Title: Car Rating

Data Summary

The dataset, *cars,* contains 50 rating of automobiles based on MPG (mile per gallon), RELIABLE, ACCEL (acceleration), BRAKE, HANDLE, RIDE, VISIB (visibility), COMFORT, QUIET, and CARGO. The ratings are range from 1 to 5. Each car is given their ORIGIN, MAKE and MODEL.

```
summary(cars[4:13])
      MPG          RELIABLE        ACCEL          BRAKE          HANDLE
 Min.   :1.00   Min.   :1.00   Min.   :2.00   Min.   :2.00   Min.   :2.0
 1st Qu.:2.00   1st Qu.:1.00   1st Qu.:3.00   1st Qu.:3.00   1st Qu.:3.0
 Median :3.00   Median :2.00   Median :4.00   Median :4.00   Median :4.0
 Mean   :2.84   Mean   :2.32   Mean   :3.94   Mean   :3.88   Mean   :3.7
 3rd Qu.:3.00   3rd Qu.:3.00   3rd Qu.:5.00   3rd Qu.:4.00   3rd Qu.:4.0
 Max.   :5.00   Max.   :5.00   Max.   :5.00   Max.   :5.00   Max.   :5.0
      RIDE          VISIB        COMFORT        QUIET          CARGO
 Min.   :1.00   Min.   :1.0   Min.   :1.0   Min.   :1.00   Min.   :1.00
 1st Qu.:3.00   1st Qu.:3.0   1st Qu.:2.0   1st Qu.:2.00   1st Qu.:3.25
 Median :4.00   Median :3.0   Median :4.0   Median :3.00   Median :4.00
 Mean   :3.82   Mean   :3.4   Mean   :3.7   Mean   :3.18   Mean   :3.78
 3rd Qu.:5.00   3rd Qu.:4.0   3rd Qu.:5.0   3rd Qu.:4.00   3rd Qu.:5.00
 Max.   :5.00   Max.   :5.0   Max.   :5.0   Max.   :5.00   Max.   :5.00

round(cor(cars[4:13]),1)
          MPG RELIABLE ACCEL BRAKE HANDLE RIDE VISIB COMFORT QUIET CARGO
MPG       1.0      0.2  -0.2   0.4    0.3 -0.2   0.7    -0.1  -0.5   0.0
RELIABLE  0.2      1.0  -0.1   0.3   -0.1  0.3   0.3     0.4   0.5   0.4
ACCEL    -0.2     -0.1   1.0   0.1    0.1  0.3  -0.1     0.1   0.0  -0.1
BRAKE     0.4      0.3   0.1   1.0    0.2  0.3   0.4     0.3   0.0   0.2
HANDLE    0.3     -0.1   0.1   0.2    1.0  0.1   0.1    -0.1   0.0  -0.1
RIDE     -0.2      0.3   0.3   0.3    0.1  1.0   0.2     0.8   0.5   0.4
VISIB     0.7      0.3  -0.1   0.4    0.1  0.2   1.0     0.3  -0.2   0.4
COMFORT  -0.1      0.4   0.1   0.3   -0.1  0.8   0.3     1.0   0.4   0.5
QUIET    -0.5      0.5   0.0   0.0    0.0  0.5  -0.2     0.4   1.0   0.3
CARGO     0.0      0.4  -0.1   0.2   -0.1  0.4   0.4     0.5   0.3   1.0
```

The summary of ratings suggests there is no skew-ness in the data since mean and median of all ratings are about the same. From the correlation test, we can see that MPG is strongly correlated with VISIB & RIDE is strongly correlated with COMFORT.

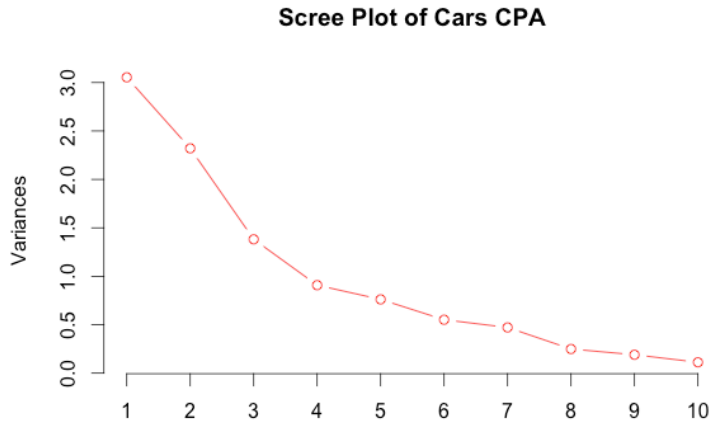
Statement of Problem

Frist, we want to the number of principal components are sufficient for the data. Then we would like to group car's model according to their ratings using Ward Hierarchical Clustering.

