Homework 2 - ISYE6501-OA

5/27/2018

Objectives

The focus of Week2 was Classification, Data Preparation, Outliers, and Change Detection (with specific application to CUSUM). This homework tests several of these aspects.

Question 4.1

Describe a situation or problem from your job, everyday life, current events, etc., for which a clustering model would be appropriate. List some (up to 5) predictors that you might use.

I want to understand my family's spending habits in order to better balance our budget as well as save for the future. Data can be extracted in csv format from every bank or credit card company's online portal. There could be several attributes/predictors that play into the spending habits

1. **day of week** There could be potentially seven clusters which show the cumulative spending for that day.
2. **category** These categories are spread across: Utilities, Groceries, Home, Automobiles, Discretionary Spending (such as eating out), Retail (clothes and other purchases)
3. **times of year** There may be some times more than others (Xmas, Memorial day, birthdays) where my family and I have more propensity to spend.
4. **location** Do we spend more in one location vs another. This may create too many clusters but worth exploring
5. **person** Who is doing the spending. This may be a predictor , or may be a lens to look at all the other predictors above. For instance, I would run 1 through 4 for myself, and then a separate (facet) view of 1-4 for my wife and so on, so separate spending patterns and look for insights.

Question 4.2

Q4.2 Use the R function kmeans to cluster the points as well as possible [using the iris data]. Report the best combination of predictors, your suggested value of k, and how well your best clustering predicts flower type.

For this question, let's understand the data first. In 1936, a biologist named Ronald Fisher measured 50 data sets across three different species of the iris flower.

Here's a sample of the iris data set:

```r
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

Some important points before beginning:

- this data set is essentially unsupervised data. We dont necessarily know the grouping. However, the biologist Ron Fisher made this data unique by adding the labels (aka Species) to each data set.
- in real life we would not have this information available
- and given a new measurement of sepal or petal length/width we would have to use the clustering process to identify which cluster that new value would belong to.
