Initial Data

col 0 col 1

0.73 56.0

0.35 42.0

0.11 64.0

0.07 16.0

0.6 95.0

0.78 25.0

0.55 9.0

0.26 74.0

0.88 49.0

0.54 32.0

0.17 21.0

0.71 57.0

0.11 62.0

0.7 31.0

0.36 59.0

0.39 87.0

0.53 7.0

0.6 86.0

0.46 76.0

0.11 42.0

---------------------------------------------

Euclidean Distance k = 2

---------------------------------------------

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 2.42, Max : 21.6, Sum : 118.01

Intra-Cluster distances in cluster 1

Min : 2.41, Max : 23.41, Sum : 120.0

Sum of intra-cluster distance

between clusters :238.01

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 2.42, Max : 21.6, Sum : 118.01

Intra-Cluster distances in cluster 1

Min : 2.41, Max : 23.41, Sum : 120.0

Sum of intra-cluster distance

between clusters :238.01

---------------------------------------------

Euclidean Distance k = 4

---------------------------------------------

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 0.53, Max : 9.4, Sum : 30.53

Intra-Cluster distances in cluster 1

Min : 2.33, Max : 5.67, Sum : 11.33

Intra-Cluster distances in cluster 2

Min : 3.01, Max : 15.0, Sum : 74.02

Intra-Cluster distances in cluster 3

Min : 5.0, Max : 7.0, Sum : 24.0

Sum of intra-cluster distance

between clusters :139.88

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 0.53, Max : 9.4, Sum : 30.53

Intra-Cluster distances in cluster 1

Min : 2.33, Max : 5.67, Sum : 11.33

Intra-Cluster distances in cluster 2

Min : 2.15, Max : 13.14, Sum : 59.19

Intra-Cluster distances in cluster 3

Min : 3.0, Max : 9.0, Sum : 32.0

Sum of intra-cluster distance

between clusters :133.05

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 0.53, Max : 9.4, Sum : 30.53

Intra-Cluster distances in cluster 1

Min : 2.33, Max : 5.67, Sum : 11.33

Intra-Cluster distances in cluster 2

Min : 2.8, Max : 9.8, Sum : 30.82

Intra-Cluster distances in cluster 3

Min : 0.28, Max : 12.0, Sum : 44.29

Sum of intra-cluster distance

between clusters :116.97

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 0.53, Max : 9.4, Sum : 30.53

Intra-Cluster distances in cluster 1

Min : 2.33, Max : 5.67, Sum : 11.33

Intra-Cluster distances in cluster 2

Min : 2.8, Max : 9.8, Sum : 30.82

Intra-Cluster distances in cluster 3

Min : 0.28, Max : 12.0, Sum : 44.29

Sum of intra-cluster distance

between clusters :116.97

---------------------------------------------

Manhattan Distance k = 2

---------------------------------------------

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 2.08, Max : 67.73, Sum : 252.62

Intra-Cluster distances in cluster 1

Min : 2.23, Max : 62.48, Sum : 305.01

Sum of intra-cluster distance

between clusters :557.63

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 2.08, Max : 67.73, Sum : 252.62

Intra-Cluster distances in cluster 1

Min : 2.23, Max : 62.48, Sum : 305.01

Sum of intra-cluster distance

between clusters :557.63

---------------------------------------------

Manhattan Distance k = 4

---------------------------------------------

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 0.05, Max : 79.58, Sum : 194.76

Intra-Cluster distances in cluster 1

Min : 25.74, Max : 47.5, Sum : 106.36

Intra-Cluster distances in cluster 2

Min : 4.2, Max : 49.06, Sum : 196.96

Intra-Cluster distances in cluster 3

Min : 5.12, Max : 53.16, Sum : 97.66

Sum of intra-cluster distance

between clusters :595.74

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 0.05, Max : 79.58, Sum : 194.76

#############################################

Euclidean Distance k = 2

#############################################

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 2.42, Max : 21.6, Sum : 118.01

Intra-Cluster distances in cluster 1

Min : 2.41, Max : 23.41, Sum : 120.0

Sum of intra-cluster distance

between clusters :238.01

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 2.42, Max : 21.6, Sum : 118.01

