COGS 9 Course Announcements

- I have no control over the waitlist.
  - cogsadvising@ucsd.edu OR their drop-in hours are your best bet
- I teach another course right after this.
  - I will be here early for questions before lecture.
  - If you have a non-waitlist-related question that you wanted to ask after class, feel free to come on up now!

Welcome to COGS 9!

Introduction to Data Science

Shannon E. Ellis, Ph.D
UC San Diego

Department of Cognitive Science
sellis@ucsd.edu

hello
my name is
Professor Ellis

Instructional Staff:
Discussion Section, Piazza, & Office Hours

TAs

Kunal

Matt

Shubham

IAs

Connor

Mincong

Yupei

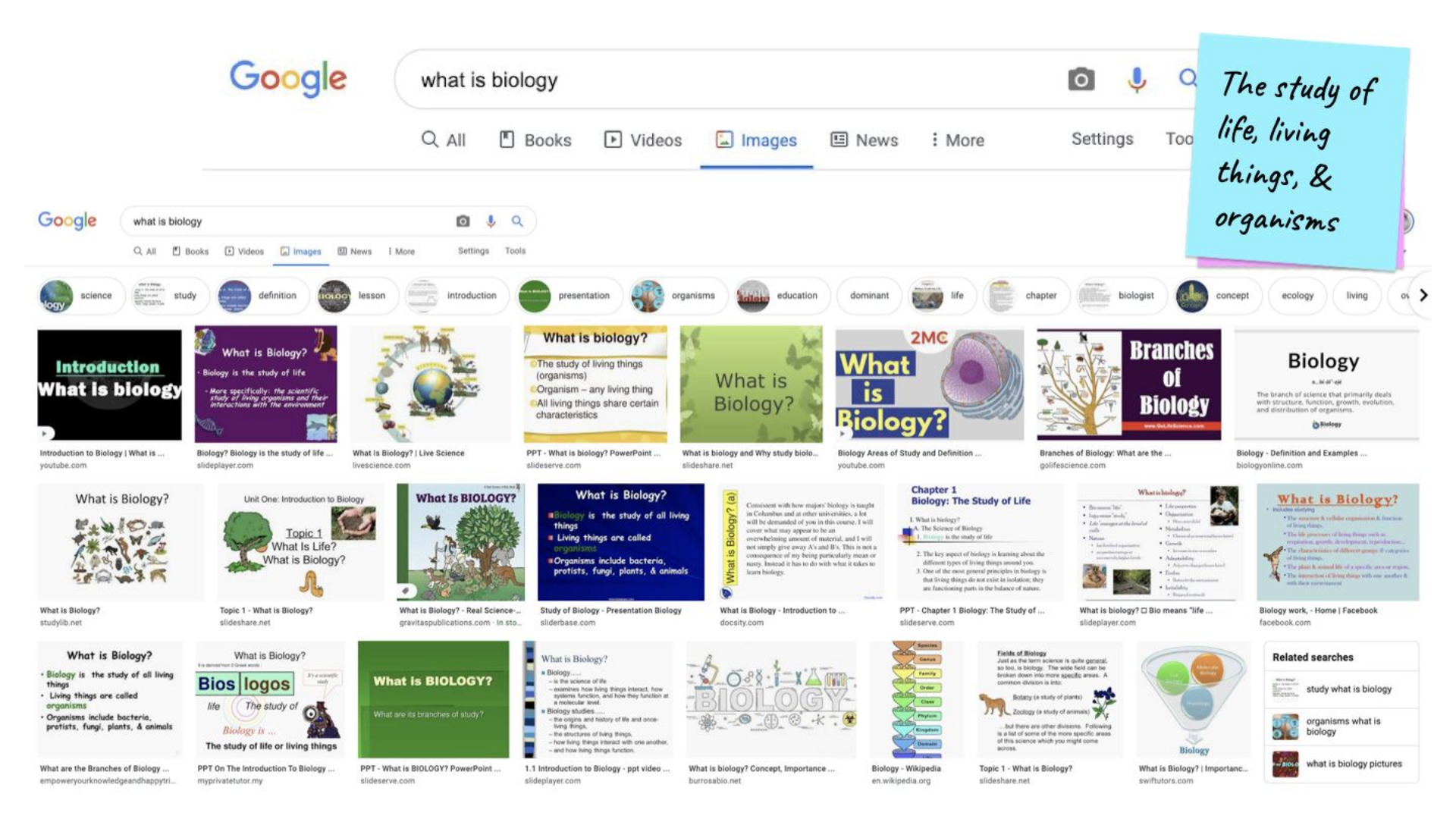
Today’s Learning Objectives:

Define data science and figure out how this class is going to work.

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?

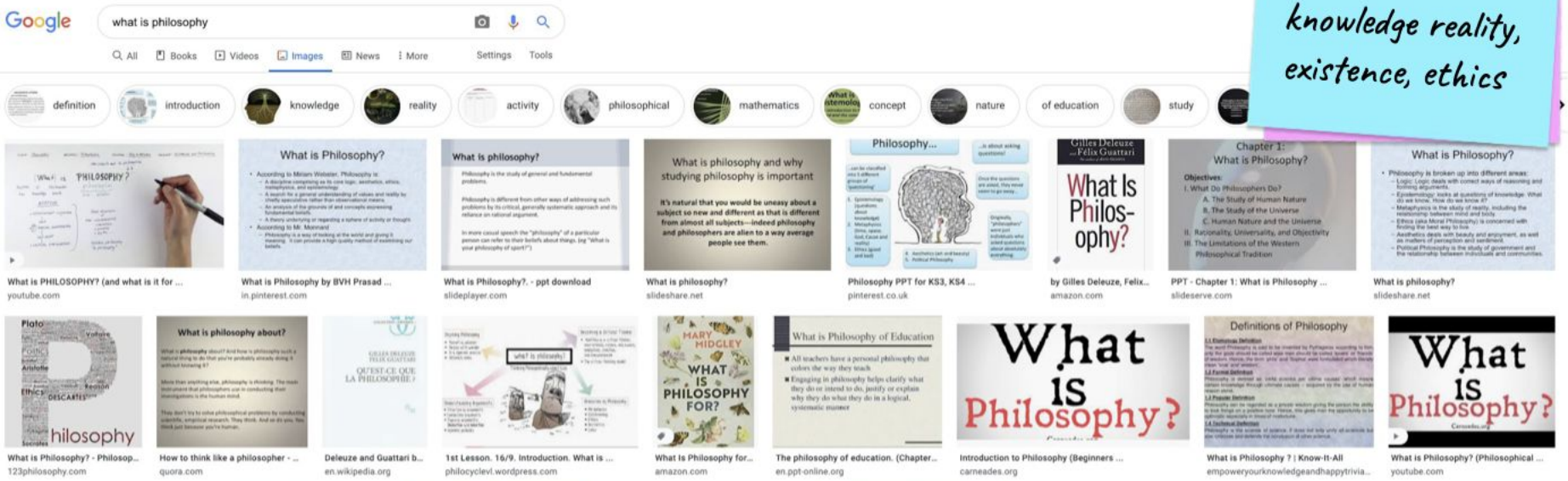

Google
what is biology
All
Books
Videos
Images
News
More
Settings
Too
The study of life, living things, & organisms
Google
what is biology
All
Books
Videos
Images
News
More
Settings
Tools
science
study
definition
lesson
introduction
presentation
organisms
education
dominant
life
chapter
biologist
concept
ecology
living
Introduction
What is biology
Introduction to Biology | What is ...
youtube.com
What is Biology?
Biology is the study of life
More specifically: the scientific study of living organisms and their interactions with the environment
Biology? Biology is the study of life ...
slideplayer.com
What Is Biology? | Live Science
livescience.com
What is biology?
The study of living things (organisms)
Organism – any living thing
All living things share certain characteristics
PPT - What is biology? PowerPoint ...
slideserve.com
What is Biology?
What is biology and Why study biolo...
slideshare.net
2MC
What is Biology?
Biology Areas of Study and Definition ...
youtube.com
Branches of Biology
Branches of Biology: What are the ...
golifescience.com
Biology
The branch of science that primarily deals with structure, function, growth, evolution, and distribution of organisms.
Biology - Definition and Examples ...
biologyonline.com
What is Biology?
What is Biology?
studylib.net
Unit One: Introduction to Biology
Topic 1
What Is Life?
What is Biology?
Topic 1 - What is Biology?
slideshare.net
What Is BIOLOGY?
What is Biology? - Real Science-...
gravitaspublications.com · In sto...
What is Biology?
Biology is the study of all living things
Living things are called organisms
Organisms include bacteria, protists, fungi, plants, & animals
Study of Biology - Presentation Biology
sliderbase.com
What is Biology? (a)
Consistent with how majors' biology is taught in Columbus and at other universities, a lot will be demanded of you in this course. I will cover what may appear to be an overwhelming amount of material, and I will not simply give away A's and B's. This is not a consequence of my being particularly mean or nasty. Instead it has to do with what it takes to learn biology.
What is Biology - Introduction to ...
docsity.com
Chapter 1
Biology: The Study of Life
1. What is biology?
A. The Science of Biology
1. Biology is the study of life
2. The key aspect of biology is learning about the different types of living things around you.
3. One of the most general principles in biology is that living things do not exist in isolation; they are functioning parts in the balance of nature.
PPT - Chapter 1 Biology: The Study of ...
slideserve.com
What is biology?
What is Biology?
What is biology? ☐ Bio means "life ...
slideplayer.com
What is Biology?
Biology work, - Home | Facebook
facebook.com
What is Biology?
Biology is the study of all living things
Living things are called organisms
Organisms include bacteria, protists, fungi, plants, & animals
What are the Branches of Biology ...
empoweryourknowledgeandhappytri...
What is Biology?
Bios logos
life
The study of
Biology is ...
The study of life or living things
PPT On The Introduction To Biology ...
myprivatetutor.my
What is BIOLOGY?
What are its branches of study?
PPT - What is BIOLOGY? PowerPoint ...
slideserve.com
What is Biology?
Biology.....
is the science of life
examines how living things interact, how systems function, and how they function at a molecular level.
Biology studies.....
the origins and history of life and once-living things,
the structures of living things,
how living things interact with one another, and how living things function.
1.1 Introduction to Biology - ppt video ...
slideplayer.com
BIOLOGY
What is biology? Concept, Importance ...
burrosabio.net
Species
Genus
Family
Order
Class
Phylum
Kingdom
Domain
Biology - Wikipedia
en.wikipedia.org
Fields of Biology
Just as the term science is quite general, so too, is biology. The wide field can be broken down into more specific areas. A common division is into:
Botany (a study of plants)
Zoology (a study of animals)
...but there are other divisions. Following is a list of some of the more specific areas of this science which you might come across.
Topic 1 - What is Biology?
slideshare.net
Biology
What is Biology? | Importanc...
swiftutors.com
Related searches
study what is biology
organisms what is biology
what is biology pictures

