

# Naum-Samples

Fernando Gutierrez

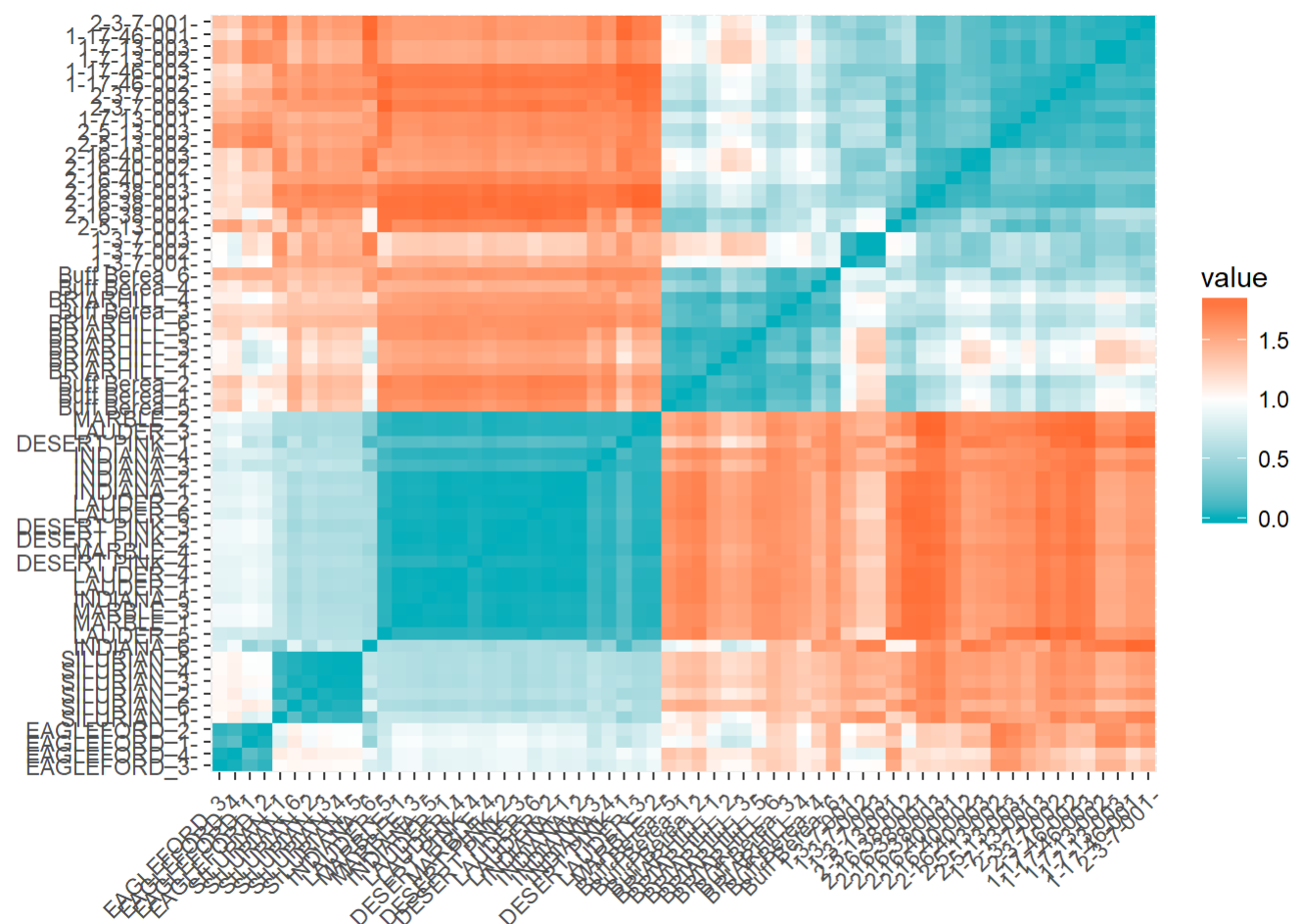
April 19, 2018

The elemental data was collected using a handheld XRF gun at two energy settings. Major elements were detected using a lower voltage setting. Traces were scanned using a high energy setting. Each sample was scanned for two minutes. In the table below, for example, the readings Buff Berea\_1 and 2 belong to the same sample.

