

Course Reminders

- Due Friday (11:59 PM)
 - D3
 - Q3
 - Project Proposal
 - Weekly Project Survey (*optional; released on Wed*)

Approaches to Analysis

Shannon E. Ellis, Ph.D
UC San Diego

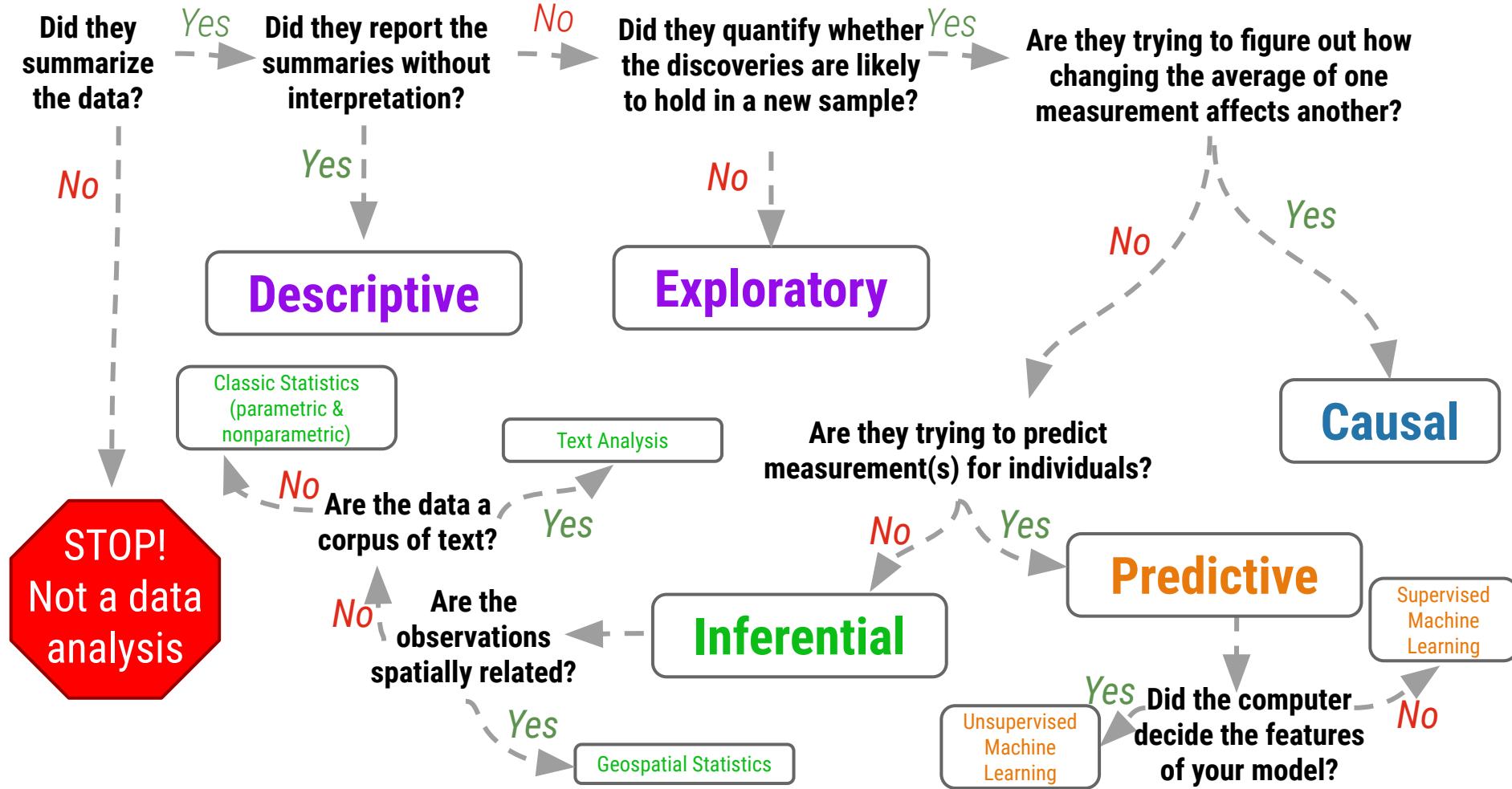


Department of Cognitive Science
sellis@ucsd.edu

*“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



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?

