

Model 2Y1X using Linear Trendline in Excel 2013

by Milo Schield

Member: International Statistical Institute  
US Rep: International Statistical Literacy Project

Slides at: [www.StatLit.org/pdf/Excel2013-Model-Trendline-Linear-2Y1X-Slides.pdf](http://www.StatLit.org/pdf/Excel2013-Model-Trendline-Linear-2Y1X-Slides.pdf)  
Data at: [www.StatLit.org/XLS/Pulse.xlsx](http://www.StatLit.org/XLS/Pulse.xlsx)

Background

BACKGROUND: The Pulse dataset (© Minitab) contains data on college students. After recording rest pulse, height, weight, etc., student flipped coins. One group (Run=1) jogged in place for a minute; the other group watched. Afterward, everyone recorded their pulse a second time (Pulse2). Thus, the meaning of Pulse2 depends on the value of RUN.

To separate the two groups, sort on RUN and select just those rows that have the desired value of RUN.

N1 Descending; N2 Same as N1; N3 Ascending; N4 Descending

Goal: Summarize association for two series of variables

A. Generate an XY chart showing association between two quantitative variables having a common X. Generate four graphs on slides 13, 23, 28 and 33.

B. Show trend-line, linear equation,  $R^2$  for each series.

C. No writing in this exercise.

Use Pulse data set at [www.StatLit.org/xls/Pulse.xlsx](http://www.StatLit.org/xls/Pulse.xlsx)

For details on using Excel Chart Trendline to build a model, see [www.StatLit.org/pdf/Excel2013Model-Trendline-Linear-Slides.pdf](http://www.StatLit.org/pdf/Excel2013Model-Trendline-Linear-Slides.pdf)

1) Pulse1 and Pulse2 vs. Ht. Select all data prior to Sort

	A	B	C	D	E	F	G	H
1	Pulse1	Pulse2	Height	Weight	Activity	Run?	Smokes?	Male?
2	48	54	68	150	1	0	1	1
3	54	56	69	145	2	0	1	1
4	54	50	69	160	2	0	0	1
5	58	70	72	145	2	1	0	1
6	58	58	66	135	3	0	0	1
7	58	56	67	125	2	0	0	0
8	60	76	71	170	3	1	0	1

1) Sort on Run: Descending

1) Results of Sort on Run

	A	B	C	D	E	F	G	H
26	78	118	69	145	2	1	0	0
27	80	96	72	155	2	1	0	1
28	80	128	68	125	2	1	0	0
29	82	100	68	138	2	1	0	0
30	84	84	72	150	3	1	0	1
31	88	110	69	150	2	1	1	0
32	90	94	74	160	1	1	1	1
33	92	84	70	153	3	1	1	1
34	96	140	61	140	2	1	0	0
35	96	116	68	116	2	1	0	0
36	100	115	63	121	2	1	0	0
37	48	54	68	150	1	0	1	1

1) Select Pulse1 for Run = 1. (Rows 2-36). Insert XY Plot

Rows 2-36

1a) Select Pulse1 series Right-mouse. "Select Data"

1b) To insert X values, select Edit

1c) Either select X values range or copy Y, paste to X and edit

Pulse1 is Y axis (A2:A36); Height is X axis (C2:C36)

1d) Add 2nd series (Pulse2 by Ht), Select data; Select "Add"

1e) Add Pulse2 (B2:B36) as Y; Add Height (C2:C36) as X

Press OK

**1f) Result: Two Series same Y axis Format axis; add trendlines**

**2a) Copy data to a new Worksheet**

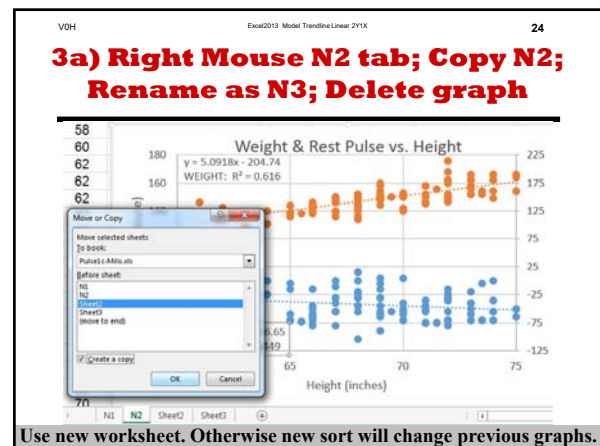
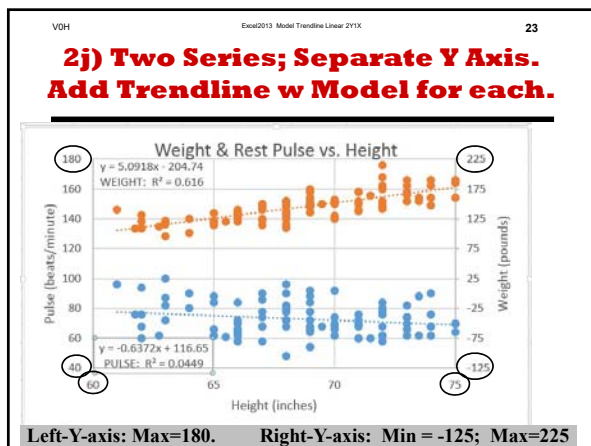
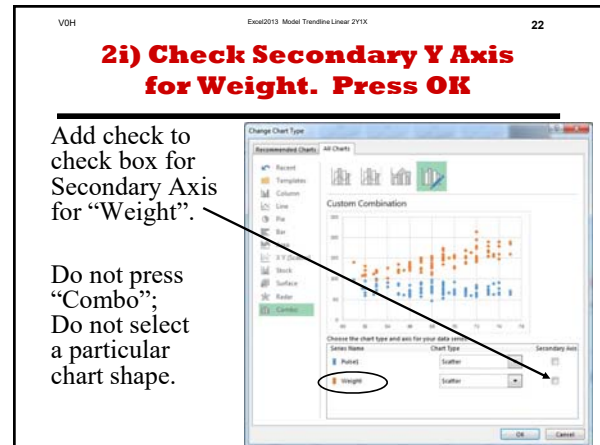
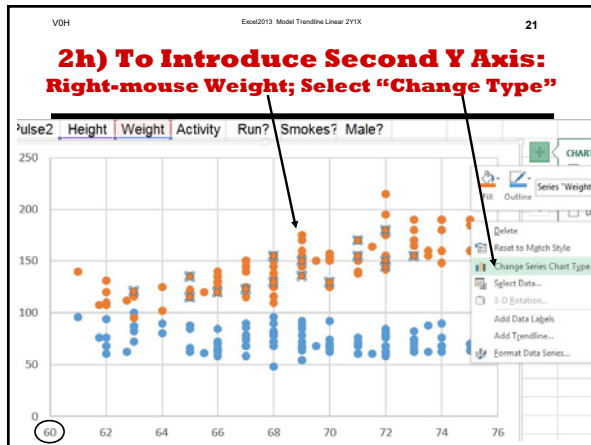
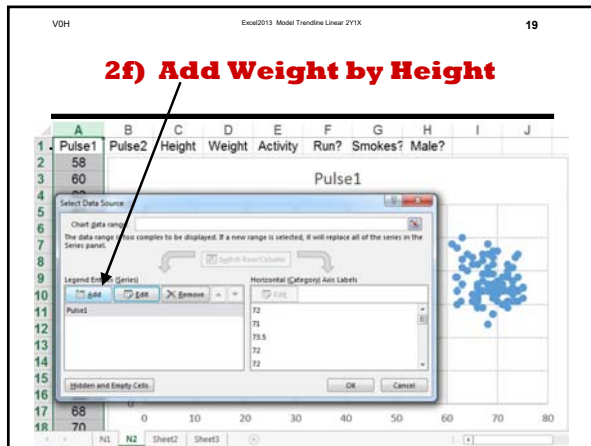
**2b) Select all Pulse1 (A2:A93) Insert XY chart**

