Name:Islam,Rafiqul   Id:17-34438-1   Sec:A
Unsupervised Learning (with vitamin and rating)

Project definition:

In this project we used a dataset of cereals along with their different percentage of calories, protein, fat etc. Besides, we did hierarchical clustering for finding out who needs which type of diet or food.

Literature Survey:

(With 'vitamin' and 'rating' columns): 77 x 11 here:

http://www.cs.umd.edu/hcil/hce/examples/cereal/cereal-updated.txt

The meaning of each column:

1. 1st column: Name of cereal

 2. Calories: calories per serving

 3. Protein: grams of protein

4. Fat: grams of fat

 5. Sodium: milligrams of sodium

 6. Fiber: grams of dietary fiber

 7. Carbo: grams of complex carbohydrates

 8. Sugars: grams of sugars

9. Potassium: milligrams of potassium

10. Vitamins: vitamins and minerals - 0, 25, or 100, indicating the typical percentage of FDA recommended

  11. Shelf: display shelf (1, 2, or 3, counting from the floor)

 12. Rating: a rating of the cereals (calculated by Consumer Reports)

Method:

The dataset has been clustered by the hierarchical clustering technique. The cluster tree has been cut in several places. Then similarities between instances of individual clusters and dissimilarities between instances of different clusters have been analyzed.

Hierarchical cluster tree with cutting point:

=== Run information ===

Scheme: weka.clusterers.HierarchicalClusterer -N 2 -L SINGLE -P -A "weka.core.EuclideanDistance -R first-last"

Relation: WithVitamin

Instances: 75

Attributes: 12

        Name

        calories

        protein

        fat

        sodium

        fiber

        carbo

        sugers

        potassium

vitamins

        shelf

        rating

Test mode: evaluate on training data

=== Clustering model (full training set) ===

Cluster 0

(((((1:0.45956,3:0.45956):0.21047,4:0.67003):0.00706,(((2:0.63675,((((((((5:0.34946,((8:0.24314,(46:0.21537,75:0.21537):0.02777):0.07118,(73:0.09803,74:0.09803):0.2163):0.03513):0.01125,35:0.36071):0.0113,(24:0.10622,36:0.10622):0.26578):0.11218,(((15:0.04606,61:0.04606):0.10861,16:0.15467):0.09347,60:0.24814):0.23605):0.04578,(((((((((6:0.20993,23:0.20993):0.01771,(((14:0.03206,18:0.03206):0.15431,(28:0.01603,72:0.01603):0.17034):0.02973,41:0.2161):0.01153):0.0032,65:0.23084):0.02228,47:0.25311):0.02191,17:0.27503):0.00006,((10:0.11214,34:0.11214):0.14118,12:0.25332):0.02177):0.07193,30:0.34702):0.17681,(((((((((((7:0.21278,48:0.21278):0.06881,50:0.28159):0.03189,((13:0.19164,58:0.19164):0.06953,19:0.26118):0.05231):0.00829,21:0.32177):0.00038,((31:0.24174,55:0.24174):0.0414,(32:0.21347,49:0.21347):0.06966):0.03902):0.00223,((20:0.27122,71:0.27122):0.002,22:0.27322):0.05116):0.01553,26:0.33991):0.02772,33:0.36763):0.03526,((43:0.17638,44:0.17638):0.15406,45:0.33044):0.07245):0.00072,(27:0.37831,51:0.37831):0.02529):0.05821,9:0.46182):0.04317,39:0.50499):0.01885):0.00614):0.01013,57:0.5401):0.00472,(((25:0.2627,67:0.2627):0.18199,42:0.44469):0.07307,59:0.51776):0.02706):0.00696,40:0.55179):0.0068,29:0.55859):0.07816):0.01932,(53:0.27195,54:0.27195):0.38412):0.0045,(62:0.29323,(63:0.10657,64:0.10657):0.18666):0.36734):0.01652):0.00506,(11:0.49375,66:0.49375):0.18839):0.08807,(((((37:0.28993,68:0.28993):0.10692,70:0.39685):0.03047,38:0.42733):0.0605,52:0.48783):0.11208,69:0.59991):0.1703)

