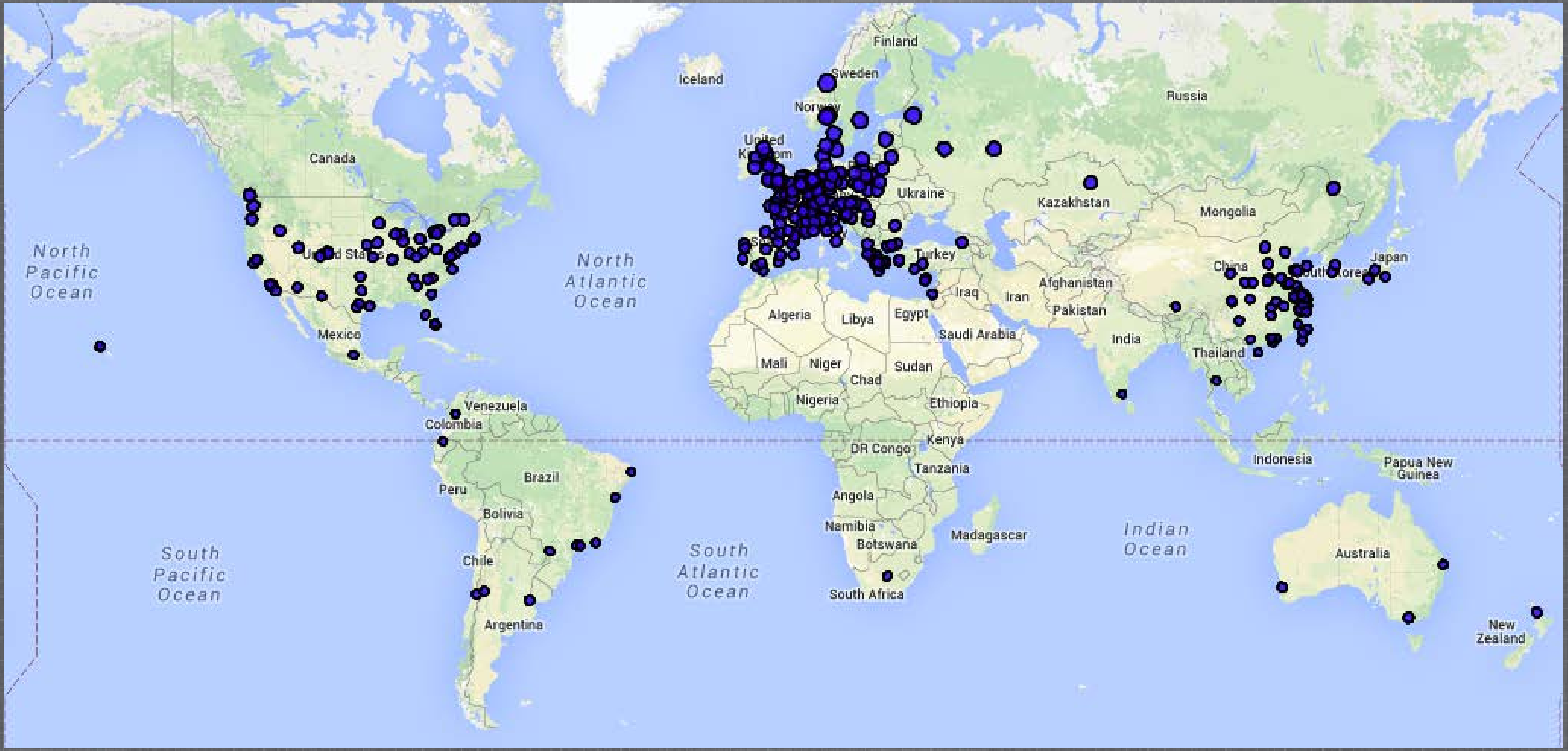
Smarter Tools for (Citi)Bike Sharing:
Analytics for Managing a Complex System

David B. Shmoys
Cornell University

Joint work with Daniel Freund, Shane Henderson, Eoin O'Mahony

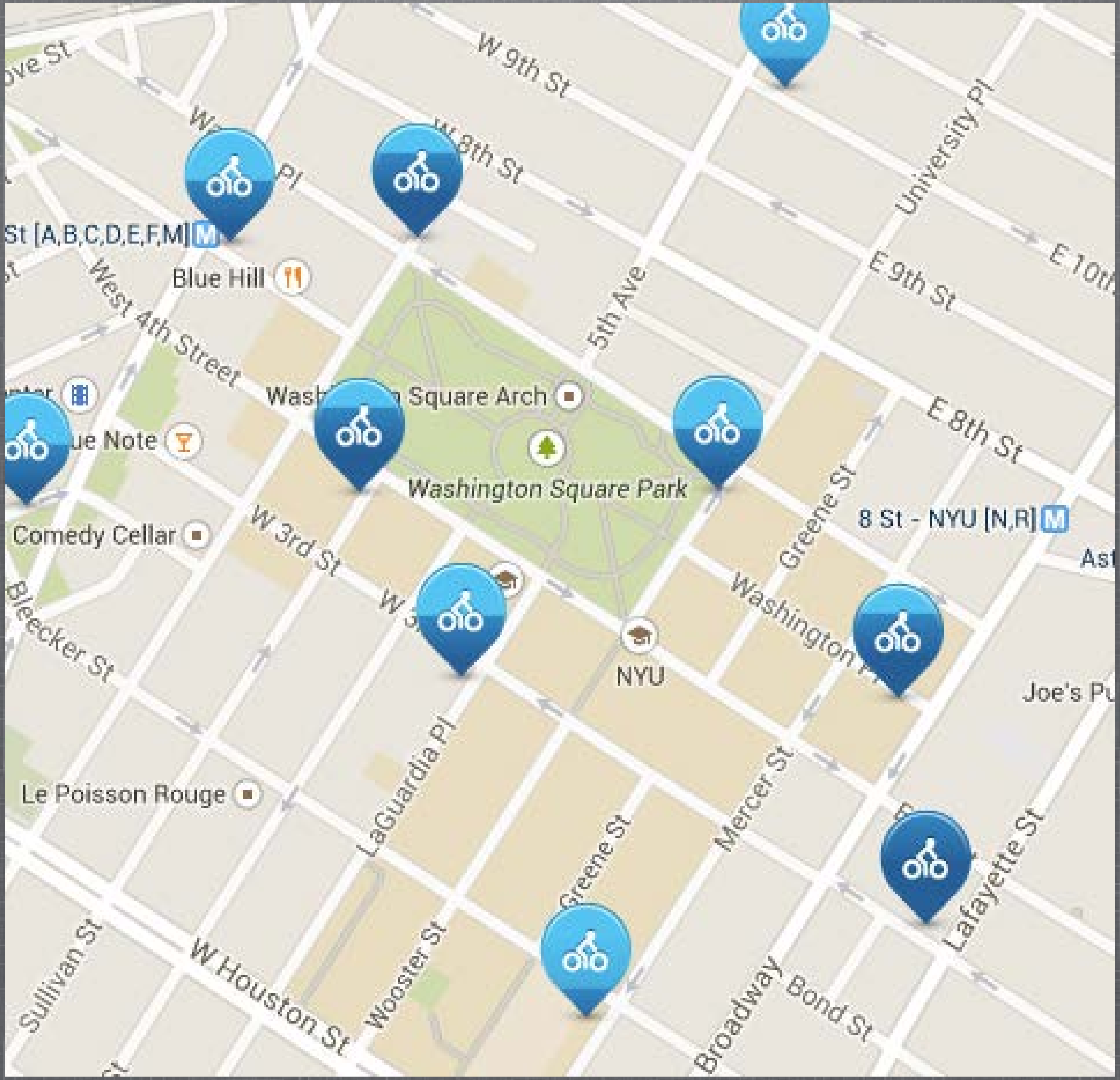
BIKE SHARING SYSTEMS



BIKE SHARING SYSTEMS

BIKE SHARING SYSTEMS

BIKE SHARING SYSTEMS



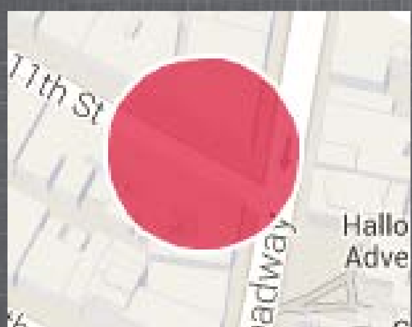
citi bike

- Initiated May 2013

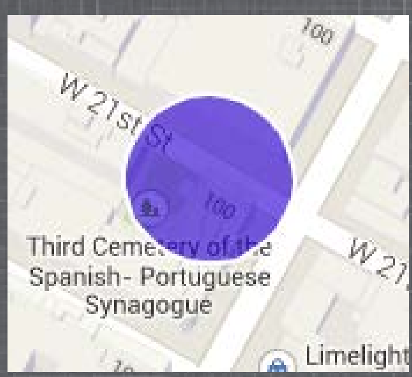
- Now
  - More than 1 million rides per month (Vélib' 3M)
  - More than 10 million rides in 2015
  - 6000+ bikes
  - 380+ stations
  - Ongoing expansion

- Citibike has worked with Cornell throughout
  - Initially routing trucks for station battery replacement
  - Current focus is rebalancing

