ISYE 6501

2022-09-12

Submission HW2 | Fall 22

- Ashish Dhiman | ashish.dhiman@gatech.edu
- Abhinav Arun | aarun60@gatech.edu
- Anshit Verma | averma373@gatech.edu

***Analysis Notes are marked with Red header: #Analysis**

```
#lib imports
library(grid)
library(reshape2)
library(ggplot2)
```

Read Data

```
data = read.table('./temps.txt', sep="\t", dec=".", header = TRUE)
head(data)
```

```
##     DAY X1996 X1997 X1998 X1999 X2000 X2001 X2002 X2003 X2004 X2005 X2006 X2007
## 1 1-Jul    98    86    91    84    89    84    90    73    82    91    93    95
## 2 2-Jul    97    90    88    82    91    87    90    81    81    89    93    85
## 3 3-Jul    97    93    91    87    93    87    87    87    86    86    93    82
## 4 4-Jul    90    91    91    88    95    84    89    86    88    86    91    86
## 5 5-Jul    89    84    91    90    96    86    93    80    90    89    90    88
## 6 6-Jul    93    84    89    91    96    87    93    84    90    82    81    87
##   X2008 X2009 X2010 X2011 X2012 X2013 X2014 X2015
## 1    85    95    87    92   105    82    90    85
## 2    87    90    84    94    93    85    93    87
## 3    91    89    83    95    99    76    87    79
## 4    90    91    85    92    98    77    84    85
## 5    88    80    88    90   100    83    86    84
## 6    82    87    89    90    98    83    87    84
```

Transform Temperature data from Fahrenheit to Celsius

```
### Transform to celsius
temp_cols = names(data)[!(names(data) %in% c("DAY"))]
#Transform to celsius
data_celsius = data
data_celsius[temp_cols] <- lapply(data_celsius[temp_cols], function(f) (f-32)*5/9)
```

EDA of Temperature Data

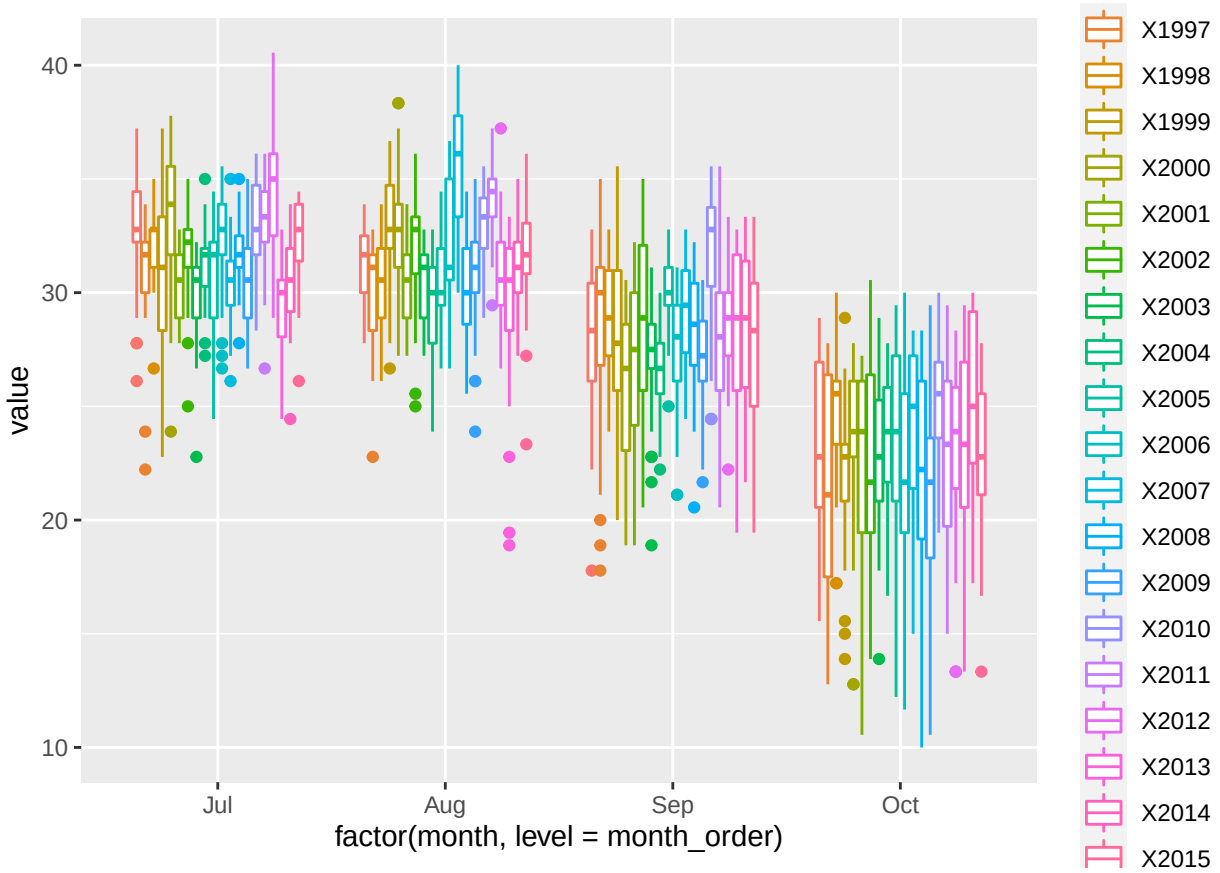
```
summary(data_celsius)
```

```
##      DAY                X1996           X1997           X1998
##  Length:123         Min.   :15.56   Min.   :12.78   Min.   :17.22
##  Class :character   1st Qu.:26.11   1st Qu.:25.83   1st Qu.:26.39
##  Mode  :character   Median :28.89   Median :28.89   Median :30.00
##                     Mean   :28.73   Mean   :27.60   Mean   :29.03
##                     3rd Qu.:32.22   3rd Qu.:31.39   3rd Qu.:31.67
##                     Max.   :37.22   Max.   :35.00   Max.   :35.00
##      X1999           X2000           X2001           X2002
##  Min.   :13.89   Min.   :12.78   Min.   :10.56   Min.   :13.89
##  1st Qu.:23.89   1st Qu.:25.00   1st Qu.:25.56   1st Qu.:25.56
##  Median :30.00   Median :30.00   Median :28.89   Median :30.56
##  Mean   :28.53   Mean   :28.91   Mean   :27.53   Mean   :28.66
##  3rd Qu.:32.78   3rd Qu.:32.78   3rd Qu.:30.56   3rd Qu.:32.78
##  Max.   :37.22   Max.   :38.33   Max.   :33.89   Max.   :36.11
##      X2003           X2004           X2005           X2006
##  Min.   :13.89   Min.   :16.67   Min.   :12.22   Min.   :11.67
##  1st Qu.:25.56   1st Qu.:25.56   1st Qu.:27.50   1st Qu.:26.11
##  Median :28.89   Median :27.78   Median :29.44   Median :29.44
##  Mean   :27.49   Mean   :27.65   Mean   :28.53   Mean   :28.36
##  3rd Qu.:30.56   3rd Qu.:30.56   3rd Qu.:31.11   3rd Qu.:32.78
##  Max.   :32.78   Max.   :35.00   Max.   :34.44   Max.   :36.67
##      X2007           X2008           X2009           X2010
##  Min.   :15.00   Min.   :10.00   Min.   :10.56   Min.   :19.44
##  1st Qu.:27.22   1st Qu.:26.39   1st Qu.:23.89   1st Qu.:27.78
##  Median :30.00   Median :29.44   Median :28.33   Median :32.22
##  Mean   :29.67   Mean   :28.06   Mean   :27.22   Mean   :30.67
##  3rd Qu.:31.94   3rd Qu.:31.39   3rd Qu.:31.11   3rd Qu.:33.89
##  Max.   :40.00   Max.   :35.00   Max.   :35.00   Max.   :36.11
##      X2011           X2012           X2013           X2014
##  Min.   :15.00   Min.   :13.33   Min.   :13.33   Min.   :17.22
##  1st Qu.:26.11   1st Qu.:26.39   1st Qu.:25.00   1st Qu.:27.50
##  Median :31.67   Median :29.44   Median :28.89   Median :30.00
##  Mean   :29.60   Mean   :29.25   Mean   :27.59   Mean   :28.86
##  3rd Qu.:34.44   3rd Qu.:32.50   3rd Qu.:31.11   3rd Qu.:31.67
##  Max.   :37.22   Max.   :40.56   Max.   :33.33   Max.   :35.00
##      X2015
##  Min.   :13.33
##  1st Qu.:25.00
##  Median :29.44
##  Mean   :28.50
##  3rd Qu.:32.22
##  Max.   :36.11
```

Visualize the temperature series with a boxplot

```
data_celsius$month = unlist(
  lapply(data_celsius$DAY, function(x) substr(x,start = nchar(x)-3+1,stop = nchar(x)+1)))
```

```
data_mod <- melt(data_celsius, id.vars='month',
                 measure.vars=temp_cols)
month_order = c('Jul', 'Aug', 'Sep', 'Oct')
ggplot(data_mod) +
geom_boxplot(aes(x=factor(month, level = month_order), y=value, color=variable))
```



#Analysis

Question 6.1

Question 6.2

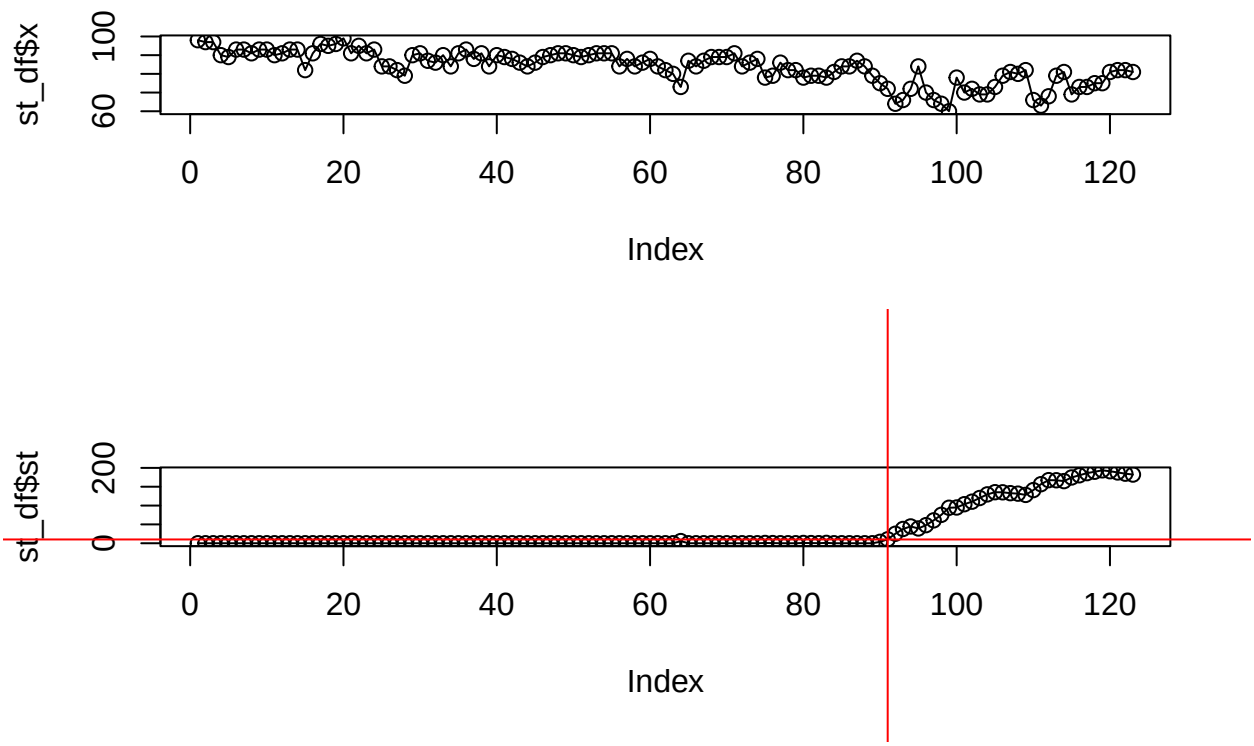
Part 1: Start of Winter through cusum

```
CUSUM_ad = function(x, year, days = data$DAY ,C = 0.30,T = 4, if_c_relative = FALSE){
  mean_x = mean(x)
  sd_x = sd(x)
  if (if_c_relative) {
    C = C * sd_x
  }
  #apply cusum
  x1 = lapply(x, function(xi)  (mean_x - xi - C))
```

```
  st_df = data.frame(st=double())
  st = 0
  for(i in 1:length(x)){
    #print (paste(i,class(st),class(x1[[i]])))
    st = max(0, st + x1[[i]])
    st_df[i,] = c(st)
  }
  #Find index where St > T
  st_df$Day = days
  st_df["anamoly"] = 0
  st_df$anamoly[st_df$st > T] = 1
  st_df$xt = x
  #plot x with T and if decrease see
  par(mfrow=c(2,1))
  plot(st_df$x,type='o')
  title(paste("Temperature (in Celsius) and S_t (bottom) for year",year), line = -1, outer = TRUE)
  plot(st_df$st,type='o')
  pushViewport(viewport())
  grid.lines(x = c(0,1), y = grconvertY(T, "user", "ndc"), gp = gpar(col = "red"))
  change_detected = which.max(st_df$st>T)
  #for zero true, which max returns 1
  if (sum(st_df$st>T) == 0){
    print ("No value greater than T")
    change_detected = NaN
  }
  grid.lines(x = grconvertX(change_detected, "user", "ndc"), y = c(0,1), gp = gpar(col = "red"))
  popViewport()
  #return first point where St touches T
  change_length = length(days)-change_detected
  return (c(as.numeric(year),as.numeric(change_detected),as.numeric(change_length)))
}
```

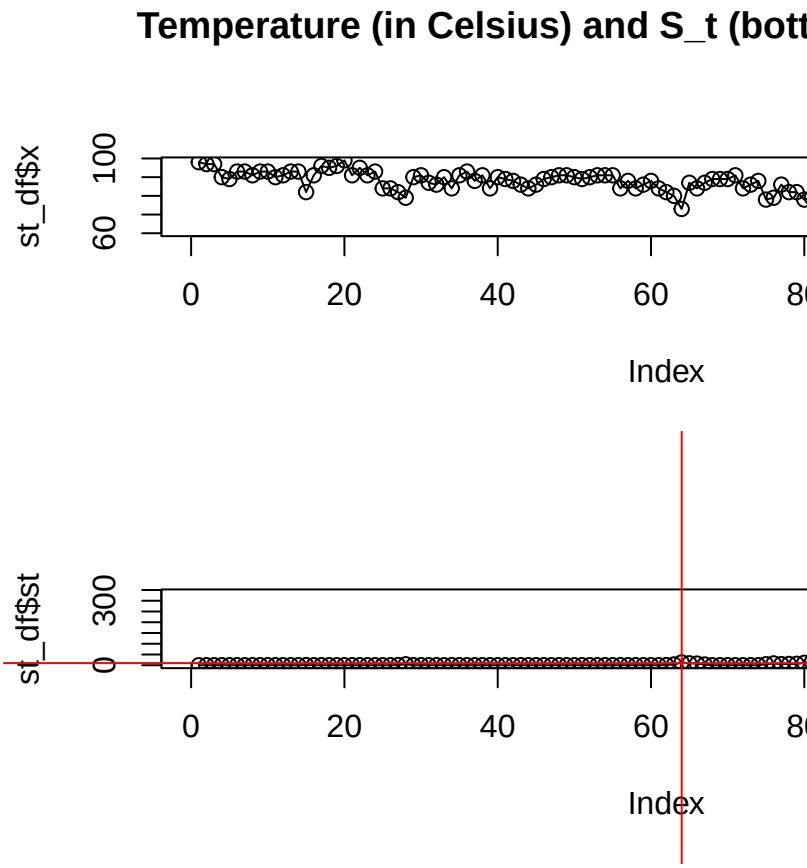
Function for Cusum Test the function for 1996

```
CUSUM_ad(data$X1996,"1996", C= 5, T = 10)
```

