MATH 189 HW1

Zijian Su
Zelong Zhou
Xiangyi Lin

Last Updated: January 16, 2023

Concrete contributions

All problems were done by Zijian Su, Zelong Zhou, Xiangyi Lin. All contributing equally to this assignment. Everyone put in enough effort.

Overview

The Motor Trend Car Road Tests dataset (mtcars.csv) contains the data extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models). The dataset contains 32 observations on 11 variables. Analyze the dataset according to the following steps:

Packages

```
#install.packages("rmarkdown")
#install.packages("ggplot2")

#tinytex::install_tinytex()

#install.packages("scatterplot3d")
```

Question 1

Calculate sample mean and sample variance of each variable.

Answer:

Means:

```
#get the data from csv file
data <- read.csv("./mtcars.csv")

#remove the first column
```

```r
data_ <-data[,-1]

#Output the average in order
for (i in colnames(data_)){
  m <- mean(data[[i]])
  print(paste(i,": ",m))
}
```

```
## [1] "mpg :  20.090625"
## [1] "cyl :  6.1875"
## [1] "disp :  230.721875"
## [1] "hp :  146.6875"
## [1] "drat :  3.5965625"
## [1] "wt :  3.21725"
## [1] "qsec :  17.84875"
## [1] "vs :  0.4375"
## [1] "am :  0.40625"
## [1] "gear :  3.6875"
## [1] "carb :  2.8125"
```

Variance:

```r
#Output the variance in order
for (i in colnames(data_)){
  v <- var(data[[i]])
  print(paste(i,": ",v))
}
```

```
## [1] "mpg :  36.3241028225806"
## [1] "cyl :  3.18951612903226"
## [1] "disp :  15360.799828629"
## [1] "hp :  4700.86693548387"
## [1] "drat :  0.285881350806452"
## [1] "wt :  0.957378967741936"
## [1] "qsec :  3.19316612903226"
## [1] "vs :  0.254032258064516"
## [1] "am :  0.248991935483871"
## [1] "gear :  0.544354838709677"
## [1] "carb :  2.60887096774194"
```

Question 2

Calculate the sample variance-covariance matrix and the sample correlation matrix. What can you say about the variance-covariance matrix and correlation matrix?

Answer:

Variance-covariance matrix:

```
cov(data_)
```

```
##              mpg         cyl        disp          hp         drat          wt
## mpg    36.324103  -9.1723790  -633.09721 -320.732056   2.19506351  -5.1166847
## cyl    -9.172379   3.1895161   199.66028  101.931452  -0.66836694   1.3673710
## disp -633.097208 199.6602823 15360.79983 6721.158669 -47.06401915 107.6842040
## hp   -320.732056 101.9314516  6721.15867 4700.866935 -16.45110887  44.1926613
## drat    2.195064  -0.6683669   -47.06402  -16.451109   0.28588135  -0.3727207
## wt     -5.116685   1.3673710   107.68420   44.192661  -0.37272073   0.9573790
## qsec    4.509149  -1.8868548   -96.05168  -86.770081   0.08714073  -0.3054816
## vs      2.017137  -0.7298387   -44.37762  -24.987903   0.11864919  -0.2736613
## am      1.803931  -0.4657258   -36.56401   -8.320565   0.19015121  -0.3381048
## gear    2.135685  -0.6491935   -50.80262   -6.358871   0.27598790  -0.4210806
## carb   -5.363105   1.5201613    79.06875   83.036290  -0.07840726   0.6757903
##              qsec          vs          am        gear        carb
## mpg    4.50914919   2.01713710   1.80393145   2.1356855 -5.36310484
## cyl   -1.88685484  -0.72983871  -0.46572581  -0.6491935  1.52016129
## disp -96.05168145 -44.37762097 -36.56401210 -50.8026210 79.06875000
## hp   -86.77008065 -24.98790323  -8.32056452  -6.3588710 83.03629032
## drat   0.08714073   0.11864919   0.19015121   0.2759879 -0.07840726
## wt    -0.30548161  -0.27366129  -0.33810484  -0.4210806  0.67579032
## qsec   3.19316613   0.67056452  -0.20495968  -0.2804032 -1.89411290
## vs     0.67056452   0.25403226   0.04233871   0.0766129 -0.46370968
## am    -0.20495968   0.04233871   0.24899194   0.2923387  0.04637097
## gear  -0.28040323   0.07661290   0.29233871   0.5443548  0.32661290
## carb  -1.89411290  -0.46370968   0.04637097   0.3266129  2.60887097
```

