Unsupervised learning

Given information:

Total instances: 77.

Total attribute: 12

Missing values: 4

@ATTRIBUTE 'calories' numeric

@ATTRIBUTE 'protein' numeric

@ATTRIBUTE 'fat' numeric

@ATTRIBUTE 'sodium' numeric

@ATTRIBUTE 'fiber' numeric

@ATTRIBUTE 'carbo' numeric

@ATTRIBUTE 'sugars' numeric

@ATTRIBUTE 'potass' numeric

@ATTRIBUTE 'vitamins' numeric

@ATTRIBUTE 'shelf' numeric

@ATTRIBUTE 'rating' numeri

Preprocessing: After getting the data, I converted the data into csv and looked for missing values. There were some missing values in the data set.

6^{th} instance -> potash

22^{th} -> potash

59^{th} -> carbo

59^{th} -> sugar

The missing data values were replaced with the average value of those instances in the data set. The average value of potash, carbo and sugar are 99,13,7. The csv file was converted in to .arff to import in weka.

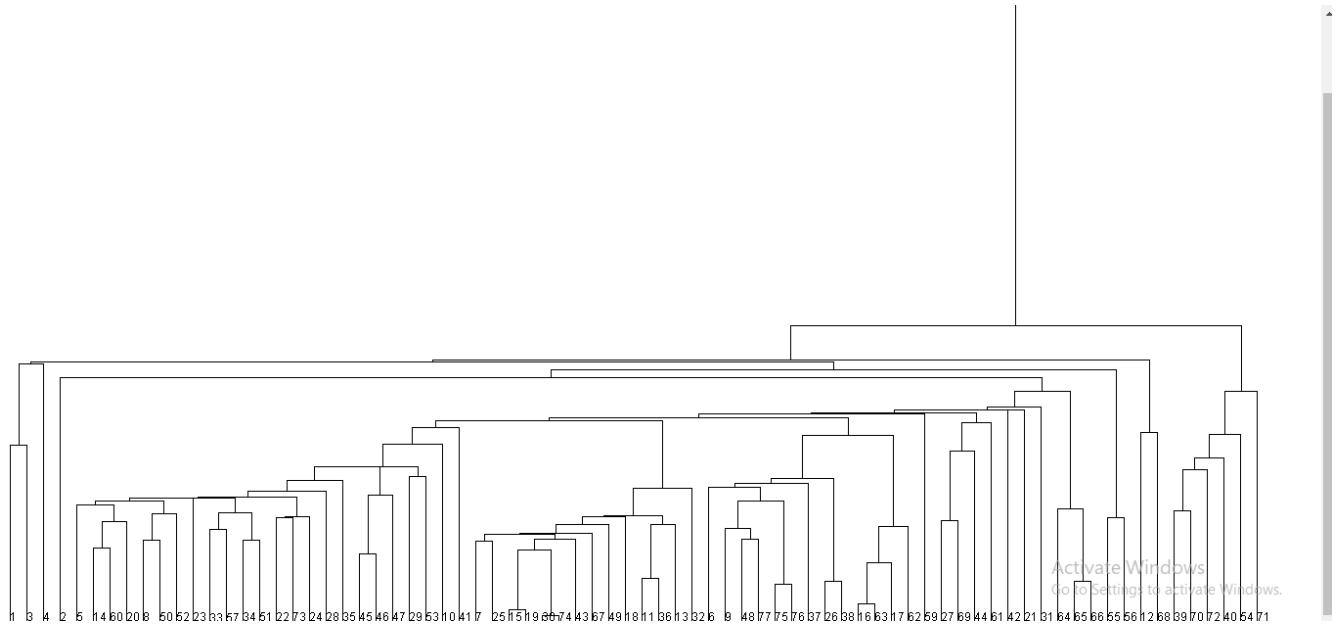

Fig: before cutting point

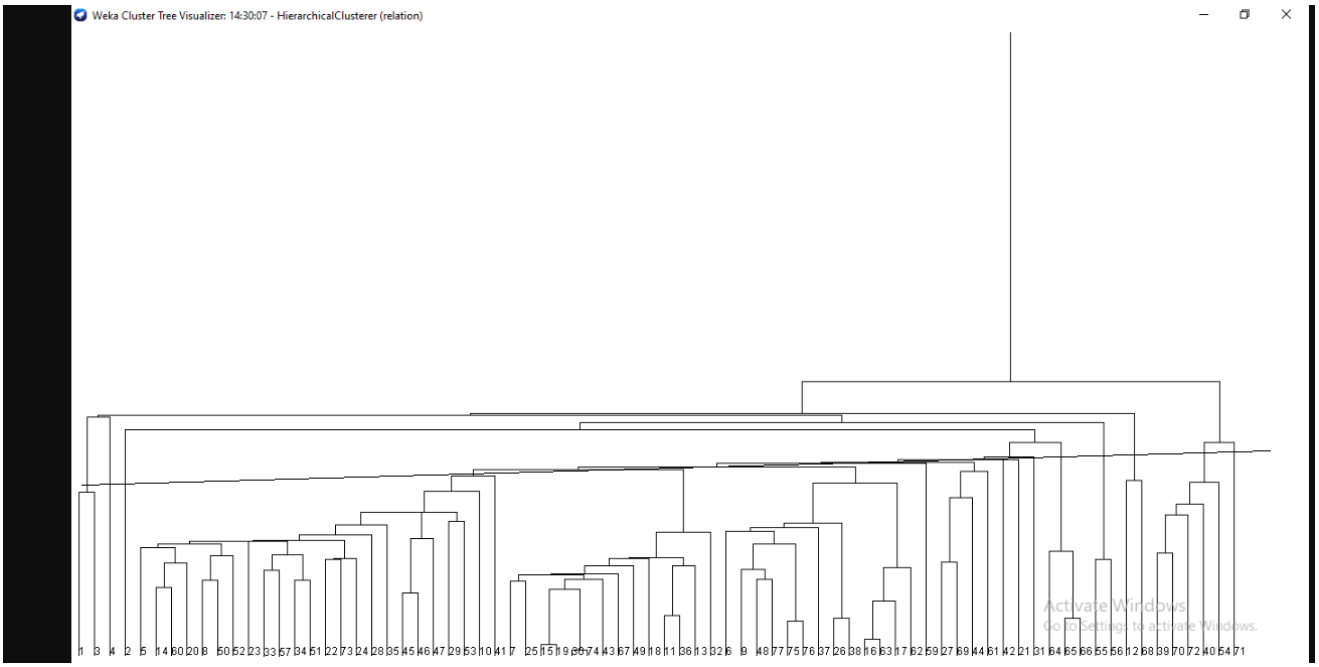

Fig: Dendogram of the cereal data set

After getting the data set we arbitrarily selected a cutting point and we get 14 clusters from the dendogram. The detailed of each cluster is given below.

Cluster 1(row :1,3):

name	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
100% Bran	70	4	1	130	10	5	6	280	25	3	68.40297
All-Bran	70	4	1	260	9	7	5	320	25	3	59.42551

Cluster 2(row :4):

name	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
All-Bran with Extra Fiber	50	4	0	140	14	8	0	330	25	3	93.70491

Cluster 3(row :2):

name	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
100% Natural Bran	120	3	5	15	2	8	8	135	0	3	33.98368

Cluster 4(row :5 ,15,60,20,8,50,52,23,33,57,34,51,22 ,73,24 ,28 ,35,46,47, 29,53,10,41)

name	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Almond Delight	110	2	2	200	1	14	8	99	25	3	34.38484
Cocoa Puffs	110	1	1	180	0	12	13	55	25	2	22.73645
Raisin Nut Bran	100	3	2	140	2.5	10.5	8	140	25	3	39.7034