Google

what is philosophy

All | Books | Videos | Images | News | More | Settings

Study of the fundamental nature of knowledge reality, existence, ethics

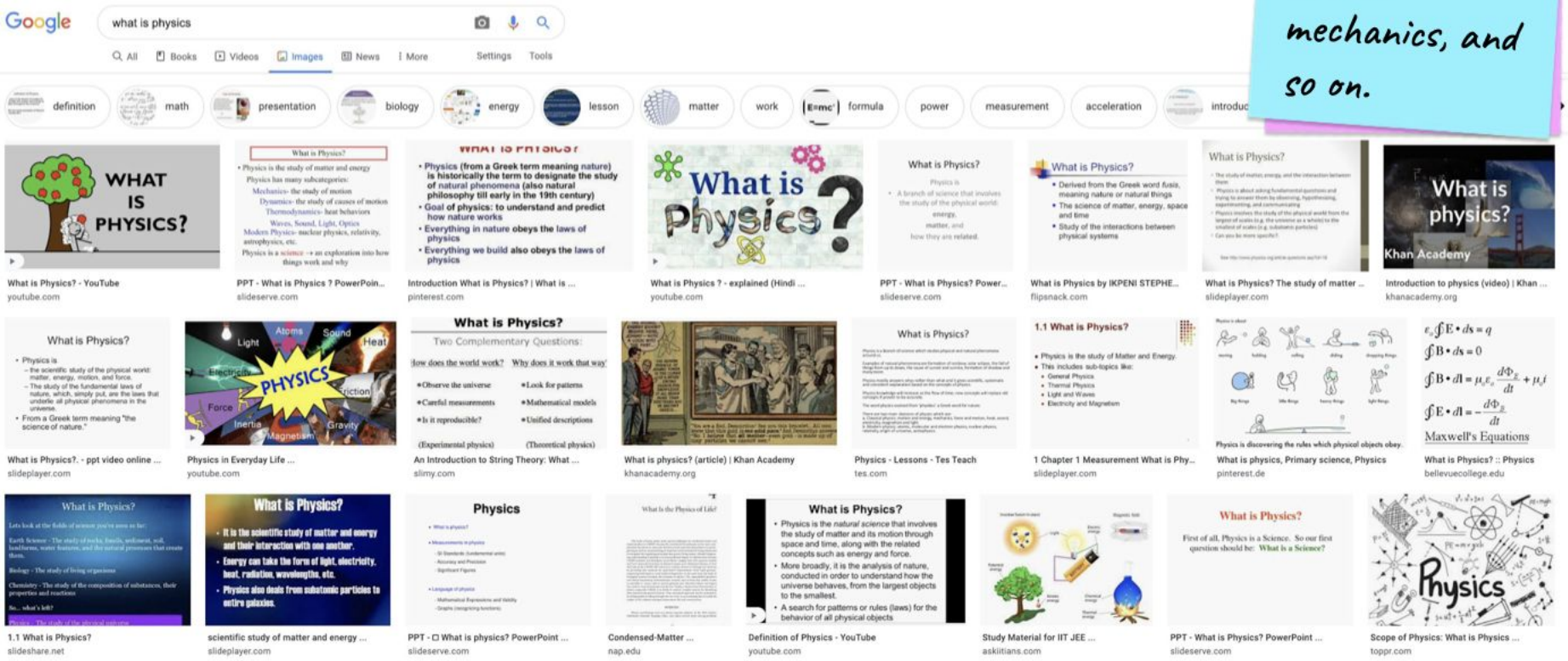


Google

what is physics

All | Books | Videos | Images | News | More | Settings

Study of matter, energy, atoms, mechanics, and so on.

