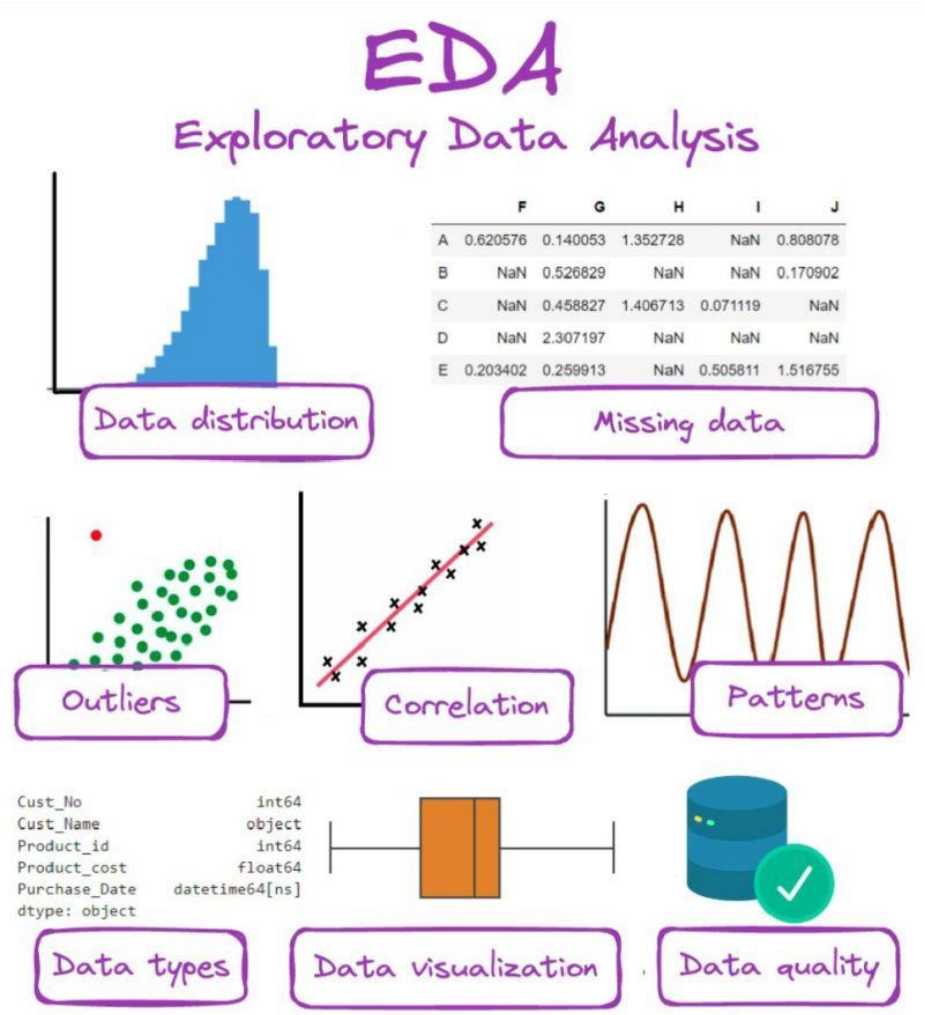
EXPERIMENT – 7



NAME : JVN GANESH

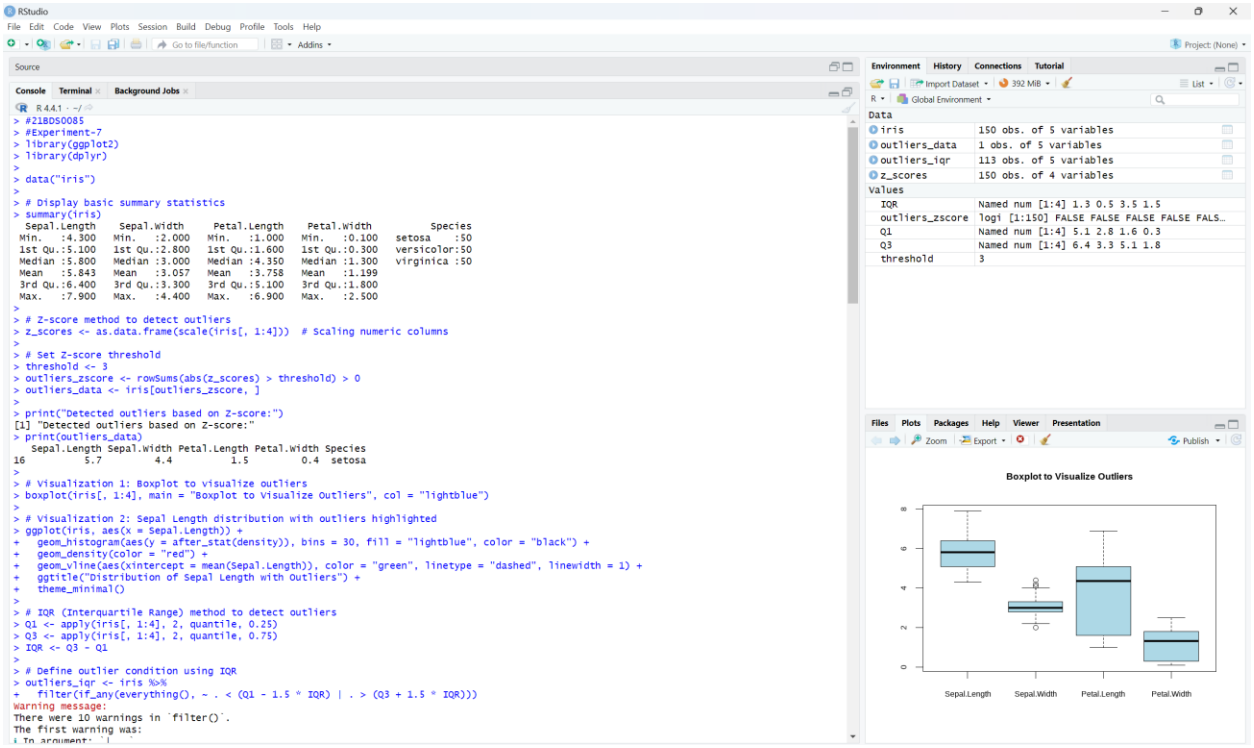
Roll No : 21BDS0085

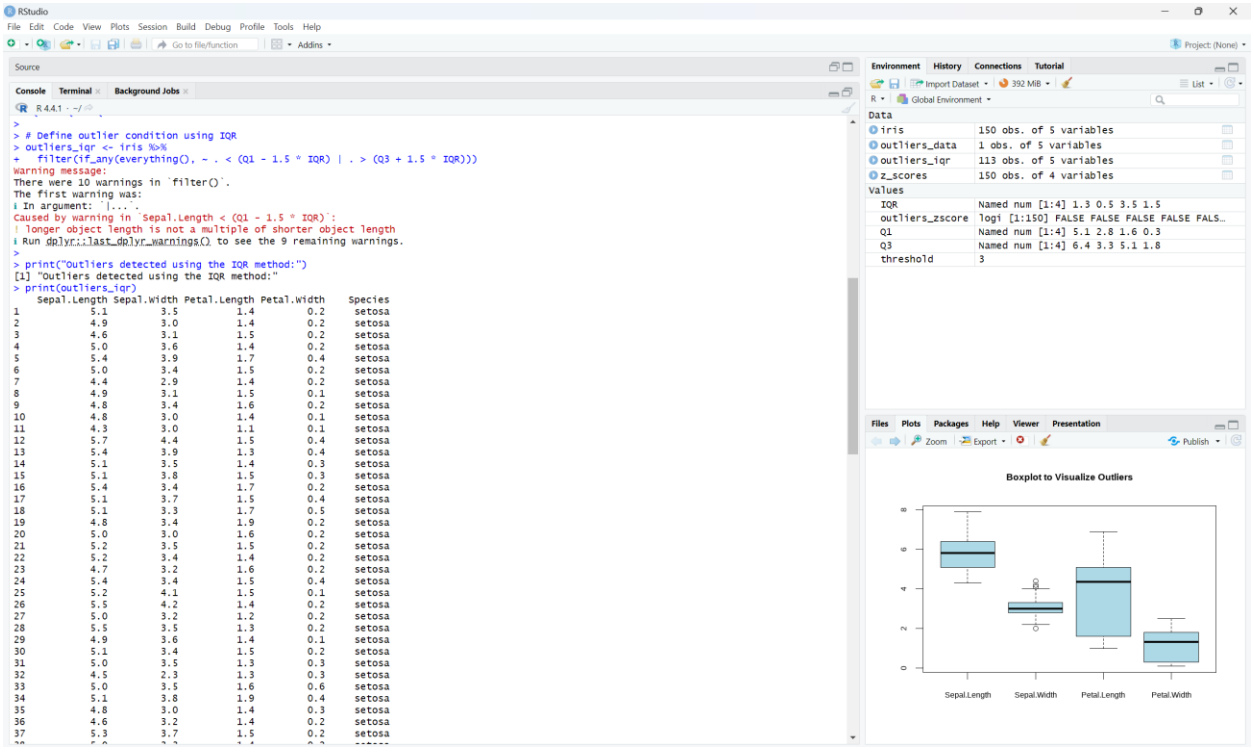
Code

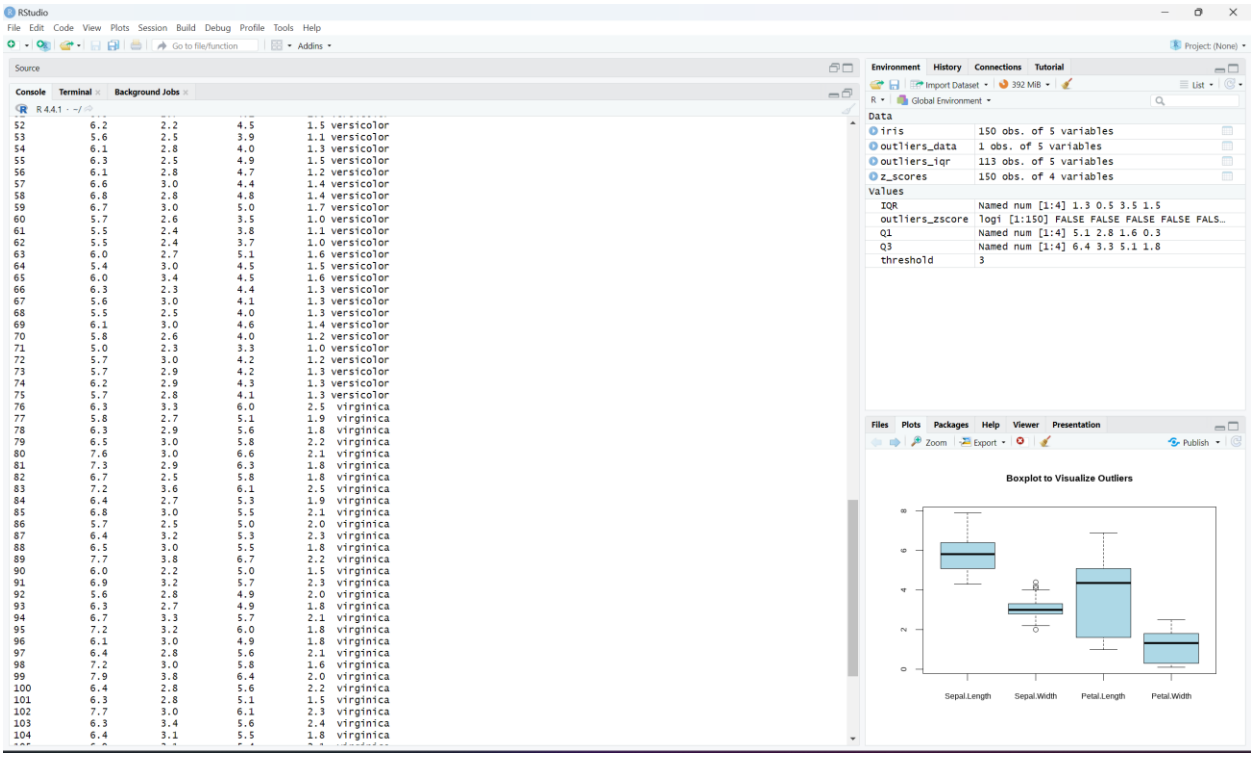
```
#21BDS0085
#Experiment-7
library(ggplot2)
library(dplyr)

data("iris")

# Display basic summary statistics
summary(iris)

# Z-score method to detect outliers
z_scores <- as.data.frame(scale(iris[, 1:4]))  # Scaling numeric columns

# Set Z-score threshold
threshold <- 3
outliers_zscore <- rowSums(abs(z_scores) > threshold) > 0
outliers_data <- iris[outliers_zscore, ]

print("Detected outliers based on Z-score:")
print(outliers_data)

