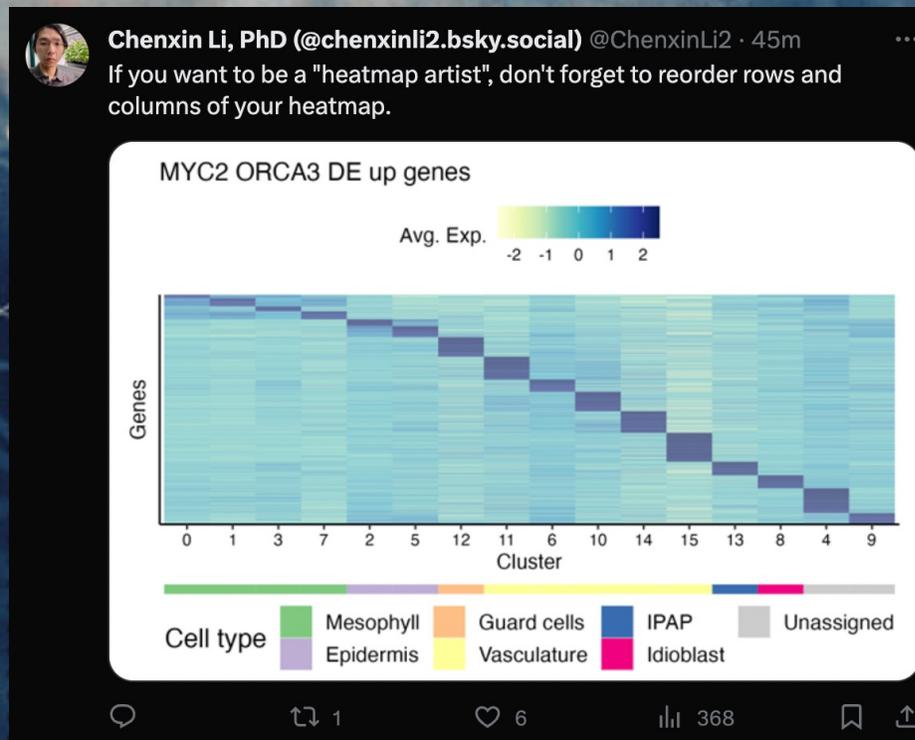


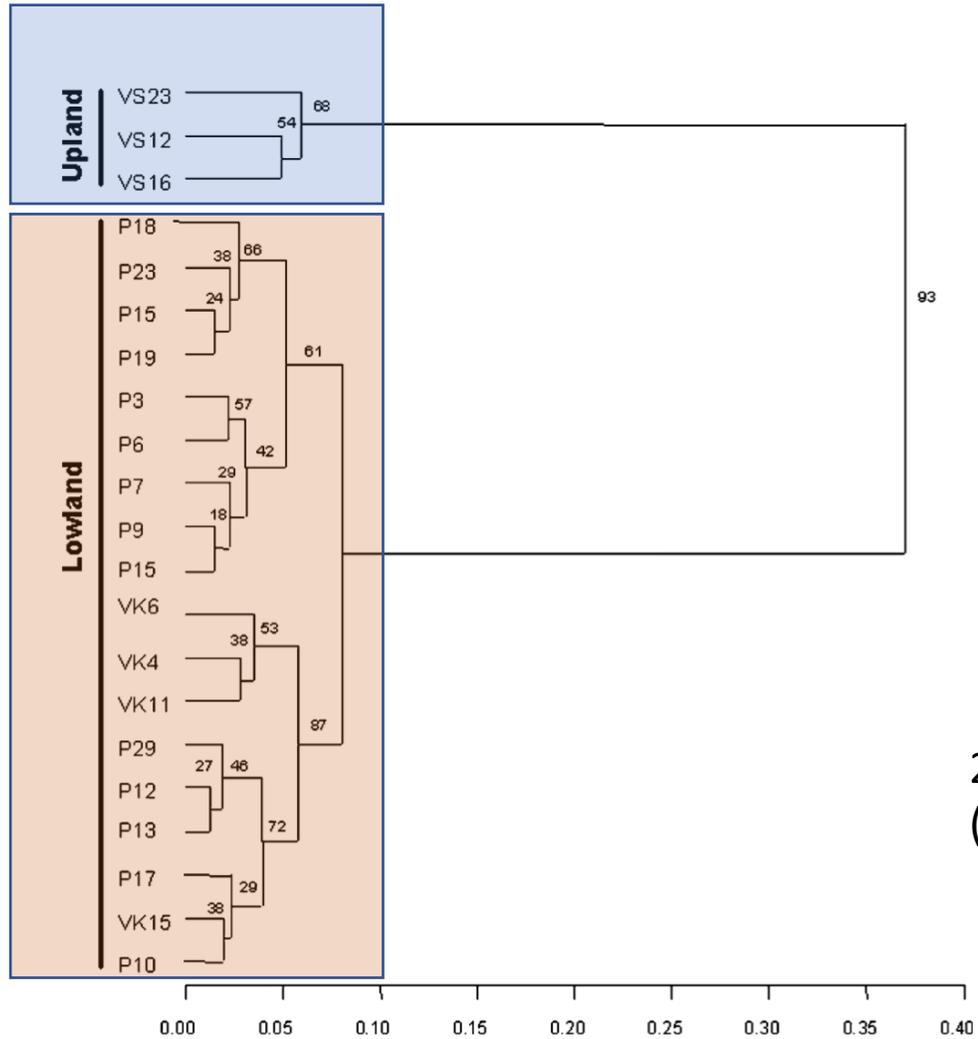
• Hierarchical clustering



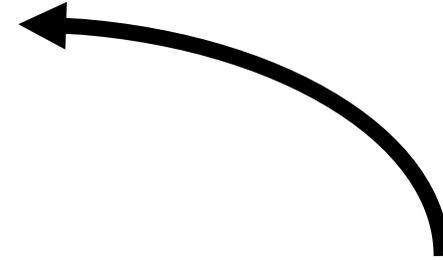
Upland ecotype



Lowland ecotype

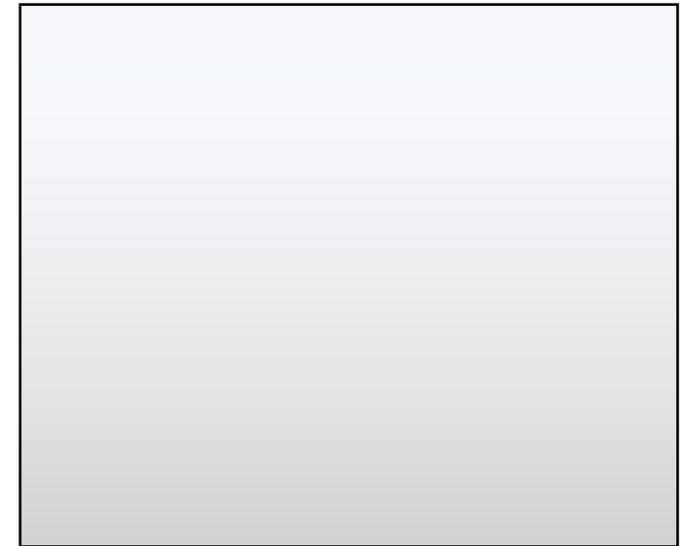


Hierarchical clustering



Data: 85 genetic markers

22 plants
(observations)



K=2



Mammals



Birds



K=6



Squirrel



Mouse



Gorilla



Duck



Really cool
looking duck



Eagle

K=5



Squirrel



Mouse



Gorilla



Duck

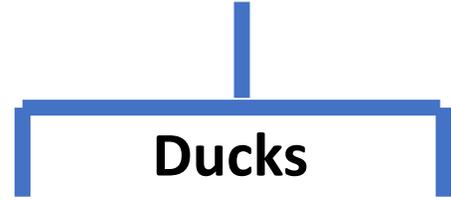


Really cool looking duck



Eagle

K=4



Squirrel



Mouse



Gorilla



Duck

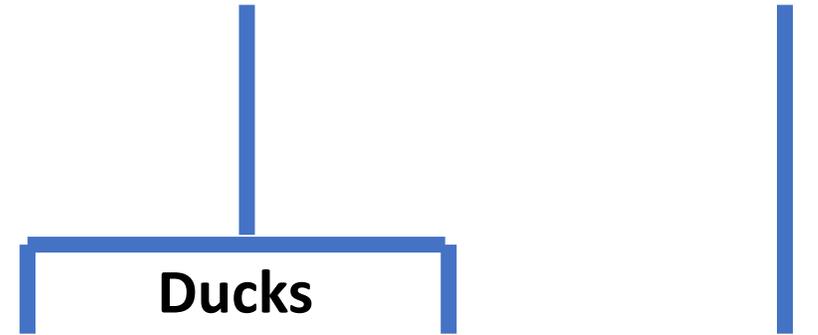
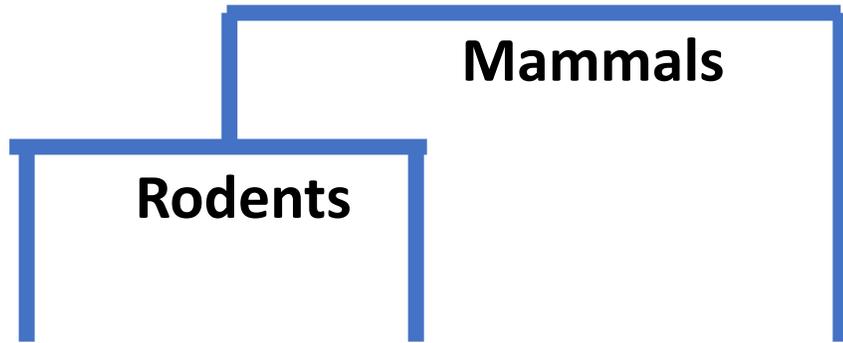


Really cool
looking duck



Eagle

K=3



Squirrel



Mouse



Gorilla



Duck

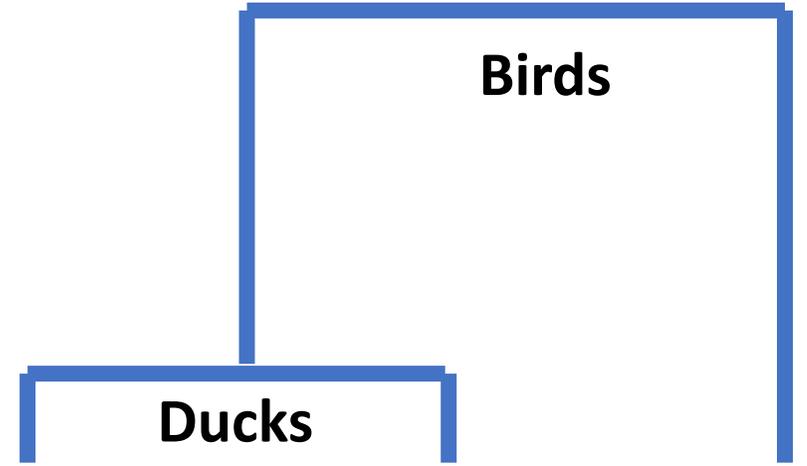
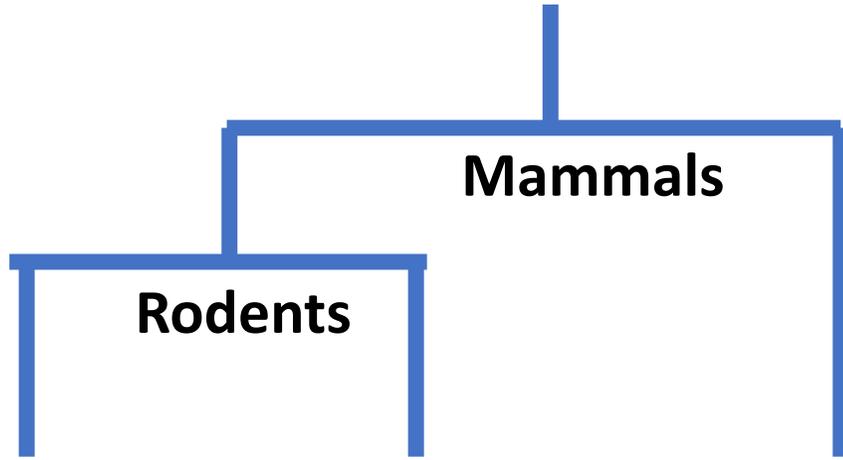


Really cool looking duck



Eagle

K=2



Squirrel



Mouse



Gorilla



Duck



Really cool looking duck



Eagle

K=1



Squirrel



Mouse



Gorilla



Duck

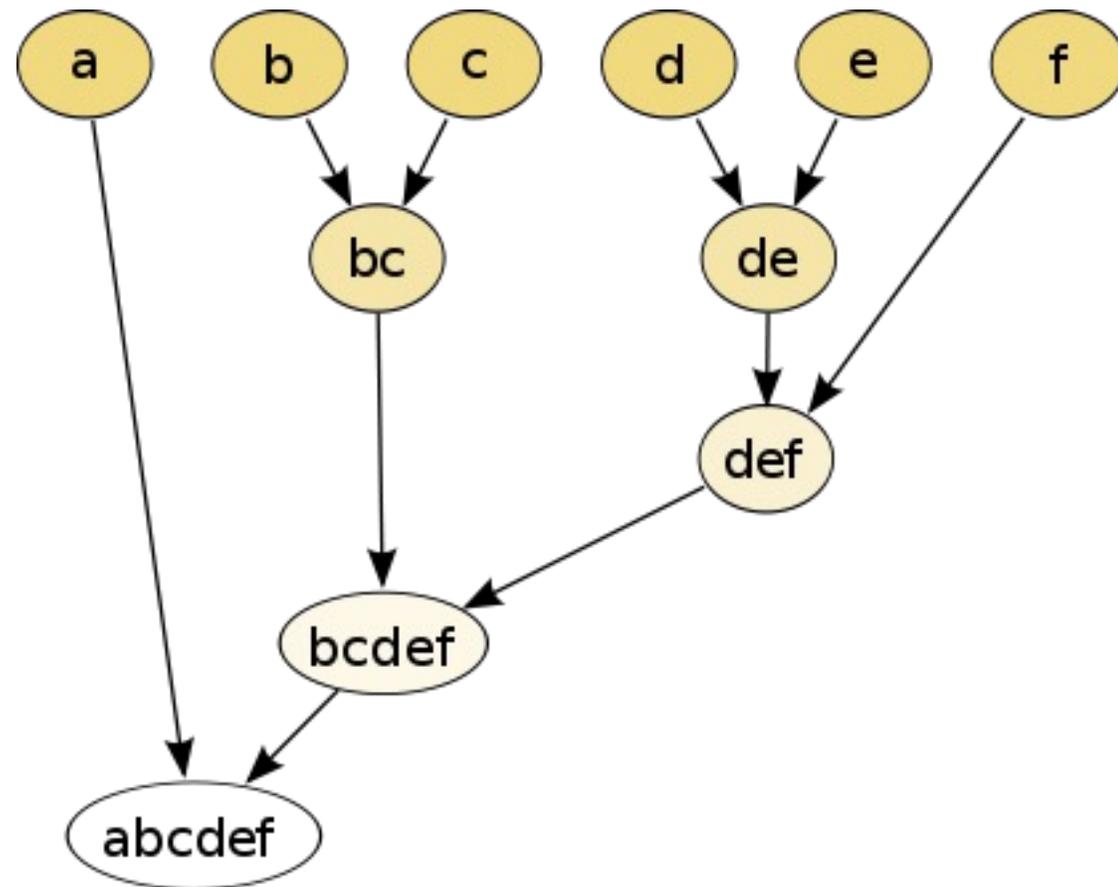


Really cool looking duck

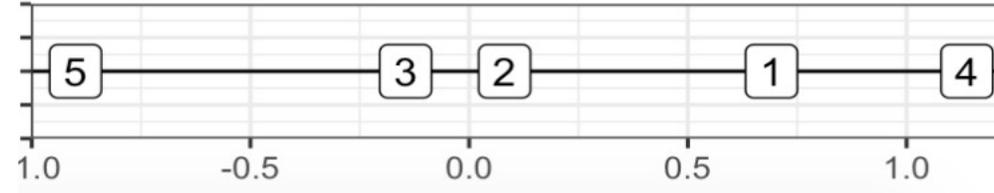


Eagle

Dendrogram



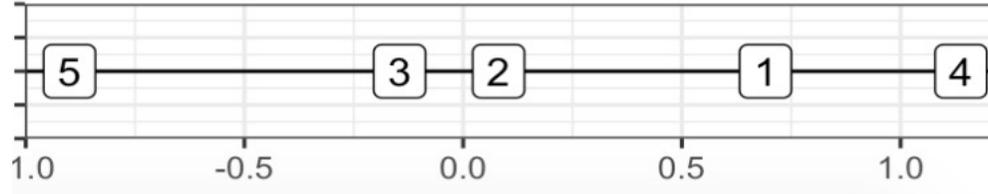
Variation



Variation



Distance matrix



Distance matrix

	1	2	3	4
2	0.6100783			
3	0.8363964	0.2263181		
4	0.4471031	1.0571814	1.2834995	
5	1.5908574	0.9807791	0.7544610	2.0379605

