

### Question 1.

*Assumption that we're using a .05 significance level when answering the questions below.*

There are a number of ways to test for outliers using the `grubbs.test` function. The answer to the outlier question is dependent on the testing method used.

I've configured the `grubbs.test` to check if lowest and highest values are two outliers on opposite tails of the sample and to treat it as a two-sided test:

```
grubbs.test(uscrime_df$Crime, type = 11, opposite = FALSE, two.sided = TRUE)
```

This results in a p-value  $< 2.2e-16$  which allows us to reject the null hypothesis and accept the alternative hypothesis that 342 (lowest-crime) and 1993 (highest-crime) are outliers.

However, if you test for a high/low outlier individually you receive different results. Checking if the highest value is an outlier with the following function:

```
grubbs.test(uscrime_df$Crime, type = 10, opposite = FALSE, two.sided = FALSE)
```

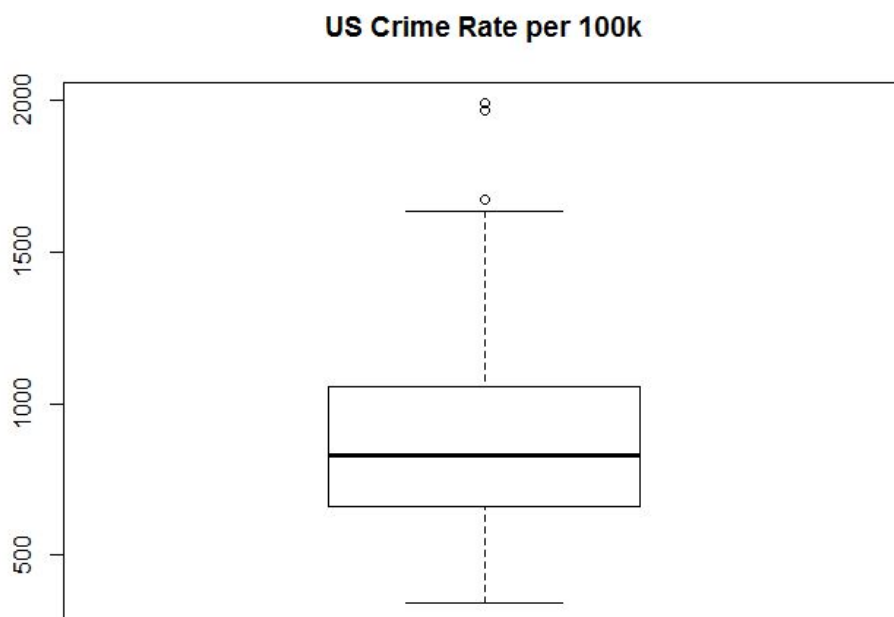
The result is a p-value of 0.07887 which doesn't allow us to reject the null hypothesis.

Checking if the lowest value is an outlier with the following function:

```
grubbs.test(uscrime_df$Crime, type = 10, opposite = TRUE, two.sided = FALSE)
```

The result is a p-value of 1, which doesn't allow us to reject the null hypothesis.

Also, if you use a box plot to plot crime, 342 doesn't appear as an outlier



## Question 2.

A change detection algorithm would be useful in tracking refrigerator temperature. We'd want to track to make sure that the temperature stays between 34 and 40 degrees. This would ensure that refrigerated food maintains a safe temperature. I'd place a low threshold (T) value with a low C because I'd want the algorithm to be more sensitive and allow me to respond quickly if temperatures exceeded the safety threshold.

## Question 3.1

Because we're interested in when summer ends in Atlanta (not when it experiences a heat wave), I've used one-sided CUSUM focused on a decrease from the mean.

$$\text{Equation: } S_t = \max\{0, S_{t-1} + (x_t - \mu - C)\}$$

With the assumption that  $\mu$  is 83° and a low C of 1 (keeping the CUSUM model more sensitive to variation), I found that a threshold (T) level of 50 produced a fairly consistent "summer ending date" across the years while also minimizing the number of "false alarms" where the threshold was briefly exceeded.

