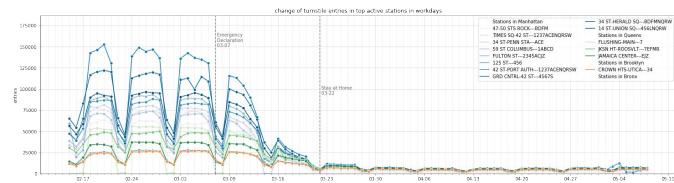
New top 10 stations in April 2020

- Jamaica center(EJZ) Queens
- Crown Hts Utica Av(34) Brooklyn
- Jackson Hts Roosevelt Av(7EFMR) Queens
- 125 street station(456) Manhattan
- Flushing Main St(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%

STATION	lines	borough	april_avg_entries
47-50 STS ROCK	BDFM	Manhattan	1685.636364
TIMES SQ-42 ST	1237ACENQRSW	Manhattan	3288.136364
34 ST-PENN STA	ACE	Manhattan	3594.818182
59 ST COLUMBUS	1ABCD	Manhattan	4384.045455
FLUSHING-MAIN	7	Queens	4433.636364
FULTON ST	2345ACJZ	Manhattan	4723.409091
125 ST	456	Manhattan	4766.590909
JKSN HT-ROOSVLT	7EFMR	Queens	4772.909091
CROWN HTS-UTICA	34	Brooklyn	4810.818182
42 ST-PORT AUTH	1237ACENQRSW	Manhattan	5197.090909
GRD CNTRL-42 ST	4567S	Manhattan	5329.227273
34 ST-HERALD SQ	BDFMNQRW	Manhattan	5360.227273
JAMAICA CENTER	EJZ	Queens	5564.409091
14 ST-UNION SQ	456LNQRW	Manhattan	6485.000000



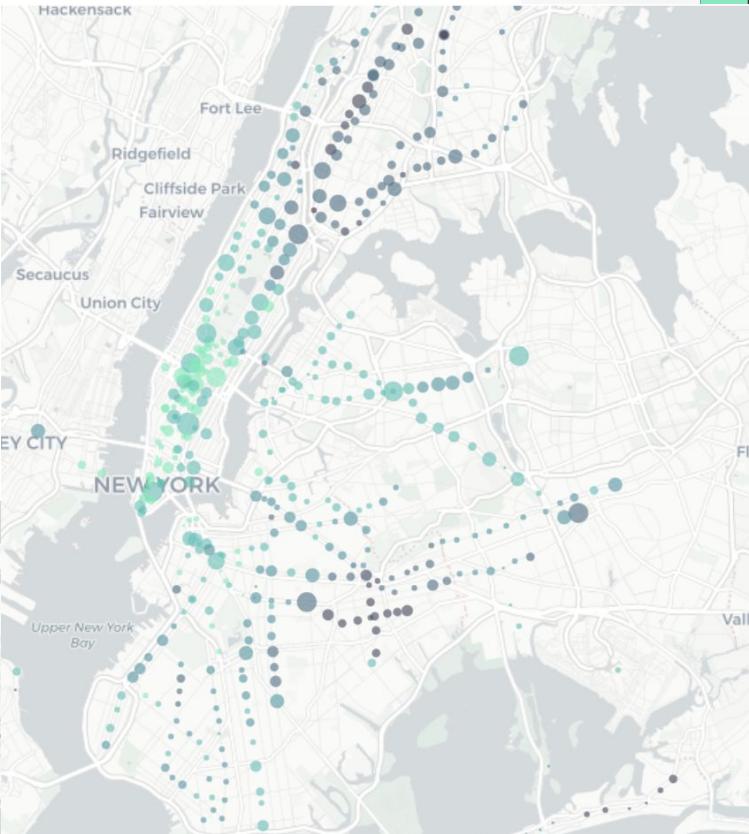
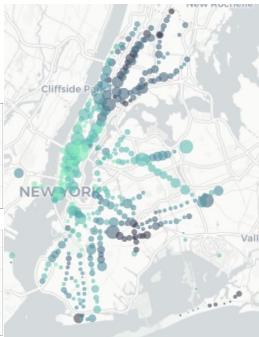
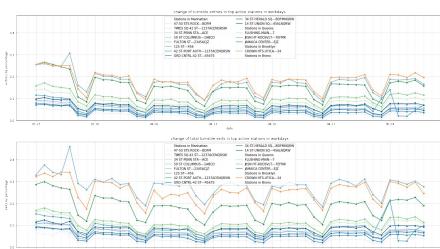
2. Ridership change by station and location

