Regression and Time Trends Exercise: Independent project.

For this exercise, we will return to the interactive web application at http://spark.rstudio.com/statmos/mod1_regression.

Recap

We now have a few new tools in our data analysis toolkit for detecting trends over time: we can use them to identify situations where we can use Ordinary Least Squares (OLS) to test whether apparent trends are likely to be due to chance. If OLS is clearly not appropriate due to temporal autocorrelation, we have learned about a few different ways of dealing with the issue that still allow us to make inference if apparent trends are real. These tools are important for monitoring our environment. Our atmosphere, lands, and oceans are dynamic and it is often difficult to identify long-term trends when there is lots of short-term variation. Although the appearance of a significant trend over time doesn't tell you much about what's driving the trend (that's a much harder problem), the first step is usually to test whether there is any trend at all.

Instructions

- For the directed exercises, we have had you focus on a few different sets of data that are appropriate for demonstrating particular statistical issues. The data sets that are accessible through the web application are much more extensive, incorporating climate data for more than 200 countries and more than a million different search terms in Google Books. Your assignment is to use these data and the tools provided to test a hypothesis about trends over time. The hypothesis should be framed in a way that you can use the tools provided to test it explicitly using linear regression. An example of a hypothesis like this is below:

  The declining sales of Nintendo-branded game systems will have resulted in the phrase "Nintendo" decreasing in frequency in Google nGrams from 1990 to 2008.

- The main challenge here is to choose an appropriate method to analyze the data, justify your choice, and correctly interpret how the model results relate to your hypothesis.

- Another important point is that where possible, scientists try and form hypotheses *before* they look at the data, not afterwards. It is easy to come up with a hypothesis to match any particular set of data, but the real challenge in science is to make predictions about the patterns you expect to see before you collect the data. This means that you should think about what hypothesis you want to test *before* you look at the data, not after.

To complete this assignment, you should answer the following questions with a short paragraph.

1. What hypothesis did you decide to test? Why was this interesting to you?

2. Paste-in a graph of your data.

3. After you looked at the data, but before fitting any models, did there appear to be a strong trend in the data over time?

4. Which analysis method (OLS, Subsampling, Lagging, GLS) did you select for the analysis? What properties of the data caused you to make this choice?

5. How do the model results relate to your hypothesis? Do they provide evidence to support or reject it? What is that evidence?