Assignment2

Prashant Kubsad

### 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.

*Answer:* I have worked previously with a bank in a consumer credit department. We were clustering consumers to target the reward structure for each group. This information helped us to categorise all the users under different buckets. Each bucket had different reward awards. For ex.: for customers who had more restaurant merchants got 3% cash back on restaurant expenses while 2% cash back on retail expenses and 1% on fuel purchase. Similarly a different group, who travelled a lot for thier work, got 3% cash back on fuel purchases and 2% on dining expenses. There were a lot of data that was used to categorise customers. We had data at inidividual level, nature of expenses, payment history. Few of the predictors used:

* Average Balance: Average balance in the credit card account on the monthly basis.
* Cash advance frequency: How frequent customer takes cash advance on the credit card.
* Number of purchases: Number of purchases in each monthly statement.
* Purchase frequency: How often does a person use this credit card. There was a calculation to normalise this number on perday basis.
* Payments : Credit card bill payment history of the customer. The risk index was caluculated based on if the customer made 100% payments all the time, or part payments to meet minimum balance.
* Number of distinct merchants in the monthly bill.: number of different merchants in the monthly statement.

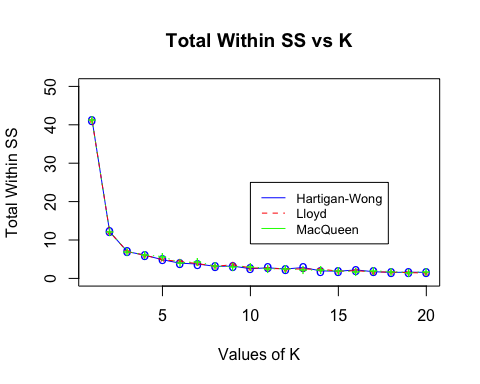
### Question 4.2

*Answer:* I am scaling the given data between range (0,1) and running kmeans() for K = 1:20 and for algorithms *Hartigan-Wong, Lloyd, MacQueen*.All the algorithms are giving almost similar response. Based on the elbow chart, we see the biggest bang for the buck is around *k=3*. I am using parameters *k=3 and algorithm=Lloyd* as my best parameters and re runnig the kmeans. The response of kmeans function has a value: ‘cluster’ which shows the cluster under which each row was classified. The ‘size’ in the response shows how many rows each cluster has: *size: 50 39 61*. By lookig at the ‘cluster’ value in the response, we can guess which cluster represents which species. I have printed that in the r output.

#cleaning environment and starting fresh.  
rm(list = ls())  
#setting seed for consistent results  
set.seed(1)  
#Read the text file into data.  
data<-read.table("iris.txt", header=TRUE)  
#normalizing data  
ndata<-data.frame(apply(data[,-5], MARGIN = 2, FUN = function(X) (X - min(X))/diff(range(X))))  
totalWithinss <-integer(length(20))  
totalWithinssForAlgorithms <-data.frame()  
for(alg in c("Hartigan-Wong", "Lloyd","MacQueen")){  
 i<-integer(0)  
 for(k in 1:20){  
 model<-kmeans(x=ndata,  
 centers=k,  
 iter.max = 50)  
 totalWithinss[k]<-model$tot.withinss  
 totalWithinssForAlgorithms[k,alg]<-model$tot.withinss  
 }  
 i<-i+1  
 totalWithinssForAlgorithms  
}  
totalWithinssForAlgorithms

## Hartigan-Wong Lloyd MacQueen  
## 1 41.166110 41.166110 41.166110  
## 2 12.127791 12.127791 12.127791  
## 3 6.982216 6.982216 6.982216  
## 4 6.049224 6.049224 6.083904  
## 5 4.859706 4.859706 5.720472  
## 6 3.961393 4.406693 4.073179  
## 7 3.749654 3.855648 4.355876  
## 8 3.143803 3.134198 3.150286  
## 9 3.202909 3.582673 3.096001  
## 10 2.532787 2.739131 3.124648  
## 11 2.750905 2.595359 2.478849  
## 12 2.454984 2.491045 2.584428  
## 13 2.826114 2.425108 2.238245  
## 14 1.961615 2.429972 2.108170  
## 15 1.871210 1.869962 1.922104  
## 16 2.204677 1.999534 1.806632  
## 17 1.746345 1.695807 1.960859  
## 18 1.560208 1.647148 1.798213  
## 19 1.615875 1.497341 1.518846  
## 20 1.586018 1.404260 1.739975

plot(1:20, totalWithinssForAlgorithms[,1], type="o", col="blue", pch="o", lty=1,  
 ylim=c(0,50),xlab="Values of K", ylab="Total Within SS",main="Total Within SS vs K")  
 points(1:20, totalWithinssForAlgorithms[,2], col="red", pch="\*")  
 lines(1:20, totalWithinssForAlgorithms[,2], col="red",lty=2)  
 points(1:20, totalWithinssForAlgorithms[,3], col="green",pch="+")  
 lines(1:20, totalWithinssForAlgorithms[,3], col="green", lty=3)  
 legend(10, 25, legend=c("Hartigan-Wong", "Lloyd" , "MacQueen"),  
 col=c("blue", "red","green"), lty=1:2, cex=0.8)



bestModel<-kmeans(x=ndata,  
 centers=3,  
 algorithm = "Lloyd")  
cat("size: ", bestModel$size)

