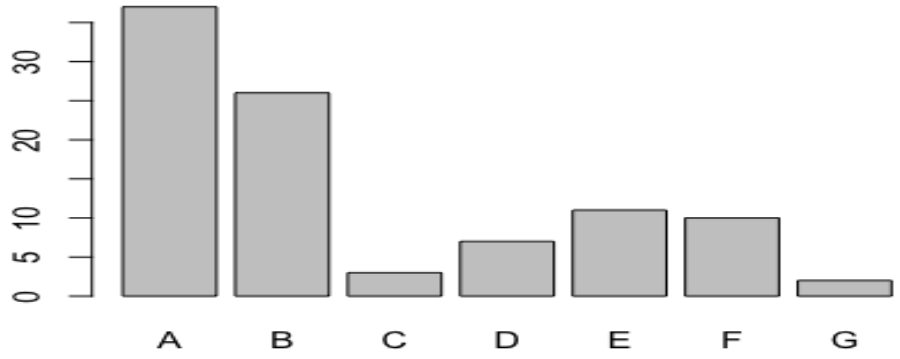
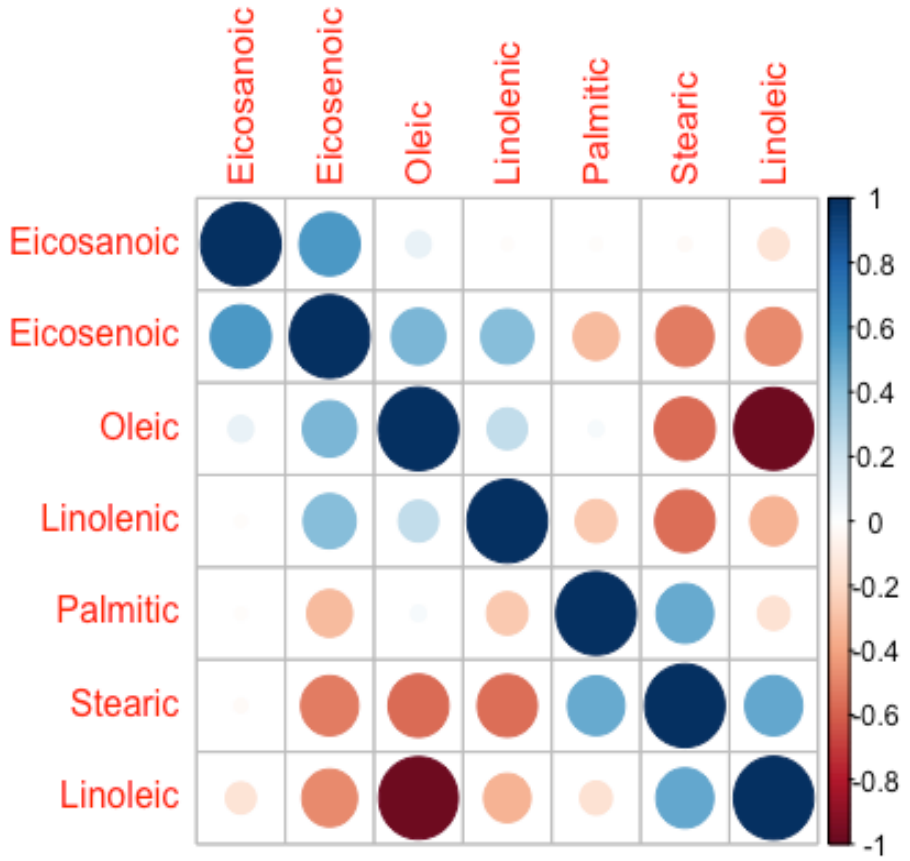
2.a.

Because the class imbalance the data set should be split using stratified sampling.




```
Based on the Correlation plot there are some correlated predictors which were removed.
```



1. Mixture Discriminant Analysis

```
74 samples
 6 predictor
 7 classes: 'A', 'B', 'C', 'D', 'E', 'F', 'G'

Pre-processing : Center and Scale
Resampling: Bootstrapped (25 reps)

Summary of sample sizes: 74, 74, 74, 74, 74, 74, ...

Resampling results across tuning parameters:

  subclasses  Accuracy   Kappa      Accuracy SD  Kappa SD
  1           0.9268027  0.9000571  0.04613226   0.06230732
  2           0.9136905  0.8766354  0.09978009   0.14252050

Kappa was used to select the optimal model using  the largest value.
The final value used for the model was subclasses = 1.
```

