Data Processing:

Since the weather data APIs found required purchasing access to historical data, we ended up pulling historical data in csv format from the NOAA Climate data archive.

The data set pulled were composed of daily weather data for the major international airports of the three cities sampled, Chicago, Phoenix, and Seattle. However, the data set contained gaps in the average daily temperature (TAVG) value. To overcome this limitation, I wrote some conditional logic that looked to see if the TAVG value was missing, and if it that was the case, set the TAVG to the average of the daily high and daily low temperature values.

The next issue arose with the data values. We needed to first get the average monthly temp for every year and then the average monthly temp across all five years. In order to accomplish that we needed to split the date data and create year and month columns for each record. We accomplished this by first splitting the date value year out and placing it into its own column. Then we split the month from the date value and placed it into its own column.

After adding all of the columns needed for processing we performed two groupby functions. First we grouped by year and month value to get the mean monthly temp for each year. Then we grouped by month and took the mean of the month value to get the average monthly temp across all five years.

Lastly, we took all three city dataframes and joined them into a single dataframe from which to plot our graphs.

Analysis:

One of the factors we compared against the health data was weather. The goal was to see if local temps have an effect on the overall health of those communities. To make the comparison, first we took the daily temperature data for a five year period (2010-2015) for all three cities. We then calculated the average monthly temperature for each individual year and lastly calculated an average monthly temperature across all five years.

Over the five year period, Phoenix and Chicago seem to experience parallel temperatures sets with Phoenix at a fairly steady 20F hotter annually. They also experience the largest annual temperature swings. Phoenix min average temp occurs in January at approximately 57F then July tops out at an average of 98F. An average 41 degree swing. Chicago’s lowest average temp is again in January at 22F. Average highs hit in July at 77F. This equates to 55 degree annual average temperature swing for Chicago.

Lastly we have Seattle. Seattle experience a totally different weather pattern than Phoenix and Chicago. Though the average low temperature seems to strike in January at 44F, their average high temperature is experienced in August at 69F. This gives Seattle the lowest average annual temperature swing of just 25F. Seattle winters average right in between those experienced in Chicago (coldest) and those of Phoenix (hottest). Seattle also experiences the coolest Summers with the coolest average annual temps of all three cities from mid April through mid September.

When comparing the weather data to the health data, I noticed no real trend in Diabetes and Cancer mortality rates and the average annual temperature swing of the three cities. Chicago, with the highest annual temperature swing, had the highest cancer and diabetes death rates but the lowest suicide deaths. Phoenix, with the second highest average temperature swings came in with the highest Suicide, second highest diabetes, and lowest Cancer deaths. Seattle, with the least annual average temperature swing came in second for cancer and suicide, and with the lowest diabetes deaths.

In conclusion, it seems to be no real correlation between the average annual temperature swing experience by people in communities and deaths from diabetes, cancer, and suicide. While Seattle, with the least average temperature swing also displayed the least extreme suicide, diabetes, and cancer death rates, it wasn’t enough data to accurately determine there is a correlation between temperature and these death rates.