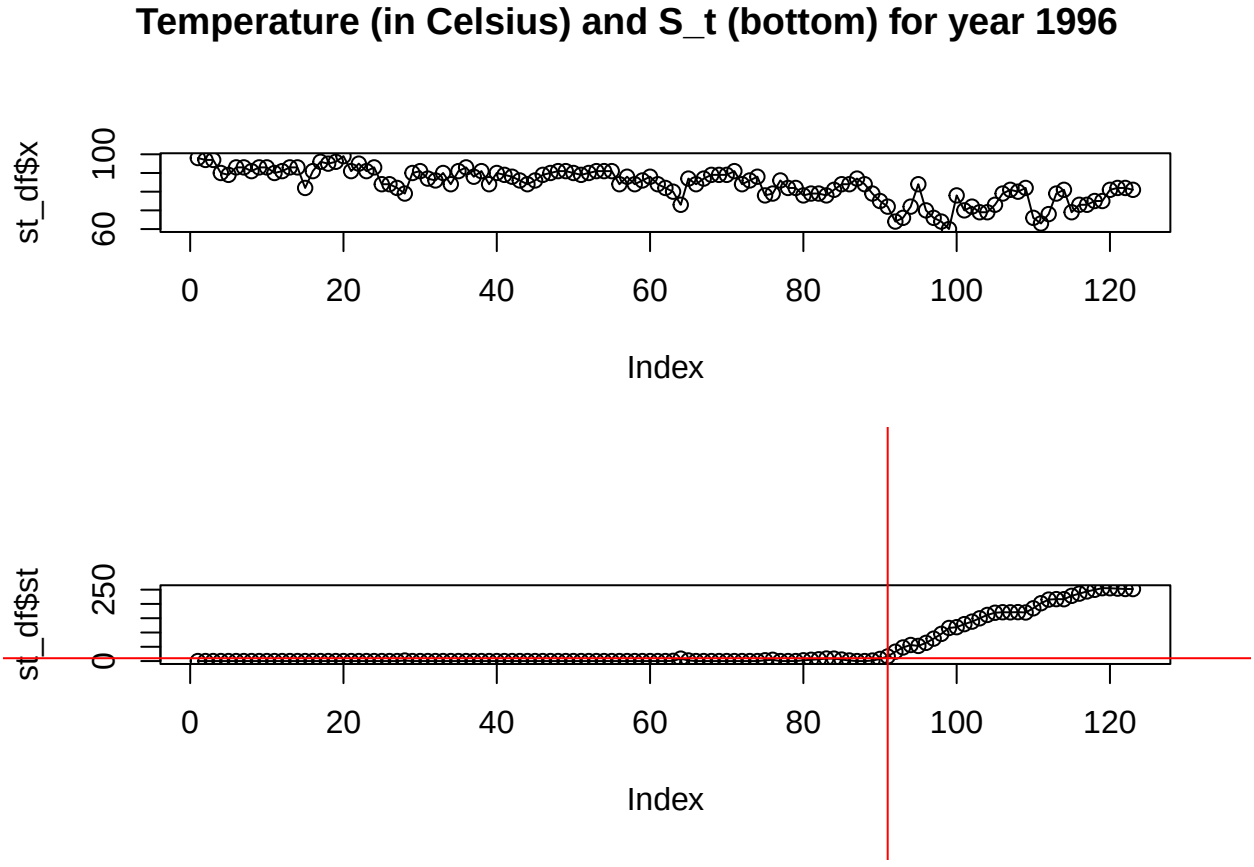


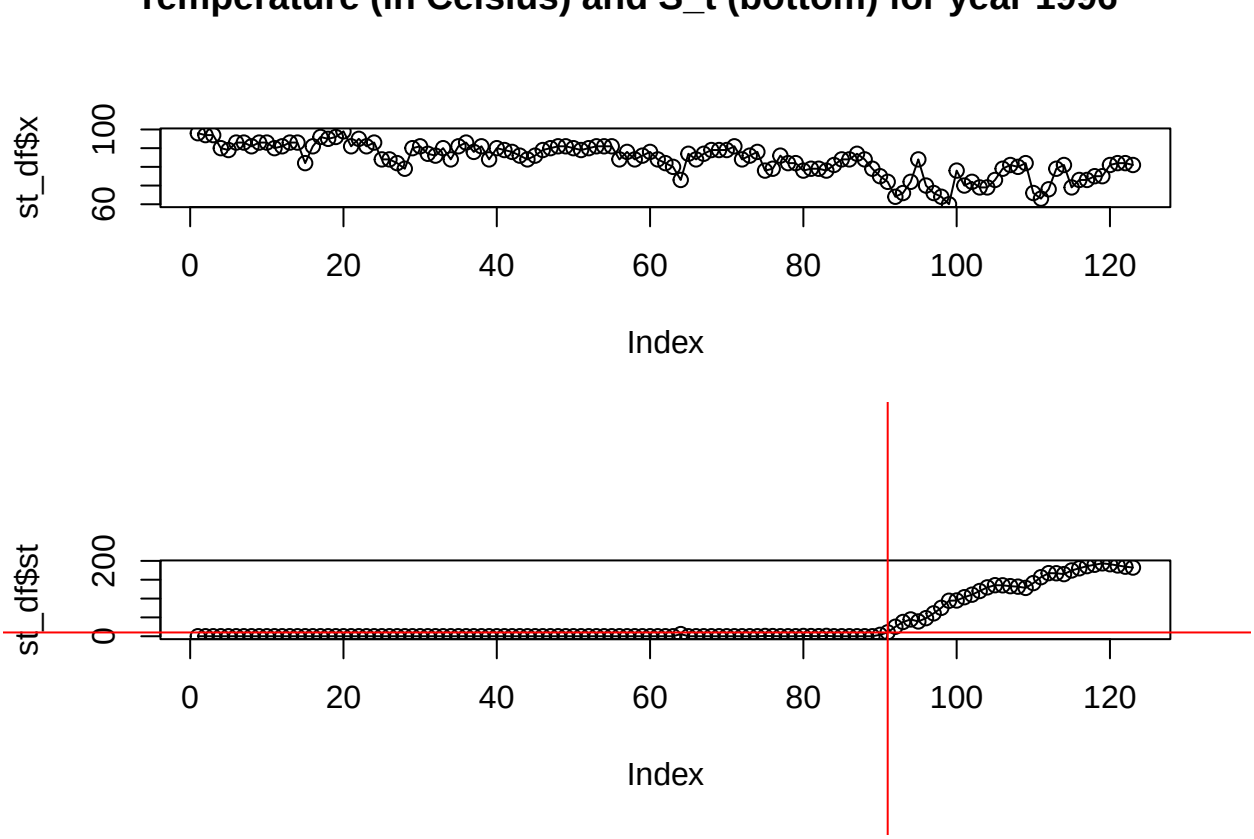

```
## [1] 1996   91   32
```

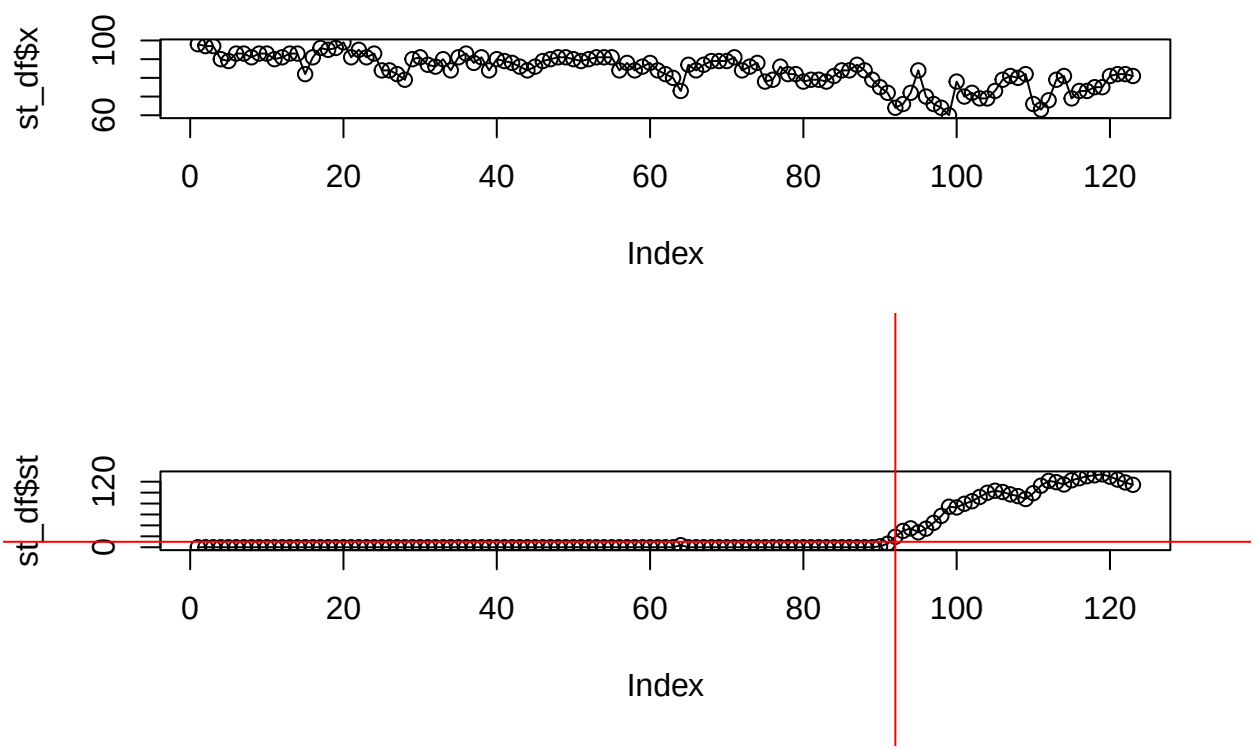
```
C_effect = lapply(seq(1,20,2),function(x) CUSUM_ad(data$X1996,"1996", C= x, T = 10))
```

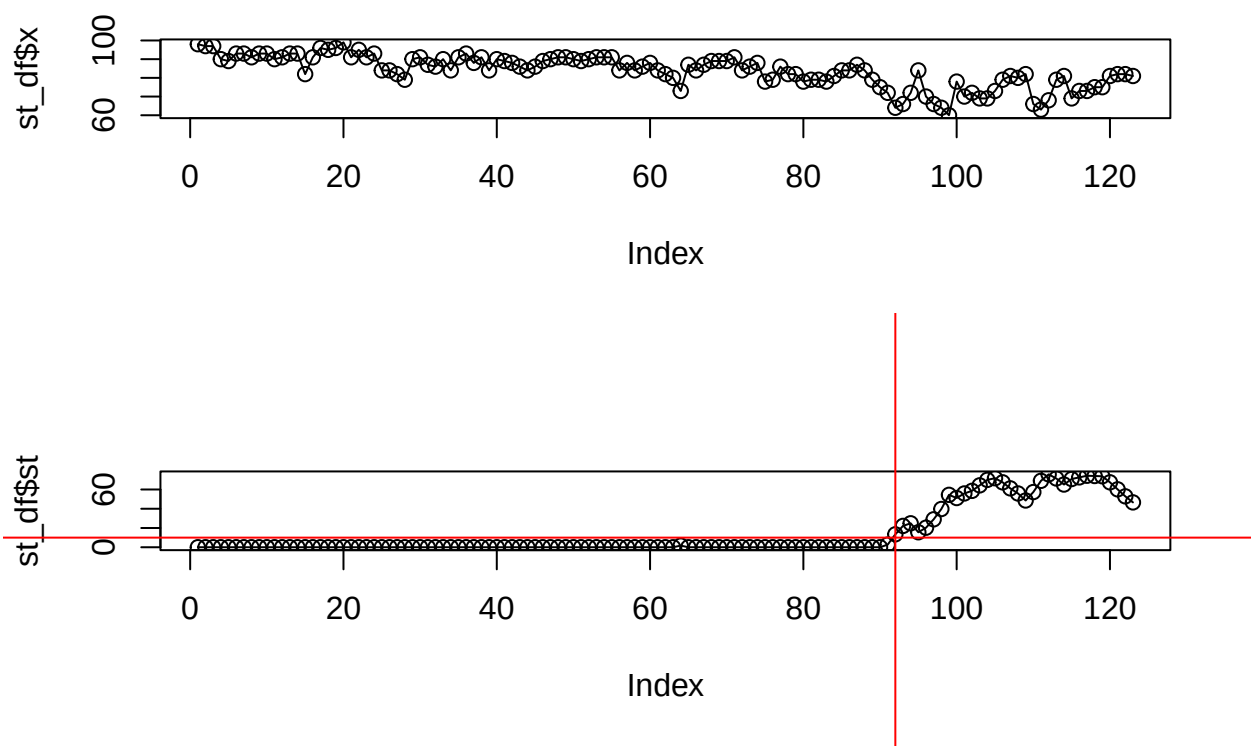


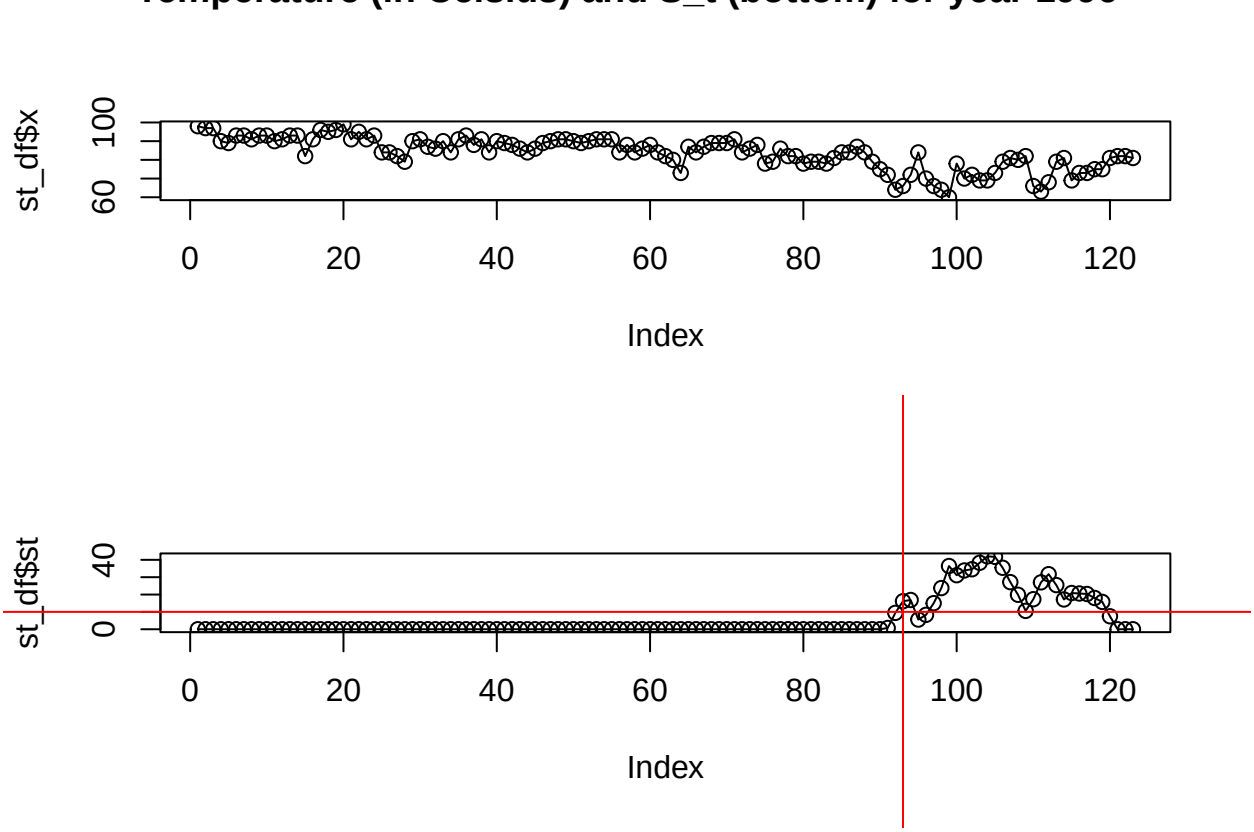
Understanding the effect of C on St calculation

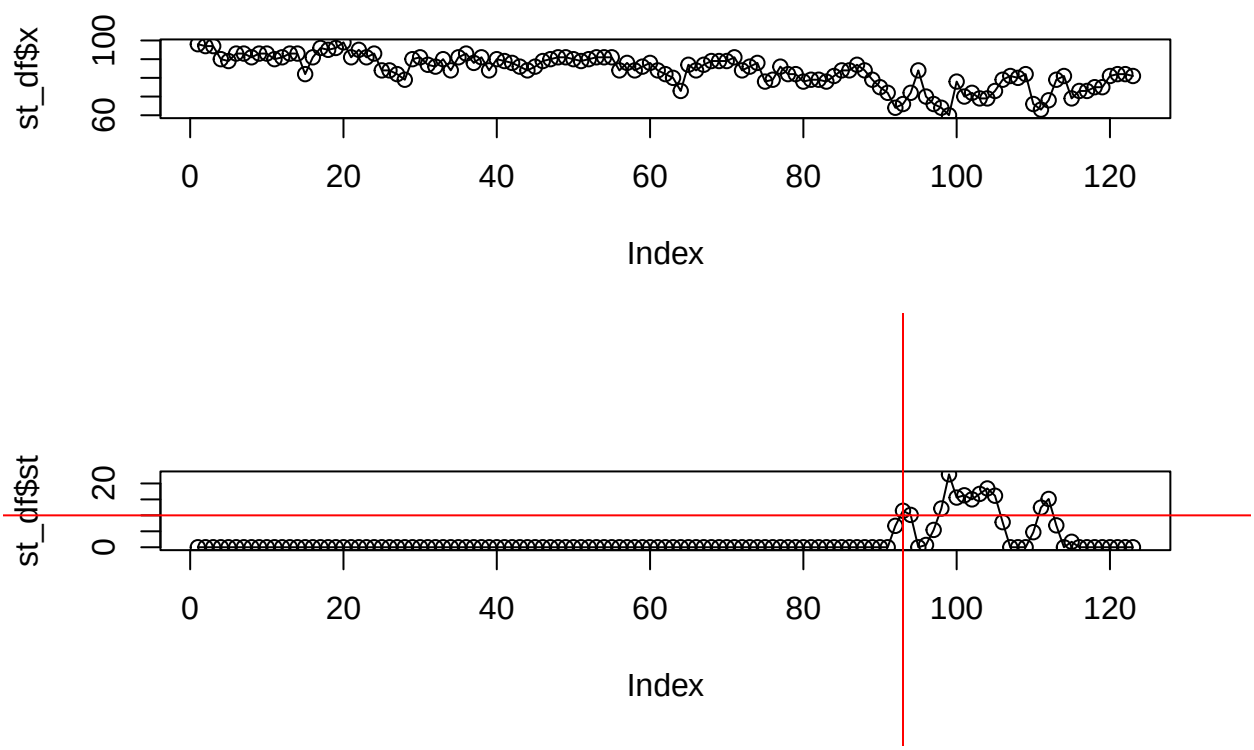


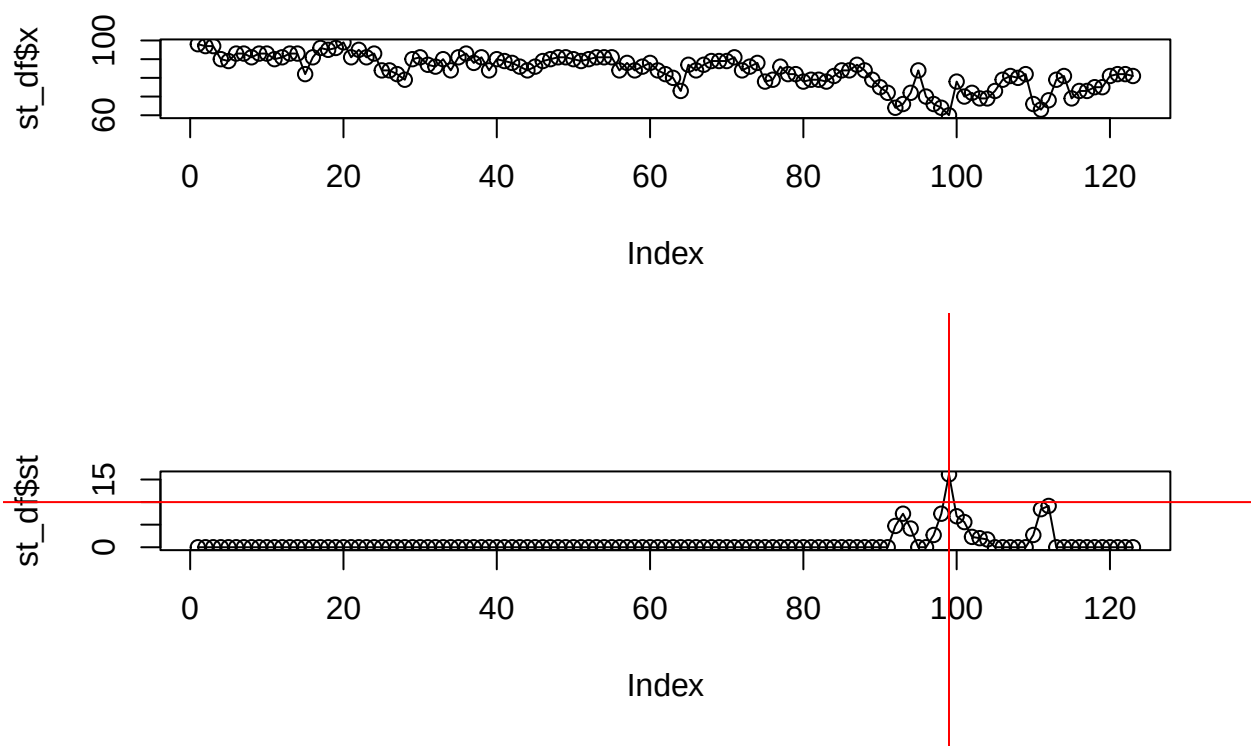

Temperature (in Celsius) and S_t (bottom) for year 1996
st_df$x
60
100
0
20
40
60
80
100
120
Index
st_df$st
0
200
0
20
40
60
80
100
120
Index