## size: 39 50 61

response<-table(bestModel$cluster,data[,5])  
cat("\ncross reference table between model's clusters (1,2,3) and species(setosa,versicolor,virginica)\n")

##   
## cross reference table between model's clusters (1,2,3) and species(setosa,versicolor,virginica)

response

##   
## setosa versicolor virginica  
## 1 0 3 36  
## 2 50 0 0  
## 3 0 47 14

# based on the table output, we can infer the cluster number for each species by finding max value  
species<-colnames(response)[apply(response,1,which.max)]  
resultTable<-as.data.frame(species)  
resultTable$clusterNumber=c(1,2,3)  
resultTable

## species clusterNumber  
## 1 virginica 1  
## 2 setosa 2  
## 3 versicolor 3

# based on the above table, we can find the accuracy by (total Correct Guesses)/(totalRows)  
cat("\ncorrect guesses for cluster 1: 36, Incorrect guesses for cluster 1: 3" );

##   
## correct guesses for cluster 1: 36, Incorrect guesses for cluster 1: 3

cat("\ncorrect guesses for cluster 2: 47, Incorrect guesses for cluster 1: 14" );

##   
## correct guesses for cluster 2: 47, Incorrect guesses for cluster 1: 14

cat("\ncorrect guesses for cluster 3: 50, Incorrect guesses for cluster 1: 0" );

##   
## correct guesses for cluster 3: 50, Incorrect guesses for cluster 1: 0

accuracy = (50+47+36)/nrow(data)  
cat("\naccuracy: ",accuracy)

##   
## accuracy: 0.8866667

Even though kmeans is not meant for classification and to calculate the accuracy, since the outcomes were provided in the test data(supervised data), we can calculate accuracy which in my case came to be *88.66%*

### Question 5.1

*Answer:* Using the grubbs.test(), for the first iteration with full data, we see the alternative hypothesis that *1993* is an outlier with *p=0.07*. There are 2 hypothesis in the grubbs test.

* Null Hypothesis: This says the outermost point is not an outlier.
* Alternative Hypothesis: This says the outermost point is an outlier.

The p-value is a probability that measures the evidence against the null hypothesis. A smaller p-value provides stronger evidence against the null hypothesis.I am going with p<0.10 as a outlier. for 1993, the p value was less than 0.1, am going to consider it as an outlier and re run the grubbs test excluding that record. This second run will tell me if the next highest record is an outlier or not. With this approach we can see that the data points *1993 and 1969 are outliers*. For the next outlier: 1674 the value of p goes up to 0.17 which inidicates it might not be an outlier. The code for this:

#cleaning environment and starting fresh.  
library(outliers)  
rm(list = ls())  
#setting seed for consistent results  
set.seed(1)  
#Read the text file into data.  
data<-read.table("uscrime.txt", header=TRUE)  
output<-grubbs.test(x=data[,16],  
 type=10,   
 opposite = FALSE,   
 two.sided = FALSE)  
cat(output$alternative, "with p value of:", output$p.value)

## highest value 1993 is an outlier with p value of: 0.07887486

#excluding the outlier row 24 and re running the test  
output<-grubbs.test(x=data[-26,16],  
 type=10,   
 opposite = FALSE,   
 two.sided = FALSE)  
cat("\n",output$alternative, "with p value of:", output$p.value)

##   
## highest value 1969 is an outlier with p value of: 0.02847821

#excluding the outlier rows 24 and 6 and re running the test  
output<-grubbs.test(x=data[-c(26,4),16],  
 type=10,   
 opposite = FALSE,   
 two.sided = FALSE)  
cat("\n",output$alternative, "with p value of:", output$p.value)

##   
## highest value 1674 is an outlier with p value of: 0.1780797

The same can be visualised with box-feather plot that was discussed in the lecture. I have created box with 25-75 percentile data and calculated the upper fence as 1.5 times the Inter Quarantile Range(IQR). IQR also known as mid-spread or middle-50% is range of the data that lies between 25th percentile to 75th percentile of data. We can establish the upper fence or(the boundary upto which feathers can extend) to 1.5 times IQR. You can see there 2 points which are clearly above this fence which are our outliers identified above. The third point which is very close to the upper fence cannot be confidently classified as an outlier.

q1<-quantile(data[,16], c(.25))  
q3<-quantile(data[,16], c(.75))  
IQR<-IQR(data[,16])  
  
cat("25th percentile: ",q1)

## 25th percentile: 658.5

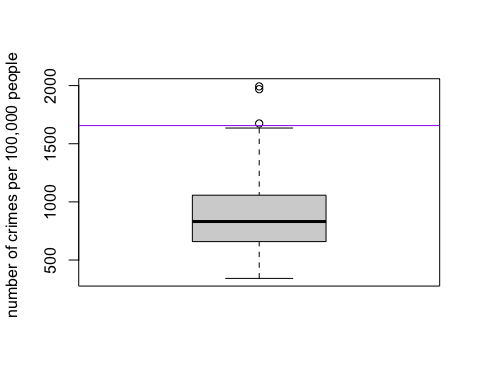
cat("755th percentile: ",q3)

## 755th percentile: 1057.5

cat("IQR: ",IQR)

