ISYE-6501
Fall 2024

Homework 3

Question 5.1

Using crime data from the file uscrime.txt, 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.

Analysis:

After using multiple instances and variations of the grubbs.test function on the last column, here are my findings. If we use the test as two sided (because we would like to check for outliers higher than the majority of the points and lower than the majority of the points) and using type=11 instead of type 10 since we are testing for the maximum and minimum values as outliers, we can successfully reject the null that there are no outliers in the data set because the p-value (2.2e-16) is smaller than .05. This means that the points 342 and 1993 are in fact outliers.

Please see the code below with the output in blue.

CODE:

```
> library(outliers)
> data <- read.table("uscrime.txt", header=TRUE)
> head(data, 10)
    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
> grubbs.test(data[,16], type=11, two.sided=TRUE)

        Grubbs test for two opposite outliers

data:  data[, 16]
G = 4.26877, U = 0.78103, p-value < 2.2e-16
alternative hypothesis: 342 and 1993 are outliers
```

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?

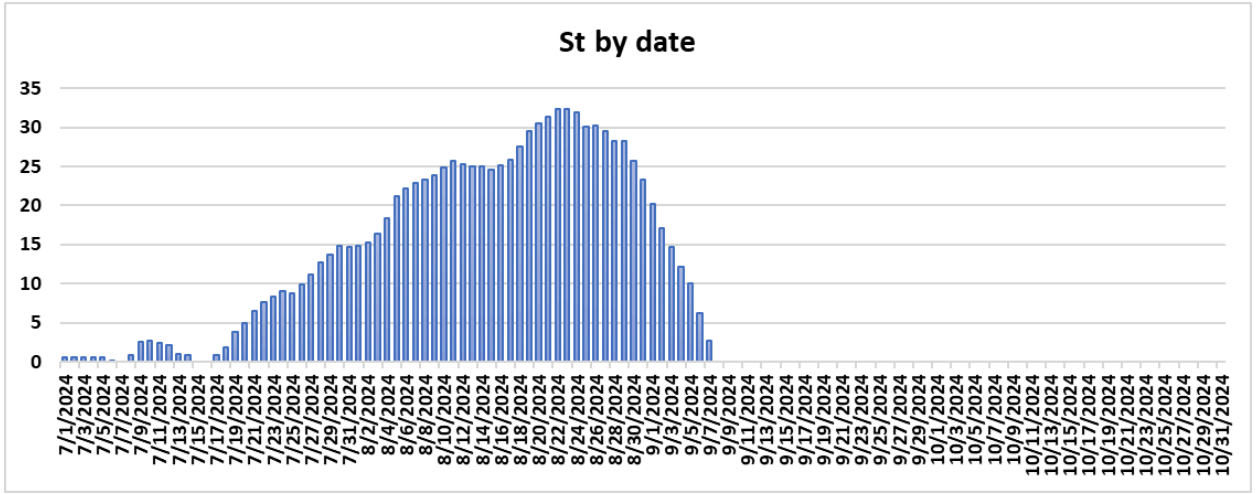
As a quality data analyst for a hospital, I can use a Change Detection model for monitoring changes in hospital-acquired infection rates throughout the months. After having looked at the data from the past few years, there is an average of 1 hospital-acquired infection per 1000 acute patient days. Having said that, I am going to choose a critical value of .5 infections per 1000 acute patient days. I decided to choose a small C value so that the model is more sensitive because some hospital-acquired infections can lead to very bad symptoms or even death in older patients. As for threshold I would have to see what the S(t) calculations show, I would like to choose a threshold that is near the high average of infections per 1000 patient days.

