

Homework 3 - Week 3

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?

I am an Analyst at Mindtree Limited. A year back I worked on a project on a Proof of Concept to one of the clients, that involved the time series data. I use Python for analysis and used the Scikit Learn package in Python. In that project, I ask was to estimate the ATM cash demand forecasts for New York, Illinois, California. I created a model to predict the demand for the number of bills for each denomination. I tried different modeling techniques like ARIMA, SARIMA, and Exponential Smoothing (Holt-Winters). In the end, I used Holt-Winters, which gave a great result, to predict the demand forecast for a few weeks. The error rate of the prediction models was less than 15%. I used α of 0.7 giving importance/weightage to my recent value but at the same time not leaving the previous outputs.

Few of the KPI's(predictors) are as follows,

1. Zipcode
2. Time of loading the money
3. Count of bills that was loaded (for each denomination)
4. Time at which each type of bill was empty (for each denomination)
5. Distance of ATMs from one and another in the same city

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).

For this Question I had a reference from : <https://fukamilab.github.io/BIO202/09-A-time-series.html>
(<https://fukamilab.github.io/BIO202/09-A-time-series.html>)

```
In [270]: # if (!require("fma")) install.packages("fma")
# if (!require("expsmooth")) install.packages("expsmooth")
# install.packages("forecast", repos='http://cran.us.r-project.org')
library(fma)
library(expsmooth)
suppressWarnings(suppressMessages(library("TTR")))
suppressWarnings(suppressMessages(library("forecast")))
```

```
In [271]: temperature <- read.table("temps.txt", sep = '\t', header = TRUE, check.names = FALSE)
```

```
In [272]: head(temperature)
```

DAY	1996	1997	1998	1999	2000	2001	2002	2003	2004	...	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
1-Jul	98	86	91	84	89	84	90	73	82	...	93	95	85	95	87	92	105	82	90	87
2-Jul	97	90	88	82	91	87	90	81	81	...	93	85	87	90	84	94	93	85	93	93
3-Jul	97	93	91	87	93	87	87	87	86	...	93	82	91	89	83	95	99	76	87	87
4-Jul	90	91	91	88	95	84	89	86	88	...	91	86	90	91	85	92	98	77	84	84
5-Jul	89	84	91	90	96	86	93	80	90	...	90	88	88	80	88	90	100	83	86	86
6-Jul	93	84	89	91	96	87	93	84	90	...	81	87	82	87	89	90	98	83	87	87

```
In [273]: dim(temperature)
```

```
123 21
```

```
In [274]: # Flatten the data frame to a single vector (time series data of frequency = 1 day).  
temp_vector <- as.vector(unlist(temperature[,2:21], recursive = TRUE, use.names = TRUE))
```

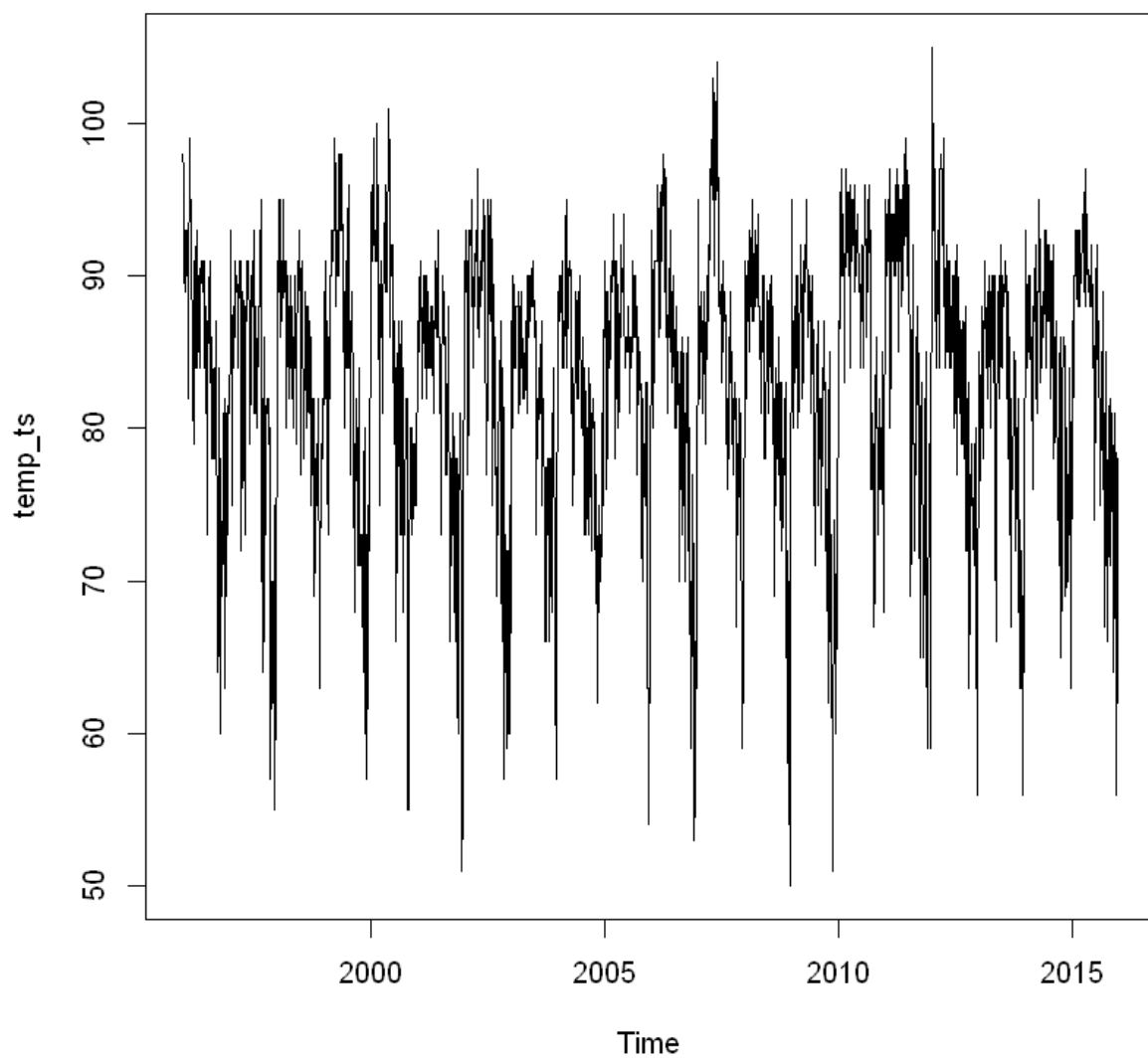
```
In [275]: # There are 123 days of data. We want frequency of 1 day. So, deltat = 1/123  
temp_ts <- ts(temp_vector, start=c(1996,1), end = c(2015,123), deltat = 1/123)
```

```
In [276]: str(temp_ts)
```

