

# R Analysis Assignment III - Final Submission

## Background

In this assignment, you will pick a dataset that aligns with your interest and conduct novel, self-directed analysis.

There are three parts to the assignment:

- Independent Proposal
- Class presentation and feedback
- Final submission

For this project, you will identify a published dataset to apply the R tools we've worked with this semester and conduct a new analysis. This may be a dataset that was originally analyzed with a different language or older R tools – or analyze the dataset with a different research question.

This final assignment will include the full workflow and conclusion of your analysis. Incorporating the feedback from your instructor and peers during the workflow, apply the skills learned throughout the course to uniquely analyze a publicly available dataset.

## Objectives

- Formulate a novel research question to investigate a publicly available dataset from a new perspective
- Apply skills and functions practiced during the course to a novel dataset to investigate a research question
- Incorporate instructor and peer feedback to improve the analysis
- Summarize the conclusion of your analysis and contrast the conclusion with the previously published dataset if available to convey the relevance of your analysis

## How to complete the assignment

The majority of this assignment will be your analysis of the dataset using R. For this assignment, you will need to submit:

- The dataset you are working with saved as plain text file with the same name as referenced in your R workflow
- A clear and annotated workflow to organize your analysis that is clean of errors when run
- A max 2 page conclusion

The instructor needs to run your R analysis to grade it so make sure the code runs smoothly, is well annotated, and only uses functions covered in class. If you would like to use a function or package not covered in the course, please email the instructor especially if this was not included in the proposal.

Well written code includes:

- Descriptive and efficient use of headers to organize the analysis
- Efficient and accurate use of markdown free text and R code cells to describe the analysis
- Efficient and accurate use of functions learned in the course to explore, wrangle, and analyze the dataset
- Properly formatted tables and/or figures with accurate axis labels AND a table/figure caption to describe the image as a whole

After your analysis is complete, write a 2 page conclusion to summarize the findings of your analysis and relate it back to the previously published analysis. You may refer to the tables and figures created in the workflow (make sure your code is well labelled with table/figure captions!)

Course grade

The final submission is worth 35% of the final course grade.

Support

You are recommended to start the analysis as soon as you're finished with the proposal and presentation. In fact, you should have already started reading in the dataset and exploring it when preparing for the proposal! The instructor is working on returning your feedback as soon as possible but don't wait, go ahead and get started.

Questions should be posted on the course discussion board so that everyone has the same information and to reduce repeated questions. Clearly indicate which assignment and question you are inquiring about in the subject. Keep in mind that everyone is working with a different dataset so more context about the dataset and problem you are encountering may be required.

You are permitted to post incomplete problematic code in your question – also include any errors that you encounter or an explanation of why the code is not behaving as you are expecting.

You are not permitted to post completed code on the discussion board or otherwise share completed code with others in the course or other platforms. When responding to your peer's posts on the discussion board, do not respond with completed code but instead respond with suggestions such as "please check the name of your object for typos".

There will be time available at the end of most coding lessons for brief questions.

Rubric

Evaluation	Does not meet expectations	Meets expectations	Exceeds expectations
------------	----------------------------	--------------------	----------------------

<p><b>Data exploration and wrangling</b> (7 pts)</p>	<p>Begins data exploration but key features of the dataset have not been identified</p> <p>Data wrangling is incomplete, still includes missing data or formatting is not ideal for moving to next steps</p> <p>Annotations are missing for data wrangling steps</p> <p>0-3 pts</p>	<p>Data is sufficiently explored to identify some areas to be cleaned</p> <p>Data wrangling applied to prepare data for next steps.</p> <p>Data wrangling steps are stated</p> <p>4-5 pts</p>	<p>Comprehensive data exploration to understand the structure of the data and identify areas that need to be cleaned</p> <p>Apply data wrangling functions to effectively and efficiently prepare the data for analysis</p> <p>Clearly explains the rationale and goals of data wrangling</p> <p>6-7 pts</p>
<p><b>Analysis Design</b> (40 pts)</p>	<p>Identified some relevant functions and packages but key tools are missing. Analysis covers less than 4 course modules.</p> <p>Analysis strategy is not ideal for the current dataset. Parts of the analysis are missing or are improperly applied to the dataset.</p> <p>Difficult to follow the analysis and understanding the organization of the tables and figures</p> <p>0-19 pts</p>	<p>Functions and skills from at least 4 course modules applied to the analysis</p> <p>Analysis strategy is acceptable for the dataset chosen</p> <p>Together, the tables and figures support a comprehensive analysis of the dataset. Some figures require additional formatting for clarity</p> <p>20-33 pts</p>	<p>Functions and skills from at least 4 relevant course modules applied effectively and accurate to the analysis.</p> <p>Analysis strategy is well suited and customized to the dataset chosen</p> <p>Together, the tables and figures created support a well organized and comprehensive analysis of the dataset. Analysis is easy to follow</p> <p>34-40 pts</p>
<p><b>Accuracy and Robustness</b> (20pts)</p>	<p>Code has trouble running</p> <p>Code is missing annotations or they are inaccurate</p> <p>Code is missing checkpoints or they are</p>	<p>Code mostly executes smoothly and free of bugs.</p> <p>Code mostly matches the accompanying annotations</p>	<p>Code executes smoothly and free of bugs.</p> <p>Code matches the accompanying annotations.</p>

	<p>inaccurate, many instances of incorrect behaviors 0-9 pts</p>	<p>Code ensures some checkpoints, code rarely leads to minor incorrect behaviors 10-14 pts</p>	<p>Code is well written to ensure checkpoints demonstrating the accuracy of the code and robustly completes the task at hand 15-20 pts</p>
Efficiency of design (5pts)	<p>Code is written with functions that attempt to address the task. Writing is not succinct. 0-1pts</p>	<p>Code is written with reasonably suited functions for the tasks and with mostly efficient writing 2-3 pts</p>	<p>Code accomplishes the task using the most efficient functions and using the most efficient writing 4-5 pts</p>
Reflection (25pts)	<p>Conclusions of the analysis are incomplete Identifies the field in which this analysis is relevant to but contribution to the field is unclear. Relation to previously published analysis unclear. 0-12 pts</p>	<p>Conclusions of the analysis are stated. Relevance of the analysis to the respective field is explained. Conclusions contrasted with previously published analysis. 13-19 pts</p>	<p>Conclusions clearly stated and references the figures that were created in the R file. Relevance of the analysis to the respective field clearly explained. Comparison with previously published analysis highlights similarities and differences between approaches and conclusions. Suggest reasonings for some differences between results. 20-25pts</p>
Formatting (3pts)	<p>Code is missing a large portion of formatting Some tables and figures are missing major aspects of formatting 0-1 pts</p>	<p>Code is mostly annotated to assist the reader with following the methodology and rationale of the analysis. Some tables and figures are missing aspects of formatting</p>	<p>Code is well annotated making it easy for the reader to follow the methodology and rationale of the analysis. Tables and figures are clearly displayed and properly formatted</p>

		2 pts	including accurate and legible labels
			3 pts