

IMSE 586 PROJECT ASSIGNMENT

Part A: Getting to Know Minitab and Cluster Analysis

Team:

Suresh Ooty Krishnaswamy

Selvamani Masilamani

Contents

Part A: Getting to Know Minitab and Cluster Analysis	1
Team Effort:	4
Data:.....	4
A1: Study of Variables using Pearson Correlation on Actual Values	5
A2: Study of Variables using Pearson Correlation on Standardized Values.....	7
A3: Cluster Analysis on Observations using “Single Linkage” and five distance measures (Actual Values).....	8
A4: Cluster Analysis on Observations using “Single Linkage” and five distance measures (Standardized Values)	13
A5: Cluster Analysis on Observations using “Euclidean Distance” measure and all linkage methods (Actual Values)	19
A6: Cluster Analysis on Observations using “Euclidean Distance” measure and all linkage methods (Standardized Values)	27
A7: Cluster Analysis on Variables (Actual Values).....	34
Findings:	34
A8: Cluster Analysis on Variables (Standardized Values).....	34
A9: Cluster Analysis on Observations for variables XC1 (Actual Values)	35
A10: Cluster Analysis on Observations for variables XC1 (Standardized values)	37
A11: Cluster Analysis on Observations for variables XC2 (dropped X5) (Actual Values)	39
A12: Cluster Analysis on Observations for variables XC2 (standardized values)	41
A13: Cluster K-means (actual Values)	43
A14: Cluster K-means (standardized Values)	45
REFERENCES:	48

TABLES

Table 1 : Comparison of A3 & A4.....	17
Table 2 : Comparison of A3 & A5	25
Table 3 : Comparison of A4, A5 & A6.....	31
Table 4 : Comparison of A7 & A8.....	35
Table 5 : Comparison of A9 & A3	35
Table 6 : Comparison of A2, A4 & A10.....	37
Table 7 : Comparison of A3 & A11	39
Table 8 : Comparison of A2, A4 & A12.....	41
Table 9 : Comparison of k-means 2, 3 & 4 clusters	43
Table 10 : Comparison of A3 & A5 only single –Euclidean	44
Table 11 : Comparison of k-means 2, 3 & 4 clusters	45

Table 12 : Comparison of A4 & A6 only single -Euclidean.....	46
--	----

FIGURES

Figure 1: A1 Scatterplot for 4 variables are correlated in Pearson's method	6
Figure 2: A1 Scatterplot for 5 variables with regression lines.....	6
Figure 3: A1 regression lines on Scatterplot.....	6
Figure 4: Standardized values & correlation matrix for the standardized values	7
Figure 5: standardized Scatterplots and Regression lines	7
Figure 6 : Euclidean distance using single linkage	9
Figure 7 : Manhattan distance using single linkage.....	10
Figure 8 : Pearson distance using single linkage.....	10
Figure 9 : Squared Euclidean distance using single linkage.....	11
Figure 10 : Squared Pearson distance using single linkage	11
Figure 11 : All distance dendrograms using single linkage	12
Figure 12 : Euclidean distance measure using single linkage for std values	13
Figure 13 : Manhattan distance measure using single linkage for std values	14
Figure 14 : Pearson distance measure using single linkage for std values	14
Figure 15 : Squared Euclidean distance measure using single linkage for std values	15
Figure 16 : Squared Pearson distance measure using single linkage for std values	15
Figure 17 : All distance measures using single linkage for std values	16
Figure 18 : Single linkage using Euclidean distance measure for Actual values.....	21
Figure 19 : Average linkage using Euclidean distance measure for Actual values.....	21
Figure 20 : Centroid linkage using Euclidean distance measure for Actual values	22
Figure 21 : Complete linkage using Euclidean distance measure for Actual values	22
Figure 22 : McQuitty linkage using Euclidean distance measure for Actual values	23
Figure 23 : Median linkage using Euclidean distance measure for Actual values	23
Figure 24 : Ward linkage using Euclidean distance measure for Actual values	24
Figure 25 : All linkages using Euclidean distance measure for Actual values.....	24
Figure 26 : Single linkage using Euclidean distance measure for Std values.....	27
Figure 27 : Average linkage using Euclidean distance measure for Std values	27
Figure 28 : Centroid linkage using Euclidean distance measure for Std values	28
Figure 29 : Complete linkage using Euclidean distance measure for Std values	28
Figure 30 : McQuitty linkage using Euclidean distance measure for Std values	29
Figure 31 : Median linkage using Euclidean distance measure for Std values	29
Figure 32 : Ward linkage using Euclidean distance measure for Std values	30
Figure 33 : All linkage comparison using Euclidean distance measure for Std values	30
Figure 34 : Cluster Variables using Actual Values.....	34
Figure 35 : Cluster of Variables using standardized values	34
Figure 36 : Single linkage using Euclidean distance measure for xc1.....	35
Figure 37 : Cluster Analysis of observations for std. Xc1 values	37
Figure 38 : Cluster Analysis of observations for act. Xc2	39
Figure 39 : Cluster Analysis of observations for std.XC2	41

Team Effort:

As the team comprises of two members, the effort has been equally divided between the members, each contributing to this part of project (Part A) equally at 50%.

Data:

A dataset of 7 observations and 5 variables were given to study the correlation of observations and correlation of variables applying the Cluster analysis methods for observations and variables.

Original data					
Case	X1	X2	X3	X4	X5
1	7	10	9	7	10
2	9	9	8	9	9
3	5	5	6	7	7
4	6	6	3	3	4
5	1	2	2	1	2
6	4	3	2	3	3
7	2	4	5	2	5

Similarity Measure : Correlation of Observations							
Case	1	2	3	4	5	6	7
1	1						
2	-0.147	1					
3	0	0	1				
4	0.087	0.516	-0.824	1			
5	0.963	-0.408	0	-0.06	1		
6	-0.466	0.791	-0.354	0.699	-0.645	1	
7	0.891	-0.516	0.165	-0.239	0.963	-0.699	1

It could be noted that there are two groups 1, 5 & 7 and 2, 4 & 6 from the 'blue' boxes with positive values

Similarity Measure : Euclidean Distance							
Case	1	2	3	4	5	6	7
1	nc						
2	3.32	nc					
3	6.86	6.63	nc				
4	10.25	10.2	6	nc			
5	15.78	16.19	10.1	7.07	nc		
6	13.11	13	7.28	3.87	3.87	nc	
7	11.27	12.16	6.32	5.1	4.9	4.36	nc

The observations that are in closest proximity are 1 & 2, 4 & 6, 5 & 6, 6 & 7.

A1: Study of Variables using Pearson Correlation on Actual Values

Pearson Correlation Method: Wikipedia: Pearson's correlation coefficient is the covariance of the two variables divided by the product of their standard deviations. The form of the definition involves a "product moment", that is, the mean (the first moment about the origin) of the product of the mean-adjusted random variables; hence the modifier product-moment in the name.

Pearson's correlation coefficient when applied to a population is commonly represented by the Greek letter ρ (rho) and may be referred to as the population correlation coefficient or the population Pearson correlation coefficient. The formula for ρ is:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

where:

- cov is the covariance
- σ_X is the standard deviation of X

Pearson Correlation is a common measure of similarity. When applied to a sample, given two objects y_r and y_s in a p-dimensional space, a similarity measure, S_{rs} , satisfies the following conditions

- $0 \leq S_{rs} \leq 1$ for all objects y_r and y_s
- $S_{rs} = 1$ if and only if $y_r = y_s$
- $S_{rs} = S_{sr}$

It is a product moment correlation between object y_r and y_s , $r, s = 1, 2, 3, \dots, n$ which is defined as

$$P_{rs} = \frac{\sum_{j=1}^p (y_{rj} - \bar{y}_{r.})(y_{sj} - \bar{y}_{s.})}{\sqrt{[\sum_{j=1}^p (y_{rj} - \bar{y}_{r.})^2 \sum_{j=1}^p (y_{sj} - \bar{y}_{s.})^2]}}$$

where $\bar{y}_{r.} = \sum_j y_{rj} / p$ and $\bar{y}_{s.} = \sum_j y_{sj} / p$

The above details were taken from IMSE 514 notes.

Advantages	Disadvantages
It is easy to interpret	The correlation coefficient only measures linear relationship between two variables, however it does not imply causality. If the relationship is not linear, then the result is inaccurate.

	X1	X2	X3	X4
X2	0.888			
X3	0.696	0.886		
X4	0.872	0.817	0.851	
X5	0.779	0.919	0.989	0.902
	0.039	0.003	0.000	0.005

Cell Contents: Pearson correlation
P-Value

Observation: The variables X2, X3 & X4 are in high correlation with X5. And in turn X2, X3 & X4 are in reasonable correlation with each other. (Values close to 0.8) Scatterplots for these variables that are found correlated as in the figure. Figure 2 shows that X1 & X5 are not correlated.

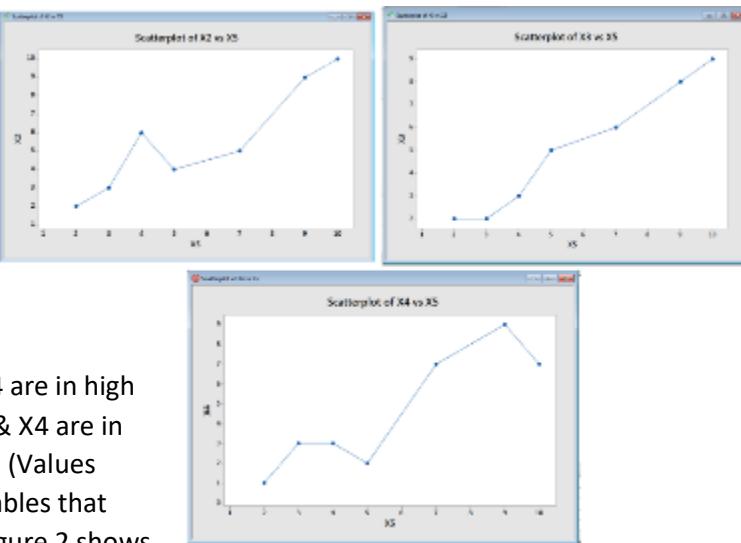


FIGURE 1: A1 SCATTERPLOT FOR 4 VARIABLES ARE CORRELATED IN PEARSON'S METHOD

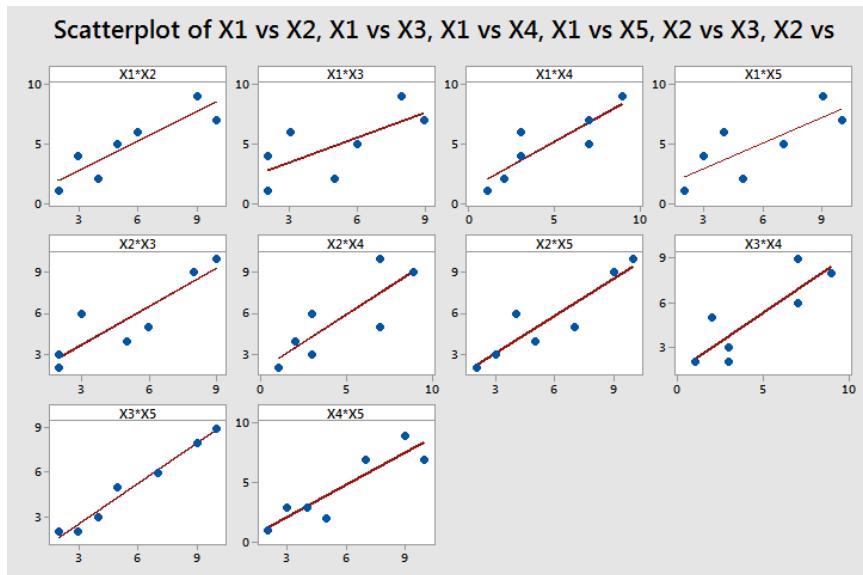


FIGURE 2: A1 SCATTERPLOT FOR 5 VARIABLES WITH REGRESSION LINES

Findings: The correlation value for the variables X3 & X5 is 0.989, which is maximum. This could be noted from the way the scatterplot has come out. When a regression is drawn in between, the residuals are very low to the regression, with 3 data points falling on the regression line. The data points for X2, X4 are little away from their regression lines as in the figure below.

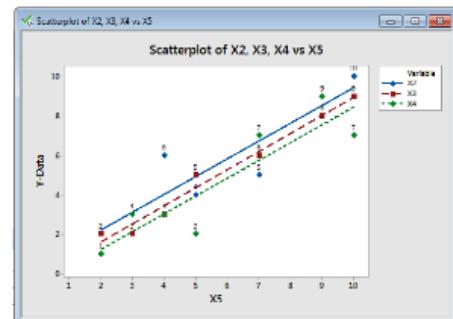


FIGURE 3: A1 regression lines on Scatterplot

A2: Study of Variables using Pearson Correlation on Standardized Values

The actual data was standardized using “Subtract Mean and divide by Standard deviation” method. The standardized values are as follows

C6	C7	C8	C9	C10
A1	A2	A3	A4	A5
0.76680	1.48011	1.41421	0.79697	1.41004
1.48248	1.14589	1.06066	1.45330	1.08103
0.05112	-0.19098	0.35355	0.79697	0.42301
0.40896	0.14324	-0.70711	-0.51569	-0.56402
-1.38024	-1.19364	-1.06066	-1.17202	-1.22204
-0.30672	-0.85942	-1.06066	-0.51569	-0.89303
-1.02240	-0.52520	0.00000	-0.84385	-0.23501

Correlation: A1, A2, A3, A4, A5				
	A1	A2	A3	A4
A2	0.888 0.008			
A3	0.696 0.083	0.886 0.008		
A4	0.872 0.010	0.817 0.025	0.851 0.015	
A5	0.779 0.039	0.919 0.003	0.989 0.000	0.902 0.005

Cell Contents: Pearson correlation
P-Value

FIGURE 4: STANDARDIZED VALUES & CORRELATION MATRIX FOR THE STANDARDIZED VALUES

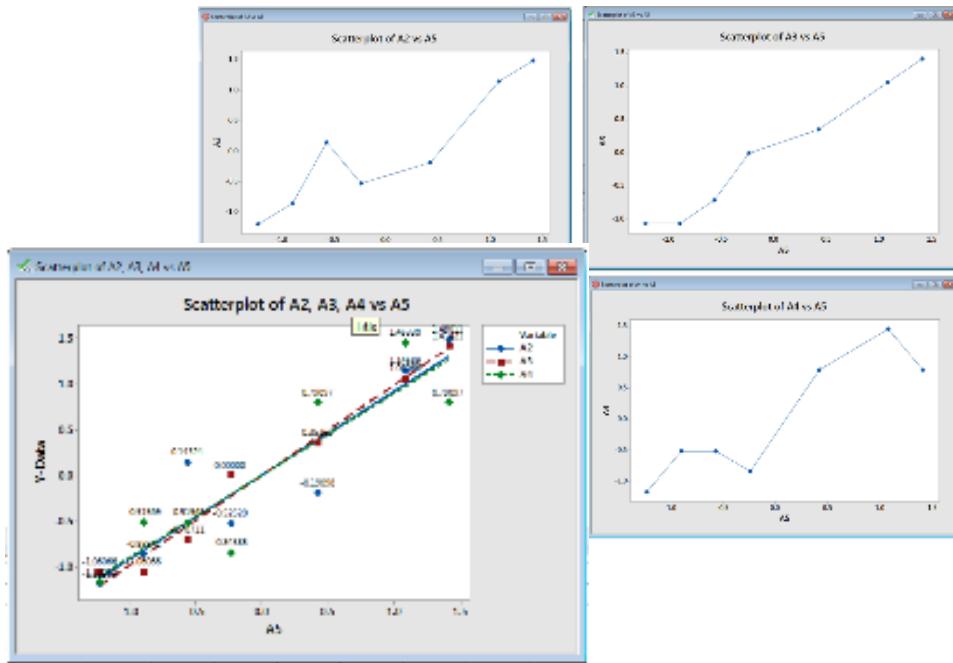


FIGURE 5: STANDARDIZED SCATTERPLOTS AND REGRESSION LINES

Observation: Study of Correlation of variables using Pearson’s method, does not deviate at all from the correlations observed using the actual values. This is because, all of the variables have been standardized with similar transformation method. If one of the variables was left out for standardization step, a different correlation matrix could have been observed. As there was no difference observed in the correlation matrix, a similar set of scatter plots is expected for the standardized values. Except for the scale of the axes, there are no differences identified. However, the regression lines of each correlation has become closure to each other and the data points are close to the line, again caused by the change in scale.

A3: Cluster Analysis on Observations using “Single Linkage” and five distance measures (Actual Values)

Using the actual values of the observations, the single linkage method was used with the following distance measures to identify the similarity between observations to form the cluster of observations.

Euclidean Distance: (from book “Data Mining Concepts & Techniques”) Euclidean distance is one of such dissimilarity measure used to find dissimilarity of objects defined by numeric attributes. It is also called as “Straight line or as the crow flies”. Let $i = (x_{i1}, x_{i2}, \dots, x_{ip})$ and $j = (x_{j1}, x_{j2}, \dots, x_{jp})$ be two objects described in p numeric attributes. The Euclidean distance $d(i, j)$ between objects i and j is defined as

$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2}$$

Manhattan Distance: (from book “Data Mining Concepts & Techniques”) Manhattan or City block distance, named so because it is the distance in blocks between any two points in a city (such as 2 blocks down and 3 blocks over for a total of 5 blocks). It is defined as

$$d(i, j) = |(x_{i1} - x_{j1})| + |(x_{i2} - x_{j2})| + \dots + |(x_{ip} - x_{jp})|$$

Pearson Distance: (See in A1 [response](#))

But put in different words, (from <http://www.improvedoutcomes.com/>), Pearson Correlation measures the similarity in shape between two profiles. The formula for the Pearson Correlation distance is:

$$d = 1 - r, \text{ where } r = Z(x) \cdot Z(y)/n$$

is the dot product of the z-scores of the vectors x and y . The z-score of x is constructed by subtracting from x its mean and dividing by its standard deviation.

Squared Euclidean Distance:

Wikipedia: The standard Euclidean distance can be squared in order to place progressively greater weight on objects that are farther apart. In this case, the equation becomes

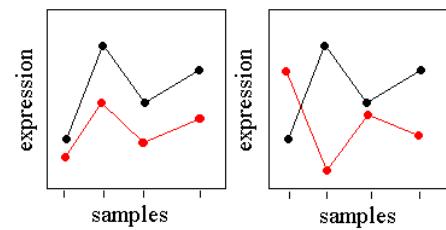
$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_i - q_i)^2 + \dots + (p_n - q_n)^2}.$$

Squared Euclidean Distance is not a metric as it does not satisfy the triangle inequality, however, it is frequently used in optimization problems in which distances only have to be compared. It is also referred to as “quadrandance” within the field of rational trigonometry.

Squared Pearson Distance:

(From <http://www.improvedoutcomes.com/>) The Pearson Squared distance measures the similarity in shape between two profiles, but can also capture inverse relationships. For example, consider the following gene profiles:

In the figure on the left, the black profile and the red profile have almost perfect Pearson correlation despite the differences in basal expression level and scale. These genes would cluster together with either Pearson Correlation or Pearson Squared distance. In the figure on the right, the black and red profiles are almost perfectly *anti-correlated*. These genes would be placed in remote clusters using Pearson Correlation, but would be put in the same cluster using Pearson Squared. The formula for the Pearson Squared distance is



$$d = 1 - 2r, \text{ where } r \text{ is the Pearson correlation defined above.}$$

Caveats: While most combinations of clustering algorithm and distance metrics provide meaningful results, there are a few combinations that are difficult to interpret. In particular, combining [K-Means clustering](#) with the Pearson Squared distance metric can lead to non-intuitive [centroid plots](#) since the centroid represents the mean of the cluster and Pearson Squared can group anti-correlated objects. In these cases, visually drilling into clusters to see the individual members through the use of Cluster Plots produce better results. Alternatively, the results of the clustering run can be visualized using the [Matrix Tree Plot](#).