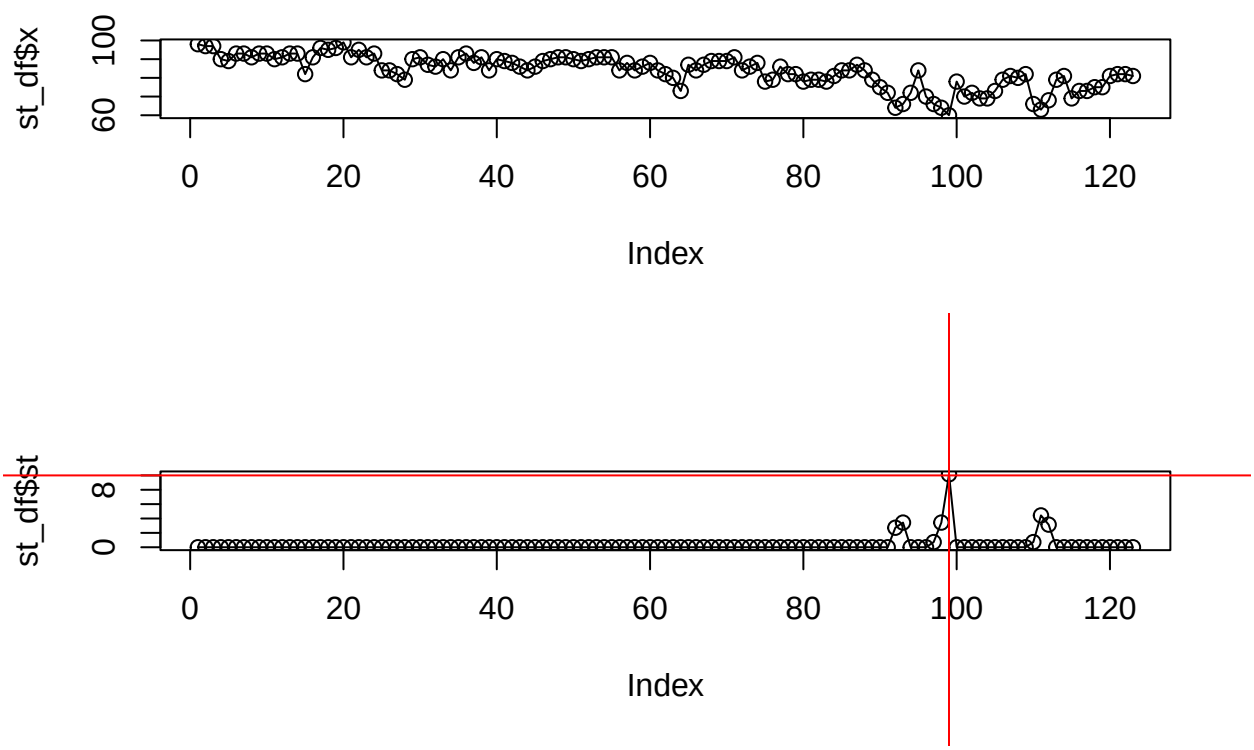

Temperature (in Celsius) and S_t (bottom) for year 1996
st_df$x
60
100
0
20
40
60
80
100
120
Index
st_df$st
0
120
0
20
40
60
80
100
120
Index

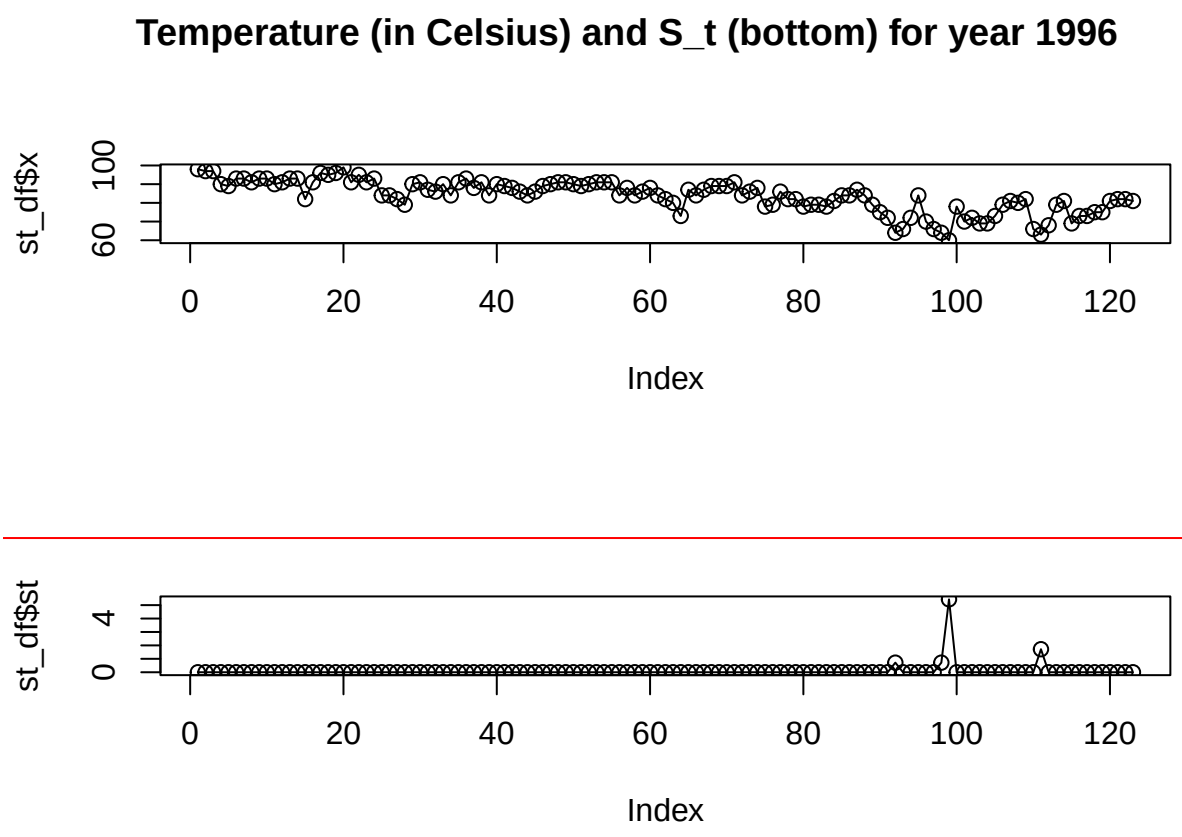

Temperature (in Celsius) and S_t (bottom) for year 1996
st_df$x
60
100
0
20
40
60
80
100
120
Index
st_df$st
0
60
0
20
40
60
80
100
120
Index


Temperature (in Celsius) and S_t (bottom) for year 1996
st_df$x
60
100
0
20
40
60
80
100
120
Index
st_df$st
0
40
0
20
40
60
80
100
120
Index


Temperature (in Celsius) and S_t (bottom) for year 1996
st_df$x
60
100
0
20
40
60
80
100
120
Index
st_df$st
0
20
0
20
40
60
80
100
120
Index


Temperature (in Celsius) and S_t (bottom) for year 1996
st_df$x
60
100
0
20
40
60
80
100
120
Index
st_df$st
0
15
0
20
40
60
80
100
120
Index

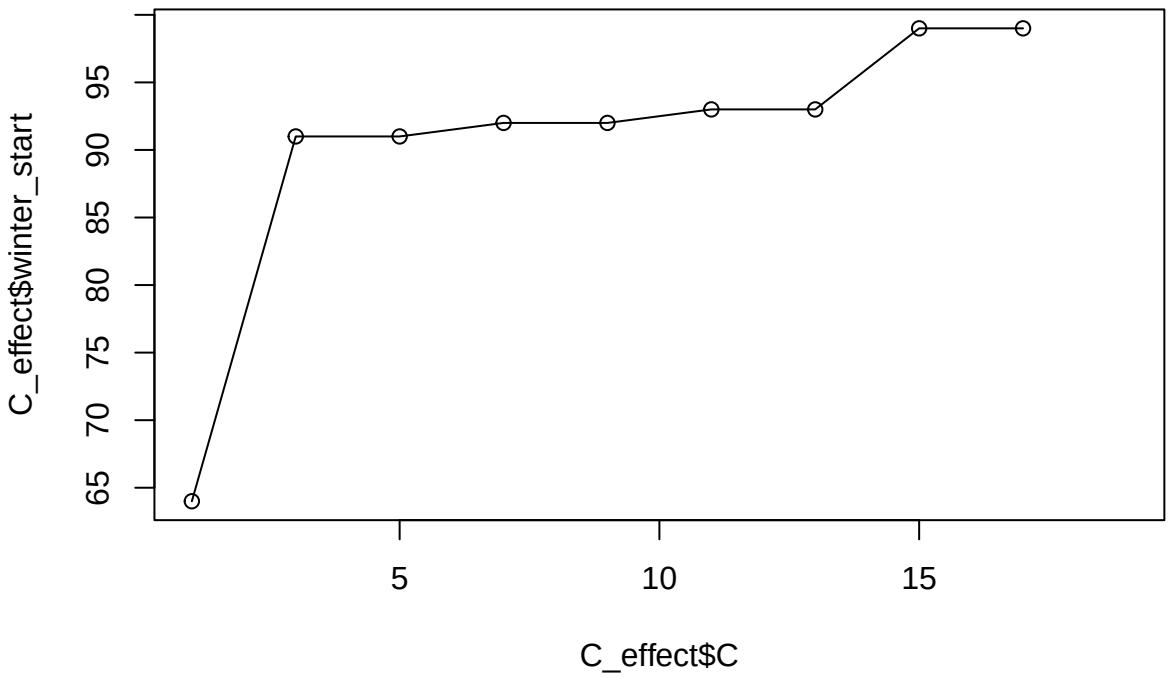

Temperature (in Celsius) and S_t (bottom) for year 1996
st_df$x
60
100
0
20
40
60
80
100
120
Index
st_df$st
0
8
0
20
40
60
80
100
120
Index




```
## [1] "No value greater than T"
```

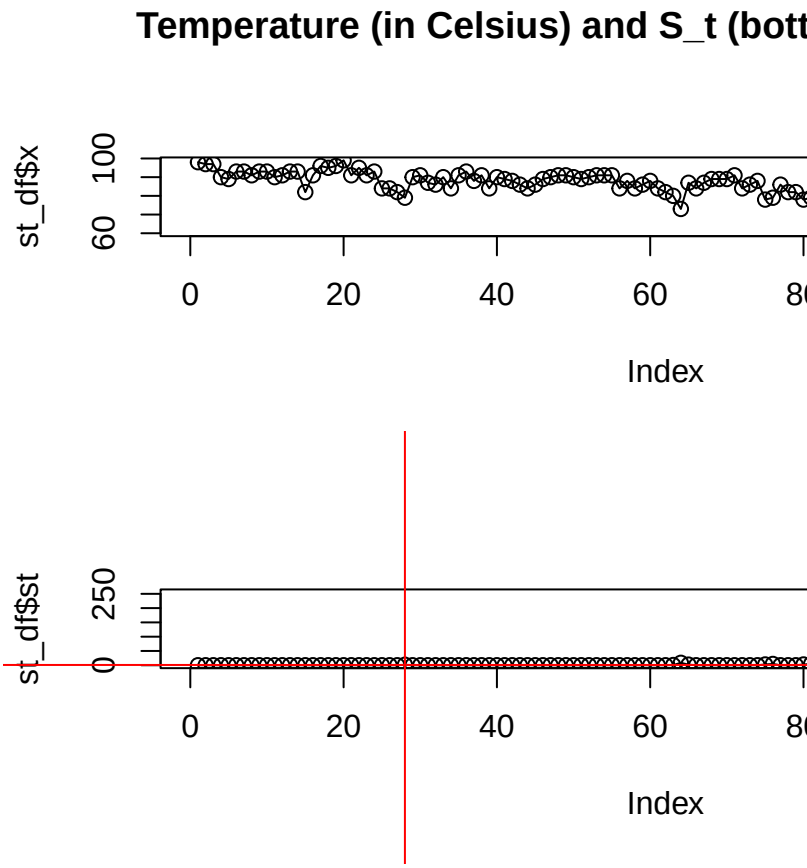
C is like the margin for noise built into the model. From the above graphs it is apparent that increasing C, makes the end part (>90) of St more grainy and abrupt. This is intuitive since St = max(0,St-1 + (mean - x - C), therefore any drop in xt (relative to mean) lower than C would be marginalised. This further warrants that St would be increasing only when we see a sustained drop in xt beyond C.

```
C_effect = as.data.frame(do.call(rbind, C_effect))
colnames(C_effect) = c("year","winter_start","winter_length")
C_effect$C = seq(1,20,2)
plot(x=C_effect$C,y=C_effect$winter_start,type='o')
```

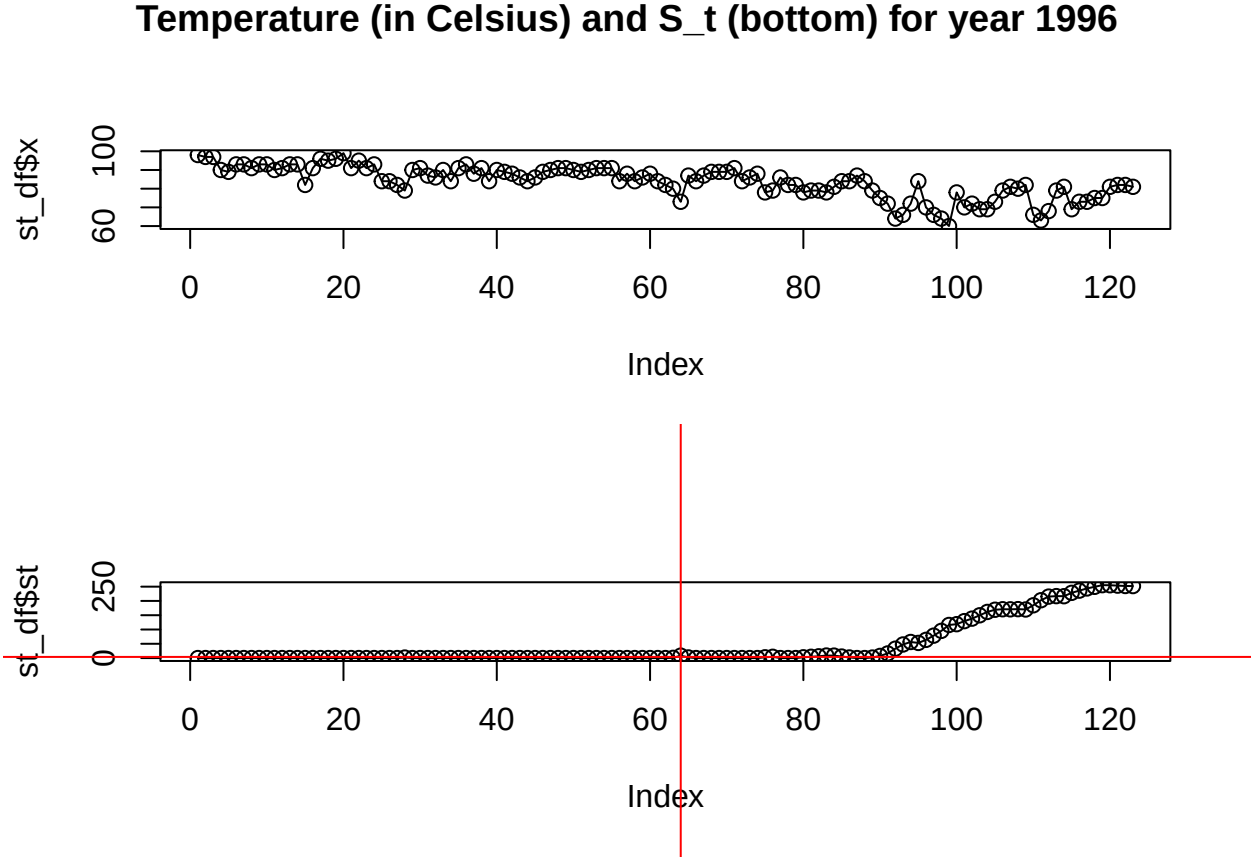


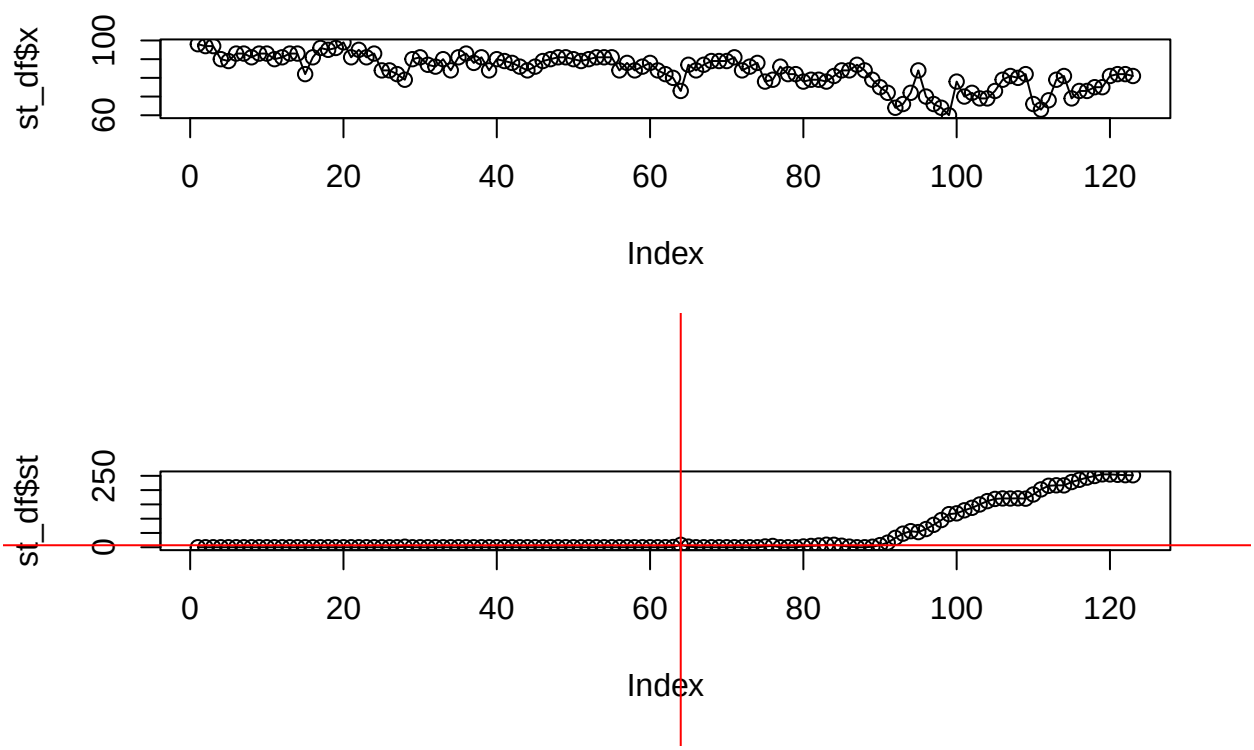
Increasing the effect of C thus makes the model less sensitive, and delays the resolution of change, winter_start increases.

