Reminders

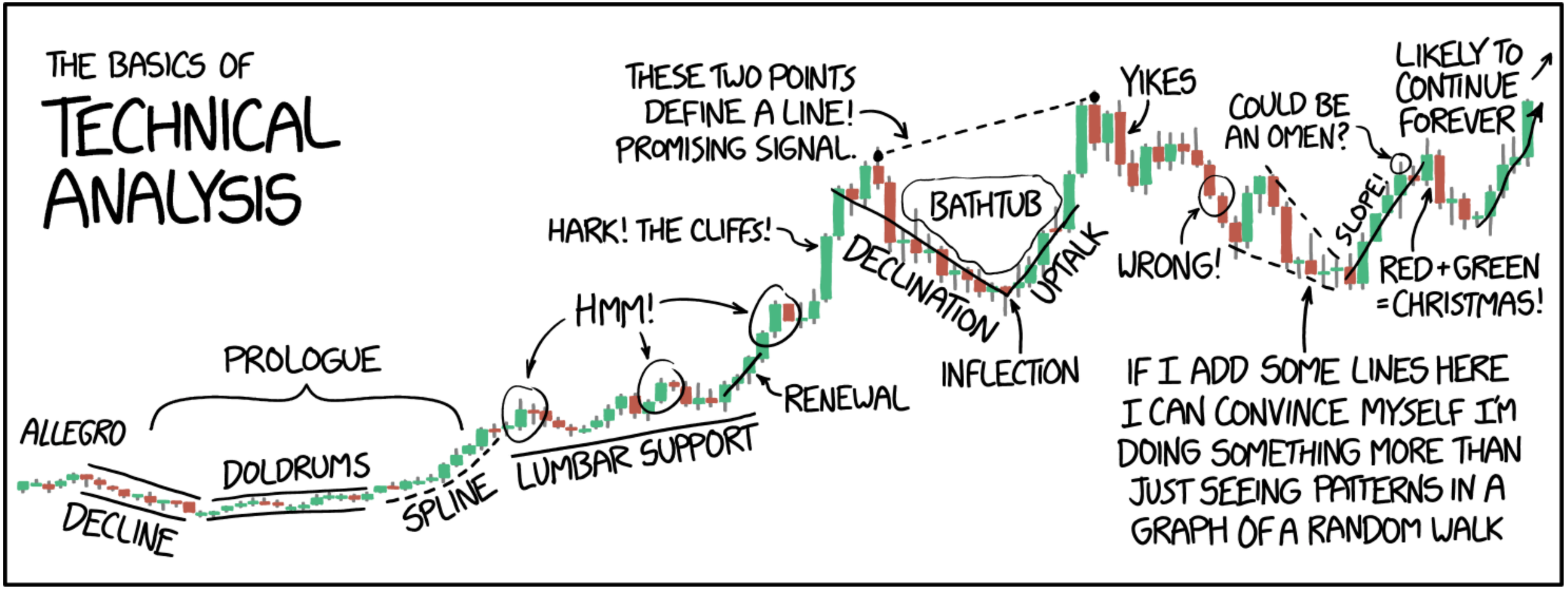
Upcoming due dates
Fri Oct 17th Discussion Lab 2
Mon Oct 20th Quiz 3
Wed Oct 22nd Project review (1 per group)

Repo invites coming today! Click accept before it expires next week!

