S&DS 363, HW6: Ordination

Lorenzo Flores (Individual Submission)

First, some useful packages for performing Ordination.

```
library(vegan)
```

```
## Loading required package: permute
```

```
## Loading required package: lattice
```

```
## This is vegan 2.5-4
```

```
library(vegan3d)
library(mgcv)
```

```
## Loading required package: nlme
```

```
## This is mgcv 1.8-24. For overview type 'help("mgcv-package")'.
```

```
library(MASS)
library(rgl)
```

Reading in Data

```
cereal <- read.csv("http://reuningscherer.net/stat660/data/cereal.attitudes.csv")
```

1. Fit Correspondence Analysis to your data.

Correspondence Analysis

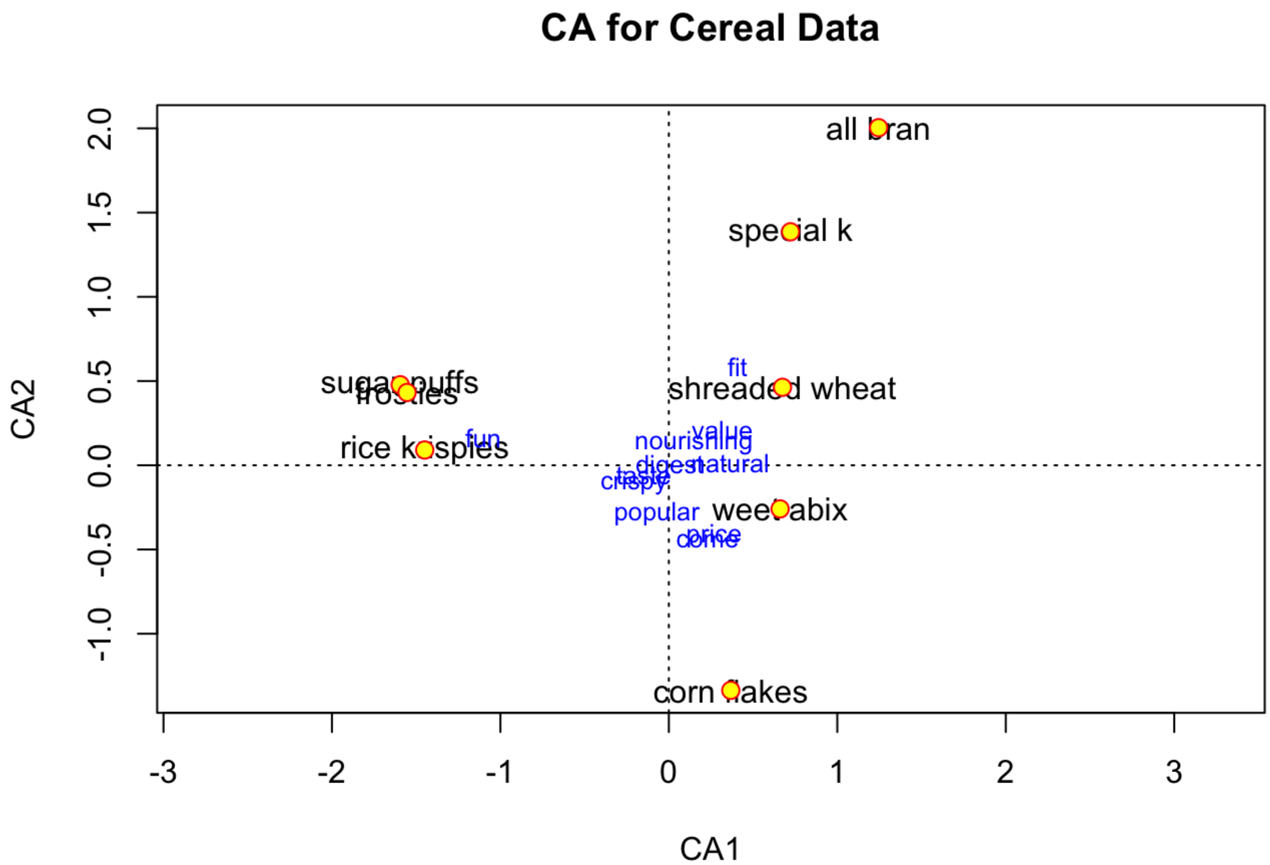
```
#Perform correspondence analysis
cerealcca <- cca(cereal[,2:12])
```

