Approaches to Analysis

C. Alex Simpkins Jr., Ph.D

RDPRobotics LLC,

UC San Diego, Department of Cognitive Science

rdprobotics@gmail.com

csimpkinsjr@ucsd.edu

Lectures :http://casimpkinsjr.radiantdolphinpress.com/pages/cogs108_ss1_23/index.html
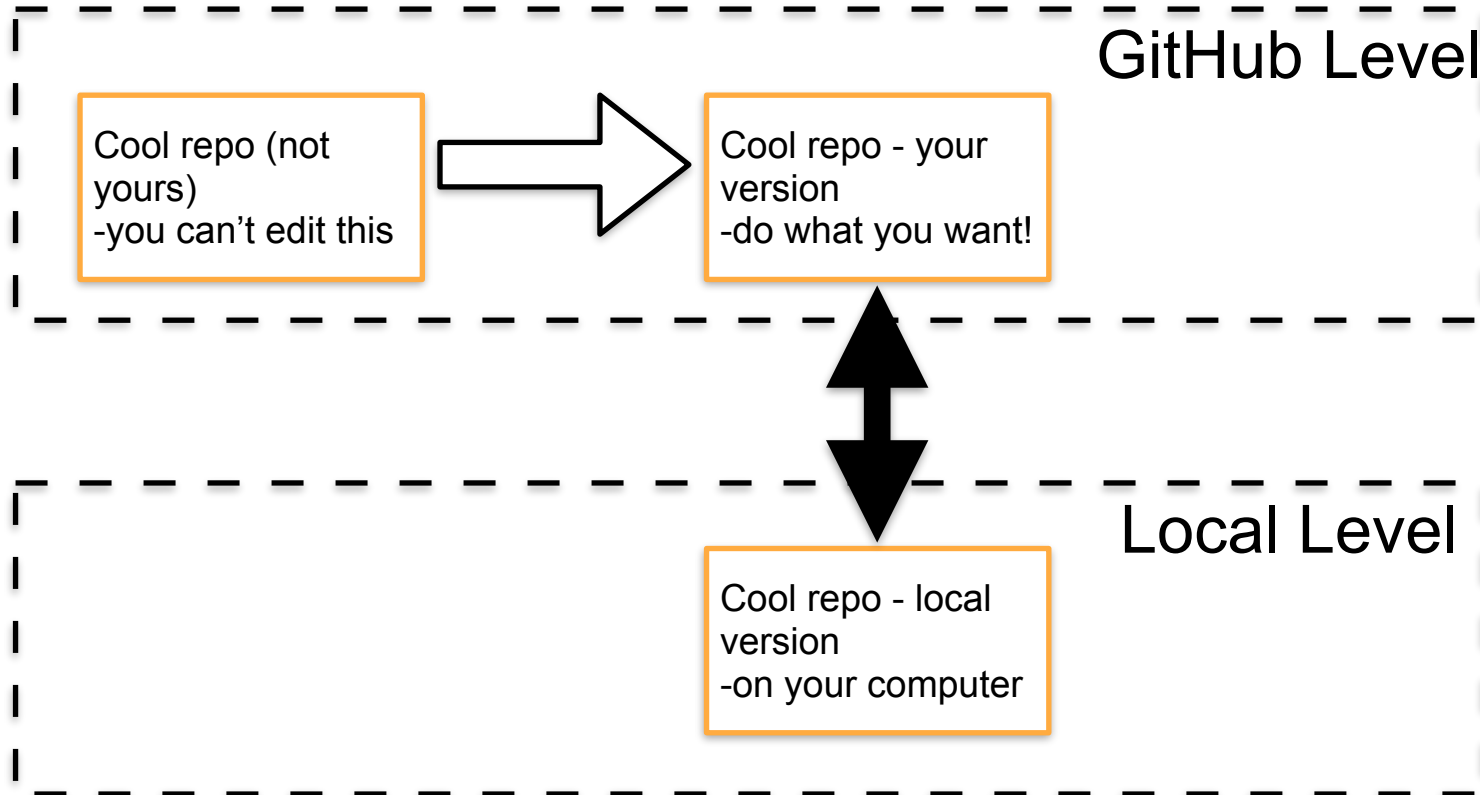
Course Reminders

