Homework 10

Daniel Caley

8/29/2021

Contents

The homework for week 10 is exercises 2, 5, 6, 7, and 8 on pages 272 and 273.

Question 2

Download and library the `nlme` package and use data `("Blackmore")` to activate the `Blackmore` data set. Inspect the data and create a box plot showing the exercise level at different ages. Run a repeated measures `ANOVA` to compare exercise levels at ages 8, 10, and 12 using `aov()`. You can use a command like, `myData <-Blackmore[Blackmore$age <=12,]`, to subset the data. Keeping in mind that the data will need to be balanced before you can conduct this analysis, try running a command like this, `table(myData$subject,myData$age)`, as the starting point for cleaning up the data set.

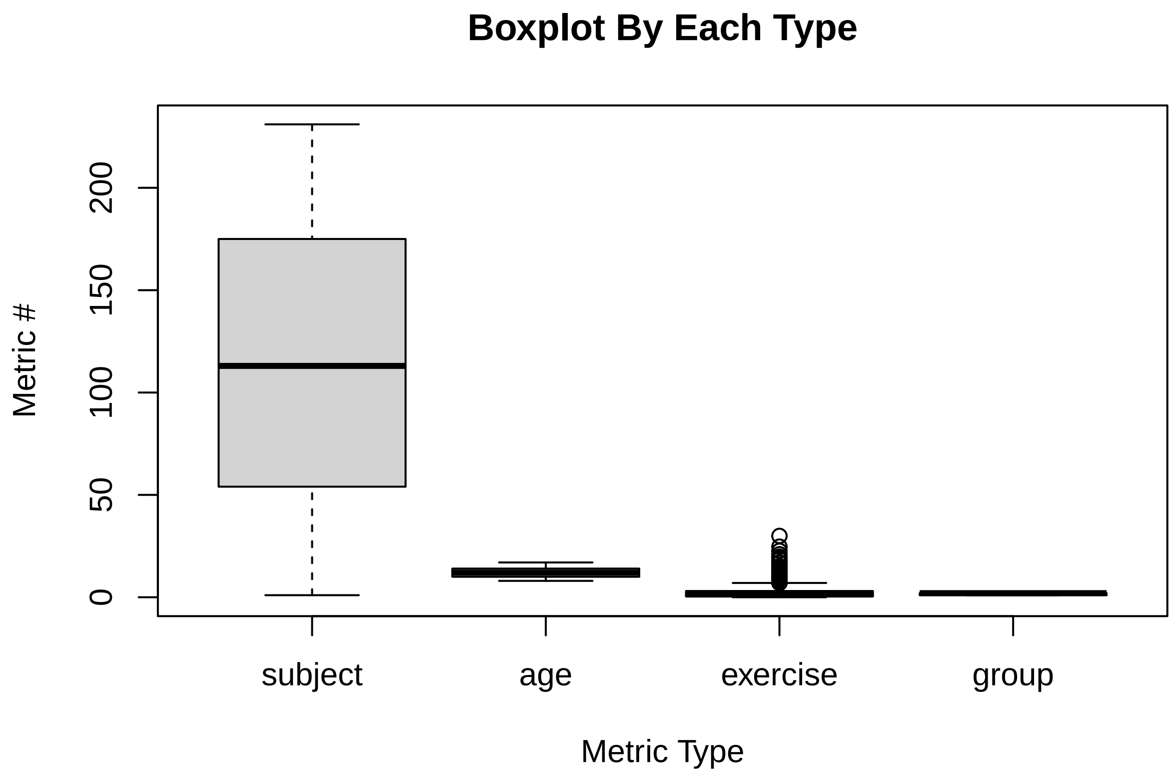
```
MyBlackmore <- Blackmore %>% mutate(age = floor(age))

str(MyBlackmore)
```

```
## 'data.frame':    945 obs. of  4 variables:
##  $ subject : Factor w/ 231 levels "100","101","102",..: 1 1 1 1 1 2 2 2 2 2 ...
##  $ age     : num  8 10 12 14 15 8 10 12 14 16 ...
##  $ exercise: num  2.71 1.94 2.36 1.54 8.63 0.14 0.14 0 0 5.08 ...
##  $ group   : Factor w/ 2 levels "control","patient": 2 2 2 2 2 2 2 2 2 2 ...
```

