Practical DataScience
(Data Visualization)
Assignment-3

Load the data set SleepStudy from package Lock5Data for Problems 1 and 2.

Problem 1: Examine the variables Stress and AlcoholUse.

a. Make a table of the variables Stress and AlcoholUse.

b. What proportion of students are in each alcohol use category?

c. What proportion of students in the high stress group report high alcohol use?

d. Display the data from this table in a bar graph that effectively compares the distribution of alcohol use for each stress group. Write the R code you used to create this graph.

e. Describe the patterns you see in the data.

Problem 2: Examine the variables Drinks and LarkOwl.

a. Find the mean, median, and standard deviation of the number of alcoholic drinks per week for the entire group of students.

b. Do the same as (a), but separately for each group of students that classify themselves as early risers (larks), night owls, or neither.

c. Create an effective display of the Drinks variable that shows how the distribution of number of drinks per week varies among larks, night owls, and neither.

d. Describe the patterns you see in the data.

Problem 3: You can find the datasets required for this problem at algorithmica repository.

Dataset 1 "salespeople.txt"contains how many cases of Teatime Chocolate Biscuits were sold by each of three salespeople in each of three months.

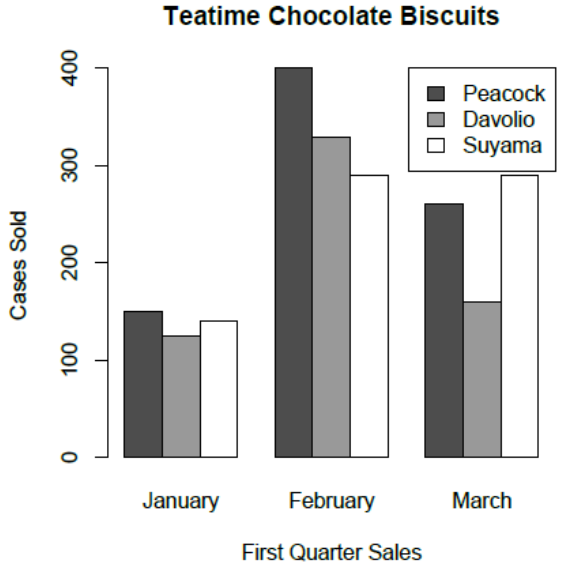
Dataset 2 "Vegetation2.txt" contains the vegetation species richness observed with particular environmental factors in the Yellowstone National Park.

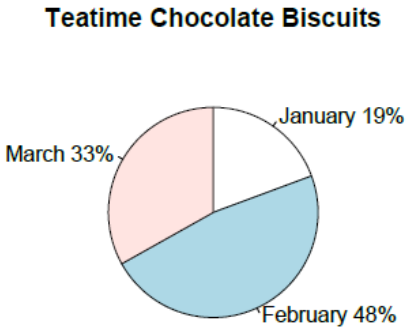
Practical DataScience
(Data Visualization)
Assignment-3

Dataset 3 "CodParasite.txt" contains the presence or absence of the parasite in fish as well as the number of parasites per fish, and the attributes of the host fish.

Write R code to make the following graphs:

a. (Dataset 1) Create a barchart that compares how much each person sold each month. A sample graph is shown below(left).