- however, since we know that these measurements are grouped into 3 species, we basically already know that the optimum clusters are 3.
- however, we shall still evaluate how the kmeans() algorithm performs for other values of k, and *objectively* attempt to come to the most optimum cluster value
- in short, **we will have selective amnesia during this exercise that k really is 3 , and try to figure it out by analysis!)**

Let's disect this data.. We have four predictors: *(the last column, Species, is the manual categorization done by the biologist, its not a predictor)*

```
colnames(irisTable)
```

```
## [1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"
## [5] "Species"
```

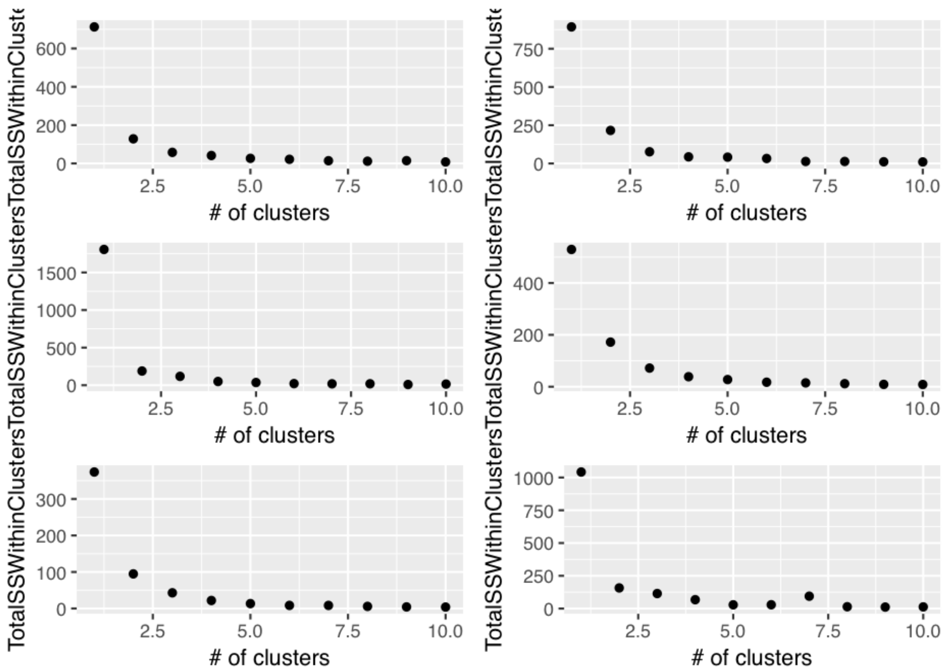
We have 4 predictors that can be combined in groups of 2 (i.e. tuple length is 2) This means we have C(4,2) = 6 of combinations we can use in 2-dimensional clustering, using the combinations formula.. n!/[(n-k)!*k!]

These are:

1. Sepal.Length plotted against Sepal.Width
2. Sepal.Length plotted against Petal.Length
3. Sepal.Length plotted against Petal.Width
4. Sepal.Width plotted against Petal.Length
5. Sepal.Width plotted against Petal.Width
6. Petal.Length plotted against Petal.Width

Against each of these combinations, we run cluster k value of 1 through 10 to see the outcome.. Specifically, we plot the number of clusters against the total sum of squares within each cluster.

The argument being that the sum of squares within a cluster needs to be minimized by choosing the appropriate number of clusters.

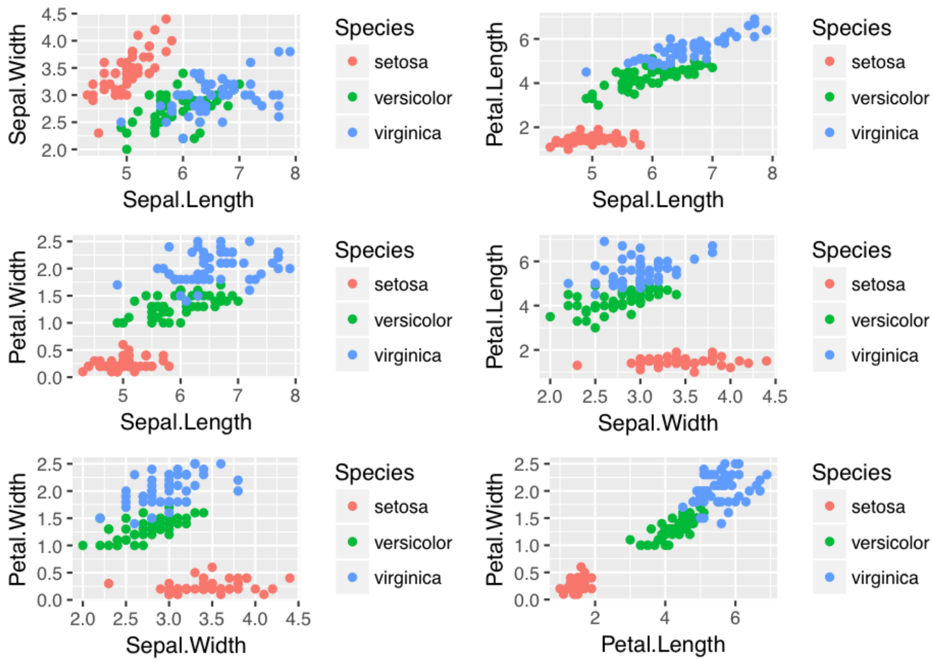


Summary

As can be seen from the above, there is diminishing improvement of the total sum of squares within each cluster after k = 3 . Implies that the optimum number of clusters witnessed within the data set is 3, as that is where the "kink" in the elbow curve lies.

Some post-analysis confirmation:

As mentioned earlier, Ronald Fisher already provided us the answer (i.e. number of clusters) within his data. Knowing that, and working backwards, if we plot the three clusters (categorized by "Species" using ggplot2, we see the following:



The clean demarcation in almost all cases of shows that indeed the number of clusters is 3.

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.

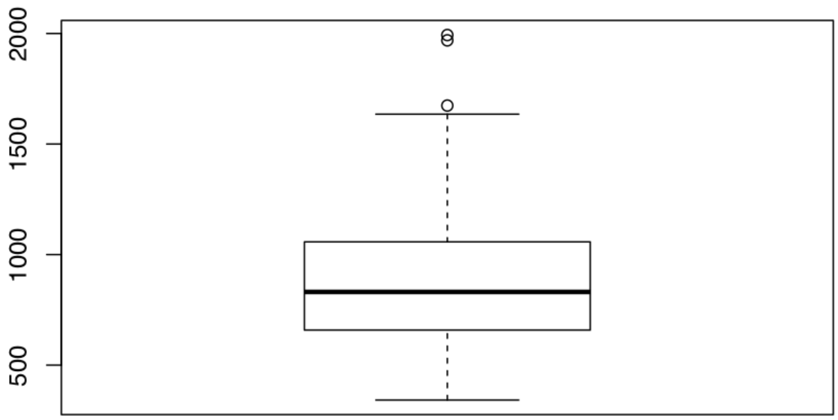
First, we pull down the data and load it…

