**Psych 10 Project Rubric**

**Grading instructions for TAs:**

* Grades should be returned within 1 week of submission.
* Download the team’s .rmd file from Canvas.
* Save as “submitter’s name\_etal\_part#\_graded “ (e.g., “king\_etal\_part1\_graded).
* Below each question, write feedback and score directly in the .rmd file (helpful to bold your feedback).
* Save and knit to .html. Check that you have provided feedback/score to every question. Calculate total score. Return .html to entire team over email, providing total score.
* Record total score for team for each team member in Canvas.

**Project Overview**

Over the next 10 weeks you will work on a team data project. The purpose of the project is for you to gain hands-on experiencing investigating a research question using real-world data. You will develop your skills in asking a research question, developing a hypothesis, cleaning and exploring your data, visualizing your data, conducting statistical tests, quantifying effect size, and interpreting your results.

You will complete the project in R Studio using an R markdown template we provide. You will complete the project in three parts. Finally, you will conduct an in-section peer review of another team’s project. In lieu of a written exam, you will conduct a 5-minute presentation of your project with your team during the exam period (example project presentations are provided here: <https://github.com/jackie-schwartz/psych10_project/tree/master/psych10_2019_presentation_slides>

**Data**

In collaboration with your team, you will choose a dataset to analyze. You are limited to datasets that contained in an R package (there are many!). Below is a list of R packages containing datasets. Google is your friend for finding out more about these packages.

1. fivethirtyeight
2. billboard
3. nycflights13
4. diamonds (part of ggplot2)
5. neiss ( use devtools::install\_github("hadley/neiss") )

**Part 1: Due Monday, 1/27 at 9am**

**The goals of Part 1 of your project are to:**

* Identify your question and explain why you are asking it
* Identify, briefly describe, and read in your dataset to R
* Identify and describe the variables you will use to test your hypothesis.
* Identify how you will need to clean/wrangle your dataset.
* Identify limitations of your project as a whole.

First, download the project template saved here: <https://github.com/jackie-schwartz/psych10_project>

|  |  |
| --- | --- |
| Submission format | Is the submission file the completed templay .Rmd? Does it successfully knit? **(-5 points deducted from total if not)** |
| Environment set up | * q1 Are the necessary libraries loaded? (+1) |
| Project question | * q2 Is the question clearly stated (i.e., phrased clearly and efficiently)? (+ 1) * q3 Is the motivation for the question clearly explained? (+ 1) * q4 Is the hypothesis clearly stated and is it testable given the data? (+1) |
| The dataset | * q5 Is it clear where the data come from and how they were obtained? (+ 1) * q6 Are the data saved to the environment? (+1) * q8 Was the glimpse () function used correctly and are the number of observations and variables correctly identified? (+1) * q9 Are the variables correctly identified? (+1) |
| Data challenges | * q11 Is the planned cleaning and data wrangling process clearly explained and is it appropriate (+1)? |
| Project limitations | * Are there at least two limitations explained and are these limitations logical (+2)? |

**Part 2: Due Monday, 2/10 at 9am**

**The goals of Part 2 of your project are to:**

* Clean and wrangle your data and describe this process.
* Visualize and summarize each of your variables and interpret what you see.

|  |  |
| --- | --- |
| Code errors | Did all code run successfully (without errors)? (-1 deducted from total if not) |
| Data cleaning | * q12 Are the steps the group described in q10 of Part 1 completed appropriately and/or did the team update their data cleaning plan based on feedback from the TA? (+3) * q12 Are each of the data cleaning steps explained in writing? (+1) |
| Histograms | * q13 Are the distributions of the variables plotted appropriately? (+2) * q13 Are the distributions correctly described in writing? (+2) |
| Summarize | * q14 Are the measures of central tendency and standard deviation calculated correctly for numeric variables? If using character variables, are frequencies calculated for each group? (+3) |

**Part 3: Due Friday, 3/6 at 9am to TA and peer review group**

**The goals of Part 3 of your project are to:**

* Test your hypothesis using the appropriate statistical tests.
* Estimate how large the effect is by using one of the following: effect size, confidence intervals, Bayes factors
* Visualize your results, including one figure that summarizes your primary result.
* Interpret your results in writing.

|  |  |
| --- | --- |
| Code errors | Did all code run successfully (without errors)? (-1 deducted from total if not) |
| Null hypothesis | * q15 Is the null hypothesis correct and clearly stated? (+1) |
| Test hypothesis | * q16 Are the appropriate statistical tests conducted to test the alternative hypothesis? (+3) * q16 Are the tests and results clearly and correctly explained in writing? (+3) |
| Effect size | * q18 Are the appropriate effect size estimates calculated, including the Bayes factor? (+3) * q18 Are the effect size estimates correctly and clearly explained in writing? (+3) |
| Visualization | * q19 Is there an appropriate, presentation quality plot that is easily correctly interpreted? (+3) * extra credit: Particularly great plot and/or additional excellent plots (+1-2) |

**Peer Review: Due Wednesday, 3/11 at 9am**

For this component you will evaluate another team’s project. Download and open the .html file for the team’s project. Please respond to the following questions:

1. What dataset is the other team analyzing and how were these data collected? Are there any issues with the data collection described? What are potential sources of bias?
2. In addition to the project limitations listed by the authors, are there additional limitations to what the authors can conclude from their analysis of this dataset?
3. After evaluating the team’s question, hypothesis, and analyses, suggest another question and/or analysis you would conduct with this data.
4. Do you agree with the team’s interpretation of their results? Why or why not? Are there alternative interpretations?
5. After evaluating the team’s presentation quality plot, identify 1 strength and 1 weakness of the plot.

**Final Write-Up: Due Monday, 3/16 at 9am**

* Integrate the above work into one cohesive .Rmd file.
* Integrate the feedback you have received for each milestone.