# 1. Get to know the data

(a) Read the data and **print a summary**.

(b) Notice that `class` has been read as numeric. Convert it to a factor using
   `vehdata$class = factor(vehdata$class, labels=c("2D", "4D", "BUS", "VAN"))`
   Show a summary of this new version of `class`.

(c) **Print the correlation matrix for the explanatory variables. Comment on any strong correlations (maybe beyond ±0.7 and note especially any beyond ±0.9)**

```
 1  # Title: STAT 452 Exercise 8 L17A1
 2  # Author: Injun Son
 3  # Date: November 21, 2020
 4
 5  library(dplyr)
 6  library(MASS)    # For ridge regression
 7  library(glmnet) # For LASSO
 8  library(nnet) # Fits neural net models
 9  library(rgl)
10  library(FNN)
11  source("Helper Functions.R")
12
13
14  ### Some of our code will be random, so we have to set the seed.
15  ### Use Mersenne-Twister for compatibility.
16  set.seed(46536737, kind = "Mersenne-Twister")
17
18  # 1. Get to know the data
19  # (a) Read the data and print a summary.
20  vehdata = read.csv("vehicle.csv")
21  print(summary(vehdata))
```

```
> print(summary(vehdata))
  Compactness      Circularity    Distance.Circularity Radius.Ratio    Pr.Axis.Aspect.Ratio Max.Length.Aspect.Ratio Scatter.Ratio
 Min.   : 73.00   Min.   :33.00   Min.   : 40.00       Min.   :104.0   Min.   : 47.00       Min.   : 2.000          Min.   :112.0
 1st Qu.: 87.00   1st Qu.:40.00   1st Qu.: 70.00       1st Qu.:141.0   1st Qu.: 57.00       1st Qu.: 7.000          1st Qu.:146.2
 Median : 93.00   Median :44.00   Median : 80.00       Median :167.0   Median : 61.00       Median : 8.000          Median :157.0
 Mean   : 93.68   Mean   :44.86   Mean   : 82.09       Mean   :168.9   Mean   : 61.69       Mean   : 8.567          Mean   :168.8
 3rd Qu.:100.00   3rd Qu.:49.00   3rd Qu.: 98.00       3rd Qu.:195.0   3rd Qu.: 65.00       3rd Qu.:10.000          3rd Qu.:198.0
 Max.   :119.00   Max.   :59.00   Max.   :112.00       Max.   :333.0   Max.   :138.00       Max.   :55.000          Max.   :265.0
  Elongatedness   Pr.Axis.Rectangularity Max.Length.Rectangularity Scaled.Variance.Along.Major.Axis Scaled.Variance.Along.Minor.Axis
 Min.   :26.00   Min.   :17.00           Min.   :118               Min.   :130.0                    Min.   : 184.0
 1st Qu.:33.00   1st Qu.:19.00           1st Qu.:137               1st Qu.:167.0                    1st Qu.: 318.2
 Median :43.00   Median :20.00           Median :146               Median :178.5                    Median : 364.0
 Mean   :40.93   Mean   :20.58           Mean   :148               Mean   :188.6                    Mean   : 439.9
 3rd Qu.:46.00   3rd Qu.:23.00           3rd Qu.:159               3rd Qu.:217.0                    3rd Qu.: 587.0
 Max.   :61.00   Max.   :29.00           Max.   :188               Max.   :320.0                    Max.   :1018.0
 Scaled.Radius.of.Gyration Skewness.About.Major.Axis Skewness.About.Minor.Axis Kurtosis.About.Minor.Axis Kurtosis.About.Major.Axis
 Min.   :109.0             Min.   : 59.00            Min.   : 0.000            Min.   : 0.0              Min.   :176.0
 1st Qu.:149.0             1st Qu.: 67.00            1st Qu.: 2.000            1st Qu.: 5.0              1st Qu.:184.0
 Median :173.0             Median : 71.50            Median : 6.000            Median :11.0             Median :188.0
 Mean   :174.7             Mean   : 72.46            Mean   : 6.377            Mean   :12.6             Mean   :188.9
 3rd Qu.:198.0             3rd Qu.: 75.00            3rd Qu.: 9.000            3rd Qu.:19.0             3rd Qu.:193.0
 Max.   :268.0             Max.   :135.00            Max.   :22.000            Max.   :41.0             Max.   :206.0
 Hollows.Ratio        class
 Min.   :181.0   Min.   :1.000
 1st Qu.:190.2   1st Qu.:1.250
 Median :197.0   Median :2.000
 Mean   :195.6   Mean   :2.478
 3rd Qu.:201.0   3rd Qu.:3.000
 Max.   :211.0   Max.   :4.000
```

```r
#(b) Notice that class has been read as numeric. Convert it to a factor using
# vehdata$class = factor(vehdata$class, labels=c("2D", "4D", "BUS", "VAN"))
# Show a summary of this new version of class.
vehdata$class = factor(vehdata$class, labels=c("2D", "4D", "BUS", "VAN"))
print(summary(vehdata$class))
```

```
 class
 2D :212
 4D :217
 BUS:218
 VAN:199
```

```r
29  # (c) Print the correlation matrix for the explanatory variables. Comment
30  # on any strong correlations (maybe beyond ±0.7 and note especially any
31  #                           beyond ±0.9)
32  mydata.cor = cor(vehdata[1:18], method = c("spearman"))
33  mydata.cor
```

```
> mydata.cor
                             Compactness Circularity Distance.Circularity Radius.Ratio Pr.Axis.Aspect.Ratio Max.Length.Aspect.Ratio
Compactness                    1.0000000  0.65436193           0.77660300    0.74637159           0.24005615              0.46260544
