EXPERIMENT 7.1

Taniya Ahmed

21BDS0059

1. **Load the in-build *AirPassengers* dataset from R.**

   CODE:

   df = data(AirPassengers)

   print("Taniya Ahmed 21BDS0059")

   OUTPUT:

```
> data_logged = log(df)
> print("Taniya Ahmed 21BDS0059")
[1] "Taniya Ahmed 21BDS0059"
```

2. **Check for the Structure and the Data type of *AirPassengers*.**

   CODE:

   df = AirPassengers

   str(df)

   class(df)

   print("Taniya Ahmed 21BDS0059")

   OUTPUT:

```
> df = AirPassengers
> str(df)
 Time-Series [1:144] from 1949 to 1961: 112 118 132 129 121 135 148 148 136 119 ...
> class(df)
[1] "ts"
> print("Taniya Ahmed 21BDS0059")
[1] "Taniya Ahmed 21BDS0059"
```

3. **Check for missing values in the dataset.**

CODE:

```
sum(is.na(df))

df

print("Taniya Ahmed 21BDS0059")
```

OUTPUT:

```
> sum(is.na(df))
[1] 0
> df
     Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
1949 112 118 132 129 121 135 148 148 136 119 104 118
1950 115 126 141 135 125 149 170 170 158 133 114 140
1951 145 150 178 163 172 178 199 199 184 162 146 166
1952 171 180 193 181 183 218 230 242 209 191 172 194
1953 196 196 236 235 229 243 264 272 237 211 180 201
1954 204 188 235 227 234 264 302 293 259 229 203 229
1955 242 233 267 269 270 315 364 347 312 274 237 278
1956 284 277 317 313 318 374 413 405 355 306 271 306
1957 315 301 356 348 355 422 465 467 404 347 305 336
1958 340 318 362 348 363 435 491 505 404 359 310 337
1959 360 342 406 396 420 472 548 559 463 407 362 405
1960 417 391 419 461 472 535 622 606 508 461 390 432
> print("Taniya Ahmed 21BDS0059")
[1] "Taniya Ahmed 21BDS0059"
```

4. **Check for the Starting date and Ending date.**

CODE:

```
start(df)

end(df)

print("Taniya Ahmed 21BDS0059")
```

OUTPUT:

```
> start(df)
[1] 1949    1
> end(df)
[1] 1960   12
> print("Taniya Ahmed 21BDS0059")
[1] "Taniya Ahmed 21BDS0059"
```

5. **Check for the frequency of the dataset.**

CODE:

```
frequency(df)

print("Taniya Ahmed 21BDS0059")
```

OUTPUT:

```
> frequency(df)
[1] 12
> print("Taniya Ahmed 21BDS0059")
[1] "Taniya Ahmed 21BDS0059"
```

6. **Check for the summary of the dataset.**

CODE:

```
summary(df)

print("Taniya Ahmed 21BDS0059")
```

OUTPUT:

```
> summary(df)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  104.0   180.0   265.5   280.3   360.5   622.0
> print("Taniya Ahmed 21BDS0059")
[1] "Taniya Ahmed 21BDS0059"
```

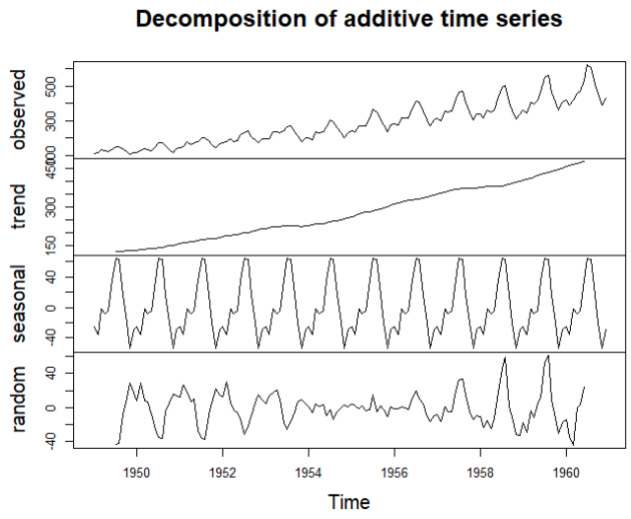
7. **Plot the decomposition of the dataset – Break data into trend, seasonal, and random.**

CODE:

```
data_decomposed = decompose(AirPassengers)

plot(data_decomposed)

print("Taniya Ahmed 21BDS0059")
```