SYSTEM IMBALANCE



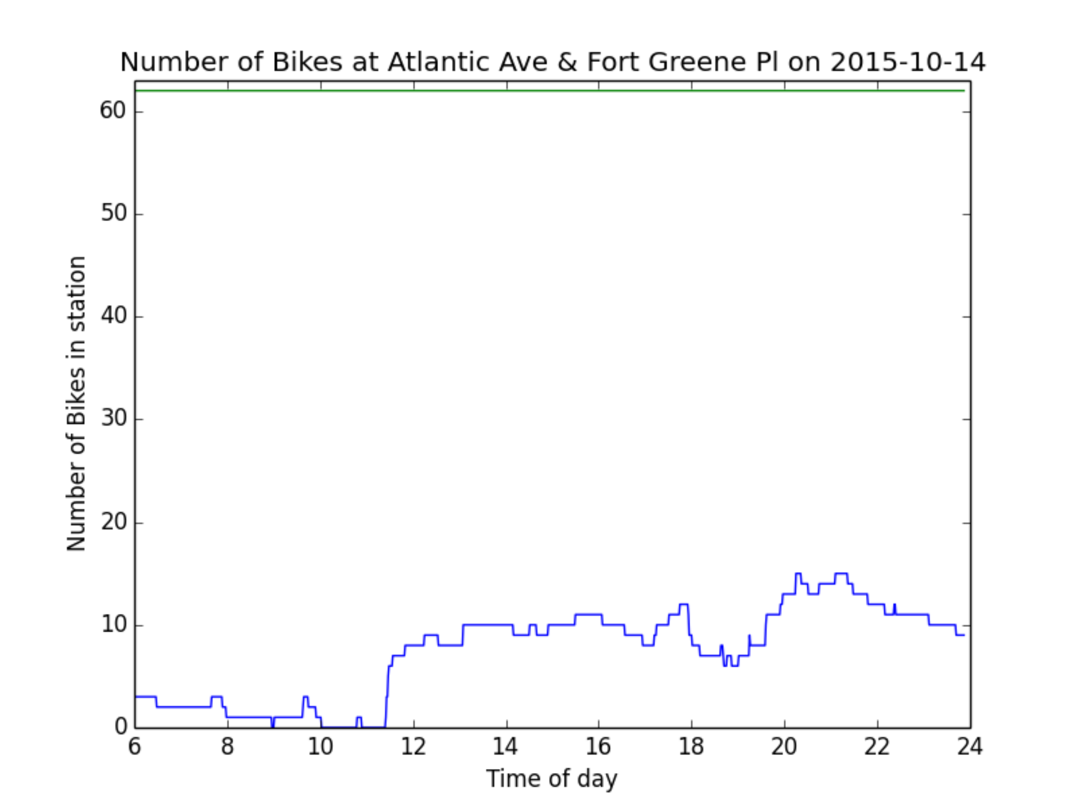
Departing

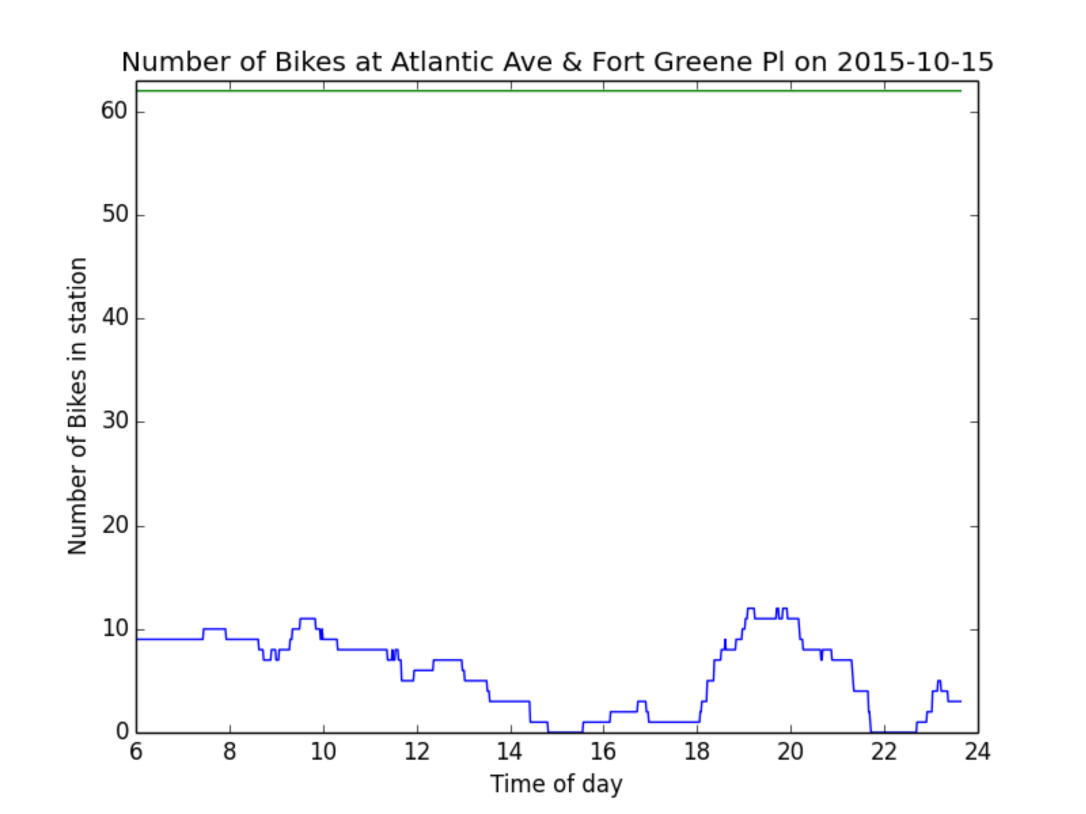


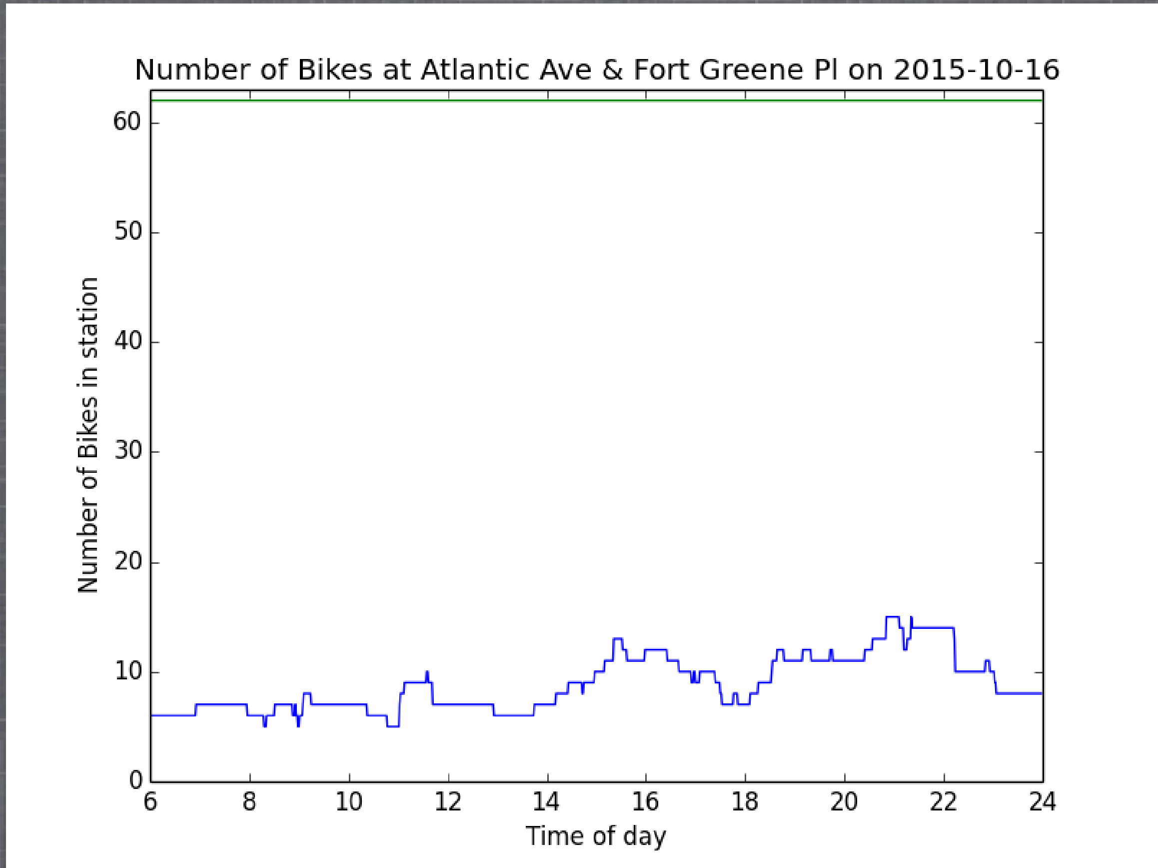
Arriving

Diameter~
Bikes

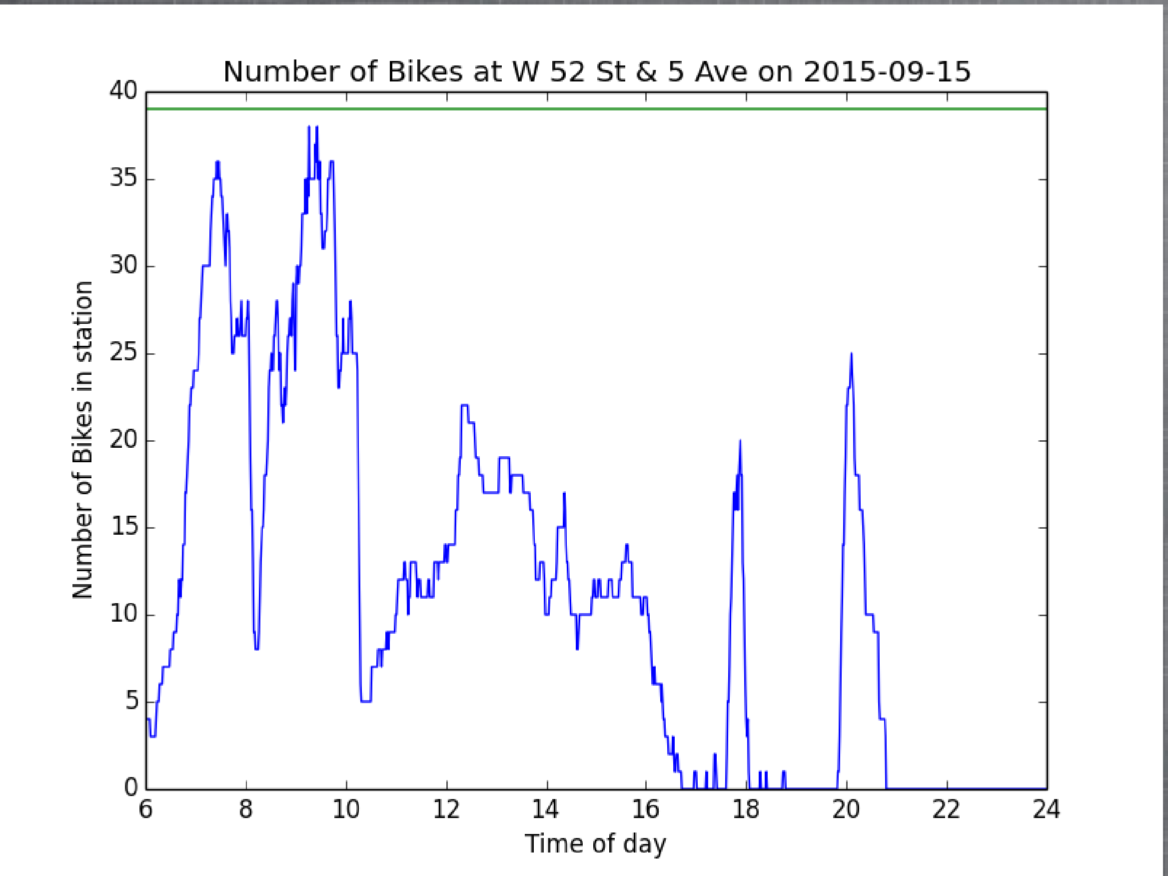
KNOWING SYSTEM STATE

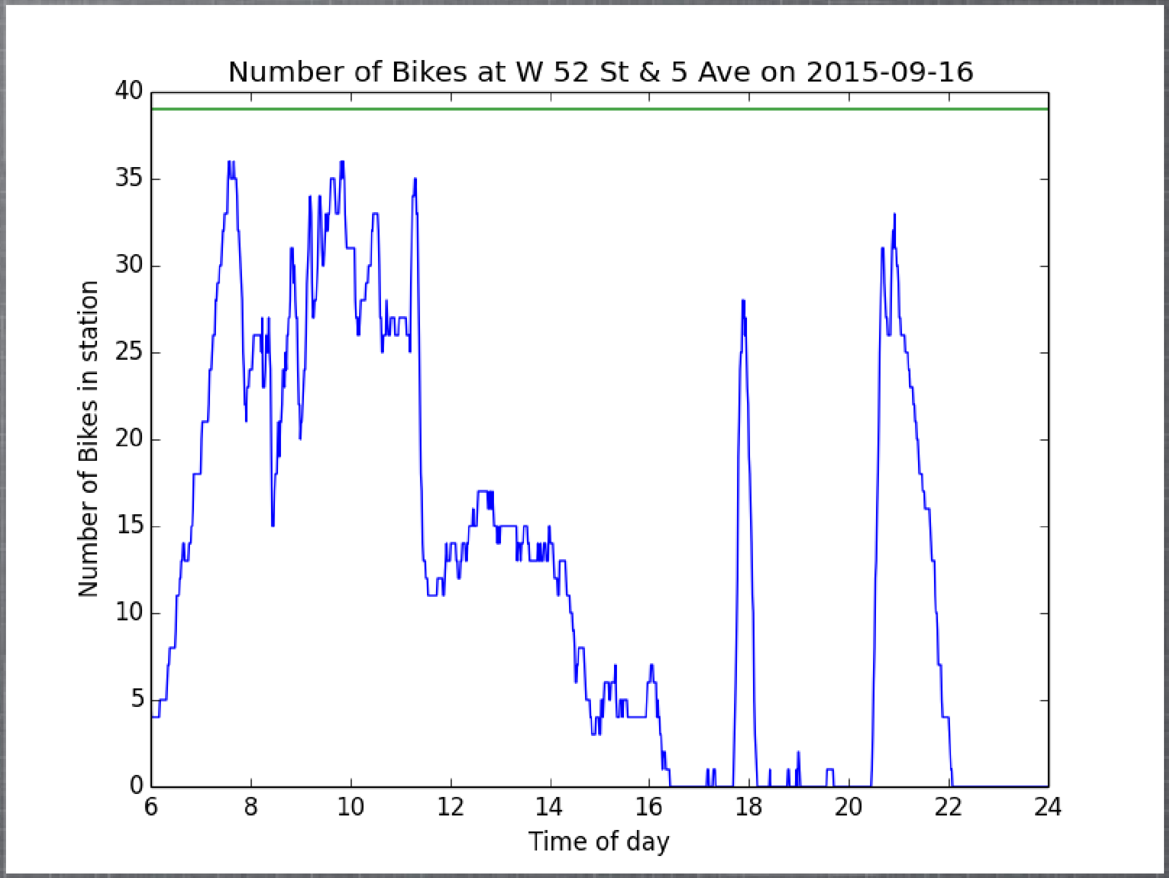


