Your task in this project is to select a real-world data set to which multiple linear regression can be appropriately applied, and then to conduct a regression analysis (including diagnostics leading up to the regression) of the data set using the techniques that you have learned through Chapter 6. In contrast with Project 2, however, not only must your data set have at least two explanatory variables, but at least one of them must be categorical. 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. *Hint:* Try consulting faculty in your department to see if they have any suitable data sets!

As before, 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 handful 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:

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thursday, November 29</td>
<td>Topic proposal due (having run diagnostics)</td>
</tr>
<tr>
<td>Monday, December 3</td>
<td>Paper due</td>
</tr>
<tr>
<td></td>
<td>Presentation date (some time in December 3-6)</td>
</tr>
</tbody>
</table>

*Topic proposal:* As with Projects 1 and 2, in about 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. In addition to what you did for the topic proposal for Project 1 though, you should have run the relevant diagnostics to make sure that multiple regression analysis is suitable for the data. (The regression assumptions should be satisfied, etc.) You do not have to include graphs in the topic proposal, but you should state what diagnostics you have done and why you therefore know that regression analysis is appropriate. I won’t collect these but will simply ask you about them in class.

*Presentation date:* This 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, diagnostics, analysis, conclusions, and bibliography. All of these should be clearly labeled, except for the introduction, which simply begins the paper without a header.

Although it is not usual for a paper in a journal, in this paper you should include all of the relevent graphs and statements about how you know this data set is suitable for multiple linear regression. This should be modeled after the diagnostics that you have done in the homeworks (as in the text and the Faraway handout).

The main difference between this paper and the paper from Project 2 is that, as part of this paper, I would like you to submit an R script file that can be run with the `source()` command. This script file should contain comments indicating what each command is designed to do. We will discuss how to write such a file in class, and I would be happy to help anyone who is still having difficulties writing the script.

About the presentation
As with the previous presentation, be absolutely sure to practice by giving your presentation to some other audience (a friend, roommate, etc.) before presenting it to the class. There is little that is different from the previous presentation in this presentation, except that now you will probably have more pictures/graphs to look at (which can be quite helpful).

As before, you should check that whatever technology you intend to use actually works for you. We have already seen difficulties where pictures don’t load. Try to avoid this.

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).