Bradley Voytek, Ph.D.

UC San Diego

Department of Cognitive Science
Halıcıoğlu Data Science Institute
Neurosciences Graduate Program

bvoytek@ucsd.edu
voyteklab.com

COGS 9
Introduction to Data Science

UC San Diego

GRADING

- Four assignments (10% each)
- Two exams (12.5% each)
- Five readings (5% each after lowest is dropped)
- Final project (10%)
- Guess lecture attendance (5% total for 3? lectures)
- Two free points *that you can* ***lose***

Guest lectures

- Attendance *required* for guest lectures

Past guest lecturers

- Franziska Bell, PhD: Senior Data Science Manager, *Uber*
- Eli Bressert, PhD: Manager, Data Engineering & Analytics, *Netflix*
- Claire Dorman, PhD: Data Scientist, *Pandora*
- Mina Doroud, PhD: Data Scientist, *Twitter* (Senior Data Scientist, *LinkedIn*)
- Carlos Gomez-Uribe, PhD: Director, Core Data Science, *Facebook* (Statistician, *Google*; VP Product Innovation, *Netflix*)
- Hiroki Hiyama, PhD: Senior Data Scientist, *Uber*
- Emi Nomura, PhD: Senior Manager, Data Science, *Pandora*
- Kevin Novak: Head of Data Science, *Uber*
- DJ Patil, PhD: US Chief Data Scientist, *Obama White House* (Head of Data Science, *LinkedIn*; co-coined the term "data science")
- Maksim Pecherskiy: Chief Data Officer, *City of San Diego*
- Sarah Rich, PhD: Data Scientist, *Twitter*
- John Myles White, PhD: Research Scientist, *Facebook* (Author: *Machine Learning for Hackers*)
- Josh Wills, PhD: Head of Data Engineering, *Slack*

This quarter's guest lecturers

- **Jenni Tremaine, Ph.D.**
  - Director, Data Insights: *Sony*
- **Emily Kubicek, Ph.D.**
  - Manager, Data Strategy: *Disney*
- **Duncan Watson-Parris, Ph.D.**
  - Assistant Professor: *UC San Diego Data Science*

Free points!

- See the syllabus.
- Everyone in the class begins the quarter two bonus points that amount to a 0.5% boost to your overall, final grade. These are "fudge" points to get folks who are very close to the next grade level up and over the mark at the end of the quarter.

Free points!

- It is possible to have these bonus points taken away.
- Any student that messages me, the TAs, or IAs with requests asking for free points, or to "bump up" their grade "just a little bit" because you're "so close", or to get another chance on an exam, assignment, or reading that wasn't already cleared with me for a verifiable excuse will be ignored.

Free points!

- *Furthermore, the free bonus points that I gave will be removed from any student who makes these kinds of grade-begging requests.* These kinds of requests are unfair to other students who do not make such unreasonable requests, and it would be very unfair to everyone else who wouldn't get the same special treatment.

Free points!

- Note that questions to get clarification, or regarding possible grading errors or mistakes, or regrade requests are *not unreasonable nor inequitable*, and will not be penalized!
- I welcome these kinds of questions.

Other stuff

- Data Science majors?

Other stuff

- 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.

Other stuff

- If something in lecture or a reading is unclear:
  - *ask in the Q&A*
  - *attend discussion sections!*
  - *post on Piazza*
  - *<u>come to our one-on-one hours</u>!*

Who is this guy?

VOYTEKlab

Cognitive Science

Eena Kosik

Quirine van Engen

Dillan Cellier

Postdocs

Pamela Riviere Ruiz

Christian Cazares

Data Science

Ryan Hammonds

Neuroscience

Andrew Bender

Blanca Burgos-Martin

Morgan Fitzgerald

Trevor McPherson

Michael (MJ) Preston

Sydney Smith

Zombies!

WIRED

THE SCIENCE OF SURVIVING THE ZOMBIE APOCALYPSE

THE ZOMBIE DISORDER

CONSCIOUSNESS DEFICIT HYPOACTIVITY DISORDER

Consciousness Deficit Hypoactivity Disorder (CDHD): The loss of rational, voluntary and conscious behavior replaced by delusional/impulsive aggression,stimulus-driven attention, the inability to coordinate motor linguistic behaviors and an insatiable appetite for human flesh.

SCANS OF THE ZOMBIE BRAIN

ZOMBIE HUMAN

Through detailed scans, the exact brain areas that have been destroyed in the zombie can be reconstructed.

The scans show significant brain tissue loss in the zombie. The gray area shows what a human brain would look like. The profile of damage corroborates the behavioral observations of zombies.