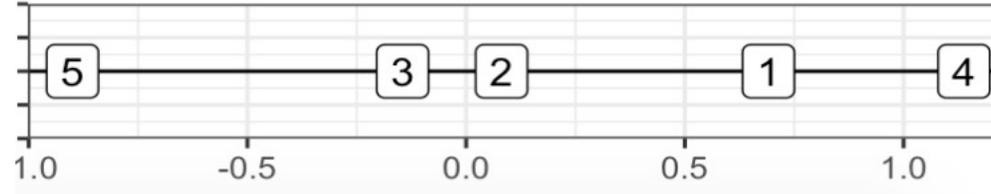
Variation



Distance matrix

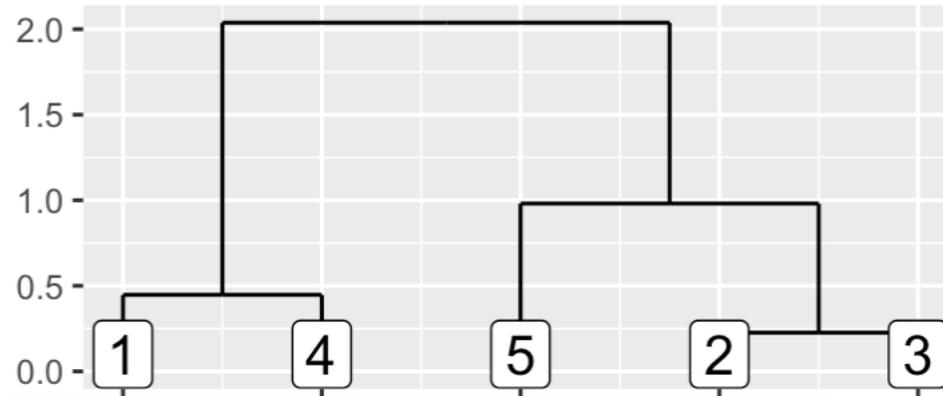


Hierarchy of distance



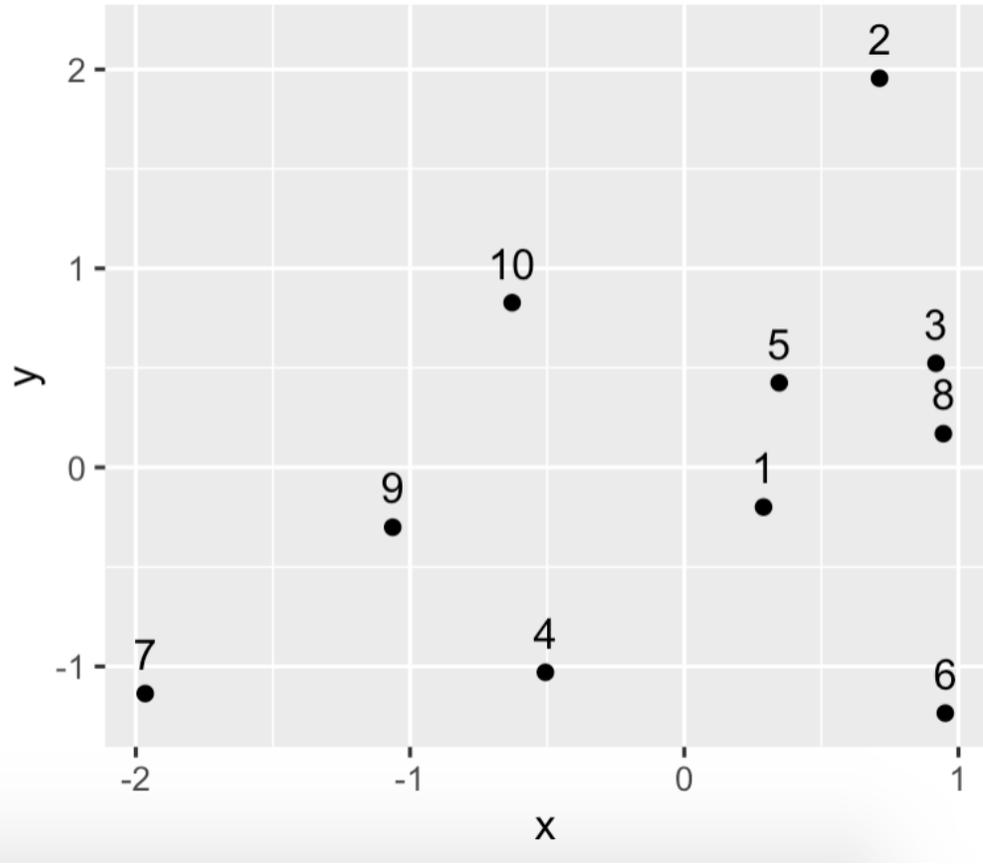
Distance matrix

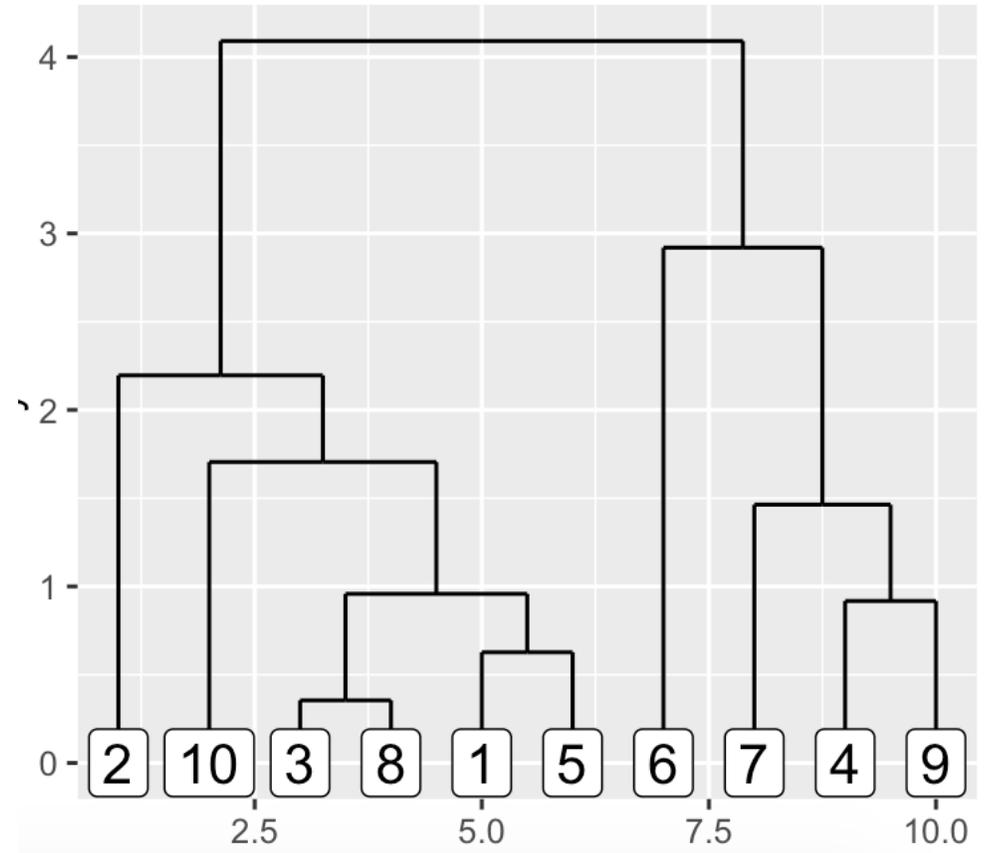
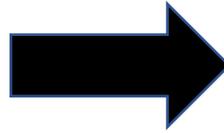
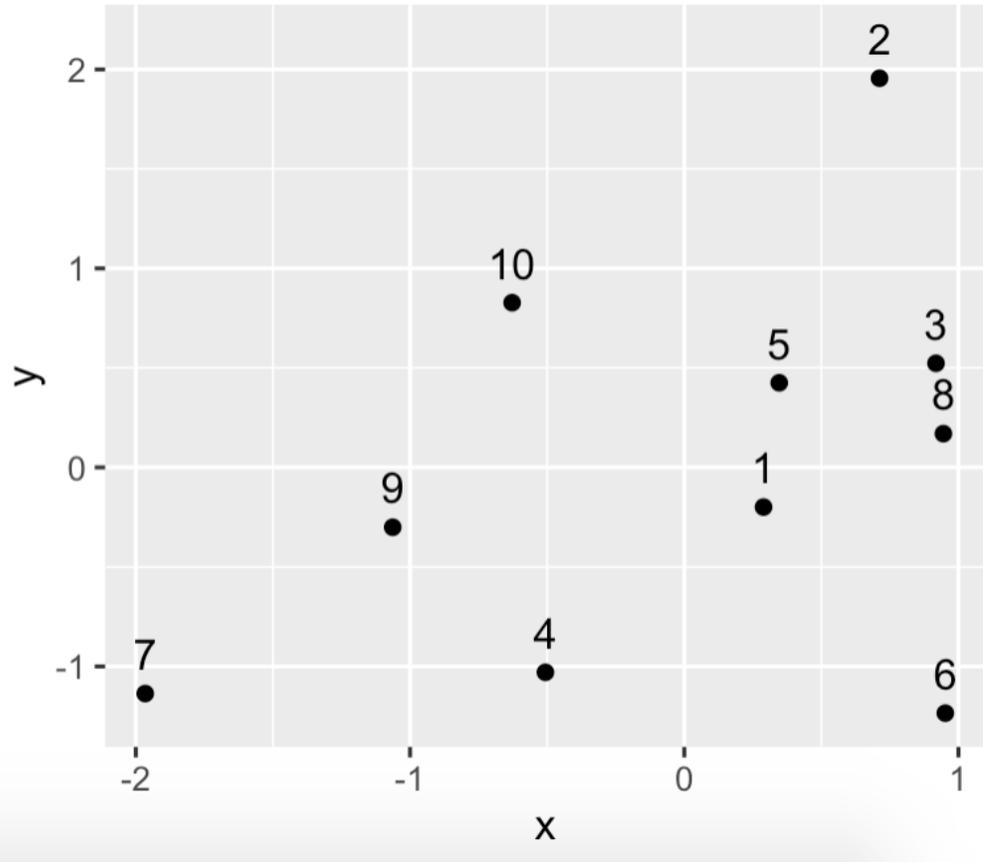
	1	2	3	4
2	0.6100783			
3	0.8363964	0.2263181		
4	0.4471031	1.0571814	1.2834995	
5	1.5908574	0.9807791	0.7544610	2.0379605



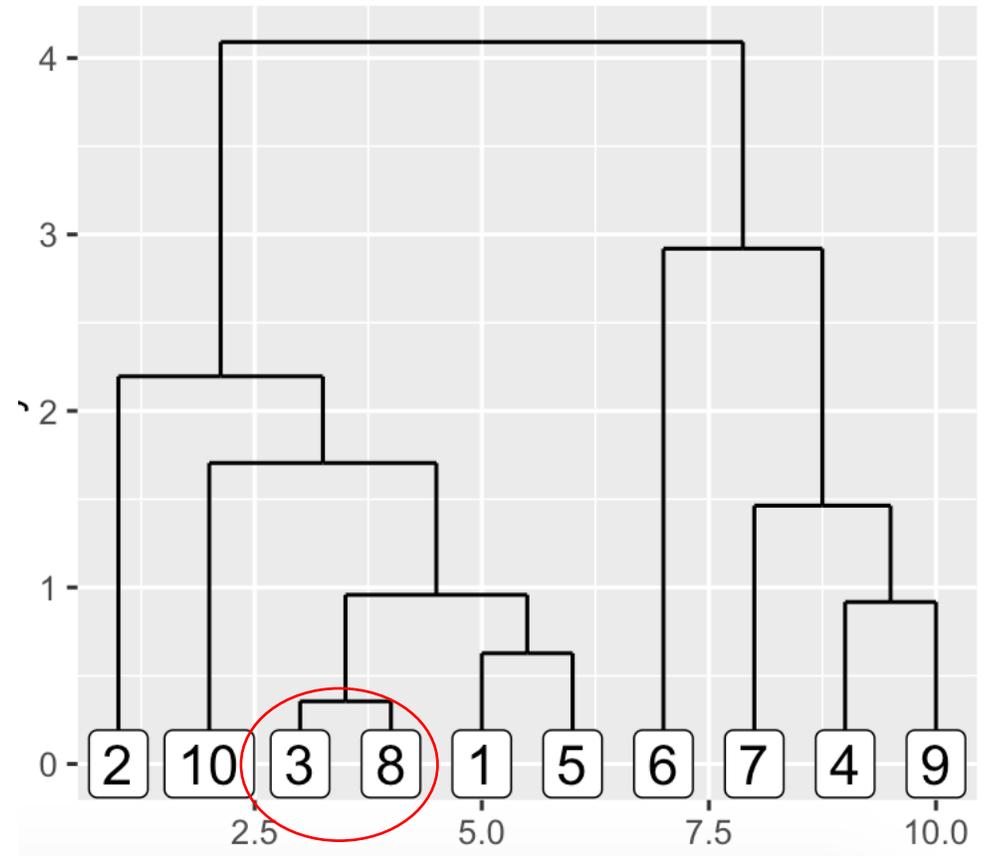
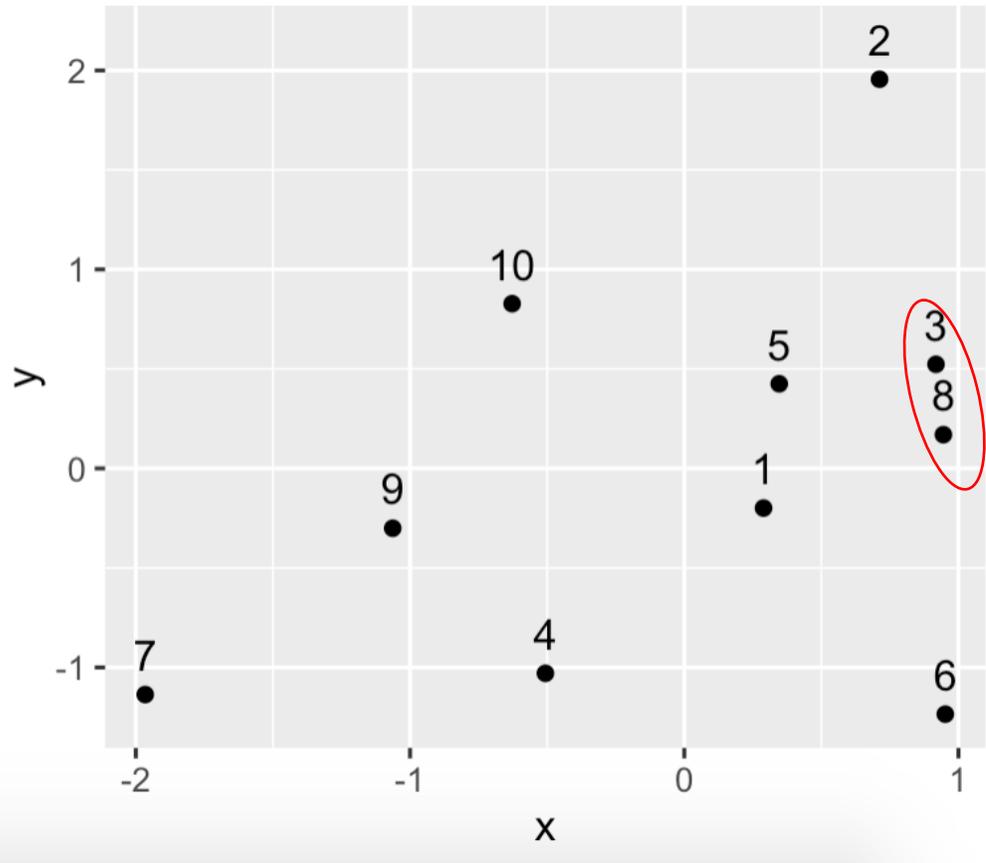
Hierarchical clustering

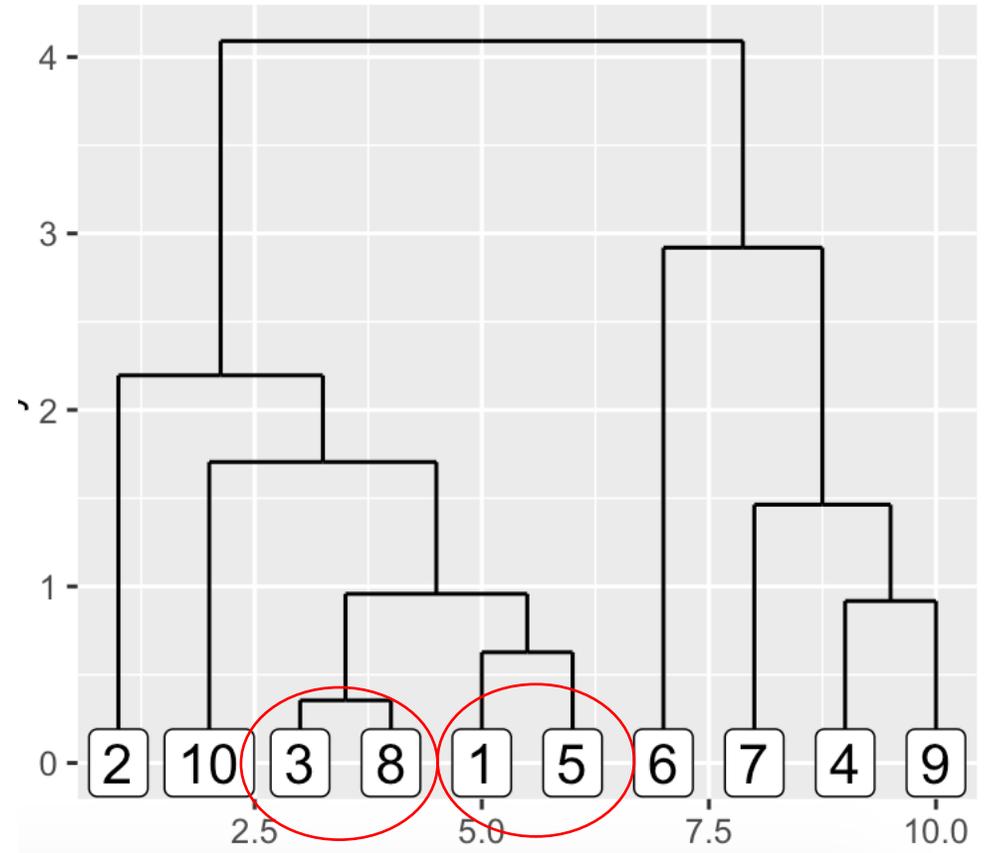
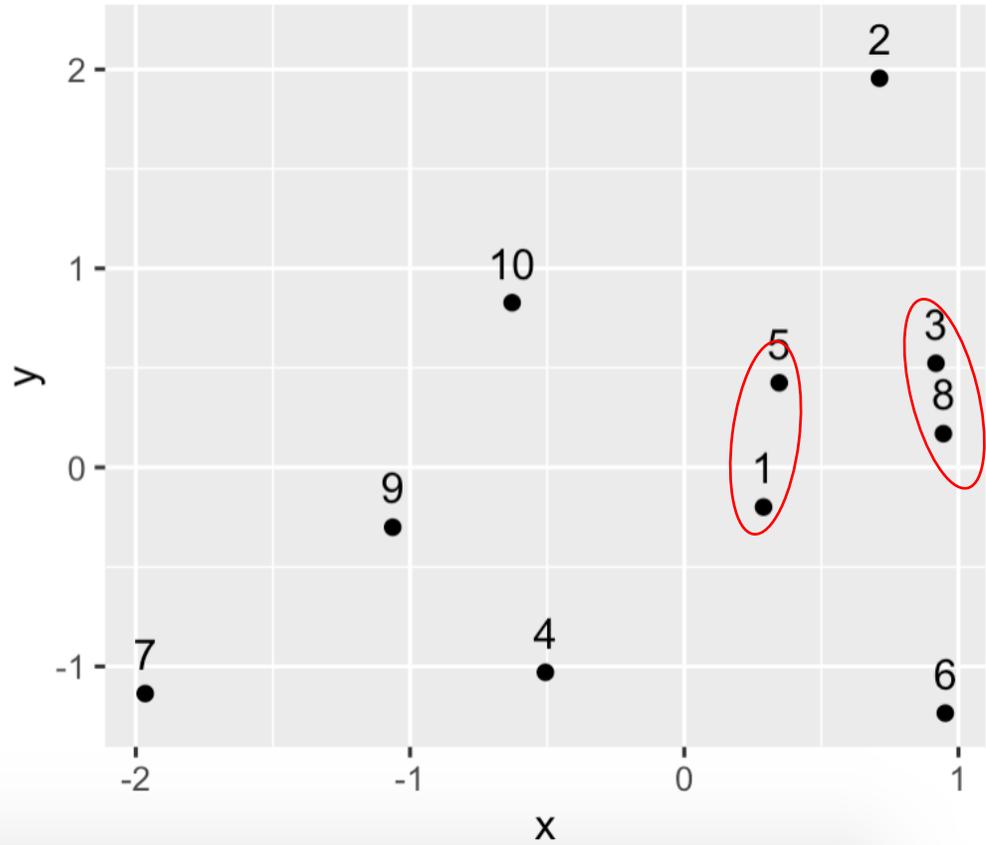
- Bottom up clustering
- Methods of merging clusters with `hclust()`
 - E.g. “complete” vs “single” linkage
- Measuring distance
 - E.g. “euclidean” vs “manhattan” distance
- Examples of hierarchical clustering in research

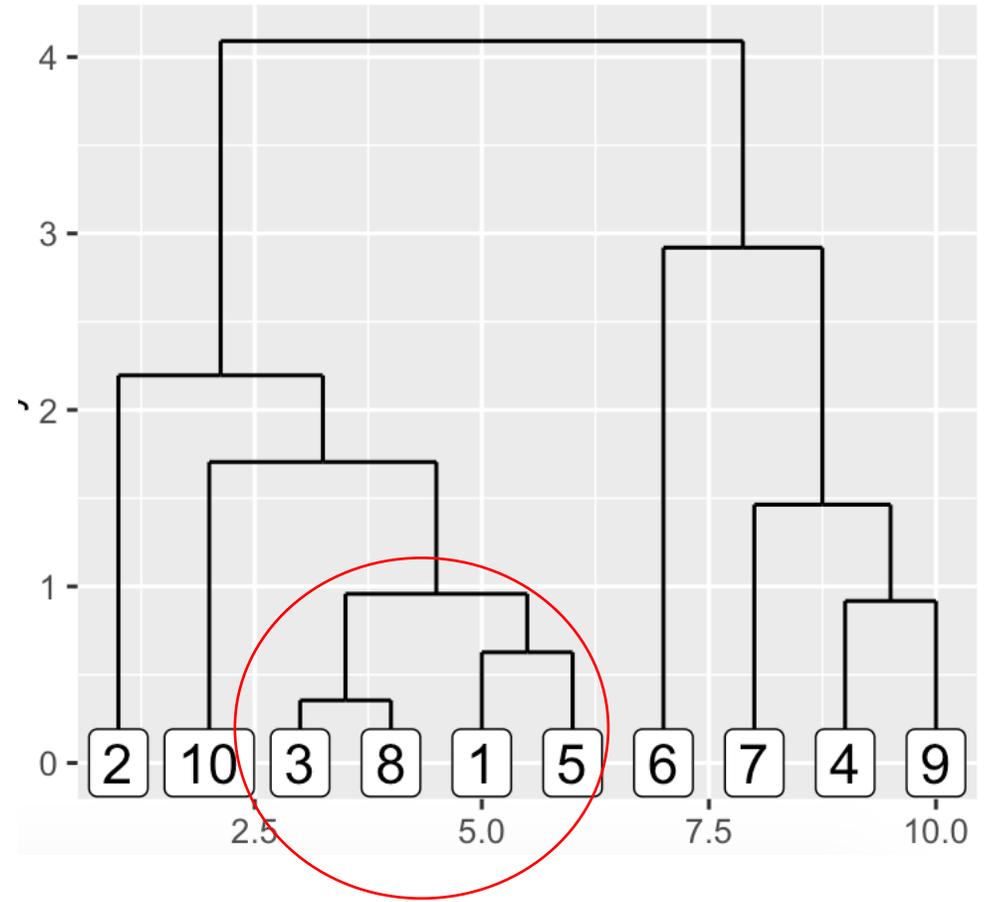
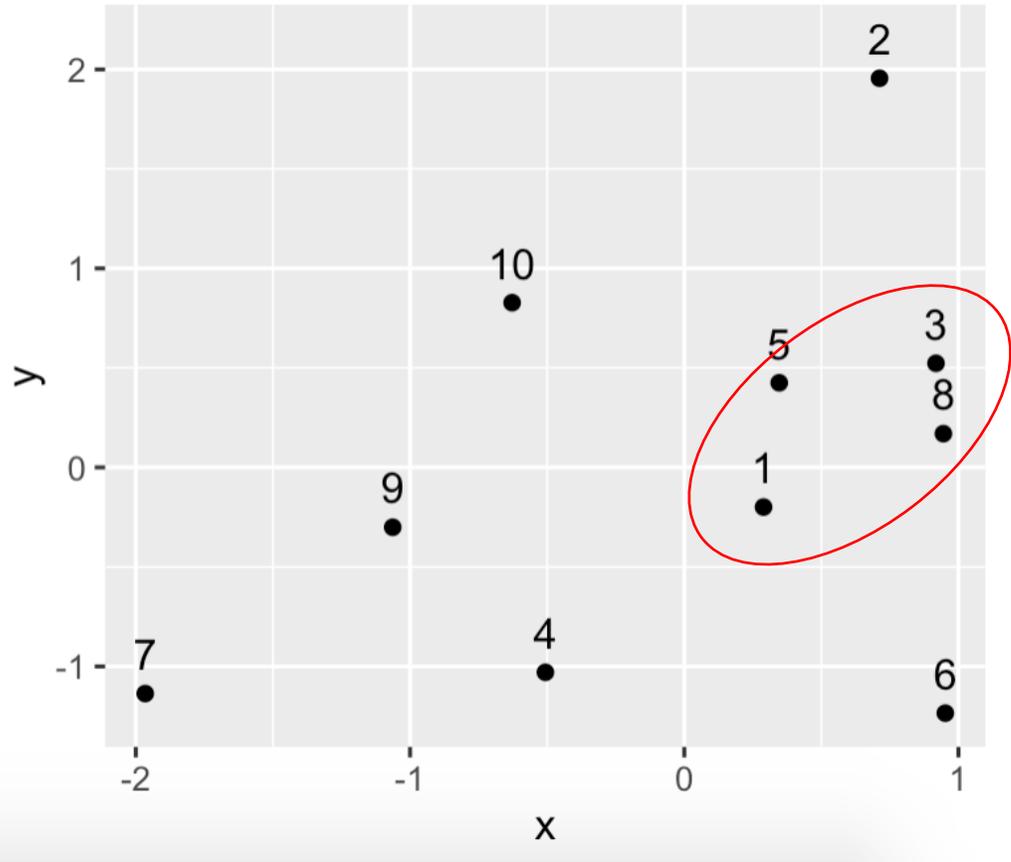