##	Sample	Mg	Al	Si	Ca	K	Ti
## 1	Buff Berea_1	0.001	2.530261	34.55315	0.2596631	0.7912687	0.22236322
## 2	Buff Berea_2	0.001	2.473600	34.40304	0.2517036	0.7861751	0.23329620
## 3	Buff Berea_3	0.001	2.116177	34.33145	0.2331608	0.5670154	0.14817508
## 4	Buff Berea_4	0.001	2.102057	34.52096	0.2315730	0.5551911	0.15836986
## 5	Buff Berea_5	0.001	2.054869	34.30847	0.2880010	0.7122518	0.27137828
## 6	Buff Berea_6	0.001	1.971308	34.13965	0.2729138	0.6847200	0.25867003
## 7	BRIARHILL_1	0.001	1.752356	35.51626	0.2066847	0.1947084	0.14871110
## 8	BRIARHILL_2	0.001	1.695216	35.18963	0.1979021	0.1867750	0.15131002
## 9	BRIARHILL_3	0.001	1.809483	35.83321	0.2186367	0.4772847	0.08609406
## 10	BRIARHILL_4	0.001	1.902852	35.76810	0.2188049	0.4866401	0.09041277
##	Fe	P	S	Th	U		
## 1	0.84474030	0.04143695	0.2568482	5.344030	11.109485		
## 2	0.93524566	0.03208158	0.2620198	5.702483	9.100555		
## 3	0.05561469	0.03204321	0.2714004	4.959338	3.583232		
## 4	0.00100000	0.03558950	0.2582115	3.235979	0.001000		
## 5	0.73431121	0.02735151	0.2170681	5.313419	11.961882		
## 6	0.72004661	0.02341854	0.2145897	5.347436	1.931609		
## 7	0.85022179	0.04649287	0.2417173	4.521085	7.949242		
## 8	0.87113482	0.04218400	0.2354193	4.356281	13.358783		
## 9	0.03217399	0.04648058	0.2576128	4.751681	11.371237		
## 10	0.00100000	0.04567802	0.2610075	4.583604	5.691024		

