# Some Simple Reports in R

We will look at some of the summary methods in R. This document will be available as a markdown doc, so you can use this to create MSoffice, pdf or html report files on your own data.

# Define datasets

data(mtcars)  
df <- mtcars  
dim(df)

## [1] 32 11

library(gmodels)  
library(Hmisc)  
library(ade4)  
library(markdown)

## Error: there is no package called 'markdown'

library(knitr)

View data

View(df)  
head(df)

## mpg cyl disp hp drat wt qsec vs am gear carb  
## Mazda RX4 21.0 6 160 110 3.90 2.620 16.46 0 1 4 4  
## Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 1 4 4  
## Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 1 4 1  
## Hornet 4 Drive 21.4 6 258 110 3.08 3.215 19.44 1 0 3 1  
## Hornet Sportabout 18.7 8 360 175 3.15 3.440 17.02 0 0 3 2  
## Valiant 18.1 6 225 105 2.76 3.460 20.22 1 0 3 1

tail(df)

## mpg cyl disp hp drat wt qsec vs am gear carb  
## Porsche 914-2 26.0 4 120.3 91 4.43 2.140 16.7 0 1 5 2  
## Lotus Europa 30.4 4 95.1 113 3.77 1.513 16.9 1 1 5 2  
## Ford Pantera L 15.8 8 351.0 264 4.22 3.170 14.5 0 1 5 4  
## Ferrari Dino 19.7 6 145.0 175 3.62 2.770 15.5 0 1 5 6  
## Maserati Bora 15.0 8 301.0 335 3.54 3.570 14.6 0 1 5 8  
## Volvo 142E 21.4 4 121.0 109 4.11 2.780 18.6 1 1 4 2

str(df)

## 'data.frame': 32 obs. of 11 variables:  
## $ mpg : num 21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...  
## $ cyl : num 6 6 4 6 8 6 8 4 4 6 ...  
## $ disp: num 160 160 108 258 360 ...  
## $ hp : num 110 110 93 110 175 105 245 62 95 123 ...  
## $ drat: num 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...  
## $ wt : num 2.62 2.88 2.32 3.21 3.44 ...  
## $ qsec: num 16.5 17 18.6 19.4 17 ...  
## $ vs : num 0 0 1 1 0 1 0 1 1 1 ...  
## $ am : num 1 1 1 0 0 0 0 0 0 0 ...  
## $ gear: num 4 4 4 3 3 3 3 4 4 4 ...  
## $ carb: num 4 4 1 1 2 1 4 2 2 4 ...

Basic Summary

summary(df)

## mpg cyl disp hp   
## Min. :10.4 Min. :4.00 Min. : 71.1 Min. : 52.0   
## 1st Qu.:15.4 1st Qu.:4.00 1st Qu.:120.8 1st Qu.: 96.5   
## Median :19.2 Median :6.00 Median :196.3 Median :123.0   
## Mean :20.1 Mean :6.19 Mean :230.7 Mean :146.7   
## 3rd Qu.:22.8 3rd Qu.:8.00 3rd Qu.:326.0 3rd Qu.:180.0   
## Max. :33.9 Max. :8.00 Max. :472.0 Max. :335.0   
## drat wt qsec vs   
## Min. :2.76 Min. :1.51 Min. :14.5 Min. :0.000   
## 1st Qu.:3.08 1st Qu.:2.58 1st Qu.:16.9 1st Qu.:0.000   
## Median :3.69 Median :3.33 Median :17.7 Median :0.000   
## Mean :3.60 Mean :3.22 Mean :17.8 Mean :0.438   
## 3rd Qu.:3.92 3rd Qu.:3.61 3rd Qu.:18.9 3rd Qu.:1.000   
## Max. :4.93 Max. :5.42 Max. :22.9 Max. :1.000   
## am gear carb   
## Min. :0.000 Min. :3.00 Min. :1.00   
## 1st Qu.:0.000 1st Qu.:3.00 1st Qu.:2.00   
## Median :0.000 Median :4.00 Median :2.00   
## Mean :0.406 Mean :3.69 Mean :2.81   
## 3rd Qu.:1.000 3rd Qu.:4.00 3rd Qu.:4.00   
## Max. :1.000 Max. :5.00 Max. :8.00

Using the describe function

library(Hmisc)  
describe(df)

