Εξόρυξη δεδομένων με χρήση τεχνικών μηχανικής μάθησης

Τμ. Μηχανικών Πληροφορικής και Υπολογιστών

Παράδειγμα

K-means

Αλγόριθμος k-means

Βήματα Αλγορίθμου:

1. Ορίζουμε την τιμή του k.
2. Κάνουμε δειγματοληψία αρχίζοντας από κάποιο δείγμα (υποσύνολο) των παρατηρήσεων που περιλαμβάνονται στο σύνολο δεδομένων.
3. Ο αλγόριθμος εντάσσει αρχικά τις παρατηρήσεις τυχαία σε k-συστάδες.
4. Υπολογίζει για τις παρατηρήσεις των συστάδων του δείγματος την κεντροειδή τους, δηλαδή τους μέσους/μέσους όρους για κάθε χαρακτηριστικό τους. Κάθε συστάδα έχει την κεντροειδή της που μπορεί να ανήκει ή όχι στις παρατηρήσεις.

Αλγόριθμος k-means

Βήματα Αλγορίθμου:

5. Στη συνέχεια, συγκρίνει τα χαρακτηριστικά όλων των παρατηρήσεων του δείγματος με τους μέσους και ανάλογα με τις αποστάσεις εντάσσει εκ νέου τις παρατηρήσεις σε συστάδες. Υπολογίζει εκ νέου τις κεντροειδείς.
6. Η διαδικασία αυτή γίνεται επανειλημμένα προκειμένου ο αλγόριθμος να «κινηθεί κυκλικά» γύρω από τα καλύτερα ταιριάσματα (matches) και στη συνέχεια να διαμορφώσει τις συστάδες των παρατηρήσεων.

Αλγόριθμος k-means (συν.)

Λαμβάνουμε υπόψη τα εξής:

- Όλες οι παρατηρήσεις αντιστοιχίζονται στην πλησιέστερη συστάδα. Η έννοια της πλησιέστερης εξαρτάται από το μέτρο απόστασης που χρησιμοποιείται.
- Ο αλγόριθμος σταματάει όταν οι κεντροειδείς δεν αλλάζουν («δεν μετακινούνται») ή όταν έχει εκτελεστεί ο προκαθορισμένος μέγιστος αριθμός βημάτων (max optimization steps) π.χ., 10 επαναλήψεις.
- Η διαδικασία επαναλαμβάνεται κάθε φορά με ένα διαφορετικό σύνολο σημείων εκκίνησης, δηλαδή κεντροειδών. Στο σύνολο των συστάδων στο οποίο καταλήγει ο αλγόριθμος, όλες οι παρατηρήσεις μιας συστάδας έχουν την ελάχιστη απόσταση από την κεντροειδή τους.

Παράδειγμα k-means

παράδειγμα: k-means cluster data mining model (χρήση RapidMiner)

Εργάζεστε σε ασφαλιστική εταιρεία. Στόχος σας είναι να εντοπίσετε και στη συνέχεια να προσπαθήσετε να προσεγγίσετε τα άτομα που ασφαλίζονται από την εταιρεία σας και τα οποία διατρέχουν υψηλό κίνδυνο στεφανιαίας νόσου εξαιτίας του βάρους τους και / ή της υψηλής χοληστερόλης. Καταλαβαίνετε ότι όσοι βρίσκονται σε χαμηλό κίνδυνο, δηλαδή άτομα με χαμηλό βάρος και χοληστερόλη, είναι απίθανο να συμμετάσχουν στα προγράμματα που θα προσφέρετε. Γνωρίζετε, επίσης, ότι υπάρχουν πιθανώς τρεις ακόμη περιπτώσεις:

- ασφαλισμένοι με υψηλό βάρος και χαμηλή χοληστερόλη,
- ασφαλισμένοι με υψηλό βάρος και υψηλή χοληστερόλη,
- ασφαλισμένοι με χαμηλό βάρος και υψηλή χοληστερόλη.

Καταλαβαίνετε ακόμη ότι είναι πιθανό να υπάρχουν πολλοί ασφαλισμένοι κάπου ενδιάμεσα. Προκειμένου να επιτύχετε το στόχο σας, πρέπει να αναζητήσετε μεταξύ των χιλιάδων κατόχων ασφαλιστηρίων συμβολαίων και να βρείτε ομάδες ατόμων με παρόμοια χαρακτηριστικά και να εκπονήσετε ασφαλιστικά προγράμματα και στρατηγικές επικοινωνίας συναφείς και ελκυστικές για τους ανθρώπους σε αυτές τις διαφορετικές ομάδες.

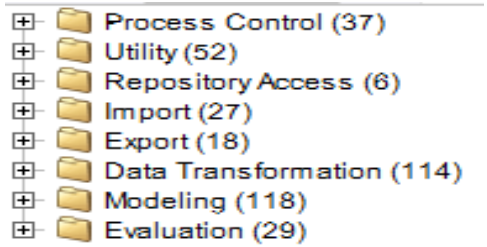
(M. North, Data Mining for the Masses, Διασκευή παραδείγματος κεφαλαίου 6)

https://sites.google.com/site/dataminingforthemasses/

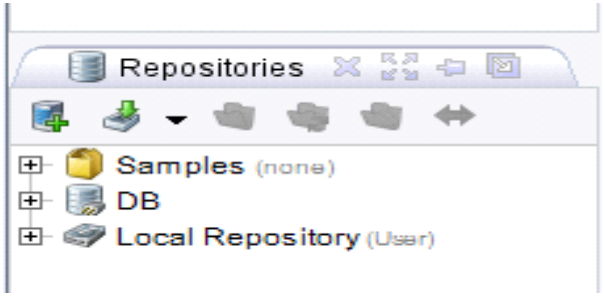
Weight,Cholesterol,Gender
102,111,1
115,135,1
115,136,1
140,167,0
130,158,1
198,227,1
114,131,1
145,176,0
191,223,0

Ορολογία εργαλείου Rapid Miner

Operators

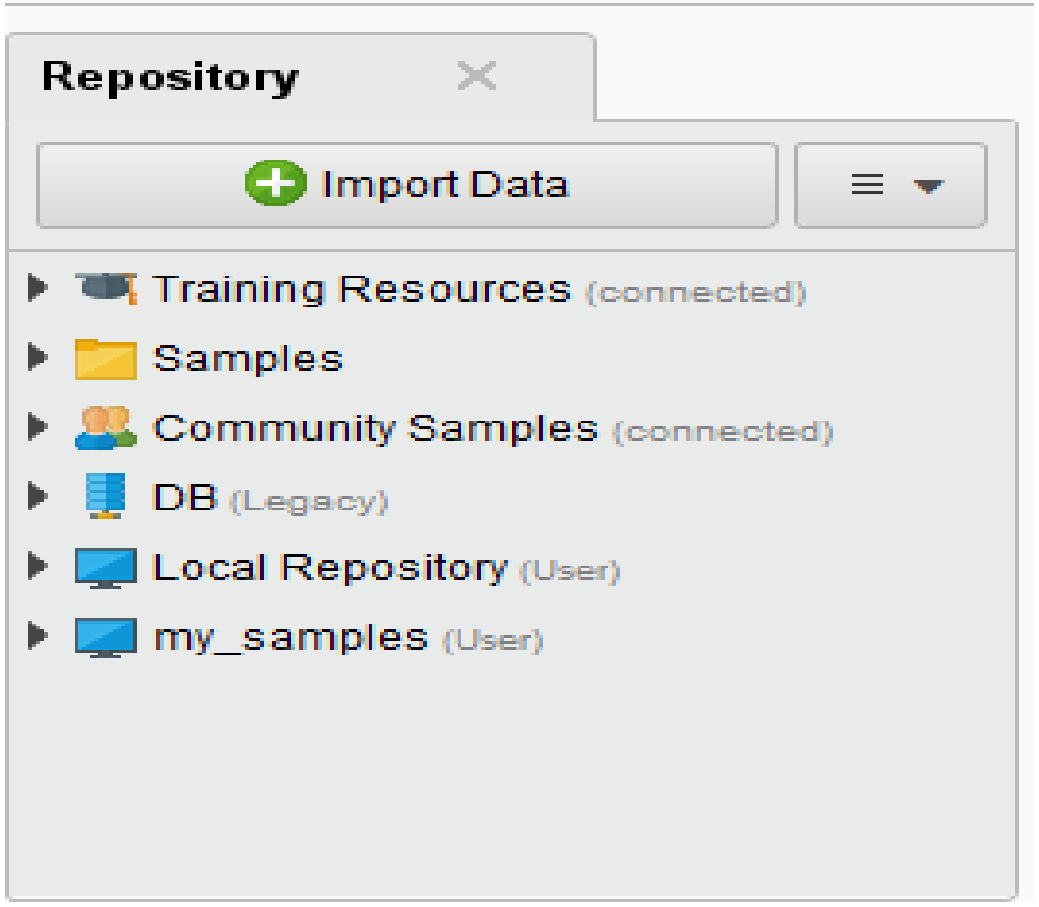


Repositories

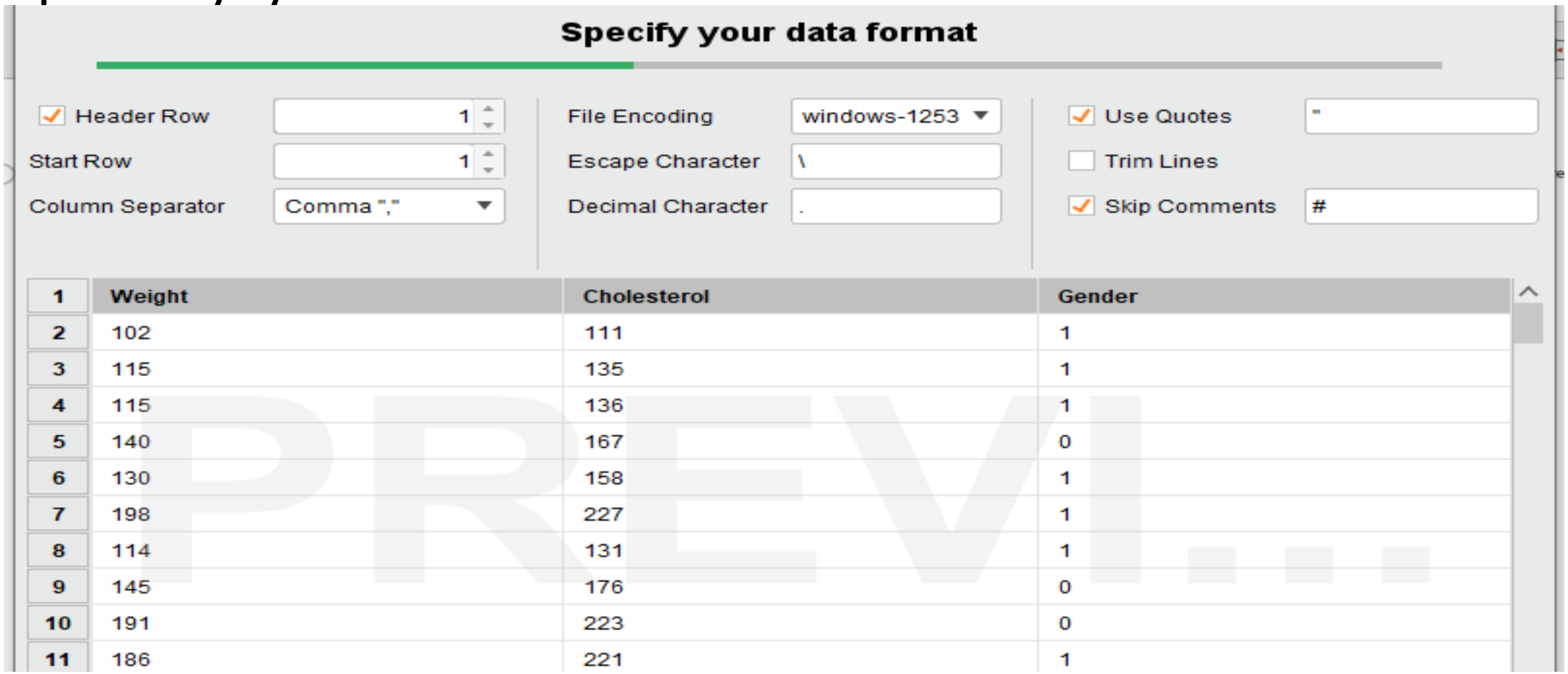


Θα πρέπει να δημιουργήσουμε ένα local repository

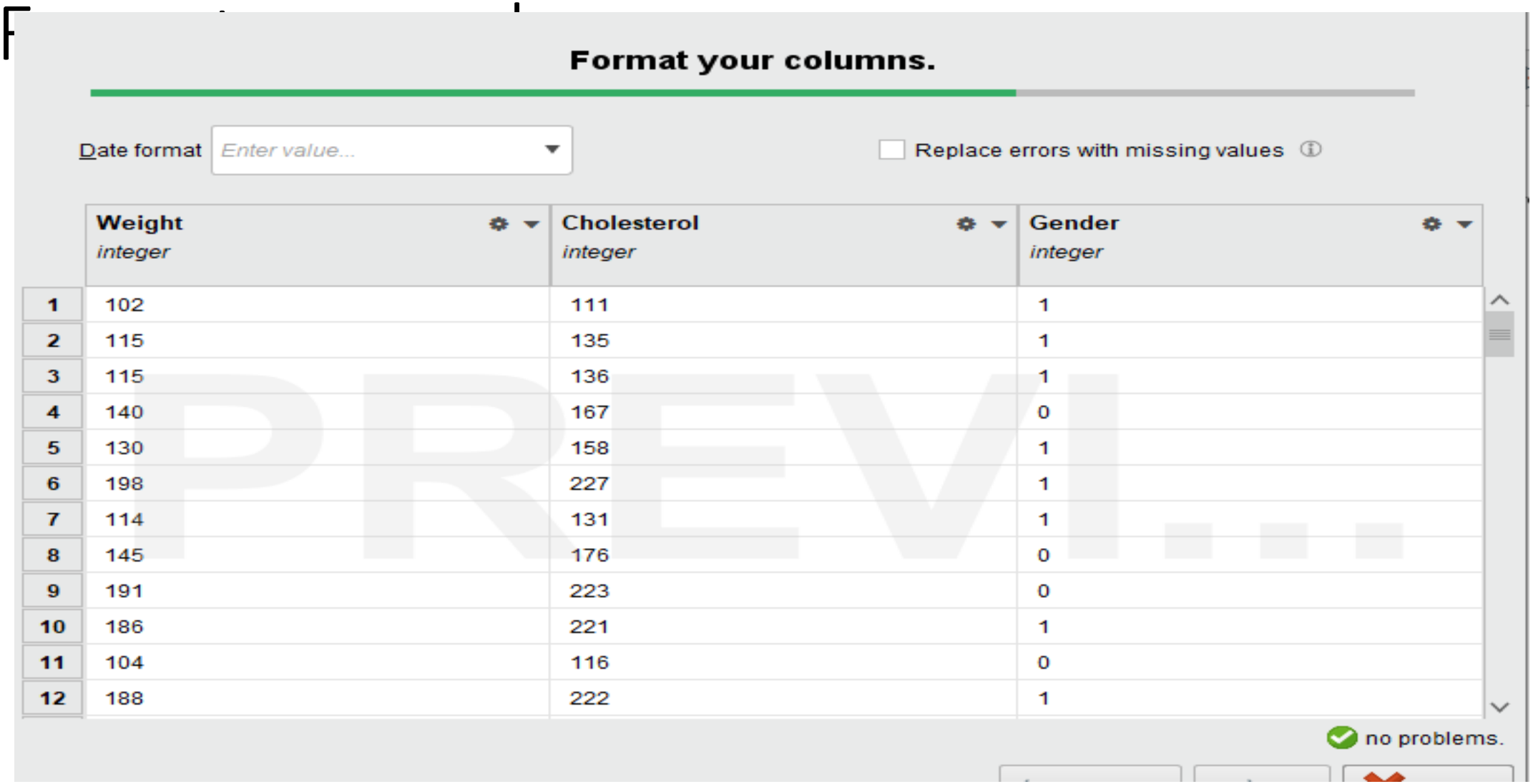
Import Data Set



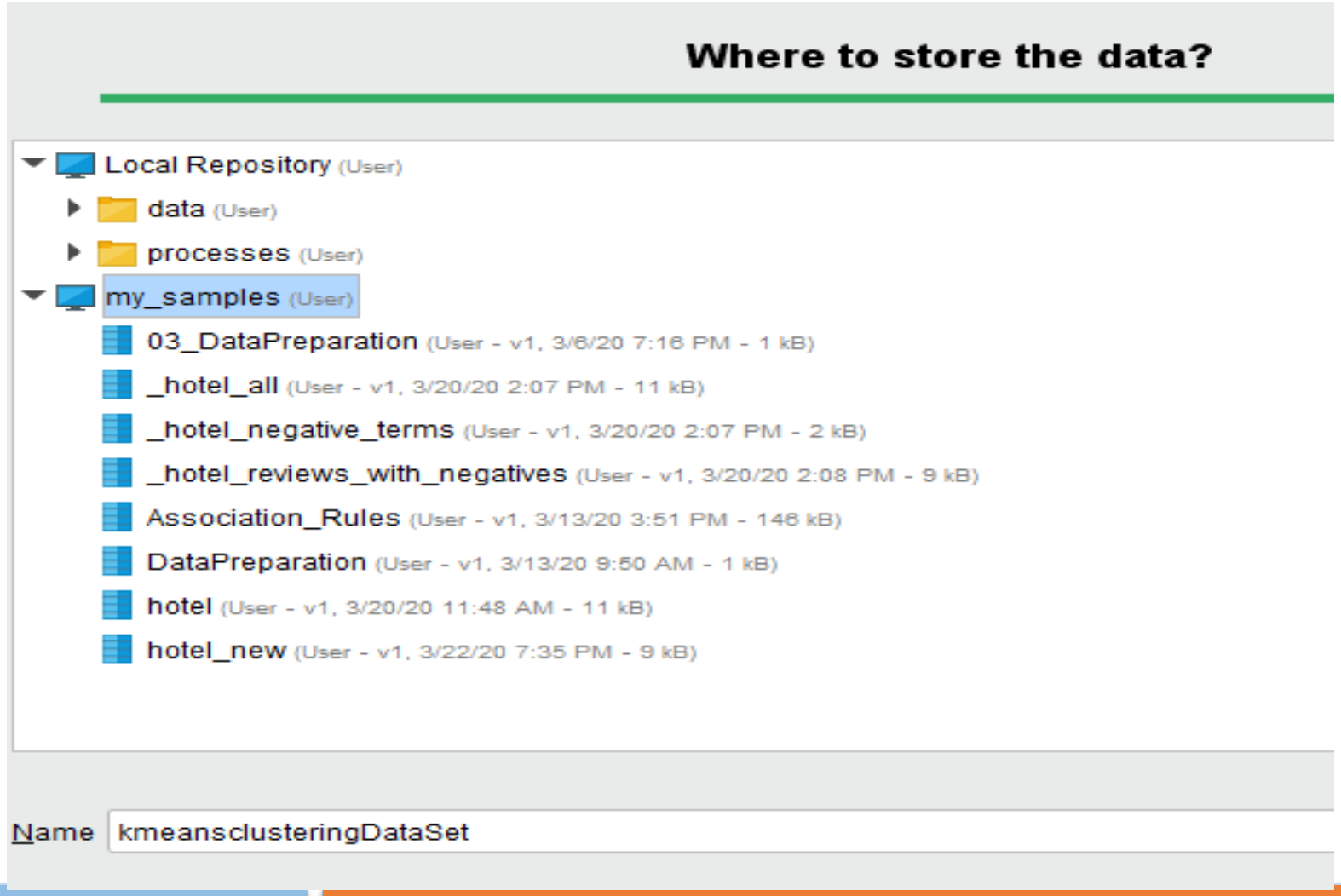
Specify your data format



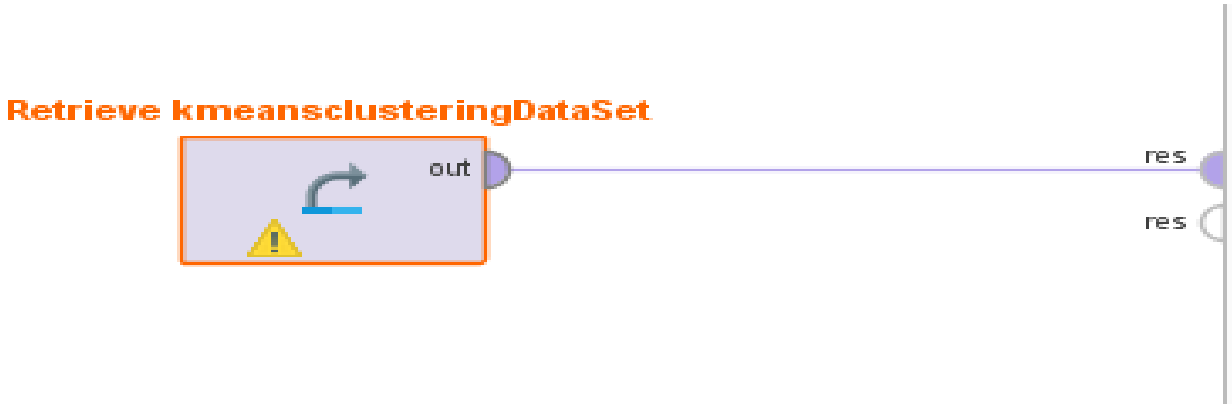
1	Weight	Cholesterol	Gender
2	102	111	1
3	115	135	1
4	115	136	1
5	140	167	0
6	130	158	1
7	198	227	1
8	114	131	1
9	145	176	0
10	191	223	0
11	186	221	1

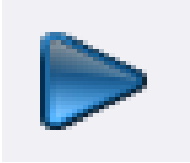

Format your columns.
Date format
Enter value...
Replace errors with missing values
Weight
integer
Cholesterol
integer
Gender
integer
1 102 111 1
2 115 135 1
3 115 136 1
4 140 167 0
5 130 158 1
6 198 227 1
7 114 131 1
8 145 176 0
9 191 223 0
10 186 221 1
11 104 116 0
12 188 222 1
no problems.

Where to store



Process Windows (Design Perspective)





Data V

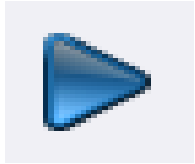
Row No.	Weight	Cholesterol	Gender
1	102	111	1
2	115	135	1
3	115	136	1
4	140	167	0
5	130	158	1
6	198	227	1
7	114	131	1
8	145	176	0
9	191	223	0
10	186	221	1
11	104	116	0
12	188	222	1
13	96	102	0
14	156	192	0
15	125	152	0

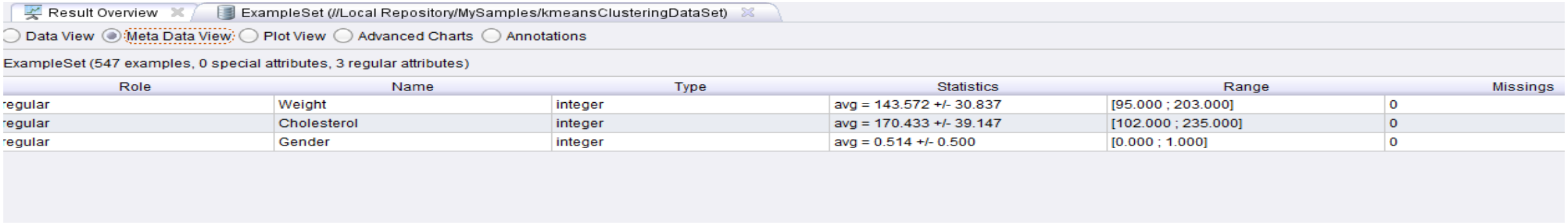
ExampleSet (547 examples, 0 special attributes, 3 regular attributes)

Meta Data View

Name	Type	Missing	Statistics			Filter (3 / 3 attributes): Search for Attributes
Weight	Integer	0	Min 95	Max 203	Average 143.572	
Cholesterol	Integer	0	Min 102	Max 235	Average 170.433	
Gender	Integer	0	Min 0	Max 1	Average 0.514	

Παλιό interface





Role	Name	Type	Statistics	Range	Missings
regular	Weight	integer	avg = 143.572 +/- 30.837	[95.000 ; 203.000]	0
regular	Cholesterol	integer	avg = 170.433 +/- 39.147	[102.000 ; 235.000]	0
regular	Gender	integer	avg = 0.514 +/- 0.500	[0.000 ; 1.000]	0

Οπτικοποίηση

Τρόποι Οπτικοποίησης των παρατηρήσεων

- Οπτικοποίηση μπορεί να γίνει μέσω πινάκων και διαγραμμάτων, π.χ., ιστογράμματα, ραβδογράμματα, διαγράμματα πίτας κ.ά.
- Διάκριση των παρατηρήσεων σε ποσοτικά ή ποιοτικά δεδομένα.

Οπτικοποίηση

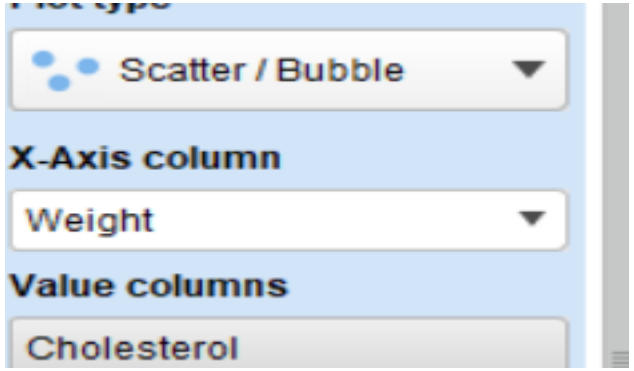
Οπτικοποίηση Ποσοτικών Δεδομένων

Ποσοτικά δεδομένα, τα οποία συνήθως παρουσιάζουμε με πίνακες, ιστογράμματα κ.λπ.