# Visualization 1: Boxplot to visualize outliers
boxplot(iris[, 1:4], main = "Boxplot to Visualize Outliers", col = "lightblue")

# Visualization 2: Sepal Length distribution with outliers highlighted
```

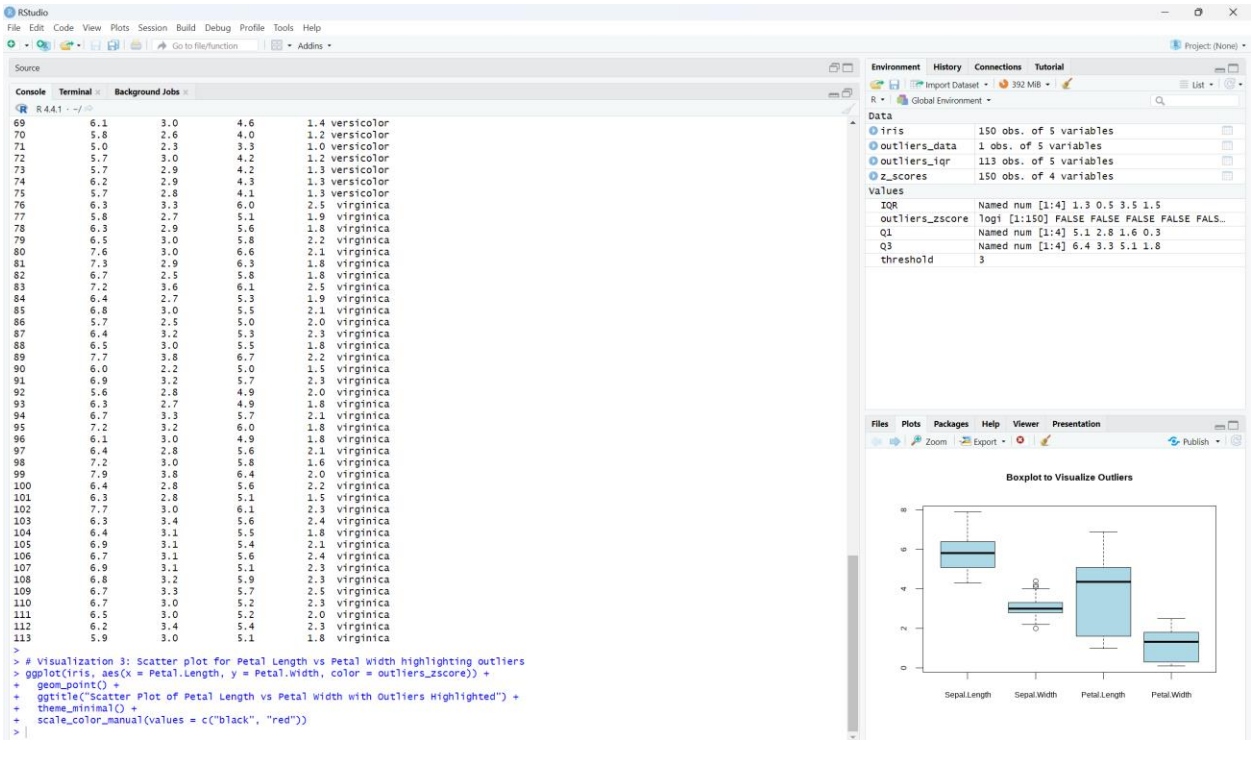
```
ggplot(iris, aes(x = Sepal.Length)) +
  geom_histogram(aes(y = after_stat(density)), bins = 30, fill = "lightblue", color = "black") +
  geom_density(color = "red") +
  geom_vline(aes(xintercept = mean(Sepal.Length)), color = "green", linetype = "dashed",
linewidth = 1) +
  ggtitle("Distribution of Sepal Length with Outliers") +
  theme_minimal()

# IQR (Interquartile Range) method to detect outliers
Q1 <- apply(iris[, 1:4], 2, quantile, 0.25)
Q3 <- apply(iris[, 1:4], 2, quantile, 0.75)
IQR <- Q3 - Q1

# Define outlier condition using IQR
outliers_iqr <- iris %>%
  filter(if_any(everything(), ~ . < (Q1 - 1.5 * IQR) | . > (Q3 + 1.5 * IQR)))

print("Outliers detected using the IQR method:")
print(outliers_iqr)

# Visualization 3: Scatter plot for Petal Length vs Petal Width highlighting outliers
ggplot(iris, aes(x = Petal.Length, y = Petal.Width, color = outliers_zscore)) +
  geom_point() +
  ggtitle("Scatter Plot of Petal Length vs Petal Width with Outliers Highlighted") +
  theme_minimal() +
  scale_color_manual(values = c("black", "red"))
```