Analysis toolbox and applying it to class projects

Data Science in Practice

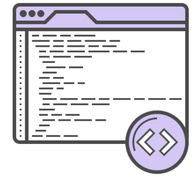
Jason G. Fleischer, PhD Dept. of Cognitive Science UC San Diego https://jgfleischer.com


THE BASICS OF
TECHNICAL ANALYSIS
THESE TWO POINTS DEFINE A LINE! PROMISING SIGNAL.
YIKES
COULD BE AN OMEN?
LIKELY TO CONTINUE FOREVER
HARK! THE CLIFFS!
BATHTUB
SLOPE!
HMM!
DECLINATION
UPTALK
WRONG!
RED+GREEN =CHRISTMAS!
PROLOGUE
ALLEGRO
DOLDRUMS
DECLINE
SPLINE
LUMBAR SUPPORT
RENEWAL
INFLECTION
IF I ADD SOME LINES HERE I CAN CONVINCE MYSELF I'M DOING SOMETHING MORE THAN JUST SEEING PATTERNS IN A GRAPH OF A RANDOM WALK

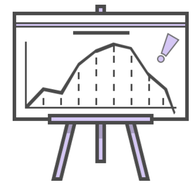
How?

Data

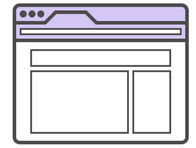
let me show you



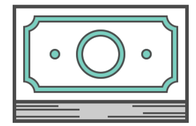
A Model!



Results!

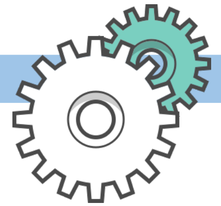


Product!

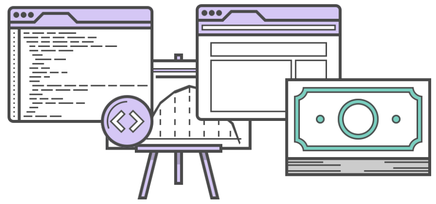


Revenue!

Data



The Analytic Approach
Your Tool Box



The Goods

Descriptive Analysis ✓

Exploratory Analysis ✓

?

Inferential Analysis ✓

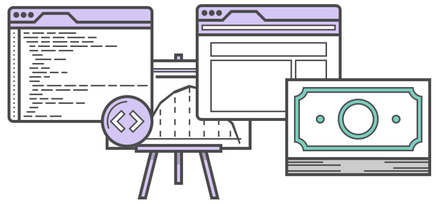
Predictive Analysis ✓

Causal Analysis ✗

Mechanistic Analysis ✗

Data

The Analytic Approach
Your Tool Box



The Goods

Descriptive Analysis ✓

Exploratory Analysis ✓

?

Inferential Analysis ✓

- Classic Statistics (parametric & nonparametric)
- Frequentist & Bayesian
- Text & Geospatial Analysis

Predictive Analysis ✓

- Statistical learning/ML
  - Supervised
  - Unsupervised
- Monte Carlo simulations

Causal Analysis ✓ ✗

variable X ⬆️
causes
variable Y ⬇️

e.x. effects of new medication on some illness by randomized trial

Mechanistic Analysis ✗

variable X ⬆️ 3.2 units
results in
variable Y ⬇️ 1.1 units

e.x. electric current governed by wire size

Summary: Analytical Approaches

Typically Less Effort →

Descriptive Analysis

- 1st thing you do on new data
- Summarize the data
- univariate plots of variables

Exploratory Analysis

- Exploring relationships
- Asking/definning questions
- univariate/bivariate/multivariate analysis and plotting
- formulate hypothesis

Inferential Analysis

- Estimating uncertainty
- test theories (infer) about the population (data gen. process)
- Building inference models

→ *Typically More Effort*

Predictive Analysis

- Building predictive models
- Use historical knowledge to predict future events
- Finding patterns

Mechanistic Analysis

- Understand precise changes one variable has on another
- typically modeled using deterministic equations
- break down complex systems into constituent parts

Causal Analysis

- Determine the average change in one variable when you alter another
- typically requires experiments (e.g. randomized studies)
- manipulate one variable observe effect on other

General question: What impacts politics in America?

Data Science question: Is there a relationship between the sentiment of political words in South Park and America's presidential approval rating?

