

Research Project Overview

Throughout the semester you will work with a co-author to analyze a data set and present the results of your research during finals week. Regular homework assignments will serve as a first draft of exploration into your research questions, and helps you build your story.

This project will be done in a few stages;

- Stage 1: Data and topic proposal
- Stage 2: Introduce and motivate your research
- Stage 3: Analyze relationships and draw conclusions
- Stage 4: Dissemination via research poster symposium

! Group project but individual responsibilities

This is a group project. Since we are not using version control like git/GitHub (you are welcome to if you want) you will have to collaborate closely to determine who will have the primary copy of the project template file that will be turned in.

Both authors are responsible for the content that is submitted, but only one submission should be made.

Stage 1: Data and Topic Proposal

Identify a research partner, propose a research topic and corresponding data set.

Start by looking through [Dr.D's curated data sets](#). There is documentation in each folder that tell you a little about the data set, and you can also ask our custom class research assistant LLM [Name TBD](#)

Other open source data repositories to consider:

- [Teaching of Statistics in the Health Sciences Resources Portal](#)
- [Data.gov](#)
- [IPUMS](#)
- [Teaching of Statistics in Health Sciences](#)
- [City of Minneapolis data](#) (includes police use of force)
- [Pew Research Center](#) is “a nonpartisan fact tank that informs the public about the issues, attitudes and trends shaping the world.”
- [Bureau of Labor Statistics](#)
- The [All of Us](#) Research Hub stores health data from participants from across the United States
- Other?

Sources not allowed: Machine learning repositories such as Kaggle or UCI. Data generated from experimental lab designs with few variables.

Criteria for choosing a data set

If you choose a data set that is not one of my in my approved list it must be pre-approved and meet the following criteria:

- You either know something about the topic or it is something you want to learn about
- File type must be a .txt, .csv, .xlsx or .xls file
- File size is less than 1 Gig
- A codebook or data dictionary that fully explains what each variable means is available.
- There are at least 200 rows (observations), but ideally between 500-10,000.
- There are 10 or more *unique* and interesting variables
 - At least 4 quantitative variables
 - Variables are not functions of each other (e.g. weight in lbs and weight in kg)
 - Unique identifiers, dates, addresses, lat/long and other non-analyzable columns do not count

Your data set must be approved before you are allowed to work with it. Add your team's names and your chosen data set to the [Project Stage 1](#) spreadsheet in our shared class Google Drive folder.

The sooner your data is approved the sooner you can work with it! If your proposed data set is turned down twice in a row get turned down, you will be required to use one of Dr. D's curated data sets.

Stage 2: Introduction and Motivation

In this stage you will identify a few research questions, ground them in prior scientific evidence and start to write an introduction to your research project.

1. A project template file has been provided for you in the **shared/Donatello Math 315** folder in RStudio.
 - You could also download it from [this link](#), and then rename it to remove the _ in the front.
2. Complete the following sections using your Homework [2](#), [3](#), and [4](#) as guidance.

Example

The template is very bare bones and has no instructions. See [this example](#) of a project after Phase 2 is completed.

Use section headers

The section headers in the template are an important part of your project organization. Do not delete them. However **DO** update generic portions like “variable 1” and “RQ1” to your specific project details.

Introduction

- **Background & Significance:** Use your topic area statement and literature synthesis from Homework 2 to introduce the general topic of your project and why it matters. Do not just copy and paste your entire assignment but consider any feedback that was given and summarize the information into the appropriate sections. This should be written in plain language and should help the reader understand the real-world issue behind your research question.
- **References:** List the data citation and all sources cited in the report using a consistent citation style. This is at the bottom of the template, but done during this stage.
- **Research Questions:** List all four research questions. Primary RQ’s go as #1 and #2, Secondary go as #3 and #4. If these questions have drifted since HW 2, briefly explain your reasoning.

Methods

- **Study Design & Data Source:** Describe the data you are using: where you got it from (with citation), the population being studied, the unit of observation, and whether the study is observational or experimental.
- **Variables of Interest:** List every variable used in the four research questions as a bullet point list. Describe what each one measures. Identify the data type, if it’s a response or explanatory variable, and for which research question.
- **Data Management:** Copy the relevant code from Homework 3 to read in your raw data, and select and clean *only* the variables mentioned above. Put all code into one code chunk, but use # comment to annotate the main pieces of code explaining what they are doing. Do not include code that produces output like `class`, `table` or `summary`. Do not export a clean data set. Below this code chunk explain using bullet points each major cleaning step in words.

Sample Characteristics

- Describe the distribution of all variables analyzed in this research. Use the same style as Homework 4 (plot, summary statistics and descriptive paragraph).

Stage 3: Analysis and conclusions

Continue filling in the same template using Homework 5-9 as a guide. Note that if you decided to change variables since Stage 2, you will need to update appropriate prior sections.

Results

- **Relationship Between Variables:** Visualize and describe the relationship between your primary response and explanatory variable, using the same style as Homework 5.
- **Statistical Analysis:** Formally test this relationship. Include your statistical method, results, and model assessment, using Homework 6-9 as a guide.
- **Discussion & Conclusion:** Interpret your results in plain language, connect them back to your research questions and the literature findings and note any limitations of your analysis.

Stage 4: Dissemination via Research Poster

During week 14 and 15 you will transfer all findings into a research poster, print the poster, and then present your research to your classmates during our class final period in a poster symposium format. - Full guidelines including examples and evaluation criteria [are written in this blog post](#). - Submit the poster file as printed to Canvas by the due date.

Draft version

This draft is graded based on how complete the poster is. You should consider this a draft that you would circulate to your colleagues for final review and comments. There is a rubric in Canvas with details on grading criteria and submission instructions.

Peer Review

You will be asked to review one other team's poster.

Final Version

Make any necessary revisions, then upload your final poster as it is printed in PDF format Canvas.

Presentation at the Poster Symposium

This will be a joint poster symposium with the other section of Math 315. When not presenting, you will walk around and learn about others research. Ask the presenters questions and fill out an evaluation form as you go. Poster scoring follows the above evaluation criteria and will be done via Google Forms. The link to this semesters form is in Canvas. Printed copies will be available upon request.