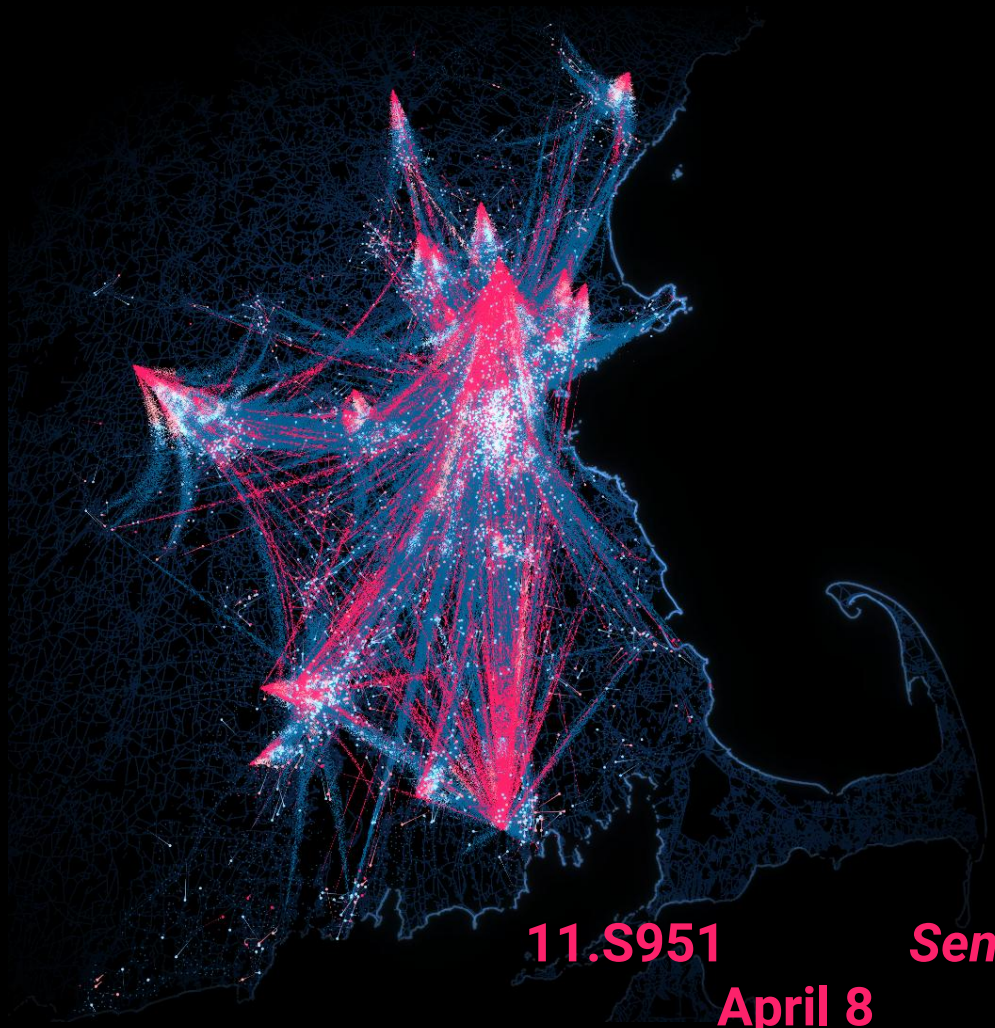


Week 9

City Scanner:
Insights from environmental data

Sanjana Paul, An Wang,
Simone Mora

11.S951 *Senseable City : Data and Analytics*

April 8

Class Agenda + Goals

Agenda

- CityScanner Recap
- Context
  - Environmental Sensing 101
  - The Bronx
- Activities
  - WAQI API Exercise
  - Time Series Analysis
  - Hotspot Analysis + Mapping
  - Exposure analysis with Twitter data

Goals

- Understanding hyperlocal air quality sensing value
- Historical context on The Bronx
- Use air quality API to access + interpret data
- Perform time series + hotspot analysis
- Make maps of environmental phenomenon

City Scanner Recap

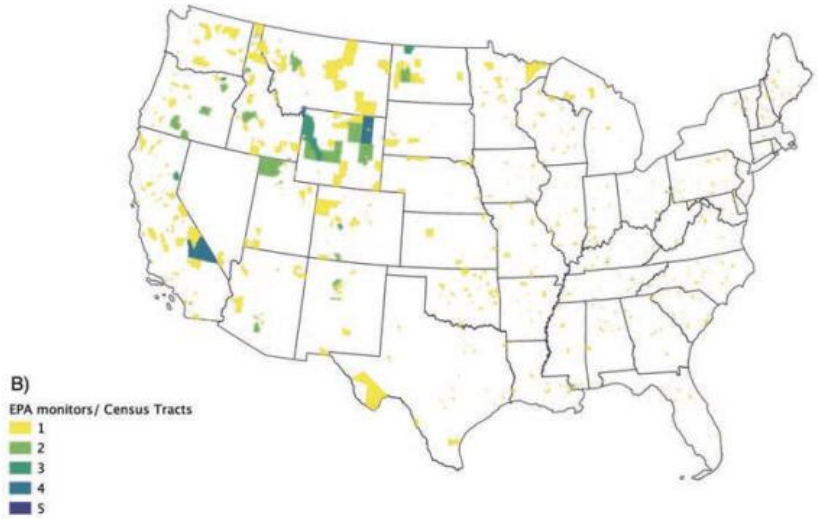
Drive-by Sensing

Can we turn urban vehicles into sensing platforms?

City Scanner

- Location
- Air quality (PM, CO, NOx)
- Temperature & Humidity
- Noise

Stationary vs. Mobile Sensors



EPA Monitors that report PM2.5 from 2015 to Feb 22 2020 per census tract in the US (deSouza and Kinney 2021)

Space coverage achieved with five city scanner sensors deployed in the Bronx for 3 months


Sparwood
2019
Cambridge
2017
New York City
2020
Stockholm
2020
Oskemen
2021
Beirut
2021

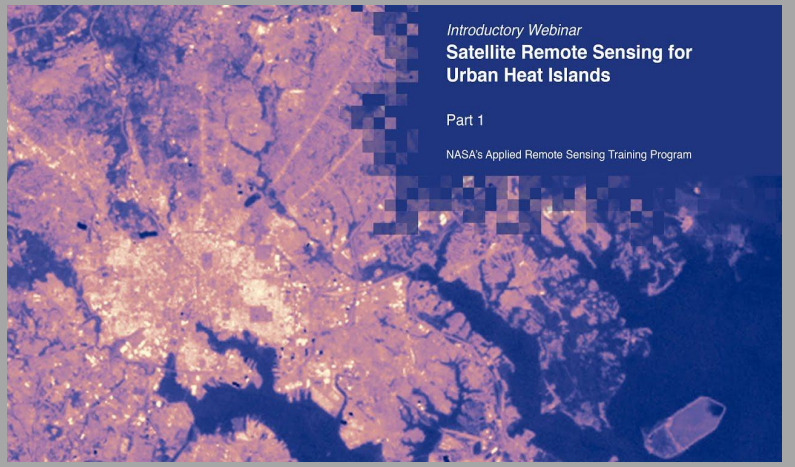
Environmental Sensing

Context + Background

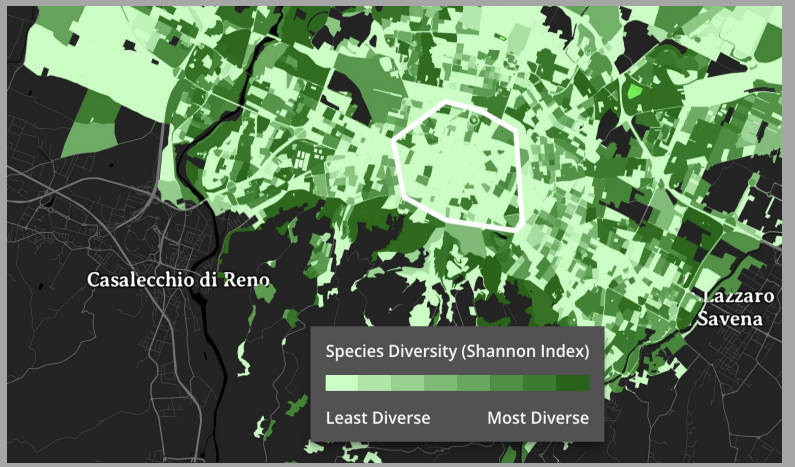
ENVIRONMENTAL SENSING 101

What information can we gather about our environment with different sensors?

Heat, noise, air quality, temperature humidity, soil health, water pollutants, tree health, biodiversity, and more!

