Locality Lookout

Identifying and predicting complaints in your locale

Dahlia Nadkarni

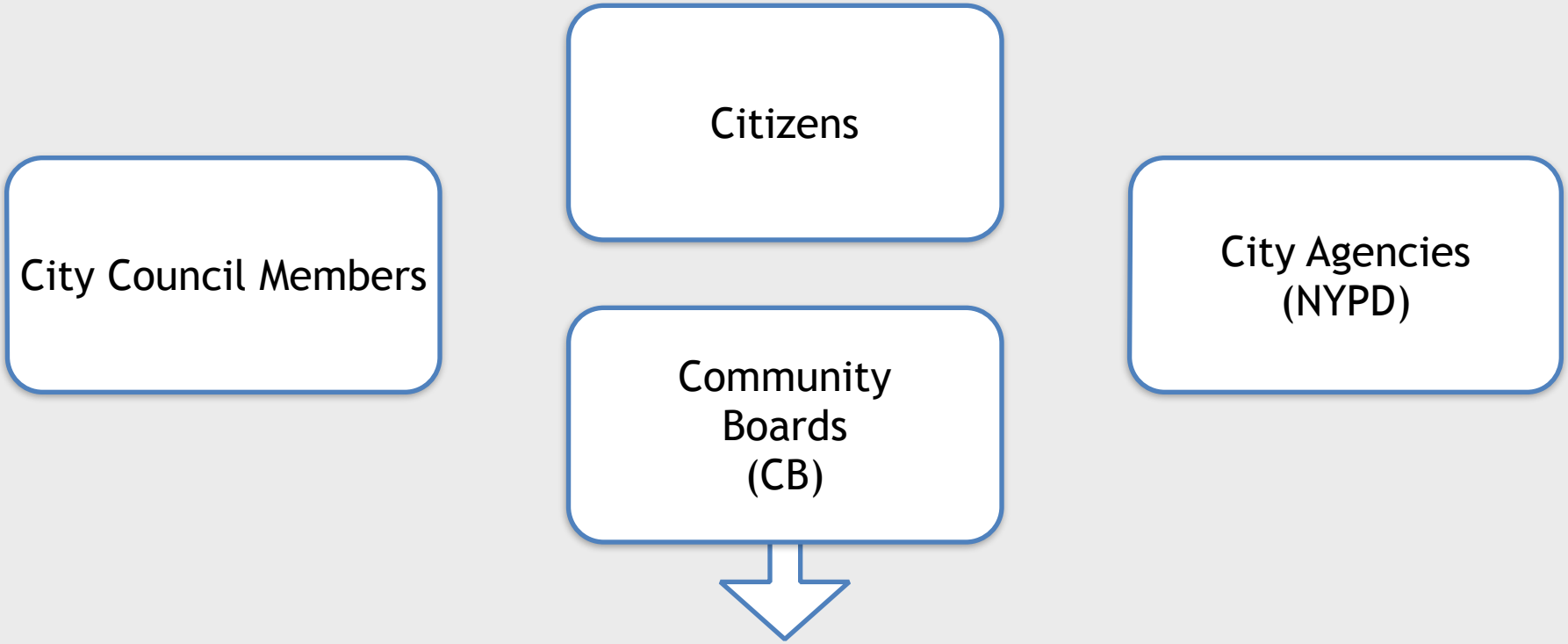
Complaint dataset

- Non-emergency complaints: Noise, Sewer, Rodent, Mold, Heating…
- Around 8 million complaints registered since 2010
- Location: 5 borough, ~ 60 districts, ~200 zip code, lat-long
- Timestamps: created, resolved, due date

