

Bradley Voytek, Ph.D.  
UC San Diego

Department of Cognitive Science  
Neurosciences Graduate Program  
Halicioglu Data Science Institute

bvoytek@ucsd.edu  
@bradleyvoytek

UC San Diego

COGS 9  
Introduction to Data Science

*Hello World!*

# GRADING

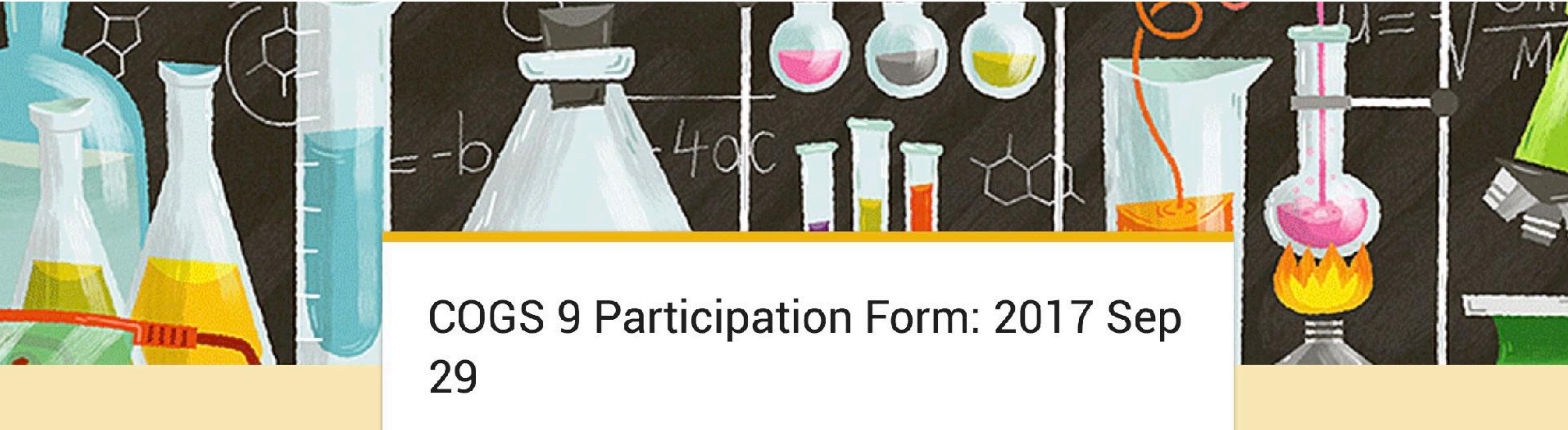
- Four assignments (15% each)
- Final project (30%)
- Participation (10%)

# Participation

- Attendance required for guest lectures

# Guest lecturers

- Franziska Bell, PhD: Senior Data Science Manager, Uber
- Claire Dorman, PhD: Data Scientist, *Pandora*
- Mina Doroud, PhD: Data Scientist, *LinkedIn*
- Carlos Gomez-Uribe, PhD:
  - Formerly: Statistician, Google; VP Product Innovation, Netflix
  - Currently: Director, News Feed Algorithm Integrity, Facebook



COGS 9 Participation Form: 2017 Sep  
29

<http://bit.ly/cogs9fa17q1>

# Course links

- Piazza: <https://piazza.com/ucsd/fall2017/cogs9/home>
- TritonEd

# Scheduling

**Lecture:**

MWF 10:00a-10:50a

Pepper Canyon Hall (PCYNH) 106

**Sections:**

TBA

**Voytek's Office Hours:**

Mondays 11:00a-12:00p and  
by appointment

# General stuff and junk

- First quarter?
- Data Science majors?
- 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 and junk

- If something in lecture or a reading is unclear:
  - *ask in class*
  - *ask during section*
  - *email your TAs*
  - *post on Piazza*
  - *come to office hours!*

# Why this course?

You are going to be analyzing lots of data because you're studying to be a:  
**cognitive scientist**

# Why this course?

You are going to be analyzing lots of data because you're studying to be a:  
**neuroscientist**

# Why this course?

You are going to be analyzing lots of data because you're studying to be a:  
**statistician**

# Why this course?

You are going to be analyzing lots of data because you're studying to be a:  
**computer scientist**

# Why this course?

You are going to be analyzing lots of data because you're studying to be a:  
**tech guru**

# Why this course?

You are going to be analyzing lots of data because you're studying to be a:  
**CEO/small business owner**

# Why this course?

You are going to be analyzing lots of data because you're studying to be a:  
**political activist**

# Why this course?

You are going to be analyzing lots of data because you're studying to be a:  
**journalist**

# Why this course?

Because.....

