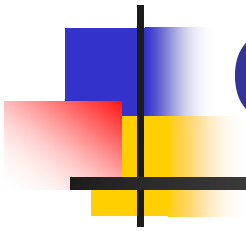
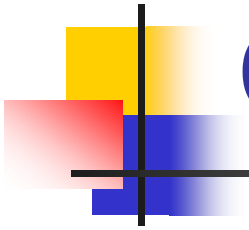


COMP1942

Clustering 2
(Hierarchical Clustering Methods)

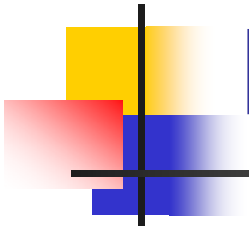
Prepared by Raymond Wong
Some parts of this notes are borrowed from LW Chan's notes
XLMiner screenshots captured by Hao Liu
Presented by Raymond Wong
raywong@cse



Clustering Methods

- K-means Clustering
  - Original k-means Clustering
  - Sequential K-means Clustering
  - Forgetful Sequential K-means Clustering
  - How to use the data mining tool
- Hierarchical Clustering Methods
  - Agglomerative methods
  - Divisive methods – polythetic approach and monothetic approach
  - How to use the data mining tool

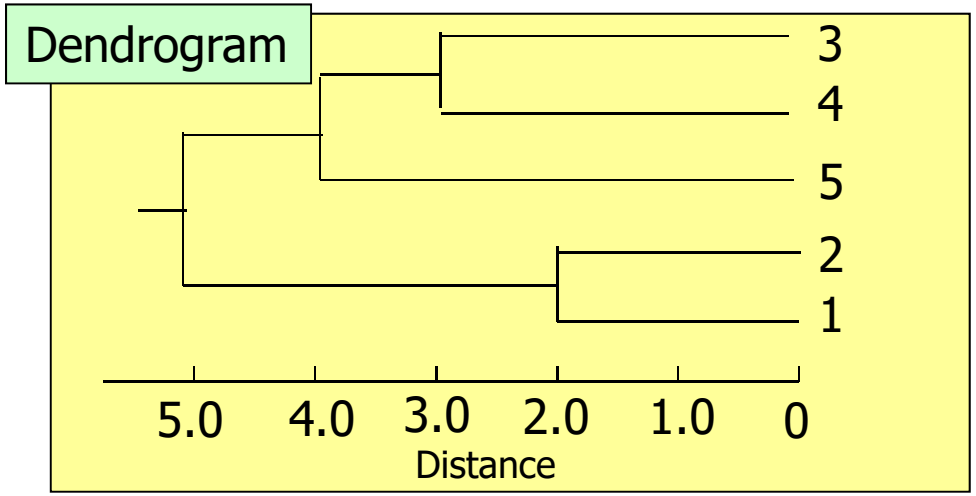
Hierarchical Clustering Methods

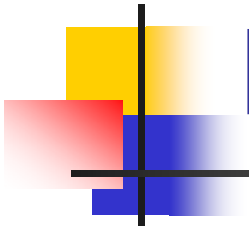


- The partition of data is not done at a single step.
- There are two varieties of hierarchical clustering algorithms
  - Agglomerative – successively fusions of the data into groups
  - Divisive – separate the data successively into finer groups

Dendrogram

- Hierarchic grouping can be represented by two-dimensional diagram known as a **dendrogram**.



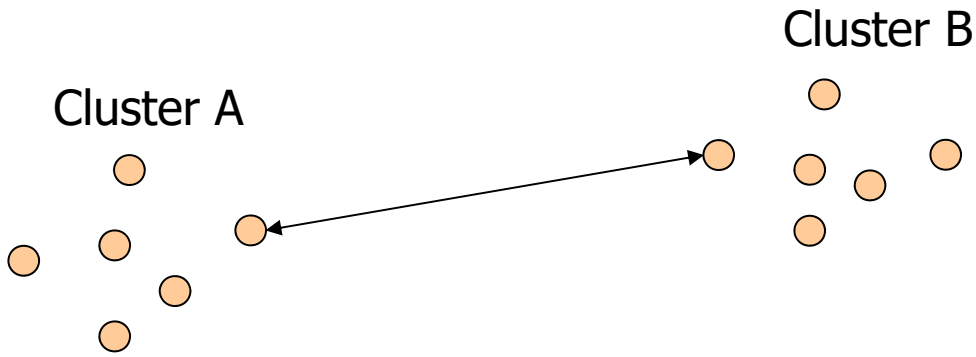


Distance

- Single Linkage
- Complete Linkage
- Group Average Linkage
- Centroid Linkage
- Median Linkage

Single Linkage

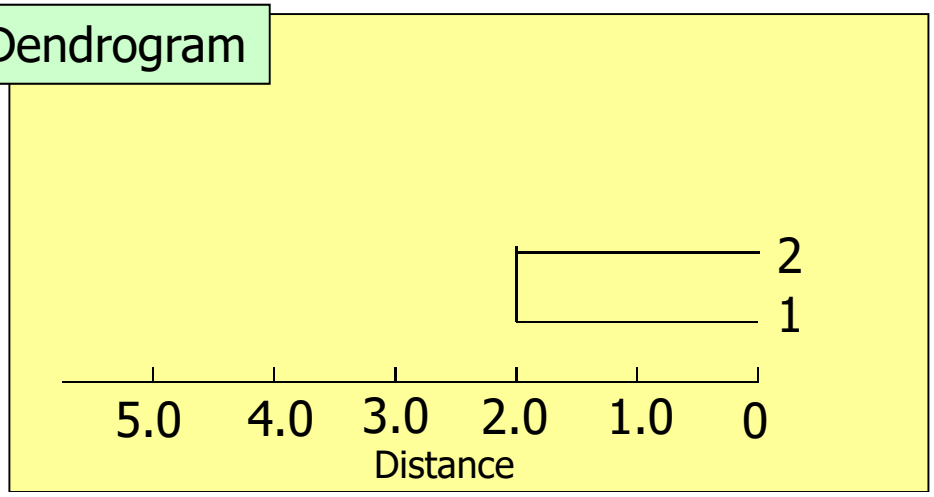
- Also, known as the **nearest neighbor** technique
- Distance between groups is defined as that of the closest pair of data, where only pairs consisting of one record from each group are considered



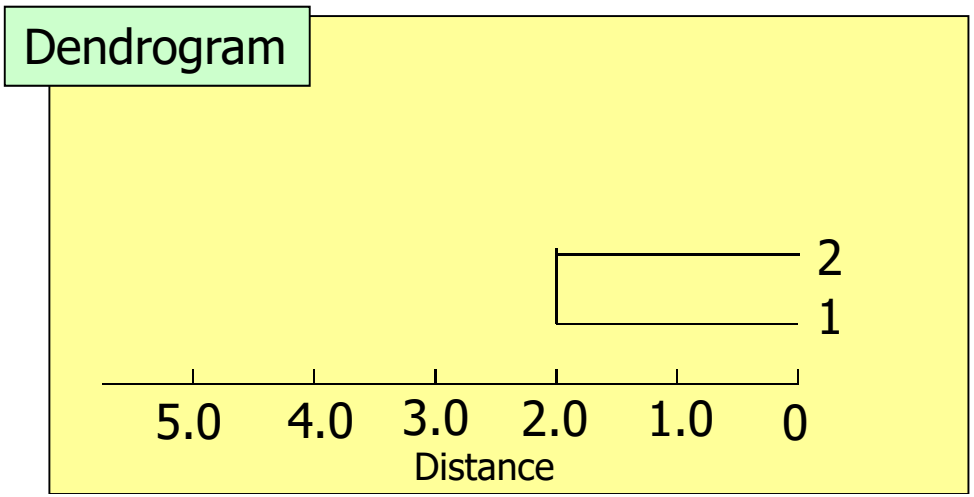
	1	2	3	4	5
1	0.0				
2	2.0	0.0			
3	6.0	5.0	0.0		
4	10.0	9.0	4.0	0.0	
5	9.0	8.0	5.0	3.0	0.0

	(12)	3	4	5
(12)	0.0			
3	5.0	0.0		
4	9.0	4.0	0.0	
5	8.0	5.0	3.0	0.0

Dendrogram

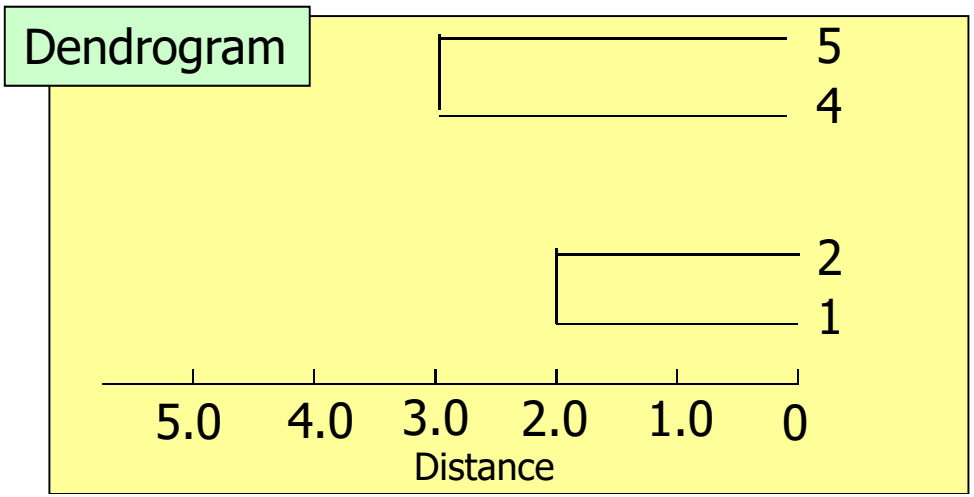


	(12)	3	4	5
(12)	0.0			
3	5.0	0.0		
4	9.0	4.0	0.0	
5	8.0	5.0	3.0	0.0

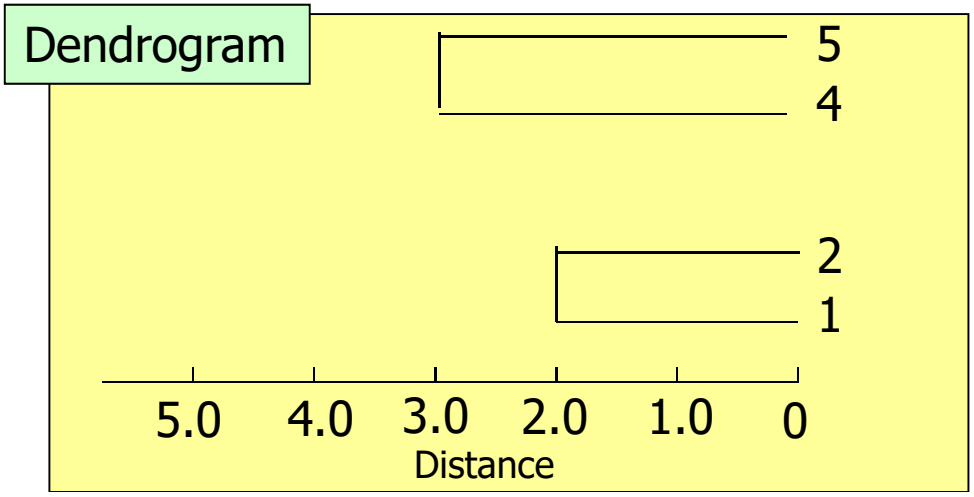


	(12)	3	4	5
(12)	0.0			
3	5.0	0.0		
4	9.0	4.0	0.0	
5	8.0	5.0	3.0	0.0

	(12)	3	(4 5)
(12)	0.0		
3	5.0	0.0	
(4 5)	8.0	4.0	0.0

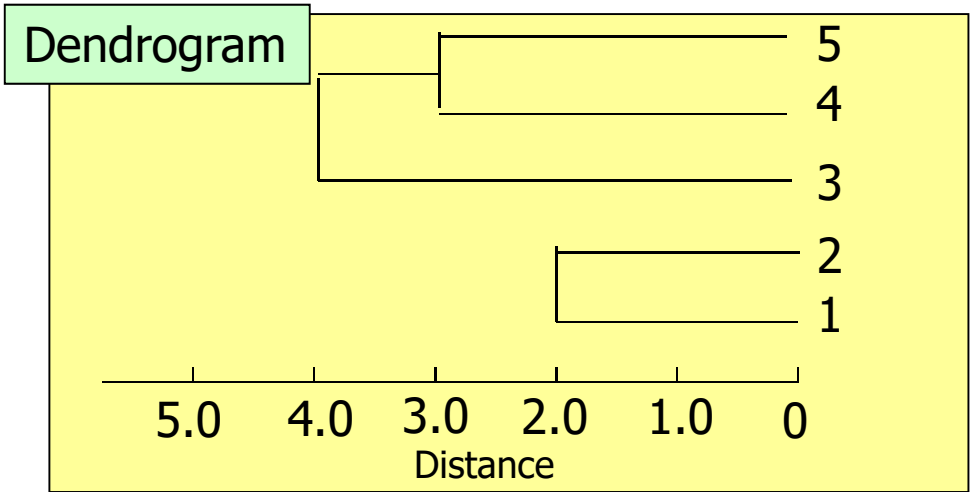


$$\begin{array}{c} \\ (12) \\ 3 \\ (4\ 5) \end{array} \begin{array}{c} \begin{array}{ccc} (12) & 3 & (4\ 5) \end{array} \\ \begin{pmatrix} 0.0 & & \\ 5.0 & 0.0 & \\ 8.0 & 4.0 & 0.0 \end{pmatrix} \end{array}$$

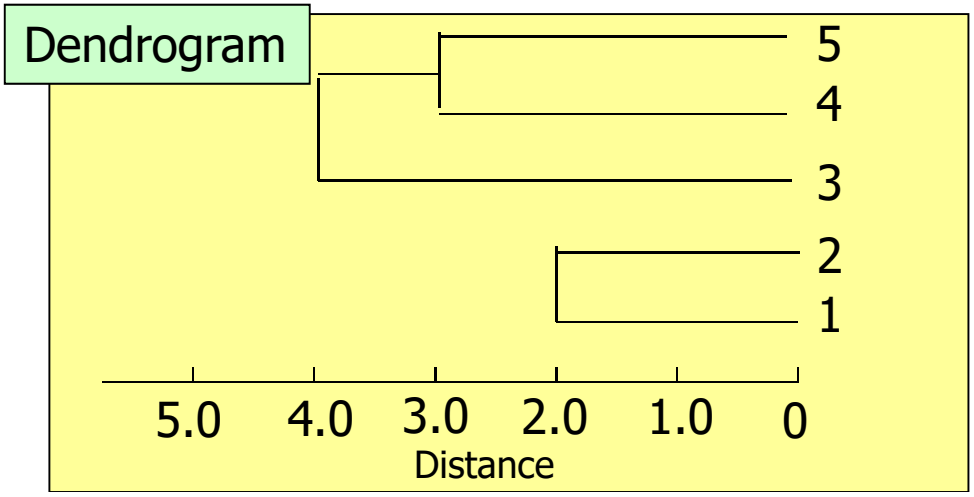


$$\begin{array}{c} \\ (12) \\ 3 \\ (4\ 5) \end{array} \begin{array}{c} \begin{array}{ccc} (12) & 3 & (4\ 5) \end{array} \\ \begin{pmatrix} 0.0 & & \\ 5.0 & 0.0 & \\ 8.0 & 4.0 & 0.0 \end{pmatrix} \end{array}$$

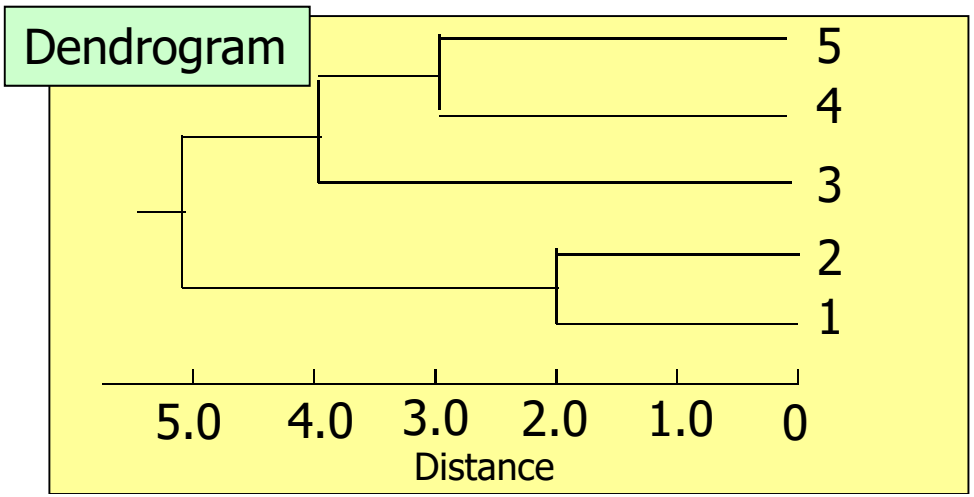
$$\begin{array}{c} \\ (12) \\ (3\ 4\ 5) \end{array} \begin{array}{c} \begin{array}{cc} (12) & (3\ 4\ 5) \end{array} \\ \begin{pmatrix} 0.0 & \\ 5.0 & 0.0 \end{pmatrix} \end{array}$$

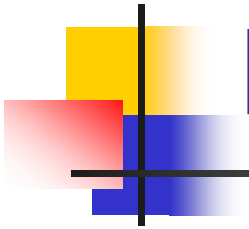


$$\begin{array}{c} \\ (12) \\ (3\ 4\ 5) \end{array} \begin{array}{c} \begin{array}{cc} (12) & (3\ 4\ 5) \end{array} \\ \begin{pmatrix} 0.0 & \\ 5.0 & 0.0 \end{pmatrix} \end{array}$$



	(12)	(3 4 5)
(12)	0.0	
(3 4 5)	5.0	0.0



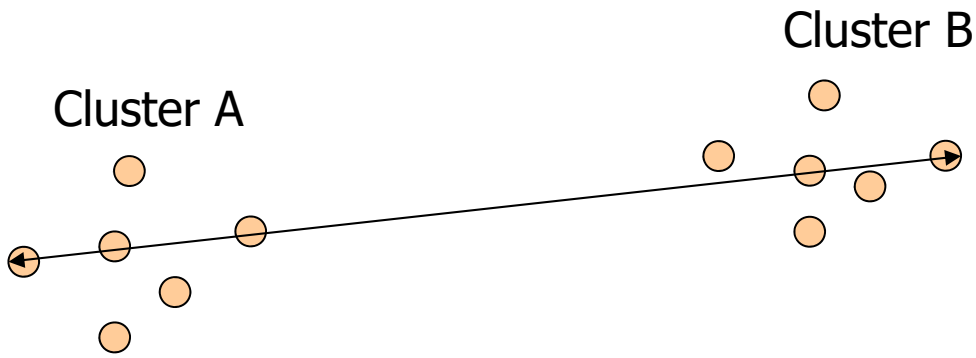


Distance

- Single Linkage
- Complete Linkage
- Group Average Linkage
- Centroid Linkage
- Median Linkage

Complete Linkage

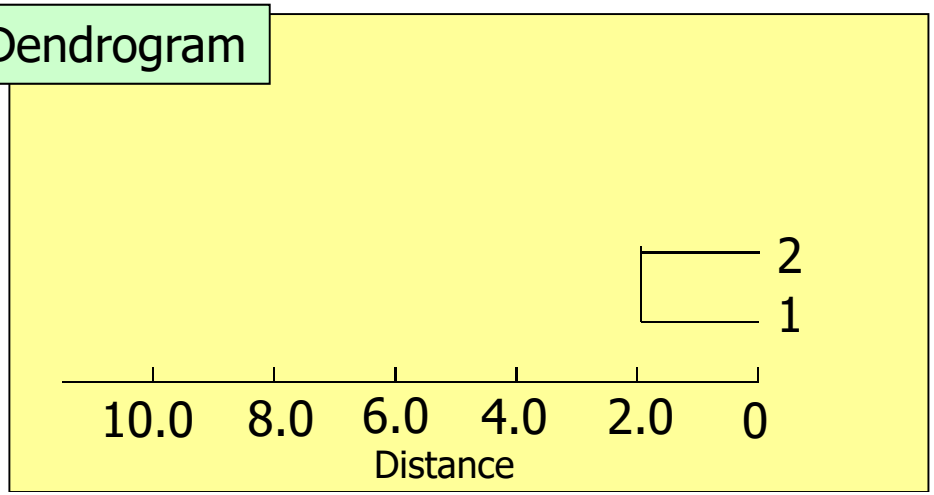
- The distance between two clusters is given by the distance between their most distant members



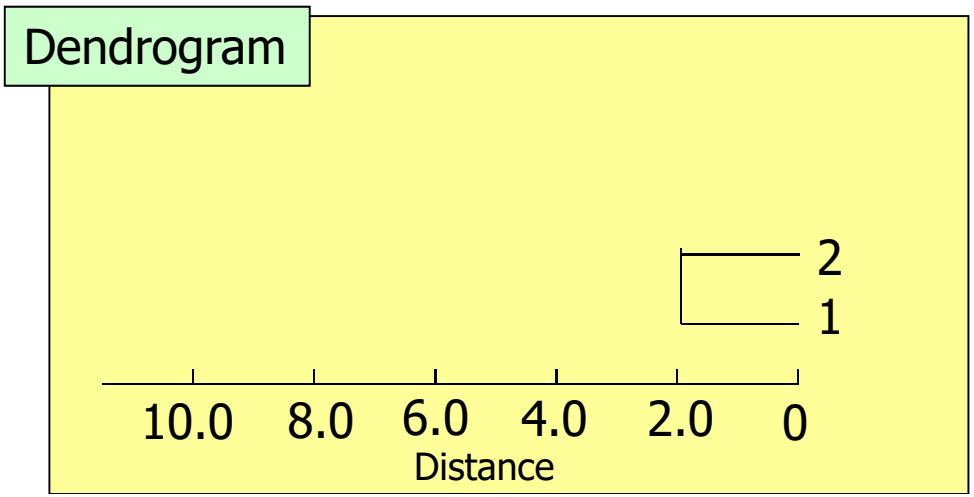
	1	2	3	4	5
1	0.0				
2	2.0	0.0			
3	6.0	5.0	0.0		
4	10.0	9.0	4.0	0.0	
5	9.0	8.0	5.0	3.0	0.0

	(12)	3	4	5
(12)	0.0			
3	6.0	0.0		
4	10.0	4.0	0.0	
5	9.0	5.0	3.0	0.0

Dendrogram

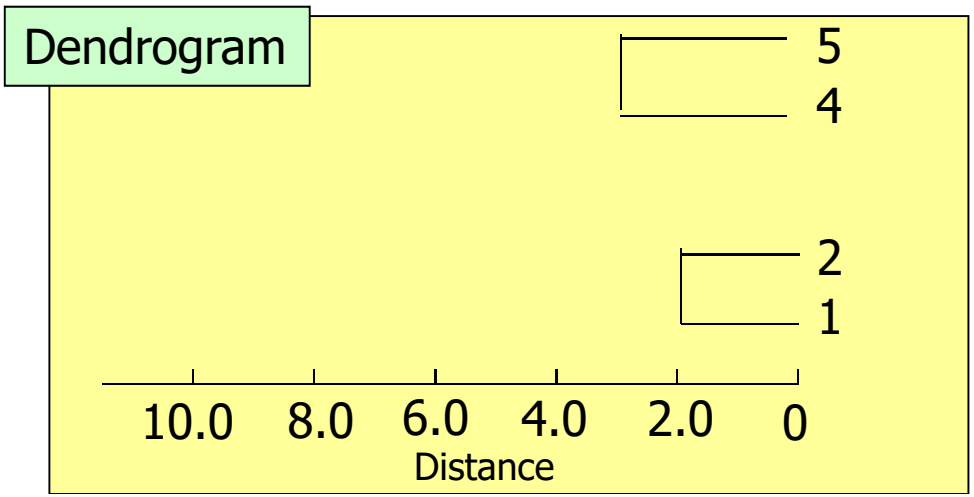


$$
\begin{array}{c} \\ (12) \\ 3 \\ 4 \\ 5 \end{array}
\begin{array}{c}
\begin{array}{cccc} (12) & 3 & 4 & 5 \end{array} \\
\begin{pmatrix}
0.0 & & & \\
6.0 & 0.0 & & \\
10.0 & 4.0 & 0.0 & \\
9.0 & 5.0 & 3.0 & 0.0
\end{pmatrix}
\end{array}
$$

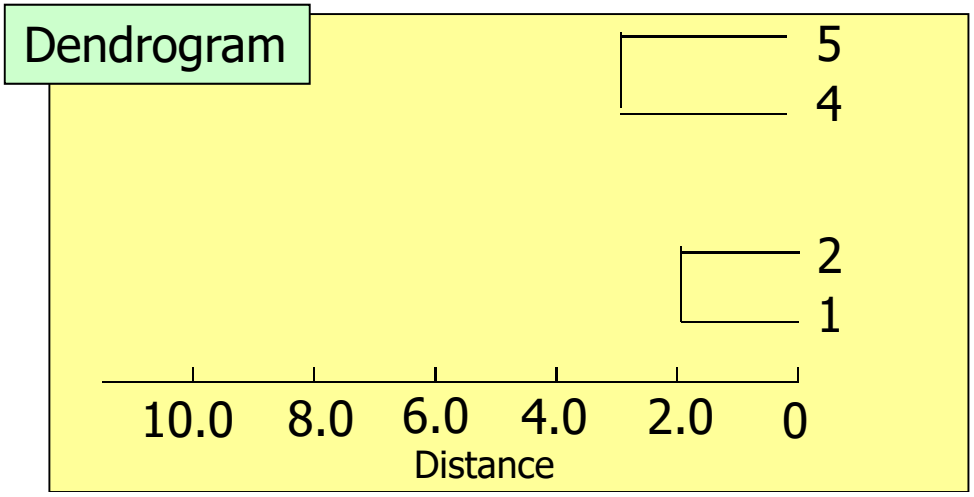


	(12)	3	4	5
(12)	0.0			
3	6.0	0.0		
4	10.0	4.0	0.0	
5	9.0	5.0	3.0	0.0

	(12)	3	(4 5)
(12)	0.0		
3	6.0	0.0	
(4 5)	10.0	5.0	0.0

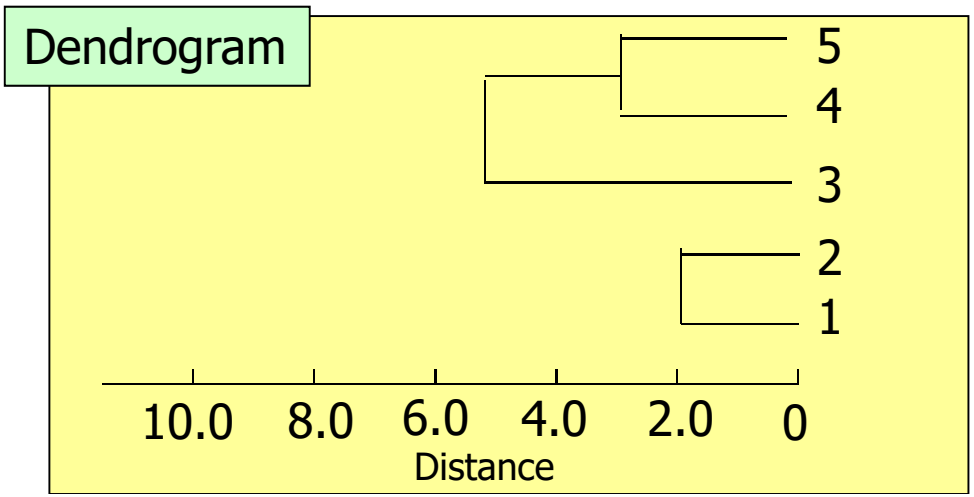


$$\begin{array}{c} \\ (12) \\ 3 \\ (4\ 5) \end{array} \begin{array}{c} \begin{array}{ccc} (12) & 3 & (4\ 5) \end{array} \\ \begin{pmatrix} 0.0 & & \\ 6.0 & 0.0 & \\ 10.0 & 5.0 & 0.0 \end{pmatrix} \end{array}$$

