**AE5**

In this assignment, the data of people who shop at an online grocery store was analyzed in order to cluster the customers based on age, income, and a "spending score" which we assume to be a score assigned by the grocer. A higher score seems to be desirable in this case. Based on a complete linkage, the data was cut into 5 clusters. Gender and customerID were left out of this analysis as Gender was already categorized. Each cluster seems to be distinct, cluster 2 seems to have an average age of 24.8, with a household income of $25,000 and a spend score of 81, whereas cluster 5 has an average age of 42, a household income of $93,000 and a spend score of 18.2. We can make no assumptions regarding the correlation between age, income and spending score. What the data does tell us is that cluster 4, with an average household income of $92,300 and a spend score of 82.1, is a very attractive group to focus on as they may be big spenders at the store.

The k-means method suggests that 3-5 clusters are ideal for this analysis. These two methods cannot be compared.
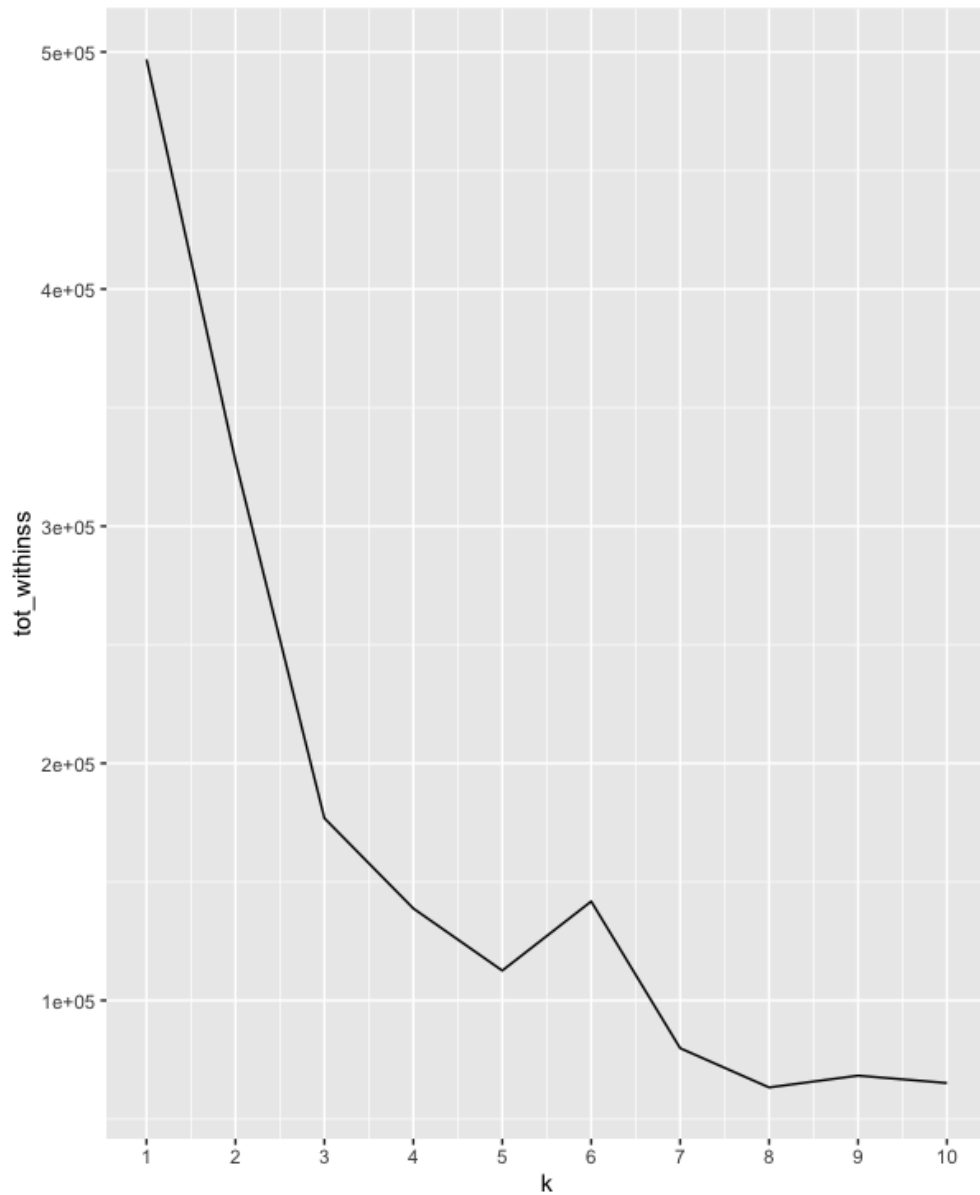
Elbow Plot

```
# A tibble: 5 x 4
  cluster    age HouseHoldIncome10k SpendingScorePerWeek
*   <int> <dbl>              <dbl>                <dbl>
1       1  45.2               26.3                 20.9
2       2  24.8               25.0                 81
3       3  42.2               54.8                 49.8
4       4  32.3               92.3                 82.1
5       5  42.0               93.8                 18.2
> clust_pick<- mutate(df_pick, cluster = cut_pick)
```

```
> count(segment, cluster)
  cluster  n
1        1 23
2        2 20
3        3 83
4        4 79
5        5 73
```