

Applied Statistics Using R

(MCA232)

# Practical 1

***BY***

**Himanshu Heda (24225013)**

**SUBMITTED TO**

**Dr. Ashish Sharma**



**Download a data set with at  
least 100 entries and a minimum 4 numeric variables.**

**Load the Dataset**

> data <- read.csv("HUBC\_stock\_data.csv")

> data

X open high low close adjclose volume ticker

1 03-01-2022 98.40 98.40 98.40 98.40 98.40 1000 HUBC

2 04-01-2022 98.40 98.40 98.40 98.40 98.40 0 HUBC

3 05-01-2022 98.80 99.00 98.70 98.90 98.90 90090 HUBC

4 06-01-2022 99.10 99.10 98.90 99.00 99.00 38390 HUBC

5 07-01-2022 99.00 99.30 98.80 99.20 99.20 54640 HUBC

6 10-01-2022 99.10 99.20 99.00 99.00 99.00 11510 HUBC

7 11-01-2022 99.00 99.20 99.00 99.00 99.00 4920 HUBC

8 12-01-2022 99.00 99.20 99.00 99.20 99.20 110 HUBC

9 13-01-2022 99.20 99.20 99.20 99.20 99.20 60 HUBC

10 14-01-2022 99.10 99.30 99.10 99.20 99.20 5700 HUBC

11 18-01-2022 99.20 99.20 99.20 99.20 99.20 0 HUBC

12 19-01-2022 99.20 99.50 99.20 99.50 99.50 48350 HUBC

13 20-01-2022 99.30 99.30 99.30 99.30 99.30 80 HUBC

14 21-01-2022 99.10 99.15 98.90 98.90 98.90 610 HUBC

15 24-01-2022 99.00 99.10 99.00 99.00 99.00 390 HUBC

16 25-01-2022 99.00 99.00 99.00 99.00 99.00 0 HUBC

17 26-01-2022 99.00 99.00 99.00 99.00 99.00 0 HUBC

18 27-01-2022 99.00 99.10 98.60 98.60 98.60 8800 HUBC

19 28-01-2022 99.00 99.20 99.00 99.20 99.20 680 HUBC

20 31-01-2022 99.20 99.20 99.20 99.20 99.20 0 HUBC

21 01-02-2022 99.20 99.20 99.20 99.20 99.20 0 HUBC

22 02-02-2022 99.00 99.00 99.00 99.00 99.00 3200 HUBC

23 03-02-2022 99.00 99.00 99.00 99.00 99.00 0 HUBC

24 04-02-2022 99.10 99.10 99.10 99.10 99.10 4350 HUBC

25 07-02-2022 99.10 99.10 99.10 99.10 99.10 0 HUBC

26 08-02-2022 99.10 99.10 99.10 99.10 99.10 0 HUBC

27 09-02-2022 99.30 99.40 99.30 99.35 99.35 12130 HUBC

28 10-02-2022 99.30 99.50 99.30 99.50 99.50 17560 HUBC

29 11-02-2022 99.50 99.50 99.50 99.50 99.50 0 HUBC

30 14-02-2022 99.50 99.50 99.50 99.50 99.50 0 HUBC

31 15-02-2022 99.50 99.50 99.50 99.50 99.50 0 HUBC

32 16-02-2022 99.40 99.40 99.40 99.40 99.40 24060 HUBC

33 17-02-2022 99.40 99.40 99.30 99.40 99.40 30080 HUBC

34 18-02-2022 99.40 99.40 99.40 99.40 99.40 0 HUBC

35 22-02-2022 99.50 99.50 99.50 99.50 99.50 1050 HUBC

36 23-02-2022 99.50 99.50 99.50 99.50 99.50 0 HUBC

37 24-02-2022 99.50 99.50 99.50 99.50 99.50 0 HUBC

38 25-02-2022 99.50 99.50 99.50 99.50 99.50 0 HUBC

39 28-02-2022 99.60 99.60 99.60 99.60 99.60 90 HUBC

40 01-03-2022 99.60 99.80 99.60 99.80 99.80 23110 HUBC