```
dataFile5_1 <- "uscrime.txt"
if (!file.exists(dataFile5_1)) {
  crimeDataURL <- paste0(c("http://www.statsci.org/data/general/uscrime.txt"))
  download.file(crimeDataURL, dataFile5_1) }

crimeDataTable <- read.table(dataFile5_1, header = TRUE )
```

Plotting box-plot:

```r
boxplot(crimeDataTable$Crime)
```



Since all the outliers on one side (i.e one tail, we just use one tailed grubb test)

Running Grubs test with one-tailed (opposite = false)

```r
grubbs.test(crimeDataTable$Crime, type = 10, opposite = FALSE)
```

```
##
##  Grubbs test for one outlier
##
## data:  crimeDataTable$Crime
## G = 2.81290, U = 0.82426, p-value = 0.07887
## alternative hypothesis: highest value 1993 is an outlier
```

- Also the p value of 0.07887 (is very close to less than 0.05), so we can reject the null hypothesis given we're playing with 90% confidence intervals.
- And agree with the alternative hypothesis is accurate so 1993 is indeed an outlier.

the highest value of 1993 is an outlier. Let's not just take grubbs word for it; let's plot this data too!

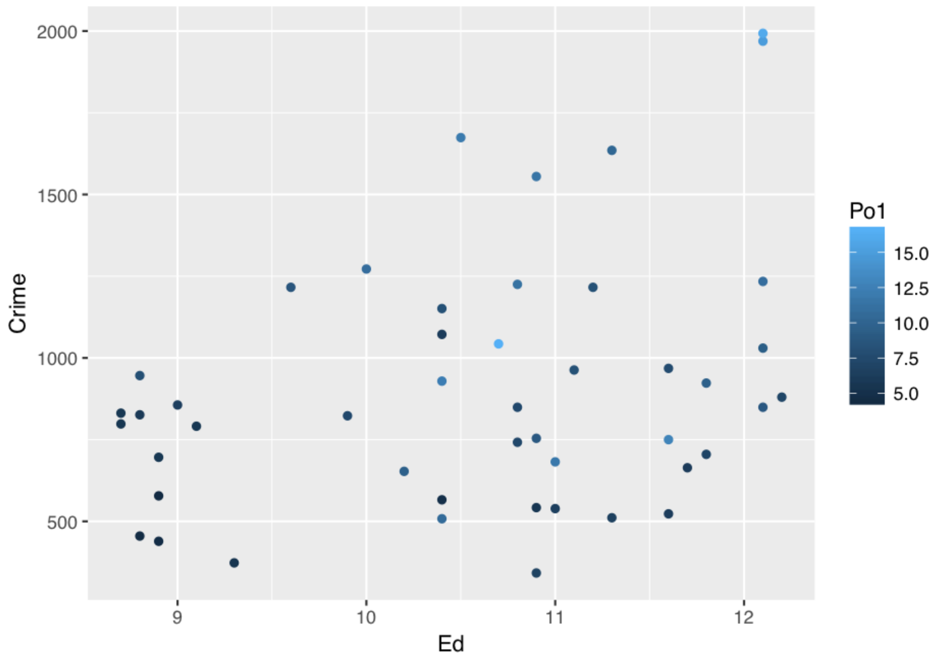
First, we need to understand the various attributes here !(http://www.statsci.org/data/general/uscrime.html)

`Variable        Description`

M percentage of males aged 14–24 in total state population So indicator variable for a southern state Ed mean years of schooling of the population aged 25 years or over Po1 per capita expenditure on police protection in 1960 Po2 per capita expenditure on police protection in 1959 LF labour force participation rate of civilian urban males in the age-group 14-24 M.F number of males per 100 females Pop state population in 1960 in hundred thousands NW percentage of nonwhites in the population U1 unemployment rate of urban males 14–24 U2 unemployment rate of urban males 35–39 Wealth wealth: median value of transferable assets or family income Ineq income inequality: percentage of families earning below half the median income Prob probability of imprisonment: ratio of number of commitments to number of offenses Time average time in months served by offenders in state prisons before their first release Crime crime rate: number of offenses per 100,000 population in 1960

This plot below shows the crime rate plotted against the state's education situation (mean years of schooling of populate >25). I colored the points based on the amount of police protection available in that state.

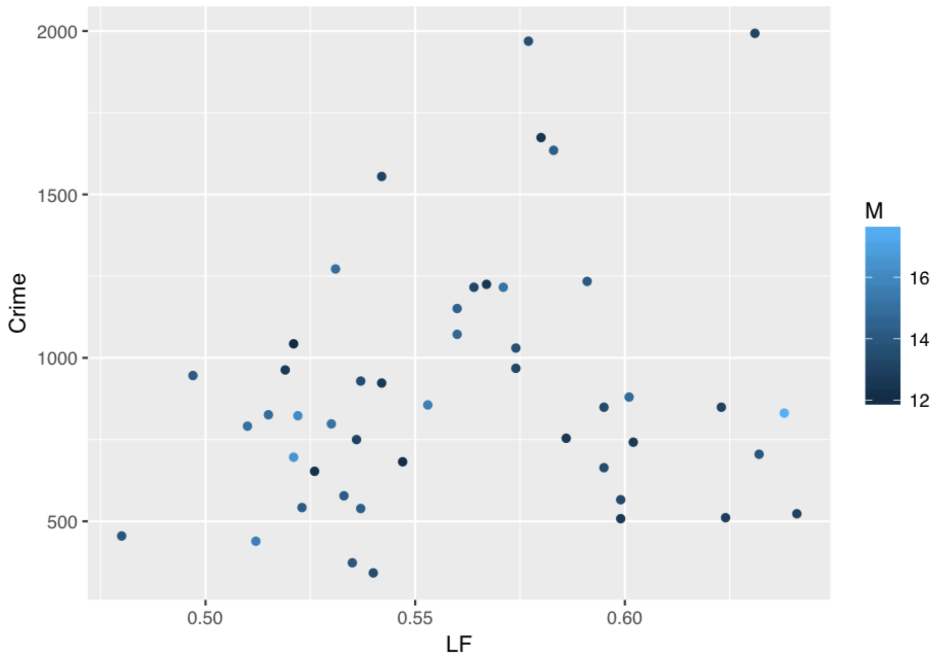
```r
ggplot(data = crimeDataTable) +
  geom_point(mapping = aes(Ed, Crime, color = Po1))
```



The statistic of 1993 crimes / 100k population is the far upper right and different from the other states with high crime rate where there is lower per capita expenditure on police protection:

This plot below shows crime rate against LF (labor force participation rate in ages 14-24)

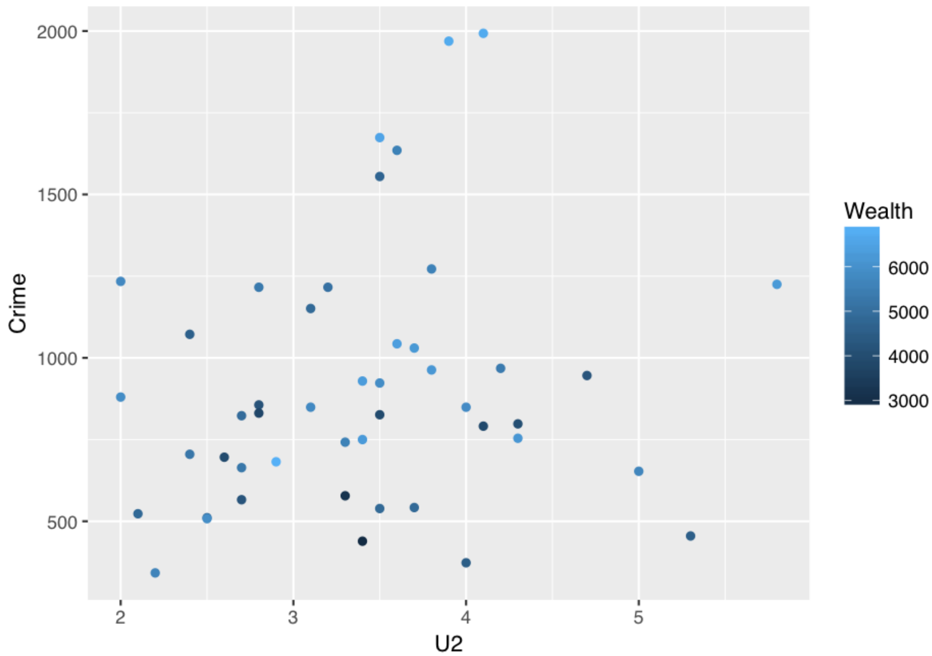
```
ggplot(data = crimeDataTable) +
  geom_point(mapping = aes(LF, Crime, color = M))
```



Again, most of the states with high youth labor force have lower crime (less than 1000/100k population). Yet, this state has the highest in the country! Another clear sign of an outlier.

Finally, this plot below shows (U2) the unemployment rate of urban males (35-39) plotted against the crime rate.