Cracklin' Oat Bran	110	3	3	140	4	10	7	160	25	3	40.44877
Basic 4	130	3	2	210	2	18	8	100	25	3	37.03856
Nutri-Grain Almond-Raisin	140	3	2	220	3	21	7	130	25	3	40.69232
Oatmeal Raisin Crisp	130	3	2	170	1.5	13.5	10	120	25	3	30.45084
Crispy Wheat & Raisins	100	2	1	140	2	11	10	120	25	3	36.1762
Grape Nuts Flakes	100	3	1	140	3	15	5	85	25	3	52.0769
Quaker Oat Squares	100	4	1	135	2	14	6	110	25	3	49.51187
Grape-Nuts	110	3	0	170	3	17	3	90	25	3	53.37101
Nutri-grain Wheat	90	3	0	170	3	18	2	90	25	3	59.64284
Crispix	110	2	0	220	1	21	3	30	25	3	46.89564
Triples	110	2	1	250	0	21	3	60	25	3	39.10617
Double Chex	100	2	0	190	1	18	5	80	25	3	44.33086
Fruit & Fibre	120	3	2	160	5	12	10	200	25	3	40.91705
Great Grains Pecan	120	3	3	75	3	13	4	100	25	3	45.81172
Great Grains Pecan	120	3	3	75	3	13	4	100	25	3	45.81172
Muesli Raisins; Peaches; & Pecans	150	4	3	150	3	16	11	170	25	3	34.13977
Mueslix Crispy Blend	160	3	2	150	3	17	13	160	25	3	30.31335
Fruitful Bran	120	3	0	240	5	14	12	190	25	3	41.01549
Post Nat. Raisin Bran	120	3	1	200	6	11	14	260	25	3	37.84059

Bran Flakes	90	3	0	210	5	13	5	190	25	3	53.31381
Kix	110	2	1	260	0	21	3	40	25	2	39.24111

Cluster 5(row :7,25,15,19,30,74,43,67,49,18,11,36,13,32)

name	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Apple Jacks	110	2	0	125	1	11	14	30	25	2	33.17409
Froot Loops	110	2	1	125	1	11	13	30	25	2	32.20758
Cocoa Puffs	110	1	1	180	0	12	13	55	25	2	22.73645
Count Chocula	110	1	1	180	0	12	13	65	25	2	22.39651
Fruity Pebbles	110	1	1	135	0	13	12	25	25	2	28.02577
Trix	110	1	1	140	0	13	12	25	25	2	27.7533
Lucky Charms	110	2	1	180	0	12	12	55	25	2	26.73452
Smacks	110	2	1	70	1	9	15	40	25	2	31.23005
Nut&Honey Crunch	120	2	1	190	0	15	9	40	25	2	29.92429
Corn Pops	110	1	0	90	1	13	12	20	25	2	35.78279
Cap'n'Crunch	120	1	2	220	0	12	12	35	25	2	18.04285
Honey Graham Ohs	120	1	2	220	1	12	11	45	25	2	21.87129
Cinnamon Toast Crunch	120	1	3	210	0	13	9	45	25	2	19.82357
Golden Grahams	110	1	1	280	0	15	9	45	25	2	23.80404

Cluster 6(row: 6,9,48,7,75,76,37,26,38,16,63,17,62)

name	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Apple Cinnamon Cheerios	110	2	2	180	1.5	10.5	10	70	25	1	29.50954
Bran Chex	90	2	1	200	4	15	6	125	25	1	49.12025
Multi-Grain Cheerios	100	2	1	220	2	15	6	90	25	1	40.10597
Apple Jacks	110	2	0	125	1	11	14	30	25	2	33.17409
Wheat Chex	100	3	1	230	3	17	3	115	25	1	49.78745
Wheaties	100	3	1	200	3	17	3	110	25	1	51.59219
Honey Nut Cheerios	110	3	1	250	1.5	11.5	10	90	25	1	31.07222
Frosted Flakes	110	1	0	200	1	14	11	25	25	1	31.43597
Honey Nut Cheerios	110	3	1	250	1.5	11.5	10	90	25	1	31.07222
Corn Chex	110	2	0	280	0	22	3	25	25	1	41.44502
Rice Krispies	110	2	0	290	0	22	3	35	25	1	40.56016
Corn Flakes	100	2	0	290	1	21	2	35	25	1	45.86332
Rice Chex	110	1	0	240	0	23	2	30	25	1	41.99893