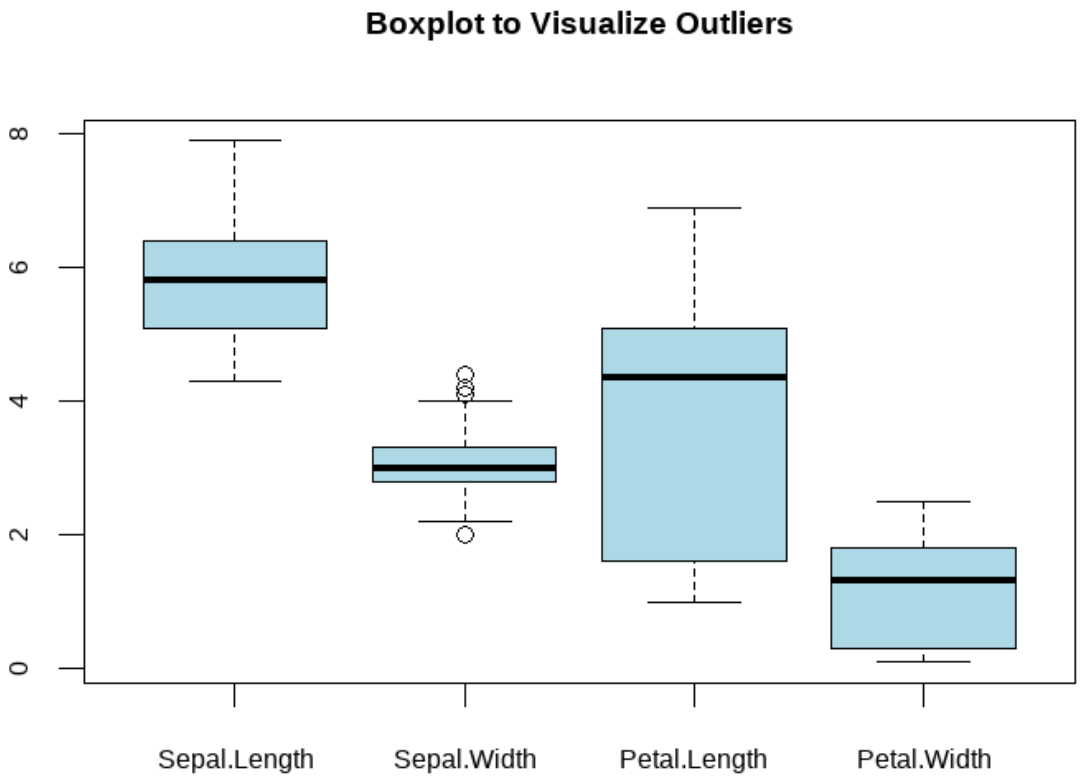
OUTPUT SCREENHOTS

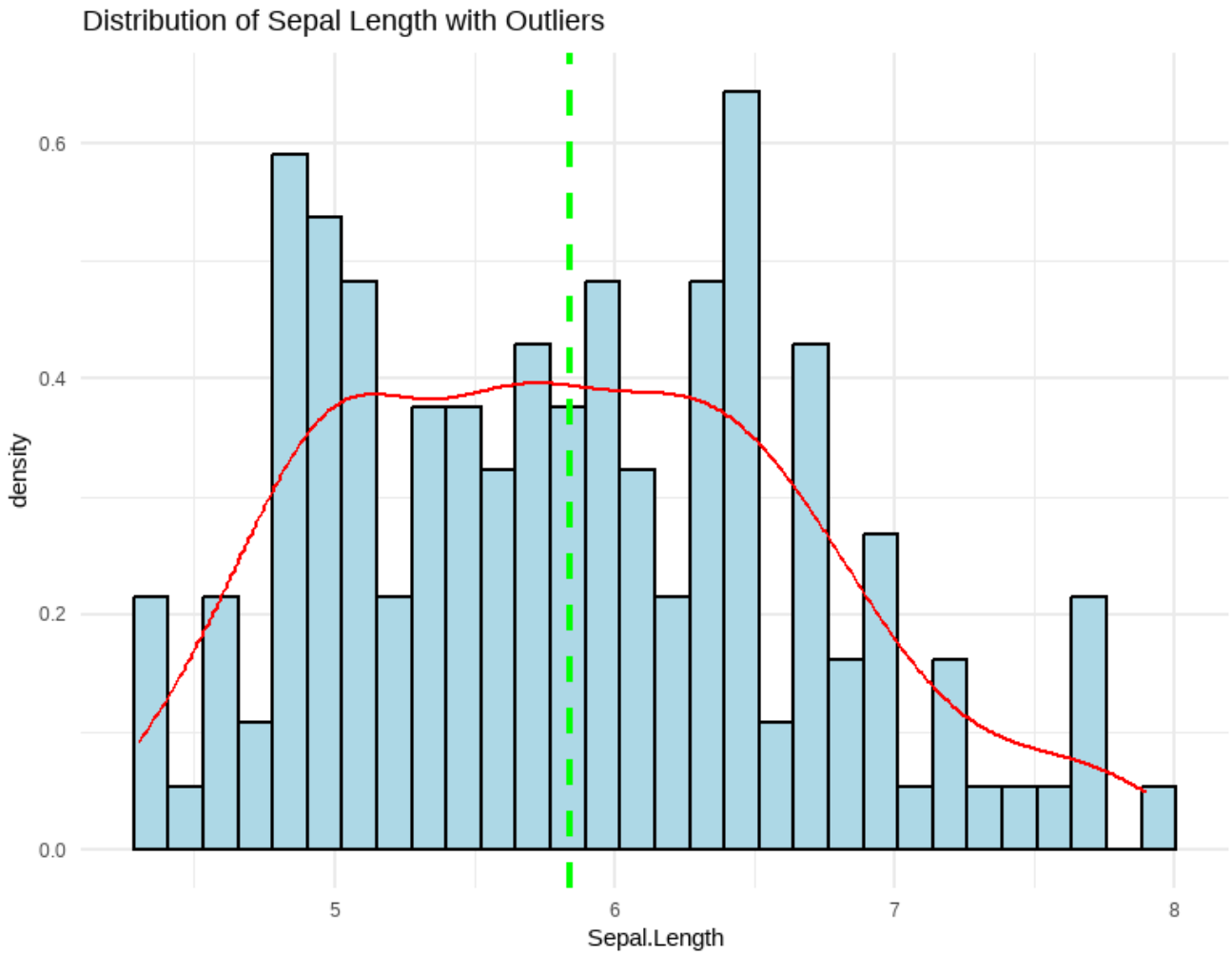


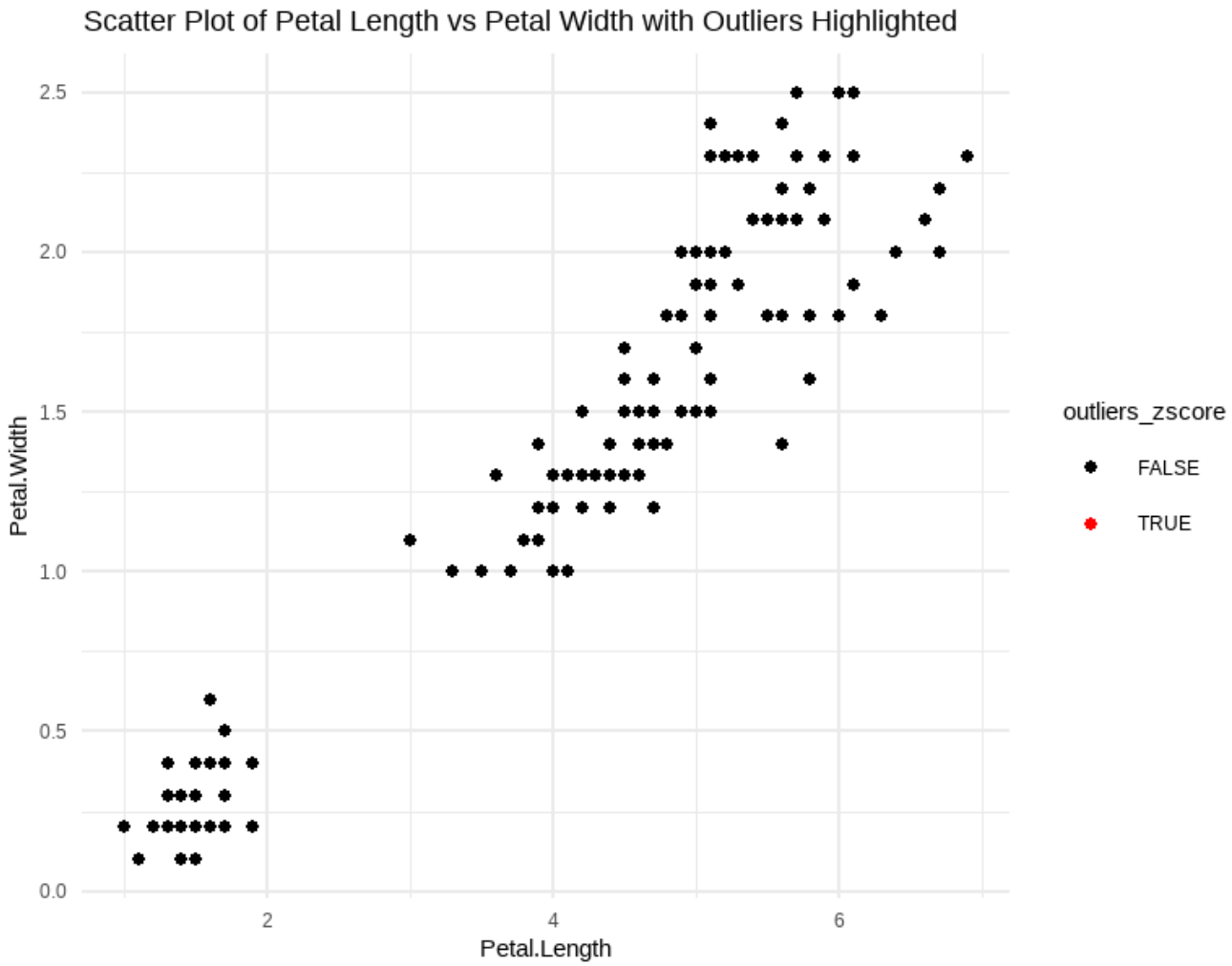



RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
R 4.4.1 · ~/
52 6.2 2.2 4.5 1.5 versicolor
53 5.6 2.5 3.9 1.1 versicolor
54 6.1 2.8 4.0 1.3 versicolor
55 6.3 2.5 4.9 1.5 versicolor
56 6.1 2.8 4.7 1.2 versicolor
57 6.6 3.0 4.4 1.4 versicolor
58 6.8 2.8 4.8 1.4 versicolor
59 6.7 3.0 5.0 1.7 versicolor
60 5.7 2.6 3.5 1.0 versicolor
61 5.5 2.4 3.8 1.1 versicolor
62 5.5 2.4 3.7 1.0 versicolor
63 6.0 2.7 5.1 1.6 versicolor
64 5.4 3.0 4.5 1.5 versicolor
65 6.0 3.4 4.5 1.6 versicolor
66 6.3 2.3 4.4 1.3 versicolor
67 5.6 3.0 4.1 1.3 versicolor
68 5.5 2.5 4.0 1.3 versicolor
69 6.1 3.0 4.6 1.4 versicolor
70 5.8 2.6 4.0 1.2 versicolor
71 5.0 2.3 3.3 1.0 versicolor
72 5.7 3.0 4.2 1.2 versicolor
73 5.7 2.9 4.2 1.3 versicolor
74 6.2 2.9 4.3 1.3 versicolor
75 5.7 2.8 4.1 1.3 versicolor
76 6.3 3.3 6.0 2.5 virginica
77 5.8 2.7 5.1 1.9 virginica
78 6.3 2.9 5.6 1.8 virginica
79 6.5 3.0 5.8 2.2 virginica
80 7.6 3.0 6.6 2.1 virginica
81 7.3 2.9 6.3 1.8 virginica
82 6.7 2.5 5.8 1.8 virginica
83 7.2 3.6 6.1 2.5 virginica
84 6.4 2.7 5.3 1.9 virginica
85 6.8 3.0 5.5 2.1 virginica
86 5.7 2.5 5.0 2.0 virginica
87 6.4 3.2 5.3 2.3 virginica
88 6.5 3.0 5.5 1.8 virginica
89 7.7 3.8 6.7 2.2 virginica
90 6.0 2.2 5.0 1.5 virginica
91 6.9 3.2 5.7 2.3 virginica
92 5.6 2.8 4.9 2.0 virginica
93 6.3 2.7 4.9 1.8 virginica
94 6.7 3.3 5.7 2.1 virginica
95 7.2 3.2 6.0 1.8 virginica
96 6.1 3.0 4.9 1.8 virginica