Question 6.2

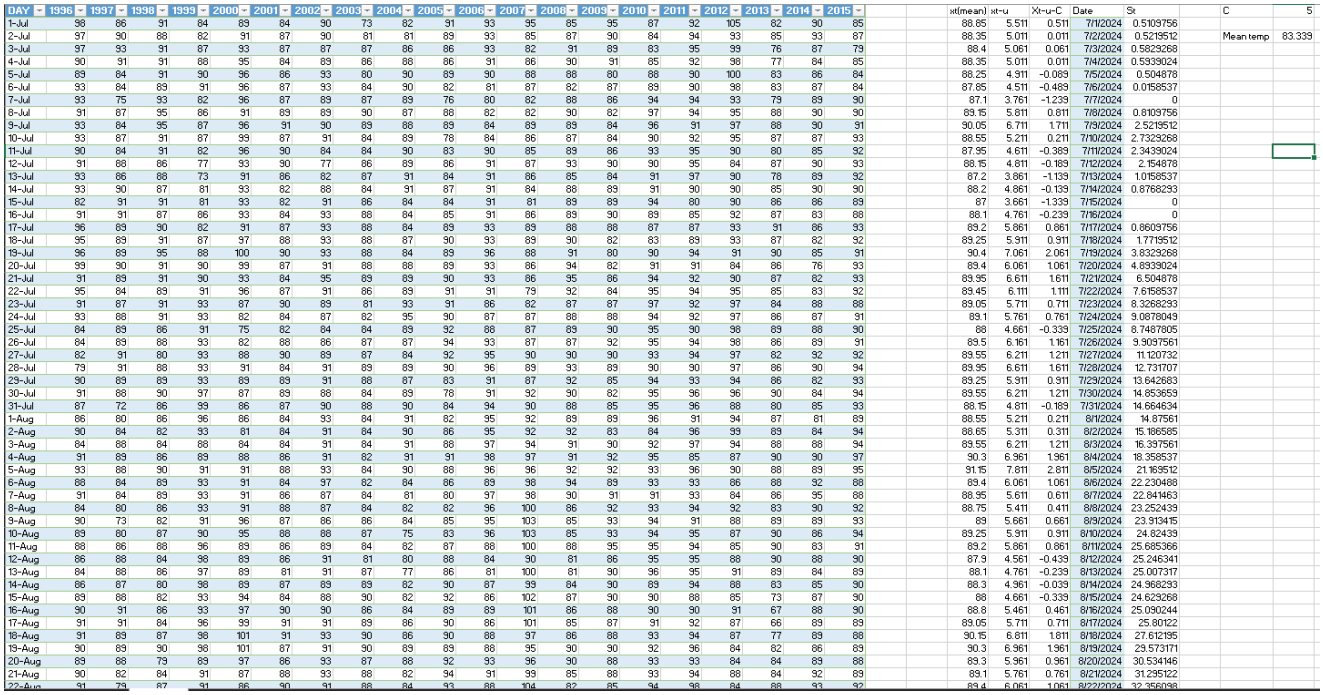
1. Using July through October daily-high-temperature data for Atlanta for 1996 through 2015, use a CUSUM approach to identify when unofficial summer ends (i.e., when the weather starts cooling off) each year. You can get the data that you need from the file temps.txt or online. You can use R if you'd like but it's straightforward enough that an Excel spreadsheet can easily do the job too.

Analysis:

After calculating the average temperature of each day over the years, I subtracted the mean temp of 83.34 degrees from each of the daily temperatures (Xt-u). I used C=5 and calculated the St of each day as you can see in the screenshot of my excel sheet below. I then created a graph of all calculated St's on the y axis and the date on the x-axis. As you can see from the values in the graph, **the temperature starts to cool off around August 26th.**



I put C=5 and calculated the St of each day as you can see in the snippet of my excel sheet below.