## df   
##   
## 11 Variables 32 Observations  
## ---------------------------------------------------------------------------  
## mpg   
## n missing unique Mean .05 .10 .25 .50 .75   
## 32 0 25 20.09 12.00 14.34 15.43 19.20 22.80   
## .90 .95   
## 30.09 31.30   
##   
## lowest : 10.4 13.3 14.3 14.7 15.0, highest: 26.0 27.3 30.4 32.4 33.9   
## ---------------------------------------------------------------------------  
## cyl   
## n missing unique Mean   
## 32 0 3 6.188   
##   
## 4 (11, 34%), 6 (7, 22%), 8 (14, 44%)   
## ---------------------------------------------------------------------------  
## disp   
## n missing unique Mean .05 .10 .25 .50 .75   
## 32 0 27 230.7 77.35 80.61 120.83 196.30 326.00   
## .90 .95   
## 396.00 449.00   
##   
## lowest : 71.1 75.7 78.7 79.0 95.1  
## highest: 360.0 400.0 440.0 460.0 472.0   
## ---------------------------------------------------------------------------  
## hp   
## n missing unique Mean .05 .10 .25 .50 .75   
## 32 0 22 146.7 63.65 66.00 96.50 123.00 180.00   
## .90 .95   
## 243.50 253.55   
##   
## lowest : 52 62 65 66 91, highest: 215 230 245 264 335   
## ---------------------------------------------------------------------------  
## drat   
## n missing unique Mean .05 .10 .25 .50 .75   
## 32 0 22 3.597 2.853 3.007 3.080 3.695 3.920   
## .90 .95   
## 4.209 4.314   
##   
## lowest : 2.76 2.93 3.00 3.07 3.08, highest: 4.08 4.11 4.22 4.43 4.93   
## ---------------------------------------------------------------------------  
## wt   
## n missing unique Mean .05 .10 .25 .50 .75   
## 32 0 29 3.217 1.736 1.956 2.581 3.325 3.610   
## .90 .95   
## 4.048 5.293   
##   
## lowest : 1.513 1.615 1.835 1.935 2.140  
## highest: 3.845 4.070 5.250 5.345 5.424   
## ---------------------------------------------------------------------------  
## qsec   
## n missing unique Mean .05 .10 .25 .50 .75   
## 32 0 30 17.85 15.05 15.53 16.89 17.71 18.90   
## .90 .95   
## 19.99 20.10   
##   
## lowest : 14.50 14.60 15.41 15.50 15.84  
## highest: 19.90 20.00 20.01 20.22 22.90   
## ---------------------------------------------------------------------------  
## vs   
## n missing unique Sum Mean   
## 32 0 2 14 0.4375   
## ---------------------------------------------------------------------------  
## am   
## n missing unique Sum Mean   
## 32 0 2 13 0.4062   
## ---------------------------------------------------------------------------  
## gear   
## n missing unique Mean   
## 32 0 3 3.688   
##   
## 3 (15, 47%), 4 (12, 38%), 5 (5, 16%)   
## ---------------------------------------------------------------------------  
## carb   
## n missing unique Mean   
## 32 0 6 2.812   
##   
## 1 2 3 4 6 8  
## Frequency 7 10 3 10 1 1  
## % 22 31 9 31 3 3  
## ---------------------------------------------------------------------------

# 1,2 and 3-way Cross Tabulations

table(df$cyl)

##   
## 4 6 8   
## 11 7 14

table(df$cyl, df$gear)

##   
## 3 4 5  
## 4 1 8 2  
## 6 2 4 1  
## 8 12 0 2

# Number of cyclinders, numbers of gear, transmission type  
table(df$cyl, df$gear, df$am)

## , , = 0  
##   
##   
## 3 4 5  
## 4 1 2 0  
## 6 2 2 0  
## 8 12 0 0  
##   
## , , = 1  
##   
##   
## 3 4 5  
## 4 0 6 2  
## 6 0 2 1  
## 8 0 0 2  
##

Crosstabulation using formula format

xtabs(cyl ~ gear, df)

## gear  
## 3 4 5   
## 112 56 30

xtabs(cyl ~ gear + am + vs, df)

## , , vs = 0  
##   
## am  
## gear 0 1  
## 3 96 0  
## 4 0 12  
## 5 0 26  
##   
## , , vs = 1  
##   
## am  
## gear 0 1  
## 3 16 0  
## 4 20 24  
## 5 0 4  
##

Create Contingency Table

`?`(ftable)  
ftable(df$cyl, df$vs, df$am, df$gear, row.vars = c(2, 4), dnn = c("Cylinders",   
 "V/S", "Transmission", "Gears"))

## Cylinders 4 6 8   
## Transmission 0 1 0 1 0 1  
## V/S Gears   
## 0 3 0 0 0 0 12 0  
## 4 0 0 0 2 0 0  
## 5 0 1 0 1 0 2  
## 1 3 1 0 2 0 0 0  
## 4 2 6 2 0 0 0  
## 5 0 1 0 0 0 0

ftable(df$cyl, df$vs, df$am, df$gear, row.vars = c(2, 3), dnn = c("Cylinders",   
 "V/S", "Transmission", "Gears"))

