Course Reminders

- Due today (11:59PM)
  - Q3
- Due Friday (11:59 PM)
  - D3
  - Project Proposal
  - Weekly Project Survey (*optional EC; released on Wed*)
- Github invite expired? Ask on Piazza in private message to instructors

Kinds of analysis

Jason G. Fleischer, Ph.D.
Asst. Teaching Professor
Department of Cognitive Science, UC San Diego
jfleischer@ucsd.edu

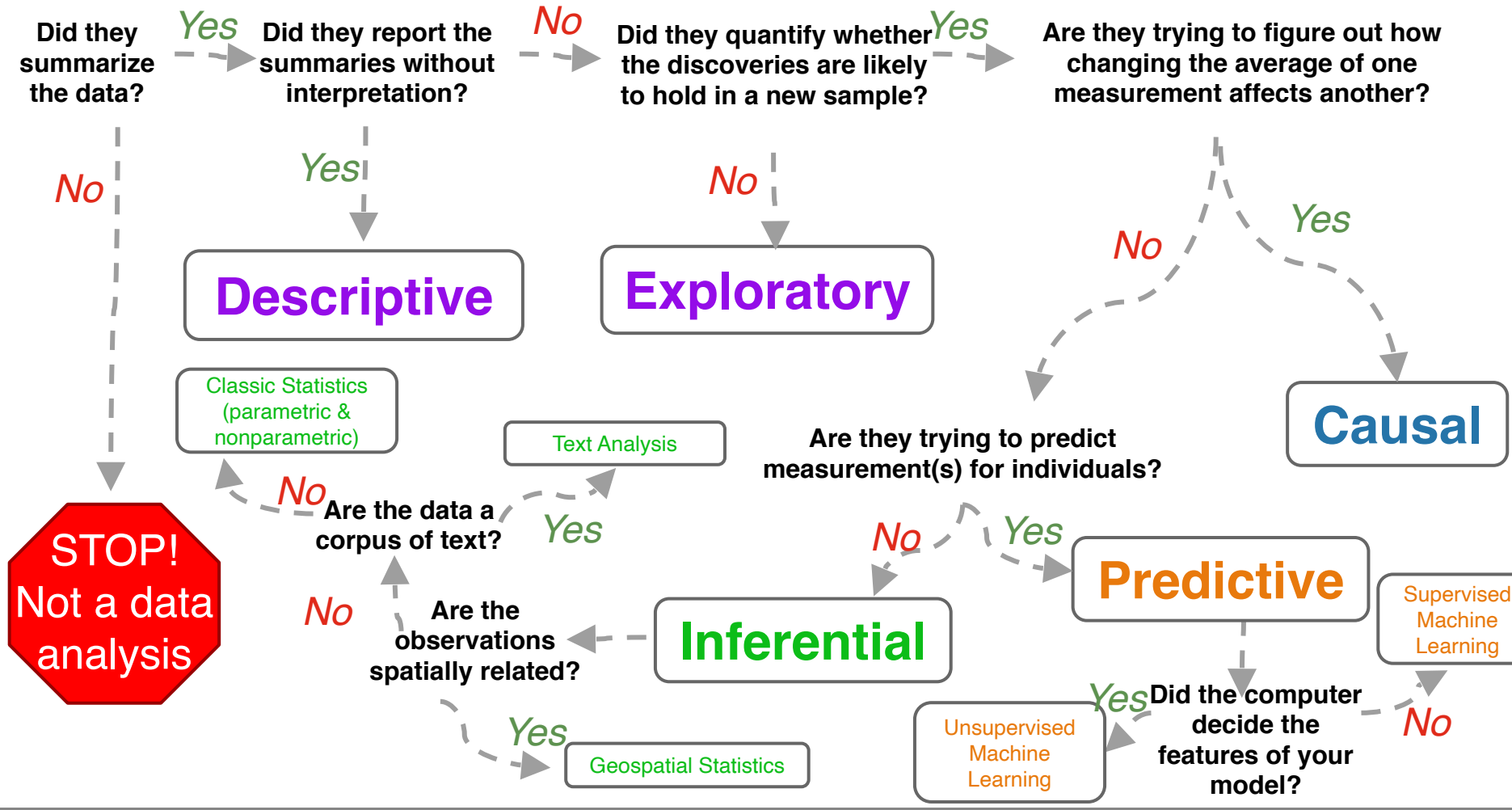
@jasongfleischer

https://jgfleischer.com

"Data science is the process of formulating a quantitative question that can be answered with data, collecting and cleaning the data, analyzing the data, and communicating the answer to the question to a relevant audience."

To do this, you have to *look at, describe, and explore* the data

https://simplystatistics.org/2015/03/17/data-science-done-well-looks-easy-and-that-is-a-big-problem-for-data-scientists/


Did they summarize the data?
Yes
Did they report the summaries without interpretation?
No
Did they quantify whether the discoveries are likely to hold in a new sample?
Yes