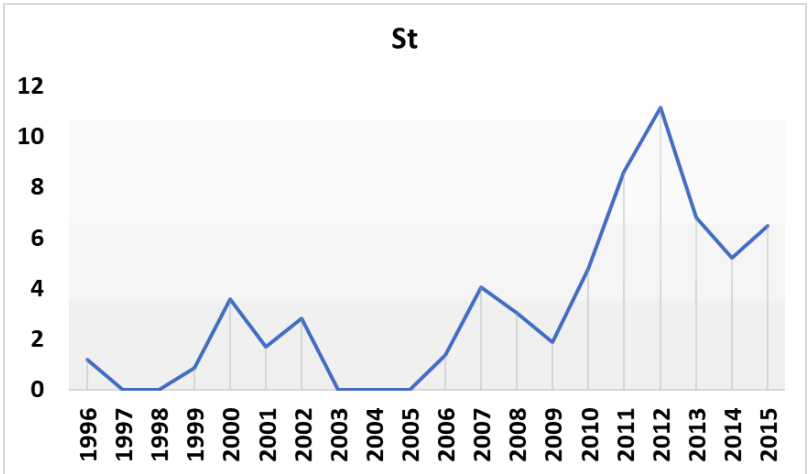


DAY	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	xt(mean)	xt-u	Xt-u-C	Date	St		C	5
1-Jul	98	86	91	84	89	84	90	73	82	91	93	95	85	95	87	92	105	82	90	85	88.85	5.511	0.511	7/1/2024	0.5109756			
2-Jul	97	90	88	82	91	87	90	81	81	89	93	85	87	90	84	94	93	85	93	87	88.35	5.011	0.011	7/2/2024	0.5219512		Mean temp	83.339
3-Jul	97	93	91	87	93	87	87	87	86	86	93	82	91	89	83	95	99	76	87	79	88.4	5.061	0.061	7/3/2024	0.5829268			
4-Jul	90	91	91	88	95	84	89	86	88	86	91	86	90	91	85	92	98	77	84	85	88.35	5.011	0.011	7/4/2024	0.5939024			
5-Jul	89	84	91	90	96	86	93	80	90	89	90	88	88	80	88	90	100	83	86	84	88.25	4.911	-0.089	7/5/2024	0.504878			
6-Jul	93	84	89	91	96	87	93	84	90	82	81	87	82	87	89	90	98	83	87	84	87.85	4.511	-0.489	7/6/2024	0.0158537			
7-Jul	93	75	93	82	96	87	89	87	89	76	80	82	88	86	94	94	93	79	89	90	87.1	3.761	-1.239	7/7/2024	0			
8-Jul	91	87	95	86	91	89	89	90	87	88	82	82	90	82	97	94	95	88	90	90	89.15	5.811	0.811	7/8/2024	0.8109756			
9-Jul	93	84	95	87	96	91	90	89	88	89	84	89	89	84	96	91	97	88	90	91	90.05	6.711	1.711	7/9/2024	2.5219512			
10-Jul	93	87	91	87	99	87	91	84	89	78	84	86	87	84	90	92	95	87	87	93	88.55	5.211	0.211	7/10/2024	2.7329268			
11-Jul	90	84	91	82	96	90	84	84	90	83	90	85	89	86	93	95	90	80	85	92	87.95	4.611	-0.389	7/11/2024	2.3439024			
12-Jul	91	88	86	77	93	90	77	86	89	86	91	87	93	90	90	95	84	87	90	93	88.15	4.811	-0.189	7/12/2024	2.154878			
13-Jul	93	86	88	73	91	86	82	87	91	84	91	86	85	84	91	97	90	78	89	92	87.2	3.861	-1.139	7/13/2024	1.0158537			
14-Jul	93	90	87	81	93	82	88	84	91	87	91	84	88	89	91	90	90	85	90	90	88.2	4.861	-0.139	7/14/2024	0.8768293			
15-Jul	82	91	91	81	93	82	91	86	84	84	91	81	88	89	94	80	90	86	86	89	87	3.661	-1.339	7/15/2024	0			
16-Jul	91	91	87	86	93	84	93	88	84	85	91	86	89	90	89	85	92	87	83	88	88.1	4.761	-0.239	7/16/2024	0			
17-Jul	96	89	90	82	91	87	93	86	84	89	93	89	88	88	87	87	93	91	86	93	89.2	5.861	0.861	7/17/2024	0.8609756			
18-Jul	95	89	91	87	97	88	93	88	87	90	93	89	90	82	83	89	93	87	82	92	89.25	5.911	0.911	7/18/2024	1.7719512			
19-Jul	96	89	95	88	100	90	93	88	84	89	96	88	91	80	90	94	91	90	85	91	90.4	7.061	2.061	7/19/2024	3.8329268			
20-Jul	99	90	91	90	99	87	91	88	88	89	93	86	94	82	91	91	84	86	76	93	89.4	6.061	1.061	7/20/2024	4.8939024			
21-Jul	91	89	91	90	93	84	95	89	89	90	93	86	95	86	94	92	90	87	82	93	89.95	6.611	1.611	7/21/2024	6.504878			
22-Jul	95	84	89	91	96	87	91	86	89	91	91	79	92	84	95	94	95	85	83	92	89.45	6.111	1.111	7/22/2024	7.6158537			
23-Jul	91	87	91	93	87	90	89	81	93	91	86	82	87	87	97	92	97	84	88	88	89.05	5.771	0.771	7/23/2024	8.3268293			
24-Jul	93	88	91	93	82	84	87	82	95	90	87	87	88	88	94	92	97	86	87	91	89.1	5.761	0.761	7/24/2024	9.0878049			
25-Jul	84	89	86	91	75	82	84	84	89	92	88	87	89	90	95	90	98	89	88	90	88	4.661	-0.339	7/25/2024	8.7487805			
26-Jul	84	89	88	93	82	88	86	87	87	94	93	87	87	92	95	94	98	86	89	91	89.5	6.161	1.161	7/26/2024	9.9097561			