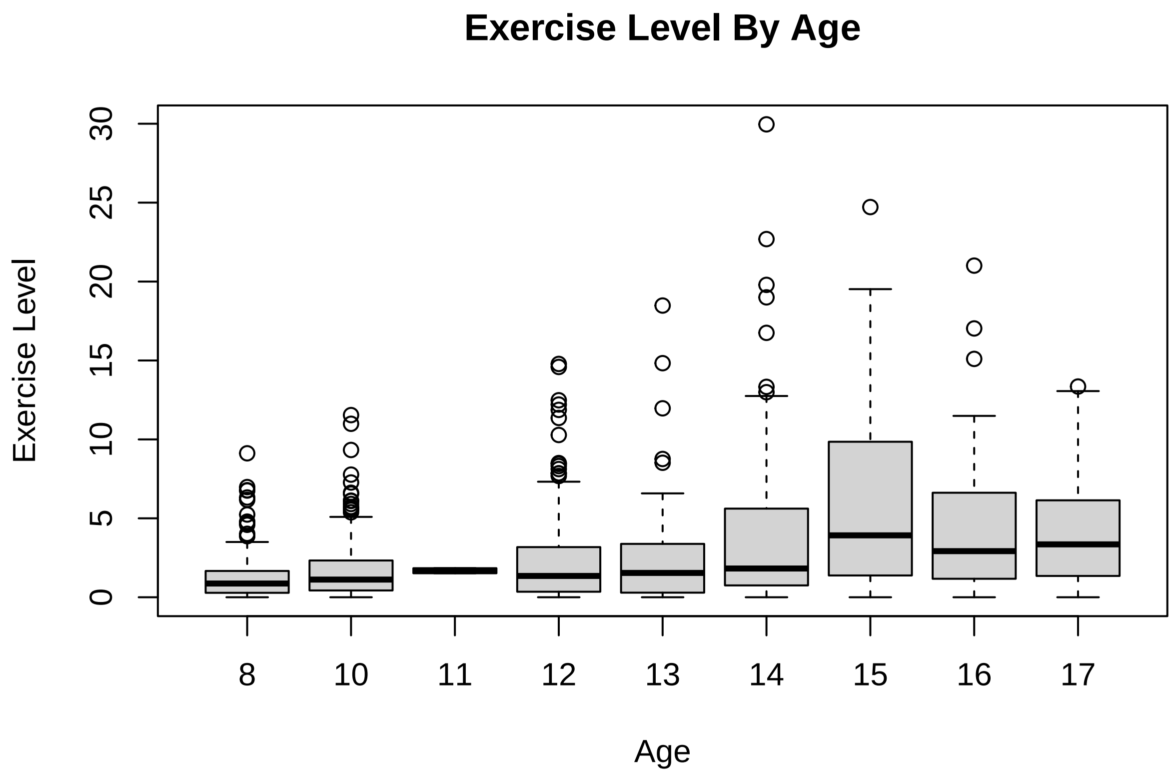
```
summary(MyBlackmore)
```

```
##     subject         age           exercise          group
##  100    :  5   Min.   : 8.00   Min.   : 0.000   control:359
##  101    :  5   1st Qu.:10.00   1st Qu.: 0.400   patient:586
##  105    :  5   Median :12.00   Median : 1.330
##  106    :  5   Mean   :11.34   Mean   : 2.531
##  107    :  5   3rd Qu.:14.00   3rd Qu.: 3.040
##  108    :  5   Max.   :17.00   Max.   :29.960
##  (Other):915
```