2. Discuss the inertia, make a two dimensional plot of the first two CA directions.

```
summary(cerealcca)
```

```
##
## Call:
## cca(X = cereal[, 2:12])
##
## Partitioning of scaled Chi-square:
##               Inertia Proportion
## Total          0.2697          1
## Unconstrained  0.2697          1
##
## Eigenvalues, and their contribution to the scaled Chi-square
##
## Importance of components:
##                          CA1     CA2     CA3      CA4       CA5       CA6
## Eigenvalue            0.1719 0.07409 0.01349 0.009194 0.0008484 0.0001575
## Proportion Explained  0.6373 0.27468 0.05001 0.034087 0.0031452 0.0005839
## Cumulative Proportion 0.6373 0.91200 0.96201 0.996098 0.9992429 0.9998268
##                             CA7
## Eigenvalue            4.672e-05
## Proportion Explained  1.732e-04
## Cumulative Proportion 1.000e+00
##
## Scaling 2 for species and site scores
## * Species are scaled proportional to eigenvalues
## * Sites are unscaled: weighted dispersion equal on all dimensions
##
##
## Species scores
##
##                  CA1       CA2      CA3      CA4       CA5       CA6
## come        0.231262 -0.442973  0.09068 -0.08996  0.021724  0.003400
## taste      -0.148625 -0.056114 -0.01393  0.01193 -0.047421  0.018916
## popular    -0.071308 -0.285575  0.02424 -0.02211  0.001408  0.008809
## digest      0.006142 -0.004481  0.02188 -0.08172 -0.042313 -0.024605
## nourishing  0.148661  0.137265 -0.16861  0.03546  0.001952  0.010523
## natural     0.366225  0.017419 -0.04978  0.09790  0.042256 -0.005702
## price       0.268622 -0.417795 -0.03489 -0.08287  0.025677 -0.002860
## value       0.318227  0.212557 -0.17019  0.07076 -0.009360 -0.011981
## crispy     -0.207558 -0.103751  0.23583  0.26891 -0.004733 -0.005989
## fit         0.408278  0.587310  0.19178 -0.09942  0.013200  0.008026
## fun        -1.100558  0.164375 -0.03585 -0.04693  0.031455 -0.003298
##
##
## Site scores (weighted averages of species scores)
##
##          CA1      CA2     CA3      CA4     CA5     CA6
## sit1  0.3677 -1.33606  0.8619 -0.04802 -0.1726  0.2250
## sit2  0.6603 -0.25909 -1.6540 -0.90165  0.3347 -0.2735
## sit3 -1.4492  0.09057  0.5041  0.57093  0.9201 -1.0846
## sit4  0.6746  0.46287 -0.7508  2.55736  0.2352  0.1759
## sit5 -1.5947  0.47803 -0.4929 -0.22296 -0.8490  2.4167
## sit6  0.7223  1.38589  0.5626 -0.17849 -2.7226 -1.1309
## sit7 -1.5540  0.43212 -0.1099 -0.70982  0.1698 -1.2885
## sit8  1.2453  2.00386  1.4516 -0.90863  1.6985  0.8354
```

