Max Wagner

602 Project – Part 2

**Summary**:

Berkley Earth has gathered climate and weather data from 1750 until present day and has compiled it into multiple csv files. I would like to look at temperature over time, which regions have changed the most, and if specific types of regions are more prone to temperature change.

**Data**:

As stated above, the data has been compiled into multiple csv files by Berkley Earth. The formatting is not perfect and needs cleaning. The csv files are split into sections by location. There are sections for; individual stations, cities, states, countries, and hemispheres.

The cleaning will be done using base R and the packages stringr, dplyr, and tidyr from R. This step will also remove data points that are not relevant to the project. The remaining cleaned csv files will be saved for later use in Python.

**Analytics and Visualization**:

The first step is to find the average overall temperature by year, and plot it as an initial visualization to base the others on. This gives a global general trend which can then be broken down into smaller segments later on. I can continue to plot temperature change over time next by hemisphere, and then by country. If a specific country or region stands out, the major cities in that region can also be plotted. The data sets also include uncertainty, which can be plotted when necessary.

In addition to the overall view, I would like to look at specifically New York City and how the climate has changed in relation to time. This will allow a closer look at data on particular months or seasons and how they are affected.

An optional extra could be using ggmap from R in order to overlay the temperature data over maps in order to further visualize the changes.

**Resources:**

<http://berkeleyearth.org/data/>

<https://cran.r-project.org/web/packages/>