ISYE HW 3- Change Detection

Ryan Cherry

2024-01-24

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.

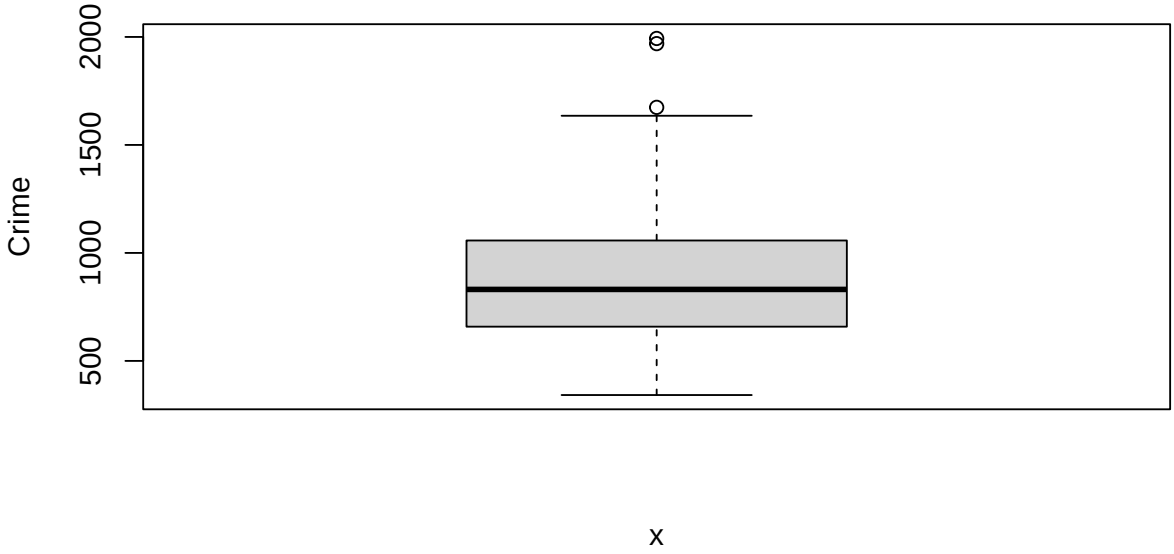
The Grubb's test is a procedure meant to determine whether a single point is an outlier or not, with the null hypothesis that there is no outlier and the alternative that there is exactly one outlier. However, the test is only reasonable if the data are normally distributed. As a result, the Crime variable was visually inspected using a box plot and a histogram.

```
#crime data set was read in
crime_data <- read.table("http://www.statsci.org/data/general/uscrime.txt", header = TRUE)
#view the column from which outliers will be investigated
summary(crime_data$Crime)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   342.0   658.5   831.0   905.1  1057.5  1993.0
```

The boxplot sown below reveals that there are three points far away from the rest of the data that may be outliers. In addition, the box plot seems to indicate that the data could be right-skewed.

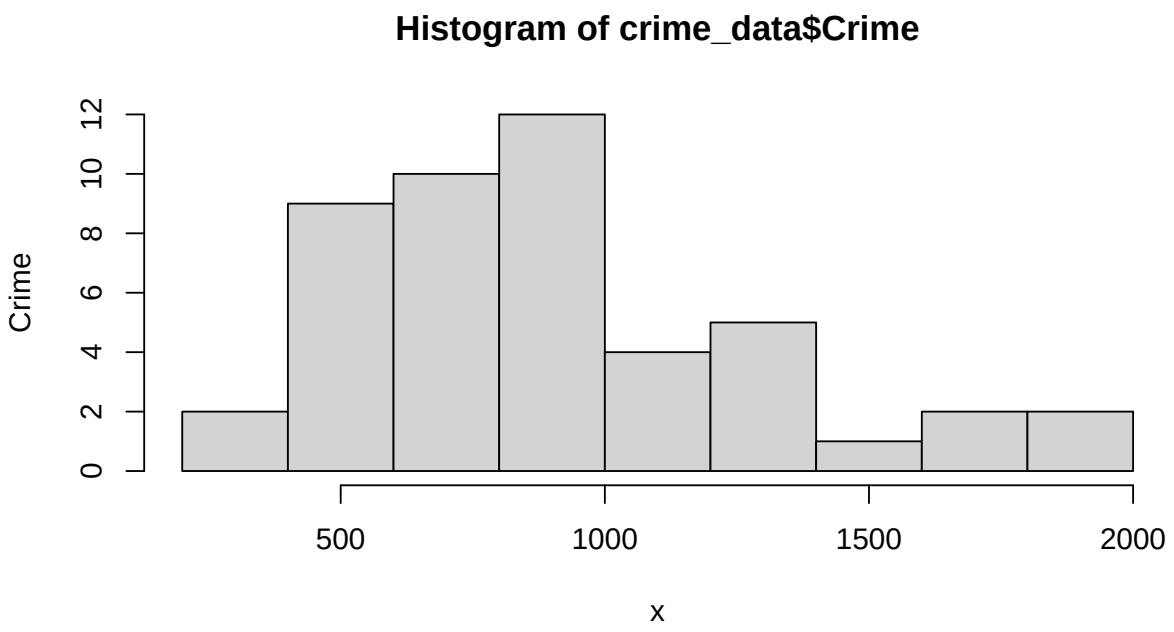
```
#visualize the Crime variable using a boxplot
boxplot(crime_data$Crime, data = crime_data, xlab = "x", ylab = "Crime")
```



A histogram was also created to get a visual assessment of the distribution and, just like in the box plot above, the Crime variable appears to be heavily skewed to the right. As a result, the Grubb's test is not appropriate to use on this data set but we will run it anyway.

```
#view the distribution of the data since the Grubbs test requires normality
hist(crime_data$Crime, data = crime_data, xlab = "x", ylab = "Crime")
```

```
## Warning in plot.window(xlim, ylim, "", ...): "data" is not a graphical parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...): "data"
## is not a graphical parameter

## Warning in axis(1, ...): "data" is not a graphical parameter

## Warning in axis(2, at = yt, ...): "data" is not a graphical parameter
```



The Grubb's test was run and we obtained a p-value above 0.05, meaning we would accept the alternative hypothesis that 1993 is an outlier. However, since the p-value is not too far from 0.05 and since the normality condition is not met, I decided to remove the data point and run the Grubb's test again to see if there were any more outliers.

The test is meant to be run until no more outliers are detected.

```
#perform grubbs test for one outlier
grubbs.test(crime_data$Crime, type = 10)
```

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

The Grubb's test was run again with 1993 removed and this time, a p-value of 0.028 was obtained, less than the cutoff of 0.05. In this case, we reject the null hypothesis and conclude that the point 1969 is an outlier.

The Grubb's test was run again after this and no more outliers were detected.

```
#We would fail to reject the null hypothesis so 1993 would not be an outlier. However, normality is req

#remove data point that was tested in first run of grubbs test

crime_data2 <- crime_data[(-26), ]

#rerun grubbs test with previous outlier removed

grubbs.test(crime_data2$Crime, type = 10)
```

```
##
##  Grubbs test for one outlier
##
## data:  crime_data2$Crime
## G = 3.06343, U = 0.78682, p-value = 0.02848
## alternative hypothesis: highest value 1969 is an outlier
```

Overall, the Grubb's test was used, even though the normality condition was not met, and one outlier was found, 1969, and another potential outlier, 1993, as well. With more time, the first outlier the Grubb's test detected, 1996, would be investigated further to see whether it truly is an outlier or actual data.

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?

My undergraduate degree was in Textile Technology, which involved gaining knowledge of various processes for turning fibers into yarn and then into a knitted item, a woven item, or a non-woven. In several processes, particularly in spinning a fiber into a yarn, a change detection model would be appropriate. For example, in one step of the spinning process, individual fibers are straightened out on a machine called a drawing frame. The finished product from the machine is a sliver of straightened fibers. We can then test several strands of sliver to determine the percentage of fibers that are straight. If that percentage falls below a certain threshold, that would be an indication that a particular batch of fiber is defective or the teeth inside the draw frame that straighten the individual fibers out are worn out and may need to be replaced. The critical value can be derived from ASTM testing standards, which are test standards used in the textile industry. A good threshold in a textile context is no more than 2 standard deviations from the mean.

QUESTION 6.2

a. 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, for example at http://www.iweathernet.com/atlanta-weather-records or https://www.wunderground.com/history/airport/KFTY/2015/7/1/CustomHistory.html . You can use R if you'd like, but it's straightforward enough that an Excel spreadsheet can easily do the job too

A quick overview of the temperature data set showed that the years included were 1996 up through 2015. The summary indicates that the average temperature for each year in the data set was in the range of 81 to 86 degrees Fahrenheit. Based on this, it seems like the average temperature has not changed much in Atlanta over this 20-year period, but a CUSUM approach will be run in order to determine when the summer ends.

```
#read in Atlanta temperature data set
Atlanta_temps <- read.table("https://prod-edxapp.edx-cdn.org/assets/courseware/v1/592f3be3e90d2bdfe6a69

#get a quick summary of mean and standard deviation for each year
summary(Atlanta_temps)
```

```
##      DAY                X1996           X1997           X1998
##  Length:123         Min.   :60.00   Min.   :55.00   Min.   :63.00
##  Class :character   1st Qu.:79.00   1st Qu.:78.50   1st Qu.:79.50
##  Mode  :character   Median :84.00   Median :84.00   Median :86.00
##                     Mean   :83.72   Mean   :81.67   Mean   :84.26
##                     3rd Qu.:90.00   3rd Qu.:88.50   3rd Qu.:89.00
##                     Max.   :99.00   Max.   :95.00   Max.   :95.00
```

```
##      X1999           X2000           X2001           X2002
##  Min.   :57.00   Min.   : 55.00   Min.   :51.00   Min.   :57.00
##  1st Qu.:75.00   1st Qu.: 77.00   1st Qu.:78.00   1st Qu.:78.00
##  Median :86.00   Median : 86.00   Median :84.00   Median :87.00
##  Mean   :83.36   Mean   : 84.03   Mean   :81.55   Mean   :83.59
##  3rd Qu.:91.00   3rd Qu.: 91.00   3rd Qu.:87.00   3rd Qu.:91.00
##  Max.   :99.00   Max.   :101.00   Max.   :93.00   Max.   :97.00
##      X2003           X2004           X2005           X2006
##  Min.   :57.00   Min.   :62.00   Min.   :54.00   Min.   :53.00
##  1st Qu.:78.00   1st Qu.:78.00   1st Qu.:81.50   1st Qu.:79.00
##  Median :84.00   Median :82.00   Median :85.00   Median :85.00
##  Mean   :81.48   Mean   :81.76   Mean   :83.36   Mean   :83.05
##  3rd Qu.:87.00   3rd Qu.:87.00   3rd Qu.:88.00   3rd Qu.:91.00
##  Max.   :91.00   Max.   :95.00   Max.   :94.00   Max.   :98.00
##      X2007           X2008           X2009           X2010
##  Min.   : 59.0   Min.   :50.00   Min.   :51.00   Min.   :67.00
##  1st Qu.: 81.0   1st Qu.:79.50   1st Qu.:75.00   1st Qu.:82.00
##  Median : 86.0   Median :85.00   Median :83.00   Median :90.00
##  Mean   : 85.4   Mean   :82.51   Mean   :80.99   Mean   :87.21
##  3rd Qu.: 89.5   3rd Qu.:88.50   3rd Qu.:88.00   3rd Qu.:93.00
##  Max.   :104.0   Max.   :95.00   Max.   :95.00   Max.   :97.00
##      X2011           X2012           X2013           X2014
##  Min.   :59.00   Min.   : 56.00   Min.   :56.00   Min.   :63.00
##  1st Qu.:79.00   1st Qu.: 79.50   1st Qu.:77.00   1st Qu.:81.50
##  Median :89.00   Median : 85.00   Median :84.00   Median :86.00
##  Mean   :85.28   Mean   : 84.65   Mean   :81.67   Mean   :83.94
##  3rd Qu.:94.00   3rd Qu.: 90.50   3rd Qu.:88.00   3rd Qu.:89.00
##  Max.   :99.00   Max.   :105.00   Max.   :92.00   Max.   :95.00
##      X2015
##  Min.   :56.0
##  1st Qu.:77.0
##  Median :85.0
##  Mean   :83.3
##  3rd Qu.:90.0
##  Max.   :97.0
```

```
Atlanta_temps1 <- data.frame(Atlanta_temps)
```

```
#obtain mean temperatures for each year up until 2015 using colMeans and saving the result to a
#data frame
Atlanta_temp_means <- data.frame(colMeans(Atlanta_temps1[, 2:21]))
Atlanta_temp_means$mean <- Atlanta_temp_means$colMeans.Atlanta_temps1...2.21..
Atlanta_temp_means
```

```
##       colMeans.Atlanta_temps1...2.21..     mean
## X1996                         83.71545 83.71545
## X1997                         81.67480 81.67480
## X1998                         84.26016 84.26016
## X1999                         83.35772 83.35772
## X2000                         84.03252 84.03252
## X2001                         81.55285 81.55285
## X2002                         83.58537 83.58537
## X2003                         81.47967 81.47967
```

```
## X2004                          81.76423 81.76423
## X2005                          83.35772 83.35772
## X2006                          83.04878 83.04878
## X2007                          85.39837 85.39837
## X2008                          82.51220 82.51220
## X2009                          80.99187 80.99187
## X2010                          87.21138 87.21138
## X2011                          85.27642 85.27642
## X2012                          84.65041 84.65041
## X2013                          81.66667 81.66667
## X2014                          83.94309 83.94309
## X2015                          83.30081 83.30081
```

```
sd(Atlanta_temps1$X1996)
```

```
## [1] 8.548339
```