Correlation matrix:

```
cor(data_)
```

```
##             mpg        cyl       disp         hp        drat         wt
## mpg   1.0000000 -0.8521620 -0.8475514 -0.7761684  0.68117191 -0.8676594
## cyl  -0.8521620  1.0000000  0.9020329  0.8324475 -0.69993811  0.7824958
## disp -0.8475514  0.9020329  1.0000000  0.7909486 -0.71021393  0.8879799
## hp   -0.7761684  0.8324475  0.7909486  1.0000000 -0.44875912  0.6587479
## drat  0.6811719 -0.6999381 -0.7102139 -0.4487591  1.00000000 -0.7124406
## wt   -0.8676594  0.7824958  0.8879799  0.6587479 -0.71244065  1.0000000
## qsec  0.4186840 -0.5912421 -0.4336979 -0.7082234  0.09120476 -0.1747159
## vs    0.6640389 -0.8108118 -0.7104159 -0.7230967  0.44027846 -0.5549157
## am    0.5998324 -0.5226070 -0.5912270 -0.2432043  0.71271113 -0.6924953
## gear  0.4802848 -0.4926866 -0.5555692 -0.1257043  0.69961013 -0.5832870
## carb -0.5509251  0.5269883  0.3949769  0.7498125 -0.09078980  0.4276059
```

```
##             qsec         vs          am       gear        carb
## mpg   0.41868403  0.6640389  0.59983243  0.4802848 -0.55092507
## cyl  -0.59124207 -0.8108118 -0.52260705 -0.4926866  0.52698829
## disp -0.43369788 -0.7104159 -0.59122704 -0.5555692  0.39497686
## hp   -0.70822339 -0.7230967 -0.24320426 -0.1257043  0.74981247
## drat  0.09120476  0.4402785  0.71271113  0.6996101 -0.09078980
## wt   -0.17471588 -0.5549157 -0.69249526 -0.5832870  0.42760594
## qsec  1.00000000  0.7445354 -0.22986086 -0.2126822 -0.65624923
## vs    0.74453544  1.0000000  0.16834512  0.2060233 -0.56960714
## am   -0.22986086  0.1683451  1.00000000  0.7940588  0.05753435
## gear -0.21268223  0.2060233  0.79405876  1.0000000  0.27407284
## carb -0.65624923 -0.5696071  0.05753435  0.2740728  1.00000000
```

We can know from the **Variance-covariance matrix** whether the relationship between the data is proportional or inverse.

