Predicting SARS-CoV-2 in Wastewater:

The Role of K-12 Transmission in NYC, 2020-2023

Lou Hines
August 2023

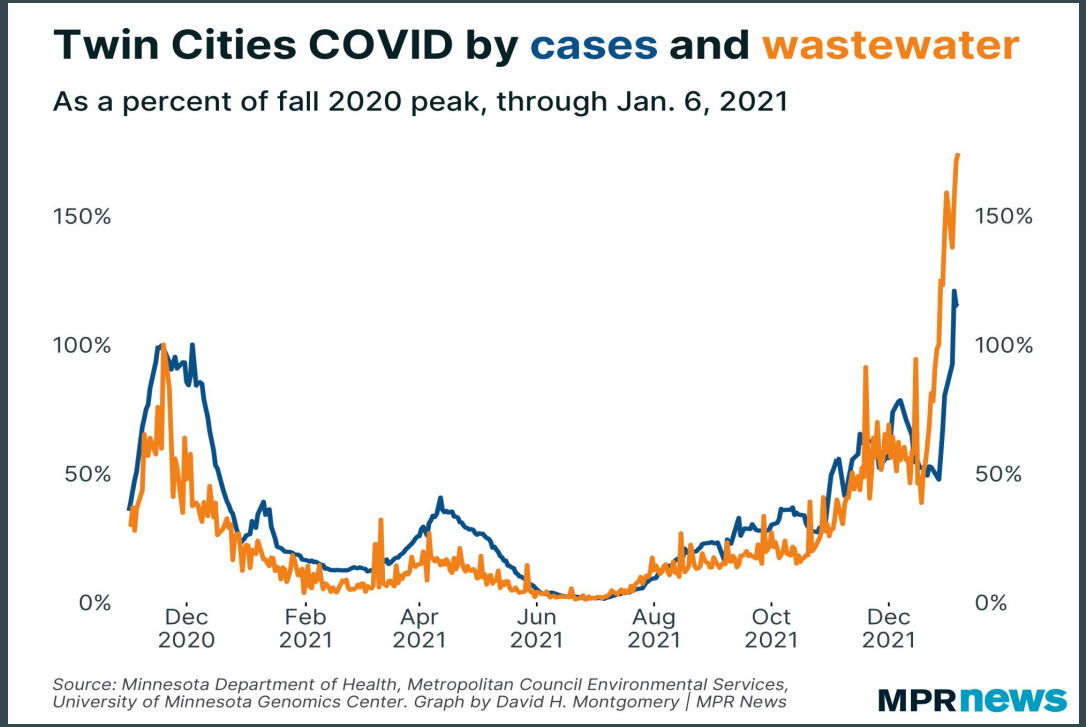
Assumptions:

Wastewater Monitoring:

- Cost-effective
- Highly Correlated with Cases

School Dynamics:

- Big role in spread



Our findings provide valuable insights that could contribute to more informed public and private health decisions.

Original Features

Gene Copies per Liter

Per Capita Daily Gene Copies*

Facility Name

Sample Date

Facility Population

Test Date (dropped)

Annotations (dropped)

Engineered Features

Facility School-Aged Population

Holidays

Season

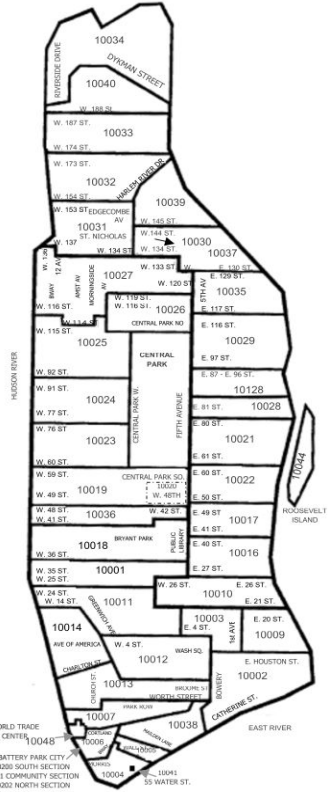
School Status*

Non-Correspondence of Zip Codes v. Facilities

ZIP CODE MAP OF MANHATTAN

Unique Zip Codes

10103	666 Fifth Ave.
10104	1290 Ave. of the Am.
10105	1345 Ave. of the Am.
10106	888 Seventh Ave.
10107	250 W. 57th St.
10110	500 Fifth Ave.
10111	630 Fifth Ave
10112	30 Rockefeller Plaza
10115	475 Riverside Drive
10118	350 Fifth Ave.
10119	1 Penn Plaza
10120	112 W. 34th St.
10121	2 Penn Plaza
10123	450 Seventh Ave.
10151	745 Fifth Ave.
10152	375 Park Ave.
10153	767 Fifth Ave.
10154	345 Park Ave.
10155	964Third Ave. (150 E. 58th St.)
10158	605 Third Ave.
10162	500 E. 77th St.
10165	60 E. 42nd St.
10166	200 Park Ave.
10167	245 Park Ave.
10168	122 E. 42nd St.
10169	230 Park Ave.
10170	420 Lexington Ave.
10171	299 Park Ave.
10172	277 Park Ave.
10173	342 Madison Ave.
10174	405 Lexington Ave.
10175	521 Fifth Ave.
10176	551 Fifth Ave.
10177	250 Park Ave.
10270	70 Pine. St.
10271	120 Broadway
10278	26 Federal Plaza
10279	233 Broadway



Wastewater Treatment Plants

Area No.	Borough	Rated Cap MGD
1 ★	**Manhattan** Wards Island	275
2	North River	170
3 ★	**Bronx** Hunts Point	200
4	**Brooklyn** Newtown Creek	310
5 ★	26th Ward	85
6 ★	Coney Island	110
7	Red Hook	60
8	Owls Head	120
9 ★	**Queens** Tallman Island	80
10 ★	Jamaica	100
11 ★	Bowery Bay	150
12 ★	Rockaway	45
13	**Staten Island** Port Richmond	60
14	Oakwood Beach	40
	Total	1805

