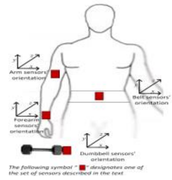


Session 18 Segmentation Analysis

Weight Lifting Excercise

5. Problem Statement

1. Use the below given data set
DataSet
2. Perform the below given activities:
a. Create classification model using different decision trees.
b. Verify model goodness of fit.
c. Apply all the model validation techniques.
d. Make conclusions

```
setwd("C:/Users/Seshan/Desktop/sv R related/acadgild/assignments/session17")

library(readr)

Weightlift <- read.csv("C:/Users/Seshan/Desktop/sv R related/acadgild/assignments/session 18 Assign/Weight lift.csv",sep=',',header=F)

#problem: find out natual grouping

df<-Weightlift

#df=Weightlift <- read.csv("C:/Users/Seshan/Desktop/sv R related/acadgild/assignments/session 18 Assign/Weightlift.csv",sep=',',header=F)

View(df)

#df = read.csv("https://archive.ics.uci.edu/ml/machine-learning-databases/00273/Example_WearableComputing_weight_lifting_exercises_biceps_curl_variations.csv",sep=',',header=F)
```

```
#problem: find out natual grouping

#View(df)

head(df)


str(df)


set.seed(1234)

ind = sample(1:nrow(df),0.8*nrow(df),replace = F)

df_train =df[ind,-1]

df_test = df[-ind,-1]


summary(df)

dim(df)

apply(df,2,range)

apply(df,2,summary)

df[] <- lapply(df, function(x) as.numeric(as.character(x)))

# KMeans - comes from Rcmdr library

# Kmeans- from amap library

# kmeans- from stats library

# steps in k-means clustering

#1- preprocessing the data (impute missing values, remove outliers, feature trasnformation)

#2- scaling or standardization of data set

#3- decide the number of clusters (value of K)

#4- iterate over the samples to create clusters

#5- decide the distance measure

#6- calculate the group accuracy

# scaling of data

df_train1 <- scale(df_train)
```

```
head(df_train1)

class(df_train1)


# screeplot approach to decide the number of clusters

km = kmeans(df_train1,1)

km$withinss

km$tot.withinss


km = kmeans(df_train1,2)

km$withinss

km$tot.withinss


km = kmeans(df_train1,3)

km$withinss

km$tot.withinss


km = kmeans(df_train1,4)

km$withinss

km$tot.withinss


km = kmeans(df_train1,5)

km$withinss

km$tot.withinss


km = kmeans(df_train1,6)

km$withinss

km$tot.withinss
```

```
km = kmeans(df_train1,7)

km$withinss

km$tot.withinss


km = kmeans(df_train1,8)

km$withinss

km$tot.withinss


km = kmeans(df_train1,9)

km$withinss

km$tot.withinss


km = kmeans(df_train1,10)

km$withinss

km$tot.withinss


dev.off()

sumsq=NULL

for (i in 1:10)

  sumsq[i] = sum(kmeans(df_train,centers=i,

                        iter.max = 1000,

                        nstart=i,

                        algorithm='Forgy')$withinss)

plot(1:10,sumsq,type='b', main='Screeplot showing within group sum of squares')


km = kmeans(df_train1,3)

km$withinss

km$tot.withinss
```

```
class(km$cluster)

summary(km)

km$centers

as.numeric(km$cluster)

length(km$cluster)

dim(df_train)

class(df_train)

df_train$cl <- km$cluster

head(df_train)
# profiles of clusters
aggregate(df_train[,1:8],list(df_train[,9]),mean)

table(df$V1)

library(cluster)

clusplot(df_train,df_train$cl,cex=0.9,color=T,shade=T, labels=4,lines=0)

#HC clustering or Hierarchical Clustering
```

```
# distance (euclidean, manhattan, cosine distance)

# Divisive method (top down)
# Agglomorative method (bottom up)



df_train = df_train[,-9]
head(df_train)

# compute the distance metrix
d1 <- dist(df_train,method='euclidean')
summary(d1)


# HC
fit <- hclust(d1,method = 'ward.D2')
plot(fit)

# single, double, average, ward, ward.D2

# agglomorative method
fit <- agnes(d1,metric='euclidean',method = 'ward')
plot(fit)

# divisive method
fit <- diana(d1,metric='euclidean')
plot(fit)
```

```
library(ggdendro)

if(require(cluster)){

  fit<- agnes(d1, metric = "manhattan", stand = TRUE)

  dg <- as.dendrogram(fit)

  ggdendrogram(dg)



  fit <- diana(d1, metric = "manhattan", stand = TRUE)

  dg <- as.dendrogram(fit)

  ggdendrogram(dg)

}
```

```
> setwd("C:/Users/Seshan/Desktop/sv R related/acadgild/assignments/session17")
> library(readr)
>
> Weightlift <- read.csv("C:/Users/Seshan/Desktop/sv R related/acadgild/assign
ments/session 18 Assign/Weight lift.csv",sep=',',header=F)
> #problem: find out natual grouping
> df<-Weightlift
> View(df)
> head(df)
         V1                   V2                   V3               V4
1 user_name raw_timestamp_part_1 raw_timestamp_part_2   cvtd_timestamp
2    eurico           1322489729                34670 28/11/2011 14:15
3    eurico           1322489729                62641 28/11/2011 14:15
4    eurico           1322489729                70653 28/11/2011 14:15
5    eurico           1322489729                82654 28/11/2011 14:15
6    eurico           1322489729                90637 28/11/2011 14:15
          V5         V6        V7          V8       V9              V10
1 new_window num_window roll_belt pitch_belt yaw_belt total_accel_belt
2         no          1       3.7        41.6    -82.8                3
3         no          1      3.66        42.8    -82.5                2
4         no          1      3.58        43.7    -82.3                1
5         no          1      3.56        44.4    -82.1                1
6         no          1      3.57        45.1    -81.9                1
                V11                V12                V13                V14
1 kurtosis_roll_belt kurtosis_picth_belt skewness_roll_belt skewness_roll_belt
2          -1.03566           -0.39133           0.005406           0.045115
3          -1.03566           -0.39133           0.005406           0.045115
4          -1.03566           -0.39133           0.005406           0.045115
```

```
5          -1.03566           -0.39133           0.005406           0.045115
6          -1.03566           -0.39133           0.005406           0.045115
          V15            V16          V17           V18            V19
1 max_roll_belt max_picth_belt max_yaw_belt min_roll_belt min_pitch_belt
2          -4.1             20           -1         -7.25             18
3          -4.1             20           -1         -7.25             18
4          -4.1             20           -1         -7.25             18
5          -4.1             20           -1         -7.25             18
6          -4.1             20           -1         -7.25             18
          V20                 V21                  V22                V23
1 min_yaw_belt amplitude_roll_belt amplitude_pitch_belt amplitude_yaw_belt
2           -1               1.345                    2                  0
3           -1               1.345                    2                  0
4           -1               1.345                    2                  0
5           -1               1.345                    2                  0
6           -1               1.345                    2                  0
                 V24           V25              V26           V27
1 var_total_accel_belt avg_roll_belt stddev_roll_belt var_roll_belt
2                  0.3         121.9              0.6          0.35
3                  0.3         121.9              0.6          0.35
4                  0.3         121.9              0.6          0.35
5                  0.3         121.9              0.6          0.35
6                  0.3         121.9              0.6          0.35
            V28               V29            V30          V31             V32
1 avg_pitch_belt stddev_pitch_belt var_pitch_belt avg_yaw_belt stddev_yaw_belt
2          25.75              0.35            0.1        -4.95             0.4
3          25.75              0.35            0.1        -4.95             0.4
4          25.75              0.35            0.1        -4.95             0.4
5          25.75              0.35            0.1        -4.95             0.4
6          25.75              0.35            0.1        -4.95             0.4
          V33          V34          V35          V36          V37          V3
8
1 var_yaw_belt gyros_belt_x gyros_belt_y gyros_belt_z accel_belt_x accel_belt_
y
2         0.17         2.02         0.18         0.02           -3           -1
8
3         0.17         1.96         0.14         0.05           -2           -1
3
4         0.17         1.88         0.08         0.05           -2            -
6
5         0.17          1.8         0.03         0.08           -6            -
5
6         0.17         1.77            0         0.13           -4            -
9
          V39          V40          V41          V42      V43       V44
1 accel_belt_z magnet_belt_x magnet_belt_y magnet_belt_z roll_arm pitch_arm
2           22           387           525          -267      132     -43.7
3           16           405           512          -254      129     -45.3
4            8           409           511          -244      125     -46.8
5            7           422           513          -221      120     -48.1
6            0           418           508          -208      115     -49.1
      V45             V46           V47          V48             V49
1 yaw_arm total_accel_arm var_accel_arm avg_roll_arm stddev_roll_arm
2   -53.6              38       65.0977     76.22175         16.1039
3     -49              38       65.0977     76.22175         16.1039
4   -43.7              35       65.0977     76.22175         16.1039
5   -38.1              35       65.0977     76.22175         16.1039
6   -31.7              34       65.0977     76.22175         16.1039
          V50           V51              V52           V53         V54
1 var_roll_arm avg_pitch_arm stddev_pitch_arm var_pitch_arm avg_yaw_arm
2     259.3599      -10.1695         10.66725      113.7978     19.0615
3     259.3599      -10.1695         10.66725      113.7978     19.0615
4     259.3599      -10.1695         10.66725      113.7978     19.0615
5     259.3599      -10.1695         10.66725      113.7978     19.0615
```

```
6     259.3599      -10.1695         10.66725      113.7978     19.0615
             V55          V56          V57          V58          V59          V60
1 stddev_yaw_arm var_yaw_arm gyros_arm_x gyros_arm_y gyros_arm_z accel_arm_x
2        35.8809    1287.463         2.65        -0.61        -0.02          143
3        35.8809    1287.463         2.79        -0.64        -0.11          146
4        35.8809    1287.463         2.91        -0.69        -0.15          156
5        35.8809    1287.463         3.08        -0.72        -0.23          158
6        35.8809    1287.463          3.2        -0.77        -0.25          163
          V61          V62          V63          V64          V65
1 accel_arm_y accel_arm_z magnet_arm_x magnet_arm_y magnet_arm_z
2          30         -346          556         -205         -374
3          35         -339          599         -206         -335
4          44         -307          613         -198         -319
5          52         -305          646         -186         -268
6          55         -288          670         -175         -241
               V66               V67              V68               V69
1 kurtosis_roll_arm kurtosis_picth_arm kurtosis_yaw_arm skewness_roll_arm
2          -1.18224          -0.96912         -0.86977           0.12353
3          -1.18224          -0.96912         -0.86977           0.12353
4          -1.18224          -0.96912         -0.86977           0.12353
5          -1.18224          -0.96912         -0.86977           0.12353
6          -1.18224          -0.96912         -0.86977           0.12353
                V70               V71          V72           V73          V74
1 skewness_pitch_arm skewness_yaw_arm max_roll_arm max_picth_arm max_yaw_arm
2           -0.10319         0.059765         8.45         77.25           38
3           -0.10319         0.059765         8.45         77.25           38
4           -0.10319         0.059765         8.45         77.25           38
5           -0.10319         0.059765         8.45         77.25           38
6           -0.10319         0.059765         8.45         77.25           38
          V75           V76          V77                V78                  V7
9
1 min_roll_arm min_pitch_arm min_yaw_arm amplitude_roll_arm amplitude_pitch_ar
m
2        -33.6         -58.6           10             36.945               121.
5
3        -33.6         -58.6           10             36.945               121.
5
4        -33.6         -58.6           10             36.945               121.
5
5        -33.6         -58.6           10             36.945               121.
5
6        -33.6         -58.6           10             36.945               121.
5
               V80           V81            V82          V83
1 amplitude_yaw_arm roll_dumbbell pitch_dumbbell yaw_dumbbell
2                27   51.23553997    11.69884724  104.2647274
3                27   55.82441814    9.645819033  100.2280531
4                27    55.4698307    6.875243852  101.0841063
5                27   55.94485974    11.07929719  99.78455638
6                27   55.21173932    11.42683324  100.4225829
                    V84                     V85                    V86
1 kurtosis_roll_dumbbell kurtosis_picth_dumbbell skewness_roll_dumbbell
2              -0.09595                 -0.4422                 0.0819
3              -0.09595                 -0.4422                 0.0819
4              -0.09595                 -0.4422                 0.0819
5              -0.09595                 -0.4422                 0.0819
6              -0.09595                 -0.4422                 0.0819
                      V87               V88                V89                V9
0
1 skewness_pitch_dumbbell max_roll_dumbbell max_picth_dumbbell max_yaw_dumbbel
l
2                  -0.216             41.85                133               -0.
1
3                  -0.216             41.85                133               -0.
```

```
1
4              -0.216          41.85               133            -0.
1
5              -0.216          41.85               133            -0.
1
6              -0.216          41.85               133            -0.
1
             V91                V92               V93                    V9
4
1 min_roll_dumbbell min_pitch_dumbbell min_yaw_dumbbell amplitude_roll_dumbbel
l
2          -26.75               20.2             -0.1                  55.7
1
3          -26.75               20.2             -0.1                  55.7
1
4          -26.75               20.2             -0.1                  55.7
1
5          -26.75               20.2             -0.1                  55.7
1
6          -26.75               20.2             -0.1                  55.7
1
                       V95                    V96                V97
1 amplitude_pitch_dumbbell amplitude_yaw_dumbbell total_accel_dumbbell
2                    54.74                      0                    4
3                    54.74                      0                    4
4                    54.74                      0                    4
5                    54.74                      0                    5
6                    54.74                      0                    4
                V98               V99                V100              V101
1 var_accel_dumbbell avg_roll_dumbbell stddev_roll_dumbbell var_roll_dumbbell
2           2.41635          -5.11805               17.058           291.001
3           2.41635          -5.11805               17.058           291.001
4           2.41635          -5.11805               17.058           291.001
5           2.41635          -5.11805               17.058           291.001
6           2.41635          -5.11805               17.058           291.001
               V102                  V103              V104             V105
1 avg_pitch_dumbbell stddev_pitch_dumbbell var_pitch_dumbbell avg_yaw_dumbbell
2           13.9312               14.1062          199.0775          64.7063
3           13.9312               14.1062          199.0775          64.7063
4           13.9312               14.1062          199.0775          64.7063
5           13.9312               14.1062          199.0775          64.7063
6           13.9312               14.1062          199.0775          64.7063
                V106             V107             V108             V109
1 stddev_yaw_dumbbell var_yaw_dumbbell gyros_dumbbell_x gyros_dumbbell_y
2             13.5747         184.5578            -0.31             0.16
3             13.5747         184.5578            -0.31             0.14
4             13.5747         184.5578            -0.31             0.16
5             13.5747         184.5578            -0.31             0.16
6             13.5747         184.5578            -0.31             0.14
              V110             V111             V112             V113
1 gyros_dumbbell_z accel_dumbbell_x accel_dumbbell_y accel_dumbbell_z
2             0.08                5               21               37
3             0.07                4               22               35
4             0.05                3               23               37
5             0.07                5               24               38
6             0.07                5               23               37
              V114             V115             V116         V117
1 magnet_dumbbell_x magnet_dumbbell_y magnet_dumbbell_z roll_forearm
2             -471              191              277         -111
3             -472              184              281         -112
4             -468              190              275         -114
5             -469              184              285         -115
6             -468              189              292         -117
          V118         V119                   V120                 V121
```

```
1 pitch_forearm yaw_forearm kurtosis_roll_forearm kurtosis_picth_forearm
2          26.5         138              -1.09475               -0.97525
3          26.2         138              -1.09475               -0.97525
4            26         137              -1.09475               -0.97525
5          25.8         137              -1.09475               -0.97525
6          25.5         137              -1.09475               -0.97525
                  V122                   V123             V124
1 skewness_roll_forearm skewness_pitch_forearm max_roll_forearm
2              -0.05065                0.17285             49.6
3              -0.05065                0.17285             49.6
4              -0.05065                0.17285             49.6
5              -0.05065                0.17285             49.6
6              -0.05065                0.17285             49.6
              V125            V126            V127             V128
1 max_picth_forearm max_yaw_forearm min_roll_forearm min_pitch_forearm
2               168            -1.1             4.65            -168.5
3               168            -1.1             4.65            -168.5
4               168            -1.1             4.65            -168.5
5               168            -1.1             4.65            -168.5
6               168            -1.1             4.65            -168.5
             V129                  V130                   V131
1 min_yaw_forearm amplitude_roll_forearm amplitude_pitch_forearm
2            -1.1                   32.2                   341.5
3            -1.1                   32.2                   341.5
4            -1.1                   32.2                   341.5
5            -1.1                   32.2                   341.5
6            -1.1                   32.2                   341.5
                  V132                V133              V134             V135
1 amplitude_yaw_forearm total_accel_forearm var_accel_forearm avg_roll_forearm
2                     0                  30           14.0772         27.85936
3                     0                  31           14.0772         27.85936
4                     0                  32           14.0772         27.85936
5                     0                  33           14.0772         27.85936
6                     0                  34           14.0772         27.85936
                V136             V137              V138                V139
1 stddev_roll_forearm var_roll_forearm avg_pitch_forearm stddev_pitch_forearm
2            45.16342         2749.163          25.35597             8.906695
3            45.16342         2749.163          25.35597             8.906695
4            45.16342         2749.163          25.35597             8.906695
5            45.16342         2749.163          25.35597             8.906695
6            45.16342         2749.163          25.35597             8.906695
              V140            V141               V142            V143
