Homework: Week 4 (Due Monday, October 21, 2024)

FW 599: Multivariate Analysis of Ecological Data

Instructions

Please submit all homework assignments to Canvas as an **R markdown document** (Markdown or Quarto) including **visible code** and **relevant output**. A tidy *.pdf document is preferable to an *.html or a "raw" *.qmd file. Note that homework questions are intended to directly accompany lab exercises, building up to the final class project. Consequently, it is in your best interest to answer them thoroughly and thoughtfully.

Questions

Question 1) Explain the different linkage methods (single, complete, average, and Ward's) used in agglomerative hierarchical clustering. How does each method affect the shape and size of the resulting clusters?

Question 2) Discuss the implications of choosing different association coefficients in agglomerative hierarchical clustering. How do these choices impact the clustering results?

Question 3) Try each of the agglomerative hierarchical clustering methods with your own dataset. Which one appears to perform best? What is the optimal number of clusters? Verify this using the appropriate non-statistical and statistical approaches.

Question 4) Do you see any evidence that noise and outliers are impacting your results? If so, how could you treat the data differently to account for these sources of error?

Question 5) Discuss the advantages and disadvantages of using divisive hierarchical clustering over agglomerative hierarchical clustering in ecological data analysis. When you run a divisive analysis (DIANA) using your data, how do the results differ from the agglomerative clustering output?

Question 6) Explain the k-means clustering algorithm and list key steps. When you run a k-means partitioning analysis using your data, how do the results compare to the hierarchical clustering output(s)? Based on the sum of squares error, do you think the resulting solution makes sense?

Question 7) Discuss the role of initialization in k-means clustering. How do different initialization methods impact the convergence and final results of the algorithm?

Question 8) Explain the concept of indicator species in ecological analysis. How are indicator species identified and used to interpret ecological data? Do there appear to be any indicator species in your data, and if so, for what site groups and are the indicator values significant?