CHAPTER 4

Data Exploration and Manipulation
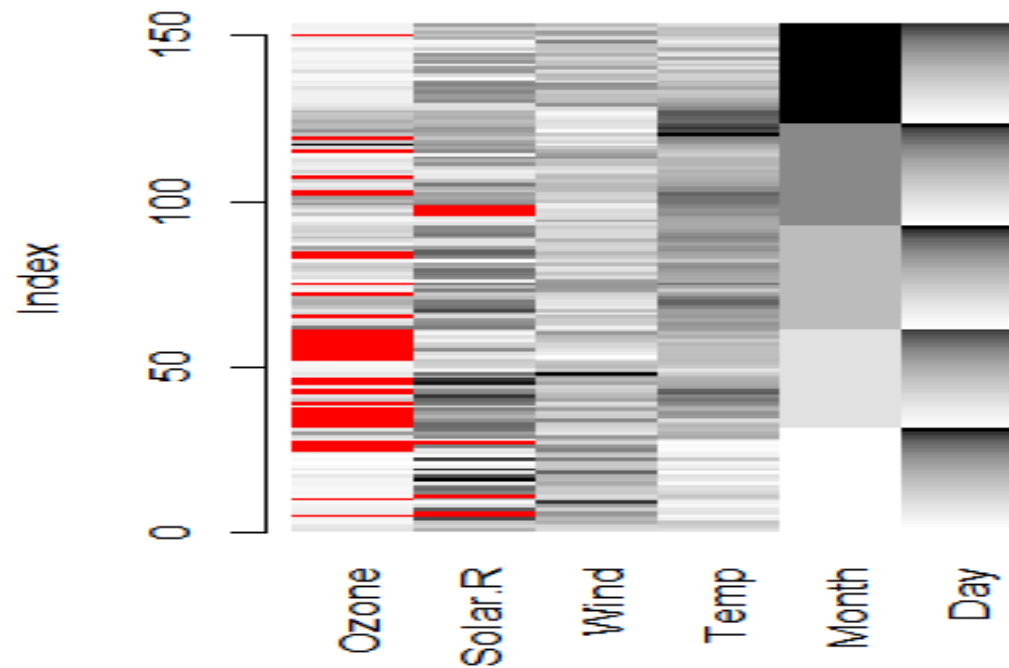

Figure 4.5 Missing values in "airquality" dataset.

CHAPTER 6
Basic Visualization
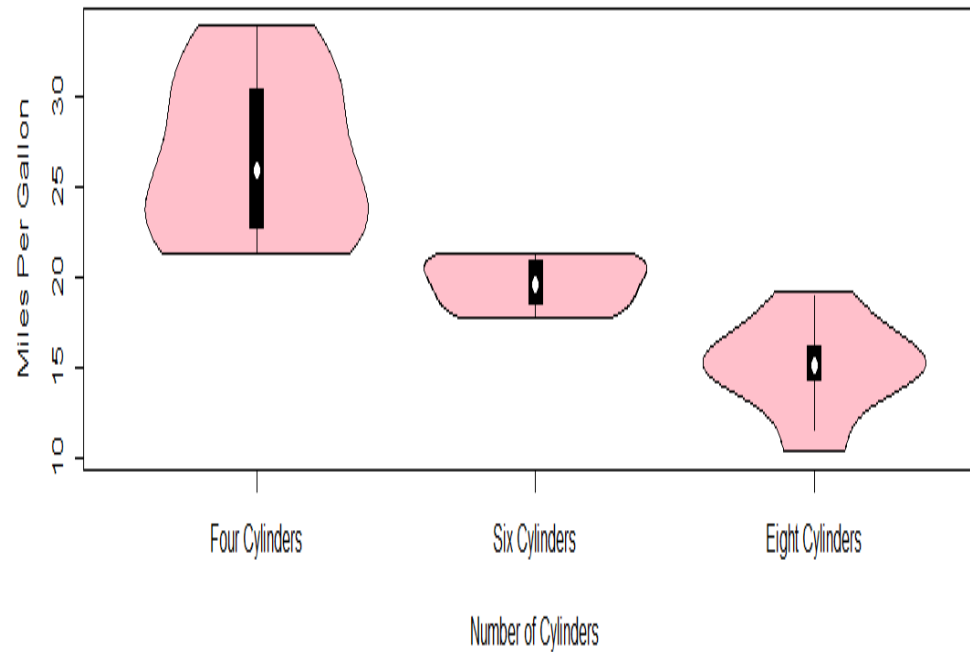

Figure 6.40 Violin plot.
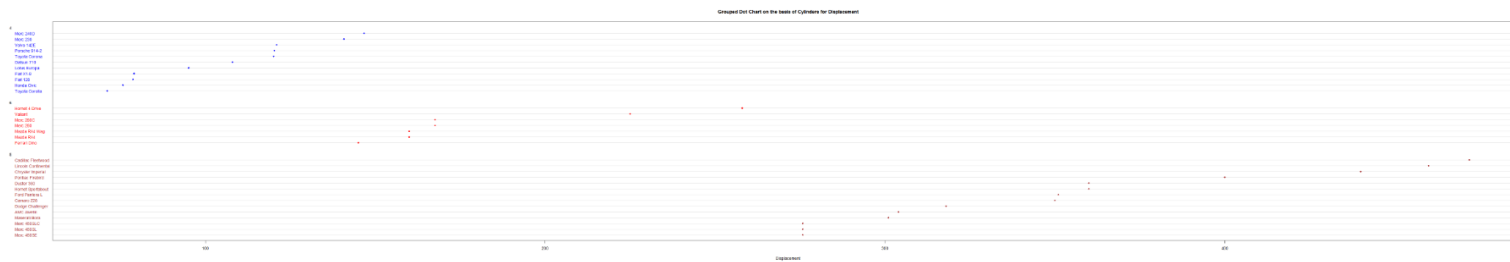
Figure 6.42 Grouped dot chart corresponding to car models for different cylinders in different colors.
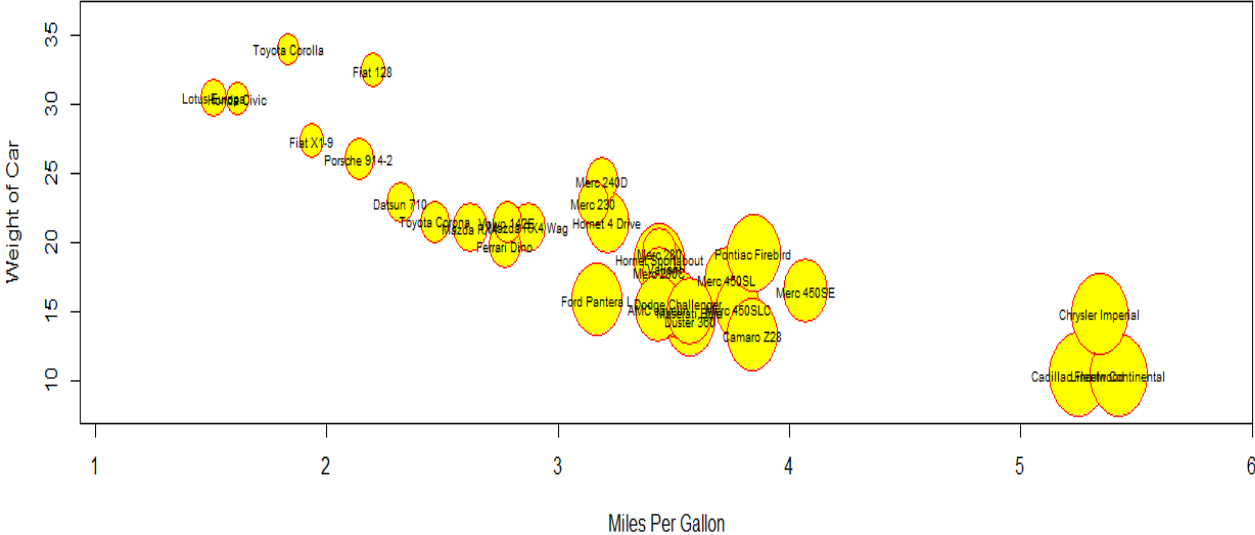
Figure 6.43 Bubble chart for different models.

Figure 6.44 Correlation between variables in "mtcars" dataset.
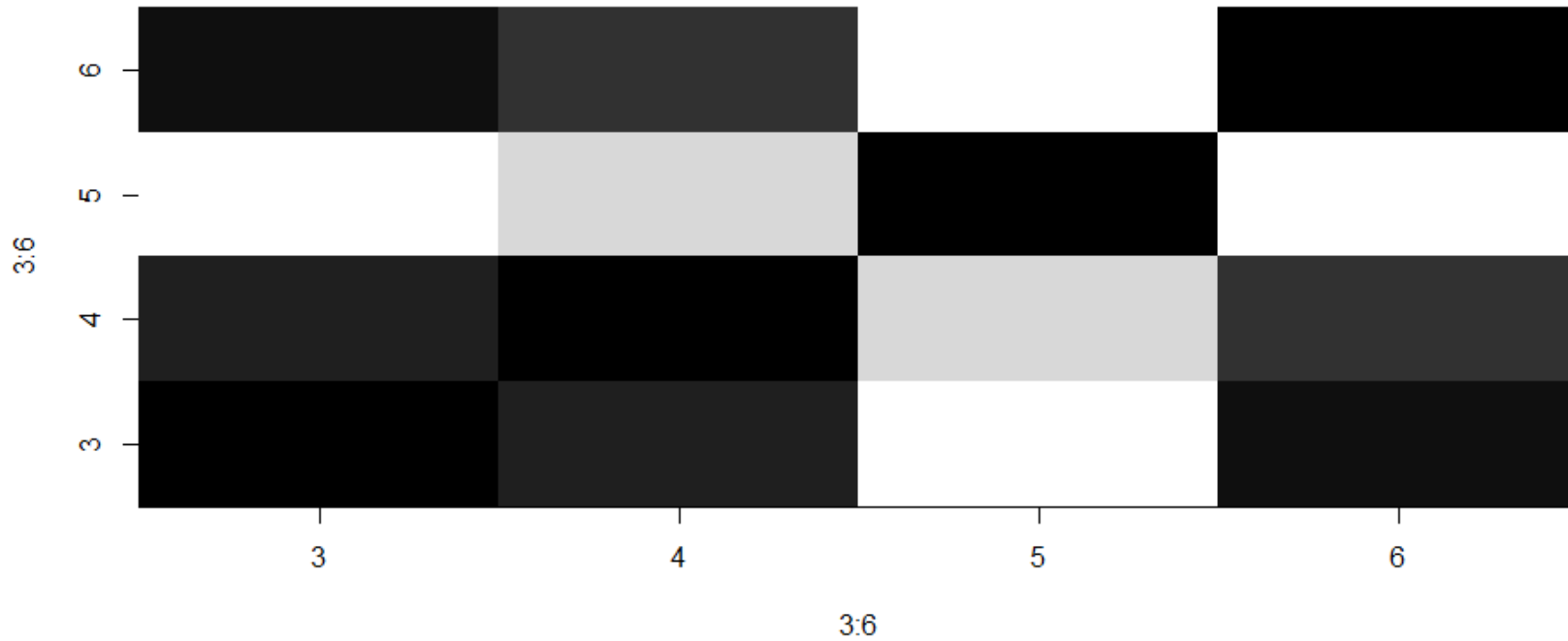
Figure 6.45 Correlation between variables in gray black scale of "mtcars" dataset.
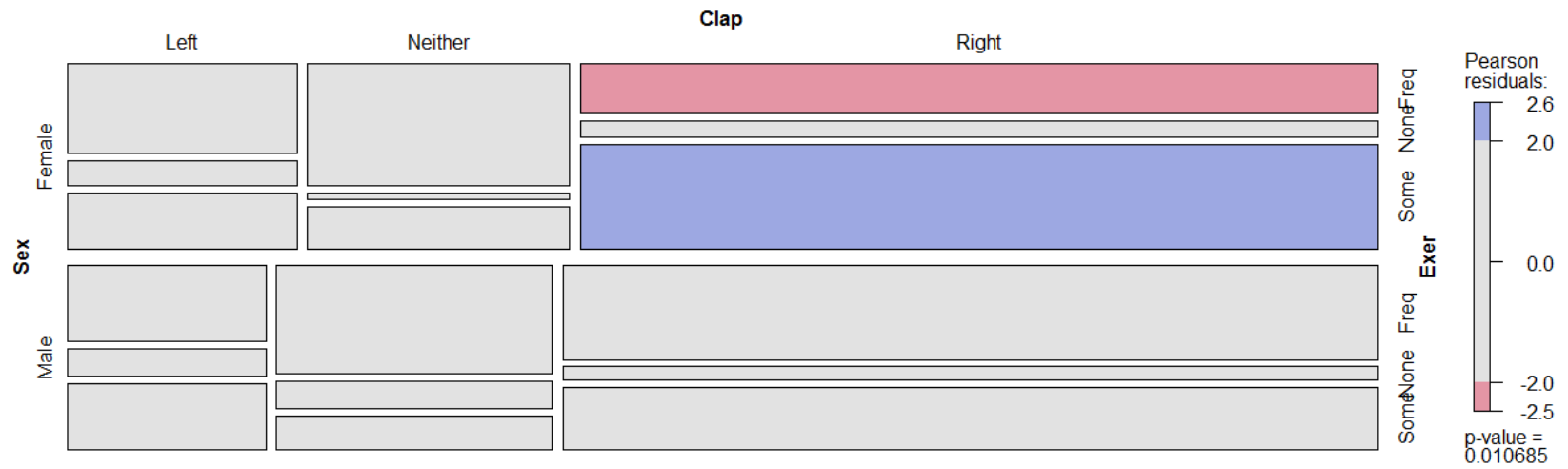
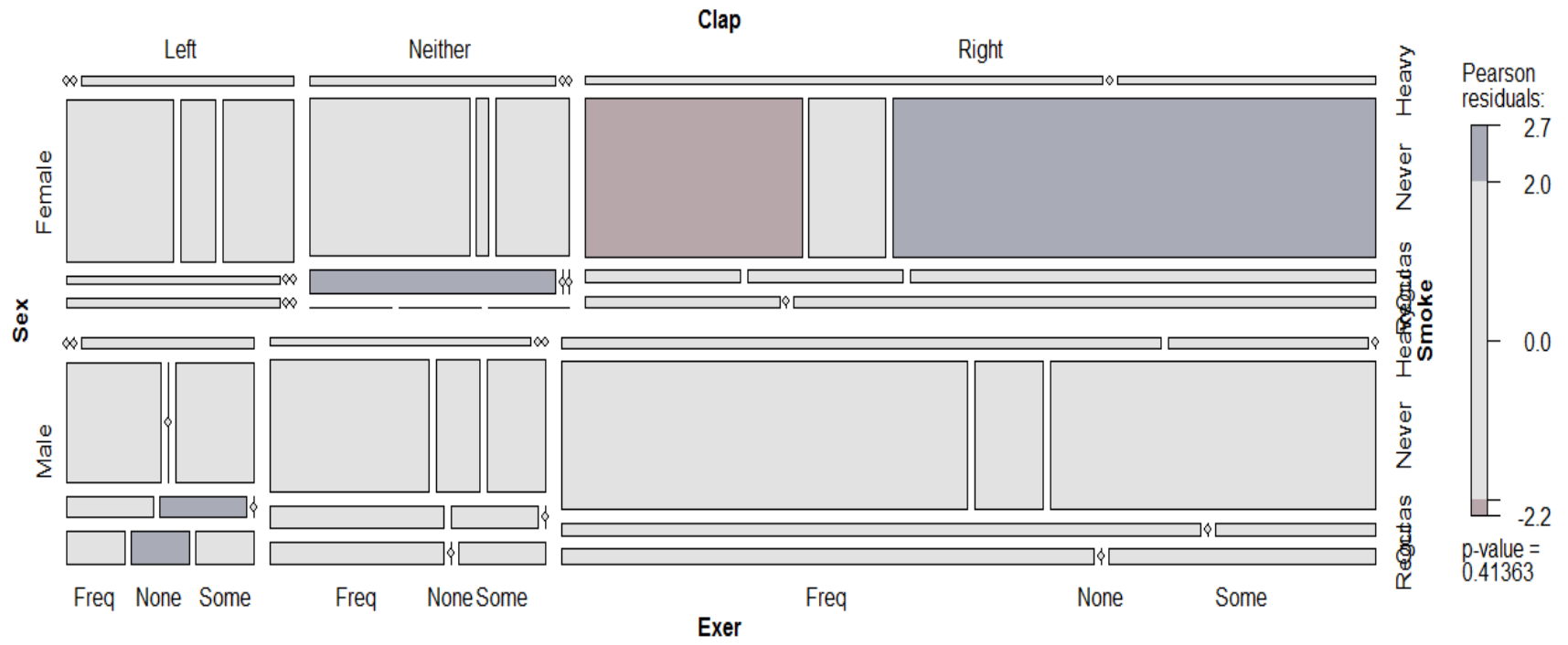**Figure 6.46** Mosaic plot.

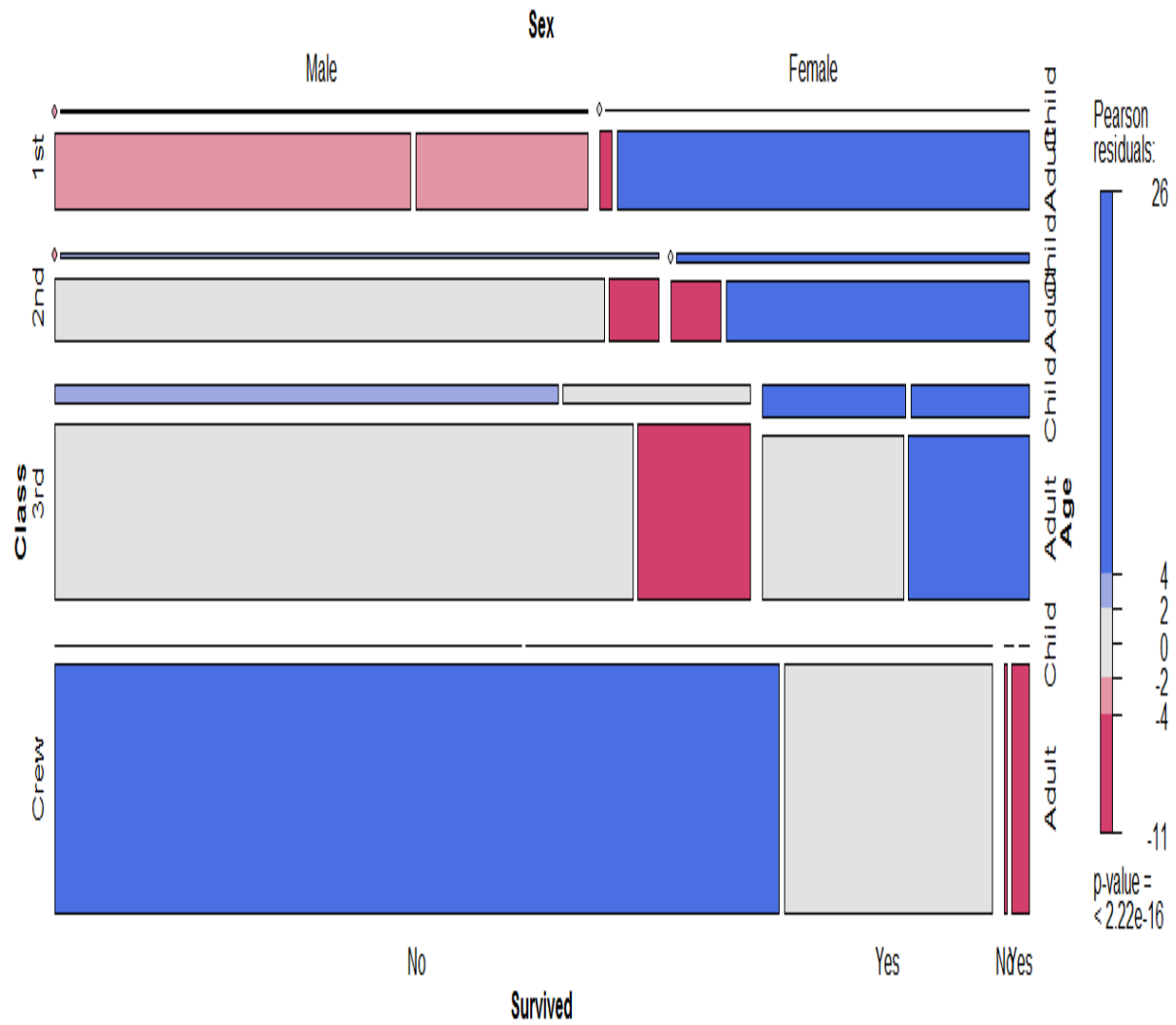**Figure 6.47** Mosaic plot for four categorical variables.

Figure 6.48 Mosaic plot showing frequency according to different categories of "Titanic" dataset.
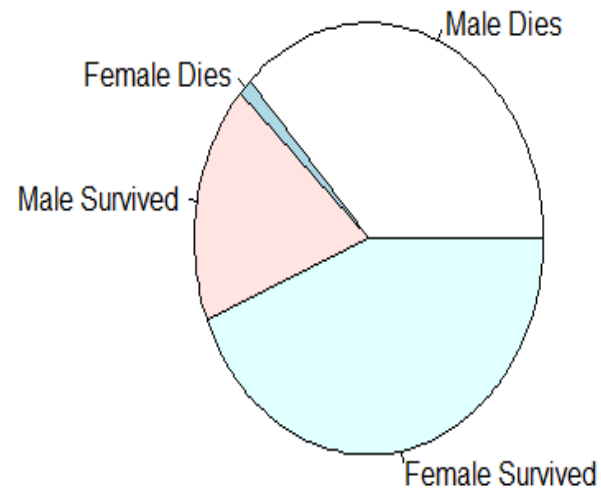
First Class

Figure 6.49 First class adults.
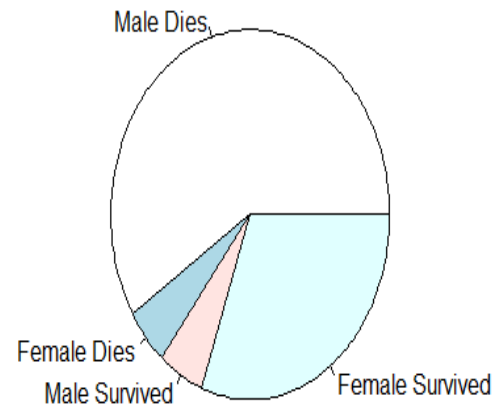
Figure 6.50 Second class adults.

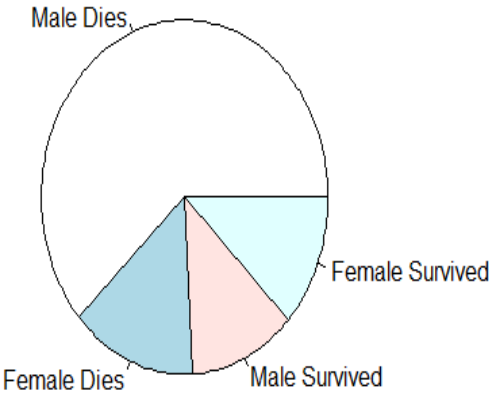**Figure 6.51** Third class adults.

In Crew
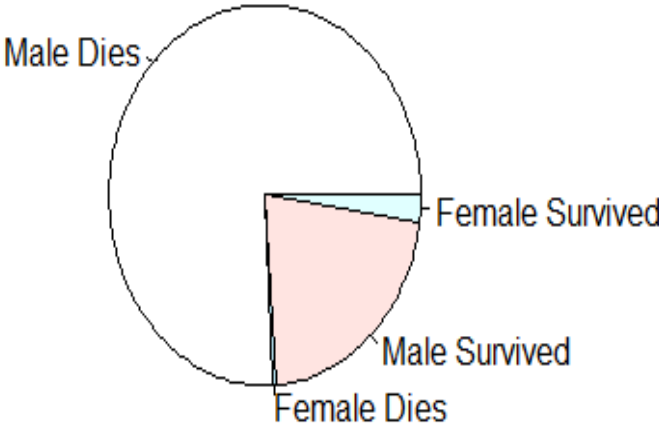
Figure 6.52 In crew.
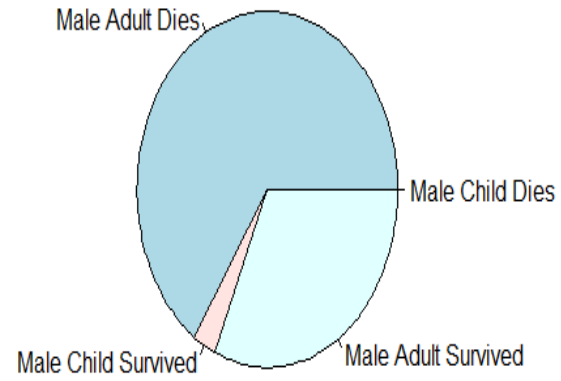
Figure 6.53 First class male.
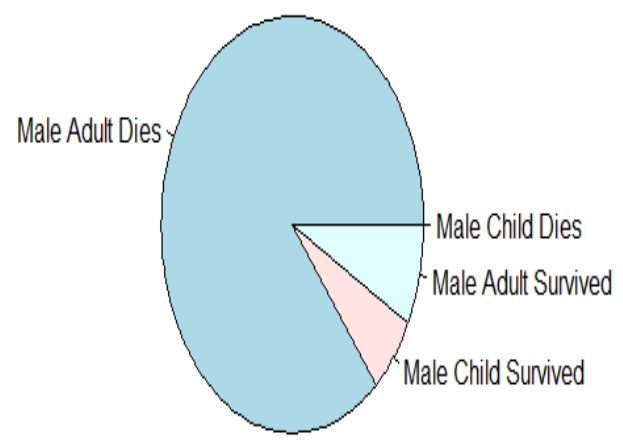
Figure 6.54 Second class male.

Figure 6.55 Female third class.
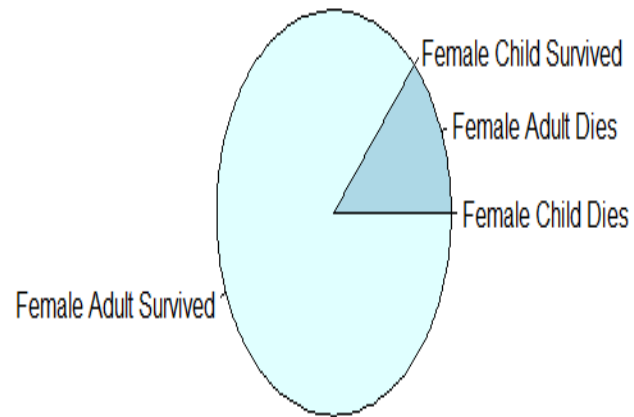
Figure 6.56 Female in crew.

Male Child

Third class Dies

First class Dies
Second class Dies
Third class Survived

Crew Dies

First class Survived

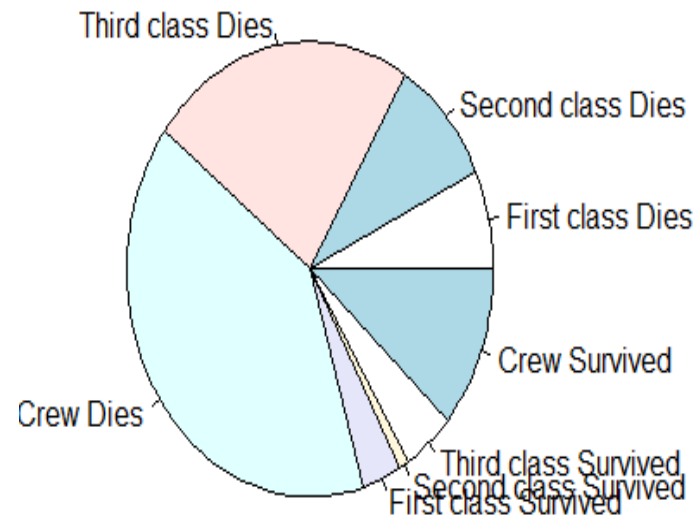Second class Survived
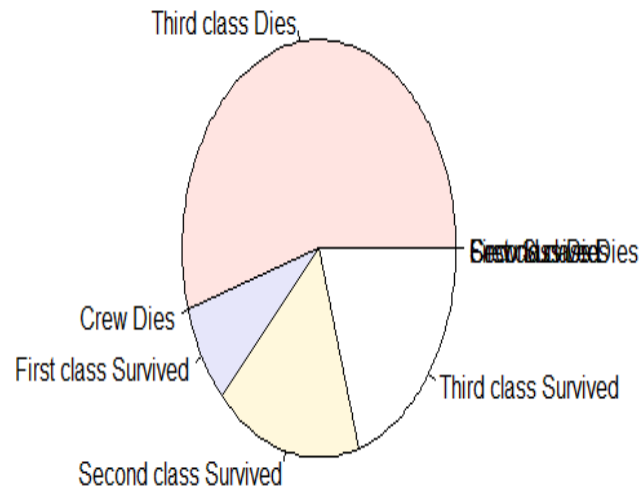
Figure 6.57 Male child.
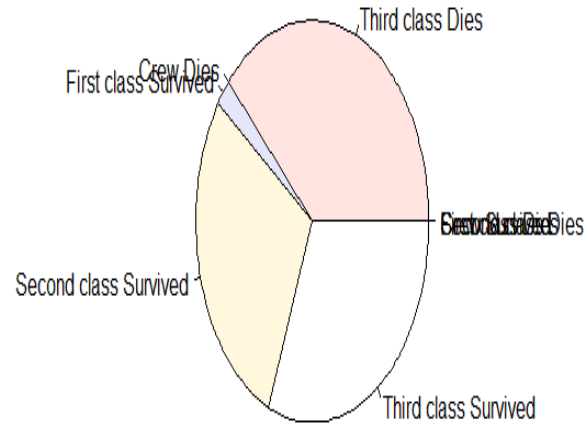
Figure 6.58 Male adult.

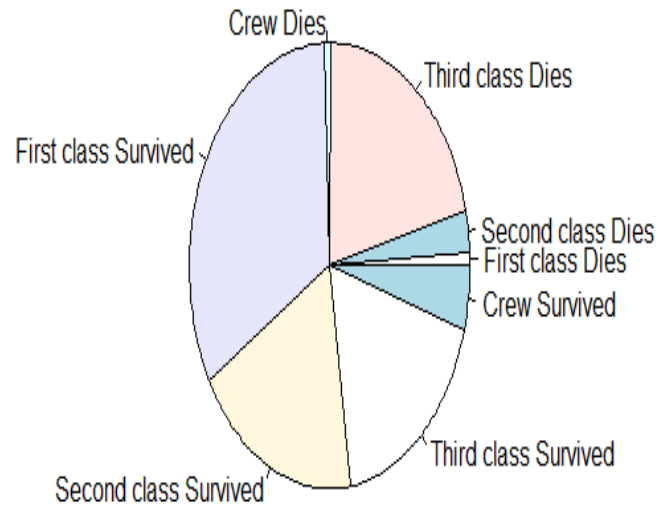**Female Child**

Figure 6.59 Female child.
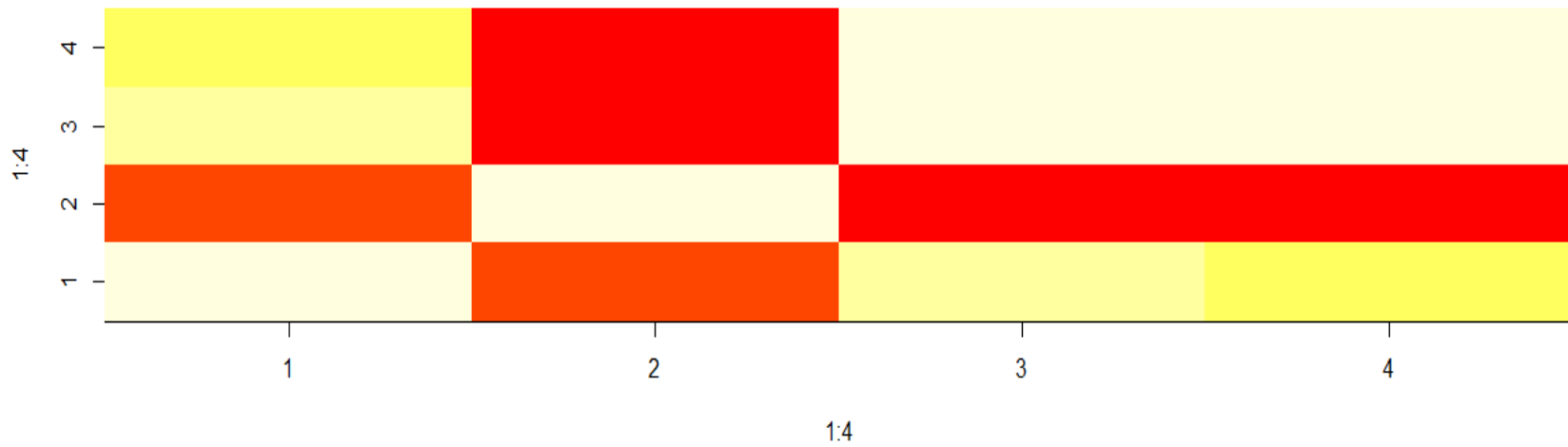
Figure 6.60 Female adult.

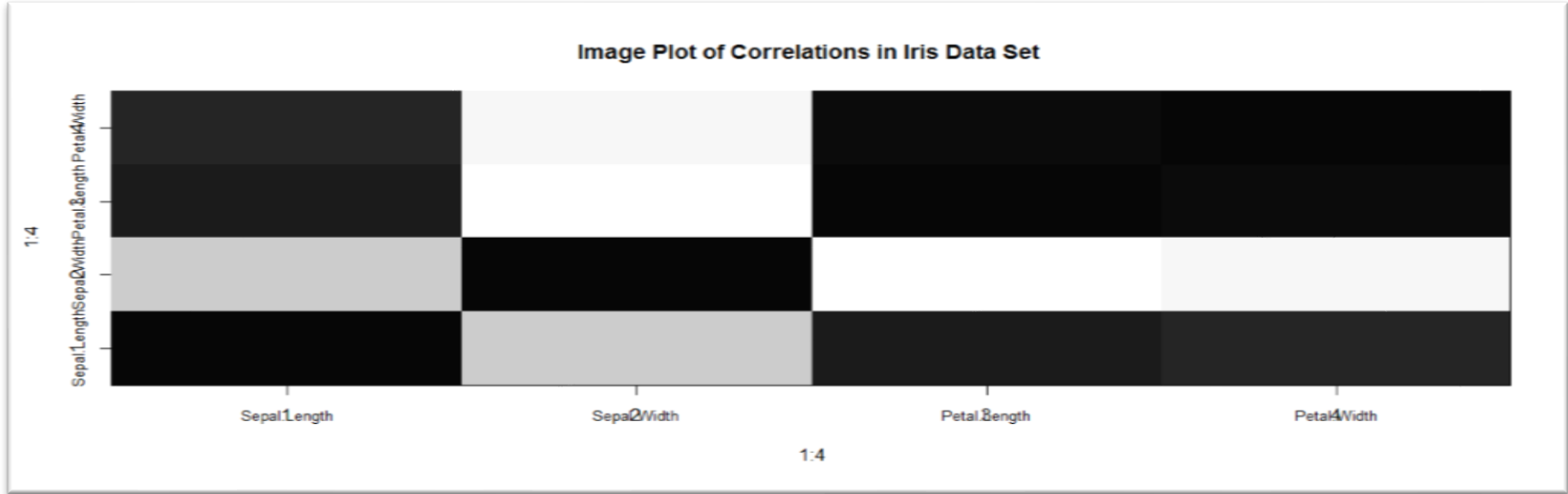**Figure 6.61** Image plot of correlations in "iris" using heat colors.