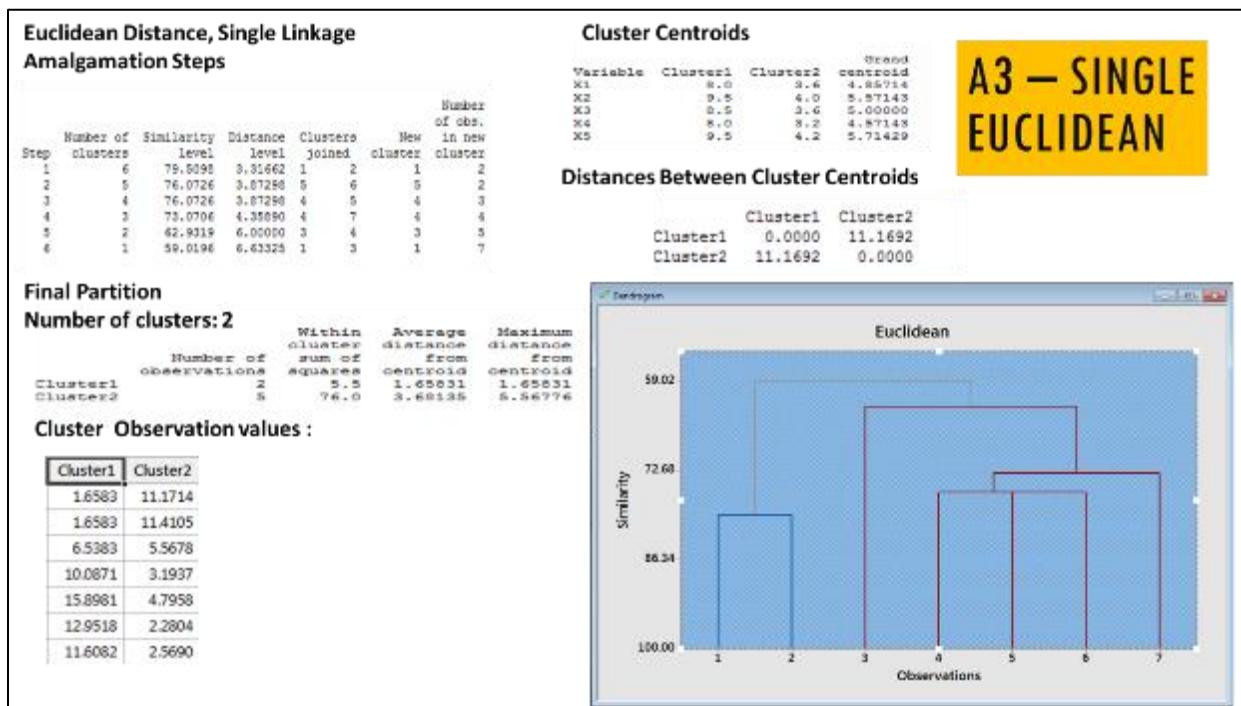


FIGURE 6 : EUCLIDEAN DISTANCE USING SINGLE LINKAGE

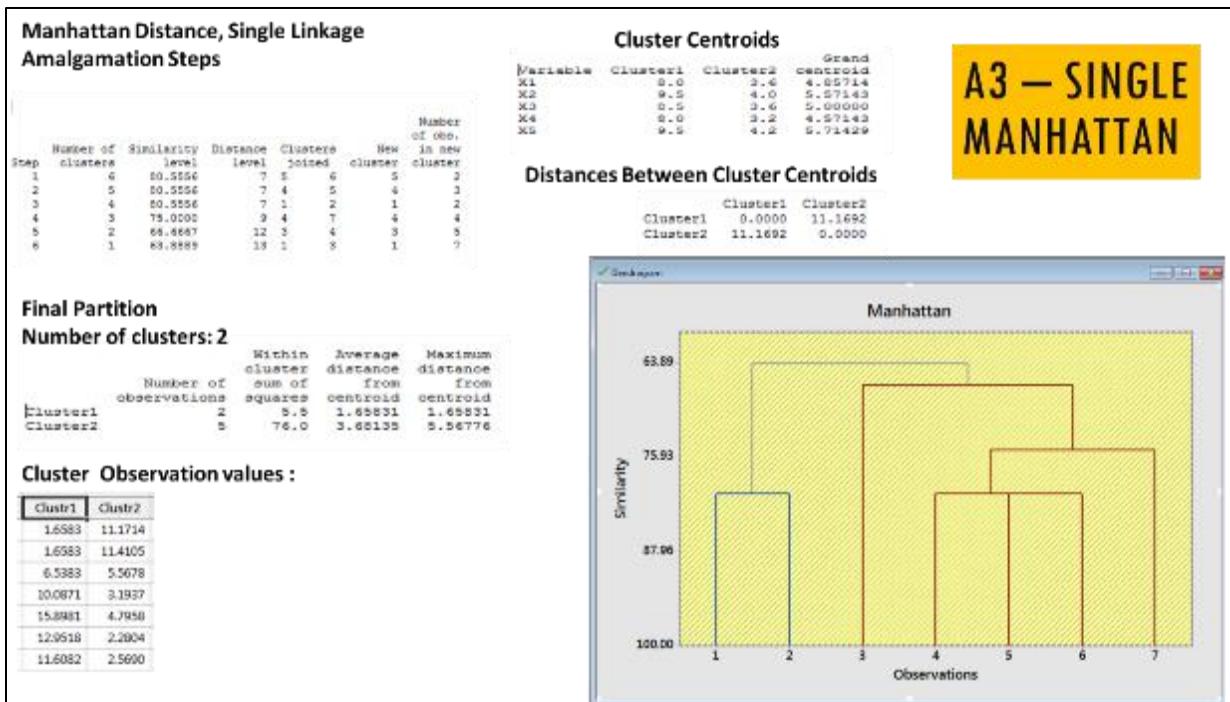


FIGURE 7 : MANHATTAN DISTANCE USING SINGLE LINKAGE

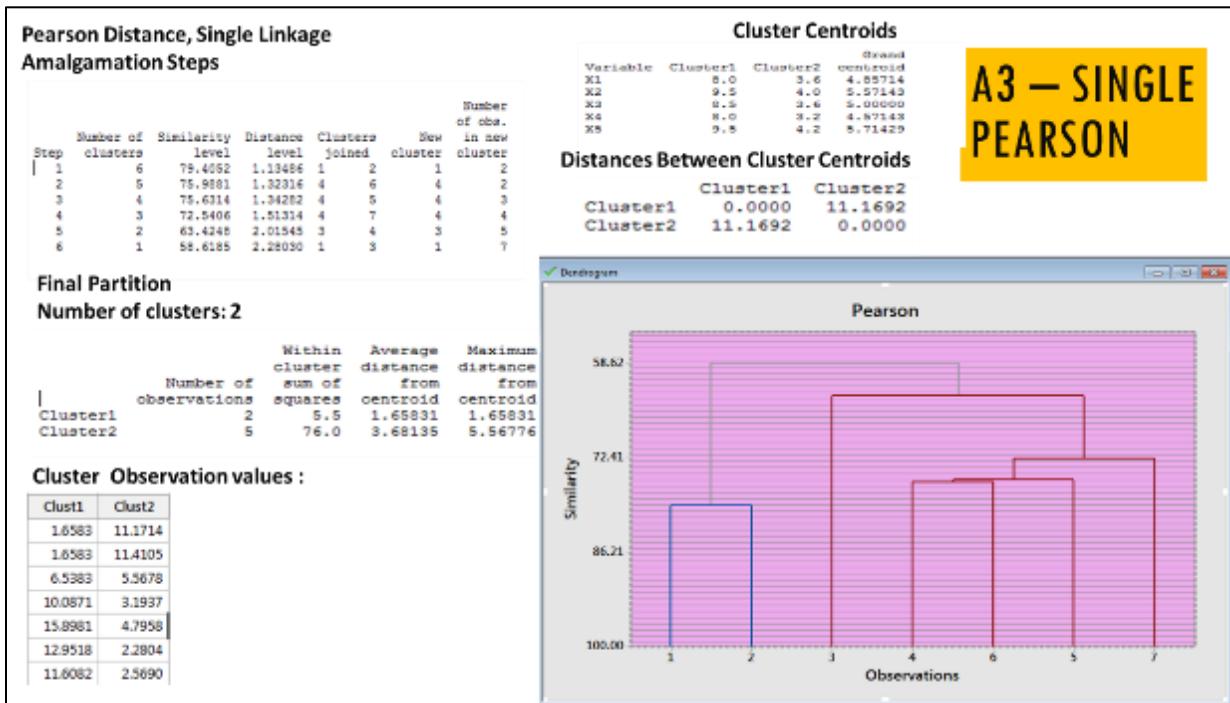


FIGURE 8 : PEARSON DISTANCE USING SINGLE LINKAGE

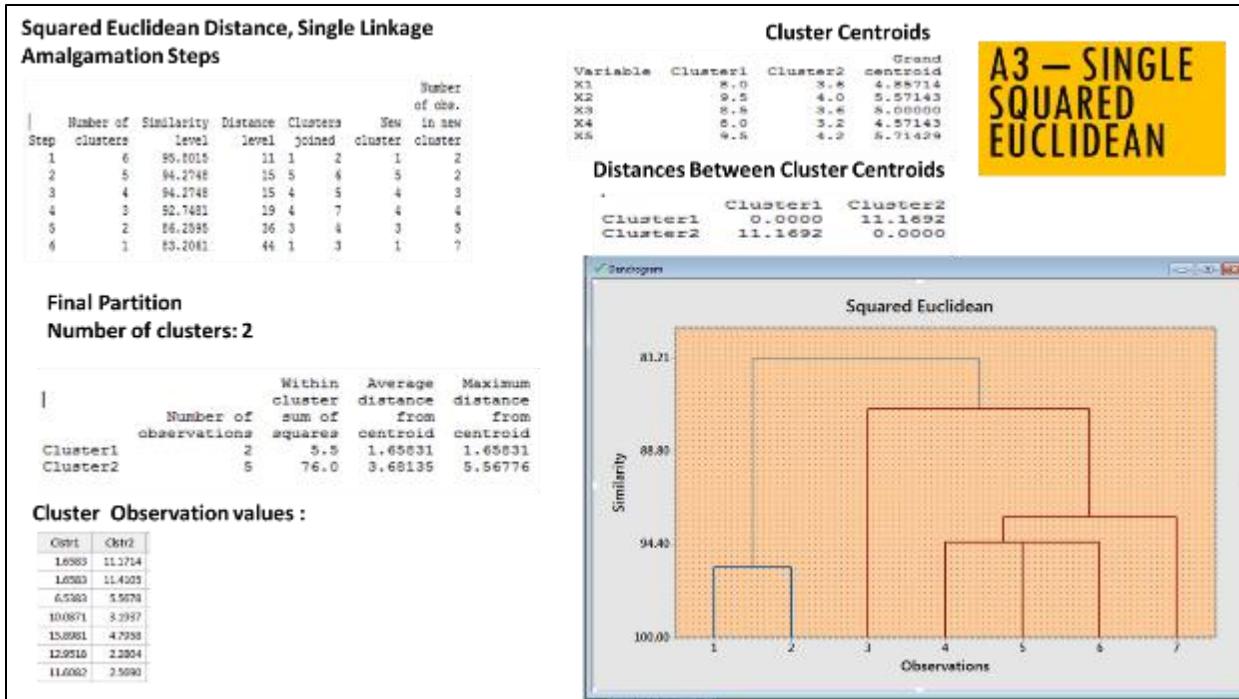


FIGURE 9 : SQUARED EUCLIDEAN DISTANCE USING SINGLE LINKAGE

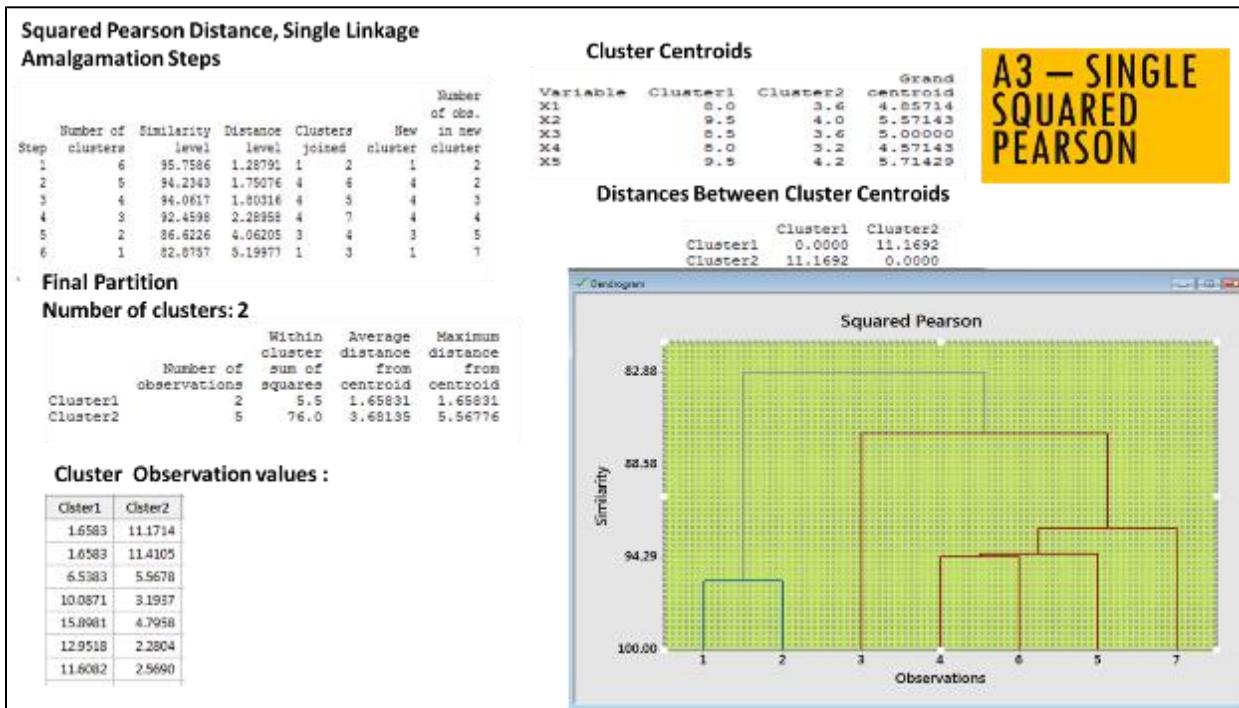


FIGURE 10 : SQUARED PEARSON DISTANCE USING SINGLE LINKAGE

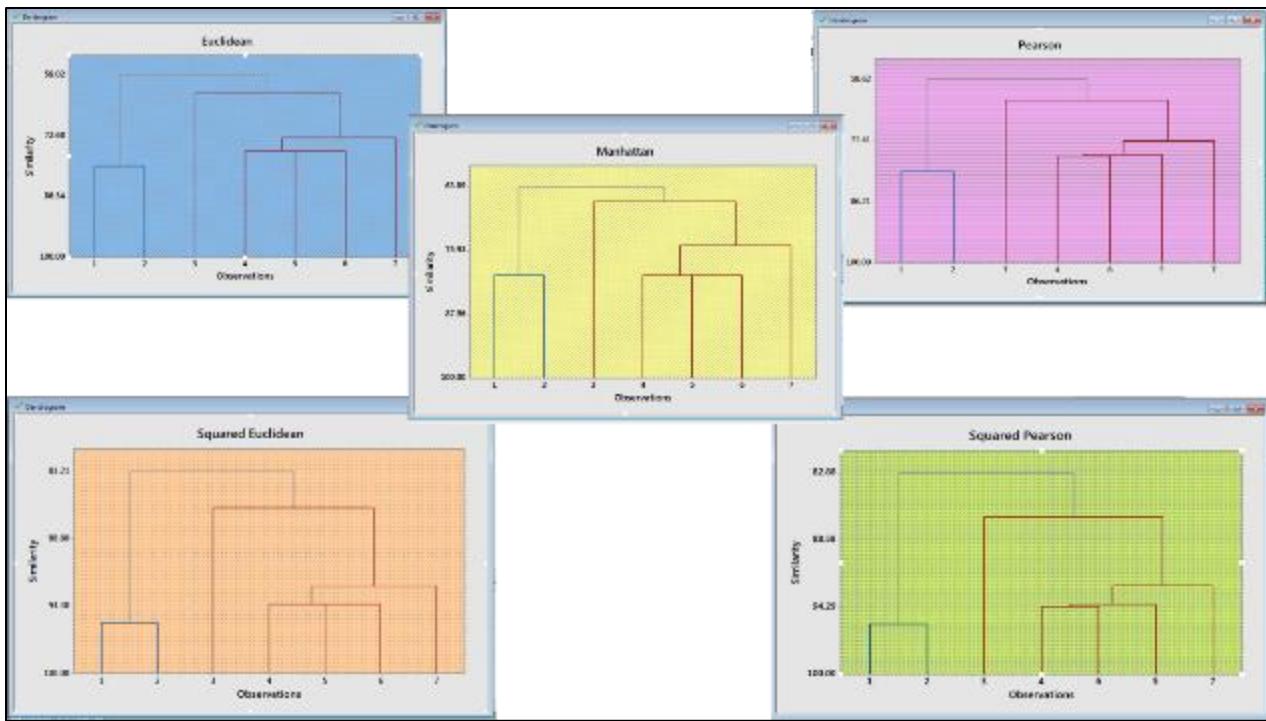


FIGURE 11 : ALL DISTANCE DENDROGRAMS USING SINGLE LINKAGE

Findings:

- All the methods produced 2 clusters, first cluster with 2 observations (1,2) and second cluster with 5 observations(3, 4,5,6,7)
- Distance between cluster centroids is 11.1692
- Within Sum-of-Squares (WSS) for Cluster 1 is 5.5 and Cluster 2 is 76.0, which is almost similar in all methods
- Maximum distance from centroid for cluster 1 is 1.6 and cluster 2 is 5.6 (for all methods)
- Pearson & Squared Pearson arranged the cluster 2 with observations 4 & 6 as the first similar set. Whereas, rest of the methods found 5 & 6 in the bottom level (similar ones).
- Each of the methods has different highest similarity levels, especially the squared methods have “highest similarity level” at step 1 above 90.

A4: Cluster Analysis on Observations using “Single Linkage” and five distance measures (Standardized Values)

Findings: When the A3 steps were repeated for the standardized values, it could be noted that there is significant reduction in “within cluster sum of squares”, “Average distance from centroid” and “Maximum distance from centroid” for all the five distance measures. This could be envisioned in a way that the clusters have been brought down in scale and closure to each other, a shrinking impact caused by the standardization. The similarity levels did not change, which makes sense that we have standardized the values, which made them to be nearer than before, but did not impact their correlation. No change in similarity levels caused the dendrograms to be very similar to that of the actual values in A3. Overall the clusters have shrunk in their sizes and have come closure, due to standardization.

In other words, if only few of the variables were standardized, it might have led to a change in similarity level due to the difference in scale among the variables.

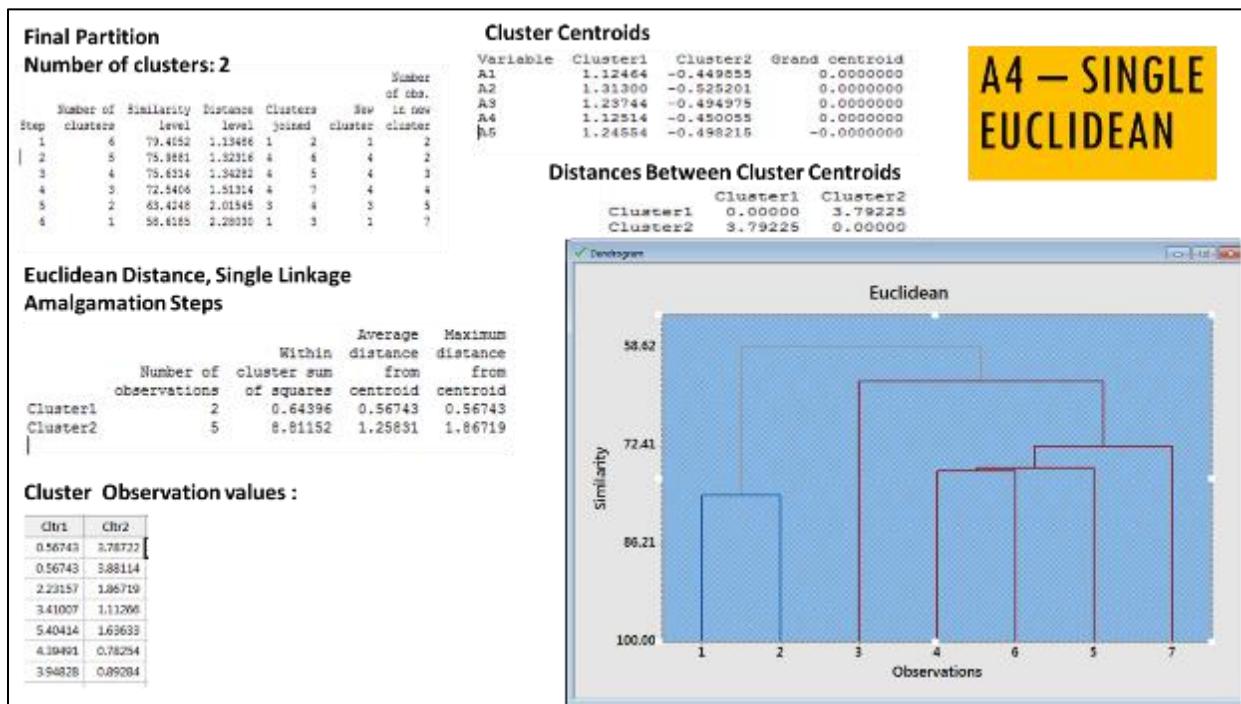


FIGURE 12 : EUCLIDEAN DISTANCE MEASURE USING SINGLE LINKAGE FOR STD VALUES

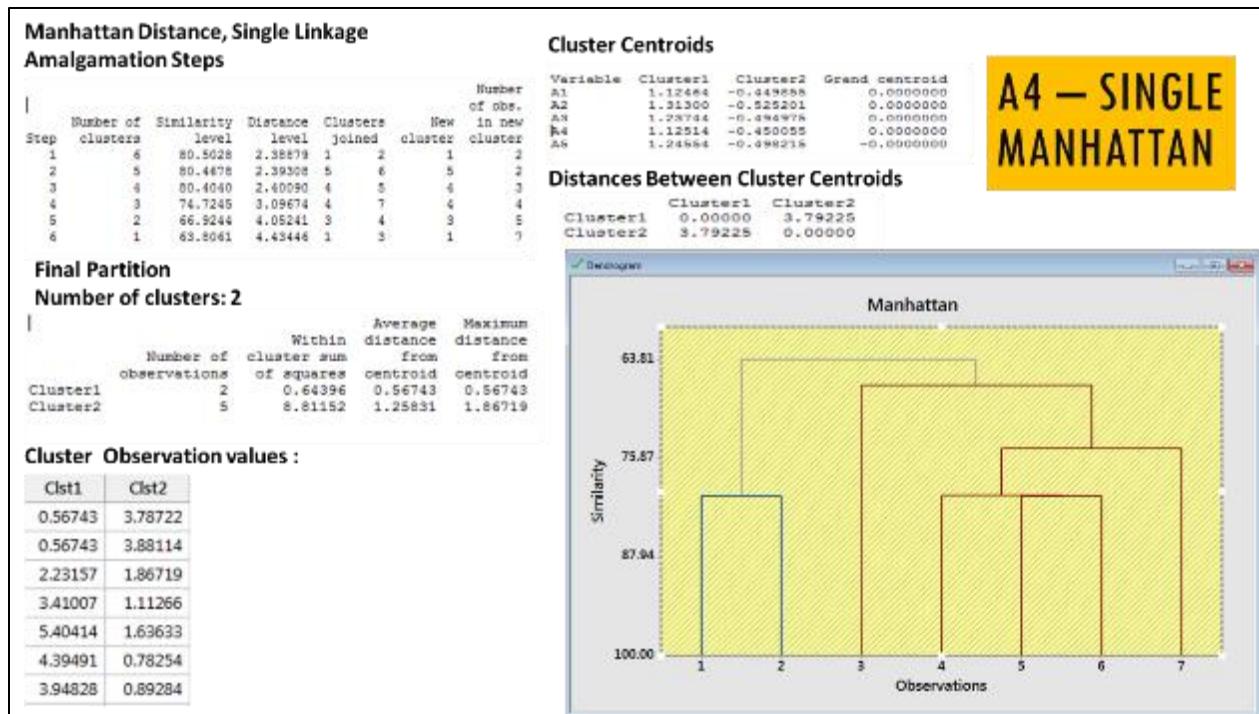


FIGURE 13 : MANHATTAN DISTANCE MEASURE USING SINGLE LINKAGE FOR STD VALUES

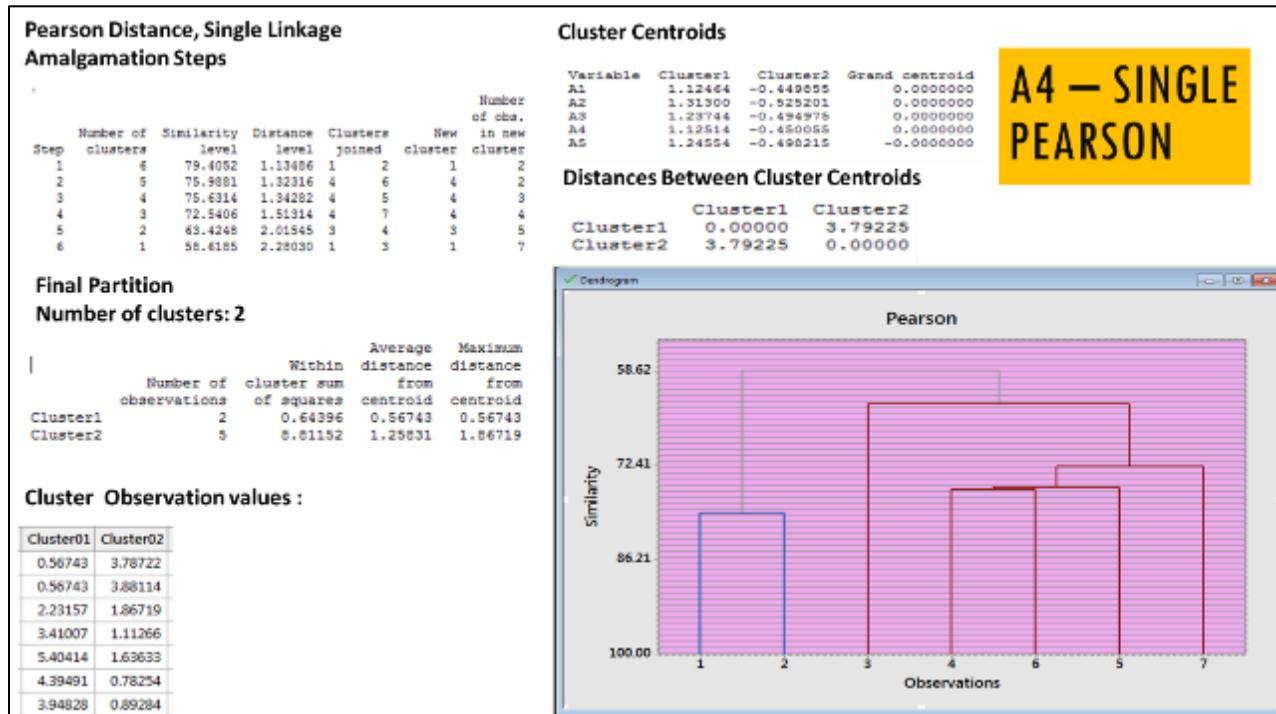


FIGURE 14 : PEARSON DISTANCE MEASURE USING SINGLE LINKAGE FOR STD VALUES

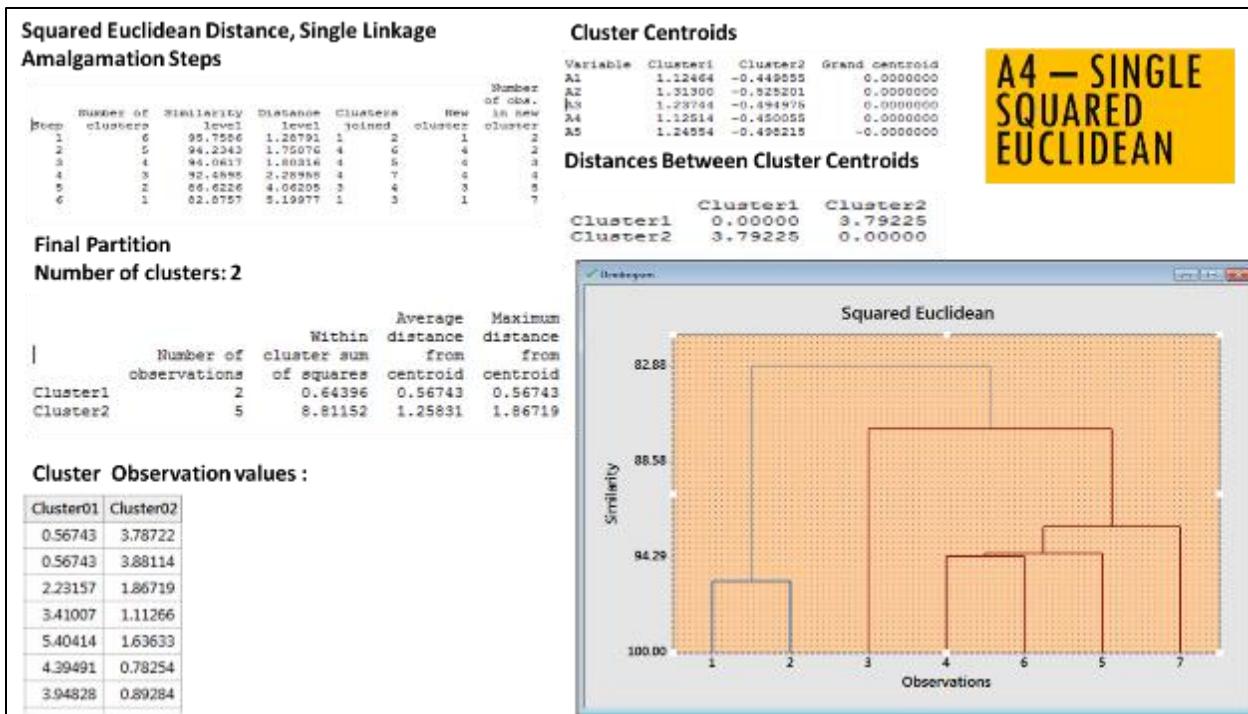


FIGURE 15 : SQUARED EUCLIDEAN DISTANCE MEASURE USING SINGLE LINKAGE FOR STD VALUES

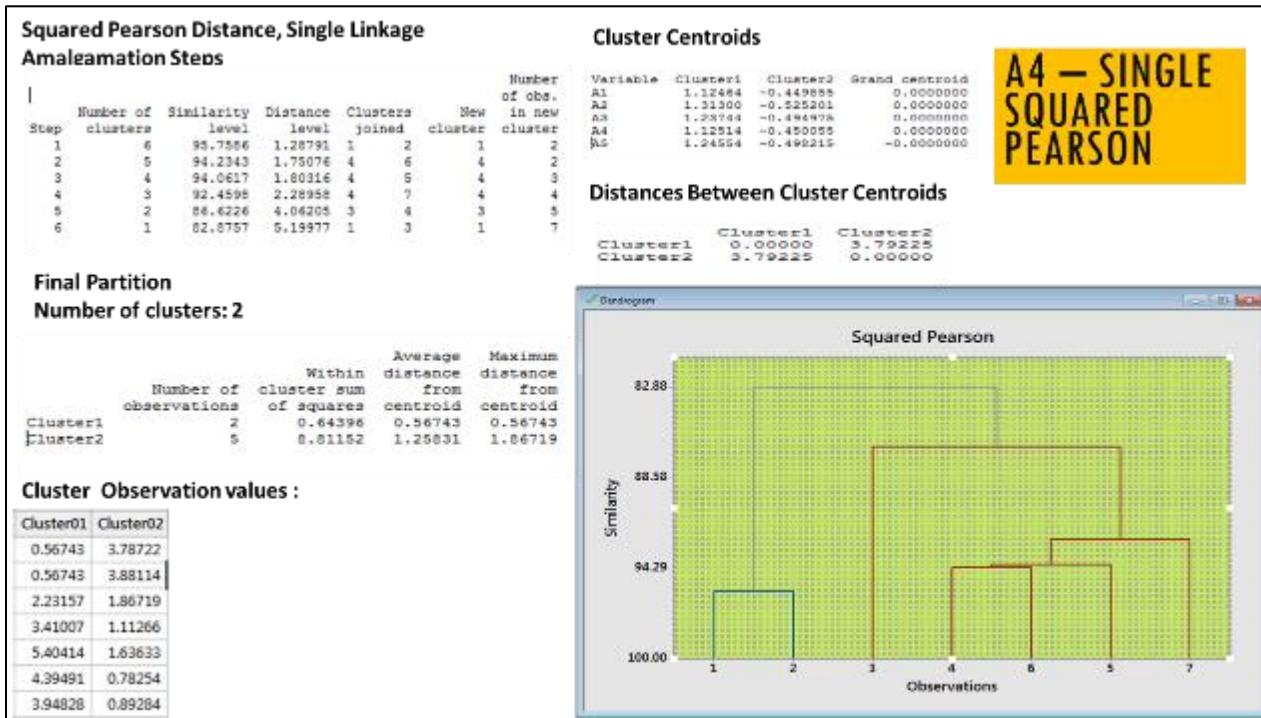


FIGURE 16 : SQUARED PEARSON DISTANCE MEASURE USING SINGLE LINKAGE FOR STD VALUES

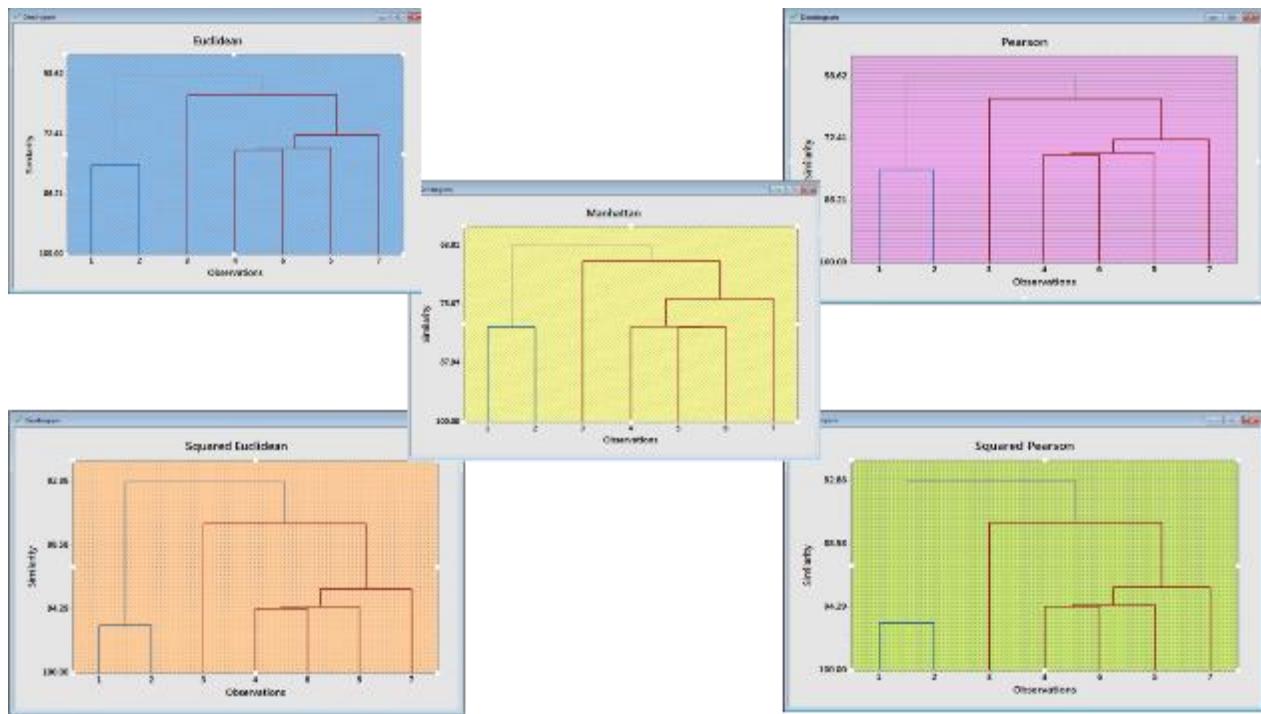


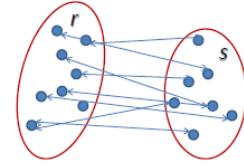
FIGURE 17 : ALL DISTANCE MEASURES USING SINGLE LINKAGE FOR STD VALUES

TABLE 1 : COMPARISON OF A3 & A4

Applying Single Linkage with	Using Actual Values (A3)	Using Standardized Values (A4)
Euclidean Distance Measure	Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[4,c[5,6]]]] Highest similarity level: 79.5 Distance btwn. Cluster Centroids: 11.1692 WSS1: 5.5 WSS2: 76.0	Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[5,c[4,6]]]] Highest similarity level: 79.4 Distance btwn. Cluster Centroids: 3.792 WSS1: 0.643 WSS2: 8.811
Manhattan Distance Measure	Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[4,c[5,6]]]] Highest similarity level: 80.55 Distance btwn. Cluster Centroids: 11.1692 WSS1: 5.5 WSS2: 76.0	Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[4,c[5,6]]]] Highest similarity level: 80.5 Distance btwn. Cluster Centroids: 3.792 WSS1: 0.643 WSS2: 8.811
Pearson Distance Measure	Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[5,c[4,6]]]] Highest similarity level: 79.4 Distance btwn. Cluster Centroids: 11.1692 WSS1: 5.5 WSS2: 76.0	Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[5,c[4,6]]]] Highest similarity level: 79.4 Distance btwn. Cluster Centroids: 3.792 WSS1: 0.643 WSS2: 8.811
Squared Euclidean	Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[4,c[5,6]]]] Highest similarity level: 95.0 Distance btwn. Cluster Centroids: 11.1692 WSS1: 5.5 WSS2: 76.0	Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[5,c[4,6]]]] Highest similarity level: 95.755 Distance btwn. Cluster Centroids: 3.792 WSS1: 0.643 WSS2: 8.811
Squared Pearson	Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[5,c[4,6]]]] Highest similarity level: 95.75 Distance btwn. Cluster Centroids: 11.1692 WSS1: 5.5 WSS2: 76.0	Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[5,c[4,6]]]] Highest similarity level: 95,7556 Distance btwn. Cluster Centroids: 3.792 WSS1: 0.643 WSS2: 8.811
Findings	<ul style="list-style-type: none"> • Pearson & Squared Pearson started cluster 2 with 4 & 6. (rest started with 5 & 6) • Squared (both) methods have high – highest similarity levels at ~95 	<ul style="list-style-type: none"> • Only Manhattan started cluster 2 with 5 & 6. (Rest started with 4 & 6) → Standardization of values caused this change. • Squared(both) methods have high – highest similarity levels at ~95 • WSS values reduced drastically (means the points inside cluster have come nearer) • Distance between cluster centroids reduced (means the clusters have come nearer)

A5: Cluster Analysis on Observations using “Euclidean Distance” measure and all linkage methods (Actual Values)

Average Linkage: (From www.saedsayad.com) In average linkage hierarchical clustering, the distance between two clusters is defined as the average distance between each point in one cluster to every point in the other cluster. For example, the distance between clusters “r” and “s” to the left is equal to the average length each arrow between connecting the points of one cluster to the other.

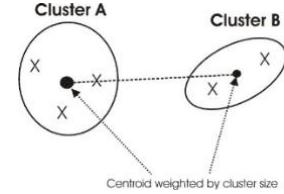


$$L(r, s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} D(x_{ri}, x_{sj})$$

Reasons to use: Whereas the single or complete linkage methods group clusters based upon single pair distances, average linkage uses a more central measure of location

Centroid Linkage:

(From <http://www.mitodb.com/>) Given two clusters of diseases $G = \{X_1, X_2, \dots, X_K\}$ and $H = \{Y_1, Y_2, \dots\}$, where each element is a disease with associated symptom vector ($X_j = (x_{j1}, x_{j2}, \dots, x_{jn})$) etc. Each cluster could potentially consist of only one disease, in the case of a disease that has not (yet) been assigned a cluster. The centroid-vector of each cluster is then defined as (for



$$\text{Centroid}_G = \left(\frac{\sum_{i=1, n} \sum_{j=1, K} x_{ji}}{K}, \frac{\sum_{i=1, n} \sum_{j=1, K} x_{j2}}{K}, \dots, \frac{\sum_{i=1, n} \sum_{j=1, K} x_{jn}}{K} \right)$$

the cluster G):

where N is the total number of disease pairs, i.e. Centroid vector consists of the average of the symptom frequencies for the constituent diseases. The distance between each cluster is then:

$$D(G, H) = D(\text{Centroid}_G, \text{Centroid}_H)$$

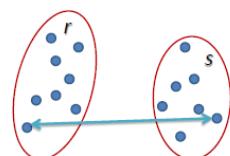
i.e. the distance between the centroid vectors of the two clusters

Advantages: Easy to implement and commonly used in the field of Biology.

Reasons to use: Like average linkage, this method is another averaging technique

Complete Linkage:

(From www.saedsayad.com) In complete linkage hierarchical clustering, the distance between two clusters is defined as the longest distance between two points in each cluster. For example, the distance between clusters “r” and “s” to the left is equal to the length of the arrow between their two furthest points.



$$L(r, s) = \max(D(x_{ri}, x_{sj}))$$

Reasons to use: Ensures that all items in a cluster are within a maximum distance and tends to produce clusters with similar diameters. The results can be sensitive to outliers

McQuitty Linkage:

The average of the distances of the soon to be joined clusters to that other cluster. For example, if clusters 1 and 3 are to be joined into a new cluster, say 1^* , then the distance from 1^* to cluster 4 is the average of the distances from 1 to 4 and 3 to 4. Also called "weighted average linkage."