New top 10 stations in April 2020

- Jamaica center(EJZ) Queens
- Crown Hts Utica Av(34) Brooklyn
- Jackson Hts Roosevelt Av(7EFMR) Queens
- 125 street station(456) Manhattan
- Flushing Main St(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%



```
if changed<0.75:  
    color='#302433'  
elif changed<0.8:  
    color='#41455C'  
elif changed<0.85:  
    color='#446C84'  
elif changed<0.9:  
    color='#3B96A3'  
elif changed<0.95:  
    color='#3CC1B5'  
else:  
    color='#62ECB8'
```

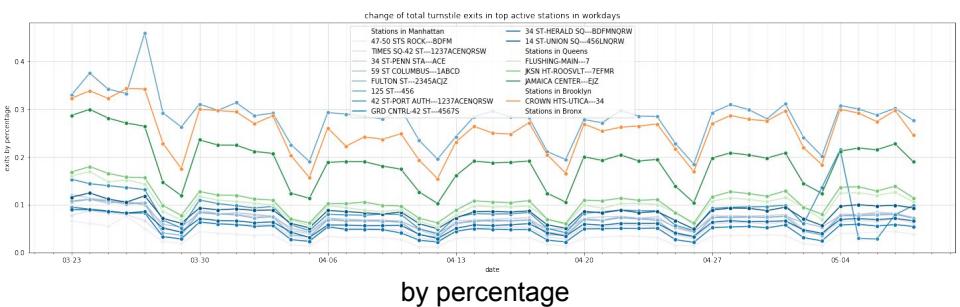
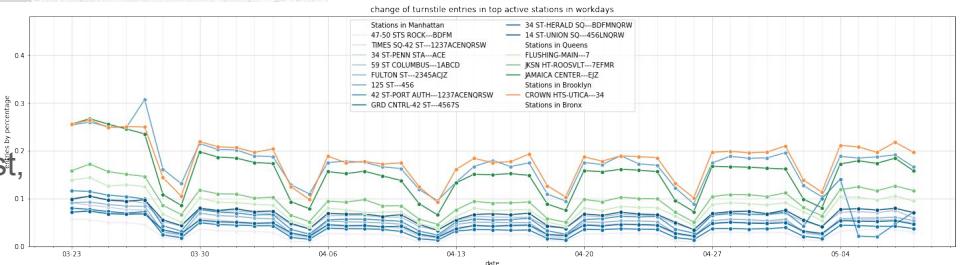
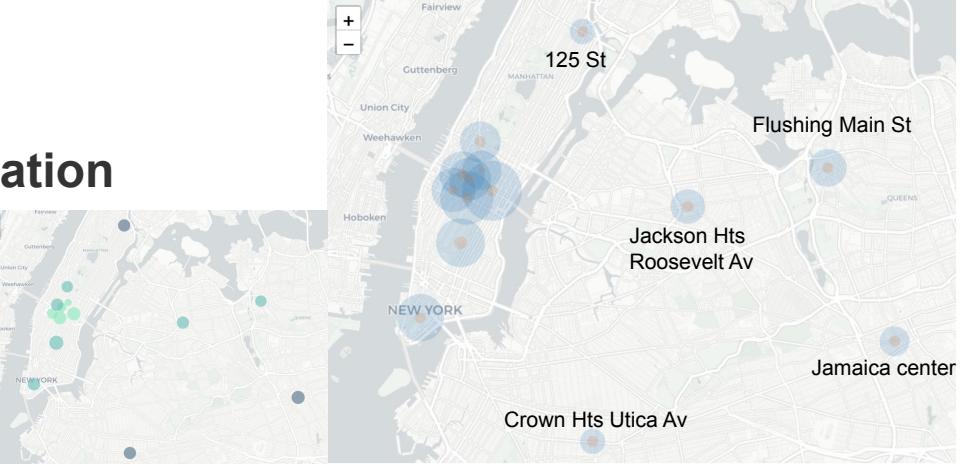
2. Ridership change by station and location