Confusion Matrix and Statistics **Testing set**

```
          Reference
Prediction A B C D E F G
         A 9 0 0 0 0 0 0
         B 0 6 0 0 0 0 0
         C 0 0 1 0 0 0 0
         D 0 0 0 1 0 0 0
         E 0 0 0 0 2 0 0
         F 0 0 0 0 0 2 0
         G 0 0 0 0 0 0 1

Overall Statistics for Testing set :

               Accuracy : 1
                 95% CI : (0.8456, 1)
    No Information Rate : 0.4091
    P-Value [Acc > NIR] : 2.884e-09

                  Kappa : 1
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: A Class: B Class: C Class: D Class: E Class: F
Sensitivity            1.0000   1.0000  1.00000  1.00000  1.00000  1.00000
Specificity            1.0000   1.0000  1.00000  1.00000  1.00000  1.00000
                     Class: G
Sensitivity           1.00000
```

```
Specificity           1.00000


variables are sorted by maximum importance across the classes
           A    B C D E      F G
Stearic    1 1.00 1 1 1 1.0000 1
Palmitic   1 1.00 1 1 1 1.0000 1
Linolenic  1 1.00 1 1 1 1.0000 1
Oleic      1 1.00 1 1 1 1.0000 1
Eicosanoic 1 0.95 1 1 1 1.0000 1
Eicosenoic 1 1.00 1 1 1 0.8542 1
```

2. Neural Network

```
74 samples
 6 predictor
 7 classes: 'A', 'B', 'C', 'D', 'E', 'F', 'G'

Pre-processing: spatial sign transformation, scaled, centered
Resampling: Bootstrapped (25 reps)

Summary of sample sizes: 74, 74, 74, 74, 74, 74, ...

Resampling results across tuning parameters:

  size  decay  Accuracy   Kappa       Accuracy SD  Kappa SD
   1    0.0    0.6987881  0.57791892  0.11334438   0.14718124
   1    0.1    0.6366933  0.47999634  0.13745402   0.15611758
   3    0.0    0.8280413  0.76343216  0.10899117   0.14861201
   3    0.1    0.9036922  0.86530845  0.09538117   0.13467173
   3    1.0    0.6731553  0.50692249  0.11053172   0.14404744
   3    2.0    0.5259045  0.27195591  0.16456447   0.22225626
   4    0.0    0.8634232  0.81252625  0.10719723   0.14398761
   4    0.1    0.9088823  0.87356650  0.09228015   0.12790248
   4    1.0    0.6886408  0.53435636  0.11699580   0.15801568
   4    2.0    0.5291096  0.27813048  0.16321931   0.21659432
   5    0.0    0.8754391  0.82796214  0.09683868   0.13396598
   5    0.1    0.9148953  0.88264541  0.08742802   0.11942961
   5    1.0    0.6948350  0.54428195  0.11903585   0.16212303
   5    2.0    0.5470168  0.30760730  0.16391903   0.21663562
     Kappa was used to select the optimal model using  the largest value.
The final values used for the model were size = 5 and decay = 0.1.

Confusion Matrix and Statistics Testing set

          Reference
Prediction A B C D E F G
         A 9 0 0 0 0 0 0
         B 0 6 0 0 0 0 1
         C 0 0 1 0 0 0 0
         D 0 0 0 0 0 0 0
         E 0 0 0 0 2 0 0
         F 0 0 0 1 0 2 0
         G 0 0 0 0 0 0 0
```

```
Overall Statistics for Testing set

               Accuracy : 0.9091
                 95% CI : (0.7084, 0.9888)
    No Information Rate : 0.4091
    P-Value [Acc > NIR] : 1.485e-06

                  Kappa : 0.8743
 Mcnemar's Test P-Value : NA



Statistics by Class:

                     Class: A Class: B Class: C Class: D Class: E Class: F
Sensitivity            1.0000   1.0000  1.00000  0.00000  1.00000  1.00000
Specificity            1.0000   0.9375  1.00000  1.00000  1.00000  0.95000

                     Class: G
Sensitivity           0.00000
Specificity           1.00000
```

3. Flexible Discriminant Analysis

```
74 samples
 6 predictor
 7 classes: 'A', 'B', 'C', 'D', 'E', 'F', 'G'

Pre-processing : Center and Scale

Resampling: Bootstrapped (25 reps)

Summary of sample sizes: 74, 74, 74, 74, 74, 74, ...

Resampling results across tuning parameters:

  nprune  Accuracy   Kappa      Accuracy SD  Kappa SD
   2      0.5788969  0.3919771  0.12983654   0.17813099
   7      0.9212303  0.8937645  0.04500279   0.06016160
  13      0.9237493  0.8969303  0.04157087   0.05550361

Tuning parameter 'degree' was held constant at a value of 1
Kappa was used to select the optimal model using  the largest value.
The final values used for the model were degree = 1 and nprune = 13.
```

