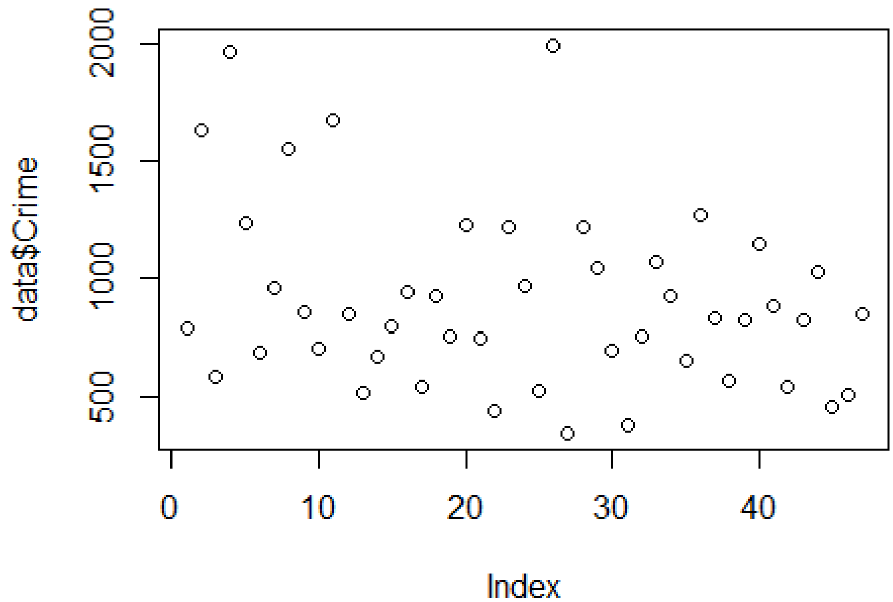
5.1

First we would like to visualize the data to see if we can identify any potential outliers. Given the various graphs, it seems we would expect the highest crime rates to be outliers, especially given the box and whiskers plot.

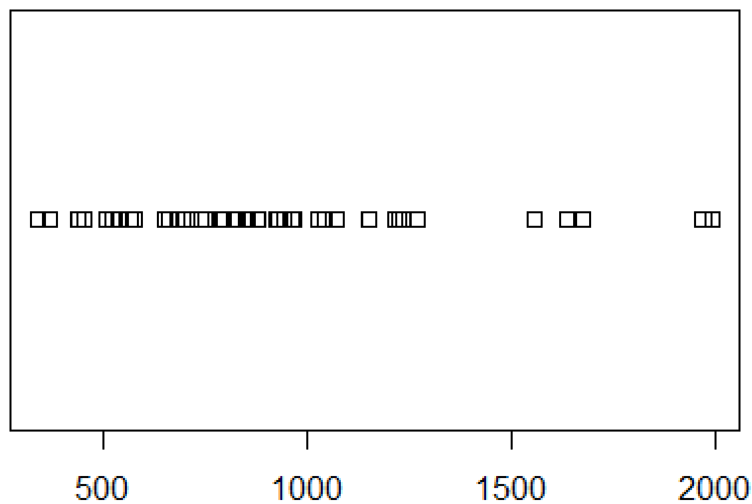
Type 10 test which just checks for outlier on the top end of the dataset yielded a p-value of 0.07887. Since p approaching 0 indicates stronger evidence of an outlier, 1993 as the outlier seems likely, especially given the graphs above.

On the other hand, type 11 test tests both tip and the tail, but yields a p-value of 1 which indicates there is weaker evidence of the outliers.

