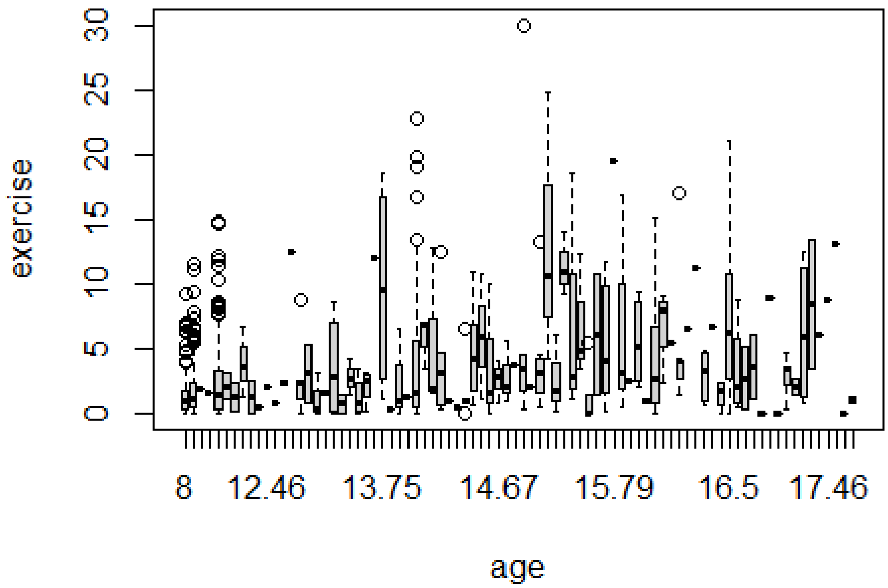
HW10_Videtti

```
##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(nlme)
library(car)

## Loading required package: carData

data("Blackmore")
str(Blackmore)

## '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 15.9 ...
##  $ 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
...

head(Blackmore)

##   subject   age exercise   group
## 1     100  8.00     2.71 patient
## 2     100 10.00     1.94 patient
## 3     100 12.00     2.36 patient
## 4     100 14.00     1.54 patient
## 5     100 15.92     8.63 patient
## 6     101  8.00     0.14 patient

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




```
myData <-Blackmore[Blackmore$age %in% c(8,10,12),]
table(myData$subject,myData$age)

##
##        8 10 12
##   100  1  1  1
##   101  1  1  1
##   102  1  1  1
##   103  1  1  1
##   104  1  1  1
##   105  1  1  1
##   106  1  1  1
##   107  1  1  1
##   108  1  1  1
##   109  1  1  0
##   110  1  1  1
##   111  1  1  1
##   112  1  1  1
##   113  1  1  1
##   114  1  1  1
##   115  1  1  1
##   116  1  1  1
##   117  1  1  1
##   118  1  1  1
##   119  1  1  1
##   120  1  1  1
```

```
##   121  1  1  1
##   122  1  1  1
##   123  1  1  1
##   124  1  1  1
##   125  1  1  1
##   126  1  1  1
##   127  1  1  1
##   128  1  1  0
##   129  1  1  1
##   130  1  1  1
##   132  1  1  0
##   133  1  1  1
##   134  1  1  1
##   135  1  1  1
##   136  1  1  1
##   137  1  1  1
##   138  1  1  1
##   139  1  1  1
##   140  1  1  1
##   141  1  1  1
##   142  1  1  1
##   143  1  1  1
##   144  1  1  1
##   145  1  1  1
##   146  1  1  1
##   147  1  1  0
##   148  1  1  1
##   149  1  1  0
##   150  1  1  1
##   151  1  1  1
##   152  1  1  1
##   153  1  1  1
##   154  1  1  1
##   155  1  1  0
##   156  1  1  1
##   157  1  1  1
##   158  1  1  0
##   159  1  1  1
##   160  1  1  1
##   161  1  1  1
##   162  1  1  0
##   163  1  1  1
##   164  1  1  1
##   165  1  1  1
##   166  1  1  1
##   167  1  1  1
##   168  1  0  0
##   169  1  1  0
##   170  1  1  1
##   171  1  1  1
```

