

# DAPT 622 Assignment 3

*Daniel Erb*

*March 2, 2019*

## Contents

<b>Part 1 - Cluster Analysis</b>	<b>1</b>
Section A . . . . .	1
Section B . . . . .	3

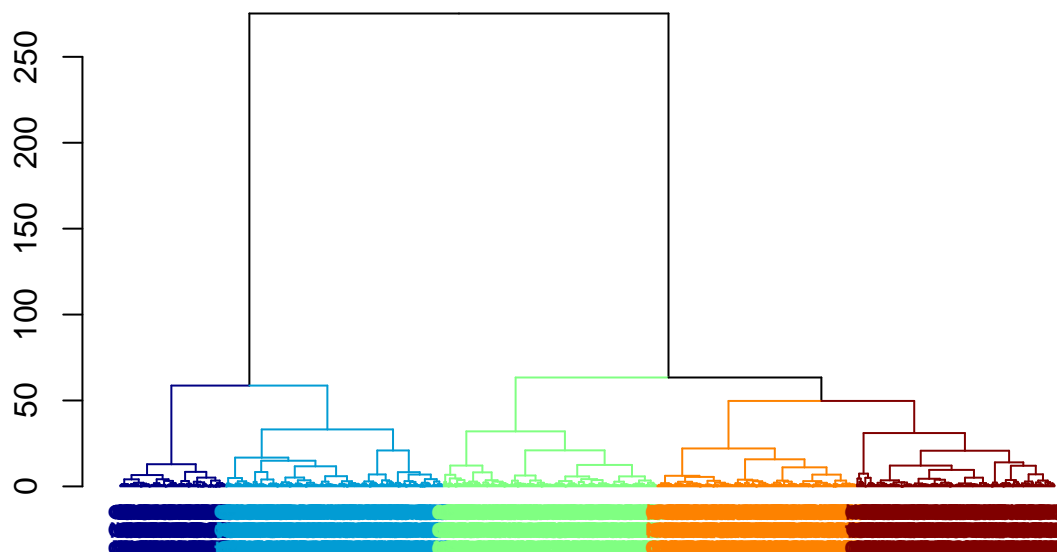
## Part 1 - Cluster Analysis

### 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?

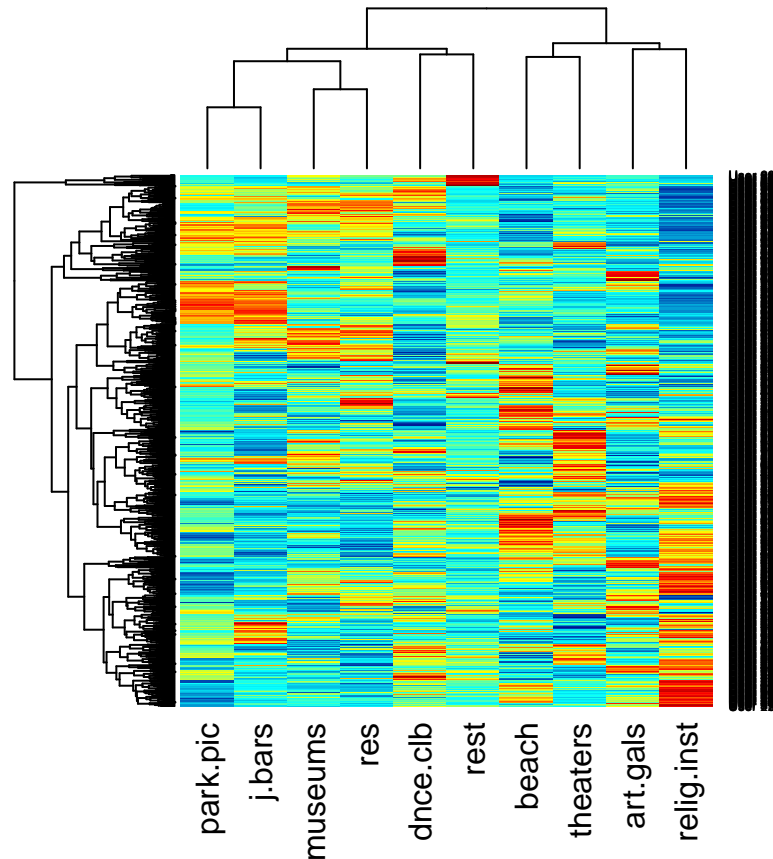
Cluster	art.gals	dnce.clb	j.bars	rest	museums	res	park.pic	beach	theaters	relig.inst
1	0.94	1.53	1.56	0.56	1.32	2.45	3.19	2.82	1.63	2.57
2	1.03	1.04	0.97	0.49	0.76	1.63	3.18	2.85	1.42	2.81
3	0.87	1.36	0.42	0.41	0.68	1.34	3.17	2.84	1.48	3.14
4	0.78	1.53	0.34	0.67	0.95	1.88	3.18	2.86	1.80	2.82
5	0.78	1.26	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 galleries, while disfavoring dance clubs and theaters, when compared to the other clusters. For cluster 3, individuals 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 interesting 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))