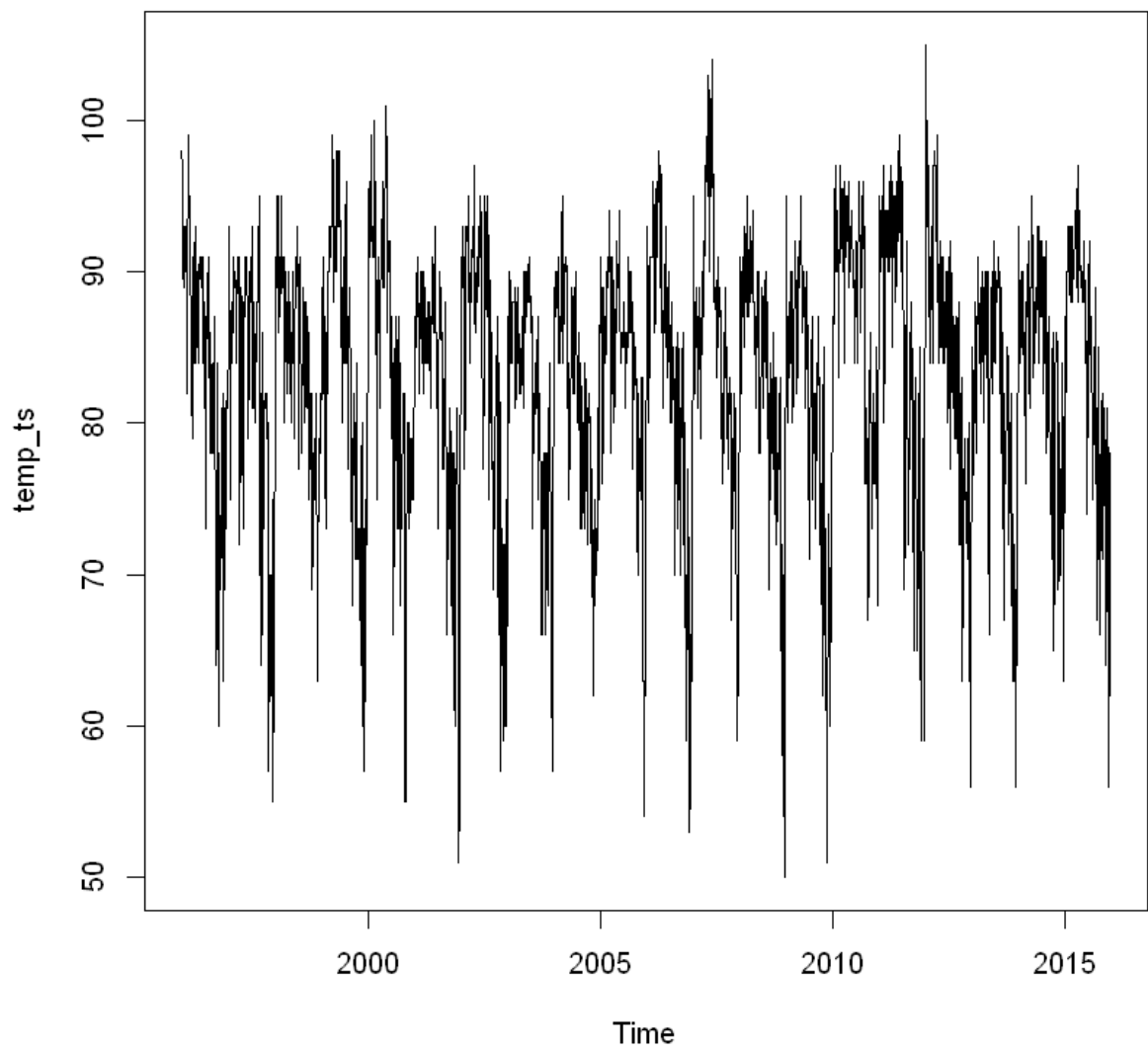
```
Time-Series [1:2460] from 1996 to 2016: 98 97 97 90 89 93 93 91 93 93 ...
```

Plotting the Time Series data across Years

```
In [277]: plot(temp_ts)
```

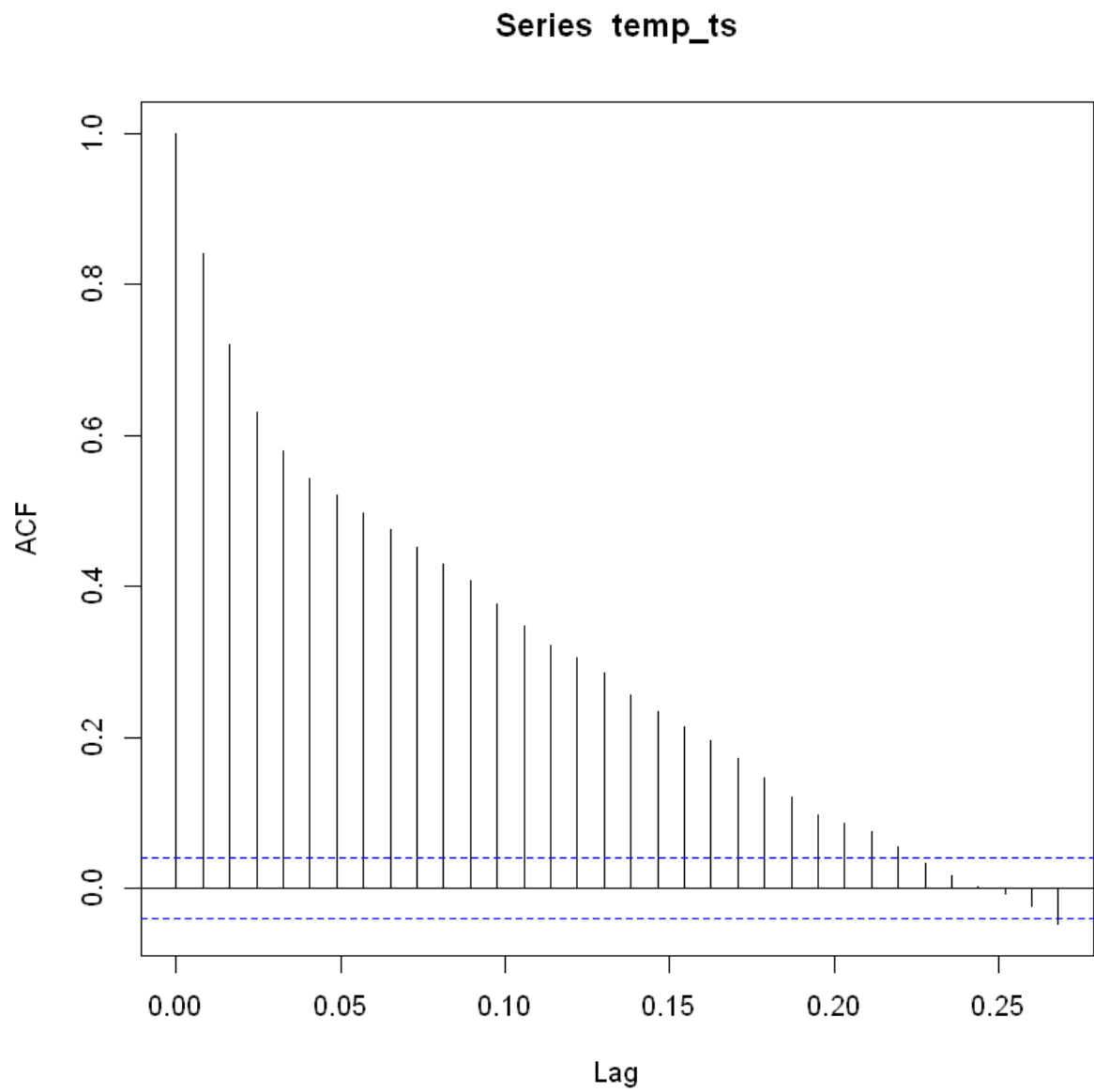


```
In [278]: plot.ts(temp_ts)
```



```
In [145]: # Auto Correlation plot showing that only one value lying outside the 95% limits and
```

```
acf(temp_ts, lag.max = NULL, type = "correlation", plot = TRUE)
```