41 02-03-2022 99.80 99.80 99.80 99.80 99.80 20 HUBC

42 03-03-2022 99.80 99.80 99.80 99.80 99.80 50 HUBC

43 04-03-2022 99.80 99.80 99.80 99.80 99.80 0 HUBC

44 07-03-2022 99.80 99.80 99.80 99.80 99.80 0 HUBC

45 08-03-2022 99.80 99.80 99.80 99.80 99.80 1760 HUBC

46 09-03-2022 99.80 99.80 99.80 99.80 99.80 0 HUBC

47 10-03-2022 99.80 99.80 99.80 99.80 99.80 0 HUBC

48 11-03-2022 99.50 99.80 99.50 99.80 99.80 7670 HUBC

49 14-03-2022 99.90 100.00 99.80 100.00 100.00 9260 HUBC

50 15-03-2022 100.00 100.00 100.00 100.00 100.00 450 HUBC

51 16-03-2022 100.00 100.00 100.00 100.00 100.00 0 HUBC

52 17-03-2022 100.00 100.00 100.00 100.00 100.00 80 HUBC

53 18-03-2022 100.00 100.10 100.00 100.10 100.10 50 HUBC

54 21-03-2022 100.10 100.10 100.10 100.10 100.10 0 HUBC

55 22-03-2022 100.10 100.10 100.10 100.10 100.10 0 HUBC

56 23-03-2022 100.70 100.70 100.00 100.30 100.30 117280 HUBC

57 24-03-2022 100.50 100.50 99.70 100.00 100.00 64730 HUBC

58 25-03-2022 99.89 100.10 99.70 99.80 99.80 2910 HUBC

59 28-03-2022 99.80 100.00 99.70 99.70 99.70 10680 HUBC

60 29-03-2022 99.86 100.40 99.70 99.90 99.90 29020 HUBC

61 30-03-2022 99.70 99.80 99.70 99.70 99.70 660 HUBC

62 31-03-2022 99.80 100.02 99.80 99.80 99.80 470 HUBC

63 01-04-2022 99.80 100.25 99.70 100.25 100.25 3870 HUBC

64 04-04-2022 100.02 100.10 99.95 100.10 100.10 600 HUBC

65 05-04-2022 100.10 100.15 100.05 100.10 100.10 1470 HUBC

66 06-04-2022 100.20 100.20 99.90 99.90 99.90 8130 HUBC

67 07-04-2022 99.91 100.22 99.91 100.18 100.18 60 HUBC

68 08-04-2022 99.90 99.90 99.90 99.90 99.90 180 HUBC

69 11-04-2022 99.90 99.90 99.90 99.90 99.90 0 HUBC

70 12-04-2022 99.90 99.90 99.90 99.90 99.90 130 HUBC

71 13-04-2022 100.00 100.00 100.00 100.00 100.00 130 HUBC

72 14-04-2022 100.00 100.15 100.00 100.15 100.15 20 HUBC

73 18-04-2022 100.15 100.15 100.15 100.15 100.15 0 HUBC

74 19-04-2022 100.30 100.30 100.10 100.30 100.30 2280 HUBC

75 20-04-2022 100.49 100.60 100.10 100.10 100.10 1670 HUBC

76 21-04-2022 100.50 100.50 100.50 100.50 100.50 180 HUBC

77 22-04-2022 100.50 100.50 100.50 100.50 100.50 0 HUBC

78 25-04-2022 100.20 100.20 100.10 100.20 100.20 7410 HUBC

79 26-04-2022 100.20 100.20 100.20 100.20 100.20 20 HUBC

80 27-04-2022 100.20 100.20 100.20 100.20 100.20 0 HUBC

81 28-04-2022 100.30 100.30 100.20 100.20 100.20 650 HUBC

82 29-04-2022 100.20 100.20 100.20 100.20 100.20 0 HUBC

83 02-05-2022 100.20 100.30 100.20 100.20 100.20 340 HUBC

84 03-05-2022 100.20 100.20 100.20 100.20 100.20 1350 HUBC

85 04-05-2022 100.62 100.62 100.62 100.62 100.62 20 HUBC