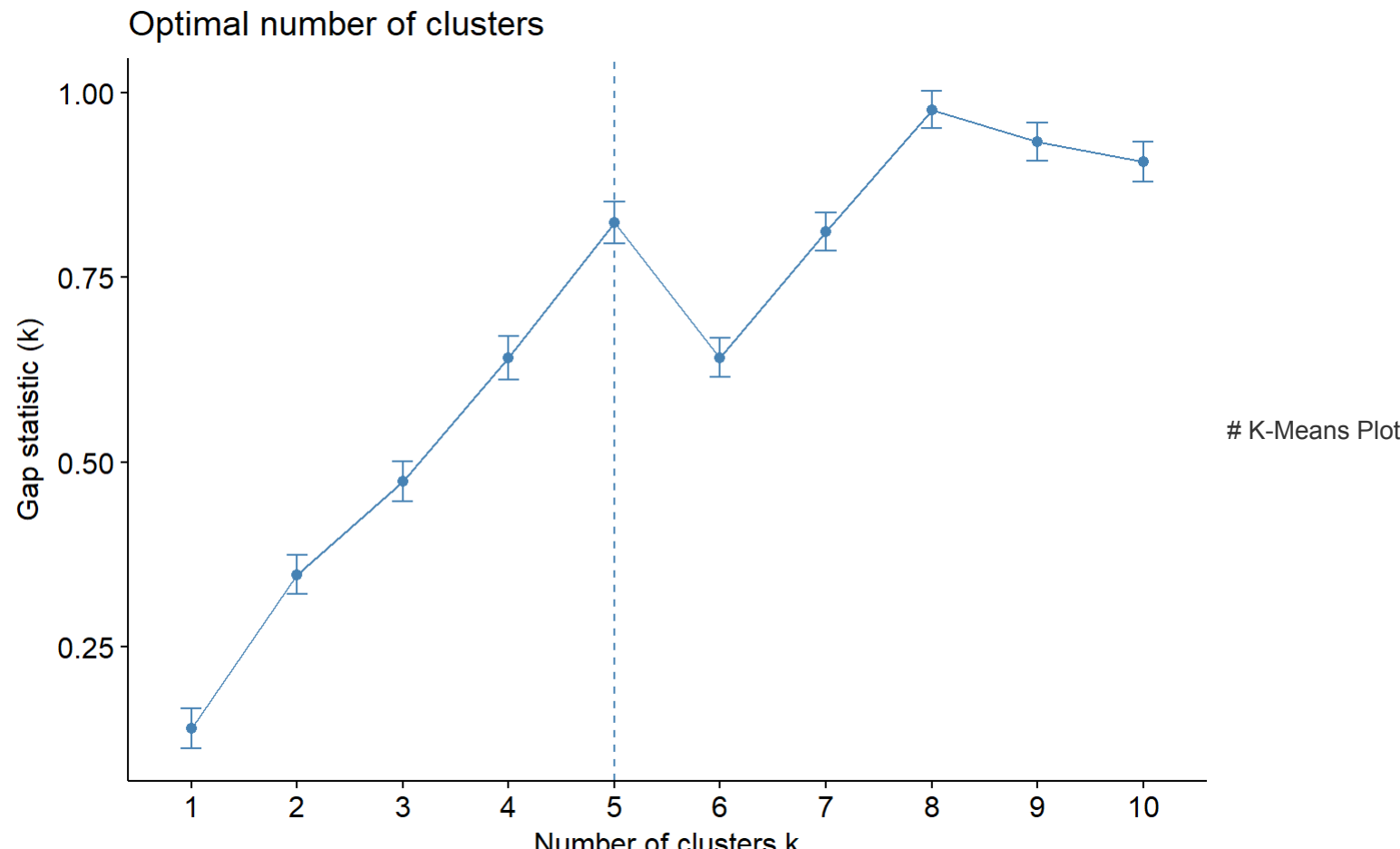
## Distance Matrix

- A distance matrix is a square matrix (two-dimensional array) containing the distances, taken pairwise, between the elements of a set.
- The 45 degree line pairs each element to itself. For example, sample 2-3-7-001 meets itself at top right corner and it is color coded blue with the least transparency in the heat map. The heat map represents the degree of similarity between elements of the matrix. The color bar value represent the distance. In this image, bright blue denotes a distance of 0 and bright Orange is maximal distance.