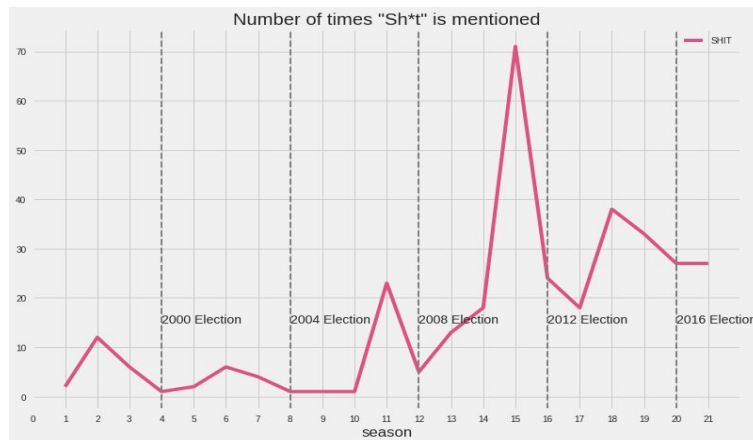
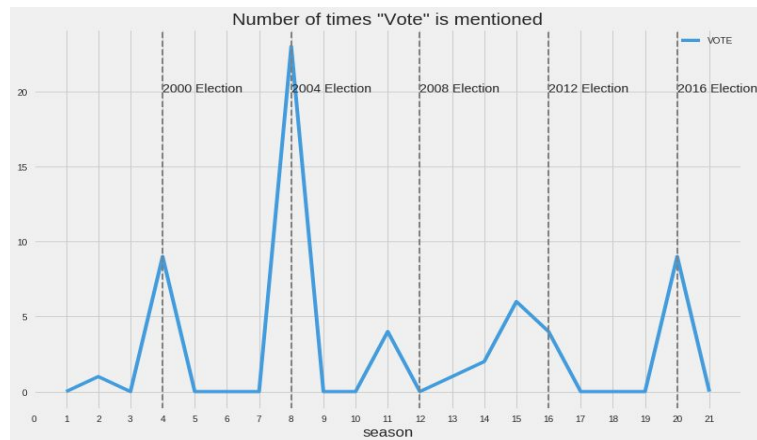
Descriptive

Exploratory

Inferential

Text Analysis

Classic Statistics
(parametric &
nonparametric)



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

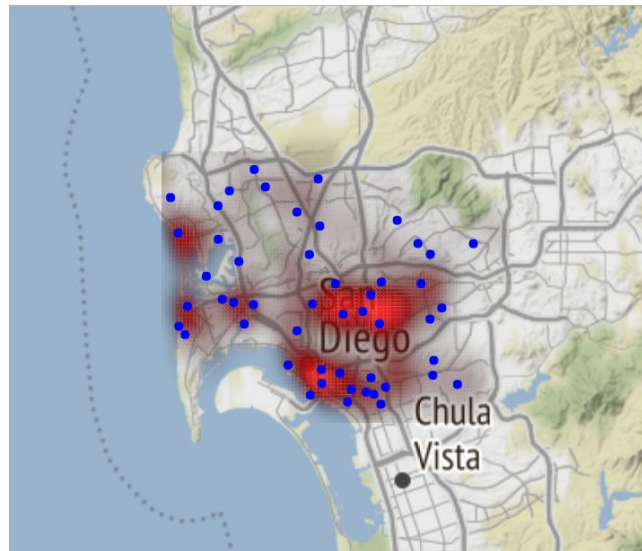
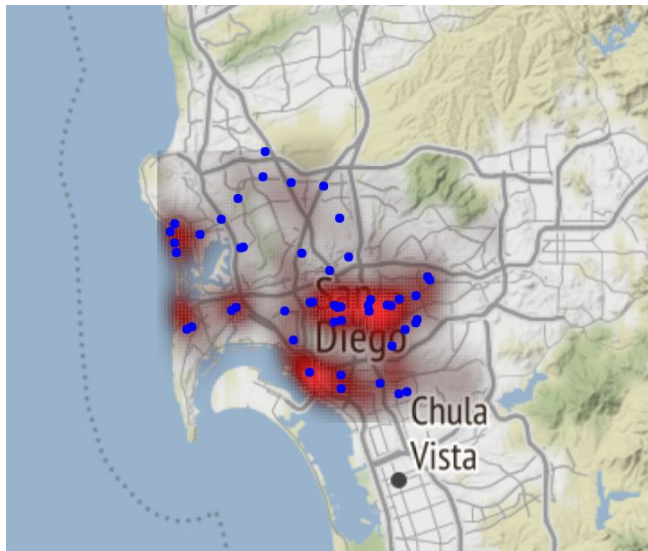
Data Science question: Where should police cars be stationed, accounting for crime levels and time of day, to make police response times equitable throughout San Diego?

Descriptive

Exploratory

Predictive

Inferential



General question: How has COVID-19 impacted students?