Unique Key	Created Date	Closed Date	Agency	Complaint Type	Incident Zip	City	Status	Due Date	Resolution Action Updated Date	Latitude	Longitude
25416040	2014-07-31 08:31:20	07/29/2013 12:00:00 AM	DSNY	Graffiti	11213	BROOKLYN	Open	08/30/2014 08:31:20 AM	07/31/2014 08:31:20 AM	40.671239	-73.928093
25409622	2014-07-31 08:27:35	07/29/2013 12:00:00 AM	DSNY	Graffiti	10458	BRONX	Open	08/30/2014 08:27:35 AM	07/31/2014 08:27:35 AM	40.856835	-73.888855
28582487	2014-07-31 02:20:50	NaN	DOT	Street Condition	NaN	NaN	Open	NaN	NaN	NaN	NaN
28581269	2014-07-31 02:12:05	NaN	NYPD	Blocked Driveway	11377	WOODSIDE	Open	07/31/2014 10:12:05 AM	NaN	40.736517	-73.919085
28582144	2014-07-31 02:10:22	07/31/2014 02:48:27 AM	NYPD	Illegal Parking	10454	BRONX	Closed	07/31/2014 10:10:22 AM	07/31/2014 02:47:07 AM	40.809466	-73.917764

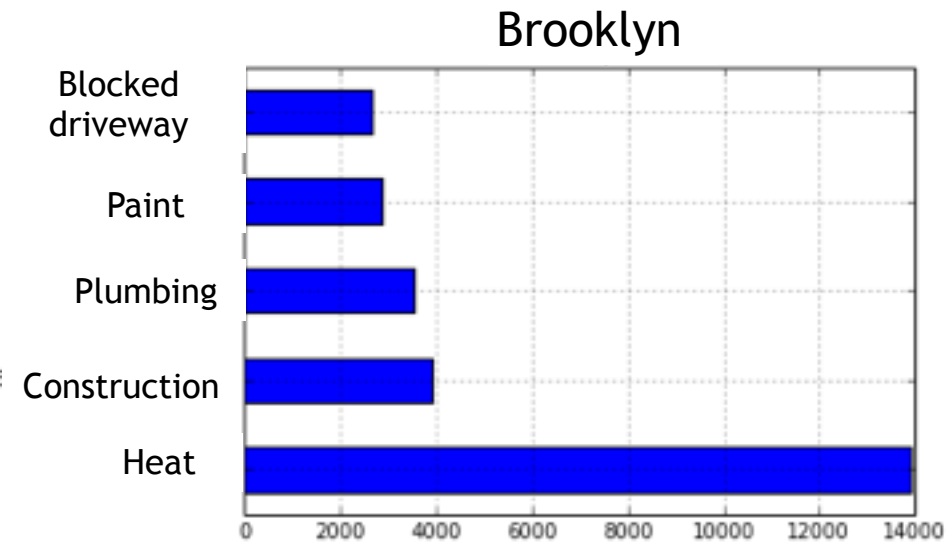
This is after preprocessing

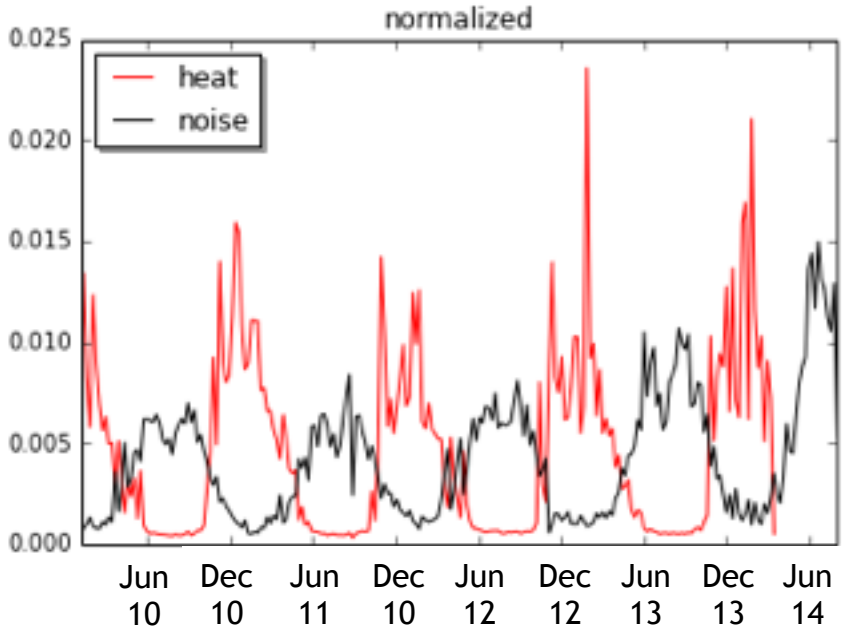
Who uses this data?



