Data Exploration and Visualization

Step	Grade %	**Due Dates**
Project Proposal	20%	October 6th
Data read-in	10%	October 20th
Presentations	70%	November 10th

Overview

This midterm project is meant to do several things. Among them is to have you apply the techincial skills you have built up to this point. To that end in this project you will be responsible for obtaining your own data and producing high quality graphics from it that can tell a story. It will culminate in a brief presentation of your data with accompanying graphics. The project is broken into two check-ins and then the midterm presentation. In total this is worth 10% of your grade.

1 Data

You are encouraged and required to find your own data. The requirements I have for your project is that you have...

- 2-3 memebers
- At least one categorical variable
- At least two numeric variables
- All your variables you should believe a priori are related
- You have a coherent research goal
- I do NOT want projects that deal with sleep and GPA. I've grown nauseous from the lack of creativity of some students so these are non-starters.
- You will need AT LEAST 40 observations
- Overall I'd prefer you run experiments if possible as they lead to generally stronger results and conclusions (and I study them) but I will not deduct points for choosing to do a survey if it is done well.
  - Done well? Survey's shouldn't be made up of your friends/roommates/etc... ("convenience" sampling isn't random enough for statisticians)
  - Need to find a way to sample people while mitigating lurking variables
    - Weighing athletes vs non-athletes is valid...unless you are only weighing people who are at the gym already. Either find a different way to sample or change your research question to be gym-going athletes vs gym-going non-athletes
  - The quality of the survey is generally a reflection of how randomly the population was sampled from and avoiding sampling only from subpopulations
  - Warning: Survey participants have notoriously bad memories (eg What did you wear yesterday? is more difficult to answer than it should be for most people)
- Avoid surveys with open ended questions. Instead use questions with limited response choices
  - WRONG: "What's your favorite ice cream?" will get 23 different responses out of 25 students
  - BETTER: "Do you prefer chocolate, vanilla, or strawberry ice cream?"
- If the data you use is from another class, please indicate that to me as I will need to touch base with the other professor to make sure they are alright with you using data generated for them for this project.

1.1 Examples of Projects from the Past

A couple of fun projects that come to mind are...

- What is the relationship between the time someone claims they can plank and how long they can actually plank. Does the relationship depend on them being an athlete?
- How many standard drinks are students consuming over the course of a weekend? Does it matter if they are in a frat/sorority?
- If students are presented with a list of random words, will they do better memorizing it by typing out the list or writing the list out by hand?
- What is the relationship between being raised with a measurement system (metric, customary) and the quality of estimating something's length with the opposing system?

1.2 Deliverable

On a single page to be turned in...

1. The group member's names
2. The proposed research goal/topic
3. A brief description of the variables you plan to collect
4. How you plan to collect the data
    - Both on the abstract ("We will survey fellow students")
    - and the technical ("We will randomly select 5 tables at the dining hall at a randomly selected time and survey one randomly chosen student at each table. We will repeat this on six separate days.")
5. What a priori beliefs you have about the data you may be interested in checking?
    - eg I suspect students who were raised with the customary system will do better at guessing distance in metric units than students raised with the metric system do at guessing distance in customary units.
    - eg We do not believe that the respondent's sex will be correlated (predict, have a relationship) with their perceptions of violent crimes in their areas

1.3 Grading

You will be graded on completion of this step; including all the parts above. You are STRONGLY encouraged to talk to me beforehand about your project ideas.

2 Data Check-in

In the middle of October I will be requesting a summary of where you are at with your project. In particular I want to see that you have successfully read in the data (or are in the process of reading in the data) you have already collected (required) and are working on exploring it (would be nice). This step is to make sure that you have made forward progress. Struggling at this stage is both common and can consume large quantities of time. If this is not done sufficiently far out from the presentation and deliverables things can be rough.

2.1 Deliverables

1. Your data saved in an easy to read format
    - eg .csv, .Rds, or an excel file

2.2 Grading

You will be graded on completion of this step. Note it is alright to still be in the process of collecting data but I want to make sure some data has been collected and progress is being made.

3 Presentations

The presentation will be capped at five minutes and one slide. The slide should contain predominately graphics. The graphics presented can be chosen because they were something you found interesting or because they help tell the story the data has to tell.