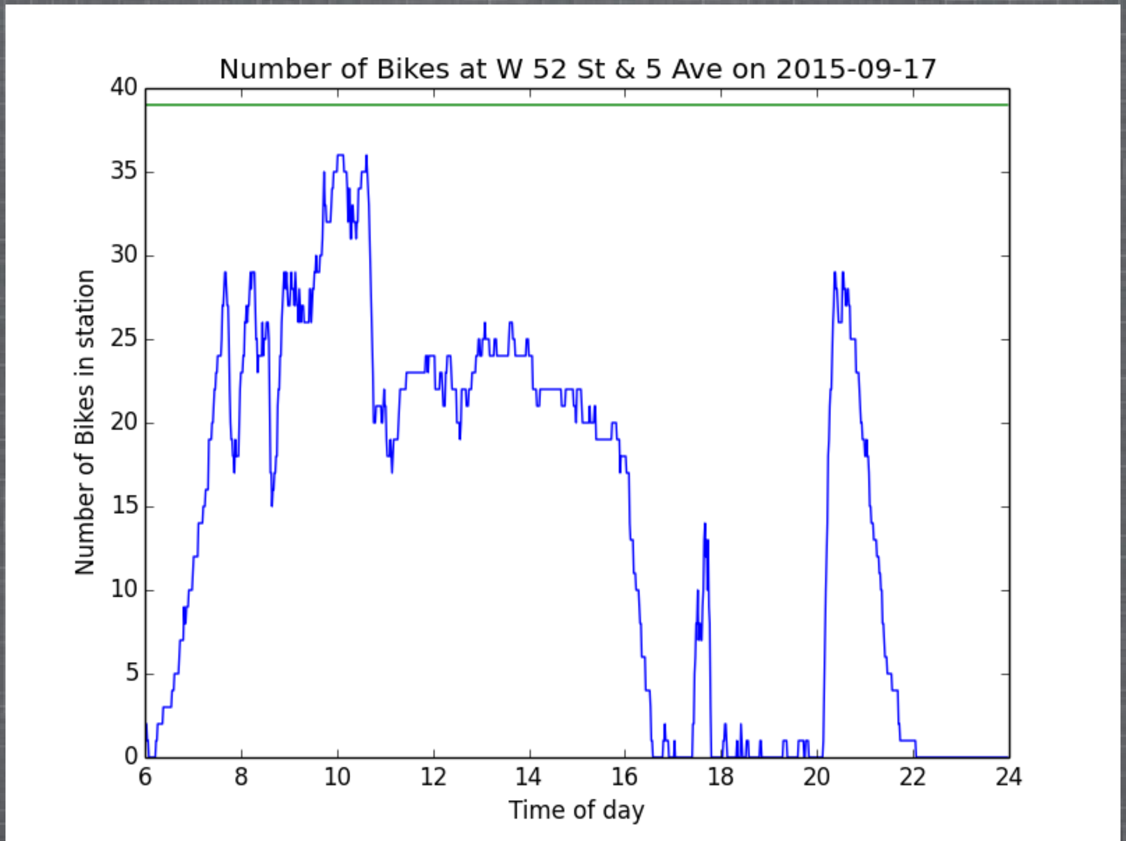




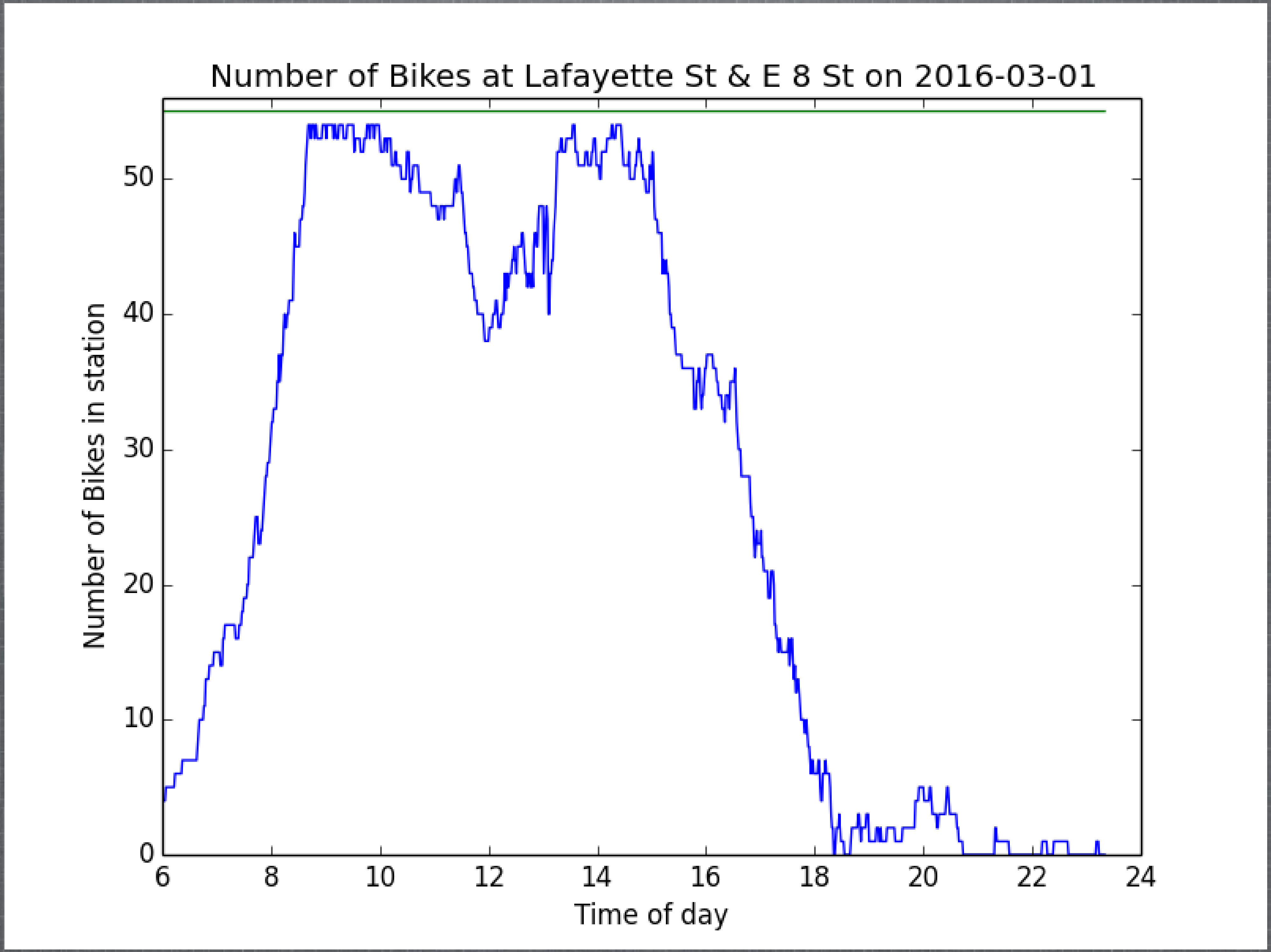
KNOWING SYSTEM STATE







KNOWING SYSTEM STATE?



CENSORED DATA

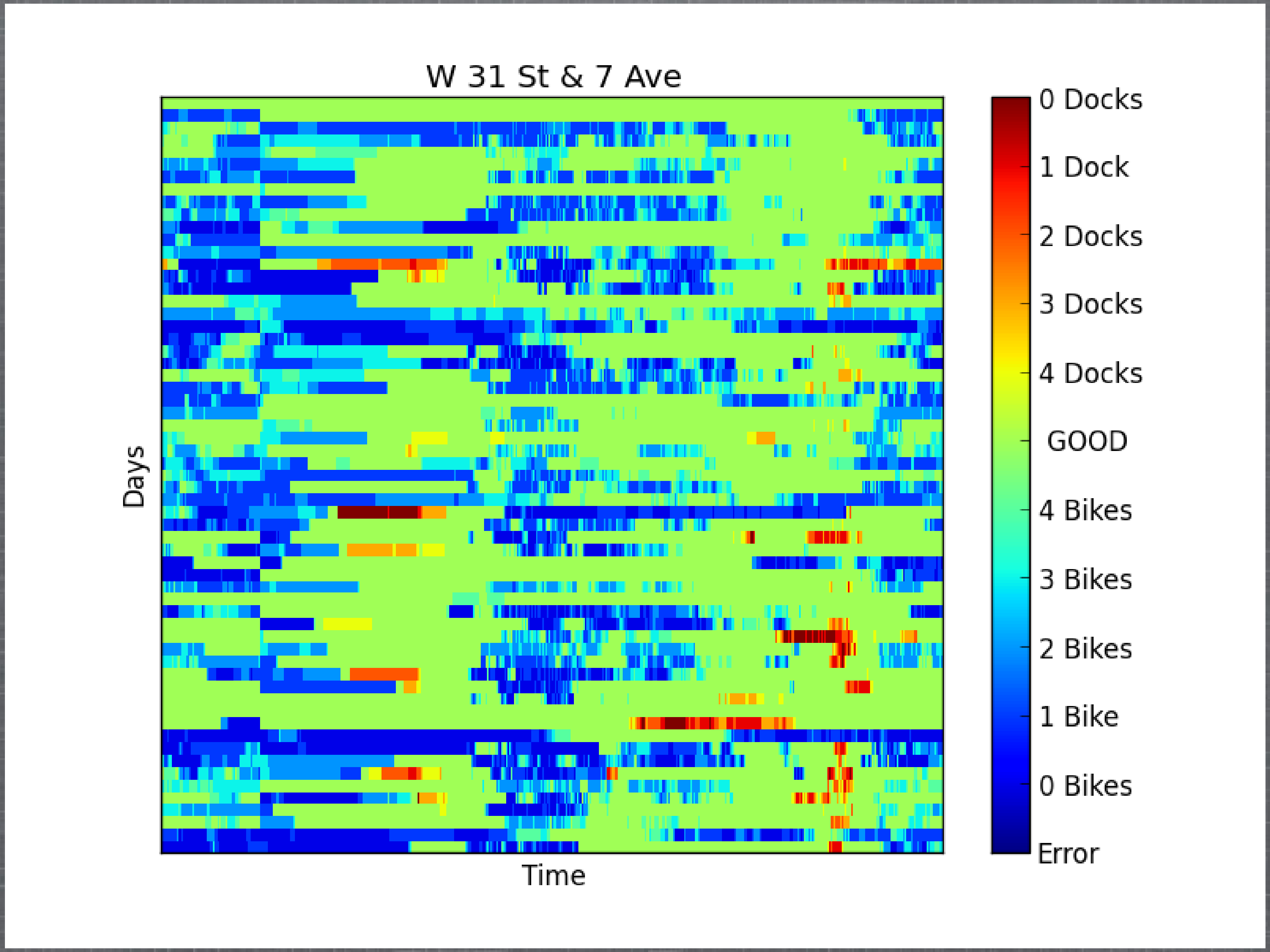
We don’t observe failed trips

Can we infer the actual demand?

