Project description

Introduction

TL;DR: *Ask a question you're curious about and answer it with a dataset of your choice. This is your project in a nutshell.*

May be too long, but please do read

The project for this class will consist of analysis on a dataset of your own choosing. The dataset may already exist, or you may collect your own data using a survey or by conducting an experiment. You can choose the data based on your teams' interests or based on work in other courses or research projects. The goal of this project is for you to demonstrate proficiency in the techniques we have covered in this class (and beyond, if you like) and apply them to a novel dataset in a meaningful way.

The goal is not to do an exhaustive data analysis i.e., do not calculate every statistic and procedure you have learned for every variable, but rather let me know that you are proficient at asking meaningful questions and answering them with results of data analysis, that you are proficient in using R, and that you are proficient at interpreting and presenting the results. Focus on methods that help you begin to answer your research questions. You do not have to apply every statistical procedure we learned. Also, critique your own methods and provide suggestions for improving your analysis. Issues pertaining to the reliability and validity of your data, and appropriateness of the statistical analysis should be discussed here.

The project is very open ended. You should create some kind of compelling visualization(s) of this data in R. There is no limit on what tools or packages you may use but sticking to packages we learned in class is required. You do not need to visualize all of the data at once. A single high-quality visualization will receive a much higher grade than a large number of poor-quality visualizations. Also pay attention to your presentation. Neatness, coherency, and clarity will count. All analyses must be done in RStudio, using R, and all components of the project **must be reproducible** (with the exception of the presentation).

You will work on the project with your lab teams.

The four milestones for the final project are

1. Milestone 1 - Working collaboratively
2. Milestone 2 - Proposals, with three dataset ideas
3. Milestone 3 - Peer review, on another team's project
4. Milestone 4 - Presentation with slides and a reproducible project writeup of your analysis, with a draft along the way.

You will not be submitting anything on Gradescope for the project. Submission of these deliverables will happen on GitHub and feedback will be provided as GitHub issues that you need to engage with and close. The collection of the documents in your GitHub repo will create a webpage for your project. To create the webpage go to the Build tab in RStudio, and click on Render Website.

Milestone 1 - Working collaboratively

For the first milestone of your project you'll practice a collaborative Git workflow with your team members. Instructions are outlined in Milestone 1: Working collaboratively. Each team member taking part in the collaborative working activity will get 5 points towards their project.

Milestone 2 - Proposal

There are two main purposes of the project proposal:

- To help you think about the project early, so you can get a head start on finding data, reading relevant literature, thinking about the questions you wish to answer, etc.
- To ensure that the data you wish to analyze, methods you plan to use, and the scope of your analysis are feasible and will allow you to be successful for this project.

Instructions and grading criteria for this milestone are outlined in Milestone 2: Proposal.

Milestone 3 - Peer review

Critically reviewing others' work is a crucial part of the scientific process, and STA 199 is no exception. You will be assigned two teams to review. This feedback is intended to help you create a high quality final project, as well as give you experience reading and constructively critiquing the work of others.

Instructions and grading criteria for this milestone are outlined in Milestone 3: Peer review.

Milestone 4 - Writeup and presentation

Instructions and grading criteria for this milestone are outlined in Milestone 4: Writeup + presentation.

Other

Reproducibility + organization

All written work (with exception of presentation slides) should be reproducible, and the GitHub repo should be neatly organized.

Points for reproducibility + organization will be based on the reproducibility of the write-up and the organization of the project GitHub repo. The repo should be neatly organized as described above, there should be no extraneous files, all text in the README should be easily readable.

Teamwork

You will be asked to fill out a survey where you rate the contribution and teamwork of each team member by assigning a contribution percentage for each team member. Filling out the survey is a prerequisite for getting credit on the team member evaluation. If you are suggesting that an individual did less than half the expected contribution given your team size (e.g., for a team of four students, if a student contributed less than 12.5% of the total effort), please provide some explanation. If any individual gets an average peer score indicating that this was the case, their grade will be assessed accordingly and penalties may apply beyond the teamwork component of the grade.

If you have concerns with the teamwork and/or contribution from any team members, please email me by the project presentation deadline. **You only need to email me if you have concerns.** Otherwise, I will assume everyone on the team equally contributed and will receive full credit for the teamwork portion of the grade.

Grading

The grade breakdown is as follows:

Total	100 pts
M1: Working collaboratively	5 pts

Total	100 pts
M2: Project proposal	10 pts
M3: Peer review	5 pts
M4: Write-up	40 pts
M4: Slides + presentation	25 pts
Reproducibility + organization	5 pts
Teamwork	10 pts

Grading summary

Grading of the project will take into account the following:

- Content - What is the quality of research and/or policy question and relevancy of data to those questions?
- Correctness - Are statistical procedures carried out and explained correctly?
- Writing and Presentation - What is the quality of the statistical presentation, writing, and explanations?
- Creativity and Critical Thought - Is the project carefully thought out? Are the limitations carefully considered? Does it appear that time and effort went into the planning and implementation of the project?

A general breakdown of scoring is as follows:

- *90%-100%*: Outstanding effort. Student understands how to apply all statistical concepts, can put the results into a cogent argument, can identify weaknesses in the argument, and can clearly communicate the results to others.
- *80%-89%*: Good effort. Student understands most of the concepts, puts together an adequate argument, identifies some weaknesses of their argument, and communicates most results clearly to others.
- *70%-79%*: Passing effort. Student has misunderstanding of concepts in several areas, has some trouble putting results together in a cogent argument, and communication of results is sometimes unclear.
- *60%-69%*: Struggling effort. Student is making some effort, but has misunderstanding of many concepts and is unable to put together a cogent argument. Communication of results is unclear.
- *Below 60%*: Student is not making a sufficient effort.

Late work policy

There is no late work accepted on this project. Be sure to turn in your work early to avoid any technological mishaps.