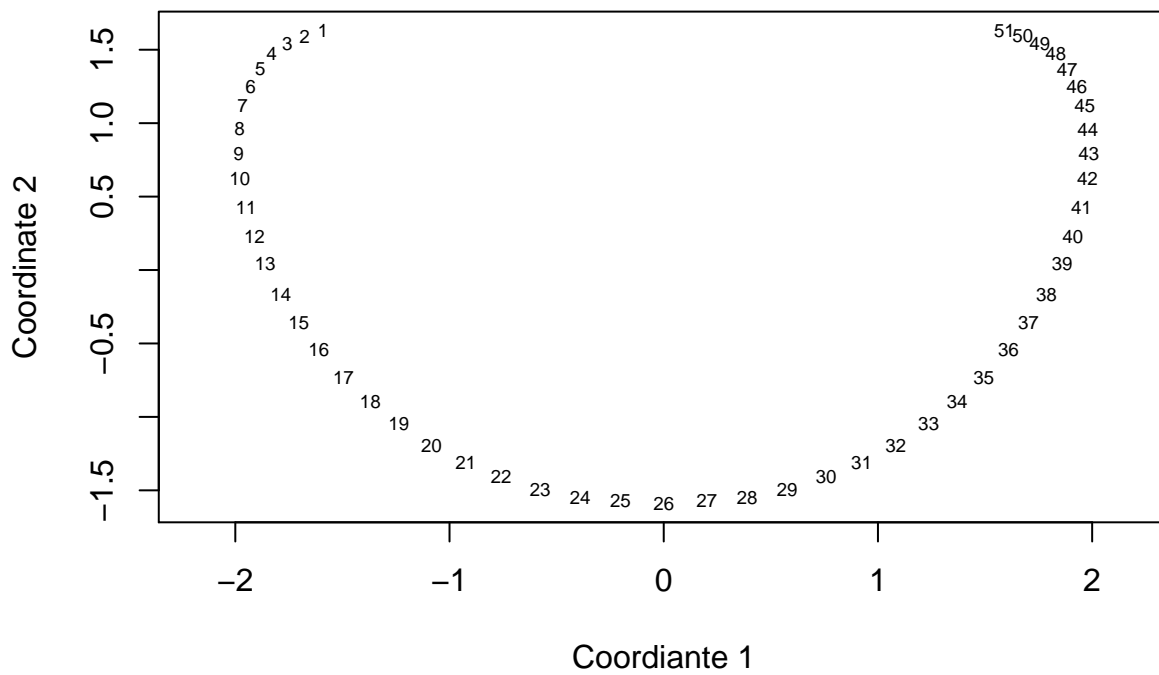


Homework 4

Cody Frisby

3/17/2017

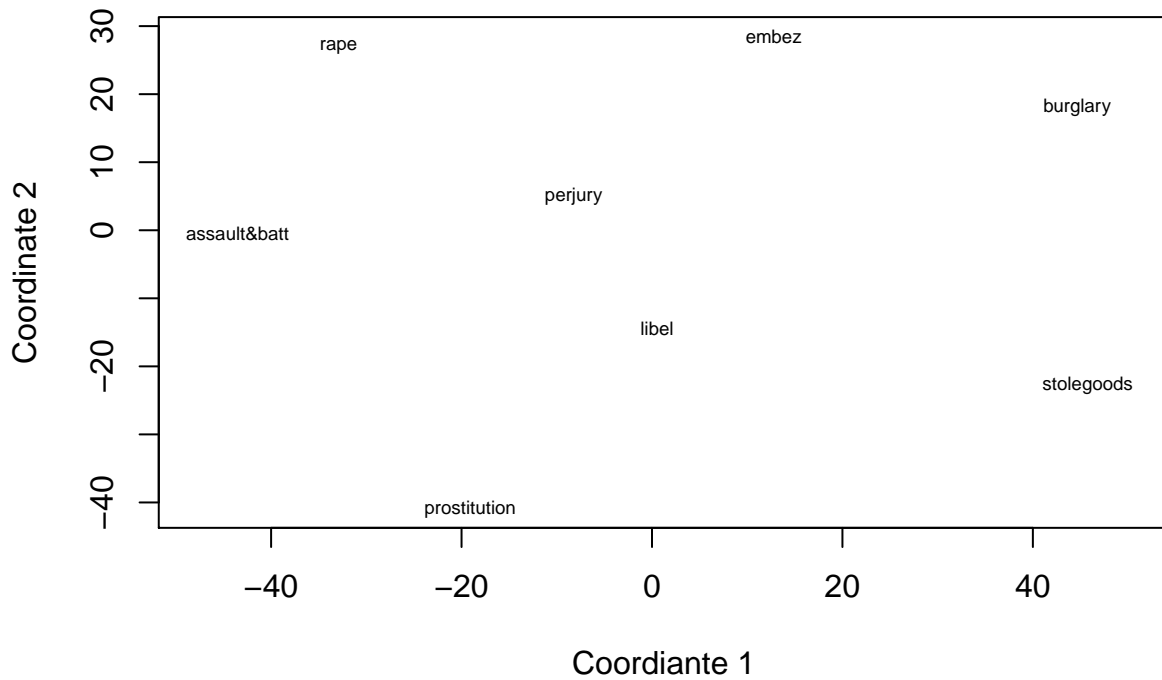
4.1



It looks like a necklace.

4.2

A plot of the two-dimensional solution is displayed here.

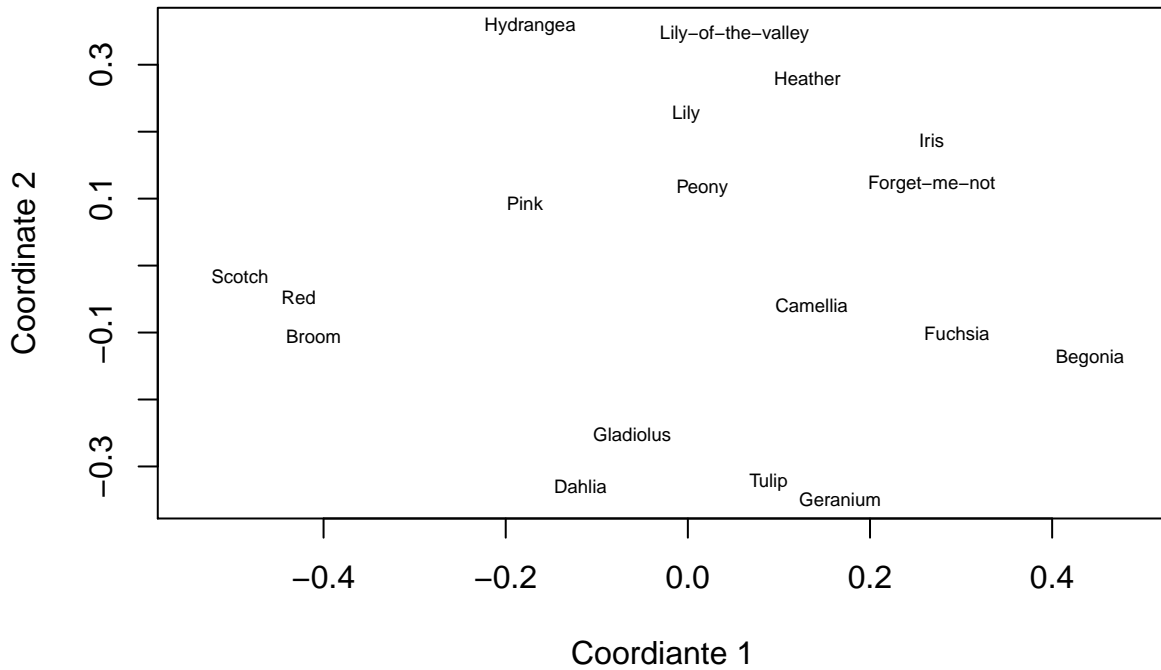


From examining this plot it appears that *rape* and *receiving stolen goods* are viewed as very dissimilar. Taking a look at the eigen values indicates that a 3 dimension solution is probably adequate.

dim	abs_eig	eig2
1	0.4269311	0.7046554
2	0.6490099	0.8953223
3	0.7556547	0.9392907
4	0.8190015	0.9548042
5	0.8739515	0.9664776
6	0.8773917	0.9665234
7	0.9147913	0.9719309
8	1.0000000	1.0000000

4.3

A two-dimensional solution using classical multidimensional scaling is shown below.



