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?

Study the trend of cryptocurrency is in my daily routine, I think exponential smoothing would be appropriate for predicting the future price fluctuations by using past historical time series data. I may choose to use past 3-5 years historical price data from bitcoin which is more stable from long run.
I also think the value of α should depends on certain scenario. As we all know, α is close to 1, there is more weigh assigned to recent past observation, whereas small values mean more of the history is taken into consideration. Since there is equation problem, we could use mean square error to decide. We can choose the best value for α so the value which results in the smallest MSE.

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.)

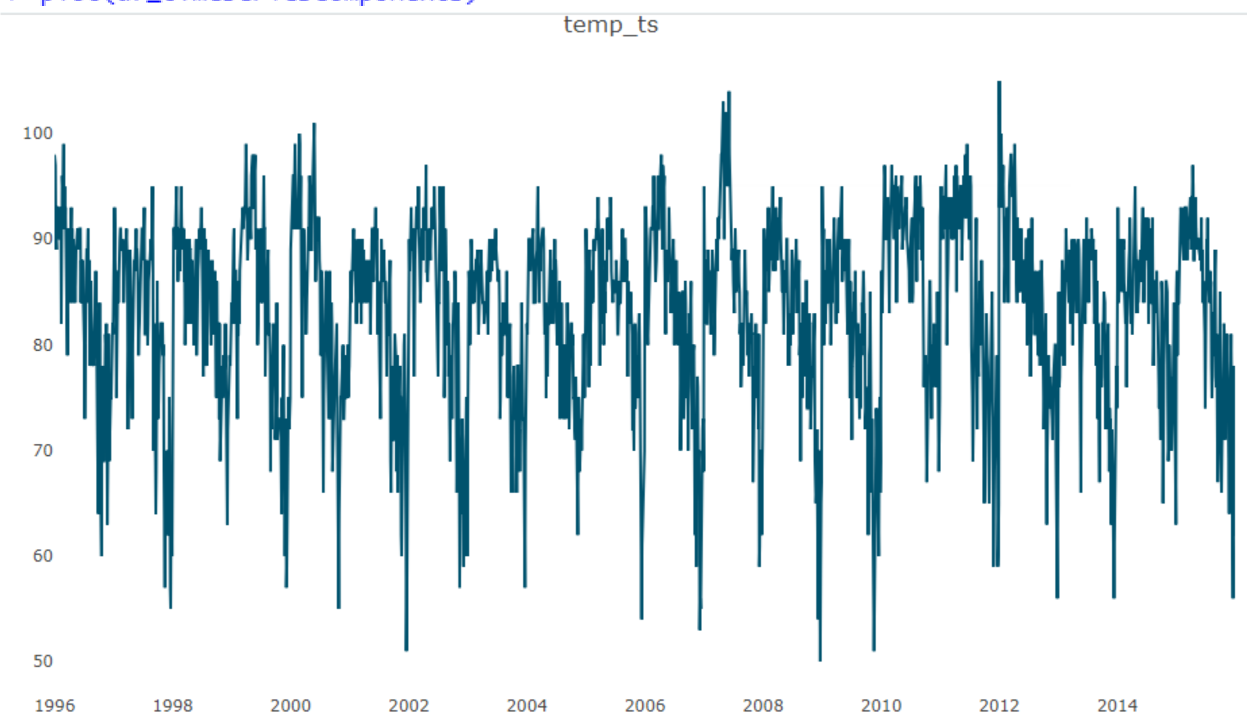
Note: in R, you can use either `HoltWinters` (simpler to use) or the `smooth` package's `es` function (harder to use, but more general). If you use `es`, the Holt-Winters model uses `model="AAM"` in the function call (the first and second constants are used "A"dditively, and the third (seasonality) is used "M"ultiplicatively; the documentation doesn't make that clear).

In conclusion, there is no significant results to support summer is ending later. More detailed analysis is required. However, we can use the CUSUM test from last week, combining the HoltWinters & CUSUM test we can conclude that there is no significant evidence suggest that the end of summer has gotten later over the 20 years. Detailed code is listed below.

