Course Reminders

- Due Thur

  - In person lecture survey on Canvas

- Due Friday (11:59 PM)
  - D3
  - Project Proposal
  - Weekly Project Survey (*optional; 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
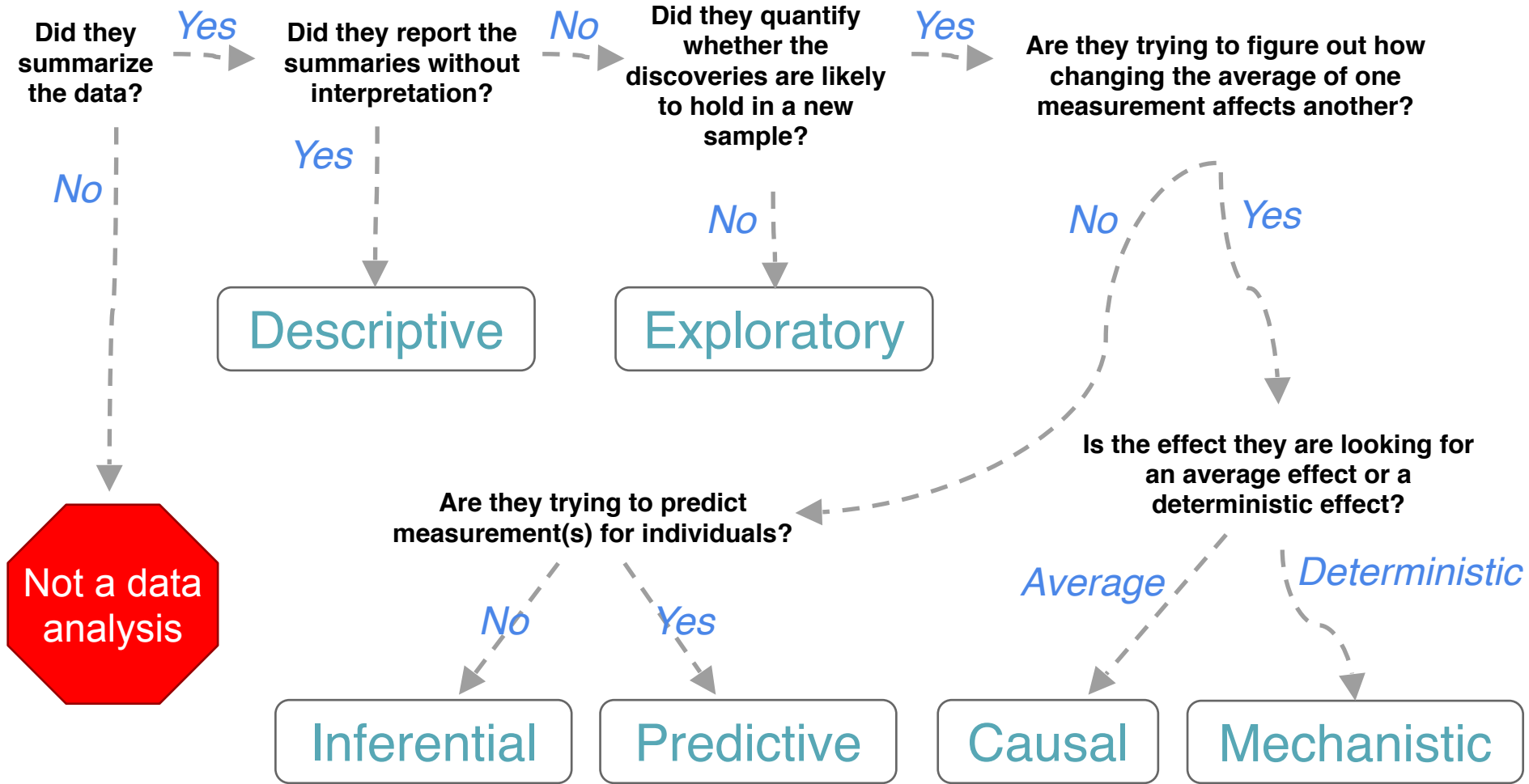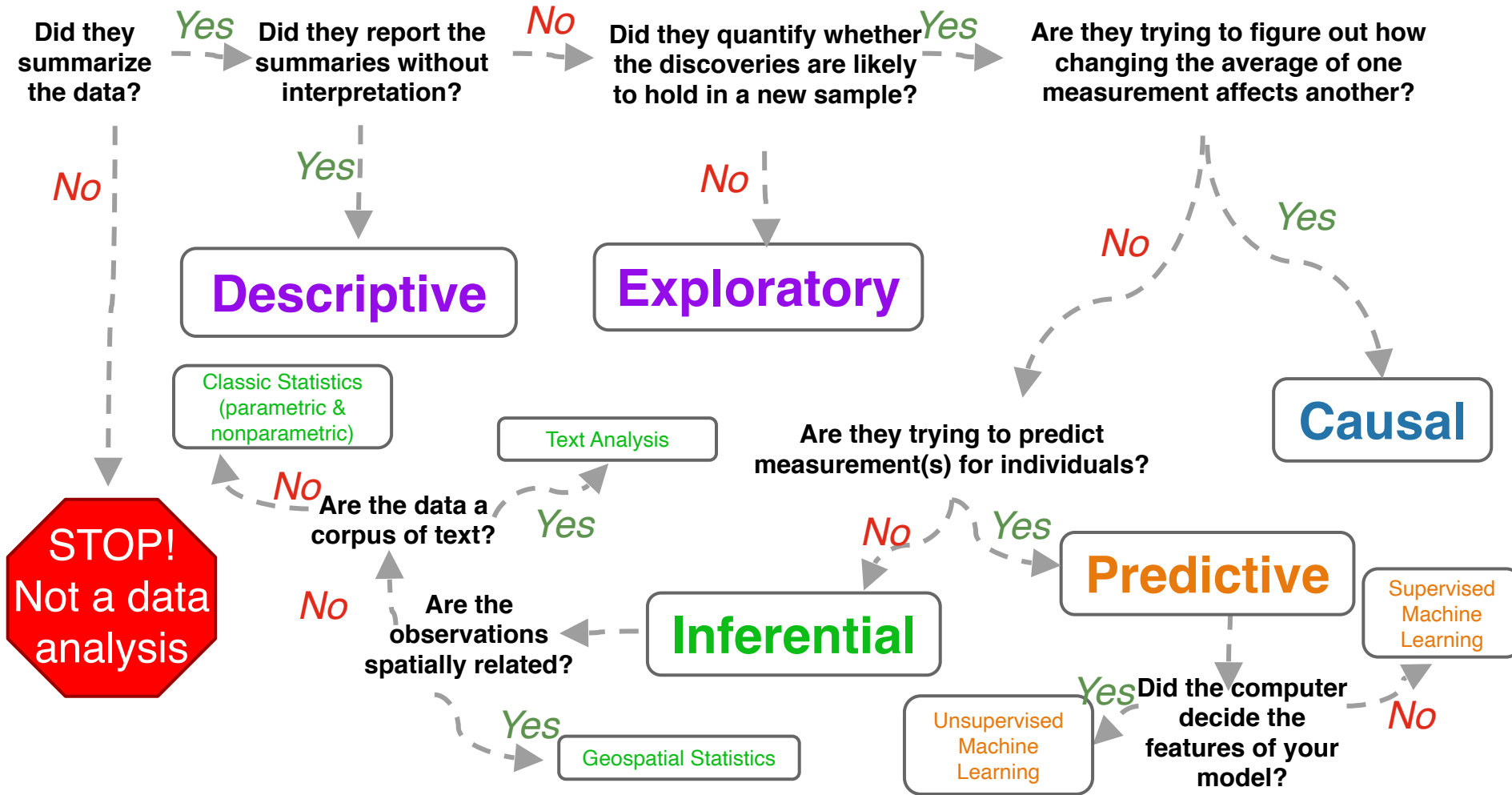**jfleischer@ucsd.edu**