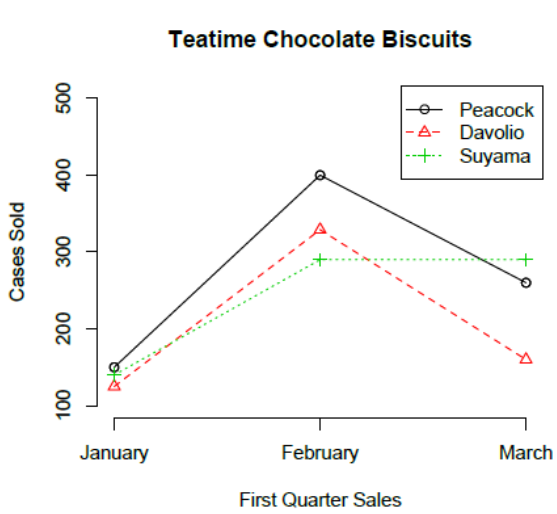


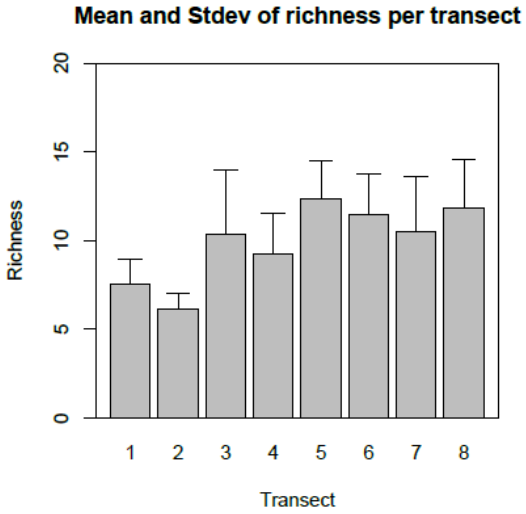


b. (Dataset 1) Make a pie chart showing the big picture of how each month's sales Figure contributes to the quarterly total. A sample graph is shown above(right).

c. (Dataset 1) Make a graph showing the changes and trends in sales over time For each salesperson, month by month. A sample graph is shown below(left).