Reasons to use: Here, distance depends on a combination of clusters rather than individual items in the clusters. Similar to average linkage, but the size of the clusters are assumed equal, so the pairwise distances are weighted accordingly.

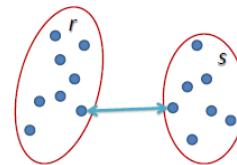
Median Linkage:

Median or **equilibrious centroid** method (WPGMC) is the modified previous. Proximity between two clusters is the proximity between their geometric centroids (squared Euclidean distance between those); while the centroids are defined so that the sub clusters of which each of these two clusters were merged recently have equalized influence on its centroid – even if the sub clusters differed in the number of objects

Reasons to use: Similar to average or centroid method, though it reduces the effect of outliers.

Single Linkage:

(From www.saedsayad.com) In single linkage hierarchical clustering, the distance between two clusters is defined as the shortest distance between two points in each cluster. For example, the distance between clusters "r" and "s" to the left is equal to the length of the arrow between their two closest points.



$$L(r,s) = \min(D(x_{ri}, x_{sj}))$$

Reasons to use: Best suited for observations or variables that are clearly separated. When they lie close together, single linkage tends to identify long chain-like clusters that can have a relatively large distance separating items at either end of the chain

Ward Linkage:

Ward's method, or minimal increase of sum-of-squares (MISSQ), sometimes incorrectly called "minimum variance" method. Proximity between two clusters is the magnitude by which the summed square in their joint cluster will be greater than the combined summed square in these two clusters: $SS_{12} - (SS_1 + SS_2)$

(Between two singleton objects this quantity = squared Euclidean distance / 2).

Reasons to use: Tends to produce clusters with similar numbers of items, but it is sensitive to outliers

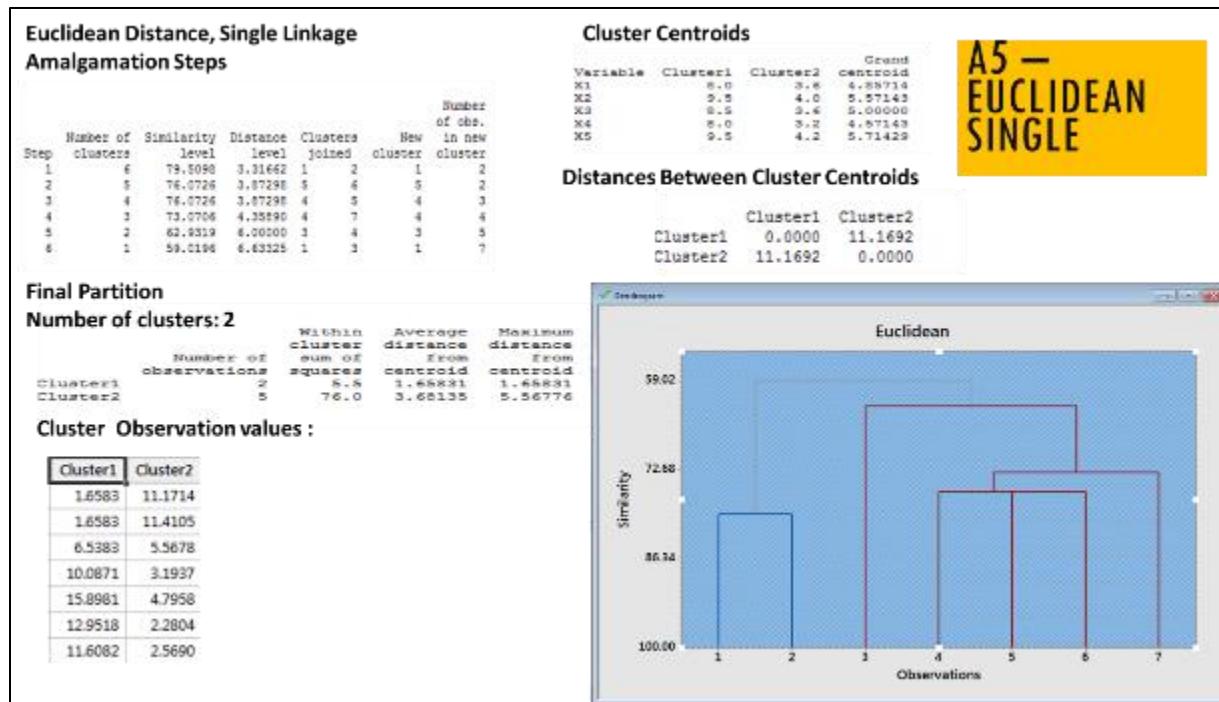


FIGURE 18 : SINGLE LINKAGE USING EUCLIDEAN DISTANCE MEASURE FOR ACTUAL VALUES

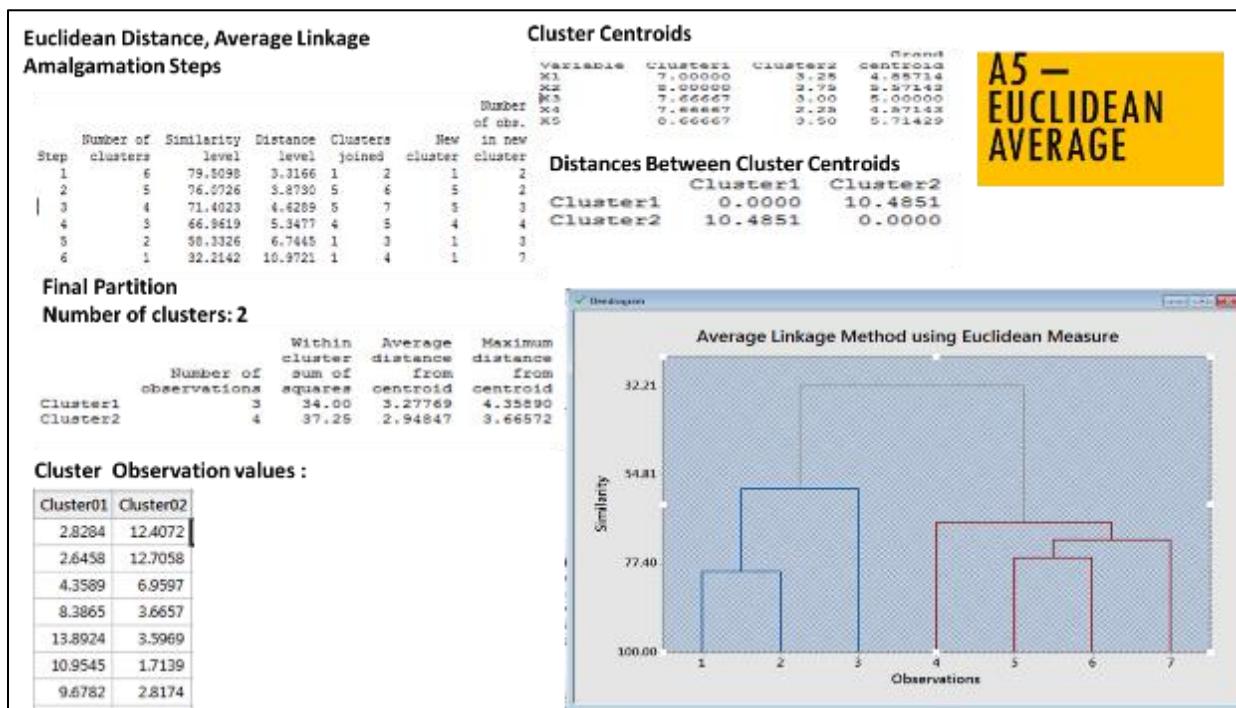


FIGURE 19 : AVERAGE LINKAGE USING EUCLIDEAN DISTANCE MEASURE FOR ACTUAL VALUES

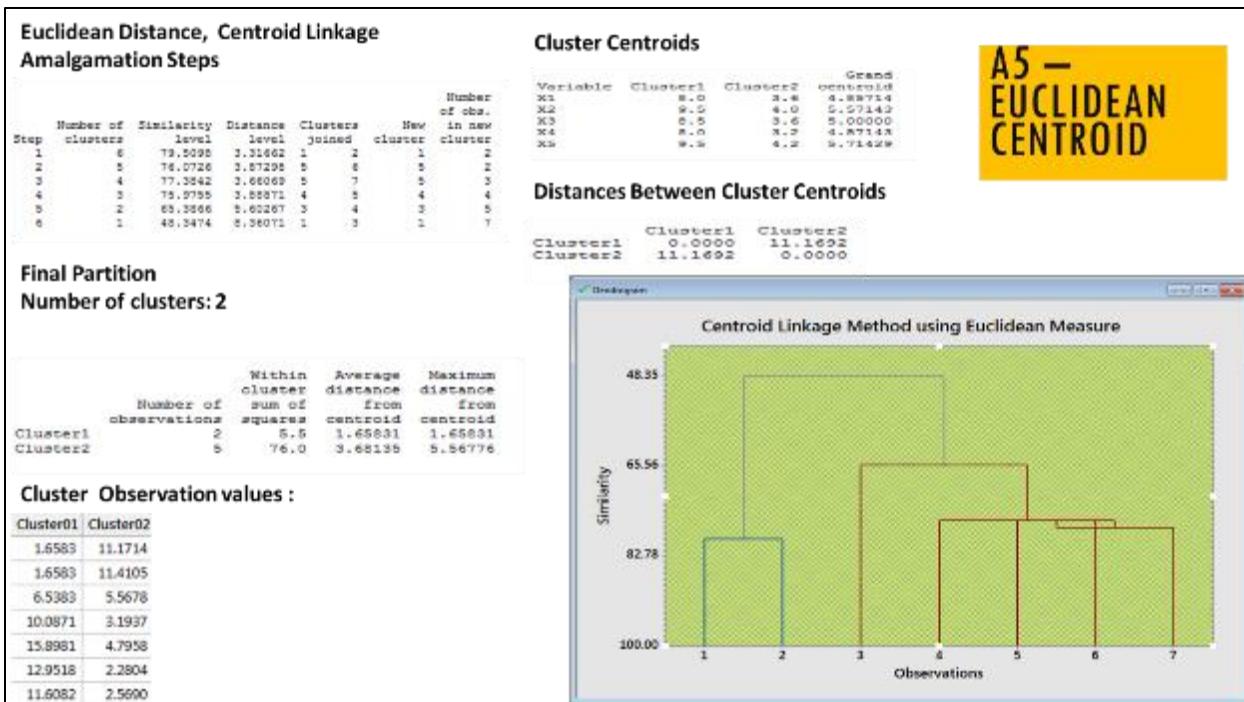


FIGURE 20 : CENTROID LINKAGE USING EUCLIDEAN DISTANCE MEASURE FOR ACTUAL VALUES

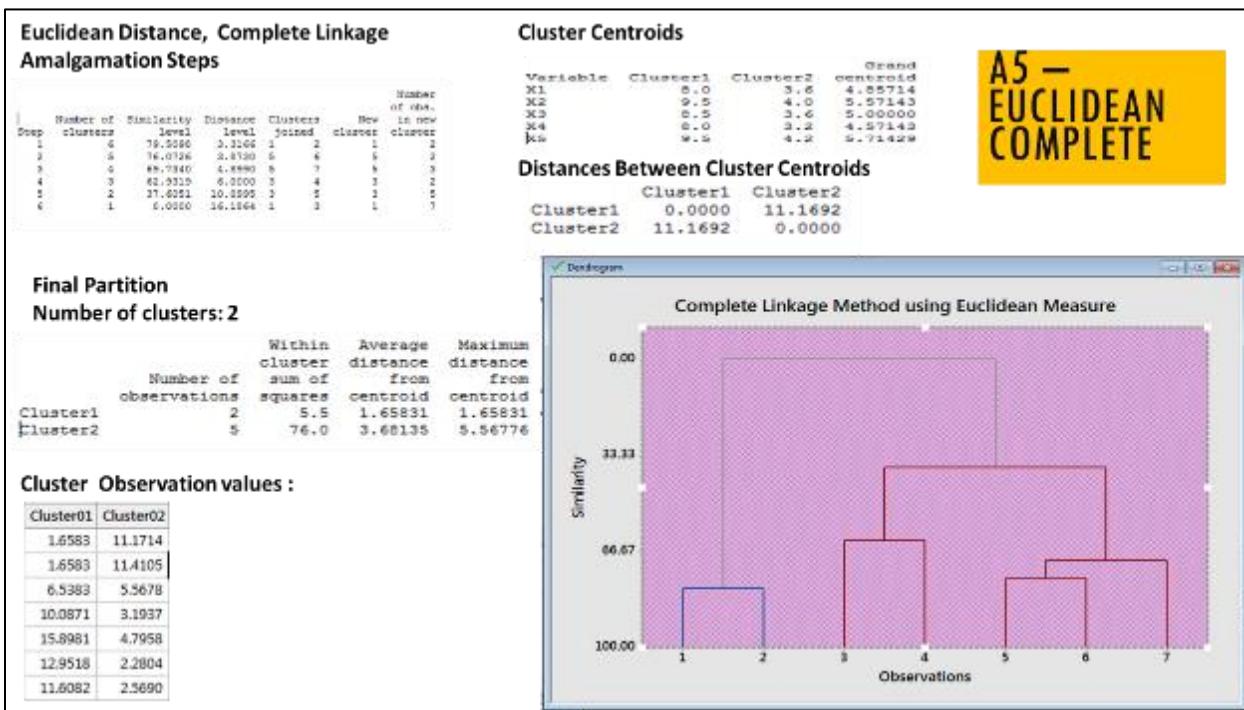


FIGURE 21 : COMPLETE LINKAGE USING EUCLIDEAN DISTANCE MEASURE FOR ACTUAL VALUES

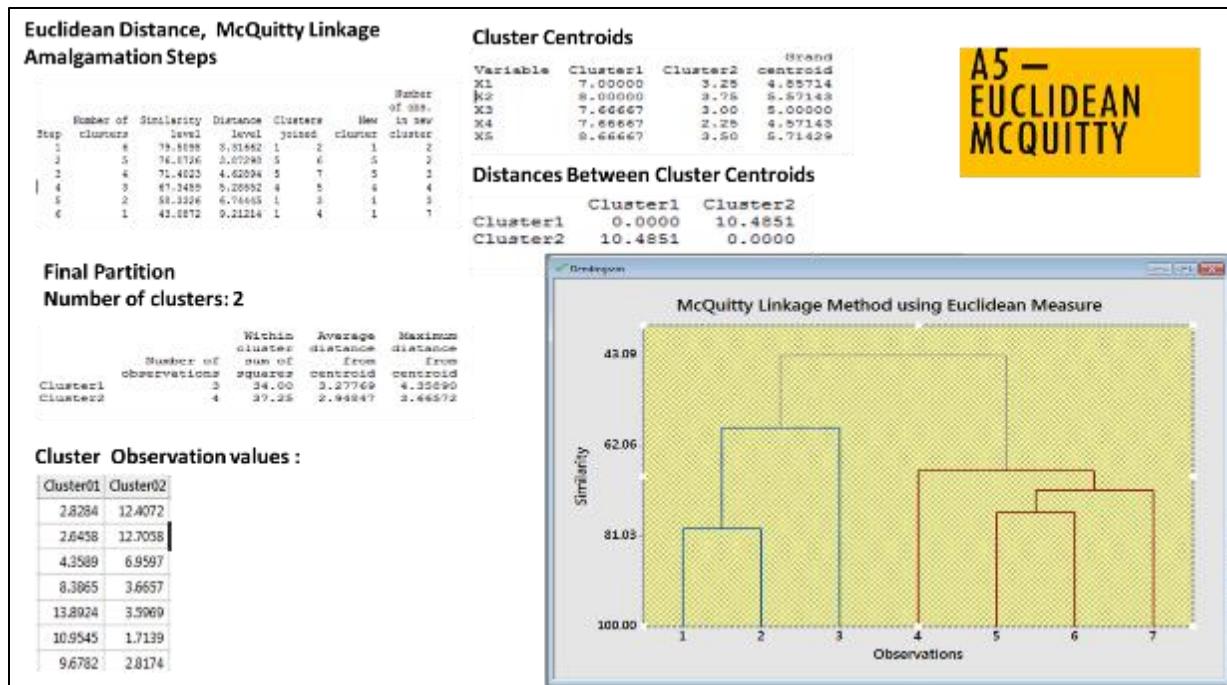


FIGURE 22 : MCQUITTY LINKAGE USING EUCLIDEAN DISTANCE MEASURE FOR ACTUAL VALUES

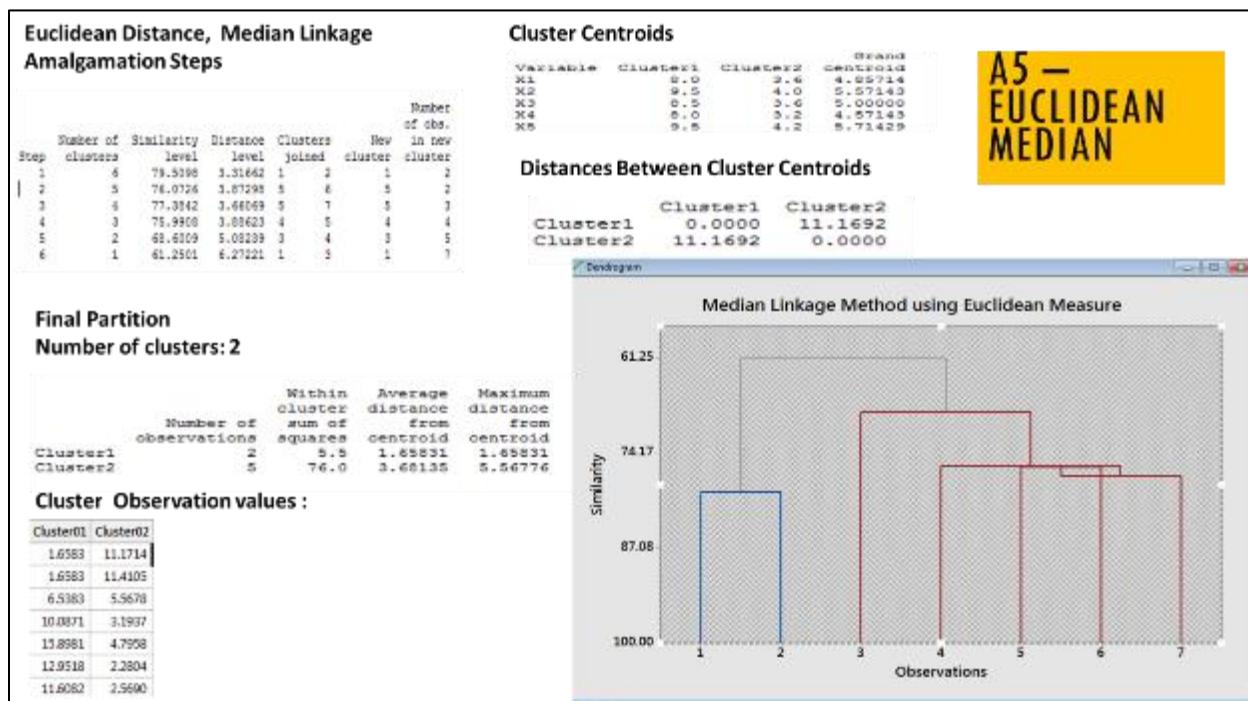


FIGURE 23 : MEDIAN LINKAGE USING EUCLIDEAN DISTANCE MEASURE FOR ACTUAL VALUES

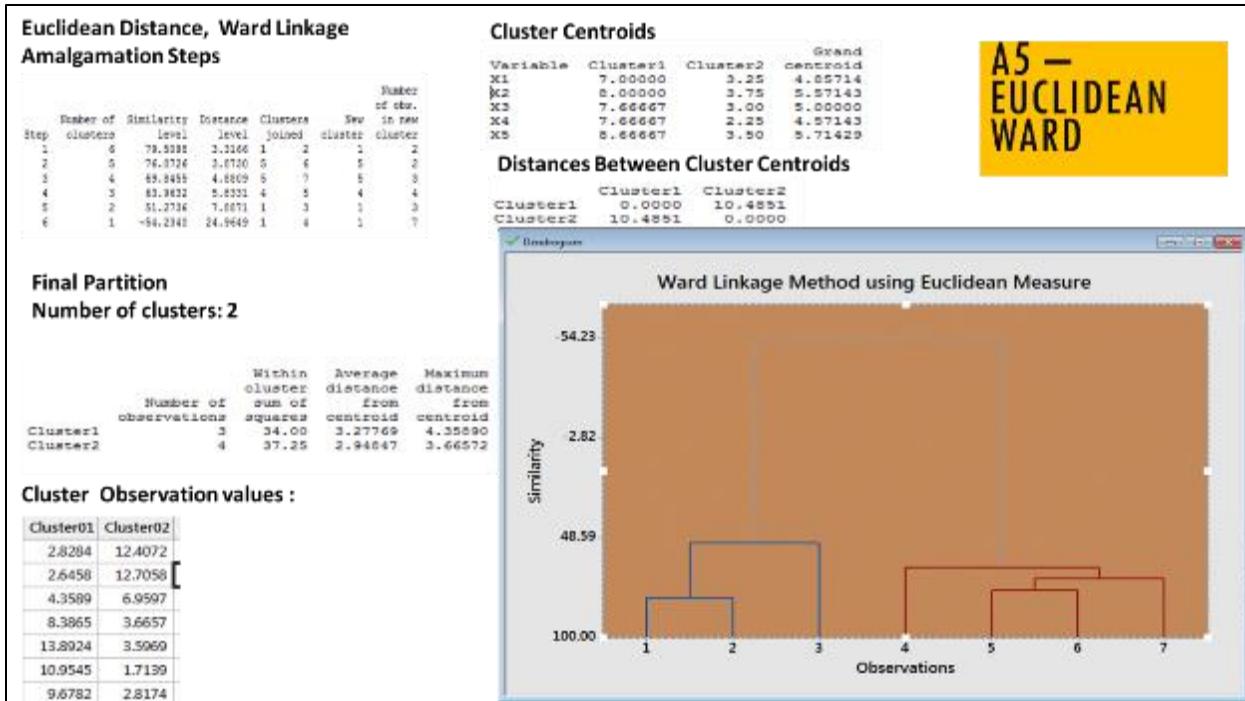


FIGURE 24 : WARD LINKAGE USING EUCLIDEAN DISTANCE MEASURE FOR ACTUAL VALUES

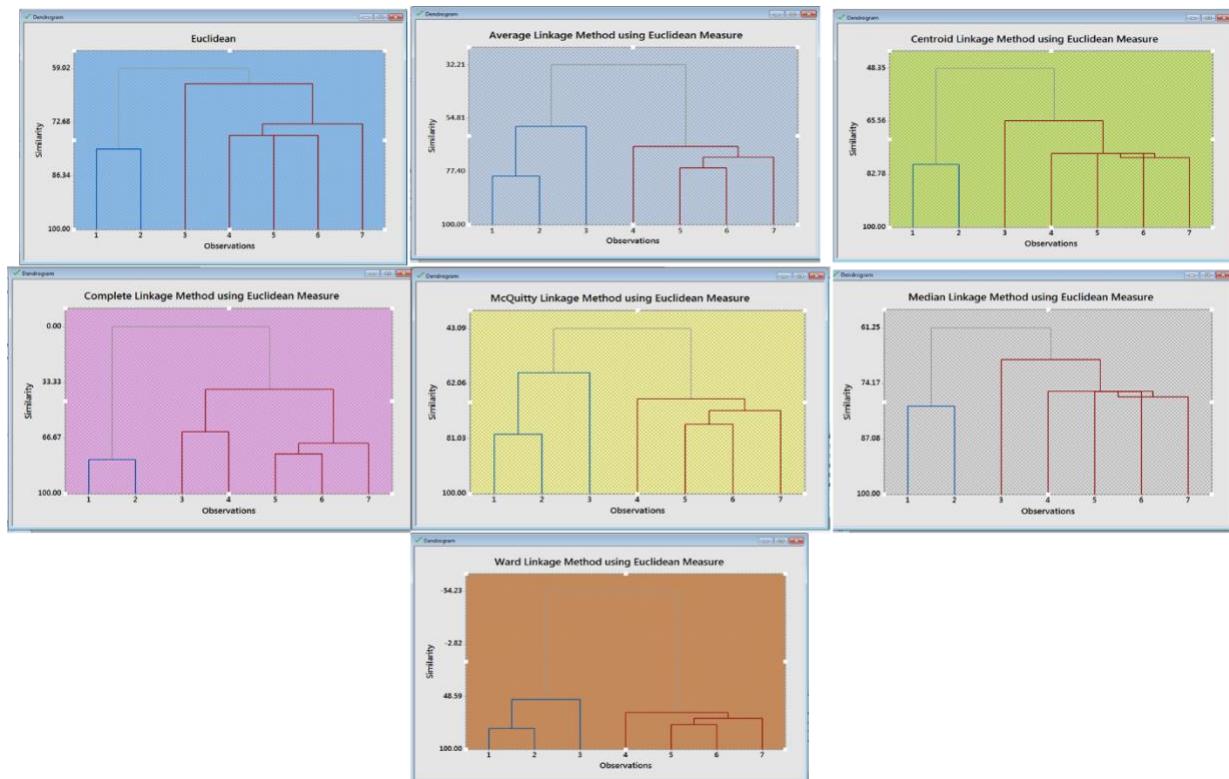


FIGURE 25 : ALL LINKAGES USING EUCLIDEAN DISTANCE MEASURE FOR ACTUAL VALUES

Findings:

Centroid & Median linkages: They tend to join at closest proximities at sub clusters, leading to unclear branches.

Average & McQuitty linkages: They tend to build the clusters from inside – out. The observations 3 & 4 are leaning to either sides, probably not showing the accurate clusters of 3 & 4.

Ward Linkages: This one seem to find high similarity for bottom levels and the similarity level spread out when it grows to Top. However, the dendrograms show that it is an accelerated Average linkage.

Complete Linkage: The only linkage that brought 3 & 4 to a new sub cluster. It tends to build the sub clusters in between other clusters.

General Observation:

1. Though complete linkage & other linkages bring a set of sub clusters, a right one could be chosen based on the actual intent study. (Note: We do not know the variable name of this data)

TABLE 2 : COMPARISON OF A3 & A5

Applying Euclidean Distance measure with	Using Single Linkage (A3)	Using Different Linkages (A5)
Single Linkage	Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[5,c[4,6]]]] Highest similarity level: 79.5 Distance btwn. Cluster Centroids: 11.1692 WSS1: 5.5 WSS2: 76.0	Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[5,c[4,6]]]] Highest similarity level: 79.5 Distance btwn. Cluster Centroids: 11.1692 WSS1: 5.5 WSS2: 76 Avg. Distance from Centroid C1 : 1.65 Avg. Distance from Centroid C2 : 3.68
Average Linkage		Cluster 1 : c[3,c[1,2]] Cluster 2 : c[4,c[7,c[5,6]]] Highest similarity level: 79.5 Distance btwn. Cluster Centroids: 10.485 WSS1: 34 WSS2: 37.25 Avg. Distance from Centroid C1 : 3.27 Avg. Distance from Centroid C2 : 2.94
Centroid Linkage		Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[4,c[5,6]]]] Highest similarity level: 79.5 Distance btwn. Cluster Centroids: 11.1692 WSS1: 5.5 WSS2: 76 Avg. Distance from Centroid C1 : 1.65 Avg. Distance from Centroid C2 : 5.96
Complete Linkage		Cluster 1 : c[1,2] Cluster 2 : c[c[3,4],c[7,c[5,6]]] Highest similarity level: 79.5 Distance btwn. Cluster Centroids: 11.1692 WSS1: 5.5 WSS2: 76

		Avg. Distance from Centroid C1 : 1.65 Avg. Distance from Centroid C2 : 3.68
McQuitty Linkage		Cluster 1 : c[3,c[1,2]] Cluster 2 : c[4,c[7,c[5,6]]] Highest similarity level: 79.5 Distance btwn. Cluster Centroids: 10.485 WSS1: 34 WSS2: 37.25 Avg. Distance from Centroid C1 : 3.27 Avg. Distance from Centroid C2 : 2.94
Median Linkage		Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[4,c[5,6]]]] Highest similarity level: 79.5 Distance btwn. Cluster Centroids: 11.1692 WSS1: 5.5 WSS2: 76 Avg. Distance from Centroid C1 : 1.65 Avg. Distance from Centroid C2 : 3.68
Ward Linkage		Cluster 1 : c[3,c[1,2]] Cluster 2 : c[4,c[7,c[5,6]]] Highest similarity level: 79.5 Lowest similarity level : -54 Distance btwn. Cluster Centroids: 10.485 WSS1: 34 WSS2: 37.25 Avg. Distance from Centroid C1 : 3.27 Avg. Distance from Centroid C2 : 2.94
Notes	<ul style="list-style-type: none"> In contrast to A5, none of the methods did give cluster one with 3 observations. It was always with 1 & 2. Also none of the methods gave the middle pair with 3 & 4 in cluster 2. 	<ul style="list-style-type: none"> Average, McQuitty & Ward linkages created cluster 1 with observations 1, 2 & 3. These clusters have higher WSS1 values (34) compared to the ones that have only 2 observations in cluster 1. The number of observations in a cluster define the WSS. The clusters in these linkages have almost equal Avg. distances for both centroids (3.27 & 2.94) – i.e., balanced clusters. Only Single linkage has observations 4 & 6 as the low level pair, in cluster 2. Complete linkage tend to create pairs in between two clusters, as it tends to push the clusters away. Ward linkage has similarity levels ranging from 79.5 to -54, lead to extend the tree length.

