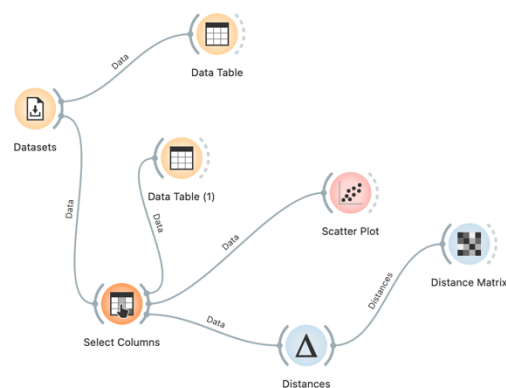


Muhammad Raflika Dwiyanasyah

1103210211

Euclidean Clustering Using ODM

Clustering berdasarkan jarak Euclidean menggunakan metode ODM (Oracle Data Mining). Jarak Euclidean digunakan untuk mengukur kedekatan antar data dalam ruang multidimensi. Biasanya, hasilnya divisualisasikan dalam grafik atau diagram untuk menunjukkan grup-grup yang terbentuk.



Ignored (5)

Filter

French

History

Biology

Physics

Physical

Features (2)

Filter

English

Algebra

Target

Metas (1)

Student

Reset

☐ Ignore new variables by default

☒ Send Automatically

16 | 16 | 2

Info

16 instances (no missing data)
2 features
No target variable.
1 meta attribute

Variables

☒ Show variable labels (if present)
☐ Visualize numeric values
☒ Color by instance classes

Selection

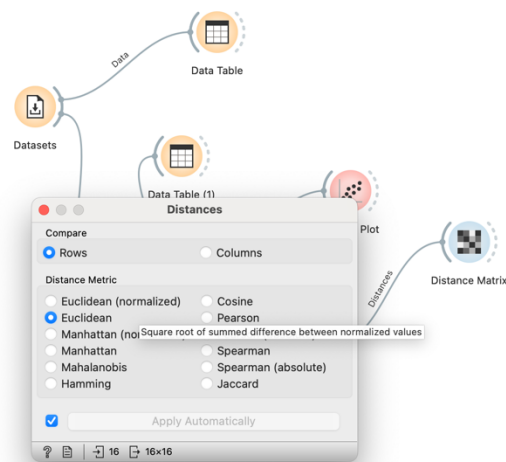
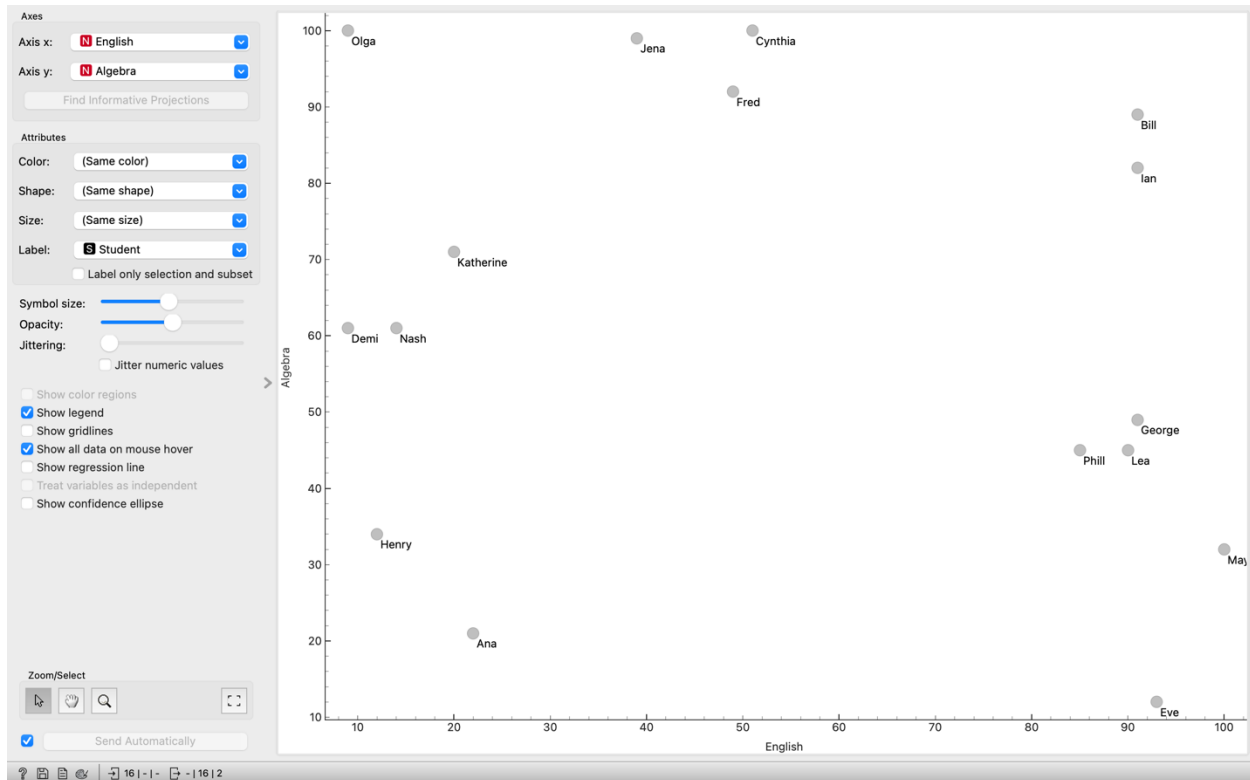
☒ Select full rows

