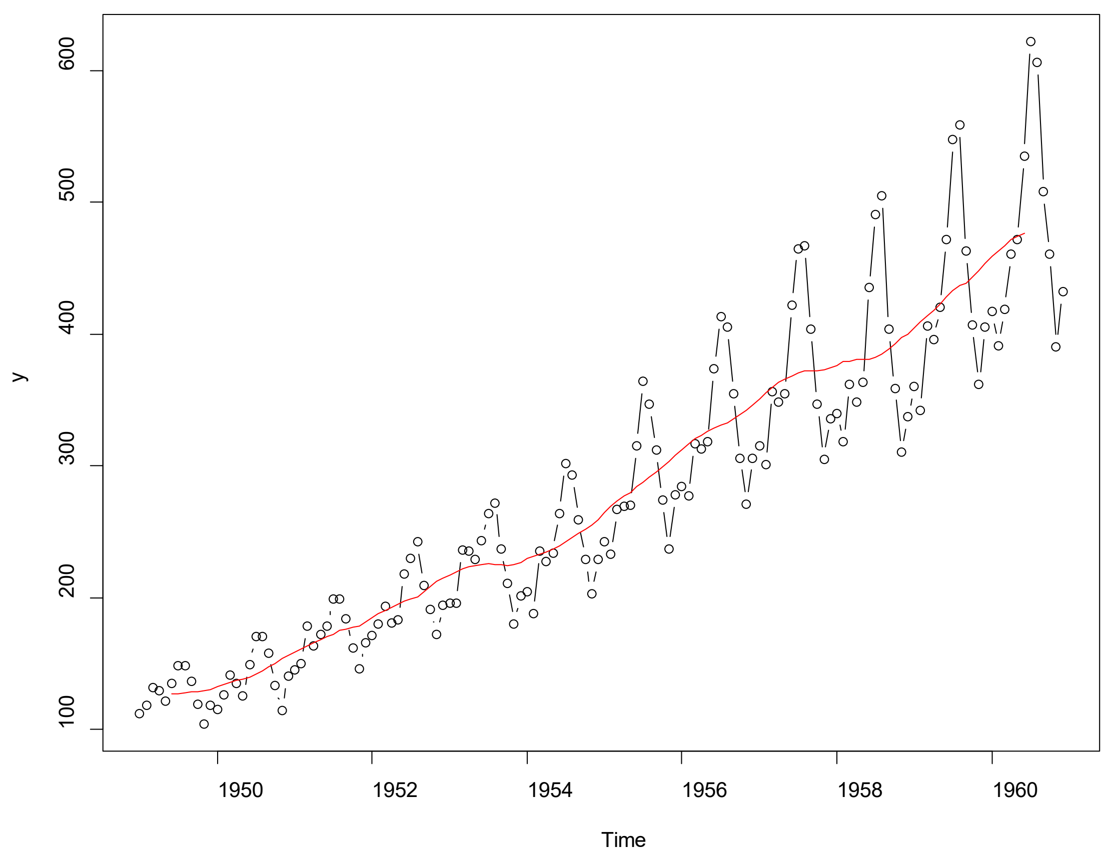
MSiA 420, HW #4

1) HW4_data.xls contains monthly airline passenger data from January 1949 through December 1960. For all of the plots in this question, include the past data along with any forecasts, if relevant.

   (a) Use an appropriate moving average filter to smooth out the seasonality in the data.

Use a centered moving average filter with period m = 12.