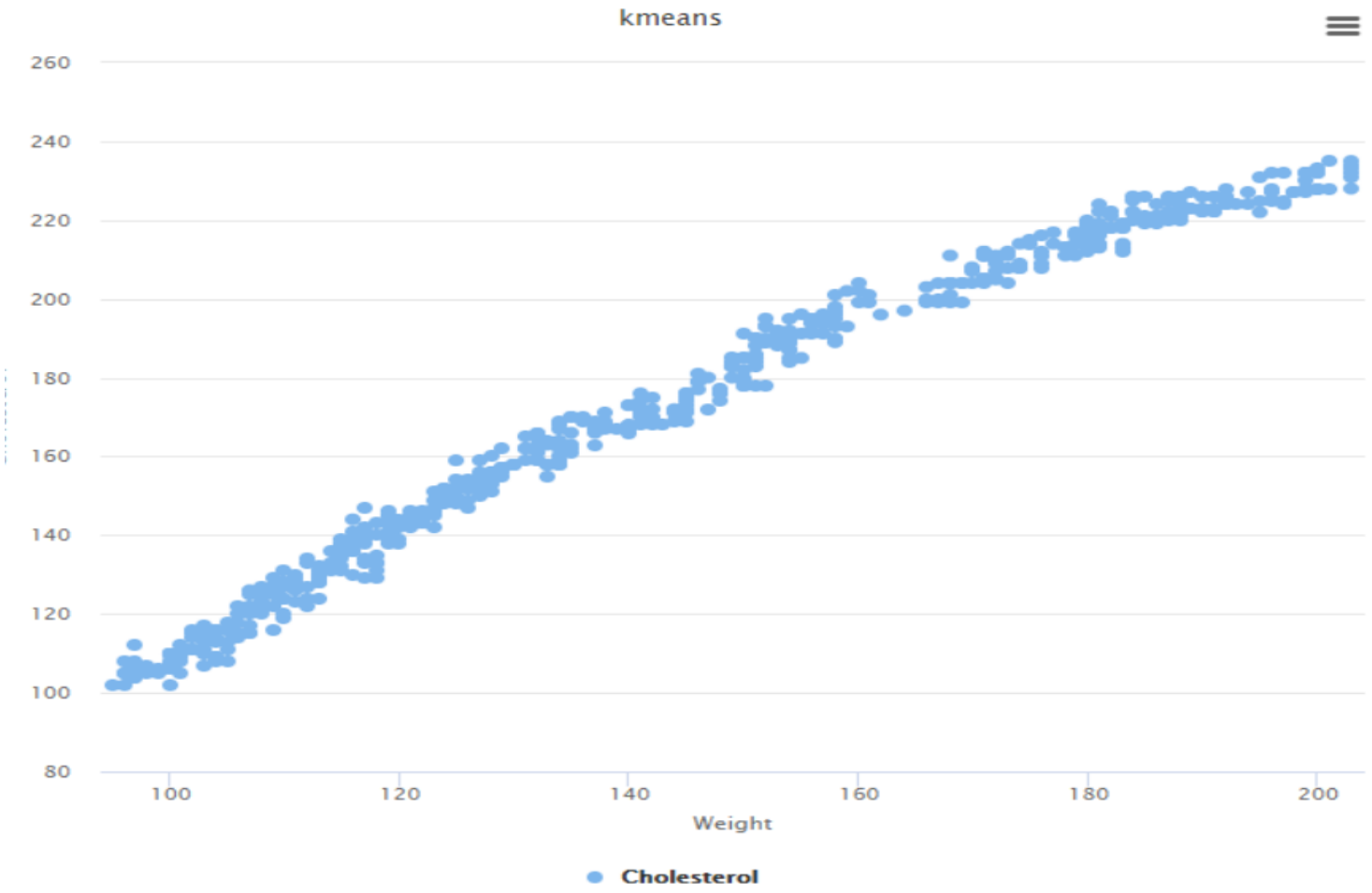
Οπτικοποίηση Ποιοτικών Δεδομένων

Ποιοτικά ή κατηγορικά δεδομένα, τα οποία συνήθως παρουσιάζουμε με πίνακες, ραβδογράμματα και διαγράμματα πίτας.

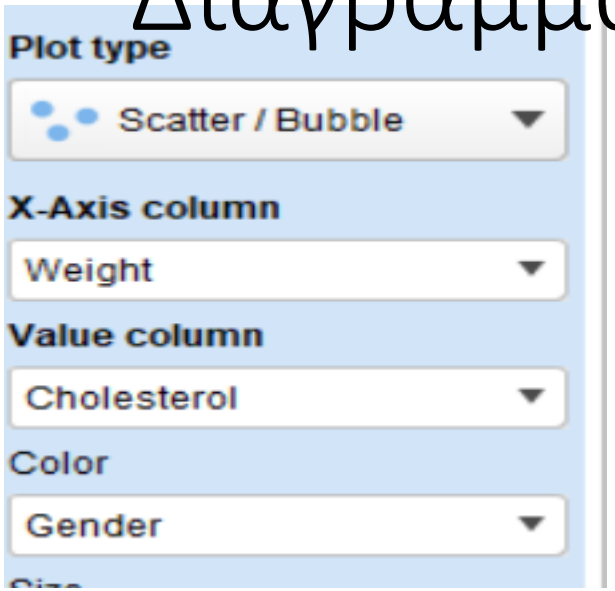
Διάγραμμα διασποράς (Scatter plot)

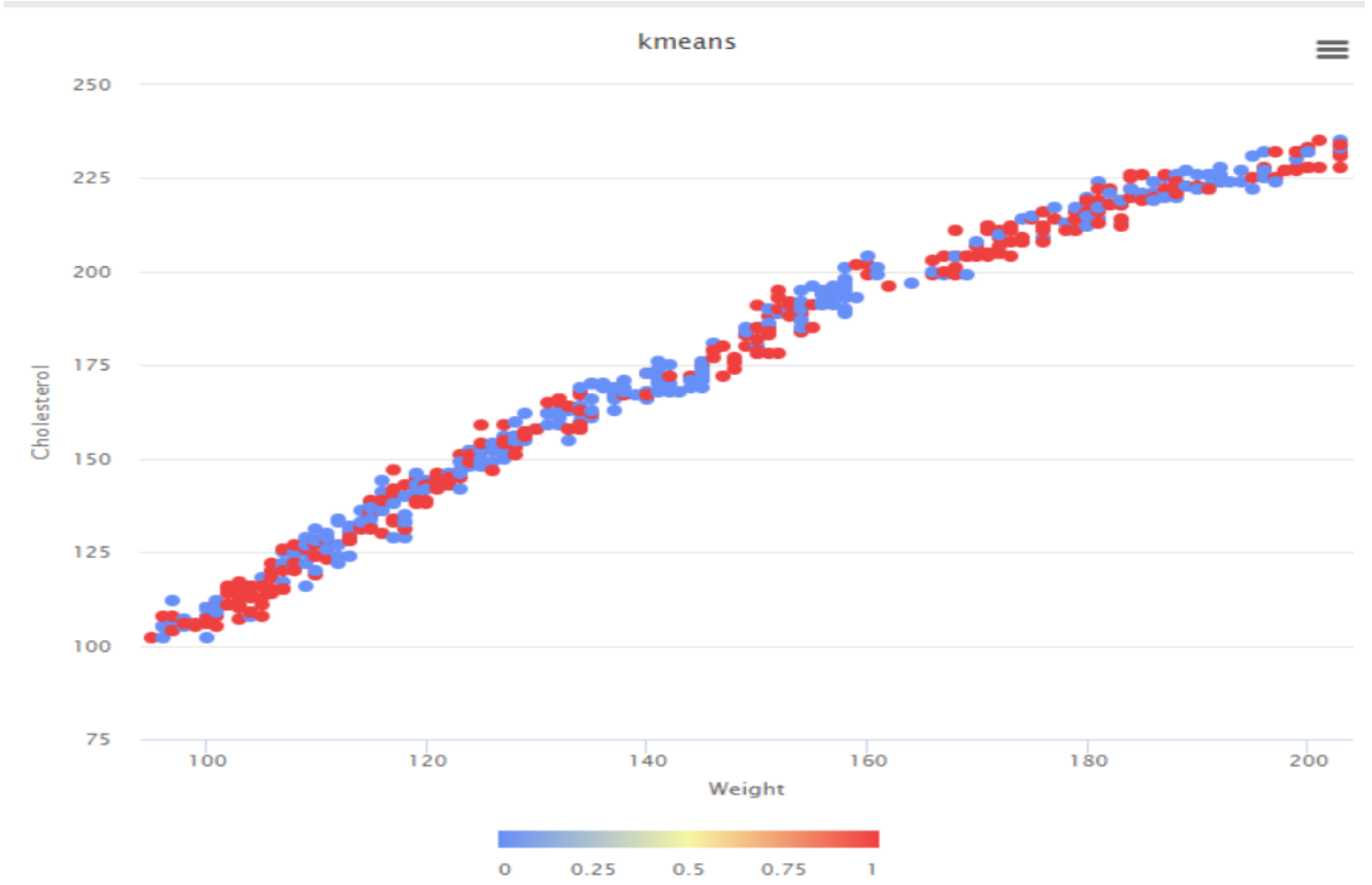


Θετική συσχέτιση
βάρους-χοληστερόλης

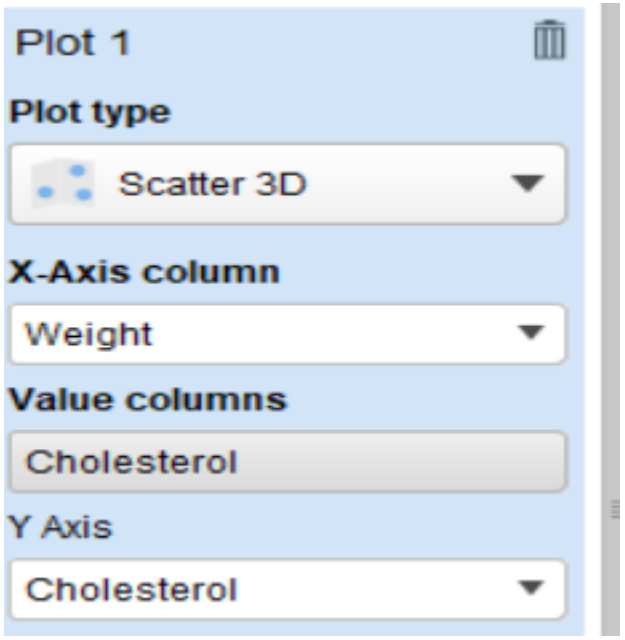


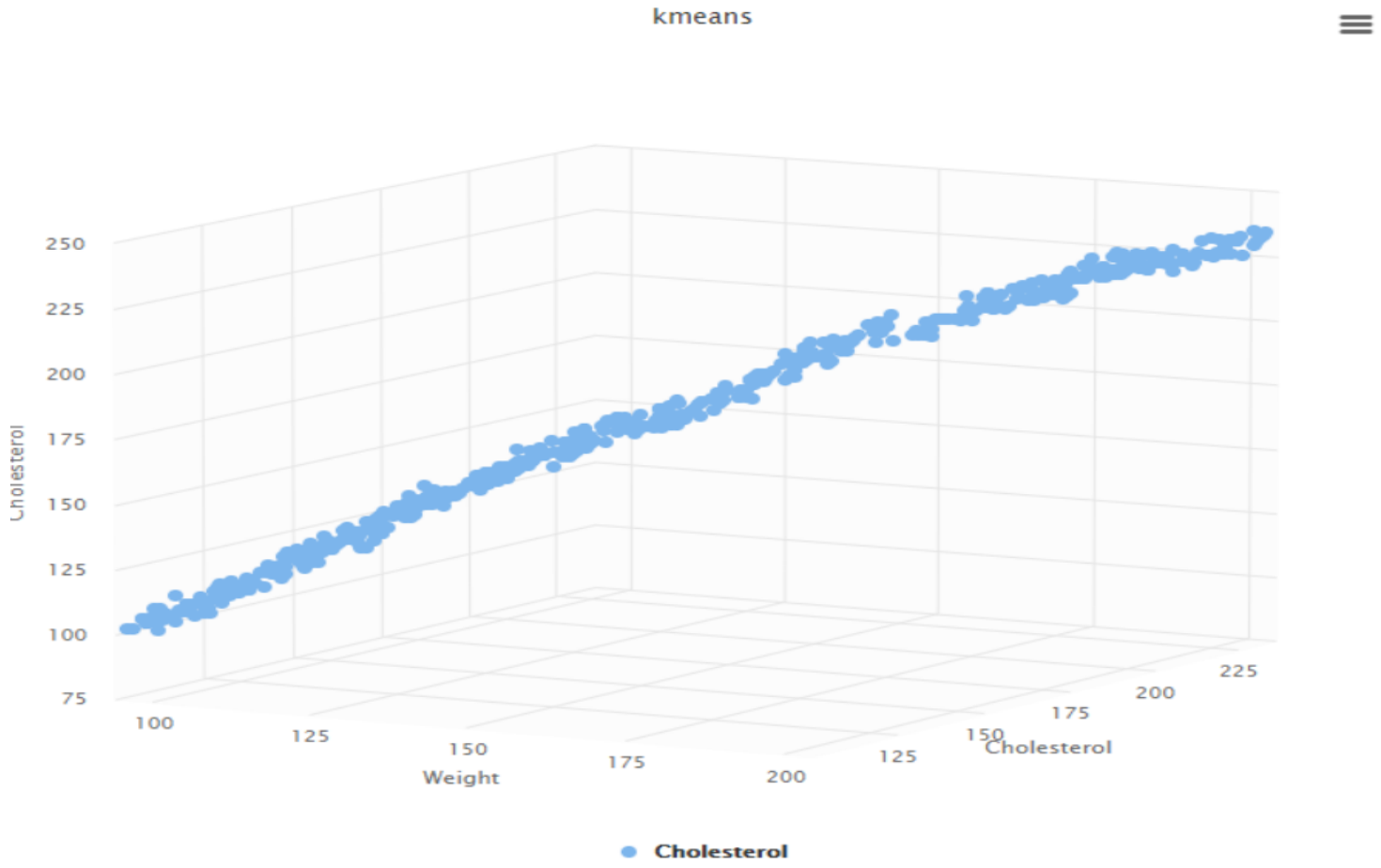
Διάγραμμα διασποράς (Scatter plot)



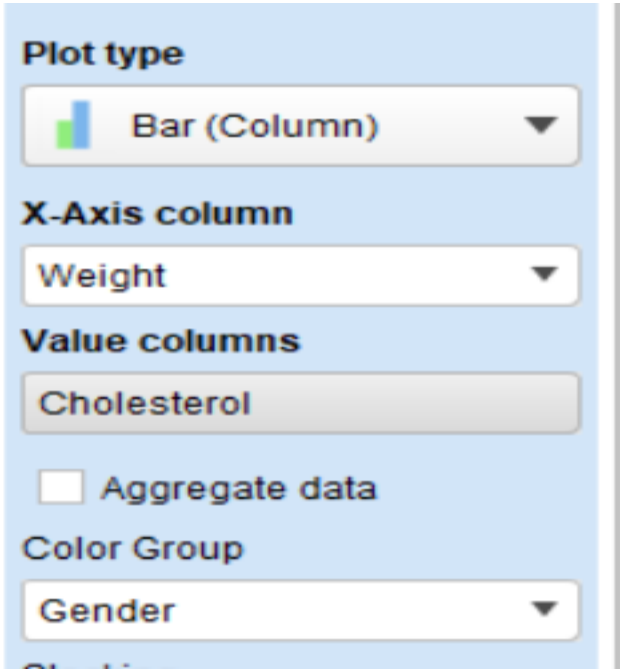


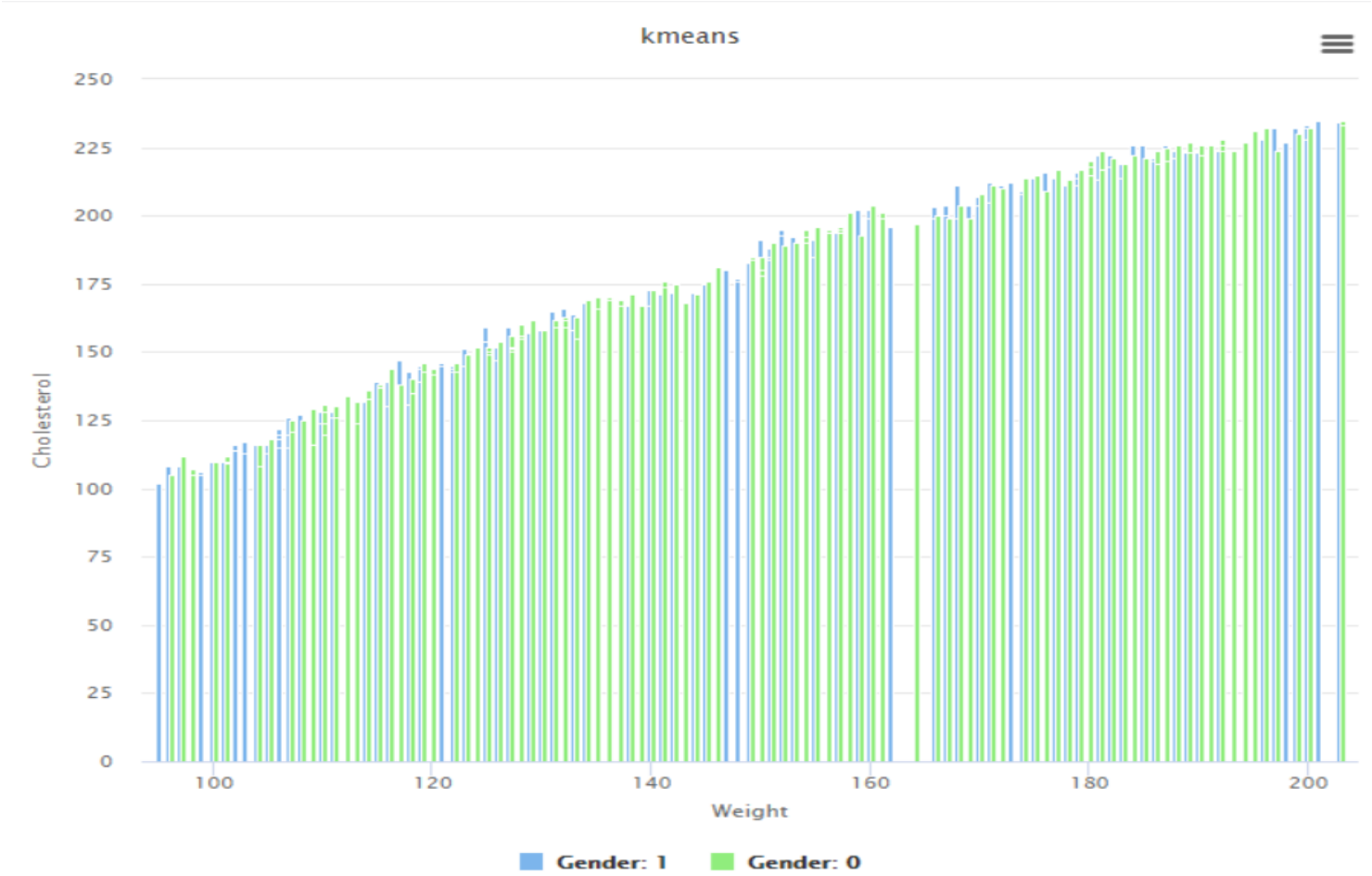
Τρισδιάστατο Διάγραμμα διασποράς Scatter (3D)



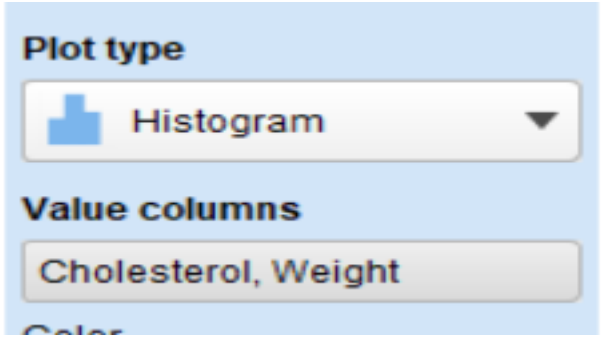


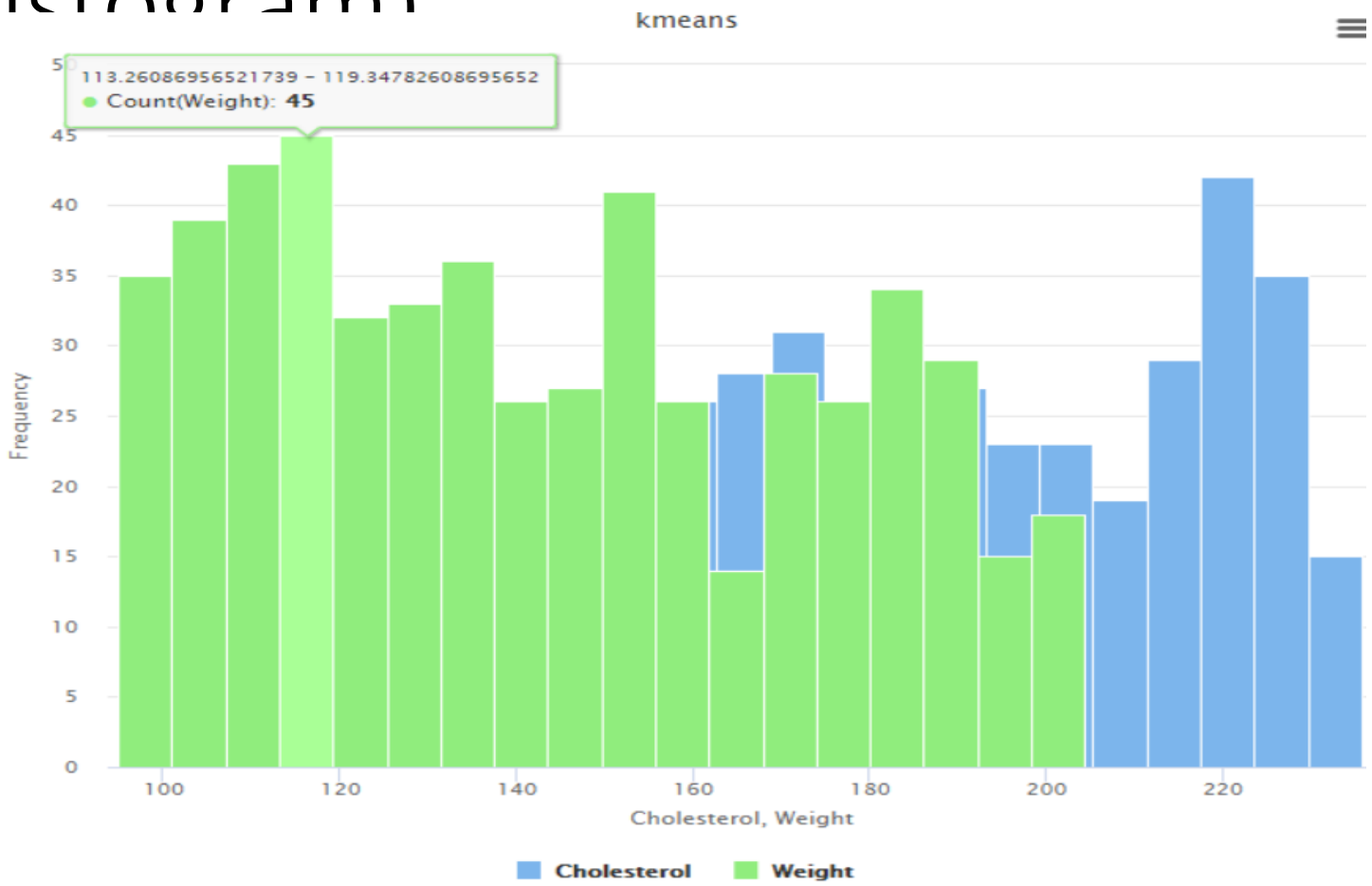
Ραβδόγραμμα - Bar (column) plot



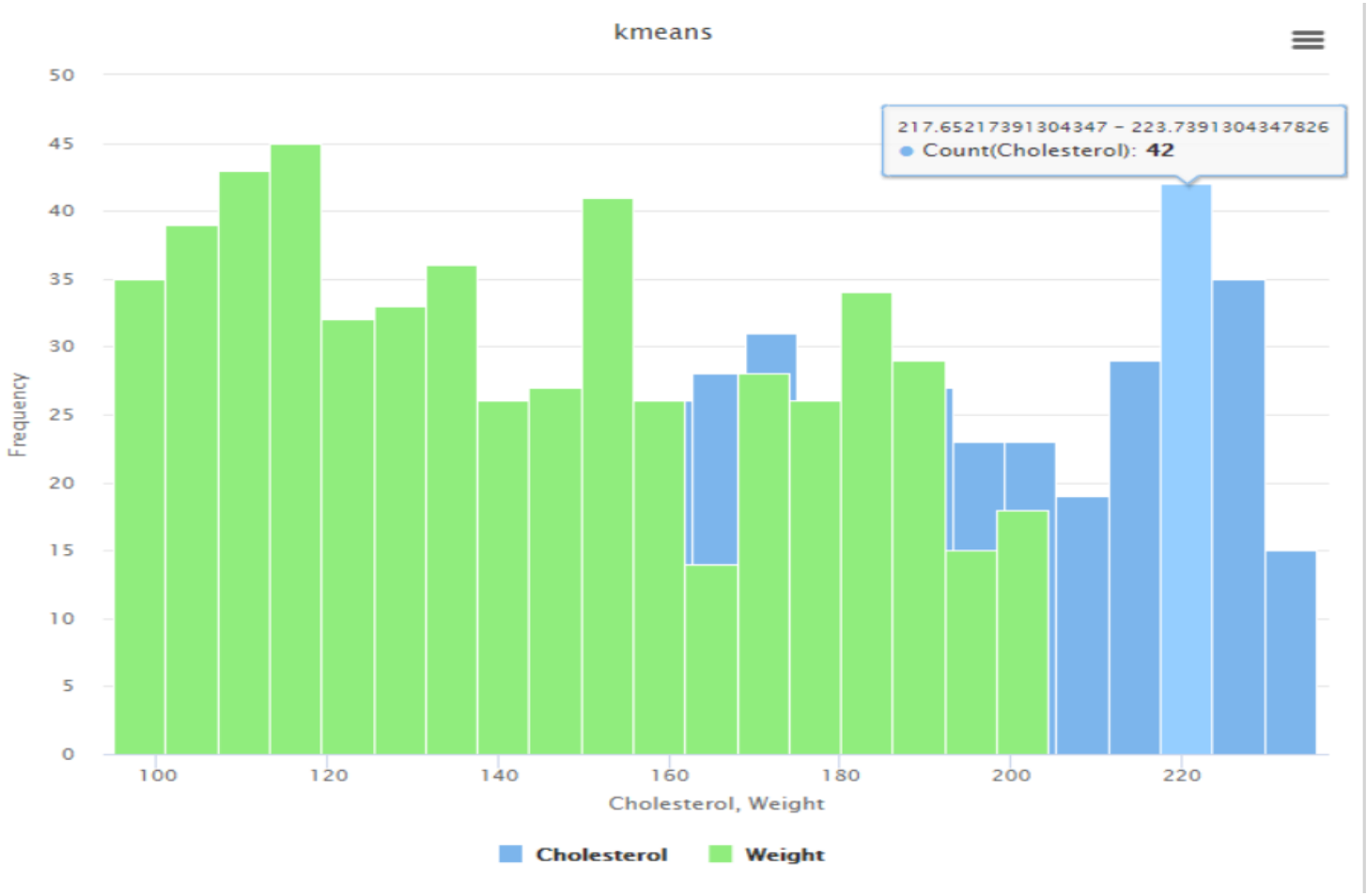


Ιστόγραμμα (Histogram)

