Your task in this project is to select a real-world data set to which simple linear regression can be appropriately applied, and then to conduct a regression analysis of the data set using the techniques that you have learned through Chapter 2. The analysis should be your own (with computations conducted in R), even if you select a data set from a book or paper with its own analysis. I would recommend that you choose a data set that you find interesting or relevant to an area of study that you enjoy.

The idea in this project is to pretend that you are a statistical consultant. It is up to you both to conduct a thorough analysis of that problem (or as thorough as possible given the material that we have studied) and to communicate your results to your clients (who the rest of the class and I will pretend to be). This communication is an important part of statistics, both in its written form and in the presentation. You may assume some familiarity with statistics on the part of your clients, but you should not get completely mired in technical terminology.

Projects will be done individually and will consist of three parts:

1. A written analysis and solution of the problem, probably a couple of pages or so.
2. A 6- to 10- minute in-class presentation.
3. Attendance at the other student presentations.

The schedule of events for these is as follows:

    Thursday, September 27  Topic proposal due
    Thursday, October 4    Paper due
    ______________________ Presentation date (TBA, some time in October 4-8)

*Topic proposal:* In a paragraph, describe the data set that you are proposing to analyze: where it comes from, how it was collected, why you selected it, and so forth. If you have a guess as to what your regression analysis may show (possibly based on what the originators of the data set found), you might also include that.

*Presentation date:* These will be decided by lot. You might write your presentation date down in the place for it above.

*About the paper*
The paper should include the following sections: introduction, background, data, analysis, conclusions, and bibliography. All of these should be clearly labeled, except for the introduction, which simply begins the paper without a header.

The introduction should state in brief what will be done in the paper. Its purpose is to allow the reader to decide whether or not to read the paper, so details need not be
provided. Instead, just the information relevant to deciding whether or not to continue reading should be included.

The background section should include information about the context of the data set. For example, how was it collected, when, and by whom? Such information is important for allowing the reader to understand and interpret the results of the paper.

The data section should give a clear depiction of the raw data, not yet analyzed. Some thought should be given as to what type of table or data display should be used. A graphical display should be saved for the next section, since it generally does not give the exact values of the data set. If the data set is too large (perhaps doesn’t fit on even on a page or two), then a suitable reference should be given as to where it can be obtained electronically. *For this paper, you should also submit your data to me by email or with a link to the web, so that I can check your computations on R myself as well.*

The analysis section should include the various things that you use R to compute, along with a little bit of interpretation or explanation of them. Some items that belong in this section are: an explanation of the regression model that you applied to the data (which variables are which), the regression line obtained, the correlation coefficient, a scatterplot of the data and the regression line (with appropriate and informative labels for the axes and for the plot itself), the standard error for both coefficients, a 95% confidence interval for both coefficients, and a $p$-value (along with a statement of whether or not this is statistically significant) for both coefficients.

The conclusions section should not include much if any technical jargon, but rather should contain information that the purported “clients” can understand about the solution or resolution of the problem that was analyzed.

For the bibliography, there is no specific format needed. If you are aware of such a format for the field of study that you are in, then go ahead and use that. If not, just be sure to put enough information about the sources for anyone interested to be able to find them.

**About the presentation**

Two brief suggestions in preparing for your presentation:

1. First and foremost, *be absolutely sure to practice by giving your presentation to some other audience (a friend, roommate, etc.) before presenting it to the class.* I cannot stress the importance of this enough—it helps in so many ways! One important thing it does is to help you get a sense of the timing of the presentation, which is difficult to impossible to judge otherwise. Also, it allows you to iron out some of the wrinkles in the talk, where it you to have trouble presenting or where something you were hoping to present simply doesn’t work well when presented (in which case it can often be removed).

2. Also, if you are using any technology at all, come into the classroom the day before your presentation or earlier and try it out. Technology is a huge source of
problems in presentations—audio not working, incompatibility of one version of Powerpoint with another, ... . The list goes on and on. You will need to try everything out that you can beforehand. Even then, things can go wrong, but at least in that case they will not be your fault. If you do not try out your technology (in the actual room) beforehand and something goes wrong with it, your grade will be affected adversely. Your grade will not be affected by technology that worked when you tried it before (on the actual equipment to be used) but then inexplicably doesn’t work the day of the presentation (such as disappearing speakers, etc.)

I will give you a handout on making statistical presentations that you may refer to as well.

A brief comment
Please see me at any point along the way if you would like any help or guidance in selecting a topic or conducting your analysis. While I certainly won’t do either one for you, I may be able to get you unstuck if you happen to get stuck at some point (including finding a topic).