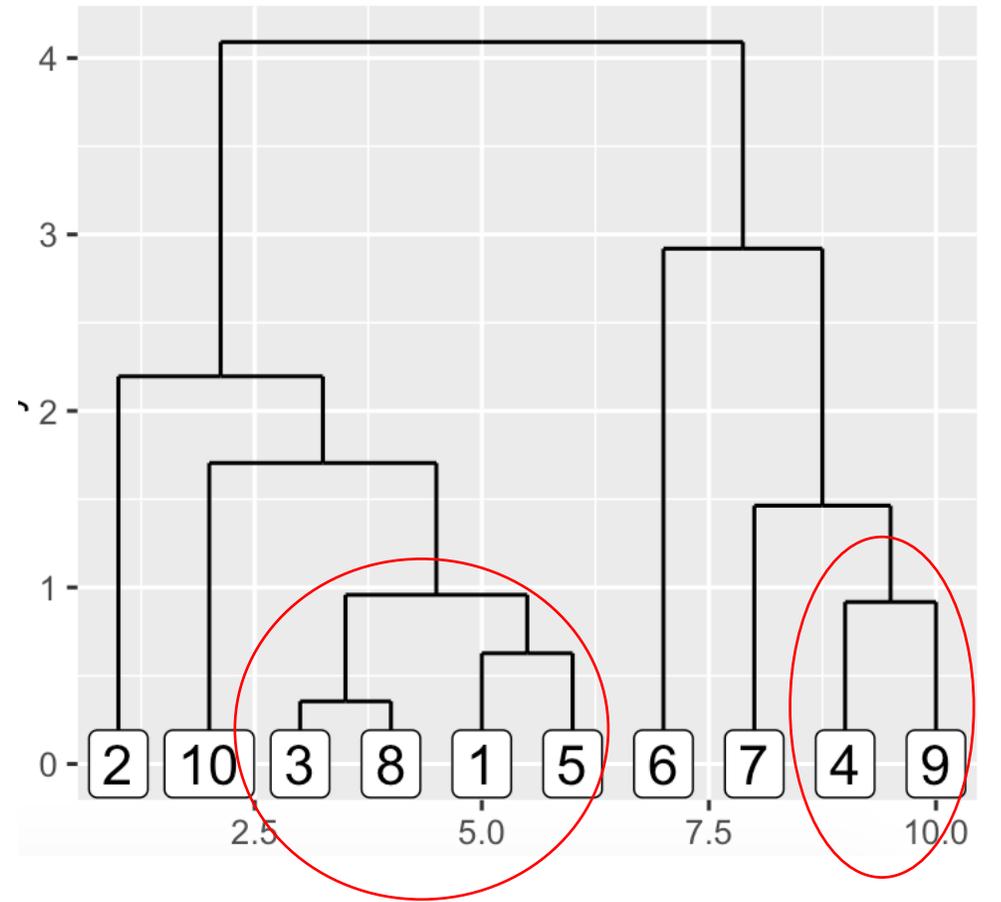
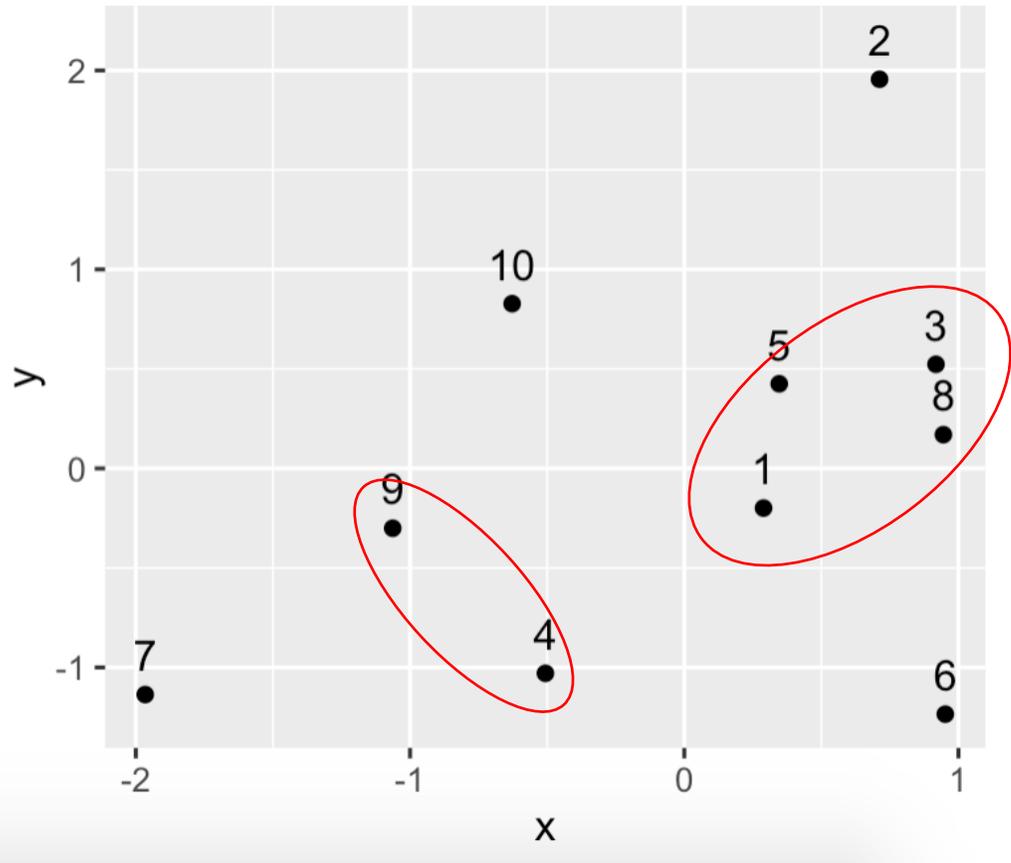


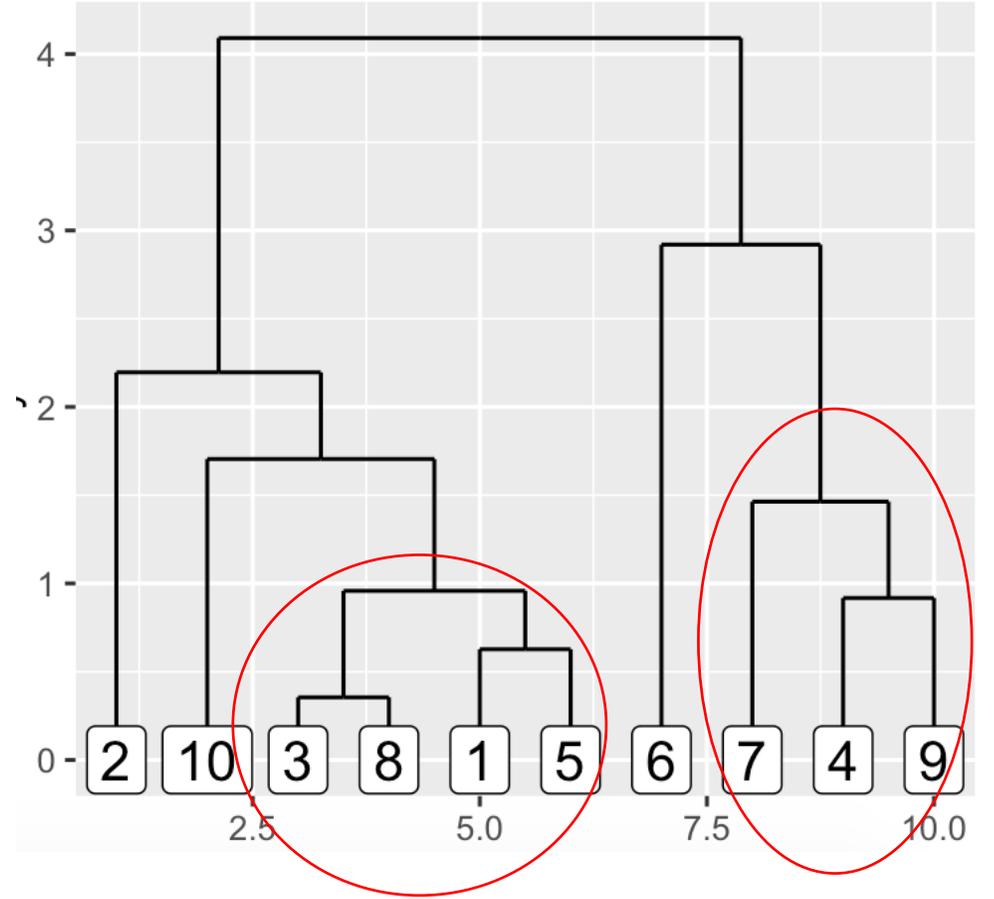
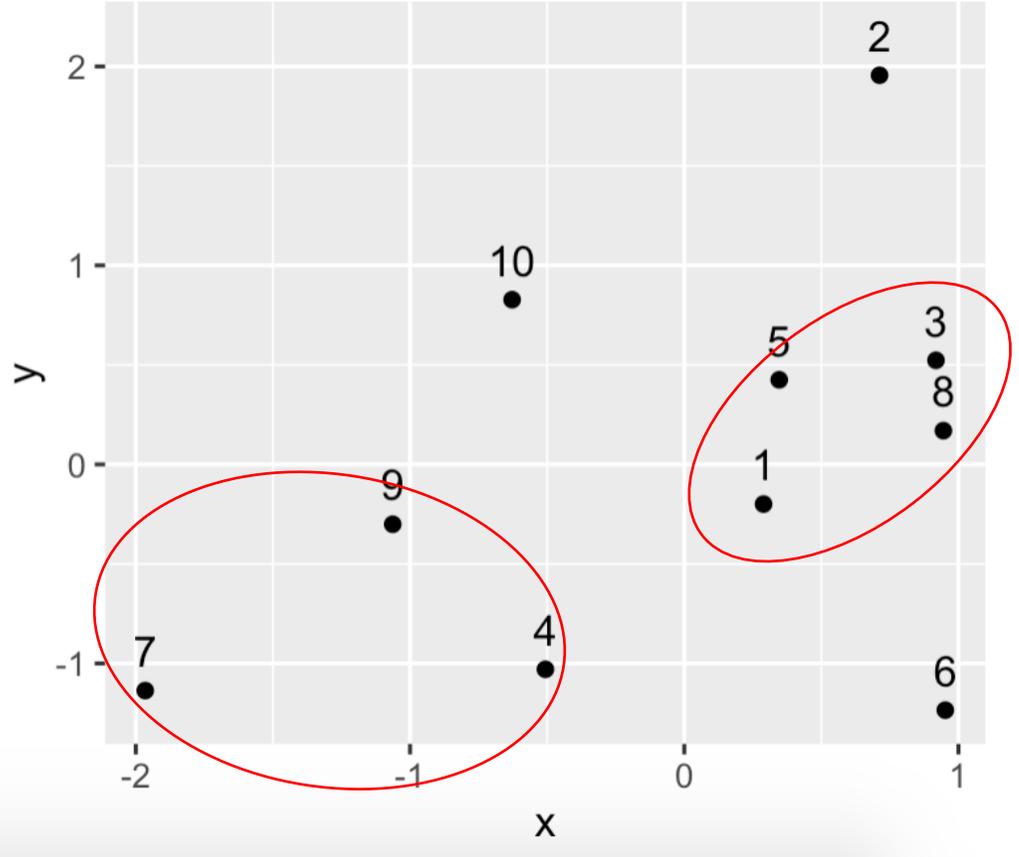
Hierarchical clustering merges the "closest things" in a bottom up fashion

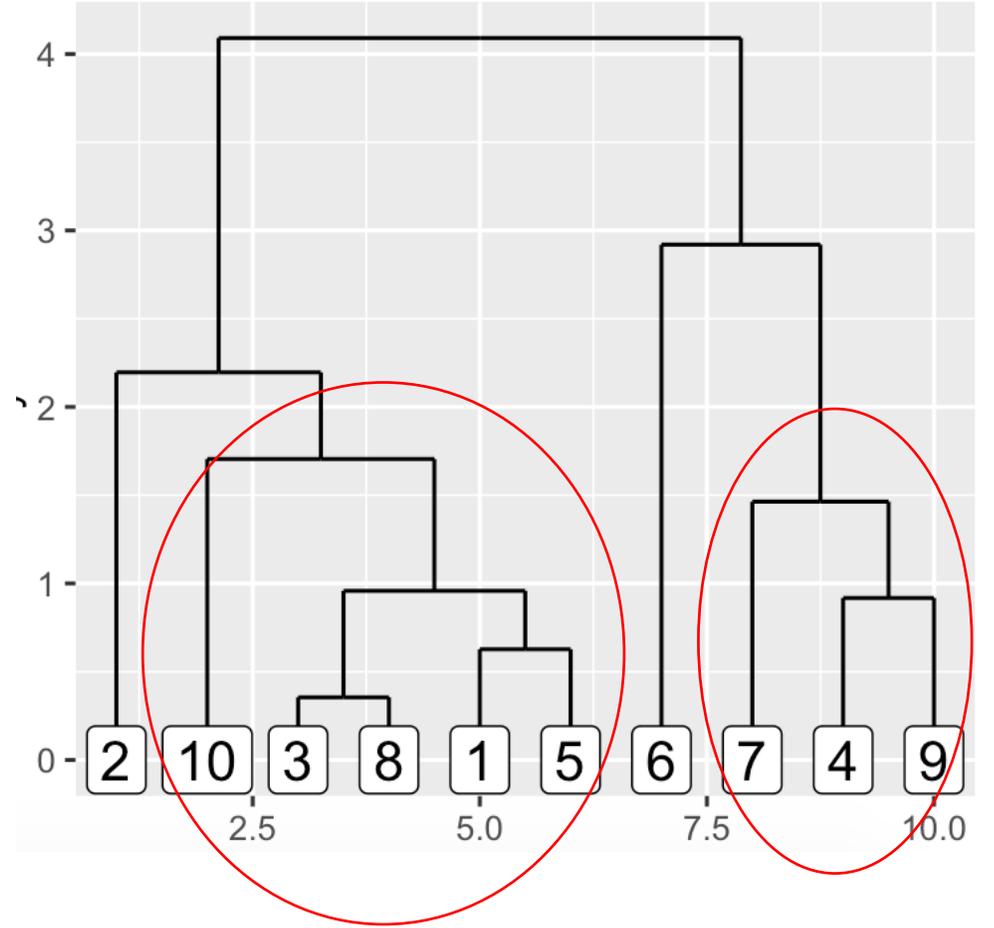
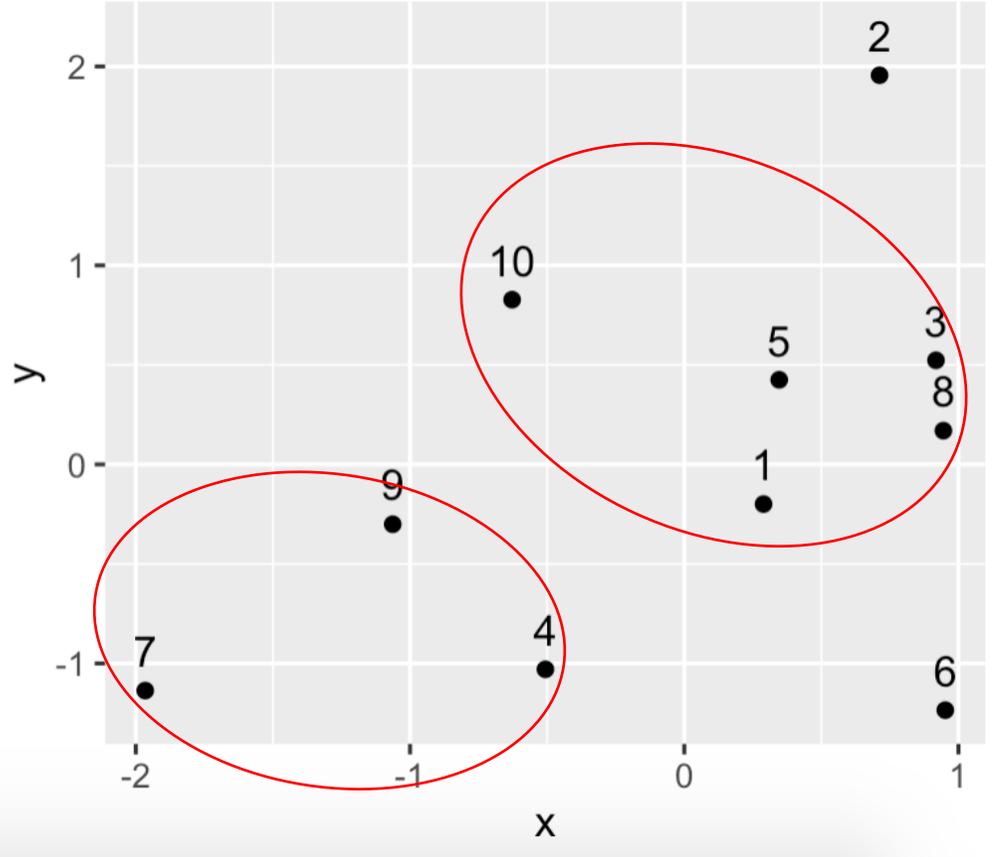


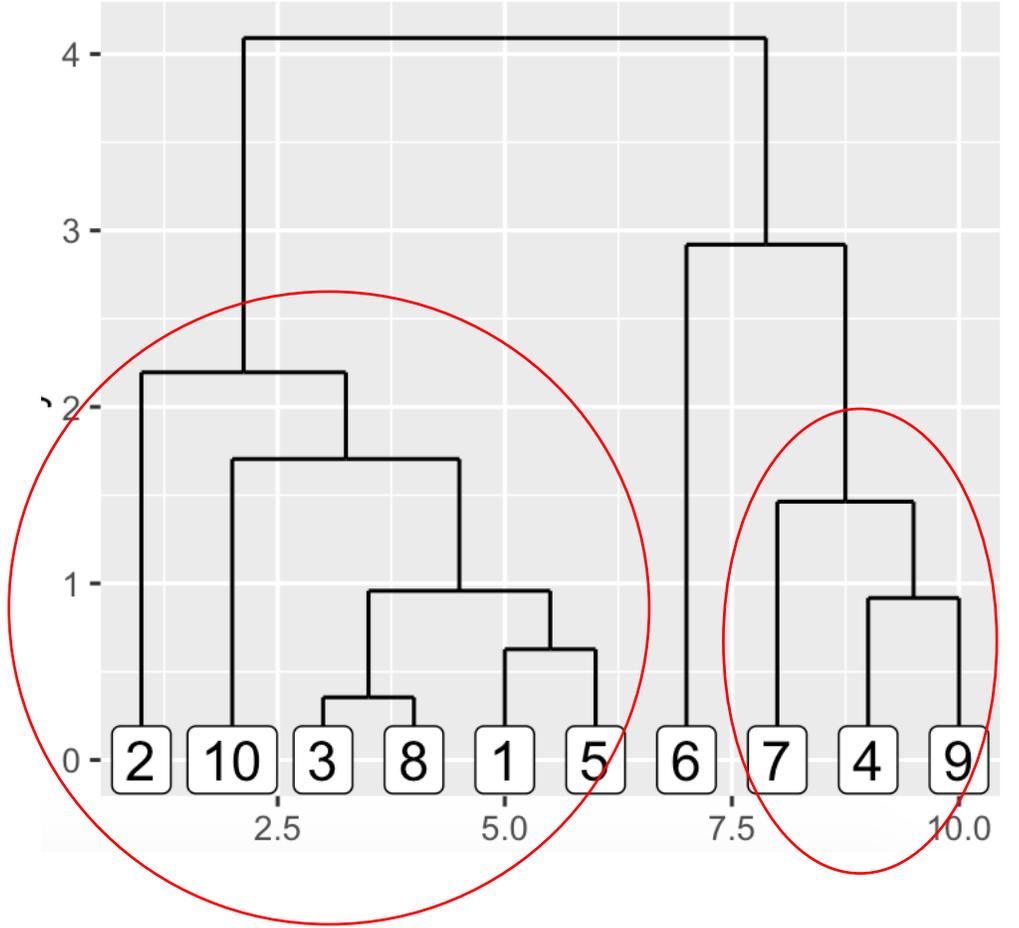
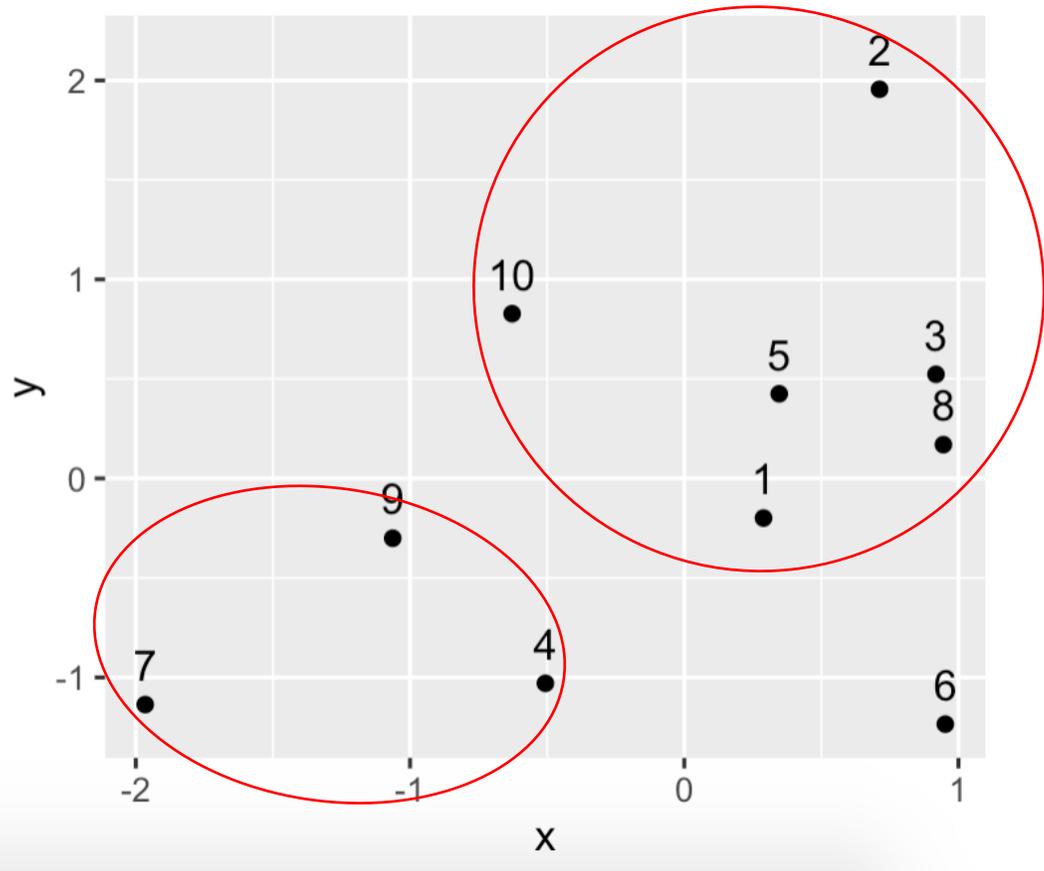