Are they trying to figure out how changing the average of one measurement affects another?
No
Yes
No
Yes
No
Descriptive
Exploratory
Causal
Classic Statistics (parametric & nonparametric)
Text Analysis
Are they trying to predict measurement(s) for individuals?
No
Are the data a corpus of text?
Yes
No
Yes
Predictive
Supervised Machine Learning
No
Are the observations spatially related?
Inferential
Yes
Did the computer decide the features of your model?
Yes
No
Unsupervised Machine Learning
Yes
Geospatial Statistics
STOP! Not a data analysis

Summary: Analytical Approaches

1. Descriptive (and Exploratory) Data Analysis are the first step(s)
2. Inference establishes relationships
    a. Classic Statistics
    b. Geospatial Analysis
    c. Text Analysis
3. Machine Learning is for prediction
    a. Supervised
    b. Unsupervised
4. Experiments best way to establish causality

Exploring Analyses

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?

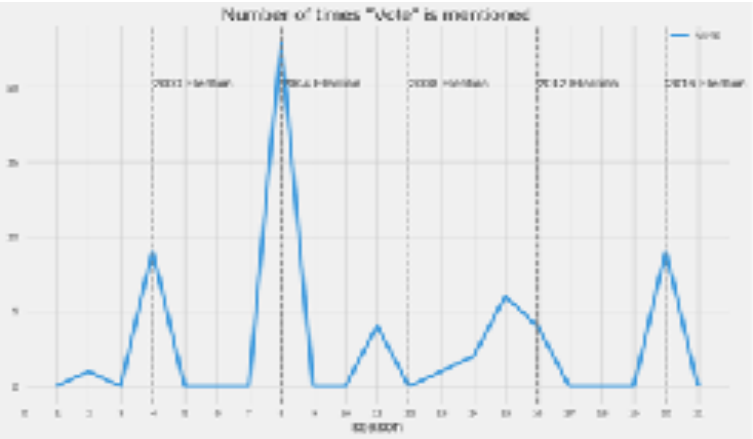
Descriptive

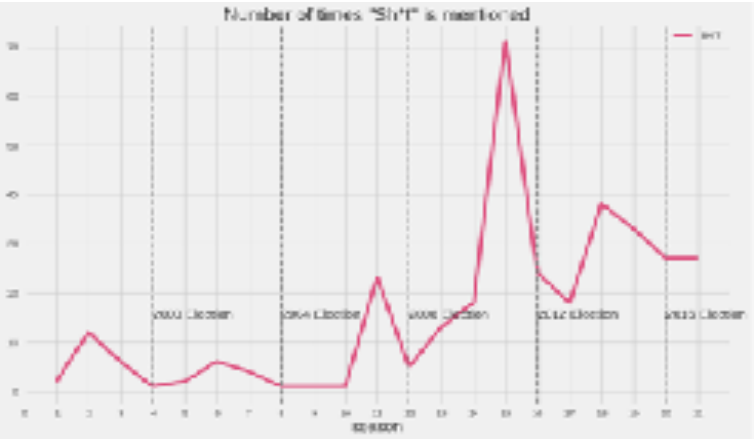
Exploratory

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?

