

Unsupervised Learning and Clustering

Open the "*Cereals.csv*" file. Answer the following questions.

1. Using all of the variables, except *name* and *rating*, run the *k*-means algorithm with $k = 5$ to identify clusters within the data.
2. Develop clustering profiles that clearly describe the characteristics of the cereals within the cluster.
3. Rerun the *k*-means algorithm with $k = 3$.
4. Which clustering solutions you prefer and why?
5. Develop clustering profiles that clearly describe the characteristics of the cereals within the cluster.
6. Use cluster membership to predict *rating*. One way to do this would be to construct a histogram of rating based on cluster membership alone. Describe how the relationship you uncovered makes sense, based on your earlier cluster profiles.