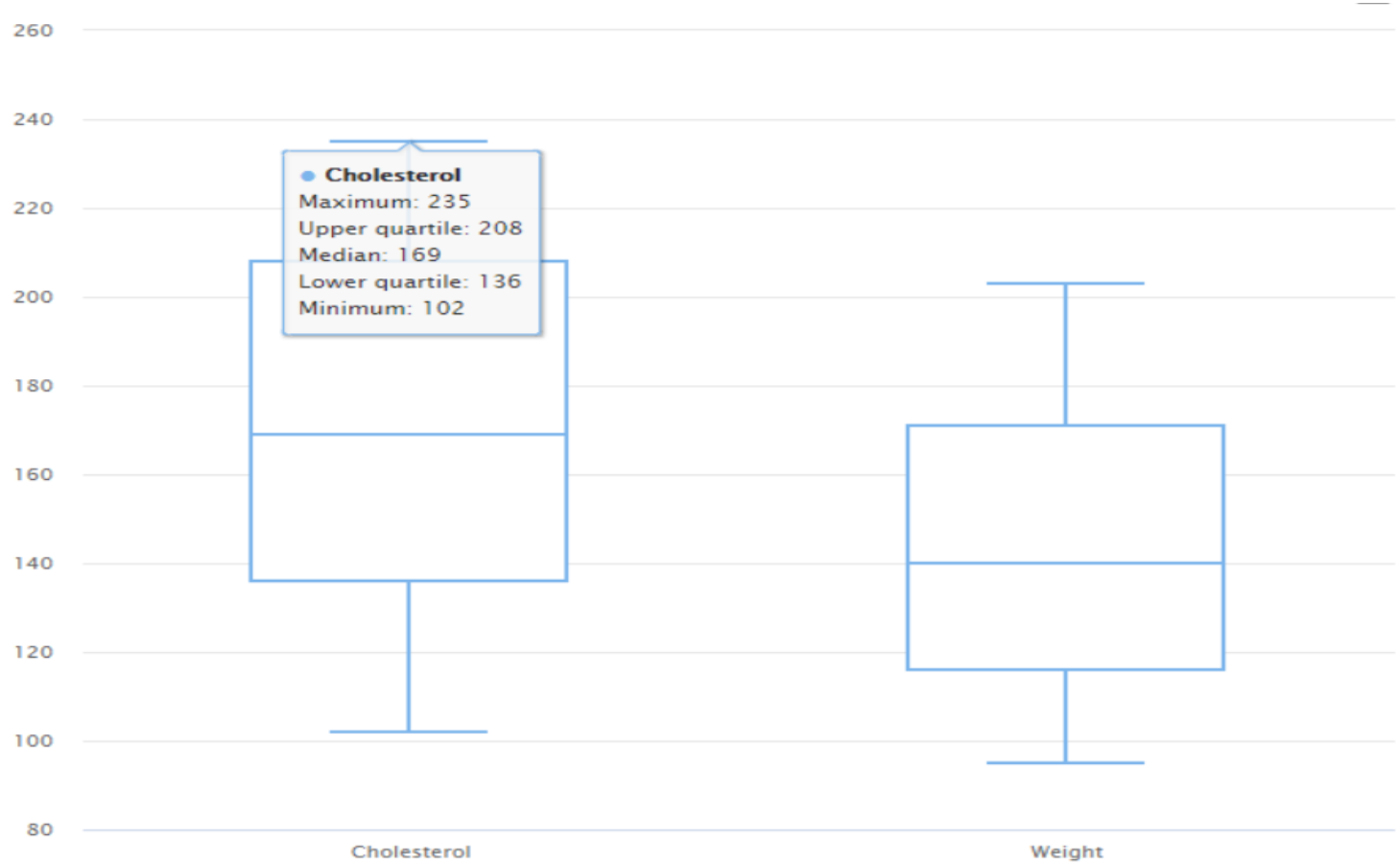




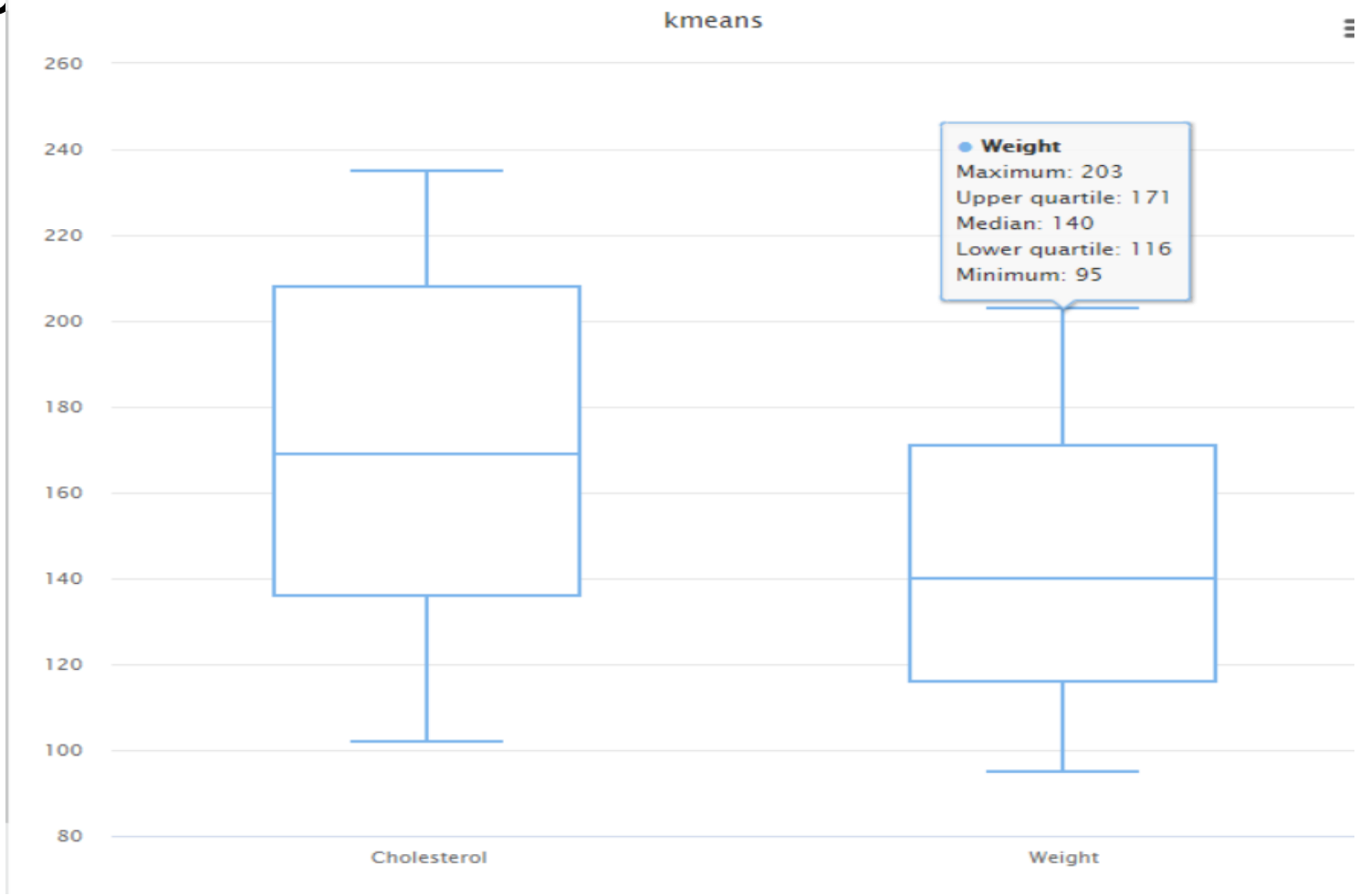
Ιστόγραμμα (Histogram)



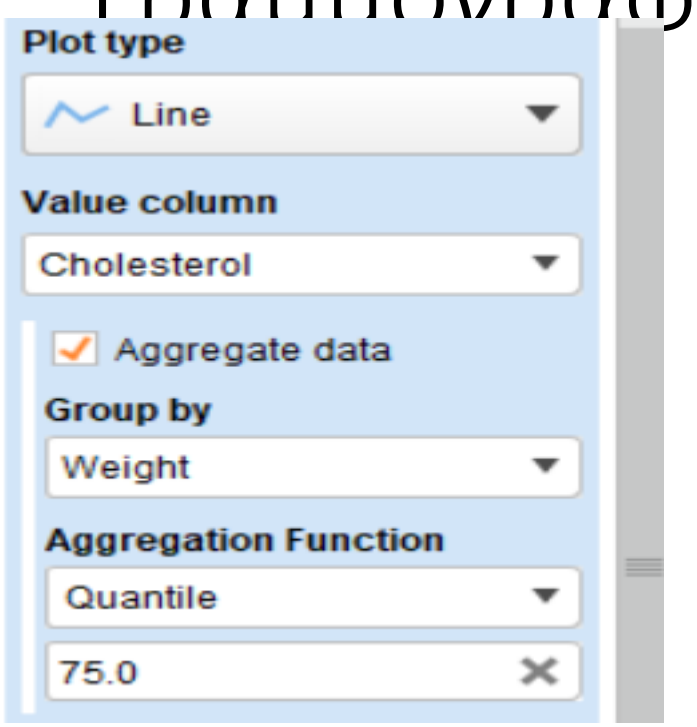
Θηκογράμματα (Boxplots)

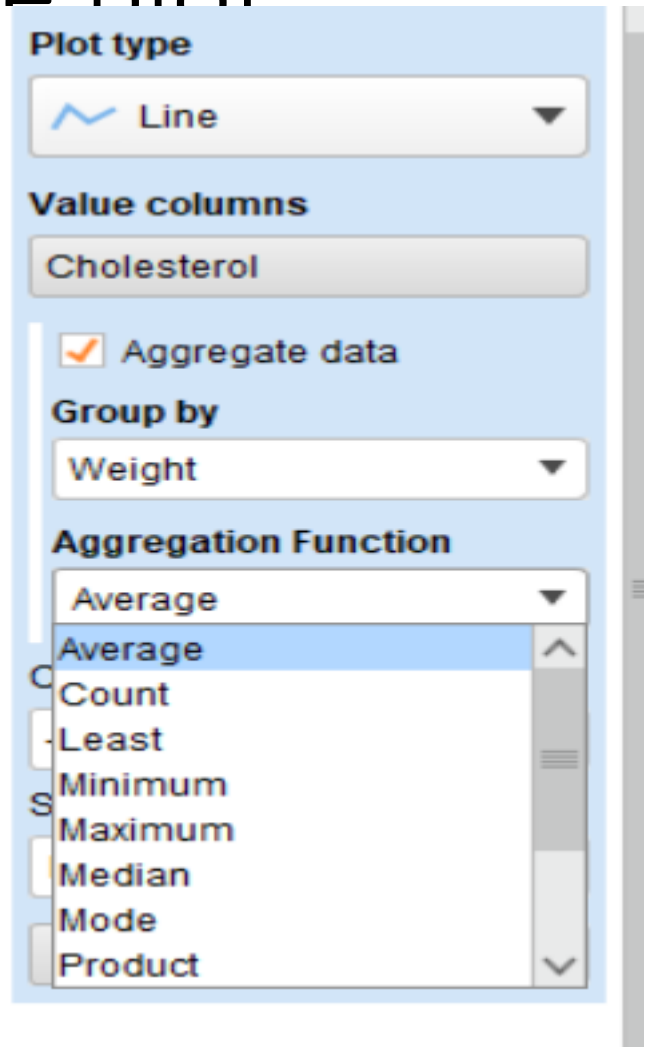


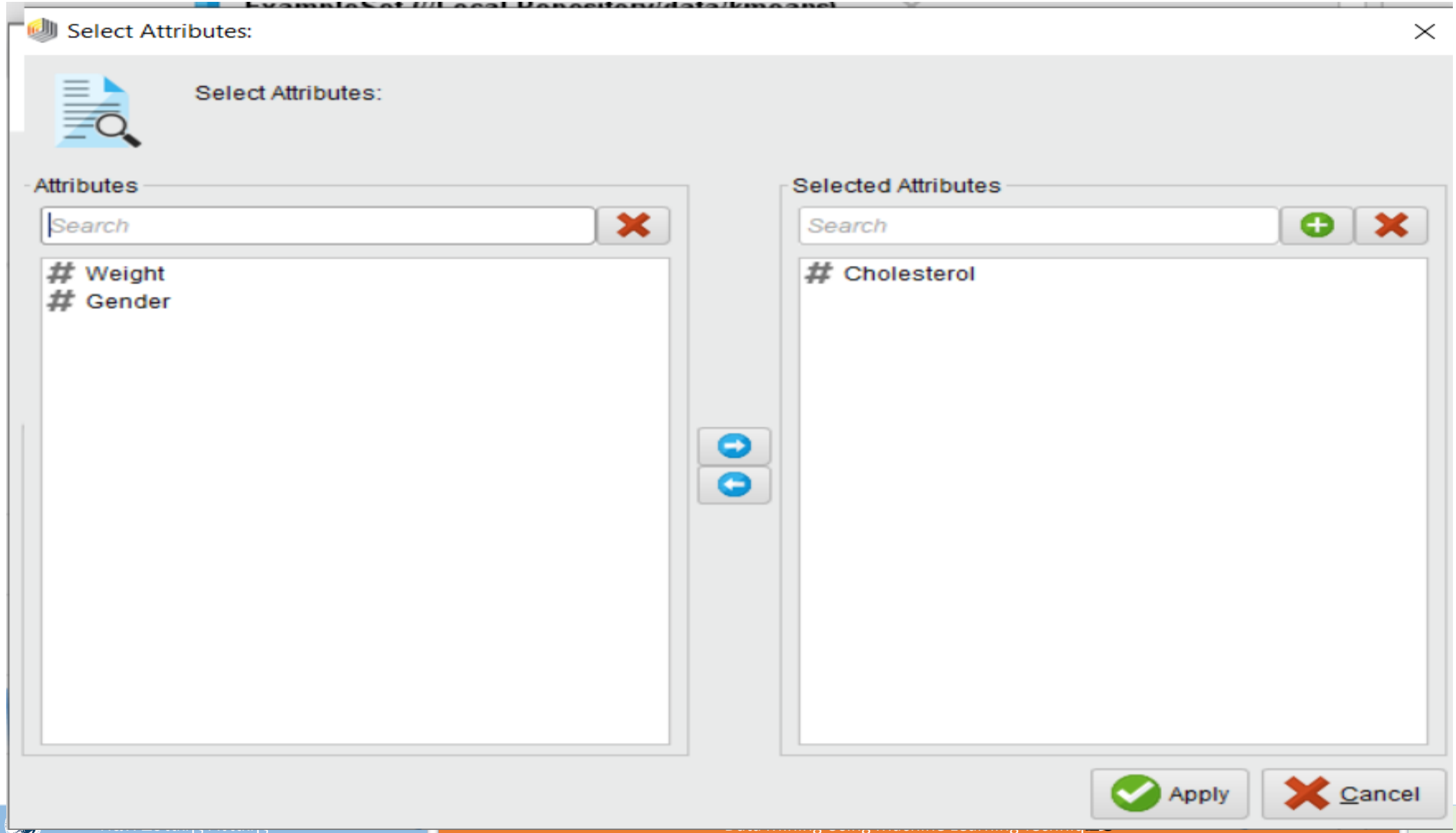
Θηκογράμματα (Boxplots)

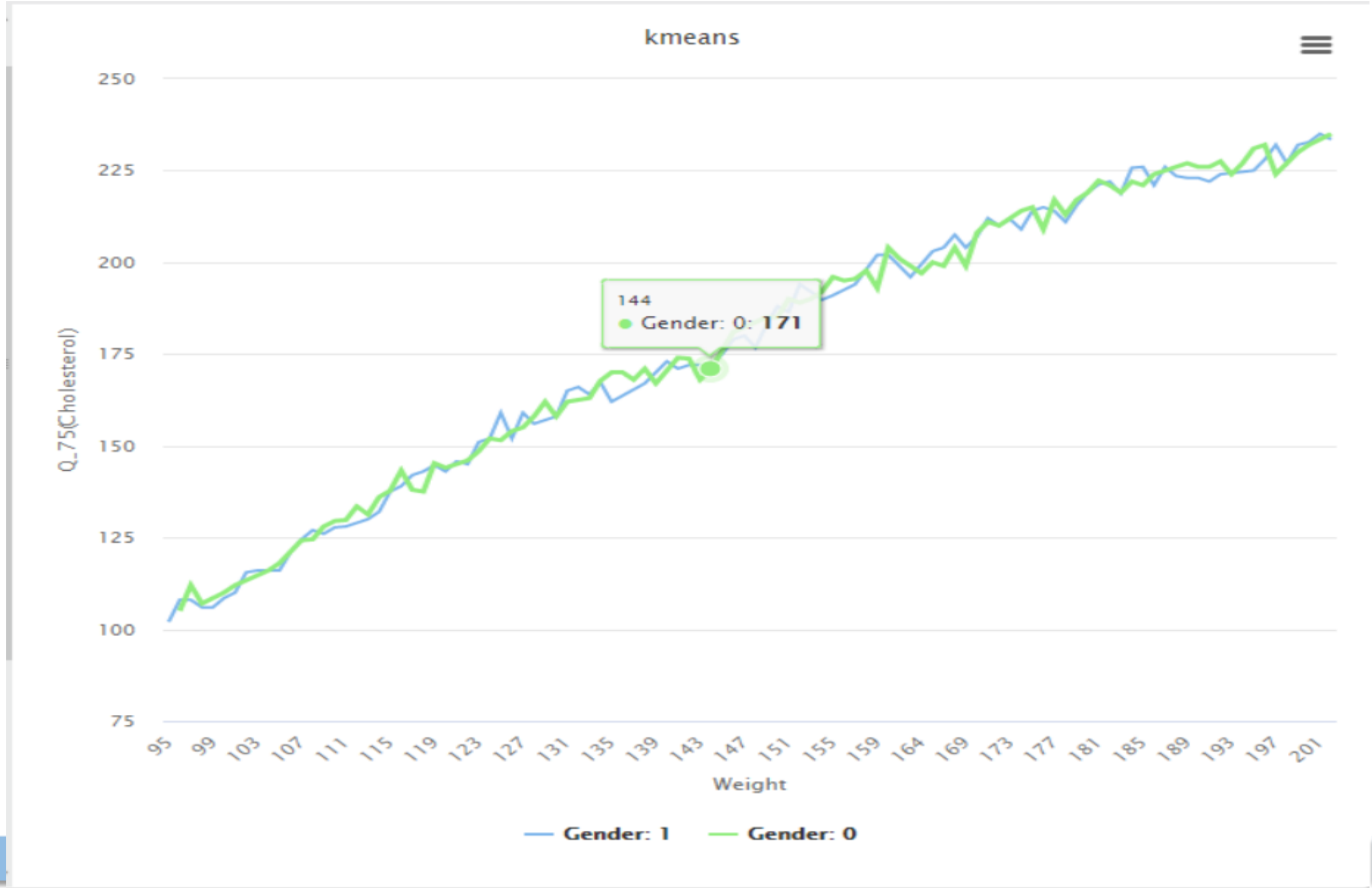


Γραμμογράφημα - Line plot

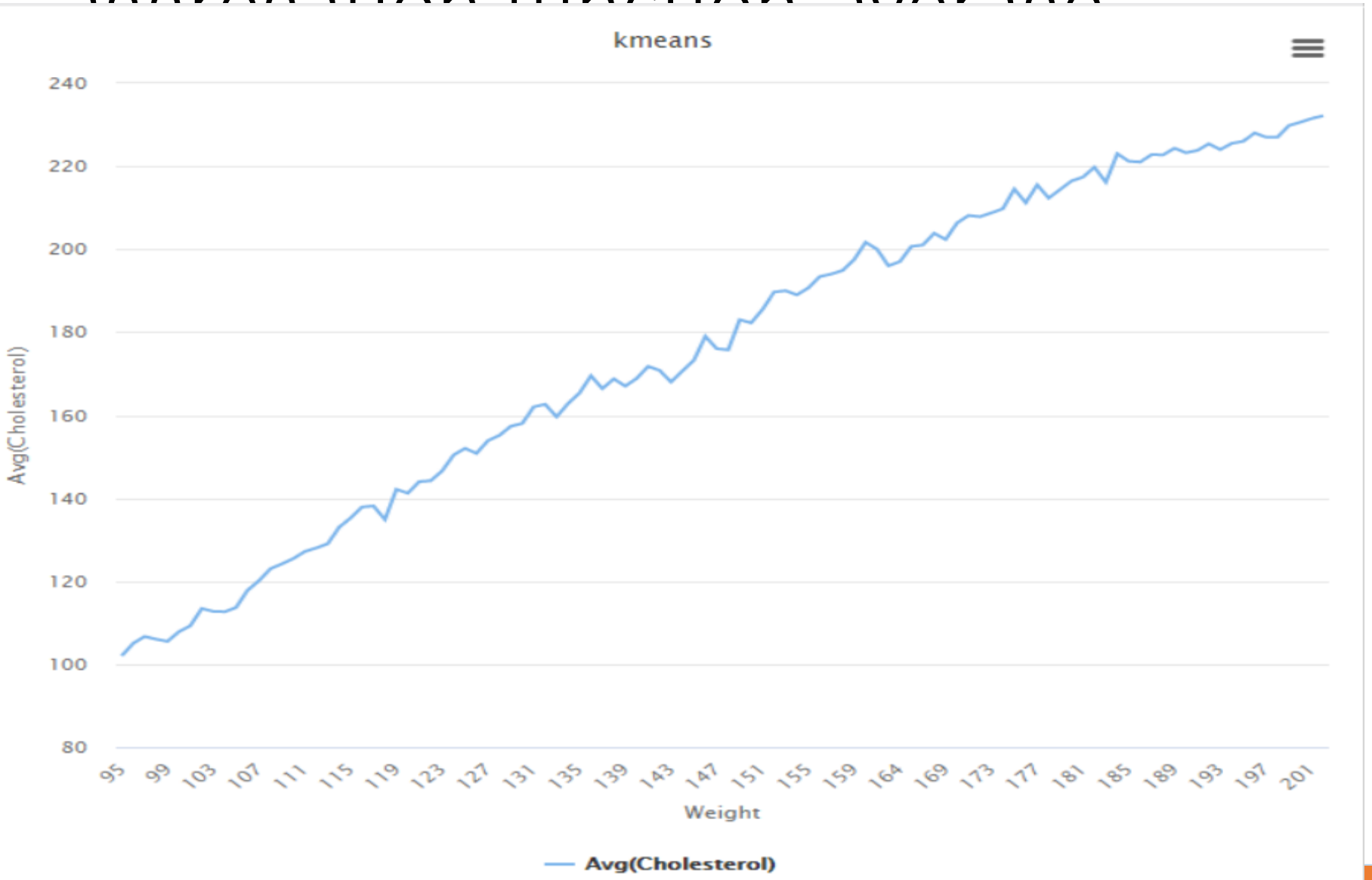




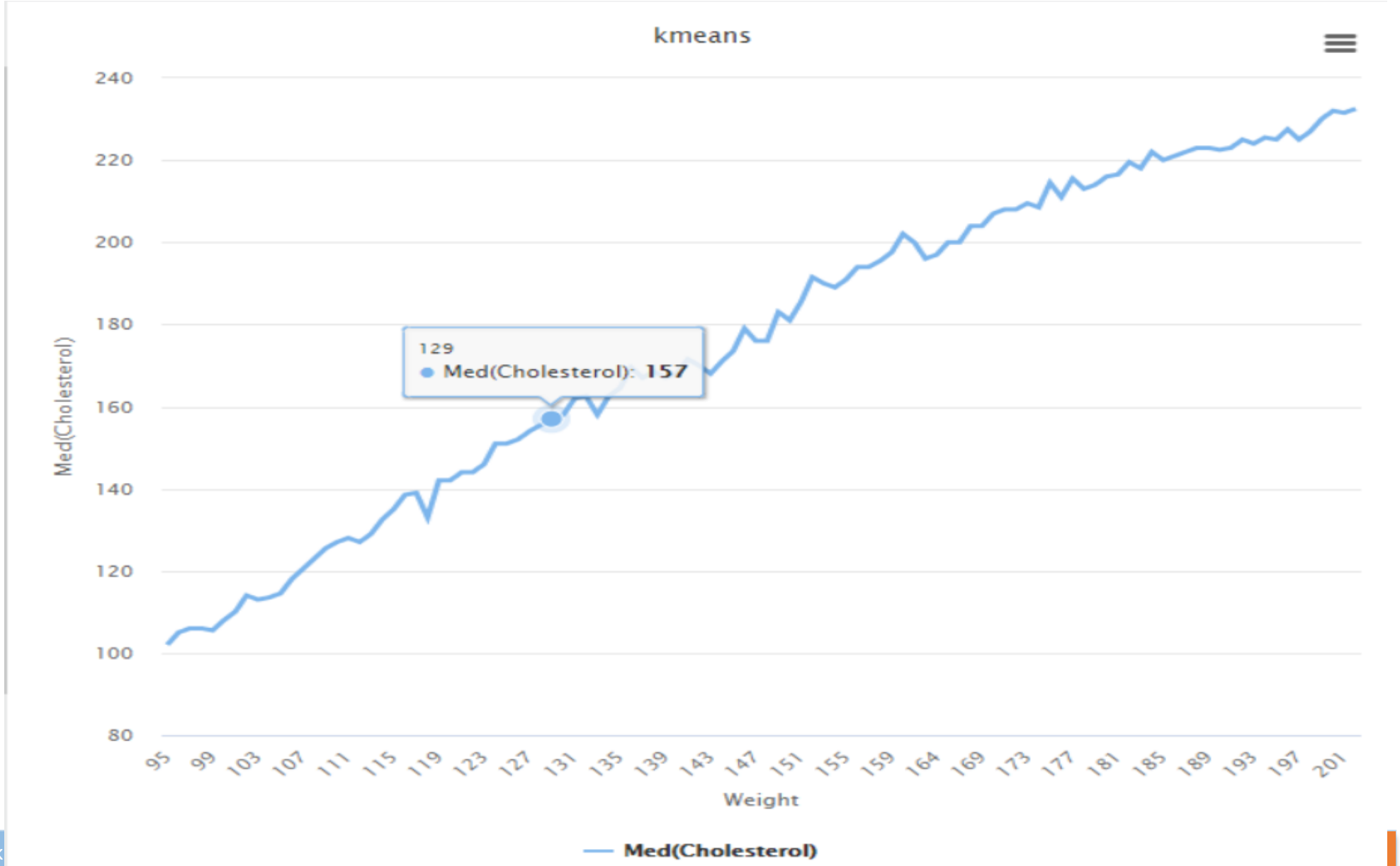

Select Attributes:
Select Attributes:
Attributes
Search
Weight
Gender
Selected Attributes
Search
Cholesterol
Apply
Cancel


kmeans
250
225
200
175
150
125
100
75
Q_75(Cholesterol)
144
Gender: 0: 171
95
99
103
107
111
115
119
123
127
131
135
139
143
147
151
155
159
164
169
173
177
181
185
189
193
197
201
Weight
Gender: 1
Gender: 0

