allgreens.R

dell

2019-10-04

setwd("/home/dell/mne/BABI/Project2/2.1")  
install.packages("tidyverse")

## Installing package into '/home/dell/R/x86\_64-pc-linux-gnu-library/3.6'  
## (as 'lib' is unspecified)

rawdata = readxl::read\_excel("allgreens.xls")  
rawdata

## # A tibble: 27 x 6  
## X1 X2 X3 X4 X5 X6  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 231 3 294 8.20 8.20 11  
## 2 156 2.20 232 6.90 4.10 12  
## 3 10 0.5 149 3 4.30 15  
## 4 519 5.5 600 12 16.1 1  
## 5 437 4.40 567 10.6 14.1 5  
## 6 487 4.80 571 11.8 12.7 4  
## 7 299 3.10 512 8.1 10.1 10  
## 8 195 2.5 347 7.70 8.4 12  
## 9 20 1.20 212 3.30 2.10 15  
## 10 68 0.600 102 4.90 4.70 8  
## # … with 17 more rows

dim(rawdata)

## [1] 27 6

str(rawdata)

## Classes 'tbl\_df', 'tbl' and 'data.frame': 27 obs. of 6 variables:  
## $ X1: num 231 156 10 519 437 487 299 195 20 68 ...  
## $ X2: num 3 2.2 0.5 5.5 4.4 ...  
## $ X3: num 294 232 149 600 567 571 512 347 212 102 ...  
## $ X4: num 8.2 6.9 3 12 10.6 ...  
## $ X5: num 8.2 4.1 4.3 16.1 14.1 ...  
## $ X6: num 11 12 15 1 5 4 10 12 15 8 ...

anyNA(rawdata)

## [1] FALSE

summary(rawdata)

## X1 X2 X3 X4   
## Min. : 0.5 Min. :0.500 Min. :102.0 Min. : 2.50   
## 1st Qu.: 98.5 1st Qu.:1.400 1st Qu.:204.0 1st Qu.: 4.80   
## Median :341.0 Median :3.500 Median :382.0 Median : 8.10   
## Mean :286.6 Mean :3.326 Mean :387.5 Mean : 8.10   
## 3rd Qu.:450.5 3rd Qu.:4.750 3rd Qu.:551.0 3rd Qu.:10.95   
## Max. :570.0 Max. :8.600 Max. :788.0 Max. :17.40   
## X5 X6   
## Min. : 1.600 Min. : 0.000   
## 1st Qu.: 4.500 1st Qu.: 4.000   
## Median :11.300 Median : 8.000   
## Mean : 9.693 Mean : 7.741   
## 3rd Qu.:14.050 3rd Qu.:12.000   
## Max. :16.300 Max. :15.000

names(rawdata)

## [1] "X1" "X2" "X3" "X4" "X5" "X6"

boxplot(rawdata, col = rep(c("Red", "Green", "Blue", "Yellow", "Purple" )))  
  
  
install.packages("purrr")

## Installing package into '/home/dell/R/x86\_64-pc-linux-gnu-library/3.6'  
## (as 'lib' is unspecified)

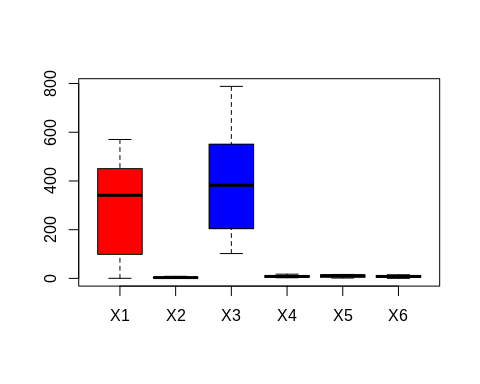
install.packages("tidyr")

## Installing package into '/home/dell/R/x86\_64-pc-linux-gnu-library/3.6'  
## (as 'lib' is unspecified)

install.packages("ggplot2")