```



Parks and picnic areas clustered together with juice bars earlier than any other categories. The next to cluster to form was comprised of museums and restaurant ratings. These two clusters then clustered sooner than any other categories clustered with any other categories. As we follow the tree upward, we see that beaches and theaters cluster together next, however the visual seems to give weaker evidence of this.

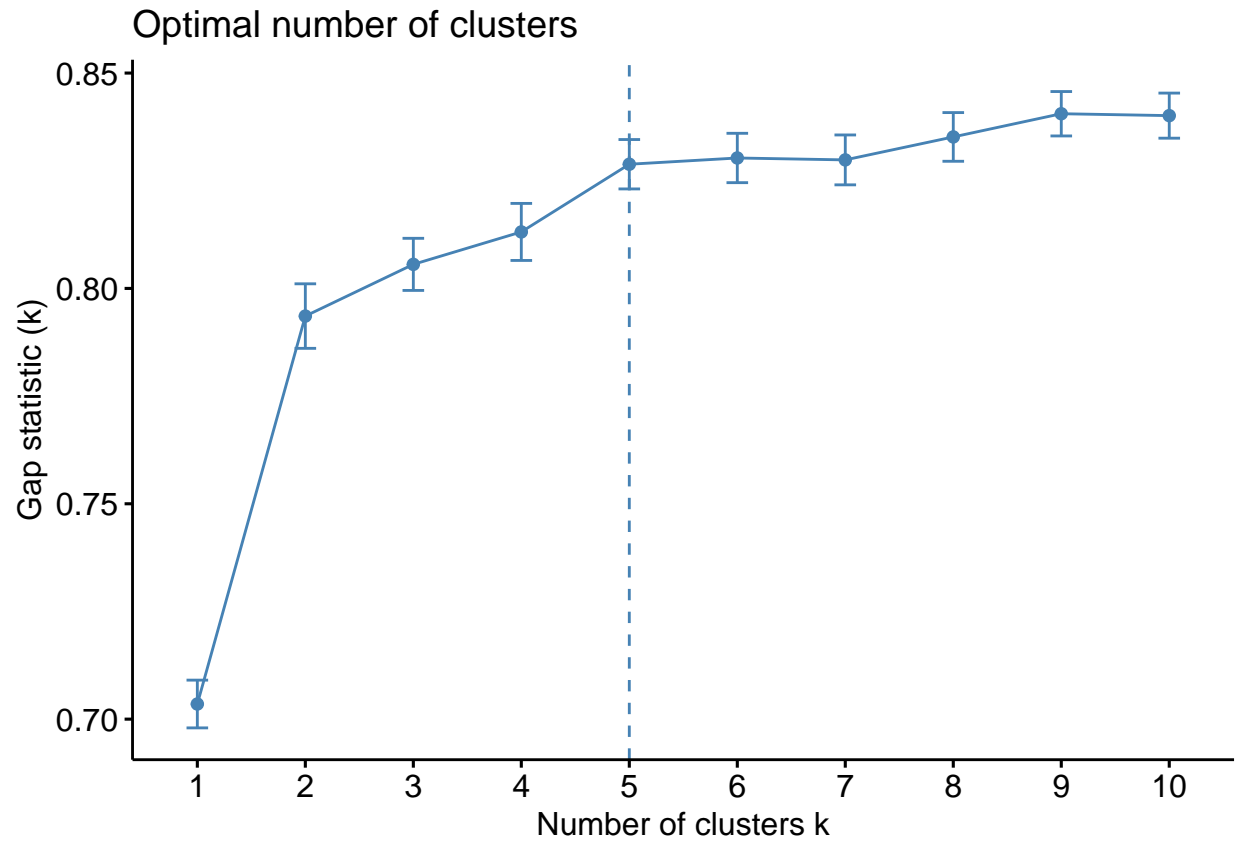
## Section B

Now perform a k-means cluster analysis. Specify the same number of clusters as selected in part a). Display the cluster summary including the cluster means. Are different clusters produced than in part a)? Describe what you see.

