Question 7.1
Describe a situation or problem from your job, everyday life, current events, etc., for which exponential smoothing would be appropriate. What data would you need? Would you expect the value of α (the first smoothing parameter) to be closer to 0 or 1, and why?

Answer:
One situation where exponential smoothing would be appropriate would be in sales forecasting. Business often compensate sales teams on a bonus model where they receive their bonus if they are able to hit a sales quota. This quota can be modeled off of historical sales. In the case of sales teams who work in large customer sales or departments where deal sizes are large but occur infrequently, I would expect my alpha to be closer to 0 since the exact date of a sale might fluctuate but period trends would nonetheless be reliable. For example, your department might expect 15 distinct deals over 1million dollars to occur in a quarter, but the exact date of these deals in the quarter might fluctuate. Thus we wouldn't want to weight the previous days data too heavily compared to the previous model overall.

Question 7.2:
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

Answer:
First, I loaded my data into R and transformed the data into a time-series vector so it was in the correct format for the HoltWinter package

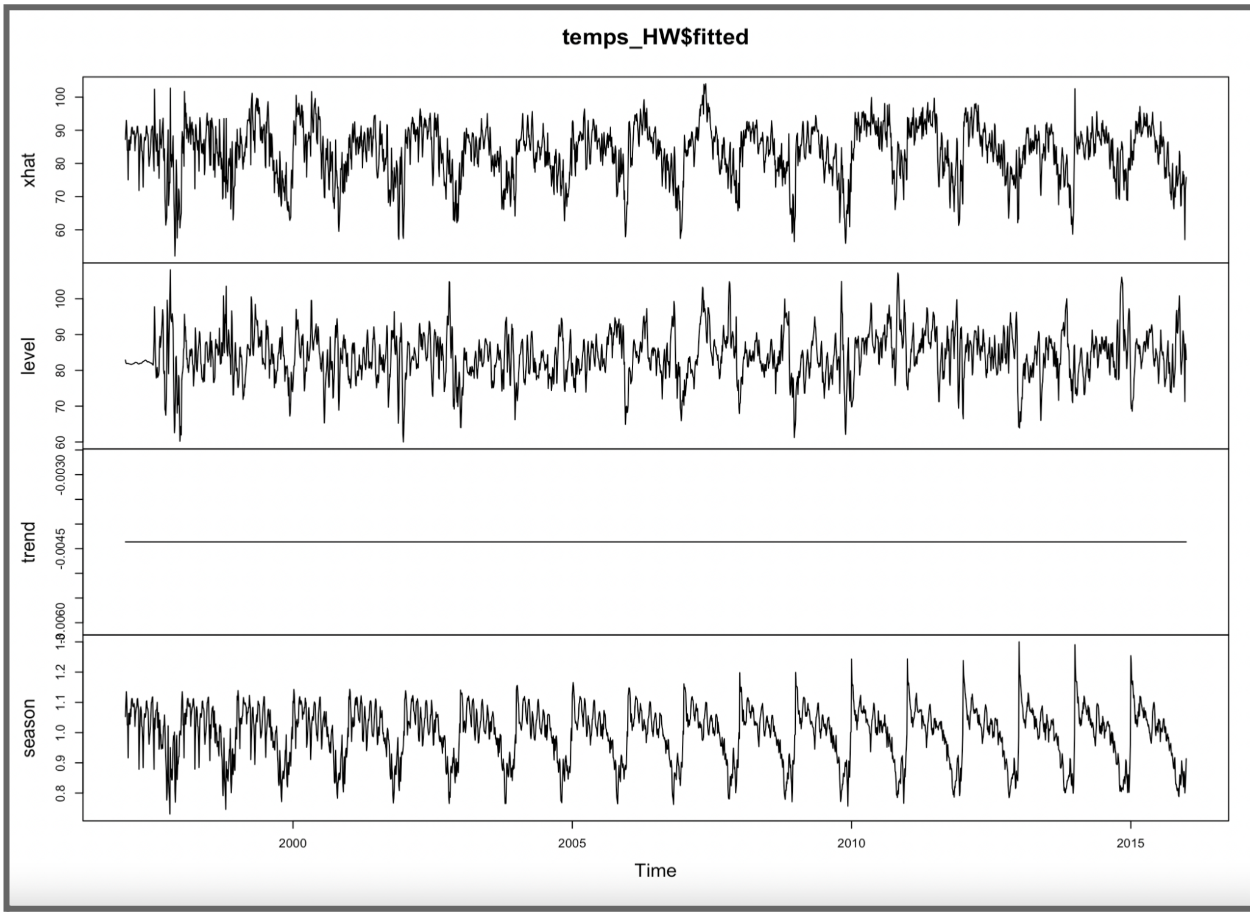
```
#convert to time series
temps_vec <- as.vector(unlist(temps[,2:21]))
temps_vec
plot(temps_vec)
temps_ts <- ts(temps_vec, start = 1996, frequency=123)
temps_ts
plot(temps_ts)
```

Then, I applied the HoltWinters function to the time series vector and plotted the fitted values

```
#holtwinters
temps_HW <- HoltWinters(temps_ts, alpha = NULL, beta = NULL, gamma = NULL, seasonal = "multiplicative")
temps_HW
plot(temps_HW)

#plot our fitted values
plot(temps_HW$fitted)

temps_HW_sf <- matrix(temps_HW$fitted[,4], nrow = 123)
temps_HW_smoothed <- matrix(temps_HW$fitted[,1],nrow=123)
```



