George_Smith_HW10_IST772

George Smith

9/11/2021

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.

```
library(car)
```

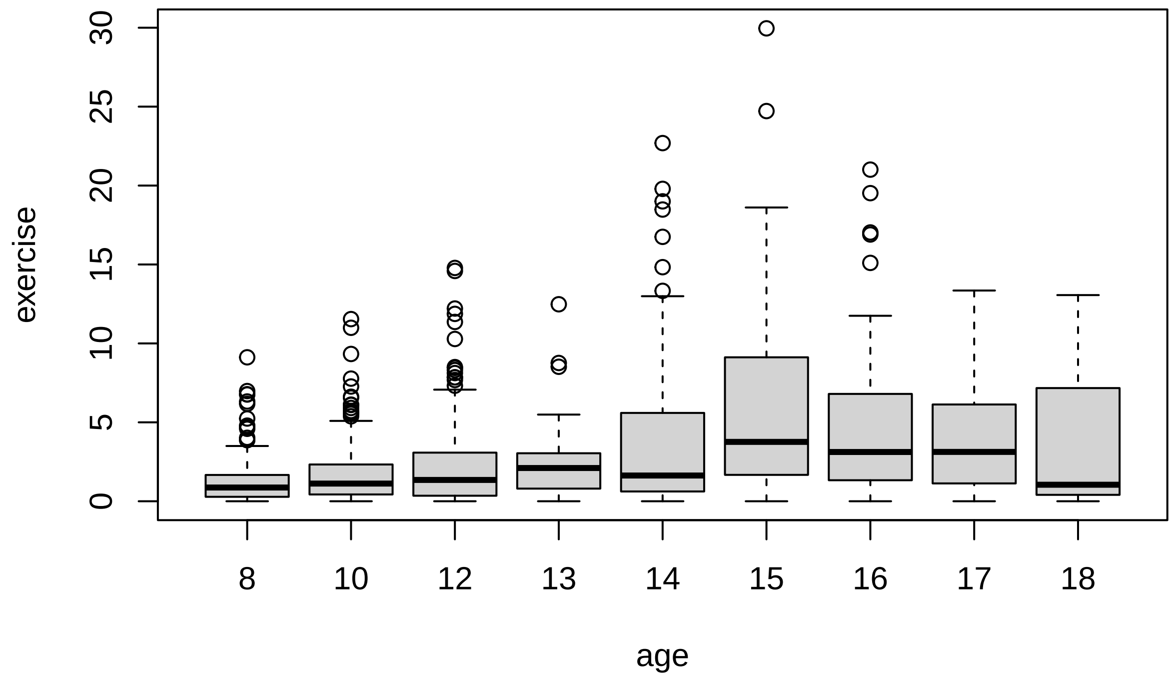
```
## Warning: package 'car' was built under R version 4.1.1
```

```
## Loading required package: carData
```

```
data("Blackmore")
bmDF <- data.frame(Blackmore)
bmDF$age <- round(bmDF$age)
summary(bmDF)
```

```
##     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   :11.43   Mean   : 2.531
##  107    :  5   3rd Qu.:14.00   3rd Qu.: 3.040
##  108    :  5   Max.   :18.00   Max.   :29.960
##  (Other):915
```

```
boxplot(exercise~age, data = bmDF)
```




```r
subBM <- bmDF[bmDF$age <= 12,]
subBM$ageFact <- as.factor(subBM$age)
list <- rowSums(table(subBM$subject,subBM$ageFact))==3
list <- list[list == TRUE]
list <- as.numeric(names(list))
```

