Hierarchical Clustering Algorithm

and Its Application in Business

Kristina Pestaria Sinaga

Master in Information System Management
Bina Nusantara University

January 07, 2021

Summary

1. What is Hierarchical Clustering?

2. Discussion
   - Customer Segmentation From an E-commerce Site

What is Hierarchical Clustering?

Types of Clustering

1. Hierarchical Clustering
   - Agglomerative
   - Divisive
2. Partial Clustering
   - K-Means
   - K-Medoids
   - Fuzzy C-Means
   - Possibilistic C-Means

Definition

This is the most common method of clustering. It creates a series of models with cluster solutions from 1 (all cases in one cluster) to n (each case is an individual cluster). This approach also works with variables instead of cases. Hierarchical clustering can group variables together in a manner similar to factor analysis.
Finally, hierarchical cluster analysis can handle nominal, ordinal, and scale data. But, remember not to mix different levels of measurement into your study.
Source: alchemer.com/

Steps to Perform Agglomerative Hierarchical Clustering

Treat each data point as single cluster. Hence, we will be having, say K clusters at start. The number of data points will also be K at start

$\Downarrow$

Now, in this step we need to form a big cluster by joining two closet datapoints. This will result in total of K-1 clusters

$\Downarrow$

Now, to form more clusters we need to join two closet clusters. This will result in total of K-2 clusters.

Steps to Perform Agglomerative Hierarchical Clustering

Treat each data point as single cluster. Hence, we will be having, say K clusters at start. The number of data points will also be K at start

$\Downarrow$

Now, in this step we need to form a big cluster by joining two closet datapoints. This will result in total of K-1 clusters

$\Downarrow$

Now, to form more clusters we need to join two closet clusters. This will result in total of K-2 clusters.

Steps to Perform Agglomerative Hierarchical Clustering

Treat each data point as single cluster. Hence, we will be having, say K clusters at start. The number of data points will also be K at start

$\Downarrow$

Now, in this step we need to form a big cluster by joining two closet datapoints. This will result in total of K-1 clusters

$\Downarrow$

Now, to form more clusters we need to join two closet clusters. This will result in total of K-2 clusters.

Steps to Perform Agglomerative Hierarchical Clustering

Now, to form one big cluster repeat the above three steps until K would become 0 i.e. no more data points left to join

$\Downarrow$

At last, after making one single big cluster, dendrograms will be used to divide into multiple clusters depending upon the problem

Source: tutorialspoint.com

Steps to Perform Agglomerative Hierarchical Clustering

Now, to form one big cluster repeat the above three steps until K would become 0 i.e. no more data points left to join

$\Downarrow$

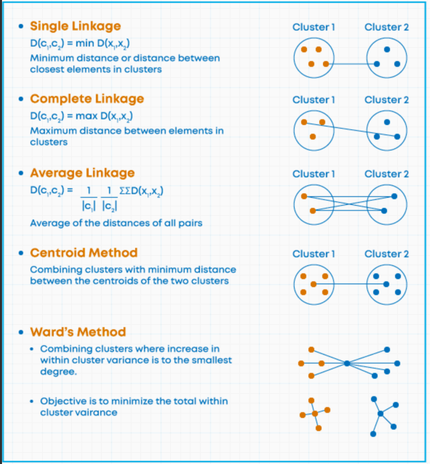
At last, after making one single big cluster, dendrograms will be used to divide into multiple clusters depending upon the problem
Source: tutorialspoint.com

Difference ways to measure the distance between two clusters

There are several ways to measure the distance between in order to decide the rules for clustering, and they are often called Linkage Methods. Some of the popular linkage methods are:

- Single Linkage
- Complete Linkage
- Average Linkage
- Centroid Linkage
- Ward's Linkage

