S&DS 361 - Data Analysis

While we are chatting

```
install.packages("dplyr")
install.packages("ggplot2")
```

And if those finish

```
install.packages("tidyverse")
```

Intro

- Introduce self
- Student introductions with neighbors
- Introduce course

```
install.packages("dplyr")
install.packages("ggplot2")
install.packages("tidyverse")
```

What is Data Analysis?

Wikipedia: Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, and is used in different business, science, and social science domains.

ChatGPT: Data analysis is the process of examining, cleaning, transforming, and modeling data with the goal of discovering useful information, drawing conclusions, and supporting decision-making. It can be applied to various fields, such as business, science, and social science, and can involve a variety of techniques, including statistical analysis, machine learning, and data visualization.

Common steps when analyzing data

- Determine the problem or goal
- Acquire data, cleaning, organizing, merging data
- Explore, visualizing, summarizing data
- Analyze data, building models
- Interpret the results
- Communicate your conclusions and recommendations
- Implement the recommendations
- Reassess and repeat

Course Objectives

- Ask interesting questions and develop a strategy to answer those questions using data.
- Build comfort with coding and working hands-on with data.
- Get experience with interactive data exploration and visualization
- Learn statistical modeling with regression, generalized linear models, mixed effects regression and other topics

Syllabus

Data

- Water usage
- Basketball, hockey, other spots
- Census
- EV charging stations
- Attitudes towards climate change survey
- Property values
- Satellite
- And others, including. . .
- Your ideas
  - Something that interests you and you think would interest a lot of students.
  - Especially something timely, about current events.
  - I'm not opposed to postponing a lesson in order to work on a data set that is timely in some way.
  - Summarize them in Ed

Technology prep

- Software prep in pset 0.
- Clone the repo for this course. Maintain same folder structure as Canvas.
- Knit settings. Preview In Viewer Pane.
- Knit directory, Project Directory
- If you have R Studio issues throughout the semester: R Studio Cloud

Rest of Week 1

- Co-DUS. Primary advisor for class of 2026 (sophomores).
- GPT 4 Walled Garden
- Github copilot
- Notes
- Github usages throughout the semester
- Active Learning, Visualizations
- Other visualization examples
- Water