# Analysis on Dataset “teeth”

by Yimi Zhao

## 1 Introduction of Dataset

The “teeth” dataset contains 32 observations across 9 variables. It records the numbers of eight types of teeth of 32 animals. The data format is:

**teeth** animal, tincisor, bincisor, tcanine, bcanine, tprmolar, bprmolar, tmolar, bmolar [“t” = top, “b” = bottom]

```
 > head(teeth)
         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
```

## 2 Data Analysis

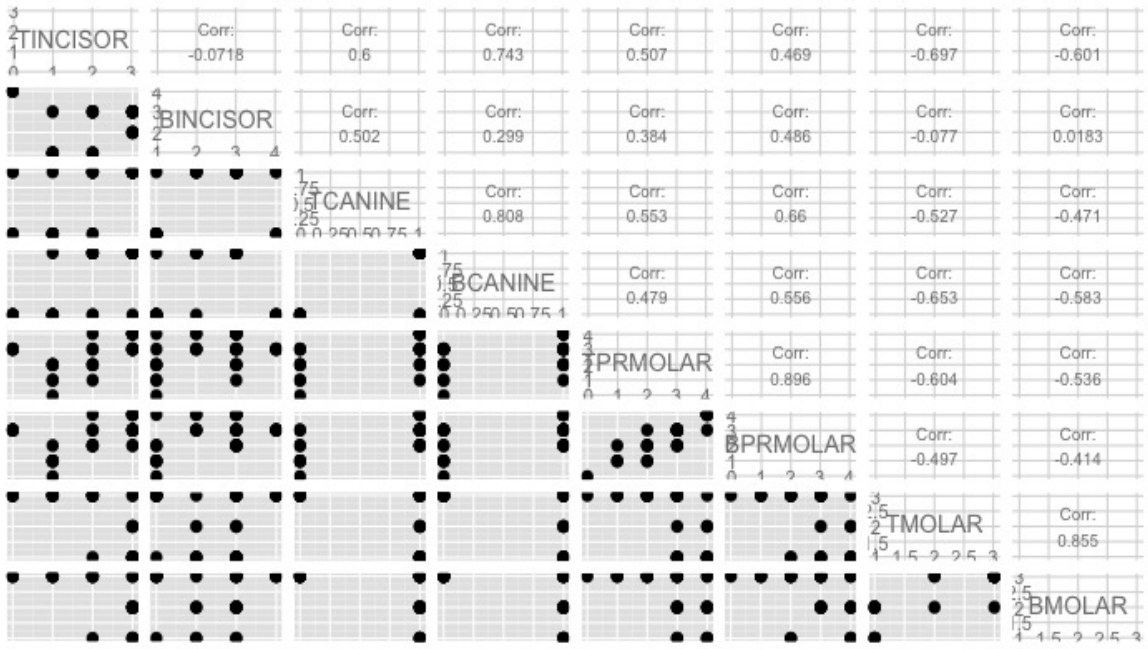
### 2.1 Data Exploration

The data summary does not have an indication of extreme observations. But we do realize that there are some zero values in TCAINE, BCANINE, TPRMOLAR, BPRMOLAR, which could be our clues to further analyze the data.

```
# data exploration
>summary(teeth[,-1])
    TINCISOR        BINCISOR        TCANINE          BCANINE         TPRMOLAR    
 Min.   :0.000   Min.   :1.000   Min.   :0.0000   Min.   :0.000   Min.   :0.000  
 1st Qu.:1.000   1st Qu.:1.750   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:2.000  
 Median :2.000   Median :3.000   Median :1.0000   Median :1.000   Median :3.000  
 Mean   :2.031   Mean   :2.469   Mean   :0.7188   Mean   :0.625   Mean   :2.812  
 3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:4.000  
 Max.   :3.000   Max.   :4.000   Max.   :1.0000   Max.   :1.000   Max.   :4.000  

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

# pair plot
>library(GGally)
>ggpairs(teeth[,-1],cex=0.8)
```