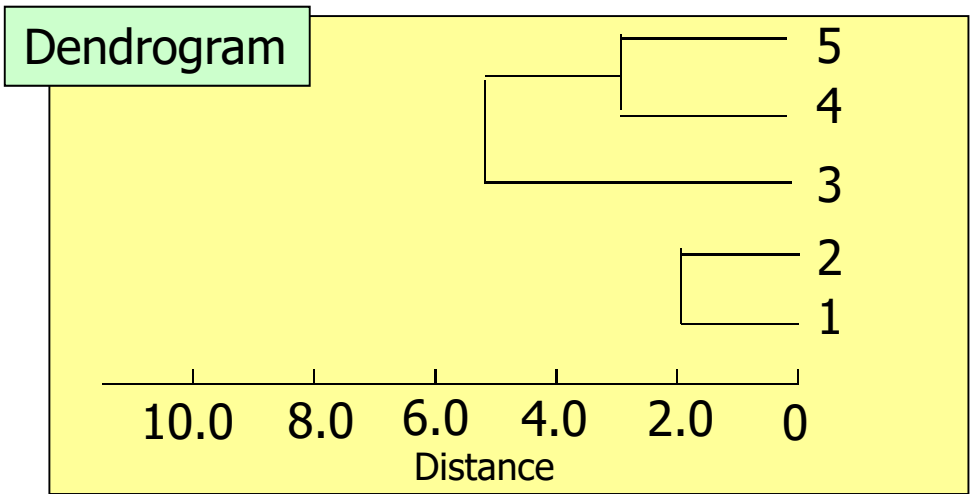


	(12)	3	(4 5)
(12)	0.0		
3	6.0	0.0	
(4 5)	10.0	5.0	0.0

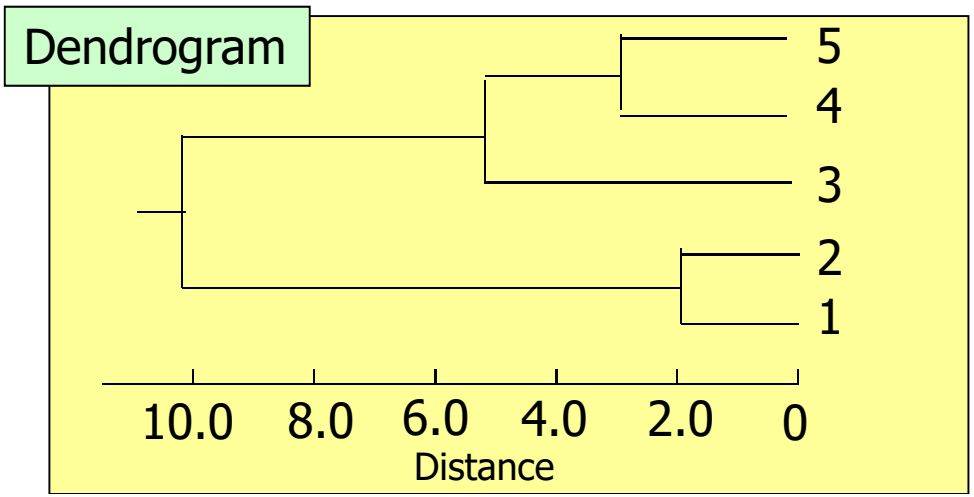
	(12)	(3 4 5)
(12)	0.0	
(3 4 5)	10.0	0.0

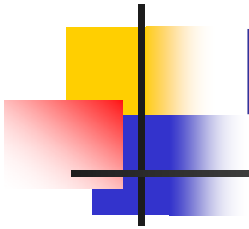


$$\begin{array}{c} \\ (12) \\ (3\ 4\ 5) \end{array} \begin{array}{c} \begin{array}{cc} (12) & (3\ 4\ 5) \end{array} \\ \begin{pmatrix} 0.0 & \\ 10.0 & 0.0 \end{pmatrix} \end{array}$$



	(12)	(3 4 5)
(12)	0.0	
(3 4 5)	10.0	0.0



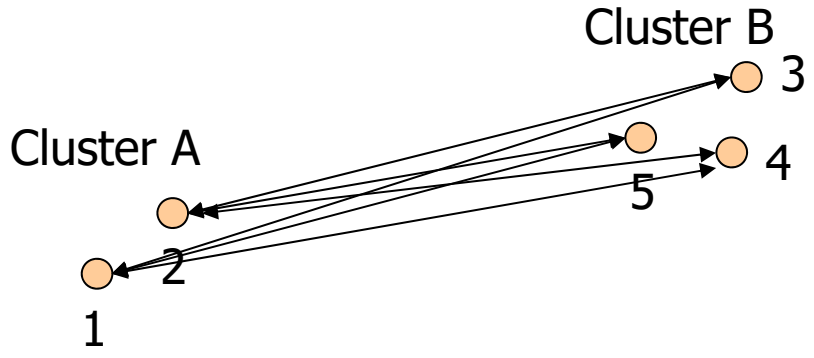


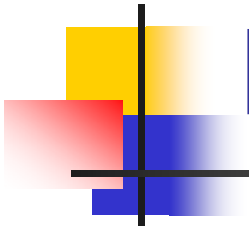
Distance

- Single Linkage
- Complete Linkage
- Group Average Linkage
- Centroid Linkage
- Median Linkage

Group Average Clustering

- The distance between two clusters is defined as the average of the distances between all pairs of records (one from each cluster).
- $d_{AB} = 1/6\ (d_{13} + d_{14} + d_{15} + d_{23} + d_{24} + d_{25})$



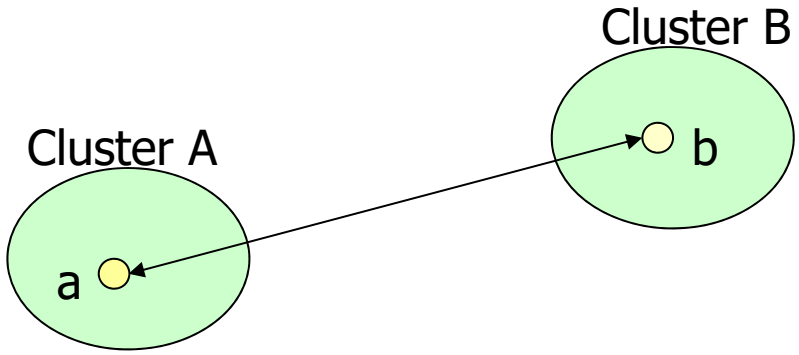


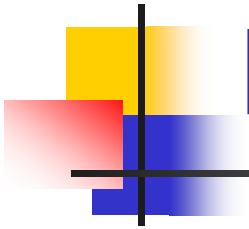
Distance

- Single Linkage
- Complete Linkage
- Group Average Linkage
- Centroid Linkage
- Median Linkage

Centroid Clustering

- The distance between two clusters is defined as the distance between the mean vectors of the two clusters.
- $d_{AB} = d_{ab}$
- where a is the mean vector of the cluster A and b is the mean vector of the cluster B.



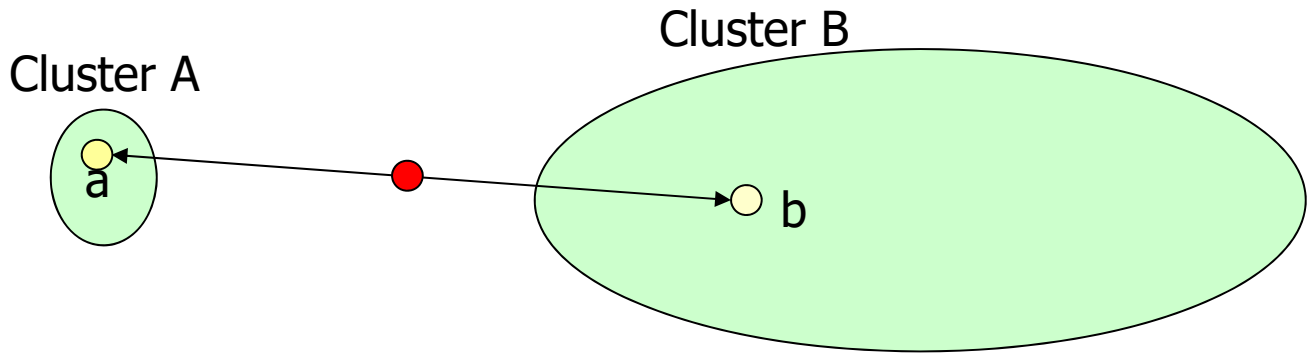


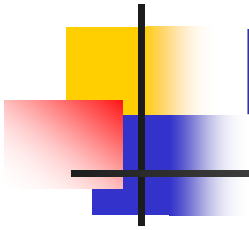
Distance

- Single Linkage
- Complete Linkage
- Group Average Linkage
- Centroid Linkage
- Median Linkage

Median Clustering

- Disadvantage of the Centroid Clustering: When a large cluster is merged with a small one, the centroid of the combined cluster would be closed to the large one, ie. The characteristic properties of the small one are lost
- After we have combined two groups, the mid-point of the original two cluster centres is used as the centre of the newly combined group

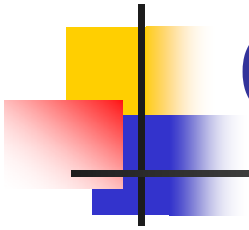




Distance

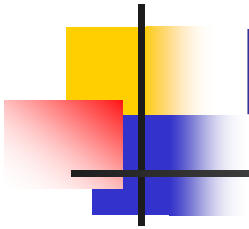
- Single Linkage
- Complete Linkage
- Group Average Linkage
- Centroid Linkage
- Median Linkage

There are other distance measurements (e.g., McQuitty's linkage/method and Ward's linkage/method).
We will not discuss them.



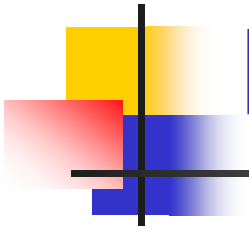
Clustering Methods

- K-means Clustering
  - Original k-means Clustering
  - Sequential K-means Clustering
  - Forgetful Sequential K-means Clustering
  - How to use the data mining tool
- Hierarchical Clustering Methods
  - Agglomerative methods
  - Divisive methods – polythetic approach and monothetic approach
  - How to use the data mining tool



Divisive Methods

- In a divisive algorithm, we start with the assumption that all the data is part of one cluster.
- We then use a distance criterion to divide the cluster in two, and then subdivide the clusters until a stopping criterion is achieved.
  - Polythetic – divide the data based on the values by all attributes
  - Monothetic – divide the data on the basis of the possession of a single specified attribute



Polythetic Approach

- Distance
  - Single Linkage
  - Complete Linkage
  - Group Average Linkage
  - Centroid Linkage
  - Median Linkage

Polythetic Approach

	1	2	3	4	5	6	7
1	0						
2	10	0					
3	7	7	0				
4	30	23	21	0			
5	29	25	22	7	0		
6	38	34	31	10	11	0	
7	42	36	36	13	17	9	0

A = {1 }

B = {2, 3, 4, 5, 6, 7}

D(1, *) = 26.0

D(2, *) = 22.5

D(3, *) = 20.7

D(4, *) = 17.3

D(5, *) = 18.5

D(6, *) = 22.2

D(7, *) = 25.5

Polythetic Approach

	1	2	3	4	5	6	7
1	0						
2	10	0					
3	7	7	0				
4	30	23	21	0			
5	29	25	22	7	0		
6	38	34	31	10	11	0	
7	42	36	36	13	17	9	0

D(2, A) = 10

D(3, A) = 7

D(4, A) = 30

D(5, A) = 29

D(6, A) = 38

D(7, A) = 42

A = {1 }

B = {2, 3, 4, 5, 6, 7}

Polythetic Approach

	1	2	3	4	5	6	7
1	0						
2	10	0					
3	7	7	0				
4	30	23	21	0			
5	29	25	22	7	0		
6	38	34	31	10	11	0	
7	42	36	36	13	17	9	0

D(2, A) = 10 | D(2, B) = 25.0

D(3, A) = 7 | D(3, B) = 23.4

D(4, A) = 30 | D(4, B) = 14.8

D(5, A) = 29 | D(5, B) = 16.4

D(6, A) = 38 | D(6, B) = 19.0

D(7, A) = 42 | D(7, B) = 22.2

A = {1 }

B = {2, 3, 4, 5, 6, 7}

Polythetic Approach

	1	2	3	4	5	6	7
1	0						
2	10	0					
3	7	7	0				
4	30	23	21	0			
5	29	25	22	7	0		
6	38	34	31	10	11	0	
7	42	36	36	13	17	9	0

D(2, A) = 10 | D(2, B) = 25.0 | Δ_2 = 15.0

D(3, A) = 7 | D(3, B) = 23.4 | Δ_3 = 16.4

D(4, A) = 30 | D(4, B) = 14.8 | Δ_4 = -15.2

D(5, A) = 29 | D(5, B) = 16.4 | Δ_5 = -12.6

D(6, A) = 38 | D(6, B) = 19.0 | Δ_6 = -19.0

D(7, A) = 42 | D(7, B) = 22.2 | Δ_7 = -19.8

A = {1, 3 }

B = {2, 4, 5, 6, 7}

Polythetic Approach

	1	2	3	4	5	6	7
1	0						
2	10	0					
3	7	7	0				
4	30	23	21	0			
5	29	25	22	7	0		
6	38	34	31	10	11	0	
7	42	36	36	13	17	9	0

A = {1, 3 }

B = {2, 4, 5, 6, 7}

D(2, A) = 10	D(2, B) = 25.0	Δ_2 = 15.0
D(3, A) = 7	D(3, B) = 23.4	Δ_3 = 16.4
D(4, A) = 30	D(4, B) = 14.8	Δ_4 = -15.2
D(5, A) = 29	D(5, B) = 16.4	Δ_5 = -12.6
D(6, A) = 38	D(6, B) = 19.0	Δ_6 = -19.0
D(7, A) = 42	D(7, B) = 22.2	Δ_7 = -19.8

Polythetic Approach

	1	2	3	4	5	6	7
1	0						
2	10	0					
3	7	7	0				
4	30	23	21	0			
5	29	25	22	7	0		
6	38	34	31	10	11	0	
7	42	36	36	13	17	9	0

D(2, A) = 8.5

D(4, A) = 25.5

D(5, A) = 25.5

D(6, A) = 34.5

D(7, A) = 39.0

A = {1, 3 }

B = {2, 4, 5, 6, 7}

Polythetic Approach

	1	2	3	4	5	6	7
1	0						
2	10	0					
3	7	7	0				
4	30	23	21	0			
5	29	25	22	7	0		
6	38	34	31	10	11	0	
7	42	36	36	13	17	9	0

D(2, A) = 8.5

D(2, B) = 29.5

D(4, A) = 25.5

D(4, B) = 13.2

D(5, A) = 25.5

D(5, B) = 15.0

D(6, A) = 34.5

D(6, B) = 16.0

D(7, A) = 39.0

D(7, B) = 18.75

A = {1, 3 }

B = {2, 4, 5, 6, 7}

Polythetic Approach

	1	2	3	4	5	6	7
1	0						
2	10	0					
3	7	7	0				
4	30	23	21	0			
5	29	25	22	7	0		
6	38	34	31	10	11	0	
7	42	36	36	13	17	9	0

D(2, A) = 8.5 | D(2, B) = 29.5 | Δ_2 = 21.0

D(4, A) = 25.5 | D(4, B) = 13.2 | Δ_4 = -12.3

D(5, A) = 25.5 | D(5, B) = 15.0 | Δ_5 = -10.5

D(6, A) = 34.5 | D(6, B) = 16.0 | Δ_6 = -18.5

D(7, A) = 39.0 | D(7, B) = 18.75 | Δ_7 = -20.25

A = {1, 3 , 2 }

B = {4, 5, 6, 7}

Polythetic Approach

	1	2	3	4	5	6	7
1	0						
2	10	0					
3	7	7	0				
4	30	23	21	0			
5	29	25	22	7	0		
6	38	34	31	10	11	0	
7	42	36	36	13	17	9	0

D(2, A) = 8.5 | D(2, B) = 29.5 | Δ_2 = 21.0

D(4, A) = 25.5 | D(4, B) = 13.2 | Δ_4 = -12.3

D(5, A) = 25.5 | D(5, B) = 15.0 | Δ_5 = -10.5

D(6, A) = 34.5 | D(6, B) = 16.0 | Δ_6 = -18.5

D(7, A) = 39.0 | D(7, B) = 18.75 | Δ_7 = -20.25

A = {1, 3 , 2 }

B = { 4, 5, 6, 7}

Polythetic Approach

	1	2	3	4	5	6	7
1	0						
2	10	0					
3	7	7	0				
4	30	23	21	0			
5	29	25	22	7	0		
6	38	34	31	10	11	0	
7	42	36	36	13	17	9	0

D(4, A) = 24.7

D(5, A) = 25.3

D(6, A) = 34.3

D(7, A) = 38.0

A = {1, 3 , 2 }

B = {4, 5, 6, 7}

Polythetic Approach

	1	2	3	4	5	6	7
1	0						
2	10	0					
3	7	7	0				
4	30	23	21	0			
5	29	25	22	7	0		
6	38	34	31	10	11	0	
7	42	36	36	13	17	9	0

D(4, A) = 24.7 D(4, B) = 10.0

D(5, A) = 25.3 D(5, B) = 11.7

D(6, A) = 34.3 D(6, B) = 10.0

D(7, A) = 38.0 D(7, B) = 13.0

A = {1, 3 , 2 }

B = {4, 5, 6, 7}

Polythetic Approach

	1	2	3	4	5	6	7
1	0						
2	10	0					
3	7	7	0				
4	30	23	21	0			
5	29	25	22	7	0		
6	38	34	31	10	11	0	
7	42	36	36	13	17	9	0

D(4, A) = 24.7 | D(4, B) = 10.0 | Δ_4 = -14.7

D(5, A) = 25.3 | D(5, B) = 11.7 | Δ_5 = -13.6

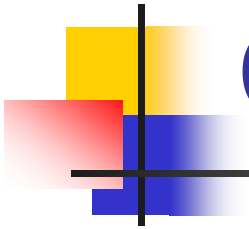
D(6, A) = 34.3 | D(6, B) = 10.0 | Δ_6 = -24.3

D(7, A) = 38.0 | D(7, B) = 13.0 | Δ_7 = -25.0

A = {1, 3 , 2 }

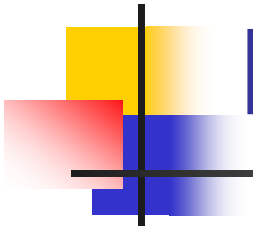
B = {4, 5, 6, 7}

All differences are negative. The process would continue on each subgroup separately.



Clustering Methods

- K-means Clustering
  - Original k-means Clustering
  - Sequential K-means Clustering
  - Forgetful Sequential K-means Clustering
  - How to use the data mining tool
- Hierarchical Clustering Methods
  - Agglomerative methods
  - Divisive methods – polythetic approach and monothetic approach
  - How to use the data mining tool



Monothetic

It is usually used when the data consists of **binary** variables.

	A	B	C
1	0	1	1
2	1	1	0
3	1	1	1
4	1	1	0
5	0	0	1

Monothetic

It is usually used when the data consists of **binary** variables.

	A	B	C
1	0	1	1
2	1	1	0
3	1	1	1
4	1	1	0
5	0	0	1

B \ A	1	0
1	a=3	b=1
0	c=0	d=1

Chi-Square Measure

$$\chi_{AB}^{\ 2} = \frac{(ad-bc)^2 N}{(a+b)(a+c)(b+d)(c+d)}$$

$$= \frac{(3-0)^2 \cdot 5}{4 \cdot 3 \cdot 2 \cdot 1}$$

$$= 1.875$$

etic

B \ A	1	0
1	a=3	b=1
0	c=0	d=1

Chi-Square Measure

It is usually used when the data consists of **binary** variables.

	A	B	C
1	0	1	1
2	1	1	0
3	1	1	1
4	1	1	0
5	0	0	1

$$\chi_{AB}^{\ 2} = \frac{(ad-bc)^2 N}{(a+b)(a+c)(b+d)(c+d)}$$

$$= \frac{(3-0)^2 \cdot 5}{4 \cdot 3 \cdot 2 \cdot 1}$$

$$= 1.875$$

etic

B \ A	1	0
1	a=3	b=1
0	c=0	d=1

Attr.	AB		
a	3		
b	1		
c	0		
d	1		
N	5		
χ^2	1.87		

Chi-Square Measure

It is usually used when the data consists of **binary** variables.

	A	B	C
1	0	1	1
2	1	1	0
3	1	1	1
4	1	1	0
5	0	0	1

$$\chi_{AB}^2 = \frac{(ad-bc)^2 N}{(a+b)(a+c)(b+d)(c+d)}$$

$$= \frac{(3-0)^2 \cdot 5}{4 \cdot 3 \cdot 2 \cdot 1}$$

$$= 1.875$$

B \ A	1	0
1	a=3	b=1
0	c=0	d=1

etic

Chi-Square Measure

It is usually used when the data consists of **binary** variables.

	A	B	C
1	0	1	1
2	1	1	0
3	1	1	1
4	1	1	0
5	0	0	1

Attr.	AB	AC	BC
a	3	1	2
b	1	2	1
c	0	2	2
d	1	0	0
N	5	5	5
χ^2	1.87	2.22	0.83

For attribute A,

$$\chi_{AB}^2 + \chi_{AC}^2 = 4.09$$

For attribute B,

$$\chi_{AB}^2 + \chi_{BC}^2 = 2.70$$

For attribute C,

$$\chi_{AC}^2 + \chi_{BC}^2 = 3.05$$

B \ A	1	0
1	a=3	b=1
0	c=0	d=1

etic

Chi-Square Measure

It is usually used when the data consists of **binary** variables.