- Due Friday 7/14 (11:59 PM)
  - D1, D2, Q1
  - Previous project review
  - Project proposal

Git discussion about Forking

- Some confusion on it, let's discuss, TAs are preparing some tutorial type links and resources
- Can't edit this but you can see it.
  - You can later ask them to incorporate changes you make to a clone you do locally, but you can't just change this one since it belongs to them
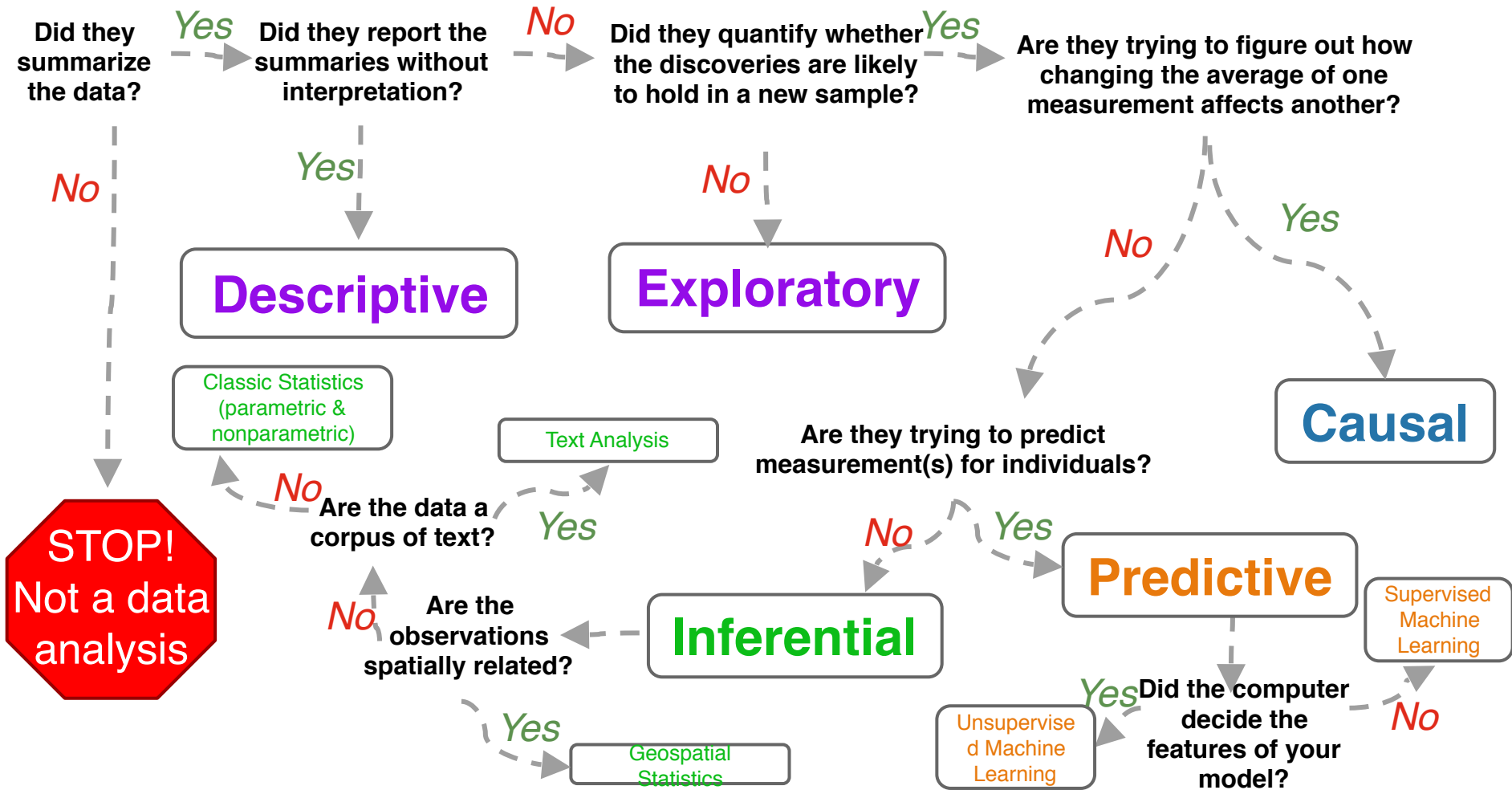 - What do we do if we want to use github to make our own version with version control?