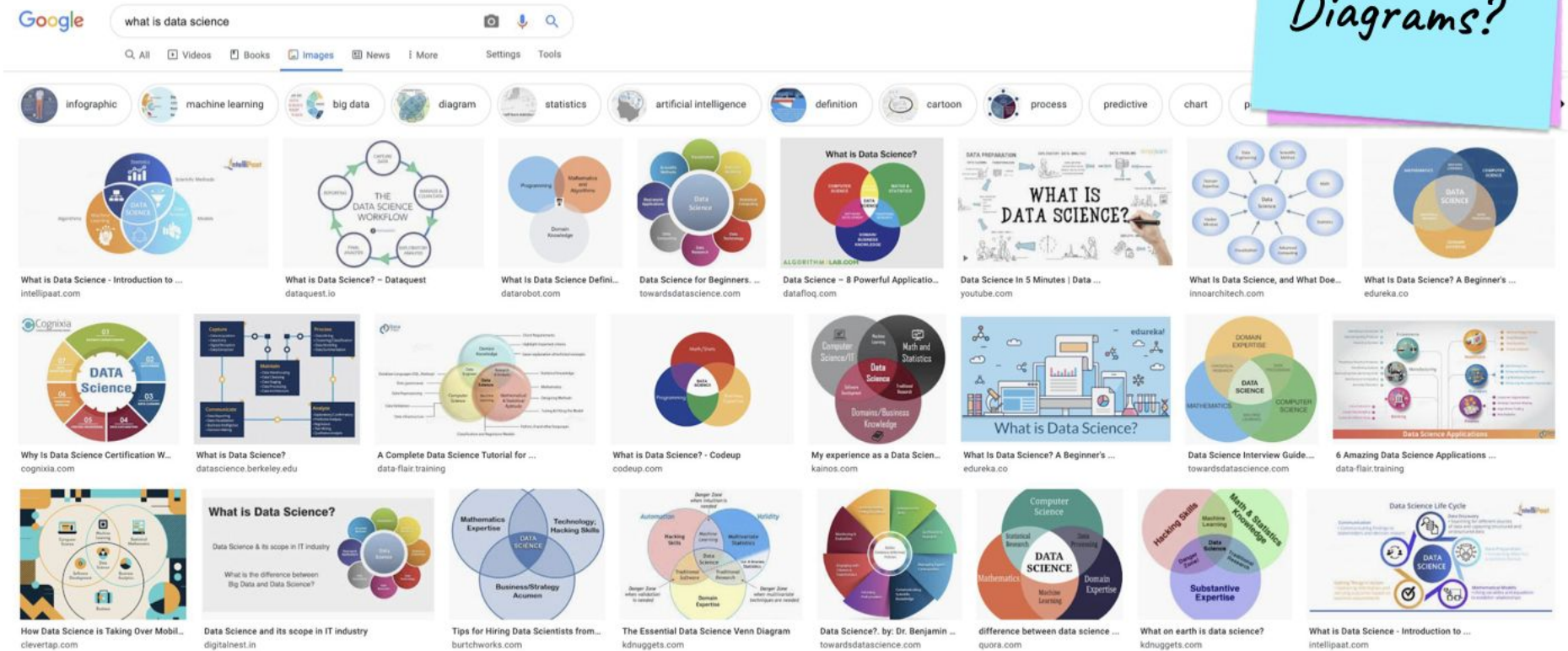


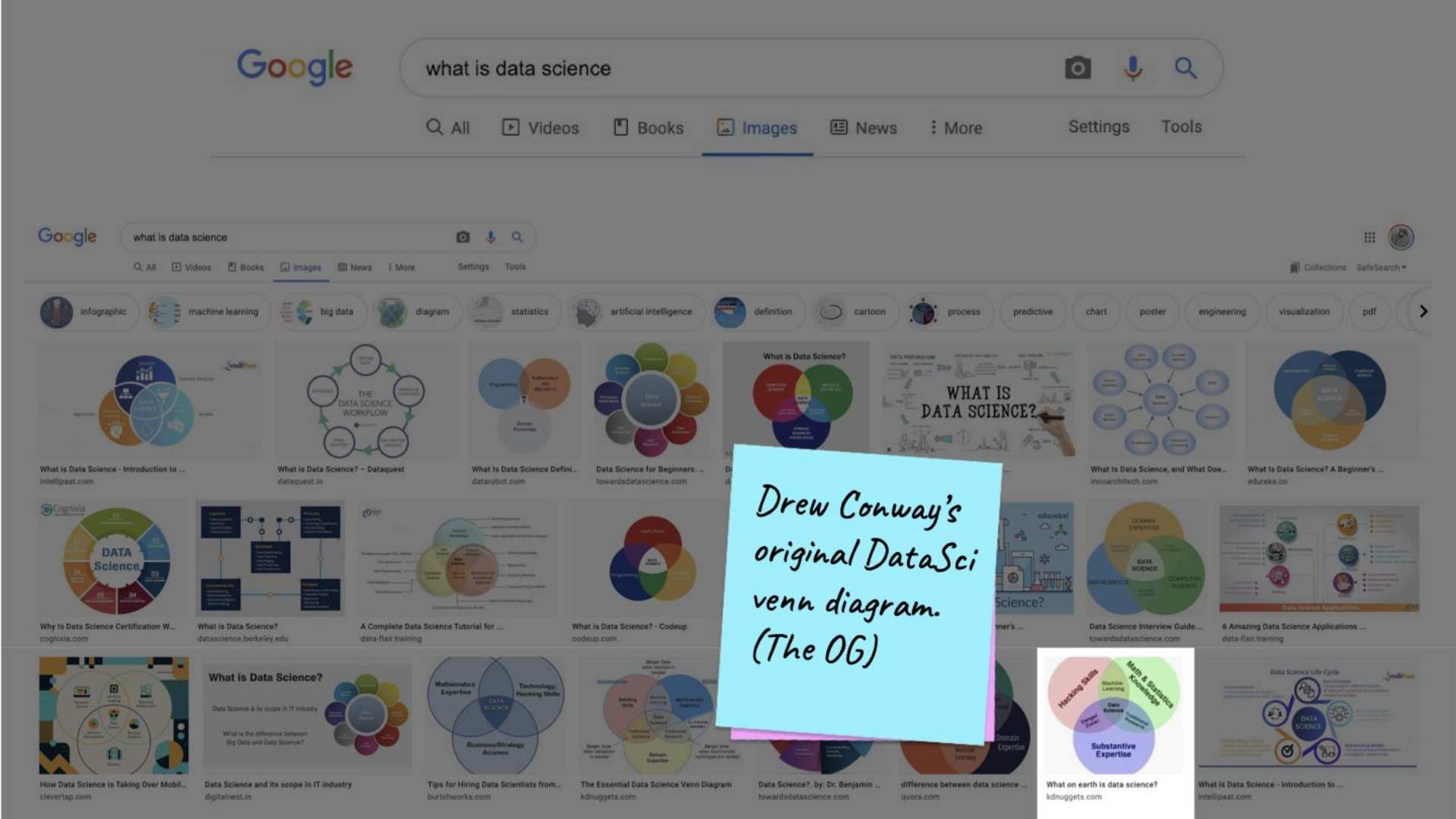
Google

what is data science

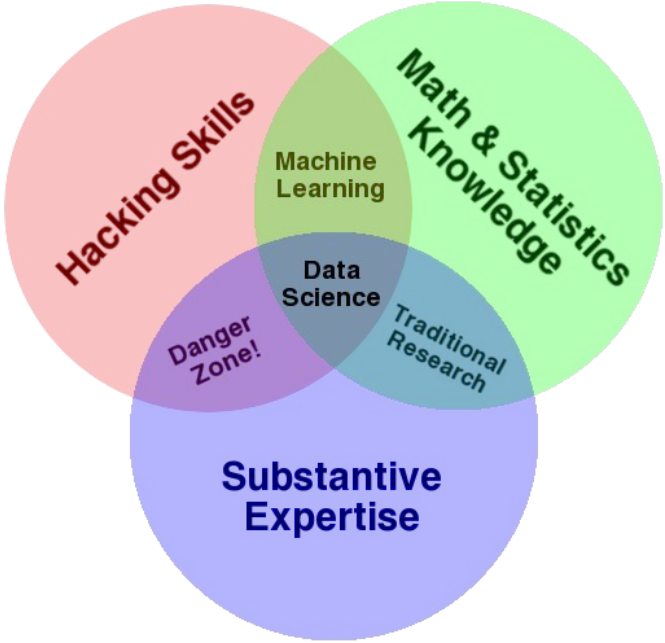
All | Videos | Books | Images | News | More | Settings

The study of Venn Diagrams?




Google
what is data science
All
Videos
Books
Images
News
More
Settings
Tools
infographic
machine learning
big data
diagram
statistics
artificial intelligence
definition
cartoon
process
predictive
chart
poster
engineering
visualization
pdf
What is Data Science - Introduction to ...
intellipaat.com
What is Data Science? – Dataquest
dataquest.io
What Is Data Science Defini...
datarobot.com
Data Science for Beginners. ...
towardsdatascience.com
What is Data Science?
WHAT IS DATA SCIENCE?
What Is Data Science, and What Doe...
innoarchitech.com
What Is Data Science? A Beginner's ...
edureka.co
Why Is Data Science Certification W...
cognixia.com
What is Data Science?
datascience.berkeley.edu
A Complete Data Science Tutorial for ...
data-flair.training
What is Data Science? - Codeup
codeup.com
Data Science Interview Guide....
towardsdatascience.com
6 Amazing Data Science Applications ...
data-flair.training
How Data Science is Taking Over Mobil...
clevertap.com
What is Data Science?
Data Science Application
Data Science and its scope in IT industry
digitalnest.in
Mathematics Expertise
Technology: Hacking Skills
Business/Strategy Acumen
Tips for Hiring Data Scientists from...
burtchworks.com
The Essential Data Science Venn Diagram
kdnuggets.com
Data Science?. by: Dr. Benjamin ...
towardsdatascience.com
difference between data science ...
quora.com
Hacking Skills
Math & Statistics Knowledge
Substantive Expertise
What on earth is data science?
kdnuggets.com
Data Science Life Cycle
What is Data Science - Introduction to ...
intellipaat.com
Drew Conway's original DataSci venn diagram. (The OG)

What is data science?



Copyright © 2014 by Steven Geringer Raleigh, NC.
Permission is granted to use, distribute, or modify this image, provided that this copyright notice remains intact.

http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

Josh Wills
@josh_wills

Follow

Data Scientist (n.): Person who is better at statistics than any software engineer and better at software engineering than any statistician.

RETWEETS 1,046 FAVORITES 532

6:55 PM - 3 May 2012

He is the head data engineer @ slack, former Cloudera & Google DataSci lead

….so….uh…what is data science?

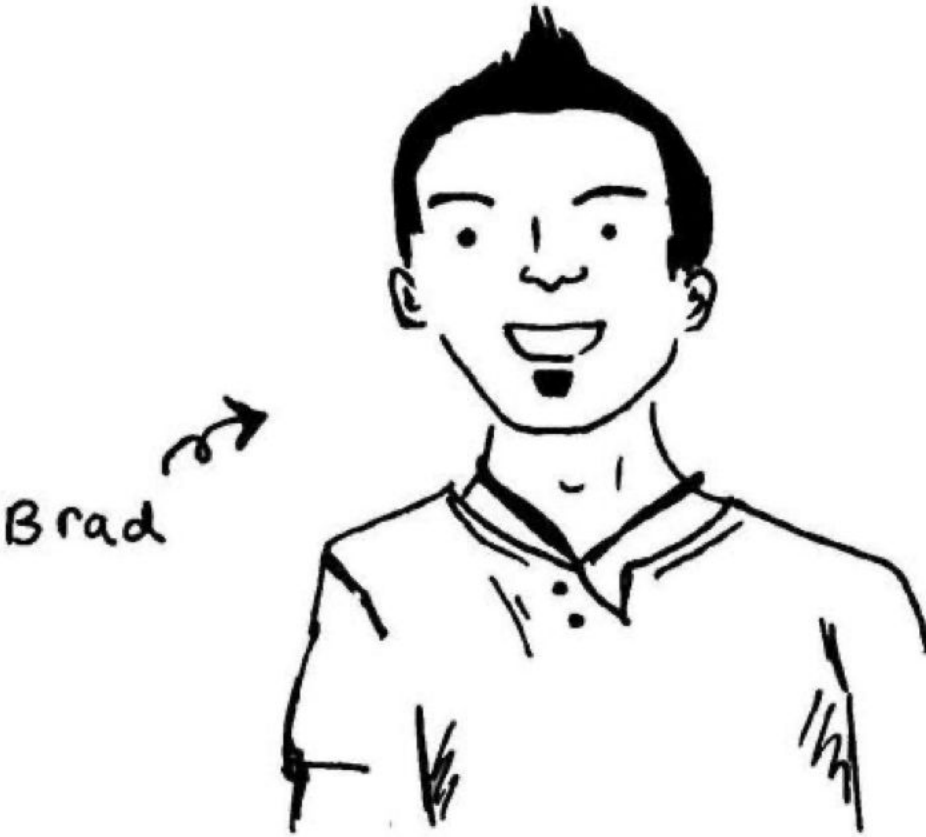
David Donoho

This coupling of scientific discovery and practice involves the collection, management, processing, analysis, visualization, and interpretation of vast amounts of heterogeneous data associated with a diverse array of scientific, translational, and interdisciplinary actions.

slide from Kyle Shannon

Bradley Voytek

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



slide from Kyle Shannon

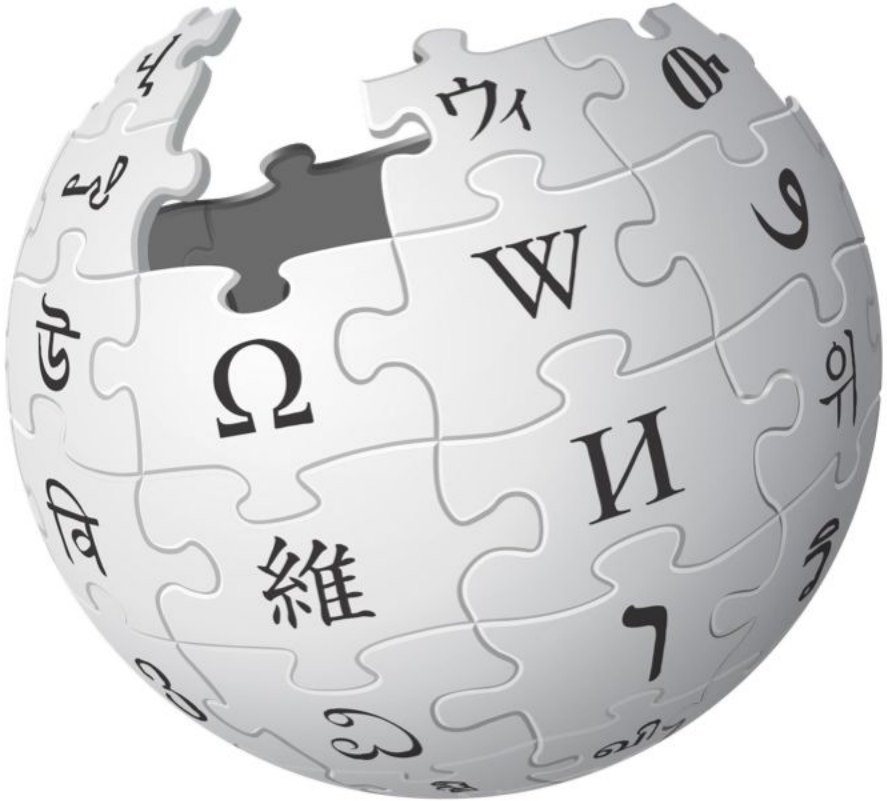
Rafael Irizarry

An umbrella term used by organizations to describe the processes used to extract value from data.

slide from Kyle Shannon

Wikipedia!

A concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with data.[3] It employs techniques and theories drawn from many fields within the context of mathematics, statistics, information science, and computer science.



slide from Kyle Shannon

Cogs 9 Definition

The scientific and ethical process of extracting value from data.

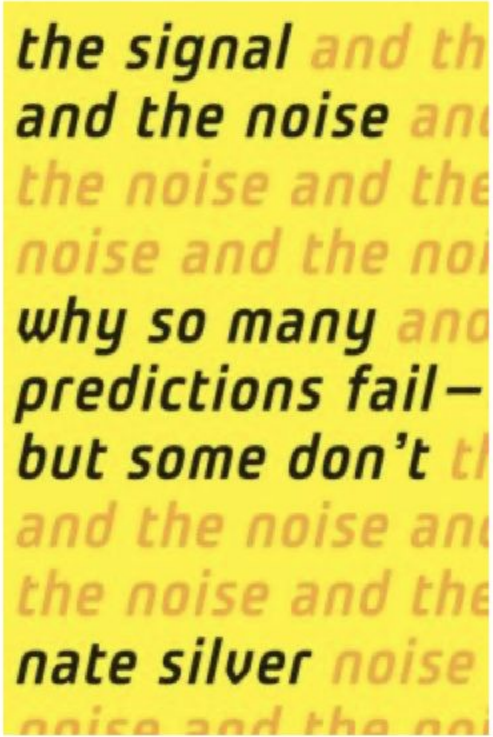
slide from Kyle Shannon

How does value get extracted from data?

slide from Kyle Shannon

A Philosophical Answer!

> Numbers have no way of speaking for themselves. We speak for them. We imbue them with meaning…. Before we demand more of our data, we need to demand more of ourselves. - Nate Silver

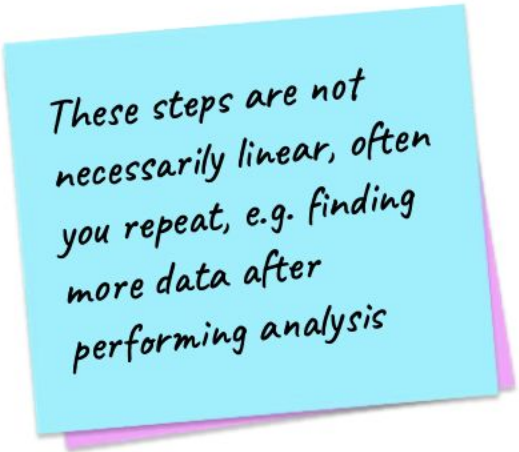


slide from Kyle Shannon

A Practical Answer!

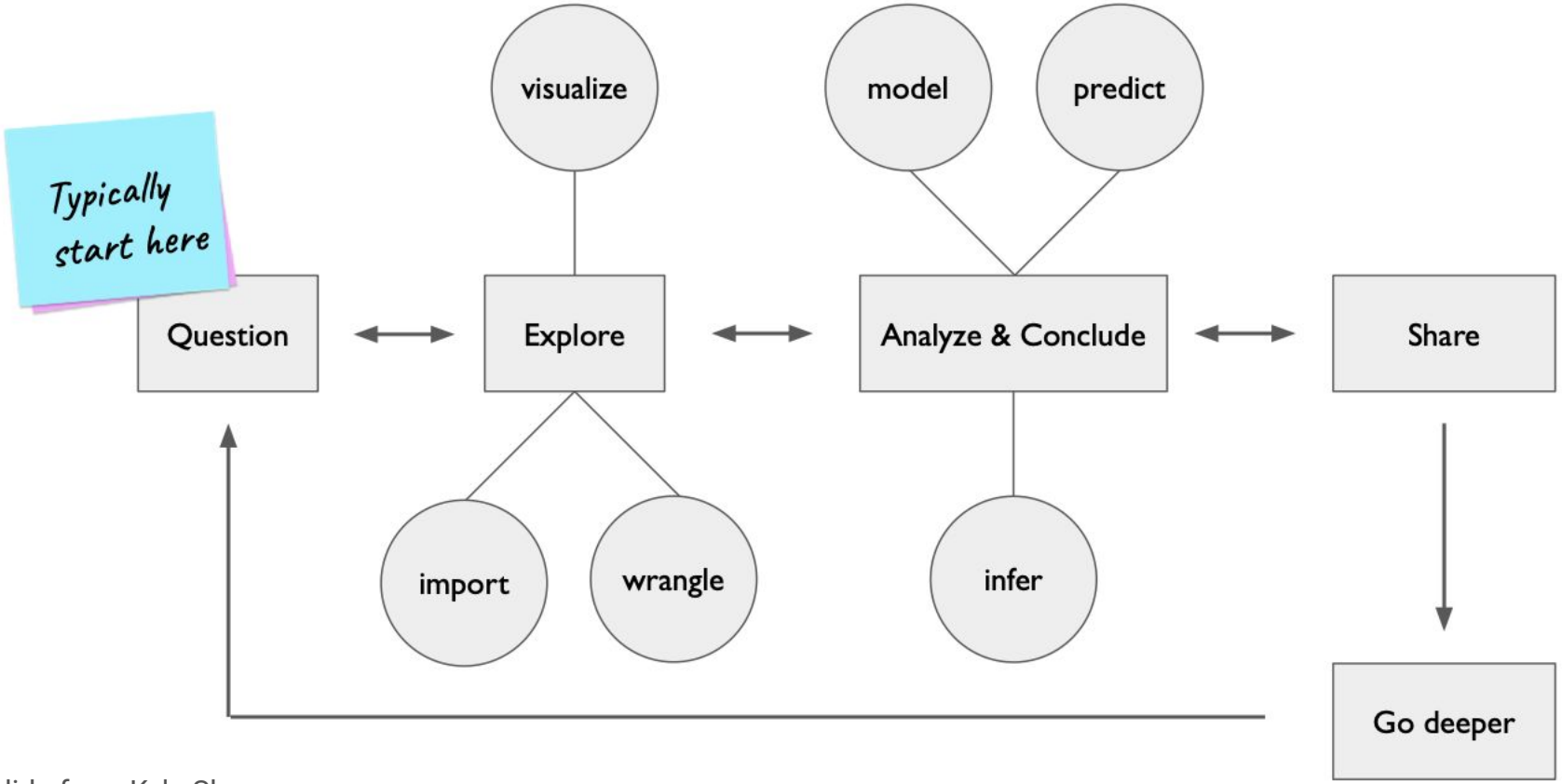
The General Data Science Process

- Craft your question and hypothesis while considering ethics
- Find, clean, and wrangle relevant data
- Explore and visualize your data
- Perform some analysis or prediction
- Generate reports and visuals about these findings
- Share these results with others



slide from Kyle Shannon

The General Data Science Process



slide from Kyle Shannon

Data scientists ask interesting questions & answer them with data

A project that…

May make you better at dating?