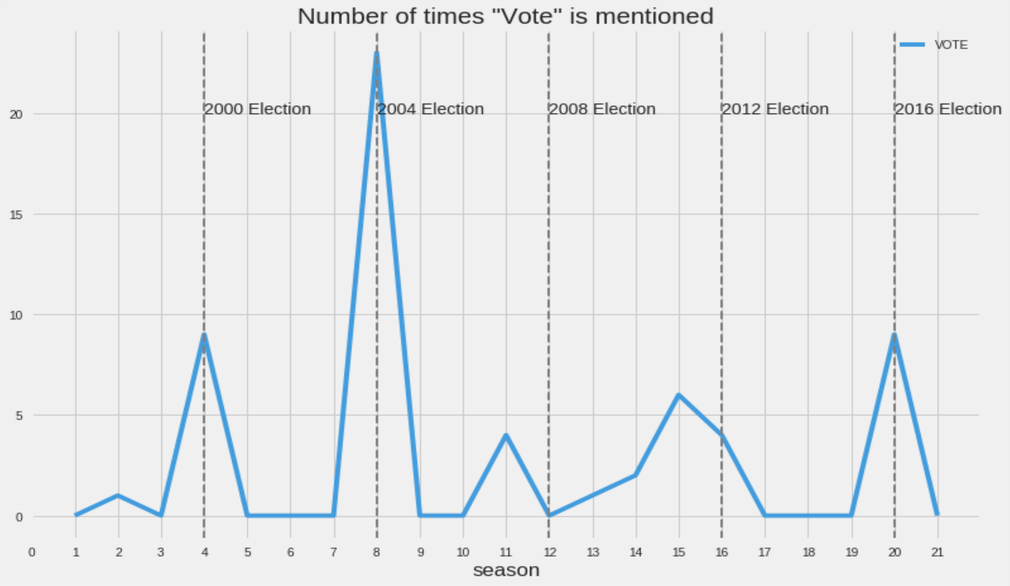
Descriptive

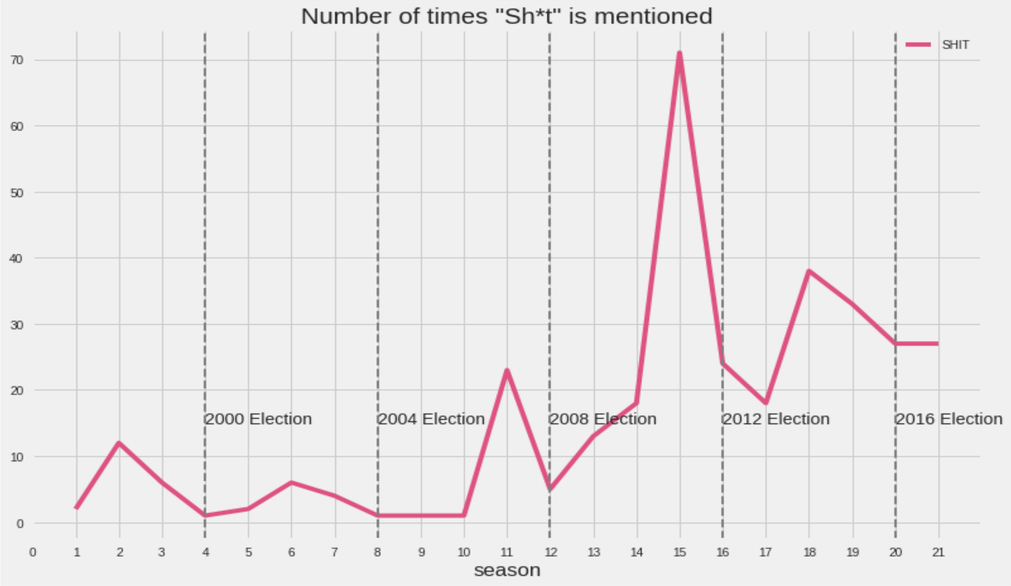
Exploratory

Inferential

Text Analysis

Classic Statistics (parametric & nonparametric)





https://github.com/COGS108/FinalProjects-Wi18/blob/master/071-FinalProject.ipynb

General question: How has COVID-19 impacted students?