Practical DataScience
(Data Visualization)
Assignment-3

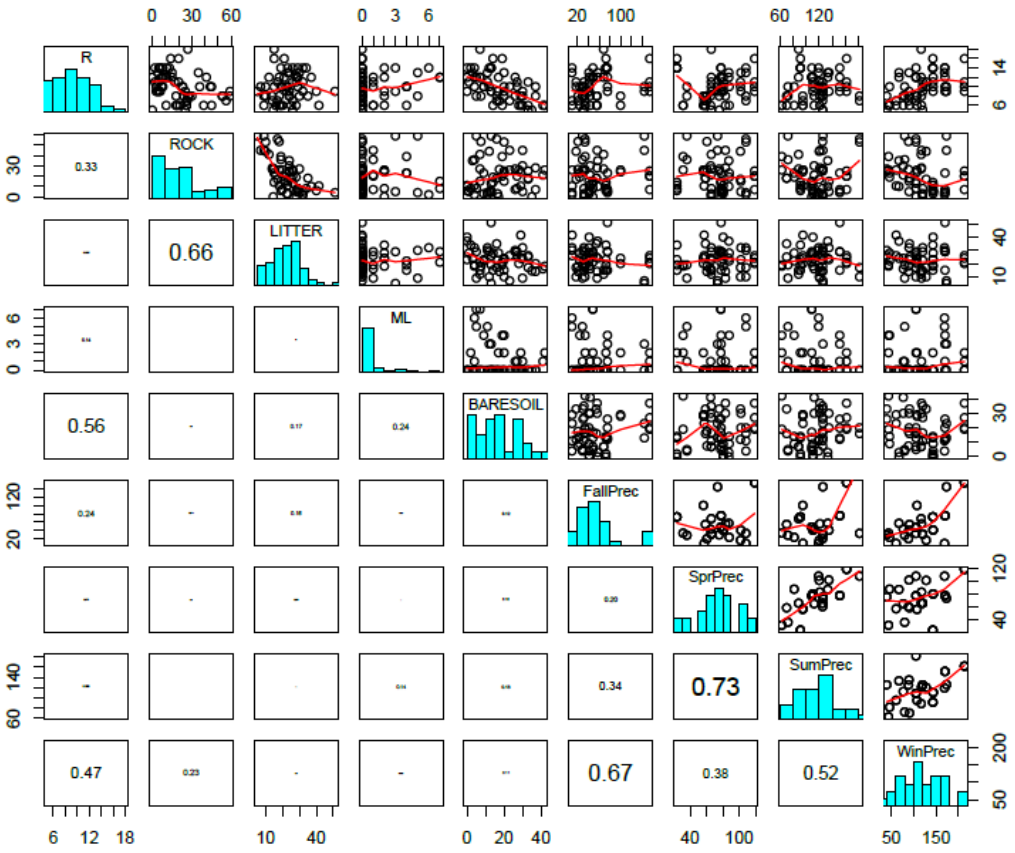




d. (Dataset 2) Make a bar chart showing the mean species richness per transect, and Add a vertical line for the standard deviation of richness per transect. A sample graph is shown above(right).

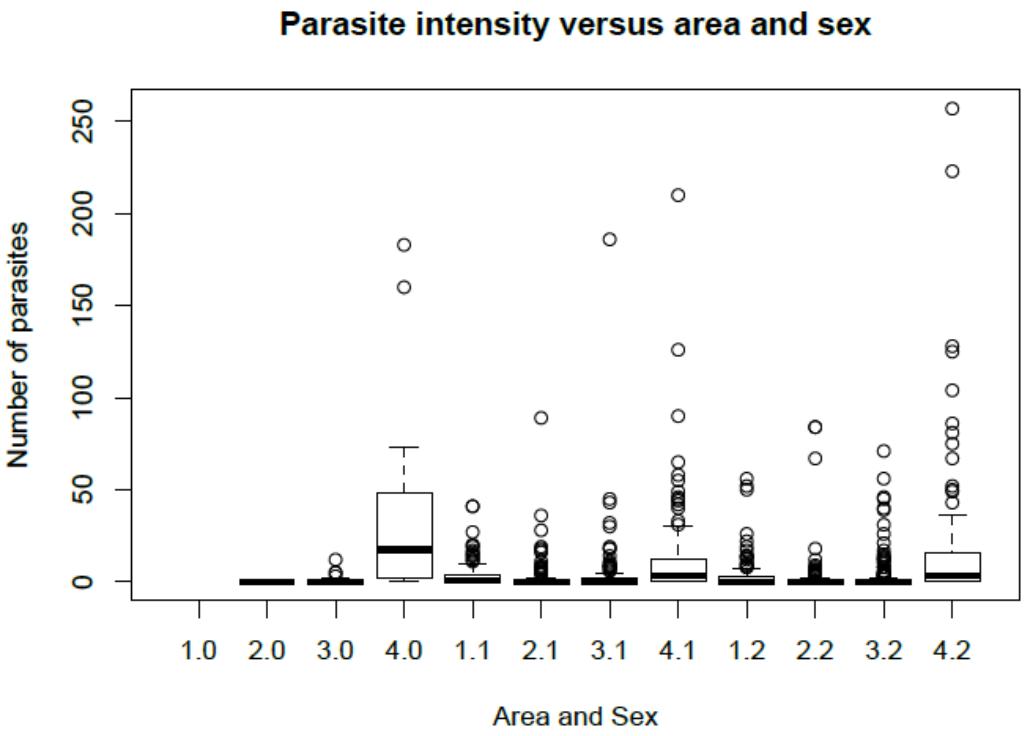
e. (Dataset 2) Make a pairplot for the selected variables (from R to WinPrec in 5th To 13th columns) in the vegetation data. Add correlation coefficients in the lower panels. A sample graph is shown below.

Practical DataScience
(Data Visualization)
Assignment-3



f. (Dataset 3) Make a boxplot of the number of parasites (Intensity) conditional on Area and sex. A sample graph is shown below.

Practical DataScience
(Data Visualization)
Assignment-3



g. (Dataset 3) Make a Cleveland dotplot for the number of parasites (Intensity), and Group the observations by area. A sample graph is shown below.

Practical DataScience
(Data Visualization)
Assignment-3