97 6.4 2.8 5.6 2.1 virginica
98 7.2 3.0 5.8 1.6 virginica
99 7.9 3.8 6.4 2.0 virginica
100 6.4 2.8 5.6 2.2 virginica
101 6.3 2.8 5.1 1.5 virginica
102 7.7 3.0 6.1 2.3 virginica
103 6.3 3.4 5.6 2.4 virginica
104 6.4 3.1 5.5 1.8 virginica
Environment History Connections Tutorial
Import Dataset 392 MiB
Global Environment
Data
iris 150 obs. of 5 variables
outliers_data 1 obs. of 5 variables
outliers_iqr 113 obs. of 5 variables
z_scores 150 obs. of 4 variables
Values
IQR Named num [1:4] 1.3 0.5 3.5 1.5
outliers_zscore logi [1:150] FALSE FALSE FALSE FALSE FALS…
Q1 Named num [1:4] 5.1 2.8 1.6 0.3
Q3 Named num [1:4] 6.4 3.3 5.1 1.8
threshold 3
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
Boxplot to Visualize Outliers
Sepal.Length
Sepal.Width
Petal.Length
Petal.Width


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
R 4.4.1 · ~/
69 6.1 3.0 4.6 1.4 versicolor
70 5.8 2.6 4.0 1.2 versicolor
71 5.0 2.3 3.3 1.0 versicolor
72 5.7 3.0 4.2 1.2 versicolor
73 5.7 2.9 4.2 1.3 versicolor
74 6.2 2.9 4.3 1.3 versicolor
75 5.7 2.8 4.1 1.3 versicolor
76 6.3 3.3 6.0 2.5 virginica
77 5.8 2.7 5.1 1.9 virginica
78 6.3 2.9 5.6 1.8 virginica
79 6.5 3.0 5.8 2.2 virginica
80 7.6 3.0 6.6 2.1 virginica
81 7.3 2.9 6.3 1.8 virginica
82 6.7 2.5 5.8 1.8 virginica
83 7.2 3.6 6.1 2.5 virginica
84 6.4 2.7 5.3 1.9 virginica
85 6.8 3.0 5.5 2.1 virginica
86 5.7 2.5 5.0 2.0 virginica
87 6.4 3.2 5.3 2.3 virginica
88 6.5 3.0 5.5 1.8 virginica
89 7.7 3.8 6.7 2.2 virginica
90 6.0 2.2 5.0 1.5 virginica
91 6.9 3.2 5.7 2.3 virginica