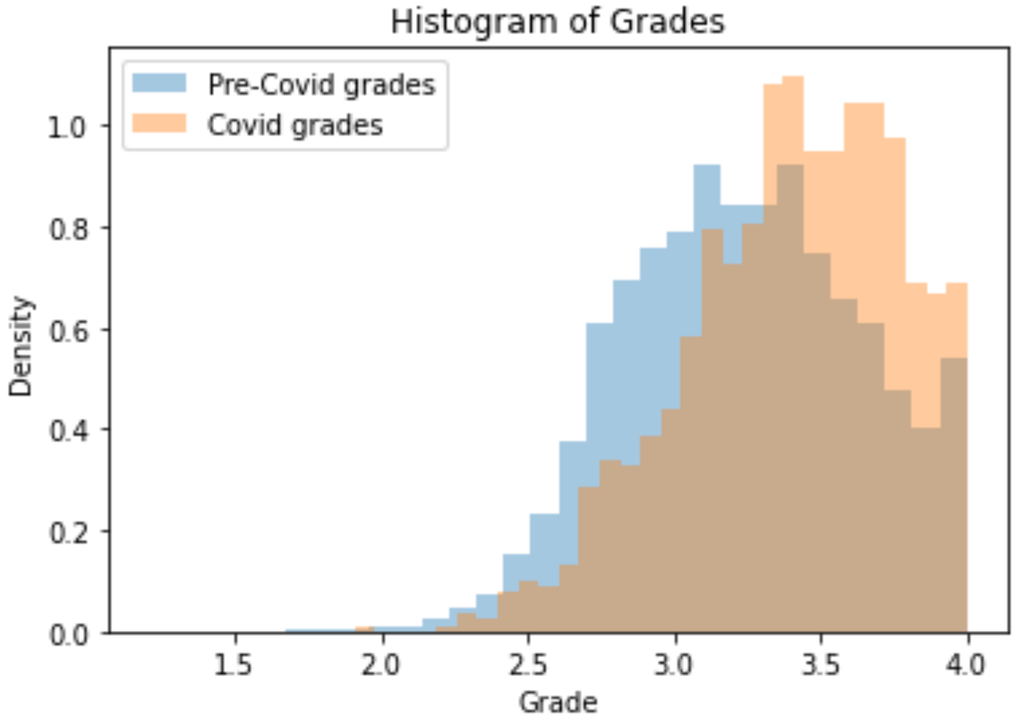
Data Science question: At UCSD, is there a difference between students' grades and how they rate their classes before COVID-19 and during remote learning, due to COVID-19?