TEDEd

Discover Create Support

Diagnosing a zombie: Brain and body (Part one) - Tim Verstynen & Bradley Voytek

738,663 Views

4,758 Questions Answered

TEDEd Animation

232,378 Views

2,946 Questions Answered

TEDEd Animation

Copyrighted Material

TIMOTHY VERSTYNEN
BRADLEY VOYTEK

DO ZOMBIES DREAM OF UNDEAD SHEEP?

A NEUROSCIENTIFIC VIEW OF THE ZOMBIE BRAIN

Copyrighted Material

WIRED

★★★★★ **Five Stars**

By rks1125 on February 6, 2015

Format: Hardcover | **Verified Purchase**

As described

I love brains.

I love data.

Measuring the brain is difficult.

Brains!

voxel-level information

architecture
- cortical thickness
- white matter connectivity
- myelination
- neuronal density

cellular
- cell type and morphology
- gene expression
- receptors
- neurotransmitters / modulators

dynamics
- single-unit
- field potential / oscillations

function
- cognitive correlates
- disease relationships
- development and aging

I care about *how* we measure things.

More specifically…

How we transform our

data

into

numbers

that we perform

statistics

on is very important.

So we *write (Python)* code
and
develop mathematical algorithms

to understand
(and help others understand)

“how the brain works”

Bycycle!

Amplitude = (| + |) / 2

Period = — + —

Rise-decay symmetry = — /(— + —)

Peak-trough symmetry = — /(— + —)

Trough time	Amplitude	Period	rdsym	ptsym
0.12	2.9	105	0.43	0.51
0.21	1.7	89	0.34	0.5
0.31	1.9	99	0.4	0.49
0.4	2.2	82	0.32	0.52
0.5	2.2	109	0.4	0.58
0.61	2.2	108	0.44	0.57

Source: Cole & Voytek, *J Neurophysiol* 2019

UC San Diego

Parameterizing neural power spectra

Source: Donoghue*, Haller*, Peterson*, … Shestyuk†, Voytek†, *Nature Neurosci* 2020 Code: https://specparam-tools.github.io/

UC San Diego

Parameterizing ~~neural~~ power spectra

Fig. 5. Spectrum of the QCM with a peak frequency of $f = 29\,\mathrm{kHz}$ for the spectral line at $\lambda = 587\,\mathrm{nm}$. This line also shows the mode with its double frequency of $f = 58\,\mathrm{kHz}$ at $\rho_{\mathrm{pol}} = 1.02$, as discussed in Section 2.5. The fit model to calculate the aperiodic (noise) and periodic (mode) components of the spectrum is described in [23].

Source: Griener, *et al.*, *Nuclear Materials and Energy* 2023

UC San Diego

What do we study?

Mini-brains in dishes.

Cortical organoids

~ 4mm diameter
~ 100,000 neurons

Source: Trujillo*, Gao*, Nagreas*… Voytek†, Muotri†, *Cell Stem Cell* 2019

UC San Diego

Cortical organoids

Source: Trujillo*, Gao*, Nagreas*… Voytek†, Muotri†, *Cell Stem Cell* 2019

UC San Diego

Brain surgery.

WARNING!

47
42

WARNING!

My path as a Data Scientist

brainSCANr

Source: Voytek & Voytek, *J Neurosci Methods* 2012

UC San Diego

brainSCANr

Source: Voytek & Voytek, *J Neurosci Methods* 2012

UC San Diego

Uber

uber
U
Bradley Voytek, Ph.D.
voytek@uber.com
[redacted]
@bradleyvoytek
http://uber.com

City dynamics

- Aggregating data into arbitrary geographical units: here neighborhoods, but under the hood you'd use e.g., hexagonal tiling.

- Turns lat/lng pairs and time into temporal profiles of demand between parts of a city.

Source: Voytek, *Uber, Inc.*

UC San Diego

Spatiotemporal dynamics

Source: Voytek, *Uber, Inc.*

Spatiotemporal dynamics

- **Looking at dynamics over time allows you to correlate neighborhoods within and between cities.**

- **Can identify "types" of neighborhoods: those with peak weekend and late night demands are more "party"-like whereas M-F peaks are more "business" regions.**

Source: Voytek, *Uber, Inc.*

UC San Diego

Rider patience

- The entire curve moves backward with each year Uber is in a city.
- From this we can infer that people are getting more accustomed to the service and expect more of it over time.

Source: Voytek, *Uber, Inc.*

UC San Diego

Qwake

Disclaimer: I'm an investor

Why are you here?

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:
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?

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

Or is it because...

DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

ntist: The Sexiest
e 21st Century

. Patil

sexiest

Seriously, why Data Science?

