데이터 과학 기초

05

군집 분석

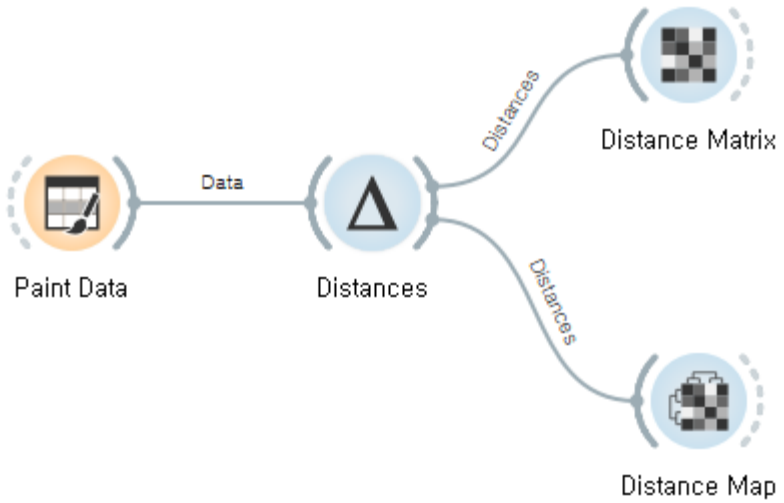
경북대학교 배준현 교수
(joonion@knu.ac.kr)

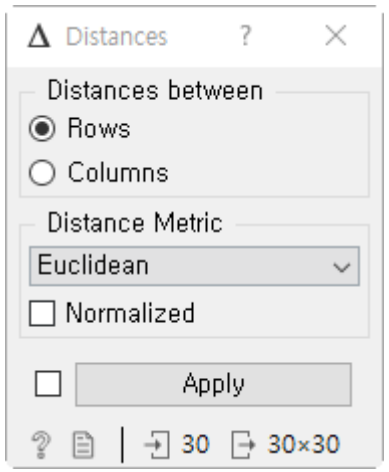
05. 군집 분석

- **군집화: *Clustering***
  - 데이터를 서로 비슷하거나, 서로 가까운 것들끼리 묶어서 나누는 것
  - 비지도 학습: 정답(라벨)이 없음. 유사도와 거리로 판단
    - 유사도(*similarity*): 두 데이터가 얼마나 가까운가를 나타내는 척도
    - 거리(*distance*): 두 데이터 사이의 거리
    - $s = 1 - d, s = similarity, d = distance$

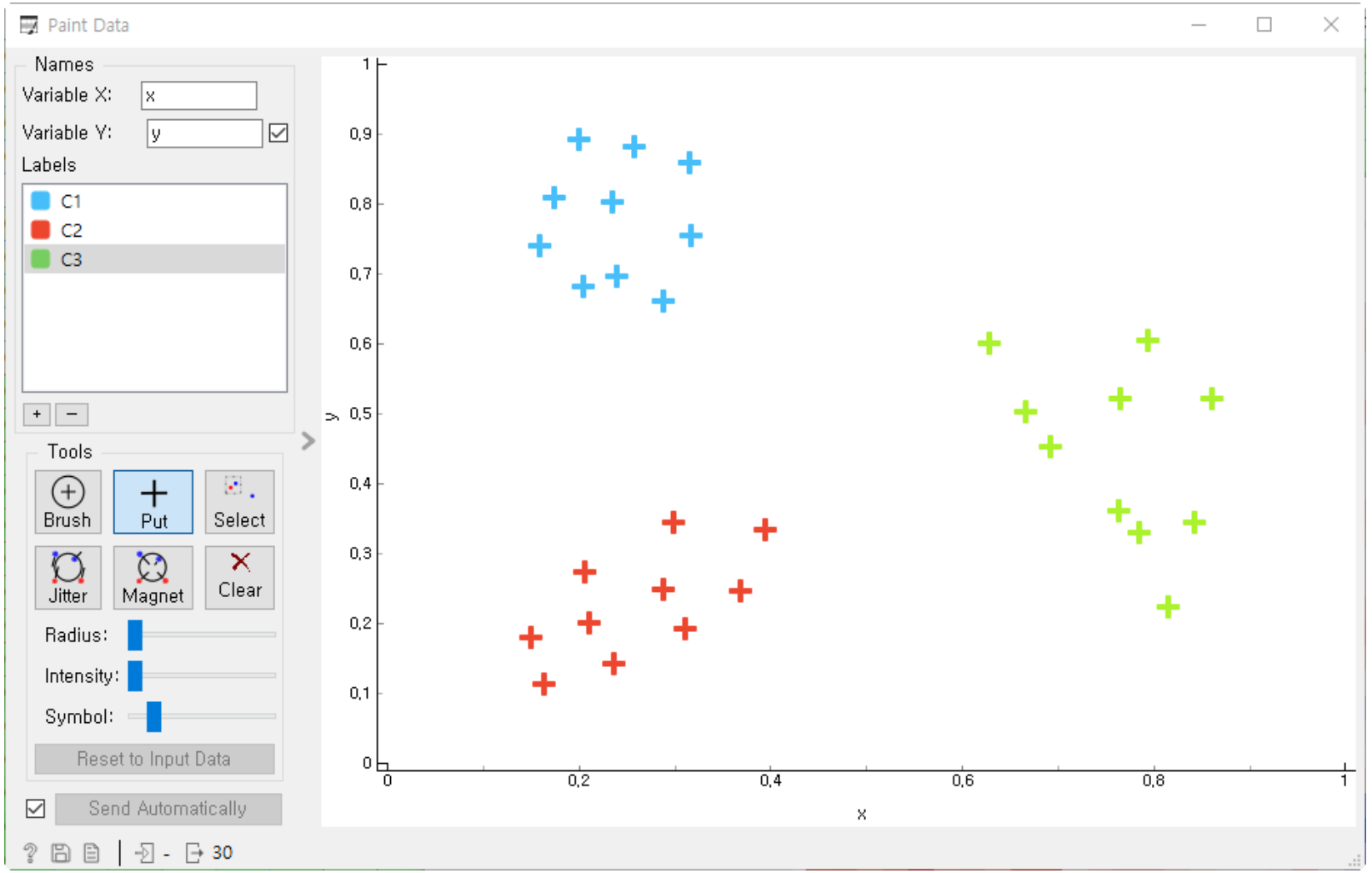
05. 군집 분석

- Orange: Distance





05. 군집 분석


Paint Data
Names
Variable X:
x
Variable Y:
y
Labels
C1
C2
C3
Tools
Brush
Put
Select
Jitter
Magnet
Clear
Radius:
Intensity:
Symbol:
Reset to Input Data
Send Automatically
30
y
x
0
0,1
0,2
0,3
0,4
0,5
0,6
0,7
0,8
0,9
1
0,2
0,4
0,6
0,8

05. 군집 분석

Distance Matrix

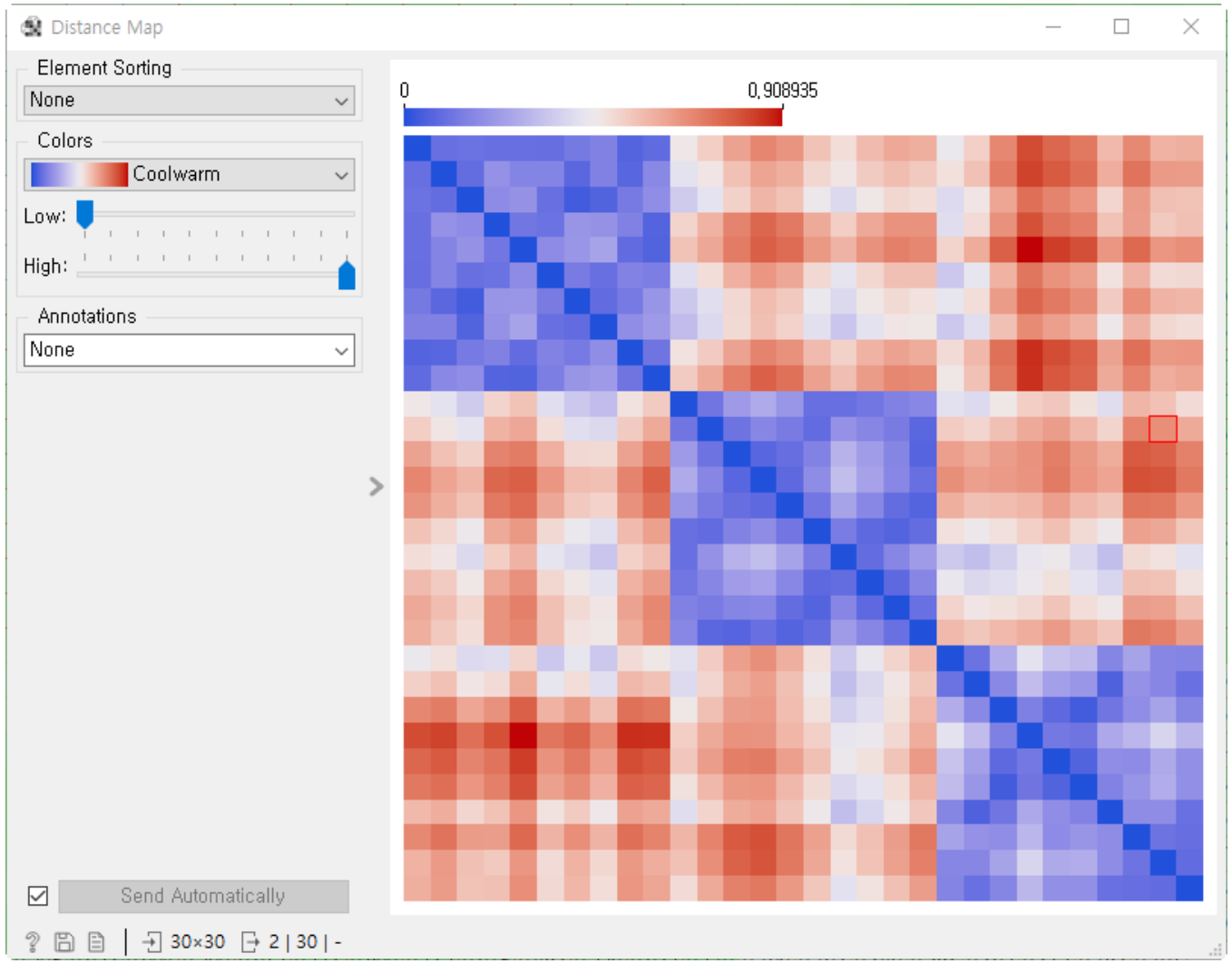
	C1	C1	C1	C1	C1	C1	C1	C1	C1	C1	C2	C2	C2	C2	C2	C2	C2	C2	C2	C2	C3	C3	C3	C3	C3	C3	C3	C3	C3	C3
C1		0.098	0.106	0.098	0.096	0.095	0.125	0.151	0.061	0.082	0.463	0.530	0.629	0.694	0.661	0.557	0.495	0.572	0.615	0.603	0.442	0.526	0.689	0.820	0.761	0.725	0.576	0.686	0.593	0.600
C1	0.098		0.092	0.196	0.158	0.159	0.074	0.151	0.070	0.173	0.420	0.469	0.561	0.627	0.603	0.509	0.470	0.537	0.569	0.542	0.490	0.560	0.714	0.835	0.790	0.749	0.606	0.735	0.649	0.644
C1	0.106	0.092		0.179	0.200	0.097	0.038	0.060	0.130	0.186	0.357	0.424	0.525	0.588	0.554	0.451	0.394	0.468	0.509	0.497	0.401	0.469	0.622	0.745	0.698	0.657	0.514	0.646	0.562	0.554
C1	0.098	0.196	0.179		0.120	0.104	0.209	0.200	0.150	0.062	0.515	0.596	0.699	0.761	0.721	0.611	0.531	0.615	0.667	0.667	0.406	0.500	0.670	0.809	0.737	0.708	0.554	0.641	0.542	0.562
C1	0.096	0.158	0.200	0.120		0.181	0.211	0.248	0.087	0.059	0.557	0.619	0.715	0.780	0.751	0.650	0.591	0.668	0.709	0.692	0.518	0.608	0.775	0.909	0.845	0.812	0.660	0.758	0.660	0.676
C1	0.095	0.159	0.097	0.104	0.181		0.134	0.098	0.153	0.140	0.411	0.494	0.599	0.660	0.618	0.507	0.428	0.511	0.563	0.565	0.348	0.431	0.596	0.729	0.667	0.632	0.482	0.592	0.500	0.505
C1	0.125	0.074	0.038	0.209	0.211	0.134		0.086	0.131	0.207	0.351	0.409	0.505	0.570	0.541	0.441	0.396	0.465	0.501	0.481	0.432	0.495	0.645	0.764	0.722	0.679	0.539	0.676	0.594	0.583
C1	0.151	0.151	0.060	0.200	0.248	0.098	0.086		0.187	0.223	0.317	0.396	0.501	0.562	0.522	0.413	0.344	0.422	0.469	0.467	0.346	0.410	0.562	0.685	0.639	0.597	0.455	0.590	0.509	0.497
C1	0.061	0.070	0.130	0.150	0.087	0.153	0.131	0.187		0.111	0.481	0.537	0.630	0.696	0.670	0.572	0.524	0.595	0.632	0.610	0.500	0.580	0.740	0.868	0.814	0.776	0.629	0.744	0.653	0.657
C1	0.082	0.173	0.186	0.062	0.059	0.140	0.207	0.223	0.111		0.539	0.611	0.711	0.775	0.740	0.634	0.565	0.645	0.692	0.683	0.465	0.558	0.726	0.863	0.795	0.764	0.611	0.703	0.604	0.622
C2	0.463	0.420	0.357	0.515	0.557	0.411	0.351	0.317	0.481	0.539		0.117	0.222	0.268	0.212	0.096	0.096	0.120	0.153	0.169	0.418	0.400	0.465	0.530	0.544	0.486	0.408	0.589	0.560	0.499
C2	0.530	0.469	0.424	0.596	0.619	0.494	0.409	0.396	0.537	0.611	0.117		0.109	0.166	0.135	0.086	0.198	0.165	0.133	0.073	0.534	0.514	0.564	0.611	0.640	0.582	0.518	0.700	0.675	0.612
C2	0.629	0.561	0.525	0.699	0.715	0.599	0.505	0.501	0.630	0.711	0.222	0.109		0.068	0.094	0.154	0.289	0.229	0.162	0.064	0.637	0.609	0.640	0.667	0.712	0.653	0.607	0.789	0.772	0.704
C2	0.694	0.627	0.588	0.761	0.780	0.660	0.570	0.562	0.696	0.775	0.268	0.166	0.068		0.079	0.184	0.320	0.245	0.167	0.099	0.674	0.636	0.649	0.661	0.717	0.658	0.628	0.808	0.800	0.727
C2	0.661	0.603	0.554	0.721	0.751	0.618	0.541	0.522	0.670	0.740	0.212	0.135	0.094	0.079		0.118	0.248	0.168	0.090	0.064	0.603	0.561	0.571	0.584	0.639	0.580	0.552	0.731	0.725	0.651
C2	0.557	0.509	0.451	0.611	0.650	0.507	0.441	0.413	0.572	0.634	0.096	0.086	0.154	0.184	0.118		0.136	0.081	0.061	0.091	0.490	0.456	0.489	0.528	0.563	0.503	0.453	0.634	0.619	0.550
C2	0.495	0.470	0.394	0.531	0.591	0.428	0.396	0.344	0.524	0.565	0.096	0.198	0.289	0.320	0.248	0.136		0.091	0.165	0.227	0.355	0.320	0.370	0.435	0.448	0.390	0.321	0.503	0.483	0.415
C2	0.572	0.537	0.468	0.615	0.668	0.511	0.465	0.422	0.595	0.645	0.120	0.165	0.229	0.245	0.168	0.081	0.091		0.079	0.165	0.439	0.393	0.411	0.447	0.484	0.425	0.384	0.564	0.556	0.483
C2	0.615	0.569	0.509	0.667	0.709	0.563	0.501	0.469	0.632	0.692	0.153	0.133	0.162	0.167	0.090	0.061	0.165	0.079		0.101	0.517	0.472	0.483	0.505	0.553	0.494	0.462	0.641	0.635	0.561
C2	0.603	0.542	0.497	0.667	0.692	0.565	0.481	0.467	0.610	0.683	0.169	0.073	0.064	0.099	0.064	0.091	0.227	0.165	0.101		0.579	0.547	0.576	0.605	0.648	0.589	0.544	0.725	0.710	0.641
C3	0.442	0.490	0.401	0.406	0.518	0.348	0.432	0.346	0.500	0.465	0.418	0.534	0.637	0.674	0.603	0.490	0.355	0.439	0.517	0.579		0.105	0.275	0.421	0.334	0.313	0.161	0.246	0.166	0.158
C3	0.526	0.560	0.469	0.500	0.608	0.431	0.495	0.410	0.580	0.558	0.400	0.514	0.609	0.636	0.561	0.456	0.320	0.393	0.472	0.547	0.105		0.172	0.316	0.237	0.210	0.056	0.195	0.163	0.101
C3	0.689	0.714	0.622	0.670	0.775	0.596	0.645	0.562	0.740	0.726	0.465	0.564	0.640	0.649	0.571	0.489	0.370	0.411	0.483	0.576	0.275	0.172		0.147	0.081	0.038	0.116	0.188	0.246	0.160
C3	0.820	0.835	0.745	0.809	0.909	0.729	0.764	0.685	0.868	0.863	0.530	0.611	0.667	0.661	0.584	0.528	0.435	0.447	0.505	0.605	0.421	0.316	0.147		0.124	0.111	0.260	0.302	0.382	0.302
C3	0.761	0.790	0.698	0.737	0.845	0.667	0.722	0.639	0.814	0.795	0.544	0.640	0.712	0.717	0.639	0.563	0.448	0.484	0.553	0.648	0.334	0.237	0.081	0.124		0.060	0.185	0.178	0.265	0.193
C3	0.725	0.749	0.657	0.708	0.812	0.632	0.679	0.597	0.776	0.764	0.486	0.582	0.653	0.658	0.580	0.503	0.390	0.425	0.494	0.589	0.313	0.210	0.038	0.111	0.060		0.154	0.206	0.275	0.193
C3	0.576	0.606	0.514	0.554	0.660	0.482	0.539	0.455	0.629	0.611	0.408	0.518	0.607	0.628	0.552	0.453	0.321	0.384	0.462	0.544	0.161	0.056	0.116	0.260	0.185	0.154		0.182	0.183	0.100
C3	0.686	0.735	0.646	0.641	0.758	0.592	0.676	0.590	0.744	0.703	0.589	0.700	0.789	0.808	0.731	0.634	0.503	0.564	0.641	0.725	0.246	0.195	0.188	0.302	0.178	0.206	0.182		0.107	0.096
C3	0.593	0.649	0.562	0.542	0.660	0.500	0.594	0.509	0.653	0.604	0.560	0.675	0.772	0.800	0.725	0.619	0.483	0.556	0.635	0.710	0.166	0.163	0.246	0.382	0.265	0.275	0.183	0.107		0.088
C3	0.600	0.644	0.554	0.562	0.676	0.505	0.583	0.497	0.657	0.622	0.499	0.612	0.704	0.727	0.651	0.550	0.415	0.483	0.561	0.641	0.158	0.101	0.160	0.302	0.193	0.193	0.100	0.096	0.088	