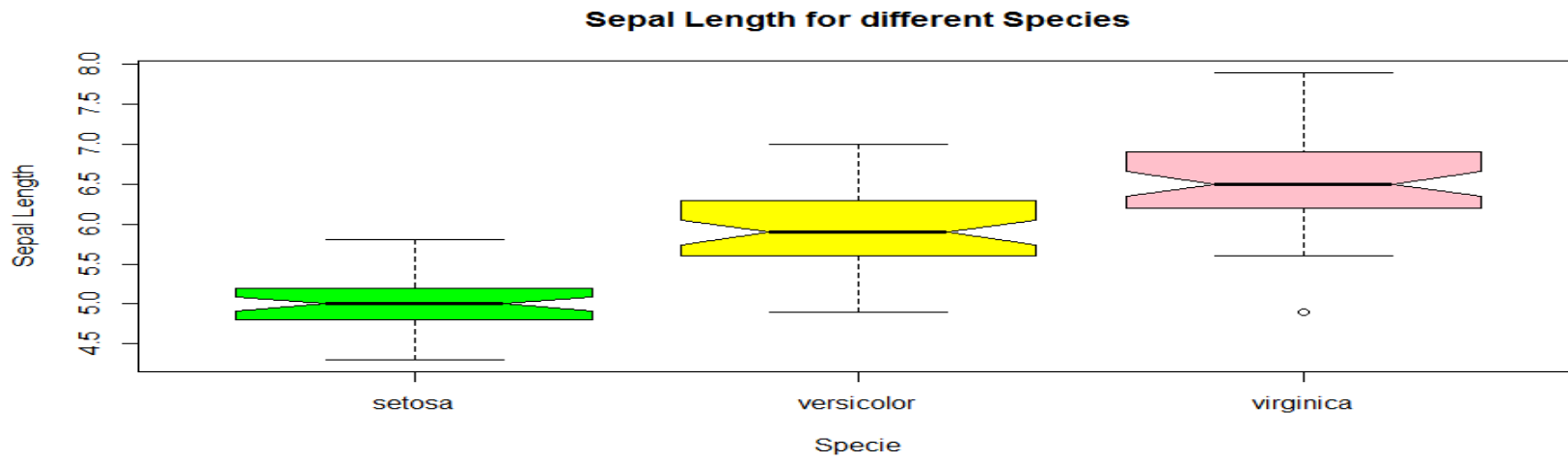**Figure 6.62** Image plot of correlations in "iris" using gray scale.

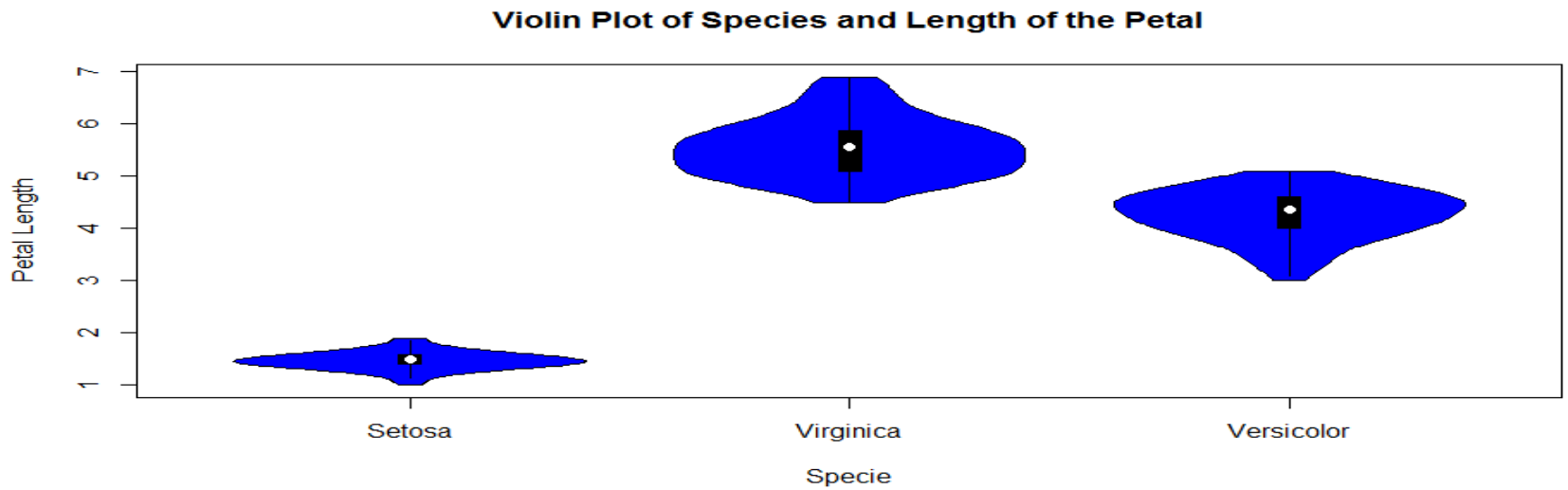**Figure 6.63** Box plot of "iris" dataset according to species.

Figure 6.64 Violin plot of "iris" dataset according to species.
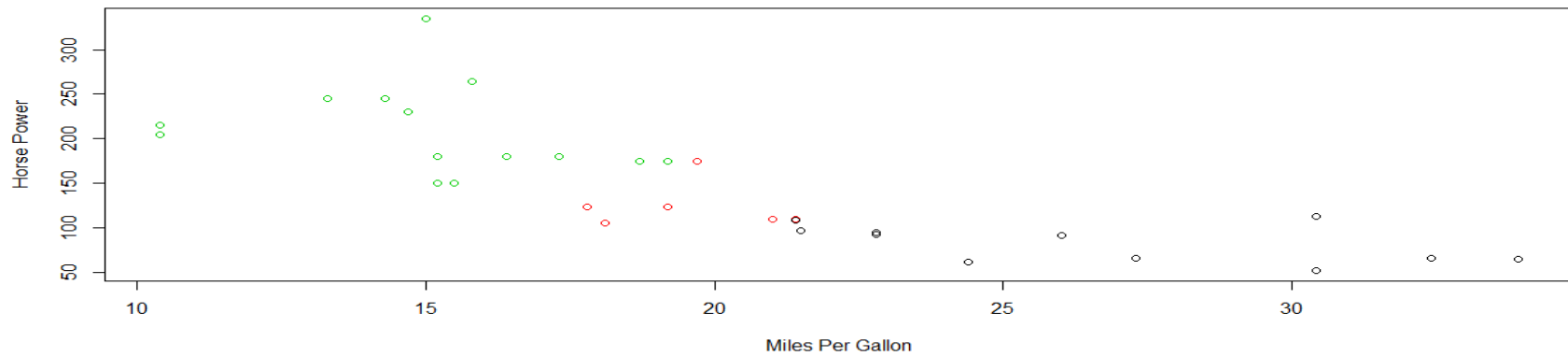
CHAPTER 7
Advanced Visualization

Figure 7.1 Chart for miles per gallon and horsepower representing number of cylinders in different colors.
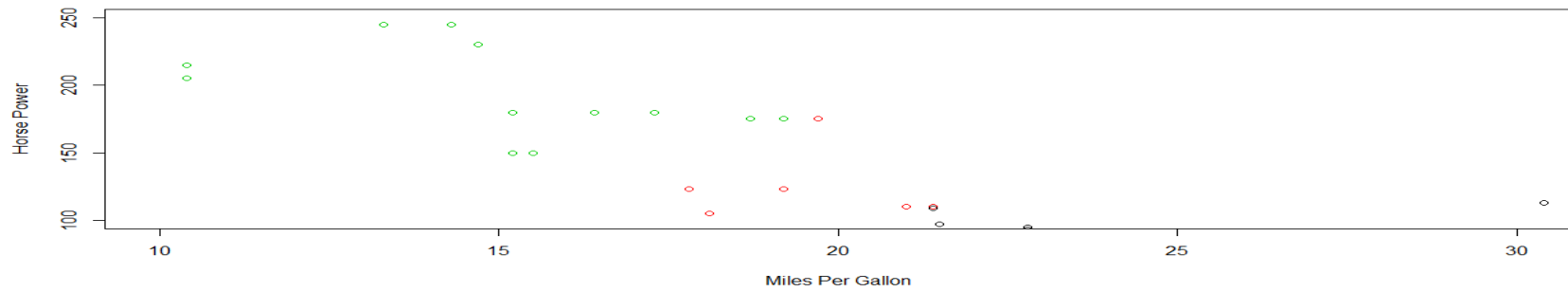
Figure 7.2 Chart for selected values of horsepower and miles per gallon representing number of cylinders in different colors.
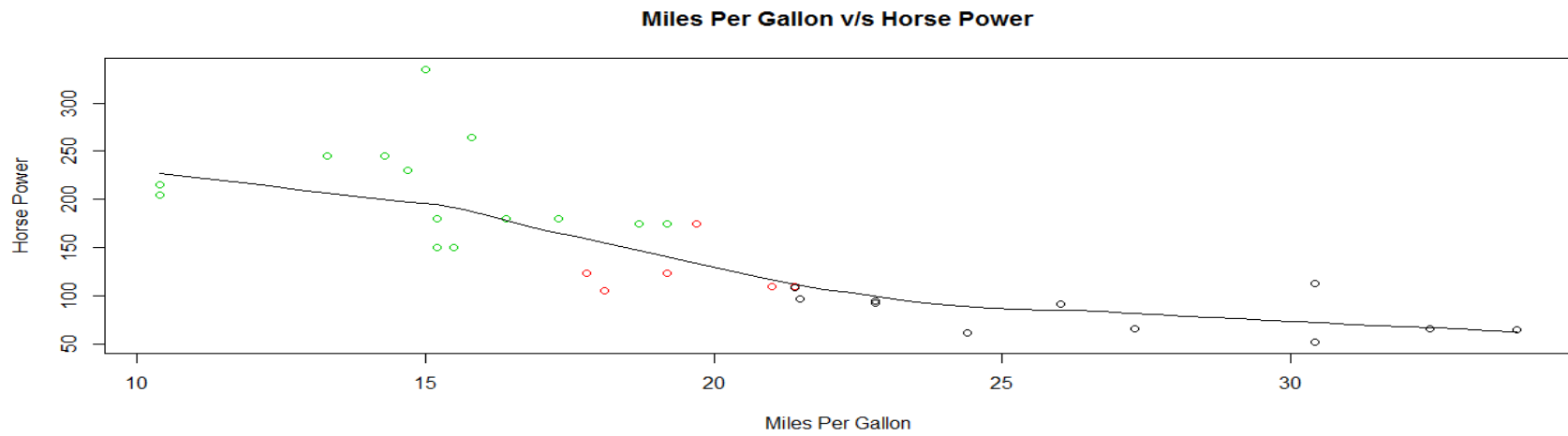
Figure 7.3 Using smooth curve on the scatter plot for "mtcars" dataset.
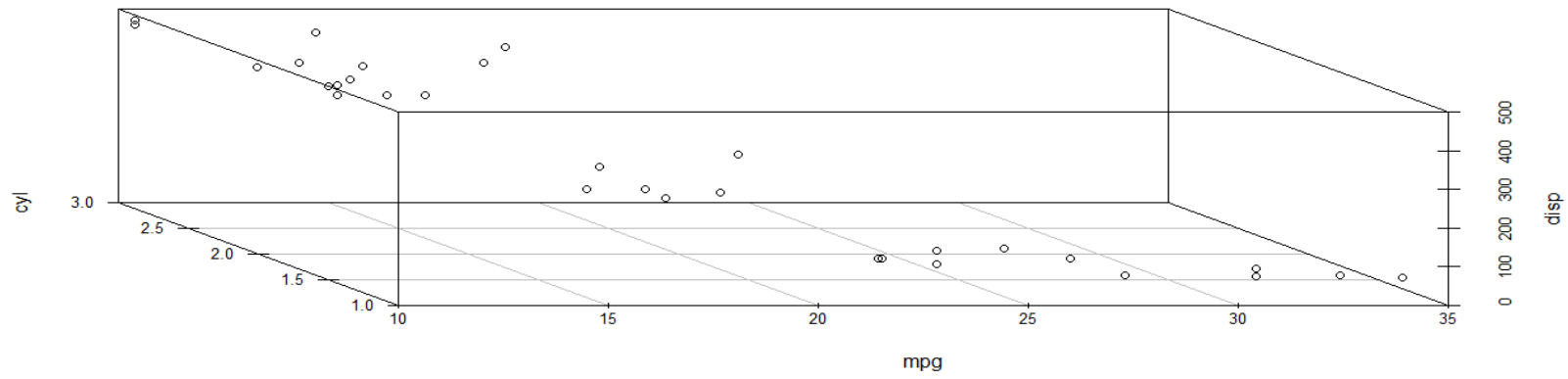
Figure 7.4 A 3D scatter plot chart for "mtcars" dataset.

Figure 7.5 A 3D scatter plot chart for "mtcars" dataset with a rotation of 120°.
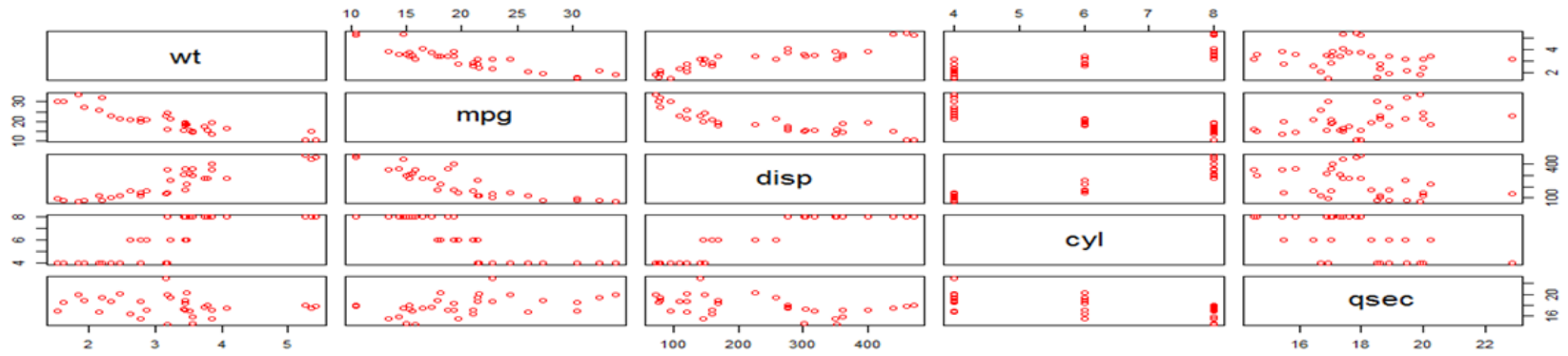
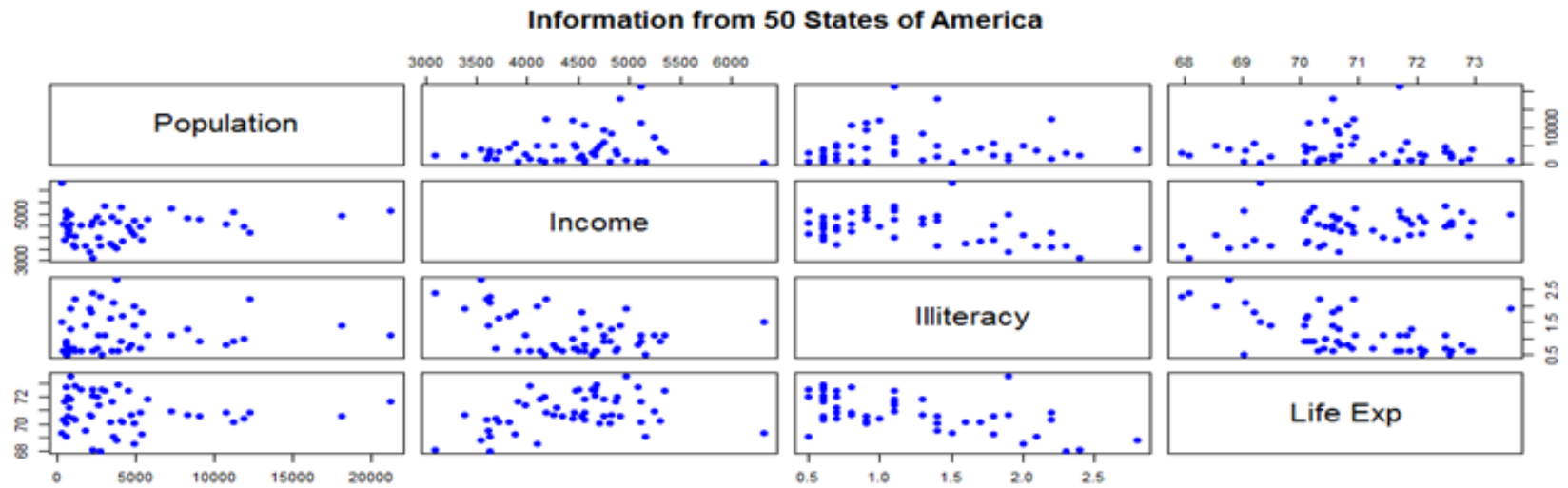**Figure 7.6** Scatter plot matrix for five variables of "mtcars" dataset.

Figure 7.7 Scatter plot matrix for four variables of "survey.x77" dataset.
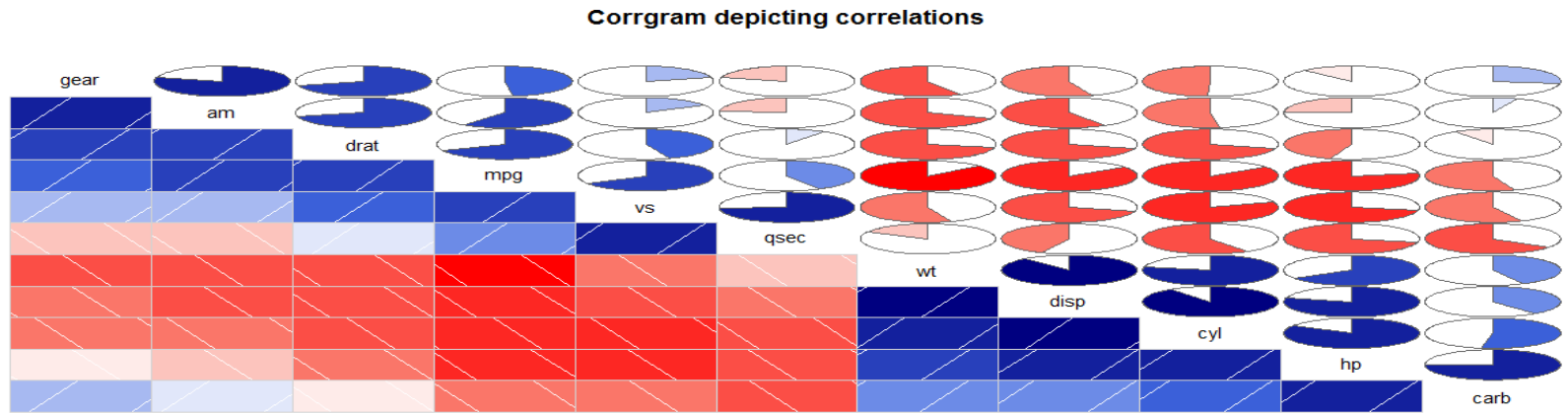
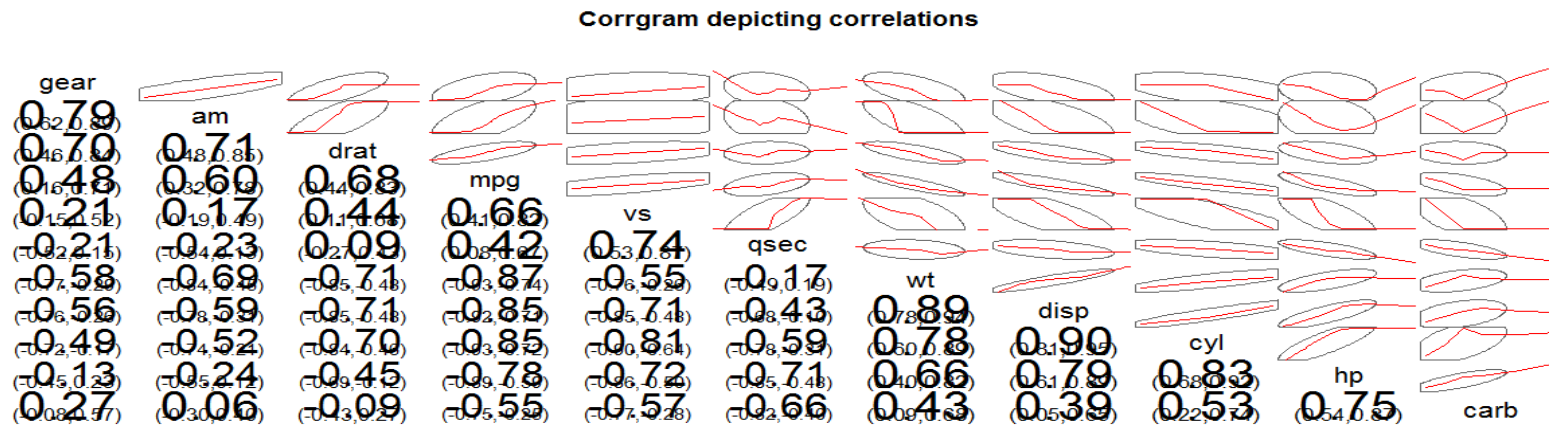**Figure 7.8** Displaying corrgram with `panel.shade` option in lower panel and `panel.pie` in upper panel.

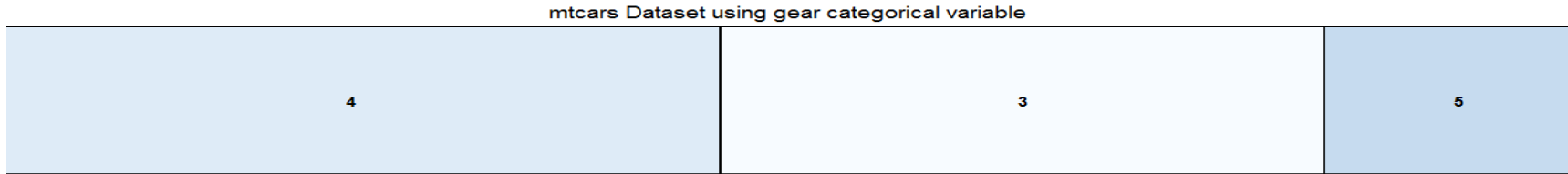**Figure 7.9** Displaying corrgram with `panel.conf` option in lower panel and `panel.ellipse` in upper panel.

Figure 7.11 Segment plot of "mtcars" dataset.

mtcars Dataset using gear categorical variable

| 4 | 3 | 5 |

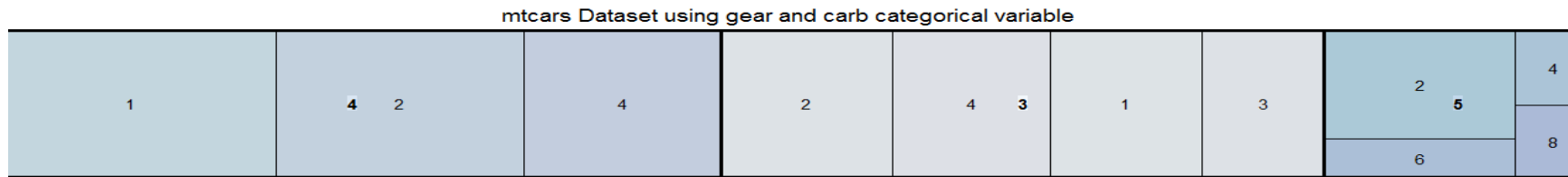**Figure 7.12** Tree map considering one categorical variable.

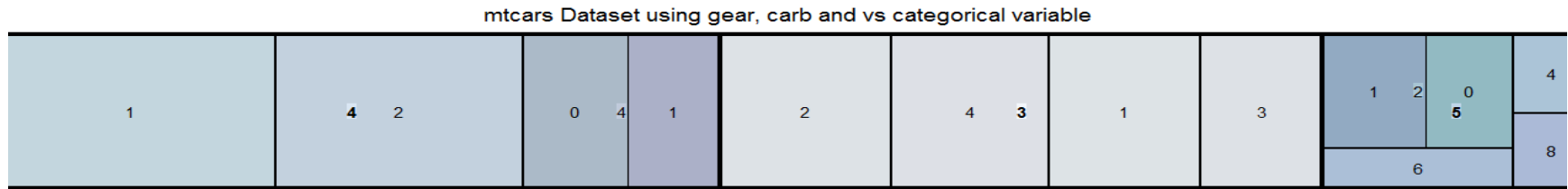**Figure 7.13** Tree map considering two categorical variables.

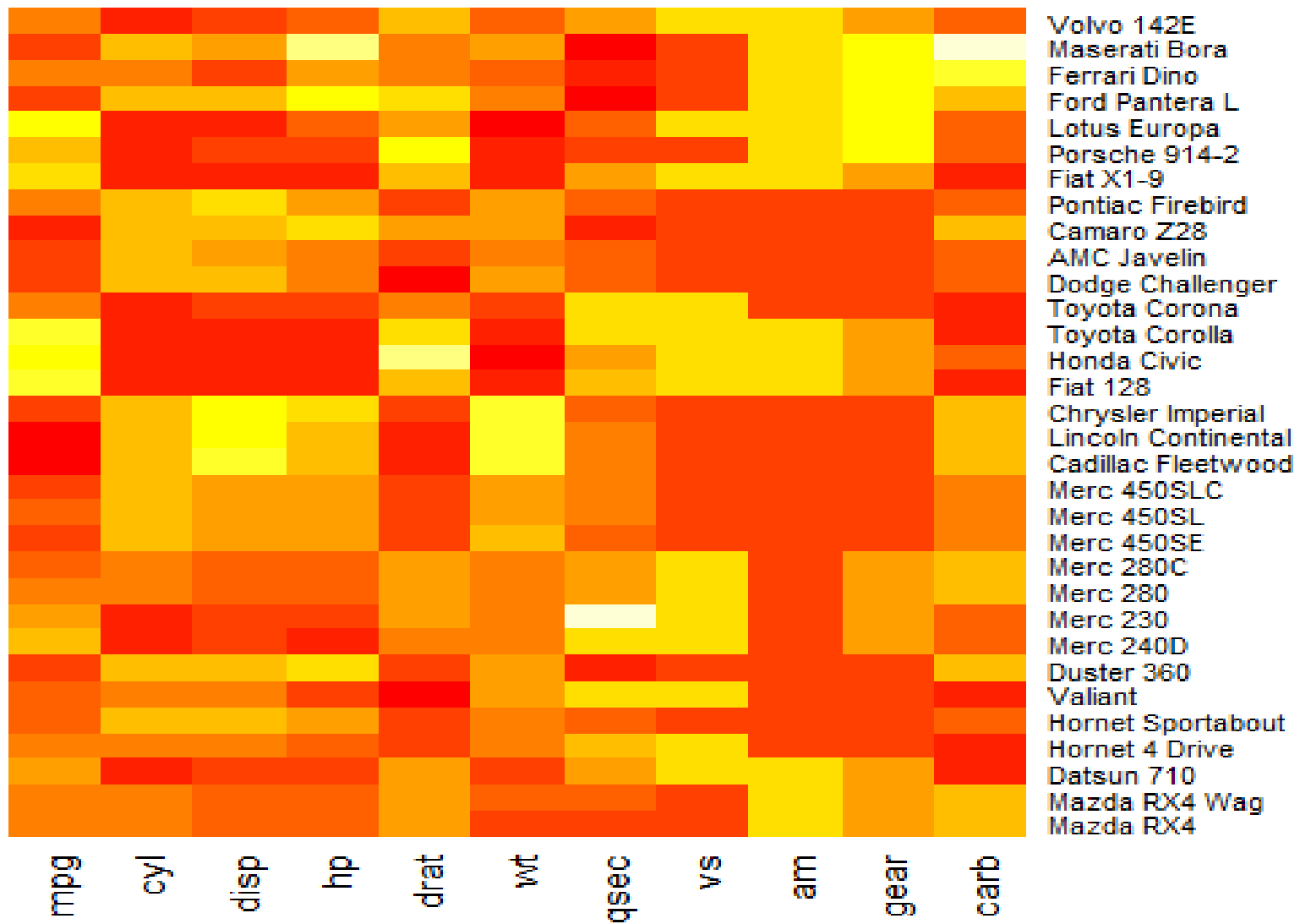**Figure 7.14** Tree map considering three categorical variables.
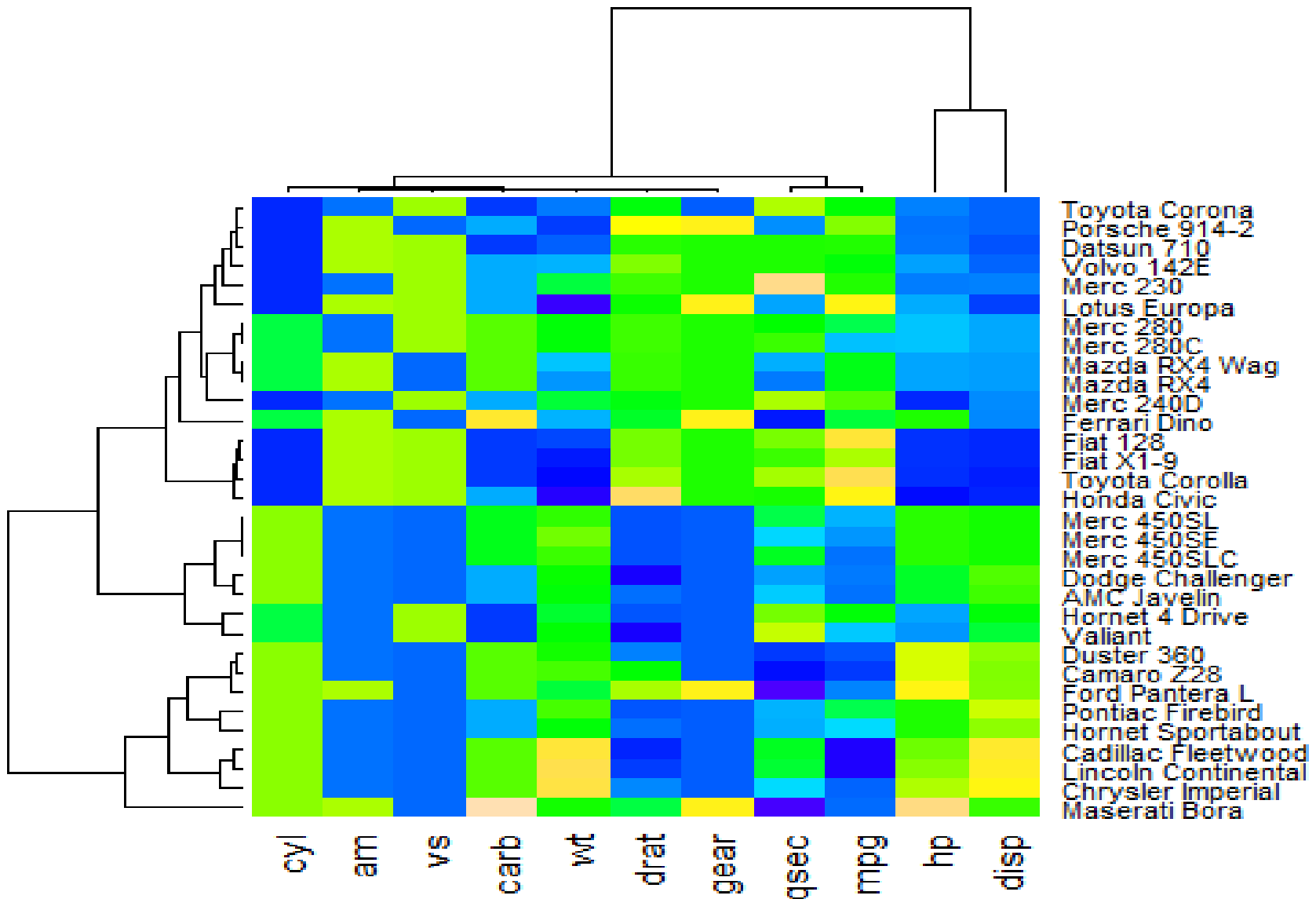
Figure 7.15 Heat map with no dendrogram.
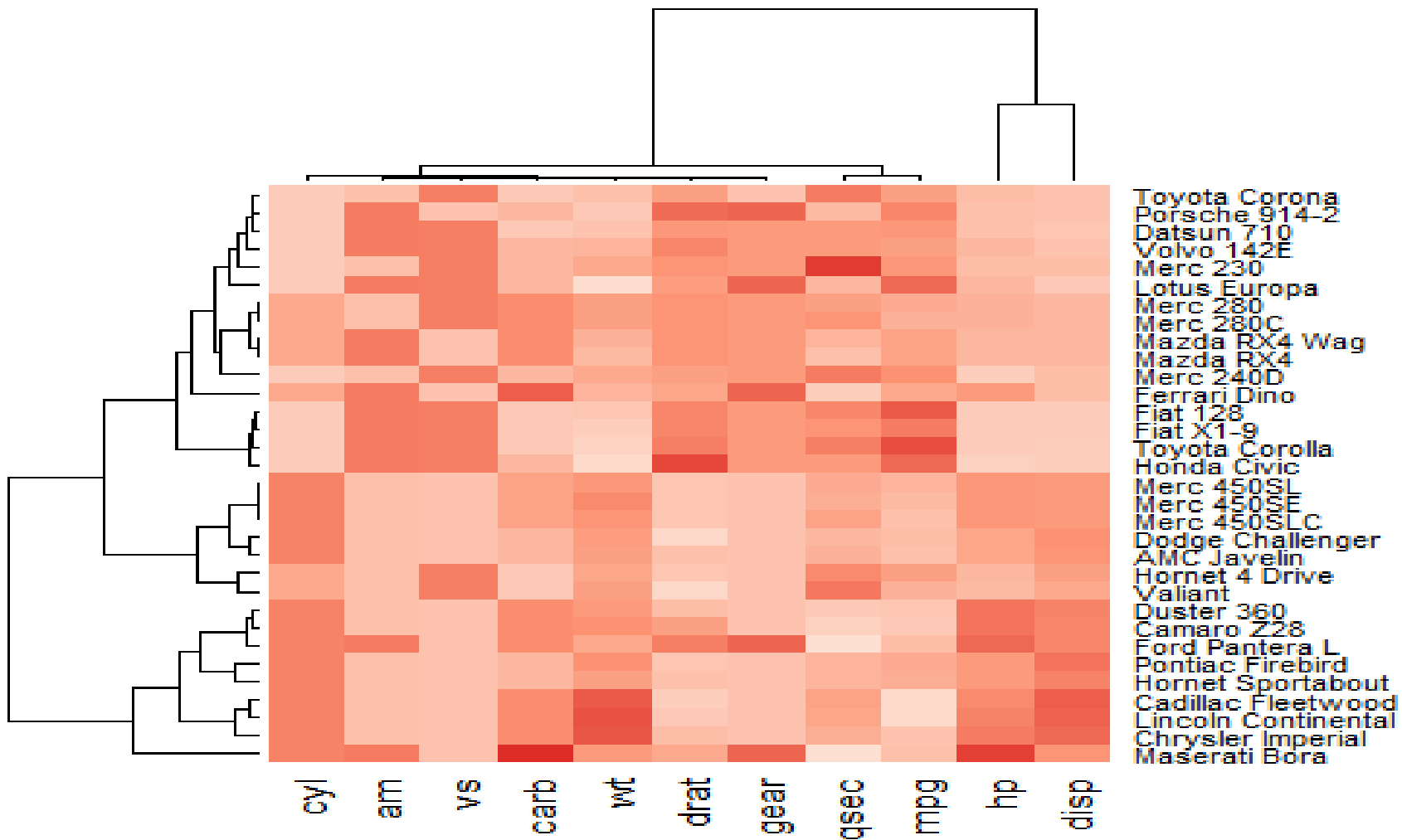
Figure 7.16 Heatmap with dendrogram for clustering.