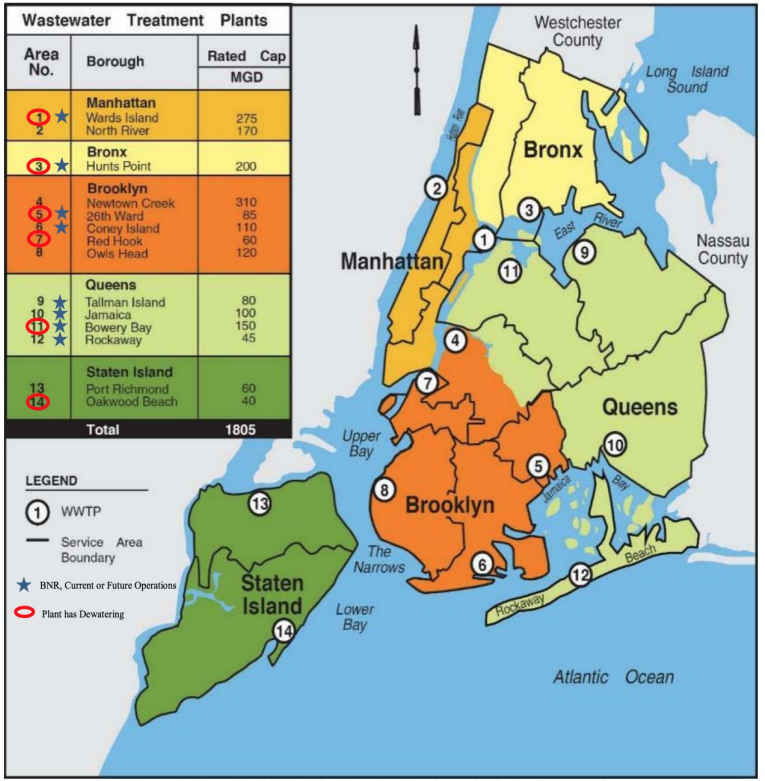
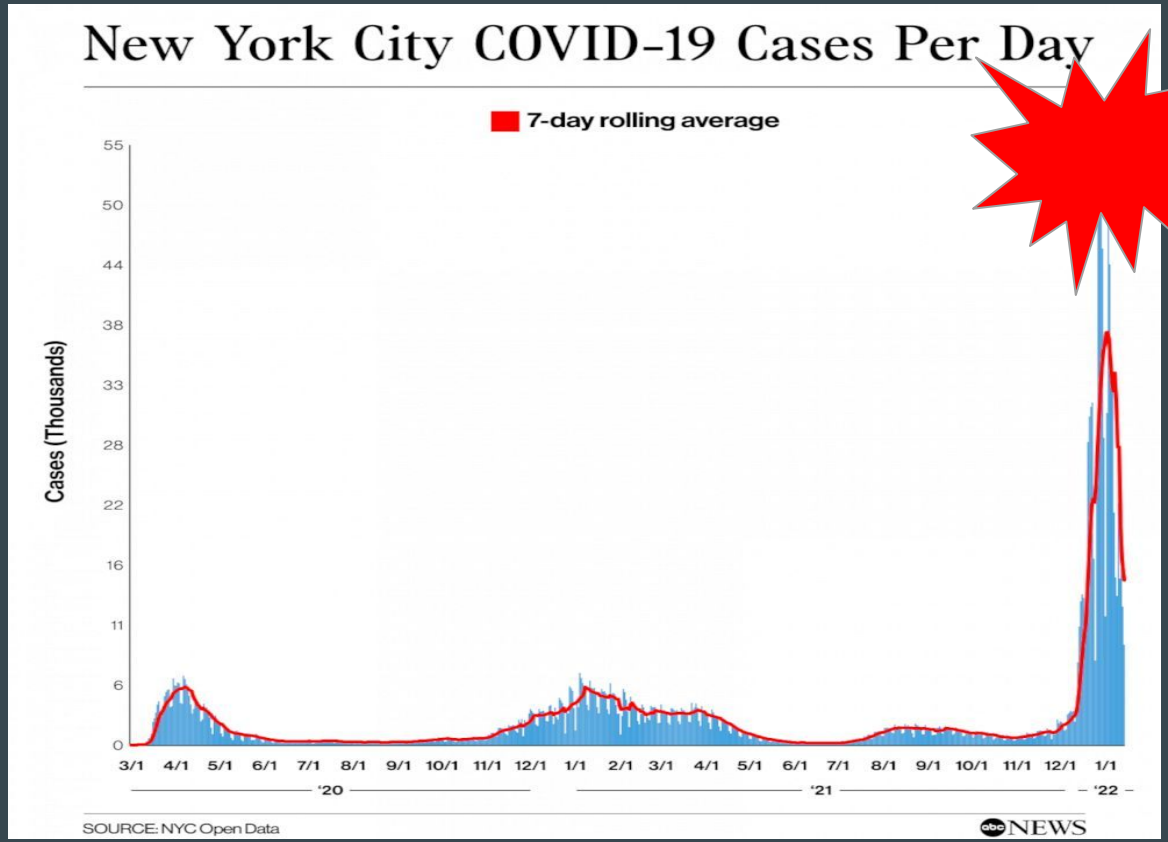


Figure 4: New York City Wastewater Treatment Plants (WWTPs)