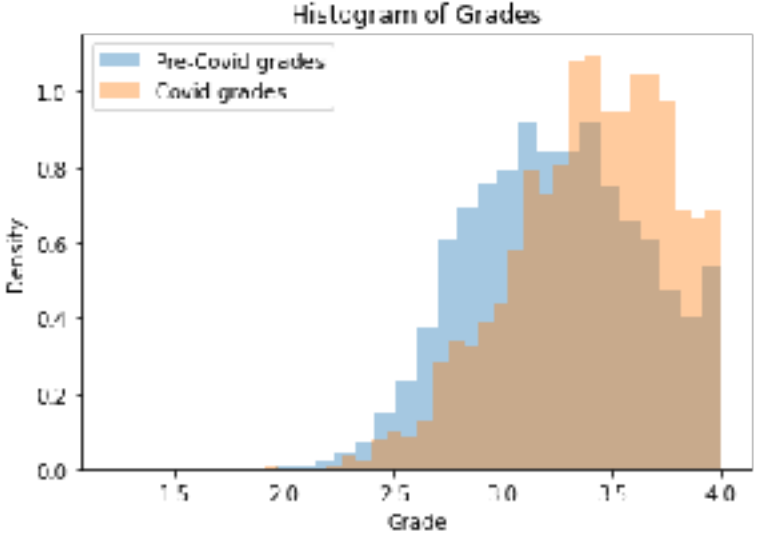
Data Science question: At UCSD, is there a difference between students' grades and how they rate their classes before COVID-19 and during remote learning, due to COVID-19?

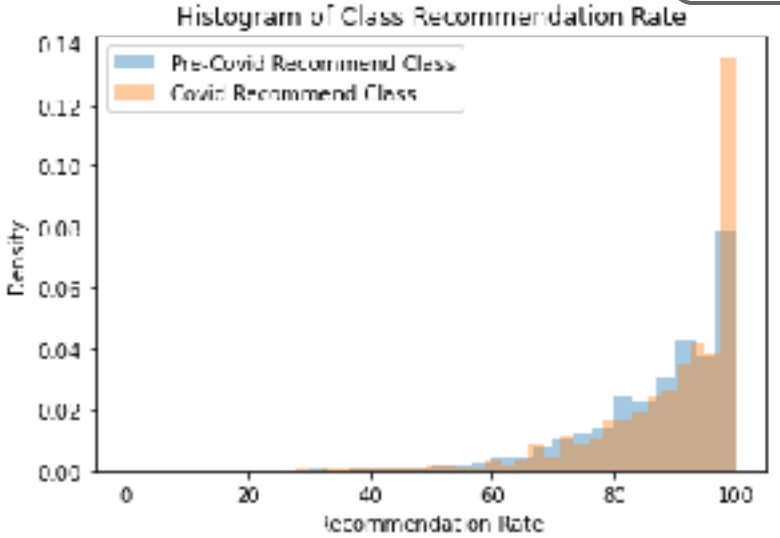
Descriptive

Exploratory

Inferential

Classic Statistics (parametric & nonparametric)

