Afra's Document

Nana Akua Afra Owusu-Addo

2023-09-16

R Markdown

##QUESTION 1 ## Student Exam Results

```
Names=c("Yaw", "Adjoa", "Kofi", "Yaa", "Ama", "Kwadwo", "Abena")
Gender=c("Male", "Female", "Male", "Female", "Female", "Male", "Female")
Maths=c(82, 78, 70, NA, 72, 85, 88)
English=c(78, NA, 85, 88, 70, NA, 77)
Science=c(NA, 77, 87, 88, NA, 70, 78)
Social=c(88, 86, 72, 90, NA, 78, NA)
```

##Creating the Dataframe

```
StudentExamResults=data.frame(Names,Gender,Maths,English,Science,Social)
```

```
StudentExamResults
```

```
##    Names Gender Maths English Science Social
## 1    Yaw   Male    82      78      NA     88
## 2  Adjoa Female    78      NA      77     86
## 3   Kofi   Male    70      85      87     72
## 4    Yaa Female    NA      88      88     90
## 5    Ama Female    72      70      NA     NA
## 6 Kwadwo   Male    85      NA      70     78
## 7  Abena Female    88      77      78     NA
```

##Question 1a

```
summary.data.frame(StudentExamResults)
```

```
##     Names              Gender              Maths          English
##  Length:7           Length:7           Min.   :70.00   Min.   :70.0
##  Class :character   Class :character   1st Qu.:73.50   1st Qu.:77.0
##  Mode  :character   Mode  :character   Median :80.00   Median :78.0
##                                        Mean   :79.17   Mean   :79.6
##                                        3rd Qu.:84.25   3rd Qu.:85.0
##                                        Max.   :88.00   Max.   :88.0
##                                        NA's   :1       NA's   :2
##     Science       Social
##  Min.   :70   Min.   :72.0
##  1st Qu.:77   1st Qu.:78.0
##  Median :78   Median :86.0
##  Mean   :80   Mean   :82.8
##  3rd Qu.:87   3rd Qu.:88.0
##  Max.   :88   Max.   :90.0
##  NA's   :2    NA's   :2
```

##Question 1b

```
is.na(StudentExamResults)
```

```
##      Names Gender Maths English Science Social
## [1,] FALSE  FALSE FALSE   FALSE    TRUE  FALSE
## [2,] FALSE  FALSE FALSE    TRUE   FALSE  FALSE
## [3,] FALSE  FALSE FALSE   FALSE   FALSE  FALSE
## [4,] FALSE  FALSE  TRUE   FALSE   FALSE  FALSE
## [5,] FALSE  FALSE FALSE   FALSE    TRUE   TRUE
## [6,] FALSE  FALSE FALSE    TRUE   FALSE  FALSE
## [7,] FALSE  FALSE FALSE   FALSE   FALSE   TRUE
```

##Yes there are missing values in the dataset.

##Question 1c

```
sum(is.na(StudentExamResults))
```

```
## [1] 7
```

##Question 1d

```
na.exclude(StudentExamResults)
```

```
##   Names Gender Maths English Science Social
## 3  Kofi   Male    70      85      87     72
```

##Question 1E

```
StudentExamResults$Maths[is.na(StudentExamResults$Maths)] <- mean(StudentExamResults$Maths, n
a.rm = TRUE)
print(StudentExamResults)
```

```
##    Names Gender    Maths English Science Social
## 1    Yaw   Male 82.00000      78      NA     88
## 2  Adjoa Female 78.00000      NA      77     86
## 3   Kofi   Male 70.00000      85      87     72
## 4    Yaa Female 79.16667      88      88     90
## 5    Ama Female 72.00000      70      NA     NA
## 6 Kwadwo   Male 85.00000      NA      70     78
## 7  Abena Female 88.00000      77      78     NA
```

```
StudentExamResults$English[is.na(StudentExamResults$English)] <-mean(StudentExamResults$Engli
sh, na.rm = TRUE)
StudentExamResults$Science[is.na(StudentExamResults$Science)] <-mean(StudentExamResults$Scien
ce, na.rm = TRUE)
StudentExamResults$Social[is.na(StudentExamResults$Social)] <-mean(StudentExamResults$Social,
na.rm = TRUE)
print(StudentExamResults)
```

```
##    Names Gender    Maths English Science Social
## 1    Yaw   Male 82.00000    78.0      80   88.0
## 2  Adjoa Female 78.00000    79.6      77   86.0
## 3   Kofi   Male 70.00000    85.0      87   72.0
## 4    Yaa Female 79.16667    88.0      88   90.0
## 5    Ama Female 72.00000    70.0      80   82.8
## 6 Kwadwo   Male 85.00000    79.6      70   78.0
## 7  Abena Female 88.00000    77.0      78   82.8
```

