Exponential Smoothing Exercise with CUSUM

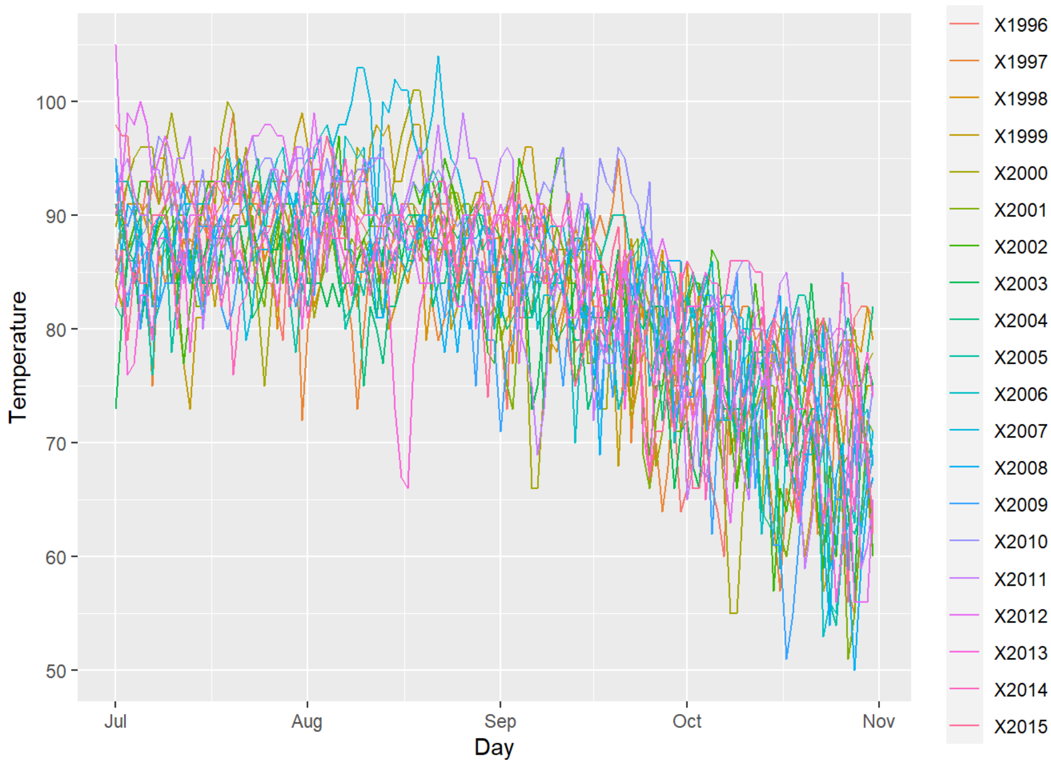
Question

Using the 20 years of daily high temperature data for Atlanta (July through October) from Question 6.2 (file temps.txt), build and use an exponential smoothing model to help make a judgment of whether the unofficial end of summer has gotten later over the 20 years. (Part of the point of this assignment is for you to think about how you might use exponential smoothing to answer this question. Feel free to combine it with other models if you'd like to. There's certainly more than one reasonable approach.) Note: in R, you can use either HoltWinters (simpler to use) or the smooth package's es function (harder to use, but more general). If you use es, the Holt-Winters model uses model="AAM" in the function call (the first and second constants are used "A"dditively, and the third (seasonality) is used "M"ultiplicatively; the documentation doesn't make that clear).