86 05-05-2022 100.20 100.30 100.20 100.30 100.30 1770 HUBC

87 06-05-2022 100.20 100.20 100.20 100.20 100.20 90 HUBC

88 09-05-2022 100.20 100.20 100.20 100.20 100.20 90 HUBC

89 10-05-2022 100.10 100.10 100.10 100.10 100.10 2980 HUBC

90 11-05-2022 100.35 100.35 100.10 100.10 100.10 200 HUBC

91 12-05-2022 100.00 100.30 100.00 100.30 100.30 1020 HUBC

92 13-05-2022 99.90 99.90 99.90 99.90 99.90 40 HUBC

93 16-05-2022 99.90 99.90 99.90 99.90 99.90 0 HUBC

94 17-05-2022 100.49 100.49 100.49 100.49 100.49 30 HUBC

95 18-05-2022 100.60 100.60 100.00 100.00 100.00 70 HUBC

96 19-05-2022 99.90 99.90 99.90 99.90 99.90 60 HUBC

97 20-05-2022 99.90 99.90 99.90 99.90 99.90 0 HUBC

98 23-05-2022 99.90 99.90 99.90 99.90 99.90 60 HUBC

99 24-05-2022 99.90 99.90 99.90 99.90 99.90 60 HUBC

100 25-05-2022 99.90 99.90 99.90 99.90 99.90 0 HUBC

101 26-05-2022 99.90 99.90 99.90 99.90 99.90 30 HUBC

102 27-05-2022 99.90 99.90 99.90 99.90 99.90 0 HUBC

103 31-05-2022 99.90 99.90 99.90 99.90 99.90 20 HUBC

104 01-06-2022 100.00 100.00 100.00 100.00 100.00 30 HUBC

105 02-06-2022 100.10 100.20 99.90 100.20 100.20 6880 HUBC

106 03-06-2022 100.00 100.20 100.00 100.20 100.20 11170 HUBC

107 06-06-2022 100.25 100.25 100.10 100.10 100.10 320 HUBC

108 07-06-2022 100.10 100.28 100.10 100.28 100.28 5140 HUBC

109 08-06-2022 100.30 100.30 100.20 100.20 100.20 130 HUBC

110 09-06-2022 100.20 100.20 100.20 100.20 100.20 0 HUBC

111 10-06-2022 100.20 100.20 100.20 100.20 100.20 40 HUBC

112 13-06-2022 100.40 100.40 100.15 100.20 100.20 5920 HUBC

113 14-06-2022 100.20 100.20 99.90 99.90 99.90 5500 HUBC

114 15-06-2022 99.90 99.90 99.90 99.90 99.90 0 HUBC

115 16-06-2022 100.00 100.30 99.90 99.90 99.90 21110 HUBC

116 17-06-2022 100.05 100.05 100.05 100.05 100.05 20 HUBC

117 21-06-2022 100.00 100.10 100.00 100.10 100.10 130 HUBC

118 22-06-2022 100.10 100.10 100.10 100.10 100.10 0 HUBC

119 23-06-2022 100.05 100.20 99.95 100.00 100.00 5130 HUBC

120 24-06-2022 100.10 100.15 100.10 100.15 100.15 140 HUBC

121 27-06-2022 100.15 100.15 100.15 100.15 100.15 0 HUBC

122 28-06-2022 100.15 100.15 100.15 100.15 100.15 0 HUBC

123 29-06-2022 100.15 100.15 100.15 100.15 100.15 0 HUBC

124 30-06-2022 100.60 100.60 100.60 100.60 100.60 20 HUBC

125 01-07-2022 100.60 100.60 100.60 100.60 100.60 0 HUBC

[ reached 'max' / getOption("max.print") -- omitted 494 rows ]

> str(data)

'data.frame': 619 obs. of 8 variables:

$ X : chr "03-01-2022" "04-01-2022" "05-01-2022" "06-01-2022" ...