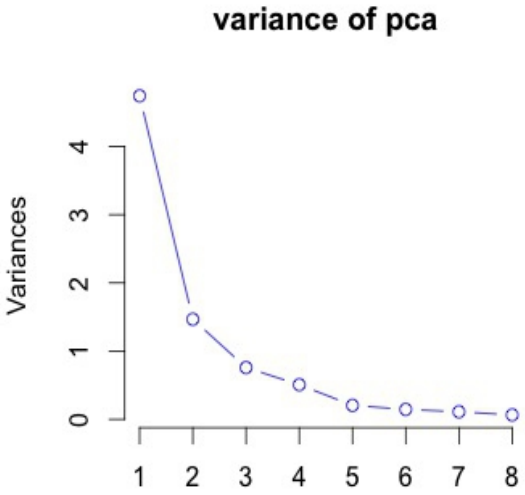

According to the pair correlations, we found strong some positive correlations: TCANINE vs BCANINE (corr=0.808), TPRMOLAR vs BPRMOLAR (corr=0.896), TMOLAR vs BMOLAR(corr=0.855) and TINCISOR vs BCANINE(corr=0.743). The result indicates the number of top and bottom teeth is highly related, the more top teeth, the more bottom teeth, except the incisor teeth.

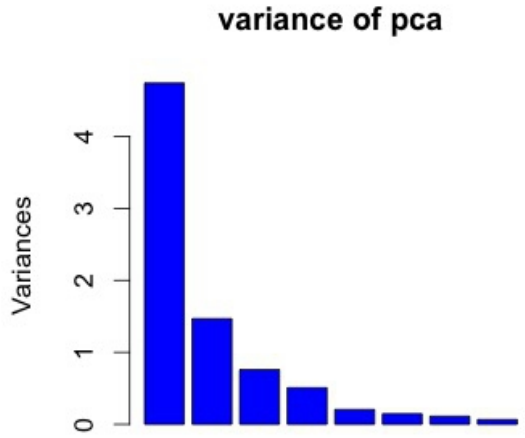
## 2.2 Principal Component Analysis

Since there are as many as eight variables recording the numbers of different types of teeth, and strong correlations exist among them. Naturally, we should consider to reduce the data dimension to simplify the further analysis.

Based on the output of principal component analysis, the first component(PC1) alone explains 59% of the variance, the second component(PC2) explains 7.7%, and the first four components (PC1+PC2+PC3+PC4) together account for 93.4% of the variance. Therefore, we may consider apply these four components in further data analysis instead.

```
#conduct pca
> teeth.pca<- prcomp(teeth[,-1],scale=T)
> summary(teeth.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
```




```
>round(teeth.pca$rotation,3)
            PC1    PC2    PC3    PC4    PC5    PC6    PC7    PC8
TINCISOR  0.360  0.330 -0.214  0.473 -0.485  0.316 -0.341  0.201
BINCISOR  0.168 -0.691 -0.153 -0.446 -0.436  0.081 -0.253  0.111
TCANINE   0.383 -0.198 -0.437  0.036  0.507  0.539  0.237 -0.133
BCANINE   0.391  0.049 -0.498  0.040  0.025 -0.735 -0.025 -0.231
TPRMOLAR  0.373 -0.193  0.553  0.214 -0.044  0.037 -0.142 -0.671
BPRMOLAR  0.371 -0.323  0.348  0.333  0.211 -0.239  0.169  0.629
TMOLAR   -0.382 -0.307 -0.156  0.383  0.374 -0.057 -0.667 -0.017
BMOLAR   -0.348 -0.369 -0.202  0.522 -0.360 -0.019  0.519 -0.175
```

The rotation table above shows how the PCs are consist of the original eight variables. Clearly, the weights of PC1 are quite similar(around 0.35), except BINCISOR. On the contrary, the PC2 and PC4 mainly explain BINCISOR (-0.691, 0.473 and -0.446). PC3 primarily relates to CANINE and PRMOLAR teeth.