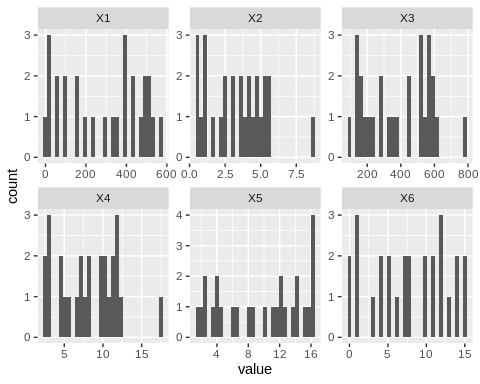
## Installing package into '/home/dell/R/x86\_64-pc-linux-gnu-library/3.6'  
## (as 'lib' is unspecified)

library(purrr)  
library(tidyr)  
library(ggplot2)

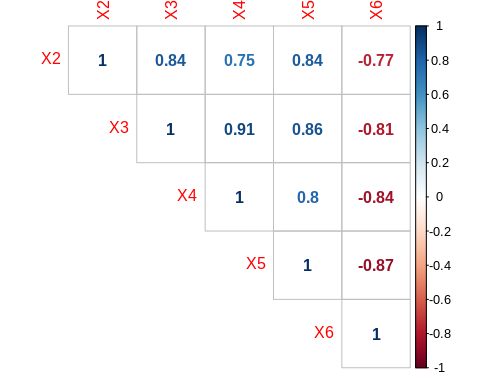


rawdata %>%  
 gather() %>%   
 ggplot(aes(value)) +  
 facet\_wrap(~ key, scales = "free") +  
 geom\_histogram()

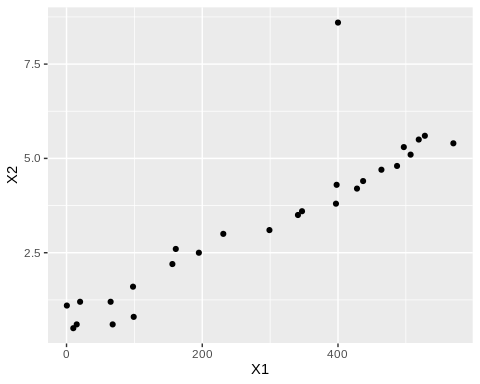
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



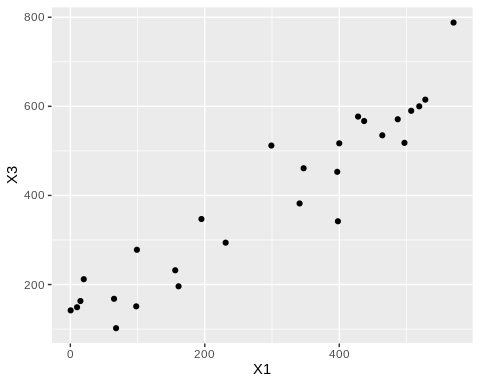
cormatrix = cor(rawdata[,2:6])  
corrplot::corrplot(cormatrix, type="upper", method="number")



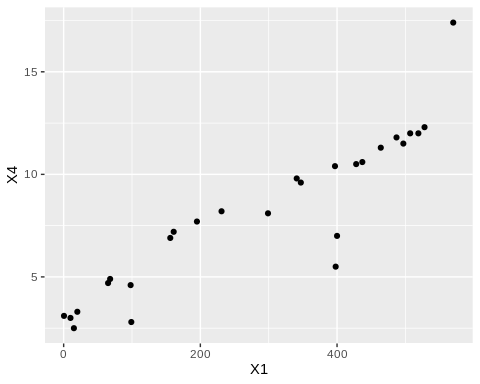
ggplot(data = rawdata) + geom\_point( mapping = aes(x = X1, y = X2))



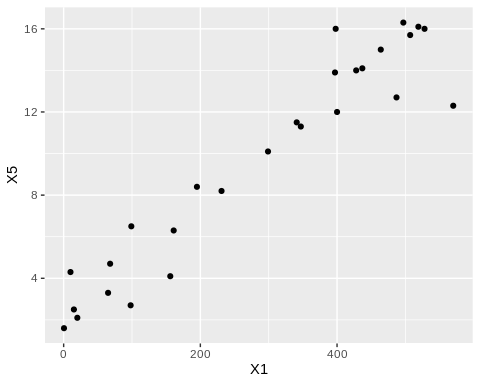
ggplot(data = rawdata) + geom\_point( mapping = aes(x = X1, y = X3))