1 var_pitch_forearm avg_yaw_forearm stddev_yaw_forearm var_yaw_forearm
2          79.33451        17.09505           74.27584        5541.956
3          79.33451        17.09505           74.27584        5541.956
4          79.33451        17.09505           74.27584        5541.956
5          79.33451        17.09505           74.27584        5541.956
6          79.33451        17.09505           74.27584        5541.956
             V144            V145            V146            V147
1 gyros_forearm_x gyros_forearm_y gyros_forearm_z accel_forearm_x
2           -0.05           -0.37           -0.43            -170
3           -0.06           -0.37           -0.59            -178
4           -0.05           -0.27           -0.72            -182
5            0.02           -0.24           -0.79            -185
6            0.08           -0.27           -0.82            -188
             V148            V149             V150             V151
1 accel_forearm_y accel_forearm_z magnet_forearm_x magnet_forearm_y
2             155             184            -1160             1400
3             164             182            -1150             1410
4             172             185            -1130             1400
5             182             188            -1120             1400
6             195             188            -1100             1400
              V152
1 magnet_forearm_z
```

```
2            -876
3            -871
4            -863
5            -855
6            -843
>
> str(df)
'data.frame':  4025 obs. of  152 variables:
 $ V1  : Factor w/ 6 levels "adelmo","carlitos",..: 6 3 3 3 3 3 3 3 3 3 ...
 $ V2  : Factor w/ 89 levels "1322489729","1322489730",..: 89 1 1 1 1 1 1 1 1
1 ...
 $ V3  : Factor w/ 3789 levels "100232","100295",..: 3789 1038 2291 2620 3101
3435 310 403 597 696 ...
 $ V4  : Factor w/ 8 levels "2/12/2011 13:35",..: 8 2 2 2 2 2 2 2 2 2 ...
 $ V5  : Factor w/ 3 levels "new_window","no",..: 1 2 2 2 2 2 2 2 2 2 ...
 $ V6  : Factor w/ 92 levels "1","10","11",..: 92 1 1 1 1 1 1 1 1 1 ...
 $ V7  : Factor w/ 870 levels "-0.03","-0.04",..: 870 665 663 662 660 661 655
652 643 618 ...
 $ V8  : Factor w/ 822 levels "-27.8","-27.9",..: 822 403 404 405 406 407 409
411 415 417 ...
 $ V9  : Factor w/ 936 levels "-0.02","-0.03",..: 936 650 648 647 645 643 643
643 646 649 ...
 $ V10 : Factor w/ 28 levels "0","1","10","11",..: 28 21 13 2 2 2 2 21 22 13 .
..
 $ V11 : Factor w/ 90 levels "-0.076054","-0.120447",..: 90 30 30 30 30 30 30
30 30 30 ...
 $ V12 : Factor w/ 79 levels "-0.06016","-0.108371",..: 79 15 15 15 15 15 15 1
5 15 15 ...
 $ V13 : Factor w/ 90 levels "-0.011683","-0.037647",..: 90 45 45 45 45 45 45
45 45 45 ...
 $ V14 : Factor w/ 81 levels "-0.045472","-0.04606",..: 81 40 40 40 40 40 40 4
0 40 40 ...
 $ V15 : Factor w/ 66 levels "-0.2","-0.4",..: 66 24 24 24 24 24 24 24 24 24 .
..
 $ V16 : Factor w/ 15 levels "10","17","19",..: 15 4 4 4 4 4 4 4 4 4 ...
 $ V17 : Factor w/ 35 levels "-0.1","-0.2",..: 35 10 10 10 10 10 10 10 10 10 .
..
 $ V18 : Factor w/ 65 levels "-1","-1.5","-10.5",..: 65 27 27 27 27 27 27 27 2
7 27 ...
 $ V19 : Factor w/ 14 levels "0","1","13","15",..: 14 7 7 7 7 7 7 7 7 7 ...
 $ V20 : Factor w/ 35 levels "-0.1","-0.2",..: 35 10 10 10 10 10 10 10 10 10 .
..
 $ V21 : Factor w/ 64 levels "0","0.1","0.2",..: 64 23 23 23 23 23 23 23 23 23
...
 $ V22 : Factor w/ 11 levels "0","1","2","21",..: 11 3 3 3 3 3 3 3 3 3 ...
 $ V23 : Factor w/ 2 levels "0","amplitude_yaw_belt": 2 1 1 1 1 1 1 1 1 1 ...
 $ V24 : Factor w/ 27 levels "0","0.1","0.2",..: 27 4 4 4 4 4 4 4 4 4 ...
 $ V25 : Factor w/ 62 levels "-0.7","-1","-1.6",..: 62 26 26 26 26 26 26 26 26
26 ...
 $ V26 : Factor w/ 33 levels "0","0.1","0.2",..: 33 7 7 7 7 7 7 7 7 7 ...
 $ V27 : Factor w/ 39 levels "0","0.1","0.2",..: 39 5 5 5 5 5 5 5 5 5 ...
 $ V28 : Factor w/ 55 levels "-30.6","-31.8",..: 55 10 10 10 10 10 10 10 10 10
...
 $ V29 : Factor w/ 23 levels "0","0.1","0.2",..: 23 5 5 5 5 5 5 5 5 5 ...
 $ V30 : Factor w/ 24 levels "0","0.1","0.2",..: 24 2 2 2 2 2 2 2 2 2 ...
 $ V31 : Factor w/ 70 levels "-0.6","-0.8",..: 70 25 25 25 25 25 25 25 25 25 .
..
 $ V32 : Factor w/ 21 levels "0","0.1","0.2",..: 21 5 5 5 5 5 5 5 5 5 ...
 $ V33 : Factor w/ 56 levels "0","0.01","0.02",..: 56 17 17 17 17 17 17 17 17
17 ...
 $ V34 : Factor w/ 136 levels "-0.02","-0.03",..: 136 135 134 133 132 130 129
131 129 128 ...
 $ V35 : Factor w/ 51 levels "-0.02","-0.03",..: 51 37 35 31 28 26 2 4 4 2 ...
 $ V36 : Factor w/ 99 levels "-0.02","-0.03",..: 99 49 51 51 53 56 58 57 62 68
```

```
...
 $ V37 : Factor w/ 154 levels "-1","-10","-11",..: 154 25 14 14 55 36 84 84 90
1 ...
 $ V38 : Factor w/ 113 levels "-1","-10","-11",..: 113 9 5 21 19 25 25 14 16 1
0 ...
 $ V39 : Factor w/ 214 levels "-1","-10","-11",..: 214 155 148 212 204 140 135
138 139 137 ...
 $ V40 : Factor w/ 247 levels "-1","-10","-11",..: 247 192 202 204 208 206 213
215 216 218 ...
 $ V41 : Factor w/ 187 levels "428","435","437",..: 187 63 51 50 52 47 49 47 4
2 46 ...
 $ V42 : Factor w/ 372 levels "-112","-115",..: 372 83 70 60 38 28 18 12 5 4 .
..
 $ V43 : Factor w/ 1238 levels "-0.02","-0.05",..: 1238 671 668 664 659 654 64
9 643 1226 1182 ...
 $ V44 : Factor w/ 1694 levels "-0.01","-0.05",..: 1694 517 533 546 557 565 57
0 573 571 564 ...
 $ V45 : Factor w/ 1670 levels "-0.18","-0.23",..: 1670 453 407 371 318 278 22
8 158 48 336 ...
 $ V46 : Factor w/ 60 levels "1","10","11",..: 60 32 32 29 29 28 27 22 21 20 .
..
 $ V47 : Factor w/ 83 levels "0","0.0278","0.1308",..: 83 62 62 62 62 62 62 62
62 62 ...
 $ V48 : Factor w/ 90 levels "-128","-128.4898",..: 90 74 74 74 74 74 74 74 74
74 ...
 $ V49 : Factor w/ 89 levels "0","0.05","0.2323",..: 89 21 21 21 21 21 21 21 2
1 21 ...
 $ V50 : Factor w/ 89 levels "0","0.0025","0.054",..: 89 30 30 30 30 30 30 30
30 30 ...
 $ V51 : Factor w/ 90 levels "-10.1695","-10.902",..: 90 1 1 1 1 1 1 1 1 1 ...
 $ V52 : Factor w/ 90 levels "0","0.0548","0.2209",..: 90 20 20 20 20 20 20 20
20 20 ...
 $ V53 : Factor w/ 90 levels "0","0.003","0.0488",..: 90 19 19 19 19 19 19 19
19 19 ...
 $ V54 : Factor w/ 89 levels "-0.0558","-1.7763",..: 89 51 51 51 51 51 51 51 5
1 51 ...
 $ V55 : Factor w/ 86 levels "0","0.1924","0.3087",..: 86 41 41 41 41 41 41 41
41 41 ...
 $ V56 : Factor w/ 86 levels "0","0.037","0.0953",..: 86 15 15 15 15 15 15 15
15 15 ...
 $ V57 : Factor w/ 543 levels "-0.02","-0.03",..: 543 457 466 473 484 491 498
510 512 504 ...
 $ V58 : Factor w/ 318 levels "-0.02","-0.03",..: 318 38 40 43 45 48 52 52 52
52 ...
 $ V59 : Factor w/ 213 levels "-0.02","-0.03",..: 213 1 7 9 14 15 18 19 13 7 .
..
 $ V60 : Factor w/ 707 levels "-1","-10","-100",..: 707 339 342 353 355 361 35
8 363 350 339 ...
 $ V61 : Factor w/ 419 levels "-1","-10","-100",..: 419 343 348 358 367 370 37
4 383 387 395 ...
 $ V62 : Factor w/ 646 levels "-1","-10","-100",..: 646 258 250 219 218 198 18
5 136 128 115 ...
 $ V63 : Factor w/ 1151 levels "-1","-10","-100",..: 1151 892 937 952 988 1015
1042 1070 1074 1091 ...
 $ V64 : Factor w/ 817 levels "-1","-10","-100",..: 817 110 111 102 89 79 78 6
6 57 38 ...
 $ V65 : Factor w/ 1038 levels "-1","-10","-100",..: 1038 245 214 201 159 138
95 24 7 290 ...
 $ V66 : Factor w/ 88 levels "-0.11926","-0.19002",..: 88 33 33 33 33 33 33 33
33 33 ...
 $ V67 : Factor w/ 84 levels "-0.10176","-0.15381",..: 84 28 28 28 28 28 28 28
28 28 ...
 $ V68 : Factor w/ 82 levels "-0.06791","-0.12096",..: 82 26 26 26 26 26 26 26
26 26 ...
```

```
 $ V69 : Factor w/ 88 levels "-0.00696","-0.01884",..: 88 44 44 44 44 44 44 44
44 44 ...
 $ V70 : Factor w/ 84 levels "-0.01247","-0.09627",..: 84 3 3 3 3 3 3 3 3 3 ..
.
 $ V71 : Factor w/ 82 levels "-0.0046","-0.008",..: 82 41 41 41 41 41 41 41 41
41 ...
 $ V72 : Factor w/ 83 levels "-0.7","-1.1",..: 83 80 80 80 80 80 80 80 80 80 .
..
 $ V73 : Factor w/ 80 levels "-1.9","-122",..: 80 63 63 63 63 63 63 63 63 63 .
..
 $ V74 : Factor w/ 39 levels "19","20","21",..: 39 17 17 17 17 17 17 17 17 17
...
 $ V75 : Factor w/ 88 levels "-10","-11.5",..: 88 14 14 14 14 14 14 14 14 14 .
..
 $ V76 : Factor w/ 84 levels "-10.9","-100",..: 84 43 43 43 43 43 43 43 43 43
...
 $ V77 : Factor w/ 30 levels "1","10","11",..: 30 2 2 2 2 2 2 2 2 2 ...
 $ V78 : Factor w/ 88 levels "0","0.1","0.57",..: 88 42 42 42 42 42 42 42 42 4
2 ...
 $ V79 : Factor w/ 78 levels "0","1","100.2",..: 78 13 13 13 13 13 13 13 13 13
...
 $ V80 : Factor w/ 41 levels "0","1","12","13",..: 41 17 17 17 17 17 17 17 17
17 ...
 $ V81 : Factor w/ 3653 levels "-0.949086746",..: 3653 2993 3032 3029 3034 302
6 3017 3028 3052 3017 ...
 $ V82 : Factor w/ 3572 levels "-0.738381151",..: 3572 1153 3542 3326 1118 112
9 1313 1328 3378 1313 ...
 $ V83 : Factor w/ 3602 levels "-10.50214329",..: 3602 698 627 644 3598 633 63
9 3591 3572 639 ...
 $ V84 : Factor w/ 90 levels "-0.0292","-0.0312",..: 90 5 5 5 5 5 5 5 5 5 ...
 $ V85 : Factor w/ 90 levels "-0.0122","-0.0334",..: 90 17 17 17 17 17 17 17 1
7 17 ...
 $ V86 : Factor w/ 90 levels "-0.0369","-0.0649",..: 90 45 45 45 45 45 45 45 4
5 45 ...
 $ V87 : Factor w/ 90 levels "-0.0084","-0.0126",..: 90 8 8 8 8 8 8 8 8 8 ...
 $ V88 : Factor w/ 85 levels "-18","-22.8",..: 85 40 40 40 40 40 40 40 40 40 .
..
 $ V89 : Factor w/ 80 levels "-17.6","-19.4",..: 80 28 28 28 28 28 28 28 28 28
...
 $ V90 : Factor w/ 44 levels "-0.1","-0.2",..: 44 1 1 1 1 1 1 1 1 1 ...
 $ V91 : Factor w/ 83 levels "-1.2","-123.3",..: 83 10 10 10 10 10 10 10 10 10
...
 $ V92 : Factor w/ 85 levels "-113.8","-118.5",..: 85 52 52 52 52 52 52 52 52
52 ...
 $ V93 : Factor w/ 44 levels "-0.1","-0.2",..: 44 1 1 1 1 1 1 1 1 1 ...
 $ V94 : Factor w/ 90 levels "0","0.54","0.9",..: 90 58 58 58 58 58 58 58 58 5
8 ...
 $ V95 : Factor w/ 90 levels "0","0.61","1.02",..: 90 75 75 75 75 75 75 75 75
75 ...
 $ V96 : Factor w/ 2 levels "0","amplitude_yaw_dumbbell": 2 1 1 1 1 1 1 1 1 1
...
 $ V97 : Factor w/ 34 levels "1","10","11",..: 34 28 28 28 29 28 28 28 28 28 .
..
 $ V98 : Factor w/ 87 levels "0","0.0196","0.0204",..: 87 47 47 47 47 47 47 47
47 47 ...
 $ V99 : Factor w/ 90 levels "-10.4132","-11.0964",..: 90 30 30 30 30 30 30 30
30 30 ...
  [list output truncated]
> set.seed(1234)
> ind = sample(1:nrow(df),0.8*nrow(df),replace = F)
> df_train =df[ind,-1]
> df_test = df[-ind,-1]
> summary(df)
         V1                 V2             V3                       V4
```

```
adelmo    : 311   1323084336:  61   160329 :   3   5/12/2011 14:23 :1585
carlitos :1580   1323095007:  61   160332 :   3   5/12/2011 11:25 :1243
eurico   :  88   1322832938:  60   188310 :   3   5/12/2011 14:22 : 456
jeremy   :   4   1323084327:  60   248321 :   3   5/12/2011 11:23 : 337
pedro    :2041   1323084346:  60   512311 :   3   2/12/2011 13:35 : 311
user_name:   1   1323084354:  60   536294 :   3   28/11/2011 14:15:  88
         V5                  V6                  V7                  V8                  V9
new_window:   1    29      :  61   123     : 520   26.1    : 175   -93.2  : 150
no        :3936    79      :  61   122     : 411   26.2    : 161   -93.1  : 140
yes       :  88    20      :  60   124     : 382   25.9    : 159   -93.3  : 140
                   37      :  60   121     : 252   26      : 132   -94.1  : 104
                   4       :  60   125     : 201   26.4    : 125   -93.8  :  96
                   43      :  60   129     : 115   25.8    : 112   -94.4  :  64
     V10                V11                 V12                V13
20     :1042   -1.03566 :3936   -0.39133 :3944   0.005406 :3936
3      : 826   -0.076054:   1   -0.15095 :   2   -0.011683:   1
19     : 780   -0.120447:   1   -2.060105:   2   -0.037647:   1
2      : 296   -0.157538:   1   0.988872 :   2   -0.068863:   1
21     : 231   -0.313326:   1   -0.06016 :   1   -0.130813:   1
4      : 217   -0.395556:   1   -0.108371:   1   -0.160041:   1
       V14               V15                V16                V17               V18
0.045115 :3944   -4.1    :3939   20      :3959   -1     :3941   -7.25  :3936
-1.717824:   2   -93.1   :   7   21      :  23   -0.7   :   9   -93.2  :   5
-0.045472:   1   -2.9    :   3   3       :  12   -1.1   :   9   -93.4  :   4
-0.04606 :   1   -91.2   :   3   4       :   9   -1.4   :   7   -93.3  :   3
-0.059758:   1   -0.8    :   2   5       :   8   -1.5   :   7   -94    :   3
-0.104399:   1   -1.5    :   2   22      :   2   -1.3   :   6   -94.4  :   3
     V19             V20             V21             V22
18     :3954   -1     :3941   1.345  :3936   2      :3949
19     :  22   -0.7   :   9   0.1    :   7   1      :  32
1      :  14   -1.1   :   9   0.5    :   5   3      :  16
3      :  12   -1.4   :   7   0.2    :   3   0      :   7
2      :   6   -1.5   :   7   0.3    :   3   4      :   7
20     :   5   -1.3   :   6   0.7    :   3   5      :   5
                  V23             V24             V25             V26
0                   :4024   0.3    :3948   121.9  :3936   0.6    :3941
amplitude_yaw_belt:   1   0.2    :  20   1.4    :   6   0.5    :  10
                          0      :  10   1.1    :   5   0.1    :   8
                          0.5    :   7   123    :   5   0      :   7
                          0.1    :   6   122.1  :   3   0.4    :   7
                          0.4    :   4   124    :   3   0.2    :   5
     V27             V28             V29             V30             V31
0.35   :3936   25.75  :3936   0.35   :3936   0.1    :3953   -4.95  :3936
0      :  19   25.9   :   6   0.1    :  15   0      :  29   -93.3  :   6
0.2    :  10   26     :   6   0.2    :  15   0.2    :  10   -3.1   :   3
0.1    :   8   26.2   :   4   0.3    :  11   0.8    :   4   -4.6   :   3
0.3    :   7   7.4    :   4   0.5    :  10   0.3    :   3   -92.9  :   3
0.6    :   5   26.4   :   3   0.7    :   5   0.4    :   3   -2.2   :   2
     V32             V33             V34             V35             V36
0.4    :3947   0.17   :3939   0.02   : 242   -0.03  :1025   -0.44  : 420
0.1    :  11   0      :  13   -0.43  : 222   -0.02  : 824   -0.43  : 370
0.3    :  10   0.01   :   5   0      : 214   0      : 742   -0.46  : 361
0      :   9   0.03   :   4   -0.42  : 198   -0.05  : 389   -0.02  : 339
0.5    :   9   0.08   :   3   -0.4   : 179   0.02   : 283   -0.41  : 231
0.2    :   6   0.12   :   3   -0.45  : 164   0.03   : 114   -0.03  : 215
     V37             V38             V39             V40             V41
-42    : 270   70     : 431   -177   : 148   1      : 229   584    : 178
-41    : 260   69     : 353   -175   : 141   0      : 216   583    : 176
-43    : 225   71     : 345   -178   : 140   -1     : 207   582    : 157
-40    : 211   3      : 258   -172   : 133   2      : 197   581    : 135
-20    : 187   4      : 256   -174   : 133   4      : 177   585    : 128
-19    : 186   68     : 225   21     : 131   -2     : 173   580    : 125
     V42             V43             V44             V45             V46
-370   :  87   -131   :  76   20.7   :  22   -162   : 119   34     : 413
```

```
-368   :  83   -130   :  48   15.3   :  17   -161   :  76   25     : 179
-375   :  83   -133   :  48   17.7   :  17   -164   :  63   3      : 116
-376   :  81   -134   :  48   -26.4  :  13   -163   :  57   29     : 114
-373   :  79   -132   :  47   13.8   :  13   -165   :  23   23     : 107
-369   :  78   117    :  47   13.9   :  13   126    :  22   19     : 104
     v47              v48              v49              v50              v51
65.0977:3936   76.22175 :3936   16.1039:3936   259.3599:3936   -10.1695:3936
0      :   8   -128     :   1   0      :   2   0       :   2   -10.902 :   1
0.0278 :   1   -128.4898:   1   0.05   :   1   0.0025  :   1   -11.7236:   1
0.1308 :   1   -129.6863:   1   0.2323 :   1   0.054   :   1   -12.2027:   1
0.1616 :   1   -130.7826:   1   0.417  :   1   0.1739  :   1   -14.164 :   1
0.1883 :   1   -131.2292:   1   0.4247 :   1   0.1804  :   1   -14.9221:   1
      v52              v53              v54              v55
10.66725:3936   113.7978:3936   19.0615  :3936   35.8809:3936
0       :   1   0       :   1   -161     :   2   0      :   5
0.0548  :   1   0.003   :   1   -0.0558  :   1   0.1924 :   1
0.2209  :   1   0.0488  :   1   -1.7763  :   1   0.3087 :   1
0.2218  :   1   0.0492  :   1   -105.7773:   1   0.3594 :   1
0.297   :   1   0.0882  :   1   -115.4516:   1   0.4871 :   1
      v56              v57              v58              v59              v60
1287.463:3936   0.02   : 190   -0.02  : 258   -0.02  : 243   -289   : 109
0       :   5   0      : 128   -0.03  :  84   0      : 172   -288   :  85
0.037   :   1   -0.02  :  39   0      :  73   -0.03  : 117   -290   :  59
0.0953  :   1   -0.06  :  31   -0.05  :  52   -0.07  :  84   -287   :  51
0.1292  :   1   -0.05  :  30   -0.22  :  51   0.02   :  83   1      :  45
0.2373  :   1   -0.22  :  25   -0.24  :  45   0.16   :  70   -1     :  43
     v61              v62              v63              v64              v65
111    : 122   -123   :  89   -367   :  43   335    :  49   510    :  41
110    : 108   -122   :  79   -368   :  42   336    :  40   512    :  41
109    :  77   -124   :  71   -369   :  33   337    :  37   513    :  41
-10    :  41   -125   :  46   -372   :  30   338    :  36   511    :  37
6      :  39   -121   :  35   -370   :  29   339    :  36   515    :  37
-9     :  38   -88    :  31   -371   :  26   334    :  35   514    :  33
      v66              v67              v68              v69
-1.18224:3938   -0.96912:3942   -0.86977:3944   0.12353 :3938
-0.11926:   1   -0.10176:   1   -0.06791:   1   -0.00696:   1
-0.19002:   1   -0.15381:   1   -0.12096:   1   -0.01884:   1
-0.20488:   1   -0.15426:   1   -0.1697 :   1   -0.03359:   1
-0.34389:   1   -0.16444:   1   -0.20332:   1   -0.08596:   1
-0.36227:   1   -0.18233:   1   -0.21951:   1   -0.1009 :   1
      v70              v71              v72              v73              v74
-0.10319:3942   0.059765:3944   8.45   :3936   77.25  :3936   38     :3944
-0.01247:   1   -0.0046 :   1   -10.1  :   2   -161   :   3   34     :  11
-0.09627:   1   -0.008  :   1   -13.9  :   2   127    :   3   39     :   6
-0.14513:   1   -0.00863:   1   -3     :   2   -162   :   2   45     :   4
-0.14758:   1   -0.05777:   1   -6.9   :   2   -164   :   2   46     :   4
-0.17734:   1   -0.08516:   1   0      :   2   180    :   2   32     :   3
     v75              v76              v77              v78              v79
-33.6  :3936   -58.6  :3936   10     :3942   36.945 :3936   121.5  :3936
-4     :   2   -138   :   2   4      :   8   1.3    :   2   0      :   5
-65.5  :   2   -161   :   2   34     :   7   45.2   :   2   1      :   4
-10    :   1   -162   :   2   6      :   6   0      :   1   125.2  :   2
-11.5  :   1   -164   :   2   3      :   5   0.1    :   1   138.8  :   2
-12.7  :   1   -180   :   2   5      :   5   0.57   :   1   158.5  :   2
     v80                  v81                    v82                    v83
27     :3941   13.2272917 :  58   0            :  93   -84.40838906:  58
0      :   8   0          :  39   -70.91576678:  58   0           :  13
29     :   5   13.05217456:  11   -70.49400371:  11   -84.87393888:  11
31     :   5   13.07948887:   9   -70.63995378:   9   -84.64918975:   9
25     :   4   13.35451266:   9   -70.67116245:   9   -84.69053461:   9
16     :   3   55.82441814:   9   9.645819033 :   9   100.2280531 :   9
      v84              v85              v86              v87              v88
-0.09595:3936   -0.4422:3936   0.0819 :3936   -0.216 :3936   41.85  :3936
-0.0292 :   1   -0.0122:   1   -0.0369:   1   -0.0084:   1   -70    :   2
```

```
 -0.0312 :   1   -0.0334:   1   -0.0649:   1   -0.0126:   1   -70.1  :   2
 -0.0363 :   1   -0.0605:   1   -0.0876:   1   -0.0902:   1   38     :   2
 -0.0947 :   1   -0.0668:   1   -0.096 :   1   -0.1013:   1   48.4   :   2
 -0.0972 :   1   -0.1238:   1   -0.1135:   1   -0.163 :   1   49.9   :   2
      v89            v90            v91            v92            v93
 133    :3936   -0.1   :3938   -26.75 :3936   20.2   :3936   -0.1   :3938
 -84.3  :   3   -0.7   :   9   -2.4   :   2   -85.3  :   2   -0.7   :   9
 124    :   2   -0.9   :   5   -70.9  :   2   -85.4  :   2   -0.9   :   5
 133.9  :   2   -0.6   :   4   -71    :   2   -85.5  :   2   -0.6   :   4
 134.6  :   2   -0.8   :   4   10.6   :   2   0      :   2   -0.8   :   4
 137    :   2   0.2    :   4   10.9   :   2   21.9   :   2   0.2    :   4
      v94            v95                           v96            v97
 55.71  :3936   54.74  :3936   0                    :4024   9      : 513
 0      :   1   0      :   1   amplitude_yaw_dumbbell:   1   10     : 482
 0.54   :   1   0.61   :   1                                 8      : 328
 0.9    :   1   1.02   :   1                                 37     : 307
 0.96   :   1   1.1    :   1                                 11     : 227
 1.08   :   1   1.2    :   1                                 5      : 217
      v98              v99             v100           v101            v102