OUTPUT:

```
> data_decomposed = decompose(AirPassengers)
> plot(data_decomposed)
> print("Taniya Ahmed 21BDS0059")
[1] "Taniya Ahmed 21BDS0059"
```

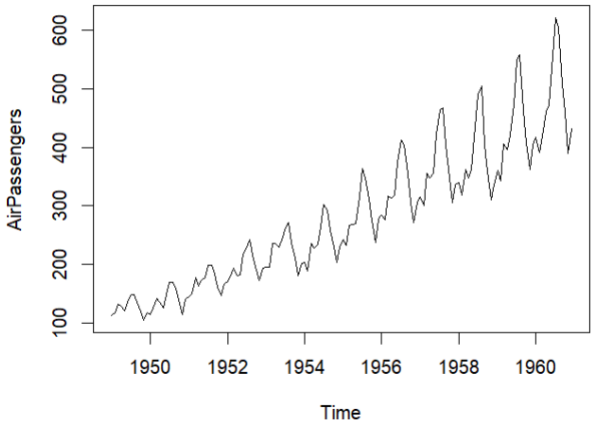


8. **Plot the dataset**

CODE:

```
plot(AirPassengers)
print("Taniya Ahmed 21BDS0059")
```

OUTPUT:

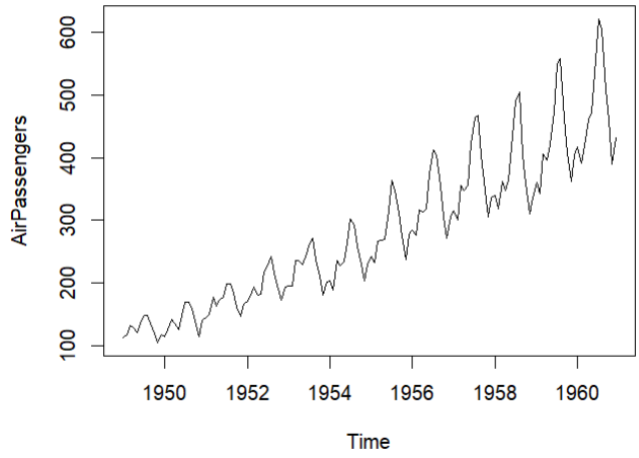


9. **Plot the time series of the dataset.**

CODE:

```
plot.ts(AirPassengers)
print("Taniya Ahmed 21BDS0059")
```

OUTPUT:

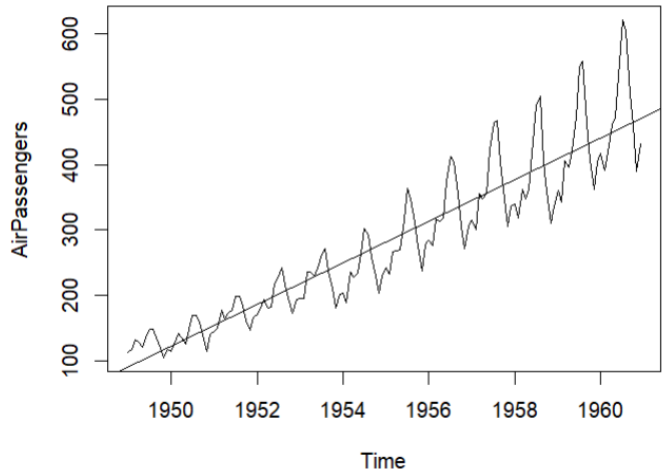


10. **Draw the regressor line for the.**

CODE:

```
abline(lm(df ~ time(df)))
print("Taniya Ahmed 21BDS0059")
```

OUTPUT:



11. **Print the cycle across the years for the dataset.**

CODE:

```
cycle(df)
print("Taniya Ahmed 21BDS0059")
```

OUTPUT:

```
> cycle(df)
     Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
1949   1   2   3   4   5   6   7   8   9  10  11  12
1950   1   2   3   4   5   6   7   8   9  10  11  12
1951   1   2   3   4   5   6   7   8   9  10  11  12
1952   1   2   3   4   5   6   7   8   9  10  11  12
1953   1   2   3   4   5   6   7   8   9  10  11  12
1954   1   2   3   4   5   6   7   8   9  10  11  12
1955   1   2   3   4   5   6   7   8   9  10  11  12
1956   1   2   3   4   5   6   7   8   9  10  11  12
1957   1   2   3   4   5   6   7   8   9  10  11  12
1958   1   2   3   4   5   6   7   8   9  10  11  12
1959   1   2   3   4   5   6   7   8   9  10  11  12
1960   1   2   3   4   5   6   7   8   9  10  11  12
> print("Taniya Ahmed 21BDS0059")
[1] "Taniya Ahmed 21BDS0059"
```