Circularity                    0.6543619  1.00000000           0.76256334    0.62954291           0.23190953              0.52787244
Distance.Circularity           0.7766030  0.76256334           1.00000000    0.80000737           0.28032255              0.62469082
Radius.Ratio                   0.7463716  0.62954291           0.80000737    1.00000000           0.68336888              0.46759998
Pr.Axis.Aspect.Ratio           0.2400562  0.23190953           0.28032255    0.68336888           1.00000000              0.20027187
Max.Length.Aspect.Ratio        0.4626054  0.52787244           0.62469082    0.46759998           0.20027187              1.00000000
Scatter.Ratio                  0.7749861  0.81564187           0.89605615    0.80385607           0.27680450              0.49969714
Elongatedness                 -0.7902441 -0.80101133          -0.89261674   -0.83308108          -0.31975330             -0.46894194
Pr.Axis.Rectangularity         0.7744207  0.82127905           0.89564202    0.78693933           0.25398372              0.50605547
Max.Length.Rectangularity      0.6567227  0.96795279           0.74751284    0.57473370           0.18322821              0.61037428
Scaled.Variance.Along.Major.Axis  0.7372092  0.78125491        0.86554455    0.83681424           0.32823063              0.46011868
Scaled.Variance.Along.Minor.Axis  0.7840289  0.80971828        0.89601324    0.81999457           0.30012371              0.48641318
Scaled.Radius.of.Gyration      0.5494204  0.94686062           0.67442589    0.55100253           0.19184677              0.41640178
Skewness.About.Major.Axis     -0.2886566  0.10072692          -0.21857364   -0.34994987          -0.25918310             -0.25138798
Skewness.About.Minor.Axis      0.1805705  0.14670658           0.09875452    0.04106367          -0.04800746              0.06114411
Kurtosis.About.Minor.Axis      0.1491559 -0.03156559           0.25250089    0.16459603          -0.02744796              0.13418976
Kurtosis.About.Major.Axis      0.3656290 -0.08996483           0.18988598    0.44559818           0.40428405              0.07892698
Hollows.Ratio                  0.4047055  0.03051612           0.32512591    0.48960693           0.39802728              0.36285398

                             Scatter.Ratio Elongatedness Pr.Axis.Rectangularity Max.Length.Rectangularity Scaled.Variance.Along.Major.Axis
Compactness                     0.77498614   -0.79024408             0.77442075               0.656722717                       0.737209208
Circularity                     0.81564187   -0.80101133             0.82127905               0.967952792                       0.781254909
Distance.Circularity            0.89605615   -0.89261674             0.89564202               0.747512842                       0.865544326
Radius.Ratio                    0.80385607   -0.83308108             0.78693933               0.574733696                       0.836814238
Pr.Axis.Aspect.Ratio            0.27680450   -0.31975330             0.25398372               0.183228208                       0.328230632
Max.Length.Aspect.Ratio         0.49969714   -0.46894194             0.50605547               0.610374284                       0.460118677
Scatter.Ratio                   1.00000000   -0.99166554             0.98363531               0.780775878                       0.956025199
Elongatedness                  -0.99166554    1.00000000            -0.97413042              -0.756358543                      -0.960603859
Pr.Axis.Rectangularity          0.98363531   -0.97413042             1.00000000               0.789824756                       0.938814313
Max.Length.Rectangularity       0.78077588   -0.75635854             0.78982476               1.000000000                       0.730371934
Scaled.Variance.Along.Major.Axis  0.95602520   -0.96060386          0.93881431               0.730371934                       1.000000000
Scaled.Variance.Along.Minor.Axis  0.99799958   -0.99679587          0.98065734               0.770336971                       0.959532618
Scaled.Radius.of.Gyration       0.75586708   -0.74193240             0.76227449               0.890518341                       0.742421148
Skewness.About.Major.Axis      -0.03619881    0.07876114            -0.02582493               0.087783216                       0.008297668
Skewness.About.Minor.Axis       0.05489723   -0.05502282             0.06369624               0.136095269                       0.025739012
Kurtosis.About.Minor.Axis       0.19849953   -0.18060919             0.19741490              -0.006077339                       0.191718529
Kurtosis.About.Major.Axis       0.08155343   -0.14406412             0.05953476              -0.092115189                       0.084881974
Hollows.Ratio                   0.17547271   -0.21895243             0.15695197               0.059069425                       0.144872010
```

