

ETC3250/5250: Introduction to Machine Learning

Hierarchical clustering

Lecturer: *Professor Di Cook*

Department of Econometrics and Business Statistics

✉ ETC3250.Clayton-x@monash.edu

CALENDAR
Week 10a

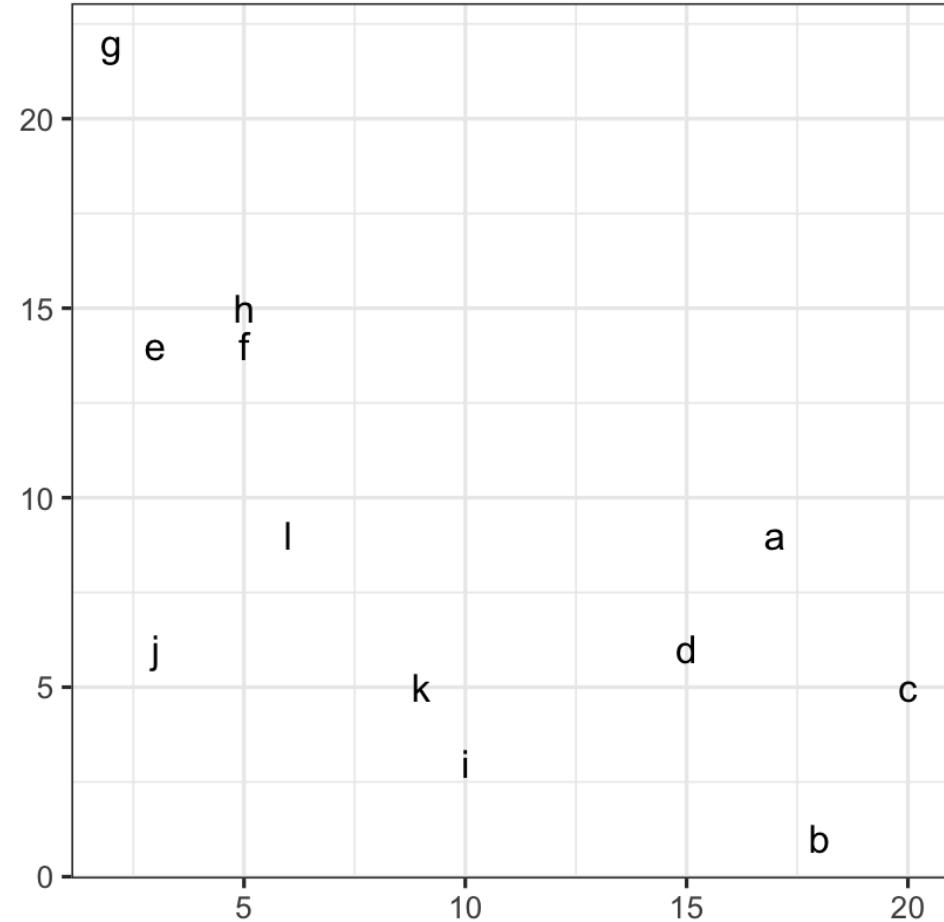


Hierarchical clustering

- **Agglomeration**: Begin with all observations in singleton clusters. Sequentially **join** points into clusters, until all are in one cluster.
- **Divisive**: Begin with all observations in one cluster, and sequentially **divide** until all observations are in singleton clusters.
- Produces a tree diagram illustrating the process, called a **dendrogram**.