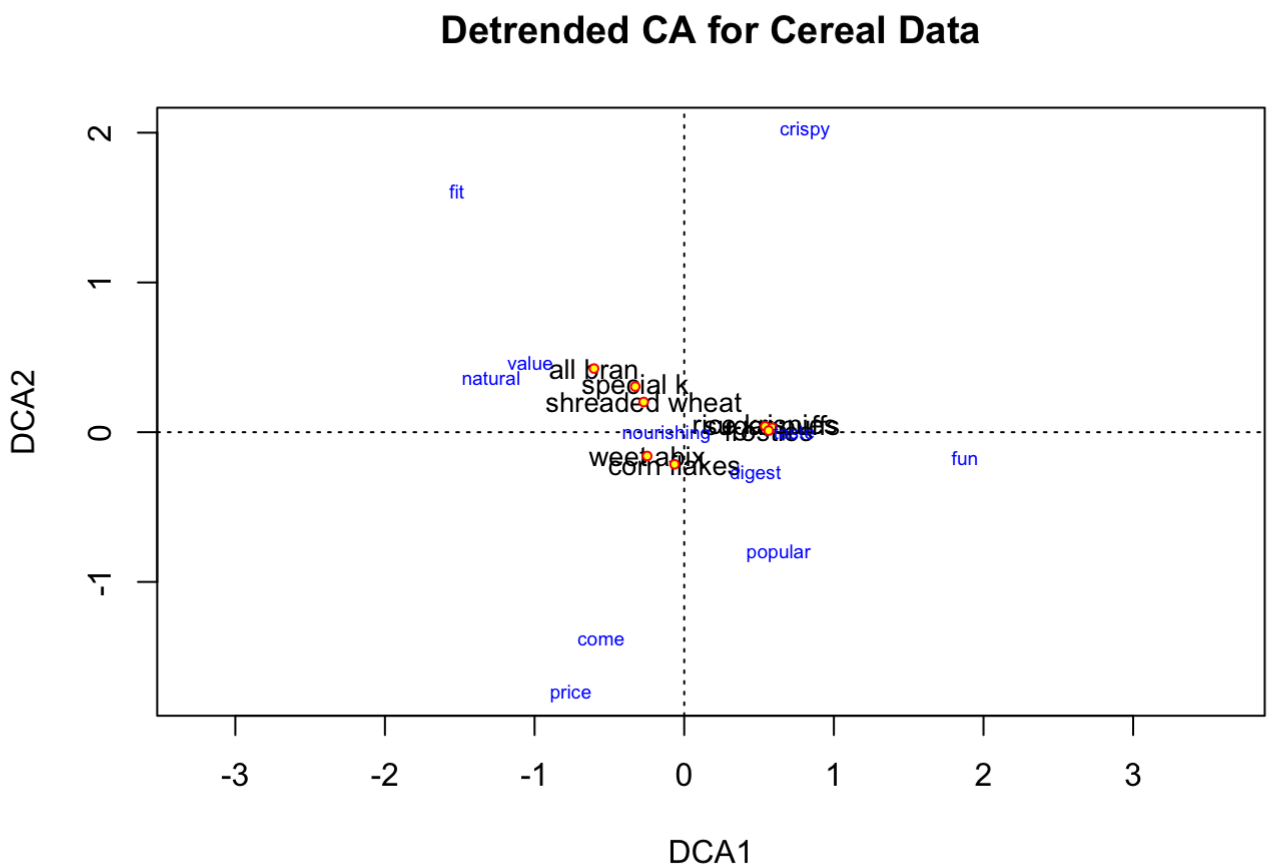
```
#plot results
plot(cerealcca, type="n" ,xlim=c(-2,2.5), main = "CA for Cereal Data")
text(cerealcca, dis="wa",labels=cereal[,1])
points(cerealcca, pch=21, col="red", bg="yellow", cex=1.2)
text(cerealcca, "species", col="blue", cex=0.8)
```



Basing off the propertion explained, which in turn was derived from the eigenvalues (the square root of the inertia), the first component accounted for 63.7%, while the second component accounted for 27.5%, giving a total of 91.2% explained in this plot alone. Furthermore, 99% of the variance was explained within 4 components.

Detrended Correspondence Analysis

```
#detrended correspondence analysis
cerealdca<-decorana(cereal[,2:12])
plot(cerealdca,type="n", main = "Detrended CA for Cereal Data")
text(cerealdca, display=c("sites"),labels=cereal[,1],cex=0.86)
points(cerealdca, pch=21, col="red", bg="yellow", cex=0.6)
text(cerealdca, "species", col="blue", cex=0.6)
```



3. Comment on whether or not there is any evidence of 'data snaking' in higher dimensional space.

Data snaking is evident in a plot when data seem to circle around areas. It appeared that in the first CA plot, the different cereal brands did circle around the plot, although I cannot think of a clear reason or point to a specific variable that would account for this. This will explored in later questions.

4. In a few sentences, describe what you conclude from your plot.

CA1 appears to be mainly controlled by variables like price and 'natural' flavors (upwards), in contrast to taste and crispiness (downwards). CA2 on the other hand contrasts popularity, if they would come back to it, and price (leftwards), with value and how fit eating the cereal keeps you (rightwards).

The plot describes 91.2% of the variance, as was noted from the inertia and eigenvalues.

We see from the plot that Rice Krispies, Frosties, and Sugar Puffs possess the same attributes (popular, great taste, crispy); Corn Flakes and Weet Abix appear close to one another, exhibiting good price and if customers would come back to them; Shredded Wheat, Special K, and All Bran exhibited increasing levels of perceived value and helping consumers stay fit.