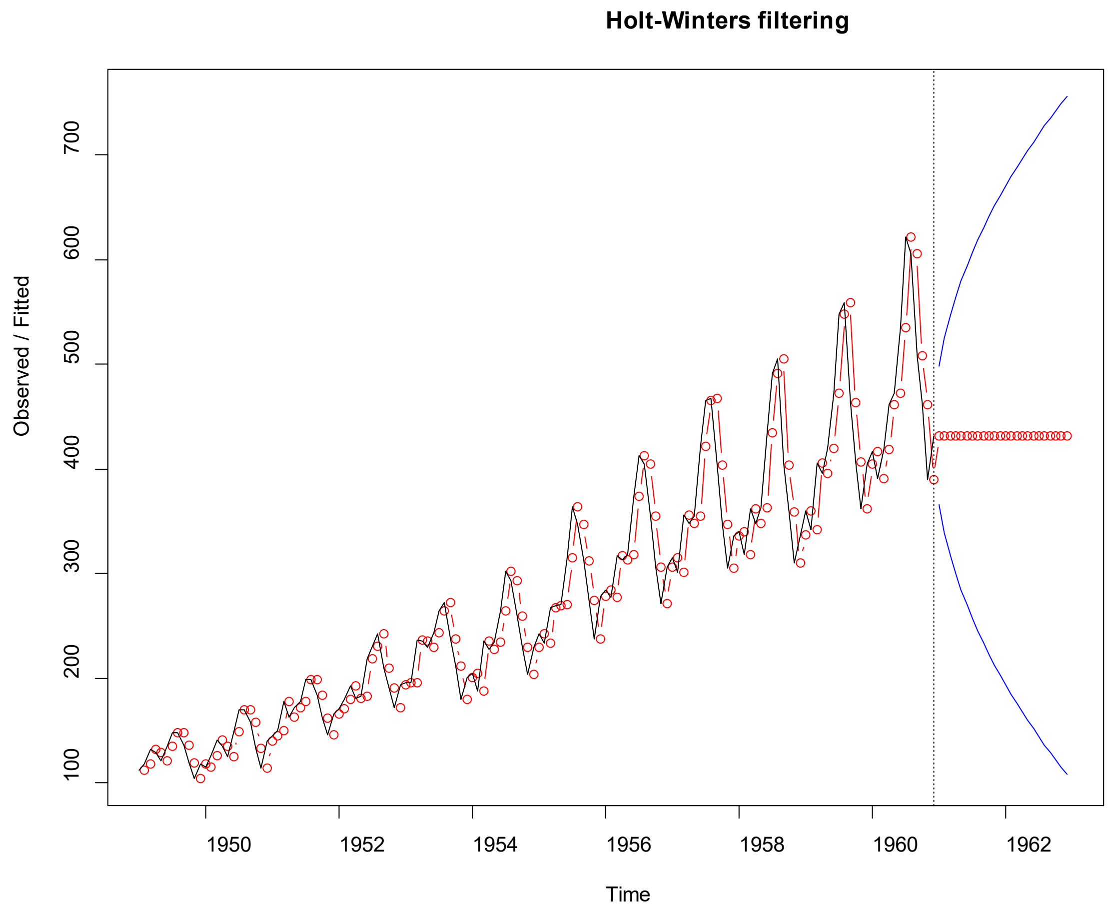
```
airpassenger<-read.csv("airpassenger.csv",header=F)
y<-ts(airpassenger[[1]], start = c(1949,1), frequency=12)
m=12;n=length(y)
MAair<-filter(y, filter=rep(1/m,m), method = "convolution", sides = 2)
plot(y,type="b")
lines(MAair,col="red")
```



(b) Calculate and plot EWMA forecasts for the next two years (1961 and 1962). What is the optimal EWMA parameter α? Explain the nature of the k-step-ahead forecasts for this optimal α value.

```
> k=24;n=length(y)  #k = prediction horizon
> EWMAair<-HoltWinters(y, seasonal = "additive", beta = FALSE, gamma = FALSE)
> EWMAairPred<-predict(EWMAair, n.ahead=k, prediction.interval = T, level = 0.95)
> plot(EWMAair,EWMAairPred,type="b")
> EWMAair
Holt-Winters exponential smoothing without trend and without seasonal component.

Call:
 HoltWinters(x = y, beta = FALSE, gamma = FALSE, seasonal = "additive")

Smoothing parameters:
 alpha:  0.9999339
 beta :  FALSE
 gamma:  FALSE

Coefficients:
    [,1]
a 431.9972
```

Hence, the optimal $\alpha = 1.00$. From the EWMA equations, for $\alpha = 1.00$, the k-step-ahead forecasts are just $\hat{y}_{t+k|t} = L_t = y_t$. That is, the best forecast of future values is the last observed time series value.

