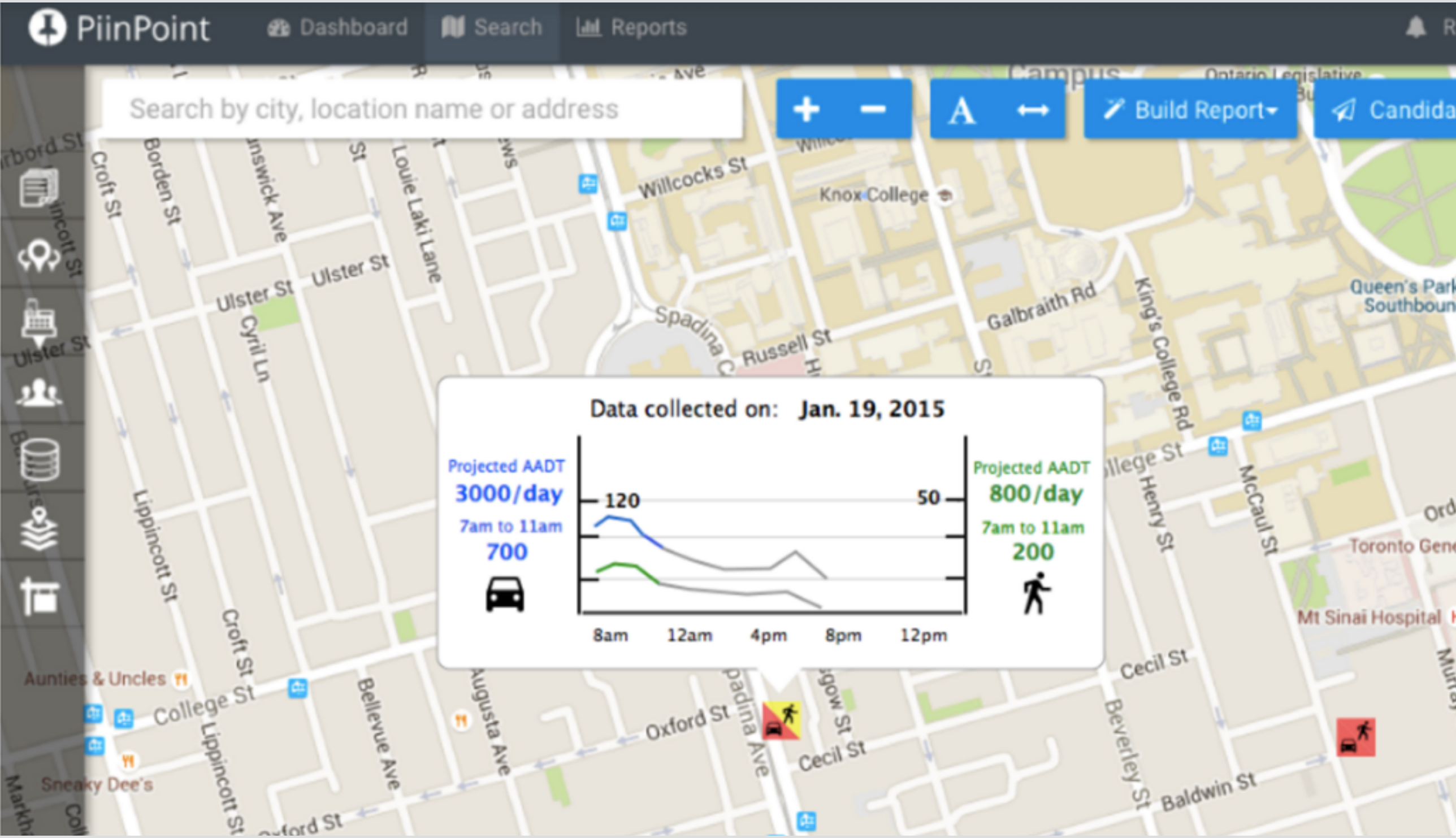
Count `em Up:

Traffic estimation for Piinpoint

Ross Fadely

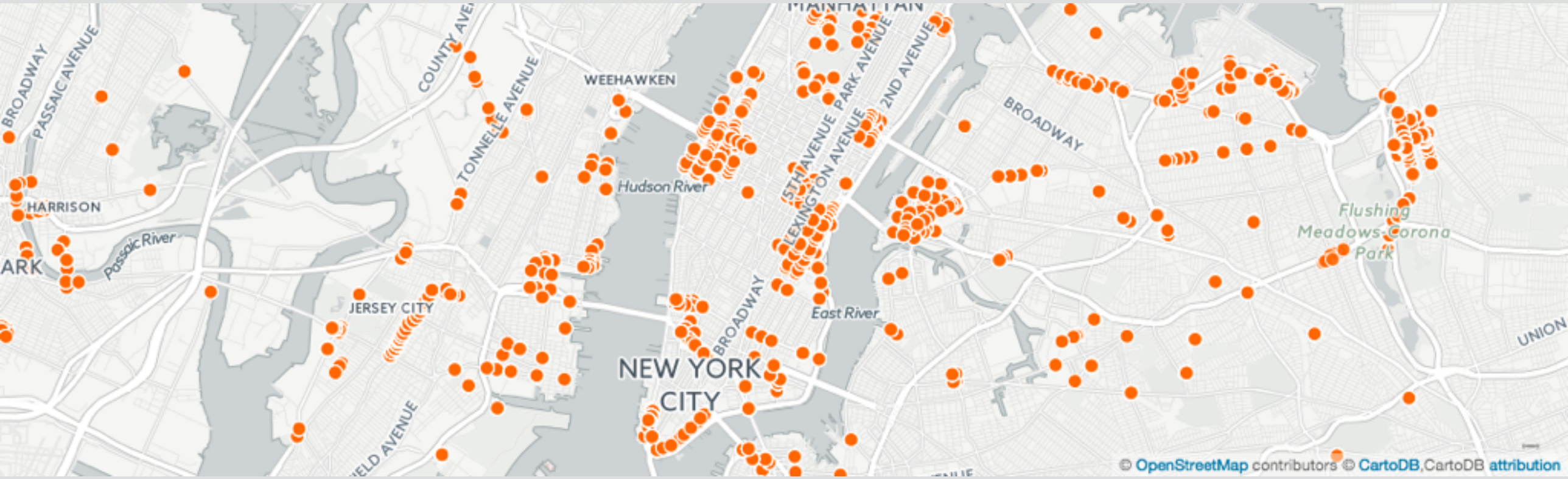


PiinPoint has you covered*

*except…

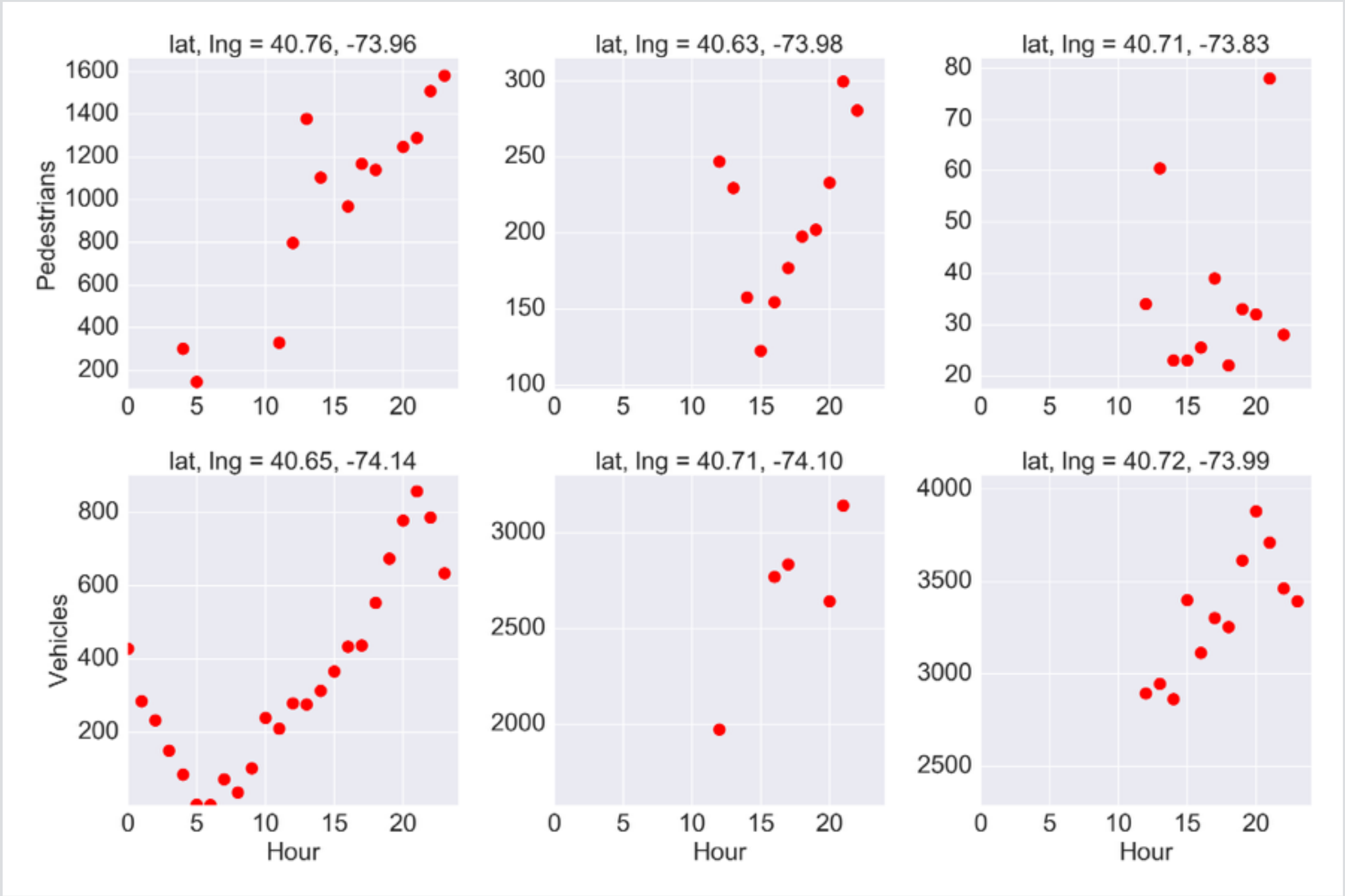
Data

WHOA!!