## IQR: 399

upperFence<- q3+(1.5 \* IQR);  
  
  
boxplot(data[,16], ylab = 'number of crimes per 100,000 people ')  
abline(h=upperFence, col="purple")



### Question 6.1

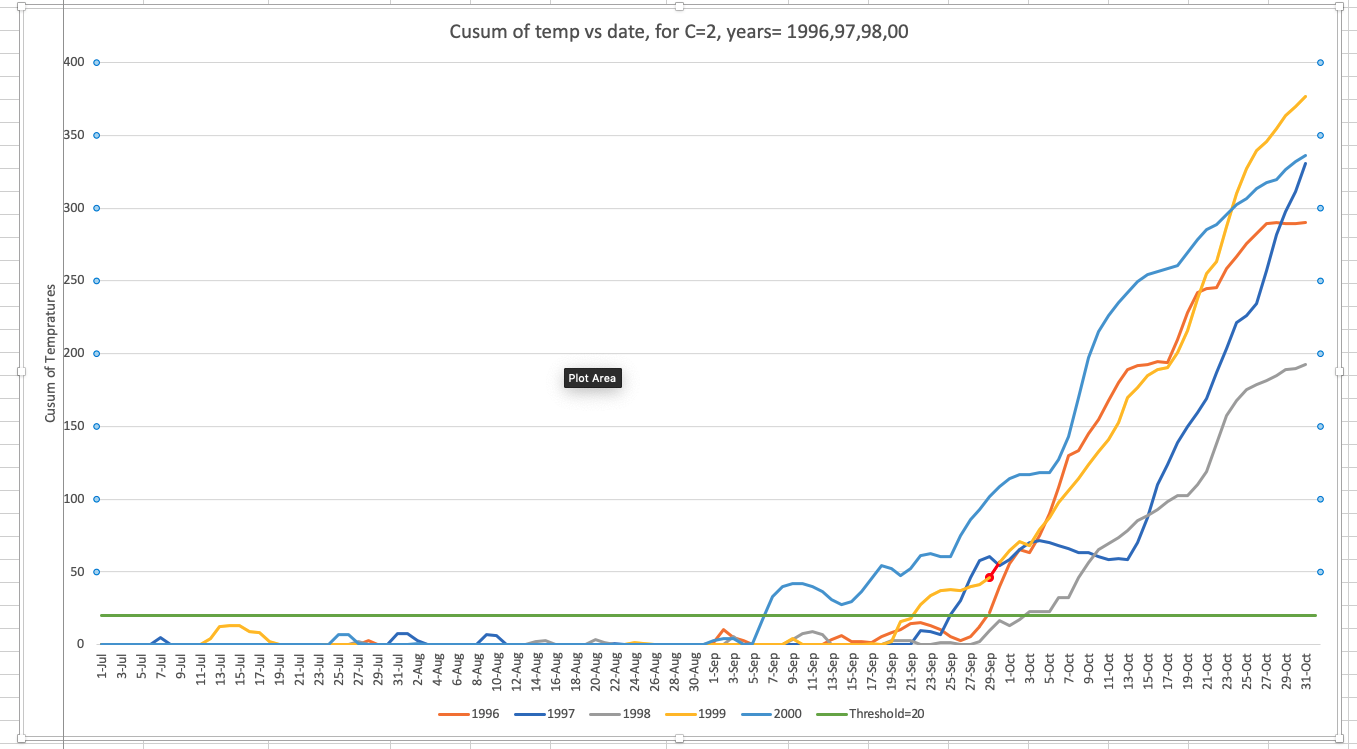
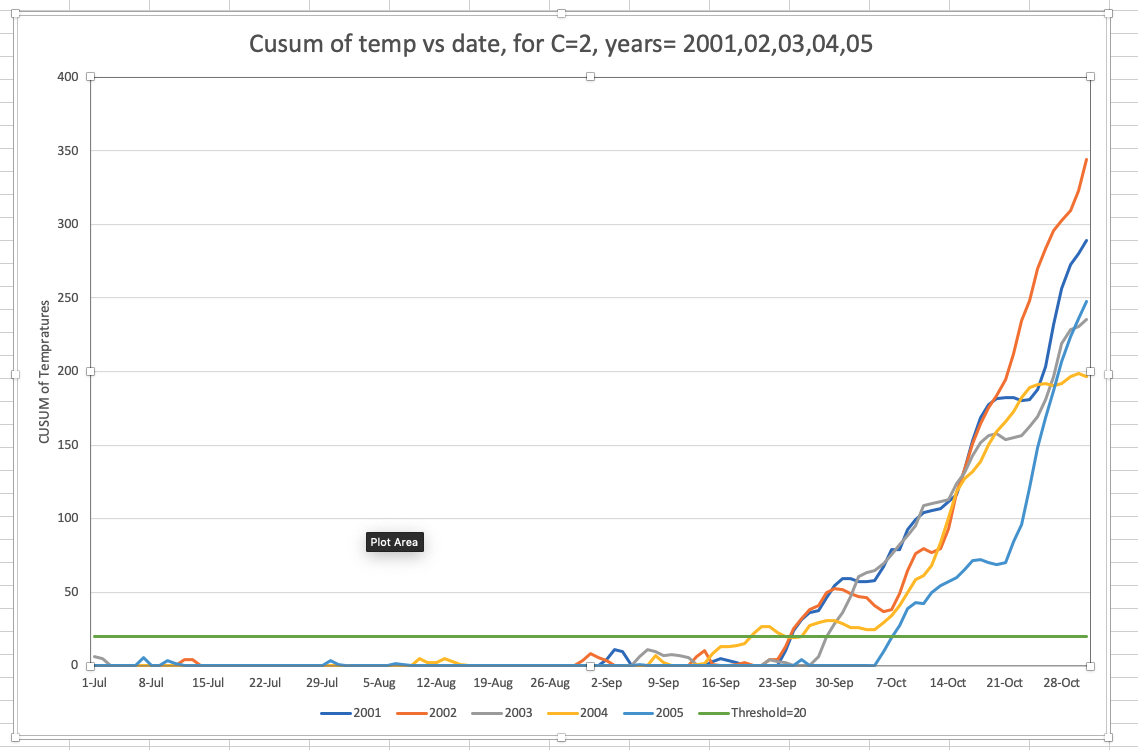
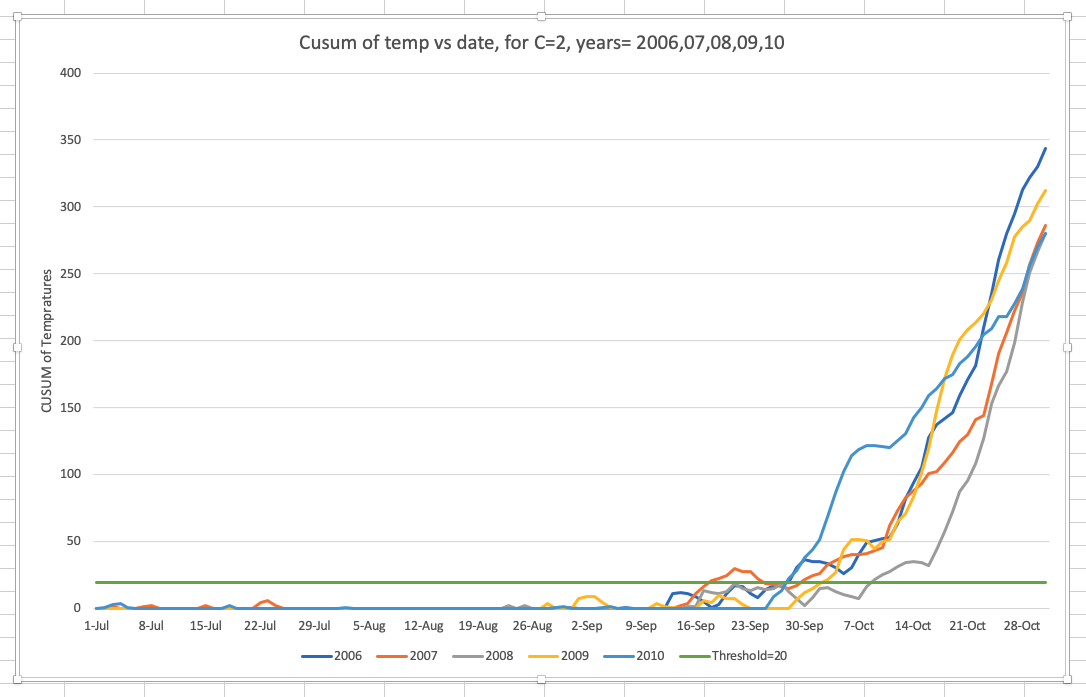
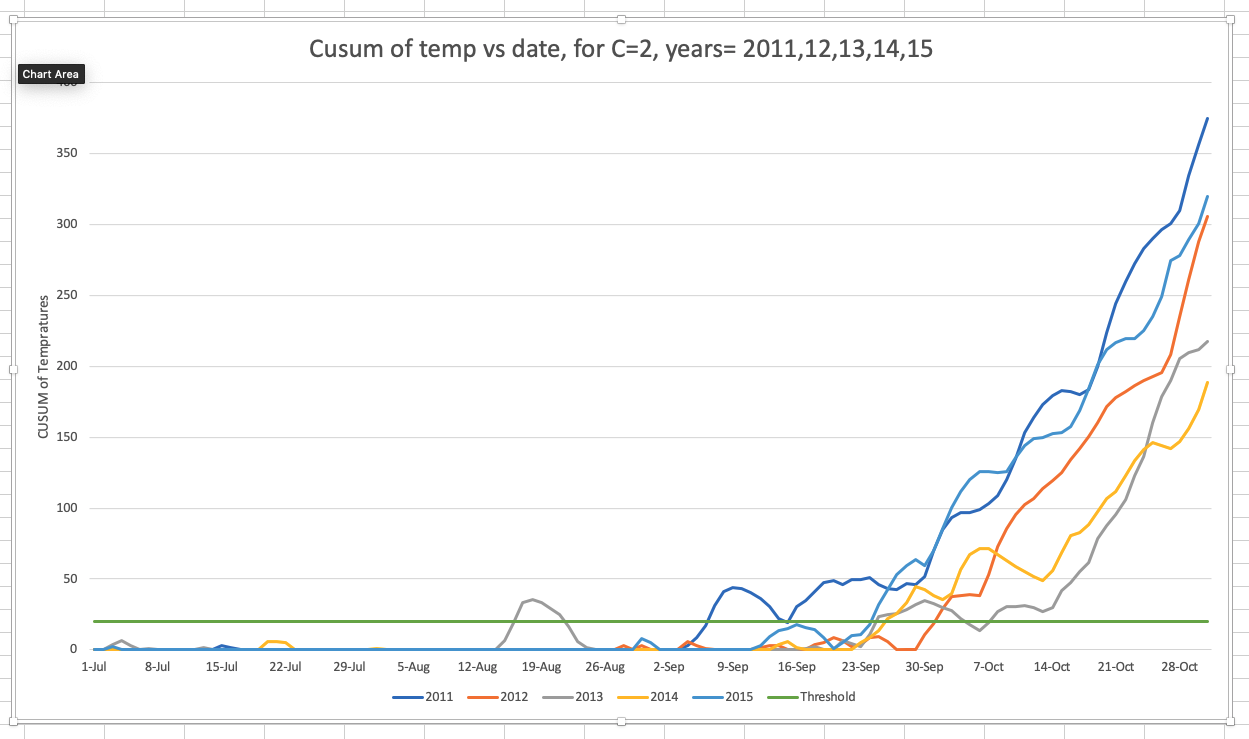
*Answer:* In my current work we work with a lot of mortgage insurance on home loans. There are various factors contribute to the monthly business a company makes. One such measure is the amount of *‘New Insurance Written (NIW)’*. We measure monthly NIW rate and keep monitoring it for changes. At the tail end of 2019, we started seeing drop in NIW but a rise in refinance of old loans. When we further analysed the change detection it was directly related to the feds slashing the interest rates. When Interest Rates fell to historic low values, more people started to refinance thier mortgage loans. Using CUSUM technique we can set a threshold of 15% above or below in new insurance written to identify the change. Once the change was identified, the company started introducing programs to get the major share of the refinance market.