Principal Component Analysis

```
> summary(pca)
Importance of components:
                          PC1    PC2    PC3     PC4     PC5     PC6     PC7
Standard deviation     1.7475 1.5233 1.1754 0.95304 0.87259 0.74256 0.68677
Proportion of Variance 0.3054 0.2321 0.1382 0.09083 0.07614 0.05514 0.04717
Cumulative Proportion  0.3054 0.5374 0.6756 0.76641 0.84255 0.89769 0.94485
                           PC8    PC9    PC10
Standard deviation     0.49911 0.4359 0.33513
Proportion of Variance 0.02491 0.0190 0.01123
Cumulative Proportion  0.96977 0.9888 1.00000

screeplot(pca, type="lines",col=3, ,main = "Scree Plot of Cars CPA")
```

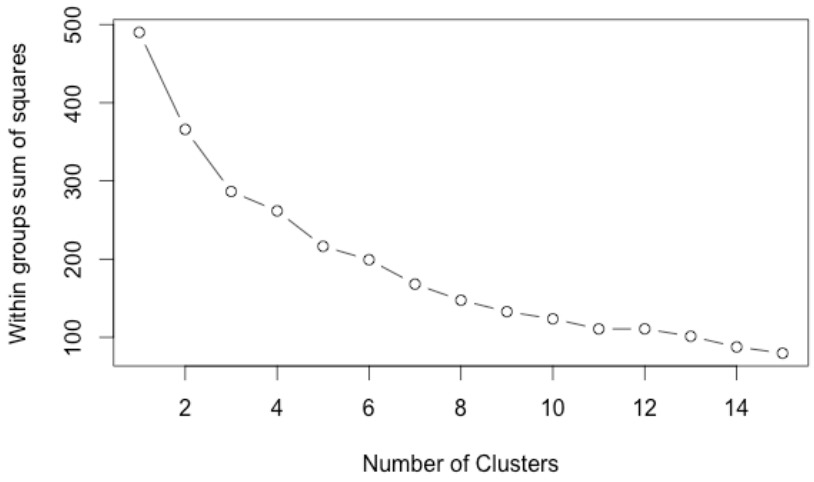



```
> round(pca$rotation[,1:6],2)
           PC1   PC2   PC3   PC4   PC5   PC6
MPG      -0.04  0.62 -0.01  0.01 -0.13  0.18
RELIABLE -0.37  0.08 -0.27 -0.25 -0.67  0.20
ACCEL    -0.04 -0.17  0.62  0.43 -0.42  0.23
BRAKE    -0.29  0.33  0.26  0.00 -0.13 -0.84
HANDLE   -0.01  0.20  0.52 -0.67  0.29  0.24
RIDE     -0.45 -0.20  0.33  0.08  0.14  0.06
VISIB    -0.28  0.48 -0.06  0.18  0.09  0.29
COMFORT  -0.48 -0.11  0.02  0.17  0.24  0.11
QUIET    -0.33 -0.39 -0.10 -0.46 -0.16 -0.06
CARGO    -0.40 -0.02 -0.28  0.17  0.38  0.01
```

Based on the cumulative proportion of variance and scree plot, first six principal components explained about 90% of the variation. The first component mainly measures reliable, ride, comfort, and cargo. The second component mainly measures MPG and visibility. The third component mainly measures acceleration and handle. The fourth component mainly measures handle. The fifth component mainly measures reliable and sixth component heavily weighted on the brake. These six components are sufficient to explain the variation in the data.