 2.41635:3936   -5.11805 :3936   17.058 :3936   291.001:3936   13.9312 :3936
 0      :   4   -10.4132 :   1   0      :   1   0      :   1   -0.6827 :   1
 0.0196 :   1   -11.0964 :   1   0.1811 :   1   0.0328 :   1   -1.7762 :   1
 0.0204 :   1   -110.9328:   1   0.2034 :   1   0.0414 :   1   -12.1577:   1
 0.0213 :   1   -12.4734 :   1   0.3015 :   1   0.0909 :   1   -17.0387:   1
 0.0217 :   1   -12.9177 :   1   0.3396 :   1   0.1153 :   1   -19.3215:   1
      v103             v104              v105            v106            v107
 14.1062:3936   199.0775:3936   64.7063  :3936   13.5747:3936   184.5578:3936
 0      :   1   0       :   1   -101.7805:   1   0      :   1   0       :   1
 0.1799 :   1   0.0324  :   1   -105.6502:   1   0.2172 :   1   0.0472  :   1
 0.2384 :   1   0.0568  :   1   -15.4101 :   1   0.256  :   1   0.0656  :   1
 0.2425 :   1   0.0588  :   1   -17.5064 :   1   0.2815 :   1   0.0793  :   1
 0.2687 :   1   0.0722  :   1   -30.8185 :   1   0.2864 :   1   0.082   :   1
      v108           v109           v110           v111           v112
 0      : 263   -0.02  : 327   0      : 247   -234   : 102   48     : 134
 0.47   : 126   -0.1   : 134   -0.02  : 169   5      :  97   47     :  90
 0.48   : 114   -0.08  : 129   -0.26  : 106   23     :  96   46     :  63
 0.45   : 109   -0.11  : 122   -0.23  : 103   17     :  95   1      :  54
 -0.02  : 103   -0.13  : 107   -0.25  : 103   18     :  94   -9     :  51
 -0.03  : 102   -0.06  : 104   -0.28  :  97   0      :  93   2      :  51
      v113           v114           v115           v116           v117
 -272   :  93   -552   :  77   290    :  77   -73    :  86   0      : 311
 -270   :  75   -558   :  46   295    :  43   -68    :  42   141    :  77
 84     :  67   -554   :  38   -533   :  37   -69    :  42   139    :  71
 -271   :  66   -555   :  38   292    :  35   -65    :  36   134    :  65
 79     :  61   533    :  37   -517   :  32   -71    :  35   140    :  61
 85     :  59   521    :  34   -522   :  32   -70    :  33   133    :  57
      v118           v119             v120             v121             v122
 0      : 312   0      : 311   -1.09475:3944   -0.97525:3944   -0.05065:3944
 -63.8  :  66   107    : 101   -1.3846 :   2   -0.0259 :   1   -0.009  :   1
 -63.5  :  53   108    :  85   -0.0699 :   1   -0.0918 :   1   -0.011  :   1
 -63.7  :  52   106    :  79   -0.0781 :   1   -0.1289 :   1   -0.0252 :   1
 -63.9  :  52   102    :  68   -0.1168 :   1   -0.1574 :   1   -0.0525 :   1
 -63.6  :  43   105    :  65   -0.1804 :   1   -0.2494 :   1   -0.0705 :   1
      v123           v124           v125           v126           v127
 0.17285:3944   49.6   :3938   168    :3938   -1.1   :3948   4.65   :3936
 -0.0428:   1   0      :   7   176    :   9   -1.3   :  14   0      :   7
 -0.0673:   1   -0.2   :   2   0      :   7   -1.5   :   6   -63.9  :   3
 -0.0732:   1   -63.4  :   2   174    :   7   -0.7   :   5   7.3    :   3
 -0.14  :   1   -63.6  :   2   180    :   6   -0.9   :   5   -63.7  :   2
 -0.2117:   1   53.4   :   2   171    :   4   -1     :   5   22.7   :   2
      v128           v129           v130           v131
 -168.5 :3936   -1.1   :3948   32.2   :3936   341.5  :3936
 -177   :   7   -1.3   :  14   0      :   8   354    :   9
 -178   :   7   -1.5   :   6   0.3    :   5   0      :   8
```

```
 0      :   7   -0.7   :   5   23.9   :   2   1      :   7
 -175   :   6   -0.9   :   5   41.6   :   2   350    :   7
 -176   :   6   -1     :   5   0.15   :   1   351    :   7
                   V132            V133           V134             V135
 0                   :4024   36     : 593   14.0772:3936   27.85936  :3936
 amplitude_yaw_forearm:   1   35     : 366   0      :   7   0         :   7
                             34     : 343   0.02083:   1   -136.38298:   1
                             37     : 242   0.03701:   1   -14.5     :   1
                             33     : 226   0.25391:   1   -145.13953:   1
                             28     : 167   0.26768:   1   -18.14035 :   1
      V136             V137             V138             V139
 45.16342:3936   2749.163:3936   25.35597:3936   8.906695:3936
 0       :   7   0       :   7   0       :   7   0        :   8
 0.04082 :   1   0.00167 :   1   -0.00125:   1   0.06088 :   1
 0.12963 :   1   0.0168  :   1   -0.59022:   1   0.06652 :   1
 0.15744 :   1   0.02479 :   1   -0.84531:   1   0.07474 :   1
 0.31045 :   1   0.09638 :   1   -1.01095:   1   0.07814 :   1
      V140             V141             V142             V143
 79.33451:3936   17.09505  :3936   74.27584:3936   5541.956:3936
 0       :   8   0         :   7   0       :   8   0       :   8
 0.00371 :   1   -1.07213  :   1   0.05774 :   1   0.00333 :   1
 0.00442 :   1   -100.77442:   1   0.20841 :   1   0.04343 :   1
 0.00559 :   1   -12.26034 :   1   0.39038 :   1   0.15239 :   1
 0.00611 :   1   -13.03158 :   1   0.44344 :   1   0.19664 :   1
     V144           V145           V146           V147           V148
 0.02   : 225   -0.02  : 173   -0.02  : 207   192    :  82   204    :  78
 0      : 132   0      : 171   -0.03  : 154   193    :  67   205    :  68
 0.03   : 112   -0.03  :  85   0      : 138   191    :  55   203    :  59
 0.06   :  92   0.02   :  82   -0.05  : 111   194    :  44   202    :  52
 0.08   :  89   0.03   :  58   -0.07  :  69   190    :  36   206    :  45
 -0.02  :  76   -0.05  :  42   0.02   :  69   195    :  26   207    :  33
     V149           V150           V151           V152
 -214   : 119   -10    :  38   655    :  51   472    :  37
 -213   :  98   -12    :  37   656    :  47   471    :  33
 -215   :  98   -11    :  35   653    :  44   469    :  32
 -216   :  70   -13    :  34   751    :  43   470    :  30
 -191   :  53   -9     :  30   654    :  41   468    :  27
 -188   :  52   -14    :  28   749    :  40   467    :  25
 [ reached getOption("max.print") -- omitted 1 row ]
> dim(df)
[1] 4025  152
```

```
$ V29 : Factor w/ 23 levels "0","0.1","0.2",..: 23 5 5 5 5 5 5 5 5 5 ...
 $ V30 : Factor w/ 24 levels "0","0.1","0.2",..: 24 2 2 2 2 2 2 2 2 2 ...
 $ V31 : Factor w/ 70 levels "-0.6","-0.8",..: 70 25 25 25 25 25 25 25 25 25
...
 $ V32 : Factor w/ 21 levels "0","0.1","0.2",..: 21 5 5 5 5 5 5 5 5 5 ...
 $ V33 : Factor w/ 56 levels "0","0.01","0.02",..: 56 17 17 17 17 17 17 17 17
17 ...
 $ V34 : Factor w/ 136 levels "-0.02","-0.03",..: 136 135 134 133 132 130 129
131 129 128 ...
 $ V35 : Factor w/ 51 levels "-0.02","-0.03",..: 51 37 35 31 28 26 2 4 4 2 ..
.
 $ V36 : Factor w/ 99 levels "-0.02","-0.03",..: 99 49 51 51 53 56 58 57 62 6
8 ...
 $ V37 : Factor w/ 154 levels "-1","-10","-11",..: 154 25 14 14 55 36 84 84 9
0 1 ...
 $ V38 : Factor w/ 113 levels "-1","-10","-11",..: 113 9 5 21 19 25 25 14 16
10 ...
 $ V39 : Factor w/ 214 levels "-1","-10","-11",..: 214 155 148 212 204 140 13
5 138 139 137 ...
```

```
 $ V40 : Factor w/ 247 levels "-1","-10","-11",..: 247 192 202 204 208 206 21
3 215 216 218 ...
 $ V41 : Factor w/ 187 levels "428","435","437",..: 187 63 51 50 52 47 49 47
42 46 ...
 $ V42 : Factor w/ 372 levels "-112","-115",..: 372 83 70 60 38 28 18 12 5 4
...
 $ V43 : Factor w/ 1238 levels "-0.02","-0.05",..: 1238 671 668 664 659 654 6
49 643 1226 1182 ...
 $ V44 : Factor w/ 1694 levels "-0.01","-0.05",..: 1694 517 533 546 557 565 5
70 573 571 564 ...
 $ V45 : Factor w/ 1670 levels "-0.18","-0.23",..: 1670 453 407 371 318 278 2
28 158 48 336 ...
 $ V46 : Factor w/ 60 levels "1","10","11",..: 60 32 32 29 29 28 27 22 21 20
...
 $ V47 : Factor w/ 83 levels "0","0.0278","0.1308",..: 83 62 62 62 62 62 62 6
2 62 62 ...
 $ V48 : Factor w/ 90 levels "-128","-128.4898",..: 90 74 74 74 74 74 74 74 7
4 74 ...
 $ V49 : Factor w/ 89 levels "0","0.05","0.2323",..: 89 21 21 21 21 21 21 21
21 21 ...
 $ V50 : Factor w/ 89 levels "0","0.0025","0.054",..: 89 30 30 30 30 30 30 30
30 30 ...
 $ V51 : Factor w/ 90 levels "-10.1695","-10.902",..: 90 1 1 1 1 1 1 1 1 1 ..
.
 $ V52 : Factor w/ 90 levels "0","0.0548","0.2209",..: 90 20 20 20 20 20 20 2
0 20 20 ...
 $ V53 : Factor w/ 90 levels "0","0.003","0.0488",..: 90 19 19 19 19 19 19 19
19 19 ...
 $ V54 : Factor w/ 89 levels "-0.0558","-1.7763",..: 89 51 51 51 51 51 51 51
51 51 ...
 $ V55 : Factor w/ 86 levels "0","0.1924","0.3087",..: 86 41 41 41 41 41 41 4
1 41 41 ...
 $ V56 : Factor w/ 86 levels "0","0.037","0.0953",..: 86 15 15 15 15 15 15 15
15 15 ...
 $ V57 : Factor w/ 543 levels "-0.02","-0.03",..: 543 457 466 473 484 491 498
510 512 504 ...
 $ V58 : Factor w/ 318 levels "-0.02","-0.03",..: 318 38 40 43 45 48 52 52 52
52 ...
 $ V59 : Factor w/ 213 levels "-0.02","-0.03",..: 213 1 7 9 14 15 18 19 13 7
...
 $ V60 : Factor w/ 707 levels "-1","-10","-100",..: 707 339 342 353 355 361 3
58 363 350 339 ...
 $ V61 : Factor w/ 419 levels "-1","-10","-100",..: 419 343 348 358 367 370 3
74 383 387 395 ...
 $ V62 : Factor w/ 646 levels "-1","-10","-100",..: 646 258 250 219 218 198 1
85 136 128 115 ...
 $ V63 : Factor w/ 1151 levels "-1","-10","-100",..: 1151 892 937 952 988 101
5 1042 1070 1074 1091 ...
 $ V64 : Factor w/ 817 levels "-1","-10","-100",..: 817 110 111 102 89 79 78
66 57 38 ...
 $ V65 : Factor w/ 1038 levels "-1","-10","-100",..: 1038 245 214 201 159 138
95 24 7 290 ...
 $ V66 : Factor w/ 88 levels "-0.11926","-0.19002",..: 88 33 33 33 33 33 33 3
3 33 33 ...
 $ V67 : Factor w/ 84 levels "-0.10176","-0.15381",..: 84 28 28 28 28 28 28 2
8 28 28 ...
```

```
 $ V68 : Factor w/ 82 levels "-0.06791","-0.12096",..: 82 26 26 26 26 26 26 2
6 26 26 ...
 $ V69 : Factor w/ 88 levels "-0.00696","-0.01884",..: 88 44 44 44 44 44 44 4
4 44 44 ...
 $ V70 : Factor w/ 84 levels "-0.01247","-0.09627",..: 84 3 3 3 3 3 3 3 3 3 .
..
 $ V71 : Factor w/ 82 levels "-0.0046","-0.008",..: 82 41 41 41 41 41 41 41 4
1 41 ...
 $ V72 : Factor w/ 83 levels "-0.7","-1.1",..: 83 80 80 80 80 80 80 80 80 80
...
 $ V73 : Factor w/ 80 levels "-1.9","-122",..: 80 63 63 63 63 63 63 63 63 63
...
 $ V74 : Factor w/ 39 levels "19","20","21",..: 39 17 17 17 17 17 17 17 17 17
...
 $ V75 : Factor w/ 88 levels "-10","-11.5",..: 88 14 14 14 14 14 14 14 14 14
...
 $ V76 : Factor w/ 84 levels "-10.9","-100",..: 84 43 43 43 43 43 43 43 43 43
...
 $ V77 : Factor w/ 30 levels "1","10","11",..: 30 2 2 2 2 2 2 2 2 2 ...
 $ V78 : Factor w/ 88 levels "0","0.1","0.57",..: 88 42 42 42 42 42 42 42 42
42 ...
 $ V79 : Factor w/ 78 levels "0","1","100.2",..: 78 13 13 13 13 13 13 13 13 1
3 ...
 $ V80 : Factor w/ 41 levels "0","1","12","13",..: 41 17 17 17 17 17 17 17 17
17 ...
 $ V81 : Factor w/ 3653 levels "-0.949086746",..: 3653 2993 3032 3029 3034 30
26 3017 3028 3052 3017 ...
 $ V82 : Factor w/ 3572 levels "-0.738381151",..: 3572 1153 3542 3326 1118 11
29 1313 1328 3378 1313 ...
 $ V83 : Factor w/ 3602 levels "-10.50214329",..: 3602 698 627 644 3598 633 6
39 3591 3572 639 ...
 $ V84 : Factor w/ 90 levels "-0.0292","-0.0312",..: 90 5 5 5 5 5 5 5 5 5 ...
 $ V85 : Factor w/ 90 levels "-0.0122","-0.0334",..: 90 17 17 17 17 17 17 17
17 17 ...
 $ V86 : Factor w/ 90 levels "-0.0369","-0.0649",..: 90 45 45 45 45 45 45 45
45 45 ...
 $ V87 : Factor w/ 90 levels "-0.0084","-0.0126",..: 90 8 8 8 8 8 8 8 8 8 ...
 $ V88 : Factor w/ 85 levels "-18","-22.8",..: 85 40 40 40 40 40 40 40 40 40
...
 $ V89 : Factor w/ 80 levels "-17.6","-19.4",..: 80 28 28 28 28 28 28 28 28 2
8 ...
 $ V90 : Factor w/ 44 levels "-0.1","-0.2",..: 44 1 1 1 1 1 1 1 1 1 ...
 $ V91 : Factor w/ 83 levels "-1.2","-123.3",..: 83 10 10 10 10 10 10 10 10 1
0 ...
 $ V92 : Factor w/ 85 levels "-113.8","-118.5",..: 85 52 52 52 52 52 52 52 52
52 ...
 $ V93 : Factor w/ 44 levels "-0.1","-0.2",..: 44 1 1 1 1 1 1 1 1 1 ...
 $ V94 : Factor w/ 90 levels "0","0.54","0.9",..: 90 58 58 58 58 58 58 58 58
58 ...
 $ V95 : Factor w/ 90 levels "0","0.61","1.02",..: 90 75 75 75 75 75 75 75 75
75 ...
 $ V96 : Factor w/ 2 levels "0","amplitude_yaw_dumbbell": 2 1 1 1 1 1 1 1 1 1
...
 $ V97 : Factor w/ 34 levels "1","10","11",..: 34 28 28 28 29 28 28 28 28 28
...
 $ V98 : Factor w/ 87 levels "0","0.0196","0.0204",..: 87 47 47 47 47 47 47 4
7 47 47 ...
```

```
 $ V99 : Factor w/ 90 levels "-10.4132","-11.0964",..: 90 30 30 30 30 30 30 3
0 30 30 ...
  [list output truncated]
>
> set.seed(1234)
> ind = sample(1:nrow(df),0.8*nrow(df),replace = F)
> df_train =df[ind,-1]
> df_test = df[-ind,-1]
>
> summary(df)
         V1                 V2               V3                      V4
 adelmo   : 311   1323084336:  61   160329 :   3   5/12/2011 14:23 :1585
 carlitos :1580   1323095007:  61   160332 :   3   5/12/2011 11:25 :1243
 eurico   :  88   1322832938:  60   188310 :   3   5/12/2011 14:22 : 456
 jeremy   :   4   1323084327:  60   248321 :   3   5/12/2011 11:23 : 337
 pedro    :2041   1323084346:  60   512311 :   3   2/12/2011 13:35 : 311
 user_name:   1   1323084354:  60   536294 :   3   28/11/2011 14:15:  88
          V5              V6               V7               V8               V9
 new_window:   1   29     :  61   123    : 520   26.1   : 175   -93.2  : 150
 no        :3936   79     :  61   122    : 411   26.2   : 161   -93.1  : 140
 yes       :  88   20     :  60   124    : 382   25.9   : 159   -93.3  : 140
                   37     :  60   121    : 252   26     : 132   -94.1  : 104
                   4      :  60   125    : 201   26.4   : 125   -93.8  :  96
                   43     :  60   129    : 115   25.8   : 112   -94.4  :  64
      V10              V11               V12               V13
 20     :1042   -1.03566 :3936   -0.39133 :3944   0.005406 :3936
 3      : 826   -0.076054:   1   -0.15095 :   2   -0.011683:   1
 19     : 780   -0.120447:   1   -2.060105:   2   -0.037647:   1
 2      : 296   -0.157538:   1   0.988872 :   2   -0.068863:   1
 21     : 231   -0.313326:   1   -0.06016 :   1   -0.130813:   1
 4      : 217   -0.395556:   1   -0.108371:   1   -0.160041:   1
        V14              V15              V16              V17              V18
 0.045115 :3944   -4.1   :3939   20     :3959   -1     :3941   -7.25  :3936
 -1.717824:   2   -93.1  :   7   21     :  23   -0.7   :   9   -93.2  :   5
 -0.045472:   1   -2.9   :   3   3      :  12   -1.1   :   9   -93.4  :   4
 -0.04606 :   1   -91.2  :   3   4      :   9   -1.4   :   7   -93.3  :   3
 -0.059758:   1   -0.8   :   2   5      :   8   -1.5   :   7   -94    :   3
 -0.104399:   1   -1.5   :   2   22     :   2   -1.3   :   6   -94.4  :   3
      V19              V20              V21              V22
 18     :3954   -1     :3941   1.345  :3936   2      :3949
 19     :  22   -0.7   :   9   0.1    :   7   1      :  32
 1      :  14   -1.1   :   9   0.5    :   5   3      :  16
 3      :  12   -1.4   :   7   0.2    :   3   0      :   7
 2      :   6   -1.5   :   7   0.3    :   3   4      :   7
 20     :   5   -1.3   :   6   0.7    :   3   5      :   5
               V23              V24              V25              V26
 0                :4024   0.3    :3948   121.9  :3936   0.6    :3941
 amplitude_yaw_belt:   1   0.2    :  20   1.4    :   6   0.5    :  10
                          0      :  10   1.1    :   5   0.1    :   8
                          0.5    :   7   123    :   5   0      :   7
                          0.1    :   6   122.1  :   3   0.4    :   7
                          0.4    :   4   124    :   3   0.2    :   5
      V27              V28              V29              V30              V31
 0.35   :3936   25.75  :3936   0.35   :3936   0.1    :3953   -4.95  :3936
 0      :  19   25.9   :   6   0.1    :  15   0      :  29   -93.3  :   6
 0.2    :  10   26     :   6   0.2    :  15   0.2    :  10   -3.1   :   3
 0.1    :   8   26.2   :   4   0.3    :  11   0.8    :   4   -4.6   :   3
```

```
 0.3    :   7   7.4    :   4   0.5    :  10   0.3    :   3   -92.9  :   3
 0.6    :   5   26.4   :   3   0.7    :   5   0.4    :   3   -2.2   :   2
      v32            v33            v34            v35            v36
 0.4    :3947   0.17   :3939   0.02   : 242   -0.03  :1025   -0.44  : 420
 0.1    :  11   0      :  13   -0.43  : 222   -0.02  : 824   -0.43  : 370
 0.3    :  10   0.01   :   5   0      : 214   0      : 742   -0.46  : 361
 0      :   9   0.03   :   4   -0.42  : 198   -0.05  : 389   -0.02  : 339
 0.5    :   9   0.08   :   3   -0.4   : 179   0.02   : 283   -0.41  : 231
 0.2    :   6   0.12   :   3   -0.45  : 164   0.03   : 114   -0.03  : 215
      v37            v38            v39            v40            v41
 -42    : 270   70     : 431   -177   : 148   1      : 229   584    : 178
 -41    : 260   69     : 353   -175   : 141   0      : 216   583    : 176
 -43    : 225   71     : 345   -178   : 140   -1     : 207   582    : 157
 -40    : 211   3      : 258   -172   : 133   2      : 197   581    : 135
 -20    : 187   4      : 256   -174   : 133   4      : 177   585    : 128
 -19    : 186   68     : 225   21     : 131   -2     : 173   580    : 125
      v42            v43            v44            v45            v46
 -370   :  87   -131   :  76   20.7   :  22   -162   : 119   34     : 413
 -368   :  83   -130   :  48   15.3   :  17   -161   :  76   25     : 179
 -375   :  83   -133   :  48   17.7   :  17   -164   :  63   3      : 116
 -376   :  81   -134   :  48   -26.4  :  13   -163   :  57   29     : 114
 -373   :  79   -132   :  47   13.8   :  13   -165   :  23   23     : 107
 -369   :  78   117    :  47   13.9   :  13   126    :  22   19     : 104
      v47              v48              v49              v50              v51
 65.0977:3936   76.22175 :3936   16.1039:3936   259.3599:3936   -10.1695:3936
 0      :   8   -128      :   1   0      :   2   0        :   2   -10.902 :   1
 0.0278 :   1   -128.4898:   1   0.05   :   1   0.0025   :   1   -11.7236:   1
 0.1308 :   1   -129.6863:   1   0.2323 :   1   0.054    :   1   -12.2027:   1
 0.1616 :   1   -130.7826:   1   0.417  :   1   0.1739   :   1   -14.164 :   1
 0.1883 :   1   -131.2292:   1   0.4247 :   1   0.1804   :   1   -14.9221:   1
       v52              v53              v54              v55
 10.66725:3936   113.7978:3936   19.0615  :3936   35.8809:3936
 0       :   1   0       :   1   -161      :   2   0      :   5
 0.0548  :   1   0.003   :   1   -0.0558   :   1   0.1924 :   1
 0.2209  :   1   0.0488  :   1   -1.7763   :   1   0.3087 :   1
 0.2218  :   1   0.0492  :   1   -105.7773:   1   0.3594 :   1
 0.297   :   1   0.0882  :   1   -115.4516:   1   0.4871 :   1
       v56            v57            v58            v59            v60
 1287.463:3936   0.02   : 190   -0.02  : 258   -0.02  : 243   -289   : 109
 0       :   5   0      : 128   -0.03  :  84   0      : 172   -288   :  85
 0.037   :   1   -0.02  :  39   0      :  73   -0.03  : 117   -290   :  59
 0.0953  :   1   -0.06  :  31   -0.05  :  52   -0.07  :  84   -287   :  51
 0.1292  :   1   -0.05  :  30   -0.22  :  51   0.02   :  83   1      :  45
 0.2373  :   1   -0.22  :  25   -0.24  :  45   0.16   :  70   -1     :  43
      v61            v62            v63            v64            v65
 111    : 122   -123   :  89   -367   :  43   335    :  49   510    :  41
 110    : 108   -122   :  79   -368   :  42   336    :  40   512    :  41
 109    :  77   -124   :  71   -369   :  33   337    :  37   513    :  41
 -10    :  41   -125   :  46   -372   :  30   338    :  36   511    :  37
 6      :  39   -121   :  35   -370   :  29   339    :  36   515    :  37
 -9     :  38   -88    :  31   -371   :  26   334    :  35   514    :  33
       v66              v67              v68              v69
 -1.18224:3938   -0.96912:3942   -0.86977:3944   0.12353 :3938
 -0.11926:   1   -0.10176:   1   -0.06791:   1   -0.00696:   1
 -0.19002:   1   -0.15381:   1   -0.12096:   1   -0.01884:   1
 -0.20488:   1   -0.15426:   1   -0.1697 :   1   -0.03359:   1
 -0.34389:   1   -0.16444:   1   -0.20332:   1   -0.08596:   1
```

```
 -0.36227:   1   -0.18233:   1   -0.21951:   1   -0.1009 :   1
       v70              v71              v72              v73              v74
 -0.10319:3942   0.059765:3944   8.45   :3936   77.25  :3936   38     :3944
 -0.01247:   1   -0.0046 :   1   -10.1  :   2   -161   :   3   34     :  11
 -0.09627:   1   -0.008  :   1   -13.9  :   2   127    :   3   39     :   6
 -0.14513:   1   -0.00863:   1   -3     :   2   -162   :   2   45     :   4
 -0.14758:   1   -0.05777:   1   -6.9   :   2   -164   :   2   46     :   4
 -0.17734:   1   -0.08516:   1   0      :   2   180    :   2   32     :   3
      v75              v76              v77              v78              v79
 -33.6  :3936   -58.6  :3936   10     :3942   36.945 :3936   121.5  :3936
 -4     :   2   -138   :   2   4      :   8   1.3    :   2   0      :   5
 -65.5  :   2   -161   :   2   34     :   7   45.2   :   2   1      :   4
 -10    :   1   -162   :   2   6      :   6   0      :   1   125.2  :   2
 -11.5  :   1   -164   :   2   3      :   5   0.1    :   1   138.8  :   2