Cool repo (not yours)
-you can't edit this

Fork it!

GitHub Level

Cool repo (not yours)
-you can't edit this

Cool repo - your version
-do what you want!

Local Level

Cool repo - local version
-on your computer

"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?

No ↓ (from "Did they summarize the data?")

Yes ↓ (from "Did they report the summaries without interpretation?")

Descriptive

No ↓ (from "Did they quantify whether...")

Exploratory

No → (from "Are they trying to figure out how changing...")

Yes → **Causal**

Classic Statistics (parametric & nonparametric)

Text Analysis

Are they trying to predict measurement(s) for individuals?

STOP! Not a data analysis

No → Are the data a corpus of text?

Yes → (to Text Analysis)

No → (to Are they trying to predict)

Yes → **Predictive**

Supervised Machine Learning

Are the observations spatially related?

No → (to Are the data a corpus of text)

Inferential

Yes → **Predictive**

Yes → Unsupervised Machine Learning

Did the computer decide the features of your model?

No → Supervised Machine Learning

Yes → Geospatial Statistics

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 the likelihood of causality

   a. Remember you *cannot* establish causality with computational methods only correlations along with statistical beliefs

   b. More you are establishing if they are NOT related or 'NOT NOT' related

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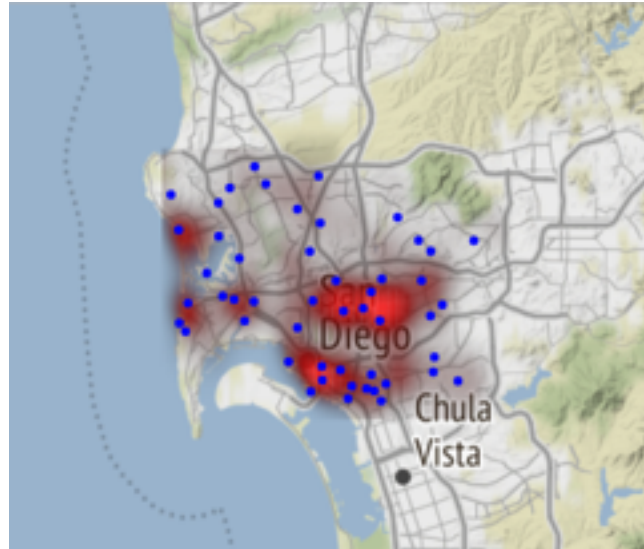
General question: Why isn't police response time always the same?

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?
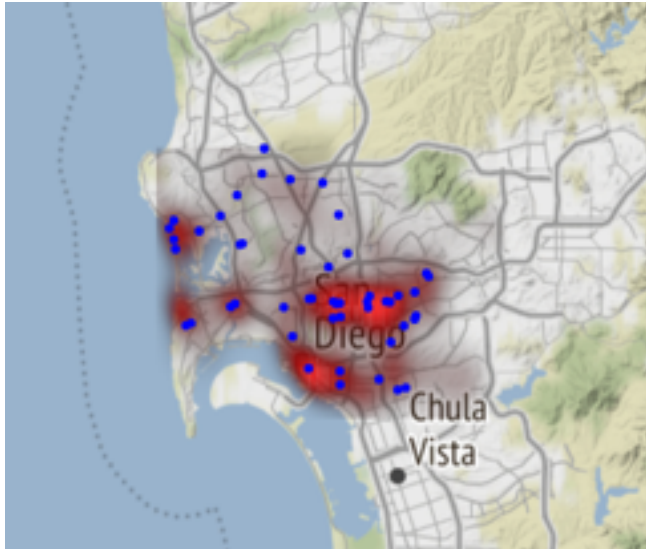
Descriptive