- **Advocacy**: “I can do better advocacy, if I had data that is easier to look at.”
- **Efficiency**: “If we had easier to understand 311 data, we could have more productive CB meetings.”
- **Effectiveness**: "I could set better agenda topics” -Brooklyn CB6 manager

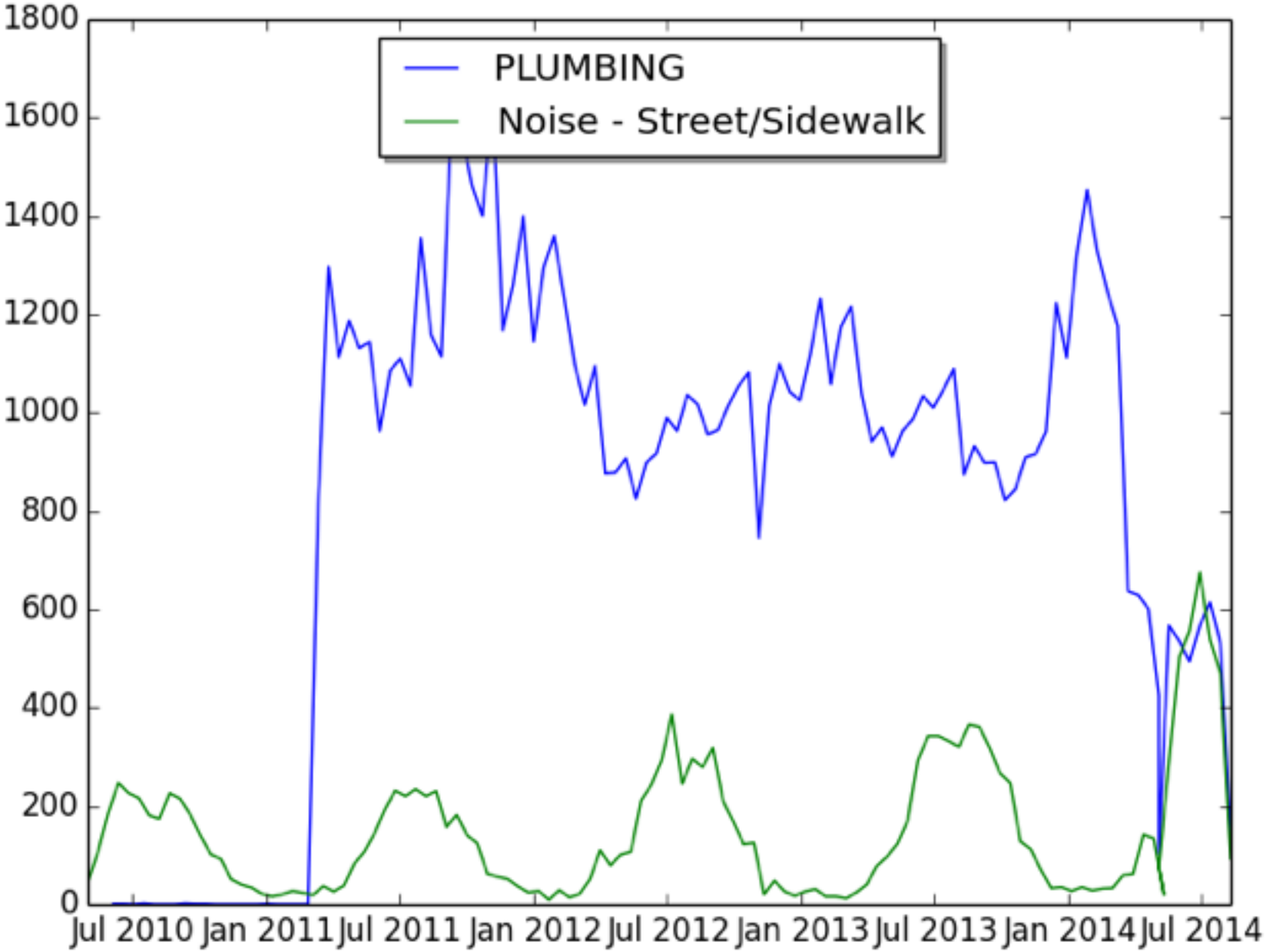
Identifying high-priority problems





How do you quantify ‘high-priority’?