	A	B	C	D	E
1	Pulse1	Pulse2	Height	Weight	Activity
2	58	70	72	145	2
3	60	76	71	170	3
4	62	76	73.5	180	3
5	62	75	72	195	2
6	62	58	72	175	3
7	62	100	66	120	2
8	62	98	62.75	112	2
9	64	88	66	140	2

**2c) Initial Chart: 1Y, no X**

**2d) Right-mouse on Data; Select "Select Data"**

**2e) Edit Pulse1 by Height series: X (C2:C93), Y (A2:A93)**



**3b) Select ALL columns and rows; From Home or Data menu, select Sort**

	A	B	C	D	E	F	G	H
1	Pulse1	Pulse2	Height	Weight	Activity	Run?	Smokes?	Male?
2	58	70	72	145	2	1	0	1
3	60	76	71	170	3	1	0	1
4	62	76	73.5	160	3	1	1	1
5	62	75	72	195	2	1	0	1
6	62	58	72	175	3	1	0	1
7	62	100	66	120	2	1	0	0
8	62	98	62.75	112	2	1	1	0
9	64	88	66	140	2	1	0	1
10	64	80	69	155	2	1	0	1
11	66	78	73	190	1	1	1	1
12	66	82	69	175	2	1	1	1
13	66	102	70	130	2	1	0	1
14	68	72	74	190	2	1	0	1
15	68	76	67	145	2	1	0	1
16	68	76	74	180	2	1	1	1
17	68	112	70	125	2	1	0	0
18	70	75	75	170	3	1	1	1

**3c) From Sort-by, select RUN; Order from Small-to-Large**

**3d) Select A1:B58. Insert XY Chart**

Select data with Run=0

	A	B	C	D	E	F	G	H
48	64	84	66	130	2	0	0	0
49	64	80	65	118	1	0	0	0
50	66	84	67	150	3	0	0	0
51	67	84	63	95	3	0	0	0
52	68	84	73.5	155	2	0	0	1
53	68	74	65	135	2	0	1	0
54	90	88	67	140	2	0	1	1
55	90	90	68	145	1	0	0	1
56	90	92	64	125	1	0	1	0
57	92	94	69	150	2	0	1	1
58	94	92	62	131	2	0	1	0
59	58	70	72	145	2	1	0	1
60	76	71	170	3	1	0	1	1
61	76	73.5	160	3	1	1	1	1
62	75	72	195	2	1	0	1	1
63	58	72	175	3	1	0	1	1
64	100	66	120	2	1	0	0	0
65	98	62.75	112	2	1	1	0	0
66	88	66	140	2	1	0	0	1
67	80	69	155	2	1	0	0	1
68	78	73	190	1	1	1	1	1
69	82	69	175	2	1	1	1	1
70	102	70	130	2	1	0	0	1
71	72	74	190	2	1	0	0	1
72	76	67	145	2	1	0	0	1
73	76	74	180	2	1	1	1	1
74	112	70	125	2	1	0	0	0
75	75	75	170	3	1	1	1	1

**3e) XY plot: Pulse2 vs Pulse1 Add "No Change" line: 45,45 - 95,95**

Repeat Measure of Rest Pulse  
Select: Run = 0

$y = 0.849x + 10.832$   
 $R^2 = 0.8521$

Dark line = 2nd=1st  
Dashed line = Trend Line

To create a diagonal line in Excel:  
Insert/ Shapes OR Chart Tools/Format/Insert Shapes: Select Line.

**4a) Right-Mouse N3 Tab; Copy N3 Rename as N4. Delete graph**

Use new worksheet. Otherwise new sort will change previous graphs.

**4b) Select all data; Sort by Run; Sort Values: Large to Small**

**4c) Select Pulse1 & 2 columns, Rows 2:36. Insert XY Chart**

**4d) Insert no-change line**

Excel: Insert/Illustrations/Shapes. Select Line

**4e) Set Intercept = 18.925  
Create trend-line and equation**

Note: This slide is slightly different from the previous slide.

**Summary of N3 and N4**

N3: Test-retest of rest pulse has slope of 0.849.  
A perfect test-retest would have a slope of 1.00.  
Explanation: Regression to the mean.

R-squared = 85%. The original resting pulse explains 85% of variability in the retest of pulse.

N4: Running in place increases pulse by 18.9 bpm over rest pulse when model has unit slope.

R-squared = 37%. Rest pulse explains 37% of the variation in pulse after running in place.

**Two Common Problems**

1. Including column heading in data selection. You may include top row when you select data. Normally Excel excludes this so the data starts in row 2. Sometimes Excel does not exclude row 1 and this really messes up your graph! Solution: Select Data and edit the X and Y data ranges manually.
2. Getting X (horizontal axis) and Y (vertical axis) reversed. There is no simple way to fix this. You need to right mouse the data, select "select data" and Edit the data in question. Manually make the changes necessary without making things worse...

**Two More Problems**

1. Sometimes your data just gets all mixed up. The most likely cause: you sorted just a part of the data instead of sorting the entire data set. Solution: Copy the data from a previous page. Paste it on the desired page and then repeat the sort.
2. Sometimes when you sort the data, a prior graph goes bad. The simplest explanation is that both graphs are 'pointing' at the same data. Solution: Make sure each graph has its own unique data on its own unique page. You may need to delete and recreate graphs too.

# **Model 2Y1X using Linear Trendline in Excel 2013**

---

by

**Milo Schield**

*Member: International Statistical Institute*

*US Rep: International Statistical Literacy Project*

**Slides at: *[www.StatLit.org/pdf](http://www.StatLit.org/pdf)***

***[/Excel2013-Model-Trendline-Linear-2Y1X-Slides.pdf](http://www.StatLit.org/pdf/Excel2013-Model-Trendline-Linear-2Y1X-Slides.pdf)***

**Data at: *[www.StatLit.org/XLS//Pulse.xlsx](http://www.StatLit.org/XLS//Pulse.xlsx)***

# Background

---

**BACKGROUND:** The Pulse dataset (© Minitab) contains data on college students. After recording rest pulse, height, weight, etc., student flipped coins. One group (Run=1) jogged in place for a minute; the other group watched. Afterward, everyone recorded their pulse a second time (Pulse2). Thus, the meaning of Pulse2 depends on the value of RUN.

