

Title: Mammal's Teeth

Data Summary

The *teeth* dataset contains 31 observations and 9 variables. The data includes the numbers of different kinds of teeth for a variety of mammals. The different types of teeth are top incisors, bottom incisors, top canines, bottom canines, top premolars, bottom premolars, top molars, and bottom molars.

Summary of teeth:

```
> summary(teeth[-1])
```

TINCISOR		BINCISOR		TCANINE		BCANINE	
Min.	:0.000	Min.	:1.000	Min.	:0.0000	Min.	:0.000
1st Qu.:	1.000	1st Qu.:	1.750	1st Qu.:	0.0000	1st Qu.:	0.000
Median	:2.000	Median	:3.000	Median	:1.0000	Median	:1.000
Mean	:2.031	Mean	:2.469	Mean	:0.7188	Mean	:0.625
3rd Qu.:	3.000	3rd Qu.:	3.000	3rd Qu.:	1.0000	3rd Qu.:	1.000
Max.	:3.000	Max.	:4.000	Max.	:1.0000	Max.	:1.000

TPRMOLAR		BPRMOLAR		TMOLAR		BMOLAR	
Min.	:0.000	Min.	:0.000	Min.	:1.000	Min.	:1.000
1st Qu.:	2.000	1st Qu.:	2.000	1st Qu.:	1.000	1st Qu.:	2.000
Median	:3.000	Median	:3.000	Median	:3.000	Median	:3.000
Mean	:2.812	Mean	:2.688	Mean	:2.219	Mean	:2.438
3rd Qu.:	4.000	3rd Qu.:	3.250	3rd Qu.:	3.000	3rd Qu.:	3.000
Max.	:4.000	Max.	:4.000	Max.	:3.000	Max.	:3.000

```
> round(cor(teeth[-1]),2)
```

	TINCISOR	BINCISOR	TCANINE	BCANINE	TPRMOLAR	BPRMOLAR	TMOLAR	BMOLAR
TINCISOR	1.00	-0.07	0.60	0.74	0.51	0.47	-0.70	-0.60
BINCISOR	-0.07	1.00	0.50	0.30	0.38	0.49	-0.08	0.02
TCANINE	0.60	0.50	1.00	0.81	0.55	0.66	-0.53	-0.47
BCANINE	0.74	0.30	0.81	1.00	0.48	0.56	-0.65	-0.58
TPRMOLAR	0.51	0.38	0.55	0.48	1.00	0.90	-0.60	-0.54
BPRMOLAR	0.47	0.49	0.66	0.56	0.90	1.00	-0.50	-0.41
TMOLAR	-0.70	-0.08	-0.53	-0.65	-0.60	-0.50	1.00	0.85
BMOLAR	-0.60	0.02	-0.47	-0.58	-0.54	-0.41	0.85	1.00

The mean and median of different teeth are about the same, so we can say the data is not skewed. From the correlation test, we can see there is high correlation between top and bottom teeth for CANINE, PRMOLAR, and MOLAR. However, the correlation for top and bottom INCISOR is very low, but TINCISOR are highly correlated with CANINE (positively) and MOLAR (negatively).

Statement of Problem

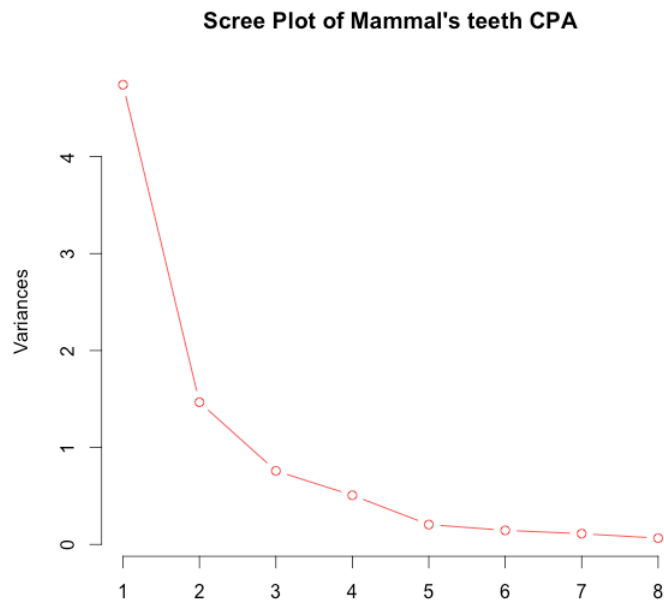
For the data, we want examine how many components can explain the data using principal component analysis. And we can to group different mammals according to their number of different types of teeth using Ward Hierarchical Clustering. The numbers of clusters are based on within group sum of squares.