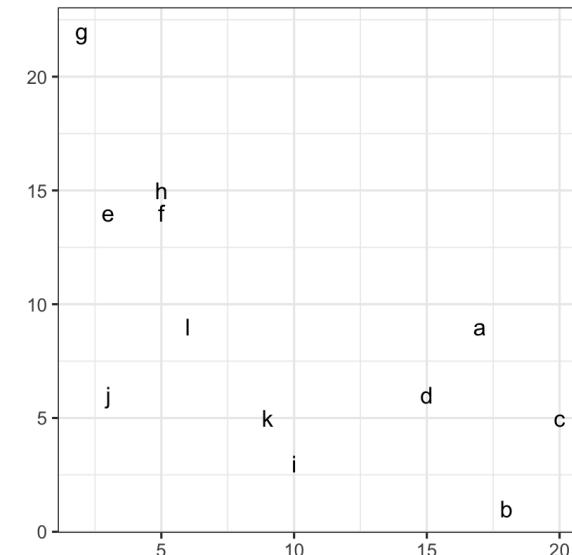
Some new data

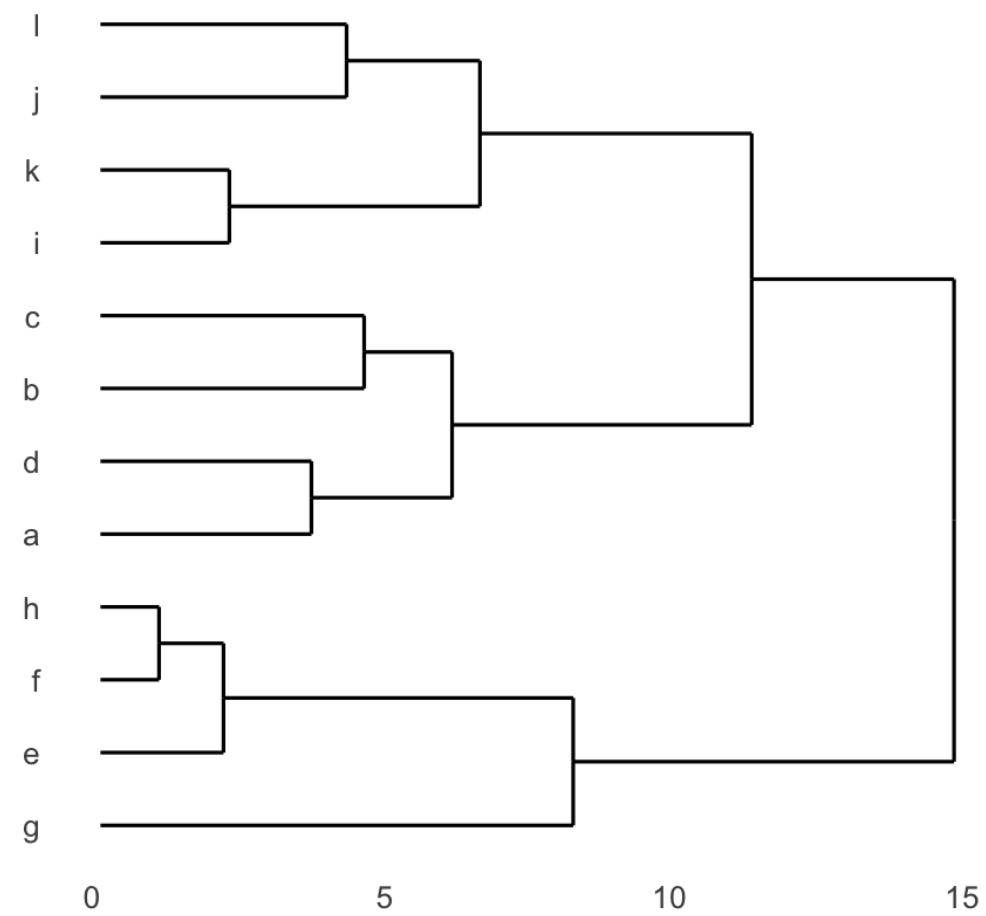
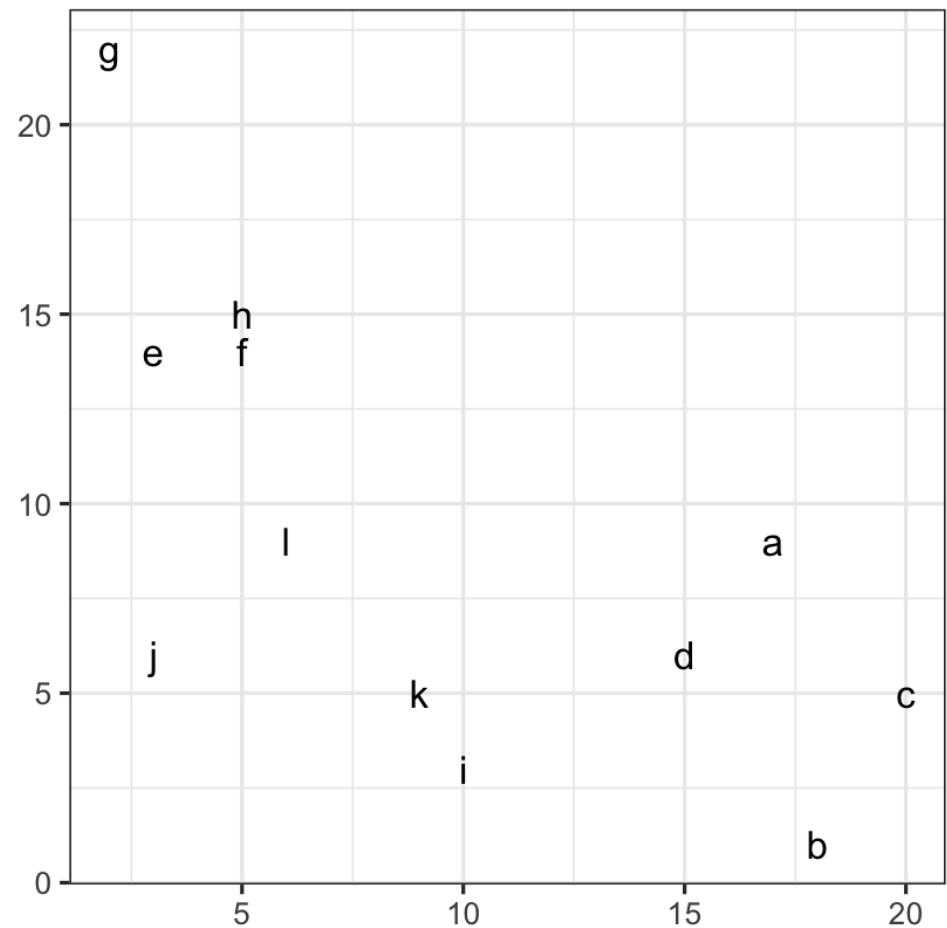
lbl	x1	x2
a	17	9
b	18	1
c	20	5
d	15	6
e	3	14
f	5	14
g	2	22
h	5	15
i	10	3
j	3	6
k	9	5
l	6	9