Time taken to build model (full training data) : 0.09 seconds

=== Model and evaluation on training set ===

Clustered Instances

0   74 (99%)

1   1 (1%)
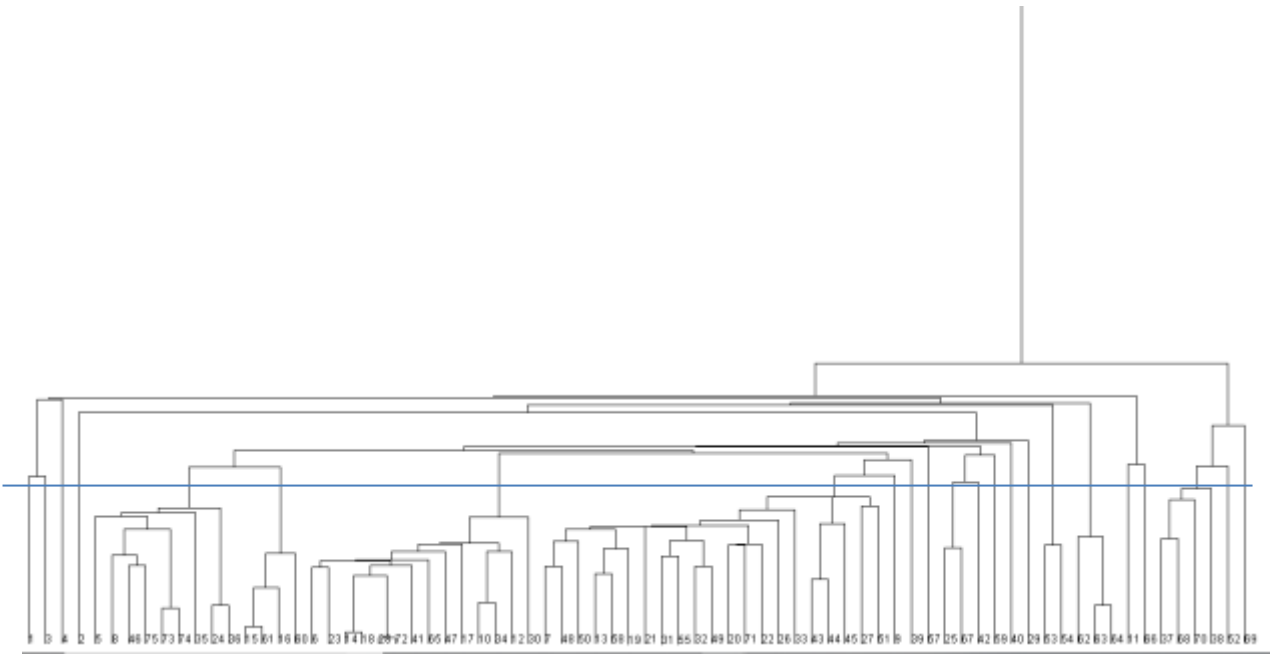
Fig 1: Hierarchical cluster tree with cutting point

Cluster Analysis:

For Cluster-1(1 instances):

Cereals	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

Findings:Calories(low),protein(medium),fat(low),sodium(low),fiber(high),carbo(low),sugars(low),potass(high),vitamin(low),rating(medium)

For Cluster- 2(1 instances):

Cereals	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
All-Bran	70	4	1	260	9	7	5	320	25	3	59.42551

Findings:Calories(low),protein(medium),fat(low),sodium(medium),fiber(high),carbo(medium),sugars(low),potass(high),vitamin(low),shelf(high),rating(medium)

For Cluster -3(1 instances):