Intra-Cluster distances in cluster 1

Min : 2.41, Max : 23.41, Sum : 120.0

Sum of intra-cluster distance

between clusters :238.01

#############################################

Euclidean Distance k = 4

#############################################

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 0.53, Max : 9.4, Sum : 30.53

Intra-Cluster distances in cluster 1

Min : 2.33, Max : 5.67, Sum : 11.33

Intra-Cluster distances in cluster 2

Min : 3.01, Max : 15.0, Sum : 74.02

Intra-Cluster distances in cluster 3

Min : 5.0, Max : 7.0, Sum : 24.0

Sum of intra-cluster distance

between clusters :139.88

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 0.53, Max : 9.4, Sum : 30.53

Intra-Cluster distances in cluster 1

Min : 2.33, Max : 5.67, Sum : 11.33

Intra-Cluster distances in cluster 2

Min : 2.15, Max : 13.14, Sum : 59.19

Intra-Cluster distances in cluster 3

Min : 3.0, Max : 9.0, Sum : 32.0

Sum of intra-cluster distance

between clusters :133.05

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 0.53, Max : 9.4, Sum : 30.53

Intra-Cluster distances in cluster 1

Min : 2.33, Max : 5.67, Sum : 11.33

Intra-Cluster distances in cluster 2

Min : 2.8, Max : 9.8, Sum : 30.82

Intra-Cluster distances in cluster 3

Min : 0.28, Max : 12.0, Sum : 44.29

Sum of intra-cluster distance

between clusters :116.97

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 0.53, Max : 9.4, Sum : 30.53

Intra-Cluster distances in cluster 1

Min : 2.33, Max : 5.67, Sum : 11.33

Intra-Cluster distances in cluster 2

Min : 2.8, Max : 9.8, Sum : 30.82

Intra-Cluster distances in cluster 3

Min : 0.28, Max : 12.0, Sum : 44.29

Sum of intra-cluster distance

between clusters :116.97

#############################################

Manhattan Distance k = 2

#############################################

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 2.08, Max : 67.73, Sum : 252.62

Intra-Cluster distances in cluster 1

Min : 2.23, Max : 62.48, Sum : 305.01

Sum of intra-cluster distance

between clusters :557.63

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 2.08, Max : 67.73, Sum : 252.62

Intra-Cluster distances in cluster 1

Min : 2.23, Max : 62.48, Sum : 305.01

Sum of intra-cluster distance

between clusters :557.63

#############################################

Manhattan Distance k = 4

#############################################

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 0.05, Max : 79.58, Sum : 194.76

Intra-Cluster distances in cluster 1

Min : 25.74, Max : 47.5, Sum : 106.36

Intra-Cluster distances in cluster 2

Min : 4.2, Max : 49.06, Sum : 196.96

Intra-Cluster distances in cluster 3

Min : 5.12, Max : 53.16, Sum : 97.66

Sum of intra-cluster distance

between clusters :595.74

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 0.05, Max : 79.58, Sum : 194.76

Intra-Cluster distances in cluster 1

Min : 25.74, Max : 47.5, Sum : 106.36

Intra-Cluster distances in cluster 2

Min : 2.37, Max : 50.89, Sum : 167.41

Intra-Cluster distances in cluster 3

Min : 3.15, Max : 51.19, Sum : 118.12

Sum of intra-cluster distance

between clusters :586.65

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 0.05, Max : 79.58, Sum : 194.76

Intra-Cluster distances in cluster 1

Min : 25.74, Max : 47.5, Sum : 106.36

Intra-Cluster distances in cluster 2

Min : 2.63, Max : 55.89, Sum : 123.58

Intra-Cluster distances in cluster 3

Min : 0.29, Max : 54.84, Sum : 202.97

Sum of intra-cluster distance

between clusters :627.66

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 0.05, Max : 79.58, Sum : 194.76

Intra-Cluster distances in cluster 1

Min : 25.74, Max : 47.5, Sum : 106.36

Intra-Cluster distances in cluster 2

Min : 2.63, Max : 55.89, Sum : 123.58

Intra-Cluster distances in cluster 3

Min : 0.29, Max : 54.84, Sum : 202.97

Sum of intra-cluster distance

between clusters :627.66

----------------------------------------

Intra-Cluster distances in cluster 1

Min : 25.74, Max : 47.5, Sum : 106.36

Intra-Cluster distances in cluster 2

Min : 2.37, Max : 50.89, Sum : 167.41

Intra-Cluster distances in cluster 3

Min : 3.15, Max : 51.19, Sum : 118.12

Sum of intra-cluster distance

between clusters :586.65

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 0.05, Max : 79.58, Sum : 194.76