WILEY

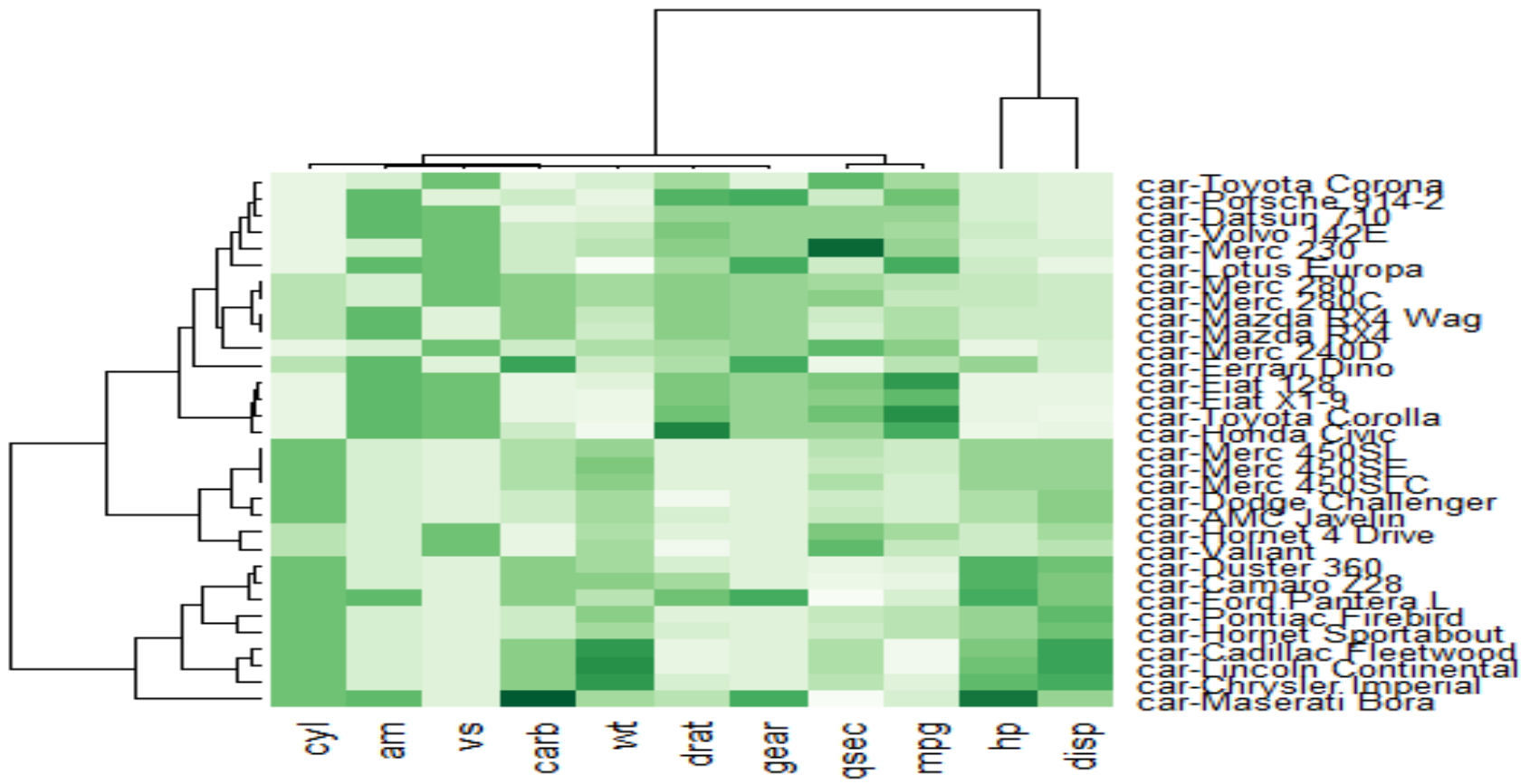**Figure 7.17** Colors from "RcolorBrewer" package.

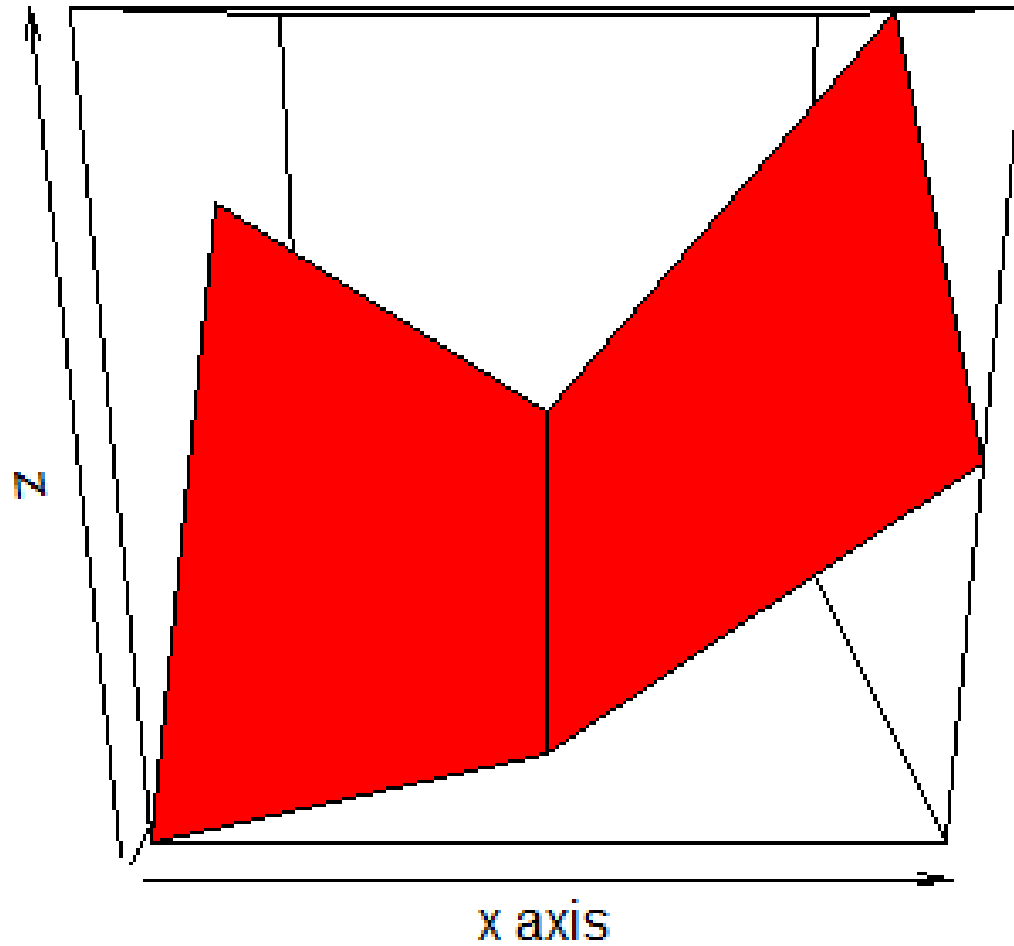**Figure 7.18** Changing label of rows and colors from "RcolorBrewer".

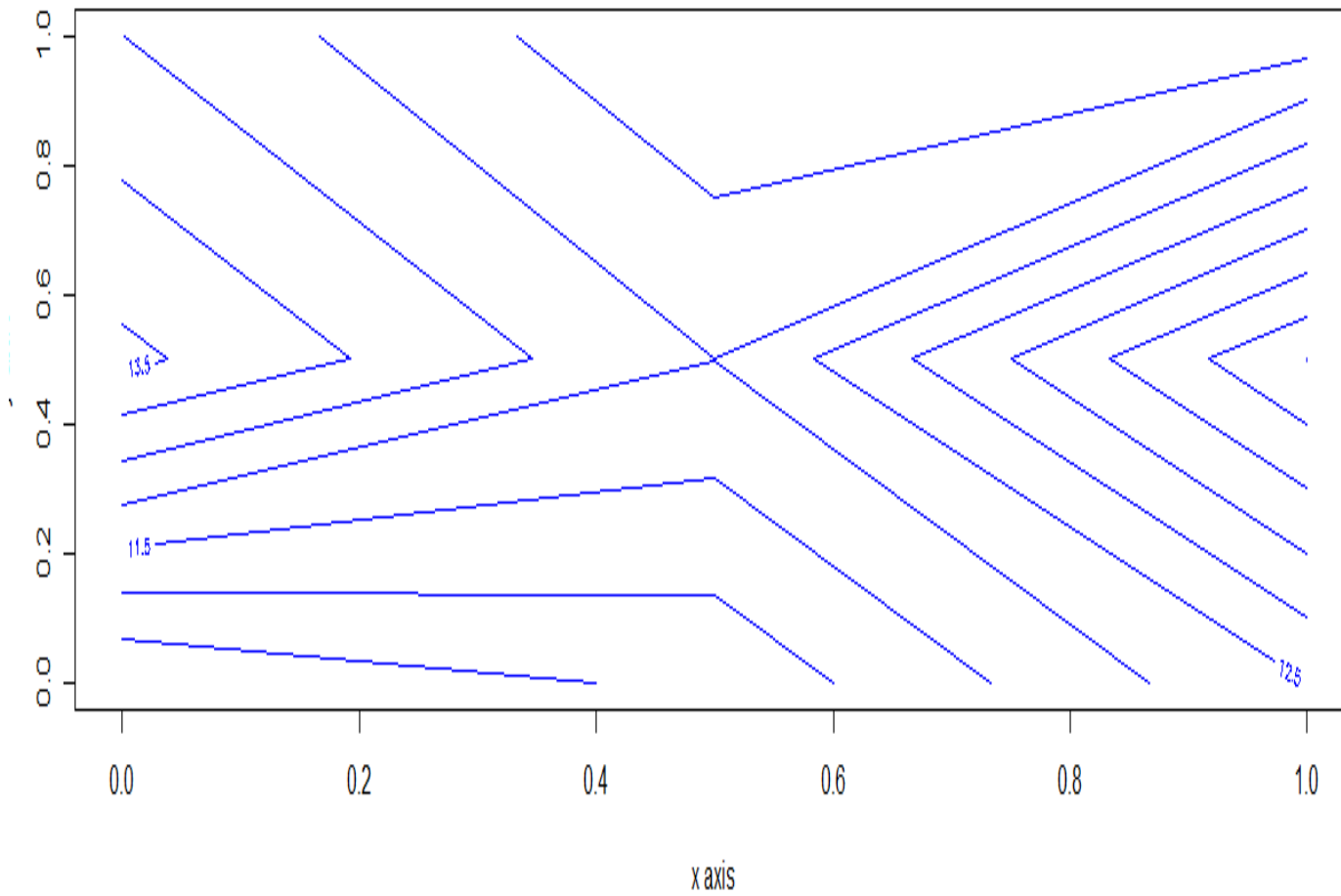**Figure 7.19** Perspective plot of a matrix.
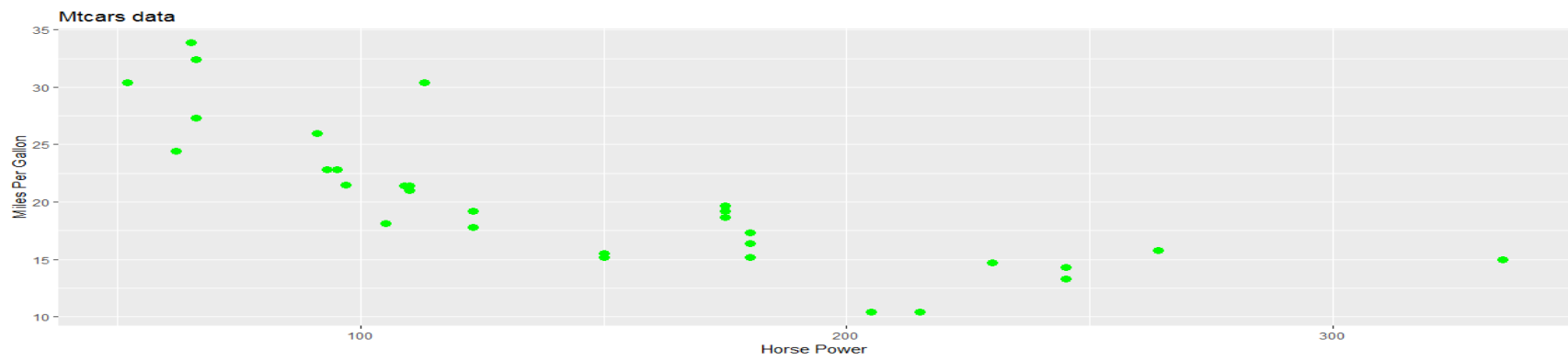
Figure 7.20 Contour plot for a matrix.

Figure 7.21 Creating a chart using `geom_point()` function.
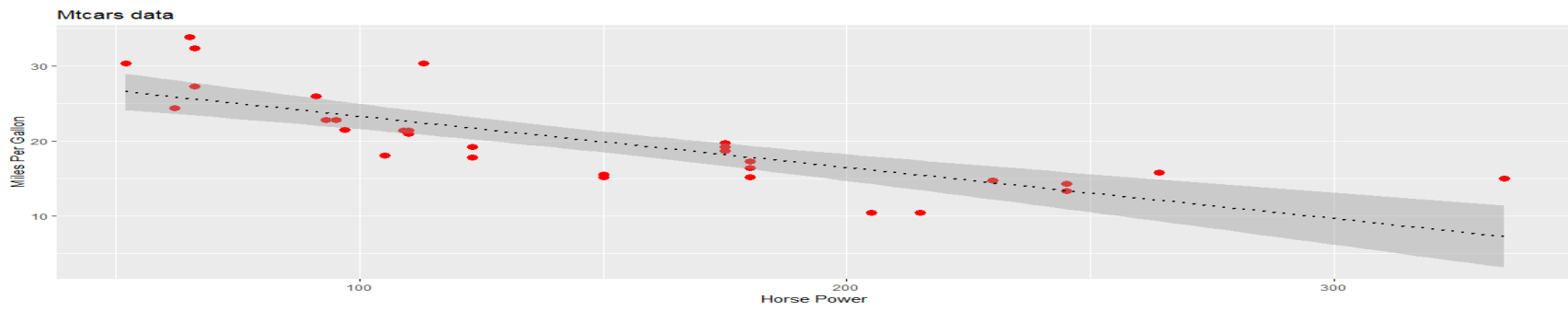
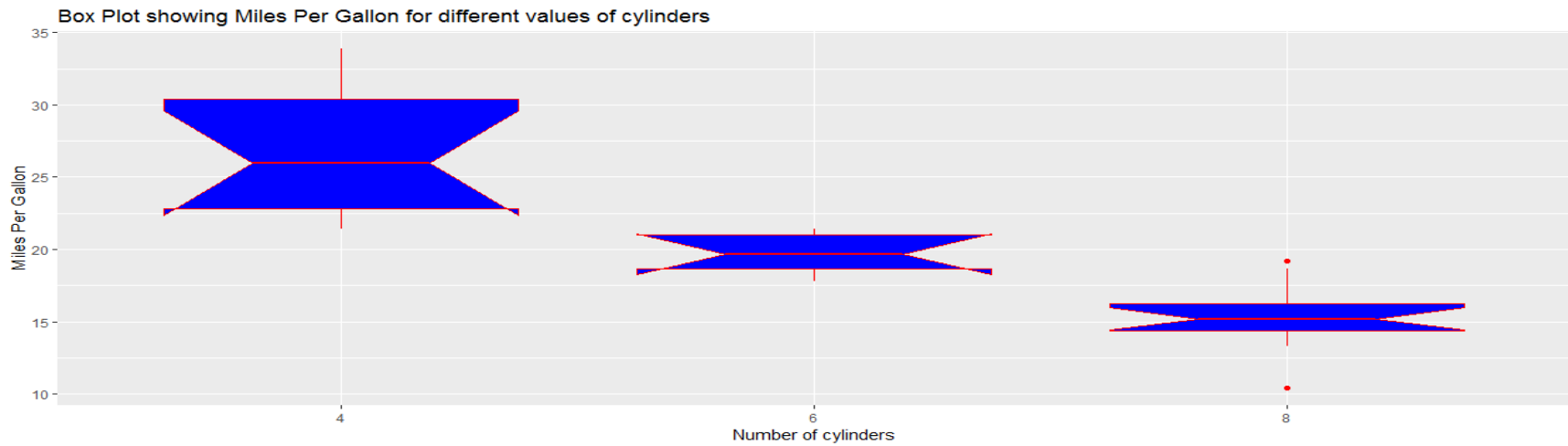**Figure 7.22** Creating a smooth line using `geom_smooth()` function.

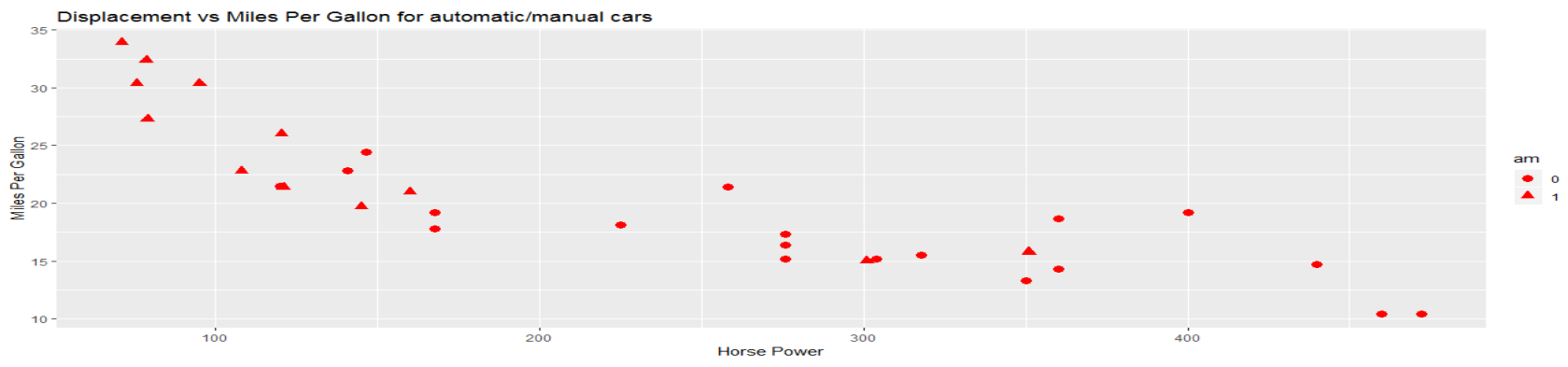**Figure 7.24** Creating a boxplot for miles per gallon on the basis of different cylinders.

Figure 7.25 Chart considering mode of transmission (am) as additional variable.
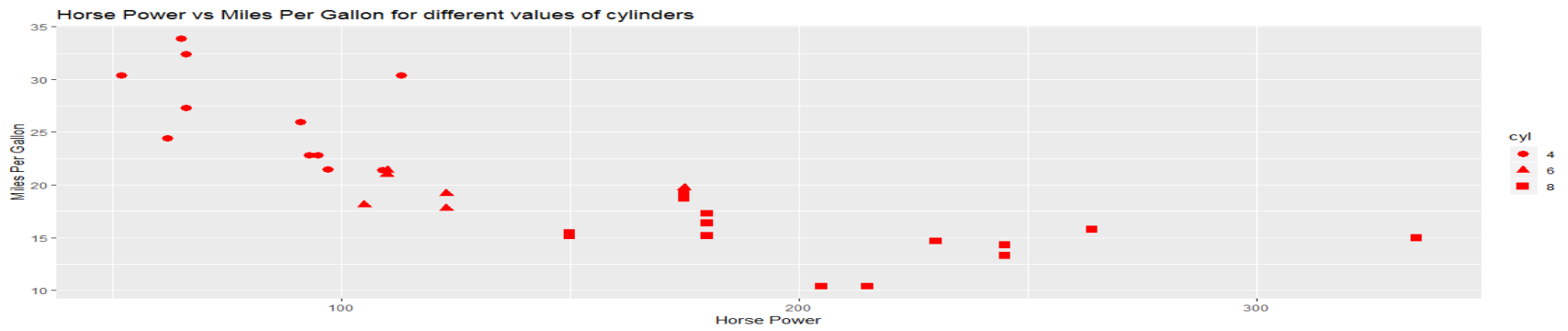
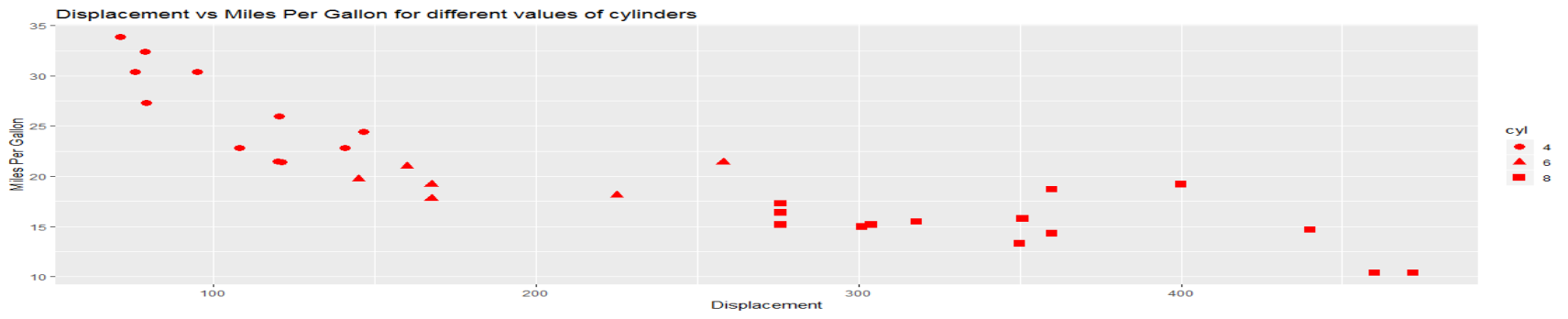**Figure 7.26** Chart considering number of cylinders (cyl) as additional variable.

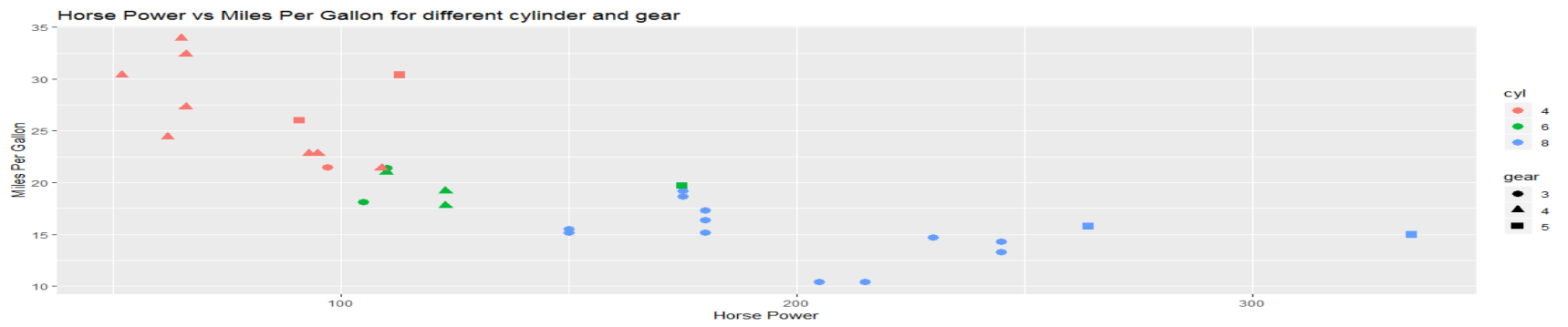**Figure 7.27** Chart considering displacement on *x*-axis.

Figure 7.28 Chart considering two additional variables (number of cylinders and gear).
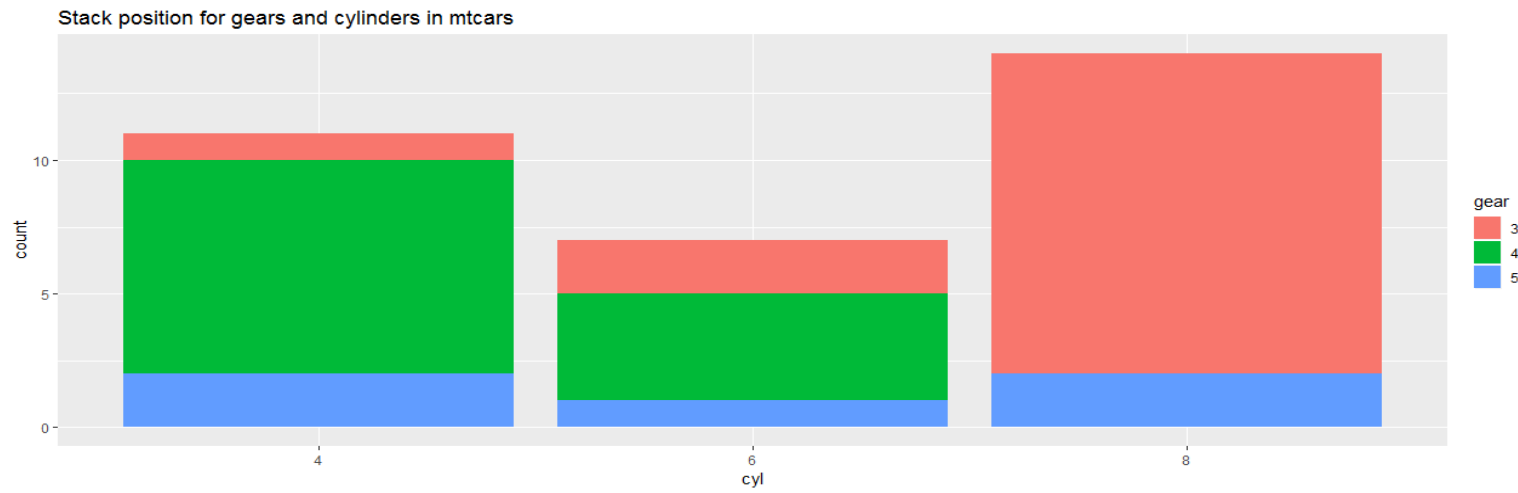
Figure 7.29 Bar plot for stack position representing number of gears and cylinders.
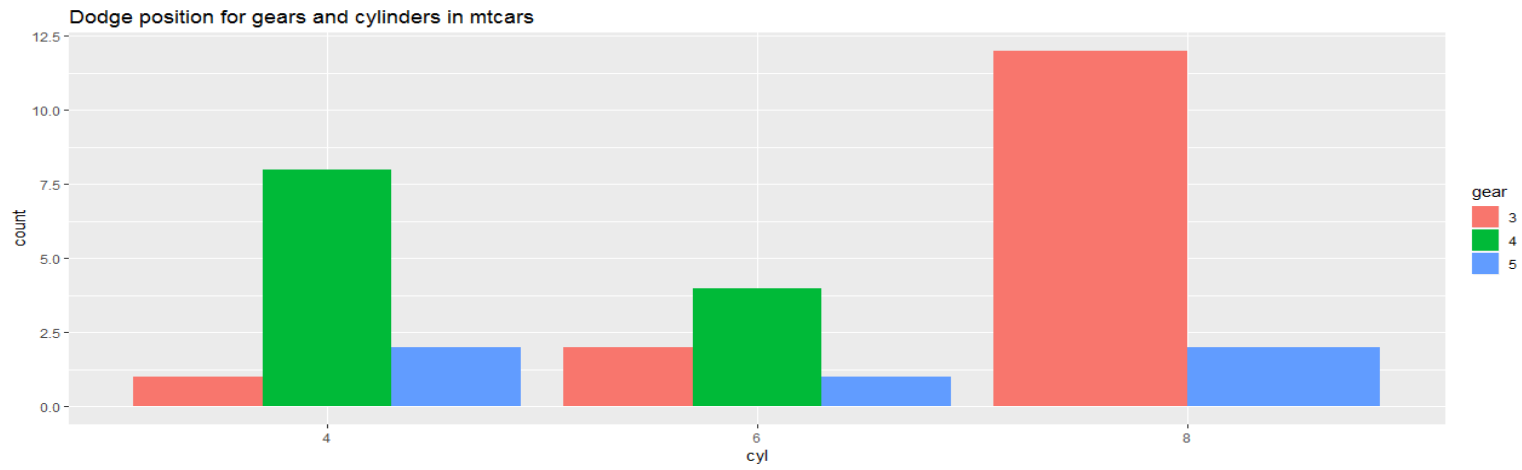
Figure 7.30 Bar plot for dodge position representing number of gears and cylinders.
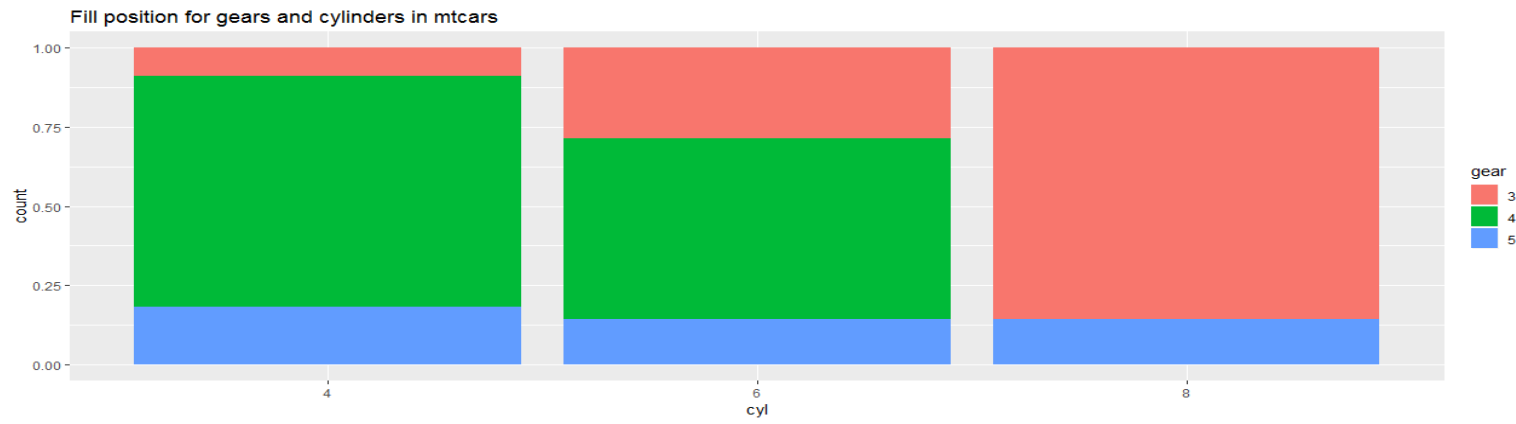
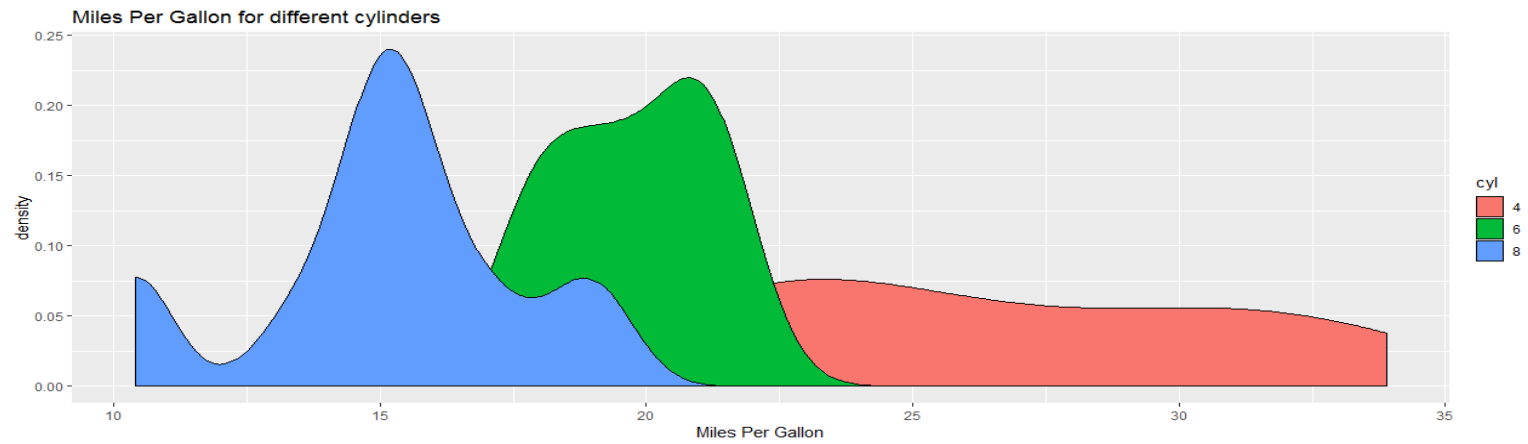**Figure 7.31** Bar plot for fill position representing number of gears and cylinders.