Cereals	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

Findings:Calories(low),protein(medium),fat(0),sodium(low),fiber(high),carbo(low),sugars(0),potass(high),vitamin(low),rating(high),shelf(high)

For Cluster -4(1 instances):

Cereals	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

Findings:Calories(high),protein(medium),fat(high),sodium(low),fiber(low),sugars(medium),potass(low),vitamin(0),rating(low),shelf(high)

For Cluster -5(9 instances):

Cereals	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Apple_	110	2	2	180	1.5	10.5	10	70	25	1	29.5095

Cereals	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Cinnamon_Cheerios											4
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
Wheaties_Honey_Gold	110	2	1	200	1	16	8	60	25	1	36.18756
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-comb	110	1	0	180	0	14	11	35	25	1	28.74241

Findings:Calories(high),protein(medium),fat(low),sodium(medium),fiber(low),carbo(high),sugars(high),potass(low),vitamin(low),rating(medium),shelf(low)

For Cluster- 6 (4 instances):

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

Findings:Calories(high),protein(low),fat(0),sodium(high),fiber(low),carbo(high),sugars(low),potass(low),vitamin(low),rating(medium),shelf(low)

For Cluster -7(14 instances):

Cereals	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	27.75335
Maypo	100	4	1	0	0	16	3	95	25	2	54.85092
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

Findings:Calories(high),protein(medium),fat(low),sodium(high),fiber(low),carbo(medium),sugars(high),potass(low),vitamin(low),rating(medium),shelf(medium)

For Cluster -8(21 inatances):

Cereals	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Basic_4	130	3	2	210	2	18	8	100	25	3	37.038

											56
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
Clusters	110	3	2	140	2	13	7	105	25	3	40.40021
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
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_Dates,_Walnuts,_and_Oats	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
Muesli	150	4	3	95	3	16	11	170	25	3	37.136

Cereals	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
_Raisins,_Dates,_&_Almonds											86
Muesli_Raisins,_Peaches,_&_Pecans	150	4	3	150	3	16	11	170	25	3	34.13977
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
Mueslix_Crispy_Blend	160	3	2	150	3	17	13	160	25	3	30.31335

Findings:Calories(high),protein(medium),fat(medium),sodium(medium),fiber(low),carbo(high),sugars(high),potass(medium),vitamin(low),rating(medium),shelf(high)

For Cluster -9(1 instances):

Cereals	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Bran_Flakes	90	3	0	210	5	13	5	190	25	3	53.31381

Findings:Calories(medium),protein(medium),fat(0),sodium(medium),fiber(high),carbo(medium),sugars(low),potass(medium),vitamins(low),rating(medium),shel(high)

For Cluster -10(1 instances):

Cereals	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Kix	110	2	1	260	0	21	3	40	25	2	39.24111

Findings:Calories(high),protein(low),fat(low),sodium(high),fiber(0),carbo(high),sugars(low),potass(low),vitamins(low),rating(low),shelf(medium)

For Cluster- 11(1 instances):

Cereals	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

Findings:Calories(high),protein(medium),fat(low),sodium(medium),fiber(high),carbo(medium),sugars(high),potass(high),vitamin(low),rating(medium),shelf(medium)

For Cluster -12(2 instances):

Cereals	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Frosted_Mini-Wheats	100	3	0	0	3	14	7	100	25	2	58.34514
Strawberry_Fruit_Wheats	90	2	0	15	3	15	5	90	25	2	59.36399

Findings:Calories(medium),protein(medium),fat(0),sodium(low),fiber(medium),carbo(medium),sugars(medium),potass(low),vitamin(low),rating(medium),shelf(medium)

For Cluster -13(1 instances):

Cereals	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Maypo	100	4	1	0	0	16	3	95	25	2	54.85092

Findings:Calories(high),protein(medium),fat(low),sodium(0),fiber(0),carbo(medium),sugars(low),potass(low),vitamin(low),rating(medium),shelf(medium)