27-Jul	82	91	80	93	88	90	89	87	84	92	95	90	90	90	93	94	97	82	92	92	89.55	6.211	1.211	7/27/2024	11.120732			
28-Jul	79	91	88	93	91	84	91	89	89	90	96	89	93	89	90	90	97	86	90	94	89.95	6.611	1.611	7/28/2024	12.731707			
29-Jul	90	89	89	93	89	89	91	88	87	83	91	87	92	85	94	93	94	86	82	93	89.25	5.911	0.911	7/29/2024	13.642683			
30-Jul	91	88	90	97	87	89	88	84	89	78	91	92	90	82	95	96	96	90	84	94	89.55	6.211	1.211	7/30/2024	14.853659			
31-Jul	87	72	86	99	86	87	90	88	90	84	94	90	88	85	95	96	88	80	85	93	88.15	4.811	-0.189	7/31/2024	14.664634			
1-Aug	86	80	86	96	86	84	93	84	91	82	95	92	89	89	96	91	94	87	81	89	88.55	5.211	0.211	8/1/2024	14.87561			
2-Aug	90	84	82	93	81	84	91	84	90	86	95	92	92	83	84	96	99	89	84	94	88.65	5.311	0.311	8/2/2024	15.186585			
3-Aug	84	88	84	88	84	84	91	84	91	88	97	94	91	90	92	97	94	88	88	94	89.55	6.211	1.211	8/3/2024	16.397561			
4-Aug	91	89	86	89	88	86	91	82	91	91	98	97	91	92	95	85	97	90	90	97	90.3	6.961	1.961	8/4/2024	18.358537			
5-Aug	93	88	90	91	91	88	93	84	90	88	96	96	92	92	93	96	90	88	89	95	91.15	7.811	2.811	8/5/2024	21.169512			
6-Aug	88	84	89	93	91	84	97	82	84	86	89	98	94	89	93	93	86	88	92	88	89.4	6.061	1.061	8/6/2024	22.230488			
7-Aug	91	84	89	93	91	86	87	84	81	80	97	98	90	91	91	93	84	86	95	88	88.95	5.611	0.611	8/7/2024	22.841463			
8-Aug	84	80	86	93	91	88	87	84	82	82	96	100	86	92	93	94	92	83	90	92	88.75	5.411	0.411	8/8/2024	23.252439			
9-Aug	90	73	82	91	96	87	86	86	84	85	95	103	85	93	94	91	88	89	89	93	89	5.661	0.661	8/9/2024	23.913415			
10-Aug	89	80	87	90	95	88	88	87	75	83	96	103	85	93	94	95	87	90	86	94	89.25	5.911	0.911	8/10/2024	24.82439			
11-Aug	88	86	88	96	89	86	89	84	82	87	88	100	88	95	95	94	85	90	83	91	89.2	5.861	0.861	8/11/2024	25.685366			
12-Aug	86	88	84	98	89	86	91	81	80	88	84	90	81	86	95	95	88	90	88	90	87.9	4.561	-0.439	8/12/2024	25.246341			
13-Aug	84	88	86	97	89	81	91	87	77	86	81	100	81	90	96	95	91	89	84	89	88.1	4.761	-0.239	8/13/2024	25.007317			
14-Aug	86	87	80	98	89	87	89	89	82	90	87	99	84	90	89	94	88	83	85	90	88.3	4.961	-0.039	8/14/2024	24.968293			
15-Aug	89	88	82	93	94	84	88	90	82	92	86	102	87	90	90	88	85	73	87	90	88	4.661	-0.339	8/15/2024	24.629268			
16-Aug	90	91	86	93	97	90	90	86	84	89	89	101	86	88	90	90	91	67	88	90	88.8	5.461	0.461	8/16/2024	25.090244			
17-Aug	91	91	84	96	99	91	91	89	86	90	86	101	85	87	91	92	87	66	89	89	89.05	5.711	0.711	8/17/2024	25.80122			
18-Aug	91	89	87	98	101	91	93	90	86	90	88	97	86	88	93	94	87	77	89	88	90.15	6.811	1.811	8/18/2024	27.612195			
19-Aug	90	89	90	98	101	87	91	90	89	89	88	95	90	90	92	96	84	82	86	89	90.3	6.961	1.961	8/19/2024	29.573171			
20-Aug	89	88	79	89	97	86	93	87	88	92	93	96	90	88	93	93	84	84	89	88	89.3	5.961	0.961	8/20/2024	30.534146			
21-Aug	90	82	84	91	87	88	93	88	82	94	91	99	85	88	93	94	88	84	92	89	89.1	5.761	0.761	8/21/2024	31.295122			
22-Aug	91	79	87	91	86	90	91	88	84	93	88	104	82	85	94	98	84	88	93	92	89.4	6.061	1.061	8/22/2024	32.256098			

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).