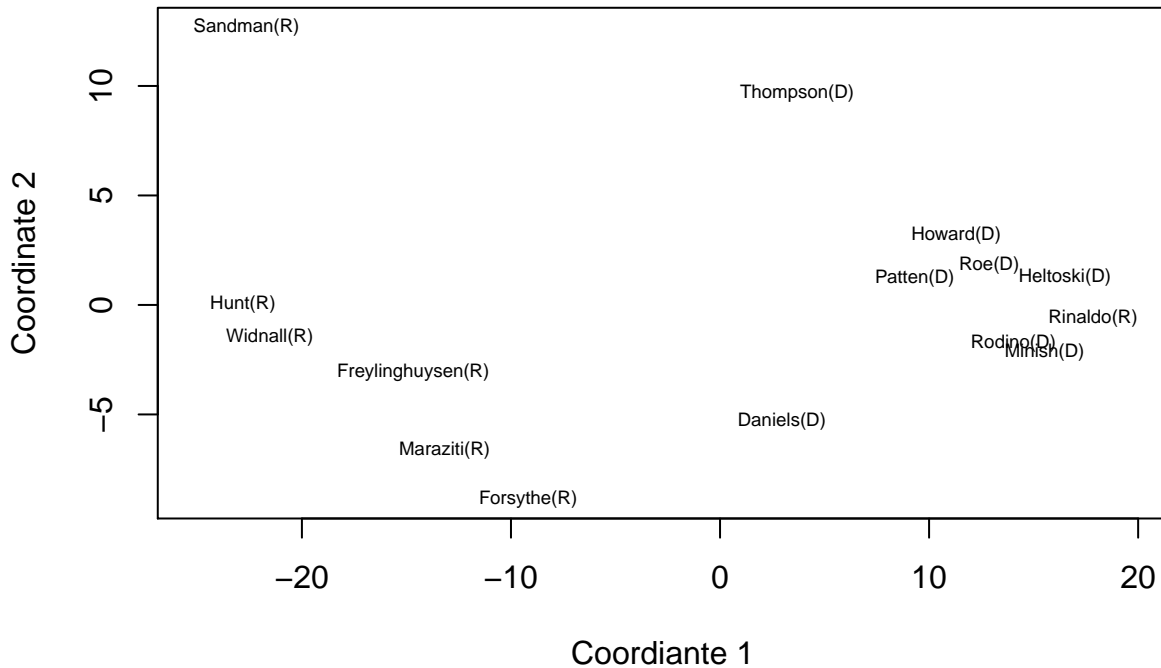
Looking at some of the values from the dissimilarity matrix, a two dimension solution does show indications of groupings of some flowers. Taking a look at some of the flowers, *Begonia* compared to *Broom* has a large value of 0.91 and they are far apart on the plot. This is desired. *Broom* and *Forget-me-not* also have a large value, 0.90, and they are far apart on the plot as well.

Checking on the eigen values, to see how appropriate a two-dimension solution might be we can see that the best solution is probably around 5 or 6 dimensions.

dim	abs_eig	eig2
1	0.2549273	0.4611160
2	0.4493002	0.7291863
3	0.5738645	0.8392805
4	0.6791096	0.9178730
5	0.7364481	0.9412006
6	0.7863903	0.9588982
7	0.8046086	0.9612532
8	0.8190510	0.9627331
9	0.8255356	0.9630315
10	0.8255356	0.9630315
11	0.8302034	0.9631861
12	0.8391005	0.9637478
13	0.8485894	0.9643866
14	0.8678594	0.9670214
15	0.8905760	0.9706829
16	0.9183028	0.9761377
17	0.9555567	0.9859851
18	1.0000000	1.0000000

4.4

For this exercise we take a look at voting data from the `HSAUR2` package.



