EXPERIMENT NO : 08

AIM: Analyze relationships between multiple variables using visualizations and correlation analysis in R for better data interpretation.

Definition

- **Correlation** measures the **strength and direction** of the relationship between two numeric variables.
- The correlation coefficient (**r**) ranges between **-1 and 1**:
- **r = 1** → Perfect positive correlation (both variables increase together)
- **r = 0** → No correlation (no relationship between variables)
- **r = -1** → Perfect negative correlation (one variable increases, the other decreases)

Types of Correlation in R

- **Pearson correlation**: Measures the **linear** relationship between two continuous variables.
- **Spearman correlation**: Measures **monotonic** relationships (useful when data is not normally distributed).
- **Kendall correlation**: Used for **ordinal** data or small datasets.

Visualizing Relationships Between Multiple Variables

A. **Scatter Plots (Pairwise Relationships)** Used for examining relationships between two numeric variables.
B. **Pair Plot (Visualizing Relationships for Multiple Variables)** A pair plot helps visualize relationships between all numeric variables.
C. **Boxplots (Comparing a Numeric Variable Across Categories)** Useful for checking **distribution differences** across categorical variables.
D. **Bar Plots (Comparing Categorical Variables) Used to compare counts or proportions in categorical variables.**

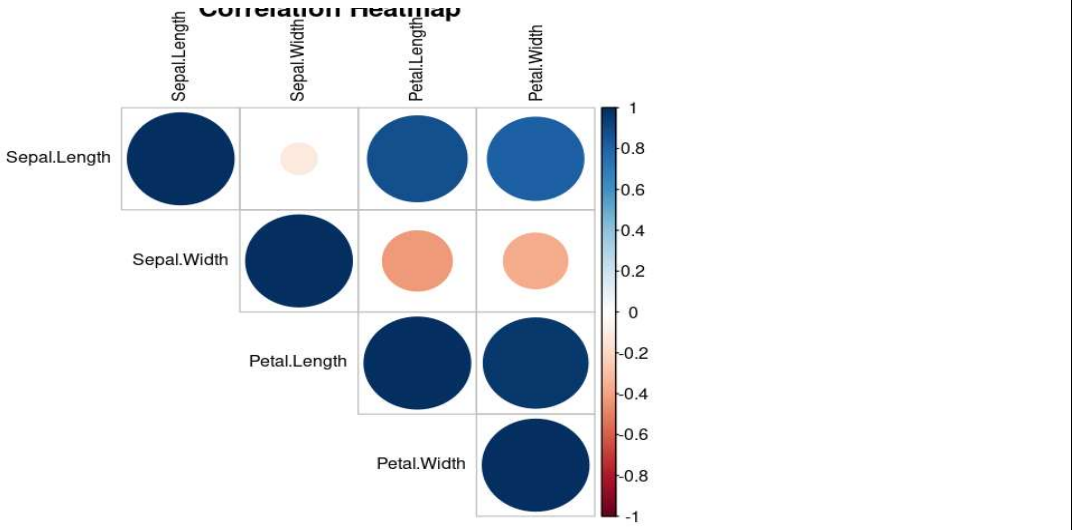
CODE:

```
library(ggplot2)

library(corrplot)

library(GGally)

# Load the Iris dataset

data(iris)

# 1. Scatter plot of Sepal.Length vs Petal.Length

ggplot(iris, aes(x = Sepal.Length, y = Petal.Length, color = Species)) +

  geom_point() +

  labs(title = "Scatter Plot of Sepal Length vs Petal Length") +
```

```
 theme_minimal()
# 2. Pair plot of all numeric variables in the Iris dataset
ggpairs(iris, aes(color = Species))
# 3. Compute and print the correlation matrix
cor_matrix <- cor(iris[, 1:4])  # Using only numeric columns
print(cor_matrix)
# 4. Correlation heatmap
corrplot(cor_matrix, method = "circle", type = "upper",
         title = "Correlation Heatmap", tl.cex = 0.8, tl.col = "black")
# 5. Pearson correlation between Sepal.Length and Petal.Length
pearson_corr <- cor(iris$Sepal.Length, iris$Petal.Length)
print(paste("Pearson Correlation between Sepal.Length and Petal.Length:", pearson_corr))
```

OUTPUT:

```
Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length    1.0000000  -0.1175698    0.8717538   0.8179411
Sepal.Width    -0.1175698   1.0000000   -0.4284401  -0.3661259
Petal.Length    0.8717538  -0.4284401    1.0000000   0.9628654
Petal.Width     0.8179411  -0.3661259    0.9628654   1.0000000
"Pearson Correlation between Sepal.Length and Petal.Length: 0.871753775886583"
```



CONCLUSION: Hence we successfully implemented relationships between multiple variables using visualizations and correlation analysis in R for better data interpretation.