To separate the two groups, sort on RUN and select just those rows that have the desired value of RUN.

N1 Descending; N2 Same as N1; N3 Ascending; N4 Descending



# **Goal: Summarize association for two series of variables**

---

- A. Generate an XY chart showing association between two quantitative variables having a common X.  
**Generate four graphs on slides 13, 23, 28 and 33.**
- B. Show trend-line, linear equation,  $R^2$  for each series.
- C. No writing in this exercise.

Use Pulse data set at [www.StatLit.org/xls/Pulse.xlsx](http://www.StatLit.org/xls/Pulse.xlsx)

For details on using Excel Chart Trendline to build a model, see [www.StatLit.org/pdf/](http://www.StatLit.org/pdf/)

[Excel2013Model-Trendline-Linear-Slides.pdf](http://www.StatLit.org/pdf/Excel2013Model-Trendline-Linear-Slides.pdf)

# 1) Pulse1 and Pulse2 vs. Ht. Select all data prior to Sort

The screenshot shows the Microsoft Excel 2013 ribbon with the **DATA** tab selected. The **Sort & Filter** group contains the **Sort** button, which is highlighted with a black arrow. The **Analysis** group contains **Data Analysis** and **Solver**. The formula bar shows the text **Pulse1**.

	A	B	C	D	E	F	G	H	I
1	Pulse1	Pulse2	Height	Weight	Activity	Run?	Smokes?	Male?	
2	48	54	68	150	1	0	1	1	
3	54	56	69	145	2	0	1	1	
4	54	50	69	160	2	0	0	1	
5	58	70	72	145	2	1	0	1	
6	58	58	66	135	3	0	0	1	
7	58	56	67	125	2	0	0	0	
8	60	76	71	170	3	1	0	1	

# 1) Sort on Run: Descending

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I
1	Pulse1	Pulse2	Height	Weight	Activity	Run?	Smokes?	Male?	
2	48	54	68	150	1	0	1	1	
3	54	56	60	145	2	0	1	1	

The Sort dialog box is open, showing the following settings:

- Sort by: Run?
- Sort On: Values
- Order: Largest to Smallest

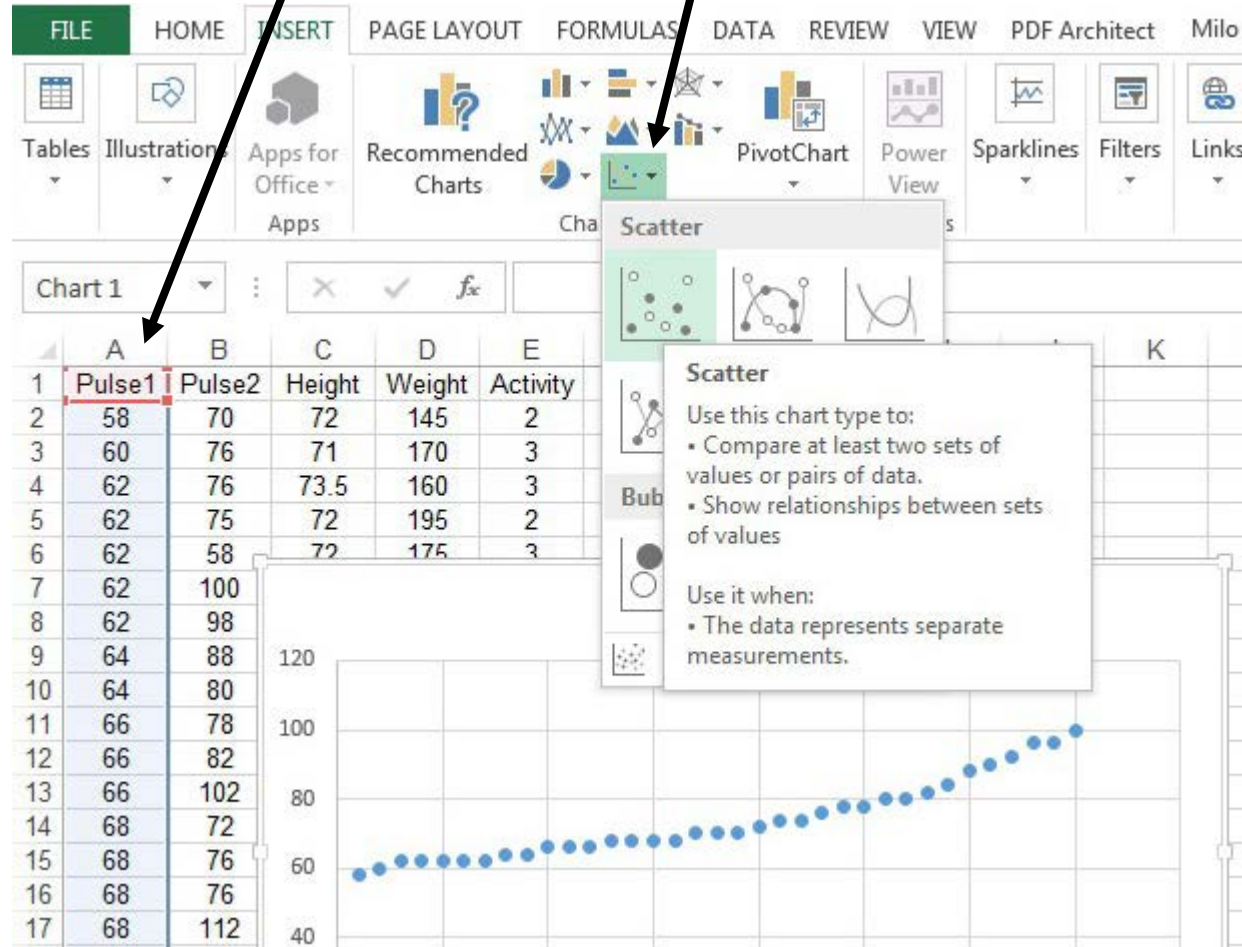
Arrows point to the 'Run?' dropdown and the 'Largest to Smallest' dropdown.