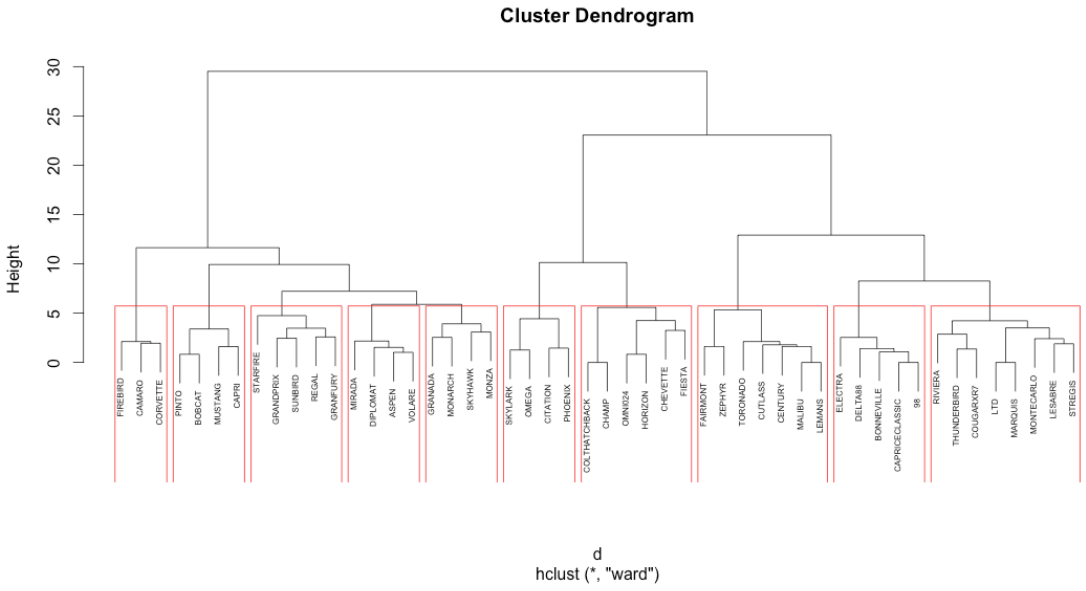
Ward Hierarchical Clustering

```
> wss <- (nrow(scd.car)-1)*sum(apply(scd.car,2,var))
> for (i in 2:15) wss[i] <- sum(kmeans(scd.car,centers=i)$withinss)
> plot(1:15, wss, type="b", xlab="Number of Clusters",ylab="Within
groups sum of squares")## 10 clusters
```



According to the plot on the left. The number of clusters that have relative minimum within sum of squares is 10. Therefore, we separate cars into 10 groups based on their ratings.

```
> d = dist(scd.car, method = "euclidean")##distance matrix
> fit = hclust(d, method="ward")
> plot(fit, cex = .5)
> plot(fit, cex = .5, labels = MODEL)
> groups = cutree(fit,k=10)## cut tree into 10 clusters
```



From the dendrogram, we can see the name of models in each group and the number of models in each group. Then we want to know what the model and the ratings are based on for these ten groups in conclusion.

Conclusion

```
> for (i in  1:10) { print(subset(car, GROUP == i) )}
      MODEL MPG RELIABLE ACCEL BRAKE HANDLE RIDE VISIB COMFORT QUIET CARGO GROUP
1   CENTURY   3        3     3     4      4    4     4       5     4     4     1
13   MALIBU   3        3     3     3      4    4     4       5     4     4     1
22 FAIRMONT   3        3     2     4      3    4     5       4     3     4     1
34   ZEPHYR   3        1     2     4      3    4     5       4     3     4     1
35  CUTLASS   3        4     4     3      4    4     4       5     4     4     1
40 TORONADO   3        3     2     3      4    4     3       5     4     4     1
48   LEMANS   3        3     3     3      4    4     4       5     4     4     1
```

Cluster 1 includes model Century, Malibu, Fairmont, Zephyr, Cutlass, Toronado, and Lemans. They are models with high rating of comfort and visibility and mediocre ratings for rest of the variables .

```
            MODEL MPG RELIABLE ACCEL BRAKE HANDLE RIDE VISIB COMFORT QUIET CARGO GROUP
2         ELECTRA   2        4     3     4      4    5     3       5     5     5     2
9  CAPRICECLASSIC   2        4     4     5      3    5     3       5     5     5     2
36        DELTA88   2        4     3     5      3    5     3       5     5     5     2
37              98   2        4     4     5      3    5     3       5     5     5     2
45     BONNEVILLE   2        3     4     5      3    5     3       5     5     5     2
```

Cluster 2 includes Electra, Carpice Classic, Delta88, 98, Bonneville. They are the models with highest rating in brake, ride, comfort, quiet and cargo and mediocre for others.

```
          MODEL MPG RELIABLE ACCEL BRAKE HANDLE RIDE VISIB COMFORT QUIET CARGO GROUP
3       LESABRE   2        3     5     4      3    5     3       5     4     5     3
5       RIVIERA   2        3     5     4      5    5     3       5     4     3     3
14  MONTECARLO   3        2     5     3      3    5     3       5     4     4     3
21     STREGIS   1        1     5     4      3    5     3       5     4     5     3
25         LTD   3        3     5     4      3    5     4       5     5     5     3
28 THUNDERBIRD   2        3     5     4      3    4     4       4     4     4     3
31   COUGARXR7   2        4     5     4      4    4     4       4     4     4     3
32     MARQUIS   3        3     5     4      3    5     4       5     5     5     3
```