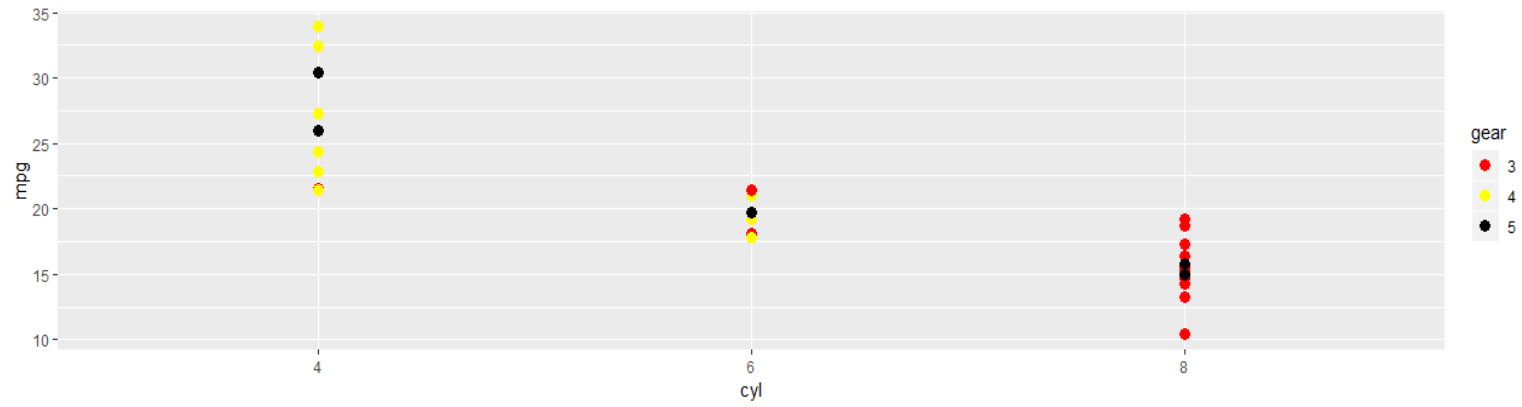**Figure 7.32** Density plot showing miles per gallon according to number of cylinders.

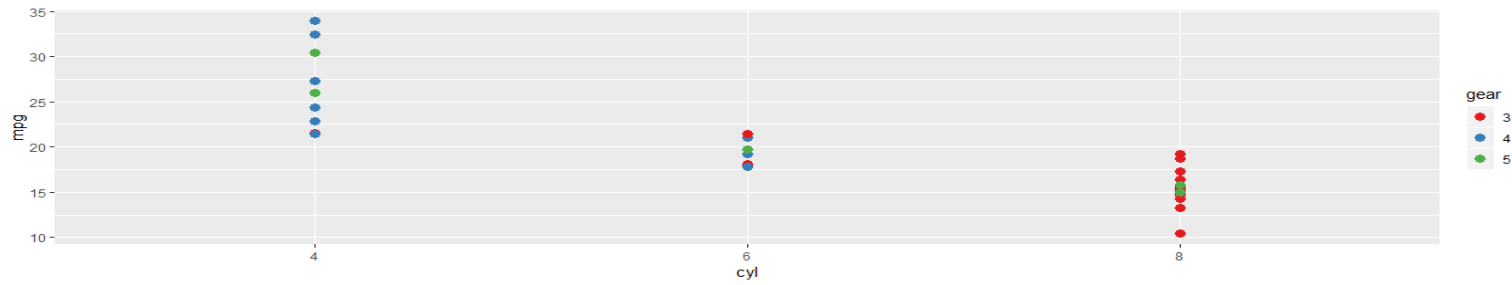**gure 7.33** Customizing colors in the chart through user-defined colors.

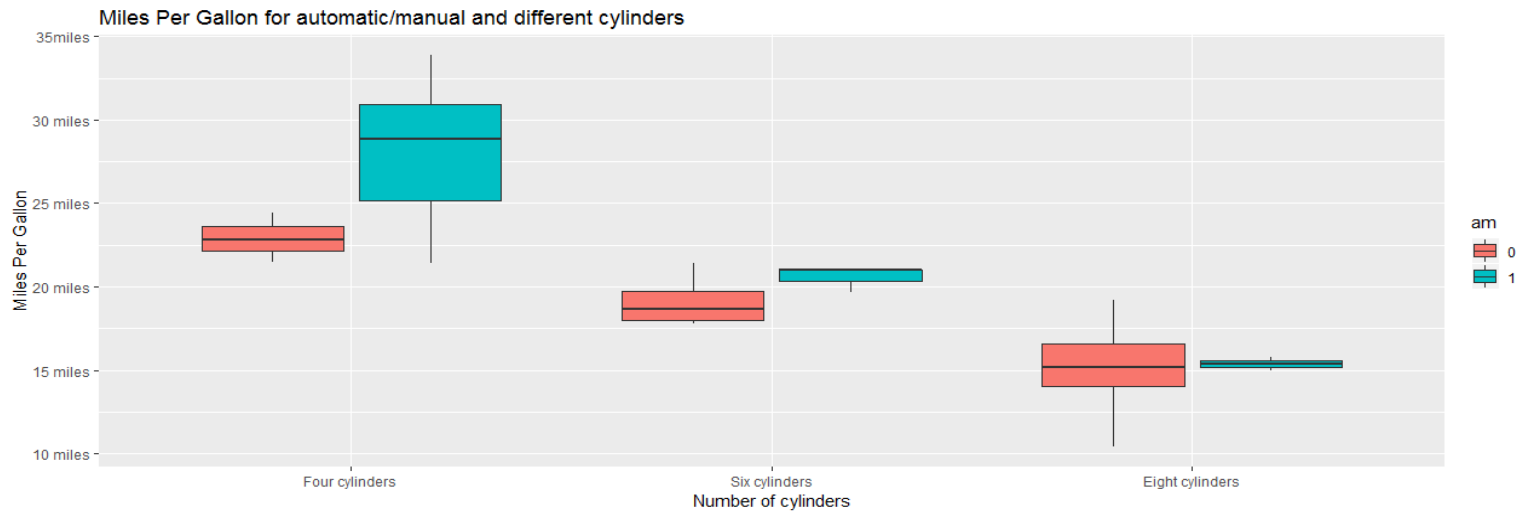**Figure 7.34** Customizing colors in the chart through user-defined colors.

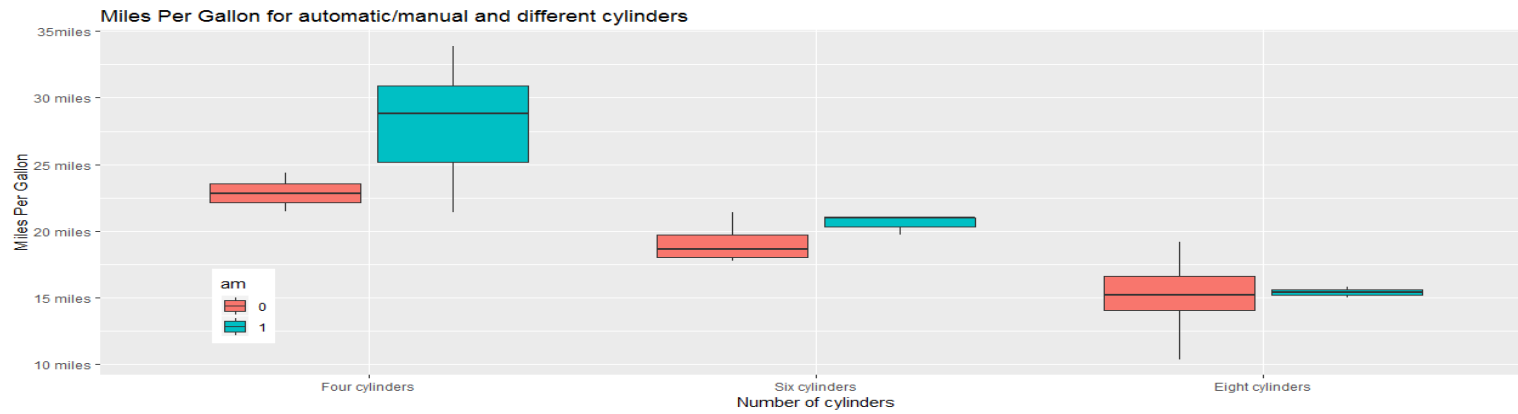**Figure 7.35** Adding scale in *x*-axis and *y*-axis.

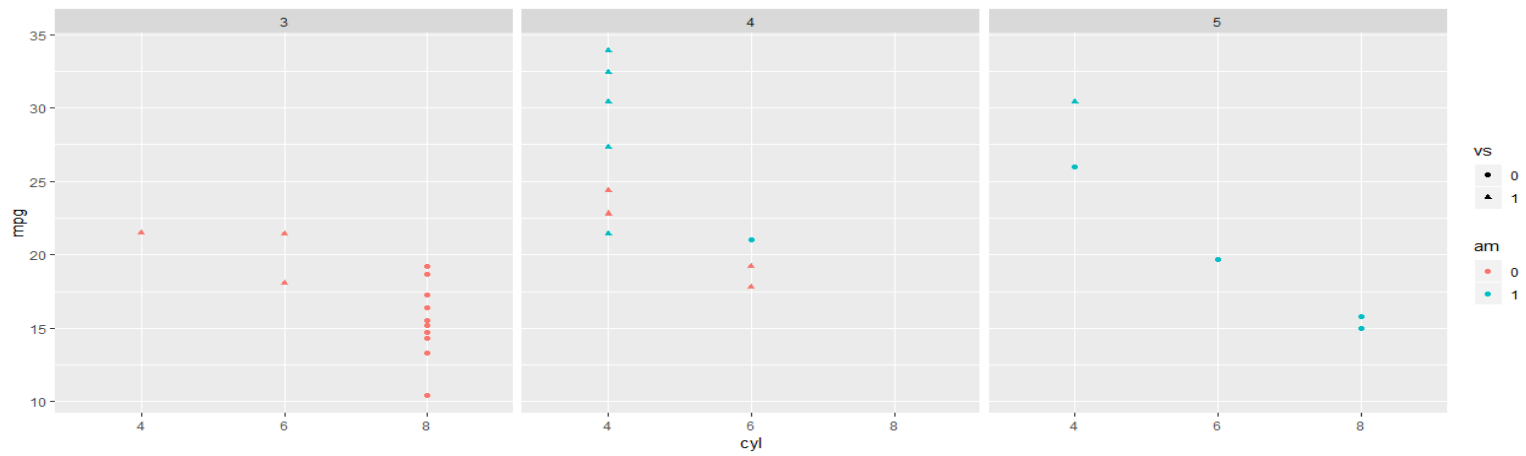**Figure 7.36** Displaying legend at specified location.

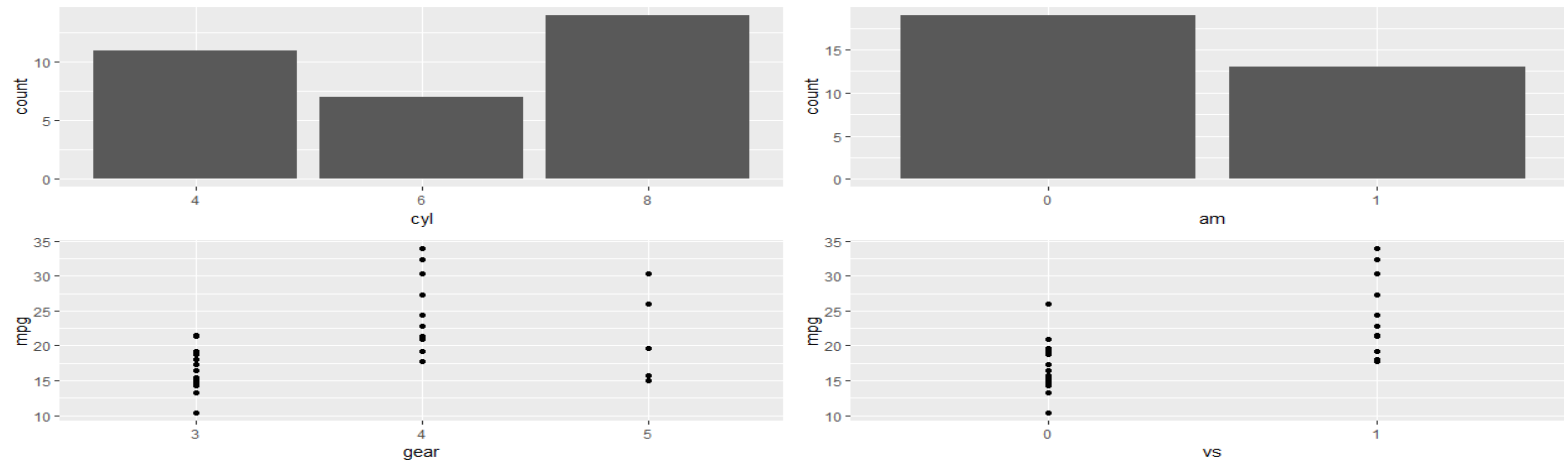**Figure 7.37** Create a plot with `facet_grid()` function.

Figure 7.38 Drawing multiple graphs in one chart.

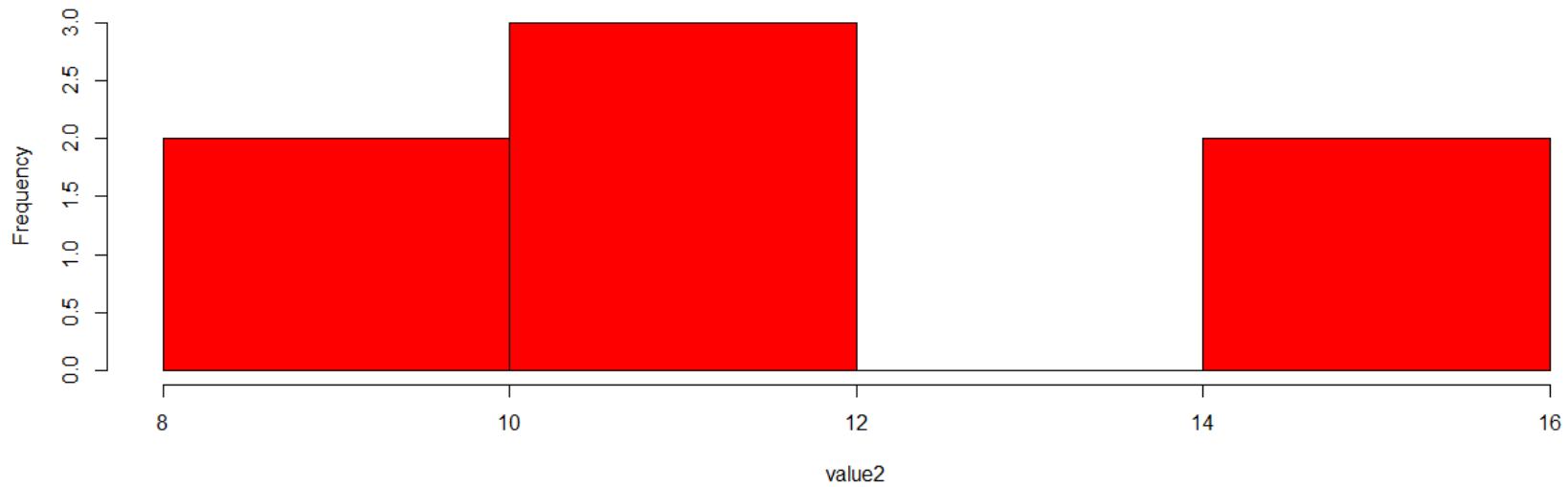# CHAPTER 8
Basic Statistics

Figure 8.4 Histogram depicting values generated randomly following a normal distribution.
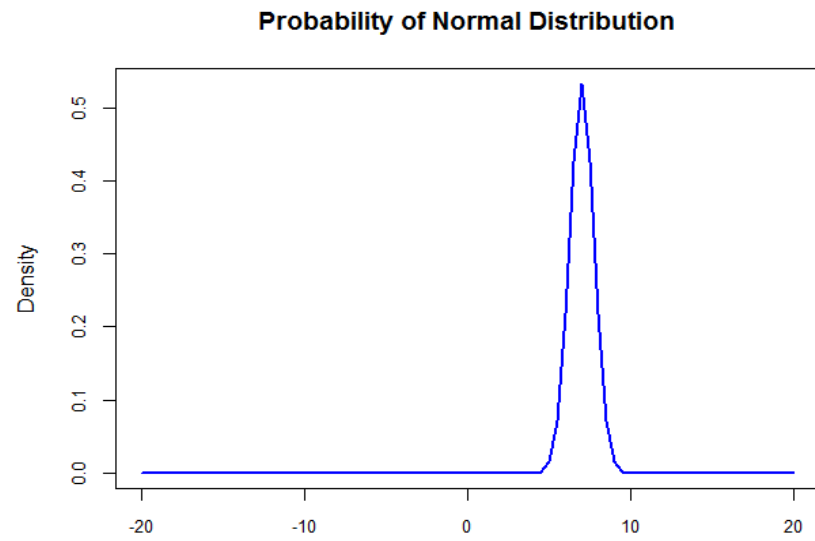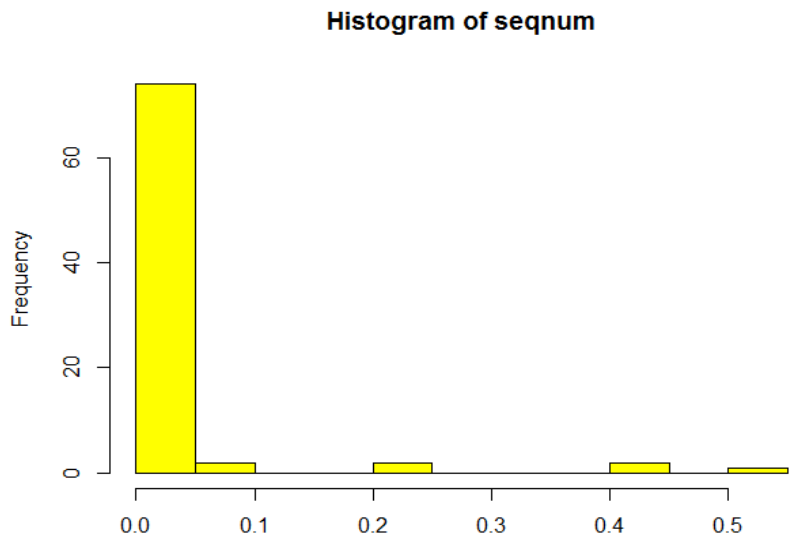
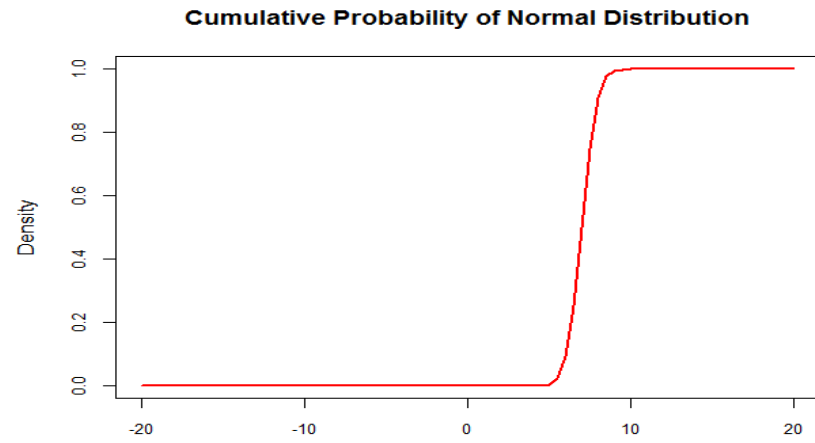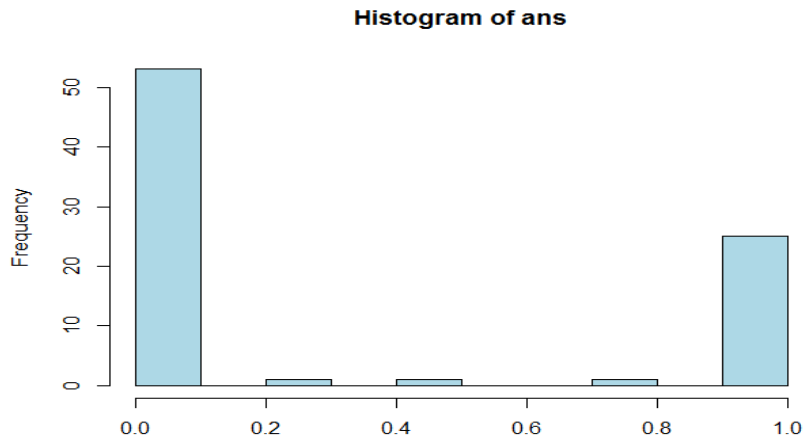**Figure 8.5** Probability density plot for a created sequence.

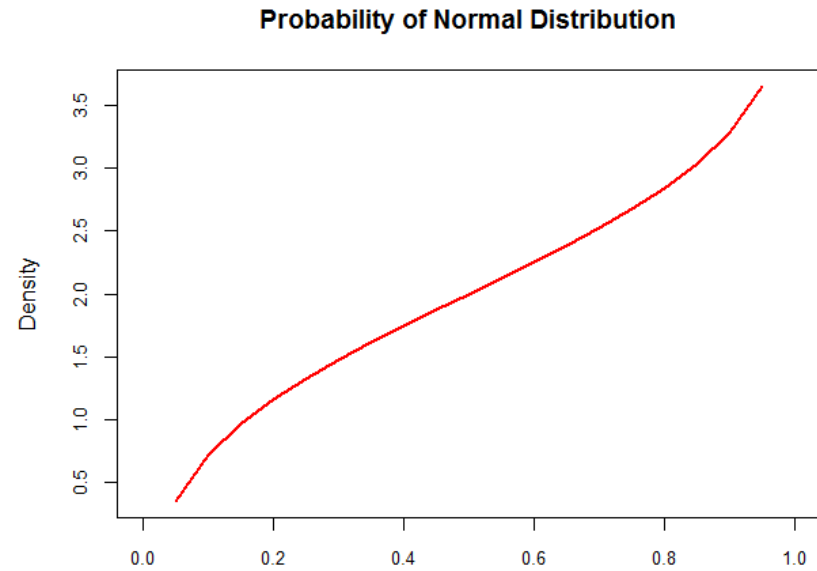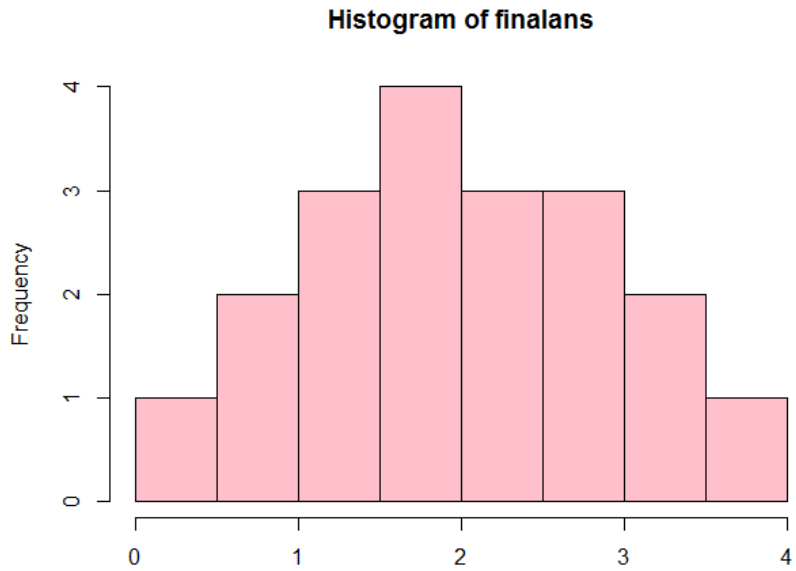**Figure 8.6** Probability of normal distributed values.

Figure 8.7 Histogram and plot of probability values from specified mean and standard deviation.
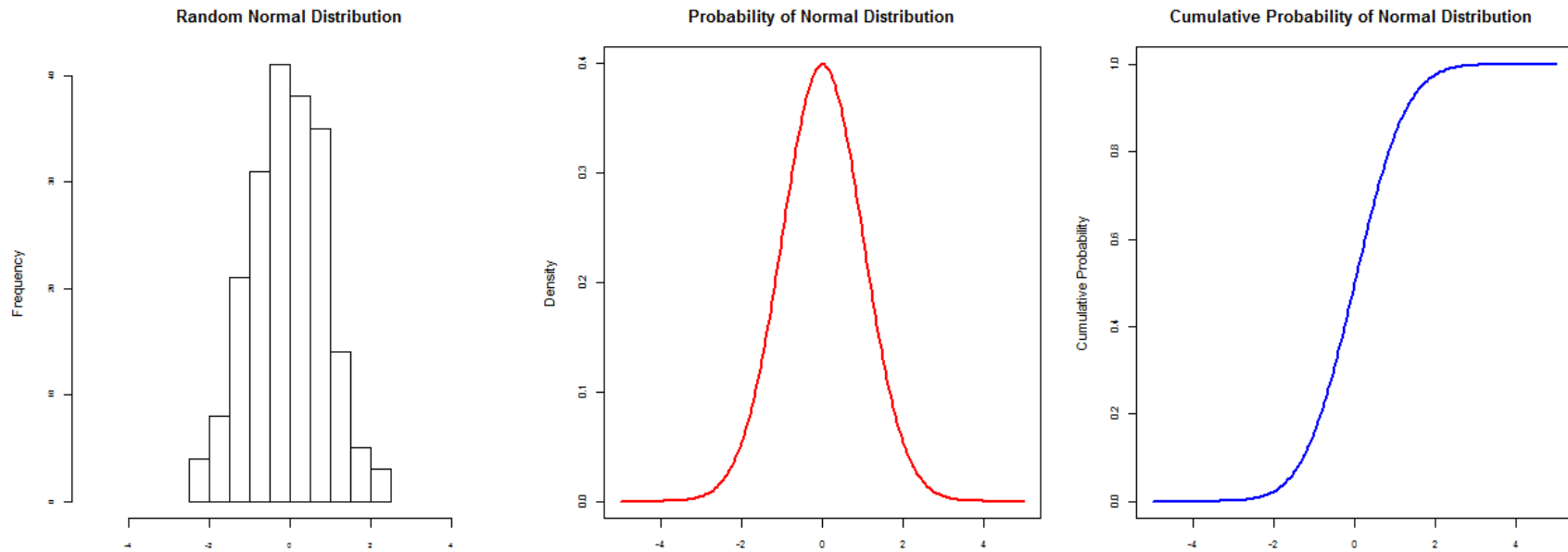
Figure 8.8 Multiple charts displaying functions of normal distribution.
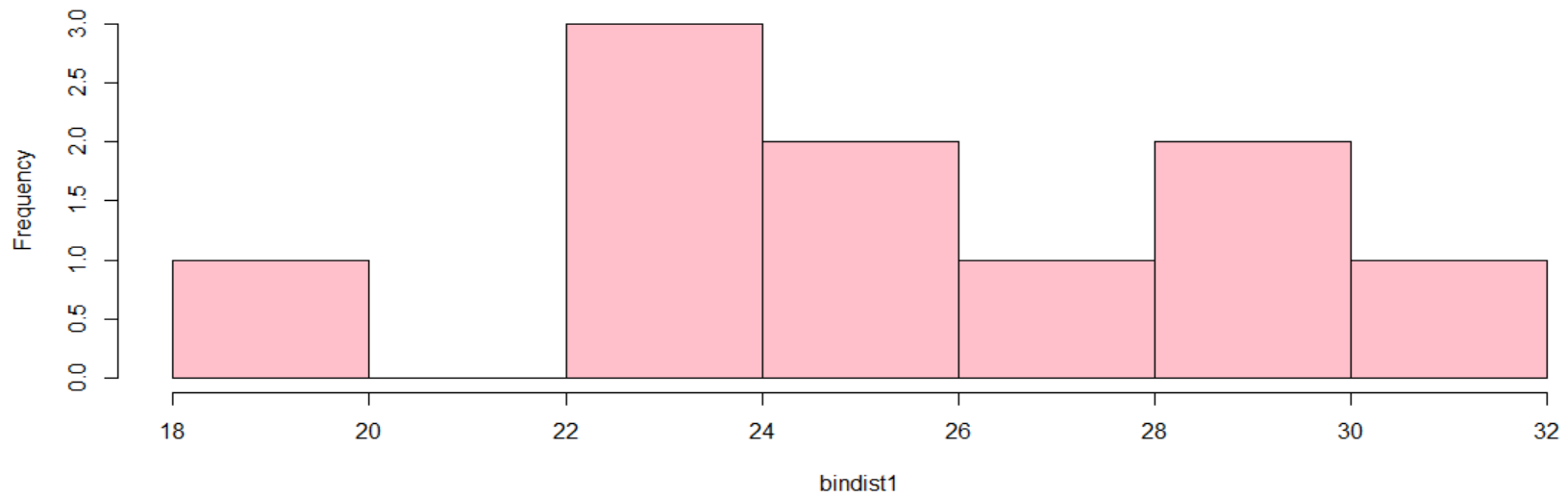
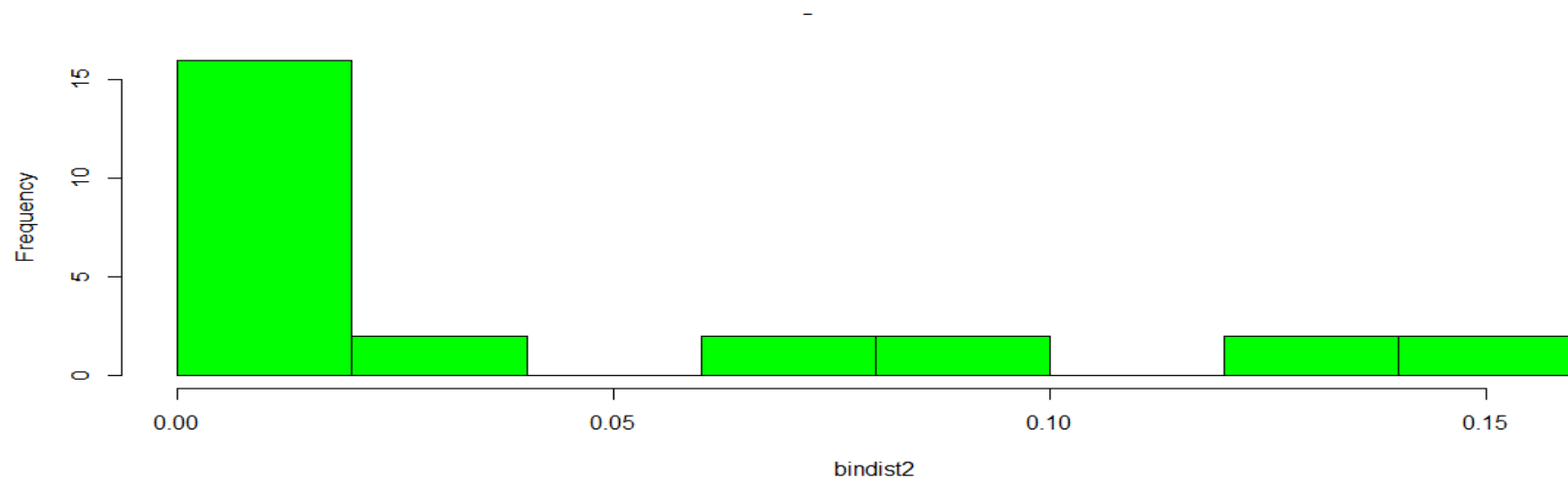**Figure 8.9** Histogram of random values following binomial distribution.

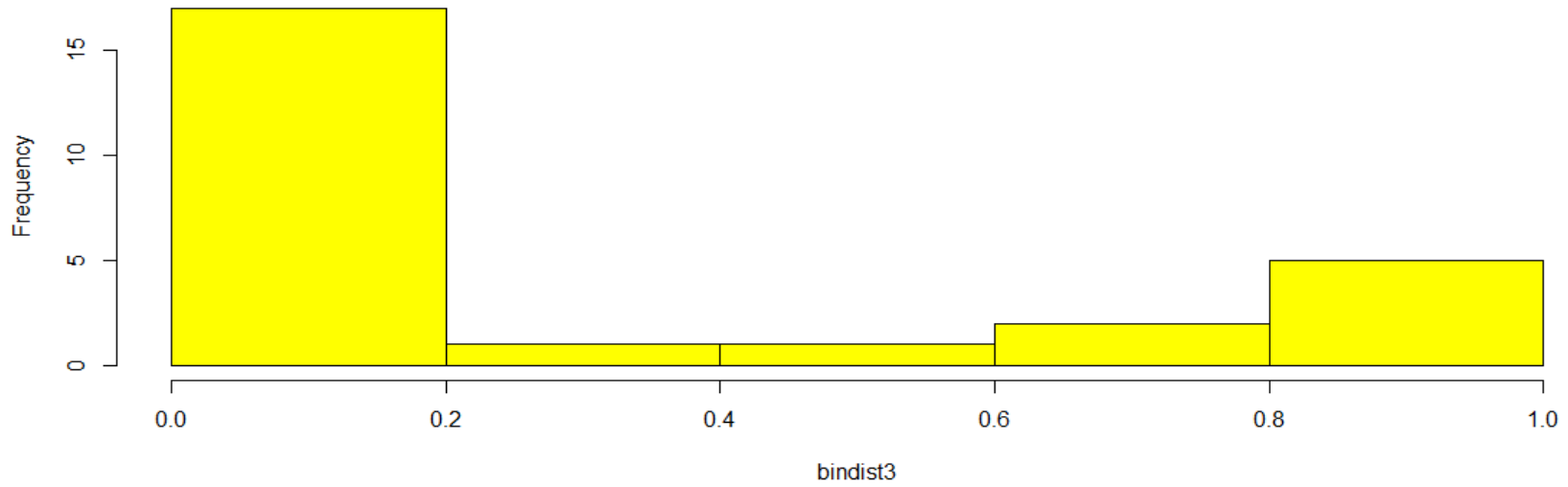**Figure 8.10** Histogram of probability density distribution following binomial distribution.