$ open : num 98.4 98.4 98.8 99.1 99 ...

$ high : num 98.4 98.4 99 99.1 99.3 ...

$ low : num 98.4 98.4 98.7 98.9 98.8 ...

$ close : num 98.4 98.4 98.9 99 99.2 ...

$ adjclose: num 98.4 98.4 98.9 99 99.2 ...

$ volume : int 1000 0 90090 38390 54640 11510 4920 110 60 5700 ...

$ ticker : chr "HUBC" "HUBC" "HUBC" "HUBC" ...

> head(data)

X open high low close adjclose volume ticker

1 03-01-2022 98.4 98.4 98.4 98.4 98.4 1000 HUBC

2 04-01-2022 98.4 98.4 98.4 98.4 98.4 0 HUBC

3 05-01-2022 98.8 99.0 98.7 98.9 98.9 90090 HUBC

4 06-01-2022 99.1 99.1 98.9 99.0 99.0 38390 HUBC

5 07-01-2022 99.0 99.3 98.8 99.2 99.2 54640 HUBC

6 10-01-2022 99.1 99.2 99.0 99.0 99.0 11510 HUBC

**1.     Draw  
the histogram, pie chart, box plot for any three variables and interpret them.**

**Draw Histogram, Pie Chart, and Box Plot**

# Install ggplot2 if not already installed

install.packages("ggplot2")

# Install ggplot2 if not already installed

**Histogram**

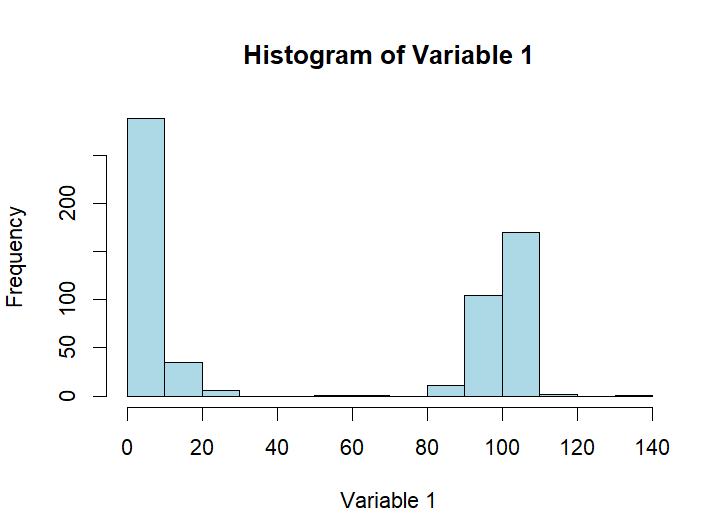
> hist(data$open,

+ main = "Histogram of Variable 1",

+ xlab = "Variable 1",

+ col = "lightblue",

+ border = "black")

****

**Pie Chart**

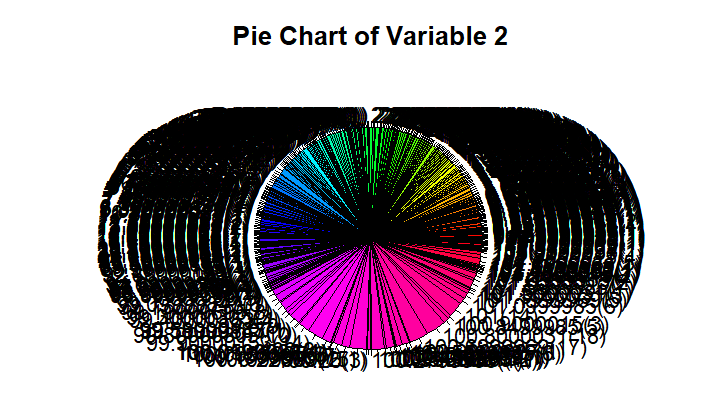
> table\_var2 <- table(data$high)

