# Question a

# Aggregate across situations (the second dimension is for situations)

person\_behavior\_aggregated <- apply(anger$data, MARGIN =3, FUN = rowSums)

dim(person\_behavior\_aggregated)

# We have 101 persons and 8 behaviors, the result is a 101 x 8 matrix

EuclideanDistance <- dist(person\_behavior\_aggregated, method = "euclidean",

diag = TRUE, upper = TRUE)

# hierarchical clustering Ward bimodal data on squared Euclidean distance

hiclust\_ward<- hclust(EuclideanDistance, "ward.D2")

par(pty="s")

plot(hiclust\_ward,hang=-1)

#Save the cluster membership variable of the 2-cluster solution

clusters <- cutree(hiclust\_ward, k = 2)

nclust <- 2

#centroid

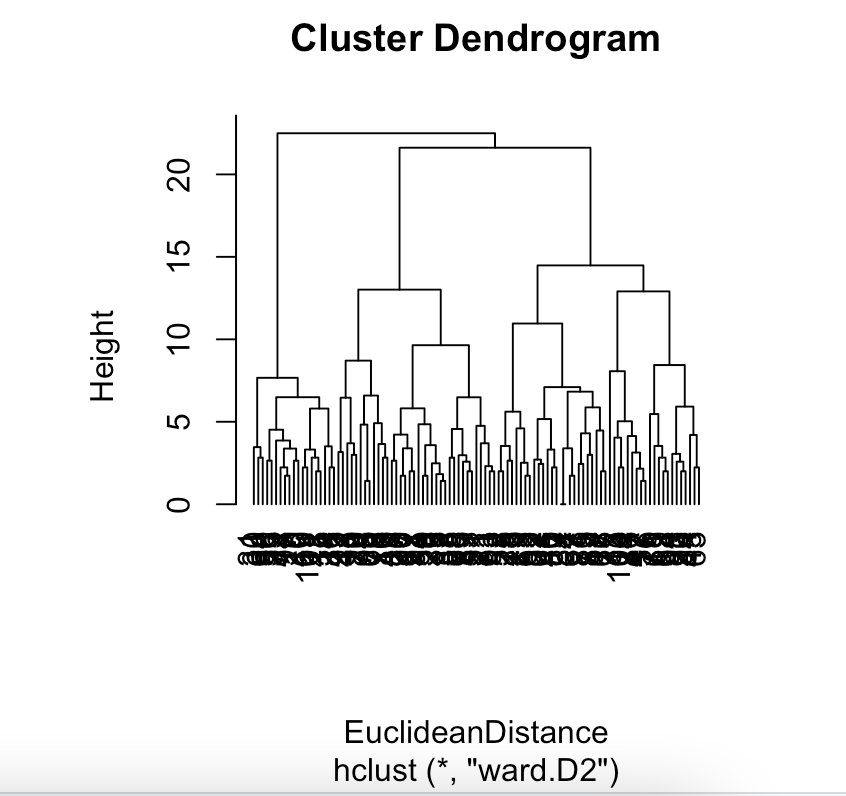
stat<-describeBy(person\_behavior\_aggregated, clusters, mat=TRUE)

hcenter <- matrix(stat[,5],nrow=nclust)

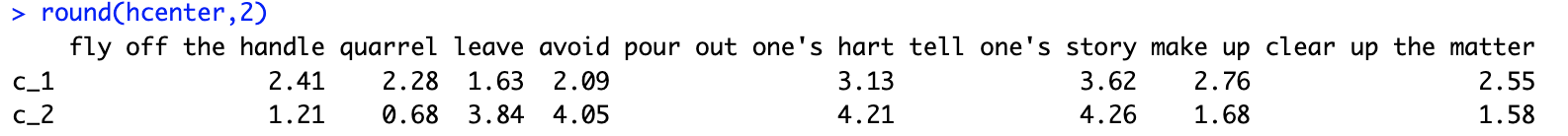
rownames(hcenter) <- paste("c\_",rep(1:nclust),sep="")

colnames(hcenter) <- c(colnames(anger$freq2))

round(hcenter,2)



From dendrogram, we can say that Ward’s method fails to capture the difference in the true modality of the two clusters.

Cluster 2 seems to have higher frequencies for behaviors related to leaving, avoiding, emotional sharing compared to Cluster 1. On the other hand, Cluster 1 generally has lower frequencies for these behaviors and higher frequencies for behaviors related to fighting and making up.

