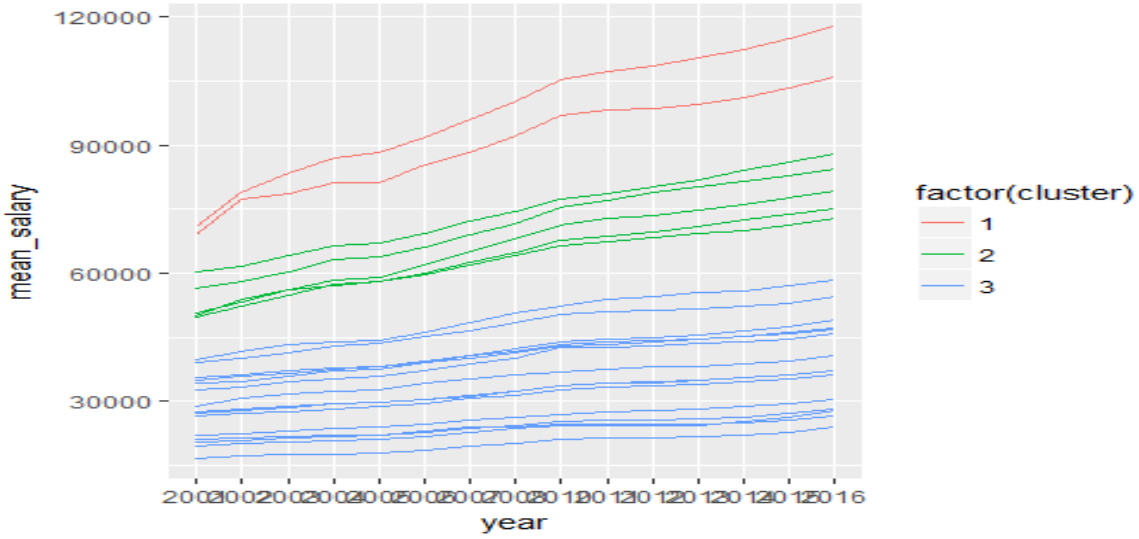
This document gives brief overview of the business problem and the programming approach to solve the same. The dataset here consists of average incomes for 20 occupations prevalent in USA over the period of 15 years starting 2001. We would like to perform an unsupervised learning here in terms of explore the likelihood of identifying any clusters among these 20 career tracks in terms of their respective incomes levels over this period.

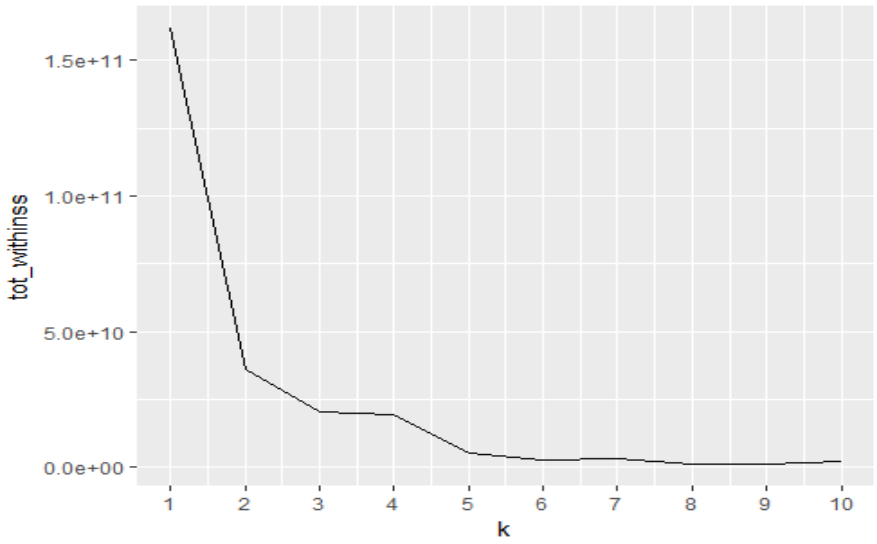
We apply Hierarchical, KMeans and Silhouette approaches to identify the possible clusters in term of income trends. We get to see 3, 2 and 7 clusters if we choose to go by the model output of Hierarchical, KMeans and Silhouette. However, as we know the cluster count is subject to the given context of the business problem and one can consume the data inferences as suitable and intended. We see Management and Legal to be falling in the same top cluster with impressive incremental pattern as compared to the other professions.

Below are the snippets of the dendrogram and plots for KMeans and Silhouette outputs.

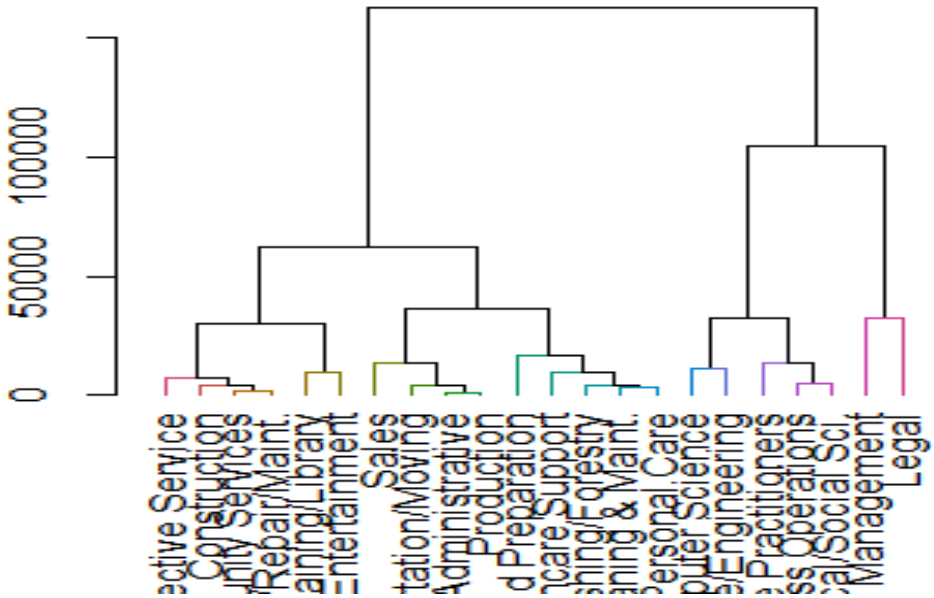
1. **Hierachical Clustering with cuts at 90000 height (income)**



2. **KMeans Clustering: Euclidean, Average**



3. Hierachical Clustering denrdogram with cuts at 90000 height (income)



4. **Silhouette Average Width approach**