Figure 8.11 Histogram of cumulative probability of event following binomial distribution.
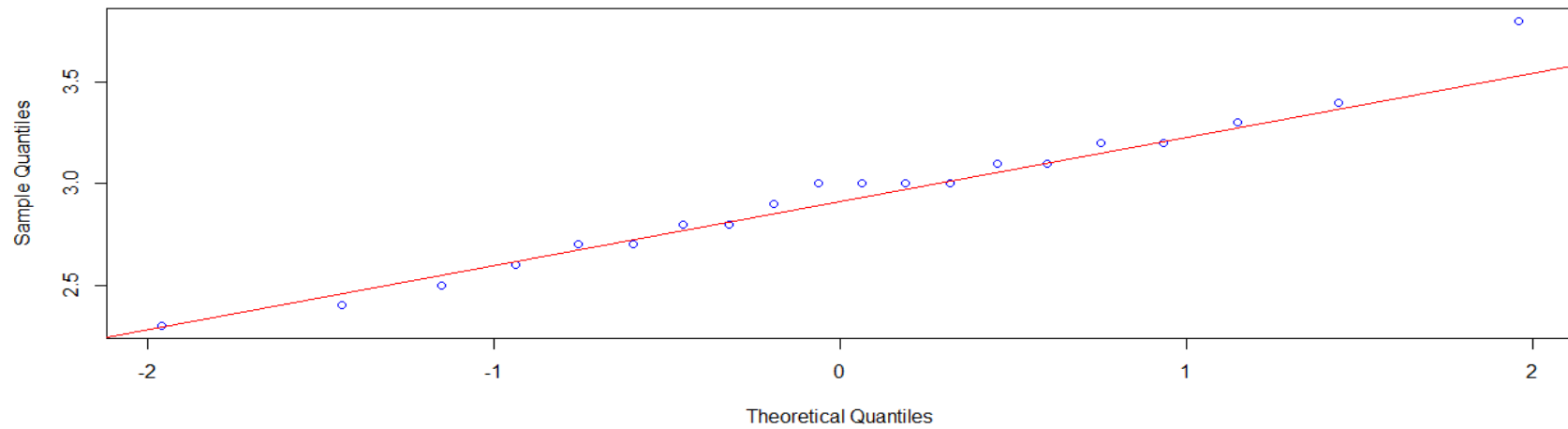
CHAPTER 9
Compare Means

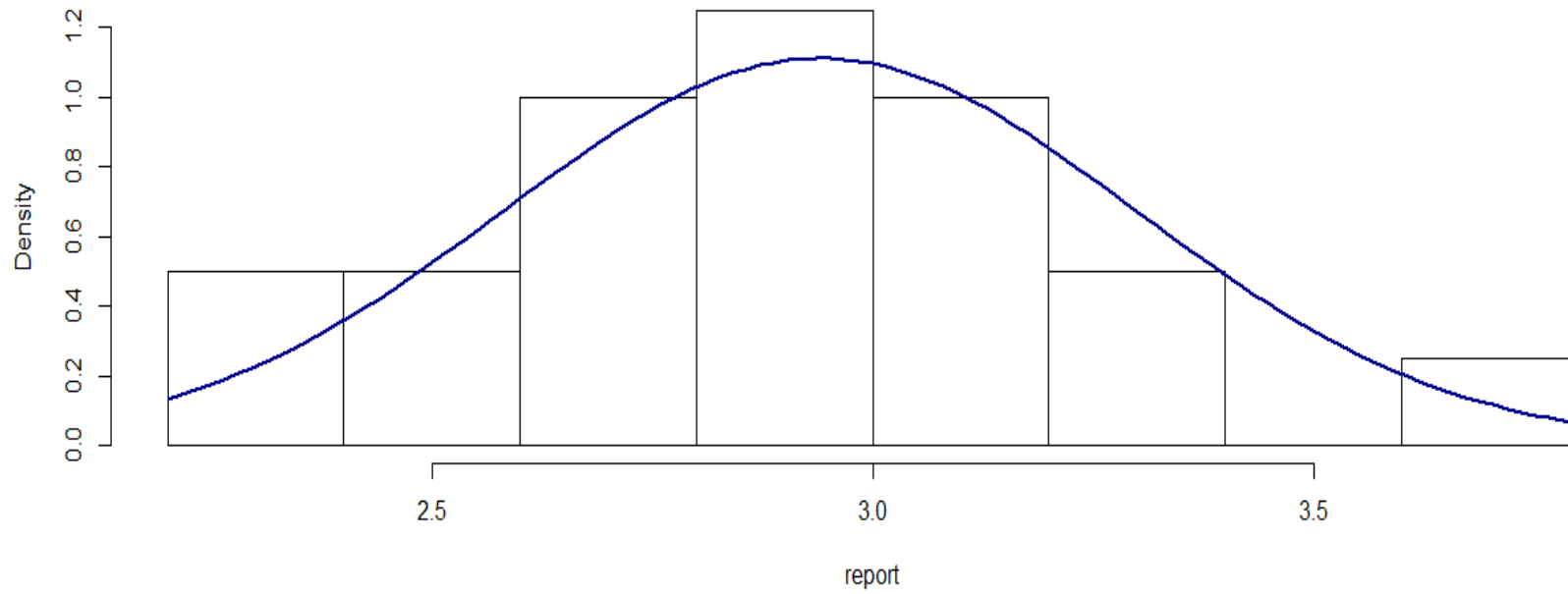**Figure 9.1** Q–Q plot depicting normality of the report data.

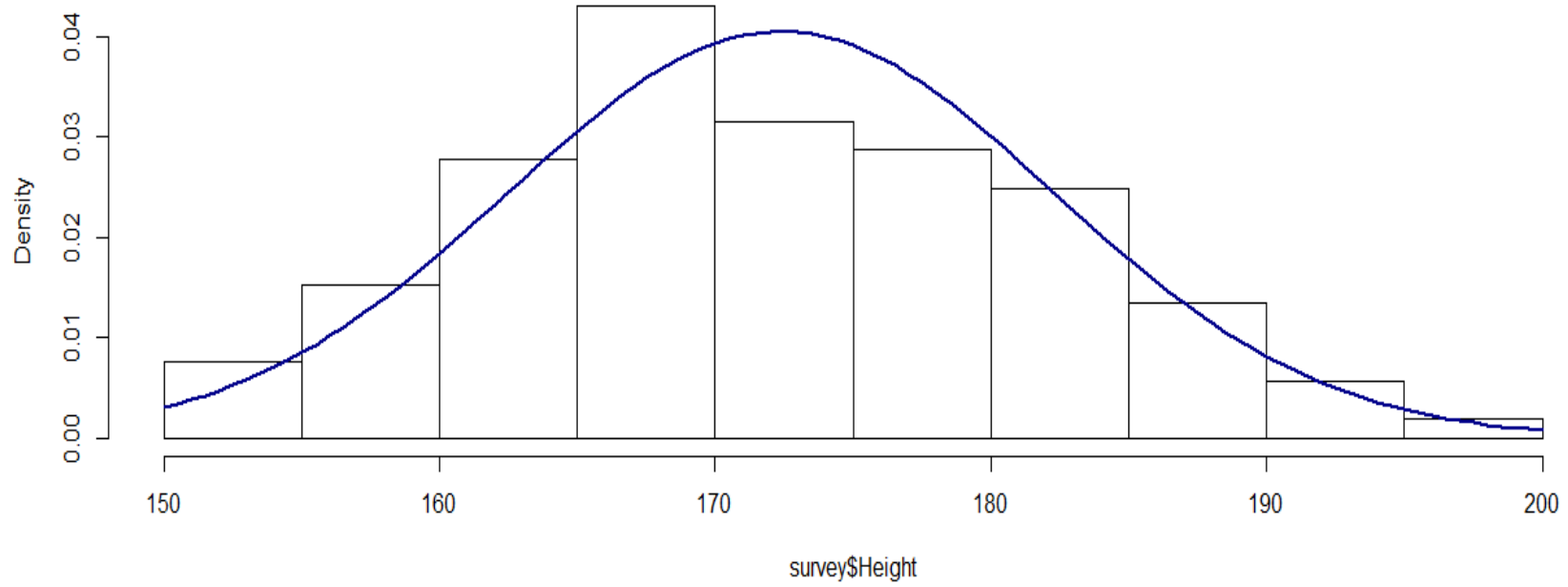**Figure 9.2** Normality curve depicting normality of report data.

Figure 9.3 Normality curve depicting normality of report data.
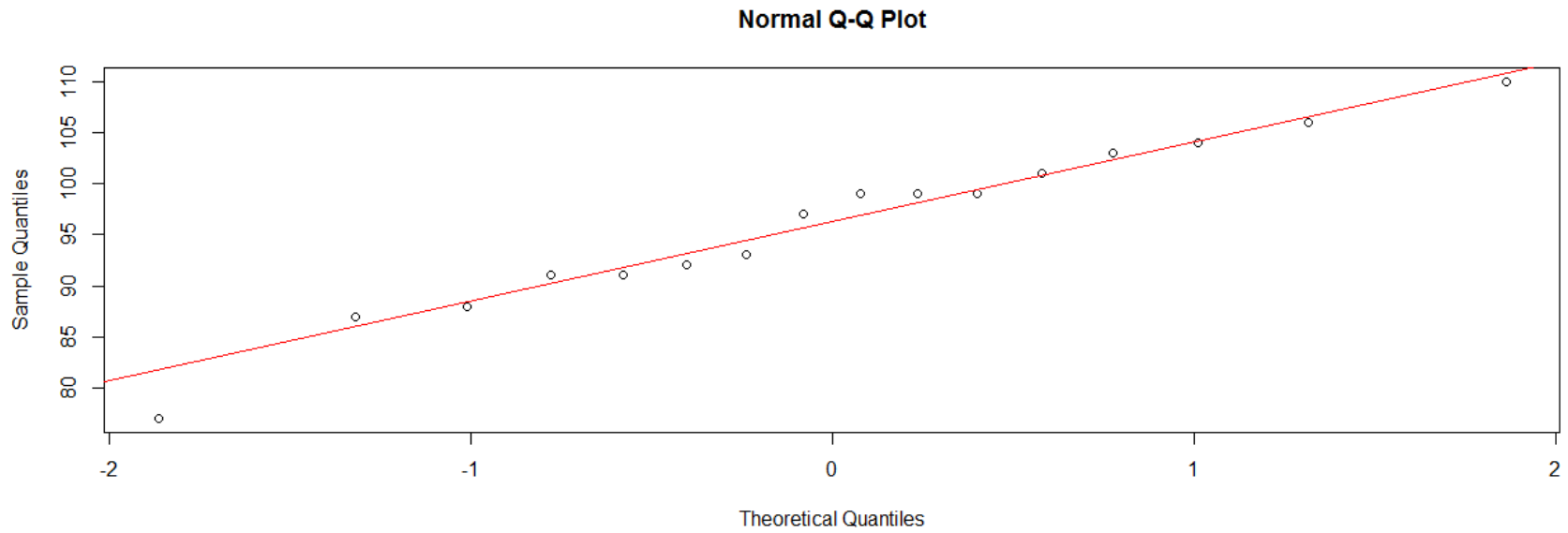
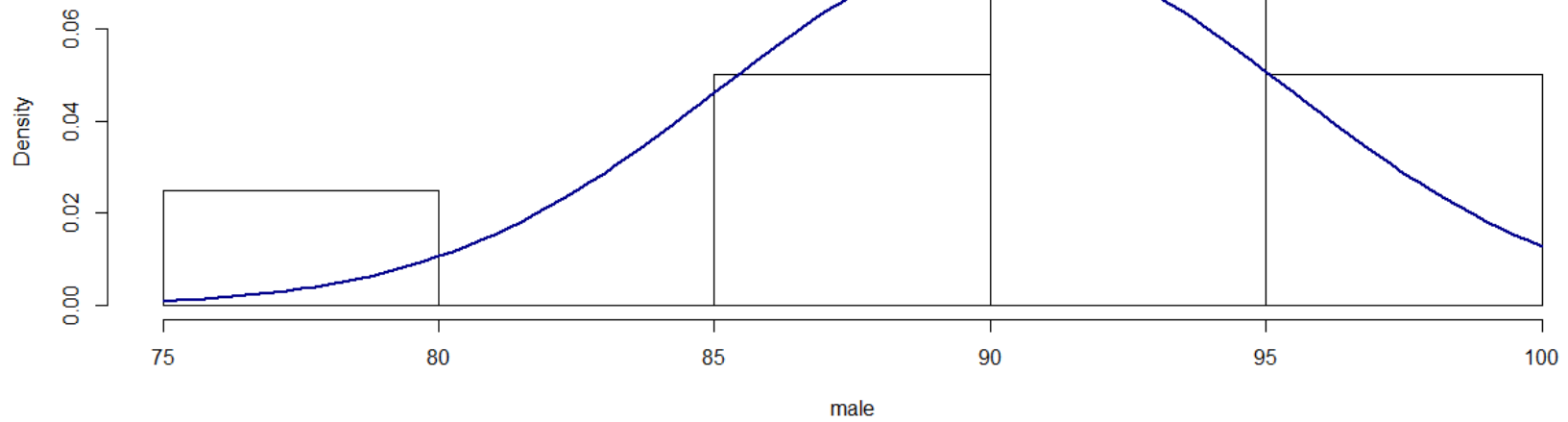**Figure 9.4** `qqnorm()` and `qqline()` functions for checking normality.

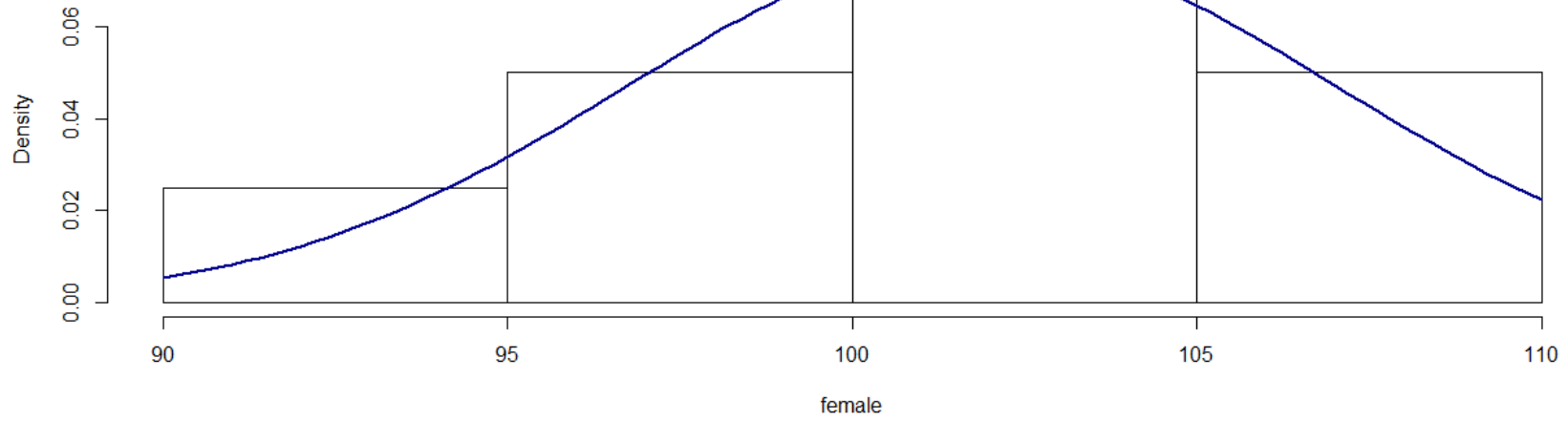**Figure 9.5** Chart displaying histogram and normality curve of male data.

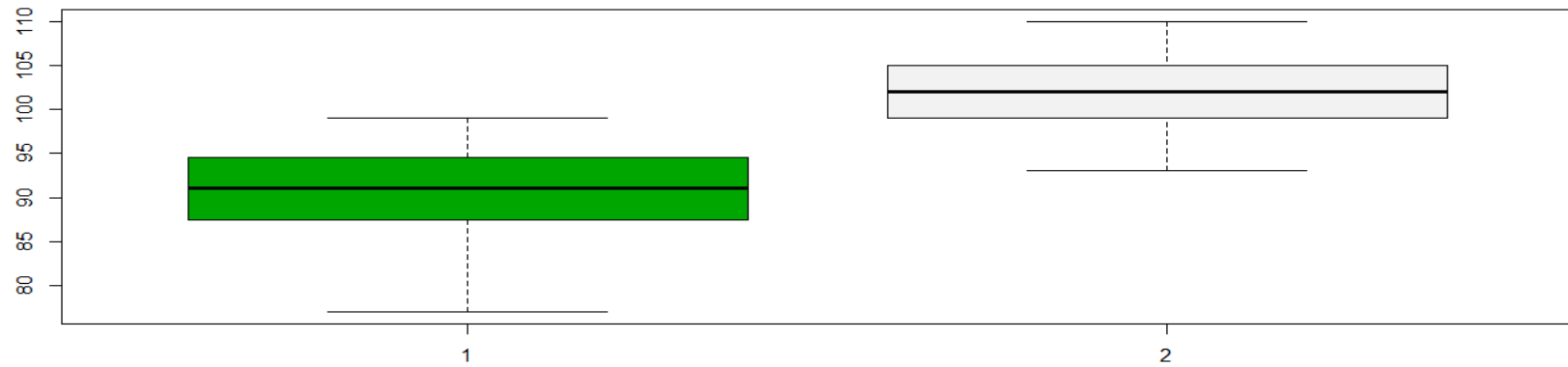**Figure 9.6** Chart displaying histogram and normality curve of female data.
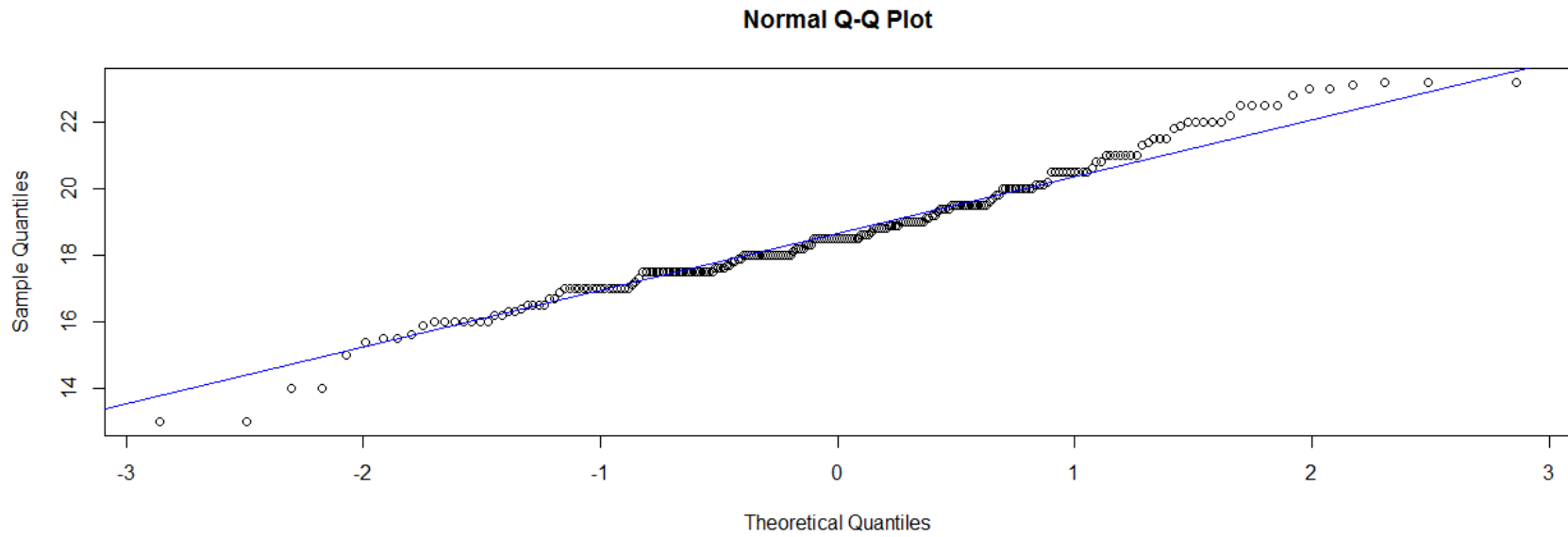
Figure 9.7 Box plot for male and female.

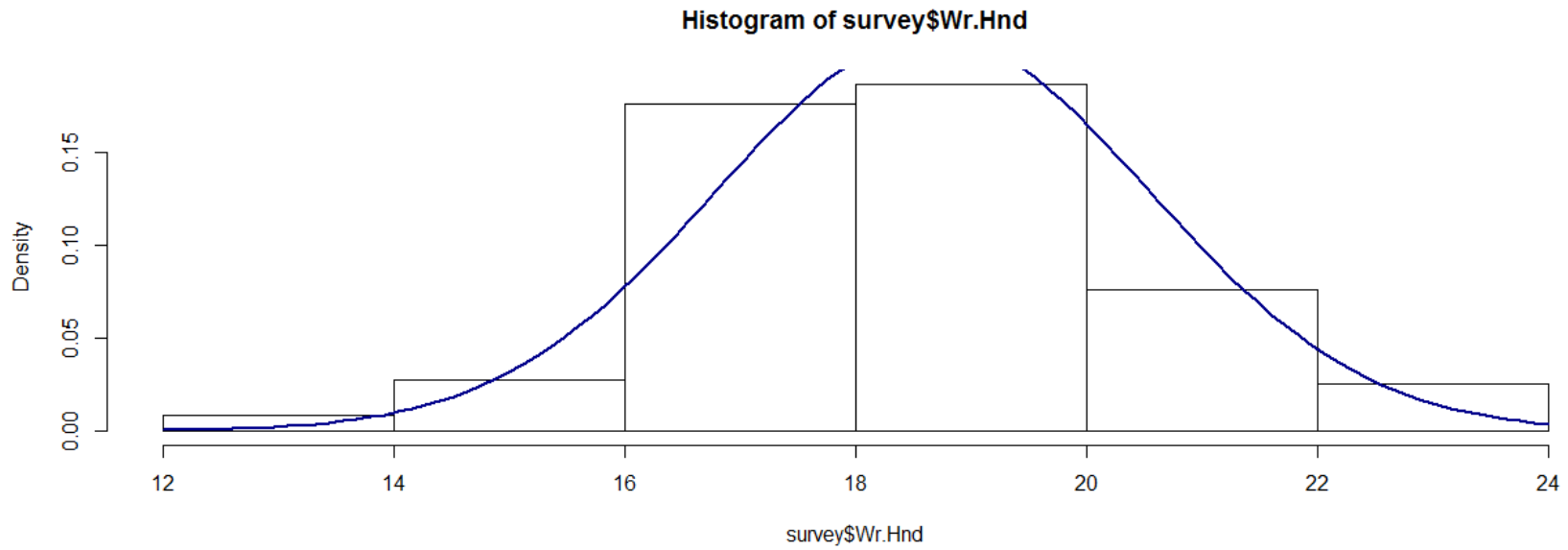**Figure 9.8** `qqnorm()` and `qqline()` functions for checking normality.

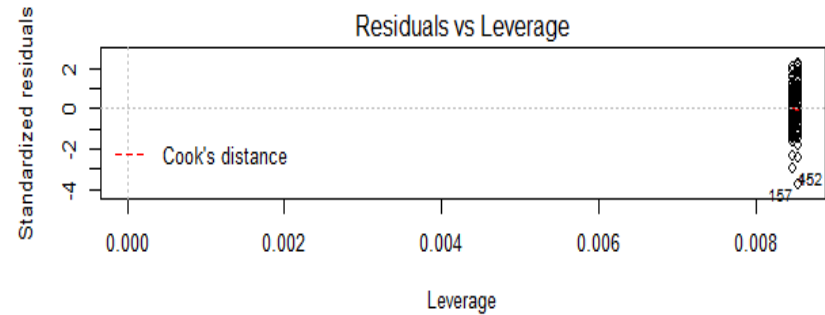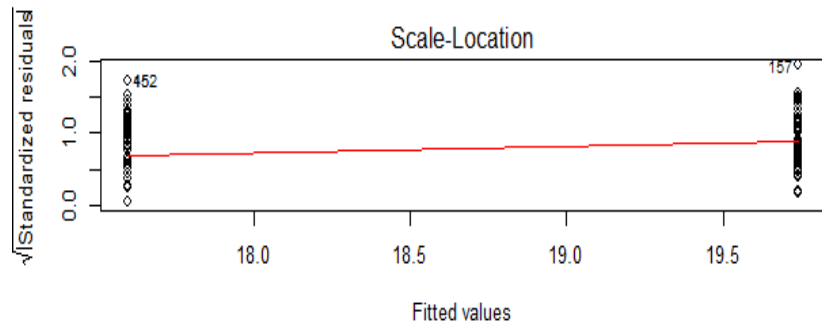**Figure 9.9** Histogram and normality curve for Wr.Hnd in "survey" dataset.
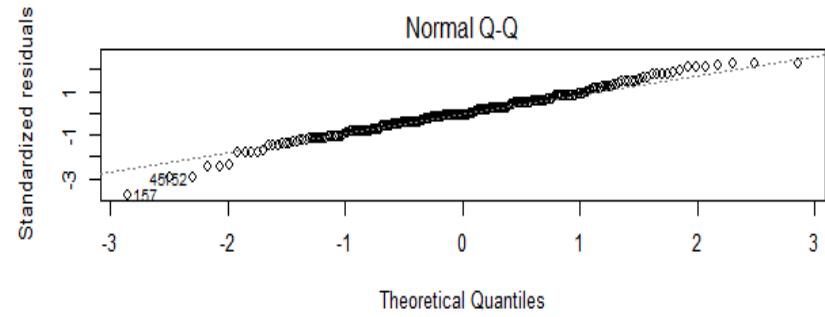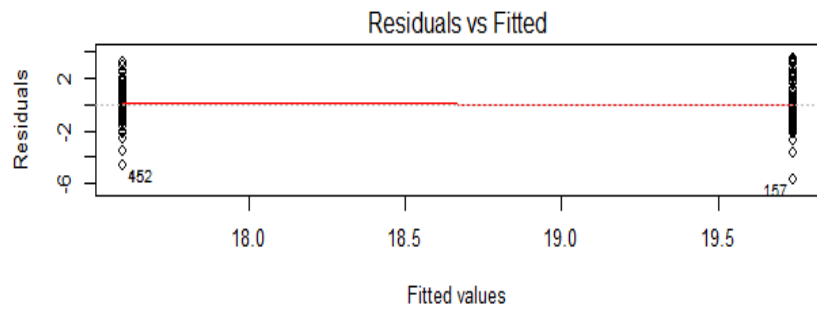
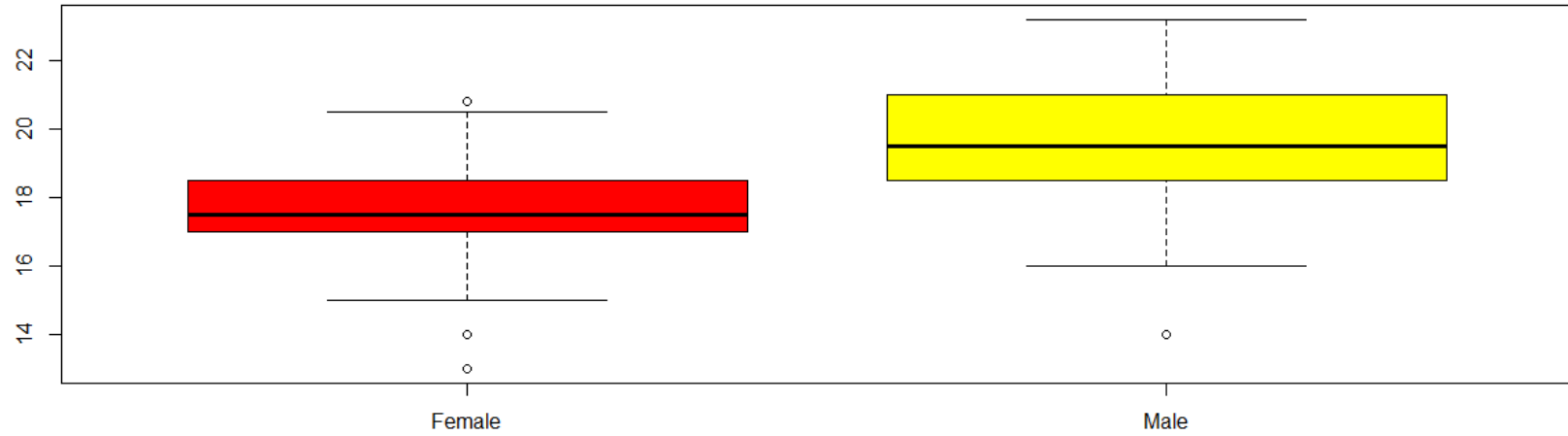**Figure 9.10** Graphical evaluation of assumptions of t-test.

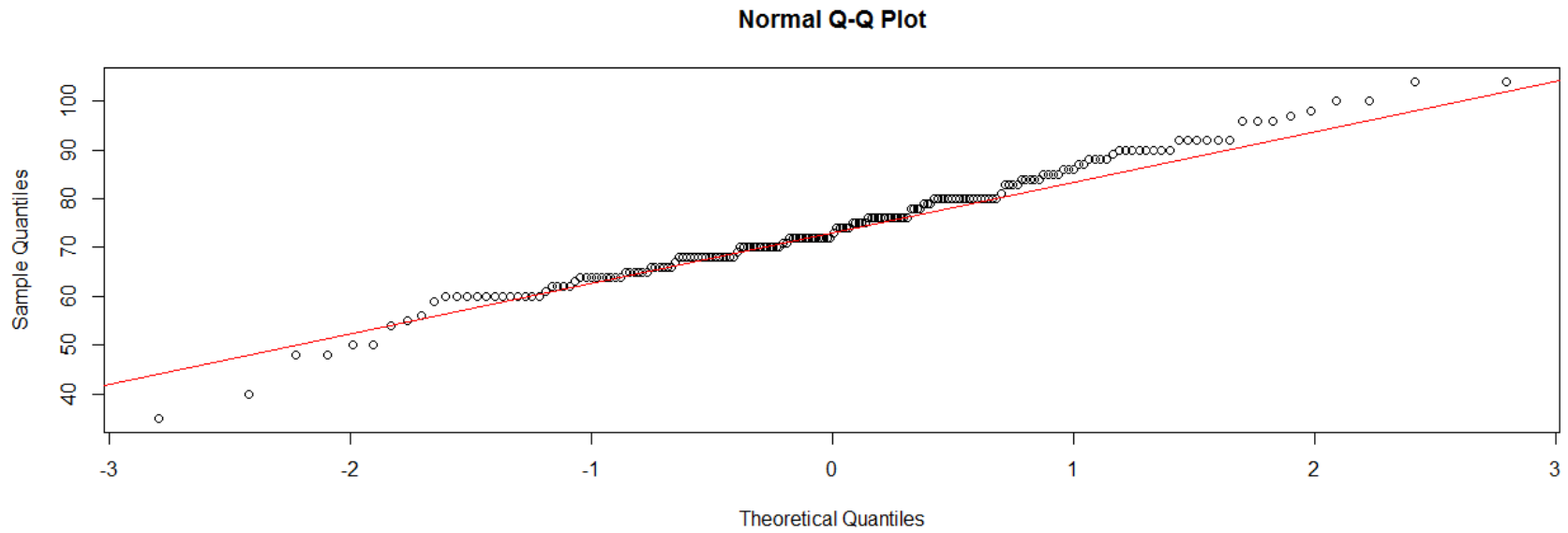**Figure 9.11** Box plot depicting different groups of Sex categorical variable with respect to Wr.Hnd.

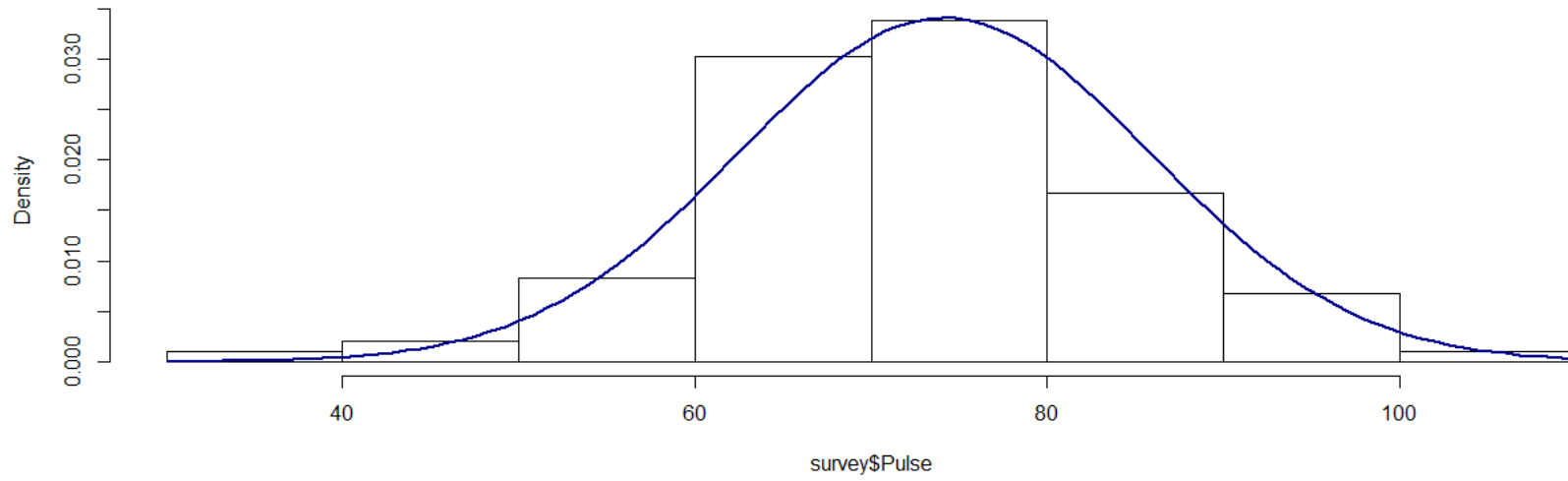**Figure 9.12** `qqnorm()` and `qqline()` functions for determining normality.

Figure 9.13 Histogram and normality curve depicting normality.

Figure 9.14 Graphical evaluation of assumptions of t-test.

WILEY

Figure 9.15 Box plot for three different groups.

Figure 9.17 `qqnorm()` and `qqline()` functions for determining normality.

