

Bradley Voytek, Ph.D.
UC San Diego
Cognitive and Neural Dynamics Laboratory

Department of Cognitive Science
Neurosciences Graduate Program
The Institute for Neural Computation

bvoytek@ucsd.edu
@bradleyvoytek



COGS 108
Data Science in Practice

Hello World!



FEATURING
KEYNOTE SPEAKER DON NORMAN
GUEST SPEAKER NATE BOLT



UCSD CSSA PRESENTS
**COGNITION
AT WORK**

ANNUAL COGNITIVE SCIENCE CONFERENCE 2017

PRICE CENTER EAST BALLROOM
SUNDAY, APRIL 9TH 2017
11AM-5PM

Proposed course order

- 2. Why data analysis? (prediction and classification)
- 3. Python!
- 4. Data Science in Python (jupyter, pandas, numpy, scipy, scikit-learn, etc.)
- 5. Data gathering (How do you find and clean data?)
- 6. Data wrangling (JSON, CSV, XML, SQL, APIs)
- 7. Data cleaning
- 8. Data privacy and HIPAA (anonymization)
- 9. Basic data visualization
 - 10. Data intuition and the “sniff test” (Fermi estimation)
 - 11. Linear modeling
 - 12. OLS (optimization)
 - 13. Distributions and outliers
 - 14. Distributions and outliers: CDF, PDFs
 - 15. Multiple linear regression and collinearities
 - 16. Model validation (bootstrapping, resampling, k-fold, leave-p-out, train/test)
 - 17. Feature selection
 - 18. Dimensionality reduction (PCA)
 - 19. Clustering (knn and k-means)
 - 20. Classification (SVM)
 - 21. Interpretability (trees!)
 - 22. Non-parametric statistics
 - 23. NLP and text-mining (tf-idf, sentiment analysis)
 - 24. Geospatial analysis
 - 25. Unsupervised learning (dbSCAN)

GRADING

- Six assignments (10% each)
- Final Project (30%)
- Participation (10%)

OFFICE HOURS

- Mondays 10:00-11:00
- CSB 169

GENERAL STUFF

- First quarter?
- This isn't high school, you're paying for this.
- As an adult, what you do with your time is up to you.

GENERAL STUFF

- If something in lecture or a reading is unclear,
*ask in class, during section, email your TAs, and/or
come to office hours!*

GENERAL STUFF

- If you email a question, the answer to which is on the syllabus... well, I'm sorry—and I'm sure you're amazing—but ***you will not get an email response from us.***

Final Project judging

- Saturday, June 10, 2017 | 1:00a

Final Project judging



Final Project judging



DJ Patil - Fireside chat

- Thursday, June 08, 2017 4:30-6:30p

Guest lectures

- UCSD faculty
- Practicing data scientists:
 - Eli Bressert, Manager, Data Engineering & Analytics, Netflix
 - Mina Doroud, Data Scientist, Twitter (Senior Data Scientist, LinkedIn)
 - Hiroki Hiyama, Senior Data Scientist, Uber
 - Emi Nomura, Data Scientist, Jawbone (Senior Manager, Data Science, Pandora)
 - Kevin Novak, Head of Data Science, Uber
 - Maksim Pecherskiy, Chief Data Officer, City of San Diego
 - Sarah Rich, Data Scientist, Twitter
 - John Myles White, Research Scientist, Facebook (Author: Machine Learning for Hackers)

What is Data Science?



Hilary Mason

@hmason

Following

Replying to @bradleyvoytek

This is the closest thing I've seen in a course description to what practical data science actually is most of the time.

RETWEET

1

LIKES

9



6:44 PM - 17 Feb 2017



1



1



9

What is Data Science?

The study of how the quantification of observable phenomena can lead to human understanding of the processes giving rise to those phenomena—or even the ability to predict future outcomes absent human understanding—and why certain phenomena require more or less data to lead to human understanding and/or prediction accuracy.

What is Data Science?



Josh Wills
@josh_wills

Following

Rule #1 of Hiring Data Scientists: Anyone who wants to do machine learning isn't qualified to do machine learning.

RETWEETS
111

LIKES
257



9:41 PM - 17 Feb 2017

What is Data Science?



Josh Wills
@josh_wills

Following

Rule #1 of Hiring Data Scientists: Anyone who wants to do machine learning isn't qualified to do machine learning.

RETWEETS
111

LIKES
257



9:41 PM - 17 Feb 2017



Josh Wills
@josh_wills

Following

Rule #2 of Hiring Data Scientists: You can get a data scientist to do anything if they believe that what they are doing is machine learning.