DATA

## Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

Harvard  
Business  
Review



UC San Diego

# ntist: The Sexiest e 21st Century

. Patil

**sexiest**

*Who is this guy?*





# A young UCSD scientist vents about how hard it is to obtain grants



UC San Diego neuroscientist Bradley Voytek. (K.C. Alfred / Union-Tribune) (K.C. Alfred / UT San Diego/Zuma Press)

# A young UCSD scientist vents about how hard it is to obtain grants



UC San Diego neuroscientist Bradley Voytek. (K.C. Alfred / Union-Tribune) (K.C. Alfred / UT San Diego/Zuma Press)



# Scientists: Advertise Your Failures!

They're a part of every career, and being upfront about them can help put things in perspective

By Elizabeth Landau on September 25, 2017

You might have encountered Bradley Voytek through his writings and fan-convention talks about the brains of zombies, or through his more serious day job as assistant professor of computational cognitive science and neuroscience at the University of California, San Diego. But some know Voytek because he's open about just how hard it was to get to where he is now.



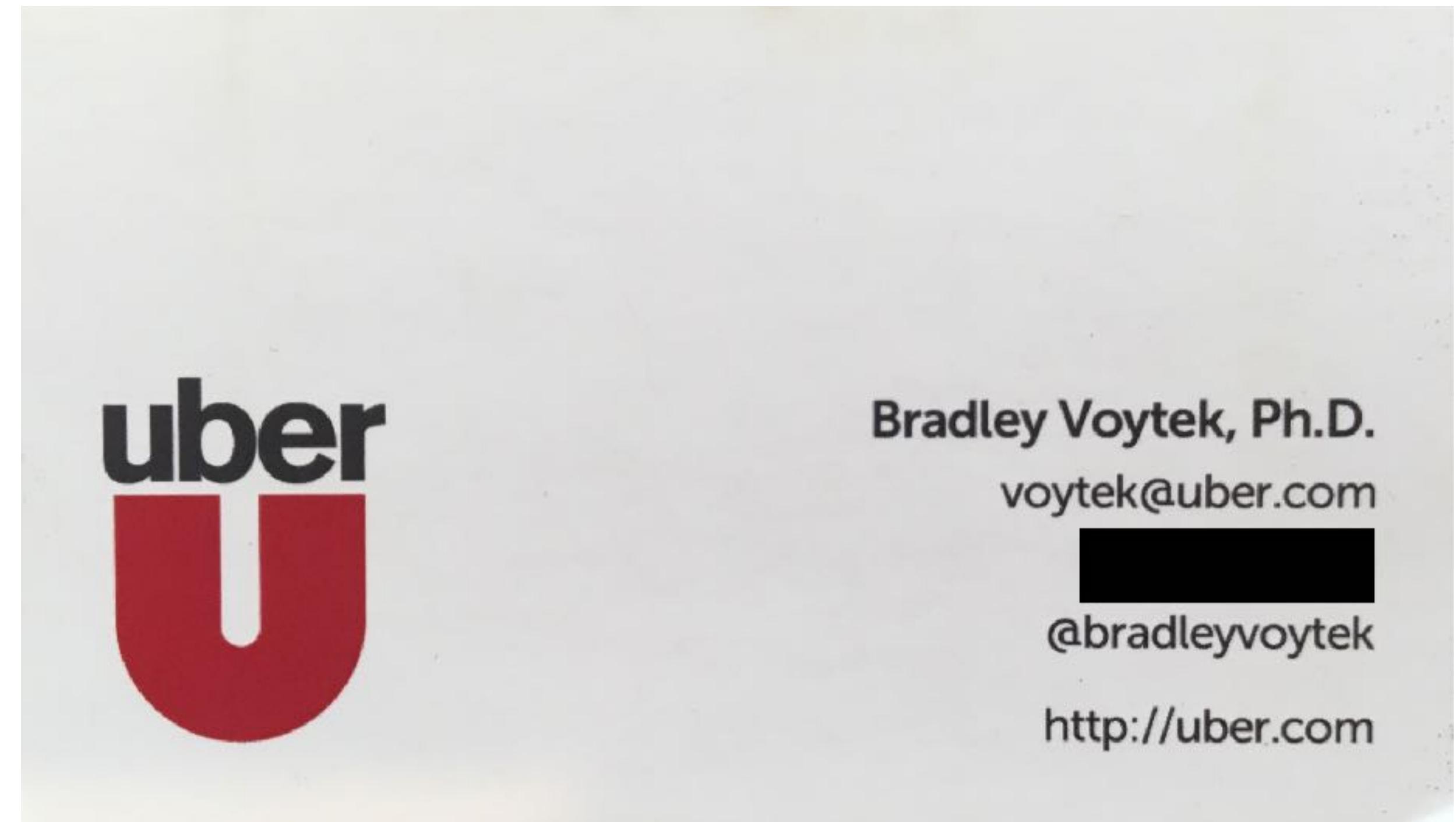
# Scientists: Advertise Your Failures!