| | Scaled.Variance.Along.Minor.Axis | Scaled.Radius.of.Gyration | Skewness.About.Major.Axis | Skewness.About.Minor.Axis |
|---|---|---|---|---|
| Compactness | 0.78402886 | 0.54942039 | -0.288656615 | 0.18057045 |
| Circularity | 0.80971828 | 0.94686062 | 0.100726923 | 0.14670658 |
| Distance.Circularity | 0.89601324 | 0.67442589 | -0.218573639 | 0.09875452 |
| Radius.Ratio | 0.81999457 | 0.55100253 | -0.349949866 | 0.04106367 |
| Pr.Axis.Aspect.Ratio | 0.30012371 | 0.19184677 | -0.259183102 | -0.04800746 |
| Max.Length.Aspect.Ratio | 0.48641318 | 0.41640178 | -0.251387983 | 0.06114411 |
| Scatter.Ratio | 0.99799958 | 0.75586708 | -0.036198814 | 0.05489723 |
| Elongatedness | -0.99679587 | -0.74193240 | 0.078761145 | -0.05502282 |
| Pr.Axis.Rectangularity | 0.98065734 | 0.76227449 | -0.025824934 | 0.06369624 |
| Max.Length.Rectangularity | 0.77033697 | 0.89051834 | 0.087783216 | 0.13609527 |
| Scaled.Variance.Along.Major.Axis | 0.95953262 | 0.74242115 | 0.008297668 | 0.02573901 |
| Scaled.Variance.Along.Minor.Axis | 1.00000000 | 0.74930141 | -0.058578783 | 0.05398693 |
| Scaled.Radius.of.Gyration | 0.74930141 | 1.00000000 | 0.216405061 | 0.17616403 |
| Skewness.About.Major.Axis | -0.05857878 | 0.21640506 | 1.000000000 | -0.06970820 |
| Skewness.About.Minor.Axis | 0.05398693 | 0.17616403 | -0.069708202 | 1.00000000 |
| Kurtosis.About.Minor.Axis | 0.19031627 | -0.06134846 | -0.099290252 | -0.02919089 |
| Kurtosis.About.Major.Axis | 0.11418929 | -0.19493296 | -0.863739778 | 0.08903370 |
| Hollows.Ratio | 0.19923711 | -0.09770540 | -0.896347051 | 0.06948044 |

| | Kurtosis.About.Minor.Axis | Kurtosis.About.Major.Axis | Hollows.Ratio |
|---|---|---|---|
| Compactness | 0.149155933 | 0.36562898 | 0.40470548 |
| Circularity | -0.031565595 | -0.08996483 | 0.03051612 |
| Distance.Circularity | 0.252500892 | 0.18988598 | 0.32512591 |
| Radius.Ratio | 0.164596032 | 0.44559818 | 0.48960693 |
| Pr.Axis.Aspect.Ratio | -0.027447961 | 0.40428405 | 0.39802728 |
| Max.Length.Aspect.Ratio | 0.134189760 | 0.07892698 | 0.36285398 |
| Scatter.Ratio | 0.198499531 | 0.08155343 | 0.17547271 |
| Elongatedness | -0.180609189 | -0.14406412 | -0.21895243 |
| Pr.Axis.Rectangularity | 0.197414902 | 0.05953476 | 0.15695197 |
| Max.Length.Rectangularity | -0.006077339 | -0.09211519 | 0.05906942 |
| Scaled.Variance.Along.Major.Axis | 0.191718529 | 0.08488197 | 0.14487201 |
| Scaled.Variance.Along.Minor.Axis | 0.190316270 | 0.11418929 | 0.19923711 |
| Scaled.Radius.of.Gyration | -0.061348462 | -0.19493296 | -0.09770540 |
| Skewness.About.Major.Axis | -0.099290252 | -0.86373978 | -0.89634705 |
| Skewness.About.Minor.Axis | -0.029190891 | 0.08903370 | 0.06948044 |
| Kurtosis.About.Minor.Axis | 1.000000000 | 0.06690761 | 0.17861724 |
| Kurtosis.About.Major.Axis | 0.066907607 | 1.00000000 | 0.89403274 |
| Hollows.Ratio | 0.178617236 | 0.89403274 | 1.00000000 |
| Hollows.Ratio | 0.17547271   -0.21895243 | 0.15695197 | 0.059069425                    0.144872010 |

It looks like Compactness, Circularity, Distance.Circularity, Radius.Ratio, Scatter.Ratio, Pr.Axis.Rectangularity , Max.Length.Rectangularity, Scaled.Variance.Along.Minor.Axis have quite many correlation with other variables.