Attr.	AB	AC	BC
a	3	1	2
b	1	2	1
χ^2		2.22	0.83

We choose attribute A for dividing the data into two groups. {2, 3, 4}, and {1, 5}

	A	B	C
1	0	1	1
2	1	1	0
3	1	1	1
4	1	1	0
5	0	0	1

For attribute A,

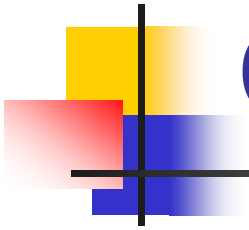
$$\chi_{AB}^2 + \chi_{AC}^2 = 4.09$$

For attribute B,

$$\chi_{AB}^2 + \chi_{BC}^2 = 2.70$$

For attribute C,

$$\chi_{AC}^2 + \chi_{BC}^2 = 3.05$$



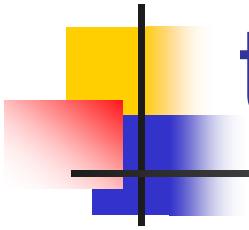
Clustering Methods

- K-means Clustering
  - Original k-means Clustering
  - Sequential K-means Clustering
  - Forgetful Sequential K-means Clustering
  - How to use the data mining tool
- Hierarchical Clustering Methods
  - Agglomerative methods
  - Divisive methods – polythetic approach and monothetic approach
  - How to use the data mining tool

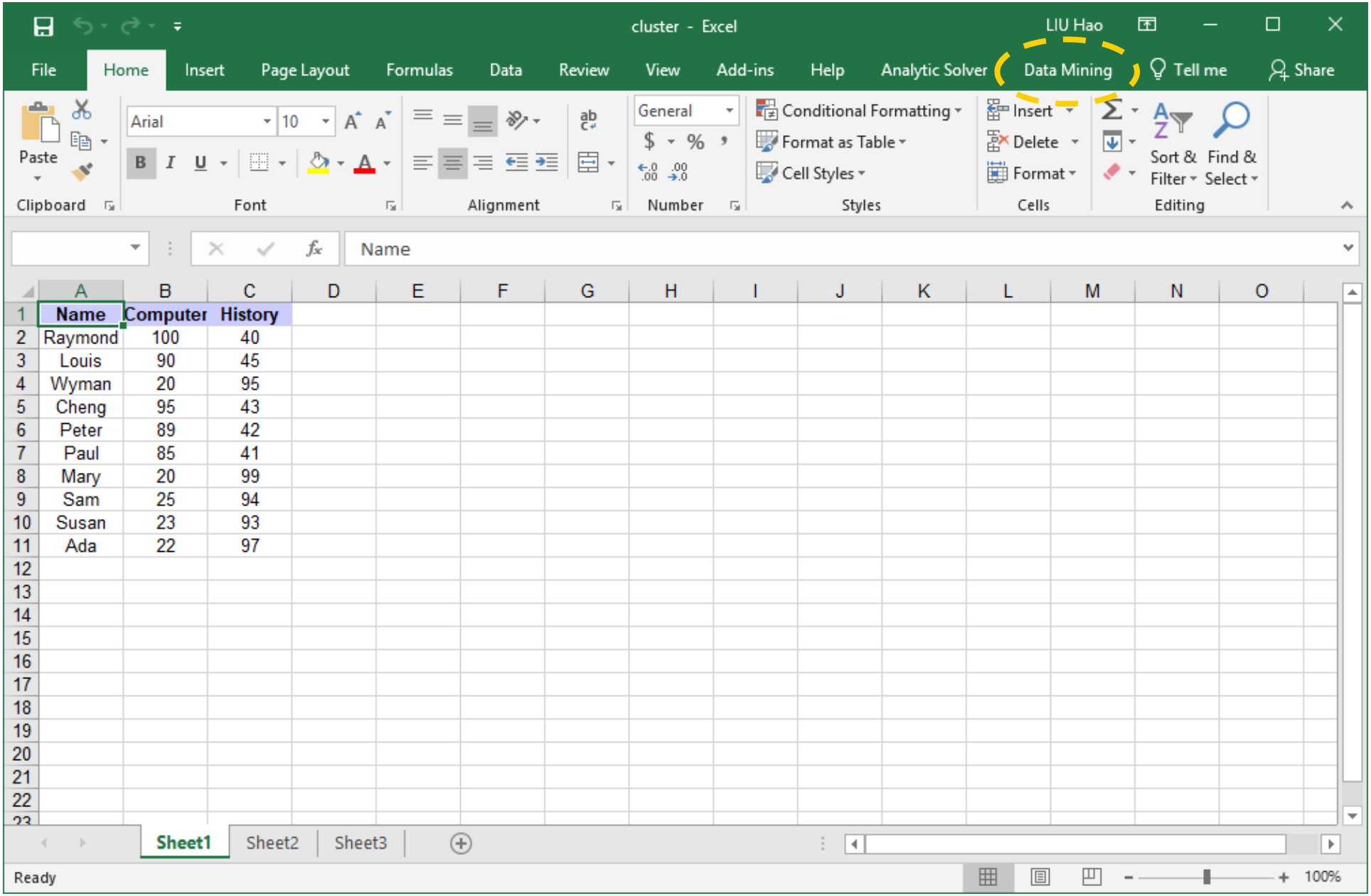
How to use the data mining tool

- We have the following 2 versions.
  - XLMiner Desktop (installed in either the CSE lab or your computer)
  - XLMiner Cloud (installed as a plugin in your Office 365 Excel)

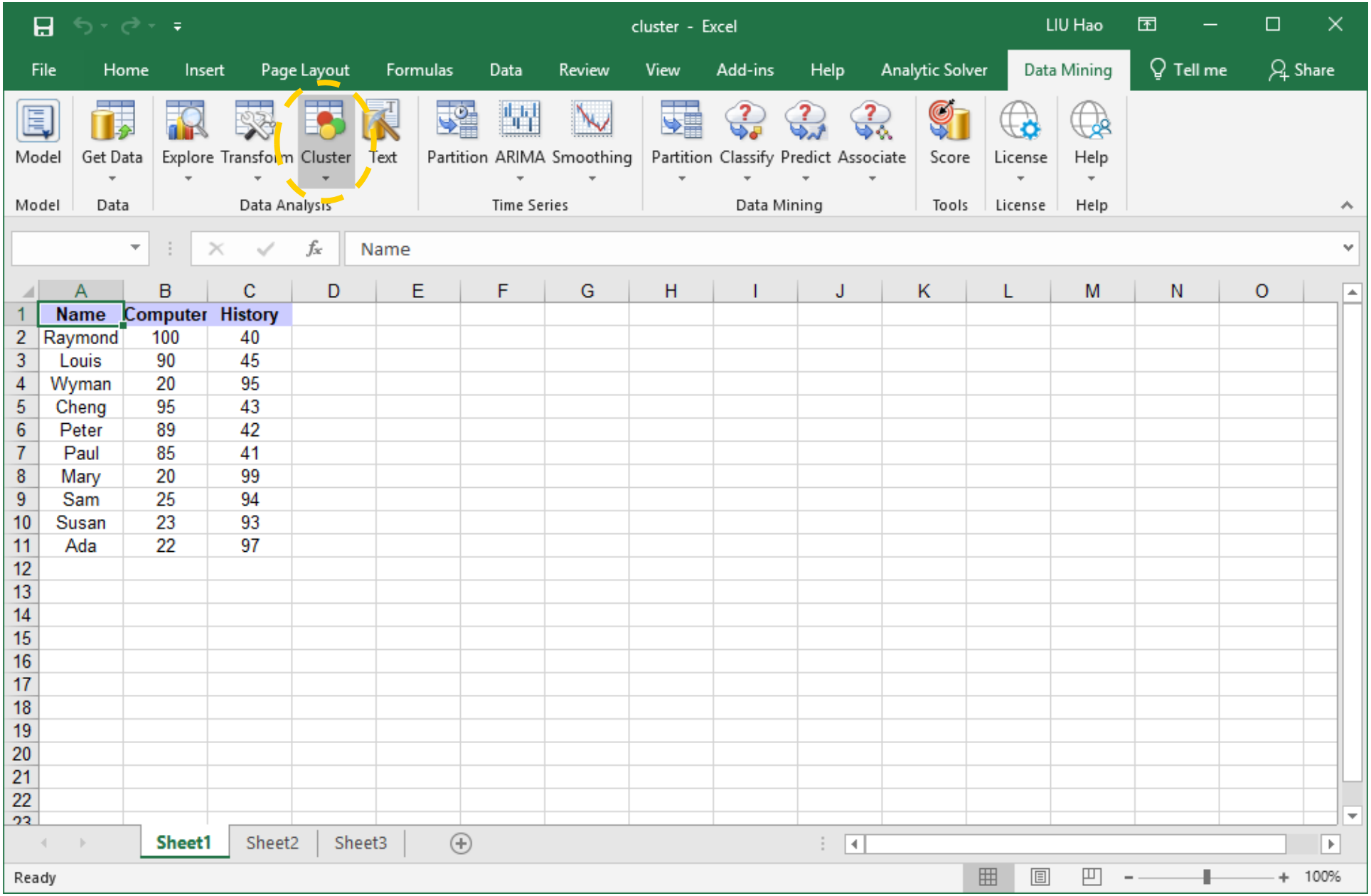
How to use the data mining tool (XLMiner Desktop)

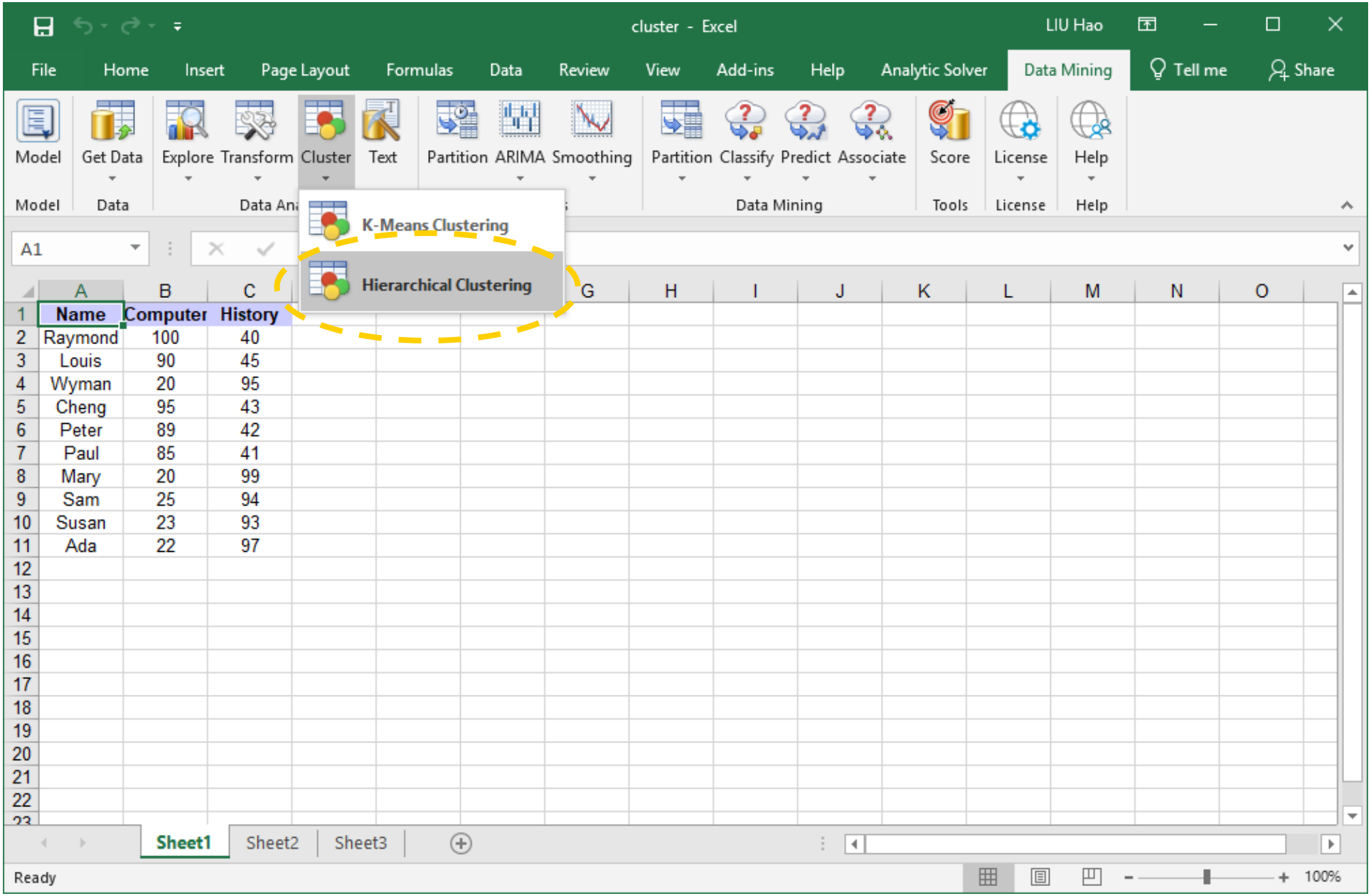


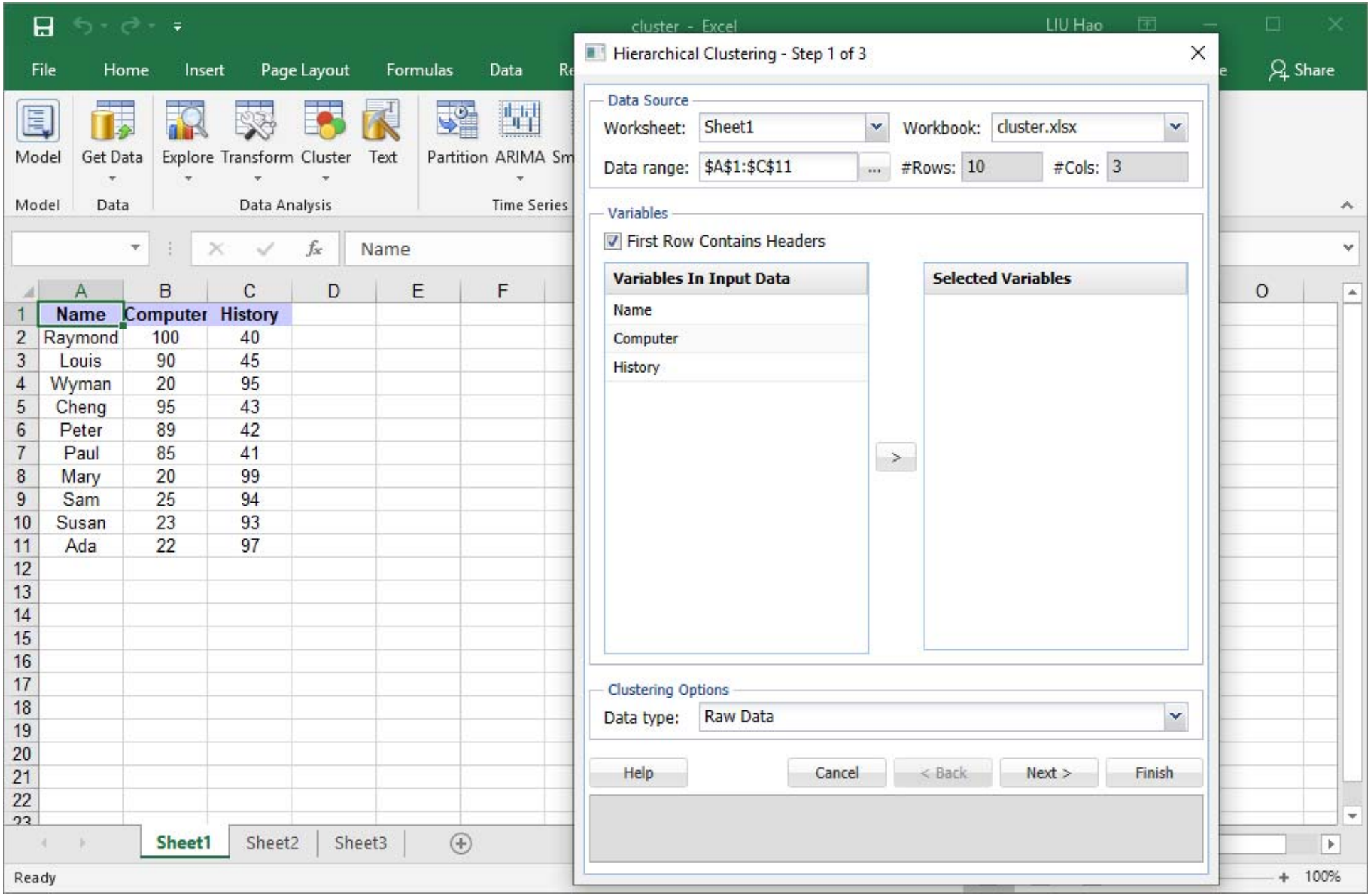
- We can use XLMiner for performing hierarchical clustering
- Open “cluster.xlsx” in MS Excel in a CSE lab machine

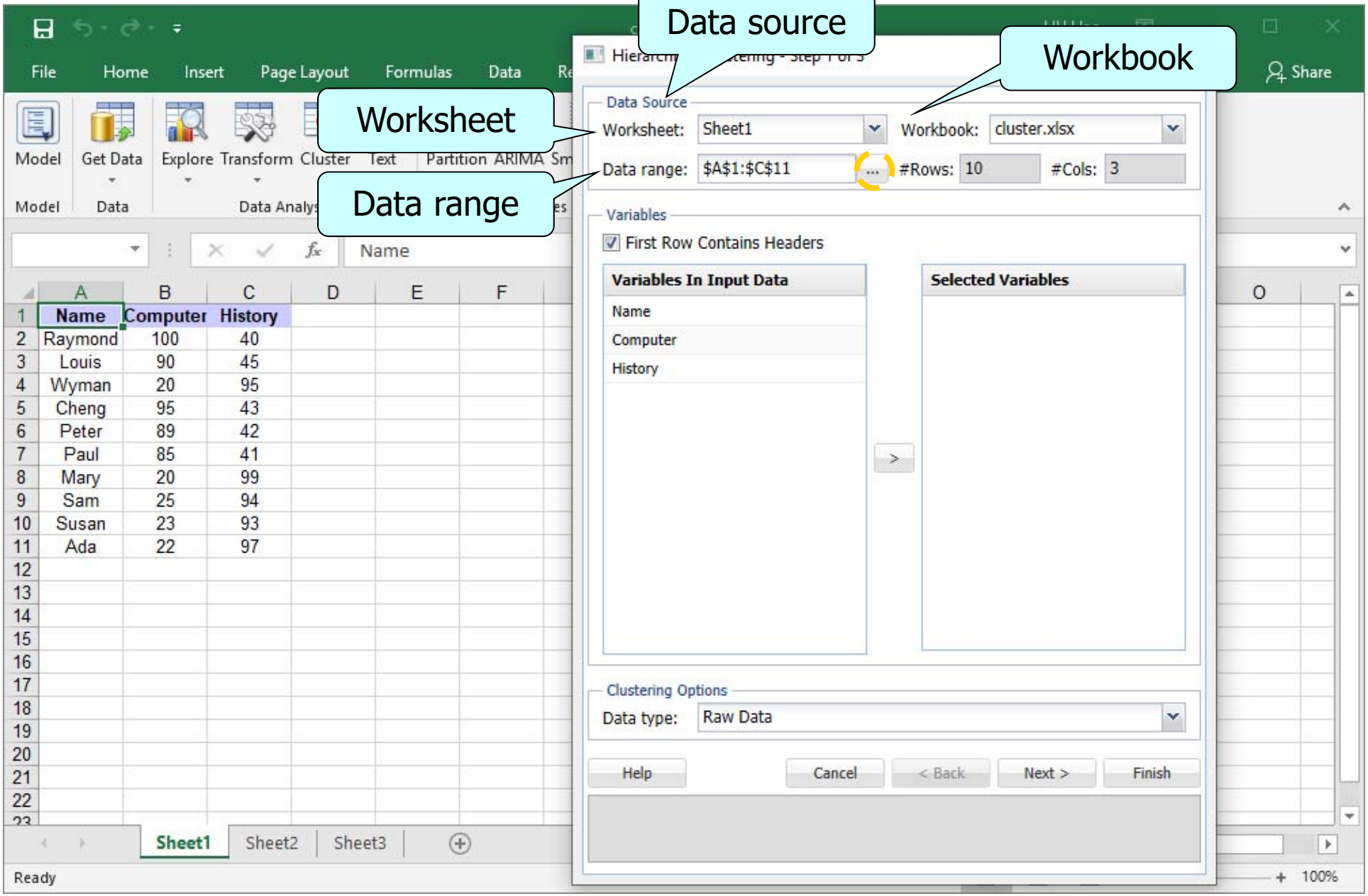


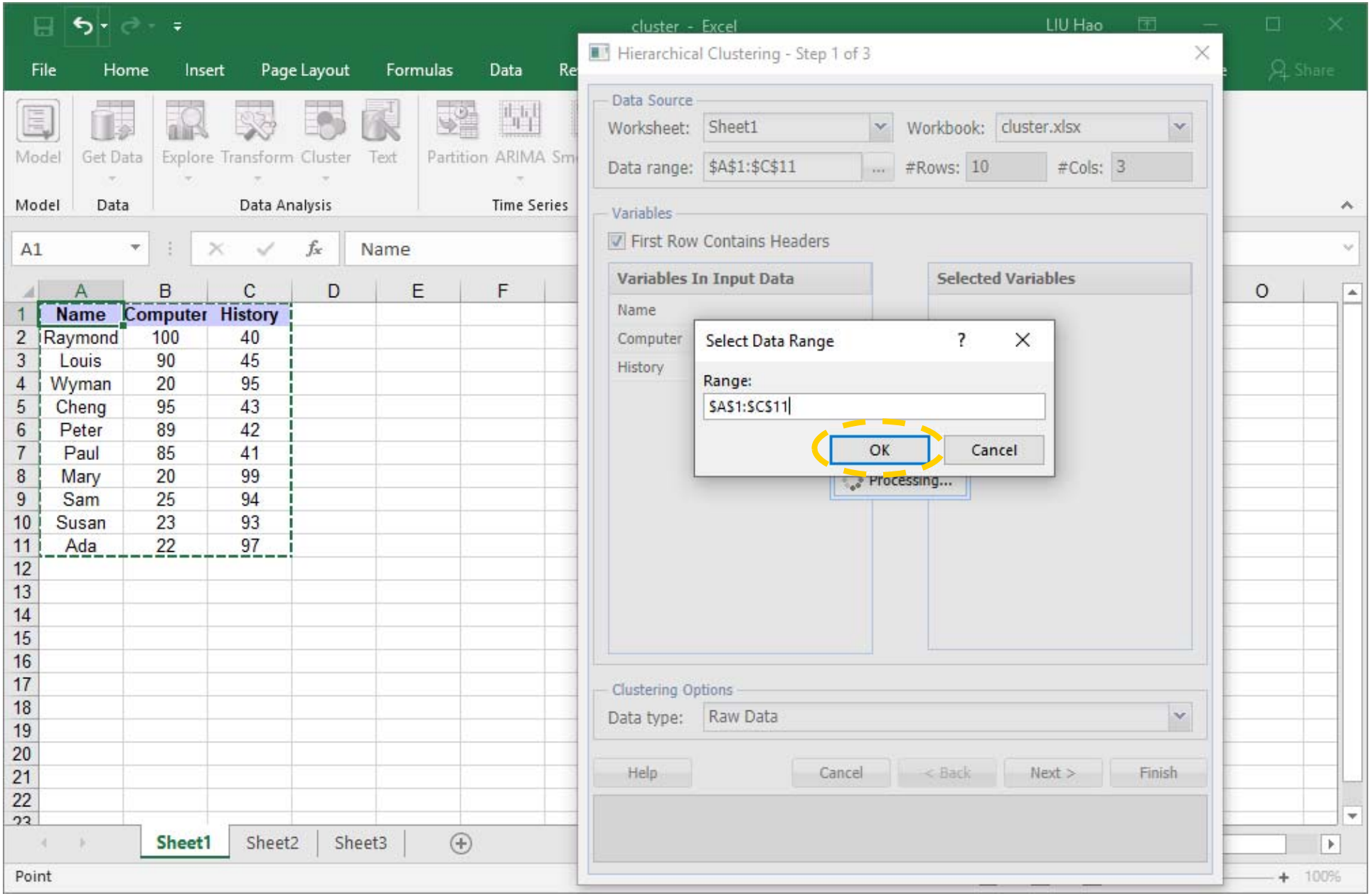
	Name	Computer	History
2	Raymond	100	40
3	Louis	90	45
4	Wyman	20	95
5	Cheng	95	43
6	Peter	89	42
7	Paul	85	41
8	Mary	20	99
9	Sam	25	94
10	Susan	23	93
11	Ada	22	97

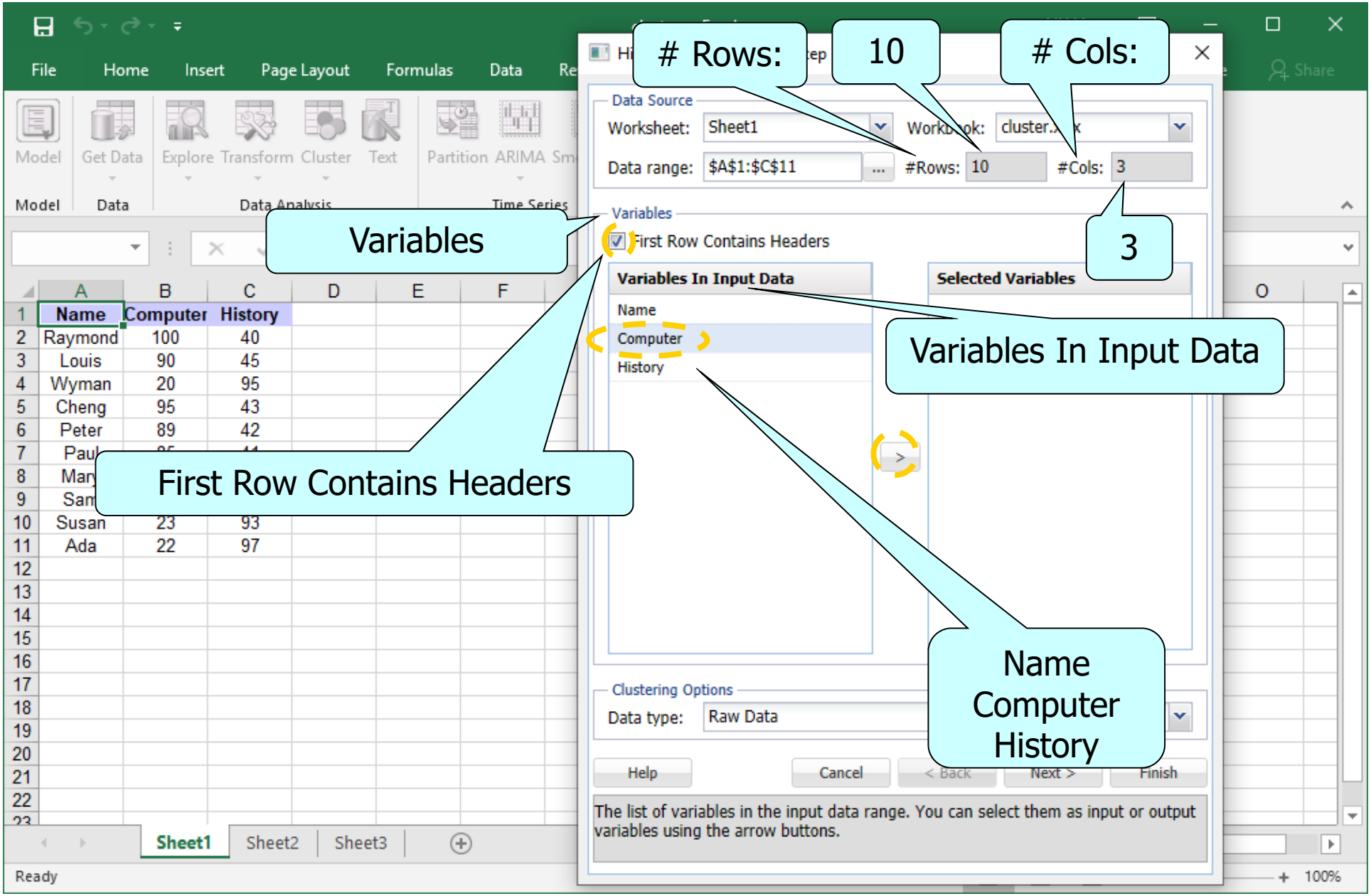

cluster - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
Model
Data
Data Analysis
Time Series
Data Mining
Tools
License
Help
Name
Name Computer History
Raymond 100 40
Louis 90 45
Wyman 20 95
Cheng 95 43
Peter 89 42
Paul 85 41
Mary 20 99
Sam 25 94
Susan 23 93
Ada 22 97
Sheet1
Sheet2
Sheet3
Ready
100%

