APPLIED STATISTICAL ANALYSIS LAB

NAME: ADITI KULKARNI

ROLL NO: 55

YEAR: SY

DIVISION: E (E3)

SRN NO: 202201893

ASSIGNMENT 5

STATEMENT: *Visualize the relationship between two scale variables creating scatter plots and to quantify this relationship with the correlation coefficient.*

THEORY:

1. *Reading Data: Reads a dataset from a CSV file.*

- *Data Exploration:*
- *Checks the number of rows and columns.*
- *Inspects the data types of attributes (columns).*

2. *Creating Cross-Tabulation Tables: Compares rainfall data between 'Jan-Feb' and 'Mar-May' months.*

3. *Creating a Scatter Plot: Generates a scatter plot to visualize the relationship between 'Jan-Feb' and 'Mar-May' rainfall.*

4. *Calculating Correlation: Computes the correlation coefficient to quantify the strength and direction of the relationship between 'Jan-Feb' and 'Mar-May' rainfall.*

5. *Correlation Matrix: Creates a correlation matrix to understand how all variables in the dataset relate to each other.*

6. *Visualizing Correlations: Uses the corrplot library to visually represent the correlations between variables in the dataset using colors.*

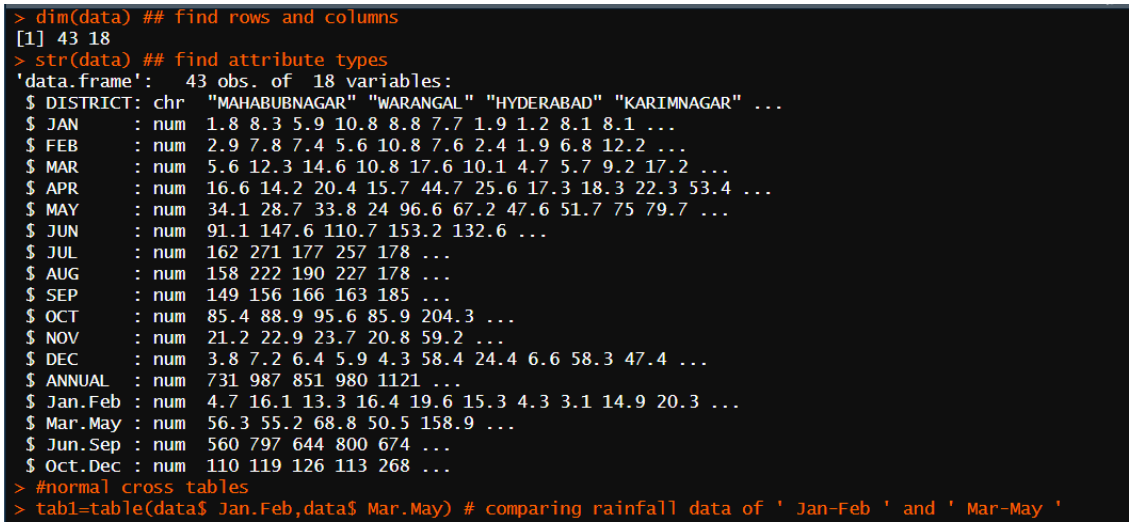
7. *Views the dataset.*

SOURCE CODE:

```
#Visualize the relationship between two scale variables creating scatter plots and to quantify this relationship with the correlation coefficient.

data<-read.csv(file.choose()) ## Read the dataset

View(data) ## View the data-frame

dim(data) ## find rows and columns

str(data) ## find attribute types

#normal cross tables

tab1=table(data$ Jan.Feb,data$ Mar.May) # comparing rainfall data of ' Jan-Feb ' and ' Mar-May '

tab1

margin.table(tab1) # row totals

margin.table(tab1) # columns totals
```

```
prop.table(tab1) # proportions based on overall totals

# Scatterplot

plot(data$Jan.Feb, data$Mar.May,

    xlab = "Jan-Feb", ylab = "Mar-May",

    main = "Scatter Plot")

# Calculate the correlation coefficient between ' Jan-Feb ' and ' Mar-May '

cor_coeff <- cor(data$ Jan.Feb, data$ Mar.May)

# Display the correlation coefficient

cat("Correlation Coefficient:", round(cor_coeff, 2))

# Create a correlation matrix for all variables in the dataset

cor(data[, c(2:13,14)])

# Visualize the correlation matrix using corrplot

library(corrplot)

corrplot(cor(data[, c(2:13,14)]), method = "color")
```