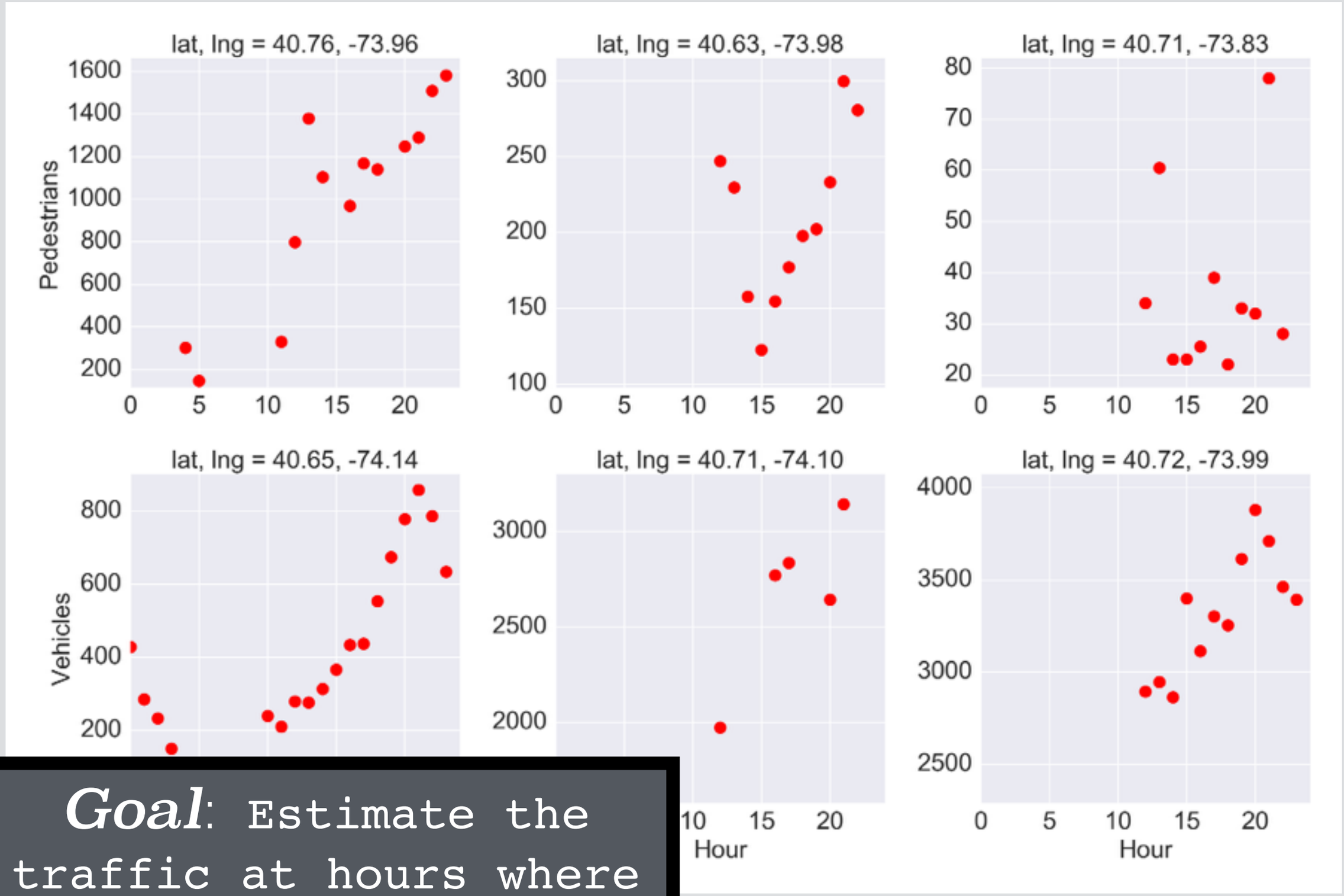
(very sparse in space)

(slightly) crazy talk



(also sparse in time)

(slightly) crazy talk

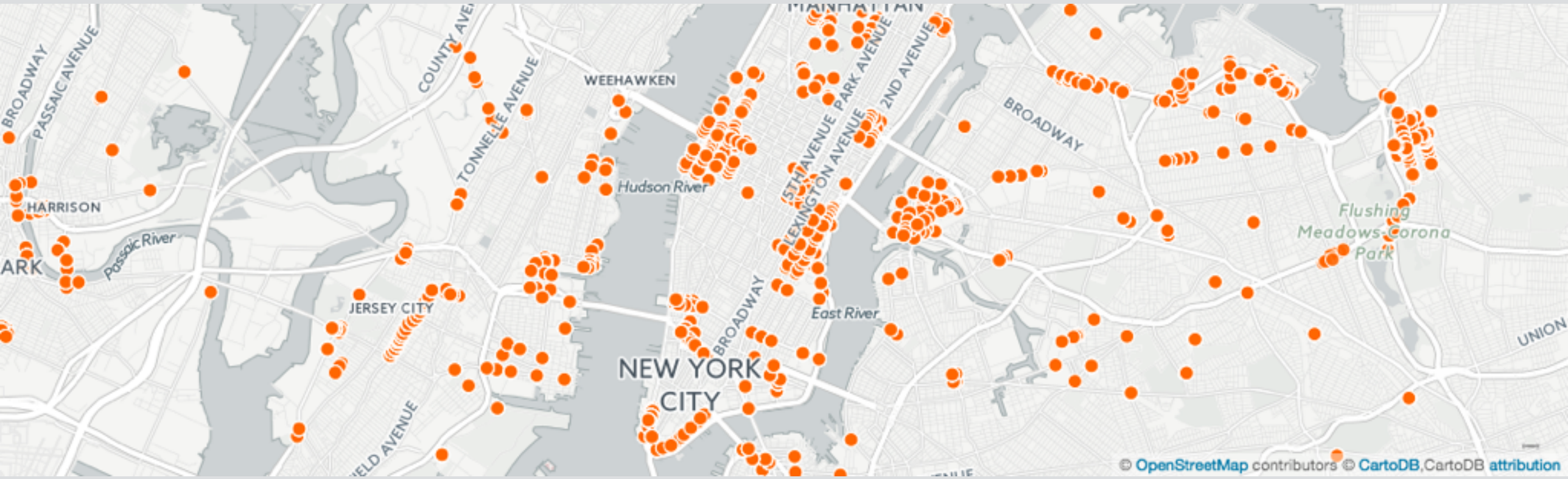


Goal: Estimate the traffic at hours where there is **NO** data.