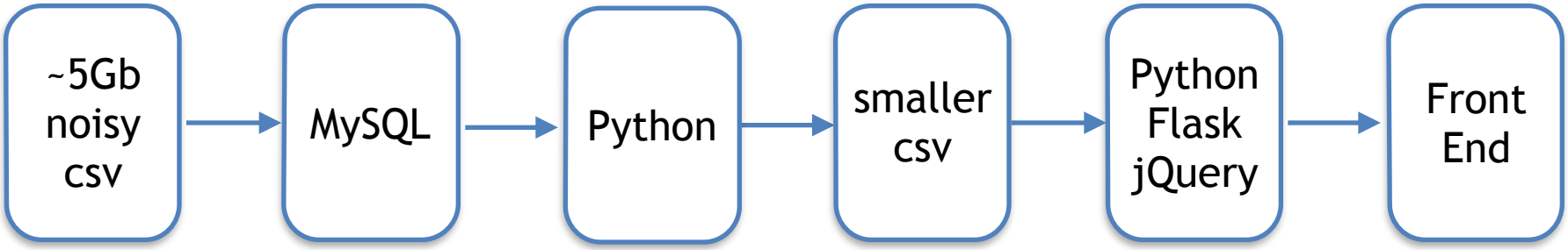
BRONX complaints in last 3 months

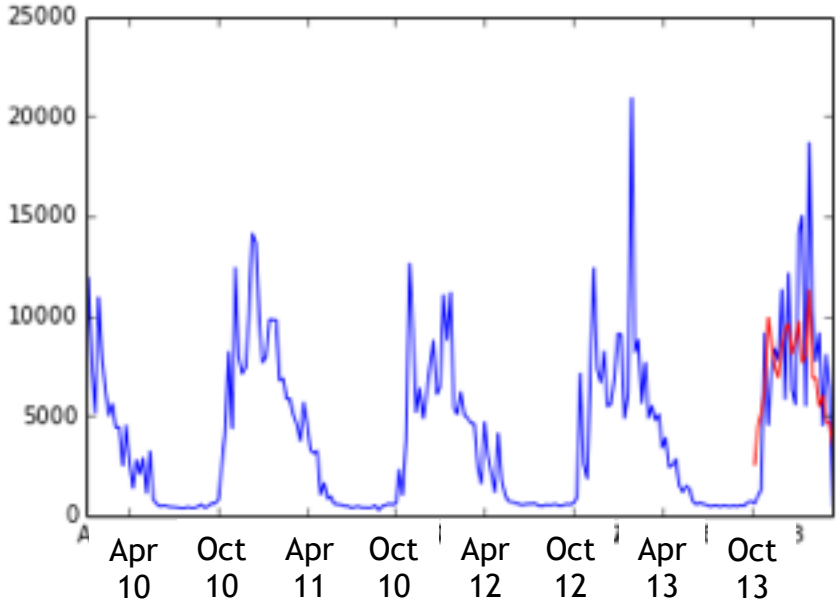


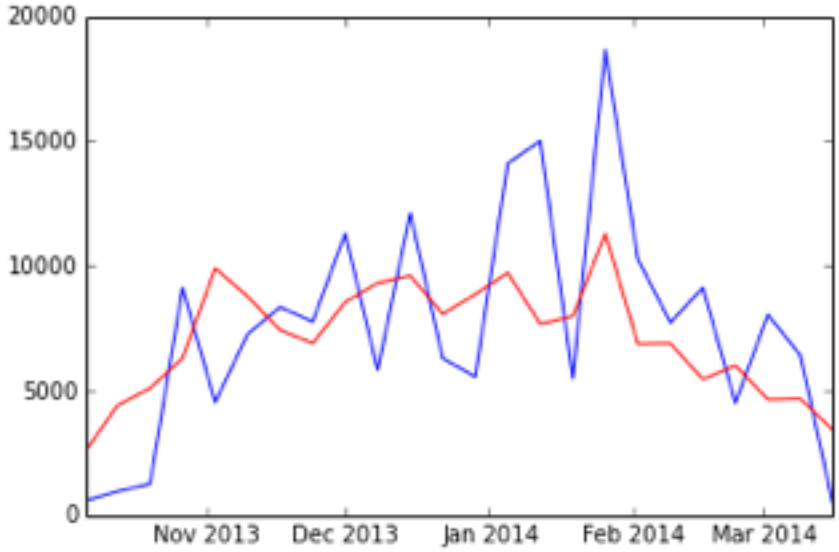
Behind scenes

- Model the time series and compare expected & actual volumes
- Priority = $\frac{(y - \hat{y})}{\hat{\sigma}_y}$
- Linear regression (plain, ridge, lasso) & Gaussian processes
- Features:
