

Clustering in R

There is a nice post at

<https://stackoverflow.com/questions/15376075/cluster-analysis-in-r-determine-the-optimal-number-of-clusters> that provide a number of examples to illustrate how to determine number of clusters.

Below are two examples:

```
# prepare data from a dataframe # var1 includes one group # var2 includes
second group # var3 includes all the values used in clustering # transpose the
dataframe data to matrix kmdata <- acast(my.data, var1 ~ var2, value.var='var3')
# kmeans my.cluster <- cascadeKM(kmdata, inf.gr = 1, sup.gr = nrow(kmdata)-1)
plot(my.cluster, sortg = TRUE, grpmts.plot = TRUE) calinski.best <-
as.numeric(which.max(my.cluster$results[2,])) cat("Calinski criterion optimal
number of clusters:", calinski.best, "\n") # sum of square error wss <-
(nrow(kmdata)-1)*sum(apply(kmdata,2,var)) for (i in 2:(nrow(kmdata)-1)) wss[i] <-
sum(kmeans(kmdata, centers=i)$withinss) plot(1:(nrow(kmdata)-1), wss,
type="b", xlab="Number of Clusters", ylab="Within groups sum of squares")
my.cluster2 <- kmeans(kmdata, 4)
```

Here is another nice article talking about clustering (

<http://www.sthda.com/english/wiki/cluster-analysis-in-r-unsupervised-machine-learning>)