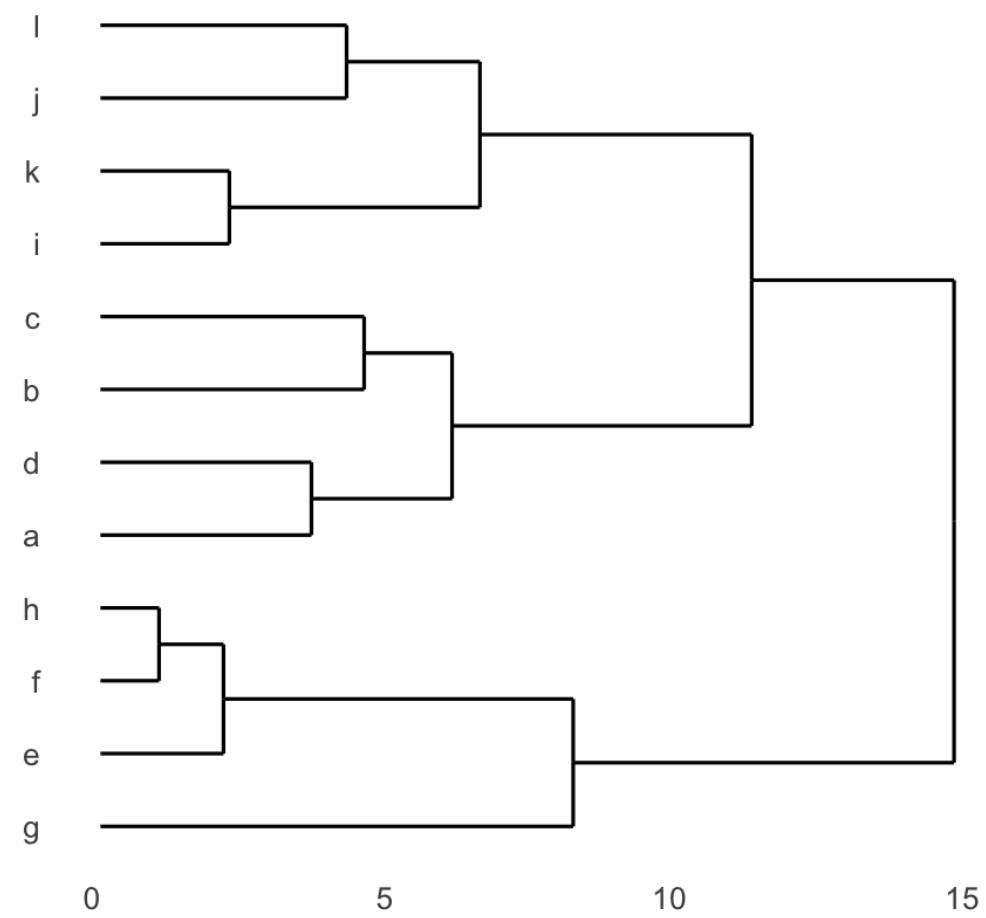
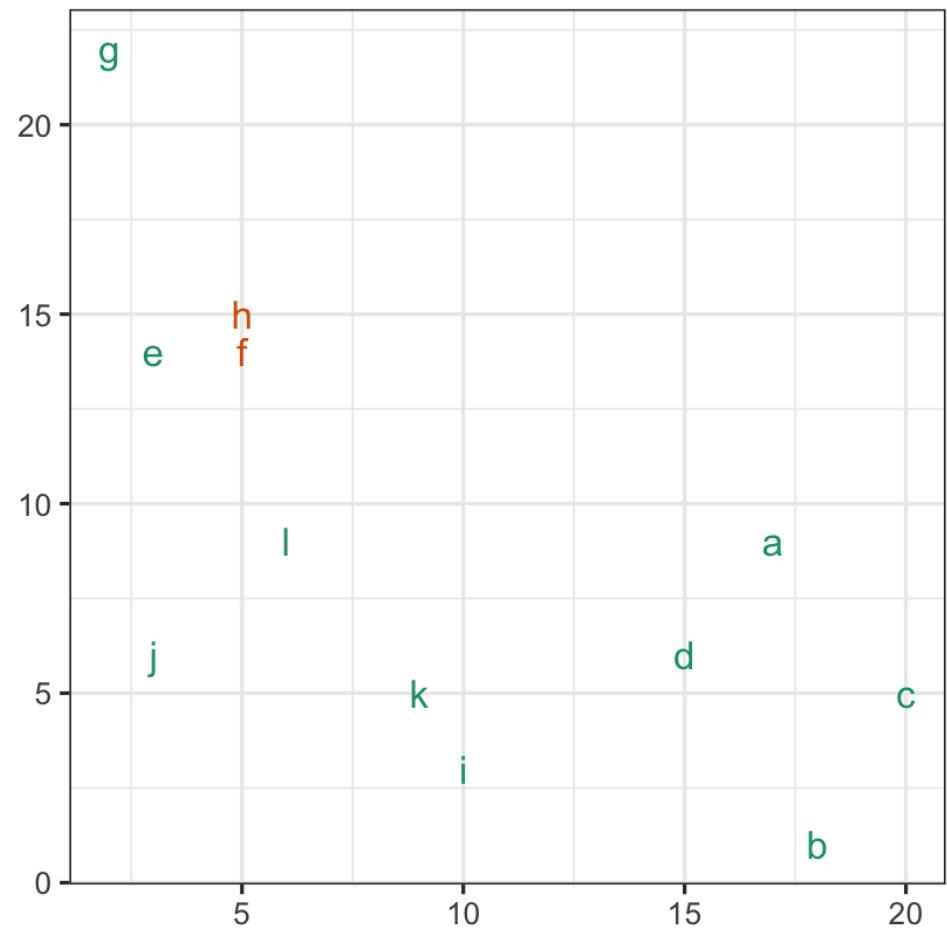