  - Categorical - for periodicity
  - Linear - for short term trends
- Predict daily, weekly, monthly volumes, use the best model
- Predict future complaint volumes - useful to allocate resources

Behind scenes



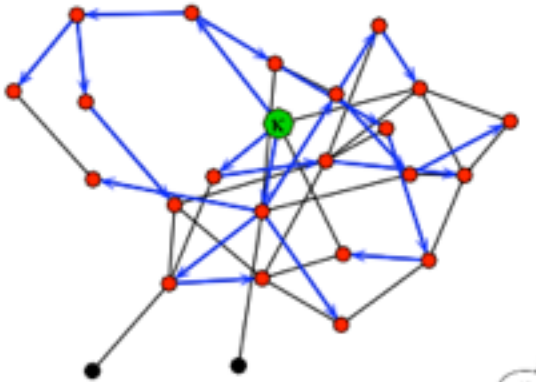


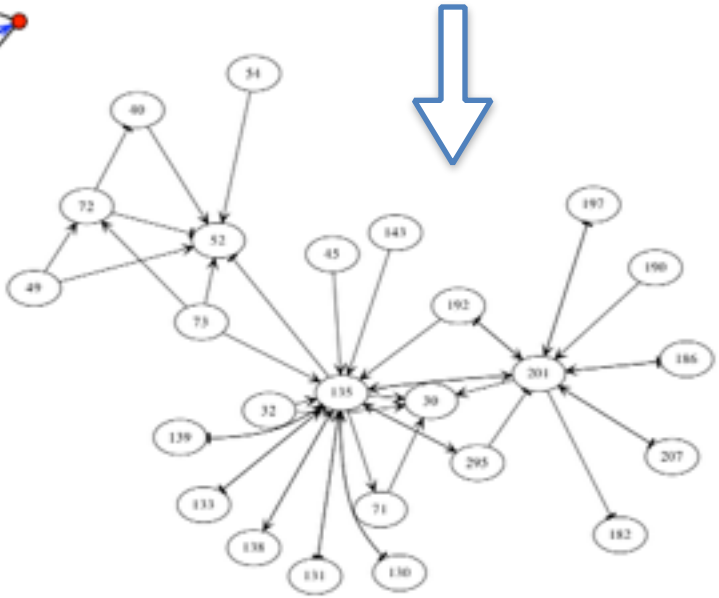


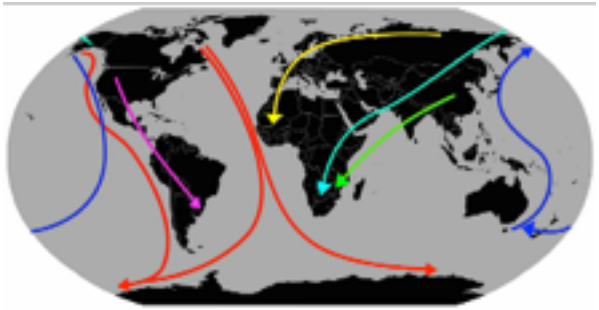
Heat complaints: Blue - actual volumes, Red - predicted