Fig. 1: Table of "Unofficial" End of Summer Dates - when T is exceeded and stays exceeded

"Unofficial" End of Summer	
Avg.	6-Oct
1996	1-Oct
1997	27-Sep
1998	9-Oct
1999	28-Sep
2000	18-Sep
2001	28-Sep
2002	29-Sep
2003	29-Sep
2004	21-Sep
2005	12-Oct
2006	30-Sep
2007	14-Oct
2008	8-Oct
2009	30-Sep
2010	4-Oct
2011	3-Oct
2012	8-Oct
2013	29-Sep
2014	4-Oct
2015	26-Sep

Fig. 2: Plot of St values over the years:

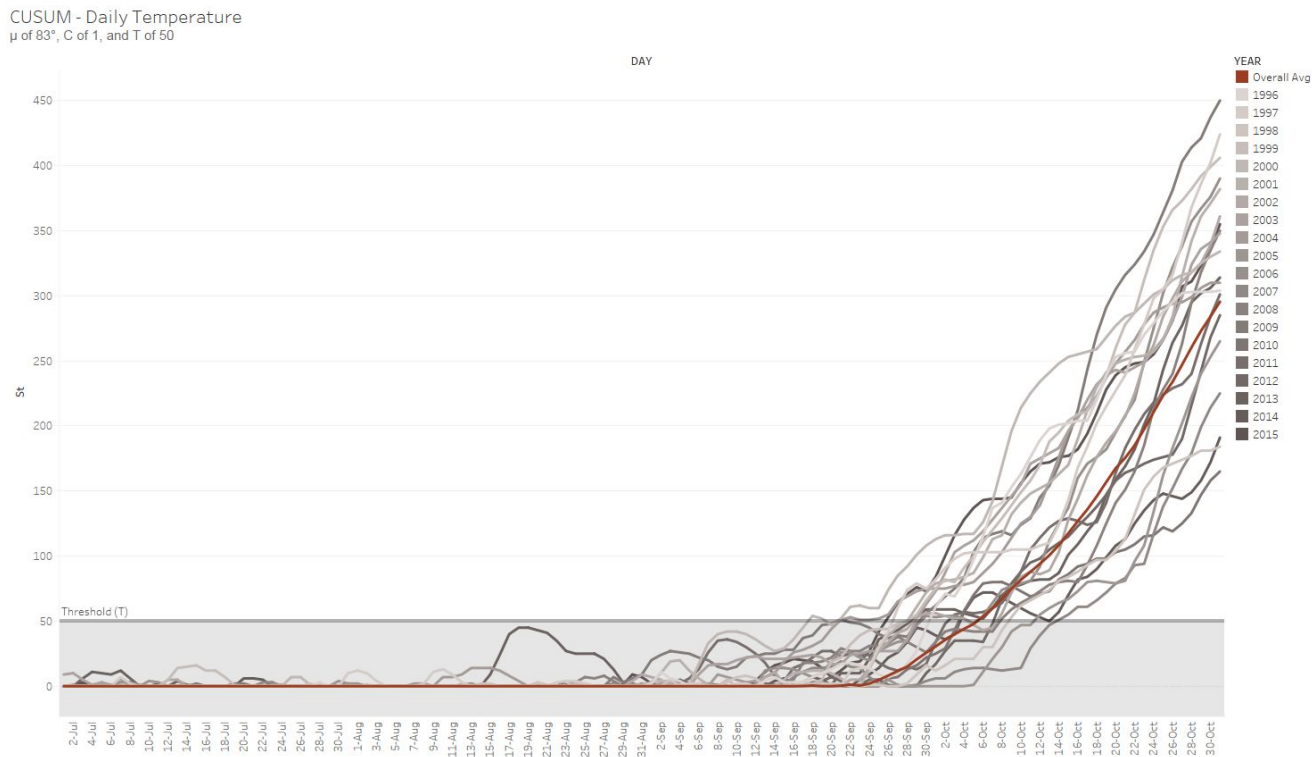


Fig. 3: Chart of St values across the years (red text exceeds threshold)

### Question 3.2

Using the tables and charts from my answer to 3.1, I'd make the judgement that Atlanta's summer climate hasn't gotten warmer. The summer end dates appear to fall within normal variation.

*Analysis was done using Alteryx and Tableau*