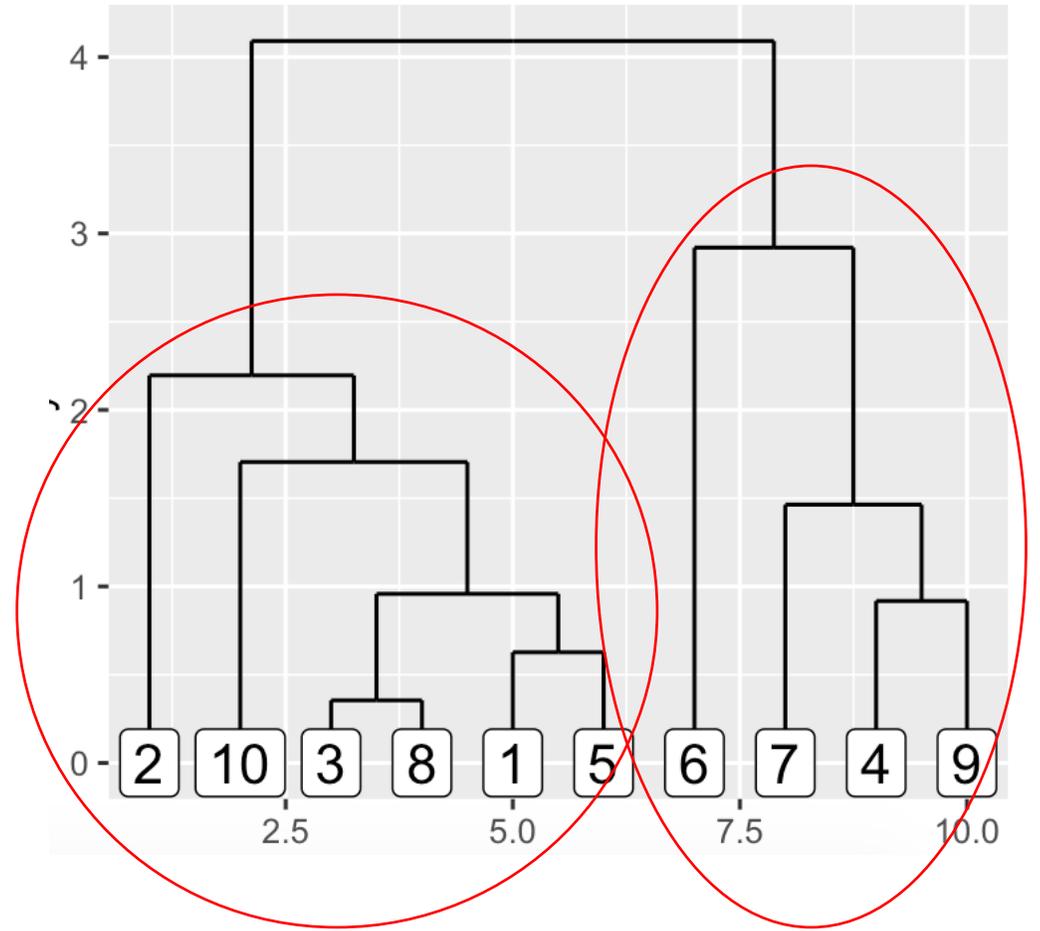
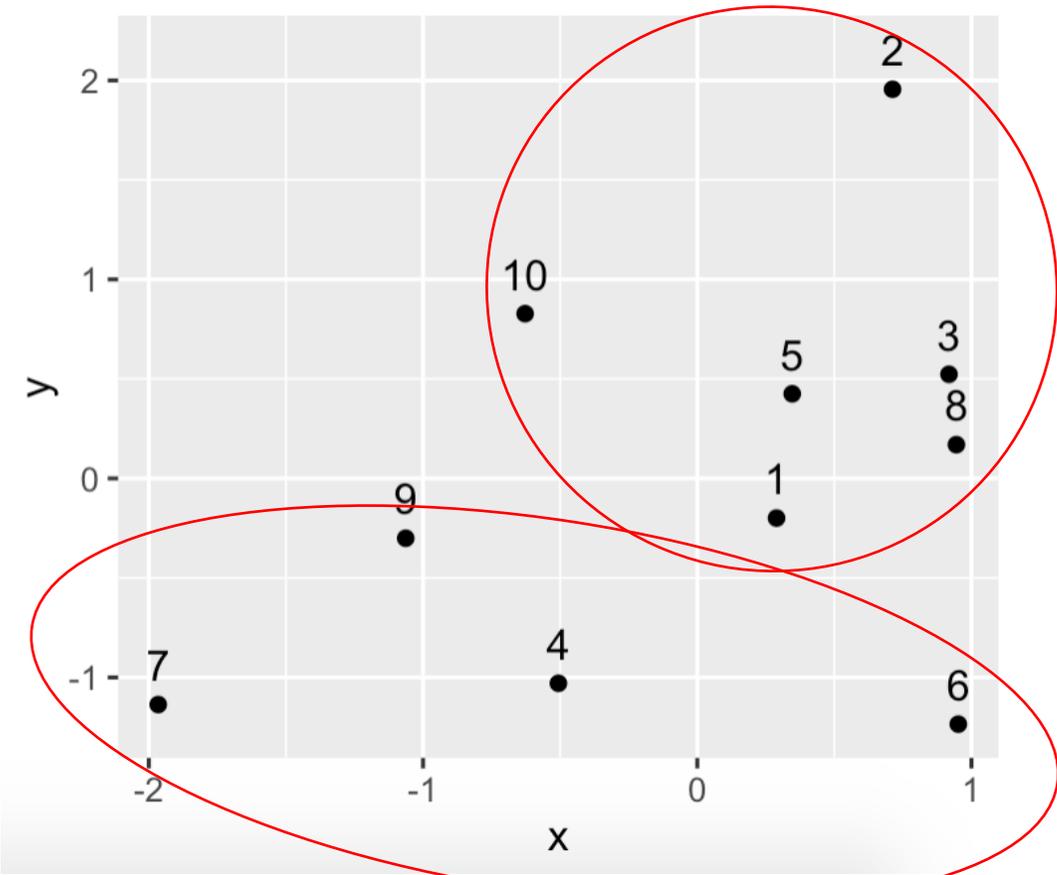


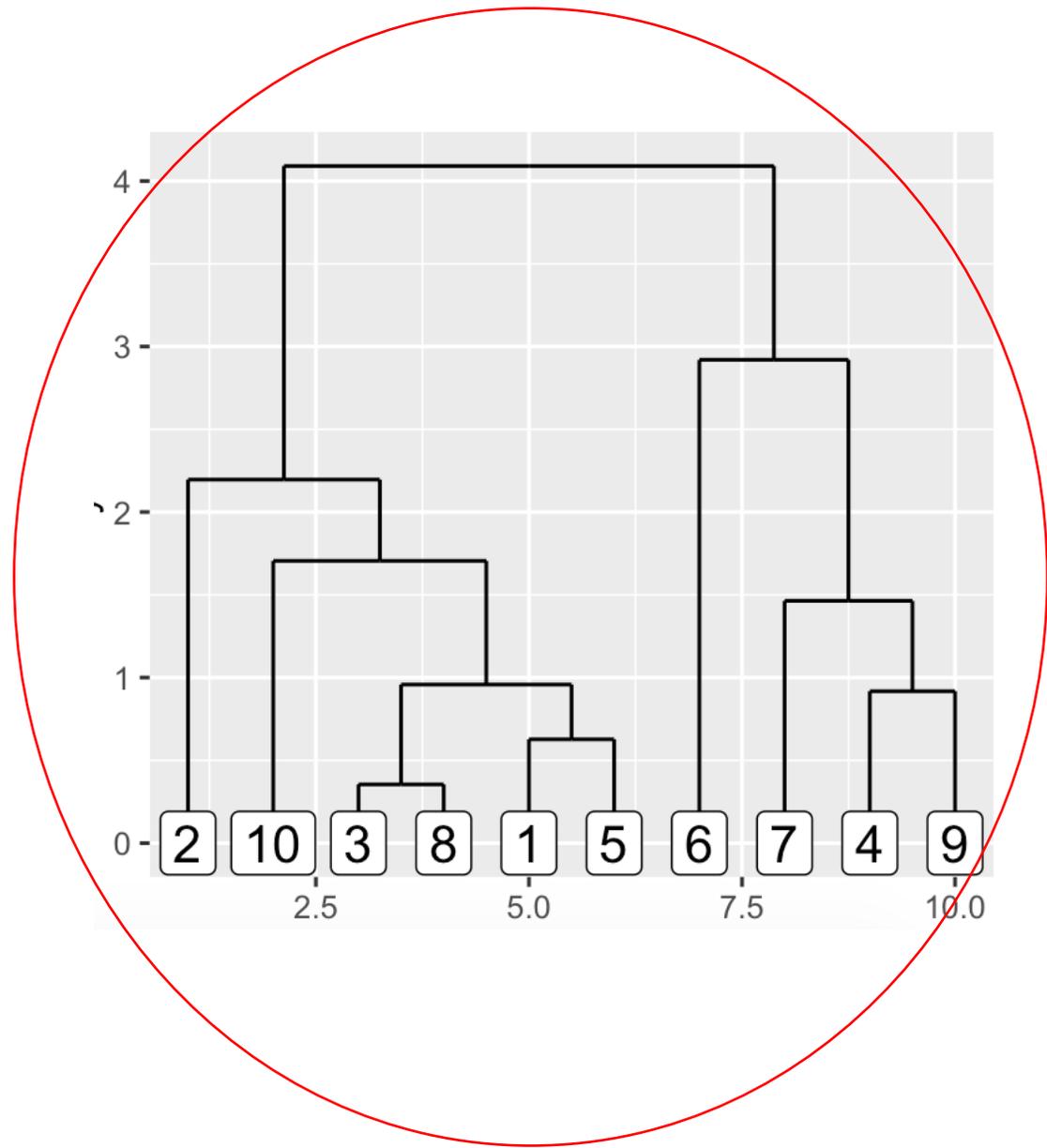
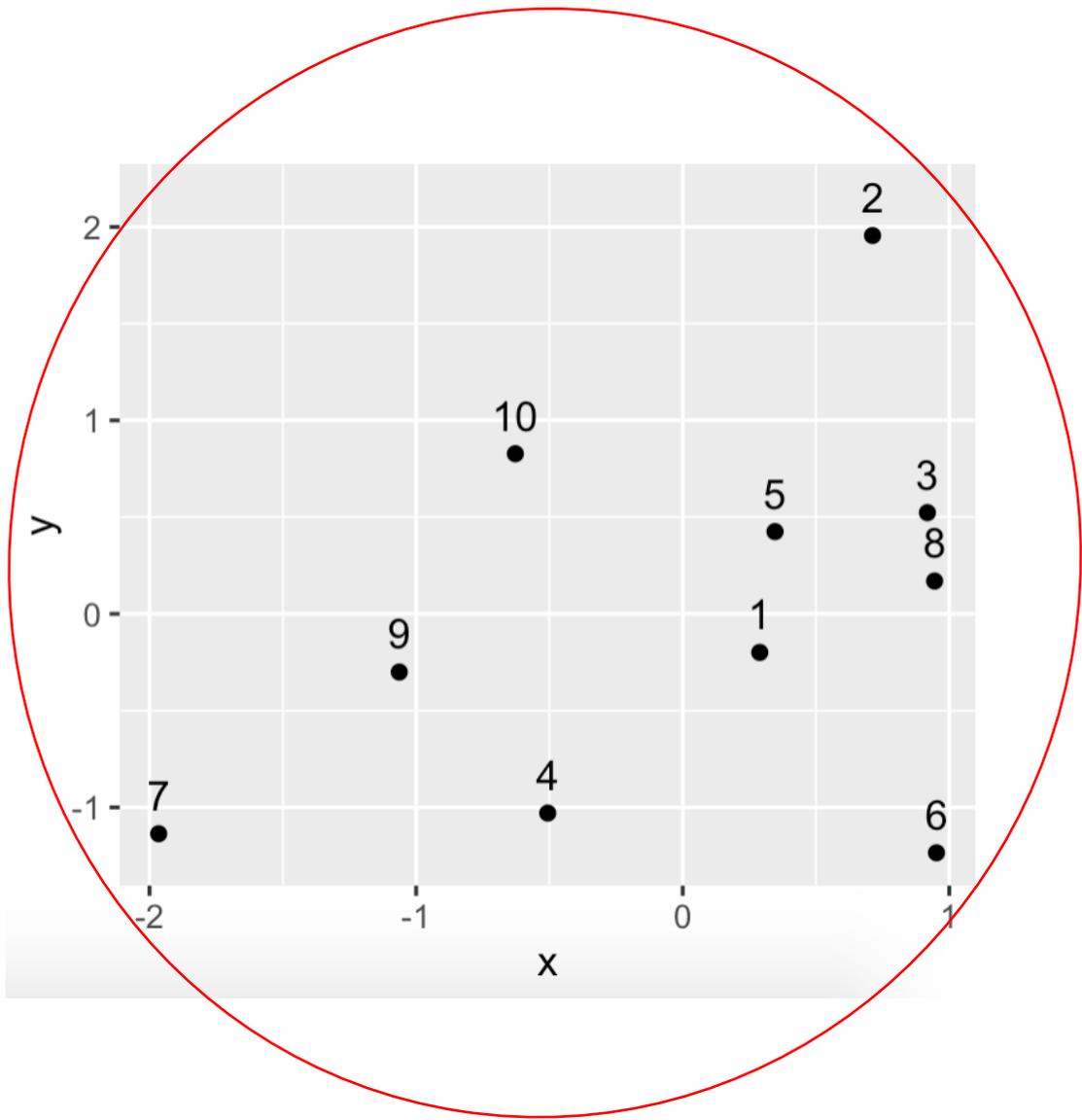


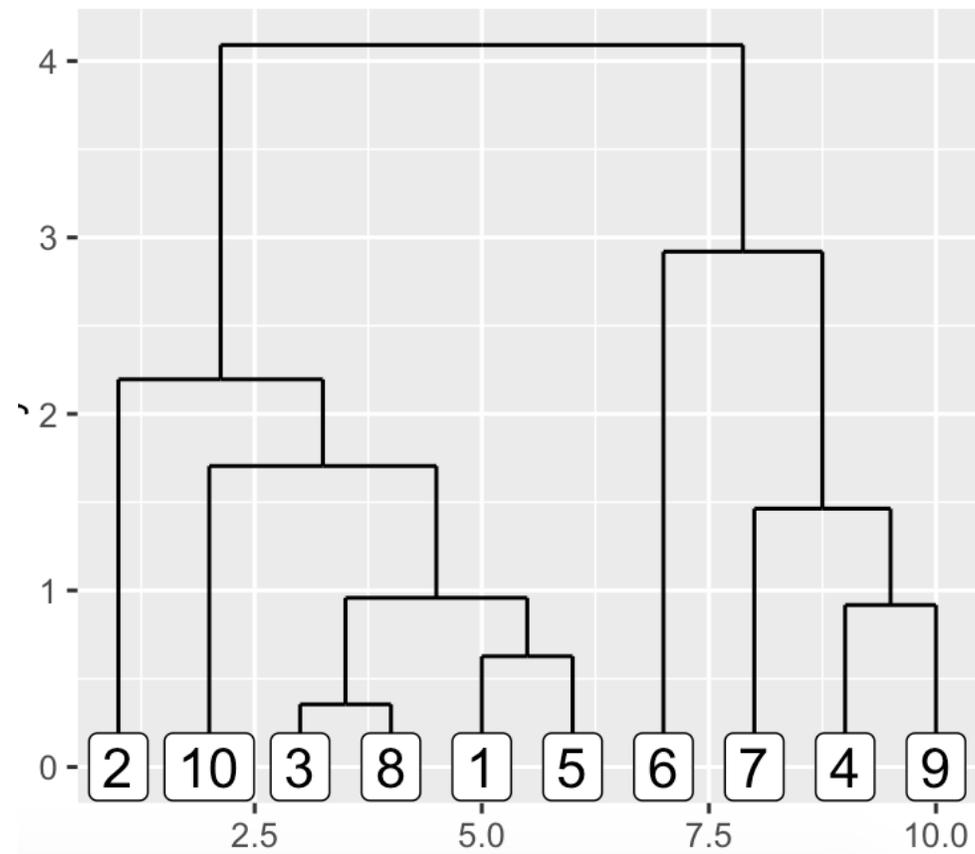
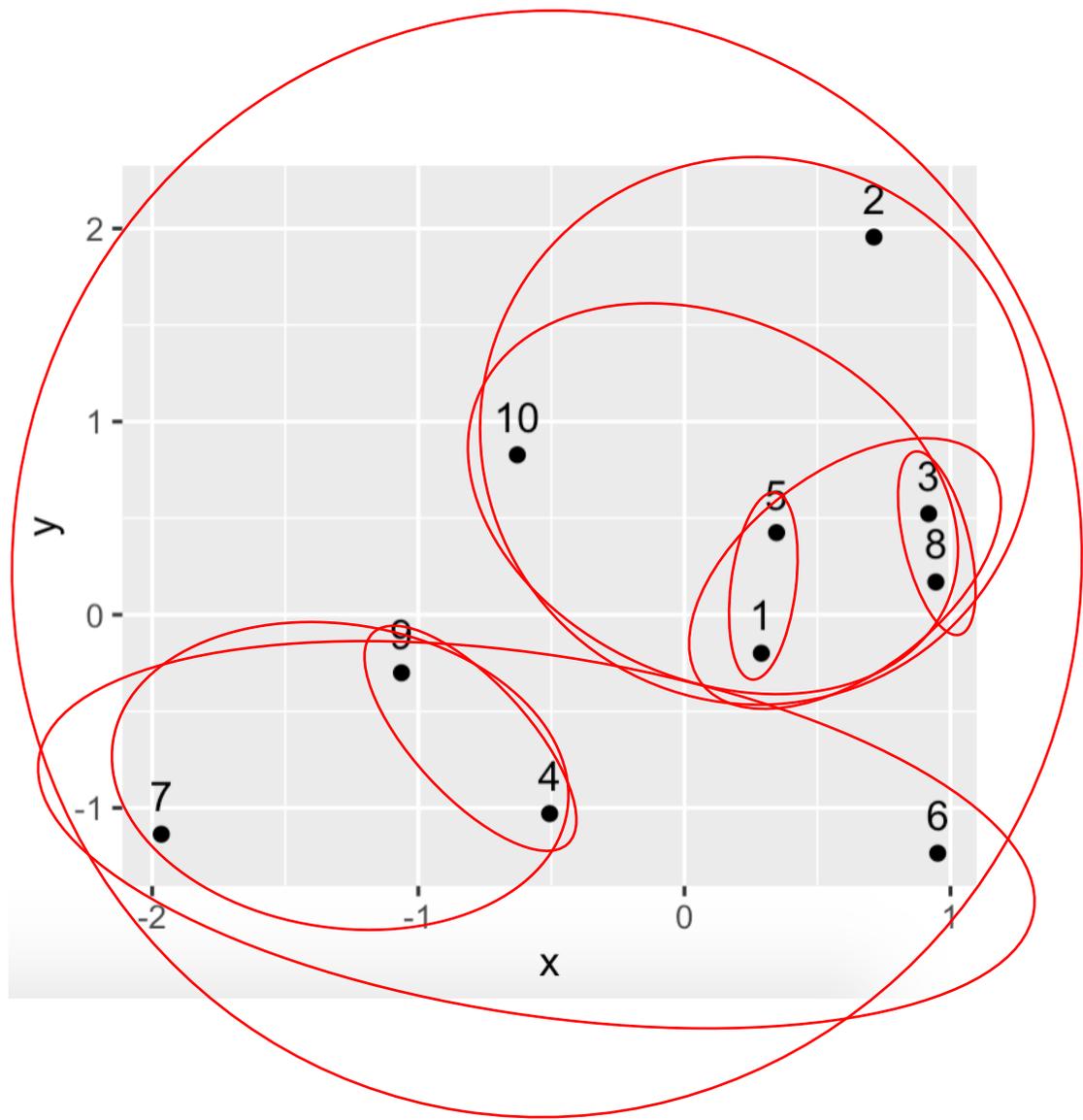


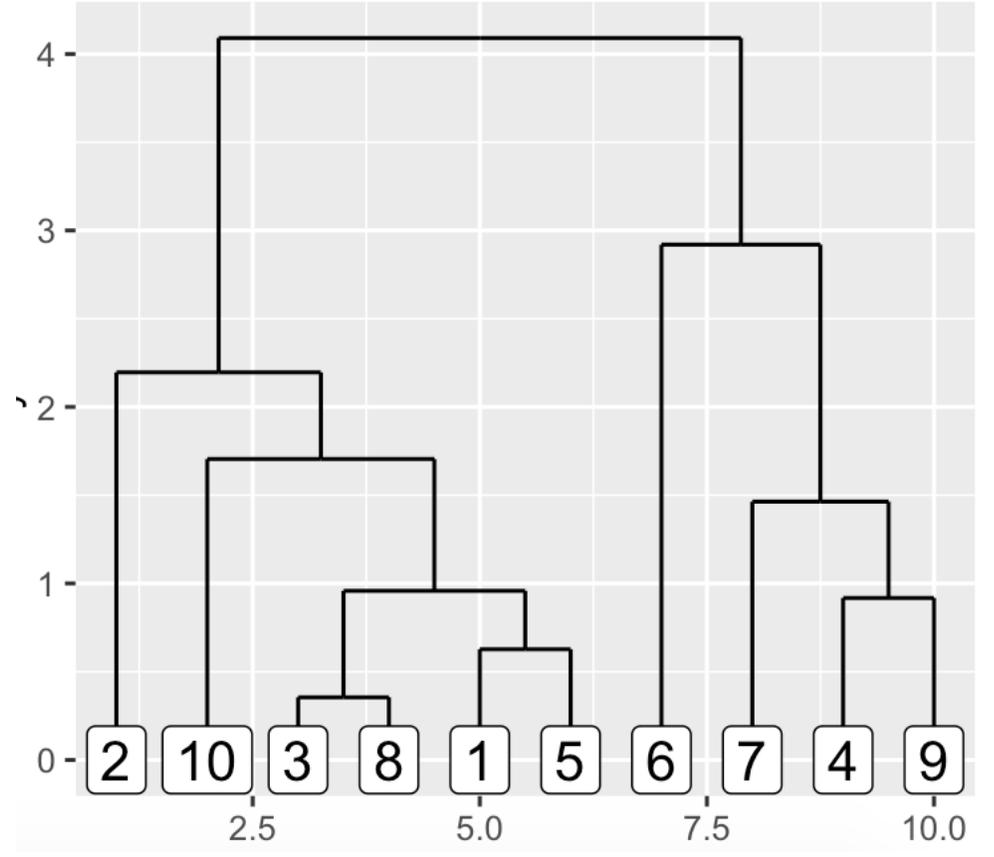
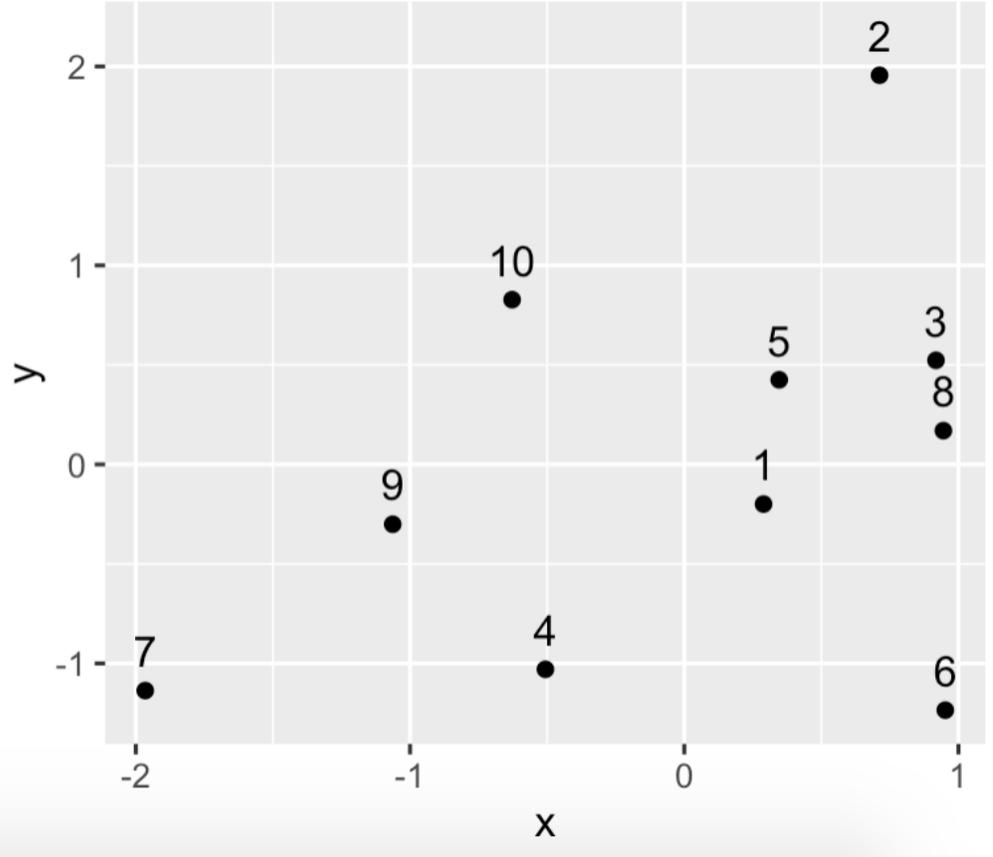