> pie(table\_var2,

+ main = "Pie Chart of Variable 2",

+ col = rainbow(length(table\_var2)),

+ labels = paste(names(table\_var2), "(", table\_var2, ")", sep = ""))

****

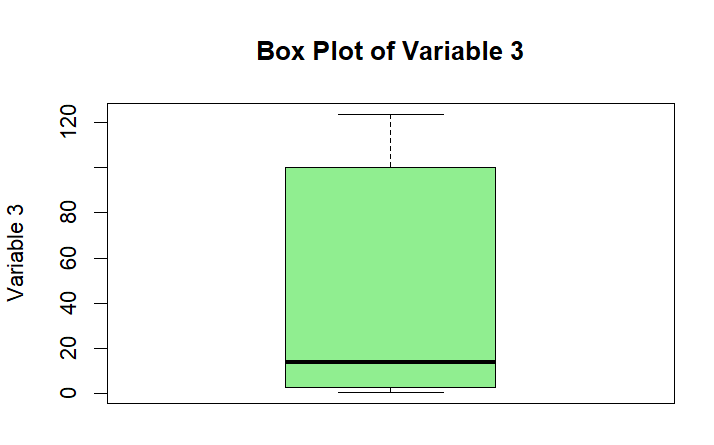
**BOX Plot**

> boxplot(data$adjclose,

+ main = "Box Plot of Variable 3",

+ ylab = "Variable 3",

+ col = "lightgreen")

****

**Line Graph between Two Variables**

> plot(data$high, data$low,

+ main = "Line Graph between Variable X and Variable Y",

+ type = "o",

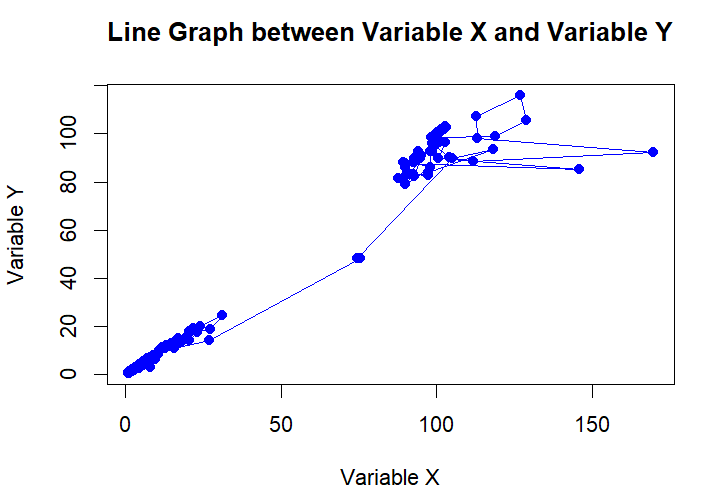
+ lty = 1,

+ pch = 19,

+ col = "blue",

+ xlab = "Variable X",

+ ylab = "Variable Y")

****

**3. Find  
the summary of at least 3 numeric variables.**

**Summary of Three Numeric Variables**

> summary(data$open)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.73 2.85 14.40 49.06 100.20 132.70

> summary(data$high)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.749 2.900 15.600 49.789 100.265 169.800

> summary(data$adjclose)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.719 2.665 13.900 48.949 100.200 123.600

**4. Find  
the following for any one of the variable  
  
(1)  2, 4 and 6 decile.  
  
(2)  5, 75, and 85  
percentile  
  
(3)  All the quartiles**

**Calculate Deciles, Percentiles, and Quartiles**

**Deciles**

> deciles <- quantile(data$open, probs = c(0.2, 0.4, 0.6))

> print(deciles)

20% 40% 60%

2.206 5.820 99.000

**Percentiles**

4

> percentiles <- quantile(data$open, probs = c(0.05, 0.75, 0.85))

> print(percentiles)

5% 75% 85%

0.9093 100.2000 100.8000

**Quartiles**

> quartiles <- quantile(data$open, probs = c(0.25, 0.5, 0.75))

> print(quartiles)

25% 50% 75%

2.85 14.40 100.20

**Analysis Report**

**1. Description of the Dataset**

The dataset used for this analysis contains stock market data for a company, with variables such as open, high, low, and adjclose. These variables represent the opening price, highest price, lowest price, and adjusted closing price of the stock on specific days. The data consists of numeric variables that help analyze stock trends over time.