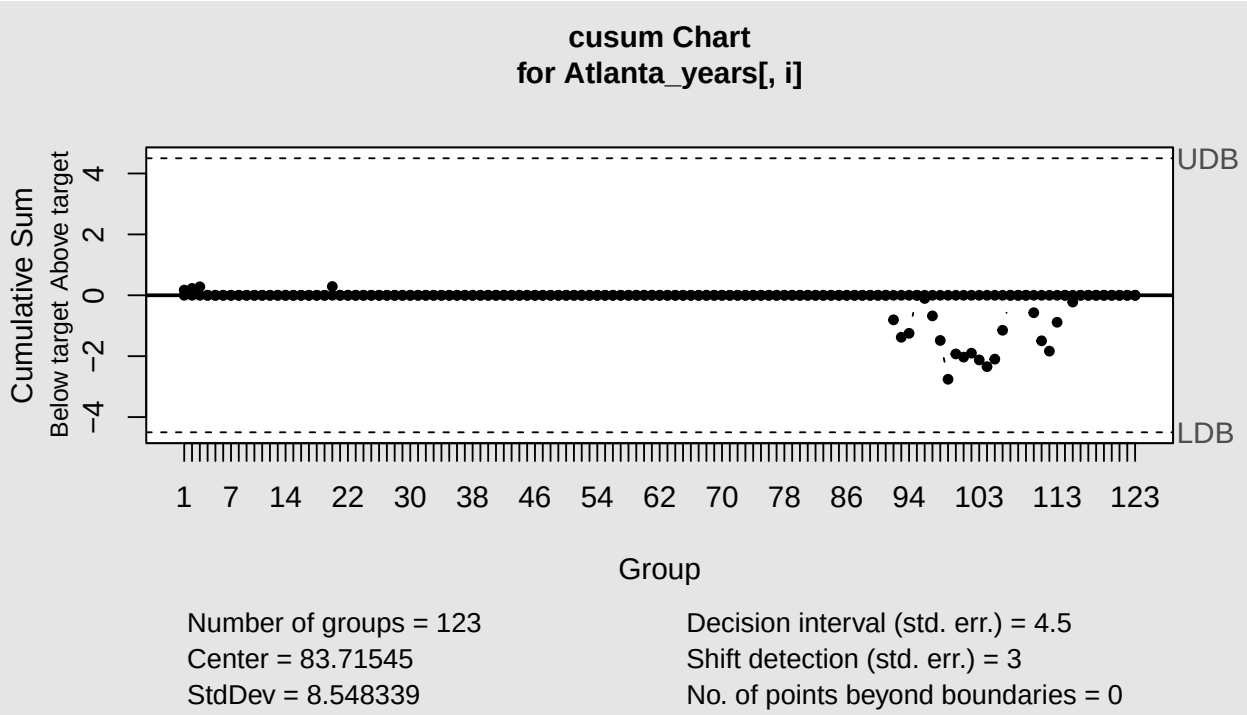
```
#used sapply to get standard deviation of each year in the data set and saved the result to a
#data frame
Atlanta_temp_sd <- data.frame(sapply(Atlanta_temps1[, 2:21], sd))
Atlanta_temp_sd$standarddeviation <- Atlanta_temp_sd$sapply.Atlanta_temps1...2.21...sd.
Atlanta_temp_sd
```

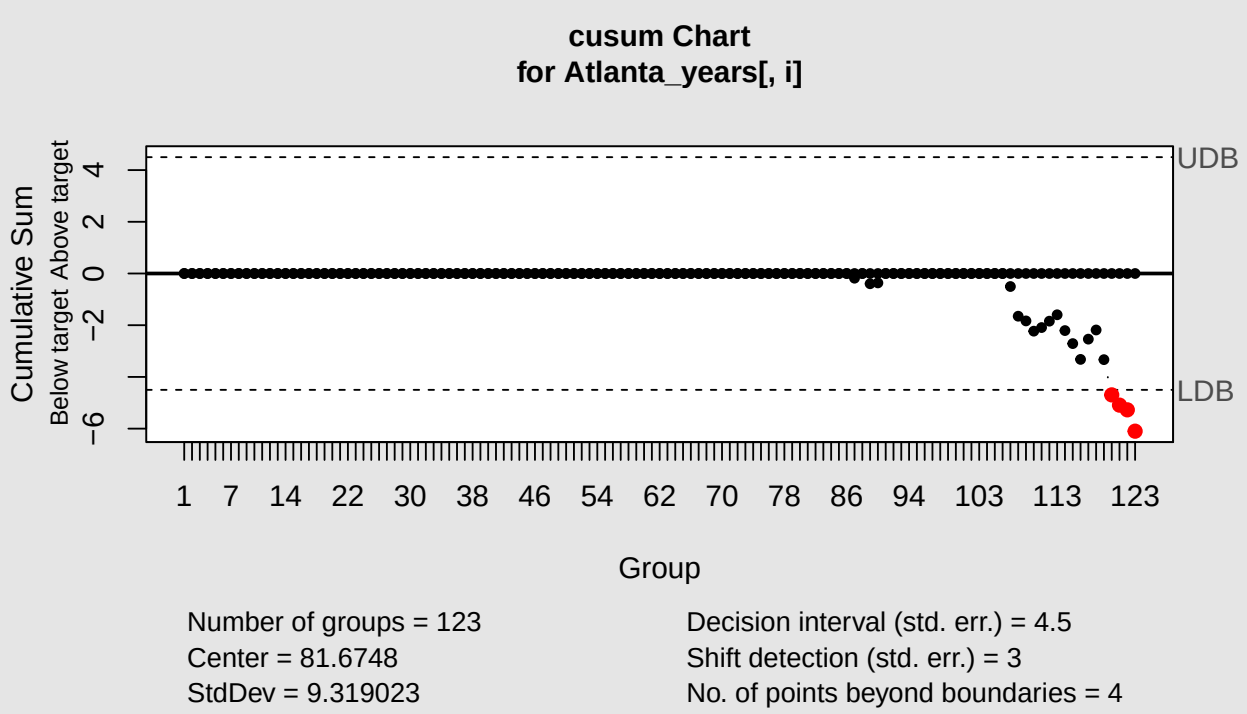
```
##       sapply.Atlanta_temps1...2.21...sd. standarddeviation
## X1996                           8.548339          8.548339
## X1997                           9.319023          9.319023
## X1998                           6.409314          6.409314
## X1999                           9.723328          9.723328
## X2000                           9.518692          9.518692
## X2001                           8.224517          8.224517
## X2002                           9.426095          9.426095
## X2003                           7.017951          7.017951
## X2004                           6.662940          6.662940
## X2005                           7.733396          7.733396
## X2006                           9.793653          9.793653
## X2007                           9.033399          9.033399
## X2008                           8.733172          8.733172
## X2009                           9.013192          9.013192
## X2010                           7.445157          7.445157
## X2011                           9.931157          9.931157
## X2012                           9.252367          9.252367
## X2013                           7.726542          7.726542
## X2014                           6.591476          6.591476
## X2015                           8.709271          8.709271
```

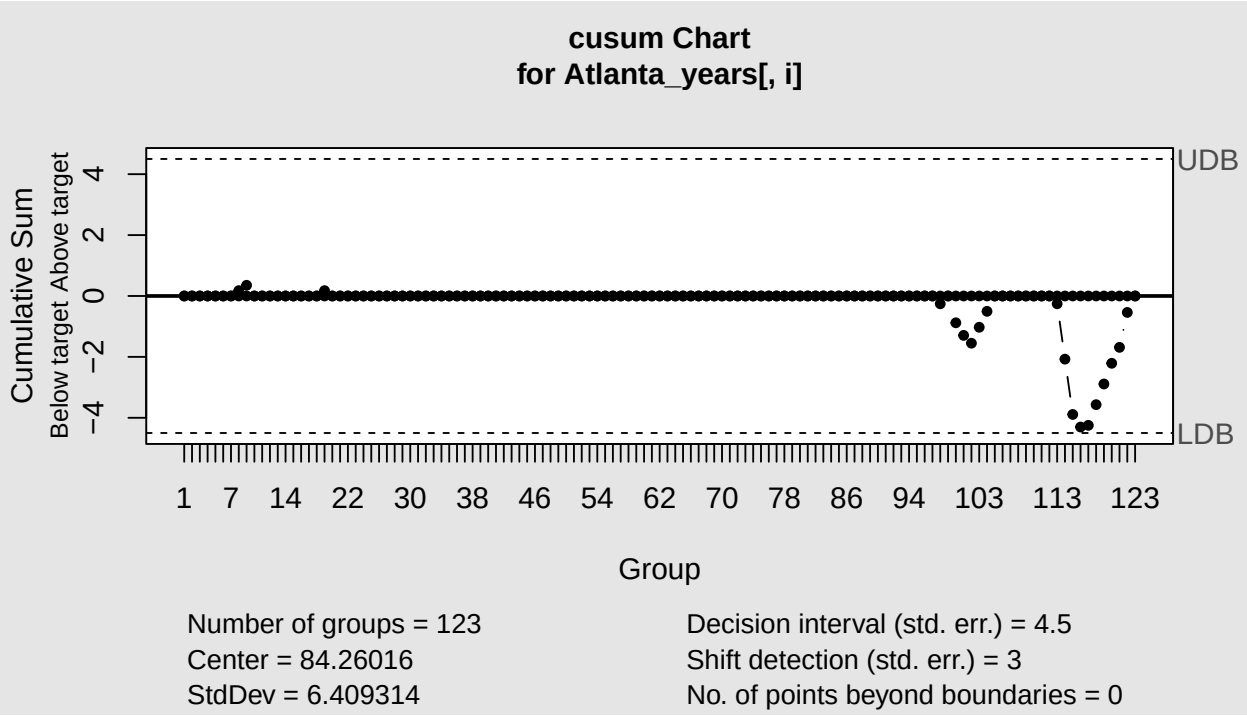
Since it seems to be most common to set the threshold at 4 or 5 standard deviations from the mean, I ran CUSUM with a threshold of 4.5 directly between those two values. The shift value was set at 3. The CUSUM model below was created to detect temperatures that fell below the average for each year. Years that had low temperatures are identified further down.

