Kursus Bigdata F2016

Problem Week 3

Prepare an R script file generating and sometimes saving plots of a selection of datatypes and examples as detailed in the following:

Q1. On the basis of the Example 4.17 in the [R script A], the following extensions of the already demonstrated clusters are implemented:

- insert a 3D random uniform point cube with dimensions 1 x 1 x 1 and center positioned in the coordinates x=-1, y=-1, z=-1.
- insert a 3D random normal point set with center x=0, y=0, z=4 and standard deviations sdx=0.1, sdy=0.1 and sdz=0.1
- insert a 2D 1 x 1 square of uniformly distributed observations with center in x=-1, y=-1, z=-1 and which is parallel to the xy plane.

Verify by visualization all the 3D clusters in the extended Example 4.17.

Q2. E.g. using the examples presented, prepare an R script which visualizes the Central Limit Theorem (CLT) in one plot with 2 rows of subplots. The upper row contains the visualization using the uniform density function with the following number of variables in the resulting sum: 1, 10, 50, 100.
The lower row contains the same for the binomial distribution.
The plot is saved in a pdf file with the name CLT_1.pdf in the RStudio working directory.

Q3. Using the examples presented, prepare an R script which generates a scatter plot of the dataset mtcars variables cyl and mpg. The plot should include a labeling of each data point with the appropriate car name. The plot is saved as a jpg file with name mtcars_1.jpg in the RStudio working directory.

Q4. Create one or more dataset from your preferred application domain(s) e.g. using the examples demonstrated earlier, and visualize/explore the dataset using the function plot3d().

Q5. Create a script which runs the [Kabacoff, 2015] p. 129 “Example comparing groups by using parallel Box plot”. Use this script on a different dataset than mtcars, preferably from your own application domain(s).

Q6. Create a script which runs the [Kabacoff, 2015] p. 134 “Example on plotting a large number of labeled values on a simple horizon scale, sorted and colored”. Use this script on a different dataset than mtcars, preferably from your own application domain(s).

Q7. Only if time permits (non mandatory). Create a spatial function with coordinates x(t), y(t), z(t), and where x(t), y(t) and z(t) are appropriately selected products of e.g. sin() and cos() functions, depending of one discretized parameter t. Explore the function created using plot3d.

The resulting R code is inserted into a Problem3_xxx.R script file where xxx are characters chosen from the persons name. Each participant keeps the script for later submission.

Course material

[Kabacoff, 2015] Robert I. Kabacoff, "R in Action", 2'Ed, Manning Publications, 2015.
[R script A] 4_R_Intro_Visualization.R 2017.02.16/jaas