3 Data Visualization

For all the figures, you must print the axis labels, the figure title and the legend (if many curves).

Exercise 3.1
You must install and use the library "ggplot2".

1. Load Iris dataset with command : data(iris)
2. Create a pie chart to plot Iris species (command "pie").
3. Create a barplot to plot Iris species (command "barplot").
4. Show the histograms of all Iris attributes (command "hist"). You must all the histograms on a same figure.
5. Plot the ecdf of Petal length and Setal width on a same figure (command "ecdf"). Each ecdf must be plotted with 100 points on the x-axis.
6. Plot the percentiles between 0 and 1 with step 0.1 of Sepal width, Sepal length, Petal width, and Petal length (command "quantile"). All the graphs must be on the same figure.
7. Plot on the same figure the boxplots of Sepal width, Sepal length, Petal width, and Petal length (command "boxplot").

Useful R commands : table, paste, par

Exercise 3.2

1. Load Iris dataset with command : data(iris)
2. Plot a common scatter plot (4×4 plots on the same figure) of Sepal width, Sepal length, Petal width, and Petal length (command "pairs"). The samples should appear like full blue circles on the scatter plot.
3. Plot the correlation matrix of Sepal width, Sepal length, Petal width, and Petal length (command "corrgram"). The package "corrgram" should be installed.
4. Plot the correlation matrix by using the "corrplot" package (command "cor" and "corrplot").
5. Plot the parallel coordinates of Sepal width, Sepal length, Petal width, and Petal length (command "parcoord" with library "MASS").
6. Plot the 2D histogram of Petal width, and Petal length (command "hexbinplot" in package "hexbin"). What is the connection between the 1D histogram and the 2D histogram ?
7. Draw a star plot of the multivariate data set including Sepal width, Sepal length, Petal width, and Petal length (command "stars"). Precise the legend of the star within the figure.

Exercise 3.3

1. Load the dataset Motor Trend Car Road Tests : data(mtcars). Describe briefly the dataset.
2. Plot a heat map of the data "mtcars" (command "heatmap" in library "stats"). Do not forget to scale the columns to have a balanced color representation. What is a heat map ?
   Useful R command : as.matrix

3. Create a 3D scatter plot of the variable "mpg" as a function of "wt" and "disp". This figure will show the linear regression between the variables. What is a 3D scatter plot ? The command is "scatter3d" and you should install the "car" package with all dependencies. You should scale the axes and plot the residuals.

Exercise 3.4

1. Create a Twitter application by following the steps described in "How to Register a Twitter App in 8 Easy Steps.pdf". If necessary, you can create a "temporary" account just for this course.
2. Connect R to your Twitter account. A R code is available on the class website.
3. Use R to collect $n = 100$ tweets with the keyword "#paris".
4. What is the data contained in a single tweet ?
5. Store the tweets into a data frame and save the data frame into a ".csv" file named "tweetParis.csv".
6. Load the tweets by reading the ".csv" file.