# Question B

data\_with\_clusters <- data.frame(person\_behavior\_aggregated, cluster = clusters)

profile\_vectors <- aggregate(. ~ cluster, data = data\_with\_clusters, sum)

profile\_vectors <- profile\_vectors[, -1]

# Define the new column names

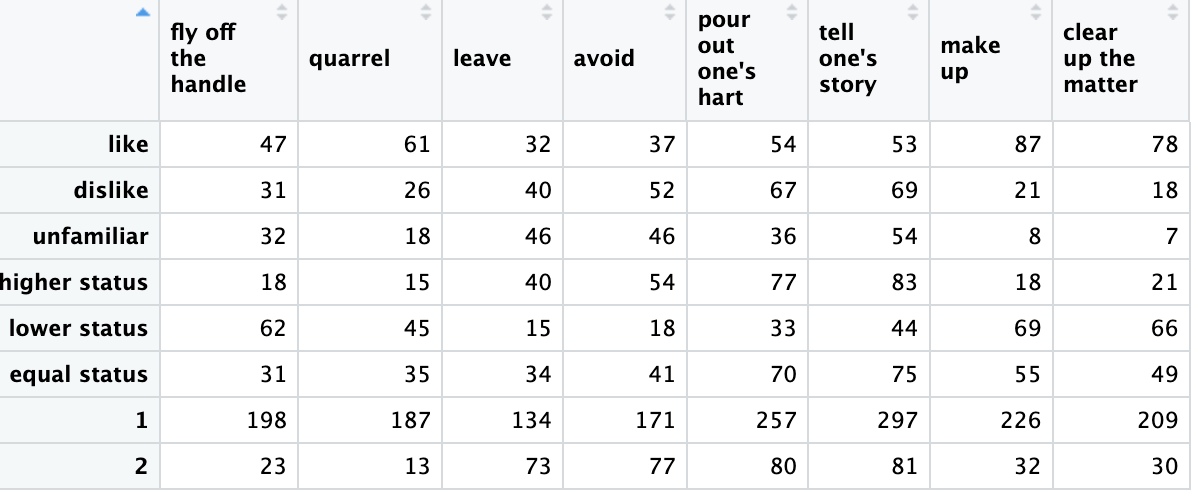
new\_column\_names <- c(

"fly off the handle","quarrel", "leave","avoid","pour out one's hart","tell one's story","make up","clear up the matter")

# Assign the new column names to 'profile\_vectors'

colnames(profile\_vectors) <- new\_column\_names

final\_freq1 <- rbind(anger$freq1, profile\_vectors)



# Question C

#H0: bahabior and situations are statistically independent

#if the Pearson-Chi square test indicates that Xand Y are statistically dependent, it is meaningful to use CA to further study the nature of the relation between Xand Y.

> chisq.test(final\_freq1)

Pearson's Chi-squared test

data: final\_freq1

X-squared = 431.48, df = 49, p-value < 2.2e-16

ca.out <- ca(final\_freq1)

summary(ca.out)

par(pty="s", cex=0.9)