 -12.7  :   1   -180   :   2   5      :   5   0.57   :   1   158.5  :   2
      v80                 v81                    v82                    v83
 27     :3941   13.2272917 :  58   0             :  93   -84.40838906:  58
 0      :   8   0          :  39   -70.91576678:  58   0           :  13
 29     :   5   13.05217456:  11   -70.49400371:  11   -84.87393888:  11
 31     :   5   13.07948887:   9   -70.63995378:   9   -84.64918975:   9
 25     :   4   13.35451266:   9   -70.67116245:   9   -84.69053461:   9
 16     :   3   55.82441814:   9   9.645819033 :   9   100.2280531 :   9
       v84              v85              v86              v87              v88
 -0.09595:3936   -0.4422:3936   0.0819 :3936   -0.216 :3936   41.85  :3936
 -0.0292 :   1   -0.0122:   1   -0.0369:   1   -0.0084:   1   -70    :   2
 -0.0312 :   1   -0.0334:   1   -0.0649:   1   -0.0126:   1   -70.1  :   2
 -0.0363 :   1   -0.0605:   1   -0.0876:   1   -0.0902:   1   38     :   2
 -0.0947 :   1   -0.0668:   1   -0.096 :   1   -0.1013:   1   48.4   :   2
 -0.0972 :   1   -0.1238:   1   -0.1135:   1   -0.163 :   1   49.9   :   2
      v89              v90              v91              v92              v93
 133    :3936   -0.1   :3938   -26.75 :3936   20.2   :3936   -0.1   :3938
 -84.3  :   3   -0.7   :   9   -2.4   :   2   -85.3  :   2   -0.7   :   9
 124    :   2   -0.9   :   5   -70.9  :   2   -85.4  :   2   -0.9   :   5
 133.9  :   2   -0.6   :   4   -71    :   2   -85.5  :   2   -0.6   :   4
 134.6  :   2   -0.8   :   4   10.6   :   2   0      :   2   -0.8   :   4
 137    :   2   0.2    :   4   10.9   :   2   21.9   :   2   0.2    :   4
      v94              v95                             v96              v97
 55.71  :3936   54.74  :3936   0                      :4024   9      : 513
 0      :   1   0      :   1   amplitude_yaw_dumbbell:   1   10     : 482
 0.54   :   1   0.61   :   1                                 8      : 328
 0.9    :   1   1.02   :   1                                 37     : 307
 0.96   :   1   1.1    :   1                                 11     : 227
 1.08   :   1   1.2    :   1                                 5      : 217
      v98              v99              v100             v101             v102
 2.41635:3936   -5.11805 :3936   17.058 :3936   291.001:3936   13.9312 :3936
 0      :   4   -10.4132 :   1   0      :   1   0      :   1   -0.6827 :   1
 0.0196 :   1   -11.0964 :   1   0.1811 :   1   0.0328 :   1   -1.7762 :   1
 0.0204 :   1   -110.9328:   1   0.2034 :   1   0.0414 :   1   -12.1577:   1
 0.0213 :   1   -12.4734 :   1   0.3015 :   1   0.0909 :   1   -17.0387:   1
 0.0217 :   1   -12.9177 :   1   0.3396 :   1   0.1153 :   1   -19.3215:   1
      v103             v104              v105             v106             v107
 14.1062:3936   199.0775:3936   64.7063  :3936   13.5747:3936   184.5578:3936
 0      :   1   0       :   1   -101.7805:   1   0      :   1   0       :   1
 0.1799 :   1   0.0324  :   1   -105.6502:   1   0.2172 :   1   0.0472  :   1
 0.2384 :   1   0.0568  :   1   -15.4101 :   1   0.256  :   1   0.0656  :   1
 0.2425 :   1   0.0588  :   1   -17.5064 :   1   0.2815 :   1   0.0793  :   1
 0.2687 :   1   0.0722  :   1   -30.8185 :   1   0.2864 :   1   0.082   :   1
```

```
     v108          v109          v110          v111          v112    
0      : 263   -0.02  : 327   0      : 247   -234   : 102   48     : 134  
0.47   : 126   -0.1   : 134   -0.02  : 169   5      :  97   47     :  90  
0.48   : 114   -0.08  : 129   -0.26  : 106   23     :  96   46     :  63  
0.45   : 109   -0.11  : 122   -0.23  : 103   17     :  95   1      :  54  
-0.02  : 103   -0.13  : 107   -0.25  : 103   18     :  94   -9     :  51  
-0.03  : 102   -0.06  : 104   -0.28  :  97   0      :  93   2      :  51  
     v113          v114          v115          v116          v117    
-272   :  93   -552   :  77   290    :  77   -73    :  86   0      : 311  
-270   :  75   -558   :  46   295    :  43   -68    :  42   141    :  77  
84     :  67   -554   :  38   -533   :  37   -69    :  42   139    :  71  
-271   :  66   -555   :  38   292    :  35   -65    :  36   134    :  65  
79     :  61   533    :  37   -517   :  32   -71    :  35   140    :  61  
85     :  59   521    :  34   -522   :  32   -70    :  33   133    :  57  
     v118          v119           v120            v121            v122     
0      : 312   0      : 311   -1.09475:3944   -0.97525:3944   -0.05065:3944  
-63.8  :  66   107    : 101   -1.3846 :   2   -0.0259 :   1   -0.009  :   1  
-63.5  :  53   108    :  85   -0.0699 :   1   -0.0918 :   1   -0.011  :   1  
-63.7  :  52   106    :  79   -0.0781 :   1   -0.1289 :   1   -0.0252 :   1  
-63.9  :  52   102    :  68   -0.1168 :   1   -0.1574 :   1   -0.0525 :   1  
-63.6  :  43   105    :  65   -0.1804 :   1   -0.2494 :   1   -0.0705 :   1  
     v123          v124          v125          v126          v127    
0.17285:3944   49.6   :3938   168    :3938   -1.1   :3948   4.65   :3936  
-0.0428:   1   0      :   7   176    :   9   -1.3   :  14   0      :   7  
-0.0673:   1   -0.2   :   2   0      :   7   -1.5   :   6   -63.9  :   3  
-0.0732:   1   -63.4  :   2   174    :   7   -0.7   :   5   7.3    :   3  
-0.14  :   1   -63.6  :   2   180    :   6   -0.9   :   5   -63.7  :   2  
-0.2117:   1   53.4   :   2   171    :   4   -1     :   5   22.7   :   2  
     v128          v129          v130          v131    
-168.5 :3936   -1.1   :3948   32.2   :3936   341.5  :3936  
-177   :   7   -1.3   :  14   0      :   8   354    :   9  
-178   :   7   -1.5   :   6   0.3    :   5   0      :   8  
0      :   7   -0.7   :   5   23.9   :   2   1      :   7  
-175   :   6   -0.9   :   5   41.6   :   2   350    :   7  
-176   :   6   -1     :   5   0.15   :   1   351    :   7  
                    v132           v133          v134             v135       
0                    :4024   36     : 593   14.0772:3936   27.85936  :3936  
amplitude_yaw_forearm:   1   35     : 366   0      :   7   0         :   7  
                             34     : 343   0.02083:   1   -136.38298:   1  
                             37     : 242   0.03701:   1   -14.5     :   1  
                             33     : 226   0.25391:   1   -145.13953:   1  
                             28     : 167   0.26768:   1   -18.14035 :   1  
      v136            v137            v138            v139      
45.16342:3936   2749.163:3936   25.35597:3936   8.906695:3936  
0       :   7   0       :   7   0       :   7   0       :   8  
0.04082 :   1   0.00167 :   1   -0.00125:   1   0.06088 :   1  
0.12963 :   1   0.0168  :   1   -0.59022:   1   0.06652 :   1  
0.15744 :   1   0.02479 :   1   -0.84531:   1   0.07474 :   1  
0.31045 :   1   0.09638 :   1   -1.01095:   1   0.07814 :   1  
      v140              v141             v142            v143      
79.33451:3936   17.09505  :3936   74.27584:3936   5541.956:3936  
0       :   8   0         :   7   0       :   8   0       :   8  
0.00371 :   1   -1.07213  :   1   0.05774 :   1   0.00333 :   1  
0.00442 :   1   -100.77442:   1   0.20841 :   1   0.04343 :   1  
0.00559 :   1   -12.26034 :   1   0.39038 :   1   0.15239 :   1  
0.00611 :   1   -13.03158 :   1   0.44344 :   1   0.19664 :   1  
     v144          v145          v146          v147          v148    
```

```
 0.02   : 225   -0.02  : 173   -0.02  : 207   192    :  82   204    :  78
 0      : 132   0      : 171   -0.03  : 154   193    :  67   205    :  68
 0.03   : 112   -0.03  :  85   0      : 138   191    :  55   203    :  59
 0.06   :  92   0.02   :  82   -0.05  : 111   194    :  44   202    :  52
 0.08   :  89   0.03   :  58   -0.07  :  69   190    :  36   206    :  45
 -0.02  :  76   -0.05  :  42   0.02   :  69   195    :  26   207    :  33
      V149           V150           V151           V152
 -214   : 119   -10    :  38   655    :  51   472    :  37
 -213   :  98   -12    :  37   656    :  47   471    :  33
 -215   :  98   -11    :  35   653    :  44   469    :  32
 -216   :  70   -13    :  34   751    :  43   470    :  30
 -191   :  53   -9     :  30   654    :  41   468    :  27
 -188   :  52   -14    :  28   749    :  40   467    :  25
 [ reached getOption("max.print") -- omitted 1 row ]
> dim(df)
[1] 4025  152
> # outlier definition
> # x > Q3+1.5*IQR - positive side outlier
> # x < Q1-1.5*IQR - negative or lower side outlier
> #par(mfrow=c(2,3))
> #(boxplot(df$V24)$out);(boxplot(df$V25)$out);(boxplot(df$V26)$out);(boxplot
(df$V27)$out);(boxplot(df$V28)$out)
> #(boxplot(df$V29)$out);(boxplot(df$V30)$out);(boxplot(df$V31)$out);(boxplot
(df$V32)$out)
>
>
> apply(df,2,range)
     V1          V2                     V3
[1,] "adelmo"    "1322489729"           "100232"
[2,] "user_name" "raw_timestamp_part_1" "raw_timestamp_part_2"
     V4                V5           V6           V7          V8
[1,] "2/12/2011 13:35" "new_window" "1"          "-0.03"     "-27.8"
[2,] "cvtd_timestamp"  "yes"        "num_window" "roll_belt" "pitch_belt"
     V9         V10                V11                   V12
[1,] "-0.02"    "0"                "-0.076054"           "-0.06016"
[2,] "yaw_belt" "total_accel_belt" "kurtosis_roll_belt" "kurtosis_picth_belt"
     V13                  V14                  V15             V16
[1,] "-0.011683"          "-0.045472"          "-0.2"          "10"
[2,] "skewness_roll_belt" "skewness_roll_belt" "max_roll_belt" "max_picth_bel
t"
     V17            V18            V19              V20
[1,] "-0.1"         "-1"           "0"              "-0.1"
[2,] "max_yaw_belt" "min_roll_belt" "min_pitch_belt" "min_yaw_belt"
     V21                  V22                   V23
[1,] "0"                  "0"                   "0"
[2,] "amplitude_roll_belt" "amplitude_pitch_belt" "amplitude_yaw_belt"
     V24                   V25             V26                V27
[1,] "0"                   "-0.7"          "0"                "0"
[2,] "var_total_accel_belt" "avg_roll_belt" "stddev_roll_belt" "var_roll_belt
"
     V28              V29                  V30              V31
[1,] "-30.6"          "0"                  "0"              "-0.6"
[2,] "avg_pitch_belt" "stddev_pitch_belt" "var_pitch_belt" "avg_yaw_belt"
     V32               V33            V34            V35
[1,] "0"               "0"            "-0.02"        "-0.02"
[2,] "stddev_yaw_belt" "var_yaw_belt" "gyros_belt_x" "gyros_belt_y"
     V36            V37            V38            V39
```

```
[1,] "-0.02"        "-1"           "-1"           "-1"
[2,] "gyros_belt_z" "accel_belt_x" "accel_belt_y" "accel_belt_z"
     V40            V41            V42            V43        V44
[1,] "-1"           "428"          "-112"         "-0.02"    "-0.01"
[2,] "magnet_belt_x" "magnet_belt_y" "magnet_belt_z" "roll_arm" "pitch_arm"
     V45        V46              V47              V48
[1,] "-0.18"   "1"               "0"              "-128"
[2,] "yaw_arm" "total_accel_arm" "var_accel_arm" "avg_roll_arm"
     V49               V50            V51              V52
[1,] "0"               "0"            "-10.1695"       "0"
[2,] "stddev_roll_arm" "var_roll_arm" "avg_pitch_arm" "stddev_pitch_arm"
     V53             V54           V55              V56           V57
[1,] "0"             "-0.0558"     "0"              "0"           "-0.02"
[2,] "var_pitch_arm" "avg_yaw_arm" "stddev_yaw_arm" "var_yaw_arm" "gyros_arm_
x"
     V58           V59           V60           V61           V62
[1,] "-0.02"       "-0.02"       "-1"          "-1"          "-1"
[2,] "gyros_arm_y" "gyros_arm_z" "accel_arm_x" "accel_arm_y" "accel_arm_z"
     V63            V64            V65            V66
[1,] "-1"           "-1"           "-1"           "-0.11926"
[2,] "magnet_arm_x" "magnet_arm_y" "magnet_arm_z" "kurtosis_roll_arm"
     V67                  V68                V69
[1,] "-0.10176"           "-0.06791"         "-0.00696"
[2,] "kurtosis_picth_arm" "kurtosis_yaw_arm" "skewness_roll_arm"
     V70                  V71                V72            V73
[1,] "-0.01247"           "-0.0046"          "-0.7"         "-1.9"
[2,] "skewness_pitch_arm" "skewness_yaw_arm" "max_roll_arm" "max_picth_arm"
     V74           V75            V76             V77
[1,] "19"          "-10"          "-10.9"         "1"
[2,] "max_yaw_arm" "min_roll_arm" "min_pitch_arm" "min_yaw_arm"
     V78                  V79                   V80
[1,] "0"                  "0"                   "0"
[2,] "amplitude_roll_arm" "amplitude_pitch_arm" "amplitude_yaw_arm"
     V81              V82              V83            V84
[1,] "-0.949086746"  "-0.738381151"   "-10.50214329" "-0.0292"
[2,] "roll_dumbbell" "pitch_dumbbell" "yaw_dumbbell" "kurtosis_roll_dumbbell"
     V85                        V86
[1,] "-0.0122"                  "-0.0369"
[2,] "kurtosis_picth_dumbbell" "skewness_roll_dumbbell"
     V87                        V88                   V89
[1,] "-0.0084"                  "-18"                 "-17.6"
[2,] "skewness_pitch_dumbbell" "max_roll_dumbbell" "max_picth_dumbbell"
     V90                 V91                   V92
[1,] "-0.1"              "-1.2"                "-113.8"
[2,] "max_yaw_dumbbell" "min_roll_dumbbell" "min_pitch_dumbbell"
     V93                V94                       V95
[1,] "-0.1"             "0"                       "0"
[2,] "min_yaw_dumbbell" "amplitude_roll_dumbbell" "amplitude_pitch_dumbbell"
     V96                      V97                    V98
[1,] "0"                      "1"                    "0"
[2,] "amplitude_yaw_dumbbell" "total_accel_dumbbell" "var_accel_dumbbell"
     V99                 V100                    V101
[1,] "-10.4132"          "0"                     "0"
[2,] "avg_roll_dumbbell" "stddev_roll_dumbbell" "var_roll_dumbbell"
     V102                 V103                    V104
[1,] "-0.6827"            "0"                     "0"
[2,] "avg_pitch_dumbbell" "stddev_pitch_dumbbell" "var_pitch_dumbbell"
```

```
     V105               V106                   V107
[1,] "-101.7805"        "0"                    "0"
[2,] "avg_yaw_dumbbell" "stddev_yaw_dumbbell" "var_yaw_dumbbell"
     V108               V109                V110
[1,] "-0.02"            "-0.02"             "-0.02"
[2,] "gyros_dumbbell_x" "gyros_dumbbell_y" "gyros_dumbbell_z"
     V111               V112                V113
[1,] "-1"               "-1"                "-1"
[2,] "accel_dumbbell_x" "accel_dumbbell_y" "accel_dumbbell_z"
     V114                V115                 V116                 V117
[1,] "-1"                 "-459"               "-1"                 "-111"
[2,] "magnet_dumbbell_x" "magnet_dumbbell_y" "magnet_dumbbell_z" "roll_forear
m"
     V118            V119          V120
[1,] "-0.05"         "-100"        "-0.0699"
[2,] "pitch_forearm" "yaw_forearm" "kurtosis_roll_forearm"
     V121                     V122                     V123
[1,] "-0.0259"                "-0.009"                 "-0.0428"
[2,] "kurtosis_picth_forearm" "skewness_roll_forearm" "skewness_pitch_forearm
"
     V124               V125                V126
[1,] "-0.2"             "-143"              "-0.1"
[2,] "max_roll_forearm" "max_picth_forearm" "max_yaw_forearm"
     V127               V128                V129
[1,] "-0.9"             "-144"              "-0.1"
[2,] "min_roll_forearm" "min_pitch_forearm" "min_yaw_forearm"
     V130                     V131                      V132
[1,] "0"                      "0"                       "0"
[2,] "amplitude_roll_forearm" "amplitude_pitch_forearm" "amplitude_yaw_forear
m"
     V133                  V134                V135
[1,] "10"                  "0"                 "-136.38298"
[2,] "total_accel_forearm" "var_accel_forearm" "avg_roll_forearm"
     V136                  V137               V138
[1,] "0"                   "0"                "-0.00125"
[2,] "stddev_roll_forearm" "var_roll_forearm" "avg_pitch_forearm"
     V139                   V140                V141
[1,] "0"                    "0"                 "-1.07213"
[2,] "stddev_pitch_forearm" "var_pitch_forearm" "avg_yaw_forearm"
     V142                 V143              V144              V145
[1,] "0"                  "0"               "-0.02"           "-0.02"
[2,] "stddev_yaw_forearm" "var_yaw_forearm" "gyros_forearm_x" "gyros_forearm_
y"
     V146              V147              V148              V149
[1,] "-0.02"           "-1"              "-10"             "-1"
[2,] "gyros_forearm_z" "accel_forearm_x" "accel_forearm_y" "accel_forearm_z"
     V150               V151               V152
[1,] "-1"               "-1"               "-100"
[2,] "magnet_forearm_x" "magnet_forearm_y" "magnet_forearm_z"
>
> apply(df,2,summary)
       V1          V2          V3          V4          V5          V6
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class  "character" "character" "character" "character" "character" "character
"
Mode   "character" "character" "character" "character" "character" "character
"
```

```
       V7          V8          V9          V10         V11         V12
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class  "character" "character" "character" "character" "character" "character
"
Mode   "character" "character" "character" "character" "character" "character
"
       V13         V14         V15         V16         V17         V18
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class  "character" "character" "character" "character" "character" "character
"
Mode   "character" "character" "character" "character" "character" "character
"
       V19         V20         V21         V22         V23         V24
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class  "character" "character" "character" "character" "character" "character
"
Mode   "character" "character" "character" "character" "character" "character
"
       V25         V26         V27         V28         V29         V30
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class  "character" "character" "character" "character" "character" "character
"
Mode   "character" "character" "character" "character" "character" "character
"
       V31         V32         V33         V34         V35         V36
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class  "character" "character" "character" "character" "character" "character
"
Mode   "character" "character" "character" "character" "character" "character
"
       V37         V38         V39         V40         V41         V42
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class  "character" "character" "character" "character" "character" "character
"
Mode   "character" "character" "character" "character" "character" "character
"
       V43         V44         V45         V46         V47         V48
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class  "character" "character" "character" "character" "character" "character
"
Mode   "character" "character" "character" "character" "character" "character
"
       V49         V50         V51         V52         V53         V54
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class  "character" "character" "character" "character" "character" "character
"
Mode   "character" "character" "character" "character" "character" "character
"
       V55         V56         V57         V58         V59         V60
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class  "character" "character" "character" "character" "character" "character
"
Mode   "character" "character" "character" "character" "character" "character
"
       V61         V62         V63         V64         V65         V66
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
```

```
Class   "character" "character" "character" "character" "character" "character
"
Mode    "character" "character" "character" "character" "character" "character
"
        V67         V68         V69         V70         V71         V72
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class   "character" "character" "character" "character" "character" "character
"
Mode    "character" "character" "character" "character" "character" "character
"
        V73         V74         V75         V76         V77         V78
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class   "character" "character" "character" "character" "character" "character
"
Mode    "character" "character" "character" "character" "character" "character
"
        V79         V80         V81         V82         V83         V84
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class   "character" "character" "character" "character" "character" "character
"
Mode    "character" "character" "character" "character" "character" "character
"
        V85         V86         V87         V88         V89         V90
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class   "character" "character" "character" "character" "character" "character
