Homework 3

Scientists are interested in the Earth's temperature change since the last glacial maximum, about 20,000 years ago. The first study to estimate the temperature change was published in 1980, and estimated a change of -1.5 degrees C, ±1.2 degrees C in tropical sea surface temperatures. The negative value means that the Earth was colder then than now. Since 1980 there have been many other studies. Different studies use different measurement techniques, or proxies. Some proxies can be used over land, others over water. The proxies are

proxy	code
"Mg/Ca"	1
"alkenone"	2
"Faunal"	3
"Sr/Ca"	4
"del 180"	5
"Ice Core"	6
"Pollen"	7
"Noble Gas"	8

Each study reports an estimate `deltaT`, a standard deviation of that estimate `sdev` as a measure of uncertainty, the `proxy` used (coded 1 to 8), whether it was a terrestrial or marine study (`T/M`), which is coded as 0 for Terrestrial, 1 for Marine, and the latitude at which data were collected (`latitude`). Data are in the file `climate.txt`, linked on the canvas page of the class (same place where you downloaded this problem set).

1. Write down a general linear model that allows a different intercept for each proxy as well as separate coefficients for latitude and latitude squared for each proxy and a variance for each observation that is proportional to the measured standard deviation squared.

2. What is the rank of the design matrix?

3. Consider the following R code:

```
X = model.matrix(deltaT ~ factor(proxy)*poly(latitude, 2), data=climate)
```

What is the rank of X? Do the columns of X span the same space as describe in part 1? Explain what the code does and provide an interpretation of what the parameters mean (use the `help()` function in R).

4. Using the above model forumula, fit the generalized linear model using the column `sdev` to construct appropriate weights. Provide an interpretation of each of the parameters in the context of the data; i.e. quantify and interpret how changes of "1 unit" (or more) relate to climate change. This interpretation should understandable by someone with little statistical backgound.

5. Using the model output, construct confidence intervals that address each of the following questions:

1. Do estimates vary systematically by proxy?
2. Does latitude affect the temperature change?
3. Do terrestrial estimates differ from marine estimates?
4. Are some proxies more precise than others?

Sumarize your results in a one page report (one page of tables & plots may be attached), with a brief summary (non-statistical but that provides quantification and measures of uncertainty of key results), introduction, model description, results, and discussion. For discussion, what other simpler models might be appropriate to try out as suggested by your intervals?

No need to fit any alternative models or carry out a formal hypothesis test at this point in time to address the questions, but use confidence intervals as a form of EDA (exploratory data analysis).

Please turn in a typed solution (pdf file by email is fine, of course). In your solutions please make sure that all graphs have a legend and/or caption that defines colors, lines etc, units are given, and axes are clearly labeled. Any graphs that are included should be referenced in the text by Figure 1, etc. (if you do not refer or discuss it do not include it!). Please do *not* submit R code, only the salient results.