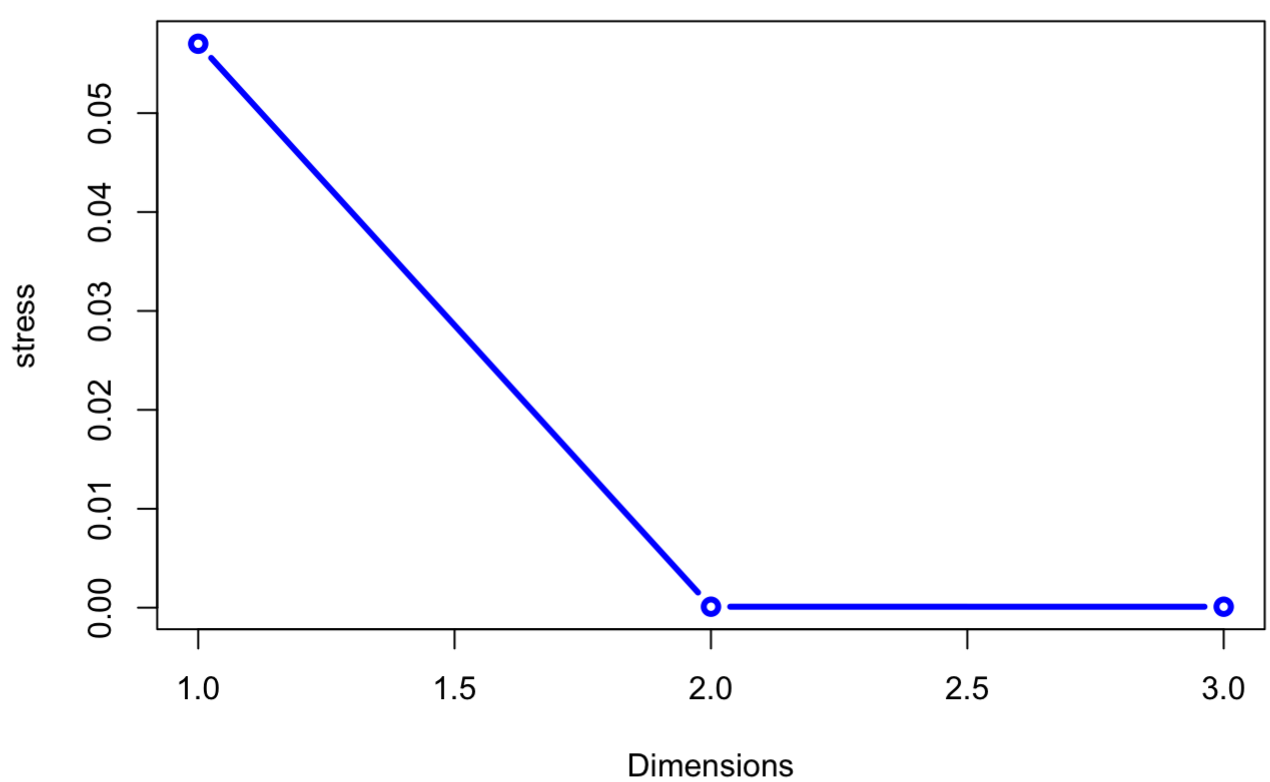
5. Perform Multidimensional Scaling (metric or non-metric) for 1, 2, and 3 dimensions.

```
rownames(cereal) <- cereal[,1]
cereal <- cereal[,-1]
```

```
cereal.mds1 <- metaMDS(cereal, distance="euclidean", k=1)
cereal.mds2 <- metaMDS(cereal, distance="euclidean", k=2)
cereal.mds3 <- metaMDS(cereal, distance="euclidean", k=3)
```