They're a part of every career, and being upfront about them can help put things in perspective

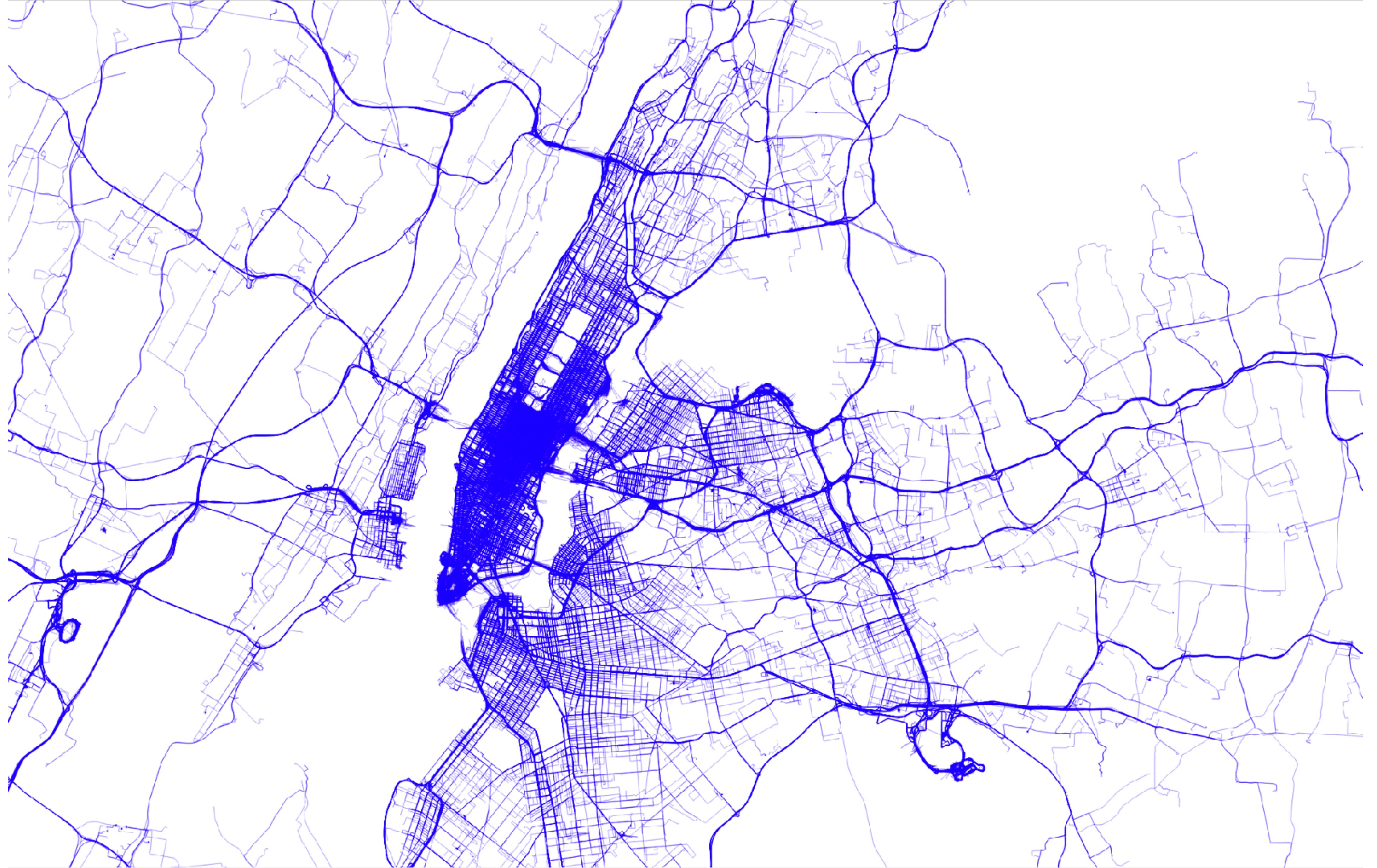
By Elizabeth Landau on September 25, 2017

Most academics are not so vocal about the setbacks they've faced, but Voytek's story has gained traction. Online discussion boards about graduate admissions have invoked Voytek's name to show that it's possible to enter a Ph.D. program—he went to the University of California, Berkeley—with a subpar transcript. One even asks: "Is the neuroscientist Bradley Voytek a real or fictional person?"

