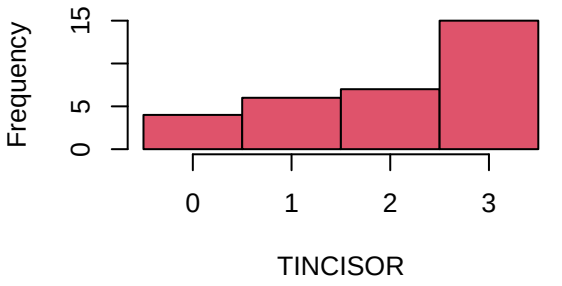
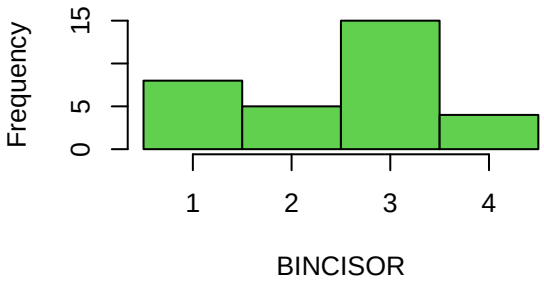
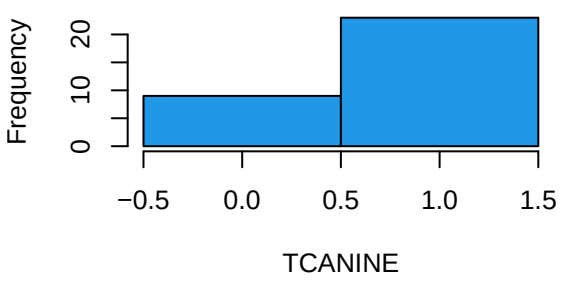
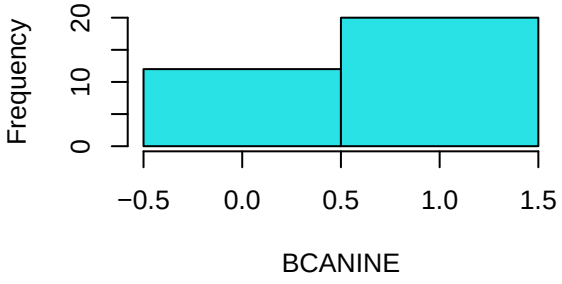
# Teeth

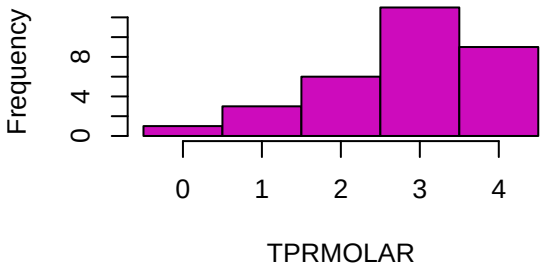
The Teeth data set has data on the number of pairs of teeth, broken down into 8 different types of tooth pairs, for 32 species of heterodont mammals (mammals with different tooth morphologies). Distributions of numbers of pairs by type are shown below.

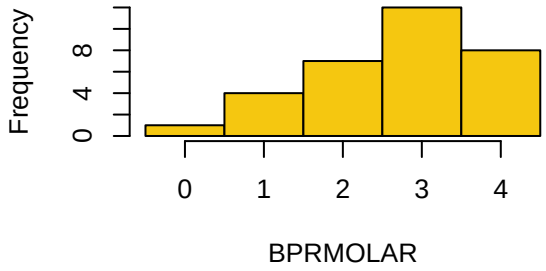


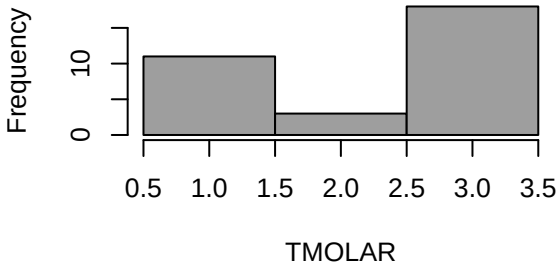



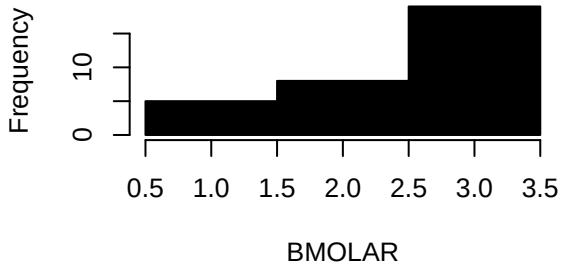






With this data set, I am particularly interested in looking for clusters of animals within the 32 species and in looking for associations between the numbers of teeth of different types.

### Principal Components

```
principal(teeth[,2:9],nfactor=5,rotate=F,scores=T)
```

```
## Specified rotation not found, rotate='none' used
```

```
## Principal Components Analysis
## Call: principal(r = teeth[, 2:9], nfactors = 5, rotate = F, scores = T)
## Standardized loadings (pattern matrix) based upon correlation matrix
##             PC1   PC2   PC3   PC4   PC5   h2     u2 com
## TINCISOR  0.78 -0.40  0.19  0.34  0.22 0.97 0.0300 2.2
## BINCISOR  0.37  0.84  0.13 -0.32  0.20 0.99 0.0088 1.9
## TCANINE   0.83  0.24  0.38  0.03 -0.23 0.95 0.0496 1.8
## BCANINE   0.85 -0.06  0.43  0.03 -0.01 0.92 0.0821 1.5
## TPRMOLAR  0.81  0.23 -0.48  0.15  0.02 0.97 0.0316 1.9
## BPRMOLAR  0.81  0.39 -0.30  0.24 -0.10 0.96 0.0371 2.0
## TMOLAR   -0.83  0.37  0.14  0.27 -0.17 0.95 0.0499 1.8
## BMOLAR   -0.76  0.45  0.18  0.37  0.16 0.97 0.0319 2.4
##
##                        PC1  PC2  PC3  PC4  PC5
## SS loadings           4.74 1.47 0.76 0.51 0.20
## Proportion Var        0.59 0.18 0.09 0.06 0.03
## Cumulative Var        0.59 0.78 0.87 0.93 0.96
## Proportion Explained  0.62 0.19 0.10 0.07 0.03
## Cumulative Proportion 0.62 0.81 0.91 0.97 1.00
##
## Mean item complexity =  1.9
## Test of the hypothesis that 5 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.02
```

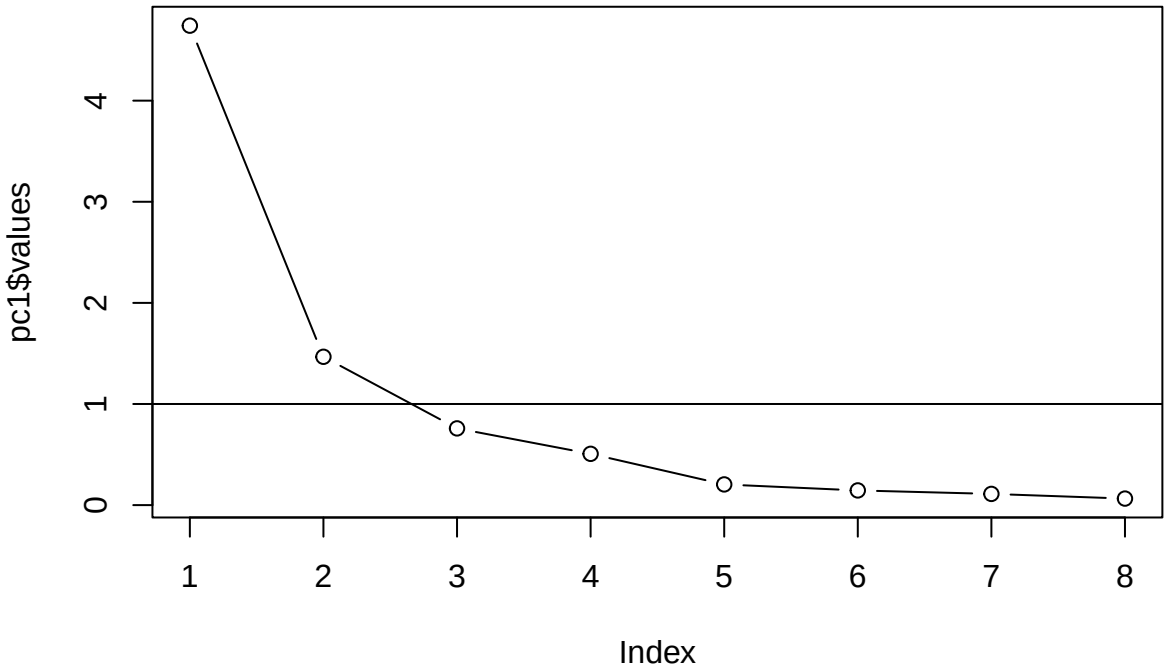
```
##  with the empirical chi square  0.69  with prob <  NA
##
## Fit based upon off diagonal values = 1
```

