

# STATS 32 Final Project Guidelines

To demonstrate your understanding of the material we have covered, you perform a data analysis project in R. **Each person should submit their own project.** While you may discuss ideas and troubleshoot code with each other, each person must submit a unique project.

## Project Minimum Requirements:

- 1 .Rmd file, as well as 1 .html or .pdf file. (Don't worry about what these file formats are and how to create them right now; we'll talk about this in Week 3.)
- Use a dataset that we have not used in class. If using an external dataset, provide the data file and/or a link to it.
- An "Introduction" section describing the problem and the datasets to be used.
- A "Data analysis" section where you have code. Code should have accompanying text that describes what the code is doing and interprets the results.
- At least 3 data visualizations (not all of the same type).
- A "Conclusion" section to summarize the results of the analysis, as well as any deviations from the project proposal.

While your analysis should not be too skimpy, it does not have to be comprehensive either. As a rough guide, the PDF version of the report should be anywhere from 5-15 pages. (While this may seem like a lot, recall that printing parts of the dataset and graphs take up a lot of space.) These links are examples of what the content of the final project could look like:

- <https://www.r-bloggers.com/network-analysis-of-game-of-thrones-family-ties/> (but shorter)
- <https://www.r-bloggers.com/sentiment-analysis-of-the-lord-of-the-rings-with-tidytext/> (but longer)
- <https://www.r-bloggers.com/text-analysis-of-trumps-tweets-confirms-he-writes-only-the-angrier-android-half/>
- <https://www.r-bloggers.com/reproducible-finance-with-r-sector-correlations/>
- <https://www.r-bloggers.com/the-heat-is-on-or-is-it-trend-analysis-of-toronto-climate-data/>

Here are some ideas for datasets which you might want to work on:

- Governmental or NGO open data websites (e.g. City of Palo Alto, State of California, World Bank)
- <https://github.com/caesar0301/awesome-public-datasets>: A compilation of links to various open datasets
- Data which you have collected yourself

## Project Proposal

For the project proposal, submit a 1-2 paragraph description of your final project. You should provide details on the problem you wish to explore, any datasets that you will use, and potential visualizations.

The project proposal should be submitted as a text file (any format will do).

In implementing your final project, you should try to stick as closely to your proposal as possible. If you are unable to do so, please provide a paragraph in your final project to explain the deviations from the proposal.

## Grading & Deadlines

Both the project proposal and final project should be submitted through Canvas.

The project proposal is due by **Oct 20 (Fri), 23:59:59** and is worth 20%.

The final project is due by **Nov 3 (Fri), 23:59:59** and is worth 80%.

Please note that these deadlines are hard and that there will be **no extensions** granted to be fair to students who submitted their work on time. If you foresee that you may have difficulty meeting these deadlines, please come and speak with me **immediately**.