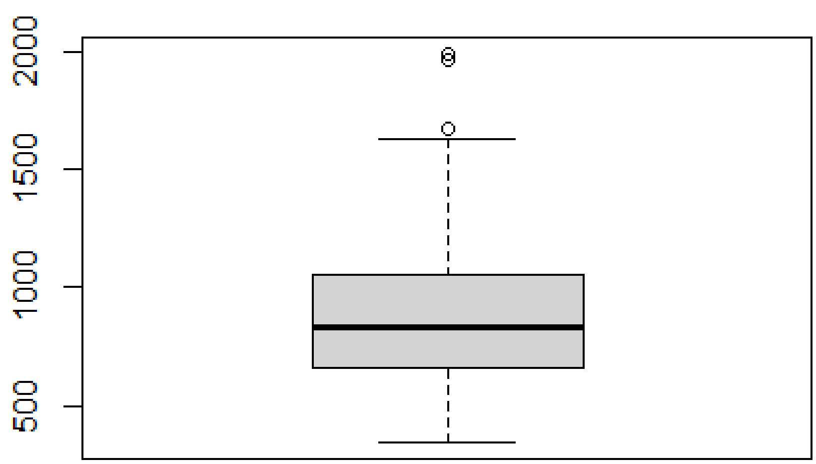
```
##
##  Type 10 Grubbs test for one outlier
##
## data:  data$Crime
## G = 2.81287, U = 0.82426, p-value = 0.07887
## alternative hypothesis: highest value 1993 is an outlier

##
##  Type 11 Grubbs test for two opposite outliers
##
## data:  data$Crime
## G = 4.26877, U = 0.78103, p-value = 1
## alternative hypothesis: 342 and 1993 are outliers
```




```
library(outliers)
data <- read.table('uscrime.txt',header=TRUE)

plot(data$Crime)

#index on the x-axis in this case doesnt really mean anything, so lets try an
other view
plot(data.frame(data$Crime))
```




```
boxplot(data$Crime)
```



```
#default type=10 (test for one outlier)
grubbs.test(data$Crime)

#type=11 (test for two outliers on opposite tails)
grubbs.test(data$Crime,type=11)
```

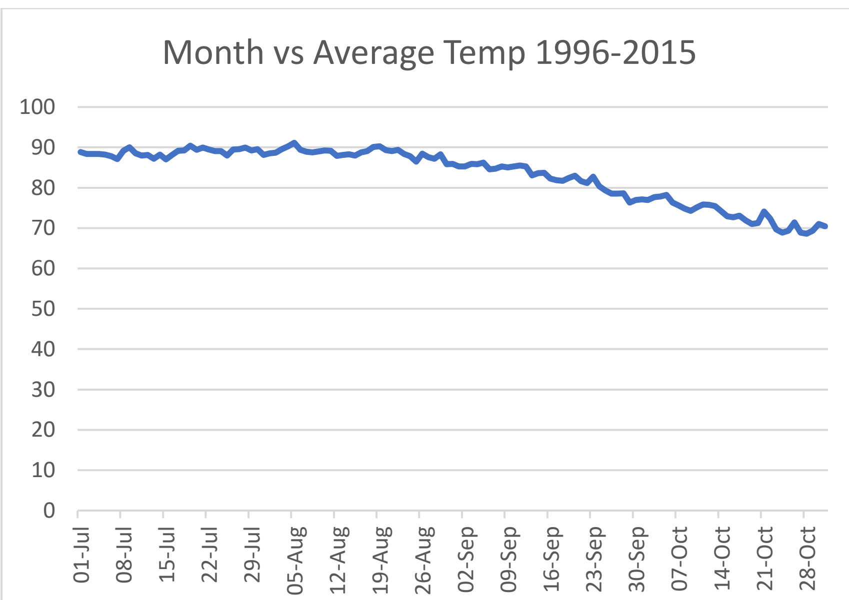
6.1

I work for a company that manufactures machines for packaging automation. When testing and qualifying these machines on our site, we typically have to test a wide range of materials to identify what yields the best results. However, each test run can have variance and outside factors influencing it.

CUSUM can be used by recording a machine performance metric (throughput or efficiency) across a wide range of materials (example: plastic with different stiffness levels) and recording machine performance while slowly changing the parameter. We can then analyze when change in machine performance occurs using CUSUM and at what plastic stiffness. This could help identify the material properties that would handle best on the machines

6.2.1

Taking an average of temperature in each month over all the years in the dataset, we get the following graph. We can see temperature consistently averages around the 90 mark until we get to middle of September after which a slow decline starts.



Given this, we can calculate our mean from July to end of august. Standard deviation is also calculated in this range to get an idea of typical deviation from the mean value we can expect. I used C = to 1 standard deviation to filter out any noise. Looking at the data and real world experience, temperatures tend to have considerable ups and downs even within seasons, so we do not need our C to be very small. T was chosen as 4x the standard deviation to avoid false positives on general variation of temperature.

With this we see unofficial start of summer typically start in the first half of September across the years.

