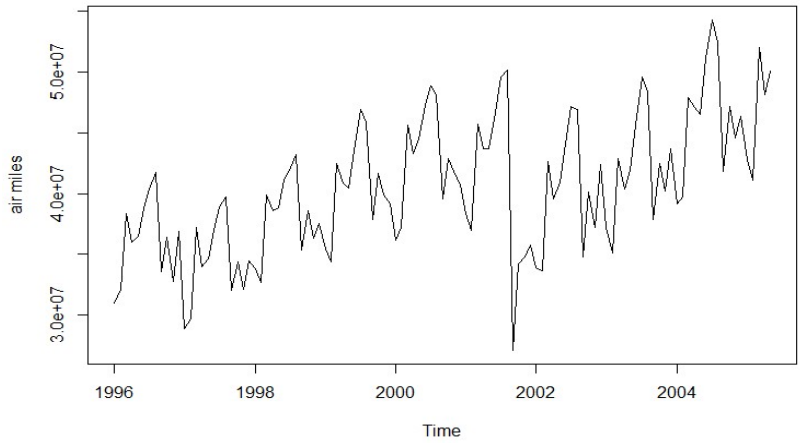
(a) Load in and plot the time series dataset. Does this data appear to come from a stationary process? Why or why not?

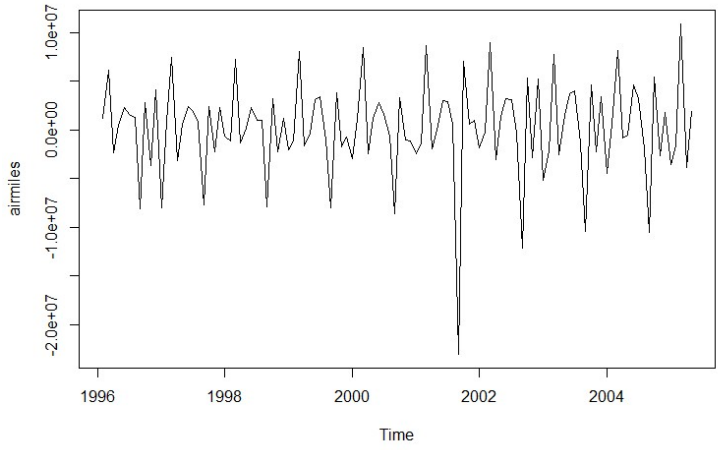
```
rm(list=ls()) # Clear the Environment / History

set.seed(135343466) # <- fix the seed so the results are reproducible!
library(TSA)


data(airmiles)
plot(airmiles, type='l', ylab="final position")
```



I think it's not from a stationary process because the value was 3.0e+0.7 at the beginning and it has increased to 5.0e+07, which means "drift up".

(b) Plot the first difference of the time series. What improvements do you see here? Is there still something in this data that needs to be accounted for?

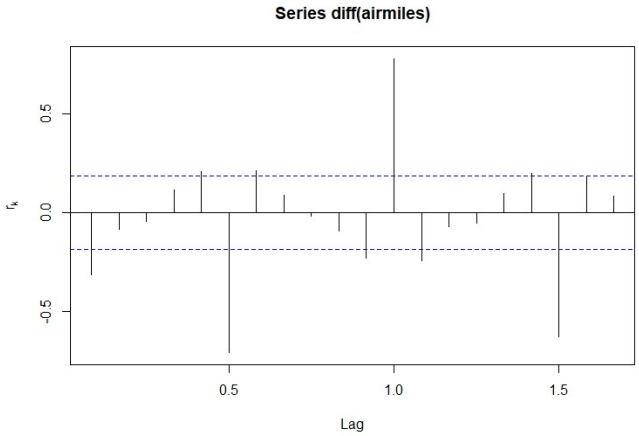
```
# (b) Plot the first difference of the time series. What improvements do you see here?
#   Is there still something in this data that needs to be accounted for?
plot(diff(airmiles))
```



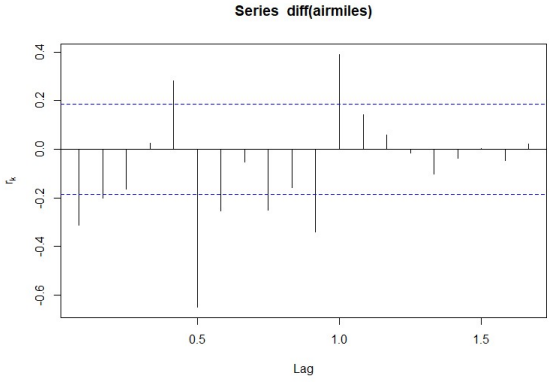
Yes there is an improvement. However, there is a big plunge near 2002.

