Visualization recap - R + Python

Project no. 1 [10 p.]

Data

The dataset consist of characteristics of red and white variants of the Portuguese "Vinho Verde" wine[1].

For more details, consult [P. Cortez et al., Modeling wine preferences by data mining from physicochemical properties, Decision Support Systems 47(4), 2009].

Exercise in R

Exercise 1.1.

Explore data by creating plots / charts / graphs suitable for:

- investigating the distribution of `alcohol` variable [0.5 p.];
- comparing the distribution of `alcohol` variable between two types of wine i.e. `red` and `white` [0.5 p.];
- comparing the distribution of `alcohol` variable in each of possible quality group defined by `response` variable [0.5 p.];
- percentage of `red` and `white` wines within each quality group [0.5 p.];
- investigating the relationship between variables describing acidity of the wines [0.5 p.].

Consider functions from `ggplot2` and / or `graphics` package. Create short report in `knitr` + `Markdown` such that plots and tables are created within `R` chunks (explicitly or via `source()` function and `R` scripts) [1.5 p.].

Exercise in Python

Exercise 1.2.

Explore data by creating plots / charts / graphs suitable for:

- investigating the distribution of `alcohol` variable [0.5 p.];
- comparing the distribution of `alcohol` variable between two types of wine i.e. `red` and `white` [0.5 p.];
- comparing the distribution of `alcohol` variable in each of possible quality group defined by `response` variable [0.5 p.];
- percentage of `red` and `white` wines within each quality group [0.5 p.];
- investigating the relationship between variables describing acidity of the wines [0.5 p.].

Consider functions and methods from `matplotlib`, `seaborn` or other `Python` packages. Create short report with `Jupyter` notebook and `Markdown` [1.5 p.].

In your report include a discussion of all the charts you have chosen and their alternatives [2 p.].

[1] http://archive.ics.uci.edu/ml/datasets/Wine+Quality