

Project for STAT 530 – Fall 2025

You will analyze a multivariate data set of your choosing, using any appropriate method or methods we have discussed in this class.

The data set could be one that is of personal or academic interest to you and that fits one of the methods of multivariate analysis studied in STAT 530. It should be a real data set which you have not analyzed before and which has not been analyzed in a textbook (using the methods of this course, anyway). When choosing a data set, make sure that each individual observation has several variables measured on it (certainly at least 4 or 5 variables!). I recommend picking a data set in which at least a few of the variables are numerical/approximately continuous. Some of the variables in your data set could be categorical, but most of the methods we have discussed in this class assume that at least some of the variables are numerical. **For 5% of the project grade, please submit details about your chosen data set (i.e., tell me what are the variables and the observations) by Tues., Oct. 28.**

You may work individually, or in groups of two or three people. Since it is very common in the working world to do data analysis in teams, **I strongly encourage you to work in a group!** To encourage this, I'll add 3 percentage points extra credit to the project grades of those working in a group.

You should write a concise report summarizing your analysis. The report should be no longer than four (typed) pages, not counting any R output, graphs, etc., which you may wish to include as support or illustration for your analysis. The style of the report is up to you, but the best reports will use some graphical and analytical tools we have discussed to reveal interesting findings about the data set.

There is a very wide choice of analyses that you could choose to do for the project (depending on the data set and nature of the research questions you are interested in). Don't try to do every kind of data analysis that we've talked about in the class!!! Just pick a small number of analyses that will reveal what you're really interested in learning from your data. There should definitely be some kind of graphical summary / visualization of the data, but which kind(s) of graphs you choose is up to you and your goals.

For extra credit, you can do the following in addition to your written report:

You can upload a video file of yourself or your group making a brief oral presentation about your data set, the data analytic approaches you took, and your conclusions. This is to (1) give you practice in speaking clearly and understandably about technical data analysis, and (2) allow others in the class to learn about your project.

If you choose not to do the video, then as an individual or group you should post a very brief one-page summary of your data set and your most important findings (you can include graphs on separate pages if appropriate) to a shared site on Blackboard that I will set up.

Part of the grade for the project will be for you to view some of the videos or brief summaries by other students and groups and to make brief comments about their projects. More details on that will be shared later.

The report will be **due on or before Friday, November 21, 2025 at 11:59 p.m.**, and the project grade will count as 15% of your overall course grade.