Descriptive_Data_Analysis_with_mtcars

Arif_Furkan

2024-06-29

Before you start; “Zorunlu Paket Yükleniyor” means Installing required package

Data Loading and Preprocessing

```
require(dplyr)
```

```
## Zorunlu paket yükleniyor: dplyr
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
require(ggplot2)
```

```
## Zorunlu paket yükleniyor: ggplot2
```

```
require(GGally)
```

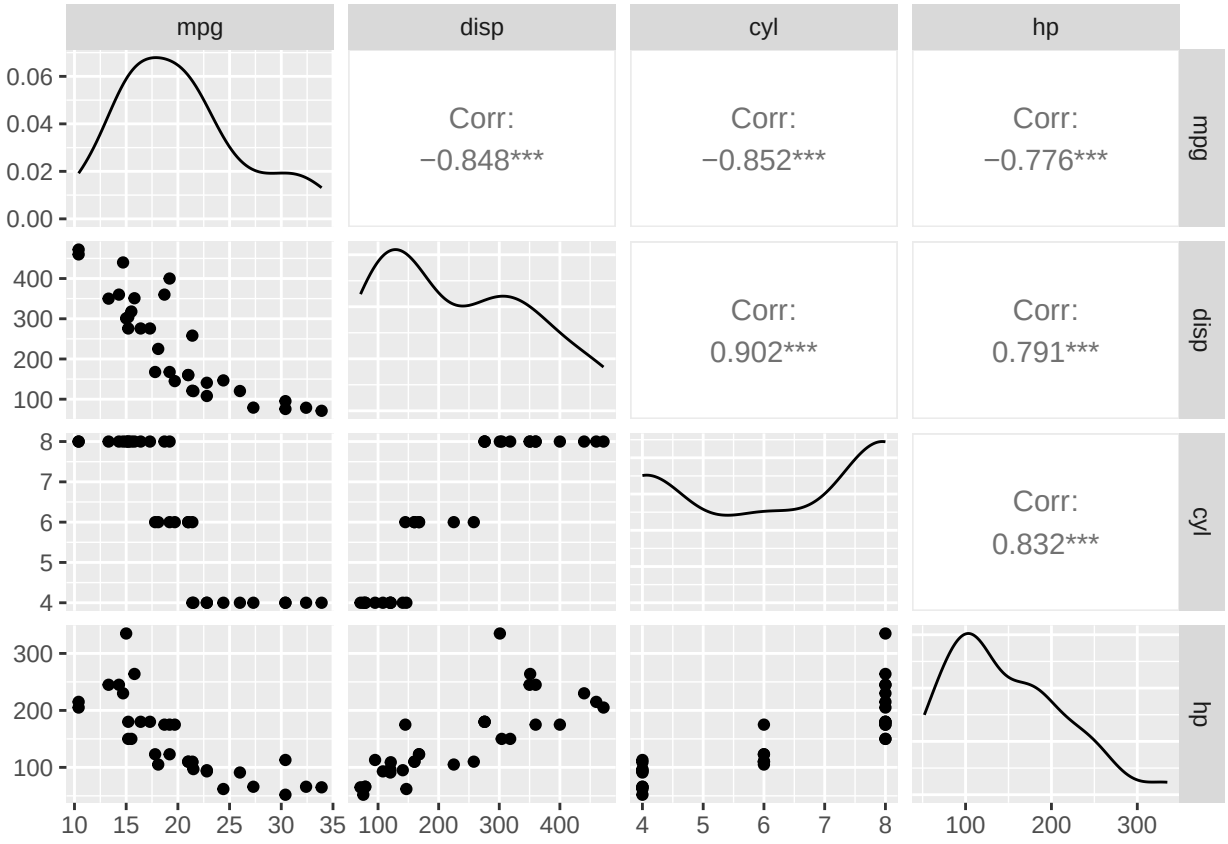
```
## Zorunlu paket yükleniyor: GGally
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
data(mtcars)
```