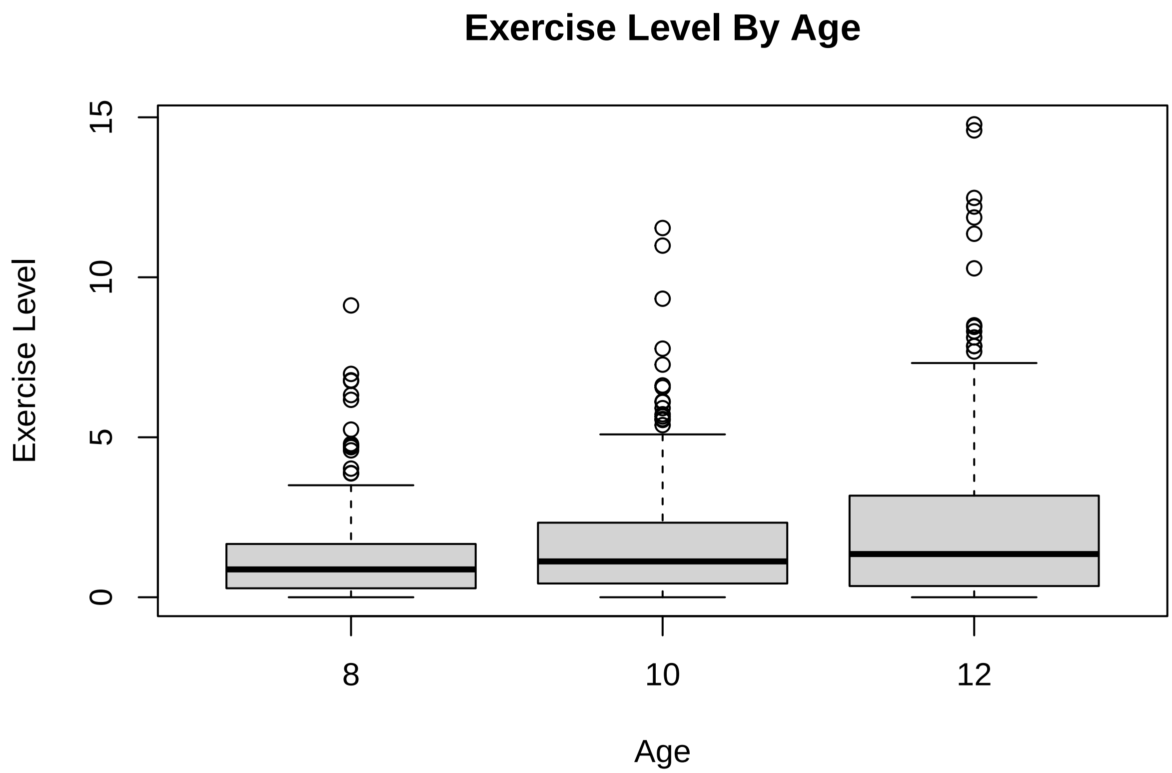
```
boxplot(MyBlackmore, main= "Boxplot By Each Type", xlab="Metric Type",
        ylab="Metric #")
```




```
boxplot(exercise~age, MyBlackmore, main= "Exercise Level By Age", xlab="Age",
        ylab="Exercise Level")
```




```
MyBlackmoreAges <- MyBlackmore[MyBlackmore$age %in% c(8,10,12),]


boxplot(exercise~age, MyBlackmoreAges,
        main= "Exercise Level By Age",
        xlab="Age",
        ylab="Exercise Level")
```




```
MyBlackmoreTable <-  table(MyBlackmoreAges$subject,MyBlackmoreAges$age)

MyBlackmoreAgesWide <- MyBlackmoreAges %>% select(-group) %>%
  pivot_wider(names_from = age, values_from = exercise)

summary(MyBlackmoreAgesWide)
```

```
##     subject          8               10               12
##  100    :  1   Min.   :0.000   Min.   : 0.000   Min.   : 0.000
##  101    :  1   1st Qu.:0.280   1st Qu.: 0.430   1st Qu.: 0.350
##  102    :  1   Median :0.870   Median : 1.120   Median : 1.350
##  103    :  1   Mean   :1.259   Mean   : 1.746   Mean   : 2.341
##  104    :  1   3rd Qu.:1.665   3rd Qu.: 2.330   3rd Qu.: 3.175
##  105    :  1   Max.   :9.120   Max.   :11.540   Max.   :14.780
##  (Other):225                   NA's   :2        NA's   :40
```

```
print("With Out Removing Rows with Missing Records")
```

```
## [1] "With Out Removing Rows with Missing Records"
```

```
summary(aov(exercise~age, data = MyBlackmoreAges))
```

```
##                Df Sum Sq Mean Sq F value   Pr(>F)
## age             1    122  121.97   27.01 2.72e-07 ***
```

```
## Residuals   649   2931    4.52
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
CompleteValues <- rowSums(MyBlackmoreTable)==3