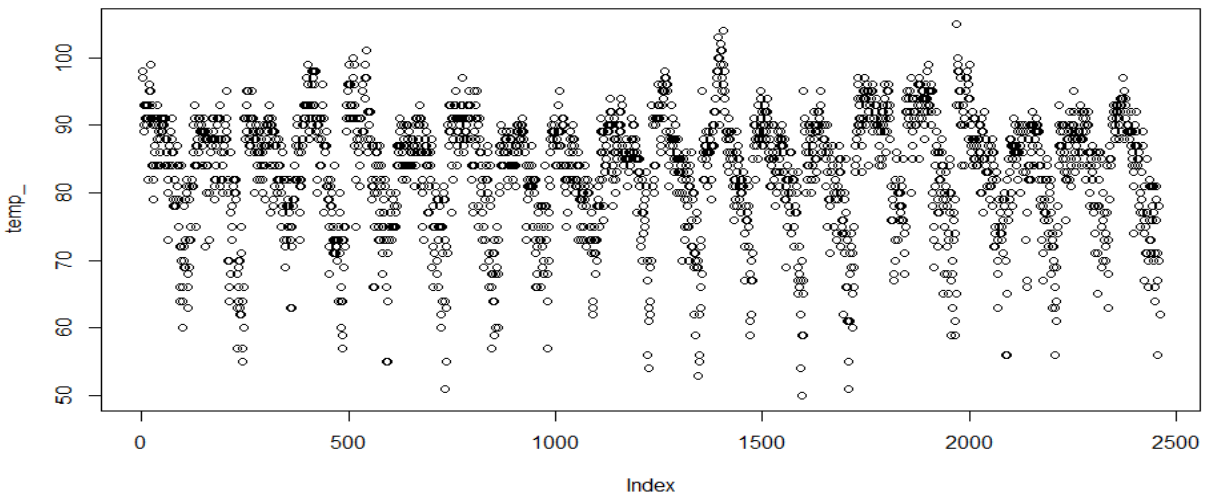
```
> # import data
> df <- read.delim("C:/Users/zhuoxun.yang001/Desktop/fff/data 7.2/temps.txt")
> # set random seed
> set.seed(9876)
> # check the head
> head(df)
    DAY X1996 X1997 X1998 X1999 X2000 X2001 X2002 X2003 X2004 X2005 X2006 X2007 X2008 X2009 X2010 X2011 X2012
1 1-Jul    98    86    91    84    89    84    90    73    82    91    93    95    85    95    87    92   105
2 2-Jul    97    90    88    82    91    87    90    81    81    89    93    85    87    90    84    94    93
3 3-Jul    97    93    91    87    93    87    87    87    86    86    93    82    91    89    83    95    99
4 4-Jul    90    91    91    88    95    84    89    86    88    86    91    86    90    91    85    92    98
5 5-Jul    89    84    91    90    96    86    93    80    90    89    90    88    88    80    88    90   100
6 6-Jul    93    84    89    91    96    87    93    84    90    82    81    87    82    87    89    90    98
  X2013 X2014 X2015
1    82    90    85
2    85    93    87
3    76    87    79
4    77    84    85
5    83    86    84
6    83    87    84
> # check the summary
> summary(df)
     DAY                X1996           X1997           X1998           X1999           X2000
 Length:123         Min.   :60.00   Min.   :55.00   Min.   :63.00   Min.   :57.00   Min.   : 55.00
 Class :character   1st Qu.:79.00   1st Qu.:78.50   1st Qu.:79.50   1st Qu.:75.00   1st Qu.: 77.00
 Mode  :character   Median :84.00   Median :84.00   Median :86.00   Median :86.00   Median : 86.00
                    Mean   :83.72   Mean   :81.67   Mean   :84.26   Mean   :83.36   Mean   : 84.03
                    3rd Qu.:90.00   3rd Qu.:88.50   3rd Qu.:89.00   3rd Qu.:91.00   3rd Qu.: 91.00