RETWEETS
105

LIKES
236



10:14 PM - 17 Feb 2017

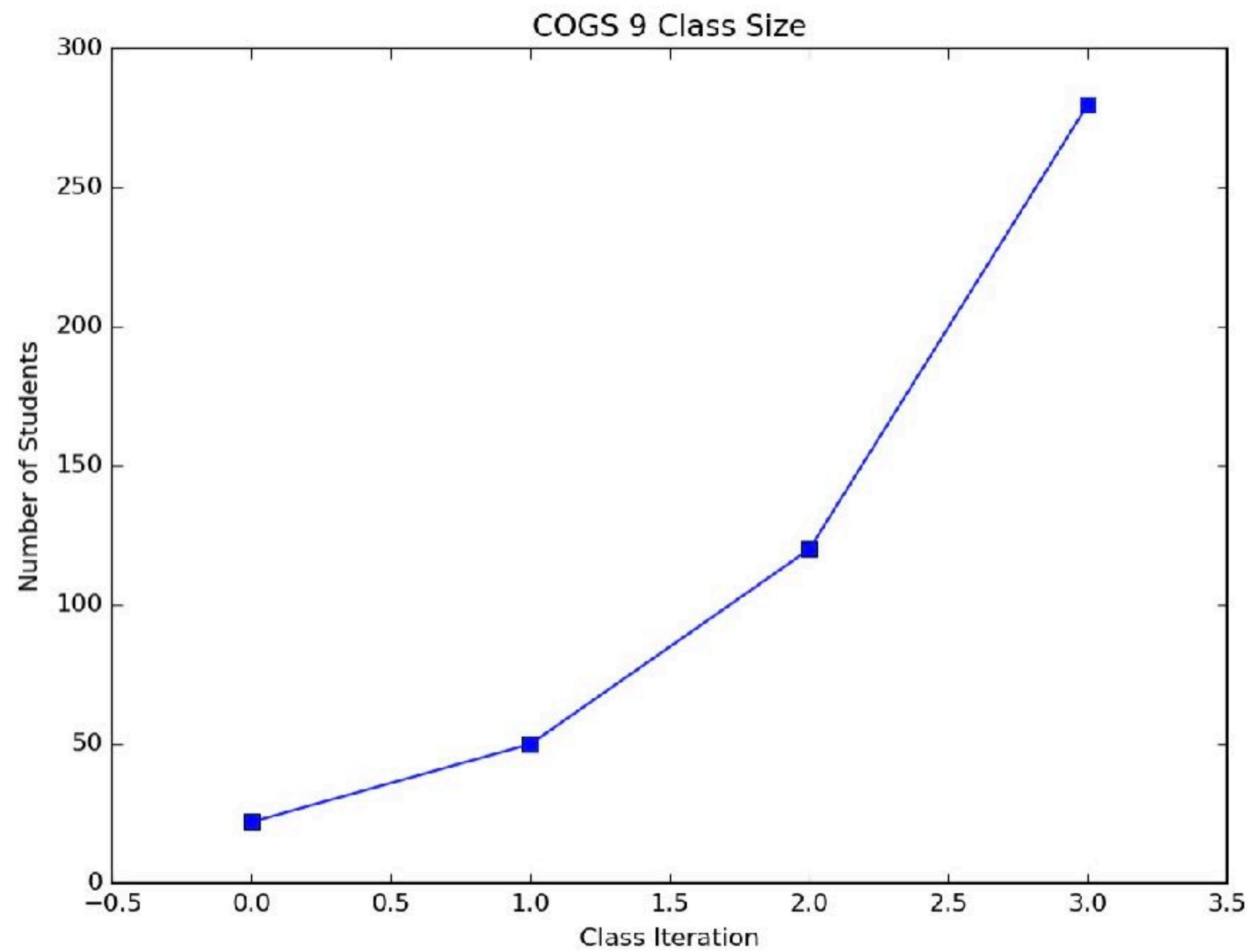
Data Science at UCSD