Restore Original Order

☒ Send Automatically

16 | 1 | 16

	Student	English	Algebra
1	Ana	22	21
2	Bill	91	89
3	Cynthia	51	100
4	Demi	9	61
5	Eve	93	12
6	Fred	49	92
7	George	91	49
8	Henry	12	34
9	Ian	91	82
10	Jena	39	99
11	Katherine	20	71
12	Lea	90	45
13	Maya	100	32
14	Nash	14	61
15	Olga	9	100
16	Phill	85	45



	Ana	Bill	Cynthia	Demi	Eve	Fred	George	Henry	Ian	Jena	Katherine	Lea	Maya	Nash	Olga	Phill
Ana		96,876	84,155	42,059	71,568	75,961	74,465	16,401	92,098	79,831	50,040	72,111	78,772	40,792	80,062	67,417
Bill	96,876		41,485	86,649	77,026	42,107	40,000	96,260	7,000	52,953	73,246	44,011	57,706	81,933	82,735	44,407
Cynthia	84,155	41,485		57,315	97,509	8,246	64,815	76,662	43,863	12,042	42,450	67,424	83,815	53,759	42,000	64,661
Demi	42,059	86,649	57,315		97,247	50,606	82,873	27,166	84,646	48,415	14,866	82,565	95,509	5,000	39,000	77,666
Eve	71,568	77,026	97,509	97,247		91,302	37,054	83,934	70,029	102,396	93,862	33,136	21,190	92,962	121,655	33,956
Fred	75,961	42,107	8,246	50,606	91,302		60,108	68,797	43,174	12,207	35,805	62,370	78,746	46,755	40,792	59,203
George	74,465	40,000	64,815	82,873	37,054	60,108		80,411	33,000	72,139	74,330	4,123	19,235	77,929	96,566	7,211
Henry	16,401	96,260	76,662	27,166	83,934	68,797	80,411		92,439	70,385	37,855	78,772	88,023	27,074	66,068	73,824
Ian	92,098	7,000	43,863	84,646	70,029	43,174	33,000	92,439		54,708	71,847	37,014	50,804	79,812	83,952	37,483
Jena	79,831	52,953	12,042	48,415	102,396	12,207	72,139	70,385	54,708		33,838	74,277	90,609	45,486	30,017	70,937
Katherine	50,040	73,246	42,450	14,866	93,862	35,805	74,330	37,855	71,847	33,838		74,673	89,000	11,662	31,016	70,007
Lea	72,111	44,011	67,424	82,565	33,136	62,370	4,123	78,772	37,014	74,277	74,673		16,401	77,666	97,908	5,000
Maya	78,772	57,706	83,815	95,509	21,190	78,746	19,235	88,023	50,804	90,609	89,000	16,401		90,758	113,600	19,849
Nash	40,792	81,933	53,759	5,000	92,962	46,755	77,929	27,074	79,812	45,486	11,662	77,666	90,758		39,319	72,780
Olga	80,062	82,735	42,000	39,000	121,655	40,792	96,566	66,068	83,952	30,017	31,016	97,908	113,600	39,319		93,814
Phill	67,417	44,407	64,661	77,666	33,956	59,203	7,211	73,824	37,483	70,937	70,007	5,000	19,849	72,780	93,814	

Labels: Student
Send Automatically

16x16

Hierarchical Clustering using ODM

metode *Hierarchical Clustering* dengan ODM. Dalam metode ini, data diorganisasikan dalam struktur hierarki berupa dendrogram. Gambar mungkin mencerminkan proses penggabungan atau pemisahan data secara bertahap berdasarkan kesamaan mereka.