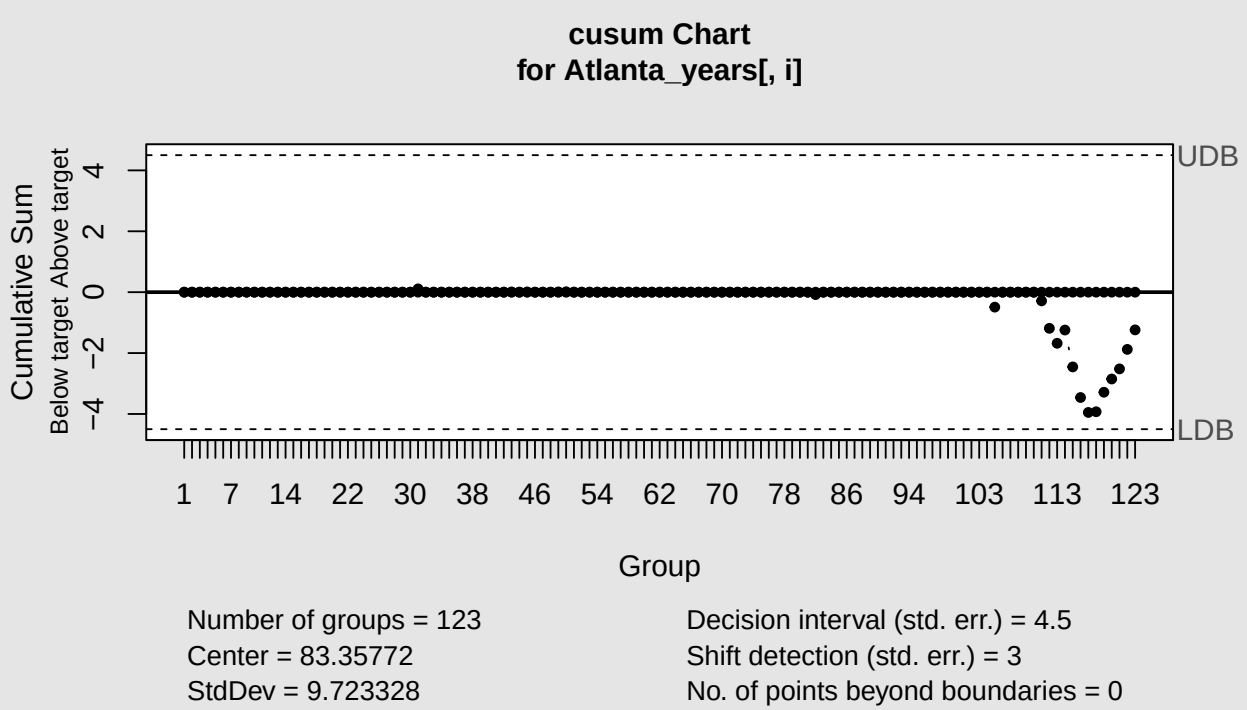
```
Atlanta_years <- Atlanta_temps1[, 2:21]
#create two vectors, one for years and another to store low temperature days that fall below a
```

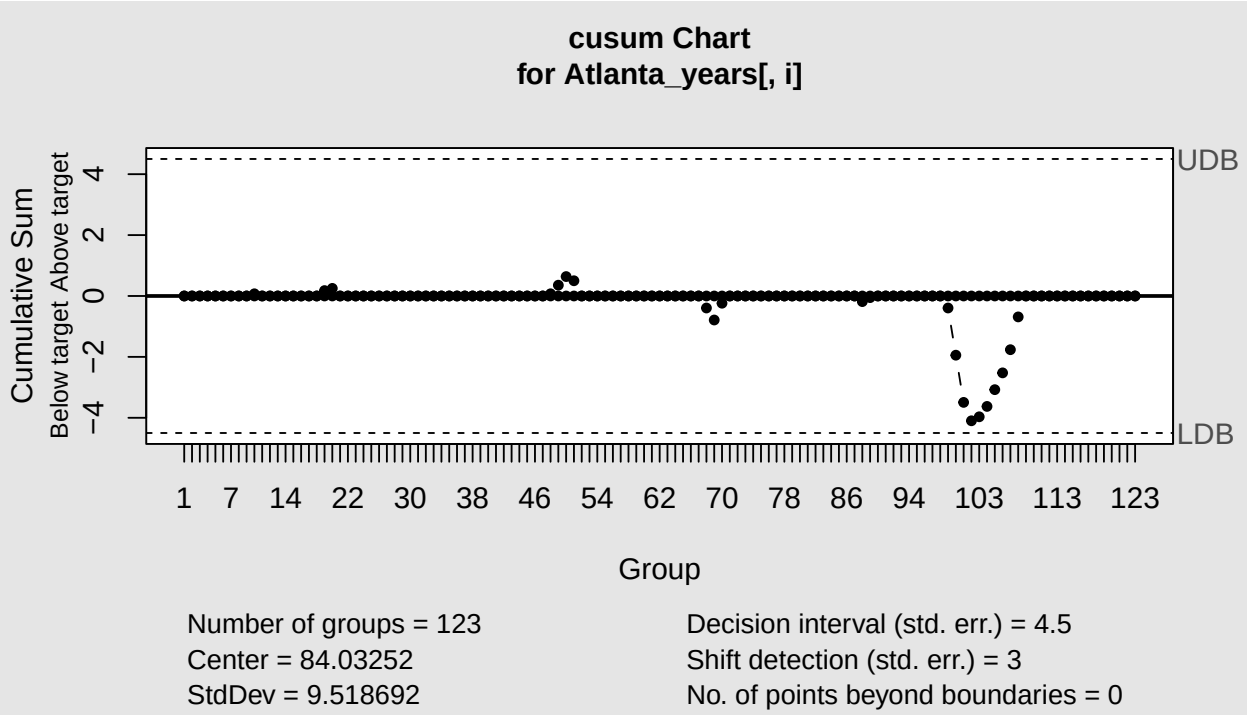
```r
#threshold indicated in the cusum function
CUMSUM_years <- vector("list", ncol(Atlanta_years))
CUMSUMlow_temps <- vector("list", ncol(Atlanta_years))

#looped over each year in order for a cusum to be run for each year in the data set
for (i in 1:ncol(Atlanta_years)) {
  #ran a cusum with a threshold of 4.5, right between the standard thresholds of 4 and 5 which
  #seem to be common, and a shift of 3
 CUMSUM_years[[i]] <- cusum(Atlanta_years[, i], center = Atlanta_temp_means$mean[i], std.dev = Atlanta_
decision.interval = 4.5, se.shift = 3, plot = TRUE)
#obtain low temperatures that fell below the threshold of 4.5
CUMSUMlow_temps[[i]] <- CUMSUM_years[[i]]$violations
}
```