## Cylinders 4 6 8   
## Gears 3 4 5 3 4 5 3 4 5  
## V/S Transmission   
## 0 0 0 0 0 0 0 0 12 0 0  
## 1 0 0 1 0 2 1 0 0 2  
## 1 0 1 2 0 2 2 0 0 0 0  
## 1 0 6 1 0 0 0 0 0 0

2 way cross tabulation in SAS format

library(gmodels)  
CrossTable(df$cyl, df$gear, format = "SAS")

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | Chi-square contribution |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 32   
##   
##   
## | df$gear   
## df$cyl | 3 | 4 | 5 | Row Total |   
## -------------|-----------|-----------|-----------|-----------|  
## 4 | 1 | 8 | 2 | 11 |   
## | 3.350 | 3.640 | 0.046 | |   
## | 0.091 | 0.727 | 0.182 | 0.344 |   
## | 0.067 | 0.667 | 0.400 | |   
## | 0.031 | 0.250 | 0.062 | |   
## -------------|-----------|-----------|-----------|-----------|  
## 6 | 2 | 4 | 1 | 7 |   
## | 0.500 | 0.720 | 0.008 | |   
## | 0.286 | 0.571 | 0.143 | 0.219 |   
## | 0.133 | 0.333 | 0.200 | |   
## | 0.062 | 0.125 | 0.031 | |   
## -------------|-----------|-----------|-----------|-----------|  
## 8 | 12 | 0 | 2 | 14 |   
## | 4.505 | 5.250 | 0.016 | |   
## | 0.857 | 0.000 | 0.143 | 0.438 |   
## | 0.800 | 0.000 | 0.400 | |   
## | 0.375 | 0.000 | 0.062 | |   
## -------------|-----------|-----------|-----------|-----------|  
## Column Total | 15 | 12 | 5 | 32 |   
## | 0.469 | 0.375 | 0.156 | |   
## -------------|-----------|-----------|-----------|-----------|  
##   
##

CrossTable(df$cyl, df$gear, expected = TRUE, format = "SAS")

## Warning: Chi-squared approximation may be incorrect

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | Expected N |  
## | Chi-square contribution |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 32   
##   
##   
## | df$gear   
## df$cyl | 3 | 4 | 5 | Row Total |   
## -------------|-----------|-----------|-----------|-----------|  
## 4 | 1 | 8 | 2 | 11 |   
## | 5.156 | 4.125 | 1.719 | |   
## | 3.350 | 3.640 | 0.046 | |   
## | 0.091 | 0.727 | 0.182 | 0.344 |   
## | 0.067 | 0.667 | 0.400 | |   
## | 0.031 | 0.250 | 0.062 | |   
## -------------|-----------|-----------|-----------|-----------|  
## 6 | 2 | 4 | 1 | 7 |   
## | 3.281 | 2.625 | 1.094 | |   
## | 0.500 | 0.720 | 0.008 | |   
## | 0.286 | 0.571 | 0.143 | 0.219 |   
## | 0.133 | 0.333 | 0.200 | |   
## | 0.062 | 0.125 | 0.031 | |   
## -------------|-----------|-----------|-----------|-----------|  
## 8 | 12 | 0 | 2 | 14 |   
## | 6.562 | 5.250 | 2.188 | |   
## | 4.505 | 5.250 | 0.016 | |   
## | 0.857 | 0.000 | 0.143 | 0.438 |   
## | 0.800 | 0.000 | 0.400 | |   
## | 0.375 | 0.000 | 0.062 | |   
## -------------|-----------|-----------|-----------|-----------|  
## Column Total | 15 | 12 | 5 | 32 |   
## | 0.469 | 0.375 | 0.156 | |   
## -------------|-----------|-----------|-----------|-----------|  
##   
##   
## Statistics for All Table Factors  
##   
##   
## Pearson's Chi-squared test   
## ------------------------------------------------------------  
## Chi^2 = 18.04 d.f. = 4 p = 0.001214   
##   
##   
##

2 way cross tabulation in SPSS format

library(gmodels)  
CrossTable(df$cyl, df$gear, format = "SPSS")