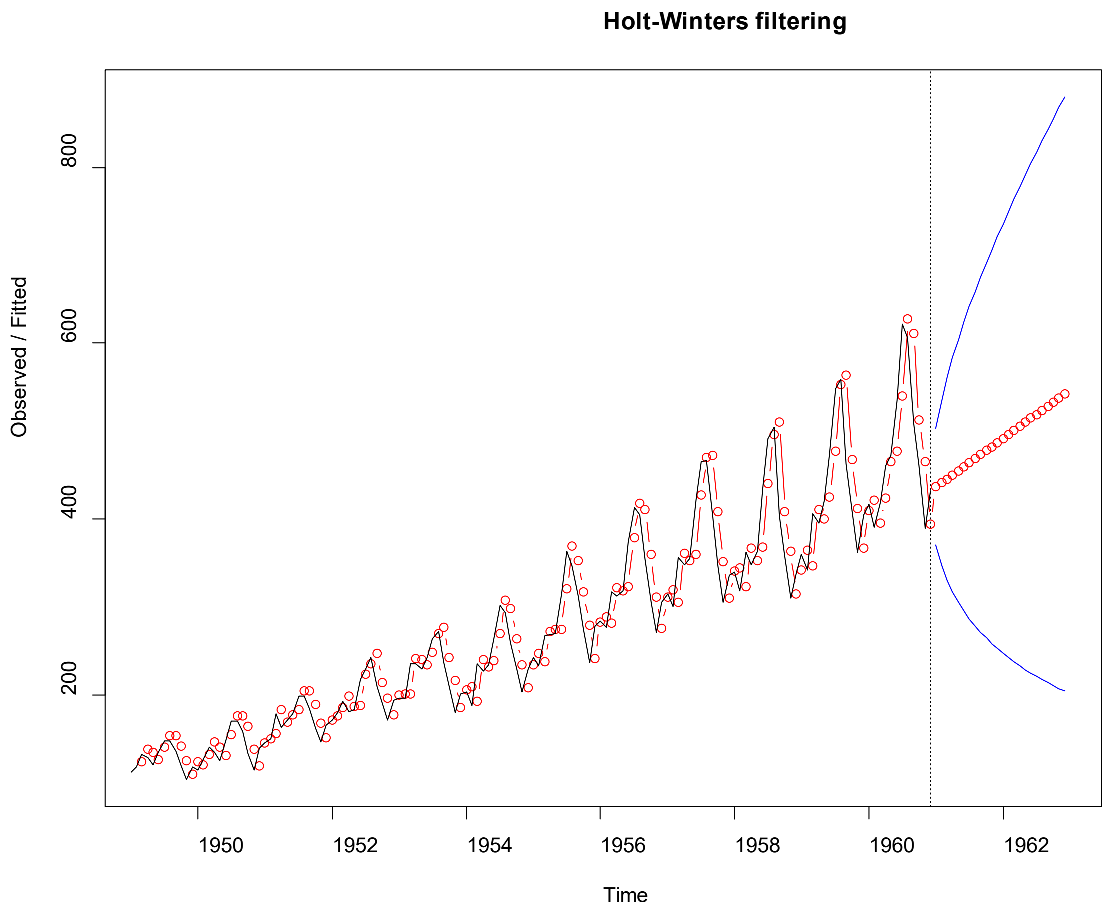


(c) Calculate and plot the Holt method forecasts for the next two years (1961 and 1962). What optimal Holt parameters α and β? Discuss the nature of the k-step-ahead forecasts and how they differ from the EWMA forecasts in part (b).

```
> k=24;n=length(y)  #k = prediction horizon
> Holtair<-HoltWinters(y, seasonal = "additive", gamma = FALSE)
> HoltairPred<-predict(Holtair, n.ahead=k, prediction.interval = T, level = 0.95)
> plot(Holtair,HoltairPred,type="b")
> Holtair
Holt-Winters exponential smoothing with trend and without seasonal component.

Call:
 HoltWinters(x = y, gamma = FALSE, seasonal = "additive")

Smoothing parameters:
 alpha:  1
 beta :  0.003218516
 gamma:  FALSE
```

Coefficients:
      [,1]
a 432.000000
b   4.597605

Hence, the optimal parameters are $\alpha = 1.00$ and $\beta = 0.0032$.