CompleteValues <- CompleteValues[CompleteValues == TRUE]
CompleteValues <- as.numeric(names(CompleteValues))
```

```
## Warning: NAs introduced by coercion
```

```r
MyBlackmoreAges <- MyBlackmoreAges[MyBlackmoreAges$subject %in% CompleteValues,]


aovBlackmore <- aov(exercise~age+Error(subject),data = MyBlackmoreAges)
summary(aovBlackmore)
```

```
##
## Error: subject
##            Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 180   2010   11.17
##
## Error: Within
##            Df Sum Sq Mean Sq F value   Pr(>F)
## age         1  115.9  115.90   58.98 1.51e-13 ***
## Residuals 361  709.3    1.96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretting Results

The Median value ranges from 0.880 to 1.480 for ages 8 through 12. All of them on average is about 1. What can be seen here is that by age increases exercise also increases. This can be seen by looking identifying the change in the box plots for the 3rd quartiles. Rather the variation is increasing by age and not as consistent in younger ages. Further proof of this change is shown when looking at analysis of variance `aov` function. The p-value is 1.51e-13 which is less than 0.05, indicating that their is significant between the age and the amount of exercise they do per week.

Question 5

Given that the `AirPassengers` data set has a substantial growth trend, use `diff()` to create a difference data set. Use `plot()` to examine and interpret the results of `differencing`. Use `cpt.var()` to find the change point in the variability of the difference time series. Plot the result and describe in your own words what the change point signifies.