Solution

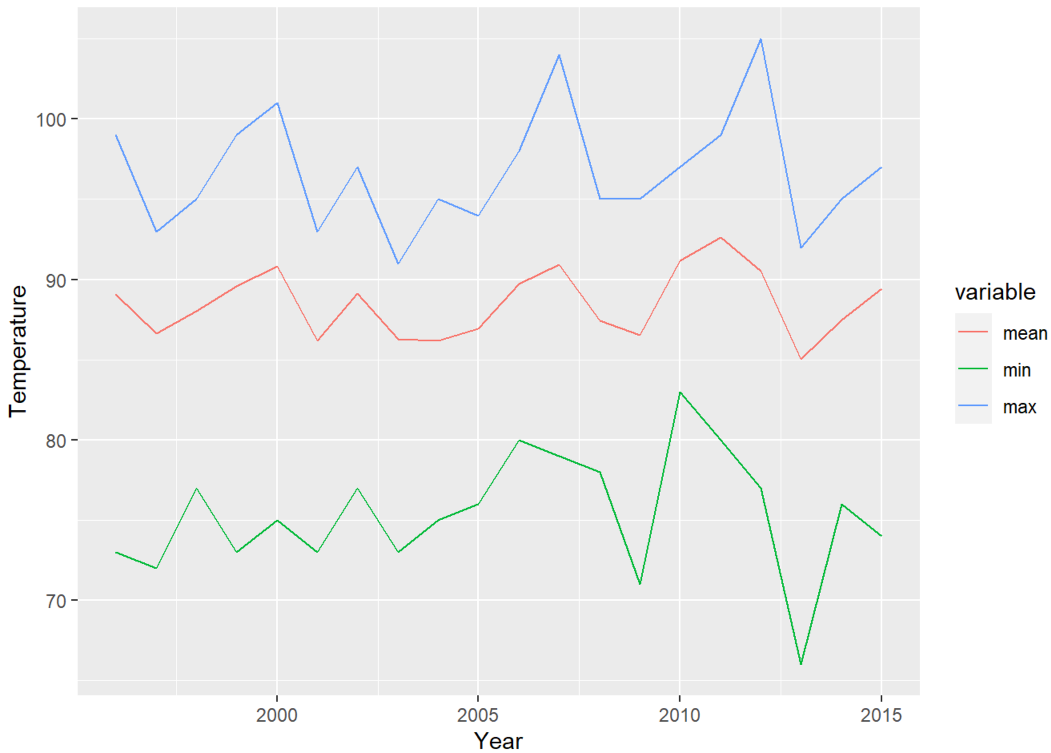
Exploratory Data Analysis

First, I do Exploratory Data Analysis to get an idea of the data we are working with. I want to plot all years on the same chart so I can see if there is a larger trend. I do this using the the reshape and dyplyr packages.



By first inspection, I can see that there is a general trend throughout the summer, with lowest temperatures in November and a peak in temperatures around August. There is an incredible amount of noise in this chart.

I am also curious if there is a trend through the years of the man, min, or max temperatures. I created a chart to look at this visually. I chose to end summer at September 4, which is after labor. This way fall temperatures wouldn't creep into the data



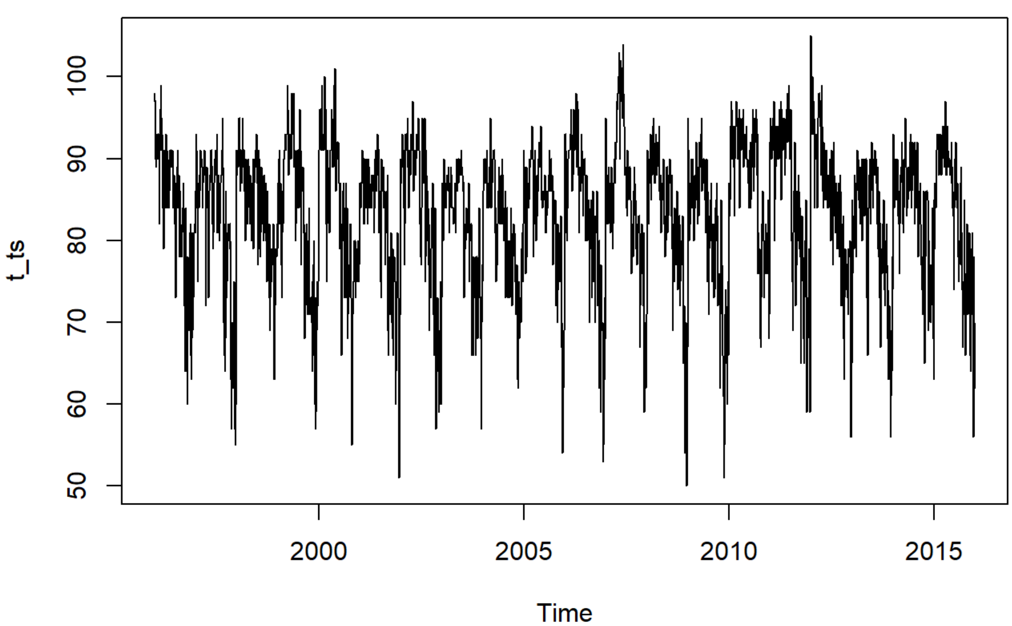
This chart is also inconclusive. The mean seems fairly consistent across the years with a small increase in 2010.

Exponential Smoothing

I then beginning exponential smoothing. I do this because of the amount of noise in the original set. It will be easier to draw conclusions about the end of summer if the data is more clear.