Confusion Matrix and Statistics **Testing set**

```
          Reference
Prediction A B C D E F G
         A 9 0 0 0 0 0 0
         B 0 6 0 0 0 0 0
         C 0 0 1 0 0 0 0
         D 0 0 0 1 0 0 0
         E 0 0 0 0 2 0 0
         F 0 0 0 0 0 2 0
         G 0 0 0 0 0 0 1
```

```
Overall Statistics
               Accuracy : 1
                  Kappa : 1
Statistics by Class:
                     Class: A Class: B Class: C Class: D Class: E Class: F Class: G
Sensitivity            1.0000   1.0000  1.00000  1.00000  1.00000  1.00000  1.00000
Specificity            1.0000   1.0000  1.00000  1.00000  1.00000  1.00000  1.00000
```

fda variable importance

```
           Overall
Palmitic    100.00
Oleic        93.04
Linolenic    84.83
Stearic      74.44
Eicosenoic   28.78
Eicosanoic    0.00
```

5. Support Vector Machines with Radial Basis Function Kernel

```
74 samples
 6 predictor
 7 classes: 'A', 'B', 'C', 'D', 'E', 'F', 'G'

Pre-processing : Center and Scale
Resampling: Bootstrapped (25 reps)

Summary of sample sizes: 74, 74, 74, 74, 74, 74, ...

Resampling results across tuning parameters:

  C        Accuracy   Kappa       Accuracy SD  Kappa SD
   0.0625  0.3675221  0.00000000  0.06691789   0.00000000
   0.1250  0.3884587  0.03673462  0.08308983   0.08448864
   0.2500  0.6864940  0.54756831  0.14405452   0.18809785
   0.5000  0.8072887  0.73374931  0.11593993   0.15174910
   1.0000  0.8937349  0.85520193  0.09786982   0.13207147
   2.0000  0.9219243  0.89510608  0.07700160   0.10067752
   4.0000  0.9315190  0.90754581  0.07822848   0.10274981
   8.0000  0.9400412  0.90573022  0.07800019   0.10220908
  16.0000  0.9315797  0.90795720  0.07901996   0.10368099

Tuning parameter 'sigma' was held constant at a value of 0.033386
Kappa was used to select the optimal model using  the largest value.
The final values used for the model were sigma = 0.033386 and C = 8.
```

Confusion Matrix and Statistics **Testing set**

```
          Reference
Prediction A B C D E F G
         A 9 0 0 0 0 0 0
         B 0 6 0 0 0 0 0
         C 0 0 1 0 0 0 0
         D 0 0 0 0 0 0 0
         E 0 0 0 0 2 0 0
         F 0 0 0 0 0 2 0
         G 0 0 0 1 0 0 1

Overall Statistics

               Accuracy : 0.9545
                 95% CI : (0.7716, 0.9988)
    No Information Rate : 0.4091
    P-Value [Acc > NIR] : 9.454e-08

                  Kappa : 0.9382
 Mcnemar's Test P-Value : NA
```

```
Statistics by Class:

                     Class: A Class: B Class: C Class: D Class: E Class: F
Sensitivity            1.0000   1.0000  1.00000  0.00000  1.00000  1.00000
Specificity            1.0000   1.0000  1.00000  1.00000  1.00000  1.00000

                     Class: G
Sensitivity           1.00000
Specificity           0.95238
```

6. k-Nearest Neighbors

```
74 samples
 6 predictor
 7 classes: 'A', 'B', 'C', 'D', 'E', 'F', 'G'

Pre-processing : Center and Scale

Resampling: Bootstrapped (25 reps)

Summary of sample sizes: 74, 74, 74, 74, 74, 74, ...

Resampling results across tuning parameters:

  k    Accuracy   Kappa         Accuracy SD  Kappa SD
    3  0.9437413   0.921563597  0.03115882   0.04654870
    5  0.8974869   0.859998705  0.06600801   0.08908393
    9  0.8526078   0.800567045  0.07350034   0.09467525
   13  0.7864498   0.710750040  0.11679834   0.15233482
   17  0.7380869   0.640733236  0.08764073   0.11421086
   21  0.6747135   0.541849740  0.10598175   0.14161552
   41  0.5024954   0.283891081  0.12272963   0.13262620
   61  0.3530650   0.001767019  0.06639894   0.01560056
   81  0.3461684  -0.002319588  0.06692049   0.01037351
  101  0.3522029   0.002561380  0.07035510   0.02483029

  401  0.3547891   0.009154930  0.06847409   0.04094209
  451  0.3461684  -0.009472656  0.06692049   0.04236301

Kappa was used to select the optimal model using  the largest value.
The final value used for the model was k = 3.
```