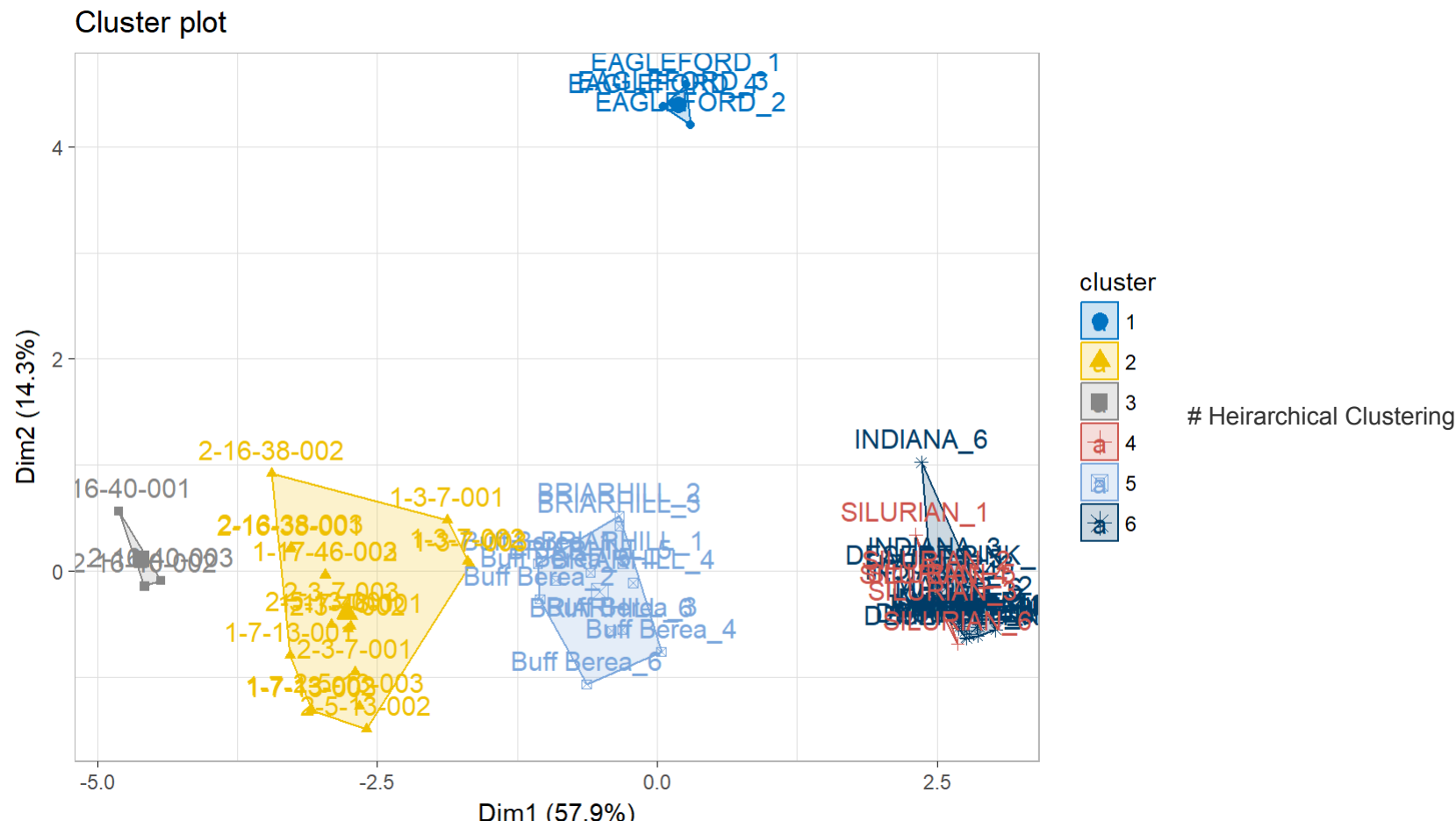


## Optimal Number Of Clusters Plot

The optimal number of clusters plot calculates number of cluster in the data. The gap statistic technique uses the output of any “clustering algorithm”, e.g. Kmeans, compares the change in within-cluster dispersion with the expected under an appropriate reference null distribution.