Reference: <https://www.statisticshowto.com/ljung-box-test/> (<https://www.statisticshowto.com/ljung-box-test/>)

```
In [146]: # the Ljung box test has a p-value < 2.2e-16

Box.test(temp_ts, lag = 10, type = "Ljung-Box", fitdf = 0)
```

Box-Ljung test

data: temp_ts
X-squared = 8350.9, df = 10, p-value < 2.2e-16

The null hypothesis of the Box Ljung Test, H_0 , is that our model does not show lack of fit (or in simple terms—the model is just fine). The alternate hypothesis, H_a , is just that the model does show a lack of fit.

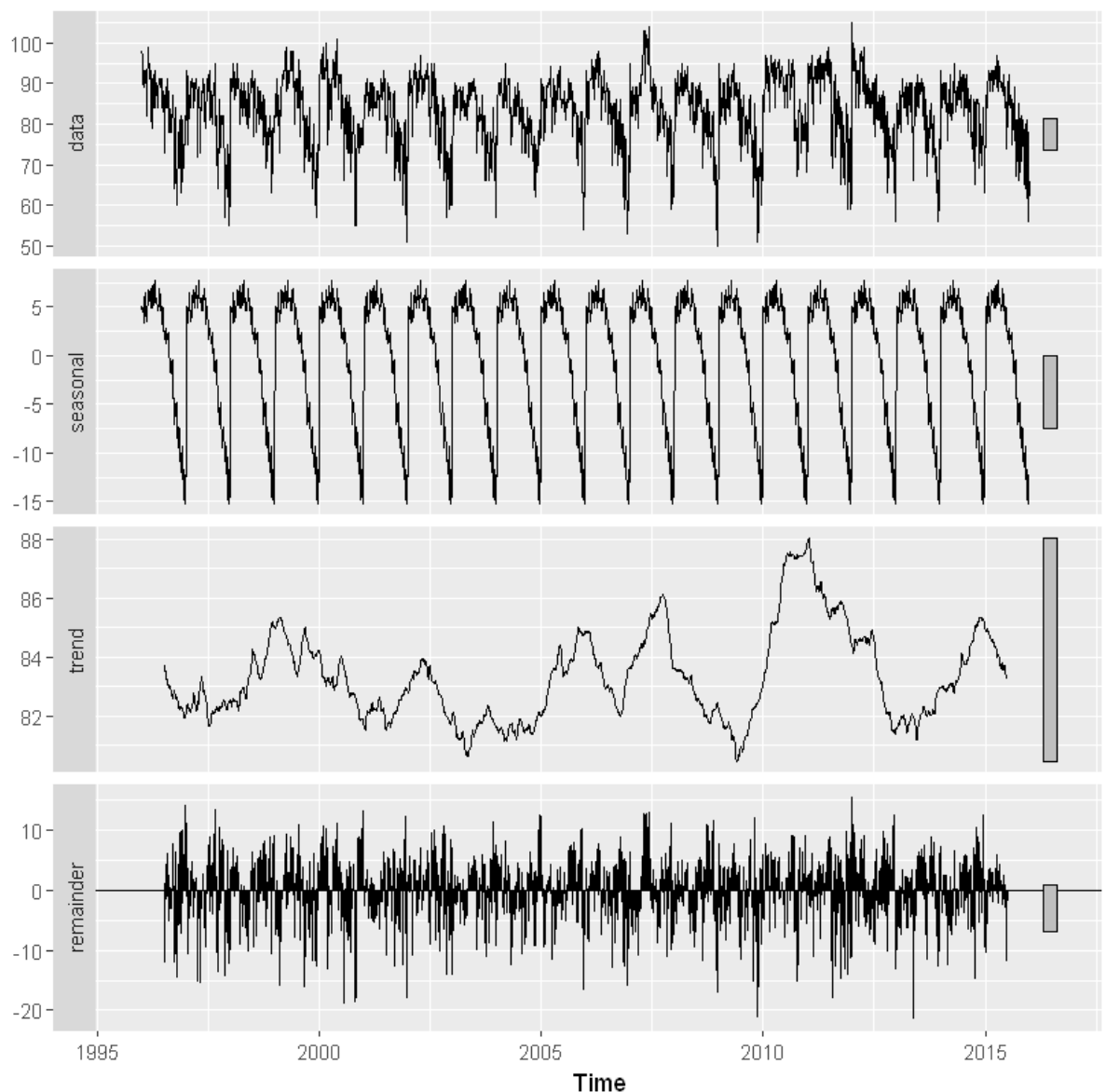
Our p-value is less than 0.05 in this test which rejects the null hypothesis that the time series isn't autocorrelated.