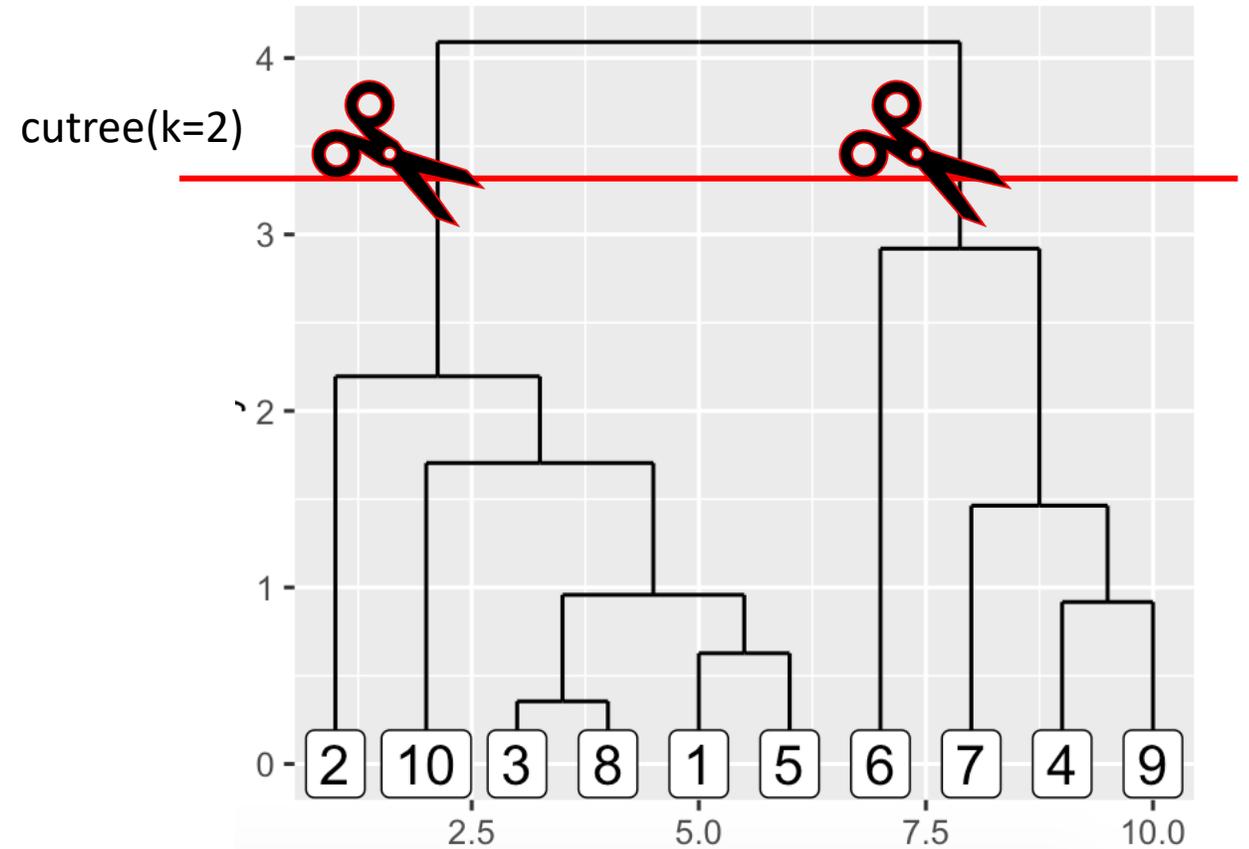
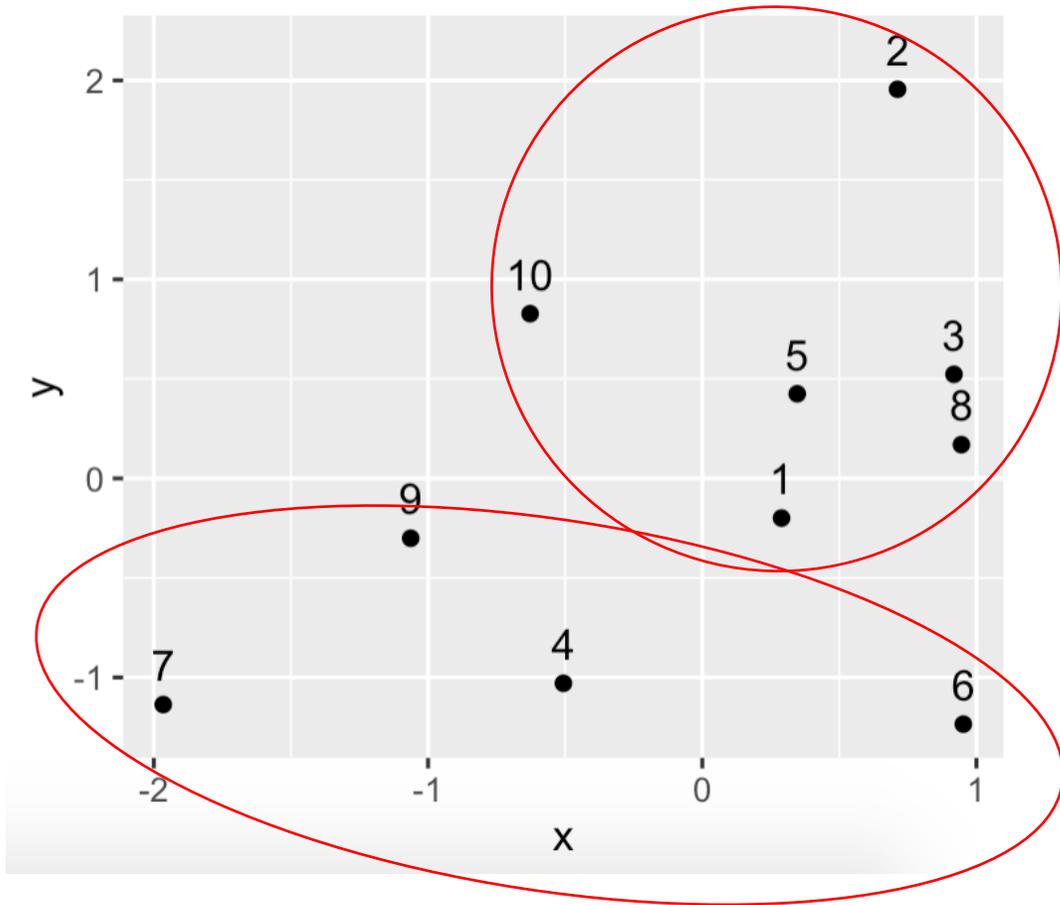




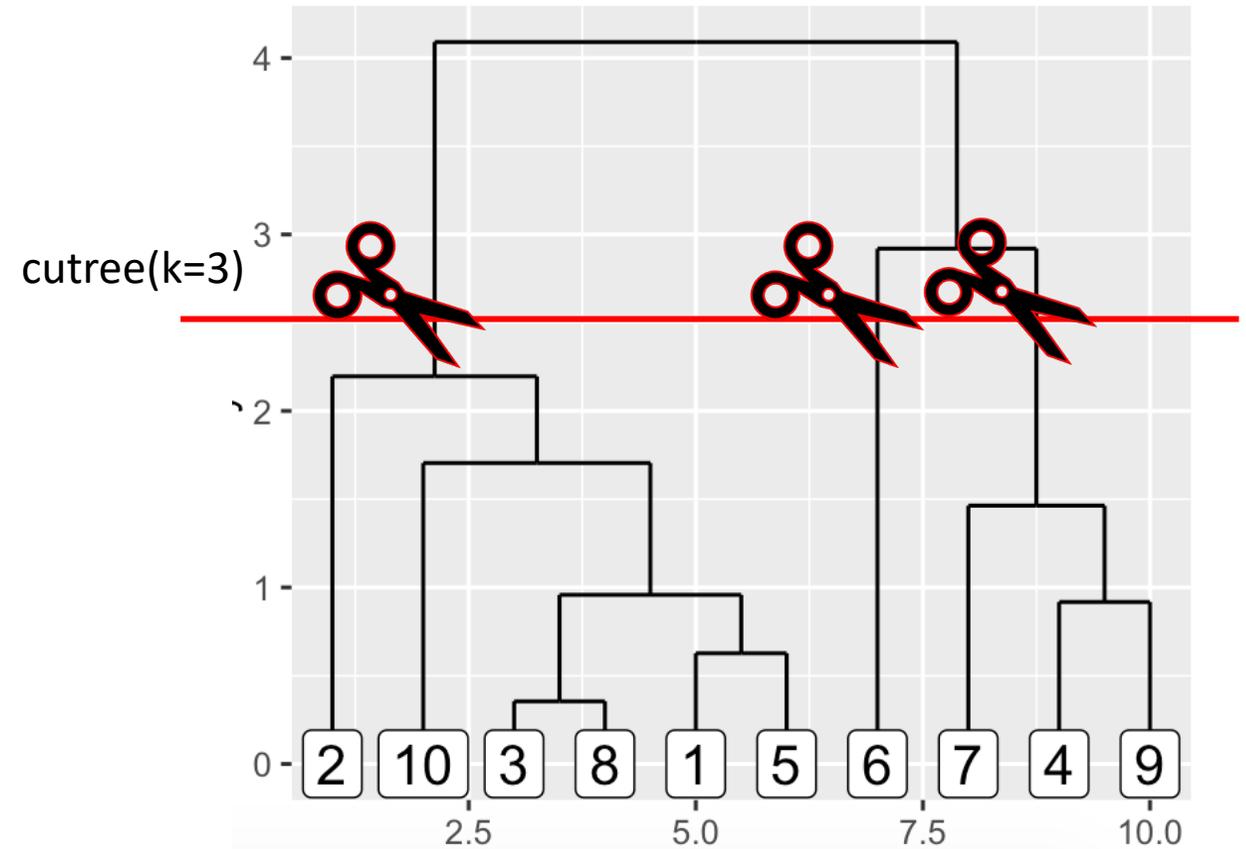
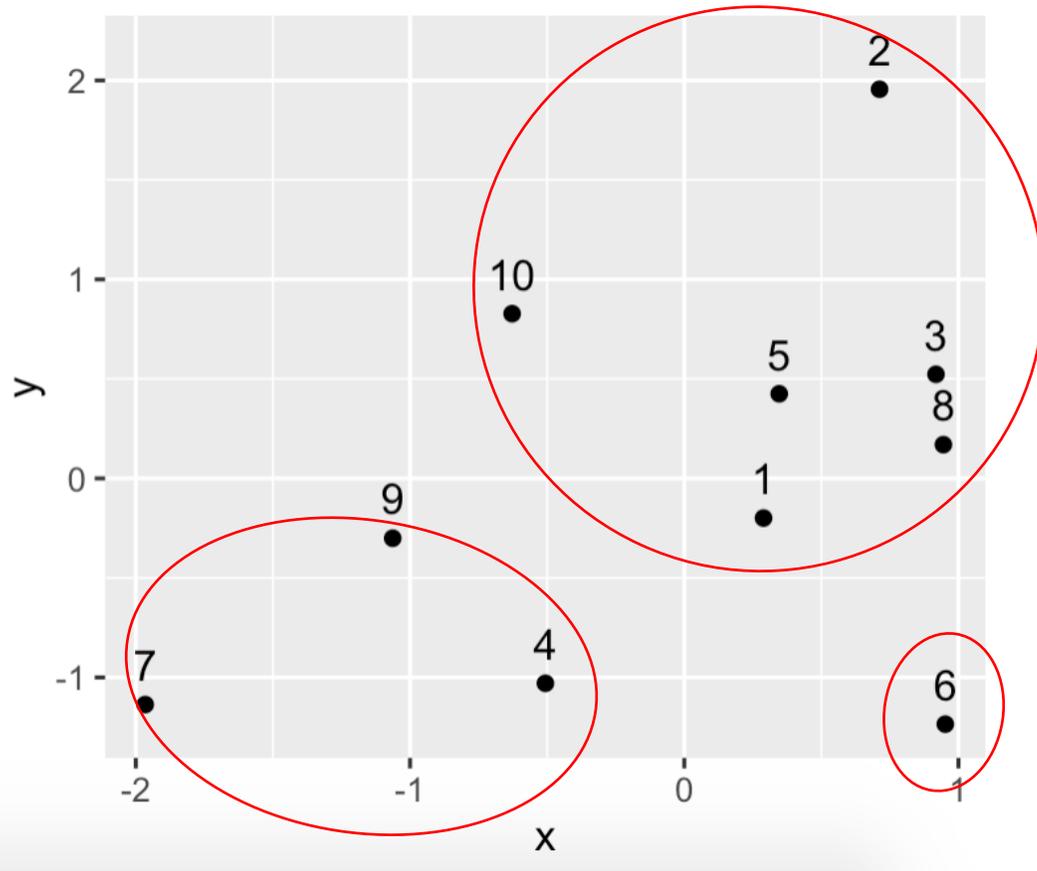




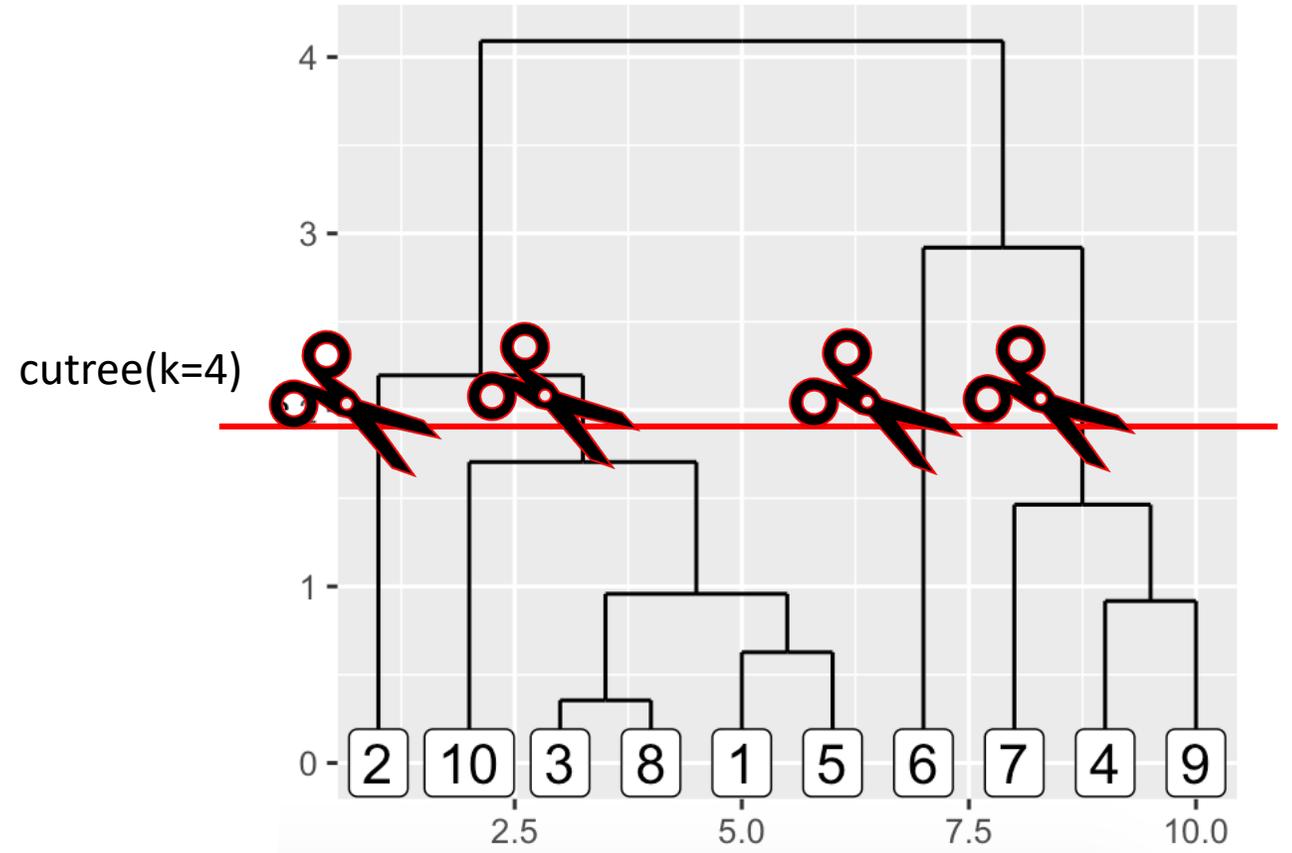
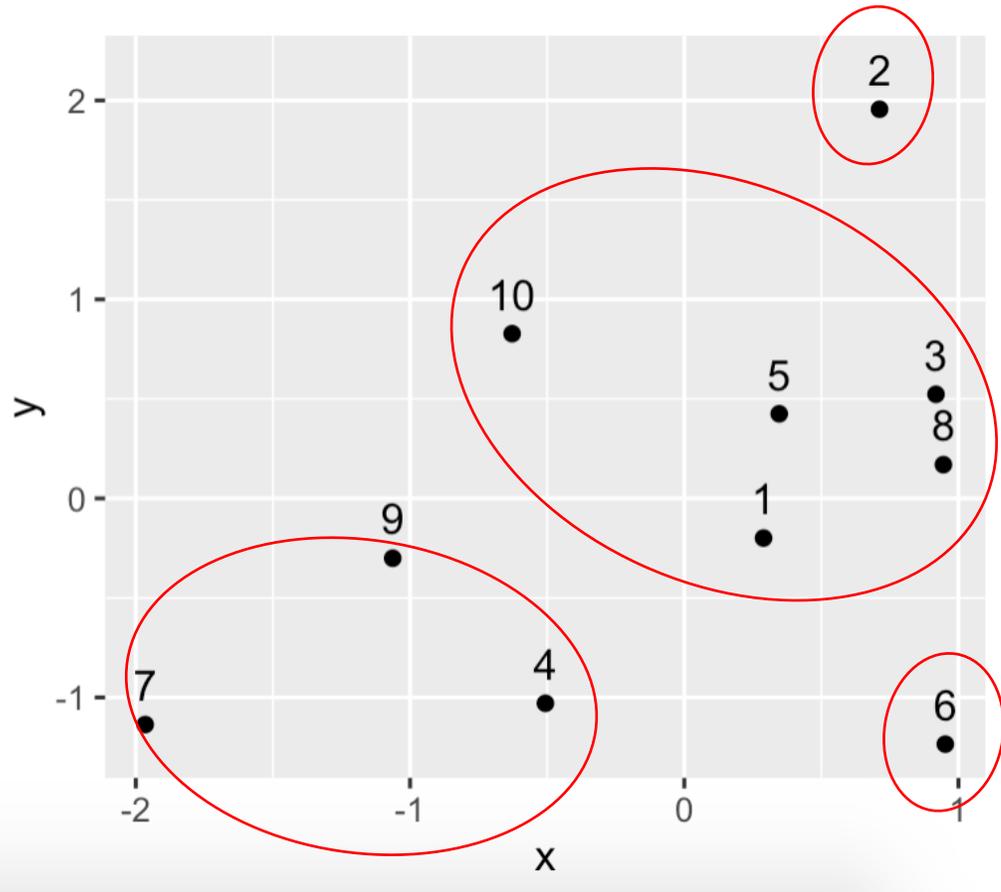
We can “cut” the tree and extract any number of clusters (k)

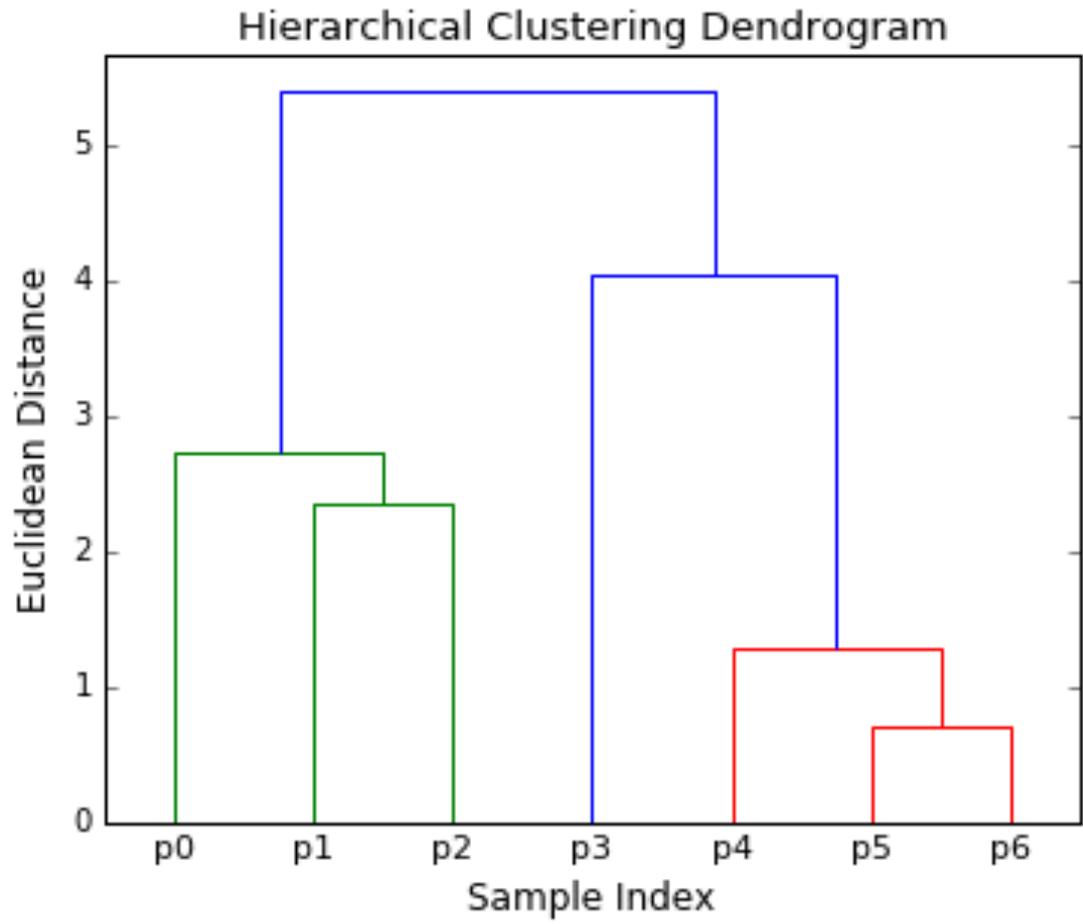
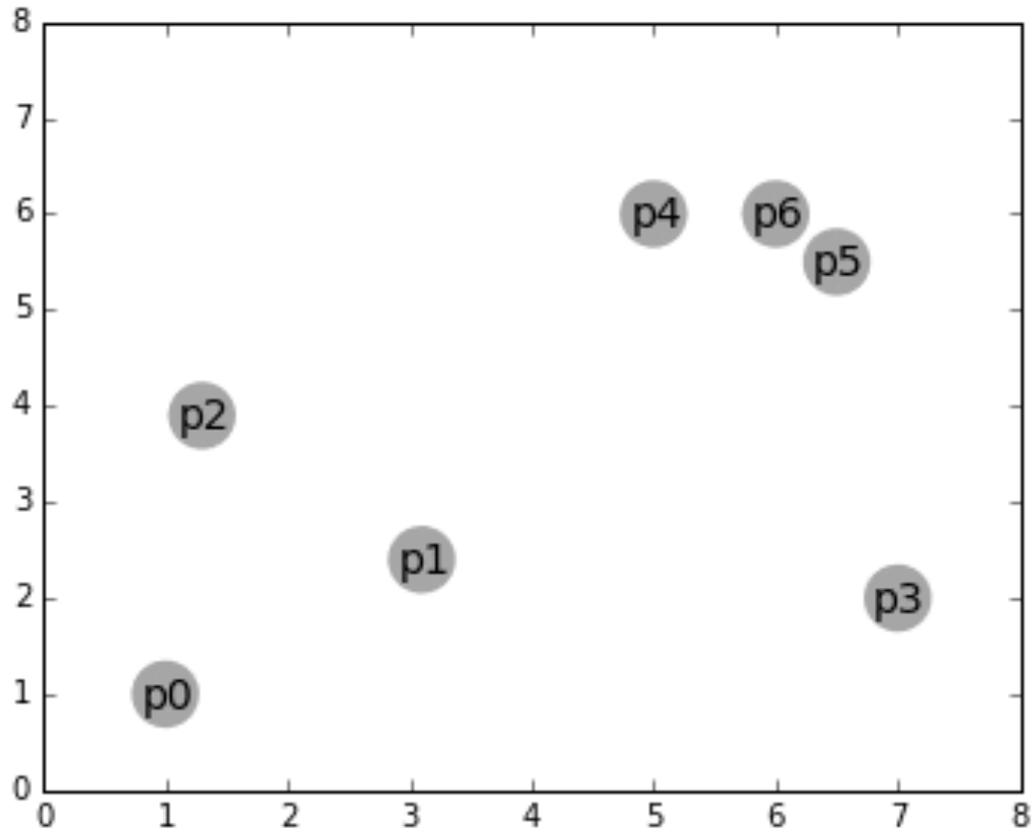