Cluster 7(row: 59)

name	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating

Raisin Bran	120	3	1	210	5	14	12	240	25	2	39.2592

Cluster 8(row: 27,69,44,61,42,21,31)

name	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Corn Flakes	100	2	0	290	1	21	2	35	25	1	45.86332
Strawberry Fruit Wheats	90	2	0	15	3	15	5	90	25	2	59.36399
Maypo	100	4	1	0	0	16	3	95	25	2	54.85092
Raisin Squares	90	2	0	0	2	15	6	110	25	3	55.33314
Life	100	4	2	150	2	12	6	95	25	2	45.32807
Cream of Wheat (Quick)	100	3	0	80	1	21	0	99	0	2	64.53382
Golden Crisp	100	2	0	45	0	11	15	40	25	1	35.25244

Cluster 9(row: 31)

name	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Golden Crisp	100	2	0	45	0	11	15	40	25	1	35.25244

Cluster 10(row :64,65,66)

name	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Shredded Wheat	80	2	0	0	3	16	0	95	0	1	68.23589

Shredded Wheat 'n'Bran	90	3	0	0	4	19	0	140	0	1	74.47295
Shredded Wheat spoon size	90	3	0	0	3	20	0	120	0	1	72.80179

Cluster 11(row :55,56)

name	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Puffed Rice	50	1	0	0	0	13	0	15	0	3	60.75611
Puffed Wheat	50	2	0	0	1	10	0	50	0	3	63.00565

Cluster 12(row :12,68)

name	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Cheerios	110	6	2	290	2	17	1	105	25	1	50.765
Special K	110	6	0	230	1	16	3	55	25	1	53.13132

Cluster 13(row :39,70,72,40,54)

name	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Just Right Crunchy Nuggets	110	2	1	170	1	17	6	60	100	3	36.52368
Total Corn Flakes	110	2	1	200	0	21	3	35	100	3	38.83975
Total Whole Grain	100	3	1	200	3	16	3	110	100	3	46.65884

Just Right Fruit & Nut	140	3	1	170	2	20	9	95	100	3	36.47151
Product 19	100	3	0	320	1	20	3	45	100	3	41.50354

Cluster 14(row :71)

name	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Total Raisin Bran	140	3	1	190	4	15	14	230	100	3	28.59279

Questions:

1. Question: Is a strong correlation between dietary fiber and potassium?

=>

There is a strong correlation between fiber and potassium in this data set. The correlation coefficient amongst these two is 0.903374. Also in the below picture we can see the positioning of the points are proportional.