(also sparse in time)

Approach

First Intuitions, Later Tweaks



- kNN can capture the local structure of the problem
- Find spatial neighbors for current, previous, and subsequent hours, compute weighted sum.

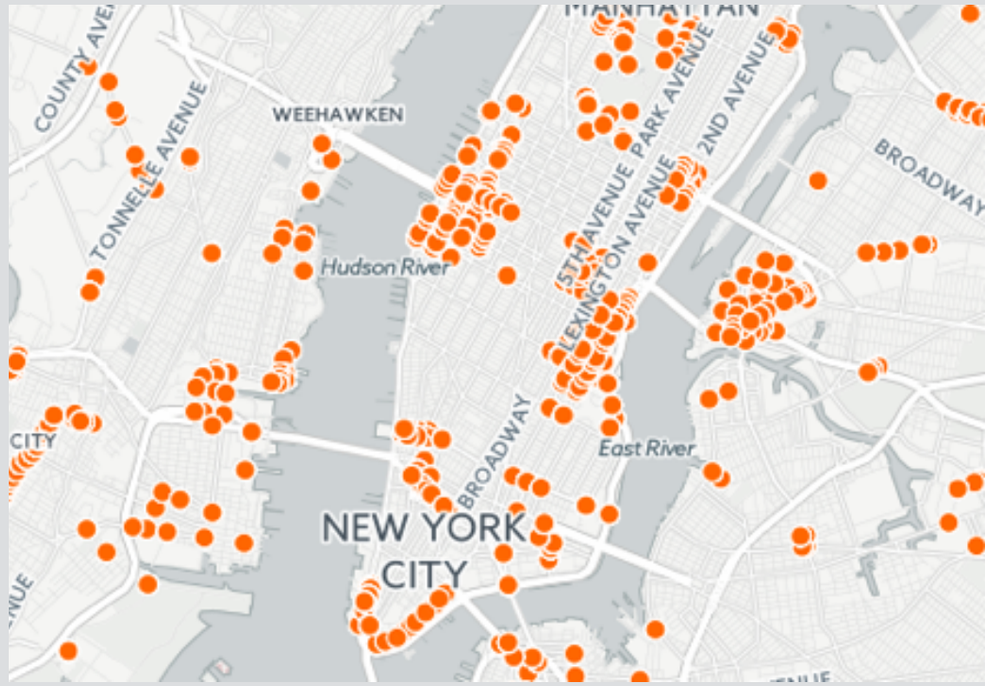
Test RMSE 5% better than PiinPoint

- Its ok, but can we do better? Try Random Forest.

Test RMSE 63% better than PiinPoint

Filling in the gaps:

Augmentation with other data

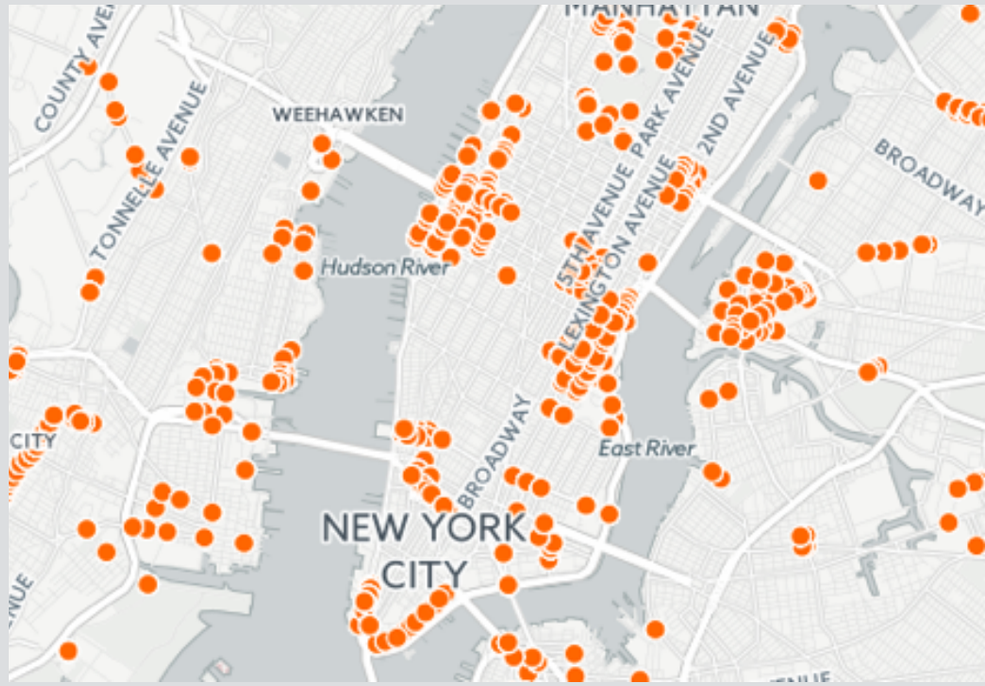


Filling in the gaps:

Augmentation with other data

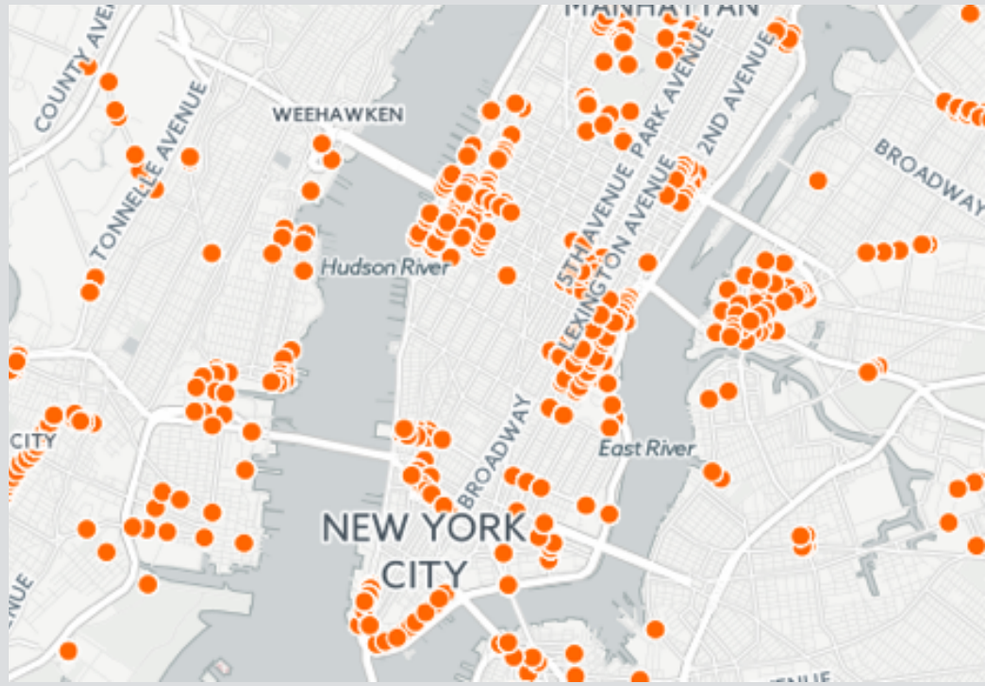
US Census -

Local estimate of population and age



Filling in the gaps:

Augmentation with other data



US Census -

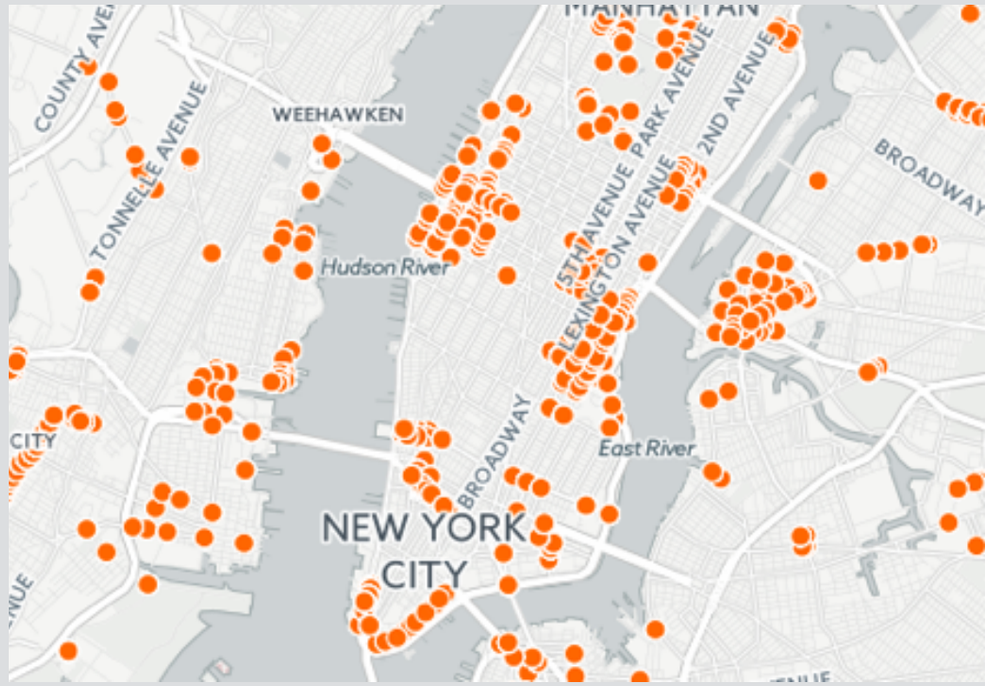
Local estimate of population and age

Google Geocode API -

Get street type (highway, road, lane, bridge, …), place into categorical bins.

Filling in the gaps:

Augmentation with other data



US Census -

Local estimate of population and age

Google Geocode API -

Get street type (highway, road, lane, bridge, …), place into categorical bins.