92 5.6 2.8 4.9 2.0 virginica
93 6.3 2.7 4.9 1.8 virginica
94 6.7 3.3 5.7 2.1 virginica
95 7.2 3.2 6.0 1.8 virginica
96 6.1 3.0 4.9 1.8 virginica
97 6.4 2.8 5.6 2.1 virginica
98 7.2 3.0 5.8 1.6 virginica
99 7.9 3.8 6.4 2.0 virginica
100 6.4 2.8 5.6 2.2 virginica
101 6.3 2.8 5.1 1.5 virginica
102 7.7 3.0 6.1 2.3 virginica
103 6.3 3.4 5.6 2.4 virginica
104 6.4 3.1 5.5 1.8 virginica
105 6.9 3.1 5.4 2.1 virginica
106 6.7 3.1 5.6 2.4 virginica
107 6.9 3.1 5.1 2.3 virginica
108 6.8 3.2 5.9 2.3 virginica
109 6.7 3.3 5.7 2.5 virginica
110 6.7 3.0 5.2 2.3 virginica
111 6.5 3.0 5.2 2.0 virginica
112 6.2 3.4 5.4 2.3 virginica
113 5.9 3.0 5.1 1.8 virginica
> # Visualization 3: Scatter plot for Petal Length vs Petal Width highlighting outliers
> ggplot(iris, aes(x = Petal.Length, y = Petal.Width, color = outliers_zscore)) +
+ geom_point() +
+ ggtitle("Scatter Plot of Petal Length vs Petal Width with Outliers Highlighted") +
+ theme_minimal() +
+ scale_color_manual(values = c("black", "red"))
Environment History Connections Tutorial
Import Dataset 392 MiB
Global Environment
Data
iris 150 obs. of 5 variables
outliers_data 1 obs. of 5 variables
outliers_iqr 113 obs. of 5 variables
z_scores 150 obs. of 4 variables
Values
IQR Named num [1:4] 1.3 0.5 3.5 1.5
outliers_zscore logi [1:150] FALSE FALSE FALSE FALSE FALS…
Q1 Named num [1:4] 5.1 2.8 1.6 0.3
Q3 Named num [1:4] 6.4 3.3 5.1 1.8
threshold 3
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
Boxplot to Visualize Outliers
Sepal.Length
Sepal.Width
Petal.Length
Petal.Width