Popular Linkage Methods



Source: dataaspirant.com

An Example of Hierarchical Clustering

Let's dive into one example to best demonstrate Hierarchical clustering

	Income	Spend
Customer 1	233	150
Customer 2	250	187
Customer 3	204	172
Customer 4	236	178
Customer 5	236	178
Customer 6	354	163
Customer 7	192	148
Customer 8	294	153
Customer 9	263	173
Customer 10	199	162

Customer Dataset

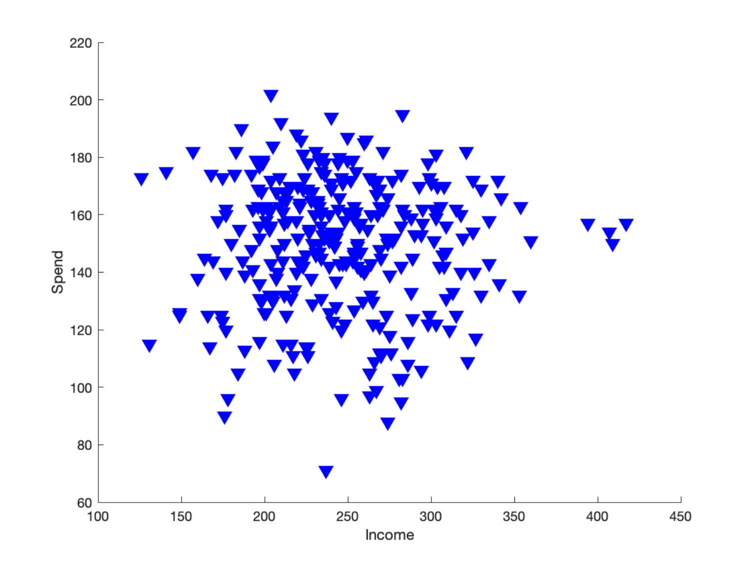
Objective

This presentation will demonstrates the concept of segmentation of a customer data set from an e-commerce site using Hierarchical clustering Algorithm in Matlab and R Language. I will use the Hierarchical clustering algorithm for creating customer segments based on their income and spend data.

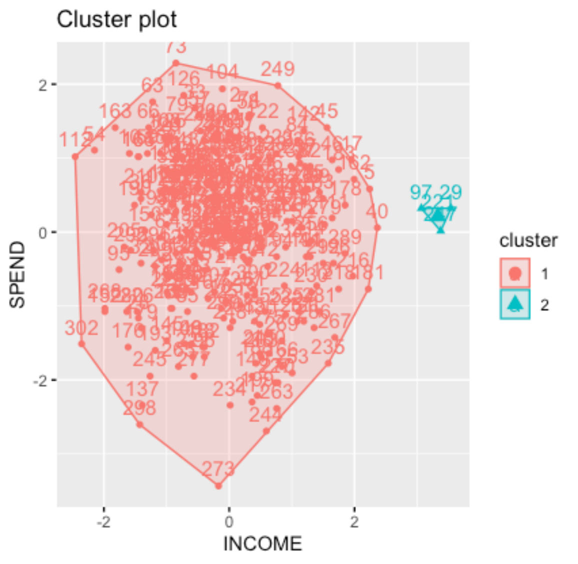
About the Dataset

The dataset consists of Annual income (in $\$000$) of 303 customers and their total spend (in $\$000$) on an e-commerce site for a period of one year.
Source: towardsdatascience.com/