                    Max.   :99.00   Max.   :95.00   Max.   :95.00   Max.   :99.00   Max.   :101.00
     X2001           X2002           X2003           X2004           X2005           X2006
 Min.   :51.00   Min.   :57.00   Min.   :57.00   Min.   :62.00   Min.   :54.00   Min.   :53.00
 1st Qu.:78.00   1st Qu.:78.00   1st Qu.:78.00   1st Qu.:78.00   1st Qu.:81.50   1st Qu.:79.00
 Median :84.00   Median :87.00   Median :84.00   Median :82.00   Median :85.00   Median :85.00
 Mean   :81.55   Mean   :83.59   Mean   :81.48   Mean   :81.76   Mean   :83.36   Mean   :83.05
 3rd Qu.:87.00   3rd Qu.:91.00   3rd Qu.:87.00   3rd Qu.:87.00   3rd Qu.:88.00   3rd Qu.:91.00
 Max.   :93.00   Max.   :97.00   Max.   :91.00   Max.   :95.00   Max.   :94.00   Max.   :98.00
     X2007           X2008           X2009           X2010           X2011           X2012
 Min.   : 59.0   Min.   :50.00   Min.   :51.00   Min.   :67.00   Min.   :59.00   Min.   : 56.00
 1st Qu.: 81.0   1st Qu.:79.50   1st Qu.:75.00   1st Qu.:82.00   1st Qu.:79.00   1st Qu.: 79.50
 Median : 86.0   Median :85.00   Median :83.00   Median :90.00   Median :89.00   Median : 85.00
 Mean   : 85.4   Mean   :82.51   Mean   :80.99   Mean   :87.21   Mean   :85.28   Mean   : 84.65
 3rd Qu.: 89.5   3rd Qu.:88.50   3rd Qu.:88.00   3rd Qu.:93.00   3rd Qu.:94.00   3rd Qu.: 90.50
 Max.   :104.0   Max.   :95.00   Max.   :95.00   Max.   :97.00   Max.   :99.00   Max.   :105.00
     X2013           X2014           X2015
 Min.   :56.00   Min.   :63.00   Min.   :56.0
 1st Qu.:77.00   1st Qu.:81.50   1st Qu.:77.0
 Median :84.00   Median :86.00   Median :85.0
 Mean   :81.67   Mean   :83.94   Mean   :83.3
 3rd Qu.:88.00   3rd Qu.:89.00   3rd Qu.:90.0
 Max.   :92.00   Max.   :95.00   Max.   :97.0
```