Factual.com API -

Places into two bins, count each for 50m, 150m, 300m

Bin #1: {Automotive, Community/Gov., Healthcare, Business/Services, Travel}

Bin #2: {Retail, Landmarks, Restaurants/Bars, Sports/Rec, Landmarks, Social, Transportation}

Finalizing a model

Try:

kNN, Random Forest, Gradient Boosting, Neural Nets, and BART

For each:

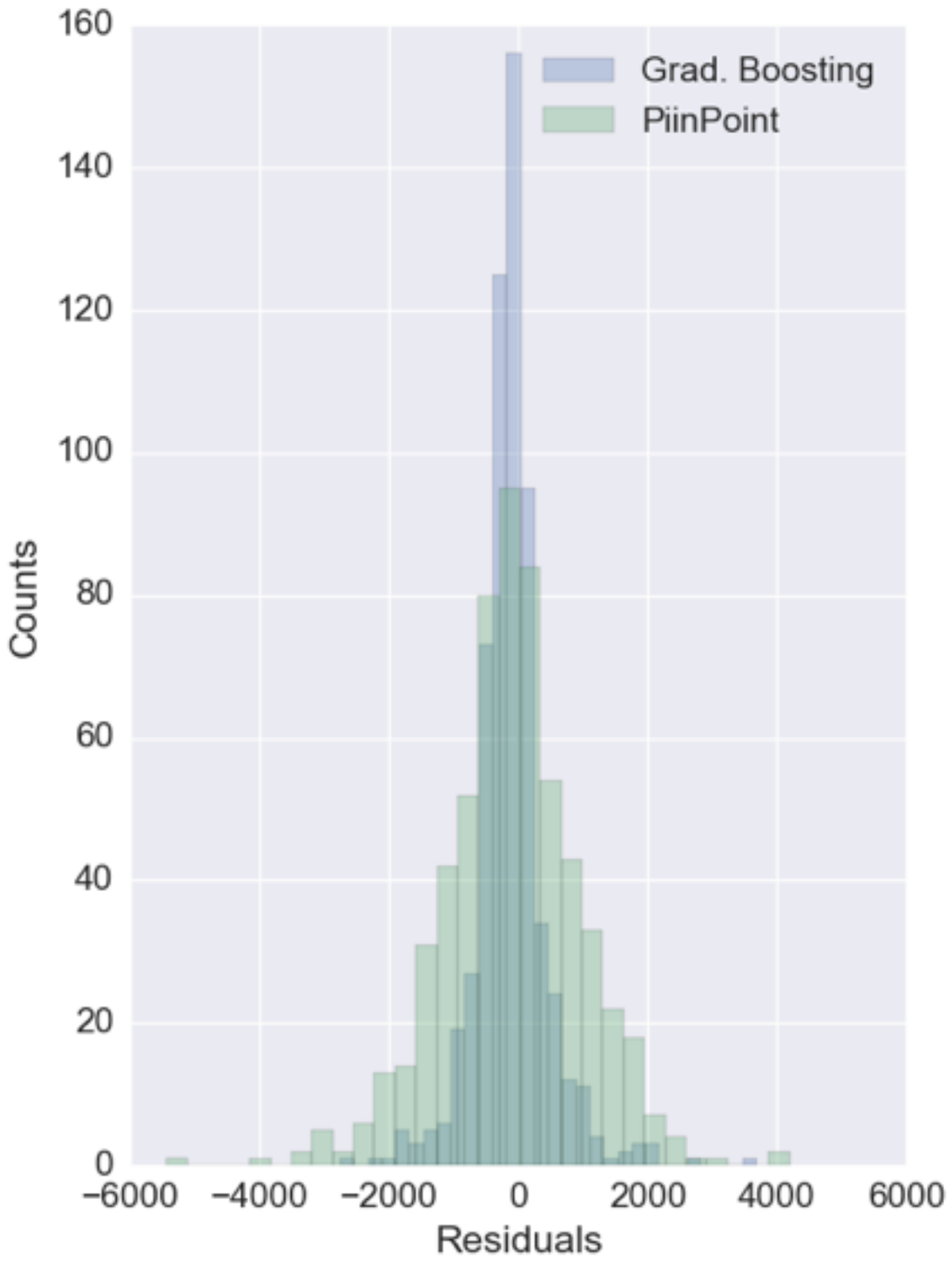
Select best features and parameters using 10-fold CV