The first two principal components explain a combined 78% of the variance. With a third component, the percentage of variance explained increases to 87%

```
pc1 <- principal(teeth[,2:9],nfactor=5,rotate=F,scores=T)
```

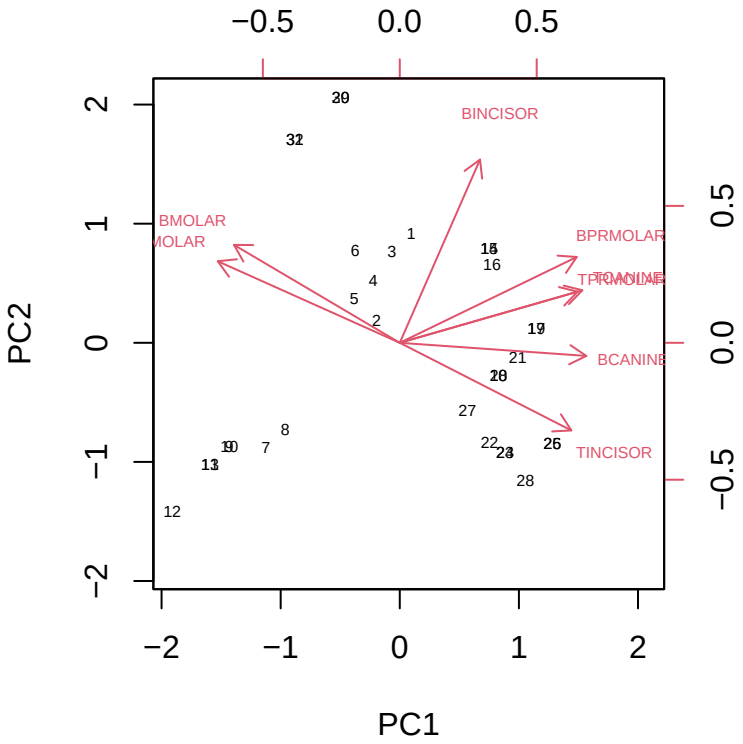
```
## Specified rotation not found, rotate='none' used
```

```
#round(pc1$loadings[,1:5],2)
plot(pc1$values, type="b")
abline(h=1)
```




Based on the skree plot, using two principal components seems appropriate.

```
biplot(pc1$scores[,1:2],pc1$loadings[,1:2], cex=0.5)
```



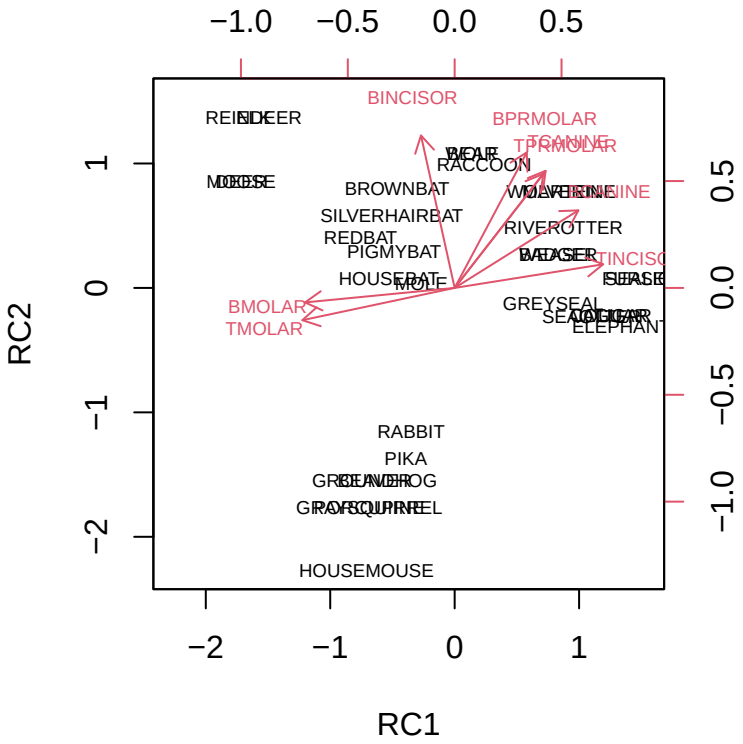

We'll rotate the principal components to see if they become more readily interpretable.

**Rotated principal Components**

```
pc2= principal(teeth[,2:9],nfactor=2,rotate="varimax",scores=T)
round(pc2$loadings[,1:2],2)
```

```
##            RC1   RC2
## TINCISOR  0.87  0.14
## BINCISOR -0.20  0.89
## TCANINE   0.53  0.68
## BCANINE   0.72  0.45
## TPRMOLAR  0.52  0.67
## BPRMOLAR  0.42  0.79
## TMOLAR   -0.89 -0.19
## BMOLAR   -0.87 -0.08
```

```
biplot(pc2$scores[,1:2],pc2$loadings[,1:2], cex=0.6,xlabs=teeth[,1])
```




```
pc2$loadings[,1:2]
```

```
##                 RC1         RC2
## TINCISOR  0.8685661  0.13819735
## BINCISOR -0.1968851  0.89167705
## TCANINE   0.5320484  0.68482970
## BCANINE   0.7234059  0.45342629
## TPRMOLAR  0.5184062  0.66675857
## BPRMOLAR  0.4218199  0.79226717
## TMOLAR   -0.8905521 -0.18874716
## BMOLAR   -0.8748296 -0.08477799
```

The loadings and the biplots show that numbers of top molars contains the same information as the number of bottom molars and likewise the number of top pre-molars lines up closely with the number of bottom pre-molars but not as closely as top pre-molars lineup with top canines. Top incisors are, on the other hand, orthogonal to bottom incisors and are strongly negatively associated with numbers of top and bottom molars.

Even with species labels shown on the plot, these principal components are somewhat difficult to comprehend. The first principal component with large negative values for top and bottom molars could perhaps be described as the extent to which teeth are located towards the front of the mouth rather than the back.

### Clusters

```
d = teeth
dst <- dist(d[,2:9], method = "euclidean")
dst2 <- dist(pc2$scores[,1:2], method = "euclidean")
fit <- hclust(dst, method="ward")
```

```
## The "ward" method has been renamed to "ward.D"; note new "ward.D2"
```

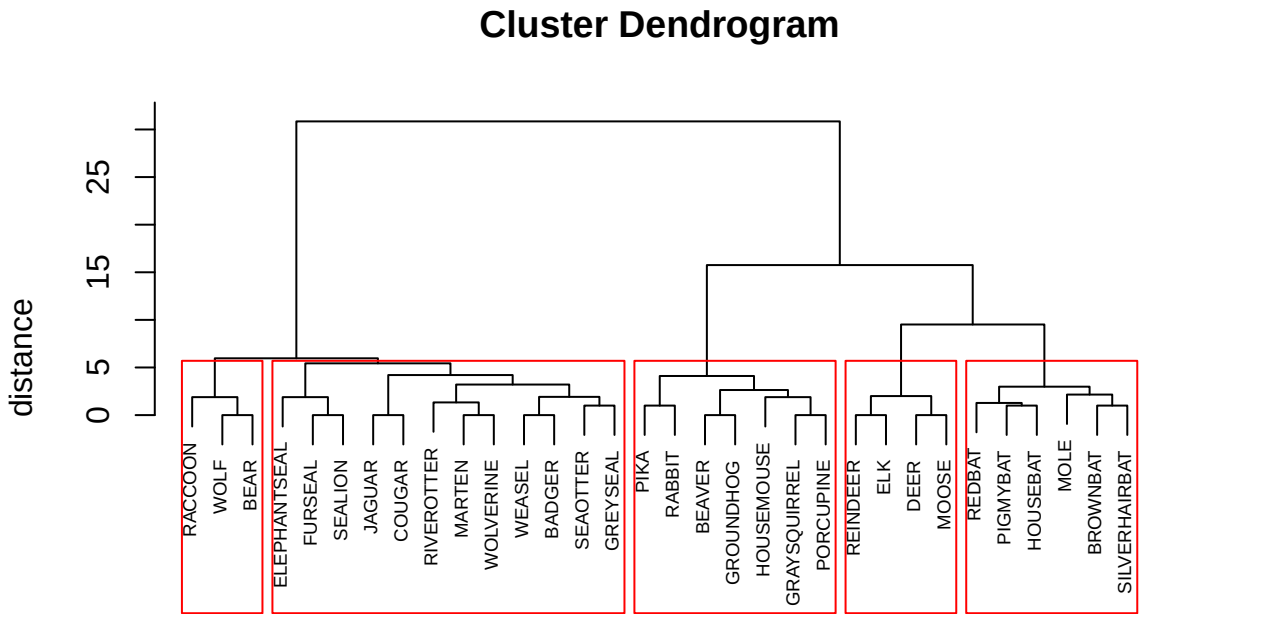
```
fit2 <- hclust(dst2, method="ward")
```

```
## The "ward" method has been renamed to "ward.D"; note new "ward.D2"
```