## 2.3 Cluster Analysis

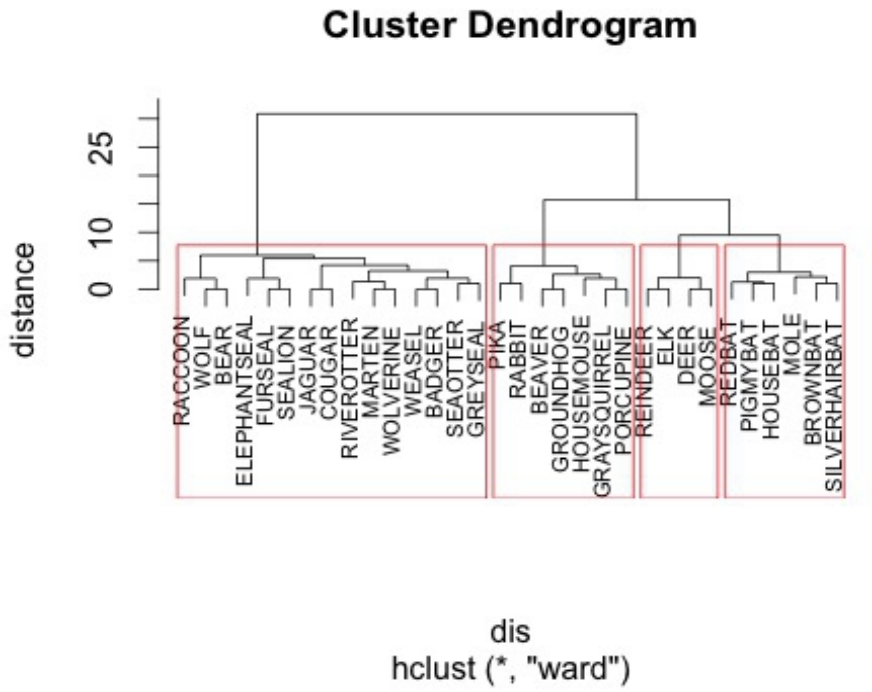
The dataset covers 32 animals, checking the records of their teeth, we found some of them actually have quite similar values. Then, it is reasonable that we may divide these animals into a few groups based on their teeth numbers. Therefore, we are going to apply cluster analysis to further explore this dataset.

### 2.3.1 Hierarchical and K-means Cluster Analysis

First, we try hierarchical cluster analysis.

```
# hierarchical cluster analysis
> teeth2<-teeth[-1]
> dis<-dist(teeth2,method='euclidean')
> dis.matrix=as.matrix(dis)
> teeth_ward<-hclust(dis, method="ward")
```

```
#plot the dendrogram cluster with city names at the bottom
> plot(teeth_ward, labels=teeth$ANIMAL, ylab='distance',cex=0.7)
> #cut the tree into 4 clusters and draw dendrogram with red borders around the clusters
> groups<-cutree(teeth_ward,k=4)
> rect.hclust(teeth_ward, k=4, border="red")
```

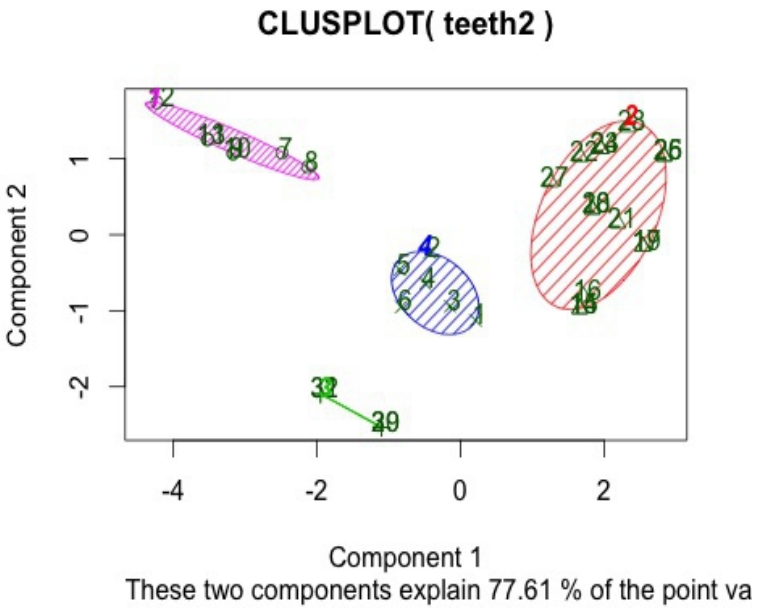



```
##analyze the result
> round(aggregate(teeth[,2:9], by=list(groups),FUN=mean),3)
  Group.1 TINCISOR BINCISOR TCANINE BCANINE TPRMOLAR BPRMOLAR TMOLAR BMOLAR
1       1    2.000    2.833     1.0   0.833    2.167    2.500  3.000    3.0
2       2    1.286    1.000     0.0   0.000    1.571    1.143  3.000    3.0
3       3    2.933    2.600     1.0   1.000    3.600    3.400  1.333    1.8
4       4    0.000    4.000     0.5   0.000    3.000    3.000  3.000    3.0
> GROUP<-as.factor(groups)
> clusters<-cbind(teeth, GROUP)
> table(clusters$GROUP)
 1  2  3  4
 6  7 15  4
```

Through Ward method, the 32 animals are nicely divided into 4 clusters with size 6, 7, 15, 4.