(c) Create the sample ACF plot, sample PACF plot and sample EACF table for the First difference of the time series. Explain any conclusions you can make from each of these visualizations. If you find that it is difficult to make a single overall conclusion about the underlying model, explain why it is difficult.

```
16 # (c) Create the sample ACF plot, sample PACF plot and sample EACF table for the
17 # first difference of the time series. Explain any conclusions you can make from
18 # each of these visualizations. If you find that it is difficult to make a single overall
19 # conclusion about the underlying model, explain why it is difficult.
20 acf(diff(airmiles), ylab=expression(r[k]))
```




```
22 pacf(diff(airmiles), ylab=expression(r[k]))
```



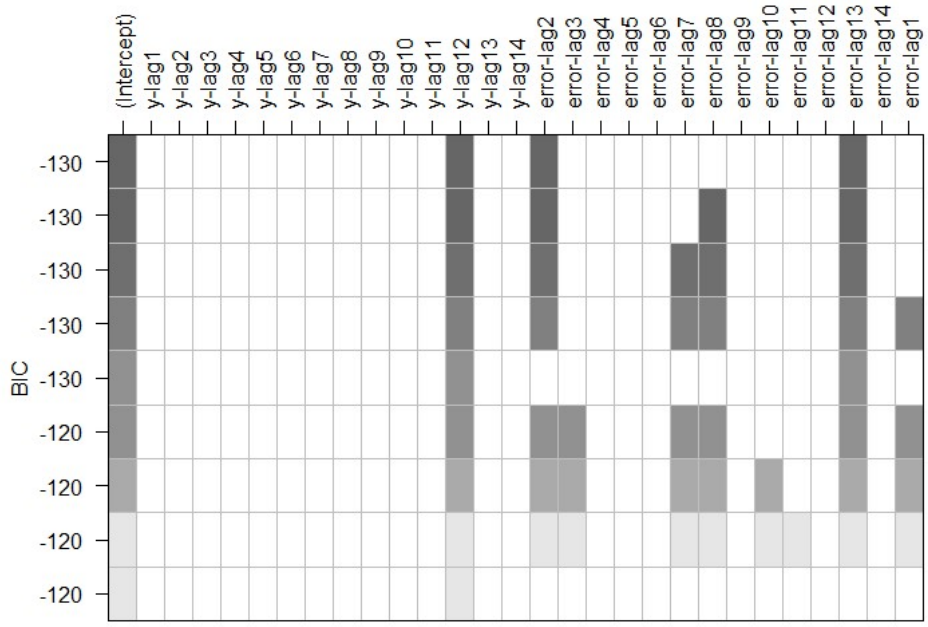

```
24 eacf(diff(airmiles))
```

```
> eacf(diff(airmiles))
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x o o o x x x o o o x  x  x  o
1 x o o o o x x o o o o  x  x  o
2 x o o o o x x o o o o  x  x  o
3 o x o o o x x o o o o  x  o  o
4 o x o o o x o o o o o  x  o  o
5 x o o x x x x o o o x  x  x  o
6 x o x o x x o o x o o  x  o  o
7 o o x o o x o o x o o  x  o  o
```

For ACF and EACF the k is quite big, so it is hard for me to choose a model.

(d) Sometimes this type of data is seasonal. So, perhaps the difficulty in choosing an appropriate ARMA model is due to the fact that this is a subset-ARMA model, i.e. some of the coefficients are zero. Create the \best subset-ARMA selection" plot.
(Hint: An example of how to create this plot is shown in Video 28. You can ignore the Warning message that R gives you.)

```
# (d) Sometimes this type of data is seasonal. So, perhaps the difficulty in choosing an
# appropriate ARMA model is due to the fact that this is a subset-ARMA model,
# i.e. some of the coefficients are zero. Create the \best subset-ARMA selection plot.
# (Hint: An example of how to create this plot is shown in Video 28. You can ignore
# the Warning message that R gives you.)

#Select and evaluate the best subset ARMA models:
armasubsets.select = armasubsets(y=diff(airmiles), nar = 14, nma = 14, y.name = 'y')

#Plot the subset ARMA selection
plot(armasubsets.select)
```



(e) Interpret the best subset-ARMA selection plot you created in part (d). What does it generally tell you about the \most important" terms in the model?

➔ It is saying that the model might be seasonal because only some of them are used. The most important terms are Y 12, and e2, e13.

(f) Write out the top choice of model based on the best subset-ARMA selection plot

in part (d).
(Hint: Remember that this ARMA model is for the first difference of the series, $\{W_t\}$, not for the original series $\{Y_t\}$.)

➔ Wt = Wt-12 + et + et-4 + et-13

(g) Does this choice of model make sense to you? Why do you think these lags were chosen to be the most important?

➔ May be the those months affect most significantly to the model.

(h) Since this ARMA model was constructed for the first difference of the time series, write out the equation for the original series, $\{Y_t\}$.

Yt = Yt-12-Yt-13 + et + et-4 +et-13