```
ggplot(data = crimeDataTable) +
  geom_point(mapping = aes(U2, Crime, color = Wealth))
```



Again as we can witness, the state with the highest , 1993/100K crime rate in in stark difference than the other states also in the 4% range of unemployment.

So, in short, Grubbs.test is fairly accurate in its estimation

---

Question6.1

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?

I am a co-founder of waada (http://waada.org). The purpose of this non profit is to help folks with mental illness using technology. The reason why CUSUM / Change Detection is so germane to this organization is that I can can use the CUSUM algorithm to detect mood changes. Unless the person has bi-polar depression , where the changes are obvious, depression in people who are prone to it creeps in gradually until its too late for the care giver to make an impact. In this situation, the "slippery slope" hits the depressed person and they stay depressed for weeks or months. Sometimes crude measures like medicine have to be taken to lift them out, but those are mostly artificial and there is no way to measure the exact quantity to be taken by the person to get better since measuring the "extent" of depression is so subjective. Therefore there is almost always a slight overdose of the medicine, which in the long term is severely adverse to the health of the patient, since he or she invariably becomes dependent on that medicine, akin to a drug addict.

The concept I have is as follows. Suppose we are able to take in physiological information (heartbeats through fitbits or iWatch wearables), phone mobility (through its gyroscope), # of calls made, length of calls, we can start creating a pattern around this person. We can also allow this person to directly enter into the phone via an app if they are feeling down or not (taking care to give something back in return, like a calming remedy , a song, breathing techniques etc so we can motivate the person to enter the data).

This data can then be used to create a "mental wellness score", at which point you have mapped the subjective "extent" of depression into objective numbers. After that, applying the CUSUM technique is simple.

Prompt identification of the onset of depression would allow us to engage in proactive measures like involving the caregiver much sooner, or providing special services through the mobile app around improving breathing techniques and a more engaged package of activities. if the onset is pretty severe, healthcare and even emergency services (suicide hotline) could be put on notice.

Specifics around C and T for this use-case

Threshold T - the threshold T is decided upon based on how costly it is to qualify that a change is detected

if the threshold is low, the cost is that the patient may not have fallen ill, and we have the cost to bear of engaging extra healthcare services and the cost of pulling in the caregiver sooner than normal.

if the threshold is too high, we may never find out until its too late. the cost can be catastrophic. This is why I'd tend to stay on the lower range of the threshold.

Dampener C - the value C depends on how costly it is to be too sensitive to change.

A low value of C will raise false alarms, if the mood is oscillating close to the border line. This may be the case if the person is taking drugs/medicine to improve their well being, and spurious random data can also cause the model to be inaccurate. The cost of false alarms is pretty much the same as a low threshold above.

If however the dampener is high, the change may get detected a little later than desired. (e.g. on Day4 when the actual depression started getting established on Day2). The costs over here, though not as catastrophic as a high Threshold value, is still high. namely the cost is that a high C renders the model useless. This is because when a person has been consistently in a depressed state for 1-2 days, its going to get much harder to pull them out of it, and the latter was the whole purpose of this initiative.

Therefore I'd err towards a mid to lower range of C.

Note: I have concatenated the answer to Question 6.2 below this Rmd. It was written in Google Docs and exported to PDF in case you are wondering why the formatting and fonts etc changed so much!

HW2 ISYE 6501 Summer 2018 OA

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

How CUSUM works:
The goal in the CUSUM[1] (cumulative sum) approach is to detect change. Specifically, looking at the formula For detecting a change that is "higher" than normal:
First we calculate the
$S(t) == \max(0,S(t-1) + x(t) - \mu - C$

and then ask: Is $S(t) => T$

- The idea here is to calculate the cumulative sum S(t) and then comparing it (for each observation) against a threshold T
  - S(t) is the maximum of either 0 or
  - the addition of prior S(t-1) and the difference for the observed data point from its average ($x(t) - \mu$)
- the C parameter is a dampener to change the models sensitivity
  - higher C means the model is less sensitive, but may be delayed (or may never) detect the change
  - lower C means the model is more sensitive
- the threshold T is decided upon based on how costly it is to qualify that a change is detected
- the value C depends on how costly it is to be too sensitive to change.

Goal of this exercise...
To evaluate all the daily temperature values of the 4month data for 15 years, and see when temperature dramatically starts to drop off

- Essentially this is the *opposite* of the above equation.
- Meaning, we are trying to measure the change has dropped below threshold:
- So , re-framing the equation:

If $S(t) == \max(0,S(t-1) + \mu - x(t) - C$
Then is Is $S(t) => T$?

---

[1] Extra reading for CUSUM here: https://itl.nist.gov/div898/handbook/pmc/section3/pmc323.htm

- extra reading talks about v masks here:

Steps followed in creating the model

- First I set up the excel sheet (google sheets) where I calculated the Average, Standard Deviation , Target and Dampener for EACH year..
    - Average and Stdev are built-in excel functions and self-explanatory
    - For the Target T, I used any change which is 2 standard deviations from the norm. I made this into a variable as seen below so I can change it and see the effect
    - For the dampener C, I started with 0 first and changed it as I saw the spurious values show up..