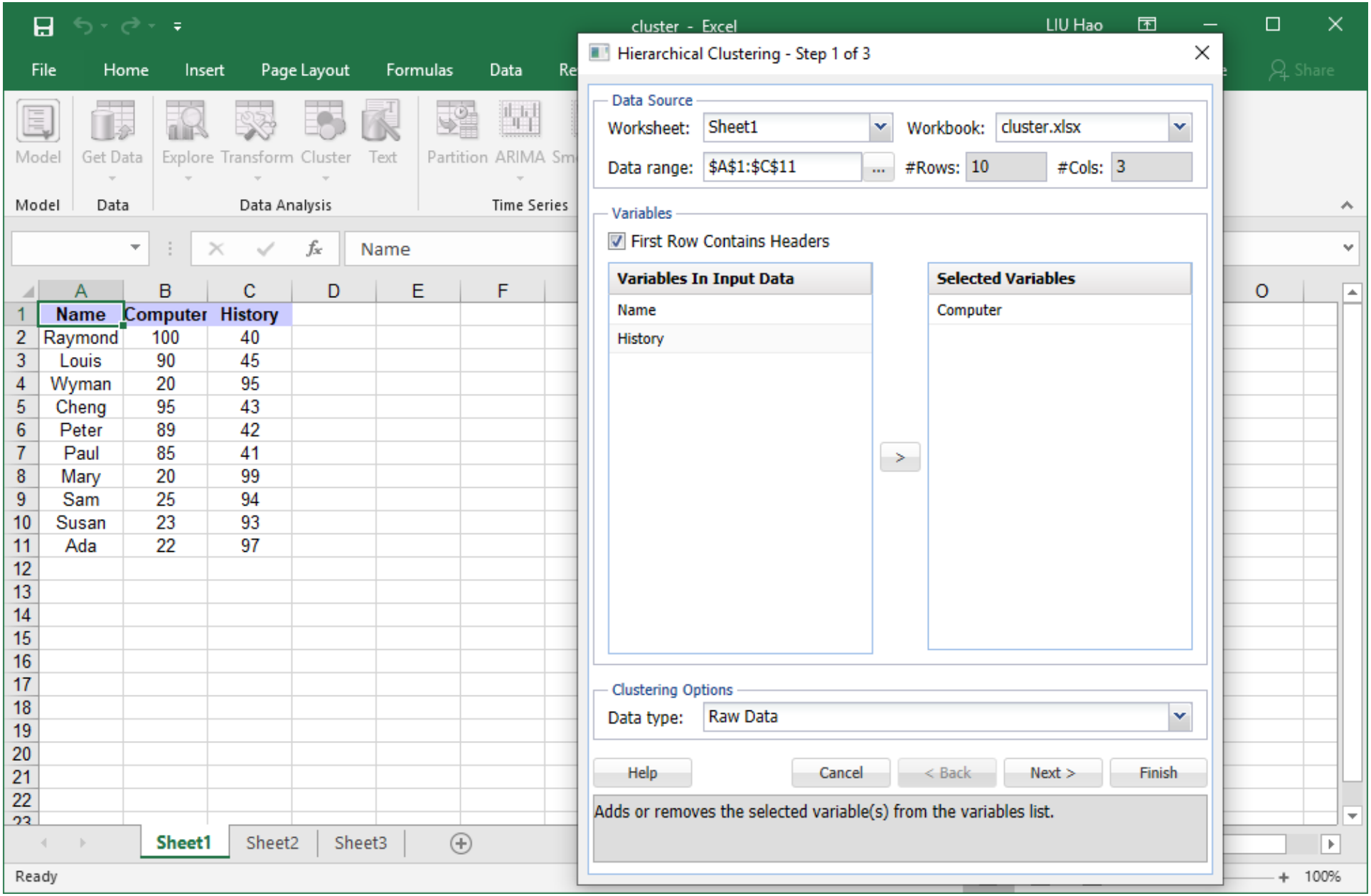

cluster - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
Model
Data
Data Mining
Tools
License
Help
K-Means Clustering
Hierarchical Clustering
A1
Name
Computer
History
Raymond 100 40
Louis 90 45
Wyman 20 95
Cheng 95 43
Peter 89 42
Paul 85 41
Mary 20 99
Sam 25 94
Susan 23 93
Ada 22 97
Sheet1
Sheet2
Sheet3
Ready
100%

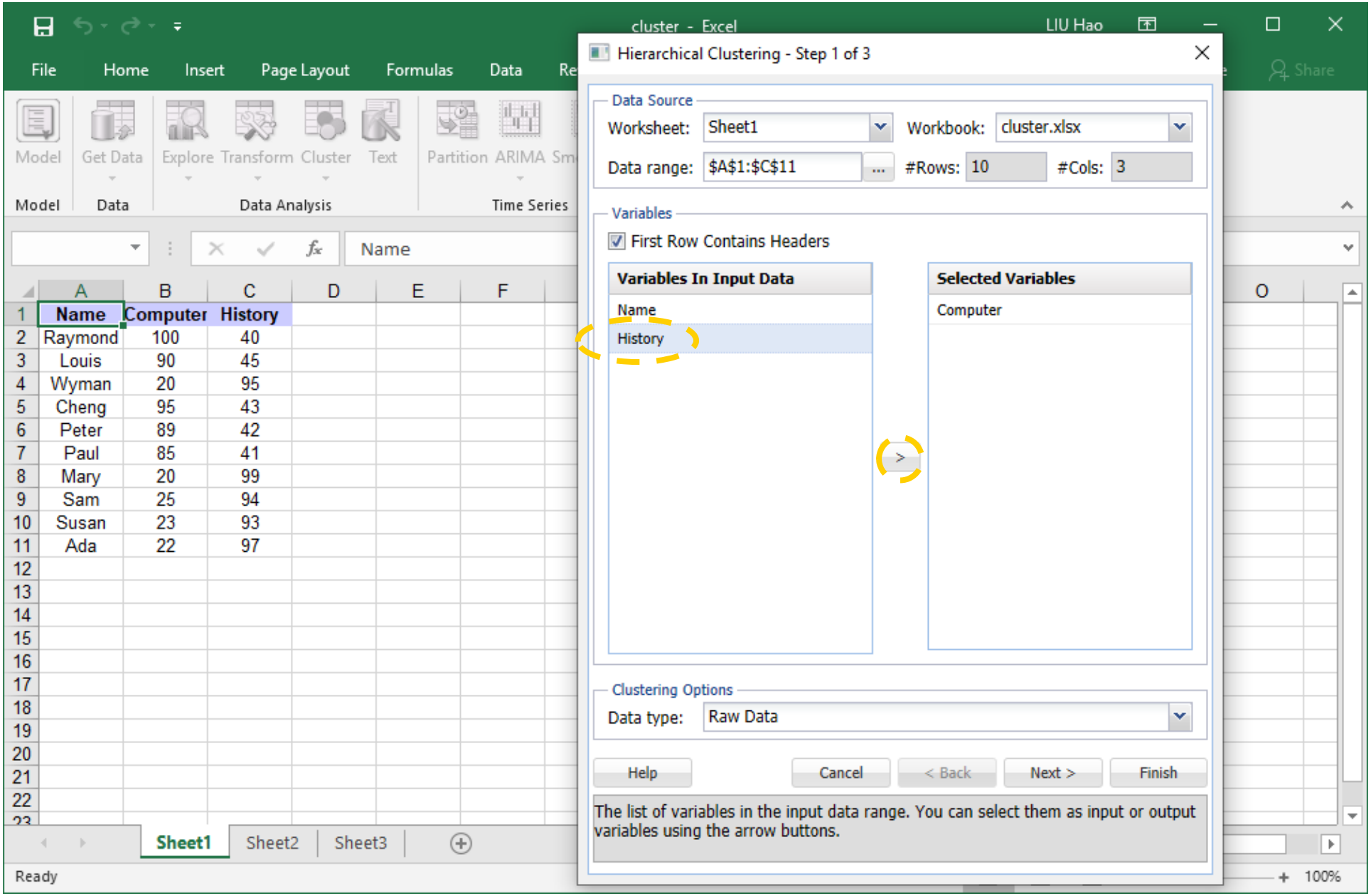

cluster - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Model
Data
Data Analysis
Time Series
Name
Name
Computer
History
Raymond 100 40
Louis 90 45
Wyman 20 95
Cheng 95 43
Peter 89 42
Paul 85 41
Mary 20 99
Sam 25 94
Susan 23 93
Ada 22 97
Sheet1
Sheet2
Sheet3
Ready
100%
Hierarchical Clustering - Step 1 of 3
Data Source
Worksheet: Sheet1
Workbook: cluster.xlsx
Data range: $A$1:$C$11
#Rows: 10
#Cols: 3
Variables
First Row Contains Headers
Variables In Input Data
Name
Computer
History
Selected Variables
>
Clustering Options
Data type: Raw Data
Help
Cancel
< Back
Next >
Finish

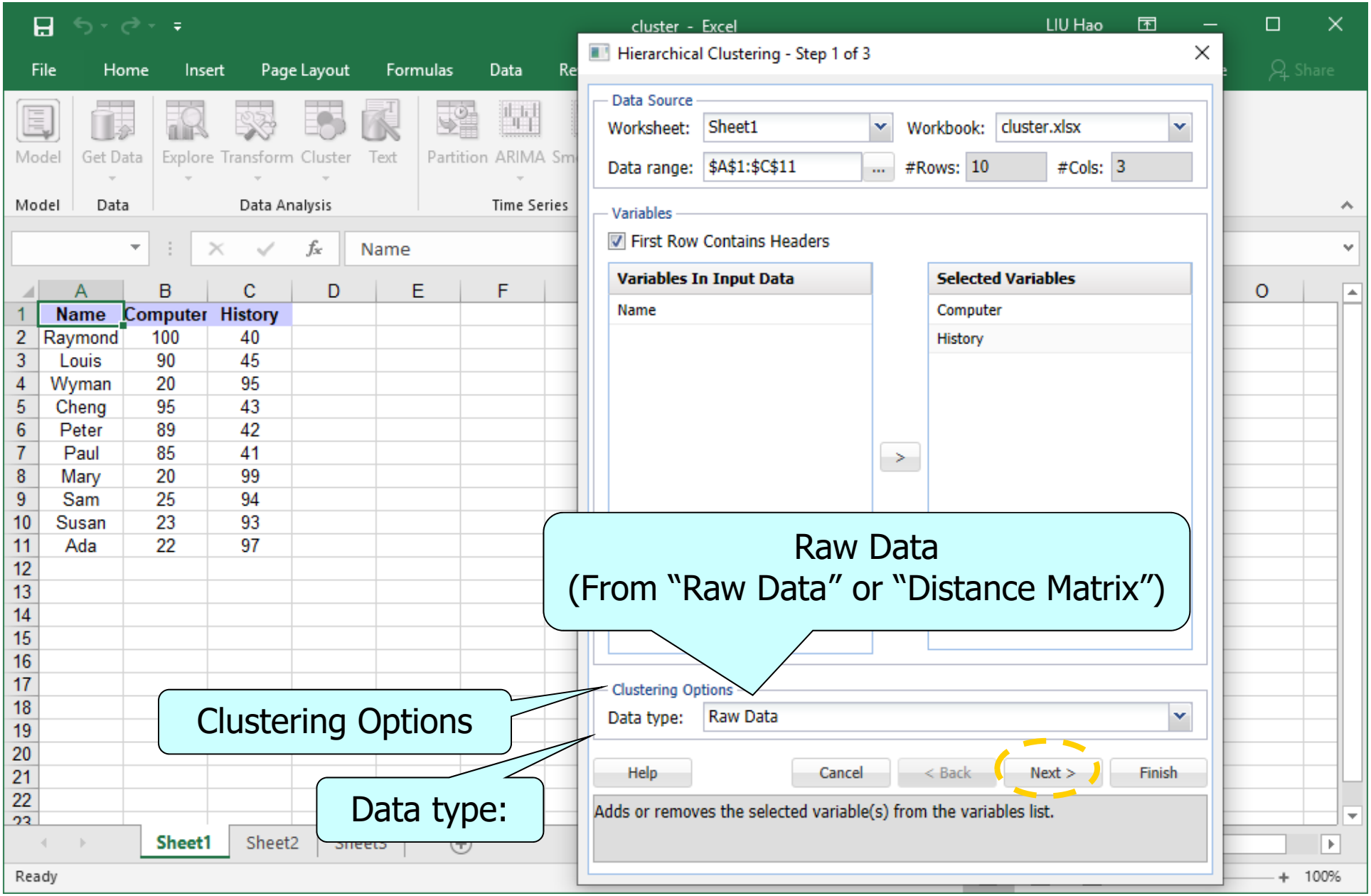

Data source
Workbook
Worksheet
Data range
File
Home
Insert
Page Layout
Formulas
Data
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Data Source
Worksheet: Sheet1
Workbook: cluster.xlsx
Data range: $A$1:$C$11
#Rows: 10
#Cols: 3
Variables
First Row Contains Headers
Variables In Input Data
Name
Computer
History
Selected Variables
Clustering Options
Data type: Raw Data
Help
Cancel
< Back
Next >
Finish
Name
Computer
History
Raymond 100 40
Louis 90 45
Wyman 20 95
Cheng 95 43
Peter 89 42
Paul 85 41
Mary 20 99
Sam 25 94
Susan 23 93
Ada 22 97
Sheet1
Sheet2
Sheet3
Ready
100%


cluster - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Model
Data
Data Analysis
Time Series
A1
Name
Name
Computer
History
Raymond 100 40
Louis 90 45
Wyman 20 95
Cheng 95 43
Peter 89 42
Paul 85 41
Mary 20 99
Sam 25 94
Susan 23 93
Ada 22 97
Sheet1
Sheet2
Sheet3
Point
Hierarchical Clustering - Step 1 of 3
Data Source
Worksheet: Sheet1
Workbook: cluster.xlsx
Data range: $A$1:$C$11
#Rows: 10
#Cols: 3
Variables
First Row Contains Headers
Variables In Input Data
Selected Variables
Name
Computer
History
Select Data Range
Range:
$A$1:$C$11
OK
Cancel
Processing...
Clustering Options
Data type: Raw Data
Help
Cancel
< Back
Next >
Finish

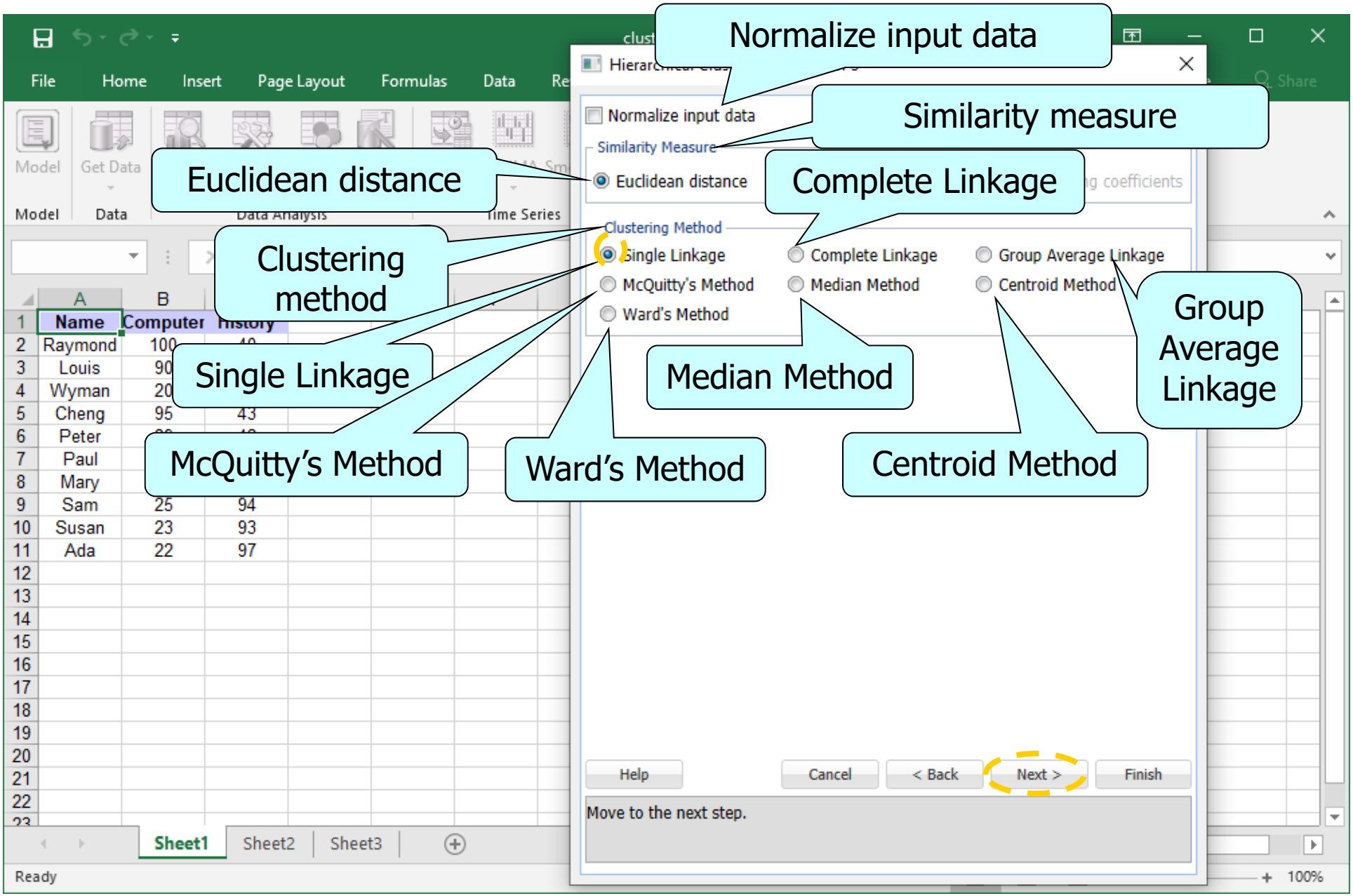

Rows:
10
Cols:
3
Variables
First Row Contains Headers
Variables In Input Data
Name
Computer
History
Data Source
Worksheet:
Sheet1
Data range:
$A$1:$C$11
#Rows:
10
#Cols:
3
Variables
First Row Contains Headers
Variables In Input Data
Selected Variables
Name
Computer
History
Clustering Options
Data type:
Raw Data
Help
Cancel
< Back
Next >
Finish
The list of variables in the input data range. You can select them as input or output variables using the arrow buttons.
Name
Computer
History
Sheet1
Sheet2
Sheet3
Ready

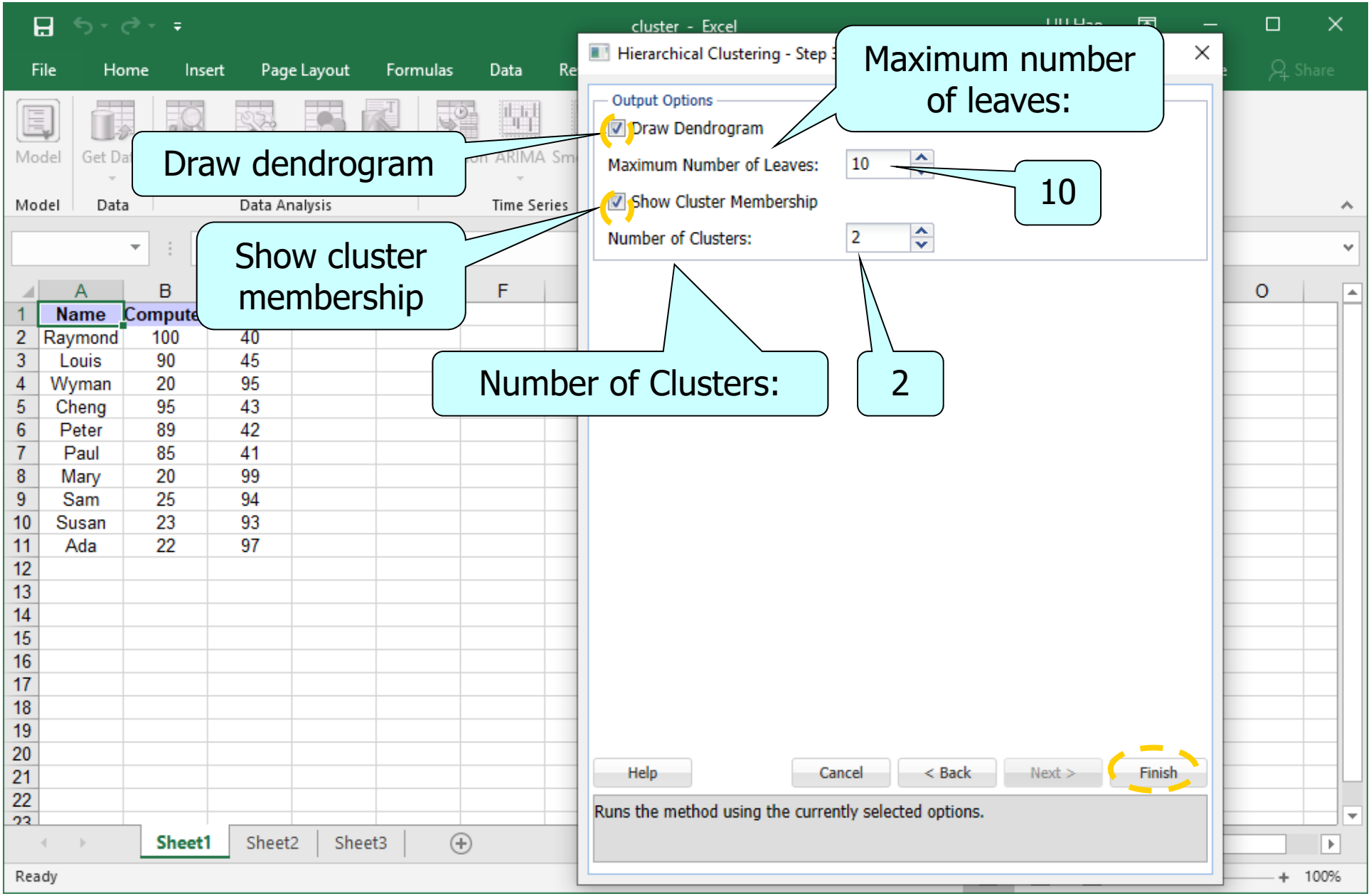

cluster - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Model
Data
Data Analysis
Time Series
Name
Name
Computer
History
Raymond 100 40
Louis 90 45
Wyman 20 95
Cheng 95 43
Peter 89 42
Paul 85 41
Mary 20 99
Sam 25 94
Susan 23 93
Ada 22 97
Sheet1
Sheet2
Sheet3
Ready
100%
Hierarchical Clustering - Step 1 of 3
Data Source
Worksheet: Sheet1
Workbook: cluster.xlsx
Data range: $A$1:$C$11
#Rows: 10
#Cols: 3
Variables
First Row Contains Headers
Variables In Input Data
Name
History
Selected Variables
Computer
>
Clustering Options
Data type: Raw Data
Help
Cancel
< Back
Next >
Finish
Adds or removes the selected variable(s) from the variables list.

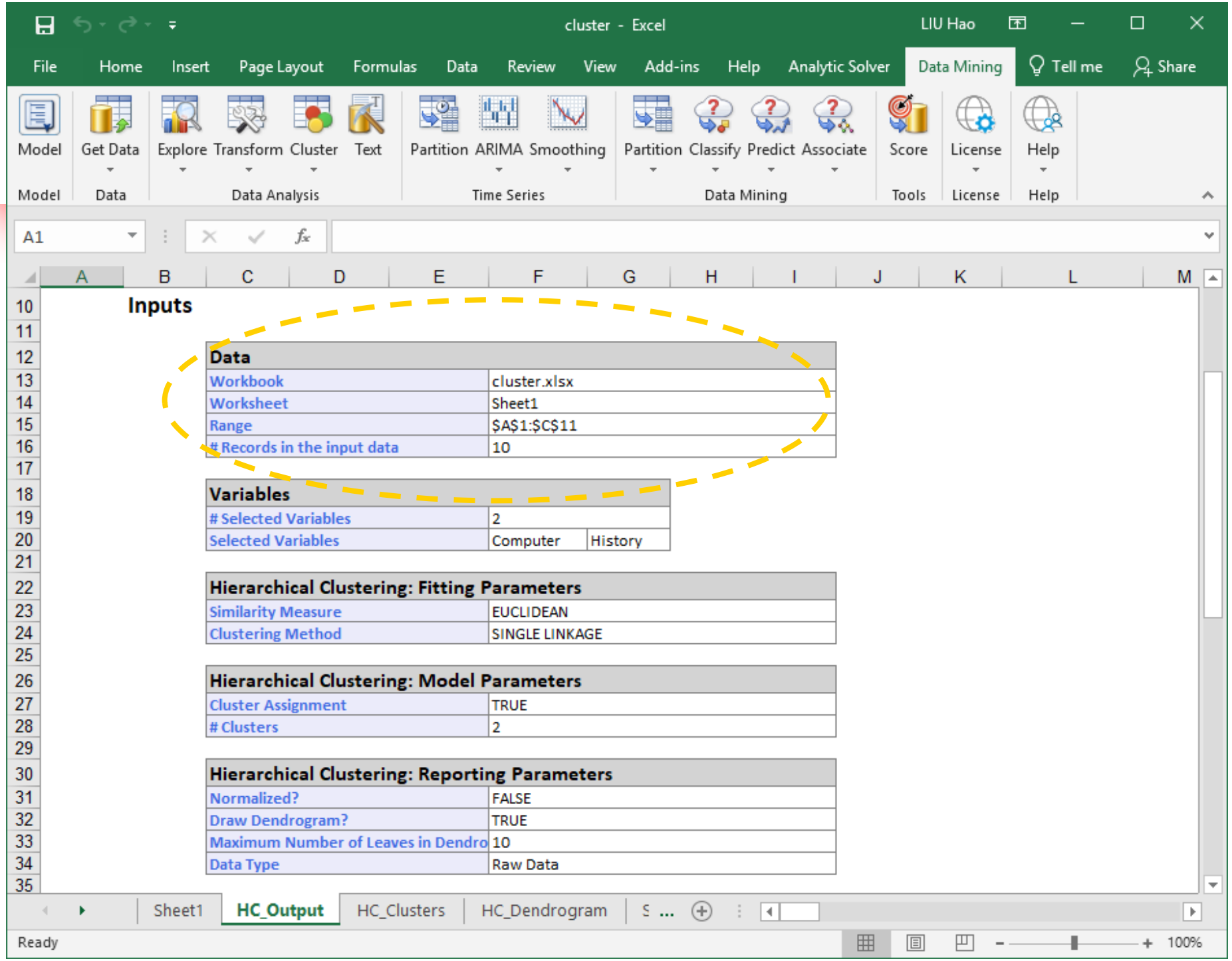

cluster - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Model
Data
Data Analysis
Time Series
Name
Name
Computer
History
Raymond 100 40
Louis 90 45
Wyman 20 95
Cheng 95 43
Peter 89 42
Paul 85 41
Mary 20 99
Sam 25 94
Susan 23 93
Ada 22 97
Sheet1
Sheet2
Sheet3
Ready
100%
Hierarchical Clustering - Step 1 of 3
Data Source
Worksheet: Sheet1
Workbook: cluster.xlsx
Data range: $A$1:$C$11
#Rows: 10
#Cols: 3
Variables
First Row Contains Headers
Variables In Input Data
Name
History
Selected Variables
Computer
>
Clustering Options
Data type: Raw Data
Help
Cancel
< Back
Next >
Finish
The list of variables in the input data range. You can select them as input or output variables using the arrow buttons.