@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, <u>analyzing the data</u>, and communicating the answer to the question to a relevant audience."

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

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?**

Did they summarize the data? — *No* → STOP! Not a data analysis

Did they report the summaries without interpretation? — *Yes* → **Descriptive**

Did they quantify whether the discoveries are likely to hold in a new sample? — *No* → **Exploratory**

Are they trying to figure out how changing the average of one measurement affects another? — *No* → **Are they trying to predict measurement(s) for individuals?**

Are they trying to figure out how changing the average of one measurement affects another? — *Yes* → **Causal**

Are they trying to predict measurement(s) for individuals? — *No* → **Inferential**

Are they trying to predict measurement(s) for individuals? — *Yes* → **Predictive**

Classic Statistics (parametric & nonparametric)

Are the data a corpus of text? — *No* → Classic Statistics (parametric & nonparametric)

Are the data a corpus of text? — *Yes* → Text Analysis

Inferential → **Are the observations spatially related?**

Are the observations spatially related? — *No* → **Are the data a corpus of text?**

Are the observations spatially related? — *Yes* → Geospatial Statistics

Predictive → **Did the computer decide the features of your model?**

Did the computer decide the features of your model? — *Yes* → Unsupervised Machine Learning

Did the computer decide the features of your model? — *No* → Supervised Machine Learning

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)

Number of times "Vote" is mentioned

Number of times "Sh*t" is mentioned
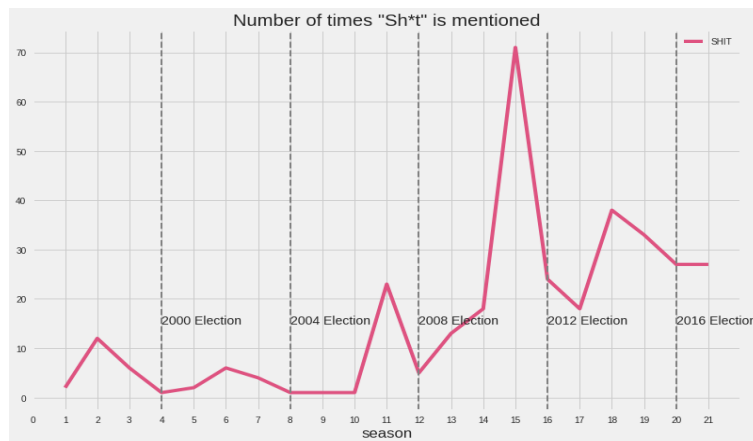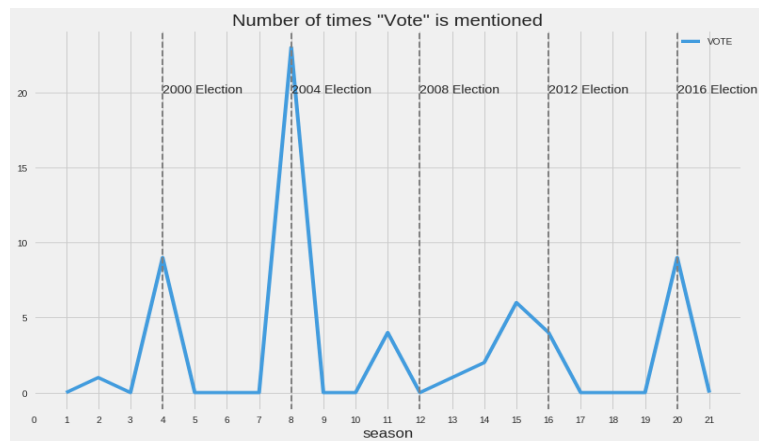
General question: How has COVID-19 impacted students?

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?