Filter...

Data

File

CSV File Import

Datasets

SQL Table

Data Table

Paint Data

Data Info

Rank

Edit Domain

Color

Feature Statistics

Save Data

Transform

Data Sampler

Select Columns

Select Rows

Transpose

Merge Data

Concatenate

Select by Data Index

Unique

Aggregate Columns

Group by

Pivot Table

Apply Domain

Preprocess

Impute

Continuize

Discretize

Select a widget to show its description.

See [workflow examples](#), [YouTube tutorials](#), or open the [welcome screen](#).

#

T

||

?

```

graph LR
    Datasets -- Data --> DataTable[Data Table]
    DataTable -- Data --> Distances
    Distances -- Distances --> HierarchicalClustering[Hierarchical Clustering]
    HierarchicalClustering -- Data --> BoxPlot[Box Plot]
    DataTable -- Data --> BoxPlot
  
```

Search for data set ...

Show data sets in English

Title	Size	Instances	Variables	Target	Tags
Course Grades	9.2 KB	16	8	none	synthetic, education
BBC3	2.6 MB	1407	3	categorical	text, classification, news
Dendritic cells and monocytes in human...	19.4 MB	1140	26595	categorical	expression, human, homo-sapiens, blood
Dendritic cells and monocytes in human...	18.1 MB	1244	26595	categorical	expression, human, homo-sapiens, blood
Breast Cancer and Docetaxel Treatment	1.8 MB	24	9486	categorical	biology
Smoking effect on B lymphocytes	1.8 MB	79	3001	categorical	genomics
HDI	45.2 KB	188	54	none	economy, geo
ParlaMint	1.7 MB	1000	18	categorical	text, classification, time, politics
SentiNews	5.0 MB	2000	8	categorical	text, sentiment
TKI resistance	1.2 MB	280	468	categorical	spectral
Elections for Rector of the University of ...	14.1 KB	26	7	none	education
Abalone	187.5 KB	4177	9	numeric	biology
Adult Census Income	5.3 MB	48842	15	categorical	economy, fairness
Advertising	15.5 KB	200	4	numeric	economy
Ames Iowa Housing	831.2 KB	2930	81	numeric	economy
Bone marrow mononuclear cells with A...	345.0 KB	1000	1004	categorical	aml, expression, sample
Bone marrow mononuclear cells with A...	2.7 MB	8390	1004	categorical	aml, expression
Roman Amphorae	23.7 KB	164	16	categorical	archaeology, image analytics
Attrition - Predict	838 bytes	3	19	none	economy, synthetic, education
Attrition - Train	182.2 KB	1470	19	categorical	economy, synthetic
Auto MPG	17.4 KB	398	9	numeric	
Bank Marketing	466.1 KB	4119	21	categorical	economy
Banking Crises	31.3 KB	211	73	none	time, economy
Dendritic cells in human	21.3 MB	8560	20120	categorical	human expression, disease

Description

Course Grades

A small dataset with grades on seven courses (English, French, History, Algebra, Biology, Physics, Physical) that was handcrafted to introduce hierarchical clustering.

?

16

Info

16 instances (no missing data)
7 features
No target variable.
1 meta attribute

Variables

☒ Show variable labels (if present)
☒ Visualize numeric values
☒ Color by instance classes

Selection

☒ Select full rows

Restore Original Order

☒ Send Automatically

	Student	English	French	History	Algebra	Biology	Physics	Physical
1	Ana	22	30	32	21	37	46	99
2	Bill	91	95	65	89	39	11	29
3	Cynthia	51	89	21	100	70	100	27
4	Demi	9	15	18	61	100	90	8
5	Eve	93	99	39	12	47	17	63
6	Fred	49	17	17	92	70	98	73
7	George	91	99	97	49	96	81	69
8	Henry	12	30	32	34	12	33	96
9	Ian	91	80	20	82	93	87	22
10	Jena	39	18	19	99	97	77	23
11	Katherine	20	50	10	71	99	78	12
12	Lea	90	100	45	45	20	15	100
13	Maya	100	98	97	32	72	22	37
14	Nash	14	4	15	61	42	51	39
15	Olga	9	22	8	100	11	92	29
16	Phill	85	90	100	45	38	92	21

Filter...

