

R Analysis Assignment- Proposal

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.

For this first proposal, identify a dataset you would like to work with, conduct a preliminary exploration of the data, and propose a plan of action for your analysis.

Objectives

- Identify a published dataset that includes metadata about the features and phenotypes that aligns with your interest.
- Explain the context of your dataset (Where did it originate from? What is the structure of the data? What are the phenotypes and features? What did the original analysis conclude?)
- Justify how your analysis will differ from the currently published analysis and how your anticipated results will contribute to our understanding of the dataset/field.
- Cite the

How to complete the assignment

Start a new markdown document and complete the assignment in a single document. This includes text clearly formatted using markdown and exploratory code written with R.

Once completed, run the code sequentially from start to end (code cells must be numbered starting from 1 in the first cell with no skips in subsequent code cells numbers). Export the completed markdown document to PDF.

Submit the following documents to the correct assignment on the course website before the due date:

- Markdown document
- PDF exported from the Markdown document
- .csv or .txt file containing your dataset
- PDF of publication of original analysis, if applicable

Course grade

This proposal is worth 15% of the final course grade

Support

You are recommended to start the assignment as soon as possible.

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.

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
Dataset selection 5pts	Published dataset is incomplete Raw data is unavailable Citation is inaccurate or incomplete 0-2 pts	Published dataset is robust and includes descriptions for the samples and features Data is available in raw format Published dataset and accompanying publication is properly cited 3-5pts	<i>Note: You may be required to select a different dataset before proceeding to the next steps of the R analysis project if it does not meet these criteria.</i>

Context of the dataset 20pts	<p>Aims, methodology, and conclusion of previous analysis is unclear</p> <p>Structure of the data unclear or not enough information available to understand previous analysis</p> <p>0-5 pts</p>	<p>Aims, methodology, and conclusions explained, but aspects remain unclear</p> <p>Structure the data explained but details missing.</p> <p>6-12 pts</p>	<p>Research question, methodology, and conclusion of original analysis summarized</p> <p>Describe the structure of the dataset and summarize the samples and features of the dataset</p> <p>13-20pts</p>
Exploration of dataset 15pts	<p>Exploratory code partially explores:</p> <ul style="list-style-type: none"> - The structure of the data - An overview of the samples and features - Data wrangling to prepare datasets for further analysis - Identify areas which may be troublesome in future analysis <p>0-5 pts</p>	<p>Exploratory code begins to explore:</p> <ul style="list-style-type: none"> - The structure of the data - An overview of the samples and features - Data wrangling to prepare datasets for further analysis - Identify areas which may be troublesome in future analysis <p>6-10 pts</p>	<p>Exploratory code thoroughly explores:</p> <ul style="list-style-type: none"> - The structure of the data - An overview of the samples and features - Data wrangling to prepare datasets for further analysis - Identify areas which may be troublesome in future analysis <p>11-15 pts</p>
Explain analysis proposal 30pts	<p>Identify a comparison in the dataset</p> <p>Propose some steps for the analysis. Direction is unclear.</p> <p>Identify some relevant functions and packages but key tools are missing. Analysis covers less than 4 course modules.</p> <p>0-10 pts</p>	<p>Write a novel research question to address in your analysis</p> <p>Propose steps required in the analysis. Consider some areas of concern</p> <p>Identify some relevant functions and packages from at least 4 course modules that will be relevant to your analysis</p> <p>11-20pts</p>	<p>Write a clear and applicable novel research question to address in your analysis</p> <p>Propose key steps in the analysis that will be required and identify areas you anticipate may cause concern</p> <p>Clearly identify functions and packages from at least 4 course modules that will be relevant to your analysis</p>

			21-30 pts
Anticipated roadblocks 10pts	Show consideration for limitations of the study	Identify an anticipated road block. Suggest how it can impact the analysis. Propose an action to minimize or address the roadblock	Identify an anticipated road block and explain why this may be problematic for your analysis Propose at least actions that can be used to minimize or address the roadblock
Efficiency of design (10pts)	Code is written with functions that attempt to address the task. Writing is not succinct. 0-4pts	Code is written with reasonably suited functions for the tasks and with mostly efficient writing 5-7 pts	Code accomplishes the task using the most efficient functions and using the most efficient writing 8-10 pts
Formatting (5pts)	Code is missing a large portion of formatting Some tables and figures are missing major aspects of formatting 0-1 pts	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 2-3 pts	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 including accurate and legible labels Description and exploration of dataset in 2 pages Proposal and anticipated roadblocks in 2 pages 4-5 pts
Export and submission (5pts)	Assignment instructions not followed, format incorrect. 0-1 pts	Assignment instructions loosely followed, inconsistencies with running code sequentially before export	Assignment instructions closely followed to run all code sequentially before exporting as a PDF for submission on course site.

		2-3 pts	4-5 pts
--	--	---------	---------