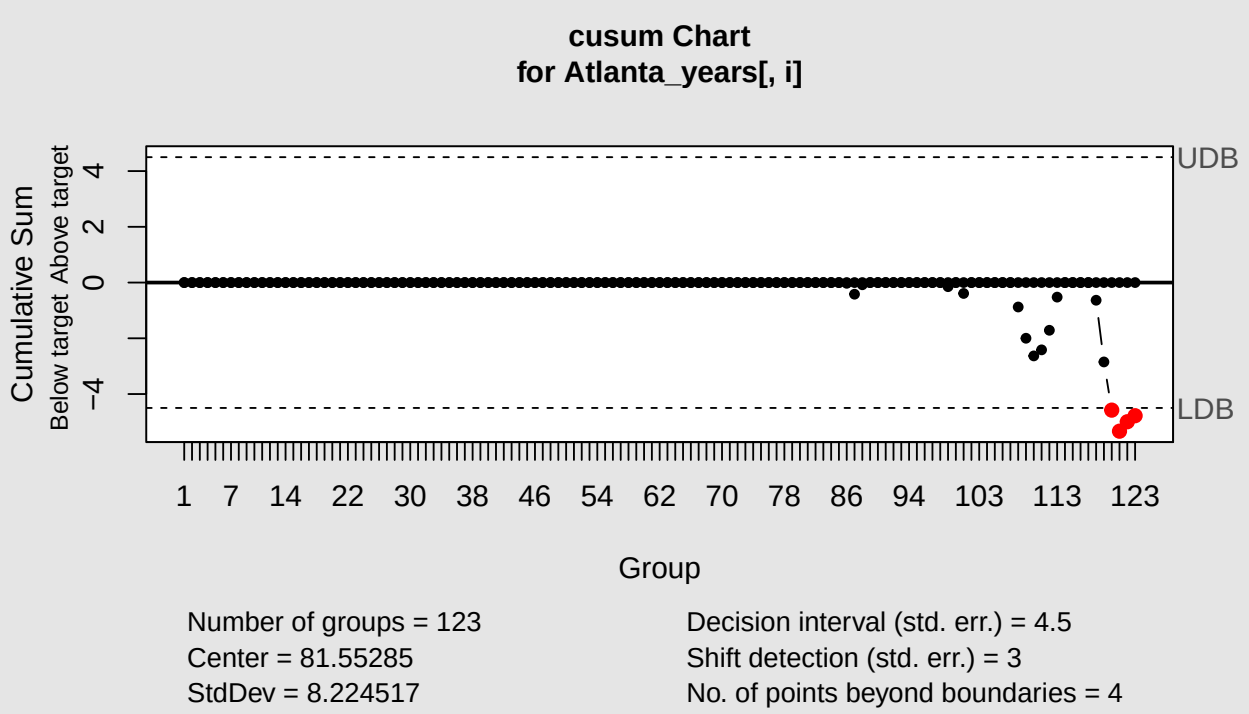


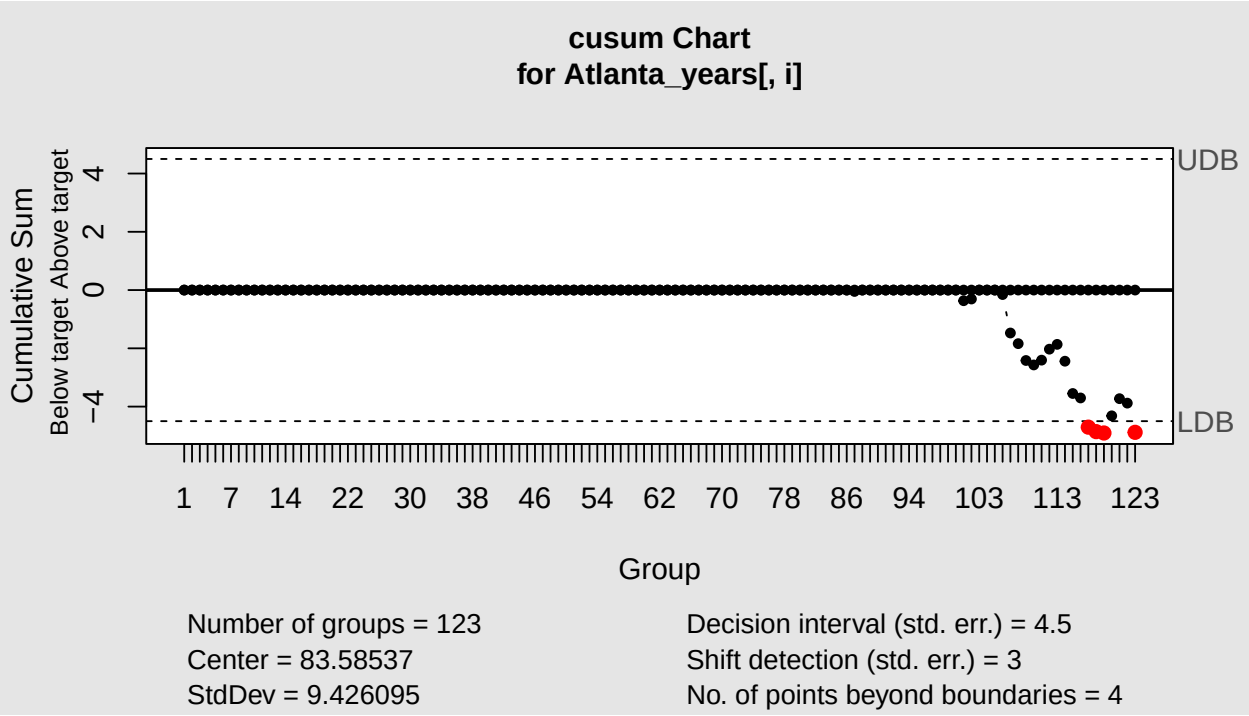

cusum Chart
for Atlanta_years[, i]
Cumulative Sum
Below target Above target
4
2
0
−2
−6
UDB
LDB
1
7
14
22
30
38
46
54
62
70
78
86
94
103
113
123
Group
Number of groups = 123
Center = 81.6748
StdDev = 9.319023
Decision interval (std. err.) = 4.5
Shift detection (std. err.) = 3
No. of points beyond boundaries = 4

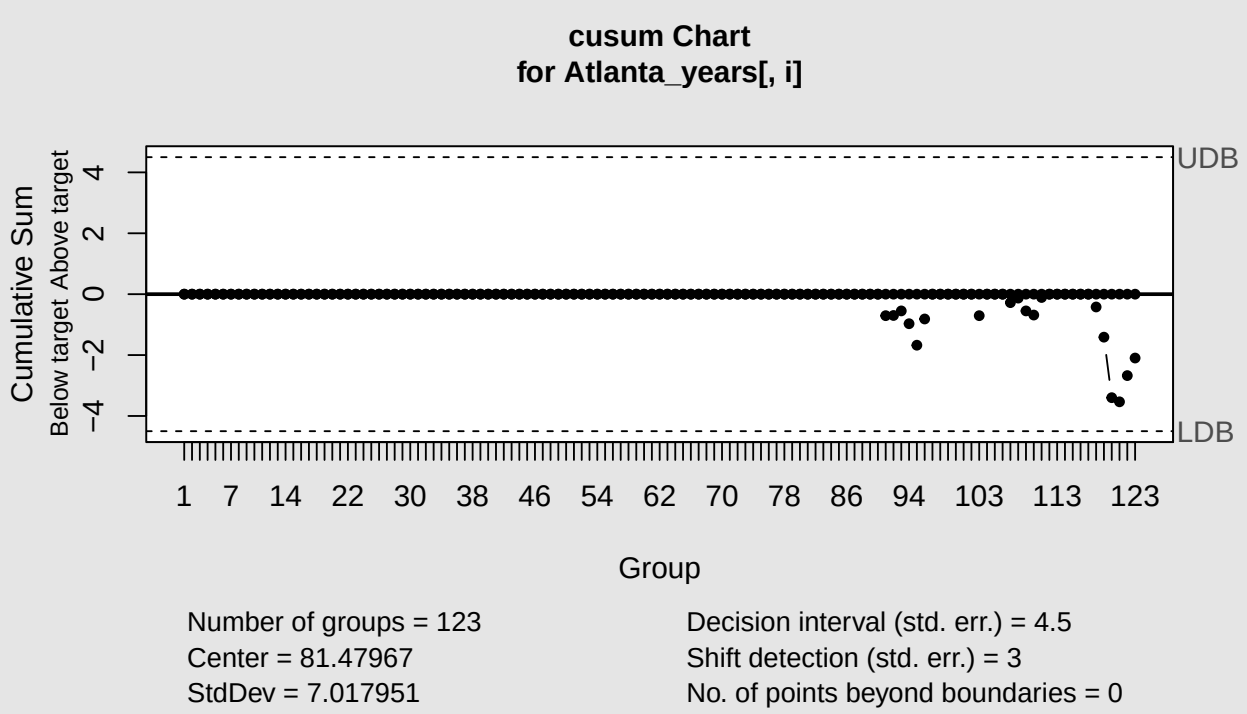

cusum Chart
for Atlanta_years[, i]
Cumulative Sum
Below target Above target
4
2
0
−2
−4
UDB
LDB
1
7
14
22
30
38
46
54
62
70
78
86
94
103
113
123
Group
Number of groups = 123
Center = 84.26016
StdDev = 6.409314
Decision interval (std. err.) = 4.5
Shift detection (std. err.) = 3
No. of points beyond boundaries = 0

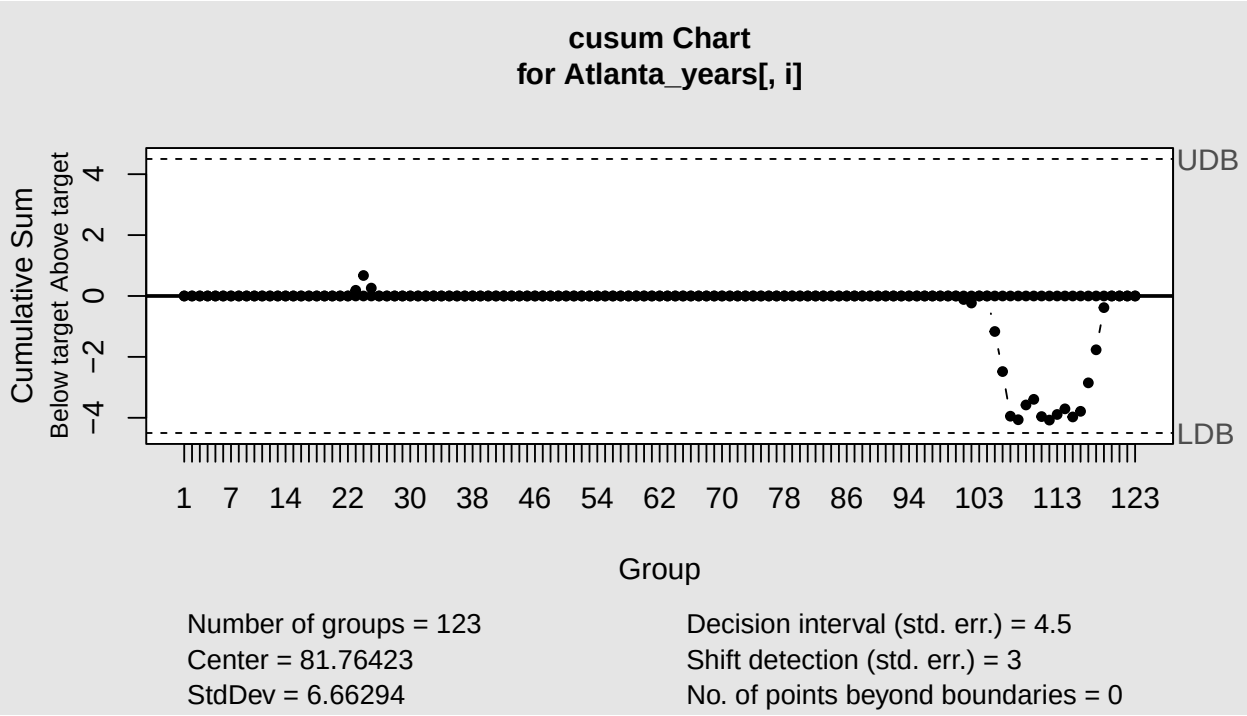

cusum Chart
for Atlanta_years[, i]
Cumulative Sum
Below target Above target
4
2
0
−2
−4
UDB
LDB
1 7 14 22 30 38 46 54 62 70 78 86 94 103 113 123
Group
Number of groups = 123
Center = 83.35772
StdDev = 9.723328
Decision interval (std. err.) = 4.5
Shift detection (std. err.) = 3
No. of points beyond boundaries = 0

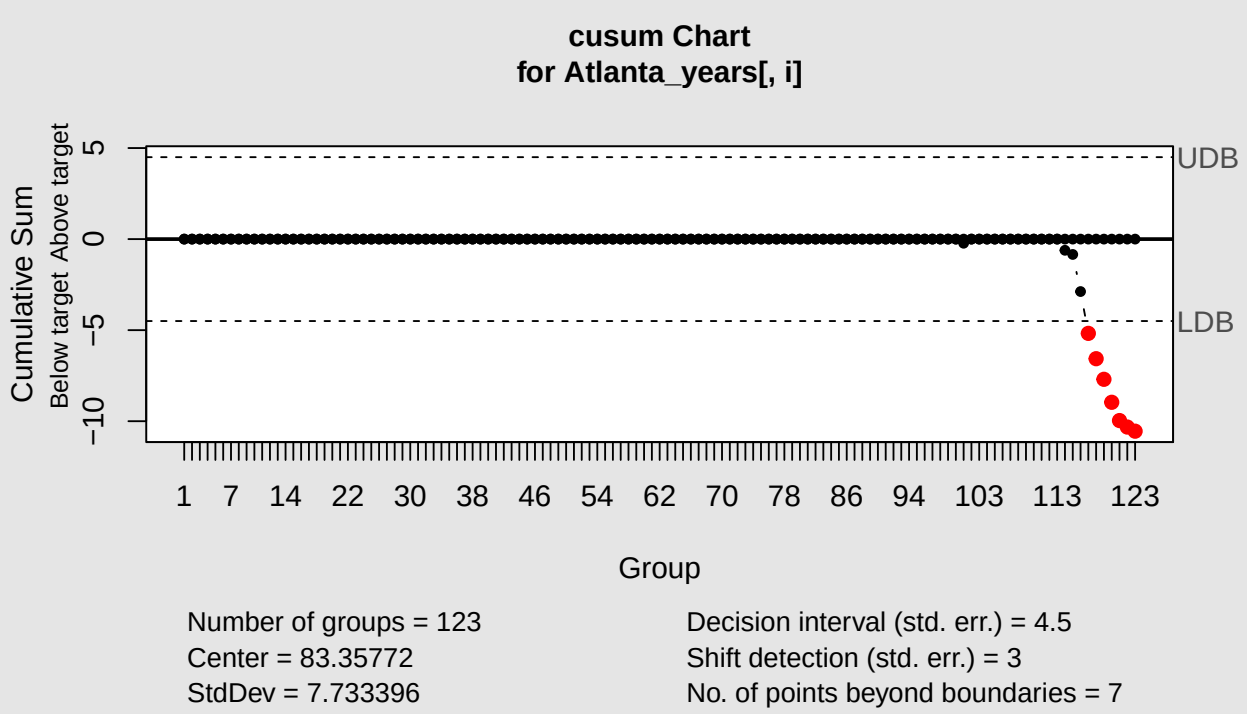

cusum Chart
for Atlanta_years[, i]
Cumulative Sum
Below target Above target
4
2
0
−2
−4
UDB
LDB
1 7 14 22 30 38 46 54 62 70 78 86 94 103 113 123
Group
Number of groups = 123
Center = 84.03252
StdDev = 9.518692
Decision interval (std. err.) = 4.5
Shift detection (std. err.) = 3
No. of points beyond boundaries = 0