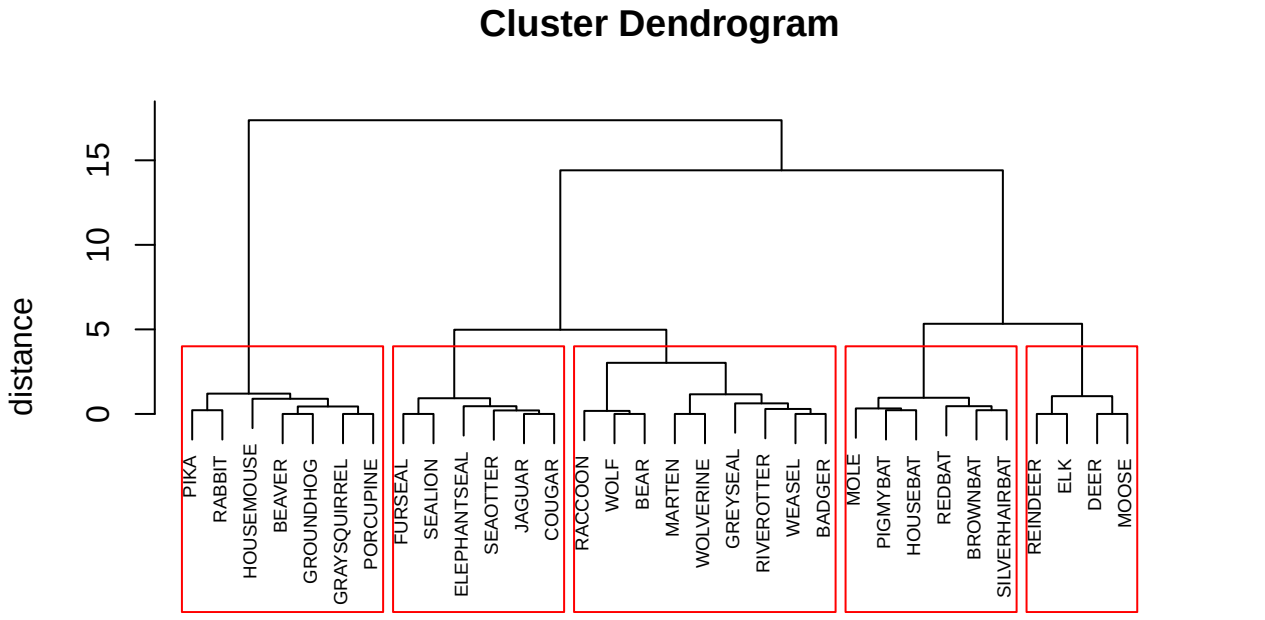
The dendrogram shown below does an impressive job of clustering animals near what we know to be similar species. There are some surprises, however, as based on their teeth, grey seals are deemed to be more similar to sea otters and even weasels and badgers than they are to other seals. Also, raccoon are deemed to be more similar to wolves and bears than to weasels, badgers and otters.

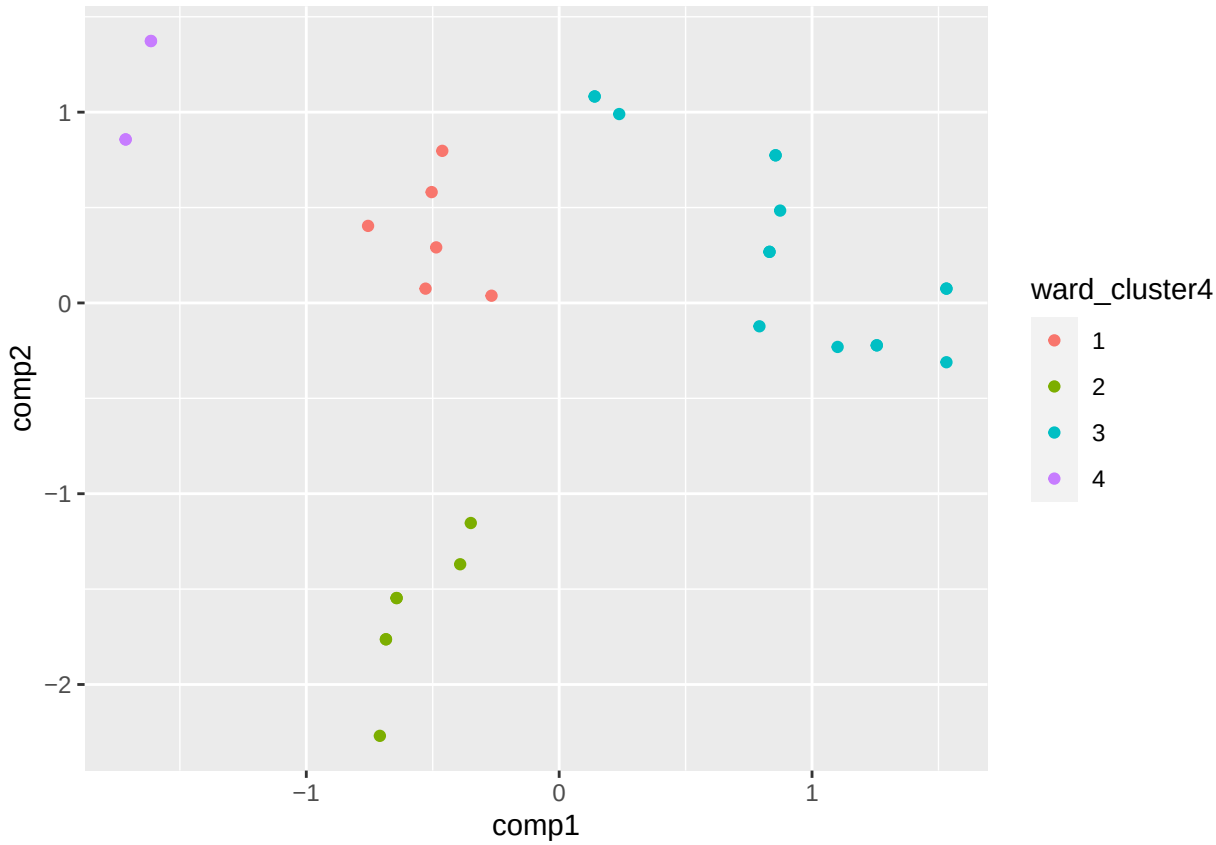
While I was tempted to split these mammals into as many as 6 clusters (and it is possible to do so without splitting up the most closely relating animals) based on the plots shown below I decided that 5 clusters was the best choice when using Ward clustering. The plots below show the mammals plotted on a graph of principal component 1 versus principal component two with the points colored by Ward clustering, first with 4 clusters, then 5 and finally with 6.

```
plot(fit, labels=d$ANIMAL, ylab='distance', cex=0.6, sub="", xlab="")
rect.hclust(fit, k=5, border="red")
```

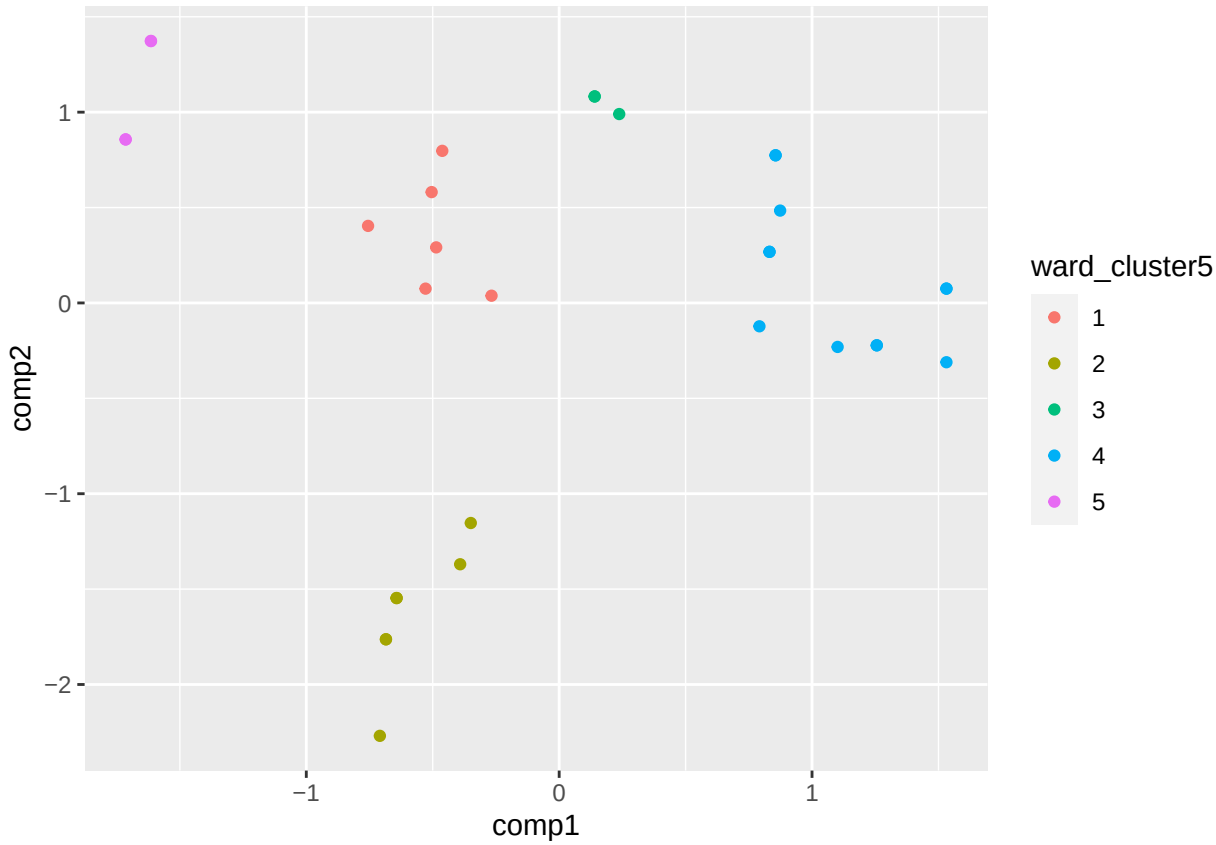



```
plot(fit2, labels=d$ANIMAL, ylab='distance', cex=0.6, sub="", xlab="")
rect.hclust(fit2, k=5, border="red")
```