Data Visualization

```
mtcars %>%
select(mpg, disp, cyl, hp) %>%
ggpairs(.)
```



Installing and Using Other Packages

```
require(gcookbook)
```

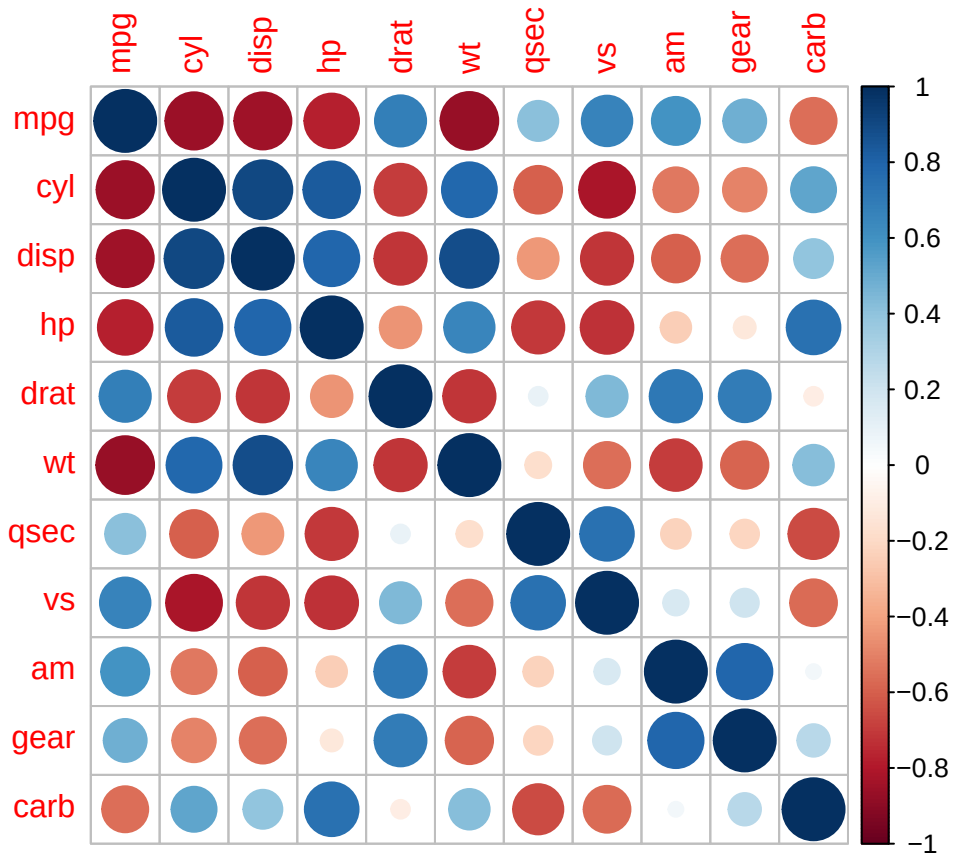
```
## Zorunlu paket yükleniyor: gcookbook
```

```
require(corrplot)
```

```
## Zorunlu paket yükleniyor: corrplot
```

```
## corrplot 0.92 loaded
```

```
mcor <- cor(mtcars)
corrplot(mcor)
```



