

**Instructions: Group Project Progress Reports II**  
**STAT 340: Data Modeling II**  
**Due December 3rd, 2021 at 11:59pm**

The purpose of this assignment is to serve as a second check-in on your group's project. Think of this as a sort of second draft of your group's final report (though strictly speaking your "final report" will take the form of a poster) or a check-in to see how your group is doing. Of course, as with your first progress report, we have no expectation that you have definitively answered your statistical question of interest or anything to that effect. Rather, the goal of this assignment is to ensure that you are making progress toward that goal, and to give us an opportunity to see potential issues well ahead of the final project deadline at the end of the semester.

By Friday, December 3 at 11:59pm, you should submit via Canvas a document in PDF format that includes:

- The names and NetIDs of all group members.
- A description of your data set and a summary of your statistical question(s) of interest (you may copy-paste this from your proposal and previous progress report, if you wish).
- A description of where this data came from, who gathered it, and some background information, such as why your group found this data set in particular to be of interest and why your reader should care. If you obtained your data set from Kaggle or a similar repository, it is not enough to simply say "This was collected by user xyz and posted to Kaggle". You should be able to explain the specific source of your data (e.g., a medical study, survey, government agency, etc.).
- A description of the variables available in your data set, with more details for the variables that are likely to be relevant to your question (you may copy-paste this from your proposal and previous progress report, if you wish).
- A block of code showing how to load the data into R.
- At least one plot summarizing some aspect of the data set that is interesting to you and your group. Your plot should be well-labeled, have an appropriate title, and your document should include a reasonably detailed description of what the plot shows as well as any discussion/interpretation of the contents of the plot (e.g., if the plot shows a clear trend in your data, discuss that trend, why it is or isn't surprising, etc.).
- At least one plot or table summarizing the output of an estimation procedure, test, model fitting or some such statistical procedure. Please explain clearly the procedure and/or method(s) you used and why this was an appropriate choice in light of your project goals.
- A brief discussion of the progress and/or challenges faced so far in answering your statistical question(s) of interest. This may include a discussion of the methods and models used; issues that arose when downloading and cleaning the data; shortcomings of the methods/models used so far, difficulties in understanding particular variables in the data, etc.
- A summary of your next steps (e.g., your goals for the remainder of November, other methods/models you want to try, etc.).

There is no particular length requirement for this document, but five pages total is a good target (including figures). If you can address all of the above in less space, that's fine. If you need a bit more space, that's fine too. Give enough detail that your instructors and TAs will see that you have thought carefully about your statistical question(s) and that you have been working consistently on your project.

We expect to see substantial progress when comparing your previous report to this one. Of course, we recognize that for many groups, progress may take the form of trying a lot of things that simply didn't work. If this is the case, we still want to hear about it. Discuss these challenges and dead ends, if you encountered them. That's an important part of research, even if it isn't as flashy and exciting as finding a small p-value, and it warrants acknowledgement and discussion.