Confusion Matrix and Statistics **Testing set**

```
          Reference
Prediction A B C D E F G
         A 9 0 0 0 0 0 1
         B 0 6 0 0 0 0 0
         C 0 0 1 0 0 0 0
         D 0 0 0 0 0 0 0
         E 0 0 0 0 2 0 0
         F 0 0 0 1 0 2 0
         G 0 0 0 0 0 0 0
```

```
Overall Statistics for Testing set :

               Accuracy : 0.9091
                 95% CI : (0.7084, 0.9888)
    No Information Rate : 0.4091
    P-Value [Acc > NIR] : 1.485e-06

                  Kappa : 0.8732
 Mcnemar's Test P-Value : NA

 Statistics by Class:

                     Class: A Class: B Class: C Class: D Class: E Class: F
 Sensitivity           1.0000   1.0000  1.00000  0.00000  1.00000  1.00000
 Specificity           0.9231   1.0000  1.00000  1.00000  1.00000  0.95000

                     Class: G
 Sensitivity          0.00000
 Specificity          1.00000
```

7.Naive Bayes Classifier for Discrete Predictors

```
 Call:
 naiveBayes.default(x = trainX, y = trainY)

 A-priori probabilities:
 trainY
         A          B          C          D          E          F
 0.37837838 0.27027027 0.02702703 0.08108108 0.12162162 0.10810811
         G
 0.01351351
   Conditional probabilities:


       Palmitic
 trainY     [,1]       [,2]
     A 10.95714 1.36474894
     B  6.29000 0.36259300
     C  9.65000 0.07071068
     D 11.90000 1.56588633
     E 10.41111 0.69362173
     F  5.11250 0.40510140
     G 10.00000         NA
```

```
      Stearic
trainY     [,1]       [,2]
     A 5.335714 0.58004743
     B 4.050000 0.40457905
     C 3.350000 0.07071068
     D 2.783333 0.14719601
     E 3.988889 0.26193723
     F 1.925000 0.20528726
     G 2.300000         NA

      Oleic
trainY     [,1]     [,2]
     A 33.38929 4.391434
     B 26.25000 1.883865
     C 58.50000 1.131371
     D 73.90000 3.055487
     E 25.81111 2.010873
     F 58.87500 4.089272
     G 36.90000       NA

      Linolenic
trainY     [,1]       [,2]
     A 1.014286 1.00764275
     B 0.635000 0.51633832
     C 0.150000 0.07071068
     D 0.700000 0.08944272
     E 6.766667 0.79056942
     F 8.312500 0.99058064
     G 2.200000         NA

      Eicosanoic
trainY      [,1]      [,2]
     A 0.4142857 0.2731358
     B 0.3550000 0.5835238
     C 1.5000000 0.0000000
     D 0.1500000 0.1224745
     E 0.3111111 0.2204793
     F 0.4375000 0.2924649
     G 0.5000000        NA

      Eicosenoic
trainY      [,1]       [,2]
     A 0.1821429 0.14920424
     B 0.2000000 0.17770466
     C 1.5000000 0.42426407
     D 0.1333333 0.08164966
     E 0.2444444 0.26977357
     F 1.0000000 0.65246784
     G 0.5000000         NA
```

Confusion Matrix and Statistics **Testing set**

```
          Reference
Prediction A B C D E F G
         A 9 0 0 1 0 0 1
         B 0 6 0 0 0 0 0
         C 0 0 1 0 0 0 0
         D 0 0 0 0 0 0 0
         E 0 0 0 0 2 0 0
         F 0 0 0 0 0 2 0
         G 0 0 0 0 0 0 0
```

Overall Statistics for Testing set :

```
               Accuracy : 0.9091
                 95% CI : (0.7084, 0.9888)
    No Information Rate : 0.4091
    P-Value [Acc > NIR] : 1.485e-06

                  Kappa : 0.8706
 Mcnemar's Test P-Value : NA
```

Statistics by Class:

	Class: A	Class: B	Class: C	Class: D	Class: E	Class: F	Class: G
Sensitivity	1.0000	1.0000	1.00000	0.00000	1.00000	1.00000	0.00000
Specificity	0.8462	1.0000	1.00000	1.00000	1.00000	1.00000	1.00000

LINEAR MODELS	Kappa	Accuracy
LDA	1	1
PLSDA	0.7413	0.8182
Penalised Models	0.8764	0.9091
NSC	0.9391	0.9545

Non – LINEAR MODELS	Kappa	Accuracy
MDA	1	1
NNet	0.8743	0.9091
FDA	1	1
SVM	0.9382	0.9545
KNN	0.8732	0.9091
Naïve Bayes	0.8706	0.9888

a).Based on Kappa and accuracy MDA and FDA are best models for this dataset from Non linear models. LDA from Linear models have similar performance. Two of the Non Linear models were able to do perfect classification and one of the linear models also achieved the same. Also performance difference in other models is very close. So I assume this is LINEAR BOUNDARY.

b).

Best predice Oil type : A

Least accurate oil type : G