```
fviz_nbclust(trip.advisor[,2:11],kmeans,method="gap_stat")
```

```
## Warning: did not converge in 10 iterations
```

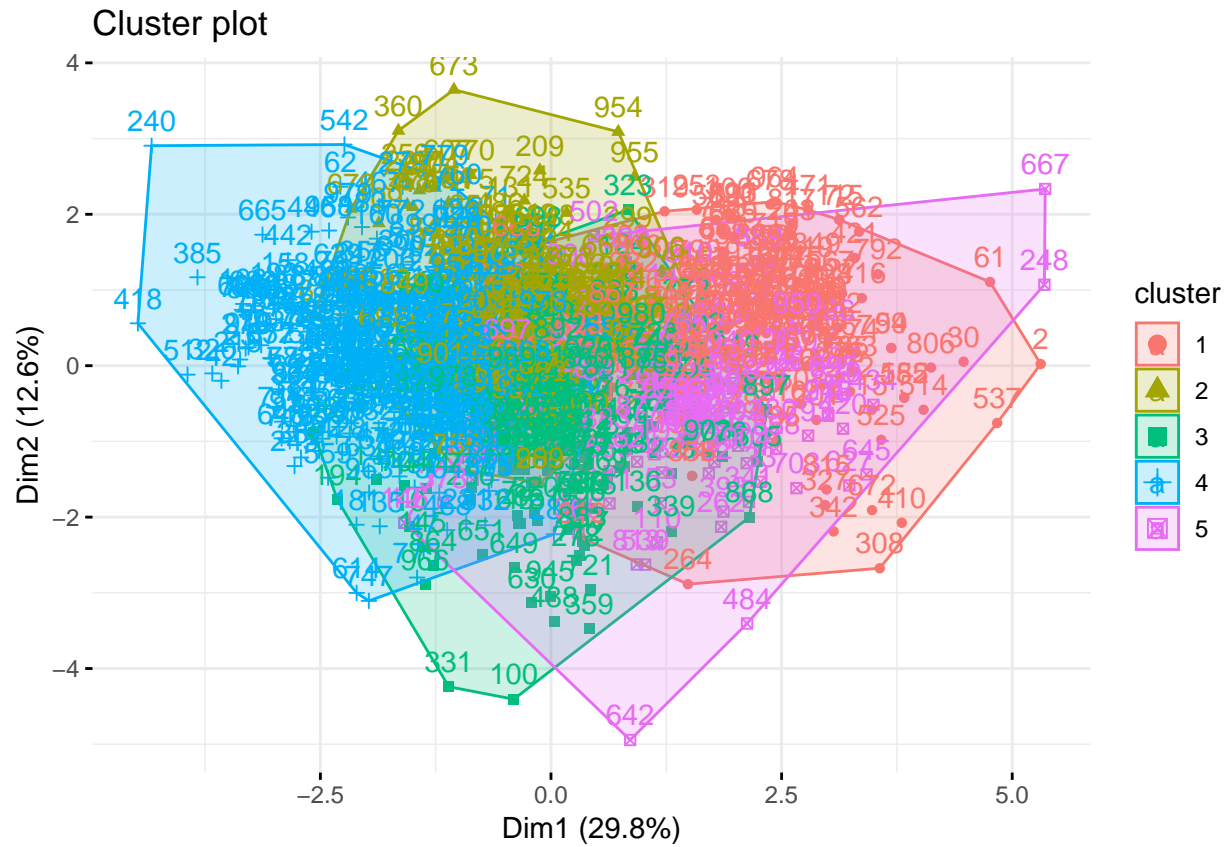
```
## Warning: did not converge in 10 iterations
```