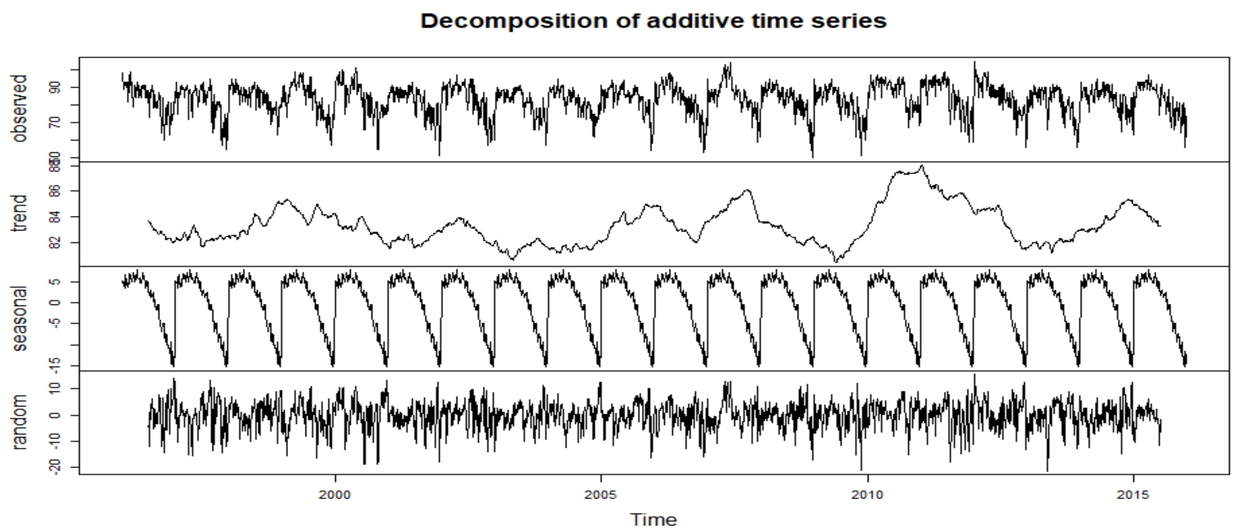
This part of code is really simple here. Import data, and check the head and summary of the dataframe.

```
> # check the structure of dataframe
> str(df)
'data.frame':	123 obs. of  21 variables:
 $ DAY  : chr  "1-Jul" "2-Jul" "3-Jul" "4-Jul" ...
 $ X1996: int  98 97 97 90 89 93 93 91 93 93 ...
 $ X1997: int  86 90 93 91 84 84 75 87 84 87 ...
 $ X1998: int  91 88 91 91 91 89 93 95 95 91 ...
 $ X1999: int  84 82 87 88 90 91 82 86 87 87 ...
 $ X2000: int  89 91 93 95 96 96 96 91 96 99 ...
 $ X2001: int  84 87 87 84 86 87 87 89 91 87 ...
 $ X2002: int  90 90 87 89 93 93 89 89 90 91 ...
 $ X2003: int  73 81 87 86 80 84 87 90 89 84 ...
 $ X2004: int  82 81 86 88 90 90 89 87 88 89 ...
 $ X2005: int  91 89 86 86 89 82 76 88 89 78 ...
 $ X2006: int  93 93 93 91 90 81 80 82 84 84 ...
 $ X2007: int  95 85 82 86 88 87 82 82 89 86 ...
 $ X2008: int  85 87 91 90 88 82 88 90 89 87 ...
 $ X2009: int  95 90 89 91 80 87 86 82 84 84 ...
 $ X2010: int  87 84 83 85 88 89 94 97 96 90 ...
 $ X2011: int  92 94 95 92 90 90 94 94 91 92 ...
 $ X2012: int  105 93 99 98 100 98 93 95 97 95 ...
 $ X2013: int  82 85 76 77 83 83 79 88 88 87 ...
 $ X2014: int  90 93 87 84 86 87 89 90 90 87 ...
 $ X2015: int  85 87 79 85 84 84 90 90 91 93 ...
> # converting to time series
> temp_<-as.vector(unlist(df[,2:21]))
> plot(temp_)
> # converting to time series
> temp_ts<-ts(temp_, start = 1996, frequency = 123)
> ts_plot(temp_ts)
> # decompose the time series
> df_timeseriescomponents <- decompose(temp_ts)
> plot(df_timeseriescomponents)
```