Other Analytics and Visualizations

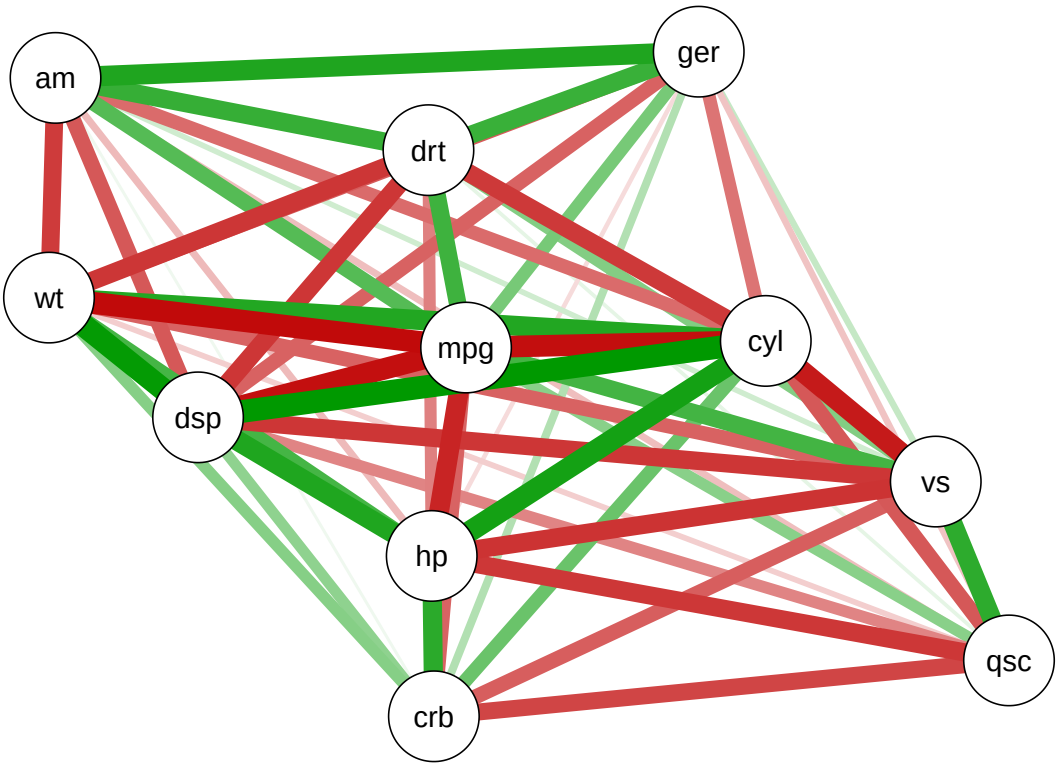
Graphs are created with the qgraph package.

Multivariate analyzes are performed with the MVA package.

```
require(qgraph)
```

```
## Zorunlu paket yükleniyor: qgraph
```

```
require(corrplot)
qgraph(mcor,layout = "spring")
```




```
require(MVA)
```

```
## Zorunlu paket yükleniyor: MVA
```

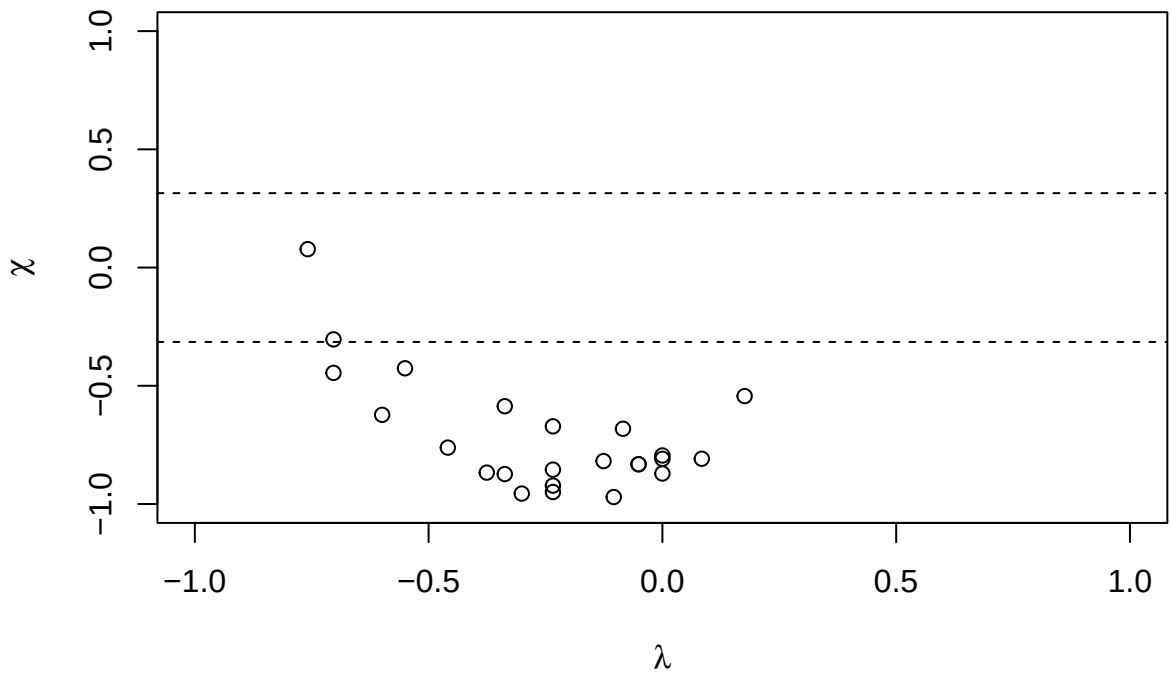
```
## Zorunlu paket yükleniyor: HSAUR2
```

```
## Zorunlu paket yükleniyor: tools
```

```
install.packages("MVA")
```

```
## Warning: package 'MVA' is in use and will not be installed
```

```r
with(mtcars, chiplot(mpg,hp))
```




```r
stem(mtcars$hp)
```

```
##
##   The decimal point is 2 digit(s) to the right of the |
##
##   0 | 5677799
##   1 | 0011111122
##   1 | 55888888
##   2 | 123
##   2 | 556
##   3 | 4
```