Rebalancing can help us

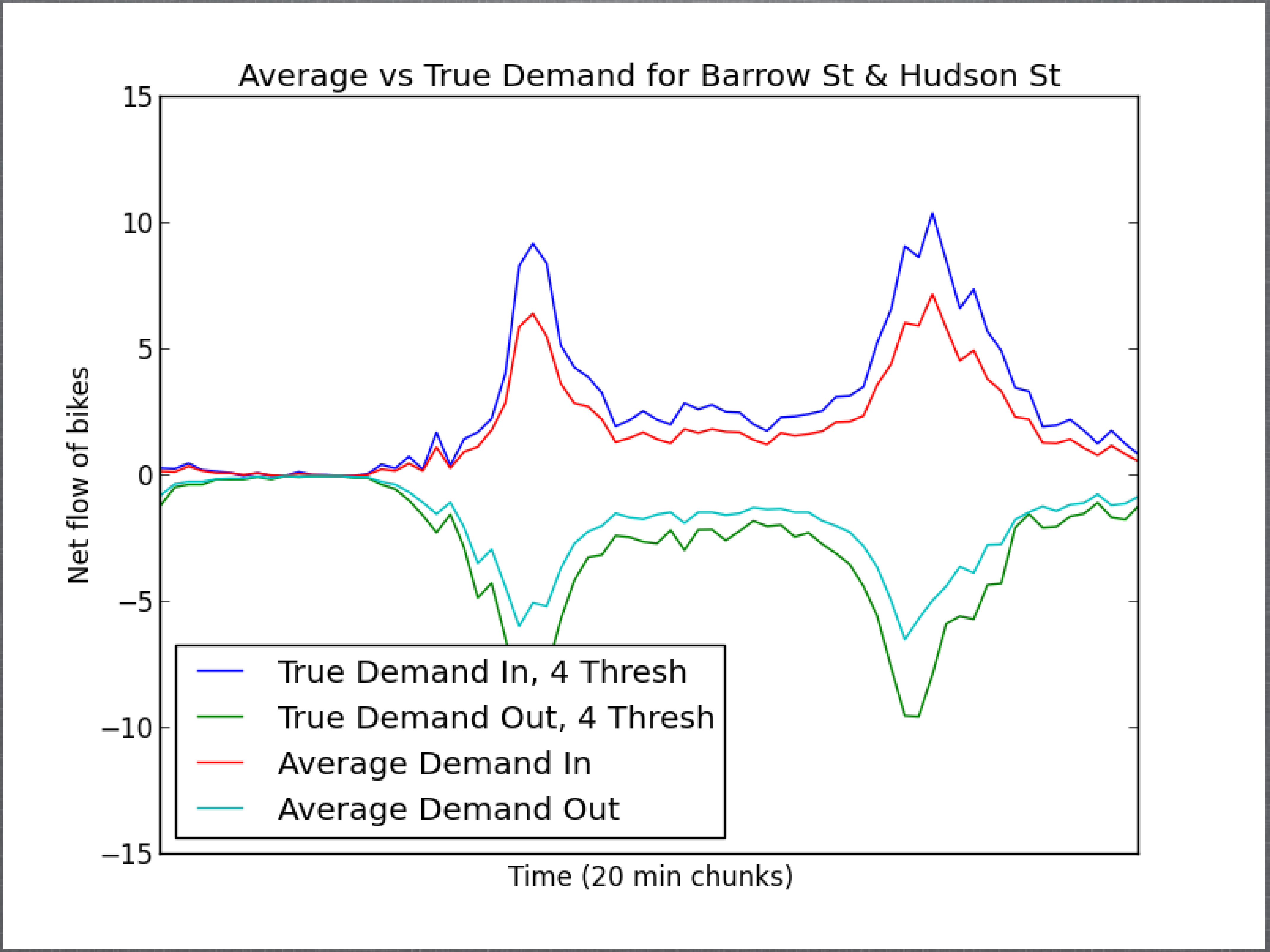
CENSORED DATA

Flow of bikes?

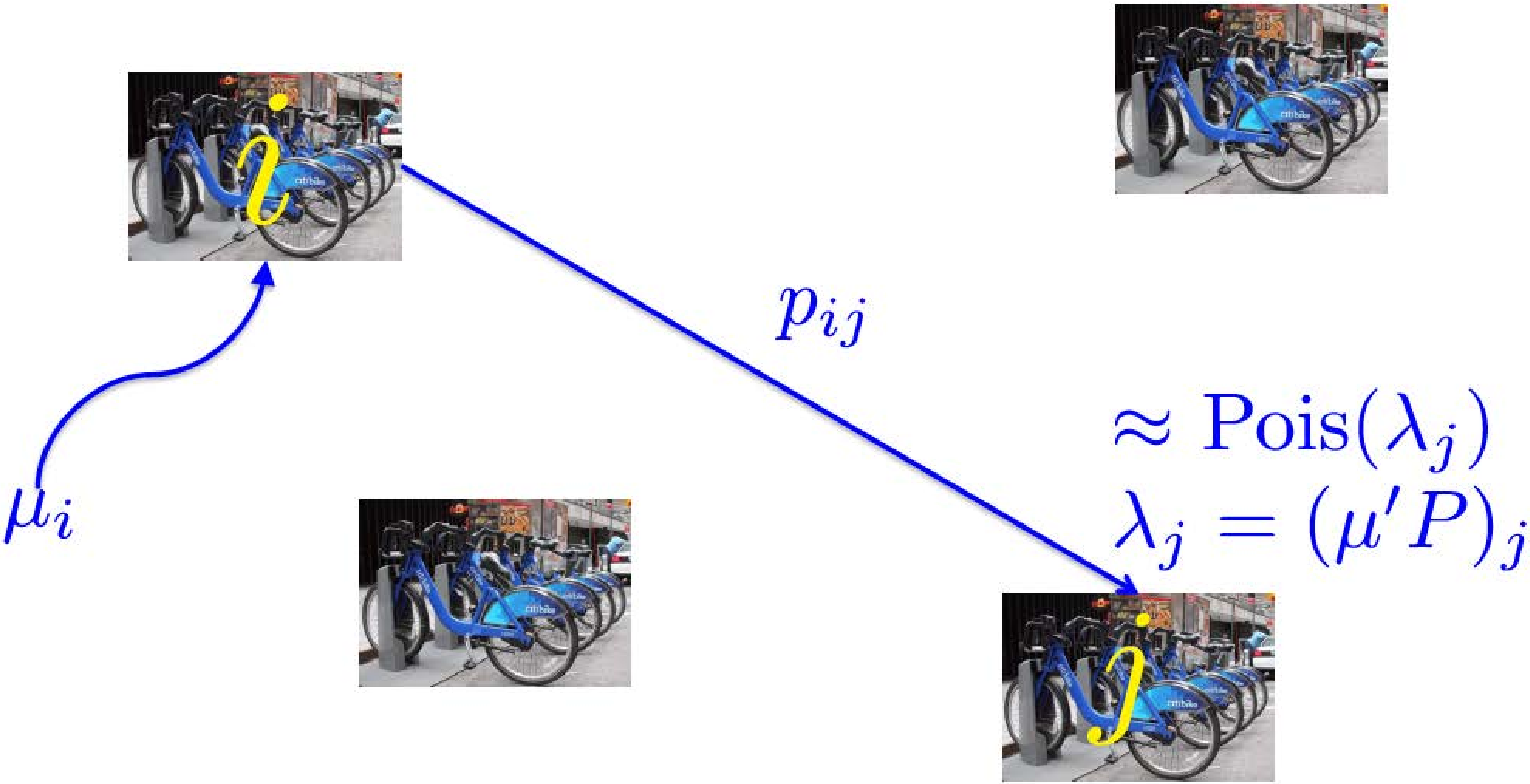


12:00am 12:00pm 12:00am

CENSORED DATA



A Model



Using fact that departure process from M/G/∞ is Poisson

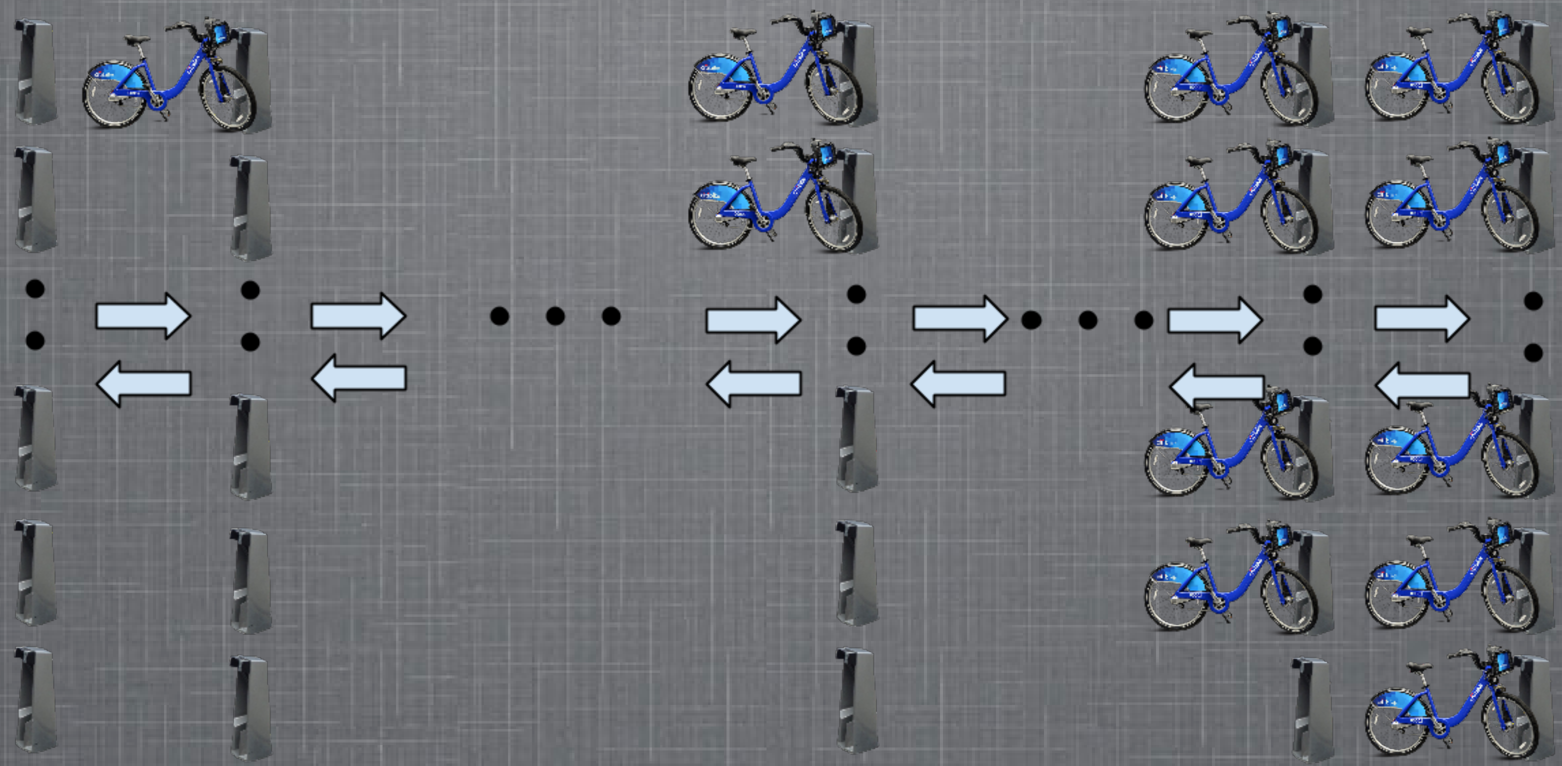
WHERE AND WHEN

Poisson flow of bikes (compute from data)

Model stations with Continuous Time Markov Chain (CTMC)