For Cluster -14(1 instances):

Cereals	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Raisin_ Square s	90	2	0	0	2	15	6	110	25	3	55.33314

Findings:Calories(medium),protein(low),fat(0),sodium(0),fiber(low),carbo(medium),sugars(low),potass(low),vitamin(low),rating(medium),shel(high)

For Cluster -15(1 instances):

Cereals	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Life	100	4	2	150	2	12	6	95	25	2	45.32807

Findings:Calories(high),protein(medium),fat(low),sodium(medium),fiber(low),carbo(medium),sugars(low),potass(low),vitamin(low),rating(medium),shelf(medium)

For Cluster -16(1 instances):

Cereals	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

Findings:Calories(high),protein(low),fat(0),sodium(low),fiber(0),carbo(medium),sugars(high),potass(low),vitamin(low),rating(low),shelf(low)

For Cluster -17(2 instances):

Cereals	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

Findings:Calories(low),protein(medium),fat(0),sodium(0),fiber(low),carbo(medium),sugars(0),potass(low),vitamin(low),rating(medium),shelf(high)

For Cluster -18(3 instances):

Cereals	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_sp oon_size	90	3	0	0	3	20	0	120	0	1	72.80179

Findings:Calories(medium),protein(medium),fat(0),sodium(0),fiber(medium),carbo(0),sugars(0),potass(low),vitamin(low),rating(medium),shelf(low)

For Cluster- 19(1 instances):

Cereals	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Cheerios	110	6	2	290	2	17	1	105	25	1	50.765

Findings:Calories(high),protein(high),fat(low),sodium(high),fiber(low),carbo(medium),sugars(low),potass(low),vitamin(low),rating(medium),shelf(low)

For Cluster -20(1 instances):

Cereals	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Special_ K	110	6	0	230	1	16	3	55	25	1	53.13132

Findings:Calories(high),protein(hith),fat(0),sodium(hith),fiber(low),carbo(medium),sugars(low),potass(low),vitamin(low),rating(medium),shelf(low)

For Cluster- 21(4 instances):

Cereals	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

Findings:Calories(high),protein(medium),fat(low),sodium(medium),fiber(medium),carbo(high),sugars(low),potass(medium),vitamin(high),rating(low),shelf(high)

For Cluster -22(1 instances):

Cereals	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
Product_19	100	3	0	320	1	20	3	45	100	3	41.50354

Findings:Calories(high),protein(medium),fat(0),sodium(high),fiber(low),carbo(20),sugars(low),potass(low),vitamin(high),rating(medium),shelf(high)

For Cluster -23(1 instances):

Cereals	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

Findings:Calories(high),protein(medium),fat(low),sodium(medium),fiber(medium),carbo(medium),sugars(high),potass(high),vitamin(high),rating(low),shelf(high)

Question:

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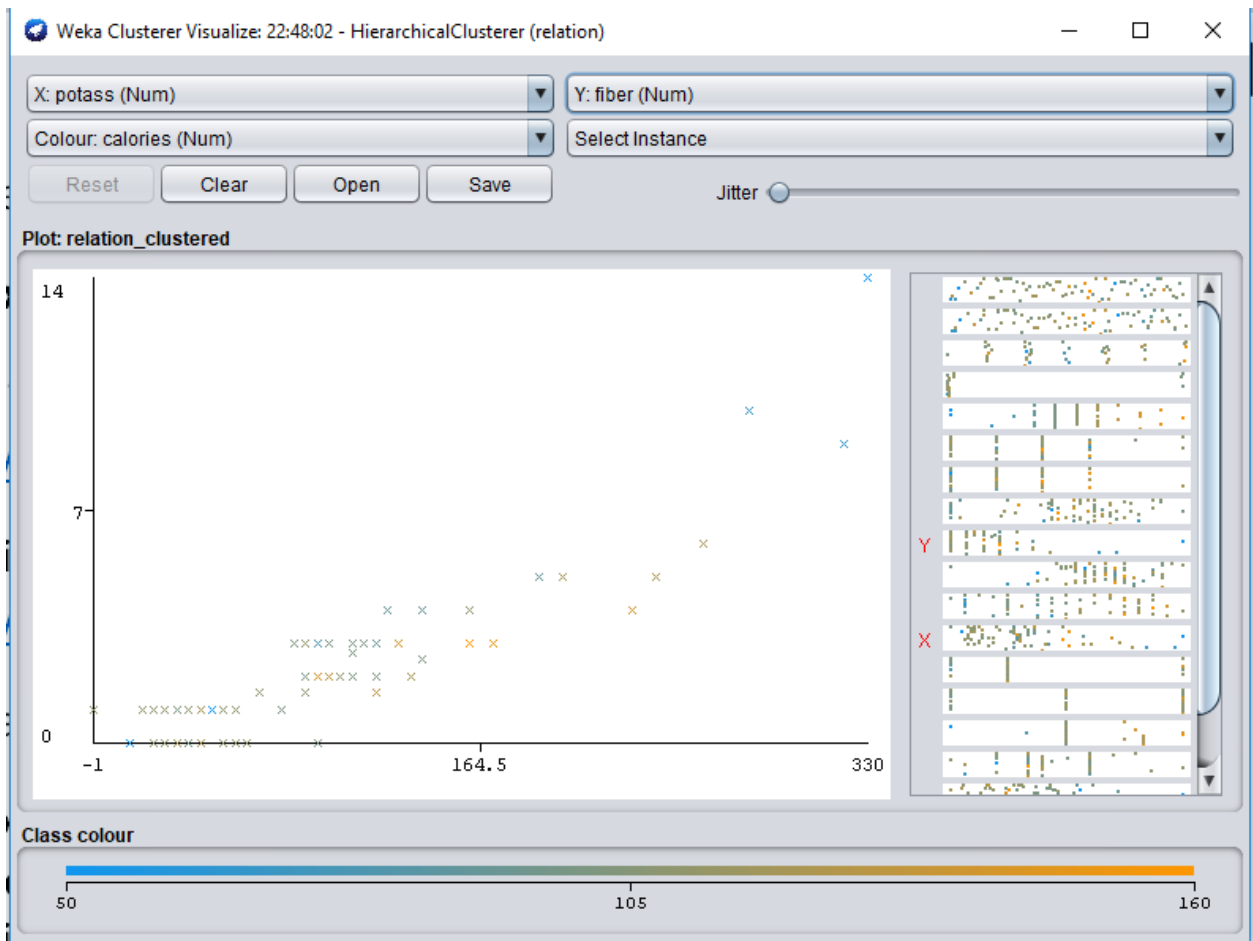
   Ans:

Figure1: Correlation between Dietary fiber(y) and potassium(x)

From figure1, we can see if potassium is increasing then dietary fiber is also increasing, so that Dietary fiber and potassium have a strong correlation between them.

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

Ans: Yes. From (1-23) clusters which are given above, anyone can get any combination of he/she likes such as-

➢ Anyone who wants to lose weight should consume low fat, low sugar and low carbohydrate. They can choose from cluster

➢ Anyone who wants to gain weight can choose from cluster

➢ Anyone with high diabetes can choose from cluster

➢ Anyone who has high blood pressure should avoid sodium and can choose from cluster