I first have to create a timeseries model from the data we have. I begin by creating a vector from the years and temperatures. I transform this into a time series data with frequency 123, which is the number of days in the season we are looking at (July 1 - October 31).

I plot the data to have a visual inspection of what the time series data gave me.



From here, I see the plot of the temperatures over time. Multiplicative seasonal data would show higher highs and lower lowers as the time increased, but this does not seem to be the case. The highs and lows appear to be fairly constant , so I will use an additive seasonal model moving forward.

Holt-Winters Smoothing

Holt-Winters Default

Next, we use the timeseries data to build the Holt_winters model for exponential smoothing. I chose additive for the seasonal data, as described above.

```
## alpha =  0.615003
## beta =  0
## gamma =  0.5495256
```

```
## NULL
```

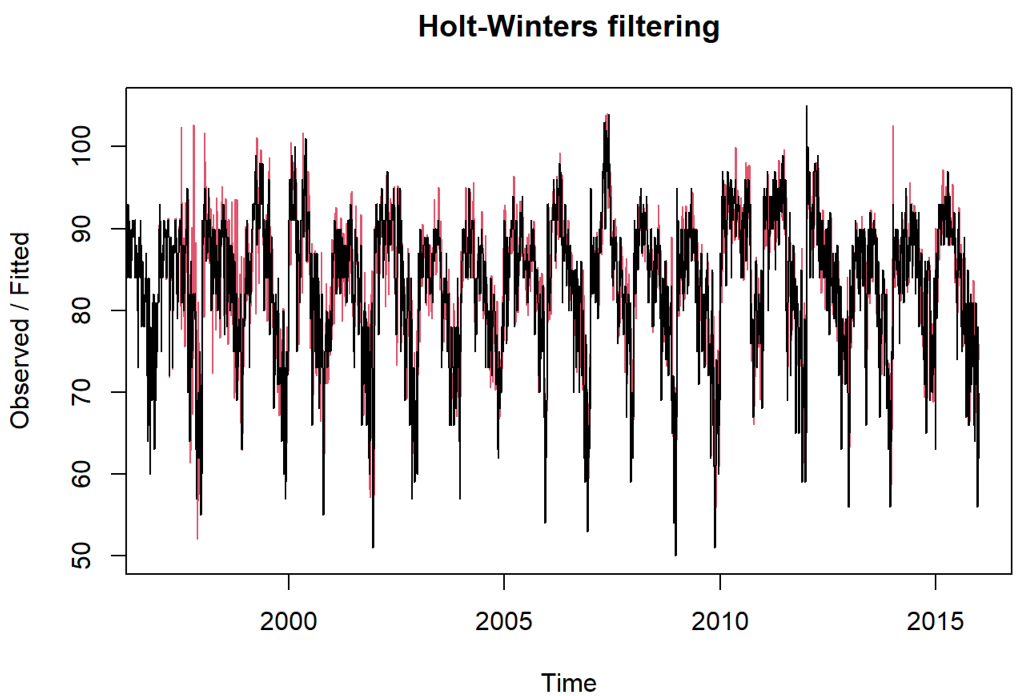
```
## Time Series:
## Start = c(1997, 1)
## End = c(1997, 6)
## Frequency = 123
##              xhat    level        trend   season
## 1997.000 87.23653 82.87739 -0.004362918 1.052653
## 1997.008 90.42182 82.15059 -0.004362918 1.100742
## 1997.016 92.99734 81.91055 -0.004362918 1.135413
## 1997.024 90.94030 81.90763 -0.004362918 1.110338
## 1997.033 83.99917 81.93634 -0.004362918 1.025231
## 1997.041 84.04496 81.93247 -0.004362918 1.025838
```

```
## Time Series:
## Start = c(2015, 118)
## End = c(2015, 123)
## Frequency = 123
##              xhat    level        trend    season
## 2015.951 76.54551 87.81303 -0.004362918 0.8717307
## 2015.959 69.70436 81.07435 -0.004362918 0.8598048
## 2015.967 57.02909 71.26750 -0.004362918 0.8002607
## 2015.976 72.14646 87.37935 -0.004362918 0.8257107
## 2015.984 73.89293 85.77627 -0.004362918 0.8615051
## 2015.992 75.83100 82.99285 -0.004362918 0.9137532
```

Based on the best fitted:

- Alpha is 0.615, suggesting there is randomness in this model to be considered.
- Beta is 0, which suggests there is no trend. From looking at the timeseries plot, this makes sense.
- Gamma is 0.549, suggesting ties to seasonality, neither explicitly favoring older nor more recent data.