New top 10 stations in April 2020

- Jamaica center(EJZ) Queens
- Crown Hts Utica Av(34) Brooklyn
- Jackson Hts Roosevelt Av(7EFMR) Queens
- 125 street station(456) Manhattan
- Flushing Main St(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)
- 125 ST (456) dropped to about 30%



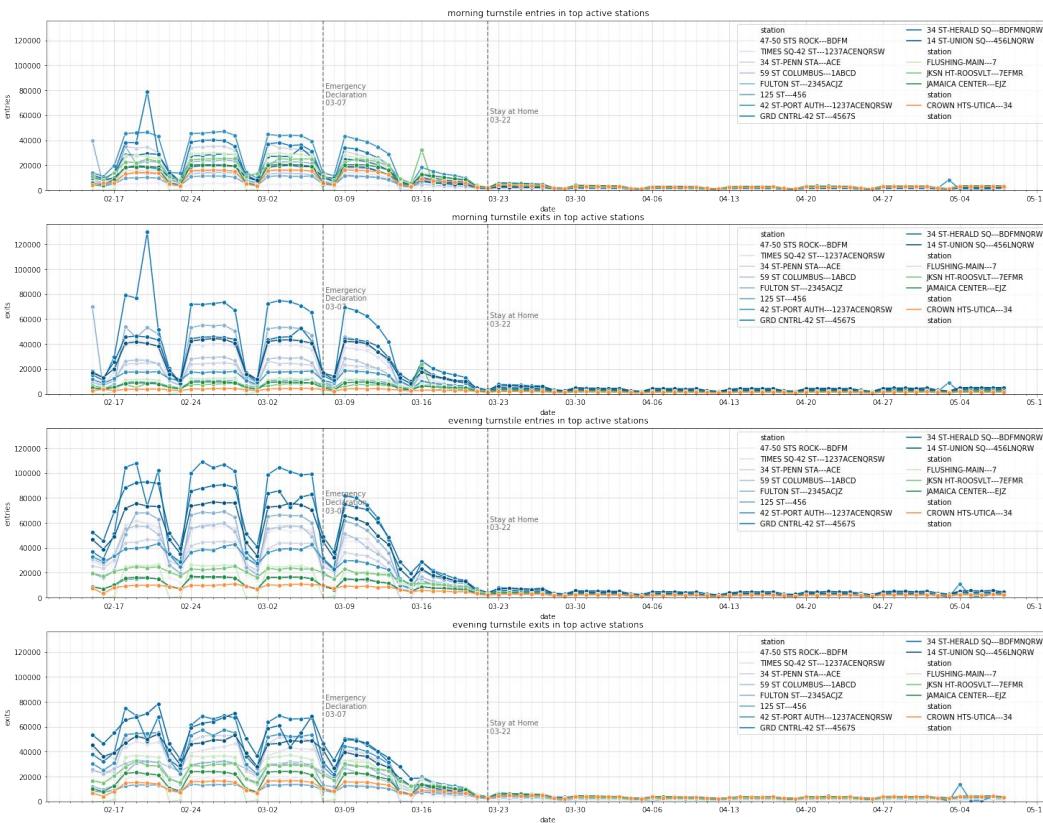
3. Ridership change by time-slot

Turnstile readings by 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.
 - Measuring individual commuter's mobilitchage?



