

Research data group assignment 1

Jing Guo

Assignment: Data Project - Identify data source and formulating research questions

Objective:

To explore data sources, formulate research questions, and reading articles to derive insights. You may also watch the videos to get an idea of the student data project here (<https://www.causeweb.org/usproc/projects/winners>).

Instructions:

- Select a Data Source:

Identify an available dataset that interests you. Consider sources such as: Government databases (e.g., data.gov) Academic repositories (e.g., Kaggle, UCI Machine Learning Repository, The Cancer Genome Atlas (TCGA)).

You may also use your own collected datasets.

- Formulate Research Questions:

Develop 3-5 research questions that you aim to answer through your analysis. Ensure that these questions are specific, measurable, and relevant to the dataset you have chosen.

- Data Variable Description:

Explain the variables in your dataset you intend to analyze.

- Preparation:

Describe how you did data preparation. This may include: - Checking and handling missing values - Encoding categorical variables