- I then created new columns just to calculate S(t) as per the above equation. (Col X in Figure 6.2.1)
- Also created a simple IF statement to see if S(t) > T. If yes, then I marked that as TRUE, otherwise I marked it FALSE (Column Y in Figure 6.2.1)
    - Initially , I set up the model with just S(1996) but then later, once this year was tuned I replicated this for S>T for all years.

Variables														
Target (T)	2	standard deviations from norm												
Dampener (C)	0	standard deviations from norm												
AVERAGE:	**83.72**	**81.67**	**84.26**	**83.36**	**84.03**	**81.55**	**83.59**	**81.48**	**81.67**	**83.94**	**83.30**			
STDDEV:	**8.55**	**9.32**	**6.41**	**9.72**	**9.52**	**8.22**	**9.43**	**7.02**	**7.73**	**6.59**	**8.71**			
Target (T)	**17.10**	**18.64**	**12.82**	**19.45**	**19.04**	**16.45**	**18.85**	**14.04**	**15.45**	**13.18**	**17.42**			
Dampener (C)	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**			
DAY	1996	1997	1998	1999	2000	2001	2002	2003	2013	2014	2015	u-x(1996)-C	S(1996)	Is S(1996) > T(1996) ?
1-Jul	98	86	91	84	89	84	90	73	82	90	85	-14.28	0	FALSE
2-Jul	97	90	88	82	91	87	90	81	85	93	87	-13.28	0	FALSE
3-Jul	97	93	91	87	93	87	87	87	76	87	79	-13.28	0	FALSE
4-Jul	90	91	91	88	95	84	89	86	77	84	85	-6.28	0	FALSE
5-Jul	89	84	91	90	96	86	93	80	83	86	84	-5.28	0	FALSE
6-Jul	93	84	89	91	96	87	93	84	83	87	84	-9.28	0	FALSE
7-Jul	93	75	93	82	96	87	89	87	79	89	90	-9.28	0	FALSE
8-Jul	91	87	95	86	91	89	89	90	88	90	90	-7.28	0	FALSE
9-Jul	93	84	95	87	96	91	90	89	88	90	91	-9.28	0	FALSE
10-Jul	93	87	91	87	99	87	91	84	87	87	93	-9.28	0	FALSE
11-Jul	90	84	91	82	96	90	84	84	80	85	92	-6.28	0	FALSE
12-Jul	91	88	86	77	93	90	77	86	87	90	93	-7.28	0	FALSE
13-Jul	93	86	88	73	91	86	82	87	78	89	92	-9.28	0	FALSE
14-Jul	93	90	87	81	93	82	88	84	85	90	90	-9.28	0	FALSE

Figure 6.2.1 - Model Snapshot

How I determined the value of C:

Started with C = 0. Immediately noticed in the chart, spurious S(t) values spiking and then dropping back down, as marked in the Orange Cells below.

Variables														
Target (T)	2	standard deviations from norm												
Dampener (C)	0	standard deviations from norm												
AVERAGE:	**83.72**	**81.67**	**84.26**	**83.36**	**84.03**	**81.55**	**83.59**	**81.48**	**81.67**	**83.94**	**83.30**			
STDDEV:	**8.55**	**9.32**	**6.41**	**9.72**	**9.52**	**8.22**	**9.43**	**7.02**	**7.73**	**6.59**	**8.71**			
Target (T)	**17.10**	**18.64**	**12.82**	**19.45**	**19.04**	**16.45**	**18.85**	**14.04**	**15.45**	**13.18**	**17.42**			
Dampener (C)	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**			
DAY	1996	1997	1998	1999	2000	2001	2002	2003	2013	2014	2015	u-x(1996)-C	S(1996)	Is S(1996) > T(1996) ?
1-Jul	98	86	91	84	89	84	90	73	82	90	85	-14.28	0	FALSE
2-Jul	97	90	88	82	91	87	90	81	85	93	87	-13.28	0	FALSE
3-Jul	97	93	91	87	93	87	87	87	76	87	79	-13.28	0	FALSE
4-Jul	90	91	91	88	95	84	89	86	77	84	85	-6.28	0	FALSE
5-Jul	89	84	91	90	96	86	93	80	83	86	84	-5.28	0	FALSE
6-Jul	93	84	89	91	96	87	93	84	83	87	84	-9.28	0	FALSE
7-Jul	93	75	93	82	96	87	89	87	79	89	90	-9.28	0	FALSE
8-Jul	91	87	95	86	91	89	89	90	88	90	90	-7.28	0	FALSE
9-Jul	93	84	95	87	96	91	90	89	88	90	91	-9.28	0	FALSE
10-Jul	93	87	91	87	99	87	91	84	87	87	93	-9.28	0	FALSE
11-Jul	90	84	91	82	96	90	84	84	80	85	92	-6.28	0	FALSE
12-Jul	91	88	86	77	93	90	77	86	87	90	93	-7.28	0	FALSE
13-Jul	93	86	88	73	91	86	82	87	78	89	92	-9.28	0	FALSE
14-Jul	93	90	87	81	93	82	88	84	85	90	90	-9.28	0	FALSE
15-Jul	82	91	91	81	93	82	91	86	86	86	89	1.72	1.71544715	FALSE
16-Jul	91	91	87	86	93	84	93	88	87	83	88	-7.28	0	FALSE
17-Jul	96	89	90	82	91	87	93	88	91	86	93	-12.28	0	FALSE
18-Jul	95	89	91	87	97	88	93	88	87	82	92	-11.28	0	FALSE
19-Jul	96	89	95	88	100	90	93	88	90	85	91	-12.28	0	FALSE
20-Jul	99	90	91	90	99	87	91	88	86	76	93	-15.28	0	FALSE
21-Jul	91	89	91	90	93	84	95	89	87	82	93	-7.28	0	FALSE
22-Jul	95	84	89	91	96	87	91	86	85	83	92	-11.28	0	FALSE
23-Jul	91	87	91	93	87	90	89	81	84	88	88	-7.28	0	FALSE
24-Jul	93	88	91	93	82	84	87	82	86	87	91	-9.28	0	FALSE
25-Jul	84	89	86	91	75	82	84	84	89	88	90	-0.28	0	FALSE
26-Jul	84	89	88	93	82	88	86	87	86	89	91	-0.28	0	FALSE
27-Jul	82	91	80	93	88	90	89	87	82	92	92	1.72	1.71544715	FALSE
28-Jul	79	91	88	93	91	84	91	89	86	90	94	4.72	6.43089430	FALSE
29-Jul	90	89	89	93	89	89	91	88	86	82	93	-6.28	0.14634146	FALSE
30-Jul	91	88	90	97	87	89	88	84	90	84	94	-7.28	0	FALSE
31-Jul	87	72	86	99	86	87	90	88	80	85	93	-3.28	0	FALSE
1-Aug	86	80	86	96	86	84	93	84	87	81	89	-2.28	0	FALSE
2-Aug	90	84	82	93	81	84	91	84	89	84	94	-6.28	0	FALSE
3-Aug	84	88	84	88	84	84	91	84	88	88	94	-0.28	0	FALSE
4-Aug	91	89	86	89	88	86	91	82	90	90	97	-7.28	0	FALSE

Figure 6.2.2 - Model Snapshot with C = 0

So I played around with the Dampener until I saw the S(1996) column become less fickle and shows lesser flopping..

<u>The Dampener I ended up finalizing was 0.7 x standard deviation for the data for each year.</u>

I then played around with the value of T:

Starting with T=2**σ** (or 2 times standard deviation)I notice that the change is detected on 30th September.

Variables														
Target (T)	2	standard deviations from norm												
Dampener (C)	0.7	standard deviations from norm												
AVERAGE:	**83.72**	**81.67**	**84.26**	**83.36**	**84.03**	**81.55**	**83.59**	**81.48**	**81.67**	**83.94**	**83.30**			
STDDEV:	**8.55**	**9.32**	**6.41**	**9.72**	**9.52**	**8.22**	**9.43**	**7.02**	**7.73**	**6.59**	**8.71**			
Target (T)	**17.10**	**18.64**	**12.82**	**19.45**	**19.04**	**16.45**	**18.85**	**14.04**	**15.45**	**13.18**	**17.42**			
Dampener (C)	**5.98**	**6.52**	**4.49**	**6.81**	**6.66**	**5.76**	**6.60**	**4.91**	**5.41**	**4.61**	**6.10**			
DAY	1996	1997	1998	1999	2000	2001	2002	2003	2013	2014	2015	u-x(1996)-C	S(1996)	Is S(1996) > T(1996) ?
1-Jul	98	86	91	84	89	84	90	73	82	90	85	-20.27	0	FALSE
2-Jul	97	90	88	82	91	87	90	81	85	93	87	-19.27	0	FALSE
3-Jul	97	93	91	87	93	87	87	87	76	87	79	-19.27	0	FALSE
4-Jul	90	91	91	88	95	84	89	86	77	84	85	-12.27	0	FALSE
5-Jul	89	84	91	90	96	86	93	80	83	86	84	-11.27	0	FALSE
6-Jul	93	84	89	91	96	87	93	84	83	87	84	-15.27	0	FALSE
22-Sep	81	70	88	72	73	87	77	75	82	82	76	-3.27	0	FALSE
23-Sep	84	80	84	75	81	88	82	81	82	77	81	-6.27	0	FALSE
24-Sep	84	82	81	78	84	69	73	80	71	78	74	-6.27	0	FALSE
25-Sep	87	66	82	81	82	66	69	82	67	77	67	-9.27	0	FALSE