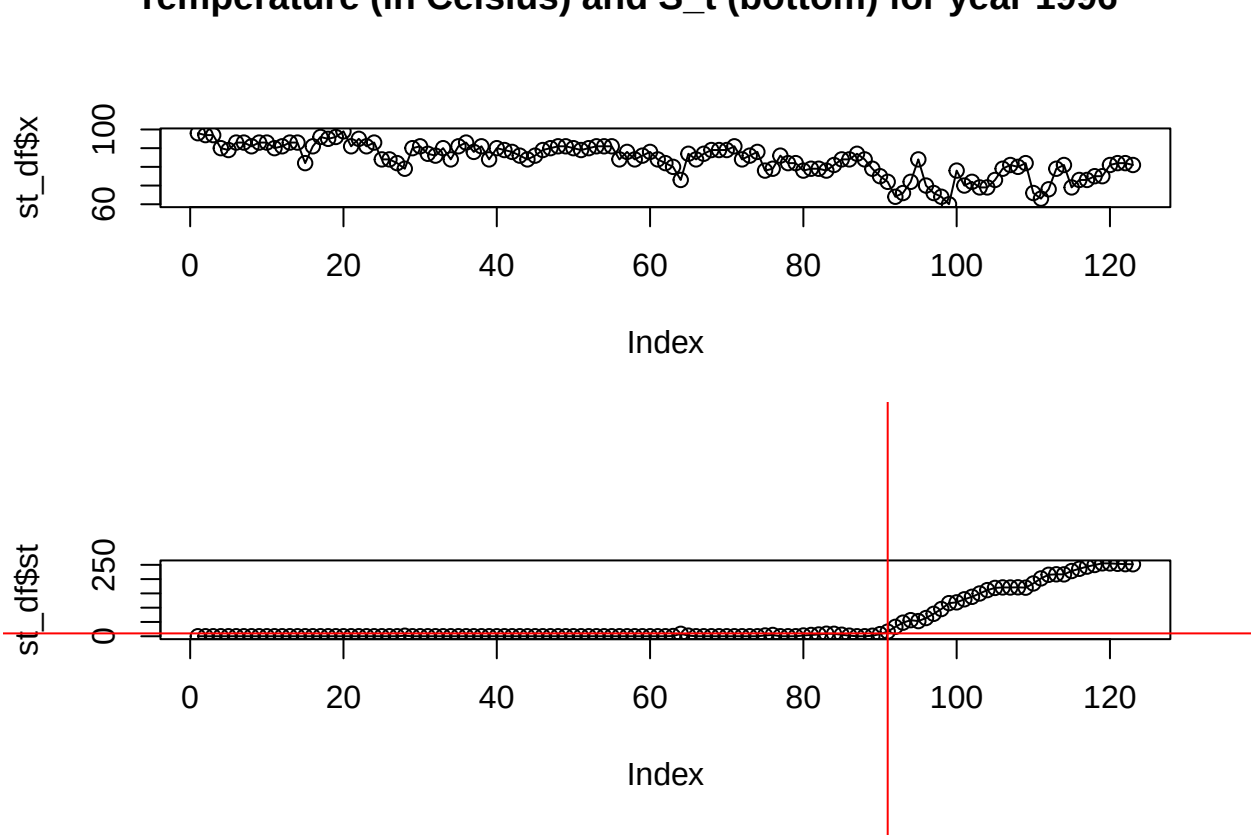
```
T_effect = lapply(seq(1,30,3),function(x) CUSUM_ad(data$X1996,"1996", C= 3, T = x))
```

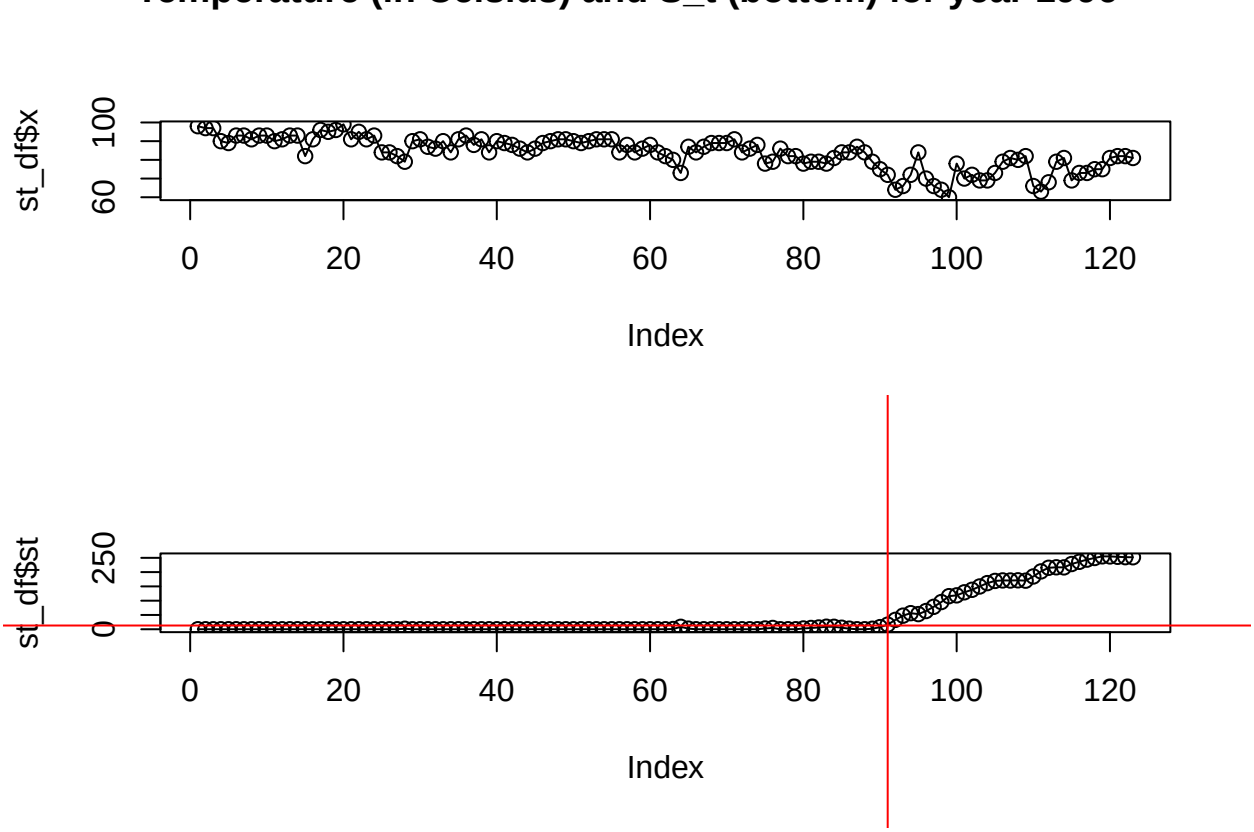


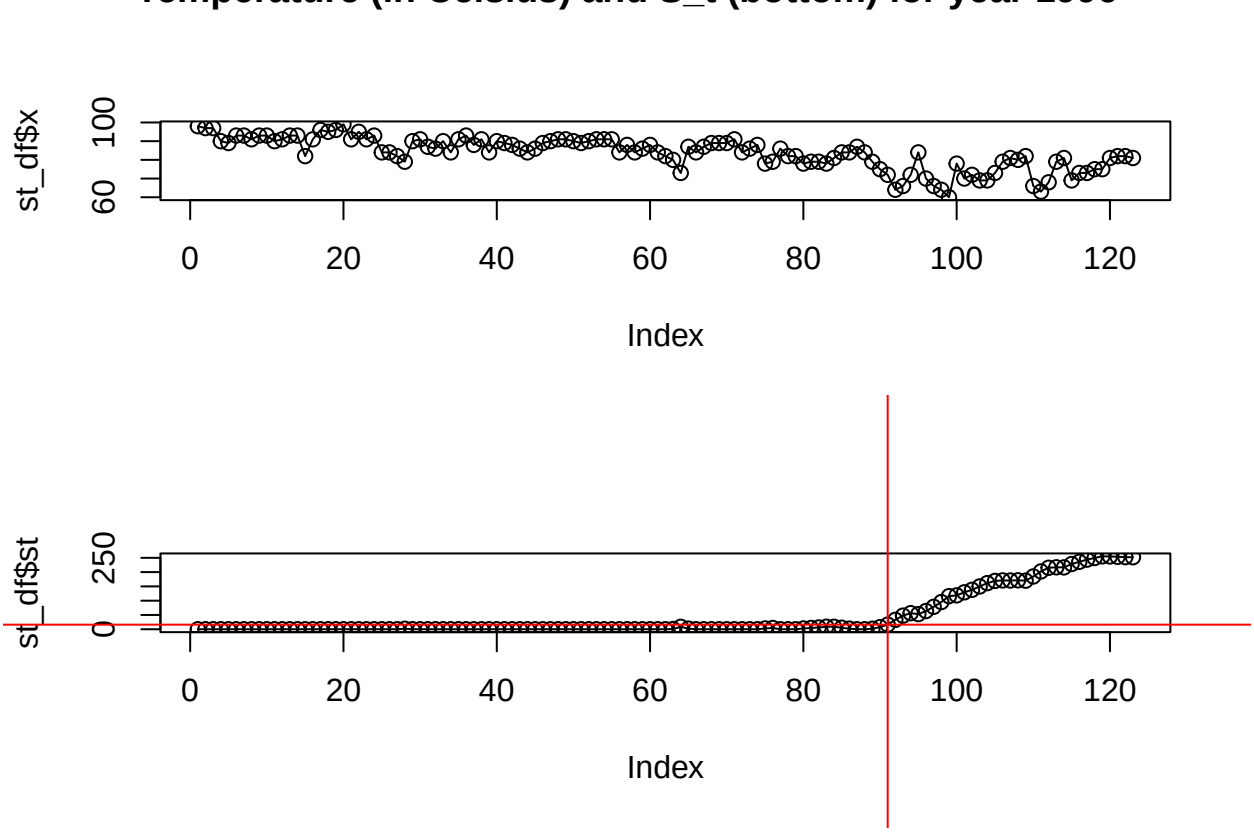
Understanding the effect of T on St calculation

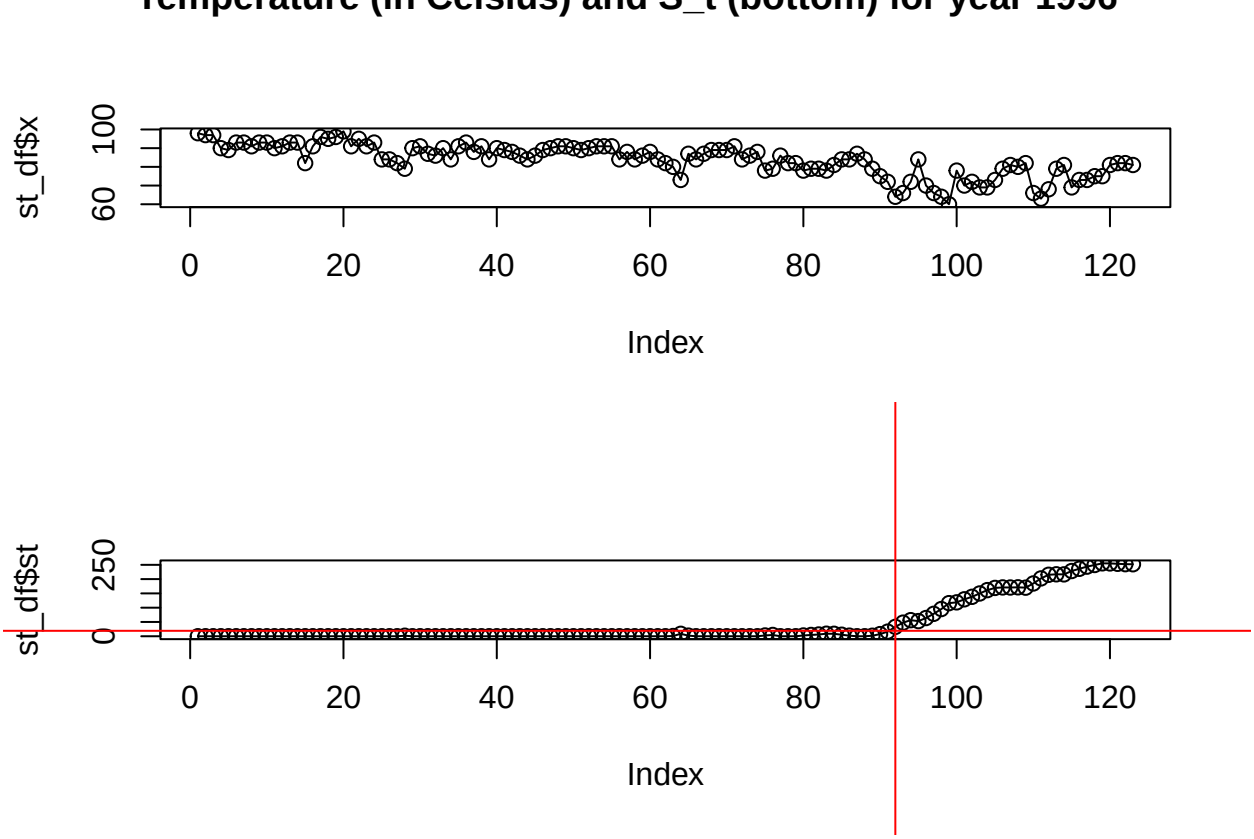


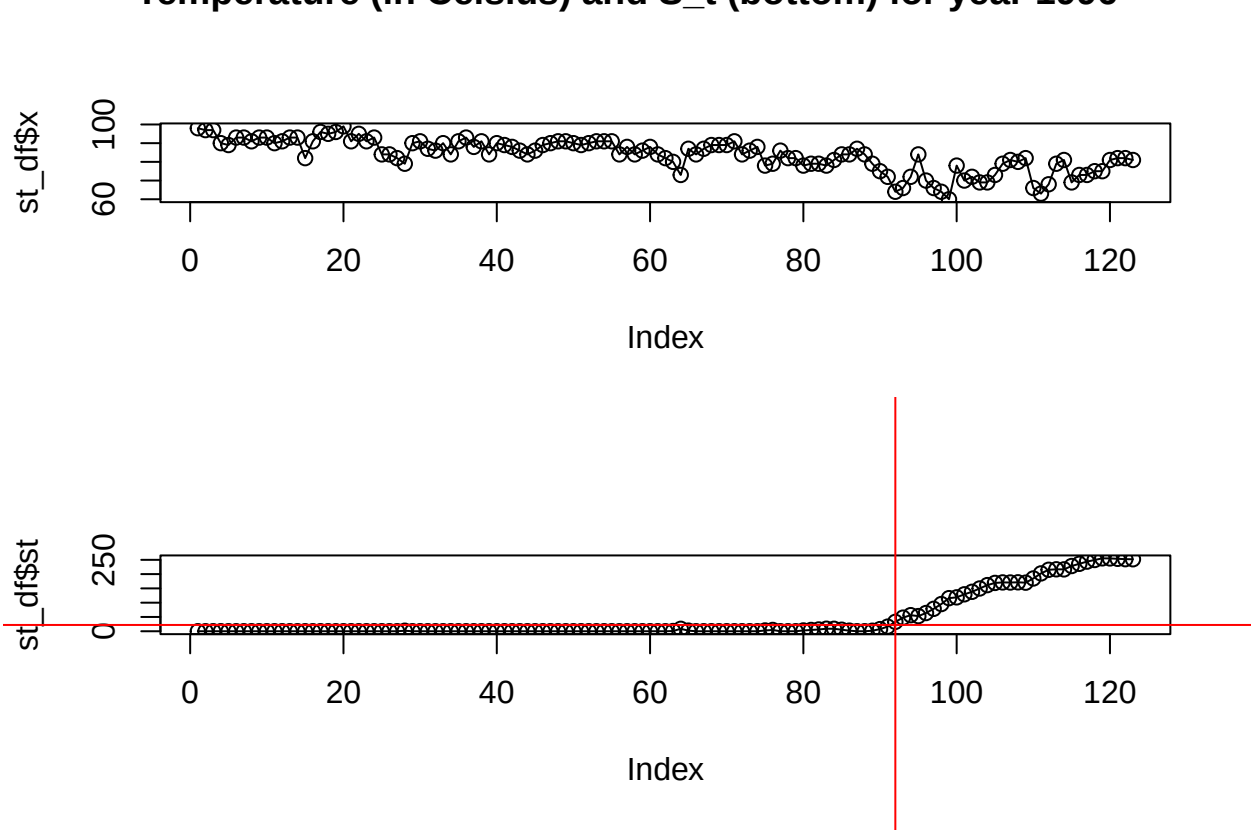

Temperature (in Celsius) and S_t (bottom) for year 1996
st_df$x
60
100
0
20
40
60
80
100
120
Index
st_df$st
0
250
0
20
40
60
80
100
120
Index

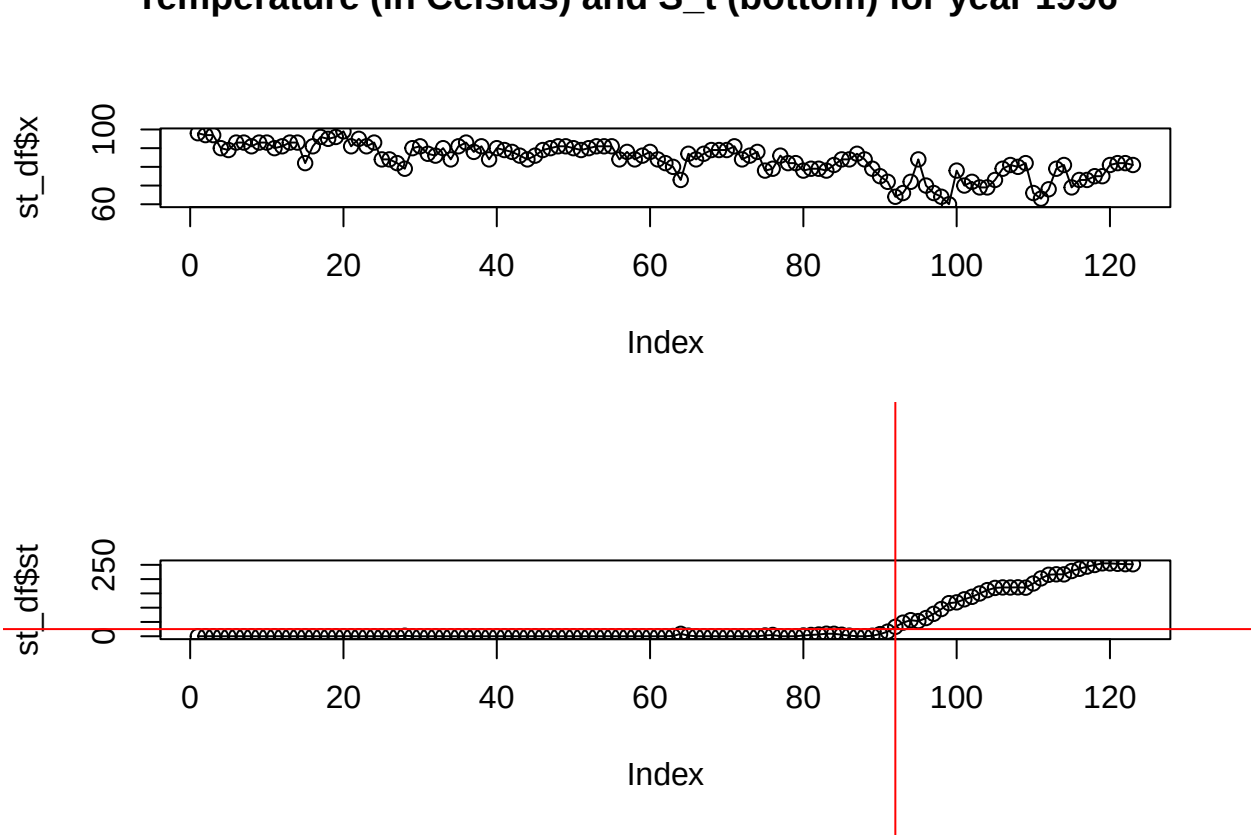

Temperature (in Celsius) and S_t (bottom) for year 1996
st_df$x
60
100
0
20
40
60
80
100
120
Index
st_df$st
0
250
0
20
40
60
80
100
120
Index