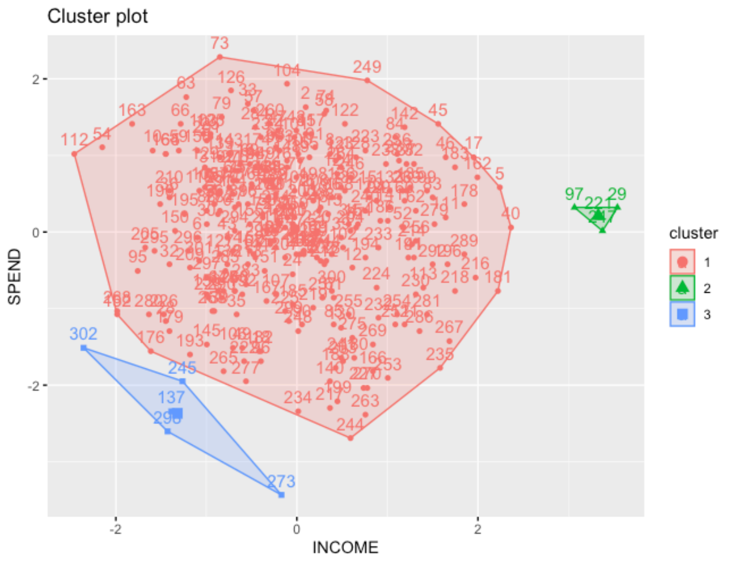
Scatterplot: E-commerce Customer



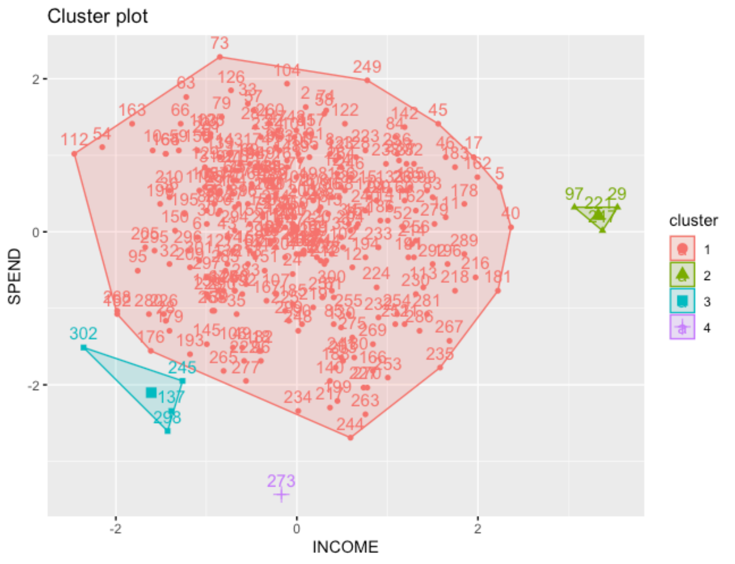
E-commerce Customer Data Scale with C=2



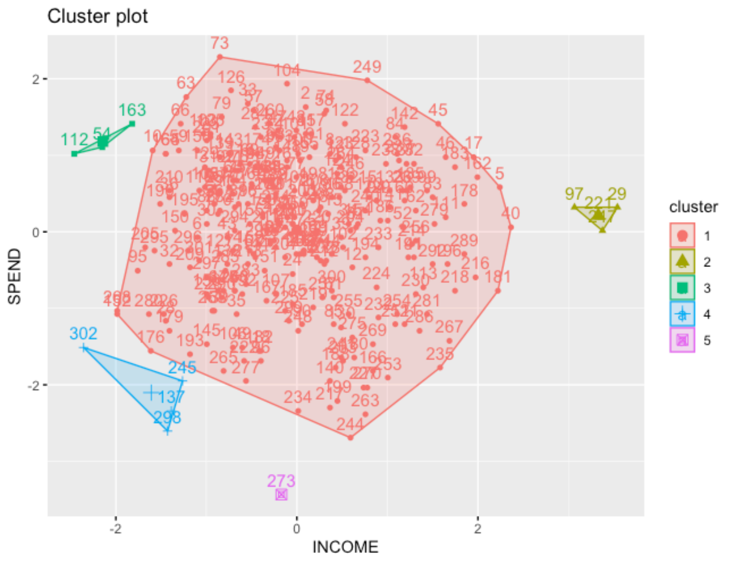
E-commerce Customer Data Scale with C=3