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

```r
summary(subBM[subBM$ageFact == 8,])
```

```
##     subject         age        exercise          group      ageFact
##  100    :  1   Min.   :8   Min.   :0.000   control: 93   8 :231
##  101    :  1   1st Qu.:8   1st Qu.:0.280   patient:138   10:  0
##  102    :  1   Median :8   Median :0.870                 12:  0
##  103    :  1   Mean   :8   Mean   :1.259
##  104    :  1   3rd Qu.:8   3rd Qu.:1.665
##  105    :  1   Max.   :8   Max.   :9.120
##  (Other):225
```

```r
summary(subBM[subBM$ageFact == 10,])
```

```
##     subject         age         exercise           group      ageFact
##  100    :  1   Min.   :10   Min.   : 0.000   control: 92   8 :  0
##  101    :  1   1st Qu.:10   1st Qu.: 0.430   patient:137   10:229
```

```
##  102     :  1   Median :10   Median : 1.120                  12:  0
##  103     :  1   Mean   :10   Mean   : 1.746
##  104     :  1   3rd Qu.:10   3rd Qu.: 2.330
##  105     :  1   Max.   :10   Max.   :11.540
##  (Other):223
```

```
summary(subBM[subBM$ageFact == 12,])
```

```
##     subject         age         exercise          group      ageFact
##  100     :  1   Min.   :12   Min.   : 0.000   control: 68   8 :  0
##  101     :  1   1st Qu.:12   1st Qu.: 0.350   patient:121   10:  0
##  102     :  1   Median :12   Median : 1.350                 12:189
##  103     :  1   Mean   :12   Mean   : 2.289
##  104     :  1   3rd Qu.:12   3rd Qu.: 3.080
##  105     :  1   Max.   :12   Max.   :14.780
##  (Other):183
```

```
subBM <- subBM[subBM$subject %in% list,]
summary(aov(exercise~ageFact+ Error(subject), data = subBM))
```

```
##
## Error: subject
##            Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 176   1931   10.97
##
## Error: Within
##            Df Sum Sq Mean Sq F value   Pr(>F)
## ageFact     2  105.2   52.60   27.82 6.09e-12 ***
## Residuals 352  665.7    1.89
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Analysis

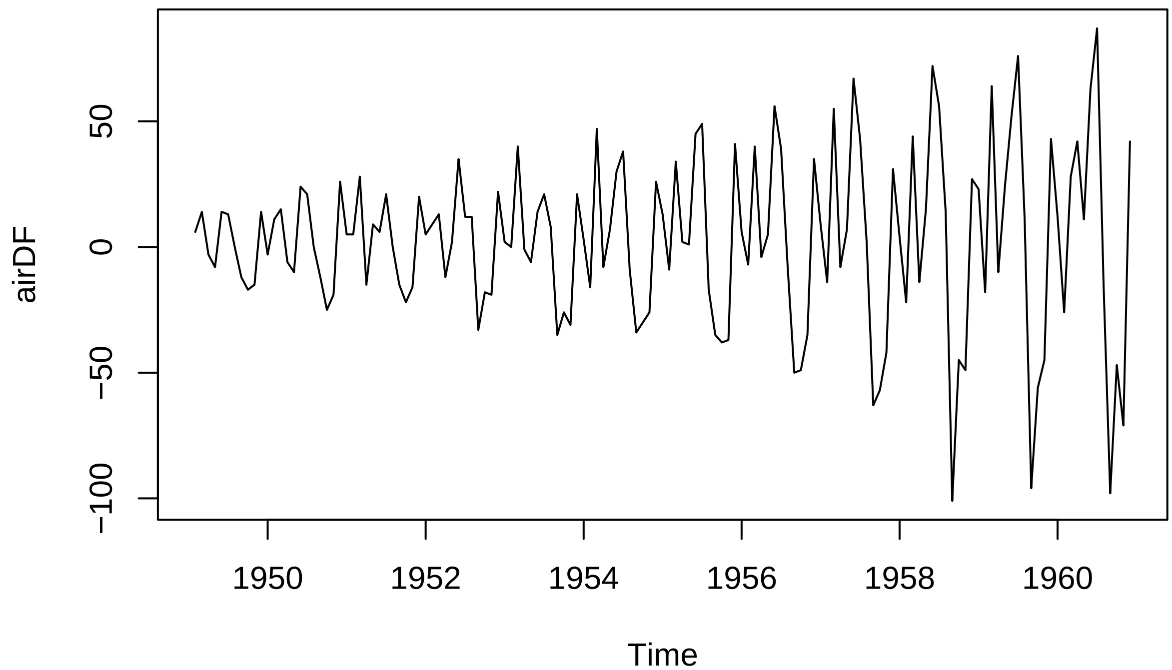
The 3rd and 4th quantiles, and the maximum value increase as age increases. This Means that on average, the age of teenagers who have yet to reach adolscence average an hour of excercise a week, but as the age increases the higher quantiles become more varied. When analyzing the variance using the aov function. The p-value,6.09e-12, indiicates that there is a signigicant difference between the age factors in regards to the amount of excercise they do in a week.

5. Given that the AirPassengers data set has a substantial growth trend, use diff() to cre-ate a differenced 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 differenced time series.Plot the result and describe in your own words what the change point signifies.

