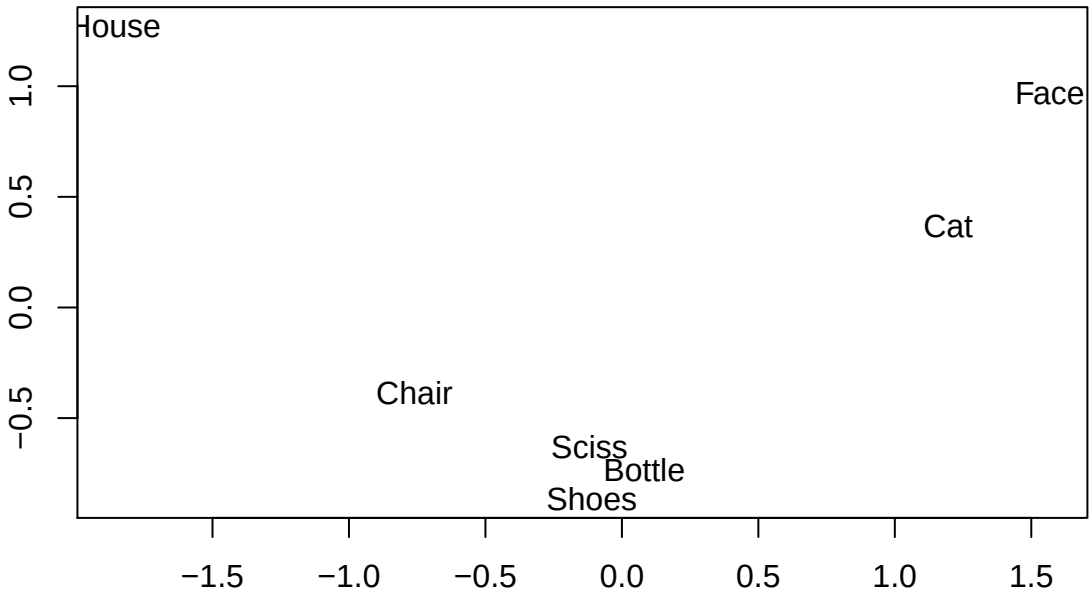
348HWw9

Shay Lebovitz

11/8/2020

1) *a)*

```
brain <- read.table('/Users/shaylebovitz/R/Stat 348/brain.txt', header = TRUE)
brain <- as.dist(brain)
brain_mds <- cmdscale(brain, eig = TRUE)
plot(brain_mds$points, type = 'n', xlab = '', ylab = '')
text(brain_mds$points, labels = labels(brain))
```



b)

```
brain_mds$GOF
```

```
## [1] 0.7217058 0.7400254
```

Both methods for goodness of fit give results greater than 0.70, so I'd say the 2-D MDS model is fairly accurate.

c)

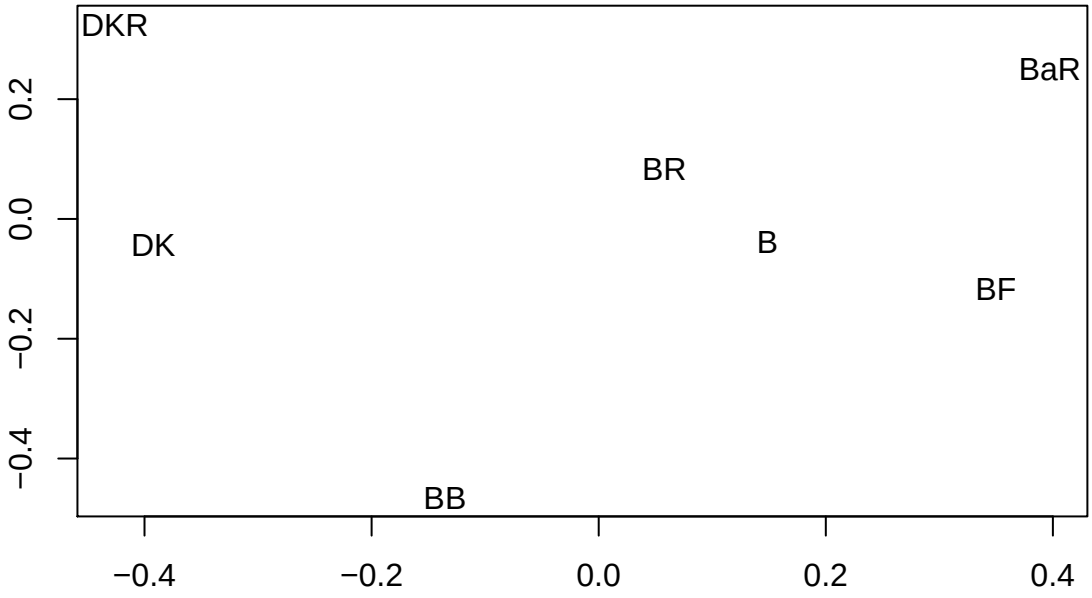
```
brain_mds$eig
```

```
## [1]  7.951622e+00  4.518817e+00  2.232862e+00  1.198964e+00  9.490998e-01
## [6]  3.330669e-16 -4.277512e-01
```

We see that the first two eigengvectors are relatively large, but not that much different than the third. Overall, the MDS analysis is appropriate in this case with regards to the ability to distinguish between categories based on a brain scan.