$n \times n$ distance matrix 

	a	b	c	d	e	f	g	h	i	j	k	l
a	0.0	8.1	5.0	3.6	14.9	13.0	19.8	13.4	9.2	14.3	8.9	11.0
b	8.1	0.0	4.5	5.8	19.8	18.4	26.4	19.1	8.2	15.8	9.8	14.4
c	5.0	4.5	0.0	5.1	19.2	17.5	24.8	18.0	10.2	17.0	11.0	14.6
d	3.6	5.8	5.1	0.0	14.4	12.8	20.6	13.5	5.8	12.0	6.1	9.5
e	14.9	19.8	19.2	14.4	0.0	2.0	8.1	2.2	13.0	8.0	10.8	5.8
f	13.0	18.4	17.5	12.8	2.0	0.0	8.5	1.0	12.1	8.2	9.8	5.1
g	19.8	26.4	24.8	20.6	8.1	8.5	0.0	7.6	20.6	16.0	18.4	13.6
h	13.4	19.1	18.0	13.5	2.2	1.0	7.6	0.0	13.0	9.2	10.8	6.1
i	9.2	8.2	10.2	5.8	13.0	12.1	20.6	13.0	0.0	7.6	2.2	7.2
j	14.3	15.8	17.0	12.0	8.0	8.2	16.0	9.2	7.6	0.0	6.1	4.2
k	8.9	9.8	11.0	6.1	10.8	9.8	18.4	10.8	2.2	6.1	0.0	5.0
l	11.0	14.4	14.6	9.5	5.8	5.1	13.6	6.1	7.2	4.2	5.0	0.0







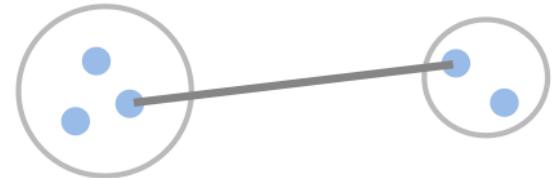


What is the **distance** between the new cluster **(f,h)** and all of the other observations?

Linkage

Between points **in** the cluster to points **not in** the cluster.

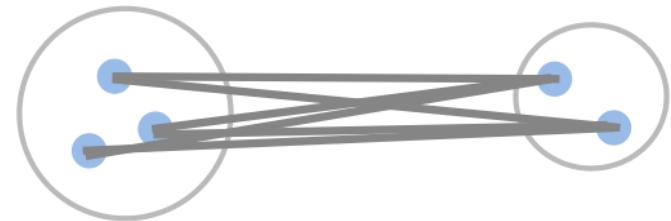
- **Single**: minimum distance between points in the different clusters
- **Complete**: maximum distance between points in the different clusters
- **Average**: average of distances between points in the different clusters
- **Centroid**: distances between the average of the different clusters
- **Wards**: minimizes the total within-cluster variance