Decomposing Time Series

Decomposing time series dismantles each sequence into its constituents—trend, irregular, and (if applicable) seasonal components

```
In [265]: temp_components <- decompose(temp_ts)
# print(temp_components$seasonal)
autoplot(temp_components)
```

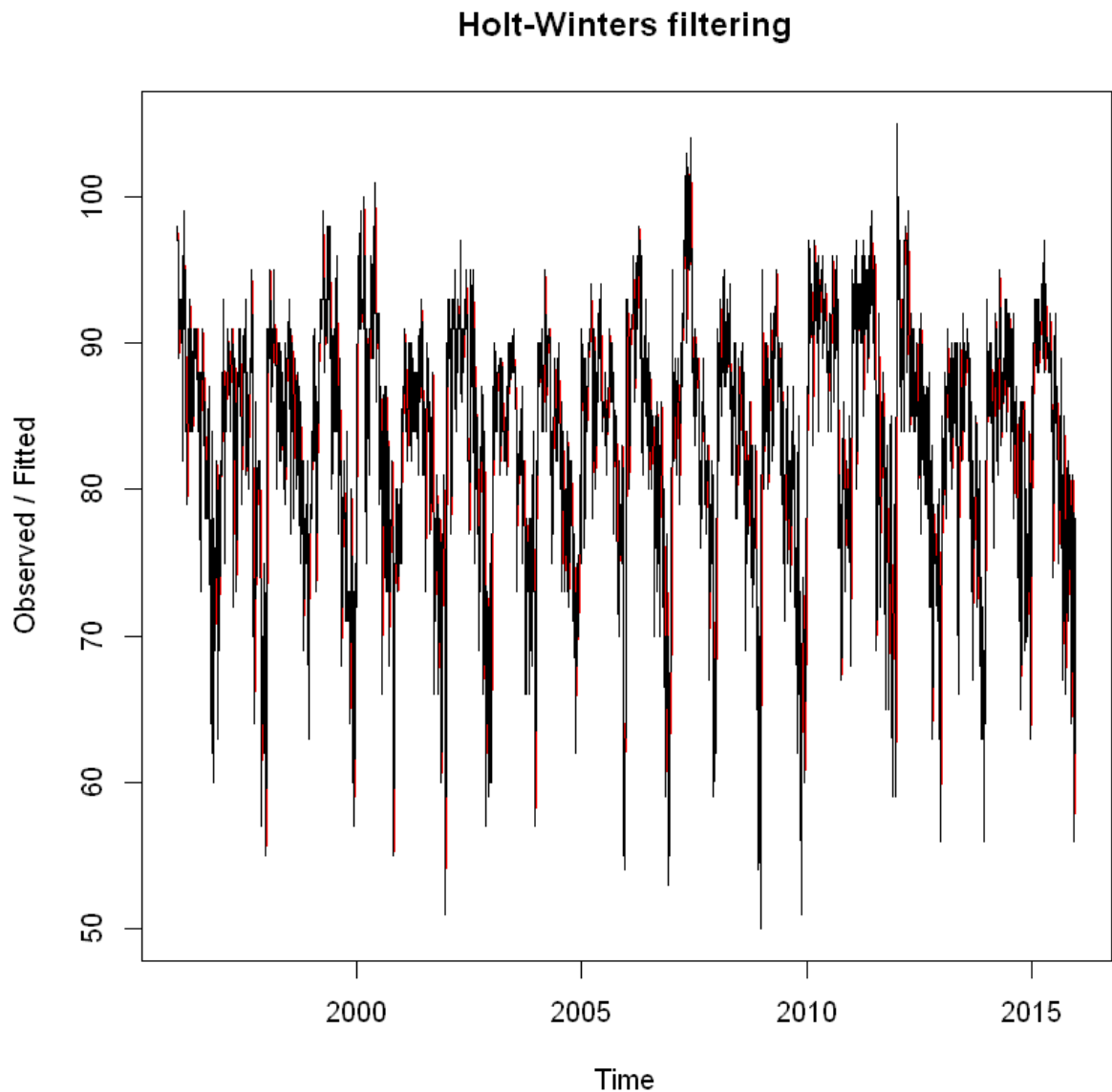
Decomposition of additive time series



Now, I am going to perform single exponential smoothing. Here, I'm using a model with no trend and seasonality. I am going to let R determine the value of alpha.

```
In [191]: temp_single_es <- HoltWinters(temp_ts, beta = FALSE, gamma = FALSE)
```

```
In [193]: plot(temp_single_es)
```



```
In [194]: temp_single_es  
temp_single_es$SSE
```

Holt-Winters exponential smoothing without trend and without seasonal component.

Call:
HoltWinters(x = temp_ts, beta = FALSE, gamma = FALSE)