Operations are performed on outliers with the OutliersO3 package.

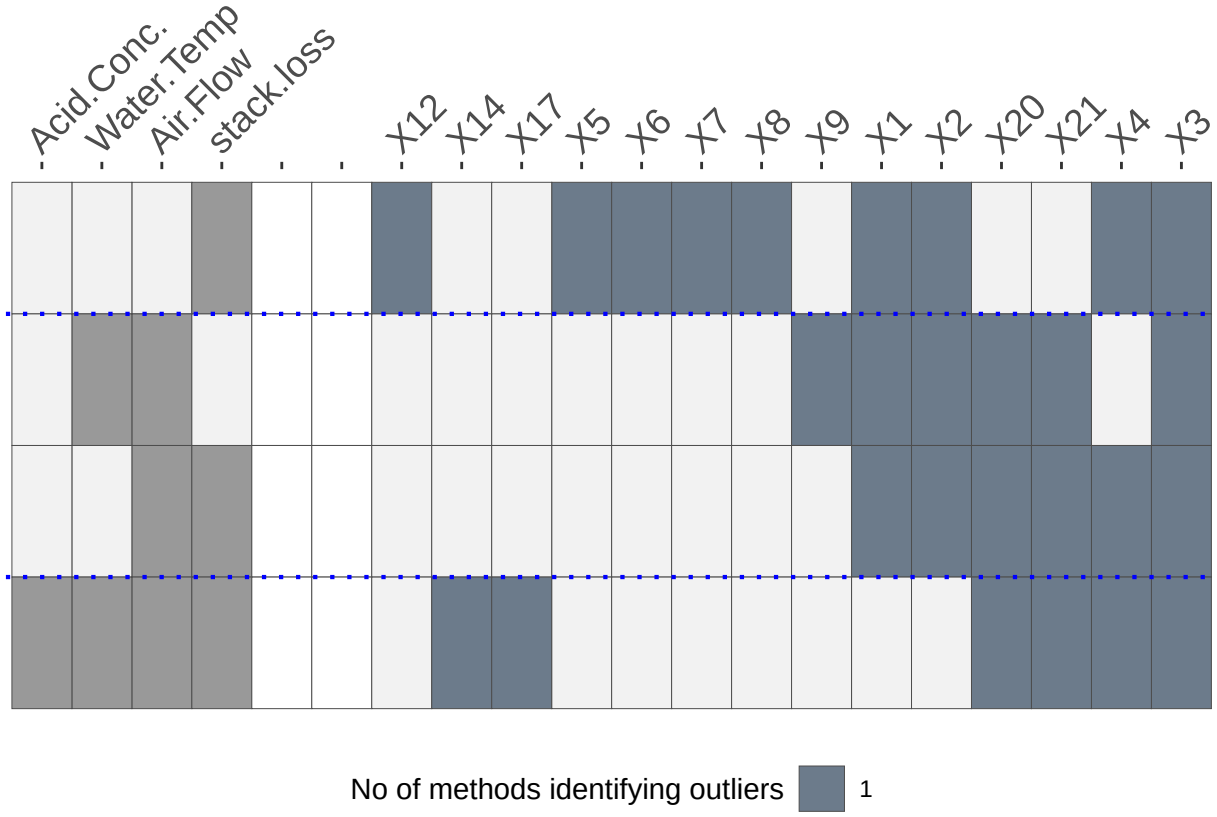
```r
require(OutliersO3)
```

```
## Zorunlu paket yükleniyor: OutliersO3
```

```
data("stackloss")
outdata <- O3prep(stackloss,method =c("HDo", "BAC", "DDC"), tols=0.05)
outmulti <- O3plotM(outdata)
outmulti$nOut
```

```
## HDo BAC DDC
##  14   0   0
```

```
outmulti$gO3
```