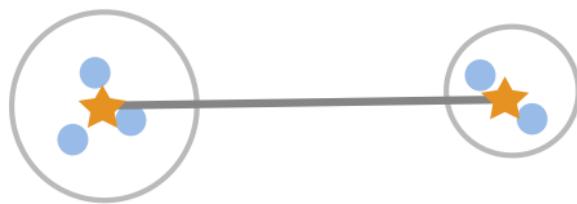
single



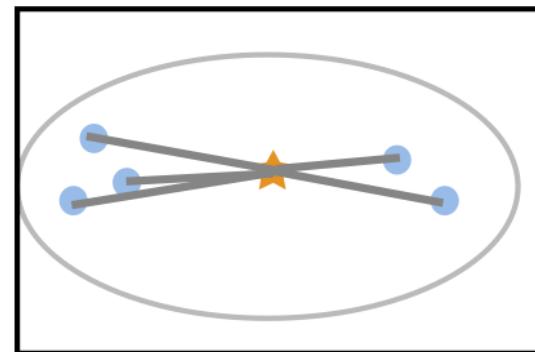
complete



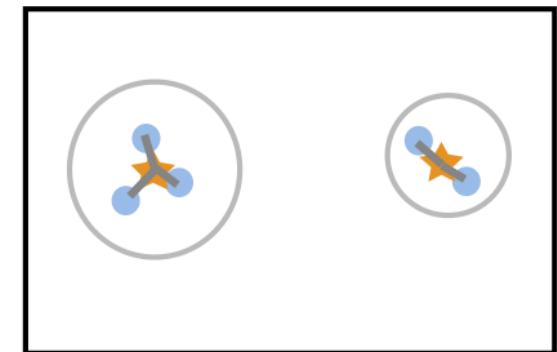
average



centroid



wards



	a	b	c	d	e	f	g	h	i	j	k	l
a	0.0	8.1	5.0	3.6	14.9	13.0	19.8	13.4	9.2	14.3	8.9	11.0
b	8.1	0.0	4.5	5.8	19.8	18.4	26.4	19.1	8.2	15.8	9.8	14.4
c	5.0	4.5	0.0	5.1	19.2	17.5	24.8	18.0	10.2	17.0	11.0	14.6
d	3.6	5.8	5.1	0.0	14.4	12.8	20.6	13.5	5.8	12.0	6.1	9.5
e	14.9	19.8	19.2	14.4	0.0	2.0	8.1	2.2	13.0	8.0	10.8	5.8
f	13.0	18.4	17.5	12.8	2.0	0.0	8.5	1.0	12.1	8.2	9.8	5.1
g	19.8	26.4	24.8	20.6	8.1	8.5	0.0	7.6	20.6	16.0	18.4	13.6
h	13.4	19.1	18.0	13.5	2.2	1.0	7.6	0.0	13.0	9.2	10.8	6.1
i	9.2	8.2	10.2	5.8	13.0	12.1	20.6	13.0	0.0	7.6	2.2	7.2
j	14.3	15.8	17.0	12.0	8.0	8.2	16.0	9.2	7.6	0.0	6.1	4.2
k	8.9	9.8	11.0	6.1	10.8	9.8	18.4	10.8	2.2	6.1	0.0	5.0
l	11.0	14.4	14.6	9.5	5.8	5.1	13.6	6.1	7.2	4.2	5.0	0.0

Distance (linkage)
between (f,h) and a

Single: 13.0 or 13.4 →
13.0

Complete: 13.0 or 13.4
→ **13.4**

Average: $(13.0 + 13.4)/2$
= **13.2**

Centroid: **13.2**

mean of (f,h) is (5, 14.5)

mean of a (17,9)

$$\sqrt{(5 - 17)^2 + (14.5 - 9)^2}$$

Wards: Your turn to
calculate it 🤷

lbl	x1	x2	cl11
a	17	9	1
b	18	1	1
c	20	5	1
d	15	6	1
e	3	14	1
f	5	14	2
g	2	22	1
h	5	15	2
i	10	3	1
j	3	6	1
k	9	5	1
l	6	9	1

Distance (linkage) between (f,h) and
a

Single: 13.0 or 13.4 → 13.0

Complete: 13.0 or 13.4 → 13.4

Average: $(13.0 + 13.4)/2 = 13.2$

Centroid: 13.2

mean of (f,h) is (5, 14.5)

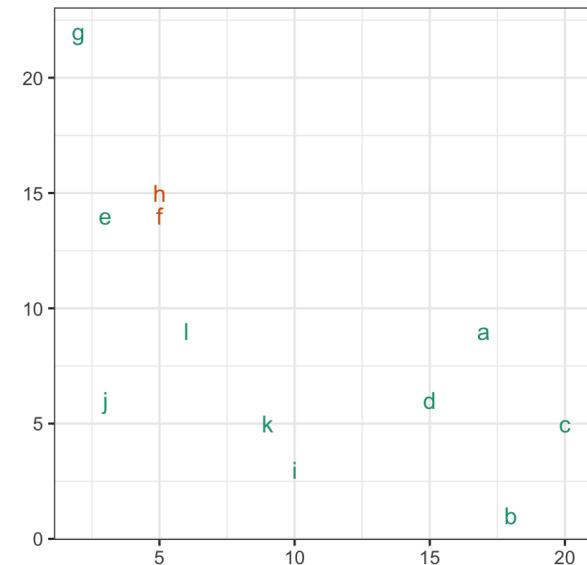
mean of a (17,9)

$$\sqrt{(5 - 17)^2 + (14.5 - 9)^2}$$

Wards: Your turn to calculate it 🤓

Single (minimum distance) linkage reduced distance matrix

	a	b	c	d	e	(f,h)	g	i	j	k	l
a	0.0	8.1	5.0	3.6	14.9	13.0	19.8	9.2	14.3	8.9	11.0
b	8.1	0.0	4.5	5.8	19.8	18.4	26.4	8.2	15.8	9.8	14.4
c	5.0	4.5	0.0	5.1	19.2	17.5	24.8	10.2	17.0	11.0	14.6
d	3.6	5.8	5.1	0.0	14.4	12.8	20.6	5.8	12.0	6.1	9.5
e	14.9	19.8	19.2	14.4	0.0	2.0	8.1	13.0	8.0	10.8	5.8
(f,h)	13.0	18.4	17.5	12.8	2.0	0.0	7.6	12.1	8.2	9.8	5.1
g	19.8	26.4	24.8	20.6	8.1	7.6	0.0	20.6	16.0	18.4	13.6
i	9.2	8.2	10.2	5.8	13.0	12.1	20.6	0.0	7.6	2.2	7.2
j	14.3	15.8	17.0	12.0	8.0	8.2	16.0	7.6	0.0	6.1	4.2
k	8.9	9.8	11.0	6.1	10.8	9.8	18.4	2.2	6.1	0.0	5.0
l	11.0	14.4	14.6	9.5	5.8	5.1	13.6	7.2	4.2	5.0	0.0