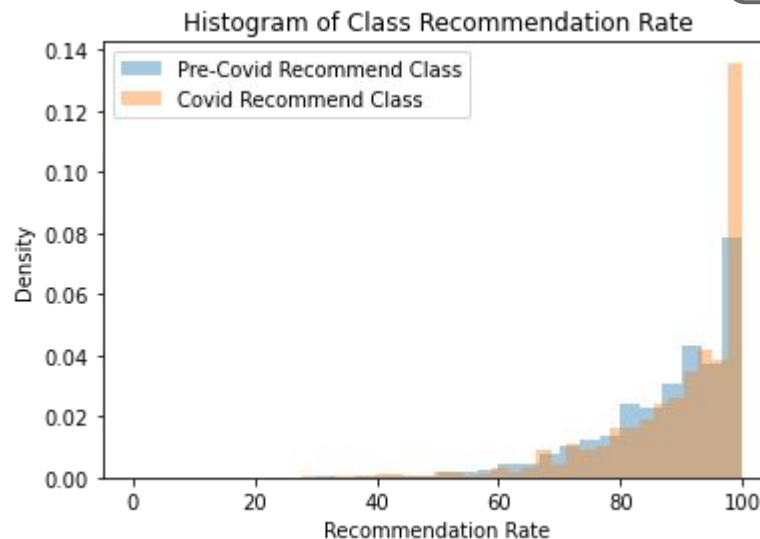
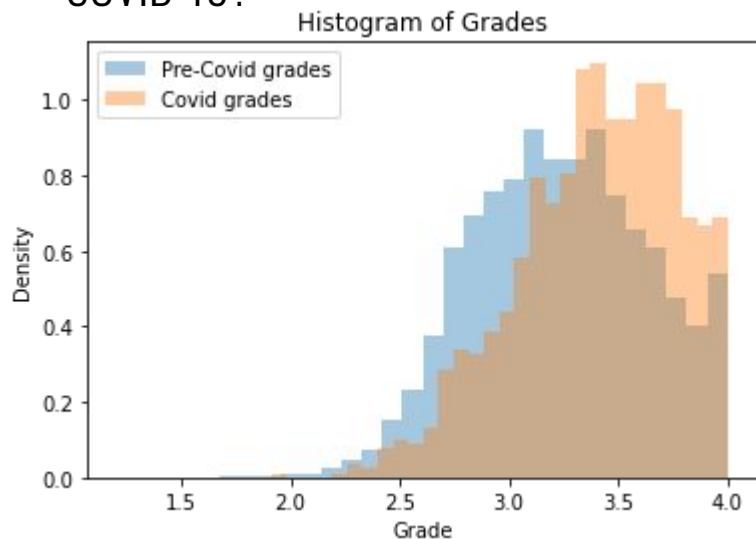
Data Science question: At UCSD, is there a difference between students' grades and how they rate their classes before COVID-19 and during remote learning, due to COVID-19?

Descriptive

Exploratory

Inferential

Classic Statistics
(parametric &
nonparametric)



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?

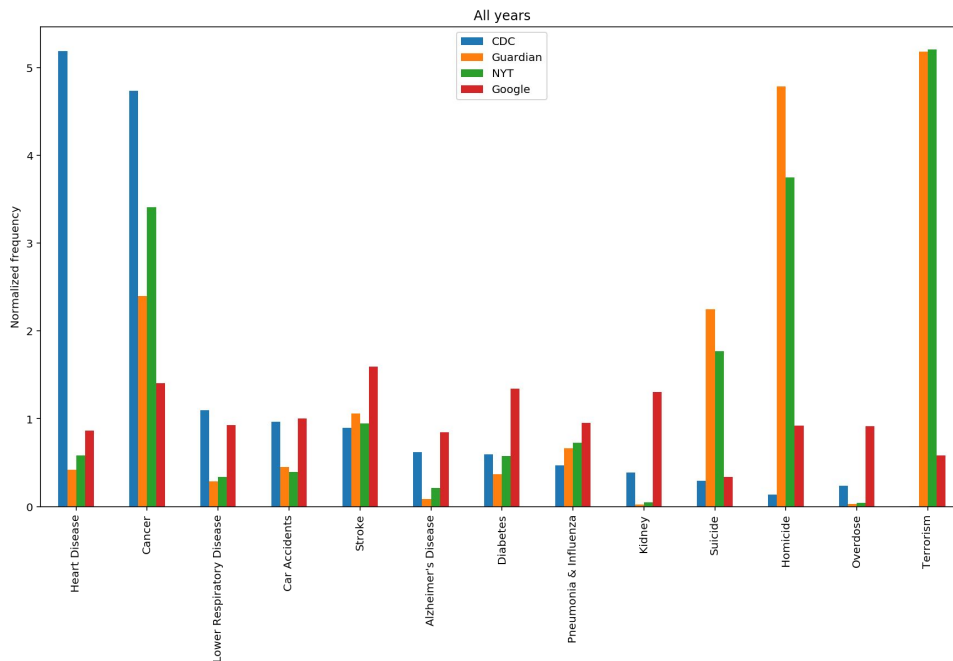
Descriptive

Exploratory

Inferential

Text Analysis

Classic Statistics
(parametric &
nonparametric)



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