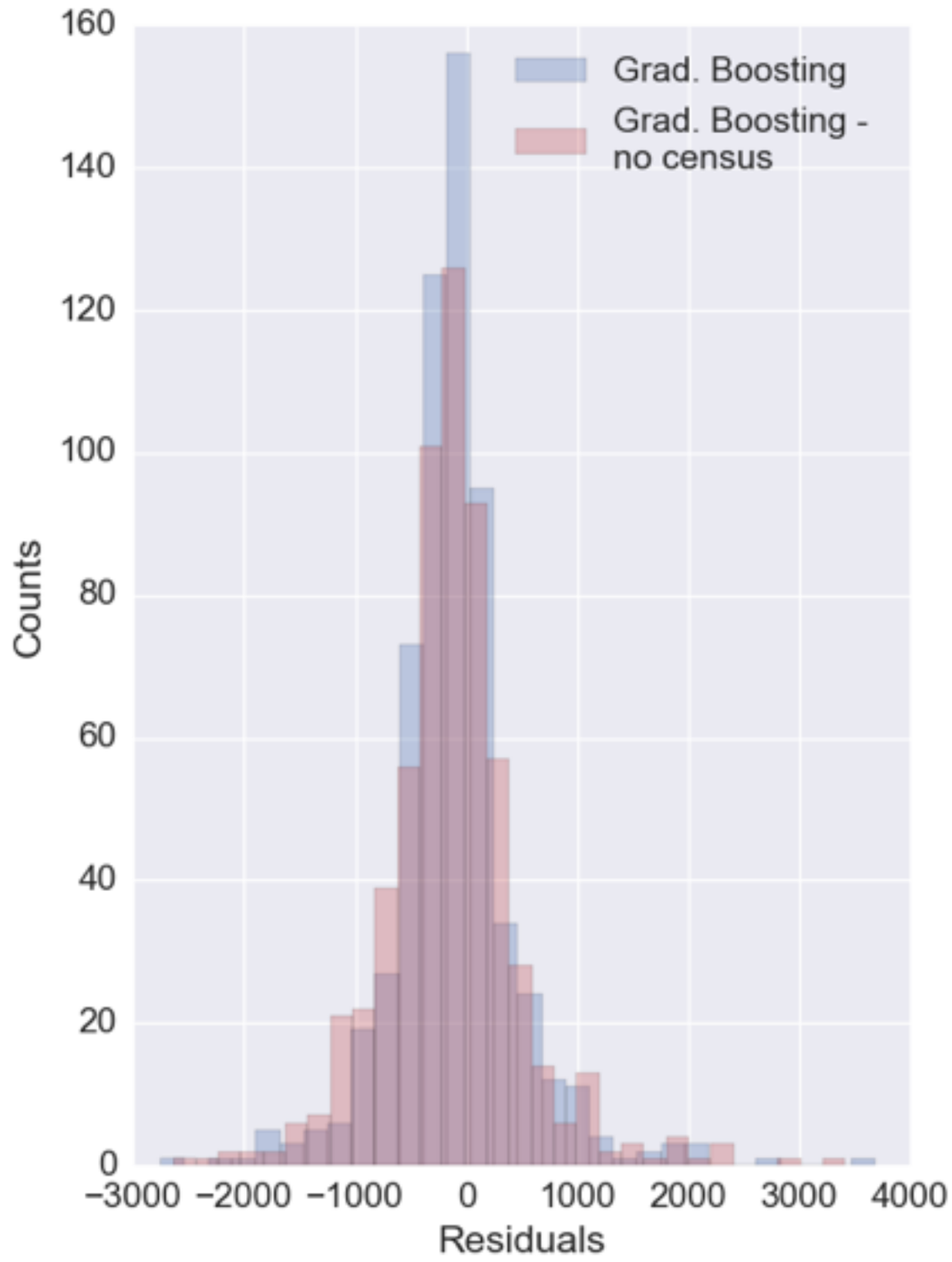
Result:

PiinPoint - Test RMSE: 1083

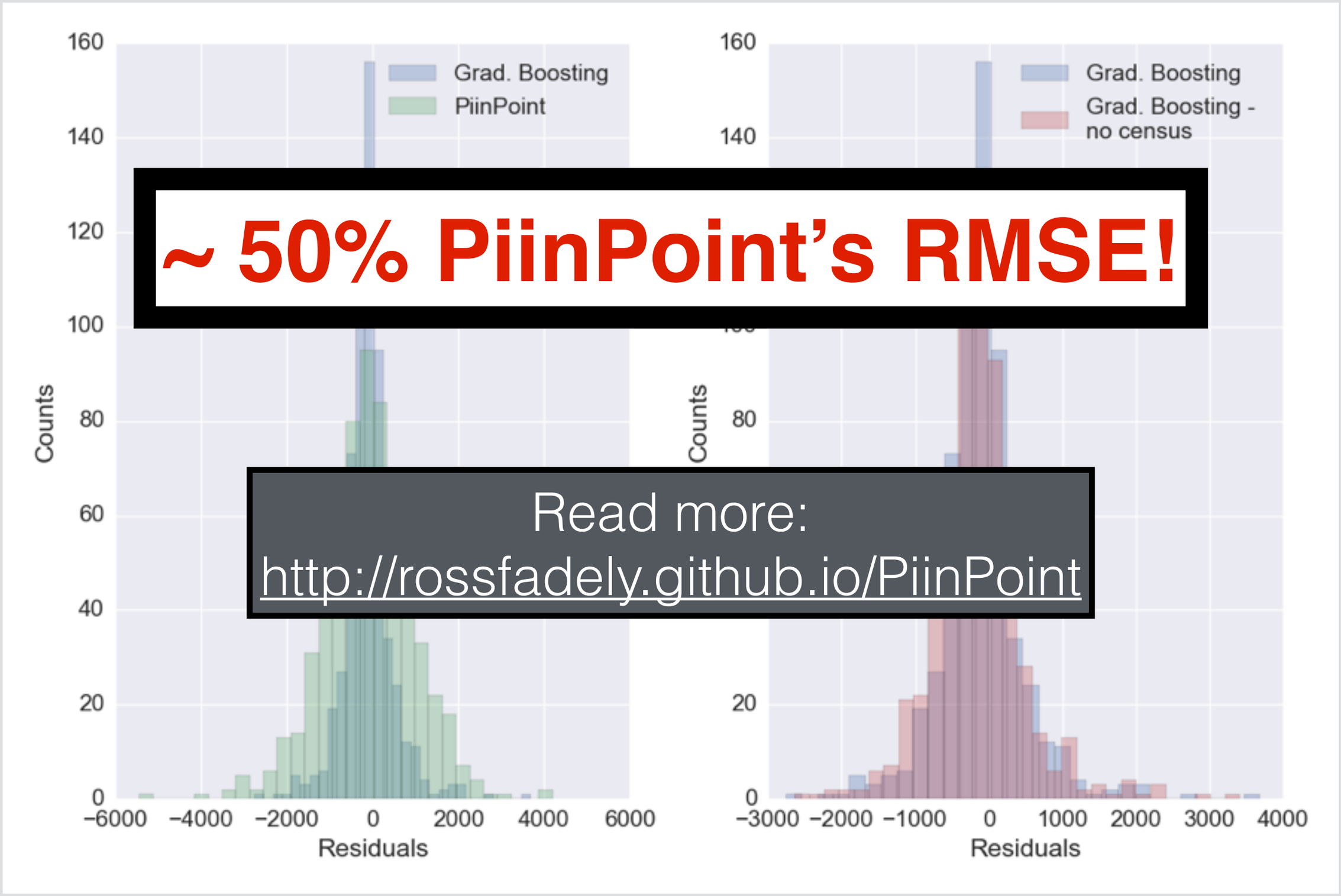
Gradient Boosting - Test RMSE: 537

Performance





Performance



~ 50% PiinPoint's RMSE!