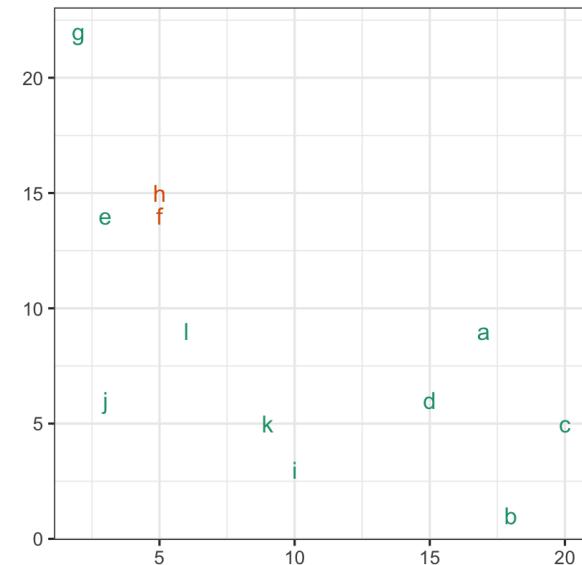


🤔 Now which **distance** is the **smallest**?

(f,h) and **e** would be joined at the next step

Complete (maximum distance) linkage reduced distance matrix

	a	b	c	d	e	(f,h)	g	i	j	k	l
a	0.0	8.1	5.0	3.6	14.9	13.4	19.8	9.2	14.3	8.9	11.0
b	8.1	0.0	4.5	5.8	19.8	19.1	26.4	8.2	15.8	9.8	14.4
c	5.0	4.5	0.0	5.1	19.2	18.0	24.8	10.2	17.0	11.0	14.6
d	3.6	5.8	5.1	0.0	14.4	13.5	20.6	5.8	12.0	6.1	9.5
e	14.9	19.8	19.2	14.4	0.0	2.2	8.1	13.0	8.0	10.8	5.8
(f,h)	13.4	19.1	18.0	13.5	2.2	0.0	8.5	13.0	9.2	10.8	6.1
g	19.8	26.4	24.8	20.6	8.1	8.5	0.0	20.6	16.0	18.4	13.6
i	9.2	8.2	10.2	5.8	13.0	13.0	20.6	0.0	7.6	2.2	7.2
j	14.3	15.8	17.0	12.0	8.0	9.2	16.0	7.6	0.0	6.1	4.2
k	8.9	9.8	11.0	6.1	10.8	10.8	18.4	2.2	6.1	0.0	5.0
l	11.0	14.4	14.6	9.5	5.8	6.1	13.6	7.2	4.2	5.0	0.0

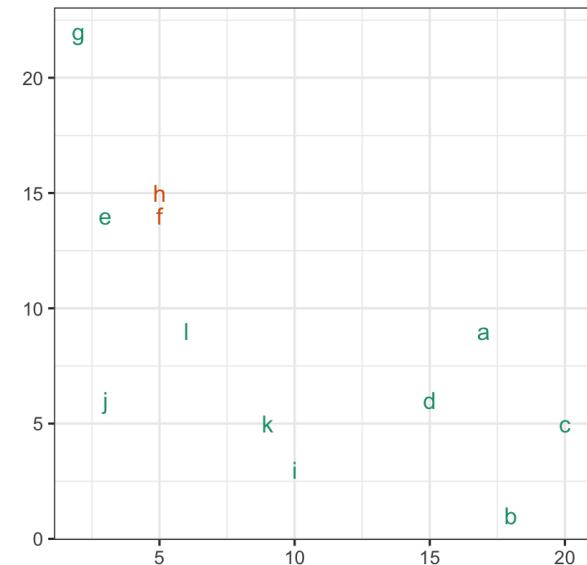


🤔 Now which **distance** is the **smallest**?

(f,h) and **e** and also **i** and **k** would be joined at the next step

Average linkage reduced distance matrix

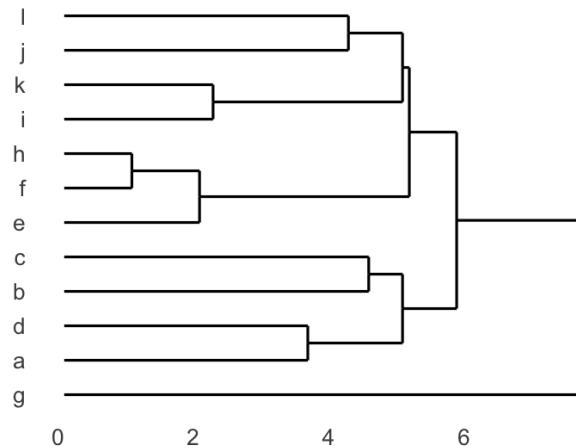
	a	b	c	d	e	(f,h)	g	i	j	k	l
a	0.0	8.1	5.0	3.6	14.9	13.0	19.8	9.2	14.3	8.9	11.0
b	8.1	0.0	4.5	5.8	19.8	18.4	26.4	8.2	15.8	9.8	14.4
c	5.0	4.5	0.0	5.1	19.2	17.5	24.8	10.2	17.0	11.0	14.6
d	3.6	5.8	5.1	0.0	14.4	12.8	20.6	5.8	12.0	6.1	9.5
e	14.9	19.8	19.2	14.4	0.0	2.0	8.1	13.0	8.0	10.8	5.8
(f,h)	13.0	18.4	17.5	12.8	2.0	0.0	8.5	12.1	8.2	9.8	5.1
g	19.8	26.4	24.8	20.6	8.1	8.5	0.0	20.6	16.0	18.4	13.6
i	9.2	8.2	10.2	5.8	13.0	12.1	20.6	0.0	7.6	2.2	7.2
j	14.3	15.8	17.0	12.0	8.0	8.2	16.0	7.6	0.0	6.1	4.2
k	8.9	9.8	11.0	6.1	10.8	9.8	18.4	2.2	6.1	0.0	5.0
l	11.0	14.4	14.6	9.5	5.8	5.1	13.6	7.2	4.2	5.0	0.0