```
MyAirPassengers <- AirPassengers

str(MyAirPassengers)
```

```
##  Time-Series [1:144] from 1949 to 1961: 112 118 132 129 121 135 148 148 136 119 ...
```

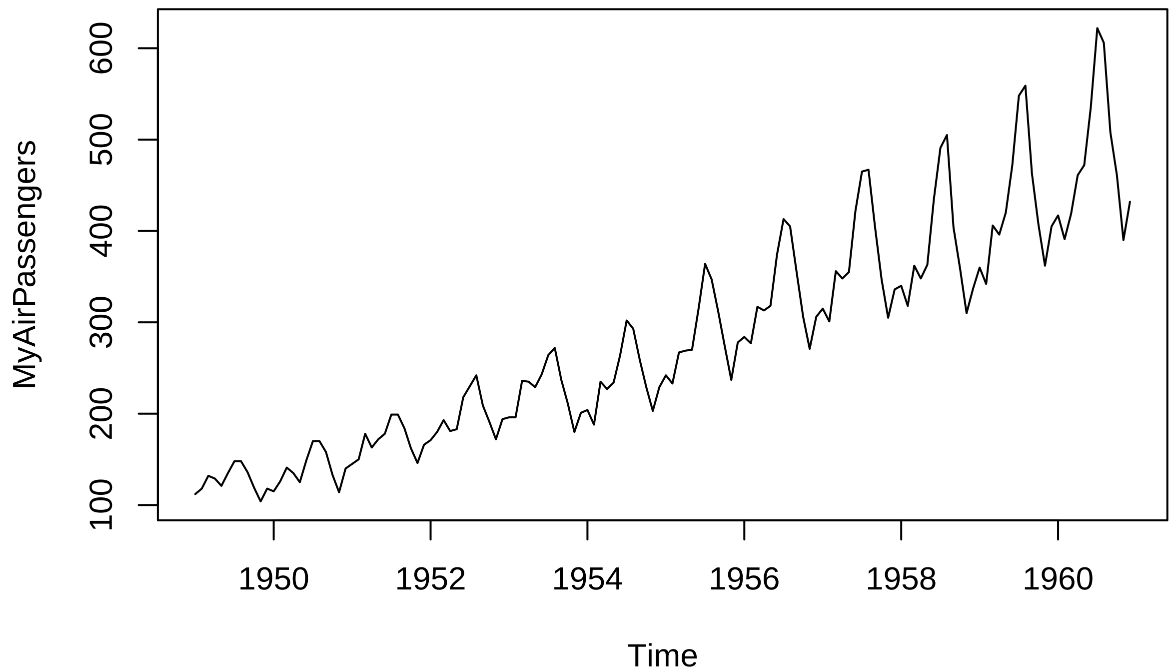
```
summary(MyAirPassengers)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   104.0   180.0   265.5   280.3   360.5   622.0
```

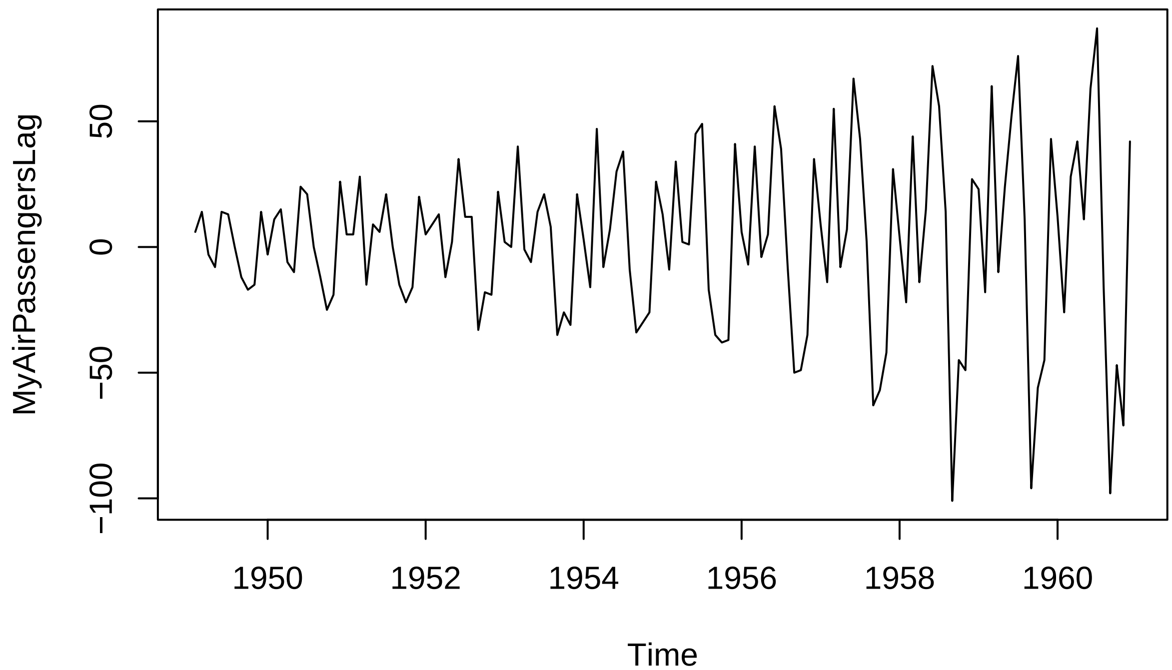
```
MyAirPassengersLag <- diff(MyAirPassengers)
MyAirPassengersCP <- cpt.var(MyAirPassengersLag)
MyAirPassengersCP
```