The locations of the above plot are slightly different from the solution in the book. But there is still two distinct groupings of the two political parties with republicans on the left hand side and democrats on the right hand side. We can still detect party affiliations in the data for the most part, **Rinaldo(R)** being the lone exception.

R code:

```
##### 4.1 #####
# function for s for ex. 4.1
f <- function(x, y) {
  i <- x
  j <- y
  s <- ifelse(i == j, 9,
    ifelse(abs(i - j) >= 1 & abs(i - j) <= 3, 8,
    ifelse(abs(i - j) >= 4 & abs(i - j) <= 6, 7,
    ifelse(abs(i - j) >= 7 & abs(i - j) <= 9, 6,
    ifelse(abs(i - j) >= 10 & abs(i - j) <= 12, 5,
    ifelse(abs(i - j) >= 13 & abs(i - j) <= 15, 4,
    ifelse(abs(i - j) >= 16 & abs(i - j) <= 18, 3,
    ifelse(abs(i - j) >= 19 & abs(i - j) <= 21, 2,
    ifelse(abs(i - j) >= 22 & abs(i - j) <= 24, 1, 0)))))))))
  return(s)
}
# create a matrix with our converted dissimilarities
m <- outer(1:51, 1:51, f)
dij <- sqrt(18 - 2 * m[lower.tri(m)])
d <- matrix(ncol = 51, nrow = 51)
diag(d) <- 0
d[lower.tri(d)] <- dij
```

```

a <- as.dist(d, diag = T, upper = F)
mds <- cmdscale(a, eig = TRUE)
x <- mds$points[,1]; y <- mds$points[,2]
plot(x, y, xlab = "Coordiante 1", ylab = "Coordinate 2", type = "n",
      xlim = range(x) * 1.1)
txt <- 1:51
text(x, y, txt, cex = 0.6)

##### 4.2 #####
# creating the matrix from the problem.
m <- diag(0, 8)
x <- c(21.1,71.2,36.4,52.1,89.9,53,90.1,54.1,36.4,54.1,75.2,73,
       93.2,36.4,52.1,36.4,75.2,71.2,70,54.1,52.1,63.4,53,36.4,
       52.1,88.3,36.4,73)
m[lower.tri(m)] <- x
m[upper.tri(m)] <- t(m[lower.tri(m)])
mds <- cmdscale(m, k = 7, eig = T)
x <- mds$points[,1]; y <- mds$points[,2]
plot(x, y, xlab = "Coordiante 1", ylab = "Coordinate 2", type = "n",
      xlim = range(x) * 1.1)
txt <- c("assault&batt", "rape", "embezz", "perjury", "libel", "burglary",
        "prostitution", "stolegoods")
text(x, y, txt, cex = 0.6)
e <- cumsum(abs(mds$eig)) / sum(abs(mds$eig))
e2 <- cumsum(mds$eig ^ 2) / sum(mds$eig ^ 2)
df <- data.frame(dim = 1:length(e), abs_eig = e, eig2 = e2)
knitr::kable(df)

##### 4.3 #####
m <- HSAUR2::gardenflowers
mds <- cmdscale(m, k = 17, eig = TRUE)
x <- mds$points[,1]; y <- mds$points[,2]
plot(x, y, xlab = "Coordiante 1", ylab = "Coordinate 2", type = "n",
      xlim = range(x) * 1.1)
txt <- labels(m)
txt <- sub(" .*", "", txt)
text(x, y, txt, cex = 0.6)
e <- cumsum(abs(mds$eig)) / sum(abs(mds$eig))
e2 <- cumsum(mds$eig ^ 2) / sum(mds$eig ^ 2)
df <- data.frame(dim = 1:length(e), abs_eig = e, eig2 = e2)
knitr::kable(df)

##### 4.4 #####
voting <- HSAUR2::voting
d <- dist(voting)
mds <- cmdscale(d)
x <- mds[,1]
y <- mds[,2]
plot(x, y, xlab = "Coordiante 1", ylab = "Coordinate 2", type = "n",
      xlim = range(x) * 1.1)
text(x, y, colnames(voting), cex = 0.6)

```