```
##    172  1   1   1
##    173  1   1   1
##    174  1   1   0
##    175  1   1   1
##    176  1   1   1
##    177  1   1   1
##    178  1   1   1
##    179  1   1   1
##    180  1   1   1
##    181  1   1   1
##    182  1   1   1
##    183  1   1   1
##    184  1   1   0
##    185  1   1   1
##    186  1   1   1
##    187  1   1   1
##    188  1   1   1
##    189  1   1   1
##    190  1   1   1
##    192  1   1   1
##    193  1   1   0
##    194  1   1   1
##    195  1   1   1
##    196  1   1   0
##    198  1   1   1
##    199  1   1   1
##    200  1   1   1
##    201  1   1   0
##    202  1   1   1
##    203  1   1   1
##    204  1   1   1
##    205  1   1   1
##    206  1   1   0
##    207a 1   1   1
##    207b 1   1   1
##    208  1   1   1
##    209  1   1   1
##    210  1   1   1
##    211  1   1   0
##    212  1   1   1
##    213  1   1   1
##    214  1   1   1
##    215  1   1   1
##    216  1   1   0
##    217  1   1   0
##    218  1   1   1
##    219  1   1   1
##    220  1   1   1
##    221  1   1   0
##    222  1   0   0
```

```
##    223  1  1  0
##    224  1  1  1
##    225  1  1  1
##    226  1  1  0
##    227  1  1  0
##    228  1  1  1
##    229a 1  1  1
##    229b 1  1  1
##    230  1  1  1
##    231  1  1  1
##    232  1  1  1
##    233  1  1  1
##    234  1  1  0
##    235  1  1  0
##    236  1  1  1
##    237  1  1  1
##    238  1  1  1
##    239  1  1  0
##    240  1  1  0
##    241  1  1  0
##    242  1  1  0
##    243  1  1  1
##    244  1  1  1
##    245  1  1  1
##    246  1  1  1
##    247  1  1  1
##    248  1  1  0
##    249  1  1  1
##    250  1  1  1
##    251  1  1  1
##    252  1  1  0
##    253  1  1  1
##    254  1  1  0
##    255  1  1  1
##    255b 1  1  1
##    256  1  1  1
##    257  1  1  1
##    258  1  1  0
##    259  1  1  1
##    260  1  1  0
##    261  1  1  0
##    262  1  1  0
##    263  1  1  0
##    264  1  1  0
##    265  1  1  0
##    266  1  1  0
##    267  1  1  1
##    268  1  1  1
##    269  1  1  0
##    270  1  1  1
```

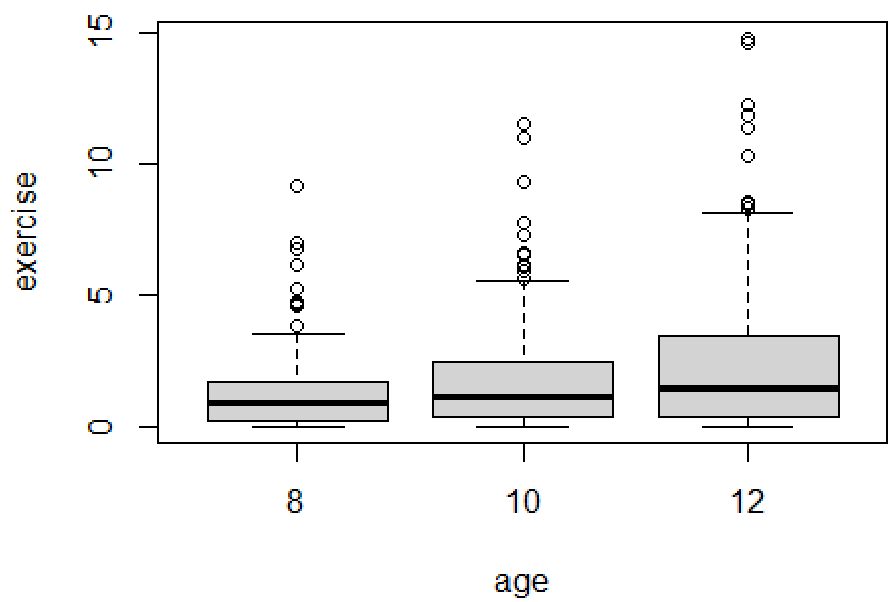
```
##    271  1  1  1
##    272  1  1  0
##    273a 1  1  1
##    273b 1  1  1
##    274  1  1  0
##    275  1  1  1
##    276  1  1  1
##    277  1  1  1
##    278  1  1  0
##    279a 1  1  0
##    279b 1  1  1
##    280a 1  1  1
##    280b 1  1  1
##    281  1  1  1
##    282  1  1  1
##    283  1  1  0
##    284  1  1  1
##    285  1  1  0
##    286  1  1  1
##    300  1  1  1
##    301  1  1  1
##    302  1  1  1
##    303  1  1  1
##    304  1  1  1
##    305  1  1  1
##    306  1  1  1
##    307  1  1  1
##    308  1  1  1
##    309  1  1  1
##    310  1  1  1
##    311  1  1  1
##    312  1  1  1
##    313  1  1  1
##    314  1  1  1
##    315  1  1  0
##    316  1  1  0
##    317  1  1  1
##    318  1  1  1
##    319  1  1  1
##    320  1  1  1
##    321  1  1  1
##    322  1  1  0
##    323  1  1  1
##    324  1  1  0
##    325  1  1  1
##    326  1  1  1
##    327  1  1  1
##    328  1  1  1
##    329  1  1  0
##    330  1  1  1
```