plot(

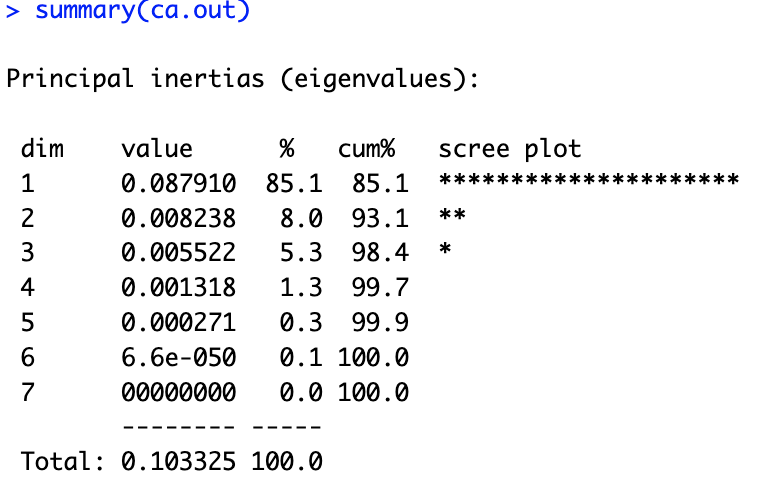
ca.out,

mass = TRUE,

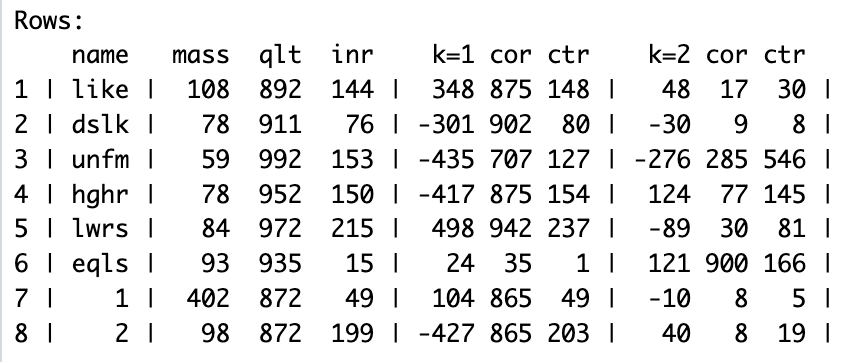
arrows = c(TRUE, FALSE),

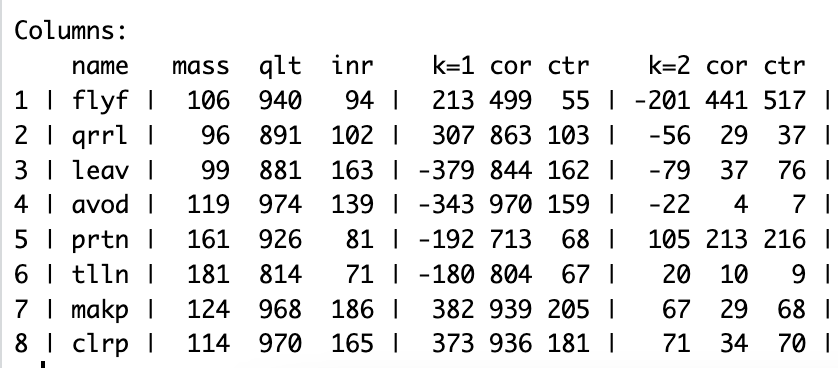
)

p-value is small enough to reject null it is meaningful to use CA to further study the nature of the relation between Xand Y.



The first three dimensions contribute to 98.4% of total inertia. This means that a three dimensions solution can well represent the dependencies between the rows and the columns of the table.

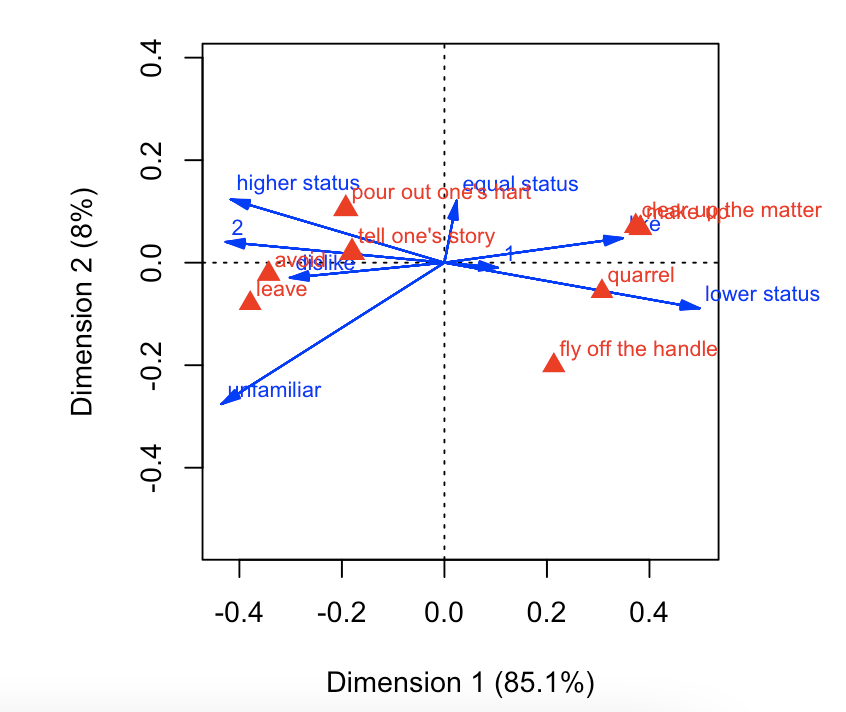




The squared correlations indicate that the first dimension in row explains most of the inertia for all situations except eqls, which was largely explained by the second dimension. We can see that the second dimension also explains some of the Unfm

We can also see that both dimensions contribute equally to the inertia of flyf. While first dimension explains most of inertia in all other behaviors.

All total quality values are above 800, which means that for all row and column points, inertia is well explained by the two dimentions. This is no surprise because the first two dimensions account for 93.1% of total inertia.



The plot shows that quarrel and fly off the handle are selected more than average in lower status.

Make up and clean up the matter are more selected if they like the person.

Avoid and leave are are selected more than average if dislike the person.

People tend to pur out their heart and tell the story if the person has higher status.