Cluster 3 includes Lesabre, Riviera, Monte Carlo, Stregis, LTD, Thunderbird, CougarxR7, and Marquis. They are models with high ratings for acceleration, ride, comfort, quite and cargo and mediocre for others.

```
        MODEL MPG RELIABLE ACCEL BRAKE HANDLE RIDE VISIB COMFORT QUIET CARGO GROUP
4       REGAL   3        2     4     4      4    4     3       4     2     4     4
39  STARFIRE   2        1     3     3      5    2     2       1     5     4     4
42  GRANFURY   2        1     3     4      3    5     3       5     3     5     4
47 GRANDPRIX   3        2     2     4      4    3     2       4     3     4     4
50   SUNBIRD   3        1     3     4      5    3     3       2     3     4     4
```

Cluster 4 includes Regal, Strafire, Granfury, Grandprix, and sunbird. They are models with high rating for cargo, low rating for reliable and mediocre for rest.

```
     MODEL MPG RELIABLE ACCEL BRAKE HANDLE RIDE VISIB COMFORT QUIET CARGO GROUP
6  SKYHAWK   3        2     3     2      4    2     3       2     2     4     5
15   MONZA   2        1     4     2      2    3     3       1     1     4     5
24 GRANADA   2        2     3     3      2    3     3       2     3     3     5
33 MONARCH   2        3     5     3      2    3     2       2     3     2     5
```

Cluster 6 includes Skyhawk, Monza, Granda, and Monarch. They are models with mediocre ratings for all the variables.

	MODEL	MPG	RELIABLE	ACCEL	BRAKE	HANDLE	RIDE	VISIB	COMFORT	QUIET	CARGO	GROUP
7	SKYLARK	4	1	4	5	5	5	5	4	2	2	6
11	CITATION	4	1	5	5	5	5	5	5	2	5	6
38	OMEGA	4	1	5	5	5	5	5	5	2	2	6
49	PHOENIX	4	1	5	5	5	5	4	4	1	5	6

Cluster 6 includes Skylark, Citation, Omega, and Phoenix. They are models with high MPG, acceleration, brake, handle, ride, visibility, comfort but low rating for reliable and quite.

	MODEL	MPG	RELIABLE	ACCEL	BRAKE	HANDLE	RIDE	VISIB	COMFORT	QUIET	CARGO	GROUP
8	CAMARO	2	2	5	4	5	4	1	2	4	1	7
12	CORVETTE	2	1	5	3	5	4	2	2	4	2	7
46	FIREBIRD	1	1	5	3	5	5	1	2	3	1	7

Cluster 7 includes Camaro, Corvette, and firebird. They are car models with high rating for acceleration and handle, but low rating for MPG, reliable, visibility and cargo.

	MODEL	MPG	RELIABLE	ACCEL	BRAKE	HANDLE	RIDE	VISIB	COMFORT	QUIET	CARGO	GROUP
10	CHEVETTE	5	3	3	5	4	2	5	2	2	3	8
17	COLTHATCHBACK	5	5	4	4	4	4	5	4	3	4	8
20	OMNI024	4	3	4	5	5	3	5	2	2	5	8
23	FIESTA	5	4	4	5	3	4	4	4	1	4	8
41	CHAMP	5	5	4	4	4	4	5	4	3	4	8
43	HORIZON	4	3	4	5	5	3	5	2	3	5	8

Cluster 8 includes Chevette, Colt hatchback, Omni023, Fiesta, Champ, Horizon. These models have high rating for all variables except mediocre rating for comfort and quiet.

	MODEL	MPG	RELIABLE	ACCEL	BRAKE	HANDLE	RIDE	VISIB	COMFORT	QUIET	CARGO	GROUP
16	ASPEN	2	1	4	3	3	3	3	4	2	4	9
18	DIPLOMAT	2	1	5	3	3	4	3	4	3	4	9
19	MIRADA	2	1	4	3	4	3	2	4	3	4	9
44	VOLARE	2	1	5	3	3	3	3	4	2	4	9

Cluster 9 includes Aspen, Diplomat, Mirada, and Volare. They are models with good rating for acceleration, comfort, and cargo but low rating for reliable.

	MODEL	MPG	RELIABLE	ACCEL	BRAKE	HANDLE	RIDE	VISIB	COMFORT	QUIET	CARGO	GROUP
26	MUSTANG	3	2	4	4	3	2	3	2	2	2	10
27	PINTO	4	1	3	4	3	1	3	2	2	2	10
29	BOBCAT	4	1	3	4	3	1	3	2	1	2	10
30	CAPRI	3	1	5	4	3	2	2	2	2	2	10

Cluster 10 includes Mustang, Pinto, Bobcat, and Capri. They are models with low/mediocre ratings for most variables.