Hawaii and PRB Report

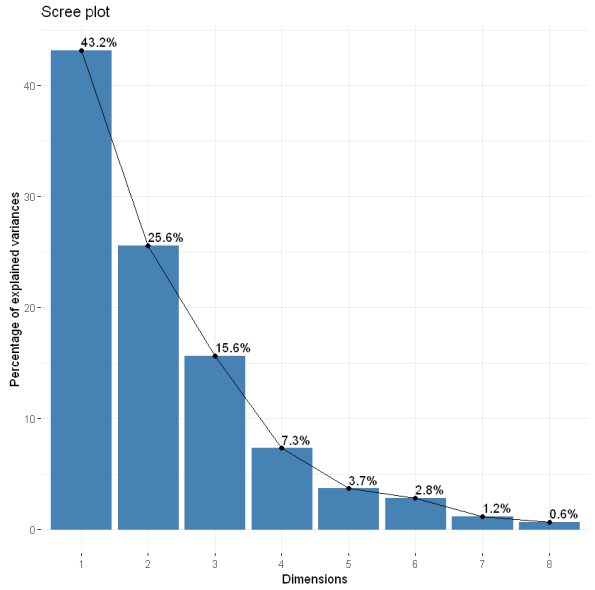
4/36/23 | Miro Manestar

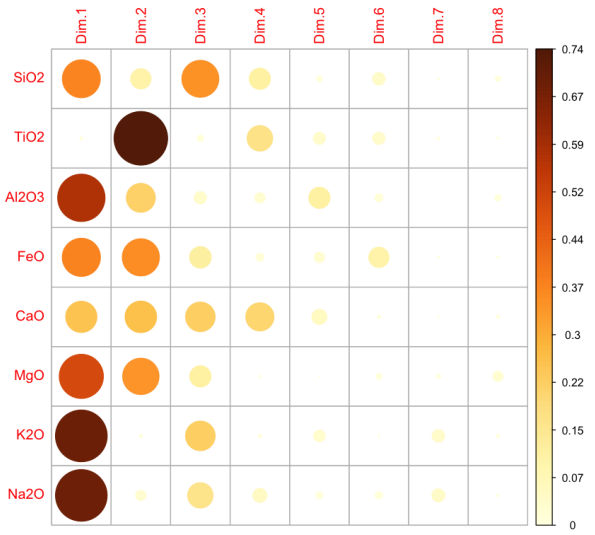
Section 1 – Overview

Analysis was done using the primary elements from the Hawaii and the PRB dataset. Samples with missing data points for any of these 8 variables was excluded. The number of usable samples was 10,476 for Hawaii and 532 for the PRB.

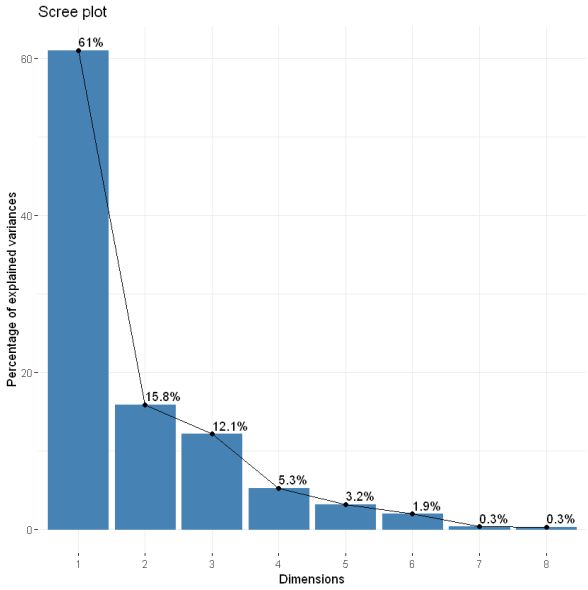
The used elements were as follows: **SiO_2, AlO_3, FeO, MgO, Na_2O, K_2O, TiO_2, CaO**

