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?

Exponential smoothing technique could be applied to analyze and forecast the stock price. I aim to predict the daily adjusted closing prices of FAANG(Facebook, Amazon, Apple, Netflix and Alphabet (formerly known as Google)). In the experiment, we will use 5 years of historical prices for FAANG. We firstly need to separate the training(3 years data) and validation dataset(2 year). Optimal alpha could be calculated by minimizing RMSE. I expect the alpha would be closer to 1, since the stock price is highly fluctuated and affected by the confidence of the stock holders in the short term. Say there is a good or bad news about this company, the stock price is inclined to be hugely impacted by the observation data and its hard to predict from the trend and seasonality factor.

The model we stick with should be Holtwinter model which includes trend and seasonality.

Question 7.2

Exponential smoothing is a great method to deal with time series dataset with some randomness and seasonality, like temps data for this question.

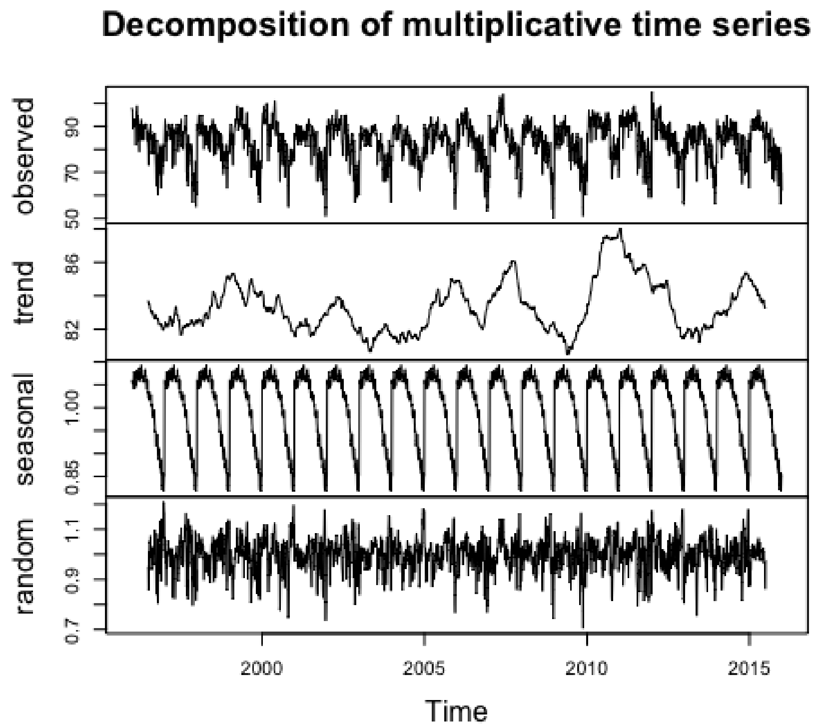
By leveraging the Holtwinter function to implement exponential smoothing on temps dataset, we could can get smoothed data which removed the noise and took seasonality factor into the consideration.

Once exported the smoothed data, we use CUSUM in excel to further investigate the unofficial summer end date and make a judgement based on the temperature of the change detected date over the 19 years(since we took 1996's data as baseline reference, there are only 19 years of temperature data available for CUSUM)

```
#Import the temps.txt

temps<- read.table('temps.txt',header=T)
temps_ve<-as.vector(unlist(temps[,2:21]))   #convert the data into vector format for ts function
temps_ts<-ts(temps_ve,start=1996,frequency = 123)

#plot
plot(decompose(temps_ts,type='multiplicative'))
```



