Summary Report

In this analysis of climate change, we investigate the linear relationship between global average temperature and time. We use linear regression to establish this relationship.

We gathered climate data from Berkeley Earth, and focused on global land temperatures for the analysis. Data are available from 1750-2013. In preparing the data, we transformed from Celsius to Fahrenheit. There were 12 missing data points in years 1750 and 1751; these years were omitted from analysis.

An initial look at the scatterplot of data suggests that linear regression is an appropriate approach to investigating this problem. We can clearly see a general overall rise in temperature across the time frame. We continue by creating 4 different models across different time periods to see if there is an increase in the rate of change.

Model Interpretation

The 1735-2013 model is Y = 0.3103+0.008518X where Y is Average Temperature and X is Time. Both the coefficient and intercept are statistically significant at the alpha <0.05 level. The overall model has a statistically significant F-statistic as well and is a good fit. The adjusted R^2 = 0.8257.

The 1850-2013 model is Y = 0.1775+0.01536X, with both the coefficient and intercept statistically significant at the alpha <0.05 level. The overall model has a statistically significant F-statistic as well and is a good fit. The adjusted R^2 = 0.7479, which is not an improvement.

The 1900-1974 model is Y 25.238+0.01142X, with both the coefficient and intercept statistically significant at the alpha <0.05 level. The overall model has a statistically significant F-statistic as well and is a good fit. The adjusted R^2 = 0.3623. This adjusted R^2 is much lower than the other two models and thus not as good.

The 1975-2013 model is Y = -49.148211+0.048966X, with both the coefficient and intercept statistically significant at the alpha <0.05 level. The overall model has a statistically significant F-statistic as well and is a good fit. The adjusted R^2 = 0.7813. While this R^2 is not as good as the first, it is still an acceptable R^2.

From R^2 perspective, model 1 has the best result. However, from an interpretation standpoint, the coefficient for time in Model 1 is close to 0 and thus shows just a small rise in temperature. This model does not provide a good fit to recent temperatures while Model 4 does.

Comparing model 3 and 4, which deal with the periods 1900-1974 and 1975-present respectively, we notice the coefficient for time changes from 0.015 to 0.049. This suggests that the rate of change is increasing more rapidly than in previous years.

Using model 4, an expected rise by approximately 0.05 degree per year is predicted. The predicted temperatures in Fahrenheit for future years are as follows:

2017: 49.6  
2020: 49.7  
2050: 51.2  
2100: 53.7  
2150: 56.1

Ultimately, based on the data we gathered, cleaned, modeled and interpreted we believe that our predictions would be reliable from the methodologies we implemented. However, after performing all of our work on the data, we believe there are a number of factors we may have left out, therefore our confidence in these predictions may be lacking. We did not factor time, but implemented as a numerical value. There also may have been other variables that could have contributed to the data and that we could have looked at. By performing some principle component regression from other variables factored in such as pollution footprints or industrial growth rate from developing countries, our models may have been better.