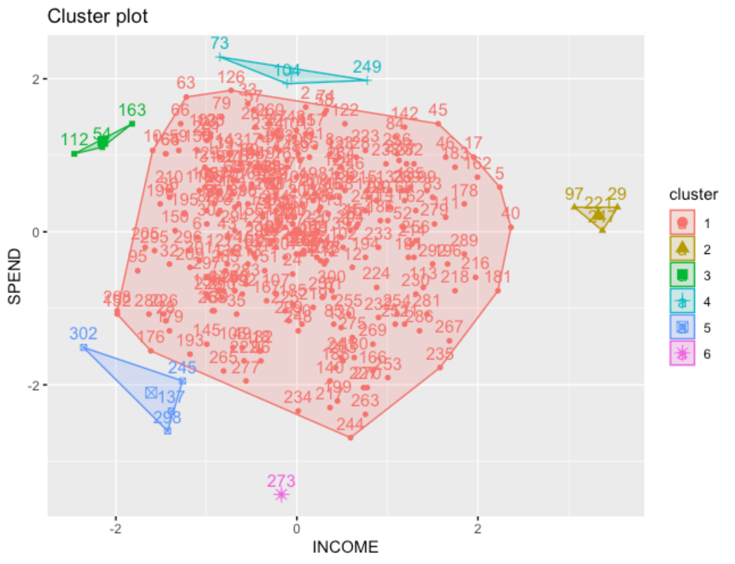
E-commerce Customer Data Scale with C=4



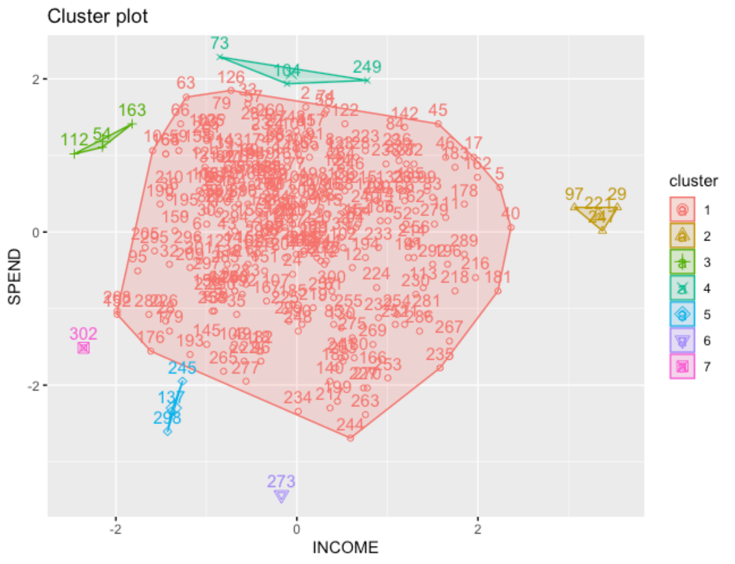
E-commerce Customer Data Scale with C=5



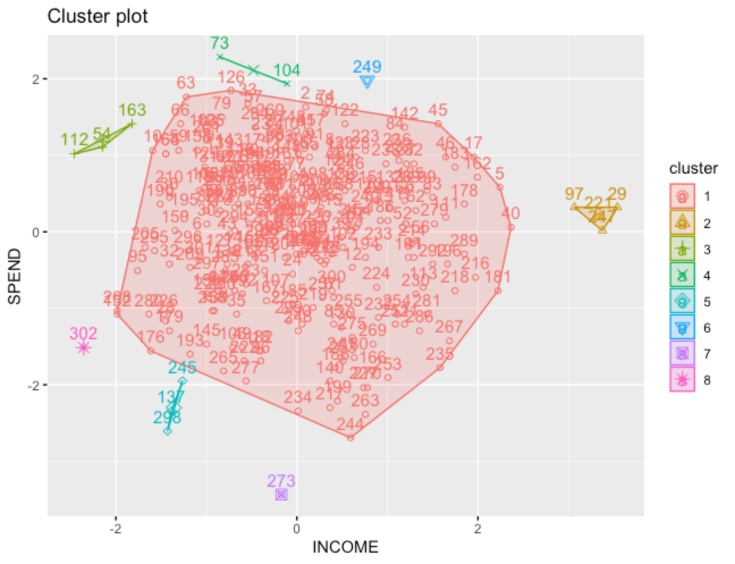
