Section Five

Class COGS 9
Teaching Assistant: Matthew Feigelis

Background

- Final Project Part 1 due Friday!

- Today

  - Review Common Faults on Part 1

  - Then, Open Office Hours // Group time // Q/A

Part 1 Sections

- Research Question

- Hypothesis

- Background

- Data

- Ethical Considerations

Research Question

- **Prompt:** Clearly state the specific data science question you're interested in answering.

- **Common Issue To Avoid:**

  - Not specific enough! Leading to a lack of clarity on what's being measured.

  - <u>Example</u>: "Does happiness relate to wealth?"

What are the top five risk factors that contribute to the development of Alzheimer's Disease, ranked from largest risk to smallest risk?

Hypothesis

- **Prompt**: Write down your group's hypothesis to your question. Provide justification how you came to this hypothesis

- **Common Issues To Avoid:**

  - 1. Not specific enough about direction of relationship. 2. Not providing justification

  - Example: "I predict happiness and wealth are related"

We predict that the 5 most influential risk factors for Alzhemier's Disease ordered from most influential to least influential are

1. Age 2. Head trauma 3. Genetics/inheritance 4. Comorbidities 5. Diet

We think this because …

Background

- **Prompt:** Provide a few paragraphs of background research and information on your topic. This should include at least 2 citations to work from others

- **Common Issue To Avoid:**

  - Not specific enough what the authors did

  - **Example:** Stating "Author 1 and author 2 did a study and found GDP was related to happiness"

  - Provide information on what the study did, what they find, and why it relates to your research question.

According to Alzheimers association, some risk factors researchers have found include non-modifiable risk factors such as age, family history, and genetics as well as modifiable risk factors such as cardiovascular health, diet, and a patient's exposure to traumatic brain injury. They found this by sampling 100,000 patients with Alzheimers across USA … (Citation here)

Data

- **Prompt:** Include a description of the perfect dataset you would need to answer this question. How many observations would you need? What variables would you collect? Give specific details about the observations and variables. Then, look online for available datasets. Again answer: How many observations are there? What variables did they collect? Then, discuss the datasets limitations, and how it differs from the ideal dataset described in part one.

- **Common Issue(s) To Avoid:**

  - Not providing the *IDEAL* dataset

  - Not describing the details of observations and variables.

    - **Example: "**The observations are 100 people". But not stating information about these people (e.g. randomly sampled across the united states, age 18-35 …)

What is the ideal data set for: Does happiness, as measured by survey self-report, relate to wealth, as measured by household income, across the United States?

Ethical Considerations

- What ethical considerations must be made when answering your specific data science question? Look over Deons data science checklist; how would you address these categories for your specific research question: Data Collection, Data Storage, Analysis, Modeling, Deployment

- Common Issue(s) To Avoid:

  - Restating bullet points in Deon's checklist without relating it to your specific research question

  - **Example**: Stating "We will make sure we don't have collection Bias by checking on sources of bias introduced during data collection and survey design"

What is the sources of collection bias for the following research question: Does happiness, as measured by survey self-report, relate to wealth, as measured by household income, across the United States?

Group work // Q/A

(A03 discussion slides linked on class website)