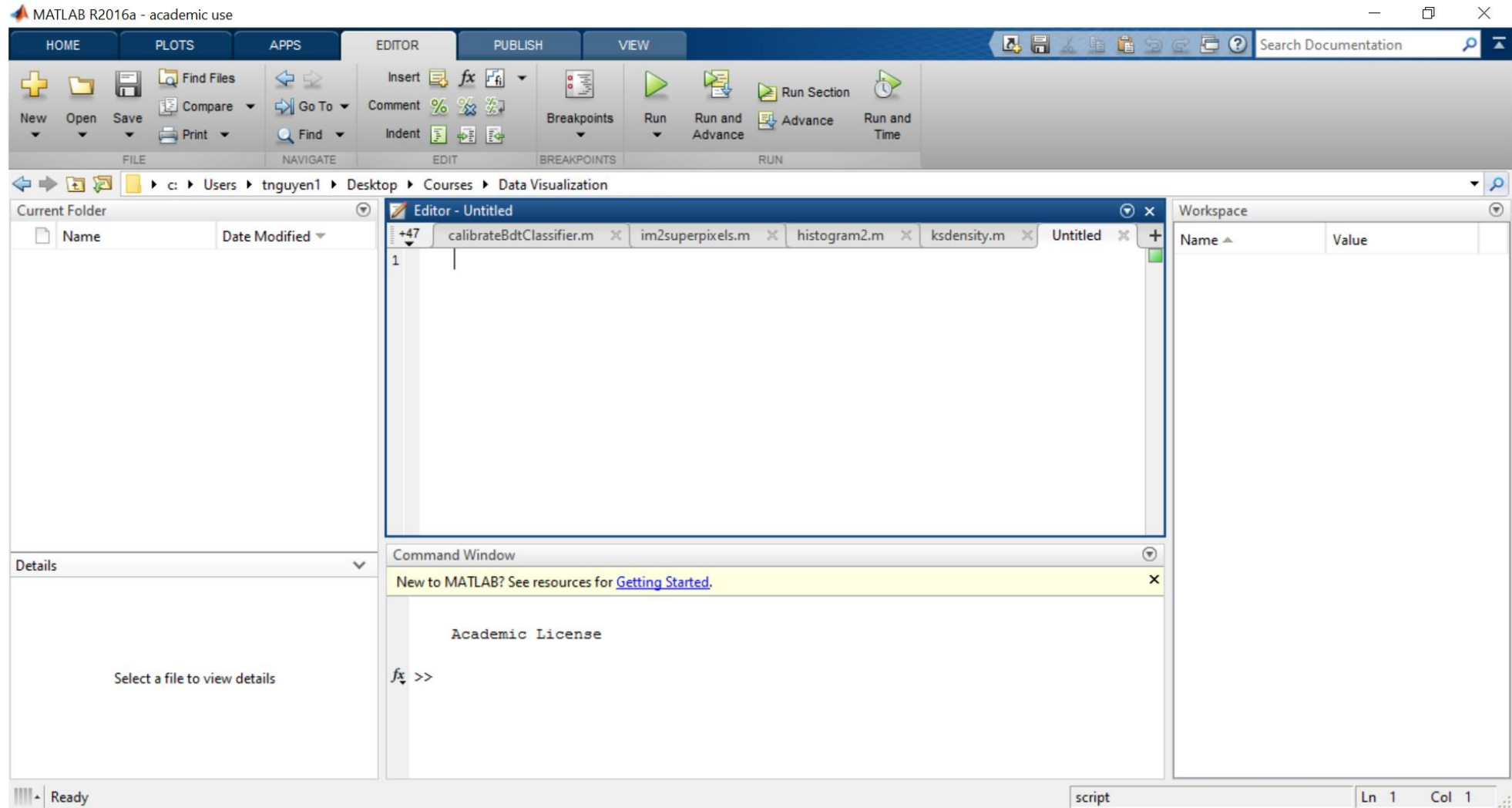# Lab 14

CPS 563 – Data Visualization
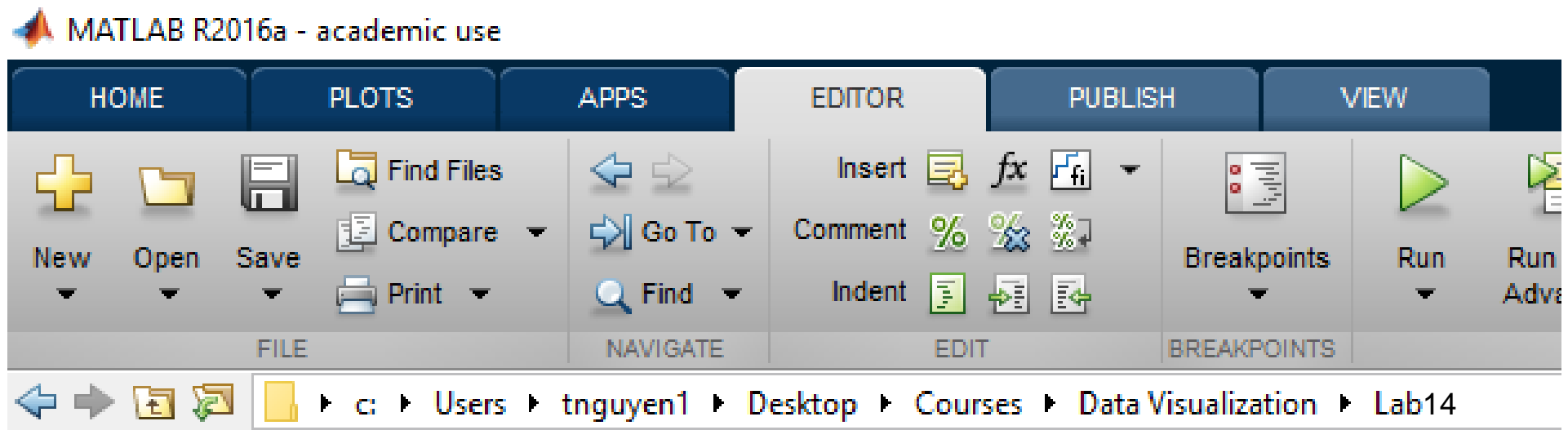
Dr. Tam Nguyen

tamnguyen@udayton.edu

Outline

- Perform hierarchical clustering
- Visualize hierarchical clustered data
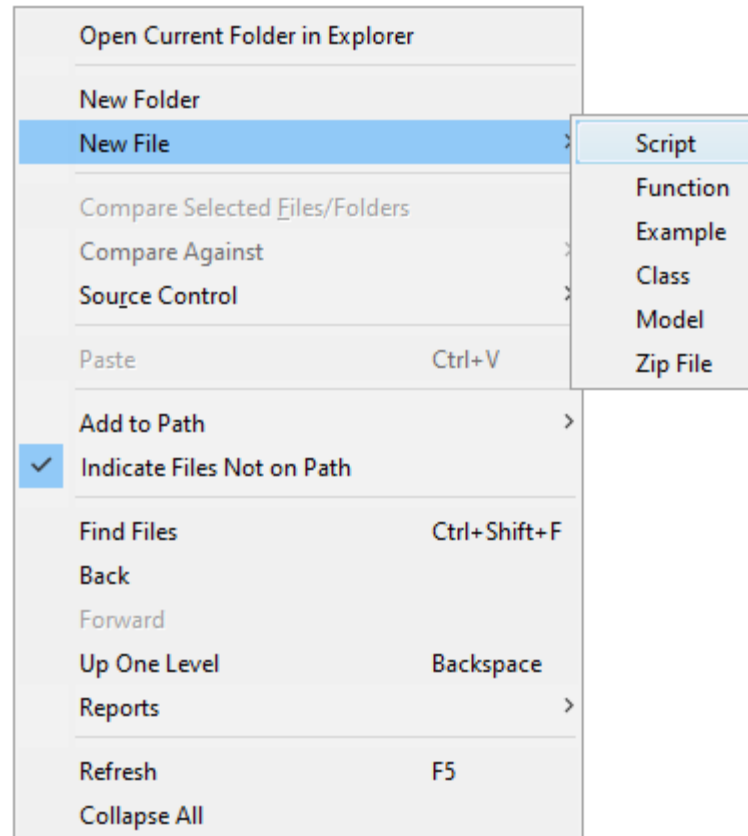
Start MATLAB

Create Lab 14 folder

Create new script file: Lab14a.m

Lab14a.m

close all;

clear all;

clc;

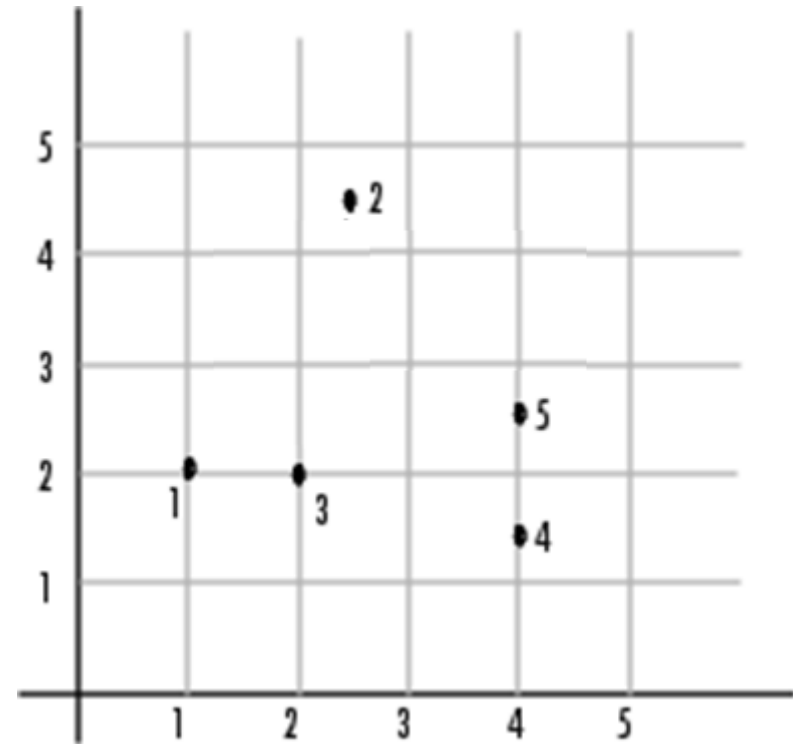# Lab14a.m – Prepare data

close all;

clear all;

clc;

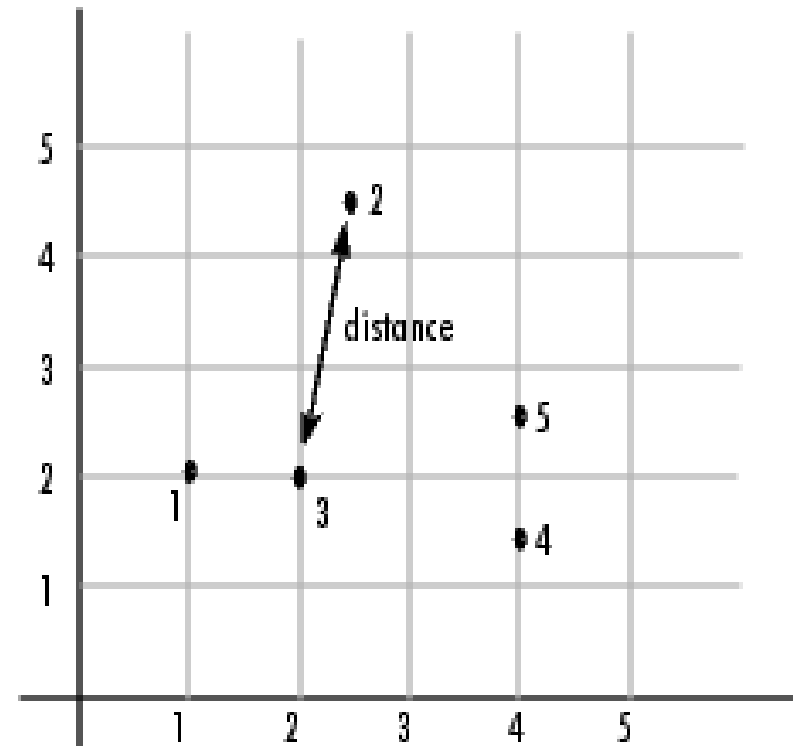**X = [1 2; 2.5 4.5; 2 2; 4 1.5; 4 2.5];**

Lab14a.m – Compute the distances

close all;

clear all;

clc;

X = [1 2; 2.5 4.5; 2 2; 4 1.5; 4 2.5];

D = pdist(X);

1×10 double									

D	1	2	3	4	5	6	7	8	9	10
1	2.9155	1.0000	3.0414	3.0414	2.5495	3.3541	2.5000	2.0616	2.0616	1.0000

Lab14a.m – Perform hierarchical clustering
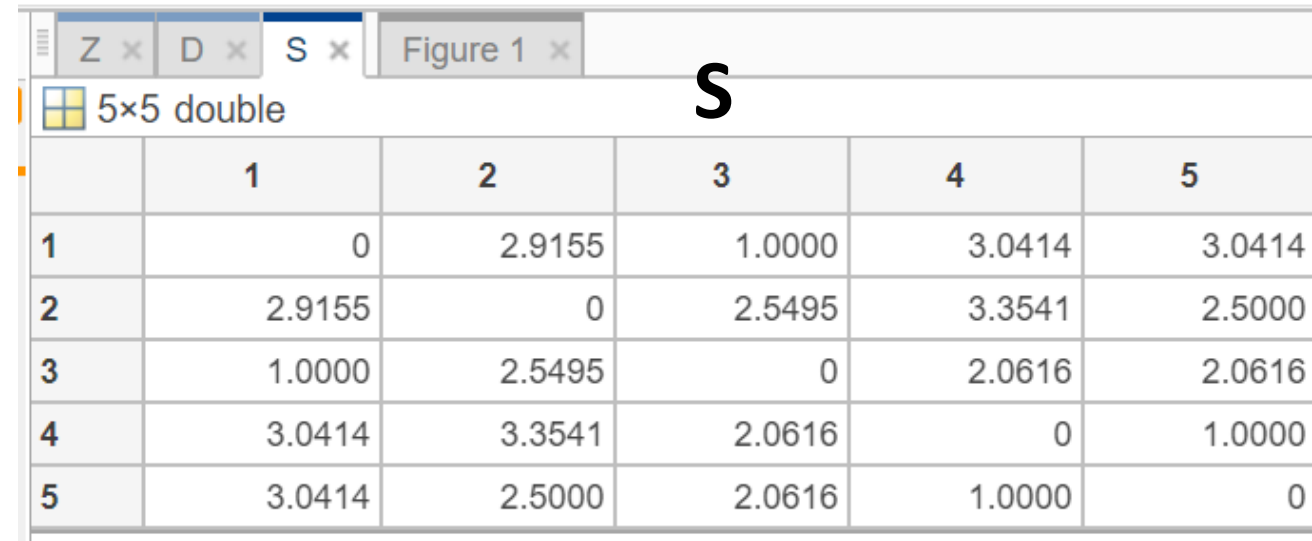
close all;

clear all;

clc;

X = [1 2; 2.5 4.5; 2 2; 4 1.5; 4 2.5];

D = pdist(X);

S = squareform(D);

S

5×5 double

	1	2	3	4	5
1	0	2.9155	1.0000	3.0414	3.0414
2	2.9155	0	2.5495	3.3541	2.5000
3	1.0000	2.5495	0	2.0616	2.0616
4	3.0414	3.3541	2.0616	0	1.0000
5	3.0414	2.5000	2.0616	1.0000	0
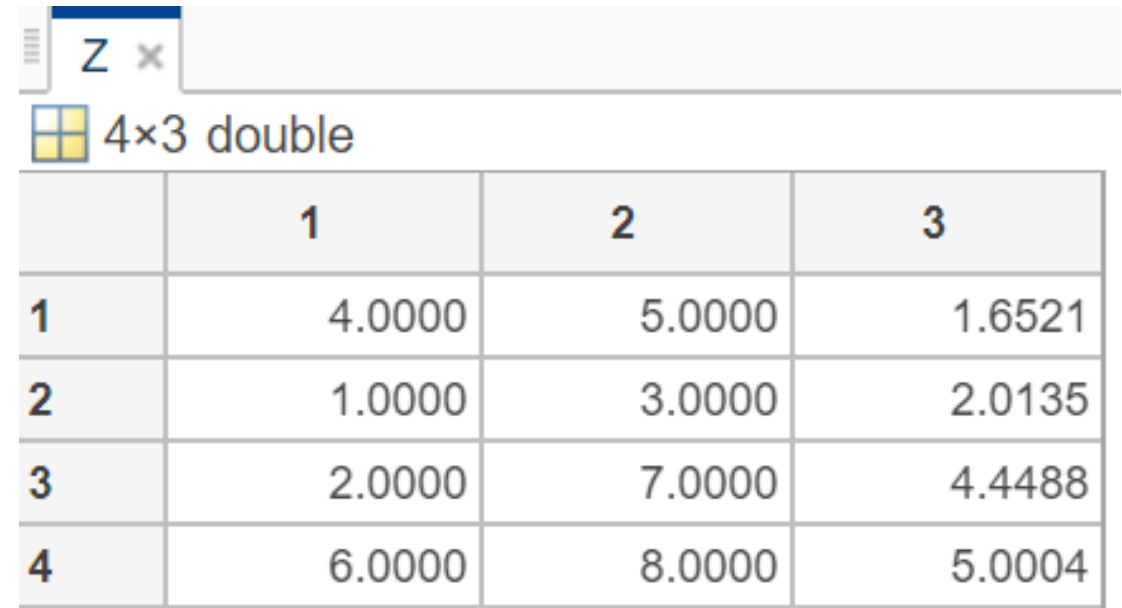
Z × D × S × Figure 1 ×

Lab14a.m – Perform hierarchical clustering

close all;

clear all;

clc;

X = [1 2; 2.5 4.5; 2 2; 4 1.5; 4 2.5];

D = pdist(X);

S = squareform(D);

Z = linkage(S, 'complete');

Z ×

4×3 double

	1	2	3
1	4.0000	5.0000	1.6521
2	1.0000	3.0000	2.0135
3	2.0000	7.0000	4.4488
4	6.0000	8.0000	5.0004

What does 'complete' mean?

Method	Description
'average'	Unweighted average distance (UPGMA)
'centroid'	Centroid distance (UPGMC), appropriate for Euclidean distances only
'complete'	Farthest distance
'median'	Weighted center of mass distance (WPGMC), appropriate for Euclidean distances only
'single'	Shortest distance
'ward'	Inner squared distance (minimum variance algorithm), appropriate for Euclidean distances only
'weighted'	Weighted average distance (WPGMA)

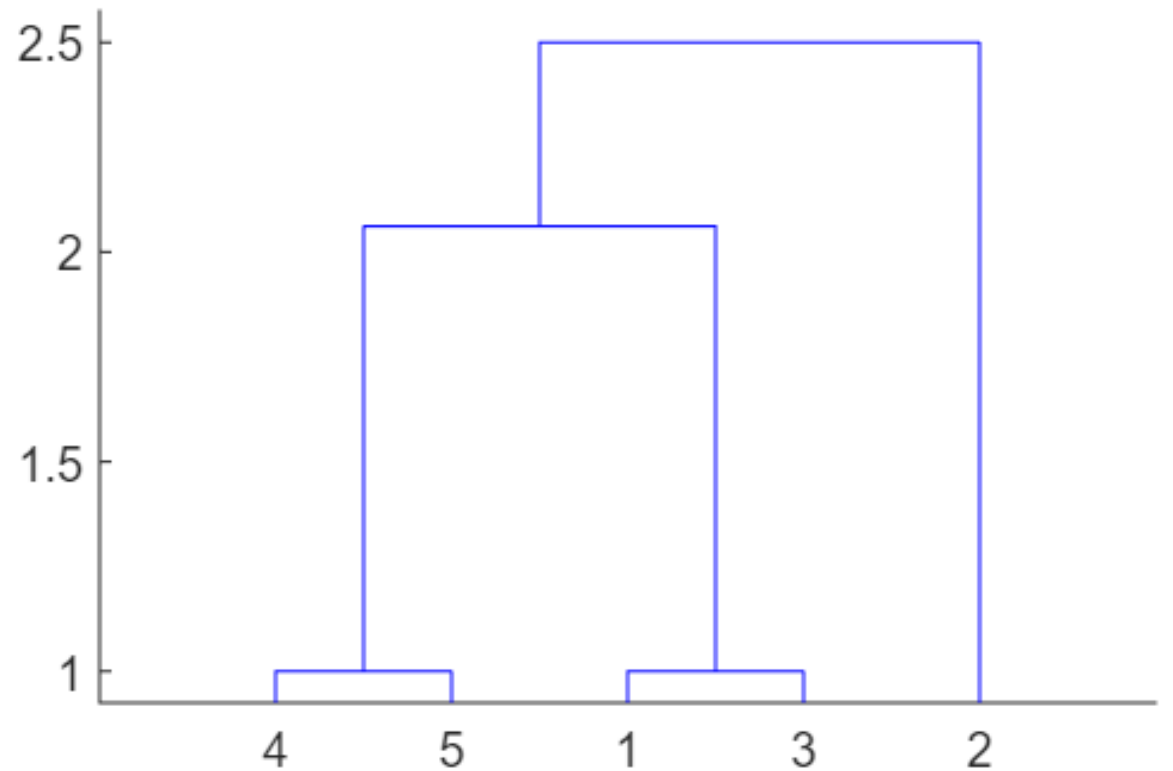# Lab14a.m – Visualize the dendrogram

close all;
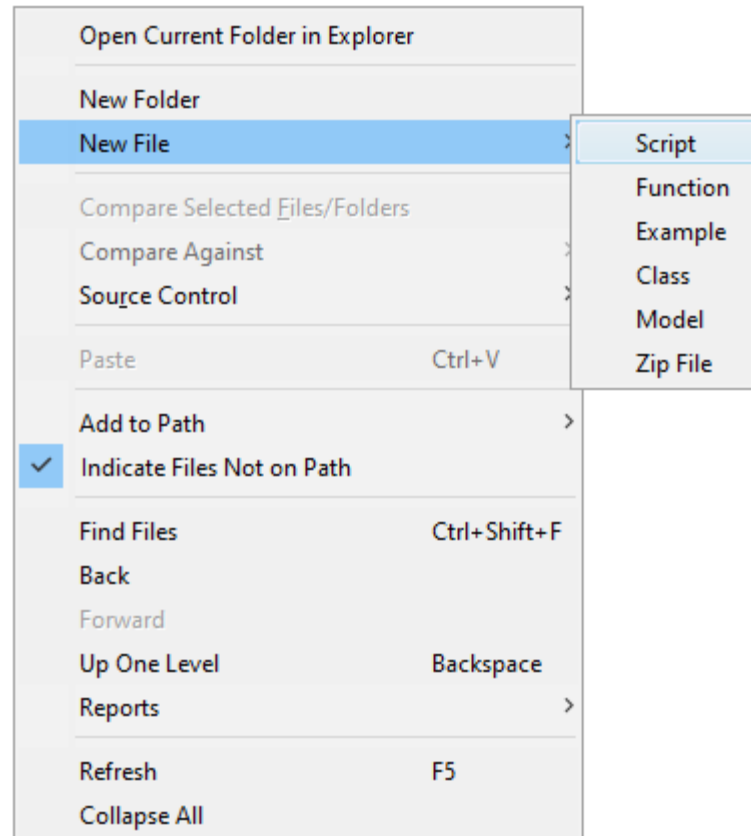
clear all;

clc;

X = [1 2; 2.5 4.5; 2 2; 4 1.5; 4 2.5];

D = pdist(X);

S = squareform(D);

Z = linkage(S, 'complete');

dendrogram(Z);

Create new script file: Lab14b.m

Lab14b.m

```
close all;
clear all;
clc;
```

Lab14b.m – Prepare data

close all;

clear all;

clc;

X = [1 2; 2.5 4.5; 2 2; 4 1.5; 4 2.5];

Lab14b.m – Compute the distances

close all;

clear all;

clc;

X = [1 2; 2.5 4.5; 2 2; 4 1.5; 4 2.5];

D = pdist(X);

Lab14b.m – Perform hierarchical clustering

close all;

clear all;

clc;

X = [1 2; 2.5 4.5; 2 2; 4 1.5; 4 2.5];

D = pdist(X);

Z = linkage(D);

Lab14b.m – Visualize the dendrogram

close all;

clear all;

clc;

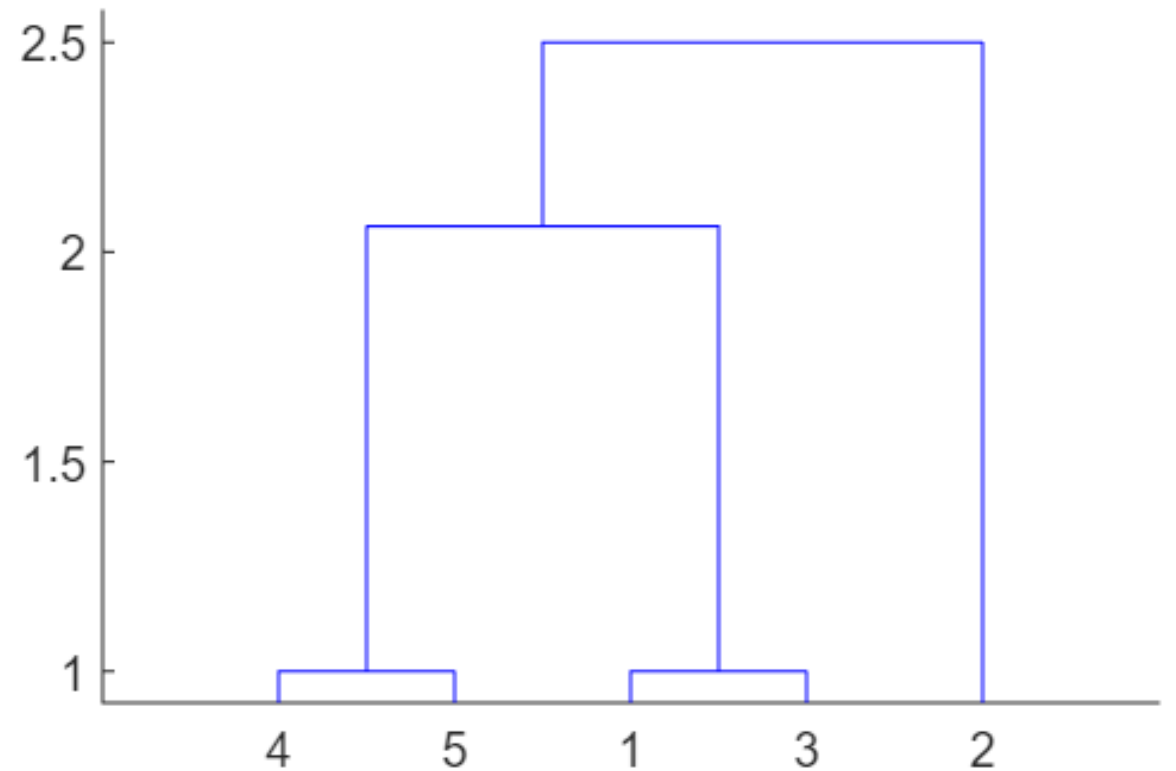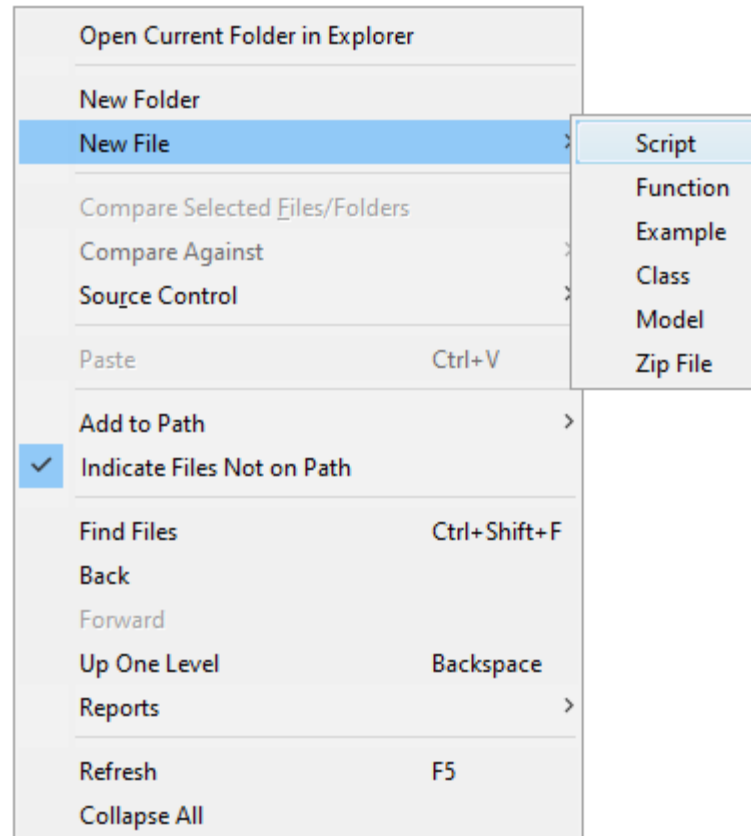X = [1 2; 2.5 4.5; 2 2; 4 1.5; 4 2.5];

D = pdist(X);

Z = linkage(D);

dendrogram(Z);

Create new script file: Lab14c.m

Lab14c.m

close all;

clear all;

clc;

Lab14c.m – Prepare data

close all;

clear all;

clc;

load fisheriris;

Lab14c.m – Compute the distances

close all;

clear all;

clc;

load fisheriris;

D = pdist(meas);

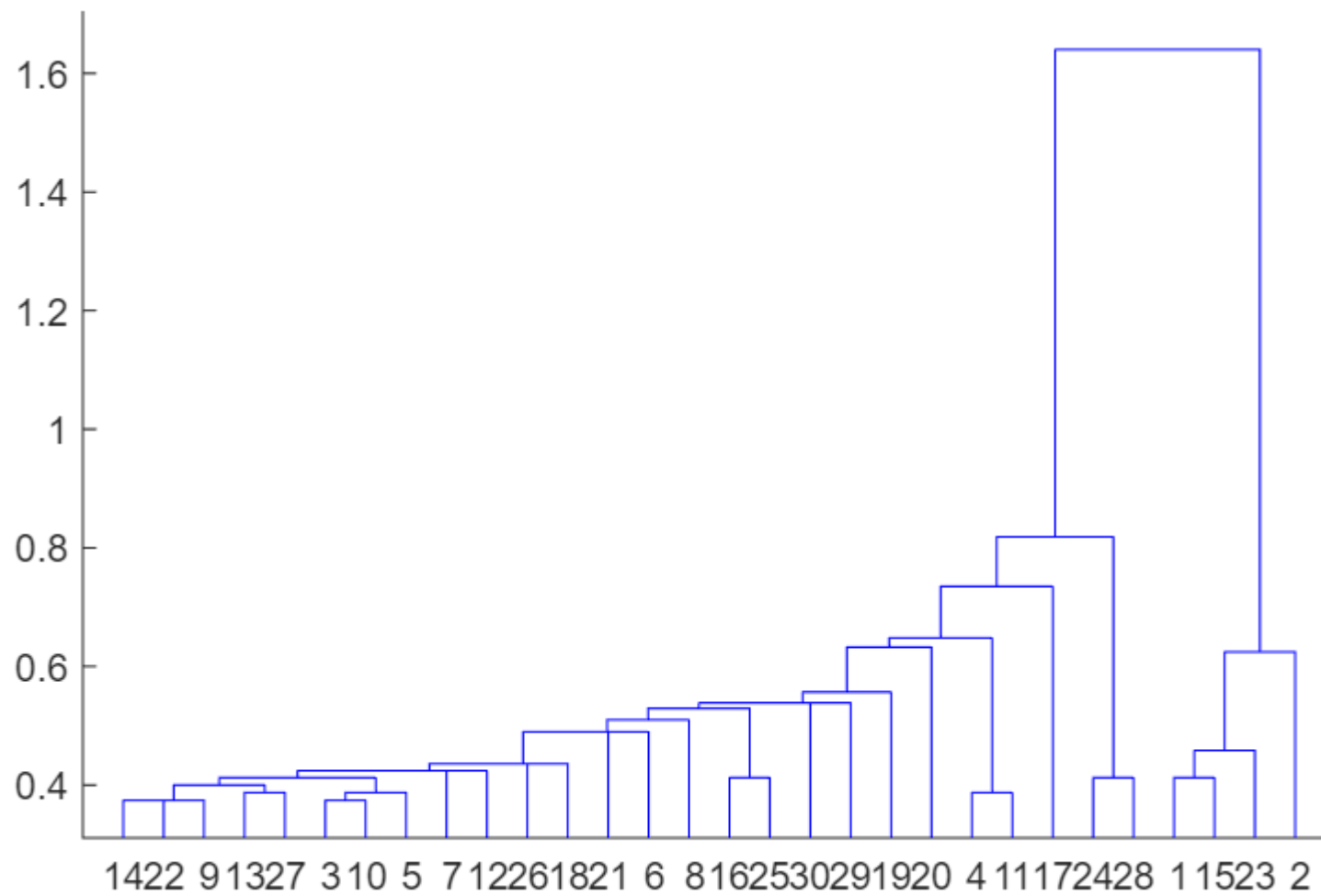# Lab14c.m – Perform hierarchical clustering

close all;

clear all;

clc;

load fisheriris;

D = pdist(meas);

Z = linkage(D);

Lab14c.m – Visualize dendrogram

close all;

clear all;

clc;

load fisheriris;

D = pdist(meas);

Z = linkage(D);

figure, dendrogram(Z);

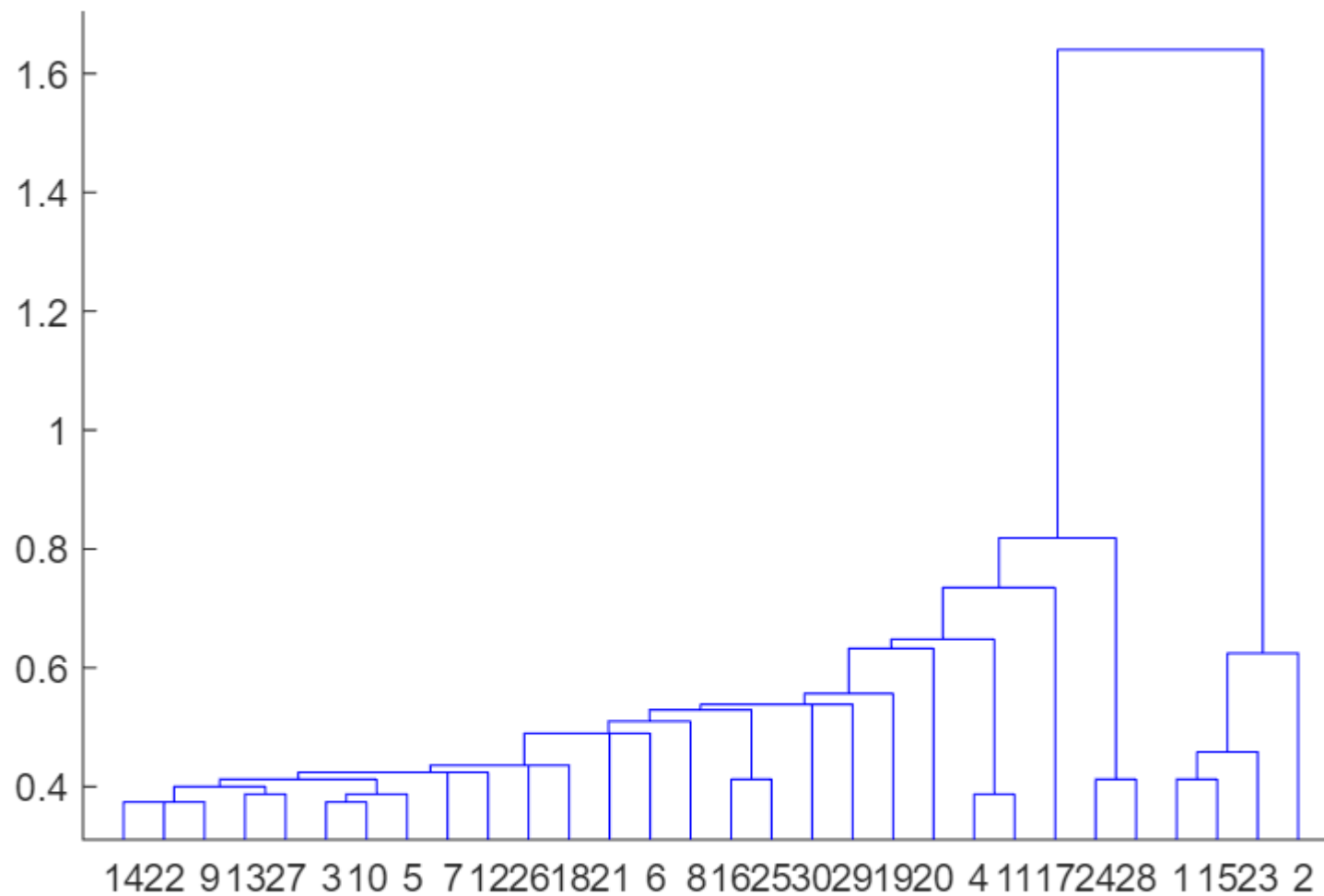# Lab14c.m-Another solution


```
close all;

clear all;

clc;

load fisheriris;

Y = pdist(meas);

Z = linkage(D, 'complete');

figure, dendrogram(Z);

Z2 = linkage(meas,'single',{'minkowski',2});;

figure, dendrogram(Z2);
```

Q&A