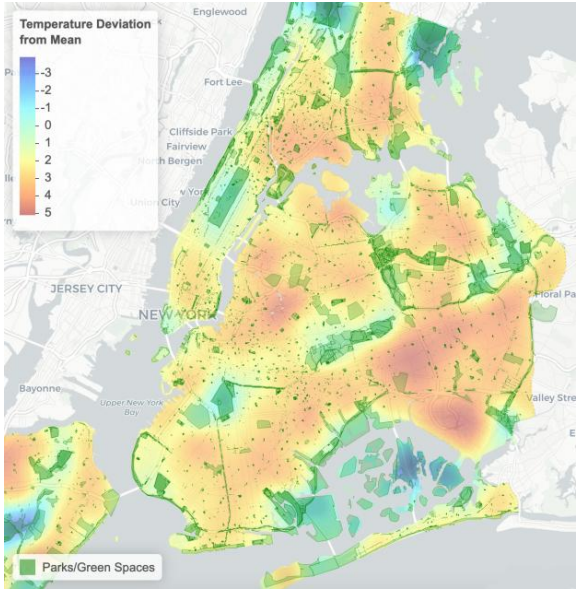


NASA ARSET

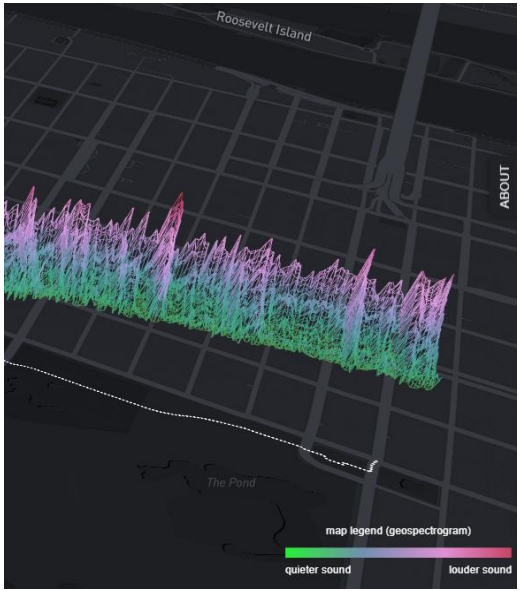


DiversiTree

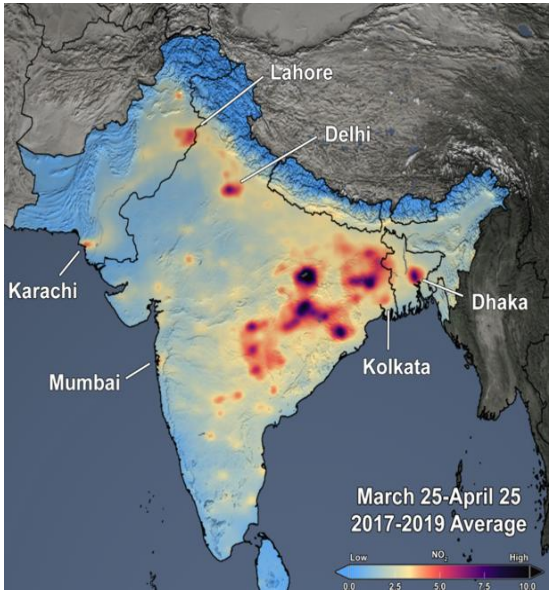
Environmental Sensing: Heat, Noise, Air Quality



https://news.climate.columbia.edu/2021/08/26/study-maps-urban-heat-islands-with-focus-on-environmental-justice/



https://senseable.mit.edu/sonic-cities/



https://aura.gsfc.nasa.gov/airquality.html

Air Quality Sensing: Why?

The Great Smog 1952 (https://www.britannica.com/event/Great-Smog-of-London)

Los Angeles Smog (https://www.britannica.com/science/smog#ref16459)

Air Quality Sensing: Why?

4.2 million per year vs. 6.2 million

91% of world population lives in places exceeding WHO AQ Standards

RESEARCH ARTICLE | SOCIAL SCIENCES | FULL ACCESS

Half of US population exposed to adverse lead levels in early childhood

Michael J. McFarland, Matt E. Hauer, and Aaron Reuben Authors Info & Affiliations

PNAS

March 7, 2022 | 119 (11) e2118631119 | https://doi.org/10.1073/pnas.2118631119

Vol. 119 | No. 11

THIS ARTICLE HAS BEEN UPDATED

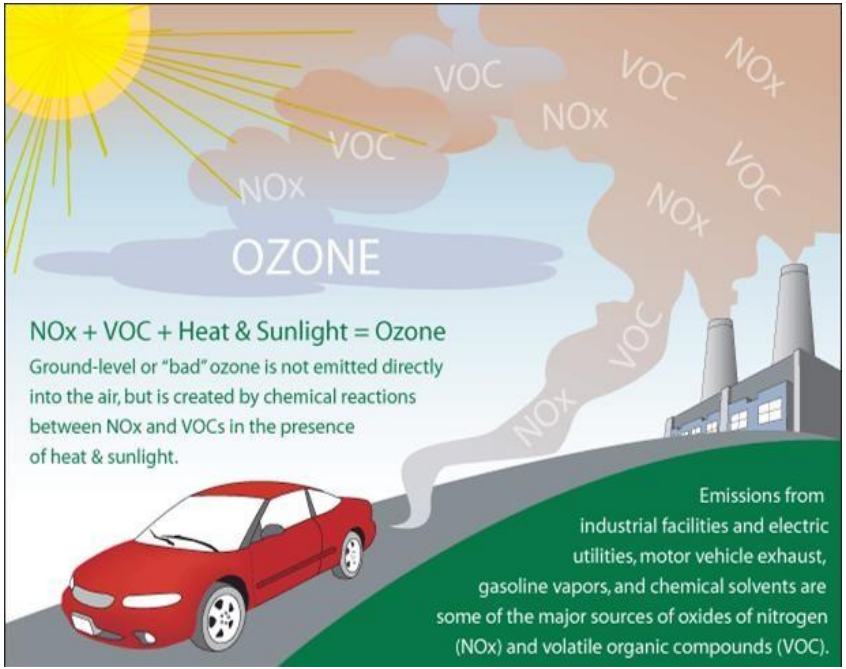
Air Quality Index: Pollutants

United States Environmental Protection Agency

Six criteria pollutants regulated by EPA

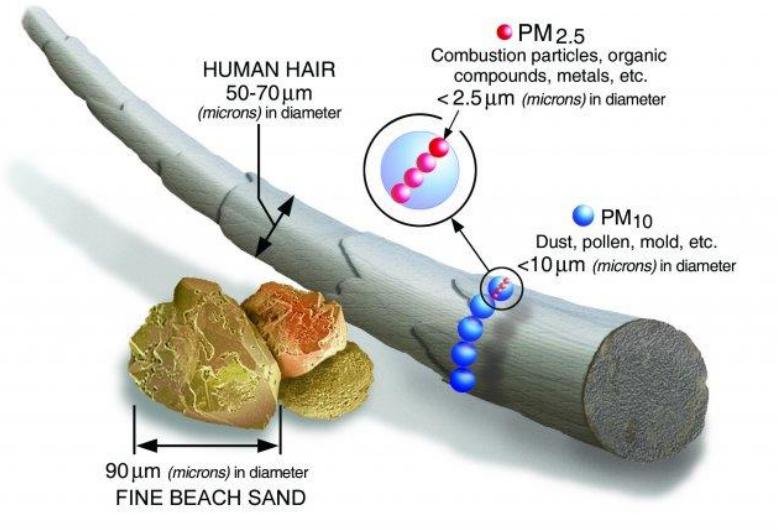
- Ground-level ozone
- Particulate matter
- Carbon monoxide
- Lead
- Sulfur dioxide
- Nitrogen dioxide

Air Quality Index: Pollutants



https://www.epa.gov/ground-level-ozone-pollution/ground-level-ozone-basics#formation

