

# Help for Excel

---

Basics to use Excel are available online on [\[Excel for Windows training\]](#). You can also use the [Excel help center](#). In addition, many tutorials can be found online. You only have to write the type of task you want to compute with Excel on Google.

*In this document you will find some help to compute some of the tasks useful for the course F78SC with Excel. An important point to keep in mind is that to compute most of the required tasks, you will need to reorder and/or reorganise the data.*

*To have access to all the necessary tools, you need to allow the Add-ins "Analysis ToolPak" and "Solver Add-In".*

*Remark: If you start copying a formula from one cell to neighbouring cells manually, cell by cell, it means that you missed a useful command... Look for it !*

# Help for Excel

---

Summary statistics

# Summary statistics

The screenshot shows a Microsoft Excel interface with the following steps highlighted:

- Step 1:** The **Data** tab is selected in the ribbon.
- Step 2:** The **Data Analysis** button is selected in the Data Tools group.
- Step 3:** The **Descriptive Statistics** option is selected in the list of analysis tools.
- Step 4:** The **OK** button is selected in the Data Analysis dialog box.

The Excel spreadsheet contains the following data:

	Gender	Age	Height
1	M	23	191
2	M	24	182
3	F	23	168
4	F	23	170
5	F	22	172
6	M	24	180
7	F	25	167
8	M	24	183
9	M	24	179
10	M	21	181
11	M	22	181
12	F	24	170
13	F	24	172
14	M	25	185
15	F	22	165
16	F	22	175
17	M	23	183

Excel status bar at the bottom: Average: 176.7058824 Count: 17 Sum: 3004

# Summary statistics

Home Insert Draw Page Layout Formulas D

From HTML From Text New Database Query Refresh All Connections Properties Edit Links

C2 A B C D E F

1 Gender Age Height  
2 M 23 191  
3 M 24 182  
4 F 23 168  
5 F 23 170  
6 F 22 172  
7 M 24 180  
8 F 25 167  
9 M 24 183  
10 M 24 179  
11 M 23 181  
12 M 23 181  
13 F 24 170  
14 F 24 172  
15 M 23 185  
16 F 23 165  
17 F 22 175  
18 M 23 183

1 – Select frame

2 - Select cells of interest

Descriptive Statistics

Input Range:  OK Cancel

Grouped By:  Columns  Rows

Labels in first row

Output options

Output Range:   
 New Worksheet Ply:  Summary statistics  
 New Workbook  
 Summary statistics: 95 %  
 Confidence Level for Mean:  
 Kth Largest: 1  
 Kth Smallest: 1

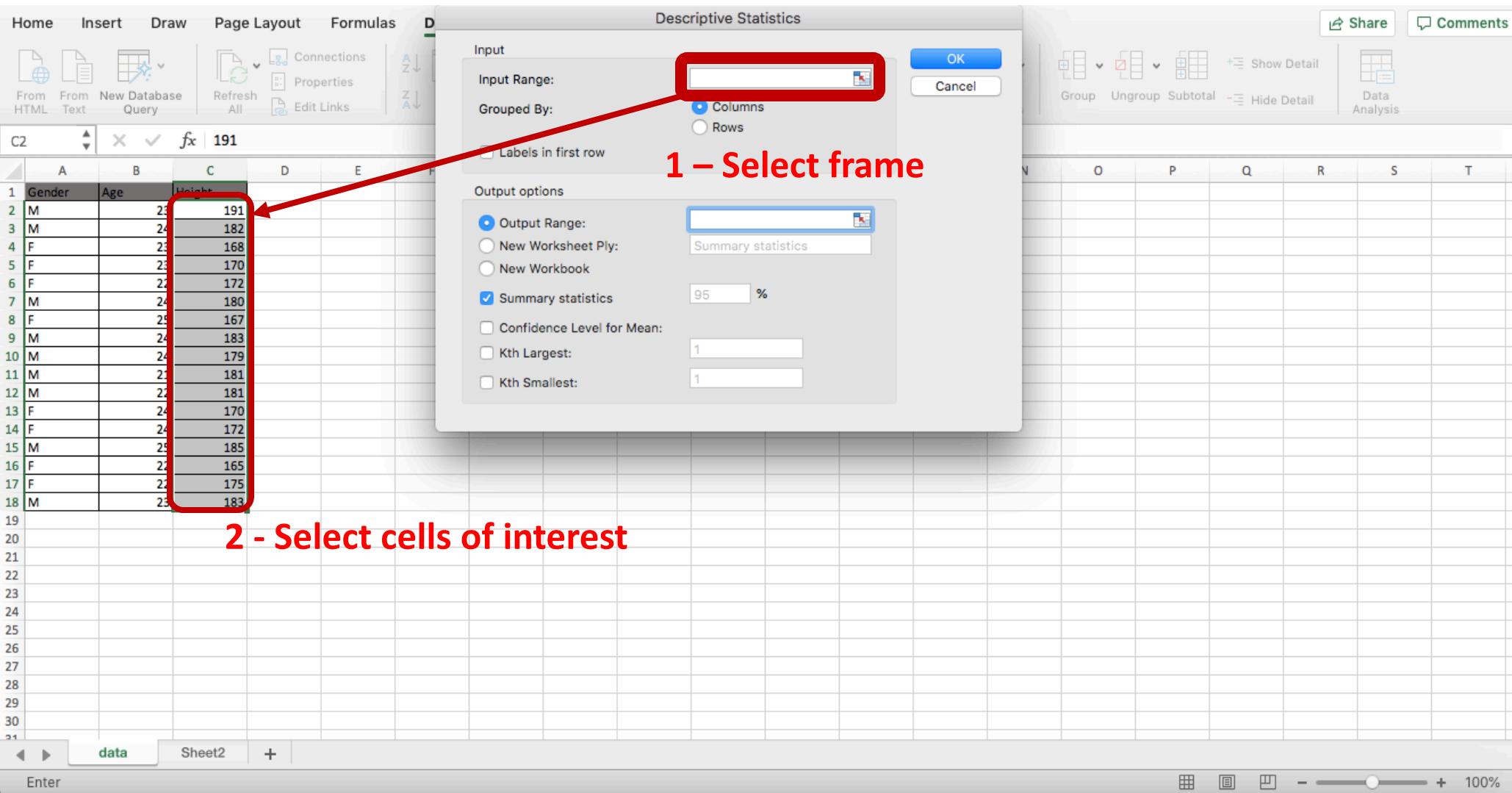
Share Comments

Group Ungroup Subtotal Hide Detail Data Analysis

N O P Q R S T

data Sheet2 +

Enter 100%



# Summary statistics

The screenshot shows the 'Descriptive Statistics' dialog box in Excel. The 'Input Range' is set to '\$C\$2:\$C\$18'. The 'Grouped By' option is set to 'Columns'. The 'Output options' section has 'Output Range:' set to '\$E\$1', 'Summary statistics' checked, and 'Confidence Level for Mean:' set to 95%. Red annotations provide instructions: 'Select Columns' points to the 'Grouped By' radio button; 'Only select if the label appears in the selected first row' points to the 'Labels in first row' checkbox; 'Select Summary statistics' points to the 'Summary statistics' checkbox; and 'Select the frame, and select the cell where the results should appear' points to the 'Output Range' dropdown.

Home Insert Draw Page Layout Formulas D

From HTML From Text New Database Query Refresh All Properties Edit Links

fx | 191

	A	B	C	D	E
1	Gender	Age	Height		
2	M	23	191		
3	M	24	182		
4	F	23	168		
5	F	23	170		
6	F	22	172		
7	M	24	180		
8	F	25	167		
9	M	24	183		
10	M	24	179		
11	M	21	181		
12	M	22	181		
13	F	24	170		
14	F	24	172		
15	M	25	185		
16	F	22	165		
17	F	22	175		
18	M	23	183		

Descriptive Statistics

Input

Input Range: \$C\$2:\$C\$18

Grouped By:  Columns  Rows

Labels in first row

Output options

Output Range: \$E\$1

New Worksheet Ply:

New Workbook

Summary statistics

95 %

Confidence Level for Mean:

Kth Largest: 1

Kth Smallest: 1

OK Cancel

Share Comments

Group Improve Subtotal Data Analysis

From HTML From Text New Database Query Refresh All Properties Edit Links

fx | 191

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

data Sheet2 +

Point 100%

# Summary statistics

Screenshot of Microsoft Excel showing summary statistics for Height.

The Data tab is selected in the ribbon. A red arrow points to the first column of the summary statistics table, labeled "Column1". A red box highlights the value "7.037E-241" in cell F17.

**Summary Statistics for Height:**

	Mean	Standard Err	Median	Mode	Standard Dev	Sample Vari	Kurtosis	Skewness	Range	Minimum	Maximum	Sum	Count
1	176.705882	1.79616566	179	170	7.40578073	54.8455882	-0.94263	0.06098873	26	165	191	3004	17
16													7.037E-241

A red box highlights the value "7.037E-241" in cell F17.

**Remove**

Excel status bar: Average: 307.0670597 Count: 27 Sum: 3991.871776

# Help for Excel

---

## Box plot

We propose to create two box plots on the same figure. One for the Height for Males, and one for the Height for Females.

# Box plot

Screenshot of Microsoft Excel showing a data table and a red annotation.

The Excel ribbon is visible at the top, with the Data tab selected. The formula bar shows "B1" and "Box plot Gender/Height". The data table in the worksheet has columns "Gender" and "Height". The first row contains the column headers "Gender" and "Height". Rows 5 through 21 contain data points, with Gender values M, F, and blank cells, and Height values ranging from 165 to 191.

A red annotation is overlaid on the right side of the data table, reading:

If it is not already the case, organise the columns such that the first column contains the categorical data

The bottom of the screen shows the Excel interface, including the status bar, tabs, and zoom controls.

	Gender	Height
5	M	191
6	M	182
7	F	168
8	F	170
9	F	172
10	M	180
11	F	167
12	M	183
13	M	179
14	M	181
15	M	181
16	F	170
17	F	172
18	M	185
19	F	165
20	F	175
21	M	183

# Box plot

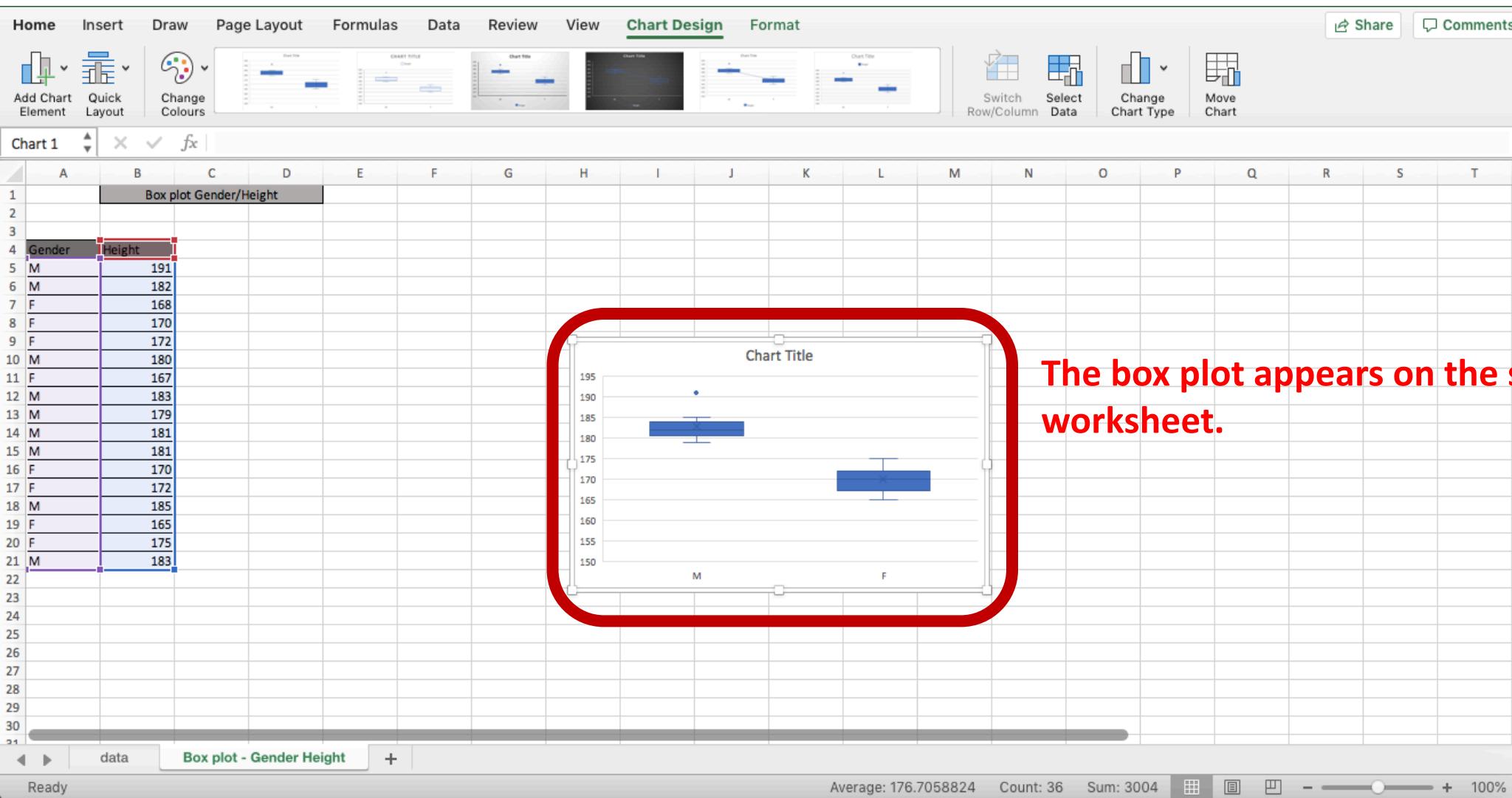
The screenshot shows a Microsoft Excel spreadsheet with the following steps highlighted:

- Step 1:** A red box highlights the "Insert" tab on the ribbon.
- Step 2:** A red box highlights the "Recommended Charts" icon in the "Charts" group.
- Step 3:** A red box highlights the "Box and Whisker" chart type in the dropdown menu.
- Step 4:** A red box highlights the "Box and Whisker" chart preview in the dropdown menu.