Here we check the structure of the dataframe, and finish the time series transformatio n steps. And plot time series to check the variations. Afterwards, we decompose the time series. As we look at the decomposition components (plots are listed below), we can visually see how they can add up to our “observed” value, in other words, the real values. It’s also very important to inspect the scales of each section to see which one is more dominant. For example, if “random” has a range significantly larger than seasonal or trend, this data is going to be very challenging to accurately forecast later on.



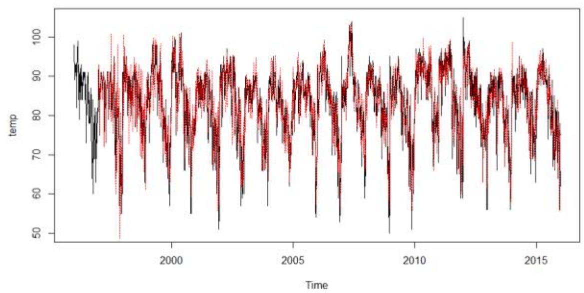


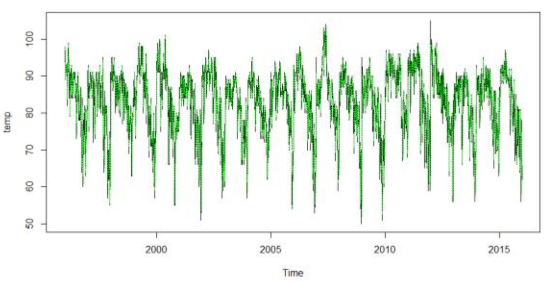
Next, we build a HoltWinters model by setting different alpha, beta, gamma parameters. Alpha is the “base value”. Higher alpha puts more weight on the most recent observations. Beta is the “trend value”. Higher beta means the trend slope is more dependent on recent trend slopes. Bamma: the “seasonal component”. Higher gamma puts more weighting on the most recent seasonal cycles. The key part here is our beta here is either 0 or approaching to 0, which means trend slope is more dependent on historical trends slopes. And out dataset puts more weighting on most recent seasonal cycles.

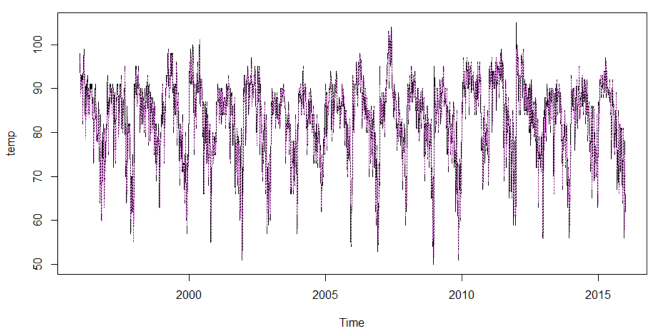
```
> # build the model
> hw0 <- HoltWinters(temp_ts)
> hw0$alpha
    alpha
0.6610618
> hw0$beta
beta
   0
> hw0$gamma
    gamma
0.6248076
> hw0$SSE
[1] 66244.25
> #
> hw1 <- HoltWinters(temp_ts,beta=FALSE,gamma=FALSE)
> hw1$alpha
[1] 0.8388021
> hw1$SSE
[1] 56198.1
> #
> hw2 <- HoltWinters(temp_ts,gamma=FALSE)
> hw2$alpha
    alpha
0.8445729
> hw2$beta
       beta
0.003720884
> hw2$SSE
[1] 56572.54
> #
> hw3 <- HoltWinters(temp_ts,seasonal="multiplicative")
> hw3$alpha
   alpha
0.615003
> hw3$beta
beta
   0
> hw3$gamma
    gamma
0.5495256
> hw3$SSE
[1] 68904.57
```

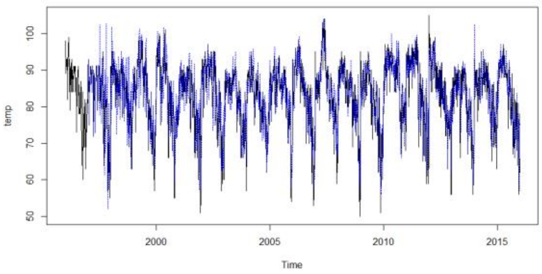
Then we plot these hw models.

```
> #
> plot(temp_ts, ylab="temp")
> lines(hw0$fitted[,1], lty=3, col="red")
> #
> plot(temp_ts, ylab="temp")
> lines(hw1$fitted[,1], lty=3, col="green", pch = 16)
> #
> plot(temp_ts, ylab="temp")
> lines(hw2$fitted[,1], lty=3, col="violet", pch = 16)
> #
> plot(temp_ts, ylab="temp")
> lines(hw3$fitted[,1], lty=3, col="blue", pch = 16)
> #
```