Labels: Class

Send Automatically

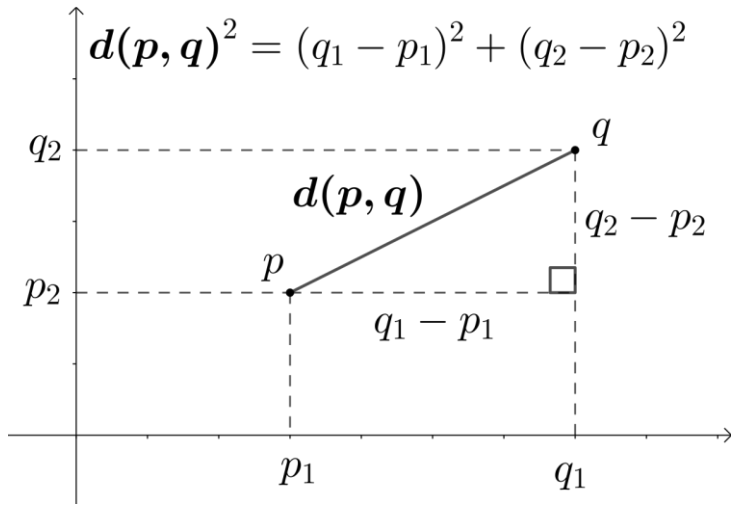
30×30

05. 군집 분석


Distance Map
Element Sorting
None
Colors
Coolwarm
Low:
High:
Annotations
None
0
0,908935
Send Automatically
30×30
2 | 30 | -

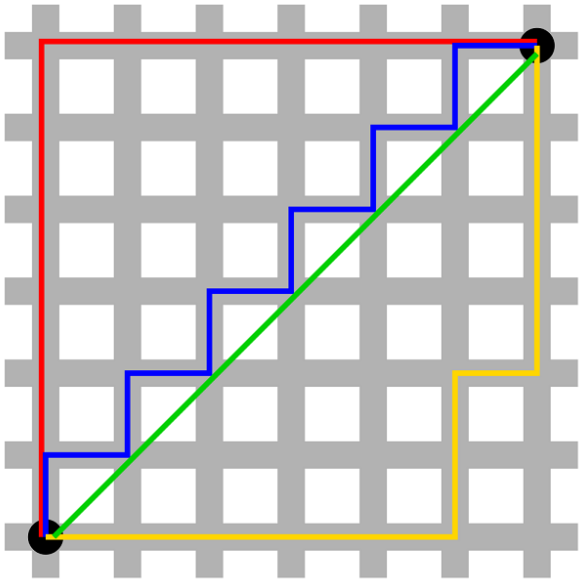
05. 군집 분석

- **유클리드 거리**: *Euclidean* Distance
  - 두 점 사이의 거리를 계산하는 가장 일반적인 방법
  - 유클리드 공간에서 두 점 사이의 거리
    - 2차원: $d = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2}$
    - 3차원: $d = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2}$
    - n차원: $P = (p_1, p_2, \cdots, p_n), Q = (q_1, q_2, \cdots, q_n)$
      - $d = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$



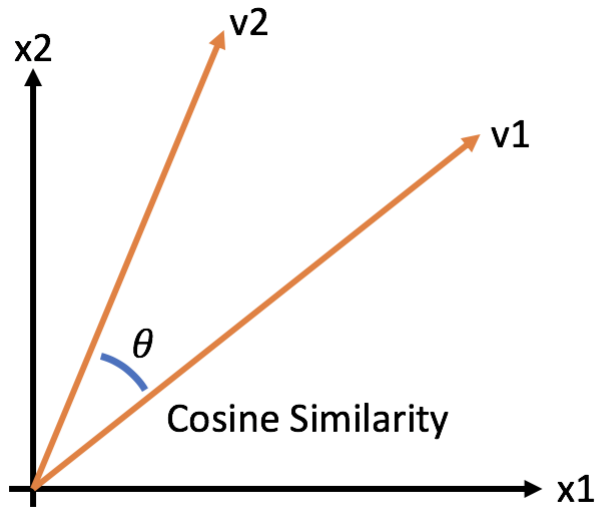
05. 군집 분석

- **맨하탄 거리:** *Manhattan* Distance
  - 택시 거리라고도 함: 맨하탄에서 택시타고 가는 거리
  - 두 점 사이의 데카르트(Cartesian) 좌표계에서의 거리차의 절대값의 총합
    - $d = \sqrt{\sum_{i=1}^{n} |p_i - q_i|}$



05. 군집 분석

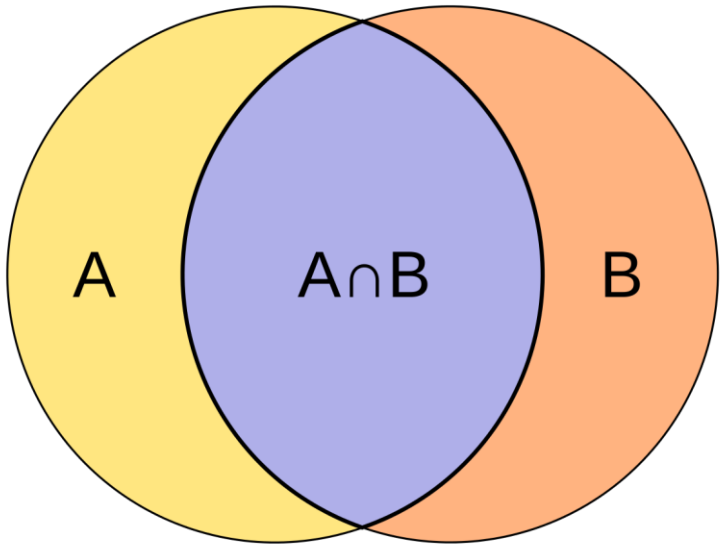
- **코사인 유사도**: *Cosine* Similarity
  - **내적공간에서 두 벡터의 방향이 이루는 각의 코사인 값**: [0, 1]
  - $s = \frac{P \cdot Q}{\|P\| \times \|Q\|} = \frac{\sum p_i \times q_i}{\sqrt{\sum {p_i}^2} \times \sqrt{\sum {q_i}^2}}$
  - **코사인 거리**: Cosine Distance
    - $d = 1 - s$



05. 군집 분석

자카드 유사도: *Jaccard* Similarity

- 두 집합 사이의 유사도를 측정하는 대표적인 방법
  - 페이스북에서 친구 추천할 때, 넷플릭스에서 영화 추천할 때.
- 전체 집합의 크기와 교집합의 크기로 유사도 측정
  - $s(A,B) = \frac{|A \cap B|}{|A \cup B|}$
- 자카드 거리: Jaccard Distance
  - $d = 1 - s$

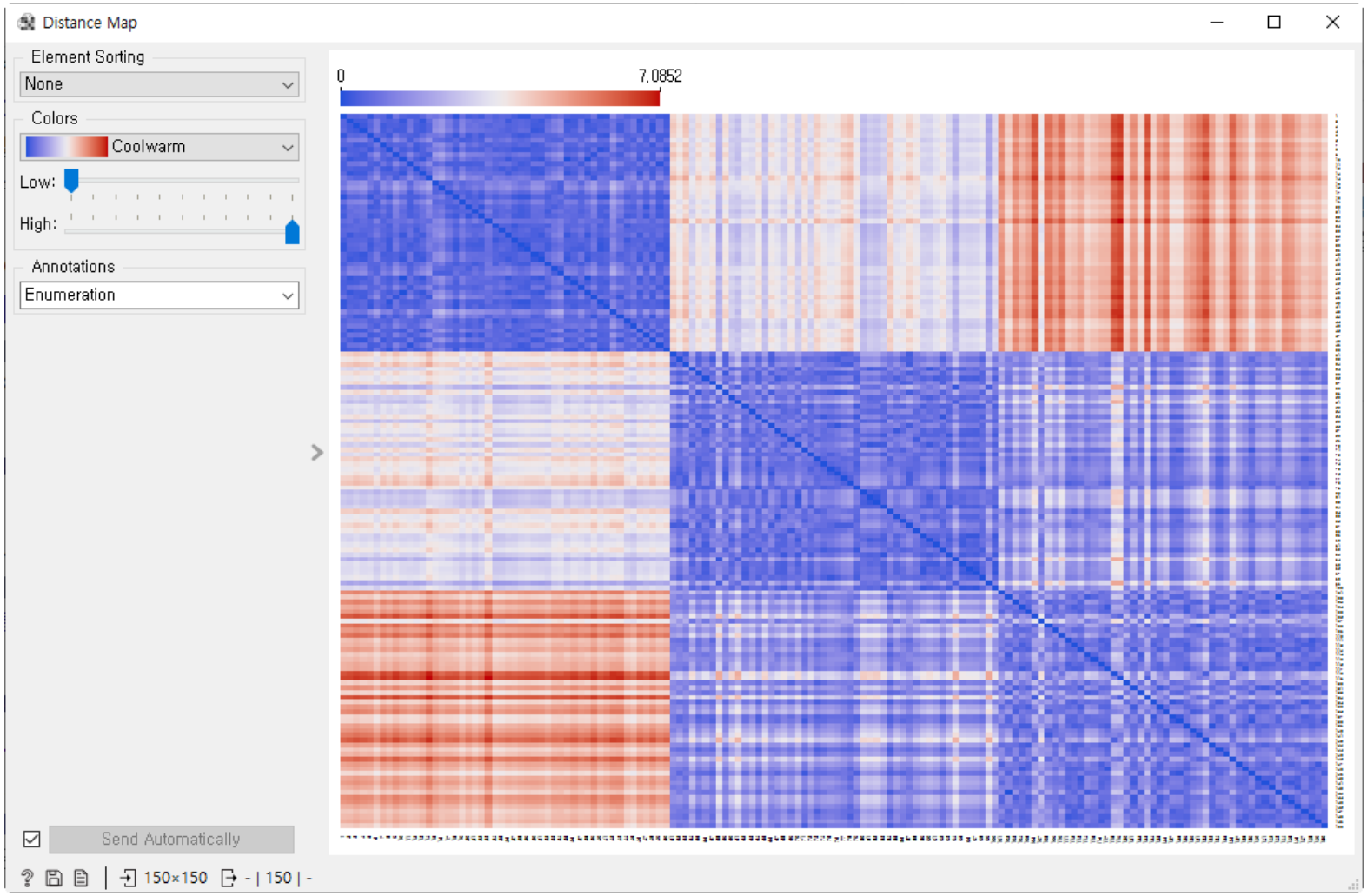


05. 군집 분석

- **기타 유사도/거리 측정을 위한 지표**: Distance Metrics
  - **피어슨 거리**: Pearson Distance
  - **스피어먼 거리**: Spearman Distance
  - **해밍 거리**: Hamming Distance
  - **마할라노비스 거리**: Mahalanobis Distance
  - **바타차야 거리**: Bhattacharyya Distance
  - **체비셰프 거리**: Chebyshev Distance