# 1) Results of Sort on Run

The screenshot shows the Microsoft Excel 2013 interface with the Data tab selected. The ribbon includes the following groups: Get External Data, Refresh, Connections, Sort & Filter (with Sort and Filter buttons), Data Tools (with Data Tools button), Outline, and Analysis Tools (with Data Analysis and Solver buttons). Below the ribbon, the formula bar shows 'H42' and '1'. The data table is as follows:

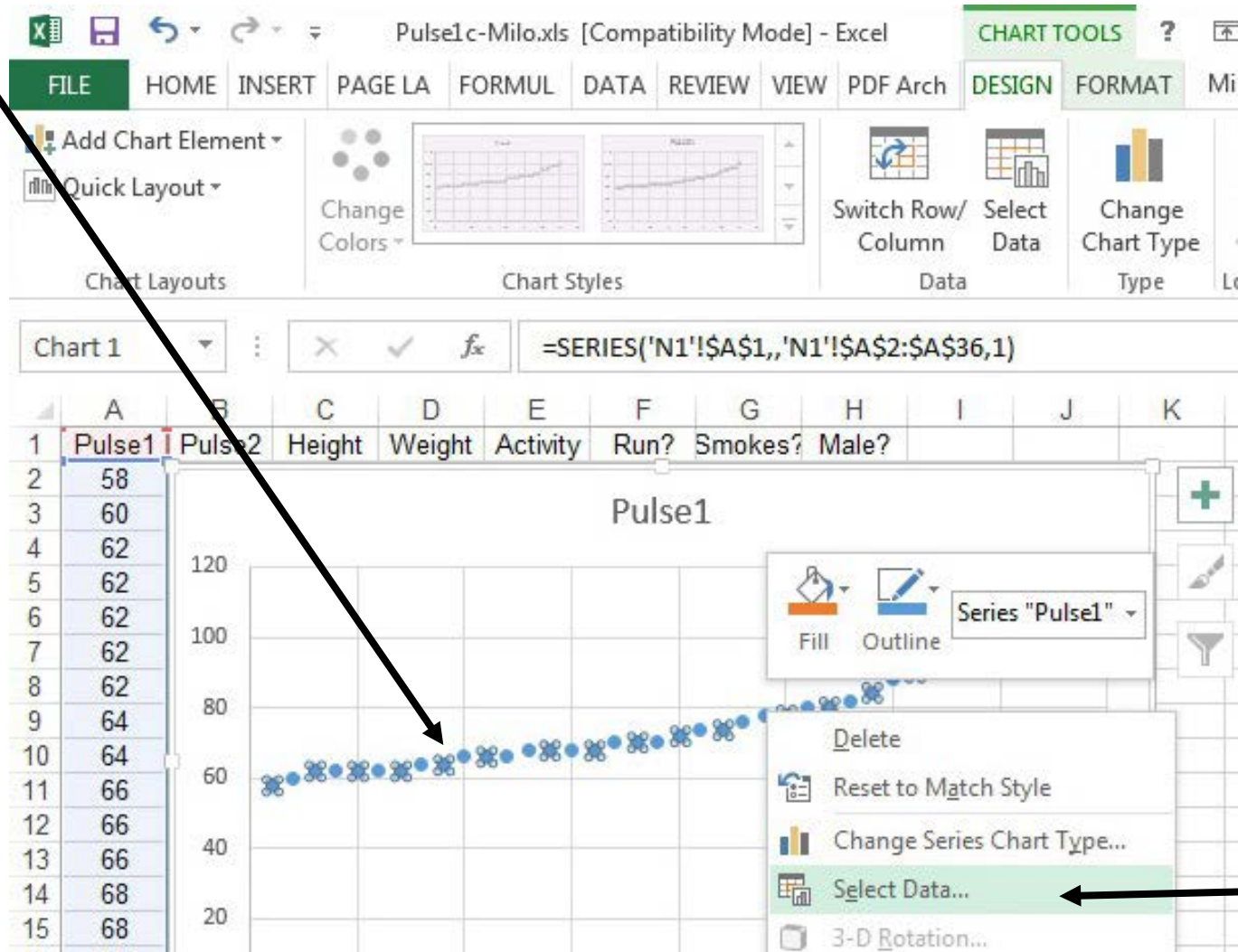
	A	B	C	D	E	F	G	H
26	78	118	69	145	2	1	0	0
27	80	96	72	155	2	1	0	1
28	80	128	68	125	2	1	0	0
29	82	100	68	138	2	1	0	0
30	84	84	72	150	3	1	0	1
31	88	110	69	150	2	1	1	0
32	90	94	74	160	1	1	1	1
33	92	84	70	153	3	1	1	1
34	96	140	61	140	2	1	0	0
35	96	116	68	116	2	1	0	0
36	100	115	63	121	2	1	1	0
37	48	54	68	150	1	0	1	1

# 1) Select Pulse1 for Run = 1. (Rows 2-36). Insert XY Plot



Rows 2-36

# 1a) Select Pulse1 series Right-mouse. "Select Data"



# 1b) To insert X values, select Edit

Chart data range: ='N1!\$A\$1:\$A\$36

Switch Row/Column

Legend Entries (Series)