```
##   331  1  1  1
##   332  1  1  1
##   333  1  1  1
##   334  1  1  1
##   335  1  1  1
##   336  1  1  1
##   337  1  1  0
##   338  1  1  1
##   340  1  1  1
##   341  1  1  0
```

```
list <- rowSums(table(myData$subject,myData$age)) == 3
list <- as.numeric(names(list[list == TRUE]))
```

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

```
Exercise2 <- myData[myData$subject %in% list,]
boxplot(exercise~age,data = Exercise2)
```

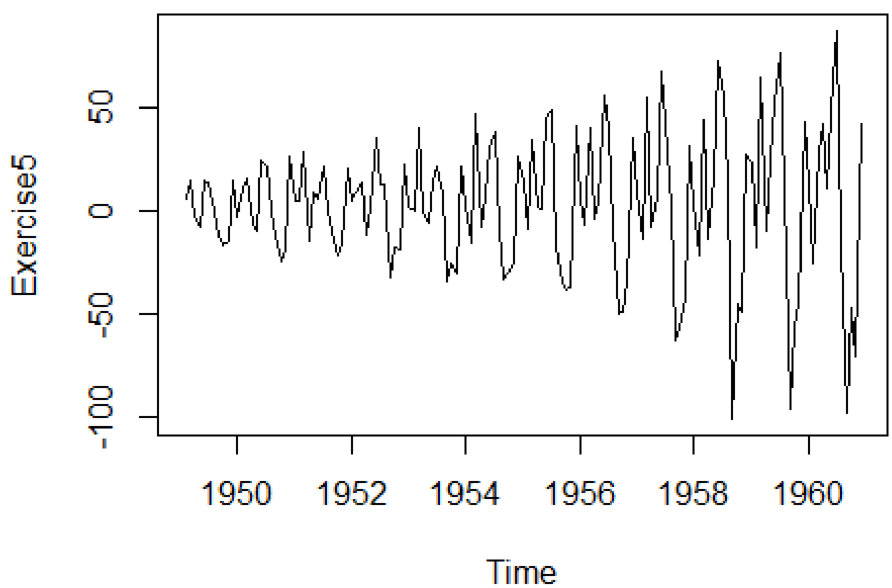