Dahlia Nadkarni

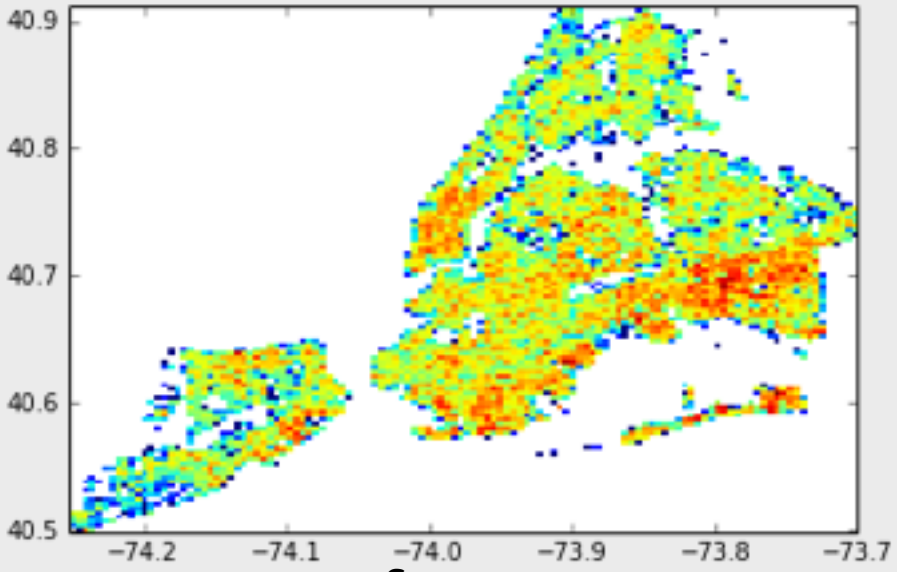
BROWN



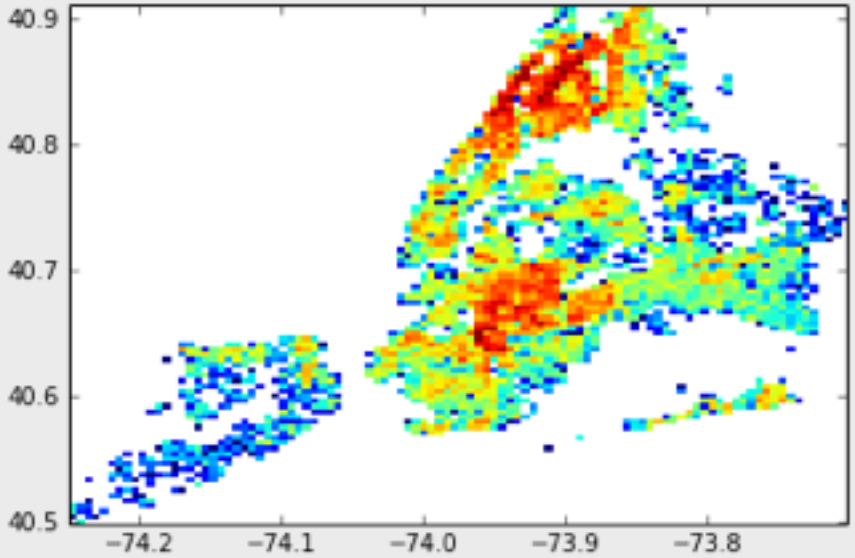




Thank you



Sewer



General Construction