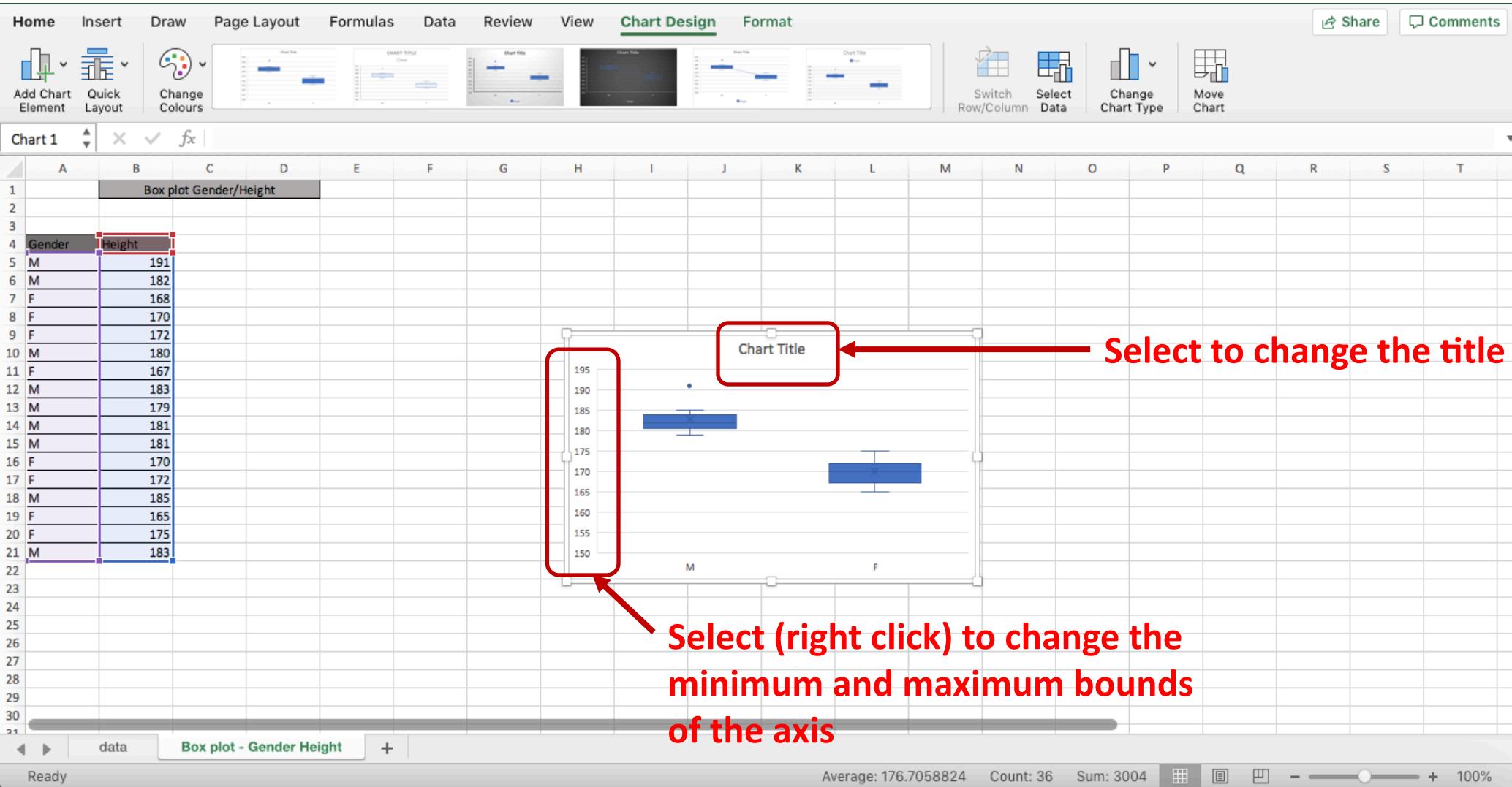
The spreadsheet contains a table with columns "Gender" and "Height". The table has 21 rows of data. The first row is a header. The data starts from row 5, with Gender values M, M, F, F, F, M, F, M, M, M, M, F, M, M, F, M, F, M, F, F, M and Height values 191, 182, 168, 170, 172, 180, 167, 183, 179, 181, 181, 170, 172, 185, 165, 175, 183 respectively.

**1 – Select the two columns of interest,  
with or without the labels**

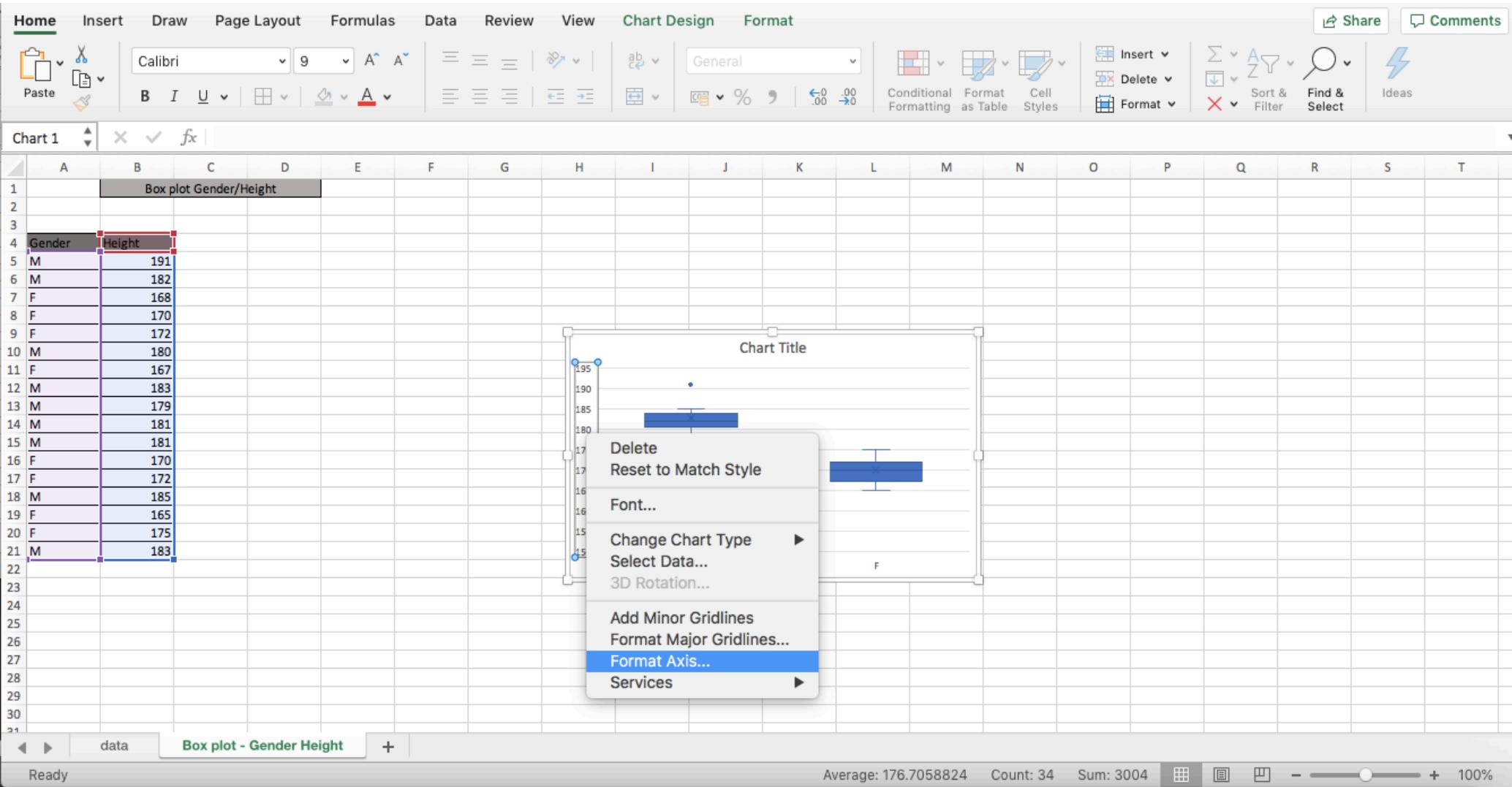
# Box plot



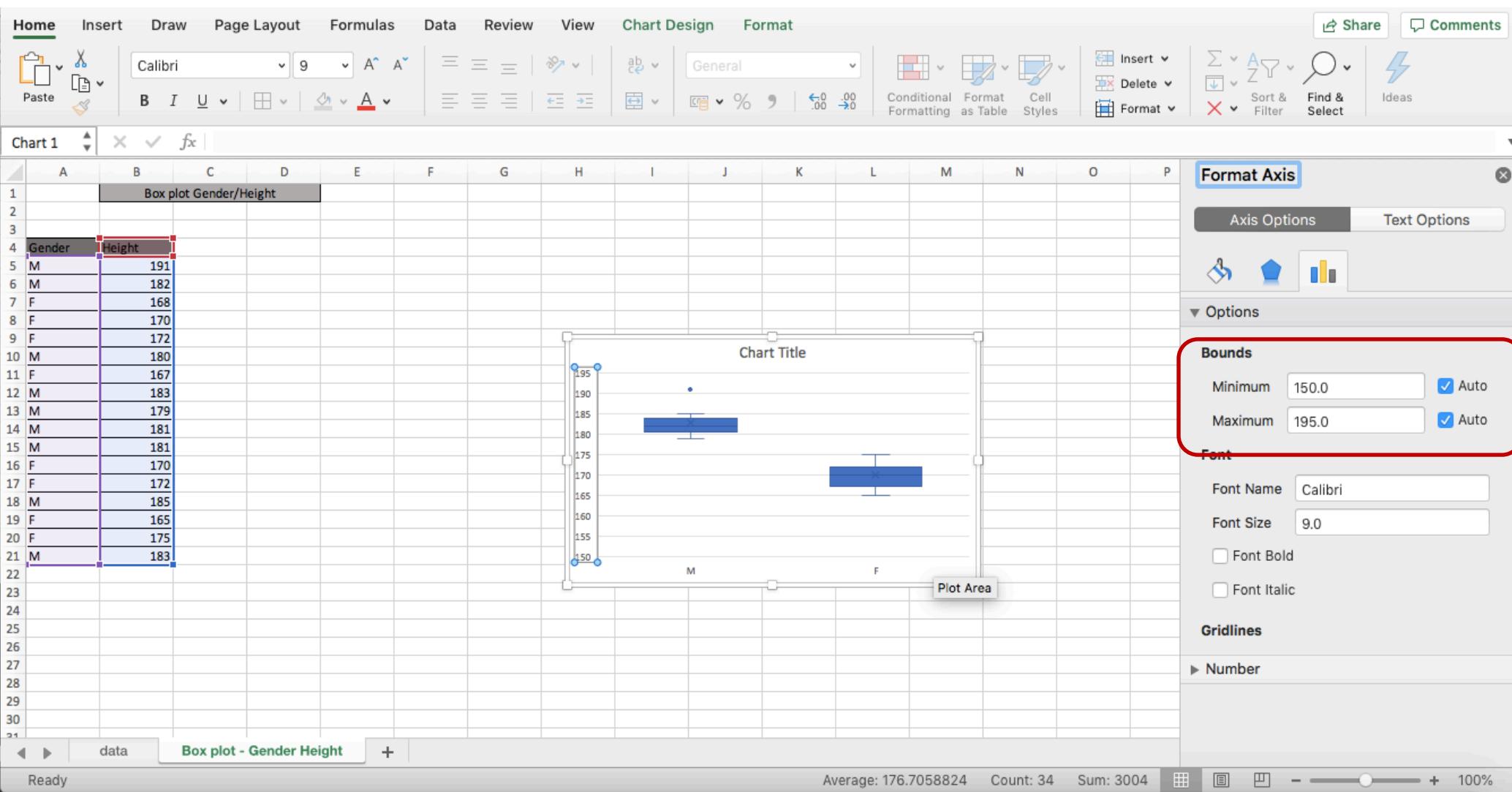
# Box plot



# Box plot

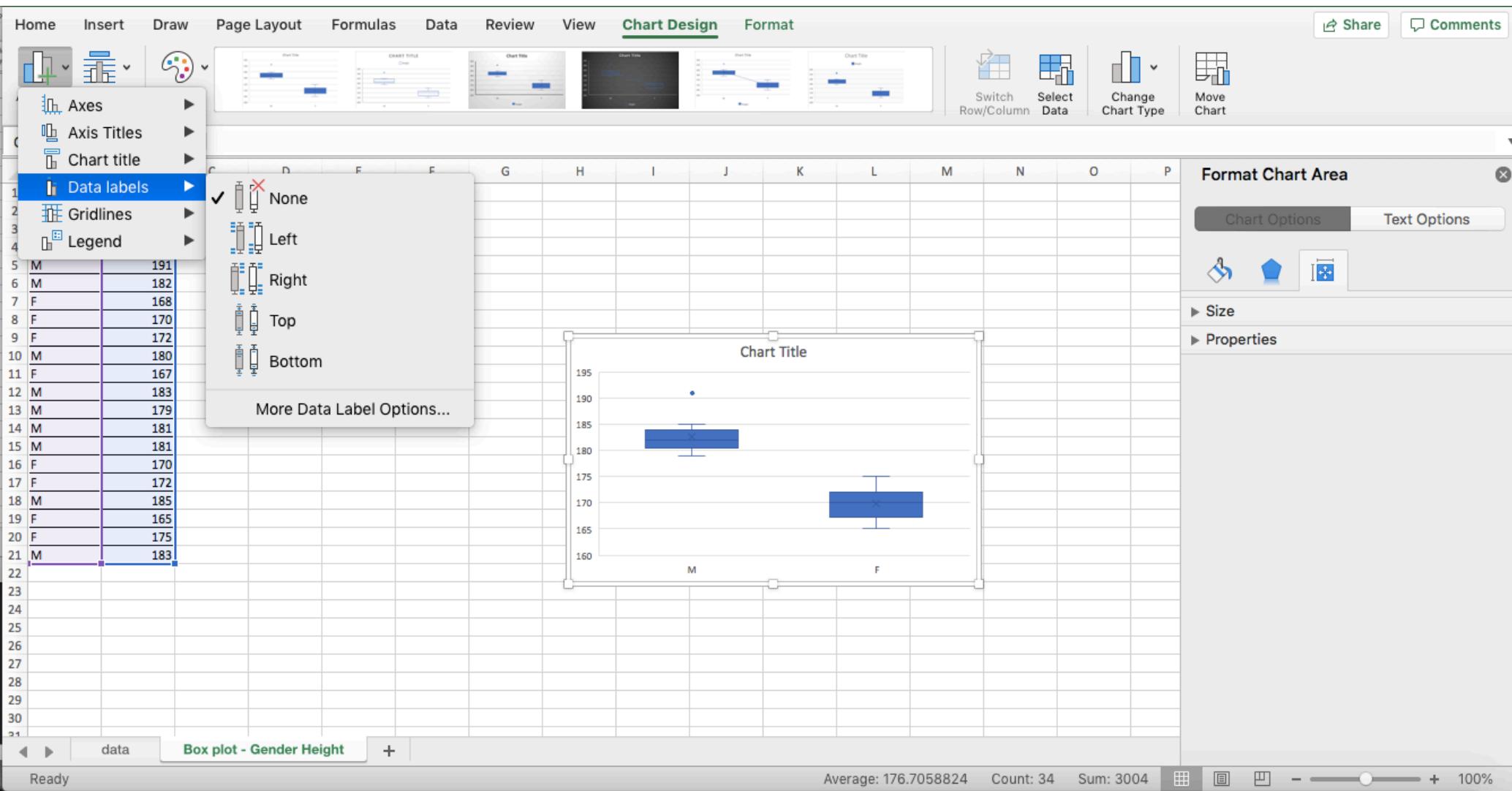


# Box plot



Modify the values  
to change the  
maximum and  
minimum bounds  
of the axis

# Box plot



Use this option to add values on the box plot

# Help for Excel

---

## t-tests

We do not provide a specific example, but we show where to find the tools to perform t-tests with Excel.