Principal Component Analysis

```
> pca = prcomp(teeth[2:9], scale. = T)
> summary(pca)
```

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Standard deviation	2.1775	1.2112	0.87104	0.71221	0.45233	0.38124	0.33293	0.2545
Proportion of Variance	0.5927	0.1834	0.09484	0.06341	0.02557	0.01817	0.01386	0.0081
Cumulative Proportion	0.5927	0.7761	0.87090	0.93431	0.95988	0.97805	0.99190	1.0000



Based on the proportion of variance and scree plot, we know the first four components are enough, which explained about 94% of the data.

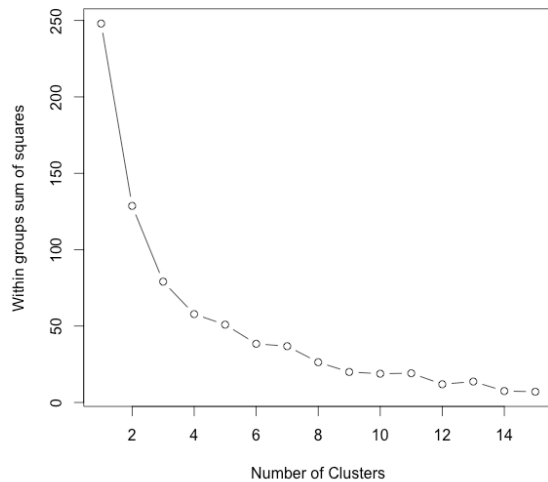
```
> pca$rotation[,1:4]
```

	PC1	PC2	PC3	PC4
TINCISOR	0.3596792	0.33035725	-0.2143159	0.47290399
BINCISOR	0.1682294	-0.69061491	-0.1527100	-0.44556742
TCANINE	0.3827299	-0.19801013	-0.4371581	0.03571043
BCANINE	0.3911168	0.04945103	-0.4979039	0.03992166
TPRMOLAR	0.3727780	-0.19259178	0.5525092	0.21353051
BPRMOLAR	0.3709015	-0.32330600	0.3483855	0.33260060
TMOLAR	-0.3815157	-0.30733168	-0.1556981	0.38285244
BMOLAR	-0.3475482	-0.36904010	-0.2021190	0.52169731

Above is the loading of first four principal components. The weight for PC1 is about the same for all variables except BINCISOR; all variables are negatively related to MOLAR teeth. PC2 mostly measure variable BINCISOR, the top and bottom incisor teeth are negatively related. PC3 is weighted heavy on CANINE teeth with opposite relationship with top MOLAR teeth. PC 4 explains mostly INCISOR and MOLAR teeth; top incisor teeth and molar teeth are opposite related to bottom incisor teeth. Therefore, 4 principal components are enough to explain the data.

Ward Hierarchical Clustering

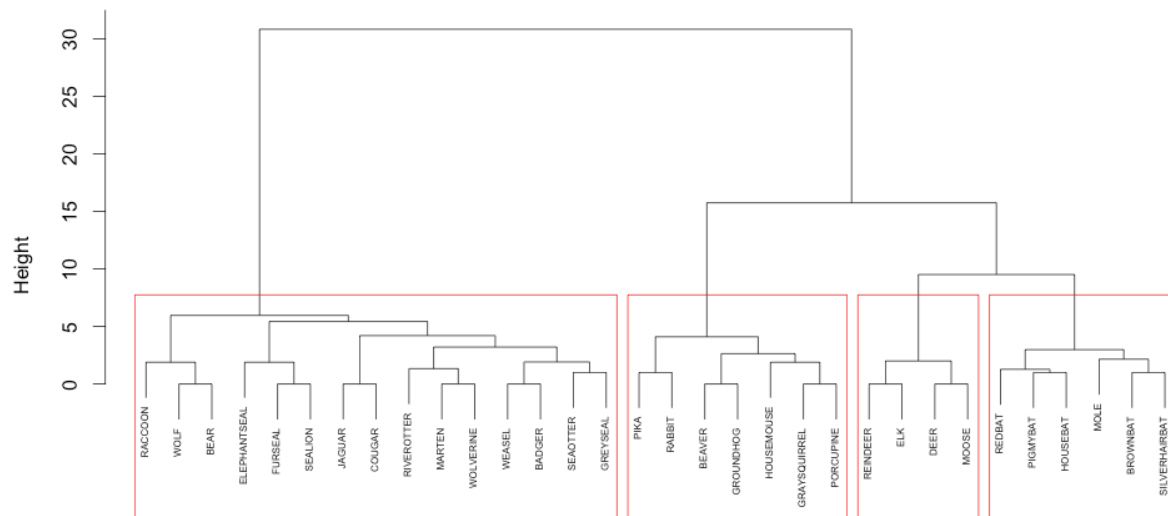
```
scaled.teeth = scale(na.omit(teeth[-1]))
wss <- (nrow(scaled.teeth)-1)*sum(apply(scaled.teeth,2,var))
for (i in 2:15) wss[i] <- sum(kmeans(scaled.teeth,centers=i)$withinss)
plot(1:15, wss, type="b", xlab="Number of Clusters",ylab="Within groups sum of squares")
```



