

Final Project

✓ Overview

For your final project, you will have the opportunity to explore a dataset of your choice and answer a question or analyze an idea. To do this, you will need to use R to create four figures and supporting text that clearly explain your findings.

Here are some examples of things you might do or think about when refining your question/idea. You don't need to do all of these and/or you might have some ideas of your own!

- You do not need to look at the whole dataset -- you could pick out two or three variables that are the most interesting to you.
- You do not need to look at the whole dataset -- you could choose to look only at certain years, countries, sports, genres, etc.
- What is the difference between the groups?
- What is happening over time?
- How are two variables related?
- What are the "Top 10"? What are the "Bottom 10"?
- What summary statistics could you calculate? Mean? Median? Min? Max? Variance?
- You might find some inspiration in [the charts in the R Gallery](https://r-graph-gallery.com/all-graphs.html) (<https://r-graph-gallery.com/all-graphs.html>).

Your work should aim to make the content accessible for a broad audience who may not necessarily be familiar with this dataset or R.

Datasets (Your Choice)

You are free to choose your own dataset, but you can find inspiration in sources such as:

- Kaggle (<https://www.kaggle.com/datasets> (<https://www.kaggle.com/datasets>).
- The "OpenIntro" Package has a bunch of datasets (<https://www.openintro.org/data> (<https://www.openintro.org/data/>).
- The "Lock5Data" Package has a bunch of datasets (<https://cran.r-project.org/web/packages/Lock5Data/Lock5Data.pdf> (<https://cran.r-project.org/web/packages/Lock5Data/Lock5Data.pdf>).
- <https://www.data.gov/> (<https://www.data.gov/>).
- Google Dataset Search: <https://datasetsearch.research.google.com/> (<https://datasetsearch.research.google.com/>).
- WHO Data: <https://apps.who.int/gho/data/node.home> (<https://apps.who.int/gho/data/node.home>).
- Web scraping from any reliable source on the web
- Downloading dataset(s) from any reliable source on the web

Requirements and Challenges

Requirements (your project should satisfy *all* the criteria below)

- Submit a .Rmd file, a .html file (where all the code as been hidden), and any raw data (.csv files, etc).
- An appropriate title that describes the main focus of your project, along with your name at the top of the project
- Provide an "Introduction" section that briefly describes the data (variables, sources, etc.) and explains your intention with the project (what questions do you hope to answer, and why is this important).
- Include a "Results" section that presents your findings in a clear and concise manner. This section should include four figures (including at least three of the challenges listed below), which can be in the form of graphs, charts, or tables.
 - Ensure that all figures are well-labelled and captioned. Figures should be referenced in the text.
 - Choose a writing style that is appropriate for you and your audience. You may aim to write in an academic style with a formal tone, concise language, and objectivity, or in a blog style with an engaging and conversational tone, and trying to be persuasive. Either way, you should be using plain language (avoiding or explaining technical terms) and facts and information that support arguments should be referenced/cited.
- Provide a "Conclusion" section that summarizes your findings and provides suggestions for future work.
- An "Acknowledgements" section which is a short paragraph to acknowledge any significant source of support during the project. This is your opportunity to thank those you have helped you personally or academically during the process of working on your final project. This could be R support from peers who gave you feedback on your project or helped you troubleshoot errors. This could be a peer tutor in the writing center. This could also be emotional support from friends or family. You do not need to acknowledge the instructor or the TAs. Acknowledgment sections are typically written in the first person.
- A "References" section (APA or MLA are both fine). [Click here to see how to reference datasets in APA style](https://apastyle.apa.org/style-grammar-guidelines/references/examples/data-set-references) (<https://apastyle.apa.org/style-grammar-guidelines/references/examples/data-set-references>).
- A "Notes" section which describes which challenges you attempted. This could be a short paragraph or a bullet point list. This section should be at the very end of your document.

Optional

- You may include a "Supplemental Materials" section at the end of your project for any additional code or figures that are interesting, but not necessary for the main article.