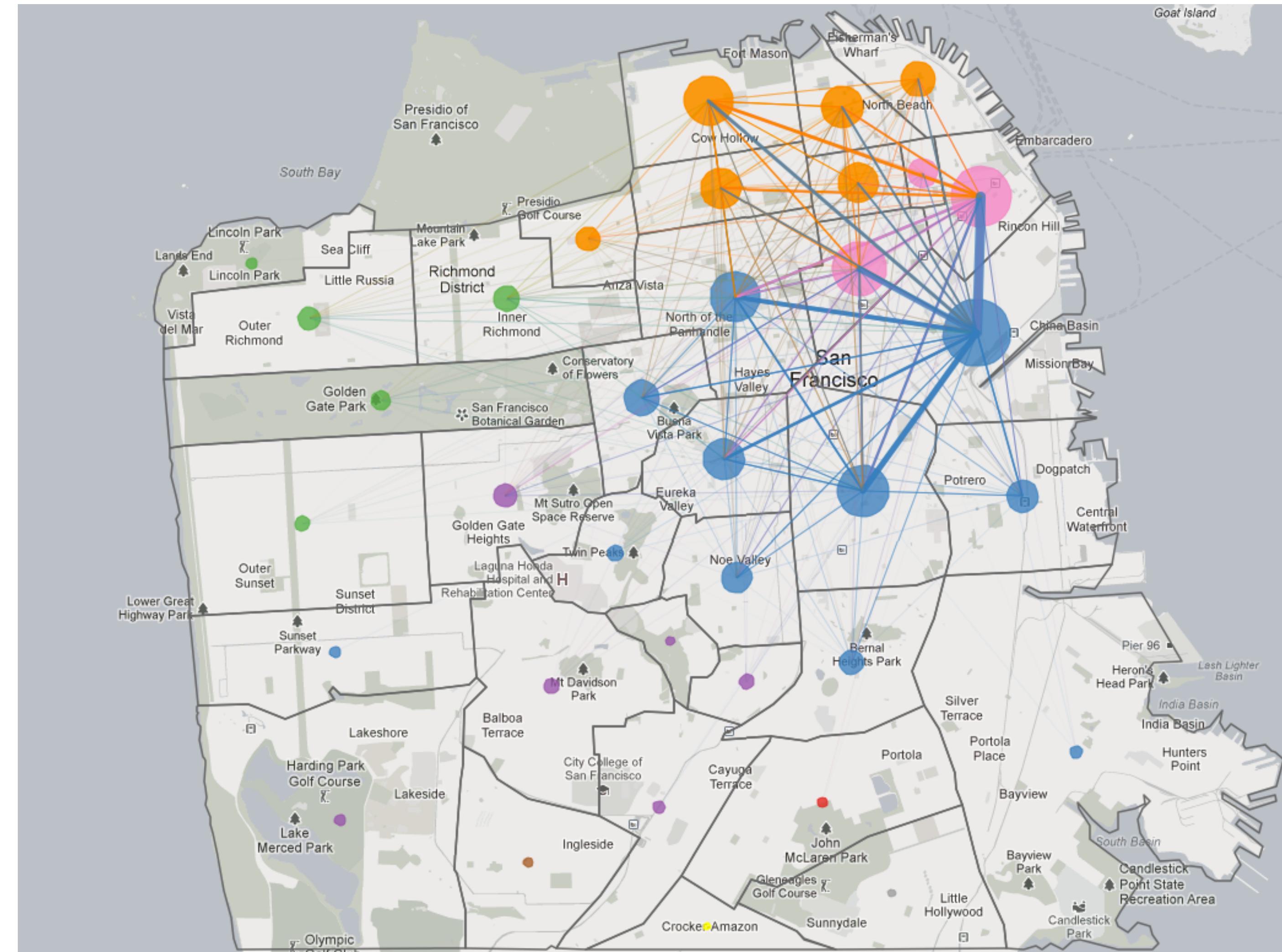
# UBER