Air Quality Index: PM and NO2

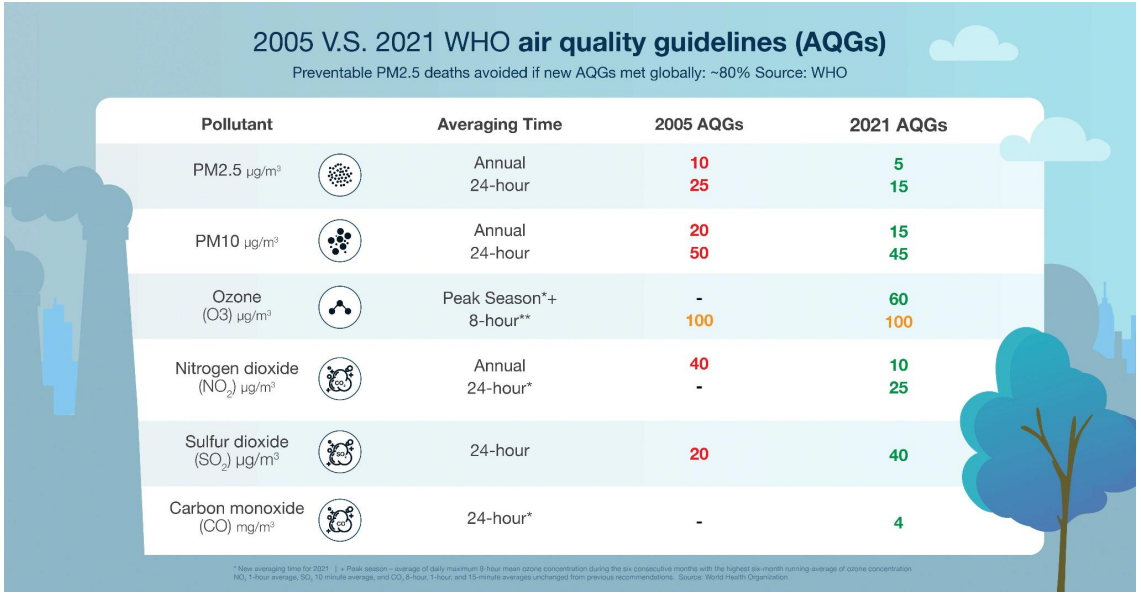


https://www.epa.gov/pm-pollution/particulate-matter-pm-basics#PM

https://phys.org/news/2018-03-german-deaths-nitrogen-dioxide.html

World Health Organization Guidelines

World Health Organization

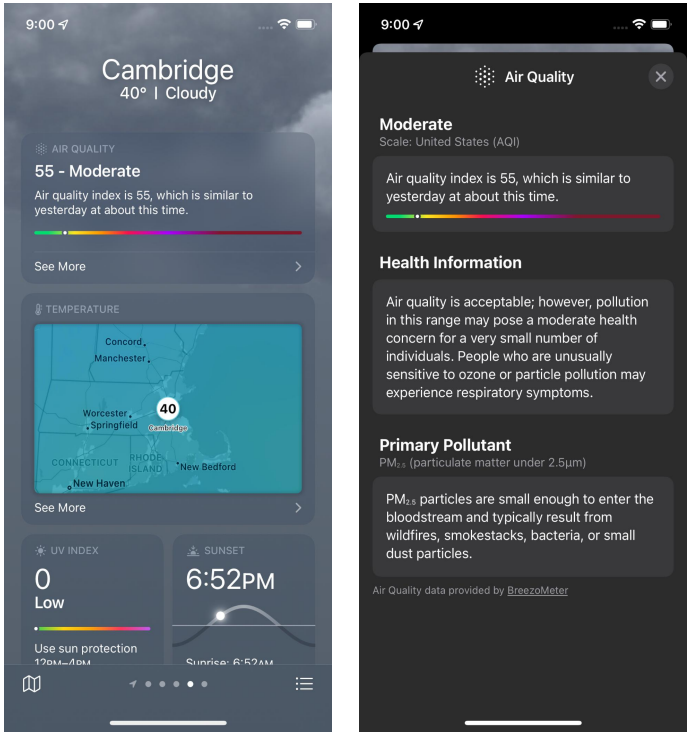


[World Health Organization]

Air Quality Index

Air Quality Index		
AQI Category and Color	**Index Value**	**Description of Air Quality**
Good Green	0 to 50	Air quality is satisfactory, and air pollution poses little or no risk.
Moderate Yellow	51 to 100	Air quality is acceptable. However, there may be a risk for some people, particularly those who are unusually sensitive to air pollution.
Unhealthy for Sensitive Groups Orange	101 to 150	Members of sensitive groups may experience health effects. The general public is less likely to be affected.
Unhealthy Red	151 to 200	Some members of the general public may experience health effects; members of sensitive groups may experience more serious health effects.
Very Unhealthy Purple	201 to 300	Health alert: The risk of health effects is increased for everyone.
Hazardous Maroon	301 and higher	Health warning of emergency conditions: everyone is more likely to be affected.

Air quality index ([United States Environmental Protection Agency])



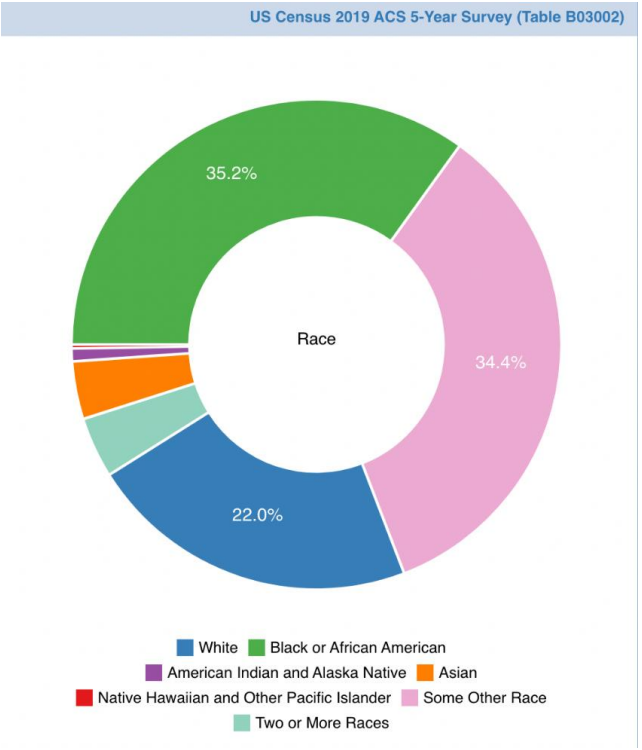
Air quality index shown in iPhone weather app

The Bronx

Historical + Environmental Context

The Bronx: Overview + Demographics

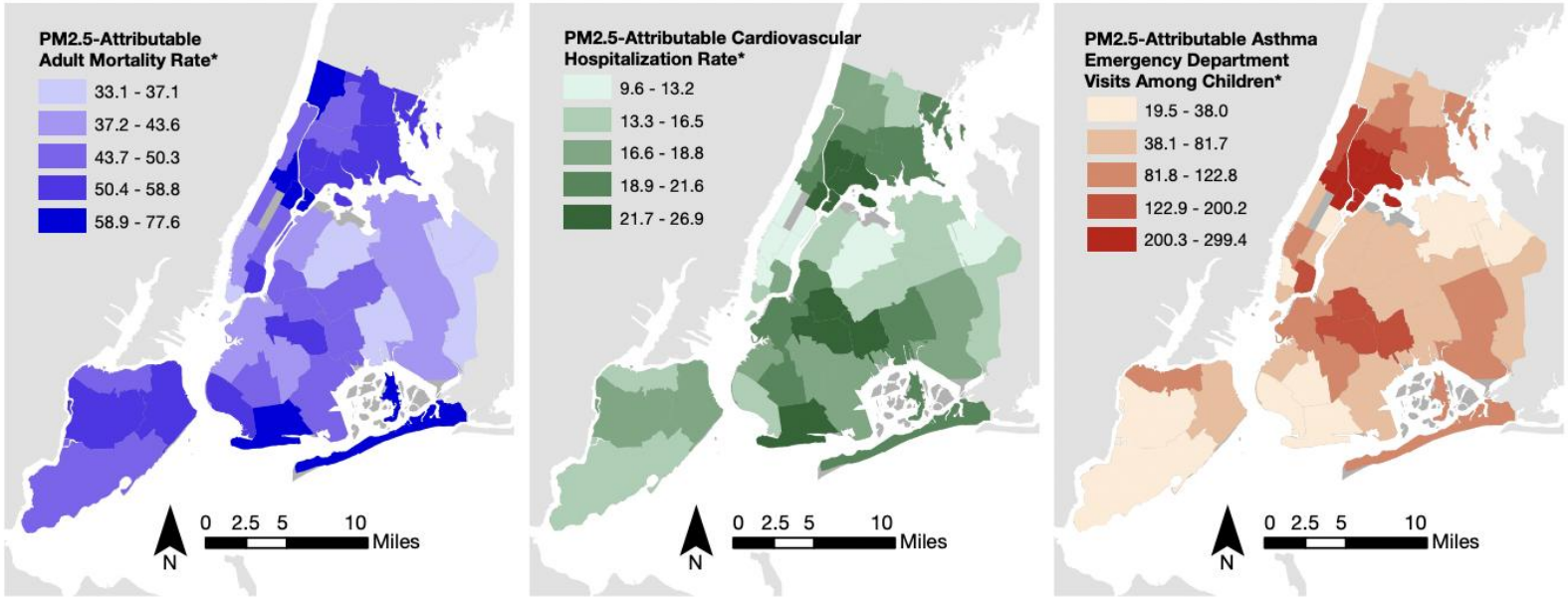
US Census 2019 ACS 5-Year Survey (Table B03002)



The Bronx: Context

The Bronx: Environmental Issues

Mortality and morbidity from selected conditions due to PM2.5 in New York City



* 2009-2011 Annual Average, Rate per 100,000 persons

The Bronx: Environmental Issues

Resources

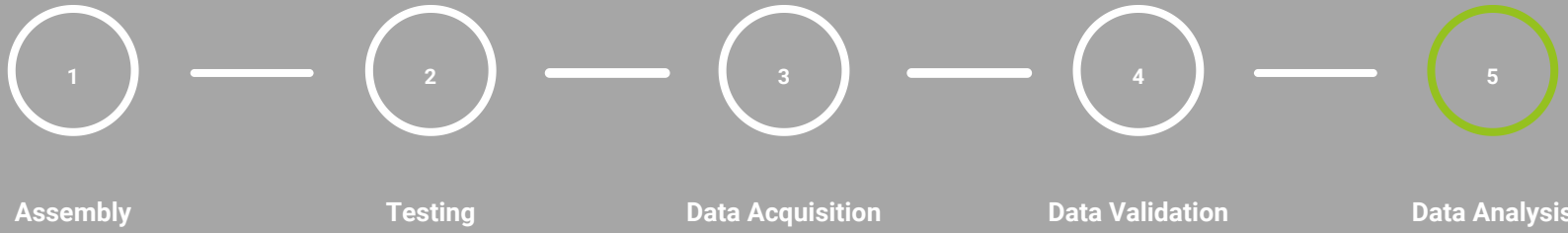
- NYCCAS Data
- South Bronx Environmental Health and Policy Study
- NYC Environmental and Health Portal
- New York Disadvantaged Communities Criteria
- Climate and Economic Justice Screening Tool
- NYC Environmental Justice Alliance
- NYC Community Health Profiles
- Potential Environmental Justice Areas in The Bronx

