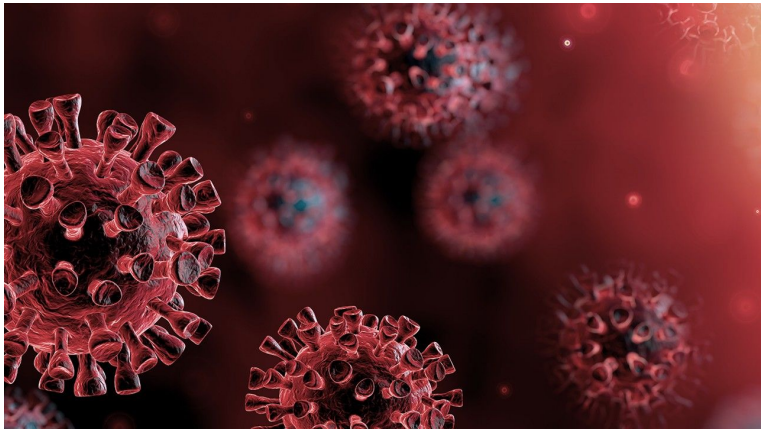
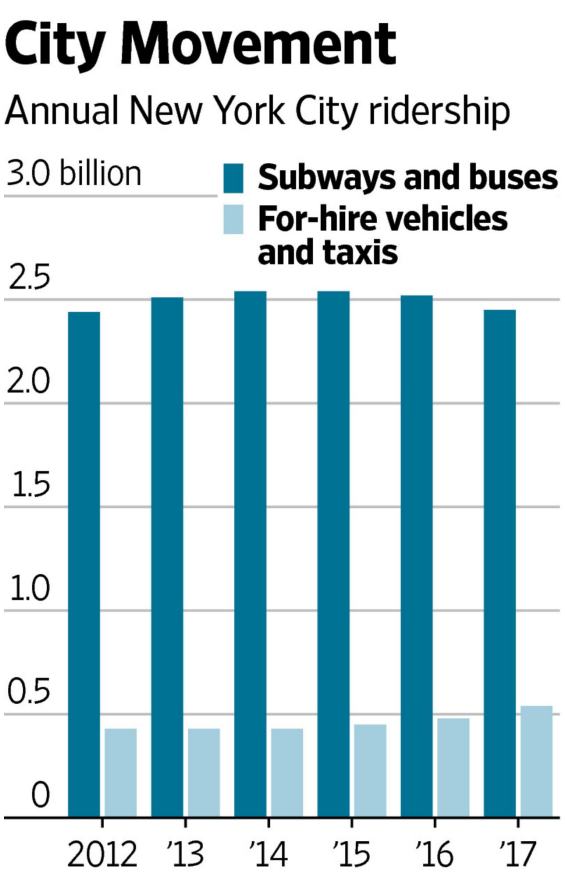
RUTGERS
UNIVERSITY | NEW BRUNSWICK

Analyzing the Impact of COVID-19 on Demand for Subway Systems

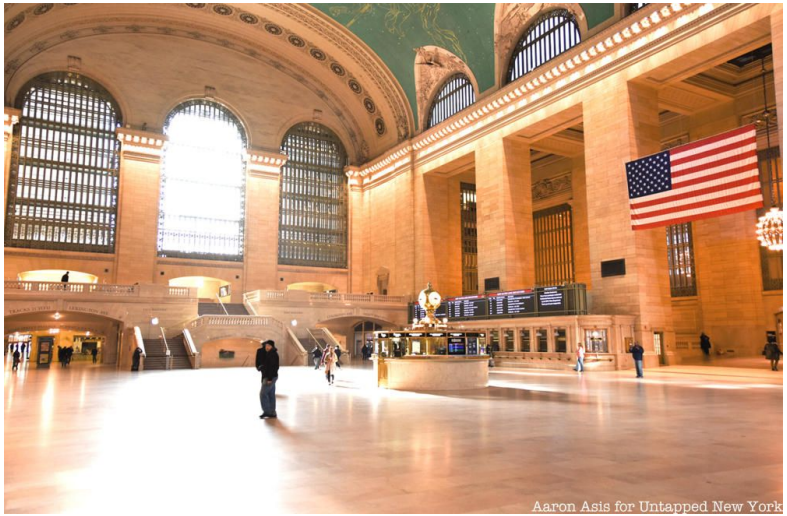
David Tian, Ilana Zane,
Daniel Pattathil, Srinivas Rajagopalan