Compute E[# Upset People] over rush-hour

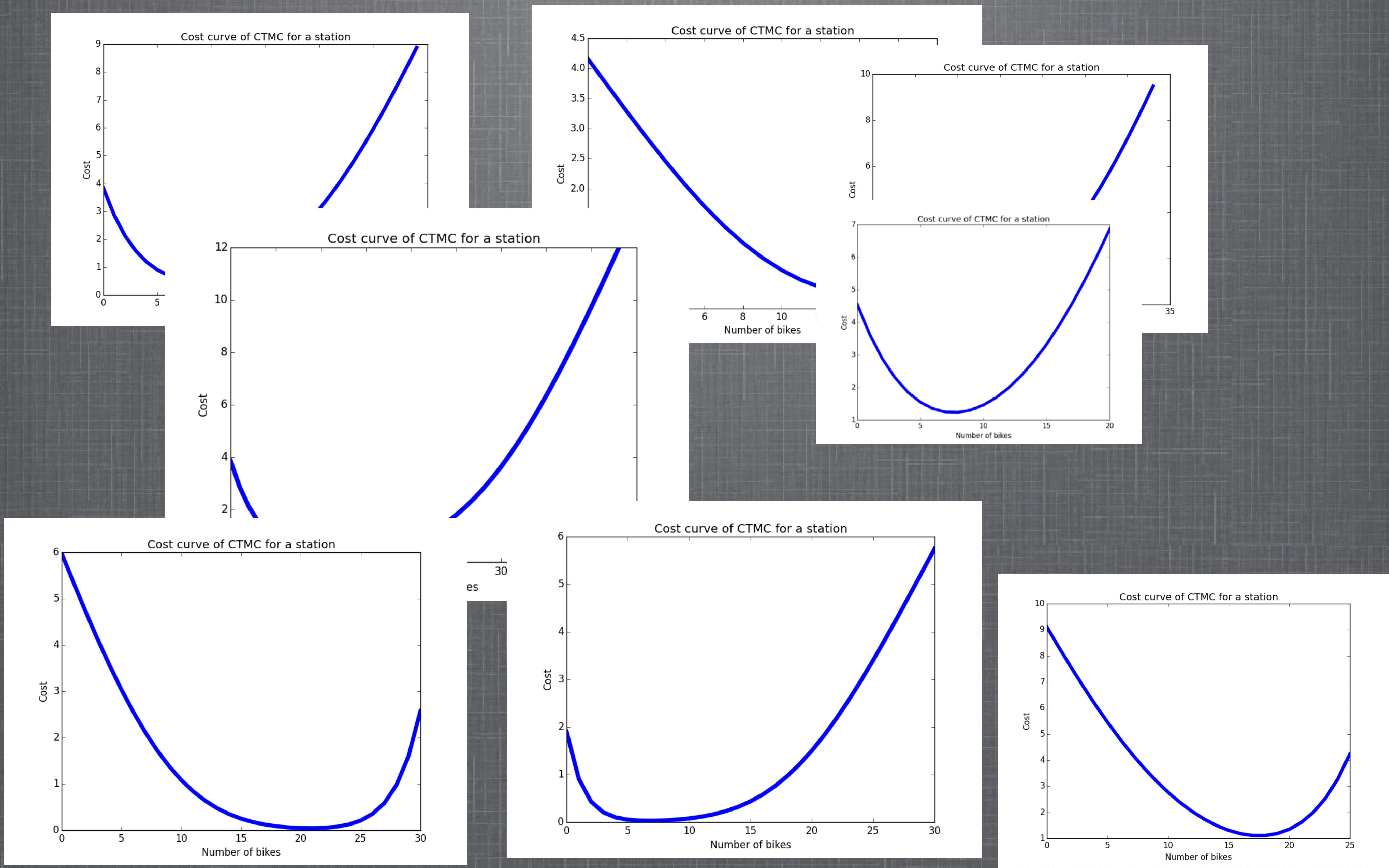
CTMC MODELING



Given Rates: λ take μ returns

$$f(i) = E\left[\lambda \int_0^t I(x(s) = 0 | x(0) = i) ds + \mu \int_0^t I(x(s) = C | x(0) = i) ds\right]$$

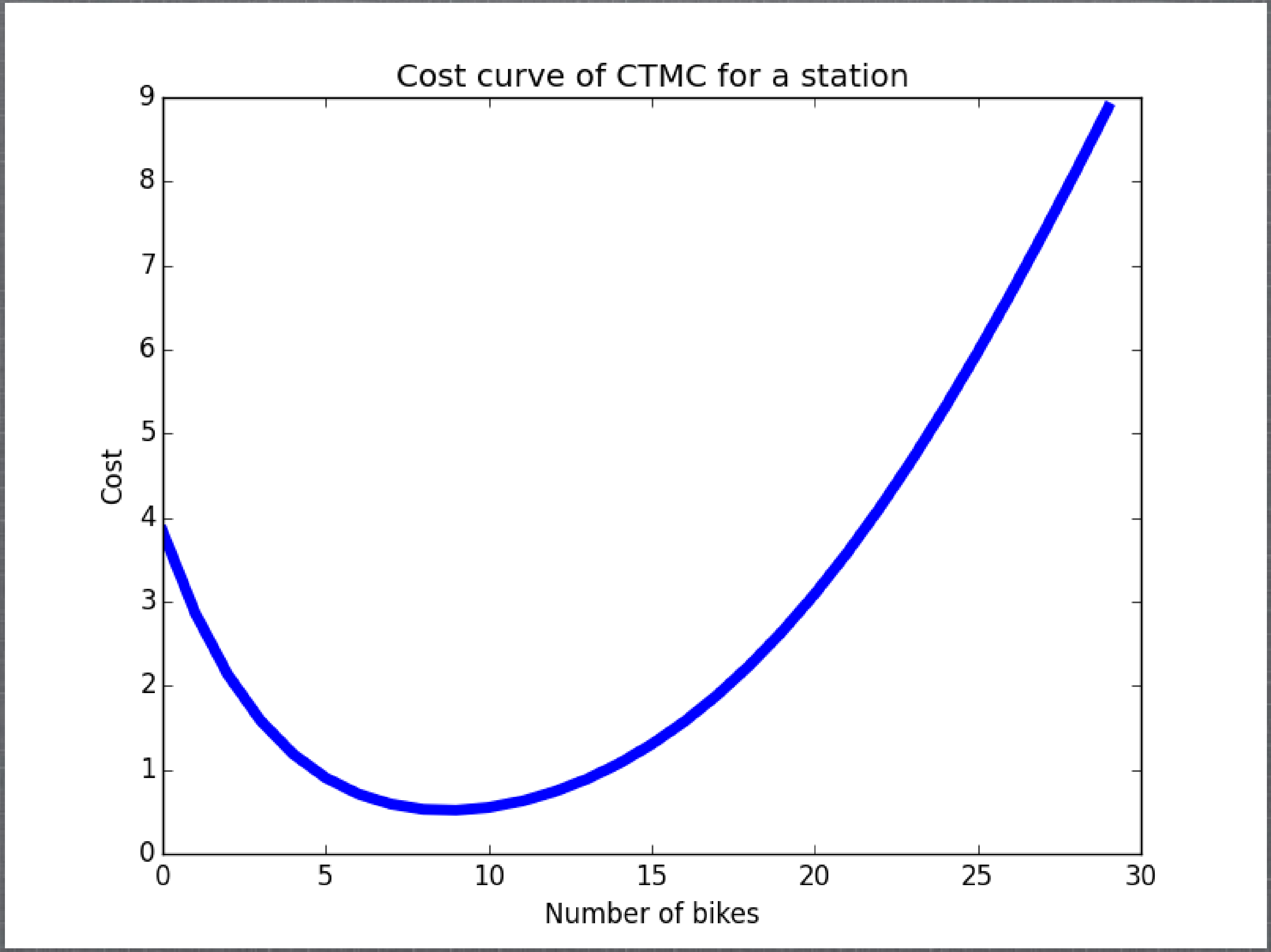
CTMC MODELING



WHERE AND WHEN

Place bikes to minimize E[#up]