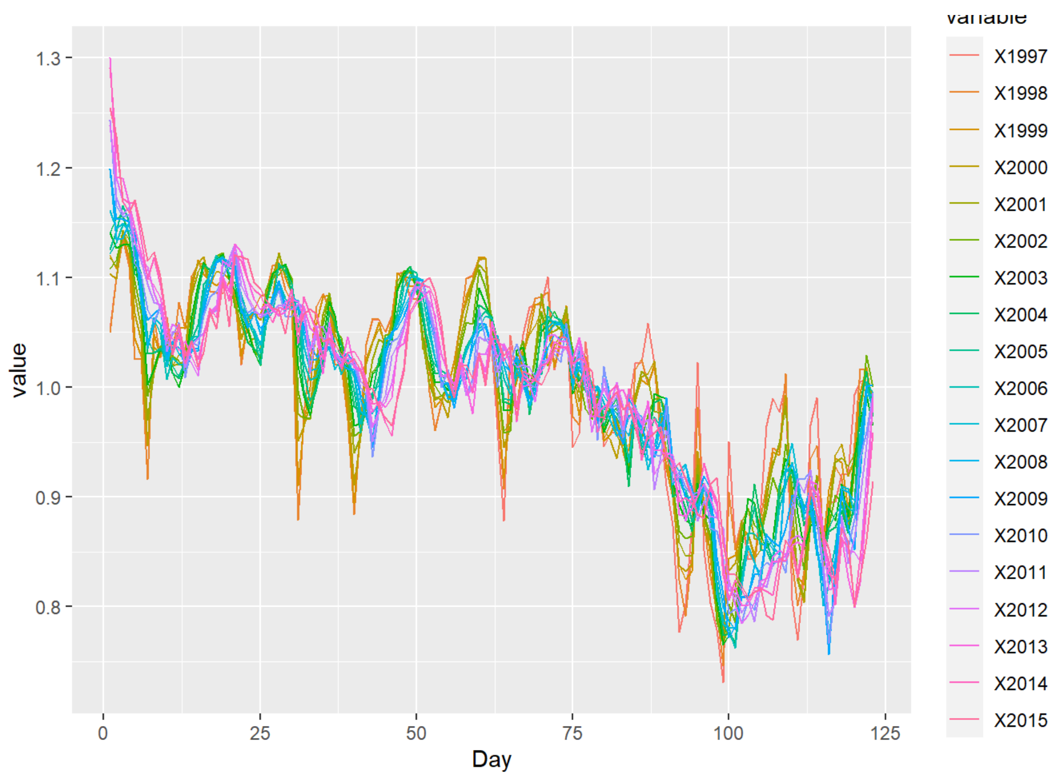
Next I plot the data. The red is the smoothed data, which fits fairly well while smoothing the sharper peaks and valleys



I want to make a matrix of just the season factors. The seasonal factors indicate how far from the baseline the data point is, according to that point across all the seasons. If the SF is greater than 1, then the temperature is higher than the baseline, while if SF is less than one, the temperature is below the baseline.

I created a function to plot the data from the Holt Winters sets.

I then run the function based on the Seasonal Factor set created earlier



As the days progress, our seasonal factor goes from it’s highest in the beginning to its lowest around day 100. This could be due to the smoothing over the course of the season.

Holt-Winters, Gamma = 0.3

I am curious what would happen to the data if I were to change the gamma to put more emphasis on the historical data. I will run another sequence that changes the gamma to 0.3. By doing this, I am saying that the history has been stable before the introduction of the data. I am quickly going to run through the code from before to generate a the graphs.

