Anthony Cunningham

DATA 824

**Assignment 13**

This assignment uses the wine dataset (from Kaggle, [link](https://www.kaggle.com/datasets/harrywang/wine-dataset-for-clustering?resource=download)), originally sourced from UCI Machine Learning repository. It contains data regarding the chemical composition of various wines.

From the description on the Kaggle project: “These data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines.”

One reason for using clustering could be to predict the type of an unknown wine. Another could be to create segments of wine that can be mapped with market segments or preference segments (e.g. perhaps more acidic wines are more favorable among American consumers, but less favorable among French consumers; clustering wines in this way could act to segment wine supplies along this, and other, lines in order to maximize productivity and future wine-making production).

Using PCA, it was found that relative Magnesium and Proline concentrations help to explain most of the differences among the 176 wines in this sample. Similar results were found using the SVD (w. imputation) and Nipals PCA methods.

Table

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart, bar chart

Description automatically generated