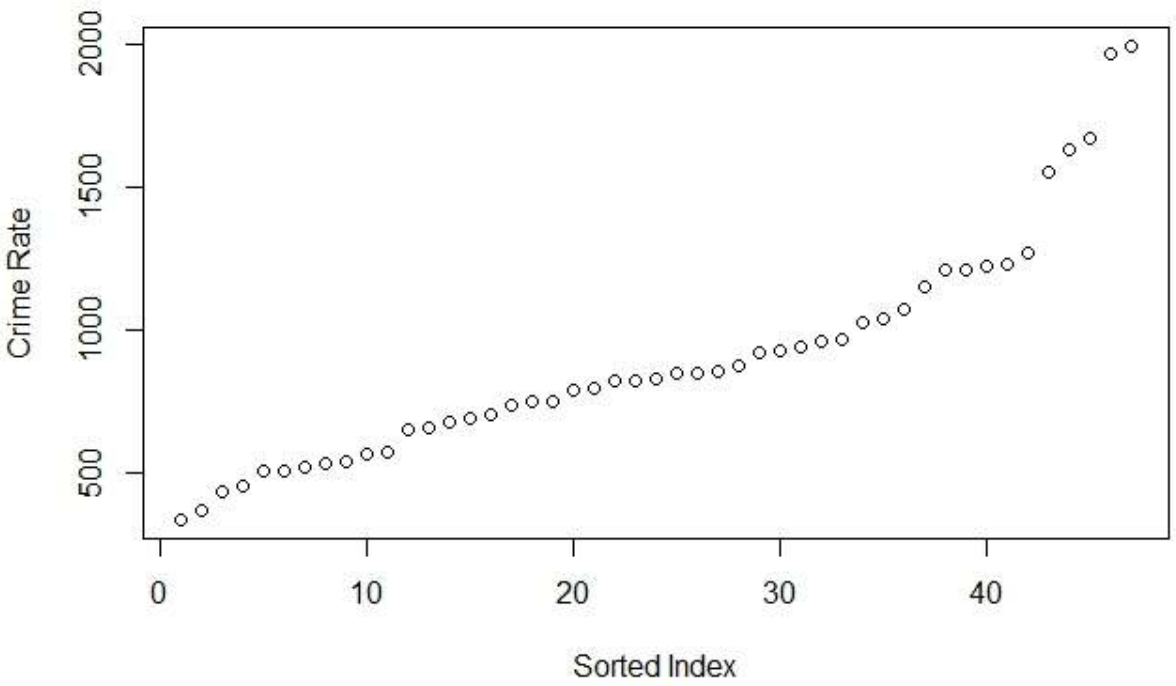
Bryson Cook
ISYE6501, Spring 2018
Homework 3

Question 5.1
Using crime data from http://www.statsci.org/data/general/uscrime.txt (description at http://www.statsci.org/data/general/uscrime.html), test to see whether there is an outlier in the last column (number of crimes per 100,000 people). Is the lowest-crime city an outlier? Is the highest-crime city an outlier? Are there others? Use the `grubbs.test` function in the `outliers` package in R.

To visually check for outliers, the data was sorted into ascending order and plotted, as shown below.



From this plot, it does not appear like the lowest crime rate is an outlier. It does appear there are outliers at the highest crime levels, though it is not certain particularly which. 1969 and 1993 are the highest crime rates, and are well above the next highest crime rate cluster, which begins at 1674 and includes the next two points. So one could also state that the top 5 results are outliers from the rest, as there is a large gap between points 42 and 45 (1272 vs 1555).

To be more thorough, the Grubbs test is applied to the upper end. If the Grubbs test determines that the point is an outlier that point is removed and the test is run on the next point until a P value of 1 is reached, signifying that the point is not an outlier. The P Values for the upper range are as follows:

```
     [,1]       [,2]
[1,] 1993 0.07887486
[2,] 1969 0.02847821
[3,] 1674 0.17807968
[4,] 1635 0.11389923
[5,] 1555 0.10815574
[6,] 1272 1.00000000
[7,]   NA         NA
[8,]   NA         NA
```

Here, the 6th point in question, 1272, was reached and determined to not be an outlier, though the previous 5 point are determined to be outliers, which reinforces the earlier visual assessment.

The process was then repeated for the lower range, with no points showing as an outlier, as the first point's pvalue equals 1. This means that no points were determined to be outliers and no data was removed.

```
     [,1] [,2]
[1,]  342    1
[2,]   NA   NA
[3,]   NA   NA
[4,]   NA   NA
[5,]   NA   NA
[6,]   NA   NA
[7,]   NA   NA
[8,]   NA   NA
```

Question 6.1

Describe a situation or problem from your job, everyday life, current events, etc., for which a Change Detection model would be appropriate. Applying the CUSUM technique, how would you choose the critical value and the threshold?