26-Sep	84	70	84	82	68	72	75	82	78	74	71	-6.27	0	FALSE
27-Sep	79	64	87	78	71	75	75	82	79	78	71	-1.27	0	FALSE
28-Sep	75	68	80	80	75	78	79	73	77	74	75	2.73	2.73160968	FALSE
29-Sep	72	77	75	77	73	71	73	66	76	71	77	5.73	8.46321937	FALSE
30-Sep	64	86	75	71	75	71	79	71	77	84	85	13.73	22.1948290	TRUE
1-Oct	66	75	86	73	77	75	82	72	82	86	71	11.73	33.9264387	TRUE
2-Oct	72	73	78	75	79	80	84	68	82	85	66	5.73	39.6580484	TRUE
3-Oct	84	75	77	84	82	81	84	66	82	78	66	-6.27	33.3896581	TRUE
4-Oct	70	78	82	71	81	80	82	77	85	65	70	7.73	41.1212678	TRUE
5-Oct	66	81	82	73	82	79	87	78	84	71	73	11.73	52.8528775	TRUE

Figure 6.2.3 - Model Snapshot with T=2 (C fixed at 0.7)

If I took T to 4**σ** or 2 times standard deviation)I notice that the change is detected on 30th September, the day of change didn't move that drastically (ie. it was Oct 2nd)

Variables														
Target (T)	4	standard deviations from norm												
Dampener (C)	0.7	standard deviations from norm												
AVERAGE:	**83.72**	**81.67**	**84.26**	**83.36**	**84.03**	**81.55**	**83.59**	**81.48**	**81.67**	**83.94**	**83.30**			
STDDEV:	**8.55**	**9.32**	**6.41**	**9.72**	**9.52**	**8.22**	**9.43**	**7.02**	**7.73**	**6.59**	**8.71**			
Target (T)	**34.19**	**37.28**	**25.64**	**38.89**	**38.07**	**32.90**	**37.70**	**28.07**	**30.91**	**26.37**	**34.84**			
Dampener (C)	**5.98**	**6.52**	**4.49**	**6.81**	**6.66**	**5.76**	**6.60**	**4.91**	**5.41**	**4.61**	**6.10**			
DAY	1996	1997	1998	1999	2000	2001	2002	2003	2013	2014	2015	u-x(1996)-C	S(1996)	Is S(1996) > T(1996) ?
1-Jul	98	86	91	84	89	84	90	73	82	90	85	-20.27	0	FALSE
2-Jul	97	90	88	82	91	87	90	81	85	93	87	-19.27	0	FALSE
3-Jul	97	93	91	87	93	87	87	87	76	87	79	-19.27	0	FALSE
4-Jul	90	91	91	88	95	84	89	86	77	84	85	-12.27	0	FALSE
5-Jul	89	84	91	90	96	86	93	80	83	86	84	-11.27	0	FALSE
6-Jul	93	84	89	91	96	87	93	84	83	87	84	-15.27	0	FALSE
22-Sep	81	70	88	72	73	87	77	75	82	82	76	-3.27	0	FALSE
23-Sep	84	80	84	75	81	88	82	81	82	77	81	-6.27	0	FALSE
24-Sep	84	82	81	78	84	69	73	80	71	78	74	-6.27	0	FALSE
25-Sep	87	66	82	81	82	66	69	82	67	77	67	-9.27	0	FALSE
26-Sep	84	70	84	82	68	72	75	82	78	74	71	-6.27	0	FALSE
27-Sep	79	64	87	78	71	75	75	82	79	78	71	-1.27	0	FALSE
28-Sep	75	68	80	80	75	78	79	73	77	74	75	2.73	2.73160968	FALSE
29-Sep	72	77	75	77	73	71	73	66	76	71	77	5.73	8.46321937	FALSE
30-Sep	64	86	75	71	75	71	79	71	77	84	85	13.73	22.1948290	FALSE
1-Oct	66	75	86	73	77	75	82	72	82	86	71	11.73	33.9264387	FALSE
2-Oct	72	73	78	75	79	80	84	68	82	85	66	5.73	39.6580484	TRUE
3-Oct	84	75	77	84	82	81	84	66	82	78	66	-6.27	33.3896581	FALSE
4-Oct	70	78	82	71	81	80	82	77	85	65	70	7.73	41.1212678	TRUE
5-Oct	66	81	82	73	82	79	87	78	84	71	73	11.73	52.8528775	TRUE
6-Oct	64	82	73	71	73	70	86	75	84	78	76	13.73	66.5844872	TRUE
7-Oct	60	82	82	73	66	68	80	73	74	82	81	17.73	84.3160968	TRUE
8-Oct	78	82	69	73	55	79	71	73	72	86	82	-0.27	84.0477065	TRUE

Figure 6.2.4 - Model Snapshot with T=4 (C fixed at 0.7)

Since 4 standard deviations is a considerable change of **DROP** in temperature, I decided to use this as my barometer (no pun intended)

Once the values of T and C were set, I expanded the model for all the rest of the years…

When the final model was set up , I changed S>T cells color to green if S>T by using conditional formatting.

I saw some interesting patterns..

With T = 4x STDEV , and C = 0.7x STDEV (my original values) it is clear that unofficially temperature noticeably starts to cool down in the first half of October.

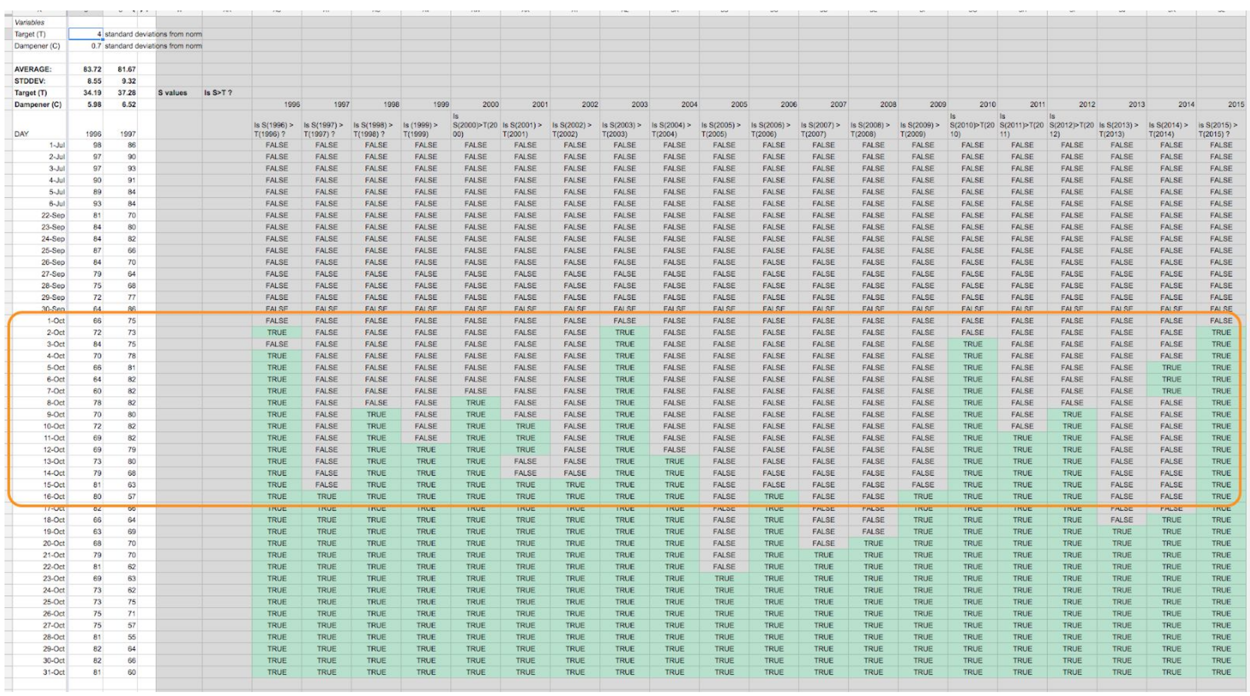


Figure 6.2.5 - S>T

(I'm hiding some columns to fit the data here)

Interestingly enough, if I increase the dampener value , C to 1, the 2015 temperature change detection gets drastically modified

Variables																								
Target (T)	4	standard deviations from norm																						
Dampener (C)	1	standard deviations from norm																						
AVERAGE:	**83.72**	**81.67**																						
STDDEV:	**8.55**	**9.32**																						
Target (T)	**34.19**	**37.28**	**S values**	**Is S>T ?**																				
Dampener (C)	**8.55**	**9.32**			1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
DAY	1996	1997			Is S(1996) > T(1996) ?	Is S(1997) > T(1997) ?	Is S(1998) > T(1998) ?	Is (1999) > T(1999)	Is S(2000)>T(2000)	Is S(2001) > T(2001)	Is S(2002) > T(2002)	Is S(2003) > T(2003)	Is S(2004) > T(2004)	Is S(2005) > T(2005)	Is S(2006) > T(2006)	Is S(2007) > T(2007)	Is S(2008) > T(2008)	Is S(2009) > T(2009)	Is S(2010)>T(2010)	Is S(2011)>T(2011)	Is S(2012)>T(2012)	Is S(2013) > T(2013)	Is S(2014) > T(2014)	is S(2015) > T(2015) ?
1-Jul	98	86			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
2-Jul	97	90			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
3-Jul	97	93			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
4-Jul	90	91			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
5-Jul	89	84			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
6-Jul	93	84			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
22-Sep	81	70			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
23-Sep	84	80			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
24-Sep	84	82			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
25-Sep	87	66			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
26-Sep	84	70			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
27-Sep	79	64			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
28-Sep	75	68			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
29-Sep	72	77			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
30-Sep	64	86			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
1-Oct	66	75			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
2-Oct	72	73			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
3-Oct	84	75			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
4-Oct	70	78			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
5-Oct	66	81			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
6-Oct	64	82			TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
7-Oct	60	82			TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
8-Oct	78	82			TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
9-Oct	70	80			TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
10-Oct	72	82			TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
11-Oct	69	82			TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