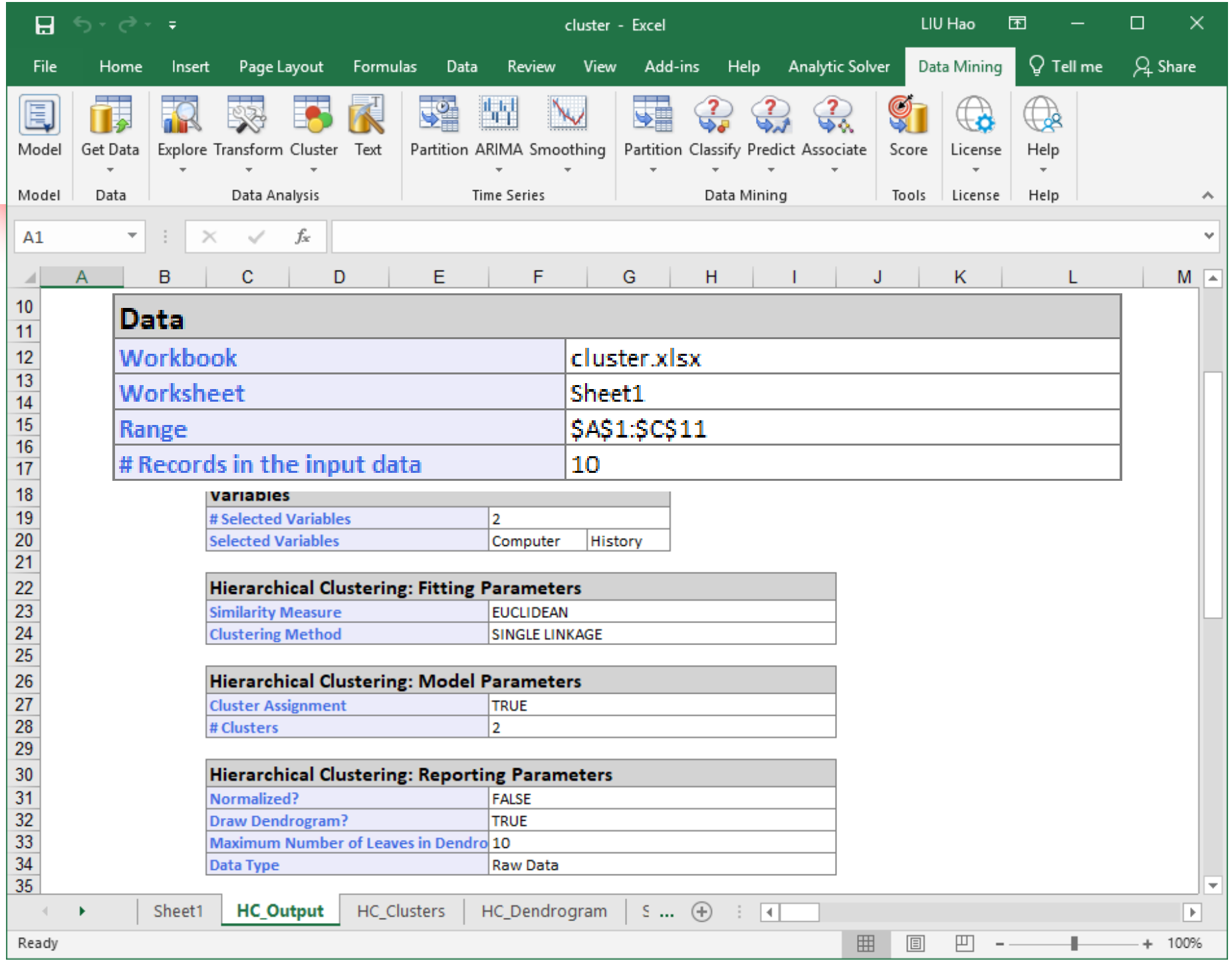


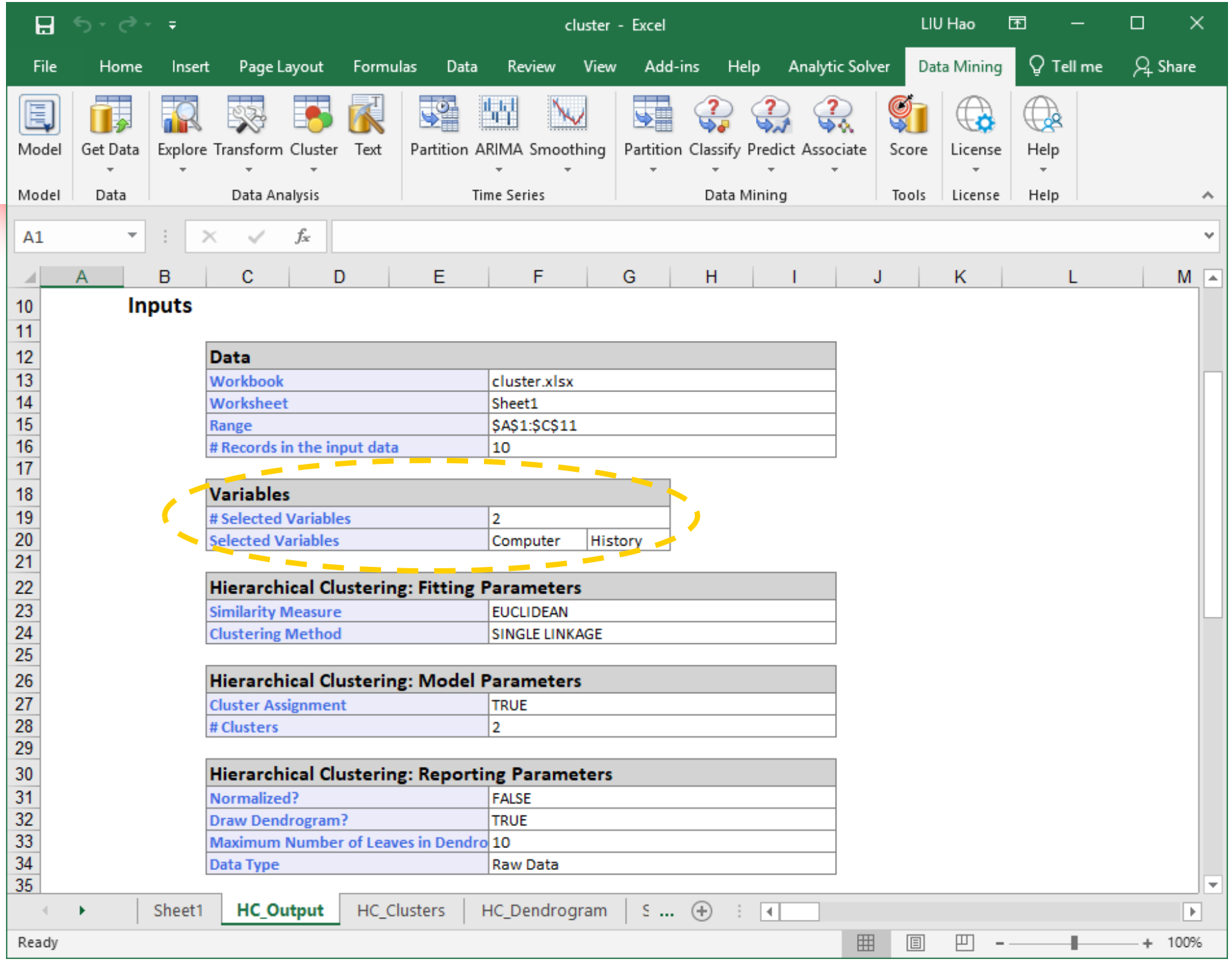
	A	B	C
1	Name	Computer	History
2	Raymond	100	40
3	Louis	90	45
4	Wyman	20	95
5	Cheng	95	43
6	Peter	89	42
7	Paul	85	41
8	Mary	20	99
9	Sam	25	94
10	Susan	23	93
11	Ada	22	97

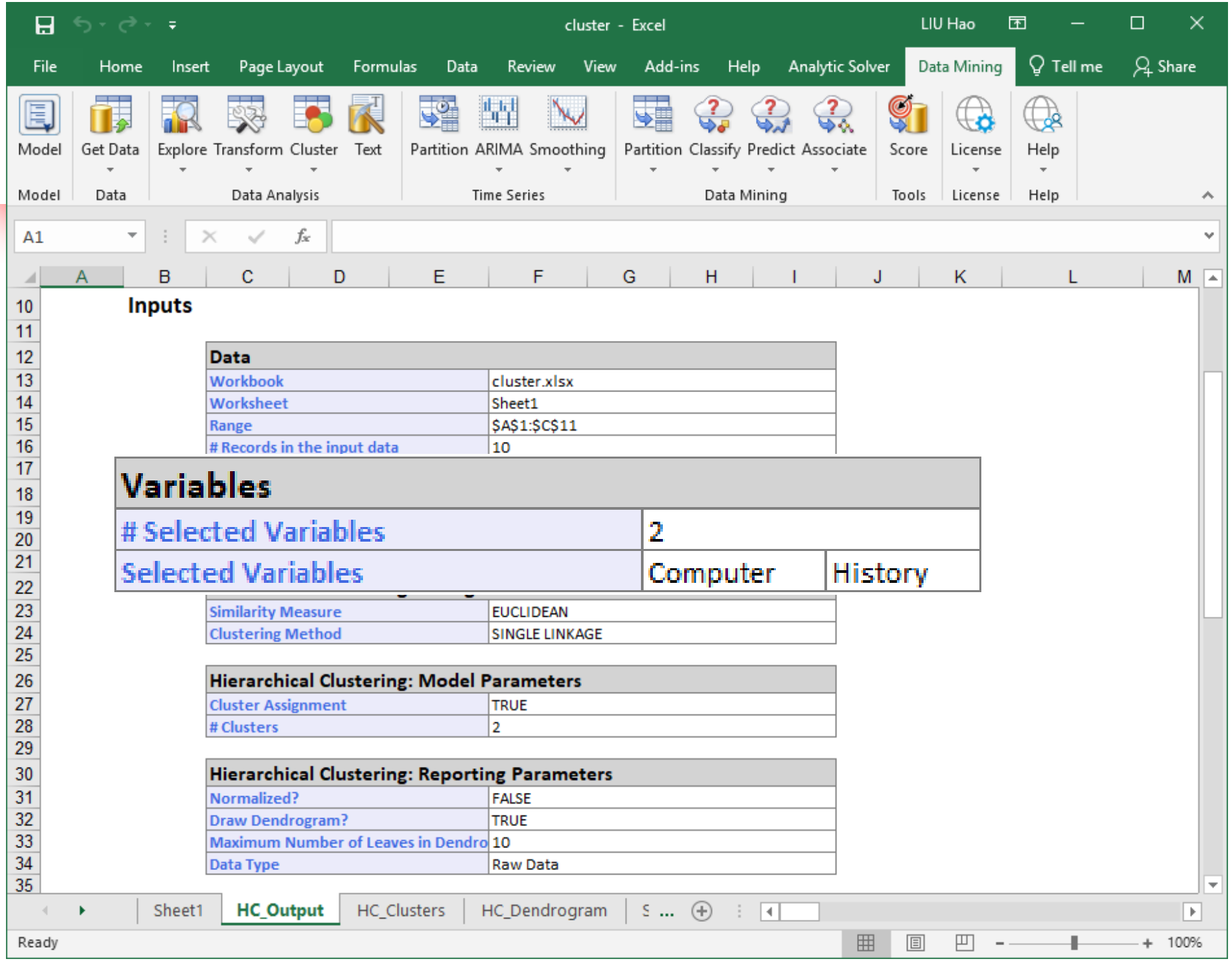

Normalize input data
Similarity measure
Euclidean distance
Complete Linkage
Clustering method
Single Linkage
Median Method
Group Average Linkage
McQuitty's Method
Ward's Method
Centroid Method
Normalize input data
Similarity Measure
Euclidean distance
Clustering Method
Single Linkage
Complete Linkage
Group Average Linkage
McQuitty's Method
Median Method
Centroid Method
Ward's Method
Help
Cancel
< Back
Next >
Finish
Move to the next step.
Name
Computer
Raymond
Louis
Wyman
Cheng
Peter
Paul
Mary
Sam
Susan
Ada
Sheet1
Sheet2
Sheet3
Ready

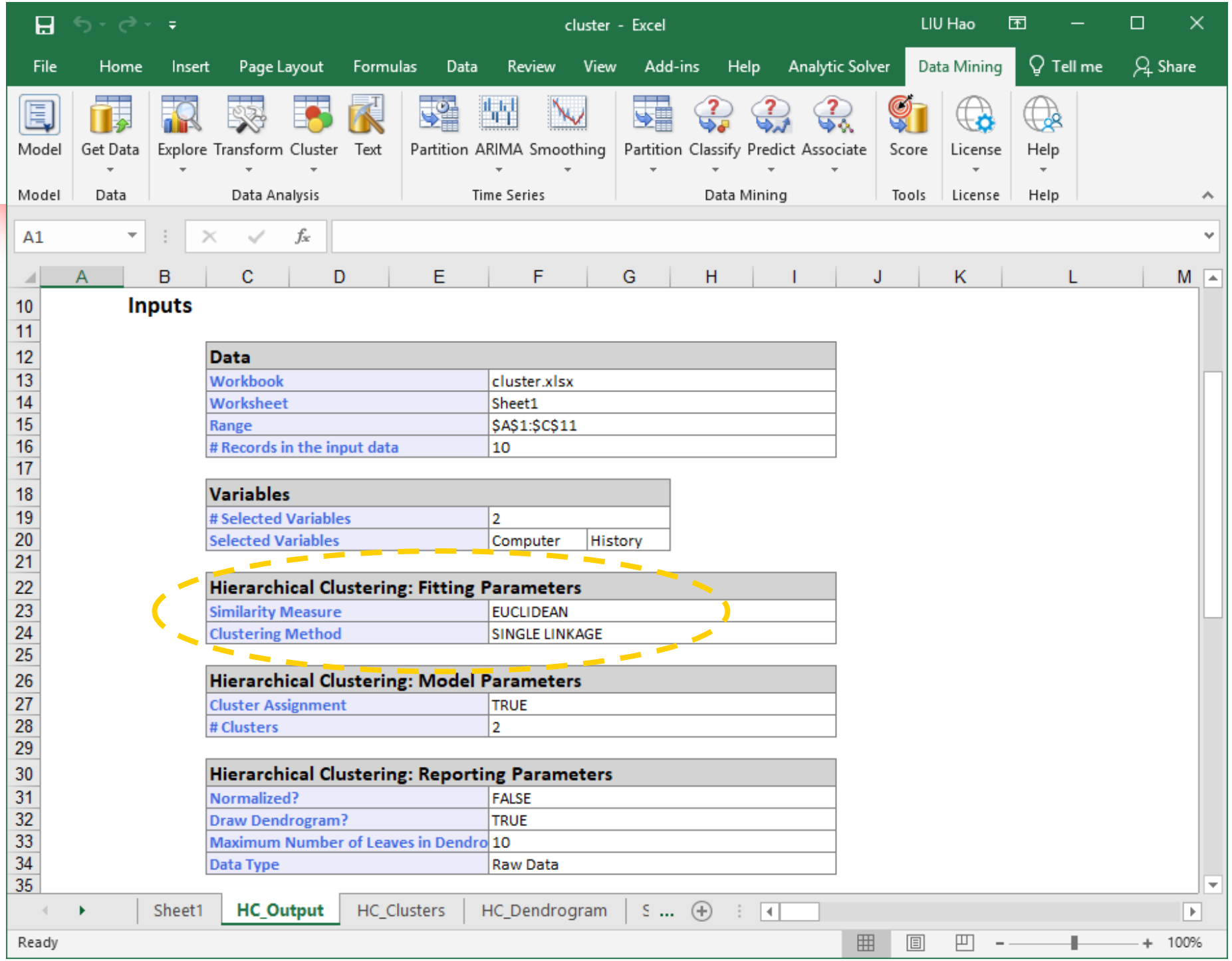

cluster - Excel
File
Home
Insert
Page Layout
Formulas
Data
Share
Model
Data
Data Analysis
Time Series
Hierarchical Clustering - Step 3
Output Options
Draw Dendrogram
Maximum Number of Leaves:
10
Show Cluster Membership
Number of Clusters:
2
Help
Cancel
< Back
Next >
Finish
Runs the method using the currently selected options.
Maximum number of leaves:
Draw dendrogram
Show cluster membership
Number of Clusters:
10
2
Name
Compute
Raymond 100 40
Louis 90 45
Wyman 20 95
Cheng 95 43
Peter 89 42
Paul 85 41
Mary 20 99
Sam 25 94
Susan 23 93
Ada 22 97
Sheet1
Sheet2
Sheet3
Ready
100%

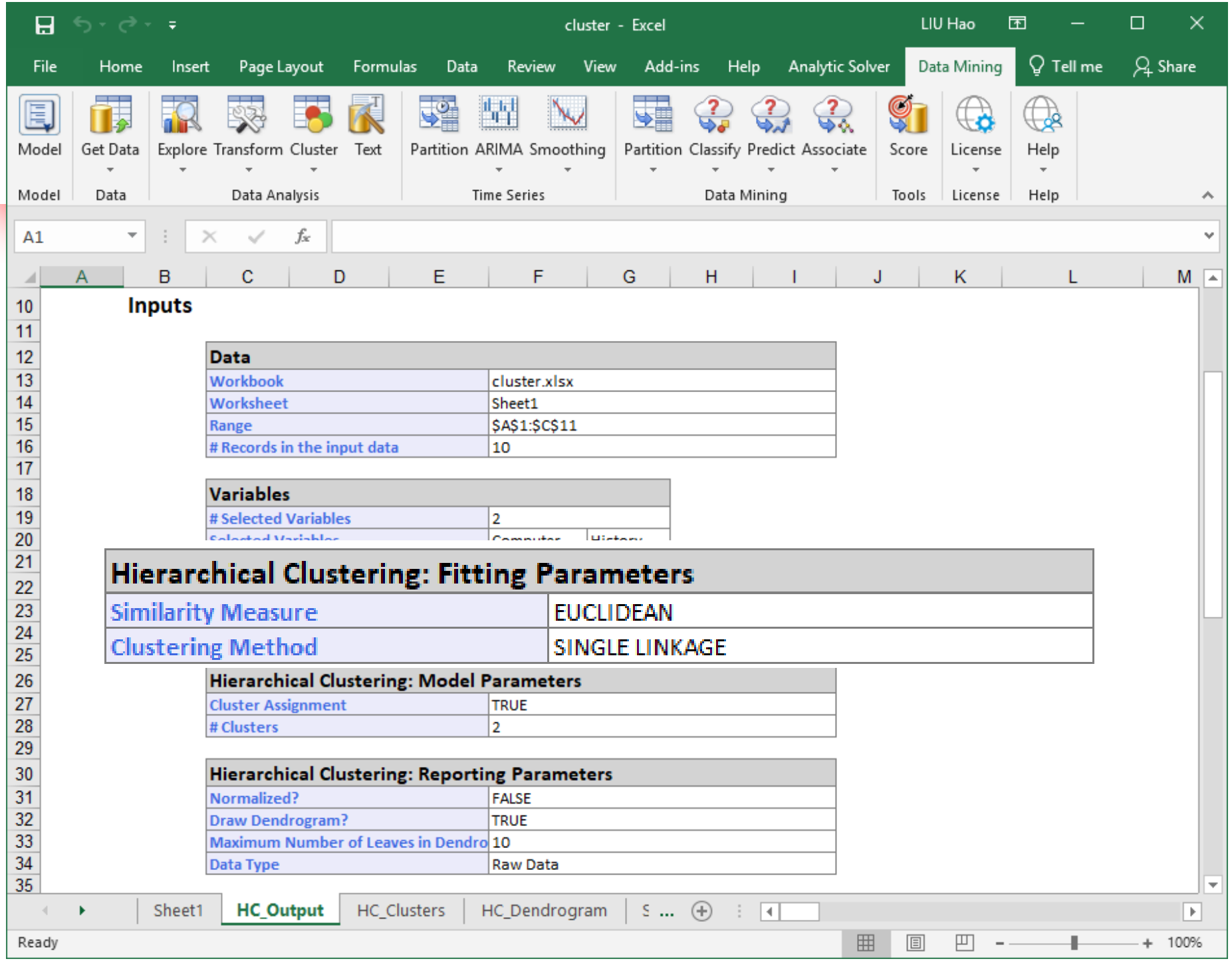

cluster - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Inputs
Data
Workbook cluster.xlsx
Worksheet Sheet1
Range $A$1:$C$11
Records in the input data 10
Variables
Selected Variables 2
Selected Variables Computer History
Hierarchical Clustering: Fitting Parameters
Similarity Measure EUCLIDEAN
Clustering Method SINGLE LINKAGE
Hierarchical Clustering: Model Parameters
Cluster Assignment TRUE
Clusters 2
Hierarchical Clustering: Reporting Parameters
Normalized? FALSE
Draw Dendrogram? TRUE
Maximum Number of Leaves in Dendro 10
Data Type Raw Data
Sheet1 HC_Output HC_Clusters HC_Dendrogram
Ready
100%

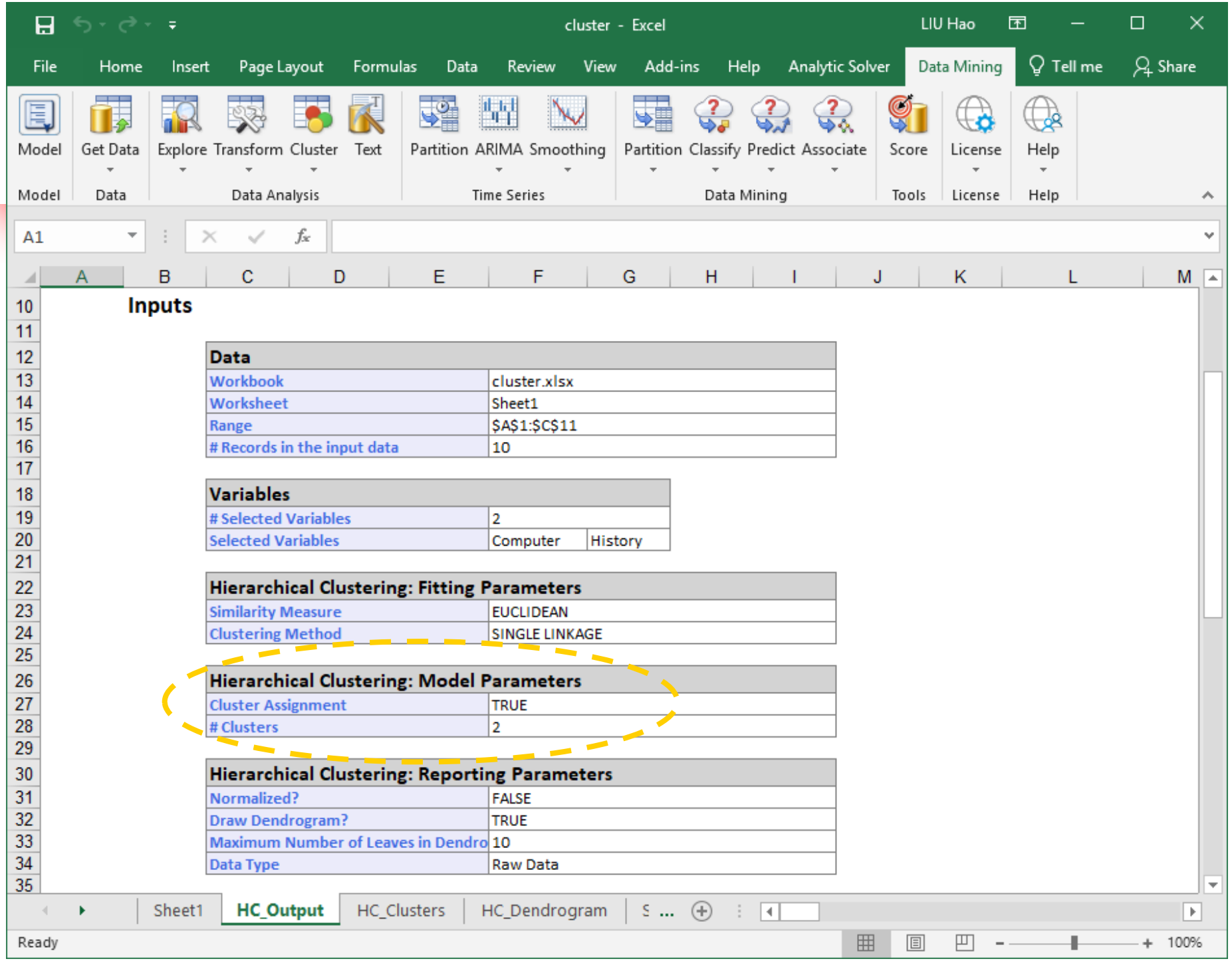

cluster - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Data
Workbook cluster.xlsx
Worksheet Sheet1
Range $A$1:$C$11
Records in the input data 10
Variables
Selected Variables 2
Selected Variables Computer History
Hierarchical Clustering: Fitting Parameters
Similarity Measure EUCLIDEAN
Clustering Method SINGLE LINKAGE
Hierarchical Clustering: Model Parameters
Cluster Assignment TRUE
Clusters 2
Hierarchical Clustering: Reporting Parameters
Normalized? FALSE
Draw Dendrogram? TRUE
Maximum Number of Leaves in Dendro 10
Data Type Raw Data
Sheet1 HC_Output HC_Clusters HC_Dendrogram
Ready
100%