Intra-Cluster distances in cluster 1

Min : 25.74, Max : 47.5, Sum : 106.36

Intra-Cluster distances in cluster 2

Min : 2.63, Max : 55.89, Sum : 123.58

Intra-Cluster distances in cluster 3

Min : 0.29, Max : 54.84, Sum : 202.97

Sum of intra-cluster distance

between clusters :627.66

----------------------------------------

Intra-Cluster distances in cluster 0

Min : 0.05, Max : 79.58, Sum : 194.76

Intra-Cluster distances in cluster 1

Min : 25.74, Max : 47.5, Sum : 106.36

Intra-Cluster distances in cluster 2

Min : 2.63, Max : 55.89, Sum : 123.58

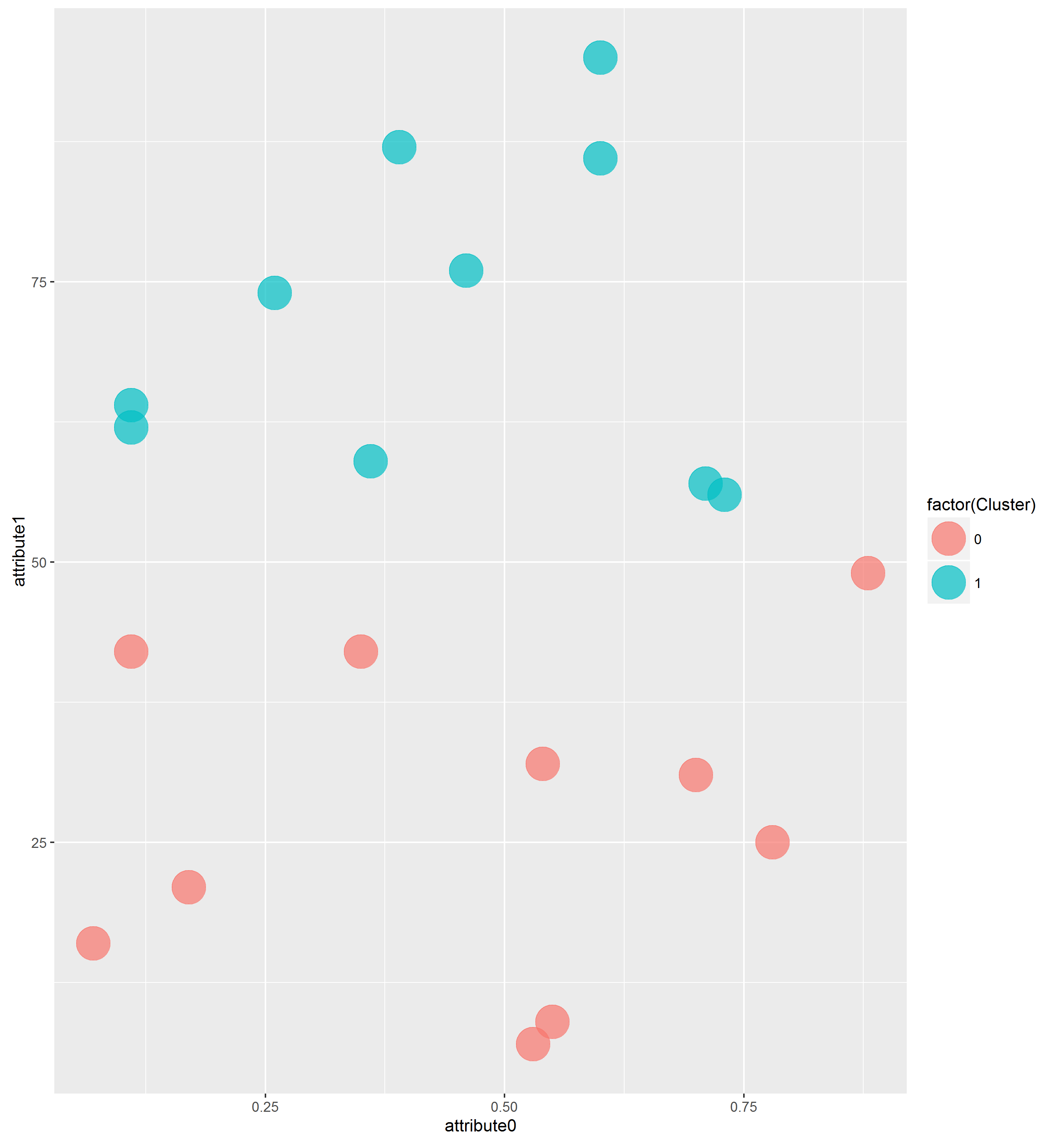