OUTPUT:



```
> dim(data) ## find rows and columns
[1] 43 18
> str(data) ## find attribute types
'data.frame':   43 obs. of  18 variables:
 $ DISTRICT: chr  "MAHABUBNAGAR" "WARANGAL" "HYDERABAD" "KARIMNAGAR" ...
 $ JAN     : num  1.8 8.3 5.9 10.8 8.8 7.7 1.9 1.2 8.1 8.1 ...
 $ FEB     : num  2.9 7.8 7.4 5.6 10.8 7.6 2.4 1.9 6.8 12.2 ...
 $ MAR     : num  5.6 12.3 14.6 10.8 17.6 10.1 4.7 5.7 9.2 17.2 ...
 $ APR     : num  16.6 14.2 20.4 15.7 44.7 25.6 17.3 18.3 22.3 53.4 ...
 $ MAY     : num  34.1 28.7 33.8 24 96.6 67.2 47.6 51.7 75 79.7 ...
 $ JUN     : num  91.1 147.6 110.7 153.2 132.6 ...
 $ JUL     : num  162 271 177 257 178 ...
 $ AUG     : num  158 222 190 227 178 ...
 $ SEP     : num  149 156 166 163 185 ...
 $ OCT     : num  85.4 88.9 95.6 85.9 204.3 ...
 $ NOV     : num  21.2 22.9 23.7 20.8 59.2 ...
 $ DEC     : num  3.8 7.2 6.4 5.9 4.3 58.4 24.4 6.6 58.3 47.4 ...
 $ ANNUAL  : num  731 987 851 980 1121 ...
 $ Jan.Feb : num  4.7 16.1 13.3 16.4 19.6 15.3 4.3 3.1 14.9 20.3 ...
 $ Mar.May : num  56.3 55.2 68.8 50.5 158.9 ...
 $ Jun.Sep : num  560 797 644 800 674 ...
 $ Oct.Dec : num  110 119 126 113 268 ...
> #normal cross tables
> tab1=table(data$ Jan.Feb,data$ Mar.May) # comparing rainfall data of ' Jan-Feb ' and ' Mar-May '
```

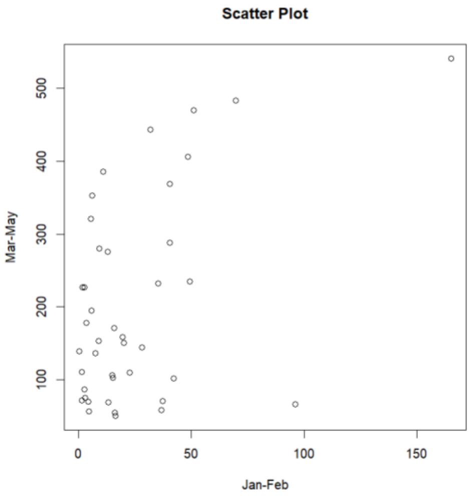
```
> tab1

      50.5 55.2 56.3 58.5 66.6 68.8 69.6 70.9 72.2 75.7 87 102.1 102.9 106.5 109.9 111.2 136.9 138.8
  0.4    0    0    0    0    0    0    0    0    0    0  0     0     0     0     0     0     0     1
  1.6    0    0    0    0    0    0    0    0    1    0  0     0     0     0     0     1     0     0
  1.8    0    0    0    0    0    0    0    0    0    0  0     0     0     0     0     0     0     0
  2.6    0    0    0    0    0    0    0    0    0    0  0     0     0     0     0     0     0     0
  2.8    0    0    0    0    0    0    0    0    0    0  1     0     0     0     0     0     0     0
  3.1    0    0    0    0    0    0    0    0    0    1  0     0     0     0     0     0     0     0
  3.6    0    0    0    0    0    0    0    0    0    0  0     0     0     0     0     0     0     0
  4.3    0    0    0    0    0    0    1    0    0    0  0     0     0     0     0     0     0     0
  4.7    0    0    1    0    0    0    0    0    0    0  0     0     0     0     0     0     0     0
  5.6    0    0    0    0    0    0    0    0    0    0  0     0     0     0     0     0     0     0
  6      0    0    0    0    0    0    0    0    0    0  0     0     0     0     0     0     0     0
  6.2    0    0    0    0    0    0    0    0    0    0  0     0     0     0     0     0     0     0
  7.5    0    0    0    0    0    0    0    0    0    0  0     0     0     0     0     1     0     0
  9      0    0    0    0    0    0    0    0    0    0  0     0     0     0     0     0     0     0
  9.4    0    0    0    0    0    0    0    0    0    0  0     0     0     0     0     0     0     0
  10.9   0    0    0    0    0    0    0    0    0    0  0     0     0     0     0     0     0     0
  13.1   0    0    0    0    0    0    0    0    0    0  0     0     0     0     0     0     0     0
  13.3   0    0    0    0    0    1    0    0    0    0  0     0     0     0     0     0     0     0
  14.9   0    0    0    0    0    0    0    0    0    0  0     0     0     1     0     0     0     0
  15.3   0    0    0    0    0    0    0    0    0    0  0     0     1     0     0     0     0     0
  16     0    0    0    0    0    0    0    0    0    0  0     0     0     0     0     0     0     0
  16.1   0    1    0    0    0    0    0    0    0    0  0     0     0     0     0     0     0     0
  16.4   1    0    0    0    0    0    0    0    0    0  0     0     0     0     0     0     0     0
  19.6   0    0    0    0    0    0    0    0    0    0  0     0     0     0     0     0     0     0
```