Thm: Cost function convex



CONVEX PROGRAM

$$\min \sum_{s} Cost(X_s)$$

$$\sum_{s} X_s \leq B$$

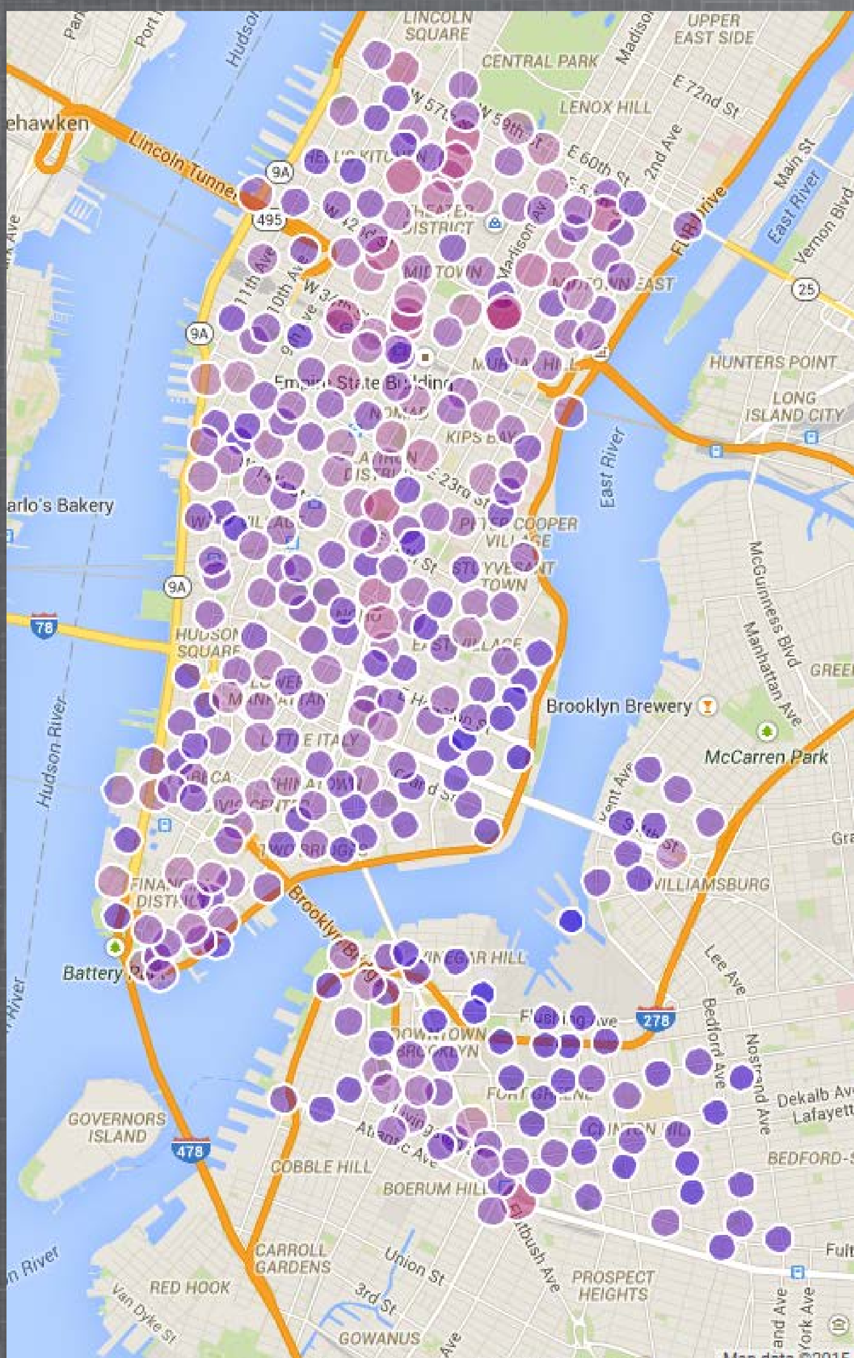
$$\forall s, 0 \leq X_s \leq C_s$$

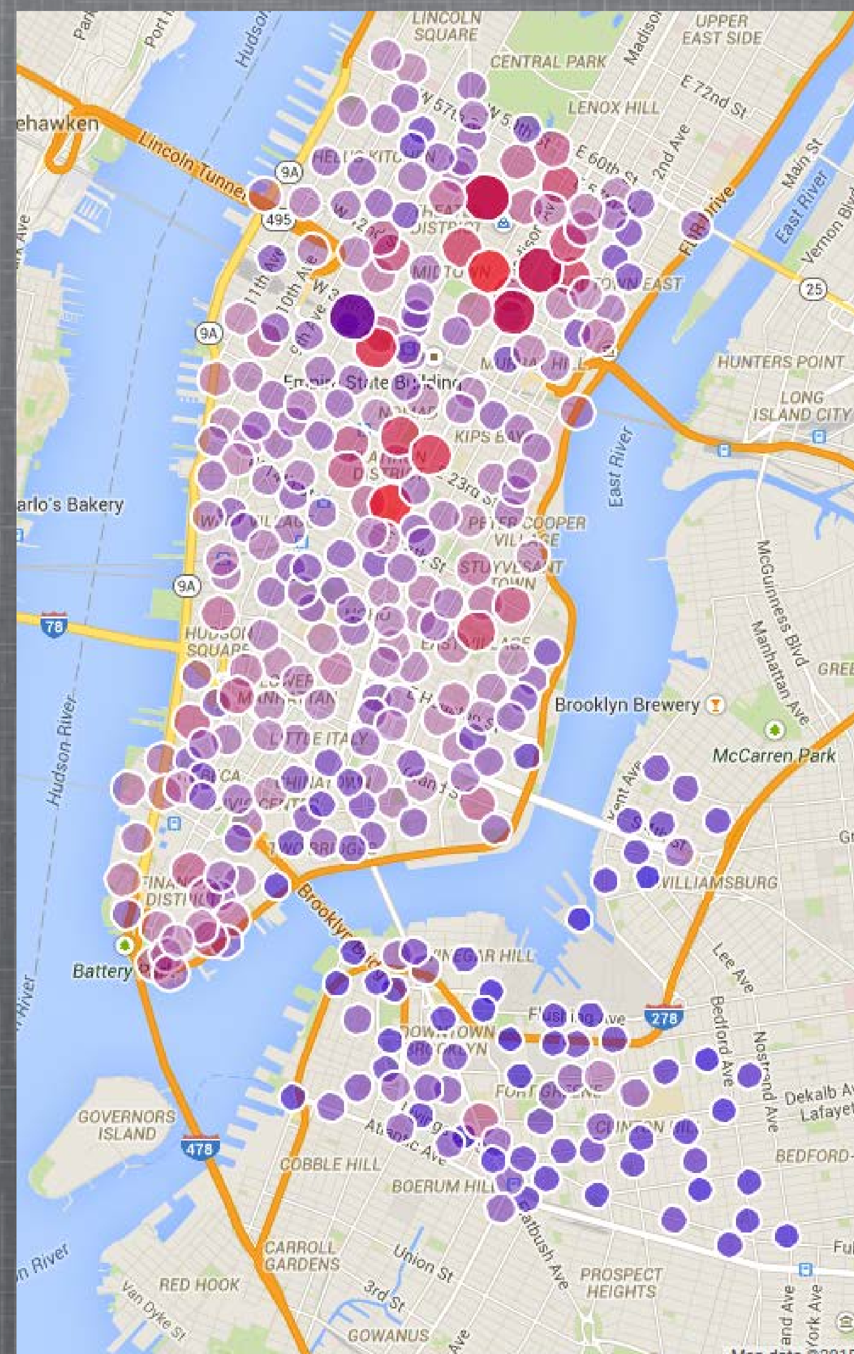
$$X_s = \text{Number of bikes at s}$$

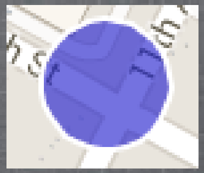
$$C_s = \text{Capacity of s}$$

HOW WE USE THIS

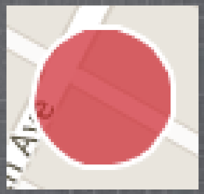
Where Should Have Racks



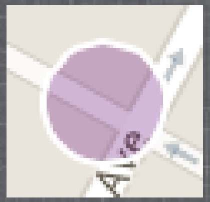




Fewer

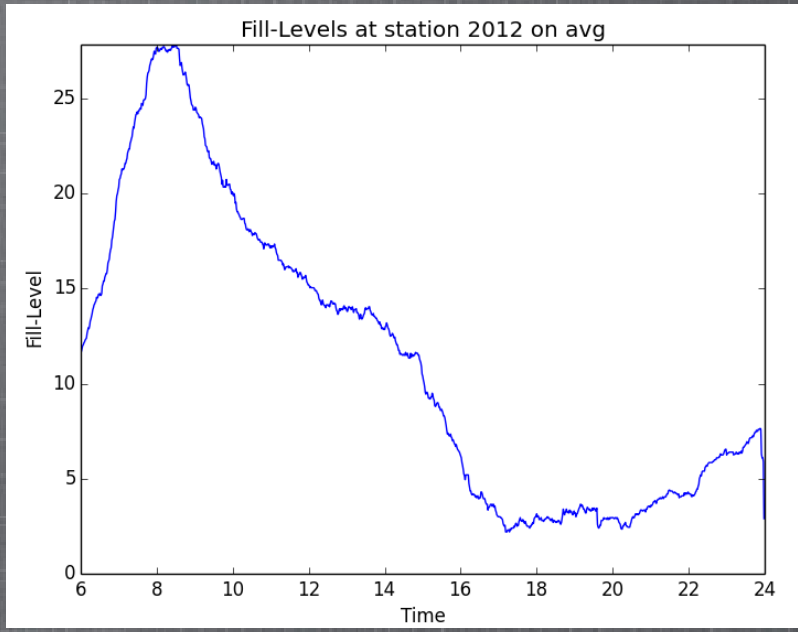


More



Same

What if λ and μ are time-dependent?



Assume that in small intervals (~30 minutes) λ , μ are constant

Find expected number of upset people for each interval, given each initial number of bikes

Find for each interval and each initial number of bikes the distribution over number of bikes at the end of the interval

Solve stochastic dynamic program

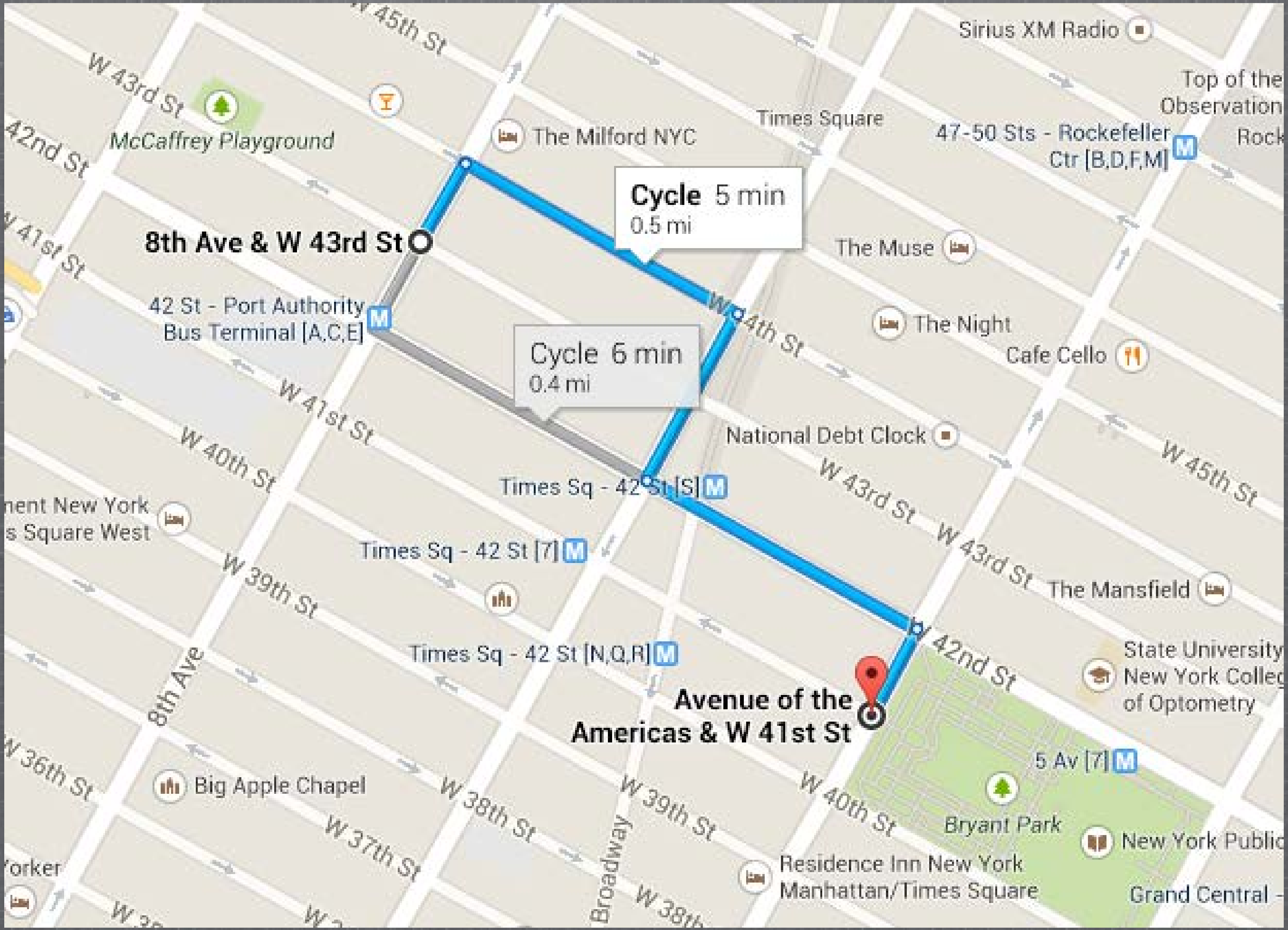
Average number of bikes over the course of a day next to Bellevue Hospital

REBALANCING WITH OPTIMIZATION

MID RUSH HOUR

REBALANCING WITH OPTIMIZATION

Mid Rush Rebalancing



MIDRUSH REBALANCING

Can’t keep system balanced

Use is high but predictable

How do we focus resources?

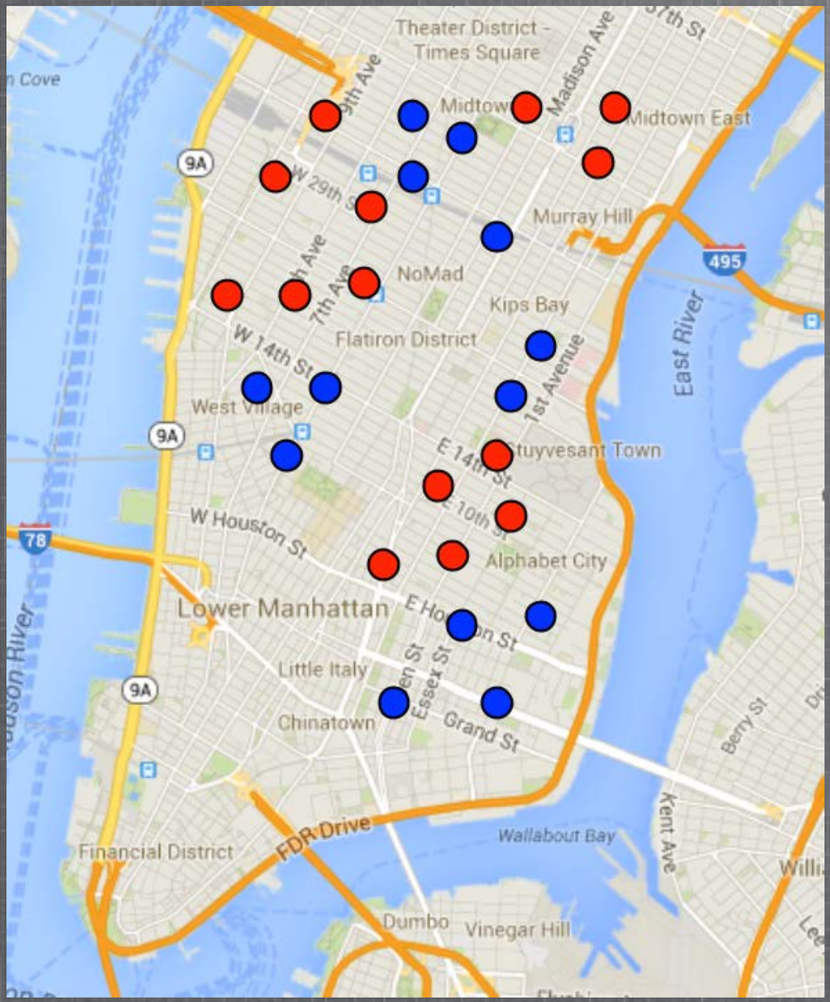
MIDRUSH REBALANCING

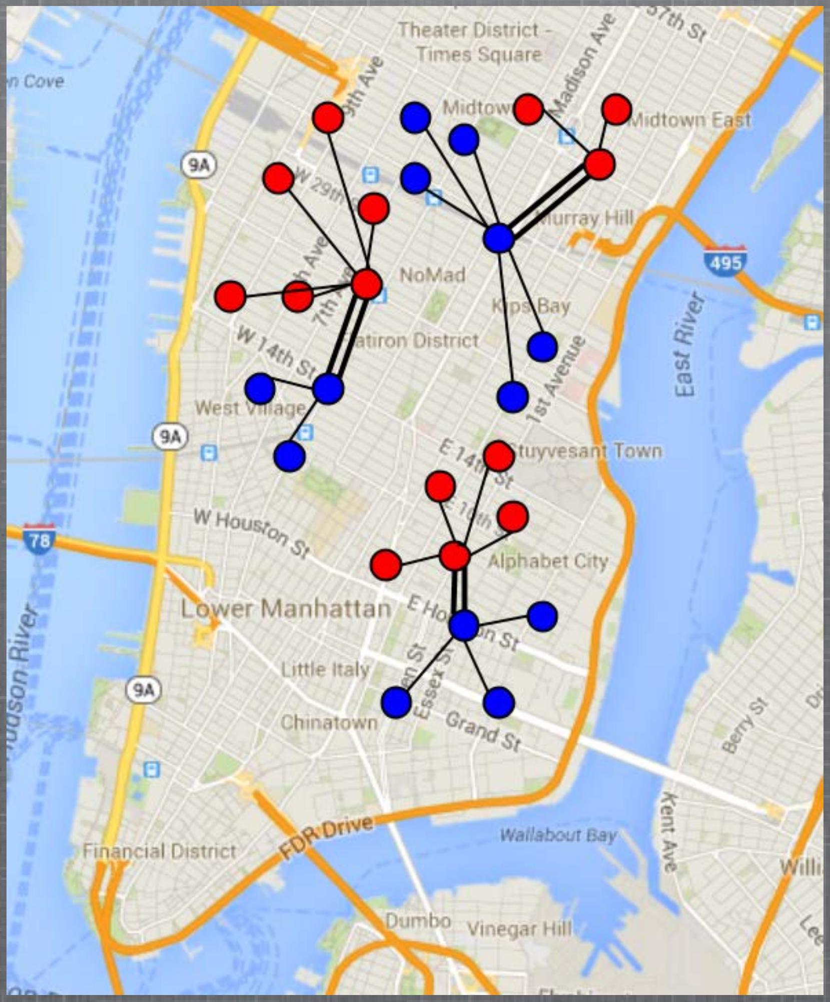
Balance areas of the city

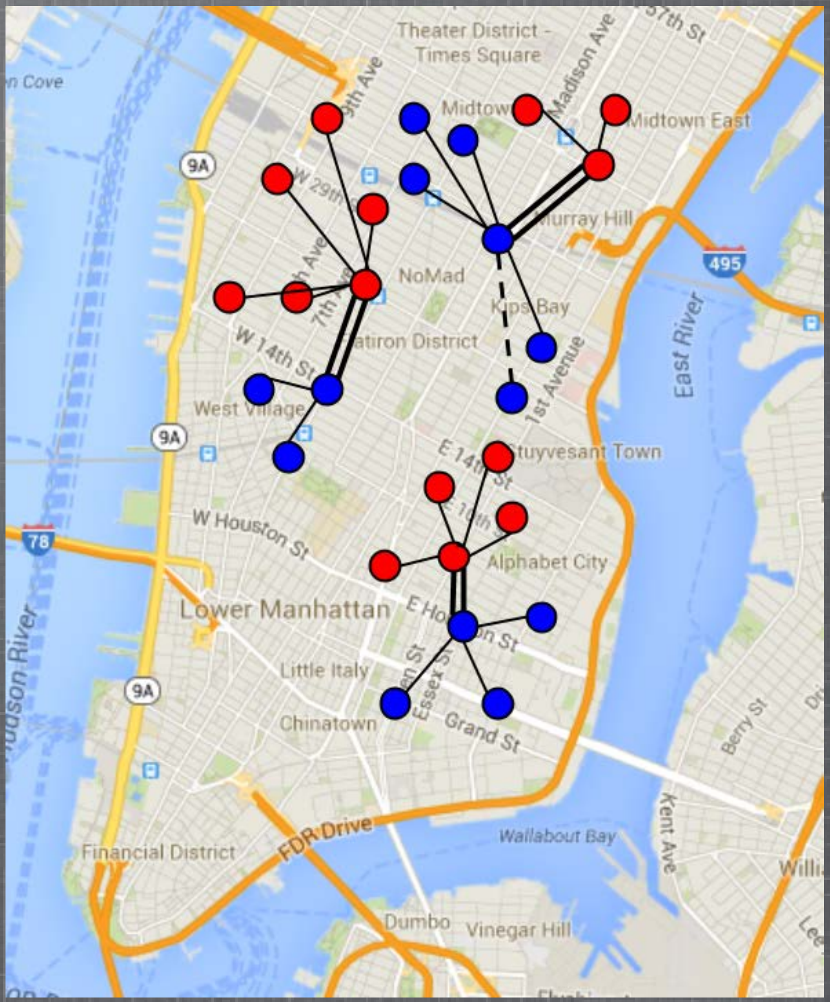
Subset of stations, k trailers

Everyone close to bike/spot

MIDRUSH REBALANCING





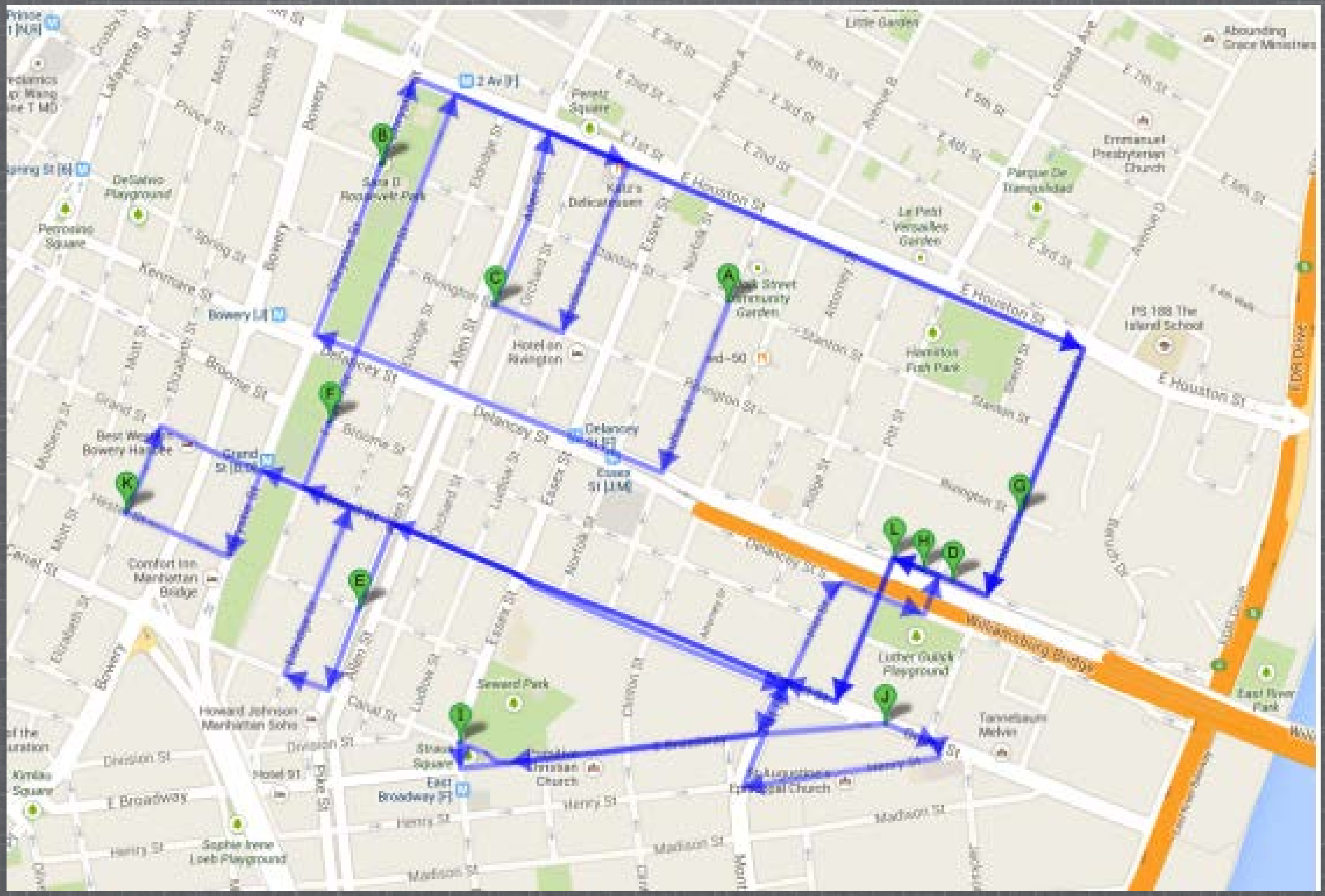


How to select 3 pairs of stations
To be rebalanced by trailers?