Data

File

CSV File Import

Datasets

SQL Table

Data Table

Paint Data

Data Info

Rank

Edit Domain

Color

Feature Statistics

Save Data

Transform

Data Sampler

Select Columns

Select Rows

Transpose

Merge Data

Concatenate

Select by Data Index

Unique

Aggregate Columns

Group by

Pivot Table

Apply Domain

Preprocess

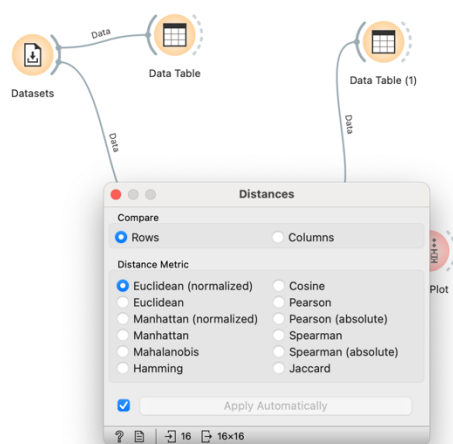
Impute

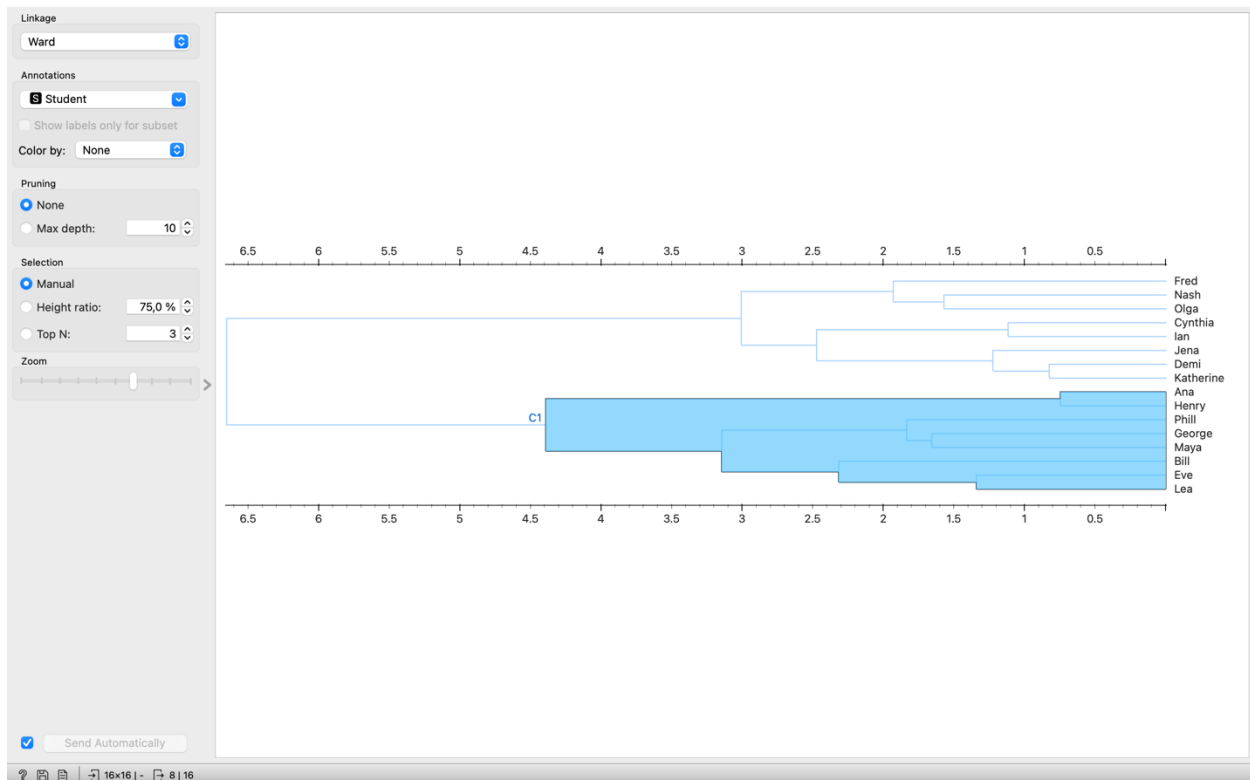
Continuize

Discretize

Distances

Compute a matrix of pairwise distances.
[more...](#)





Info

16 instances (no missing data)

7 features

Target with 2 values

1 meta attribute

Variables

Show variable labels (if present)

