

36-402: Guidelines for Constructing a Data Analysis Report

1. **You must not communicate at all with your classmates, your friends, or anybody other than the professors and the TAs about the Data Analysis Exam, until the solutions have been posted.** This rule applies to Piazza, online discussions and emails as well. Evidence of illegal communication will be taken seriously and handled according to the university's policy for Academic Integrity and Plagiarism.
2. The use of the Piazza Discussion Board is restricted to clarifying questions that do not give away your approach or solution. You are not allowed to answer other students' Piazza questions during the exam period.
3. You are of course allowed to consult your class notes, your homework assignments, or any of the other course materials in order to complete this exam. You are permitted to use the internet as a resource but you may not use it to communicate with anybody else or ask questions about your assignment. Everything you write must be in your own words, and you must explicitly cite any sources used other than the class materials.
4. The Data Exam must be submitted electronically through Canvas by **Thursday, March 7, 6:00 pm**. Your submission should include two parts: the "data analysis report" and the "R code" that was used for the data analysis.
 - The report/write-up must be typed and submitted as a pdf-file. Please do not include R commands and error messages in your report.¹ Your report should look professional and have the same format as a scientific report (see below for details).
 - The code would be a .rmd file if you use R Markdown. If you are not using a knitting program, your code would be a .R file.
5. Your report must contain the following sections:
 - Introduction
 - Exploratory Data Analysis
 - Initial Modeling/Diagnostics
 - Model Inference and Results
 - Conclusion and Discussion

Clearly separate the sections by bold headings and appropriate spacing.

6. Different exercises/exams will have different word limits; be sure to adhere to these limits. The word count only includes the text of the report, not words in the captions or other parts of figures or tables. The Appendix is not included in the word count.
7. Use a standard font, in a size of at least 11 point.
8. Use either 1.5 spacing, or double spacing.
9. Use margins of at least one inch on each side.

¹If you use R Markdown to generate your report you may find this website helpful for examples of how to use the *echo* command: <https://yihui.name/knitr/options/>.

10. Your report should not include mathematical formulas or derivations. If there is some reason why a particular derivation is necessary, then put it in the Appendix.
11. Figures should be embedded into the report (not just added at the end of the report), and referenced to in the main body of the text.
12. All figures must have appropriate captions, labels, and so forth. The font size must be sufficient for the plot to be readable.
13. The use of color in plots is nice, but not required. Your score will not be affected by whether or not you use color. Do not use color in the text of your document.
14. Do not just paste R output into the body of the report. Instead, format it into a table.
15. Your language should be very clear and precise. Do not make claims for which you have no evidence. Do not say “will” or “would” when you really mean “may” or “might.” Do not use language that implies causation. You are studying associations between variables only.
16. Grammar will be marked; as the semester continues, writing style and syntax will be checked more closely in data analysis reports. Move away from wordy phrases (*e.g.* *This is because, this is due to, the reason that this is, this means that, I believe that this, I think the reason is that*). Do not start any sentence with This, These, etc without a noun following it. (*These results show vs. This shows*). Write strong sentences using careful scientific language. Use your words wisely. The report should consist of paragraphs built from well-formed sentences.
17. The instructions for the exercise will include a list of specific questions/issues that must be addressed in your report. You can raise other issues as you see fit. **But**, for the listed issues, you must do the following: In your final pdf-document, highlight **the sentence** that corresponds most closely to each issue and write down the **question/issue number** that is being addressed in the margin.
18. I may add additional clarifications as I see fit...

The grading of the data analysis reports will use the following structure. First, each of the listed questions/issues to address will be worth three points. Then, each of the following elements will be rated from zero to ten:

- Clarity and appropriateness of report text
- Clarity and appropriateness of figures
- Grammar
- Completeness of statistical methods: EDA, Diagnostics, and Model Analysis

Finally, points will be subtracted for any deviation of the guidelines listed above, for example, too many words, incorrect structure to the document, font too small, failure to include R code, and so forth.