Name: Joshua Damm

Student Number: s4036018

I kept 3 principal components from the previous exercise based on a combination of PVE, the elbow plot, and Kaiser's rule. Together they explained about 50% of the variance. Then I ran kmeans clustering for the first 3 selected principal components. For k = 2 the within sum of squares was 36.6% for k = 2 and 49.1% for k = 3. K = 2 and k = 3 also revealed the lowest withinSS and could explain the highest amount of variance. Especially PC1 does a good job in separating the data. K = 3 seems to find a good fit to separate between PC1 and PC2. For both k = 2 and k = 3 PC2 and PC3 seem explain not enough variation in the data to gain meaningful insights. For k = 2 48 observations are assigned to cluster 1 and 74 to cluster 2. For k = 3 33 observations are assigned to cluster 1, 54 to cluster 2 and 35 to cluster 3. There seems to be one cluster in the data that is loaded more than others, indicated by the heavier loadings on one cluster in both k = 2 and k = 3.

A diagram of a number of dots

Description automatically generatedA diagram of a number of dots

Description automatically generatedA diagram of a number of dots

Description automatically generatedA diagram of a number of dots

Description automatically generatedA diagram of a number of dots

Description automatically generatedA diagram of a number of dots

Description automatically generated

A diagram of links between several links

Description automatically generated with medium confidence

Finally ran hierarchical clustering with complete, average and single linkage. I chose to cut of the dendogram at 3 clusters. Interestingly, Identity, hidden.meanings, , upset, essential, and sensations seem to cluster high on every clustering method. Is is possible that these variables are related to other variables and explain variations in underlying variables. Further research is needed to fit these empirical results to the clinical literature and practice.