Add Edit Remove

Pulse1

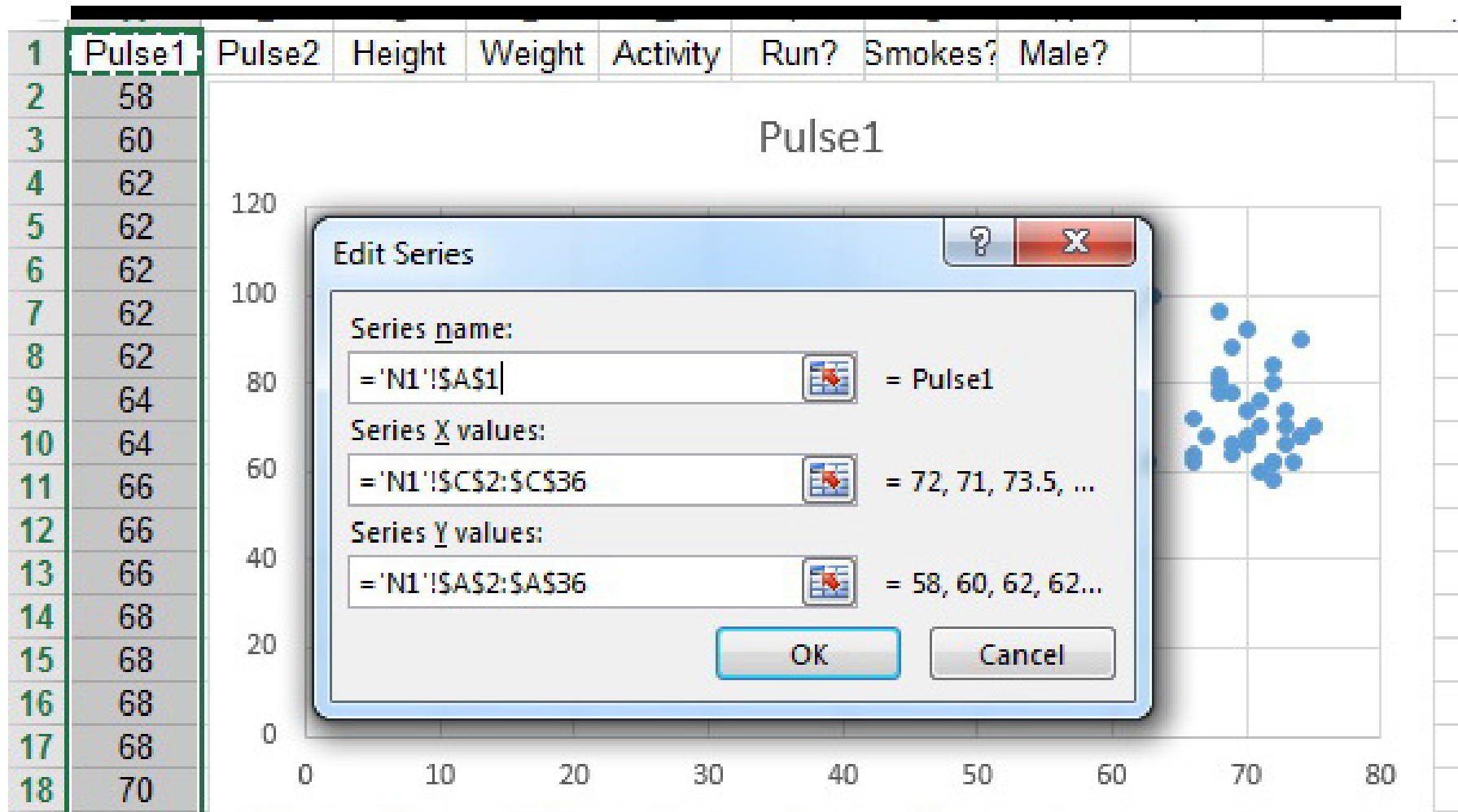
Horizontal (Category) Axis Labels

Edit

1 2 3 4 5

Hidden and Empty Cells OK Cancel

# 1c) Either select X values range or copy Y, paste to X and edit



**Pulse1 is Y axis (A2:A36); Height is X axis (C2:C36)**



# 1d) Add 2<sup>nd</sup> series (Pulse2 by Ht), Select data; Select "Add"

The screenshot shows an Excel spreadsheet with the following data:

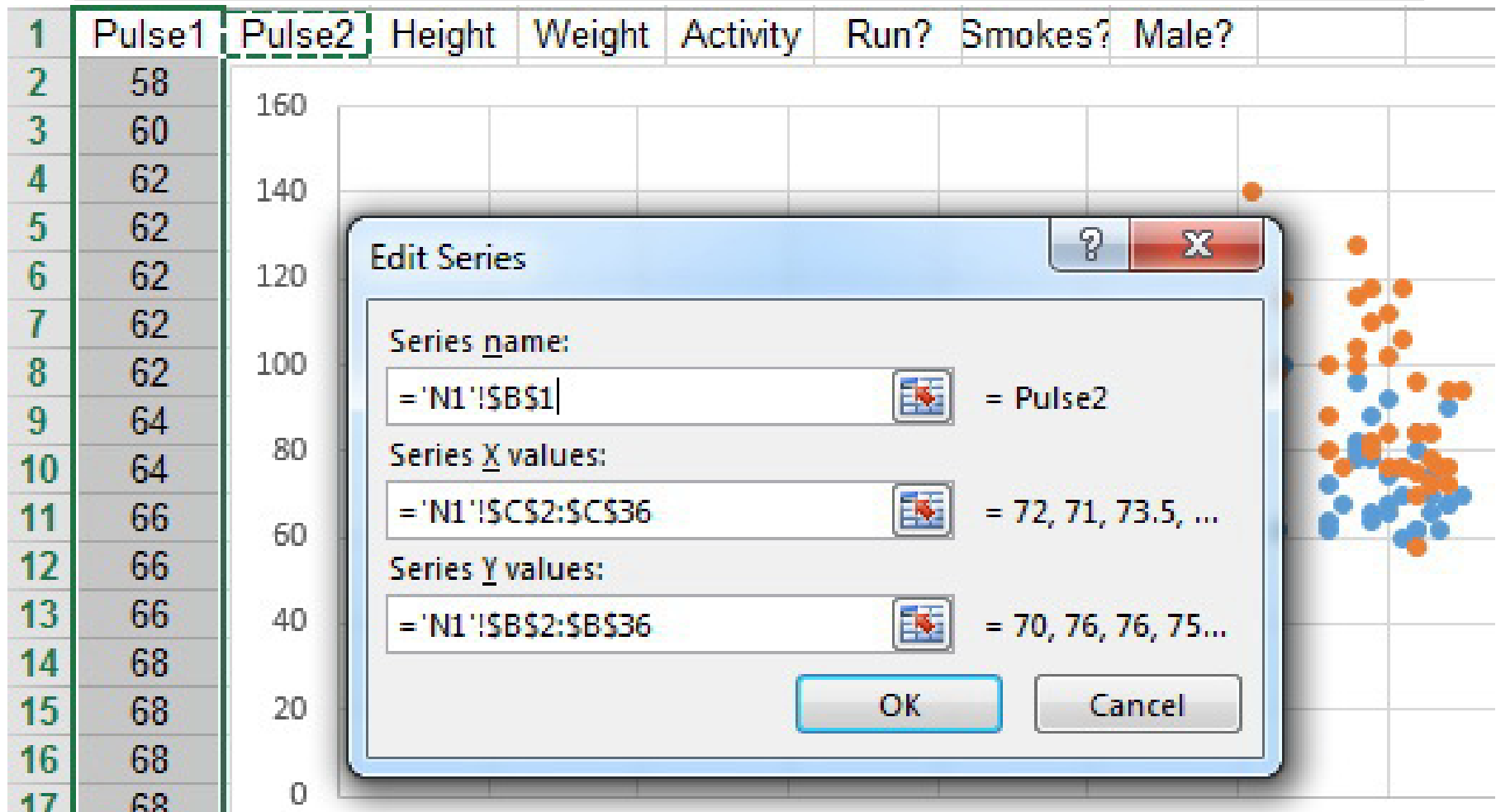
	Pulse1	Pulse2	Height	Weight	Activity	Run?	Smokes?	Male?
1								
2	58							
3	60							
4	62							

The 'Select Data Source' dialog box is open, showing the 'Legend Entries (Series)' panel with 'Pulse1' listed. An arrow points to the 'Add' button in the 'Legend Entries (Series)' panel.