E-commerce Customer Data Scale with C=6



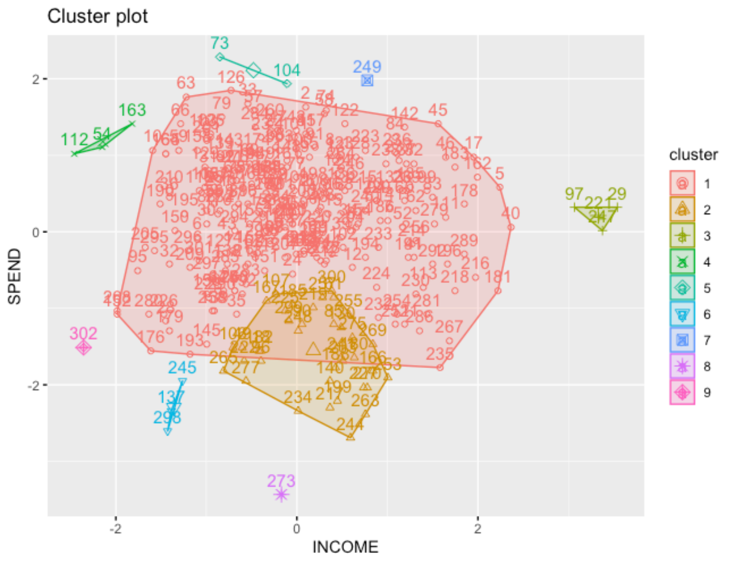
E-commerce Customer Data Scale with C=7



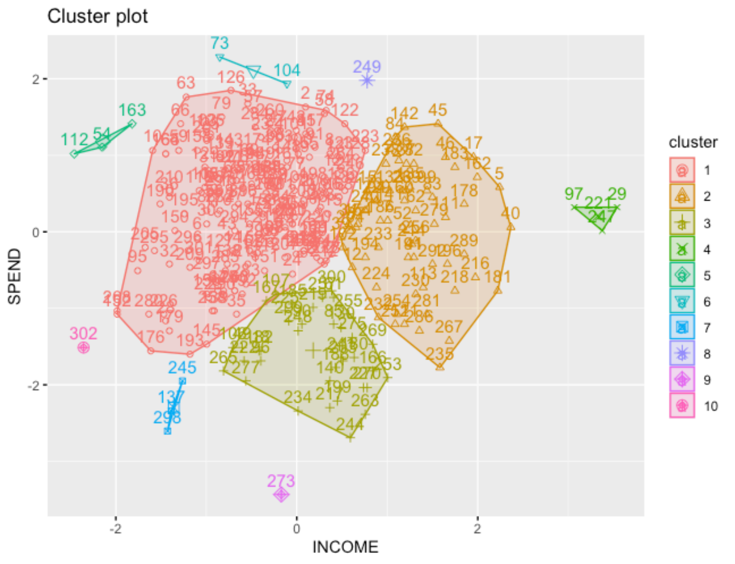
E-commerce Customer Data Scale with C=8



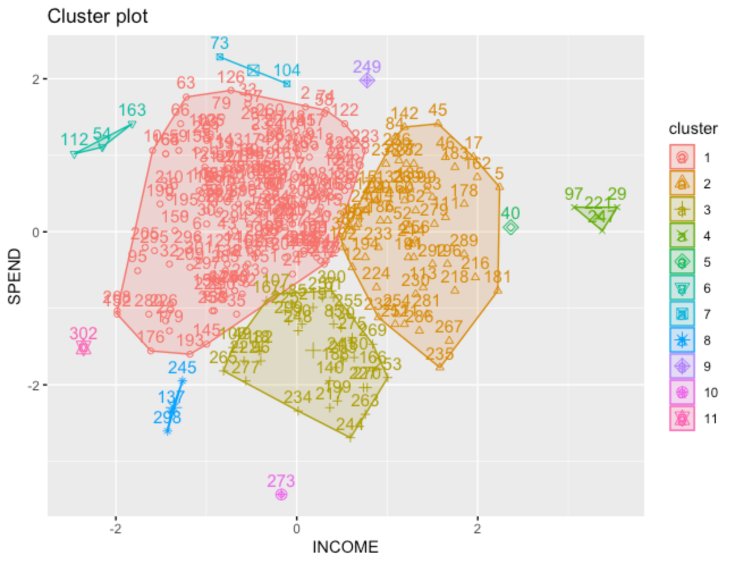