```
library(changepoint)
```

```
## Warning: package 'changepoint' was built under R version 4.1.1

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Successfully loaded changepoint package version 2.2.2
##  NOTE: Predefined penalty values changed in version 2.2.  Previous penalty values with a postfix 1 i
```

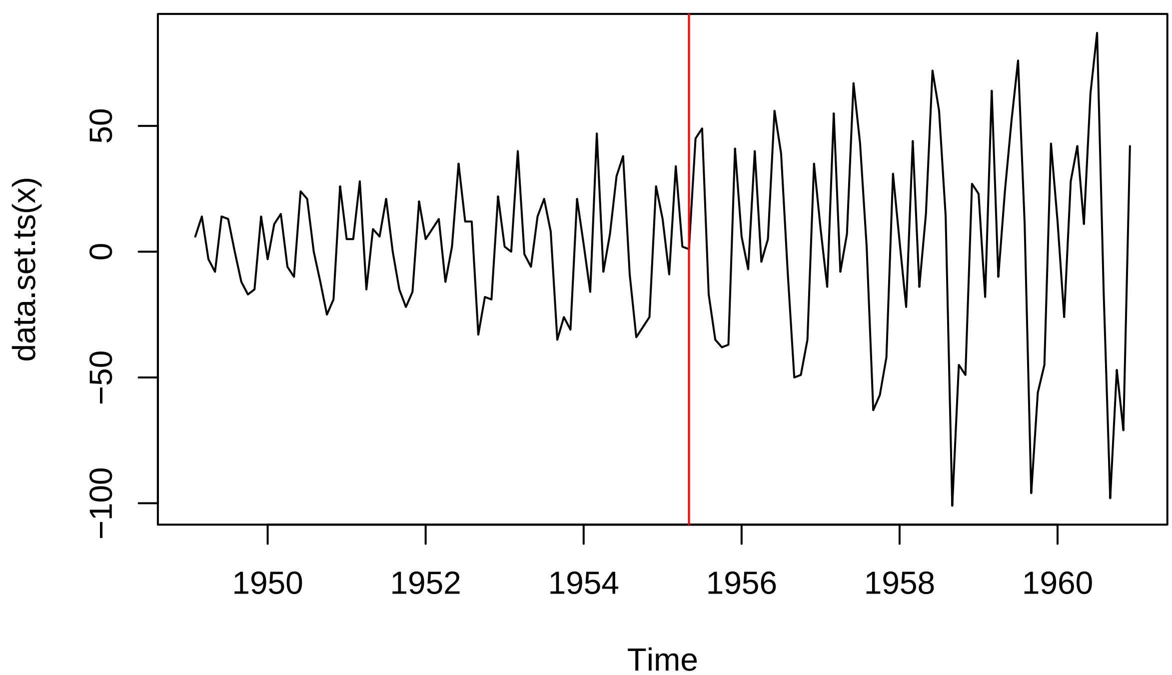
```
data("AirPassengers")
airDF <- diff(AirPassengers)
plot(airDF)
```