The 'Horizontal (Category) Axis Labels' panel shows the following values:

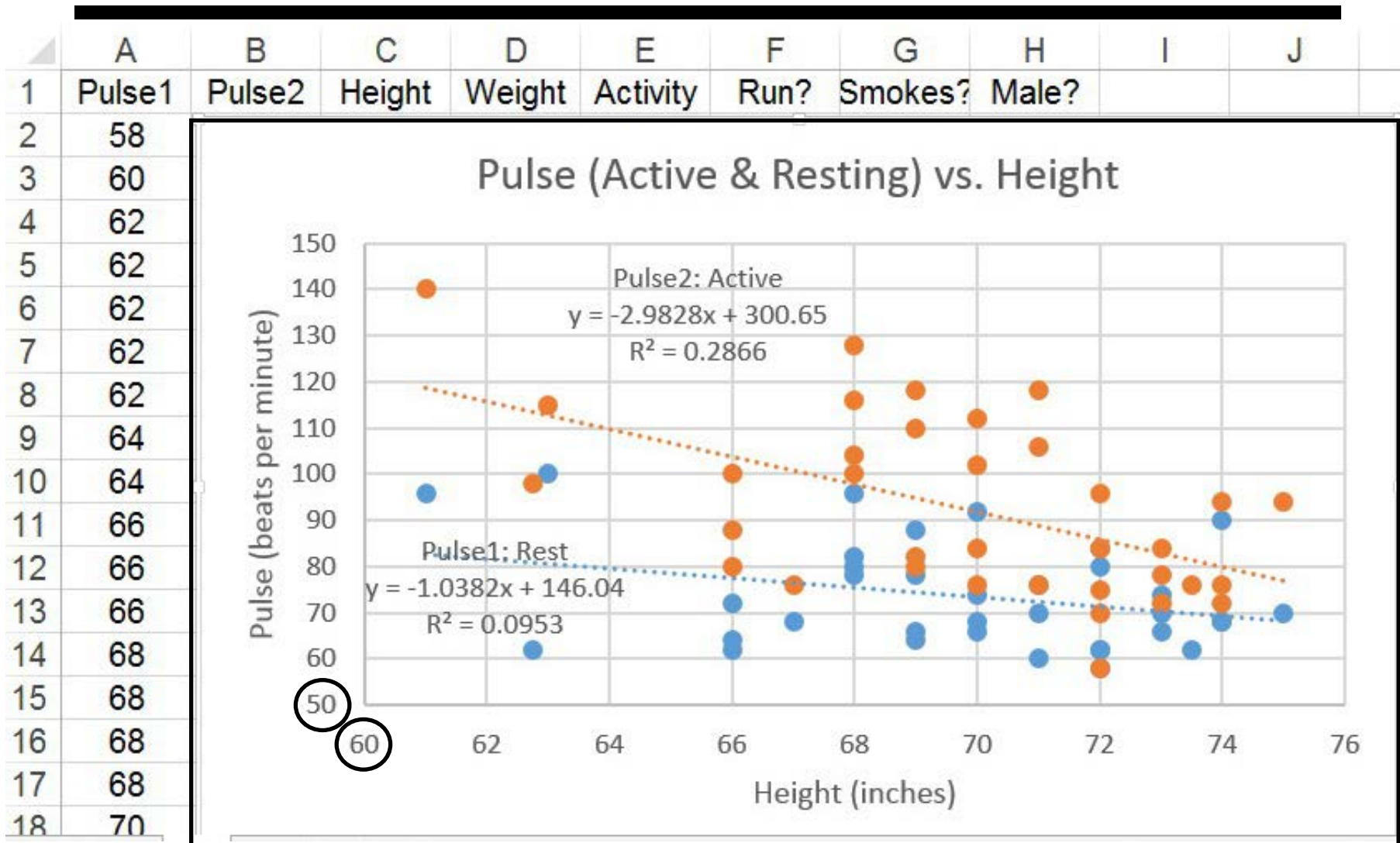
Horizontal (Category) Axis Labels
72
71
73.5
72
72

# 1e) Add Pulse2 (B2:B36) as Y; Add Height (C2:C36) as X



Press OK

# 1f) Result: Two Series same Y axis Format axis; add trendlines



## 2a) Copy data to a new Worksheet

Right mouse N1 tab.  
Select Move or Copy.  
Insert before Sheet2.  
Rename N1(2) as N2.  
Delete graph

Use new worksheet. Otherwise new sort will change previous graphs.

## 2b) Select all Pulse1 (A2:A93) Insert XY chart

The screenshot shows the Microsoft Excel 2013 ribbon with the 'INSERT' tab selected. The 'Charts' group is active, and a tooltip for 'Insert Scatter (X, Y) or Bubble Chart' is displayed. The spreadsheet data is visible below the ribbon.

	A	B	C	D	E			
1	Pulse1	Pulse2	Height	Weight	Activity			
2	58	70	72	145	2			
3	60	76	71	170	3			
4	62	76	73.5	160	3			
5	62	75	72	195	2	1	0	1
6	62	58	72	175	3	1	0	1
7	62	100	66	120	2	1	0	0
8	62	98	62.75	112	2	1	1	0
9	64	88	66	140	2	1	0	1

**Insert Scatter (X, Y) or Bubble Chart**  
Use this chart type to show the relationship between sets of values.

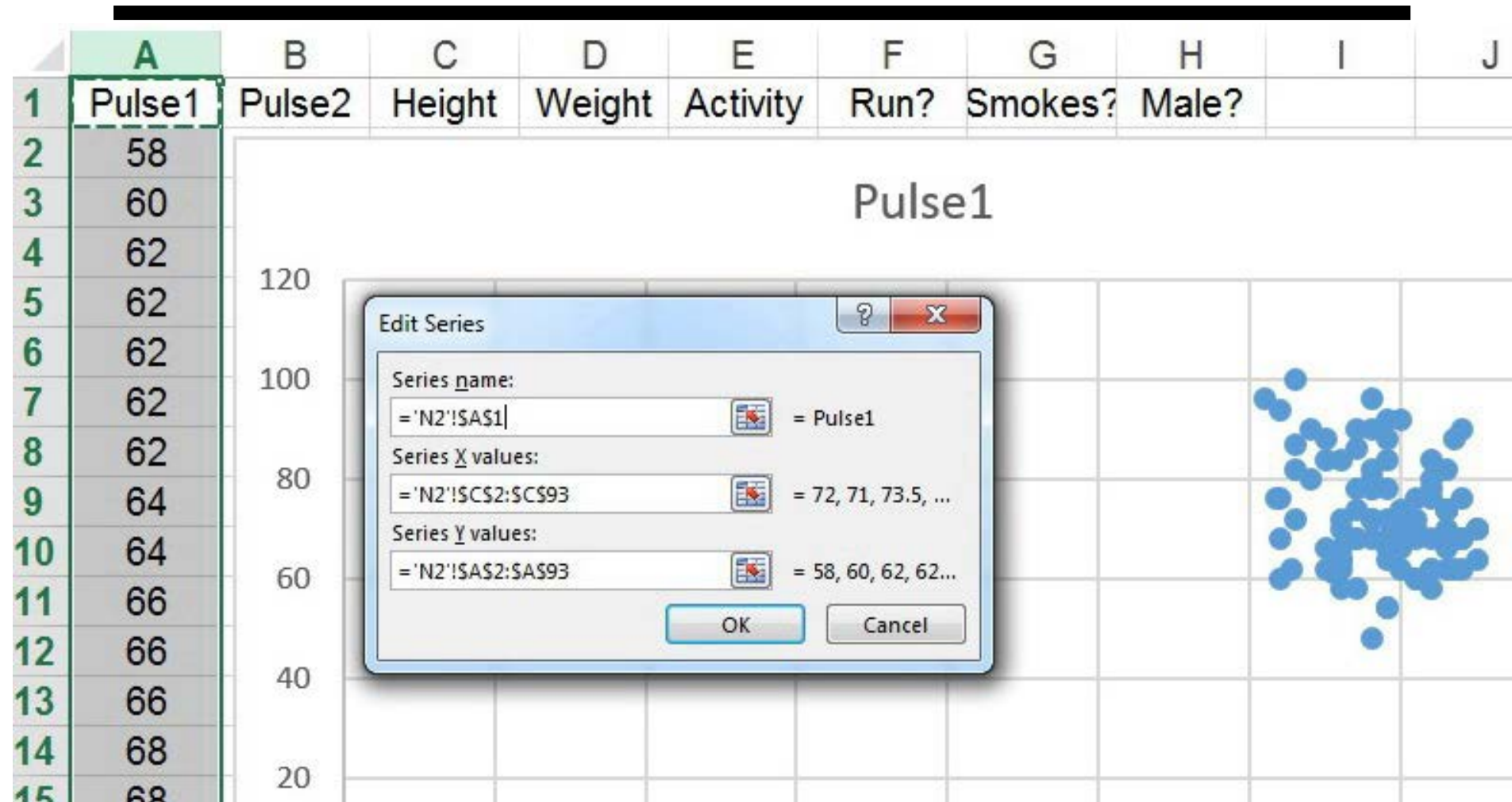
Click the arrow to see the different types of scatter and bubble charts available and pause the pointer on the icons to see a preview in your document.