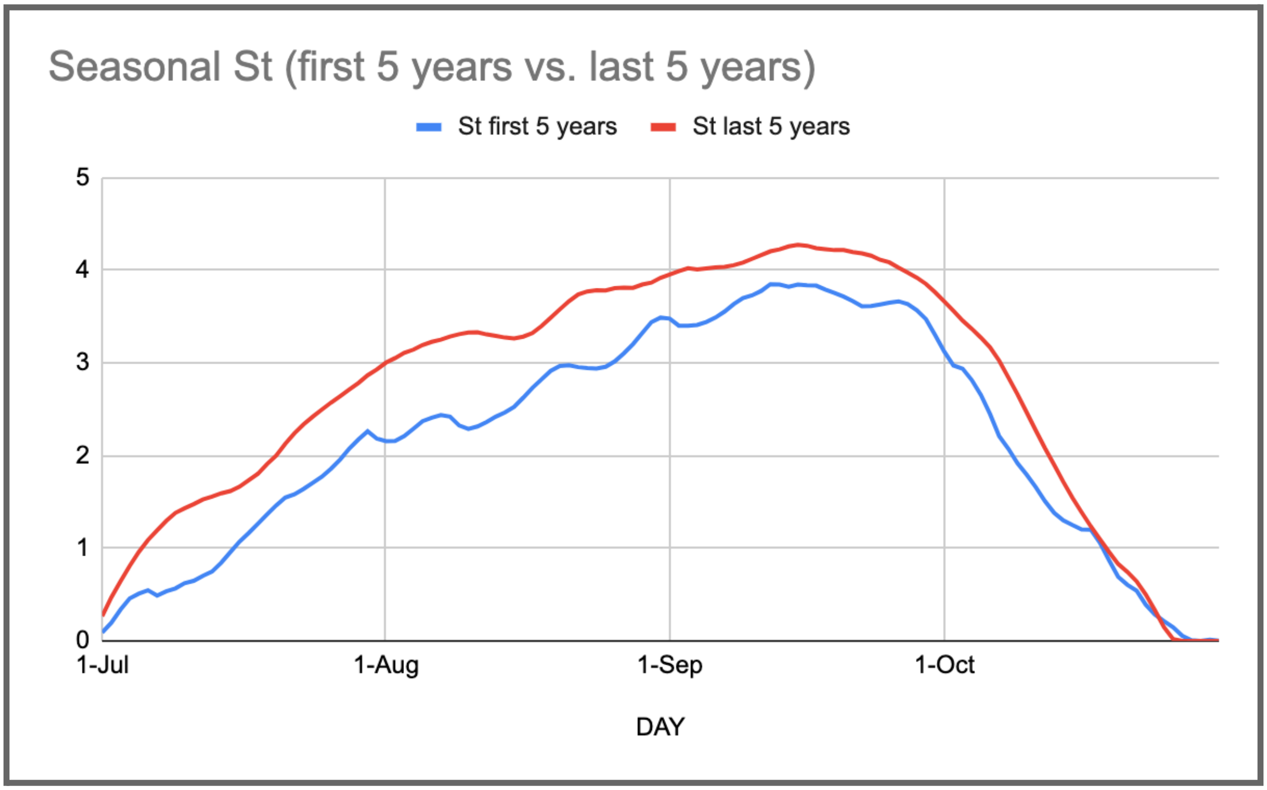
Already from this basic plot of fitted values, I had a theory that the seasonal fitted value would be useful to this analysis. I noticed that the volatility of the seasonal trend began to increase as we moved closer to 2015, especially noticing that the peak was increasing. This led me to think that the model had to apply a greater seasonal factor to the data in more recent years.

In order to validate this assumption, I created a CUSUM control chart of the seasonal fitted value. For my analysis, I compared the first 5 years (1997-2001) in my data set to the last five years (2011-2015). My theory here was that 5 year periods would minimize the risk of an outlier year while still capturing the overall behavior of two distinct periods. There is a 10 year gap between my two periods, and i am hoping that this would be enough time to show that behavior has changed (if it has).

For Xt in my first five year period (1997-2001), I used the daily seasonal fitted value for the five years. Thus, for July 1, my Xt was the average of the seasonal value on July 1 between 1997 and 2001. Similarly, for Xt in my last five year period (2011-2015), i took the daily seasonal value for the last five years; For july 1, I used the average seasonal value on July 1 from 2011 - 2015.

For u, I used the average seasonal value for the whole first five year period (1997-2001). Thus, it would be the average of the seasonal values from July 1 to Oct 31 from 1997 - 2001. My theory here was I wanted to test how my last five year period (2011-2015) compared to my first five. Keeping U the same allows me to see if my control chart shows any distinct differences. The vertex (maximum value) of the parabola will show when the summer period ends. The St value will show the relative difference in the scale of the seasonal factor applied between the two different periods.

Below is the output of the control chart:



From this chart, I am not noticing that the summer period is lasting longer as the vertex of both parabolas are relatively similar. It is noticeable that the last 5 years has a higher St throughout the period. This would mean that the model had to apply a stronger seasonal factor in later years than in earlier years, suggesting that temperatures were getting warmer in the summer period. However, the question the problem set posed was whether summer lasted longer, and the data is not overwhelmingly supporting this hypothesis (although there is some directional validity since the seasonal factor in the later period stays above the seasonal factor in the earlier period until mid October).