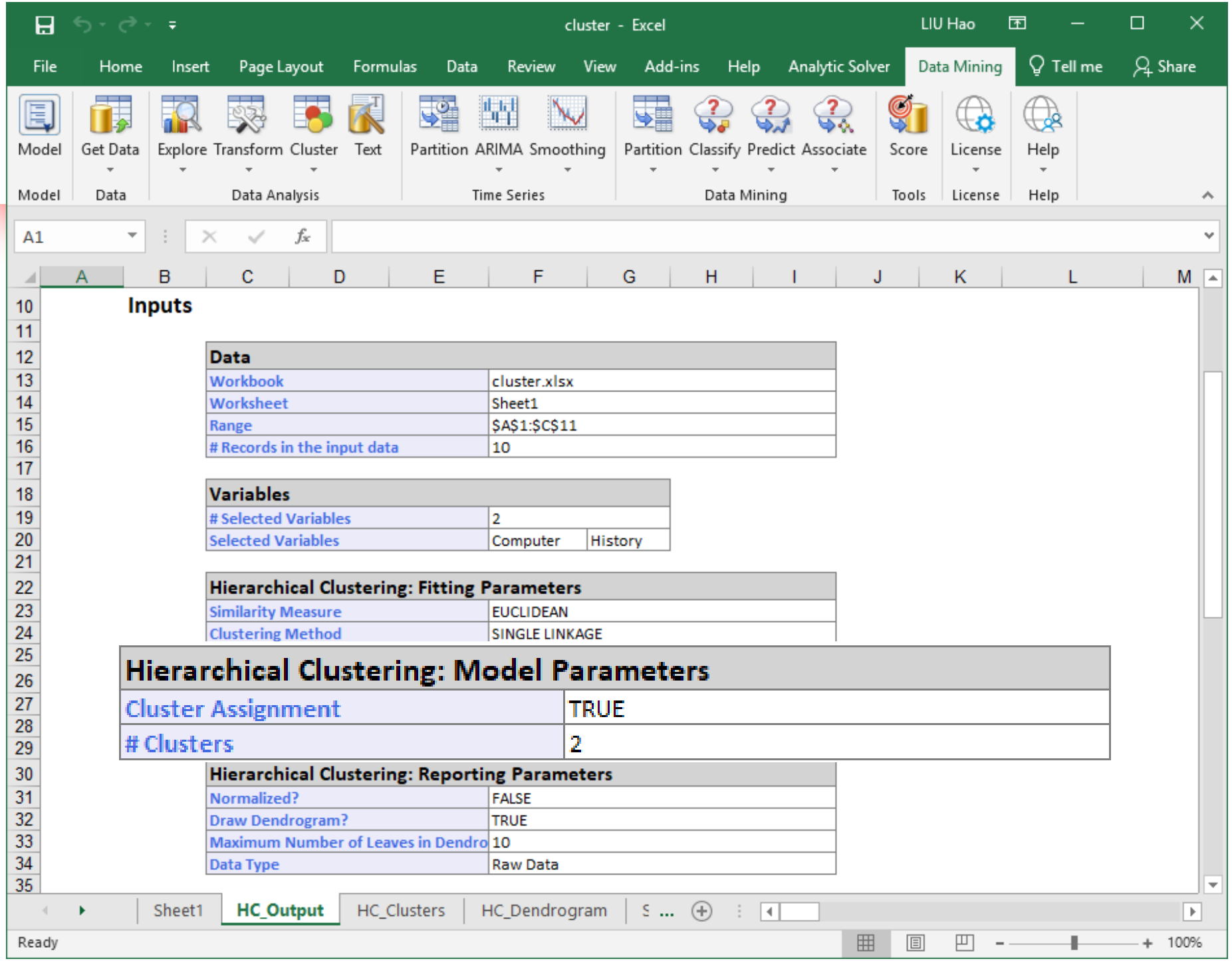

cluster - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Inputs
Data
Workbook cluster.xlsx
Worksheet Sheet1
Range $A$1:$C$11
Records in the input data 10
Variables
Selected Variables 2
Selected Variables Computer History
Hierarchical Clustering: Fitting Parameters
Similarity Measure EUCLIDEAN
Clustering Method SINGLE LINKAGE
Hierarchical Clustering: Model Parameters
Cluster Assignment TRUE
Clusters 2
Hierarchical Clustering: Reporting Parameters
Normalized? FALSE
Draw Dendrogram? TRUE
Maximum Number of Leaves in Dendro 10
Data Type Raw Data
Sheet1 HC_Output HC_Clusters HC_Dendrogram
Ready
100%

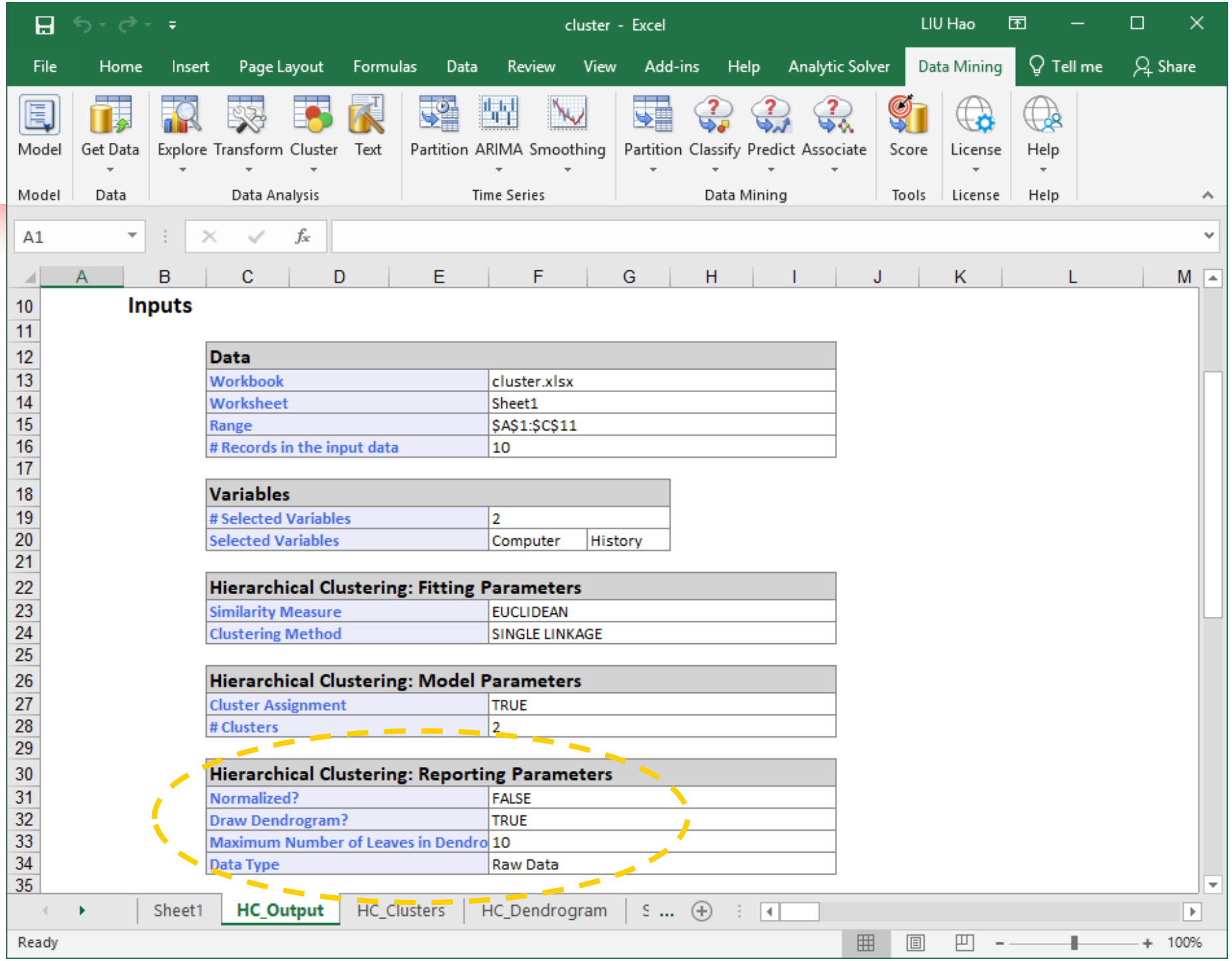

cluster - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Inputs
Data
Workbook cluster.xlsx
Worksheet Sheet1
Range $A$1:$C$11
Records in the input data 10
Variables
Selected Variables 2
Selected Variables Computer History
Similarity Measure EUCLIDEAN
Clustering Method SINGLE LINKAGE
Hierarchical Clustering: Model Parameters
Cluster Assignment TRUE
Clusters 2
Hierarchical Clustering: Reporting Parameters
Normalized? FALSE
Draw Dendrogram? TRUE
Maximum Number of Leaves in Dendro 10
Data Type Raw Data
Sheet1 HC_Output HC_Clusters HC_Dendrogram
Ready
100%

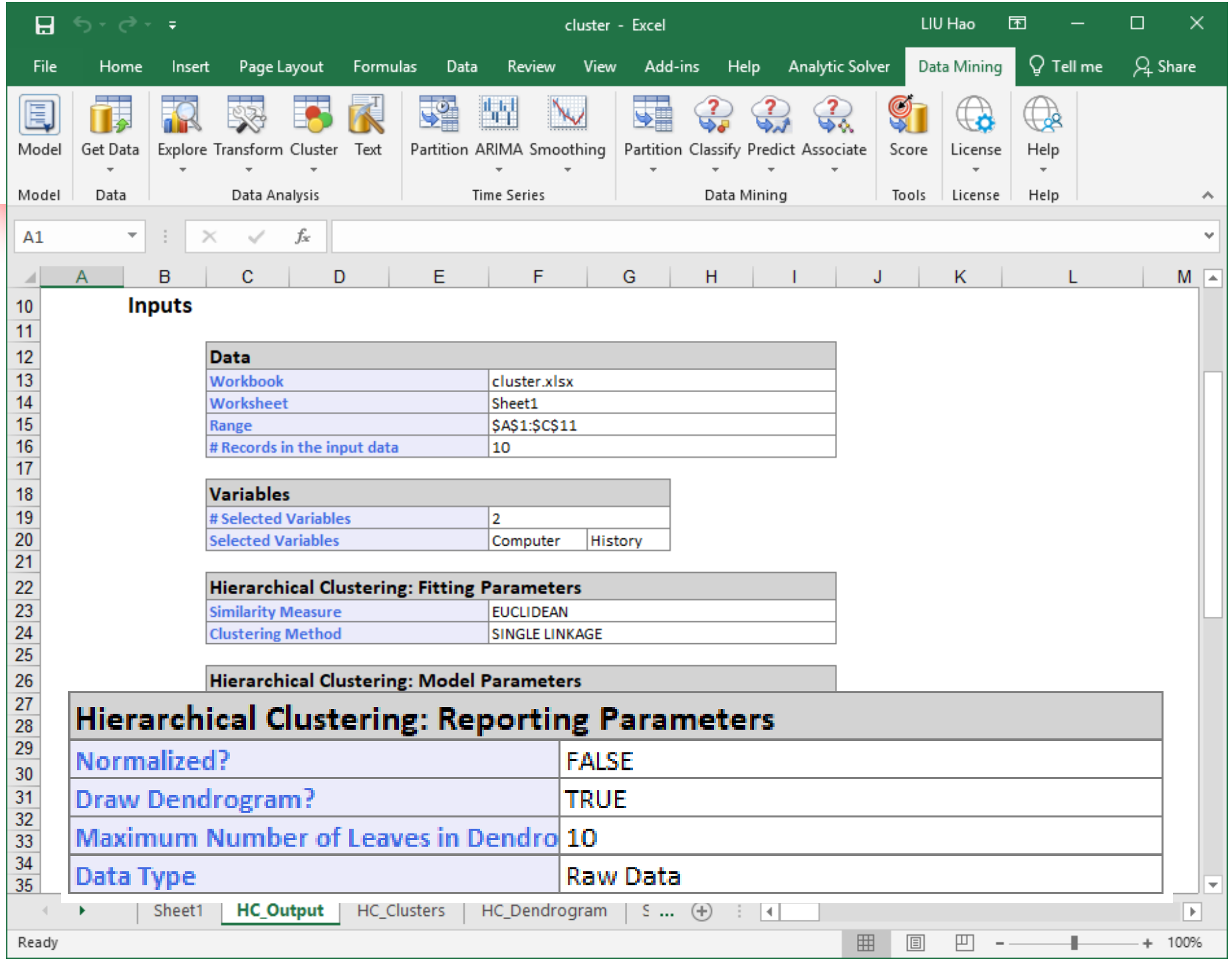

cluster - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Inputs
Data
Workbook cluster.xlsx
Worksheet Sheet1
Range $A$1:$C$11
Records in the input data 10
Variables
Selected Variables 2
Selected Variables Computer History
Hierarchical Clustering: Fitting Parameters
Similarity Measure EUCLIDEAN
Clustering Method SINGLE LINKAGE
Hierarchical Clustering: Model Parameters
Cluster Assignment TRUE
Clusters 2
Hierarchical Clustering: Reporting Parameters
Normalized? FALSE
Draw Dendrogram? TRUE
Maximum Number of Leaves in Dendro 10
Data Type Raw Data
Sheet1 HC_Output HC_Clusters HC_Dendrogram
Ready
100%

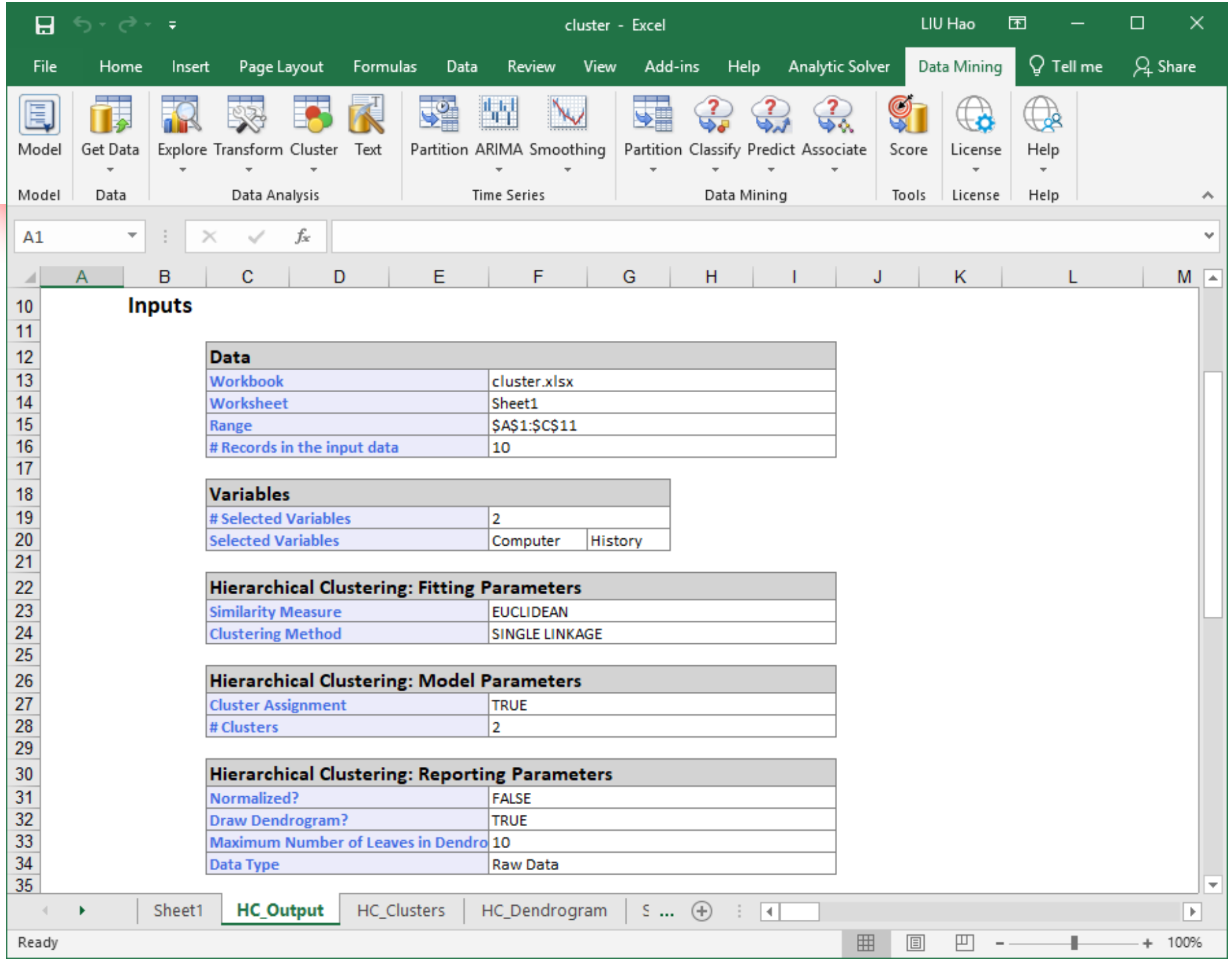

cluster - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Inputs
Data
Workbook cluster.xlsx
Worksheet Sheet1
Range $A$1:$C$11
Records in the input data 10
Variables
Selected Variables 2
Hierarchical Clustering: Fitting Parameters
Similarity Measure EUCLIDEAN
Clustering Method SINGLE LINKAGE
Hierarchical Clustering: Model Parameters
Cluster Assignment TRUE
Clusters 2
Hierarchical Clustering: Reporting Parameters
Normalized? FALSE
Draw Dendrogram? TRUE
Maximum Number of Leaves in Dendro 10
Data Type Raw Data
Sheet1 HC_Output HC_Clusters HC_Dendrogram
Ready
100%

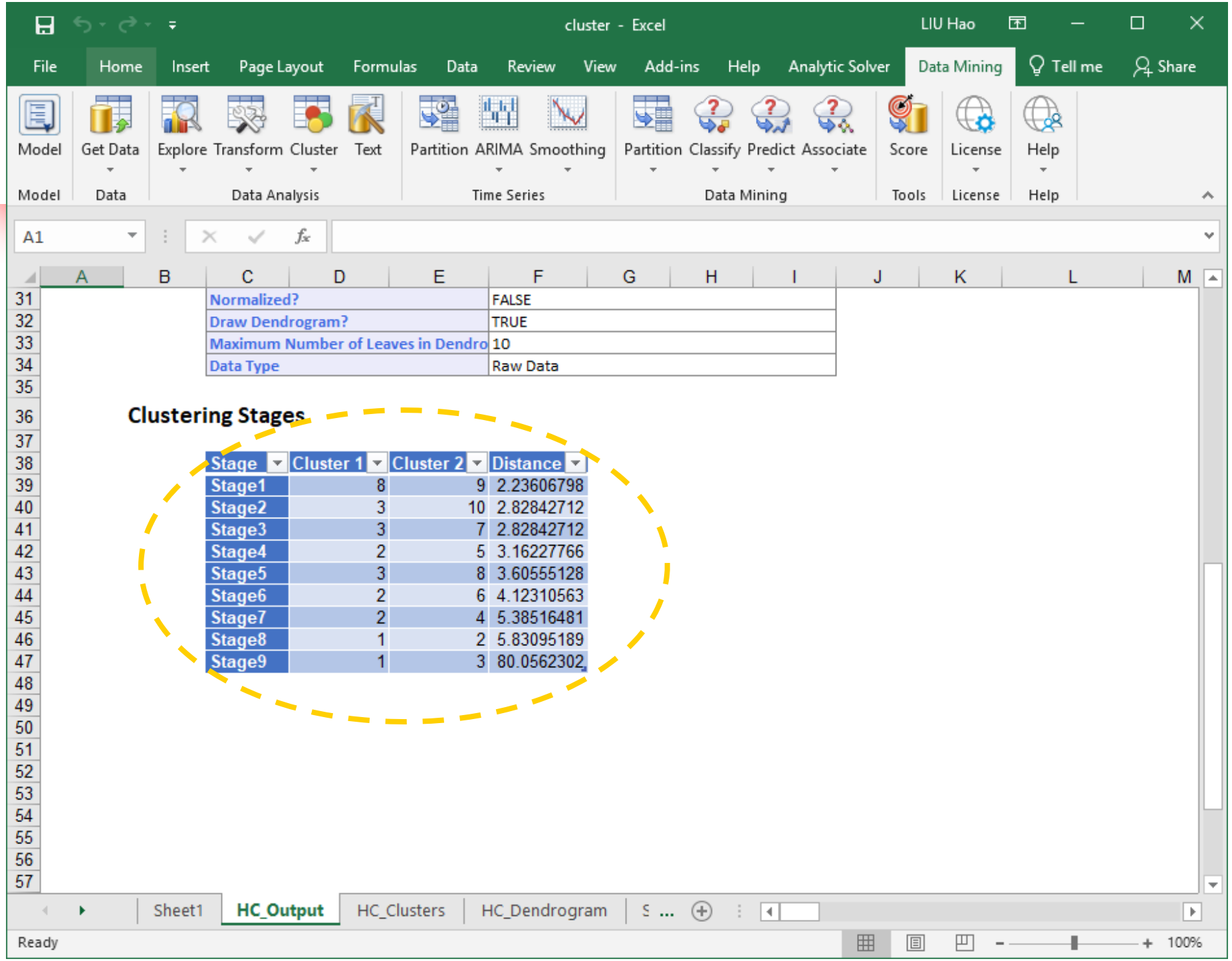

cluster - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Inputs
Data
Workbook cluster.xlsx
Worksheet Sheet1
Range $A$1:$C$11
Records in the input data 10
Variables
Selected Variables 2
Selected Variables Computer History
Hierarchical Clustering: Fitting Parameters
Similarity Measure EUCLIDEAN
Clustering Method SINGLE LINKAGE
Hierarchical Clustering: Model Parameters
Cluster Assignment TRUE
Clusters 2
Hierarchical Clustering: Reporting Parameters
Normalized? FALSE
Draw Dendrogram? TRUE
Maximum Number of Leaves in Dendro 10
Data Type Raw Data
Sheet1 HC_Output HC_Clusters HC_Dendrogram
Ready
100%

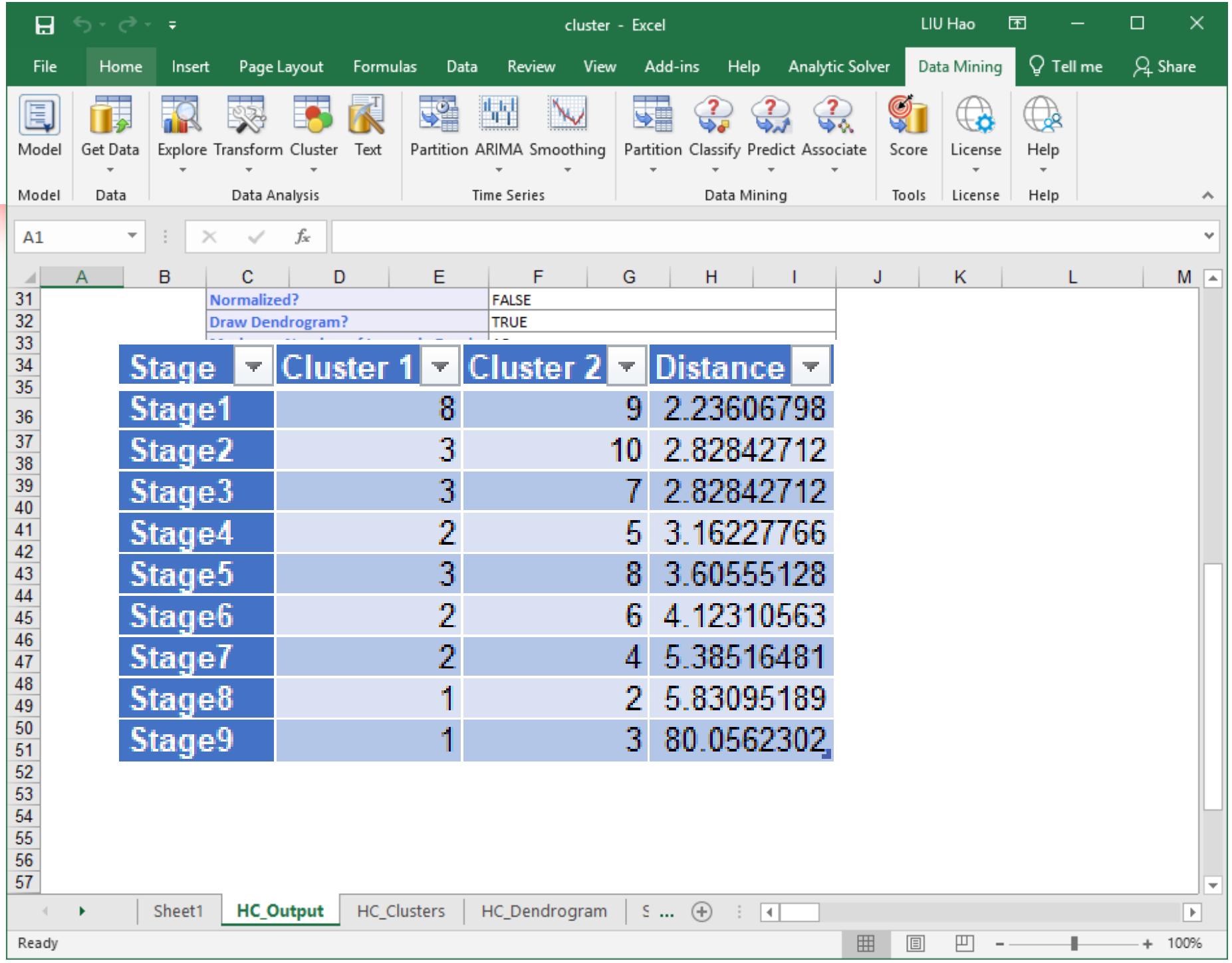

cluster - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Inputs
Data
Workbook cluster.xlsx
Worksheet Sheet1
Range $A$1:$C$11
Records in the input data 10
Variables
Selected Variables 2
Selected Variables Computer History
Hierarchical Clustering: Fitting Parameters
Similarity Measure EUCLIDEAN
Clustering Method SINGLE LINKAGE
Hierarchical Clustering: Model Parameters
Cluster Assignment TRUE
Clusters 2
Hierarchical Clustering: Reporting Parameters
Normalized? FALSE
Draw Dendrogram? TRUE
Maximum Number of Leaves in Dendro 10
Data Type Raw Data
Sheet1 HC_Output HC_Clusters HC_Dendrogram
Ready
100%


cluster - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Inputs
Data
Workbook cluster.xlsx
Worksheet Sheet1
Range $A$1:$C$11
Records in the input data 10
Variables
Selected Variables 2
Selected Variables Computer History
Hierarchical Clustering: Fitting Parameters
Similarity Measure EUCLIDEAN
Clustering Method SINGLE LINKAGE
Hierarchical Clustering: Model Parameters
Cluster Assignment TRUE
Clusters 2
Hierarchical Clustering: Reporting Parameters
Normalized? FALSE
Draw Dendrogram? TRUE
Maximum Number of Leaves in Dendro 10
Data Type Raw Data
Sheet1 HC_Output HC_Clusters HC_Dendrogram
Ready
100%