### Question 6.2.1

*Answer:* I have followed the approach mentioned in the lecture 6.2 to findout when the unofficial winter starts for each year.Lets take year 1996. I have taken value of c=0 and c=2 and have tabulated the , , and , since this is a decreasing trend we have to use instead of . You can see that we start seeing the cumulative sum start growing around *Sept 13th*. It is always better to consider a threshold value as explained in the lecture to avoid reporting false positives as change detected. To accomodate that, we have 2 parameters we can fine tune. The value *‘C’* and the Threshold. Below I have generated table with the above formula calculations on the year ‘1996’. For the sake of presentation I have omitted few dates just in printing, i have not omitted them from calculations. The full spread sheet with all the calculations is in another attachment. We see the value of C=2 and Threhold = 20 smoothes the cumulative curve and am going with those as my parameters. *With C=2 and Threshold=20* i am seeing the unofficial end of summer for the *year 1996 is 29-Sept*, because thats when cumulative sum goes beyond the threshold value.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| DAY | 1996 | (1996)mu-xt | (1996)mu-xt-c | (1996)St, c=0 | (1996)mu-xt-c | (1996)St,C=2 |
| mu(avg) | 83.71544715 |  | c=0 |  | c=2 |  |
| 1-Jul | 98 | -14.28455285 | -14.28455285 | 0 | -16.28455285 | 0 |
| 2-Jul | 97 | -13.28455285 | -13.28455285 | 0 | -15.28455285 | 0 |
| 3-Jul | 97 | -13.28455285 | -13.28455285 | 0 | -15.28455285 | 0 |
| 4-Jul | 90 | -6.284552846 | -6.284552846 | 0 | -8.284552846 | 0 |
| 5-Jul | 89 | -5.284552846 | -5.284552846 | 0 | -7.284552846 | 0 |
| 6-Jul | 93 | -9.284552846 | -9.284552846 | 0 | -11.28455285 | 0 |
| 24-Aug | 91 | -7.284552846 | -7.284552846 | 0 | -9.284552846 | 0 |
| 25-Aug | 84 | -0.284552846 | -0.284552846 | 0 | -2.284552846 | 0 |
| 26-Aug | 88 | -4.284552846 | -4.284552846 | 0 | -6.284552846 | 0 |
| 27-Aug | 84 | -0.284552846 | -0.284552846 | 0 | -2.284552846 | 0 |
| 28-Aug | 86 | -2.284552846 | -2.284552846 | 0 | -4.284552846 | 0 |
| 29-Aug | 88 | -4.284552846 | -4.284552846 | 0 | -6.284552846 | 0 |
| 30-Aug | 84 | -0.284552846 | -0.284552846 | 0 | -2.284552846 | 0 |
| 31-Aug | 82 | 1.715447154 | 1.715447154 | 1.715447154 | -0.284552846 | 0 |
| 1-Sep | 80 | 3.715447154 | 3.715447154 | 5.430894309 | 1.715447154 | 1.715447154 |
| 2-Sep | 73 | 10.71544715 | 10.71544715 | 16.14634146 | 8.715447154 | 10.43089431 |
| 3-Sep | 87 | -3.284552846 | -3.284552846 | 12.86178862 | -5.284552846 | 5.146341463 |
| 4-Sep | 84 | -0.284552846 | -0.284552846 | 12.57723577 | -2.284552846 | 2.861788618 |
| 5-Sep | 87 | -3.284552846 | -3.284552846 | 9.292682927 | -5.284552846 | 0 |
| 6-Sep | 89 | -5.284552846 | -5.284552846 | 4.008130081 | -7.284552846 | 0 |
| 7-Sep | 89 | -5.284552846 | -5.284552846 | 0 | -7.284552846 | 0 |
| 8-Sep | 89 | -5.284552846 | -5.284552846 | 0 | -7.284552846 | 0 |
| 9-Sep | 91 | -7.284552846 | -7.284552846 | 0 | -9.284552846 | 0 |