A6: Cluster Analysis on Observations using “Euclidean Distance” measure and all linkage methods (Standardized Values)

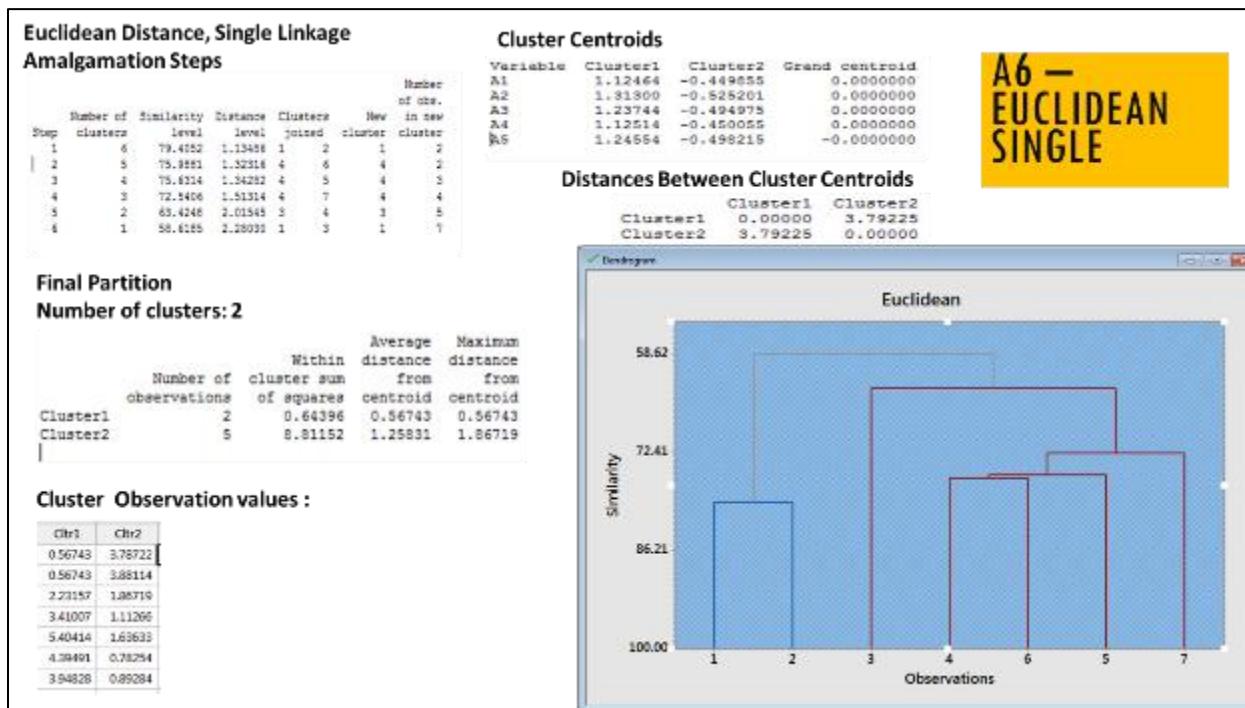


FIGURE 26 : SINGLE LINKAGE USING EUCLIDEAN DISTANCE MEASURE FOR STD VALUES

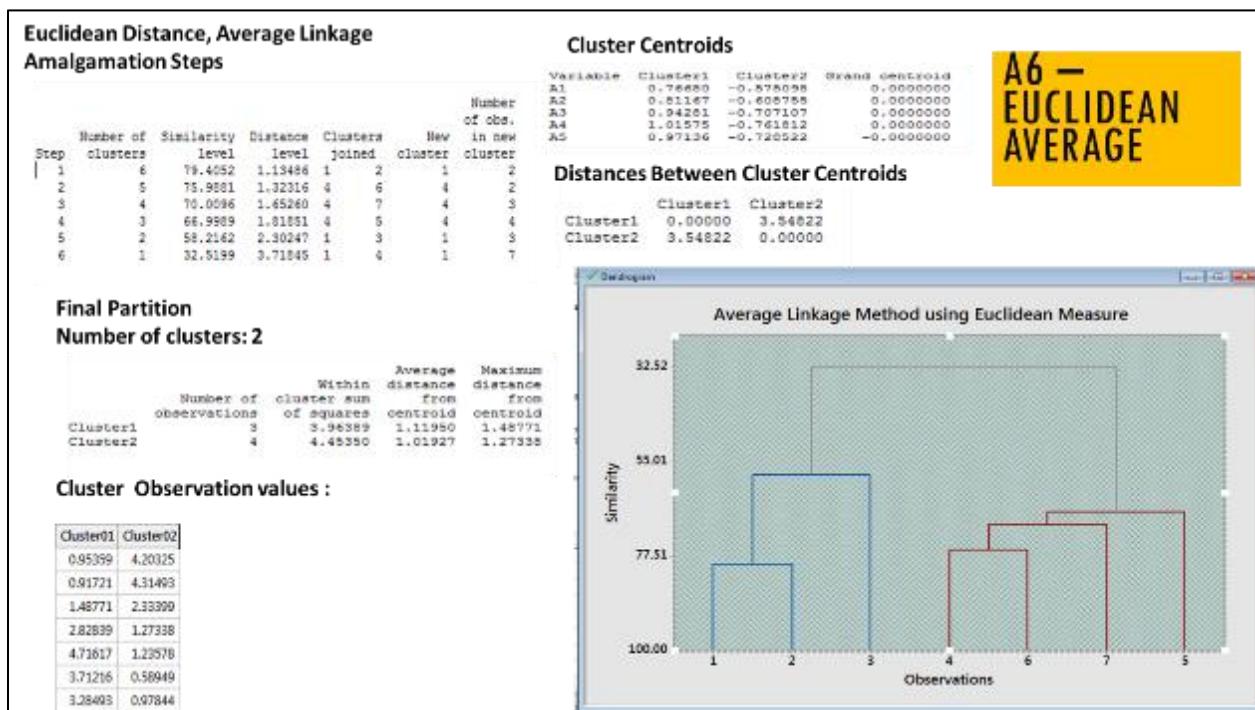


FIGURE 27 : AVERAGE LINKAGE USING EUCLIDEAN DISTANCE MEASURE FOR STD VALUES

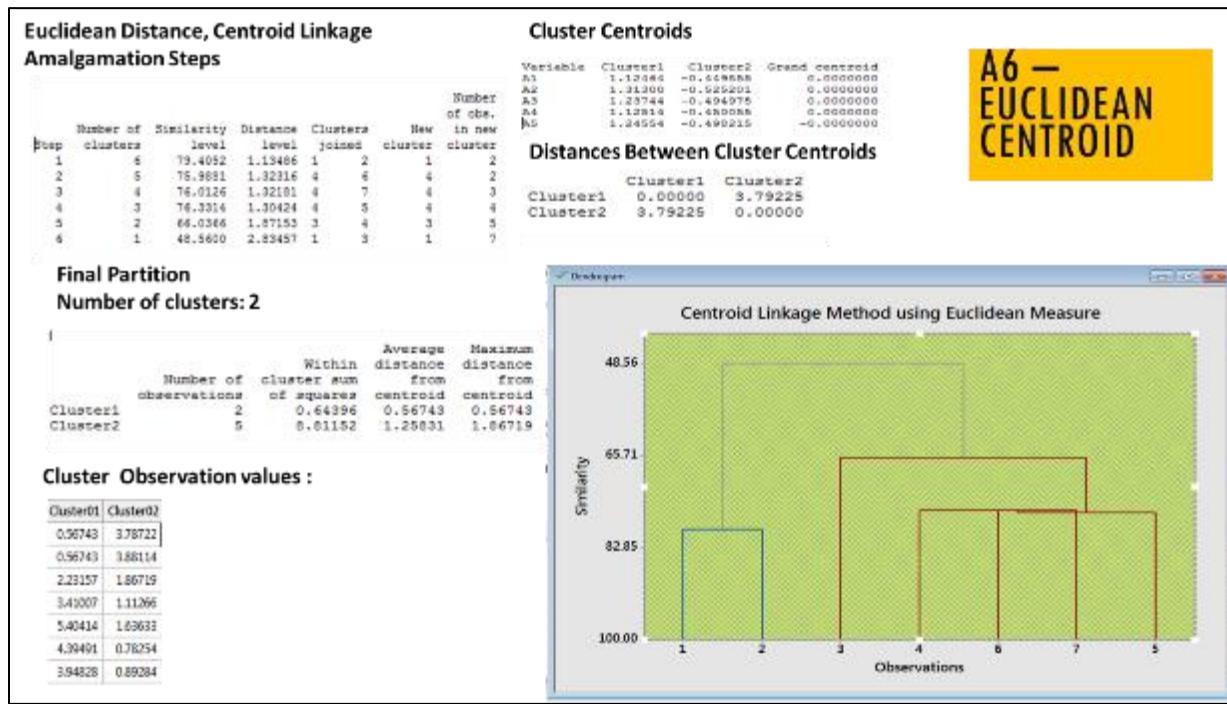


FIGURE 28 : CENTROID LINKAGE USING EUCLIDEAN DISTANCE MEASURE FOR STD VALUES

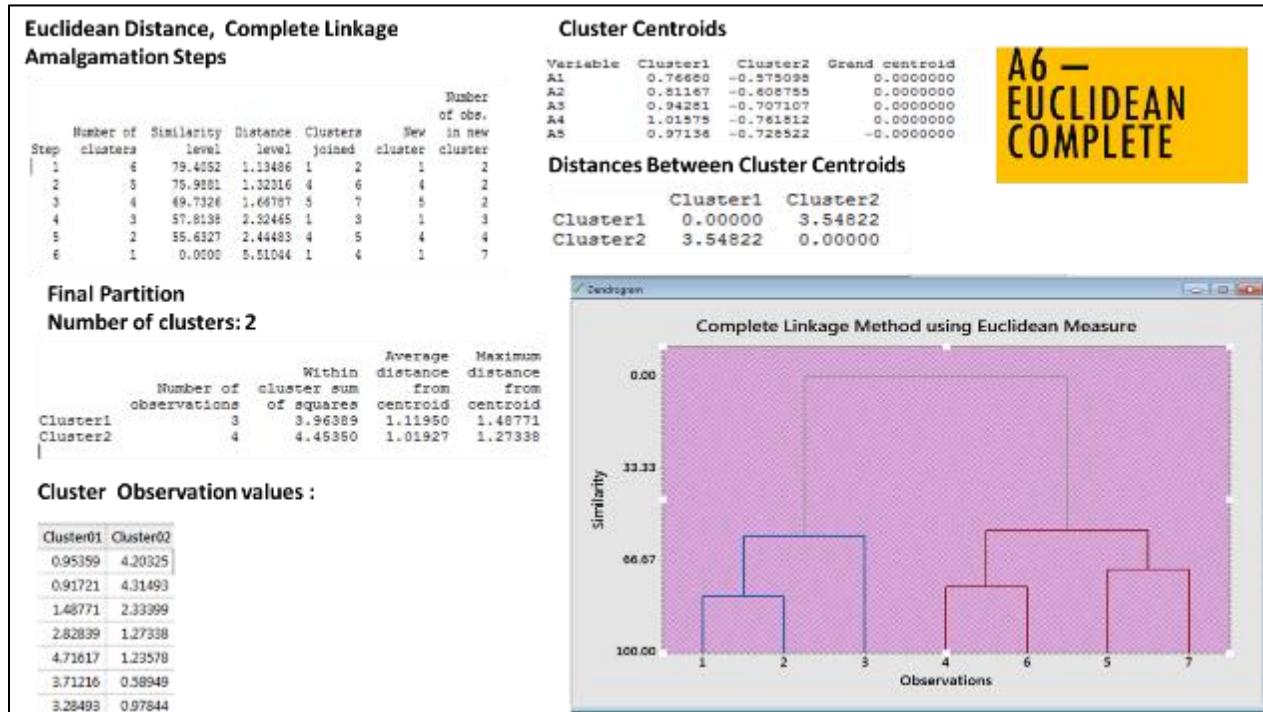


FIGURE 29 : COMPLETE LINKAGE USING EUCLIDEAN DISTANCE MEASURE FOR STD VALUES

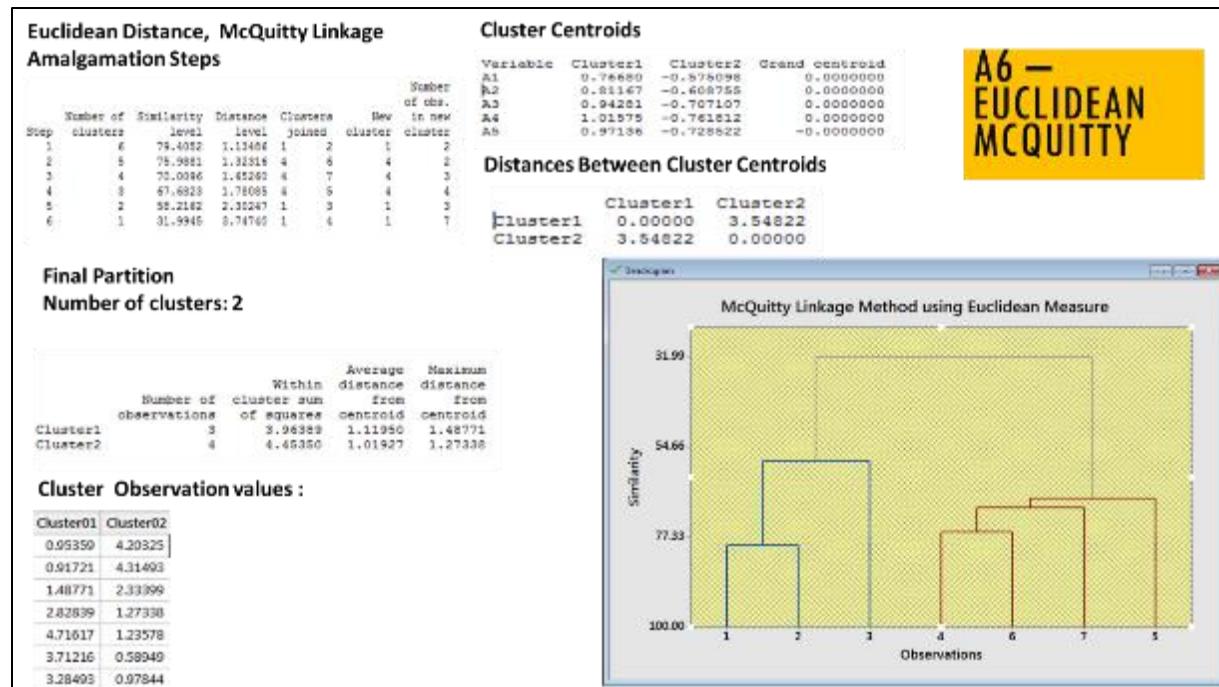


FIGURE 30 : MCQUITTY LINKAGE USING EUCLIDEAN DISTANCE MEASURE FOR STD VALUES

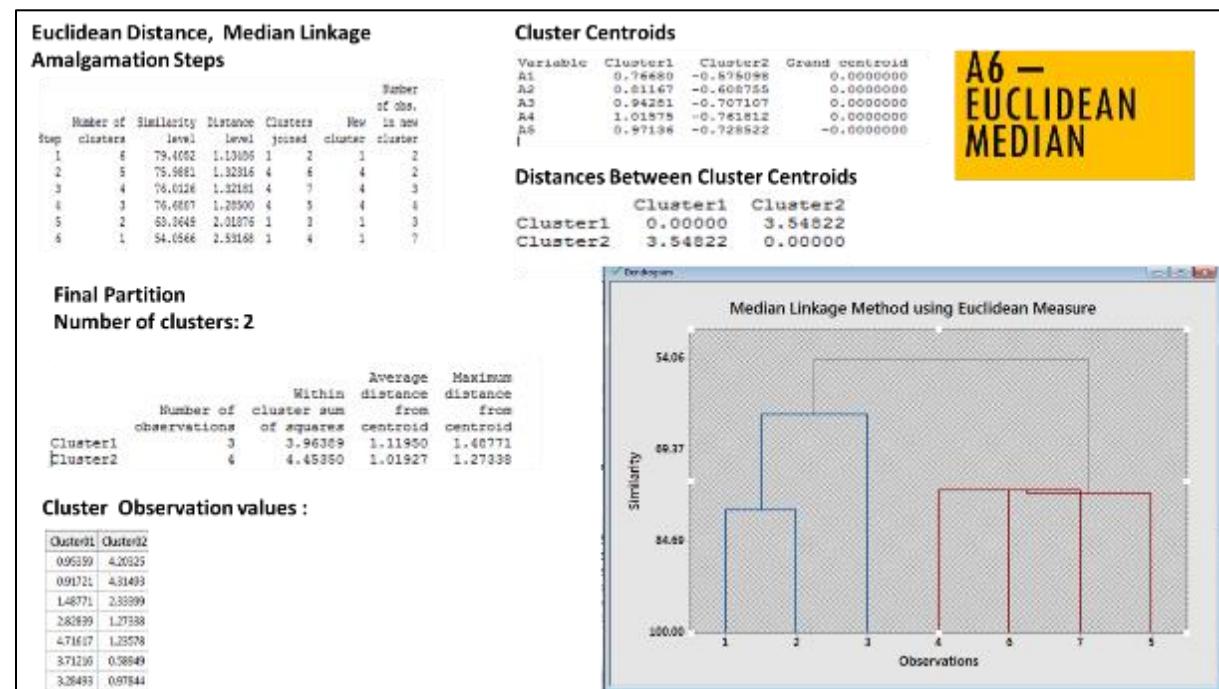


FIGURE 31 : MEDIAN LINKAGE USING EUCLIDEAN DISTANCE MEASURE FOR STD VALUES

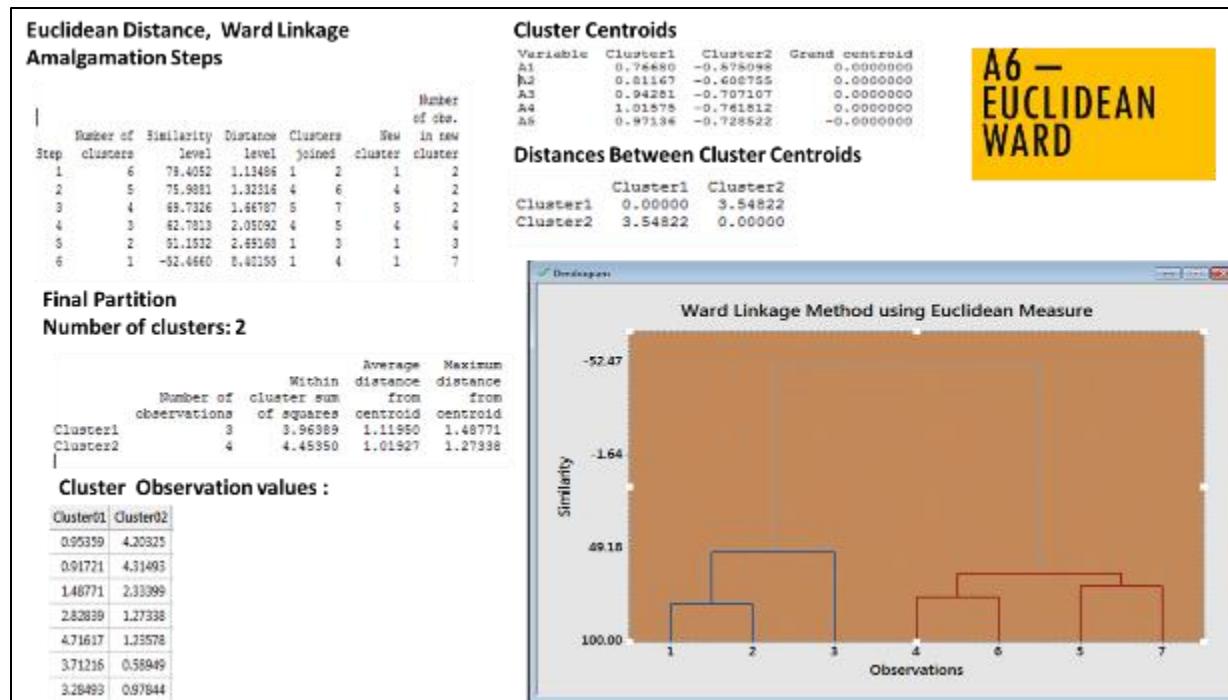


FIGURE 32 : WARD LINKAGE USING EUCLIDEAN DISTANCE MEASURE FOR STD VALUES

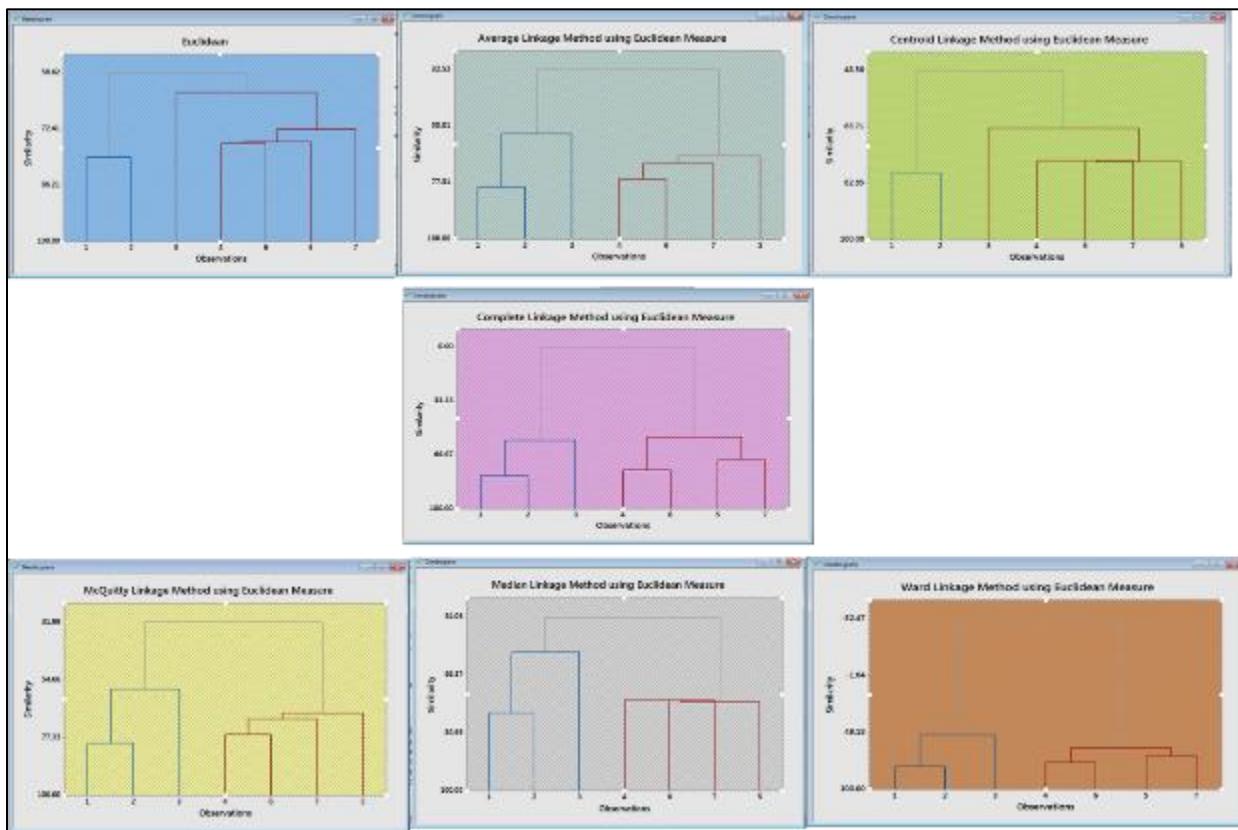


FIGURE 33 : ALL LINKAGE COMPARISON USING EUCLIDEAN DISTANCE MEASURE FOR STD VALUES

TABLE 3 : COMPARISON OF A4, A5 & A6

Applying Euclidean Distance measure with	Using Single Linkage on Standardized values (A4)	Using Different Linkages on actual values (A5)	Using Different Linkages on standardized values (A6)
Single Linkage	Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[5,c[4,6]]]] Highest similarity level: 79.4 Distance btwn. Cluster Centroids: 11.1692 WSS1: 5.5 WSS2: 76 Avg. Distance from Centroid C1 : 1.65 Avg. Distance from Centroid C2 : 3.68	Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[5,c[4,6]]]] Highest similarity level: 79.4 Distance btwn. Cluster Centroids: 11.1692 WSS1: 5.5 WSS2: 76 Avg. Distance from Centroid C1 : 1.65 Avg. Distance from Centroid C2 : 3.68	Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[5,c[4,6]]]] Highest similarity level: 79.4 Distance btwn. Cluster Centroids: 3.79 WSS1: 0.64 WSS2: 8.81 Avg. Distance from Centroid C1 : 0.56 Avg. Distance from Centroid C2 : 1.25
Average Linkage		Cluster 1 : c[3,c[1,2]] Cluster 2 : c[4,c[7,c[5,6]]] Highest similarity level: 79.4 Distance btwn. Cluster Centroids: 10.485 WSS1: 34 WSS2: 37.25 Avg. Distance from Centroid C1 : 3.27 Avg. Distance from Centroid C2 : 2.94	Cluster 1 : c[3,c[1,2]] Cluster 2 : c[5,c[7,c[4,6]]] Highest similarity level: 79.4 Distance btwn. Cluster Centroids: 3.54 WSS1: 3.9 WSS2: 4.4 Avg. Distance from Centroid C1 : 1.11 Avg. Distance from Centroid C2 : 1.01
Centroid Linkage		Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[4,c[5,6]]]] Highest similarity level: 79.4 Distance btwn. Cluster Centroids: 11.1692 WSS1: 5.5 WSS2: 76 Avg. Distance from Centroid C1 : 1.65 Avg. Distance from Centroid C2 : 5.96	Cluster 1 : c[1,2] Cluster 2 : c[3,c[5,c[7,c[4,6]]]] Highest similarity level: 79.4 Distance btwn. Cluster Centroids: 3.79 WSS1: 0.64 WSS2: 8.81 Avg. Distance from Centroid C1 : 0.56 Avg. Distance from Centroid C2 : 1.26
Complete Linkage		Cluster 1 : c[1,2] Cluster 2 : c[c[3,4],c[7,c[5,6]]] Highest similarity level: 79.4 Distance btwn. Cluster Centroids: 11.1692 WSS1: 5.5 WSS2: 76 Avg. Distance from Centroid C1 : 1.65 Avg. Distance from Centroid C2 : 3.68	Cluster 1 : c[3, c[1,2]] Cluster 2 : c[c[4,6],c[5,7]] Highest similarity level: 79.4 Distance btwn. Cluster Centroids: 3.54 WSS1: 3.9 WSS2: 4.4 Avg. Distance from Centroid C1 : 1.1 Avg. Distance from Centroid C2 : 1.0

McQuitty Linkage		<p>Cluster 1 : c[3,c[1,2]]</p> <p>Cluster 2 : c[4,c[7,c[5,6]]]</p> <p>Highest similarity level: 79.4</p> <p>Distance btwn. Cluster Centroids: 10.485</p> <p>WSS1: 34</p> <p>WSS2: 37.25</p> <p>Avg. Distance from Centroid C1 : 3.27</p> <p>Avg. Distance from Centroid C2 : 2.94</p>	<p>Cluster 1 : c[3,c[1,2]]</p> <p>Cluster 2 : c[5,c[7,c[4,6]]]</p> <p>Highest similarity level: 79.4</p> <p>Distance btwn. Cluster Centroids: 3.54</p> <p>WSS1: 3.9</p> <p>WSS2: 4.4</p> <p>Avg. Distance from Centroid C1 : 1.1</p> <p>Avg. Distance from Centroid C2 : 1.0</p>
Median Linkage		<p>Cluster 1 : c[1,2]</p> <p>Cluster 2 : c[3,c[7,c[4,c[5,6]]]]</p> <p>Highest similarity level: 79.5</p> <p>Distance btwn. Cluster Centroids: 11.1692</p> <p>WSS1: 5.5</p> <p>WSS2: 76</p> <p>Avg. Distance from Centroid C1 : 1.65</p> <p>Avg. Distance from Centroid C2 : 3.68</p>	<p>Cluster 1 : c[3,c[1,2]]</p> <p>Cluster 2 : c[5,c[7,c[4,6]]]</p> <p>Highest similarity level: 79.4</p> <p>Distance btwn. Cluster Centroids: 3.54</p> <p>WSS1: 3.9</p> <p>WSS2: 4.4</p> <p>Avg. Distance from Centroid C1 : 1.1</p> <p>Avg. Distance from Centroid C2 : 1.0</p>
Ward Linkage		<p>Cluster 1 : c[3,c[1,2]]</p> <p>Cluster 2 : c[4,c[7,c[5,6]]]</p> <p>Highest similarity level: 79.5</p> <p>Lowest similarity level : -54</p> <p>Distance btwn. Cluster Centroids: 10.485</p> <p>WSS1: 34</p> <p>WSS2: 37.25</p> <p>Avg. Distance from Centroid C1 : 3.27</p> <p>Avg. Distance from Centroid C2 : 2.94</p>	<p>Cluster 1 : c[3,c[1,2]]</p> <p>Cluster 2 : c[c[4,6],c[5,7]]</p> <p>Highest similarity level: 79.4</p> <p>Distance btwn. Cluster Centroids: 3.54</p> <p>WSS1: 3.9</p> <p>WSS2: 4.4</p> <p>Avg. Distance from Centroid C1 : 1.1</p> <p>Avg. Distance from Centroid C2 : 1.0</p>
Findings	<ul style="list-style-type: none"> Single Linkage Euclidean distance method, does not seem to impact the arrangement of pairs in A4, A5 & A6. 	<ul style="list-style-type: none"> Except for single linkage using actual values, rest of the linkages have different cluster 2. 	<ul style="list-style-type: none"> Standardizing values brought the points nearer and except single linkage Euclidean method, rest of the linkages reordered the pairs in cluster 2. Clusters have come nearer, which is evident in distance between cluster centroids value. This is the impact of standardization Smaller WSS values & Avg. distance from centroids indicate the cluster shrinkage, again an impact of standardization.

A7: Cluster Analysis on Variables (Actual Values)

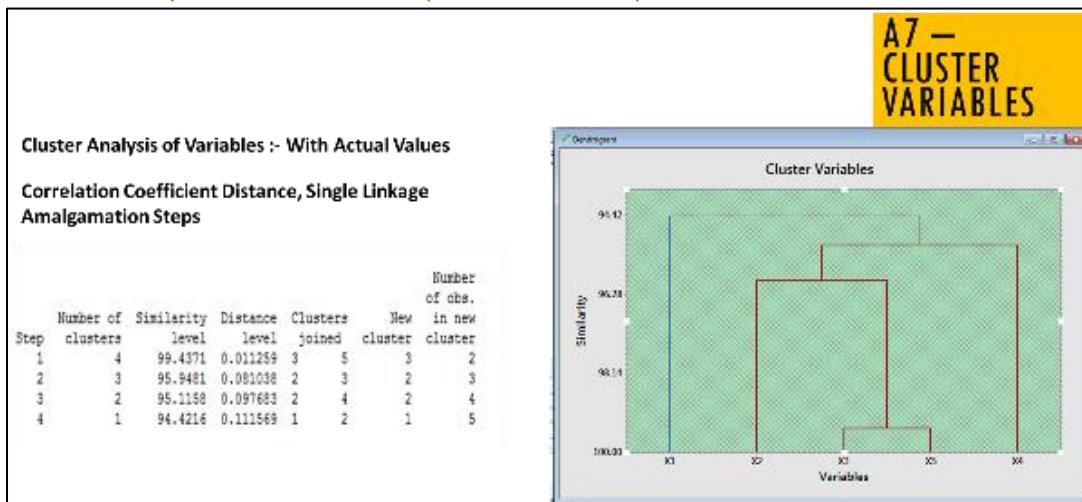


FIGURE 34 : CLUSTER VARIABLES USING ACTUAL VALUES

Findings:

The clusters on variables reflect the observation made in A1. The variable X1 is the least correlated to the other variables. The most correlated variables X3 & X5 was found as the first similar pair, which is expected.