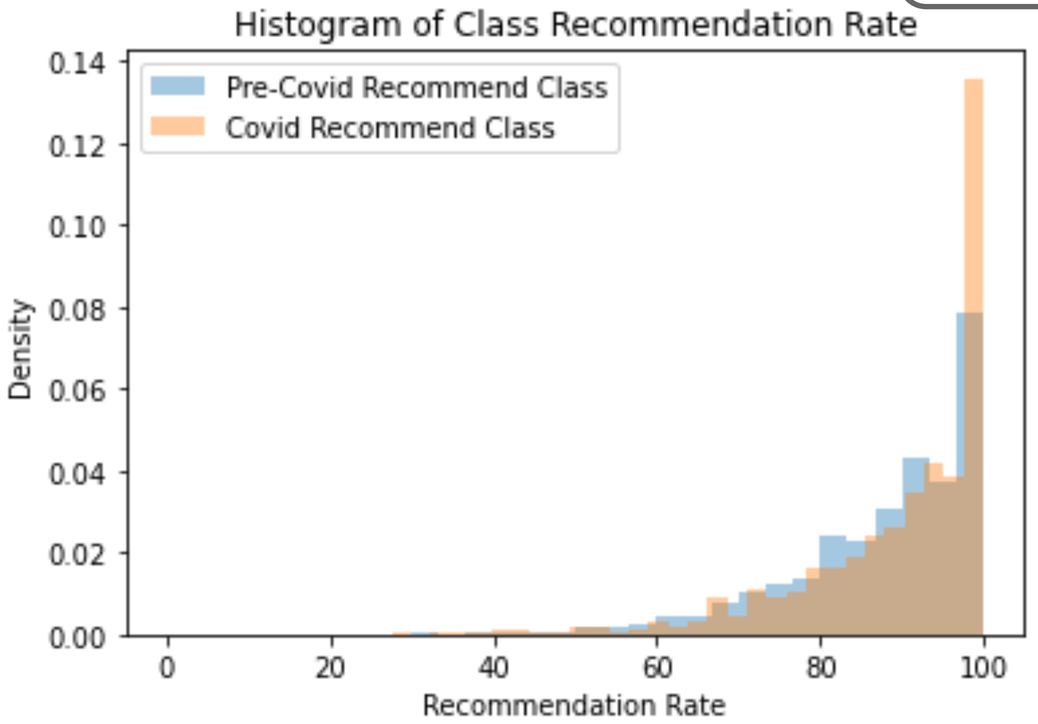
Descriptive

Exploratory

Inferential

Classic Statistics (parametric & nonparametric)