Data Science

50 Best Jobs in America

This report ranks jobs according to each job's Glassdoor Job Score, determined by combining three factors: number of job openings, salary, and overall job satisfaction rating.

United States | 2018

0 Shares

1 Data Scientist

4.8 / 5
Job Score

4.2 / 5
Job Satisfaction

$110,000
Median Base Salary

4,524
Job Openings

View Jobs

Source: Glassdoor, 2018

UC San Diego

Data Science

	Job Title	Median Base Salary	Job Satisfaction	Job Openings
#1	Enterprise Architect	$144,997	4.1/5	14,021
#2	Full Stack Engineer	$101,794	4.3/5	11,252
#3	Data Scientist	$120,000	4.1/5	10,071
#4	Devops Engineer	$120,095	4.2/5	8,548
#5	Strategy Manager	$140,000	4.2/5	6,977
#6	Machine Learning Engineer	$130,489	4.3/5	6,801
#7	Data Engineer	$113,960	4.0/5	11,821
#8	Software Engineer	$116,638	3.9/5	64,155
#9	Java Developer	$107,099	4.1/5	10,201
#10	Product Manager	$125,317	4.0/5	17,725

Source: Glassdoor, 2022

Data Science

Rank	Gen Z	Avg Rating	Non Gen Z	Avg Rating
1	Corporate Recruiter	4.79	Corporate Recruiter	4.59
2	Marketing Manager	4.56	Realtor	4.56
3	Social Media Manager	4.46	Technical Account Manager	4.56
4	Data Scientist	4.44	Full Stack Engineer	4.56
5	Product Manager	4.40	SEO Analyst	4.50

Source: Glassdoor, 2022

It’s just a good, fun job.

Okay so this is going to be your life.

What is Data 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), Professor (Medical University of South Carolina)

Console Sources Network Performance Memory Application Security Lighthouse

```
▼<div class="bp9cbjyn j83agx80 cbu4d94t ta1jpn5t t9j0dne7">
  <img data-visualcompletion="media-vc-image" alt="May be an image of one or more people, beard, lake, tree and nature" class="ji
  ug0 r9f5tntg r0294ipz" referrerpolicy="origin-when-cross-origin" src="https://scontent-lax3-1.xx.fbcdn.net/v/t1.6435-9/20659569
  8…lax3-1.xx&oh=00_AT8KPlj9Dd0Ivs0t5_xnDGsUexIr-kH4xymZqLqZtRRR4Q&oe=61F8EE78"> == $0
```

Trouble defining a field

“The only generally agreed upon definition of mathematics is "Mathematics is what mathematicians do." which is followed by "Mathematicians are people who do mathematics." What is true about defining mathematics is also true about many other fields: there is often no clear, sharp definition of the field.”

Source: Richard Hamming, 1968 Turing Award lecture

Trouble defining a field

"[Computer science] is not really about computers—and it's not about computers in the same sense that physics is not really about particle accelerators, and biology is not about microscopes and Petri dishes…"

Source: Hal Abelson (MIT)

Trouble defining a field

"…when some field is just getting started and you don't really understand it very well, *it's very easy to confuse the* ***essence*** *of what you're doing with the tools that you use.*"

Source: Hal Abelson (MIT)

What is the *essence* of what we do?

Source: NASA

Source: *Vox*

Sir Isaac Newton

$$\mathbf{F} = m\mathbf{a}$$

force = mass*acceleration

How is it that observing the world,

How is it that observing the world,

quantifying it, and

How is it that observing the world,

quantifying it, and

fitting equations to those numbers...

Can teach us the
fundamental laws of the universe?

Defining Data Science

By turning many observations into numbers we can fit equations to them, or learn relationships between them, which lets us understand how observations relate and sometimes guess what might happen in the future.

Defining Data Science

By turning many observations into numbers we can fit equations to them, or learn relationships between them, which lets us understand how observations relate and sometimes guess what might happen in the future.

quantification, wrangling, and storage

Defining Data Science

By turning many observations into numbers we can fit equations to them, or learn relationships between them, which lets us understand how observations relate and sometimes guess what might happen in the future.

mathematical modeling and parameterization

Defining Data Science

By turning many observations into numbers we can fit equations to them, or learn relationships between them, which lets us understand how observations relate and sometimes guess what might happen in the future.

machine learning and artificial intelligence

Defining Data Science

By turning many observations into numbers we can fit equations to them, or learn relationships between them, which lets us understand how observations relate and sometimes guess what might happen in the future.

inference

Defining Data Science

By turning many observations into numbers we can fit equations to them, or learn relationships between them, which lets us understand how observations relate and sometimes guess what might happen in the future.

prediction