```
summary(aov(exercise ~ age + Error(subject), data = Exercise2))
```

```
##
## Error: subject
##            Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 165   1892   11.47
##
## Error: Within
```

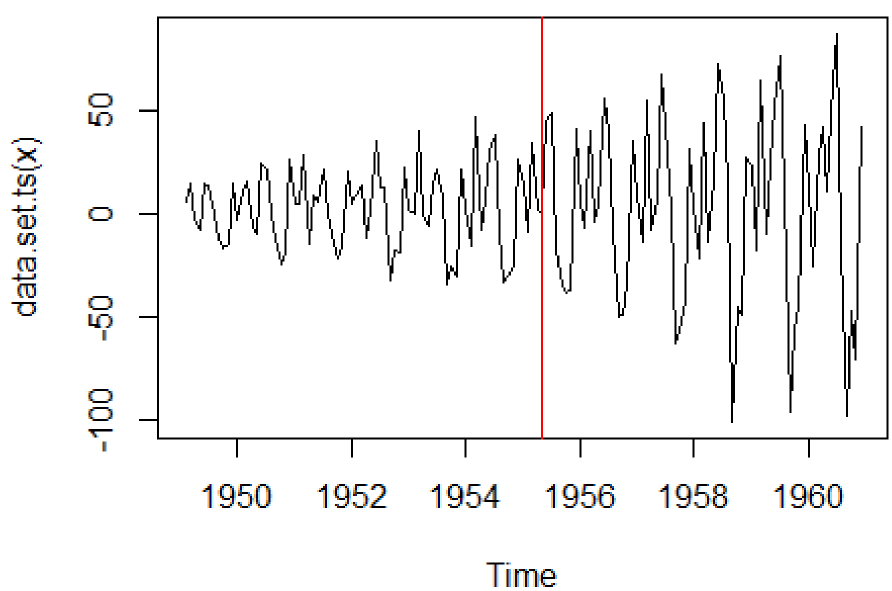
```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## age          1  102.7  102.66   54.23 1.43e-12 ***
## Residuals 331  626.6    1.89                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##5. Given that the AirPassengers data set has a substantial growth trend,
use diff() to create 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.
Exercise5 <- diff(AirPassengers)
plot(Exercise5)
```




```
#We see in this plot that month-to-month changes in airline passengers
increases in magnitude gradually over time between 1949 and 1960

library(changepoint)

## 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 2.2.3
##  See NEWS for details of changes.

cpt.var(Exercise5)

## 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.max param.est
##
## Created on  : Sun May 29 03:33:04 2022
##
## summary(.)  :
## ----------
## Created Using changepoint version 2.2.3
## 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(Exercise5))
```

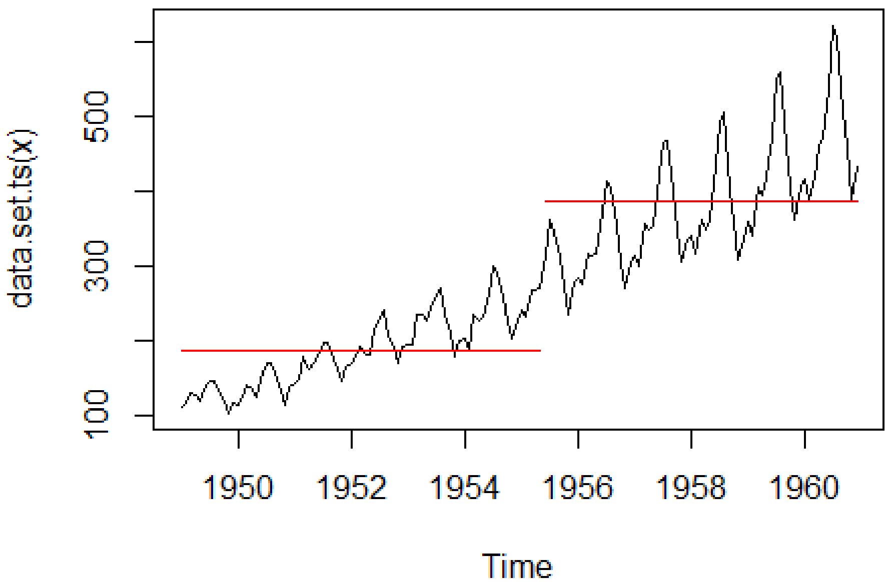



```
#This change point is at 76, meaning the 76th month after the start, or 6
years and 4 months after January 1949, or April of 1955. This was found by
```

looking for the an instance of a change in variance that results in the MBIC > 14.88853, given the type of test statistic we used (in this case, the default of "Normal"). This means our test found that between 1949 and 1960, April 1955 was the first month with a month-to-month change in the variance that was significant enough to reach that MBIC penalty. Our red line in the plot confirms that the change point is at about April 1955.

##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?