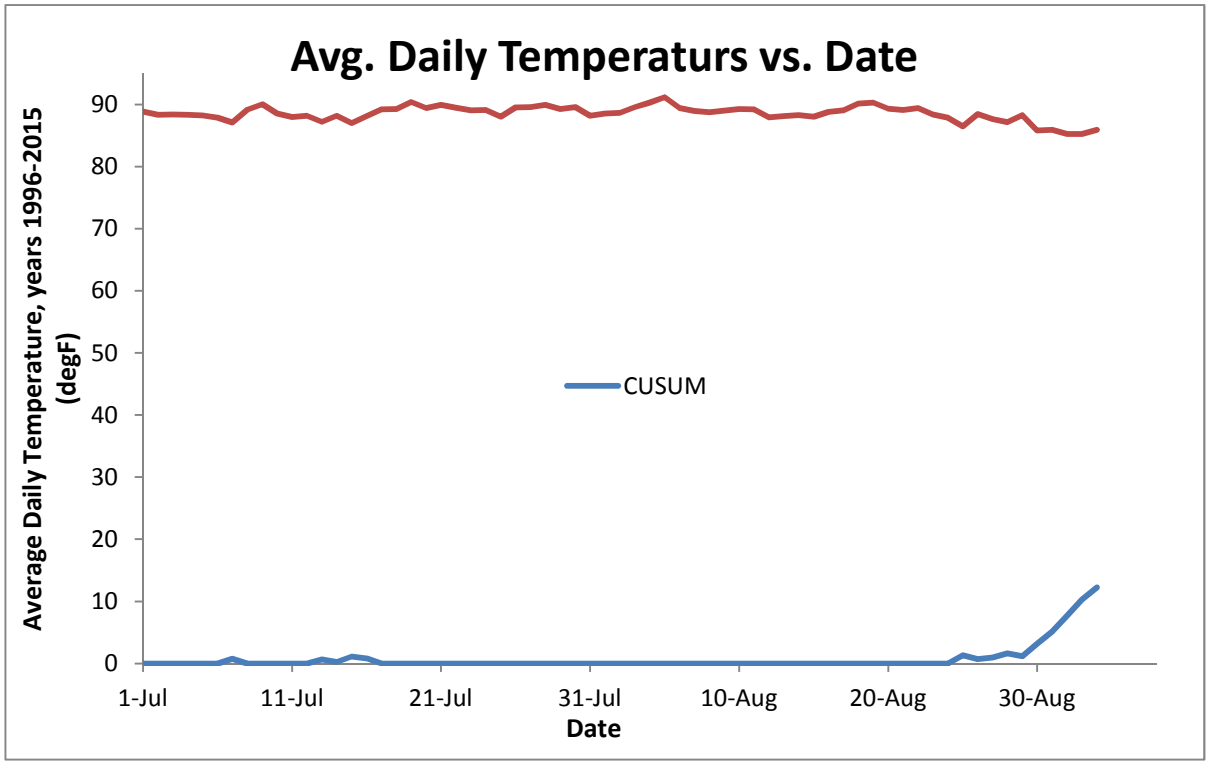
I work for a natural gas turbine OEM and the CUSUM approach would be good for implementing analytics to track the degradation, or decrease in output over time, of a unit. We have programs that record historical data of the gas turbines and a CUSUM approach could be applied to this data. When the threshold is reached, it could signify when a unit has degraded enough that a compressor wash or other maintenance is advisable to recover the lost performance or signal that the degradation is higher sooner than predicted and thus should be investigated further. The thresholds could be based off our OEM degradation curves, and since we would be using corrected performance (i.e. correcting for ambient conditions such as temperature, barometric pressure, etc.), the C value would be small as there wouldn't be as much variation in the data.

Question 6.2

1. Using July through October daily-high-temperature data for Atlanta for 1996 through 2015, use a CUSUM approach to identify when unofficial summer ends (i.e., when the weather starts cooling off) each year. You can get the data that you need from the file `temps.txt` or online, for example at http://www.iweathernet.com/atlanta-weather-records or https://www.wunderground.com/history/airport/KFTY/2015/7/1/CustomHistory.html . You can use R if you'd like, but it's straightforward enough that an Excel spreadsheet can easily do the job too.

I averaged the years together to get the daily average, then applied the CUSUM method. With mu equal to the average of 1 JUL – 15 AUG, C equal to 1 degF, and T equal to 2.5 degF, the unofficial last day of summer is 29 AUG.

- There is a slight cooling in the middle of July that, if C = 0, makes it appear that the summer has cooled down. However, even moving to C = 1 removes these false positives.



2. Use a CUSUM approach to make a judgment of whether Atlanta's summer climate has gotten warmer in that time (and if so, when).

If we take 29 AUG as the unofficial last day of summer, then applying the CUSUM method to the average temperature between 1 JUL and 29 AUG of each year should give an estimate of the change in summer temperatures over the years. However, using a threshold of 2 degF and a C value of 1 degF, it doesn't appear that a good determination can be made. Applying both and increasing and decreasing CUSUM approach to the data, there are years when the average summer temperature rises (2011-2012), but there are also years when the average temperature falls (2004-2005 and 2013-2014), as shown below.