Smoothing parameters:
alpha: 0.8388021
beta : FALSE
gamma: FALSE

Coefficients:
[,1]
a 63.30952

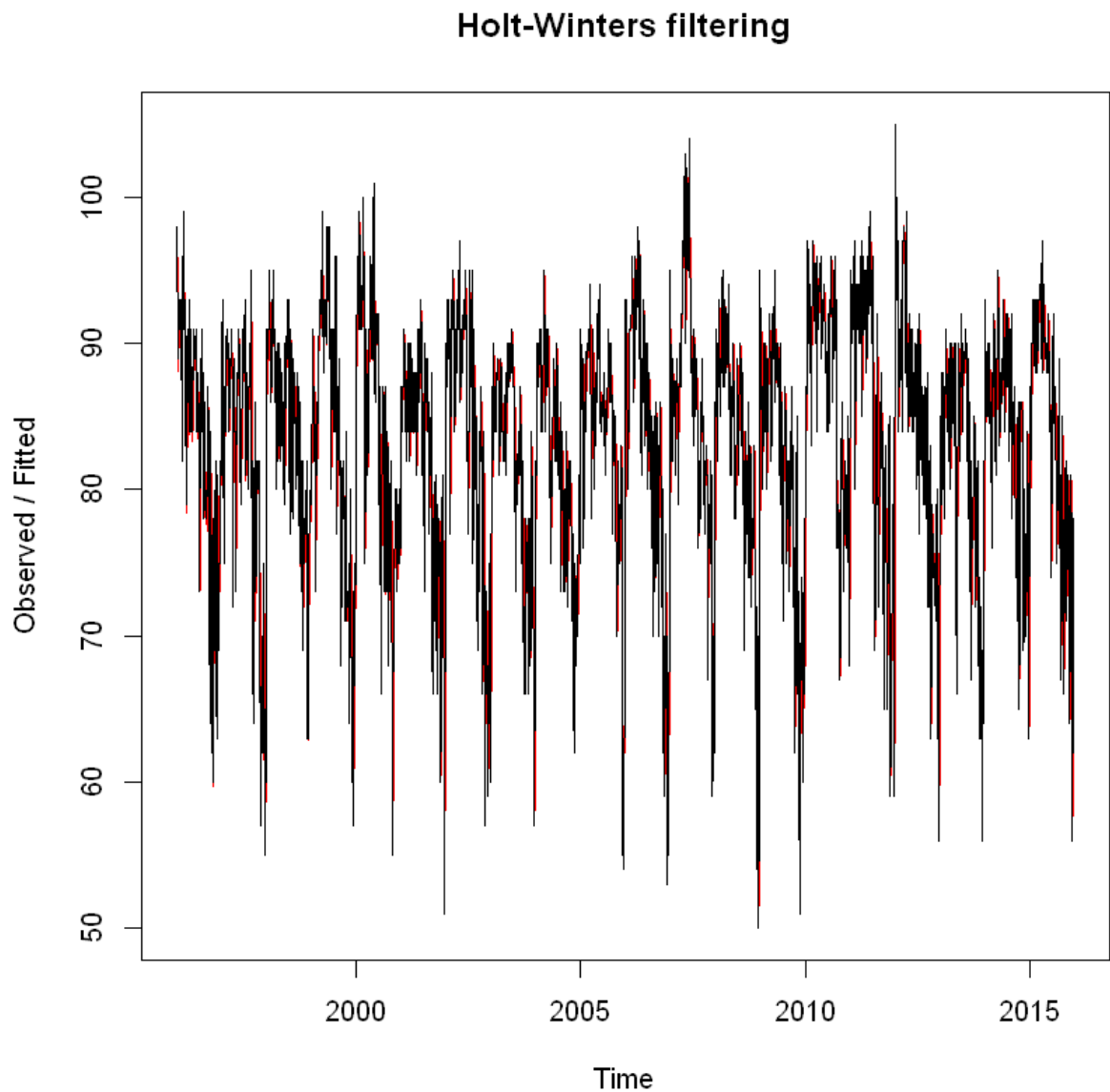
56198.0955314733

The estimated value of alpha is 0.8388021. This is high value indicating that the estimate of the current value of the

level is based mostly upon very recent observations in the time series. The value of the SSE for the in-sample forecast errors is 56198.0955314733.

I am going to perform double exponential smoothing (gamma = FALSE). I am going to let R determine the value of alpha.

```
In [176]: temp_double_es <- HoltWinters(temp_ts, gamma = FALSE)
plot(temp_double_es)
```




```
In [177]: temp_double_es  
temp_double_es$SSE
```

Holt-Winters exponential smoothing with trend and without seasonal component.

Call:

```
HoltWinters(x = temp_ts, gamma = FALSE)
```

Smoothing parameters:

alpha: 0.8445729

beta : 0.003720884

gamma: FALSE

Coefficients:

[,1]

a 63.2530022

b -0.0729933

56572.5375681139

The estimated value of alpha is 0.8445729. Beta is 0.0037. This means that the trend value from the recent observations has relatively very little weight when forecasting for future values. The value of the sum-of-squared-errors for the in-sample forecast errors is 56572.5375681139.

Next, I am going to check if the data can be described using an additive model. I am going to use Holt-Winters triple exponential smoothing to estimate the level (alpha), slope (beta) and seasonal (gamma) components.

In [266]: *# additive model*

```
temp_add_hw <- HoltWinters(temp_ts)
temp_add_hw
temp_add_hw$SSE
plot(temp_add_hw)
```

Holt-Winters exponential smoothing with trend and additive seasonal component.

Call:

```
HoltWinters(x = temp_ts)
```

Smoothing parameters:

alpha: 0.6610618

beta : 0

gamma: 0.6248076

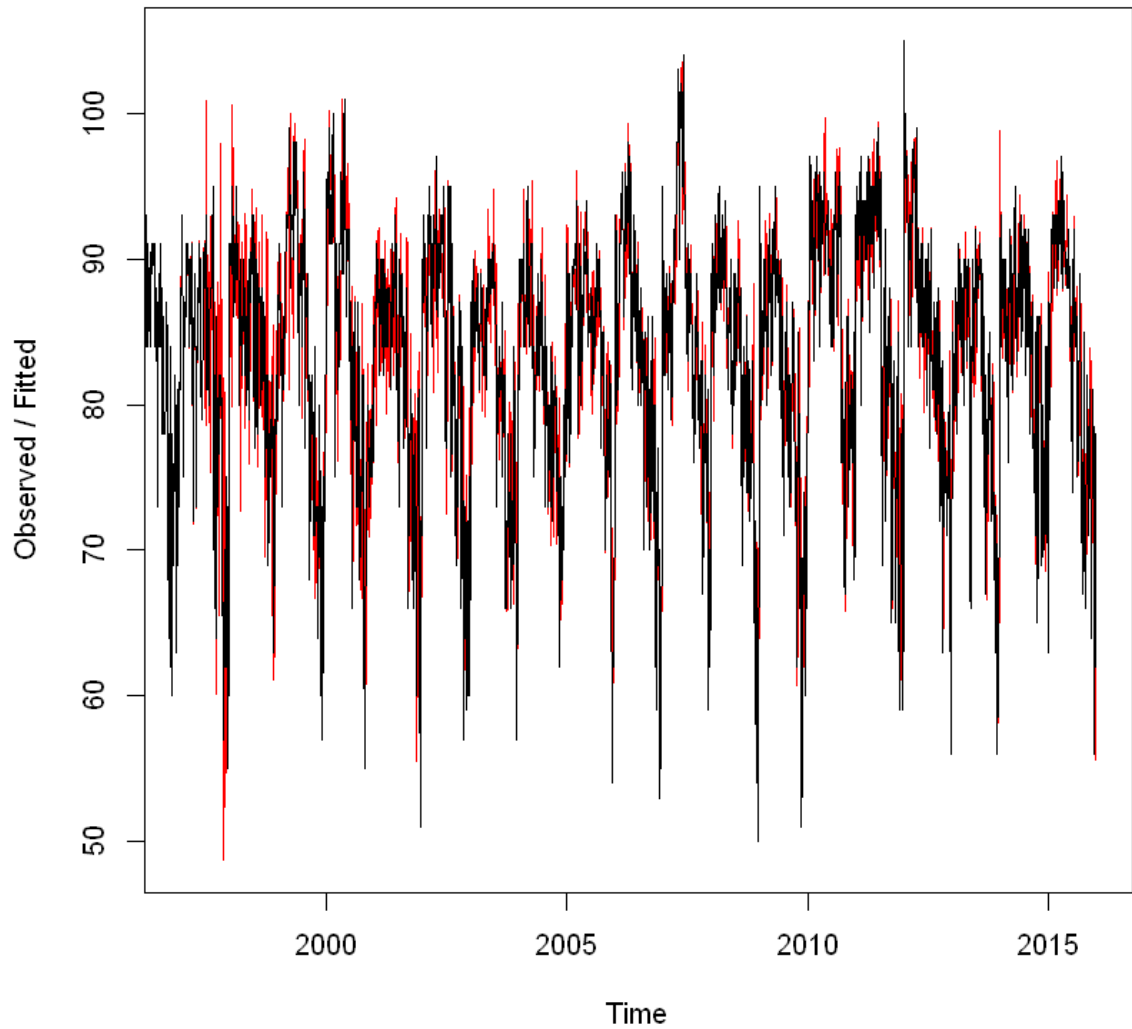
Coefficients:

```
      [,1]
a    71.477236414
b    -0.004362918
s1   18.590169842
s2   17.803098732
s3   12.204442890
s4   13.233948865
s5   12.957258705
s6   11.525341233
s7   10.854441534
s8   10.199632666
s9    8.694767348
s10   5.983076192
s11   3.123493477
s12   4.698228193
s13   2.730023168
s14   2.995935818
s15   1.714600919
s16   2.486701224
s17   6.382595268
s18   5.081837636
s19   7.571432660
s20   6.165047647
s21   9.560458487
s22   9.700133847
s23   8.808383245
s24   8.505505527
s25   7.406809208
s26   6.839204571
s27   6.368261304
s28   6.382080380
s29   4.552058253
s30   6.877476437
s31   4.823330209
s32   4.931885957
s33   7.109879628
s34   6.178469084
s35   4.886891317
s36   3.890547248
s37   2.148316257
s38   2.524866001
s39   3.008098232
s40   3.041663870
s41   2.251741386
s42   0.101091985
s43  -0.123337548
s44  -1.445675315
s45  -1.802768181
s46  -2.192036338
s47  -0.180954242
s48   1.538987281
s49   5.075394760
s50   6.740978049
s51   7.737089782
s52   8.579515859
s53   8.408834158
s54   4.704976718
s55   1.827215229
```