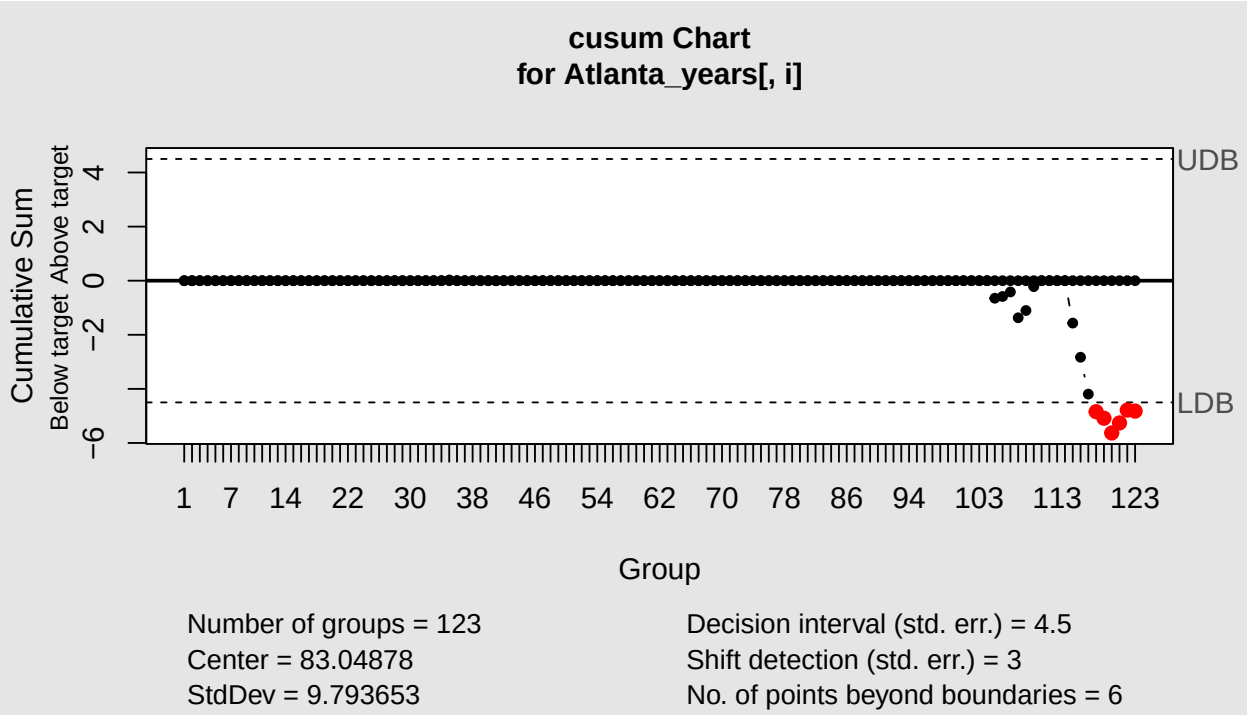

cusum Chart
for Atlanta_years[, i]
Cumulative Sum
Below target Above target
4
2
0
−4
UDB
LDB
1 7 14 22 30 38 46 54 62 70 78 86 94 103 113 123
Group
Number of groups = 123
Center = 81.55285
StdDev = 8.224517
Decision interval (std. err.) = 4.5
Shift detection (std. err.) = 3
No. of points beyond boundaries = 4

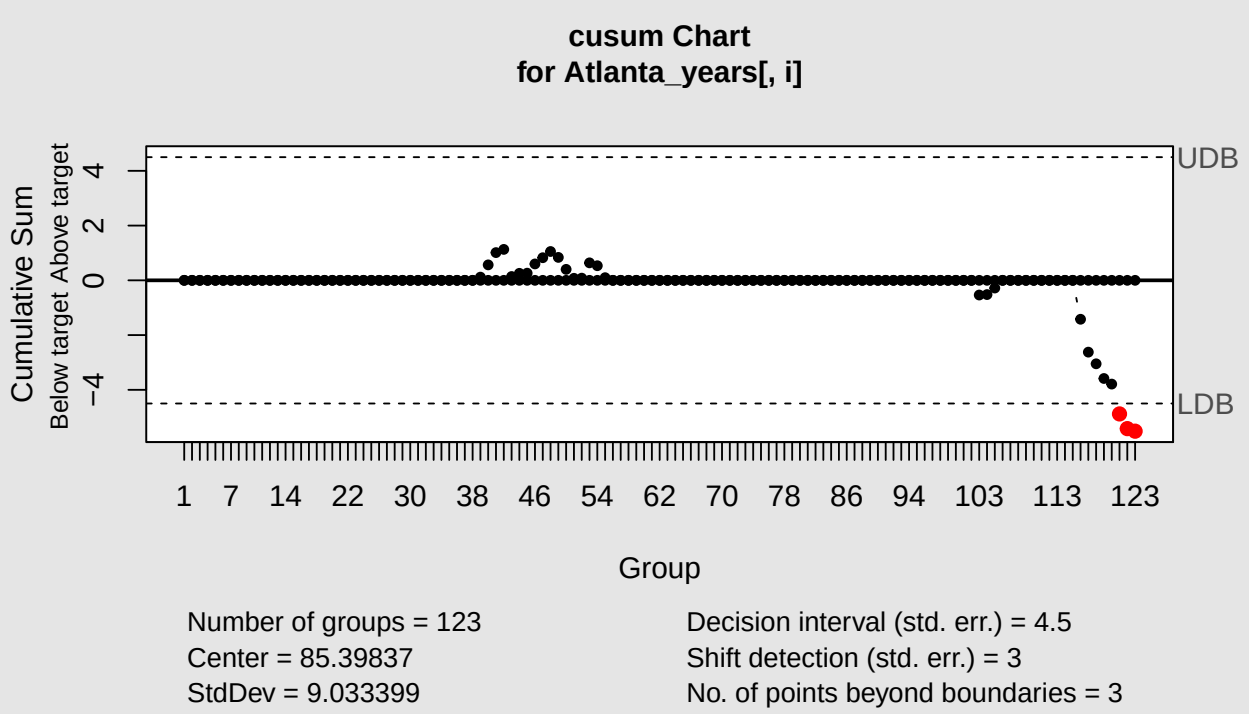

cusum Chart
for Atlanta_years[, i]
Cumulative Sum
Below target Above target
4
2
0
−4
UDB
LDB
1 7 14 22 30 38 46 54 62 70 78 86 94 103 113 123
Group
Number of groups = 123
Center = 83.58537
StdDev = 9.426095
Decision interval (std. err.) = 4.5
Shift detection (std. err.) = 3
No. of points beyond boundaries = 4

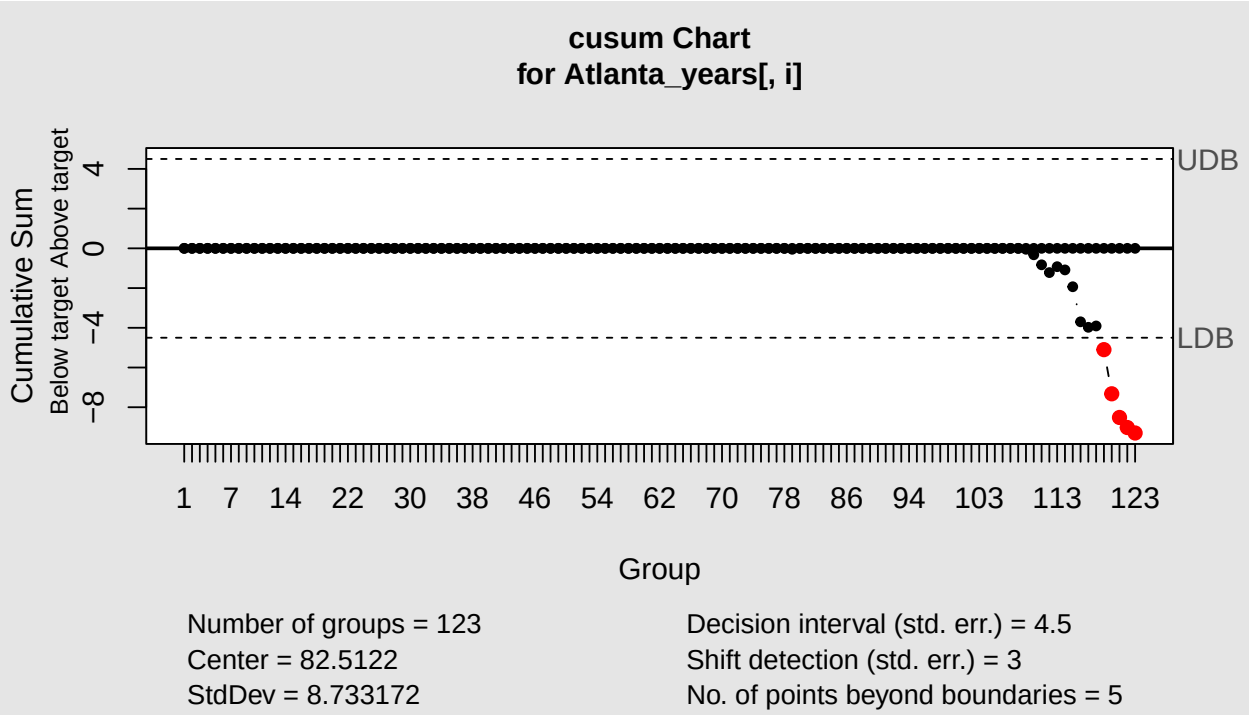

cusum Chart
for Atlanta_years[, i]
Cumulative Sum
Below target Above target
4
2
0
−2
−4
UDB
LDB
1 7 14 22 30 38 46 54 62 70 78 86 94 103 113 123
Group
Number of groups = 123
Center = 81.47967
StdDev = 7.017951
Decision interval (std. err.) = 4.5
Shift detection (std. err.) = 3
No. of points beyond boundaries = 0