Introduction



Introduction



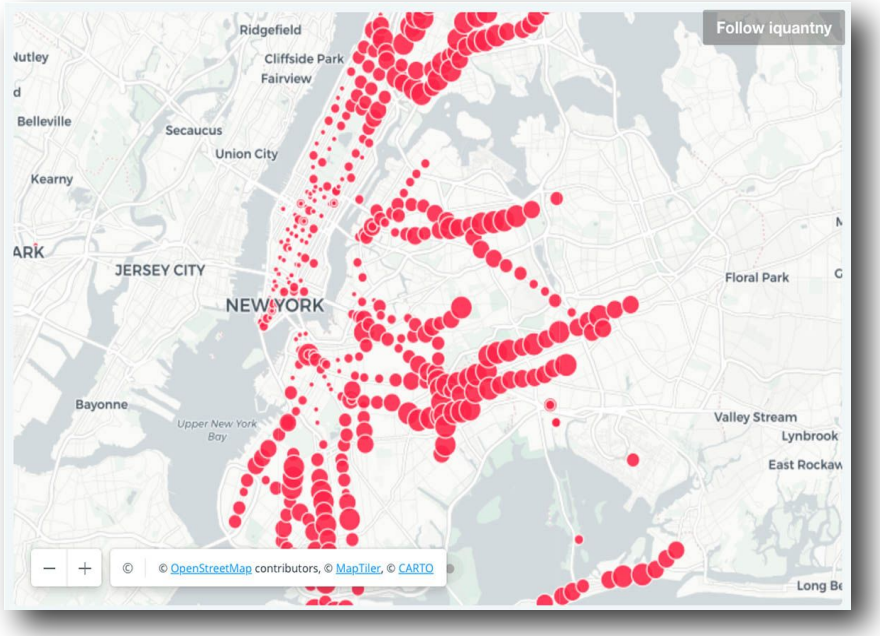
Reduced demand due to social distancing

Residents still need access to essential goods

Research Question: How can we model ridership changes during epidemics?

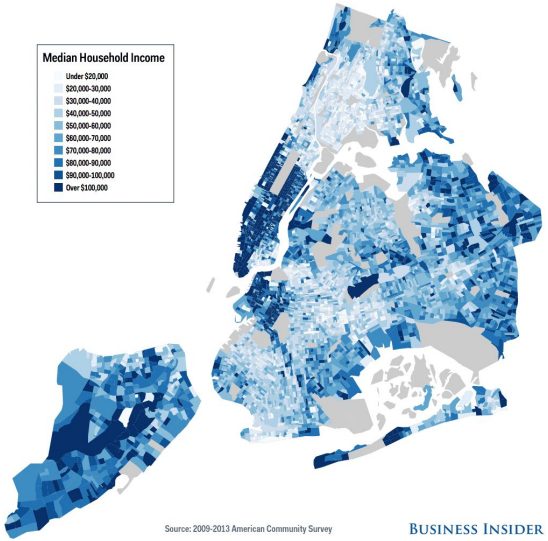
Introduction - Motivation

- Observed that ridership changes vary greatly between stations
- Considered factors such as confirmed COVID-19 cases, geographic area, socioeconomic status, and essential workers

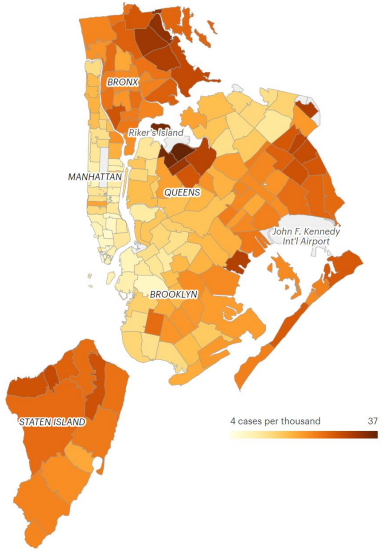


Design - Data Sources

MTA Turnstile Counts



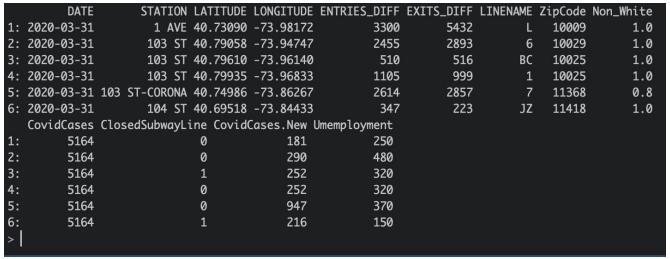
Number of returns with unemployment compensation