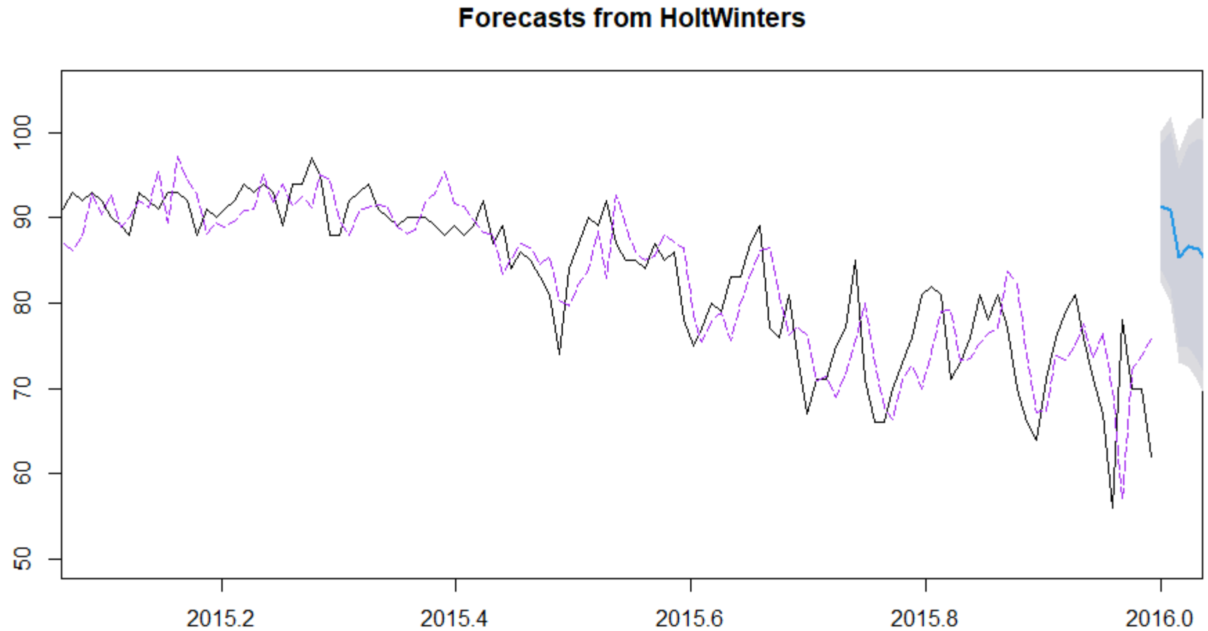




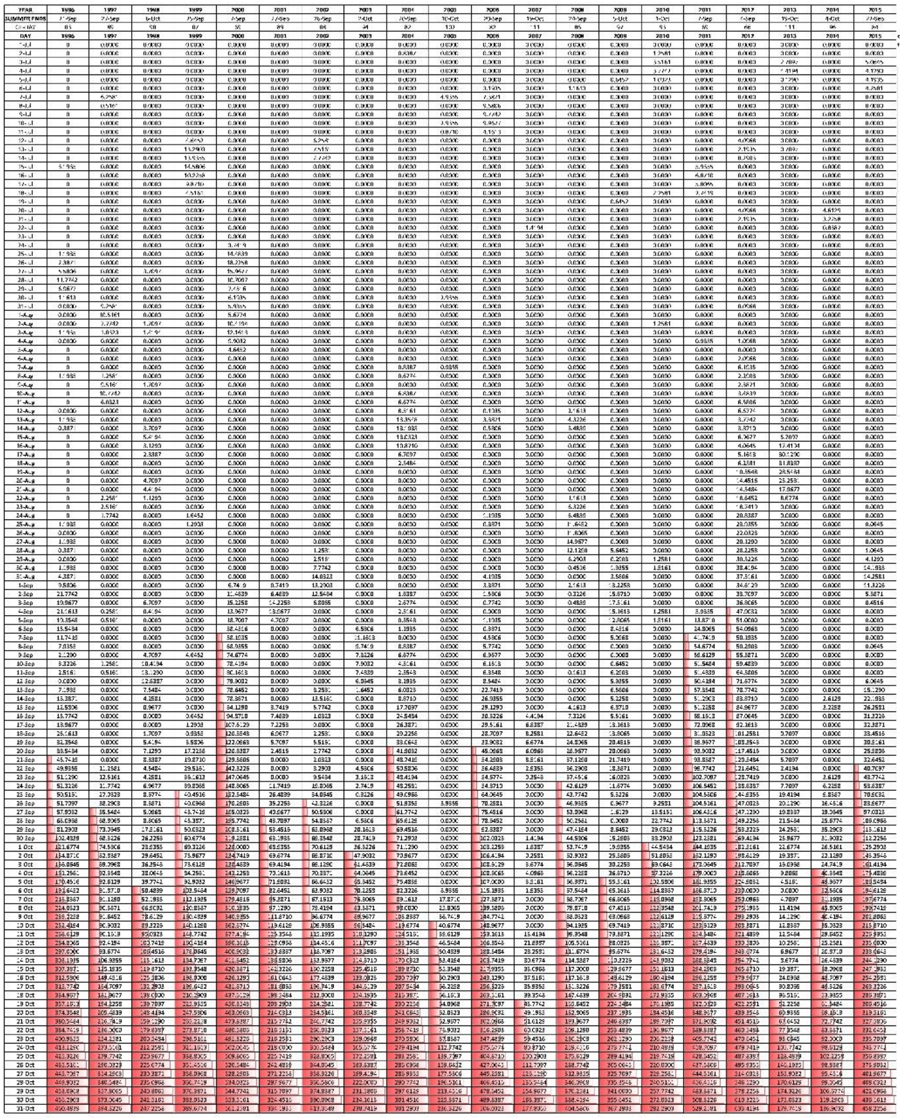
Four models fits look to follow our data quite well, so now it's time to see how they do predicting future temperature.

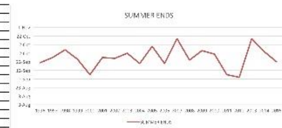
```
#
hw3_for <- forecast(hw3, h=12, level=c(90, 95))
plot(hw3_for, xlim=c(2015, 2016))
lines(hw3_for$fitted, lty=5, col="purple", pch = 16)
```

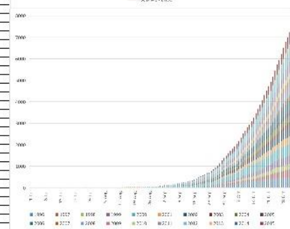
The key part here is which model should we choose to forecast. SSE is an indicator for one hand. But if you check the $fitted column for each model, you will see only hw0 and hw3 have seasonal factors. As we saw in decomposition of our dataset, the data have fluctuating seasonal trends and hw3 season factor is more smooth. So I choose hw3 for forecast.