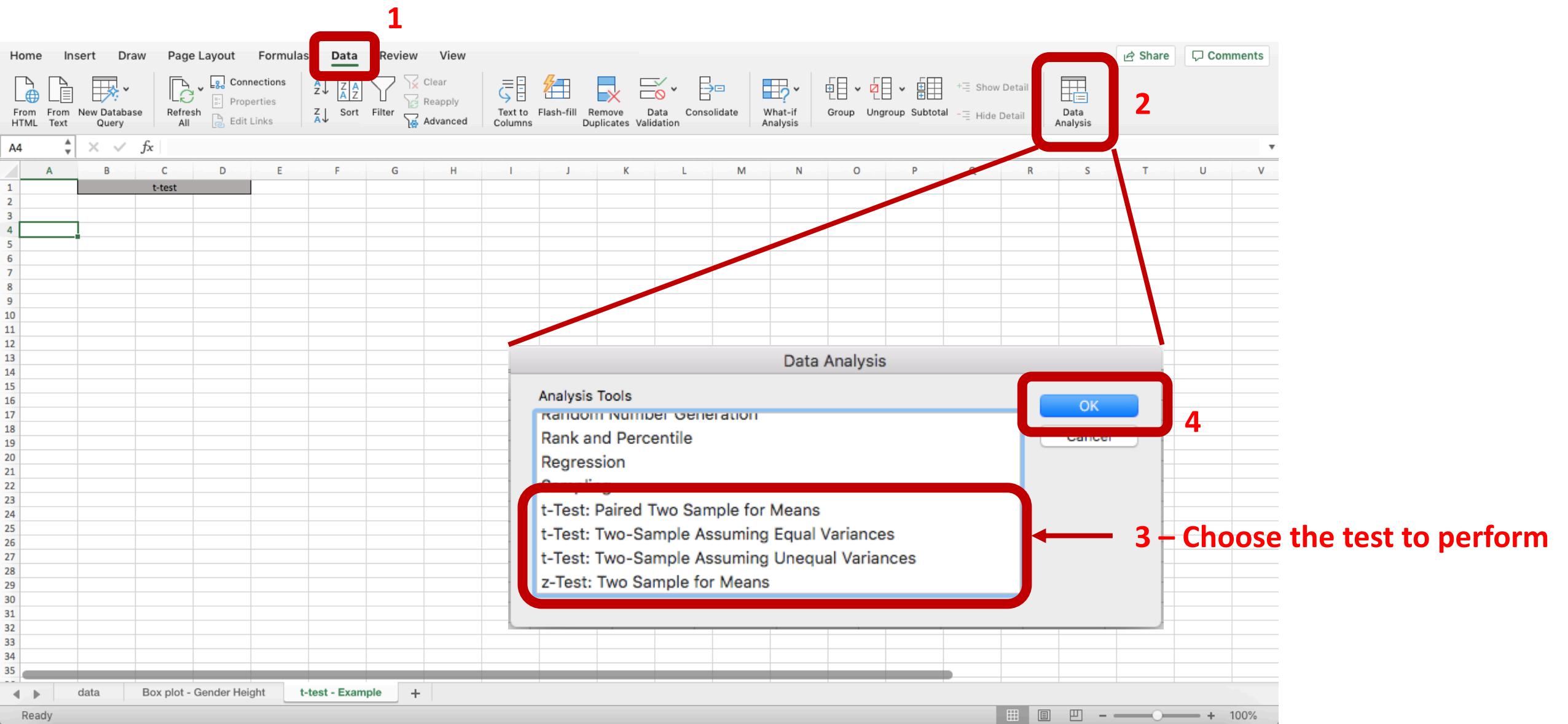
Most of the time, to compute a t-test with Excel, data will need to be reorganised.

# t-test

The screenshot shows a Microsoft Excel spreadsheet titled "t-test". The "Data" tab is selected in the ribbon. The range A1:A4 is selected, and the formula bar shows "fx". The cell A1 contains "t-test". The status bar at the bottom indicates "Ready". A red box highlights the tab bar, specifically the "t-test - Example" tab.

Most of the time, you may need to create a new worksheet to copy only the data you need to perform the t-test, and organise them if necessary

# t-test



# t-test

Home Insert Draw Page Layout Formulas **Data** Review View

From HTML From Text New Database Query Refresh All Properties Edit Links

Connections Reapply

A4 X ✓ fx |

Q R S T U V

**t-Test: Paired Two Sample for Means**

**Input**

Variable 1 Range:

Variable 2 Range:

Hypothesized Mean Difference:

Labels

Alpha: 0.05

Output options

Output Range:

New Worksheet Ply:

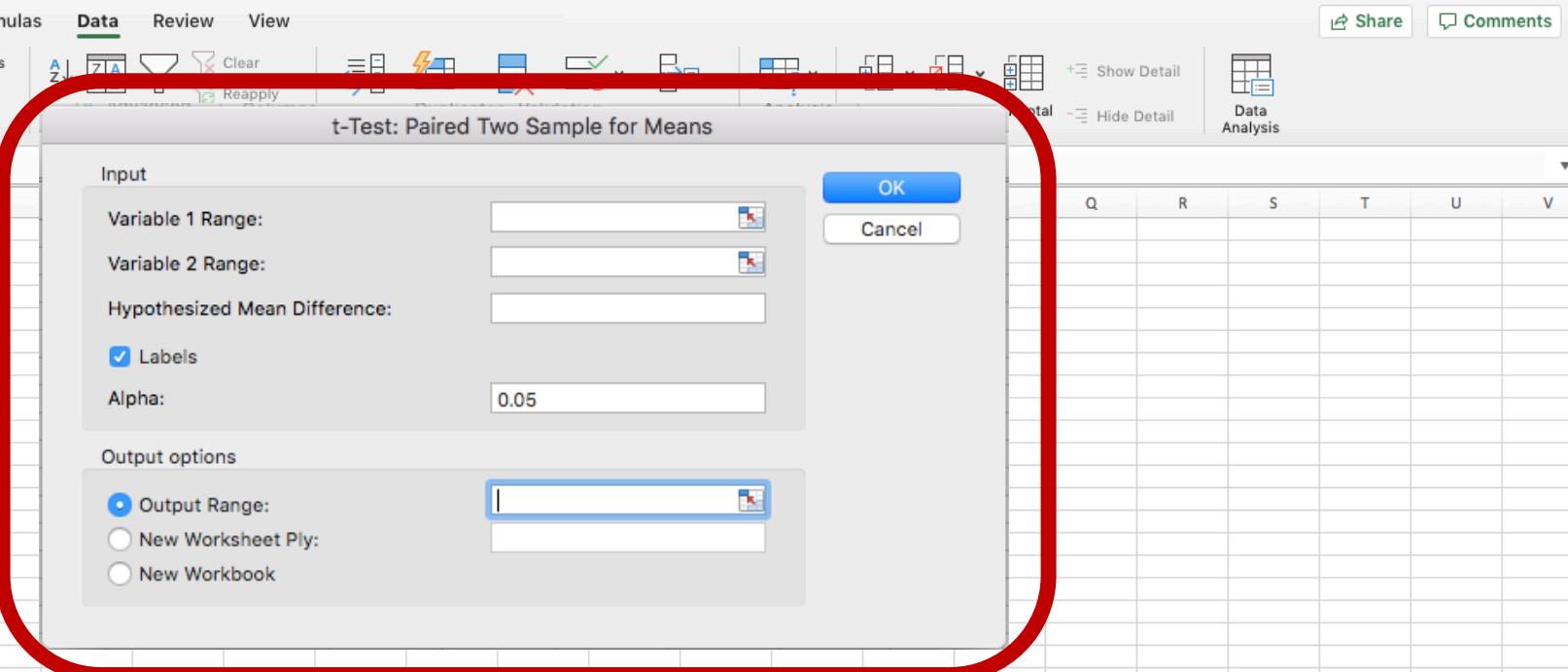
New Workbook

OK Cancel

Example of pop-up window appearing when selecting a t-test

data Box plot - Gender Height t-test - Example +

Ready 100%



# t-test

Home Insert Draw Page Layout Formulas **Data** Review View

From HTML From Text New Database Query Refresh All Properties Edit Links

A4 **t-test**

**Hypothesized value for the test (often will be 0)**

**Example of pop-up window appearing when selecting a t-test**

**Select in the worksheet the variables to be used to compute the t-test**

**t-Test: Paired Two Sample for Means**

**Input**

Variable 1 Range:  Variable 2 Range:   
Hypothesized Mean Difference:

OK Cancel

Labels

Alpha: 0.05

Output options

Output Range:   
 New Worksheet Ply:  
 New Workbook

# t-test

Home Insert Draw Page Layout Formulas **Data** Review View

From HTML From Text New Database Query Refresh All Properties Edit Links

A4 **X** **fx**

**Select if the labels appears in the first cells of your selected columns**

**OK** **Cancel**

**t-Test: Paired Two Sample for Means**

**Input**

Variable 1 Range:

Variable 2 Range:

Hypothesized Mean Difference:

Labels

Alpha:

**Level of significance**

Output options

Output Range:

New Worksheet Ply:

New Workbook

**Select the cell in the worksheet where the result of the t-test should appear**

Q R S T U V

Example of pop-up window appearing when selecting a t-test

data Box plot - Gender Height **t-test - Example** +

Ready 100%

# t-test

Home Insert Draw Page Layout Formulas **Data** Review View

From HTML From Text New Database Query Refresh All Properties Edit Links

A4 **X** **fx**

**t-Test: Paired Two Sample for Means**

**Input**

Variable 1 Range:

Variable 2 Range:

Hypothesized Mean Difference:

Labels

Alpha: 0.05

Output options

Output Range:

New Worksheet Ply:

New Workbook

OK Cancel

**Remark: Excel will provide both 1-sided and 2-sided tests (denoted one-tail and two-tail, respectively)**

Example of pop-up window appearing when selecting a t-test

data Box plot - Gender Height **t-test - Example** +

Ready 100%

The screenshot shows a Microsoft Excel interface with a 't-Test: Paired Two Sample for Means' dialog box open. The dialog box has fields for 'Variable 1 Range', 'Variable 2 Range', 'Hypothesized Mean Difference', 'Labels' (checked), 'Alpha' (0.05), and 'Output options' (radio buttons for 'Output Range', 'New Worksheet Ply', and 'New Workbook'). The 'Output Range' option is selected. The background shows a spreadsheet with data in columns A, B, C, D, and E, and a chart titled 'Box plot - Gender Height'. A red box highlights the text 'Remark: Excel will provide both 1-sided and 2-sided tests (denoted one-tail and two-tail, respectively)'. Another red box highlights the text 'Example of pop-up window appearing when selecting a t-test'.

# Help for Excel

---

## Dot plot for discrete data

The steps to create a dot plot with Excel are very similar to the steps that should be done if the dot plot was done manually, except that the calculations are done automatically.

In particular, an auxiliary column must be added to first count the occurrence of each value appearing in the dataset of interest.

# Dot plot for discrete data

Screenshot of Microsoft Excel showing a blank worksheet titled "Dot plot - Height Male". The Data tab is selected in the ribbon. A red box highlights the tab bar and the title cell "Dot plot - Height Male".

**Most of the time, you may need to create a new worksheet to copy only the data you need to create a dot plot, and organise them if necessary**

# Dot plot for discrete data

Screenshot of Microsoft Excel showing a dot plot for male height data.

The Data tab is selected in the ribbon.

The worksheet title is "Dot plot - Height Male".

The data range is A5:B22, containing the following data:

Gender	Height
M	191
M	182
F	168
F	170
F	172
M	180
F	167
M	183
M	179
M	181
M	181
F	170
F	172
M	185
F	165
F	175
M	183

A red box highlights the following text:

In this example we want to create a dot plot for the height of males. Thus we copy the two columns: Gender and Height in the new worksheet.

Bottom status bar: Average: 176.7058824 Count: 36 Sum: 3004

# Dot plot for discrete data

Home Insert Draw Page Layout Formulas Data Review View Share Comments

Calibri (Body) 12 A A Wrap Text General Conditional Formatting Insert Sort & Filter

B I U Merge & Centre Format as Table Delete Find & Select Cell Styles Ideas

E5 fx Height of males

A B C D E F G H I J K L M N O P Q R S T U

1 Dot plot - Height Male

2

3

4

5 Gender Height Height of males

6 M 191

7 M 182

8 F 168

9 F 170

10 F 172

11 M 180

12 F 167

13 M 183

14 M 179

15 M 181

16 M 181

17 F 170

18 F 172

19 M 185

20 F 165

21 F 175

22 M 183

23

24

25

26

27

28

29

30

31

32

33

34

35

data Box plot - Gender Height t-test - Example Dot plot - Height Male + 100% Ready

Since we only are interested in the height of males, we create a new column that will contain these data.

Gender	Height	Height of males
M	191	
M	182	
F	168	
F	170	
F	172	
M	180	
F	167	
M	183	
M	179	
M	181	
M	181	
F	170	
F	172	
M	185	
F	165	
F	175	
M	183	

# Dot plot for discrete data

Home Insert Draw Page Layout Formulas **Data** Review View

From HTML From Text New Database Query Refresh All Properties Edit Links Sort Filter Advanced

A6 x fx M

Gender	Height
M	191
M	182
F	168
F	170
F	172
M	180
F	167
M	183
M	179
M	181
M	181
F	170
F	172
M	185
F	165
F	175
M	183

Height of males

We need to order the data to separate the males and the females.

1 - Select the cells containing the data Gender

Count: 17

# Dot plot for discrete data

Screenshot of Microsoft Excel showing a data sorting dialog box.

The Data tab is selected in the ribbon, and the Sort icon (highlighted with a red box) is selected.

The current selection is "Dot plot - Height Male" (cell A6).

The data table includes columns "Gender" and "Height".

A "Sort Warning" dialog box is displayed:

- Data outside your current selection won't be sorted.
- What do you want to do?
  - Expand the selection
  - Continue with the current selection
- Buttons: Cancel, Sort

The status bar at the bottom shows "Count: 17".

Gender	Height
M	191
M	182
F	168
F	170
F	172
M	180
F	167
M	183
M	179
M	181
M	181
F	170
F	172
M	185
F	165
F	175
M	183

# Dot plot for discrete data

A14 Share | Comments

Home Insert Draw Page Layout Formulas **Data** Review View

From HTML From Text New Database Query Refresh All Properties Sort Filter Clear Reapply Advanced Text to Columns Flash-fill Remove Duplicates Data Validation Consolidate What-if Analysis Group Ungroup Subtotal Hide Detail Data Analysis

Gender	Height
F	168
F	170
F	172
F	167
F	170
F	172
F	165
F	175
M	191
M	182
M	180
M	183
M	179
M	181
M	181
M	185
M	183

Height of males

- Ordering in alphabetical order will give the data on females (F) first and male (M) after
- Ordering in reversed alphabetical order will give the data on male (M) first and females (F) after

data Box plot - Gender Height t-test - Example Dot plot - Height Male +

Average: 182.7777778 Count: 18 Sum: 1645 100%

# Dot plot for discrete data

Screenshot of Microsoft Excel showing a dot plot for discrete data. The data is organized into two columns: "Gender" and "Height". A new column "Height of males" has been created and populated with the heights of males from the original data.

The data table:

Gender	Height	Height of males
F	168	191
F	170	182
F	172	180
F	167	183
F	170	179
F	172	181
F	165	181
F	175	181
M	191	185
M	182	183
M	180	183
M	183	183
M	179	181
M	181	181
M	181	185
M	183	183

A red box highlights the "Height of males" column, and a red arrow points from the "Height" column to it, indicating the copied data. A red callout box contains the instruction: "Copy the Height of Males in the new column "Height of males".

