Introduction:

Question 3: The questions of interest here is to compare and contrast data from two different data sets. The first data set is a series of over 570,000 temperature observations from countries all over the world. The second data set contains over 200,000 observations from 99 large cities around the world. For each data set we analyzed the differences in maximum and minimum temperatures from Janurary 1, 1900 to December 31, 2012. We also created a subset of the temperatures recorded in the US for the previously described time period. We will produce a graph showing the annual change in average temperature over time, as well as show when the greatest year to year change occurred. The final piece of explanatory analysis was to demonstrate graphically the greatest temperature differential for each of the 99 cities in the city data set and graph that difference for the top 20 cities who had the greatest difference. We will then compare the city data chart to the country data chart and comment on the findings.

Code Explanations:

Tim – If you want to do this for your parts, I’ll do it for mine.

City Temp Data:

CityTemperature.R

Here we are importing the raw data from our working directory and putting it into a data frame for ease of manipulation in subsequent steps.

CityDataCleanup.R

This section takes the raw data, changes the date formats to a consistent format and then creates a new data frame with the dates specified in the question of interest. We also change a variable name to make it easier to work with as we begin to summarize and manipulate the data.

CityMaxandMin.R

Short section to calculate Maximum and Minimum temperatures for each city in the data set, and then create an output data frame to generate the graphics from.

Top20Cities\_Plot.R

Generate the graphics indicated in the question of interest section.

Answers to Questions of Interest:

I’ll do 3 iii, and iv

iii – The questions of interest for this section pertain to the cities of the world with the largest temperature variations. 3 of the top 5 are in China, with 1 each in Russia and Canada. While outside the scope of the current question of interest, it’s interesting that all these cities all above 40 degrees North latitude, and have generally extremely cold winters, while summers are usually mild, although occasionally get very warm, hence the differential.

Iv – The comparison of the breakdown between country temperature differential and city differential are interesting, as the top 3 cities don’t appear in the top 2 countries at all. After that, Russia and Canada are represented in the next couple of rankings, with major cities coming in the top 5 for each of them. Further investigation of both data sets shows why this might be. Neither of the top 2 countries have a city listed in the cities list, and so their data doesn’t appear. Looking at a map shows the latitude of both countries to be almost entirely abouve 40 degrees North, which is strongly correlated with wide temperature variations. As with the city data referenced above, cold winters and occasionally oppressively hot summers account for much of the difference here.

In general, addressing the questions of interest in the data sets was an opportunity to extend our data science skills by cleaning, pre-processing, analyzing and plotting data. Thank you for the opportunity to present our work.