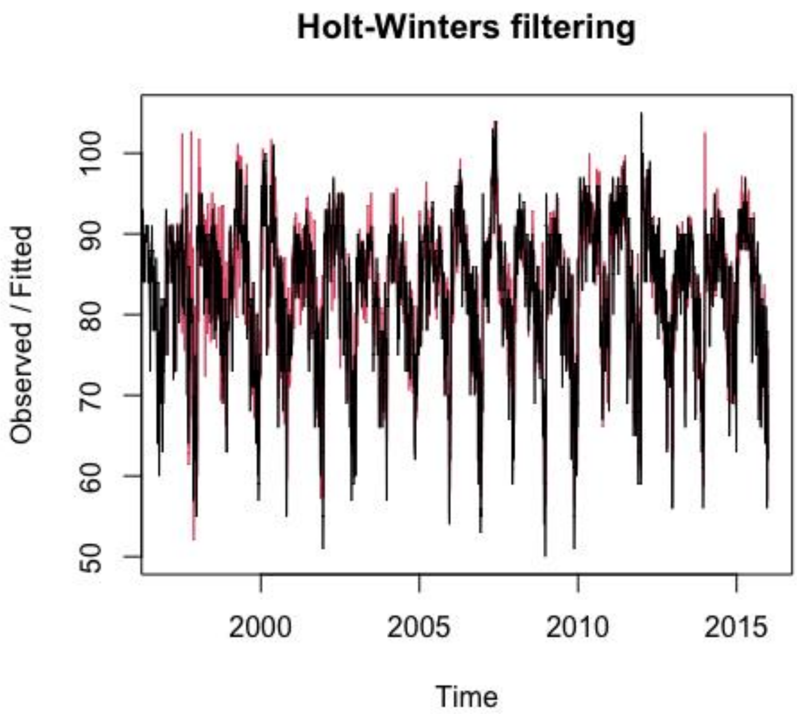
We can draw a brief conclusion that the temps data includes trend and seasonality. So the triple exponential smoothing is appropriate.

 # Use HoltWinter function (seasonal="multiplicative")

> temps_HWM<-HoltWinters(temps_ts,seasonal="multiplicative")

> plot(temps_HWM)

From the below graph, we could tell that the fitted value (RED) is becoming more smoothing as the time goes by, because more observations are fed into the model.



#check the model parameter (alpha, beta and gamma)

> temps_HWM
Holt-Winters exponential smoothing with trend and multiplicative seasonal component.

Call:
HoltWinters(x = temps_ts, seasonal = "multiplicative")

Smoothing parameters:
 alpha: 0.615003
 beta : 0
 gamma: 0.5495256

Alpha, beta and gamma are calculated by the Holt-Winters model itself through minimizing the mean square errors.

Check the fitted value for temps_HWM

```
> head(temps_HWM$fitted)
         xhat    level        trend   season
[1,] 87.23653 82.87739 -0.004362918 1.052653
```

[2,] 90.42182 82.15059 -0.004362918 1.100742
[3,] 92.99734 81.91055 -0.004362918 1.135413
[4,] 90.94030 81.90763 -0.004362918 1.110338
[5,] 83.99917 81.93634 -0.004362918 1.025231
[6,] 84.04496 81.93247 -0.004362918 1.025838

Xhat is the combination of the level and trend and seasonality, which is going to be considered as my "smoothed value"