Excel ribbon tabs: Home, Insert, Draw, Page Layout, Formulas, Data, Review, View.

Excel status bar: Average: 182.7777778 Count: 9 Sum: 1645.

# Dot plot for discrete data

Screenshot of Microsoft Excel showing a dot plot for discrete data. The data is organized into two columns: 'Gender' and 'Height'. A third column, 'Cumulative sum', is added to the right of the data.

The 'Cumulative sum' column shows the running total of heights for each gender. For example, for females (F), the cumulative sum starts at 168 and increases by 1 for each subsequent row. For males (M), it starts at 191 and increases by 1 for each subsequent row.

A red box highlights the 'Cumulative sum' column, and a red callout box contains the text: "Create a new column to find the occurrence of each value appearing in ‘Height of males’".

Gender	Height	Height of males	Cumulative sum
F	168		
F	170		19
F	172		18
F	167		18
F	170		17
F	172		18
F	165		18
F	175		18
M	191		18
M	182		
M	180		
M	183		
M	179		
M	181		
M	181		
M	185		
M	183		

Excel ribbon tabs: Home, Insert, Draw, Page Layout, Formulas, Data, Review, View.

Cell F5 is selected.

Bottom tabs: data, Box plot - Gender Height, t-test - Example, Dot plot - Height Male, RiverMill - Plot 25.

Bottom right: 100% zoom.

# Dot plot for discrete data

2 – Write in the formula window

1 – Select the first cell of the column

COUNT  =

Gender	Height
F	168
F	170
F	172
F	167
F	170
F	172
F	165
F	175
M	191
M	182
M	180
M	183
M	179
M	181
M	181
M	185
M	183

data Box plot - Gender Height t-test - Example Dot plot - Height Male + RiverMill - Plot 25 Edit 100%

# Dot plot for discrete data

Use the formula below, and adapt to select the first cell in the column “Height of males”

Gender	Height	Height of males	Cummulative sum
F	168	191	=COUNTIF(\$E\$6:\$E6,\$E6)
F	170	182	
F	172	180	
F	167	183	
F	170	179	
F	172	181	
F	165	181	
F	175	185	
M	191	183	
M	182		
M	180		
M	183		
M	179		
M	181		
M	181		
M	185		
M	183		

The formula in cell E6 is =COUNTIF(\$E\$6:\$E6,\$E6)

Excel ribbon tabs: Home, Insert, Draw, Page Layout, Formulas, Data, Review, View.

Formulas bar: COUNT, fx, =COUNTIF(\$E\$6:\$E6,\$E6)

Cells A1 to U35 are empty.

Bottom navigation: data, Box plot - Gender Height, t-test - Example, Dot plot - Height Male, +, Edit, 100%.

# Dot plot for discrete data

Home Insert Draw Page Layout Formulas Data Review View

Calibri (Body) 12 A A Wrap Text General Conditional Formatting Insert Sort & Filter Share Comments

B I U Merge & Centre % Cell Styles Delete Find & Select Ideas

F6 x fx =COUNTIF(\$E\$6:\$E\$6,\$E6)

Gender	Height	Height of males	Cummulative sum
F	168	191	1
F	170	182	
F	172	180	
F	167	183	
F	170	179	
F	172	181	
F	165	181	
F	175	185	
M	191	183	
M	182		
M	180		
M	183		
M	179		
M	181		
M	181		
M	185		
M	183		

Click on the small square on the bottom right corner of the cell, and pull to the bottom of the column

data Box plot - Gender Height t-test - Example Dot plot - Height Male +

Ready 100%

# Dot plot for discrete data

Home Insert Draw Page Layout Formulas Data Review View

Calibri (Body) 12 A A Wrap Text General Conditional Formatting Insert Sort & Filter Share Comments

B I U Merge & Centre Format as Table Delete Find & Select Ideas

F6 x fx =COUNTIF(\$E\$6:\$E\$6,\$E6)

Gender	Height	Height of males	Cummulative sum
F	168	191	1
F	170	182	1
F	172	180	1
F	167	183	1
F	170	179	1
F	172	181	1
F	165	181	2
F	175	185	1
M	191	183	2
M	182		
M	180		
M	183		
M	179		
M	181		
M	181		
M	185		
M	183		

data Box plot - Gender Height t-test - Example Dot plot - Height Male +

Average: 1.222222222 Count: 9 Sum: 11

Ready 100%

# Dot plot for discrete data

2

3

E6      x      fx | 191

A B C D E F G H I J K L M N O P Q R S T U

1 Dot plot - Height Male

2

3

4

5 Gender Height

6 F 168

7 F 170

8 F 172

9 F 167

10 F 170

11 F 172

12 F 165

13 F 175

14 M 191

15 M 182

16 M 180

17 M 183

18 M 179

19 M 181

20 M 181

21 M 185

22 M 183

23

24

25

26

27

28

29

30

31

32

33

34

35

Height of males Cumulative sum

191 1

182 1

180 1

183 1

179 1

181 1

181 2

185 1

183 2

1 – Select all the cells in the created tab

data Box plot - Gender Height t-test - Example Dot plot - Height Male +

Average: 92 Count: 18 Sum: 1656

Ready

Gender	Height	Height of males	Cumulative sum
F	168	191	1
F	170	182	1
F	172	180	1
F	167	183	1
F	170	179	1
F	172	181	1
F	165	181	2
F	175	185	1
M	191	183	2
M	182		
M	180		
M	183		
M	179		
M	181		
M	181		
M	185		
M	183		

# Dot plot for discrete data

Home Insert Draw Page Layout Formulas Data Review View

E6 A B C D E F G

1 Dot plot - Height Male

2

3

4

5 Gender Height

Gender	Height
F	168
F	170
F	172
F	167
F	170
F	172
F	165
F	175
M	191
M	182
M	180
M	183
M	179
M	181
M	181
M	185
M	183

6 Height of males Cumulative sum

191	1
182	1
180	1
183	1
179	1
181	1
181	2
185	1
183	2

78 180 182 184 186 188 190 192

Scatter

Chart Title

191

182

180

183

179

181

181

185

183

178 180 182 184 186 188 190 192

Series1 Series2

Clustered Column

Chart Title

9 8 7 6 5 4

Ready data Box plot - Gender Height t-test - Example Dot plot - Height Male +

Average: 92 Count: 18 Sum: 1656

Share Comments

Select the Scatter chart

The screenshot shows a Microsoft Excel spreadsheet titled "Dot plot - Height Male". The data is organized into two columns: "Gender" and "Height". A secondary table in columns E and F shows the "Height of males" and their "Cumulative sum". The ribbon is visible at the top, with the "Insert" tab selected. A red box highlights the "Scatter" chart icon in the ribbon's chart section. The chart area displays a scatter plot with discrete data points at various heights. A text annotation "Select the Scatter chart" is overlaid on the right side of the chart area.

# Dot plot for discrete data

Home Insert Draw Page Layout Formulas Data Review View **Chart Design** Format

Add Chart Element Quick Layout Change Colours

Chart Title Chart Title Chart Title Chart Title Chart Title Chart Title Chart Title

Share Comments

Chart 1 ▾ × ✓ fx |

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Dot plot - Height Male																			
2																				
3																				
4																				
5	Gender	Height		Height of males	Cummulative sum															
6	F	168		191	1															
7	F	170		182	1															
8	F	172		180	1															
9	F	167		183	1															
10	F	170		179	1															
11	F	172		181	1															
12	F	165		181	2															
13	F	175		185	1															
14	M	191		183	2															
15	M	182																		
16	M	180																		
17	M	183																		
18	M	179																		
19	M	181																		
20	M	181																		
21	M	185																		
22	M	183																		
23																				
24																				
25																				
26																				
27																				
28																				
29																				
30																				
31																				
32																				
33																				
34																				
35																				

data Box plot - Gender Height t-test - Example Dot plot - Height Male +

Average: 92 Count: 18 Sum: 1656

Ready

Dot plot for discrete data

# Help for Excel

---

## Regression

Example to perform a regression, with analysis of the results.

# Regression

The screenshot shows a Microsoft Excel spreadsheet with data in columns A and B. The Data tab is selected in the ribbon. A red box labeled '1' highlights the Data tab. A red box labeled '2' highlights the 'Data Analysis' button in the Data tab's ribbon group. A red box labeled '3 – Select Regression' points to the 'Regression' option in the 'Analysis Tools' dialog box. A red box labeled '4' points to the 'OK' button in the same dialog box.

1

2

3 – Select Regression

4

Data Analysis

Analysis Tools

- Rank and Percentile
- Regression**
- Sampling

t-Test: Paired Two-Sample for Means  
t-Test: Two-Sample Assuming Equal Variances  
t-Test: Two-Sample Assuming Unequal Variances  
z-Test: Two Sample for Means

OK Cancel

	A	B	C	D	E	F	G
1	Weight	Height					
2	80	191					
3	76	182					
4	60	168					
5	57	170					
6	65	172					
7	76	180					
8	55	167					
9	80	183					
10	79	179					
11	75	181					
12	77	181					
13	62	170					
14	64	172					
15	78	185					
16	56	165					
17	68	175					
18	78	183					
19							
20							
21							
22							
23							
24							
25							
26							
27							
28							
29							
30							
31							
32							
33							
34							
35							

data Box plot - Gender Height t-test - Example Dot plot - Height Male Regression +

# Regression

The screenshot shows a Microsoft Excel interface with the following details:

- Excel ribbon:** Home, Insert, Draw, Page Layout, Formulas, **Data** (selected), Review, View.
- Toolbar:** From HTML, From Text, New Database Query (highlighted), Refresh, Connections, Properties, Edit Links.
- Worksheet:** A data table with columns A (Weight) and B (Height). Row 14 is selected, with a green border around cells A14:B14.
- Dialog Box:** "Regression" dialog box is open, centered over the worksheet.
  - Input:** Input Y Range: (empty), Input X Range: (empty).
    - Labels
    - Constant is Zero
    - Confidence Level: 95 %
  - Output options:**
    - Output Range: (empty)
    - New Worksheet Ply: (empty)
    - New Workbook
  - Residuals:**
    - Residuals
    - Standardized Residuals
    - Residual Plots
    - Line Fit Plots
  - Normal Probability:**
    - Normal Probability Plots
- Bottom tabs:** data, Box plot - Gender Height, t-test - Example, Dot plot - Height Male, **Regression** (highlighted), +.

In this example we create a regression showing height against weight

# Regression

Home Insert Draw Page Layout Formulas **Data** Review View

From HTML From Text New Database Query Refresh All Properties Edit Links Sort Filter

Regression

**Input Y Range:** \$B\$1:\$B\$18

**Input X Range:**

Labels  Constant is Zero

Confidence Level: 95 %

**Output options:**

Output Range:

New Worksheet Ply:

New Workbook

**Residuals:**

Residuals  Residual Plots

Standardized Residuals  Line Fit Plots

**Normal Probability:**

Normal Probability Plots

OK Cancel

**1 – Select the Y-axis box**

**2 – select the associated column  
(here weight)**

	A	B	C	D	E	F	G
1	Weight	Height					
2	80	191					
3	76	182					
4	60	168					
5	57	170					
6	65	172					
7	76	180					
8	55	167					
9	80	183					
10	79	179					
11	75	181					
12	77	181					
13	62	170					
14	64	172					
15	78	185					
16	56	165					
17	68	175					
18	78	183					

18R x 1C

data Box plot - Gender Height t-test - Example Dot plot - Height Male Regression +

# Regression

Home Insert Draw Page Layout Formulas **Data** Review View

From HTML From Text New Database Query Refresh All Properties Edit Links Sort Filter

A1 A B C D E F G P Q R S T U V

Weight Height

	A	B
1	80	191
2	76	182
3	60	168
4	57	170
5	65	172
6	76	180
7	55	167
8	80	183
9	79	179
10	75	181
11	77	181
12	62	170
13	64	172
14	78	185
15	56	165
16	68	175
17	78	183
18		

18R x 1C

2 – select the associated column  
(here height)

Regression

Input

Input Y Range: \$B\$1:\$B\$18

Input X Range: \$A\$1:\$A\$18

Labels  Constant is Zero

Confidence Level: 95 %

Output options

Output Range: [ ]

New Worksheet Ply: [ ]

New Workbook

Residuals

Residuals  Residual Plots

Standardized Residuals  Line Fit Plots

Normal Probability

Normal Probability Plots

OK Cancel

1 – Select the X-axis box

# Regression

Screenshot of Microsoft Excel showing the Data Analysis Regression dialog box. The dialog box is titled "Regression" and contains the following settings:

- Input**:
  - Input Y Range: \$B\$1:\$B\$18
  - Input X Range: \$A\$1:\$A\$18
  - Labels
  - Constant is Zero
- Output options**:
  - Output Range: (empty)
  - New Worksheet Ply: (empty)
  - New Workbook
  - Residuals**:
    - Residuals
    - Standardized Residuals
    - Residual Plots
    - Line Fit Plots
  - Normal Probability**:
    - Normal Probability Plots

A red box highlights the "Labels" checkbox in the Input section of the dialog. A red arrow points from this highlighted box to a red text annotation that reads: "Tick the box if you selected the label cells in the columns".

The Excel worksheet shows data in columns A and B, with column C containing labels. The data ranges from row 1 to 18. The "Regression" tab is selected in the ribbon.

	A	B
1	Weight	Height
2	80	191
3	76	182
4	60	168
5	57	170
6	65	172
7	76	180
8	55	167
9	80	183
10	79	179
11	75	181
12	77	181
13	62	170
14	64	172
15	78	185
16	56	165
17	68	175
18	78	183

# Regression

Home Insert Draw Page Layout Formulas **Data** Review View

From HTML From Text New Database Query Refresh All Properties Edit Links Sort Filter

**Regression**

**Input**

Input Y Range: \$B\$1:\$B\$18  
Input X Range: \$A\$1:\$A\$18  
 Labels  Constant is Zero  
 Confidence Level: 95 %

**Output Options**

Output Range: \$E\$1  
 New Worksheet Ply:  
 New Workbook  
Residuals  
 Residuals  Residual Plots  
 Standardized Residual  Line Fit Plots  
Normal Probability  
 Normal Probability Plots

OK Cancel

E1 A B C D E F G

	A	B	C	D	E	F	G
1	Weight	Height					
2	80	191					
3	76	182					
4	60	168					
5	57	170					
6	65	172					
7	76	180					
8	55	167					
9	80	183					
10	79	179					
11	75	181					
12	77	181					
13	62	170					
14	64	172					
15	78	185					
16	56	165					
17	68	175					
18	78	183					
19							
20							
21							
22							
23							
24							
25							
26							
27							
28							
29							
30							
31							
32							
33							
34							
35							

Share Comments

P Q R S T U V

Select the cell(s) where the result should appear

data Box plot - Gender Height t-test - Example Dot plot - Height Male Regression +

# Regression

Screenshot of Microsoft Excel showing the Regression dialog box. The data range A1:B18 is selected for both Input Y Range and Input X Range. The 'Labels' checkbox is checked. The 'Output options' section shows 'Output Range' set to \$E\$1, and the 'Residuals' checkbox is checked. A red box highlights the 'Residuals' and 'Residual Plots' checkboxes. A red arrow points from the text below to this highlighted area.