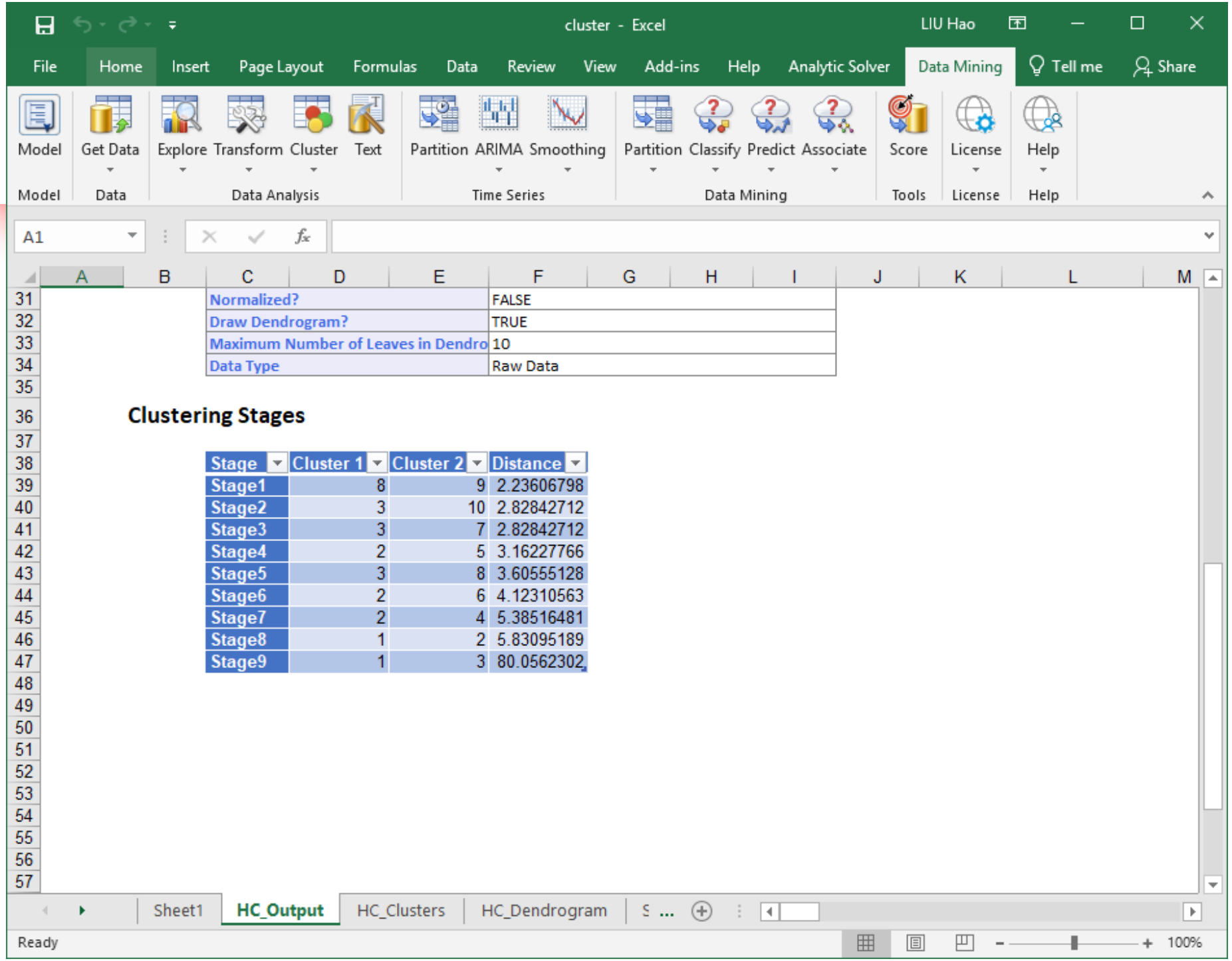

cluster - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Inputs
Data
Workbook cluster.xlsx
Worksheet Sheet1
Range $A$1:$C$11
Records in the input data 10
Variables
Selected Variables 2
Selected Variables Computer History
Hierarchical Clustering: Fitting Parameters
Similarity Measure EUCLIDEAN
Clustering Method SINGLE LINKAGE
Hierarchical Clustering: Model Parameters
Hierarchical Clustering: Reporting Parameters
Normalized? FALSE
Draw Dendrogram? TRUE
Maximum Number of Leaves in Dendro 10
Data Type Raw Data
Sheet1 HC_Output HC_Clusters HC_Dendrogram
Ready
100%

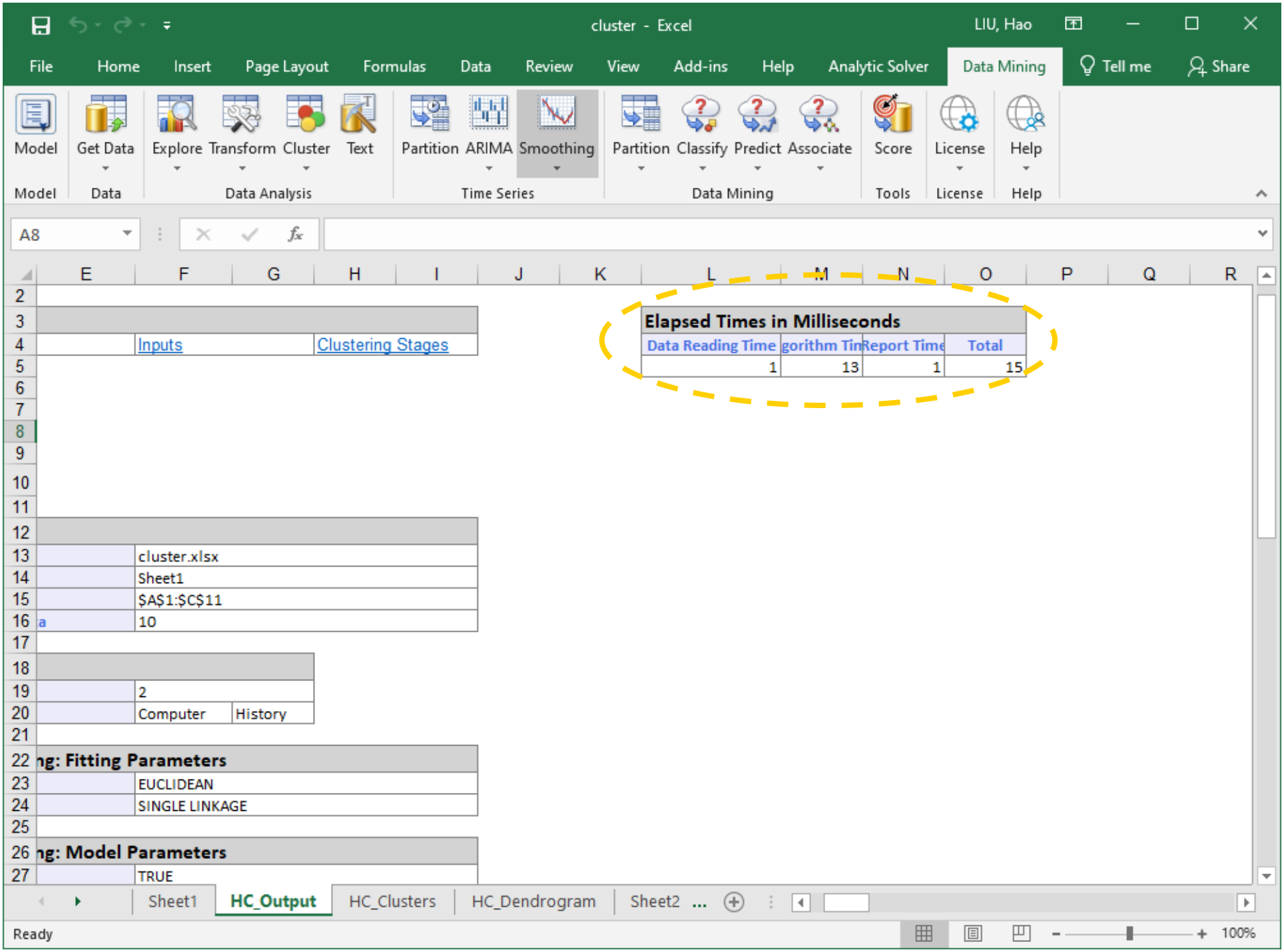

cluster - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Inputs
Data
Workbook cluster.xlsx
Worksheet Sheet1
Range $A$1:$C$11
Records in the input data 10
Variables
Selected Variables 2
Selected Variables Computer History
Hierarchical Clustering: Fitting Parameters
Similarity Measure EUCLIDEAN
Clustering Method SINGLE LINKAGE
Hierarchical Clustering: Model Parameters
Cluster Assignment TRUE
Clusters 2
Hierarchical Clustering: Reporting Parameters
Normalized? FALSE
Draw Dendrogram? TRUE
Maximum Number of Leaves in Dendro 10
Data Type Raw Data
Sheet1 HC_Output HC_Clusters HC_Dendrogram
Ready
100%

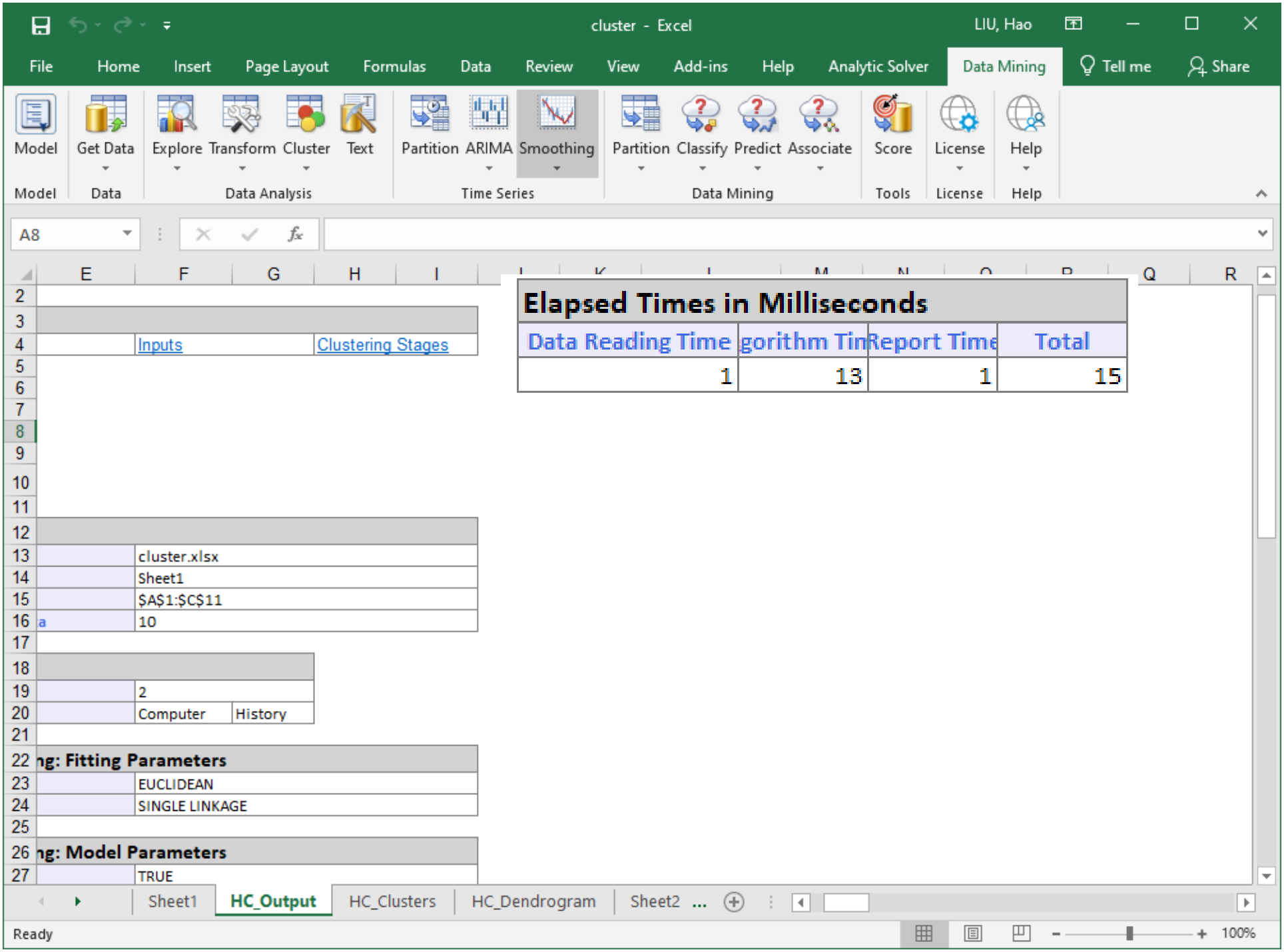

cluster - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
Model
Data
Data Analysis
Time Series
Data Mining
Tools
License
Help
A1
Normalized? FALSE
Draw Dendrogram? TRUE
Maximum Number of Leaves in Dendro 10
Data Type Raw Data
Clustering Stages
Stage Cluster 1 Cluster 2 Distance
Stage1 8 9 2.23606798
Stage2 3 10 2.82842712
Stage3 3 7 2.82842712
Stage4 2 5 3.16227766
Stage5 3 8 3.60555128
Stage6 2 6 4.12310563
Stage7 2 4 5.38516481
Stage8 1 2 5.83095189
Stage9 1 3 80.0562302
Sheet1
HC_Output
HC_Clusters
HC_Dendrogram
Ready
100%

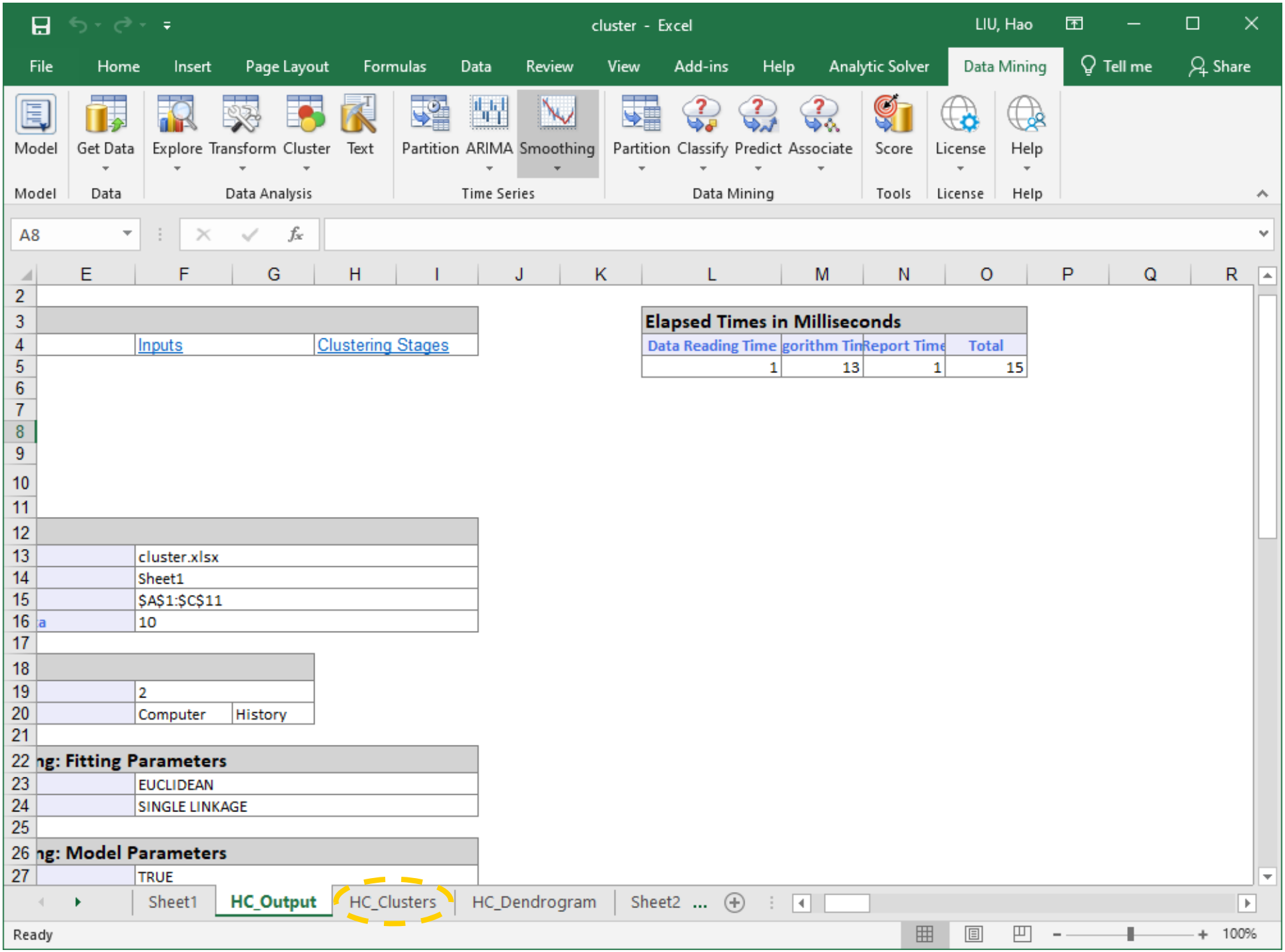


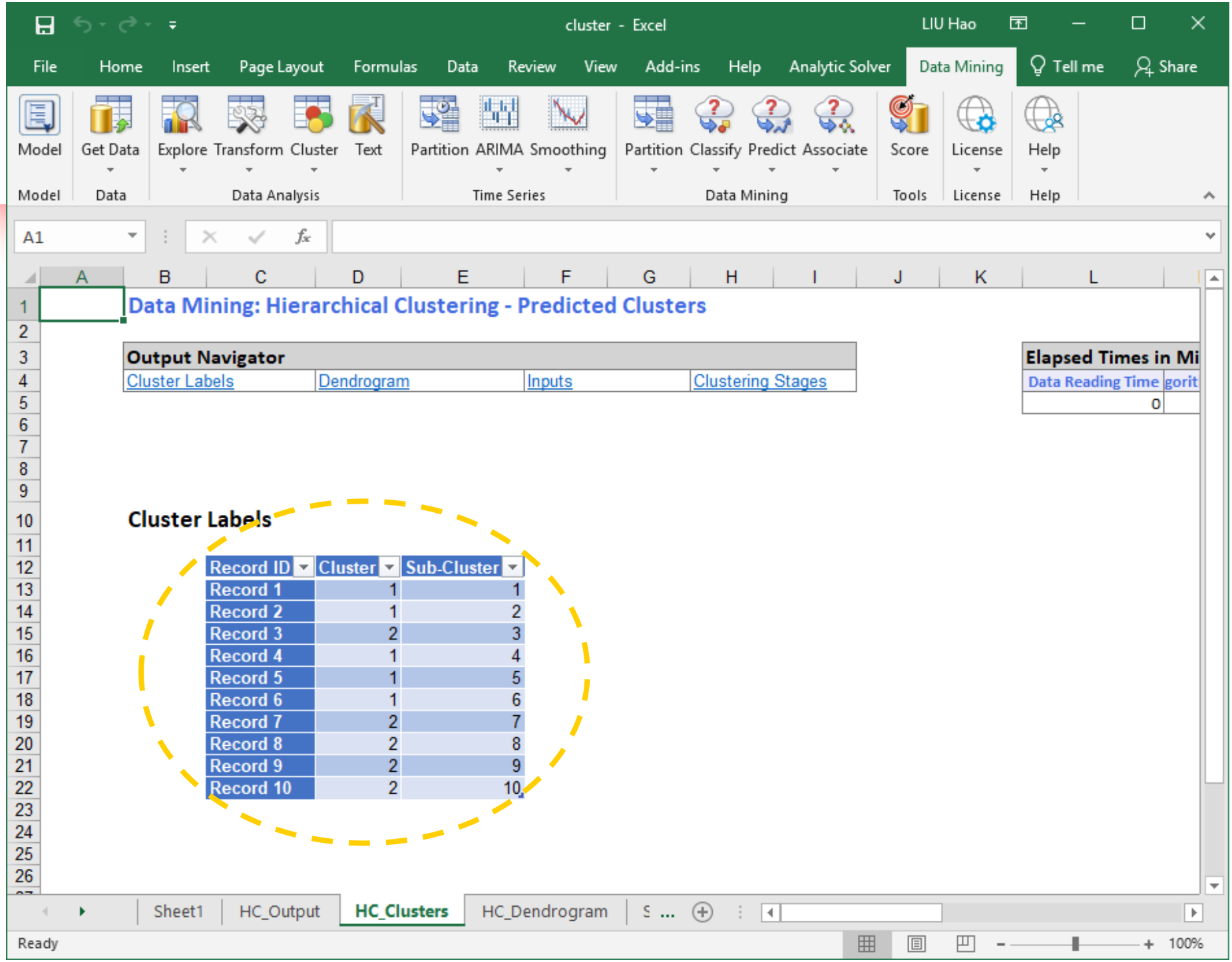
Stage	Cluster 1	Cluster 2	Distance
Stage1	8	9	2.23606798
Stage2	3	10	2.82842712
Stage3	3	7	2.82842712
Stage4	2	5	3.16227766
Stage5	3	8	3.60555128
Stage6	2	6	4.12310563
Stage7	2	4	5.38516481
Stage8	1	2	5.83095189
Stage9	1	3	80.0562302

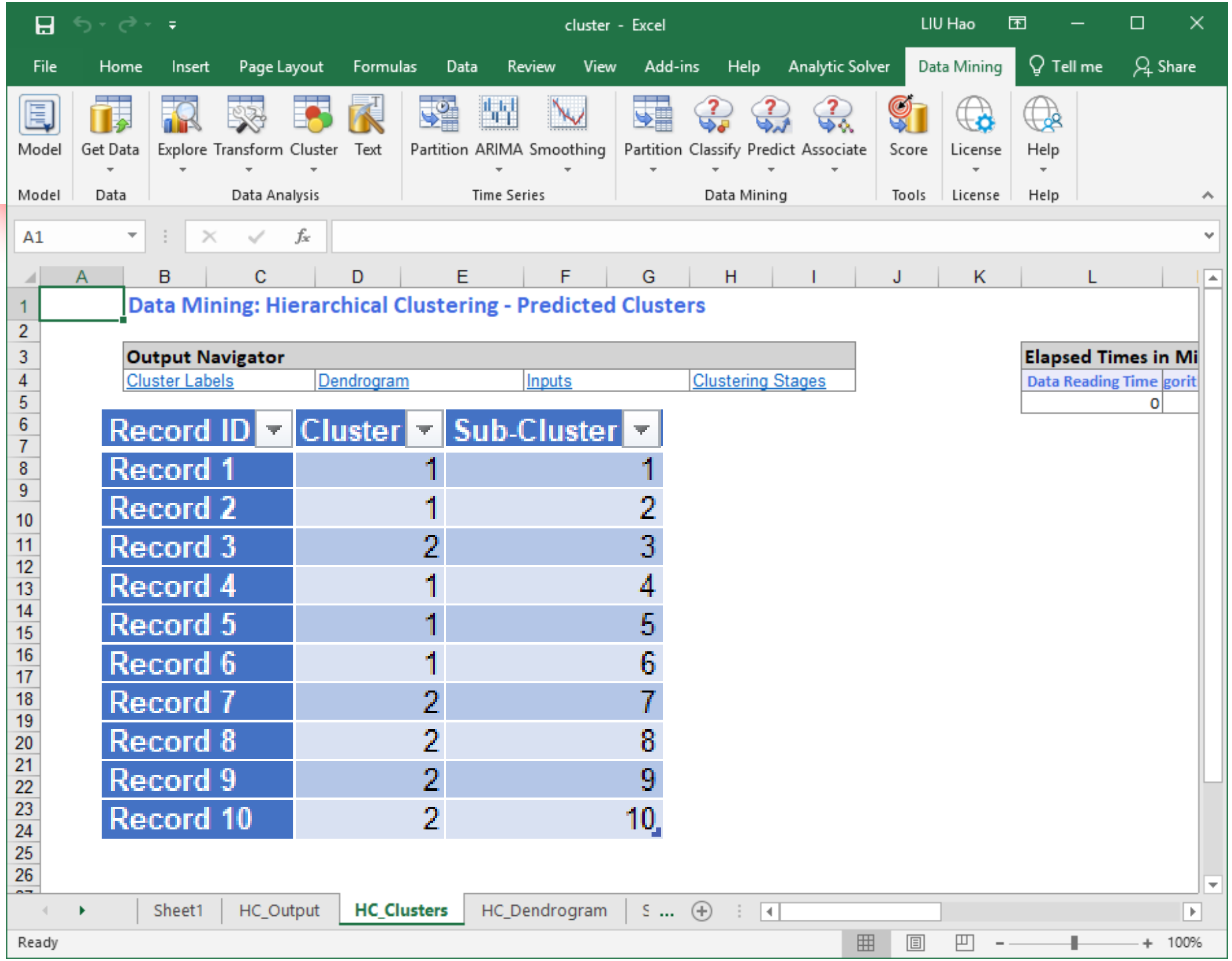

cluster - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
Model
Data
Data Analysis
Time Series
Data Mining
Tools
License
Help
A1
Normalized? FALSE
Draw Dendrogram? TRUE
Maximum Number of Leaves in Dendro 10
Data Type Raw Data
Clustering Stages
Stage Cluster 1 Cluster 2 Distance
Stage1 8 9 2.23606798
Stage2 3 10 2.82842712
Stage3 3 7 2.82842712
Stage4 2 5 3.16227766
Stage5 3 8 3.60555128
Stage6 2 6 4.12310563
Stage7 2 4 5.38516481
Stage8 1 2 5.83095189
Stage9 1 3 80.0562302
Sheet1
HC_Output
HC_Clusters
HC_Dendrogram
Ready
100%