Challenges (at least 2)

In order to make your project stand out, you should attempt at least two of the following challenges:

- Use the `lubridate` package
- Use web scraping with `rvest`
- Join tables
- Reshape data with `pivot_wider` or `pivot_longer`
- Perform Regression (Simple or Multiple)
- Use `map(s)`
- Perform text analysis using `stringr`
- Use a type of plot not discussed in class -- we discussed scatterplots, line graphs, bar graphs, and boxplots (ie. correlogram, diverging bar plot, dumbbell chart, alluvial diagram, heatmap, hierarchical dendrogram, etc)
- Use of an R packages that we have not discussed in class (ie. `gganimate`, etc). Be sure to check with the instructor to ensure your package will meet the criteria for a challenge.

Note that choosing more than 2 criteria does not mean that you will get a higher or better mark. There is no preference in the grading scheme for which 2 challenges you choose. You should pick the ones that seem most appropriate for your data.

★ Things to consider

- **Fit** Your chart(s)/graph(s)/table(s) should feel like they go together and complement each other, but not be redundant (presenting the exact same information). There may be some variables they have in common or be different ways at getting at the same question.
- **Accessibility to General Audience** Are your labels clear? Are your axes clear? Is your title clear? Do you include the source of your data?
- **Color** Is your color palette consistent throughout your project? Do you choose appropriate colors (ie. divergent/qualitative/sequential)? Are they color-blind friendly?
- **Ordering** Consider the ordering in any charts or graphs. Reorder them if you think it appropriate.
- **Aesthetics** Be thoughtful about your choice of theme for any graphs/charts. Be sure your tables are prettied up with `kableExtra`. etc.

To Hide Code:

There are two things you will want to do to make sure your code is well-hidden and displaying nicely:

1. You should use "code folding". You specify this in the YAML header (the very top of the document). Be careful because the YAML header is very, very picky of spacing. Something like this:

```
---
title: "Insert your appropriate/descriptive title name here"
author: "Your Name Here"
output:
  html_document:
    code_folding: hide
---
```

2. You should add a chunk at the very beginning of your project that sets all your chunks to not display any warnings or messages. Note that you could go through each individual chunk and specify no warnings or messages and that would work the same, but it is simpler to just set the default for all chunks at the same time.

```
```{r setup, include=FALSE}
knitr::opts_chunk$set(
 echo = TRUE,
 message = FALSE,
 warning = FALSE
)
```
```

Points 50

Submitting a file upload

| Due | For | Available from | Until |
|--------|----------|----------------|-------|
| May 19 | Everyone | Apr 19 at 12am | - |

Final Project Rubric

| Criteria | Ratings | | | | Pts |
|--|---|---|--|-------------------------------------|-------|
| Submitted a .html file | 1 pts
Full Marks | 0.5 pts
Part Marks | 0 pts
No Marks | | 1 pts |
| Included a Title and Name at the top of the document | 1 pts
Full Marks | 0.5 pts
Part Marks | 0 pts
No Marks | | 1 pts |
| Figure 1 | 3 pts
Full Marks
Figure is correct, complete, and convincing.
Figure is appealing, informative, and crisp. | 2 pts
Part Marks
Figure is partially correct but incomplete or unconvincing. Figure is readable and clear | 1 pts
Part Marks
Figure is illogical, incorrect, or incoherent.
Figure is cluttered, disjoint, or illegible | 0 pts
No Marks
Figure Missing | 3 pts |
| Figure 1 - labels/axes/text readable? title? | 1 pts
Full Marks | 0.5 pts
Part Marks | 0 pts
No Marks | | 1 pts |
| Figure 1 - Color Scheme is appropriate | 1 pts
Full Marks | 0.5 pts
Part Marks | 0 pts
No Marks | | 1 pts |
| Figure 1 - ordering is appropriate | 1 pts
Full Marks | 0.5 pts
Part Marks | 0 pts
No Marks | | 1 pts |
| Figure 2 | 3 pts
Full Marks
Figure is correct, complete, and convincing.
Figure is appealing, informative, and crisp. | 2 pts
Part Marks
Figure is partially correct but incomplete or unconvincing. Figure is readable and clear | 1 pts
Part Marks
Figure is illogical, incorrect, or incoherent.
Figure is cluttered, disjoint, or illegible | 0 pts
No Marks
Figure Missing | 3 pts |
| Figure 2 - labels/axes/text readable? title? | 1 pts
Full Marks | 0.5 pts
Part Marks | 0 pts
No Marks | | 1 pts |
| Figure 2 - Color Scheme is appropriate | 1 pts
Full Marks | 0.5 pts
Part Marks | 0 pts
No Marks | | 1 pts |
| Figure 2 - ordering is appropriate | 1 pts
Full Marks | 0.5 pts
Part Marks | 0 pts
No Marks | | 1 pts |
| Figure 3 | 3 pts
Full Marks
Figure is correct, complete, and convincing.
Figure is appealing, informative, and crisp. | 2 pts
Part Marks
Figure is partially correct but incomplete or unconvincing. Figure is readable and clear | 1 pts
Part Marks
Figure is illogical, incorrect, or incoherent.
Figure is cluttered, disjoint, or illegible | 0 pts
No Marks
Figure Missing | 3 pts |
| Figure 3 - labels/axes/text readable? title? | 1 pts
Full Marks | 0.5 pts
Part Marks | 0 pts
No Marks | | 1 pts |
| Figure 3 - Color Scheme is appropriate | 1 pts
Full Marks | 0.5 pts
Part Marks | 0 pts
No Marks | | 1 pts |
| Figure 3 - ordering is appropriate | 1 pts
Full Marks | 0.5 pts
Part Marks | 0 pts
No Marks | | 1 pts |