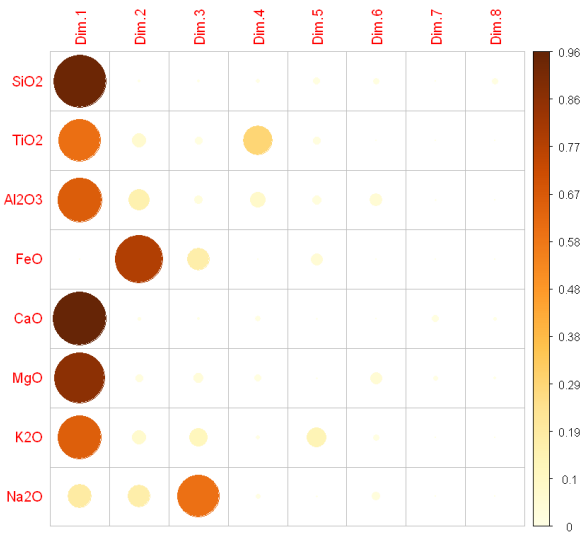
Hawaii Dataset (10,476 Samples)





PRB Dataset (532 Samples)

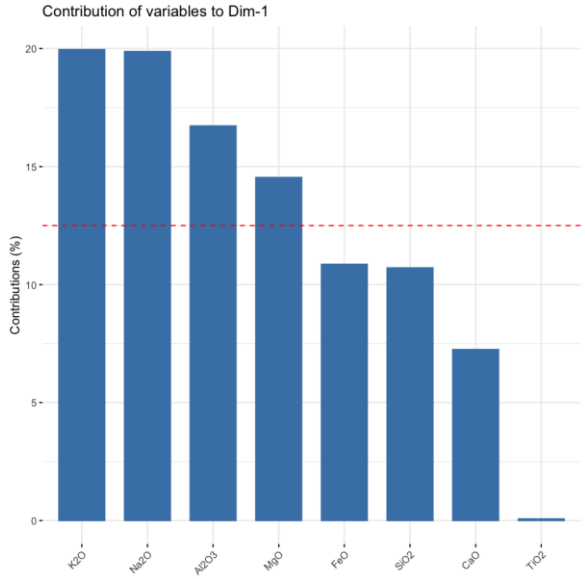


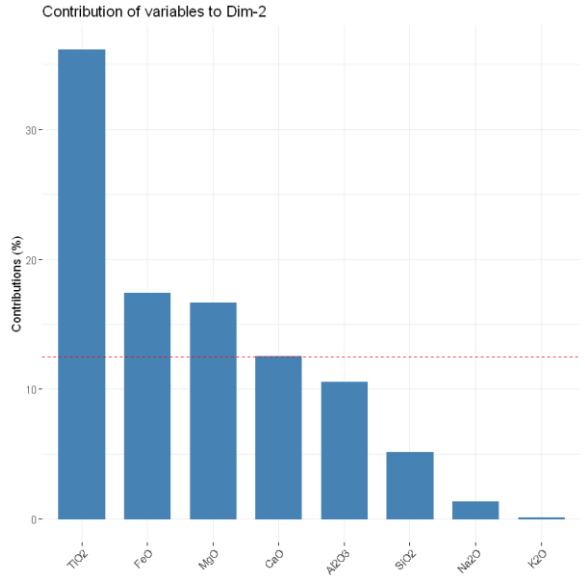


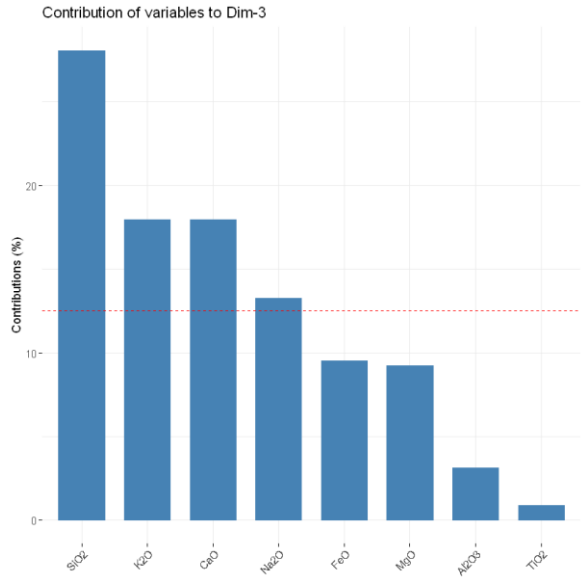
From the scree plots, only the first three dimensions in both datasets contain most of the variation. Thus, the accompanying analysis was done only up to three dimensions. PCA contributed to the K-Means clusters by reducing