```
## Class 'cpt' : Changepoint Object
##        ~~   : S4 class containing 12 slots with names
##               cpttype date version data.set method test.stat pen.type pen.value minseglen cpts ncpts
##
## Created on  : Tue May 25 17:44:21 2021
##
## summary(.)  :
## ----------
## Created Using changepoint version 2.2.2
## Changepoint type      : Change in variance
## Method of analysis    : AMOC
## Test Statistic  : Normal
## Type of penalty       : MBIC with value, 14.88853
## Minimum Segment Length : 2
## Maximum no. of cpts   : 1
## Changepoint Locations : 76
```

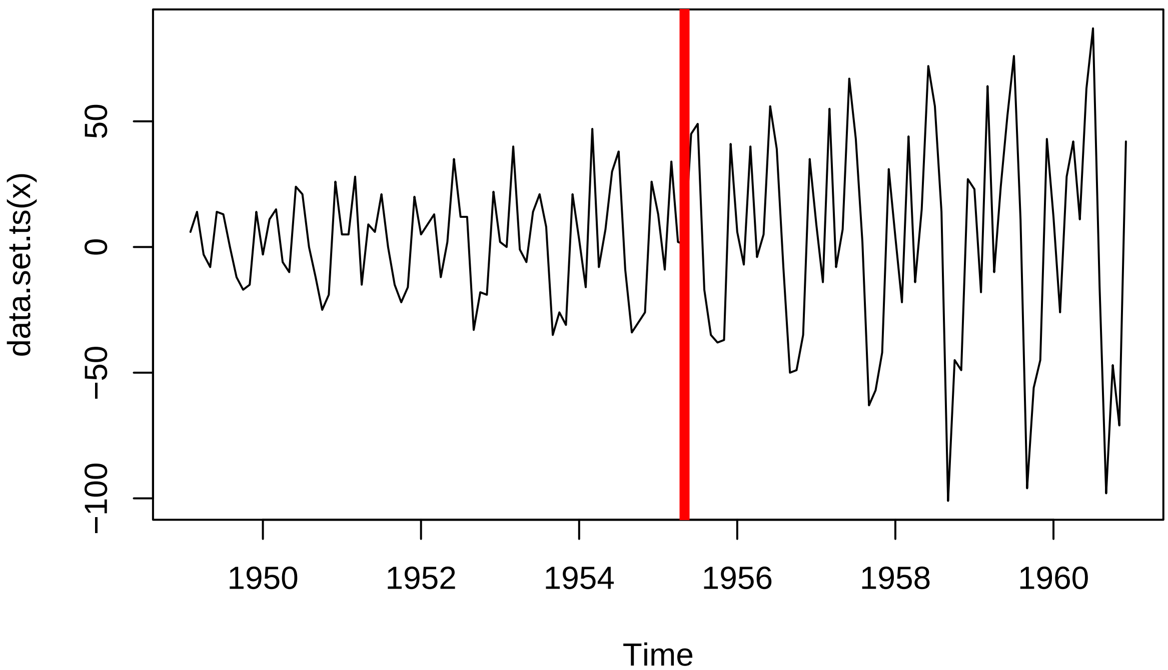
```
plot(MyAirPassengers)
```