```
cpt.mean(AirPassengers)

## 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.max param.est
##
## Created on  : Sun May 29 03:33:04 2022
##
## summary(.)  :
## ----------
## Created Using changepoint version 2.2.3
## Changepoint type      : Change in mean
## Method of analysis    : AMOC
## Test Statistic  : Normal
## Type of penalty       : MBIC with value, 14.90944
## Minimum Segment Length : 1
## Maximum no. of cpts   : 1
## Changepoint Locations : 77

plot(cpt.mean(AirPassengers))
```



#This change point is at 77, meaning the 77th month after the start, or 6 years and 5 months after January 1949, or May of 1955. This was found by looking for the an instance of a change in mean that results in the MBIC > 14.90944, given the type of test statistic we used (in this case, the default of "Normal"). This means our test found that between 1949 and 1960, May 1955 was the first month with a month-to-month change in the mean that was significant enough to reach that MBIC penalty. Our red lines in the plot confirm that the change point is at about May 1955.

#This is very similar to the change point found in the variance in Exercise 5, in fact, only one element (or in this case, one month) off. These change points suggest that international air travel had a very significant change around this period in time in terms of popularity.

##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.

#The golden age of air travel is usually a reference to air travel in the 1950's and 1960's. Airplanes became more commercialized, faster, safer, and

more accommodating than ever before around this time. This can be pinpointed by our data to be estimated as having the biggest effect on people wanting to fly internationally in early-to-mid 1955.

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

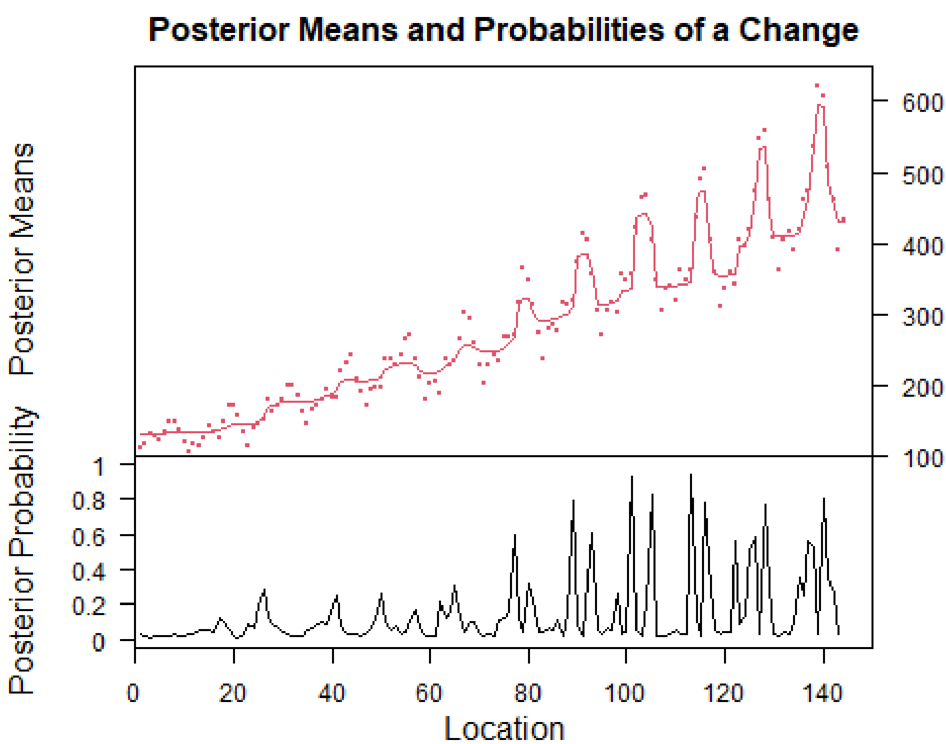
```
library(bcp)