"
Mode    "character" "character" "character" "character" "character" "character
"
        V91         V92         V93         V94         V95         V96
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class   "character" "character" "character" "character" "character" "character
"
Mode    "character" "character" "character" "character" "character" "character
"
        V97         V98         V99         V100        V101        V102
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class   "character" "character" "character" "character" "character" "character
"
Mode    "character" "character" "character" "character" "character" "character
"
        V103        V104        V105        V106        V107        V108
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class   "character" "character" "character" "character" "character" "character
"
Mode    "character" "character" "character" "character" "character" "character
"
        V109        V110        V111        V112        V113        V114
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class   "character" "character" "character" "character" "character" "character
"
Mode    "character" "character" "character" "character" "character" "character
"
        V115        V116        V117        V118        V119        V120
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class   "character" "character" "character" "character" "character" "character
"
```

```
Mode   "character" "character" "character" "character" "character" "character
"
       V121        V122        V123        V124        V125        V126
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class  "character" "character" "character" "character" "character" "character
"
Mode   "character" "character" "character" "character" "character" "character
"
       V127        V128        V129        V130        V131        V132
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class  "character" "character" "character" "character" "character" "character
"
Mode   "character" "character" "character" "character" "character" "character
"
       V133        V134        V135        V136        V137        V138
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class  "character" "character" "character" "character" "character" "character
"
Mode   "character" "character" "character" "character" "character" "character
"
       V139        V140        V141        V142        V143        V144
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class  "character" "character" "character" "character" "character" "character
"
Mode   "character" "character" "character" "character" "character" "character
"
       V145        V146        V147        V148        V149        V150
Length "4025"      "4025"      "4025"      "4025"      "4025"      "4025"
Class  "character" "character" "character" "character" "character" "character
"
Mode   "character" "character" "character" "character" "character" "character
"
       V151        V152
Length "4025"      "4025"
Class  "character" "character"
Mode   "character" "character"
> df[] <- lapply(df, function(x) as.numeric(as.character(x)))
There were 50 or more warnings (use warnings() to see the first 50)
> # KMeans - comes from Rcmdr library
> # Kmeans- from amap library
> # kmeans- from stats library
>
> # steps in k-means clustering
> #1- preprocessing the data (impute missing values, remove outliers, feature
trasnformation)
> #2- scaling or standardization of data set
> #3- decide the number of clusters (value of K)
> #4- iterate over the samples to create clusters
> #5- decide the distance measure
> #6- calculate the group accuracy
>
> # scaling of data
> df_train1 <- scale(df_train)
Error in colMeans(x, na.rm = TRUE) : 'x' must be numeric
>
> head(df_train1)
             V2          V3           V4           V5          V6          V7
```

```
21   1.27299870 -0.6409610 -0.37105619 -1.03850116  1.9078725  1.0812457
111 -1.89747567  1.2570894 -2.16243759  0.02067199  0.5274917  1.3655732
108 -0.36101501 -0.5330150 -0.37105619  0.92853470 -1.1434956 -1.4777024
110 -1.71456368 -0.8928350  1.26455292  0.17198244 -0.4169794  0.6705503
150  0.07797375  1.4010174 -0.05951160  0.62591379  0.9634014 -1.4303145
177  0.18772094  0.2226069 -0.02056852  0.17198244  1.4719628 -1.0512111
            v8         v9        v10        v11         v12        v13
21   1.1479385 -1.0099319  0.8615852  0.1924509  0.60908324  1.5417428
111  0.5540868 -1.0099319  3.4201656 -1.0084453 -0.83737711  0.2967363
108 -0.2712665  0.9086010  0.0490630 -0.8337695 -0.28431874 -0.2427665
110  0.8963064 -0.6102376  1.5358057 -1.1176177  0.05602487  0.9192395
150 -0.6436820 -0.2105432 -0.7980346  1.8300367 -1.60315024 -1.7506077
177 -1.3583172  1.3082953 -0.2448280  1.7863678 -1.47552138 -1.3494389
            v14
21   0.07490999
111 -0.62907105
108 -0.86803709
110 -0.24801710
150 -0.66782230
177  0.26866624
>
> class(df_train1)
[1] "matrix"
> # screeplot approach to decide the number of clusters
> km = kmeans(df_train1,1)
> km$withinss
[1] 1833
> km$tot.withinss
[1] 1833
>
> km = kmeans(df_train1,2)
> km$withinss
[1] 685.8406 600.9150
> km$tot.withinss
[1] 1286.756
>
> km = kmeans(df_train1,3)
> km$withinss
[1] 346.0231 298.1925 336.7207
> km$tot.withinss
[1] 980.9363
>
> km = kmeans(df_train1,4)
> km$withinss
[1] 275.1233 210.9994 209.2629 216.7761
> km$tot.withinss
[1] 912.1616
>
> km = kmeans(df_train1,5)
> km$withinss
[1] 176.7833 119.2848 138.2213 309.2322 112.0491
> km$tot.withinss
[1] 855.5707
>
> km = kmeans(df_train1,6)
> km$withinss
[1] 172.84625 114.91540 137.94960 143.94848  94.49965 149.72160
```

```
> km$tot.withinss
[1] 813.881
>
> km = kmeans(df_train1,7)
> km$withinss
[1]  74.27123  78.06204  79.70444  58.01375  81.96519 298.19248  96.74012
> km$tot.withinss
[1] 766.9492
>
> km = kmeans(df_train1,8)
> km$withinss
[1]  33.42481 196.45068 124.67682  58.35021 126.40989  34.71128  77.80059
[8]  70.22846
> km$tot.withinss
[1] 722.0527
>
> km = kmeans(df_train1,9)
> km$withinss
[1]  44.271295  78.062037 110.262048 107.569538   4.597058  47.484979 138.221
277
[8]  37.955215 104.487311
> km$tot.withinss
[1] 672.9108
>
> km = kmeans(df_train1,10)
> km$withinss
 [1]  50.42743  57.36509  74.27123  52.76306  39.50401  81.96519 233.91216
 [8]  23.12942  30.22351  33.42481
> km$tot.withinss
[1] 676.9859
> dev.off()
null device
          1
> sumsq=NULL
> for (i in 1:10)
+   sumsq[i] = sum(kmeans(df_train,centers=i,
+                           iter.max = 1000,
+                           nstart=i,
+                           algorithm='Forgy')$withinss)
Error in do_one(nmeth) : NA/NaN/Inf in foreign function call (arg 1)
In addition: Warning message:
In storage.mode(x) <- "double" : NAs introduced by coercion
> plot(1:10,sumsq,type='b', main='Screeplot showing within group sum of squar
es')
> km = kmeans(df_train1,3)
> km$withinss
[1] 346.0231 336.7207 298.1925
> km$tot.withinss
[1] 980.9363
>
> class(km$cluster)
[1] "integer"
>
> summary(km)
             Length Class  Mode
cluster      142    -none- numeric
centers       39    -none- numeric
```

```
totss          1    -none- numeric
withinss       3    -none- numeric
tot.withinss   1    -none- numeric
betweenss      1    -none- numeric
size           3    -none- numeric
iter           1    -none- numeric
ifault         1    -none- numeric
>
> km$centers
          V2         V3          V4         V5          V6          V7
1 -1.0688290 -0.3745716 -0.45159755  0.2834020 -0.69107412  0.03476227
2  0.7873041 -0.3366416  0.30126292 -0.6685043  0.63578373  0.81778071
3  0.1178079  0.7627367  0.08674128  0.5102454 -0.07309503 -0.99715372
          V8          V9         V10         V11        V12        V13
1  0.1176195 -0.02704718  0.05220622 -0.94949223  0.4882226  0.4014300
2  0.9257427 -0.57705540  0.59737684  0.08451446  0.5280108  0.7409713
3 -1.2053248  0.70608916 -0.75462324  0.82885315 -1.0992526 -1.2652089
         V14
1 -0.8160752
2  1.0661743
3 -0.4577762
>
> as.numeric(km$cluster)
  [1] 2 1 1 1 3 3 2 2 1 1 1 1 2 3 2 3 3 2 2 2 2 3 2 2 2 1 1 3 3 3 1 2 2 2 2 1
2
 [38] 3 3 1 1 1 3 3 3 1 1 1 2 1 2 3 1 3 2 3 1 1 3 1 1 2 3 3 2 3 2 2 2 1 2 2 1
1
 [75] 1 2 1 2 2 3 1 2 3 2 1 3 2 3 2 3 3 2 2 1 2 3 3 2 1 2 1 2 2 2 1 2 1 3 2 1
2
[112] 3 1 3 1 2 3 2 1 3 3 2 2 1 3 1 1 2 3 3 3 3 3 2 2 1 3 2 3 3 3 1
>
> length(km$cluster)
[1] 142
>
> dim(df_train)
[1] 3220  151
>
> class(df_train)
[1] "data.frame"
>
> df_train$cl <- km$cluster
Error in `$<-.data.frame`(`*tmp*`, cl, value = c(`21` = 2L, `111` = 1L,  :
  replacement has 142 rows, data has 3220
>
> head(df_train)
             V2     V3               V4 V5 V6   V7   V8    V9 V10      V11
458  1323084232 976322 5/12/2011 11:23 no 12 1.52  8.1 -94.3   3 -1.03566
2505 1323094981 340386 5/12/2011 14:23 no 57  124 26.1  -2.1  19 -1.03566
2452 1323094980 324269 5/12/2011 14:23 no 56  123 26.4 -3.24  19 -1.03566
2508 1323094981 352685 5/12/2011 14:23 no 57  124 26.1  -2.1  20 -1.03566
3462 1323095007 104321 5/12/2011 14:23 no 79  123 26.7 -3.73  20 -1.03566
2575 1323094982 972320 5/12/2011 14:23 no 58  123 26.3 -3.51  20 -1.03566
          V12      V13      V14  V15 V16 V17   V18 V19 V20   V21 V22 V23 V24
458  -0.39133 0.005406 0.045115 -4.1  20  -1 -7.25  18  -1 1.345   2   0 0.3
2505 -0.39133 0.005406 0.045115 -4.1  20  -1 -7.25  18  -1 1.345   2   0 0.3
2452 -0.39133 0.005406 0.045115 -4.1  20  -1 -7.25  18  -1 1.345   2   0 0.3
2508 -0.39133 0.005406 0.045115 -4.1  20  -1 -7.25  18  -1 1.345   2   0 0.3
```

```
3462 -0.39133 0.005406 0.045115 -4.1  20  -1 -7.25  18  -1 1.345   2   0 0.3
2575 -0.39133 0.005406 0.045115 -4.1  20  -1 -7.25  18  -1 1.345   2   0 0.3
       V25 V26  V27   V28  V29 V30   V31 V32  V33   V34   V35   V36 V37 V38
458  121.9 0.6 0.35 25.75 0.35 0.1 -4.95 0.4 0.17  0.02     0 -0.02 -21   3
2505 121.9 0.6 0.35 25.75 0.35 0.1 -4.95 0.4 0.17 -0.43 -0.05 -0.44 -42  69
2452 121.9 0.6 0.35 25.75 0.35 0.1 -4.95 0.4 0.17 -0.47 -0.03 -0.46 -33  68
2508 121.9 0.6 0.35 25.75 0.35 0.1 -4.95 0.4 0.17 -0.43 -0.03 -0.46 -44  71
3462 121.9 0.6 0.35 25.75 0.35 0.1 -4.95 0.4 0.17 -0.59     0 -0.56 -45  70
2575 121.9 0.6 0.35 25.75 0.35 0.1 -4.95 0.4 0.17 -0.37 -0.03 -0.41 -43  70
      V39 V40 V41  V42  V43   V44  V45 V46     V47      V48     V49      V50
458    22  -9 599 -317 -129  20.7 -161  34 65.0977 76.22175 16.1039 259.3599
2505 -171   1 579 -366  118 -23.4 90.1  17 65.0977 76.22175 16.1039 259.3599
2452 -169   1 582 -365  139 -24.7 62.3  24 65.0977 76.22175 16.1039 259.3599
2508 -172   2 579 -378  120 -22.1 79.6  18 65.0977 76.22175 16.1039 259.3599
3462 -177  -9 583 -372  102 -27.8  110  15 65.0977 76.22175 16.1039 259.3599
2575 -174   1 587 -377  146 -1.77 72.3  23 65.0977 76.22175 16.1039 259.3599
         V51      V52      V53     V54     V55      V56   V57   V58   V59  V
60
458  -10.1695 10.66725 113.7978 19.0615 35.8809 1287.463 -0.02     0 -0.02 -2
89
2505 -10.1695 10.66725 113.7978 19.0615 35.8809 1287.463 -2.28  0.88 -0.38
76
2452 -10.1695 10.66725 113.7978 19.0615 35.8809 1287.463  1.67 -1.01  0.11  1
76
2508 -10.1695 10.66725 113.7978 19.0615 35.8809 1287.463 -2.52   0.9 -0.43  1
02
3462 -10.1695 10.66725 113.7978 19.0615 35.8809 1287.463 -2.57  0.98  0.02
24
2575 -10.1695 10.66725 113.7978 19.0615 35.8809 1287.463  1.33 -1.01 -0.18  1
61
     V61  V62  V63 V64 V65      V66      V67      V68     V69      V70      V
71
458  109 -123 -373 335 511 -1.18224 -0.96912 -0.86977 0.12353 -0.10319 0.0597
65
2505  46  142    8 349 575 -1.18224 -0.96912 -0.86977 0.12353 -0.10319 0.0597
65
2452  61  140  305 341 490 -1.18224 -0.96912 -0.86977 0.12353 -0.10319 0.0597
65
2508  37  144  139 310 578 -1.18224 -0.96912 -0.86977 0.12353 -0.10319 0.0597
65
3462  44  137 -165 378 538 -1.18224 -0.96912 -0.86977 0.12353 -0.10319 0.0597
65
2575  67  138  151 391 512 -1.18224 -0.96912 -0.86977 0.12353 -0.10319 0.0597
65
      V72   V73 V74   V75   V76 V77    V78   V79 V80          V81          V8
2
458  8.45 77.25  38 -33.6 -58.6  10 36.945 121.5  27  13.02493063 -70.3175607
5
2505 8.45 77.25  38 -33.6 -58.6  10 36.945 121.5  27 -26.97447709  15.2872136
3
2452 8.45 77.25  38 -33.6 -58.6  10 36.945 121.5  27 -5.418908377  26.3001823
8
2508 8.45 77.25  38 -33.6 -58.6  10 36.945 121.5  27  1.186057995  38.9287866
3
3462 8.45 77.25  38 -33.6 -58.6  10 36.945 121.5  27 -24.95711462  16.9207471
4
```

```
2575 8.45 77.25  38 -33.6 -58.6  10 36.945 121.5  27 -6.488101883  30.7383386
6
              v83      v84     v85    v86    v87   v88 v89  v90    v91  v92
458  -85.05658246 -0.09595 -0.4422 0.0819 -0.216 41.85 133 -0.1 -26.75 20.2
2505  125.8836028 -0.09595 -0.4422 0.0819 -0.216 41.85 133 -0.1 -26.75 20.2
2452  130.2009178 -0.09595 -0.4422 0.0819 -0.216 41.85 133 -0.1 -26.75 20.2
2508  118.1308197 -0.09595 -0.4422 0.0819 -0.216 41.85 133 -0.1 -26.75 20.2
3462  126.7219718 -0.09595 -0.4422 0.0819 -0.216 41.85 133 -0.1 -26.75 20.2
2575  125.6198141 -0.09595 -0.4422 0.0819 -0.216 41.85 133 -0.1 -26.75 20.2
      v93   v94   v95 v96 v97      v98      v99   v100    v101    v102    v103
458  -0.1 55.71 54.74   0  37 2.41635 -5.11805 17.058 291.001 13.9312 14.1062
2505 -0.1 55.71 54.74   0   8 2.41635 -5.11805 17.058 291.001 13.9312 14.1062
2452 -0.1 55.71 54.74   0   9 2.41635 -5.11805 17.058 291.001 13.9312 14.1062
2508 -0.1 55.71 54.74   0   9 2.41635 -5.11805 17.058 291.001 13.9312 14.1062
3462 -0.1 55.71 54.74   0   9 2.41635 -5.11805 17.058 291.001 13.9312 14.1062
2575 -0.1 55.71 54.74   0   9 2.41635 -5.11805 17.058 291.001 13.9312 14.1062
         v104    v105    v106     v107 v108  v109  v110 v111 v112 v113 v114
458  199.0775 64.7063 13.5747 184.5578    0 -0.02     0 -234   47 -272 -553
2505 199.0775 64.7063 13.5747 184.5578 0.55 -0.05 -0.23   12  -21   75  534
2452 199.0775 64.7063 13.5747 184.5578 0.42 -0.08 -0.56   24   -5   89  513
2508 199.0775 64.7063 13.5747 184.5578 0.56 -0.13 -0.23   32    1   78  531
3462 199.0775 64.7063 13.5747 184.5578 0.32 -0.47  0.28   15  -22   85  513
2575 199.0775 64.7063 13.5747 184.5578 0.42 -0.08 -0.52   28   -6   88  524
     v115 v116 v117  v118 v119     v120     v121     v122    v123 v124 v125
458   300  -68   27 -63.8 -151 -1.09475 -0.97525 -0.05065 0.17285 49.6  168
2505 -517  -92  112  65.2  116 -1.09475 -0.97525 -0.05065 0.17285 49.6  168
2452 -534  -63   88  46.2 83.6 -1.09475 -0.97525 -0.05065 0.17285 49.6  168
2508 -518  -88  125  73.7  130 -1.09475 -0.97525 -0.05065 0.17285 49.6  168
3462 -532  -67  140  48.6  164 -1.09475 -0.97525 -0.05065 0.17285 49.6  168
2575 -522  -64 89.4  18.5 73.4 -1.09475 -0.97525 -0.05065 0.17285 49.6  168
     v126 v127   v128 v129 v130  v131 v132 v133    v134     v135     v136
458  -1.1 4.65 -168.5 -1.1 32.2 341.5    0   36 14.0772 27.85936 45.16342
2505 -1.1 4.65 -168.5 -1.1 32.2 341.5    0   35 14.0772 27.85936 45.16342
2452 -1.1 4.65 -168.5 -1.1 32.2 341.5    0   37 14.0772 27.85936 45.16342
2508 -1.1 4.65 -168.5 -1.1 32.2 341.5    0   32 14.0772 27.85936 45.16342
3462 -1.1 4.65 -168.5 -1.1 32.2 341.5    0   27 14.0772 27.85936 45.16342
2575 -1.1 4.65 -168.5 -1.1 32.2 341.5    0   33 14.0772 27.85936 45.16342
         v137     v138     v139     v140     v141     v142     v143  v144  v1
45
458  2749.163 25.35597 8.906695 79.33451 17.09505 74.27584 5541.956  0.03 -0.
05
2505 2749.163 25.35597 8.906695 79.33451 17.09505 74.27584 5541.956  0.47 -2.
47
2452 2749.163 25.35597 8.906695 79.33451 17.09505 74.27584 5541.956 -0.53  2.
81
2508 2749.163 25.35597 8.906695 79.33451 17.09505 74.27584 5541.956  0.16 -2.
39
3462 2749.163 25.35597 8.906695 79.33451 17.09505 74.27584 5541.956  0.51 -3.
53
2575 2749.163 25.35597 8.906695 79.33451 17.09505 74.27584 5541.956 -0.39   1
.4
      v146 v147 v148 v149 v150 v151 v152
458  -0.02  190  202 -213  -18  662  462
2505  -0.3  -88  283 -178 -654  499  491
2452  0.57   -6  335 -140 -556  713  680
2508 -0.44 -120  230 -176 -687  399  476
3462  -1.1 -118  174 -155 -707  214  558
```

```
2575  0.05  110  269 -147 -308  790  709
> # profiles of clusters
> aggregate(df_train[,1:8],list(df_train[,9]),mean)
   Group.1 V2 V3 V4 V5 V6 V7 V8 V9
1        0 NA NA NA NA NA NA NA NA
2        1 NA NA NA NA NA NA NA NA
3       10 NA NA NA NA NA NA NA NA
4       11 NA NA NA NA NA NA NA NA
5       12 NA NA NA NA NA NA NA NA
6       13 NA NA NA NA NA NA NA NA
7       14 NA NA NA NA NA NA NA NA
8       15 NA NA NA NA NA NA NA NA
9       16 NA NA NA NA NA NA NA NA
10      17 NA NA NA NA NA NA NA NA
11      18 NA NA NA NA NA NA NA NA
12      19 NA NA NA NA NA NA NA NA
13       2 NA NA NA NA NA NA NA NA
14      20 NA NA NA NA NA NA NA NA
15      21 NA NA NA NA NA NA NA NA
16      22 NA NA NA NA NA NA NA NA
17      23 NA NA NA NA NA NA NA NA
18      24 NA NA NA NA NA NA NA NA
19      25 NA NA NA NA NA NA NA NA
20      26 NA NA NA NA NA NA NA NA
21       3 NA NA NA NA NA NA NA NA
22       4 NA NA NA NA NA NA NA NA
23       5 NA NA NA NA NA NA NA NA
24       6 NA NA NA NA NA NA NA NA
25       7 NA NA NA NA NA NA NA NA
26       8 NA NA NA NA NA NA NA NA
27       9 NA NA NA NA NA NA NA NA
There were 50 or more warnings (use warnings() to see the first 50)
>
> table(df$V1)
< table of extent 0 >
>
> library(cluster)
>
> clusplot(df_train,df_train$cl,cex=0.9,color=T,shade=T, labels=4,lines=0)
Error in is.list(s.x.2d) :
  cannot use 'cor = TRUE' with a constant variable
> #HC clustering or Hierarchical Clustering
> # distance (euclidean, manhattan, cosine distance)
>
> # Divisive method (top down)
> # Agglomorative method (bottom up)
>
>
>
> df_train = df_train[,-9]
> head(df_train)
             V2     V3              V4 V5 V6   V7   V8    V9      V11      V1
2
458  1323084232 976322 5/12/2011 11:23 no 12 1.52  8.1 -94.3 -1.03566 -0.3913
3
2505 1323094981 340386 5/12/2011 14:23 no 57  124 26.1  -2.1 -1.03566 -0.3913
3
```

```
2452 1323094980 324269 5/12/2011 14:23 no 56  123 26.4 -3.24 -1.03566 -0.3913
3
2508 1323094981 352685 5/12/2011 14:23 no 57  124 26.1  -2.1 -1.03566 -0.3913
3
3462 1323095007 104321 5/12/2011 14:23 no 79  123 26.7 -3.73 -1.03566 -0.3913
3
2575 1323094982 972320 5/12/2011 14:23 no 58  123 26.3 -3.51 -1.03566 -0.3913
3
          V13      V14  V15 V16 V17   V18 V19 V20   V21 V22 V23 V24   V25 V26
458  0.005406 0.045115 -4.1  20  -1 -7.25  18  -1 1.345   2   0 0.3 121.9 0.6
2505 0.005406 0.045115 -4.1  20  -1 -7.25  18  -1 1.345   2   0 0.3 121.9 0.6
2452 0.005406 0.045115 -4.1  20  -1 -7.25  18  -1 1.345   2   0 0.3 121.9 0.6
2508 0.005406 0.045115 -4.1  20  -1 -7.25  18  -1 1.345   2   0 0.3 121.9 0.6
3462 0.005406 0.045115 -4.1  20  -1 -7.25  18  -1 1.345   2   0 0.3 121.9 0.6
2575 0.005406 0.045115 -4.1  20  -1 -7.25  18  -1 1.345   2   0 0.3 121.9 0.6
      V27   V28  V29 V30   V31 V32  V33   V34   V35   V36 V37 V38  V39 V40 V4
1
458  0.35 25.75 0.35 0.1 -4.95 0.4 0.17  0.02     0 -0.02 -21   3   22  -9 59
9
2505 0.35 25.75 0.35 0.1 -4.95 0.4 0.17 -0.43 -0.05 -0.44 -42  69 -171   1 57
9
2452 0.35 25.75 0.35 0.1 -4.95 0.4 0.17 -0.47 -0.03 -0.46 -33  68 -169   1 58
2
2508 0.35 25.75 0.35 0.1 -4.95 0.4 0.17 -0.43 -0.03 -0.46 -44  71 -172   2 57
9
3462 0.35 25.75 0.35 0.1 -4.95 0.4 0.17 -0.59     0 -0.56 -45  70 -177  -9 58
3
2575 0.35 25.75 0.35 0.1 -4.95 0.4 0.17 -0.37 -0.03 -0.41 -43  70 -174   1 58
7
      V42  V43   V44  V45 V46     V47      V48     V49      V50      V51
458  -317 -129  20.7 -161  34 65.0977 76.22175 16.1039 259.3599 -10.1695
2505 -366  118 -23.4 90.1  17 65.0977 76.22175 16.1039 259.3599 -10.1695
2452 -365  139 -24.7 62.3  24 65.0977 76.22175 16.1039 259.3599 -10.1695
2508 -378  120 -22.1 79.6  18 65.0977 76.22175 16.1039 259.3599 -10.1695
3462 -372  102 -27.8  110  15 65.0977 76.22175 16.1039 259.3599 -10.1695
2575 -377  146 -1.77 72.3  23 65.0977 76.22175 16.1039 259.3599 -10.1695
          V52      V53     V54     V55      V56   V57   V58   V59  V60 V61  V
62
458  10.66725 113.7978 19.0615 35.8809 1287.463 -0.02     0 -0.02 -289 109 -1
23
2505 10.66725 113.7978 19.0615 35.8809 1287.463 -2.28  0.88 -0.38   76  46  1
42
2452 10.66725 113.7978 19.0615 35.8809 1287.463  1.67 -1.01  0.11  176  61  1
40
2508 10.66725 113.7978 19.0615 35.8809 1287.463 -2.52   0.9 -0.43  102  37  1
44
3462 10.66725 113.7978 19.0615 35.8809 1287.463 -2.57  0.98  0.02   24  44  1
37
2575 10.66725 113.7978 19.0615 35.8809 1287.463  1.33 -1.01 -0.18  161  67  1
38
      V63 V64 V65      V66      V67      V68     V69      V70      V71  V72
458  -373 335 511 -1.18224 -0.96912 -0.86977 0.12353 -0.10319 0.059765 8.45
2505    8 349 575 -1.18224 -0.96912 -0.86977 0.12353 -0.10319 0.059765 8.45
2452  305 341 490 -1.18224 -0.96912 -0.86977 0.12353 -0.10319 0.059765 8.45
2508  139 310 578 -1.18224 -0.96912 -0.86977 0.12353 -0.10319 0.059765 8.45
3462 -165 378 538 -1.18224 -0.96912 -0.86977 0.12353 -0.10319 0.059765 8.45
2575  151 391 512 -1.18224 -0.96912 -0.86977 0.12353 -0.10319 0.059765 8.45
```