```
plot(MyAirPassengersLag)
```




```
plot(MyAirPassengersCP,cpt.col="red",cpt.width=5)
```



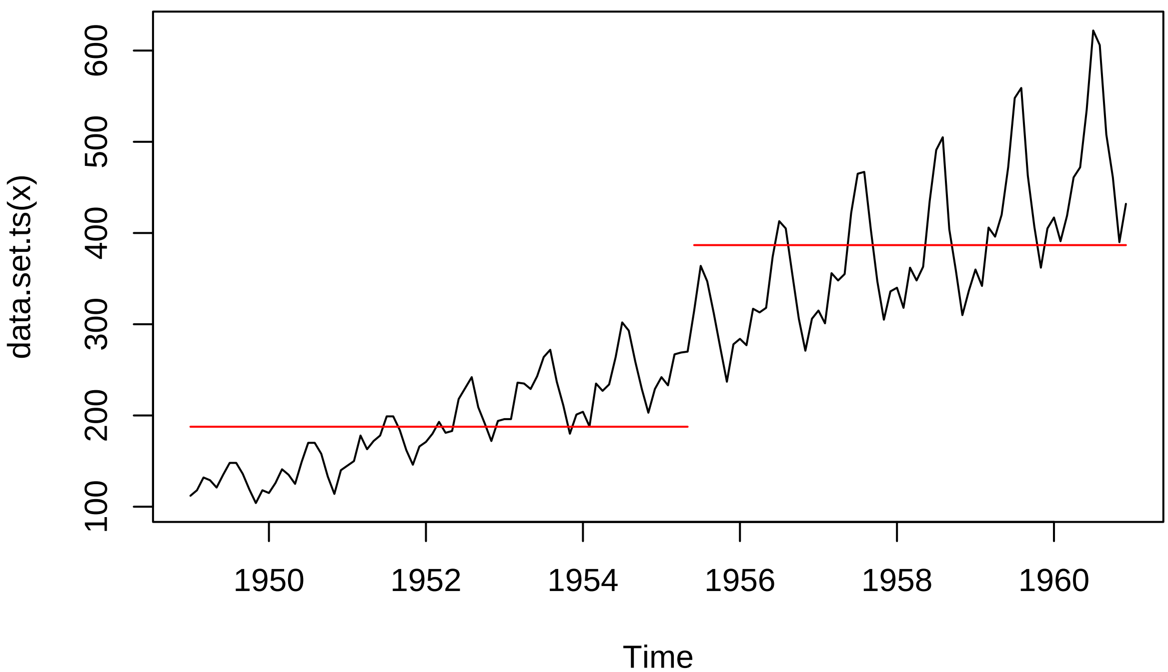
Interpretting the Results

The trend in the Airpassenger data set is positive, but also appears to have seasonality associated with the plots. As time goes on though the volatility increases. This is noticed by looking at the lag of the dataset, which is a way to compare the last period to the current period. The change point appears to have occurs sometime after 1955 based on the plot above. Where a significant amount of passengers increased.

Question 6

Use `cpt.mean()` on the `AirPassengers` time series. Plot and interpret the results. Compare the change point of the mean that you uncovered in this case to the change point in the variance that you uncovered in Exercise 5. What do these change points suggest about the history of air travel?