A8: Cluster Analysis on Variables (Standardized Values)

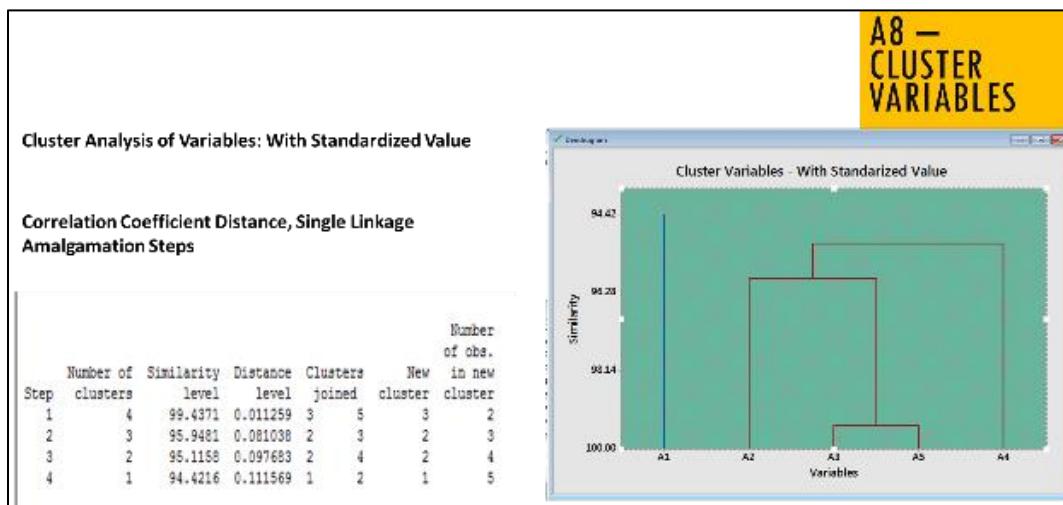


FIGURE 35 : CLUSTER OF VARIABLES USING STANDARDIZED VALUES

TABLE 4 : COMPARISON OF A7 & A8

Analysis on Variables using Single Linkage on actual values (A7)	Analysis on Variables using Single Linkage on actual values (A8)
Cluster 1 : c[1] Cluster 2 : c[2,c[4,c[3,5]]] Highest similarity level: 99.4 Lowest similarity level: 94.4 <ul style="list-style-type: none"> No differences observed between A7 and A8. As all the variables were standardized in A8, it is expected to retain the correlation among the variables. 	Cluster 1 : c[1] Cluster 2 : c[2,c[4,c[3,5]]] Highest similarity level: 99.4 Lowest similarity level: 94.4

A9: Cluster Analysis on Observations for variables XC1 (Actual Values)

The variables X3 & X5 have the high correlation, so the variable X3 is dropped from this analysis. XC1 → {X1, X2, X4, X5}.

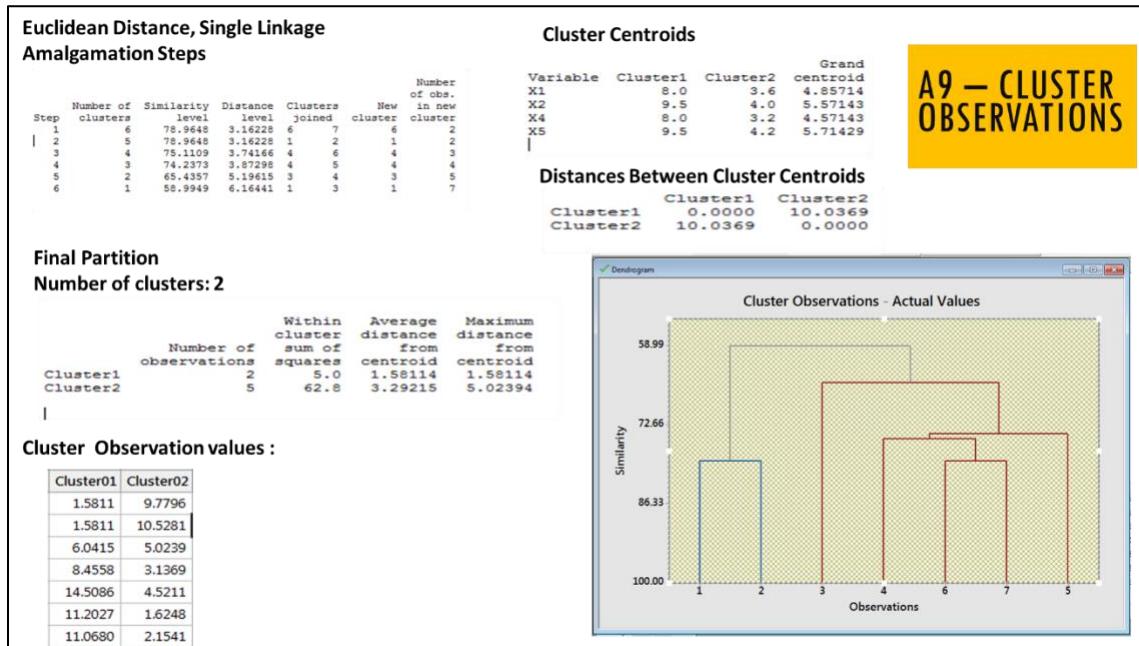


FIGURE 36 : SINGLE LINKAGE USING EUCLIDEAN DISTANCE MEASURE FOR XC1

TABLE 5 : COMPARISON OF A9 & A3

Applying Euclidean Distance measure with	With Xc1 (X1, X2, X4 & X5) variables (A9)	With 5 variables (A3)
Single Linkage	Cluster 1 : c[1,2] Cluster 2 : c[3,c[5,c[4,c[6,7]]]] Highest similarity level: 83.2 Distance btwn. Cluster Centroids: 10.03 WSS1: 5 WSS2: 62.8 Avg. Distance from Centroid C1 : 1.58	Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[4,c[5,6]]]] Highest similarity level: 79.5 Distance btwn. Cluster Centroids: 11.1692 WSS1: 5.5 WSS2: 76.0 Avg. Distance from Centroid C1 : 1.6

	Avg. Distance from Centroid C2 : 3.2	Avg. Distance from Centroid C2 : 3.6																																																																																																																																										
	<p>Euclidean Distance, Single Linkage Amalgamation Steps</p> <table border="1"> <thead> <tr> <th>Step</th> <th>Number of clusters</th> <th>Similarity level</th> <th>Distance level</th> <th>Clusters joined</th> <th>New cluster</th> <th>Number of obs. in new cluster</th> </tr> </thead> <tbody> <tr><td>1</td><td>6</td><td>78.9648</td><td>3.16228</td><td>6</td><td>7</td><td>6</td></tr> <tr><td>2</td><td>5</td><td>78.9648</td><td>3.16228</td><td>1</td><td>2</td><td>2</td></tr> <tr><td>3</td><td>4</td><td>75.1109</td><td>3.74166</td><td>4</td><td>6</td><td>4</td></tr> <tr><td>4</td><td>3</td><td>74.2377</td><td>3.87298</td><td>4</td><td>5</td><td>4</td></tr> <tr><td>5</td><td>2</td><td>65.4357</td><td>5.19615</td><td>3</td><td>4</td><td>5</td></tr> <tr><td>6</td><td>1</td><td>58.9949</td><td>6.16441</td><td>1</td><td>3</td><td>7</td></tr> </tbody> </table> <p>Final Partition Number of clusters: 2</p> <table border="1"> <thead> <tr> <th></th> <th>Within cluster distance</th> <th>Average distance from centroid</th> <th>Maximum distance from centroid</th> </tr> </thead> <tbody> <tr><td>Number of observations</td><td>Number of squares</td><td>centroid</td><td>centroid</td></tr> <tr><td>Cluster1</td><td>2</td><td>5.0</td><td>1.58114</td></tr> <tr><td>Cluster2</td><td>5</td><td>62.8</td><td>3.29215</td></tr> <tr><td></td><td></td><td></td><td>5.02394</td></tr> </tbody> </table>	Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster	1	6	78.9648	3.16228	6	7	6	2	5	78.9648	3.16228	1	2	2	3	4	75.1109	3.74166	4	6	4	4	3	74.2377	3.87298	4	5	4	5	2	65.4357	5.19615	3	4	5	6	1	58.9949	6.16441	1	3	7		Within cluster distance	Average distance from centroid	Maximum distance from centroid	Number of observations	Number of squares	centroid	centroid	Cluster1	2	5.0	1.58114	Cluster2	5	62.8	3.29215				5.02394	<p>Euclidean Distance, Single Linkage Amalgamation Steps</p> <table border="1"> <thead> <tr> <th>Step</th> <th>Number of clusters</th> <th>Similarity level</th> <th>Distance level</th> <th>Clusters joined</th> <th>New cluster</th> <th>Number of obs. in new cluster</th> </tr> </thead> <tbody> <tr><td>1</td><td>6</td><td>79.5098</td><td>3.31642</td><td>1</td><td>2</td><td>1</td></tr> <tr><td>2</td><td>5</td><td>76.0726</td><td>3.87298</td><td>5</td><td>6</td><td>2</td></tr> <tr><td>3</td><td>4</td><td>76.0726</td><td>3.87298</td><td>4</td><td>5</td><td>4</td></tr> <tr><td>4</td><td>3</td><td>75.1109</td><td>4.16228</td><td>4</td><td>7</td><td>4</td></tr> <tr><td>5</td><td>2</td><td>65.9319</td><td>4.00000</td><td>3</td><td>4</td><td>3</td></tr> <tr><td>6</td><td>1</td><td>59.0196</td><td>6.63325</td><td>1</td><td>3</td><td>7</td></tr> </tbody> </table> <p>Final Partition Number of clusters: 2</p> <table border="1"> <thead> <tr> <th></th> <th>Within cluster distance</th> <th>Average distance from centroid</th> <th>Maximum distance from centroid</th> </tr> </thead> <tbody> <tr><td>Number of observations</td><td>Number of squares</td><td>centroid</td><td>centroid</td></tr> <tr><td>Cluster1</td><td>2</td><td>5.5</td><td>1.65831</td></tr> <tr><td>Cluster2</td><td>5</td><td>76.0</td><td>3.68135</td></tr> <tr><td></td><td></td><td></td><td>5.56776</td></tr> </tbody> </table>	Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster	1	6	79.5098	3.31642	1	2	1	2	5	76.0726	3.87298	5	6	2	3	4	76.0726	3.87298	4	5	4	4	3	75.1109	4.16228	4	7	4	5	2	65.9319	4.00000	3	4	3	6	1	59.0196	6.63325	1	3	7		Within cluster distance	Average distance from centroid	Maximum distance from centroid	Number of observations	Number of squares	centroid	centroid	Cluster1	2	5.5	1.65831	Cluster2	5	76.0	3.68135				5.56776
Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster																																																																																																																																						
1	6	78.9648	3.16228	6	7	6																																																																																																																																						
2	5	78.9648	3.16228	1	2	2																																																																																																																																						
3	4	75.1109	3.74166	4	6	4																																																																																																																																						
4	3	74.2377	3.87298	4	5	4																																																																																																																																						
5	2	65.4357	5.19615	3	4	5																																																																																																																																						
6	1	58.9949	6.16441	1	3	7																																																																																																																																						
	Within cluster distance	Average distance from centroid	Maximum distance from centroid																																																																																																																																									
Number of observations	Number of squares	centroid	centroid																																																																																																																																									
Cluster1	2	5.0	1.58114																																																																																																																																									
Cluster2	5	62.8	3.29215																																																																																																																																									
			5.02394																																																																																																																																									
Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster																																																																																																																																						
1	6	79.5098	3.31642	1	2	1																																																																																																																																						
2	5	76.0726	3.87298	5	6	2																																																																																																																																						
3	4	76.0726	3.87298	4	5	4																																																																																																																																						
4	3	75.1109	4.16228	4	7	4																																																																																																																																						
5	2	65.9319	4.00000	3	4	3																																																																																																																																						
6	1	59.0196	6.63325	1	3	7																																																																																																																																						
	Within cluster distance	Average distance from centroid	Maximum distance from centroid																																																																																																																																									
Number of observations	Number of squares	centroid	centroid																																																																																																																																									
Cluster1	2	5.5	1.65831																																																																																																																																									
Cluster2	5	76.0	3.68135																																																																																																																																									
			5.56776																																																																																																																																									
	<p>Cluster Observation values :</p> <table border="1"> <thead> <tr><th>Cluster01</th><th>Cluster02</th></tr> </thead> <tbody> <tr><td>1.5811</td><td>9.7796</td></tr> <tr><td>1.5811</td><td>10.5281</td></tr> <tr><td>6.0415</td><td>5.0239</td></tr> <tr><td>8.4558</td><td>3.1369</td></tr> <tr><td>14.5086</td><td>4.5211</td></tr> <tr><td>11.2027</td><td>1.6248</td></tr> <tr><td>11.0680</td><td>2.1541</td></tr> </tbody> </table> <p>Cluster Centroids</p> <table border="1"> <thead> <tr> <th>Variable</th> <th>Cluster1</th> <th>Cluster2</th> <th>Grand centroid</th> </tr> </thead> <tbody> <tr><td>X1</td><td>8.5</td><td>3.0</td><td>4.45144</td></tr> <tr><td>X2</td><td>8.5</td><td>4.0</td><td>5.57143</td></tr> <tr><td>X3</td><td>8.0</td><td>3.2</td><td>4.57143</td></tr> <tr><td>X5</td><td>9.5</td><td>4.2</td><td>5.71429</td></tr> </tbody> </table> <p>Distances Between Cluster Centroids</p> <table border="1"> <thead> <tr> <th></th> <th>Cluster1</th> <th>Cluster2</th> </tr> </thead> <tbody> <tr><td>Cluster1</td><td>0.0000</td><td>10.0369</td></tr> <tr><td>Cluster2</td><td>10.0369</td><td>0.0000</td></tr> </tbody> </table>	Cluster01	Cluster02	1.5811	9.7796	1.5811	10.5281	6.0415	5.0239	8.4558	3.1369	14.5086	4.5211	11.2027	1.6248	11.0680	2.1541	Variable	Cluster1	Cluster2	Grand centroid	X1	8.5	3.0	4.45144	X2	8.5	4.0	5.57143	X3	8.0	3.2	4.57143	X5	9.5	4.2	5.71429		Cluster1	Cluster2	Cluster1	0.0000	10.0369	Cluster2	10.0369	0.0000	<p>Cluster Observation values :</p> <table border="1"> <thead> <tr><th>Cluster1</th><th>Cluster2</th></tr> </thead> <tbody> <tr><td>1.6583</td><td>11.1714</td></tr> <tr><td>1.6583</td><td>11.4105</td></tr> <tr><td>6.5383</td><td>5.5678</td></tr> <tr><td>10.0871</td><td>3.1937</td></tr> <tr><td>15.8981</td><td>4.7958</td></tr> <tr><td>12.9518</td><td>2.2804</td></tr> <tr><td>11.6082</td><td>2.5690</td></tr> </tbody> </table> <p>Cluster Centroids</p> <table border="1"> <thead> <tr> <th>Variable</th> <th>Cluster1</th> <th>Cluster2</th> <th>Grand centroid</th> </tr> </thead> <tbody> <tr><td>X1</td><td>8.0</td><td>3.6</td><td>4.85714</td></tr> <tr><td>X2</td><td>9.5</td><td>4.0</td><td>5.57143</td></tr> <tr><td>X3</td><td>8.5</td><td>3.6</td><td>5.00000</td></tr> <tr><td>X4</td><td>8.0</td><td>3.2</td><td>4.57143</td></tr> <tr><td>X5</td><td>9.5</td><td>4.2</td><td>5.71429</td></tr> </tbody> </table> <p>Distances Between Cluster Centroids</p> <table border="1"> <thead> <tr> <th></th> <th>Cluster1</th> <th>Cluster2</th> </tr> </thead> <tbody> <tr><td>Cluster1</td><td>0.0000</td><td>11.1692</td></tr> <tr><td>Cluster2</td><td>11.1692</td><td>0.0000</td></tr> </tbody> </table>	Cluster1	Cluster2	1.6583	11.1714	1.6583	11.4105	6.5383	5.5678	10.0871	3.1937	15.8981	4.7958	12.9518	2.2804	11.6082	2.5690	Variable	Cluster1	Cluster2	Grand centroid	X1	8.0	3.6	4.85714	X2	9.5	4.0	5.57143	X3	8.5	3.6	5.00000	X4	8.0	3.2	4.57143	X5	9.5	4.2	5.71429		Cluster1	Cluster2	Cluster1	0.0000	11.1692	Cluster2	11.1692	0.0000																																												
Cluster01	Cluster02																																																																																																																																											
1.5811	9.7796																																																																																																																																											
1.5811	10.5281																																																																																																																																											
6.0415	5.0239																																																																																																																																											
8.4558	3.1369																																																																																																																																											
14.5086	4.5211																																																																																																																																											
11.2027	1.6248																																																																																																																																											
11.0680	2.1541																																																																																																																																											
Variable	Cluster1	Cluster2	Grand centroid																																																																																																																																									
X1	8.5	3.0	4.45144																																																																																																																																									
X2	8.5	4.0	5.57143																																																																																																																																									
X3	8.0	3.2	4.57143																																																																																																																																									
X5	9.5	4.2	5.71429																																																																																																																																									
	Cluster1	Cluster2																																																																																																																																										
Cluster1	0.0000	10.0369																																																																																																																																										
Cluster2	10.0369	0.0000																																																																																																																																										
Cluster1	Cluster2																																																																																																																																											
1.6583	11.1714																																																																																																																																											
1.6583	11.4105																																																																																																																																											
6.5383	5.5678																																																																																																																																											
10.0871	3.1937																																																																																																																																											
15.8981	4.7958																																																																																																																																											
12.9518	2.2804																																																																																																																																											
11.6082	2.5690																																																																																																																																											
Variable	Cluster1	Cluster2	Grand centroid																																																																																																																																									
X1	8.0	3.6	4.85714																																																																																																																																									
X2	9.5	4.0	5.57143																																																																																																																																									
X3	8.5	3.6	5.00000																																																																																																																																									
X4	8.0	3.2	4.57143																																																																																																																																									
X5	9.5	4.2	5.71429																																																																																																																																									
	Cluster1	Cluster2																																																																																																																																										
Cluster1	0.0000	11.1692																																																																																																																																										
Cluster2	11.1692	0.0000																																																																																																																																										
Findings	<ul style="list-style-type: none"> 6&7 pair was the first nearest pair found at similarity level 78.9 Comparatively the clusters are nearer (10.03) WSS values indicate that clusters differ in size. The cluster 2 has shrunk a bit compared with A3. Has a sense of clear division of clusters and sub-clusters (may be caused by dropping one of the correlated variables – X3) 	<ul style="list-style-type: none"> 5 & 6 was the nearest pair found at similarity level 79.5 Comparatively the clusters are farther to each other (11.16) WSS values indicate that the clusters differ in size The division of clusters are clumsy (esp. 4,5,6) 																																																																																																																																										

A10: Cluster Analysis on Observations for variables XC1 (Standardized values)

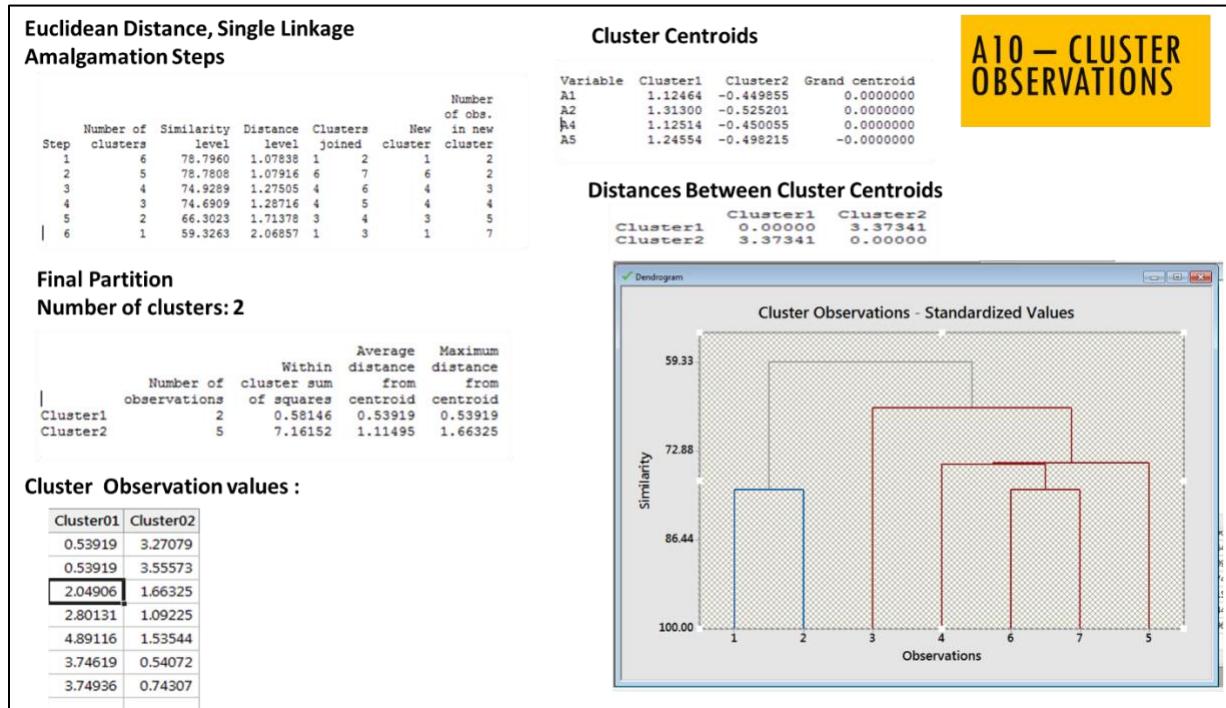


FIGURE 37 : CLUSTER ANALYSIS OF OBSERVATIONS FOR STD. XC1 VALUES

TABLE 6 : COMPARISON OF A2, A4 & A10

Analysis on Observations using all 5 variables on std.values					Analysis on Observations using all 5 variables on std.values (A4)	Analysis on Observations using XC1 (4 variables) on std.values (A10)																																													
<table border="1"> <thead> <tr> <th>C6</th><th>C7</th><th>C8</th><th>C9</th><th>C10</th></tr> <tr> <th>A1</th><th>A2</th><th>A3</th><th>A4</th><th>A5</th></tr> </thead> <tbody> <tr> <td>0.76680</td><td>1.48011</td><td>1.41421</td><td>0.79697</td><td>1.41004</td></tr> <tr> <td>1.48248</td><td>1.14589</td><td>1.06066</td><td>1.45330</td><td>1.08103</td></tr> <tr> <td>0.05112</td><td>-0.19098</td><td>0.35355</td><td>0.79697</td><td>0.42301</td></tr> <tr> <td>0.40896</td><td>0.14324</td><td>-0.70711</td><td>-0.51569</td><td>-0.56402</td></tr> <tr> <td>-1.38024</td><td>-1.19364</td><td>-1.06066</td><td>-1.17202</td><td>-1.22204</td></tr> <tr> <td>-0.30672</td><td>-0.85942</td><td>-1.06066</td><td>-0.51569</td><td>-0.89303</td></tr> <tr> <td>-1.02240</td><td>-0.52520</td><td>0.00000</td><td>-0.84385</td><td>-0.23501</td></tr> </tbody> </table>					C6	C7	C8	C9	C10	A1	A2	A3	A4	A5	0.76680	1.48011	1.41421	0.79697	1.41004	1.48248	1.14589	1.06066	1.45330	1.08103	0.05112	-0.19098	0.35355	0.79697	0.42301	0.40896	0.14324	-0.70711	-0.51569	-0.56402	-1.38024	-1.19364	-1.06066	-1.17202	-1.22204	-0.30672	-0.85942	-1.06066	-0.51569	-0.89303	-1.02240	-0.52520	0.00000	-0.84385	-0.23501	Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[5,c[4,6]]]] Highest similarity level: 79.4 Distance btwn. Cluster Centroids: 3.792 WSS1: 0.643 WSS2: 8.811 Avg. Distance from Centroid C1 : 0.56 Avg. Distance from Centroid C2 : 1.25	Cluster 1 : c[1,2] Cluster 2 : c[3,c[5,c[4,c[6,7]]]] Highest similarity level: 78.79 Distance btwn. Cluster Centroids: 3.37 WSS1: 0.5 WSS2: 7.1 Avg. Distance from Centroid C1 : 0.5 Avg. Distance from Centroid C2 : 1.1
C6	C7	C8	C9	C10																																															
A1	A2	A3	A4	A5																																															
0.76680	1.48011	1.41421	0.79697	1.41004																																															
1.48248	1.14589	1.06066	1.45330	1.08103																																															
0.05112	-0.19098	0.35355	0.79697	0.42301																																															
0.40896	0.14324	-0.70711	-0.51569	-0.56402																																															
-1.38024	-1.19364	-1.06066	-1.17202	-1.22204																																															
-0.30672	-0.85942	-1.06066	-0.51569	-0.89303																																															
-1.02240	-0.52520	0.00000	-0.84385	-0.23501																																															