City Scanner

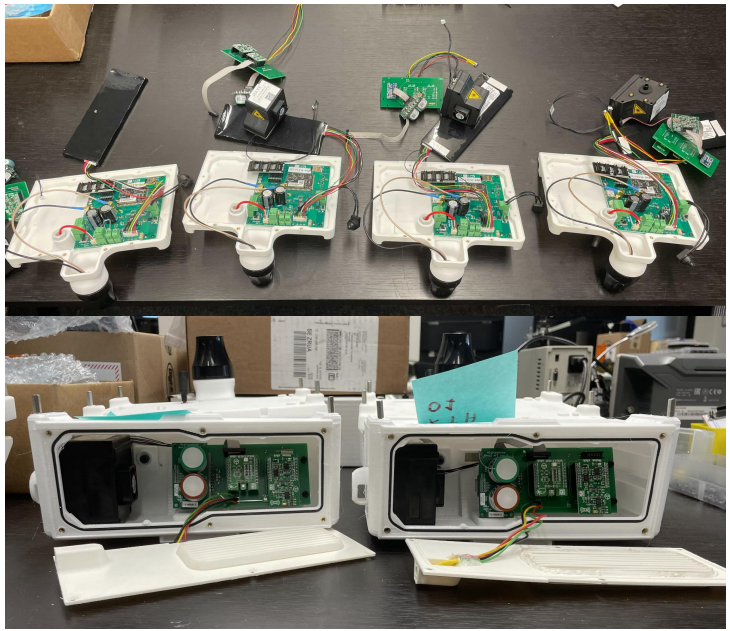
Data Collection Pipeline Tour

CITYSCANNER DATA COLLECTION PIPELINE TOUR

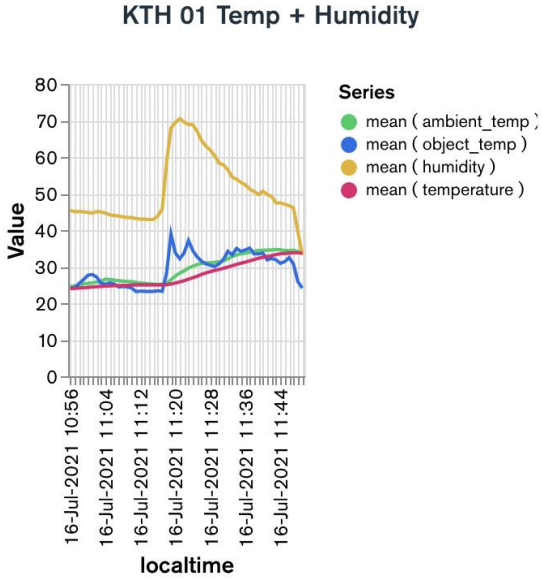
What does an environmental data collection pipeline look like in practice?
City Scanner: Data pipeline from hardware assembly to dynamic maps



Part 1: Assembly



Part 2: Testing



Part 3: Data Acquisition

Part 4: Data Validation

```
print("Total invalid timstamps for the collecti
```

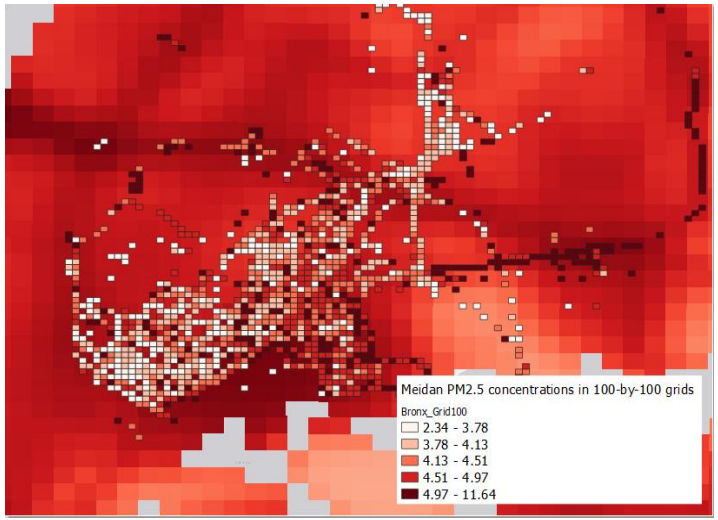
```
# of invalid timestamps for KTH01: 0
# of invalid timestamps for KTH02: 0
# of invalid timestamps for KTH03: 1
# of invalid timestamps for KTH04: 0
# of invalid timestamps for KTH05: 74
Total invalid timstamps for the collection: 75
```

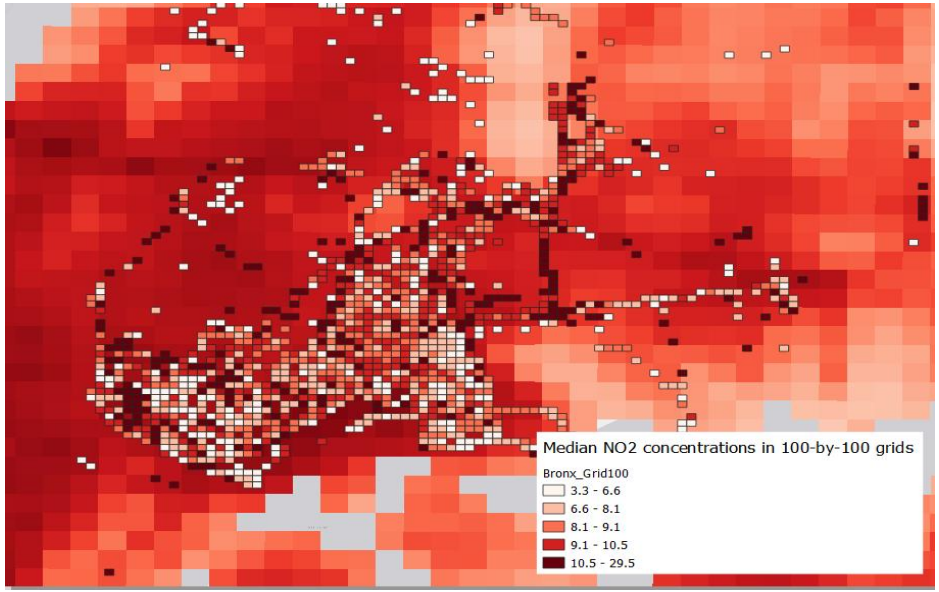
```
#find NAN/zero lat/lon values before filtering
##NOTE - this block must be run before data fil

zerolon1=(device1['longitude']==0).sum()
zerolat1=(device1['latitude']==0).sum()
print("Number of lat = 0 for KTH1:", zerolat1),
```

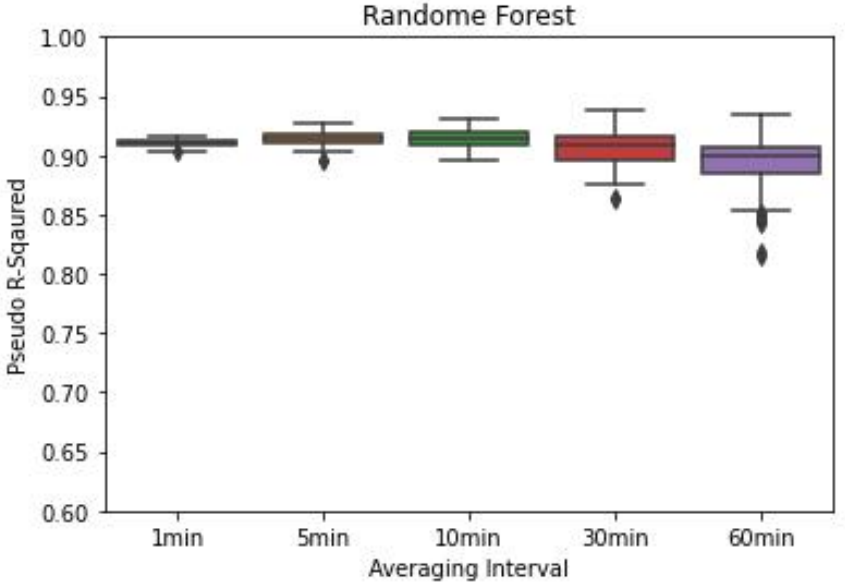
localtime	ambientIR	gas_op1_r	gas_op1_w	gas_op2_r	gas_op2_w
2021-07-26 00:00:00+00:00	27.894032	278.498540	285.698994	252.880636	420.083036
2021-07-27 00:00:00+00:00	29.039576	279.636421	286.367491	243.619637	396.145507
2021-07-28 00:00:00+00:00	24.004650	384.980391	388.208495	350.313130	528.279259
2021-07-29 00:00:00+00:00	23.968180	341.854801	346.847965	322.335673	425.801422
2021-07-30 00:00:00+00:00	21.251523	346.355689	352.861879	326.958869	473.290794
2021-07-31 00:00:00+00:00	21.646791	352.544802	352.500273	308.294471	776.294315