##Question 1f

```
StudentExamResults$English[is.na(StudentExamResults$English)] <-median(StudentExamResults$Eng
lish, na.rm = TRUE)
StudentExamResults$Maths[is.na(StudentExamResults$Maths)] <-median(StudentExamResults$Maths,
na.rm = TRUE)
StudentExamResults$Social[is.na(StudentExamResults$Social)] <-median(StudentExamResults$Socia
l, na.rm = TRUE)
StudentExamResults$Science[is.na(StudentExamResults$Science)] <-median(StudentExamResults$Sci
ence, na.rm = TRUE)
print(StudentExamResults)
```

```
##    Names Gender    Maths English Science Social
## 1    Yaw   Male 82.00000    78.0      80   88.0
## 2  Adjoa Female 78.00000    79.6      77   86.0
## 3   Kofi   Male 70.00000    85.0      87   72.0
## 4    Yaa Female 79.16667    88.0      88   90.0
## 5    Ama Female 72.00000    70.0      80   82.8
## 6 Kwadwo   Male 85.00000    79.6      70   78.0
## 7  Abena Female 88.00000    77.0      78   82.8
```

##Question 2

```
mtcars
```

```
##                      mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4           21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag       21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710          22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive      21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout   18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant             18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360          14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 240D           24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
## Merc 230            22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## Merc 280            19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
## Merc 280C           17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
## Merc 450SE          16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## Merc 450SL          17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## Merc 450SLC         15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## Cadillac Fleetwood  10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
## Chrysler Imperial   14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Fiat 128            32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
## Honda Civic         30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
## Toyota Corolla      33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
## Toyota Corona       21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
## Dodge Challenger    15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
## AMC Javelin         15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
## Camaro Z28          13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## Pontiac Firebird    19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## Fiat X1-9           27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
## Porsche 914-2       26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
## Lotus Europa        30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
## Ford Pantera L      15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
## Ferrari Dino        19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## Maserati Bora       15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## Volvo 142E          21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
```

```
library(corrr)
```

```
library(ggplot2)
```

```
library(ggcorrplot)
```

```
library(FactoMineR)
```

```
library("readxl")
```

##Data Type

```
str(mtcars)
```

```
## '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 ...
```

```
mtcars<-mtcars[,1:7]
mtcars
```

```
##                     mpg cyl  disp  hp drat    wt  qsec
## Mazda RX4          21.0   6 160.0 110 3.90 2.620 16.46
## Mazda RX4 Wag      21.0   6 160.0 110 3.90 2.875 17.02
## Datsun 710         22.8   4 108.0  93 3.85 2.320 18.61
## Hornet 4 Drive     21.4   6 258.0 110 3.08 3.215 19.44
## Hornet Sportabout  18.7   8 360.0 175 3.15 3.440 17.02
## Valiant            18.1   6 225.0 105 2.76 3.460 20.22
## Duster 360         14.3   8 360.0 245 3.21 3.570 15.84
## Merc 240D          24.4   4 146.7  62 3.69 3.190 20.00
## Merc 230           22.8   4 140.8  95 3.92 3.150 22.90
## Merc 280           19.2   6 167.6 123 3.92 3.440 18.30
## Merc 280C          17.8   6 167.6 123 3.92 3.440 18.90
## Merc 450SE         16.4   8 275.8 180 3.07 4.070 17.40
## Merc 450SL         17.3   8 275.8 180 3.07 3.730 17.60
## Merc 450SLC        15.2   8 275.8 180 3.07 3.780 18.00
## Cadillac Fleetwood 10.4   8 472.0 205 2.93 5.250 17.98
## Lincoln Continental 10.4  8 460.0 215 3.00 5.424 17.82
## Chrysler Imperial  14.7   8 440.0 230 3.23 5.345 17.42
## Fiat 128           32.4   4  78.7  66 4.08 2.200 19.47
## Honda Civic        30.4   4  75.7  52 4.93 1.615 18.52
## Toyota Corolla     33.9   4  71.1  65 4.22 1.835 19.90
## Toyota Corona      21.5   4 120.1  97 3.70 2.465 20.01
## Dodge Challenger   15.5   8 318.0 150 2.76 3.520 16.87
## AMC Javelin        15.2   8 304.0 150 3.15 3.435 17.30
## Camaro Z28         13.3   8 350.0 245 3.73 3.840 15.41
## Pontiac Firebird   19.2   8 400.0 175 3.08 3.845 17.05
## Fiat X1-9          27.3   4  79.0  66 4.08 1.935 18.90
## Porsche 914-2      26.0   4 120.3  91 4.43 2.140 16.70
## Lotus Europa       30.4   4  95.1 113 3.77 1.513 16.90
## Ford Pantera L     15.8   8 351.0 264 4.22 3.170 14.50
## Ferrari Dino       19.7   6 145.0 175 3.62 2.770 15.50
## Maserati Bora      15.0   8 301.0 335 3.54 3.570 14.60
## Volvo 142E         21.4   4 121.0 109 4.11 2.780 18.60
```