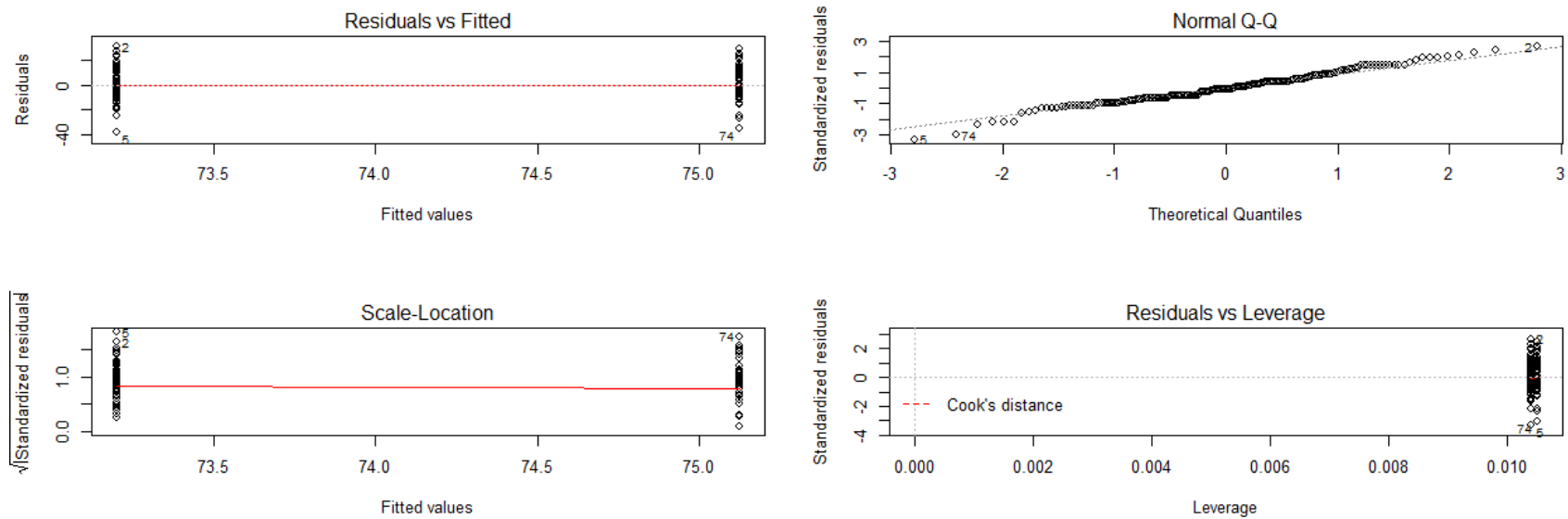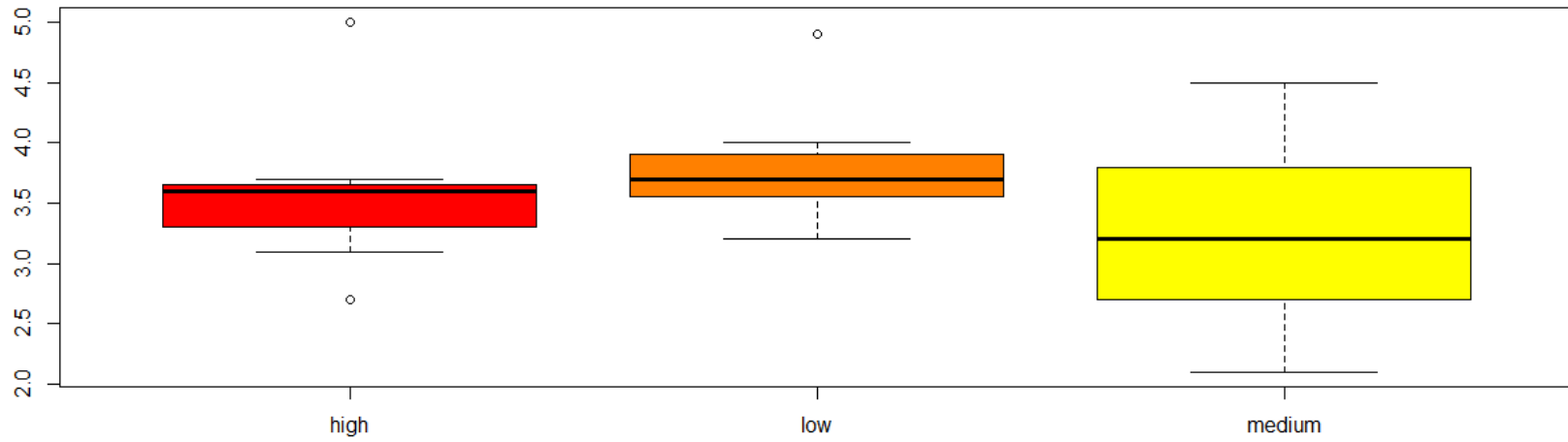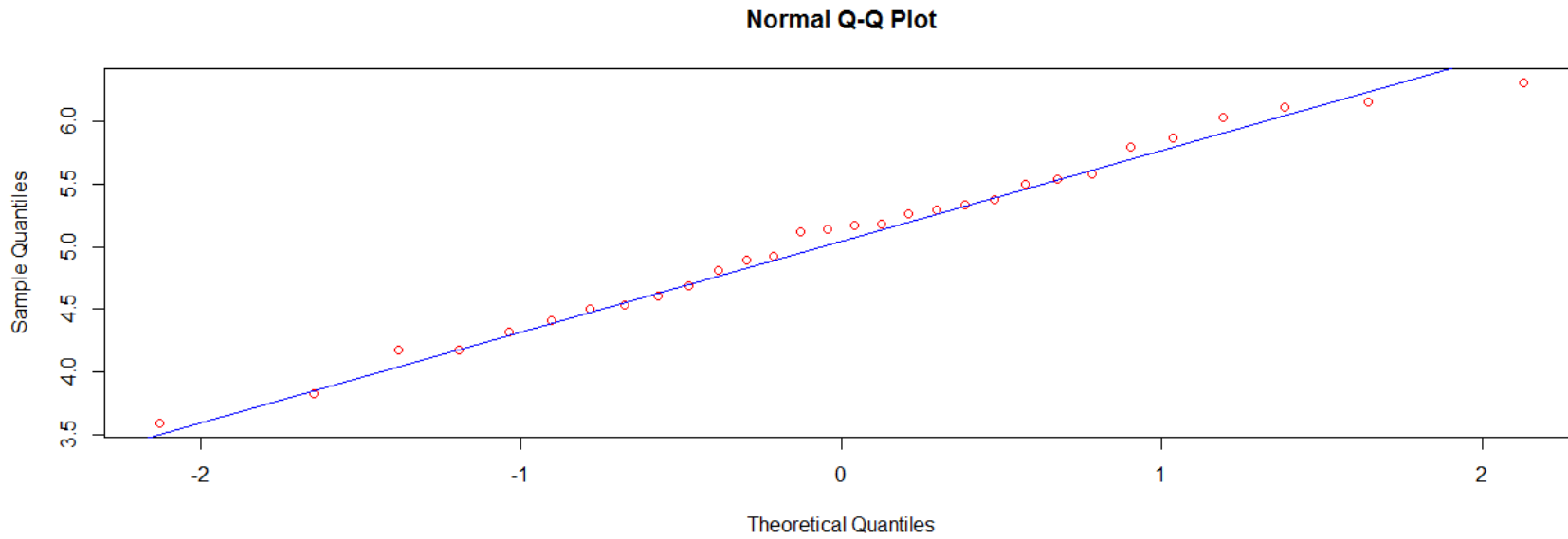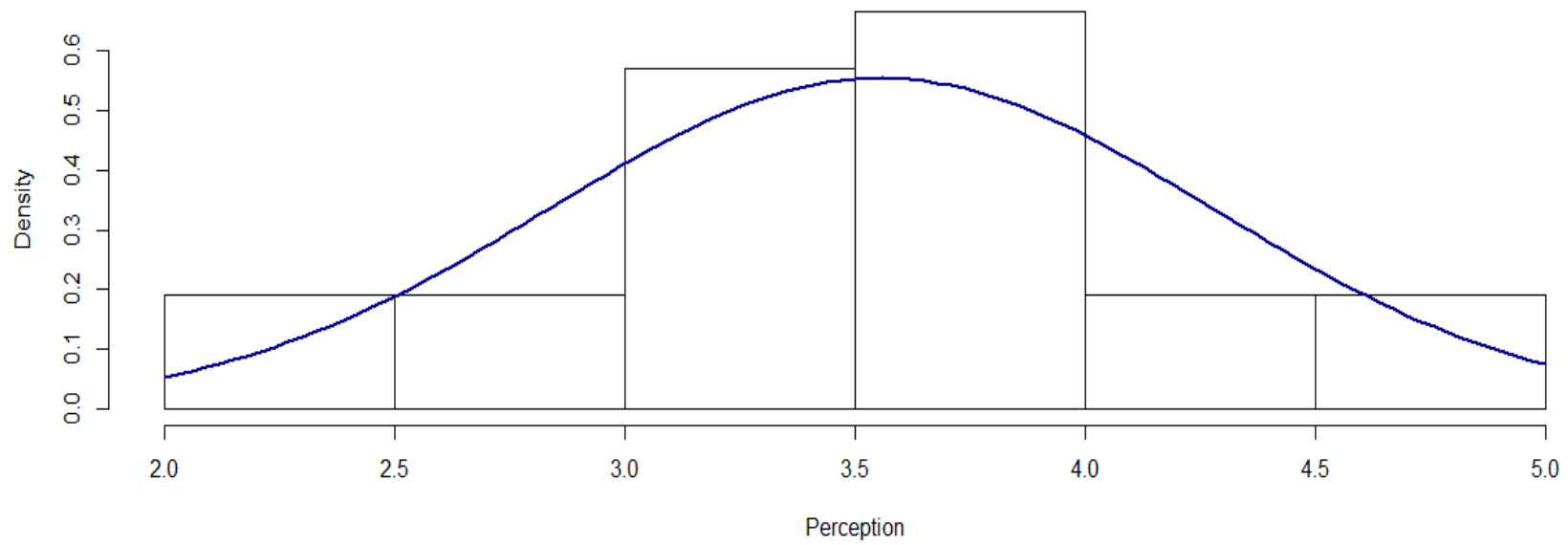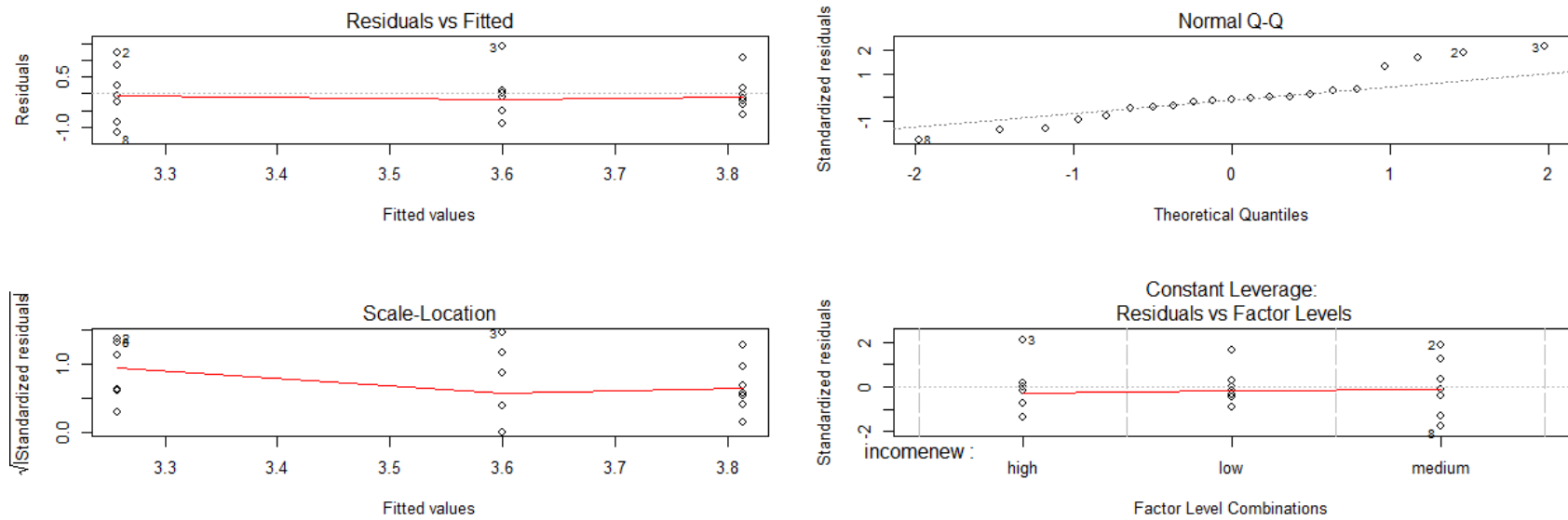**Figure 9.18** Histogram and normality curve depicting normality of PlantGrowth.

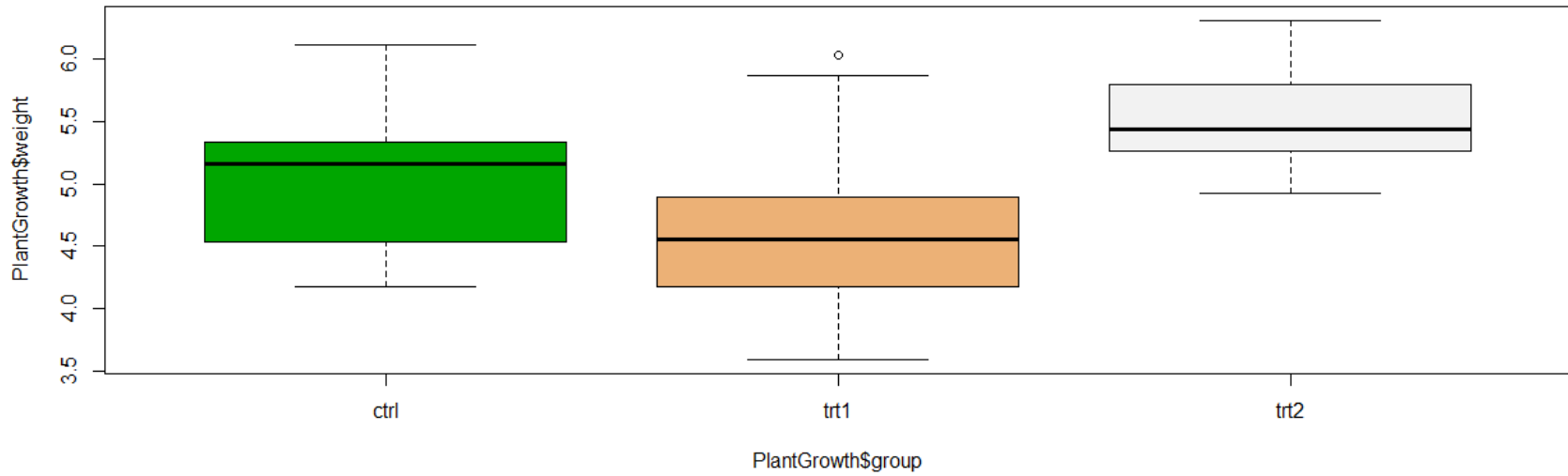**Figure 9.19** Graphical evaluation of assumptions of *t*-test.

Figure 9.20 Box plot for three different groups.
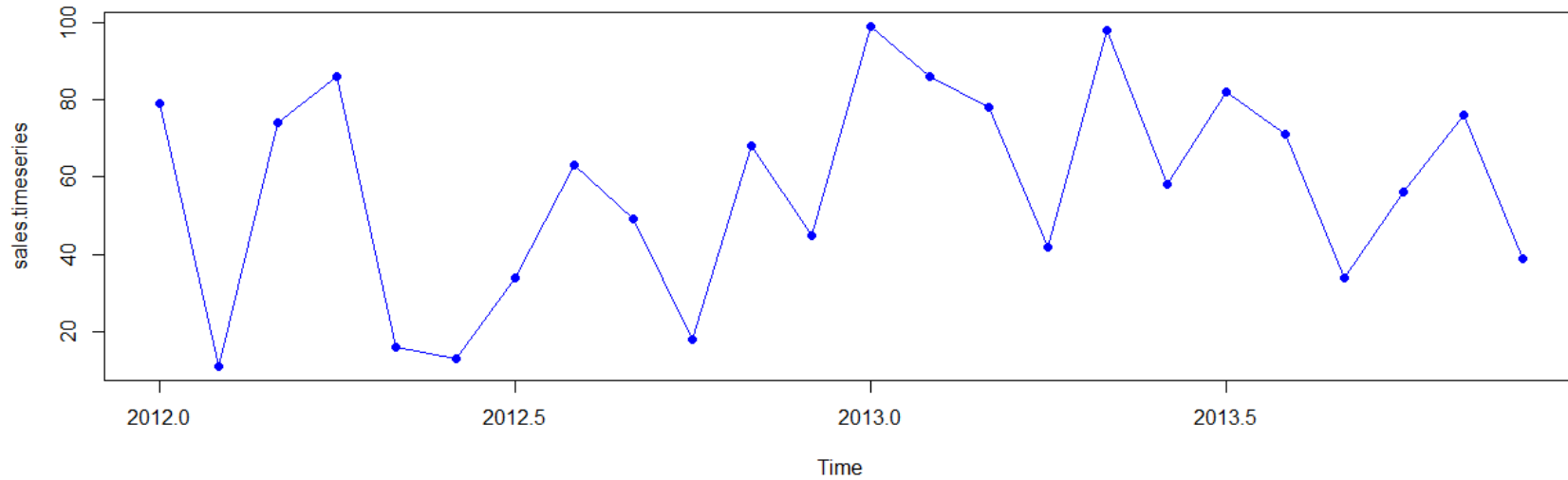
CHAPTER 10
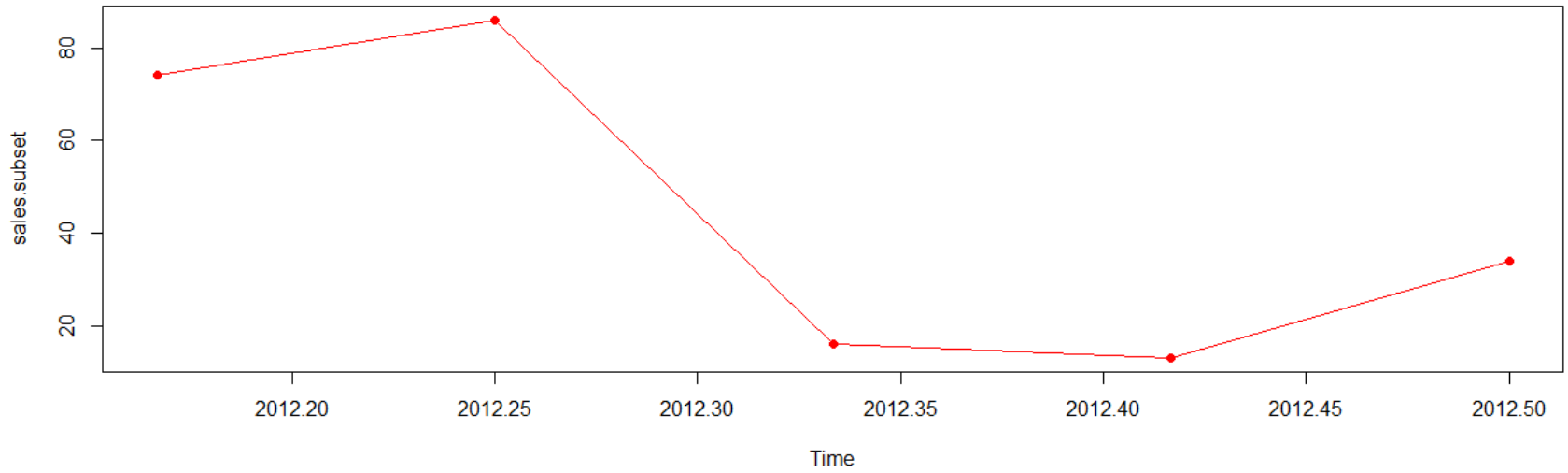# Time-Series Models

Figure 10.1 Time series for sales.

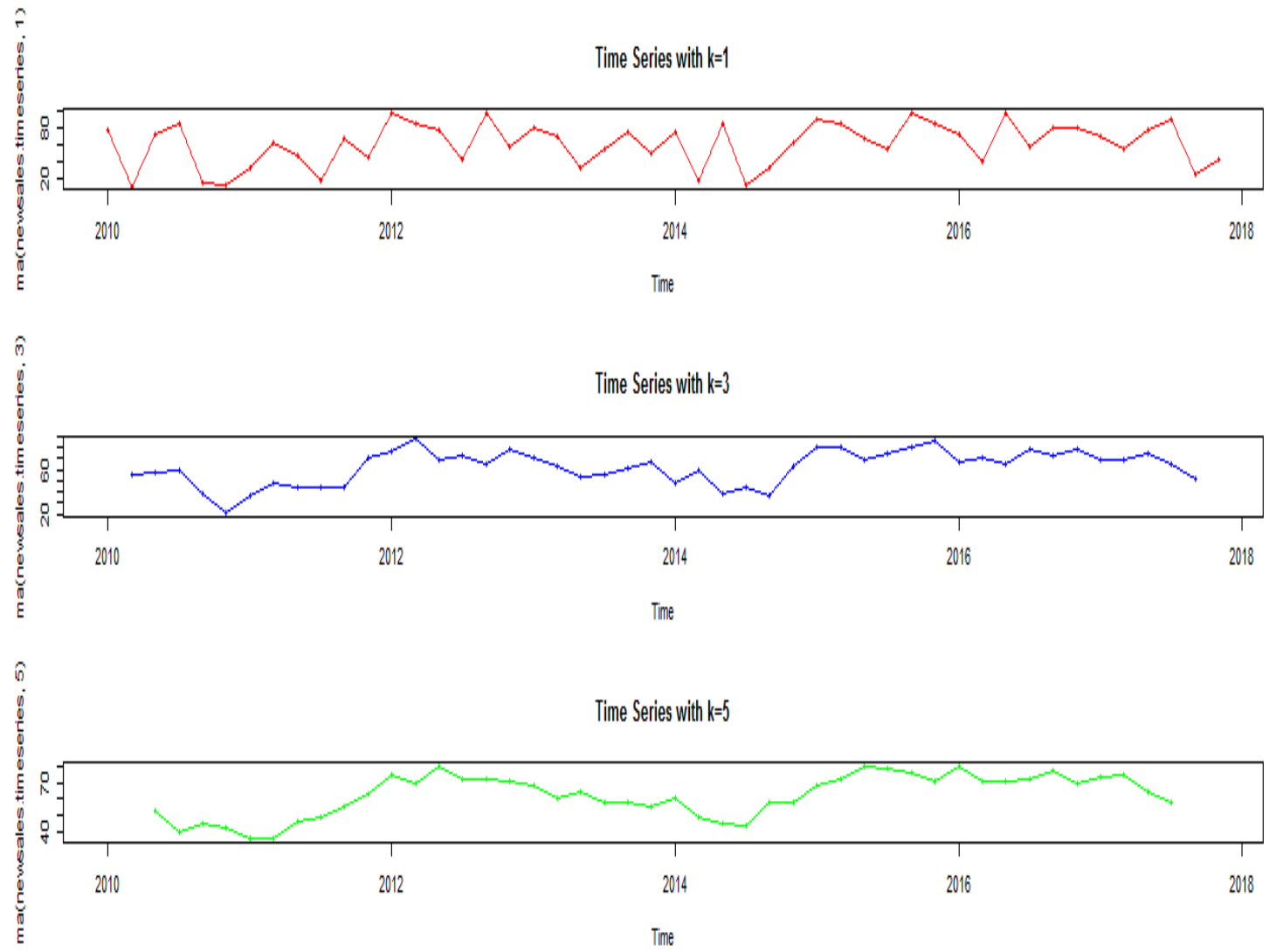**Figure 10.2** Time series for selected values of the time-series object.

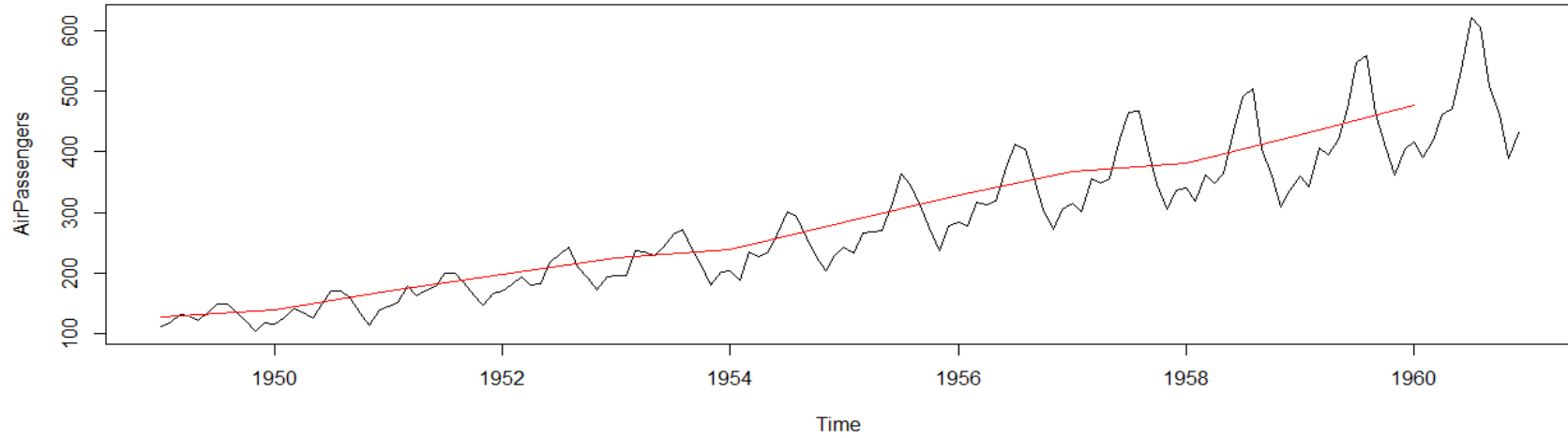**Figure 10.4** Smoothing effect on user-defined data for time series.
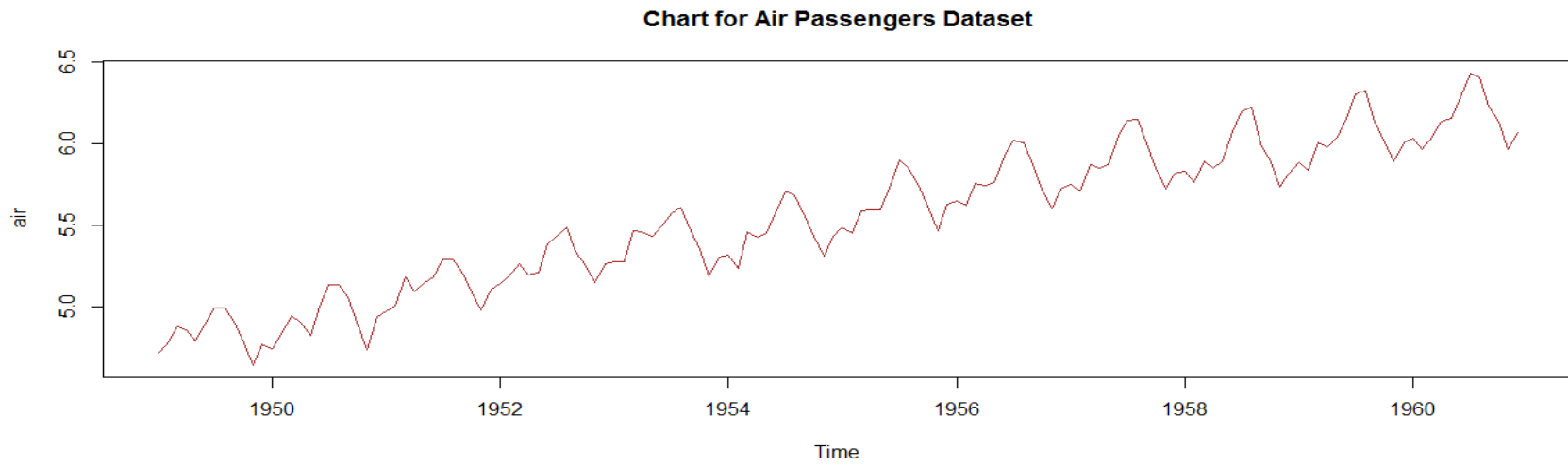
Figure 10.6 Display data of "AirPassengers" dataset.

Chart for Air Passengers Dataset

Figure 10.7 Using `log()` function to display data of "AirPassengers" dataset.
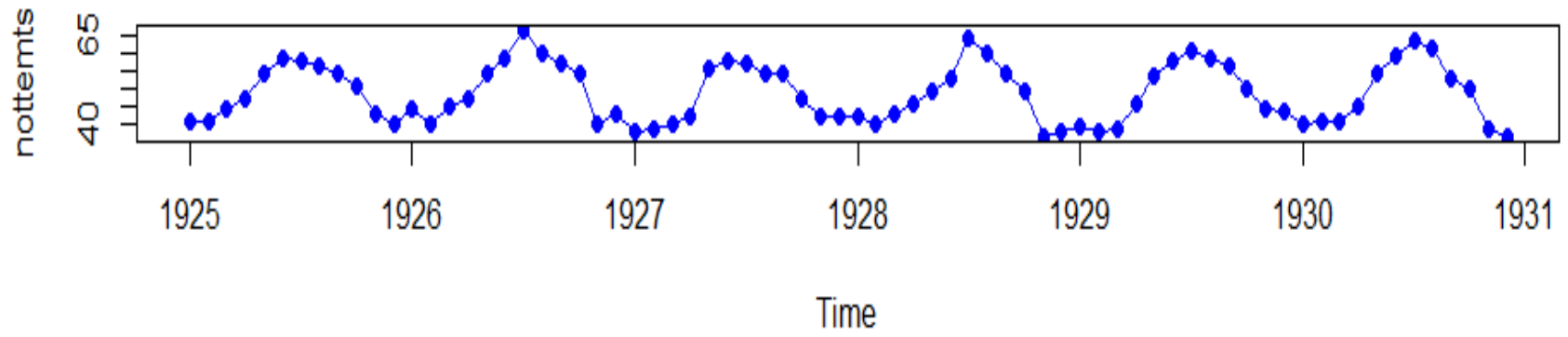
Figure 10.20 Time series for selected years.

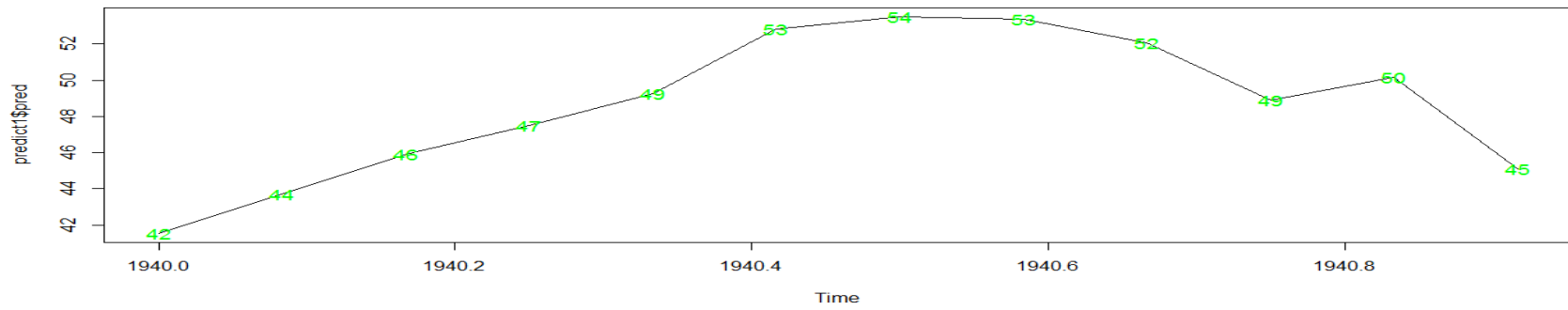**Figure 10.21** Prediction of values through ARIMA modeling.
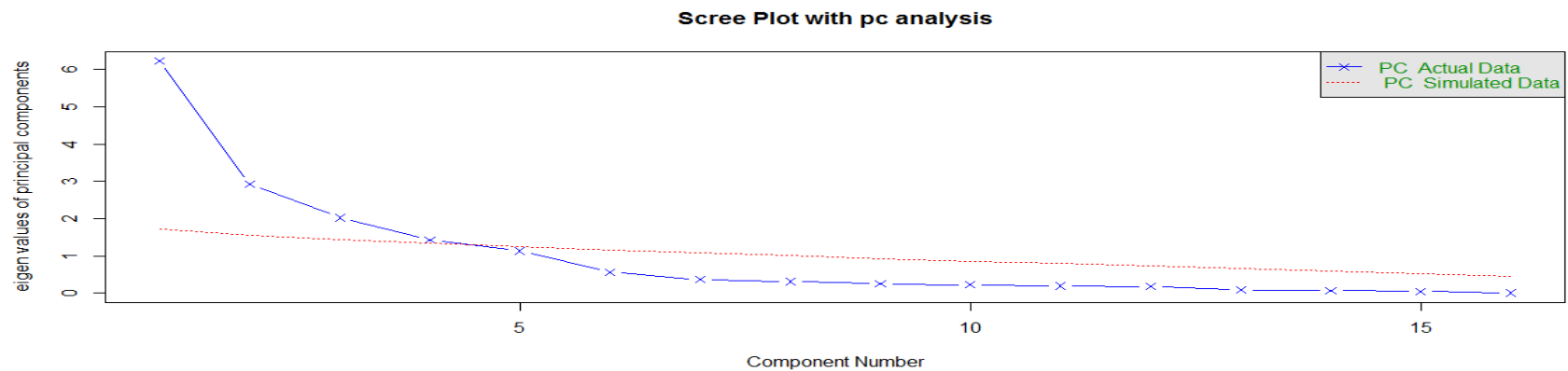
CHAPTER 11
Unsupervised Machine Learning Algorithms

Scree Plot with pc analysis

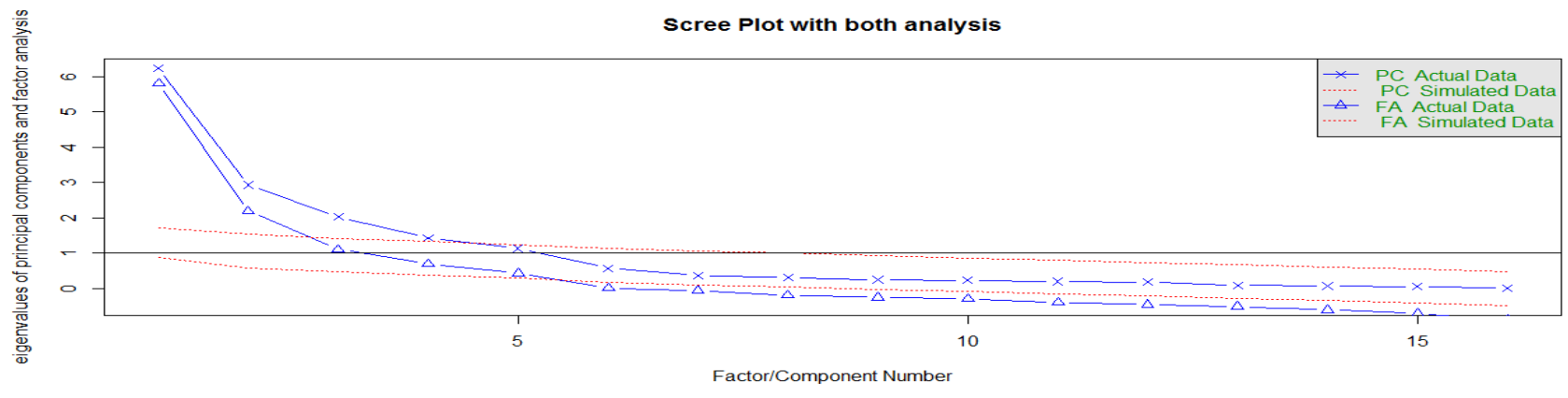**Figure 11.1** Scree plot with "pc" analysis.