amount of noise the algorithm would have to content with, and thus allowing the clusters to only work on data which had a higher degree of variance, and therefore an increased likelihood of serving as good candidates for categorical discriminators.

Contributions

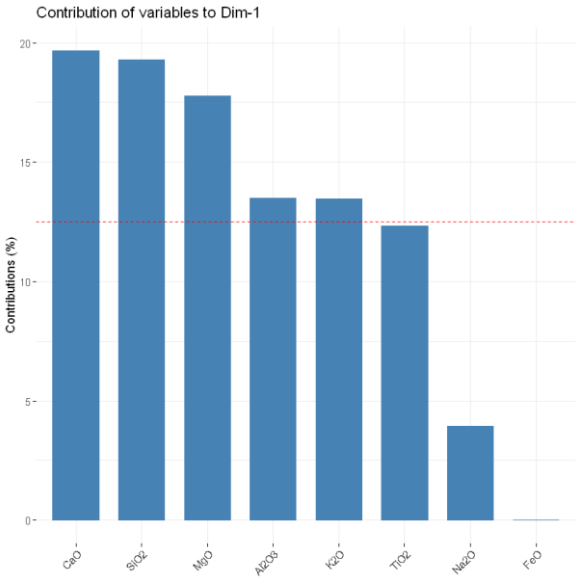
Hawaii Dataset (10,476 Samples)

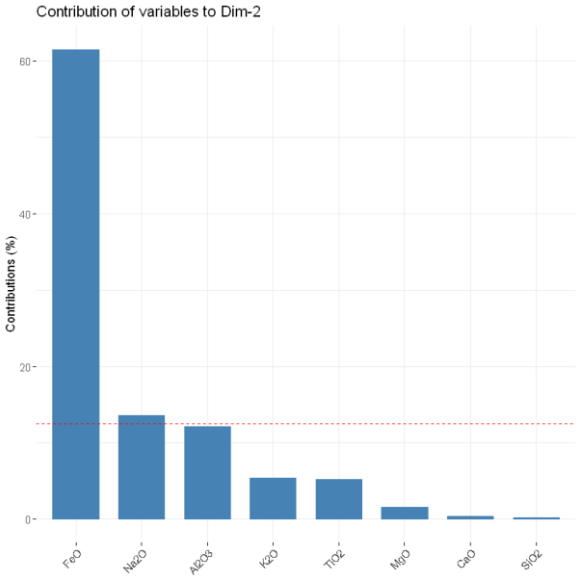


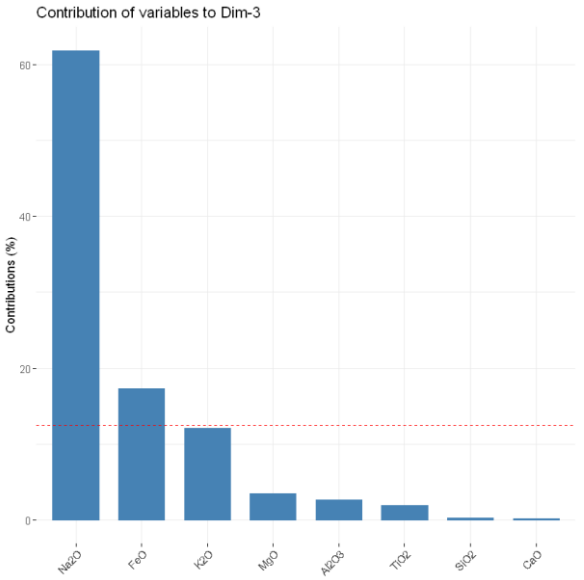




PRB Dataset (532 Samples)