- You will most likely have to describe the research goal/question to the audience as well as the variables you collected verbally during the presentation. Practice this.
- Your target audience for the talk is your peers; intelligent, bright but maybe not familiar with what you were up to for your project
- The goal is for you to make a publication ready graphic
- Sign-ups will be posted closer to the schedule times

3.1 Producton of Graphics

I have not had time in this class to go over coding in R or how to build a high quality graphic yourself. You have the theory, can graph by hand, and, as indicated by your test scores, can interpret them just fine. As such, this is the only time in the course (or my teaching career) where I am giving the green light to use AI. Note that there are several points I want respected....

- At least verbally in the presentation I want you to give reference to which AI you used (e.g. "We created this scatterplot using ChatGPT"). In the final paper you will be required to formally cite the AI.
- You are fully responsible for making sure the graphic produced correctly graphs your data (i.e. the computer isn't plotting points randomly without connection to your data file). Any observation in your data has a corresponding point on the graphic.
- Because you are outsourcing the production of the graphic I expect it be all the higher quality (see subsection 3.3 for grading details)
  - I'm a virtual luddite so I'm unsure how AI works but I suspect this will be an iterative process to get the graphic you want. By that I mean you will probably need to keep adding detail/giving instructions to the AI to get it perfect
  - (I wasn't even sure AI could take in data and graph it to be honest until I checked ten minutes before writing this)
- The use of AI is strictly limited in this project to this single step (creating a publication ready graphic).
  - Again, you wouldn't take a robot to the gym to lift weights for you so why use AI when it comes to exercising your mind in an academic setting?
  - AI is being used in this step as a labor saving device and not a replacement to learning. I firmly believe you already have learned statistical graphics, this is a bridge to go from theory to high-end publication ready graphics

3.2 Deliverables

1. 1 slide presentation
   - A slide that tells a story the data has to tell
   - The slide should convey something of scientific interest mostly through graphics.
   - Things like summary tables or cross tables can be acceptable if used parsimoniously.

2. Being present when others are presenting.
    - We are doing presentations before the final report is due so you have time to tweak things or improve things. If you are not paying attention/not engaging with the speakers then you are making that harder on yourself and others.
    - There may be feedback surveys to presenters that will be filled out

3.3 Grading

- Presenting
    - Was there sufficient information presented such that the audience can follow what is being discussed? Is there ambiguity in what was studied?
    - Did you have a clear story/result that you conveyed? Or was it just a "Here's our data"?
    - Did you speak clearly and loudly enough?
    - Was there "pizzazz"?
    - Did you act professionally? Did you avoid f-bombs?
        * A little disappointed I need to mention this specifically to be honest
- Conveying Information
    - Was the type of graph an acceptable choice for the data presented? Was it a good choice?
    - Did the presentation allow for quick interpretations? Or did it take a while for the audience to understand what the graph had to say?
    - Did the presentation convey information easily? Or was the graph confusing and hard to read?
    - Does the graphic address your research question? Did it answer it? NOTE: Being right or wrong on your a priori beleif is irrelevant; the point is that you used the data to answer the question graphically.
- Labeling
    - Axis labelled well?
    - Title and subtitle?
    - Were the labels useful/informative/interesting labels or were they "flat"?
    - Could they be read by the audience?
    - If there is a legend, are the different parts clearly labelled (eg is it clear the GDP of the US is the red, dotted line?)
- Accessibility?
    - Was the graph colorblind friendly?
    - If needed, was there redundant coding?
    - Was the text made readably large?
    - Was it generally accessible?
- Aesthetics?
    - Did the audience enjoy looking at your presentation?
    - Was there some amount of entertainment in the graphics?
    - Do the aesthetic decisions that were made still allow for accessibility? Are they balanced well (if needed)?
    - Were the aesthetic decisions well tied to the research topic, if applicable? (eg Rainbows for LGBT stuff is good, Rainbows for political demographics of Iowa is probably less justified)
- Listening?
    - Were you a polite audience member?
    - Were you an attentive listener for other students?
    - Were you physically present??
    - UNDECIDED: Critiquing/reviewing a few of your other classmates in a survey form might be required. I'm undecided.