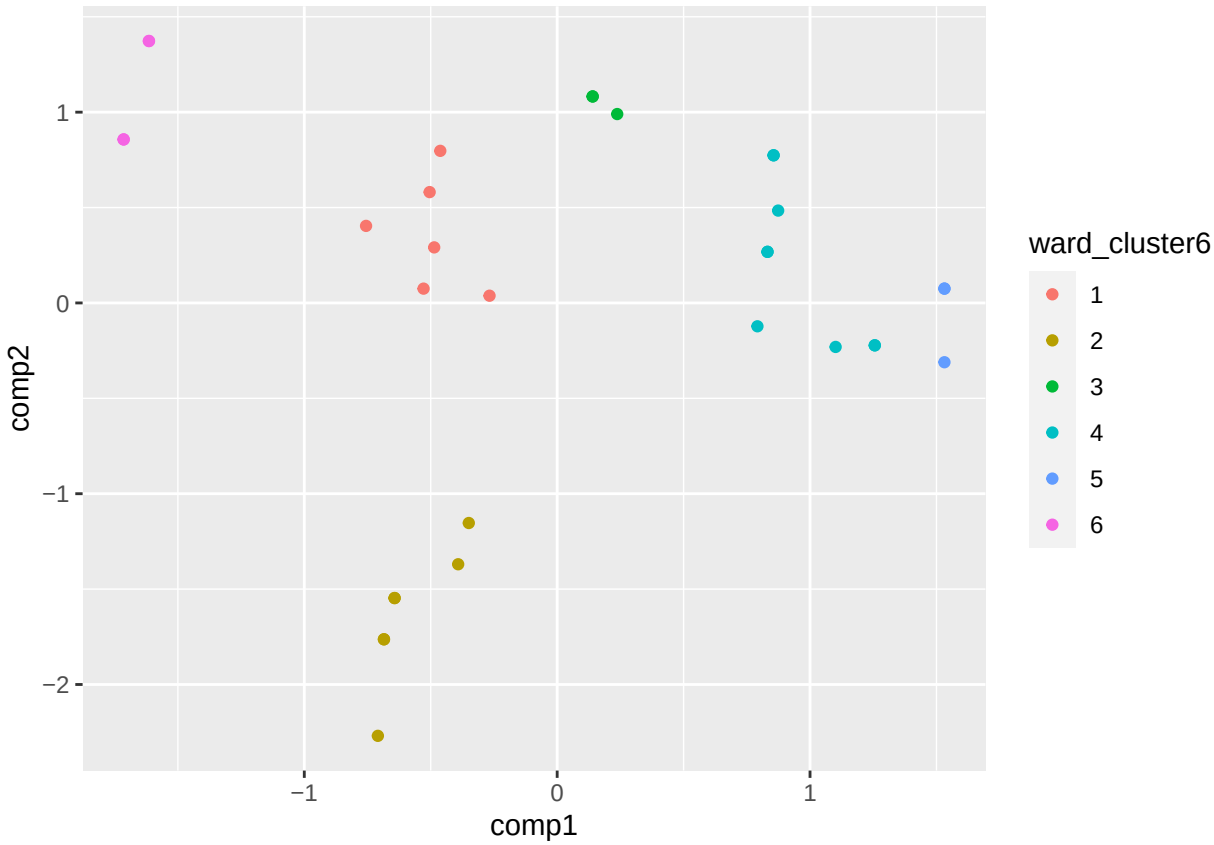





```
ggplot(teeth_plus, aes(x = comp1, y = comp2, color = ward_cluster5)) + geom_point()
```

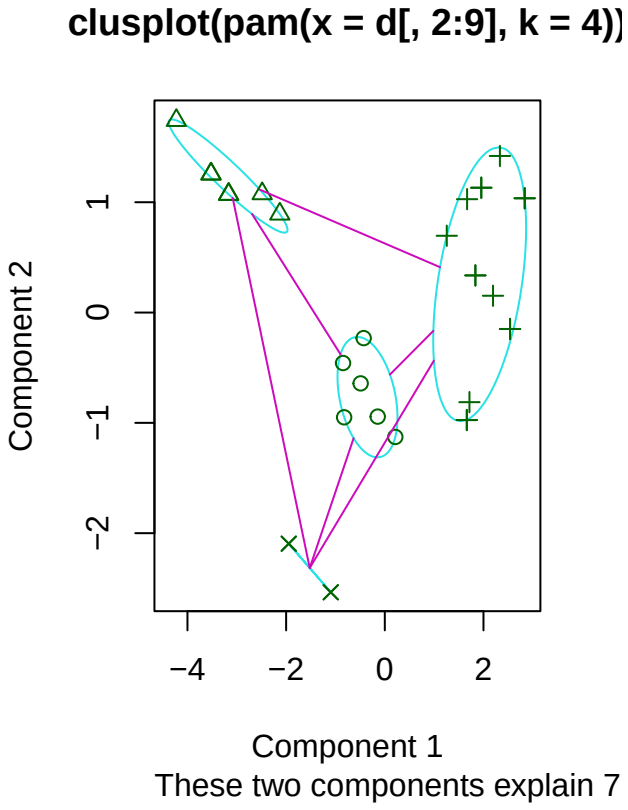


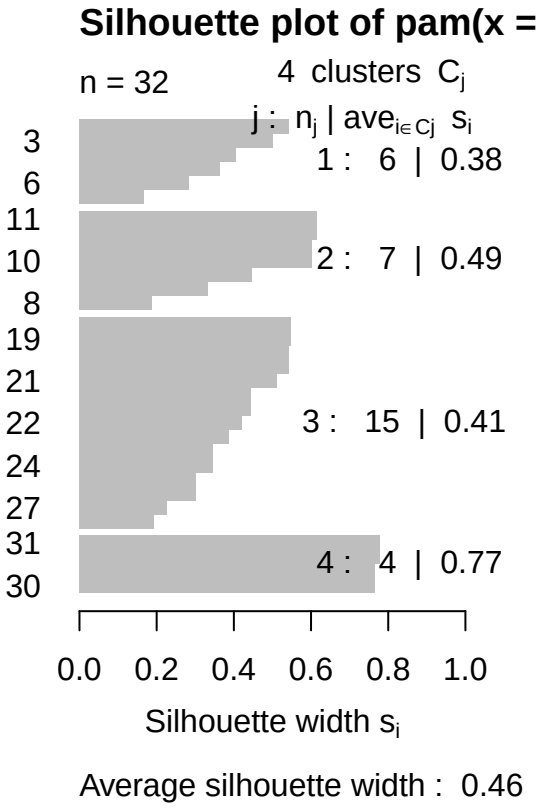

```
ggplot(teeth_plus, aes(x = comp1, y = comp2, color = ward_cluster6)) + geom_point()
```



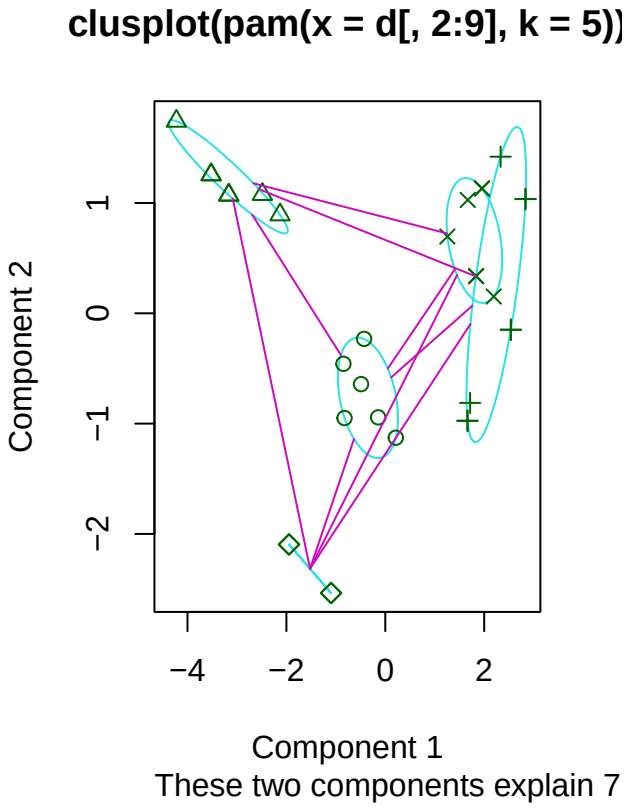

Using clustering around mediods, shown below, 4 clusters appears to be the more natural choice. In the graph with 5 clusters, two of the clusters overlap and in the 6 cluster plot there are a least a couple of species that appear as if they could have been placed in a different cluster nearly as easily.

