

Instructions: Group Project Final Reports
STAT 340: Data Modeling II
Due December 18th, 2021 at 11:59pm

This assignment serves as the final report for your group's project. This is your group's chance to demonstrate all you've done on your data set this semester and show off some of the things you've learned. As with previous iterations of your report, we have no expectation that you have definitively answered your statistical question, but we do have the expectation that your report feel reasonably complete.

We note that we have previously referred to a final presentation for the end of the semester. Our hope had been that we could end the semester with a poster session, like what is common at machine learning conferences, where you could share your work with your classmates. Unfortunately, both due to the uncertainty surrounding COVID-19 and the logistical issues of arranging a suitable space, we have decided to restrict the scope of the final project to a report.

By Saturday, December 18 at 11:59pm, you should submit via Canvas a document in PDF format that includes:

- The name of your group and the names and NetIDs of all group members.
- An abstract of at most 300 words giving an overview of your data, the question(s) you tried to answer, and a brief summary of your findings.
- A brief introduction describing your data set: where it came from, who gathered it and a brief description of why your group found this data interesting 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 previous drafts, if you wish).
- A discussion of your statistical question of interest.
- A summary, including as many plots, tables, R model outputs, etc as necessary to show how you tried to answer your statistical question. You should explain the reasoning behind your decisions to use particular methods, models or tests. This summary need not include only successes— if you tried something that turned out not to work, take some time to discuss it, why it might not have worked, and how you might fix the problem given more time.
- A *brief* summary of your findings and potential future research questions.

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. You will be assessed primarily on the clarity and thoroughness of your presentation, along with the technical correctness of your decisions. Of course, as always, if you can defend your choice to use a particular method in a reasonable manner, that is enough.

We expect that this final report will incorporate feedback from previous reports. In addition, we expect to see substantial progress when comparing your previous report to this final 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 acknowledgment and discussion.