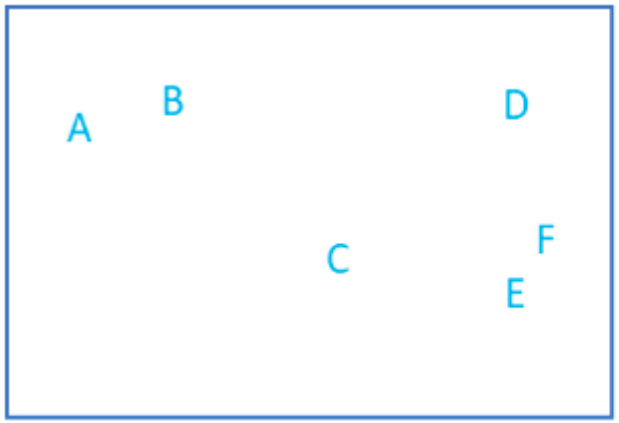
05. 군집 분석

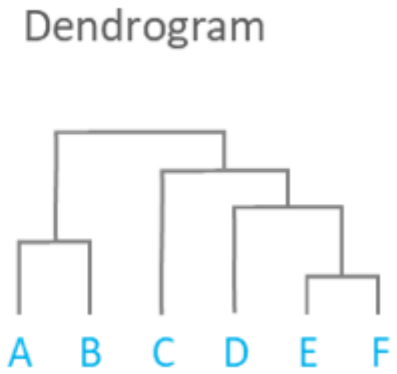
- IRIS 데이터셋에서의 유사도와 거리:



05. 군집 분석

- **계층적 군집화**: *Hierarchical* Clustering
  - 군집간의 거리를 이용하여 계층적으로 군집을 분석하는 방법
    - 병합적 방법: *agglomerative*, *bottom-up* approach
    - 분할적 방법: *divisive*(partitioning), *top-down* approach
  - 덴드로그램: *Dendrogram*
    - 군집의 계층적 구조를 그림으로 보여주는 방법



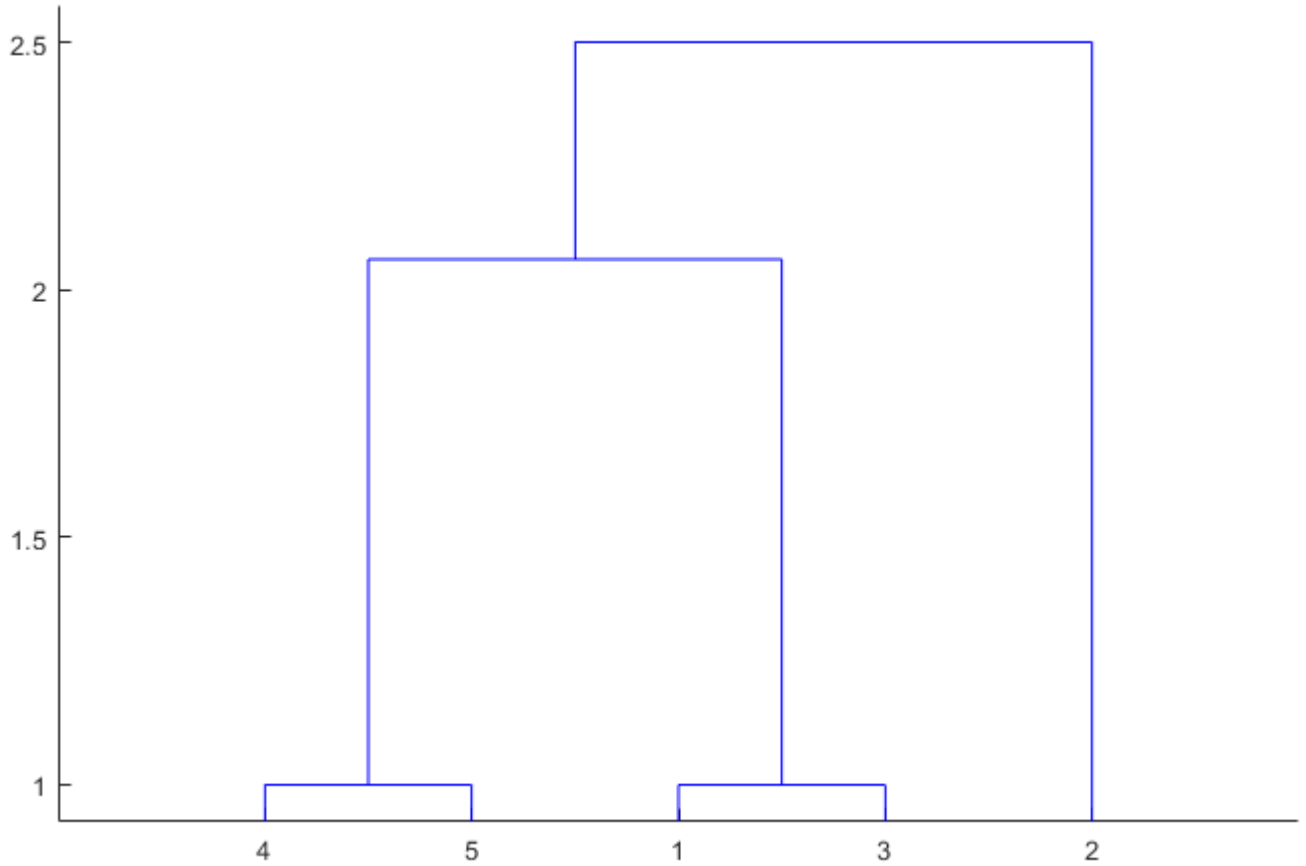


05. 군집 분석

- **군집 간의 거리를 측정하는 방법: *Linkage* Method**
  - **최단 연결법: *Single* Linkage**
    - **두 군집에 속하는 데이터 중에서 가장 가까운 데이터 간의 거리로 연결**
  - **최장 연결법: *Complete* Linkage**
    - **두 군집에 속하는 데이터 중에서 가장 먼 데이터 간의 거리로 연결**
  - **평균 연결법: *Average* Linkage**
    - **두 군집에 속하는 모든 데이터 간의 거리의 평균 거리로 연결**
  - **중심 연결법: *Centroid* Linkage**
    - **두 군집에서의 중심점(centroid)을 찾아서 두 중심점의 거리로 연결**
  - **Ward의 연결법: *Ward* Linkage**
    - **두 군집을 합쳤을 때의 분산이 최소화 되는 군집을 합치는 방법**

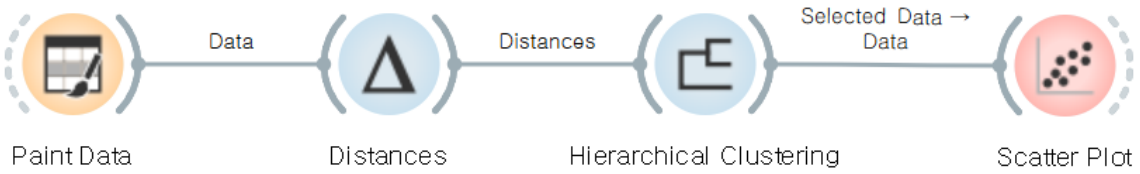
05. 군집 분석

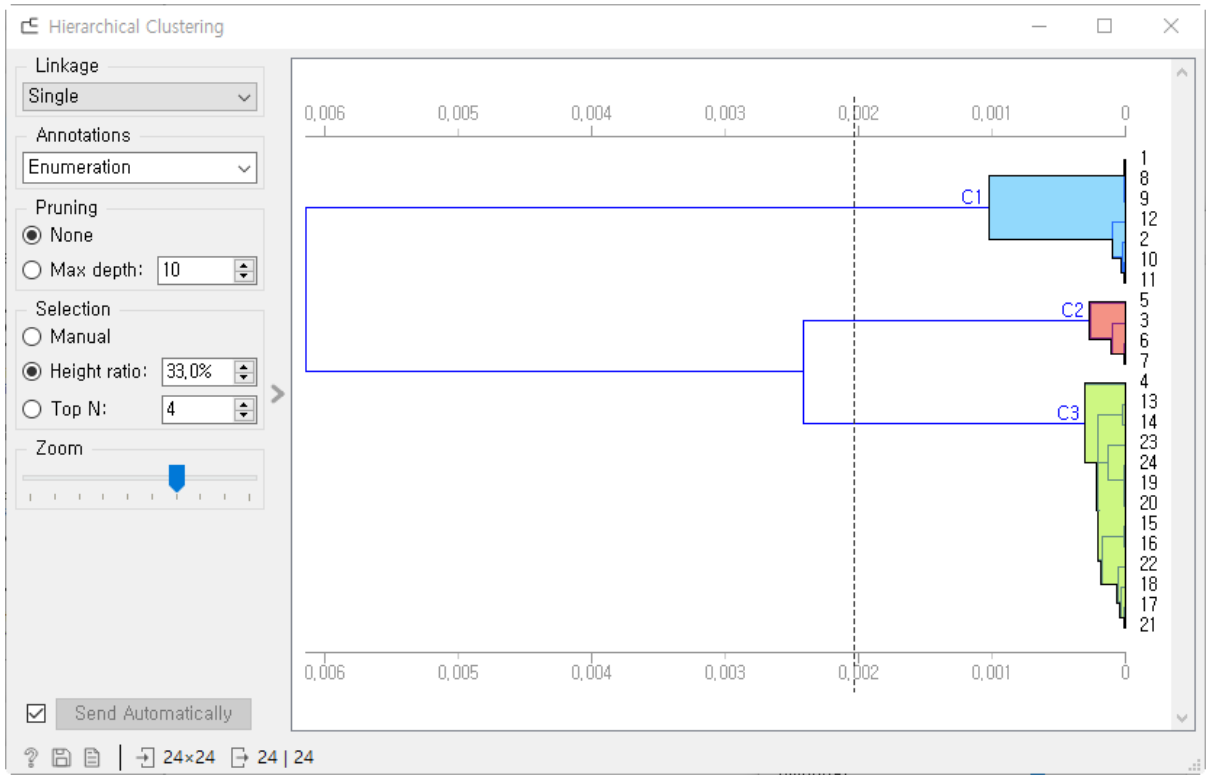
- 덴드로그램으로 군집을 나누는 방법:



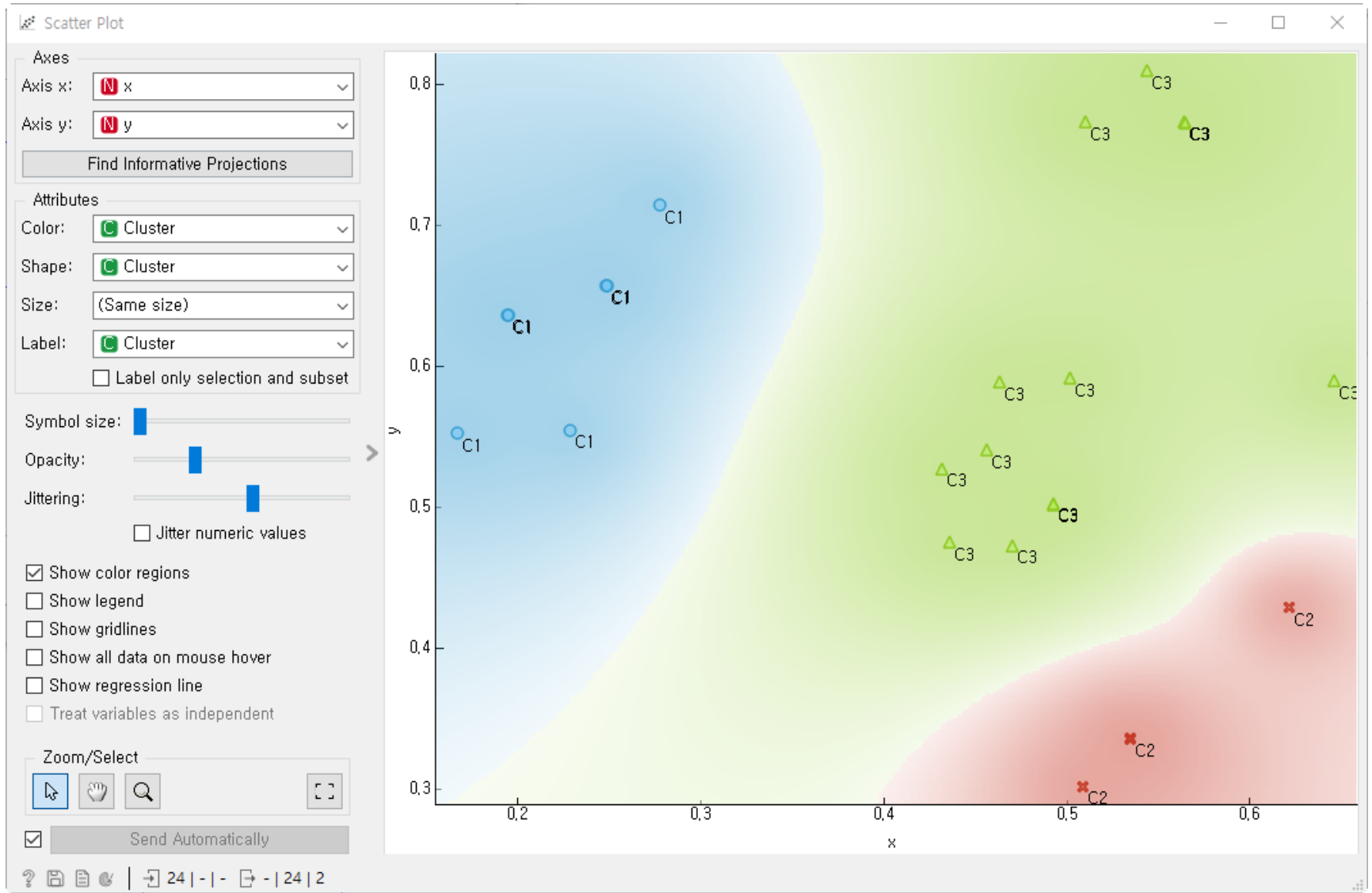
05. 군집 분석

- Orange: Hierarchical Clustering



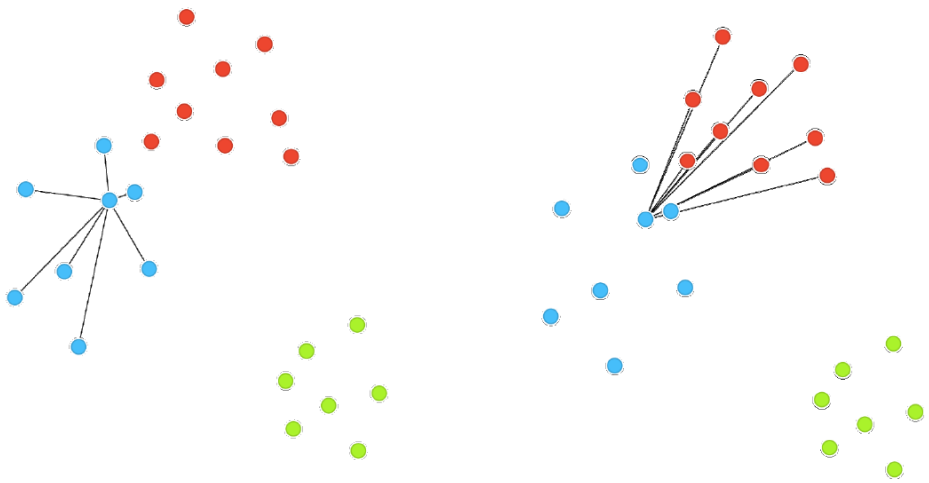


05. 군집 분석

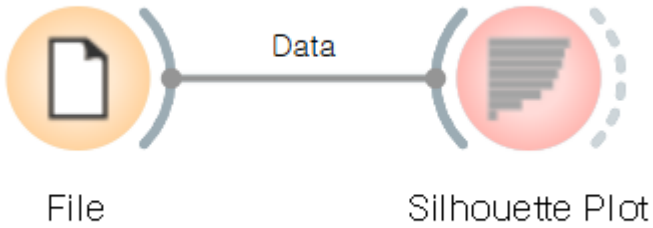

Scatter Plot
Axes
Axis x: x
Axis y: y
Find Informative Projections
Attributes
Color: Cluster
Shape: Cluster
Size: (Same size)
Label: Cluster
Label only selection and subset
Symbol size:
Opacity:
Jittering:
Jitter numeric values
Show color regions
Show legend
Show gridlines
Show all data on mouse hover
Show regression line
Treat variables as independent
Zoom/Select
Send Automatically
24 | - | -
- | 24 | 2
0,8
0,7
0,6
0,5
0,4
0,3
y
0,2
0,3
0,4
0,5
0,6
x
C1
C2
C3

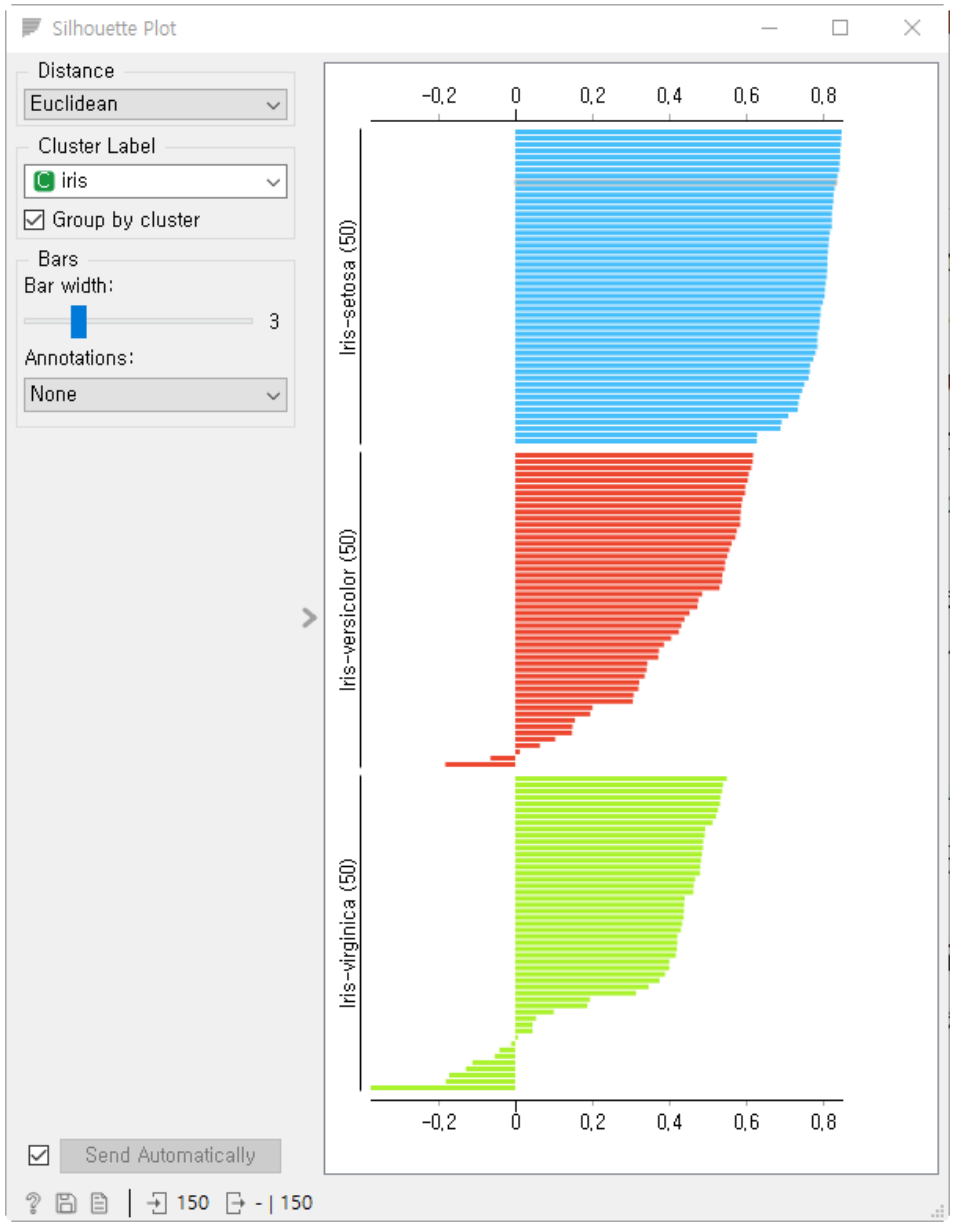
05. 군집 분석

- **군집 모델의 평가**: Cluster Evaluation
  - 비지도 학습: 군집이 얼마나 서로 잘 구분이 되었는 지로 평가
  - 실루엣 점수: *Silhouette Score*
    - 같은 군집의 데이터와는 가깝고, 다른 군집 데이터와의 거리는 멀수록 좋다.
    - $a(i)$: 같은 군집의 데이터와의 거리의 평균값
    - $b(i)$: 가장 가까운 다른 군집의 데이터와의 거리의 평균값
    - $s(i) = \frac{b(i)-a(i)}{\max\{a(i), b(i)\}}$



- Orange: Silhouette Plot



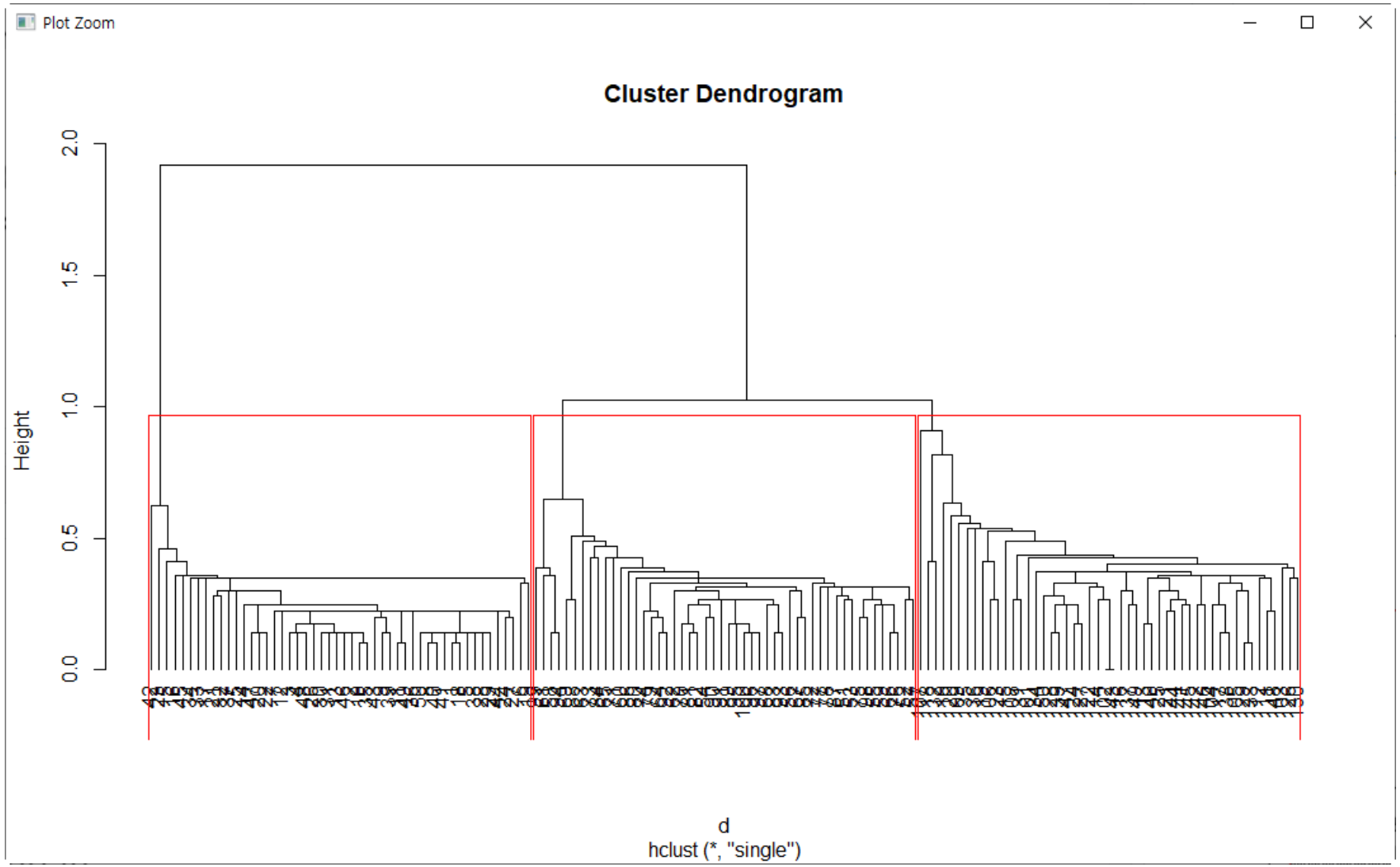


05. 군집 분석

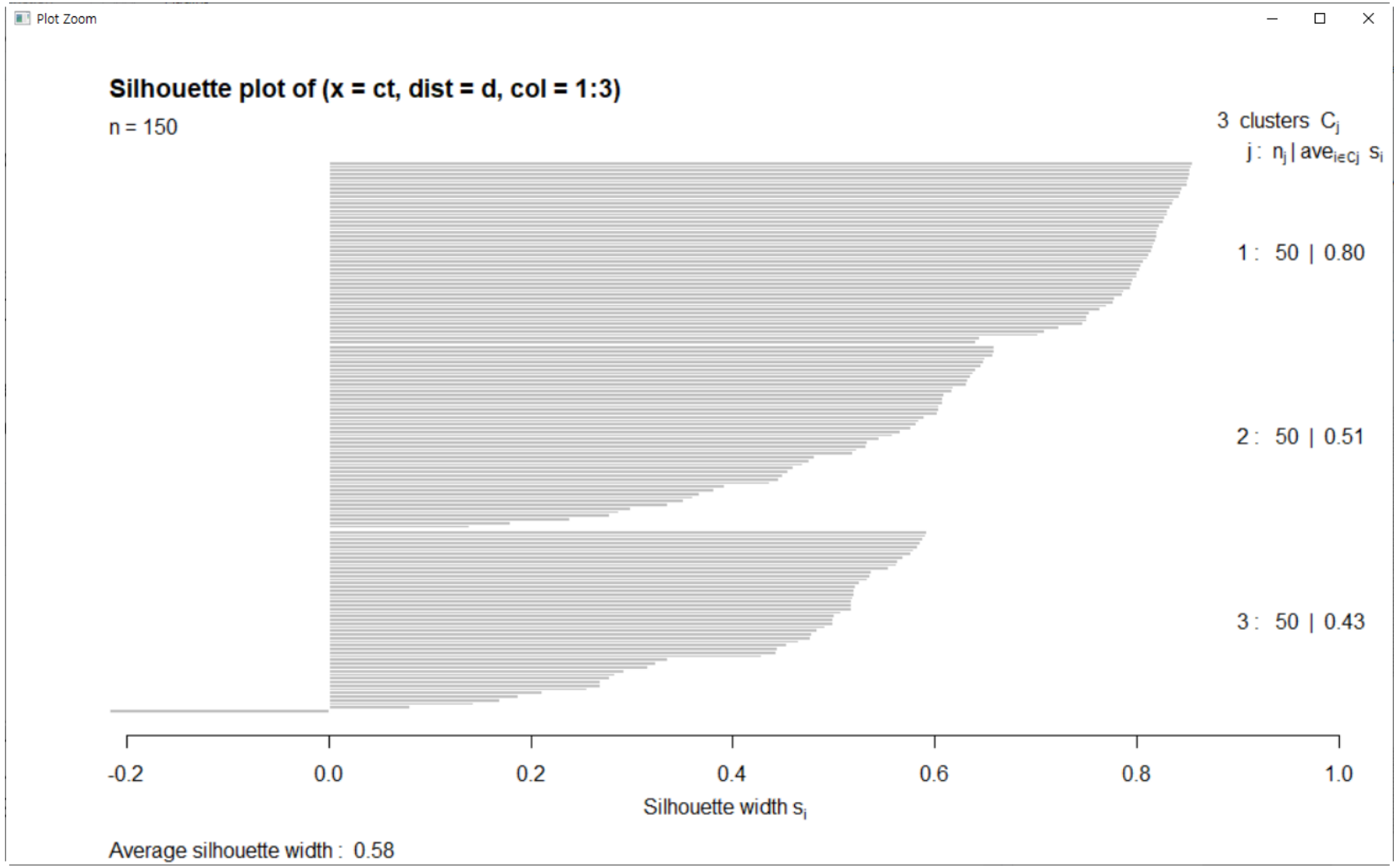
- R: hclust()

```
#install.packages("cluster")
library(cluster)

df <- iris
df$Species <- as.numeric(df$Species)
head(df)
d <- dist(df, method="euclidean")
model <- hclust(d, method="single")
model

plot(model, hang=-1)
rect.hclust(model, k=3)

ct <- cutree(model, k=3)
plot(silhouette(ct, dist=d, col=1:3))
```