Part 5: Data Analysis

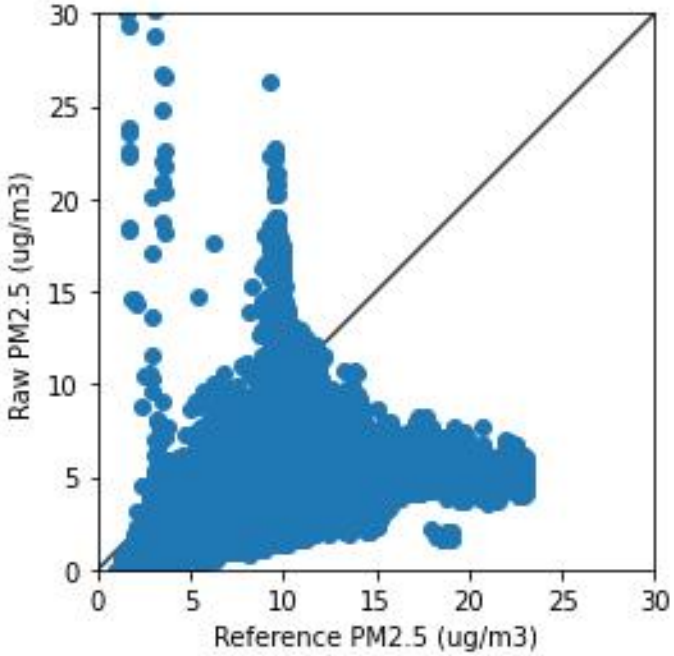




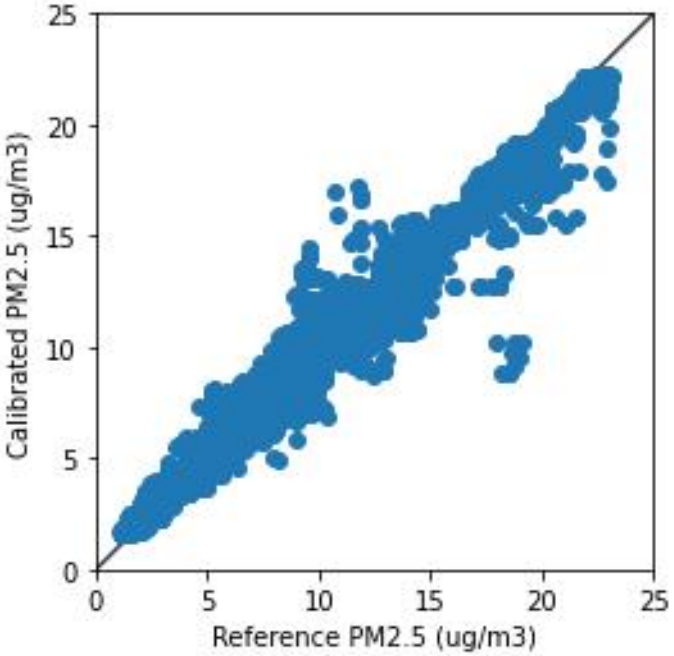
Analysis Method: Colocation + Calibration



Analysis Method: Colocation + Calibration



Before Calibration



After Calibration

Validation Method: Background Correction

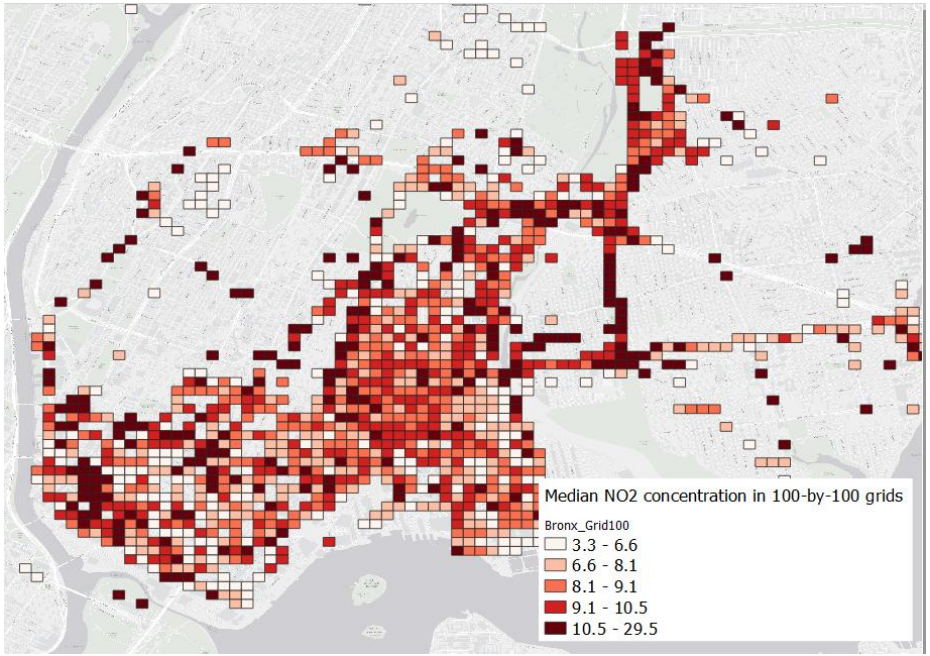
Background correction

- (Hourly) Multiplicative factor
- (Hourly) Lowest 10th percentile
- (Time series) Spline of minimums
- Background time-of-day correction

- Additive background correction factor

$$PM_{2.5,norm\ i} = PM_{2.5,OPC\ i} - PM_{2.5,\ bkg,i} + PM_{2.5,\ bkg,median}$$

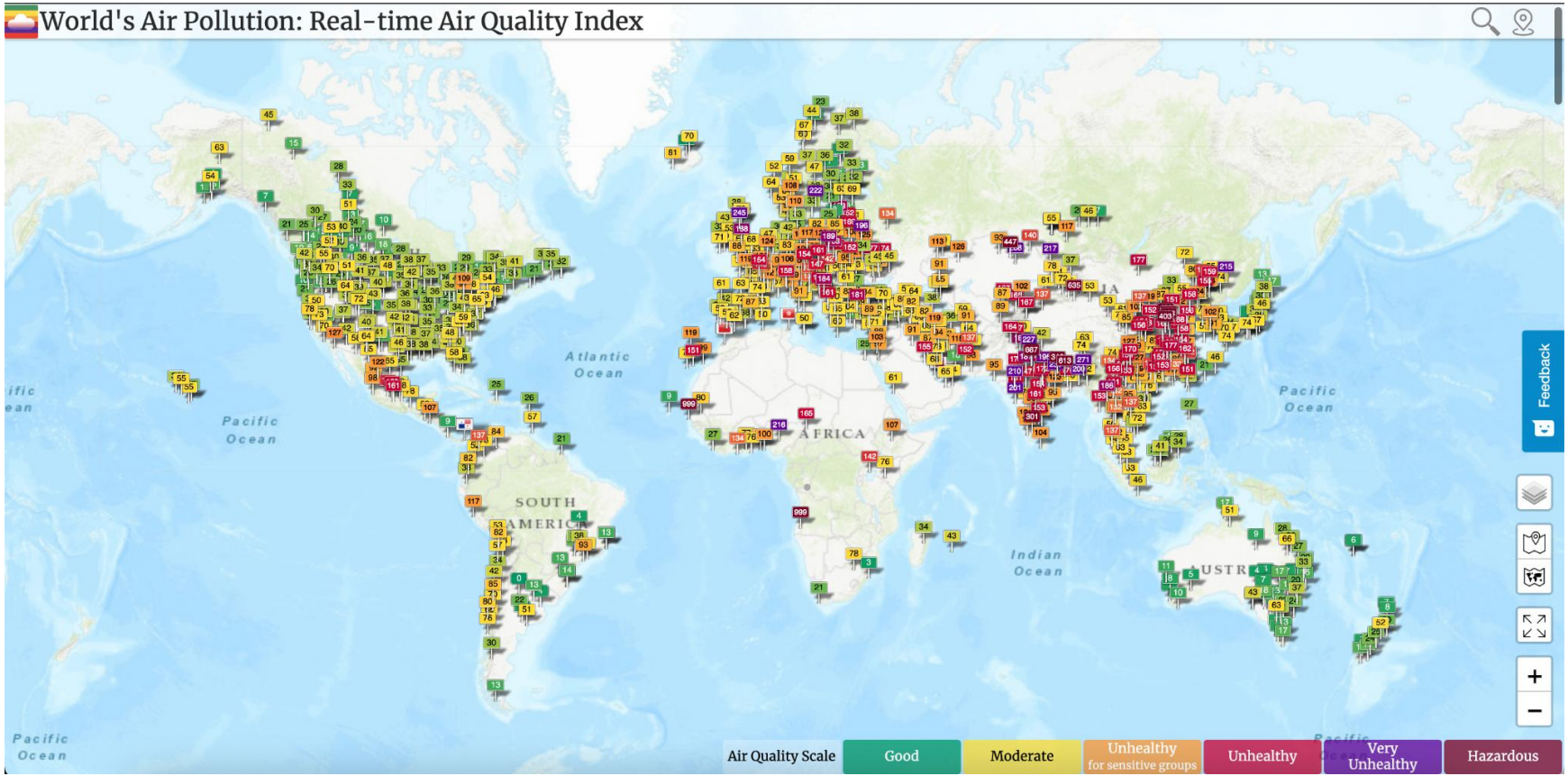
$$PM_{2.5,norm\ i} = PM_{2.5,OPC,\ i} \times PM_{2.5,\ bkg,median} / PM_{2.5,bkg,i}$$