Crosstabs are created with the gmodels package.

```
require(gmodels)
```

```
## Zorunlu paket yükleniyor: gmodels
```

```
CrossTable(mtcars$vs,mtcars$gear,prop.t=TRUE, prop.r=TRUE, prop.c=TRUE,
expected=FALSE,chisq=FALSE, format="SPSS")
```

```
##
##    Cell Contents
```

```
## |-------------------------|
## |                   Count |
## | Chi-square contribution |
## |             Row Percent |
## |          Column Percent |
## |           Total Percent |
## |-------------------------|
##
## Total Observations in Table:  32
##
##              | mtcars$gear
##    mtcars$vs |        3  |        4  |        5  | Row Total |
## -------------|-----------|-----------|-----------|-----------|
##            0 |       12  |        2  |        4  |       18  |
##              |    1.504  |    3.343  |    0.501  |           |
##              |   66.667% |   11.111% |   22.222% |   56.250% |
##              |   80.000% |   16.667% |   80.000% |           |
##              |   37.500% |    6.250% |   12.500% |           |
## -------------|-----------|-----------|-----------|-----------|
##            1 |        3  |       10  |        1  |       14  |
##              |    1.934  |    4.298  |    0.645  |           |
##              |   21.429% |   71.429% |    7.143% |   43.750% |
##              |   20.000% |   83.333% |   20.000% |           |
##              |    9.375% |   31.250% |    3.125% |           |
## -------------|-----------|-----------|-----------|-----------|
## Column Total |       15  |       12  |        5  |       32  |
##              |   46.875% |   37.500% |   15.625% |           |
## -------------|-----------|-----------|-----------|-----------|
##
##
```

Data analysis and summarization processes are performed with packages such as DescTools and dataMaid.

```
require(readr)
```

```
## Zorunlu paket yükleniyor: readr
```

```
require(DescTools)
```