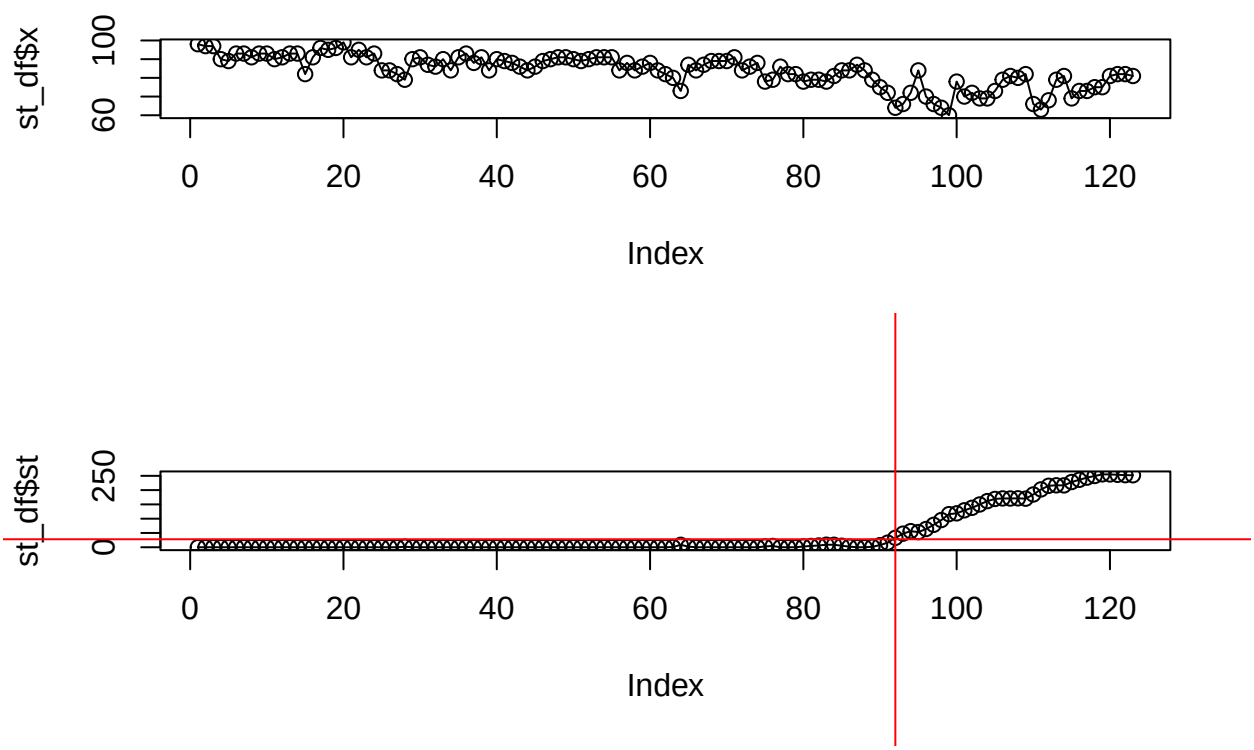

Temperature (in Celsius) and S_t (bottom) for year 1996
st_df$x
60
100
0
20
40
60
80
100
120
Index
st_df$st
0
250
0
20
40
60
80
100
120
Index


Temperature (in Celsius) and S_t (bottom) for year 1996
st_df$x
60
100
0
20
40
60
80
100
120
Index
st_df$st
0
250
0
20
40
60
80
100
120
Index

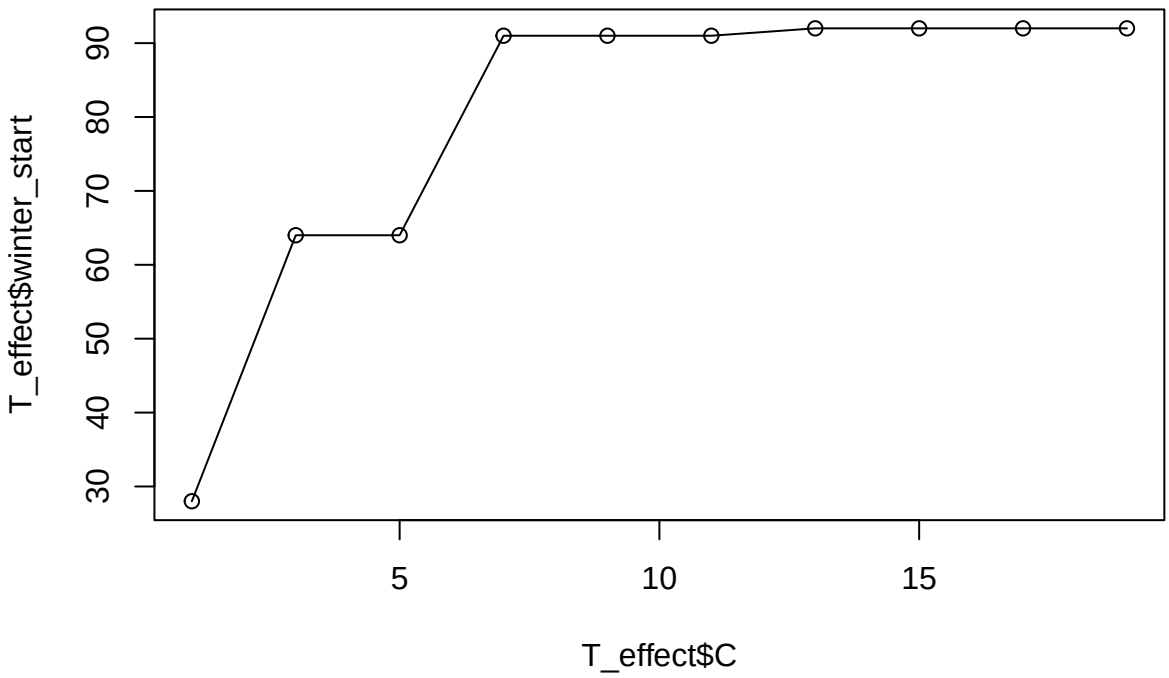

Temperature (in Celsius) and S_t (bottom) for year 1996
st_df$x
60
100
0
20
40
60
80
100
120
Index
st_df$st
0
250
0
20
40
60
80
100
120
Index


Temperature (in Celsius) and S_t (bottom) for year 1996
st_df$x
60
100
0
20
40
60
80
100
120
Index
st_df$st
0
250
0
20
40
60
80
100
120
Index


Temperature (in Celsius) and S_t (bottom) for year 1996
st_df$x
60
100
0
20
40
60
80
100
120
Index
st_df$st
0
250
0
20
40
60
80
100
120
Index



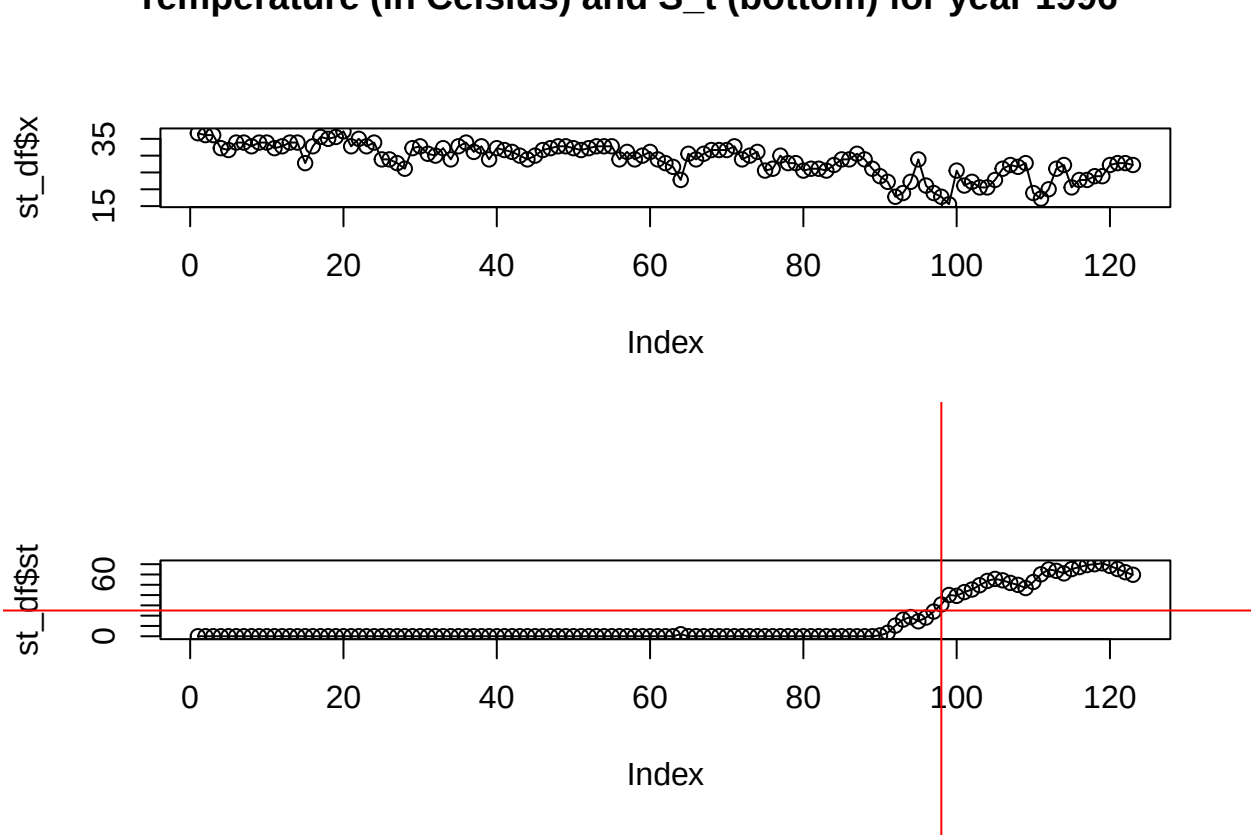

```
T_effect = as.data.frame(do.call(rbind, T_effect))
colnames(T_effect) = c("year","winter_start","winter_length")
T_effect$C = seq(1,20,2)
plot(x=T_effect$C,y=T_effect$winter_start,type='o')
```

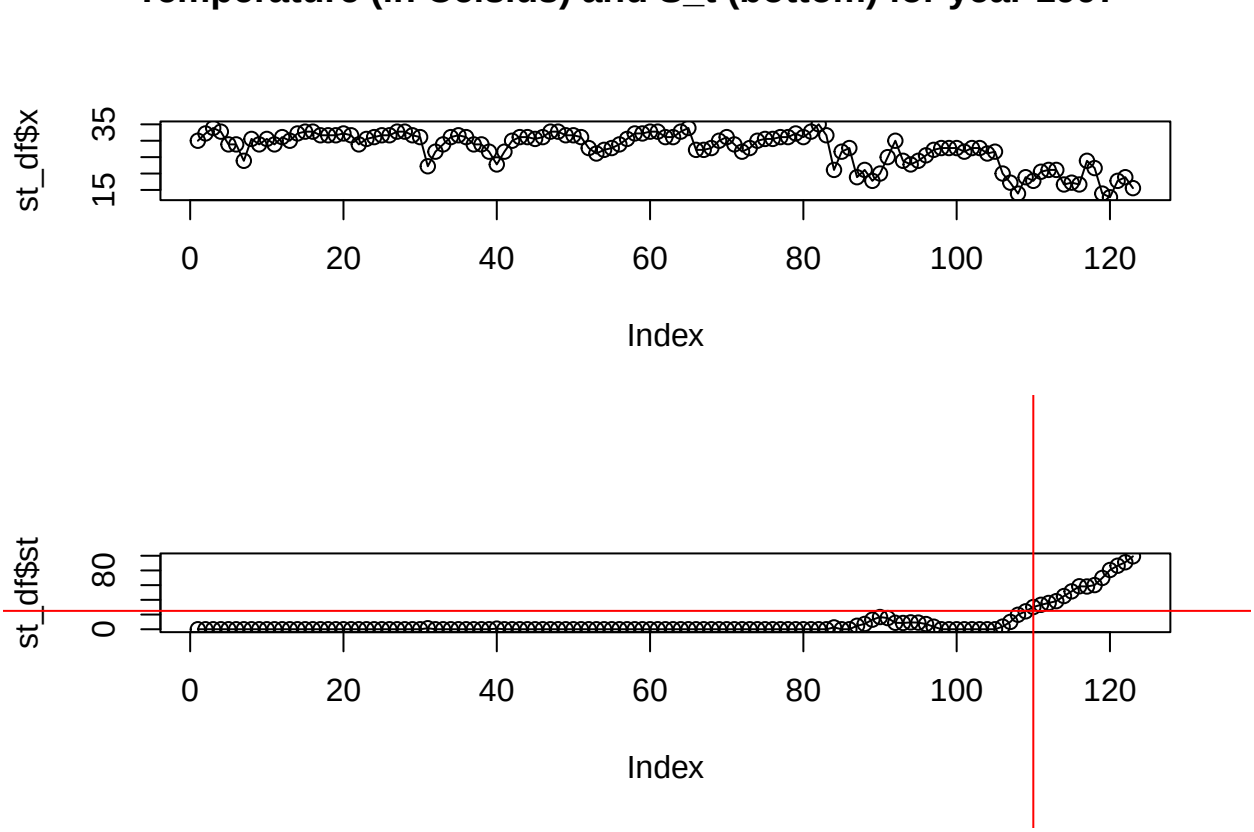


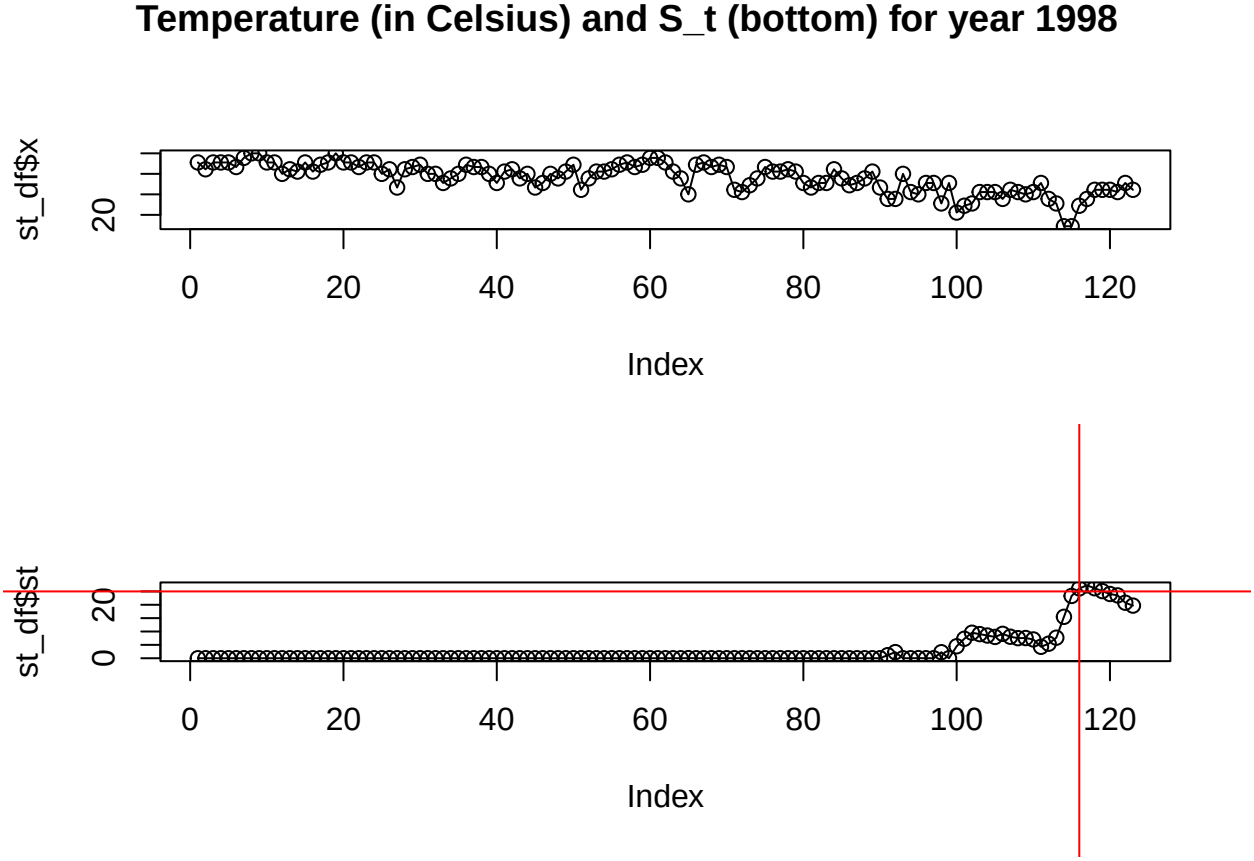
Increasing the effect of T too thus makes the model less sensitive, and delays the resolution of change, winter_start increases.