Exploratory

Predictive

Inferential

Geospatial Analysis

General question: 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?
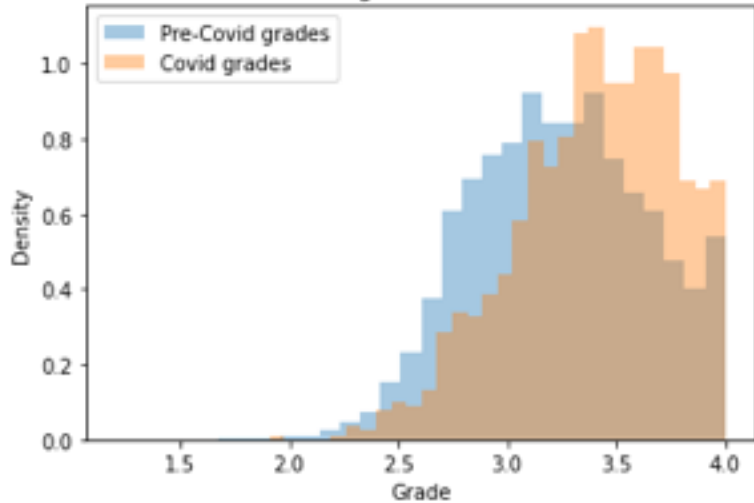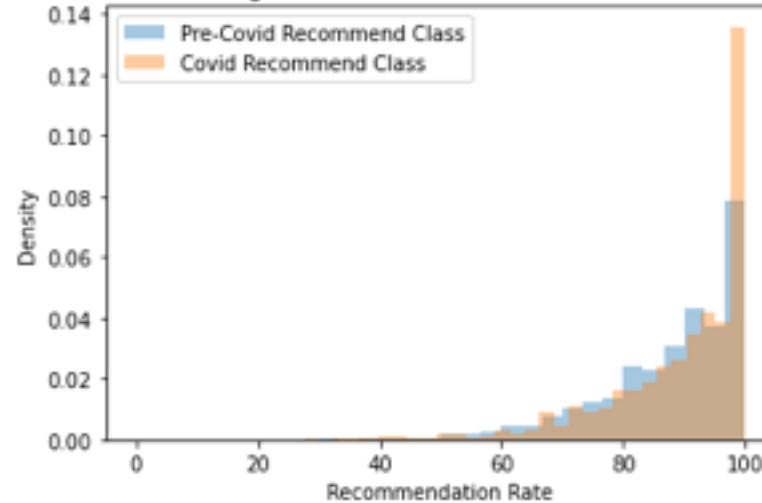