Dividing the Data

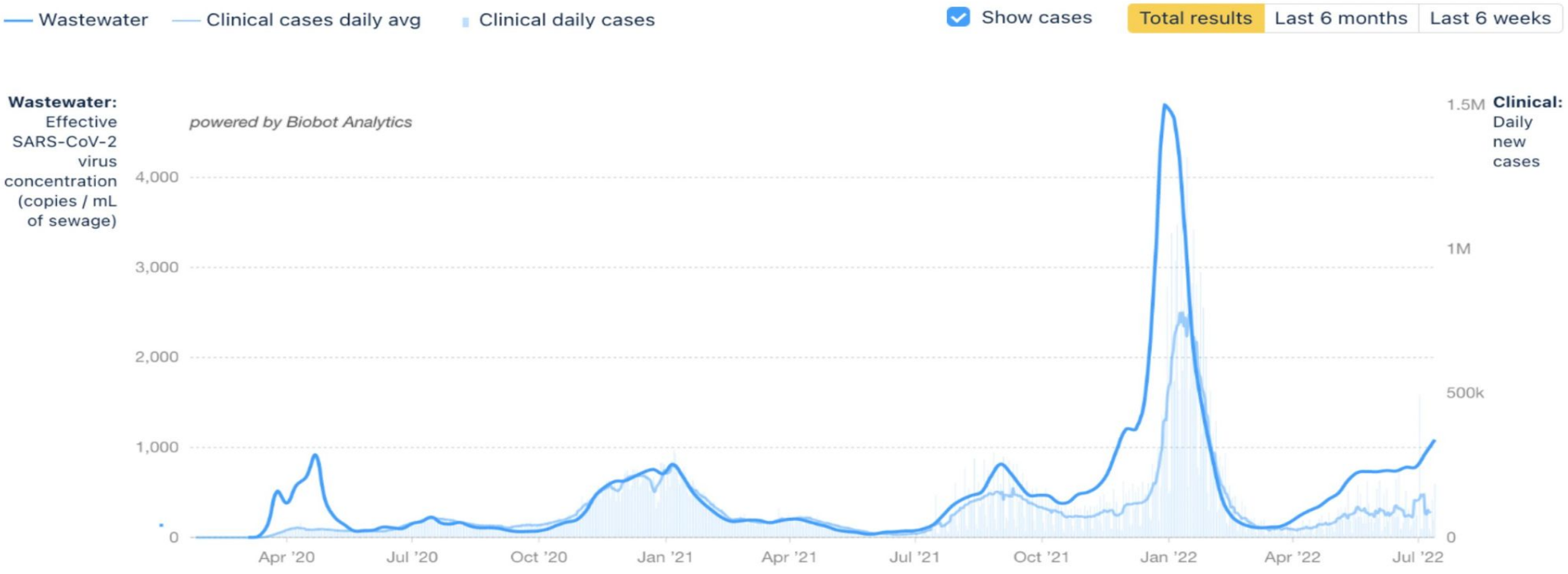

New York City COVID-19 Cases Per Day
7-day rolling average
Cases (Thousands)
55
50
44
38
33
28
22
16
11
6
0
3/1 4/1 5/1 6/1 7/1 8/1 9/1 10/1 11/1 12/1 1/1 2/1 3/1 4/1 5/1 6/1 7/1 8/1 9/1 10/1 11/1 12/1 1/1
'20
'21
'22
SOURCE: NYC Open Data
abc NEWS

OMICRON

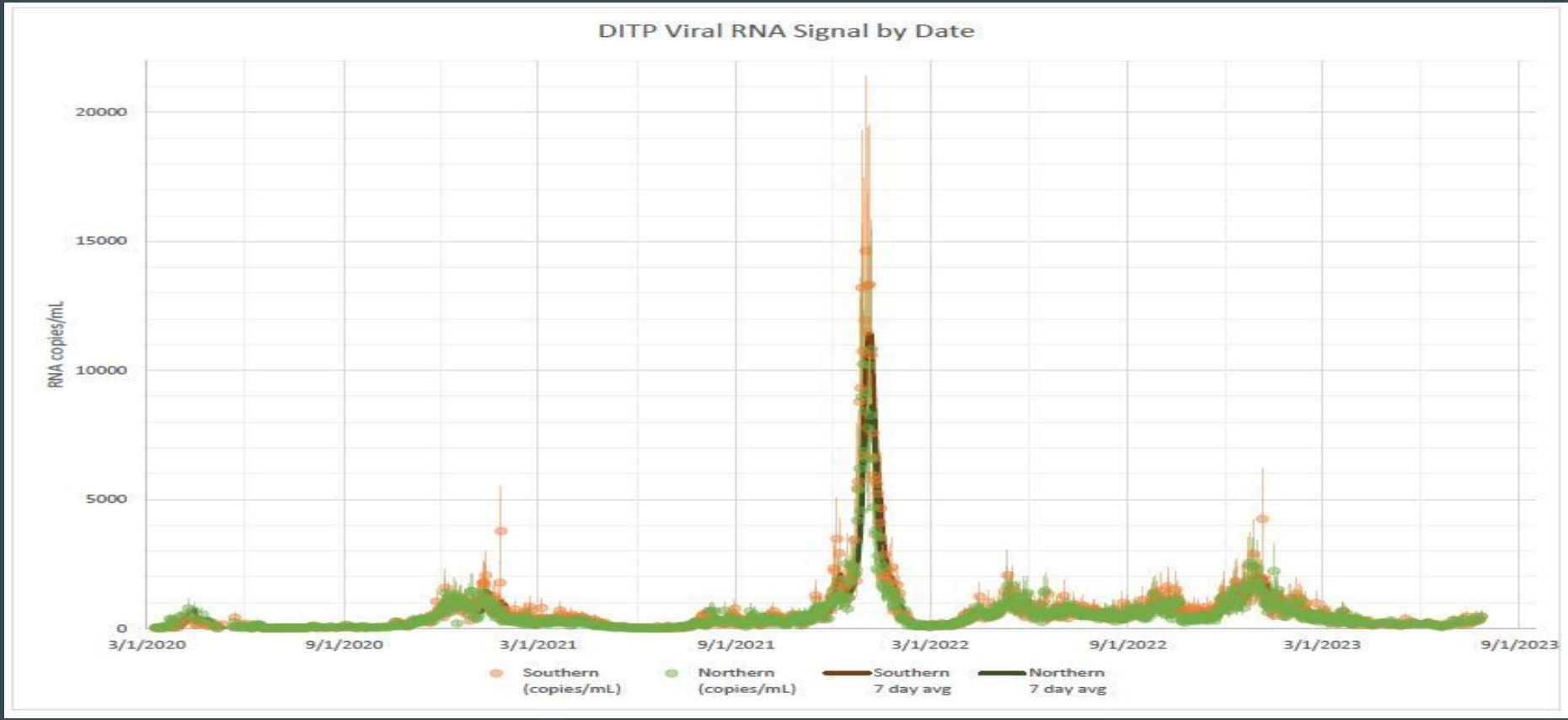
Covid-19 Wastewater Monitoring in the U.S.