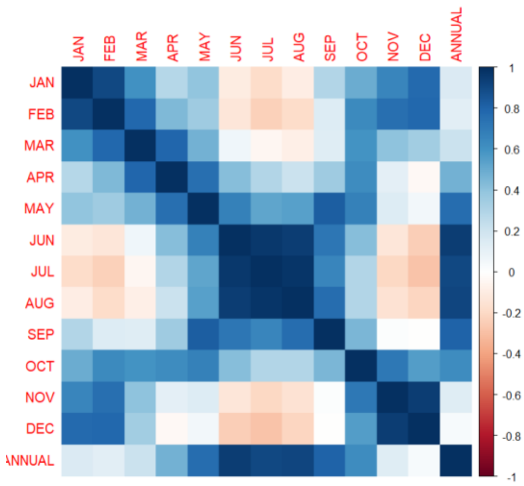
```
      144.8 150.3 153.7 158.9 170.8 178.1 194.7 227.1 227.3 232.4 235.3 275.4 279.9 288.3 320.8 352.6
0.4       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
1.6       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
1.8       0     0     0     0     0     0     0     0     1     0     0     0     0     0     0     0
2.6       0     0     0     0     0     0     0     1     0     0     0     0     0     0     0     0
2.8       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
3.1       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
3.6       0     0     0     0     0     1     0     0     0     0     0     0     0     0     0     0
4.3       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
4.7       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
5.6       0     0     0     0     0     0     0     0     0     0     0     0     0     0     1     0
6         0     0     0     0     0     0     1     0     0     0     0     0     0     0     0     0
6.2       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     1
7.5       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
9         0     0     1     0     0     0     0     0     0     0     0     0     0     0     0     0
9.4       0     0     0     0     0     0     0     0     0     0     0     0     1     0     0     0
10.9      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
13.1      0     0     0     0     0     0     0     0     0     0     0     1     0     0     0     0
13.3      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
14.9      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
15.3      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
16        0     0     0     0     1     0     0     0     0     0     0     0     0     0     0     0
16.1      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
16.4      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
19.6      0     0     0     1     0     0     0     0     0     0     0     0     0     0     0     0
```

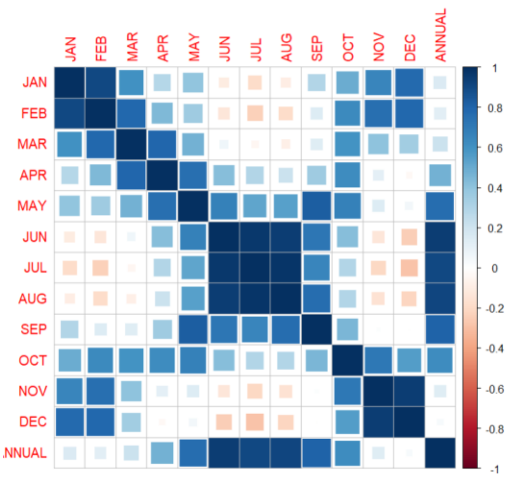
```
       368.8 385.3 405.6 443.6 469.6 483.5 540.7
  0.4      0     0     0     0     0     0     0
  1.6      0     0     0     0     0     0     0
  1.8      0     0     0     0     0     0     0
  2.6      0     0     0     0     0     0     0
  2.8      0     0     0     0     0     0     0
  3.1      0     0     0     0     0     0     0
  3.6      0     0     0     0     0     0     0
  4.3      0     0     0     0     0     0     0
  4.7      0     0     0     0     0     0     0
  5.6      0     0     0     0     0     0     0
  6        0     0     0     0     0     0     0
  6.2      0     0     0     0     0     0     0
  7.5      0     0     0     0     0     0     0
  9        0     0     0     0     0     0     0
  9.4      0     0     0     0     0     0     0
  10.9     0     1     0     0     0     0     0
  13.1     0     0     0     0     0     0     0
  13.3     0     0     0     0     0     0     0
  14.9     0     0     0     0     0     0     0
  15.3     0     0     0     0     0     0     0
  16       0     0     0     0     0     0     0
  16.1     0     0     0     0     0     0     0
  16.4     0     0     0     0     0     0     0
  19.6     0     0     0     0     0     0     0
 [ reached getOption("max.print") -- omitted 15 rows ]
> margin.table(tab1) # row totals
[1] 43
> margin.table(tab1) # columns totals
[1] 43
```