##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Chi-square contribution |  
## | Row Percent |  
## | Column Percent |  
## | Total Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 32   
##   
## | df$gear   
## df$cyl | 3 | 4 | 5 | Row Total |   
## -------------|-----------|-----------|-----------|-----------|  
## 4 | 1 | 8 | 2 | 11 |   
## | 3.350 | 3.640 | 0.046 | |   
## | 9.091% | 72.727% | 18.182% | 34.375% |   
## | 6.667% | 66.667% | 40.000% | |   
## | 3.125% | 25.000% | 6.250% | |   
## -------------|-----------|-----------|-----------|-----------|  
## 6 | 2 | 4 | 1 | 7 |   
## | 0.500 | 0.720 | 0.008 | |   
## | 28.571% | 57.143% | 14.286% | 21.875% |   
## | 13.333% | 33.333% | 20.000% | |   
## | 6.250% | 12.500% | 3.125% | |   
## -------------|-----------|-----------|-----------|-----------|  
## 8 | 12 | 0 | 2 | 14 |   
## | 4.505 | 5.250 | 0.016 | |   
## | 85.714% | 0.000% | 14.286% | 43.750% |   
## | 80.000% | 0.000% | 40.000% | |   
## | 37.500% | 0.000% | 6.250% | |   
## -------------|-----------|-----------|-----------|-----------|  
## Column Total | 15 | 12 | 5 | 32 |   
## | 46.875% | 37.500% | 15.625% | |   
## -------------|-----------|-----------|-----------|-----------|  
##   
##

CrossTable(df$cyl, df$gear, expected = TRUE, format = "SPSS")

## Warning: Chi-squared approximation may be incorrect

##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Expected Values |  
## | Chi-square contribution |  
## | Row Percent |  
## | Column Percent |  
## | Total Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 32   
##   
## | df$gear   
## df$cyl | 3 | 4 | 5 | Row Total |   
## -------------|-----------|-----------|-----------|-----------|  
## 4 | 1 | 8 | 2 | 11 |   
## | 5.156 | 4.125 | 1.719 | |   
## | 3.350 | 3.640 | 0.046 | |   
## | 9.091% | 72.727% | 18.182% | 34.375% |   
## | 6.667% | 66.667% | 40.000% | |   
## | 3.125% | 25.000% | 6.250% | |   
## -------------|-----------|-----------|-----------|-----------|  
## 6 | 2 | 4 | 1 | 7 |   
## | 3.281 | 2.625 | 1.094 | |   
## | 0.500 | 0.720 | 0.008 | |   
## | 28.571% | 57.143% | 14.286% | 21.875% |   
## | 13.333% | 33.333% | 20.000% | |   
## | 6.250% | 12.500% | 3.125% | |   
## -------------|-----------|-----------|-----------|-----------|  
## 8 | 12 | 0 | 2 | 14 |   
## | 6.562 | 5.250 | 2.188 | |   
## | 4.505 | 5.250 | 0.016 | |   
## | 85.714% | 0.000% | 14.286% | 43.750% |   
## | 80.000% | 0.000% | 40.000% | |   
## | 37.500% | 0.000% | 6.250% | |   
## -------------|-----------|-----------|-----------|-----------|  
## Column Total | 15 | 12 | 5 | 32 |   
## | 46.875% | 37.500% | 15.625% | |   
## -------------|-----------|-----------|-----------|-----------|  
##   
##   
## Statistics for All Table Factors  
##   
##   
## Pearson's Chi-squared test   
## ------------------------------------------------------------  
## Chi^2 = 18.04 d.f. = 4 p = 0.001214   
##   
##   
##   
## Minimum expected frequency: 1.094   
## Cells with Expected Frequency < 5: 6 of 9 (66.67%)  
##

# Categorical Data

The library *vcd* is very useful

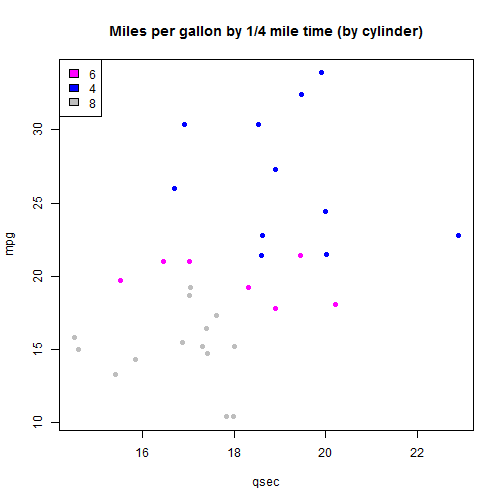
# Some Plots for Exploring Data

* scatterplot

attach(df)