Facebook pioneer donates \$75 million to UCSD for data science

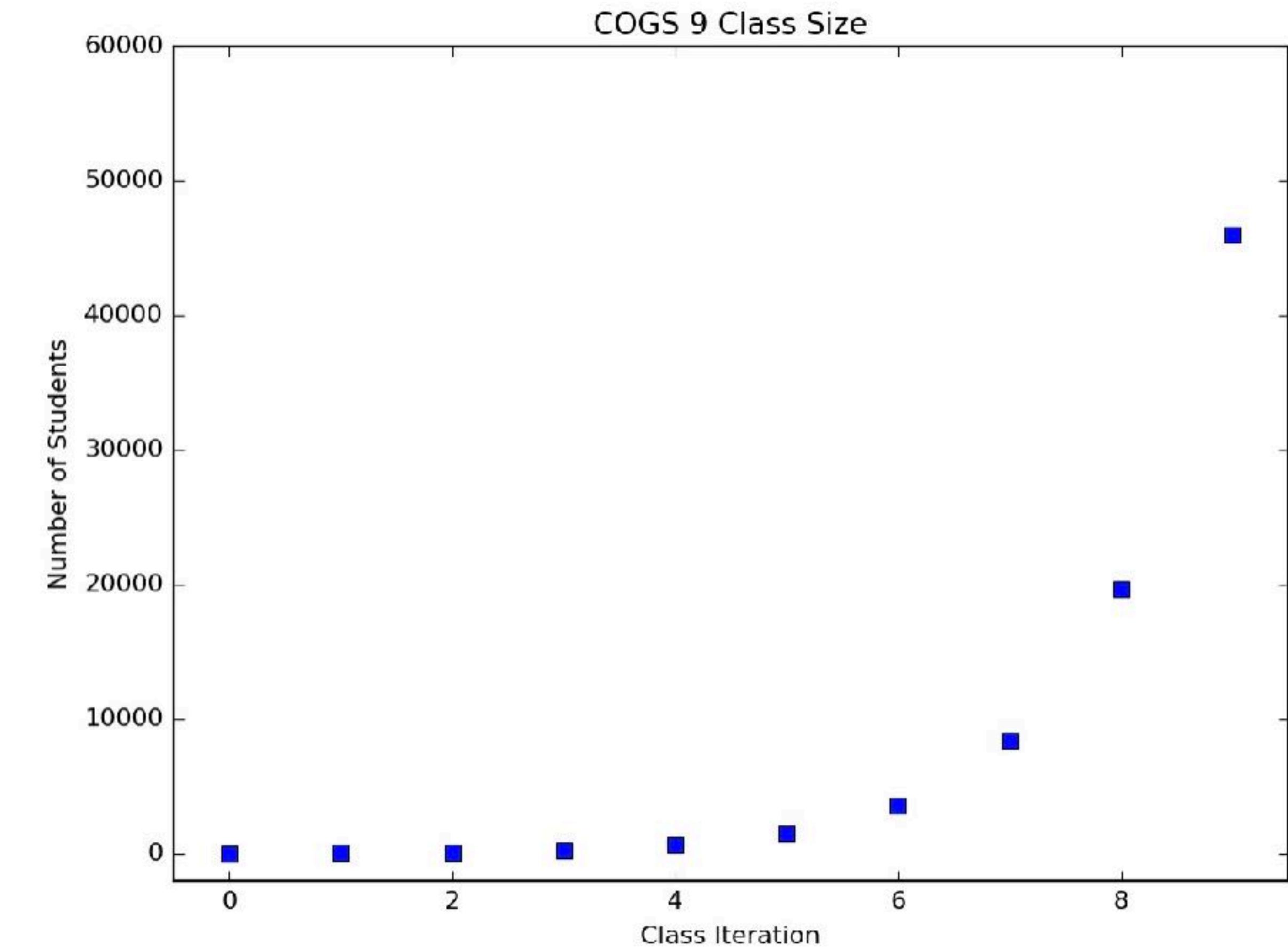
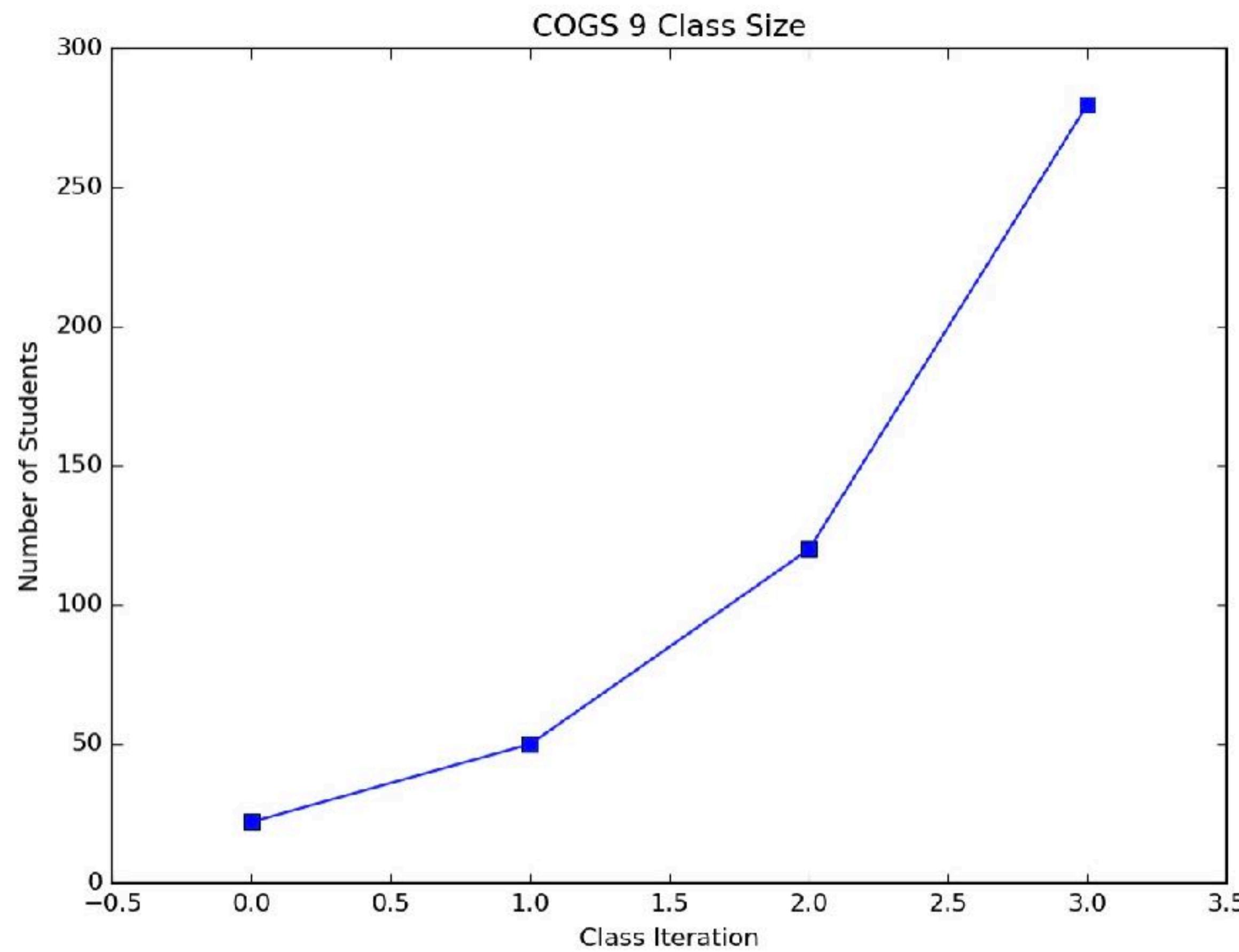