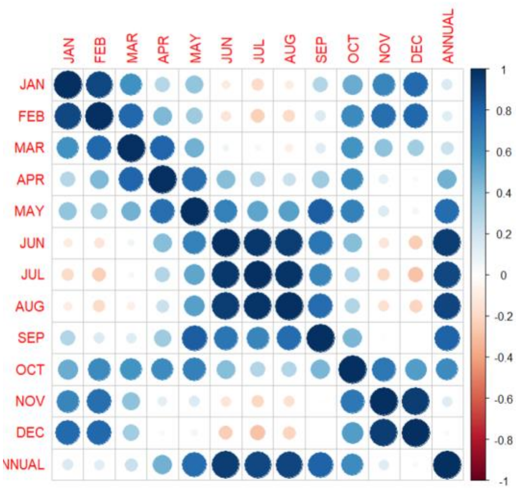
```
> # Display the correlation coefficient
> cat("Correlation Coefficient:", round(cor_coeff, 2))
Correlation Coefficient: 0.45> # Create a correlation matrix for all variables in the dataset
> cor(data[, c(2:13,14)])
              JAN         FEB         MAR         APR         MAY         JUN         JUL         AUG
JAN     1.00000000  0.9042358  0.60925074  0.28750217 0.39678652 -0.10365754 -0.18648139 -0.09684422
FEB     0.90423581  1.0000000  0.78940884  0.44059998 0.35695894 -0.13277562 -0.23602384 -0.18960915
MAR     0.60925074  0.7894088  1.00000000  0.79133651 0.47647806  0.06476721 -0.04772801 -0.08025767
APR     0.28750217  0.4406000  0.79133651  1.00000000 0.75810201  0.42562719  0.29223031  0.21022771
MAY     0.39678652  0.3569589  0.47647806  0.75810201 1.00000000  0.67672981  0.52550458  0.54333387
JUN    -0.10365754 -0.1327756  0.06476721  0.42562719 0.67672981  1.00000000  0.96816676  0.94302454
JUL    -0.18648139 -0.2360238 -0.04772801  0.29223031 0.52550458  0.96816676  1.00000000  0.97042673
AUG    -0.09684422 -0.1896091 -0.08025767  0.21022771 0.54333387  0.94302454  0.97042673  1.00000000
SEP     0.29647359  0.1453510  0.13212703  0.35899011 0.82198642  0.72864084  0.65247945  0.76342646
OCT     0.49151878  0.6313636  0.59627163  0.62511210 0.67538935  0.42431852  0.29824261  0.29867090
NOV     0.65311234  0.7594154  0.40078011  0.11385873 0.14417791 -0.13199923 -0.20184956 -0.15754658
DEC     0.77054224  0.7881165  0.34930931 -0.03628229 0.05873504 -0.24357309 -0.28765859 -0.21679597
ANNUAL  0.15908038  0.1279530  0.21573662  0.47930226 0.76462152  0.94990172  0.90757295  0.91639280
                SEP         OCT          NOV          DEC      ANNUAL
JAN     0.2964735907 0.4915188  0.65311234  0.7705422364 0.15908038
FEB     0.1453509911 0.6313636  0.75941543  0.7881164731 0.12795304
MAR     0.1321270290 0.5962716  0.40078011  0.3493093059 0.21573662
APR     0.3589901116 0.6251121  0.11385873 -0.0362822941 0.47930226
MAY     0.8219864161 0.6753893  0.14417791  0.0587350383 0.76462152
JUN     0.7286408435 0.4243185 -0.13199923 -0.2435730876 0.94990172
JUL     0.6524794530 0.2982426 -0.20184956 -0.2876585861 0.90757295
AUG     0.7634264646 0.2986709 -0.15754658 -0.2167959723 0.91639280
SEP     1.0000000000 0.4556494  0.01233041 -0.0009489843 0.80916097
OCT     0.4556494073 1.0000000  0.71968655  0.5582286668 0.62558667
NOV     0.0123304056 0.7196866  1.00000000  0.9443727175 0.13451852
DEC    -0.0009489843 0.5582287  0.94437272  1.0000000000 0.03452972
ANNUAL  0.8091609721 0.6255867  0.13451852  0.0345297211 1.00000000
> # Visualize the correlation matrix using corrplot
> library(corrplot)
corrplot 0.92 loaded
```

