# Rubric for Final Projects (36-315)

Ron Yurko

## HTML report, due Monday, May 2nd at 11:59 PM EST (100 points)

This report (produced from RMarkdown) will act as a public-facing document of your work. Though due on May 2nd, this report will be made public for the statistics department to see on May 4th. After that date, we will leave the reports publicly available unless any member of the team asks for the reports to be taken down after May 4th - if that's the case, we will make the report unavailable after December May 4th.

HTML reports should start with a short paragraph and/or bulleted list describing the data, followed by a short paragraph and/or bulleted list describing the three main research questions of the project. After that, the middle part of your report should consist of your graphs (6 to 10), bookended by statements motivating how the graphs fit into the overarching narrative of the project as well as their interpretation and main takeaways. For example, an HTML report could take the following format:

> We wanted to learn about [INSERT RESEARCH QUESTION 1], which suggests we should examine [NSERT RELEVANT VARIABLES].
>
> [ INSERT YOUR AMAZING GRAPH HERE ]
>
> The above graph suggests that [INSERT MAIN TAKEAWAYS, INTERPRETATIONS, AND CONCLUSIONS].
>
> We also wanted to better understand [INSERT RESEARCH QUESTION 2]. To do this, we plotted [INSERT RELEVANT VARIABLES] using [INSERT GRAPH CHOICE EXCEPT PIE CHARTS HERE]:
>
> [ INSERT YOUR NEXT AMAZING GRAPH HERE ]

And so on. (Please do not just copy-and-paste the above format; it's supposed to be a broad idea of how you should structure your document, rather than literal instruction.) Additionally, **you need to include at least one formal statistical analysis to complement a graph in your report**, e.g., $t$-test, KS test, regression analysis, etc. At the end of your report, you should include a short paragraph (or paragraphs) discussing the main conclusions of your project and potential directions for future work.

We will be grading final reports based on *basic items* and *detailed items*. The *basic items* are in bold below, and will generally be graded in an all-or-nothing fashion. For example, the first *basic item*, "Does the project provide a description of the dataset?", is worth 5pts. If the answer to that question is **Yes**, then the team is awarded 5pts; if the answer is **No**, the team is deducted 5pts. The basic items are meant to set a minimum grade that every team should receive based on very minimum effort; the basic items add up to 60pts. The remaining 40pts are divided into *detailed items* that we will rate on a 0-to-5 scale. Each *detailed item* is under a *basic item* and is worth 5pts each. Thus, you'll find 8 *detailed items* below, adding up to the remaining 40pts.

**Basic item (5pts): Does the report provide a description of the dataset?**

**Detailed item (rated 0 to 5): What is the quality of this description? Does it clearly communicate what the rows and columns (i.e., subjects and variables) are in the dataset in a way that is understandable to a CMU statistics undergraduate?**

**Basic item (9pts): Does the report have at least three clearly stated research questions? (minus 3pts for each question that is missing or not clearly stated)**

**Detailed item (rated 0 to 5): What is the quality of the research questions? Are they well-motivated by real-world/scientific interests? Or are they shallow? For example, a shallow question would be in the form, *What does the distribution of this variable look like?* A more interesting question would motivate why we would like to inspect particular distributions with real-world context.**

**Basic item (16 pts): Does the report have 6 to 10 graphs that follow the data and graphs requirements on Canvas (at least two of the different kinds of graphs, no more than two univariate graphs, no more than 3 of the same type of graph)? (each criterion is worth 4pts)**

**Detailed item (0 to 5): What is the quality of these graphs? Are these graphs easily readable, interpretable, and properly labeled?**

**Detailed item (0 to 5): Are the graphs well-motivated, given the research questions of the project? In other words, do the graphs address/answer the research questions of the project, or do they only provide tangential (or even irrelevant) details?**

**Basic item (10 pts): Does the report provide further descriptions/interpretations for each graph? (minus 2.5pts for each description that is missing, up to 10pts)**

**Detailed item (0 to 5): What is the quality of these descriptions? After reading these descriptions, is it crystal clear what is being displayed in the graph and what the main takeaways are?**

**Basic item (5 pts): Does the report include any kind of statistical analyses and/or summary statistics/tables to complement their graphs?**

**Detailed item (0 to 5): Are the statistical analyses appropriate given the type of data and research questions of interest? Are the statistical analyses interpreted correctly?**

**Basic item (10 pts): Does the report provide some clear conclusions that can be made from their graphs and analyses?**

**Detailed item (0 to 5): What is the accuracy of the claims being made? Are the claims well-supported by the graphs and analyses presented? Are the conclusions well-aligned with the research questions of the project?**

**Basic item (5 pts): Does the report discuss questions that have not been answered by the project, but could be answered with future work?**

**Detailed item (0 to 5): Did the team provide adequate reasons as to why these questions were left as future work (e.g., they need more data, need more nuanced statistical techniques they haven't learned, etc.)? Are these future-work questions well-motivated given what the team has completed for this project?**

## Presentations on Wednesday, May 4th (100 points)

Presentations should be 12-15 minutes, with 5-8 minutes for follow-up questions. All presentations will happen remotely via Zoom, and all team members must be present for the presentation and participate. **Teams must use slides for the presentation, and can NOT just walk through the HTML report** (slide format is the team's choice). The TAs and I will be looking for the following things during your presentation:

- Do they provide a clear description of the dataset? (**10pts**)
- Do they state thoughtful research questions they would like to answer with their dataset? (**10pts**)
- Do they showcase graphs that are of high quality (correctly labeled, interpretable, and informative)? (**20pts**)
- Are they able to clearly explain how to interpret their graphs, as well as pinpoint what the main takeaways of each graph are? (**20pts**)
- Are the graphs (and their main takeaways) relevant to the team's research questions? (**20pts**)
- Do they state the main conclusions and takeaways that can be made from the project overall? (**10pts**)
- Do they clearly communicate a cohesive narrative for the project? And if there are any follow-up questions, do they answer them clearly and correctly? (**10pts**)

Most presentations will happen on Wednesday, May 4th (Reading Day during finals week), but some presentations may happen on May 5th or 6th if any team members have conflicts with May 4th. When we get closer to finals week, we will create a sequence of 20-minute time slots that each team will sign up for to present. All presentations must happen by May 6th at the absolute latest.