Visualize numeric values

Color by instance classes

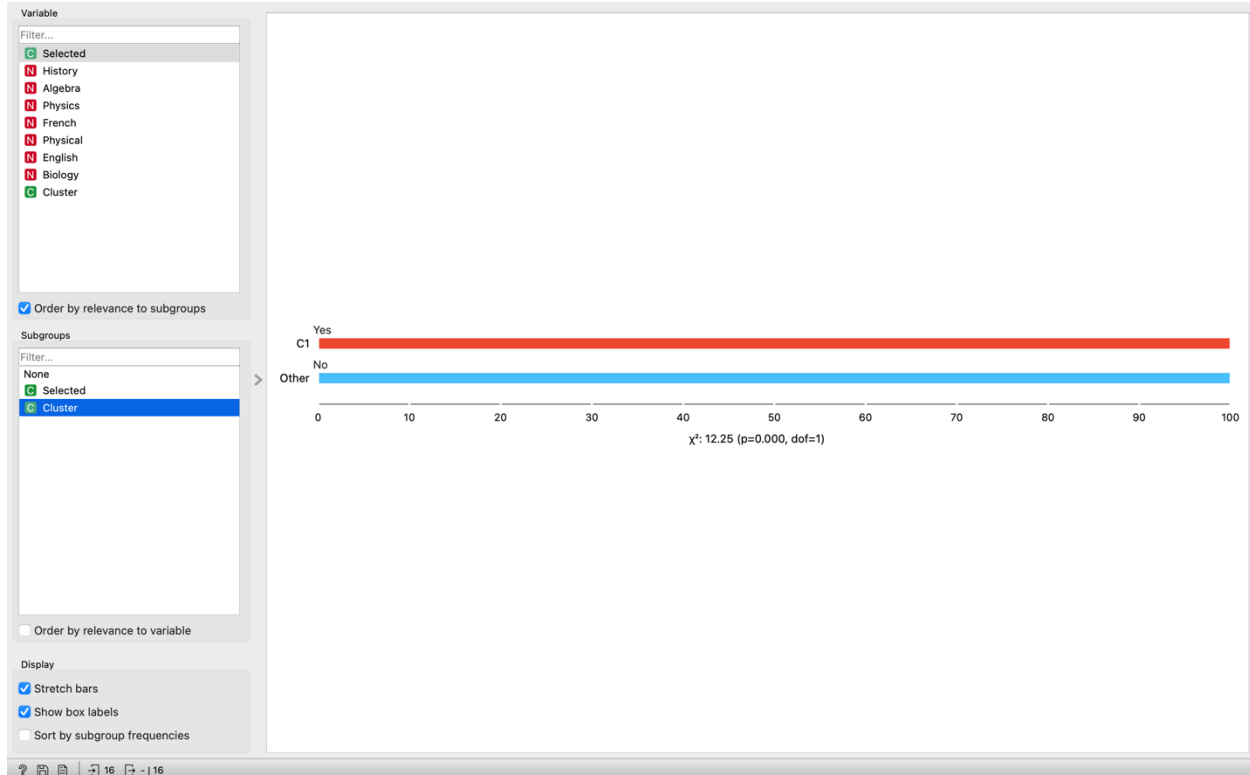
Selection

Select full rows

Restore Original Order