E-commerce Customer Data Scale with C=9



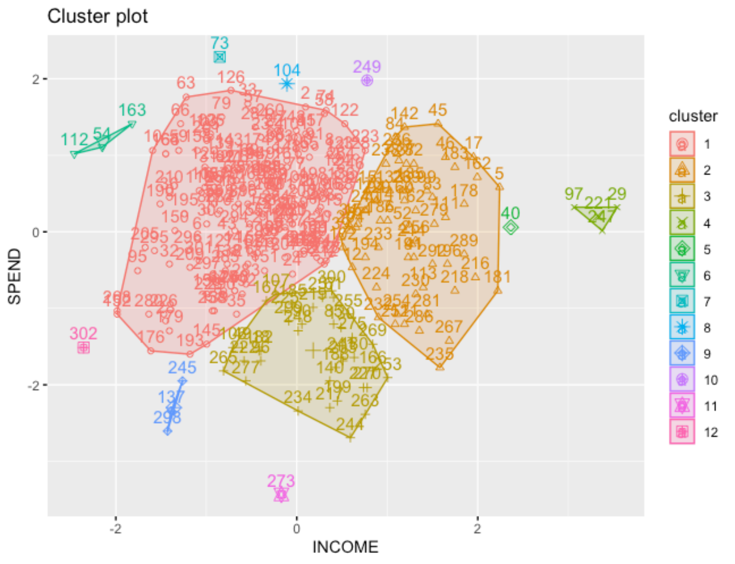
E-commerce Customer Data Scale with C=10



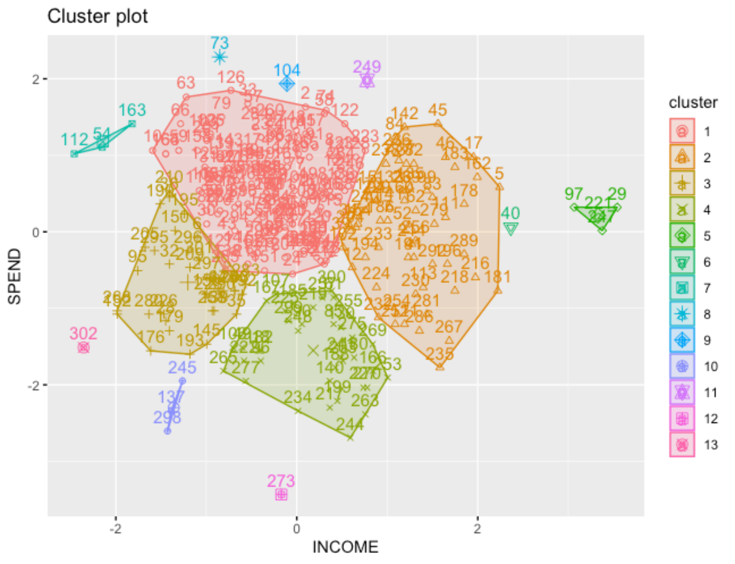
E-commerce Customer Data Scale with C=11