```
kmean.clust <- kmeans(trip.advisor[,2:11],k.clusts)
agg.vals <- as.data.frame(aggregate(mydata,by=list(kmean.clust$cluster),FUN=mean))
agg.vals <- rename(agg.vals, Cluster = Group.1)
kable(agg.vals, digits = 2, format = "markdown", caption = "K-Means Groupings")
```

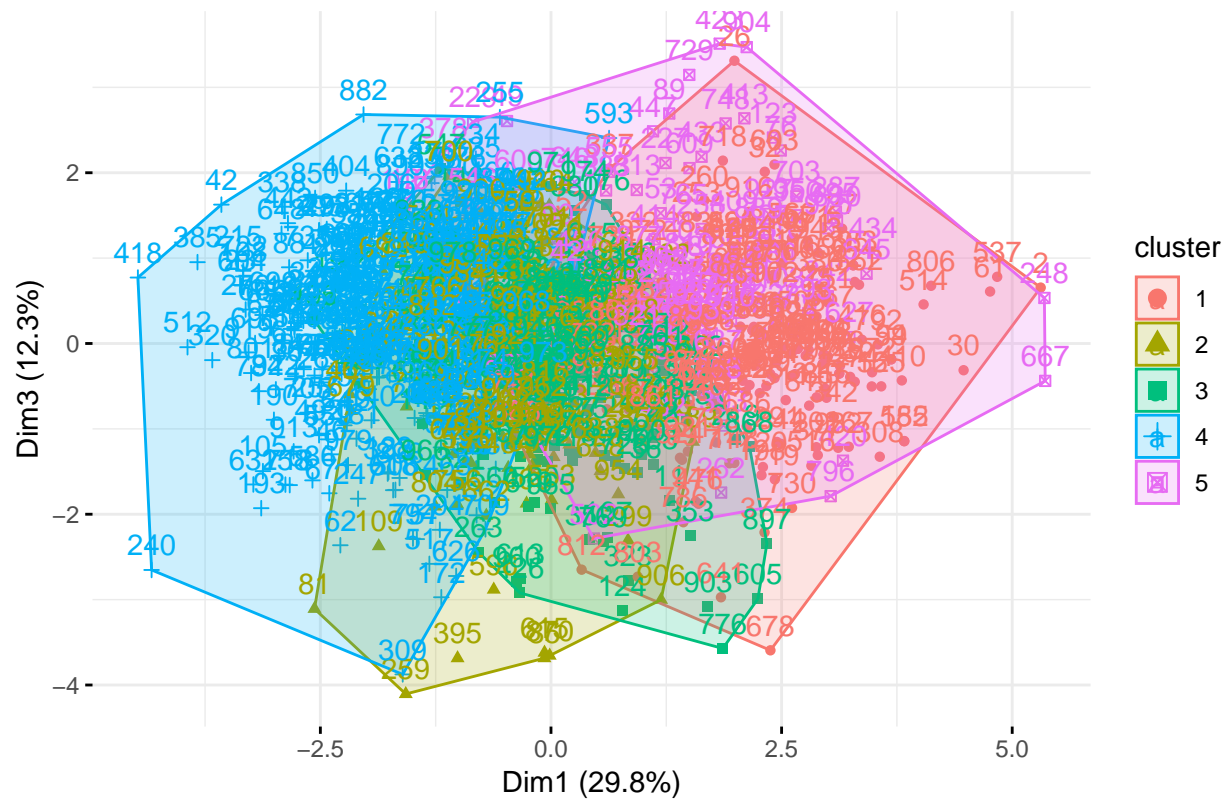
Cluster	art.gals	dnce.clb	j.bars	rest	museums	res	park.pic	beach	theaters	relig.inst
1	0.83	1.34	2.26	0.56	1.16	2.16	3.19	2.80	1.57	2.54
2	0.98	1.12	1.29	0.47	0.83	1.67	3.18	2.84	1.40	2.85
3	0.88	1.24	0.47	0.57	1.12	2.18	3.18	2.88	1.70	2.80
4	0.87	1.29	0.37	0.45	0.62	1.35	3.18	2.84	1.59	3.01
5	0.95	2.09	1.17	0.74	1.30	2.35	3.18	2.80	1.56	2.60

```
fviz_cluster(kmean.clust,data=trip.advisor[,2:11],
  axes = c(1,2),
  ggtheme = theme_minimal())
```



```
fviz_cluster(kmean.clust,data=trip.advisor[,2:11],
  axes = c(1,3),
  ggtheme = theme_minimal())
```

Cluster plot



```
fviz_cluster(kmean.clust,data=trip.advisor[,2:11],
             axes = c(2,3),
             ggtheme = theme_minimal())
```

Cluster plot