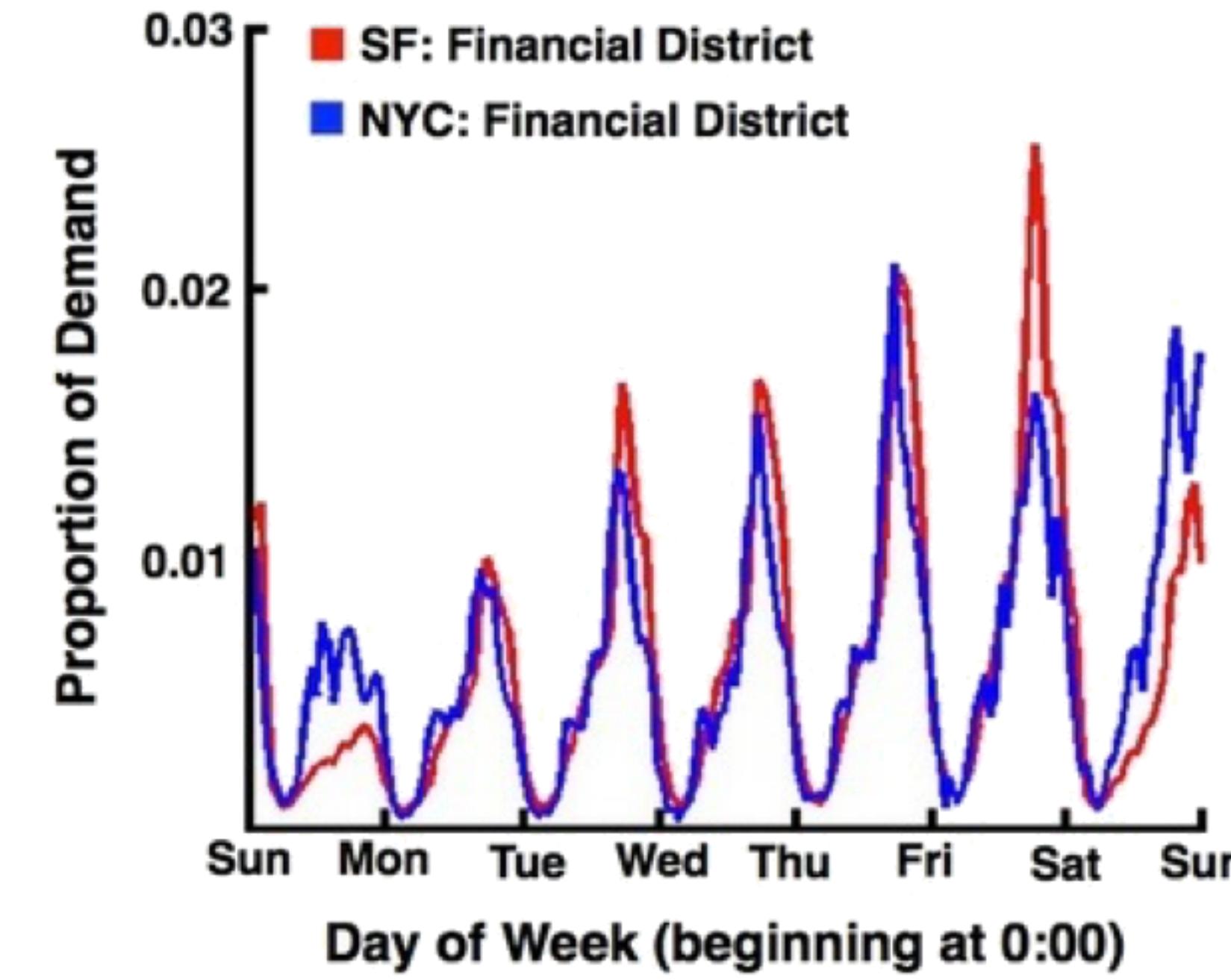
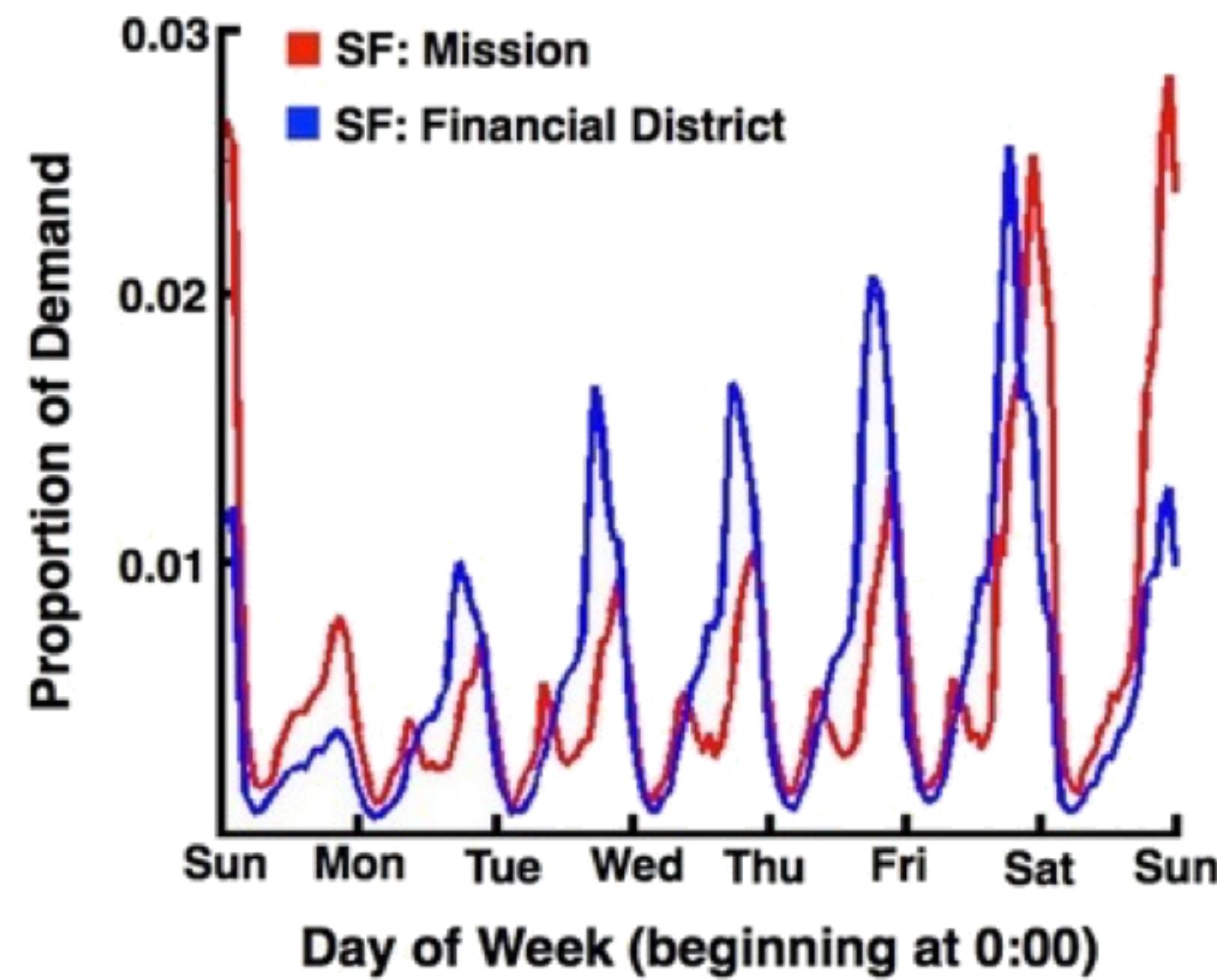