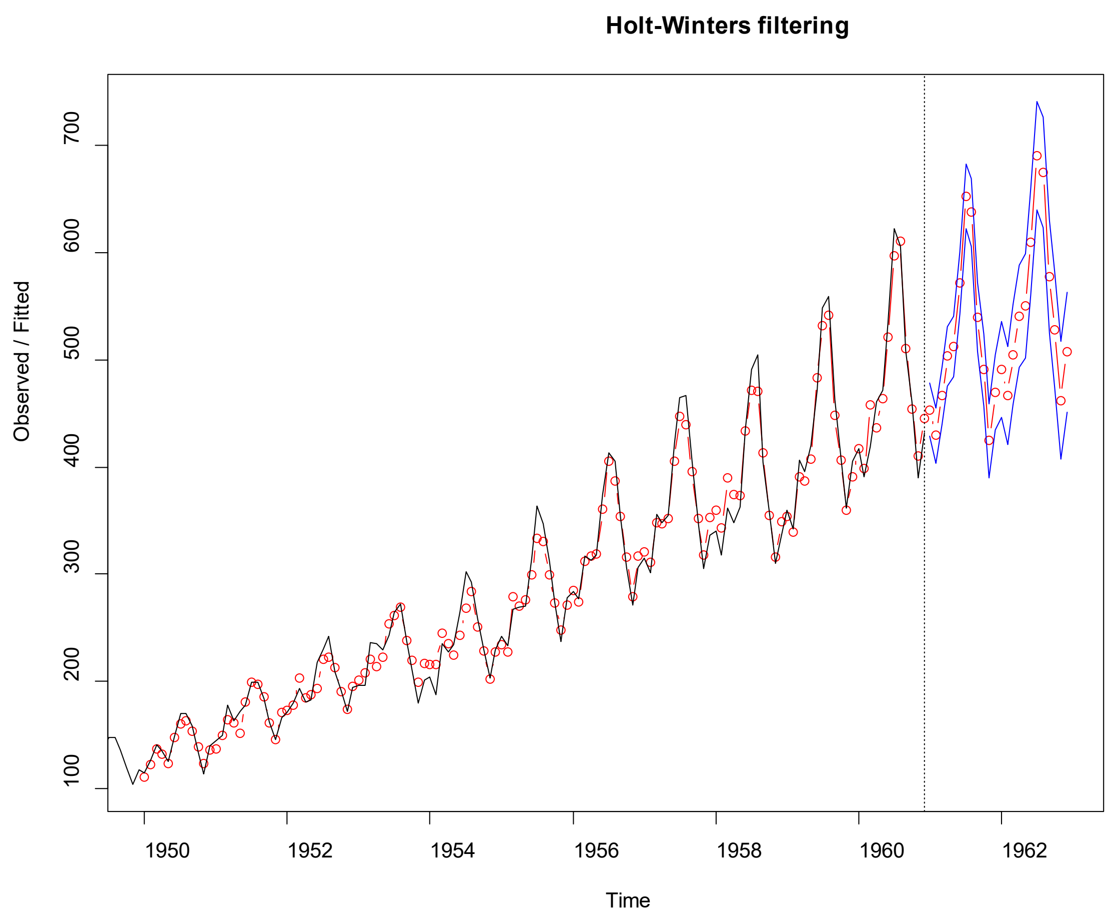
(d) Calculate and plot the Holt-Winters forecasts for the next two years (1961 and 1962) for an additive model. What optimal Holt-Winters parameters α, β, and γ? Interpret the seasonality coefficients.

```
> k=24;n=length(y)  #k = prediction horizon
> HWair<-HoltWinters(y, seasonal = "additive")
> HWairPred<-predict(HWair, n.ahead=k, prediction.interval = T, level = 0.95)
> plot(HWair,HWairPred,type="b")
> HWair
```

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

```
Call:
 HoltWinters(x = y, seasonal = "additive")

Smoothing parameters:
 alpha:  0.2479595
 beta :  0.03453373
 gamma:  1

Coefficients:
          [,1]
a   477.827781
b     3.127627
s1  -27.457685
s2  -54.692464
s3  -20.174608
s4   12.919120
s5   18.873607
s6   75.294426
s7  152.888368
s8  134.613464
s9   33.778349
s10 -18.379060
s11 -87.772408
s12 -45.827781
```

Hence, the optimal parameters are α = 0.25, β = 0.0345, and γ = 1. The seasonality coefficients indicate that the passengers are highest in month 7 (followed closely by month 8) and lowest in months November and February. Months December and January are slightly higher, perhaps because of holiday travel.