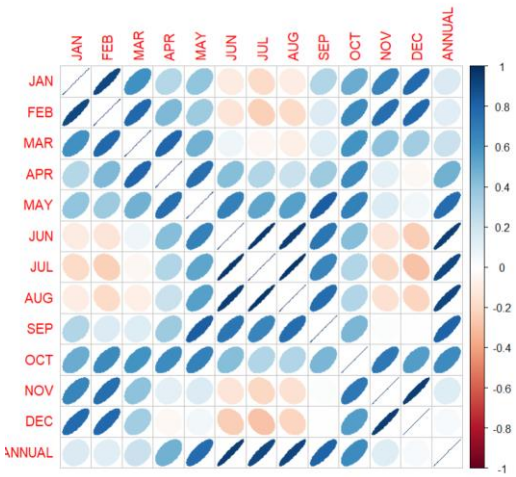
GRAPHS:

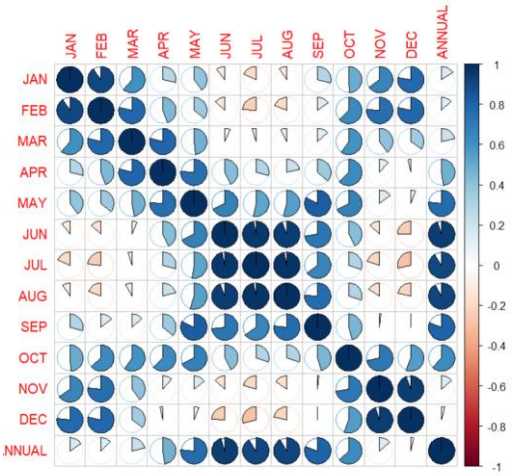

Scatter Plot
Mar-May
Jan-Feb
0
50
100
150
100
200
300
400
500

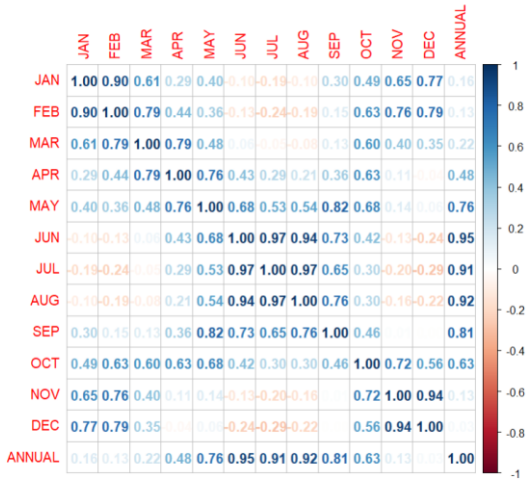

JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
ANNUAL

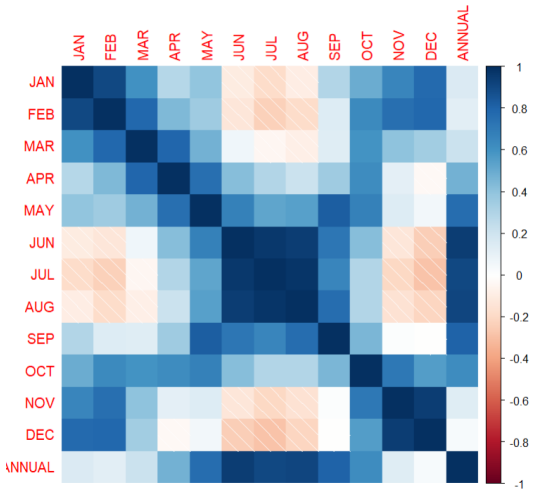

JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
ANNUAL


JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
ANNUAL









CONCLUSION: *This code reads a dataset, analyzes the relationship between 'Jan-Feb' and 'Mar-May' rainfall using a scatter plot and correlation coefficient. It also creates a correlation matrix and visualizes the correlations among all dataset variables using colors. Overall, it helps explore data patterns and assess the strength of the relationship between specific weather variables.*