##rem vs am

```
colSums(is.na(mtcars))
```

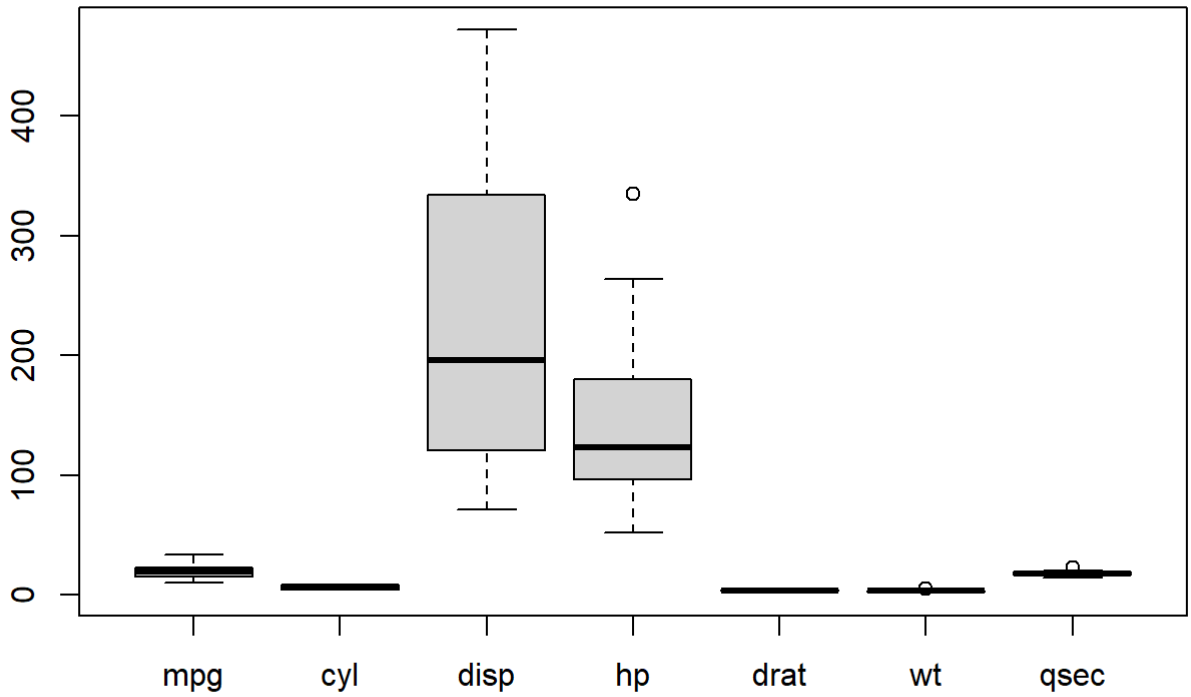
```
##  mpg  cyl disp   hp drat   wt qsec
##    0    0    0    0    0    0    0
```

