

Stat 6021: Project 1

Background information: You will be working in assigned groups of 3-4 students. Each group will work on the same data set. The data set that you will be working with describes more than 1,000 different diamonds that are for sale on <http://www.bluenile.com>. A .csv file of the data will be provided to you on Collab. Please note that the .csv file contains a subset of the diamonds on Blue Nile as well as from other internet resources. The variables are:

- carat
- clarity
- color
- cut
- price

Detailed descriptions of these variables can be found on the [diamond education page on Blue Nile](#).

Tasks: You have been approached by Blue Nile to perform the following tasks:

1. Use data visualizations to explore how price is related to the other variables (carat, clarity, color, cut), as well as how the other variables may relate to each other. Address the various claims on the [diamond education page on Blue Nile](#).
2. Fit an appropriate simple linear regression for price against carat.

Deliverables: Your group will submit (via Assignments on Collab, one submission per group)

- A report (.html or .pdf file)
- An R script containing your code (.R or .Rmd file)

Each of you will also complete the **Project 1 Evaluation Questions**, which is also **due the same day as the report**. Please note that this evaluation is specifically for the project.

Report sections: The report should include the following sections:

1. An executive summary that describes the high-level results of the analysis. This executive summary should be written in a way that can be understood by a wide variety of readers, including readers with no background in statistics. A way to think about this is how newspaper articles report results from various studies, so avoid technical jargon. This section should be no more than 2 pages.
2. A description of the data and the variables, as well as the data visualizations you created to address how price is related to the other variables as well as the claims made on the diamond education page. Be sure to provide contextual commentary on the visualizations.
3. A description of how you fitted the regression of price against carat, and the conclusions reached. If possible, be sure to provide some contextual commentary on the linear regression equation that you propose.

The audience for sections 2 and 3 is another classmate your client may hire to review your report.

Report guidelines:

1. Include the names of the group members and group number in the heading of your report.
2. Have sections that are clearly labeled.
3. Aim for no more than 20 pages. If you go over this limit a bit, that is fine.
4. Do not use appendices as a way to work around the page limit. Anything that belongs in the main body of the report should be in the main body and not be tucked away in an appendix. I will not read anything in the appendix.
5. The report should contain correct grammar, clear explanations, and professional presentation.
6. I should be able to repeat your analysis without looking at your R code.
7. Your report does not need to include any R code. Relevant output from R (e.g. graphs, results from hypothesis tests, etc) should be included if the output is referenced to in the report.
8. The text in your document should be readable after printing out. Difficult to read documents will be penalized.