Data Exploration and Visualization

Laboratory - 2

Date - 16 March 2021 (Tuesday)

Conducted By -

Abhishek Agarwal

Research Scholar

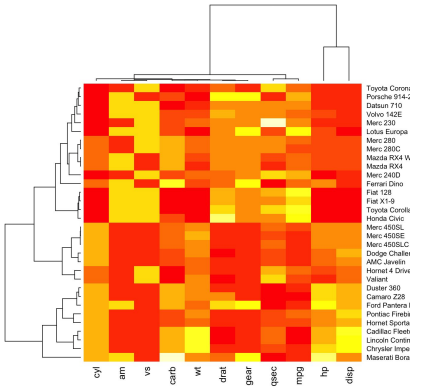
Doctoral School 3

Warsaw Technical University

Heat Map

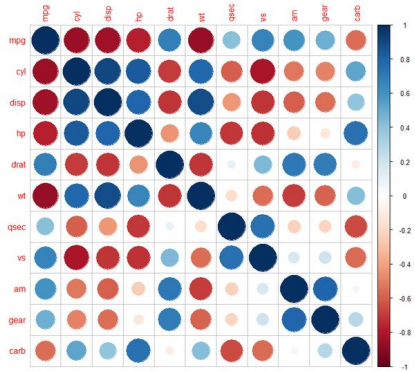
- The `heatmap()` function is natively provided in R. I
- It produces high quality matrix and offers statistical tools to normalize input data, run clustering algorithm and visualize the result with dendrograms
- Note that it takes as input a matrix. If you have a data frame, you can convert it to a matrix with `as.matrix()`, but you need numeric variables only.
- How to read it: each column is a variable. Each observation is a row. Each square is a value, the closer to yellow the higher. You can transpose the matrix with `t(data)` to swap X and Y axis.
- For Detail Study of the Function -

  https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/heatmap



Correlogram

- Correlogram is a graph of correlation matrix.
- It is very useful to highlight the most correlated variables in a data table. In this plot, correlation coefficients is colored according to the value.
- Correlation matrix can be also reordered according to the degree of association between variables.
- CORRPLOT PACKAGE - The corrplot package is a graphical display of a correlation matrix, confidence interval. It also contains some algorithms to do matrix reordering
- CORRGRAM PACKAGE - The corrgram package allows to build correlogram. The output allows to check the relationship between each pair of a set of numeric variable.
    - Relationship can be visualized with different methods:
        - panel.ellipse to display ellipses
        - panel.shade for coloured squares
        - panel.pie for pie charts
        - panel.pts for scatter plots
- Lesser known ways to build correlogram with R, like the ellipse package, ggcorr package.



3d Graphs

- **plot3D** s an R package containing many functions for 2D and 3D plotting: *scatter3D*, *points3D*, *lines3D*,*text3D*, *ribbon3d*, *hist3D*, etc.
- In addition to the x, y (and z) values, an additional data dimension can be represented by a color variable (argument colvar).
- A color key (argument colkey) can be written next to the figure. It is possible to log-transform the color key, rescale it, adjust its position
- The resolution of a figure can be increased (argument resfac).
- Either the facets can be colored, just the border, or both.
- Functions that are based on the persp function, for visualising 3D data - scatter3D, points3D, lines3D, text3D, ribbon3D, hist3D, ect.
- For Detail Study of the Package -

  https://www.rdocumentation.org/packages/rgl/versions/0.105.22/topics/plot3d