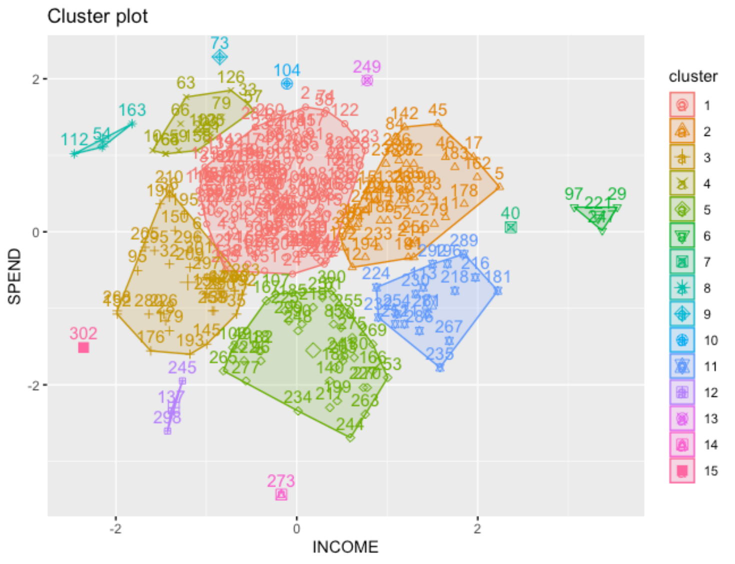
E-commerce Customer Data Scale with C=12



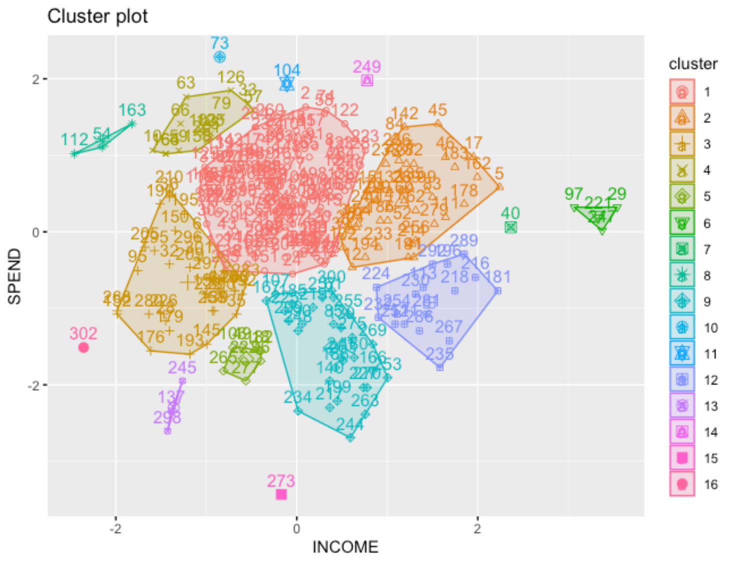
E-commerce Customer Data Scale with C=13



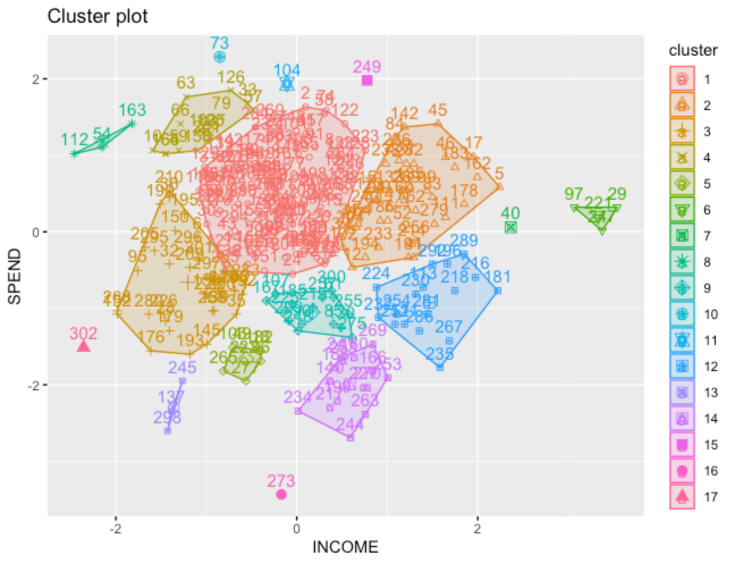
E-commerce Customer Data Scale with C=15