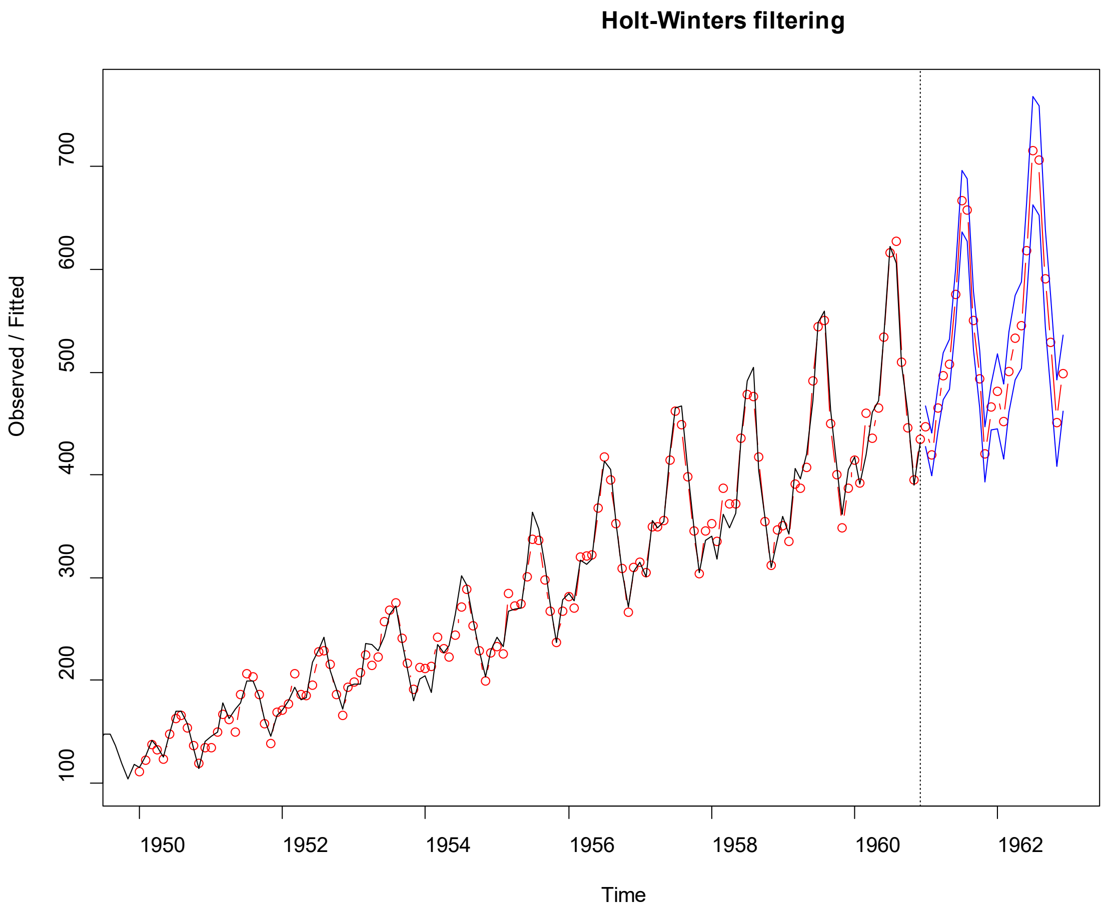


(e) Repeat part (d) but for a Holt-Winters multiplicative model?

```
> k=24;n=length(y)  #k = prediction horizon
> HWair<-HoltWinters(y, seasonal = "multiplicative")
> HWairPred<-predict(HWair, n.ahead=k, prediction.interval = T, level = 0.95)
> plot(HWair,HWairPred,type="b")
> HWair
Holt-Winters exponential smoothing with trend and multiplicative seasonal component.

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

Smoothing parameters:
 alpha:  0.2755925
 beta :  0.03269295
 gamma:  0.8707292

Coefficients:
          [,1]
```

a	469.3232206
b	3.0215391
s1	0.9464611
s2	0.8829239
s3	0.9717369
s4	1.0304825
s5	1.0476884
s6	1.1805272
s7	1.3590778
s8	1.3331706
s9	1.1083381
s10	0.9868813
s11	0.8361333
s12	0.9209877

Hence, the optimal parameters are $\alpha = 0.276$, $\beta = 0.0327$, and $\gamma = 0.871$. In terms of the most and least active months for travel, the seasonality coefficients have identical interpretation as for the additive model in part (d). The only difference is that for the multiplicative model, the seasonalities multiply the level estimates, and they are scaled so that their product is 1.



(f) Which method do you think produces the most appropriate forecasts? Explain.

Definitely the Holt-Winters model, because the data are clearly strongly seasonal. Between the additive and multiplicative Holt-Winters models, the multiplicative appears to fit better (the amplitude of the seasonalities grows as the level grows). Also see the residuals in the decomposition plots in Problem 2 for further evidence that the multiplicative model is superior to the additive model.