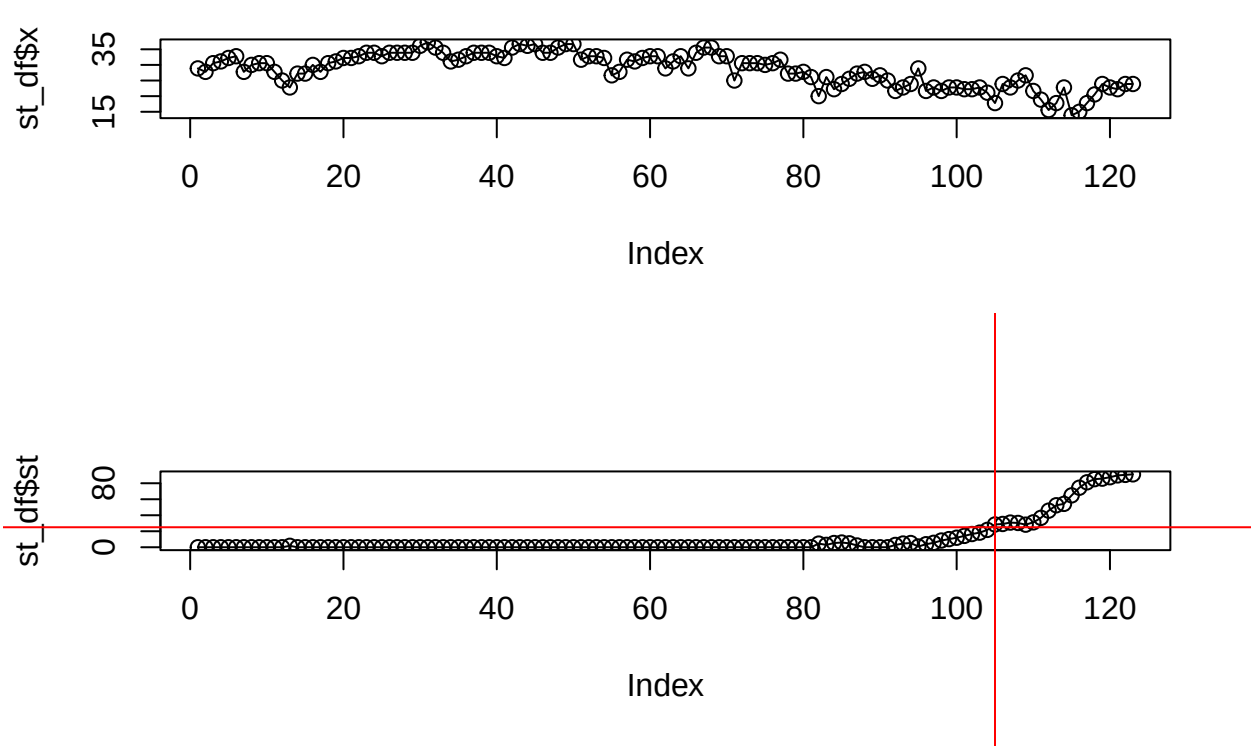
Basis the above inferences we have picked C= and T= (basis hit and trial)

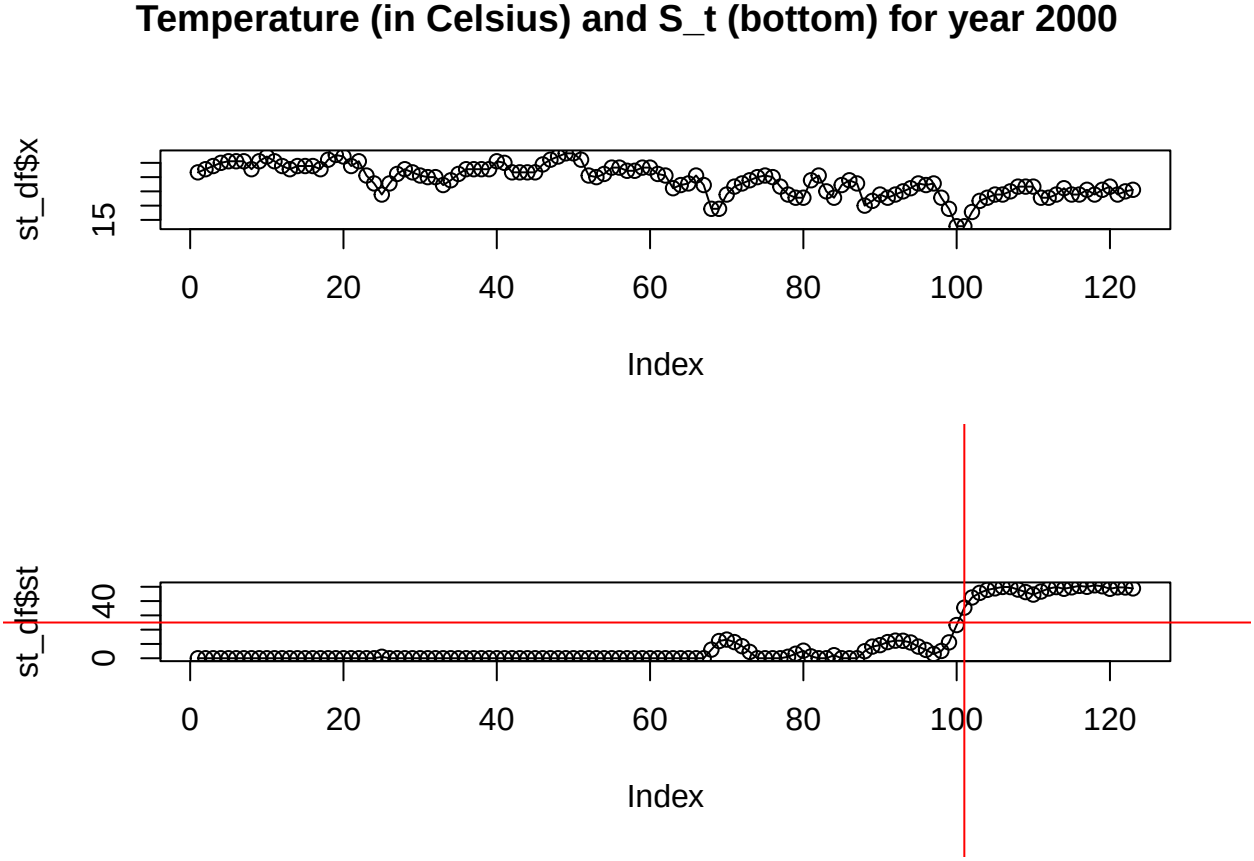
```
winter_start_days = lapply(temp_cols, function(col)
  CUSUM_ad(x=data_celsius[,col], year = substr(col,2,5), C = 4, T = 25)
  )
```

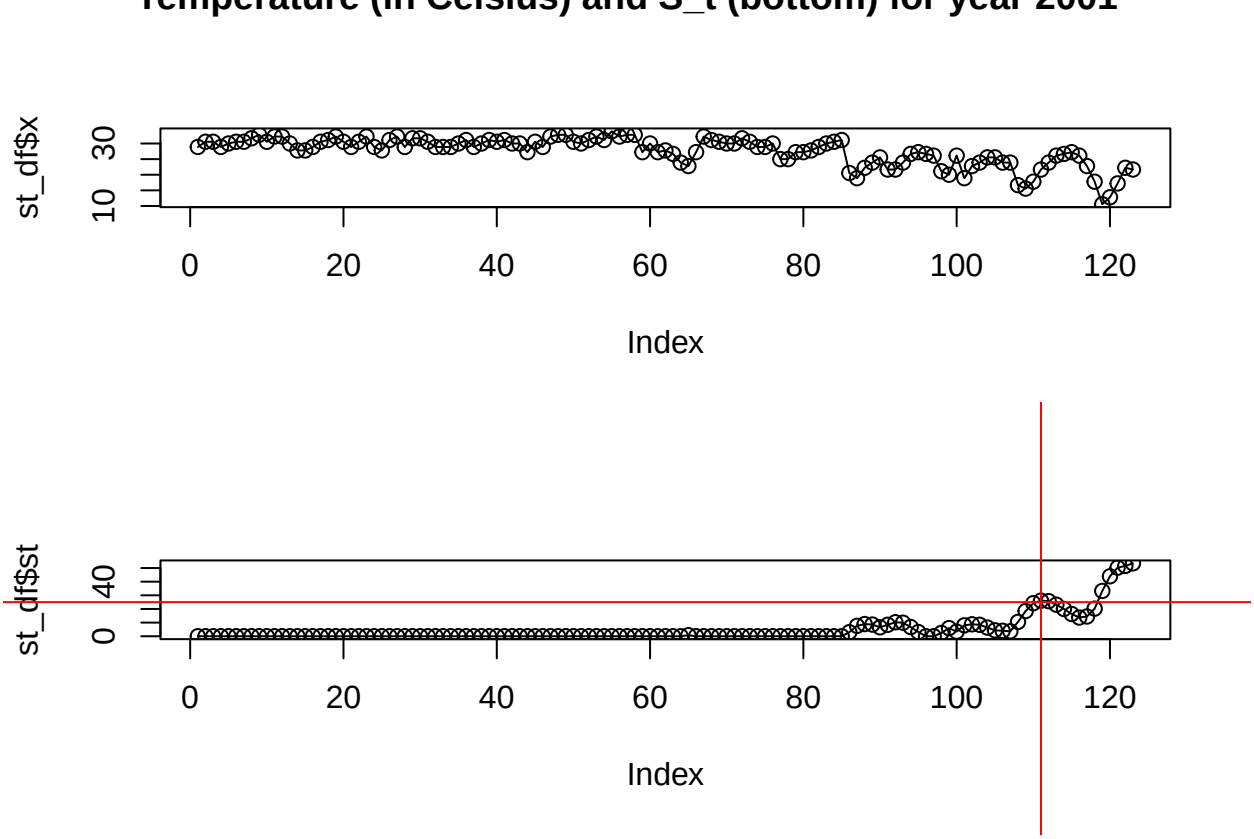

Temperature (in Celsius) and S_t (bottom) for year 1996
st_df$x
15
35
0
20
40
60
80
100
120
Index
st_df$st
0
60
0
20
40
60
80
100
120
Index

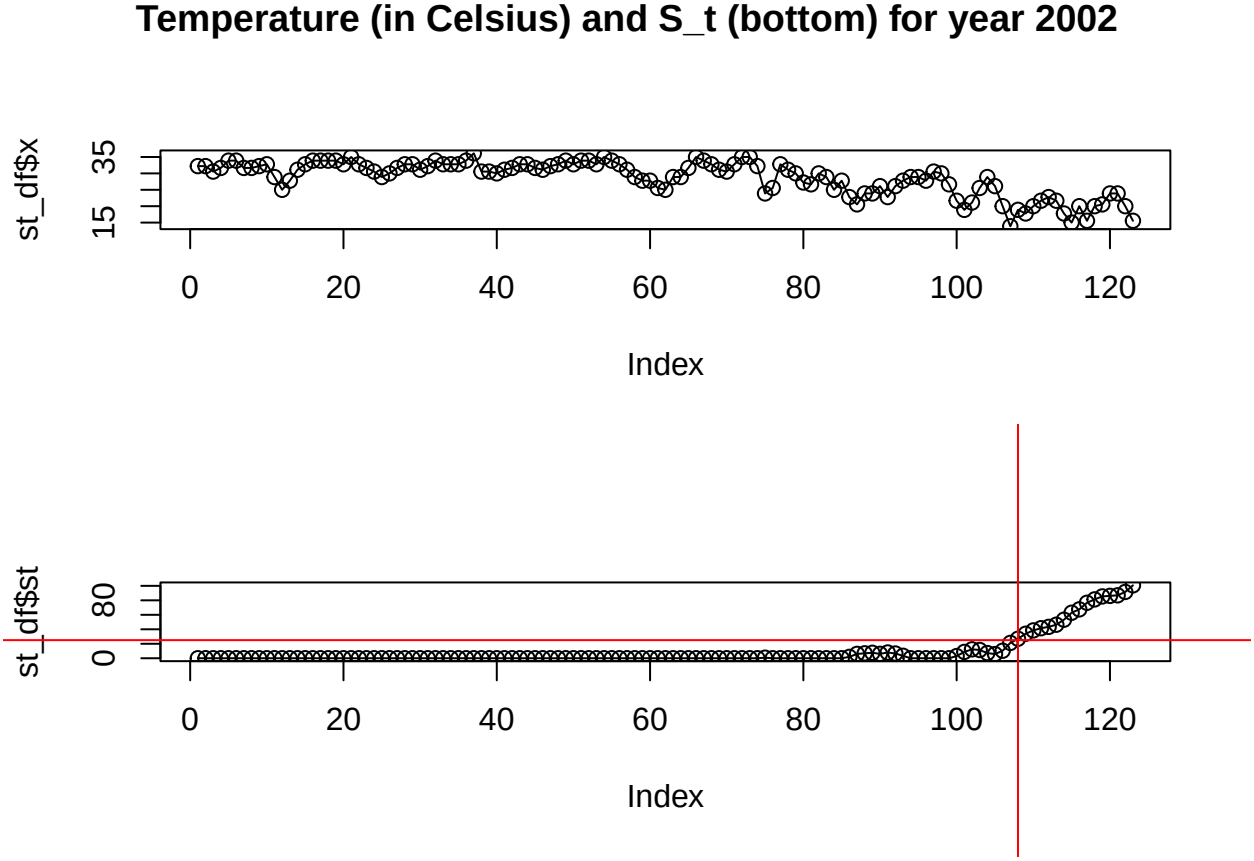

Temperature (in Celsius) and S_t (bottom) for year 1997
st_df$x
15
35
0
20
40
60
80
100
120
Index
st_df$st
0
80
0
20
40
60
80
100
120
Index

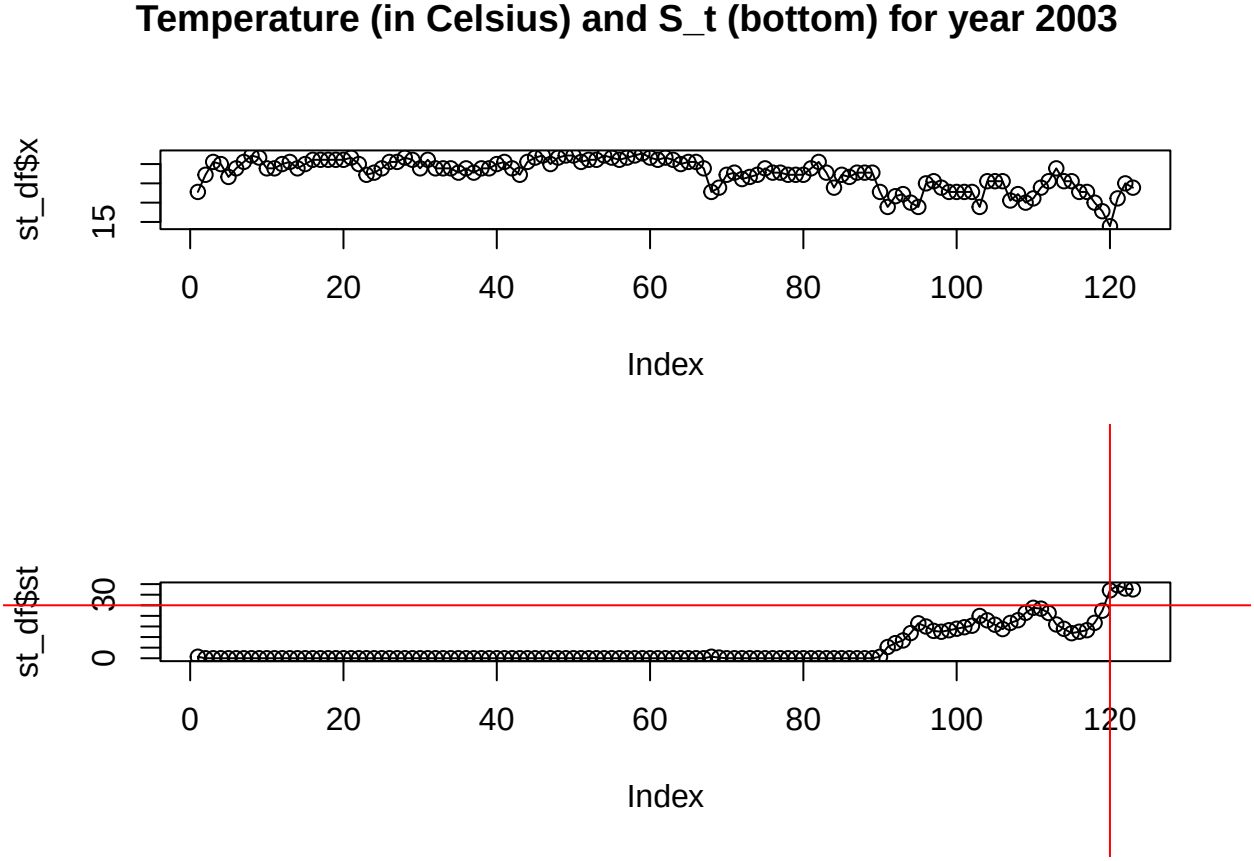

Temperature (in Celsius) and S_t (bottom) for year 1998
st_df$x
20
0
20
40
60
80
100
120
Index
st_df$st
0
20
0
20
40
60
80
100
120
Index

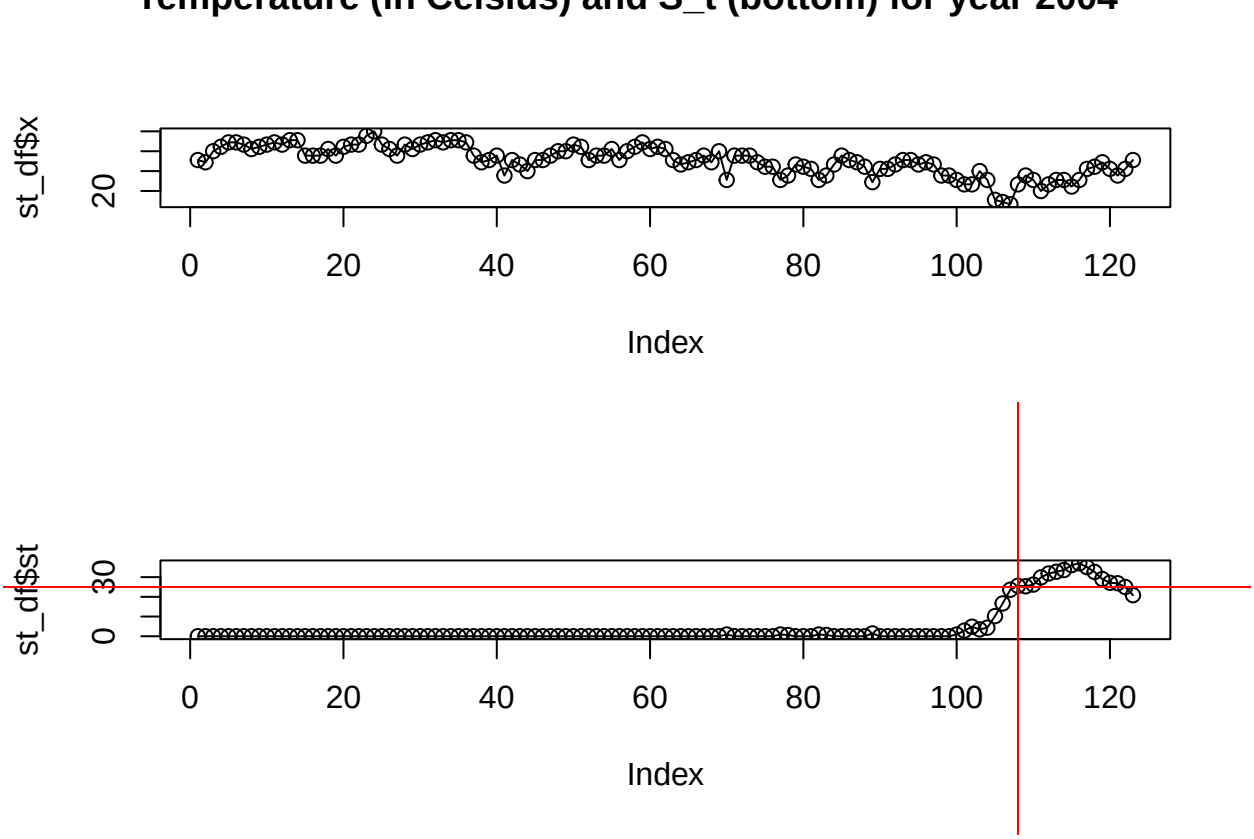

Temperature (in Celsius) and S_t (bottom) for year 1999
st_df$x
15
35
0
20
40
60
80
100
120
Index
st_df$st
0
80
0
20
40
60
80
100
120
Index