| Criteria | Ratings | | | | | | Pts |
|---|--|--|--|--|---|---|-------|
| Figure 4 | 3 pts
Full Marks
Figure is correct, complete, and convincing.
Figure is appealing, informative, and crisp. | 2 pts
Part Marks
Figure is partially correct but incomplete or unconvincing. Figure is readable and clear | | 1 pts
Part Marks
Figure is illogical, incorrect, or incoherent.
Figure is cluttered, disjoint, or illegible | | 0 pts
No Marks
Figure Missing | 3 pts |
| Figure 4 - labels/axes/text readable? title? | 1 pts
Full Marks | 0.5 pts
Part Marks | | 0 pts
No Marks | | | 1 pts |
| Figure 4 - Color Scheme is appropriate | 1 pts
Full Marks | 0.5 pts
Part Marks | | 0 pts
No Marks | | | 1 pts |
| Figure 4 - ordering is appropriate | 1 pts
Full Marks | 0.5 pts
Part Marks | | 0 pts
No Marks | | | 1 pts |
| References Section | 1 pts
Full Marks
Document is well-organized. | 0.5 pts
Part Marks
Some effort to organization, but could be cleared (perhaps adding headers or rethinking the order of figures) | | | 0 pts
No Marks
Did you even try to organize this? | | 1 pts |
| Notes Section | 1 pts
Full Marks
Notes section included which clearly expands which challenges (at least 2) were included in this project | | | 0.5 pts
Part Marks
Notes section incomplete | 0 pts
No Marks
No notes section | | 1 pts |
| Challenges
Were the appropriate number of challenges included? | 2 pts
Full Marks
2 challenges | 1 pts
Part Marks
only 1 challenge | | 0 pts
No Marks
No challenges | | | 2 pts |
| Questions and Analysis | 5 pts
Full Marks
Questions and analysis well motivated, interesting, insightful, and novel. For example, the figures explore a comparison between groups or looks at relationships over time. There is a strong connection between your analysis in all three figures. | 4 pts
Part Marks | 3 pts
Part Marks
Questions and analysis appropriate, coherent, and motivated. For example, the figures explore a comparison between groups or looks at relationships over time. There is a connection between your analysis in all figures OR two are strongly related and one is only weakly related. | 2 pts
Part Marks | 1 pts
Part Marks
Questions and Analysis overly simplistic, unrelated, or unmotivated. There is little connection between your analysis in the three figures. The figures chosen are simplistic (for example a table that only compares a mean between two groups, or a barplot that only has two bars). | 0 pts
No Marks
No questions. | 5 pts |
| Overall Cohesion | 3 pts
Full Marks | 2 pts
Part Marks | | 1 pts
Part Marks | | 0 pts
No Marks | 3 pts |