Histogram of Grades
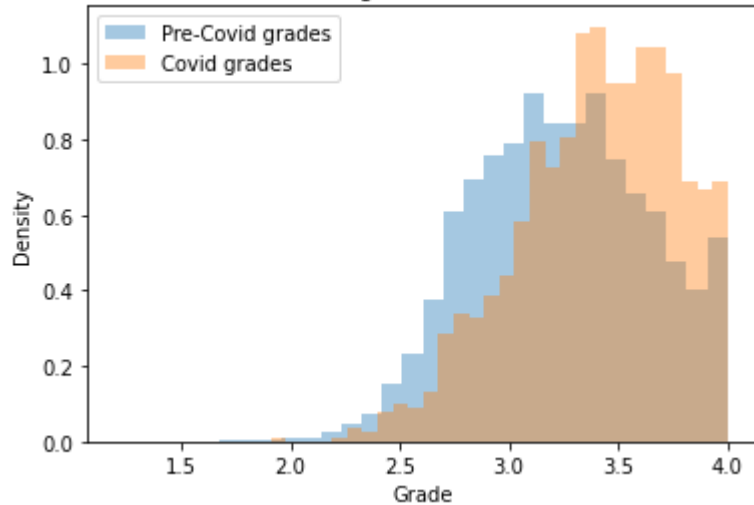
Histogram of Class Recommendation Rate
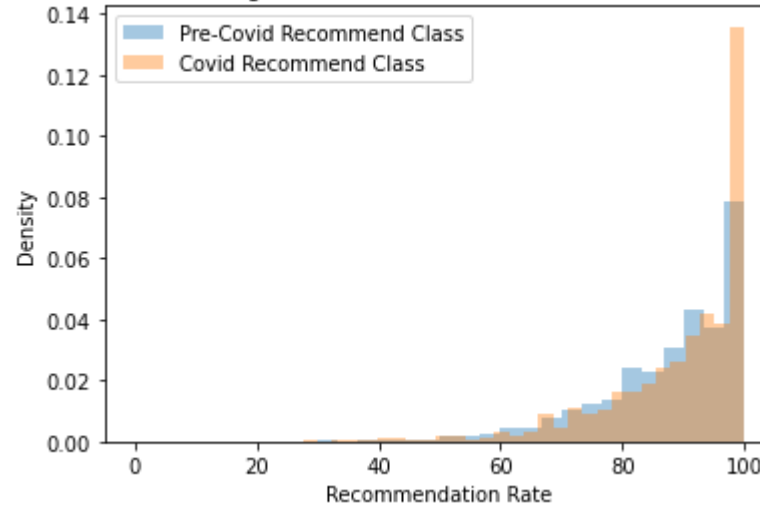
General question: Why isn't police response time always the same?

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?
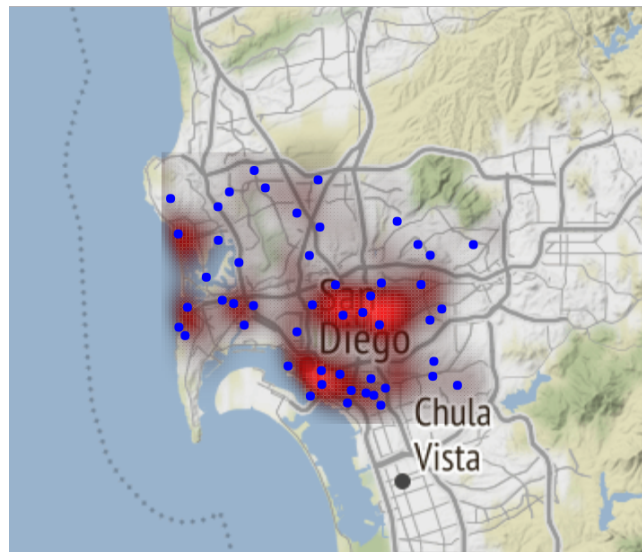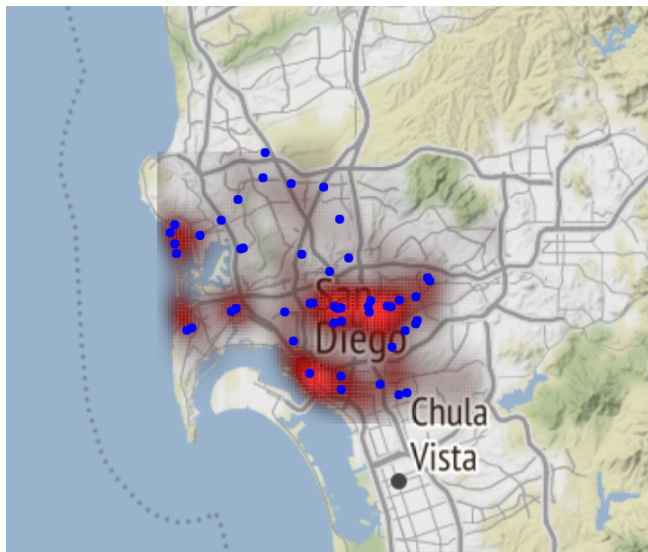

Descriptive

Exploratory

Predictive

Inferential

Geospatial Analysis

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)