# Question a

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

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

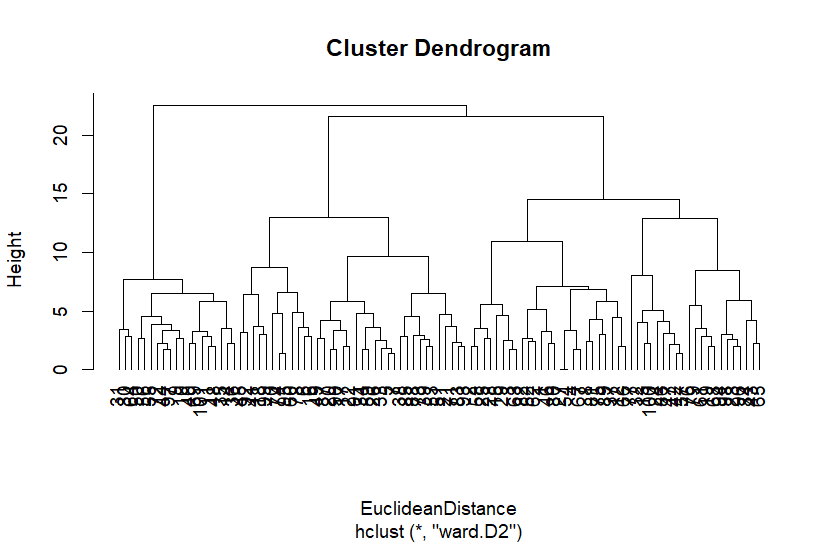
> # compute squared Euclidean distances

> EuclideanDistance <- dist(person\_behavior\_aggregated, method = "euclidean", diag = TRUE, upper = TRUE)

> # cluster 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)

> clusters

> #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)

[,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]

c\_1 2.41 2.28 1.63 2.09 3.13 3.62 2.76 2.55

c\_2 1.21 0.68 3.84 4.05 4.21 4.26 1.68 1.58

From the dendrogram, we can say that Ward’s method fails to capture the difference in the true modality of the two clusters. Maybe 3 clusters would be a better idea based on the output.

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

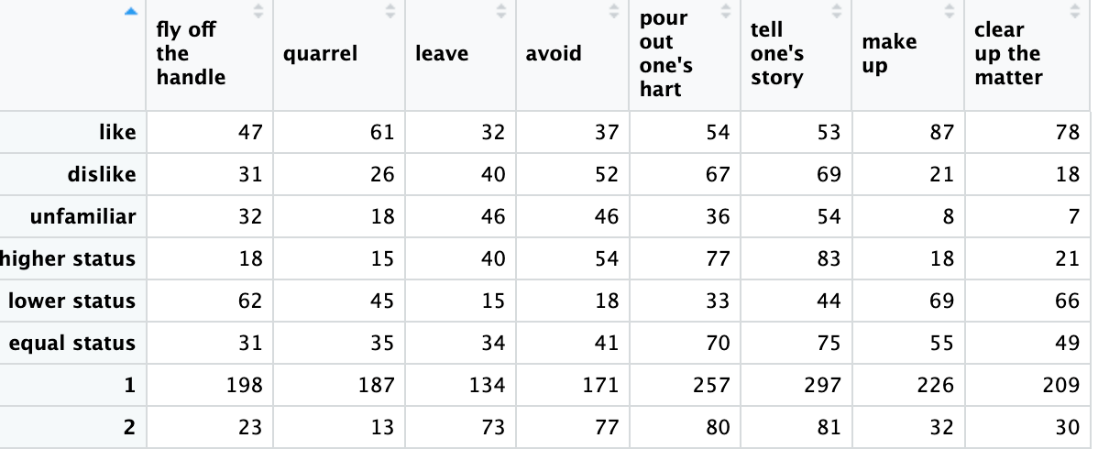
new\_column\_names <- c(

"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

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.

> #p-value is small enough to reject null

> ca.out <- ca(final\_freq1)

> summary(ca.out)

Principal inertias (eigenvalues):

dim value % cum% scree plot

1 0.087910 85.1 85.1 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

2 0.008238 8.0 93.1 \*\*

3 0.005522 5.3 98.4 \*

4 0.001318 1.3 99.7

5 0.000271 0.3 99.9

6 6.6e-050 0.1 100.0

7 00000000 0.0 100.0

-------- -----

Total: 0.103325 100.0

Rows:

name mass qlt inr k=1 cor ctr k=2 cor ctr

1 | like | 108 892 144 | 348 875 148 | 48 17 30 |

2 | dslk | 78 911 76 | -301 902 80 | -30 9 8 |

3 | unfm | 59 992 153 | -435 707 127 | -276 285 546 |

4 | hghr | 78 952 150 | -417 875 154 | 124 77 145 |

5 | lwrs | 84 972 215 | 498 942 237 | -89 30 81 |

6 | eqls | 93 935 15 | 24 35 1 | 121 900 166 |

7 | 1 | 402 872 49 | 104 865 49 | -10 8 5 |

8 | 2 | 98 872 199 | -427 865 203 | 40 8 19 |

Columns:

name mass qlt inr k=1 cor ctr k=2 cor ctr

1 | flyf | 106 940 94 | 213 499 55 | -201 441 517 |

2 | qrrl | 96 891 102 | 307 863 103 | -56 29 37 |

3 | leav | 99 881 163 | -379 844 162 | -79 37 76 |

4 | avod | 119 974 139 | -343 970 159 | -22 4 7 |

5 | prtn | 161 926 81 | -192 713 68 | 105 213 216 |

6 | tlln | 181 814 71 | -180 804 67 | 20 10 9 |

7 | makp | 124 968 186 | 382 939 205 | 67 29 68 |

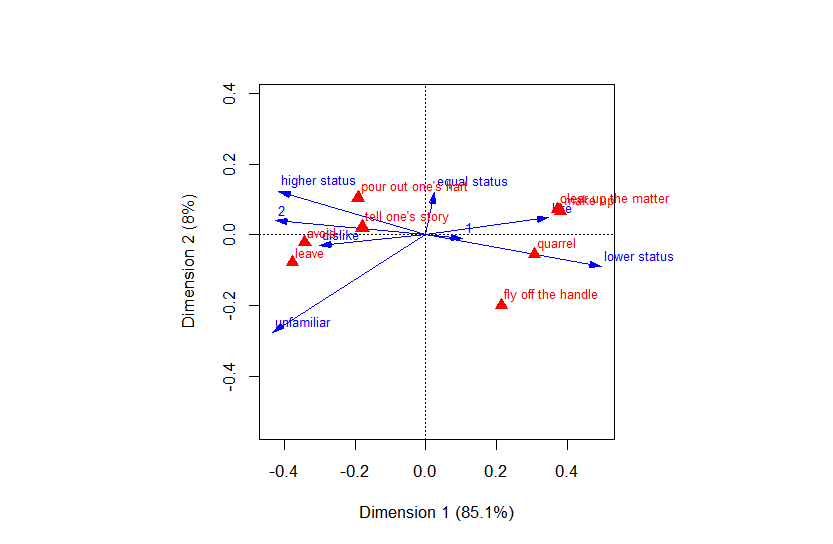
8 | clrp | 114 970 165 | 373 936 181 | 71 34 70 |

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 put out their heart and tell the story if the person has higher status.