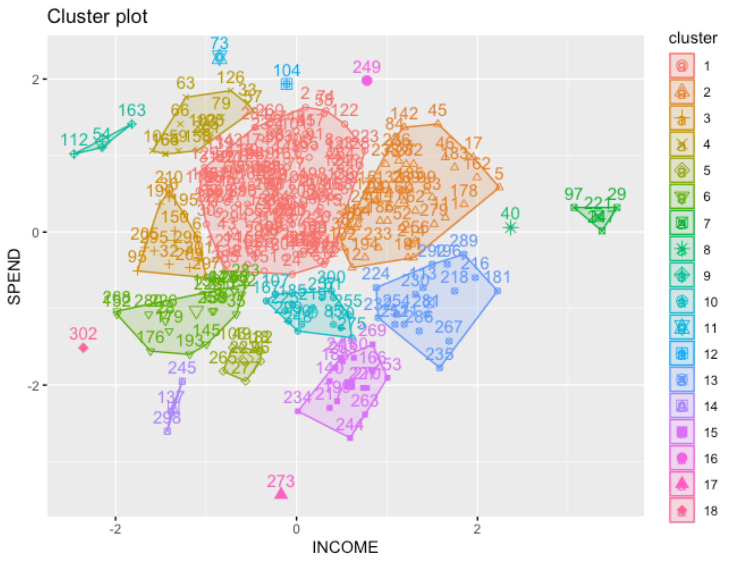
E-commerce Customer Data Scale with C=16



E-commerce Customer Data Scale with C=17



E-commerce Customer Data Scale with C=18



Discussion

Customer Segmentation From an E-commerce Site

E-commerce Customer Segmentation with C=2

Please interpret them as the following customer segments:

1. Cluster 1: Customers with medium annual income and low annual spend
2. Cluster 2: Customers with high annual income and medium to high annual spend

E-commerce Customer Segmentation with C=3

Please interpret them as the following customer segments:

1. Cluster 1: Customers with medium annual income and low annual spend
2. Cluster 2: Customers with high annual income and medium to high annual spend
3. Cluster 3: Customers with low annual income

E-commerce Customer Segmentation with C=4

Please interpret them as the following customer segments:

1. Cluster 1: Customers with medium annual income and low annual spend
2. Cluster 2: Customers with high annual income and medium to high annual spend
3. Cluster 3: Customers with low annual income
4. Cluster 4: Customers with medium annual income but high annual spend

E-commerce Customer Segmentation with C=6

Please interpret them as the following customer segments:

1. Cluster 1: Customers with Medium income, low annual spend
2. Cluster 2: Customers with Very high income, high annual spend
3. Cluster 3: Customers with High income, high annual spend
4. Cluster 4: Customers with Low income, high annual spend
5. Cluster 5: Customers with Medium income, low annual spend
6. Cluster 6: Customers with Low income, low annual spend

Perspective

Please Tell Me Your Conclusion Based on Segmentation Results

References

- Asuncion, A. & Newman, D. *UCI machine learning repository.* 2007.
- Cheng, C.-H. & Chen, Y.-S. Classifying the segmentation of customer value via RFM model and RS theory. *Expert systems with applications* **36,** 4176–4184 (2009).
- Park, H.-S. & Jun, C.-H. A simple and fast algorithm for K-medoids clustering. *Expert systems with applications* **36,** 3336–3341 (2009).
- You, Z. *et al.* A decision-making framework for precision marketing. *Expert Systems with Applications* **42,** 3357–3367 (2015).

The End