missing variable

```
head(mtcars)
```

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

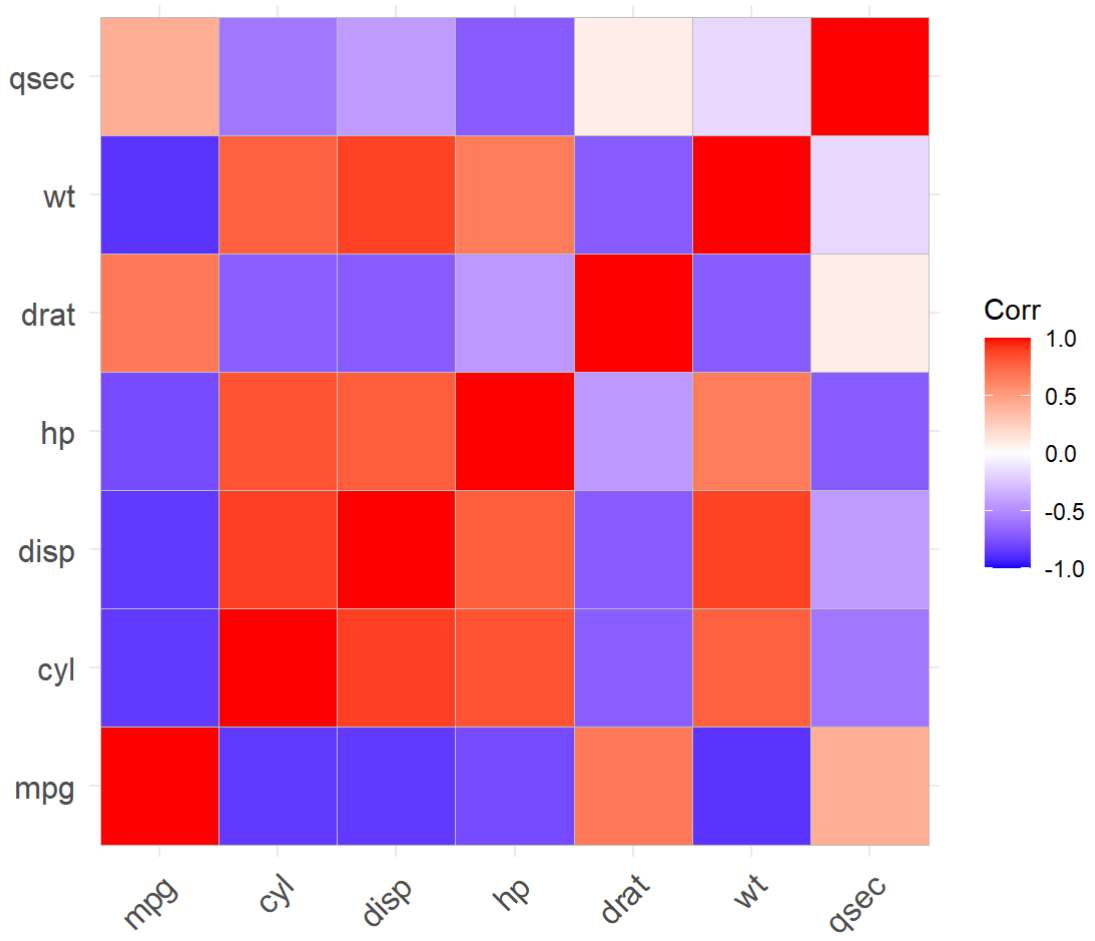
```
boxplot(mtcars)
```




```
data_normalized<-scale(mtcars)
head(data_normalized)
```

```
##                           mpg        cyl        disp         hp       drat
## Mazda RX4           0.1508848 -0.1049878 -0.57061982 -0.5350928  0.5675137
## Mazda RX4 Wag       0.1508848 -0.1049878 -0.57061982 -0.5350928  0.5675137
## Datsun 710          0.4495434 -1.2248578 -0.99018209 -0.7830405  0.4739996
## Hornet 4 Drive      0.2172534 -0.1049878  0.22009369 -0.5350928 -0.9661175
## Hornet Sportabout  -0.2307345  1.0148821  1.04308123  0.4129422 -0.8351978
## Valiant            -0.3302874 -0.1049878 -0.04616698 -0.6080186 -1.5646078
##                           wt       qsec
## Mazda RX4         -0.610399567 -0.7771651
## Mazda RX4 Wag     -0.349785269 -0.4637808
## Datsun 710        -0.917004624  0.4260068
## Hornet 4 Drive    -0.002299538  0.8904872
## Hornet Sportabout  0.227654255 -0.4637808
## Valiant            0.248094592  1.3269868
```

```
corr_matrix<-cor(data_normalized)
ggcorrplot(corr_matrix)
```




```
data.pca<-princomp(corr_matrix)
summary(data.pca)
```

```
## Importance of components:
##                           Comp.1     Comp.2      Comp.3       Comp.4
## Standard deviation     1.8649595 0.43576447 0.113625670 0.0502191811
## Proportion of Variance 0.9438274 0.05152968 0.003503537 0.0006843733
## Cumulative Proportion  0.9438274 0.99535710 0.998860634 0.9995450075
##                              Comp.5       Comp.6       Comp.7
## Standard deviation     0.0355057006 0.0203967195 1.593461e-08
## Proportion of Variance 0.0003420975 0.0001128949 6.890280e-17
## Cumulative Proportion  0.9998871051 1.0000000000 1.000000e+00
```

