21BDS0340

Abhinav Dinesh Srivatsa

Exploratory Data Analysis Lab

**Experiment 7.1**

**Code:**

library(lubridate)

setwd("~/College Work/Year 4 Semester 1 (Sem 7)/Exploratory Data Analysis Lab/Experiment 7-1")

**Output:**

> library(lubridate)

>

> setwd("~/College Work/Year 4 Semester 1 (Sem 7)/Exploratory Data Analysis Lab/Experiment 7-1")

**Code:**

# 1. loading the data

data = read.csv("airpassengers.csv")

data$Month = as.Date(paste0(data$Month, "-01"), format = "%Y-%m-%d")

**Output:**

> # 1. loading the data

> data = read.csv("airpassengers.csv")

> data$Month = as.Date(paste0(data$Month, "-01"), format = "%Y-%m-%d")

**Code:**

# 2. structure and data types of data

head(data)

typeof(data$Month)

typeof(data$X.Passengers)

**Output:**

> # 2. structure and data types of data

> head(data)

Month X.Passengers

1 1949-01-01 112

2 1949-02-01 118

3 1949-03-01 132

4 1949-04-01 129

5 1949-05-01 121

6 1949-06-01 135

> typeof(data$Month)

[1] "double"

> typeof(data$X.Passengers)

[1] "integer"

**Code:**

# 3. checking for missing values in the data

sum(is.na(data))

**Output:**

> # 3. checking for missing values in the data

> sum(is.na(data))

[1] 0

**Code:**

# 4. checking start and end date

min(data$Month)

max(data$Month)

**Output:**

> # 4. checking start and end date

> min(data$Month)

[1] "1949-01-01"

> max(data$Month)

[1] "1960-12-01"

**Code:**

# 5. checking frequency of the data

frequency(data)

**Output:**

> # 5. checking frequency of the data

> frequency(data)

[1] 1

**Code:**

# 6. checking summary of the data

summary(data)

**Output:**

> # 6. checking summary of the data

> summary(data)