We can “cut” the tree and extract any number of clusters (k)

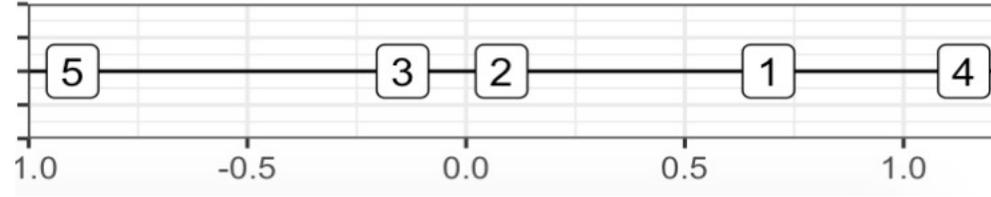


We can “cut” the tree and extract any number of clusters (k)



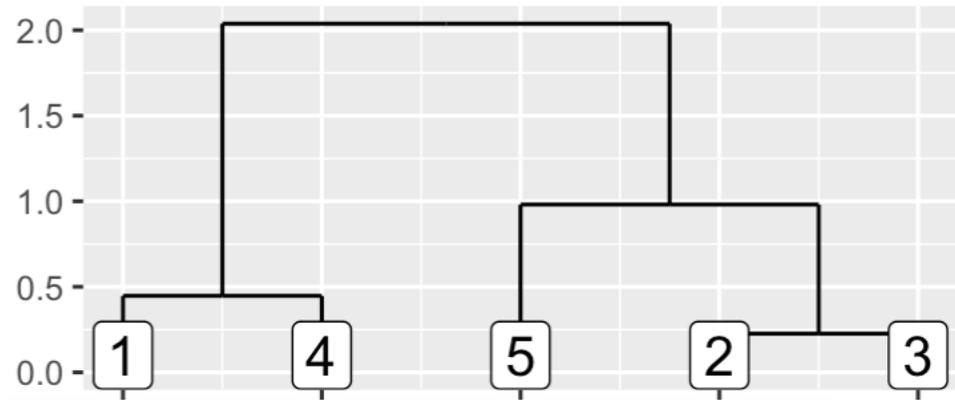


The default method `hclust()` merges clusters isn't always intuitive!

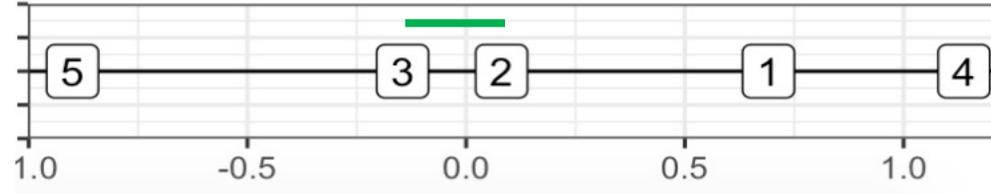


Distance matrix

```
      1      2      3      4
2 0.6100783
3 0.8363964 0.2263181
4 0.4471031 1.0571814 1.2834995
5 1.5908574 0.9807791 0.7544610 2.0379605
```

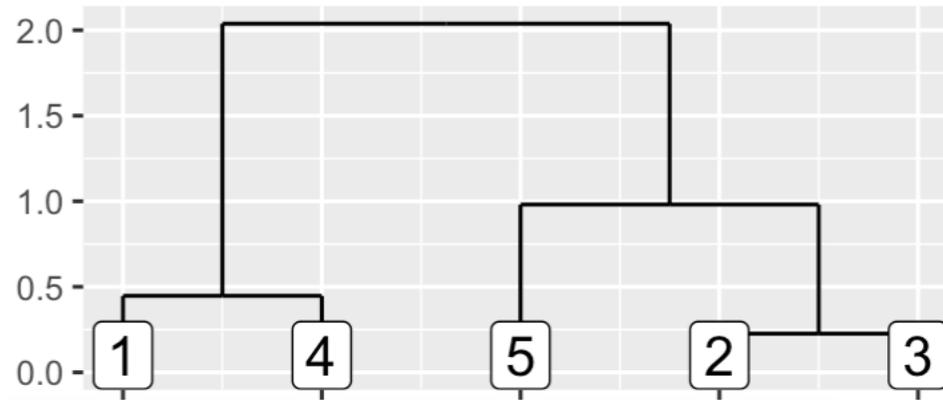


The default method `hclust()` merges clusters isn't always intuitive!

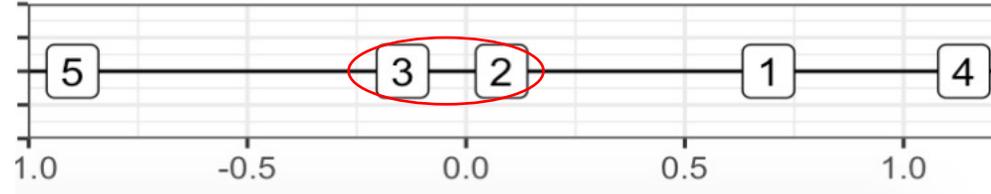


Distance matrix

	1	2	3	4
2	0.6100783			
3	0.8363964	<u>0.2263181</u>		
4	0.4471031	1.0571814	1.2834995	
5	1.5908574	0.9807791	0.7544610	2.0379605

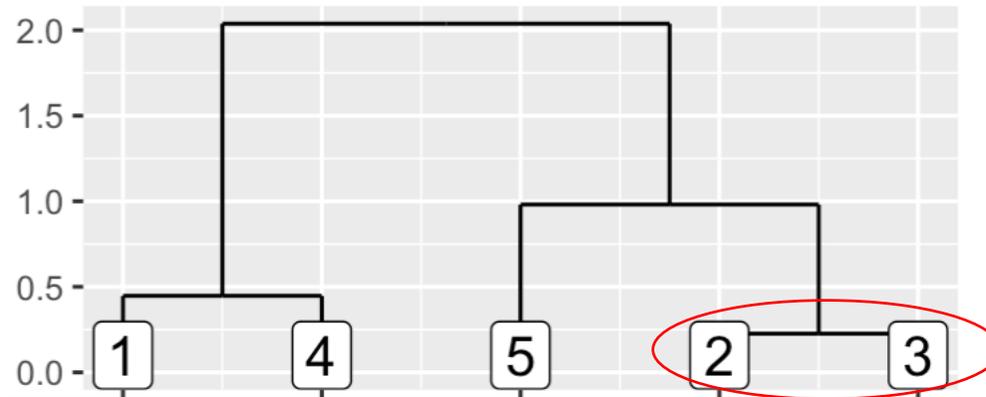


The default method `hclust()` merges clusters isn't always intuitive!

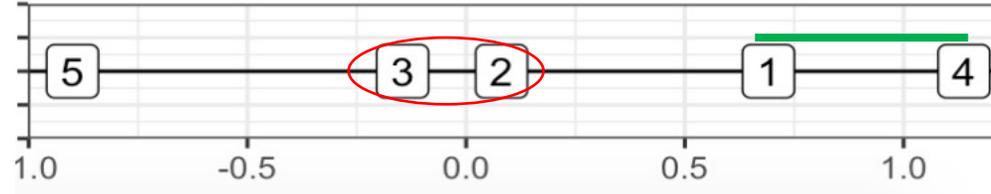


Distance matrix

	1	2	3	4
2	0.6100783			
3	0.8363964	0.2263181		
4	0.4471031	1.0571814	1.2834995	
5	1.5908574	0.9807791	0.7544610	2.0379605

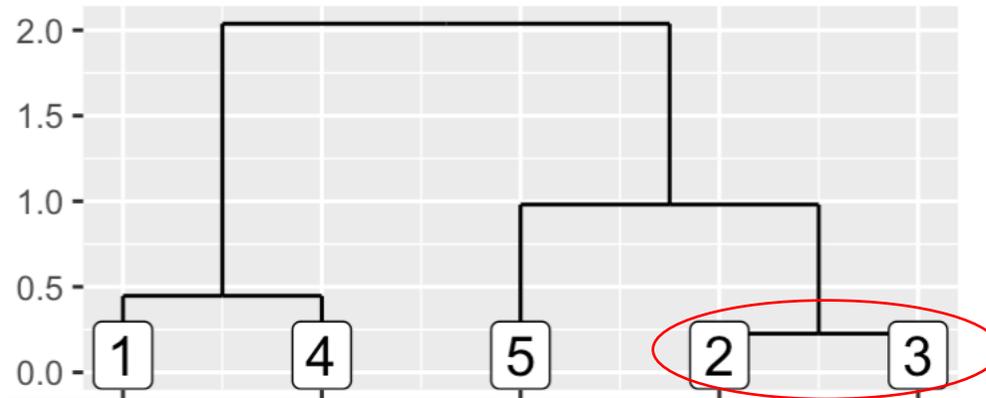


The default method `hclust()` merges clusters isn't always intuitive!

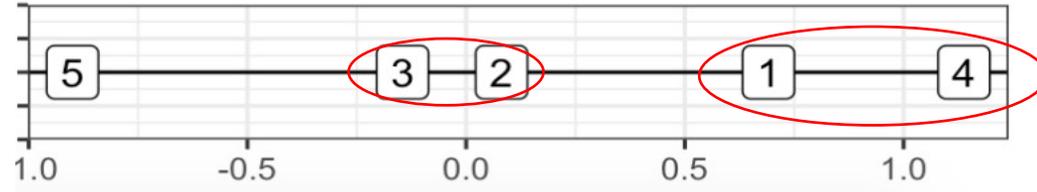


Distance matrix

	1	2	3	4
2	0.6100783			
3	0.8363964	0.2263181		
4	<u>0.4471031</u>	1.0571814	1.2834995	
5	1.5908574	0.9807791	0.7544610	2.0379605

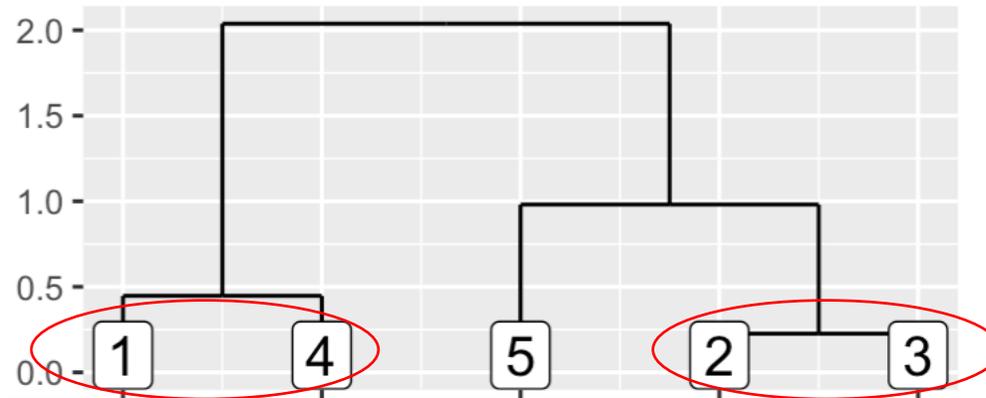


The default method `hclust()` merges clusters isn't always intuitive!

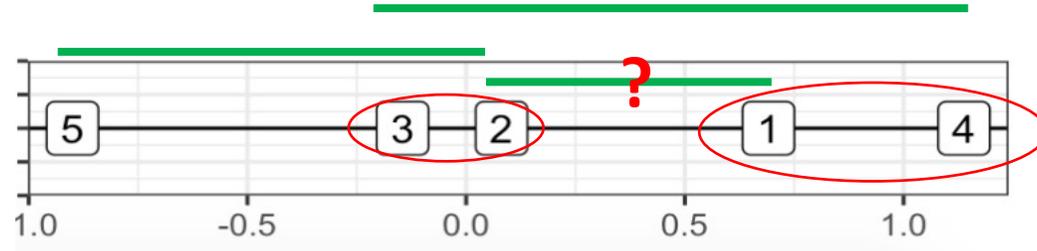


Distance matrix

```
      1      2      3      4
2 0.6100783
3 0.8363964 0.2263181
4 0.4471031 1.0571814 1.2834995
5 1.5908574 0.9807791 0.7544610 2.0379605
```

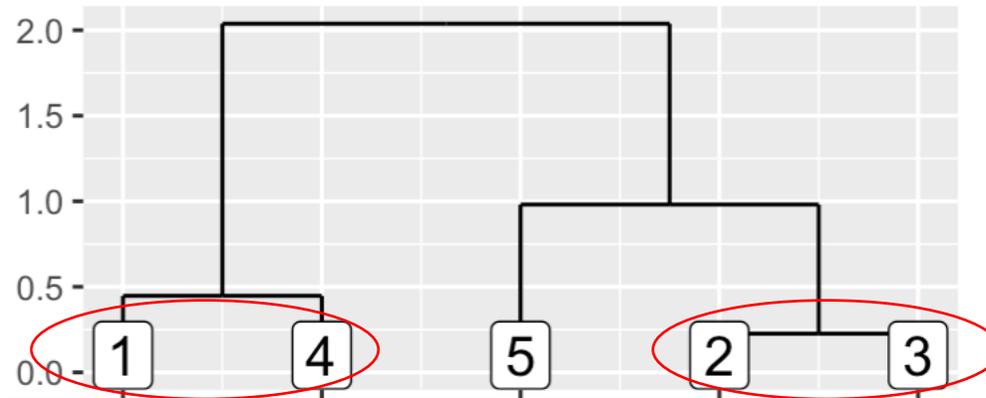


The default method `hclust()` merges clusters isn't always intuitive!



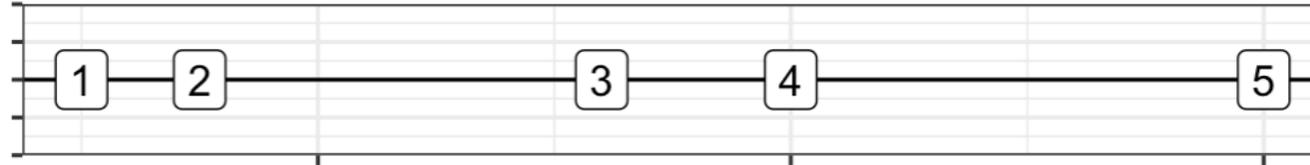
Distance matrix

	1	2	3	4
2	<u>0.6100783</u>			
3	0.8363964	0.2263181		
4	0.4471031	1.0571814	<u>1.2834995</u>	
5	1.5908574	<u>0.9807791</u>	0.7544610	2.0379605



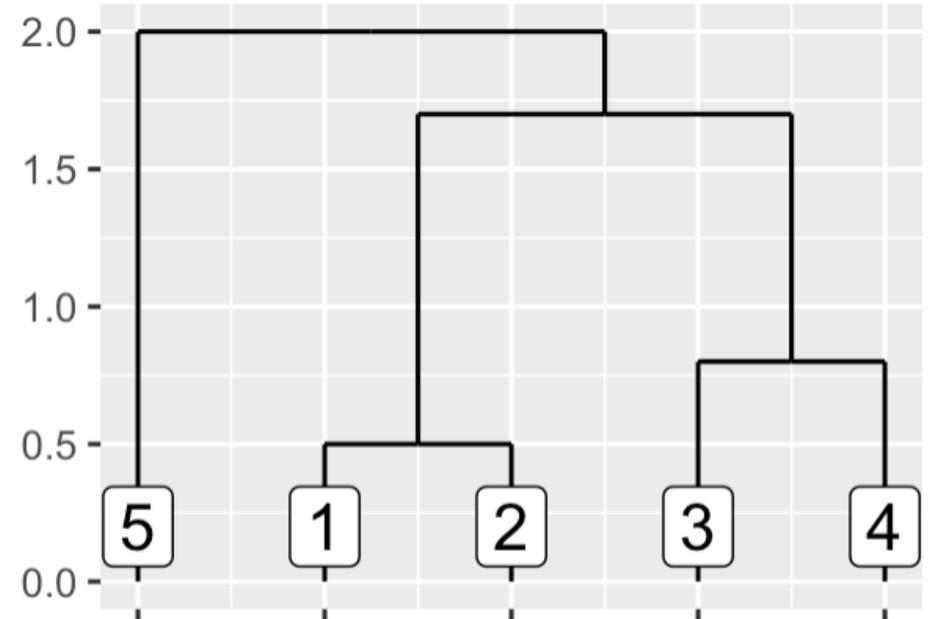
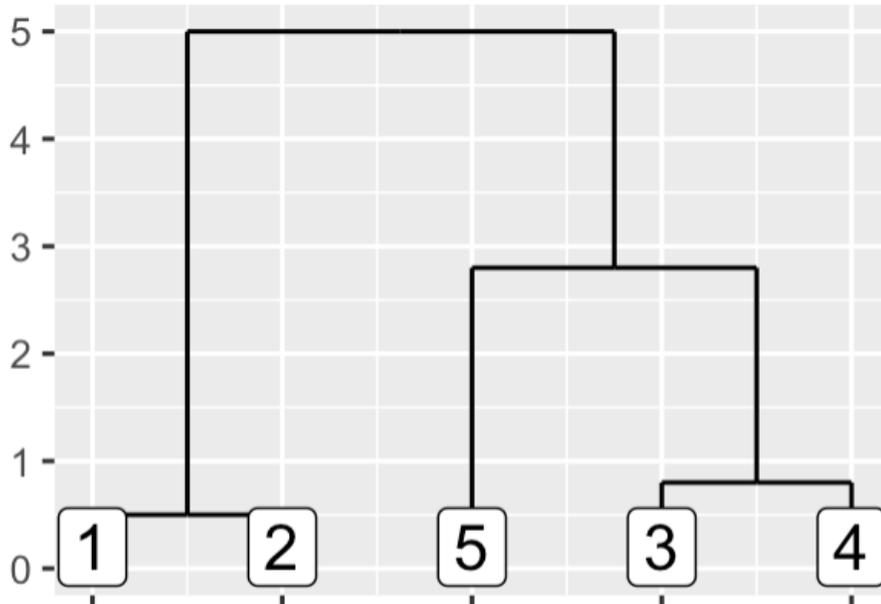
Why doesn't cluster 2/3 get merged with cluster 1/4 since the distance between 1 and 2 is smaller than 5 and 3?

`hclust()` in R uses "complete" linkage as default method for building clusters. This means it looks at the biggest difference between clusters to determine when they get merged. In this example, point 5 gets merged with cluster 2/3 because it is closer to pt 2 than the distance between point 3 and point 4, the furthest points in their respective clusters.



“complete” linkage
 Default of `hclust(method=)`

“single” linkage
 Uses closest points between clusters to merge



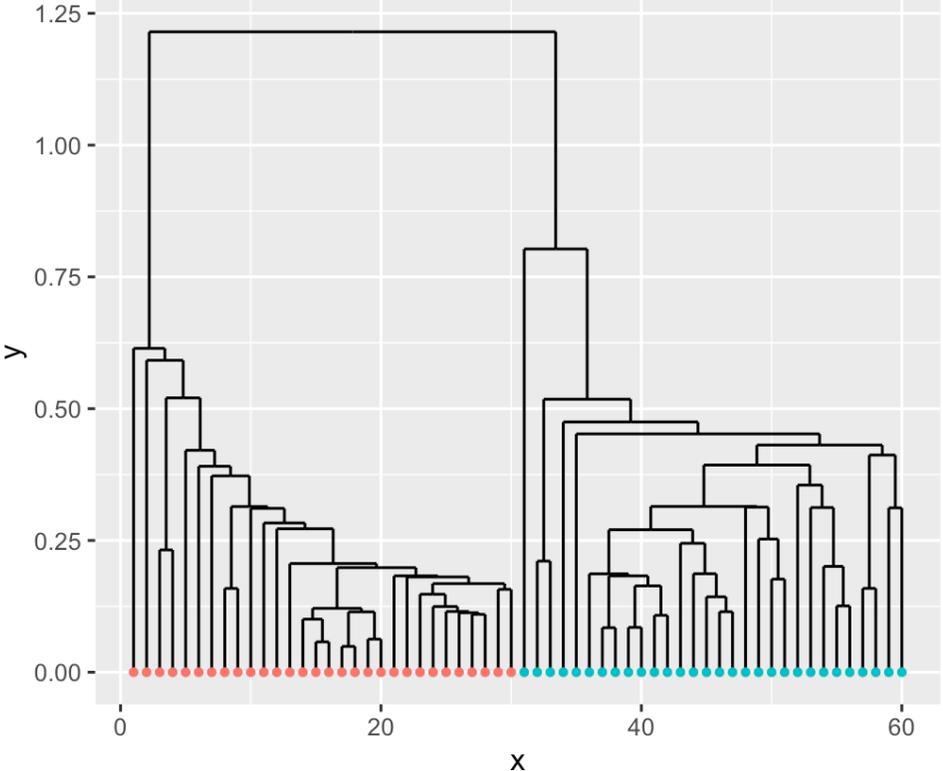
With “complete” linkage, cluster 3/4 merges with point 5 because the distance between 1 and 4 > distance between 4 and 5

With “single” linkage, cluster 1/2 merges with 3/4 because the distance between 3 and 4 < distance between 4 and 5

Different hierarchical clustering methods yield distinct tree topologies

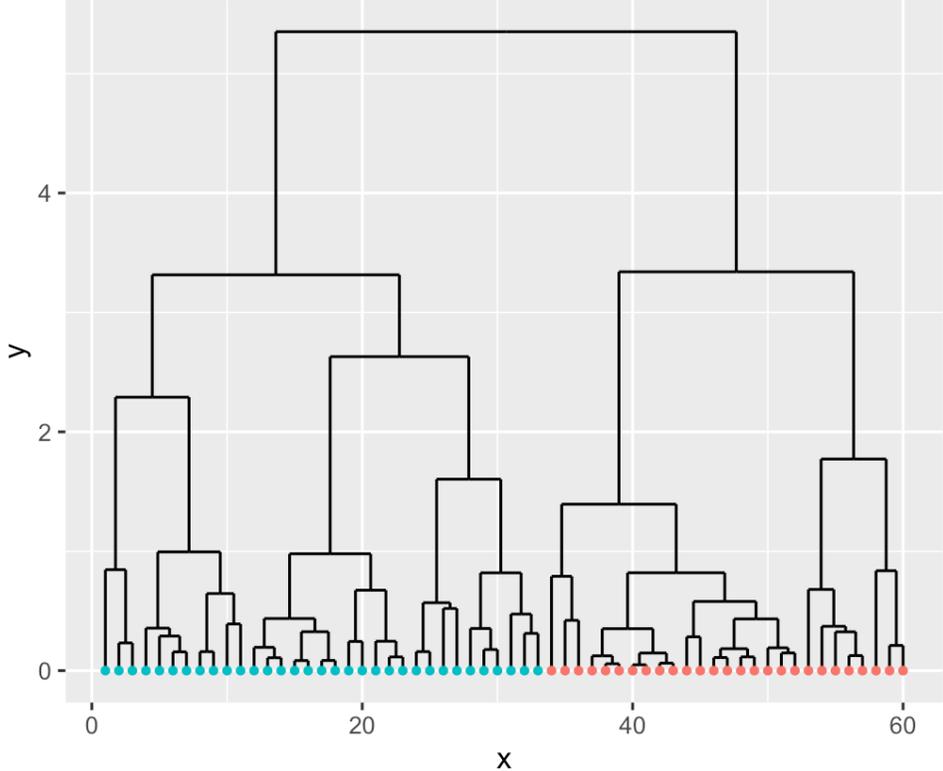
“Single”

Proximity between two clusters is the proximity between their two closest objects. Metaphor is a chain.