Analysis:

As previously stated the temperature starts to cool off around August 26th so I will check if the summer's (July 1 - August 26) have gotten warmer over the years. I did the same steps as the previous problem but analyzed the difference by year and not by day in the year. **After looking at the plot of the St calculations, I can see that 2010 was the first year where the temperature started to increase, however the temperatures started to decrease again starting 2013. Therefore there is too much variance to accurately determine that the summer temperatures have steadily increased.**



Please find below a screenshot of the excel sheet for this problem.

DAY	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
1-Jul	98	86	91	84	89	84	90	73	82	91	93	95	85	95	87	92	105	82	90	85
2-Jul	97	90	88	82	91	87	90	81	81	89	93	85	87	90	84	94	93	85	93	87
3-Jul	97	93	91	87	93	87	87	87	86	86	93	82	91	89	83	95	99	76	87	79
4-Jul	90	91	91	88	95	84	89	86	88	86	91	86	90	91	85	92	98	77	84	85
5-Jul	89	84	91	90	96	86	93	80	90	89	90	88	88	80	88	90	100	83	86	84
6-Jul	93	84	89	91	96	87	93	84	90	82	81	87	82	87	89	90	98	83	87	84
7-Jul	93	75	93	82	96	87	89	87	89	76	80	82	88	86	94	94	93	79	89	90
8-Jul	91	87	95	86	91	89	89	90	87	88	82	82	90	82	97	94	95	88	90	90
9-Jul	93	84	95	87	96	91	90	89	88	89	84	89	89	84	96	91	97	88	90	91
10-Jul	93	87	91	87	99	87	91	84	89	78	84	86	87	84	90	92	95	87	87	93
11-Jul	90	84	91	82	96	90	84	84	90	83	90	85	89	86	93	95	90	80	85	92
12-Jul	91	88	86	77	93	90	77	86	89	86	91	87	93	90	90	95	84	87	90	93
13-Jul	93	86	88	73	91	86	82	87	91	84	91	86	85	84	91	97	90	78	89	92
14-Jul	93	90	87	81	93	82	88	84	91	87	91	84	88	89	91	90	90	85	90	90
15-Jul	82	91	91	81	93	82	91	86	84	84	91	81	89	89	94	80	90	86	86	89
16-Jul	91	91	87	86	93	84	93	88	84	85	91	86	89	90	89	85	92	87	83	88
17-Jul	96	89	90	82	91	87	93	88	84	89	93	89	88	88	87	87	93	91	86	93
18-Jul	95	89	91	87	97	88	93	88	87	90	93	89	90	82	83	89	93	87	82	92
19-Jul	96	89	95	88	100	90	93	88	84	89	96	88	91	80	90	94	91	90	85	91
20-Jul	99	90	91	90	99	87	91	88	88	89	93	86	94	82	91	91	84	86	76	93
21-Jul	91	89	91	90	93	84	95	89	89	90	93	86	95	86	94	92	90	87	82	93
22-Jul	95	84	89	91	96	87	91	86	89	91	91	79	92	84	95	94	95	85	83	92
23-Jul	91	87	91	93	87	90	89	81	93	91	86	82	87	87	97	92	97	84	88	88
24-Jul	93	88	91	93	82	84	87	82	95	90	87	87	88	88	94	92	97	86	87	91
25-Jul	84	89	86	91	75	82	84	84	89	92	88	87	89	90	95	90	98	89	88	90
26-Jul	84	89	88	93	82	88	86	87	87	94	93	87	87	92	95	94	98	86	89	91
27-Jul	82	91	80	93	88	90	89	87	84	92	95	90	90	90	93	94	97	82	92	92
28-Jul	79	91	88	93	91	84	91	89	89	90	96	89	93	89	90	90	97	86	90	94
29-Jul	90	89	89	93	89	89	91	88	87	83	91	87	92	85	94	93	94	86	82	93