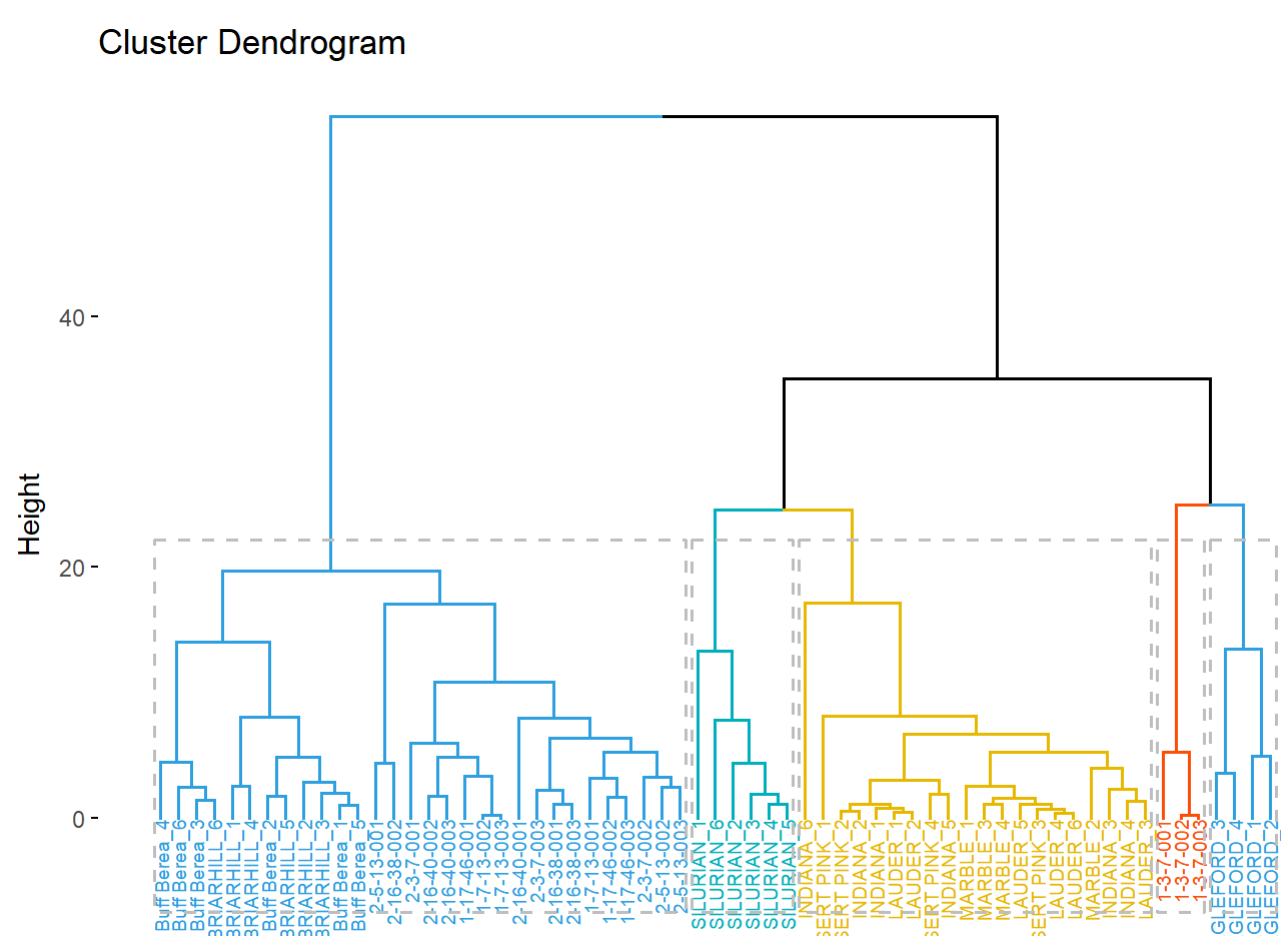


The data in our set are clustered by the k-means method, which aims to partition the points into k groups such that the sum of squares from points to the assigned cluster centers is minimized