➢ Anyone who has low blood pressure can choose from

➢ Anyone in need of high protein and potassium can choose from cluster

3. See other correlation between the data given in the files.

Ans: Other correlation are given below-

There is correlation between Calories and Carbohydrates. If the value of Calories goes higher then there is a possibility of higher carbohydrate value. This confirms low carbohydrate value for low calories.
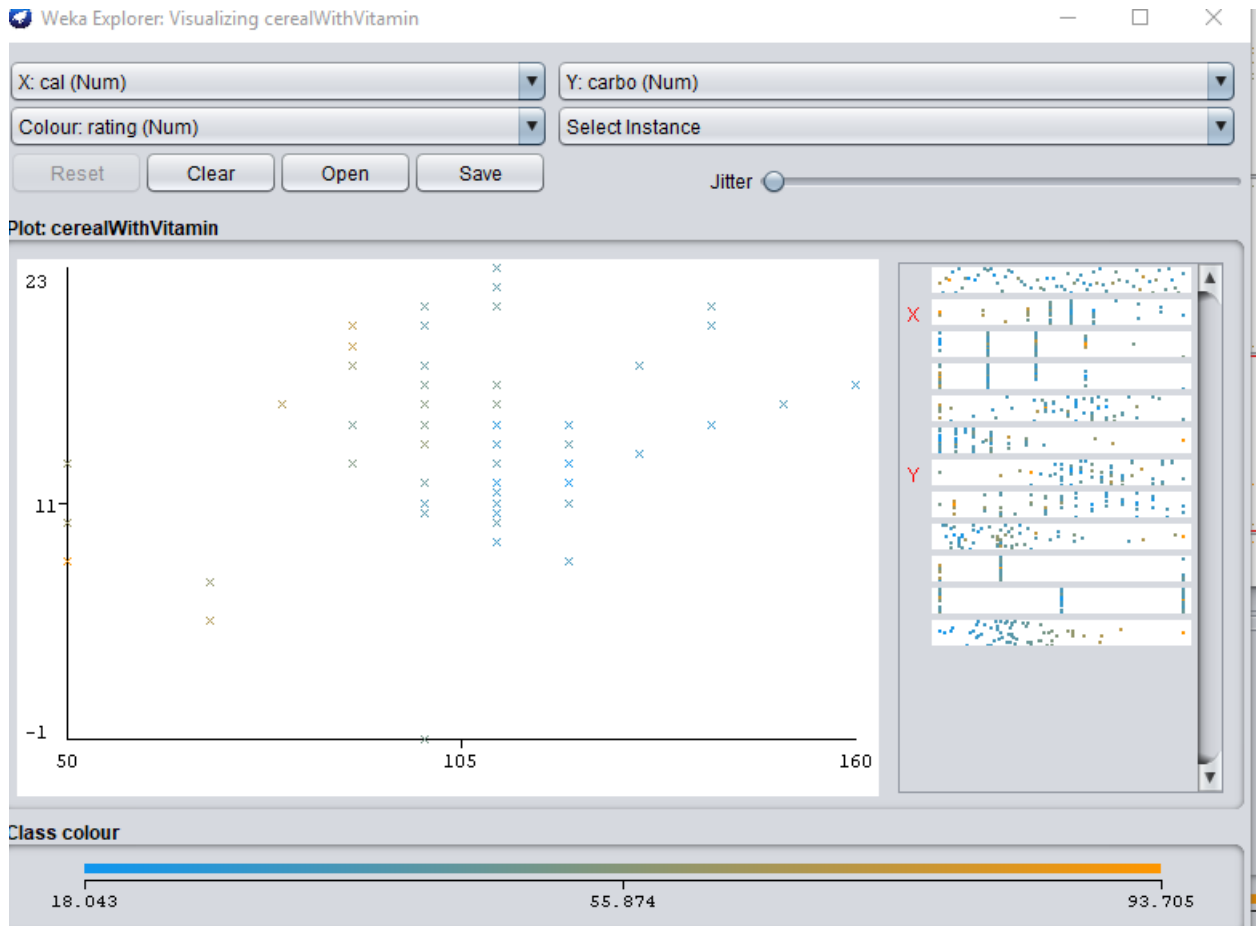

Figure2: Correlation between calories(x) and Carbohydrates (y)