05. 군집 분석

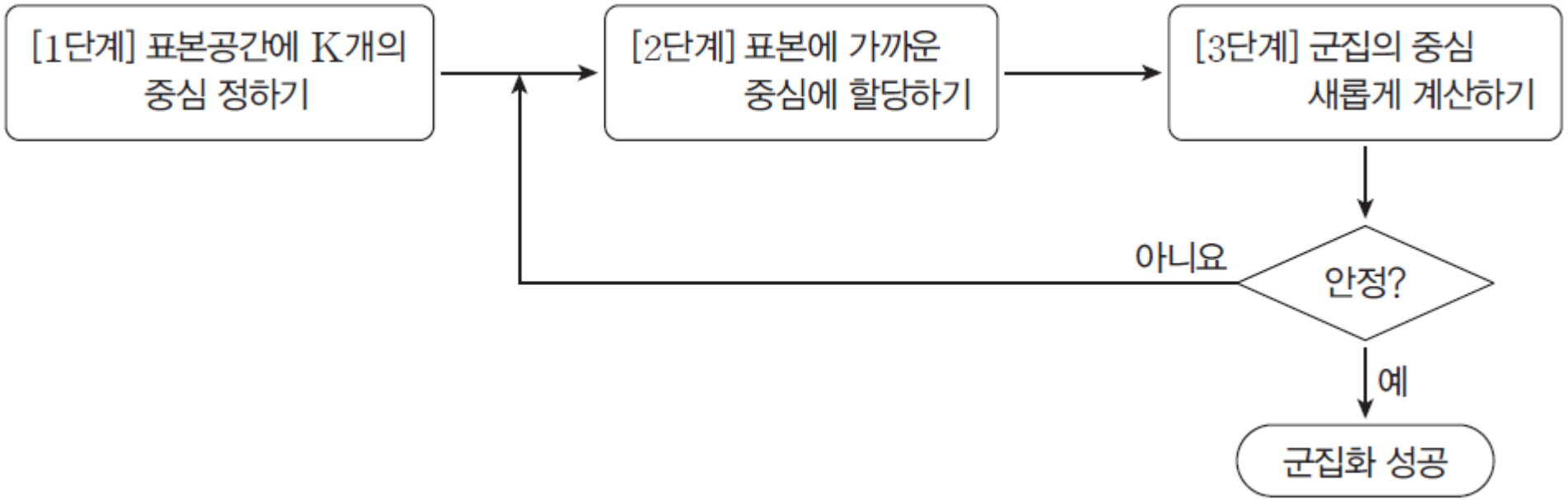

Plot Zoom
Cluster Dendrogram
Height
2.0
1.5
1.0
0.5
0.0
d
hclust (*, "single")

05. 군집 분석


Plot Zoom
Silhouette plot of (x = ct, dist = d, col = 1:3)
n = 150
3 clusters C_j
j : n_j | $ave_{i \in Cj}$ s_i
1 : 50 | 0.80
2 : 50 | 0.51
3 : 50 | 0.43
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Silhouette width s_i
Average silhouette width : 0.58

k-평균 군집화: *k-means* clustering

- 가장 일반적으로 널리 사용되는 군집화 알고리즘

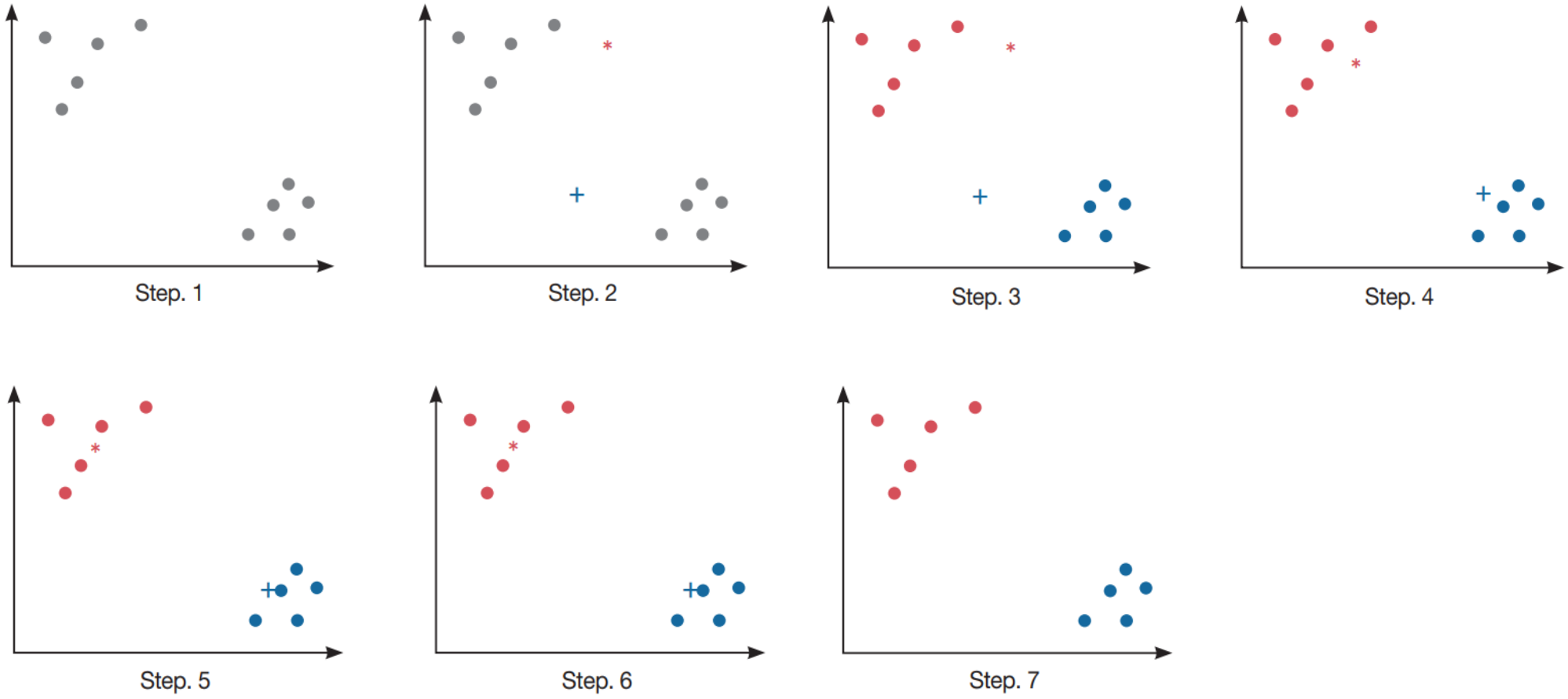


05. 군집 분석

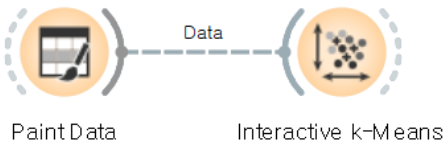
- k-평균 군집화: k-means clustering
  - 먼저 찾고자 하는 클러스터의 개수 k를 지정한다.
  - k개의 중심점(*centroid*) 위치를 임의로 지정한다.
  - 각 데이터들을 가장 가까운 중심점에 속하는 군집으로 결정한다.
  - 군집이 결정된 데이터들로 다시 중심점을 찾는다.
  - 다시, 각 데이터들을 가장 가까운 중심점에 속하는 군집으로 재편한다.
  - 중심점 이동이 없을 때까지 (혹은, 임계값 이하가 될 때까지) 반복한다.
  - 알고리즘의 동작이 멈추었을 때의 군집을 각 데이터의 군집으로 분류한다

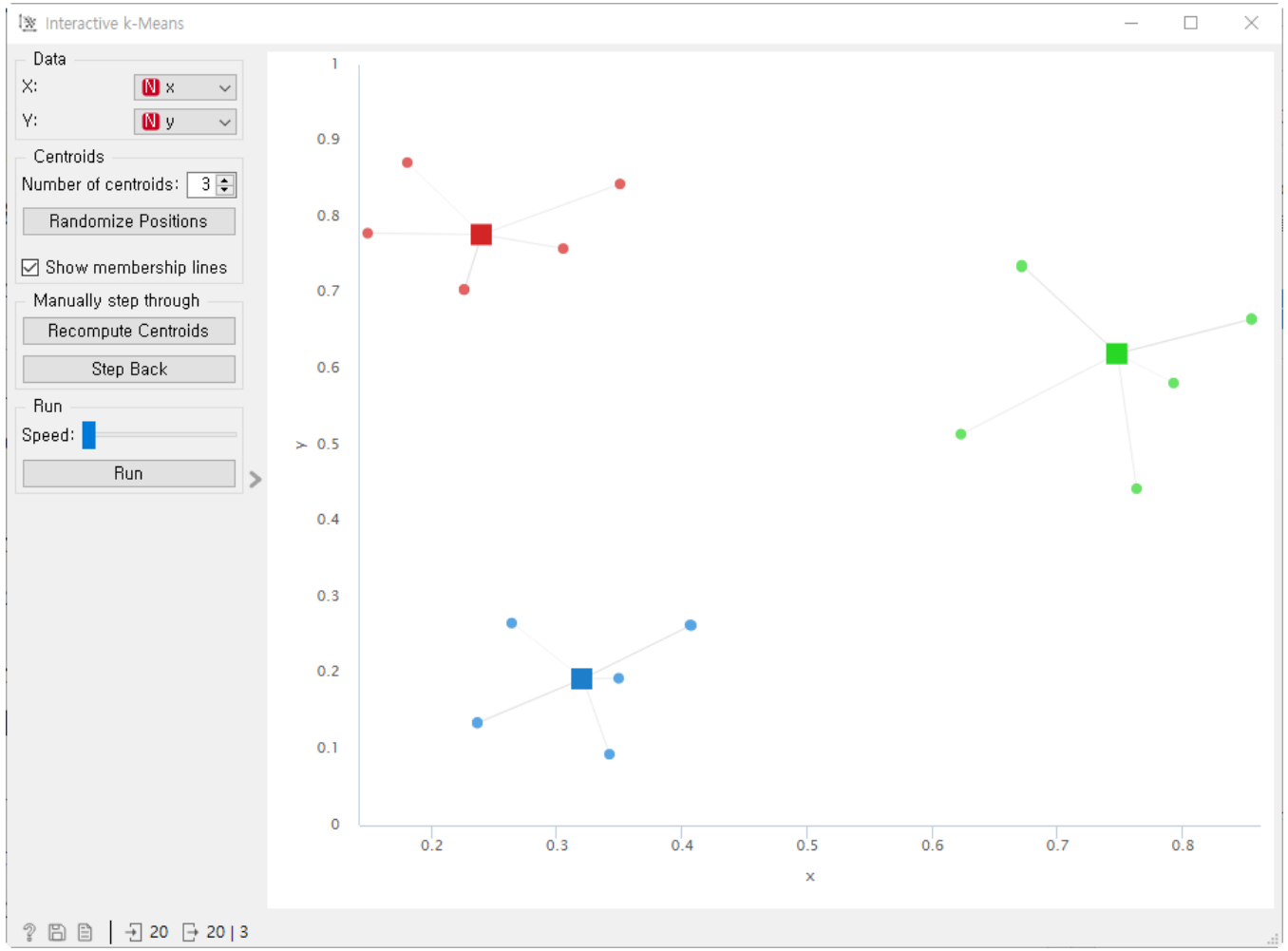
05. 군집 분석

- k=2 (군집이 두 개)인 경우:



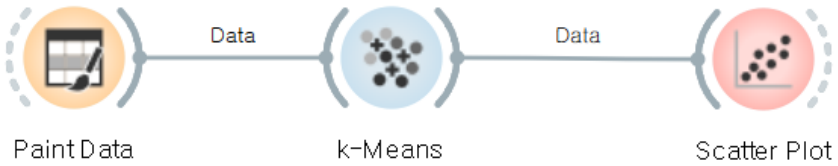
- Orange: Interactive k-Means

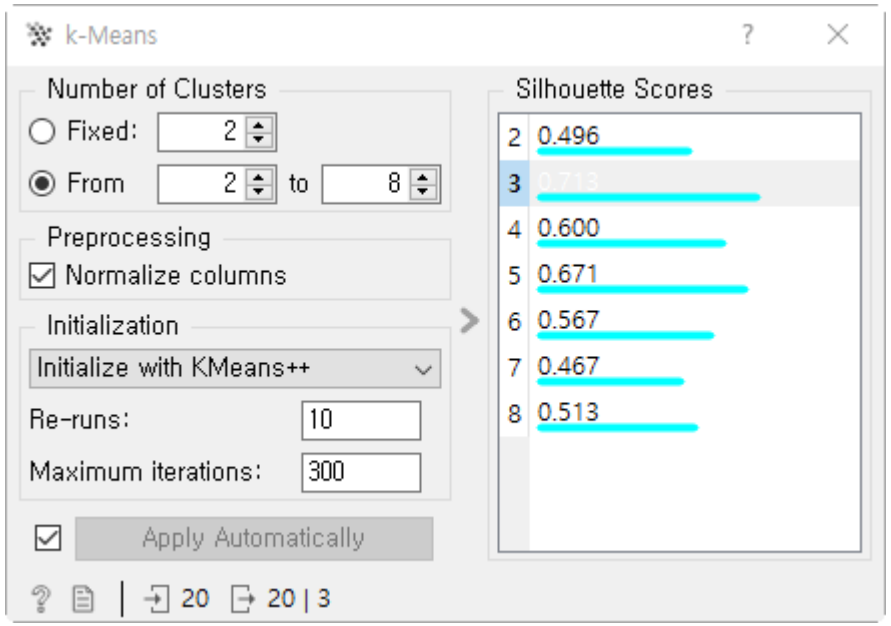




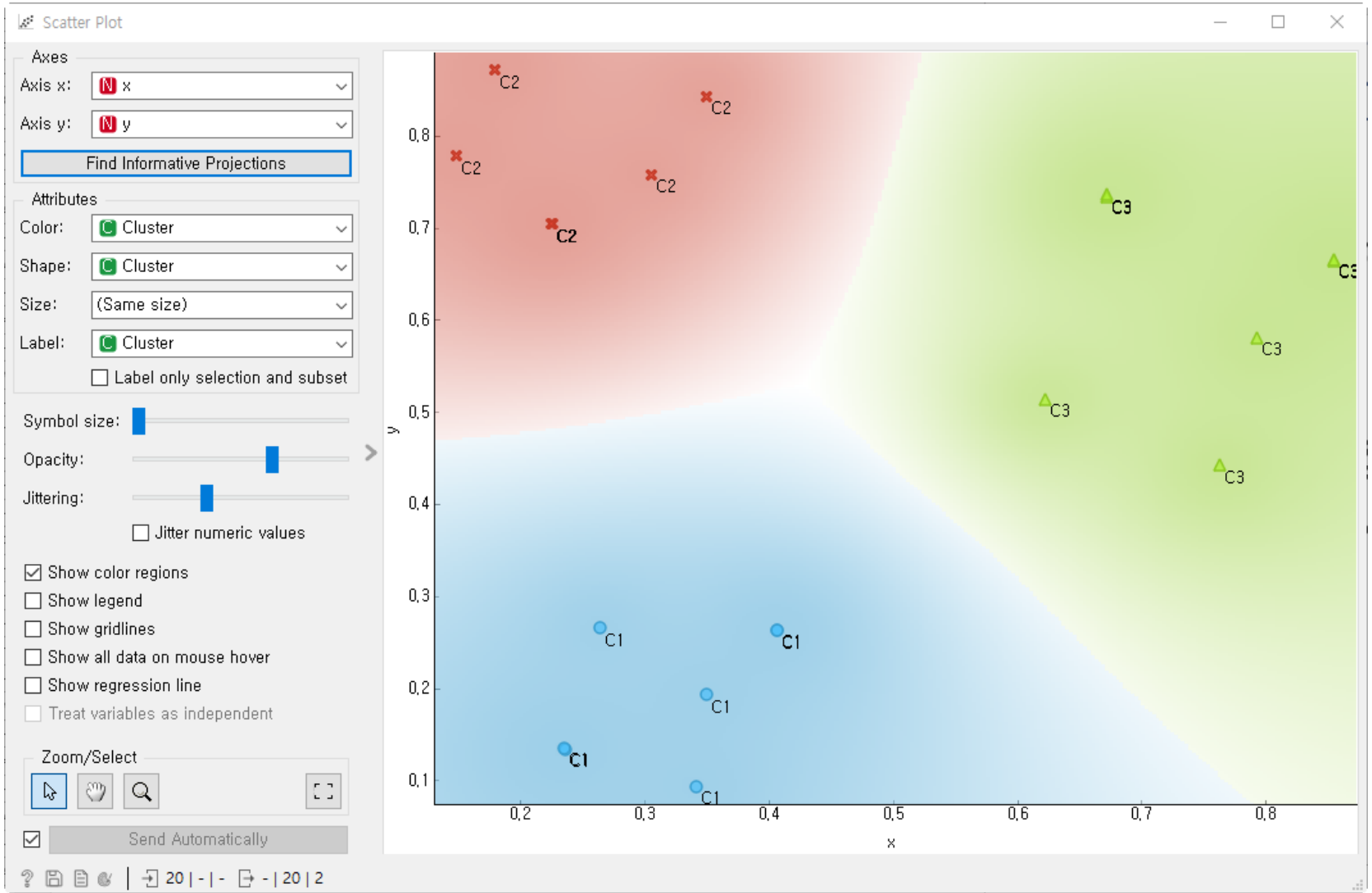
05. 군집 분석

- **적절한 클러스터의 개수 k를 어떻게 선택할까?**
  - 실루엣 점수가 가장 높아지는 k를 선택하기





05. 군집 분석


Scatter Plot
Axes
Axis x: x
Axis y: y
Find Informative Projections
Attributes
Color: Cluster
Shape: Cluster
Size: (Same size)
Label: Cluster
Label only selection and subset
Symbol size:
Opacity:
Jittering:
Jitter numeric values
Show color regions
Show legend
Show gridlines
Show all data on mouse hover
Show regression line
Treat variables as independent
Zoom/Select
Send Automatically
20 | - | -
- | 20 | 2
C1
C2
C3
0,1
0,2
0,3
0,4
0,5
0,6
0,7
0,8
y
x

05. 군집 분석

- ZOO dataset: zoo.tab
  - UCI M/L Repository: https://archive.ics.uci.edu/ml/datasets/zoo

Zoo Data Set

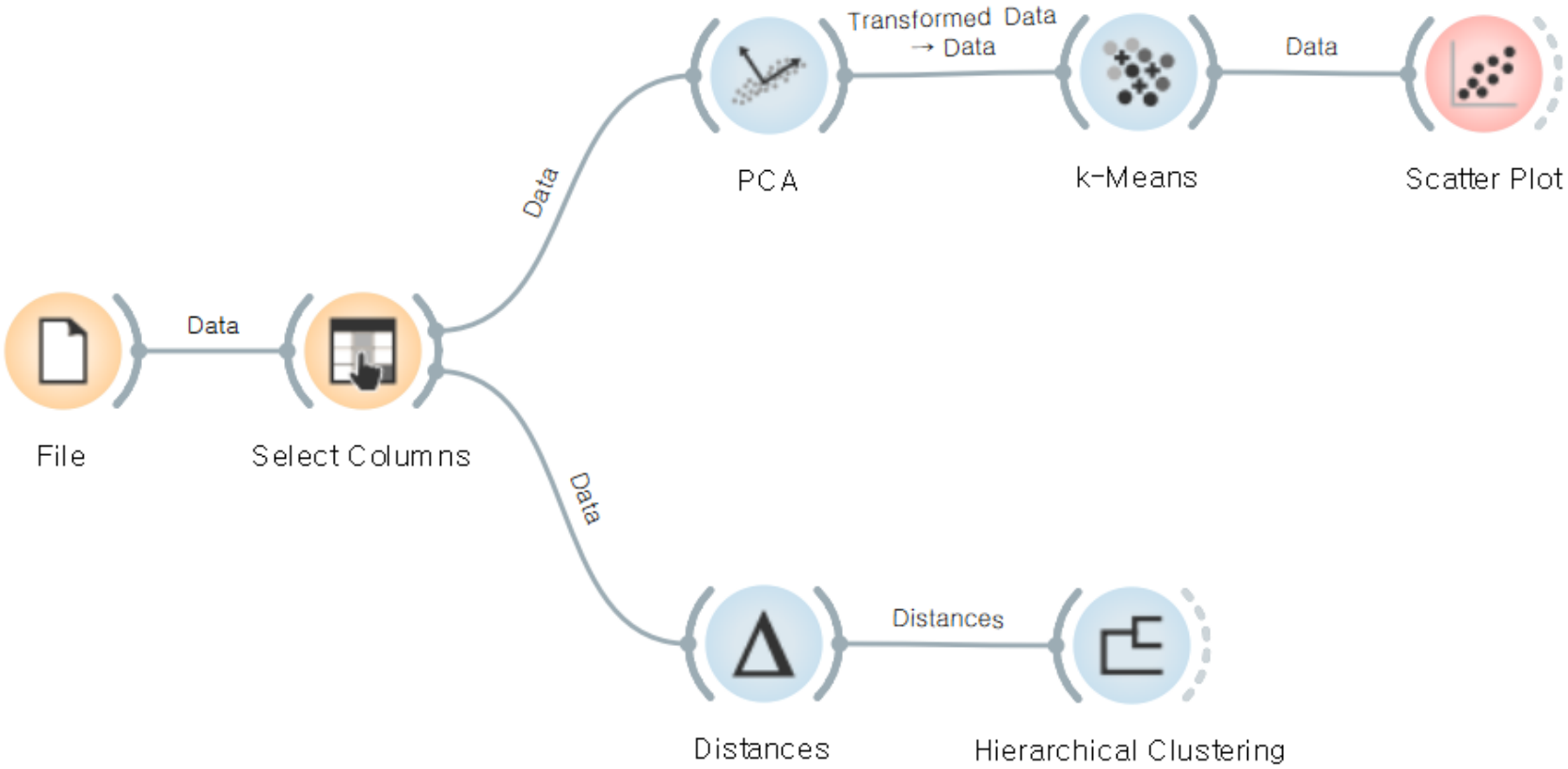
Download: Data Folder, Data Set Description

Abstract: Artificial, 7 classes of animals

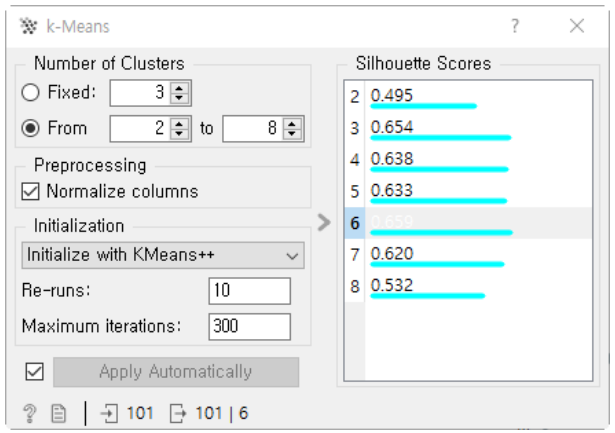
Data Set Characteristics:	Multivariate	**Number of Instances:**	101	**Area:**	Life
Attribute Characteristics:	Categorical, Integer	**Number of Attributes:**	17	**Date Donated**	1990-05-15
Associated Tasks:	Classification	**Missing Values?**	No	**Number of Web Hits:**	365821

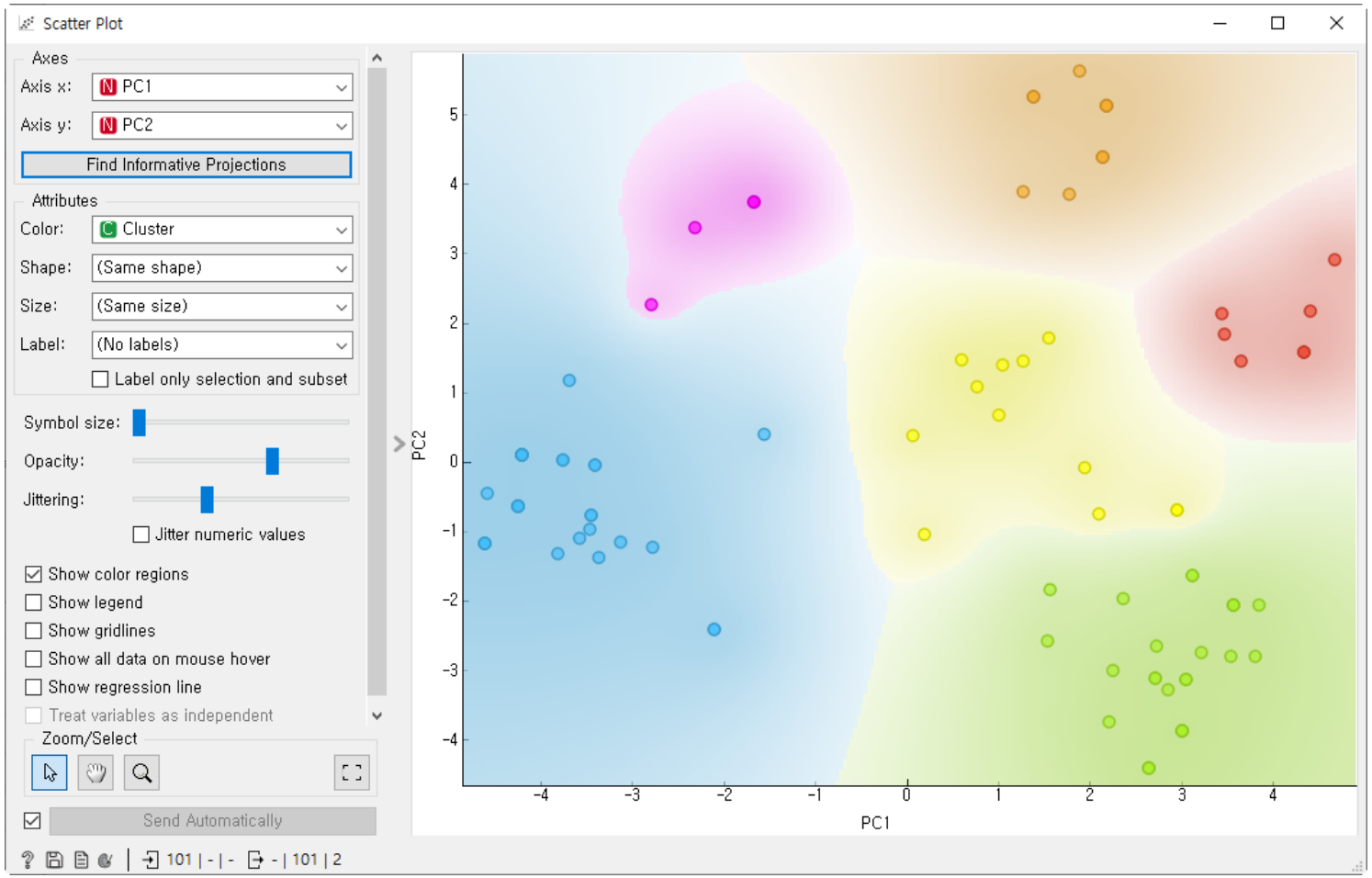
05. 군집 분석

- ZOO dataset 클러스터링:

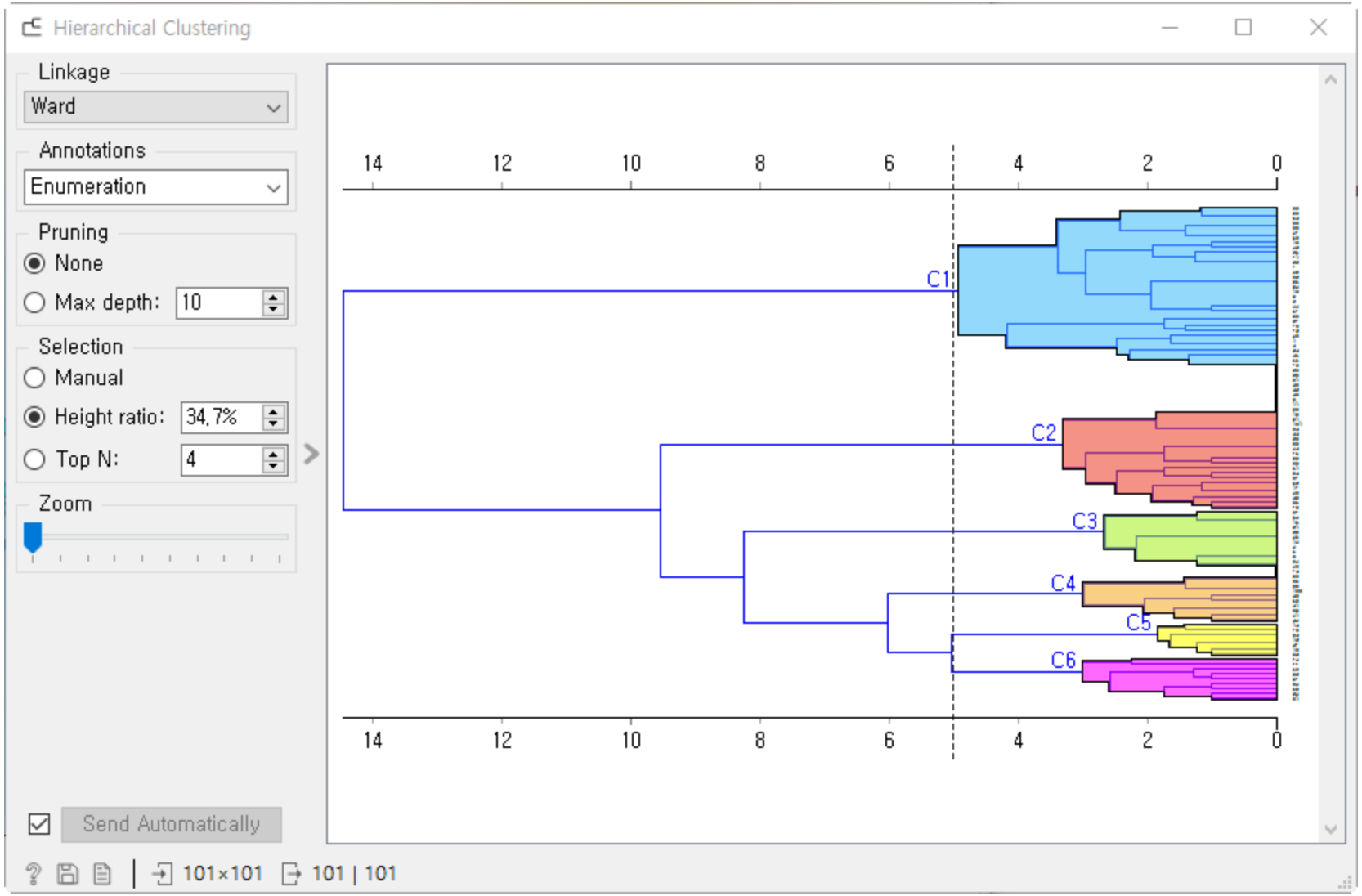


05. 군집 분석


k-Means
Number of Clusters
Fixed: 3
From 2 to 8
Preprocessing
Normalize columns
Initialization
Initialize with KMeans++
Re-runs: 10
Maximum iterations: 300
Apply Automatically
Silhouette Scores
2 0.495
3 0.654
4 0.638
5 0.633
6
7 0.620
8 0.532
101
101 | 6


Scatter Plot
Axes
Axis x: PC1
Axis y: PC2
Find Informative Projections
Attributes
Color: Cluster
Shape: (Same shape)
Size: (Same size)
Label: (No labels)
Label only selection and subset
Symbol size:
Opacity:
Jittering:
Jitter numeric values
Show color regions
Show legend
Show gridlines
Show all data on mouse hover
Show regression line
Treat variables as independent
Zoom/Select
Send Automatically
101 | - | -
- | 101 | 2
PC2
PC1

05. 군집 분석


Hierarchical Clustering
Linkage
Ward
Annotations
Enumeration
Pruning
None
Max depth: 10
Selection
Manual
Height ratio: 34.7%
Top N: 4
Zoom
Send Automatically
101×101
101 | 101
14
12
10
8
6
4
2
0
C1
C2
C3
C4
C5
C6

05. 군집 분석

- R: kmeans()

```
df <- iris[, 1:4]
model <- kmeans(df, center = 3)
model
model$centers
model$cluster

clusplot(df, model$cluster, lines=0, color=T, shade=T)
subset(df, model$cluster==1)
```

05. 군집 분석

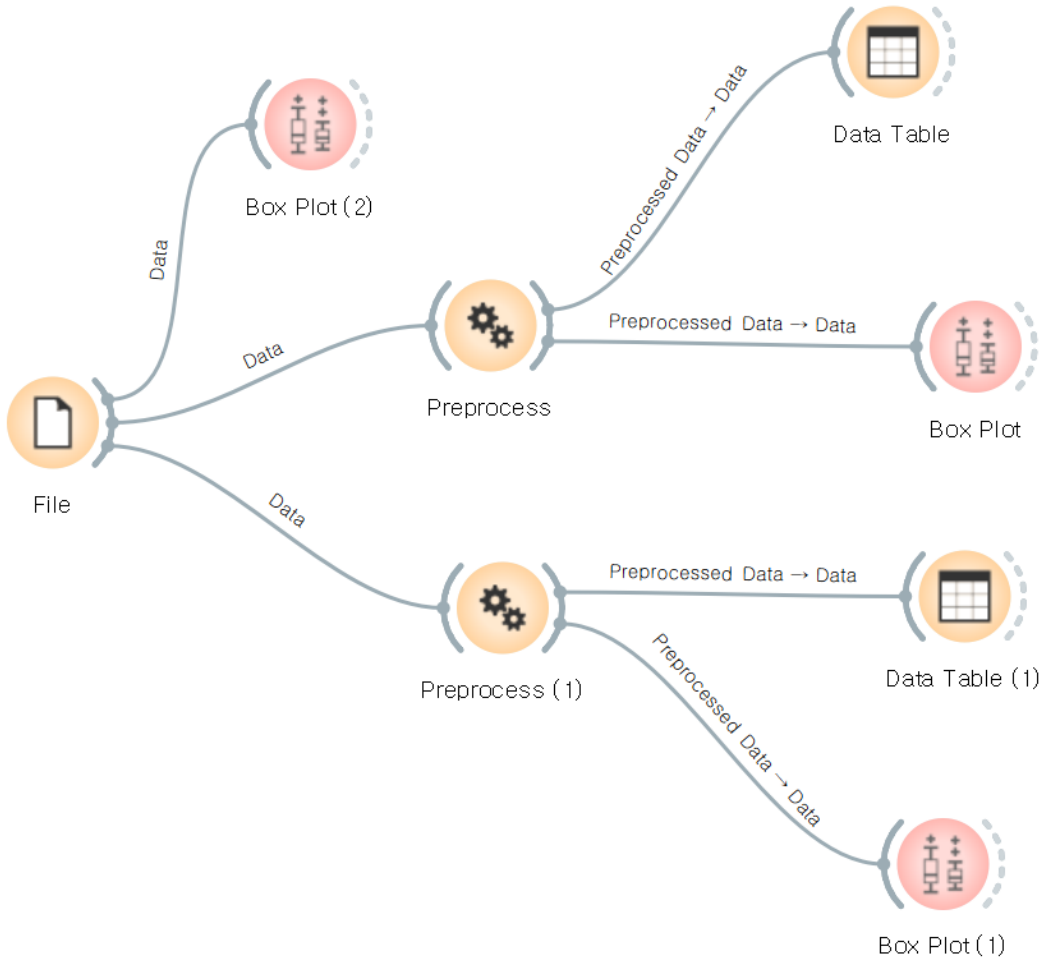
```
> model
K-means clustering with 3 clusters of sizes 62, 38, 50

Cluster means:
  Sepal.Length Sepal.Width Petal.Length Petal.Width
1     5.901613    2.748387     4.393548    1.433871
2     6.850000    3.073684     5.742105    2.071053
3     5.006000    3.428000     1.462000    0.246000

Clustering vector:
  [1] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
 [40] 3 3 3 3 3 3 3 3 3 3 3 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2
 [79] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 2 2 2 1 2 2 2 2 2 2 1 1 2 2
[118] 2 2 1 2 1 2 1 2 2 1 1 2 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 1

Within cluster sum of squares by cluster:
[1] 39.82097 23.87947 15.15100
 (between_SS / total_SS =  88.4 %)
```

05. 군집 분석

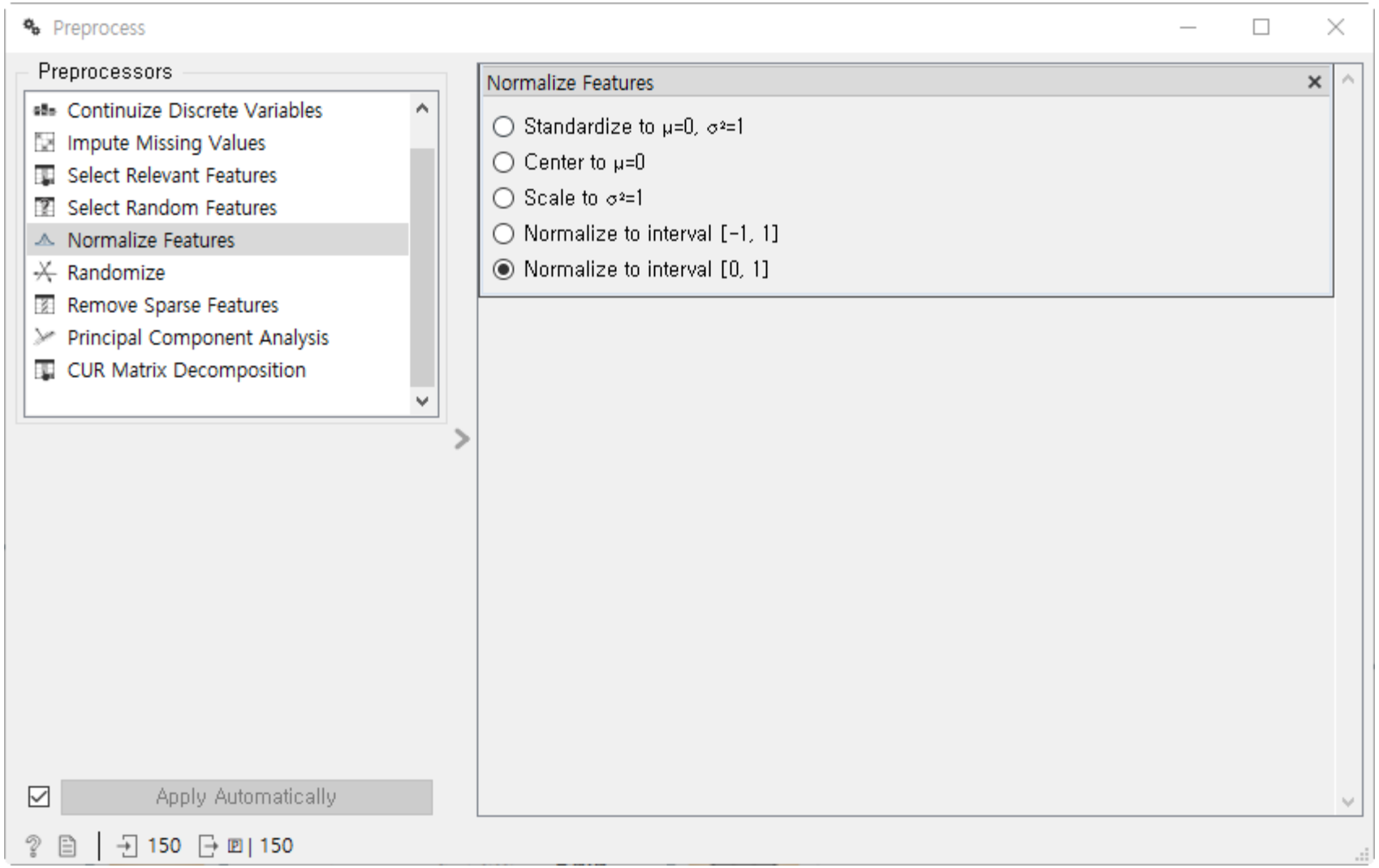
- **데이터 정규화:** *Normalization*
  - 데이터의 특성들이 비슷한 영향력을 행사할 수 있도록 변환해 줌
  - 단위의 차이, 값의 범위 차이 등으로 인한 차이 발생을 완화함
  - Min-Min 정규화: 값의 범위를 [0, 1] 구간에 있도록 변환
    - $X = \frac{X - MIN}{MAX - MIN}$
  - Z-Score 정규화: 값의 분포가 정규 분포를 따르도록 변환
    - $X = \frac{X - \mu}{\sigma}$

05. 군집 분석

- Orange: Preprocess



05. 군집 분석

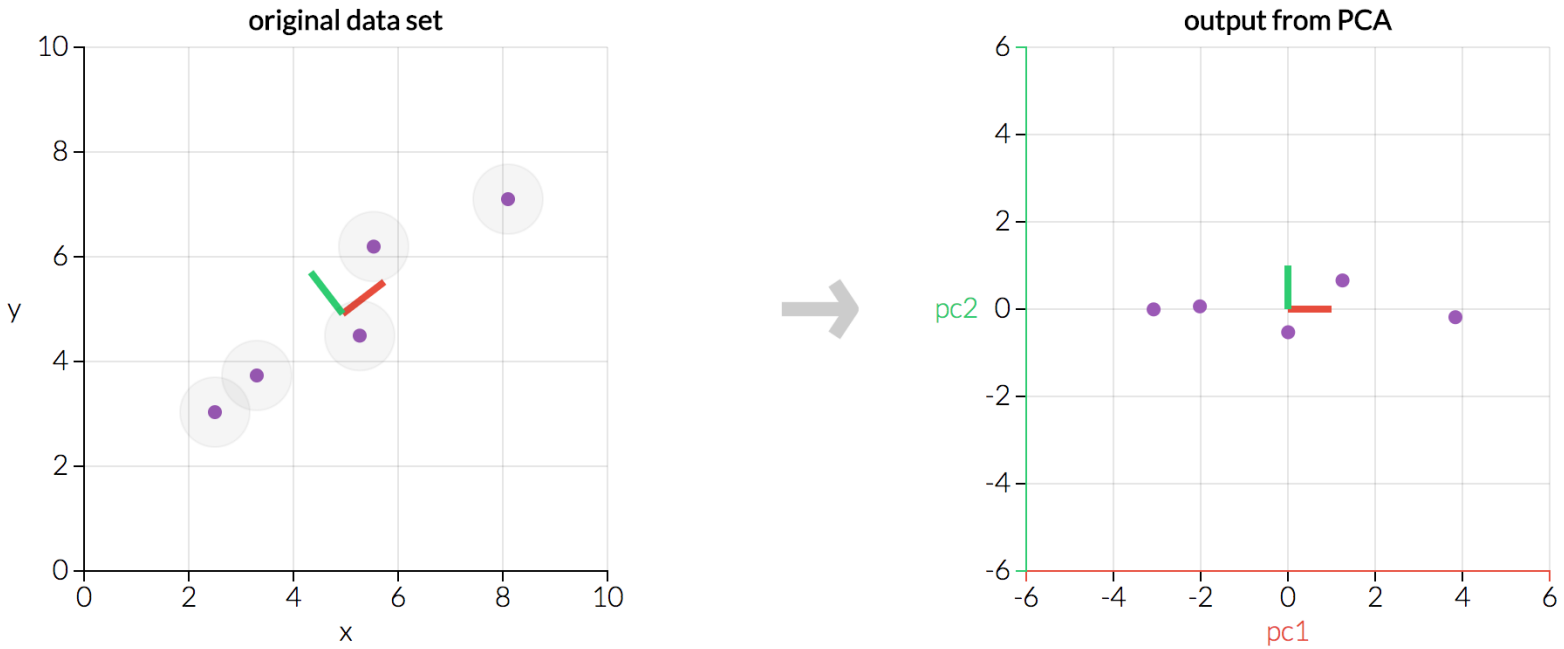

Preprocess
Preprocessors
Continuize Discrete Variables
Impute Missing Values
Select Relevant Features
Select Random Features
Normalize Features
Randomize
Remove Sparse Features
Principal Component Analysis
CUR Matrix Decomposition
Normalize Features
Standardize to μ=0, σ²=1
Center to μ=0
Scale to σ²=1
Normalize to interval [-1, 1]
Normalize to interval [0, 1]
Apply Automatically
150
150