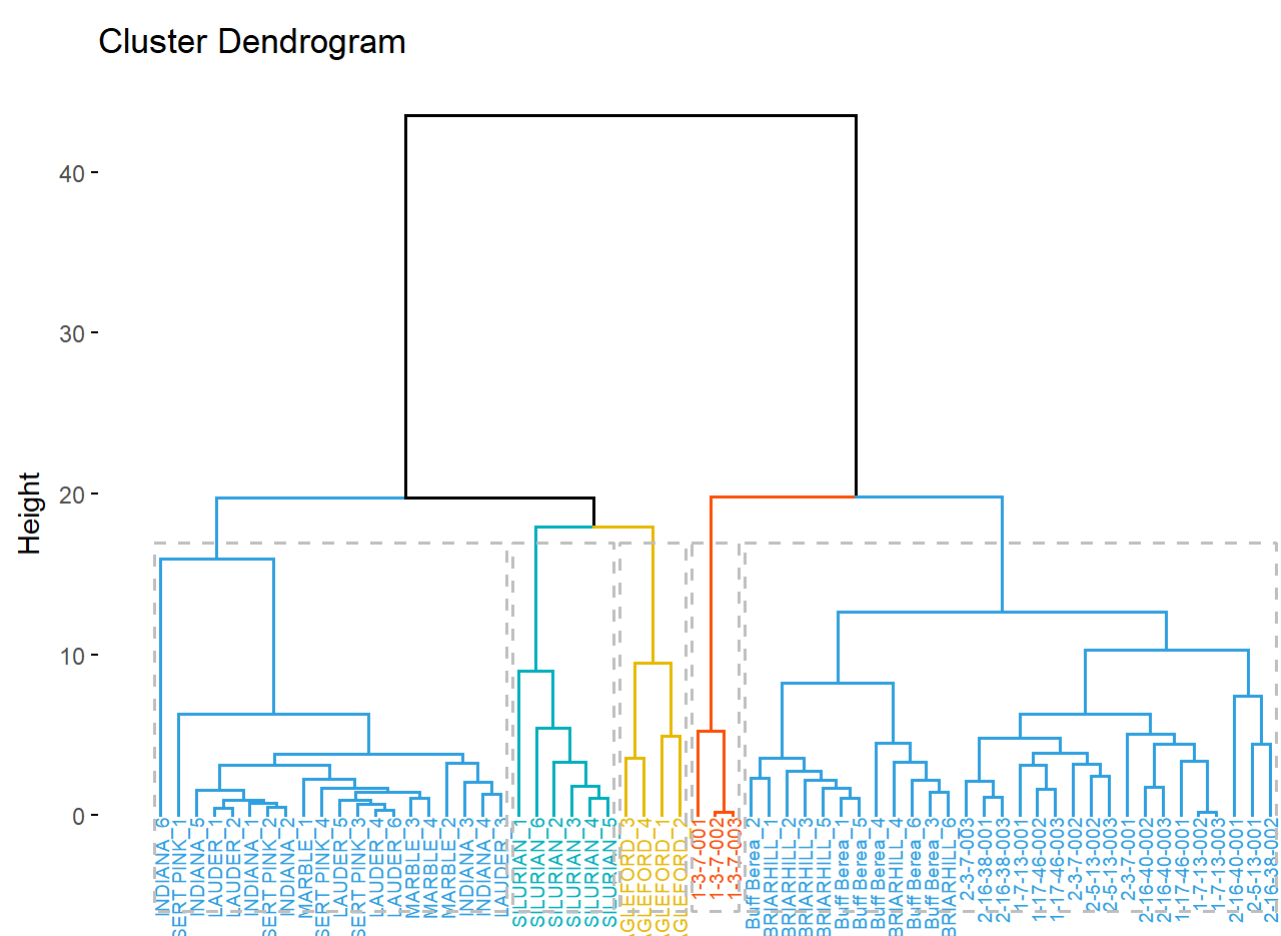


This function performs a hierarchical cluster analysis using a set of dissimilarities for the n objects being clustered. Initially, each object is assigned to its own cluster and then the algorithm proceeds iteratively, at each stage joining the two most similar clusters, continuing until there is just a single cluster. The vertical scale represents the numerical distance between samples.

Hierarchical clustering using the “complete method”



Hierarchical clustering using the “average method”



Hierarchical clustering using the “complete method”