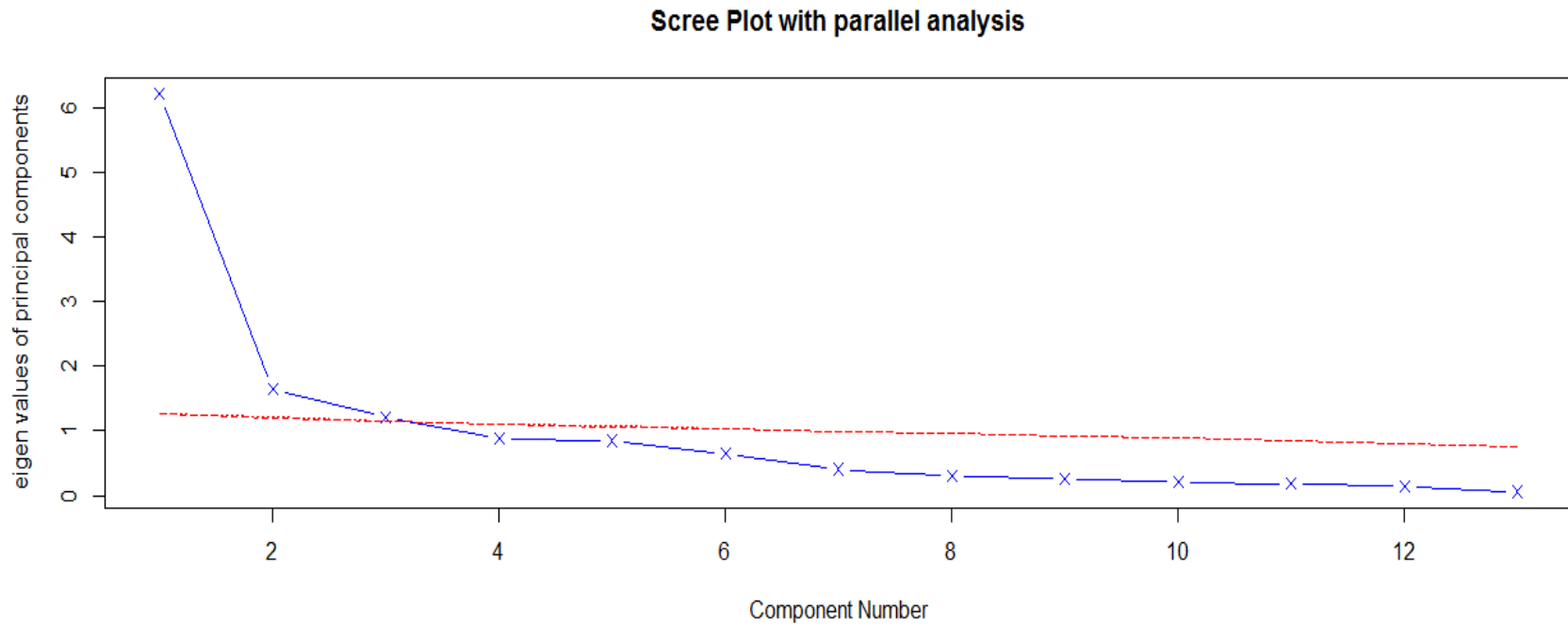**Figure 11.2** Scree plot with "both" analysis.

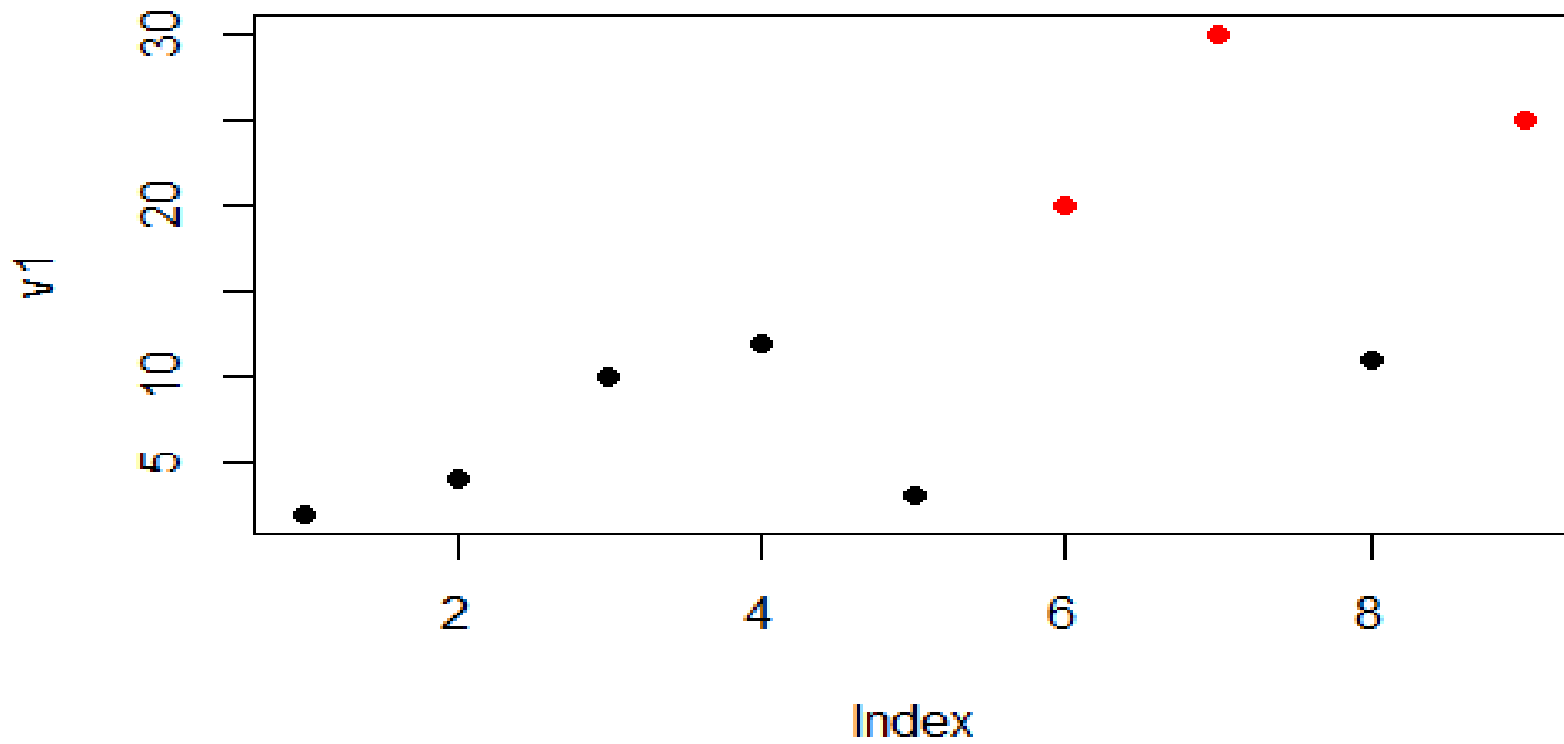**Figure 11.3** Scree plot with parallel analysis for determining number of factors.

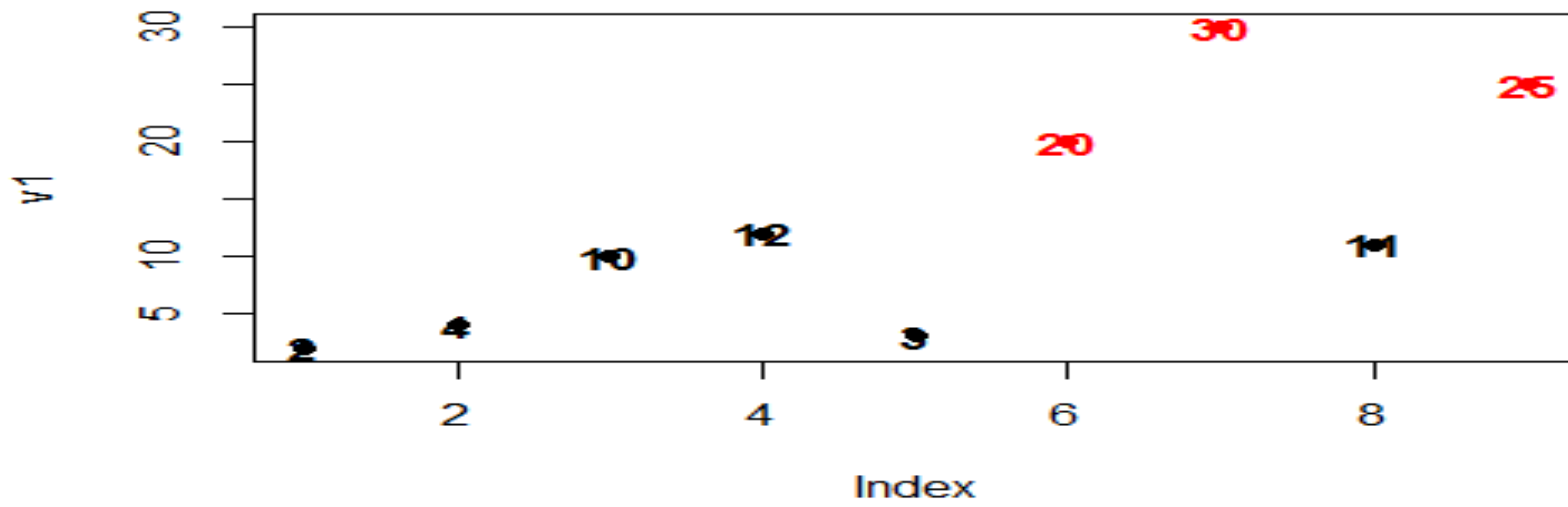**Figure 11.4** Displaying dots of different colors for different clusters.

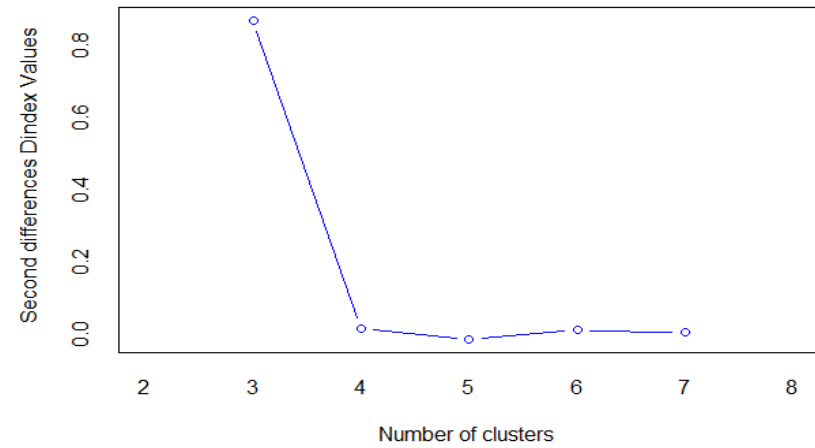**Figure 11.5** Adding text to the above chart.

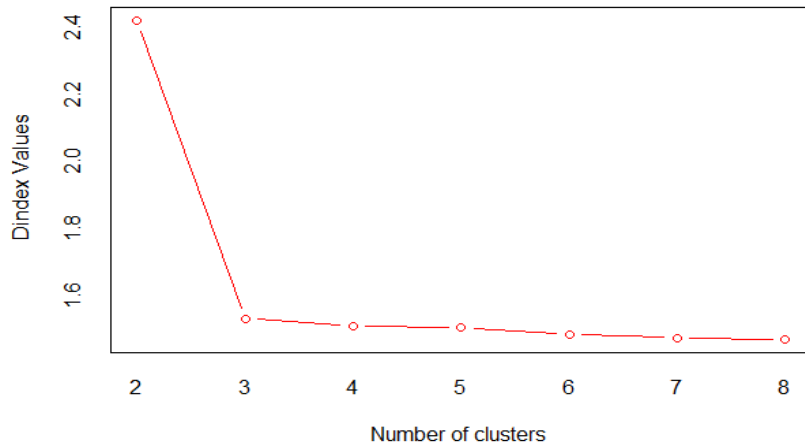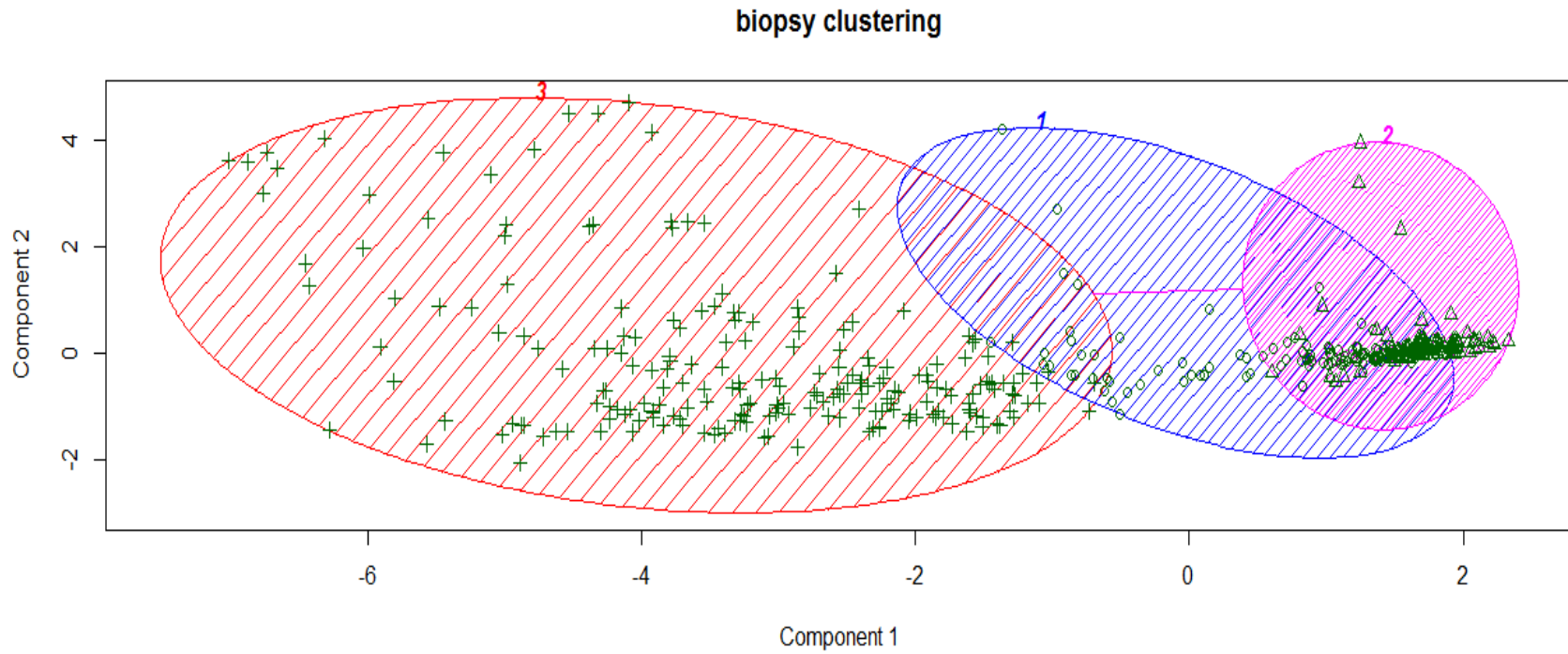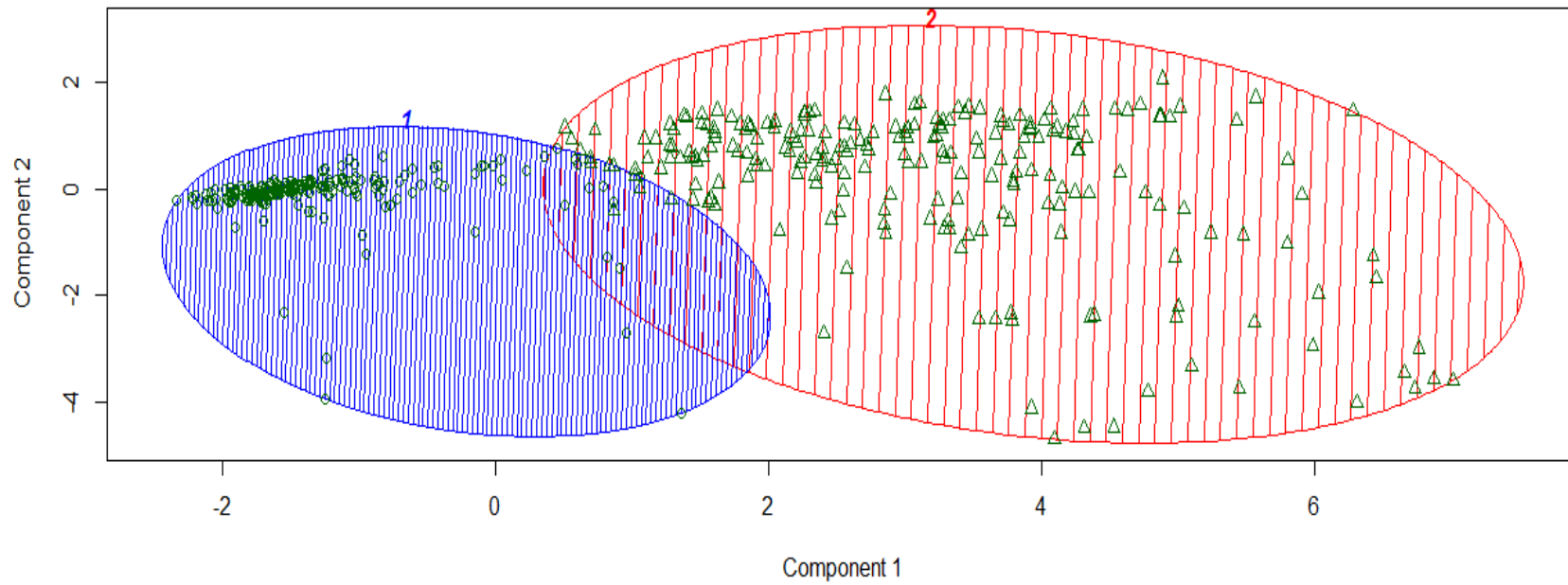**Figure 11.6** Determining optimum number of clusters for "biopsy" dataset.

Figure 11.8 Cluster plot for "biopsy" dataset.

Figure 11.9 Cluster analysis considering 2 clusters for "biopsy" dataset.
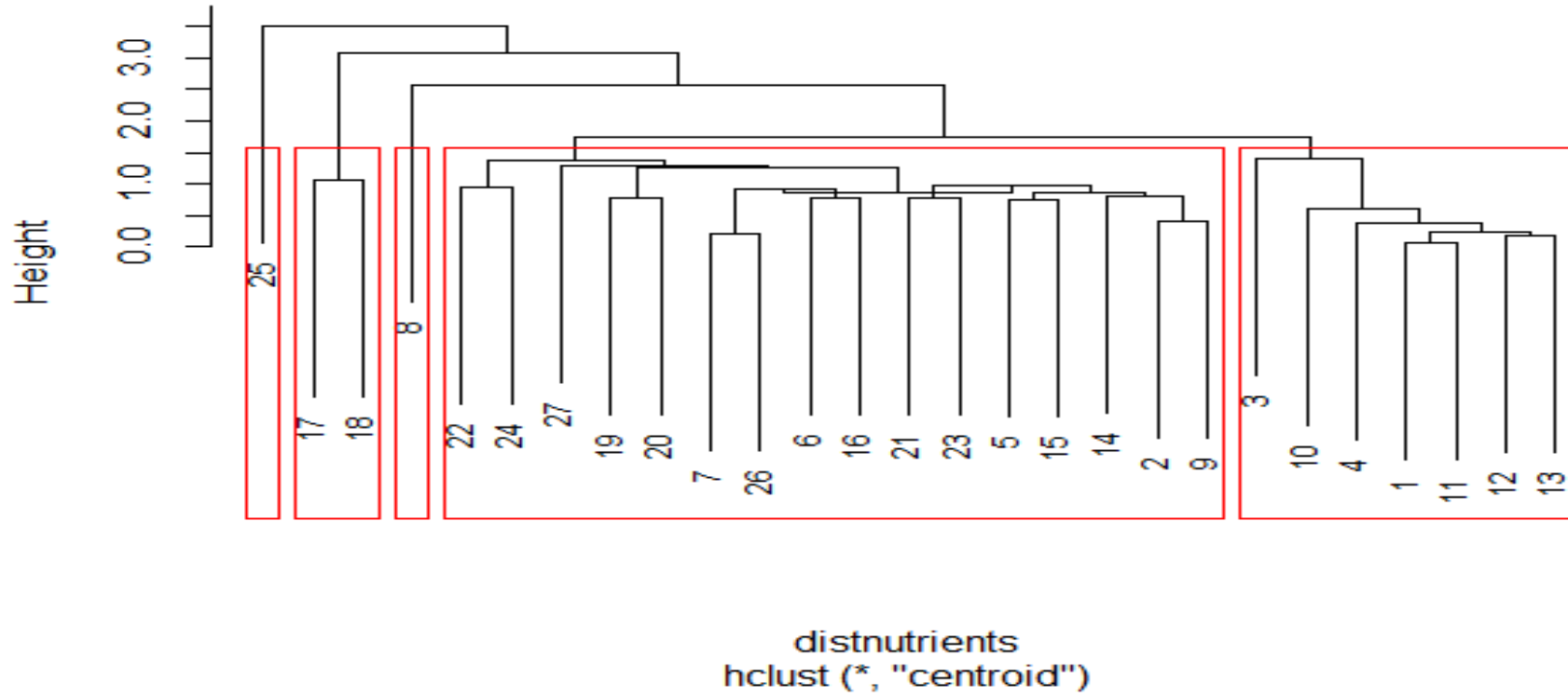
Figure 11.11 Hierarchical cluster analysis with specified number of clusters.
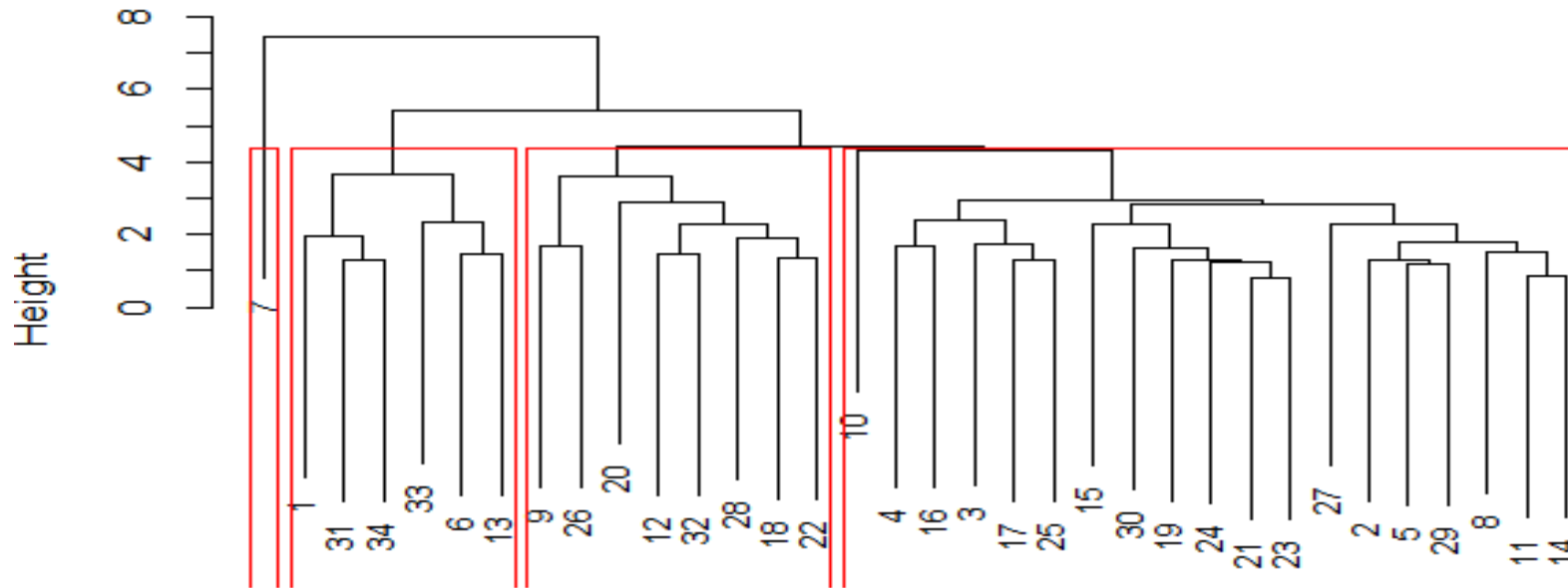
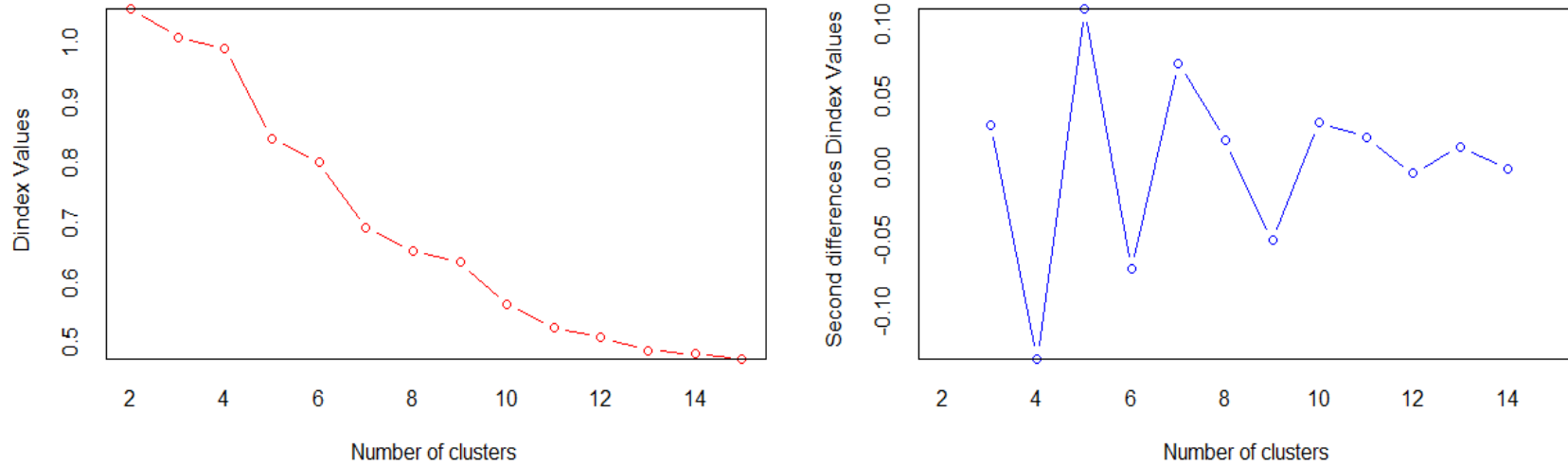**Figure 11.13** Hierarchical cluster analysis for "stock" dataset.

Figure 11.14 Determining optimum number of clusters using `Nbclust()` function.
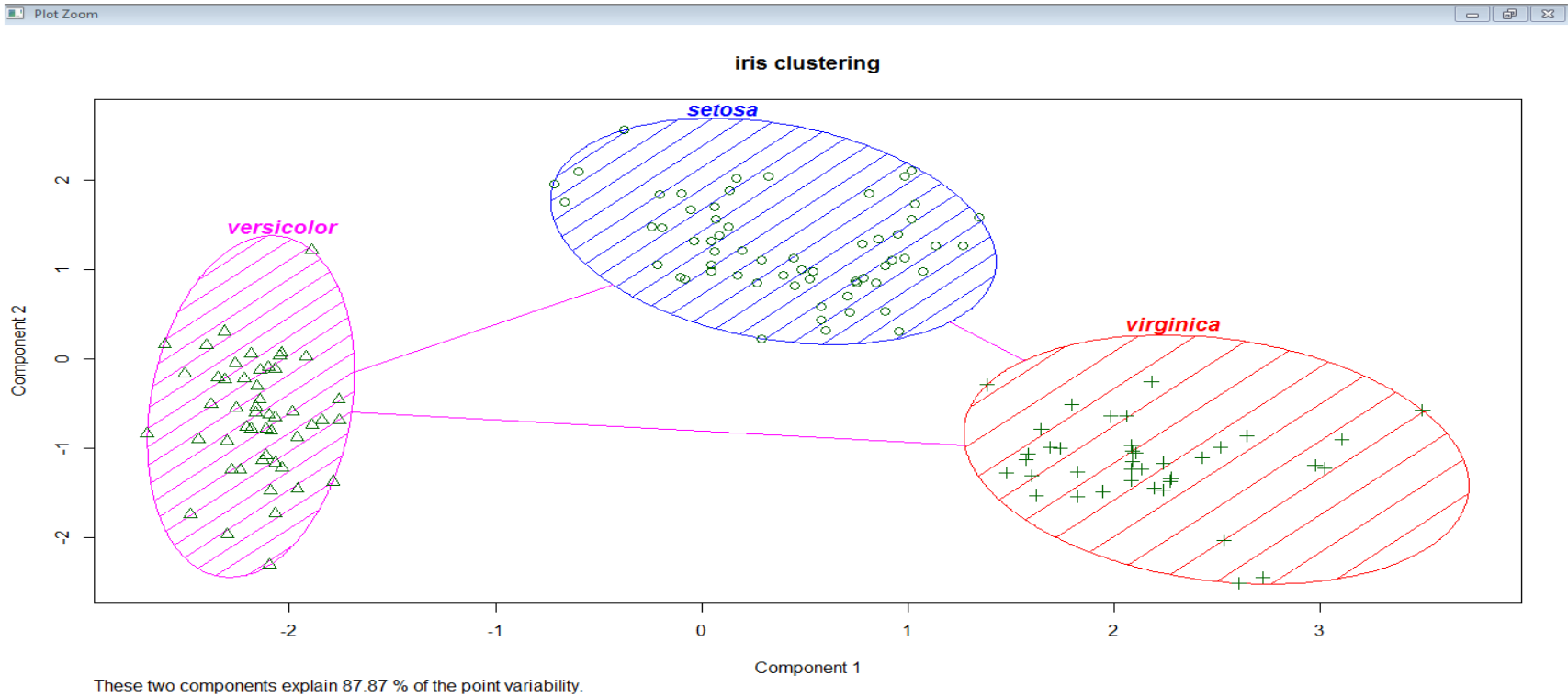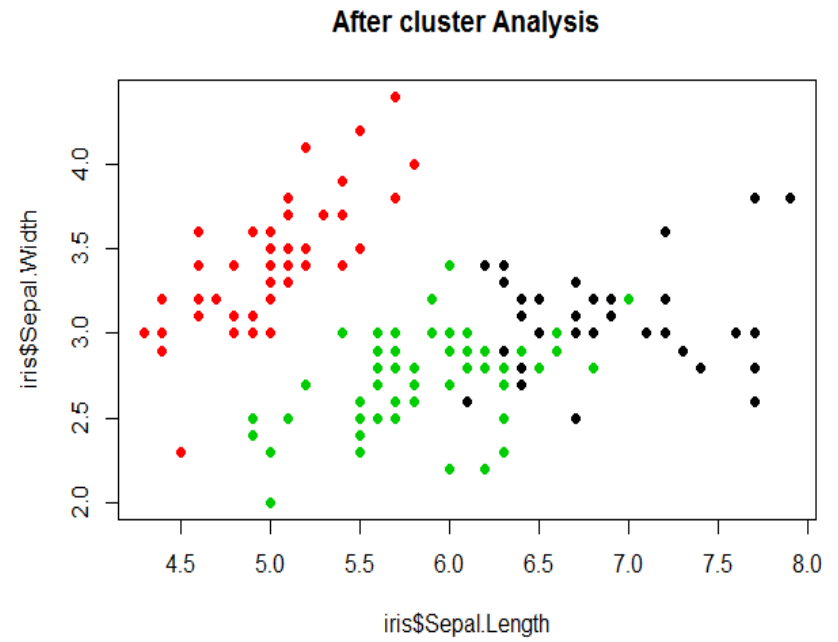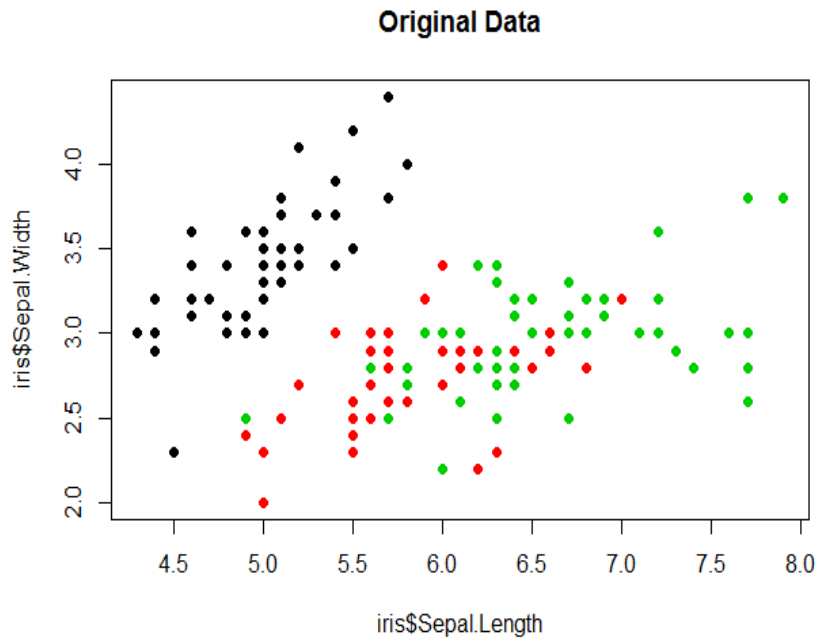
Figure 11.16 Cluster analysis.