2) *a)*

```
movies <- read.table('/Users/shaylebovitz/R/Stat 348/movies.txt', header = TRUE)
movies <- as.dist(movies)
movies_mds <- cmdscale(movies, eig = TRUE)
plot(movies_mds$points, type = 'n', xlab = '', ylab = '')
text(movies_mds$points, labels = labels(movies))
```



b)

```
movies_mds$GOF
```

```
## [1] 0.7180194 0.8190584
```

```
movies_mds$eig
```

```
## [1]  6.594746e-01  4.080729e-01  1.686167e-01  6.721957e-02 -1.665335e-16
## [6] -4.371786e-02 -1.396931e-01
```

We see that the first two eigenvalues are relatively large compared to the others, and the goodness-of-fits for each method are above 0.70, and thus we can say that the 2D MDS model is fairly accurate for the data.

c) Because 'distance' in this case is a measure of the similarity in reviewers' ratings for a given movie (0 for all reviewers ratings are the same, 1 if all different), the axes in this case can be viewed as the variability, or perhaps the contentiousness/polarizability of a movie.

d) Since she didn't like "Batman Forever", we would want to find the movie with the greatest distance away from it.

```
movies
```

```
##         B    BR    BF   BaR    DK    BB
## BR  0.571
## BF  0.250 0.500
## BaR 0.375 0.500 0.444
## DK  0.500 0.500 0.778 0.778
## BB  0.500 0.625 0.556 0.889 0.222
## DKR 0.667 0.667 0.857 0.857 0.143 0.857
```

That movie would be "Dark Knight Rises" (DKR, distance = 0.857). To find one she wouldn't like, we need

to find the most similar one. That would be "Batman" (B, distance = 0.250).

3) *a)*

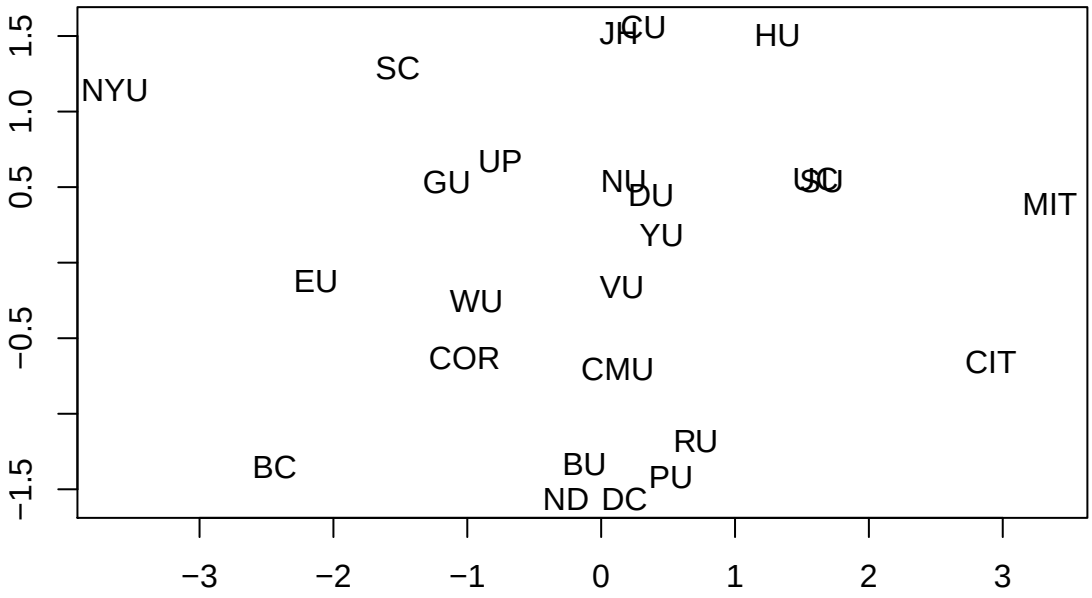
```
colleges <- read.table('/Users/shaylebovitz/R/Stat 348/Colleges.txt', header = TRUE)

fnc <- function(x) {
  x/sd(x)
}
colleges <- apply(colleges, 2, fnc)
colleges <- as.data.frame(colleges)
```

b)

```
col_dist <- dist(colleges)
```

c)

```
colleges_mds <- cmdscale(col_dist, eig = TRUE)
plot(colleges_mds$points, type = 'n', xlab = '', ylab = '')
text(colleges_mds$points, labels = labels(col_dist))
```



d)

```
colleges_mds$GOF
```

```
## [1] 0.6940049 0.6940049
```

```
colleges_mds$eig
```

```
##  [1]  5.794127e+01  2.533933e+01  1.943377e+01  1.169702e+01  5.588615e+00
##  [6]  1.170022e-14  1.285613e-15  3.624999e-16 -5.522422e-17 -1.320100e-16
## [11] -1.983912e-16 -3.222912e-16 -3.585603e-16 -5.187469e-16 -8.359932e-16
## [16] -9.099191e-16 -1.473222e-15 -1.519925e-15 -1.605773e-15 -1.825994e-15
## [21] -1.884079e-15 -2.749155e-15 -2.967755e-15 -3.156702e-15 -5.411358e-15
```

The goodness of fit by both methods is 0.69, which is right below the usual lowest accepted value of 0.70. I think it would still be fine, as 0.70 is somewhat arbitrarily chosen. We see that the third eigenvalue is relatively similar in magnitude to the second, so perhaps a 3D MDS representation would be more appropriate.

e) It is hard to tell exactly from the distance matrix, but from the plot it looks like the four most similar colleges would be Duke (DU), Penn (UP), Yale (YU), and either Georgetown (GU), U.Chicago (UC), or Stanford (SU). The least similar would be Notre Dame (ND), Dartmouth (DC), Princeton (PU), and Boston College (BC).