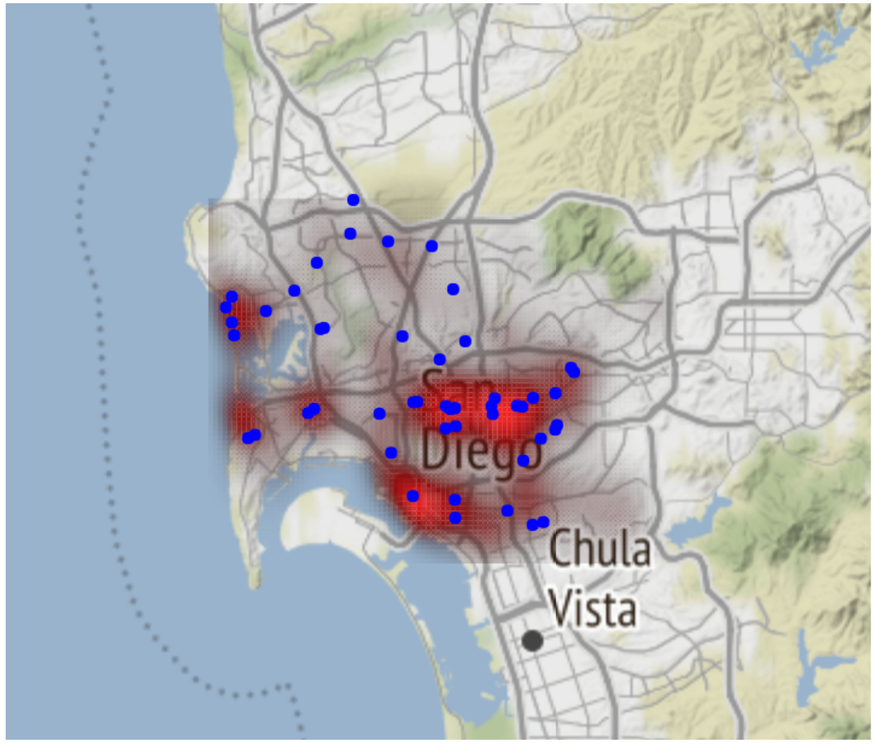


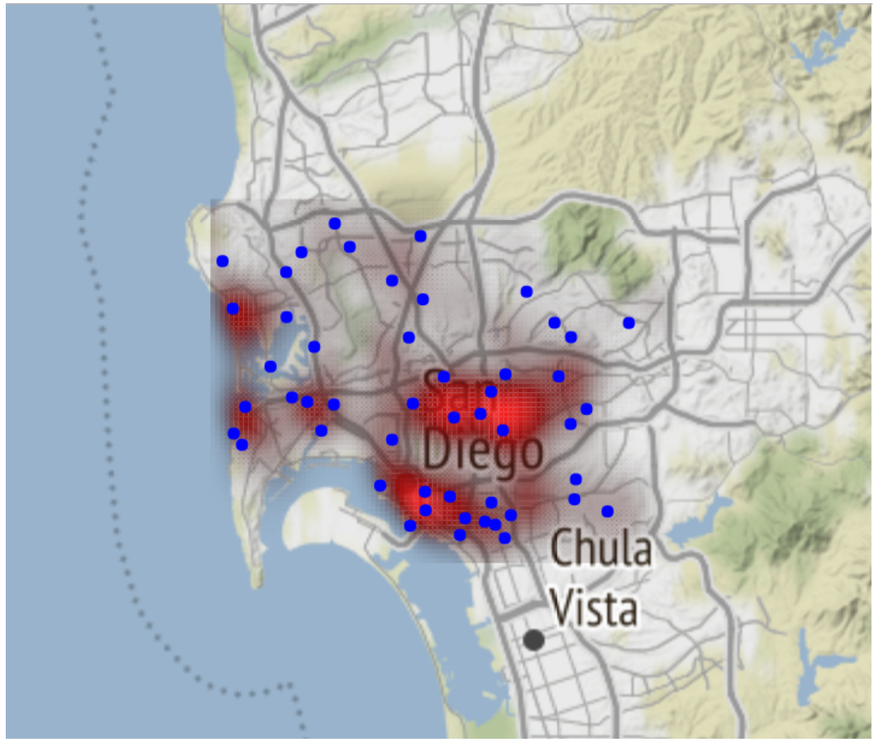


https://github.com/COGS108/FinalProjects-Wi21/blob/main/FinalProject_group084.ipynb

General question: Why isn't police response time always the same?

Data Science question: Where should police cars be stationed, accounting for crime levels and time of day, to make police response times equitable throughout San Diego?