<p>Correlation: A1, A2, A3, A4, A5</p> <table border="1"> <thead> <tr> <th></th> <th>A1</th> <th>A2</th> <th>A3</th> <th>A4</th> </tr> </thead> <tbody> <tr> <td>A2</td> <td>0.888 0.008</td> <td></td> <td></td> <td></td> </tr> <tr> <td>A3</td> <td>0.696 0.083</td> <td>0.886 0.008</td> <td></td> <td></td> </tr> <tr> <td>A4</td> <td>0.872 0.010</td> <td>0.817 0.029</td> <td>0.851 0.015</td> <td></td> </tr> <tr> <td>A5</td> <td>0.779 0.039</td> <td>0.919 0.003</td> <td>0.989 0.000</td> <td>0.902 0.005</td> </tr> </tbody> </table> <p>Cell Contents: Pearson correlation P-Value</p>		A1	A2	A3	A4	A2	0.888 0.008				A3	0.696 0.083	0.886 0.008			A4	0.872 0.010	0.817 0.029	0.851 0.015		A5	0.779 0.039	0.919 0.003	0.989 0.000	0.902 0.005	<p>Euclidean Distance, Single Linkage Amalgamation Steps</p> <table border="1"> <thead> <tr> <th>Step</th> <th>Number of clusters</th> <th>Similarity level</th> <th>Distance level</th> <th>Clusters joined</th> <th>New cluster</th> <th>Number of obs. in new cluster</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>6</td> <td>79.9881</td> <td>1.32216</td> <td>1</td> <td>2</td> <td>2</td> </tr> <tr> <td>2</td> <td>5</td> <td>75.9981</td> <td>1.32216</td> <td>4</td> <td>5</td> <td>2</td> </tr> <tr> <td>3</td> <td>4</td> <td>75.4314</td> <td>1.34282</td> <td>4</td> <td>5</td> <td>3</td> </tr> <tr> <td>4</td> <td>3</td> <td>72.5406</td> <td>1.81914</td> <td>4</td> <td>7</td> <td>4</td> </tr> <tr> <td>5</td> <td>2</td> <td>63.2424</td> <td>1.81914</td> <td>3</td> <td>4</td> <td>5</td> </tr> <tr> <td>6</td> <td>1</td> <td>56.6188</td> <td>2.28030</td> <td>1</td> <td>3</td> <td>7</td> </tr> </tbody> </table> <p>Final Partition Number of clusters: 2</p> <table border="1"> <thead> <tr> <th>Cluster</th> <th>Number of observations</th> <th>Average distance from centroid</th> <th>Within cluster sum of squares</th> <th>Maximum distance from centroid</th> </tr> </thead> <tbody> <tr> <td>Cluster1</td> <td>2</td> <td>0.64396</td> <td>0.56743</td> <td>0.56743</td> </tr> <tr> <td>Cluster2</td> <td>5</td> <td>8.81152</td> <td>1.25831</td> <td>1.86719</td> </tr> </tbody> </table>	Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster	1	6	79.9881	1.32216	1	2	2	2	5	75.9981	1.32216	4	5	2	3	4	75.4314	1.34282	4	5	3	4	3	72.5406	1.81914	4	7	4	5	2	63.2424	1.81914	3	4	5	6	1	56.6188	2.28030	1	3	7	Cluster	Number of observations	Average distance from centroid	Within cluster sum of squares	Maximum distance from centroid	Cluster1	2	0.64396	0.56743	0.56743	Cluster2	5	8.81152	1.25831	1.86719	<p>Euclidean Distance, Single Linkage Amalgamation Steps</p> <table border="1"> <thead> <tr> <th>Step</th> <th>Number of clusters</th> <th>Similarity level</th> <th>Distance level</th> <th>Clusters joined</th> <th>New cluster</th> <th>Number of obs. in new cluster</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>6</td> <td>79.7980</td> <td>1.07838</td> <td>1</td> <td>2</td> <td>2</td> </tr> <tr> <td>2</td> <td>5</td> <td>79.9289</td> <td>1.27905</td> <td>4</td> <td>5</td> <td>2</td> </tr> <tr> <td>3</td> <td>4</td> <td>74.9289</td> <td>1.27905</td> <td>4</td> <td>6</td> <td>3</td> </tr> <tr> <td>4</td> <td>3</td> <td>74.6909</td> <td>1.28716</td> <td>4</td> <td>5</td> <td>4</td> </tr> <tr> <td>5</td> <td>2</td> <td>66.3023</td> <td>1.71378</td> <td>3</td> <td>4</td> <td>5</td> </tr> <tr> <td>6</td> <td>1</td> <td>59.3263</td> <td>2.06937</td> <td>1</td> <td>3</td> <td>7</td> </tr> </tbody> </table> <p>Final Partition Number of clusters: 2</p> <table border="1"> <thead> <tr> <th>Cluster</th> <th>Number of observations</th> <th>Average distance from centroid</th> <th>Within cluster sum of squares</th> <th>Maximum distance from centroid</th> </tr> </thead> <tbody> <tr> <td>Cluster1</td> <td>2</td> <td>0.58146</td> <td>0.53919</td> <td>0.53919</td> </tr> <tr> <td>Cluster2</td> <td>5</td> <td>7.16152</td> <td>1.11495</td> <td>1.66325</td> </tr> </tbody> </table>	Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster	1	6	79.7980	1.07838	1	2	2	2	5	79.9289	1.27905	4	5	2	3	4	74.9289	1.27905	4	6	3	4	3	74.6909	1.28716	4	5	4	5	2	66.3023	1.71378	3	4	5	6	1	59.3263	2.06937	1	3	7	Cluster	Number of observations	Average distance from centroid	Within cluster sum of squares	Maximum distance from centroid	Cluster1	2	0.58146	0.53919	0.53919	Cluster2	5	7.16152	1.11495	1.66325
	A1	A2	A3	A4																																																																																																																																																							
A2	0.888 0.008																																																																																																																																																										
A3	0.696 0.083	0.886 0.008																																																																																																																																																									
A4	0.872 0.010	0.817 0.029	0.851 0.015																																																																																																																																																								
A5	0.779 0.039	0.919 0.003	0.989 0.000	0.902 0.005																																																																																																																																																							
Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster																																																																																																																																																					
1	6	79.9881	1.32216	1	2	2																																																																																																																																																					
2	5	75.9981	1.32216	4	5	2																																																																																																																																																					
3	4	75.4314	1.34282	4	5	3																																																																																																																																																					
4	3	72.5406	1.81914	4	7	4																																																																																																																																																					
5	2	63.2424	1.81914	3	4	5																																																																																																																																																					
6	1	56.6188	2.28030	1	3	7																																																																																																																																																					
Cluster	Number of observations	Average distance from centroid	Within cluster sum of squares	Maximum distance from centroid																																																																																																																																																							
Cluster1	2	0.64396	0.56743	0.56743																																																																																																																																																							
Cluster2	5	8.81152	1.25831	1.86719																																																																																																																																																							
Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster																																																																																																																																																					
1	6	79.7980	1.07838	1	2	2																																																																																																																																																					
2	5	79.9289	1.27905	4	5	2																																																																																																																																																					
3	4	74.9289	1.27905	4	6	3																																																																																																																																																					
4	3	74.6909	1.28716	4	5	4																																																																																																																																																					
5	2	66.3023	1.71378	3	4	5																																																																																																																																																					
6	1	59.3263	2.06937	1	3	7																																																																																																																																																					
Cluster	Number of observations	Average distance from centroid	Within cluster sum of squares	Maximum distance from centroid																																																																																																																																																							
Cluster1	2	0.58146	0.53919	0.53919																																																																																																																																																							
Cluster2	5	7.16152	1.11495	1.66325																																																																																																																																																							
<p>Cluster Observation values :</p> <table border="1"> <thead> <tr> <th>Clrt1</th> <th>Clrt2</th> </tr> </thead> <tbody> <tr> <td>0.56743</td> <td>3.78722</td> </tr> <tr> <td>0.56743</td> <td>3.88114</td> </tr> <tr> <td>2.23157</td> <td>1.86719</td> </tr> <tr> <td>3.41007</td> <td>1.11266</td> </tr> <tr> <td>5.40414</td> <td>1.63633</td> </tr> <tr> <td>4.39491</td> <td>0.78254</td> </tr> <tr> <td>3.94828</td> <td>0.69284</td> </tr> </tbody> </table> <p>Cluster Centroids</p> <table border="1"> <thead> <tr> <th>Variable</th> <th>Cluster1</th> <th>Cluster2</th> <th>Grand centroid</th> </tr> </thead> <tbody> <tr> <td>A1</td> <td>1.12464</td> <td>-0.449855</td> <td>0.0000000</td> </tr> <tr> <td>A2</td> <td>1.31300</td> <td>-0.525201</td> <td>0.0000000</td> </tr> <tr> <td>A3</td> <td>1.12514</td> <td>-0.450055</td> <td>0.0000000</td> </tr> <tr> <td>A4</td> <td>1.12514</td> <td>-0.450055</td> <td>0.0000000</td> </tr> <tr> <td>A5</td> <td>1.24554</td> <td>-0.498215</td> <td>-0.0000000</td> </tr> </tbody> </table> <p>Distances Between Cluster Centroids</p> <table border="1"> <thead> <tr> <th>Cluster1</th> <th>Cluster1</th> <th>Cluster2</th> </tr> </thead> <tbody> <tr> <td>Cluster1</td> <td>0.00000</td> <td>3.79225</td> </tr> <tr> <td>Cluster2</td> <td>3.79225</td> <td>0.00000</td> </tr> </tbody> </table>	Clrt1	Clrt2	0.56743	3.78722	0.56743	3.88114	2.23157	1.86719	3.41007	1.11266	5.40414	1.63633	4.39491	0.78254	3.94828	0.69284	Variable	Cluster1	Cluster2	Grand centroid	A1	1.12464	-0.449855	0.0000000	A2	1.31300	-0.525201	0.0000000	A3	1.12514	-0.450055	0.0000000	A4	1.12514	-0.450055	0.0000000	A5	1.24554	-0.498215	-0.0000000	Cluster1	Cluster1	Cluster2	Cluster1	0.00000	3.79225	Cluster2	3.79225	0.00000	<p>Cluster Observation values :</p> <table border="1"> <thead> <tr> <th>Cluster01</th> <th>Cluster02</th> </tr> </thead> <tbody> <tr> <td>0.53919</td> <td>3.27079</td> </tr> <tr> <td>0.53919</td> <td>3.55573</td> </tr> <tr> <td>2.04006</td> <td>1.66325</td> </tr> <tr> <td>2.80131</td> <td>1.09225</td> </tr> <tr> <td>4.89116</td> <td>1.53544</td> </tr> <tr> <td>3.74619</td> <td>0.54072</td> </tr> <tr> <td>3.74936</td> <td>0.74307</td> </tr> </tbody> </table> <p>Cluster Centroids</p> <table border="1"> <thead> <tr> <th>Variable</th> <th>Cluster1</th> <th>Cluster2</th> <th>Grand centroid</th> </tr> </thead> <tbody> <tr> <td>A1</td> <td>1.12464</td> <td>-0.449855</td> <td>0.0000000</td> </tr> <tr> <td>A2</td> <td>1.31300</td> <td>-0.525201</td> <td>0.0000000</td> </tr> <tr> <td>A3</td> <td>1.12514</td> <td>-0.450055</td> <td>0.0000000</td> </tr> <tr> <td>A4</td> <td>1.12514</td> <td>-0.450055</td> <td>0.0000000</td> </tr> <tr> <td>A5</td> <td>1.24554</td> <td>-0.498215</td> <td>-0.0000000</td> </tr> </tbody> </table> <p>Distances Between Cluster Centroids</p> <table border="1"> <thead> <tr> <th>Cluster1</th> <th>Cluster1</th> <th>Cluster2</th> </tr> </thead> <tbody> <tr> <td>Cluster1</td> <td>0.00000</td> <td>3.37341</td> </tr> <tr> <td>Cluster2</td> <td>3.37341</td> <td>0.00000</td> </tr> </tbody> </table>	Cluster01	Cluster02	0.53919	3.27079	0.53919	3.55573	2.04006	1.66325	2.80131	1.09225	4.89116	1.53544	3.74619	0.54072	3.74936	0.74307	Variable	Cluster1	Cluster2	Grand centroid	A1	1.12464	-0.449855	0.0000000	A2	1.31300	-0.525201	0.0000000	A3	1.12514	-0.450055	0.0000000	A4	1.12514	-0.450055	0.0000000	A5	1.24554	-0.498215	-0.0000000	Cluster1	Cluster1	Cluster2	Cluster1	0.00000	3.37341	Cluster2	3.37341	0.00000	<p>Dendrogram</p> <p>Dendrogram</p>																																																							
Clrt1	Clrt2																																																																																																																																																										
0.56743	3.78722																																																																																																																																																										
0.56743	3.88114																																																																																																																																																										
2.23157	1.86719																																																																																																																																																										
3.41007	1.11266																																																																																																																																																										
5.40414	1.63633																																																																																																																																																										
4.39491	0.78254																																																																																																																																																										
3.94828	0.69284																																																																																																																																																										
Variable	Cluster1	Cluster2	Grand centroid																																																																																																																																																								
A1	1.12464	-0.449855	0.0000000																																																																																																																																																								
A2	1.31300	-0.525201	0.0000000																																																																																																																																																								
A3	1.12514	-0.450055	0.0000000																																																																																																																																																								
A4	1.12514	-0.450055	0.0000000																																																																																																																																																								
A5	1.24554	-0.498215	-0.0000000																																																																																																																																																								
Cluster1	Cluster1	Cluster2																																																																																																																																																									
Cluster1	0.00000	3.79225																																																																																																																																																									
Cluster2	3.79225	0.00000																																																																																																																																																									
Cluster01	Cluster02																																																																																																																																																										
0.53919	3.27079																																																																																																																																																										
0.53919	3.55573																																																																																																																																																										
2.04006	1.66325																																																																																																																																																										
2.80131	1.09225																																																																																																																																																										
4.89116	1.53544																																																																																																																																																										
3.74619	0.54072																																																																																																																																																										
3.74936	0.74307																																																																																																																																																										
Variable	Cluster1	Cluster2	Grand centroid																																																																																																																																																								
A1	1.12464	-0.449855	0.0000000																																																																																																																																																								
A2	1.31300	-0.525201	0.0000000																																																																																																																																																								
A3	1.12514	-0.450055	0.0000000																																																																																																																																																								
A4	1.12514	-0.450055	0.0000000																																																																																																																																																								
A5	1.24554	-0.498215	-0.0000000																																																																																																																																																								
Cluster1	Cluster1	Cluster2																																																																																																																																																									
Cluster1	0.00000	3.37341																																																																																																																																																									
Cluster2	3.37341	0.00000																																																																																																																																																									
<ul style="list-style-type: none"> Standardization brought the values to similar scale But no difference in correlation identified. 	<ul style="list-style-type: none"> Standardized values reduced the size of the cluster , evident from smaller WSS values. Clusters moved nearer than using the actual values. The pairs at bottom levels are still clumsy 	<ul style="list-style-type: none"> Dropping correlated variables on standardized rearranged the pairs at the bottom. 6 & 7 is the new bottom pair. Clusters moved very little compared to A4, evident from Distance between cluster centroids. 																																																																																																																																																									

A11: Cluster Analysis on Observations for variables XC2 (dropped X5) (Actual Values)

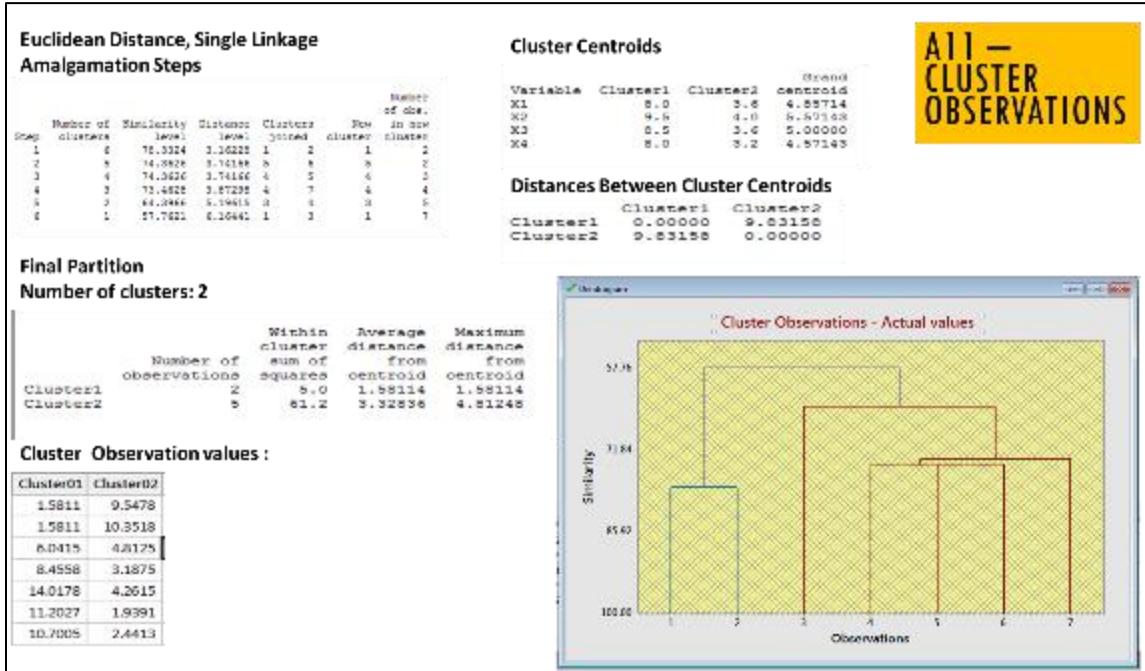


FIGURE 38 : CLUSTER ANALYSIS OF OBSERVATIONS FOR ACT. Xc2

TABLE 7 : COMPARISON OF A3 & A11

Applying Euclidean Distance measure with	With 5 variables (A3)	With Xc2 (X1,X2, X3, X4) variables (A11)																																																																																																																																
Single Linkage	Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[4,c[5,6]]]] Highest similarity level: 79.5 Distance btwn. Cluster Centroids: 11.1692 WSS1: 5.5 WSS2: 76.0 Avg. Distance from Centroid C1 : 1.6 Avg. Distance from Centroid C2 : 3.6	Cluster 1 : c[1,2] Cluster 2 : c[3,c[7,c[4,c[5,6]]]] Highest similarity level: 78.33 Distance btwn. Cluster Centroids: 9.8 WSS1: 5 WSS2: 61.2 Avg. Distance from Centroid C1 : 1.5 Avg. Distance from Centroid C2 : 3.3																																																																																																																																
	Euclidean Distance, Single Linkage Amalgamation Steps <table border="1"> <thead> <tr> <th>Step</th> <th>Number of clusters</th> <th>Similarity level</th> <th>Distance</th> <th>Clusters joined</th> <th>New cluster</th> <th>Number of obs.</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>6</td> <td>79.5098</td> <td>3.16462</td> <td>1, 2</td> <td>1</td> <td>2</td> </tr> <tr> <td>2</td> <td>5</td> <td>76.0726</td> <td>3.87298</td> <td>5, 6</td> <td>5</td> <td>2</td> </tr> <tr> <td>3</td> <td>4</td> <td>76.0726</td> <td>3.87298</td> <td>4, 5</td> <td>4</td> <td>3</td> </tr> <tr> <td>4</td> <td>3</td> <td>73.4828</td> <td>3.87298</td> <td>4, 7</td> <td>4</td> <td>4</td> </tr> <tr> <td>5</td> <td>2</td> <td>64.3966</td> <td>5.19615</td> <td>3</td> <td>3</td> <td>5</td> </tr> <tr> <td>6</td> <td>1</td> <td>59.0196</td> <td>6.16325</td> <td>1, 3</td> <td>1</td> <td>7</td> </tr> </tbody> </table> Final Partition Number of clusters: 2 <table border="1"> <thead> <tr> <th>Cluster</th> <th>Observation</th> <th>Within cluster distance</th> <th>Average distance from centroid</th> <th>Maximum distance from centroid</th> </tr> </thead> <tbody> <tr> <td>Cluster1</td> <td>1</td> <td>2</td> <td>5.5</td> <td>1.65831</td> </tr> <tr> <td>Cluster2</td> <td>5</td> <td>76.0</td> <td>3.68135</td> <td>5.56776</td> </tr> </tbody> </table>	Step	Number of clusters	Similarity level	Distance	Clusters joined	New cluster	Number of obs.	1	6	79.5098	3.16462	1, 2	1	2	2	5	76.0726	3.87298	5, 6	5	2	3	4	76.0726	3.87298	4, 5	4	3	4	3	73.4828	3.87298	4, 7	4	4	5	2	64.3966	5.19615	3	3	5	6	1	59.0196	6.16325	1, 3	1	7	Cluster	Observation	Within cluster distance	Average distance from centroid	Maximum distance from centroid	Cluster1	1	2	5.5	1.65831	Cluster2	5	76.0	3.68135	5.56776	Euclidean Distance, Single Linkage Amalgamation Steps <table border="1"> <thead> <tr> <th>Step</th> <th>Number of clusters</th> <th>Similarity level</th> <th>Distance</th> <th>Clusters joined</th> <th>New cluster</th> <th>Number of obs.</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>6</td> <td>78.3324</td> <td>3.16228</td> <td>1, 2</td> <td>1</td> <td>2</td> </tr> <tr> <td>2</td> <td>5</td> <td>74.3620</td> <td>3.74164</td> <td>3, 4</td> <td>5</td> <td>6</td> </tr> <tr> <td>3</td> <td>4</td> <td>74.8826</td> <td>3.87298</td> <td>4, 5</td> <td>4</td> <td>3</td> </tr> <tr> <td>4</td> <td>3</td> <td>73.4828</td> <td>3.87298</td> <td>4, 7</td> <td>4</td> <td>4</td> </tr> <tr> <td>5</td> <td>2</td> <td>64.3966</td> <td>5.19615</td> <td>3</td> <td>3</td> <td>5</td> </tr> <tr> <td>6</td> <td>1</td> <td>57.7621</td> <td>6.16441</td> <td>1</td> <td>1</td> <td>7</td> </tr> </tbody> </table> Final Partition Number of clusters: 2 <table border="1"> <thead> <tr> <th>Cluster</th> <th>Observation</th> <th>Within cluster distance</th> <th>Average distance from centroid</th> <th>Maximum distance from centroid</th> </tr> </thead> <tbody> <tr> <td>Cluster1</td> <td>2</td> <td>5.0</td> <td>1.58114</td> <td>1.58114</td> </tr> <tr> <td>Cluster2</td> <td>5</td> <td>61.2</td> <td>3.32836</td> <td>4.81248</td> </tr> </tbody> </table>	Step	Number of clusters	Similarity level	Distance	Clusters joined	New cluster	Number of obs.	1	6	78.3324	3.16228	1, 2	1	2	2	5	74.3620	3.74164	3, 4	5	6	3	4	74.8826	3.87298	4, 5	4	3	4	3	73.4828	3.87298	4, 7	4	4	5	2	64.3966	5.19615	3	3	5	6	1	57.7621	6.16441	1	1	7	Cluster	Observation	Within cluster distance	Average distance from centroid	Maximum distance from centroid	Cluster1	2	5.0	1.58114	1.58114	Cluster2	5	61.2	3.32836	4.81248
Step	Number of clusters	Similarity level	Distance	Clusters joined	New cluster	Number of obs.																																																																																																																												
1	6	79.5098	3.16462	1, 2	1	2																																																																																																																												
2	5	76.0726	3.87298	5, 6	5	2																																																																																																																												
3	4	76.0726	3.87298	4, 5	4	3																																																																																																																												
4	3	73.4828	3.87298	4, 7	4	4																																																																																																																												
5	2	64.3966	5.19615	3	3	5																																																																																																																												
6	1	59.0196	6.16325	1, 3	1	7																																																																																																																												
Cluster	Observation	Within cluster distance	Average distance from centroid	Maximum distance from centroid																																																																																																																														
Cluster1	1	2	5.5	1.65831																																																																																																																														
Cluster2	5	76.0	3.68135	5.56776																																																																																																																														
Step	Number of clusters	Similarity level	Distance	Clusters joined	New cluster	Number of obs.																																																																																																																												
1	6	78.3324	3.16228	1, 2	1	2																																																																																																																												
2	5	74.3620	3.74164	3, 4	5	6																																																																																																																												
3	4	74.8826	3.87298	4, 5	4	3																																																																																																																												
4	3	73.4828	3.87298	4, 7	4	4																																																																																																																												
5	2	64.3966	5.19615	3	3	5																																																																																																																												
6	1	57.7621	6.16441	1	1	7																																																																																																																												
Cluster	Observation	Within cluster distance	Average distance from centroid	Maximum distance from centroid																																																																																																																														
Cluster1	2	5.0	1.58114	1.58114																																																																																																																														
Cluster2	5	61.2	3.32836	4.81248																																																																																																																														

	<p>Cluster Observation values :</p> <table border="1"> <thead> <tr> <th>Cluster1</th> <th>Cluster2</th> </tr> </thead> <tbody> <tr><td>1.6583</td><td>11.1714</td></tr> <tr><td>1.6583</td><td>11.4105</td></tr> <tr><td>6.5383</td><td>5.5678</td></tr> <tr><td>10.0871</td><td>3.1937</td></tr> <tr><td>15.8981</td><td>4.7958</td></tr> <tr><td>12.9518</td><td>2.2804</td></tr> <tr><td>11.6082</td><td>2.5690</td></tr> </tbody> </table> <p>Cluster Centroids</p> <table border="1"> <thead> <tr> <th>Variable</th> <th>Cluster1</th> <th>Cluster2</th> <th>Grand</th> </tr> </thead> <tbody> <tr><td>X1</td><td>8.0</td><td>3.6</td><td>4.85714</td></tr> <tr><td>X2</td><td>8.5</td><td>4.0</td><td>5.57143</td></tr> <tr><td>X3</td><td>8.5</td><td>3.6</td><td>5.00000</td></tr> <tr><td>X4</td><td>8.0</td><td>3.2</td><td>4.57143</td></tr> <tr><td>X5</td><td>9.0</td><td>4.2</td><td>5.71429</td></tr> </tbody> </table> <p>Distances Between Cluster Centroids</p> <table border="1"> <thead> <tr> <th></th> <th>Cluster1</th> <th>Cluster2</th> </tr> </thead> <tbody> <tr><td>Cluster1</td><td>0.0000</td><td>11.1692</td></tr> <tr><td>Cluster2</td><td>11.1692</td><td>0.0000</td></tr> </tbody> </table>	Cluster1	Cluster2	1.6583	11.1714	1.6583	11.4105	6.5383	5.5678	10.0871	3.1937	15.8981	4.7958	12.9518	2.2804	11.6082	2.5690	Variable	Cluster1	Cluster2	Grand	X1	8.0	3.6	4.85714	X2	8.5	4.0	5.57143	X3	8.5	3.6	5.00000	X4	8.0	3.2	4.57143	X5	9.0	4.2	5.71429		Cluster1	Cluster2	Cluster1	0.0000	11.1692	Cluster2	11.1692	0.0000	<p>Cluster Observation values :</p> <table border="1"> <thead> <tr> <th>Cluster01</th> <th>Cluster02</th> </tr> </thead> <tbody> <tr><td>1.5811</td><td>9.5478</td></tr> <tr><td>1.5811</td><td>10.3518</td></tr> <tr><td>6.0415</td><td>4.8125</td></tr> <tr><td>8.4558</td><td>3.1875</td></tr> <tr><td>14.0178</td><td>4.2615</td></tr> <tr><td>11.2027</td><td>1.9391</td></tr> <tr><td>10.7005</td><td>2.4413</td></tr> </tbody> </table> <p>Cluster Centroids</p> <table border="1"> <thead> <tr> <th>Variable</th> <th>Cluster01</th> <th>Cluster02</th> <th>Grand</th> </tr> </thead> <tbody> <tr><td>X1</td><td>8.0</td><td>3.6</td><td>4.85714</td></tr> <tr><td>X2</td><td>9.5</td><td>4.0</td><td>5.57143</td></tr> <tr><td>X3</td><td>8.5</td><td>3.2</td><td>5.00000</td></tr> <tr><td>X4</td><td>8.0</td><td>3.2</td><td>4.57143</td></tr> </tbody> </table> <p>Distances Between Cluster Centroids</p> <table border="1"> <thead> <tr> <th></th> <th>Cluster01</th> <th>Cluster02</th> </tr> </thead> <tbody> <tr><td>Cluster01</td><td>0.00000</td><td>9.83158</td></tr> <tr><td>Cluster02</td><td>9.83158</td><td>0.00000</td></tr> </tbody> </table>	Cluster01	Cluster02	1.5811	9.5478	1.5811	10.3518	6.0415	4.8125	8.4558	3.1875	14.0178	4.2615	11.2027	1.9391	10.7005	2.4413	Variable	Cluster01	Cluster02	Grand	X1	8.0	3.6	4.85714	X2	9.5	4.0	5.57143	X3	8.5	3.2	5.00000	X4	8.0	3.2	4.57143		Cluster01	Cluster02	Cluster01	0.00000	9.83158	Cluster02	9.83158	0.00000
Cluster1	Cluster2																																																																																															
1.6583	11.1714																																																																																															
1.6583	11.4105																																																																																															
6.5383	5.5678																																																																																															
10.0871	3.1937																																																																																															
15.8981	4.7958																																																																																															
12.9518	2.2804																																																																																															
11.6082	2.5690																																																																																															
Variable	Cluster1	Cluster2	Grand																																																																																													
X1	8.0	3.6	4.85714																																																																																													
X2	8.5	4.0	5.57143																																																																																													
X3	8.5	3.6	5.00000																																																																																													
X4	8.0	3.2	4.57143																																																																																													
X5	9.0	4.2	5.71429																																																																																													
	Cluster1	Cluster2																																																																																														
Cluster1	0.0000	11.1692																																																																																														
Cluster2	11.1692	0.0000																																																																																														
Cluster01	Cluster02																																																																																															
1.5811	9.5478																																																																																															
1.5811	10.3518																																																																																															
6.0415	4.8125																																																																																															
8.4558	3.1875																																																																																															
14.0178	4.2615																																																																																															
11.2027	1.9391																																																																																															
10.7005	2.4413																																																																																															
Variable	Cluster01	Cluster02	Grand																																																																																													
X1	8.0	3.6	4.85714																																																																																													
X2	9.5	4.0	5.57143																																																																																													
X3	8.5	3.2	5.00000																																																																																													
X4	8.0	3.2	4.57143																																																																																													
	Cluster01	Cluster02																																																																																														
Cluster01	0.00000	9.83158																																																																																														
Cluster02	9.83158	0.00000																																																																																														
Findings	<ul style="list-style-type: none"> Clusters are unevenly sized , positioned farther 	<ul style="list-style-type: none"> Having XC2 (X1,X2,X3,X4) did bring the clusters nearer,evident from distance between cluster centroids 																																																																																														

A12: Cluster Analysis on Observations for variables XC2 (standardized values)

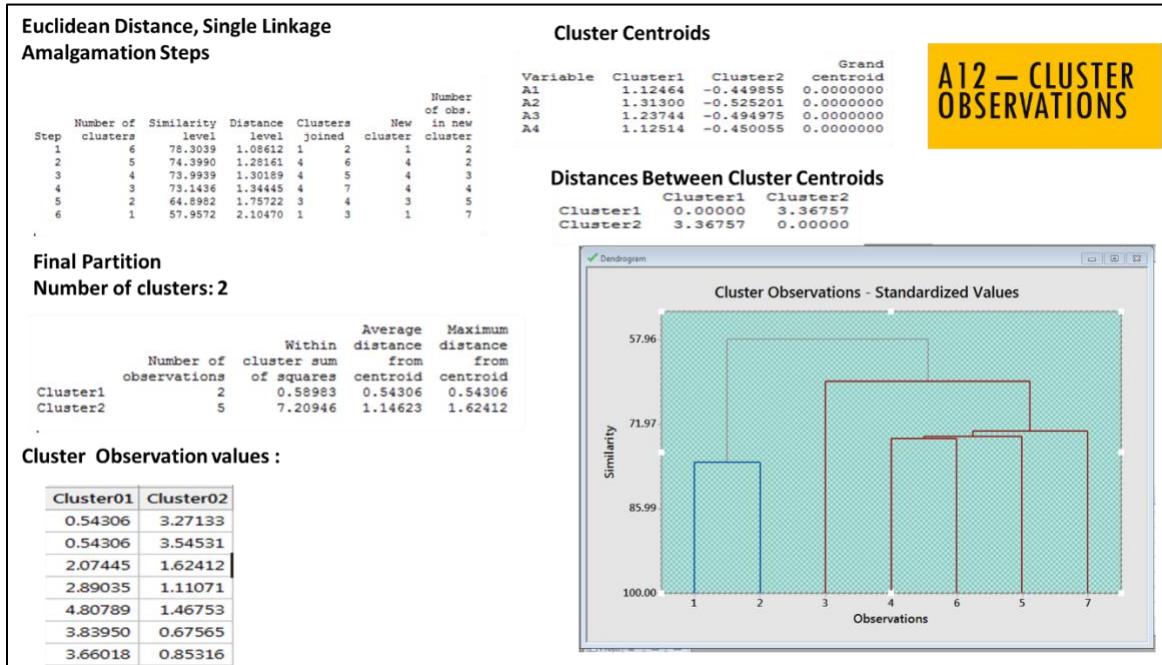


FIGURE 39 : CLUSTER ANALYSIS OF OBSERVATIONS FOR STD.XC2

TABLE 8 : COMPARISON OF A2, A4 & A12

Analysis on Observations using all 5 variables on std.values					Analysis on Observations using all 5 variables on std.values (A4)	Analysis on Observations using XC2 (4 variables) on std.values (A12)																																													
<table border="1"> <thead> <tr> <th>C6</th><th>C7</th><th>C8</th><th>C9</th><th>C10</th></tr> <tr> <th>A1</th><th>A2</th><th>A3</th><th>A4</th><th>A5</th></tr> </thead> <tbody> <tr> <td>0.76680</td><td>1.48011</td><td>1.41421</td><td>0.79697</td><td>1.41004</td></tr> <tr> <td>1.48248</td><td>1.14589</td><td>1.06066</td><td>1.45330</td><td>1.08103</td></tr> <tr> <td>0.05112</td><td>-0.19098</td><td>0.35355</td><td>0.79697</td><td>0.42301</td></tr> <tr> <td>0.40896</td><td>0.14324</td><td>-0.70711</td><td>-0.51569</td><td>-0.56402</td></tr> <tr> <td>-1.38024</td><td>-1.19364</td><td>-1.06066</td><td>-1.17202</td><td>-1.22204</td></tr> <tr> <td>-0.30672</td><td>-0.85942</td><td>-1.06066</td><td>-0.51569</td><td>-0.89303</td></tr> <tr> <td>-1.02240</td><td>-0.52520</td><td>0.00000</td><td>-0.84385</td><td>-0.23501</td></tr> </tbody> </table>					C6	C7	C8	C9	C10	A1	A2	A3	A4	A5	0.76680	1.48011	1.41421	0.79697	1.41004	1.48248	1.14589	1.06066	1.45330	1.08103	0.05112	-0.19098	0.35355	0.79697	0.42301	0.40896	0.14324	-0.70711	-0.51569	-0.56402	-1.38024	-1.19364	-1.06066	-1.17202	-1.22204	-0.30672	-0.85942	-1.06066	-0.51569	-0.89303	-1.02240	-0.52520	0.00000	-0.84385	-0.23501	<p>Cluster 1 : c[1,2]</p> <p>Cluster 2 : c[3,c[7,c[5,c[4,6]]]]</p> <p>Highest similarity level: 79.4</p> <p>Distance btwn. Cluster Centroids: 3.792</p> <p>WSS1: 0.643</p> <p>WSS2: 8.811</p> <p>Avg. Distance from Centroid C1 : 0.56</p> <p>Avg. Distance from Centroid C2 : 1.25</p>	<p>Cluster 1 : c[1,2]</p> <p>Cluster 2 : c[3,c[7,c[5,c[4,6]]]]</p> <p>Highest similarity level: 78.3</p> <p>Distance btwn. Cluster Centroids: 3.36</p> <p>WSS1: 0.5</p> <p>WSS2: 7.2</p> <p>Avg. Distance from Centroid C1 : 0.5</p> <p>Avg. Distance from Centroid C2 : 1.1</p>
C6	C7	C8	C9	C10																																															
A1	A2	A3	A4	A5																																															
0.76680	1.48011	1.41421	0.79697	1.41004																																															
1.48248	1.14589	1.06066	1.45330	1.08103																																															
0.05112	-0.19098	0.35355	0.79697	0.42301																																															
0.40896	0.14324	-0.70711	-0.51569	-0.56402																																															
-1.38024	-1.19364	-1.06066	-1.17202	-1.22204																																															
-0.30672	-0.85942	-1.06066	-0.51569	-0.89303																																															
-1.02240	-0.52520	0.00000	-0.84385	-0.23501																																															