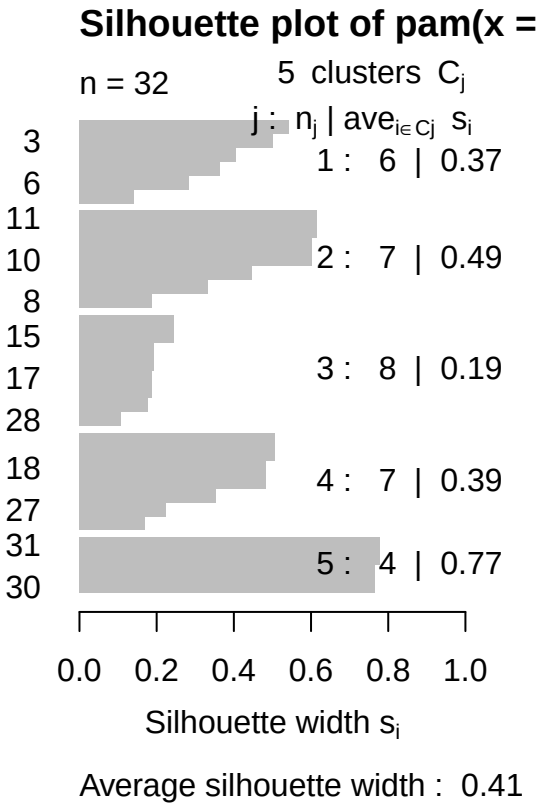
```
par(mfrow=c(1,2))
fit.pam <- pam(d[,2:9],4)
plot(fit.pam)
```



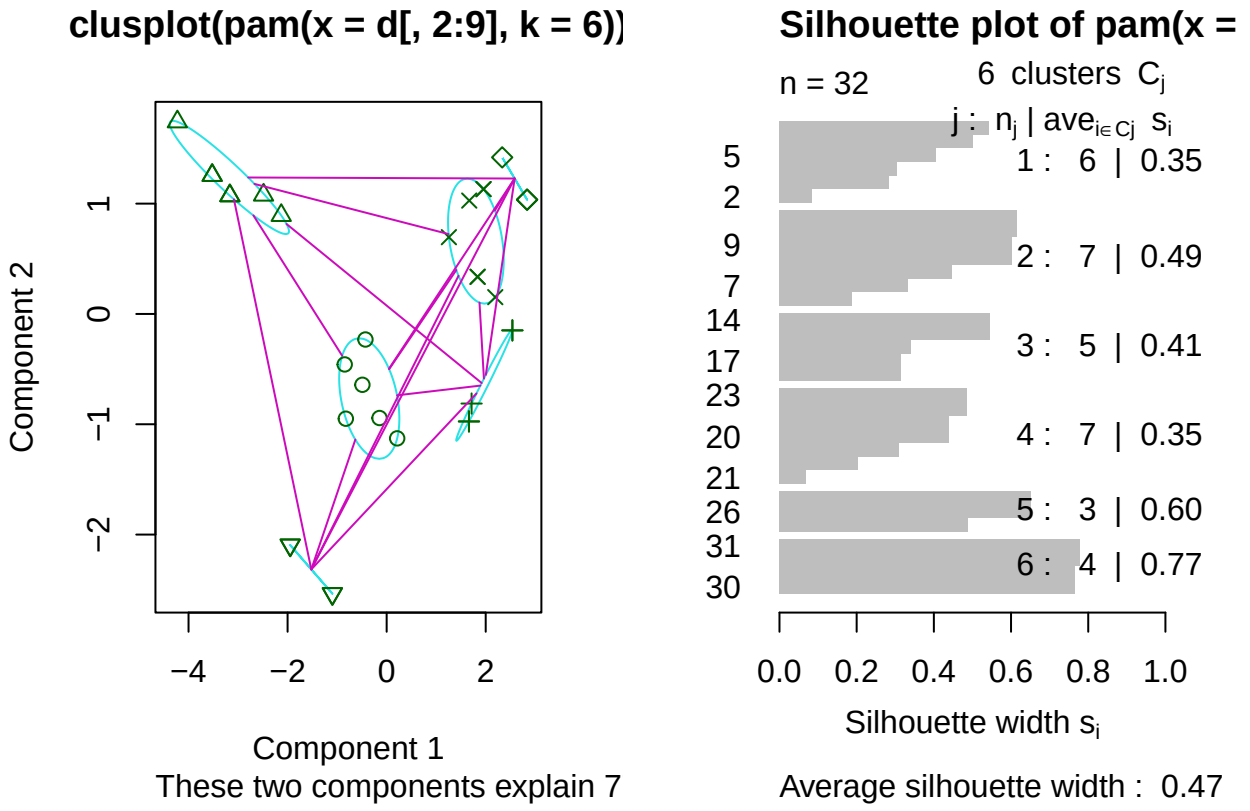




```
fit.pam <- pam(d[,2:9],5)
plot(fit.pam)
```



```
fit.pam <- pam(d[,2:9],6)
plot(fit.pam)
```




**Examinging the Clusters**

Finally, let's look at the distribution of number of teeth of each type for each cluster. Here, again, are the clusters along with my best attempt to name them.

Cluster 1 (Bats):

```
##           ANIMAL
## 1       BROWNBAT
## 2           MOLE
## 3 SILVERHAIRBAT
## 4       PIGMYBAT
## 5       HOUSEBAT
## 6         REDBAT
```

Cluster 2 (Gnawing Animals):

```
##           ANIMAL
## 7           PIKA
## 8         RABBIT
## 9         BEAVER
## 10    GROUNDHOG
## 11 GRAYSQUIRREL
## 12   HOUSEMOUSE
## 13    PORCUPINE
```

Cluster 3 (Misfits):

```
##     ANIMAL
## 14    WOLF
## 15    BEAR
## 16 RACCOON
```

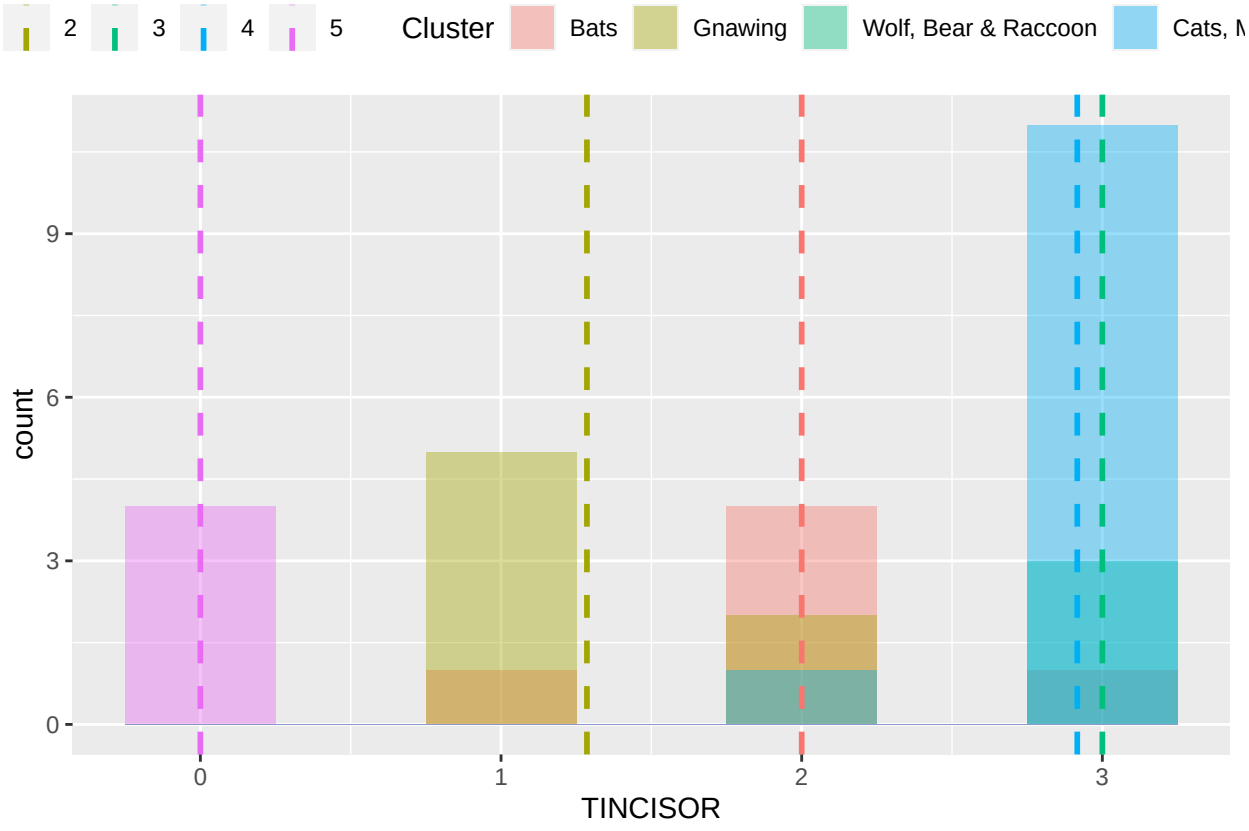
Cluster 4 (Cats, Mustelids and Seals):