Descriptive

Exploratory

Predictive

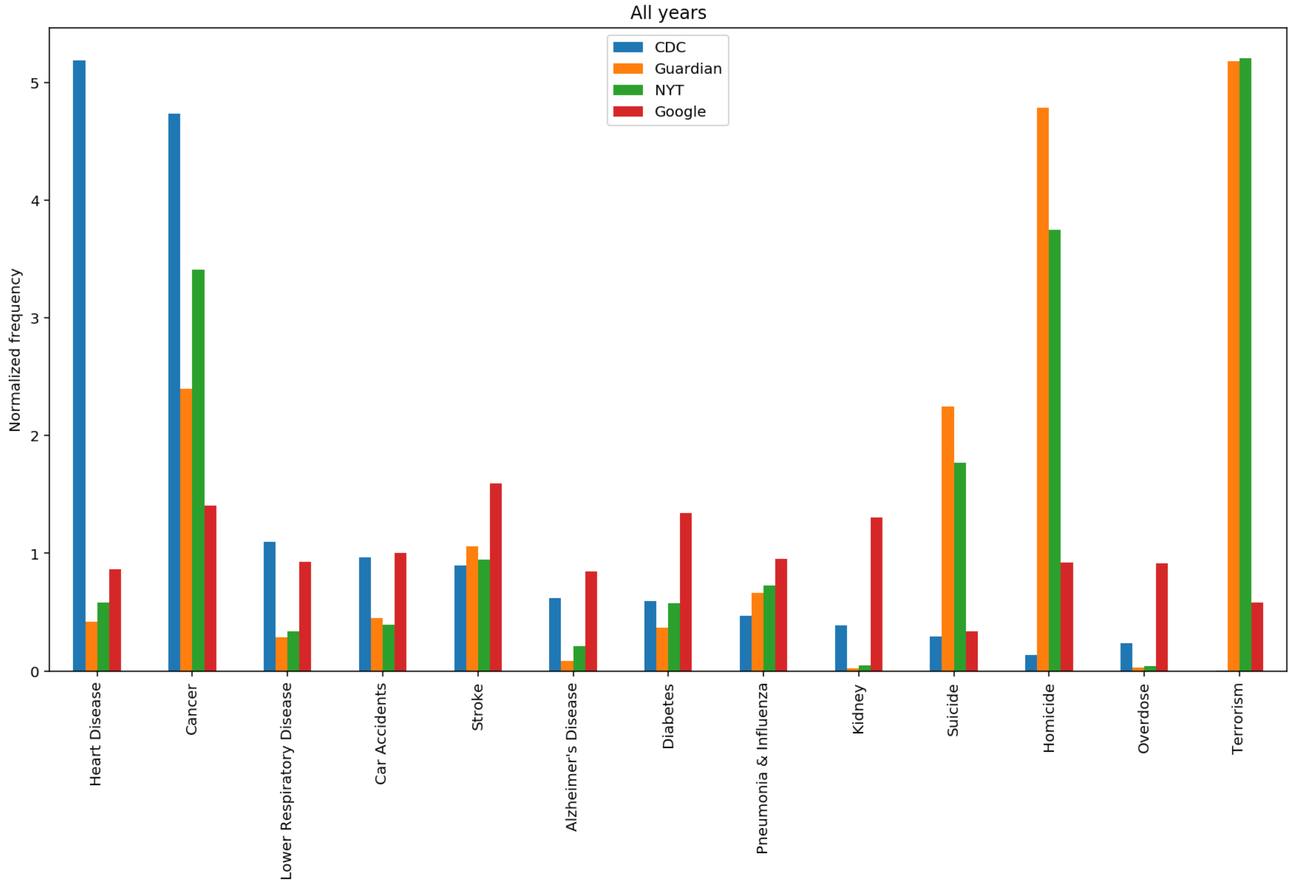
Inferential

Geospatial Analysis

https://github.com/COGS108/FinalProjects-Sp17/blob/master/029-FinalProject.ipynb

General question: What gets too much attention in the news?

Data Science Question: Is there a relationship over time between cause of death terms in the *NYT*, The Guardian, and Google trends data relative to data from the CDC?