Intra-Cluster distances in cluster 3

Min : 0.29, Max : 54.84, Sum : 202.97

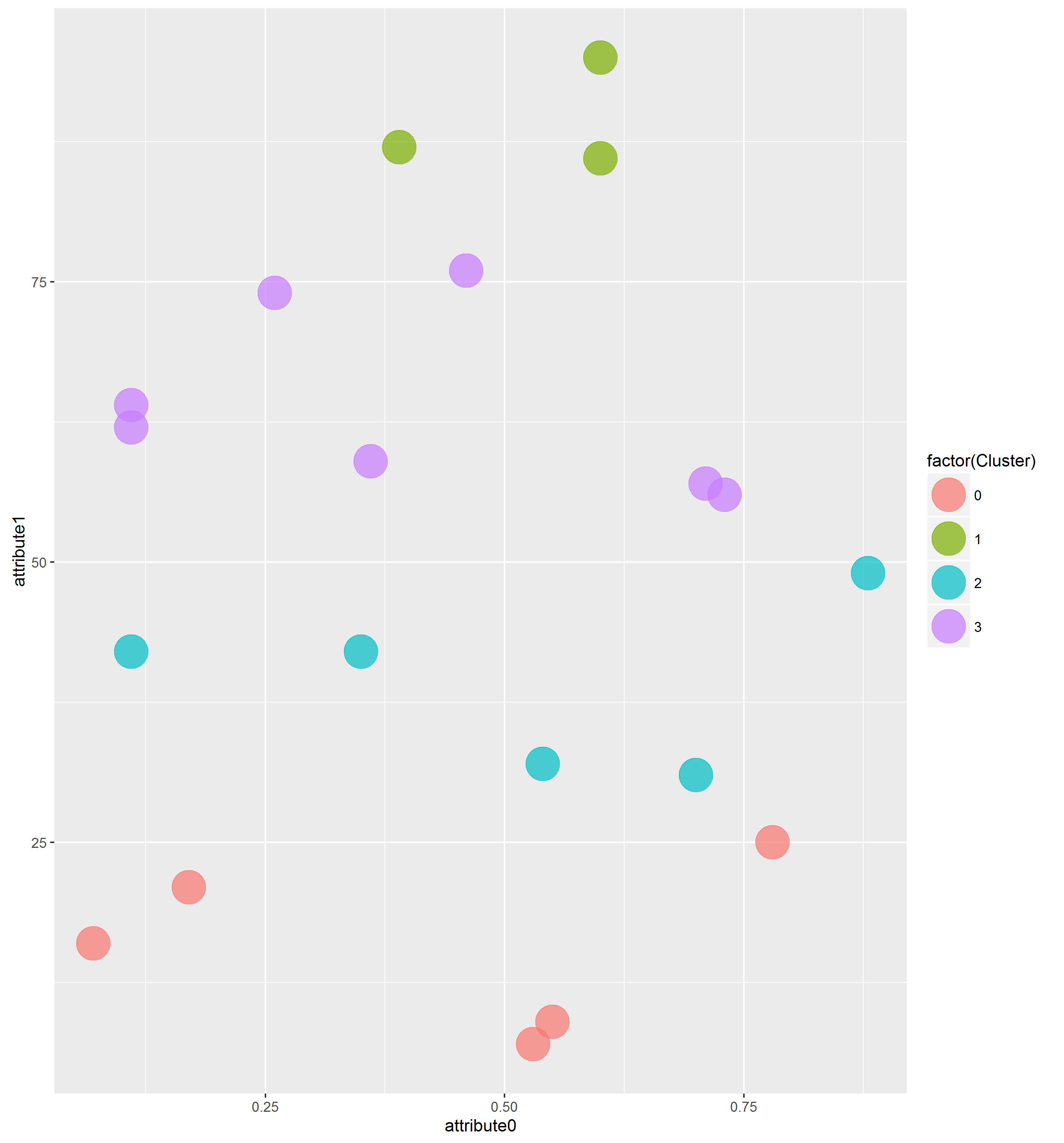
Sum of intra-cluster distance

between clusters :627.66

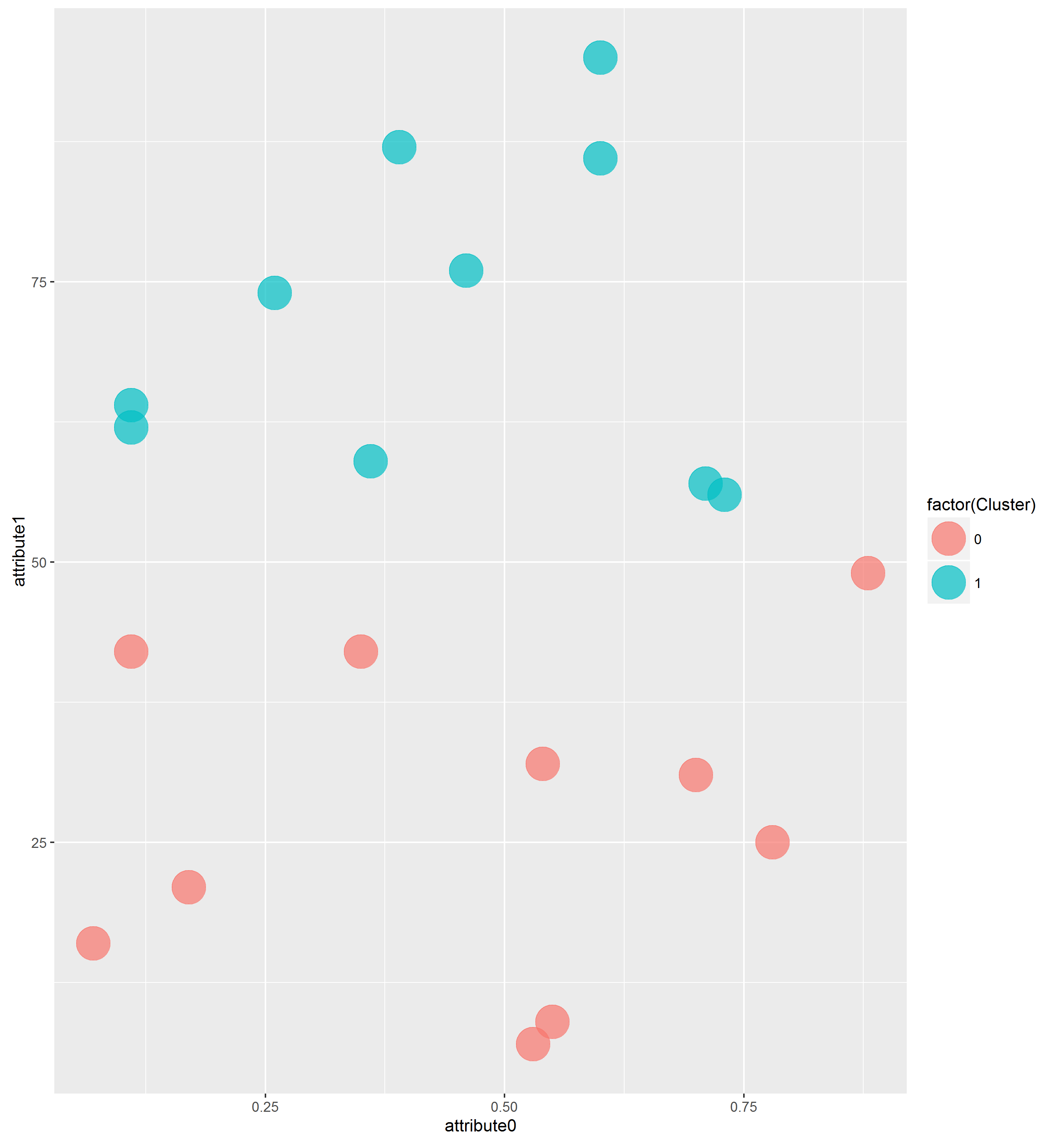
Euclidean Distance with k = 2



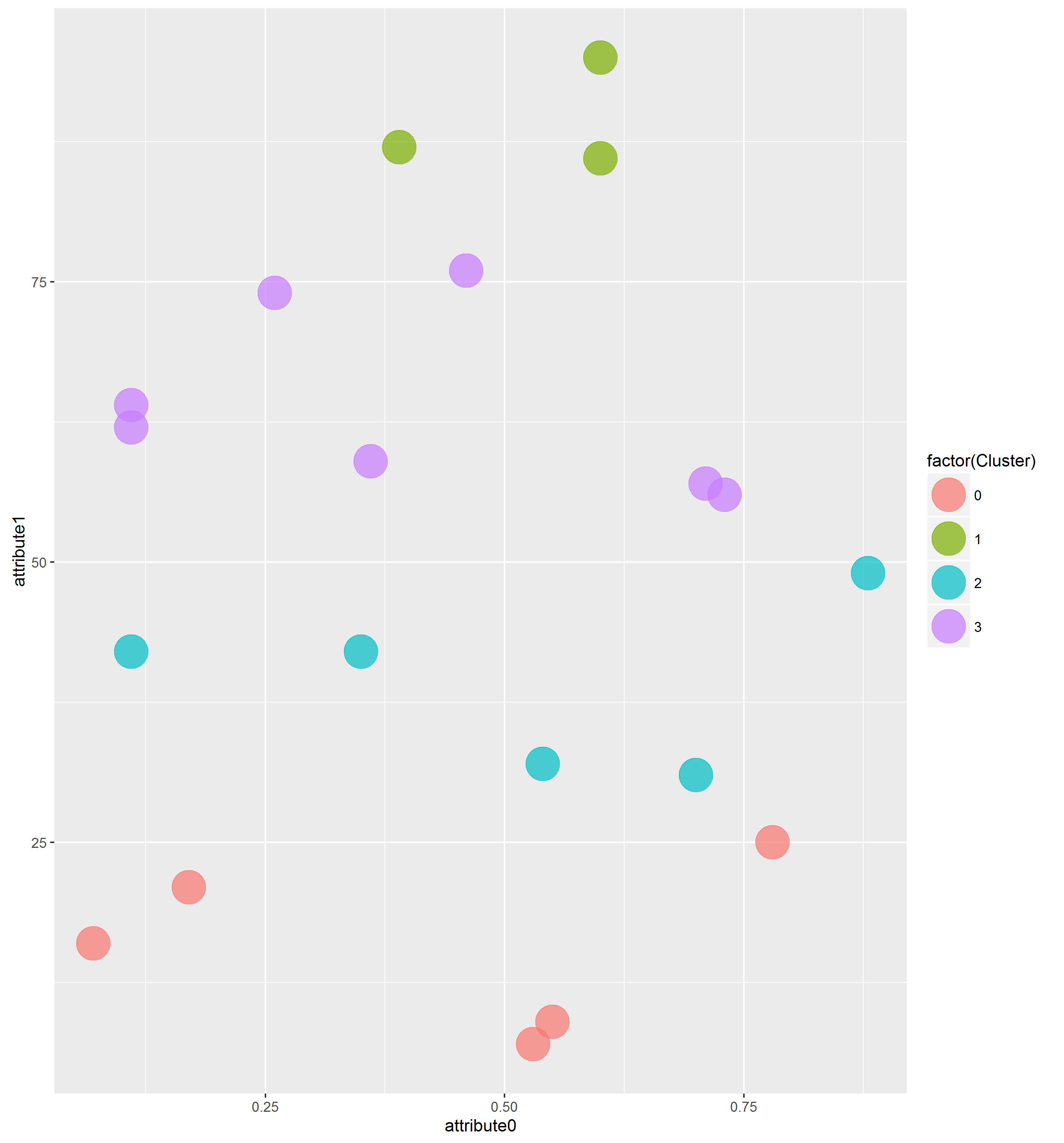
Euclidean Distance with k = 4



Manhattan Distance with k = 2



Manhattan Distance with k = 4



**Source Code**

**Main.java**

**import** java.util.Random;

**import** java.io.FileWriter;

**import** java.io.IOException;

public class Main {

public static void main(String[] args) {

// Load Data

double[][] data = setValues();