This chart shows the SARS-CoV-2 virus concentration present in samples of wastewater taken from across the United States. The level of virus in wastewater is a leading indicator, meaning it precedes the change in clinical case counts or hospitalizations.

Week of July 13, 2022



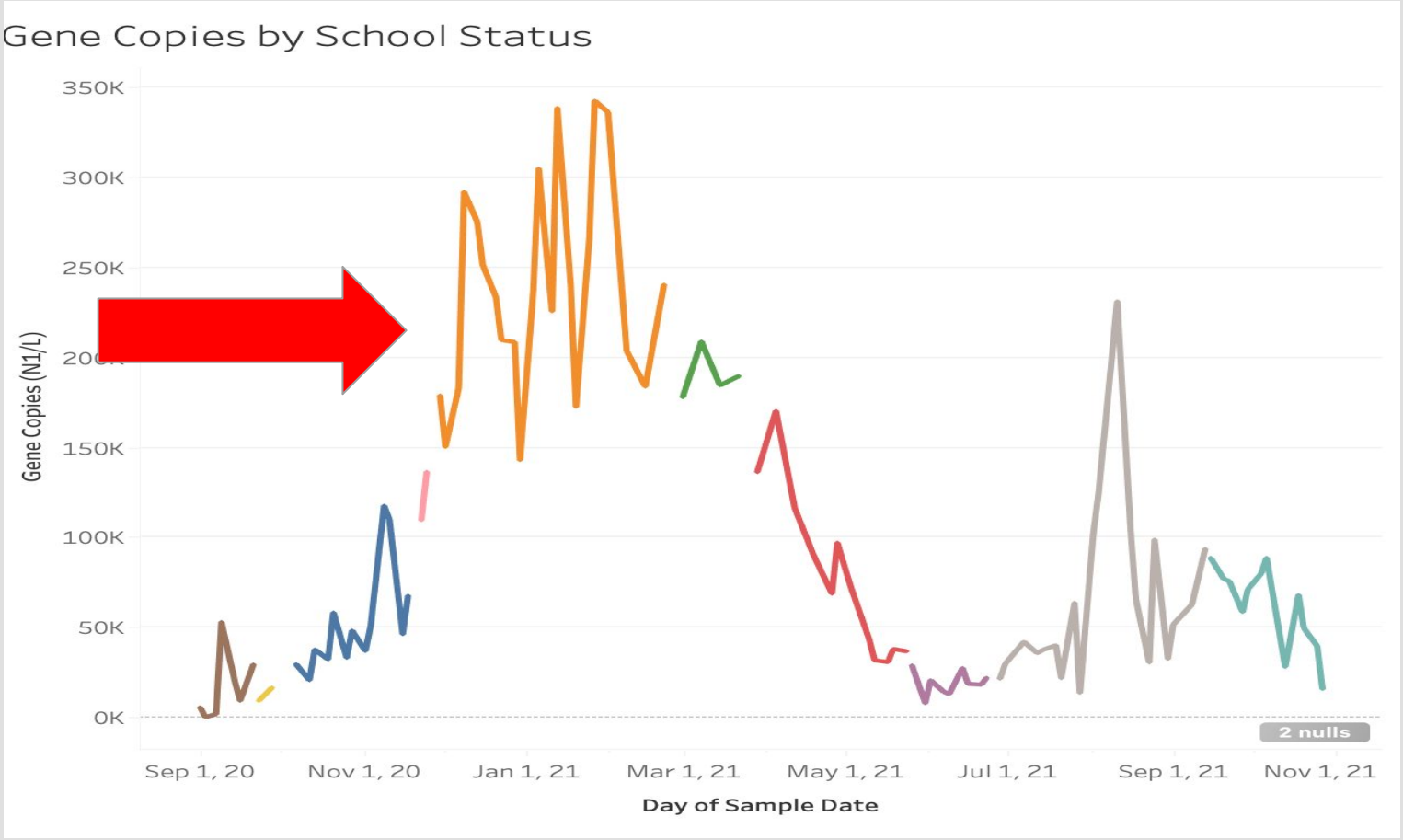
BIO BOT



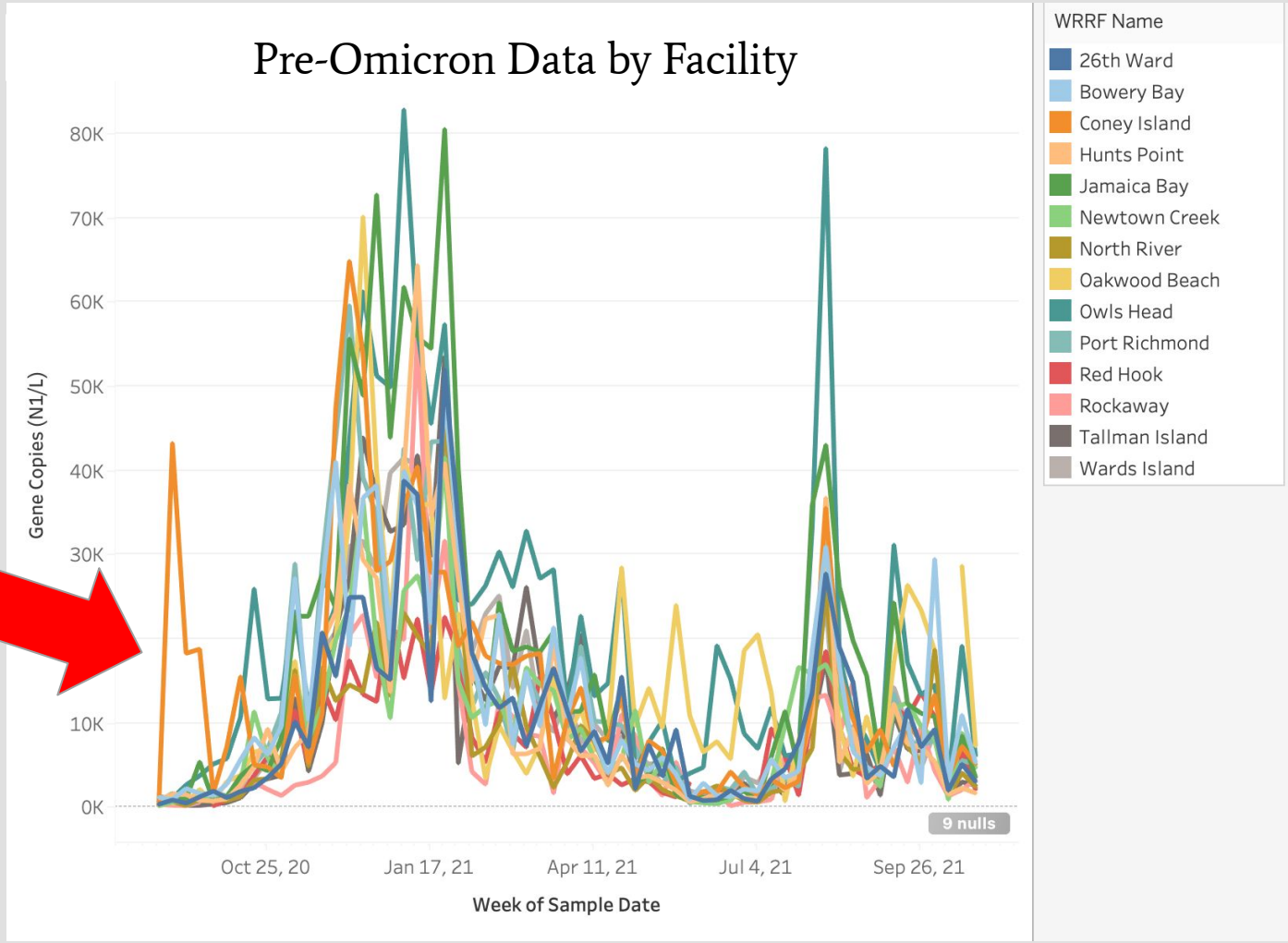
Wastewater COVID-19 Tracking

Massachusetts Water Resources Authority

Preliminary Analysis: Are we on to Something?