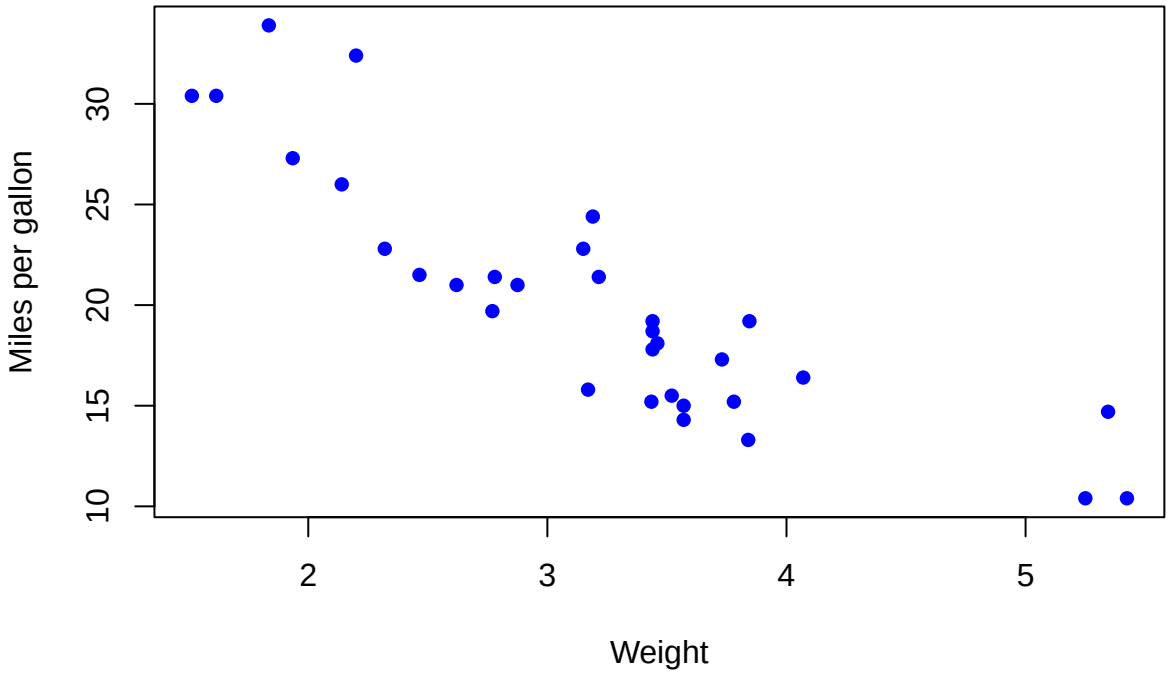
We can find the relationship between variables and variables from the **Correlation matrix**, and the degree to which they influence each other.

Question 3

Draw a scatter plot between wt (Weight) and mpg (Miles per gallon).

Answer:

```
wt <- data$wt
mpg <- data$mpg
plot(wt, mpg, pch = 16, col = "BLUE",xlab="Weight",ylab="Miles per gallon")
```



Question 4

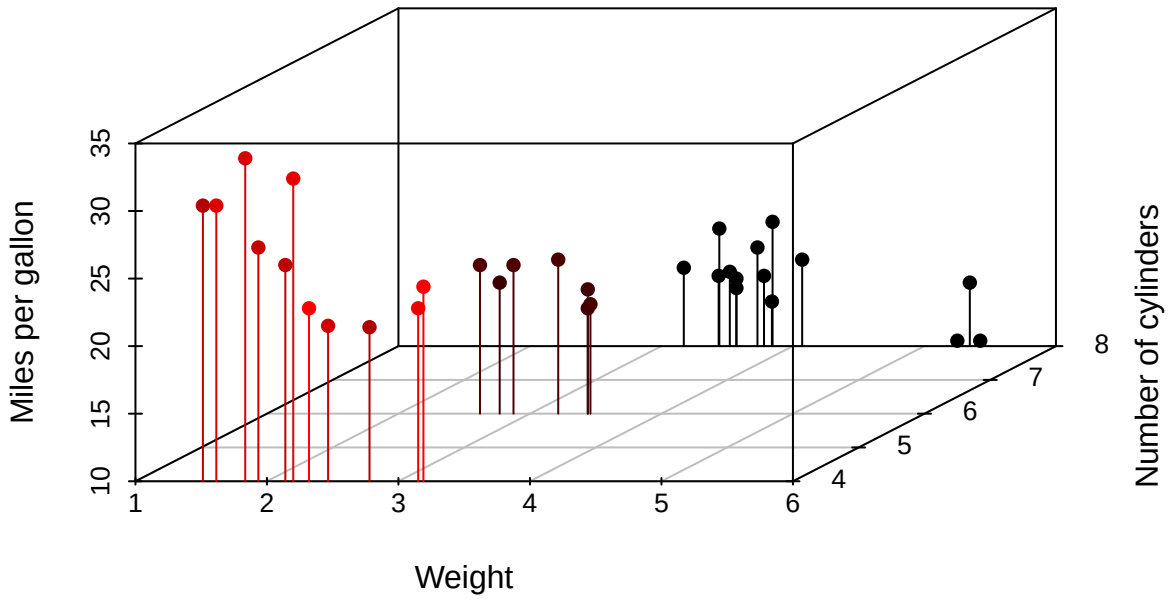
Draw a scatter plot to show the relationship between wt (Weight), mpg (Miles per gallon) and cyl (Number of cylinders). You can use 3D scatter plot or add cyl to your 2D scatter plot as the color of points.