30-Jul	91	88	90	97	87	89	88	84	89	78	91	92	90	82	95	96	96	90	84	94
31-Jul	87	72	86	99	86	87	90	88	90	84	94	90	88	85	95	96	88	80	85	93
1-Aug	86	80	86	96	86	84	93	84	91	82	95	92	89	89	96	91	94	87	81	89
2-Aug	90	84	82	93	81	84	91	84	90	86	95	92	92	83	84	96	99	89	84	94
3-Aug	84	88	84	88	84	84	91	84	91	88	97	94	91	90	92	97	94	88	88	94
4-Aug	91	89	86	89	88	86	91	82	91	91	98	97	91	92	95	85	87	90	90	97
5-Aug	93	88	90	91	91	88	93	84	90	88	96	96	92	92	93	96	90	88	89	95
6-Aug	88	84	89	93	91	84	97	82	84	86	89	98	94	89	93	93	86	88	92	88
7-Aug	91	84	89	93	91	86	87	84	81	80	97	98	90	91	91	93	84	86	95	88
8-Aug	84	80	86	93	91	88	87	84	82	82	96	100	86	92	93	94	92	83	90	92
9-Aug	90	73	82	91	96	87	86	86	84	85	95	103	85	93	94	91	88	89	89	93
10-Aug	89	80	87	90	95	88	88	87	75	83	96	103	85	93	94	95	87	90	86	94
11-Aug	88	86	88	96	89	86	89	84	82	87	88	100	88	95	95	94	85	90	83	91
12-Aug	86	88	84	98	89	86	91	81	80	88	84	90	81	86	95	95	88	90	88	90
13-Aug	84	88	86	97	89	81	91	87	77	86	81	100	81	90	96	95	91	89	84	89
14-Aug	86	87	80	98	89	87	89	89	82	90	87	99	84	90	89	94	88	83	85	90
15-Aug	89	88	82	93	94	84	88	90	82	92	86	102	87	90	90	88	85	73	87	90
16-Aug	90	91	86	93	97	90	90	86	84	89	89	101	86	88	90	90	91	67	88	90
17-Aug	91	91	84	96	99	91	91	89	86	90	86	101	85	87	91	92	87	66	89	89
18-Aug	91	89	87	98	101	91	93	90	86	90	88	97	86	88	93	94	87	77	89	88
19-Aug	90	89	90	98	101	87	91	90	89	89	88	95	90	90	92	96	84	82	86	89
20-Aug	89	88	79	89	97	86	93	87	88	92	93	96	90	88	93	93	84	84	89	88
21-Aug	90	82	84	91	87	88	93	88	82	94	91	99	85	88	93	94	88	84	92	89
22-Aug	91	79	87	91	86	90	91	88	84	93	88	104	82	85	94	98	84	88	93	92
23-Aug	91	81	87	90	88	88	95	90	84	87	87	98	78	81	93	92	88	90	93	87
24-Aug	91	82	88	80	92	93	93	89	87	85	83	95	83	86	90	93	86	84	88	89
25-Aug	84	84	90	82	92	90	91	88	82	84	85	94	78	87	89	95	85	82	84	84
26-Aug	88	87	91	89	90	91	88	89	86	84	88	92	83	90	90	99	90	82	86	86

Mean	88.8351																			
C	0																			
Xt(u)	90.0351	86.2281	87.9825	89.7018	91.5439	86.9649	89.9649	85.9474	86.3684	87.1228	90.2105	91.4912	87.8246	87.7018	91.7018	92.6842	91.386	84.4737	87.2456	90.1228
Xt - u	1.2	-2.607	-0.8526	0.86667	2.70877	-1.8702	1.12982	-2.8877	-2.4667	-1.7123	1.37544	2.65614	-1.0105	-1.1333	2.86667	3.84912	2.55088	-4.3614	-1.5895	1.28772
Xt - u - C	1.2	-2.607	-0.8526	0.86667	2.70877	-1.8702	1.12982	-2.8877	-2.4667	-1.7123	1.37544	2.65614	-1.0105	-1.1333	2.86667	3.84912	2.55088	-4.3614	-1.5895	1.28772
Date	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
St	1.2	0	0	0.86667	3.57544	1.70526	2.83509	0	0	0	1.37544	4.03158	3.02105	1.88772	4.75439	8.60351	11.1544	6.79298	5.20351	6.49123