s56 -1.275747384
s57 1.389899699
s58 1.376842871
s59 0.509553410
s60 1.886439429
s61 -0.806454923
s62 5.221873550
s63 5.383073482
s64 4.265584552
s65 3.841481452
s66 -0.231239928
s67 0.542761270
s68 0.780131779
s69 1.096690727
s70 0.690525998
s71 2.301303414
s72 2.965913580
s73 4.393732595
s74 2.744547070
s75 1.035278911
s76 1.170709479
s77 2.796838283
s78 2.000312540
s79 0.007337449
s80 -1.203916069
s81 0.352397232
s82 0.675108103
s83 -3.169643942
s84 -1.913321175
s85 -1.647780450
s86 -5.281261301
s87 -5.126493027
s88 -2.637666754
s89 -2.342133004
s90 -3.281910970
s91 -4.242033198
s92 -2.596010530
s93 -7.821281290
s94 -8.814741200
s95 -8.996689798
s96 -7.835655534
s97 -5.749139155
s98 -5.196182693
s99 -8.623793296
s100 -11.809355220
s101 -13.129428554
s102 -16.095143067
s103 -15.125436350
s104 -13.963606549
s105 -12.953304848
s106 -16.097179844
s107 -15.489223470
s108 -13.680122300
s109 -11.921434142
s110 -12.035411347
s111 -12.837047727
s112 -9.095808127
s113 -5.433029341
s114 -6.800835107
s115 -8.413639598
s116 -10.912409484
s117 -13.553826535
s118 -10.652543677
s119 -12.627298331
s120 -9.906981556
s121 -12.668519900
s122 -9.805502547
s123 -7.775306633