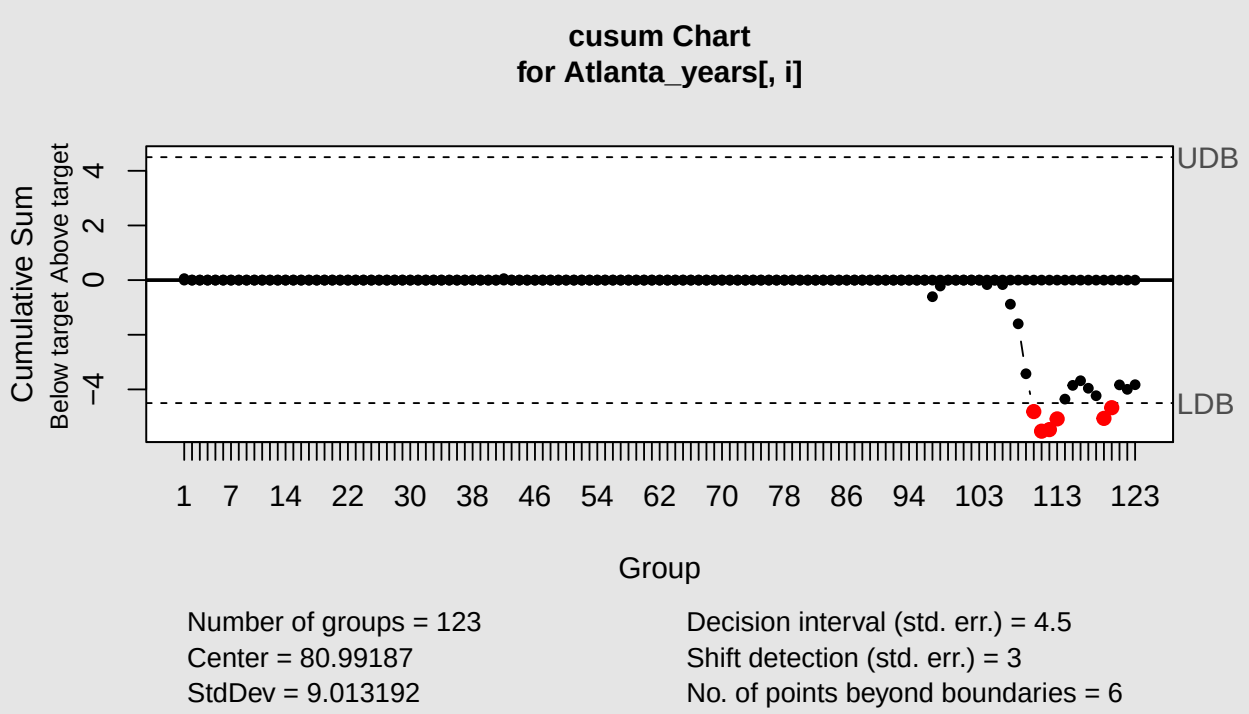

cusum Chart
for Atlanta_years[, i]
Cumulative Sum
Below target Above target
4
2
0
−2
−4
UDB
LDB
1 7 14 22 30 38 46 54 62 70 78 86 94 103 113 123
Group
Number of groups = 123
Center = 81.76423
StdDev = 6.66294
Decision interval (std. err.) = 4.5
Shift detection (std. err.) = 3
No. of points beyond boundaries = 0

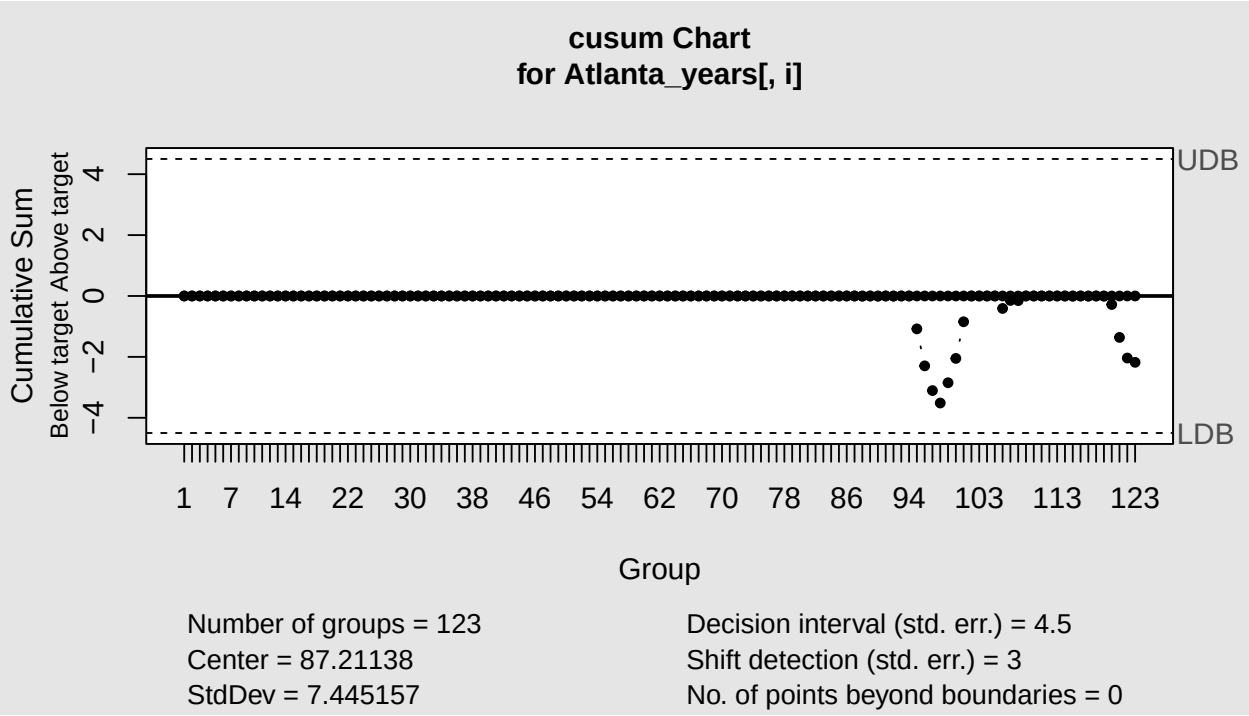

cusum Chart
for Atlanta_years[, i]
Cumulative Sum
Below target Above target
5
0
−5
−10
UDB
LDB
1 7 14 22 30 38 46 54 62 70 78 86 94 103 113 123
Group
Number of groups = 123
Center = 83.35772
StdDev = 7.733396
Decision interval (std. err.) = 4.5
Shift detection (std. err.) = 3
No. of points beyond boundaries = 7

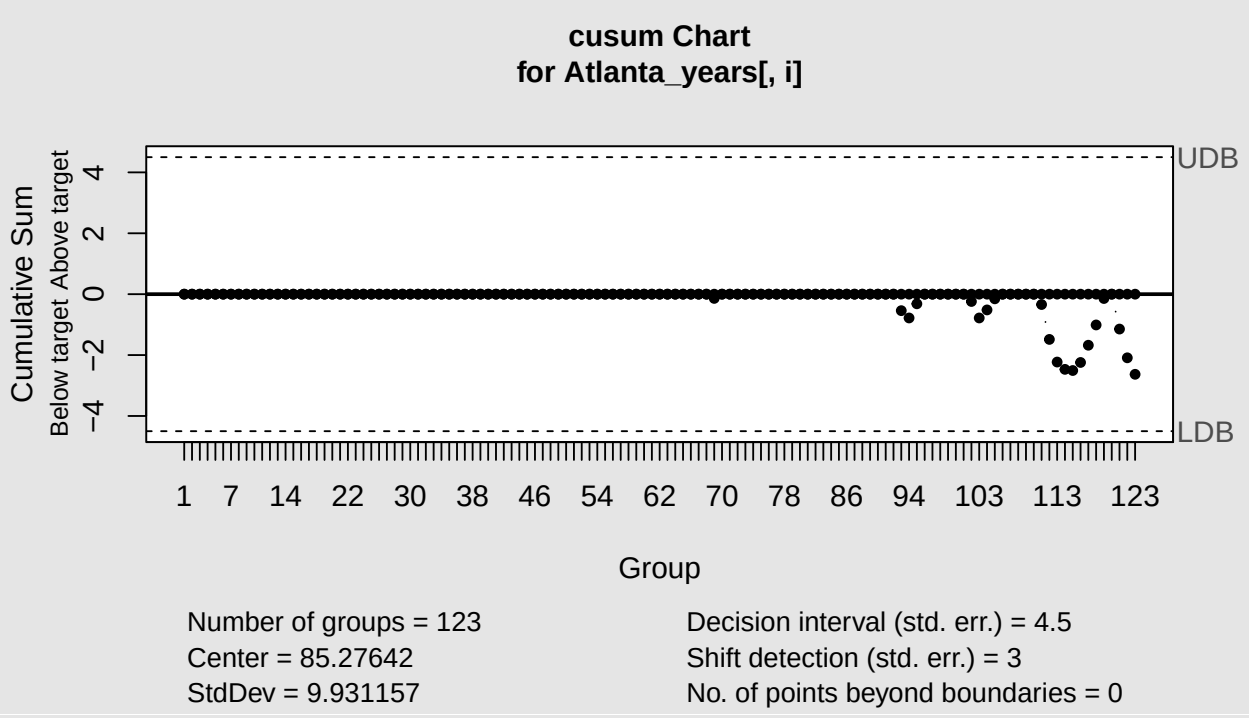

cusum Chart
for Atlanta_years[, i]
Cumulative Sum
Below target Above target
4
2
0
−2
−6
UDB
LDB
1 7 14 22 30 38 46 54 62 70 78 86 94 103 113 123
Group
Number of groups = 123
Center = 83.04878
StdDev = 9.793653
Decision interval (std. err.) = 4.5
Shift detection (std. err.) = 3
No. of points beyond boundaries = 6