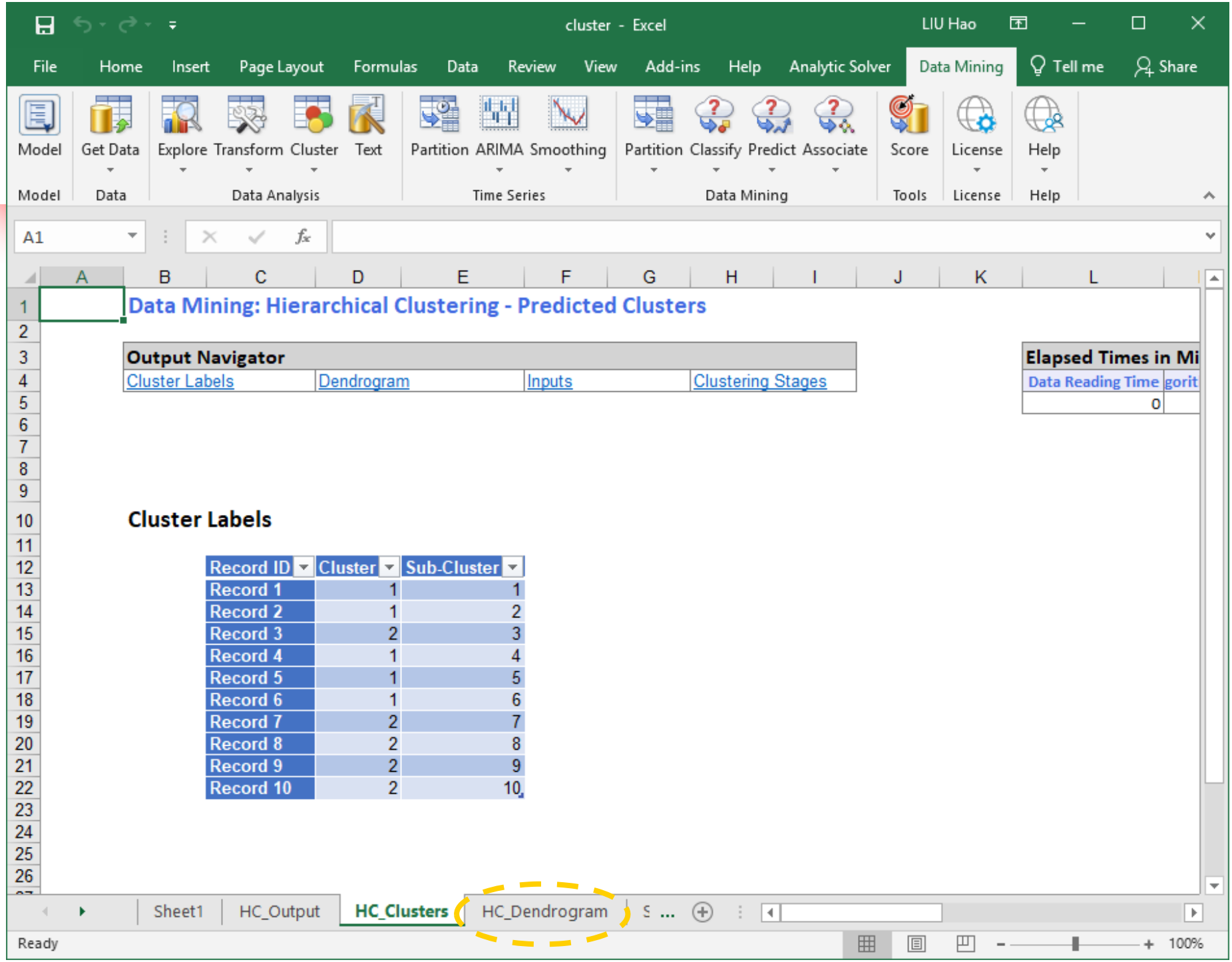

cluster - Excel
LIU, Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
Model
Data
Data Analysis
Time Series
Data Mining
Tools
License
Help
A8
Elapsed Times in Milliseconds
Data Reading Time
gorithm Tin
Report Time
Total
1
13
1
15
Inputs
Clustering Stages
cluster.xlsx
Sheet1
$A$1:$C$11
a
10
2
Computer
History
ng: Fitting Parameters
EUCLIDEAN
SINGLE LINKAGE
ng: Model Parameters
TRUE
Sheet1
HC_Output
HC_Clusters
HC_Dendrogram
Sheet2
Ready
100%

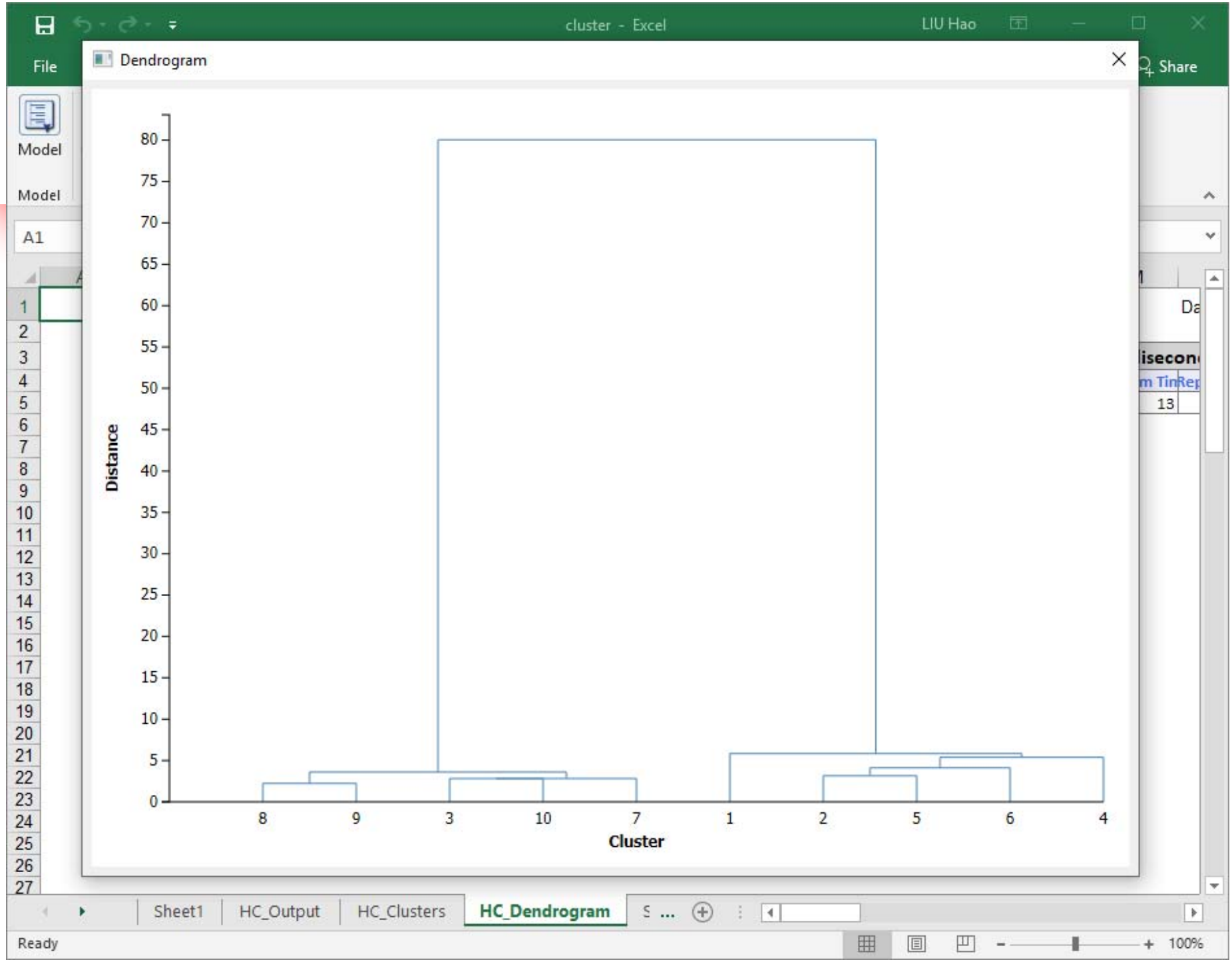

cluster - Excel
LIU, Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
Model
Data
Data Analysis
Time Series
Data Mining
Tools
License
Help
A8
Elapsed Times in Milliseconds
Data Reading Time
gorithm Tin
Report Time
Total
1
13
1
15
Inputs
Clustering Stages
cluster.xlsx
Sheet1
$A$1:$C$11
a
10
2
Computer
History
ng: Fitting Parameters
EUCLIDEAN
SINGLE LINKAGE
ng: Model Parameters
TRUE
Sheet1
HC_Output
HC_Clusters
HC_Dendrogram
Sheet2
Ready
100%


cluster - Excel
LIU, Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
Model
Data
Data Analysis
Time Series
Data Mining
Tools
License
Help
A8
Inputs
Clustering Stages
Elapsed Times in Milliseconds
Data Reading Time
gorithm Tin
Report Time
Total
1
13
1
15
cluster.xlsx
Sheet1
$A$1:$C$11
a
10
2
Computer
History
ng: Fitting Parameters
EUCLIDEAN
SINGLE LINKAGE
ng: Model Parameters
TRUE
Sheet1
HC_Output
HC_Clusters
HC_Dendrogram
Sheet2
Ready
100%


cluster - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Data Mining: Hierarchical Clustering - Predicted Clusters
Output Navigator
Cluster Labels Dendrogram Inputs Clustering Stages
Elapsed Times in Mi
Data Reading Time gorit
0
Cluster Labels
Record ID Cluster Sub-Cluster
Record 1 1 1
Record 2 1 2
Record 3 2 3
Record 4 1 4
Record 5 1 5
Record 6 1 6
Record 7 2 7
Record 8 2 8
Record 9 2 9
Record 10 2 10
Sheet1 HC_Output HC_Clusters HC_Dendrogram
Ready
100%


cluster - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Data Mining: Hierarchical Clustering - Predicted Clusters
Output Navigator
Cluster Labels Dendrogram Inputs Clustering Stages
Elapsed Times in Mi
Data Reading Time gorit
0
Record ID Cluster Sub-Cluster
Record 1 1 1
Record 2 1 2
Record 3 2 3
Record 4 1 4
Record 5 1 5
Record 6 1 6
Record 7 2 7
Record 8 2 8
Record 9 2 9
Record 10 2 10
Sheet1 HC_Output HC_Clusters HC_Dendrogram
Ready
100%


cluster - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Data Mining: Hierarchical Clustering - Predicted Clusters
Output Navigator
Cluster Labels Dendrogram Inputs Clustering Stages
Elapsed Times in Mi
Data Reading Time gorit
0
Cluster Labels
Record ID Cluster Sub-Cluster
Record 1 1 1
Record 2 1 2
Record 3 2 3
Record 4 1 4
Record 5 1 5
Record 6 1 6
Record 7 2 7
Record 8 2 8
Record 9 2 9
Record 10 2 10
Sheet1 HC_Output HC_Clusters HC_Dendrogram
Ready
100%

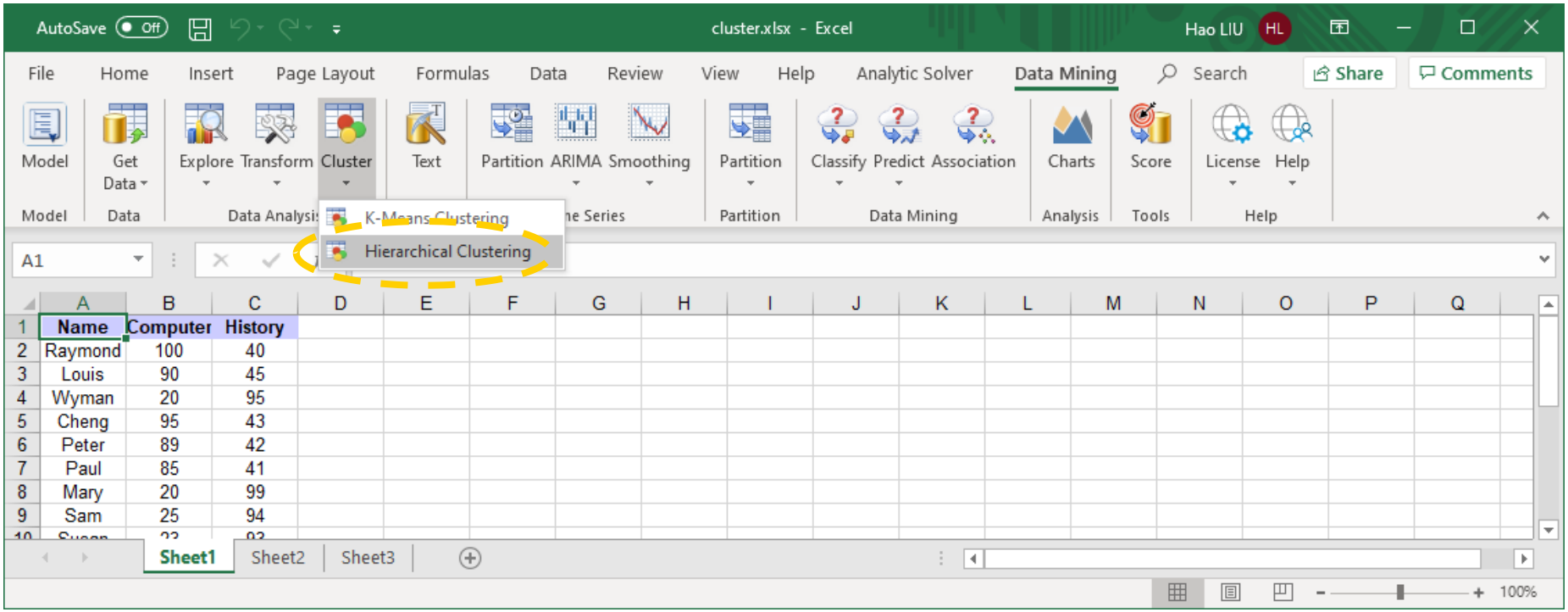

cluster - Excel
LIU Hao
File
Share
Model
Model
A1
Dendrogram
Distance
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
8
9
3
10
7
1
2
5
6
4
Cluster
Sheet1
HC_Output
HC_Clusters
HC_Dendrogram
Ready
100%

How to use the data mining tool

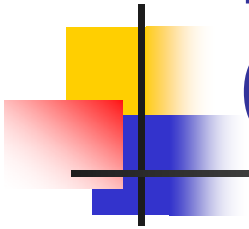
- We have the following 2 versions.
  - XLMiner Desktop (installed in either the CSE lab machine or your computer)
  - XLMiner Cloud (installed as a plugin in your Office 365 Excel)

Where can I find the data mining tool? (XLMiner Cloud)

- The way of opening Hierarchical Clustering in XLMiner Cloud plugin in your Office 365 Excel
  - "Data Mining" Tag → Cluster → Hierarchical Clustering

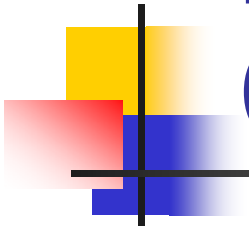


How to use the data mining tool (XLMiner Cloud)

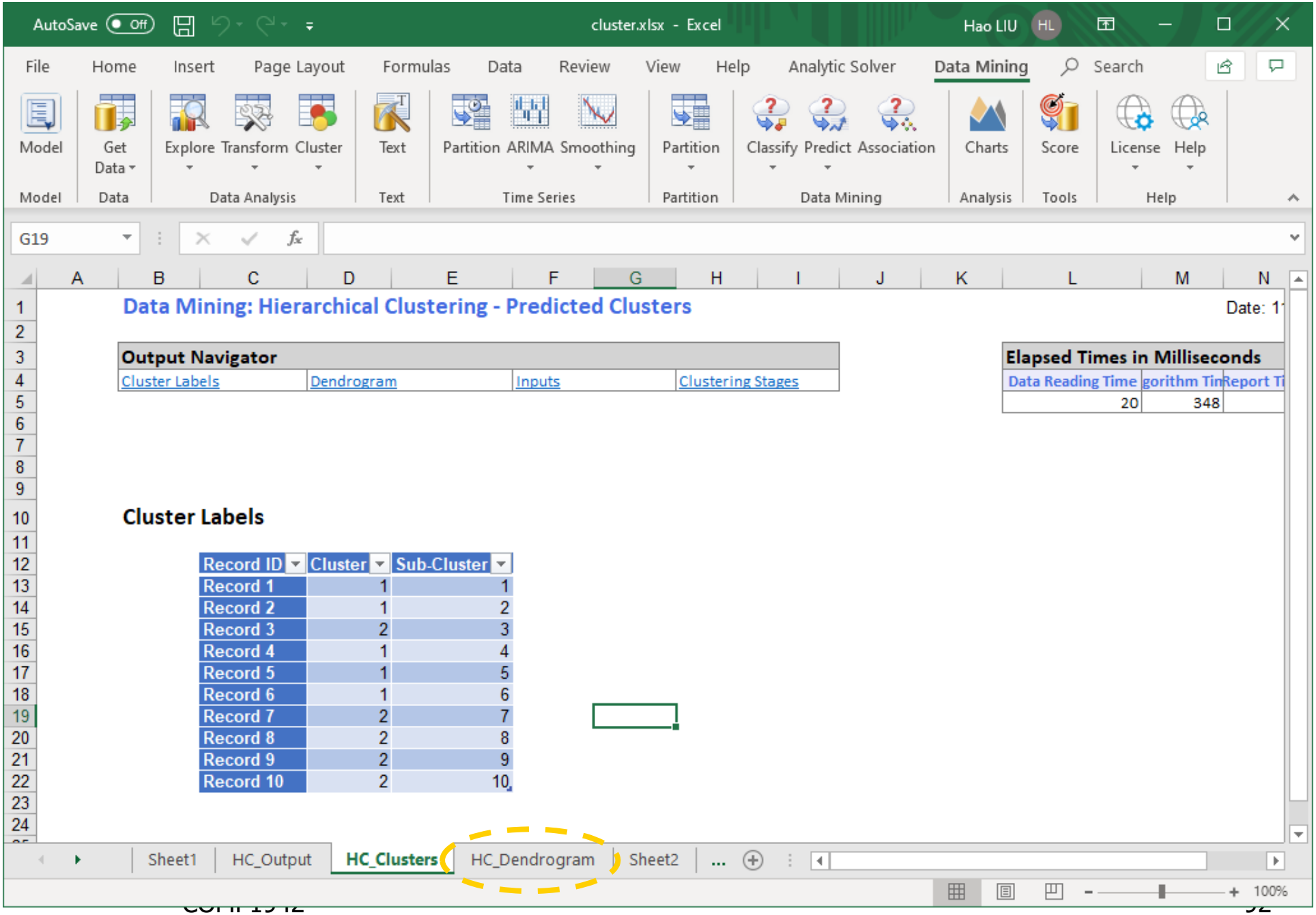


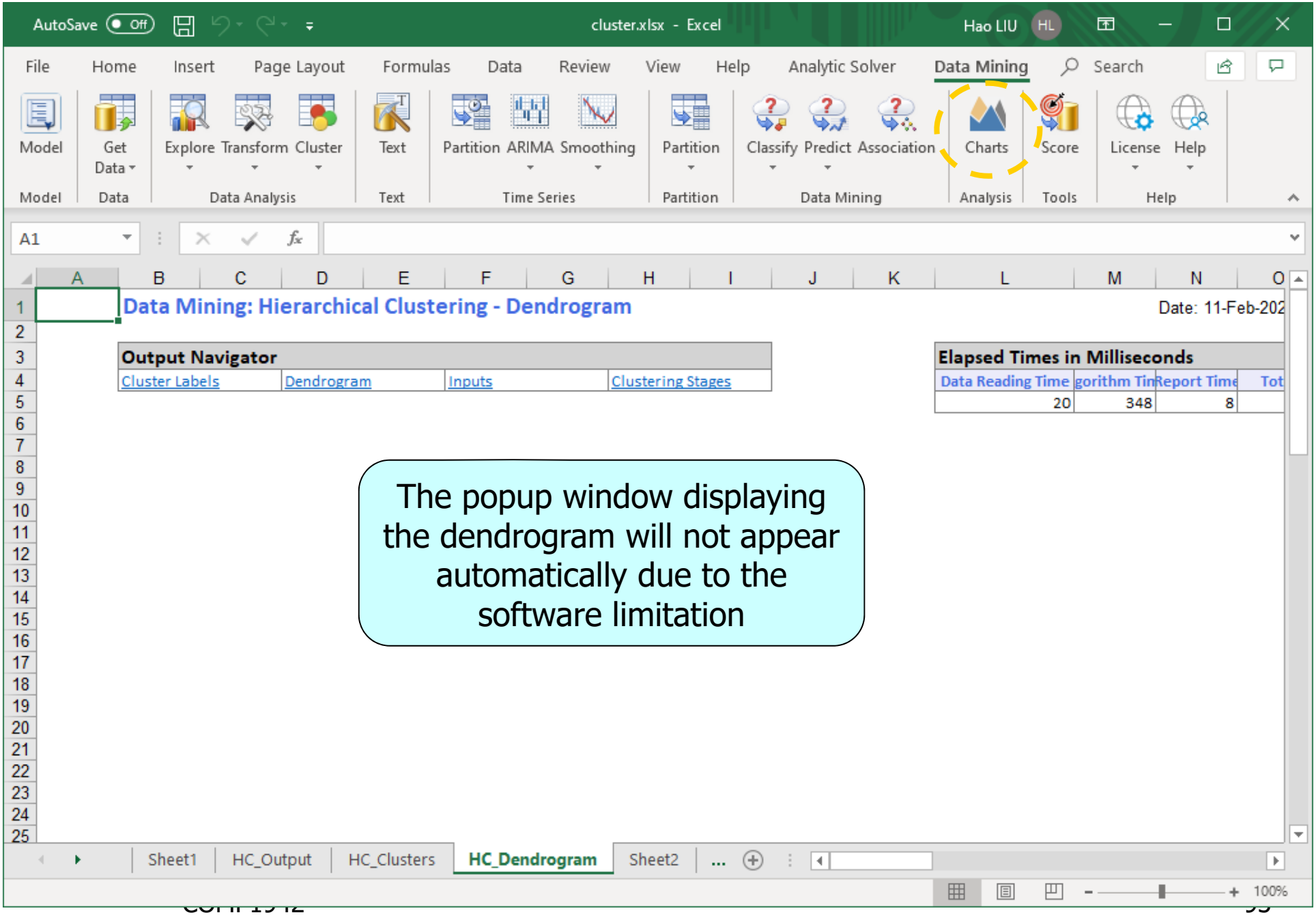
- The steps of performing “hierarchical clustering” in XLMiner Cloud is similar to the steps in XLMiner Desktop.
- The clustering result of XLMiner Cloud is the same as that from XLMiner Desktop.

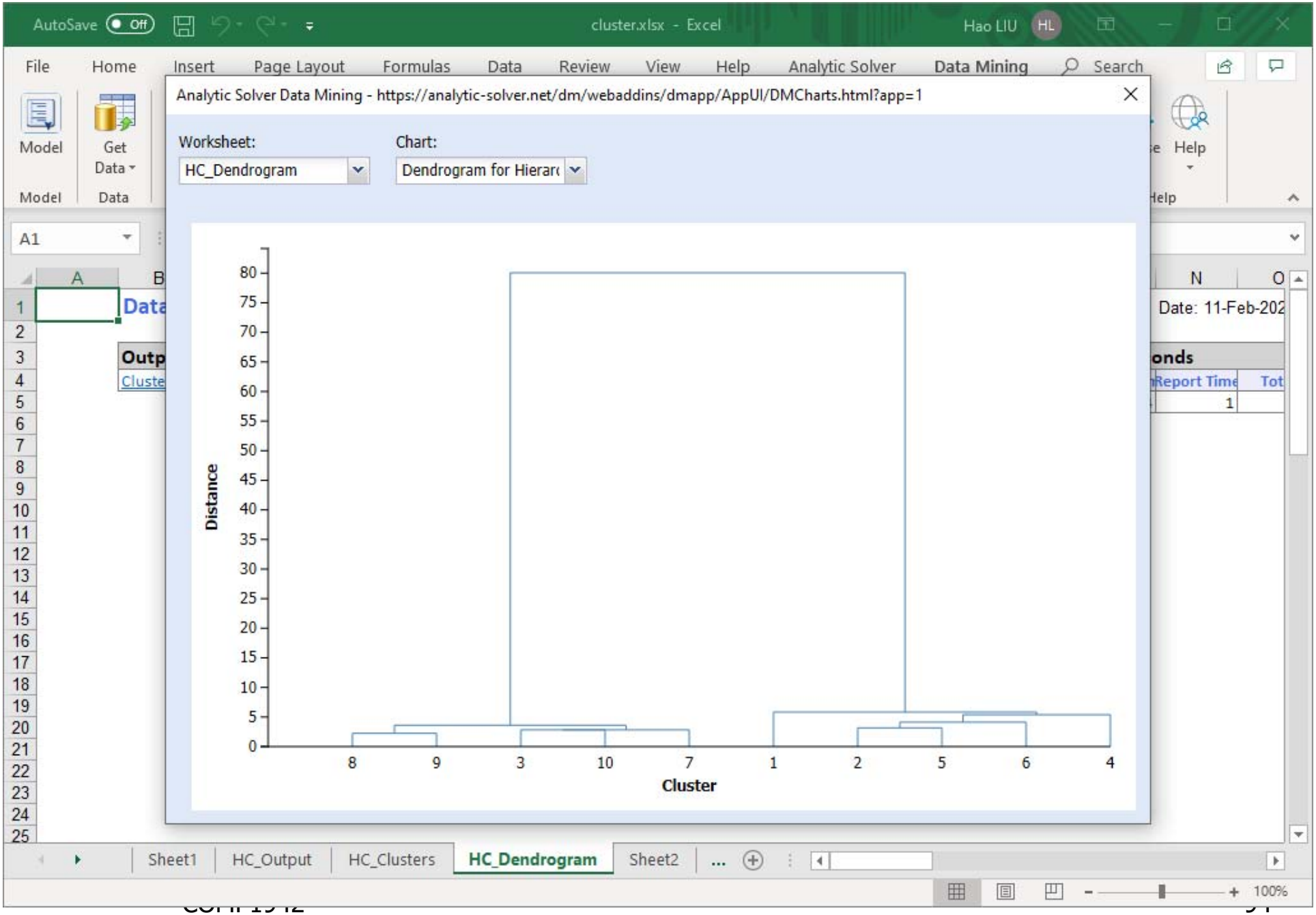
How to use the data mining tool (XLMiner Cloud)



- The output format of XLMiner Cloud is similar to the output in XLMiner Desktop.
- However, to display the dendrogram, you need to call out the “Charts” window.


AutoSave Off
cluster.xlsx - Excel
Hao LIU HL
File Home Insert Page Layout Formulas Data Review View Help Analytic Solver Data Mining Search
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Association Charts Score License Help
Model Data Data Analysis Text Time Series Partition Data Mining Analysis Tools Help
G19
Data Mining: Hierarchical Clustering - Predicted Clusters
Date: 1
Output Navigator
Cluster Labels Dendrogram Inputs Clustering Stages
Elapsed Times in Milliseconds
Data Reading Time gorithm Tin Report Ti
20 348
Cluster Labels
Record ID Cluster Sub-Cluster
Record 1 1 1
Record 2 1 2
Record 3 2 3
Record 4 1 4
Record 5 1 5
Record 6 1 6
Record 7 2 7
Record 8 2 8
Record 9 2 9
Record 10 2 10
Sheet1 HC_Output HC_Clusters HC_Dendrogram Sheet2
100%


AutoSave Off
cluster.xlsx - Excel
Hao LIU HL
File Home Insert Page Layout Formulas Data Review View Help Analytic Solver Data Mining Search
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Association Charts Score License Help
Model Data Data Analysis Text Time Series Partition Data Mining Analysis Tools Help
A1
Data Mining: Hierarchical Clustering - Dendrogram
Date: 11-Feb-202
Output Navigator
Cluster Labels Dendrogram Inputs Clustering Stages
Elapsed Times in Milliseconds
Data Reading Time gorithm Tin Report Time Tot
20 348 8
The popup window displaying the dendrogram will not appear automatically due to the software limitation
Sheet1 HC_Output HC_Clusters HC_Dendrogram Sheet2
100%


cluster.xlsx - Excel
Analytic Solver Data Mining - https://analytic-solver.net/dm/webaddins/dmapp/AppUI/DMCharts.html?app=1
Worksheet:
HC_Dendrogram
Chart:
Dendrogram for Hierar
Distance
Cluster
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
8
9
3
10
7
1
2
5
6
4
Sheet1
HC_Output
HC_Clusters
HC_Dendrogram
Sheet2