

PROPOSAL

Due: Sunday, November 17 11:59 PM

PURPOSE

The purpose of the team project is to provide an opportunity to complete a data analysis project from start to finish. This includes:

- Selecting data
- Writing research questions
- Performing exploratory data analysis
- Developing appropriate statistical models
- Communicating results through a written report and presentation
- Working with a team

TEAMS

See [this sheet](#) for your team assignment. You are responsible for dividing the work equitably among team members. Every team member is required to contribute to each portion of the project. Learn from each other and, most importantly, be kind. You will receive individual grades for the project that incorporate feedback from fellow team members.

DATA

Your first step is to select the dataset that you are most interested in analyzing. Below is a list of datasets or sources that you can consider using to find a dataset. You are not required to use these.

Your dataset must have at least 500 observations, 10 variables, and a mix of numeric and categorical variables. You may not use a dataset that we have used in class.

1. [R Data Sources for Regression Analysis](#)
2. [FiveThirtyEight data](#)
3. [World Health Organization](#)

4. The National Bureau of Economic Research
5. International Monetary Fund
6. General Social Survey
7. United Nations Data
8. Pew Research
9. World Inequality Database

You are welcome to specify a subset of a particular dataset. Be sure to look at the data to understand its scope.

PROPOSAL

Use this template for your proposal

Your proposal should list the following:

1. Team member names
2. 2 or 3 datasets of interest (list in order of preference with 1 being the top choice). For each dataset, provide:
 - The source of the data, when and how it was originally collected, and a brief description of the observations (i.e., what does each row represent?). This information is likely found on the website where you found the data, so be sure to cite your source
 - TWO research questions you are interested in exploring for each dataset. Explicitly state the outcome variables in the dataset that you will use to answer each question. You are required to use two different types of outcome variables for your research questions. For example, one research question may use a continuous outcome variable while the other uses an ordinal outcome variable. Variable types may include: continuous, binary, ordinal, nominal, time-to-event. The research questions must be related to inference, not prediction (though you will be able to assess predictions in your final report). Consult [this helpful guide](#) for writing a good research question and [this blog post](#) about the distinction between inference and prediction. Additionally, you will be required to incorporate at least one interaction term into one of your models, so you might consider incorporating this into your research question.
 - a `glimpse()` of each dataset to show the variables
 - Brief exploratory data analysis. The EDA must not exceed 2 pages (1 for each research question) and should include:
 - One plot for each of your outcome variables by themselves. Be sure to use appropriate axis labels

- exploratory plots for your primary relationship of interest (dependent variable and primary independent variable, if applicable). Whether you have one primary independent variable of interest or multiple will depend on your research question. Be sure to use appropriate axis labels

4. The proposal includes a team charter. Discuss the provided questions with your group and include your agreed-upon answers.

Use github to share code among group members. As a group, plan the tables and figures you want to generate, and split them up among the group members. Then you can consolidate your code to generate the report.

Submit one proposal per group. One person will submit and select the other group members in the Gradescope submission. After you submit, I will provide feedback to help you decide which dataset to choose for your project report.