Gene Copies by School Status
350K
300K
250K
150K
100K
50K
0K
Gene Copies (N1/L)
2 nulls
Sep 1, 20
Nov 1, 20
Jan 1, 21
Mar 1, 21
May 1, 21
Jul 1, 21
Sep 1, 21
Nov 1, 21
Day of Sample Date

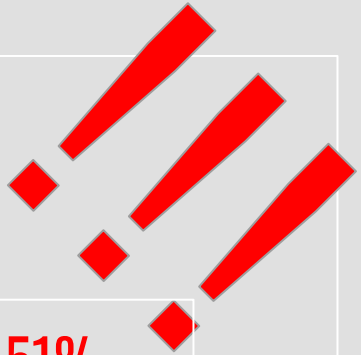
Looks promising so far, but what's this? A SUMMER SURGE?



Linear Regressions, Pre-Omicron*

Original Features: R2 = 4.5%
All Features: R2 = 53%

School-only Features: R2 = 51%

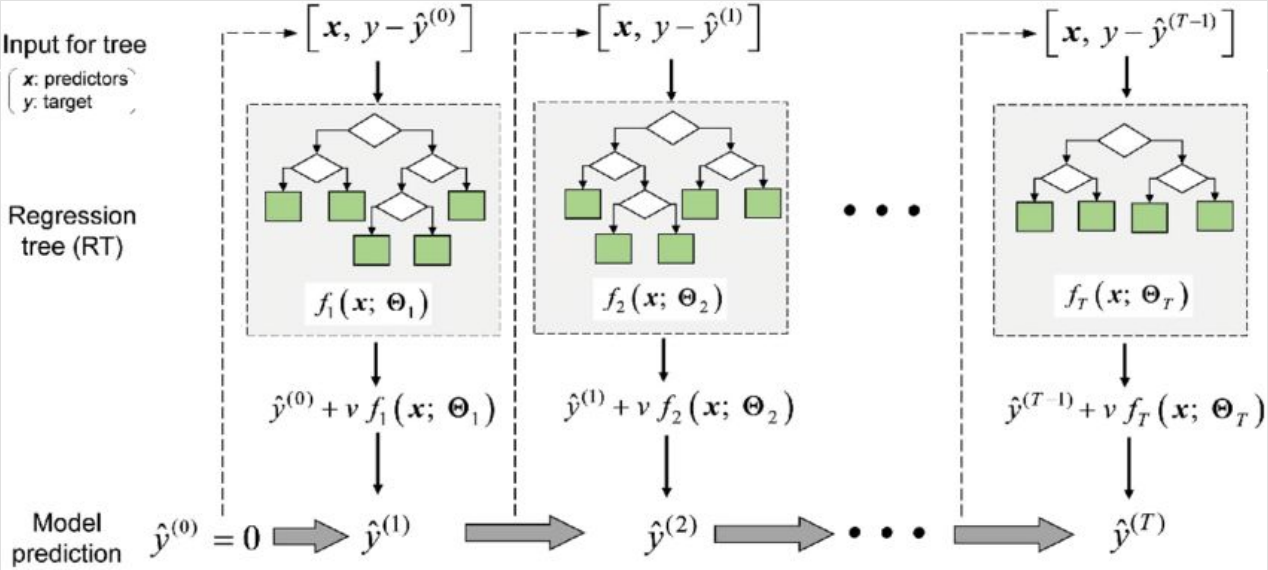


All scores are test scores.