Read more:
http://rossfadely.github.io/PiinPoint

About me

Here are places I've worked

RUTGERS UNIVERSITY

HAVERFORD

NYU

I built models to improve this

About me

Sometimes I climb these

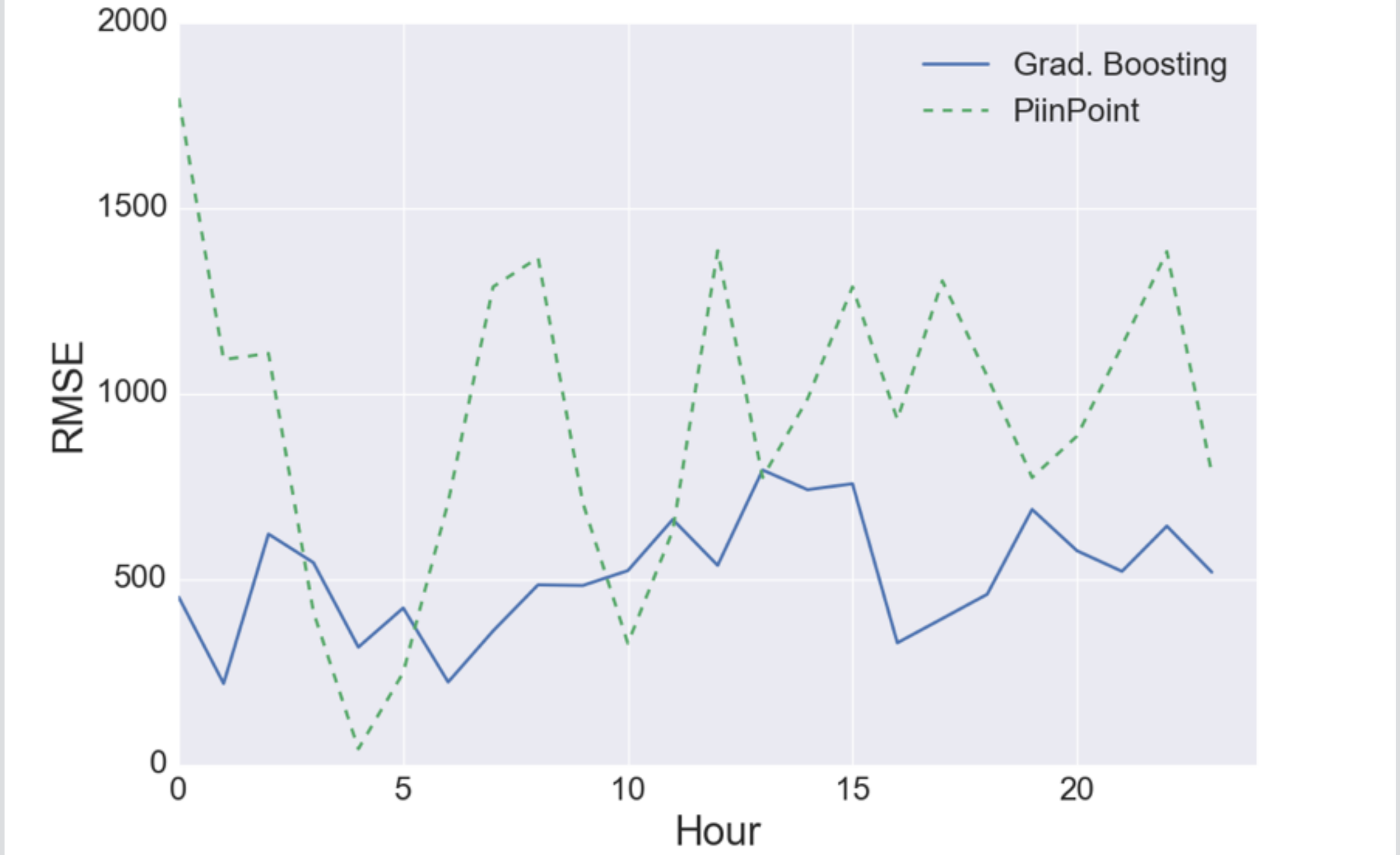
Mostly I climb here

Finalizing a model

Model	Test RMSE Peds	Cars
PiinPoint	**622**	**1083**
kNN only	749	1029
Random Forest (kNN)	471	683
Random Forest (kNN+ C)	416	633
Gradient Boosting (kNN + C)	**334**	560
Gradient Boosting (kNN + CS)	357	572
Gradient Boosting (kNN + CSF)	368	**537**
Neural Nets, B.A.R.T.	485-629	621-1123

C	S	F	L
U.S. Census (pop, med. age)	Street types	factual.com counts	Lat, Lon

Performance



Performance