DAPT 622 Assignment 3

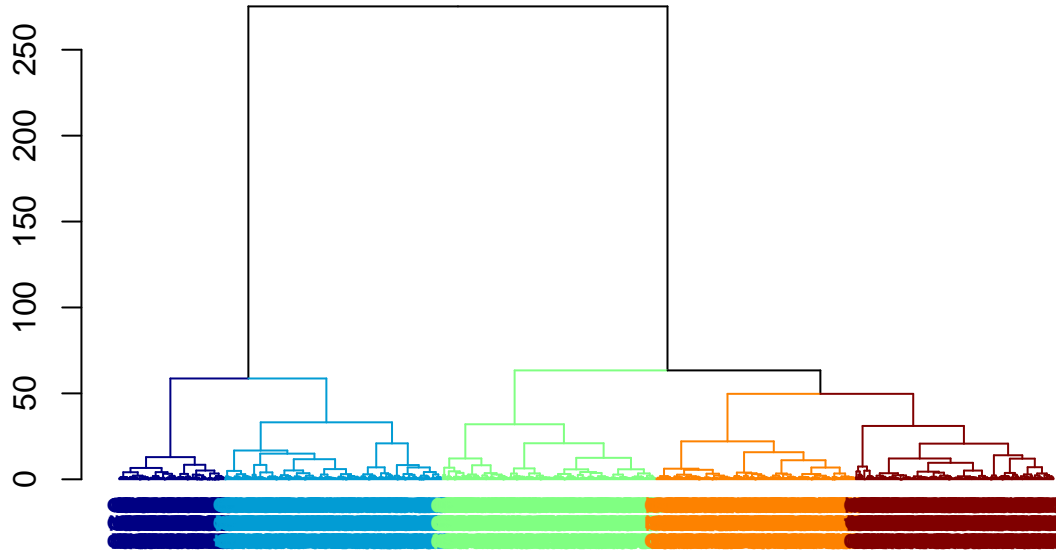*Daniel Erb*

March 2, 2019

Contents

Part 1 - Cluster Analysis

Section A

Perform a heirarchical cluster analysis (via Ward's method) using all the variables except User ID.

Subsection i

Select an appropriate number of clusters. Provide a dendrogram with the clusters highlighted.

Subsection ii

Provide a table summarizing the clusters via their means. Are there any distinguishing qualities regarding the clusters?

Group.1	art.gals	dance.clubs	j.bars	rest	mus	res	park.pic	beach	theaters	relig.inst
1	0.94	1.53	1.56	0.56	1.32	2.45	3.19	2.82	1.63	2.57
2	1.03	1.04	0.97	0.49	0.76	1.63	3.18	2.85	1.42	2.81
3	0.87	1.36	0.42	0.41	0.68	1.34	3.17	2.84	1.48	3.14
4	0.78	1.53	0.34	0.67	0.95	1.88	3.18	2.86	1.80	2.82
5	0.78	1.26	2.36	0.55	0.97	1.91	3.19	2.80	1.47	2.56

For this dataset, 5 clusters were chosen as they seem to separate the groups into by usable chunks without being too specific as to be confusing. When looking at the average scores for each type of destination, across the individuals of cluster 1, we see that they tend to rate dance clubs, museums, and restaurants more positively than the individuals within other clusters, while only giving lower scores to religious institutions. Cluster 2, seems to favor art galaries, while disfavoring dance clubs and theaters, when compared to the other clusters. For cluster 3, individiduals within this group tend to rate juice bars, museums and restaurants lower than other clusters, while rating religious institutions higher than others. Cluster 4 tended to give particularly low ratings to juice bars. This in contrast to rating restaurants and theaters higher than other clusters. Cluster 5 had an average rating for juice bars that tended to be one or two whole ratings higher. They also tended to rate religious institutions lower than other clusters. One intersting note for the averages, both beaches and parks seemed consistently rated across all the clusters.

Subsection iii

Perform a "two-way" cluster analysis (i.e., cluster the variables) and provide the dendrogram showing the variable clusters and a heat map of the data. Which variables cluster together?

```
mydata <- as.matrix(trip.advisor[,2:11])
rownames(mydata) <- trip.advisor[,1]
datascaled <- scale(mydata)
heatmap(datascaled, col = colormap(colormap = colormaps$jet))
```

park.pic
j.bars
mus
res
ance.clubs
rest
beach
theaters
art.gals
relig.inst