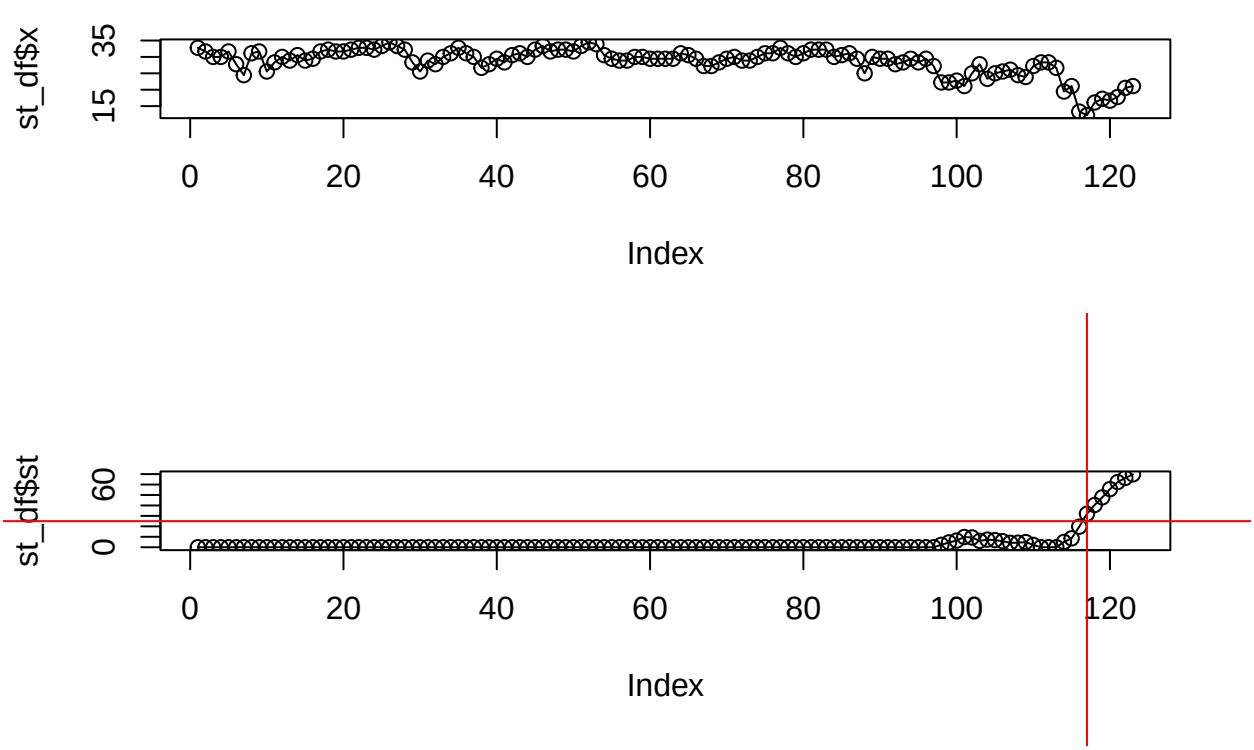

Temperature (in Celsius) and S_t (bottom) for year 2000
st_df$x
15
0
20
40
60
80
100
120
Index
st_df$st
0
40
0
20
40
60
80
100
120
Index

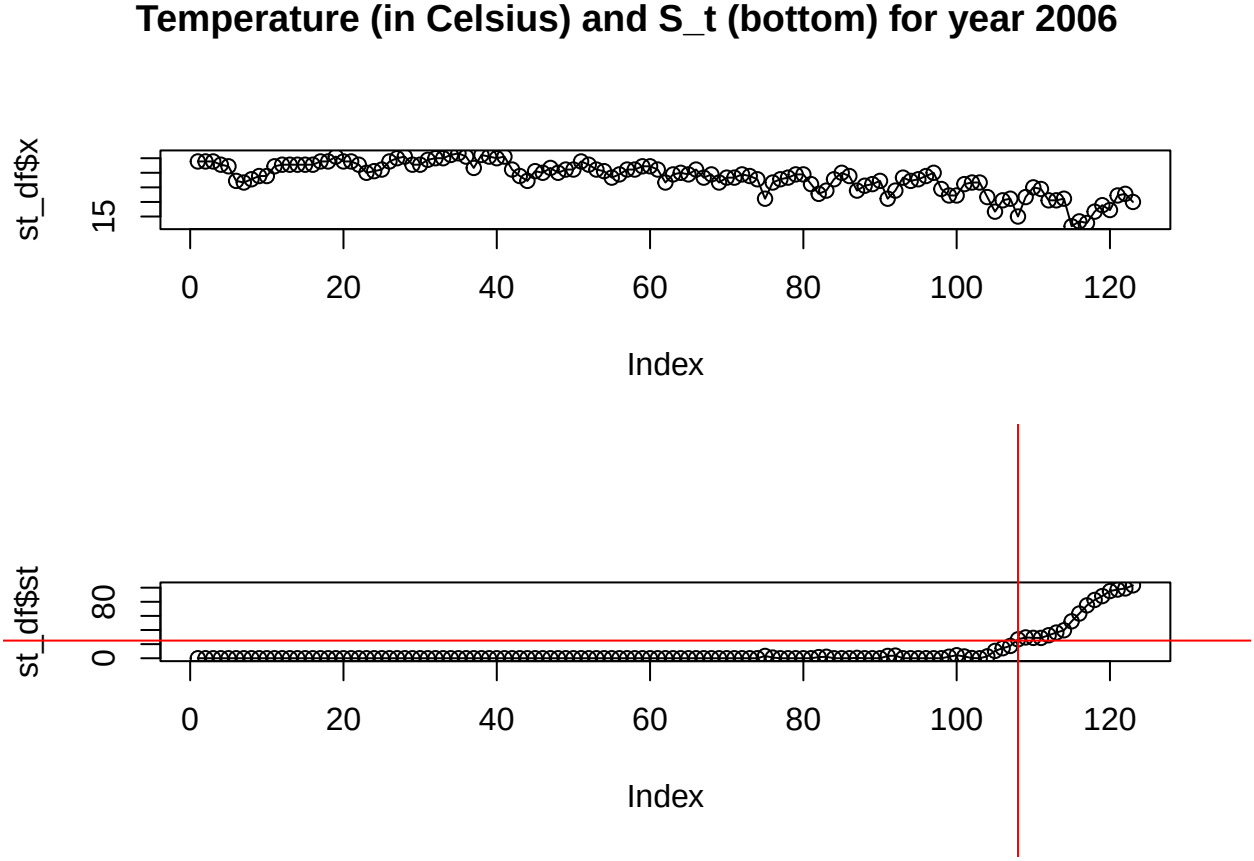

Temperature (in Celsius) and S_t (bottom) for year 2001
st_df$x
10
30
0
20
40
60
80
100
120
Index
st_df$st
0
40
0
20
40
60
80
100
120
Index

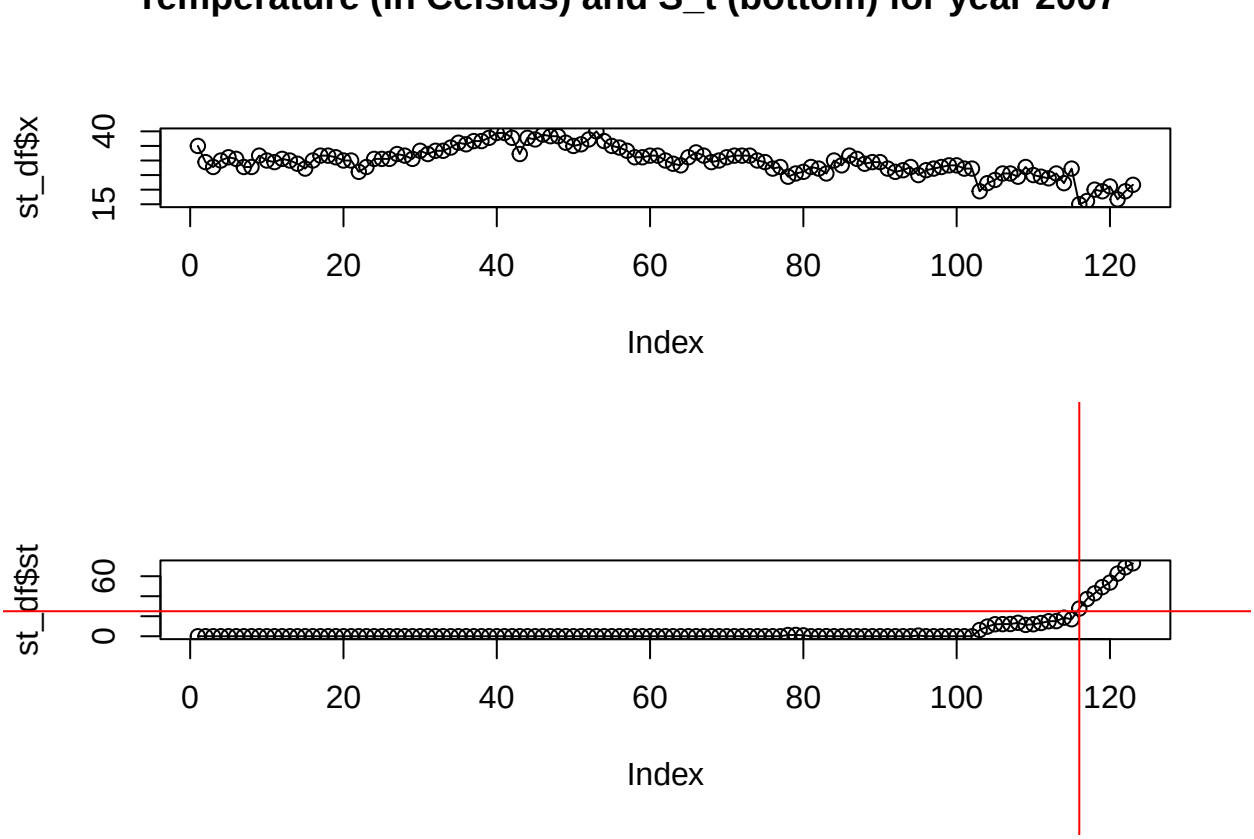

Temperature (in Celsius) and S_t (bottom) for year 2002
st_df$x
15
35
0
20
40
60
80
100
120
Index
st_df$st
0
80
0
20
40
60
80
100
120
Index

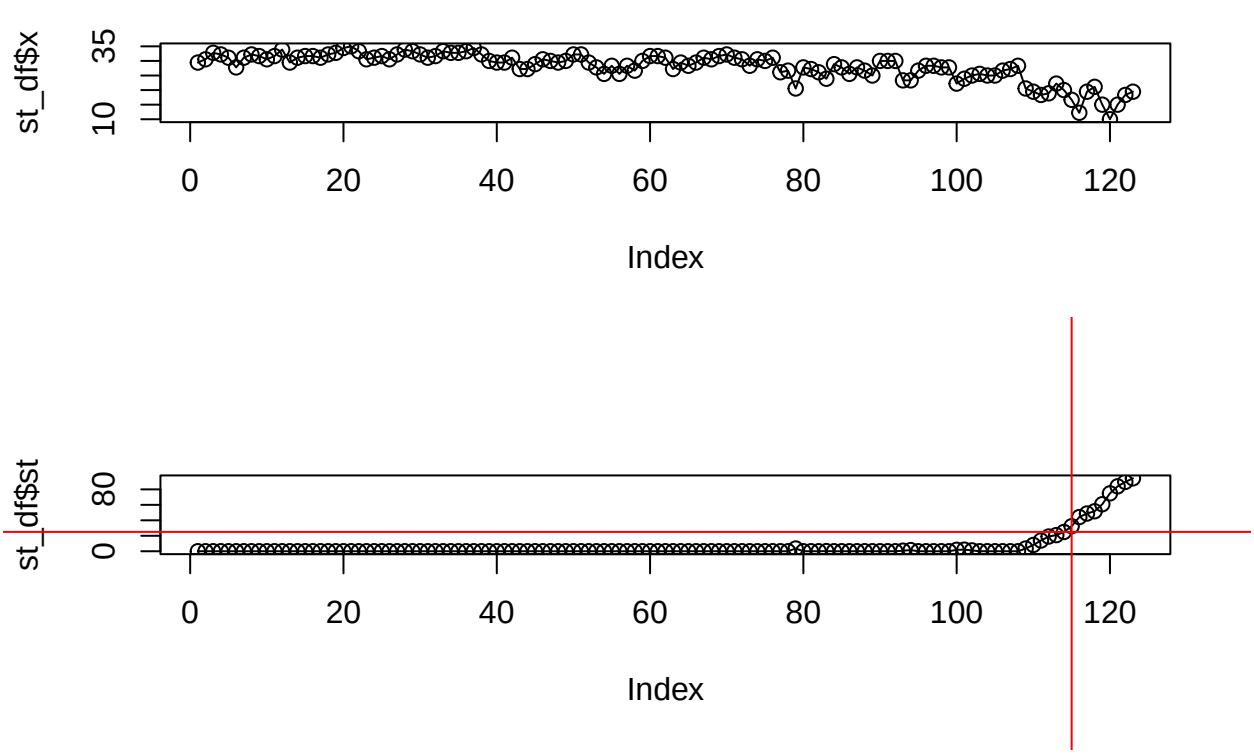

Temperature (in Celsius) and S_t (bottom) for year 2003
st_df$x
15
0
20
40
60
80
100
120
Index
st_df$st
0
30
0
20
40
60
80
100
120
Index

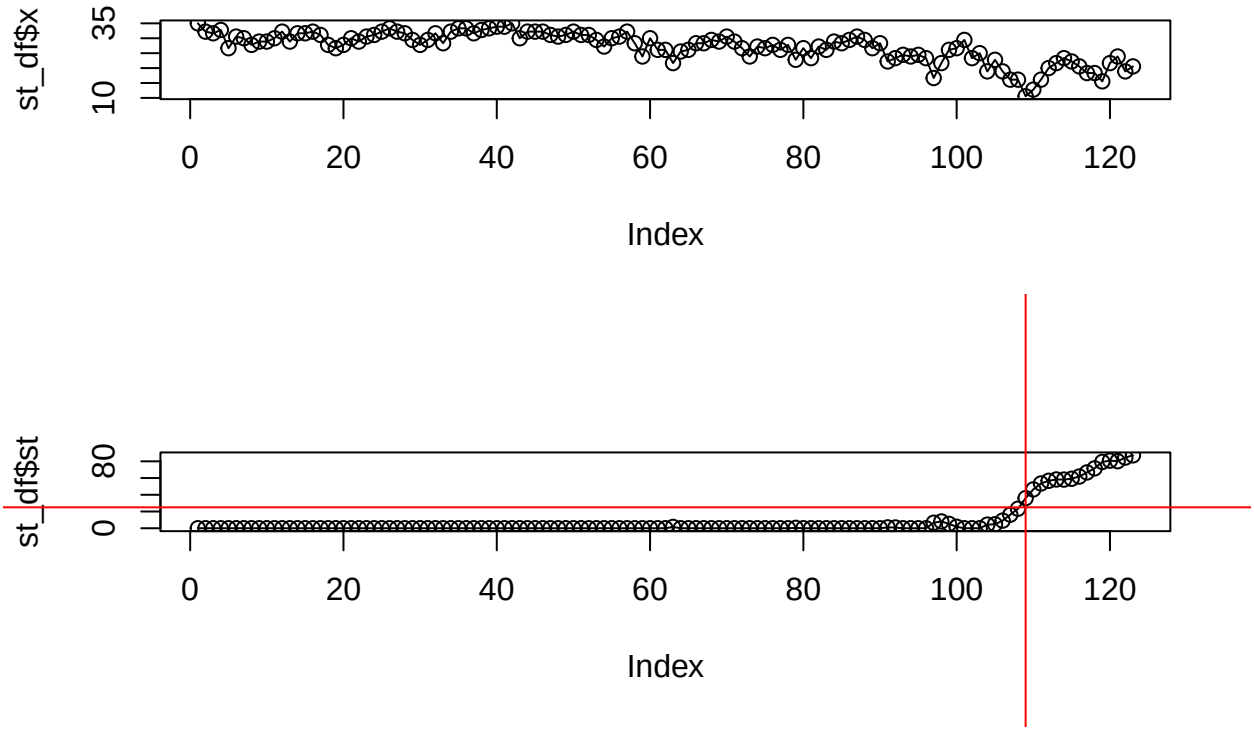

Temperature (in Celsius) and S_t (bottom) for year 2004
st_df$x
20
0
20
40
60
80
100
120
Index
st_df$st
0
30
0
20
40
60
80
100
120
Index

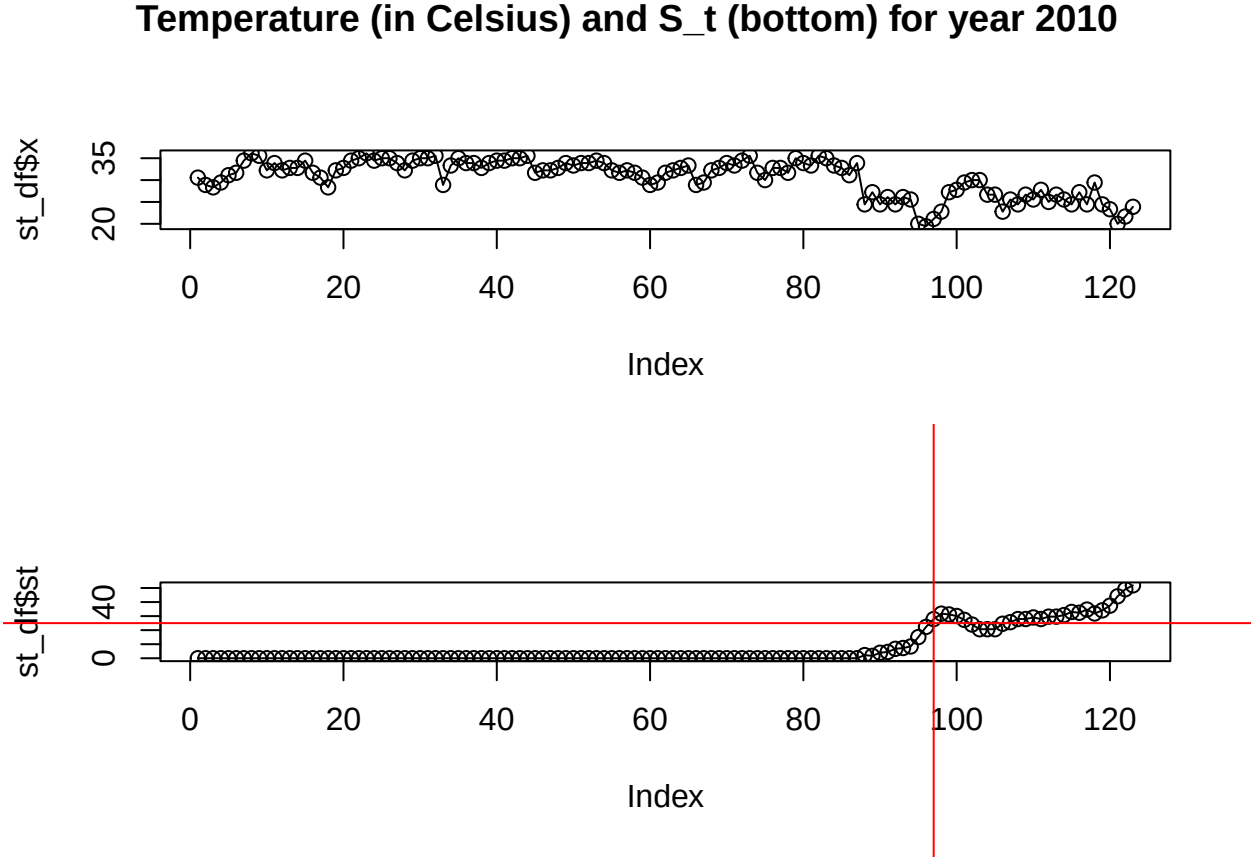

Temperature (in Celsius) and S_t (bottom) for year 2005
st_df$x
15
35
0
20
40
60
80
100
120
Index
st_df$st
0
60
0
20
40
60
80
100
120
Index