Taner Haliçoglu is donating \$75 million to UC San Diego to make his alma mater a national leader in data science. (Erik Jepsen / UC San Diego)

Data Science at UCSD



Data Science at UCSD



UCSD Data Science Major

- Joint effort between Computer Science, Math, and **Cognitive Science**

UCSD Cognitive Science

Machine Learning and Neural Computation

This area of specialization is intended for majors interested in computational and mathematical approaches to modeling cognition or building cognitive systems, theoretical neuroscience, as well as software engineering and data science. Allowed electives include advanced courses in neural networks, artificial intelligence, and computer science.

Who is this guy?

*Wait... Who are **you**?*

bit.ly/COGSI08SpI7

Who is this guy?

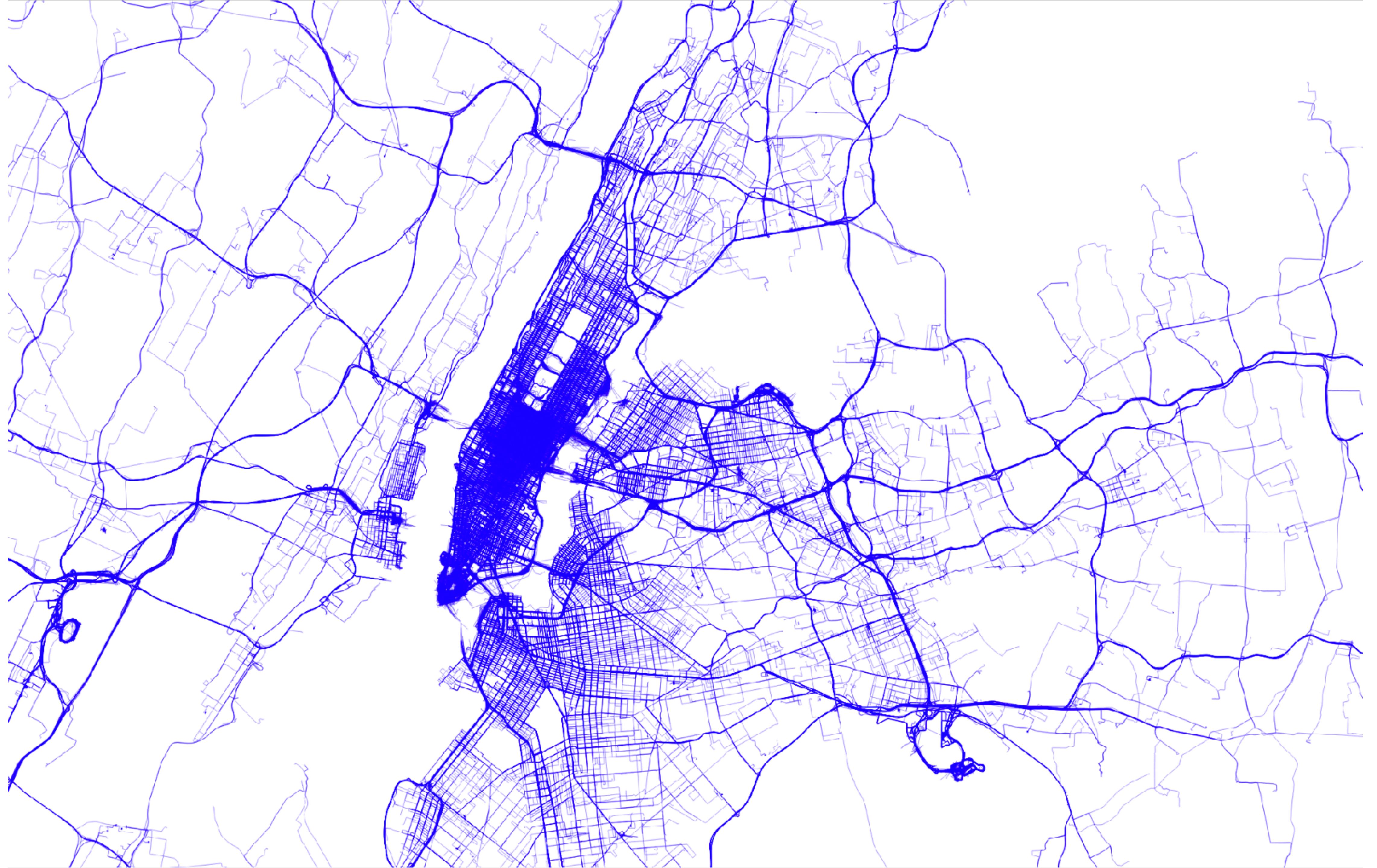




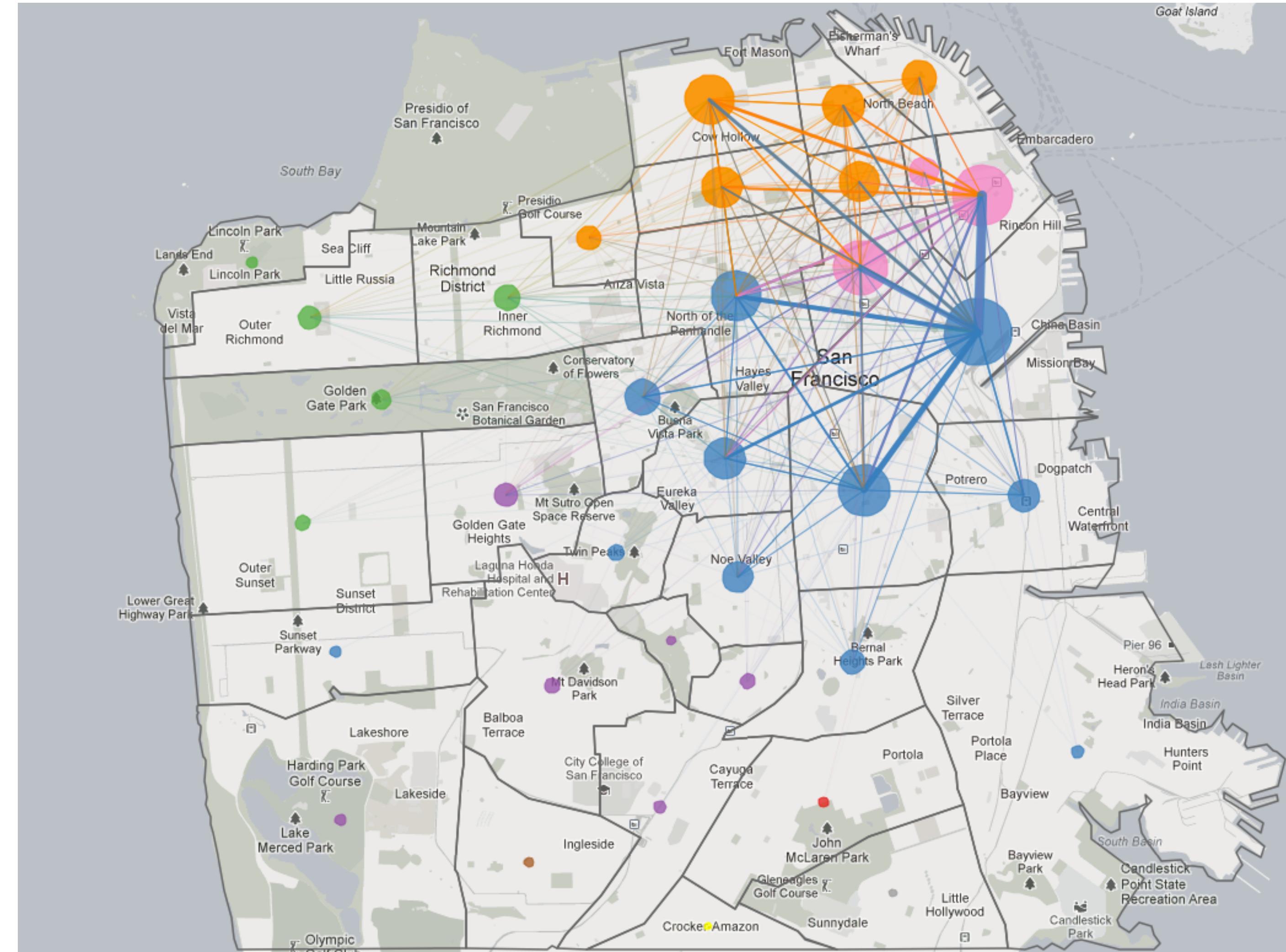
U B E R



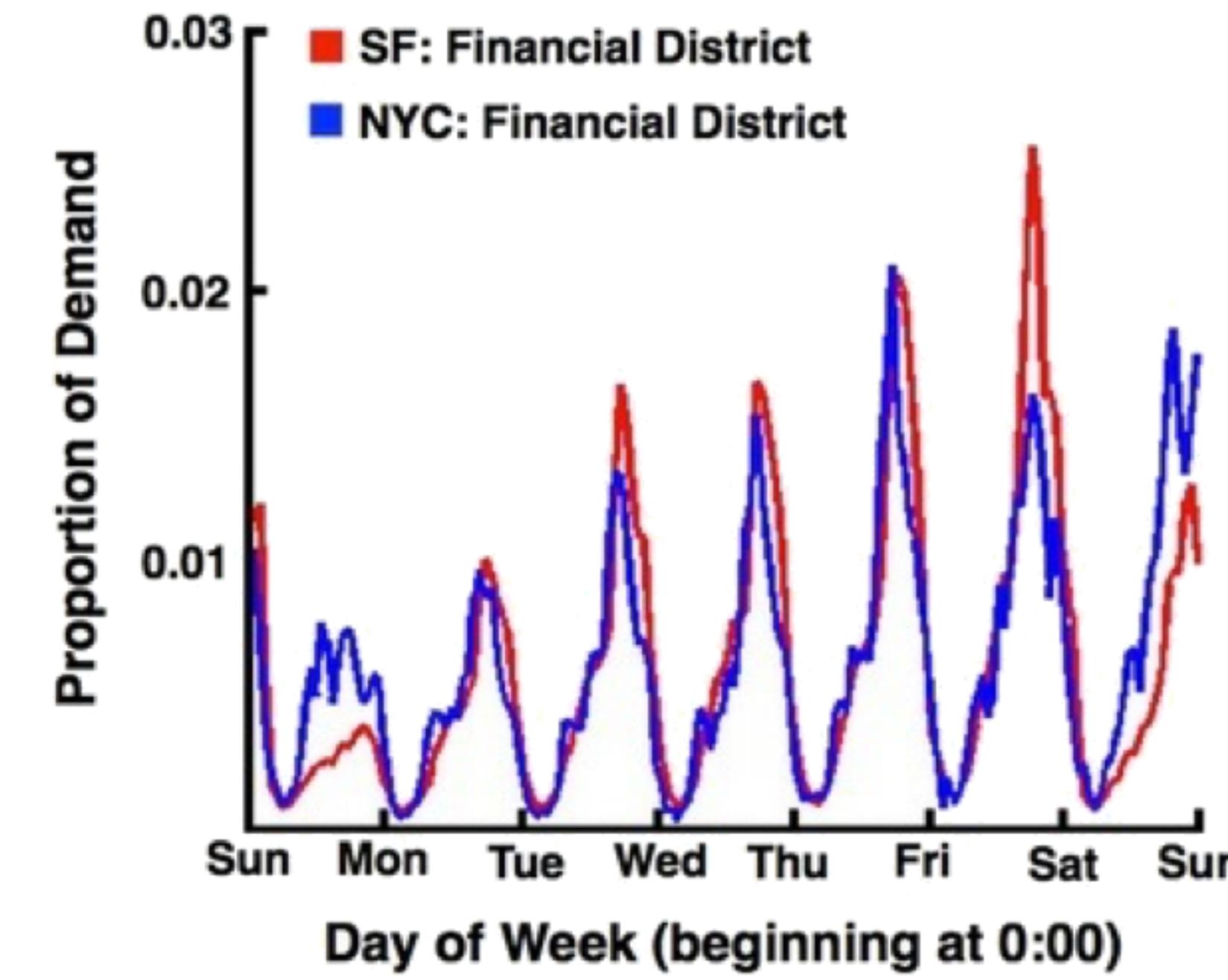
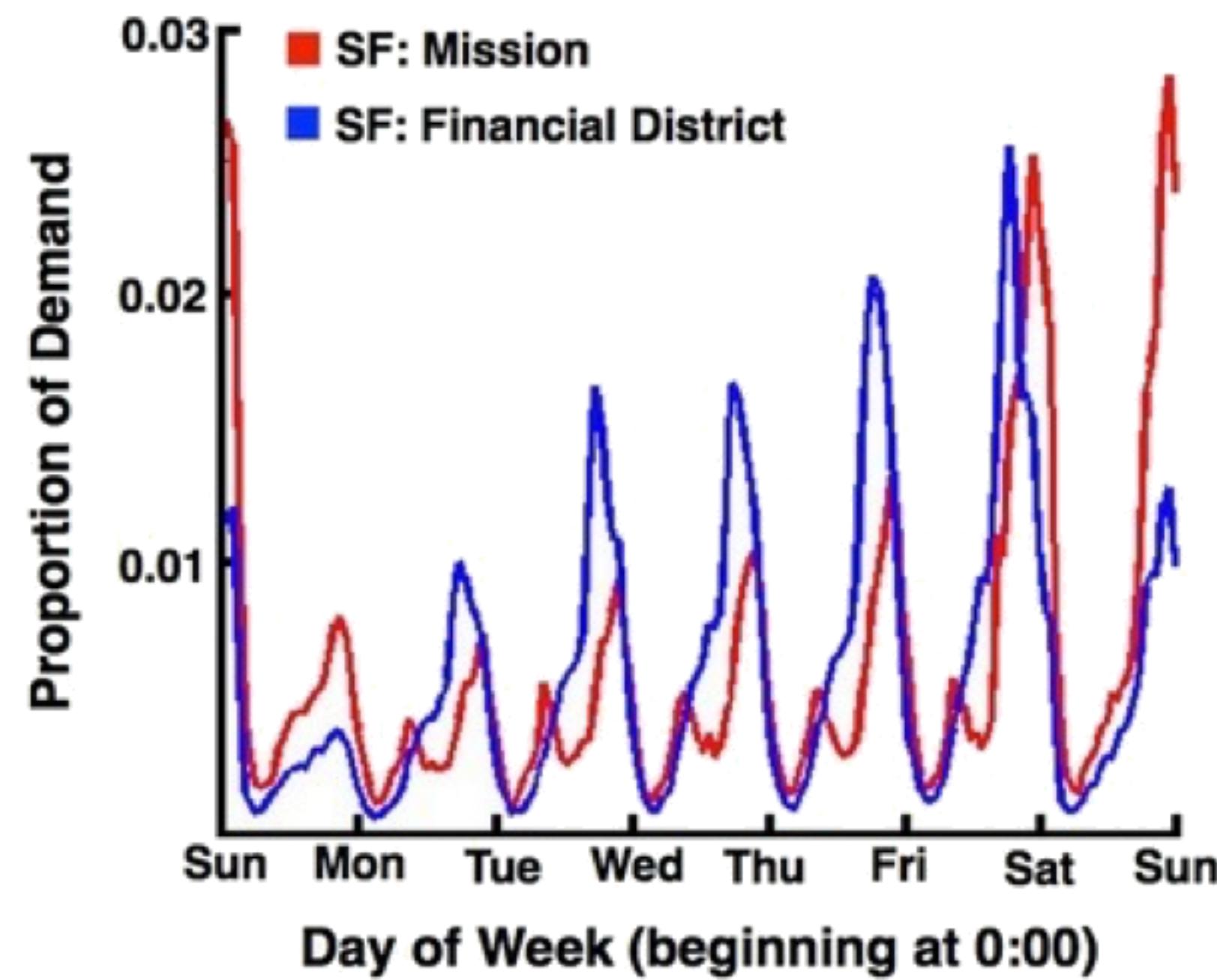