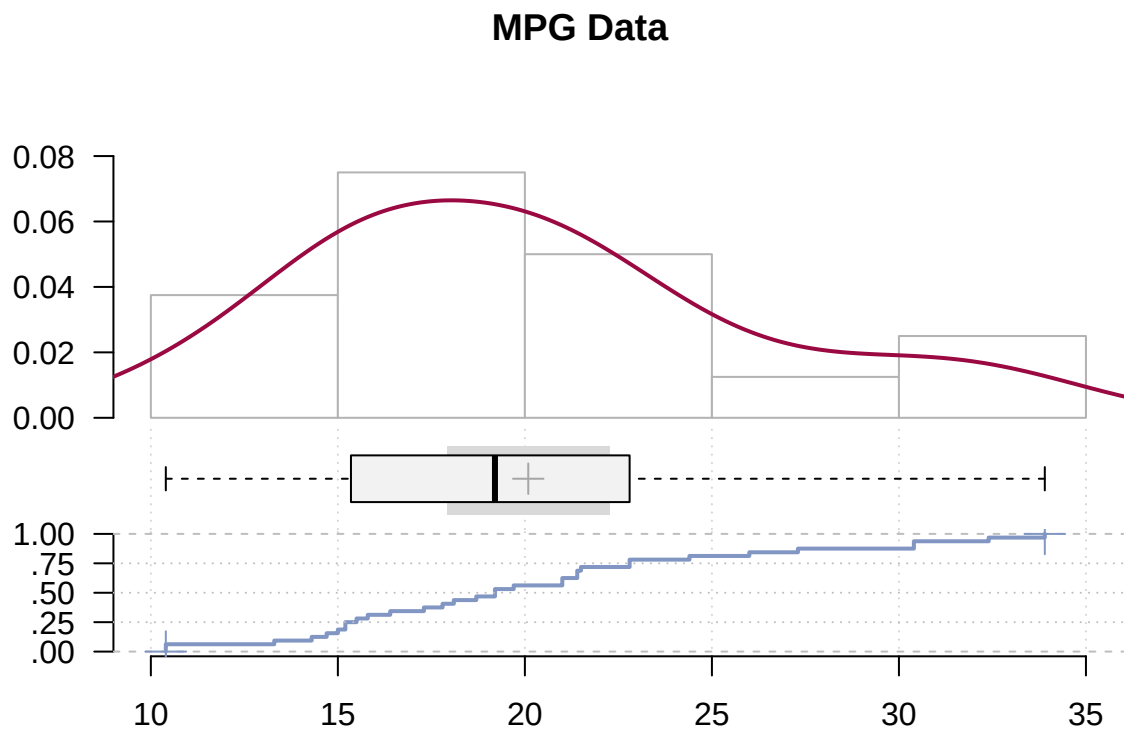
```
## Zorunlu paket yükleniyor: DescTools
```

```
## Registered S3 method overwritten by 'DescTools':
##   method         from
##   reorder.factor gdata
```

```
Desc(mtcars$mpg, plotit = TRUE, main= "MPG Data")
```

```
## ------------------------------------------------------------------------------
```

```
## MPG Data
##
##   length       n     NAs  unique      0s    mean  meanCI'
##       32      32       0      25       0  20.091  17.918
##           100.0%    0.0%            0.0%          22.264
##
##      .05     .10     .25  median     .75     .90     .95
##   11.995  14.340  15.425  19.200  22.800  30.090  31.300
##
##    range      sd   vcoef     mad     IQR    skew    kurt
##   23.500   6.027   0.300   5.411   7.375   0.611  -0.373
##
## lowest : 10.4 (2), 13.3, 14.3, 14.7, 15.0
## highest: 26.0, 27.3, 30.4 (2), 32.4, 33.9
##
## ' 95%-CI (classic)
```




```
Desc(as.factor(mtcars$gear), plotit=FALSE, main ="Gear Variable")
```

```
## ------------------------------------------------------------------------------
## Gear Variable
##
##   length       n    NAs unique levels  dupes
##       32      32      0      3      3      y
##           100.0%   0.0%
```

```
##
##    level  freq   perc  cumfreq  cumperc
## 1      3    15  46.9%       15    46.9%
## 2      4    12  37.5%       27    84.4%
## 3      5     5  15.6%       32   100.0%
```

```
Desc(as.factor(gear)~mpg, data = mtcars, plotit = FALSE,
main = "Gear ve Mpg Data")
```

```
## ------------------------------------------------------------------------------
## Gear ve Mpg Data
##
## Summary:
## n pairs: 32, valid: 32 (100.0%), missings: 0 (0.0%), groups: 3
##
##
##                3        4        5
## mean      16.107   24.533   21.380
## median    15.500   22.800   19.700
## sd         3.372    5.277    6.659
## IQR        3.900    7.075   10.200
## n             15       12        5
## np       46.875%  37.500%  15.625%
## NAs            0        0        0
## 0s             0        0        0
##
## Kruskal-Wallis rank sum test:
##   Kruskal-Wallis chi-squared = 14.323, df = 2, p-value = 0.0007758
##
##
##
##
##
## Proportions of as.factor(gear) in the quantiles of mpg:
##
##           Q1       Q2       Q3       Q4
##   3   87.5%   66.7%   25.0%    0.0%
##   4    0.0%   22.2%   62.5%   71.4%
##   5   12.5%   11.1%   12.5%   28.6%
```

```
require(dataMaid)
```

```
## Zorunlu paket yükleniyor: dataMaid
```

```
##
## Attaching package: 'dataMaid'
```

```
## The following object is masked from 'package:dplyr':
##
##     summarize
```

```r
dataMaid::summarize(mtcars["mpg"], reportstyleOutput = TRUE)
```

```
## $mpg
##      Feature                    Result
## [1,] "Variable type"            "numeric"
## [2,] "Number of missing obs."   "0 (0 %)"
## [3,] "Number of unique values" "25"
## [4,] "Median"                   "19.2"
## [5,] "1st and 3rd quartiles"    "15.43; 22.8"
## [6,] "Min. and max."            "10.4; 33.9"
```