	V73	V74	V75	V76	V77	V78	V79	V80	V81	V82
458	77.25	38	-33.6	-58.6	10	36.945	121.5	27	13.02493063	-70.31756075
2505	77.25	38	-33.6	-58.6	10	36.945	121.5	27	-26.97447709	15.28721363
2452	77.25	38	-33.6	-58.6	10	36.945	121.5	27	-5.418908377	26.30018238
2508	77.25	38	-33.6	-58.6	10	36.945	121.5	27	1.186057995	38.92878663
3462	77.25	38	-33.6	-58.6	10	36.945	121.5	27	-24.95711462	16.92074714
2575	77.25	38	-33.6	-58.6	10	36.945	121.5	27	-6.488101883	30.73833866

	V83	V84	V85	V86	V87	V88	V89	V90	V91	V92
458	-85.05658246	-0.09595	-0.4422	0.0819	-0.216	41.85	133	-0.1	-26.75	20.2
2505	125.8836028	-0.09595	-0.4422	0.0819	-0.216	41.85	133	-0.1	-26.75	20.2
2452	130.2009178	-0.09595	-0.4422	0.0819	-0.216	41.85	133	-0.1	-26.75	20.2
2508	118.1308197	-0.09595	-0.4422	0.0819	-0.216	41.85	133	-0.1	-26.75	20.2
3462	126.7219718	-0.09595	-0.4422	0.0819	-0.216	41.85	133	-0.1	-26.75	20.2
2575	125.6198141	-0.09595	-0.4422	0.0819	-0.216	41.85	133	-0.1	-26.75	20.2

