

Final Project

Due: Friday, December 10 @ 5pm

The final project is the chance for you to show off everything you've learned about statistical analysis and data analysis in R this quarter! For this assignment, I would like you to choose a data set of your choice and perform a thorough statistical analysis of your choice that fits the data (think about our ice cream reports in class). Some ideas of datasets could be:

- Data you have collected from your rotation or other old data from the lab (make sure to get permission from your PI for this!)
- Data previously published that you want to do a different analysis
- Any random data from the internet

Guidelines for the analysis can be found below:

1. Choose a dataset of your choice to perform a statistical analysis
2. Import the data into R and explore the data - what are the variables? What is the spread and center of the distributions?
3. Write up a short (~1-2 paragraph) proposal about what data you plan to use and what statistical test(s) you would like to perform. This proposal is designed to get you thinking about the project early and get feedback from Katie before the final project is due. (**DUE: Friday, November 12 @ 5pm**)
4. Select **at least two** analyses to perform with your data and run the analyses in R, generating an Rmarkdown report of your analysis (for an example, see the Ice Cream reports from class). For each analysis, be sure to state your:
 - Hypotheses
 - Assumptions
 - Methods
 - Results
 - Conclusions
5. Write up a 1-2 page report detailing:
 - What the data is and why you chose it
 - What was your research question
 - How you went about answering the research question (and why)
 - Your results (with figures/tables if appropriate)
 - What you conclude from your results (be sure to relate this back to the research question)
 - What, if any, other statistical tests you could do with this data in the future **OR** what other type of data you would like to collect for this project so that you *could* answer a different research question of interest?

Projects will be graded on:

- Selecting an appropriate statistical analysis for the data that will aim to answer the highlighted research question
- Showing thorough knowledge of the methods used (i.e. did you state your assumptions and check that your data meets them? Did you do the test correctly?)
- Correct interpretation of the results
- Some amount of thought/discussion about how to improve the analysis or future directions of the project
- Cleanliness/readability and reproducibility of the Rmarkdown report (Note: your code does not have to be "perfect", but I expect text/comments explaining each step. Think: if you had to send this report to a colleague, would they be able to understand it (*even if they had little statistical background*)? Would they be able to exactly replicate your analysis? Don't leave a single step out!)
- Clearness of the writing: your audience for the written report is a scientist in a completely different field from you who has never taken a statistic course before. Be sure to stay away from jargon, clearly explain all ideas and conclusions, and don't assume the reader knows what you are talking about!