Boxplot to Visualize Outliers
0
2
4
6
8
Sepal.Length
Sepal.Width
Petal.Length
Petal.Width


Distribution of Sepal Length with Outliers
density
0.0
0.2
0.4
0.6
5
6
7
8
Sepal.Length



OUTPUT

```
> #21BDS0085
> #Experiment-7
> library(ggplot2)
> library(dplyr)
>
> data("iris")
>
```

```
> # Display basic summary statistics
> summary(iris)
 Sepal.Length    Sepal.Width     Petal.Length    Petal.Width          Species
 Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100   setosa   :50
 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300   versicolor:50
 Median :5.800   Median :3.000   Median :4.350   Median :1.300   virginica :50
 Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
 3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
 Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
>
> # Z-score method to detect outliers
> z_scores <- as.data.frame(scale(iris[, 1:4]))  # Scaling numeric columns
>
> # Set Z-score threshold
> threshold <- 3
> outliers_zscore <- rowSums(abs(z_scores) > threshold) > 0
> outliers_data <- iris[outliers_zscore, ]
>
> print("Detected outliers based on Z-score:")
[1] "Detected outliers based on Z-score:"
> print(outliers_data)
   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
16          5.7         4.4          1.5         0.4  setosa
>
> # Visualization 1: Boxplot to visualize outliers
```

```
> boxplot(iris[, 1:4], main = "Boxplot to Visualize Outliers", col = "lightblue")
>
> # Visualization 2: Sepal Length distribution with outliers highlighted
> ggplot(iris, aes(x = Sepal.Length)) +
+   geom_histogram(aes(y = after_stat(density)), bins = 30, fill = "lightblue", color = "black") +
+   geom_density(color = "red") +
+   geom_vline(aes(xintercept = mean(Sepal.Length)), color = "green", linetype = "dashed", linewidth = 1) +
+   ggtitle("Distribution of Sepal Length with Outliers") +
+   theme_minimal()
>
> # IQR (Interquartile Range) method to detect outliers
> Q1 <- apply(iris[, 1:4], 2, quantile, 0.25)
> Q3 <- apply(iris[, 1:4], 2, quantile, 0.75)
> IQR <- Q3 - Q1
>
> # Define outlier condition using IQR
> outliers_iqr <- iris %>%
+   filter(if_any(everything(), ~ . < (Q1 - 1.5 * IQR) | . > (Q3 + 1.5 * IQR)))
Warning message:
There were 10 warnings in `filter()`.
The first warning was:
ℹ In argument: `|...`.
Caused by warning in `Sepal.Length < (Q1 - 1.5 * IQR)`:
! longer object length is not a multiple of shorter object length
ℹ Run dplyr::last_dplyr_warnings() to see the 9 remaining warnings.
```

```
>
> print("Outliers detected using the IQR method:")
[1] "Outliers detected using the IQR method:"
> print(outliers_iqr)
   Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
1           5.1         3.5          1.4         0.2     setosa
2           4.9         3.0          1.4         0.2     setosa
3           4.6         3.1          1.5         0.2     setosa
4           5.0         3.6          1.4         0.2     setosa
5           5.4         3.9          1.7         0.4     setosa
6           5.0         3.4          1.5         0.2     setosa
7           4.4         2.9          1.4         0.2     setosa
8           4.9         3.1          1.5         0.1     setosa
9           4.8         3.4          1.6         0.2     setosa
10          4.8         3.0          1.4         0.1     setosa
11          4.3         3.0          1.1         0.1     setosa
12          5.7         4.4          1.5         0.4     setosa
13          5.4         3.9          1.3         0.4     setosa
14          5.1         3.5          1.4         0.3     setosa
15          5.1         3.8          1.5         0.3     setosa
16          5.4         3.4          1.7         0.2     setosa
17          5.1         3.7          1.5         0.4     setosa
18          5.1         3.3          1.7         0.5     setosa
19          4.8         3.4          1.9         0.2     setosa
20          5.0         3.0          1.6         0.2     setosa
```