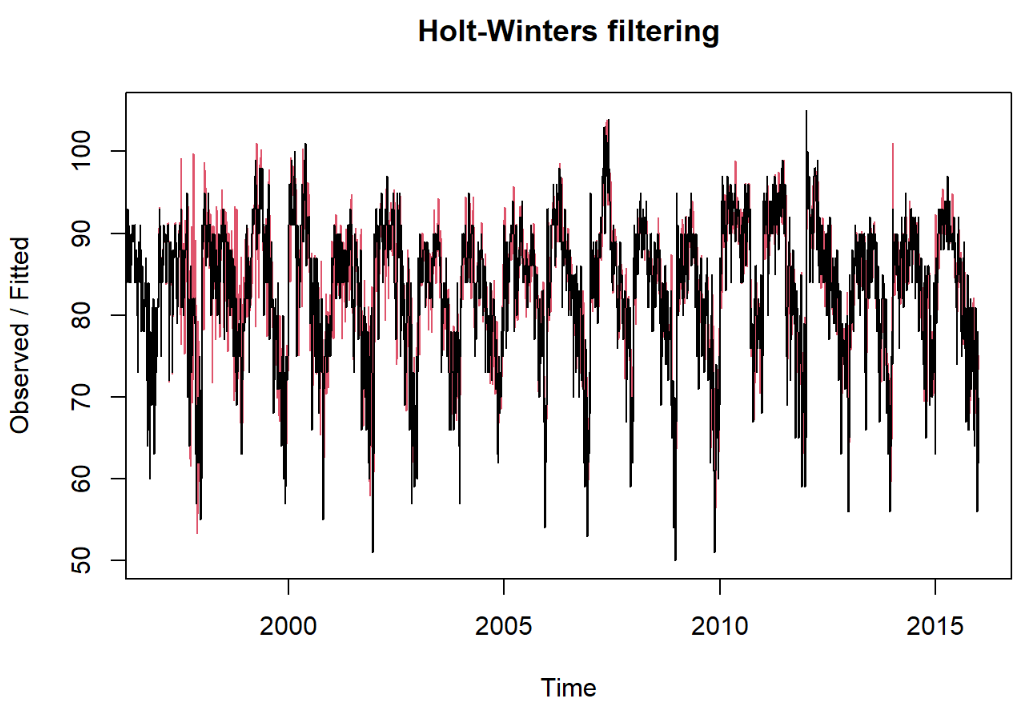
```
## alpha =  0.4646543
## beta =  0
## gamma =  0.3
```

```
## NULL
```

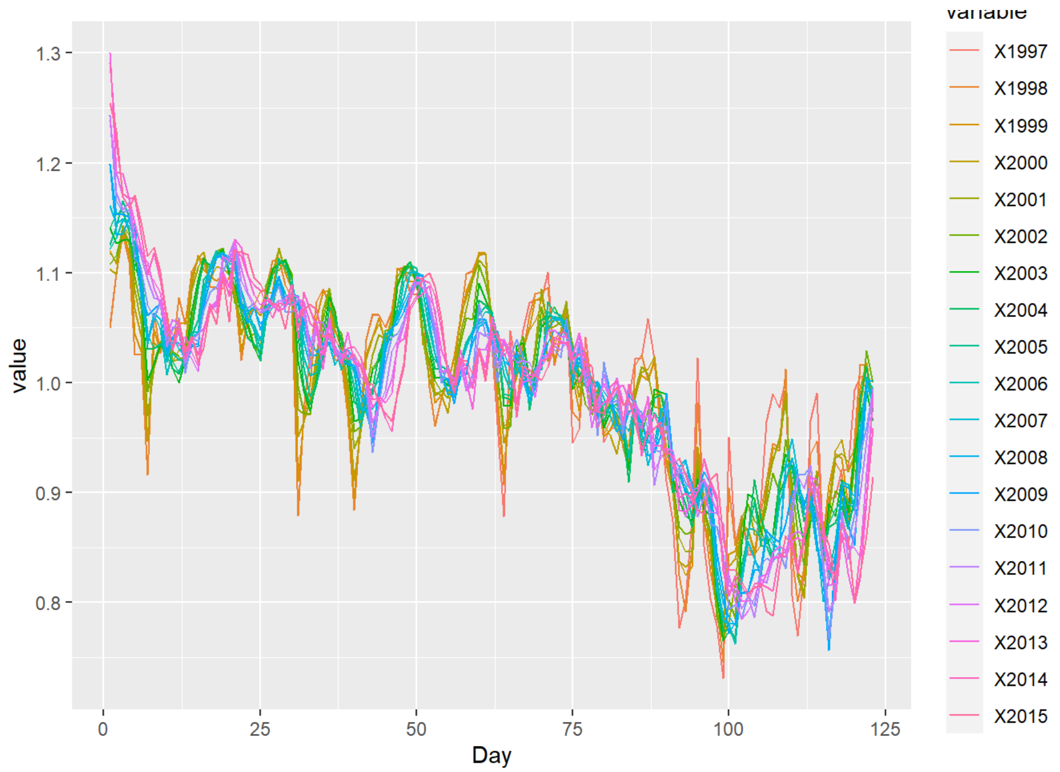
```
## Time Series:
## Start = c(1997, 1)
## End = c(1997, 6)
## Frequency = 123
##              xhat    level        trend   season
## 1997.000 87.23653 82.87739 -0.004362918 1.052653
## 1997.008 90.61622 82.32721 -0.004362918 1.100742
## 1997.016 93.17011 82.06272 -0.004362918 1.135413
## 1997.024 91.03036 81.98874 -0.004362918 1.110338
## 1997.033 84.03540 81.97167 -0.004362918 1.025231
## 1997.041 84.06424 81.95127 -0.004362918 1.025838
```

```
## Time Series:
## Start = c(2015, 118)
## End = c(2015, 123)
## Frequency = 123
##              xhat    level        trend    season
## 2015.951 77.58818 87.52388 -0.004362918 0.8865243
## 2015.959 71.28863 81.96993 -0.004362918 0.8697388
## 2015.967 60.73867 73.79768 -0.004362918 0.8230917
## 2015.976 70.84302 83.53774 -0.004362918 0.8480805
## 2015.984 73.30975 83.07149 -0.004362918 0.8825362
## 2015.992 75.07363 81.32455 -0.004362918 0.9231857
```