68	31-Aug	4.142857	0	0	0.714286	0	0.785714	22	0	0	0	5.25	0.482143	0	11.64286	2.196429	0	10.64286	0	0	19.375
69	01-Sep	9.214286	0	0	0	7.571429	2.678571	23	0	0	0	5.5	2.964286	1.910714	23.30357	0	0	6.053571	0	0	17.57143
70	02-Sep	21.28571	0	0	0	13.14286	9.571429	24	0	1.375	0	4.75	6.446429	0	27.96429	0	0	3.464286	0	0	12.76786
71	03-Sep	19.35714	0	5.928571	0.714286	17.71429	18.46429	20	0	1.75	0	5	4.928571	0	31.625	0	0	4.875	0	0	8.964286
72	04-Sep	20.42857	0.214286	0	0	17.28571	19.35714	10	0	1.125	0	2.25	0.410714	0	31.28571	2.732143	7.571429	14.28571	0	0.267857	2.160714
73	05-Sep	18.5	0.428571	0	0	22.85714	11.25	2	0	0	1.178571	4.5	0	0	30.94643	4.464286	17.14286	15.69643	0	0	0.357143
74	06-Sep	14.57143	0	0	0	43.42857	5.142857	0	7.892857	0.375	2.357143	4.75	1.482143	0	28.60714	1.196429	29.71429	17.10714	0	0	0.553571
75	07-Sep	10.64286	0	0	0	64	0.035714	0	13.78571	0	1.535714	10	1.964286	0	27.26786	0	48.28571	11.51786	0	0	0.75
76	08-Sep	6.714286	0	0	0	75.57143	0	0	13.67857	8.375	0	12.25	0.446429	0	22.92857	0	62.85714	9.928571	0	0	1.946429
77	09-Sep	0.785714	0	3.928571	7.714286	82.14286	0	0	12.57143	5.75	0	14.5	0	0	21.58929	0	69.42857	13.33929	0	0	0.142857
78	10-Sep	1.857143	1.214286	8.857143	5.428571	86.71429	0	0	14.46429	3.125	0	14.75	0	0	24.25	0	73	15.75	0	0	0.339286
79	11-Sep	0.928571	0.428571	10.78571	3.142857	89.28571	0	0	15.35714	0.5	0	16	0	0	31.91071	0	74.57143	19.16071	0	0	0
80	12-Sep	0	0	9.714286	0.857143	89.85714	0	0	15.25	0.875	0	19.25	0.482143	0	33.57143	0	75.14286	24.57143	0	0	7.196429
81	13-Sep	7.071429	0	3.642857	0	89.42857	0	10	12.14286	3.25	0	34.5	1.964286	0	36.23214	0.732143	73.71429	29.98214	0	0	17.39286
82	14-Sep	13.14286	0	0	0	90	0	17	11.03571	5.625	0	39.75	7.446429	0	36.89286	0	69.28571	33.39286	0	4.267857	25.58929
83	15-Sep	12.21429	0	0	0	96.57143	4.892857	11	9.928571	14	0	43	11.92857	3.910714	40.55357	0	70.85714	32.80357	0	6.535714	30.78571
84	16-Sep	15.28571	0	0	3.714286	108.1429	9.785714	8	9.821429	20.375	0	45.25	22.41071	6.821429	41.21429	0	86.42857	33.21429	0	2.803571	36.98214
85	17-Sep	18.35714	0	0	7.428571	121.7143	10.67857	7	9.714286	21.75	0	45.5	30.89286	20.73214	50.875	0	95	36.625	0.517857	0	39.17857
86	18-Sep	25.42857	0	0.928571	10.14286	135.2857	11.57143	11	9.607143	24.125	0	45.75	38.375	21.64286	53.53571	0	105.5714	44.03571	2.035714	0	41.375
87	19-Sep	31.5	0	3.857143	15.85714	137.8571	11.46429	16	6.5	27.5	0	52	42.85714	23.55357	62.19643	0	116.1429	49.44643	0	0	39.57143
88	20-Sep	37.57143	0	4.785714	32.57143	137.4286	9.357143	15	0.392857	35.875	0	64.25	48.33929	27.46429	63.85714	0	121.7143	56.85714	0	0	35.76786
89	21-Sep	44.64286	0	5.714286	38.28571	147	5.25	16	0	42.25	0	74.5	56.82143	35.375	67.51786	0	123.2857	58.26786	6.517857	0	43.96429
90	22-Sep	48.71429	11.21429	0.642857	51	160.5714	0.142857	24	5.892857	43.625	0	77.75	57.30357	34.28571	66.17857	0	130.8571	57.67857	4.035714	0.267857	53.16071

6.2.2

Looking at these changes over the years, we do not see any sort of trend indicating summer climate has gotten warmer in that time.