## The following object(s) are masked from 'df (position 3)':  
##   
## am, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs, wt  
## The following object(s) are masked from 'df (position 4)':  
##   
## am, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs, wt  
## The following object(s) are masked from 'df (position 5)':  
##   
## am, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs, wt  
## The following object(s) are masked from 'df (position 6)':  
##   
## am, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs, wt  
## The following object(s) are masked from 'df (position 7)':  
##   
## am, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs, wt  
## The following object(s) are masked from 'df (position 8)':  
##   
## am, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs, wt  
## The following object(s) are masked from 'df (position 9)':  
##   
## am, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs, wt  
## The following object(s) are masked from 'df (position 10)':  
##   
## am, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs, wt  
## The following object(s) are masked from 'df (position 11)':  
##   
## am, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs, wt  
## The following object(s) are masked from 'df (position 12)':  
##   
## am, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs, wt  
## The following object(s) are masked from 'df (position 13)':  
##   
## am, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs, wt  
## The following object(s) are masked from 'df (position 14)':  
##   
## am, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs, wt  
## The following object(s) are masked from 'df (position 15)':  
##   
## am, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs, wt  
## The following object(s) are masked from 'df (position 16)':  
##   
## am, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs, wt  
## The following object(s) are masked from 'mtcars':  
##   
## am, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs, wt

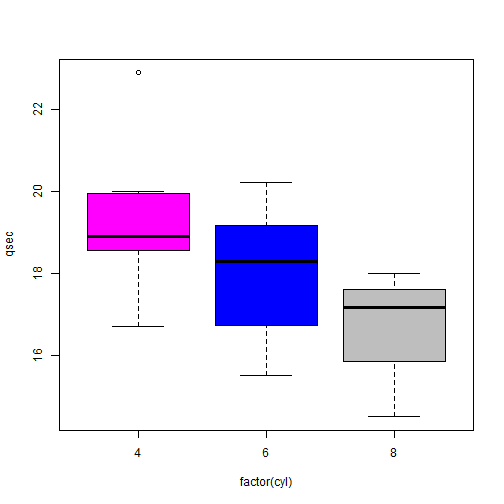
plot(qsec, mpg, col = cyl, pch = 19, main = "Miles per gallon by 1/4 mile time (by cylinder)")  
legend("topleft", legend = unique(cyl), fill = unique(cyl))



plot of chunk scatterplot

* boxplot

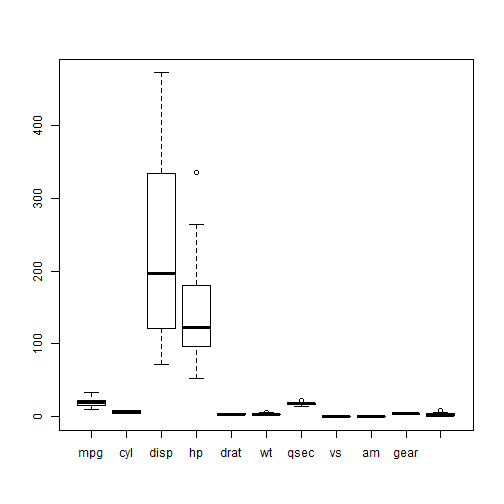
plot(qsec ~ factor(cyl), col = unique(cyl))



plot of chunk boxplot

* boxplot all of the columns

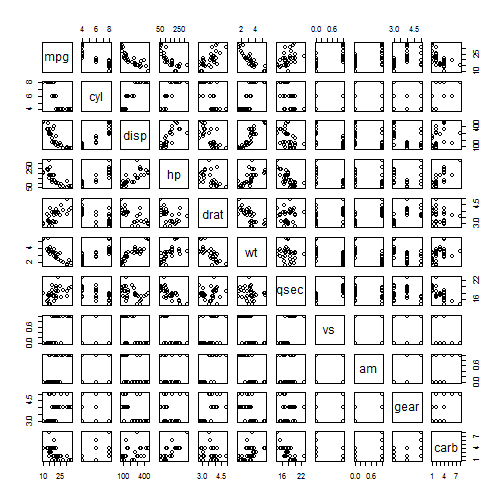
boxplot(df)



plot of chunk boxplotALL

* Correlation across

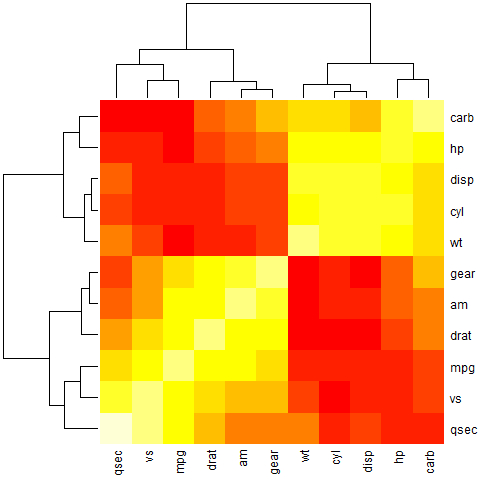
plot(df)