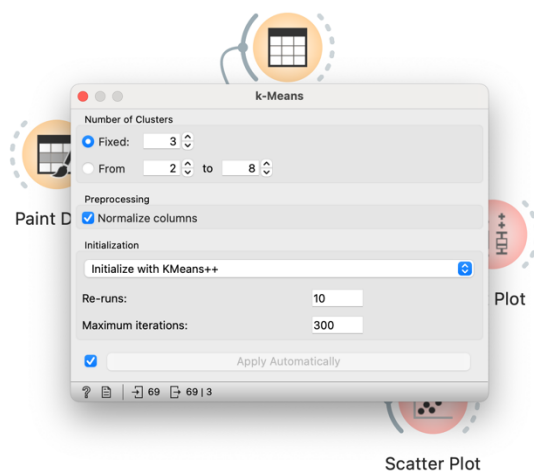
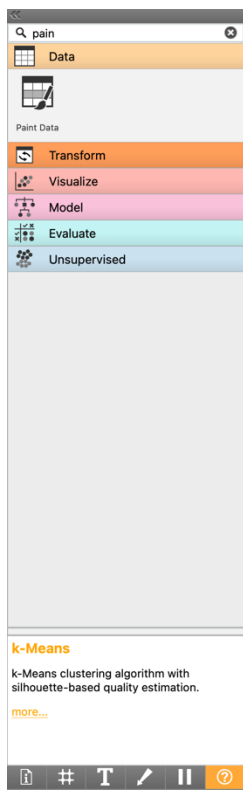
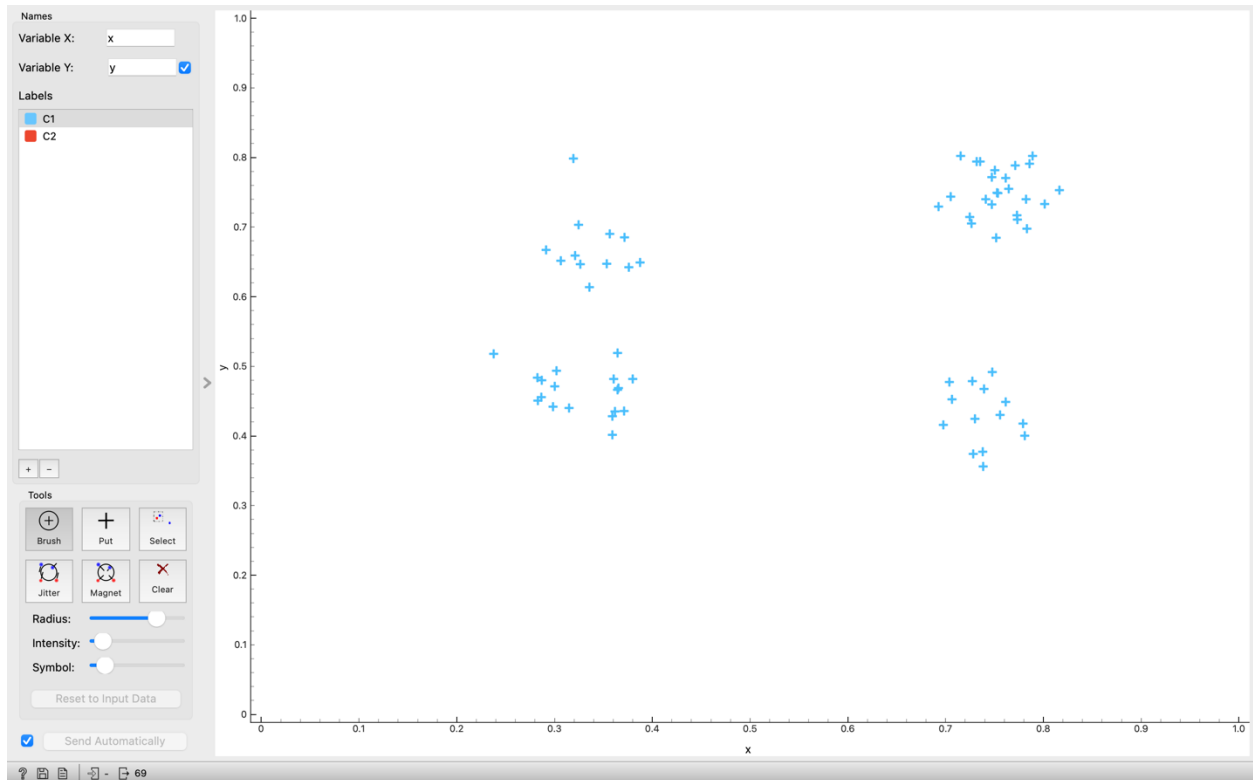
Send Automatically