```
cpt.var(airDF)
```

```
## 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  : Fri Sep 03 15:14:26 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(cpt.var(airDF))
```

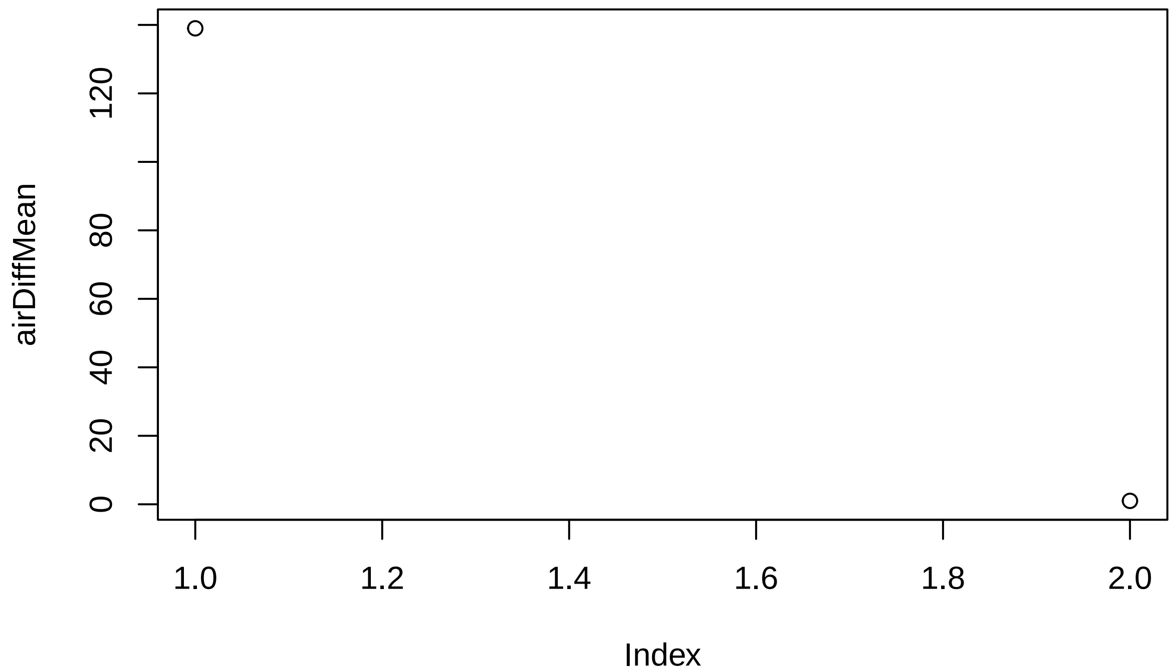


Analysis

The diff(Airpassenger) function evaluates the differencein time between current month vs previous month.The redline in the above plot indicates the changepoint of the dataset. The changepoint will search through the data and find where a major shift occured in the mean level of the data. In this case, the major shift occured in June 1955 where the amount of international passengers increased by 45,000 in one month.

6. Use cpt.mean() on the AirPassengers time series. Plot and interpret the results. Com-pare 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?