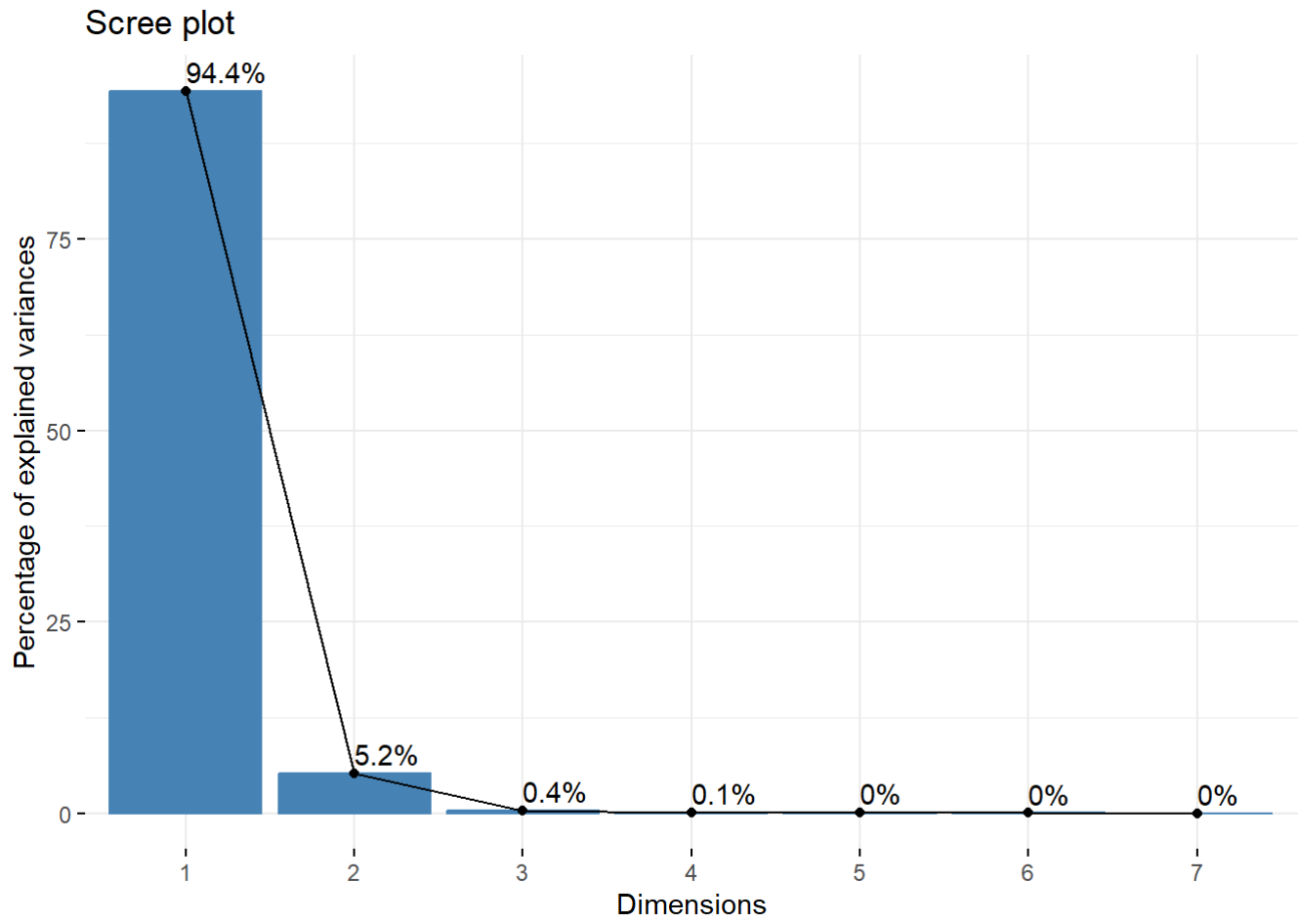
```
data.pca$loadings[,1:2]
```

```
##          Comp.1      Comp.2
## mpg   0.4147089  0.08799510
## cyl  -0.4250016  0.07636210
## disp -0.4184468 -0.07914108
## hp   -0.3862459  0.33967115
## drat  0.3364781  0.46133393
## wt   -0.3855243 -0.31782102
## qsec  0.2475077 -0.74225784
```