Input

Input Y Range: \$B\$1:\$B\$18

Input X Range: \$A\$1:\$A\$18

Labels       Constant is Zero

Confidence Level: 95 %

Output options

Output Range: \$E\$1

New Worksheet Ply:

New Workbook

Residuals       Residual Plots

Standardized Residuals       Line Fit Plots

Normal Probability

Normal Probability Plots

OK Cancel

Share Comments

Connections Properties Edit Links

From HTML From Text New Database Query Refresh All Sort Filter

A B C D E F G P Q R S T U V

Weight Height

	A	B	C	D	E	F	G	P	Q	R	S	T	U	V
1	Weight	Height												
2	80	191												
3	76	182												
4	60	168												
5	57	170												
6	65	172												
7	76	180												
8	55	167												
9	80	183												
10	79	179												
11	75	181												
12	77	181												
13	62	170												
14	64	172												
15	78	185												
16	56	165												
17	68	175												
18	78	183												
19														
20														
21														
22														
23														
24														
25														
26														
27														
28														
29														
30														
31														
32														
33														
34														
35														

data Box plot - Gender Height t-test - Example Dot plot - Height Male Regression +

Select the additional information you want to show (here residual values, residual plots, line fit plots)

+ Validate

# Regression

SUMMARY OUTPUT

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V							
1	Weight	Height			SUMMARY OUTPUT																								
2	80	191			Regression Statistics																								
3	76	182			Multiple R 0.94638177																								
4	60	168			R Square 0.89563845																								
5	57	170			Adjusted R S 0.88868102																								
6	65	172			Standard Err 2.47090169																								
7	76	180			Observations 17																								
8	55	167			ANOVA																								
9	80	183			df SS MS F Significance F																								
10	79	179			Regression 1 785.949084 785.949084 128.7311 9.2622E-09																								
11	75	181			Residual 15 91.5803273 6.10535515																								
12	77	181			Total 16 877.529412																								
13	62	170			Coefficients Standard Error t Stat P-value Lower 95% Upper 95% Lower 95.0% Upper 95.0%																								
14	64	172			Intercept 123.923571 4.69051552 26.4200323 5.4096E-14 113.925974 133.921168 113.925974 133.921168																								
15	78	185			Weight 0.75657613 0.06668235 11.3459729 9.2622E-09 0.61444606 0.89870619 0.61444606 0.89870619																								
16	56	165			RESIDUAL OUTPUT																								
17	68	175			Observation Predicted Height Residuals																								
18	78	183			1 184.449662 6.55033845																								
19					2 181.423357 0.57664296																								
20					3 169.318139 -1.318139																								
21					4 167.048411 2.95158941																								
22					5 173.10102 -1.1010196																								
23					6 181.423357 -1.423357																								
24					7 165.535258 1.46474167																								
25					8 184.449662 -1.4496616																								
26					9 183.693085 -4.6930854																								
27					10 180.666781 0.33321909																								
28					11 182.179933 -1.1799332																								

Weight Residual Plot

Weight Line Fit Plot

Height

Predicted Height

Weight

Residuals

Height

Predicted Height

Weight

REGRESSION

# Regression

SUMMARY OUTPUT

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	Weight	Height			SUMMARY OUTPUT																	
2	80	191																				
3	76	182																				
4	60	168																				
5	57	170																				
6	65	172																				
7	76	180																				
8	55	167																				
9	80	183																				
10	79	179																				
11	75	181																				
12	77	181																				
13	62	170																				
14	64	172																				
15	78	185																				
16	56	165																				
17	68	175																				
18	78	183																				
19																						
20																						
21																						
22																						
23																						
24																						
25																						
26																						
27																						
28																						
29																						
30																						
31																						
32																						
33																						
34																						
35																						

RESIDUAL OUTPUT

	Observation	Predicted Height	Residuals
1	1	184.449662	6.55033845
2	2	181.423357	0.57664296
3	3	169.318139	-1.318139
4	4	167.048411	2.95158941
5	5	173.10102	-1.1010196
6	6	181.423357	-1.423357
7	7	165.535258	1.46474167
8	8	184.449662	-1.4496616
9	9	183.693085	-4.6930854
10	10	180.666781	0.33321909
11	11	182.179933	-1.1799332

Regression coefficients

Regression linear equation: Height = a + b x Weight

Where

- a: given by the cell "Intercept"
- b: given by the cell "Weight"

Weight Residual Plot

Weight Line Fit Plot

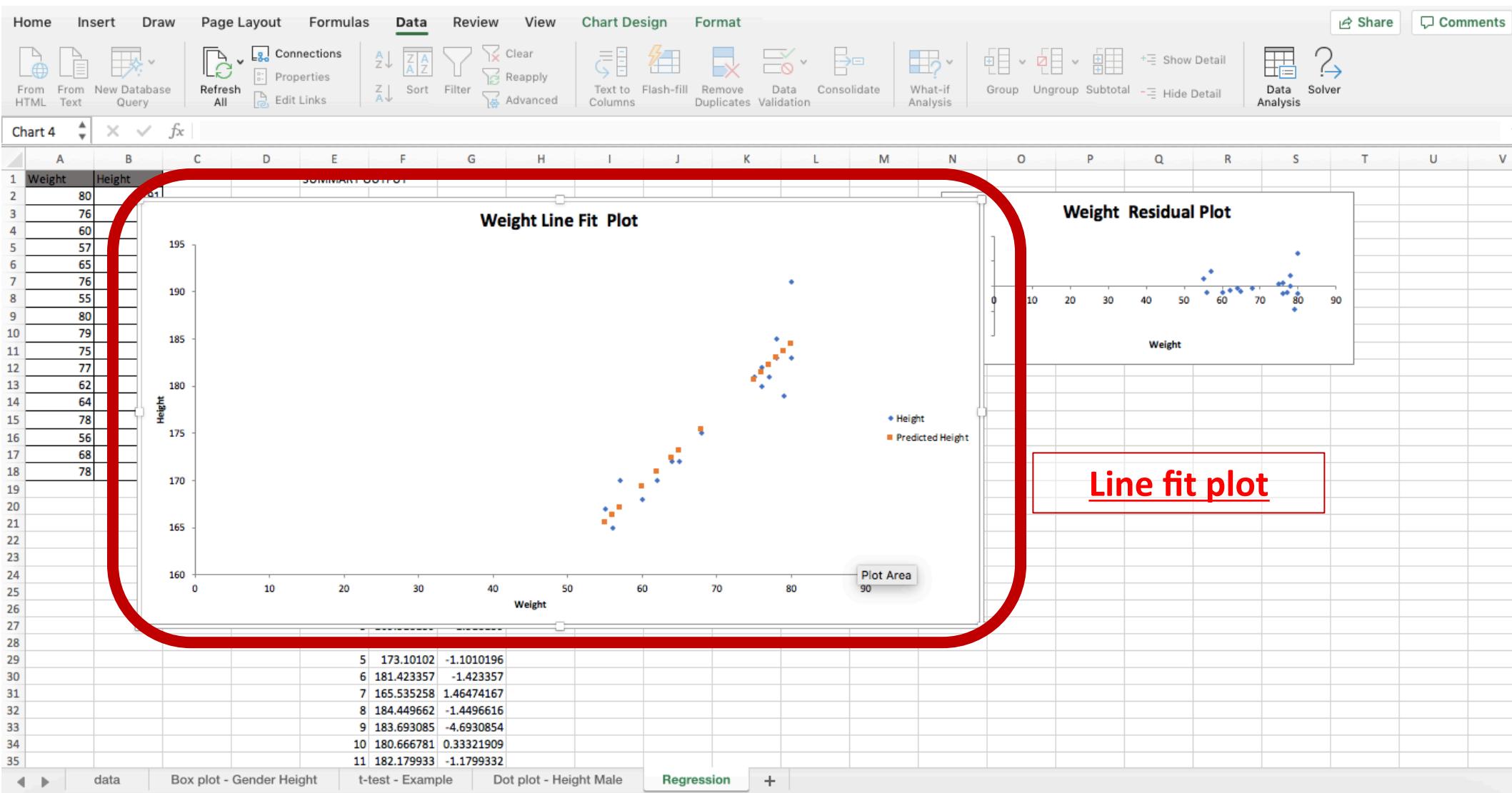
Residuals

Height

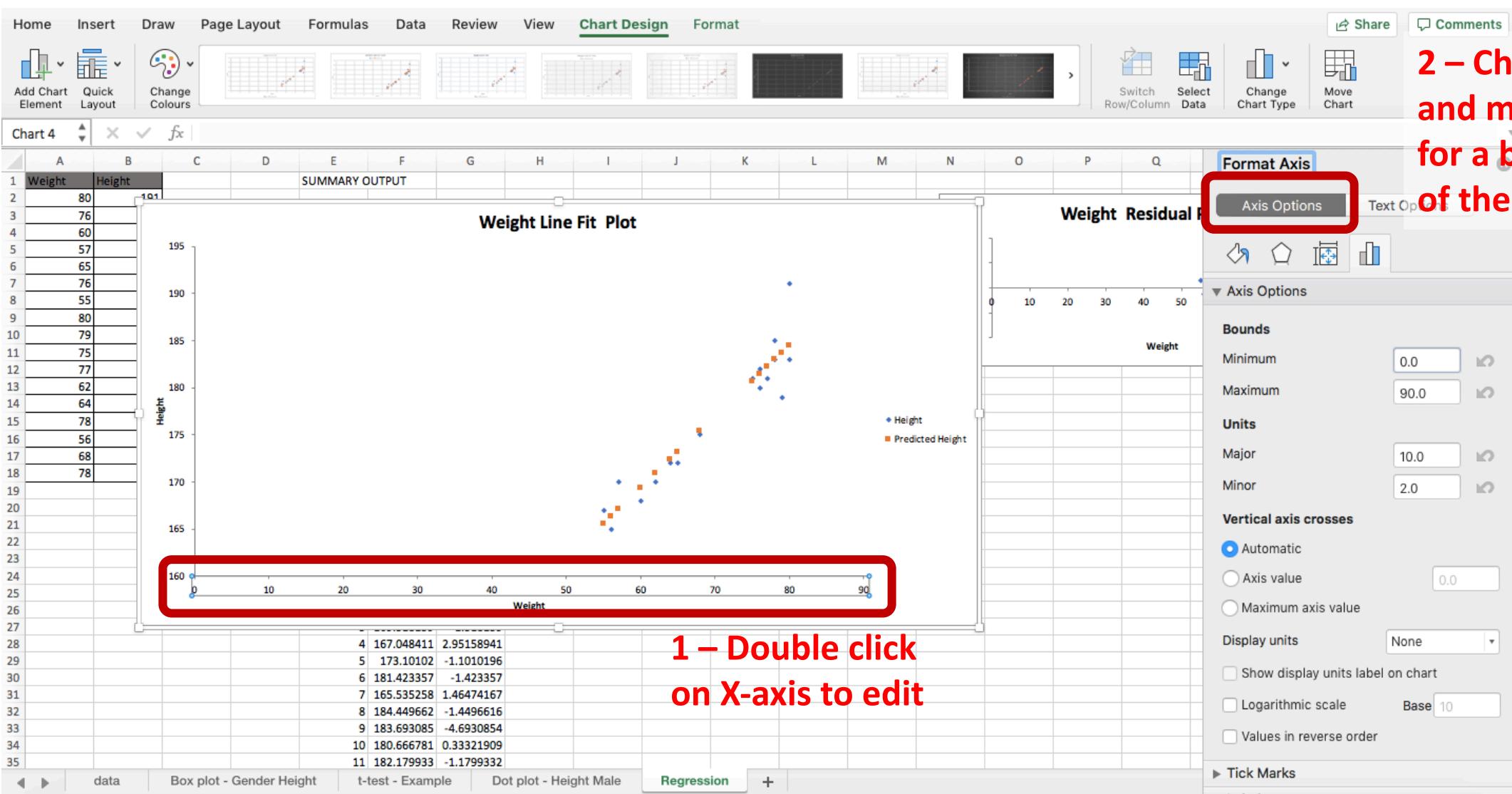
Weight

Legend: Height (blue diamond), Predicted Height (orange square)