plot of chunk pairs

Or calculate correlation and view on heatmap

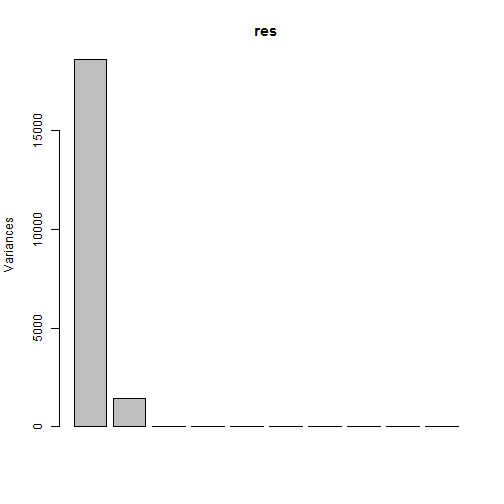
heatmap(cor(df))



plot of chunk heatmap

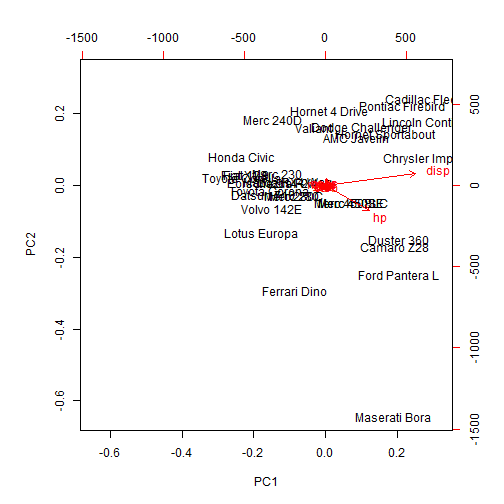
Basic principcal component analysis

res <- prcomp(df)  
screeplot(res)



plot of chunk prcomp

biplot(res)



plot of chunk prcomp

Or using fast.prcomp (optimized for big wide datasets)

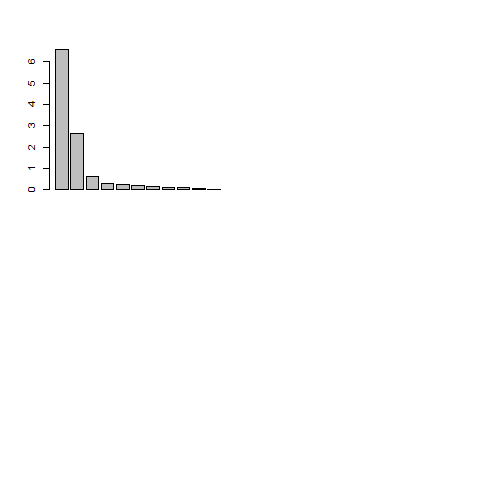
res <- fast.prcomp(df)  
s.class(res$li, factor(cyl), col = unique(cyl))

## Error: undefined columns selected

s.arrow(res$li, cpoint = cyl)

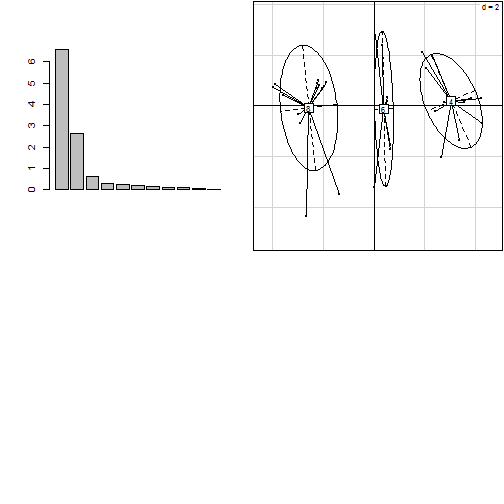
## Error: Non convenient selection for xax

library(ade4)  
res <- dudi.pca(df, scan = FALSE)  
par(mfrow = c(2, 2))  
barplot(res$eig)



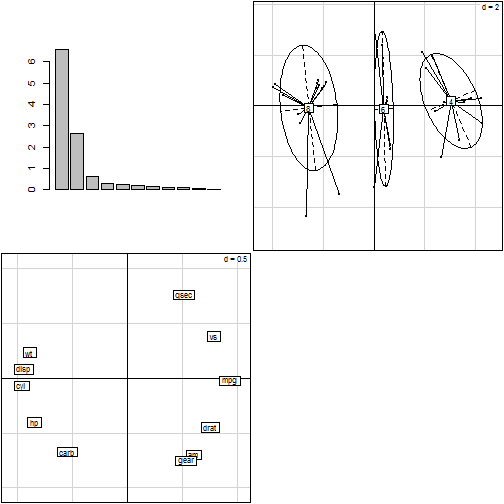
plot of chunk dudi.pca

s.class(res$li, factor(cyl))