| 10-Sep | 84 | -0.284552846 | -0.284552846 | 0 | -2.284552846 | 0 |
| 11-Sep | 86 | -2.284552846 | -2.284552846 | 0 | -4.284552846 | 0 |
| 12-Sep | 88 | -4.284552846 | -4.284552846 | 0 | -6.284552846 | 0 |
| 13-Sep | 78 | 5.715447154 | 5.715447154 | 5.715447154 | 3.715447154 | 3.715447154 |
| 14-Sep | 79 | 4.715447154 | 4.715447154 | 10.43089431 | 2.715447154 | 6.430894309 |
| 15-Sep | 86 | -2.284552846 | -2.284552846 | 8.146341463 | -4.284552846 | 2.146341463 |
| 16-Sep | 82 | 1.715447154 | 1.715447154 | 9.861788618 | -0.284552846 | 1.861788618 |
| 17-Sep | 82 | 1.715447154 | 1.715447154 | 11.57723577 | -0.284552846 | 1.577235772 |
| 18-Sep | 78 | 5.715447154 | 5.715447154 | 17.29268293 | 3.715447154 | 5.292682927 |
| 19-Sep | 79 | 4.715447154 | 4.715447154 | 22.00813008 | 2.715447154 | 8.008130081 |
| 20-Sep | 79 | 4.715447154 | 4.715447154 | 26.72357724 | 2.715447154 | 10.72357724 |
| 21-Sep | 78 | 5.715447154 | 5.715447154 | 32.43902439 | 3.715447154 | 14.43902439 |
| 22-Sep | 81 | 2.715447154 | 2.715447154 | 35.15447154 | 0.715447154 | 15.15447154 |
| 23-Sep | 84 | -0.284552846 | -0.284552846 | 34.8699187 | -2.284552846 | 12.8699187 |
| 24-Sep | 84 | -0.284552846 | -0.284552846 | 34.58536585 | -2.284552846 | 10.58536585 |
| 25-Sep | 87 | -3.284552846 | -3.284552846 | 31.30081301 | -5.284552846 | 5.300813008 |
| 26-Sep | 84 | -0.284552846 | -0.284552846 | 31.01626016 | -2.284552846 | 3.016260163 |
| 27-Sep | 79 | 4.715447154 | 4.715447154 | 35.73170732 | 2.715447154 | 5.731707317 |
| 28-Sep | 75 | 8.715447154 | 8.715447154 | 44.44715447 | 6.715447154 | 12.44715447 |
| 29-Sep | 72 | 11.71544715 | 11.71544715 | 56.16260163 | 9.715447154 | 22.16260163 |
| 30-Sep | 64 | 19.71544715 | 19.71544715 | 75.87804878 | 17.71544715 | 39.87804878 |
| 1-Oct | 66 | 17.71544715 | 17.71544715 | 93.59349593 | 15.71544715 | 55.59349593 |
| 2-Oct | 72 | 11.71544715 | 11.71544715 | 105.3089431 | 9.715447154 | 65.30894309 |
| 3-Oct | 84 | -0.284552846 | -0.284552846 | 105.0243902 | -2.284552846 | 63.02439024 |
| 4-Oct | 70 | 13.71544715 | 13.71544715 | 118.7398374 | 11.71544715 | 74.7398374 |
| 5-Oct | 66 | 17.71544715 | 17.71544715 | 136.4552846 | 15.71544715 | 90.45528455 |
| 6-Oct | 64 | 19.71544715 | 19.71544715 | 156.1707317 | 17.71544715 | 108.1707317 |
| 7-Oct | 60 | 23.71544715 | 23.71544715 | 179.8861789 | 21.71544715 | 129.8861789 |
| 8-Oct | 78 | 5.715447154 | 5.715447154 | 185.601626 | 3.715447154 | 133.601626 |

Similar calculations are done on each year in the attached excel spread sheet. I have generated the graphs of cumulative sum growth vs dates for each year. Since 20 lines in one graph was difficult to read, i have split it into 4 graphs with 5 years in each.The green horizontal line is the threshold line with value 20. Wherever this line crosses each line, we can consider that point as unofficial end of summer for that year. For ex. in the first graph, this line intersects the 1996 year’s line(orange line) at Sept 29th.