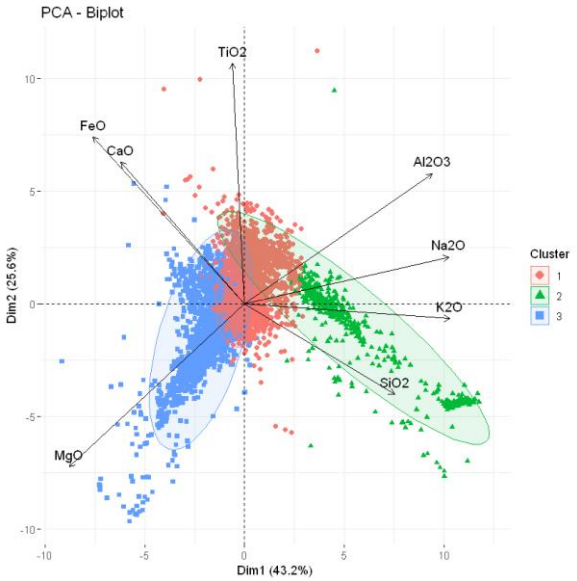


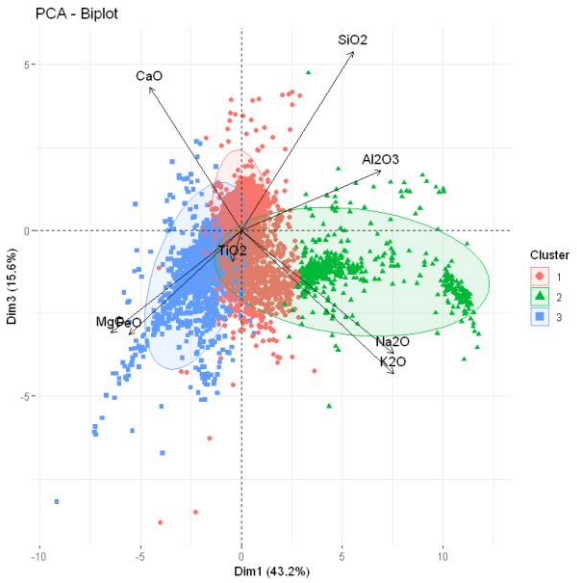


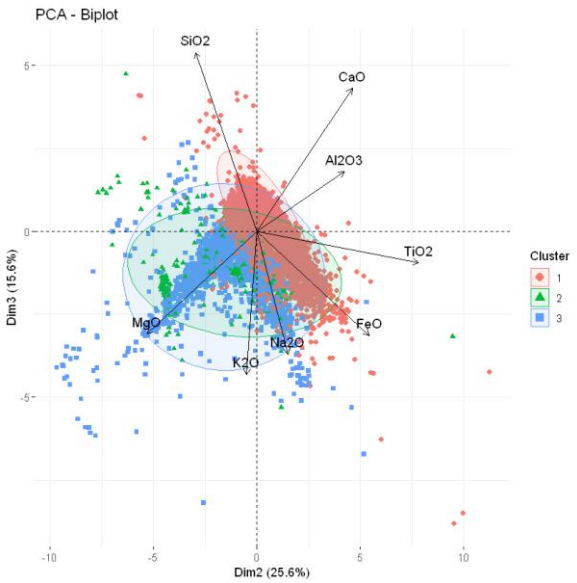


PCA Plots

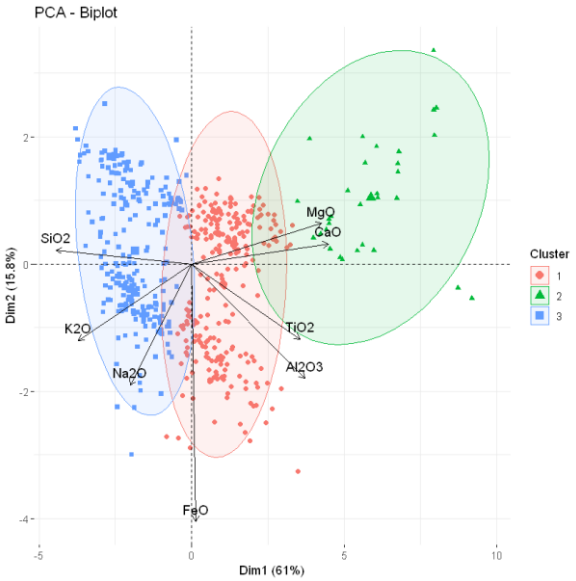
Hawaii Dataset (10,476 Samples)

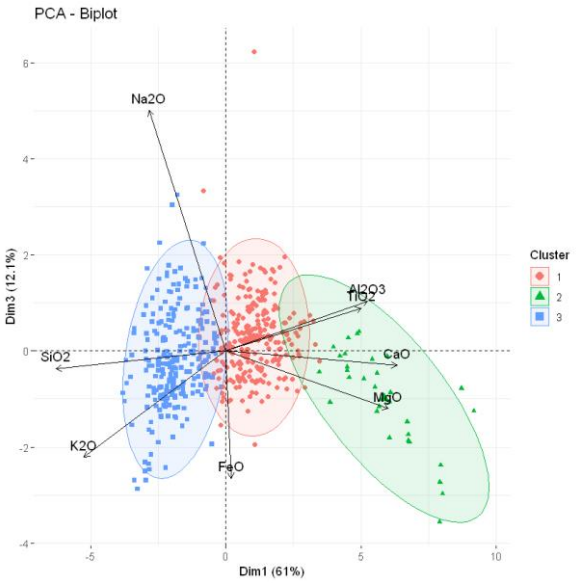


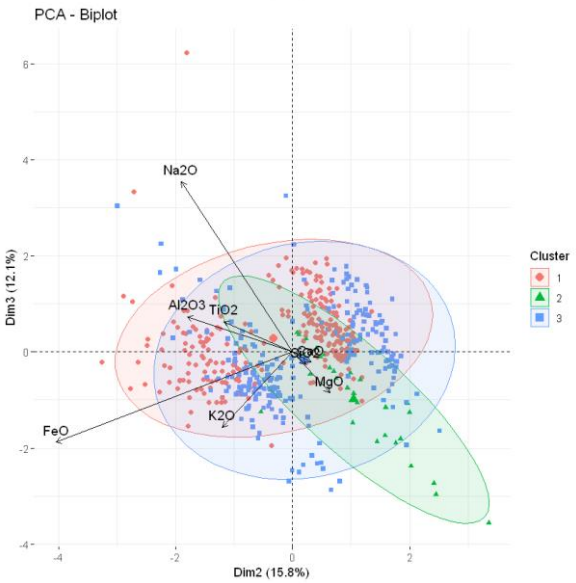




PRB Dataset (532 Samples)







Feature Averages by Cluster (Unnormalized)

Features Averages by Cluster – Hawaii (Unnormalized)

Cluster	SiO_2	TiO_2	Al_2O_3	FeO	CaO	MgO	K_2O	Na_2O
(Red) 1	49.91	2.64	13.78	11.42	10.60	7.30	0.54	2.50
(Green) 2	54.24	2.00	17.13	8.38	5.05	2.77	2.74	5.44
(Blue) 3	45.52	1.96	10.05	11.93	9.10	16.97	0.46	1.91

Features Averages by Cluster – PRB (Unnormalized)

Cluster	SiO_2	TiO_2	Al_2O_3	FeO	CaO	MgO	K_2O	Na_2O
(Red) 1	64.06	0.76	16.60	11.12	4.92	2.02	2.31	3.56
(Green) 2	50.64	0.92	18.62	7.94	10.69	6.13	0.47	2.28
(Blue) 3	72.07	0.33	14.65	9.33	2.13	0.59	3.83	3.62