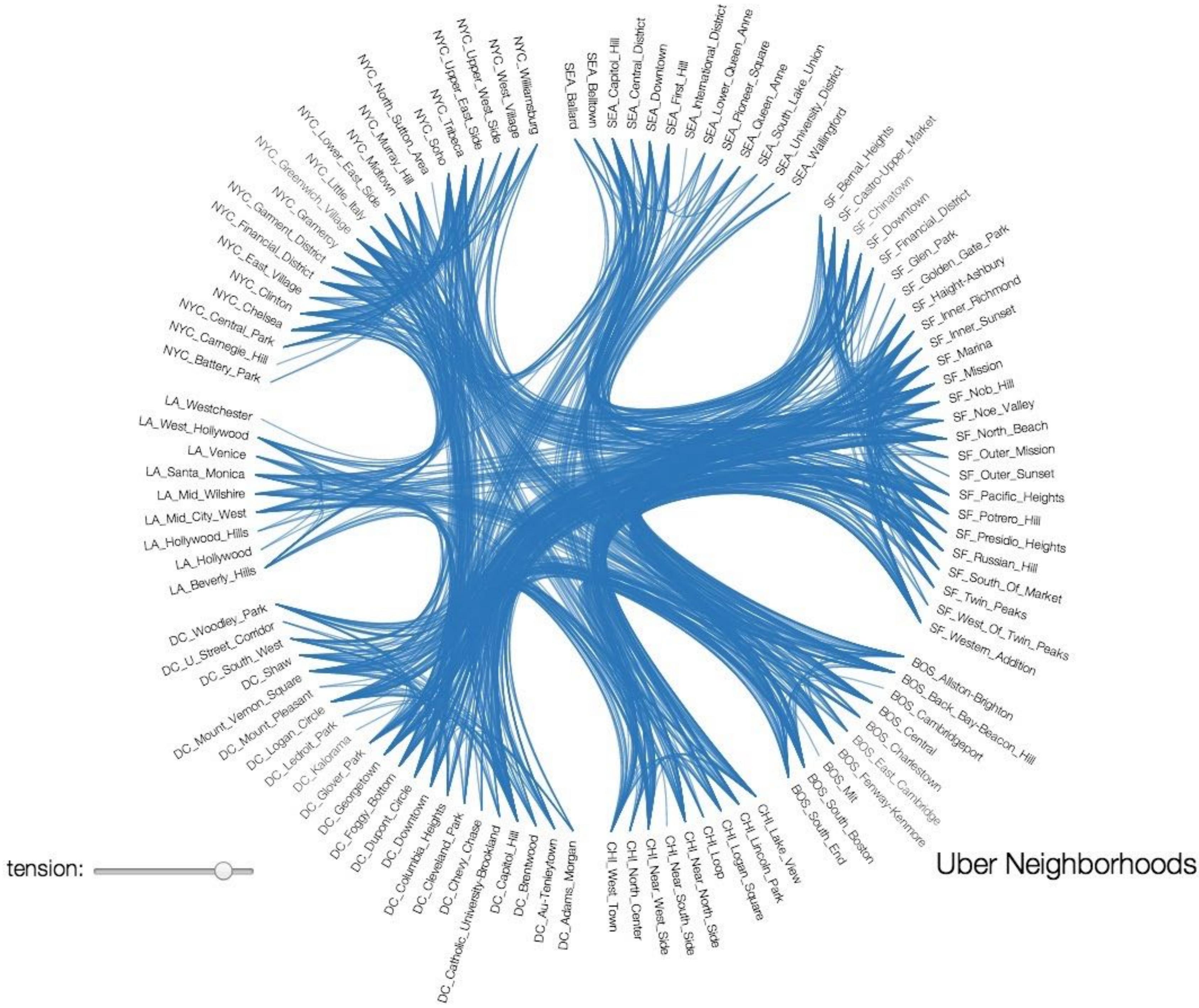


# City dynamics



# Spatiotemporal dynamics

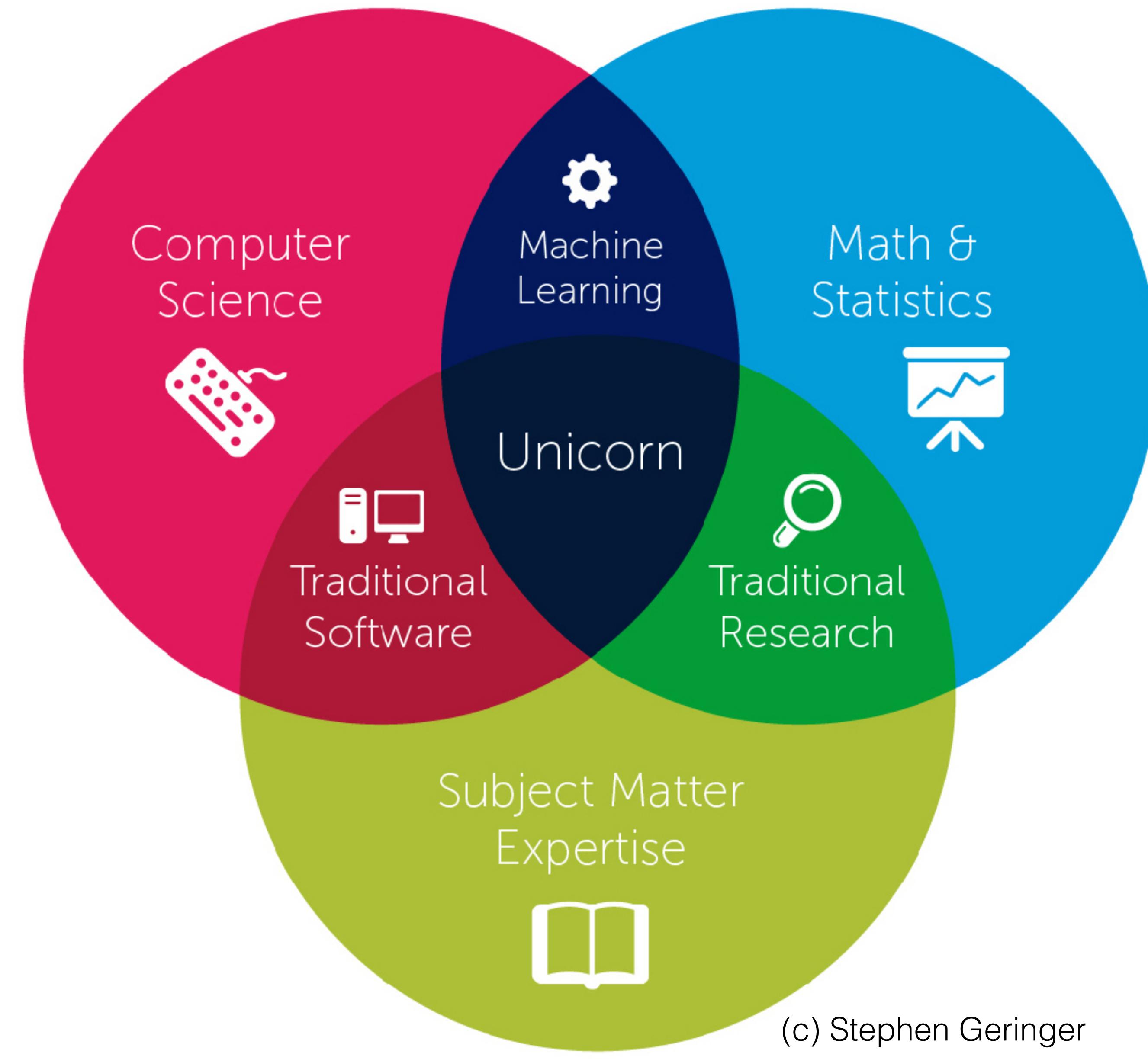




# A New UCSD Major & Minor

- Joint effort between the departments of Computer Science, Math, and Cognitive Science
- Industry & research demand for the “data scientist”

# Data Science



# UCSD Data Science Major

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

# What is Data Science?

“The best minds of my generation are thinking about how to make people click ads. That sucks.”



