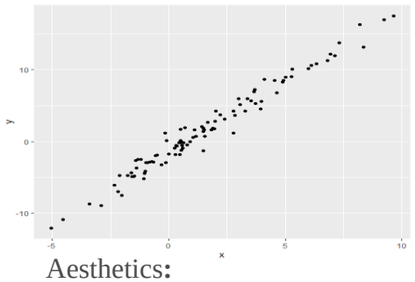
Scatterplots : : **CHEAT SHEET**

Basics

Scatterplot is one of the good methods to plot two continuous variables to investigate how they are related. It can be used to compare one continuous and one categorical variable, or two categorical variables.

package: library(ggplot2)

Sample scatterplot:
ggplot(data, aes(x = x, y = y)) + geom_point()

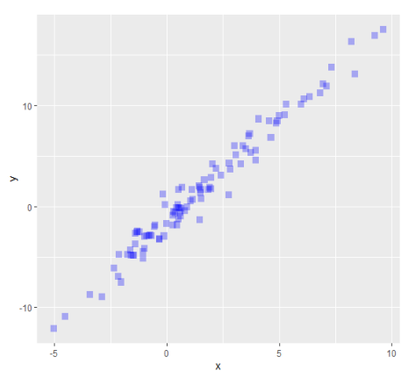


Aesthetics:
geom_point(color, size, alpha, stroke, shape)

alpha: alpha sets the transparency of points, with value between 0 and 1. The smaller the value, the more transparent points are.

stroke: stroke sets the width of the scatter edge

Example:
ggplot(data = df, aes(x = x, y = y)) + geom_point(size = 3, alpha = 0.3, shape = 15, color = "blue", stroke = 0.3)

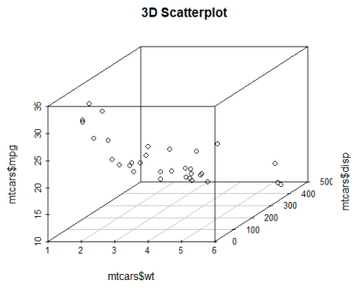


3D Scatterplots

package: library(scatterplot3d)

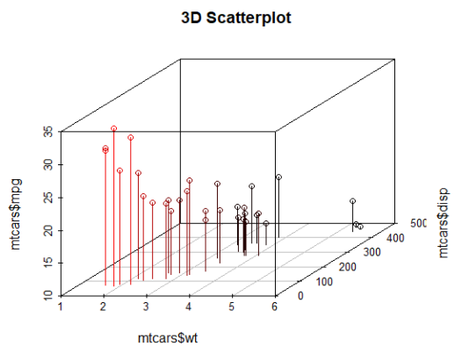
default function:
scatterplot3d(dfx, dfy, df$z)

Example:
scatterplot3d(mtcars$wt, mtcars$disp, mtcars$mpg, main="3D Scatterplot")



To make readers find the locations of these points more clearly in this 3D scatterplot, we can use parameter: **type = "h"** to add the vertical and **highlight.3d=TRUE** to add color.

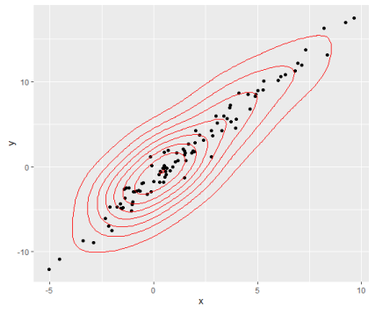
Example:
scatterplot3d(mtcars$wt, mtcars$disp, mtcars$mpg, highlight.3d=TRUE, type="h", main="3D Scatterplot")



Density contour lines, Hexagonal and Square heatmap

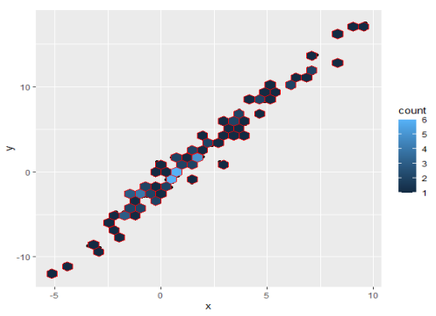
Using **geom_density2d()** to add density contour lines:

ggplot(data = df, aes(x = x, y = y)) + geom_point() + geom_density2d(color = "red")



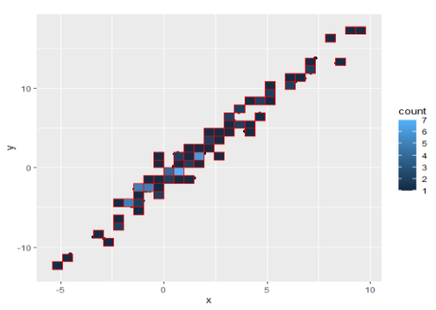
Using **geom_hex()** to make hexagonal heatmap:

ggplot(data = df, aes(x = x, y = y)) + geom_point() + geom_hex(color = "red", bins = 30)



Using geom_bin_2d() to make Square heatmap:

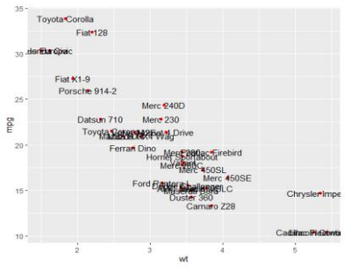
ggplot(data = df, aes(x = x, y = y)) + geom_point() + geom_bin_2d(color = "red", bins = 30)



Add text

Using **geom_text()** to add texts for all points:

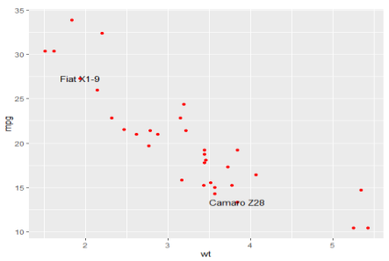
ggplot(mtcars, aes(wt, mpg)) + geom_point(color = 'red') + geom_text(aes(wt, mpg, label = rownames(mtcars)))



Even though this graph looks a little messy, we can use it to figure out which points are outliers

Using annotate() to label some specific points:

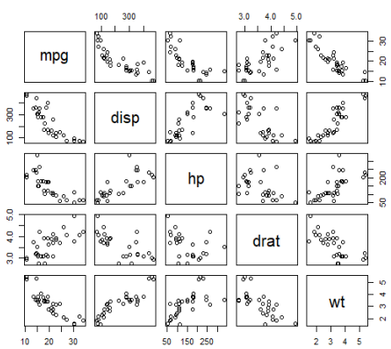
ggplot(mtcars, aes(wt, mpg)) + geom_point(color = 'red') + annotate("text", x = 3.840, y = 13.3, label = "Camaro Z28") + annotate("text", x = 1.935, y = 27.3, label = "Fiat X1-9")



Matrix

Using pairs() to create a scatterplot matrix which can show more relationship from the data:

pairs(~mpg+disp+hp+drat+wt,data=mtcars)



Source: https://zhuanlan.zhihu.com/p/123376781