// Run k-means clustering

String fileName;

int [] clusterArr;

kmeans km;

km = **new** kmeans(data, 2, "euclidean");

clusterArr = km.start();

fileName = "E:\\Google Drive\\Class\\4342\\proj1\\euclidean2.csv";

saveCSV(data, clusterArr, fileName);

km = **new** kmeans(data, 4, "euclidean");

clusterArr = km.start();

fileName = "E:\\Google Drive\\Class\\4342\\proj1\\euclidean4.csv";

saveCSV(data, clusterArr, fileName);

km = **new** kmeans(data, 2, "manhattan");

clusterArr = km.start();

fileName = "E:\\Google Drive\\Class\\4342\\proj1\\manhattan2.csv";

saveCSV(data, clusterArr, fileName);

km = **new** kmeans(data, 4, "manhattan");

clusterArr = km.start();

fileName = "E:\\Google Drive\\Class\\4342\\proj1\\manhattan4.csv";

saveCSV(data, clusterArr, fileName);

}

// Save data into csv

private static void saveCSV(double[][] data, int[] clusterArr, String fileName) {

**try** {

FileWriter writer = **new** FileWriter(fileName);

// write header

**for** (int z = 0; z < data[0].length; z++) {

writer.append("attribute" + z);

writer.append(',');

}

writer.append("Cluster");

writer.append('\n');

// write observation

**for** (int i = 0; i < data.length; i++) {

**for** (int j = 0; j < data[0].length; j++) {

writer.append(String.valueOf(data[i][j]));

writer.append(',');

}

writer.append(String.valueOf(clusterArr[i]));

writer.append("\n");

}

writer.flush();

writer.close();

} **catch** (IOException e) {

e.printStackTrace();

}

}

// Arbitrarily generate random observations

public static double[][] setValues() {

int colLength = 2;

int rowLength = 20;

double [][] data = **new** double [rowLength][colLength];

Random randnum = **new** Random(28); // set seed

**for** (int i = 0; i < rowLength; i++) {

// Assign random numbers

**for** (int j = 0; j < colLength; j++) {

**if** (j == 0) data [i][j] = Math.floor(randnum.nextDouble() \* 100.0)/ 100.0;

**else** data [i][j] = Math.floor(randnum.nextDouble() \* 100.0);

}

}

// Print Data

System.out.println("Initial Data");

**for** (int j = 0; j < colLength; j++)

System.out.print("col " + j + "\t");

System.out.println("");

**for** (int i = 0; i < rowLength; i++) {

**for** (int j = 0; j < colLength; j++) {

System.out.print(data[i][j]);

System.out.print("\t");

}

System.out.println("");

}

System.out.println("");

**return** data;

}

}

**Kmeans.java**

**import** java.util.Random;

public class kmeans {

public int k;

public int rowLength;

public int colLength;

public int[] cluster;

public double[][] data;

public double[][] centroidDist;

public double[][] intraDist;

public double[][] centroid;

public double[][] clustrMean;

public double sumIntraDist;

public boolean globalConverge;

public String method;

public kmeans (double[][] data, int k, String method) {

**this**.data = data;

**this**.k = k;

**this**.method = method;

**this**.rowLength = data.length;

**this**.colLength = data[0].length;

**this**.centroid = **new** double [k][colLength];

**this**.centroidDist = **new** double [rowLength][k];

**this**.cluster = **new** int [rowLength];

}

public int[] start() {

centroidInit();

**while** (!globalConverge){

System.out.println("----------------------------------------");

calcDist(method);

assignCluster();

calcCentroid();

System.out.println("\nSum of intra-cluster distance \nbetween clusters :" + sumIntraDist);

}

**return** cluster;

}

// Select random row as centroid

private void centroidInit() {

**for** (int i = 0; i < k; i++) {

int idx;

**for** (int j = 0; j < colLength; j++) {

Random randnum = **new** Random(i\*2);

idx = randnum.nextInt(rowLength - 1) + 0;

centroid[i][j] = data[idx][j];

}

}

}

// Caclulate centroidDists

private void calcDist(String method) {

**for** (int z = 0; z < k; z++) {

**for** (int i = 0; i < rowLength; i++) {

double d = 0.0;