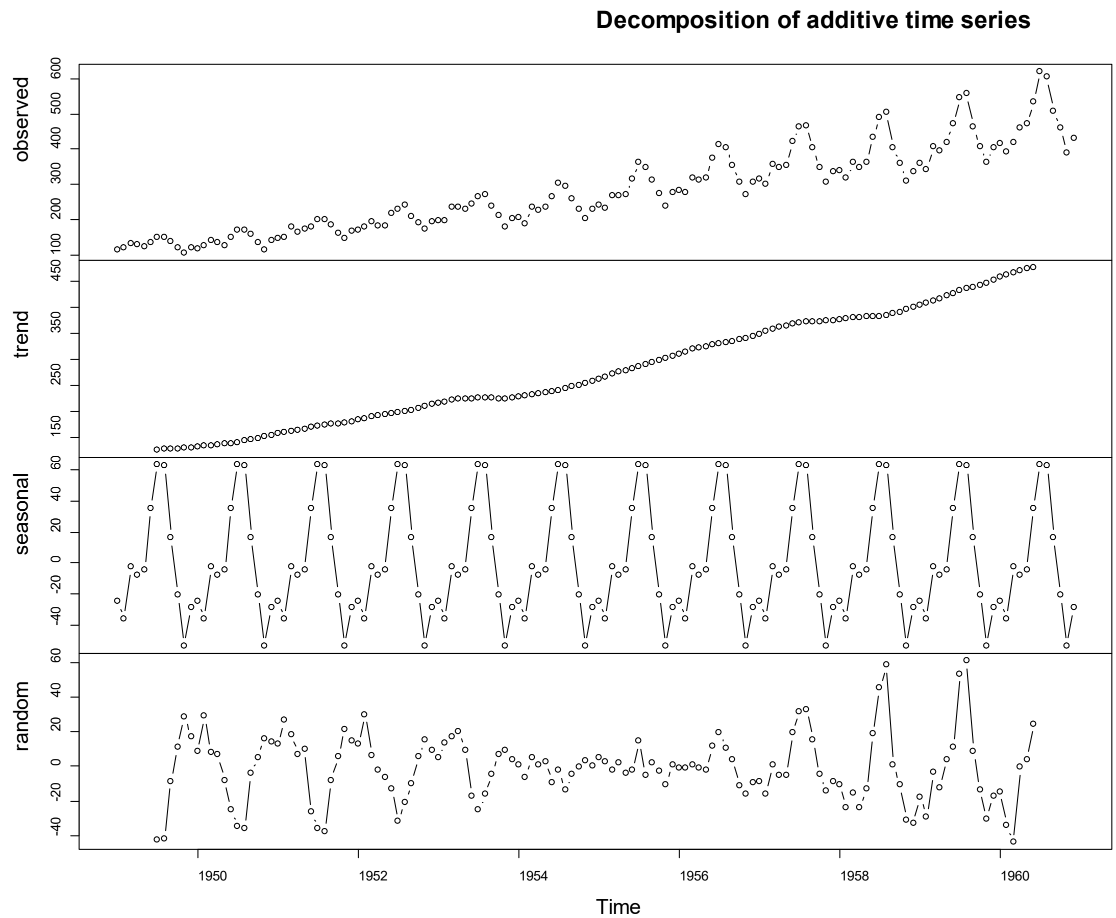
2) This problem pertains to the same airline passenger data that you analyzed in Problem 1.

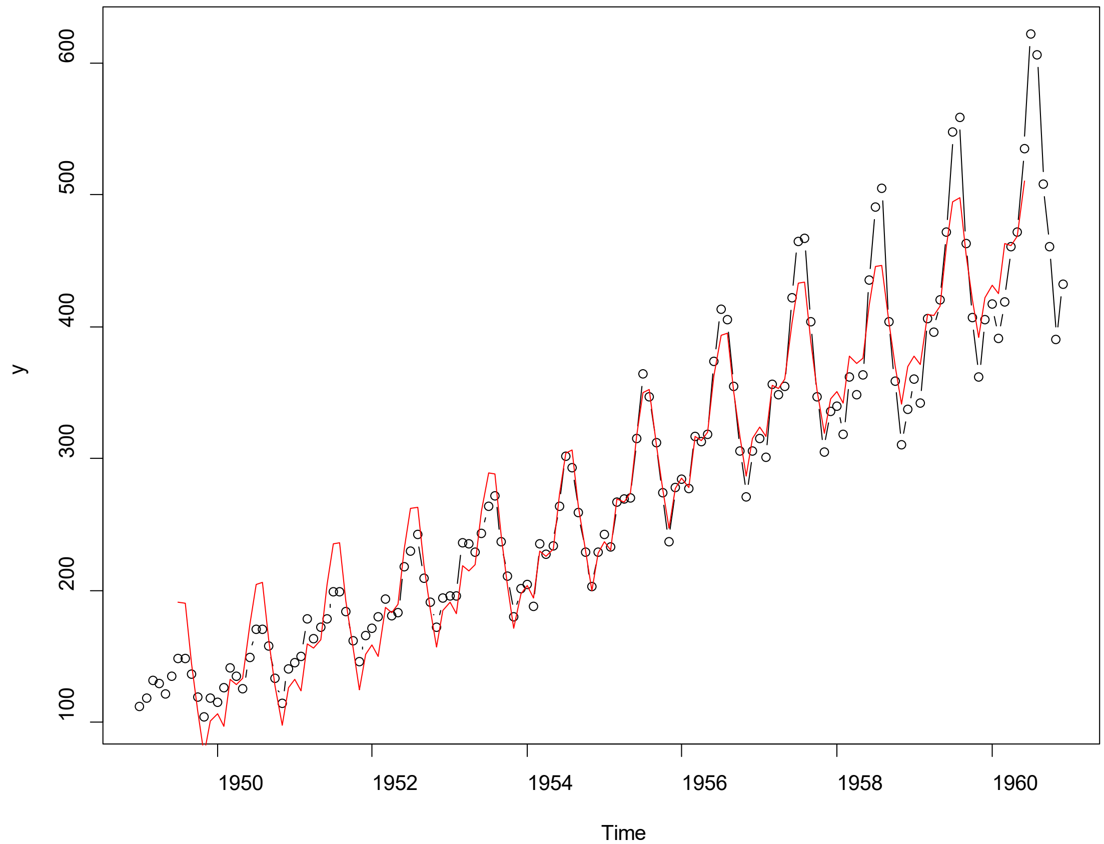
(a) Fit an additive decomposition model to the data. Interpret the trend and seasonality indices. Construct a plot of the original time series and the fitted values, both on the same plot. Discuss the extent to which the variability in the data are accounted for by the trend and seasonality.

```
> k=24;n=length(y)  #k = prediction horizon
> Decair<-decompose(y, type = "additive")
> plot(Decair,type="b")
```

```
> y_hat<-Decair$trend+Decair$seasonal
> plot(y,type="b")
> lines(y_hat,col="red")
```