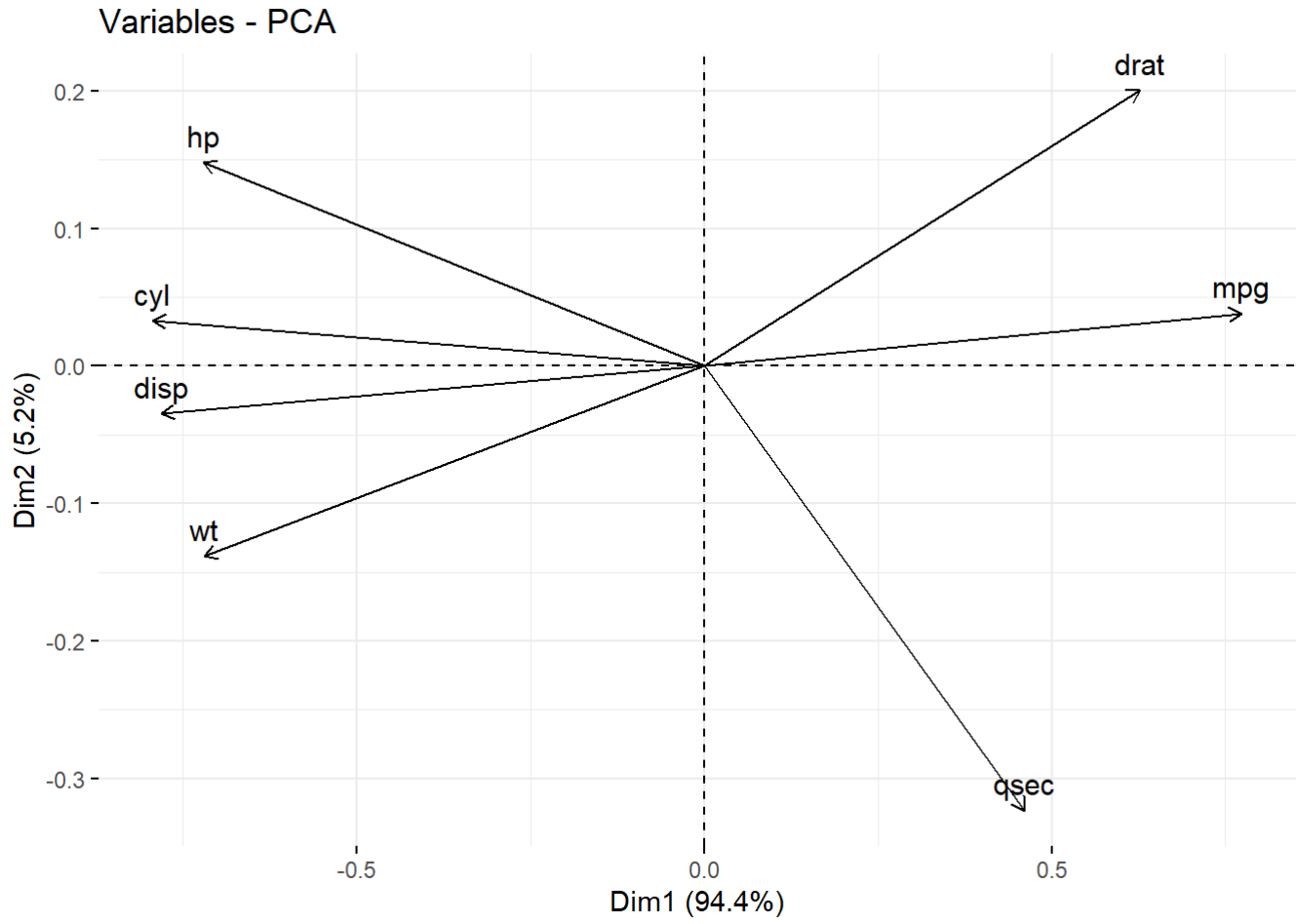
```
library(factoextra)
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

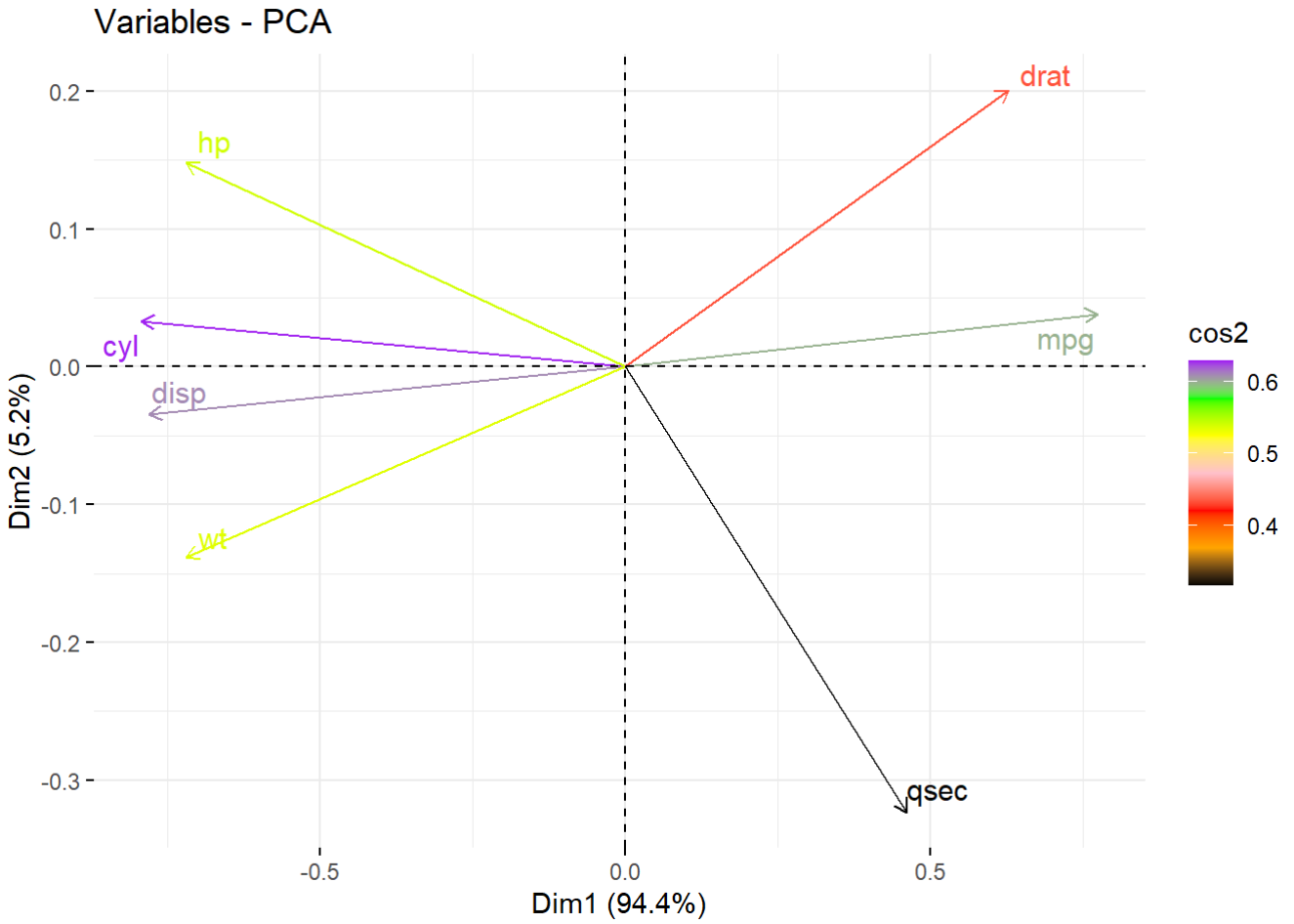
```
fviz_eig(data.pca, addlabels = TRUE)
