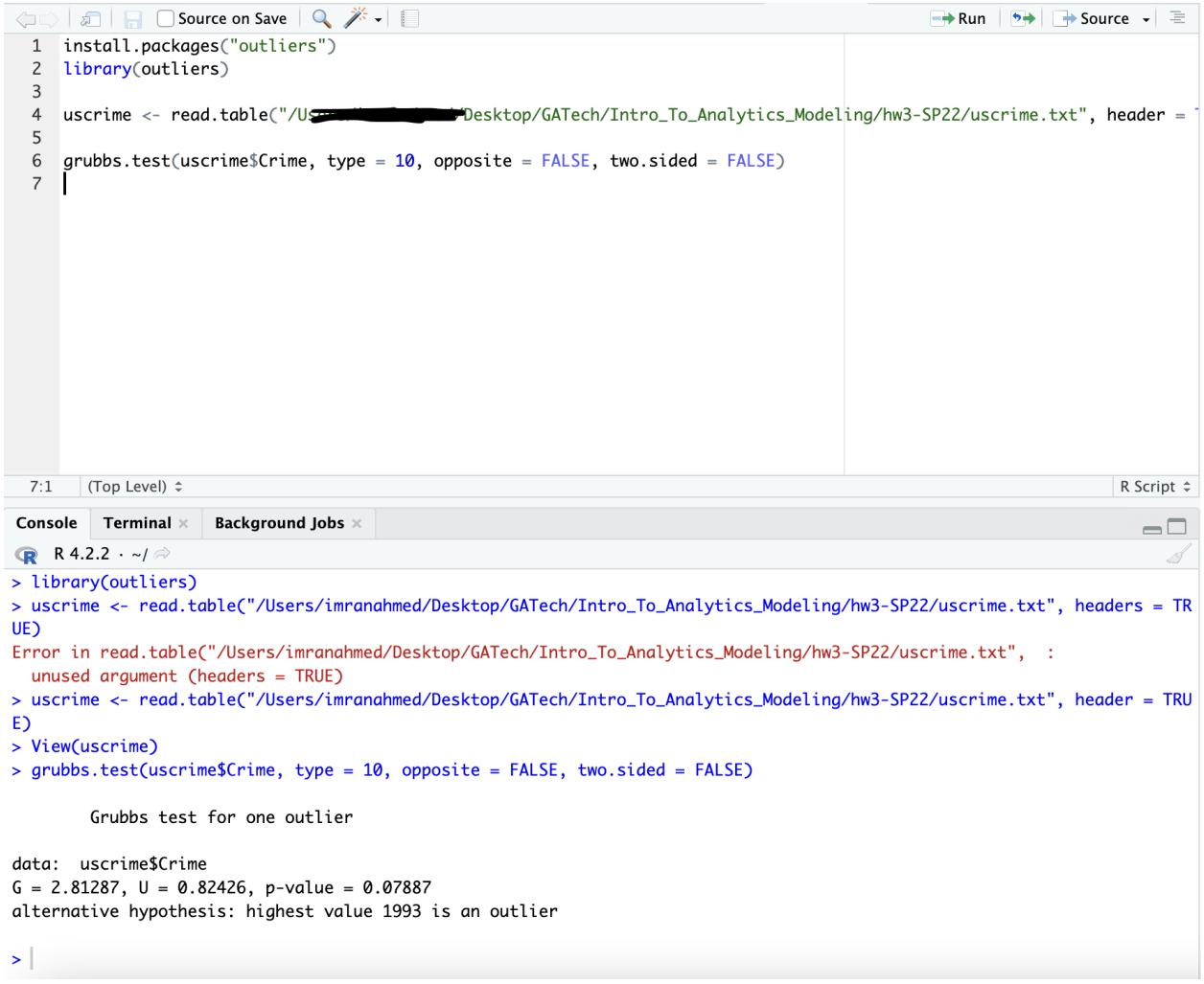
Question 5.1

Using crime data from the file `uscrime.txt` (http://www.statsci.org/data/general/uscrime.txt, description at http://www.statsci.org/data/general/uscrime.html), test to see whether there are any outliers in the last column (number of crimes per 100,000 people). Use the `grubbs.test` function in the `outliers` package in R.

ANSWER:

I loaded the outliers package and used the grubbs.test function on the data. The highest value in the data set was 1993, but the p-value was .07, so it was not significant at 95% confidence. Therefore, there is no outlier at the high end of the data. I also tested if the lowest value was an outlier (by changing the code to opposite = TRUE) and the p-value reflected that the lowest value was not statistically significant either. Therefore there seem to be no outliers in this data set.