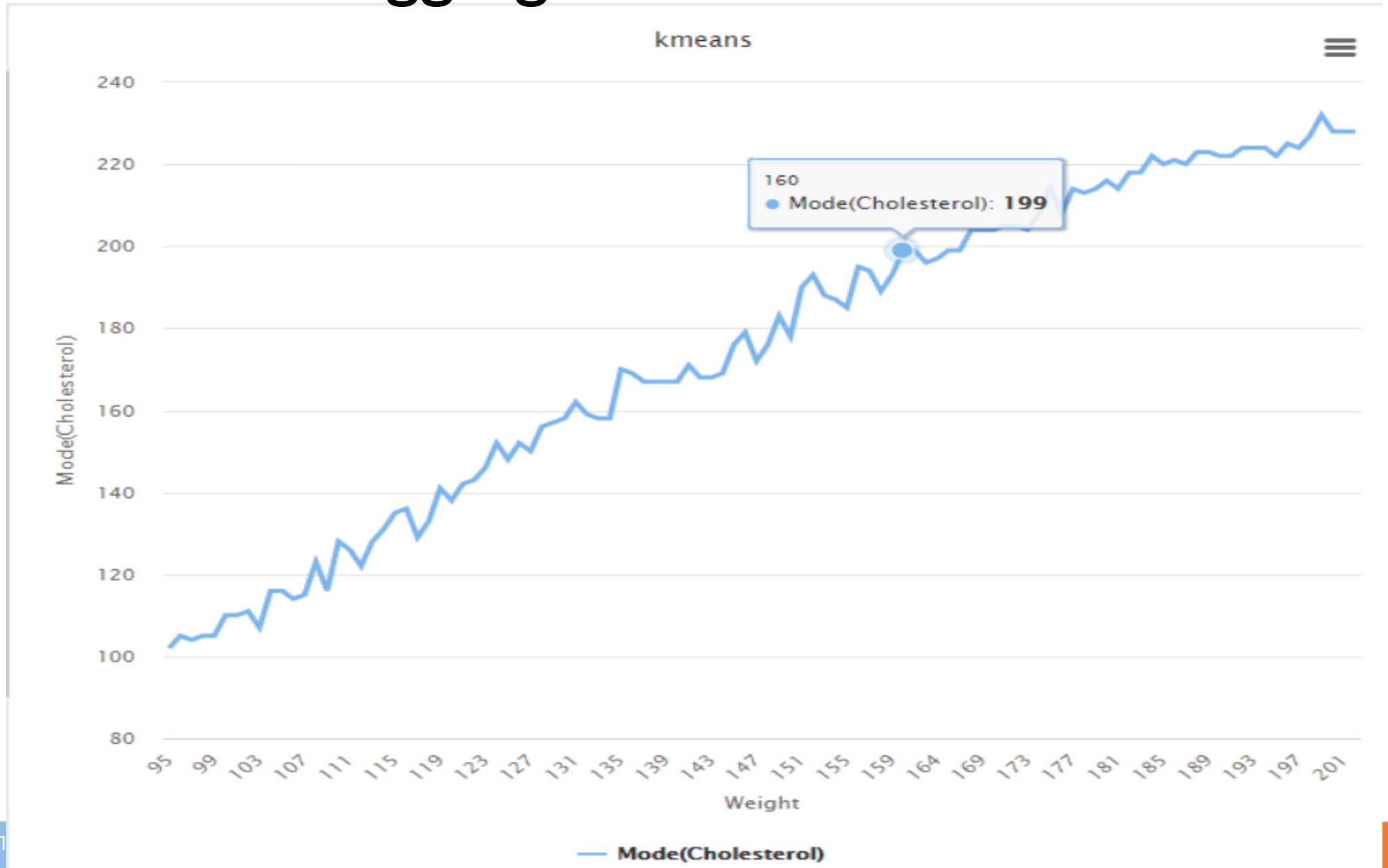
Line – aggregation function average



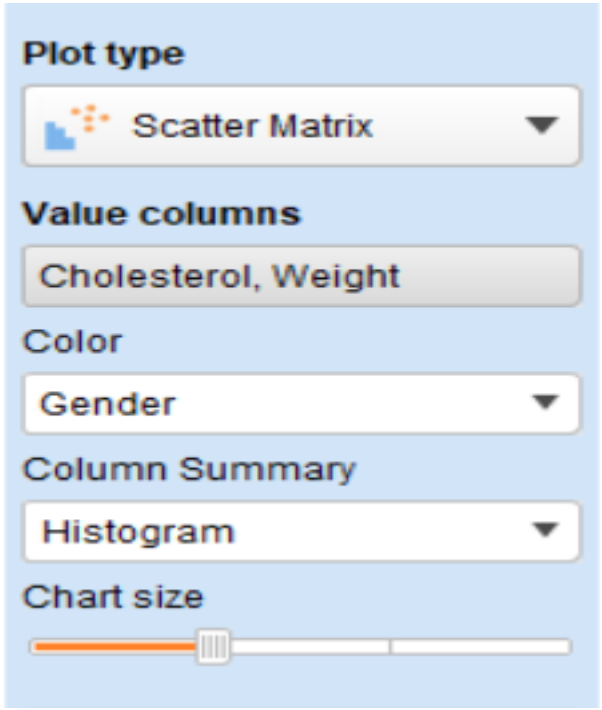
Line – aggregation function - Median



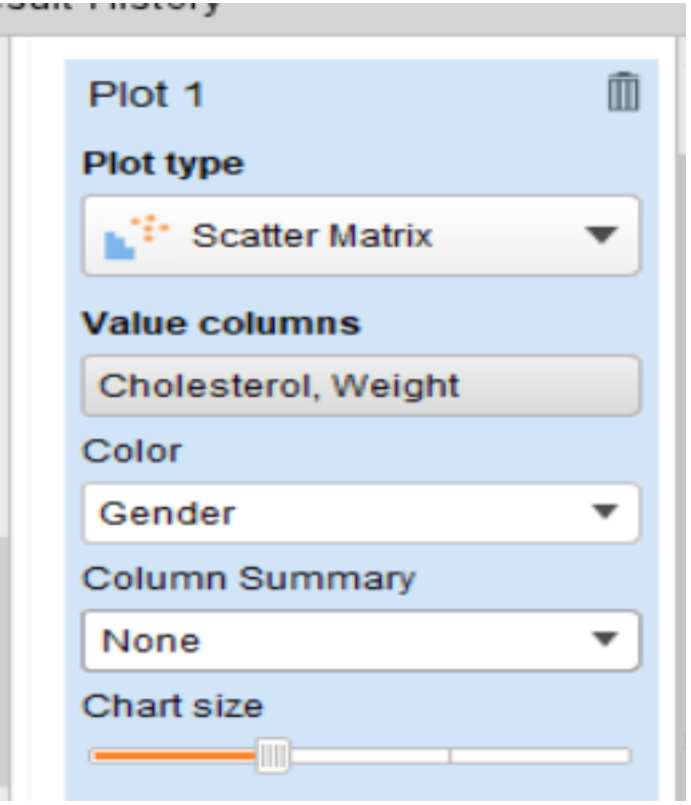
Line – aggregation function - Mode

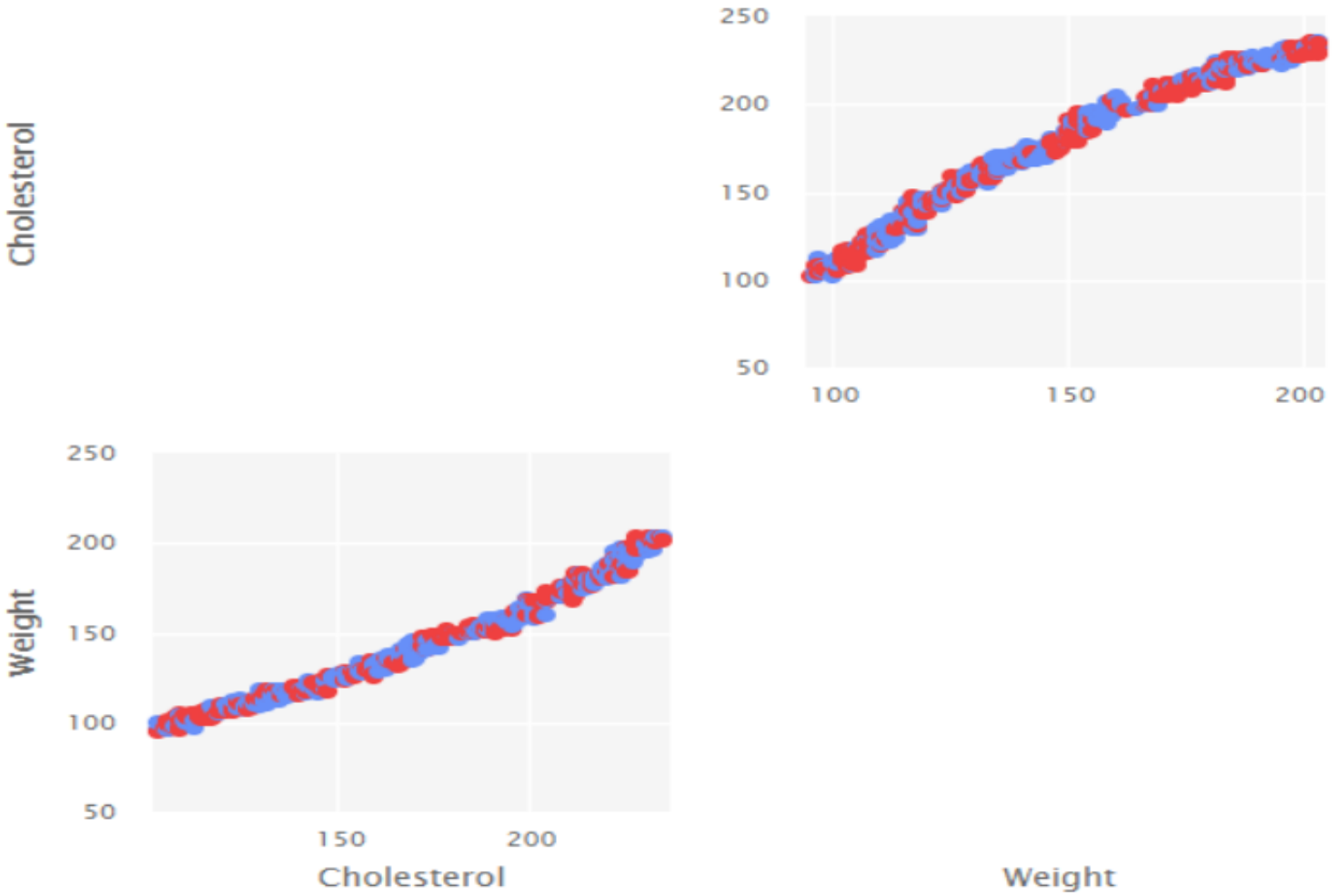


Πίνακας διασποράς (Scatter matrix)

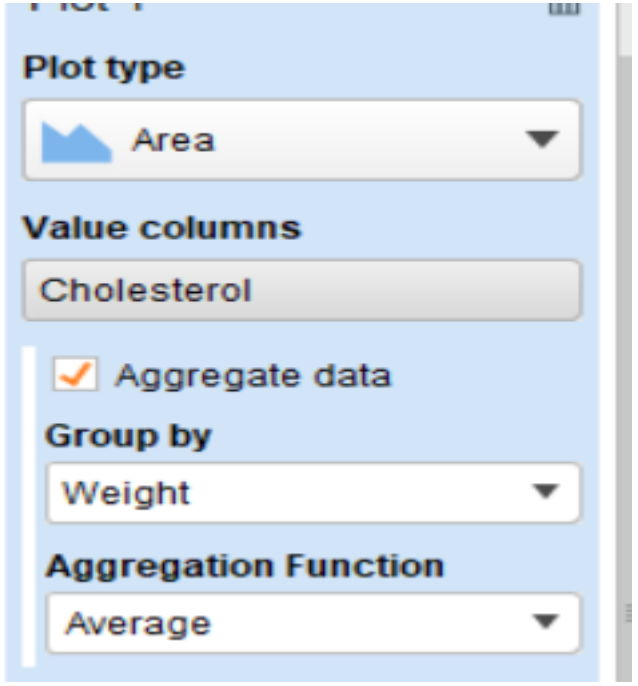


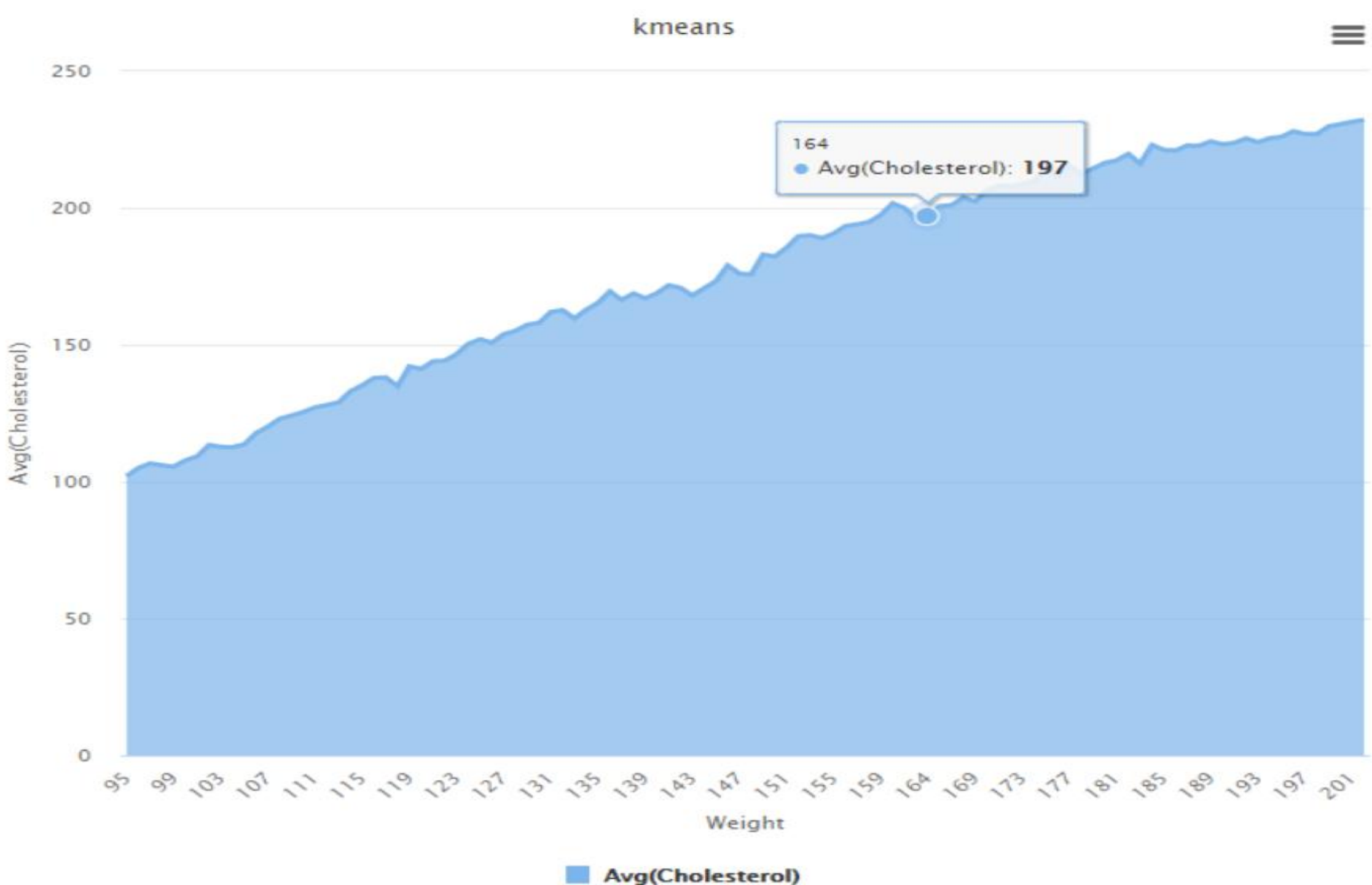
Πίνακας διασποράς (Scatter matrix)

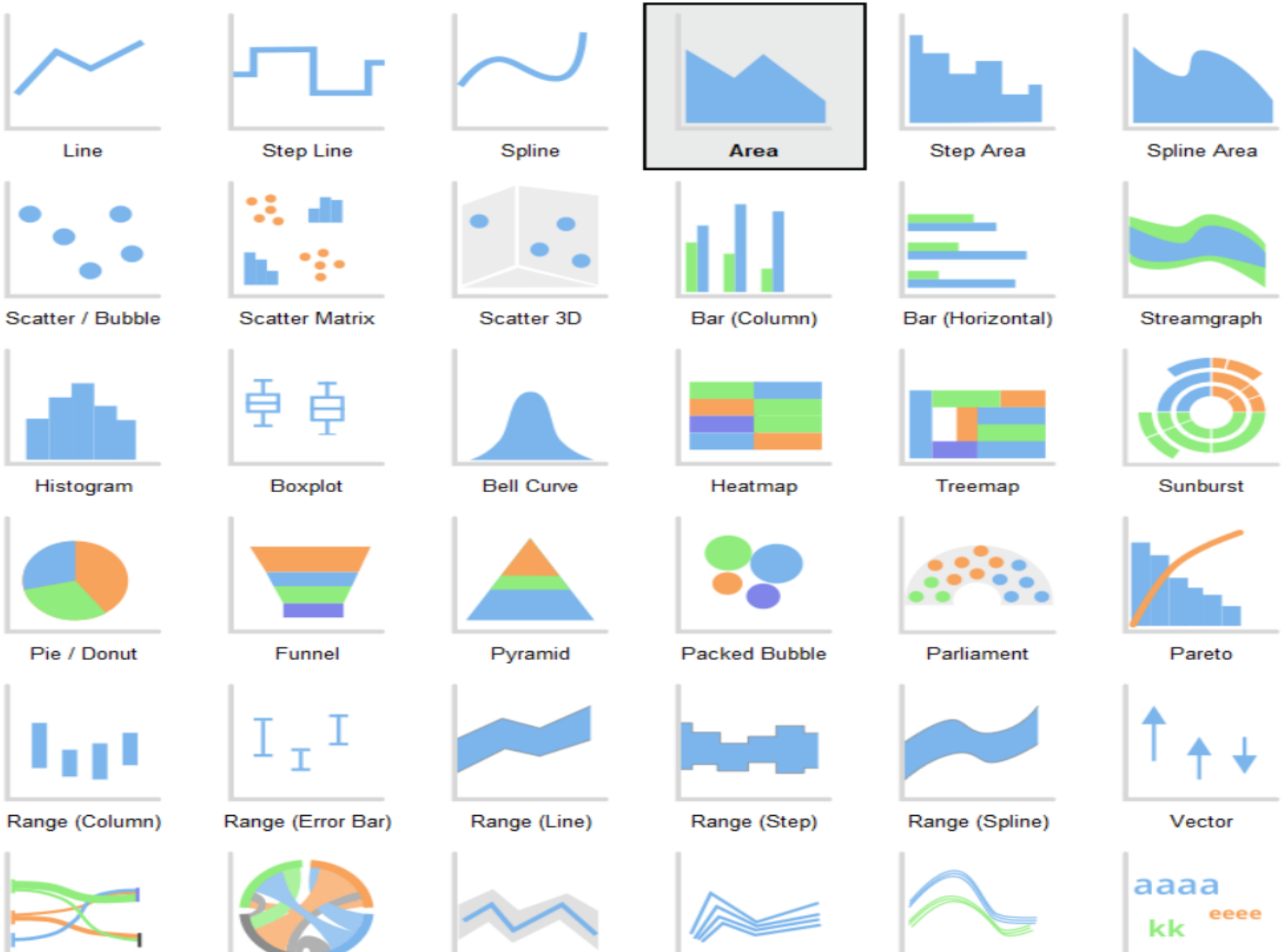




Περιοχή (area plot)



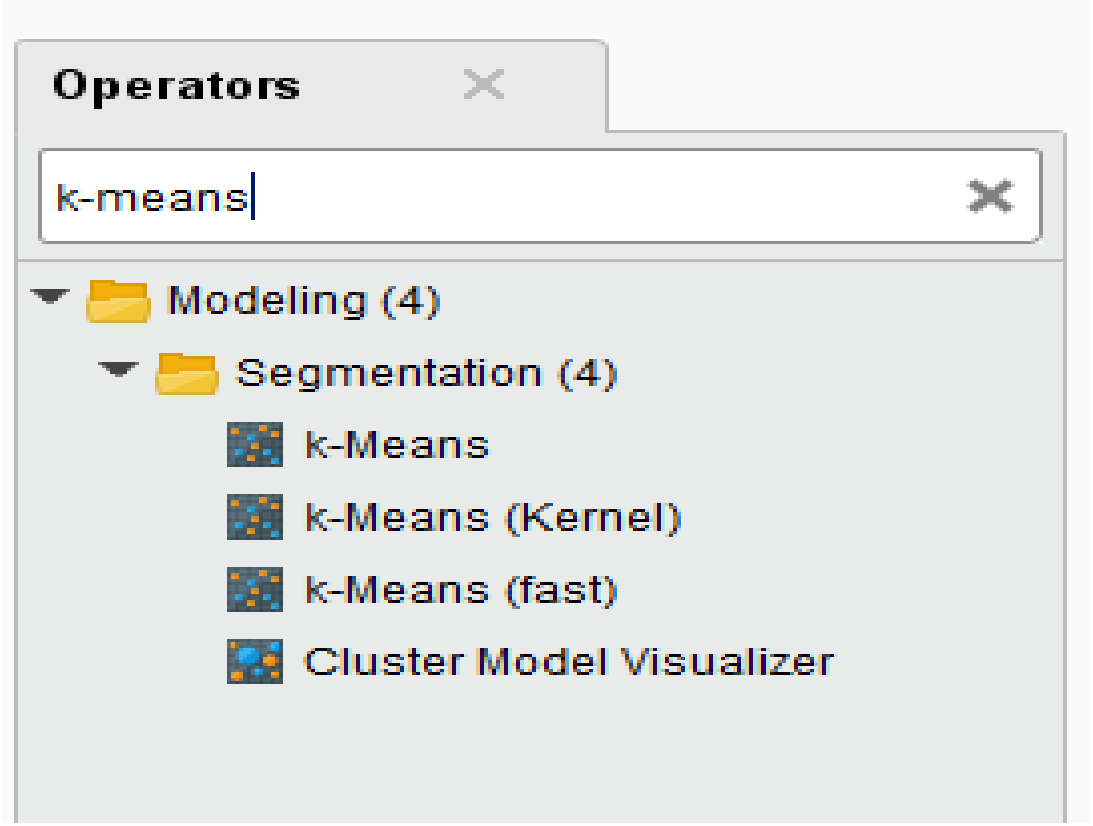

kmeans
250
200
150
100
50
0
Avg(Cholesterol)
164
Avg(Cholesterol): 197
95
99
103
107
111
115
119
123
127
131
135
139
143
147
151
155
159
164
169
173
177
181
185
189
193
197
201
Weight
Avg(Cholesterol)