66244.2504058467

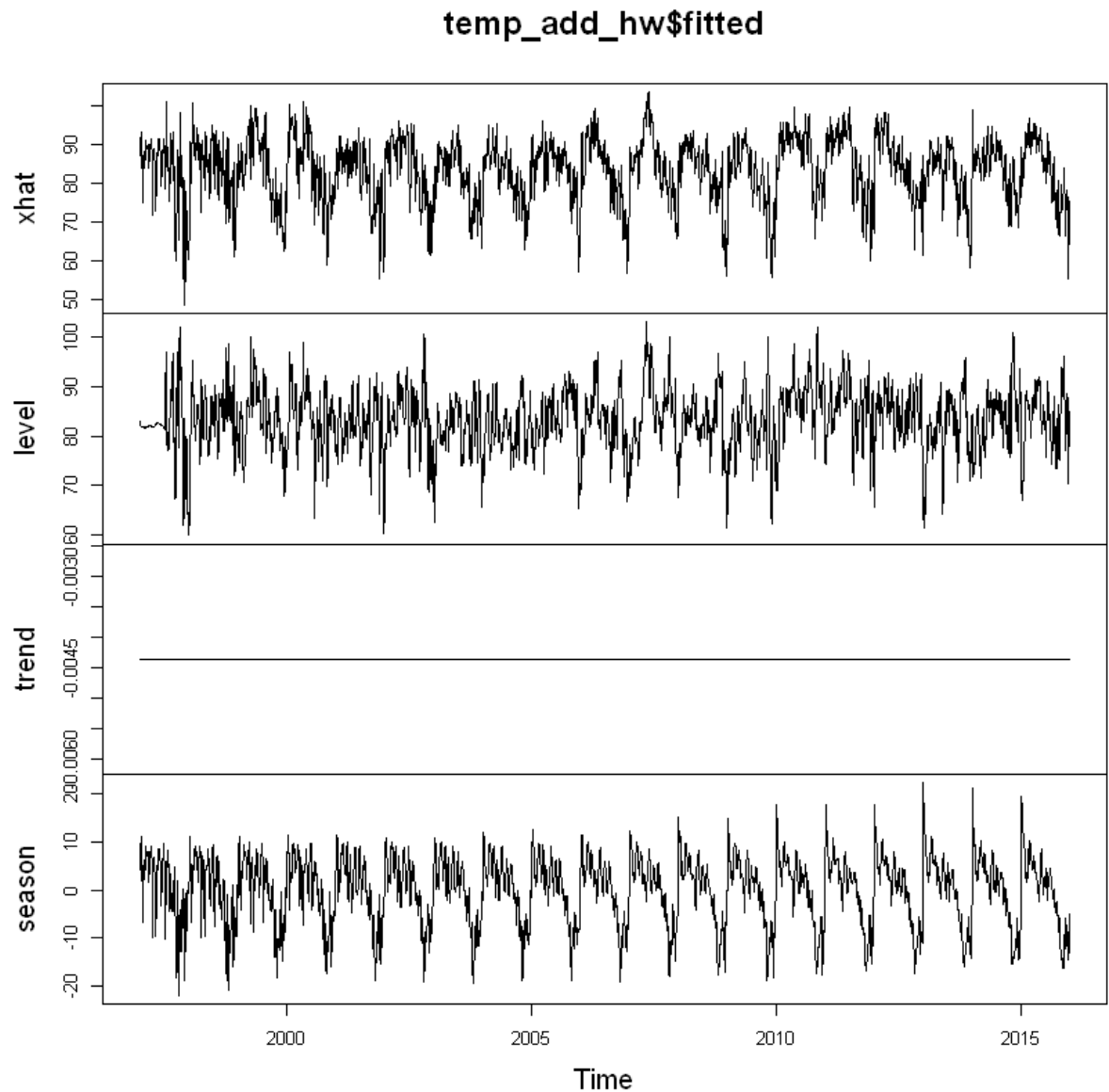
Holt-Winters filtering



The value of beta is zero, suggesting no trend from recent observations on forecasting future values. The level parameter is 0.6610618, and the seasonal smoothing parameter, gamma is 0.6248076. SSE is 66244.2504058466.

The forecasts made by HoltWinters function are stored in a named element of this list variable called fitted.

```
In [267]: # print(head(temp_add_hw$fitted,100))  
plot(temp_add_hw$fitted)
```



I am going to check if the data can be described using an multiplicative model. I am going to use Holt-Winters triple

exponential smoothing to estimate the level (α), slope (β) and seasonal (γ) components.

```
In [268]: temp_mul_hw <- HoltWinters(temp_ts,alpha = NULL,beta = NULL,gamma = NULL, seasonal = "multiplicative")
temp_mul_hw
temp_mul_hw$SSE
plot(temp_mul_hw)
```

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

Call:

```
HoltWinters(x = temp_ts, alpha = NULL, beta = NULL, gamma = NULL, seasonal = "multiplicative")
```

Smoothing parameters:

```
alpha: 0.615003
beta : 0
gamma: 0.5495256
```

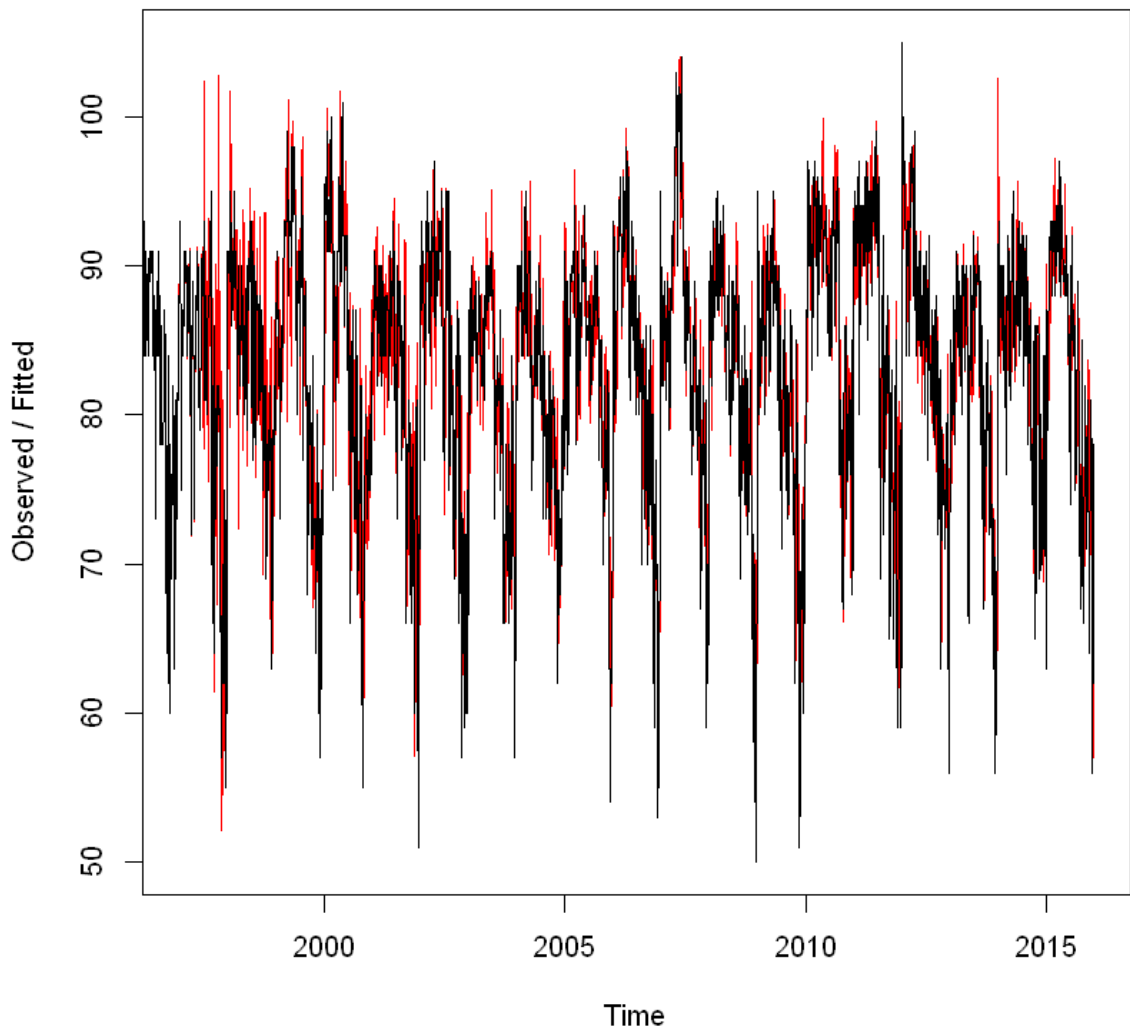
Coefficients:

```
      [,1]
a  73.679517064
b  -0.004362918
s1   1.239022317
s2   1.234344062
s3   1.159509551
s4   1.175247483
s5   1.171344196
s6   1.151038408
s7   1.139383104
s8   1.130484528
s9   1.110487514
s10  1.076242879
s11  1.041044609
s12  1.058139281
s13  1.032496529
s14  1.036257448
s15  1.019348815
s16  1.026754142
s17  1.071170378
s18  1.054819556
s19  1.084397734
s20  1.064605879
s21  1.109827336
s22  1.112670130
s23  1.103970506
s24  1.102771209
s25  1.091264692
s26  1.084518342
s27  1.077914660
s28  1.077696145
s29  1.053788854
s30  1.079454300
s31  1.053481186
s32  1.054023885
s33  1.078221405
s34  1.070145761
s35  1.054891375
s36  1.044587771
s37  1.023285461
s38  1.025836722
s39  1.031075732
s40  1.031419152
s41  1.021827552
s42  0.998177248
s43  0.996049257
s44  0.981570825
s45  0.976510542
s46  0.967977608
s47  0.985788411
s48  1.004748195
s49  1.050965934
s50  1.072515008
s51  1.086532279
s52  1.098357400
s53  1.097158461
s54  1.054827180
s55  1.022866587
s56  0.987259326
s57  1.016923524
```