The Best Questions For A First Date

How asking certain questions can reveal much more

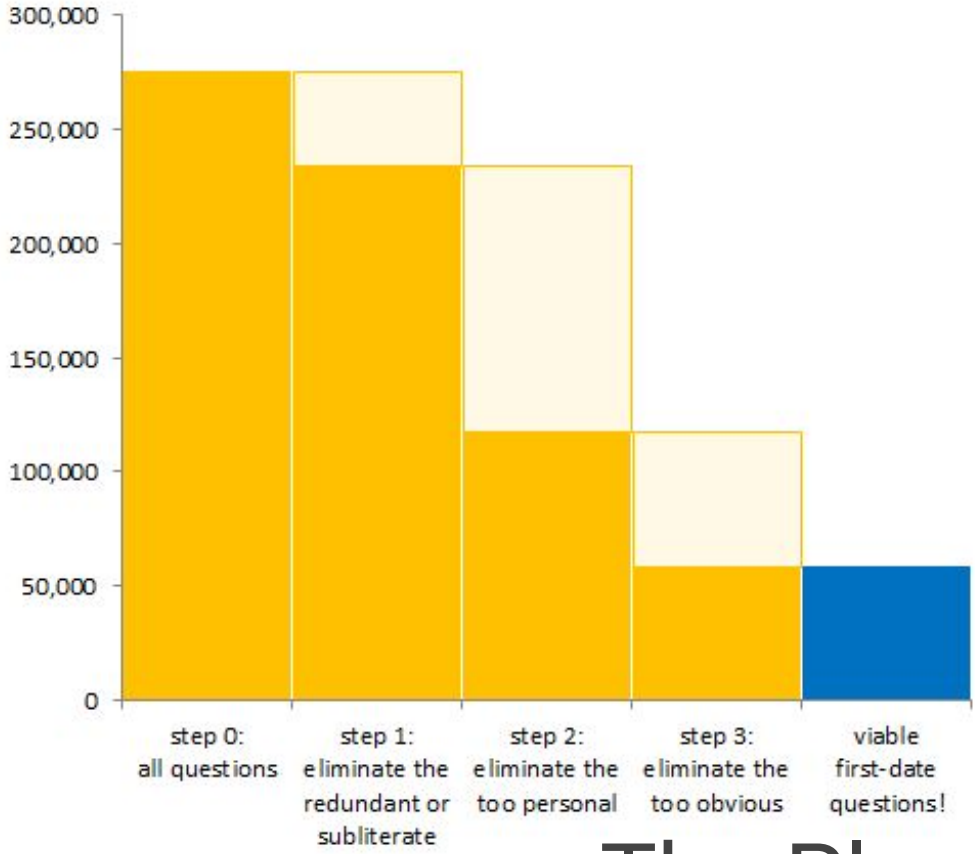
OkCupid Follow

Apr 19, 2011 · 5 min read

This post is our attempt to end the mystery. We took OkCupid's database of 275,294 match questions—probably the biggest collection of relationship concerns on earth—and the 776 *million* answers people have given us, and we asked:

> *What questions are **easy to bring up**, yet correlate to the deeper, unspeakable, issues people actually care about?*

https://theblog.okcupid.com/the-best-questions-for-a-first-date-dba6adaa9df2



The Plan

https://theblog.okcupid.com/the-best-questions-for-a-first-date-dba6adaa9df2

If you want to know…

Do my date and I have long-term potential?

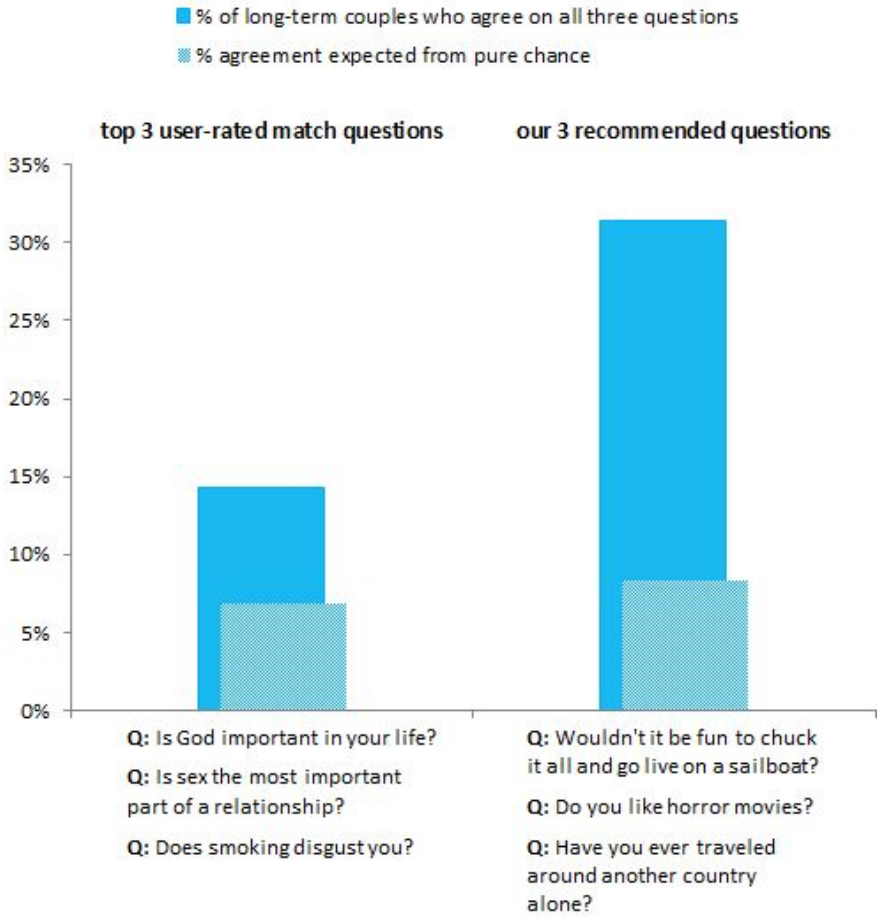
Ask your date (and yourself!)…

Do you like horror movies?

Have you ever traveled around another country alone?

Wouldn't it be fun to chuck it all and go live on a sailboat?

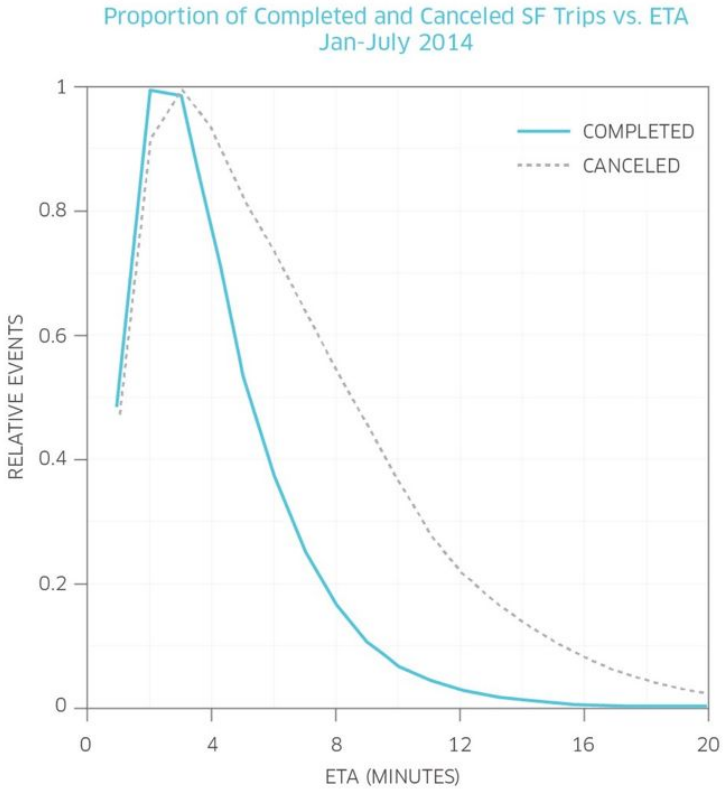
https://theblog.okcupid.com/the-best-questions-for-a-first-date-dba6adaa9df2



https://theblog.okcupid.com/the-best-questions-for-a-first-date-dba6adaa9df2

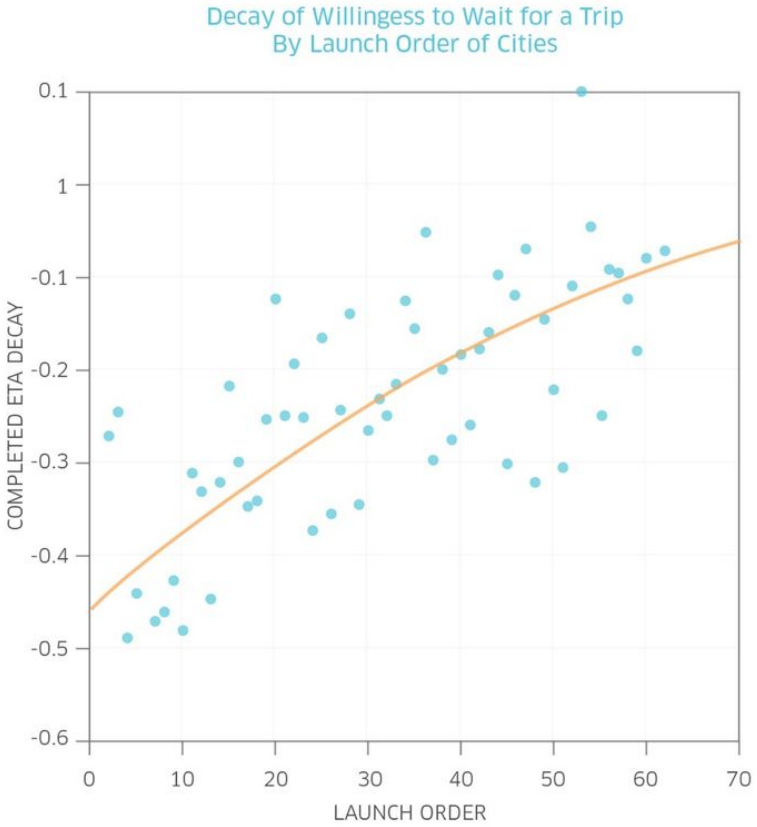
A gig-economy example:
How do I improve ride sharing?