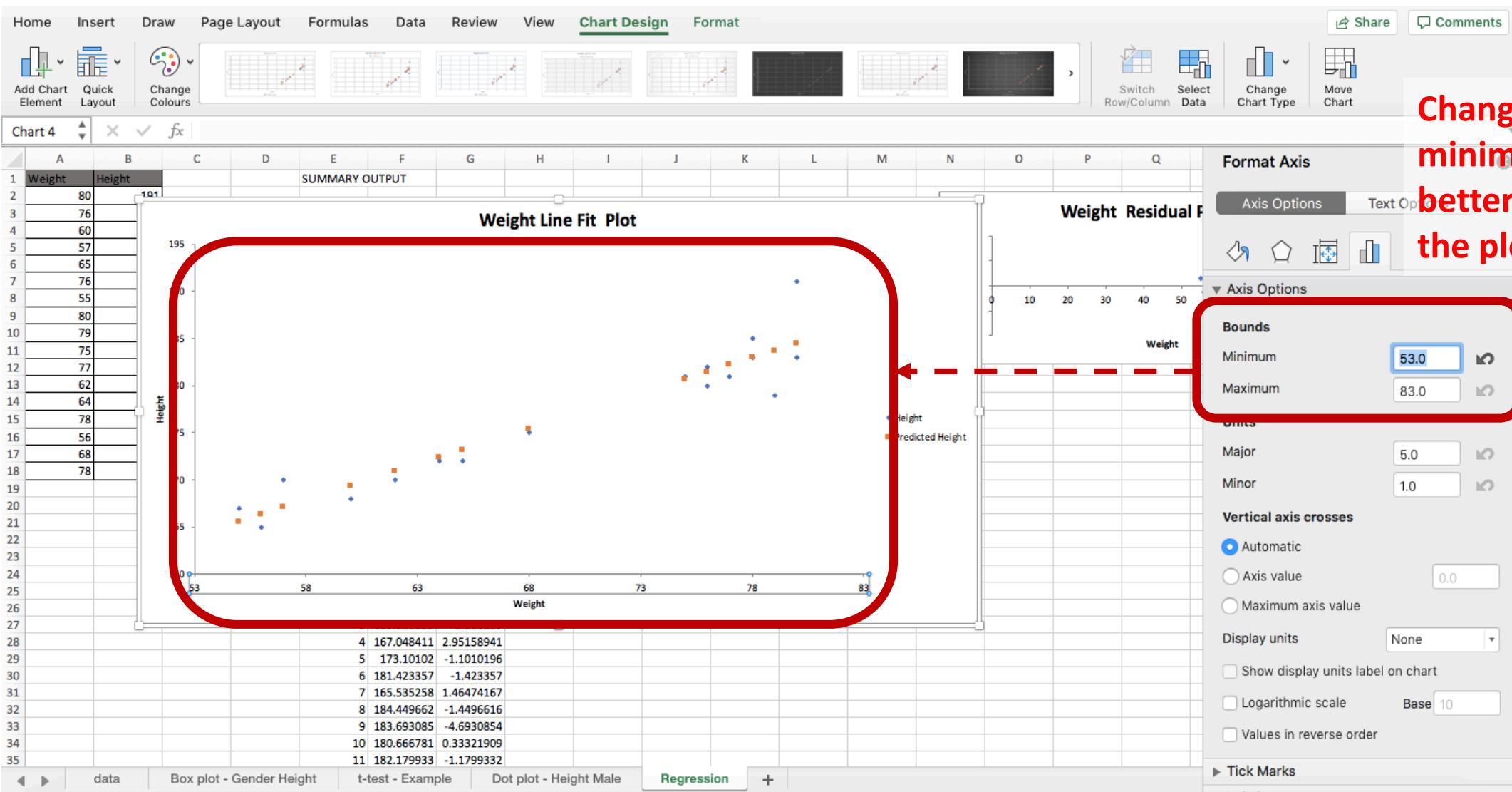
# Regression – Edit line fit plot



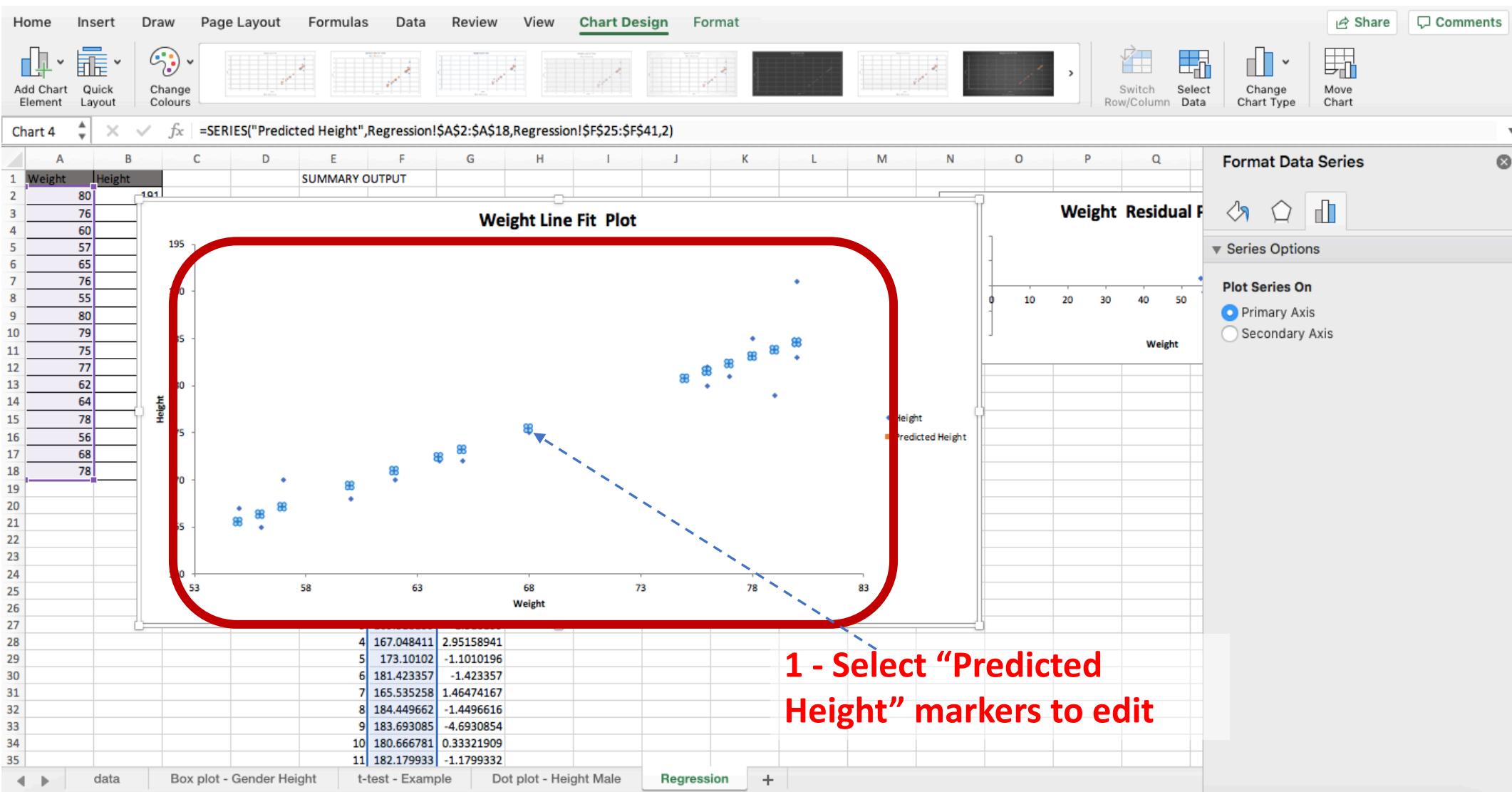
# Regression – Edit line fit plot



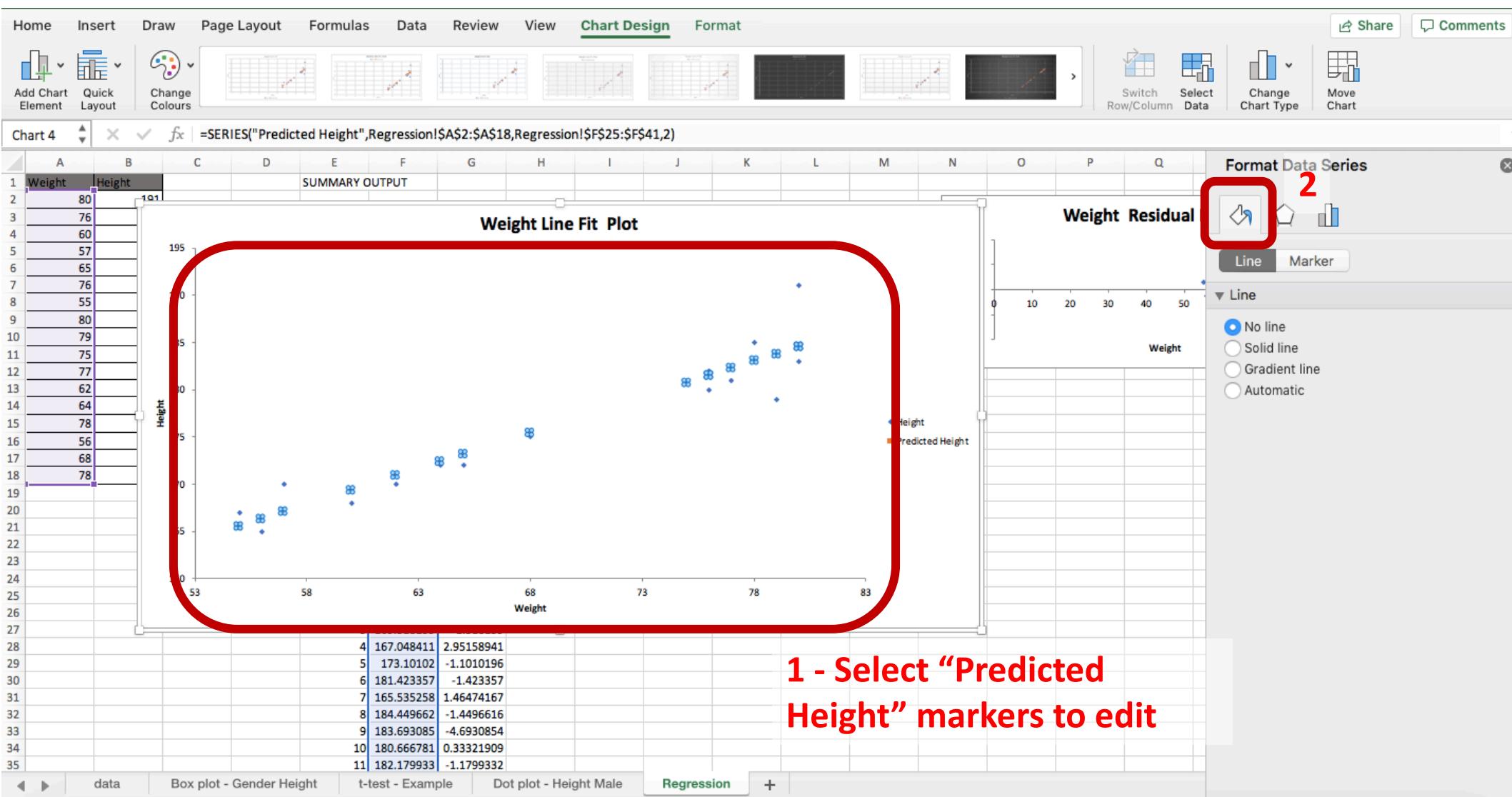
# Regression – Edit line fit plot



# Regression – Edit line fit plot



# Regression – Edit line fit plot



# Regression – Edit line fit plot

Home Insert Draw Page Layout Formulas Data Review View **Chart Design** Format Share Comments

Add Chart Element Quick Layout Change Colours

Chart 4 =SERIES("Predicted Height",Regression!\$A\$2:\$A\$18,Regression!\$F\$25:\$F\$41,2)

A B C D E F G H I J K L M N O P Q

1 Weight Height SUMMARY OUTPUT

2 80 195

3 76

4 60

5 57

6 65

7 76

8 55

9 80

10 79

11 75

12 77

13 62

14 64

15 78

16 56

17 68

18 78

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

Weight Line Fit Plot

Height

Predicted Height

Weight Residual

Weight

1 - Select "Predicted Height" markers to edit

3 – Select “Solid line” to plot the fitted line

Format Data Series

Line Marker

No line Solid line Gradient line Automatic

Colour

Transparency 0%

Width 1.5 pt

Compound type

Dash type

Cap type Round

Join type Round

Beg Join type

Begin Arrow size

End Arrow type

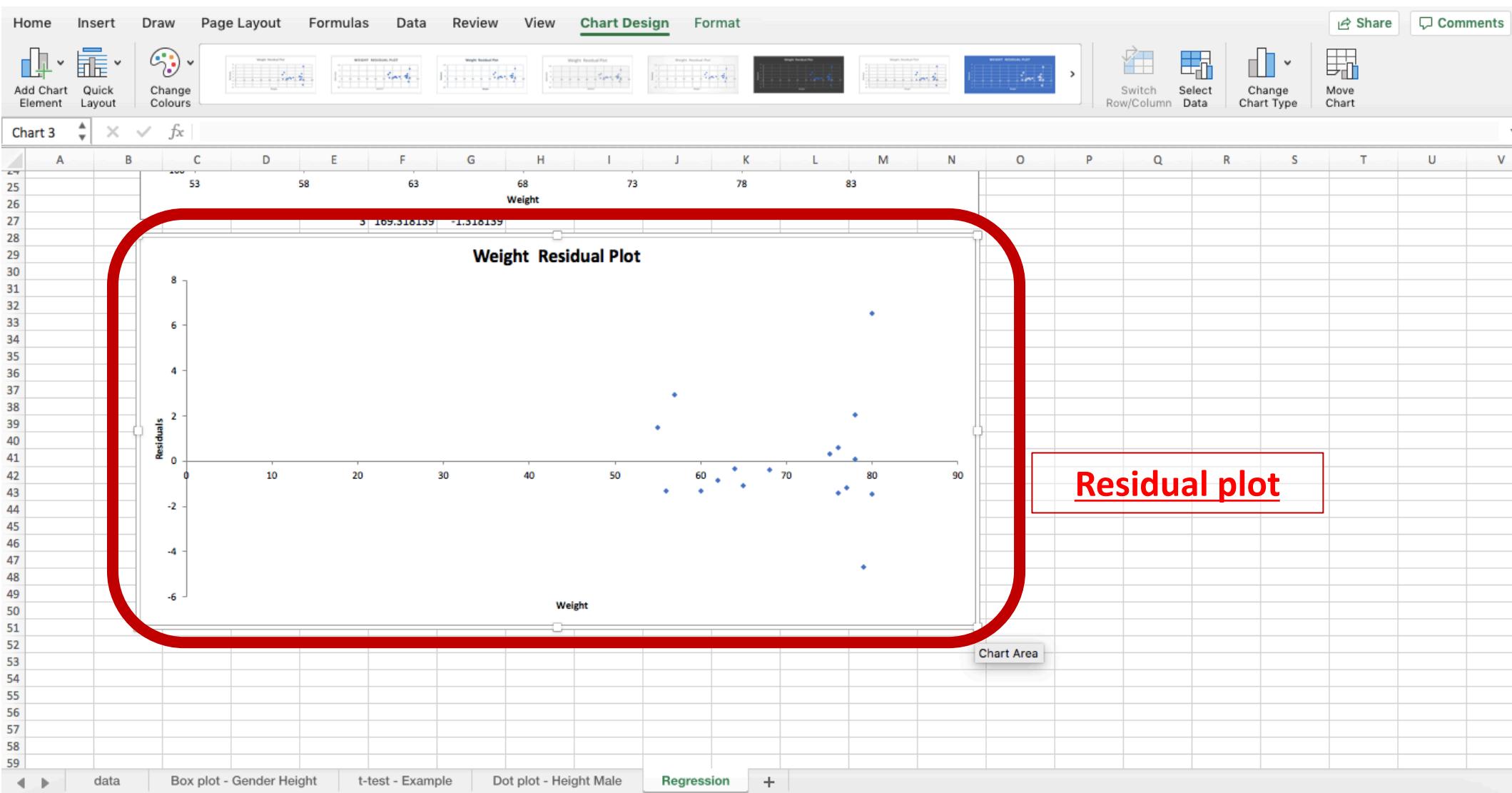
End Arrow size

Smoothed line

data Box plot - Gender Height t-test - Example Dot plot - Height Male Regression

	4	167.048411	2.95158941
5	173.10102	-1.1010196	
6	181.423357	-1.423357	
7	165.535258	1.46474167	
8	184.449662	-1.44966616	
9	183.693085	-4.6930854	
10	180.666781	0.33321909	
11	182.179933	-1.1799332	

# Regression – Edit residual plot



# Regression – Edit residual plot

Home Insert Draw Page Layout Formulas Data Review View **Chart Design** Format Share Comments

Add Chart Element Quick Layout Change Colours

Weight Residual Plot Weight Residual Plot

Switch Row/Column Select Data Change Chart Type Move Chart

Chart 3 A B C D E F G H I J K L M N O P Q

53 58 63 68 73 78 83

Weight

109.318139 -1.318139

Weight Residual Plot

Residuals

Weight

8 6 4 2 0 -2 -4 -6

1 – Double click on the X-Axis

Format Axis

Axis Options Text Options

Bounds Minimum 0.0 Maximum 90.0

Units Major 10.0 Minor 2.0

Vertical axis crosses

Automatic Axis value Maximum axis value

Display units None

Show display units label on chart Logarithmic scale Base 10

Values in reverse order

2 – Change the bounds if necessary

data Box plot - Gender Height t-test - Example Dot plot - Height Male Regression +

# Regression – Edit residual plot

Screenshot of a Microsoft Excel spreadsheet showing a Weight Residual Plot. The chart is titled "Weight Residual Plot" and displays residuals on the Y-axis against weight on the X-axis. The X-axis has major ticks at 53, 58, 63, 68, 73, 78, and 83. The Y-axis has major ticks from -6 to 8. A red box highlights the chart area. A red dashed arrow points from the "Format Axis" dialog box to the chart, indicating where to change axis settings.

The "Format Axis" dialog box is open, showing the "Axis Options" tab. It includes:

- A "Scale and Labels" icon (highlighted with a red box).
- "Axis Options" section:
  - Bounds:
    - Minimum: 53.0
    - Maximum: 83.0
- "Vertical axis crosses":
  - Automatic (selected)
  - Axis value: 0.0
  - Maximum axis value
- "Display units": None
- Checkboxes:
  - Show display units label on chart
  - Logarithmic scale: Base 10
  - Values in reverse order
- "Tick Marks"

A red callout box with the text "Change the bounds if necessary" is positioned near the "Bounds" section of the dialog box.

