ISYE 6501 HW 3

Question 5.1

```
> setwd("~/Desktop/ISYE 6501/ISYE HW 3/crime_data")
> library(outliers)
> crimedata <- read.table("uscrime.txt", stringsAsFactors = FALSE, header = T)
> head(crimedata, 20)
```

	M	So	Ed	Po1	Po2	LF	M.F	Pop	NW	U1	U2	Wealth	Ineq	Prob	Time	Crime
1	15.1	1	9.1	5.8	5.6	0.510	95.0	33	30.1	0.108	4.1	3940	26.1	0.084602	26.2011	791
2	14.3	0	11.3	10.3	9.5	0.583	101.2	13	10.2	0.096	3.6	5570	19.4	0.029599	25.2999	1635
3	14.2	1	8.9	4.5	4.4	0.533	96.9	18	21.9	0.094	3.3	3180	25.0	0.083401	24.3006	578
4	13.6	0	12.1	14.9	14.1	0.577	99.4	157	8.0	0.102	3.9	6730	16.7	0.015801	29.9012	1969
5	14.1	0	12.1	10.9	10.1	0.591	98.5	18	3.0	0.091	2.0	5780	17.4	0.041399	21.2998	1234
6	12.1	0	11.0	11.8	11.5	0.547	96.4	25	4.4	0.084	2.9	6890	12.6	0.034201	20.9995	682
7	12.7	1	11.1	8.2	7.9	0.519	98.2	4	13.9	0.097	3.8	6200	16.8	0.042100	20.6993	963
8	13.1	1	10.9	11.5	10.9	0.542	96.9	50	17.9	0.079	3.5	4720	20.6	0.040099	24.5988	1555
9	15.7	1	9.0	6.5	6.2	0.553	95.5	39	28.6	0.081	2.8	4210	23.9	0.071697	29.4001	856
10	14.0	0	11.8	7.1	6.8	0.632	102.9	7	1.5	0.100	2.4	5260	17.4	0.044498	19.5994	705
11	12.4	0	10.5	12.1	11.6	0.580	96.6	101	10.6	0.077	3.5	6570	17.0	0.016201	41.6000	1674
12	13.4	0	10.8	7.5	7.1	0.595	97.2	47	5.9	0.083	3.1	5800	17.2	0.031201	34.2984	849
13	12.8	0	11.3	6.7	6.0	0.624	97.2	28	1.0	0.077	2.5	5070	20.6	0.045302	36.2993	511
14	13.5	0	11.7	6.2	6.1	0.595	98.6	22	4.6	0.077	2.7	5290	19.0	0.053200	21.5010	664
15	15.2	1	8.7	5.7	5.3	0.530	98.6	30	7.2	0.092	4.3	4050	26.4	0.069100	22.7008	798
16	14.2	1	8.8	8.1	7.7	0.497	95.6	33	32.1	0.116	4.7	4270	24.7	0.052099	26.0991	946
17	14.3	0	11.0	6.6	6.3	0.537	97.7	10	0.6	0.114	3.5	4870	16.6	0.076299	19.1002	539
18	13.5	1	10.4	12.3	11.5	0.537	97.8	31	17.0	0.089	3.4	6310	16.5	0.119804	18.1996	929
19	13.0	0	11.6	12.8	12.8	0.536	93.4	51	2.4	0.078	3.4	6270	13.5	0.019099	24.9008	750
20	12.5	0	10.8	11.3	10.5	0.567	98.5	78	9.4	0.130	5.8	6260	16.6	0.034801	26.4010	1225

I used the grubbs test function and used type = 11 because it will test for 2 outliers on opposite ends, used false for the opposite argument because it will test the largest value from the mean, and applied it to column 16 it found 1993 as the outlier from the crime data

```
>grubbs.test(crimedata[,16], type = 11, opposite = FALSE, two.sided = FALSE )
```