```



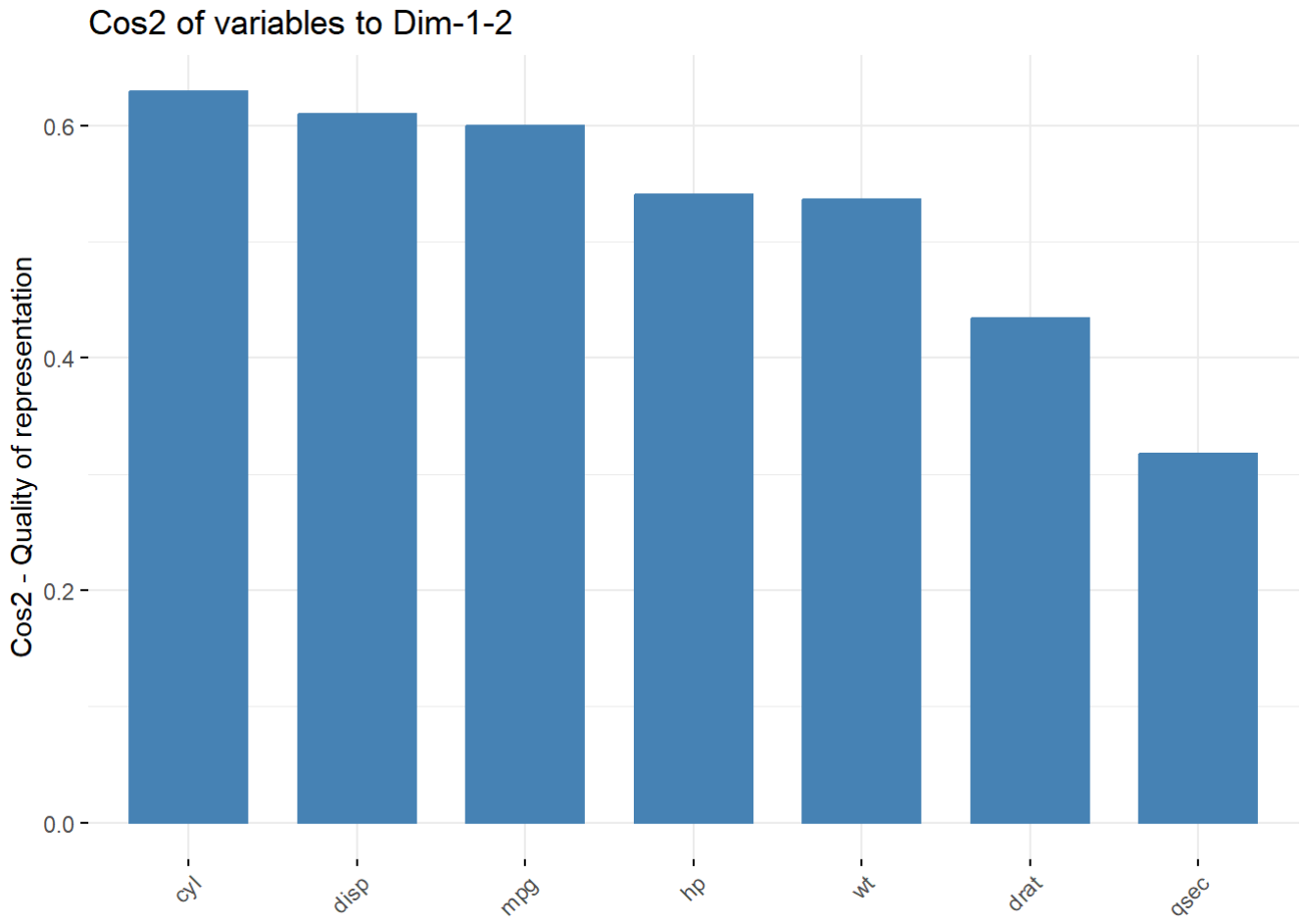

```
fviz_pca_var(data.pca, col.var = "black")
```



```
fviz_pca_var(data.pca, col.var = "cos2", gradient.cols=c("black", "orange", "red", "pink", "yellow", "green", "purple"),repel = TRUE)
```




```
fviz_cos2(data.pca, choice="var", axes=1:2)
```


Cos2 of variables to Dim-1-2
Cos2 - Quality of representation
0.0
0.2
0.4
0.6
cyl
disp
mpg
hp
wt
drat
qsec