Below the chart, the formula bar shows:  $\text{Residuals} = 109.318139 - 1.318139 \times \text{Weight}$ .

The Excel ribbon tabs visible are Home, Insert, Draw, Page Layout, Formulas, Data, Review, View, Chart Design (selected), and Format.

Cells in the spreadsheet include:

- C25: 53
- C26: 58
- C27: 63
- C28: 68
- C29: 73
- C30: 78
- C31: 83
- C32: 109.318139
- C33: -1.318139

# Regression – Plot histogram of residuals

Screenshot of Microsoft Excel showing a regression analysis and residual plots.

The Excel ribbon is visible at the top, showing the Home tab is selected. The formula bar shows the value  $6.55033844572014$  in cell G25.

Data in cells A15 to H18:

	A	B	C	D	E	F	G	H
15	78	185						
16	56	165						
17	68	175						
18	78	183						

Coefficients table (G16 to H18):

	Coefficients	Standard Error	t Stat
Intercept	123.9235712	4.690515515	26.420
Weight	0.756576129	0.06668235	11.345

RESIDUAL OUTPUT table (G25 to H41):

Observation	Predicted Height	Residuals
1	184.4496616	6.550338446
2	181.423357	0.576642961
3	169.318139	-1.318138977
4	167.0484106	2.95158941
5	173.1010196	-1.101019621
6	181.423357	-1.423357039
7	165.5352583	1.464741667
8	184.4496616	-1.449661554
9	183.6930854	-4.693085425
10	180.6667809	0.33321909
11	182.1799332	-1.179933168
12	170.8312912	-0.831291235
13	172.3444435	-0.344443492
14	182.9365093	2.063490703
15	166.2918345	-1.291834461
16	175.370748	-0.370748008
17	182.9365093	0.063490703

Weight Residual Plot (bottom right):

Use residual values to plot a histogram

Regression tab is selected in the bottom ribbon.

# Regression – Plot histogram of residuals

2

3

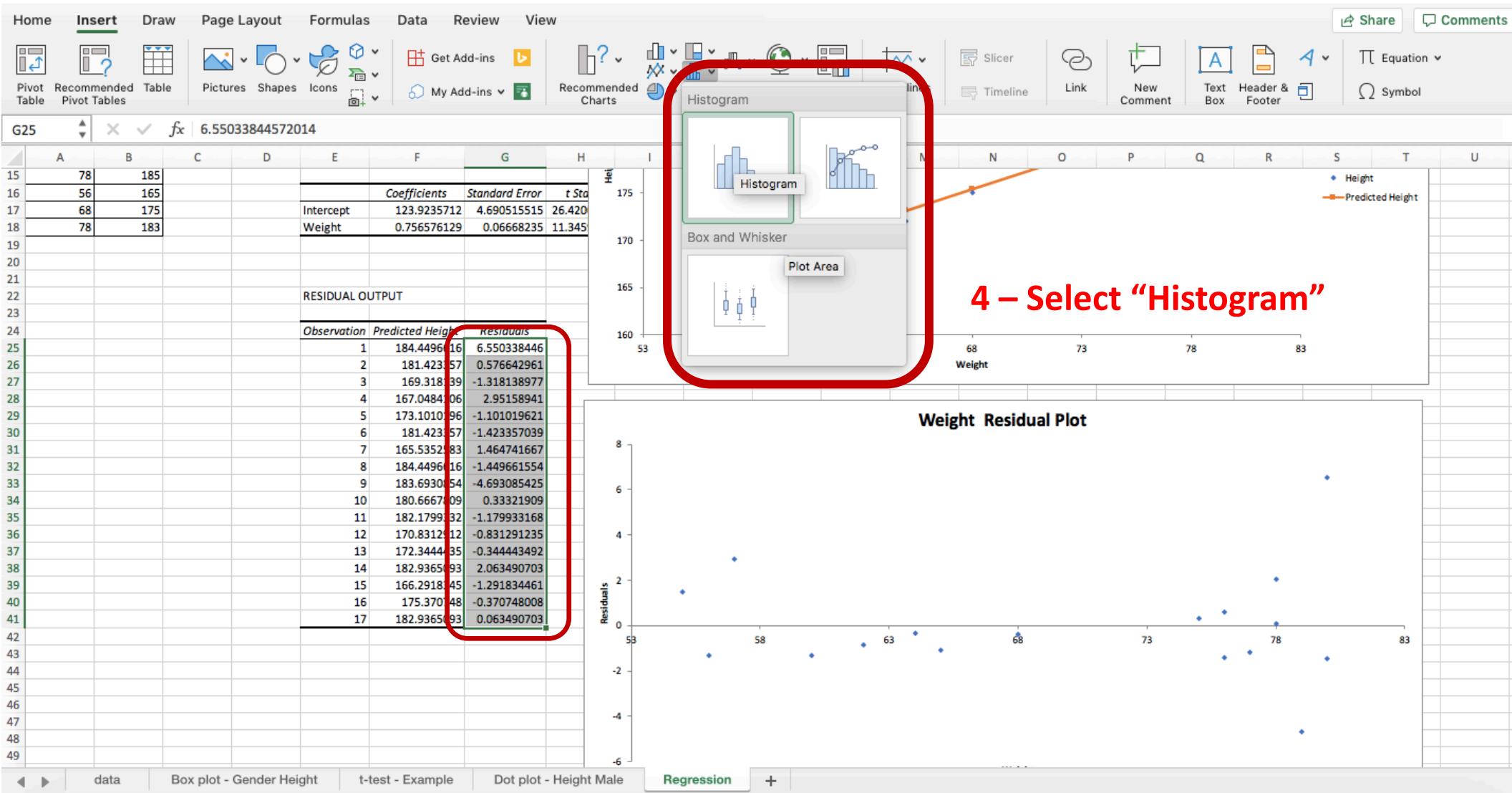
1 – Select the residual values

The screenshot shows a Microsoft Excel spreadsheet with the following details:

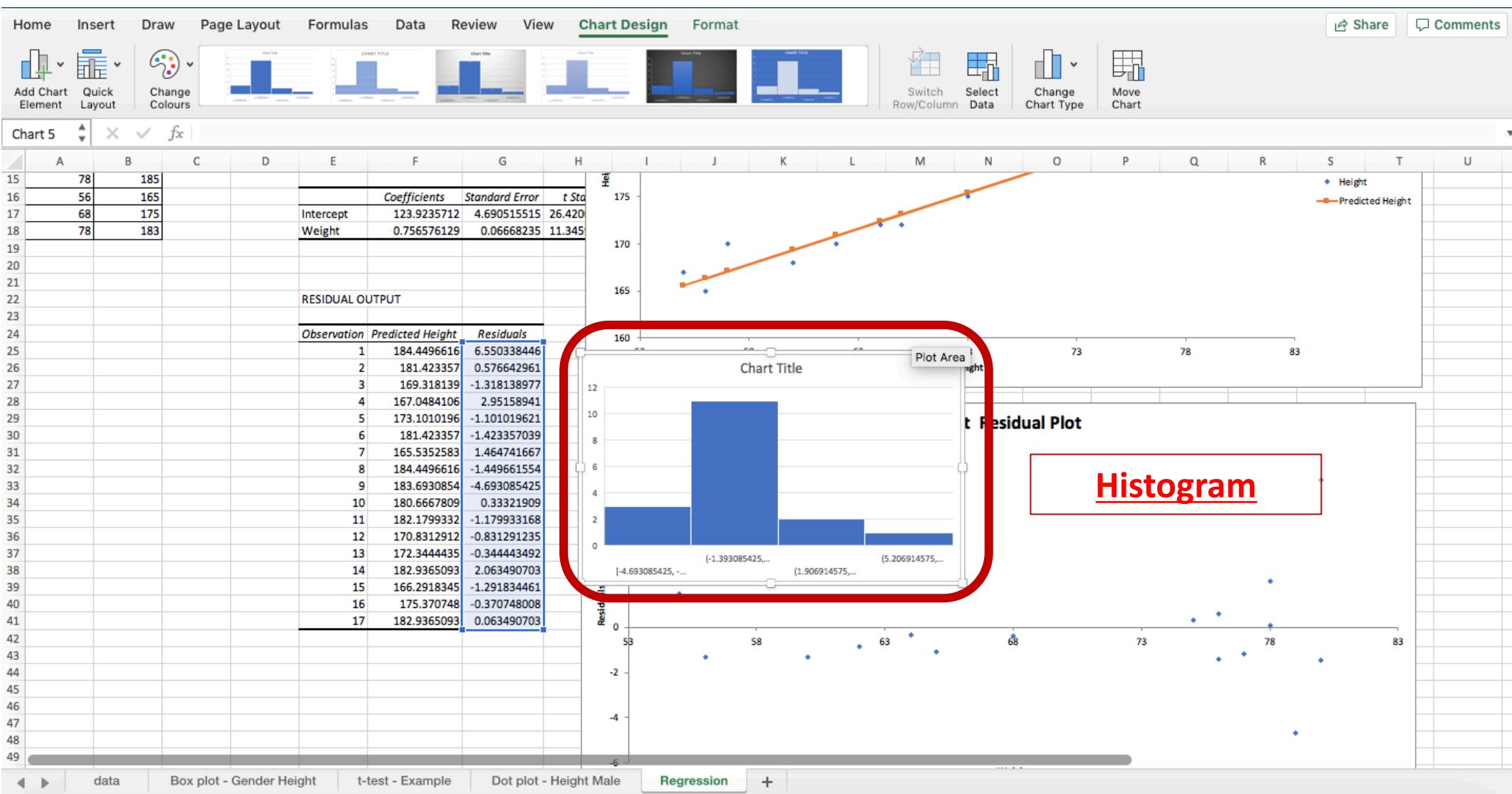
- Excel ribbon:** The "Insert" tab is selected (highlighted with a red box labeled 2). Other tabs like Home, Draw, Page Layout, etc., are visible.
- Data:** A table of coefficients is shown in rows 15-18, and a residual output table is shown from row 25 to 41.
- Chart 1:** A scatter plot titled "Weight Residual Plot" showing Residuals on the Y-axis (ranging from -6 to 8) versus weight on the X-axis (ranging from 53 to 83). The data points show a clear upward trend, indicating a non-linear relationship.
- Chart 2:** A scatter plot showing Height (blue dots) and Predicted Height (orange squares) versus weight. The predicted height follows a linear trend, while the actual height data points show a non-linear, increasing trend.
- Residual Output Table:** A table starting at row 25 with columns for Observation, Predicted Height, and Residuals. The data includes:

Observation	Predicted Height	Residuals
1	184.4496	6.550338446
2	181.423	0.576642961
3	169.318	-1.318138977
4	167.0484	2.95158941
5	173.1010	-1.101019621
6	181.423	-1.423357039
7	165.5352	1.464741667
8	184.4496	-1.449661554
9	183.6930	-4.693085425
10	180.6667	0.33321909
11	182.1799	-1.179933168
12	170.8312	-0.831291235
13	172.3444	-0.344443492
14	182.9365	2.063490703
15	166.2918	-1.291834461
16	175.370	-0.370748008
17	182.9365	0.063490703

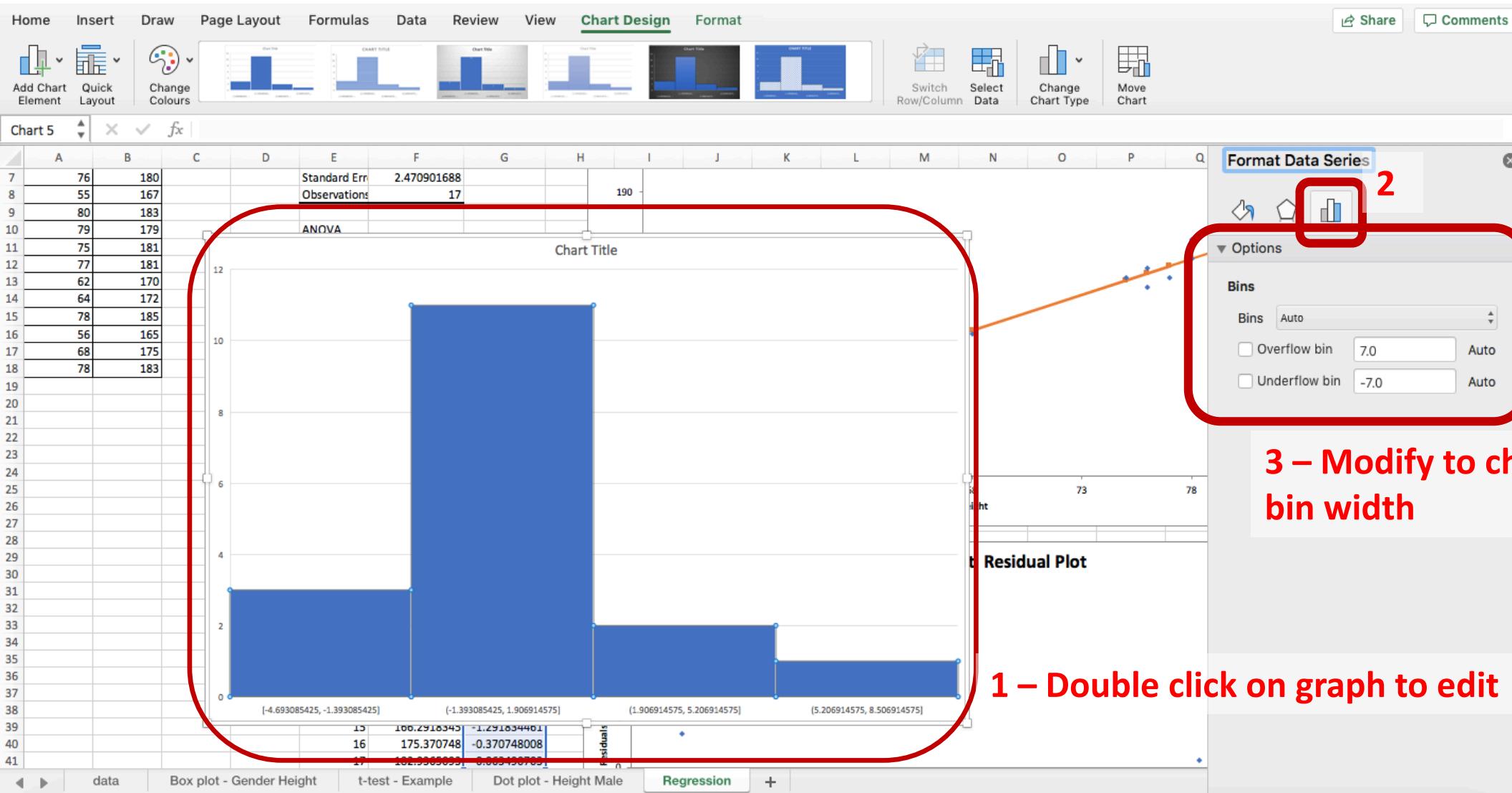
# Regression – Plot histogram of residuals