“Complete”

Proximity between two clusters is the proximity between their two most distant objects. Metaphor is a circle.

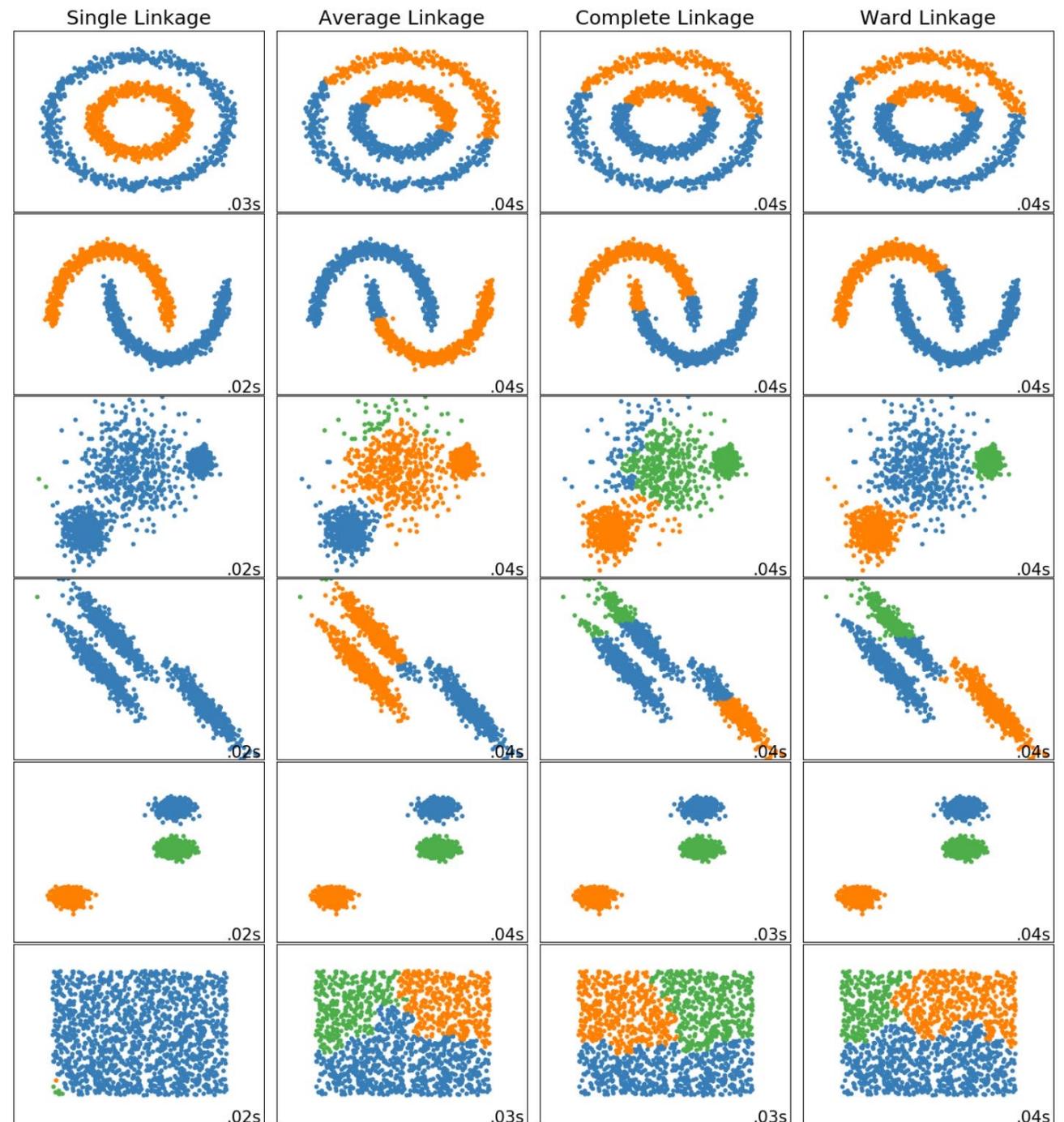


“Single” Proximity between two clusters is the proximity between their two closest objects.

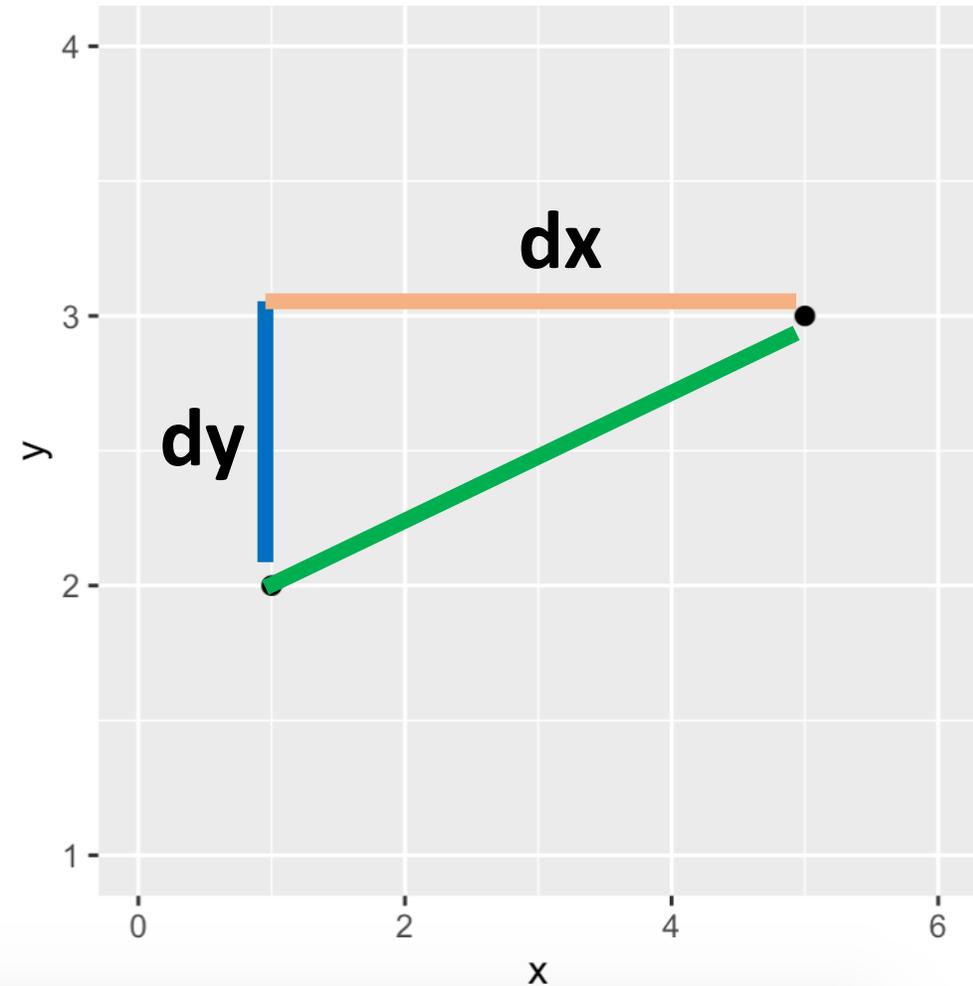
“Average” Proximity between two clusters is the arithmetic mean of all the proximities in their joint cluster.

“Complete” Proximity between two clusters is the proximity between their two most distant objects.

“Ward” Proximity between two clusters is the magnitude by which the summed square in their joint cluster will be greater than the combined summed square in these two clusters.



Measuring distance



Euclidean distance = $\sqrt{dx^2 + dy^2}$

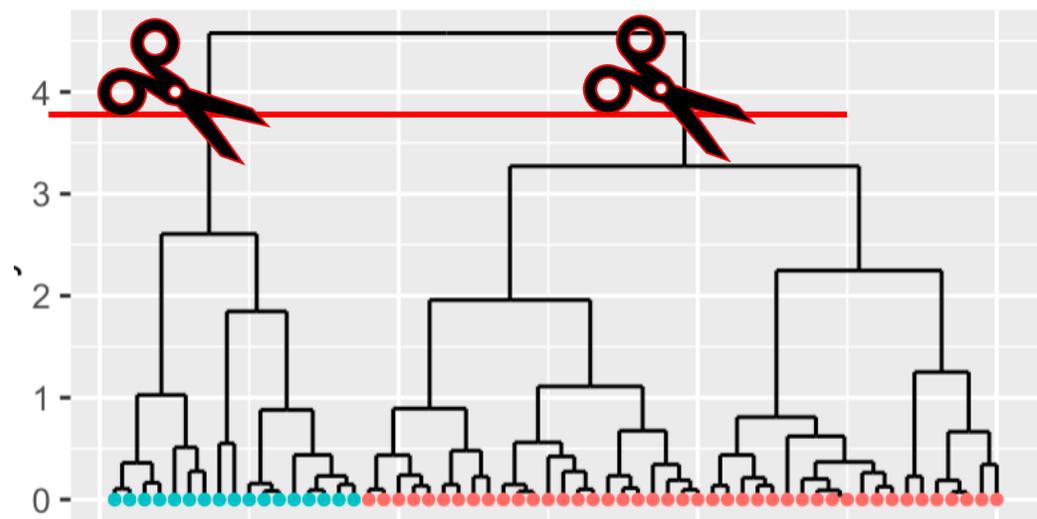


Manhattan distance = $\text{abs}(dx) + \text{abs}(dy)$



“euclidean”

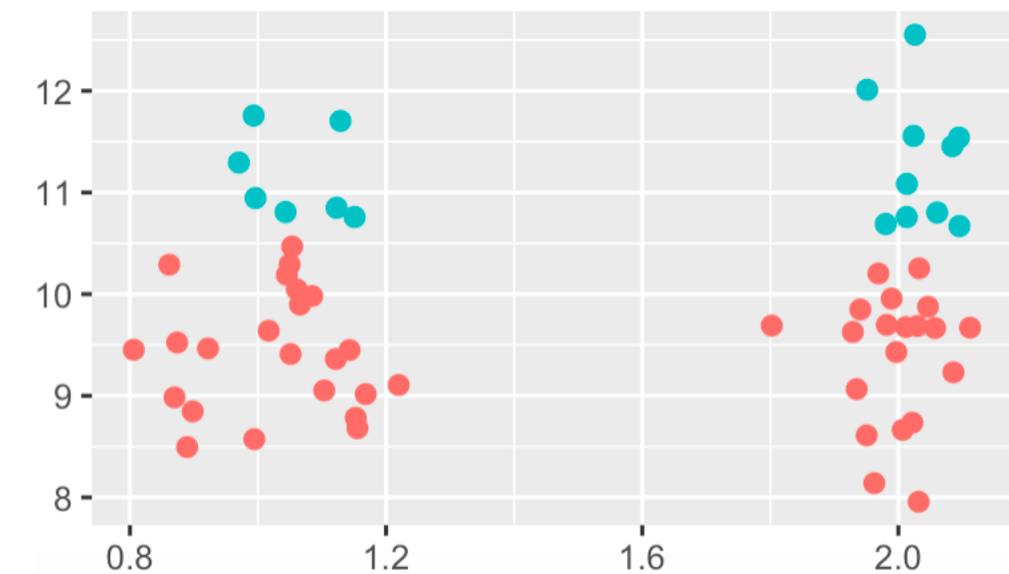
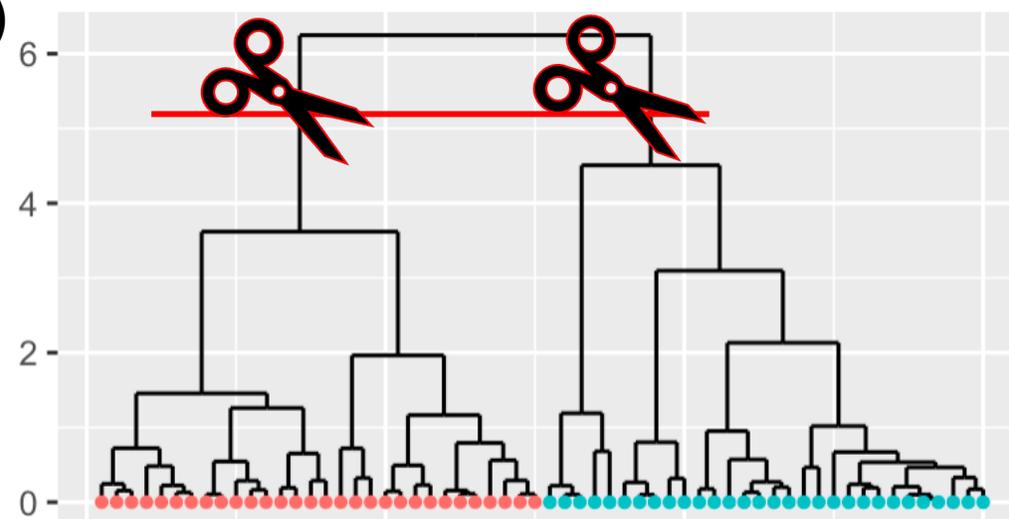
Default for `dist(method=)` in R



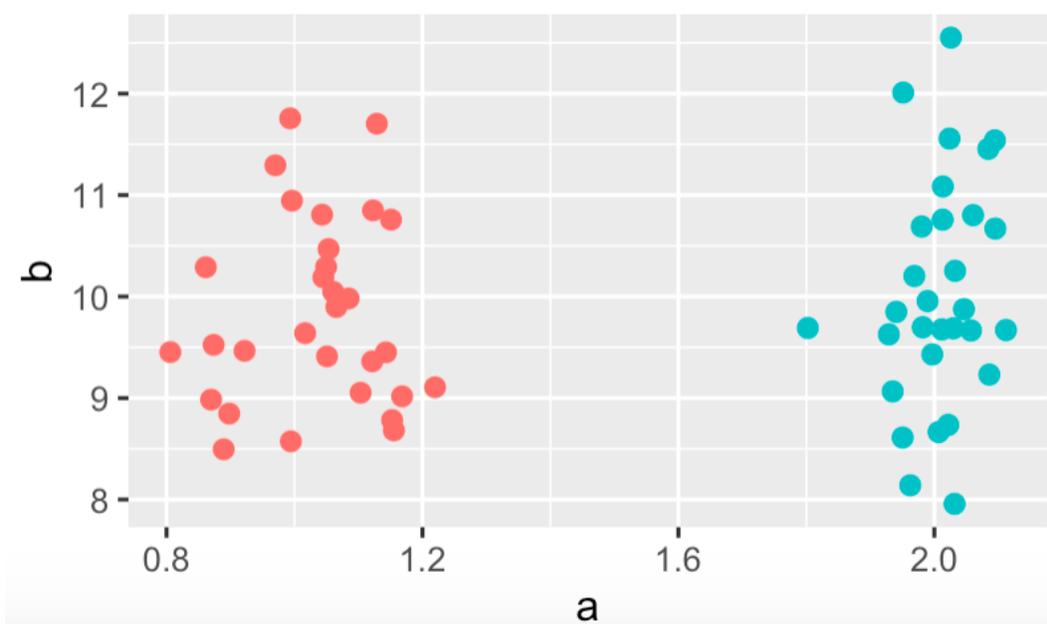
`cutree(k=2)`

“mahnattan”

Distance = sum of absolute values of differences between all variables



hlust_2



hlust_2



BIO1 = Annual Mean Temperature

BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))

BIO3 = Isothermality (BIO2/BIO7) (×100)

BIO4 = Temperature Seasonality (standard deviation ×100)

BIO5 = Max Temperature of Warmest Month

BIO6 = Min Temperature of Coldest Month

BIO7 = Temperature Annual Range (BIO5-BIO6)

BIO8 = Mean Temperature of Wettest Quarter

BIO9 = Mean Temperature of Driest Quarter

BIO10 = Mean Temperature of Warmest Quarter

BIO11 = Mean Temperature of Coldest Quarter

BIO12 = Annual Precipitation

BIO13 = Precipitation of Wettest Month

BIO14 = Precipitation of Driest Month

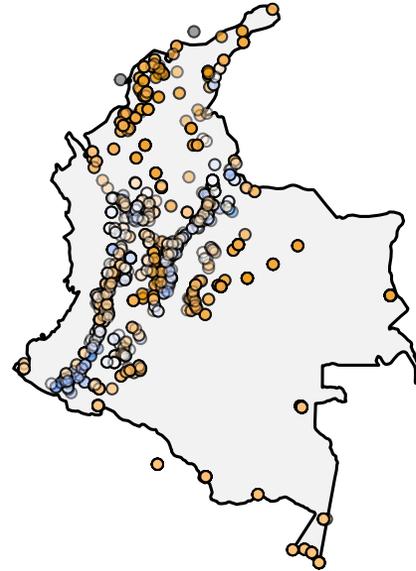
BIO15 = Precipitation Seasonality (Coefficient of Variation)

BIO16 = Precipitation of Wettest Quarter

BIO17 = Precipitation of Driest Quarter

BIO18 = Precipitation of Warmest Quarter

BIO19 = Precipitation of Coldest Quarter

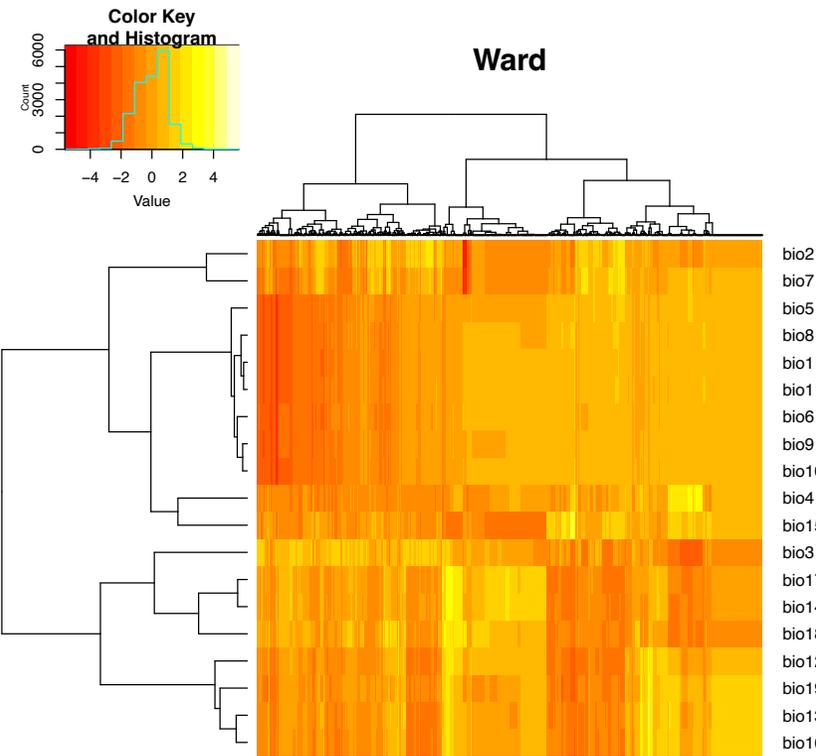
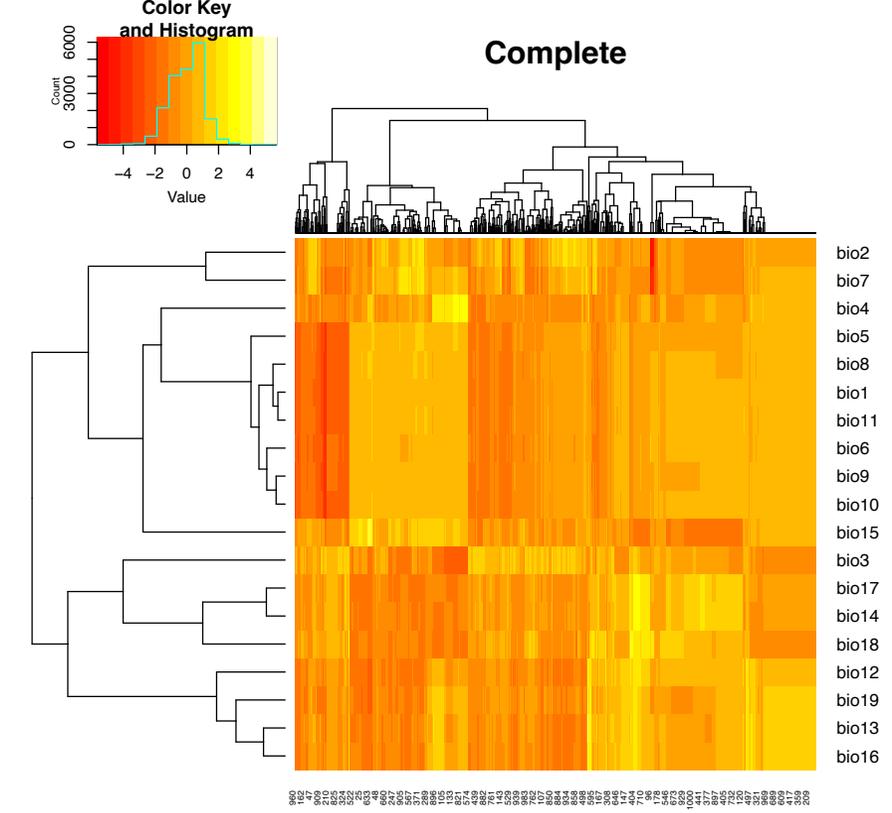
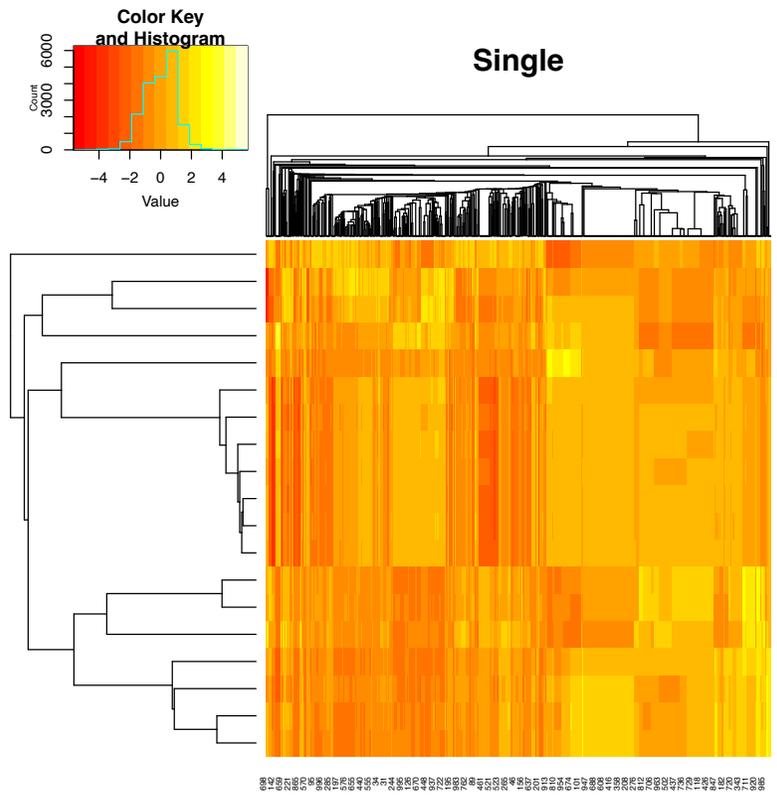


cassava landraces



19 bioclimatic variables

	bio1	bio2	bio3	bio4
1	-1.1666815	-0.14426182	0.94909285	-1.0560
2	-0.8190065	-0.44644337	-0.16748697	-0.2723
3	-1.1014924	1.36664596	0.05582899	-0.3888
4	-0.7538175	-0.34571619	0.72577688	-0.1686
5	-0.8841956	0.05719255	0.94909285	-1.0042
6	-2.6443000	-1.45371522	0.94909285	-0.2723
7	-1.5578158	0.25864692	1.17240881	-1.3022
8	-1.6881939	-0.74862493	1.39572477	-1.1079
9	-1.1014924	-0.84935211	0.72577688	-0.6220
10	-1.9924095	-0.34571619	-0.83743486	-1.2504