To confirm our analysis, we continue to apply k-means cluster analysis. And it turns out we got the same clusters with exactly size.

```
# K-means
>teeth.k<-kmeans(teeth2,centers=4)
> teeth.k$size
[1]  7 15  4  6
> round(aggregate(teeth2,by=list(teeth.k$cluster),FUN=mean),3)
  Group.1 TINCISOR BINCISOR TCANINE BCANINE TPRMOLAR BPRMOLAR TMOLAR BMOLAR
1       1    1.286    1.000     0.0   0.000    1.571    1.143  3.000    3.0
2       2    2.933    2.600     1.0   1.000    3.600    3.400  1.333    1.8
3       3    0.000    4.000     0.5   0.000    3.000    3.000  3.000    3.0
4       4    2.000    2.833     1.0   0.833    2.167    2.500  3.000    3.0
> clusplot(teeth2, teeth.k$cluster, color=TRUE, shade=TRUE,labels=2, lines=0)
```




### 2.3.2 Summary and Conclusion

Since both hierarchical and k-means analysis yield the same result, in the summary, we just quote the result form k-means.

**Summary of Cluster Analysis(Mean of teeth number)**

| Group | Size | TINCISOR | BINCISOR | TCANINE | BCANINE | TPRMOLAR | BPRMOLAR | TMOLAR | BMOLAR |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 7 | 1.29 | 1.00 | 0.0 | 0.00 | 1.57 | 1.14 | 3.00 | 3.0 |
| 2 | 15 | 2.93 | 2.60 | 1.0 | 1.00 | 3.60 | 3.40 | 1.33 | 1.8 |
| 3 | 4 | 0.00 | 4.00 | 0.5 | 0.00 | 3.00 | 3.00 | 3.00 | 3.0 |
| 4 | 6 | 2.00 | 2.83 | 1.0 | 0.83 | 2.17 | 2.50 | 3.00 | 3.0 |

According to the summary table, we come up with the following conclusions:

- These 32 animals can be divided into four groups with size 7,15,4,and 6, and each group has similar teeth record.
- The seven animals in group 1 all have no CANINE teeth(top and bottom), three MOLAR teeth(top and bottom), one bottom INCISOR tooth. Furthermore, they also have 1.29 top INCISOR, 1.57 top PRMOLAR and 1.14 bottom PRMOLAR teeth on average.

| | ANIMAL | TINCISOR | BINCISOR | TCANINE | BCANINE | TPRMOLAR | BPRMOLAR | TMOLAR | BMOLAR | GROUP |
|---|---|---|---|---|---|---|---|---|---|---|
| 7 | PIKA | 2 | 1 | 0 | 0 | 2 | 2 | 3 | 3 | 1 |
| 8 | RABBIT | 2 | 1 | 0 | 0 | 3 | 2 | 3 | 3 | 1 |
| 9 | BEAVER | 1 | 1 | 0 | 0 | 2 | 1 | 3 | 3 | 1 |
| 10 | GROUNDHOG | 1 | 1 | 0 | 0 | 2 | 1 | 3 | 3 | 1 |
| 11 | GRAYSQUIRREL | 1 | 1 | 0 | 0 | 1 | 1 | 3 | 3 | 1 |
| 12 | HOUSEMOUSE | 1 | 1 | 0 | 0 | 0 | 0 | 3 | 3 | 1 |
| 13 | PORCUPINE | 1 | 1 | 0 | 0 | 1 | 1 | 3 | 3 | 1 |

- For group 2, the fifteen animals do not miss any of the eight kinds of teeth. They all have one top and bottom CANINE tooth. Furthermore, they tend to have around 3 INCISOR teeth(top and bottom respectively), around 3.5 PRMOLAR teeth, 1.33 top MOLAR tooth and 1.8 on

average, and bottom MOLAR tooth on average.