	V93	V94	V95	V96	V97	V98	V99	V100	V101	V102	V103
458	-0.1	55.71	54.74	0	37	2.41635	-5.11805	17.058	291.001	13.9312	14.1062
2505	-0.1	55.71	54.74	0	8	2.41635	-5.11805	17.058	291.001	13.9312	14.1062
2452	-0.1	55.71	54.74	0	9	2.41635	-5.11805	17.058	291.001	13.9312	14.1062
2508	-0.1	55.71	54.74	0	9	2.41635	-5.11805	17.058	291.001	13.9312	14.1062
3462	-0.1	55.71	54.74	0	9	2.41635	-5.11805	17.058	291.001	13.9312	14.1062
2575	-0.1	55.71	54.74	0	9	2.41635	-5.11805	17.058	291.001	13.9312	14.1062

	V104	V105	V106	V107	V108	V109	V110	V111	V112	V113	V114
458	199.0775	64.7063	13.5747	184.5578	0	-0.02	0	-234	47	-272	-553
2505	199.0775	64.7063	13.5747	184.5578	0.55	-0.05	-0.23	12	-21	75	534
2452	199.0775	64.7063	13.5747	184.5578	0.42	-0.08	-0.56	24	-5	89	513
2508	199.0775	64.7063	13.5747	184.5578	0.56	-0.13	-0.23	32	1	78	531
3462	199.0775	64.7063	13.5747	184.5578	0.32	-0.47	0.28	15	-22	85	513
2575	199.0775	64.7063	13.5747	184.5578	0.42	-0.08	-0.52	28	-6	88	524

	V115	V116	V117	V118	V119	V120	V121	V122	V123	V124	V125
458	300	-68	27	-63.8	-151	-1.09475	-0.97525	-0.05065	0.17285	49.6	168
2505	-517	-92	112	65.2	116	-1.09475	-0.97525	-0.05065	0.17285	49.6	168
2452	-534	-63	88	46.2	83.6	-1.09475	-0.97525	-0.05065	0.17285	49.6	168
2508	-518	-88	125	73.7	130	-1.09475	-0.97525	-0.05065	0.17285	49.6	168
3462	-532	-67	140	48.6	164	-1.09475	-0.97525	-0.05065	0.17285	49.6	168
2575	-522	-64	89.4	18.5	73.4	-1.09475	-0.97525	-0.05065	0.17285	49.6	168

	V126	V127	V128	V129	V130	V131	V132	V133	V134	V135	V136
458	-1.1	4.65	-168.5	-1.1	32.2	341.5	0	36	14.0772	27.85936	45.16342
2505	-1.1	4.65	-168.5	-1.1	32.2	341.5	0	35	14.0772	27.85936	45.16342
2452	-1.1	4.65	-168.5	-1.1	32.2	341.5	0	37	14.0772	27.85936	45.16342
2508	-1.1	4.65	-168.5	-1.1	32.2	341.5	0	32	14.0772	27.85936	45.16342
3462	-1.1	4.65	-168.5	-1.1	32.2	341.5	0	27	14.0772	27.85936	45.16342
2575	-1.1	4.65	-168.5	-1.1	32.2	341.5	0	33	14.0772	27.85936	45.16342