Source: Jeff Hammerbacher, Cofounder (Cloudera) and Assistant Professor (Icahn School of Medicine at Mount Sinai)

**“So WTF are we?”**

BIG DATA

# Data science

BIG DATA

# Data science still woefully short on science

BIG DATA

# Data science still woefully short on science

**KD**nuggets™ The Data Science Delusion

BIG DATA

# Data science still woefully short on science



**“data science isn’t science”**

# Data Science (cf Computer Science)

“The first computer science degree program in the United States was formed at Purdue University in 1962.”



# Data Science (cf Computer Science)

“Since practical computers became available, many applications of computing have become distinct areas of study in their own rights.”



# What is Data Science?



The image shows the header section of the NYU Data Science website. It features the NYU logo (a white square with a purple torch icon) and the text "DATA SCIENCE AT NYU" in white on a dark purple background. Below the header is a navigation bar with links: "About", "What is data science?", "Research", "Academics", "News", and "Contact Us".

**NYU**

**DATA SCIENCE AT NYU**

[About](#)   [What is data science?](#)   [Research](#)   [Academics](#)   [News](#)   [Contact Us](#)



Source: NYU Data Science

# What is Data Science?



The image shows the header of the NYU Data Science website. It features the NYU logo (a torch icon) and the text "DATA SCIENCE AT NYU". Below the header is a navigation bar with links: "About", "What is data science?", "Research", "Academics", "News", and "Contact Us". The link "What is data science?" is highlighted with a yellow oval.

NYU

DATA SCIENCE AT NYU

About    What is data science?    Research    Academics    News    Contact Us



Source: NYU Data Science

# What is Data Science?

## **What is Data Science?**

There is much debate among scholars and practitioners about what data science is, and what it isn't. Does it deal only with big data? What constitutes big data? Is data science really that new? How is it different from statistics and analytics?

# What is Data Science?

## What is Data Science?

There is much debate among scholars and practitioners about what data science is, and what it isn't. Does it deal only with big data? What constitutes big data? Is data science really that new? How is it different from statistics and analytics?

????!!!!???

# What is Data Science?



Data Science Studies  
Berkeley

This working group designs and conducts research projects across disciplines and methods to understand the challenges posed by data scientists' practices in the academic context. In collaboration with our partner institutions, it also develops innovative quantitative and qualitative metrics to measure the evolution of data science environments.

# What is Data Science?

## A New Field Emerges

There is significant and growing demand for data-savvy professionals in businesses, public agencies, and nonprofits. The supply of professionals who can work effectively with data at scale is limited, and is reflected by rapidly rising salaries for data engineers, data scientists, statisticians, and data analysts.

# What is Data Science?

## A New Field Emerges

There is significant and growing demand for data-savvy professionals in businesses, public agencies, and nonprofits. The supply of professionals who can work effectively with data at scale is limited, and is reflected by rapidly rising salaries for data engineers, data scientists, statisticians, and data analysts.

The field of data science is emerging at the intersection of the fields of social science and statistics, information and computer science, and design. The UC Berkeley School of Information is ideally positioned to bring these disciplines together and to provide students with the research and professional skills to succeed in leading edge organizations.



# What is Data Science?

## A New Field Emerges

There is significant and growing demand for data-savvy professionals in businesses, public agencies, and nonprofits. The supply of professionals who can work effectively with data at scale is limited, and is reflected by rapidly rising salaries for data scientists, and data analysts.



The field of data science is emerging from the intersection of social science and statistics, information and computer science, and design. The UC Berkeley School of Information is ideally positioned to bring these disciplines together and to provide students with the research and professional skills to succeed in leading edge organizations.

# What *isn't* 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 isn't 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

# What isn't 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

**DATA SCIENCE ISN'T  
MACHINE LEARNING**

Following

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

# What is Data Science?

"Good data scientists understand, in a deep way, that the heavy lifting of cleanup and preparation isn't something that gets in the way of solving the problem: it is the problem."

# Data Science - Defining a field

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.

# Data Science - Defining a field

**Data Science is different from “using data to come to conclusions in science”**

# Data Science - Defining a field

The scientific method uses data to come to conclusions about physical phenomena and make predictions about it...

# Data Science - Defining a field

In contrast **Data Science** is the study of:

- 1) How and why data can be used that way, what kinds of data are there.
- 2) What makes "good" versus "bad" quality data for different questions, etc.

# Data Science vs. Data Engineering

Data Science is the **empirical study of data** whereas  
Data Engineering is the **application of data science**  
**methods and techniques to draw conclusions.**

# Data Science vs. Data Engineering

This is directly comparable to the difference between Computer Science (the scientific study) and Computer Engineering (the application and, commonly, job title).

*The Human Side of Data Science*

*or*

*Why is Cognitive Science a core DS department?*



Bradley Voytek, Ph.D.  
UC San Diego

Department of Cognitive Science  
Neurosciences Graduate Program  
Halicioglu Data Science Institute

bvoytek@ucsd.edu  
@bradleyvoytek

UC San Diego