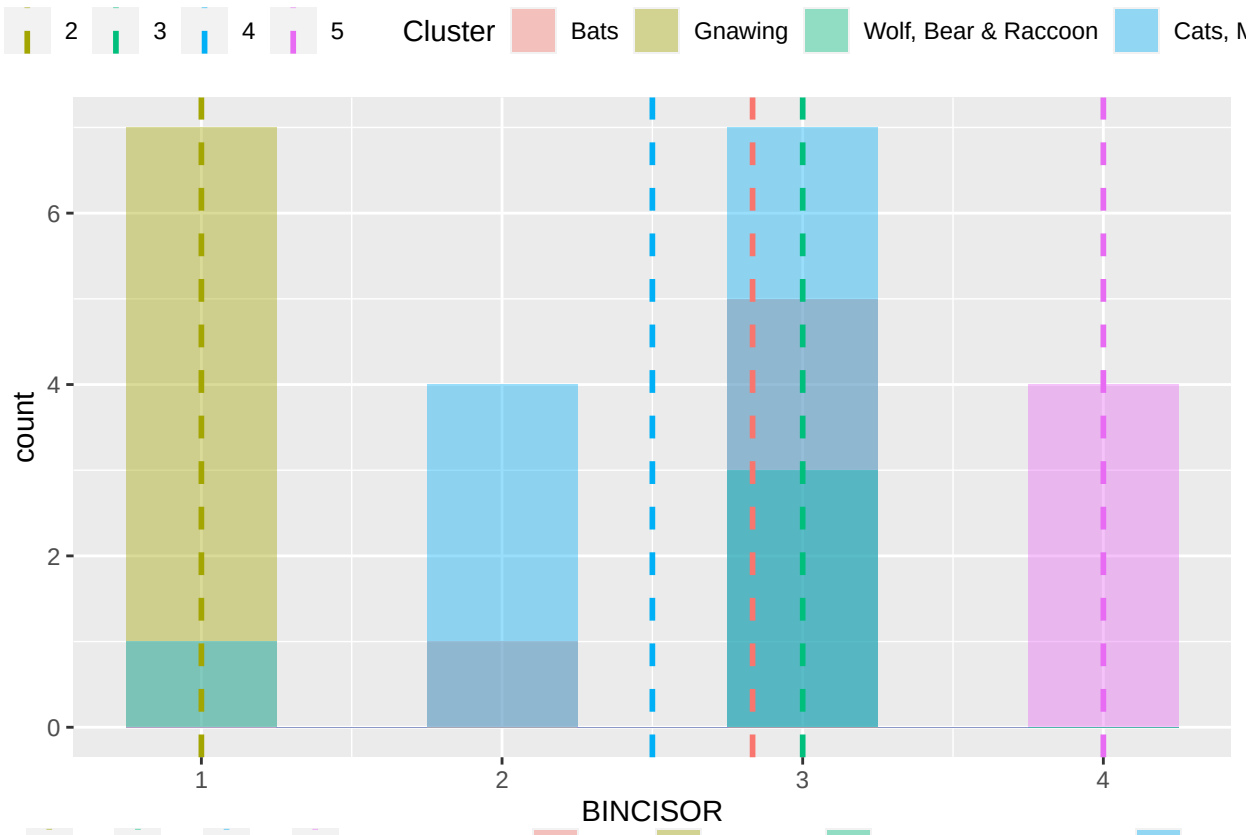
```
##          ANIMAL
## 17       MARTEN
## 18       WEASEL
## 19    WOLVERINE
## 20       BADGER
## 21   RIVEROTTER
## 22     SEAOTTER
## 23       JAGUAR
## 24       COUGAR
## 25      FURSEAL
## 26      SEALION
## 27     GREYSEAL
## 28 ELEPHANTSEAL
```

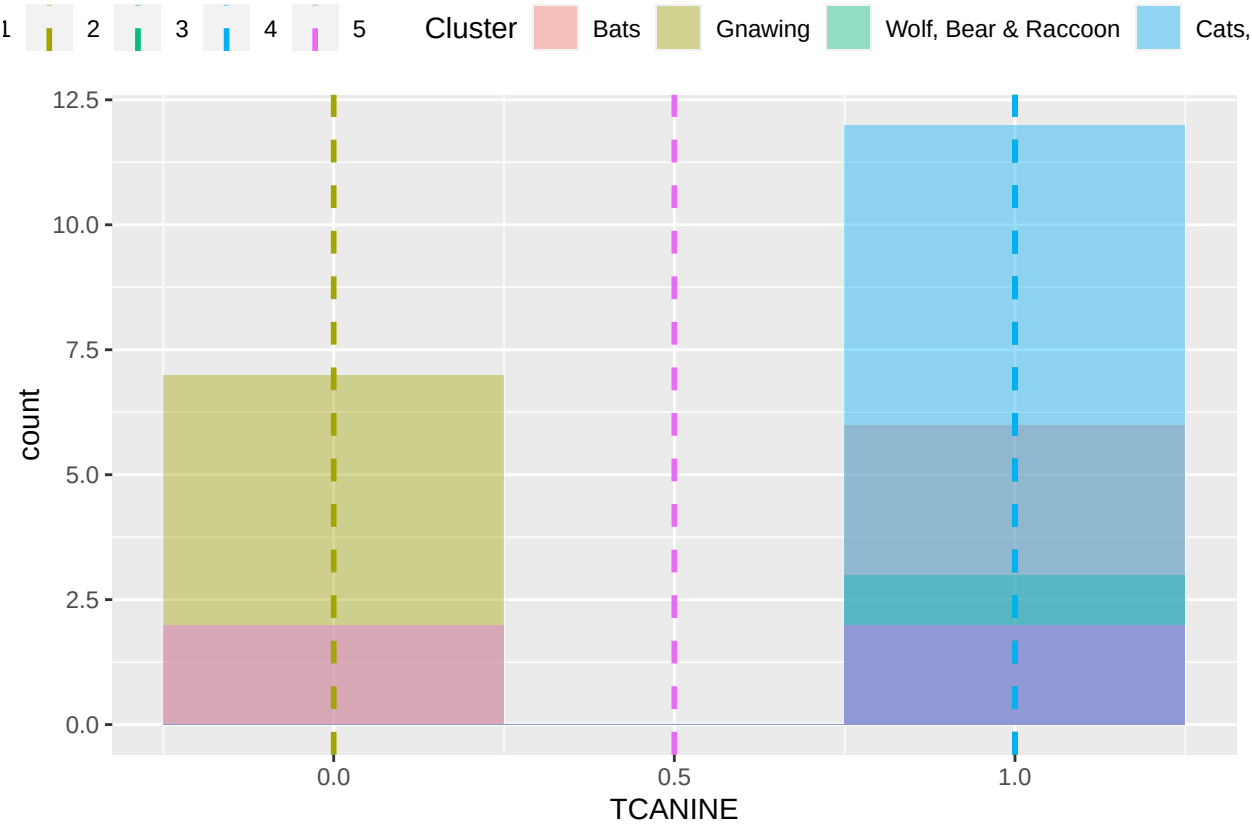
Cluster 5 (Grazing Animals):