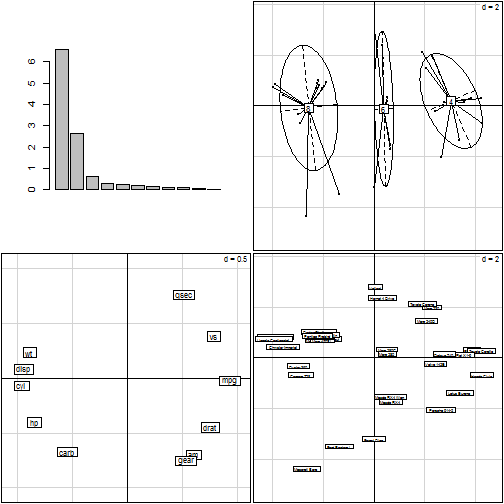
plot of chunk dudi.pca

s.label(res$co)



plot of chunk dudi.pca

s.label(res$li, clabel = 0.5)



plot of chunk dudi.pca

# Missing Data

df[sample(1:nrow(df), 2), sample(1:ncol(df), 2)] <- NA  
summary(df)

## mpg cyl disp hp   
## Min. :10.4 Min. :4.00 Min. : 71.1 Min. : 52.0   
## 1st Qu.:15.3 1st Qu.:4.00 1st Qu.:125.4 1st Qu.: 96.5   
## Median :18.9 Median :6.00 Median :196.3 Median :123.0   
## Mean :20.1 Mean :6.19 Mean :228.7 Mean :146.7   
## 3rd Qu.:22.8 3rd Qu.:8.00 3rd Qu.:314.5 3rd Qu.:180.0   
## Max. :33.9 Max. :8.00 Max. :472.0 Max. :335.0   
## NA's :2 NA's :2   
## drat wt qsec vs   
## Min. :2.76 Min. :1.51 Min. :14.5 Min. :0.000   
## 1st Qu.:3.08 1st Qu.:2.58 1st Qu.:16.9 1st Qu.:0.000   
## Median :3.69 Median :3.33 Median :17.7 Median :0.000   
## Mean :3.60 Mean :3.22 Mean :17.8 Mean :0.438   
## 3rd Qu.:3.92 3rd Qu.:3.61 3rd Qu.:18.9 3rd Qu.:1.000   
## Max. :4.93 Max. :5.42 Max. :22.9 Max. :1.000   
##   
## am gear carb   
## Min. :0.000 Min. :3.00 Min. :1.00   
## 1st Qu.:0.000 1st Qu.:3.00 1st Qu.:2.00   
## Median :0.000 Median :4.00 Median :2.00   
## Mean :0.406 Mean :3.69 Mean :2.81   
## 3rd Qu.:1.000 3rd Qu.:4.00 3rd Qu.:4.00   
## Max. :1.000 Max. :5.00 Max. :8.00   
##

# Analyzing >1 Dataset

Often we have 2 or more tables either reflecting different time points of the same sample population or different measuments on the same population.

*Merge Data* There are several function for manipulating data, see the plyr library for functions. Also see the function reshape and stack which make it easier to convert a "wide" table into a narrow one.

x1 <- data.frame(Case = sample(letters, 10), A1 = rnorm(10), B1 = 1:10,   
 C1 = rep(1:5, 2))  
x1

## Case A1 B1 C1  
## 1 f -0.4227 1 1  
## 2 w 1.1173 2 2  
## 3 c 0.2895 3 3  
## 4 u 0.2005 4 4  
## 5 l -0.2262 5 5  
## 6 x 1.1932 6 1  
## 7 g -0.4561 7 2  
## 8 e -0.6621 8 3  
## 9 o 0.2095 9 4  
## 10 h 0.2013 10 5

x2 <- data.frame(A1 = seq(1, 10, 2), Case = sample(letters, 10),   
 D1 = rnorm(10, 4), E1 = rep(1:5, 2), B1 = c(rep(c("Non-Smoker", "Smoker"),   
 each = 4), NA, NA))  
x2

## A1 Case D1 E1 B1  
## 1 1 z 4.567 1 Non-Smoker  
## 2 3 f 4.649 2 Non-Smoker  
## 3 5 y 4.286 3 Non-Smoker  
## 4 7 d 3.085 4 Non-Smoker  
## 5 9 r 3.391 5 Smoker  
## 6 1 c 4.558 1 Smoker  
## 7 3 j 2.966 2 Smoker  
## 8 5 b 5.230 3 Smoker  
## 9 7 q 2.708 4 <NA>  
## 10 9 w 2.815 5 <NA>

merge(x1, x2, "Case")