21	5.2	3.5	1.5	0.2	setosa
22	5.2	3.4	1.4	0.2	setosa
23	4.7	3.2	1.6	0.2	setosa
24	5.4	3.4	1.5	0.4	setosa
25	5.2	4.1	1.5	0.1	setosa
26	5.5	4.2	1.4	0.2	setosa
27	5.0	3.2	1.2	0.2	setosa
28	5.5	3.5	1.3	0.2	setosa
29	4.9	3.6	1.4	0.1	setosa
30	5.1	3.4	1.5	0.2	setosa
31	5.0	3.5	1.3	0.3	setosa
32	4.5	2.3	1.3	0.3	setosa
33	5.0	3.5	1.6	0.6	setosa
34	5.1	3.8	1.9	0.4	setosa
35	4.8	3.0	1.4	0.3	setosa
36	4.6	3.2	1.4	0.2	setosa
37	5.3	3.7	1.5	0.2	setosa
38	5.0	3.3	1.4	0.2	setosa
39	6.4	3.2	4.5	1.5	versicolor
40	6.9	3.1	4.9	1.5	versicolor
41	5.5	2.3	4.0	1.3	versicolor
42	5.7	2.8	4.5	1.3	versicolor
43	6.3	3.3	4.7	1.6	versicolor
44	4.9	2.4	3.3	1.0	versicolor
45	5.2	2.7	3.9	1.4	versicolor

46	5.0	2.0	3.5	1.0 versicolor
47	5.9	3.0	4.2	1.5 versicolor
48	6.1	2.9	4.7	1.4 versicolor
49	5.6	2.9	3.6	1.3 versicolor
50	6.7	3.1	4.4	1.4 versicolor
51	5.8	2.7	4.1	1.0 versicolor
52	6.2	2.2	4.5	1.5 versicolor
53	5.6	2.5	3.9	1.1 versicolor
54	6.1	2.8	4.0	1.3 versicolor
55	6.3	2.5	4.9	1.5 versicolor
56	6.1	2.8	4.7	1.2 versicolor
57	6.6	3.0	4.4	1.4 versicolor
58	6.8	2.8	4.8	1.4 versicolor
59	6.7	3.0	5.0	1.7 versicolor
60	5.7	2.6	3.5	1.0 versicolor
61	5.5	2.4	3.8	1.1 versicolor
62	5.5	2.4	3.7	1.0 versicolor
63	6.0	2.7	5.1	1.6 versicolor
64	5.4	3.0	4.5	1.5 versicolor
65	6.0	3.4	4.5	1.6 versicolor
66	6.3	2.3	4.4	1.3 versicolor
67	5.6	3.0	4.1	1.3 versicolor
68	5.5	2.5	4.0	1.3 versicolor
69	6.1	3.0	4.6	1.4 versicolor
70	5.8	2.6	4.0	1.2 versicolor

71	5.0	2.3	3.3	1.0	versicolor
72	5.7	3.0	4.2	1.2	versicolor
73	5.7	2.9	4.2	1.3	versicolor
74	6.2	2.9	4.3	1.3	versicolor
75	5.7	2.8	4.1	1.3	versicolor
76	6.3	3.3	6.0	2.5	virginica
77	5.8	2.7	5.1	1.9	virginica
78	6.3	2.9	5.6	1.8	virginica
79	6.5	3.0	5.8	2.2	virginica
80	7.6	3.0	6.6	2.1	virginica
81	7.3	2.9	6.3	1.8	virginica
82	6.7	2.5	5.8	1.8	virginica
83	7.2	3.6	6.1	2.5	virginica
84	6.4	2.7	5.3	1.9	virginica
85	6.8	3.0	5.5	2.1	virginica
86	5.7	2.5	5.0	2.0	virginica
87	6.4	3.2	5.3	2.3	virginica
88	6.5	3.0	5.5	1.8	virginica
89	7.7	3.8	6.7	2.2	virginica
90	6.0	2.2	5.0	1.5	virginica
91	6.9	3.2	5.7	2.3	virginica
92	5.6	2.8	4.9	2.0	virginica
93	6.3	2.7	4.9	1.8	virginica
94	6.7	3.3	5.7	2.1	virginica
95	7.2	3.2	6.0	1.8	virginica

```
96      6.1     3.0     4.9     1.8 virginica
97      6.4     2.8     5.6     2.1 virginica
98      7.2     3.0     5.8     1.6 virginica
99      7.9     3.8     6.4     2.0 virginica
100     6.4     2.8     5.6     2.2 virginica
101     6.3     2.8     5.1     1.5 virginica
102     7.7     3.0     6.1     2.3 virginica
103     6.3     3.4     5.6     2.4 virginica
104     6.4     3.1     5.5     1.8 virginica
105     6.9     3.1     5.4     2.1 virginica
106     6.7     3.1     5.6     2.4 virginica
107     6.9     3.1     5.1     2.3 virginica
108     6.8     3.2     5.9     2.3 virginica
109     6.7     3.3     5.7     2.5 virginica
110     6.7     3.0     5.2     2.3 virginica
111     6.5     3.0     5.2     2.0 virginica
112     6.2     3.4     5.4     2.3 virginica
113     5.9     3.0     5.1     1.8 virginica
>
> # Visualization 3: Scatter plot for Petal Length vs Petal Width highlighting outliers
> ggplot(iris, aes(x = Petal.Length, y = Petal.Width, color = outliers_zscore)) +
+   geom_point() +
+   ggtitle("Scatter Plot of Petal Length vs Petal Width with Outliers Highlighted") +
+   theme_minimal() +
+   scale_color_manual(values = c("black", "red"))
```