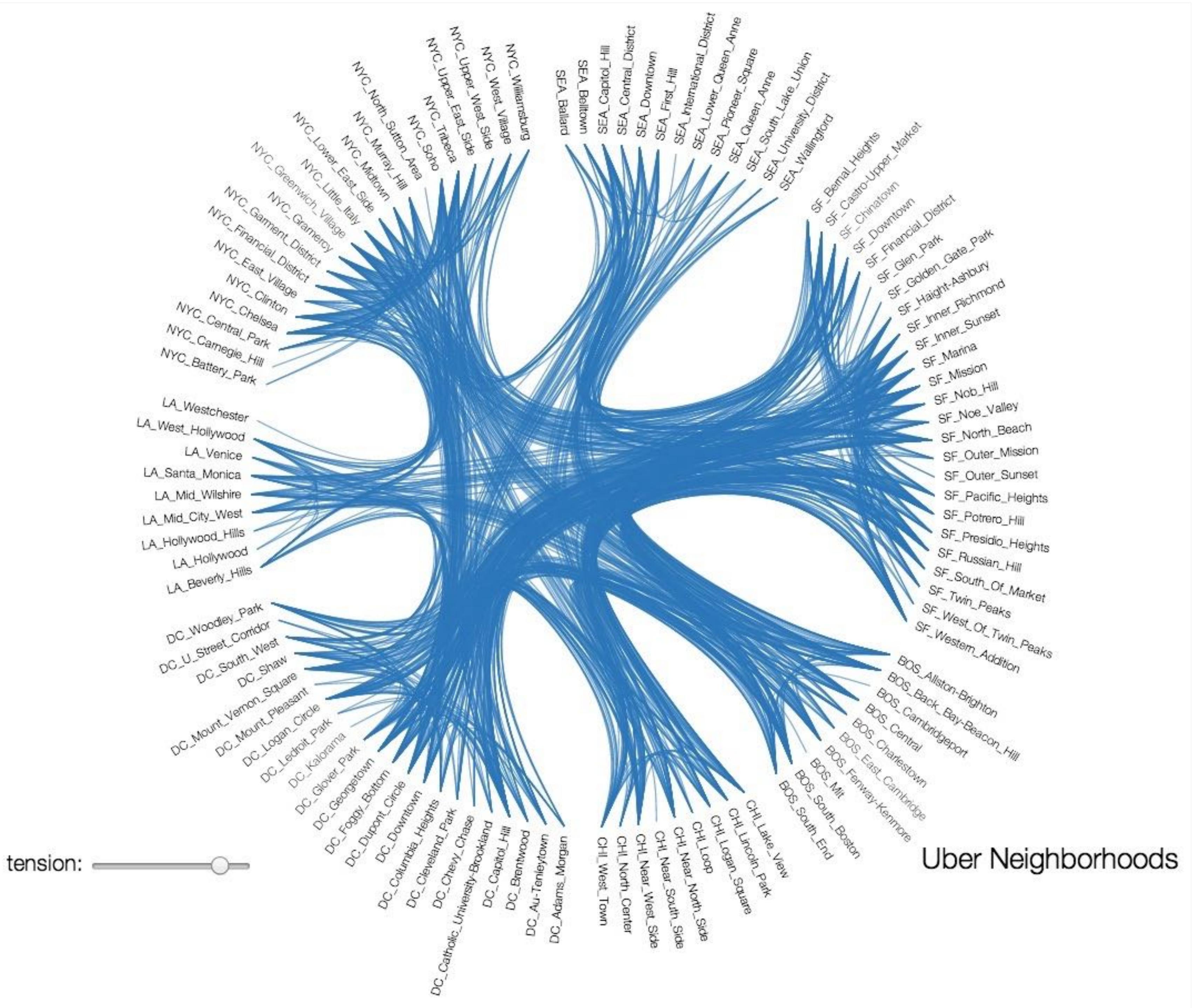


City dynamics

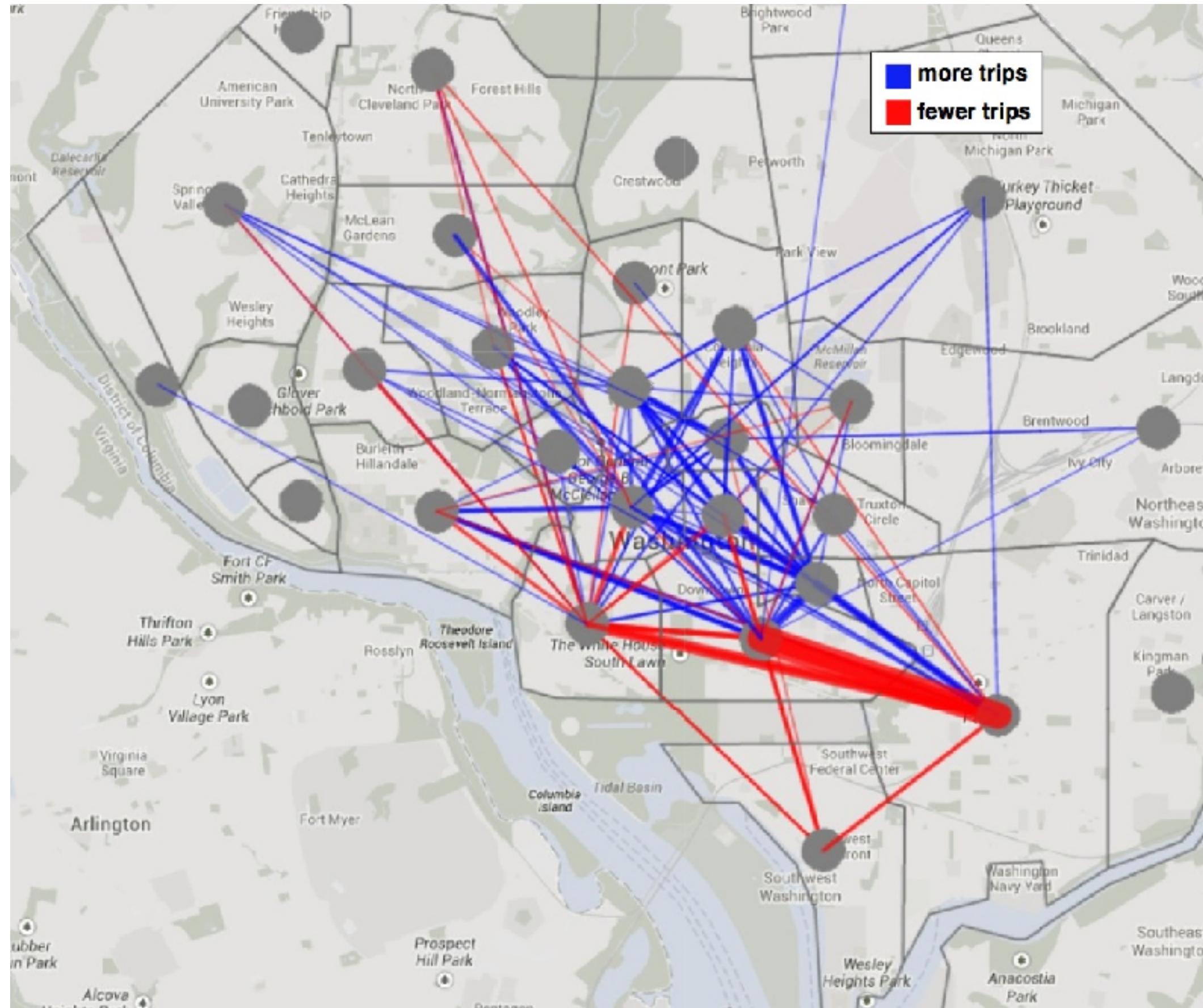


Spatiotemporal dynamics

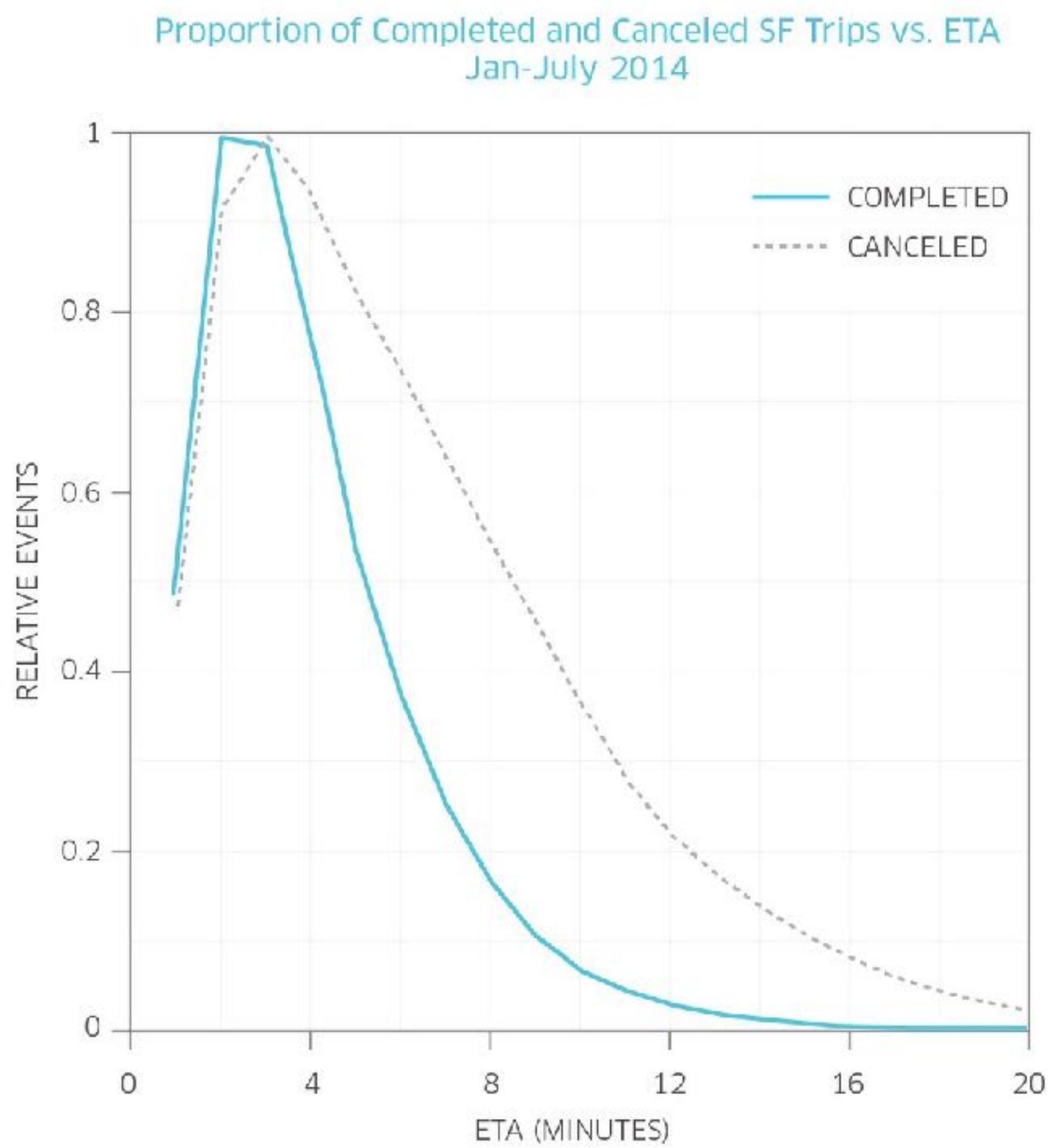




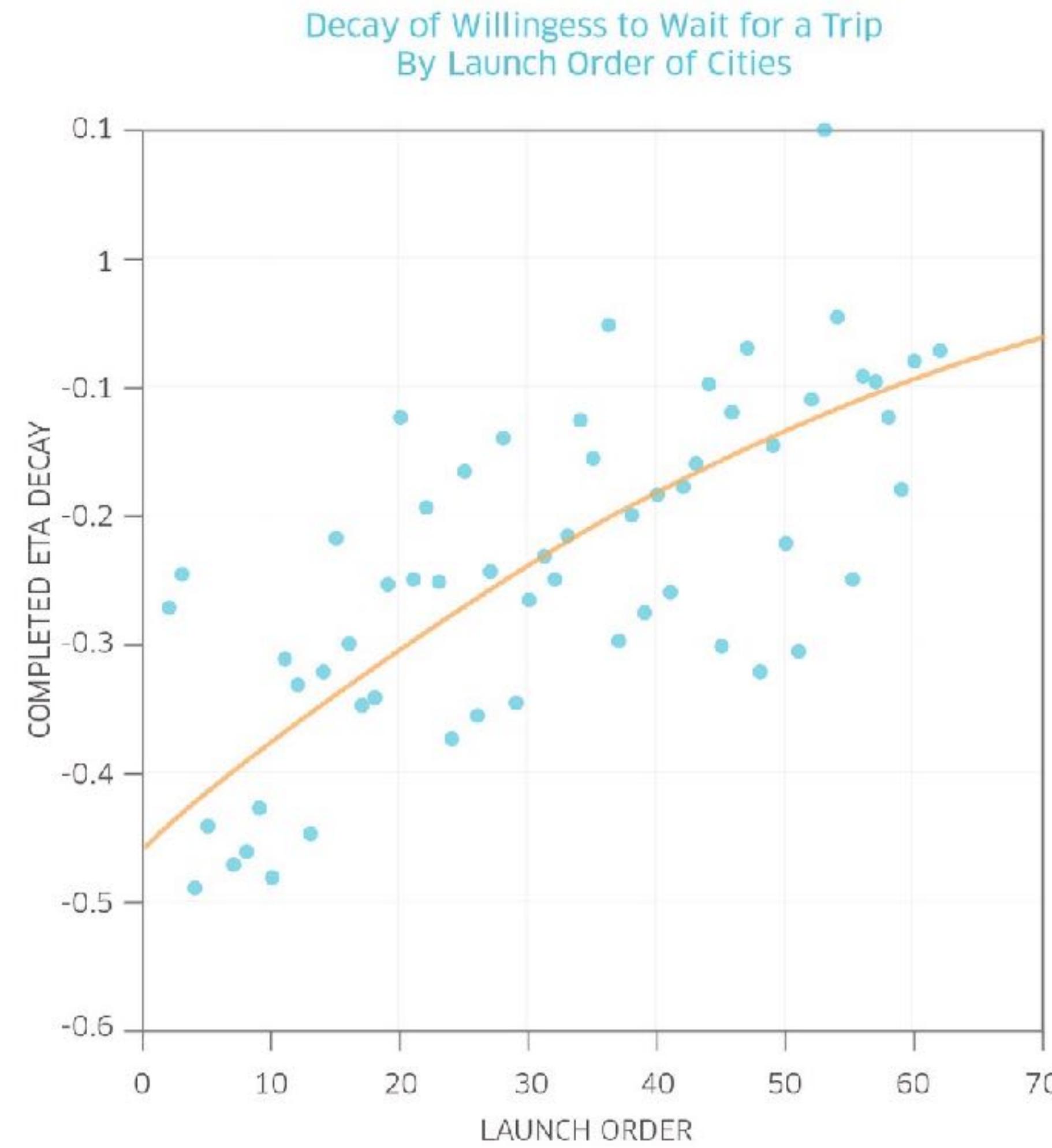
US government shutdown



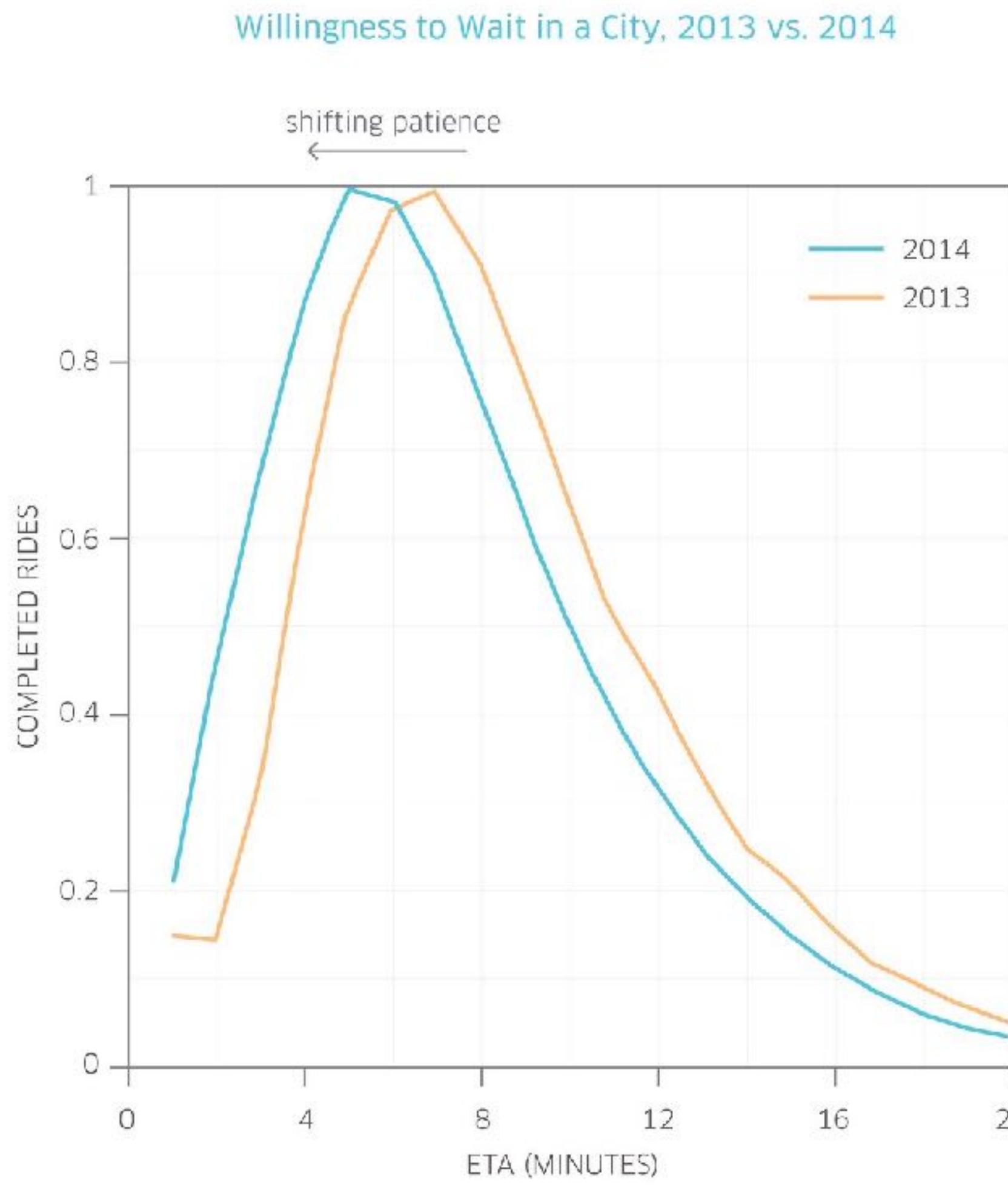