**2. Aim/Objective**

The objective of this analysis is to:

1. Visualize the data using histograms, pie charts, box plots, and line graphs.
2. Summarize key numeric variables to understand the distribution and trends in the dataset.
3. Compute statistical measures such as deciles, percentiles, and quartiles for one variable.
4. Interpret the findings to derive insights into stock market behavior.

**3. Results and Visualizations**

**3.1 Histogram**

* **Code Used:**

hist(data$open,

main = "Histogram of Variable 1",

xlab = "Variable 1",

col = "lightblue",

border = "black")

* **Result:** The histogram shows the frequency distribution of the stock's opening prices (open). The majority of the values are clustered in the lower price range, indicating the stock's opening prices are typically low.

**3.2 Pie Chart**

* **Code Used:**

table\_var2 <- table(data$high)

pie(table\_var2,

main = "Pie Chart of Variable 2",

col = rainbow(length(table\_var2)),

labels = paste(names(table\_var2), "(", table\_var2, ")", sep = ""))

* **Result:** The pie chart represents the distribution of the stock's highest prices (high). Each segment corresponds to the frequency of specific high values in the dataset. The larger segments signify prices that occurred more frequently.

**3.3 Box Plot**

* **Code Used:**

boxplot(data$adjclose,

main = "Box Plot of Variable 3",

ylab = "Variable 3",

col = "lightgreen")

* **Result:** The box plot highlights the distribution of the stock's adjusted closing prices (adjclose). It identifies potential outliers, the median value, and the interquartile range. Most adjusted closing prices are concentrated in the lower range.

**3.4 Line Graph**

* **Code Used:**

plot(data$high, data$low,

main = "Line Graph between Variable X and Variable Y",

type = "o",

lty = 1,

pch = 19,

col = "blue",

xlab = "Variable X",

ylab = "Variable Y")

* **Result:** The line graph illustrates the relationship between the stock's highest (high) and lowest (low) prices. The trend shows how fluctuations in the stock's highest prices correlate with its lowest prices over time.

**4. Summary of Variables**

* **Summary of open:**

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.73 2.85 14.40 49.06 100.20 132.70

* **Summary of high:**

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.749 2.900 15.600 49.789 100.265 169.800

* **Summary of adjclose:**

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.719 2.665 13.900 48.949 100.200 123.600

**5. Statistical Measures**

* **Deciles (for open):**

20% 40% 60%

2.206 5.820 99.000

* **Percentiles (for open):**

5% 75% 85%

0.9093 100.2000 100.8000

* **Quartiles (for open):**

25% 50% 75%

2.85 14.40 100.20

**6. Interpretation**

1. **Histogram Interpretation:** The histogram shows that the opening prices of the stock are skewed towards the lower end, with only a few occurrences of higher prices.
2. **Pie Chart Interpretation:** The pie chart reveals which stock's highest prices occurred most frequently, indicating the stock's popular trading ranges.
3. **Box Plot Interpretation:** The box plot identifies outliers and suggests that the adjusted closing prices are mostly concentrated below the median, reflecting a potentially undervalued stock.
4. **Line Graph Interpretation:** The line graph demonstrates a positive correlation between the stock's highest and lowest prices. When the highest prices increase, the lowest prices tend to increase as well.
5. **Statistical Measures Interpretation:** The deciles, percentiles, and quartiles provide insights into the distribution of the stock's opening prices. For example, 20% of the stock's opening prices are below 2.206, while 75% of the prices are below 100.200.

**7. Conclusion**

This analysis provides a comprehensive understanding of the stock's price distribution and trends. The visualizations and statistical measures show that the stock has frequent low prices, but there are outliers at the higher end. These insights can help traders and investors in making informed decisions.