Question 6.1

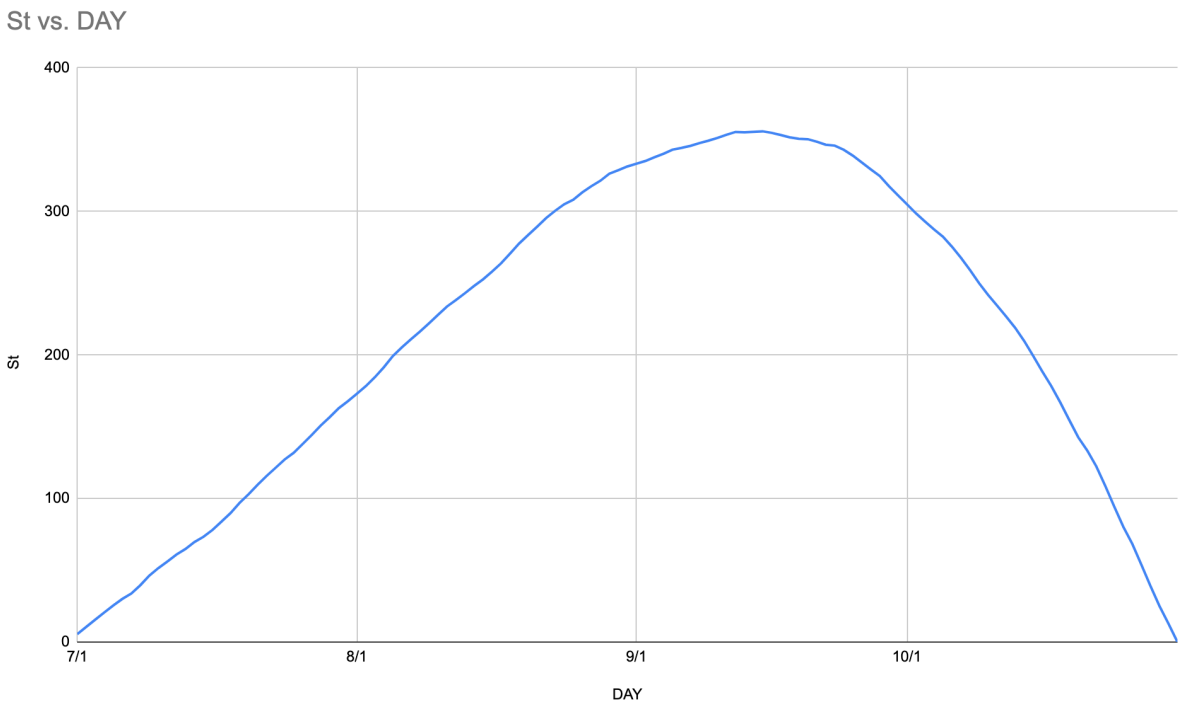
Describe a situation or problem from your job, everyday life, current events, etc., for which a Change Detection model would be appropriate. Applying the CUSUM technique, how would you choose the critical value and the threshold?

In my job, I often have to understand if advertising creative has started to wear out. Wear out happens when the creative is no longer effective because it has been in the market for too long and lacks a freshness. You can use CUSUM by plotting the KPI metric (Clickthrough rate, View Rate, Conversion Rate, etc.) over time and see how the performance trends. If the creative begins to fall too far below your benchmark rates, you can have an appropriate C and T that flags this and signals wearout.

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.**

ANSWER: For this question. I used the daily average from 1999 to 2015 to measure Xt. In other words, for July 1, my Xt was the average temperature for July 1 from 1995 - 2015. For U, I used the average for every day from 1995 - 2015. In other words, it was just the average daily temperature for the whole data set. For the purposes of understanding when 'unofficial summer' ends, I actually did not feel the need to estimate a value of C, since I would be able to use the relative change in St to understand when summer ends. For example, in the chart below, the vertex of the parabola represents the end of unofficial summer. When St starts to decline, this is when the daily temperature starts to drop below the average temperature. This was around September 15, which subjectively feels appropriate for a southern state which may have a summer season that lasts a bit into the fall.



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).**

For this approach I created an St control chart for every year and compared the years to each other. To do this, I used the temperature for the year as Xt. My U was the avg temperature for every day in the whole data set. Then I compared the charts to the average chart that I made above. For this example, I used a C = 5, but the comparative approach would've worked with any C. You can See that St accumulates much quicker for 2015 than it does for the average of the whole data set. This indicated that in 2015, the difference between your daily temperature and the average for the whole data set is greater, meaning you have more days where the temperature is over 5 degrees higher than the average for the whole data set.

St(1996), St(2015), St (All) and St(1999)