By decreasing the gamma, the alpha also decreased, suggesting more randomness in the model.



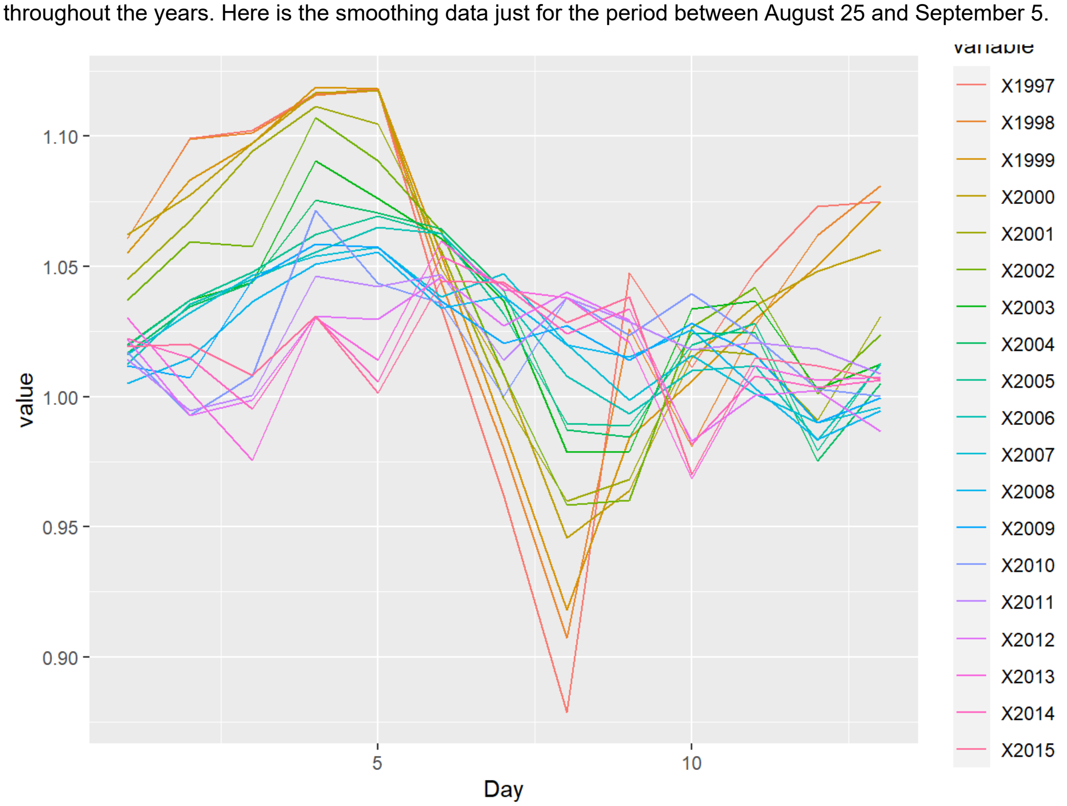
In this graph, we can see the predictive values are lagging behind the actuals.



Chose a Set

In the end, I see no advantage to using the gamma function that ties more strongly to history. The assumption it makes that historical data is similar to the year 1997 is too large. There is not enough data here to make such an assumption, as we are only given 20 years. I am going to step with the "Default" data, which gives equal wait to all data.

After the larger smoothing, what I really want to focus on is the end of the summer, as we are trying to ascertain if the end of summer has shifted throughout the years. Here is the smoothing data just for the period between August 25 and September 5.



