Introduction to R

Statement

Datasets:

- usedcars.csv contains actual data about used cars advertised for sale on a popular U.S. website. Source: B. Lantz (2015). *Machine Learning with R. Second edition*. PACKT.

Practice:

Complete the following tasks to begin learning R (Hint: In addition to the command line, you can open a new R script file and write and execute code (Ctrl+Enter) from the file):

1. Let us begin with the basics. Create a vector with values 2002, 2004, 2006, 2008 using `c` and `seq` functions.
2. Use function `length` to get the size of the vector.
3. Try the different methods for selecting the elements of the vector.
4. Load the data set **usedcars.csv** into a variable named *fdata*.
5. Use `str` and `summary` functions on *fdata*. What types of variables are in the dataset? What are the average values of the numeric variables?
6. Use `View` and `head` functions on *fdata.*
7. Access the elements number 5 to 20 of variable *color*.
8. Create a new dataset removing row numbers 10 and 100.
9. Create a new dataset only with columns *year*, *price* and *mileage*.
10. Obtain statistics for variables *year* and *price*.
11. Use function by() to calculate statistics filtering by classes.
12. Filter from this dataset the rows that have a year that matches the values of the vector created in step 1.
13. Create a new column in the dataset named PM resulting from multiplying the values of *price* and *mileage* in each row.
14. Plot the price values with a solid line.
15. Plot a scatterplot between variables *mileage* (x axis) and *price* (y axis).
16. Plot a boxplot of mileage values.
17. Plot a histogram of the *prices* data.