Line
Step Line
Spline
Area
Step Area
Spline Area
Scatter / Bubble
Scatter Matrix
Scatter 3D
Bar (Column)
Bar (Horizontal)
Streamgraph
Histogram
Boxplot
Bell Curve
Heatmap
Treemap
Sunburst
Pie / Donut
Funnel
Pyramid
Packed Bubble
Parliament
Pareto
Range (Column)
Range (Error Bar)
Range (Line)
Range (Step)
Range (Spline)
Vector
aaaa
kk
eeee

k-Means – centroid

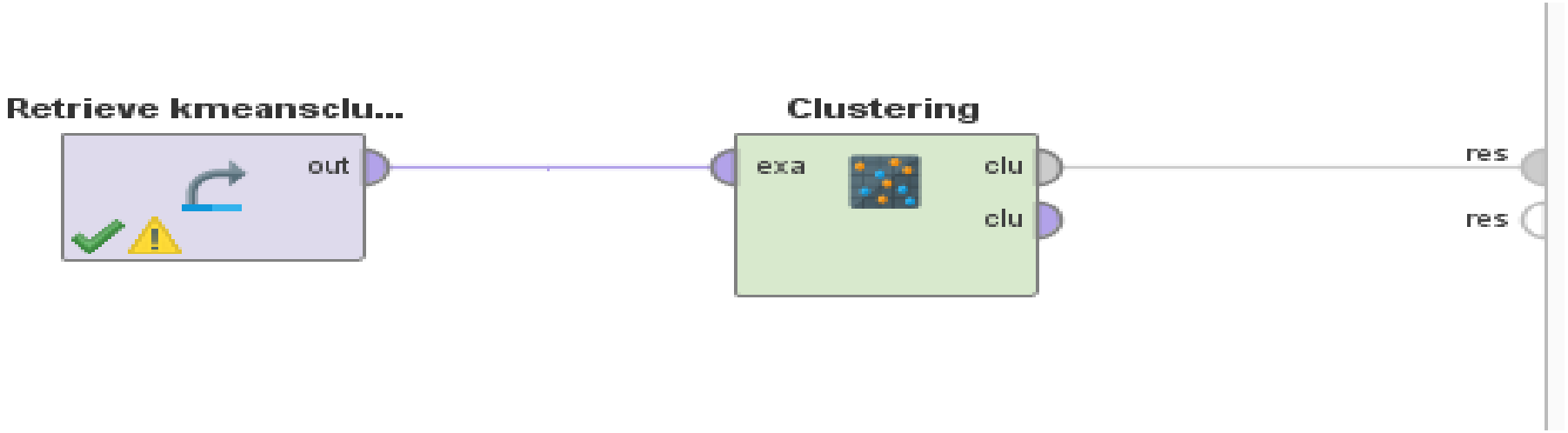
Μοντέλο Ανάλυσης

K-means operators

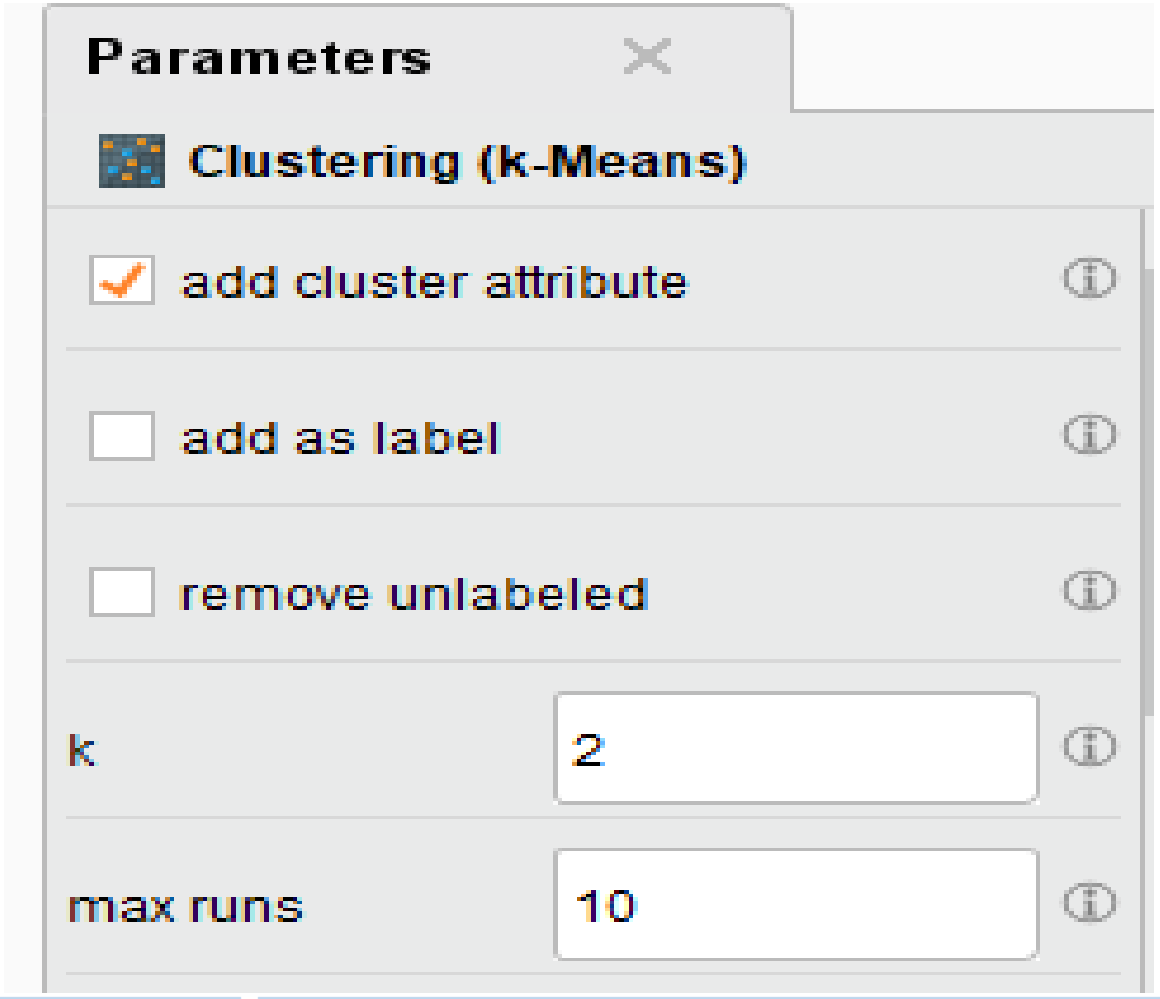


https://docs.rapidminer.com/latest/studio/operators/modeling/segmentation/k_means.html

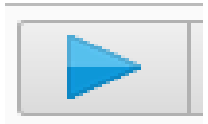
model



Parameter pane: 2 clusters, max runs 10



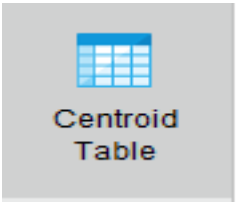
Initial report



Cluster Model

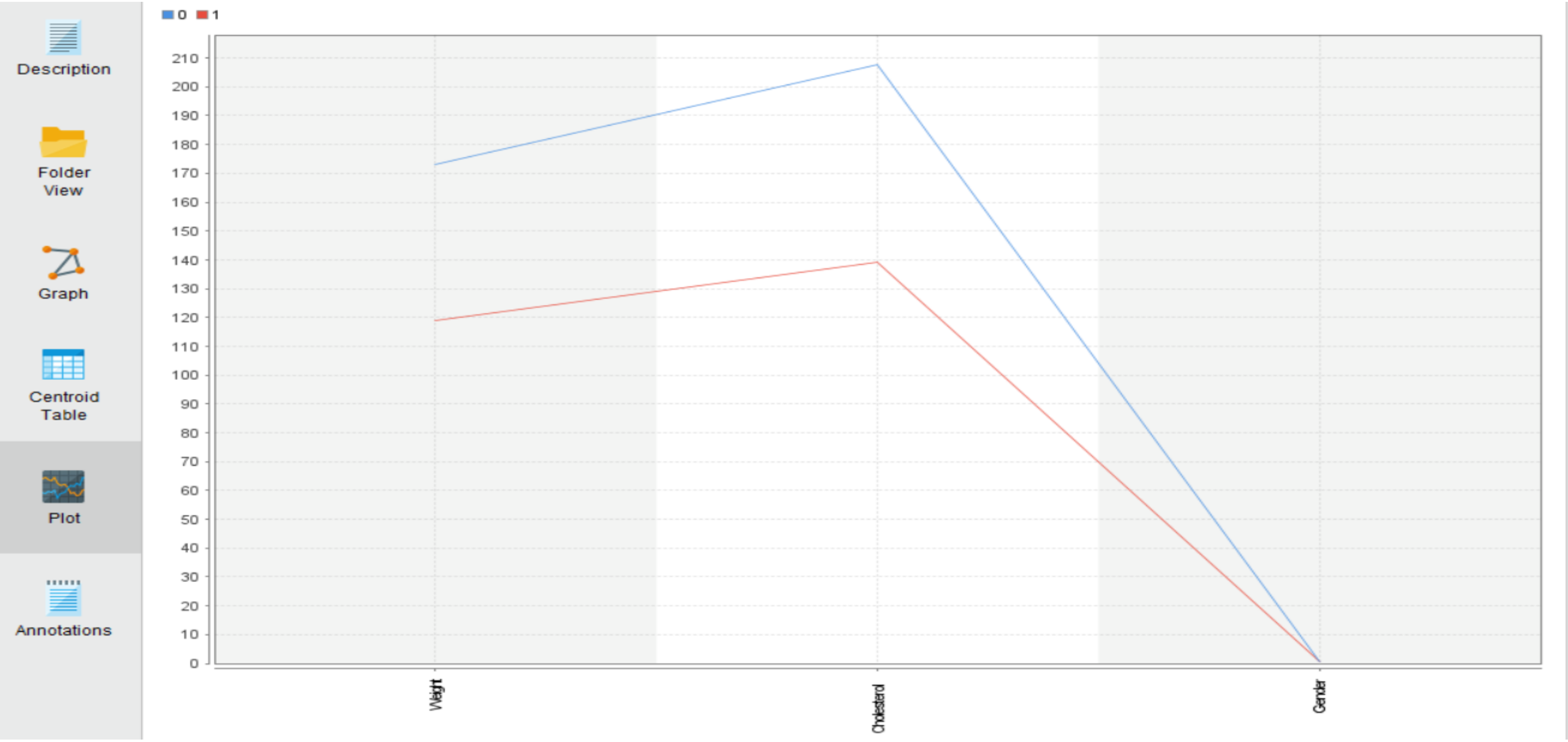
```
Cluster 0: 249 items
Cluster 1: 298 items
Total number of items: 547
```

Centroid Table

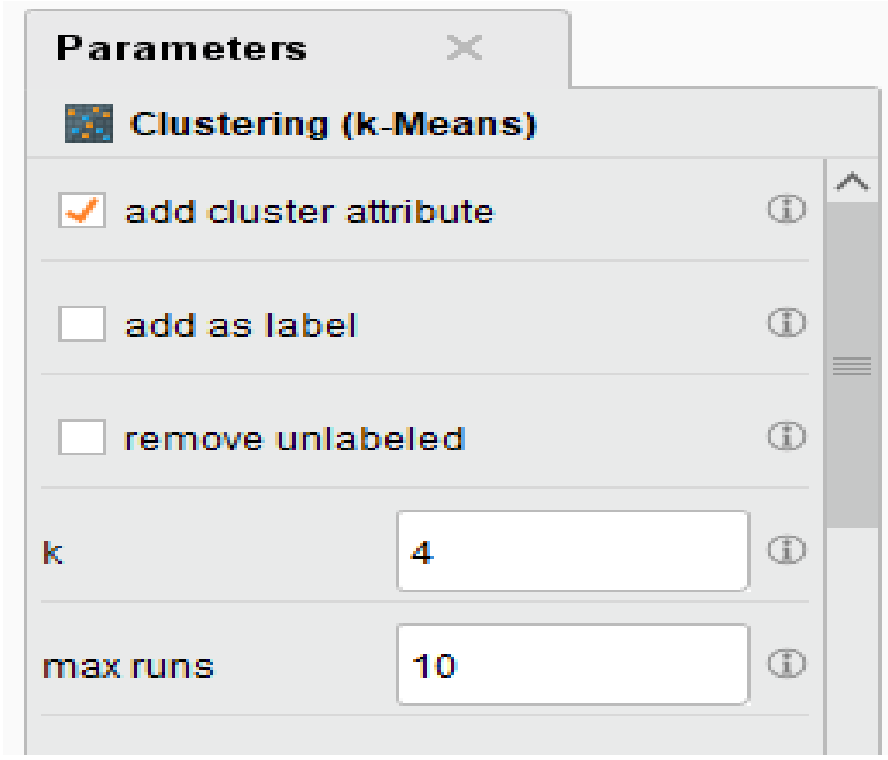


Attribute	cluster_0	cluster_1
Weight	172.892	119.074
Cholesterol	207.635	139.349
Gender	0.554	0.480

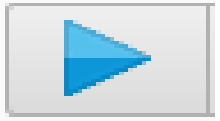
Χρήσιμο μόνο αν θέλουμε να προσδιορίσουμε άτομα υψηλού και άτομα χαμηλού κινδύνου για στεφανιαία νόσο


Description
Folder View
Graph
Centroid Table
Plot
Annotations
0
1
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
Weight
Cholesterol
Gender

Parameter pane: 4 clusters, max runs 10



4 clusters



Cluster Model

```
Cluster 0: 118 items
Cluster 1: 140 items
Cluster 2: 154 items
Cluster 3: 135 items
Total number of items: 547
```

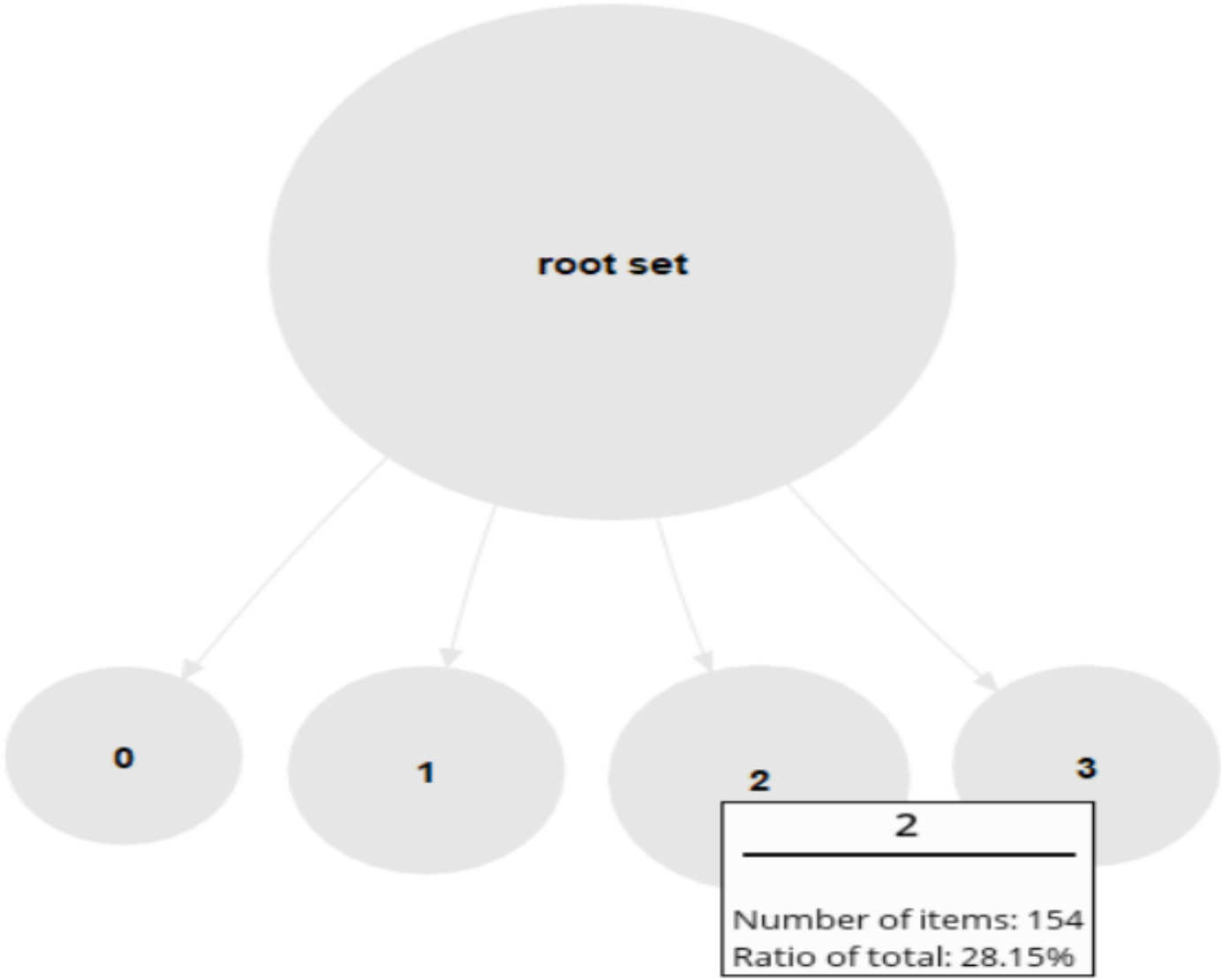
Ισορροπημένες συστάδες

Zoom

Tree

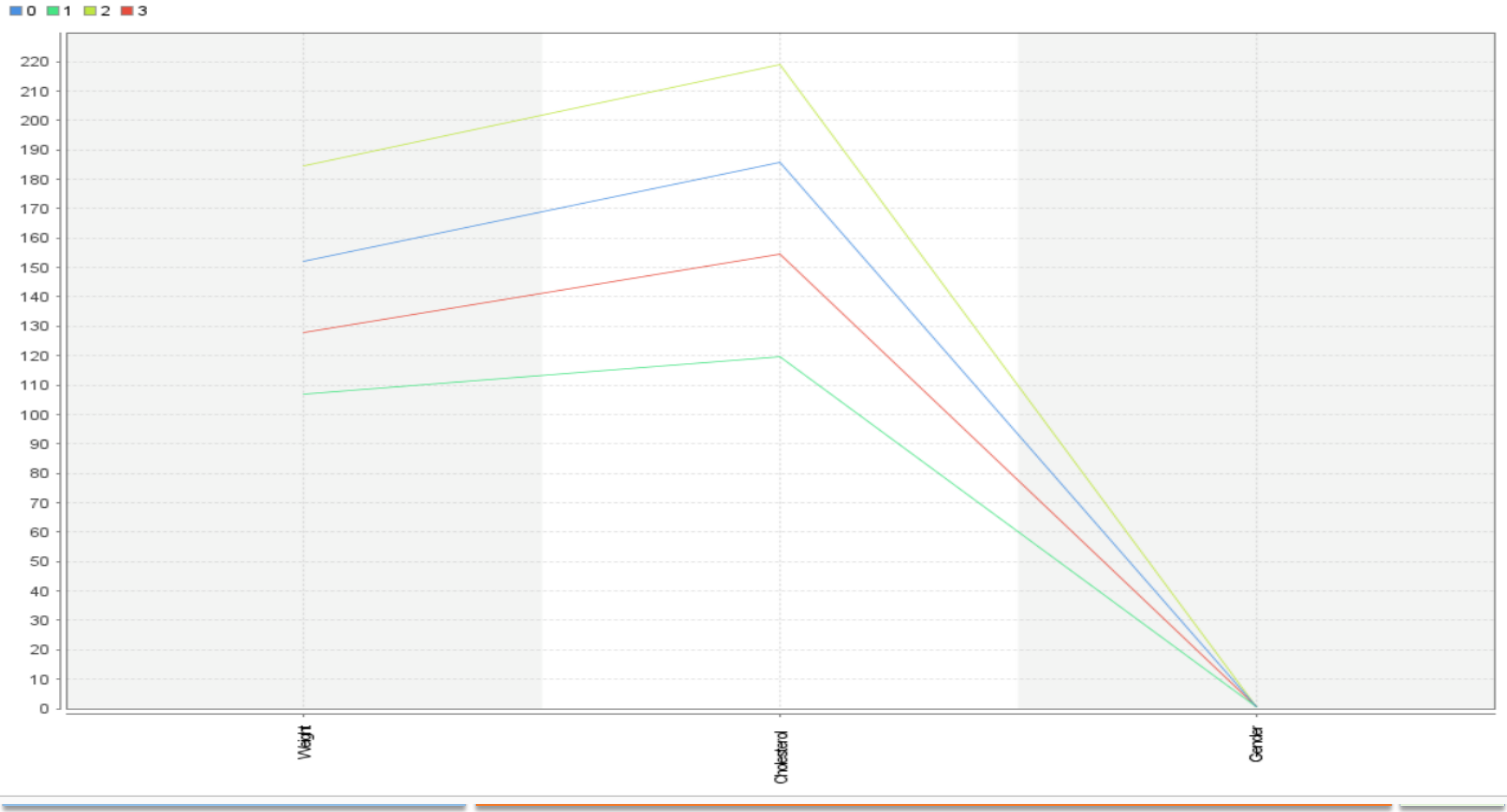
Node Labels

Edge Labels

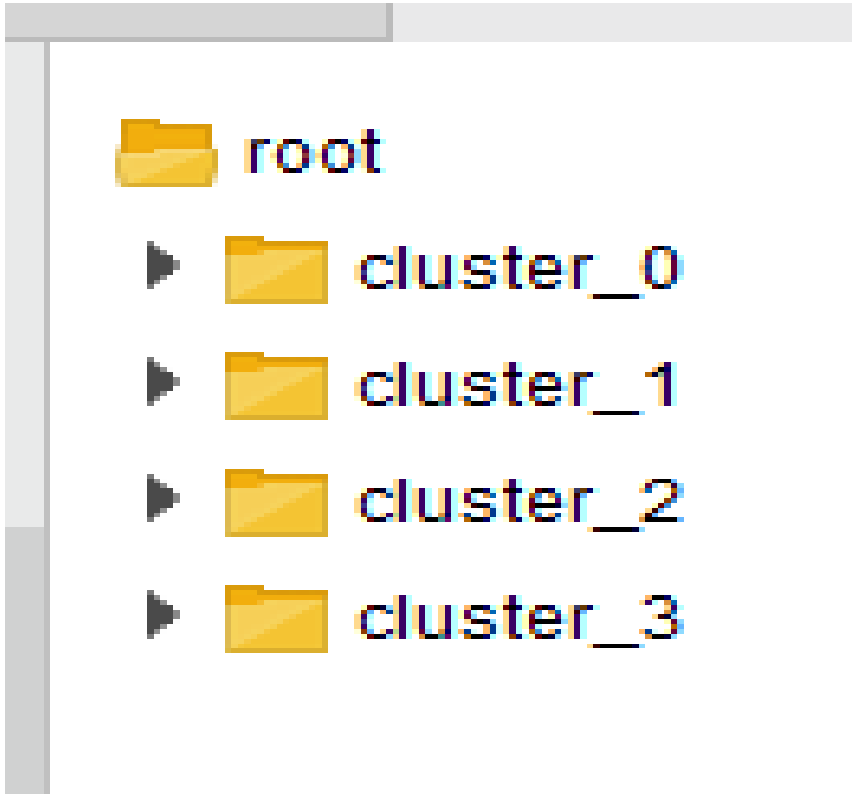


Centroid table (θα εστιάσουμε στα στοιχεία μόνο της cluster_2)

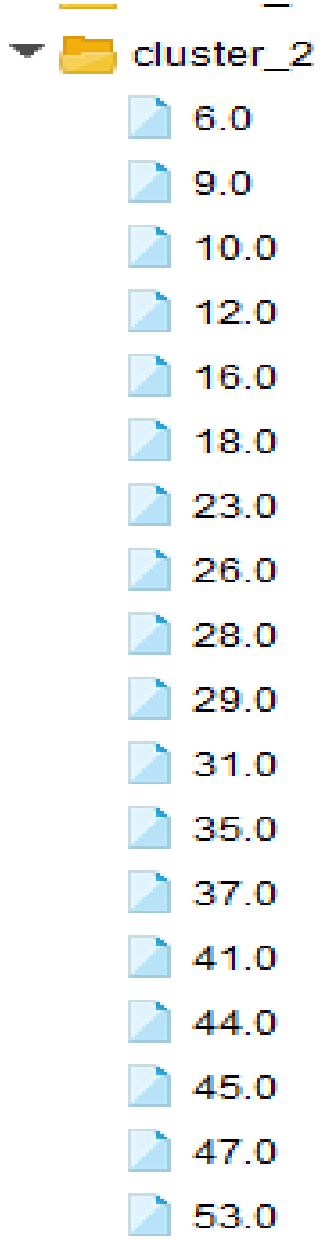
Attribute	cluster_0	cluster_1	cluster_2	cluster_3
Weight	152.093	106.850	184.318	127.726
Cholesterol	185.907	119.536	218.916	154.385
Gender	0.441	0.543	0.591	0.459


0
1
2
3
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
Weight
Cholesterol
Gender