Rider patience



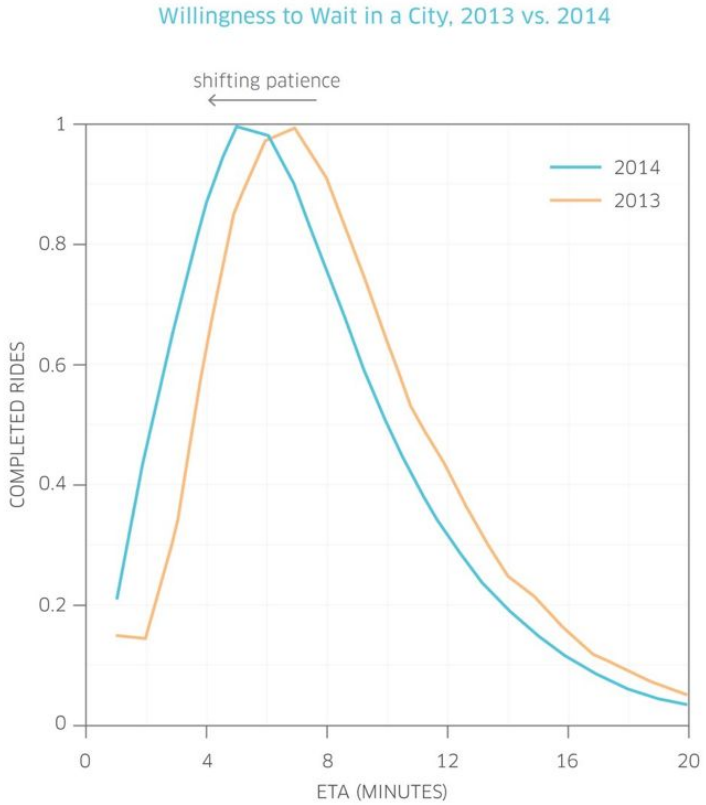
Source: Uber, Inc. (Voytek)

Rider patience



Source: Uber, Inc. (Voytek)

Rider patience



Source: Uber, Inc. (Voytek)

Where industries do data scientists work in?

DR PEPPER SNAPPLE GROUP

slide from Kyle Shannon

Elvert Barnes, Flickr

slide from Kyle Shannon

slide from Kyle Shannon

slide from Kyle Shannon

slide from Kyle Shannon

A/B Testing

Data Science team at AirBnB uses A/B testing by exposing users of their website to various recommendation and ranking algorithms to test the effectiveness of their algorithms.

Image Recognition and Analysis

AirBnB does analysis on photos to find out which ones work the best for their users, what features in the photos make them most looked-for and what kind of photos on the website get more number of clicks.

1. recommendations
2. marketing
3. Price setting
4. Learning about users
5. Image analysis
6. Geographic reach

Natural Language Processing

AirBnB uses natural language processing technology that analyses the review boards or the messages boards through sentiment analysis. This helps AirBnB understand about the true feeling behind the reviews.

Predictive Modelling

Using predictive modelling, AirBnB can do market specific forecast with multiple variables.

Regression Analysis

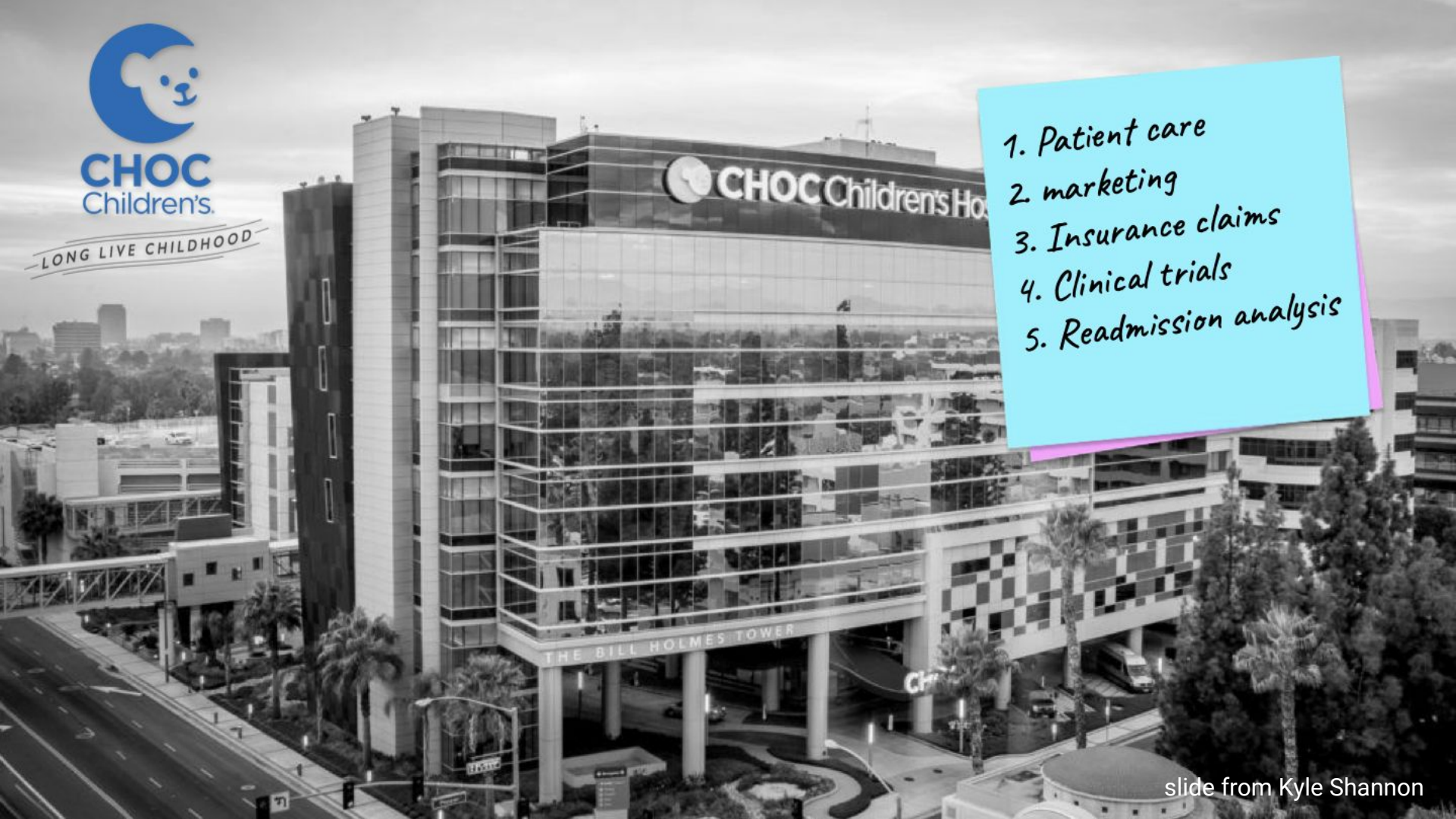
AirBnB uses regression analysis technique to find out which features of a particular listing have major impact on the bookings made.

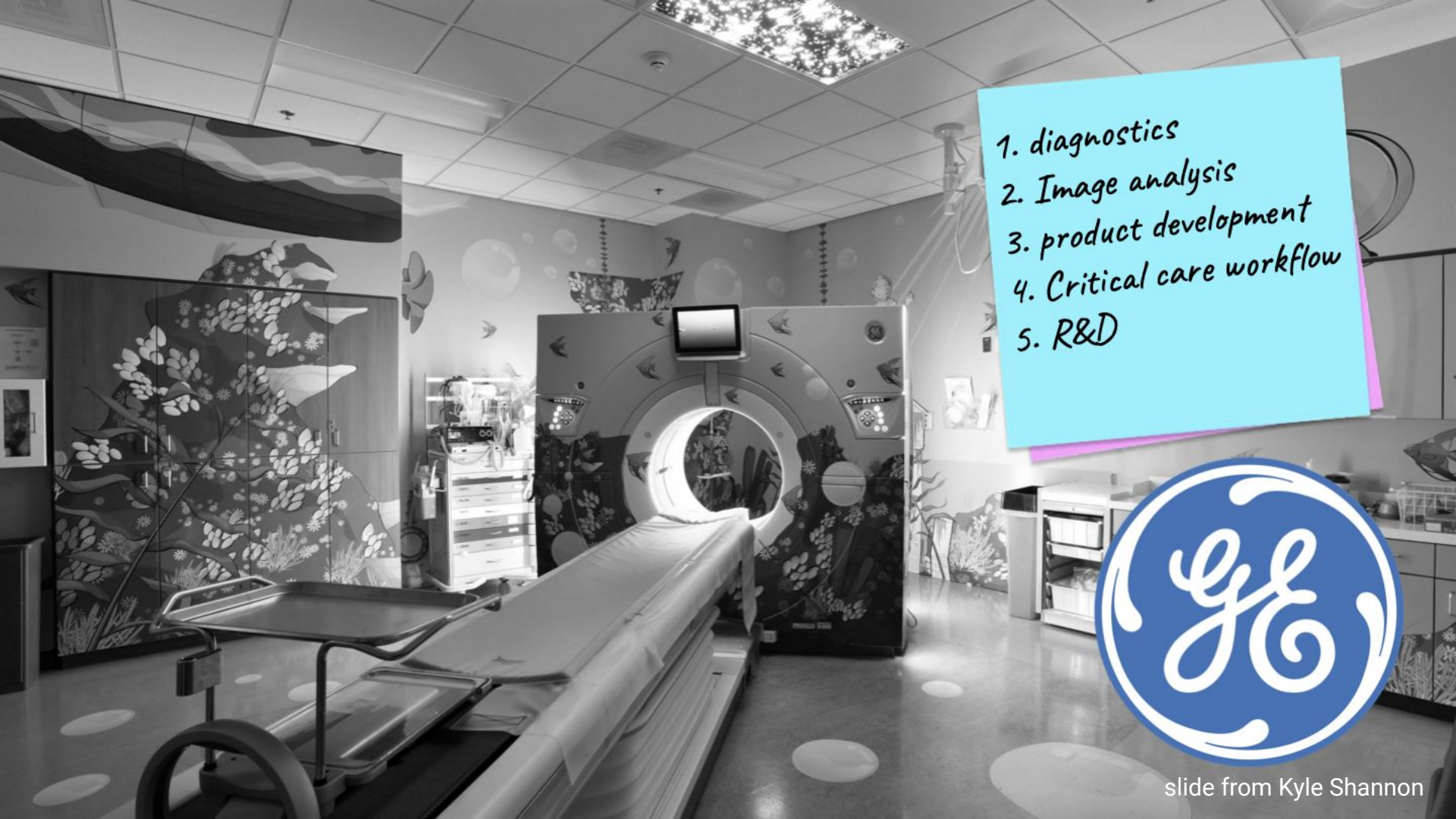
Collaborative Filtering

AirBnB data science team uses collaborative filtering technique to model host preferences.

https://www.dezyre.com/article/how-data-science-increased-airbnbs-valuation-to-25-5-bn/199

slide from Kyle Shannon


CHOC
Children's
LONG LIVE CHILDHOOD
CHOC Children's Ho
THE BILL HOLMES TOWER
1. Patient care
2. marketing
3. Insurance claims
4. Clinical trials
5. Readmission analysis
slide from Kyle Shannon



slide from Kyle Shannon

That’s awesome.
Now, what will we do
in *this* course?