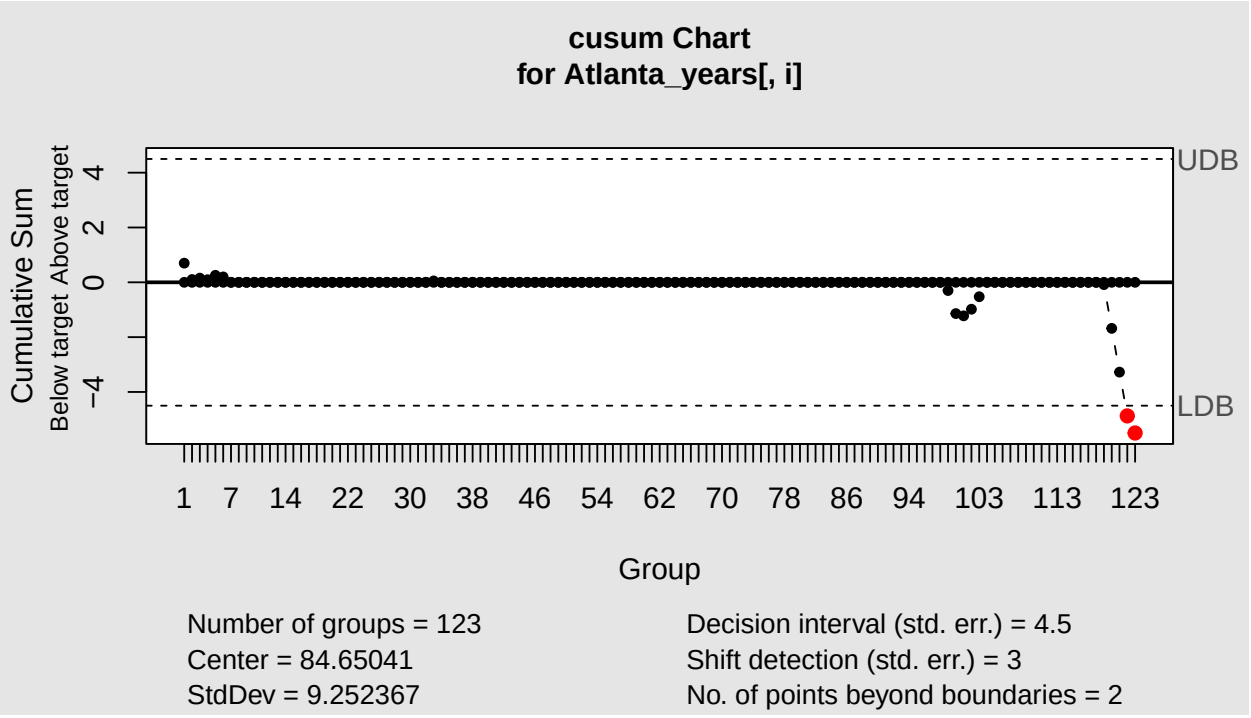

cusum Chart
for Atlanta_years[, i]
Cumulative Sum
Below target Above target
4
2
0
−4
UDB
LDB
1
7
14
22
30
38
46
54
62
70
78
86
94
103
113
123
Group
Number of groups = 123
Center = 85.39837
StdDev = 9.033399
Decision interval (std. err.) = 4.5
Shift detection (std. err.) = 3
No. of points beyond boundaries = 3

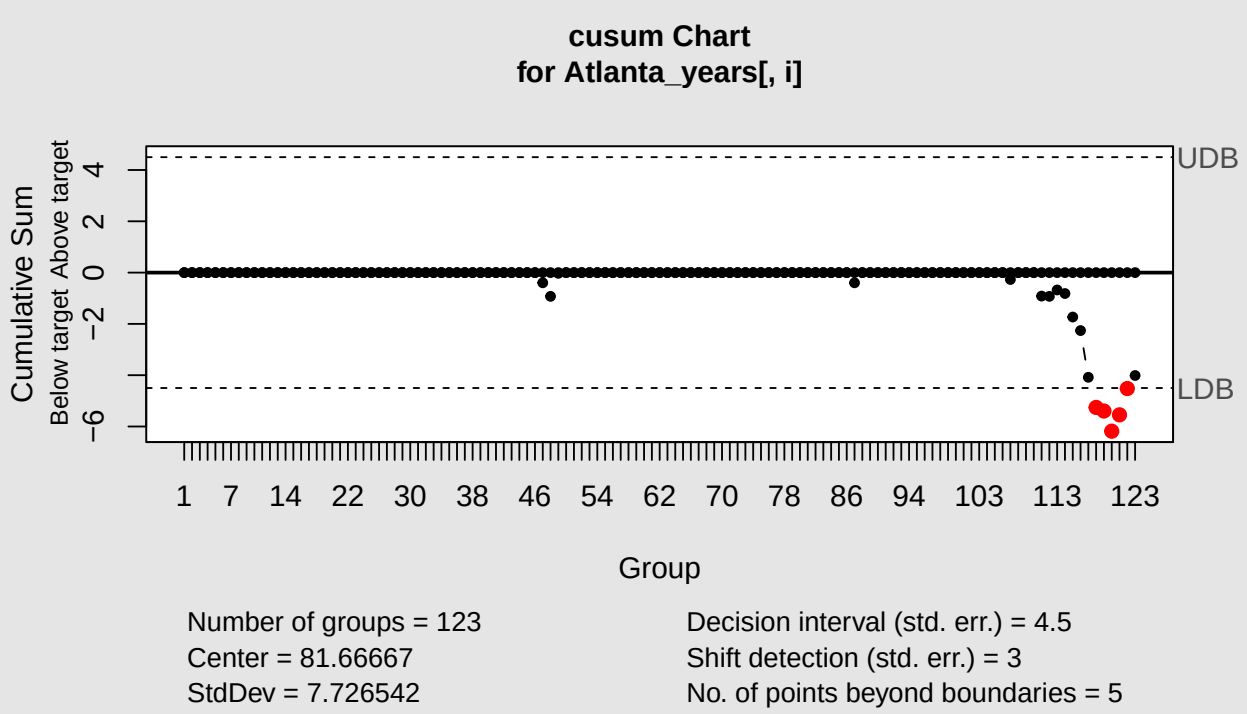

cusum Chart
for Atlanta_years[, i]
Cumulative Sum
Below target Above target
4
0
−4
−8
UDB
LDB
1
7
14
22
30
38
46
54
62
70
78
86
94
103
113
123
Group
Number of groups = 123
Center = 82.5122
StdDev = 8.733172
Decision interval (std. err.) = 4.5
Shift detection (std. err.) = 3
No. of points beyond boundaries = 5

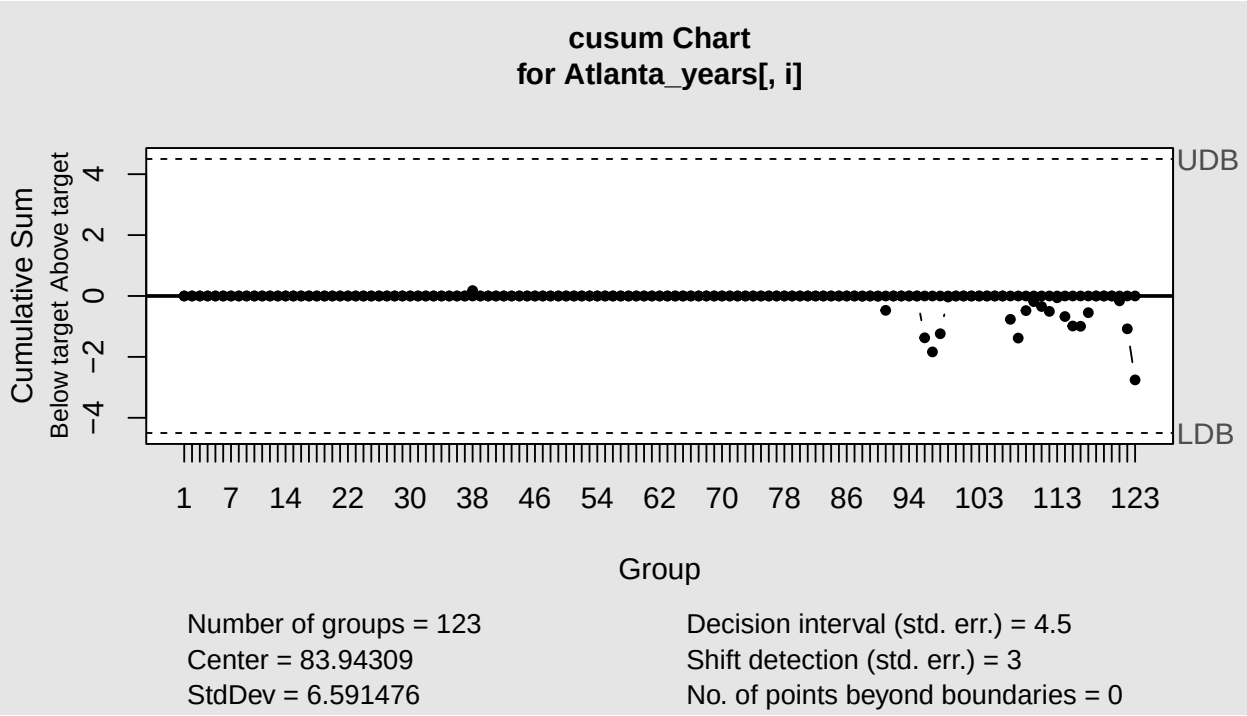

cusum Chart
for Atlanta_years[, i]
Cumulative Sum
Below target Above target
4
2
0
−4
UDB
LDB
1 7 14 22 30 38 46 54 62 70 78 86 94 103 113 123
Group
Number of groups = 123
Center = 80.99187
StdDev = 9.013192
Decision interval (std. err.) = 4.5
Shift detection (std. err.) = 3
No. of points beyond boundaries = 6

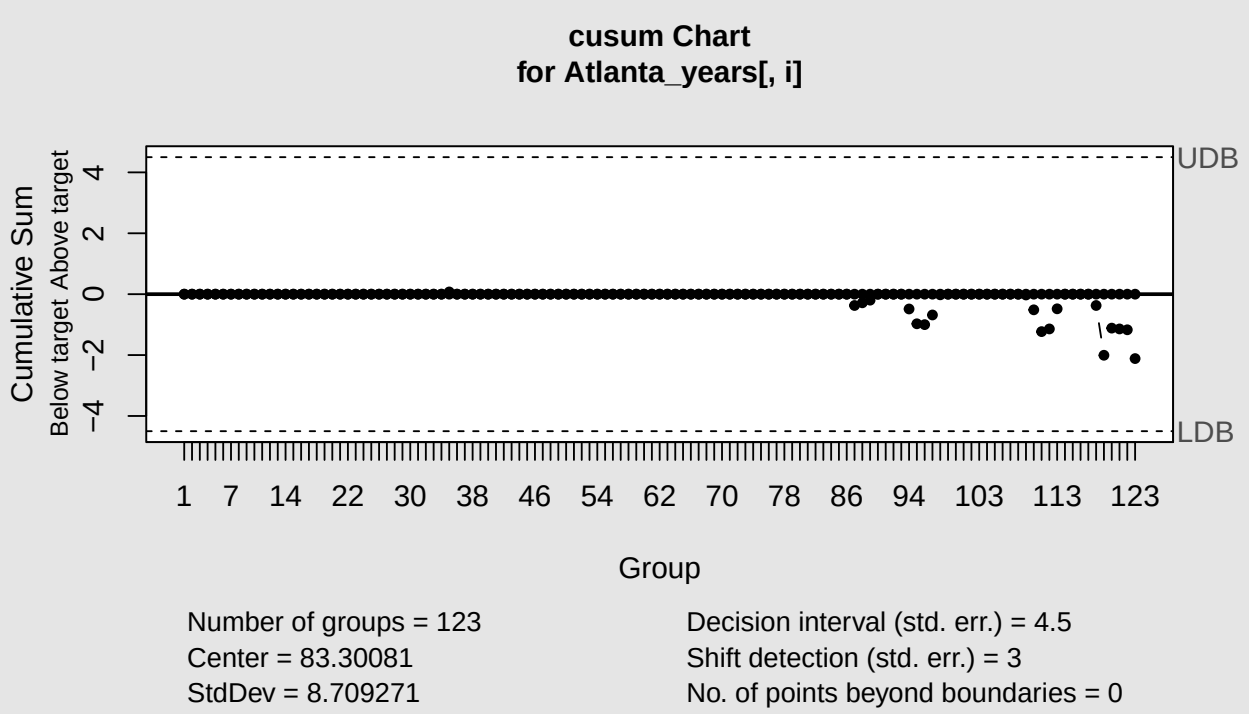

cusum Chart
for Atlanta_years[, i]
Cumulative Sum
Below target Above target
4
2
0
−2
−4
UDB
LDB
1 7 14 22 30 38 46 54 62 70 78 86 94 103 113 123
Group
Number of groups = 123
Center = 87.21138
StdDev = 7.445157
Decision interval (std. err.) = 4.5
Shift detection (std. err.) = 3
No. of points beyond boundaries = 0