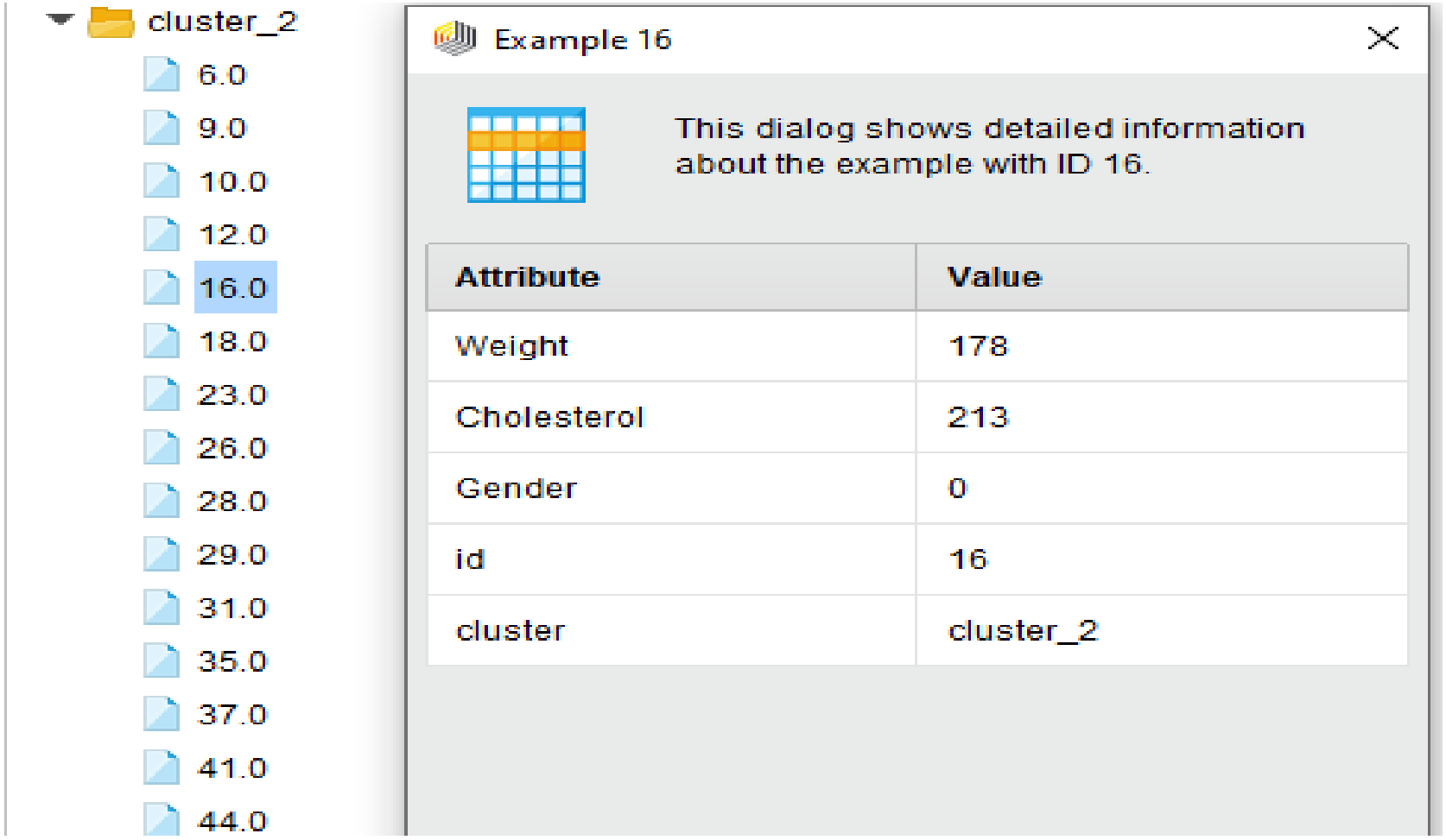
Folder view



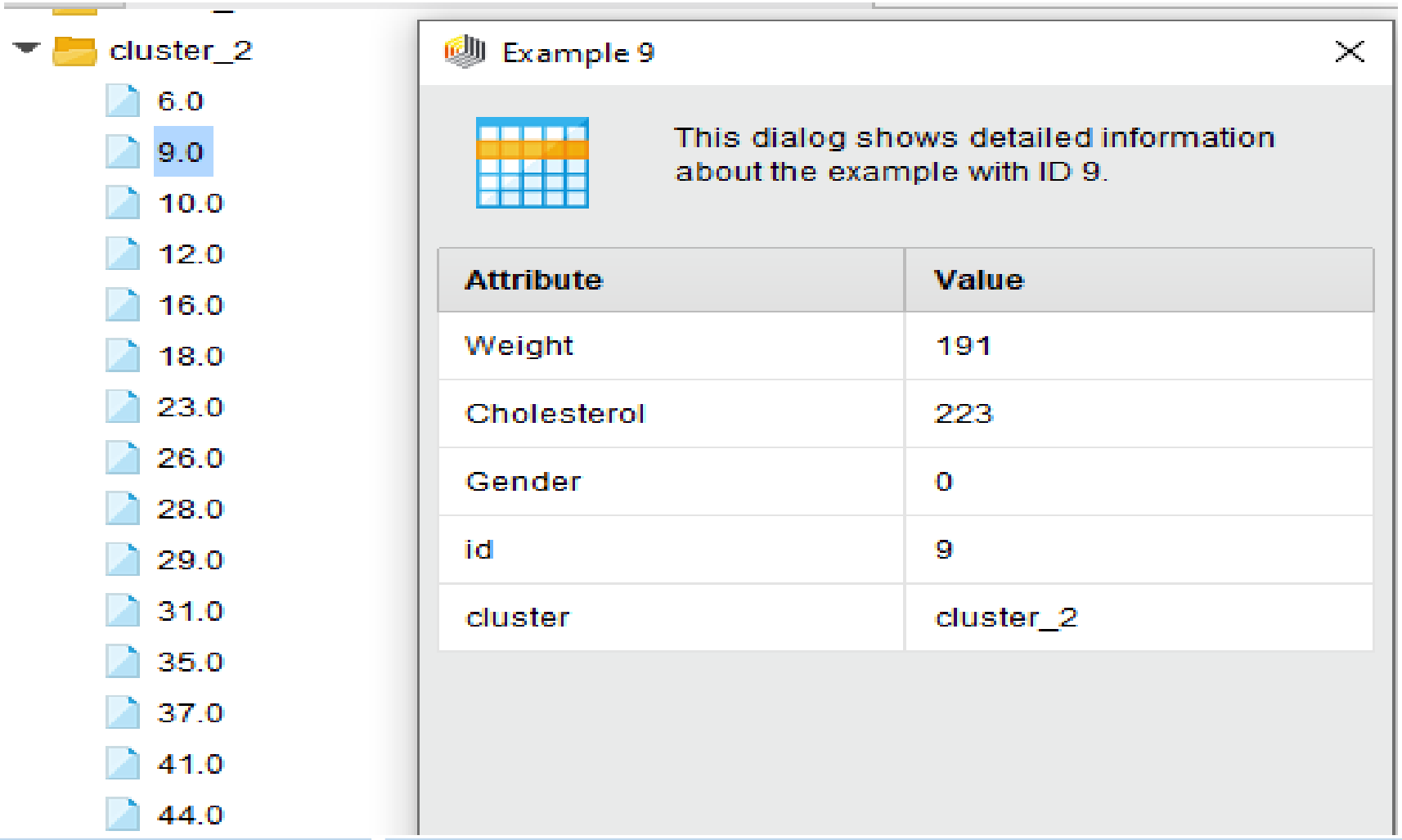
cluster_2


cluster_2
6.0
9.0
10.0
12.0
16.0
18.0
23.0
26.0
28.0
29.0
31.0
35.0
37.0
41.0
44.0
45.0
47.0
53.0

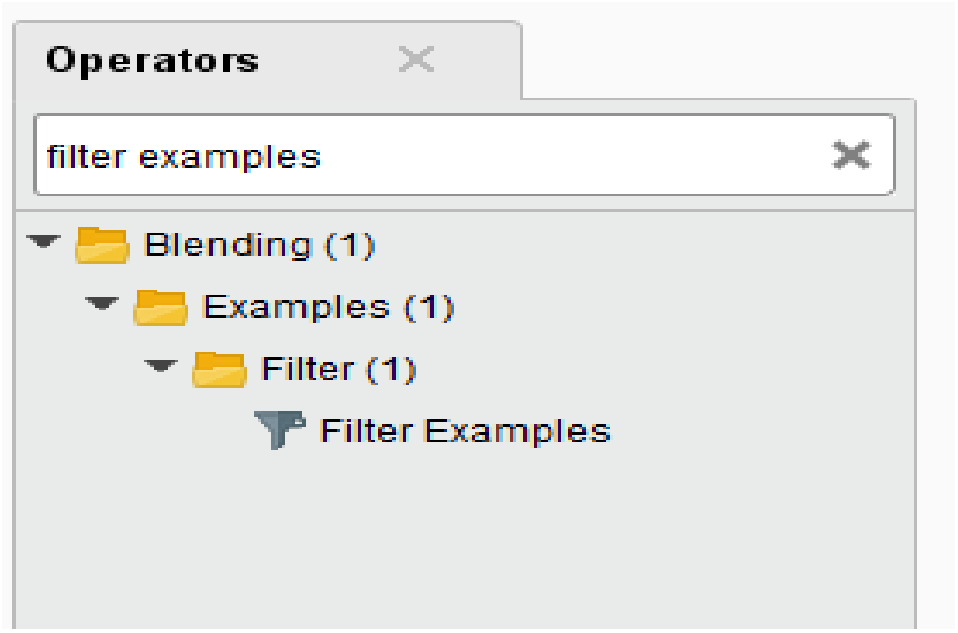
κάτω από μέσο όρο συστάδας



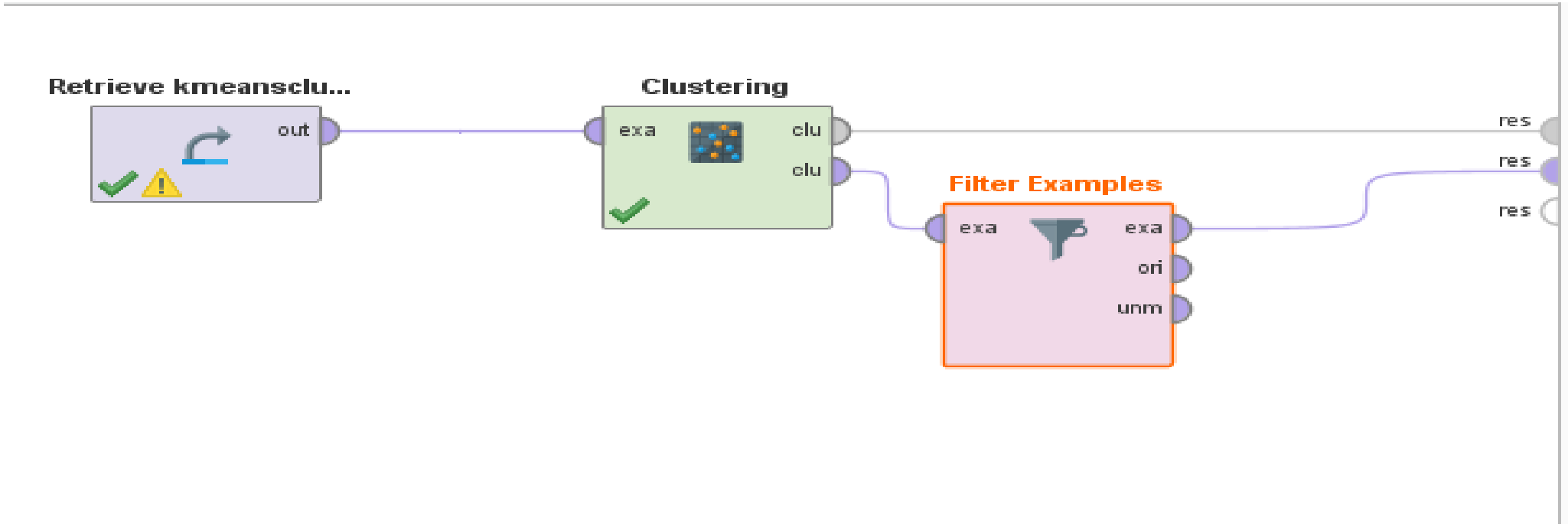
Ενδιαφέρει περισσότερο



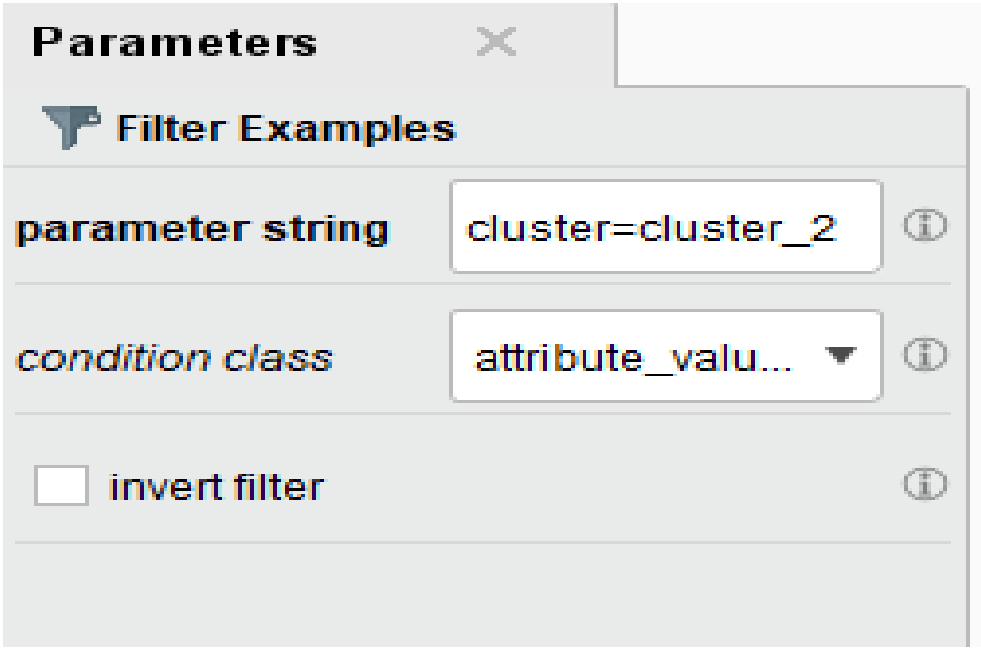
Filter


Operators
filter examples
Blending (1)
Examples (1)
Filter (1)
Filter Examples

Filter Examples



Parameters

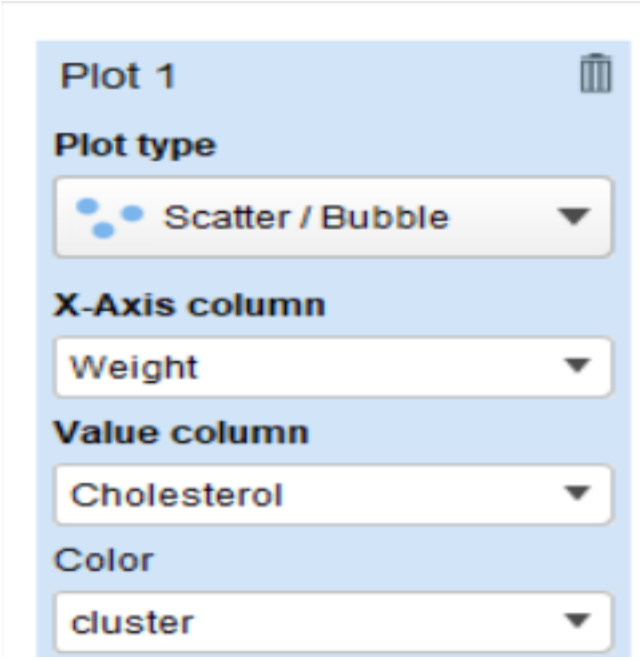

Parameters
Filter Examples
parameter string
cluster=cluster_2
condition class
attribute_valu...
invert filter

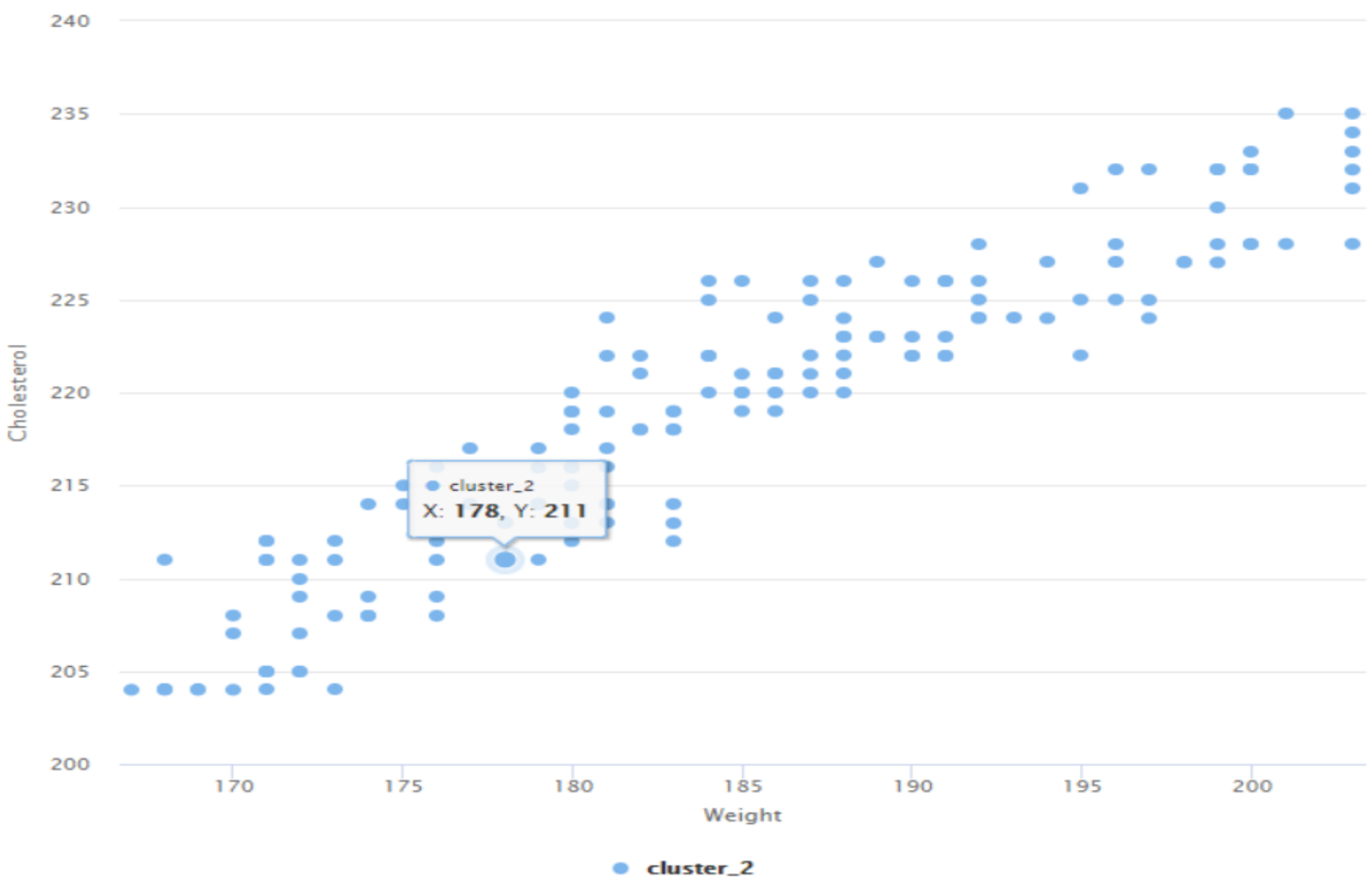
Data view: Filtered results for cluster 2 observations

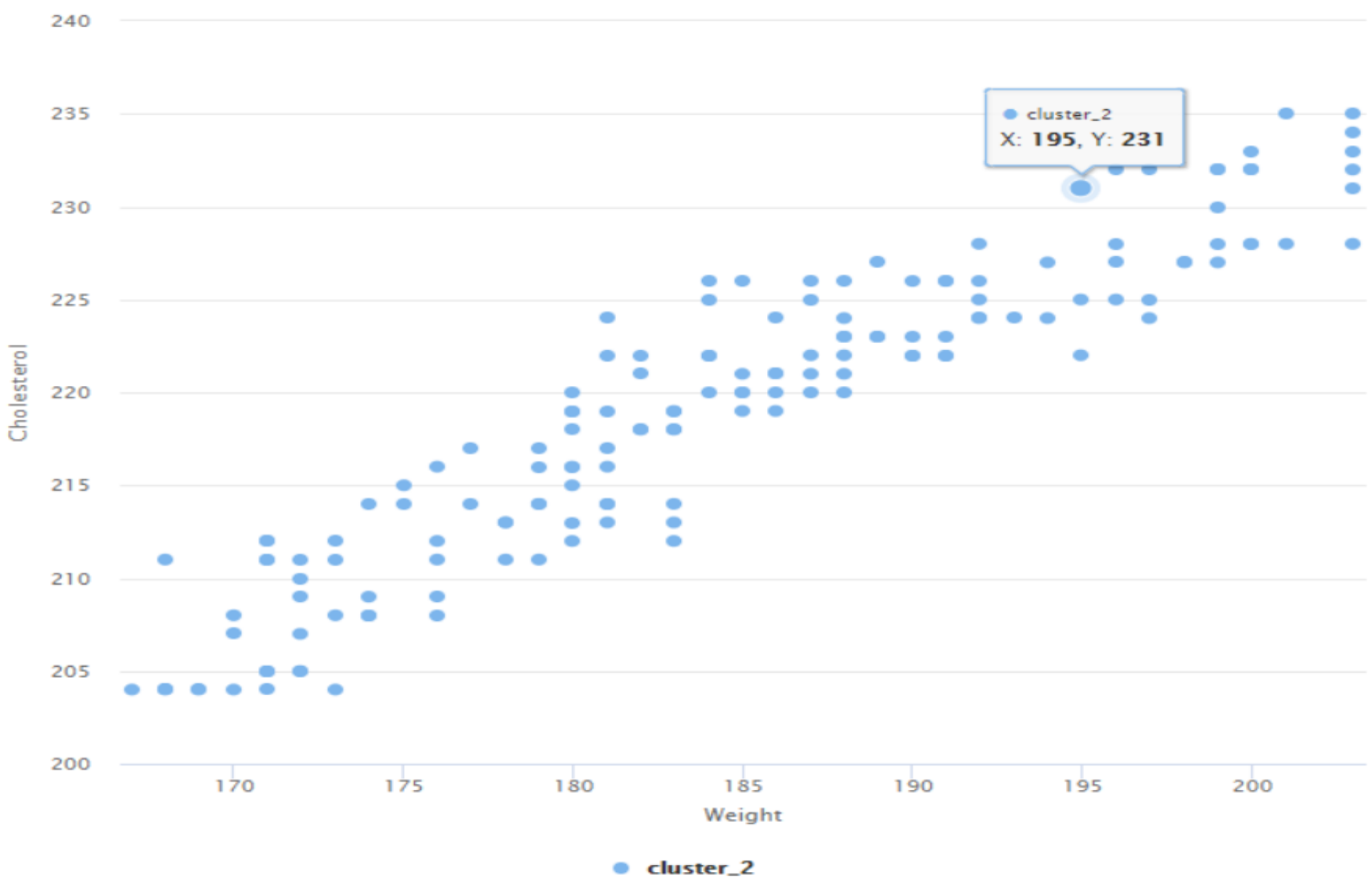
Row No.	id	cluster	Weight	Cholesterol	Gender
1	6	cluster_2	198	227	1
2	9	cluster_2	191	223	0
3	10	cluster_2	186	221	1
4	12	cluster_2	188	222	1
5	16	cluster_2	178	213	0
6	18	cluster_2	168	204	1
7	23	cluster_2	199	228	1
8	26	cluster_2	183	218	0
9	28	cluster_2	190	222	0
10	29	cluster_2	174	208	1
11	31	cluster_2	169	204	1
12	35	cluster_2	178	213	0
13	37	cluster_2	195	225	1
14	41	cluster_2	197	225	1
15	44	cluster_2	193	224	0

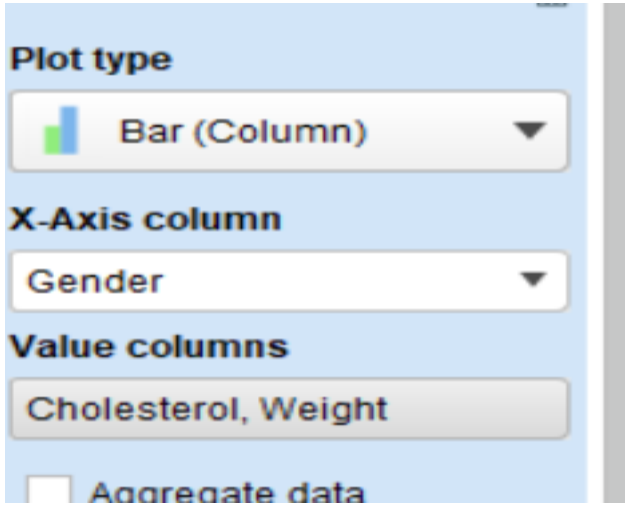
ExampleSet (154 examples, 2 special attributes, 3 regular attributes)