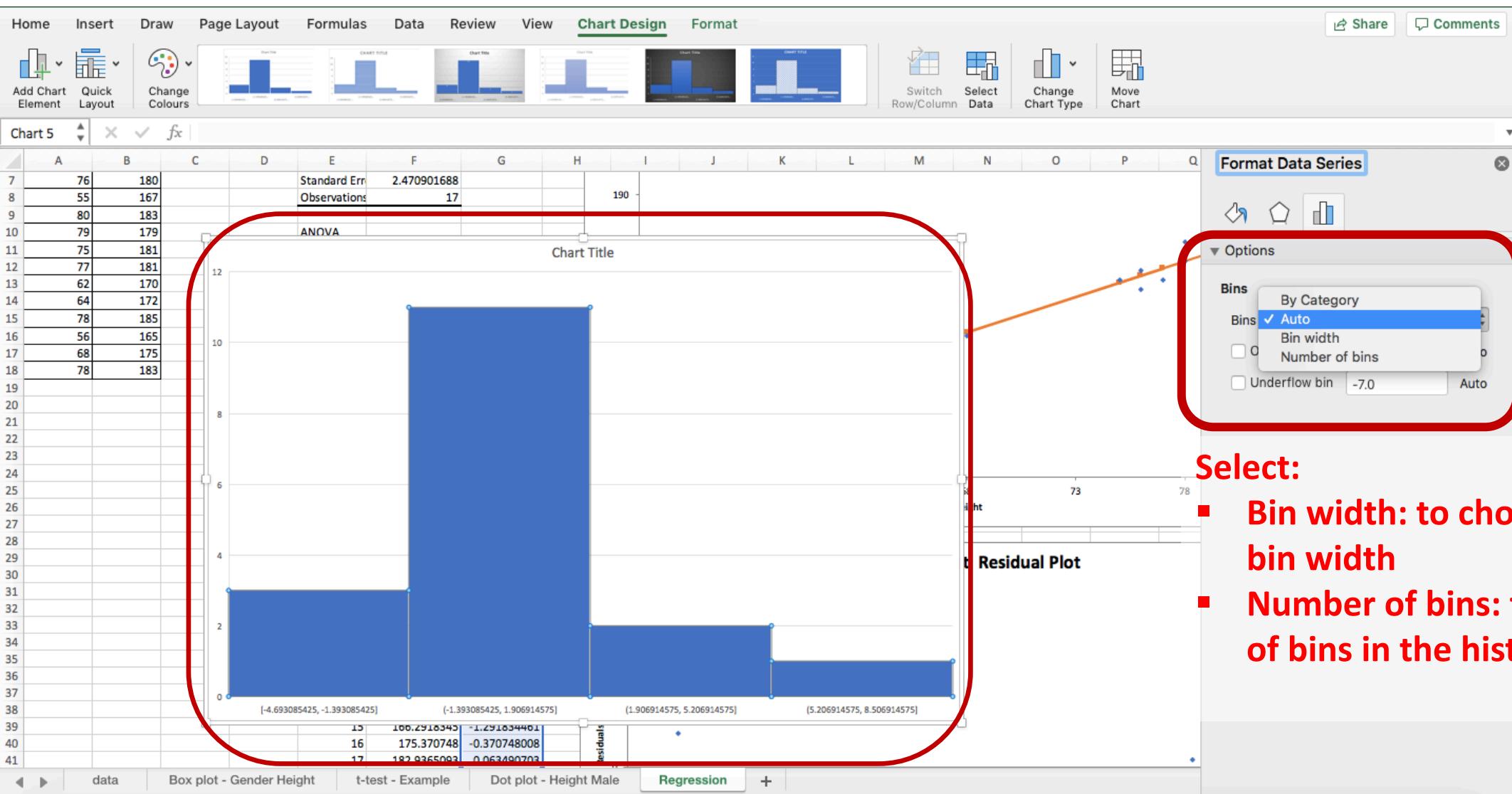
# Regression – Plot histogram of residuals



# Regression – Plot histogram of residuals



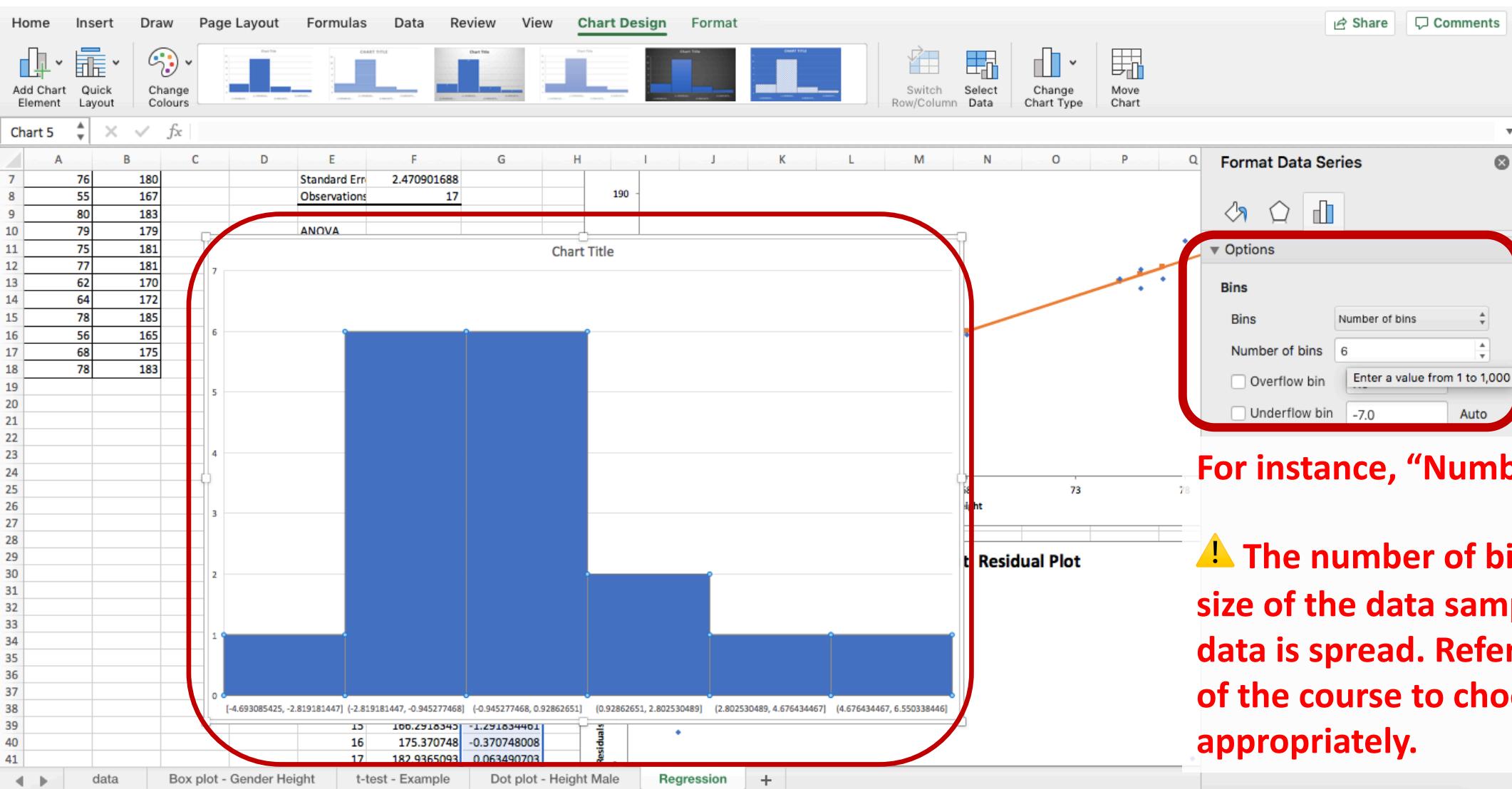
# Regression – Plot histogram of residuals



Select:

- **Bin width:** to choose directly the bin width
- **Number of bins:** to fix the number of bins in the histogram

# Regression – Plot histogram of residuals



# Regression – Plot histogram of residuals

Home Insert Draw Page Layout Formulas Data Review View **Chart Design** Format Share Comments

Add Chart Element Quick Layout Change Colours

Chart 5

A B C D E F G H I J K L M N O P Q

7	76	180			Standard Err	2.470901688									
8	55	167			Observations	17									
9	80	183													
10	79	179													
11	75	181													
12	77	181													
13	62	170													
14	64	172													
15	78	185													
16	56	165													
17	68	175													
18	78	183													
19															
20															
21															
22															
23															
24															
25															
26															
27															
28															
29															
30															
31															
32															
33															
34															
35															
36															
37															
38															
39															
40															
41															

Box plot - Gender Height t-test - Example Dot plot - Height Male Regression

1 – Select X-axis to edit

2 – Select to change the way the bin values are shown

Format Axis

Axis Options Text Options

Options

Gap Width 0.0

Font

Font Name Calibri

Font Size 9.0

Font Bold

Font Italic

Font Underline No Underline

Number Format

Category  Linked To Source

General

Number

Currency

Accounting

Date

Time

Percentage

Fraction

Scientific

Text

Plot Area

# Regression – Plot histogram of residuals

Screenshot of Microsoft Excel showing a regression analysis and residual plots.

The Excel ribbon is visible at the top, with the "Chart Design" tab selected. The chart area contains a histogram of residuals and a residual plot.

**Chart Data:**

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
7		76	180			Standard Err	2.470901688										
8		55	167			Observations	17										
9		80	183														
10		79	179														
11		75	181														
12		77	181														
13		62	170														
14		64	172														
15		78	185														
16		56	165														
17		68	175														
18		78	183														

**Format Axis Dialog:**

Number Format settings are highlighted with a red box:

- Category: Number
- Decimal places: 2
- Use 1000 Separator (,)
- Negative numbers: -1,234.00
- Format Code: #,##0.00

**Text in Red:**

For instance, “Number” with 2 decimal places.

# Regression – Plot histogram of residuals

Home Insert Draw Page Layout Formulas Data Review View **Chart Design** Format Share Comments

Add Chart Element Quick Layout Change Colours

Chart Title Chart Subtitle Chart Area Chart Legend

Switch Row/Column Select Data Change Chart Type Move Chart

Chart 5 Standard Err 2.470901688 Observations 17

ANOV

76 180  
55 167  
80 183  
79 179  
75 181  
77 181  
62 170  
64 172  
78 185  
56 165  
68 175  
78 183

7 6 5 4 3 2 1 0

[-4.69, -2.82] [-2.82, -0.95] [-0.95, 0.93] [0.93, 2.80] [2.80, 4.68] [4.68, 6.55]

Chart Title

Double click to edit title

78 73 78

sight

t Residual Plot

15 100.2918343 -1.291834401  
16 175.370748 -0.370748008  
17 182.9365093 0.063490703

Residuals

data Box plot - Gender Height t-test - Example Dot plot - Height Male Regression

Format Chart Title

Title Options Text Options

Alignment

Vertical alignment Middle

Text direction Horizontal

Resize shape to fit text

Allow text to overflow shape

Left margin

Right margin

Top margin

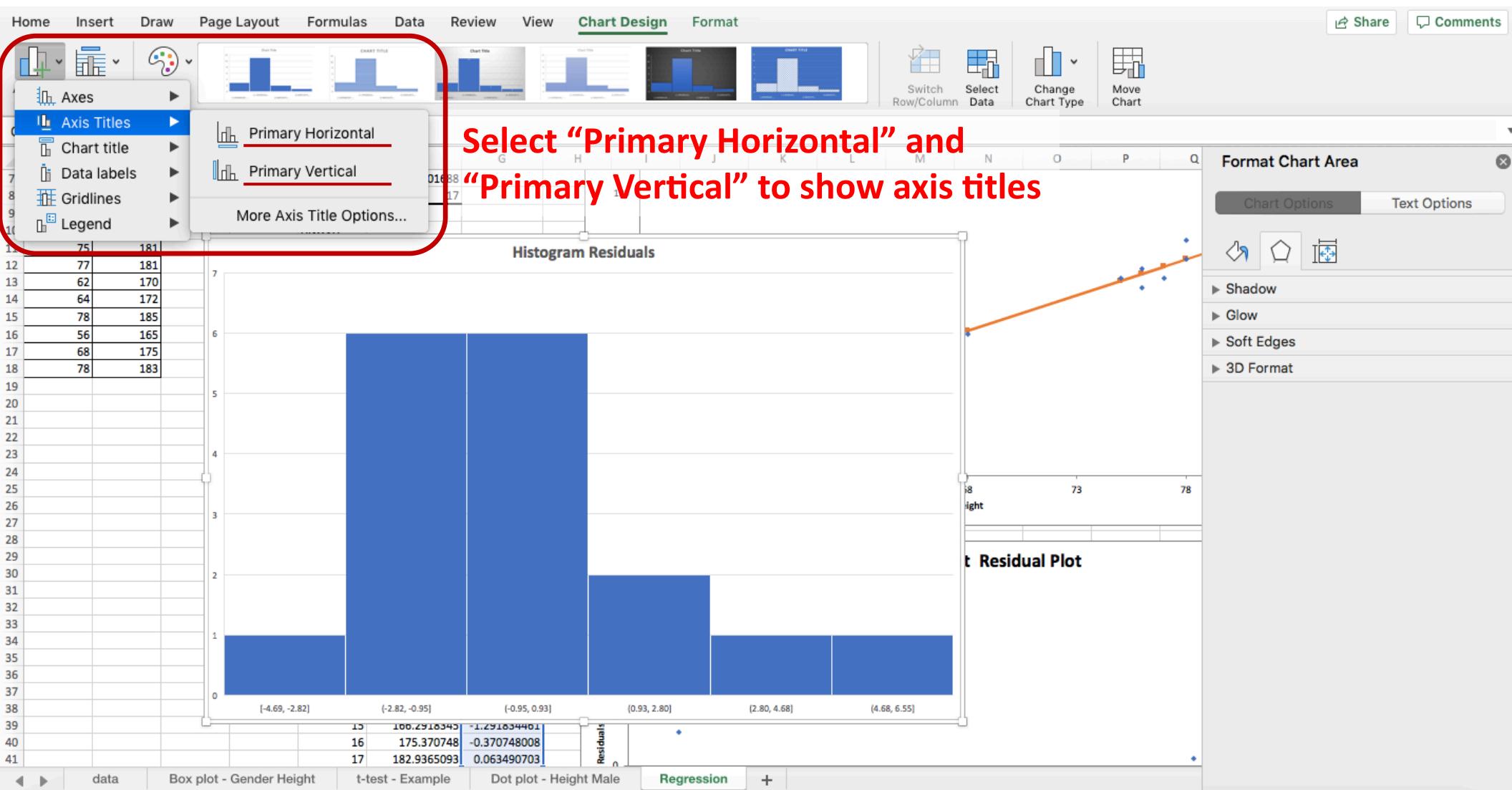
Bottom margin

Wrap text in shape

Columns...

The screenshot shows a Microsoft Excel spreadsheet with several tabs at the bottom: 'data', 'Box plot - Gender Height', 't-test - Example', 'Dot plot - Height Male', 'Regression' (which is selected), and '+'. The main area contains a histogram of residuals and a residual plot. The histogram has blue bars and a vertical line at zero. The residual plot shows points forming a positive linear trend. A red box highlights the 'Chart Title' area in the chart design ribbon, and red text says 'Double click to edit title'. The 'Format Chart Title' dialog is open on the right, showing options for title and text alignment.

# Regression – Plot histogram of residuals



# Regression – Plot histogram of residuals