Air Quality API Exercise

World Air Quality Index API

Accessing (Global) EPA Data



Accessing (Global) EPA Data

COLab Notebook
- Link: https://drive.google.com/drive/folders/1yF9dP7UStdY7DwDzfj9jZvaZnX9AZ1Kr

WAQI
- Site Link: https://waqi.info/
- API Link: https://aqicn.org/data-platform/token/

Get API Token:
- https://aqicn.org/data-platform/token/

Notebook Setup:
- Python libraries

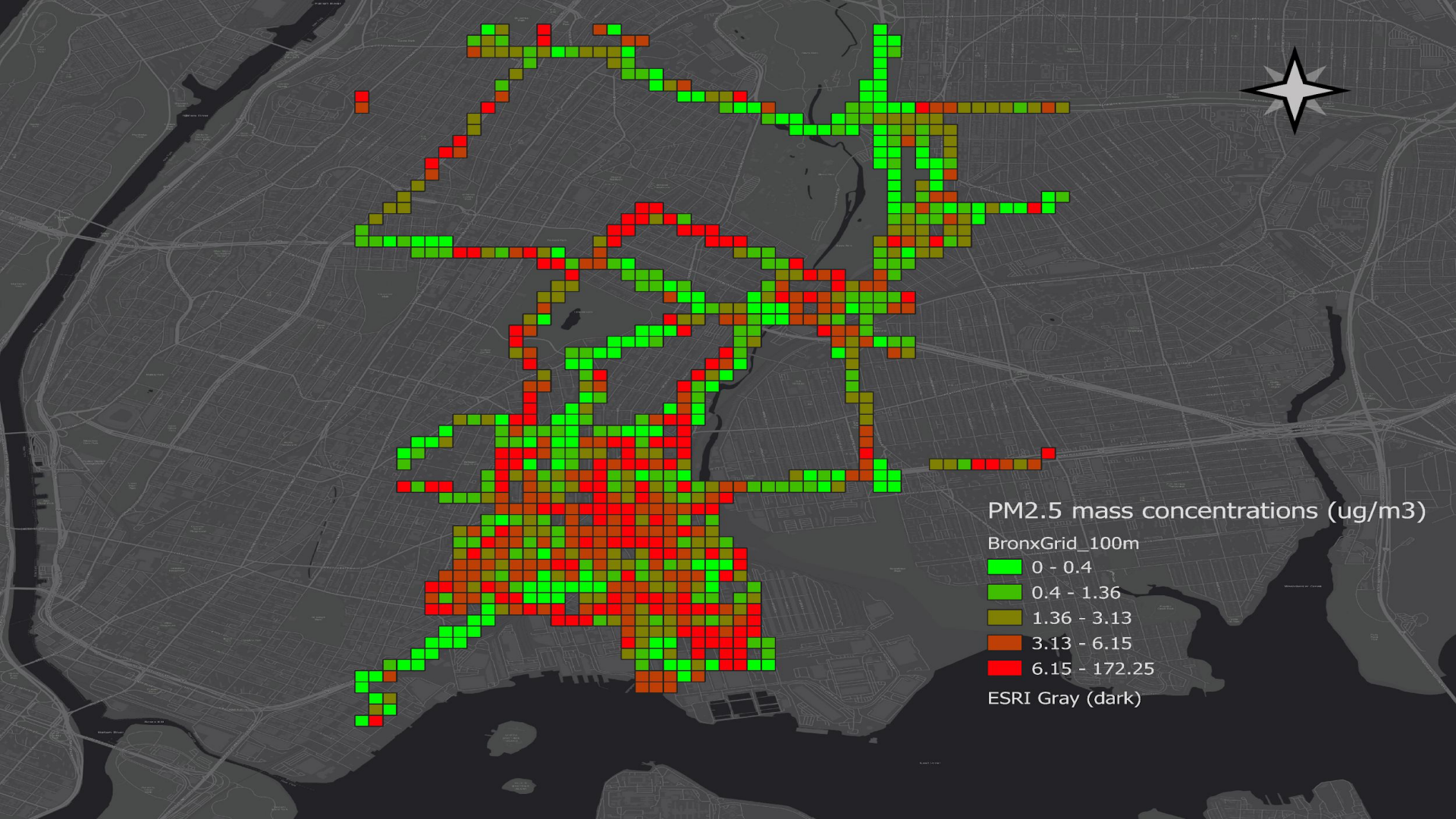
Authentication
- Input custom token into notebook

Getting the Data
- Run sample code
- Modify + experiment!

Bronx Data

CityScanner 2021 Deployment

	A	B	C	D	E	F	G	H	I	J
1	deviceID	time	lat	long	pm1	pm25	pm10	bin0	bin1	bin2
2	35003f00065	1579575798	40.807384	-73.943069	0.26	0.27	0.27	6	1	0
3	35003f00065	1579575800	40.807392	-73.943077	0.35	0.36	0.36	9	1	0
4	35003f00065	1579575802	40.8074	-73.943069	0.32	0.33	0.33	8	1	0
5	35003f00065	1579575804	40.8074	-73.943069	0.5	0.93	0.95	8	0	0
6	35003f00065	1579575815	40.807407	-73.943062	0.42	0.5	0.5	63	17	2
7	35003f00065	1579575806	40.807407	-73.943062	0.51	0.61	0.61	6	2	1
8	35003f00065	1579575816	3.001805	-73.943062	0.37	0.47	0.47	5	0	1
9	35003f00065	1579575822	40.807384	-73.943108	0.59	0.69	0.69	6	3	1
10	35003f00065	1579575820	40.807388	-73.9431	0.14	0.14	0.14	2	1	0
11	35003f00065	1579575818	40.807392	-73.943092	0.21	0.22	0.22	7	0	0
12	35003f00065	1579575824	40.807377	-73.943115	0.21	0.22	0.22	7	0	0
13	35003f00065	1579575828	40.807369	-73.943146	0.28	0.28	0.28	9	0	0
14	35003f00065	1579575826	40.807377	-73.94313	0.31	3.85	8.39	8	0	0
15	35003f00065	1579575830	40.807365	-73.943161	0.56	0.65	0.65	10	1	1
16	35003f00065	1579575832	40.807362	-73.943199	0.21	0.22	0.22	7	0	0
17	35003f00065	1579575834	40.80735	-73.94323	0.42	0.44	0.44	9	2	0
18	35003f00065	1579575836	40.807346	-73.943253	0.56	0.99	1.01	3	3	0
19	35003f00065	1579575838	40.807327	-73.943283	0.42	0.44	0.44	9	2	0
20	35003f00065	1579575842	40.807327	-73.943283	0.18	0.19	0.19	6	0	0
21	35003f00065	1579575840	40.807327	-73.943283	0.36	3.02	39.4	7	2	0
22	35003f00065	1579575844	40.807327	-73.943283	0.38	0.39	0.39	10	1	0
23	35003f00065	1579575846	40.807327	-73.943283	0.54	1.67	1.83	5	2	0
24	35003f00065	1579575848	40.807327	-73.943283	0.54	0.63	0.64	7	2	1
25	35003f00065	1579575850	40.807327	-73.943283	0.24	0.24	0.24	8	0	0
26	35003f00065	1579575852	40.807327	-73.943283	0.43	0.45	0.45	7	3	0
27	35003f00065	1579575854	40.807327	-73.943283	0.31	0.32	0.32	8	1	0
28	35003f00065	1579575856	40.807327	-73.943283	0.64	0.74	0.74	8	3	1
29	35003f00065	1579575858	40.807327	-73.943283	0.47	3.18	39.08	5	2	1
30	35003f00065	1579575860	40.807327	-73.943283	0.51	0.6	0.6	6	2	1
31	35003f00065	1579575862	40.807327	-73.943283	0.45	0.53	0.54	9	0	1


PM2.5 mass concentrations (ug/m3)
BronxGrid_100m
0 - 0.4
0.4 - 1.36
1.36 - 3.13
3.13 - 6.15
6.15 - 172.25
ESRI Gray (dark)

Data Description -- Common


```
deviceID:"1f004a000d504e5354303420"

time:1631239374

latitude:40.84767

longitude:-73.8693
```


deviceID:

- Unique identifier for each City Scanner Device

Time:

- Epoch time
- # of seconds that have elapsed since January 1 1970 (midnight UTC/GMT) not counting leap seconds

Latitude/Longitude:

- Unit: Degrees

Data Description -- raw data

```
bin0:5945
bin1:418
bin2:101
...
bin23:0
```

```
PM1:1.63
PM25:3.81
PM10:21.58
gas_op2_w:654
temperature:23.5
humidity:71.9
noise:32
```