Best Model

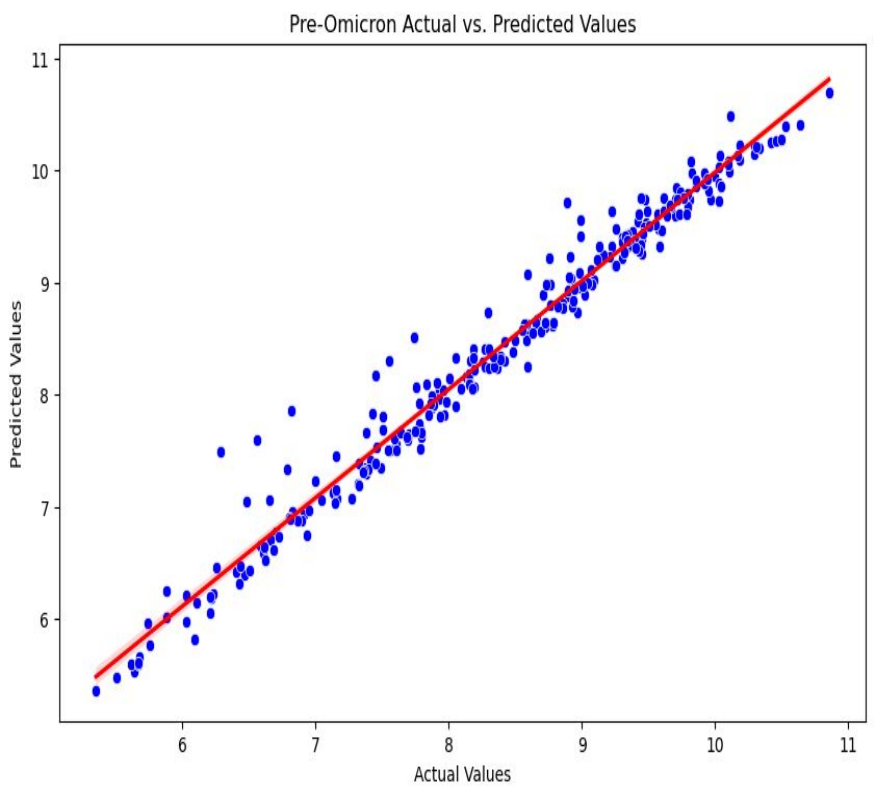
Gradient Boosting Regression (with Grid Search):

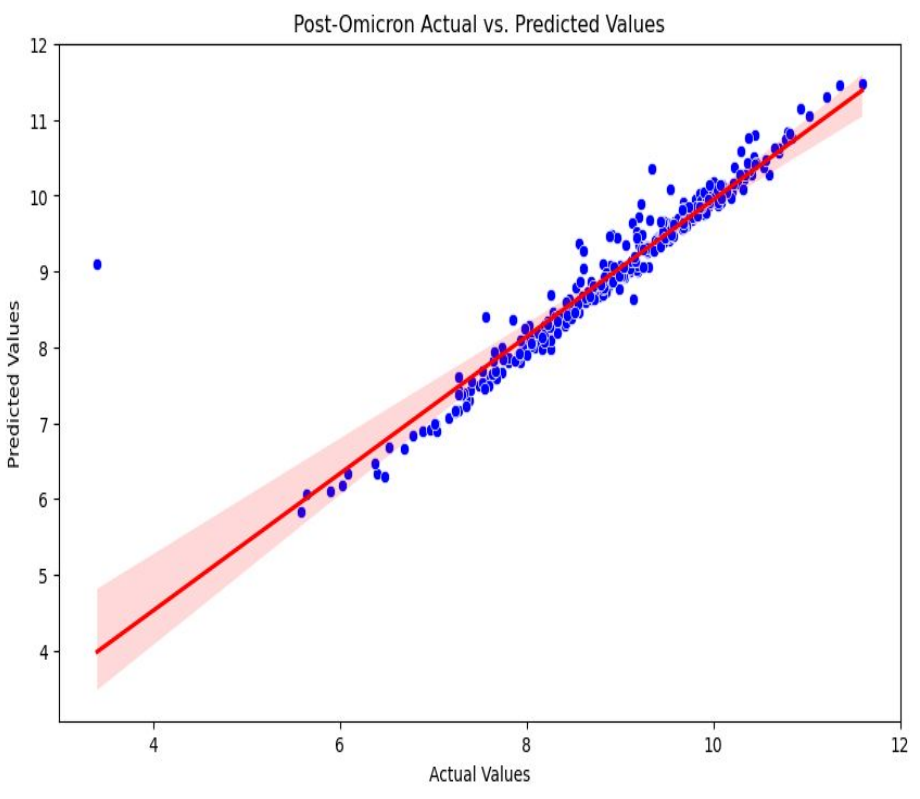
All features less dates. Train R2: 97.29, Test R2: 97.26



Our Model Worked Well Across Both Datasets:

Train R2: 96.5, Test R2: 92.8 - Some Overfitting

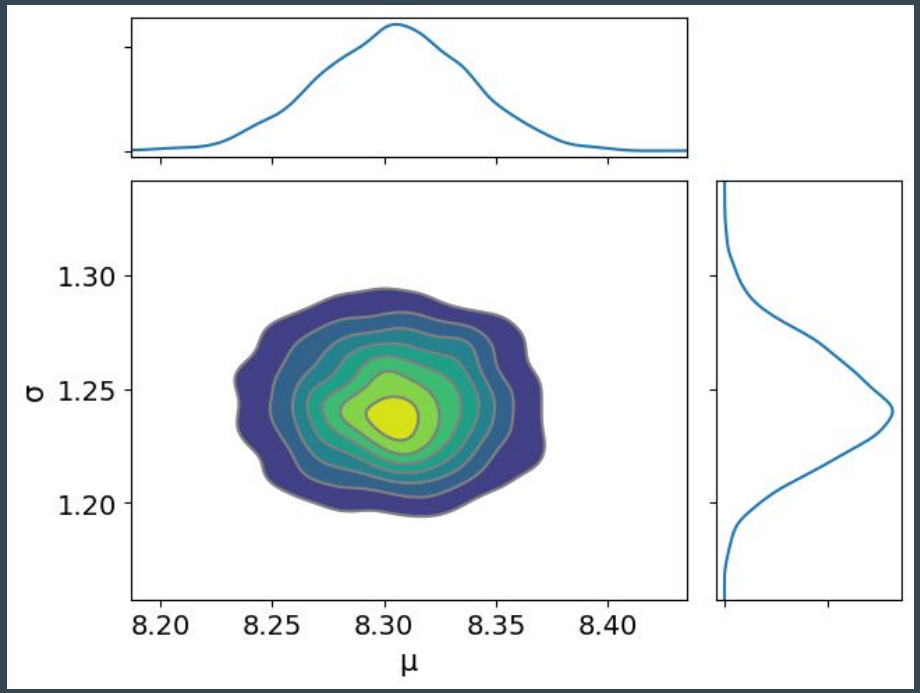


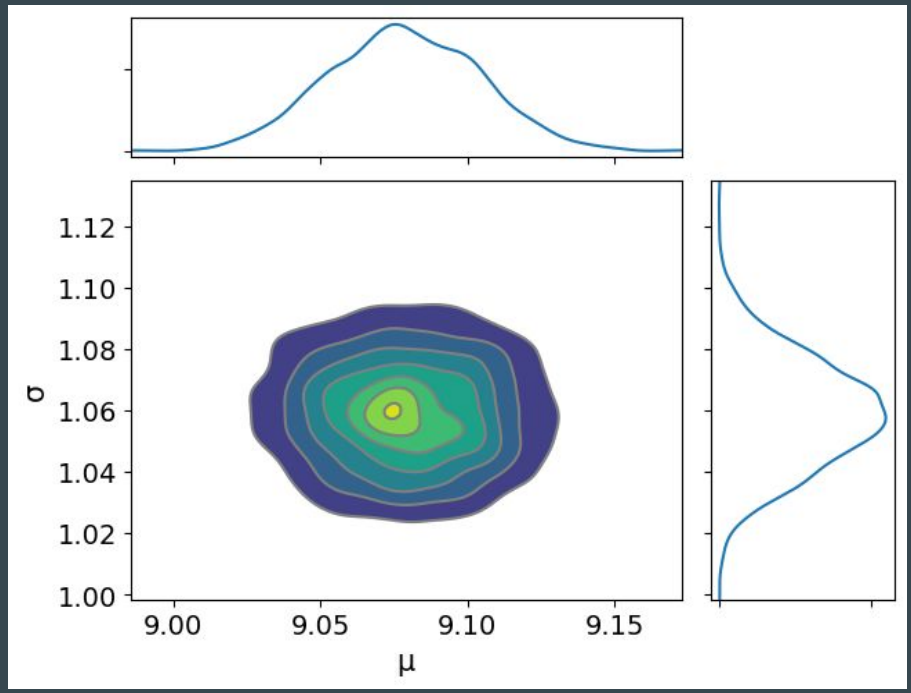


Key Predictors:

- Wastewater Facility
- Holidays
- School-Aged Population
- Per Capita Daily Gene Copies
- Season