From the plots below, the residual variance is substantially smaller than the variance of the original time series. However, there is still a clear seasonality present in the residuals, overestimated in the early years and underestimated in the later years. This indicates that a multiplicative model may be better.

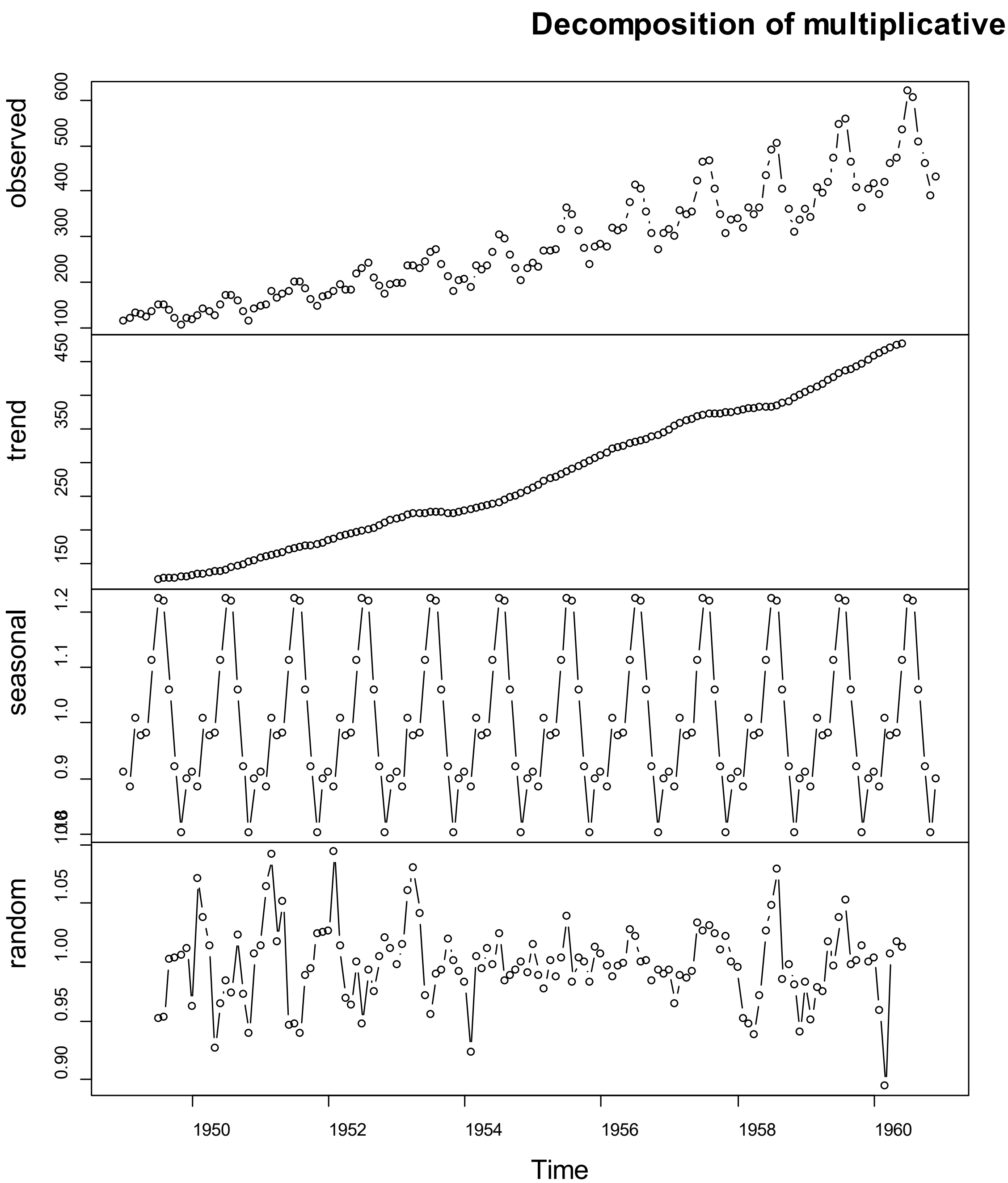


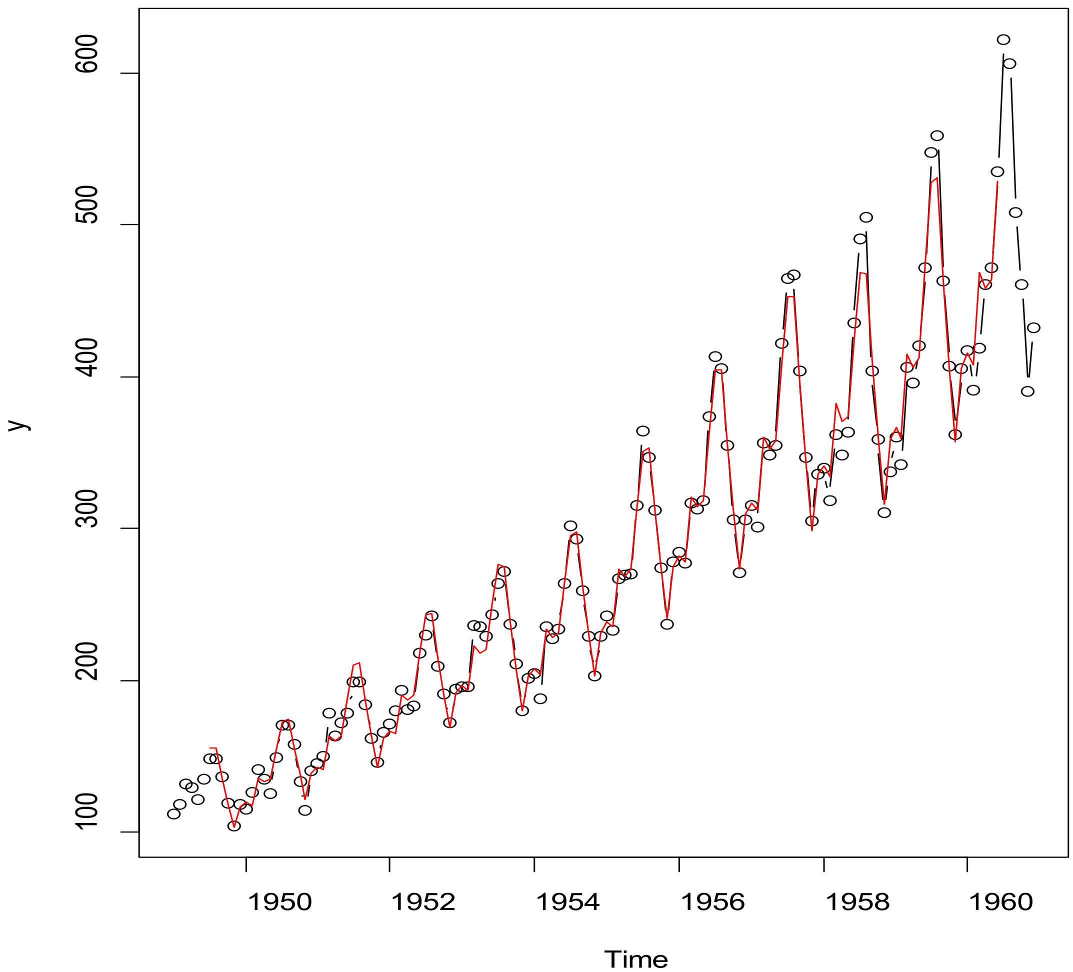


(b) Repeat for a multiplicative decomposition model. Which model – additive or multiplicative – do you think better represents the data?

```
> k=24;n=length(y)  #k = prediction horizon
> Decair<-decompose(y, type = "multiplicative")
> plot(Decair,type="b")
> y_hat<-Decair$trend*Decair$seasonal
> plot(y,type="b")
> lines(y_hat,col="red")
```

From the plots below (especially the second plot), the multiplicative model clearly fits better than the additive model.


Decomposition of multiplicative
observed
trend
seasonal
random
1950
1952
1954
1956
1958
1960
Time


600
500
400
300
200
100
y
1950
1952
1954
1956
1958
1960
Time