12-Oct	69	79			TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
13-Oct	73	80			TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
14-Oct	79	68			TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
15-Oct	81	63			TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
16-Oct	80	57			TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
17-Oct	82	66			TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
18-Oct	66	64			TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
19-Oct	63	69			TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
20-Oct	68	70			TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
21-Oct	79	70			TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
22-Oct	81	62			TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE
23-Oct	69	63			TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE
24-Oct	73	62			TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE
25-Oct	73	75			TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE
26-Oct	75	71			TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE
27-Oct	75	57			TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
28-Oct	81	55			TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
29-Oct	82	64			TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
30-Oct	82	66			TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
31-Oct	81	60			TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

Figure 6.2.6 with bigger dampener (C=1)

This tells us that temperature was drastically but steadily dropping in 2015 starting early October, but when the dampener has a high value, these changes are not detected.

Another interesting observation was that if we decrease the threshold T to 2xSTDEV, the change (to cooler season) is detected sooner, by about a week), but the year over year pattern doesn't change.

Variables																								
Target (T)	2	standard deviations from norm																						
Dampener (C)	0.7	standard deviations from norm																						
AVERAGE:	**83.72**	**81.67**																						
STDDEV:	**8.55**	**9.32**																						
Target (T)	**17.10**	**18.64**	**S values**	**Is S>T ?**																				
Dampener (C)	**5.98**	**6.52**			1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
DAY	1996	1997			Is S(1996) > T(1996) ?	Is S(1997) > T(1997) ?	Is S(1998) > T(1998) ?	Is (1999) > T(1999)	Is S(2000)>T(2000)	Is S(2001) > T(2001)	Is S(2002) > T(2002)	Is S(2003) > T(2003)	Is S(2004) > T(2004)	Is S(2005) > T(2005)	Is S(2006) > T(2006)	Is S(2007) > T(2007)	Is S(2008) > T(2008)	Is S(2009) > T(2009)	Is S(2010)>T(2010)	Is S(2011)>T(2011)	Is S(2012)>T(2012)	Is S(2013) > T(2013)	Is S(2014) > T(2014)	is S(2015) > T(2015) ?
1-Jul	98	86			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
2-Jul	97	90			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
3-Jul	97	93			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
4-Jul	90	91			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
5-Jul	89	84			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
6-Jul	93	84			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
22-Sep	81	70			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
23-Sep	84	80			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
24-Sep	84	82			FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
25-Sep	87	66			FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
26-Sep	84	70			FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
27-Sep	79	64			FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
28-Sep	75	68			FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE
29-Sep	72	77			FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE
30-Sep	64	86			TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE
1-Oct	66	75			TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE
2-Oct	72	73			TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE
3-Oct	84	75			TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE
4-Oct	70	78			TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE
5-Oct	66	81			TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE
6-Oct	64	82			TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE
7-Oct	60	82			TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE
8-Oct	78	82			TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE
9-Oct	70	80			TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE
10-Oct	72	82			TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE
11-Oct	69	82			TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE
12-Oct	69	79			TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE
13-Oct	73	80			TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE
14-Oct	79	68			TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE
15-Oct	81	63			TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE
16-Oct	80	57			TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE
17-Oct	82	66			TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE
18-Oct	66	64			TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
19-Oct	63	69			TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
20-Oct	68	70			TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
21-Oct	79	70			TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
22-Oct	81	62			TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
23-Oct	69	63			TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
24-Oct	73	62			TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
25-Oct	73	75			TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
26-Oct	75	71			TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
27-Oct	75	57			TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
28-Oct	81	55			TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
29-Oct	82	64			TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
30-Oct	82	66			TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
31-Oct	81	60			TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

Figure 6.2.7 with lower threshold T = 2xSTDEV (C back to 0.7)

Question 6.2.1 :Use a CUSUM approach to make a judgment of whether Atlanta’s summer climate has gotten warmer in that time (and if so, when).

While unofficially winter came sooner in 2003, 2010, and 2015, the pattern here doesn’t show any noticable trend of temperature changes over 15 years.

Variables																									
