Noah Tigner

GEOG 495

Exercise 8

1. There are four components in this dataset. Component 4 appears to be the least important component. Each variable is related to at least two components to some degree.
2. Many of the vectors in the biplot are at right-angles revealing the structure of the data. There were roughly 10 loadings with significant correlations, consistent with the interpretation of the components.
3. There appears to be less of an overall pattern in the vector angles of the Factor Analysis, as the angles of the PCA are closer to right-angles. The dimensions are -2 to 3 and -0.5 to 0.5. The variable RcL (ReachLen) has loadings of very weak magnitudes and could likely be eliminated.
4. After rotating the factors, there are more loadings with correlations above 0.7. The biplot shows that the vector angles are closer to right-angles, and each of the three components are more separated and distinct from each other. For these reasons the rotated factors seem more interpretable.
5. The dimensions are -2 to 2 and -1 to 1. Again, RcL seems redundant, as well as BFW to a lesser extent, since its two loadings are of weak magnitude.
6. According to the Wilks Lambda statistic there is a large approximate F-statistic and a small p-value, so the groups are significantly different.
7. The probability of a stream being in the first group (healthy) is higher than the second (unhealthy), with a probability of 0,583 compared to 0.416, respectively. Knowing the value of variables DeepPools, LrgLWD, and PoolDepth would likely be enough to correctly classify the healthiness of a stream.

Noah Tigner

GEOG 495

Exercise 8

8. Surprisingly, using two clusters provided more homogeneity than using 3. Using 4, 5, 6, 7, and even 8 clusters led to bimodal histograms. Increasing the number of clusters did not seem to make them more homogenous.