| | ANIMAL | TINCISOR | BINCISOR | TCANINE | BCANINE | TPRMOLAR | BPRMOLAR | TMOLAR | BMOLAR | GROUP |
|---|---|---|---|---|---|---|---|---|---|---|
| 14 | WOLF | 3 | 3 | 1 | 1 | 4 | 4 | 2 | 3 | 2 |
| 15 | BEAR | 3 | 3 | 1 | 1 | 4 | 4 | 2 | 3 | 2 |
| 16 | RACCOON | 3 | 3 | 1 | 1 | 4 | 4 | 3 | 2 | 2 |
| 17 | MARTEN | 3 | 3 | 1 | 1 | 4 | 4 | 1 | 2 | 2 |
| 18 | WEASEL | 3 | 3 | 1 | 1 | 3 | 3 | 1 | 2 | 2 |
| 19 | WOLVERINE | 3 | 3 | 1 | 1 | 4 | 4 | 1 | 2 | 2 |
| 20 | BADGER | 3 | 3 | 1 | 1 | 3 | 3 | 1 | 2 | 2 |
| 21 | RIVEROTTER | 3 | 3 | 1 | 1 | 4 | 3 | 1 | 2 | 2 |
| 22 | SEAOTTER | 3 | 2 | 1 | 1 | 3 | 3 | 1 | 2 | 2 |
| 23 | JAGUAR | 3 | 3 | 1 | 1 | 3 | 2 | 1 | 1 | 2 |
| 24 | COUGAR | 3 | 3 | 1 | 1 | 3 | 2 | 1 | 1 | 2 |
| 25 | FURSEAL | 3 | 2 | 1 | 1 | 4 | 4 | 1 | 1 | 2 |
| 26 | SEALION | 3 | 2 | 1 | 1 | 4 | 4 | 1 | 1 | 2 |
| 27 | GREYSEAL | 3 | 2 | 1 | 1 | 3 | 3 | 2 | 2 | 2 |
| 28 | ELEPHANTSEAL | 2 | 1 | 1 | 1 | 4 | 4 | 1 | 1 | 2 |

- Group 3 covers four animals, they are quite close to each other: no top INCISOR and bottom CANINE teeth, 4 bottom INCISOR teeth, 3 PRMOLAR teeth(top and bottom) and MOLAR teeth(top and bottom), either one or zero top CANINE tooth.

| | ANIMAL | TINCISOR | BINCISOR | TCANINE | BCANINE | TPRMOLAR | BPRMOLAR | TMOLAR | BMOLAR | GROUP |
|---|---|---|---|---|---|---|---|---|---|---|
| 29 | REINDEER | 0 | 4 | 1 | 0 | 3 | 3 | 3 | 3 | 3 |
| 30 | ELK | 0 | 4 | 1 | 0 | 3 | 3 | 3 | 3 | 3 |
| 31 | DEER | 0 | 4 | 0 | 0 | 3 | 3 | 3 | 3 | 3 |
| 32 | MOOSE | 0 | 4 | 0 | 0 | 3 | 3 | 3 | 3 | 3 |

- Group 4 consists of six animals, as for INCISOR and CANINE teeth, this group is pretty similar with group 2. The most difference relies on top MOLAR teeth, group 4 have 3 top MOLAR teeth while group 2 posses 1.33 on average.

| | ANIMAL | TINCISOR | BINCISOR | TCANINE | BCANINE | TPRMOLAR | BPRMOLAR | TMOLAR | BMOLAR | GROUP |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | BROWNBAT | 2 | 3 | 1 | 1 | 3 | 3 | 3 | 3 | 4 |
| 2 | MOLE | 3 | 2 | 1 | 0 | 3 | 3 | 3 | 3 | 4 |
| 3 | SILVERHAIRBAT | 2 | 3 | 1 | 1 | 2 | 3 | 3 | 3 | 4 |
| 4 | PIGMYBAT | 2 | 3 | 1 | 1 | 2 | 2 | 3 | 3 | 4 |
| 5 | HOUSEBAT | 2 | 3 | 1 | 1 | 1 | 2 | 3 | 3 | 4 |
| 6 | REDBAT | 1 | 3 | 1 | 1 | 2 | 2 | 3 | 3 | 4 |