## Loading required package: grid

set.seed(1)
Exercise8 <- bcp(AirPassengers)
Exercise8

##
## Bayesian Change Point (bcp) summary:
##
##
## Probability of a change in mean and posterior means:
##
##     Probability    X1
## 1         0.026 130.8
## 2         0.020 131.3
## 3         0.012 131.7
## 4         0.018 131.9
## 5         0.024 132.0
## 6         0.018 132.3
## 7         0.016 132.5
## 8         0.028 132.6
## 9         0.018 132.6
## 10        0.014 132.5
## 11        0.028 132.4
## 12        0.026 132.5
## 13        0.048 132.9
## 14        0.052 133.9
## 15        0.050 135.1
## 16        0.046 136.1
## 17        0.120 137.0
## 18        0.098 140.6
## 19        0.050 143.2
## 20        0.018 144.0
## 21        0.010 143.8
## 22        0.034 143.9
## 23        0.084 144.3
```

```
## 24         0.068 146.6
## 25         0.188 148.8
## 26         0.284 155.8
## 27         0.118 167.1
## 28         0.092 171.0
## 29         0.070 173.6
## 30         0.044 175.7
## 31         0.028 176.7
## 32         0.018 176.8
## 33         0.020 176.9
## 34         0.020 176.7
## 35         0.052 176.7
## 36         0.064 177.6
## 37         0.092 179.1
## 38         0.098 181.8
## 39         0.086 185.1
## 40         0.192 187.9
## 41         0.244 194.7
## 42         0.080 204.2
## 43         0.030 206.6
## 44         0.028 207.0
## 45         0.028 206.5
## 46         0.020 206.0
## 47         0.046 206.1
## 48         0.064 206.7
## 49         0.106 208.4
## 50         0.266 211.8
## 51         0.094 223.0
## 52         0.056 226.0
## 53         0.074 227.3
## 54         0.034 229.9
## 55         0.038 230.9
## 56         0.110 230.9
## 57         0.170 226.9
## 58         0.066 219.7
## 59         0.022 217.6
## 60         0.022 217.4
## 61         0.024 217.4
## 62         0.218 217.9
## 63         0.126 226.8
## 64         0.160 231.1
## 65         0.302 237.5
## 66         0.118 252.0
## 67         0.036 257.4
## 68         0.096 256.6
## 69         0.098 252.3
## 70         0.032 248.2
## 71         0.022 247.4
## 72         0.026 247.5
## 73         0.024 248.1
```

```
## 74         0.110 248.6
## 75         0.126 254.2
## 76         0.160 260.1
## 77         0.598 268.5
## 78         0.132 313.9
## 79         0.046 322.5
## 80         0.316 320.8
## 81         0.230 303.0
## 82         0.036 290.6
## 83         0.038 289.6
## 84         0.064 290.4
## 85         0.048 292.1
## 86         0.112 293.5
## 87         0.024 298.4
## 88         0.148 299.4
## 89         0.796 309.3
## 90         0.072 381.0
## 91         0.020 385.4
## 92         0.338 385.4
## 93         0.612 360.5
## 94         0.048 314.5
## 95         0.034 312.8
## 96         0.062 314.1
## 97         0.058 316.1
## 98         0.264 318.3
## 99         0.032 332.4
## 100        0.044 333.6
## 101        0.932 336.7
## 102        0.058 436.5
## 103        0.020 440.2
## 104        0.198 440.3
## 105        0.832 421.9
## 106        0.024 340.3
## 107        0.018 339.3
## 108        0.014 339.6
## 109        0.028 339.9
## 110        0.054 340.3
## 111        0.034 342.0
## 112        0.026 342.8
## 113        0.944 343.9
## 114        0.124 462.4
## 115        0.016 473.6
## 116        0.778 474.2
## 117        0.300 384.4
## 118        0.054 355.3
## 119        0.032 353.0
## 120        0.044 353.4
## 121        0.044 354.8
## 122        0.562 356.7
## 123        0.092 394.7
```

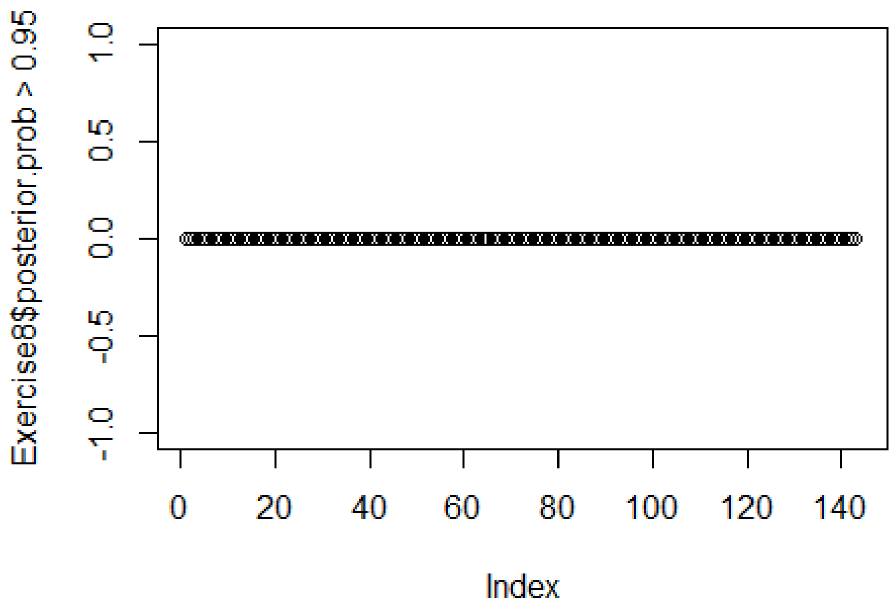
```
## 124         0.130 400.2
## 125         0.510 409.0
## 126         0.580 469.1
## 127         0.034 533.8
## 128         0.774 534.8
## 129         0.308 440.6
## 130         0.028 410.9
## 131         0.022 410.2
## 132         0.038 410.6
## 133         0.030 411.3
## 134         0.104 411.7
## 135         0.358 416.6
## 136         0.250 442.7
## 137         0.562 462.7
## 138         0.522 537.2
## 139         0.034 594.2
## 140         0.810 593.6
## 141         0.354 487.9
## 142         0.264 449.3
## 143         0.026 429.9
## 144            NA 430.0
```