<p>Correlation: A1, A2, A3, A4, A5</p> <table border="1"> <thead> <tr> <th></th> <th>A1</th> <th>A2</th> <th>A3</th> <th>A4</th> </tr> </thead> <tbody> <tr> <td>A2</td> <td>0.888 0.008</td> <td></td> <td></td> <td></td> </tr> <tr> <td>A3</td> <td>0.696 0.083</td> <td>0.886 0.008</td> <td></td> <td></td> </tr> <tr> <td>A4</td> <td>0.872 0.010</td> <td>0.817 0.029</td> <td>0.851 0.015</td> <td></td> </tr> <tr> <td>A5</td> <td>0.779 0.039</td> <td>0.919 0.003</td> <td>0.989 0.000</td> <td>0.902 0.005</td> </tr> </tbody> </table> <p>Cell Contents: Pearson correlation P-Value</p>		A1	A2	A3	A4	A2	0.888 0.008				A3	0.696 0.083	0.886 0.008			A4	0.872 0.010	0.817 0.029	0.851 0.015		A5	0.779 0.039	0.919 0.003	0.989 0.000	0.902 0.005	<p>Euclidean Distance, Single Linkage Amalgamation Steps</p> <table border="1"> <thead> <tr> <th>Step</th> <th>Number of clusters</th> <th>Similarity level</th> <th>Distance level</th> <th>Clusters joined</th> <th>New cluster</th> <th>Number of obs. in new cluster</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>6</td> <td>79.9882</td> <td>1.32216</td> <td>1</td> <td>2</td> <td>2</td> </tr> <tr> <td>2</td> <td>5</td> <td>75.9981</td> <td>1.34282</td> <td>4</td> <td>5</td> <td>2</td> </tr> <tr> <td>3</td> <td>4</td> <td>75.4314</td> <td>1.34282</td> <td>4</td> <td>5</td> <td>3</td> </tr> <tr> <td>4</td> <td>3</td> <td>72.5406</td> <td>1.81914</td> <td>4</td> <td>7</td> <td>4</td> </tr> <tr> <td>5</td> <td>2</td> <td>63.4242</td> <td>1.81914</td> <td>3</td> <td>4</td> <td>5</td> </tr> <tr> <td>6</td> <td>1</td> <td>56.6188</td> <td>2.28030</td> <td>1</td> <td>3</td> <td>7</td> </tr> </tbody> </table> <p>Final Partition Number of clusters: 2</p> <table border="1"> <thead> <tr> <th>Cluster</th> <th>Number of observations</th> <th>Average distance from centroid</th> <th>Maximum distance from centroid</th> </tr> </thead> <tbody> <tr> <td>Cluster1</td> <td>2</td> <td>0.64396</td> <td>0.56743</td> </tr> <tr> <td>Cluster2</td> <td>5</td> <td>8.81152</td> <td>1.25831</td> </tr> </tbody> </table>	Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster	1	6	79.9882	1.32216	1	2	2	2	5	75.9981	1.34282	4	5	2	3	4	75.4314	1.34282	4	5	3	4	3	72.5406	1.81914	4	7	4	5	2	63.4242	1.81914	3	4	5	6	1	56.6188	2.28030	1	3	7	Cluster	Number of observations	Average distance from centroid	Maximum distance from centroid	Cluster1	2	0.64396	0.56743	Cluster2	5	8.81152	1.25831	<p>Euclidean Distance, Single Linkage Amalgamation Steps</p> <table border="1"> <thead> <tr> <th>Step</th> <th>Number of clusters</th> <th>Similarity level</th> <th>Distance level</th> <th>Clusters joined</th> <th>New cluster</th> <th>Number of obs. in new cluster</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>6</td> <td>78.3039</td> <td>1.08612</td> <td>1</td> <td>2</td> <td>2</td> </tr> <tr> <td>2</td> <td>5</td> <td>74.3990</td> <td>1.20141</td> <td>4</td> <td>5</td> <td>2</td> </tr> <tr> <td>3</td> <td>4</td> <td>73.9981</td> <td>1.20141</td> <td>4</td> <td>5</td> <td>3</td> </tr> <tr> <td>4</td> <td>3</td> <td>73.1456</td> <td>1.34445</td> <td>4</td> <td>7</td> <td>4</td> </tr> <tr> <td>5</td> <td>2</td> <td>64.9986</td> <td>1.79722</td> <td>3</td> <td>4</td> <td>5</td> </tr> <tr> <td>6</td> <td>1</td> <td>57.0772</td> <td>2.07770</td> <td>1</td> <td>3</td> <td>7</td> </tr> </tbody> </table> <p>Final Partition Number of clusters: 2</p> <table border="1"> <thead> <tr> <th>Cluster</th> <th>Number of observations</th> <th>Average distance from centroid</th> <th>Maximum distance from centroid</th> </tr> </thead> <tbody> <tr> <td>Cluster1</td> <td>2</td> <td>0.58983</td> <td>0.54306</td> </tr> <tr> <td>Cluster2</td> <td>5</td> <td>7.20946</td> <td>1.14623</td> </tr> </tbody> </table>	Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster	1	6	78.3039	1.08612	1	2	2	2	5	74.3990	1.20141	4	5	2	3	4	73.9981	1.20141	4	5	3	4	3	73.1456	1.34445	4	7	4	5	2	64.9986	1.79722	3	4	5	6	1	57.0772	2.07770	1	3	7	Cluster	Number of observations	Average distance from centroid	Maximum distance from centroid	Cluster1	2	0.58983	0.54306	Cluster2	5	7.20946	1.14623
	A1	A2	A3	A4																																																																																																																																																	
A2	0.888 0.008																																																																																																																																																				
A3	0.696 0.083	0.886 0.008																																																																																																																																																			
A4	0.872 0.010	0.817 0.029	0.851 0.015																																																																																																																																																		
A5	0.779 0.039	0.919 0.003	0.989 0.000	0.902 0.005																																																																																																																																																	
Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster																																																																																																																																															
1	6	79.9882	1.32216	1	2	2																																																																																																																																															
2	5	75.9981	1.34282	4	5	2																																																																																																																																															
3	4	75.4314	1.34282	4	5	3																																																																																																																																															
4	3	72.5406	1.81914	4	7	4																																																																																																																																															
5	2	63.4242	1.81914	3	4	5																																																																																																																																															
6	1	56.6188	2.28030	1	3	7																																																																																																																																															
Cluster	Number of observations	Average distance from centroid	Maximum distance from centroid																																																																																																																																																		
Cluster1	2	0.64396	0.56743																																																																																																																																																		
Cluster2	5	8.81152	1.25831																																																																																																																																																		
Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster																																																																																																																																															
1	6	78.3039	1.08612	1	2	2																																																																																																																																															
2	5	74.3990	1.20141	4	5	2																																																																																																																																															
3	4	73.9981	1.20141	4	5	3																																																																																																																																															
4	3	73.1456	1.34445	4	7	4																																																																																																																																															
5	2	64.9986	1.79722	3	4	5																																																																																																																																															
6	1	57.0772	2.07770	1	3	7																																																																																																																																															
Cluster	Number of observations	Average distance from centroid	Maximum distance from centroid																																																																																																																																																		
Cluster1	2	0.58983	0.54306																																																																																																																																																		
Cluster2	5	7.20946	1.14623																																																																																																																																																		
<p>Cluster Observation values :</p> <table border="1"> <thead> <tr> <th>Clrt1</th> <th>Clrt2</th> </tr> </thead> <tbody> <tr> <td>0.56743</td> <td>3.78722</td> </tr> <tr> <td>0.56743</td> <td>3.88114</td> </tr> <tr> <td>2.23157</td> <td>1.86719</td> </tr> <tr> <td>3.41007</td> <td>1.11266</td> </tr> <tr> <td>5.40414</td> <td>1.63633</td> </tr> <tr> <td>4.39491</td> <td>0.78254</td> </tr> <tr> <td>3.94828</td> <td>0.69284</td> </tr> </tbody> </table> <p>Cluster Centroids</p> <table border="1"> <thead> <tr> <th>Variable</th> <th>Cluster1</th> <th>Cluster2</th> <th>Grand centroid</th> </tr> </thead> <tbody> <tr> <td>A1</td> <td>1.12464</td> <td>-0.449855</td> <td>0.000000</td> </tr> <tr> <td>A2</td> <td>1.31300</td> <td>-0.525201</td> <td>0.000000</td> </tr> <tr> <td>A3</td> <td>1.12514</td> <td>-0.450078</td> <td>0.000000</td> </tr> <tr> <td>A4</td> <td>1.12514</td> <td>-0.450055</td> <td>0.000000</td> </tr> <tr> <td>A5</td> <td>1.24554</td> <td>-0.498215</td> <td>-0.000000</td> </tr> </tbody> </table> <p>Distances Between Cluster Centroids</p> <table border="1"> <thead> <tr> <th>Cluster1</th> <th>Cluster1</th> <th>Cluster2</th> </tr> </thead> <tbody> <tr> <td>Cluster1</td> <td>0.00000</td> <td>3.79225</td> </tr> <tr> <td>Cluster2</td> <td>3.79225</td> <td>0.00000</td> </tr> </tbody> </table>	Clrt1	Clrt2	0.56743	3.78722	0.56743	3.88114	2.23157	1.86719	3.41007	1.11266	5.40414	1.63633	4.39491	0.78254	3.94828	0.69284	Variable	Cluster1	Cluster2	Grand centroid	A1	1.12464	-0.449855	0.000000	A2	1.31300	-0.525201	0.000000	A3	1.12514	-0.450078	0.000000	A4	1.12514	-0.450055	0.000000	A5	1.24554	-0.498215	-0.000000	Cluster1	Cluster1	Cluster2	Cluster1	0.00000	3.79225	Cluster2	3.79225	0.00000	<p>Cluster Observation values :</p> <table border="1"> <thead> <tr> <th>Clrt1</th> <th>Clrt2</th> </tr> </thead> <tbody> <tr> <td>0.56743</td> <td>3.78722</td> </tr> <tr> <td>0.56743</td> <td>3.88114</td> </tr> <tr> <td>2.23157</td> <td>1.86719</td> </tr> <tr> <td>3.41007</td> <td>1.11266</td> </tr> <tr> <td>5.40414</td> <td>1.63633</td> </tr> <tr> <td>4.39491</td> <td>0.78254</td> </tr> <tr> <td>3.94828</td> <td>0.69284</td> </tr> </tbody> </table> <p>Cluster Centroids</p> <table border="1"> <thead> <tr> <th>Variable</th> <th>Cluster1</th> <th>Cluster2</th> <th>Grand centroid</th> </tr> </thead> <tbody> <tr> <td>A1</td> <td>1.12464</td> <td>-0.449855</td> <td>0.000000</td> </tr> <tr> <td>A2</td> <td>1.31300</td> <td>-0.525201</td> <td>0.000000</td> </tr> <tr> <td>A3</td> <td>1.12514</td> <td>-0.450078</td> <td>0.000000</td> </tr> <tr> <td>A4</td> <td>1.12514</td> <td>-0.450055</td> <td>0.000000</td> </tr> </tbody> </table> <p>Distances Between Cluster Centroids</p> <table border="1"> <thead> <tr> <th>Cluster1</th> <th>Cluster1</th> <th>Cluster2</th> </tr> </thead> <tbody> <tr> <td>Cluster1</td> <td>0.00000</td> <td>3.86787</td> </tr> <tr> <td>Cluster2</td> <td>3.86787</td> <td>0.00000</td> </tr> </tbody> </table>	Clrt1	Clrt2	0.56743	3.78722	0.56743	3.88114	2.23157	1.86719	3.41007	1.11266	5.40414	1.63633	4.39491	0.78254	3.94828	0.69284	Variable	Cluster1	Cluster2	Grand centroid	A1	1.12464	-0.449855	0.000000	A2	1.31300	-0.525201	0.000000	A3	1.12514	-0.450078	0.000000	A4	1.12514	-0.450055	0.000000	Cluster1	Cluster1	Cluster2	Cluster1	0.00000	3.86787	Cluster2	3.86787	0.00000	<p>Cluster Observation values :</p> <table border="1"> <thead> <tr> <th>Cluster01</th> <th>Cluster02</th> </tr> </thead> <tbody> <tr> <td>0.54306</td> <td>3.27133</td> </tr> <tr> <td>0.54306</td> <td>3.54531</td> </tr> <tr> <td>2.07445</td> <td>1.62412</td> </tr> <tr> <td>2.89035</td> <td>1.11071</td> </tr> <tr> <td>4.80789</td> <td>1.46753</td> </tr> <tr> <td>3.83950</td> <td>0.67565</td> </tr> <tr> <td>3.66018</td> <td>0.85316</td> </tr> </tbody> </table> <p>Cluster Centroids</p> <table border="1"> <thead> <tr> <th>Variable</th> <th>Cluster1</th> <th>Cluster2</th> <th>Grand centroid</th> </tr> </thead> <tbody> <tr> <td>A1</td> <td>1.12464</td> <td>-0.449855</td> <td>0.000000</td> </tr> <tr> <td>A2</td> <td>1.31300</td> <td>-0.525201</td> <td>0.000000</td> </tr> <tr> <td>A3</td> <td>1.12514</td> <td>-0.450078</td> <td>0.000000</td> </tr> <tr> <td>A4</td> <td>1.12514</td> <td>-0.450055</td> <td>0.000000</td> </tr> </tbody> </table> <p>Distances Between Cluster Centroids</p>	Cluster01	Cluster02	0.54306	3.27133	0.54306	3.54531	2.07445	1.62412	2.89035	1.11071	4.80789	1.46753	3.83950	0.67565	3.66018	0.85316	Variable	Cluster1	Cluster2	Grand centroid	A1	1.12464	-0.449855	0.000000	A2	1.31300	-0.525201	0.000000	A3	1.12514	-0.450078	0.000000	A4	1.12514	-0.450055	0.000000																	
Clrt1	Clrt2																																																																																																																																																				
0.56743	3.78722																																																																																																																																																				
0.56743	3.88114																																																																																																																																																				
2.23157	1.86719																																																																																																																																																				
3.41007	1.11266																																																																																																																																																				
5.40414	1.63633																																																																																																																																																				
4.39491	0.78254																																																																																																																																																				
3.94828	0.69284																																																																																																																																																				
Variable	Cluster1	Cluster2	Grand centroid																																																																																																																																																		
A1	1.12464	-0.449855	0.000000																																																																																																																																																		
A2	1.31300	-0.525201	0.000000																																																																																																																																																		
A3	1.12514	-0.450078	0.000000																																																																																																																																																		
A4	1.12514	-0.450055	0.000000																																																																																																																																																		
A5	1.24554	-0.498215	-0.000000																																																																																																																																																		
Cluster1	Cluster1	Cluster2																																																																																																																																																			
Cluster1	0.00000	3.79225																																																																																																																																																			
Cluster2	3.79225	0.00000																																																																																																																																																			
Clrt1	Clrt2																																																																																																																																																				
0.56743	3.78722																																																																																																																																																				
0.56743	3.88114																																																																																																																																																				
2.23157	1.86719																																																																																																																																																				
3.41007	1.11266																																																																																																																																																				
5.40414	1.63633																																																																																																																																																				
4.39491	0.78254																																																																																																																																																				
3.94828	0.69284																																																																																																																																																				
Variable	Cluster1	Cluster2	Grand centroid																																																																																																																																																		
A1	1.12464	-0.449855	0.000000																																																																																																																																																		
A2	1.31300	-0.525201	0.000000																																																																																																																																																		
A3	1.12514	-0.450078	0.000000																																																																																																																																																		
A4	1.12514	-0.450055	0.000000																																																																																																																																																		
Cluster1	Cluster1	Cluster2																																																																																																																																																			
Cluster1	0.00000	3.86787																																																																																																																																																			
Cluster2	3.86787	0.00000																																																																																																																																																			
Cluster01	Cluster02																																																																																																																																																				
0.54306	3.27133																																																																																																																																																				
0.54306	3.54531																																																																																																																																																				
2.07445	1.62412																																																																																																																																																				
2.89035	1.11071																																																																																																																																																				
4.80789	1.46753																																																																																																																																																				
3.83950	0.67565																																																																																																																																																				
3.66018	0.85316																																																																																																																																																				
Variable	Cluster1	Cluster2	Grand centroid																																																																																																																																																		
A1	1.12464	-0.449855	0.000000																																																																																																																																																		
A2	1.31300	-0.525201	0.000000																																																																																																																																																		
A3	1.12514	-0.450078	0.000000																																																																																																																																																		
A4	1.12514	-0.450055	0.000000																																																																																																																																																		
<ul style="list-style-type: none"> Standardization brought the values to similar scale But no difference in correlation identified. 	<ul style="list-style-type: none"> Standardized values reduced the size of the cluster , evident from smaller WSS values. Clusters moved nearer than using the actual values. The pairs at bottom levels are still clumsy 	<ul style="list-style-type: none"> Cluster size has reduced further,evident from WSS values Clusters moved nearer But the pairs are not clearly divided, may be an effect of non-correlated variables 																																																																																																																																																			

A13: Cluster K-means (actual Values)

TABLE 9 : COMPARISON OF K-MEANS 2, 3 & 4 CLUSTERS

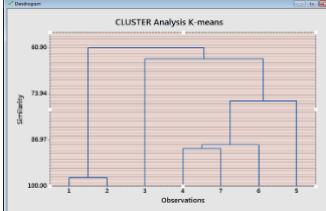
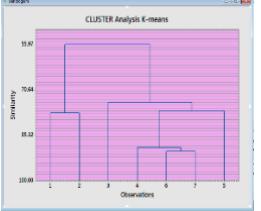
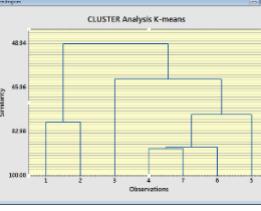
A13 – CLUSTER ANALYSIS K-MEANS WITH 2 CLUSTERS	A13 – CLUSTER ANALYSIS K-MEANS WITH 3 CLUSTERS	A13 – CLUSTER ANALYSIS K-MEANS WITH 4 CLUSTERS																																																																																																																														
Final Partition Number of clusters: 2 <table border="1"> <thead> <tr> <th></th> <th>Within cluster sum of squares</th> <th>Average distance from centroid</th> <th>Maximum distance from centroid</th> </tr> </thead> <tbody> <tr> <td>Number of observations</td> <td>4</td> <td>37.250</td> <td>2.940</td> </tr> <tr> <td>Cluster1</td> <td>3</td> <td>34.000</td> <td>3.270</td> </tr> <tr> <td>Cluster2</td> <td>1</td> <td>3.250</td> <td>4.5714</td> </tr> </tbody> </table>		Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid	Number of observations	4	37.250	2.940	Cluster1	3	34.000	3.270	Cluster2	1	3.250	4.5714	Final Partition Number of clusters: 3 <table border="1"> <thead> <tr> <th></th> <th>Within cluster sum of squares</th> <th>Average distance from centroid</th> <th>Maximum distance from centroid</th> </tr> </thead> <tbody> <tr> <td>Number of observations</td> <td>1</td> <td>0.000</td> <td>0.000</td> </tr> <tr> <td>Cluster1</td> <td>1</td> <td>0.000</td> <td>0.000</td> </tr> <tr> <td>Cluster2</td> <td>5</td> <td>76.000</td> <td>3.651</td> </tr> <tr> <td>Cluster3</td> <td>3</td> <td>5.568</td> <td>4.5714</td> </tr> </tbody> </table>		Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid	Number of observations	1	0.000	0.000	Cluster1	1	0.000	0.000	Cluster2	5	76.000	3.651	Cluster3	3	5.568	4.5714	Final Partition Number of clusters: 4 <table border="1"> <thead> <tr> <th></th> <th>Within cluster sum of squares</th> <th>Average distance from centroid</th> <th>Maximum distance from centroid</th> </tr> </thead> <tbody> <tr> <td>Number of observations</td> <td>1</td> <td>0.000</td> <td>0.000</td> </tr> <tr> <td>Cluster1</td> <td>1</td> <td>0.000</td> <td>0.000</td> </tr> <tr> <td>Cluster2</td> <td>1</td> <td>0.000</td> <td>0.000</td> </tr> <tr> <td>Cluster3</td> <td>4</td> <td>37.250</td> <td>2.948</td> </tr> <tr> <td>Cluster4</td> <td>3</td> <td>3.250</td> <td>4.5666</td> </tr> </tbody> </table>		Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid	Number of observations	1	0.000	0.000	Cluster1	1	0.000	0.000	Cluster2	1	0.000	0.000	Cluster3	4	37.250	2.948	Cluster4	3	3.250	4.5666																																																																		
	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid																																																																																																																													
Number of observations	4	37.250	2.940																																																																																																																													
Cluster1	3	34.000	3.270																																																																																																																													
Cluster2	1	3.250	4.5714																																																																																																																													
	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid																																																																																																																													
Number of observations	1	0.000	0.000																																																																																																																													
Cluster1	1	0.000	0.000																																																																																																																													
Cluster2	5	76.000	3.651																																																																																																																													
Cluster3	3	5.568	4.5714																																																																																																																													
	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid																																																																																																																													
Number of observations	1	0.000	0.000																																																																																																																													
Cluster1	1	0.000	0.000																																																																																																																													
Cluster2	1	0.000	0.000																																																																																																																													
Cluster3	4	37.250	2.948																																																																																																																													
Cluster4	3	3.250	4.5666																																																																																																																													
Cluster Centroids <table border="1"> <thead> <tr> <th>Variable</th> <th>Cluster1</th> <th>Cluster2</th> <th>Grand centroid</th> </tr> </thead> <tbody> <tr> <td>X1</td> <td>3.2500</td> <td>7.0000</td> <td>4.8571</td> </tr> <tr> <td>X2</td> <td>3.7500</td> <td>8.0000</td> <td>5.5714</td> </tr> <tr> <td>X3</td> <td>3.0000</td> <td>7.6667</td> <td>5.5714</td> </tr> <tr> <td>X4</td> <td>2.2500</td> <td>7.6667</td> <td>4.5714</td> </tr> <tr> <td>X5</td> <td>3.5000</td> <td>8.6667</td> <td>5.7143</td> </tr> </tbody> </table>	Variable	Cluster1	Cluster2	Grand centroid	X1	3.2500	7.0000	4.8571	X2	3.7500	8.0000	5.5714	X3	3.0000	7.6667	5.5714	X4	2.2500	7.6667	4.5714	X5	3.5000	8.6667	5.7143	Cluster Centroids <table border="1"> <thead> <tr> <th>Variable</th> <th>Cluster1</th> <th>Cluster2</th> <th>Cluster3</th> <th>Grand centroid</th> </tr> </thead> <tbody> <tr> <td>X1</td> <td>7.0000</td> <td>9.0000</td> <td>3.6000</td> <td>4.8571</td> </tr> <tr> <td>X2</td> <td>10.0000</td> <td>9.0000</td> <td>4.0000</td> <td>5.5714</td> </tr> <tr> <td>X3</td> <td>9.0000</td> <td>5.0000</td> <td>3.2000</td> <td>5.5714</td> </tr> <tr> <td>X4</td> <td>7.0000</td> <td>9.0000</td> <td>3.2000</td> <td>4.5714</td> </tr> <tr> <td>X5</td> <td>10.0000</td> <td>9.0000</td> <td>4.2000</td> <td>5.7143</td> </tr> </tbody> </table>	Variable	Cluster1	Cluster2	Cluster3	Grand centroid	X1	7.0000	9.0000	3.6000	4.8571	X2	10.0000	9.0000	4.0000	5.5714	X3	9.0000	5.0000	3.2000	5.5714	X4	7.0000	9.0000	3.2000	4.5714	X5	10.0000	9.0000	4.2000	5.7143	Cluster Centroids <table border="1"> <thead> <tr> <th>Variable</th> <th>Cluster1</th> <th>Cluster2</th> <th>Cluster3</th> <th>Cluster4</th> <th>Grand centroid</th> </tr> </thead> <tbody> <tr> <td>X1</td> <td>7.0000</td> <td>8.0000</td> <td>3.2500</td> <td>4.8571</td> </tr> <tr> <td>X2</td> <td>10.0000</td> <td>9.0000</td> <td>4.0000</td> <td>5.5714</td> </tr> <tr> <td>X3</td> <td>9.0000</td> <td>8.0000</td> <td>4.0000</td> <td>3.0000</td> <td>5.0000</td> </tr> <tr> <td>X4</td> <td>7.0000</td> <td>9.0000</td> <td>7.0000</td> <td>2.2500</td> <td>4.5714</td> </tr> <tr> <td>X5</td> <td>10.0000</td> <td>9.0000</td> <td>7.0000</td> <td>3.5000</td> <td>5.7143</td> </tr> </tbody> </table>	Variable	Cluster1	Cluster2	Cluster3	Cluster4	Grand centroid	X1	7.0000	8.0000	3.2500	4.8571	X2	10.0000	9.0000	4.0000	5.5714	X3	9.0000	8.0000	4.0000	3.0000	5.0000	X4	7.0000	9.0000	7.0000	2.2500	4.5714	X5	10.0000	9.0000	7.0000	3.5000	5.7143																																						
Variable	Cluster1	Cluster2	Grand centroid																																																																																																																													
X1	3.2500	7.0000	4.8571																																																																																																																													
X2	3.7500	8.0000	5.5714																																																																																																																													
X3	3.0000	7.6667	5.5714																																																																																																																													
X4	2.2500	7.6667	4.5714																																																																																																																													
X5	3.5000	8.6667	5.7143																																																																																																																													
Variable	Cluster1	Cluster2	Cluster3	Grand centroid																																																																																																																												
X1	7.0000	9.0000	3.6000	4.8571																																																																																																																												
X2	10.0000	9.0000	4.0000	5.5714																																																																																																																												
X3	9.0000	5.0000	3.2000	5.5714																																																																																																																												
X4	7.0000	9.0000	3.2000	4.5714																																																																																																																												
X5	10.0000	9.0000	4.2000	5.7143																																																																																																																												
Variable	Cluster1	Cluster2	Cluster3	Cluster4	Grand centroid																																																																																																																											
X1	7.0000	8.0000	3.2500	4.8571																																																																																																																												
X2	10.0000	9.0000	4.0000	5.5714																																																																																																																												
X3	9.0000	8.0000	4.0000	3.0000	5.0000																																																																																																																											
X4	7.0000	9.0000	7.0000	2.2500	4.5714																																																																																																																											
X5	10.0000	9.0000	7.0000	3.5000	5.7143																																																																																																																											
Distances Between Cluster Centroids <table border="1"> <thead> <tr> <th></th> <th>Cluster1</th> <th>Cluster2</th> </tr> </thead> <tbody> <tr> <td>Cluster1</td> <td>0.0000</td> <td>10.4851</td> </tr> <tr> <td>Cluster2</td> <td>10.4851</td> <td>0.0000</td> </tr> </tbody> </table>		Cluster1	Cluster2	Cluster1	0.0000	10.4851	Cluster2	10.4851	0.0000	Distances Between Cluster Centroids <table border="1"> <thead> <tr> <th></th> <th>Cluster1</th> <th>Cluster2</th> <th>Cluster3</th> </tr> </thead> <tbody> <tr> <td>Cluster1</td> <td>0.0000</td> <td>3.1166</td> <td>11.14</td> </tr> <tr> <td>Cluster2</td> <td>3.1166</td> <td>0.0000</td> <td>11.4105</td> </tr> <tr> <td>Cluster3</td> <td>11.1714</td> <td>11.4105</td> <td>0.0000</td> </tr> </tbody> </table>		Cluster1	Cluster2	Cluster3	Cluster1	0.0000	3.1166	11.14	Cluster2	3.1166	0.0000	11.4105	Cluster3	11.1714	11.4105	0.0000	Distances Between Cluster Centroids <table border="1"> <thead> <tr> <th></th> <th>Cluster1</th> <th>Cluster2</th> <th>Cluster3</th> <th>Cluster4</th> </tr> </thead> <tbody> <tr> <td>Cluster1</td> <td>0.0000</td> <td>3.3164</td> <td>6.8557</td> <td>12.4072</td> </tr> <tr> <td>Cluster2</td> <td>3.3164</td> <td>0.0000</td> <td>4.0000</td> <td>12.4072</td> </tr> <tr> <td>Cluster3</td> <td>6.8557</td> <td>4.0000</td> <td>0.0000</td> <td>6.9597</td> </tr> <tr> <td>Cluster4</td> <td>12.4072</td> <td>12.7058</td> <td>6.9597</td> <td>0.0000</td> </tr> </tbody> </table>		Cluster1	Cluster2	Cluster3	Cluster4	Cluster1	0.0000	3.3164	6.8557	12.4072	Cluster2	3.3164	0.0000	4.0000	12.4072	Cluster3	6.8557	4.0000	0.0000	6.9597	Cluster4	12.4072	12.7058	6.9597	0.0000																																																																												
	Cluster1	Cluster2																																																																																																																														
Cluster1	0.0000	10.4851																																																																																																																														
Cluster2	10.4851	0.0000																																																																																																																														
	Cluster1	Cluster2	Cluster3																																																																																																																													
Cluster1	0.0000	3.1166	11.14																																																																																																																													
Cluster2	3.1166	0.0000	11.4105																																																																																																																													
Cluster3	11.1714	11.4105	0.0000																																																																																																																													
	Cluster1	Cluster2	Cluster3	Cluster4																																																																																																																												
Cluster1	0.0000	3.3164	6.8557	12.4072																																																																																																																												
Cluster2	3.3164	0.0000	4.0000	12.4072																																																																																																																												
Cluster3	6.8557	4.0000	0.0000	6.9597																																																																																																																												
Cluster4	12.4072	12.7058	6.9597	0.0000																																																																																																																												
<table border="1"> <thead> <tr> <th>Step</th> <th>Number of clusters</th> <th>Similarity level</th> <th>Distance between centroids</th> <th>New in new cluster</th> <th>New in old cluster</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>6</td> <td>97.5810</td> <td>0.35009</td> <td>1</td> <td>2</td> </tr> <tr> <td>2</td> <td>5</td> <td>89.3224</td> <td>1.34335</td> <td>4</td> <td>2</td> </tr> <tr> <td>3</td> <td>4</td> <td>88.4913</td> <td>1.34335</td> <td>4</td> <td>3</td> </tr> <tr> <td>4</td> <td>3</td> <td>75.0885</td> <td>3.48960</td> <td>4</td> <td>5</td> </tr> <tr> <td>5</td> <td>2</td> <td>64.0492</td> <td>5.20307</td> <td>3</td> <td>4</td> </tr> <tr> <td>6</td> <td>1</td> <td>60.9034</td> <td>5.45826</td> <td>1</td> <td>3</td> </tr> </tbody> </table>	Step	Number of clusters	Similarity level	Distance between centroids	New in new cluster	New in old cluster	1	6	97.5810	0.35009	1	2	2	5	89.3224	1.34335	4	2	3	4	88.4913	1.34335	4	3	4	3	75.0885	3.48960	4	5	5	2	64.0492	5.20307	3	4	6	1	60.9034	5.45826	1	3	<table border="1"> <thead> <tr> <th>Step</th> <th>Number of clusters</th> <th>Similarity level</th> <th>Distance between centroids</th> <th>New in new cluster</th> <th>New in old cluster</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>6</td> <td>90.4725</td> <td>2.04982</td> <td>6</td> <td>2</td> </tr> <tr> <td>2</td> <td>5</td> <td>89.2718</td> <td>2.30363</td> <td>4</td> <td>4</td> </tr> <tr> <td>3</td> <td>4</td> <td>87.3204</td> <td>2.30363</td> <td>1</td> <td>2</td> </tr> <tr> <td>4</td> <td>3</td> <td>77.3546</td> <td>4.05517</td> <td>4</td> <td>5</td> </tr> <tr> <td>5</td> <td>2</td> <td>74.5585</td> <td>9.44602</td> <td>3</td> <td>4</td> </tr> <tr> <td>6</td> <td>1</td> <td>58.9666</td> <td>9.45919</td> <td>1</td> <td>5</td> </tr> </tbody> </table>	Step	Number of clusters	Similarity level	Distance between centroids	New in new cluster	New in old cluster	1	6	90.4725	2.04982	6	2	2	5	89.2718	2.30363	4	4	3	4	87.3204	2.30363	1	2	4	3	77.3546	4.05517	4	5	5	2	74.5585	9.44602	3	4	6	1	58.9666	9.45919	1	5	<table border="1"> <thead> <tr> <th>Step</th> <th>Number of clusters</th> <th>Similarity level</th> <th>Distance between centroids</th> <th>New in new cluster</th> <th>New in old cluster</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>6</td> <td>88.4137</td> <td>2.3981</td> <td>4</td> <td>2</td> </tr> <tr> <td>2</td> <td>5</td> <td>88.9702</td> <td>2.4965</td> <td>4</td> <td>3</td> </tr> <tr> <td>3</td> <td>4</td> <td>79.3204</td> <td>2.4965</td> <td>2</td> <td>1</td> </tr> <tr> <td>4</td> <td>3</td> <td>76.3112</td> <td>5.3815</td> <td>4</td> <td>5</td> </tr> <tr> <td>5</td> <td>2</td> <td>62.7574</td> <td>8.4294</td> <td>3</td> <td>3</td> </tr> <tr> <td>6</td> <td>1</td> <td>48.9446</td> <td>11.5561</td> <td>1</td> <td>5</td> </tr> </tbody> </table>	Step	Number of clusters	Similarity level	Distance between centroids	New in new cluster	New in old cluster	1	6	88.4137	2.3981	4	2	2	5	88.9702	2.4965	4	3	3	4	79.3204	2.4965	2	1	4	3	76.3112	5.3815	4	5	5	2	62.7574	8.4294	3	3	6	1	48.9446	11.5561	1	5
Step	Number of clusters	Similarity level	Distance between centroids	New in new cluster	New in old cluster																																																																																																																											
1	6	97.5810	0.35009	1	2																																																																																																																											
2	5	89.3224	1.34335	4	2																																																																																																																											
3	4	88.4913	1.34335	4	3																																																																																																																											
4	3	75.0885	3.48960	4	5																																																																																																																											
5	2	64.0492	5.20307	3	4																																																																																																																											
6	1	60.9034	5.45826	1	3																																																																																																																											
Step	Number of clusters	Similarity level	Distance between centroids	New in new cluster	New in old cluster																																																																																																																											
1	6	90.4725	2.04982	6	2																																																																																																																											
2	5	89.2718	2.30363	4	4																																																																																																																											
3	4	87.3204	2.30363	1	2																																																																																																																											
4	3	77.3546	4.05517	4	5																																																																																																																											
5	2	74.5585	9.44602	3	4																																																																																																																											
6	1	58.9666	9.45919	1	5																																																																																																																											
Step	Number of clusters	Similarity level	Distance between centroids	New in new cluster	New in old cluster																																																																																																																											
1	6	88.4137	2.3981	4	2																																																																																																																											
2	5	88.9702	2.4965	4	3																																																																																																																											
3	4	79.3204	2.4965	2	1																																																																																																																											
4	3	76.3112	5.3815	4	5																																																																																																																											
5	2	62.7574	8.4294	3	3																																																																																																																											
6	1	48.9446	11.5561	1	5																																																																																																																											
 <p>Cluster 1 : c[1,2] Cluster 2 : c[3,c[5,c[6,c[4,7]]]] Highest similarity level: 97.58 Distance btwn. Cluster Centroids: 10.4851 WSS1: 37.25 WSS2: 34</p>	 <p>Cluster 1 : c[1,2] Cluster 2 : c[3,c[5,c[4,c[6,7]]]] Highest similarity level: 90.47 Distance btwn. Cluster Centroids: 3.3, 11.1 & 11.4 WSS1: 0 WSS2: 0 WSS3 : 6</p>	 <p>Cluster 1 : c[1,2] Cluster 2 : c[3,c[5,c[6,c[4,7]]]] Highest similarity level: 89.4 Distance btwn. Cluster Centroids: 3.3, 6.8, 6.6, 6.9, 12.4 & 12.7 WSS1: 0 WSS2: 0 WSS3 : 0 WSS4: 37.25</p>																																																																																																																														
<ul style="list-style-type: none"> K-means cluster could be used when the number of clusters need to be controlled 2 & 4 clusters gave similar pairs, but the 3 cluster arranged the pairs differently The similarity measures are 97.5(2 cluster), 90.4(3 cluster) & 89.4 (4 cluster). The similarity measure goes down as the number of clusters went up. WSS1 from 2 cluster and WSS4 from 4 cluster are equal, may be because they have equal number of observations in it (4 nos) 																																																																																																																																