Figure 11.17 Comparison of sepal of original and data after cluster analysis.
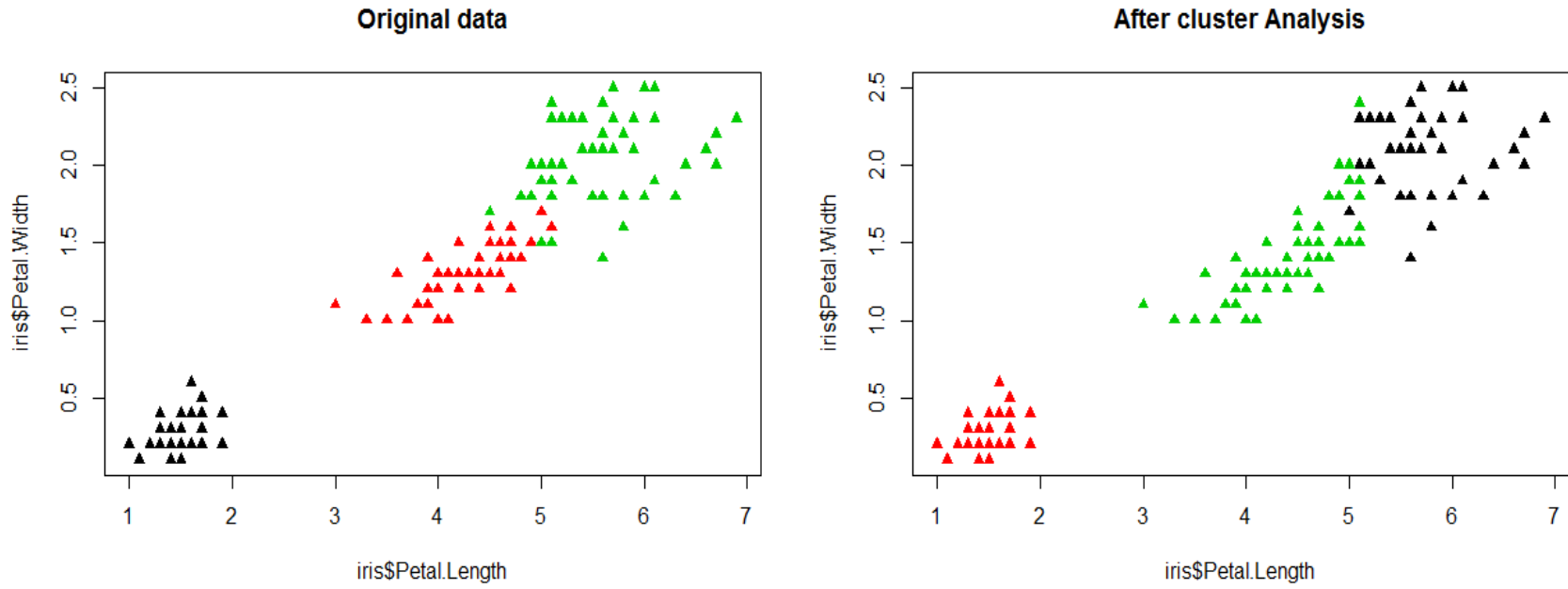
Figure 11.18 Comparison of petal of original and data after cluster analysis.
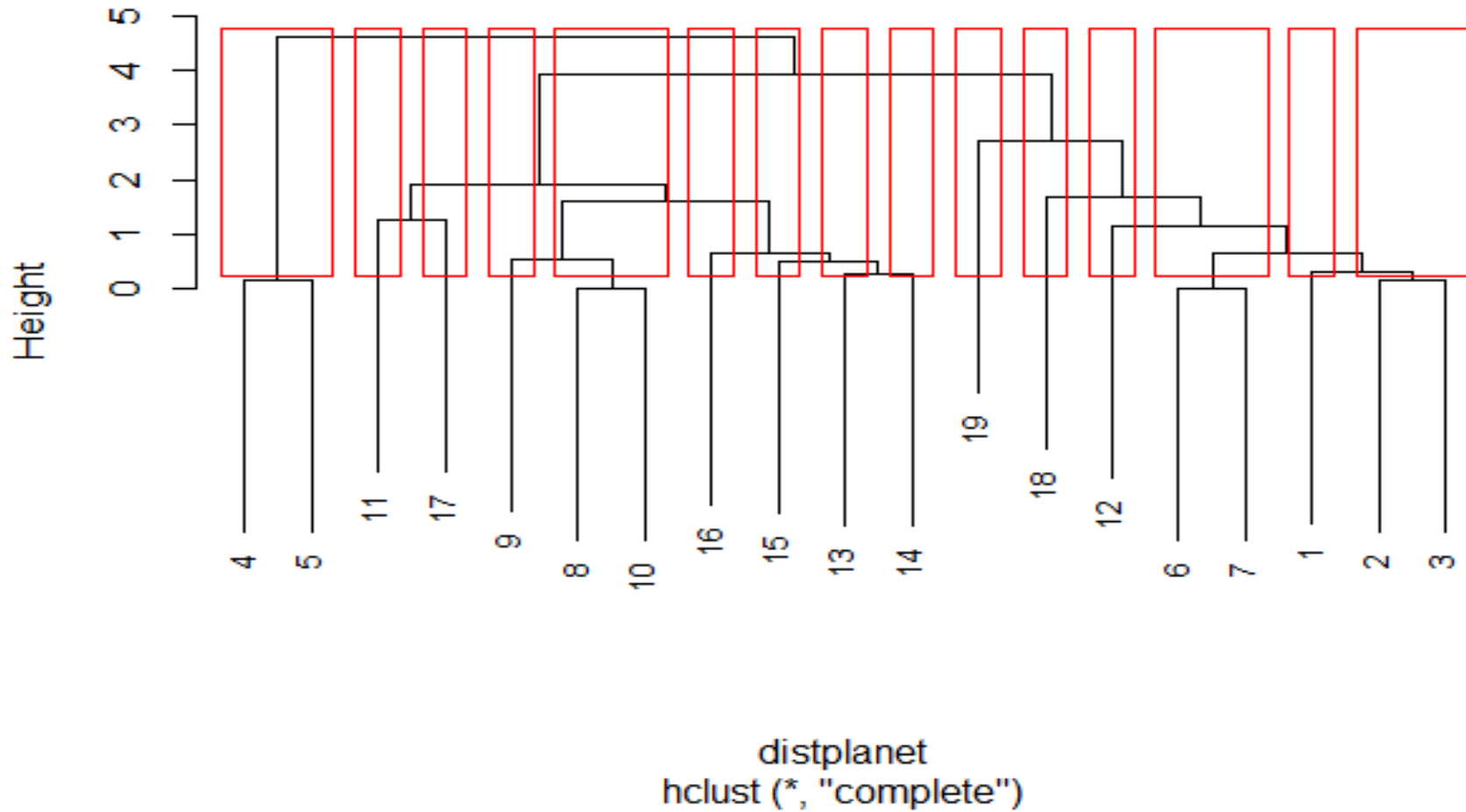
Figure 11.20 Hierarchical clustering for "planet" dataset.
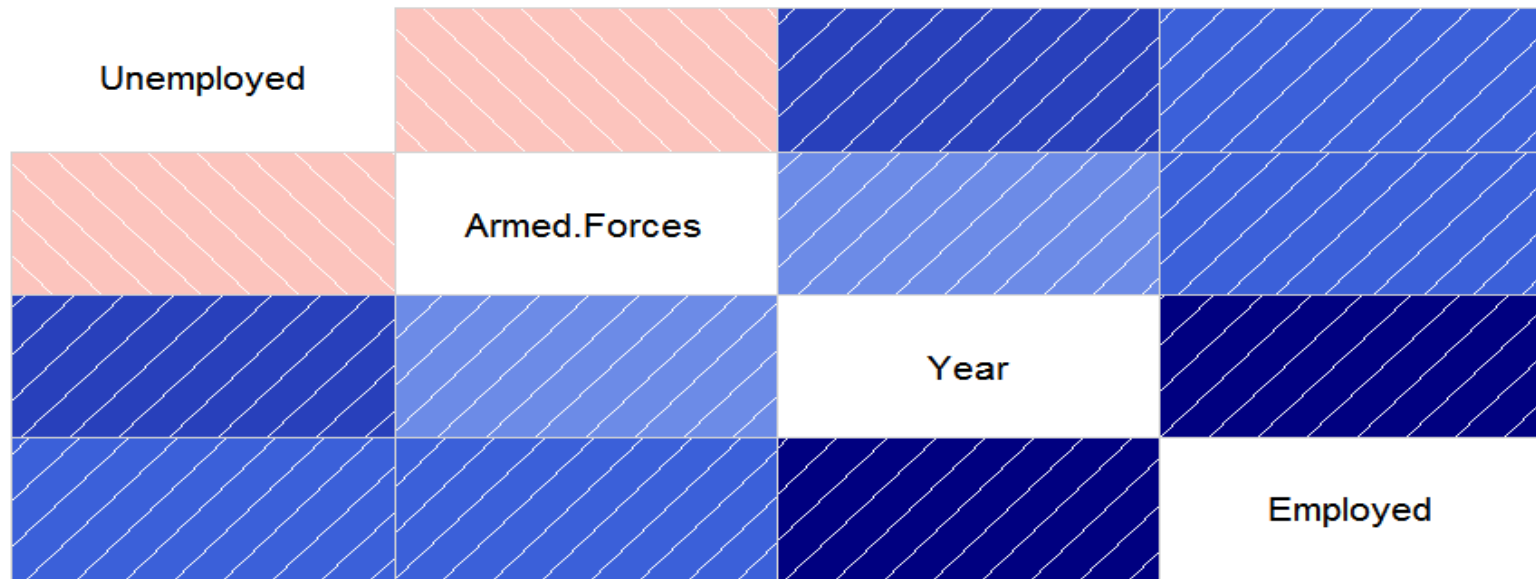
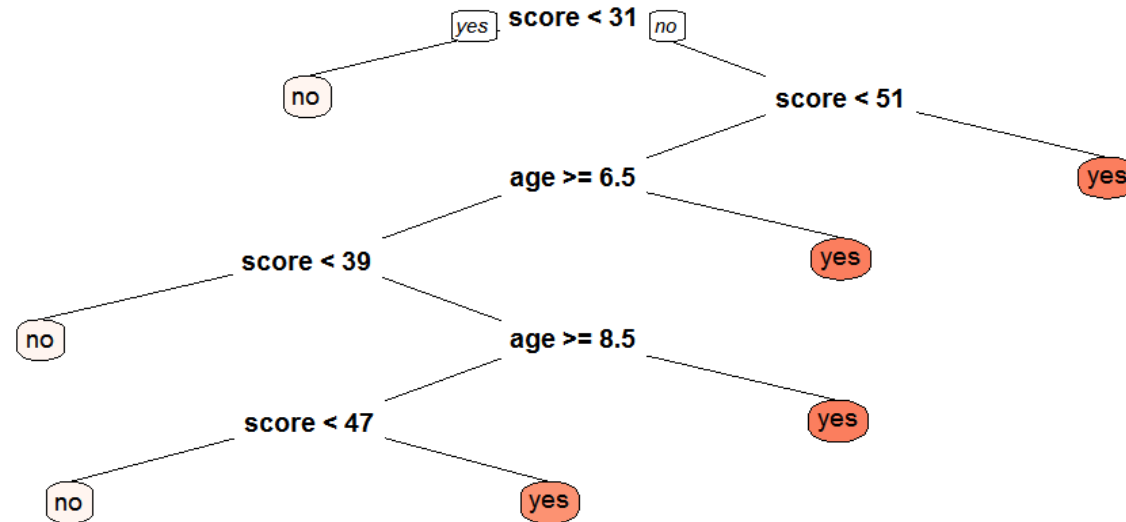**Figure 12.6** Correlation between variables of "longley" dataset.

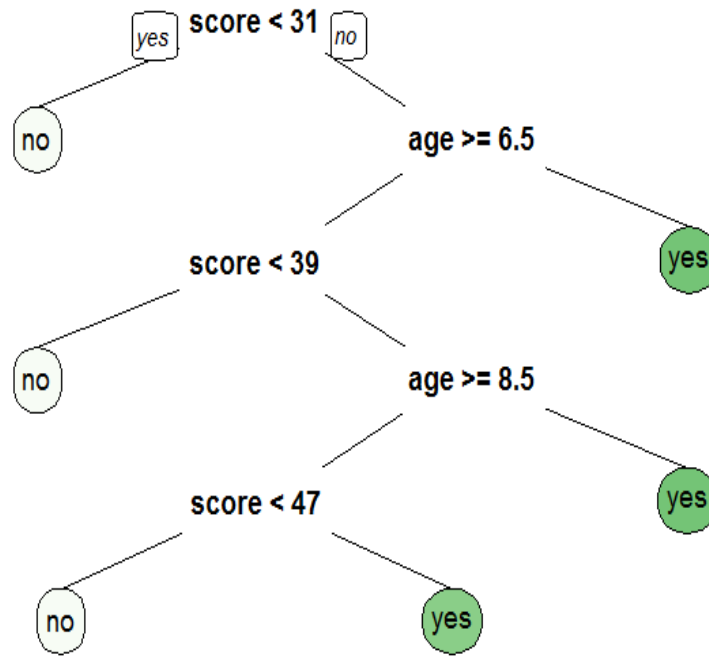**Figure 13.4** Decision tree of training dataset of "readingSkills".

Figure 13.5 Using Gini index for prediction of the model.

CHAPTER 14

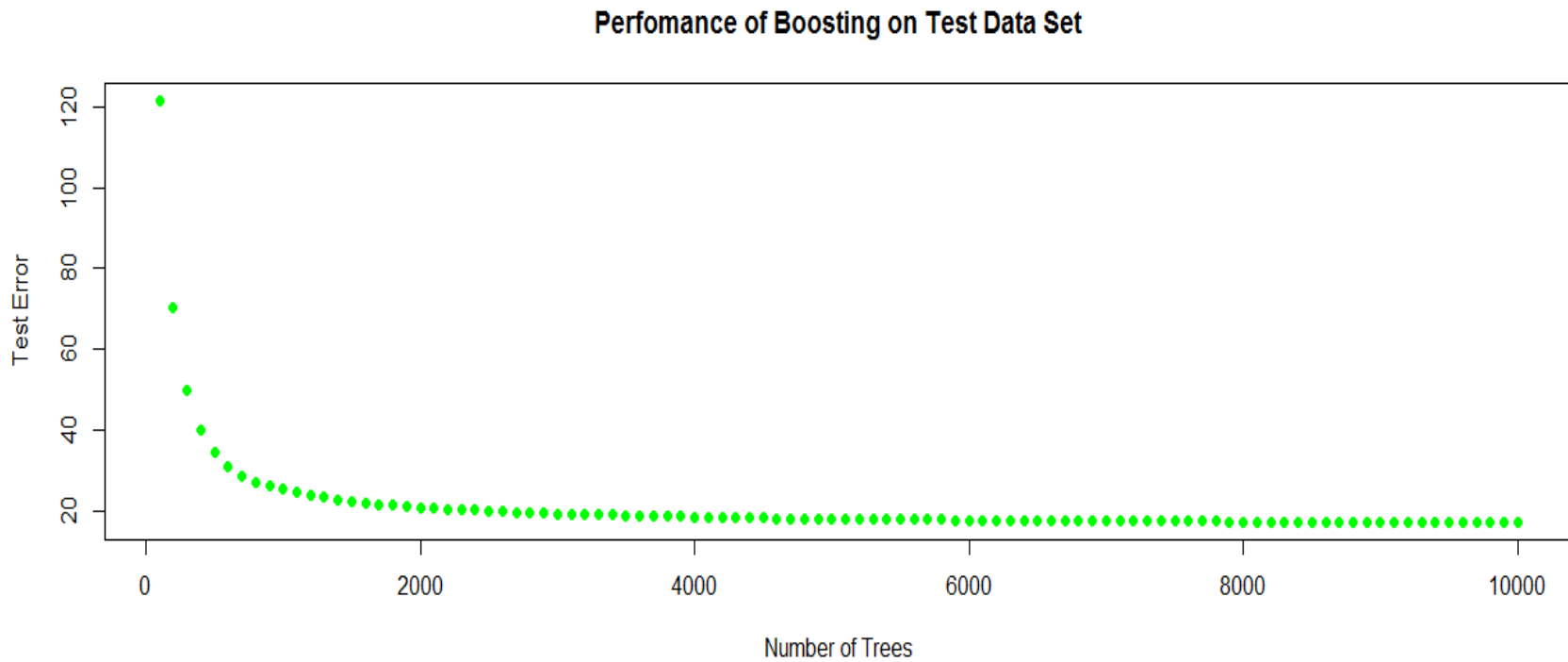# Supervised Machine Learning Ensemble Techniques

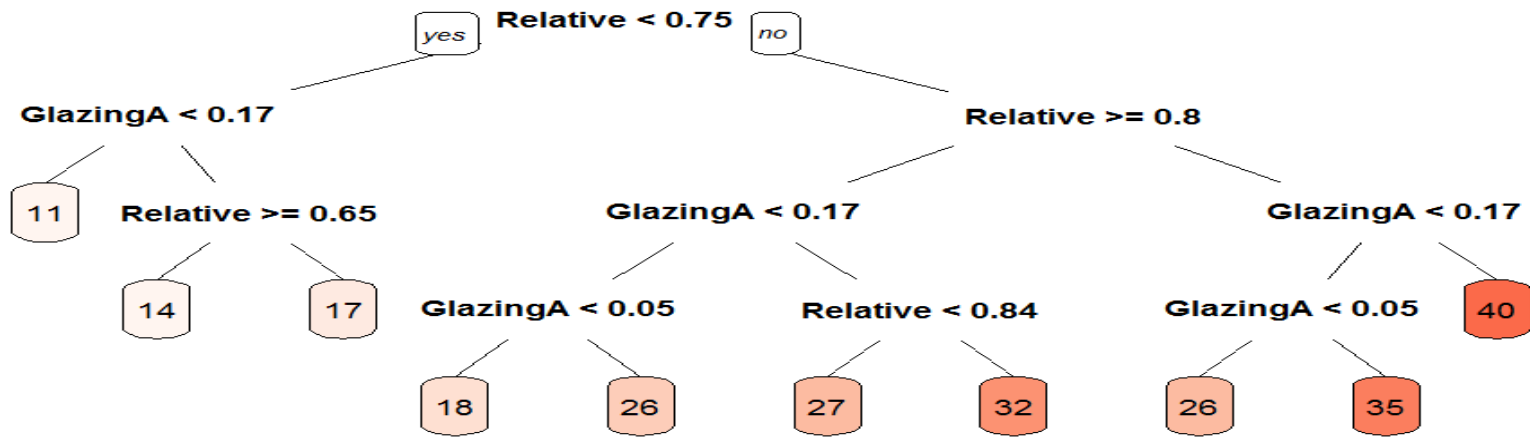**Figure 14.7** Relationship between number of trees and error value.

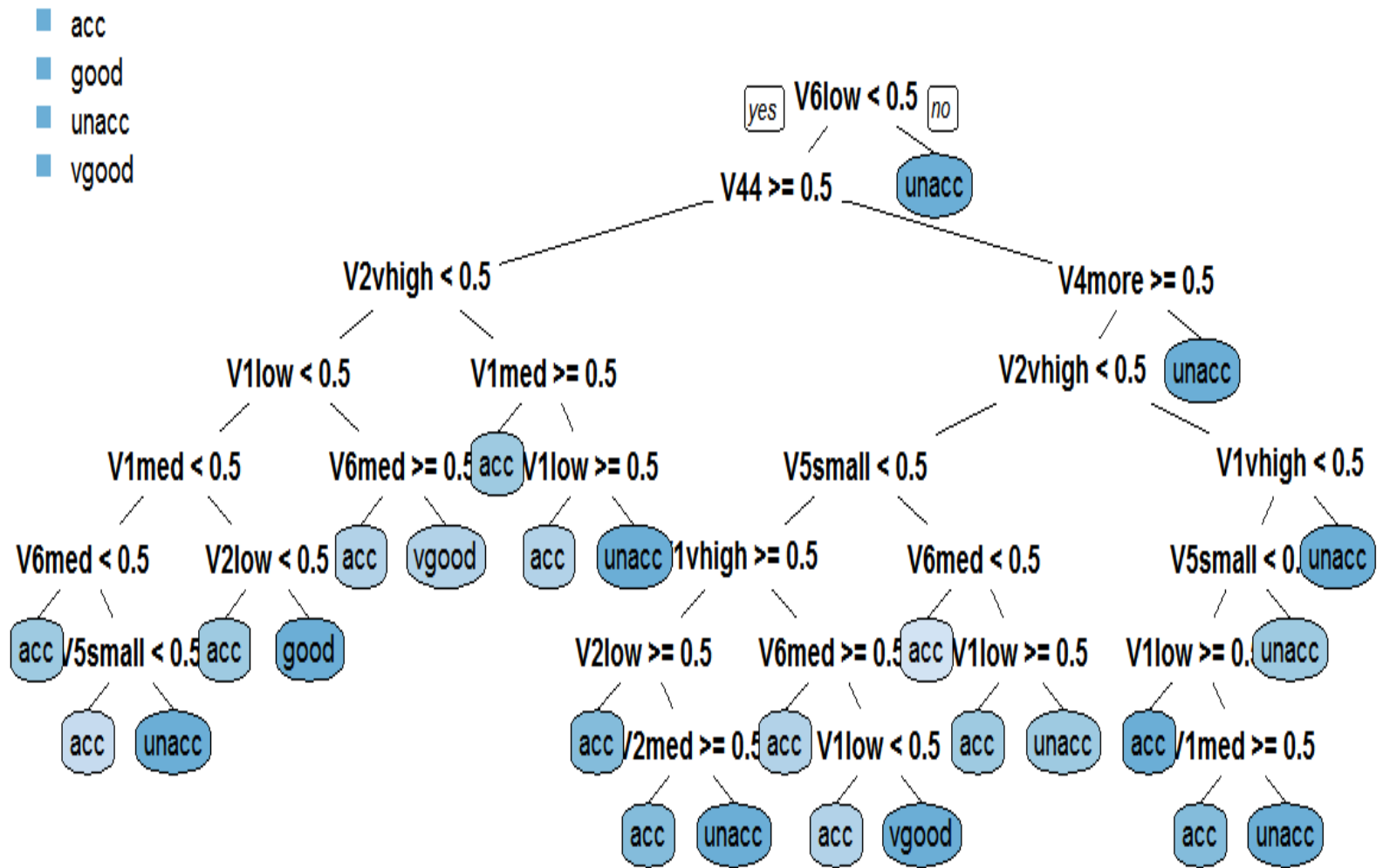**Figure 14.12** Decision tree for "Energy" dataset.

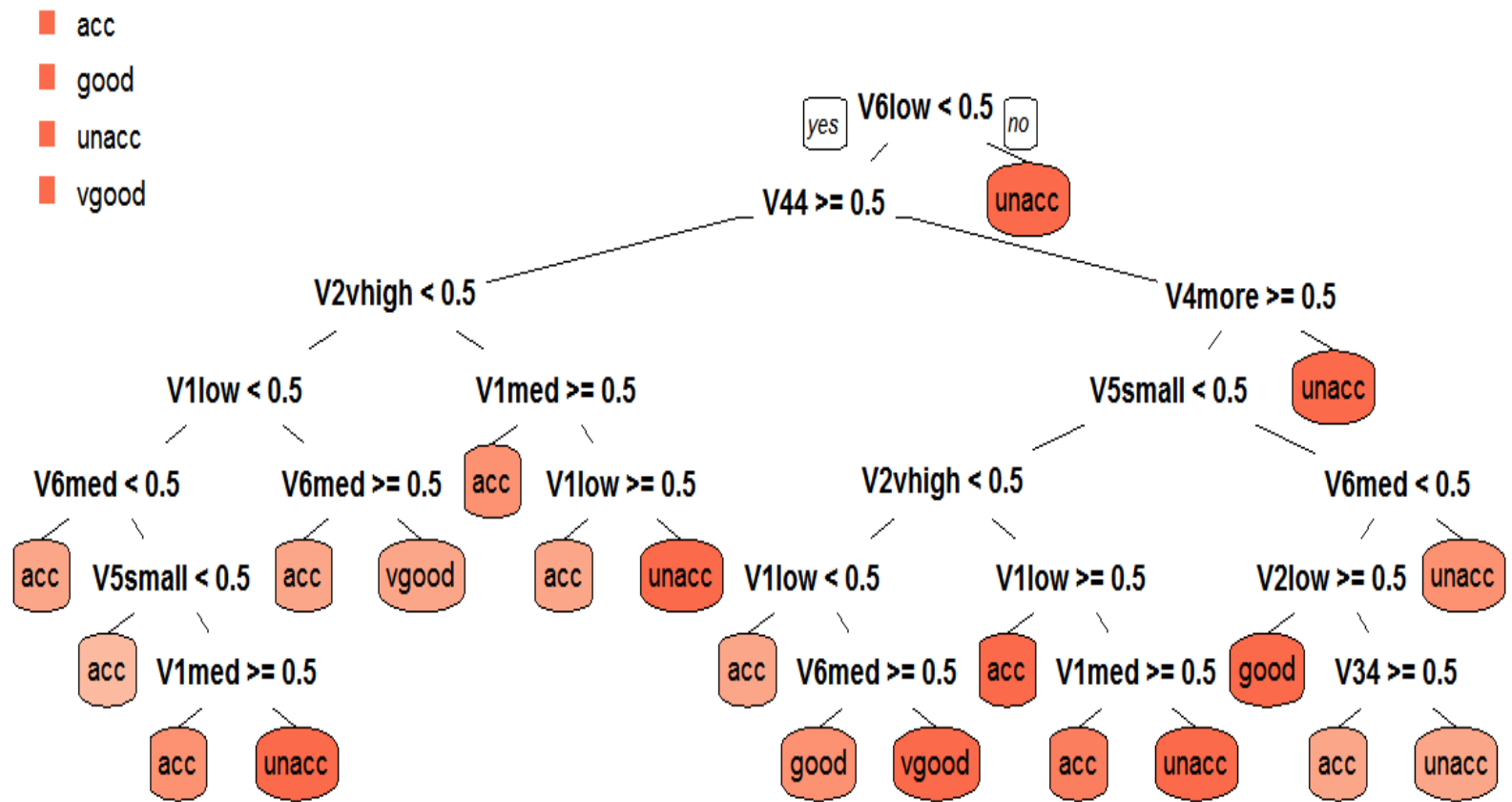**Figure 14.13** Decision tree for "car" dataset with class as criterion.

Figure 14.14 Decision tree for "car" dataset with chi-square index.

Chapter 15
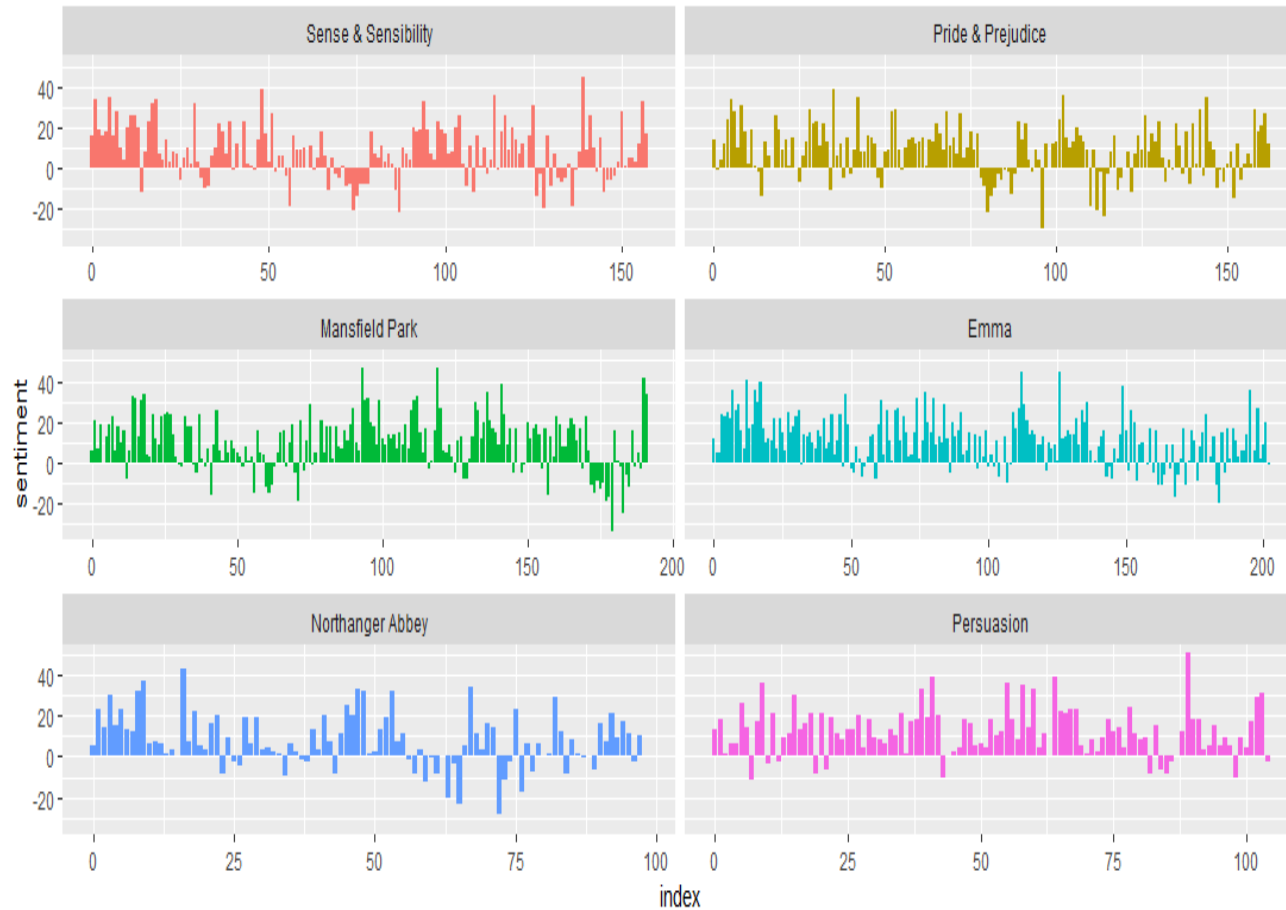# Machine Learning for Text Data

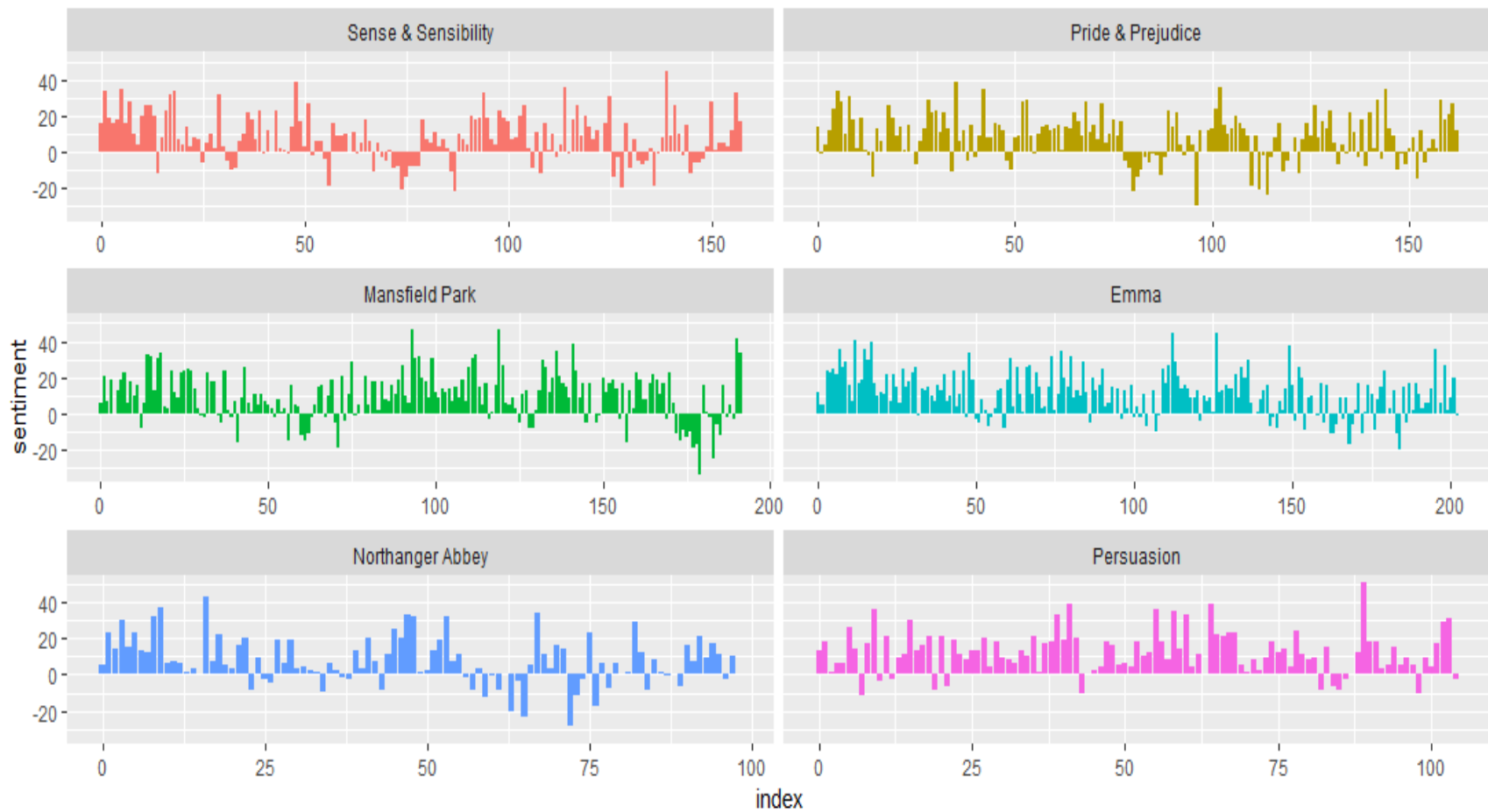**Figure 15.4** Sadness sentiment through the 6 books by Jane Austen.

Figure 15.6 Joy sentiment through the 6 books by Jane Austen.