This graph is hard to read. The data I am looking for is the SF during this time period chronologically. If the SF is above one, that means there is a general increase from the base data, while an SF below one suggests a decrease in temperature. A positive SF, thus, would suggest a longer summer, as the later summer days are warmer than they were historically.

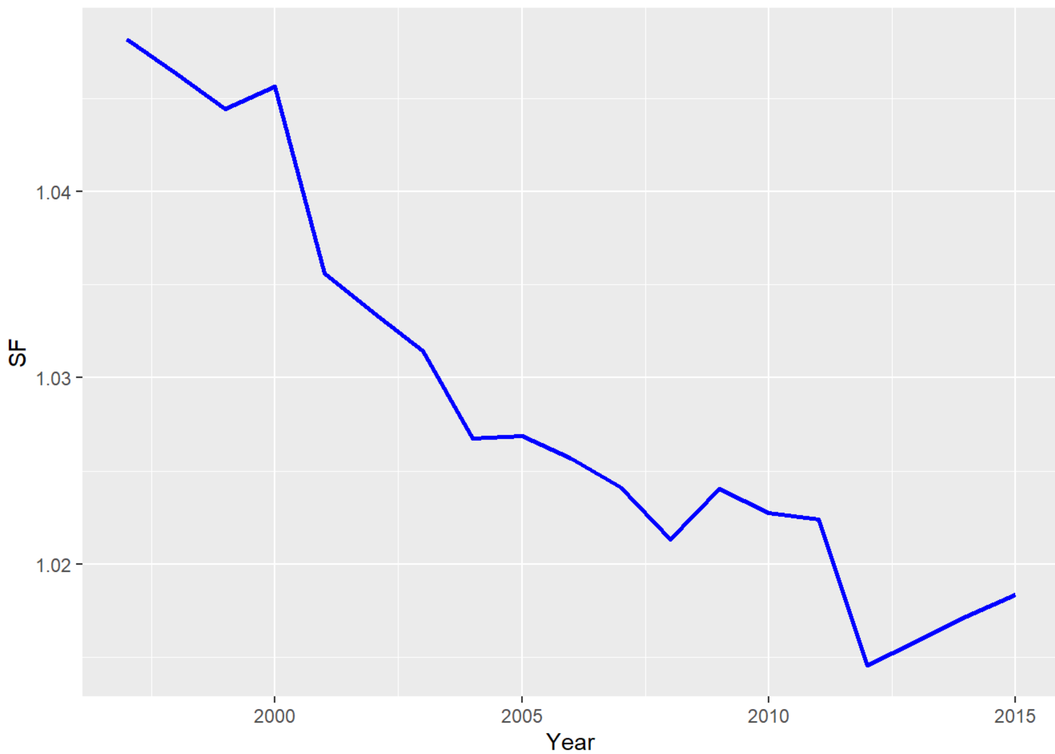
To move forward I will be using the CUSUM method with this data

CUSUM

I am using the CUSUM to see if there is a change in the seasonal factor at the end of the summer. If there is a positive change, this would suggest hotter days at the end of summer.

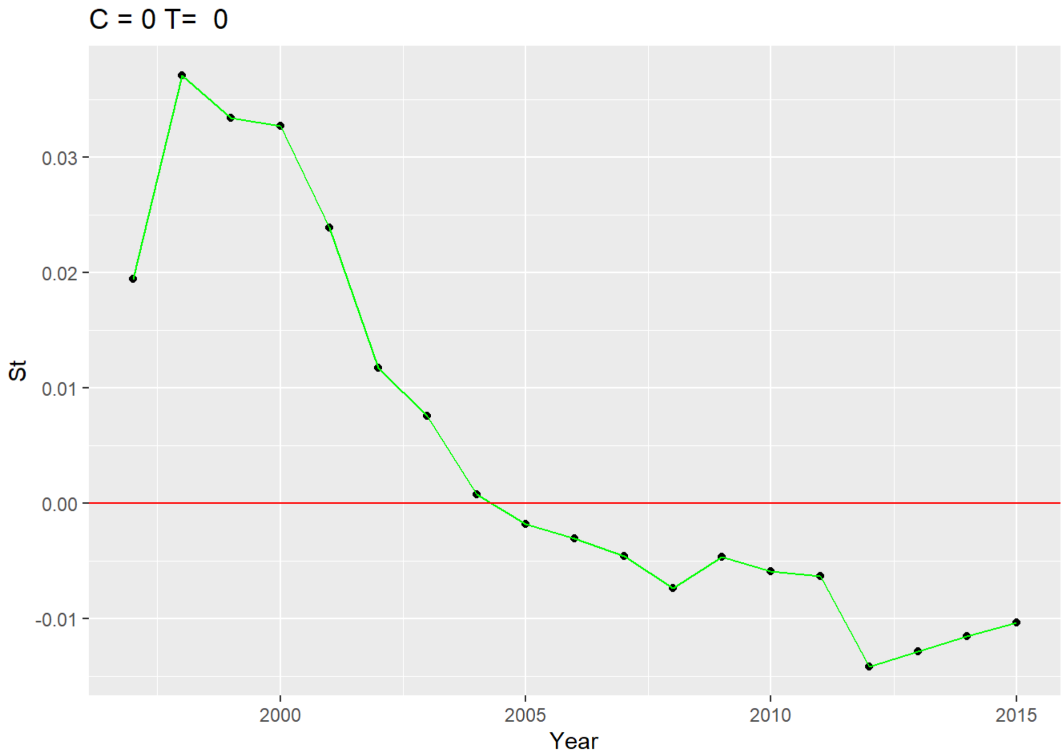
To begin, I am going to find the mean of the SF across only this period, as this is the only relevant period for End of Summer data analysis.