## Case A1.x B1.x C1 A1.y D1 E1 B1.y  
## 1 c 0.2895 3 3 1 4.558 1 Smoker  
## 2 f -0.4227 1 1 3 4.649 2 Non-Smoker  
## 3 w 1.1173 2 2 9 2.815 5 <NA>

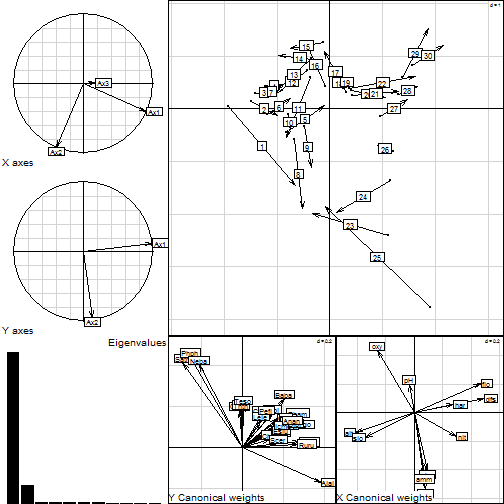
# Multivariate methods for exploring covariance across studies

Lets look at the doubs data in the ade4 package. This data set gives environmental variables, fish species and spatial coordinates for 30 sites

require(ade4)  
data(doubs)  
lapply(doubs, head)

## $env  
## dfs alt slo flo pH har pho nit amm oxy bdo  
## 1 3 934 6.176 84 79 45 1 20 0 122 27  
## 2 22 932 3.434 100 80 40 2 20 10 103 19  
## 3 102 914 3.638 180 83 52 5 22 5 105 35  
## 4 185 854 3.497 253 80 72 10 21 0 110 13  
## 5 215 849 3.178 264 81 84 38 52 20 80 62  
## 6 324 846 3.497 286 79 60 20 15 0 102 53  
##   
## $fish  
## Cogo Satr Phph Neba Thth Teso Chna Chto Lele Lece Baba Spbi Gogo Eslu  
## 1 0 3 0 0 0 0 0 0 0 0 0 0 0 0  
## 2 0 5 4 3 0 0 0 0 0 0 0 0 0 0  
## 3 0 5 5 5 0 0 0 0 0 0 0 0 0 1  
## 4 0 4 5 5 0 0 0 0 0 1 0 0 1 2  
## 5 0 2 3 2 0 0 0 0 5 2 0 0 2 4  
## 6 0 3 4 5 0 0 0 0 1 2 0 0 1 1  
## Pefl Rham Legi Scer Cyca Titi Abbr Icme Acce Ruru Blbj Alal Anan  
## 1 0 0 0 0 0 0 0 0 0 0 0 0 0  
## 2 0 0 0 0 0 0 0 0 0 0 0 0 0  
## 3 0 0 0 0 0 0 0 0 0 0 0 0 0  
## 4 2 0 0 0 0 1 0 0 0 0 0 0 0  
## 5 4 0 0 2 0 3 0 0 0 5 0 0 0  
## 6 1 0 0 0 0 2 0 0 0 1 0 0 0  
##   
## $xy  
## x y  
## 1 88 7  
## 2 94 14  
## 3 102 18  
## 4 100 28  
## 5 106 39  
## 6 112 51  
##   
## $species  
## Scientific French English code  
## 1 Cottus gobio chabot european bullhead Cogo  
## 2 Salmo trutta fario truite fario brown trout Satr  
## 3 Phoxinus phoxinus vairon minnow Phph  
## 4 Nemacheilus barbatulus loche franche stone loach Neba  
## 5 Thymallus thymallus ombre grayling Thth  
## 6 Telestes soufia agassizi blageon blageon Teso  
##

dudi1 <- dudi.pca(doubs$env, scale = TRUE, scannf = FALSE, nf = 3)  
dudi2 <- dudi.pca(doubs$fish, scale = FALSE, scannf = FALSE, nf = 2)  
coin1 <- coinertia(dudi1, dudi2, scan = FALSE, nf = 2)  
plot(coin1)



plot of chunk coinertia

# s.arrow(coin1$l1, clab = 0.7)

# How to Process this document

require(knitr)  
dir(pattern="Rmd")  
knit("Reports.Rmd")  
knit2html("Reports.Rmd")  
knit2pdf("Reports.Rmd")  
purl("Reports.Rmd")

Or use pandoc to convert markdown file {} system("pandoc -s Reports.md -o Reports.pdf") system("pandoc -s Reports.md -o Reports.docx") system("pandoc -s Reports.md -o Reports.html") dir()