6. Discuss the stress (or SStress) of each dimensional solution. Make a scree plot if you're able.

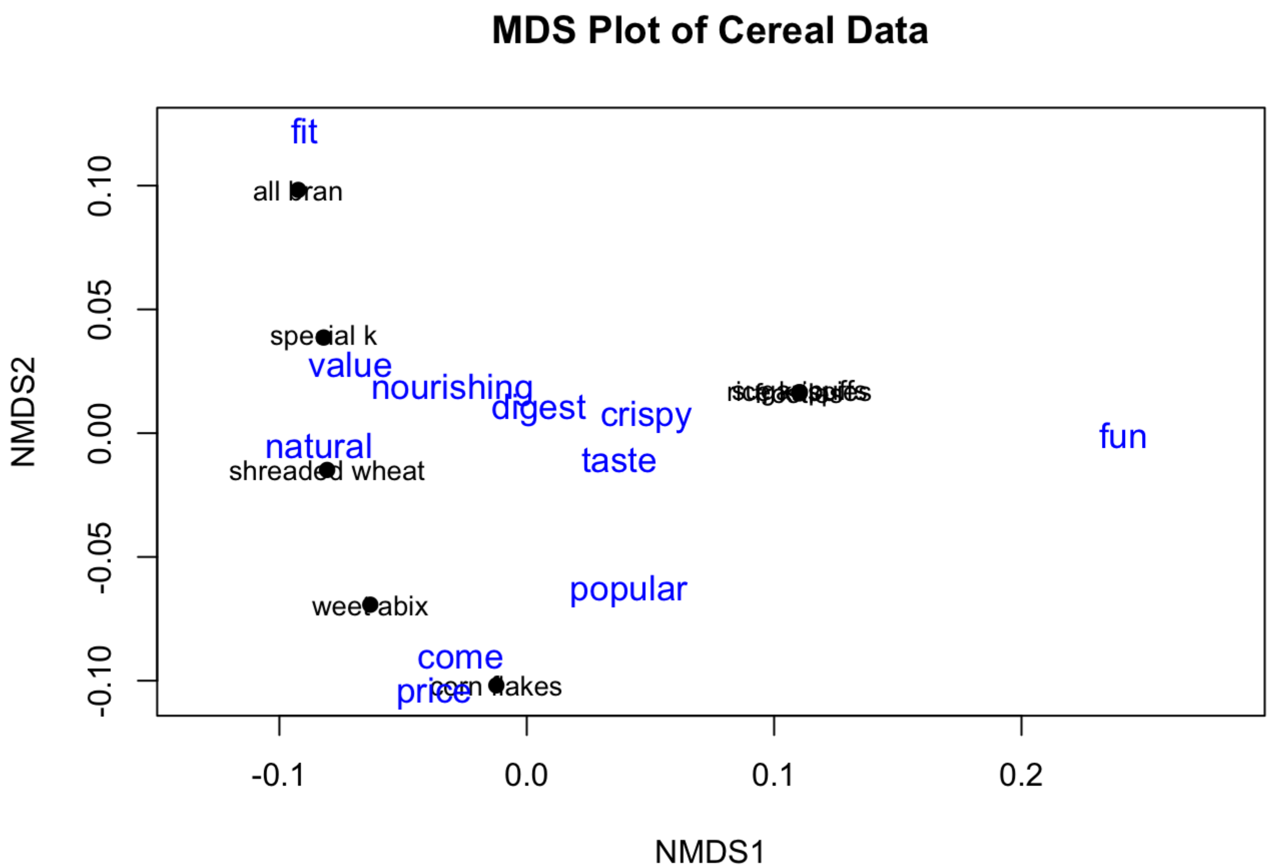
```
stress <- c(cereal.mds1$stress, cereal.mds2$stress, cereal.mds3$stress)
plot(c(1:3),stress,col="blue",lwd=3, type="b", xlab="Dimensions")
```



Stress was slightly above 0.05 at one dimension, then dropped to nearly zero for second and third dimensions. Based on the scree plot, it recommends us to use two dimensions.

7. Make a two dimensional plot of your results.

```
plot(cereal.mds2, type="n", main = "MDS Plot of Cereal Data")
points(cereal.mds2, pch=19, col="black", cex=1)
text(cereal.mds2, "sites", labels=rownames(cereal),cex=0.86)
text(cereal.mds2, "species", col="blue", cex=1.1)
```



The plot shows that All Bran is associated with helping consumers remain fit; Shreaded Wheat and Special K are associated with value, being nourishing, and having a natural taste; Rice Krispies, Sugar Puffs, and Frosties are similar to one another and associated with crispiness and taste, and Weet Abix and Corn Flakes are cereals people would come back to, and have good price. It is interesting to note that the variable fun is not strongly associated with any of the cereals. The multidimensional plot yields similar results to the original CA, with the same groups of cereals being clustered together and associated with the same traits.