```
plot(Exercise8)
```

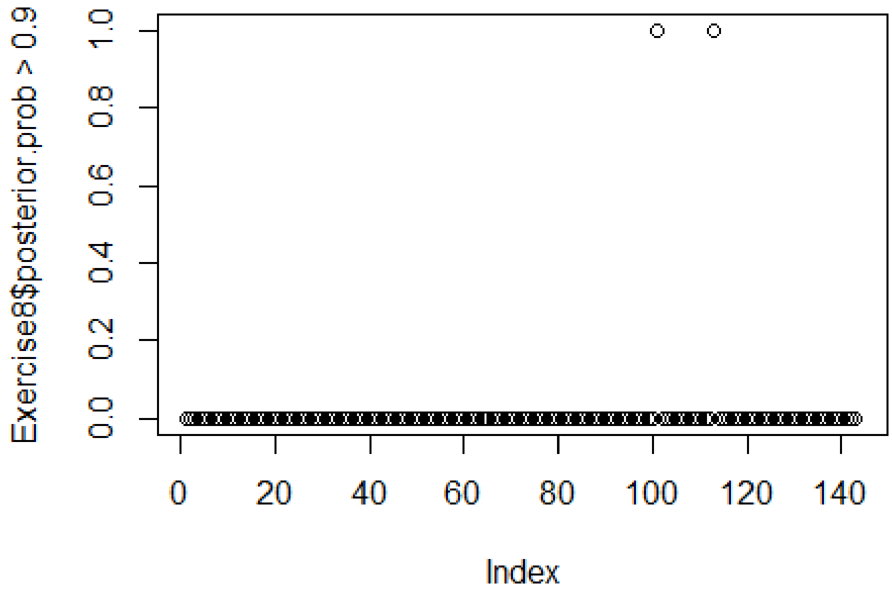



```
#In our plot, the upper pane shows the original time series and the lower
pane shows the probabilities of a mean change at each point in time. We see
large probabilities a little over 100 and a little below 120, but rather than
```

```
eyeballing it, let's look at those probabilities greater than 0.95.
plot(Exercise8$posterior.prob > 0.95)
```




```
#This seems to show us that none of the probabilities are above 0.95. Let's
check for those above 0.90 instead.
plot(Exercise8$posterior.prob > 0.90)
```




```
#Even these points are those we identified before, somewhere between 100 and
120. We could calculate these, but what is more important instead is that we
see these are nowhere near the change points that we found in exercises 5 and
6.
```