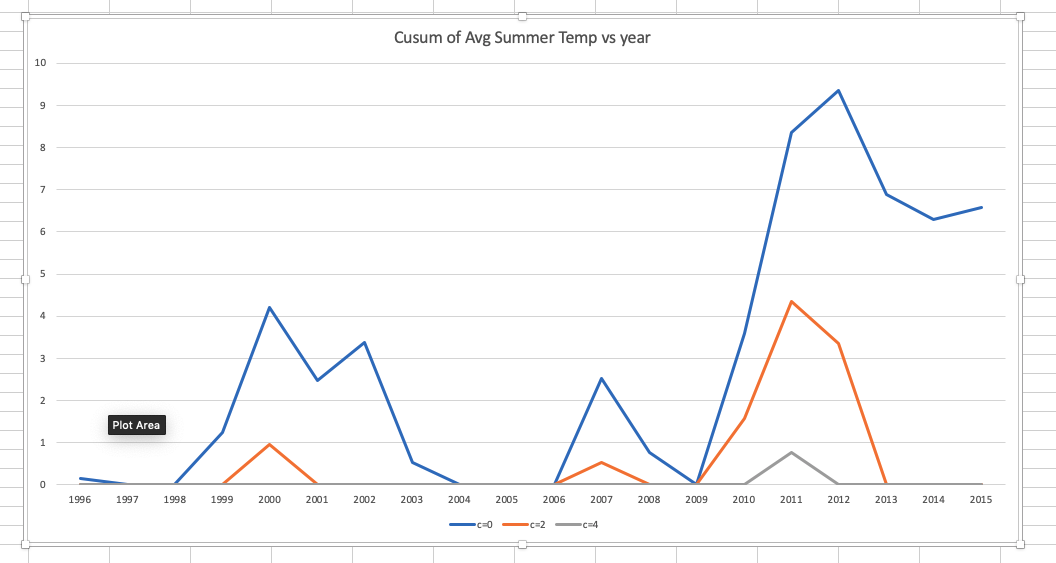
    In the attached spread sheet you can see when the cumulative sum goes past threshold for each year. The results for each year is as below.:

|  |  |
| --- | --- |
| Year | Unofficial End of Summer |
| 1996 | 29- Sept |
| 1997 | 25 -Sep |
| 1998 | 3-Oct |
| 1999 | 22-Sept |
| 2000 | 7-Sept |
| 2001 | 25-Sept |
| 2002 | 25-Sept |
| 2003 | 30-Sept |
| 2004 | 20-Sept |
| 2005 | 8-Oct |
| 2006 | 29-Sept |
| 2007 | 18-Sept |
| 2008 | 9-Oct |
| 2009 | 3-Oct |
| 2010 | 28-Sept |
| 2011 | 7-Sept |
| 2012 | 2-Oct |
| 2013 | 25-Sept |
| 2014 | 26-Sept |
| 2015 | 25-Sept |

### Question 6.2

*Answer:* In the previous question, we have determined the unofficial end of summer for each year. To determine the summer temprature changes we have to consider only the summer tempratures. Based on the dates given above, I have calculated the average temprature upto that date for each year. Ex.: For year 1996, 29-Sept is unofficial end of summer, so average summer temp for 1996 will be average of tempratures until 28-Sept which is *87.53*. Similarly I have captured for each year. We can run change detection on these averages to identify change. This is a increasing trend measurement so we wil calculate instead of . With both C=2 and threshold=1 we can see the average summer tempratures started to increase in the *year 2010*.