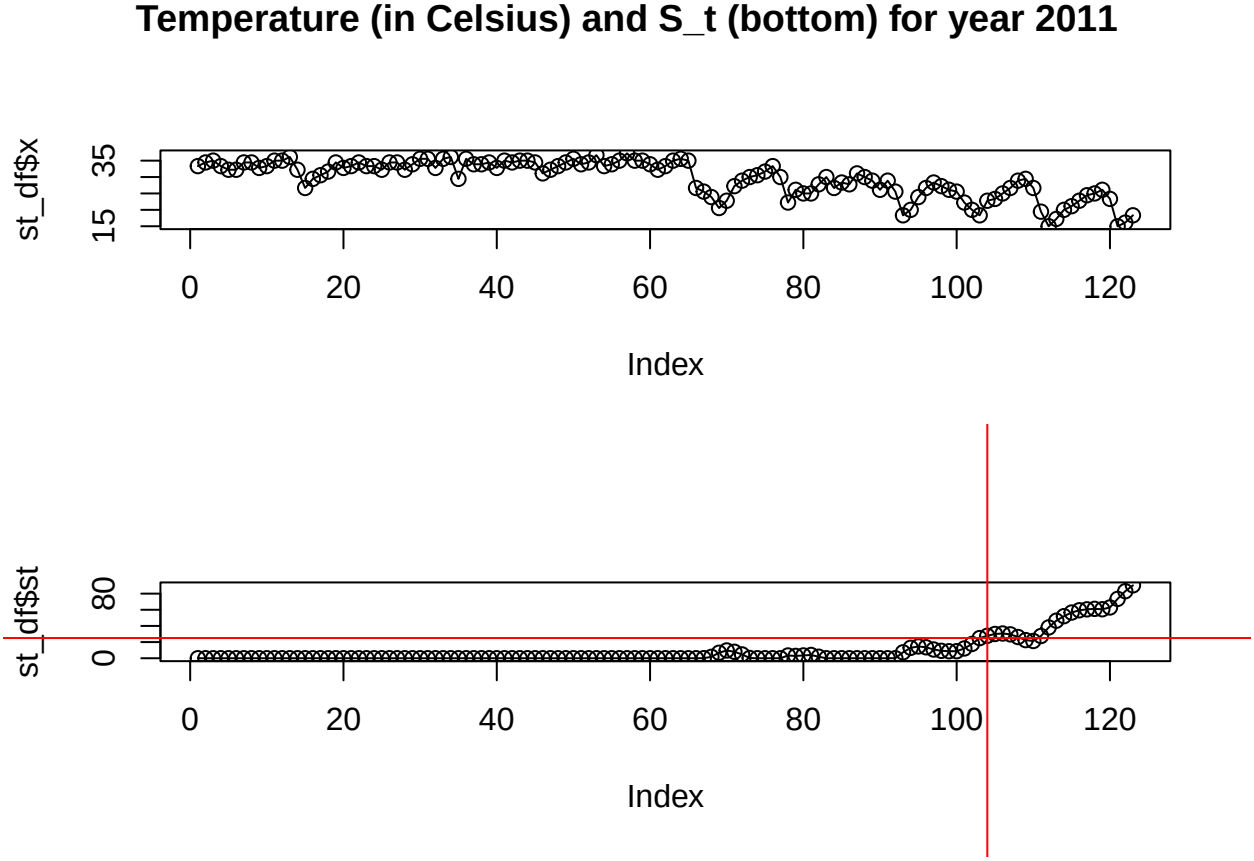

Temperature (in Celsius) and S_t (bottom) for year 2006
st_df$x
15
0
20
40
60
80
100
120
Index
st_df$st
0
80
0
20
40
60
80
100
120
Index

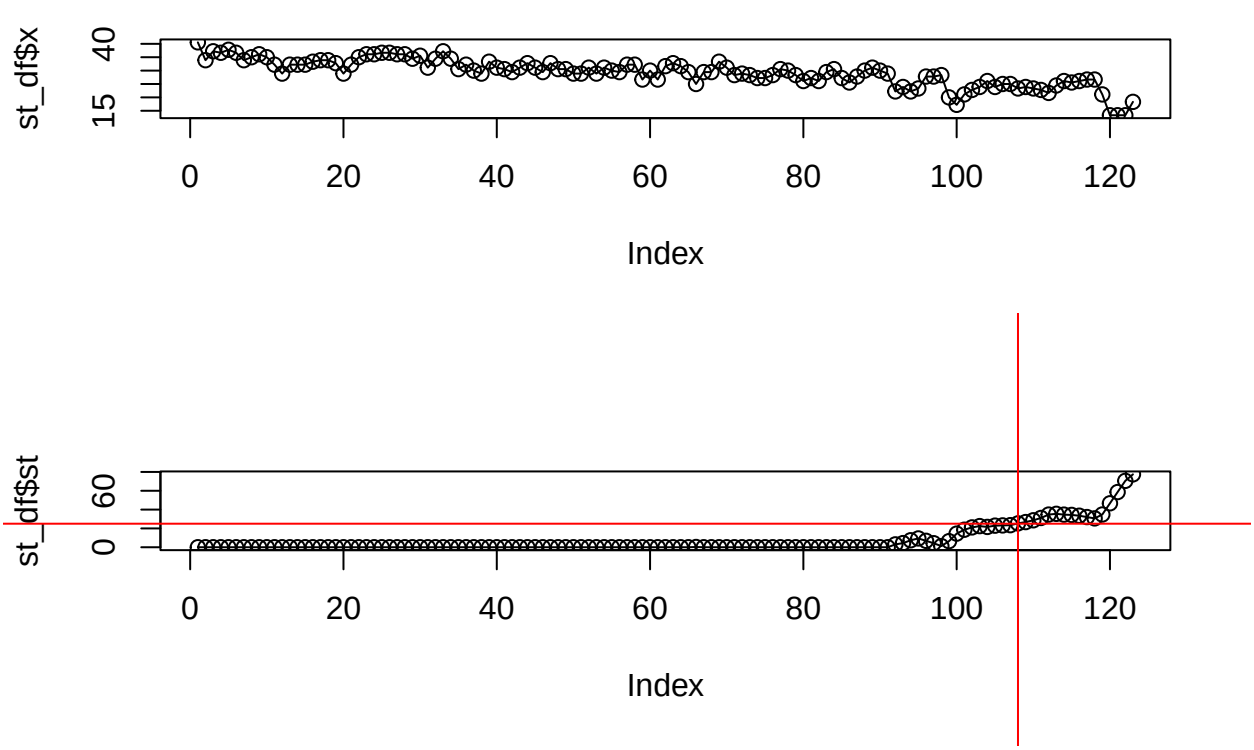

Temperature (in Celsius) and S_t (bottom) for year 2007
st_df$x
15
40
0
20
40
60
80
100
120
Index
st_df$st
0
60
0
20
40
60
80
100
120
Index

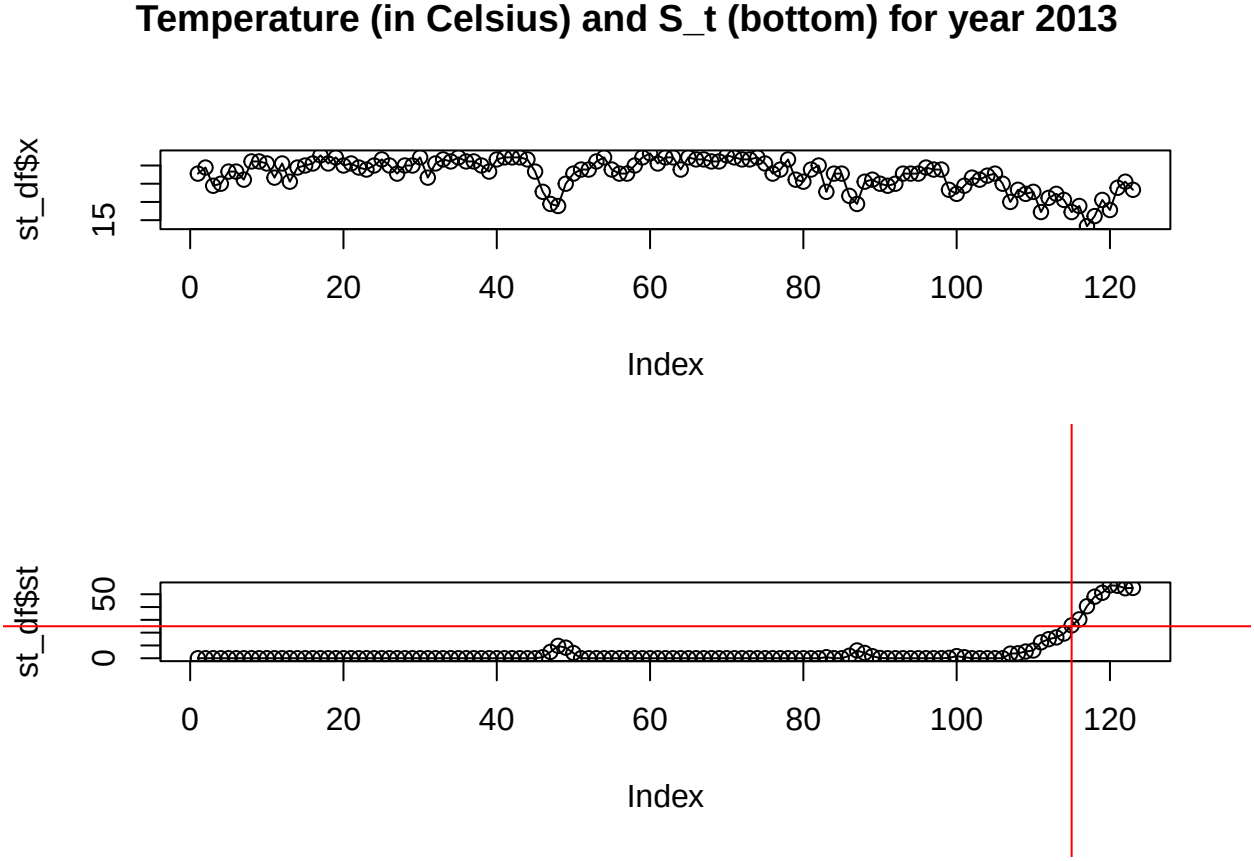

Temperature (in Celsius) and S_t (bottom) for year 2008
st_df$x
10
35
0
20
40
60
80
100
120
Index
st_df$st
0
80
0
20
40
60
80
100
120
Index

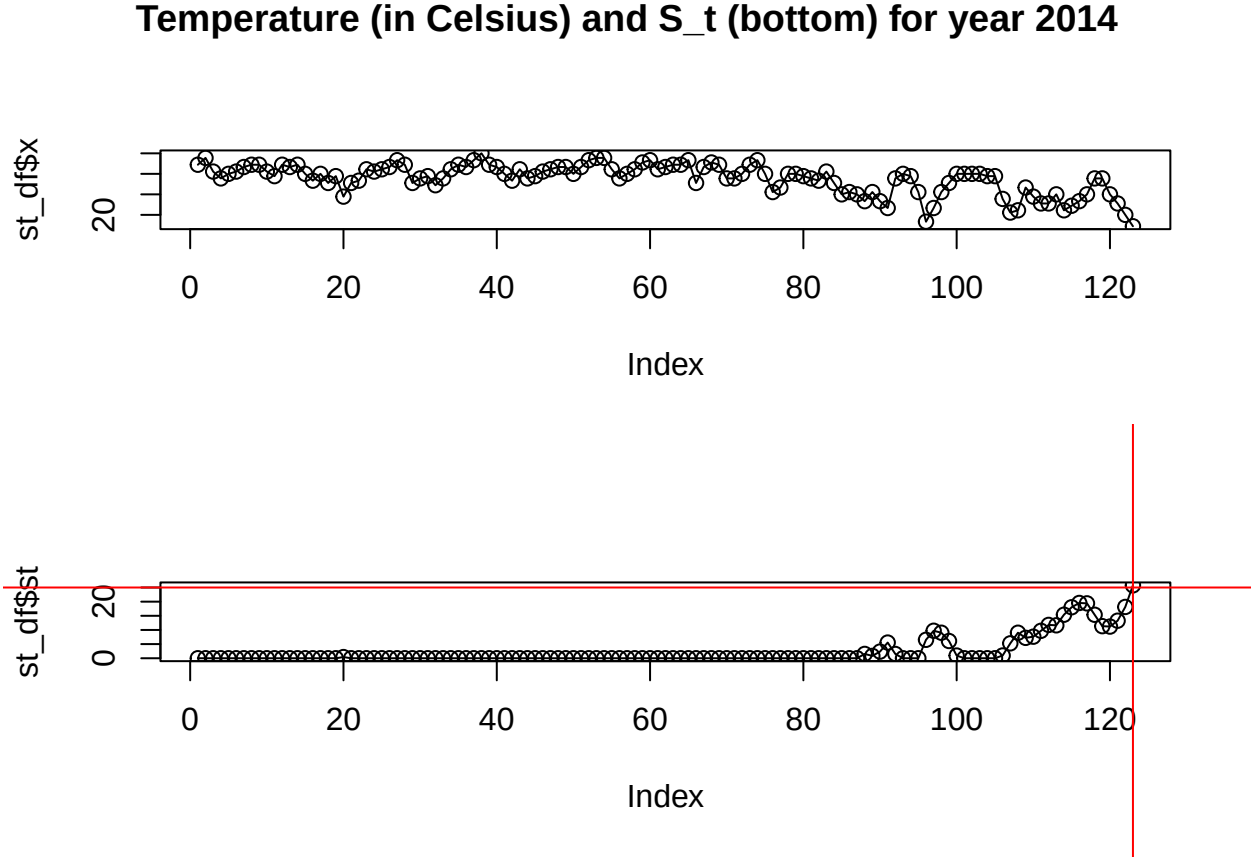

Temperature (in Celsius) and S_t (bottom) for year 2009
st_df$x
10
35
0
20
40
60
80
100
120
Index
st_df$st
0
80
0
20
40
60
80
100
120
Index

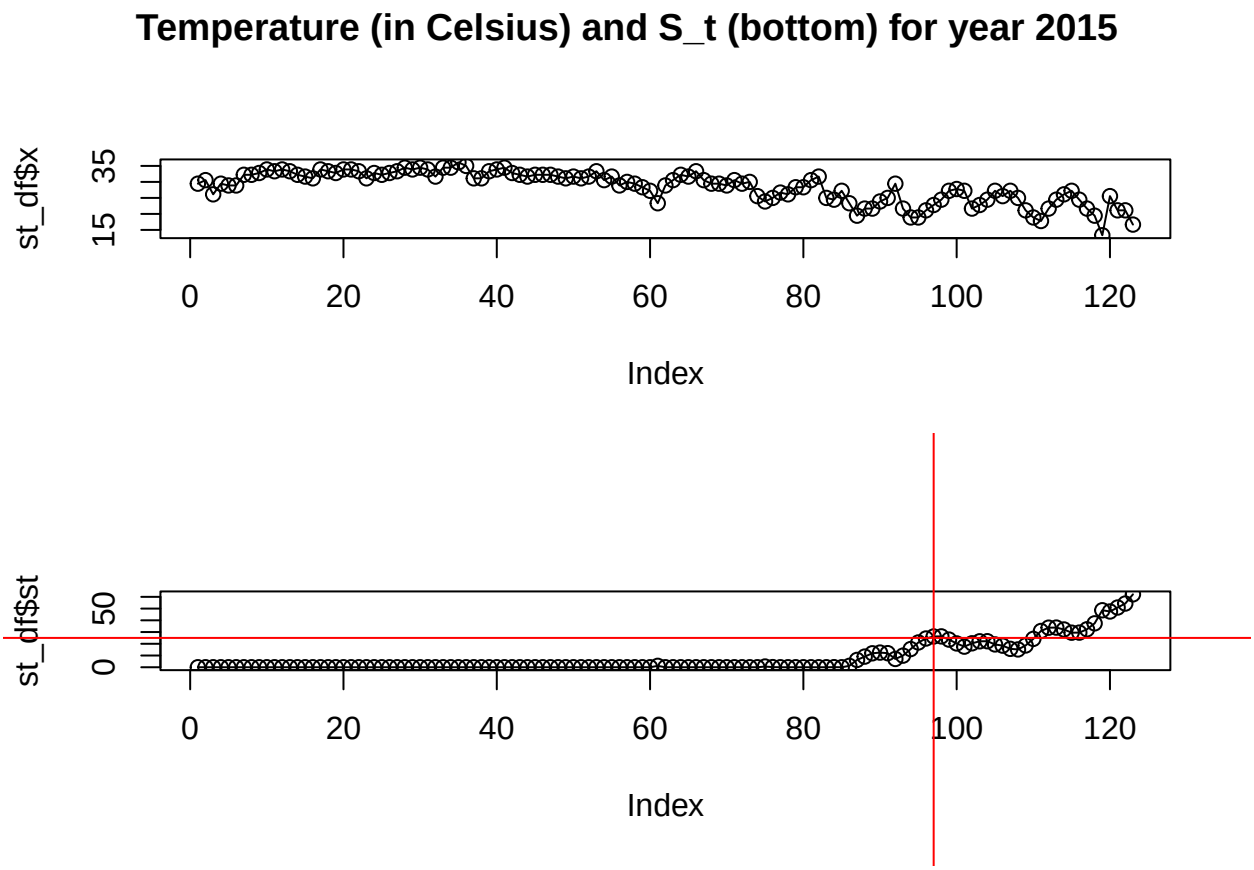

Temperature (in Celsius) and S_t (bottom) for year 2010
st_df$x
20
35
0
20
40
60
80
100
120
Index
st_df$st
0
40
0
20
40
60
80
100
120
Index


Temperature (in Celsius) and S_t (bottom) for year 2011
st_df$x
15
35
0
20
40
60
80
100
120
Index
st_df$st
0
80
0
20
40
60
80
100
120
Index


Temperature (in Celsius) and S_t (bottom) for year 2012
st_df$x
15
40
0
20
40
60
80
100
120
Index
st_df$st
0
60
0
20
40
60
80
100
120
Index


Temperature (in Celsius) and S_t (bottom) for year 2013
st_df$x
15
0
20
40
60
80
100
120
Index
st_df$st
50
0
0
20
40
60
80
100
120
Index


Temperature (in Celsius) and S_t (bottom) for year 2014
st_df$x
20
0
20
40
60
80
100
120
Index
st_df$st
0
20
0
20
40
60
80
100
120
Index



Analyse the effect of Model over the years

```
change_df <- as.data.frame(do.call(rbind, winter_start_days))
colnames(change_df) = c("year","winter_start","winter_length")
change_df["year_bucket"] = "a.1996_2000"
change_df[change_df$year>=2000,"year_bucket"] = "b.2000_2004"
change_df[change_df$year>=2004,"year_bucket"] = "c.2004_2008"
change_df[change_df$year>=2008,"year_bucket"] = "d.2008_2012"
change_df[change_df$year>=2012,"year_bucket"] = "e.2012_2016"
change_df
```

```
##    year winter_start winter_length year_bucket
## 1  1996           98            25 a.1996_2000
## 2  1997          110            13 a.1996_2000
## 3  1998          116             7 a.1996_2000
## 4  1999          105            18 a.1996_2000
## 5  2000          101            22 b.2000_2004
## 6  2001          111            12 b.2000_2004
## 7  2002          108            15 b.2000_2004
## 8  2003          120             3 b.2000_2004
## 9  2004          108            15 c.2004_2008
## 10 2005          117             6 c.2004_2008
## 11 2006          108            15 c.2004_2008
## 12 2007          116             7 c.2004_2008
```

```
## 13 2008          115             8 d.2008_2012
## 14 2009          109            14 d.2008_2012
## 15 2010           97            26 d.2008_2012
## 16 2011          104            19 d.2008_2012
## 17 2012          108            15 e.2012_2016
## 18 2013          115             8 e.2012_2016
## 19 2014          123             0 e.2012_2016
## 20 2015           97            26 e.2012_2016
```

Part 2: Has Atlanta become warmer over the years ?