Grey area is significance level between (90, 95). According to exponentially smoothed model, the summer might ended unofficially later but we can conclude that there is not enough evidence that end of summer has gotten later. Also I will attach a CUSUM results from last week.







C = 1		T = 4		mu = 88.75 (avg July temp)
YEAR	summer ends	Number of days	AVG(Tempture)	CUSUM
1996	21-Sep	83	88.12	0
1997	27-Sep	89	85.89	0.00
1998	6-Oct	98	86.55	0.00
1999	25-Sep	87	88.16	0.00
2000	7-Sep	69	90.34	0.59
2001	27-Sep	89	85.27	0.00
2002	26-Sep	88	88.08	0.00
2003	2-Oct	94	84.27	0.00
2004	20-Sep	82	85.04	0.00
2005	10-Oct	102	85.86	0.00
2006	20-Sep	82	88.28	0.00
2007	19-Oct	111	87.25	0.00
2008	24-Sep	86	86.54	0.00
2009	5-Oct	97	84.53	0.00
2010	1-Oct	93	90.52	0.77
2011	7-Sep	69	92.16	3.18
2012	4-Sep	66	90.75	4.19
2013	19-Oct	111	83.32	0.00
2014	4-Oct	96	86.11	0.00
2015	22-Sep	84	88.06	0.00