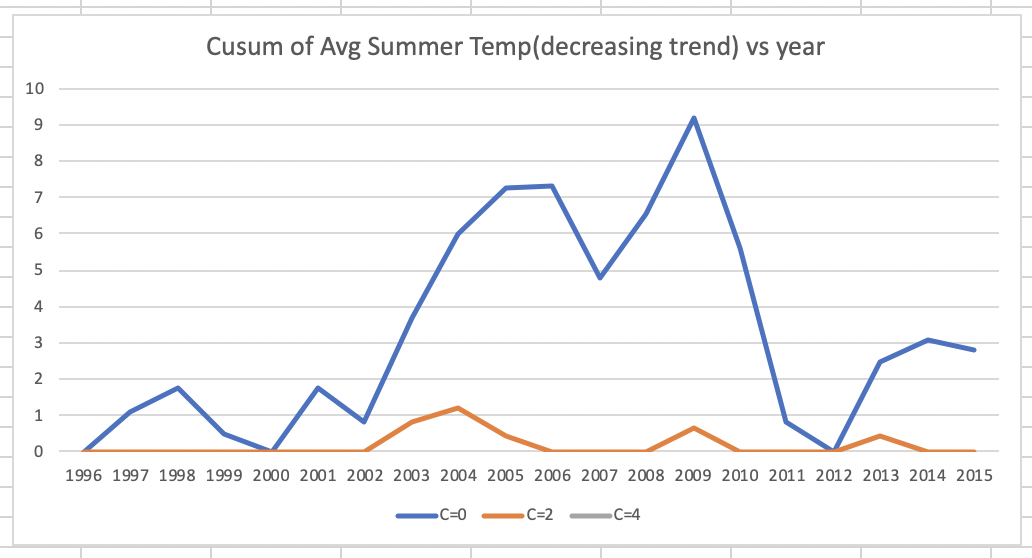
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Year | Avg Summer Temp(Xt) | Xt-mu | xt-mu-c, c=0 | *St, c=0* | xt-mu-c,c=2 | *St, c=2* | xt-mu-c, c=4 | *St,c=4* |
|  |  | =87.386 |  |  |  |  |  |  |
| 1996 | 87.53 | 0.144 | 0.144 | 0.144 | -1.856 | 0 | -3.856 | 0 |
| 1997 | 86.3 | -1.086 | -1.086 | 0 | -3.086 | 0 | -5.086 | 0 |
| 1998 | 86.74 | -0.646 | -0.646 | 0 | -2.646 | 0 | -4.646 | 0 |
| 1999 | 88.64 | 1.254 | 1.254 | 1.254 | -0.746 | 0 | -2.746 | 0 |
| 2000 | 90.34 | 2.954 | 2.954 | 4.208 | 0.954 | 0.954 | -1.046 | 0 |
| 2001 | 85.65 | -1.736 | -1.736 | 2.472 | -3.736 | 0 | -5.736 | 0 |
| 2002 | 88.3 | 0.914 | 0.914 | 3.386 | -1.086 | 0 | -3.086 | 0 |
| 2003 | 84.54 | -2.846 | -2.846 | 0.54 | -4.846 | 0 | -6.846 | 0 |
| 2004 | 85.04 | -2.346 | -2.346 | 0 | -4.346 | 0 | -6.346 | 0 |
| 2005 | 86.15 | -1.236 | -1.236 | 0 | -3.236 | 0 | -5.236 | 0 |
| 2006 | 87.34 | -0.046 | -0.046 | 0 | -2.046 | 0 | -4.046 | 0 |
| 2007 | 89.92 | 2.534 | 2.534 | 2.534 | 0.534 | 0.534 | -1.466 | 0 |
| 2008 | 85.61 | -1.776 | -1.776 | 0.758 | -3.776 | 0 | -5.776 | 0 |
| 2009 | 84.73 | -2.656 | -2.656 | 0 | -4.656 | 0 | -6.656 | 0 |
| 2010 | 90.97 | 3.584 | 3.584 | 3.584 | 1.584 | *1.584* | -0.416 | 0 |
| 2011 | 92.16 | 4.774 | 4.774 | 8.358 | 2.774 | 4.358 | 0.774 | *0.774* |
| 2012 | 88.38 | 0.994 | 0.994 | 9.352 | -1.006 | 3.352 | -3.006 | 0 |
| 2013 | 84.93 | -2.456 | -2.456 | 6.896 | -4.456 | 0 | -6.456 | 0 |
| 2014 | 86.78 | -0.606 | -0.606 | 6.29 | -2.606 | 0 | -4.606 | 0 |
| 2015 | 87.67 | 0.284 | 0.284 | 6.574 | -1.716 | 0 | -3.716 | 0 |



cusum\_chart\_1.png

Just out of curiosity I did cusum on the decreasing trend on the avg temp. I found out that the summers of Atlanta were getting cooler in around 2002/03 before they started to get warmer agian in 2010. The details of the decreasing cusum:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Year | Avg Summer Temp(Xt) | mu-xt | mu-xt-c, c=0 | St, c=0 | mu-xt-c,c=2 | St, c=2 |
| 1996 | 87.53 | -0.144 | -0.144 | 0 | -2.144 | 0 |
| 1997 | 86.3 | 1.086 | 1.086 | 1.086 | -0.914 | 0 |
| 1998 | 86.74 | 0.646 | 0.646 | 1.732 | -1.354 | 0 |
| 1999 | 88.64 | -1.254 | -1.254 | 0.478 | -3.254 | 0 |
| 2000 | 90.34 | -2.954 | -2.954 | 0 | -4.954 | 0 |
| 2001 | 85.65 | 1.736 | 1.736 | 1.736 | -0.264 | 0 |
| 2002 | 88.3 | -0.914 | -0.914 | 0.822 | -2.914 | 0 |
| 2003 | 84.54 | 2.846 | 2.846 | 3.668 | 0.846 | 0.846 |
| 2004 | 85.04 | 2.346 | 2.346 | 6.014 | 0.346 | 1.192 |
| 2005 | 86.15 | 1.236 | 1.236 | 7.25 | -0.764 | 0.428 |
| 2006 | 87.34 | 0.046 | 0.046 | 7.296 | -1.954 | 0 |
| 2007 | 89.92 | -2.534 | -2.534 | 4.762 | -4.534 | 0 |
| 2008 | 85.61 | 1.776 | 1.776 | 6.538 | -0.224 | 0 |
| 2009 | 84.73 | 2.656 | 2.656 | 9.194 | 0.656 | 0.656 |
| 2010 | 90.97 | -3.584 | -3.584 | 5.61 | -5.584 | 0 |
| 2011 | 92.16 | -4.774 | -4.774 | 0.836 | -6.774 | 0 |
| 2012 | 88.38 | -0.994 | -0.994 | 0 | -2.994 | 0 |
| 2013 | 84.93 | 2.456 | 2.456 | 2.456 | 0.456 | 0.456 |
| 2014 | 86.78 | 0.606 | 0.606 | 3.062 | -1.394 | 0 |
| 2015 | 87.67 | -0.284 | -0.284 | 2.778 | -2.284 | 0 |



cusum\_chart\_7.png

Attachment details

* Sheet 6.2.1 : Cusum of temp for each year
* Sheet 6.2.1.Experiment: Experiment with C values in cusum of summer temp for the year 1996
* Sheet 6.2.2: Cusum of average summer tempratures for all the years to determin when summers got hotter
* Sheet 6.2.2.Experiment: Experiment on decreasing trend of average summer temp to see if summers got cooler