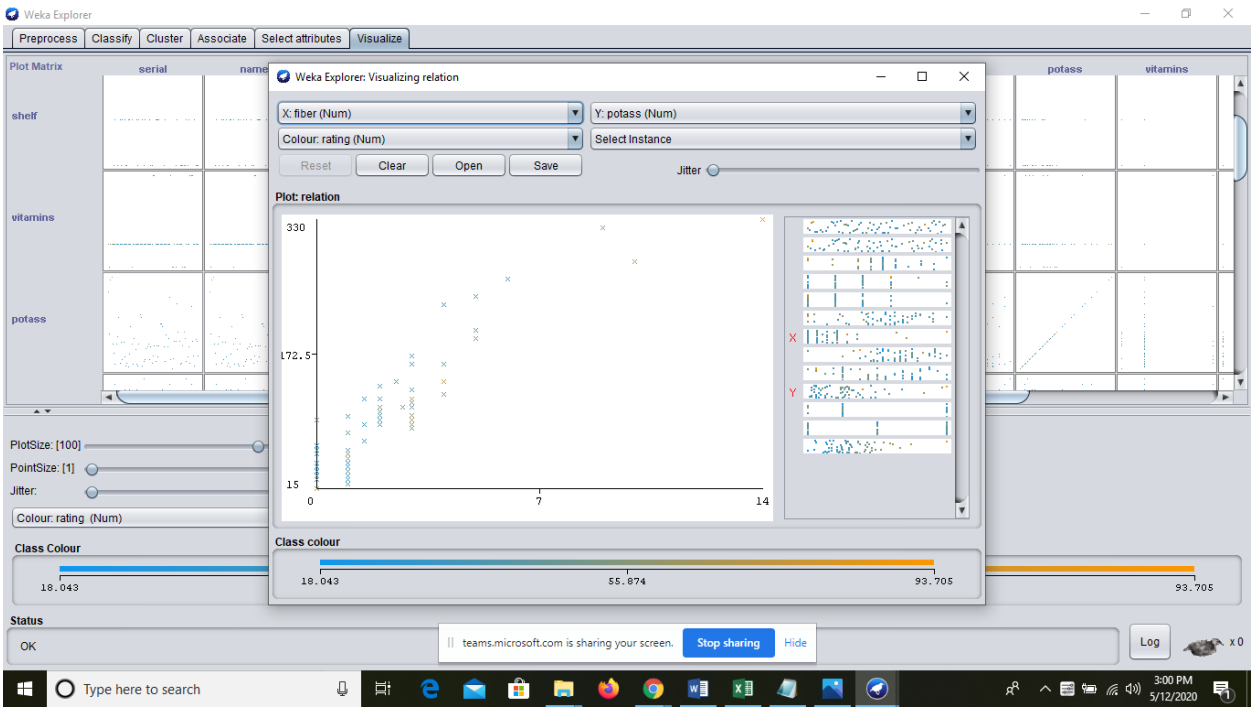

Fig : correlation between fiber and potassium

2. Question: Are groups of cereals from which we can choose according to our preferences?

   ⇨ There are total 14 clusters in after choosing the cutting point. I tried to make a pattern out of the clusters and found a suitable group of people to be consumers of a particular cluster.

Considering: for calories 1. 50 to 80 -> low calories

2. 90 – 110 -> medium calories

3.120-160 -> high calories

Protein

1. **1,2 -> low**
2. **3 - > medium**
3. **4 ->High**

Fat

1. **0 -> no**

2. **1 -> low**
3. **2,3 -> medium**
4. **4-5 -> high**

Sodium

1. **0 -> no**
2. **15 to 140 -> low**
3. **150 – 200 -> medium**
4. **210-290 -> high**

Fiber

1. **0 -> no**
2. **1 to 3 -> low**
3. **4 to 6 -> medium**
4. **7 to 14 -> high**

Carbo

1. **5 to 14 -> low**
2. **15 to 20 -> medium**
3. **21 to 23 -> high**

Sugar

1. **0 -> no**
2. **1 to 6 -> low**
3. **7 to 12 -> medium**
4. **12 to 15 -> high**

Potash

1. **20 to 100 -> low**
2. **105 to 190 -> medium**
3. **200 to 330 -> high**

Vitamins:

1. **0 -> low**
2. **25 -> medium**

100 -> high

Cluster Analysis:

Cluster No	No of instance	Id number	Result after analysis	Evaluation
1	2	1,3	calories: low, protein: high, fat: low, sodium: mid, fiber: high, carbo: low, sugar: low, potassium: high, vitamin: low	Suited for children as they need more protein ,fiber .high potassium ensures muscle build and nerve system.
2	1	4	calories: low, protein: high, fat: low, sodium: high, fiber: high, carbo: low, sugar: low, potassium: high, vitamin: mid	Better for body building program. Low calories, carbo, sugar ensures no further fat gain. Fiber ensures better digestion, potassium helps to build muscle construction. Also highest rated cereal .
3	1	2	calories: low, protein: high, fat: low, sodium: low high, fiber: high, carbo: low, sugar: low, potassium: high, vitamin: mid	Better for people with heart problem and high blood pressure because it has more fiber, low carbo ,sodium and sugar.
4	23	5,8,10,14,20,22,2324,28, 29,33,34,35,45,46,47,50, 51,52,53,57,60,73	calories: mid, protein: mid, fat: low, sodium: mid, fiber: low, carbo: mid, sugar: mid, potassium: mid, vitamin: low	Adult with no heart or blood pressure.
5	14	**7,25,15,19,30,74, 43,67,49,18,**	calories: mid, protein: low, fat:	Better for persons with