Mass-spectrometry-based draft of the *Arabidopsis* proteome

<https://doi.org/10.1038/s41586-020-2094-2>

Received: 4 June 2019

Accepted: 17 January 2020

Published online: 11 March 2020

Check for updates

Julia Mergner¹, Martin Frejno¹, Markus List², Michael Papacek³, Xia Chen⁴, Ajeet Chaudhary⁴, Patroklos Samaras¹, Sandra Richter⁵, Hiromasa Shikata^{6,7}, Maxim Messerer⁸, Daniel Lang⁸, Stefan Altmann⁹, Philipp Cyprys¹⁰, Daniel P. Zolg¹, Toby Mathieson¹¹, Marcus Bantscheff¹¹, Rashmi R. Hazarika^{12,13}, Tobias Schmidt¹, Corinna Dawid¹⁴, Andreas Dunkel¹⁴, Thomas Hofmann¹⁴, Stefanie Sprunck¹⁰, Pascal Falter-Braun^{9,15}, Frank Johannes^{12,13}, Klaus F. X. Mayer^{8,16}, Gerd Jürgens⁵, Mathias Wilhelm¹, Jan Baumbach², Erwin Grill⁵, Kay Schneitz⁴, Claus Schwechheimer⁶ & Bernhard Kuster^{1,13,17,18}

Gene expression data

>50 tissues

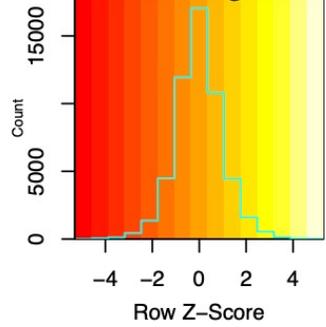


adult, adult vascular leaf	adult, cauline leaf	adult, cotyledon	adult, fruit septum	adult, fruit valve	adult, petiole, rosette leaf 7 petiole	adult, root	adult, rosette leaf, rosette leaf 1
3.8066625	3.4657359	3.8501476	3.3322045	3.0910425	3.8066625	3.5835189	3.7841890
2.6390573	3.1354942	2.9957323	2.3025851	2.4849066	2.8332133	2.9957323	3.3322045
2.8332133	3.1780538	3.1354942	3.0910425	2.9957323	2.9957323	3.4011974	3.0445224
2.7725887	1.9459101	1.7917595	1.9459101	2.6390573	1.0986123	0.6931472	1.6094379
3.0445224	3.4965076	2.9444390	3.5263605	3.5263605	2.9444390	3.2958369	3.3322045
4.7791235	4.8675345	4.3174881	4.3040651	4.1431347	4.3174881	4.5747110	4.3944492
3.7376696	3.6635616	3.4657359	3.5553481	4.0604430	3.6375862	3.7376696	3.3672958
3.6635616	3.6375862	3.5553481	3.7376696	3.7135721	3.2580965	3.4965076	3.7612001
5.8141305	5.4026774	4.9416424	5.3565863	6.1377271	3.8066625	2.8903718	5.3082677
3.6375862	5.4424177	3.9318256	4.3307333	5.1357984	5.5683445	2.9957323	5.4026774
0.0000000	1.3862944	0.0000000	0.6931472	0.6931472	1.0986123	0.6931472	1.0986123

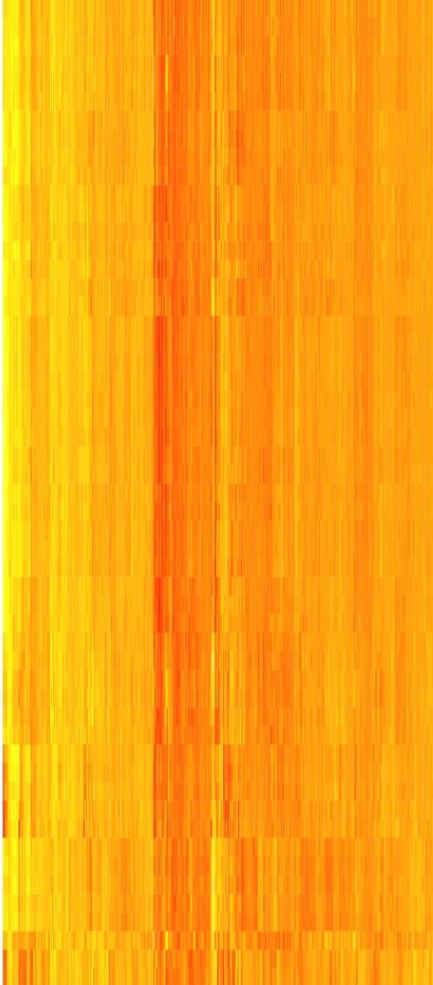
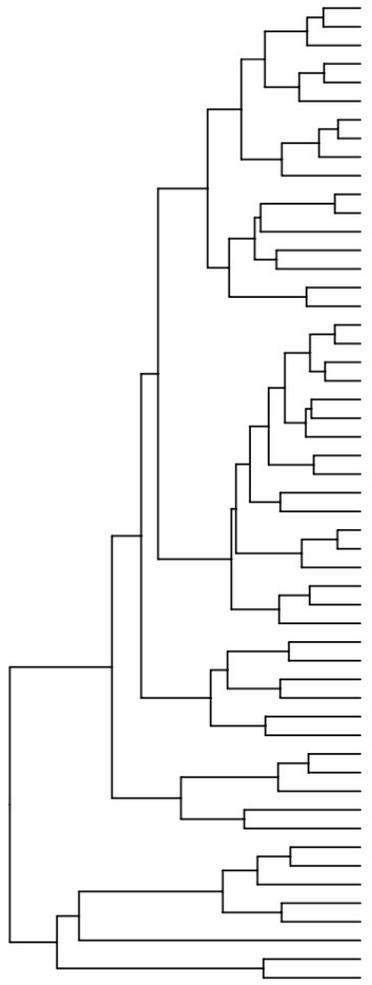
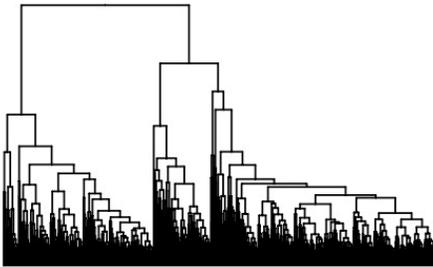
1000 genes



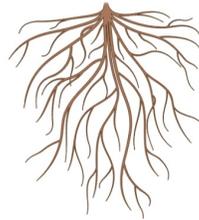
Color Key
and Histogram



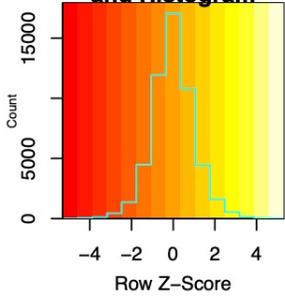
Euclidean



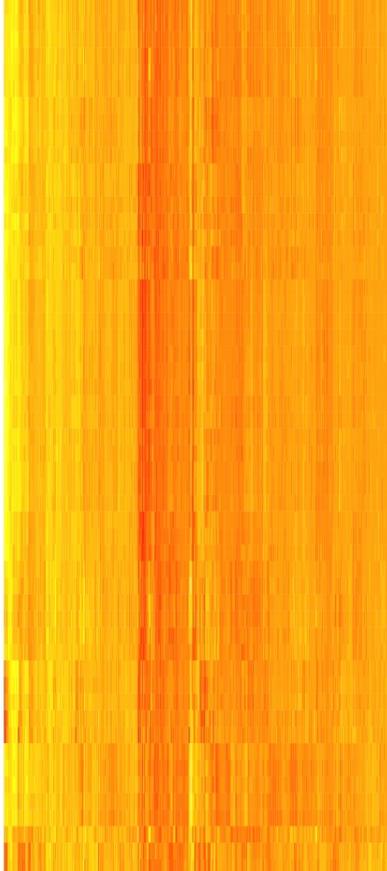
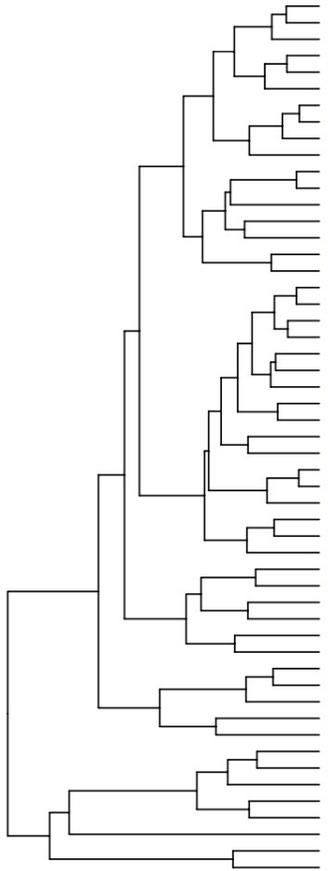
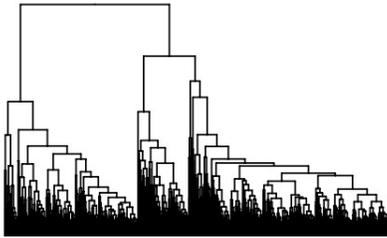
- flower stage 9, flower
- flower stage 10, flower
- flower stage 11, flower
- flower stage 15, flower
- flower stage 13, flower
- flower stage 12, flower
- siliqua stage 2, siliqua
- siliqua stage 1, siliqua
- siliqua stage 3, siliqua
- flower stage 15, carpel
- adult, stem node, first node
- adult, stem internode, second internode
- flower stage 15, flower pedicel
- seedling, hypocotyl
- adult, petiole, rosette leaf 7 petiole
- siliqua stage 4, siliqua
- siliqua stage 5, siliqua
- adult, rosette leaf, rosette leaf 4
- adult, rosette leaf, rosette leaf 3
- adult, rosette leaf, rosette leaf 5
- adult, rosette leaf, rosette leaf 6
- adult, rosette leaf, rosette leaf 9
- adult, rosette leaf, rosette leaf 7
- adult, rosette leaf, rosette leaf 8
- adult, rosette leaf, rosette leaf 7 proximal part
- adult, rosette leaf, rosette leaf 7 distal part
- seedling, mixed shoot apical meristem, cotyledon and first leave
- seedling, cotyledon
- adult, rosette leaf, rosette leaf 11
- adult, rosette leaf, rosette leaf 10
- adult, rosette leaf, rosette leaf 12
- adult, rosette leaf, rosette leaf 1
- adult, rosette leaf, rosette leaf 2
- adult, cauline leaf
- adult, fruit valve
- adult, fruit septum
- flower stage 15, petal
- flower stage 15, sepal
- adult, adult vascular leaf
- adult, cotyledon
- adult, root
- seedling, root tip, upper zone
- seedling, root tip
- callus, plant callus
- cell culture, root
- embryo stage 7, seed
- embryo stage 8, seed
- embryo stage 6, seed
- embryo stage 10, seed
- embryo stage 9, seed
- flower stage 15, stamen
- seed imbibition stage, seed
- dry seed stage, seed



Color Key
and Histogram

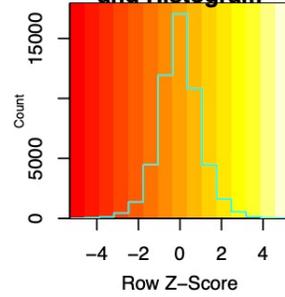


Euclidean

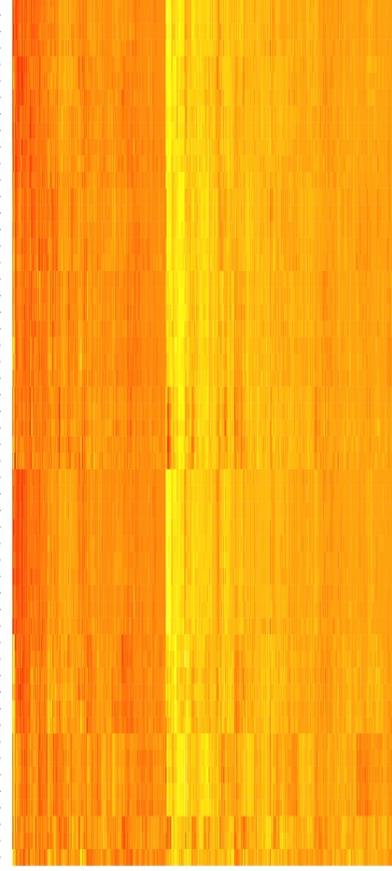
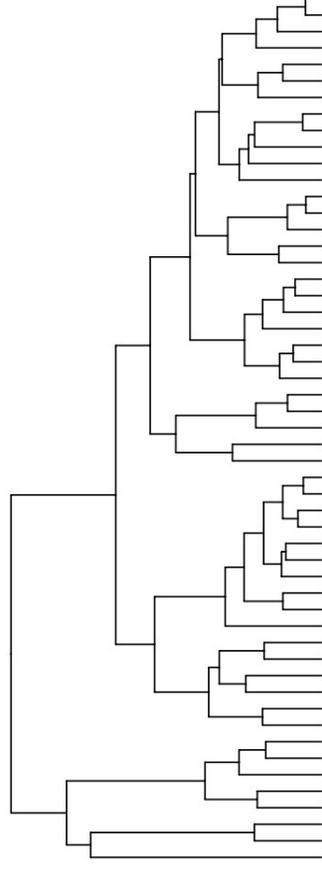
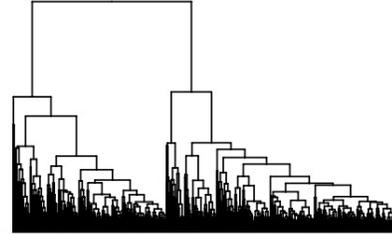


flower stage 9, flower
 flower stage 10, flower
 flower stage 11, flower
 flower stage 15, flower
 flower stage 13, flower
 flower stage 12, flower
 silique stage 2, silique
 silique stage 1, silique
 silique stage 3, silique
 flower stage 15, carpel
 adult, stem node, first node
 adult, stem internode, second internode
 flower stage 15, flower pedicel
 seedling, hypocotyl
 adult, petiole, rosette leaf 7 petiole
 silique stage 4, silique
 silique stage 5, silique
 adult, rosette leaf, rosette leaf 4
 adult, rosette leaf, rosette leaf 3
 adult, rosette leaf, rosette leaf 5
 adult, rosette leaf, rosette leaf 6
 adult, rosette leaf, rosette leaf 9
 adult, rosette leaf, rosette leaf 7
 adult, rosette leaf, rosette leaf 8
 adult, rosette leaf, rosette leaf 7 proximal part
 adult, rosette leaf, rosette leaf 7 distal part
 seedling, mixed shoot apical meristem, cotyledon and first leave
 seedling, cotyledon
 adult, rosette leaf, rosette leaf 11
 adult, rosette leaf, rosette leaf 10
 adult, rosette leaf, rosette leaf 12
 adult, rosette leaf, rosette leaf 1
 adult, rosette leaf, rosette leaf 2
 adult, cauline leaf
 adult, fruit valve
 adult, fruit septum
 flower stage 15, petal
 flower stage 15, sepal
 adult, adult vascular leaf
 adult, cotyledon
 adult, root
 seedling, root tip, upper zone
 seedling, root tip
 callus, plant callus
 cell culture, root
 embryo stage 7, seed
 embryo stage 8, seed
 embryo stage 6, seed
 embryo stage 10, seed
 embryo stage 9, seed
 flower stage 15, stamen
 seed imbibition stage, seed
 dry seed stage, seed

Color Key
and Histogram



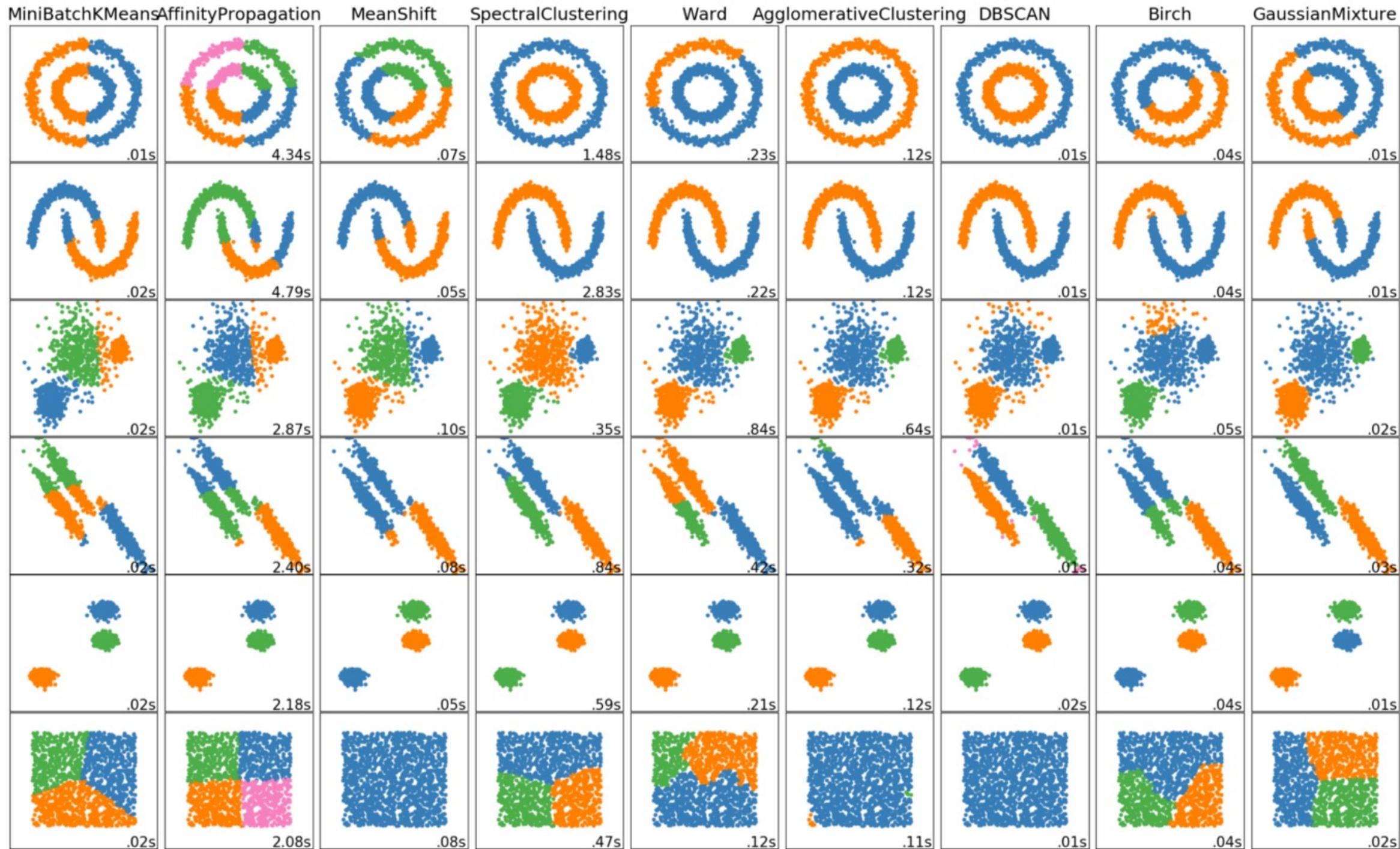
Manhattan



adult, rosette leaf, rosette leaf 11
 adult, rosette leaf, rosette leaf 10
 adult, rosette leaf, rosette leaf 12
 seedling, mixed shoot apical meristem, cotyledon and first leave
 adult, rosette leaf, rosette leaf 1
 adult, rosette leaf, rosette leaf 2
 adult, cauline leaf
 adult, stem node, first node
 adult, stem internode, second internode
 adult, petiole, rosette leaf 7 petiole
 flower stage 15, flower pedicel
 seedling, hypocotyl
 silique stage 2, silique
 silique stage 1, silique
 silique stage 3, silique
 silique stage 4, silique
 silique stage 5, silique
 flower stage 10, flower
 flower stage 11, flower
 flower stage 9, flower
 flower stage 15, carpel
 flower stage 15, flower
 flower stage 13, flower
 flower stage 12, flower
 adult, root
 seedling, root tip, upper zone
 seedling, root tip
 callus, plant callus
 cell culture, root
 adult, rosette leaf, rosette leaf 4
 adult, rosette leaf, rosette leaf 3
 adult, rosette leaf, rosette leaf 5
 adult, rosette leaf, rosette leaf 6
 adult, rosette leaf, rosette leaf 9
 adult, rosette leaf, rosette leaf 7
 adult, rosette leaf, rosette leaf 8
 adult, rosette leaf, rosette leaf 7 proximal part
 adult, rosette leaf, rosette leaf 7 distal part
 seedling, cotyledon
 adult, fruit valve
 adult, fruit septum
 adult, adult vascular leaf
 adult, cotyledon
 flower stage 15, petal
 flower stage 15, sepal
 embryo stage 7, seed
 embryo stage 8, seed
 embryo stage 6, seed
 embryo stage 10, seed
 embryo stage 9, seed
 seed imbibition stage, seed
 dry seed stage, seed
 flower stage 15, stamen



Other Clustering methods



Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