Confirmed COVID-19 Cases

Design

- Data Cleaning
  - MTA turnstile data given as cumulative counts
  - Calculate diff to get entries and exits per day, remove faulty entries
  - Assign geographic coordinates to each station



	DATE	STATION	LATITUDE	LONGITUDE	ENTRIES_DIFF	EXITS_DIFF	LINENAME	ZipCode	Non_White
1:	2020-03-31	1 AVE	40.73090	-73.98172	3300	5432	L	10009	1.0
2:	2020-03-31	103 ST	40.79058	-73.94747	2455	2893	6	10029	1.0
3:	2020-03-31	103 ST	40.79610	-73.96140	510	516	BC	10025	1.0
4:	2020-03-31	103 ST	40.79935	-73.96833	1105	999	1	10025	1.0
5:	2020-03-31	103 ST-CORONA	40.74986	-73.86267	2614	2857	7	11368	0.8
6:	2020-03-31	104 ST	40.69518	-73.84433	347	223	JZ	11418	1.0

	CovidCases	ClosedSubwayLine	CovidCases.New	Umemployment
1:	5164	0	181	250
2:	5164	0	290	480
3:	5164	1	252	320
4:	5164	0	252	320
5:	5164	0	947	370
6:	5164	1	216	150

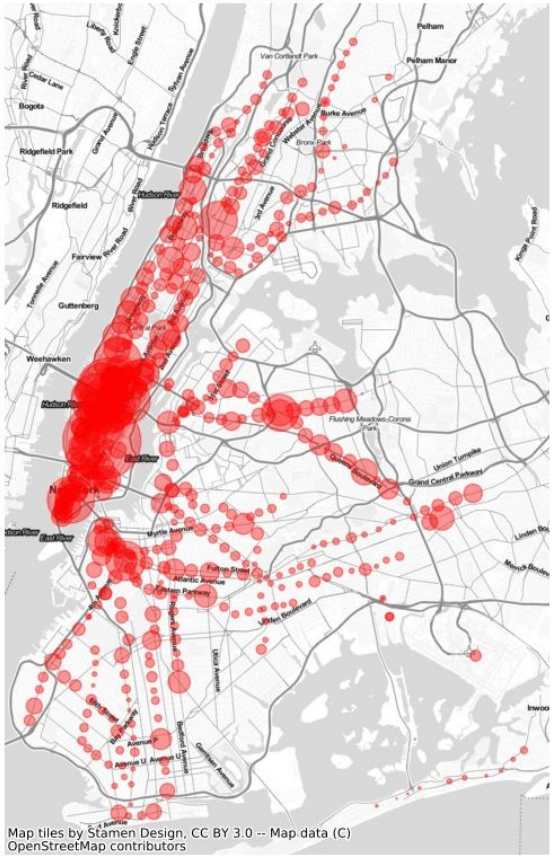
>

- Linear Regression Model
  - whether subway is closed/open
  - percentage of non-white residents
  - unemployment compensation returns

Evaluation - Visualization

- We first verified the integrity and accuracy of our data with an animated visualization
- Absolute ridership numbers from February to April 2020
- Lines up with social distancing measures implemented in March

MTA Ridership by Number of Entries
02/01/2020



Evaluation - Linear Regression Model

- Overall, model has significant p-value (< 2.2e-16)
- Most significant predictors of EXITS_DIFF were coronavirus cases and number of unemployment returns

Predictor	p-value in model
Coronavirus Cases (In Zip Code of Station)	0.287047
Closed Subway Line	< 2e-16
Percent of Non-White Residents	< 2e-16
Number of Unemployment Returns In Zip Code (2017)	0.000213

Evaluation - Model

- We observed that there was correlation between the number of unemployment returns and coronavirus cases (R = .51)
  - This could indicate that areas of greater coronavirus cases are in greater poverty areas.

- We did not observe any significant differences between VIF for each predictor.
  - This shows that there was not multicollinearity between our predictors

Evaluation - Recommendations

- Our analysis shows that coronavirus cases are linked towards poorer socioeconomic areas.
- However, the correlation between coronavirus cases and the number of exits for each station was poor (R = -0.07).
- Accuracy of the data should be noted: some stations had abnormally high values for number of exits, these datapoints were removed
  - 942 data points were greater than Q3 + 1.5*IQR
- If more people use the subway, there is less of an opportunity to practice social distancing
- This can lead to higher numbers of confirmed cases

Conclusion

- Based on our data and visualizations we see that there is a correlation between confirmed cases, economic status, and subway volume
- We can use this data to predict future demand for the following weeks and how to better prepare the subway infrastructure in the case of another pandemic