```
airLagMean <-  cpt.mean(MyAirPassengers)
plot(airLagMean)
```



Intepretting the Results

Using the cpt.mean function shows that the confidence value is a 1 meaning that there has been a shift in the mean over time.

Question 7

Find historical information about air travel on the Internet and/or in reference materials that sheds light on the results from Exercises 5 and 6. Write a mini-article (less than 250 words) that interprets your statistical findings from Exercises 5 and 6 in the context of the historical information you found.

Mini-Article

In 1950 commercial jet airliners begin taking the sky kicking off the wave of traveling for business or pleasure. Prior to this period planes were used for military purposes and little travel was taken as cabins were quite load with the propellers creating so much noise. The advent of jet engines meant farther and quicker travel. The industry couldn't stay in war times and had to adapt for commercial use.

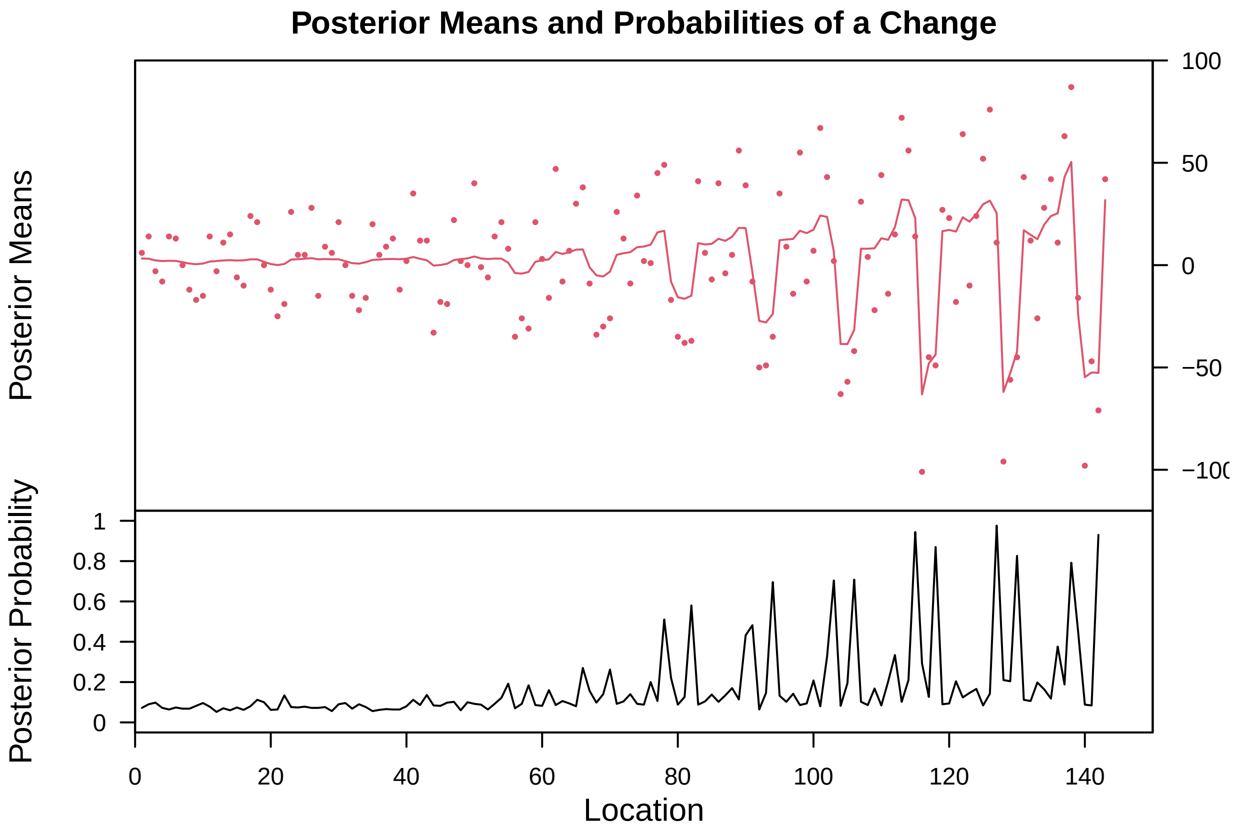
In just 5 short years the technology really booms and airtravel becomes a thing for many. As more competition enters the market and the middle class begins to grow along with airliners lowering their prices and increased wages allowed consumers to take to the sky.

referencing the Flying Boat Museum Website

Question 8

Use `bcp()` on the `AirPassengers` time series. Plot and interpret the results. Make sure to contrast these results with those from Exercise 6.

```
MyAirBCP <-  bcp(MyAirPassengersLag)

plot(MyAirBCP)
```



Interpretting the Results

Looking at the time series data through a Bayesian approach and specifically the Posterior Probability, shows that that the probability increased around 70 months but was still quite low. By month 110 or roughly 10 years the probability increased to 1. The Posterior Means also shows that the shifted around 70 months as the data become wider as the cumulative months go on.