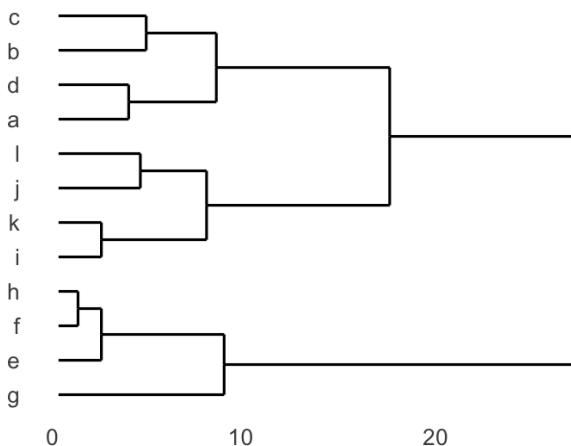
🤔 Now which **distance** is the **smallest**?

(f,h) and e would be joined at the next step

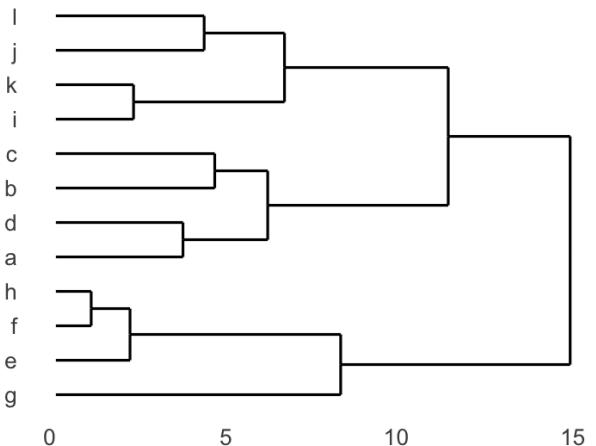
single



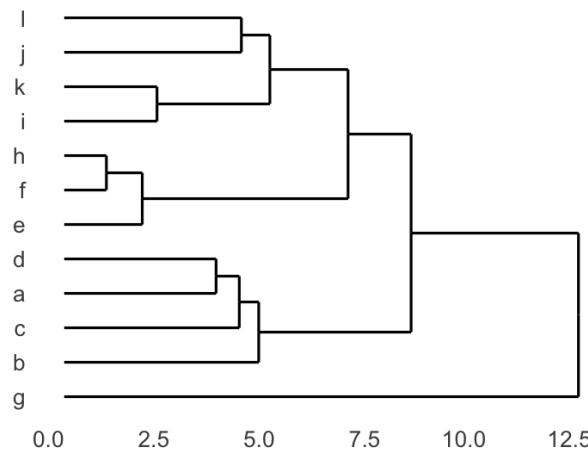
complete



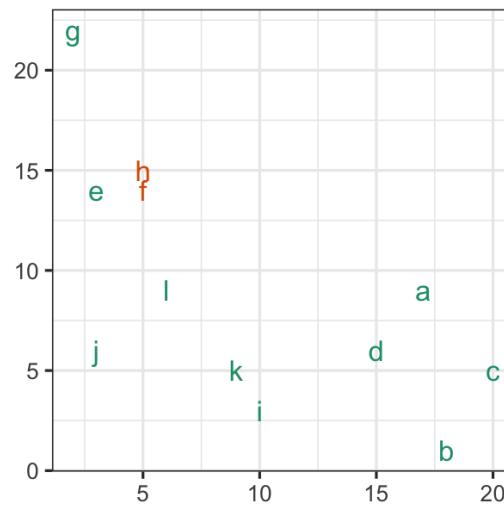
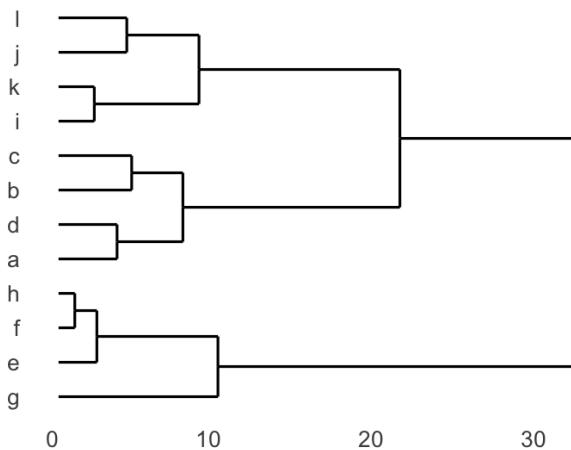
average



