

## Project 1 STAT 6021

The body of the project – report and presentation – will be completed in teams of three or four. You are free to form your own groups.

The topic of the project will be diamonds! The data set was recently scraped from bluenile.com, and it describes more than 210,000 different diamonds they have for sale. All groups will work on this same data set, but groups are free to come up with their own research hypotheses and questions.

Each project should feature:

- Clear central analytic goals and/or questions to answer – the more interesting and challenging, the better. The more “practical,” the better.
- A “substantial” computational component for the analysis.
- Analytic methods developed in this course.

The deliverables are

- A report (.doc, .docx, or .pdf)
- reproducible R-code (.r file)
- A presentation document (.pdf, .ppt, etc.),

and they are due on Collab at **Thursday, July 18th at 9am.**

### Component 1: The Report

Each team will prepare a project report. This report should be less than 10 pages and should include the following:

- An executive summary that describes the high-level goals/questions of the project, the nature/characteristics of the data used in the analysis, and the results of the analysis, including any recommendations.
- A detailed description of the data, the analysis (especially this), and the conclusions reached, making use of graphics wherever appropriate.
- Correct grammar, clear explanations, and professional presentation.

I strongly recommend that you attach the R script as a **separate file**, and that you only describe the results you obtained. For example, don’t copy/paste code or software output.

Pretend that the audience of the report is comprised of **both** clients who have no experience analyzing data, as well as data scientists and statisticians to whom the client may give the report for a second opinion.

The printed project report and commented/annotated R code are due in Collab on **Thursday, July 18th at 9am**. You should submit a .pdf or .doc file for the report, and a .r file for the code.

### **Component 2: Presentation**

Each team will give an 8-minute presentation. This presentation should be designed to be understandable by anyone familiar with the course material, but who has not read your project report yet. Each team is free to organize who talks about which topics during the presentation. Not everyone needs to speak, but all team members should make equal contributions to the project as a whole.

Presentations will take place on **Thursday, July 18th at 9am** in class. You must be present for all presentations unless you have prior permission due to a schedule conflict. Each team should bring their presentation slides on a flash drive or be prepared to **quickly** access them online if using an internet-based presentation. If you're unsure about whether or not your materials will work on the machine we have in our classroom, you should figure that out before class starts. Again, the presentation slides should be submitted to Collab by the beginning of class on **Thursday, July 18th at 9am**.