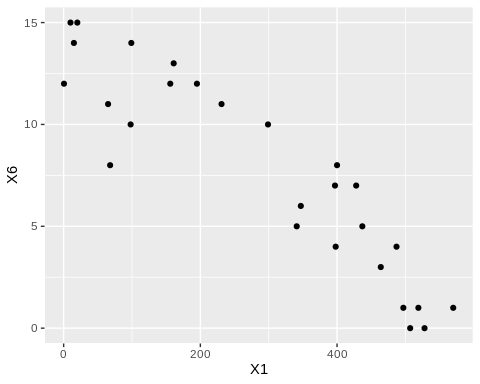
ggplot(data = rawdata) + geom\_point( mapping = aes(x = X1, y = X4))



ggplot(data = rawdata) + geom\_point( mapping = aes(x = X1, y = X5))



ggplot(data = rawdata) + geom\_point( mapping = aes(x = X1, y = X6))



lm.X2 = lm(X1 ~ X2, data = rawdata)  
lm.X2$coefficients

## (Intercept) X2   
## 2.577006 85.388873

lm.X2.eqn = paste0("X1 = ",  
 round(lm.X2$coefficients[1],3), " + ",   
 round(lm.X2$coefficients[2],3), " \* X2 ")  
lm.X2.eqn

## [1] "X1 = 2.577 + 85.389 \* X2 "

lm.X3 = lm(X1 ~ X3, data = rawdata)  
lm.X3$coefficients

## (Intercept) X3   
## -81.5043523 0.9499252

lm.X3.eqn = paste0("X1 = ",  
 round(lm.X3$coefficients[1],3), " + ",   
 round(lm.X3$coefficients[2],3), " \* X3 ")  
lm.X3.eqn

## [1] "X1 = -81.504 + 0.95 \* X3 "

lm.X4 = lm(X1 ~ X4, data = rawdata)  
lm.X4$coefficients

## (Intercept) X4   
## -90.14962 46.50910

lm.X4.eqn = paste0("X1 = ",  
 round(lm.X4$coefficients[1],3), " + ",   
 round(lm.X4$coefficients[2],3), " \* X4 ")  
lm.X4.eqn

## [1] "X1 = -90.15 + 46.509 \* X4 "

lm.X5 = lm(X1 ~ X5, data = rawdata)  
lm.X5$coefficients

## (Intercept) X5   
## -58.82321 35.63518

lm.X5.eqn = paste0("X1 = ",  
 round(lm.X5$coefficients[1],3), " + ",   
 round(lm.X5$coefficients[2],3), " \* X5 ")  
lm.X5.eqn

## [1] "X1 = -58.823 + 35.635 \* X5 "

lm.X6 = lm(X1 ~ X6, data = rawdata)  
lm.X6$coefficients

## (Intercept) X6   
## 563.59262 -35.78709

lm.X6.eqn = paste0("X1 = ",  
 round(lm.X6$coefficients[1],3), " + ",   
 round(lm.X6$coefficients[2],3), " \* X6 ")  
lm.X6.eqn

## [1] "X1 = 563.593 + -35.787 \* X6 "

lm = lm(X1 ~ ., data = rawdata)  
lm$coefficients

## (Intercept) X2 X3 X4 X5 X6   
## -18.8594142 16.2015736 0.1746352 11.5262690 13.5803129 -5.3109714

lm.eqn = paste0("X1 = ",  
 round(lm$coefficients[1],3), " + (",   
 round(lm$coefficients[2],3), " \* X2) + (",  
 round(lm$coefficients[3],3), " \* X3) + (",  
 round(lm$coefficients[4],3), " \* X4) + (",  
 round(lm$coefficients[5],3), " \* X5) + (",  
 round(lm$coefficients[2],3), " \* X6) ")  
lm.eqn

## [1] "X1 = -18.859 + (16.202 \* X2) + (0.175 \* X3) + (11.526 \* X4) + (13.58 \* X5) + (16.202 \* X6) "

predict(lm, newdata = data.frame( X2 = 3, X3 = 294, X4 = 8.2, X5 = 8.2, X6 = 11))

## 1   
## 228.5413