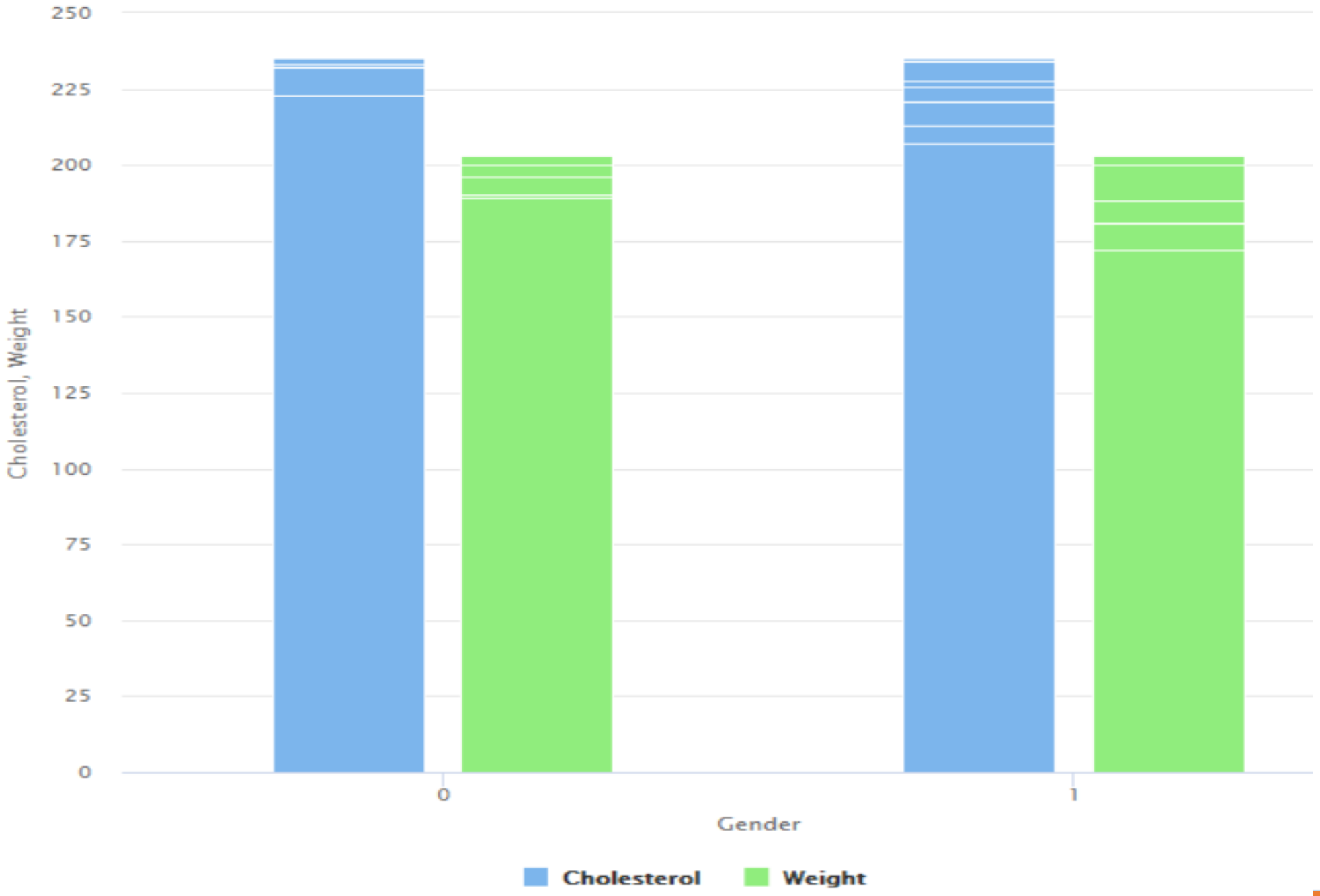
Name	Type	Missing	Statistics			Filter (5 / 5 attributes): Search for Attributes
Id **id**	Integer	0	Min 6	Max 543	Average 271.727	
Cluster **cluster**	Nominal	0	Least cluster_3 (0)	Most cluster_2 (154)	Values cluster_2 (154), cluster_0 (0), ...[2 more]	
Weight	Integer	0	Min 167	Max 203	Average 184.318	
Gender	Integer	0	Min 0	Max 1	Average 0.591	
Cholesterol	Integer	0	Min 204	Max 235	Average 218.916	

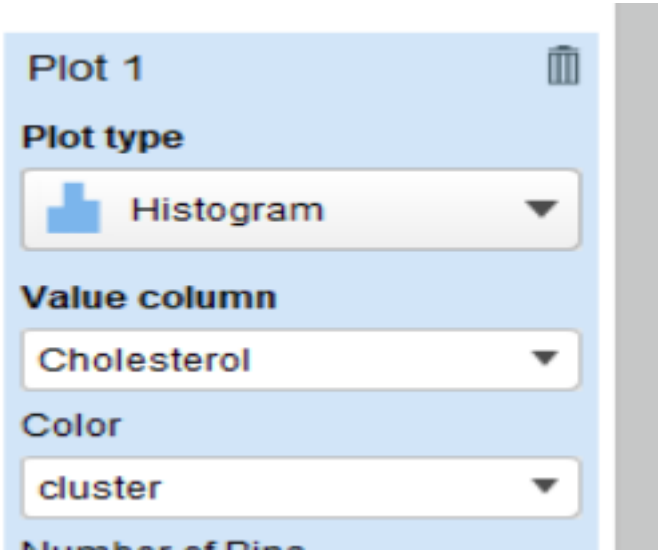

Plot 1
Plot type
Scatter / Bubble
X-Axis column
Weight
Value column
Cholesterol
Color
cluster

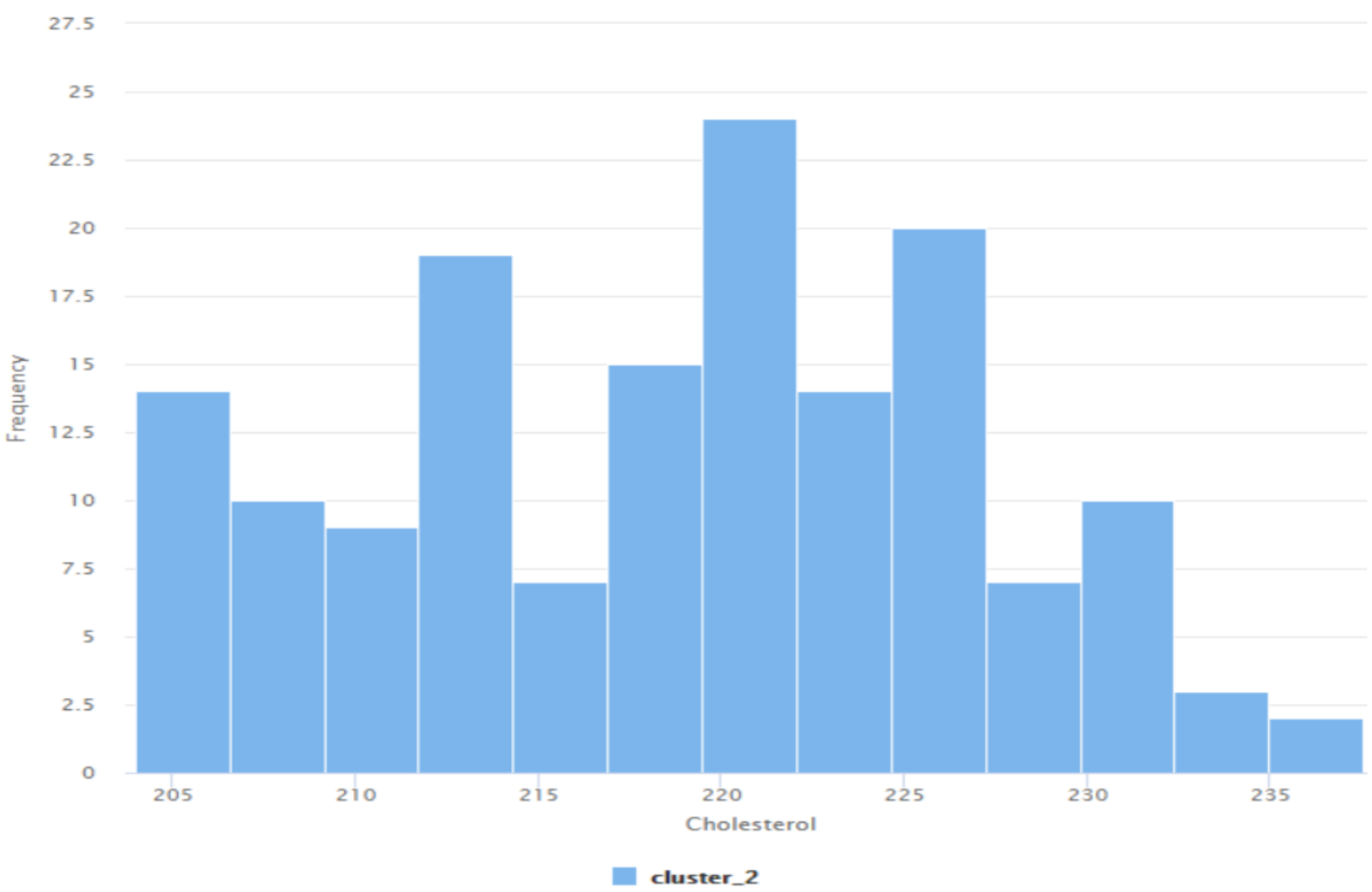

240
235
230
225
220
215
210
205
200
Cholesterol
170
175
180
185
190
195
200
Weight
cluster_2
X: 178, Y: 211
cluster_2

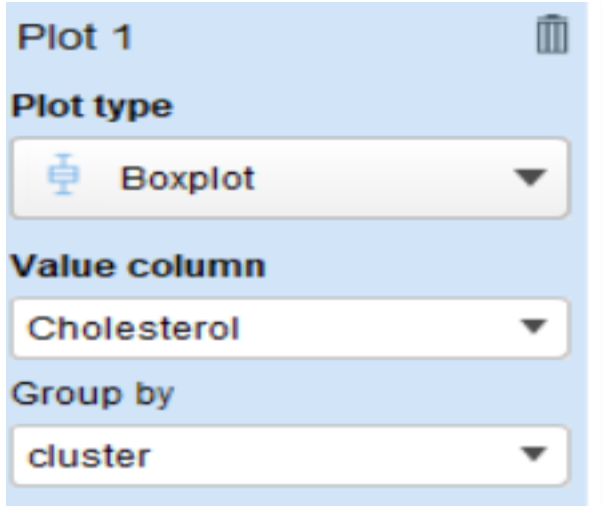

cluster_2
X: 195, Y: 231
240
235
230
225
220
215
210
205
200
Cholesterol
170
175
180
185
190
195
200
Weight
cluster_2

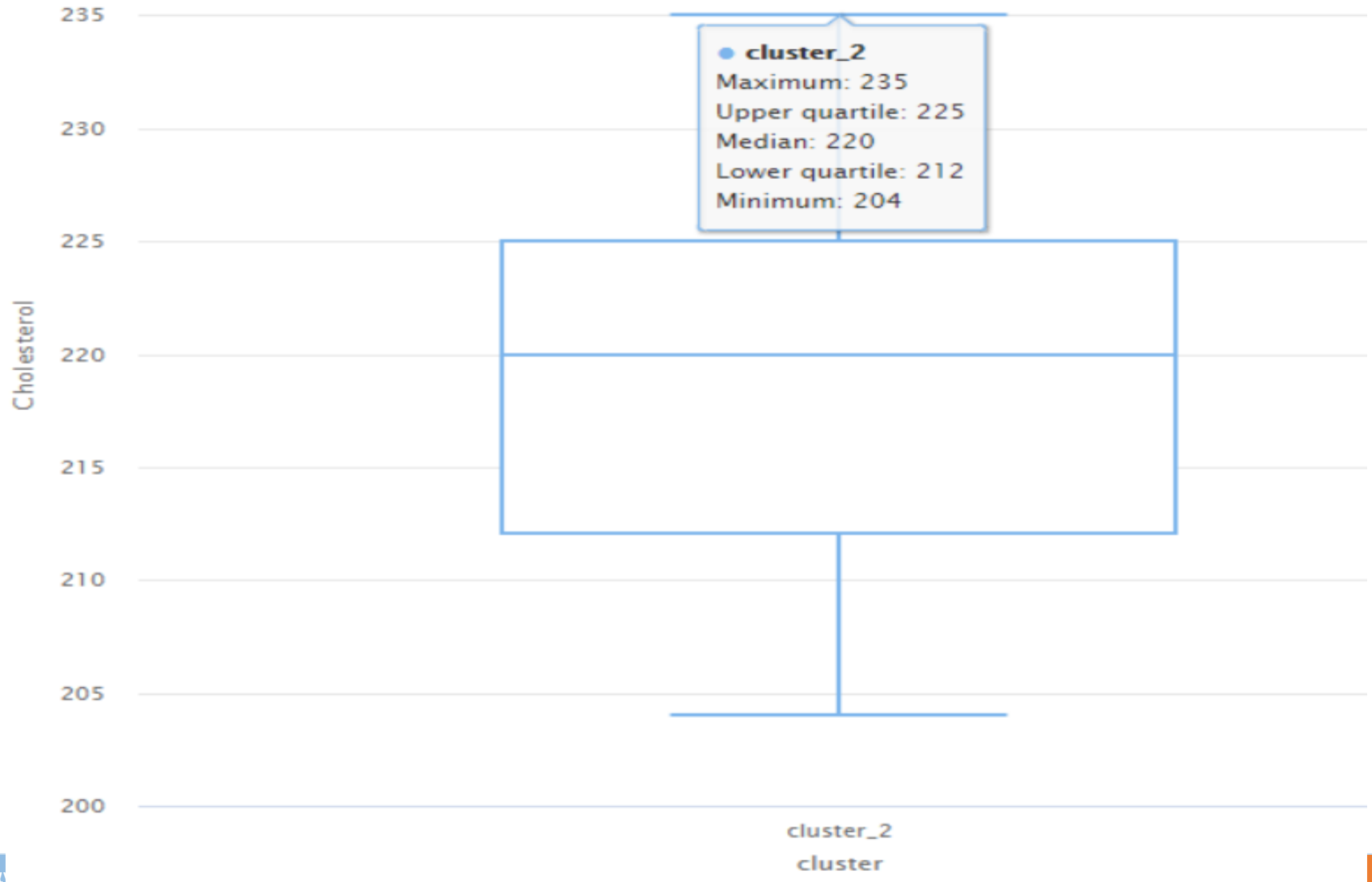

Plot type
Bar (Column)
X-Axis column
Gender
Value columns
Cholesterol, Weight
Aggregate data

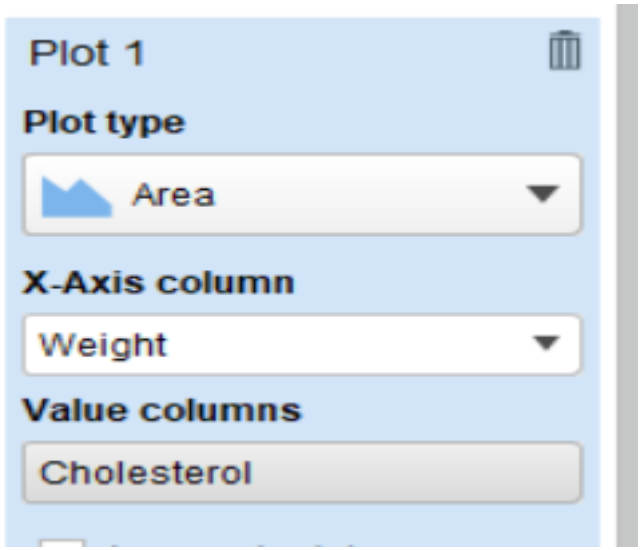

250
225
200
175
150
125
100
75
50
25
0
Cholesterol, Weight
0
1
Gender
Cholesterol
Weight

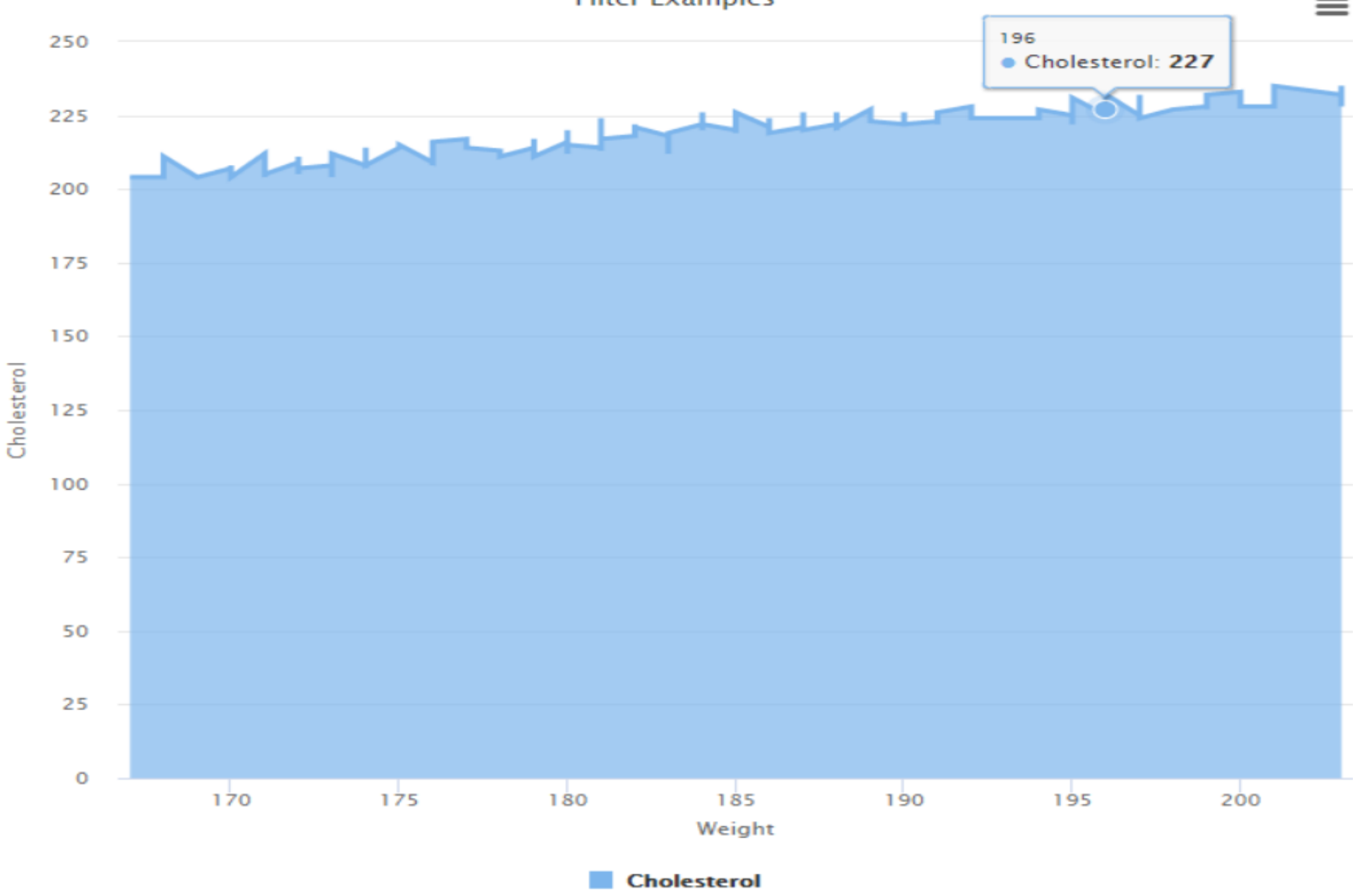

Plot 1
Plot type
Histogram
Value column
Cholesterol
Color
cluster

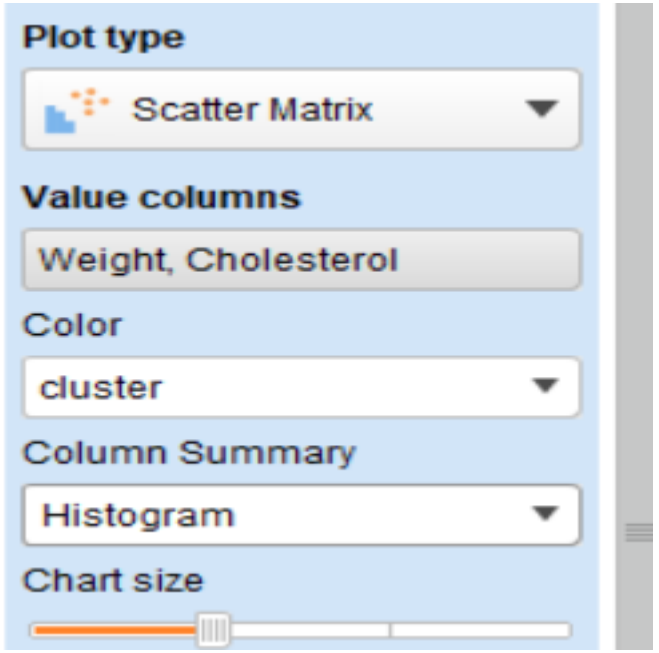

27.5
25
22.5
20
17.5
15
12.5
10
7.5
5
2.5
0
Frequency
205
210
215
220
225
230
235
Cholesterol
cluster_2

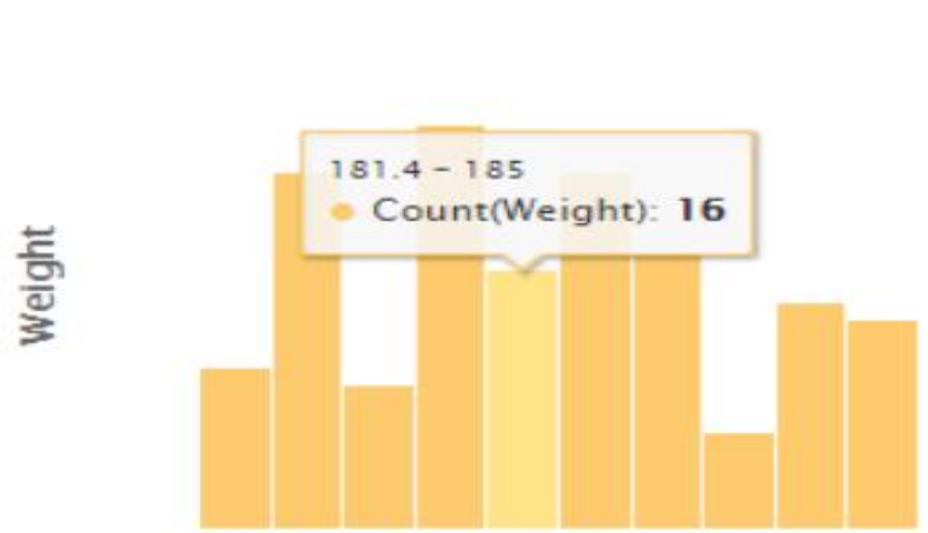

Plot 1
Plot type
Boxplot
Value column
Cholesterol
Group by
cluster

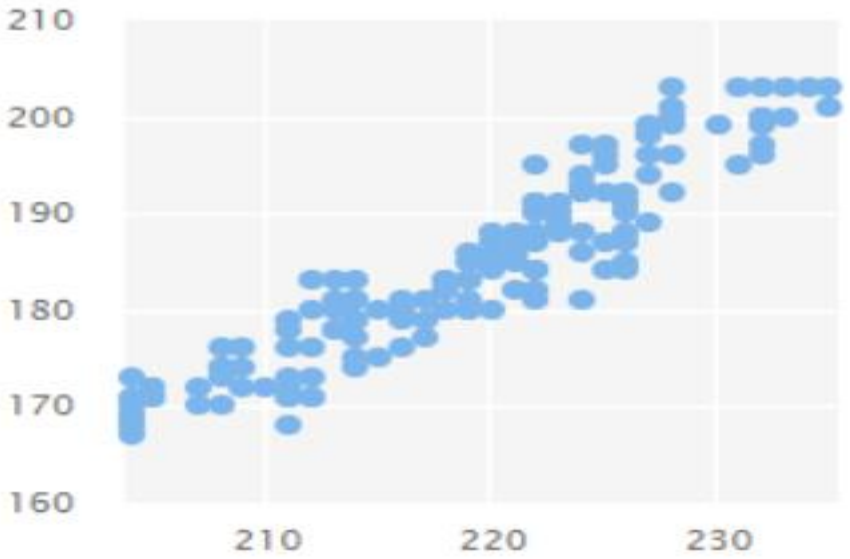

235
230
225
220
215
210
205
200
Cholesterol
cluster_2
Maximum: 235
Upper quartile: 225
Median: 220
Lower quartile: 212
Minimum: 204
cluster_2
cluster

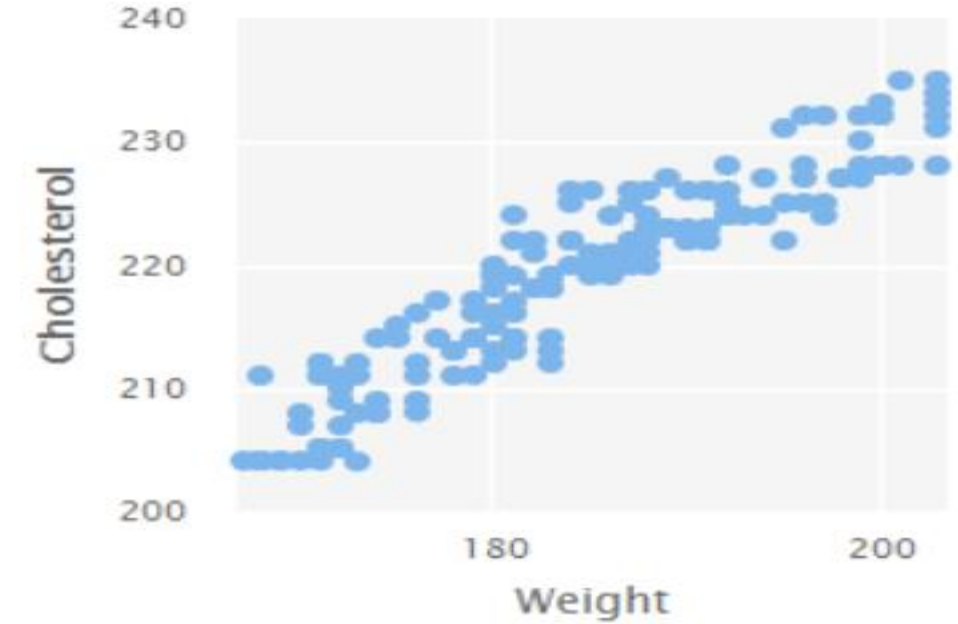

Plot 1
Plot type
Area
X-Axis column
Weight
Value columns
Cholesterol


196
Cholesterol: 227
250
225
200
175
150
125
100
75
50
25
0
Cholesterol
170
175
180
185
190
195
200
Weight
Cholesterol


Plot type
Scatter Matrix
Value columns
Weight, Cholesterol
Color
cluster
Column Summary
Histogram
Chart size


Weight
181.4 – 185
Count(Weight): 16


210
200
190
180
170
160
210
220
230


Cholesterol
240
230
220
210
200
180
200
Weight

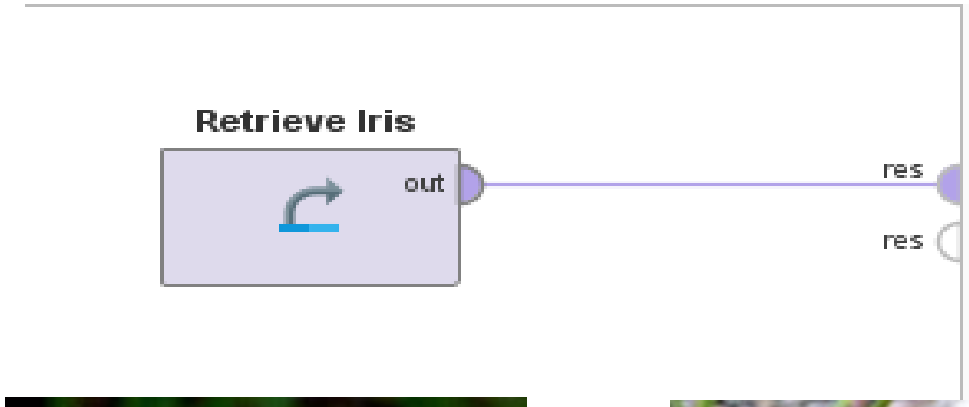

Cholesterol

Iris Data Set

Συστάδες (Clustering) του γνωστού συνόλου δεδομένων Iris

Το Rapidminer προσφέρει ένα σχετικό Tutorial Process. Εδώ θα δούμε μια απλοποιημένη μορφή.