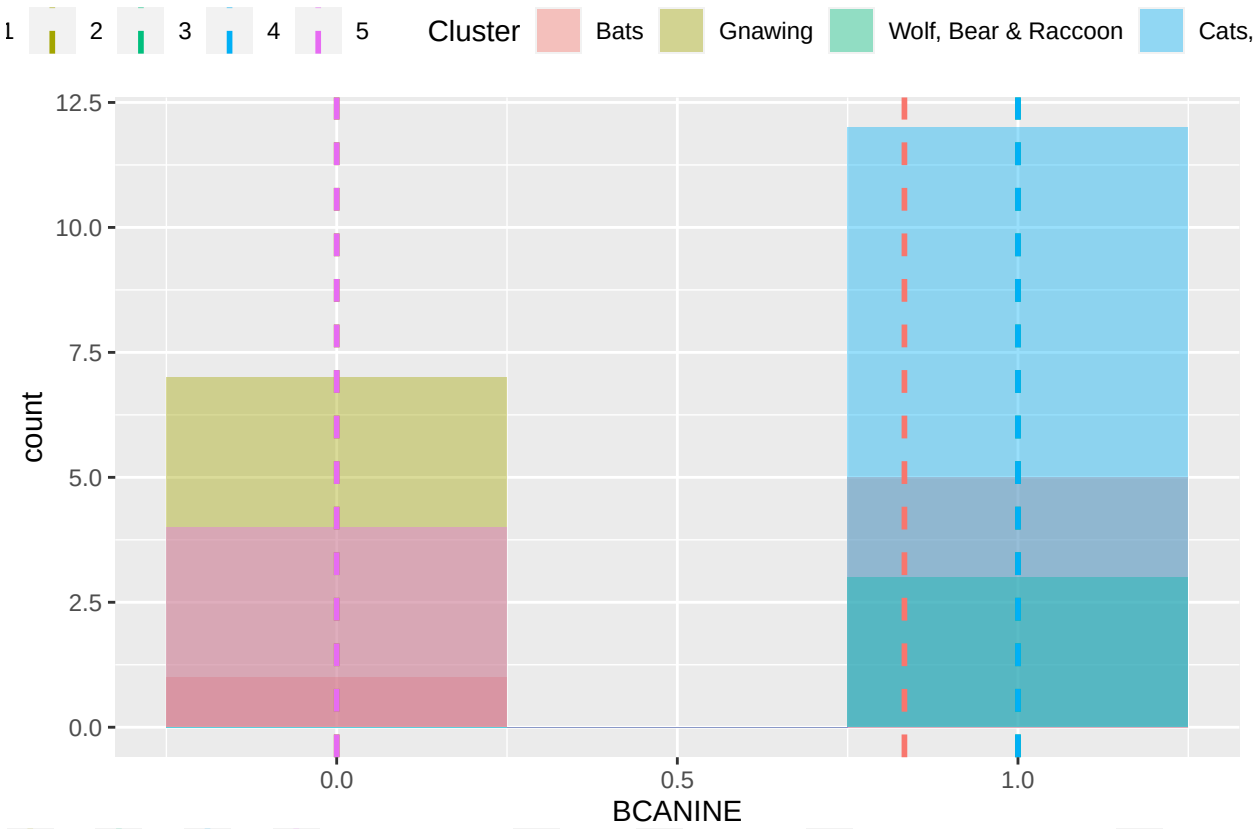
```
##      ANIMAL
## 29 REINDEER
## 30      ELK
## 31     DEER
## 32    MOOSE
```

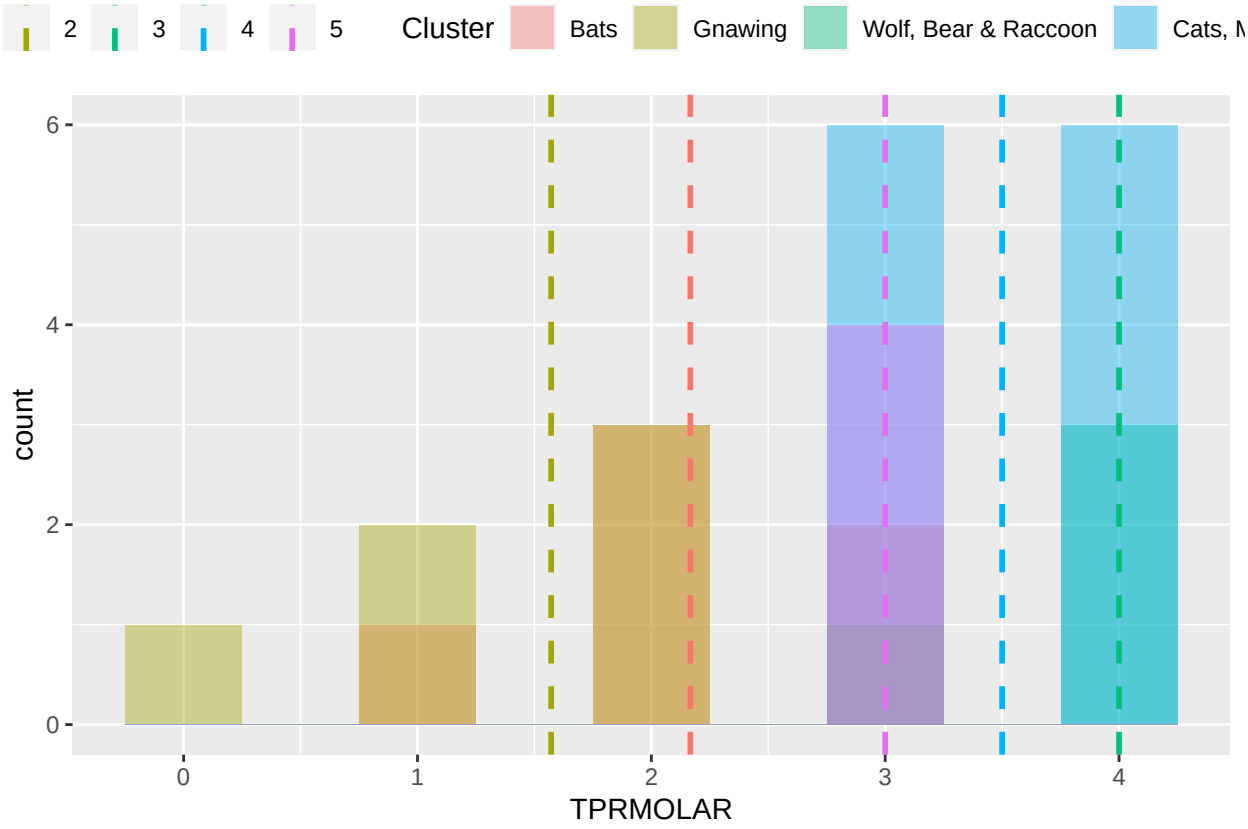
And here are the distributions of numbers of teeth pairs by type:

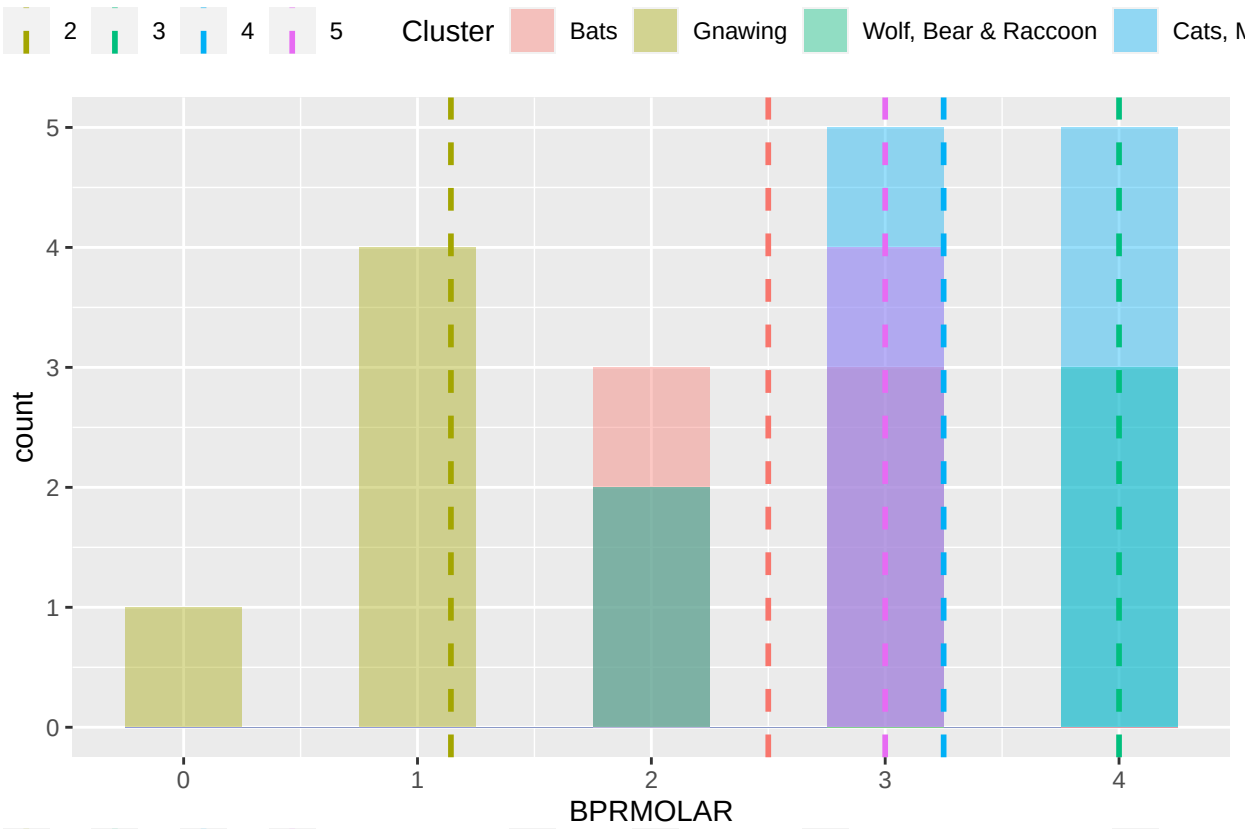


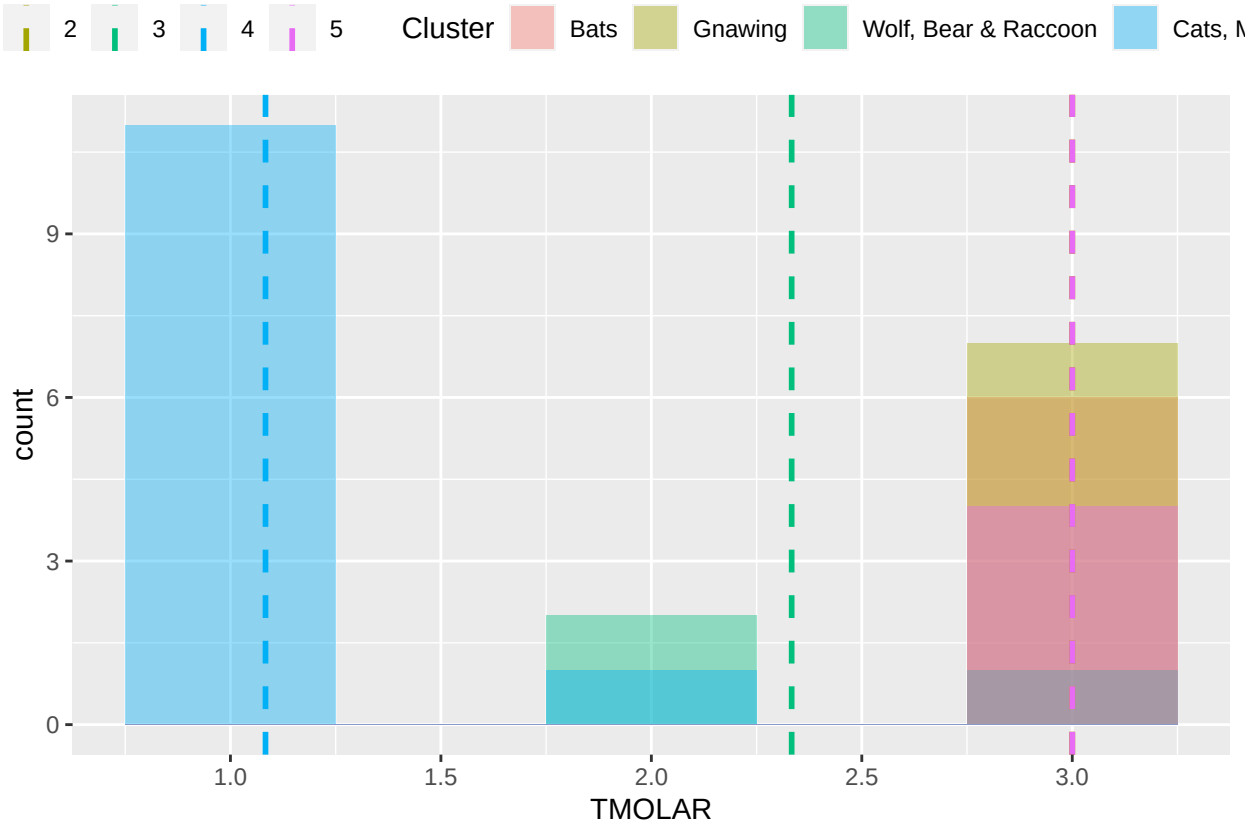

2
3
4
5
Cluster
Bats
Gnawing
Wolf, Bear & Raccoon
Cats, M
count
6
4
2
0
1
2
3
4
BINCISOR

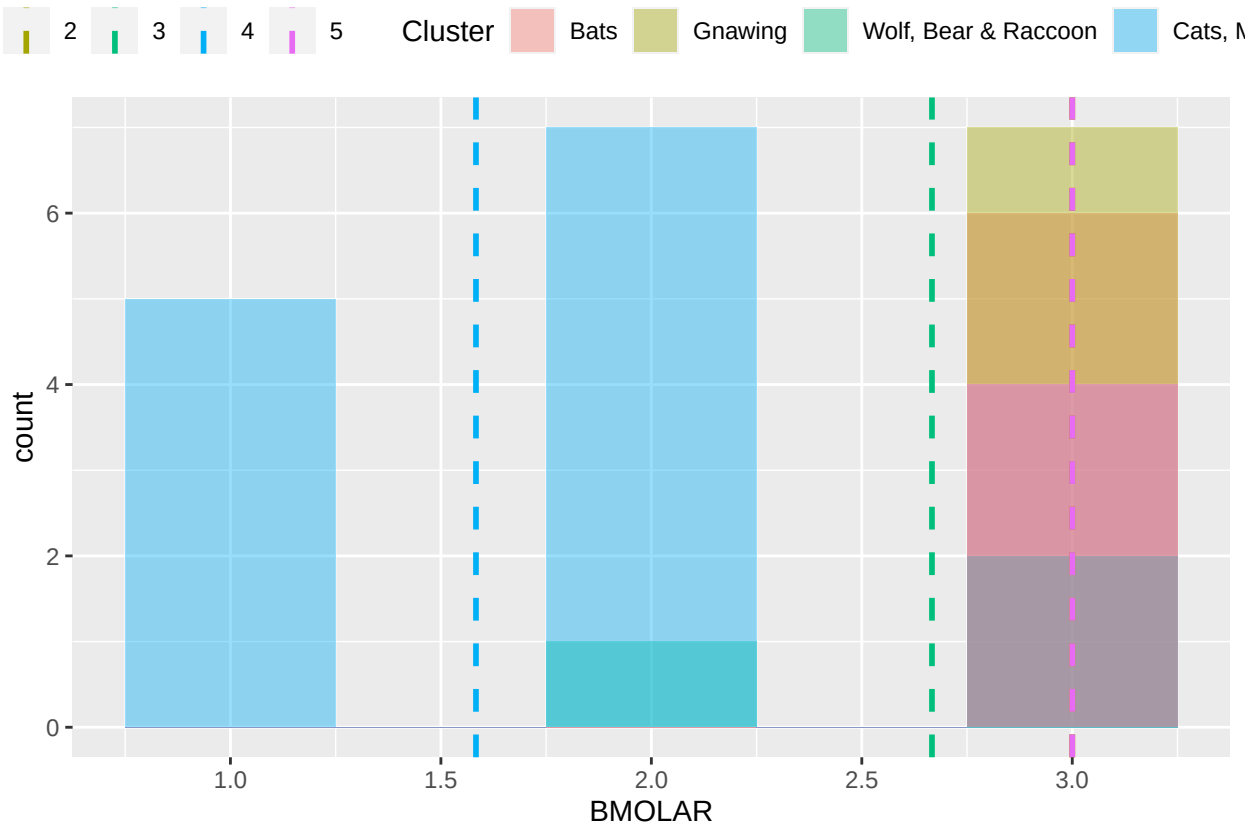

1
2
3
4
5
Cluster
Bats
Gnawing
Wolf, Bear & Raccoon
Cats,
count
12.5
10.0
7.5
5.0
2.5
0.0
0.0
0.5
1.0
TCANINE


1
2
3
4
5
Cluster
Bats
Gnawing
Wolf, Bear & Raccoon
Cats,
12.5
10.0
7.5
5.0
2.5
0.0
count
0.0
0.5
1.0
BCANINE


2
3
4
5
Cluster
Bats
Gnawing
Wolf, Bear & Raccoon
Cats, M
6
4
2
0
count
0
1
2
3
4
TPRMOLAR


2
3
4
5
Cluster
Bats
Gnawing
Wolf, Bear & Raccoon
Cats, M
count
BPRMOLAR
0
1
2
3
4
5


2
3
4
5
Cluster
Bats
Gnawing
Wolf, Bear & Raccoon
Cats, M
count
TMOLAR
1.0
1.5
2.0
2.5
3.0
0
3
6
9


2
3
4
5
Cluster
Bats
Gnawing
Wolf, Bear & Raccoon
Cats, M
count
6
4
2
0
1.0
1.5
2.0
2.5
3.0
BMOLAR