#Export the xhat into excel

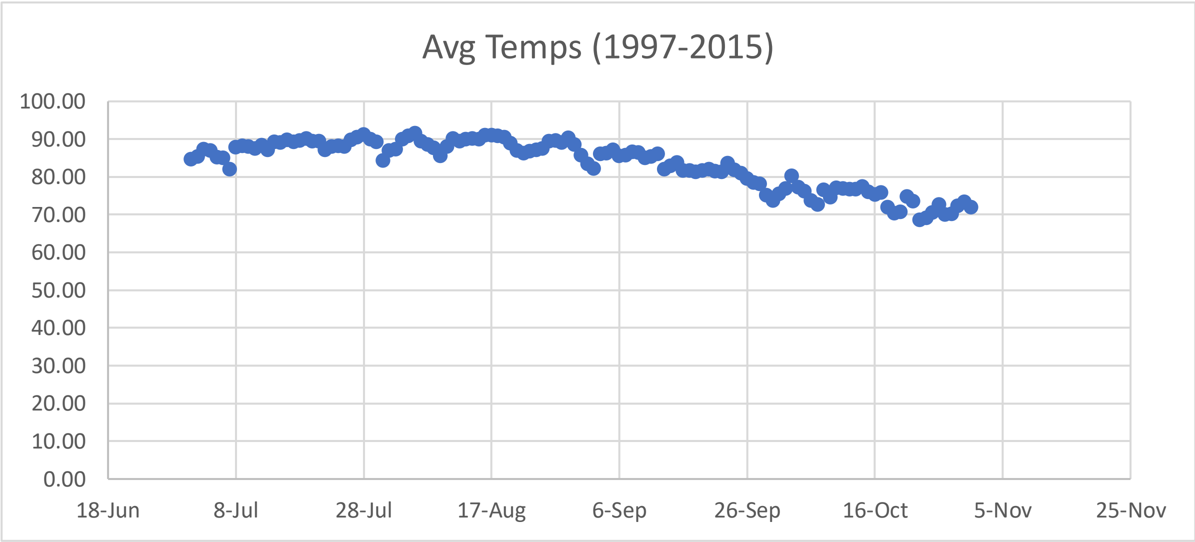
```
>temps_HWM_smooth<-matrix(temps_HWM$fitted[,1],nrow=123)
> write.csv(temps_HWM_smooth,"Smoothdata.csv",row.name=T)
```

Implement CUSUM in excel

I. I tried different C and T combination using CUSUM and came up with the following results.

Parameter	C	T	Summer end	Avg temp on detect data	standard deviation
1	0	35	07-Aug	88.61120127	4.235691997
2	5	35	26-Sep	79.46264099	6.712919784
3	1	35	17-Aug	90.93759186	5.846407757
4	2	35	09-Sep	86.42913745	6.451779609
5	3	35	17-Sep	81.69935631	4.685353574
6	4	35	23-Sep	83.6763016	5.088086498
7	6	35	29-Sep	75.13972108	5.153528443
8	2	25	15-Aug	89.95430607	5.486649392
9	3	25	26-Aug	89.42347429	4.72865707
10	4	25	17-Sep	81.69935631	4.685353574
11	5	25	22-Sep	81.36407128	6.65795145

As you can see in the table, the unofficial summer end date varies from Aug 07 to Sep 29. For better evaluation, i average the each day's temperature over 19 years and plot the change of average temperature from July 1 to Oct 31.



Georgia Tech

As we may noticed, there is a clear decline in temperature between Sep 06- Sep 30 in the plot graph. This interval should be when the unofficial summer end day is chosen from.

From the table, the parameter combinations in orange has their summer end day lying between Sep 06- Sep 30. After considering the optimal temperature and standard deviation, we choose C=4 and T= 35 as our optimal parameter for CUSUM.

By populating the optimal C and T, we can get the change detected day for 19 years (see below table)

Run CUSUM again on these 19 data points to detect if there is a change in the date of unofficial summer end.(C=3 T=20)

	Days until change detected	Change detected day	CUSUM
1997	91	29-Sep	0.00
1998	93	01-Oct	0.00
1999	87	25-Sep	0.00
2000	70	08-Sep	0.00
2001	91	29-Sep	12.95
2002	89	27-Sep	4.89
2003	94	02-Oct	0.00
2004	79	17-Sep	0.00
2005	101	09-Oct	3.95
2006	78	16-Sep	0.00
2007	106	14-Oct	4.95
2008	84	22-Sep	0.00
2009	82	20-Sep	0.00
2010	94	02-Oct	0.95
2011	72	10-Sep	0.00
2012	57	26-Aug	10.95
2013	93	01-Oct	36.89
2014	92	30-Sep	26.84
2015	80	18-Sep	17.79

There is a clear jump happening in 2013. Given that the conclusion is that unofficial summer end day is getting late since 2013.