OVERNIGHT

REBALANCING WITH OPTIMIZATION

Overnight Rebalancing



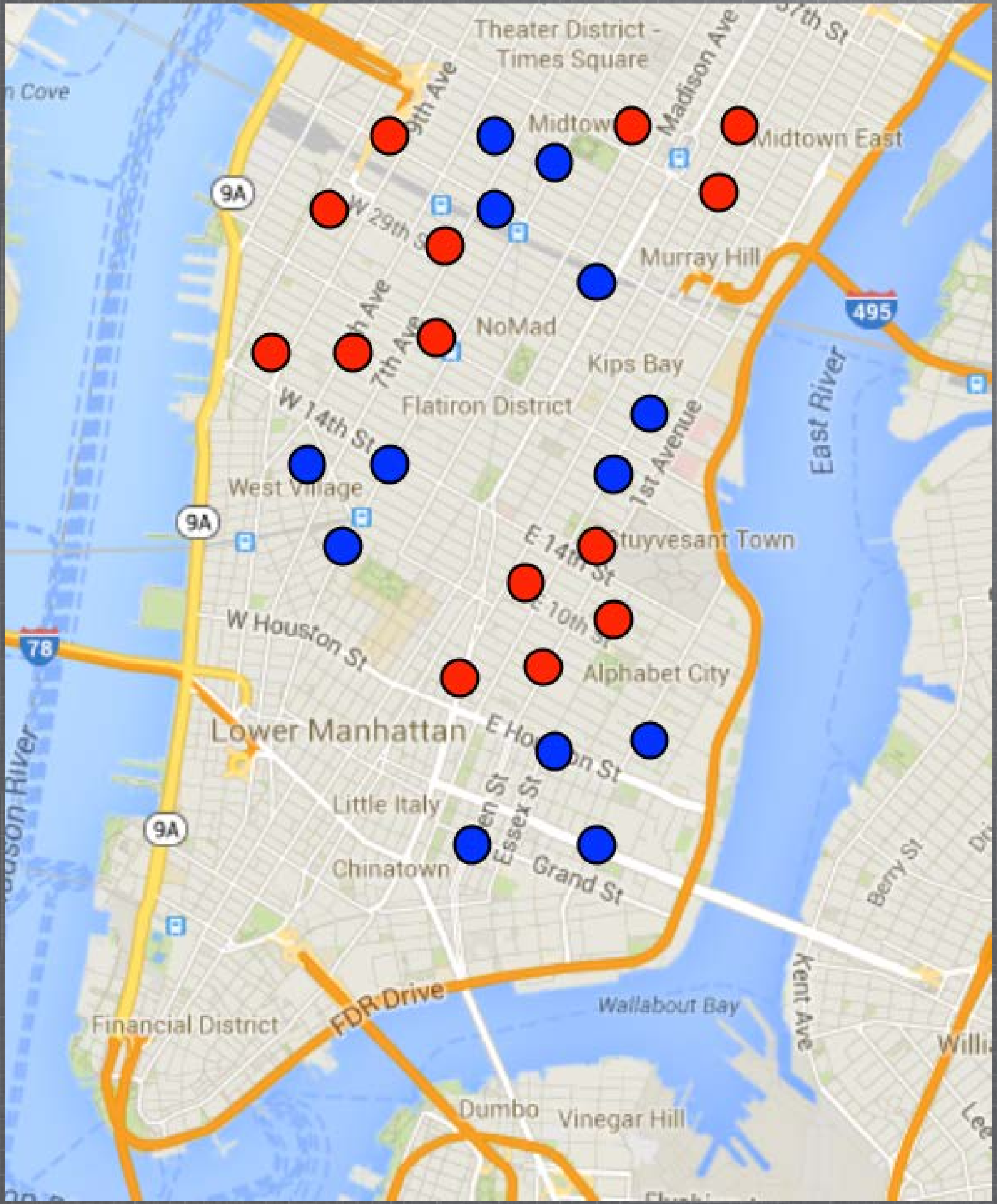
OVERNIGHT REBALANCING

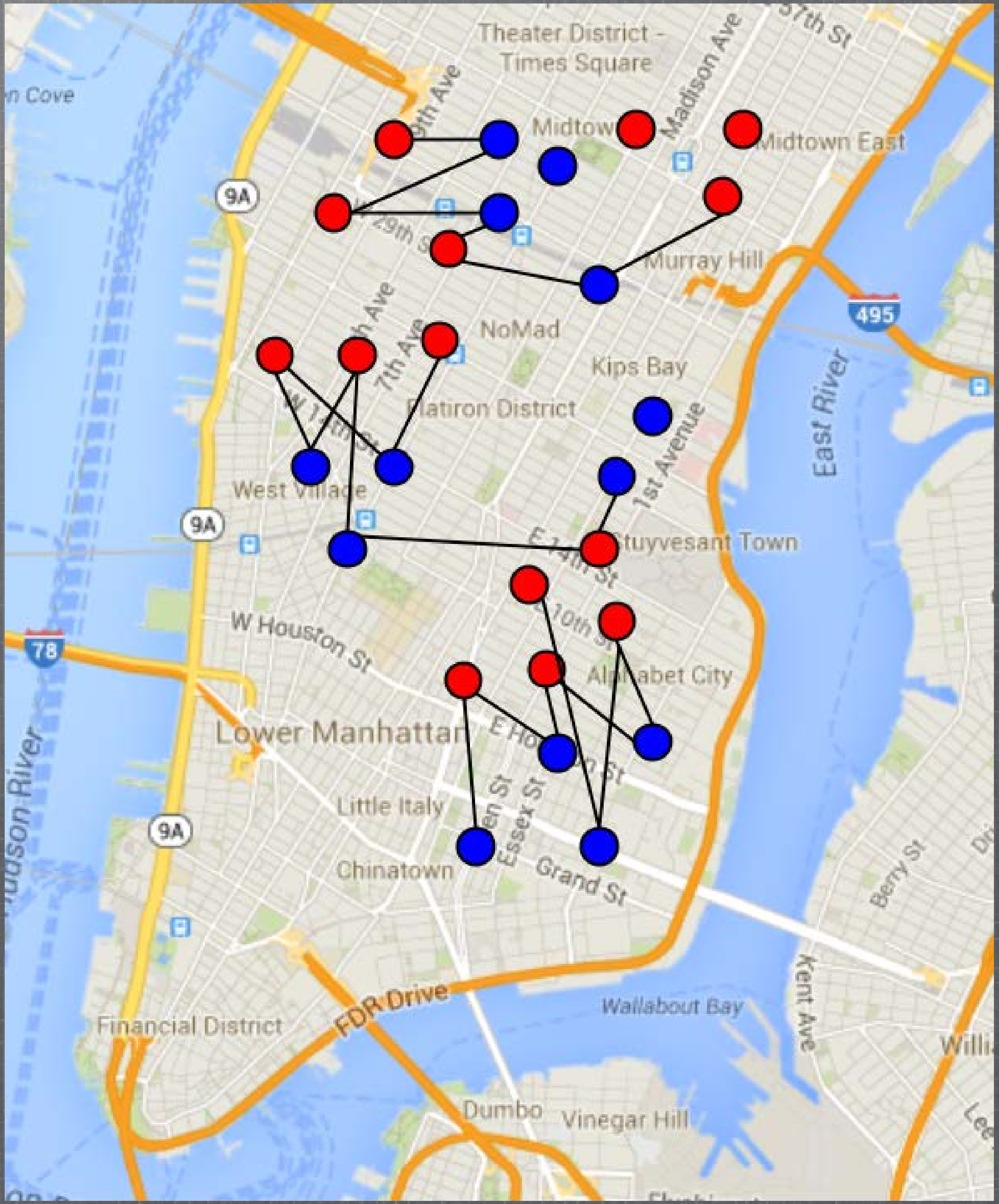
Know where bikes need to be for the AM rush hour

Traffic is light, trucks work

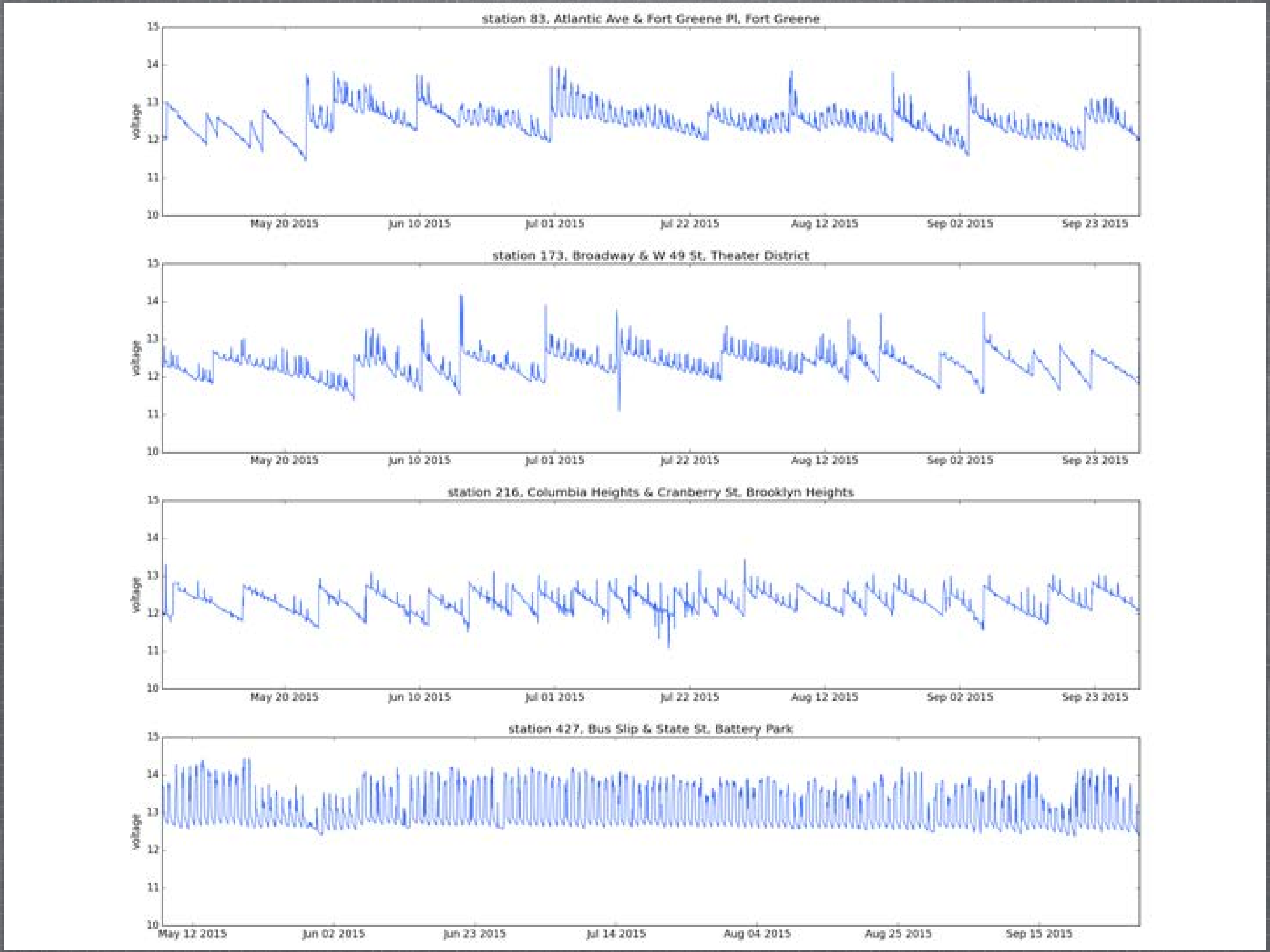
Given an 8 hour shift balance as much as possible

OVERNIGHT REBALANCING

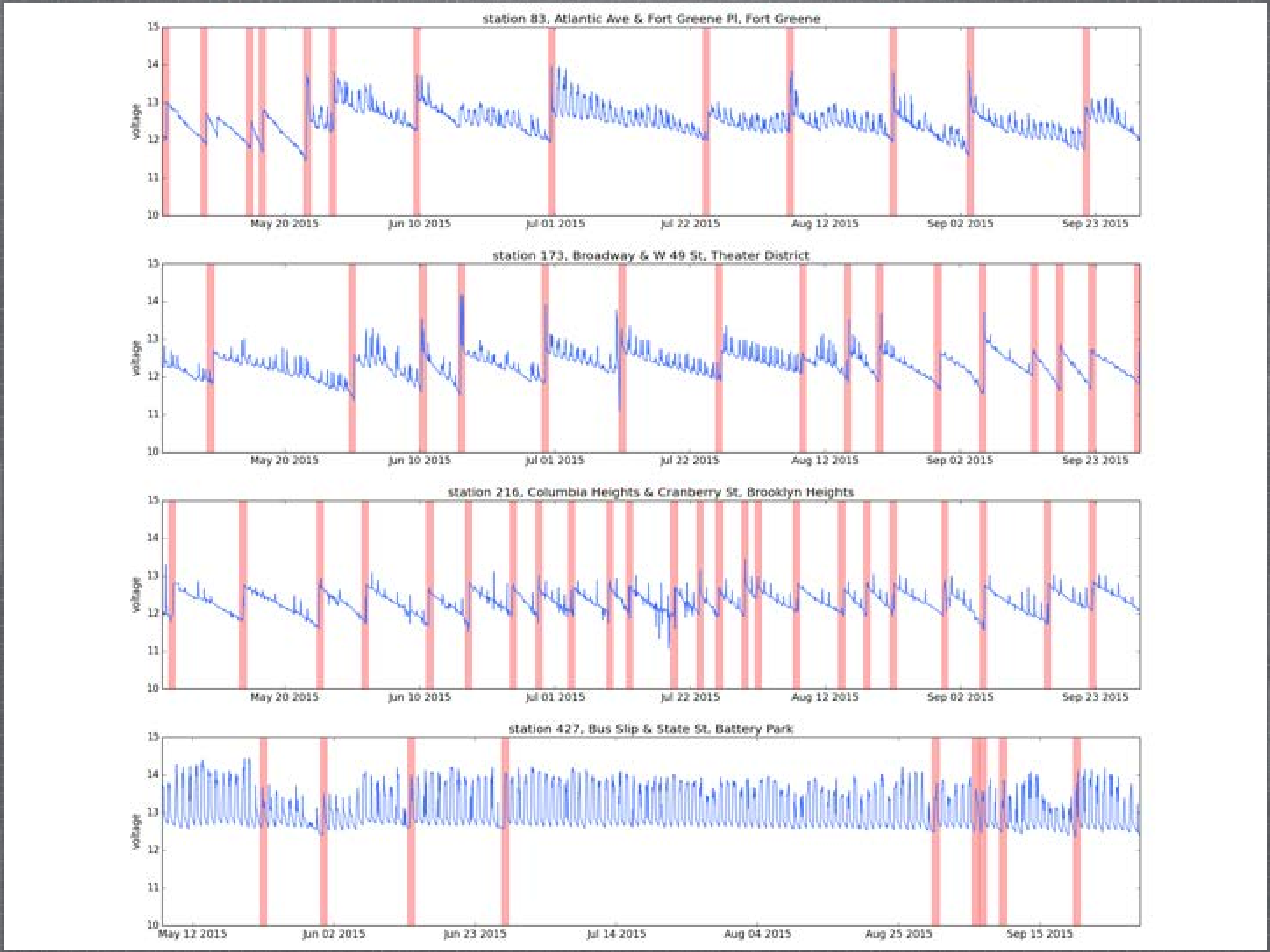




Zen and the art of battery maintenance



Zen and the art of battery maintenance



SUMMARY

Data now exists to build sophisticated models of complex operational systems

Focused on managing operations, but now working on system design as well

Mix of data + math + analytics

New challenges await!

THANK YOU