Based on the within groups sum of square, four clusters reduced the sum of square tremendously. Therefore, we want to group mammal into 4 clusters.

```
d = dist(teeth[-1], method = "euclidean")##distance matrix
fit = hclust(d, method="ward")
plot(fit, labels =ANIMAL, cex =.5 )
groups = cutree(fit,k=4)## cut tree into 4 clusters
rect.hclust(fit, k=4, border="red")
```

Cluster Dendrogram



d
hclust (*, "ward")

```

GROUP = as.factor(groups)
teeth = cbind(teeth, GROUP)

for(i in 1:4){print(subset(teeth, GROUP == i))}
GROUP = 1
  ANIMAL TINCISOR BINCISOR TCANINE BCANINE TPRMOLAR BPRMOLAR TMOLAR BMOLAR
1  BROWNBAT      2        3        1        1          3          3          3          3
2    MOLE       3        2        1        0          3          3          3          3
3 SILVERHAIRBAT  2        3        1        1          2          3          3          3
4   PIGMYBAT    2        3        1        1          2          2          3          3
5   HOUSEBAT    2        3        1        1          1          2          3          3
6   REDBAT     1        3        1        1          2          2          3          3
GROUP = 2
  ANIMAL TINCISOR BINCISOR TCANINE BCANINE TPRMOLAR BPRMOLAR TMOLAR BMOLAR
7    PIKA      2        1        0        0          2          2          3          3
8   RABBIT    2        1        0        0          3          2          3          3
9   BEAVER    1        1        0        0          2          1          3          3
10  GROUNDHOG  1        1        0        0          2          1          3          3
11 GRAYSQUIRREL 1        1        0        0          1          1          3          3
12  HOUSEMOUSE 1        1        0        0          0          0          3          3
13  PORCUPINE  1        1        0        0          1          1          3          3
GROUP = 3
  ANIMAL TINCISOR BINCISOR TCANINE BCANINE TPRMOLAR BPRMOLAR TMOLAR BMOLAR
14   WOLF      3        3        1        1          4          4          2          3
15   BEAR      3        3        1        1          4          4          2          3
16  RACCOON    3        3        1        1          4          4          3          2
17  MARTEN     3        3        1        1          4          4          1          2
18  WEASEL     3        3        1        1          3          3          1          2
19  WOLVERINE  3        3        1        1          4          4          1          2
20  BADGER     3        3        1        1          3          3          1          2
21  RIVERTOTTER 3        3        1        1          4          3          1          2
22  SEATOTTER  3        2        1        1          3          3          1          2
23  JAGUAR     3        3        1        1          3          2          1          1
24  COUGAR     3        3        1        1          3          2          1          1
25  FURSEAL    3        2        1        1          4          4          1          1
26  SEALION    3        2        1        1          4          4          1          1
27  GREYSEAL   3        2        1        1          3          3          2          2
28 ELEPHANTSEAL 2        1        1        1          4          4          1          1
GROUP = 4
  ANIMAL TINCISOR BINCISOR TCANINE BCANINE TPRMOLAR BPRMOLAR TMOLAR BMOLAR
29 REINDEER    0        4        1        0          3          3          3          3
30   ELK       0        4        1        0          3          3          3          3
31   DEER      0        4        0        0          3          3          3          3
32  MOOSE      0        4        0        0          3          3          3          3

```

Conclusion

Cluster 1 includes all the bats family and the mole. They have all different type of teeth and with only one top canine tooth and one-bottom canine tooth (except for mole with no bottom canine tooth).

Cluster 2 contains rabbit family and small to median-rodents. They tend to be herbivorous with no top and bottom canine tooth.

Cluster 3 contains mostly mid-size to large mammals. They are most likely to be carnivorous with one top canine and one bottom canine.

Cluster 4 contains the deer family. They are herbivorous with no top incisor and four bottom incisor and no canine teeth (except for reindeer and elk with one top canine tooth).