cusum Chart
for Atlanta_years[, i]
Cumulative Sum
Below target Above target
4
2
0
−2
−4
UDB
LDB
1
7
14
22
30
38
46
54
62
70
78
86
94
103
113
123
Group
Number of groups = 123
Center = 85.27642
StdDev = 9.931157
Decision interval (std. err.) = 4.5
Shift detection (std. err.) = 3
No. of points beyond boundaries = 0


cusum Chart
for Atlanta_years[, i]
Cumulative Sum
Below target Above target
4
2
0
−4
UDB
LDB
1
7
14
22
30
38
46
54
62
70
78
86
94
103
113
123
Group
Number of groups = 123
Center = 84.65041
StdDev = 9.252367
Decision interval (std. err.) = 4.5
Shift detection (std. err.) = 3
No. of points beyond boundaries = 2


cusum Chart
for Atlanta_years[, i]
Cumulative Sum
Below target Above target
4
2
0
−2
−6
UDB
LDB
1 7 14 22 30 38 46 54 62 70 78 86 94 103 113 123
Group
Number of groups = 123
Center = 81.66667
StdDev = 7.726542
Decision interval (std. err.) = 4.5
Shift detection (std. err.) = 3
No. of points beyond boundaries = 5


cusum Chart
for Atlanta_years[, i]
Cumulative Sum
Below target Above target
4
2
0
−2
−4
UDB
LDB
1 7 14 22 30 38 46 54 62 70 78 86 94 103 113 123
Group
Number of groups = 123
Center = 83.94309
StdDev = 6.591476
Decision interval (std. err.) = 4.5
Shift detection (std. err.) = 3
No. of points beyond boundaries = 0



From the CUSUM model above, the years 1997, 2001, 2002, 2005, 2006, 2007, 2008, 2009, 2012, and 2013 had days where the temperature fell far below the threshold.

```
CUMSUMlow_temps
```

```
## [[1]]
## [[1]]$lower
## integer(0)
##
## [[1]]$upper
## integer(0)
##
##
## [[2]]
## [[2]]$lower
## [1] 120 121 122 123
##
## [[2]]$upper
## integer(0)
##
##
## [[3]]
## [[3]]$lower
## integer(0)
##
## [[3]]$upper
## integer(0)
##
##
## [[4]]
## [[4]]$lower
```

```
## integer(0)
##
## [[4]]$upper
## integer(0)
##
##
## [[5]]
## [[5]]$lower
## integer(0)
##
## [[5]]$upper
## integer(0)
##
##
## [[6]]
## [[6]]$lower
## [1] 120 121 122 123
##
## [[6]]$upper
## integer(0)
##
##
## [[7]]
## [[7]]$lower
## [1] 117 118 119 123
##
## [[7]]$upper
## integer(0)
##
##
## [[8]]
## [[8]]$lower
## integer(0)
##
## [[8]]$upper
## integer(0)
##
##
## [[9]]
## [[9]]$lower
## integer(0)
##
## [[9]]$upper
## integer(0)
##
##
## [[10]]
## [[10]]$lower
## [1] 117 118 119 120 121 122 123
##
## [[10]]$upper
## integer(0)
##
##
```

```
## [[11]]
## [[11]]$lower
## [1] 118 119 120 121 122 123
##
## [[11]]$upper
## integer(0)
##
##
## [[12]]
## [[12]]$lower
## [1] 121 122 123
##
## [[12]]$upper
## integer(0)
##
##
## [[13]]
## [[13]]$lower
## [1] 119 120 121 122 123
##
## [[13]]$upper
## integer(0)
##
##
## [[14]]
## [[14]]$lower
## [1] 110 111 112 113 119 120
##
## [[14]]$upper
## integer(0)
##
##
## [[15]]
## [[15]]$lower
## integer(0)
##
## [[15]]$upper
## integer(0)
##
##
## [[16]]
## [[16]]$lower
## integer(0)
##
## [[16]]$upper
## integer(0)
##
##
## [[17]]
## [[17]]$lower
## [1] 122 123
##
## [[17]]$upper
## integer(0)
```

```
##
##
## [[18]]
## [[18]]$lower
## [1] 118 119 120 121 122
##
## [[18]]$upper
## integer(0)
##
##
## [[19]]
## [[19]]$lower
## integer(0)
##
## [[19]]$upper
## integer(0)
##
##
## [[20]]
## [[20]]$lower
## integer(0)
##
## [[20]]$upper
## integer(0)
```

From the years that had low temperature days, the indexes corresponding to them were outputted above. Those days had indexes of 110 (Oct. 18), 111 (Oct. 19), 112 (Oct. 20), 113 (Oct. 21), 118 (Oct. 26), 119 (Oct. 27), 120 (Oct. 28), 121 (Oct. 29), 122 (Oct. 30), and 123 (Oct. 31). These days fall within the last two weeks of October from October 18th-31st.

The low temperature days between October 18th-21st only occurred during one year, 2009, and the ones between October 26th-31st occurred much more frequently. As a result, summer in Atlanta ends during the last week of October, between October 26th and October 31st.

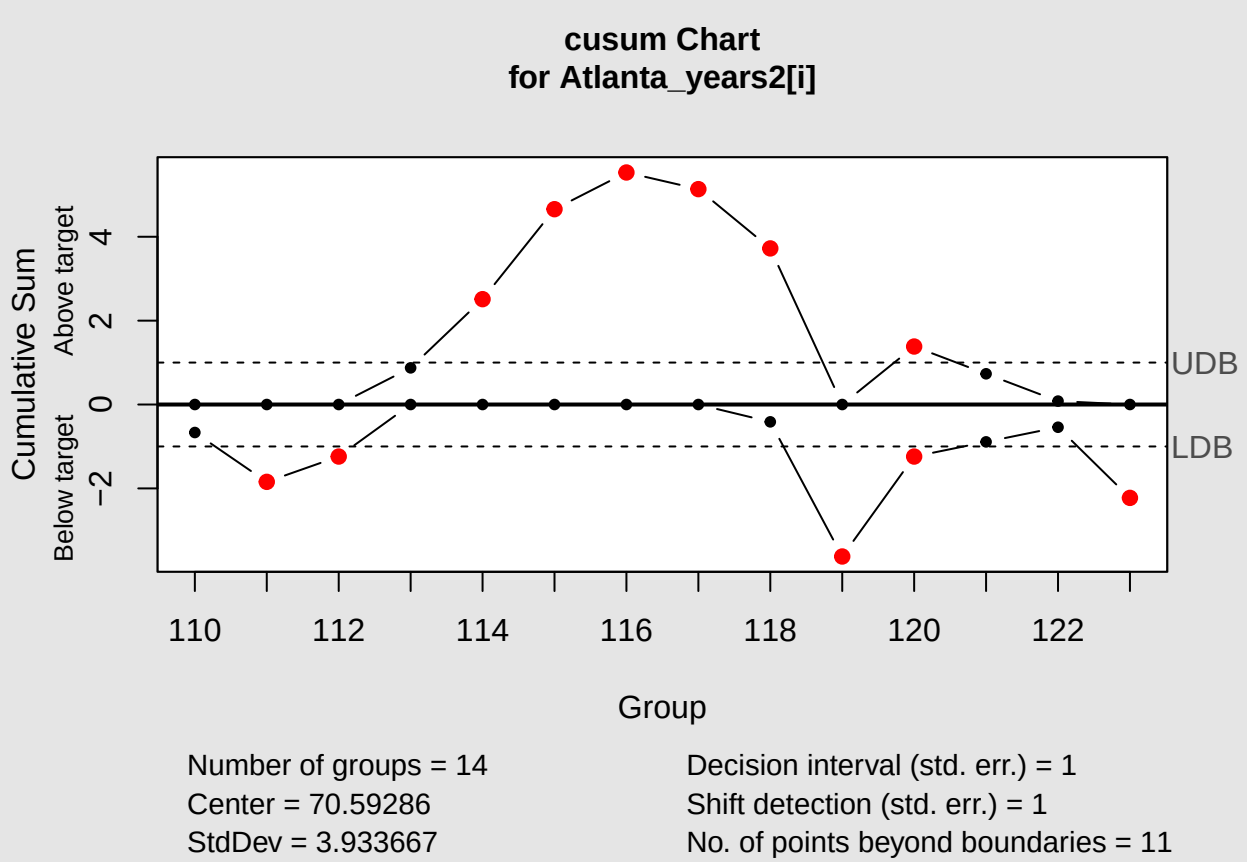
b. Use a CUSUM approach to make a judgment of whether Atlanta's summer climate has gotten #warmer in that time (and if so, when).

The last two weeks of October were used from each of the 20 years in the data set to determine whether Atlanta's summer climate has gotten warmer. To do this, the average temperature over the time from October 18th-31st was taken for each year. The average of the 20 means was then taken to serve as the critical point for the CUSUM model. The standard deviation of this two-week period was also taken.

```
#create a new dataset just containing the low temperature days identified in the previous question
Atlanta_temps2 <- Atlanta_temps1[110:123, ]

#calculated the mean of the same 14 day period for each year
Atlanta_temp_means1 <- data.frame(colMeans(Atlanta_temps2[, 2:21]))

Atlanta_temp_means1$mean <- Atlanta_temp_means1$colMeans.Atlanta_temps1...2.21..
```

Since the time period for this situation was only 14 days, compared to 123 days in 6.1.a., the threshold for this CUSUM model was lowered to only one standard deviation. The shift was lowered to 1 as well.

```
Atlanta_years2 <- Atlanta_temps2[, 2:21]

#obtained an overall mean for the same 14 day period from the means of each individual year
#to be used in the next cusum iteration
days_mean <- mean(Atlanta_temp_means1$colMeans.Atlanta_temps2...2.21..)
#obtained the standard deviation from the mean for the 14 day period for each year in the dataset
days_sd <- sd(Atlanta_temp_means1$colMeans.Atlanta_temps2...2.21..)

#ran cusum with a much tighter threshold of 1 and a shift of 1
CUSUM_model_2 <- cusum(Atlanta_years2[i], center = days_mean, std.dev = days_sd, decision.interval = 1,
```




```
#obtained data points that had temperatures both higher and lower than the threshold above
CUSUM_model_2$violations
```

```
## $lower
## [1]  2  3 10 11 14
##
## $upper
## [1]  5  6  7  8  9 11
```

From the CUSUM output above, we have six points where the temperature got higher than the threshold (5, 6, 7, 8, 9, 11) and five points where the temperature fell below the threshold (5, 6, 7, 8, 9, 11). Since the days fluctuated between stretches of exceeding the threshold and falling below the threshold, it is difficult to tell whether summer's have gotten hotter over time in Atlanta.

When looking at the average temperature of this two-week period each year, there has not been any period of steady increase in the average temperature. Instead, the average temperature has fluctuated repeatedly between the upper 60's and mid-70's over this 20 year period. It would be helpful to see the average temperature for each year since 2015 for the best answer, but based on the available data, it is inconclusive whether the summer temperature's have gotten warmer over time in Atlanta.