Scheduling

Lecture: TuTh 9:30-10:50 AM

Discussion Sections

	Date & Time	Section Location	Who
A01	W 9A	CENTER 122	Kunal & Connor
A02	W 1P	DIB 122	Kunal & Yupei
A03	W 3P	CSB 004	Matt & Shubham
A04	W 5P	MANDEB-150	Matt & Mincong
A05	W 6P	MANDEB-150	Shubham & Mincong

* Discussion Section starts week 2

What does it mean that this is a “**survey**” course?

Covered

- Concepts & light theory
- Examples & Stories
- Communication
- Ethics & questions
- Technical approaches

Not covered

- Rigorous mathematical theory
- Programming
- In-depth Statistics/Probability
- In-depth Technical Skills

Course Objectives

- Comprehend **core data science concepts** and examine their **applications**
- Discuss **data privacy** and **ethical** concerns with real-world examples
- Identify **data science questions** and the appropriate analytic approach to answering those questions
- **Communicate** data-related topics and projects
- Demonstrate how to **think critically** about data, and how to **approach problems** with a “data-first” mindset
- Describe potential **pitfalls of data analyses**, how to identify them, and how to avoid them

Topics introduced

Data Science & Ethics

Version Control & Programming

Data & Data Wrangling

Data Visualization & Communication

Approaches to Data Analysis

COGS 9: General Plan

Week	Topic(s)
1	Data Science, Ethics & Reproducibility
2	Questions, Reproducibility & Bias
3	Data
4	Data Wrangling & Programming
5	Data Visualization
6	Communication
7	Inference
8	Machine Learning
9	Text, Longitudinal, and Geospatial Analysis
10	Future of Data Science

Course links

Course Website	https://cogs9.github.io/ucsd-cogs9/	Course materials, syllabus, etc.
Piazza	https://piazza.com/class/lfvh3bqe743ix	Questions & discussion
Canvas	https://canvas.ucsd.edu/courses/44979	Grades
Gradescope	https://www.gradescope.com/courses/527643	Assignment, reading quizzes, & project submission
Course Podcast	https://podcast.ucsd.edu/	listening to me speak at a rate even faster than normal
Anonymous Feedback	Submit via Google Form	if I ever offend you, use an example you hate, or to provide general feedback

Attendance is *not* required, but is incentivized

ClassQuestion will be used in class **lecture** Q&A

- Starting Thurs, if you attend 75% of all lectures, your lowest assignment score will be replaced with full credit
- To get credit for lecture attendance, you must respond to 50% of that days' questions

Why in-class questions?

- There are a whole lot of you! Checks understanding & Provides me with feedback
- Aids in critical thinking & allows for application of concepts
- Give you all a break from listening to me (we humans need this!)
- Helps you prep for assignments/exams
- Take attendance without wasting your time

You'll need to use your UCSD email when logging in/registering.

Attendance is *not* required, but is incentivized

ClassQuestion will be used in class **lecture** Q&A

- Starting Thurs, if you attend 75% of all lectures, your lowest assignment score will be replaced with full credit
- To get credit for lecture attendance, you must respond to 50% of that days' questions

Discussion Section

- Week 2: group project formation
- Reading quiz weeks: showing up/participating gives you full credit on that week's reading quiz

Discussion Section Switching

You *should* be signed up for a section you can regularly attend. Try to attend that section.

If you are signed up for a section you *cannot* attend, you may attend a different section. *However*, if one section becomes too crowded, we'll revisit this policy.

General grading:

3 Assignments	30	60 (20 each)
5 Reading Quizzes (lowest quiz score dropped)	20	40 (10 each)
Final Project pt. 1	10	20
Final Project pt. 2	20	40
1 Final Exam	20	40
Extra Credit	N/A	5 (bonus)

(3) Assignments

Each assignment will be released and submitted on Gradescope.

Assignments will be due on Fridays 11:59pm (see calendar)

Final Project

The final project is a report on how you would handle a *complicated data science project* from start to finish.

You will be given a Google Doc skeleton for your final project. You will submit these as PDFs on Gradescope. (We'll discuss details in coming lectures!)

(5) Readings

- Take reading quiz before Friday at 11:59 PM (see calendar).
  a. 60 min time limit
  b. 10 questions

- Readings will focus on:
  1. Data Science
  2. Ethics
  3. Tidy Data
  4. Data Visualization
  5. Algorithms & Bias

(1) Final Exam

Cumulative; Primarily Multiple-Choice; Take-home; Due on the date of the final exam

Course Help

If something in lecture or a reading is unclear:

- *ask in class*
- *ask during section*
- *post on Piazza*
- *ask a classmate*
- *come to office hours!*

Why come to office hours?

- Struggling with material
- Discuss readings/lectures
- Ask questions about work/grad school/jobs
- Sit and listen to others
- Talk about cool DS stuff

CLASS CONDUCT

Be respectful. **Take care of one another**. Remember **we're all human**.

This classroom is **welcoming to all**.
(regardless of gender, gender identity and expression, age, sexual orientation, disability, physical appearance, body size, race, ethnicity, religion (or lack thereof), political beliefs/leanings, or technology choices)

If you have a concern, please speak with Prof Ellis, your TAs, or IAs. If you are uncomfortable doing so, the OPHD and/or CARE are a few resources on campus.

A message for first-gen students, transfer students, and those who don't have older siblings/friends who have attended UCSD/university

If you are struggling, come to office hours. Ask questions on Piazza. Reach out to me to ask for better approaches. Your classmates ARE doing this. And, you're not alone.

If you need a bit longer on something b/c you fell sick, a family thing came up, work called you in for an extra shift, etc., ask for an extension. Your classmates ARE doing this.

What’s your goal?

Why this course?

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

data scientist

Why this course?

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

software engineer

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:

natural scientist or physician

(i.e. neuroscientist, biologist, chemist, etc.)

Why this course?

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

statistician or biostatistician

Why this course?

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

social scientist

(ie.e linguist, psychologist, etc.)

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?

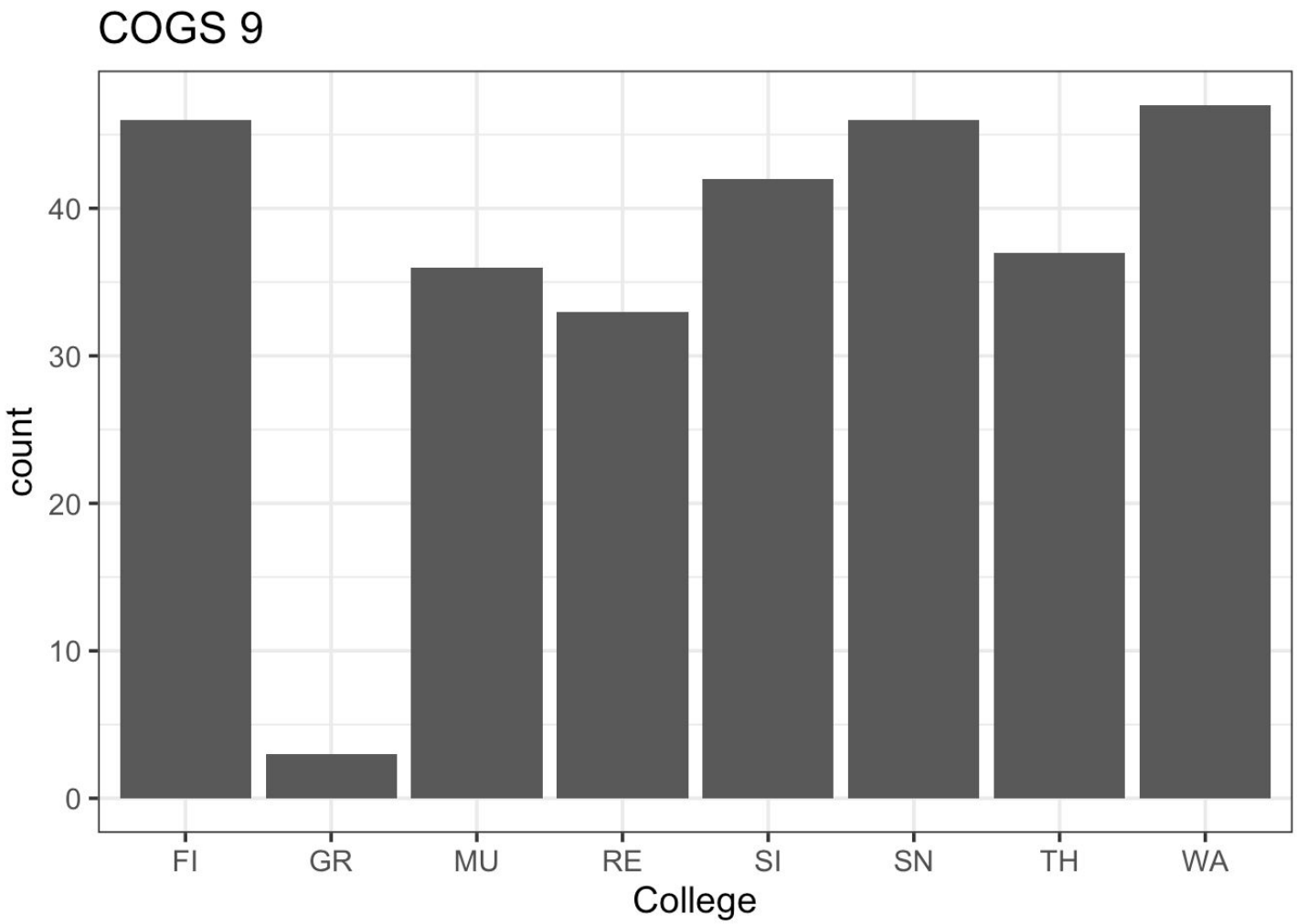
You are going to be analyzing lots of data because you're studying to be:

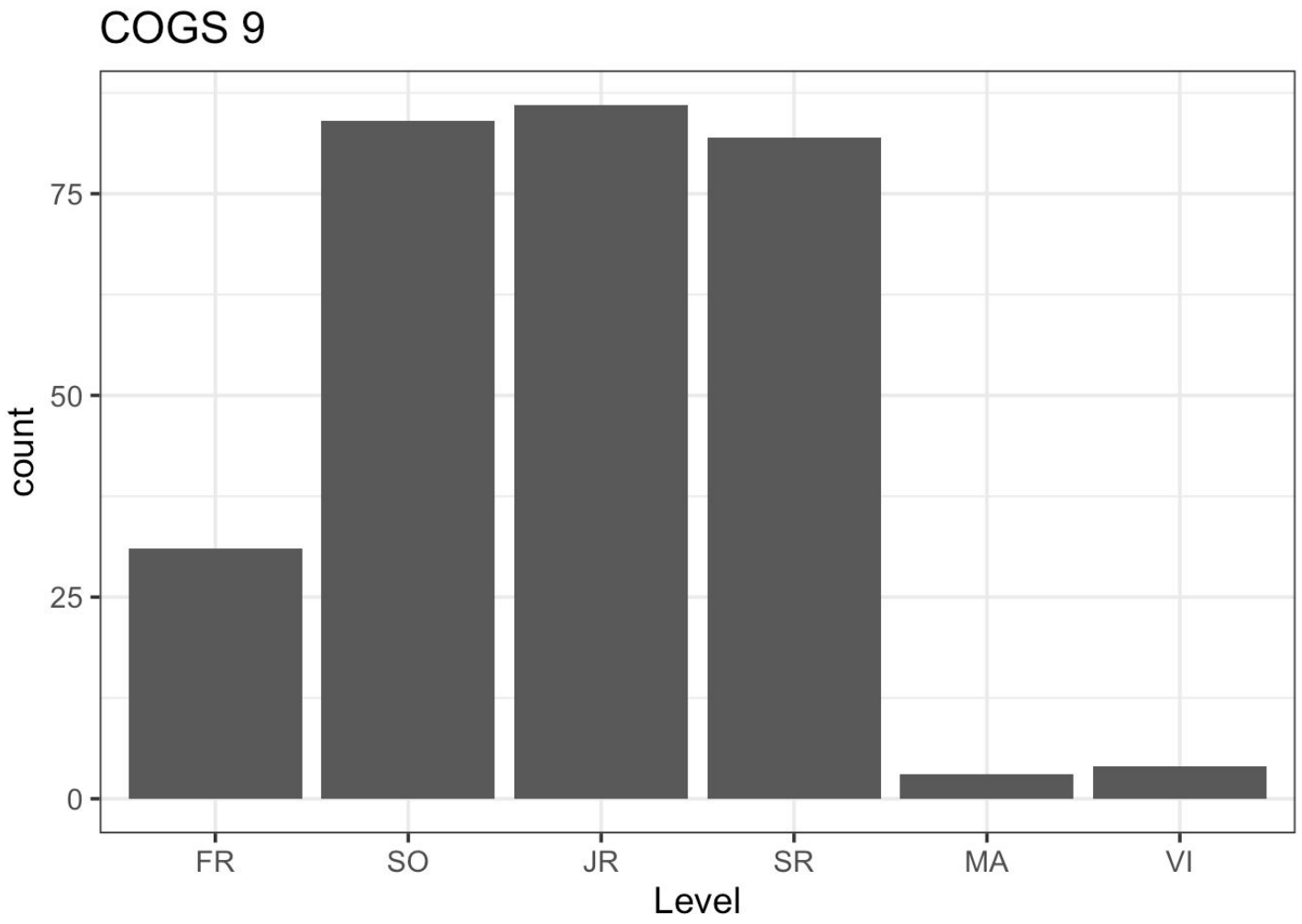
something else really awesome

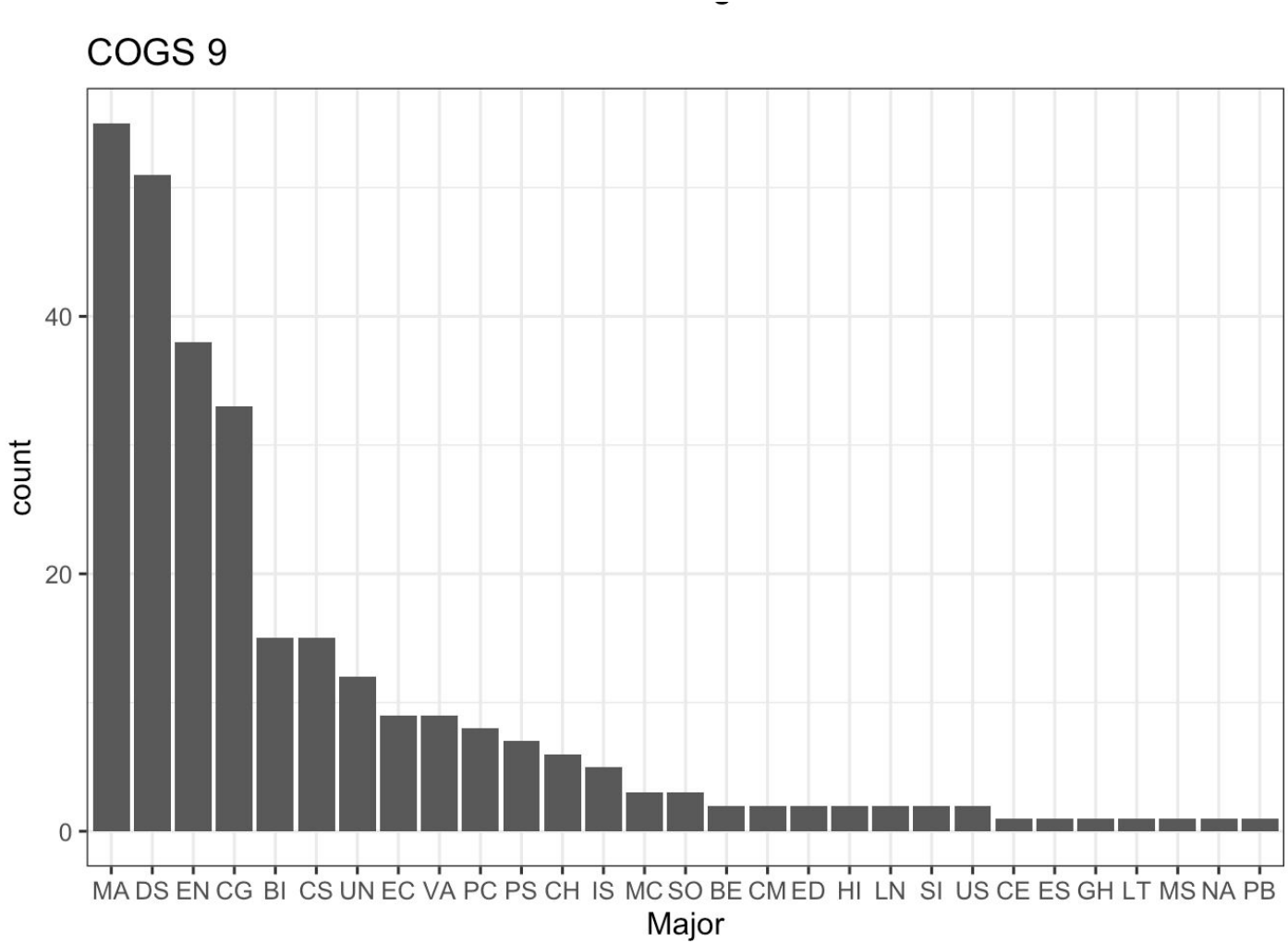
Great! You *Will* Learn Something Useful In This Class

No matter your life's path. Data will be a part or have an affect on it. We will teach you about data and the how/what/whys...

I’m excited to have you all in this class!


COGS 9
count
0
10
20
30
40
FI
GR
MU
RE
SI
SN
TH
WA
College


COGS 9
count
75
50
25
0
FR
SO
JR
SR
MA
VI
Level


COGS 9
count
40
20
0
MA DS EN CG BI CS UN EC VA PC PS CH IS MC SO BE CM ED HI LN SI US CE ES GH LT MS NA PB
Major

Pre-course Survey: https://forms.gle/PzC1j6odSpnPhQ6Q9

Please complete before 11:59 PM Friday (4/7)

COGS9 Student Survey (Spring 2023)

This survey will collect data from students in COGS9 so that I can get to know you all a bit better and to collect data that will be used throughout the course. Your name and PID are included for credit purposes; however, when used in class, the data will be anonymized. In other words, your name will never be attached to your data for in-class use.

How you respond will NOT affect how you do in this class. Also, many questions are NOT required. Please answer honestly, but do NOT answer anything that makes you uncomfortable.