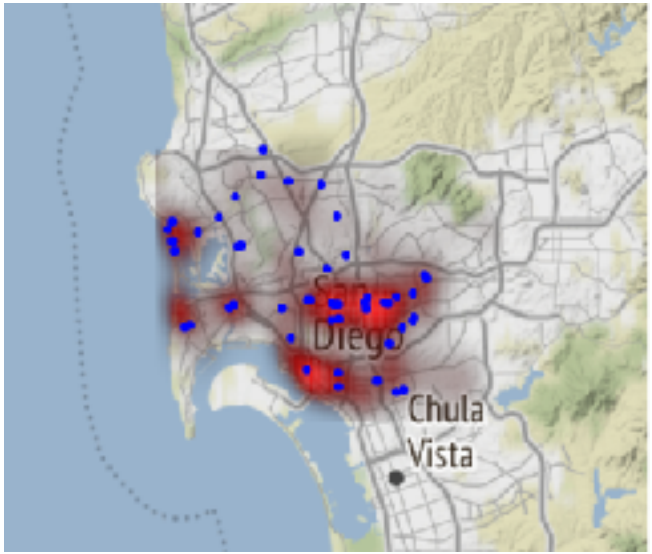


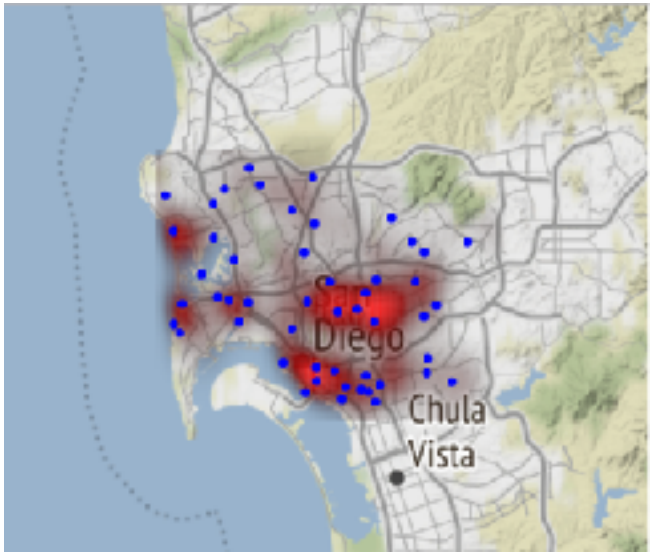


https://github.com/COGS108/FinalProjects-Wi21/blob/main/FinalProject_group084.ipynb

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

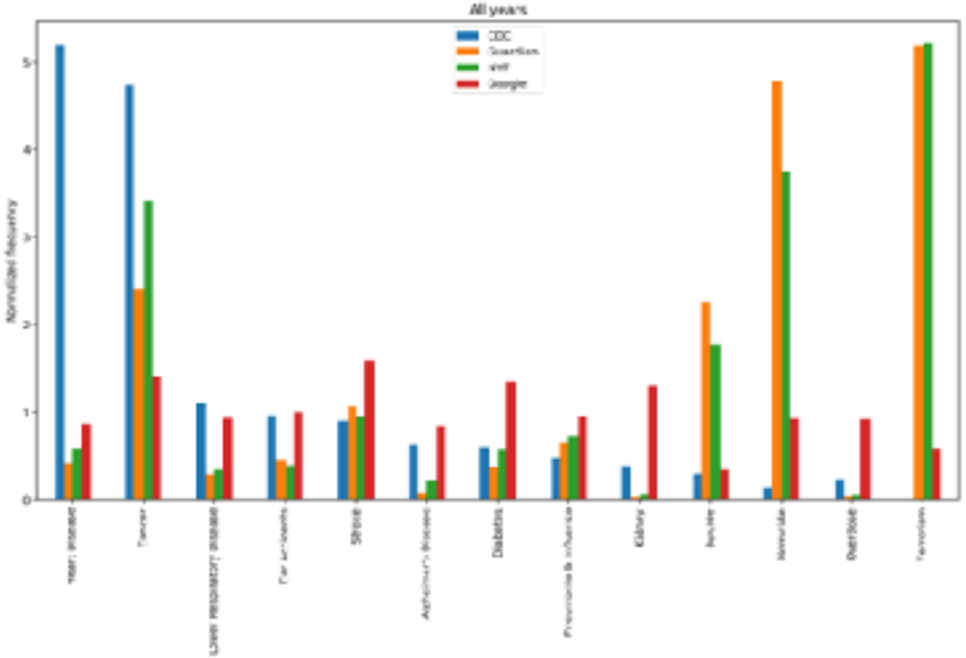
Inferential

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?



Descriptive

Exploratory

Inferential

Text Analysis

Classic Statistics (parametric & nonparametric)

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