Final Project Requirements

For the final group project, you get to pick your own dataset. First, I will provide some sources where you can find datasets online (but you may use any source you want for finding datasets). Then, I will discuss requirements for the dataset, as well as guidelines on the type of graphs that you should make for the project. These requirements and guidelines will help you decide (as a group) which dataset to choose for your project. The dataset you choose should be something the group is excited about, because you'll be working with this dataset for the remainder of the class. Your project will culminate into a short in-class presentation (during the final week of class) and a public-facing HTML file (due Monday April 28th).

Sources

Here are some repositories with many, many datasets to choose from (as well as links to other data sources):

- RStudio #TidyTuesday Project
- Kaggle
- Data is Plural
- FiveThirtyEight
- SCORE Sports Data Repository
- Carnegie Mellon Data Repository
- The US Government
- Google Public Data Explorer
- UCI Machine Learning Repository
- More links on GitHub

You can use other sources, these are just suggestions.

Data Requirements

1. Your data must be contain a mix of categorical and quantitative variables and be complex enough that you can create at least **at least six or eight interesting graphs depending on your group size** (groups of size 3 must make at least 6 graphs, while groups of size 4 must make at least 8 graphs). So datasets with only a few variables will not work.
2. You CANNOT use any of the datasets that were used in any previous assignments in this course or any other course you have taken. You must use a dataset that everyone in your group has never worked with before.

Graph Requirements

In the group project itself, you will have some requirements on the graphs you must use in your presentation and report.

Report Requirements

For the report, you are required to make **at least six or eight** graphs (depending on your group size) with the following requirements:

- **at least two graphs** for each research question (you will have three to four research questions in total, depending on your group size, with details covered in the report and presentation rubrics),
- **at least three** graphs from the first half of the semester (Weeks 1-7)
- **at least three** graphs from the second half of the semester (Weeks 8-14)

In other words, you can NOT just turn in six or eight 'basic' EDA graphs. You are required to make at least three graphs that we will be learning about in the second half of the class.

You are only allowed **a maximum of two** 1D graphs (i.e., you are only allowed two graphs that display a single variable such as a 1D histogram or 1D bar chart).

Note: a facetted graph counts as one graph, and NOT as the number of facet panels.

Your graphs should tell a somewhat cohesive story. Come up with some general questions you want to answer with your dataset. **You should be able to come up with at least three or four interesting, overarching questions for your dataset.** You'll use your graphs to walk the viewer through a comprehensive analysis of those questions, as well as demonstrate your findings and conclusions.

Presentation Requirements

For the presentation, each member of your group is required to make a **single graph** with the following requirements:

- **Your graph must account for at least two variables from your dataset**, i.e., you cannot just present a 1D graph.
- **Your graph must address one of your group's research questions and must be the only graph addressing that research question in your presentation**, i.e., each member is uniquely responsible for a single research question.
- **Your graph must contain at least one unique variable that is not featured in the graphs presented by your group members**, i.e., every member of your group will be presenting graphs that contain a distinct variable from the rest of the group (this should be easy since each graph is directed towards a different research question).
- **Your graph must be a different type of graph from the others in your project**, i.e., every member of your group must present a unique type of graph.
- **You will be required to use some form of text / annotation in your graph**, we will discuss how to do this later in the course.

Remember, **the presentation portion of the final project is graded individually**.

The graphs that each member of your group makes for the final presentation can (*and should*) be included in your report! You are not required to include the presentation graphs in your report, but I would strongly encourage you to take advantage of double-dipping in this case for your project.