12. **Make the dataset to stationary i.e. constant mean and variance and plot it.**

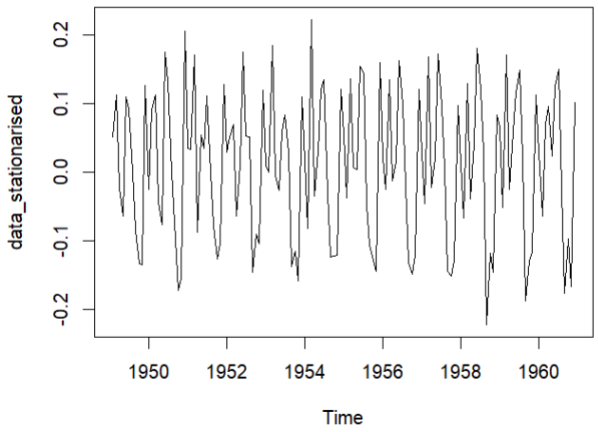
CODE:

```
data_logged = log(df)

data_stationarised = diff(data_logged)

print("Taniya Ahmed 21BDS0059")

plot(data_stationarised)

print("Taniya Ahmed 21BDS0059")
```

OUTPUT:

```
> data_logged = log(df)
> data_stationarised = diff(data_logged)
> print(data_stationarised)
              Jan          Feb          Mar
1949                0.052185753  0.112117298
1950 -0.025752496  0.091349779  0.112477983
1951  0.035091320  0.033901552  0.171148256
1952  0.029675768  0.051293294  0.069733338
1953  0.010256500  0.000000000  0.185717146
1954  0.014815086 -0.081678031  0.223143551
1955  0.055215723 -0.037899273  0.136210205
1956  0.021353124 -0.024956732  0.134884268
1957  0.028987537 -0.045462374  0.167820466
1958  0.011834458 -0.066894235  0.129592829
1959  0.066021101 -0.051293294  0.171542423
1960  0.029199155 -0.064378662  0.069163360
              Apr          May          Jun
1949 -0.022989518 -0.064021859  0.109484233
1950 -0.043485112 -0.076961041  0.175632569
1951 -0.088033349  0.053744276  0.034289073
1952 -0.064193158  0.010989122  0.175008910
1953 -0.004246291 -0.025863511  0.059339440
1954 -0.034635497  0.030371098  0.120627988
1955  0.007462721  0.003710579  0.154150680
1956 -0.012698583  0.015848192  0.162204415
1957 -0.022728251  0.019915310  0.172887525
1958 -0.039441732  0.042200354  0.180943197
1959 -0.024938948  0.058840500  0.116724274
1960  0.095527123  0.023580943  0.125287761
              Jul          Aug          Sep
1949  0.091937495  0.000000000 -0.084557388
1950  0.131852131  0.000000000 -0.073203404
1951  0.111521274  0.000000000 -0.078369067
1952  0.053584246  0.050858417 -0.146603474
1953  0.082887660  0.029852963 -0.137741925
1954  0.134477914 -0.030254408 -0.123344547
1955  0.144581229 -0.047829088 -0.106321592
1956  0.099191796 -0.019560526 -0.131769278
1957  0.097032092  0.004291852 -0.144914380
1958  0.121098097  0.028114301 -0.223143551
```

```
1959  0.149296301  0.019874186 -0.188422419
1960  0.150673346 -0.026060107 -0.176398538
              Oct           Nov           Dec
1949 -0.133531393 -0.134732594  0.126293725
1950 -0.172245905 -0.154150680  0.205443974
1951 -0.127339422 -0.103989714  0.128381167
1952 -0.090060824 -0.104778951  0.120363682
1953 -0.116202008 -0.158901283  0.110348057
1954 -0.123106058 -0.120516025  0.120516025
1955 -0.129875081 -0.145067965  0.159560973
1956 -0.148532688 -0.121466281  0.121466281
1957 -0.152090098 -0.129013003  0.096799383
1958 -0.118092489 -0.146750091  0.083510633
1959 -0.128913869 -0.117168974  0.112242855
1960 -0.097083405 -0.167251304  0.102278849
> print("Taniya Ahmed 21BDS0059")
[1] "Taniya Ahmed 21BDS0059"
>
```



13. **Plot box plot across months for seasonal effect.**

CODE:

```
boxplot(df ~ cycle(df),
    main = "Boxplot across months",
    xlab = "Month", ylab = "No. of passengers")
```

OUTPUT: