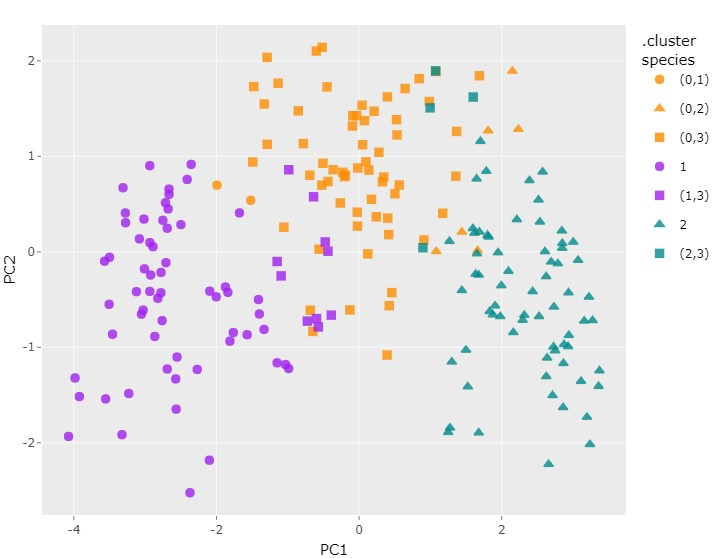
**Advanced Clustering Methods****Introduction**

This R script performs advanced clustering analysis on a dataset of seeds. The analysis includes Principal Component Analysis (PCA), K-means clustering, hierarchical clustering, and visualization of clustering results. Below is a brief overview of the steps performed in this script.

**Steps**

**Data Preprocessing**

1. Load necessary libraries: tidyverse, tidymodels, dplyr, plotly, factoextra, cluster.
2. Read the dataset from a CSV file, exclude the "groove\_length" column, and convert the "species" column to a factor.

**Principal Component Analysis (PCA)**

1. Create a PCA recipe to extract two principal components.
2. Preprocess the data using the recipe and obtain the PCA features.
3. Visualize the variance explained by each principal component.
4. Visualize the PCA scores in a 2D scatter plot.

**K-means Clustering**

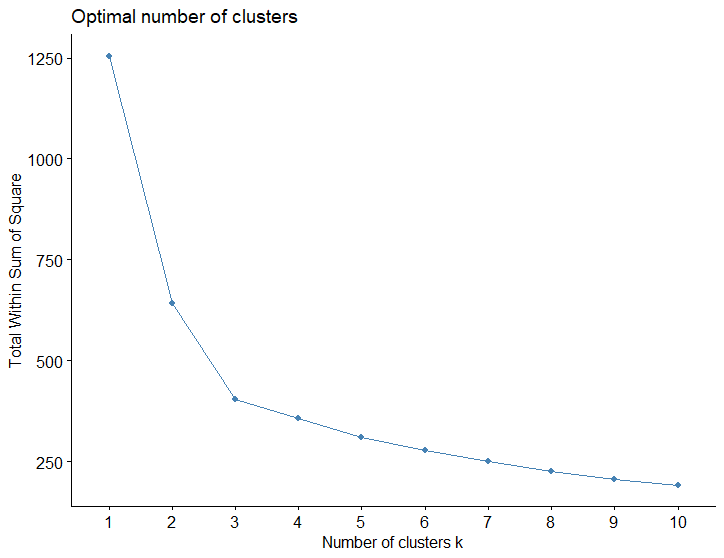
1. Preprocess the data by removing the target column (species) and normalizing the data.
2. Create 10 K-means models with cluster counts ranging from 1 to 10.
3. Plot the total within-cluster sum of squares (tot.withinss) to determine an appropriate number of clusters (K).

**Hierarchical Clustering**

1. Calculate the Euclidean distance matrix between observations.
2. Perform hierarchical clustering using complete, average, and Ward linkage methods.
3. Visualize the hierarchical clustering dendrograms.

**Optimal Number of Clusters**

1. Determine the optimal number of clusters using the within-cluster sum of squares (WSS) method.
2. Visualize the clustering structure for three clusters based on the Ward linkage.



**Comparing K-means and Hierarchical Clustering**

1. Group the data into three clusters using hierarchical clustering.
2. Compare the cluster assignments between K-means and hierarchical clustering.
3. Plot the hierarchical clustering results on the PCA data.

**Running the Script**

To run this script, follow these steps:

1. Ensure that you have the required libraries installed (tidyverse, tidymodels, dplyr, plotly, factoextra, cluster).
2. Load the script in your R environment.
3. Modify the file path in the read\_csv function to point to your dataset (seeds.csv).
4. Execute the script step by step or all at once.
5. The script generates various plots and outputs to help you analyze the clustering results.

Please note that this script assumes you have the necessary R packages and a dataset named "seeds.csv." Adjustments may be needed based on your specific dataset and requirements.

For additional details and interactivity, consider using RMarkdown or exporting the script as an RMarkdown document.