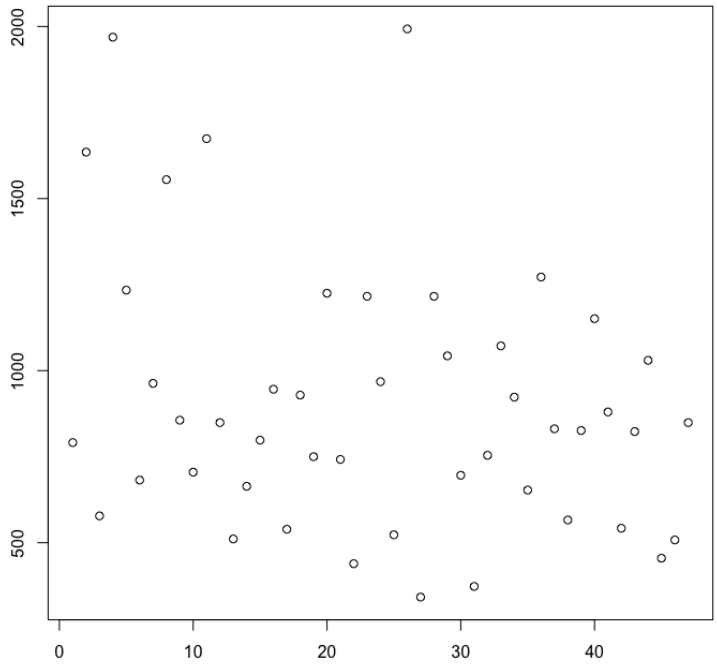
```
        Grubbs test for two opposite outliers

data:  crimedata[, 16]
G = 4.26877, U = 0.78103, p-value = 1
alternative hypothesis: 342 and 1993 are outliers
```

I plotted the crime data using plot

>plot(crimedata$Crime)

The p-value is >0.05, we reject the null hypothesis and accept <u>1993(max) as the outlier</u>, seen on the plot graph, and reject 342(min)- not an outlier



Question 6.2

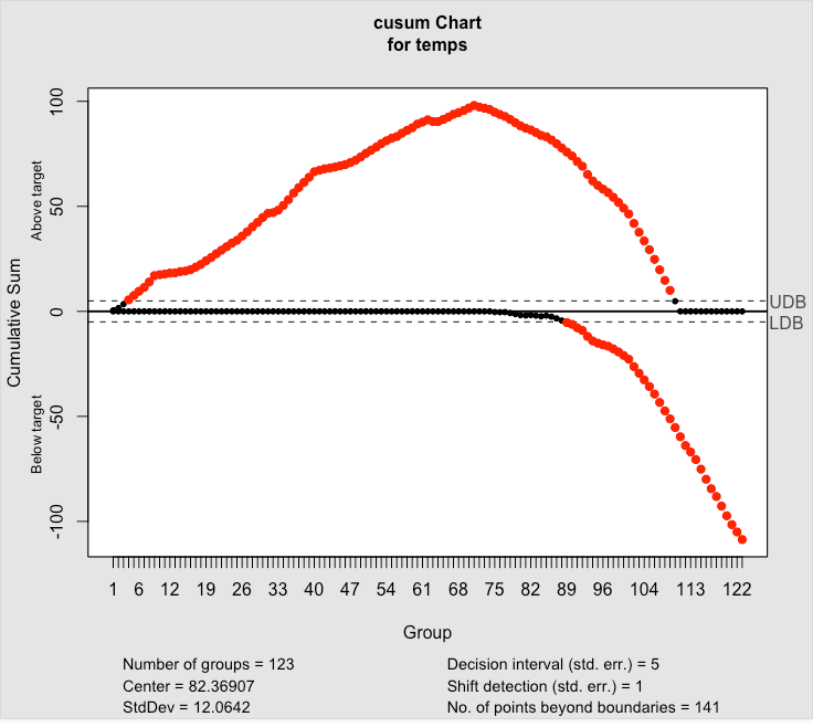
I set up the session, installed the qcc & data table package, and modified the data so the average temps of each day are in the data

> setwd("~/Desktop/ISYE 6501/ISYE HW 3/temps")
> library(qcc)
> library(“DT”)
>temps <- read.table("temps.txt", stringsAsFactors = FALSE, header = T)
>temps<-cbind(temps,rowMeans(temps[,-1]))

I plotted my average data as a cusum chart and it looks like the temperature exceeds the lower bound limit at group 89, and when I print the average temperature as a data table, group 89 is <u>September 27th</u> at 78.55 degrees F, which is the date the temperature starts to cool off, the <u>unofficial end of Summer in Atlanta.</u>

>cusum(temps)

```
List of 14
 $ call             : language cusum(data = temps)
 $ type             : chr "cusum"
 $ data.name        : chr "temps"
 $ data             : num [1:123, 1:22] 2 46 90 101 105 109 113 117 121 6 ...
  ..- attr(*, "dimnames")=List of 2
 $ statistics       : Named num [1:123] 84.9 86.4 88.5 88.9 89 ...
  ..- attr(*, "names")= chr [1:123] "1" "2" "3" "4" ...
 $ sizes            : int [1:123] 22 22 22 22 22 22 22 22 22 22 ...
 $ center           : num 82.4
 $ std.dev          : num 12.1
 $ pos              : num [1:123] 0.485 1.562 3.435 5.484 7.566 ...
 $ neg              : num [1:123] 0 0 0 0 0 0 0 0 0 0 ...
 $ head.start       : num 0
 $ decision.interval: num 5
 $ se.shift         : num 1
 $ violations       :List of 2
 - attr(*, "class")= chr "cusum.qcc"
```