Month X.Passengers

Min. :1949-01-01 Min. :104.0

1st Qu.:1951-12-24 1st Qu.:180.0

Median :1954-12-16 Median :265.5

Mean :1954-12-16 Mean :280.3

3rd Qu.:1957-12-08 3rd Qu.:360.5

Max. :1960-12-01 Max. :622.0

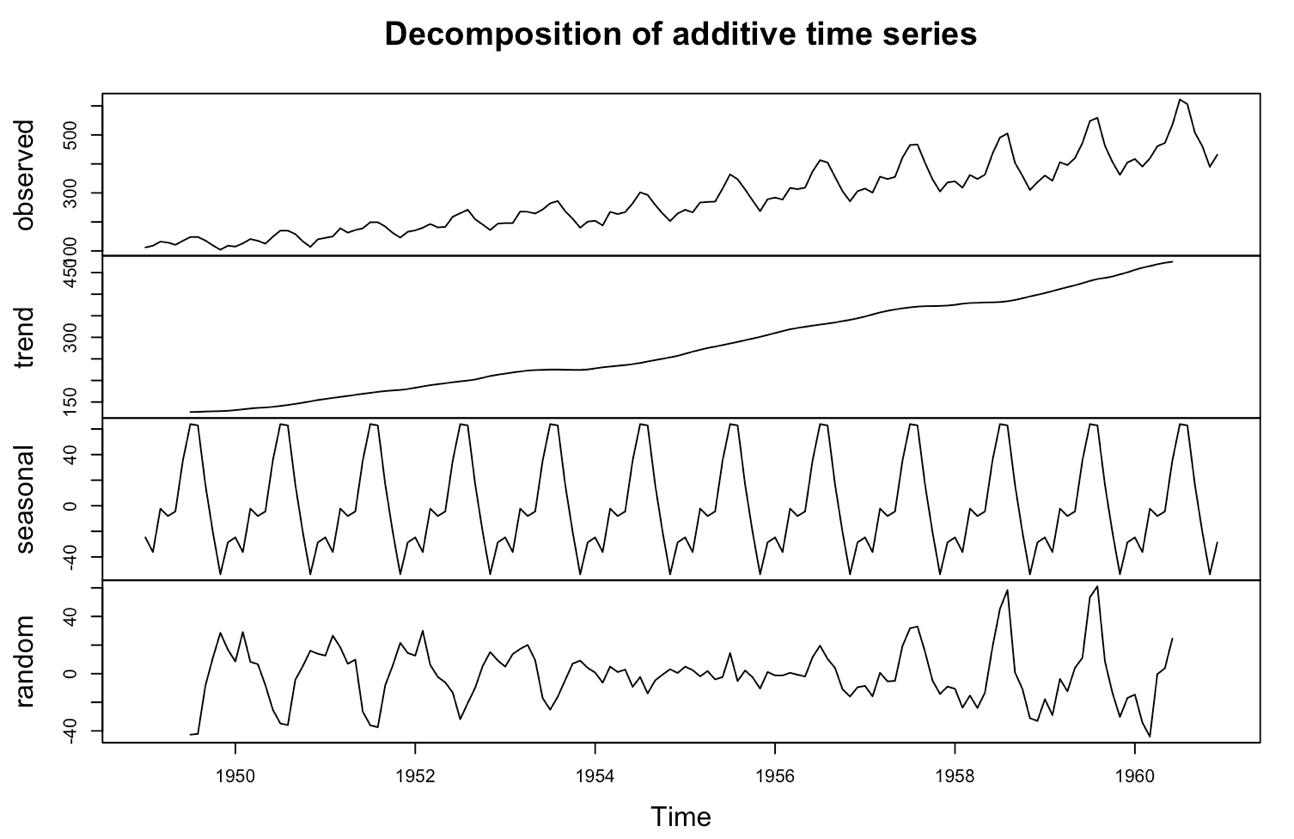
**Code:**

# 7. plotting with decompose

ts.data = ts(data$X.Passengers, start = c(1949, 1), frequency = 12)

plot(decompose(ts.data))

**Output:**

****

**Code:**

# 8. plotting the data

plot(data)

**Output:**

**A graph of growth of the year

Description automatically generated with medium confidence**

**Code:**

# 9. plotting the time series of the data

plot.ts(data)

**Output:**

**A graph with lines and numbers

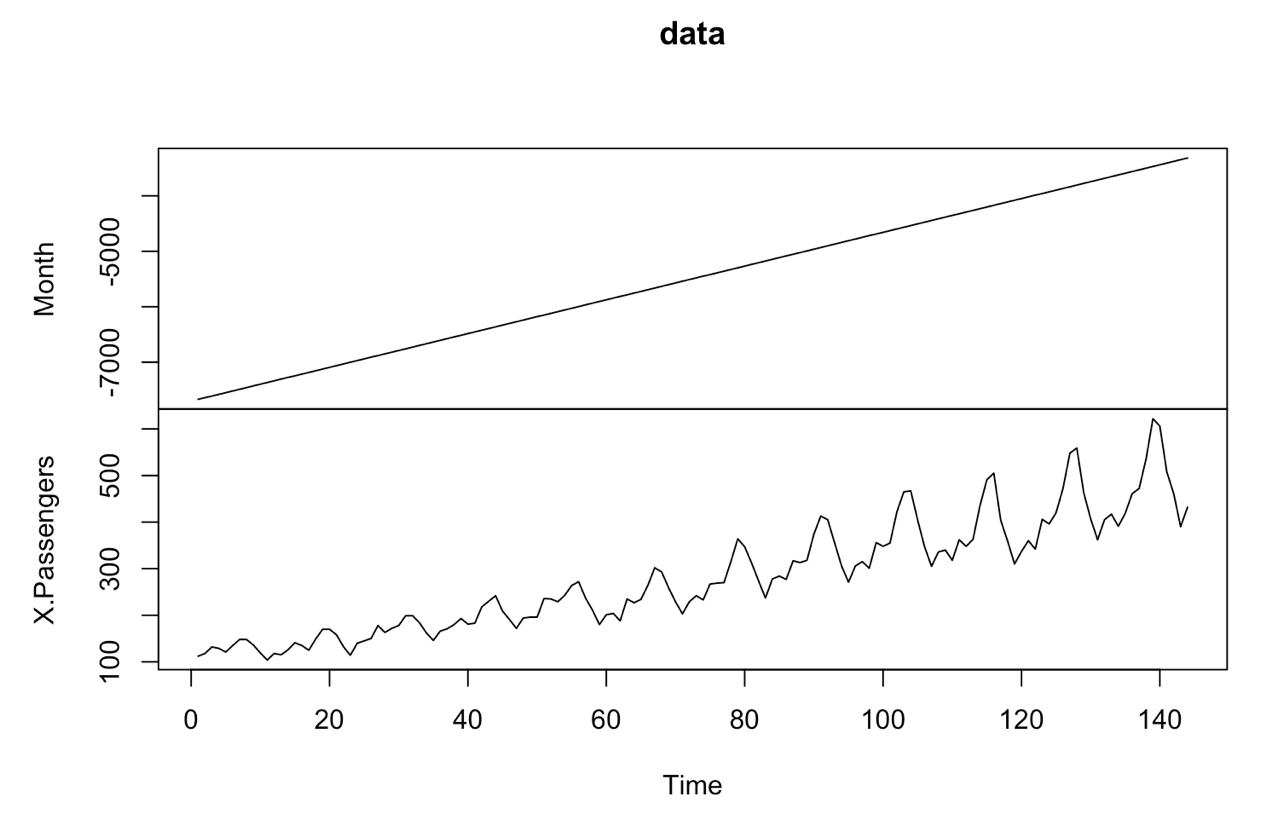
Description automatically generated**

**Code:**

# 10. regression line for the data

abline(lm(data$X.Passengers~data$Month))

**Output:**

****

**Code:**

# 11. cycles for the data

cycle(ts.data)

**Output:**

> # 11. cycles for the data

> cycle(ts.data)

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

**Code:**

# 12. making the data stationary and plotting it

plot(log(ts.data))

plot(diff(log(ts.data)))

**Output:**

**A graph showing time and time

Description automatically generated**

**Code:**

# 13. box plot

boxplot(ts.data~cycle(ts.data))

**Output:**

**A graph of a graph

Description automatically generated with medium confidence**