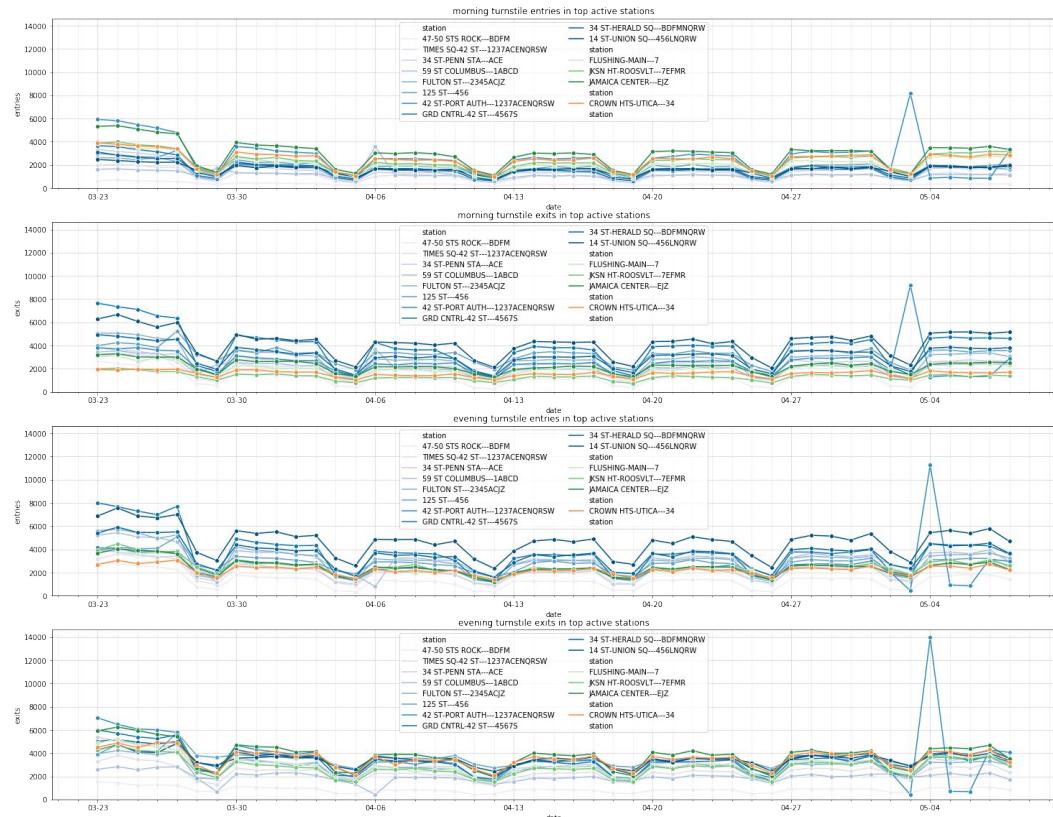
3. Ridership change by time-slot

Turnstile readings by 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.
- Measuring individual commuter's mobilitichage: may be underestimating neighborhood inequality if only use whole day change



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.

Size: avg morning entries in April

Color: change percentage



top regular morning exits:			
STATION	lines	borough	regular_avg_exits_morning
GRD CNTRL-42 ST	4567S	Manhattan	73430.870968
FULTON ST	2345ACJZ	Manhattan	52597.258065
34 ST-HERALD SQ	BDFMNRQW	Manhattan	44987.354839
47-50 STS ROCK	BDFM	Manhattan	43230.387097
14 ST-UNION SQ	456LNQRW	Manhattan	42185.193548
TIMES SQ-42 ST	1237ACENQRSW	Manhattan	39242.225806
LEXINGTON AV/53	GEM	Manhattan	29526.161290
59 ST COLUMBUS	1ABC	Manhattan	26755.000000
59 ST	456NQRW	Manhattan	26422.161290
42 ST-BRYANT PK	7BDFM	Manhattan	25293.322581
5 AV/53 ST	EM	Manhattan	24195.580645
34 ST-PENN STA	ACE	Manhattan	24057.451613
42 ST-PORT AUTH	1237ACENQRSW	Manhattan	17302.258065
CANAL ST	6JNQRWZ	Manhattan	16446.354839
WALL ST	23	Manhattan	15878.096774

top april moring exits:

STATION	lines	borough	april_avg_exits_morning
14 ST-UNION SQ	456LNQRW	Manhattan	4355.954545
GRD CNTRL-42 ST	4567S	Manhattan	3924.500000
125 ST	456	Manhattan	3381.363636
34 ST-HERALD SQ	BDFMNRQW	Manhattan	3173.954545
FULTON ST	2345ACJZ	Manhattan	2903.409091
42 ST-PORT AUTH	1237ACENQRSW	Manhattan	2598.500000
34 ST-PENN STA	ACE	Manhattan	2476.863636
59 ST	456NQRW	Manhattan	2302.863636
JAMAICA CENTER	EJZ	Queens	2264.681818
96 ST	123	Manhattan	2238.863636
59 ST COLUMBUS	1ABC	Manhattan	2166.681818
TIMES SQ-42 ST	1237ACENQRSW	Manhattan	2074.545455
125 ST	ABCD	Manhattan	2045.545455
86 ST	456	Manhattan	2034.636364
FLUSHING-MAIN	7	Queens	1964.772727

top regular moring entries:

STATION	lines	borough	regular_avg_entries_morning
42 ST-PORT AUTH	1237ACENQRSW	Manhattan	46693.322581
GRD CNTRL-42 ST	4567S	Manhattan	38710.548387
34 ST-PENN STA	ACE	Manhattan	35357.064516
34 ST-HERALD SQ	BDFMNRQW	Manhattan	28983.709677
FLUSHING-MAIN	7	Queens	28649.806452
34 ST-PENN STA	123ACE	Manhattan	26336.419355
JKSM HT-ROOSVLT	7EFMR	Queens	25472.870968
FULTON ST	2345ACJZ	Manhattan	22748.354839
JAMAICA CENTER	EJZ	Queens	19477.709677
14 ST-UNION SQ	456LNQRW	Manhattan	19375.870968
JOURNAL SQUARE	1	NaN	18438.645161
ATL AV-BARCLAY	2345BDNQR	Brooklyn	17970.193548
KEW GARDENS	EF	Queens	16760.548387
GROVE STREET	1	NaN	16012.354839
CROWN HTS-UTICA	34	Brooklyn	15812.580645

top april moring entries:

STATION	lines	borough	april_avg_entries_morning
JAMAICA CENTER	EJZ	Queens	3119.227273
42 ST-PORT AUTH	1237ACENQRSW	Manhattan	2755.636364
CROWN HTS-UTICA	34	Brooklyn	2574.000000
JKSM HT-ROOSVLT	7EFMR	Queens	2326.409091
FLUSHING-MAIN	7	Queens	2068.500000
125 ST	456	Manhattan	1902.136364
161/YANKEE STAD	4BD	Bronx	1847.227273
34 ST-HERALD SQ	BDFMNRQW	Manhattan	1709.136364
14 ST-UNION SQ	456LNQRW	Manhattan	1651.454545
103 ST-CORONA	7	Queens	1636.181818
GRD CNTRL-42 ST	4567S	Manhattan	1629.954545
3 AV-149 ST	25	Bronx	1627.181818
FULTON ST	2345ACJZ	Manhattan	1567.727273
34 ST-PENN STA	123ACE	Manhattan	1560.500000
JOURNAL SQUARE	1	NaN	1543.318182

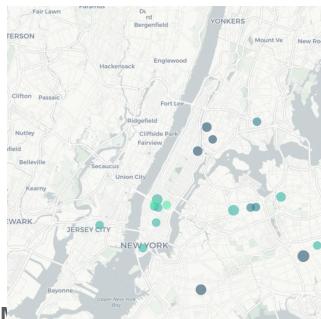
3. Ridership change by time-slot

4 hour time interval

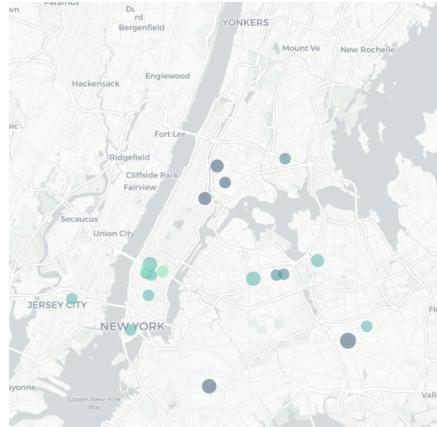
- Morning: before 12 p.m.
- Evening: after 12 p.m.

The flow of commuters

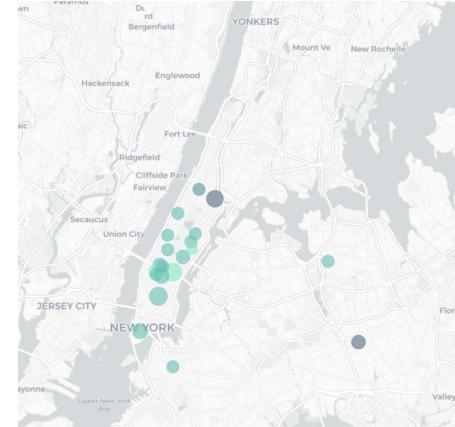
- From elsewhere to big stations in Manhattan.
- After the outbreak, Manhattan stations are still the top target, although less dominant
- But number of commuters in Manhattan may dropped the most.



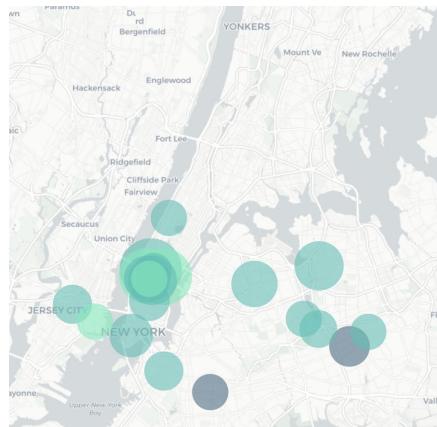
Size: avg morning entries in April



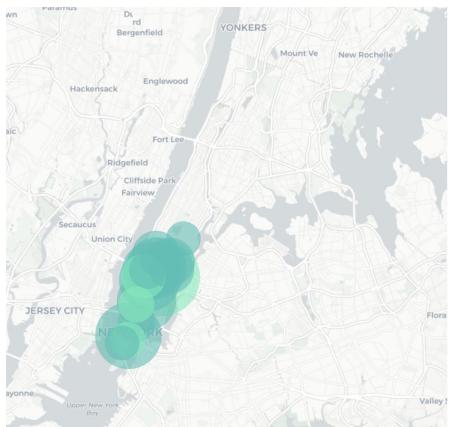
Size: avg morning exits in April



Size: avg morning entries in Jan&Feb



Size: avg morning exits in Jan&Feb



Brief discussion

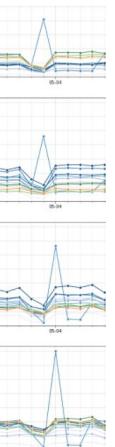
1. Overall ridership change over time
2. Ridership change by station and location
3. Ridership change by time-slot

Pros:

- Better data cleaning
- Workday and holidays, flow to and from work

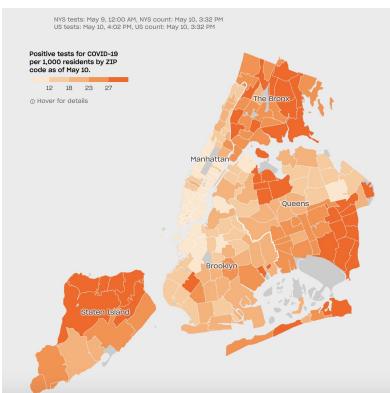
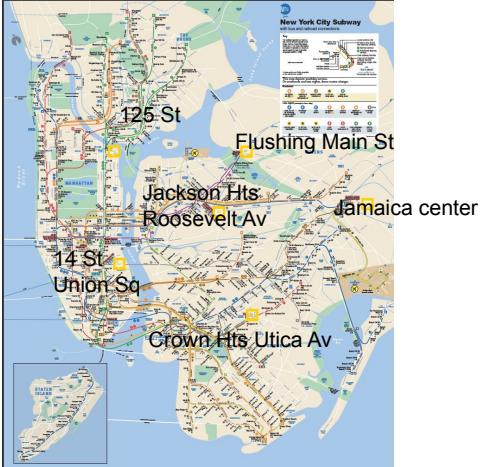
Limitations:

- Data quality: How to explain outliers?
- Limitation of turnstile dataset: Areas not covered by subway system
- Static plots: Use more interactive demonstration in the future?



42 St-Port Authority May 3

Line 456
Bronx -
Brooklyn



Positive test per 1000 residents by zip code
https://projects.thecity.nyc/2020_03_covid-19-tracker/

4. Ridership by Spatial Distribution - Methodology

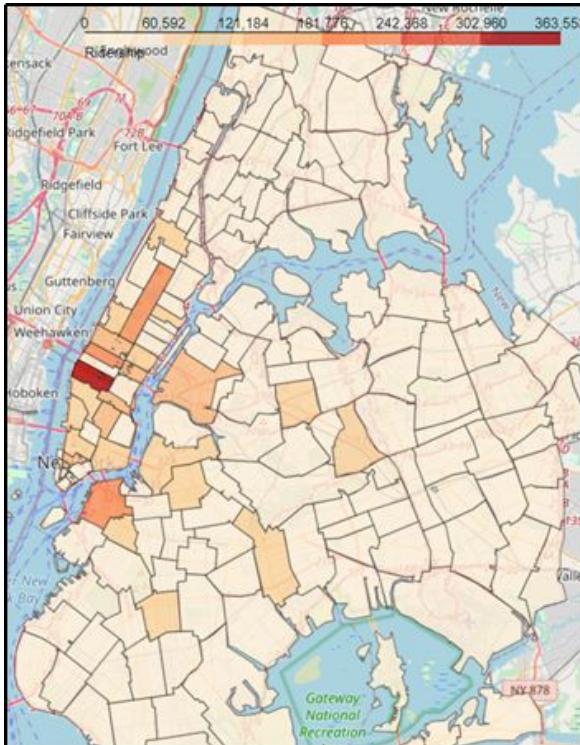
Spatial Aggregation: Quadtree Spatial Index

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

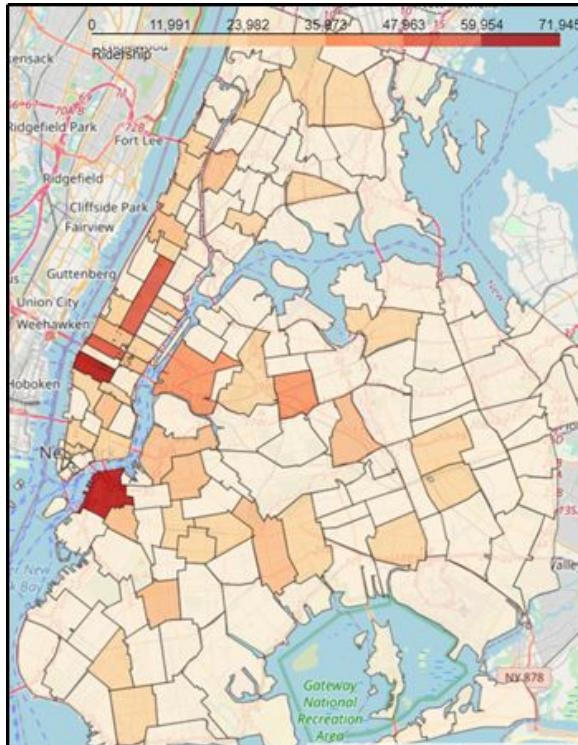
Quadtree

- 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

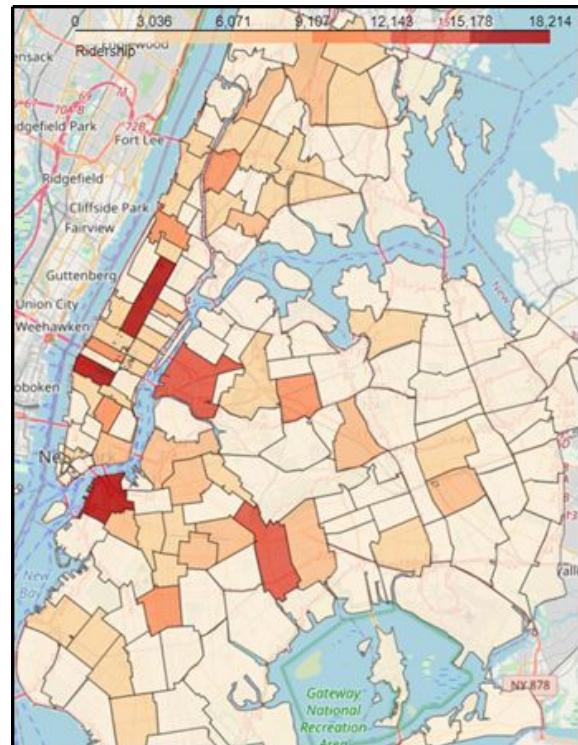
4. Ridership by Spatial Distribution - Overview



02 / 19 Ridership Map



03 / 18 Ridership Map



04 / 22 Ridership Map

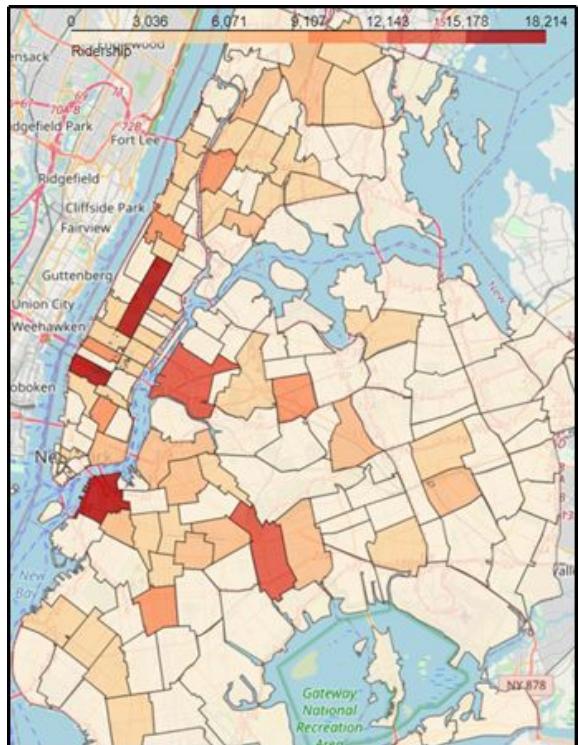
4. Ridership by Spatial Distribution - Finds

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



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