Target (T)	4	standard deviations from norm																							
Dampener (C)	0.7	standard deviations from norm																							
AVERAGE:	**83.72**	**81.67**																							
STDDEV:	**8.55**	**9.32**																							
Target (T)	**34.19**	**37.28**		**S values**	**Is S>T ?**																				
Dampener (C)	**5.98**	**6.52**				1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
DAY	1996	1997				Is S(1996) > T(1996) ?	Is S(1997) > T(1997) ?	Is S(1998) > T(1998) ?	Is (1999) > T(1999)	Is S(2000)>T(2000)	Is S(2001) > T(2001)	Is S(2002) > T(2002)	Is S(2003) > T(2003)	Is S(2004) > T(2004)	Is S(2005) > T(2005)	Is S(2006) > T(2006)	Is S(2007) > T(2007)	Is S(2008) > T(2008)	Is S(2009) > T(2009)	Is S(2010)>T(2010)	Is S(2011)>T(2011)	Is S(2012)>T(2012)	Is S(2013) > T(2013)	Is S(2014) > T(2014)	Is S(2015) > T(2015) ?
1-Jul	98	86				FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
2-Jul	97	90				FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
3-Jul	97	93				FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
4-Jul	90	91				FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
5-Jul	89	84				FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
6-Jul	93	84				FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
22-Sep	81	70				FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
23-Sep	84	80				FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
24-Sep	84	82				FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
25-Sep	87	66				FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
26-Sep	84	70				FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
27-Sep	79	64				FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
28-Sep	75	68				FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
29-Sep	72	77				FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
30-Sep	64	86				FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
1-Oct	66	75				FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
2-Oct	72	73				TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
3-Oct	84	75				FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE
4-Oct	70	78				TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE
5-Oct	66	81				TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE
6-Oct	64	82				TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE
7-Oct	60	82				TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE
8-Oct	78	82				TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE
9-Oct	70	80				TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE
10-Oct	72	82				TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE
11-Oct	69	82				TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE
12-Oct	69	79				TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE
13-Oct	73	80				TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE
14-Oct	79	68				TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE
15-Oct	81	63				TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE
16-Oct	80	57				TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE
17-Oct	82	66				TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE
18-Oct	66	64				TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE
19-Oct	63	69				TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
20-Oct	68	70				TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
21-Oct	79	70				TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
22-Oct	81	62				TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
23-Oct	69	63				TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
24-Oct	73	62				TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
25-Oct	73	75				TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
26-Oct	75	71				TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
27-Oct	75	57				TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
28-Oct	81	55				TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
29-Oct	82	64				TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
30-Oct	82	66				TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
31-Oct	81	60				TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

Figure 6.2.8 eye-balling the situation...

However to prove this via a CUSUM model… I first, counted how many days in before the change is detected by using COUNTIFS(range, FALSE)

=COUNTIFS(AS10:AS132, FALSE)

A	B	C ◂ ▸ V	W	AR	AS	AT	AU	AV	AW
Variables									
Target (T)	4	standard deviations from norm							
Dampener (C)	0.7	standard deviations from norm							
AVERAGE:	**83.72**	**81.67**							
STDDEV:	**8.55**	**9.32**							
Target (T)	**34.19**	**37.28**	**S values**	**Is S>T ?**					
Dampener (C)	**5.98**	**6.52**			1996	1997	1998	1999	2
DAY	1996	1997			Is S(1996) > T(1996) ?	Is S(1997) > T(1997) ?	Is S(1998) > T(1998) ?	Is (1999) > T(1999)	Is S(2000)>T 00)
1-Jul	98	86			FALSE	FALSE	FALSE	FALSE	FALSE
2-Jul	97	90			FALSE	FALSE	FALSE	FALSE	FALSE
28-Oct	81	55			TRUE	TRUE	TRUE	TRUE	TRUE
29-Oct	82	64			TRUE	TRUE	TRUE	TRUE	TRUE
30-Oct	82	66			TRUE	TRUE	TRUE	TRUE	TRUE
31-Oct	81	60			TRUE	TRUE	TRUE	TRUE	TRUE
			Days in before winter(unoff)		94	107	100	103	

Figure 6.2.8 Summarizing yearly data (I'm hiding many of the rows to fit the snapshot in cleanly)

The above highlighted cell shows that it was 94 days after 7/1 when the change was detected (that is the temperature got notably cooler)

I transposed the table to make it line up against years , and then just like I did for each year, I calculated, the Average, Std Dev, T, and C for each of these years as one data set (I kept T=4xSTDEV and C=0.7xSTDEV like before)

I was then able to evalue S(t) for each of the value, and compared that with T .

Year	Days after 7/1 before it gets cold	S(t)	Is S(t) > T ?
1996	94	0	FALSE
1997	107	4.771410584	FALSE
1998	100	0	FALSE
1999	103	0	FALSE
2000	99	0	FALSE
2001	103	0	FALSE

2002	106	0	FALSE
2003	93	0	FALSE
2004	104	5.771410584	FALSE
2005	114	0.5428211678	FALSE
2006	107	0	FALSE
2007	112	0	FALSE
2008	111	0	FALSE
2009	107	0	FALSE
2010	94	0	FALSE
2011	102	4.771410584	FALSE
2012	100	1.542821168	FALSE
2013	110	0.3142317517	FALSE
2014	106	0	FALSE
2015	93	0	FALSE
AVG:	**103.25**		
STDEV	**6.39798488**		
Target (T)	**25.59193952**		
Dampener (C)	**4.478589416**		

Table 6.2.9 CUSUM model against changes to winter start Year over year.

Clearly, the temperature has _not_ gotten noticeably cooler over the years.