```
airDiffMean <- cpt.mean(airDF, class = FALSE)
plot(airDiffMean)
```




```
airDiffMean["conf.value"]
```

```
## conf.value
##          1
```

Analysis

Plotting the cpt.mean of the data,and viewing confidence level, indicates that there has been a shift in the mean over time.

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.

#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.

Analysis:

The early 1900s are often referred to as The Golden Age of Air travel. Air travel was viewed as a luxury and passengers, first class or not, were treated similarly. Yet, by the 1950s, this trend began to shift. Air travel became seen as a necessity rather than a luxury. While in previous generations, international travellers would take journey's on ships by 1955 air passengers increased by 19% while sea passengers only increased by 4%. By the changepoint, June 1955, Air travel finally exceeded sea travel in Europe by 7,000 passengengers.

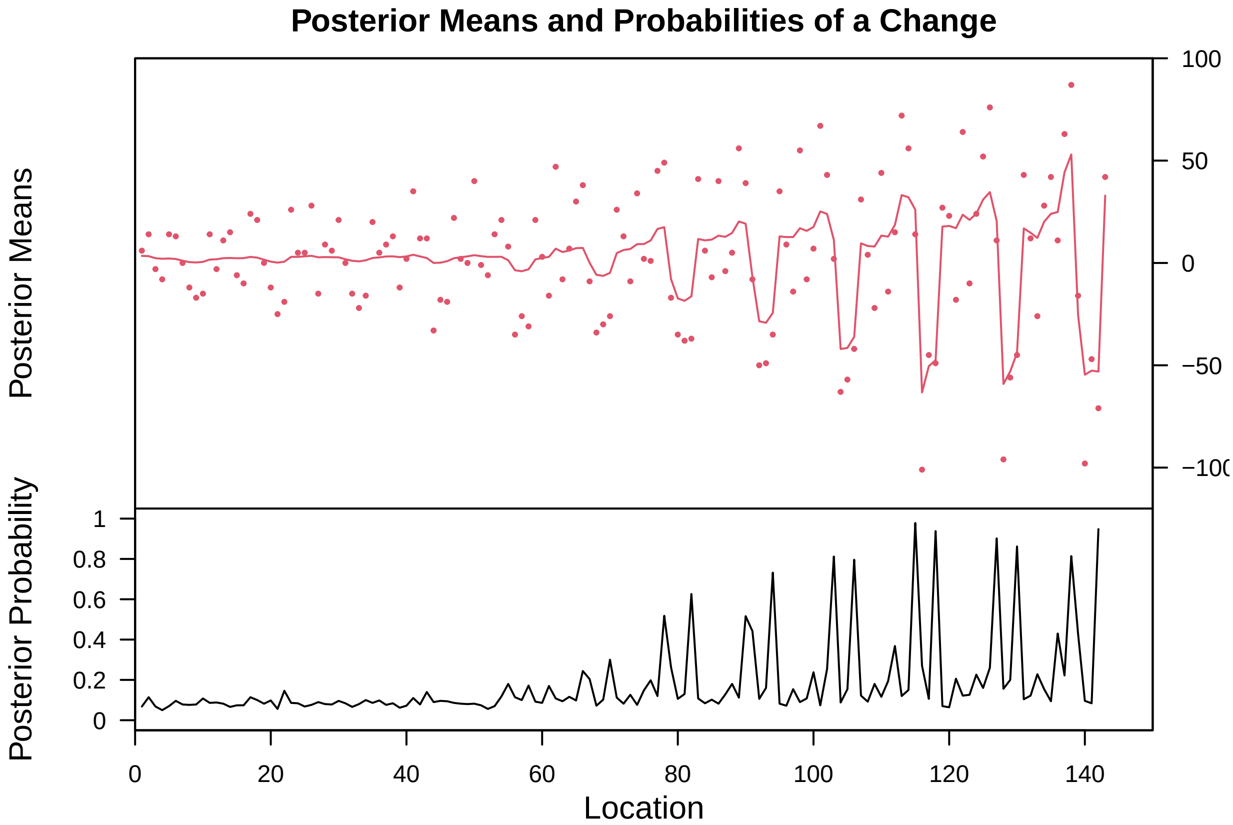
8. Use bcp() on the AirPassengers time series. Plot and interpret the results. Make sure to contrast these results with those from Exercise 6.

```
library(bcp)
```

```
## Warning: package 'bcp' was built under R version 4.1.1
```

```
## Loading required package: grid
```

```
bcpAD <- bcp(as.vector(airDF))
plot(bcpAD)
```



Analysis

Interpreting the time series with a bayesian approach seems to net the same results. The posterior probability and the posterior mean plots show a shift in probablity and mean difference around the 76th location, The 76th was the changepoint in frequentist method of analyzing the data, so it can be inferred that this is definitely a point of emphasis and a real shift occured in the data at this point.