**if** (method == "euclidean") {

**for** (int j = 0; j < colLength; j++)

d += Math.pow(data[i][j] - centroid[z][j], 2);

d = Math.floor(Math.sqrt(d) \* 100.0)/ 100.0;

} **else** **if** (method == "manhattan") {

**for** (int j = 0; j < colLength; j++)

d += (data[i][j] - centroid[z][j]);

d = Math.abs(Math.floor(d \* 100.0)/ 100.0);

}

centroidDist[i][z] = d;

}

}

}

// Assign closer points into cluster

private void assignCluster() {

boolean localConverge = **true**;

**for** (int i = 0; i < rowLength; i++) {

double min = centroidDist[i][0];

int minColIdx = 0;

**for**(int j = 0; j < k; j++) {

**if** (min > centroidDist[i][j]) {

min = centroidDist[i][j];

minColIdx = j;

}

}

**if** (cluster[i] != minColIdx) {

cluster[i] = minColIdx;

localConverge = **false**;

}

}

**if** (localConverge)

globalConverge = **true**;

}

// Cacluate mean between points w/ same cluster

private void calcCentroid() {

sumIntraDist = 0;

**for** (int z = 0; z < k; z++) {

double [][] temp = **new** double[rowLength][colLength];

int tmpRowLength = 0;

**for** (int j = 0; j < colLength; j++) {

int idxI = 0;

double sum = 0.0;

double mean = 0.0;

double counter = 0.0;

**for** (int i = 0; i < rowLength; i++) {

**if** (cluster[i] == z) {

sum += data[i][j];

counter++;

temp[idxI][j] = data[i][j]; // store points in each cluster

idxI++;

}

}

mean = sum/counter;

centroid[z][j] = Math.floor(mean \* 100.0)/ 100.0; // assigned new centroid pts

tmpRowLength = idxI;

}

calcIntraDist(temp, method, tmpRowLength, z); // calculate intra cluster dist

System.out.println();

}

}

// Cacluate intra cluster distance between points w/ same cluster

private void calcIntraDist(double[][] temp, String method, int tmpRowLength, int z) {

double min = 10\*100;

double max = 0;

double sum = 0;

System.out.println("Intra-Cluster distances in cluster " + z);

// finding intra cluster min, max, sum

**for** (int i = 0; i < tmpRowLength; i++) {

double d = 0.0;

**if** (method == "euclidean") {

**for** (int j = 0; j < colLength; j++)

d += Math.pow(temp[i][j] - centroid[z][j], 2);

d = Math.floor(Math.sqrt(d) \* 100.0)/ 100.0;

} **else** **if** (method == "manhattan") {

**for** (int j = 0; j < colLength; j++)

d += (data[i][j] - centroid[z][j]);

d = Math.abs(Math.floor(d \* 100.0)/ 100.0);

}

**if** (min > d) min = d;

**if** (max < d) max = d;

sum += d;

}

sum = Math.floor(sum \* 100.0)/ 100.0;

System.out.println("Min : " + min + ", Max : " + max + ", Sum : " + sum);

sumIntraDist += sum;

sumIntraDist = Math.floor(sumIntraDist \* 100.0)/ 100.0;

}

}

**Plot.R**

library(ggplot2)

setwd("E:/Google Drive/Class/4342/proj1")

data1 <- read.csv("euclidean2.csv")

data2 <- read.csv("euclidean4.csv")

data3 <- read.csv("manhattan2.csv")

data4 <- read.csv("manhattan4.csv")

ggplot(data1, aes(attribute0, attribute1)) + geom\_point(aes(colour = factor(Cluster)), size = 10, alpha = 7/10)

ggsave(file="euclidean2.png")

ggplot(data2, aes(attribute0, attribute1)) + geom\_point(aes(colour = factor(Cluster)), size = 10, alpha = 7/10)

ggsave(file="euclidean4.png")

ggplot(data3, aes(attribute0, attribute1)) + geom\_point(aes(colour = factor(Cluster)), size = 10, alpha = 7/10)

ggsave(file="manhattan2.png")

ggplot(data4, aes(attribute0, attribute1)) + geom\_point(aes(colour = factor(Cluster)), size = 10, alpha = 7/10)

ggsave(file="manhattan4.png")