Rider patience



Rider patience

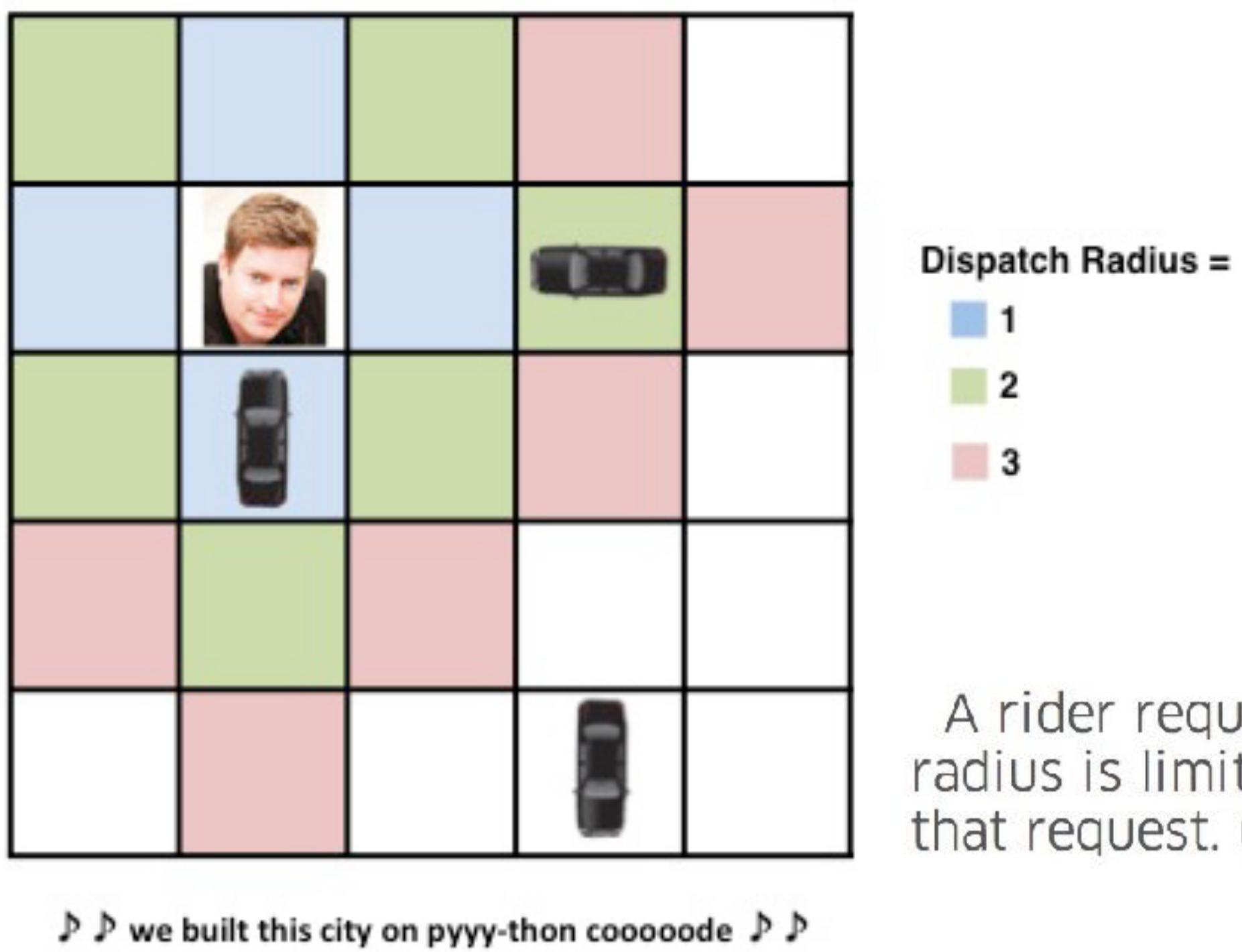


Rider patience



Simulating cities

Uberg



A rider requests a car in the simulated city of Uberg. If the dispatch radius is limited to 1 unit, then only cars within the blue zone will see that request. (Similar to how taxi street hails work: taxis can only pick up the passengers they see.)

Bradley Voytek, Ph.D.
UC San Diego
Cognitive and Neural Dynamics Laboratory

Department of Cognitive Science
Neurosciences Graduate Program
The Institute for Neural Computation

bvoytek@ucsd.edu
@bradleyvoytek