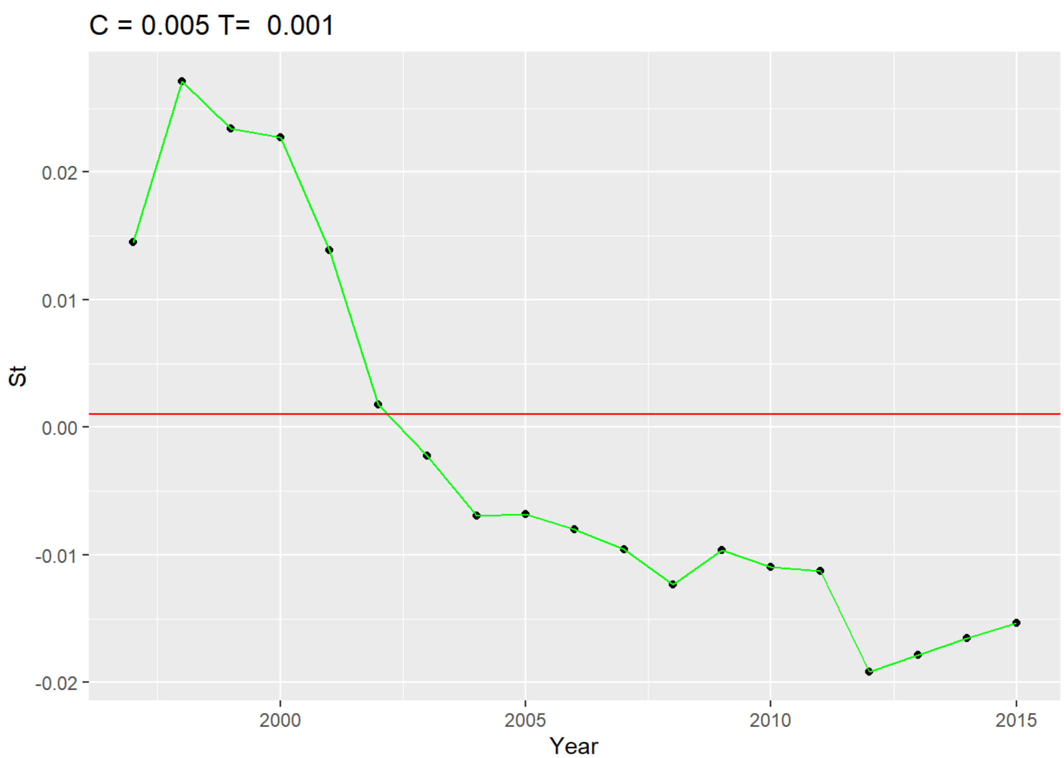
```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



This quick visualization of the data suggests that the SF is decreasing through the years, but let's see in CUSUM if that's actually true. I use the small data set with just the last days of summer.

I created a function for the last homework that let me quickly compute and plot the CUSUM.





I made two charts with different C and T values. The C value is fairly low because the data is sensitive. The lower T value detects change more quickly, which is needed when there is such a small period here.

Conclusion

Based on the Exponential Smoothing and subsequent CUSUM method, I detect a change in about 2003 that shifts the SF to a lower value. This suggests that after 2003, that “final summer” week gets cooler, not warmer. The S_t never increases, after going below the threshold, thus our SF will never increase.

Thus, the end of summer is NOT getting longer. If anything, the data suggests the opposite.