	Selected	Student	English	French	History	Algebra	Biology	Physics	Physical
1	No	Ana	22	30	32	21	37	46	99
2	No	Bill	91	95	65	89	39	11	29
3	No	Cynthia	51	89	21	100	70	100	27
4	No	Demi	9	15	18	61	100	90	8
5	No	Eve	93	99	39	12	47	17	63
6	No	Fred	49	17	17	92	70	98	73
7	No	George	91	99	97	49	96	81	69
8	No	Henry	12	30	32	34	12	33	96
9	No	Ian	91	80	20	82	93	87	22
10	No	Jena	39	18	19	99	97	77	23
11	No	Katherine	20	50	10	71	99	78	12
12	No	Lea	90	100	45	45	20	15	100
13	No	Maya	100	98	97	32	72	22	37
14	No	Nash	14	4	15	61	42	51	39
15	No	Olga	9	22	8	100	11	92	29
16	No	Phill	85	90	100	45	38	92	21



K-Means Clustering using ODM

K-Means Clustering yang dilakukan menggunakan ODM. Metode ini membagi data ke dalam kkk cluster berdasarkan jarak centroid cluster. Gambar biasanya berupa visualisasi cluster dengan titik-titik data yang dikelompokkan di sekitar centroid.



Info

69 instances (no missing data)
2 features
No target variable.
2 meta attributes

Variables

☒ Show variable labels (if present)
☐ Visualize numeric values
☒ Color by instance classes

Selection

☒ Select full rows

Restore Original Order

☒ Send Automatically

	Cluster	Silhouette	x	y
1	C3	0.666216	0.364605	0.466357
2	C3	0.647629	0.359366	0.428411
3	C3	0.675806	0.282824	0.450486
4	C3	0.666802	0.365526	0.468965
5	C3	0.682697	0.300349	0.471344
6	C3	0.668385	0.314999	0.440188
7	C3	0.685062	0.286761	0.479933
8	C3	0.651157	0.361847	0.435123
9	C3	0.686463	0.301873	0.493519
10	C3	0.671182	0.298494	0.441738
11	C3	0.675415	0.364337	0.519233
12	C3	0.66316	0.380007	0.481683
13	C3	0.6782	0.335684	0.613673
14	C3	0.66077	0.291232	0.667493
15	C3	0.655093	0.376122	0.642196
16	C3	0.6679	0.306235	0.65195
17	C3	0.662641	0.320973	0.65905
18	C3	0.667848	0.326535	0.646819
19	C3	0.547176	0.319153	0.7985
20	C3	0.630258	0.324462	0.7036
21	C3	0.630184	0.35641	0.690119
22	C2	0.726273	0.729812	0.424786
23	C2	0.715397	0.728313	0.37408
24	C2	0.726355	0.755887	0.429924
25	C2	0.721083	0.78072	0.40008
26	C2	0.717576	0.739435	0.467464
27	C2	0.70867	0.738225	0.356159
28	C2	0.72293	0.779085	0.417573
29	C2	0.699993	0.747612	0.491548
30	C1	0.729451	0.747119	0.77146
31	C1	0.718806	0.715373	0.802281
32	C1	0.727235	0.747222	0.732483
33	C1	0.727904	0.750578	0.781873
34	C1	0.698522	0.751704	0.684872
35	C1	0.720799	0.772805	0.71696
36	C1	0.715027	0.692796	0.729474
37	C1	0.71994	0.789027	0.802025
38	C1	0.72973	0.761619	0.770271
39	C1	0.721061	0.801488	0.732867
40	C1	0.720153	0.816343	0.753276
41	C1	0.728544	0.740878	0.740254
42	C1	0.730571	0.753509	0.748916

Variable

Filter...

Cluster
Silhouette
x
y

☐ Order by relevance to subgroups

Subgroups

Filter...

None
Cluster

☐ Order by relevance to variable

Display

☒ Annotate
☐ No comparison
☐ Compare medians
☐ Compare means

0.693488 ± 0.0358

0.6679 0.70867 0.721963

0.55 0.6 0.65 0.7 0.75