05. 군집 분석

- **차원 축소: *Dimension Reduction***
  - 고차원의 데이터를 2차원(또는 3차원) 데이터로 축소하는 기법
  - 차원 축소가 필요한 이유:
    - 시각화: 2, 3차원의 데이터는 산점도 등으로 시각화가 가능
    - 노이즈 제거: feature의 수가 줄어들기 때문에 노이즈도 같이 제거됨
    - 성능의 향상: 차원의 수가 작으면 처리 속도도 늘어남
  - 차원 축소의 단점:
    - 정보의 손실: 차원의 축소 과정에서 정보의 왜곡 손실 현상이 발생함

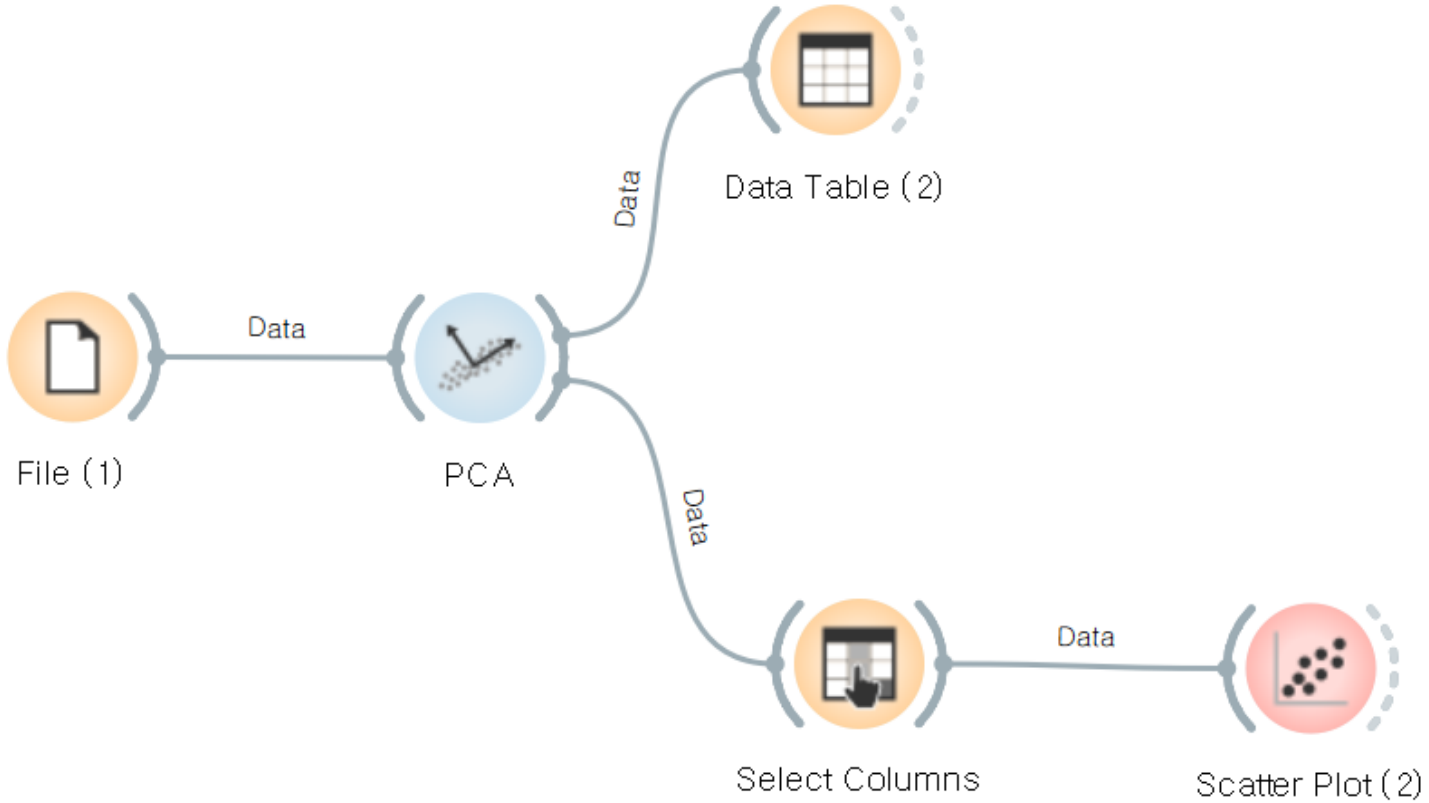
05. 군집 분석

- **주성분 분석: PCA, *Principal Component Analysis***
  - **어떤 데이터 분포를 분산이 가장 큰 방향으로 정사영하여 주성분을 구함**

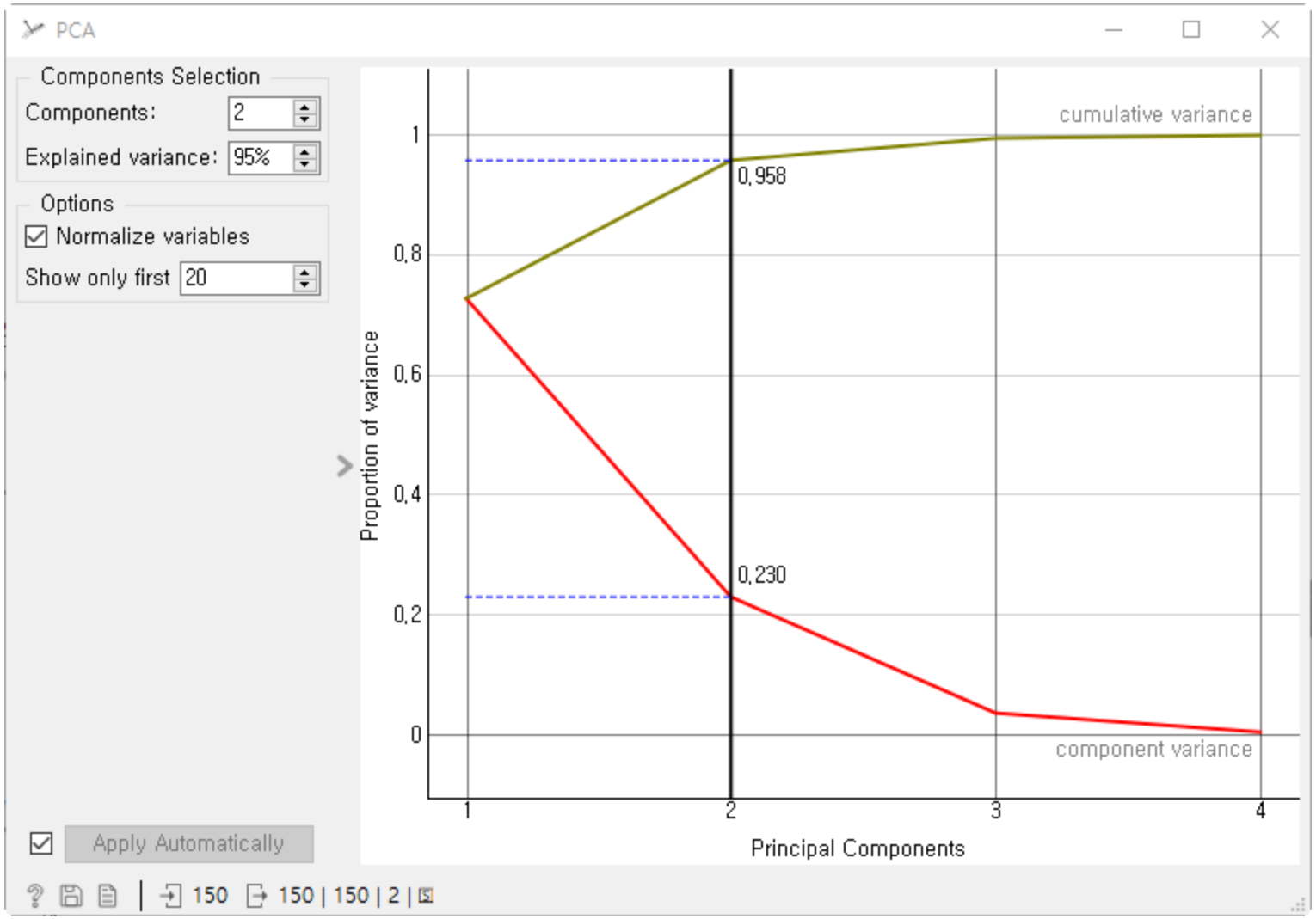


05. 군집 분석

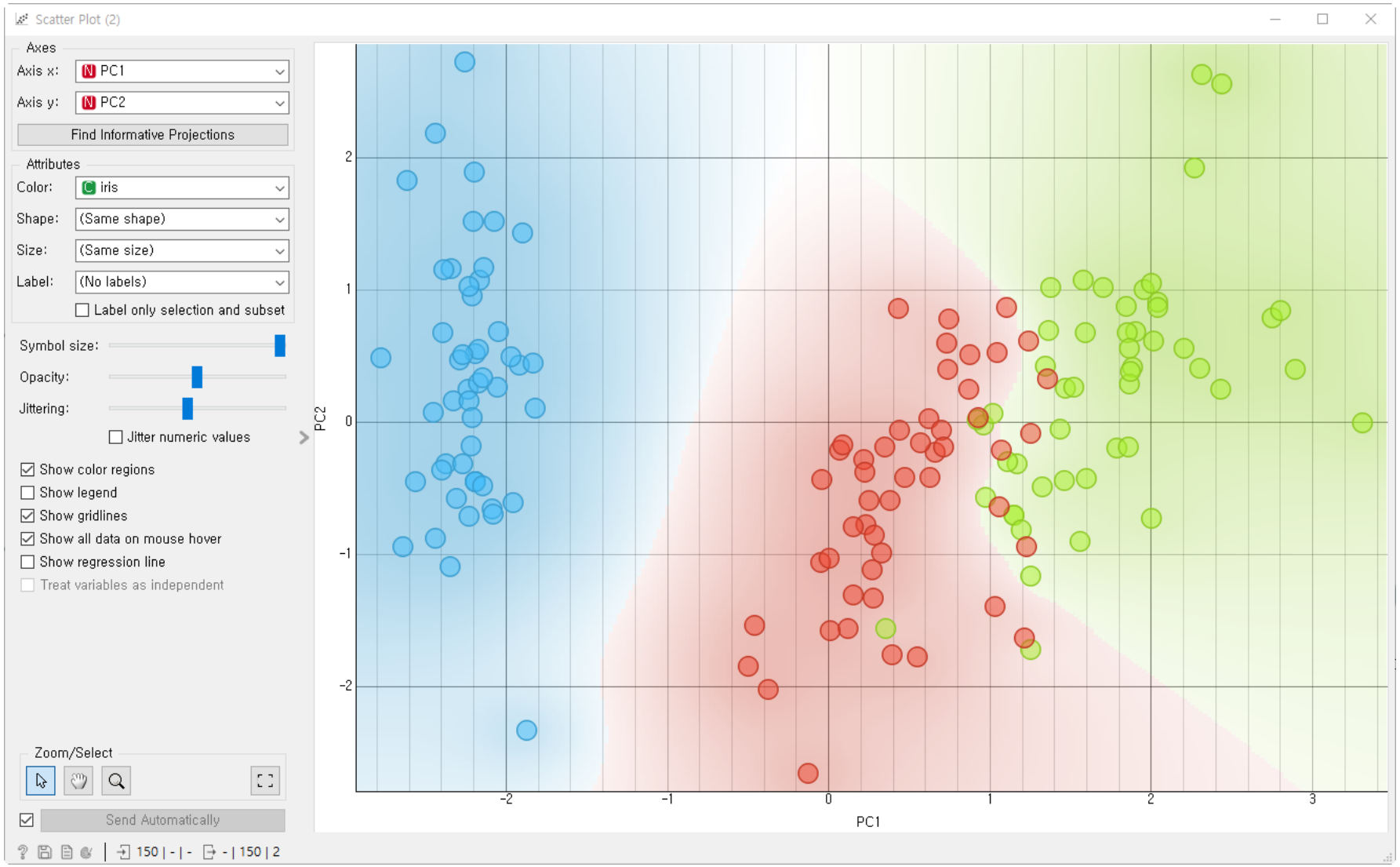
- Orange: PCA



05. 군집 분석


PCA
Components Selection
Components: 2
Explained variance: 95%
Options
Normalize variables
Show only first 20
Apply Automatically
150
150 | 150 | 2 |
cumulative variance
component variance
0,958
0,230
Proportion of variance
Principal Components
1
0,8
0,6
0,4
0,2
0
1
2
3
4

05. 군집 분석


Scatter Plot (2)
Axes
Axis x: PC1
Axis y: PC2
Find Informative Projections
Attributes
Color: iris
Shape: (Same shape)
Size: (Same size)
Label: (No labels)
Label only selection and subset
Symbol size:
Opacity:
Jittering:
Jitter numeric values
Show color regions
Show legend
Show gridlines
Show all data on mouse hover
Show regression line
Treat variables as independent
Zoom/Select
Send Automatically
150 | - | -
- | 150 | 2
PC2
PC1
2
1
0
-1
-2
-2
-1
0
1
2
3

Any Questions?