“The data set consists of 50 samples from each of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample: the length and the width of the sepals and petals, in centimeters” (Wikipedia).



Iris setosa

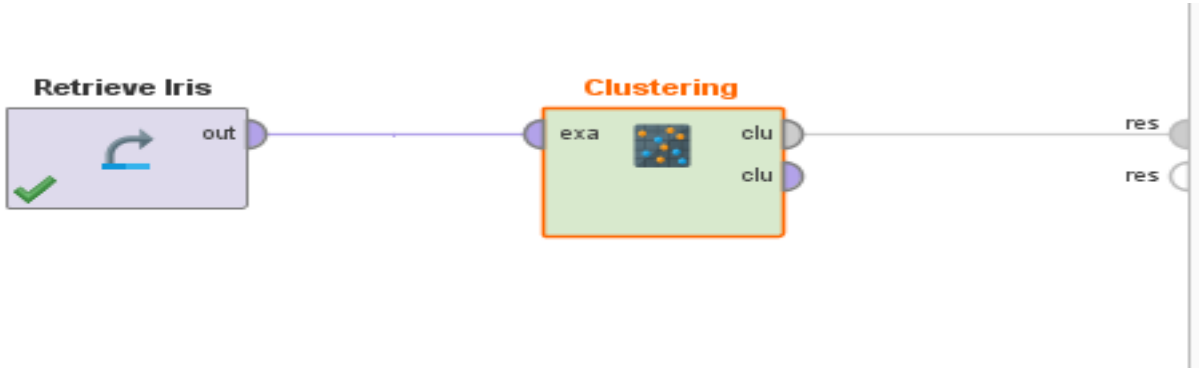
Iris virginica

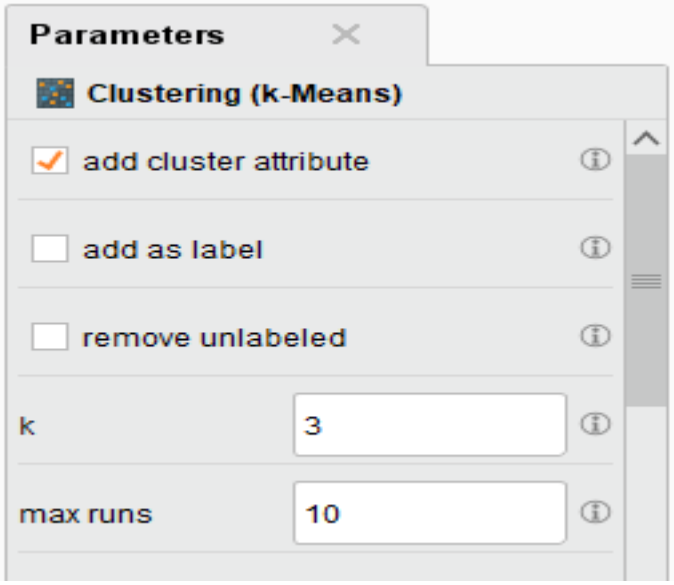
Iris versicolor

Row No.	id	label	a1	a2	a3	a4
1	id_1	Iris-setosa	5.100	3.500	1.400	0.200
2	id_2	Iris-setosa	4.900	3	1.400	0.200
3	id_3	Iris-setosa	4.700	3.200	1.300	0.200
4	id_4	Iris-setosa	4.600	3.100	1.500	0.200
5	id_5	Iris-setosa	5	3.600	1.400	0.200
6	id_6	Iris-setosa	5.400	3.900	1.700	0.400
7	id_7	Iris-setosa	4.600	3.400	1.400	0.300
8	id_8	Iris-setosa	5	3.400	1.500	0.200
9	id_9	Iris-setosa	4.400	2.900	1.400	0.200
10	id_10	Iris-setosa	4.900	3.100	1.500	0.100
11	id_11	Iris-setosa	5.400	3.700	1.500	0.200
12	id_12	Iris-setosa	4.800	3.400	1.600	0.200
13	id_13	Iris-setosa	4.800	3	1.400	0.100
14	id_14	Iris-setosa	4.300	3	1.100	0.100
15	id_15	Iris-setosa	5.800	4	1.200	0.200

ExampleSet (150 examples, 2 special attributes, 4 regular attributes)

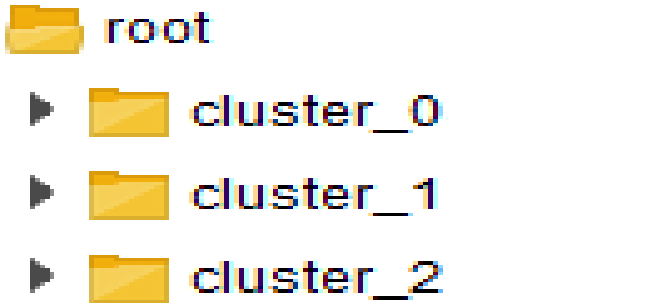
Name	Type	Missing	Statistics		Filter (6 / 6 attributes): Search for Attributes
Id **id**	Nominal	0	Least: id_99 (1)	Most: id_1 (1)	Values: id_1 (1), id_10 (1), ...[148 more]
Label **label**	Nominal	0	Least: Iris-virginica (50)	Most: Iris-setosa (50)	Values: Iris-setosa (50), Iris-versicolor (50), ...[1 more]
a1	Real	0	Min: 4.300	Max: 7.900	Average: 5.843
a2	Real	0	Min: 2	Max: 4.400	Average: 3.054
a3	Real	0	Min: 1	Max: 6.900	Average: 3.759
a4	Real	0	Min: 0.100	Max: 2.500	Average: 1.199

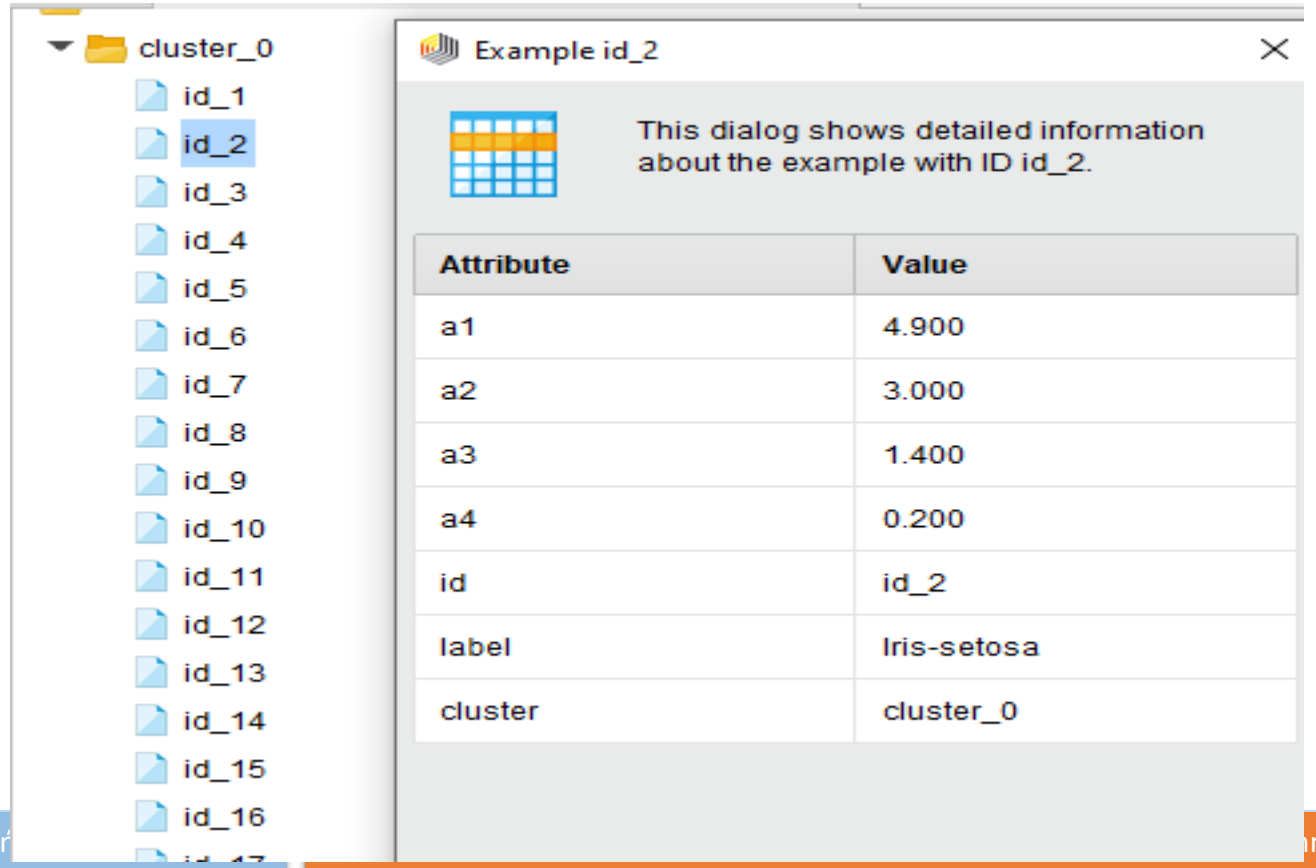

Retrieve Iris
out
Clustering
exa
clu
clu
res
res

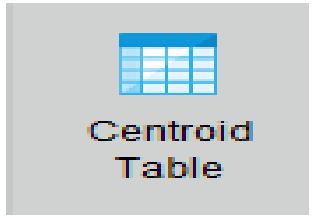

Parameters
Clustering (k-Means)
add cluster attribute
add as label
remove unlabeled
k
3
max runs
10

Cluster Model

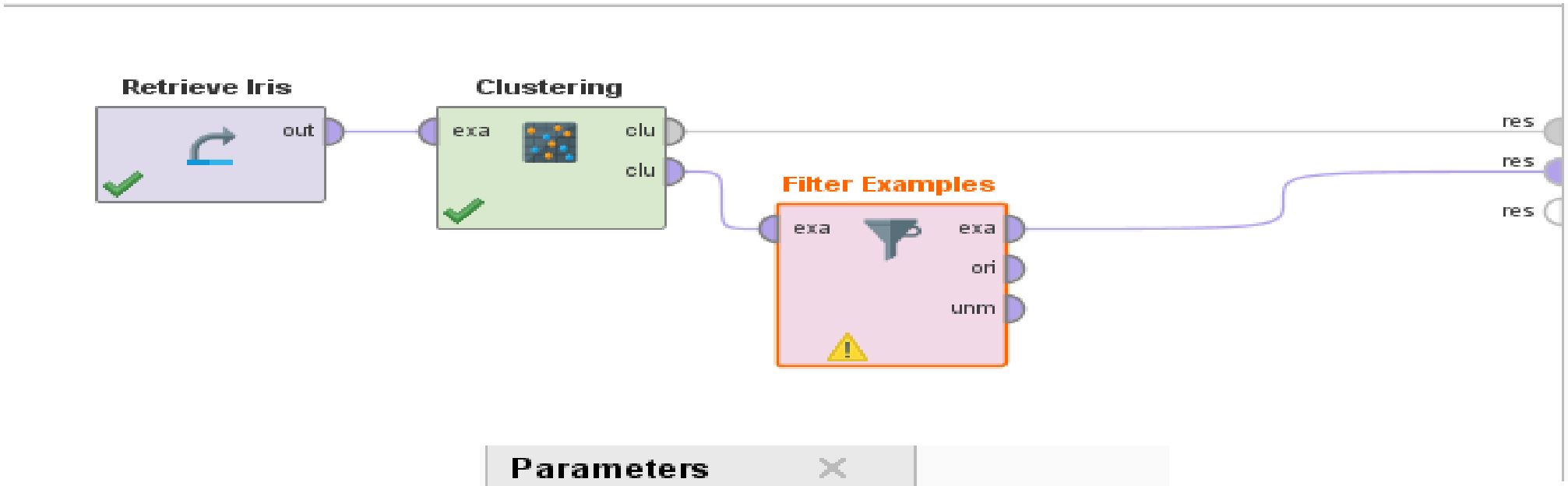
```
Cluster 0: 50 items
Cluster 1: 39 items
Cluster 2: 61 items
Total number of items: 150
```

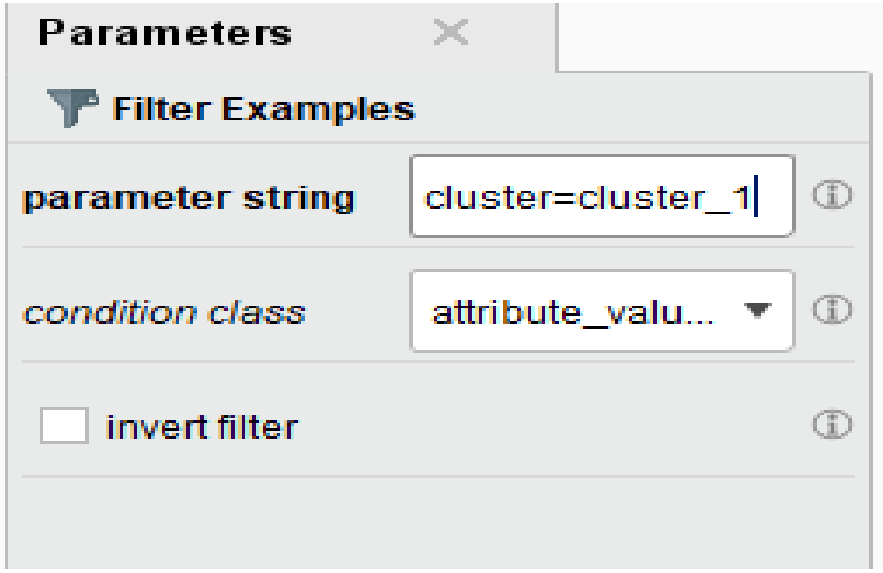






Attribute	cluster_0	cluster_1	cluster_2
a1	5.006	6.854	5.884
a2	3.418	3.077	2.741
a3	1.464	5.715	4.389
a4	0.244	2.054	1.434


Retrieve Iris
out
Clustering
exa
clu
clu
Filter Examples
exa
exa
ori
unm
res
res
res


Parameters
Filter Examples
parameter string
cluster=cluster_1
condition class
attribute_valu...
invert filter

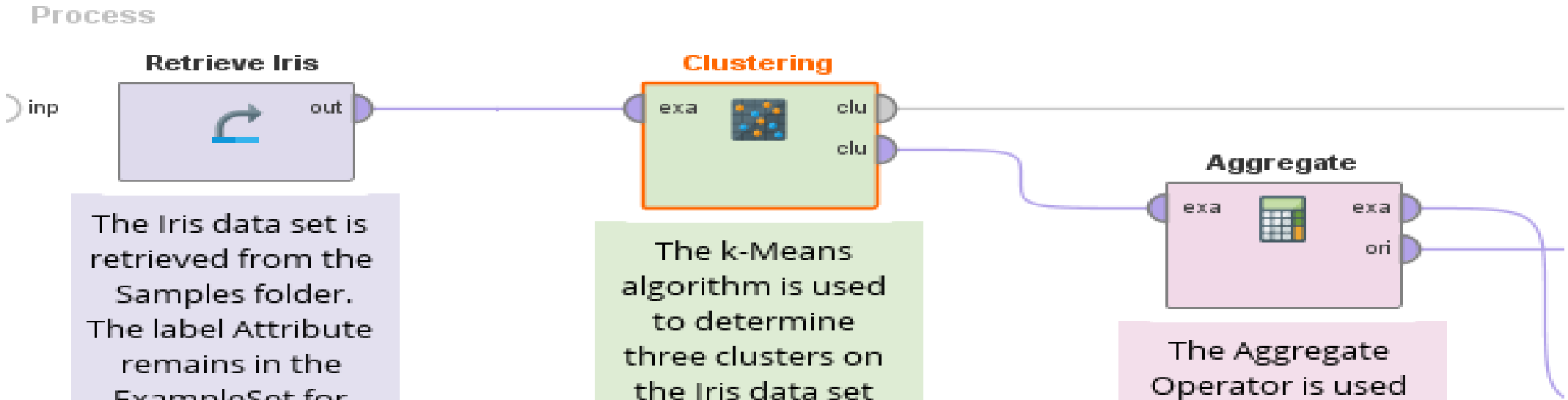
Row No.	id	label	cluster	a1	a2	a3	a4
1	id_51	Iris-versicolor	cluster_1	7	3.200	4.700	1.400
2	id_53	Iris-versicolor	cluster_1	6.900	3.100	4.900	1.500
3	id_78	Iris-versicolor	cluster_1	6.700	3	5	1.700
4	id_101	Iris-virginica	cluster_1	6.300	3.300	6	2.500
5	id_103	Iris-virginica	cluster_1	7.100	3	5.900	2.100
6	id_104	Iris-virginica	cluster_1	6.300	2.900	5.600	1.800
7	id_105	Iris-virginica	cluster_1	6.500	3	5.800	2.200
8	id_106	Iris-virginica	cluster_1	7.600	3	6.600	2.100
9	id_108	Iris-virginica	cluster_1	7.300	2.900	6.300	1.800
10	id_109	Iris-virginica	cluster_1	6.700	2.500	5.800	1.800
11	id_110	Iris-virginica	cluster_1	7.200	3.600	6.100	2.500
12	id_111	Iris-virginica	cluster_1	6.500	3.200	5.100	2
13	id_112	Iris-virginica	cluster_1	6.400	2.700	5.300	1.900
14	id_113	Iris-virginica	cluster_1	6.800	3	5.500	2.100
15	id_116	Iris-virginica	cluster_1	6.400	3.200	5.300	2.300

ExampleSet (39 examples, 3 special attributes, 4 regular attributes)

Με όμοιο τρόπο βλέπουμε τα αποτελέσματα της συσταδοποίησης

Cluster_0	50 examples Iris_serosa	
Cluster_1	36 examples Iris_virginica	3 examples Iris_versicolor
Cluster_2	47 examples Iris_versicolor	14 examples Iris_virginica

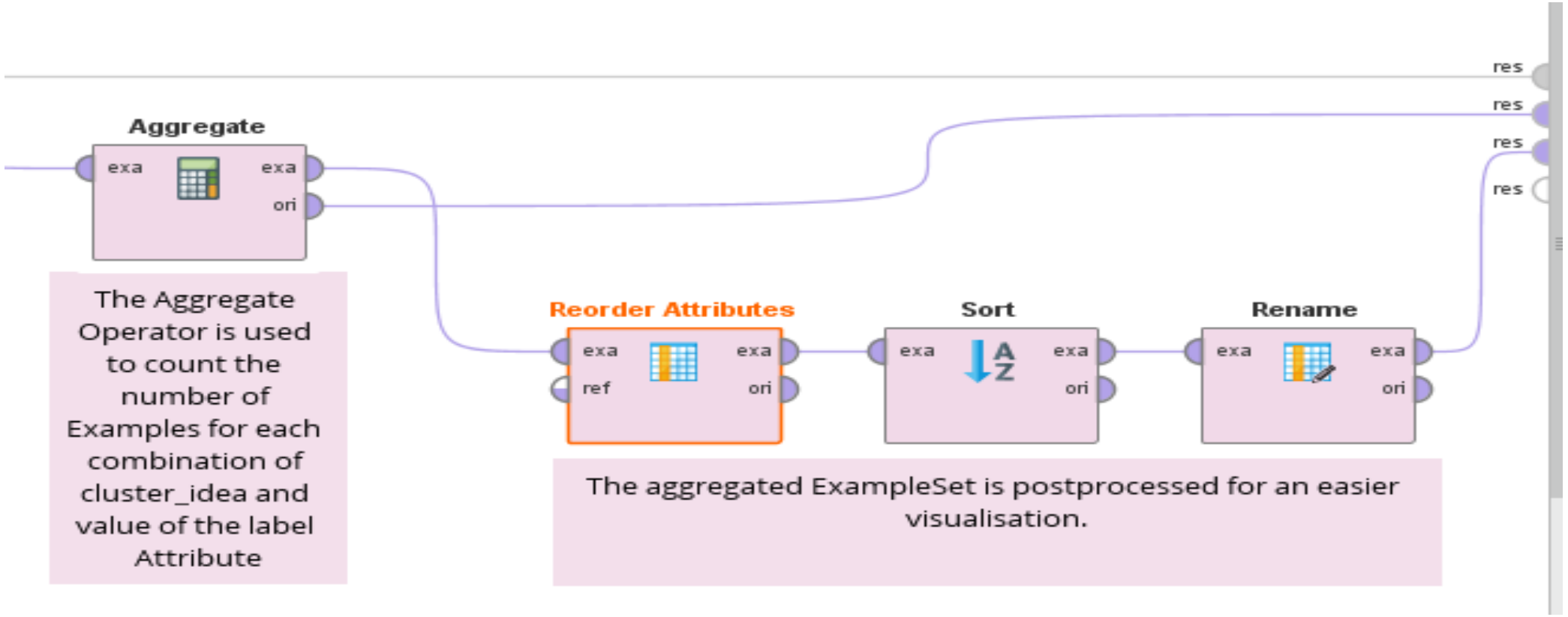
Από το tutorial του RapidMiner



The Iris data set is retrieved from the Samples folder. The label Attribute remains in the ExampleSet for comparison the results of the Clustering. It is not used in the Clustering itself.

The k-Means algorithm is used to determine three clusters on the Iris data set and assign each Example to one cluster.

The Aggregate Operator is used to count the number of Examples for each combination of cluster_idea and value of the label Attribute

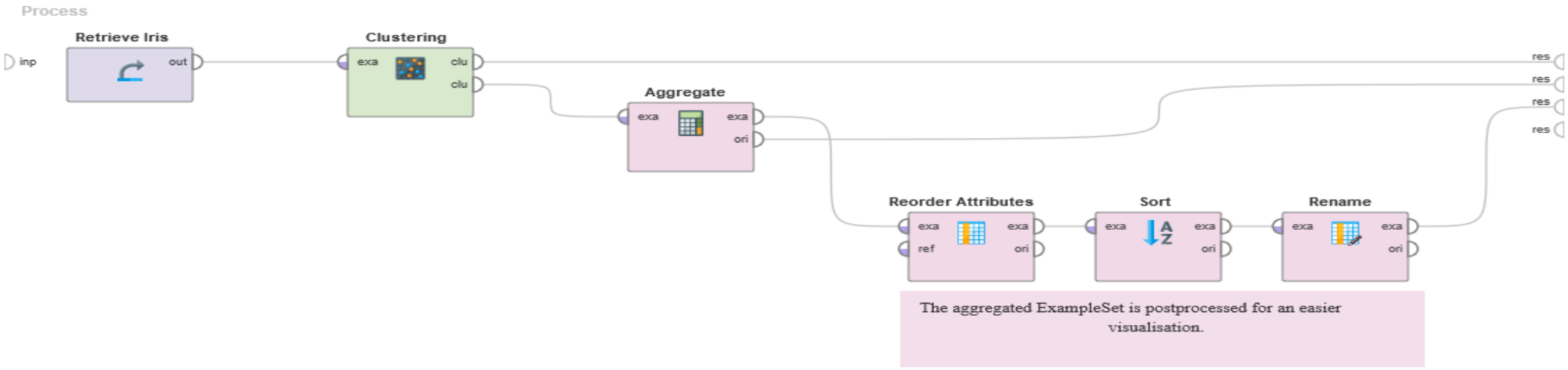

Aggregate
exa
exa
ori
res
res
res
res
The Aggregate Operator is used to count the number of Examples for each combination of cluster_idea and value of the label Attribute
Reorder Attributes
exa
ref
exa
ori
Sort
exa
exa
ori
Rename
exa
exa
ori
The aggregated ExampleSet is postprocessed for an easier visualisation.

Look into the results of the process:
ExampleSet (Rename):
- cluster_0 consist mainly of iris_virginica Examples (36) with only a few (3) iris_versicolor Examples
- cluster_1 consists completely of iris_setosa Examples (50). Also iris_setosa Example cannot be found in other clusters.
- cluster_2 consists most of iris_versicolor Examples (47) but with also some (14) iris_virginica Examples

ExampleSet (Clustering):
- You can visualize the assignment of the Examples to the clusters by using the 'Scatter' Chart, plotting two of the Attributes a1,a2,a3,a4 on x-and y-axis and the cluster Attribute as Color Column

Cluster Model (Clustering):
- The Cluster Model consist information which Example is assigned to which cluster
- the size of the clusters can be visualized as a graph
- the position of the centroids is listed

Row No.	label	cluster	count
1	Iris-versicolor	cluster_0	3
2	Iris-virginica	cluster_0	36
3	Iris-setosa	cluster_1	50
4	Iris-versicolor	cluster_2	47
5	Iris-virginica	cluster_2	14



Look into the results of the process:
ExampleSet (Rename):
- cluster_0 consist mainly of iris_virginica Examples (36) with only a few (3) iris_versicolor Examples
- cluster_1 consists completely of iris_setosa Examples (50). Also iris_setosa Example cannot be found in other clusters.
- cluster_2 consists most of iris_versicolor Examples (47) but with also some (14) iris_virginica Examples

ExampleSet (Clustering):
- You can visualize the assignment of the Examples to the clusters by using the 'Scatter' Chart, plotting two of the Attributes a1,a2,a3,a4 on x-and y-axis and the cluster Attribute as Color Column

Cluster Model (Clustering):
- The Cluster Model consist information which Example is assigned to which cluster
- the size of the clusters can be visualized as a graph
- the position of the centroids is listed

ΤΕΛΟΣ