2. Split the data using the code below, where set1 will be the training set for future analyses and set2 the test set:
set . seed (46685326 , kind =" Mersenne - Twister ")
perm <- sample (x= nrow (vehdata))
set1 <- vehdata [which (perm <= 3* nrow (vehdata)/4) ,]
set2 <- vehdata [which (perm > 3* nrow (vehdata)/4) ,]

```
#####
# 2. Split the data using the code below, where set1 will be the training set for future
# analyses and set2 the test set:
set.seed(46685326, kind = "Mersenne-Twister")
perm <- sample (x= nrow ( vehdata ))
set1 <- vehdata [ which ( perm <= 3* nrow ( vehdata )/4) , ]
set2 <- vehdata [ which ( perm > 3* nrow ( vehdata )/4) , ]

print(head(set1, 6))
print(head(set2, 6))
```

```
> print(head(set1, 6))
  Compactness Circularity Distance.Circularity Radius.Ratio Pr.Axis.Aspect.Ratio Max.Length.Aspect.Ratio Scatter.Ratio Elongatedness Pr.Axis.Rectangularity Max.Length.Rectangularity
1          95          48                  83          178                  72                     10           162           42                    20                     159
2          91          41                  84          141                  57                      9           149           45                    19                     143
3         104          50                 106          209                  66                     10           207           32                    23                     158
4          93          41                  82          159                  63                      9           144           46                    19                     143
5          85          44                  70          205                 103                     52           149           45                    19                     144
7          97          43                  73          173                  65                      6           153           42                    19                     143
  Scaled.Variance.Along.Major.Axis Scaled.Variance.Along.Minor.Axis Scaled.Radius.of.Gyration Skewness.About.Major.Axis Skewness.About.Minor.Axis Kurtosis.About.Minor.Axis
1                              176                              379                       184                       70                        6                       16
2                              170                              330                       158                       72                        9                       14
3                              223                              635                       220                       73                       14                        9
4                              160                              309                       127                       63                        6                       10
5                              241                              325                       188                      127                        9                       11
7                              176                              361                       172                       66                       13                        1
  Kurtosis.About.Major.Axis Hollows.Ratio class
1                       187           197   VAN
2                       189           199   VAN
3                       188           196    4D
4                       199           207   VAN
5                       180           183   BUS
7                       200           204   BUS
```

```
> print(head(set2, 6))
   Compactness Circularity Distance.Circularity Radius.Ratio Pr.Axis.Aspect.Ratio Max.Length.Aspect.Ratio Scatter.Ratio Elongatedness Pr.Axis.Rectangularity Max.Length.Rectangularity
6          107          57                 106          172                  50                      6           255           26                    28                     169
11          86          36                  70          143                  61                      9           133           50                    18                     130
12          90          34                  66          136                  55                      6           123           54                    17                     118
14          89          42                  85          144                  58                     10           152           44                    19                     144
15          94          49                  79          203                  71                      5           174           37                    21                     154
16          96          55                 103          201                  65                      9           204           32                    23                     166
   Scaled.Variance.Along.Major.Axis Scaled.Variance.Along.Minor.Axis Scaled.Radius.of.Gyration Skewness.About.Major.Axis Skewness.About.Minor.Axis Kurtosis.About.Minor.Axis
6                               280                              957                       264                       85                        5                        9
11                              153                              266                       127                       66                        2                       10
12                              148                              224                       118                       65                        5                       26
14                              173                              345                       161                       72                        8                       13
15                              196                              465                       206                       71                        6                        2
16                              227                              624                       246                       74                        6                        2
   Kurtosis.About.Major.Axis Hollows.Ratio class
6                        181           183   BUS
11                       194           202   VAN
12                       196           202    4D
14                       187           197   VAN
15                       197           199   BUS
16                       186           194    2D
```

3. Run a KNN analysis on the training data using *m* = 1.
(a) **Show the confusion matrix for the test data. Comment on how well separated the four classes are. In particular, are there classes that are easier/harder to separate?**
(b) Compute and **report the test misclassification rate and approximate standard error.**

```
50 ##### 3. Run a KNN analysis on the training data using m = 1.
51 # (a) Show the confusion matrix for the test data. Comment on how well
52 # separated the four classes are. In particular, are there classes that are
53 # easier/harder to separate?
54 X.train.raw = set1[,-19]
55 X.valid.raw = set2[,-19]
56 Y.train = as.matrix(set1[,19])
57 Y.valid = as.matrix(set2[,19])
58
59 ### Rescale x1 using the means and SDs of x2
60 scale.1 <- function(x1,x2){
61   for(col in 1:ncol(x1)){
62     a <- mean(x2[,col])
63     b <- sd(x2[,col])
64     x1[,col] <- (x1[,col]-a)/b
65   }
66   x1
67 }
68
69 ### Rescale our training and validation sets
70 X.train = scale.1(X.train.raw, X.train.raw)
71 X.valid = scale.1(X.valid.raw, X.train.raw) # Watch the order
72
73 pred.knn = knn(X.train, X.valid, cl = set1[,19], k=1)
74
75 ### Let's make a confusion matrix.
76 table(pred.knn, Y.valid, dnn = c("Predicted", "Observed"))
```

```
> table(pred.knn, Y.valid, dnn = c("Predicted", "Observed"))
         Observed
Predicted 2D 4D BUS VAN
      2D  25 18   1   0
      4D  26 29   1   2
      BUS  0  1  46   6
      VAN  4  6   1  46
```

```
80 # (b) Compute and report the test misclassification rate and approximate standard
81 # error
82
83 ### Next, let's get the misclassification rate
84 (misclass.knn = mean(pred.knn != Y.valid))
```

```
> (misclass.knn = mean(pred.knn != Y.valid))
[1] 0.3113208
```