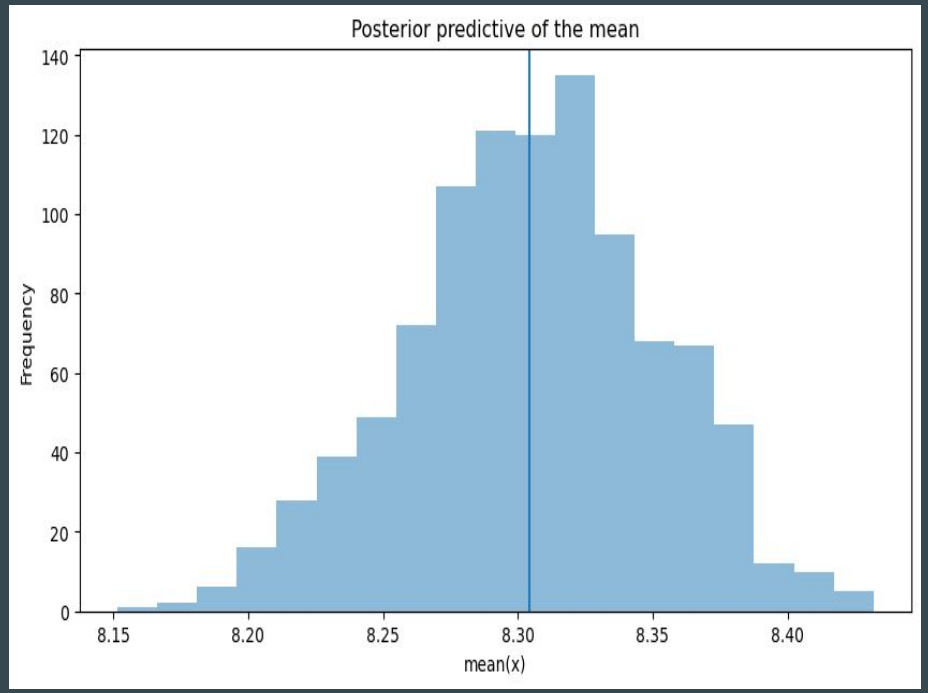
Bayesian Analysis: Checking our Priors

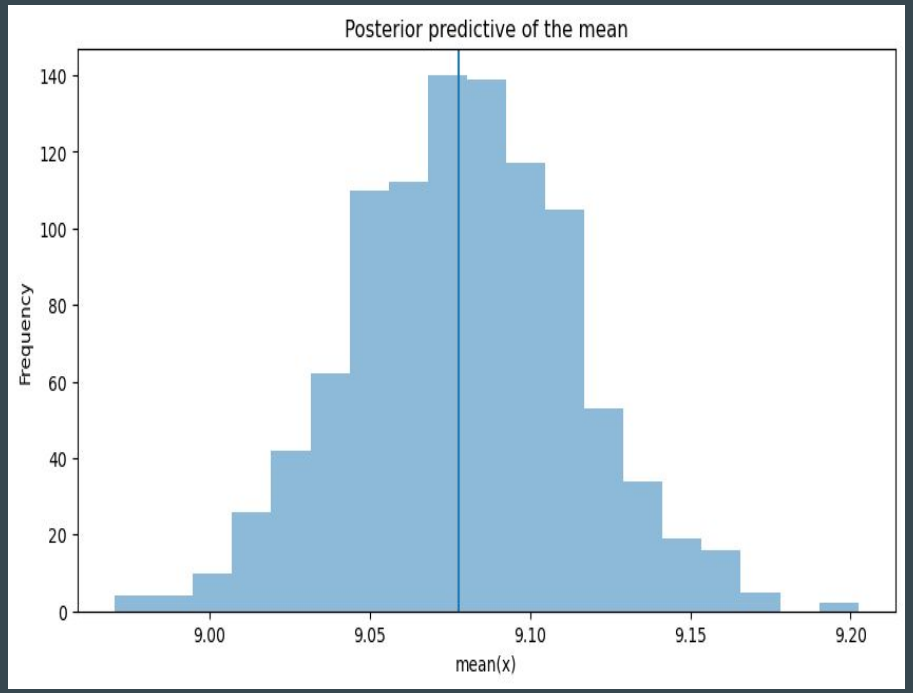




Gelman Rubin = 1 (both) , MSCE = ~.0, .000

Bayesian Analysis: Checking Priors





Recommendations: School MAP

School **Monitoring**:

- Passive monitoring via air sampling
- CO2 concentrations
- School wastewater sampling

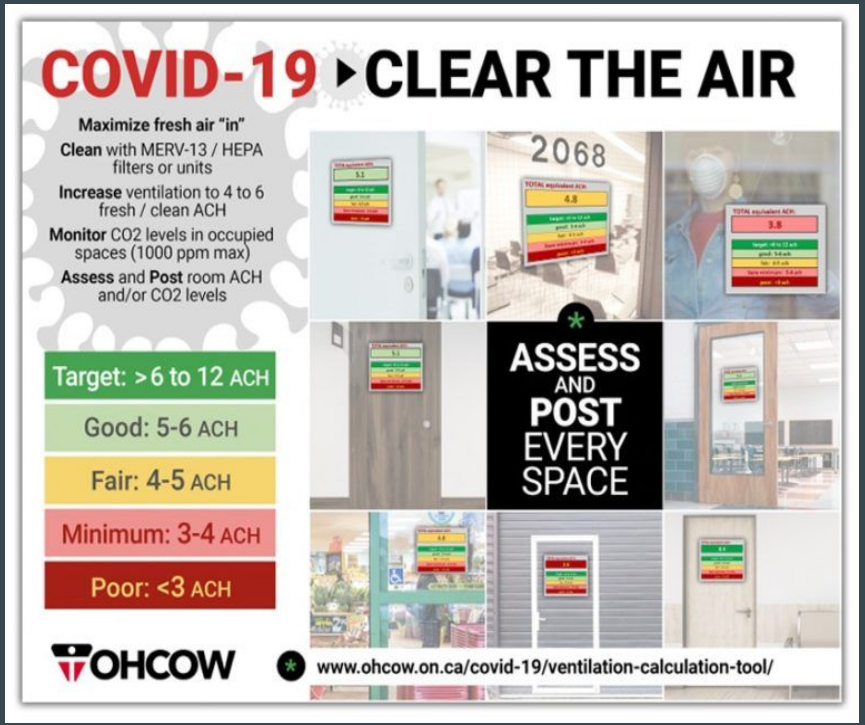
School **Analysis**:

- Enhanced predictive modelling

School **Prevention**:

- Air quality ASHRAE standards (ventilation and air purification)
- PPE

AQ Monitoring and Cleaning



Further Research:

Better understanding of city's data generation process

More granular NYC school data - maybe scrape Twitter?

Labor dynamics: in-person or remote work

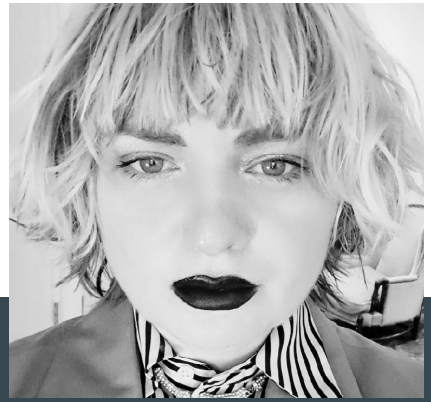
Unrecorded population shifts during 2020-2021

SIRS or other epidemiological modelling

Interactive public-use dashboards

Questions?

Contact Info



Lou Hines, Data Scientist

lou.m.hines@gmail.com

https://www.linkedin.com/in/lou-hines-data-scientist/