Answer:

```
library("scatterplot3d")
```

```
## Warning: package 'scatterplot3d' was built under R version 4.1.3
```

```
wt <- data$wt
mpg <- data$mpg
cyl <- data$cyl
scatterplot3d(wt,cyl,mpg,pch = 16,angle = 45,xlab="Weight",ylab="Number of cylinders",zlab="Miles per ga
```

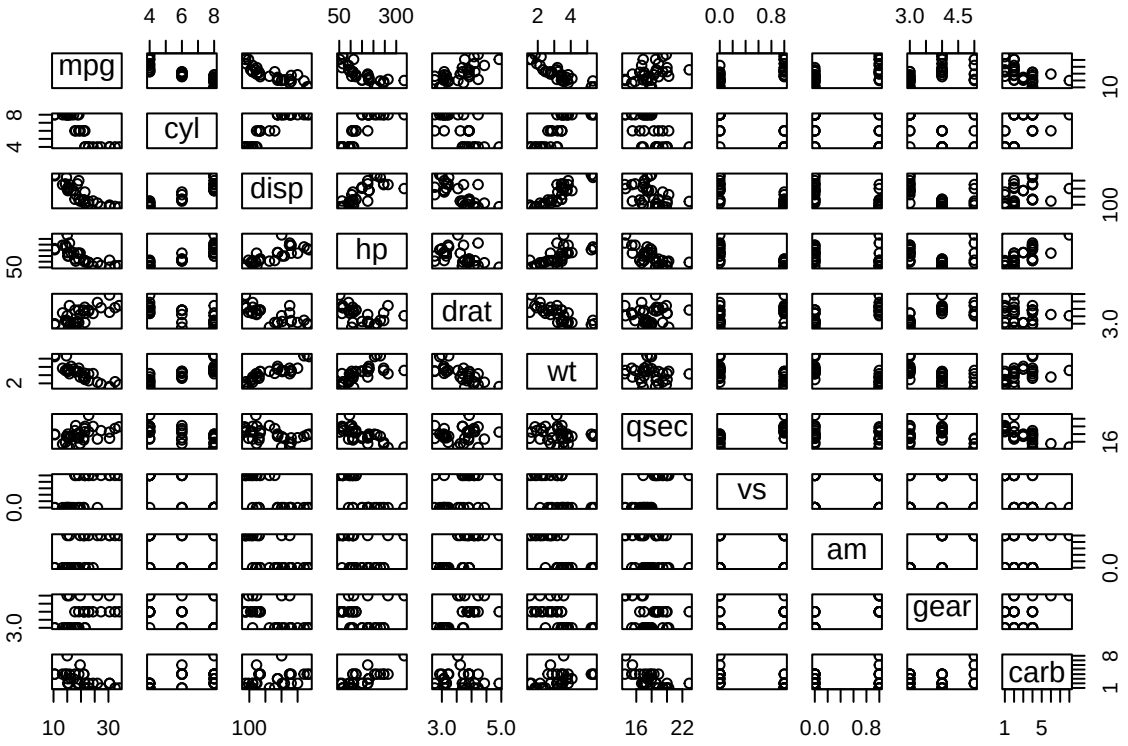


Question 5

Draw pairwise scatter plot for all variables.

Answer:

```
pairs(data_)
```



Question 6

One engineer suggests that the relationship between wt and mpg is subject to the number of cylinders. According to the plot you draw in 4, what is your opinion towards this suggestion?

Answer:

I think his idea is reasonable. From the plot of Q4, we can know that the weight of cyl=4 vehicles is relatively light, and the mpg is relatively high. And cly=8 car weight is relatively large, mpg is relatively low. A simple conclusion can be drawn: the more cyl, the heavier the car and the lower the mpg.