Exploratory
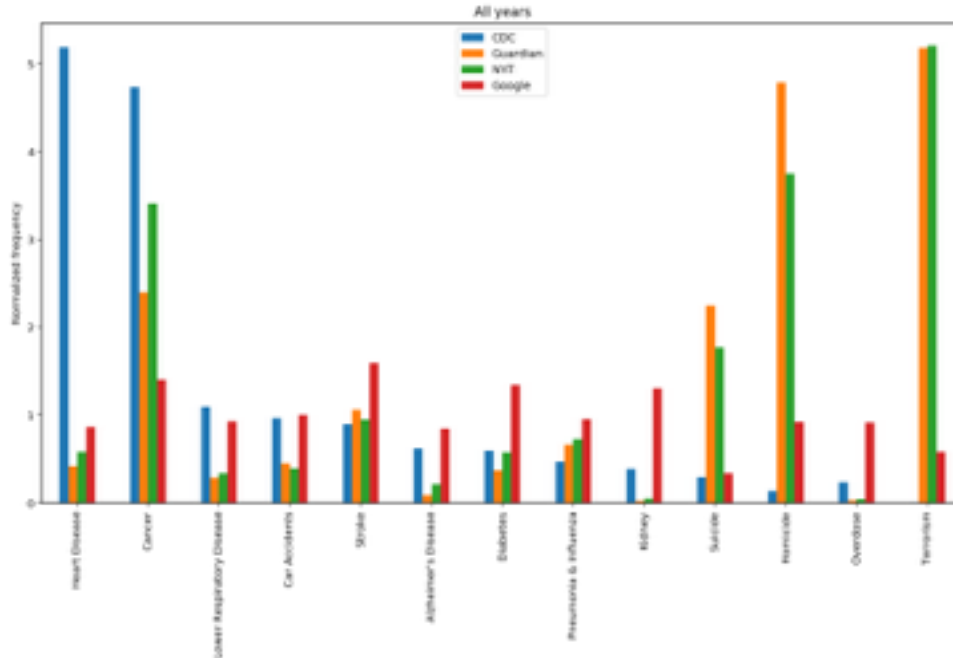
Predictive

Inferential

Classical Statistics

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?
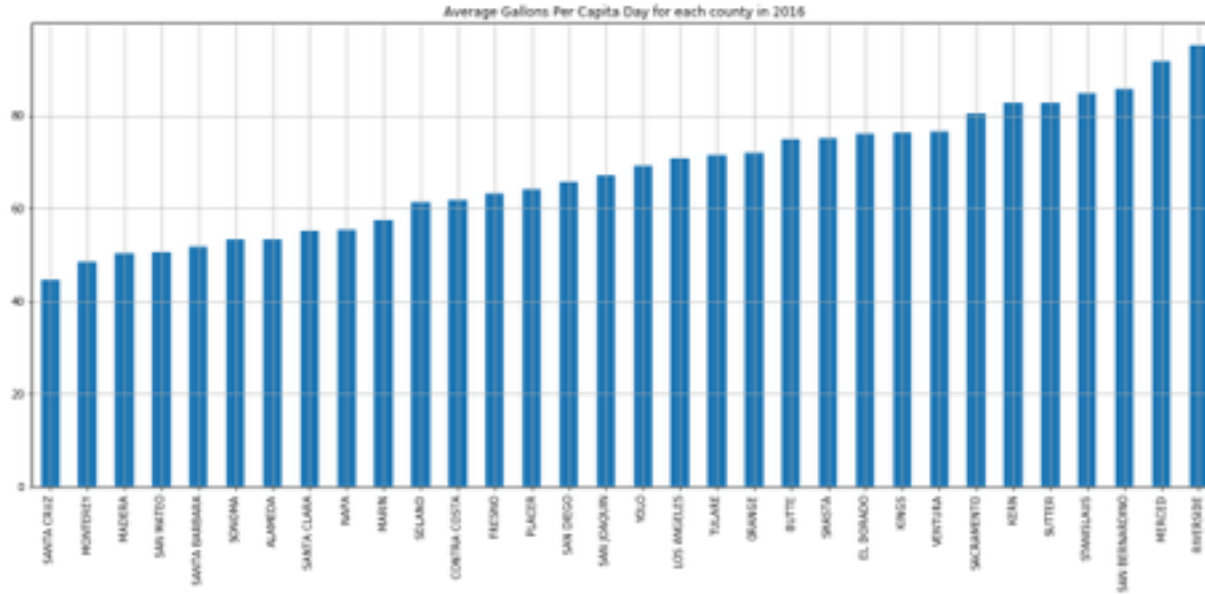

Descriptive

Exploratory

Inferential

Text Analysis

Classic Statistics (parametric & nonparametric)

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

Average Gallons Per Capita Day for each county in 2016

Descriptive

Exploratory

Predictive