

# TUTORIAL 9 HELP SHEET

## EXERCISE 9A: CLUSTER ANALYSIS ON SIMPLE DATA

### PART A

HINT 1: Think about if the data needs to be standardised, if you should set a seed, <sup>Set seed (2018) to get same as me.</sup> & the types of variables that can be used in clustering.

HINT 2: Check lecture 9 slide 36 for an example on making a hierarchical clustering, and slide 38 to see how to cut the tree and get the clusters. Check slide 47 to see how to get the k-means clusters.

HINT 3: Try `hclust(dist(data), method)` and `cutree(hclustier, # of clusters)` to get the hierarchical cluster and `kmeans(data, # clusters)`

HINT 4: To make the confusion table check lecture 9 slide 49 or use `count()` and `pivot_wider()` on a tibble that has a variable for each cluster.

### PART B

HINT 5: Map the cluster labels basically just means to specify which cluster in words is which cluster in k-means (basically correcting the labels). Sometimes you won't be able to map the cluster labels.

HINT 6: Agreement is  $\frac{\text{\# observations that are in the same cluster (with a label correction)}}{\text{total \# of observations}}$ .

## EXERCISE 9B: CLUSTER STATISTICS GRADUATE PROGRAMS

### PART A & B

HINT 7: Part A & B of Exercise 9B are almost identical to that of Exercise 9A, so you can just re-use the above hints. The main difference is that your second clustering will use different data instead of a different method.

### PART C

HINT 8: Try `ggscatmat(data, columns, color)` for the scatter plot matrix and `ggparcoord(data, columns, groupColumn, order)` to make the parallel coordinate plot.

HINT 9: To discuss a clustering think about whether the clustering has picked up on natural separations in the data, characteristics of each cluster, and if the number of observations in each cluster is the same.

### PART D

HINT 10: This should be a numerical summary of what you learned in the earlier plots. Keep in mind it only makes sense to use the 6 variables seen by both clusterings.

HINT 11: Try `select()`, `group-by`, and `summarise()`.