24 Bins:
- Separate particle count by size
- Unit: # (count)

PM1:
- Particulate matter ~1 micron in diameter
- Units: ug/m3

PM2.5
- Particulate matter ~2.5 microns in diameter
- Units: ug/m3

PM10:
- Particulate matter ~10 microns in diameter
- Units: ug/m3

Gas_op2_w
- Electric signal for NO2
- Units: mv

temperature
- Ambient temperature
- Units: Degrees celsius

humidity
- Ambient humidity
- Units: % Relative humidity out of 100%

noise:
- Units: Voltage level in mV

Data Description -- Calibrated NO2

```
tmpf: 20

dwpf: 12.78

relh: 63.12

drct: 310

sknt: 7.20

mslp: 101.1

vsby: 16.1

feel: 20

Calib_logNO2: 2.43

Calib_NO2: 11.32

Spline_10min: 11.22

Spline_dmean: 4.75

Bckadj_NO2: 4.85
```

tmpf:
- Temperature at nearest weather station
- Units: Degrees celsius

dwpf:
- Dewpoint at nearest weather station
- Units: Degrees celsius

relh:
- Relative humidity at nearest weather station
- Units: %

drct:
- Wind direction with reference to the true north as 0
- Units: Degrees

sknt:
- Wind speed at nearest weather station
- Units: m/s

mslp:
- Air pressure at nearest weather station
- Units: kpa

vsby:
- Visibility at nearest weather station
- Units: km

feel:
- Feel like temperature at nearest weather station
- Units: Degrees celsius

Calib_logNO2:
- Calibrated NO2 in log form
- Units: log ppb

Calib_NO2
- Calibrated NO2
- Units: ppb

Spline_10min
- Spline regressed NO2 using 10 min minimum values
- Units: ppb

Spline_dmean
- Daily median NO2
- Units: ppb

Bckadj_NO2
- Background adjusted NO2 after calibration
- Units: ppb

Data Description -- Calibrated PM2.5

```
tmpf: 20

dwpf: 12.78

relh: 63.12

drct: 310

sknt: 7.20

mslp: 101.1

vsby: 16.1

feel: 20

Calib_logPM: 1.34

Calib_PM:3.82

Spline_10min: 4.37

Spline_dmean: 4.02

Bckadj_PM: 3.47
```

tmpf:
- Temperature at nearest weather station
- Units: Degrees celsius

dwpf:
- Dewpoint at nearest weather station
- Units: Degrees celsius

relh:
- Relative humidity at nearest weather station
- Units: %

drct:
- Wind direction with reference to the true north as 0
- Units: Degrees

sknt:
- Wind speed at nearest weather station
- Units: m/s

mslp:
- Air pressure at nearest weather station
- Units: kpa

vsby:
- Visibility at nearest weather station
- Units: km

feel:
- Feel like temperature at nearest weather station
- Units: Degrees celsius

Calib_logPM:
- Calibrated PM in log form
- Units: log ug/m3

Calib_PM
- Calibrated PM
- Units: ug/m3

Spline_10min
- Spline regressed PM using 10 min minimum values
- Units: ug/m3

Spline_dmean
- Daily median PM
- Units: ug/m3

Bckadj_PM
- Background adjusted PM after calibration
- Units: ug/m3

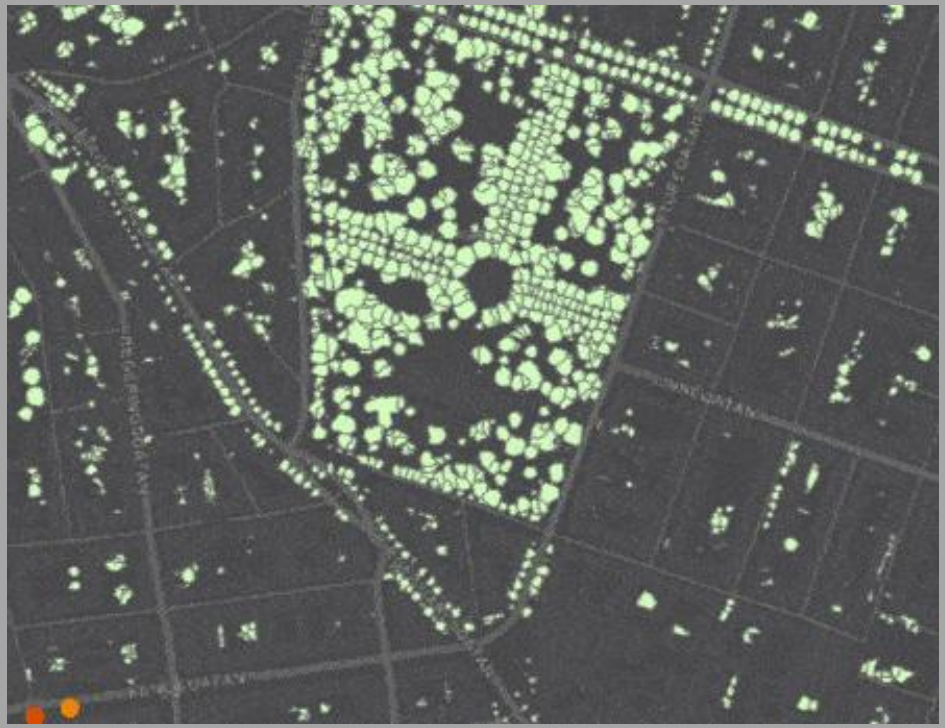
Activity: Methodology

Understanding Spatial Environmental Patterns

DATA ANALYSIS METHODS

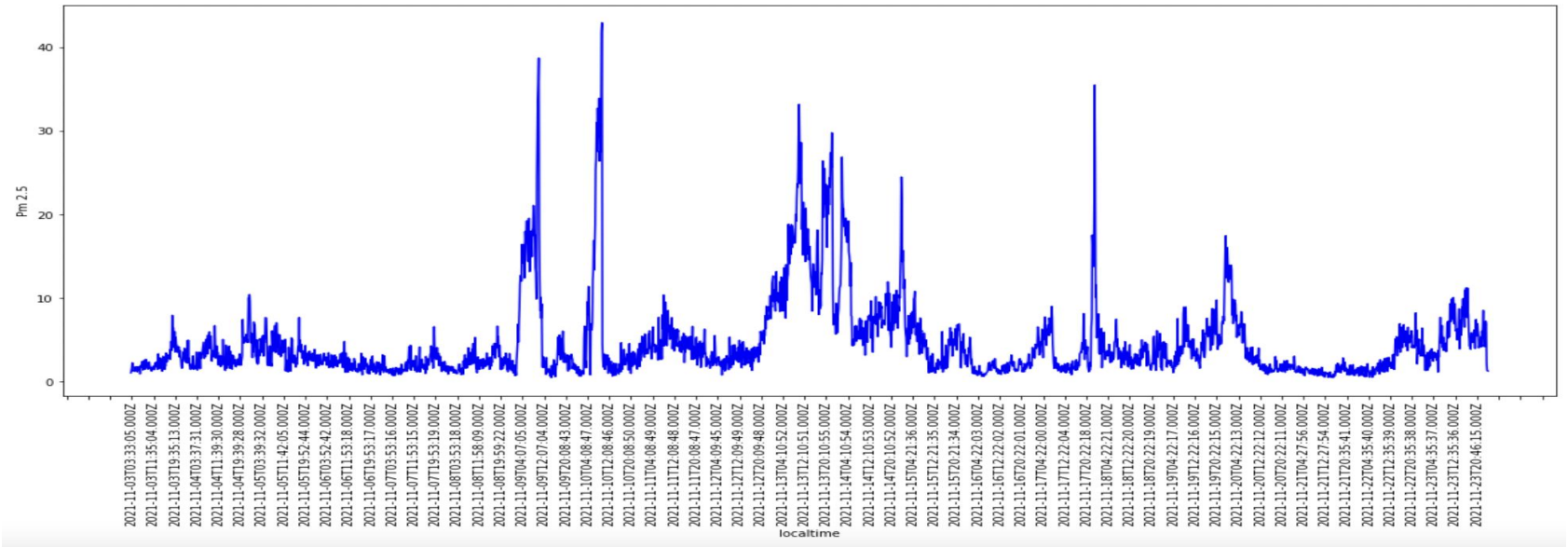
How do we understand the insights this data can provide?