s58 1.016604903
s59 1.004320951
s60 1.019102781
s61 0.983848662
s62 1.055888360
s63 1.056122844
s64 1.043478958
s65 1.039475693
s66 0.991019224
s67 1.001437488
s68 1.002221759
s69 1.003949213
s70 0.999566344
s71 1.018636837
s72 1.026490773
s73 1.042507768
s74 1.022500795
s75 1.002503740
s76 1.004560984
s77 1.025536556
s78 1.015357769
s79 0.992176558
s80 0.979377825
s81 0.998058079
s82 1.002553395
s83 0.955429116
s84 0.970970220
s85 0.975543504
s86 0.931515830
s87 0.926764603
s88 0.958565273
s89 0.963250387
s90 0.951644060
s91 0.937362688
s92 0.954257999
s93 0.892485444
s94 0.879537700
s95 0.879946892
s96 0.890633648
s97 0.917134959
s98 0.925991769
s99 0.884247686
s100 0.846648167
s101 0.833696369
s102 0.800001437
s103 0.807934782
s104 0.819343668
s105 0.828571029
s106 0.795608740
s107 0.796609993
s108 0.815503509
s109 0.830111282
s110 0.829086181
s111 0.818367239
s112 0.863958784
s113 0.912057203
s114 0.898308248
s115 0.878723779
s116 0.848971946
s117 0.813891909
s118 0.846821392
s119 0.819121827
s120 0.851036184
s121 0.820416491
s122 0.851581233
s123 0.874038407

68904.569331748

Holt-Winters filtering



In []:

Again, the value of beta is zero, suggesting no trend from recent observations on forecasting future values. The level parameter is 0.615003, and the seasonal smoothing parameter, gamma is 0.5495256. SSE is 68904.569331748.

I am going to write the fitted values to a csv file to perform CUSUM approach to detect unofficial end of summer.

```
In [211]: df_temp1 <- matrix(temp_mul_hw$fitted[,4], nrow = 123) # taking season and performing CUSUM
```

```
In [214]: # install.packages("xlsx", repos='http://cran.us.r-project.org')
suppressWarnings(suppressMessages(require(xlsx)))
```

```
In [245]: write.csv(df_temp1, file = 'smoothed_temperature.csv')
```

Next, I am going to try to predict the temperatures for July 1 through Oct 31 for 2016 and 2017 using the Holt-Winters multiplicative model. To do this, I am using the predict() function that inputs the HW object, prediction interval, number of predictions, and confidence level.

```
In [244]: predicted_temp <- predict(temp_mul_hw, n.ahead = 123*2, prediction.interval = FALSE, level = 0.95)
print(predicted_temp)
```

```
Time Series:
Start = c(2016, 1)
End = c(2017, 123)
Frequency = 123
fit
[1,] 91.28516
[2,] 90.93510
[3,] 85.41693
[4,] 86.57116
[5,] 86.27852
[6,] 84.77782
[7,] 83.91440
[8,] 83.25410
[9,] 81.77658
[10,] 79.25010
[11,] 76.65370
[12,] 77.90779
[13,] 76.01528
[14,] 76.28765
[15,] 75.00000
```

```
In [236]: new_df <- t(as.data.frame(matrix(round(predicted_temp),ncol =123,byrow = T)))
```

```
In [237]: head(new_df)
```

```
V1 91 91
V2 91 90
V3 85 85
V4 87 86
V5 86 86
V6 85 84
```

```
In [238]: # names(new_df) <- c('X2016', 'X2017')
temperature_data <- cbind(temperature, new_df)
```

```
In [240]: head(temperature_data)
```

	DAY	1996	1997	1998	1999	2000	2001	2002	2003	2004	...	2008	2009	2010	2011	2012	2013	2014	2015	1
V1	1-Jul	98	86	91	84	89	84	90	73	82	...	85	95	87	92	105	82	90	85	91
V2	2-Jul	97	90	88	82	91	87	90	81	81	...	87	90	84	94	93	85	93	87	91
V3	3-Jul	97	93	91	87	93	87	87	87	86	...	91	89	83	95	99	76	87	79	85
V4	4-Jul	90	91	91	88	95	84	89	86	88	...	90	91	85	92	98	77	84	85	87
V5	5-Jul	89	84	91	90	96	86	93	80	90	...	88	80	88	90	100	83	86	84	86
V6	6-Jul	93	84	89	91	96	87	93	84	90	...	82	87	89	90	98	83	87	84	85

```
In [241]: names(temperature_data) <- c('DAY', 'X1996', 'X1997', 'X1998', 'X1999', 'X2000', 'X2001', 'X2002', 'X2003', 'X2004', 'X2005', 'X2006', 'X2007', 'X2008', 'X2009', 'X2010', 'X2011', 'X2012', 'X2013', 'X2014', 'X2015', 'X2016', 'X2017')
```

```
In [242]: head(temperature_data)
```

	DAY	X1996	X1997	X1998	X1999	X2000	X2001	X2002	X2003	X2004	...	X2008	X2009	X2010	X2011	X2012	X2013
V1	1-Jul	98	86	91	84	89	84	90	73	82	...	85	95	87	92	105	88
V2	2-Jul	97	90	88	82	91	87	90	81	81	...	87	90	84	94	93	88
V3	3-Jul	97	93	91	87	93	87	87	87	86	...	91	89	83	95	99	78
V4	4-Jul	90	91	91	88	95	84	89	86	88	...	90	91	85	92	98	78
V5	5-Jul	89	84	91	90	96	86	93	80	90	...	88	80	88	90	100	88
V6	6-Jul	93	84	89	91	96	87	93	84	90	...	82	87	89	90	98	88

In this problem, I have used Holt Winters approach to exponentially smoothe data, and use the smoothed data to predict unofficial end of summer using CUSUM approach. In addition, I have also used the smoothed data to predict temperatures of each day (July1 to Oct 31) for the next two years.

CUSUM APPROCH