centroid



wards

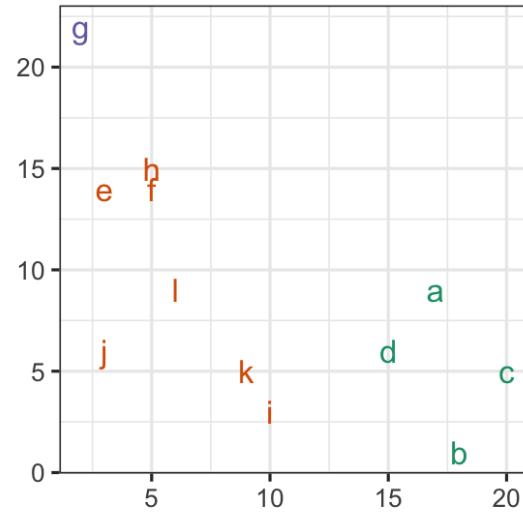


Dendrogram

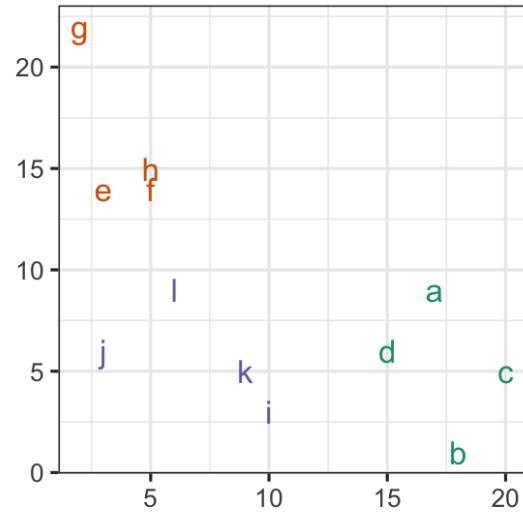
- Each **leaf** of the dendrogram represents one observation
- Leaves **fuse** into branches and branches fuse, either with leaves or other branches.
- Fusions **lower in the tree** mean the groups of observations are more similar to each other.

Cut the tree to partition the data into k clusters.

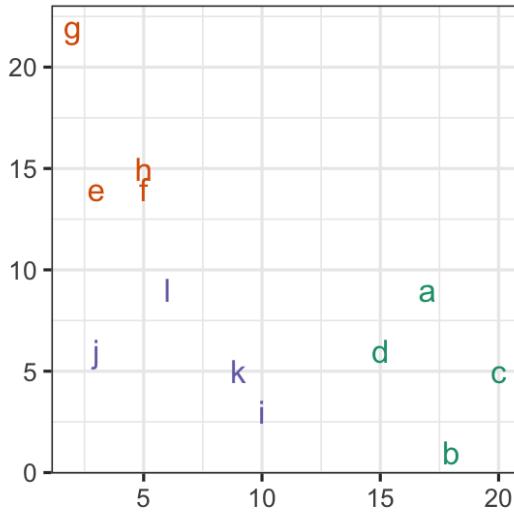
single



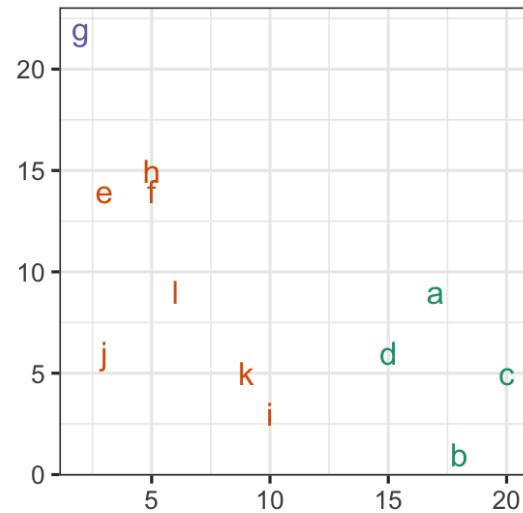
complete



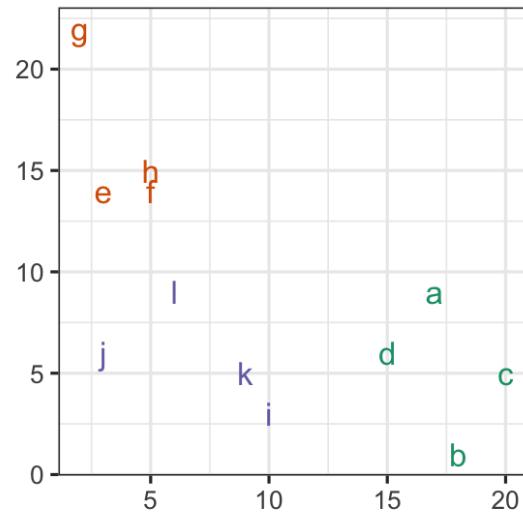
average



centroid



wards



Pros and cons

- Single linkage tends to "chain" the data into long stringy clusters, can avoid confusion from nuisance variables but gets confused by "inliers" (outliers between clusters)
- Complete linkage tends to be confused by nuisance variables, but not by inliers
- Wards tends to create spherical homogeneously shaped clusters, a little similar to k -means

No one perfect method for all problems, but Wards tends to be a good starting point.



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#).

Lecturer: *Professor Di Cook*

Department of Econometrics and Business Statistics

✉ ETC3250.Clayton-x@monash.edu

CALENDAR
Week 10a