>colnames(temps)[22] <- "Average Temperatures"
>datatable(temps)

Show 10 entries Search:

	DAY	X1996	X1997	X1998	X1999	X2000	X2001	X2002	X2003	X2004	X2005	X2006	X2007	X2008	X2009	X2010	X2011	X2012	X2013	X2014	X2015	Average Temperatures
81	19-Sep	79	91	80	79	84	82	80	84	78	90	79	82	81	74	92	77	81	84	84	87	82.4
82	20-Sep	79	95	82	68	87	84	86	87	73	90	73	81	79	81	96	82	79	86	83	89	83
83	21-Sep	78	89	82	79	77	86	84	82	75	90	75	78	75	79	95	86	85	73	87	77	81.6
84	22-Sep	81	70	88	72	73	87	77	75	80	86	82	86	84	84	92	80	87	82	82	76	81.2
85	23-Sep	84	80	84	75	81	88	82	81	84	87	86	83	82	83	91	83	81	82	77	81	82.75
86	24-Sep	84	82	81	78	84	69	73	80	82	88	84	89	78	85	88	82	78	71	78	74	80.4
87	25-Sep	87	66	82	81	82	66	69	82	81	85	75	87	82	87	93	88	82	67	77	67	79.3
88	26-Sep	84	70	84	82	68	72	75	82	79	77	78	84	80	85	76	86	86	78	74	71	78.55
89	27-Sep	79	64	87	78	71	75	75	82	72	86	79	85	77	80	81	84	88	79	78	71	78.55
90	28-Sep	75	68	80	80	75	78	79	73	78	85	81	85	86	83	76	79	86	77	74	75	78.65

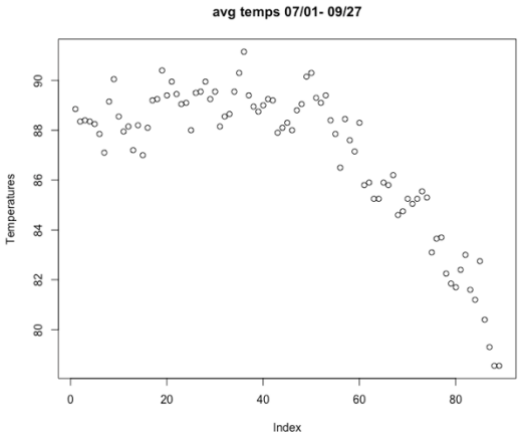
Showing 81 to 90 of 123 entries Previous 1 … 8 9 10 … 13 Next

>Summer_Data_DT <- head(temps, 89) View(Sep_27th_DT)

>plot(Summer_Data_DT[,22], main = "avg temps 07/01- 09/27", ylab = "Temperatures")

>cusum(Summer_Data_DT)

I created a separate data table from July 1st to September 27th (when unofficial summer ends) and plotted it, and used locator() to find the highest y value which came out to 91.14, in the average temperatures column, that is equivalent to 91.15 degrees F, on August 5th. I Also graphed it using cusum, which confirmed that summer has not been getting warmer. The temperatures since then have been around the same or lower, so summer in Atlanta has not gotten warmer. The plots on the graph also look consistently the same.




```
List of 14
 $ call             : language cusum(data = Summer_Data_DT)
 $ type             : chr "cusum"
 $ data.name        : chr "Summer_Data_DT"
 $ data             : num [1:89, 1:22] 2 35 66 73 76 79 82 85 88 5 ...
  ..- attr(*, "dimnames")=List of 2
 $ statistics       : Named num [1:89] 84.9 85.9 87.4 87.7 87.7 ...
  ..- attr(*, "names")= chr [1:89] "1" "2" "3" "4" ...
 $ sizes            : Named int [1:89] 22 22 22 22 22 22 22 22 22 22 ...
  ..- attr(*, "names")= chr [1:89] "1" "2" "3" "4" ...
 $ center           : num 85.1
 $ std.dev          : num 13.7
 $ pos              : num [1:89] 0 0 0.273 0.64 1.02 ...
 $ neg              : num [1:89] 0 0 0 0 0 0 0 0 0 0 ...
 $ head.start       : num 0
 $ decision.interval: num 5
 $ se.shift         : num 1
 $ violations       :List of 2
 - attr(*, "class")= chr "cusum.qcc"
```