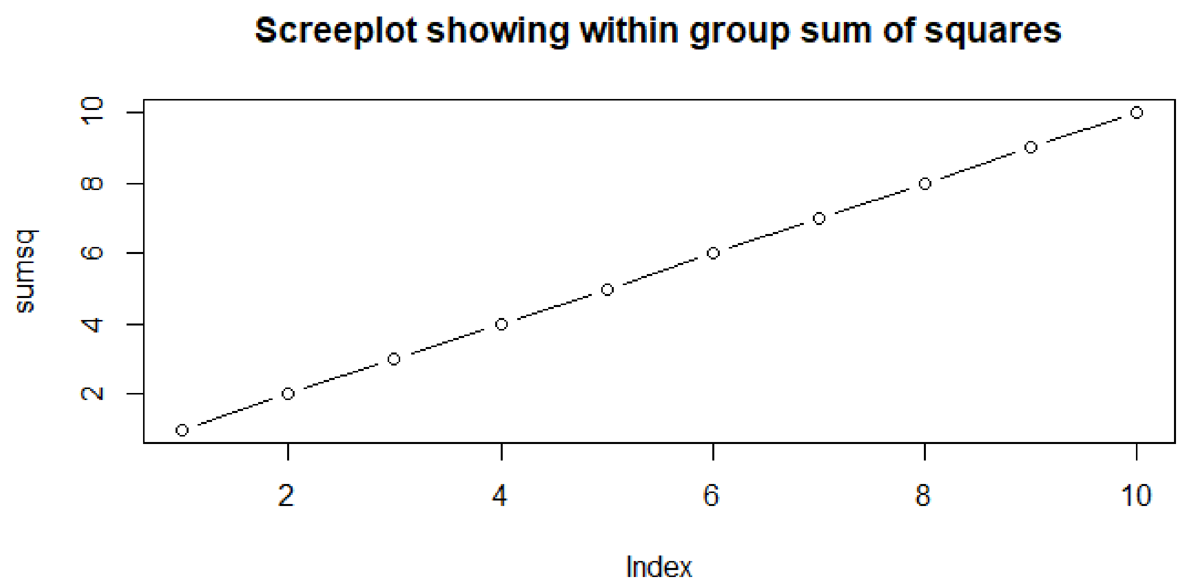
	V137	V138	V139	V140	V141	V142	V143	V144	V145
458	2749.163	25.35597	8.906695	79.33451	17.09505	74.27584	5541.956	0.03	-0.05
2505	2749.163	25.35597	8.906695	79.33451	17.09505	74.27584	5541.956	0.47	-2.47
2452	2749.163	25.35597	8.906695	79.33451	17.09505	74.27584	5541.956	-0.53	2.81
2508	2749.163	25.35597	8.906695	79.33451	17.09505	74.27584	5541.956	0.16	-2.39
3462	2749.163	25.35597	8.906695	79.33451	17.09505	74.27584	5541.956	0.51	-3.53
2575	2749.163	25.35597	8.906695	79.33451	17.09505	74.27584	5541.956	-0.39	1.4

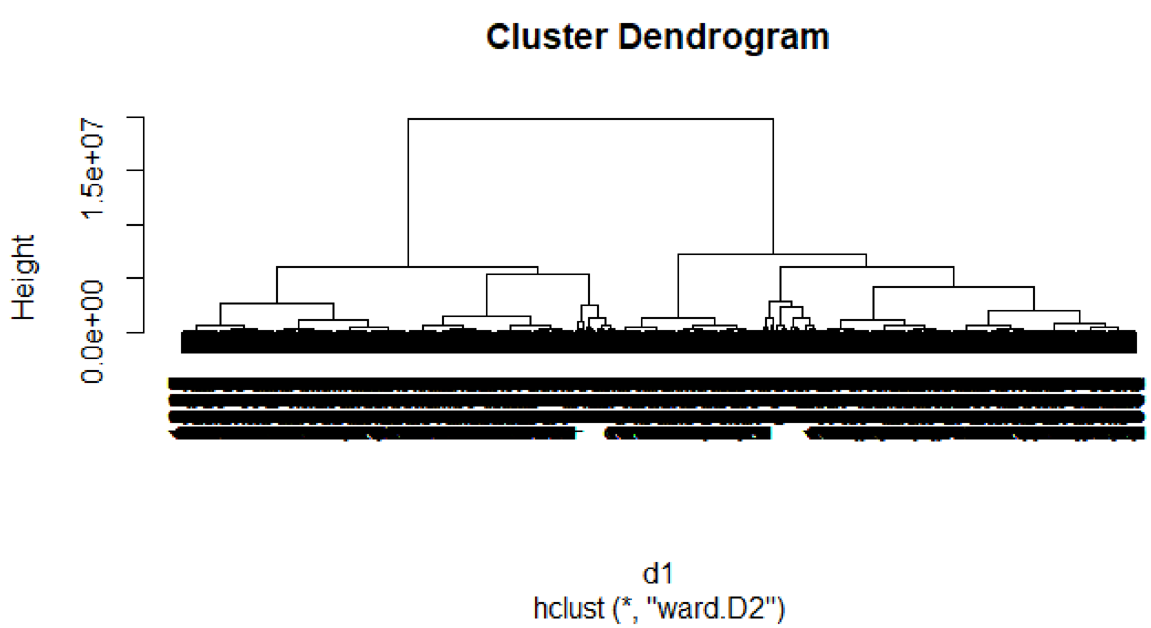
	V146	V147	V148	V149	V150	V151	V152

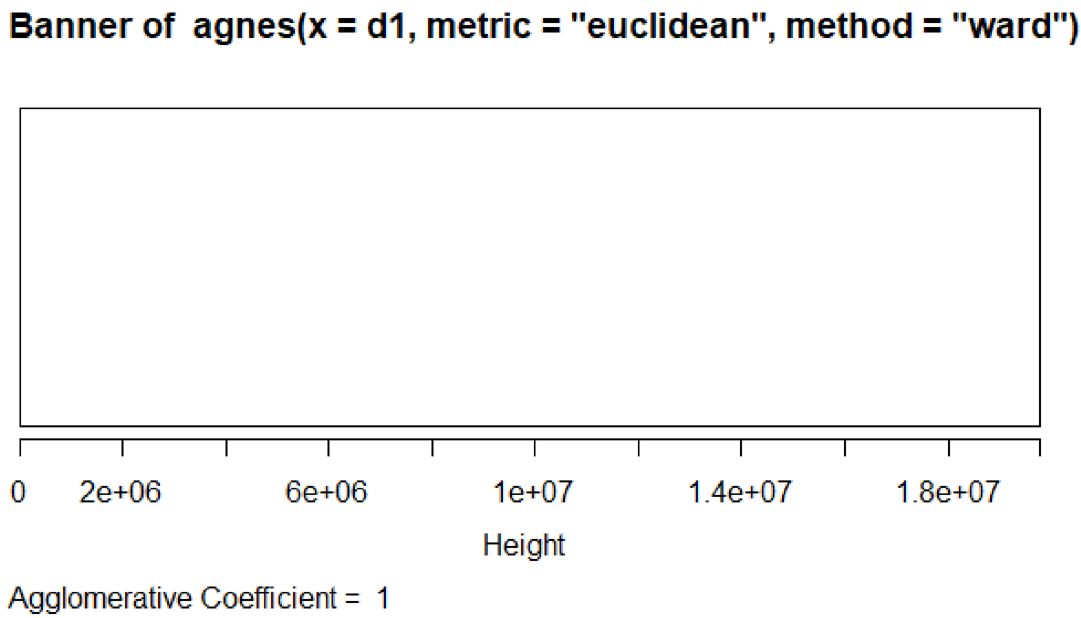
```
458  -0.02  190  202 -213  -18  662  462
2505  -0.3  -88  283 -178 -654  499  491
2452  0.57   -6  335 -140 -556  713  680
2508 -0.44 -120  230 -176 -687  399  476
3462  -1.1 -118  174 -155 -707  214  558
2575  0.05  110  269 -147 -308  790  709
>
> # compute the distance metrix
> d1 <- dist(df_train,method='euclidean')
Warning message:
In dist(df_train, method = "euclidean") : NAs introduced by coercion
> summary(d1)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
     16.5  169114.5  330259.8  365890.1  543588.0 1164944.1
> # HC
> fit <- hclust(d1,method = 'ward.D2')
> plot(fit)
>
> # single, double, average, ward, ward.D2
>
> # agglomorative method
> fit <- agnes(d1,metric='euclidean',method = 'ward')
> plot(fit)
Hit <Return> to see next plot:
Hit <Return> to see next plot: # divisive method
> fit <- diana(d1,metric='euclidean')
> plot(fit)
Hit <Return> to see next plot:
Hit <Return> to see next plot:
```

```
library(ggdendro)

if(require(cluster)){

  fit<- agnes(d1, metric = "manhattan", stand = TRUE)

  dg <- as.dendrogram(fit)

  ggdendrogram(dg)



  fit <- diana(d1, metric = "manhattan", stand = TRUE)

  dg <- as.dendrogram(fit)

  ggdendrogram(dg)

}
```

A **Scree Plot** is a simple line segment **plot** that shows the fraction of total variance in the data as explained or represented by each PC. The PCs are ordered, and by definition are therefore assigned a number label, by decreasing order of contribution to total variance.







Conclusion:-

The term *cluster analysis* encompasses a number of different algorithms and methods for grouping objects of similar kind into respective categories. In other words cluster analysis is an exploratory data analysis tool which aims at sorting different objects into groups in a way that the degree of association between two objects is maximal if they belong to the same group and minimal otherwise. Given the above, cluster analysis can be used to discover structures in data without providing an explanation/interpretation. In other words, cluster analysis simply discovers structures in data without explaining why they exist.

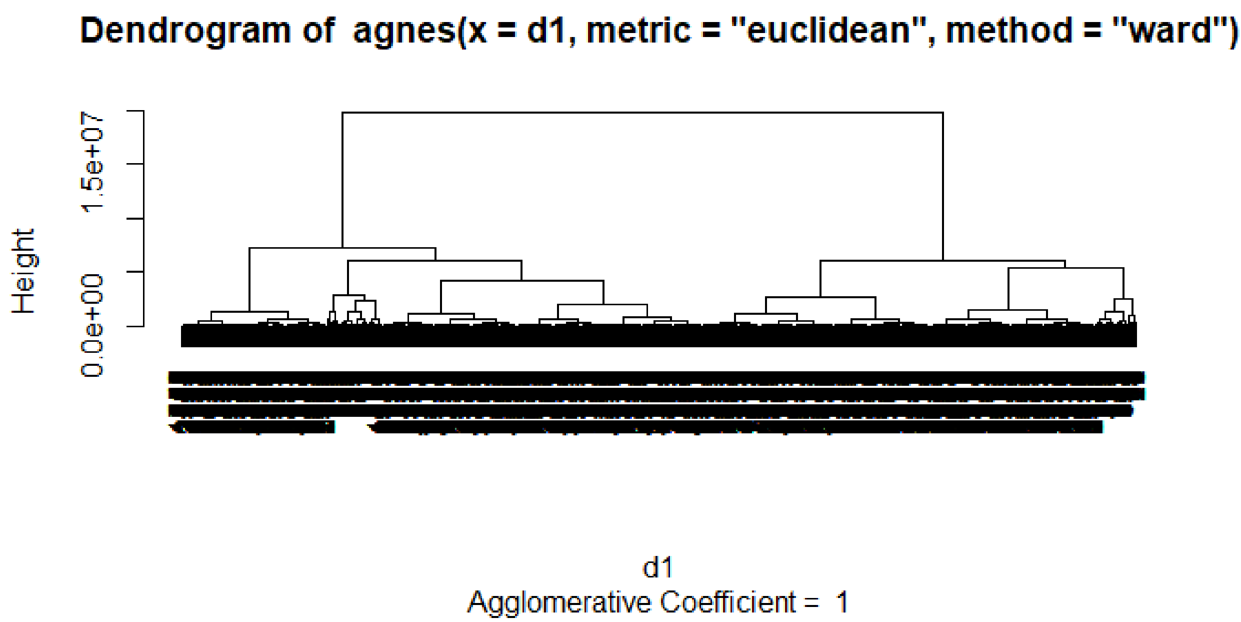
k-Means Clustering

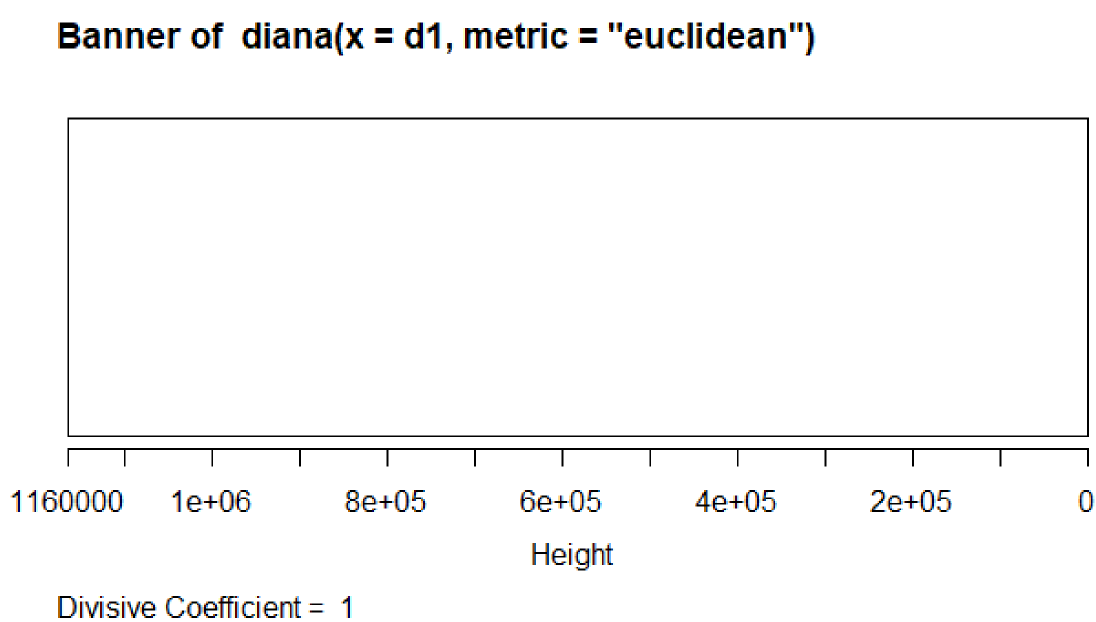
In general, the *k*-means method will produce exactly *k* different clusters of greatest possible distinction. It should be mentioned that the best number of clusters *k* leading to the greatest separation (distance) is not known as *a priori* and must be computed from the data

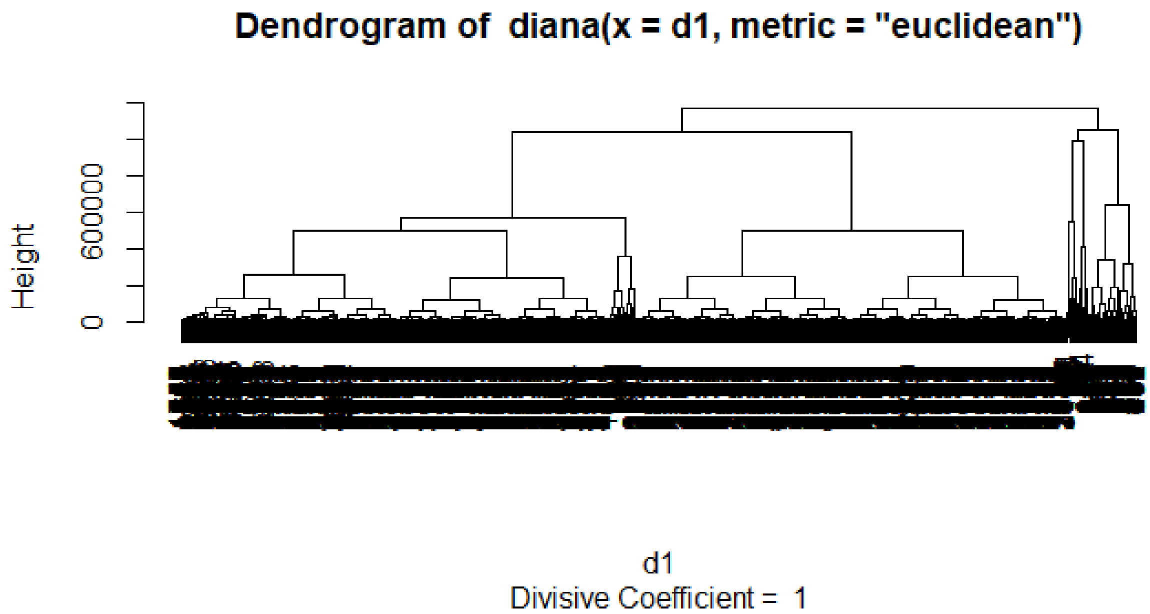
INTERPRETATION OF RESULTS

Usually, as the result of a k-means clustering analysis, we would examine the means for each cluster on each dimension to assess how distinct our k clusters are. Ideally, we would obtain very different means for most, if not all dimensions, used in the analysis. The magnitude of the F values from the analysis of variance performed on each dimension is another indication of how well the respective dimension discriminates between clusters.

Euclidean distance. This is probably the most commonly chosen type of distance. It simply is the geometric distance in the multidimensional space.







A **dendrogram** or **tree diagram** allows to illustrate the **hierarchical organisation** of several entities. For example, we often use it to make family trees. It is constituted of a root node, which give birth to several nodes that end by giving leaf nodes (the

bottom of the tree). Dendrogram can be made with 2 types of dataset. **i/** a **numeric matrix**where several variables describe the features of individuals. We can then calculate the distance between individuals and cluster them. **ii/** A **hierarchical**

gg dendroplot given below provides a good output. As the data is large it needs lot of cleaning and removal of missing data


0
300000
600000
900000
1200000