		11,36,13,32	low, sodium: mid, fiber: no, carbo: mid, sugar: mid, potassium: low, vitamin: mid	digestion problem like constipation and bloating issues. People with these issues cannot manage to consume fiber and potassium yet they should get balanced diet of carbo ,vitamin and calories
6	13	**6,9,48,7,75,76,37, 26,38,16,63,17,62**	calories: mid, protein: low, fat: no, sodium: high, fiber: low, carbo: mid, sugar: low, potassium: mid, vitamin: low	Adults with bone density problem as it has high sodium, protein(un)
7	1	59	calories: high, protein: mid, fat: low, sodium: high, fiber: mid, carbo: mid, sugar: mid, potassium: high, vitamin: low	This can be a good breakfast for a pregnant woman because they need fibers and fiber and they should take more calories than usual. Also high sodium,potas is preferable in these times
8	6	**17,69,44,61,42,21**	calories: mid, protein: low, fat: no, sodium: low, fiber: Low, carbo: mid, sugar: low, potassium: no, vitamin: low	For weight loss program
9	1	**31**	calories: high, protein: low, fat: no, sodium: mid,	No decision

			fiber: mid, carbo: low, sugar: high, potassium: low, vitamin: mid	
10	3	**64,65,66**	calories: mid, protein: mid, fat: no, sodium: no, fiber: low, carbo: mid, sugar: no, potassium: mid, vitamin: no	Good rating, popular amongst consumers so kept in shelf 1
11	2	**55,56**	calories: low, protein: low, fat: no, sodium: no, fiber: low, carbo: low, sugar: no, potassium: low, vitamin: no	Better for weight losing program as low calories, carbo, sugar helps to is necessary for burning existing fat
12	2	**12,68**	calories: mid, protein: high, fat: mid, sodium: high, fiber: low, carbo: mid, sugar: low, potassium: mid vitamin: low	No decision
13	5	**39,70,72,40,54**	calories: mid, protein: high, fat: low, sodium: mid, fiber: low, carbo: mid, sugar: low, potassium: low, vitamin: high	Can be used for Old people as they need high vitamins and a balanced amount of sodium, calories, carbo. Also they should take low sugar and potassium reduces any problem related to heart and blood pressure
14	1	**71**	calories: high, protein: mid, fat: low, sodium: mid,	Better suited for people with low blood

			fiber: mid, carbo: mid, sugar: high, potassium: high, vitamin: high	pressure because they need high potassium increases blood volume. Increased salt, sugar is not harmful for them. Also they need lots of vitamin B12 and relatively low carbohydrate.

3. See other correlation between the data given in the files.
=>
Let's suppose:
A= calories
B = protein
C=fat
D= sodium
E =Fiber
F = carbo
G = Sugar
H= potash
I = vitamin
J=shelf

Instances	Correlation value	comment
A &B	0.019066	Value is almost 0 .No correlation at all
A & C	0.49861	Almost 0.5 which means moderately correlated
A & D	0.300649	Very week correlation
A & E	-0.29341	Negative correlation .So if one attribute decreases the other increases and vice versa
A & F	0.250681	Very week correlation
A & G	0.56234	moderately correlated
A & H	-0.07206	No correlation at all
A & I	0.265356	Very week
A & J	0.097234	Very week
B&C	0.208431	Very week
B&D	-0.05467	Very week
B&E	0.50033	moderately correlated
B & F	-0.13086	Very week
B & G	-0.32914	Very week
B & H	0.563706	moderately correlated
B& I	0.007335	Very week
B& J	0.13386	Very week
C & D	-0.00541	No correlation
C & E	0.016719	No correlation
C & F	-0.31804	Very week negative correlation
C & G	0.270819	Very week
C & H	0.200445	Very week
C & I	-0.03116	No correlation
C & J	0.263691	Very week
D&E	-0.07068	No correlation
D&F	0.355983	Very week correlation
D&G	0.101451	No correlation at all
D&H	-0.04263	No correlation
D&I	0.361477	Very week correlation

D&J	-0.06972	No correlation
E &D	-0.07068	No correlation
E&G	-0.14121	No correlation at all
E&H	0.911528	Strong correlation
E&I	-0.03224	No correlation
E&J	0.297539	Very week
F&G	-0.33167	Very week
F&H	-0.3383	Very week
F&I	0.258148	Very week
F&J	-0.10179	No correlation at all
G &H	-0.0026	No correlation
G & I	0.125137	No correlation at all
G & J	0.100438	No correlation at all
H&I	-0.00543	No correlation
H&J	0.385784	Very week correlation
J & I	0.29926	Very week correlation