TABLE 10 : COMPARISON OF A3 & A5 ONLY SINGLE –EUCLIDEAN

For simplicity, all A5 values are not shown in this table. To see all the linkages, please refer A5.

Using Single Linkage (A3)	Using Single Linkage (A5)
Cluster 1 : c[1,2]	Cluster 1 : c[1,2]
Cluster 2 : c[3,c[7,c[4,c[5,6]]]]	Cluster 2 : c[3,c[7,c[5,c[4,6]]]]
Highest similarity level: 79.5	Highest similarity level: 79.5
Distance btwn. Cluster Centroids: 11.1692	Distance btwn. Cluster Centroids: 11.1692
WSS1: 5.5	WSS1: 5.5
WSS2: 76.0	WSS2: 76

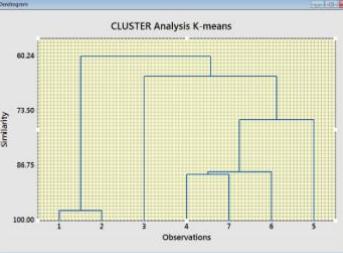
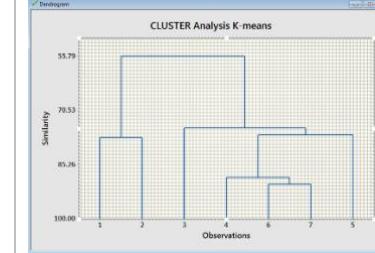
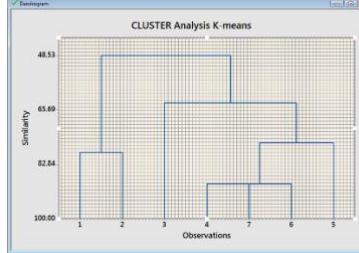
Findings against A3 & A5:

- When A3 & A5 are compared with 2 clusters, A13 has least distance between cluster centroids (10) i.e., the clusters are nearer.
- The clusters in A13 are slightly bigger than that of A3 & A5 (evident from average distance from centroid)
- The A13 clusters are of equal size compared to A3 & A5 (evident for WSS values)

A14: Cluster K-means (standardized Values)

All the variables were considered for K-means clustering.

TABLE 11 : COMPARISON OF K-MEANS 2, 3 & 4 CLUSTERS

A14 – CLUSTER ANALYSIS K-MEANS WITH 2 CLUSTERS	A14 – CLUSTER ANALYSIS K-MEANS WITH 3 CLUSTERS	A14 – CLUSTER ANALYSIS K-MEANS WITH 4 CLUSTERS																																																																																																																																																																								
Final Partition Number of clusters: 2 <table border="1"> <thead> <tr> <th></th> <th>Within cluster sum of squares</th> <th>Average distance from centroid</th> <th>Maximum distance from centroid</th> </tr> </thead> <tbody> <tr> <td>Number of observations</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Cluster1</td> <td>4</td> <td>4.454</td> <td>1.019</td> </tr> <tr> <td>Cluster2</td> <td>3</td> <td>3.964</td> <td>1.120</td> </tr> <tr> <td></td> <td></td> <td></td> <td>1.273</td> </tr> <tr> <td></td> <td></td> <td></td> <td>4.488</td> </tr> </tbody> </table>		Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid	Number of observations				Cluster1	4	4.454	1.019	Cluster2	3	3.964	1.120				1.273				4.488	Final Partition Number of clusters: 3 <table border="1"> <thead> <tr> <th></th> <th>Within cluster sum of squares</th> <th>Average distance from centroid</th> <th>Maximum distance from centroid</th> </tr> </thead> <tbody> <tr> <td>Number of observations</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Cluster1</td> <td>1</td> <td>0.000</td> <td>0.000</td> </tr> <tr> <td>Cluster2</td> <td>1</td> <td>0.000</td> <td>0.000</td> </tr> <tr> <td>Cluster3</td> <td>5</td> <td>8.812</td> <td>1.258</td> </tr> <tr> <td></td> <td></td> <td></td> <td>1.867</td> </tr> </tbody> </table>		Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid	Number of observations				Cluster1	1	0.000	0.000	Cluster2	1	0.000	0.000	Cluster3	5	8.812	1.258				1.867	Final Partition Number of clusters: 4 <table border="1"> <thead> <tr> <th></th> <th>Within cluster sum of squares</th> <th>Average distance from centroid</th> <th>Maximum distance from centroid</th> </tr> </thead> <tbody> <tr> <td>Number of observations</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Cluster1</td> <td>1</td> <td>0.000</td> <td>0.000</td> </tr> <tr> <td>Cluster2</td> <td>1</td> <td>0.000</td> <td>0.000</td> </tr> <tr> <td>Cluster3</td> <td>1</td> <td>0.000</td> <td>0.000</td> </tr> <tr> <td>Cluster4</td> <td>4</td> <td>4.454</td> <td>1.019</td> </tr> <tr> <td></td> <td></td> <td></td> <td>1.273</td> </tr> </tbody> </table>		Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid	Number of observations				Cluster1	1	0.000	0.000	Cluster2	1	0.000	0.000	Cluster3	1	0.000	0.000	Cluster4	4	4.454	1.019				1.273																																																																																												
	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid																																																																																																																																																																							
Number of observations																																																																																																																																																																										
Cluster1	4	4.454	1.019																																																																																																																																																																							
Cluster2	3	3.964	1.120																																																																																																																																																																							
			1.273																																																																																																																																																																							
			4.488																																																																																																																																																																							
	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid																																																																																																																																																																							
Number of observations																																																																																																																																																																										
Cluster1	1	0.000	0.000																																																																																																																																																																							
Cluster2	1	0.000	0.000																																																																																																																																																																							
Cluster3	5	8.812	1.258																																																																																																																																																																							
			1.867																																																																																																																																																																							
	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid																																																																																																																																																																							
Number of observations																																																																																																																																																																										
Cluster1	1	0.000	0.000																																																																																																																																																																							
Cluster2	1	0.000	0.000																																																																																																																																																																							
Cluster3	1	0.000	0.000																																																																																																																																																																							
Cluster4	4	4.454	1.019																																																																																																																																																																							
			1.273																																																																																																																																																																							
Cluster Centroids <table border="1"> <thead> <tr> <th>Variable</th> <th>Cluster1</th> <th>Cluster2</th> <th>Grand centroid</th> </tr> </thead> <tbody> <tr> <td>A1</td> <td>-0.5751</td> <td>0.7668</td> <td>0.0000</td> </tr> <tr> <td>A2</td> <td>-0.6088</td> <td>0.8117</td> <td>0.0000</td> </tr> <tr> <td>A3</td> <td>-0.7071</td> <td>0.9428</td> <td>0.0000</td> </tr> <tr> <td>A4</td> <td>-0.7618</td> <td>1.0187</td> <td>0.0000</td> </tr> <tr> <td>A5</td> <td>-0.7285</td> <td>0.9714</td> <td>-0.0000</td> </tr> </tbody> </table> Distances Between Cluster Centroids <table border="1"> <thead> <tr> <th>Cluster1</th> <th>Cluster2</th> </tr> </thead> <tbody> <tr> <td>Cluster1</td> <td>0.0000</td> </tr> <tr> <td>Cluster2</td> <td>3.5482</td> </tr> <tr> <td>Cluster2</td> <td>0.0000</td> </tr> </tbody> </table>	Variable	Cluster1	Cluster2	Grand centroid	A1	-0.5751	0.7668	0.0000	A2	-0.6088	0.8117	0.0000	A3	-0.7071	0.9428	0.0000	A4	-0.7618	1.0187	0.0000	A5	-0.7285	0.9714	-0.0000	Cluster1	Cluster2	Cluster1	0.0000	Cluster2	3.5482	Cluster2	0.0000	Cluster Centroids <table border="1"> <thead> <tr> <th>Variable</th> <th>Cluster1</th> <th>Cluster2</th> <th>Cluster3</th> <th>Grand centroid</th> </tr> </thead> <tbody> <tr> <td>A1</td> <td>0.7668</td> <td>1.1349</td> <td>3.7872</td> <td>0.0000</td> </tr> <tr> <td>A2</td> <td>0.8117</td> <td>1.1459</td> <td>0.5452</td> <td>0.0000</td> </tr> <tr> <td>A3</td> <td>0.9428</td> <td>1.0607</td> <td>-0.4980</td> <td>0.0000</td> </tr> <tr> <td>A4</td> <td>1.0187</td> <td>1.4533</td> <td>-0.4501</td> <td>0.0000</td> </tr> <tr> <td>A5</td> <td>0.9714</td> <td>1.0510</td> <td>-0.4982</td> <td>-0.0000</td> </tr> </tbody> </table> Distances Between Cluster Centroids <table border="1"> <thead> <tr> <th>Cluster1</th> <th>Cluster2</th> <th>Cluster3</th> </tr> </thead> <tbody> <tr> <td>Cluster1</td> <td>0.0000</td> <td>1.1349</td> </tr> <tr> <td>Cluster2</td> <td>1.1349</td> <td>0.0000</td> </tr> <tr> <td>Cluster3</td> <td>3.7872</td> <td>3.8811</td> </tr> <tr> <td>Cluster3</td> <td>0.0000</td> <td></td> </tr> </tbody> </table>	Variable	Cluster1	Cluster2	Cluster3	Grand centroid	A1	0.7668	1.1349	3.7872	0.0000	A2	0.8117	1.1459	0.5452	0.0000	A3	0.9428	1.0607	-0.4980	0.0000	A4	1.0187	1.4533	-0.4501	0.0000	A5	0.9714	1.0510	-0.4982	-0.0000	Cluster1	Cluster2	Cluster3	Cluster1	0.0000	1.1349	Cluster2	1.1349	0.0000	Cluster3	3.7872	3.8811	Cluster3	0.0000		Cluster Centroids <table border="1"> <thead> <tr> <th>Variable</th> <th>Cluster1</th> <th>Cluster2</th> <th>Cluster3</th> <th>Cluster4</th> <th>Grand centroid</th> </tr> </thead> <tbody> <tr> <td>A1</td> <td>0.7466</td> <td>1.4522</td> <td>5.0511</td> <td>-0.1511</td> <td>0.0000</td> </tr> <tr> <td>A2</td> <td>0.8103</td> <td>1.1459</td> <td>0.5452</td> <td>0.6093</td> <td>0.0000</td> </tr> <tr> <td>A3</td> <td>1.4142</td> <td>1.0607</td> <td>0.3536</td> <td>0.7071</td> <td>0.0000</td> </tr> <tr> <td>A4</td> <td>0.7970</td> <td>1.4533</td> <td>0.3790</td> <td>-0.7618</td> <td>0.0000</td> </tr> <tr> <td>A5</td> <td>1.4100</td> <td>1.0810</td> <td>0.4230</td> <td>-0.7285</td> <td>-0.0000</td> </tr> </tbody> </table> Distances Between Cluster Centroids <table border="1"> <thead> <tr> <th>Cluster1</th> <th>Cluster2</th> <th>Cluster3</th> <th>Cluster4</th> </tr> </thead> <tbody> <tr> <td>Cluster1</td> <td>0.0000</td> <td>1.1349</td> <td>2.3246</td> <td>4.2033</td> </tr> <tr> <td>Cluster2</td> <td>1.1349</td> <td>0.0000</td> <td>2.2803</td> <td>4.3149</td> </tr> <tr> <td>Cluster3</td> <td>3.7872</td> <td>3.8811</td> <td>0.0000</td> <td>2.3340</td> </tr> <tr> <td>Cluster4</td> <td>4.2033</td> <td>4.3149</td> <td>2.3340</td> <td>0.0000</td> </tr> </tbody> </table>	Variable	Cluster1	Cluster2	Cluster3	Cluster4	Grand centroid	A1	0.7466	1.4522	5.0511	-0.1511	0.0000	A2	0.8103	1.1459	0.5452	0.6093	0.0000	A3	1.4142	1.0607	0.3536	0.7071	0.0000	A4	0.7970	1.4533	0.3790	-0.7618	0.0000	A5	1.4100	1.0810	0.4230	-0.7285	-0.0000	Cluster1	Cluster2	Cluster3	Cluster4	Cluster1	0.0000	1.1349	2.3246	4.2033	Cluster2	1.1349	0.0000	2.2803	4.3149	Cluster3	3.7872	3.8811	0.0000	2.3340	Cluster4	4.2033	4.3149	2.3340	0.0000																															
Variable	Cluster1	Cluster2	Grand centroid																																																																																																																																																																							
A1	-0.5751	0.7668	0.0000																																																																																																																																																																							
A2	-0.6088	0.8117	0.0000																																																																																																																																																																							
A3	-0.7071	0.9428	0.0000																																																																																																																																																																							
A4	-0.7618	1.0187	0.0000																																																																																																																																																																							
A5	-0.7285	0.9714	-0.0000																																																																																																																																																																							
Cluster1	Cluster2																																																																																																																																																																									
Cluster1	0.0000																																																																																																																																																																									
Cluster2	3.5482																																																																																																																																																																									
Cluster2	0.0000																																																																																																																																																																									
Variable	Cluster1	Cluster2	Cluster3	Grand centroid																																																																																																																																																																						
A1	0.7668	1.1349	3.7872	0.0000																																																																																																																																																																						
A2	0.8117	1.1459	0.5452	0.0000																																																																																																																																																																						
A3	0.9428	1.0607	-0.4980	0.0000																																																																																																																																																																						
A4	1.0187	1.4533	-0.4501	0.0000																																																																																																																																																																						
A5	0.9714	1.0510	-0.4982	-0.0000																																																																																																																																																																						
Cluster1	Cluster2	Cluster3																																																																																																																																																																								
Cluster1	0.0000	1.1349																																																																																																																																																																								
Cluster2	1.1349	0.0000																																																																																																																																																																								
Cluster3	3.7872	3.8811																																																																																																																																																																								
Cluster3	0.0000																																																																																																																																																																									
Variable	Cluster1	Cluster2	Cluster3	Cluster4	Grand centroid																																																																																																																																																																					
A1	0.7466	1.4522	5.0511	-0.1511	0.0000																																																																																																																																																																					
A2	0.8103	1.1459	0.5452	0.6093	0.0000																																																																																																																																																																					
A3	1.4142	1.0607	0.3536	0.7071	0.0000																																																																																																																																																																					
A4	0.7970	1.4533	0.3790	-0.7618	0.0000																																																																																																																																																																					
A5	1.4100	1.0810	0.4230	-0.7285	-0.0000																																																																																																																																																																					
Cluster1	Cluster2	Cluster3	Cluster4																																																																																																																																																																							
Cluster1	0.0000	1.1349	2.3246	4.2033																																																																																																																																																																						
Cluster2	1.1349	0.0000	2.2803	4.3149																																																																																																																																																																						
Cluster3	3.7872	3.8811	0.0000	2.3340																																																																																																																																																																						
Cluster4	4.2033	4.3149	2.3340	0.0000																																																																																																																																																																						
<table border="1"> <thead> <tr> <th>Step</th> <th>clusters</th> <th>Number of obs.</th> <th>Similarity level</th> <th>Distance level</th> <th>Clusters joined</th> <th>New cluster</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>6</td> <td>97.5981</td> <td>0.11746</td> <td>1</td> <td>2</td> <td>1</td> </tr> <tr> <td>2</td> <td>5</td> <td>88.8851</td> <td>0.54353</td> <td>4</td> <td>7</td> <td>4</td> </tr> <tr> <td>3</td> <td>4</td> <td>88.1852</td> <td>0.57776</td> <td>4</td> <td>6</td> <td>4</td> </tr> <tr> <td>4</td> <td>3</td> <td>75.5826</td> <td>1.19404</td> <td>4</td> <td>5</td> <td>4</td> </tr> <tr> <td>5</td> <td>2</td> <td>65.0423</td> <td>1.70947</td> <td>3</td> <td>4</td> <td>3</td> </tr> <tr> <td>6</td> <td>1</td> <td>60.2449</td> <td>1.94407</td> <td>1</td> <td>3</td> <td>1</td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>7</td> </tr> </tbody> </table>	Step	clusters	Number of obs.	Similarity level	Distance level	Clusters joined	New cluster	1	6	97.5981	0.11746	1	2	1	2	5	88.8851	0.54353	4	7	4	3	4	88.1852	0.57776	4	6	4	4	3	75.5826	1.19404	4	5	4	5	2	65.0423	1.70947	3	4	3	6	1	60.2449	1.94407	1	3	1							7	<table border="1"> <thead> <tr> <th>Step</th> <th>clusters</th> <th>Number of obs.</th> <th>Similarity level</th> <th>Distance level</th> <th>Clusters joined</th> <th>New cluster</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>6</td> <td>90.5749</td> <td>0.68760</td> <td>6</td> <td>7</td> <td>6</td> </tr> <tr> <td>2</td> <td>5</td> <td>88.7985</td> <td>0.81720</td> <td>4</td> <td>6</td> <td>4</td> </tr> <tr> <td>3</td> <td>4</td> <td>77.9625</td> <td>1.21651</td> <td>2</td> <td>1</td> <td>2</td> </tr> <tr> <td>4</td> <td>3</td> <td>77.0228</td> <td>1.65853</td> <td>4</td> <td>5</td> <td>4</td> </tr> <tr> <td>5</td> <td>2</td> <td>75.3457</td> <td>1.79864</td> <td>3</td> <td>4</td> <td>3</td> </tr> <tr> <td>6</td> <td>1</td> <td>55.7903</td> <td>3.22529</td> <td>1</td> <td>3</td> <td>1</td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>7</td> </tr> </tbody> </table>	Step	clusters	Number of obs.	Similarity level	Distance level	Clusters joined	New cluster	1	6	90.5749	0.68760	6	7	6	2	5	88.7985	0.81720	4	6	4	3	4	77.9625	1.21651	2	1	2	4	3	77.0228	1.65853	4	5	4	5	2	75.3457	1.79864	3	4	3	6	1	55.7903	3.22529	1	3	1							7	<table border="1"> <thead> <tr> <th>Step</th> <th>clusters</th> <th>Number of obs.</th> <th>Similarity level</th> <th>Distance level</th> <th>Clusters joined</th> <th>New cluster</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>6</td> <td>88.9600</td> <td>0.84760</td> <td>4</td> <td>7</td> <td>4</td> </tr> <tr> <td>2</td> <td>5</td> <td>88.9373</td> <td>0.84934</td> <td>4</td> <td>6</td> <td>4</td> </tr> <tr> <td>3</td> <td>4</td> <td>79.2321</td> <td>1.21651</td> <td>1</td> <td>2</td> <td>1</td> </tr> <tr> <td>4</td> <td>3</td> <td>76.1581</td> <td>1.32201</td> <td>4</td> <td>5</td> <td>4</td> </tr> <tr> <td>5</td> <td>2</td> <td>63.5004</td> <td>2.02227</td> <td>3</td> <td>4</td> <td>3</td> </tr> <tr> <td>6</td> <td>1</td> <td>48.5321</td> <td>3.95147</td> <td>1</td> <td>3</td> <td>1</td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>7</td> </tr> </tbody> </table>	Step	clusters	Number of obs.	Similarity level	Distance level	Clusters joined	New cluster	1	6	88.9600	0.84760	4	7	4	2	5	88.9373	0.84934	4	6	4	3	4	79.2321	1.21651	1	2	1	4	3	76.1581	1.32201	4	5	4	5	2	63.5004	2.02227	3	4	3	6	1	48.5321	3.95147	1	3	1							7
Step	clusters	Number of obs.	Similarity level	Distance level	Clusters joined	New cluster																																																																																																																																																																				
1	6	97.5981	0.11746	1	2	1																																																																																																																																																																				
2	5	88.8851	0.54353	4	7	4																																																																																																																																																																				
3	4	88.1852	0.57776	4	6	4																																																																																																																																																																				
4	3	75.5826	1.19404	4	5	4																																																																																																																																																																				
5	2	65.0423	1.70947	3	4	3																																																																																																																																																																				
6	1	60.2449	1.94407	1	3	1																																																																																																																																																																				
						7																																																																																																																																																																				
Step	clusters	Number of obs.	Similarity level	Distance level	Clusters joined	New cluster																																																																																																																																																																				
1	6	90.5749	0.68760	6	7	6																																																																																																																																																																				
2	5	88.7985	0.81720	4	6	4																																																																																																																																																																				
3	4	77.9625	1.21651	2	1	2																																																																																																																																																																				
4	3	77.0228	1.65853	4	5	4																																																																																																																																																																				
5	2	75.3457	1.79864	3	4	3																																																																																																																																																																				
6	1	55.7903	3.22529	1	3	1																																																																																																																																																																				
						7																																																																																																																																																																				
Step	clusters	Number of obs.	Similarity level	Distance level	Clusters joined	New cluster																																																																																																																																																																				
1	6	88.9600	0.84760	4	7	4																																																																																																																																																																				
2	5	88.9373	0.84934	4	6	4																																																																																																																																																																				
3	4	79.2321	1.21651	1	2	1																																																																																																																																																																				
4	3	76.1581	1.32201	4	5	4																																																																																																																																																																				
5	2	63.5004	2.02227	3	4	3																																																																																																																																																																				
6	1	48.5321	3.95147	1	3	1																																																																																																																																																																				
						7																																																																																																																																																																				
 <p>Cluster 1 : c[1,2] Cluster 2 : c[3,c[5,c[6,c[4,7]]]] Highest similarity level: 97.59 Distance btwn. Cluster Centroids: 3.54 WSS1: 4.45 WSS2: 3.96</p>	 <p>Cluster 1 : c[1,2] Cluster 2 : c[3,c[5,c[4,c[6,7]]]] Highest similarity level: 90.47 Distance btwn. Cluster Centroids: 1.1, 3.7, 3.8 WSS1: 0 WSS2: 0 WSS3 : 8.8</p>	 <p>Cluster 1 : c[1,2] Cluster 2 : c[3,c[5,c[6,c[4,7]]]] Highest similarity level: 88.96 Distance btwn. Cluster Centroids: 1.1, 2.2, 2.3, 2.3, 4.2 & 4.3 WSS1: 0 WSS2: 0 WSS3 : 0 WSS4: 4.45</p>																																																																																																																																																																								

- WSS1 from 2 cluster and WSS4 from 4 cluster are equal, may be because they have equal number of observations in it (4 nos)
- Compared to actual values (A13), A14 have smaller, nearer clusters.
- In cluster 4 approach, the similarity levels at step 1 & step 2 are almost equal causing a visually no-division between pair 4 & 7 with observation 6.

TABLE 12 : COMPARISON OF A4 & A6 ONLY SINGLE -EUCLIDEAN

Using Single Linkage (A4)	Using Single Linkages (A6)
Cluster 1 : c[1,2]	Cluster 1 : c[1,2]
Cluster 2 : c[3,c[7,c[5,c[4,6]]]]	Cluster 2 : c[3,c[7,c[5,c[4,6]]]]
Highest similarity level: 79.4	Highest similarity level: 79.4
Distance btwn. Cluster Centroids: 3.792	Distance btwn. Cluster Centroids: 3.79
WSS1: 0.643	WSS1: 0.64
WSS2: 8.811	WSS2: 8.81

When A14 is compared to A4 & A6: (All of them are standardized values)

- Cluster k – means has generally found the similarities at high levels (around 90 or above), when compared to A4 & A6 (~80 for all linkage methods in A6). This could be interpreted that cluster k-means could either find the pairs early or find pairs that are even far away, which is could be brute forced by the number of cluster (k – value) selection.
- From WSS values, it could be noted that cluster k-means tends to bring the clusters nearer.

A15: General Findings & Insights

Observation on Distance Measures

- Euclidean, Pearson & Manhattan do catch the similar pairs at different levels, but the highest similarity levels are almost close.
- But the Squared distance methods (both the sq. Euclidean & Sq. Pearson) have higher – highest similarity levels when compared to the above set. This could lead to find pairs that are even little far away. So, the Squared distance methods shall be used with a caution.
- Manhattan distance measure could be applied to shortest distances in traffic.

Observation on Linkages

- Average, McQuitty & Ward linkages look similar to each other, in the way it found pairs and arranged the tree structure.
- Complete linkage tends position the clusters farther first, and build pairs in between the clusters. (It was noted that pair 3 & 4 was formed only by this linkage method)
- Centroid linkage tends to find the smallest dissimilarity, hence a commonly used method in biology (genomic cluster analysis)
- Median linkage also works like Centroid linkage.

Actual Values & Standardized Values

- Standardization of values help to transform the values and bring all the values to similar scale. In this project, as all the values were transformed, there was no difference found on the variable correlations (evident from A7 & A8).
- Also, when it comes to clustering behavior on the standardized values, it was noted that cluster size and the distance between the clusters were scaled down, when compared to actual values.

Clustering of Observations & Variables

- Clustering of Variables: It could be noted that clusters got formed based on the set of highly correlated variables in to the same cluster. (X2 to X5).
- Clustering of Observations: It varied based on the linkage method, distance measure and the selection of variables. A suitable linkage method & distance measure could be chosen based on the study (for e.g., apply centroid method in biological data).
- General: Different combinations could be tried for different purposes.
- WSS → represent the size of the cluster
- Distance between centroid → represent how far away the clusters are
- Similarity level → represent how near the values are.

REFERENCES:

1. http://www.lx.it.pt/~afred/tutorials/B_Clustering_Algorithms.pdf
2. <http://stats.stackexchange.com/questions/195446/choosing-the-right-linkage-method-for-hierarchical-clustering>