Overview of key CityScanner analysis practices



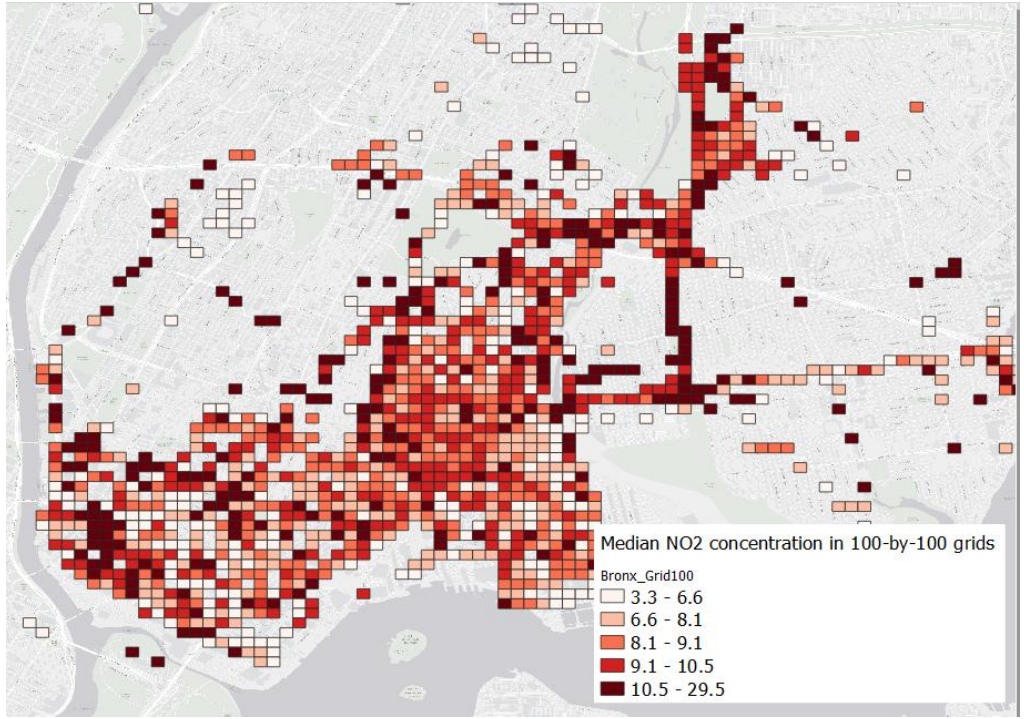
Analysis Method: Time Series

- Parameter as a function of time



Analysis Method: Mapping

- What can maps tell us?
- What info do you need to make a map?
- What tools can you use for mapping?



Analysis Method: Hotspot Detection

"A hot-spot analysis is defined in 40 CFR 93.101 as an estimation of likely future localized pollutant concentrations and a comparison of those concentrations to the relevant NAAQS." - United States EPA

Analysis Method: Hotspot Detection

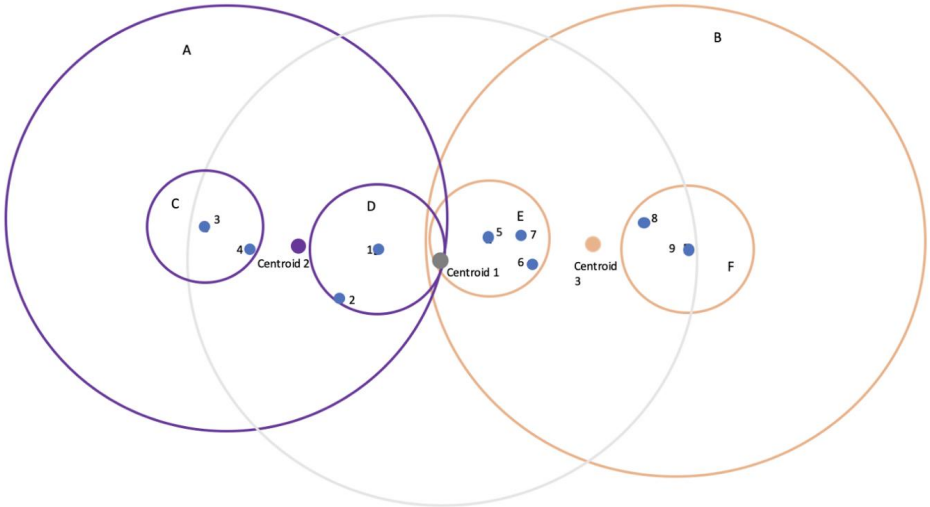
Clustering

- DBSCAN (from scikit learn)
  - Density-Based Spatial Clustering of Applications with Noise
  - Finds core samples of high density and expands clusters from them
  - Good for data which contains clusters of similar density.

Analysis Method: Hotspot Detection

Clustering

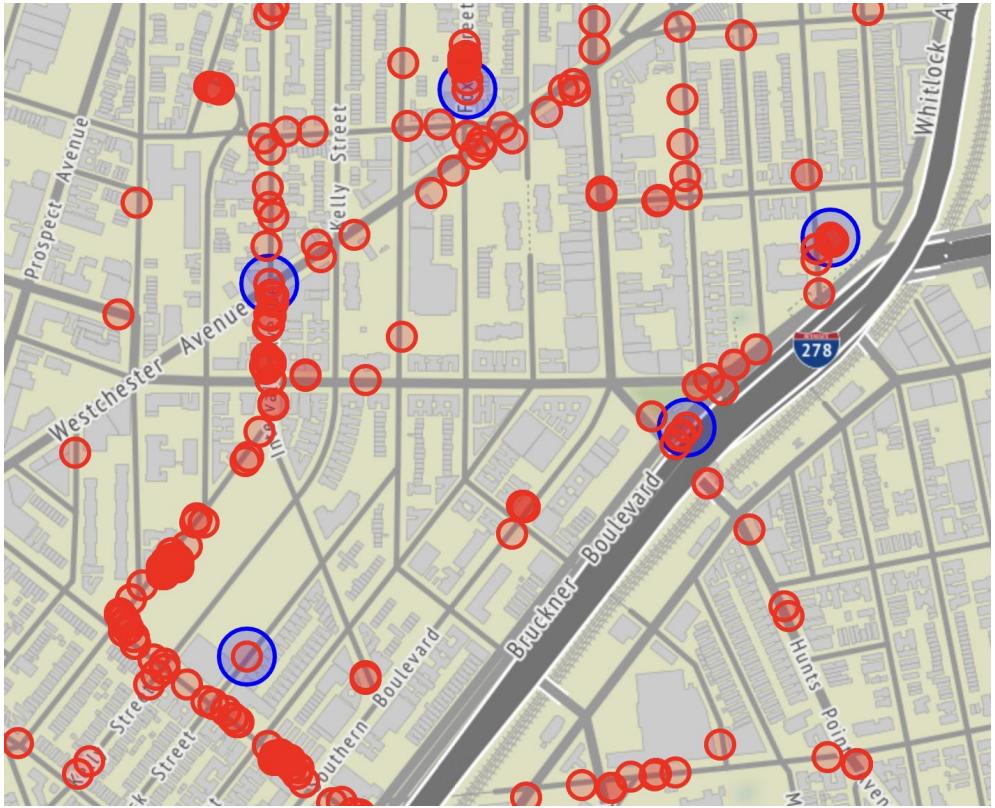
- DBSCAN → Ball_tree algorithm
- Divides groups of points into clusters until desired size is reached



Analysis Method: Hotspot Detection

Hotspot Detection

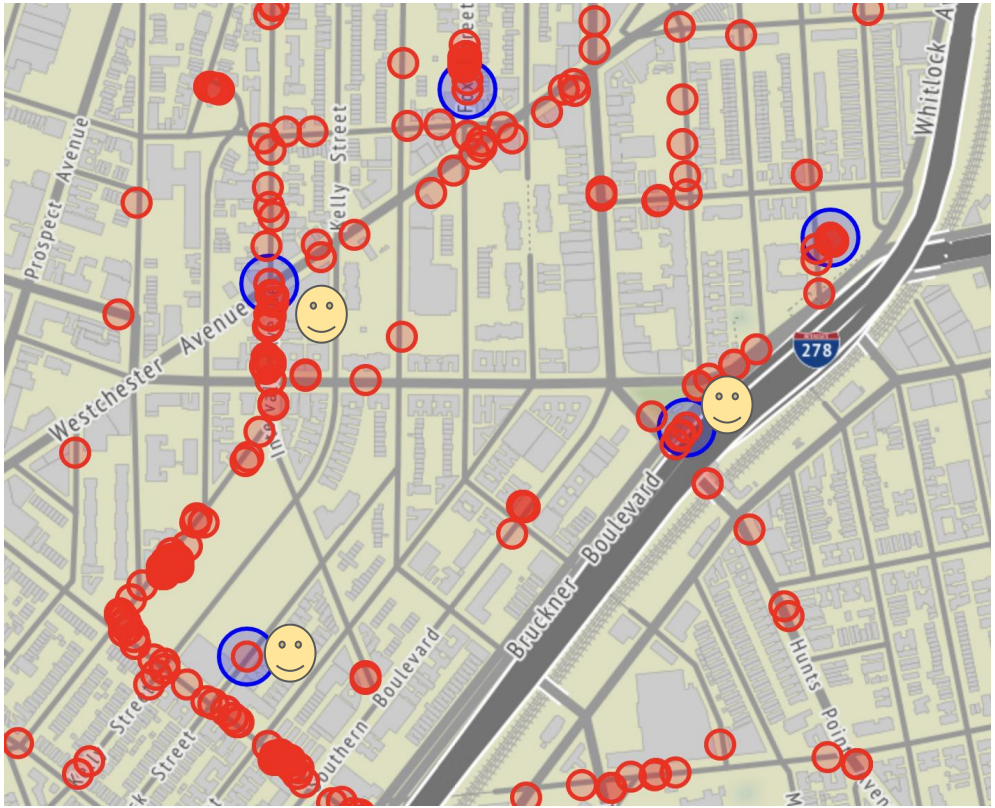
- Bottom-up hierarchical clustering “agglomerative”
- Merge clusters of data into smaller clusters



Analysis Method: Personal Exposure

Twitter Data Fusion

- Combine twitter data with Cityscanner data to better understand personal exposure
  - Count the number of people exposed to a certain hotspot



Discussion

Understanding Spatial Environmental Patterns