https://github.com/COGS108/FinalProjects-Wi18/blob/master/010-FinalProject.ipynb

Descriptive

Exploratory

Inferential

Text Analysis

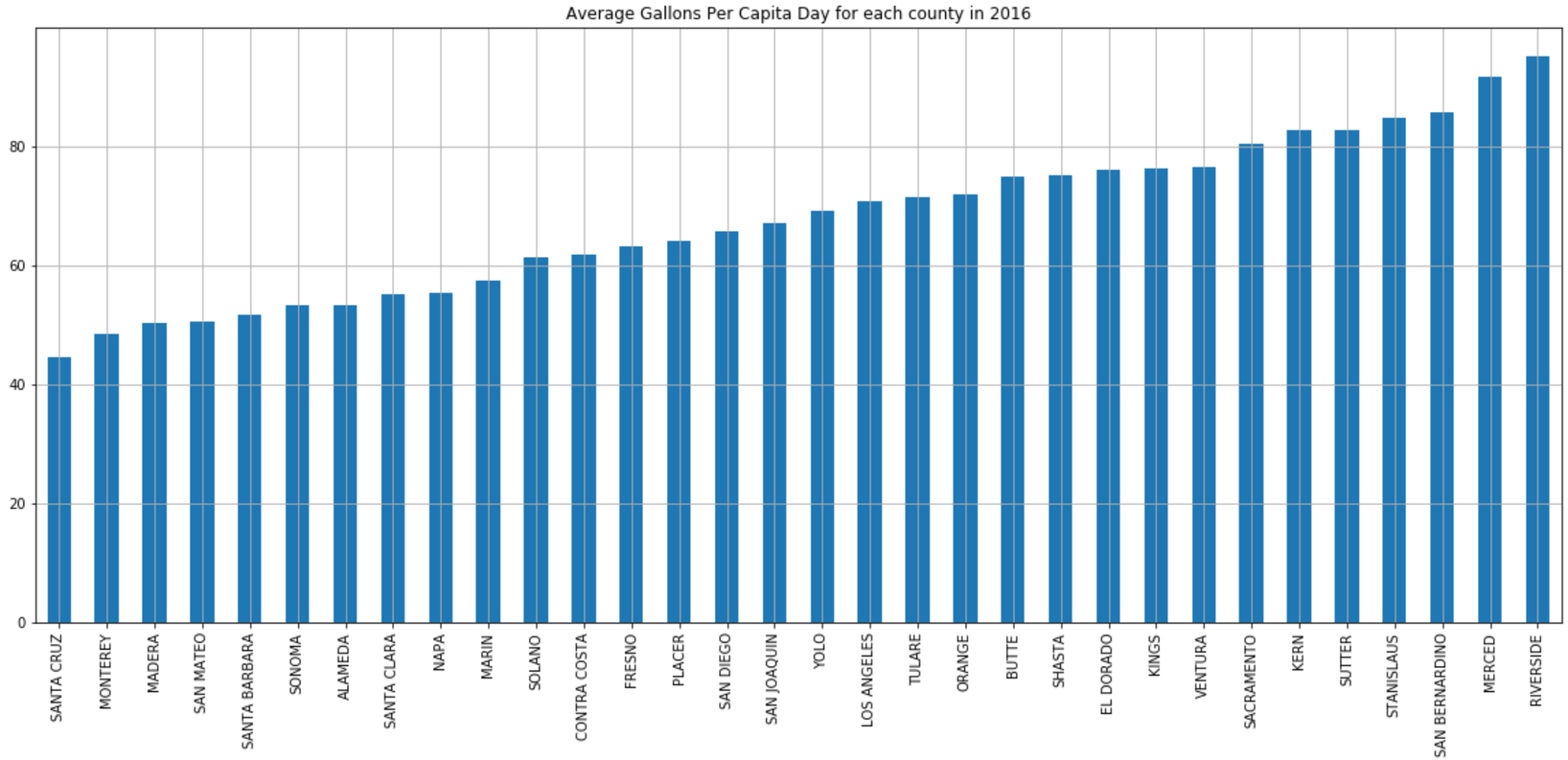
Classic Statistics (parametric & nonparametric)

In case of the total drought in California, how many desalination plant projects we need to supply residential use water for population who live in urban areas in California?

Descriptive

Exploratory

Predictive



https://github.com/COGS108/FinalProjects-Sp17/blob/master/018-FinalProject.ipynb

Project proposals

Due October 29

- Good research question that is unambiguous and detailed enough to tell us what data you will need and what kinds of analysis it takes to answer the question
- So think about how to get data, but also what you will do with it
- Brainstorm then refine

Project ideas I will throw your way

- How many e-scooter accidents are there at UCSD? Are there more e-scooter accidents than e-bike?
  - EMS or hospital admissions data for UCSD Medical and / or Scripps Green (the two closest hospitals) e.g. https://www.sandiegocounty.gov/content/sdc/ems/CoSD-LEMSIS.html
- Is crime on campus changing year on year, in terms of both what crimes are committed and their rates? Are most crimes committed by other students or by outsiders? Are crime rates positively correlated with the student population size?
  - http://ucsdpd.com and its data is here https://github.com/axelsagundo/ucsd-crime-logs-dataset
- Is there a relationship between a student's sleep onset time, sleep duration, and their academic performance? Is there a relationship between their normal work time of day and their academic performance? (And/or also ask about wellbeing)
  - Setup a survey asking for GPA, typical work schedule for assignments, and upload of Apple Watch sleep data. If also wellbeing or depression there a variety of short surveys you can include