## 2c) Initial Chart: 1Y, no X





## 2e) Edit Pulse1 by Height series: X (C2:C93), Y (A2:A93)



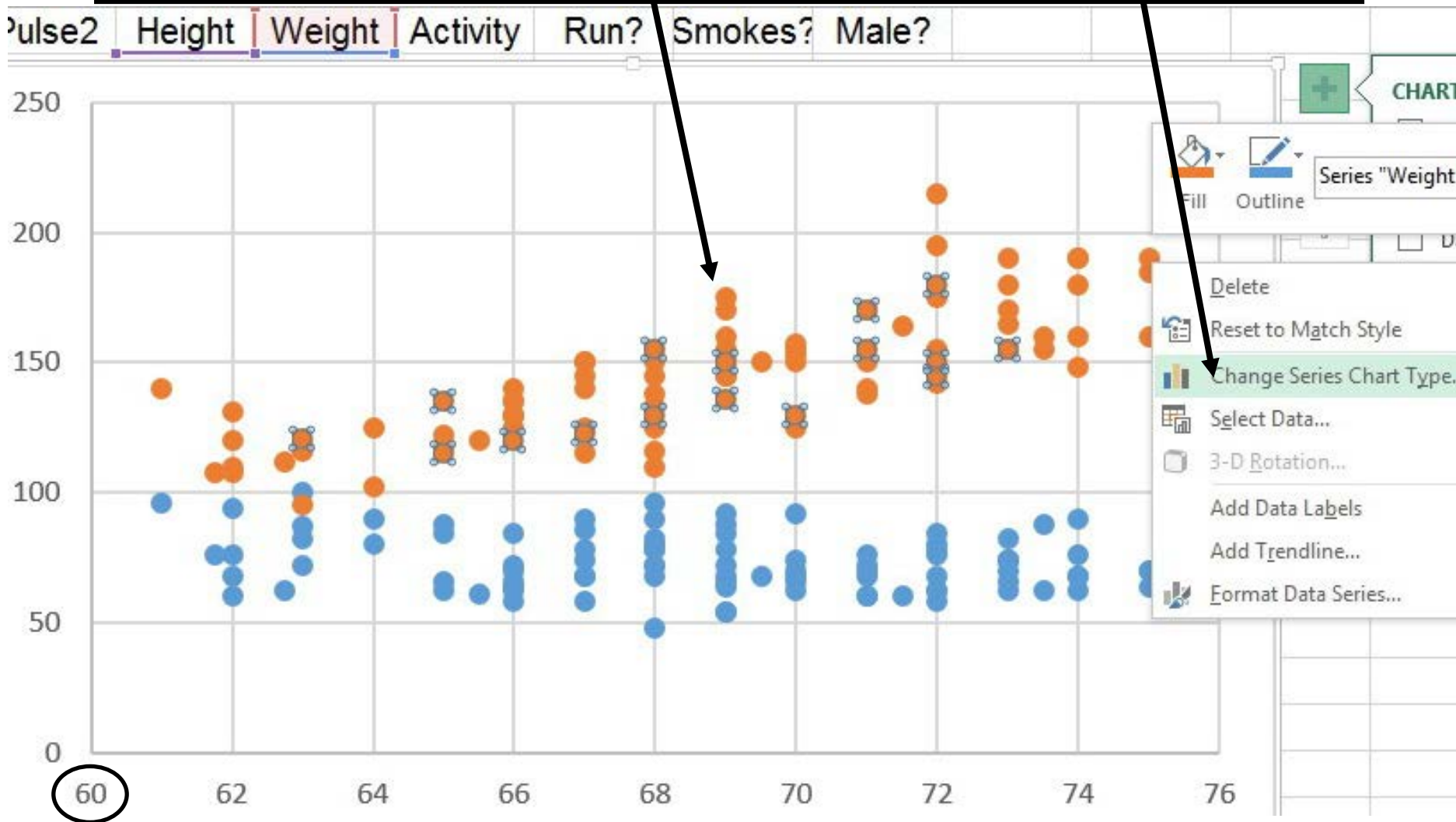




## 2g) Edit Weight by Height Series X (C2:C93), Y (D2:D93)



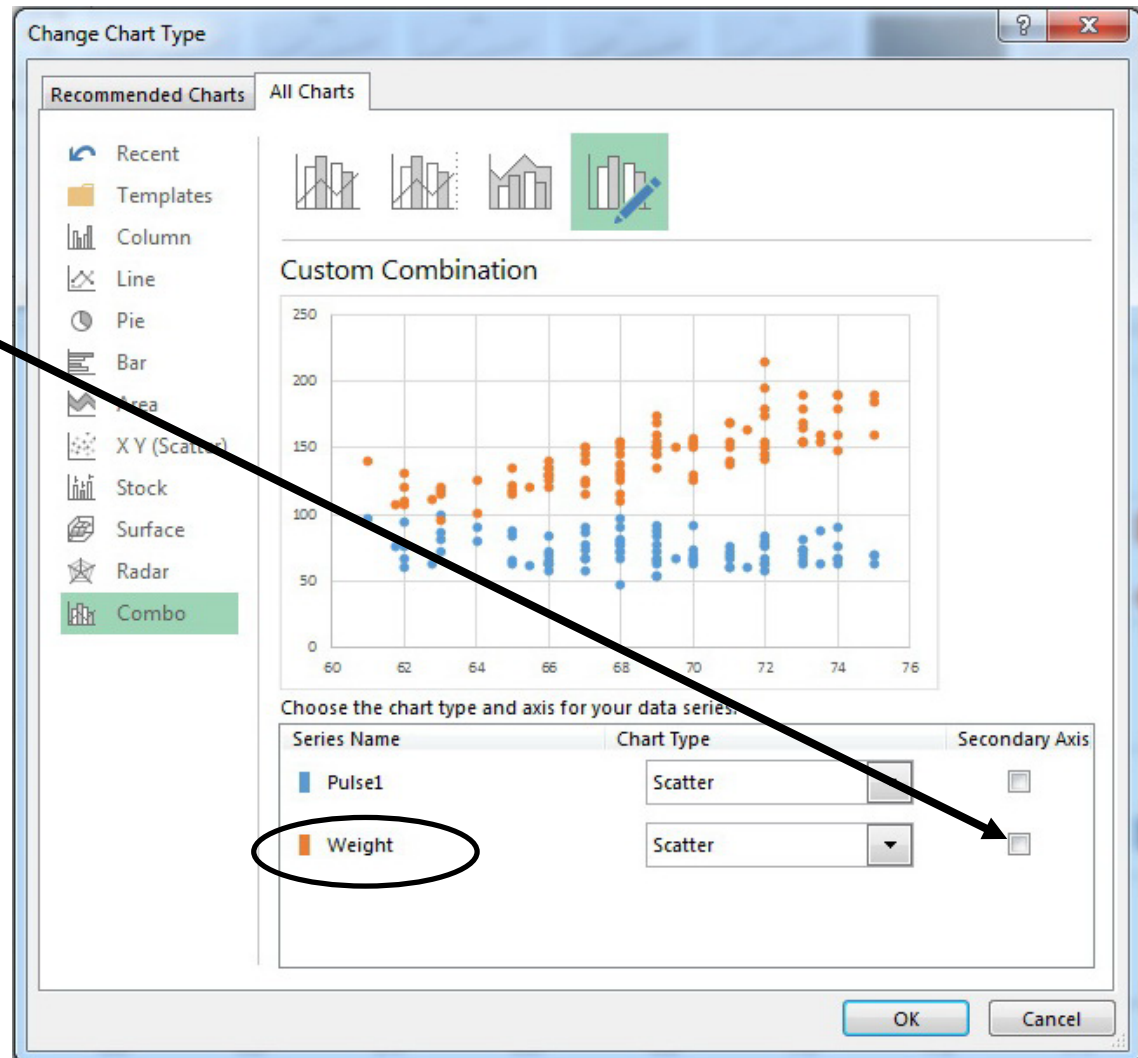
# 2h) To Introduce Second Y Axis: Right-mouse Weight; Select "Change Type"



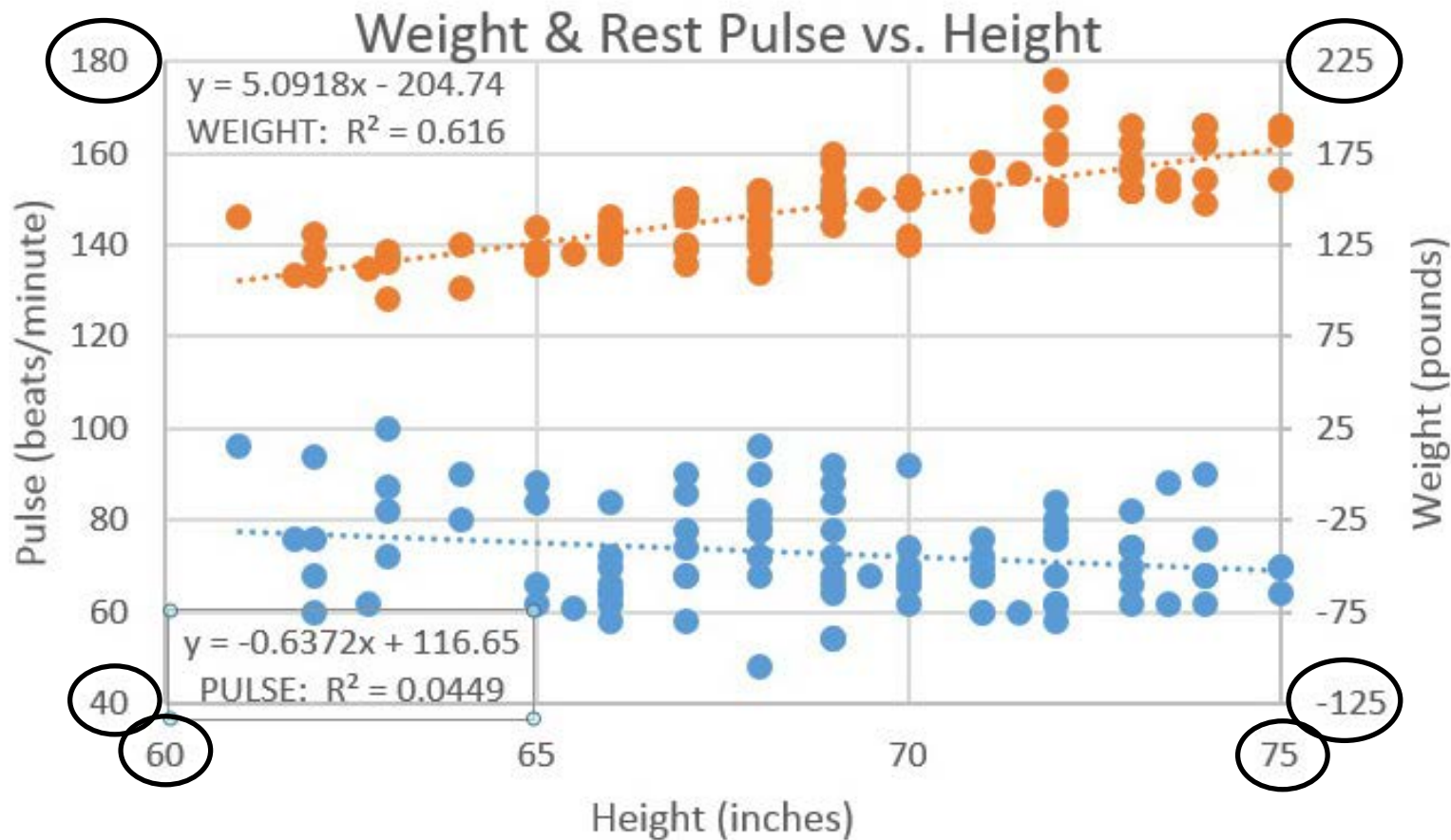
## 2i) Check Secondary Y Axis for Weight. Press OK

Add check to check box for Secondary Axis for “Weight”.

Do not press “Combo”;  
Do not select a particular chart shape.



## 2j) Two Series; Separate Y Axis. Add Trendline w Model for each.



Left-Y-axis: Max=180.

Right-Y-axis: Min = -125; Max=225

## 3a) Right Mouse N2 tab; Copy N2; Rename as N3; Delete graph



Use new worksheet. Otherwise new sort will change previous graphs.

## 3b) Select ALL columns and rows; From Home or Data menu, select Sort

	A	B	C	D	E	F	G	H
1	Pulse1	Pulse2	Height	Weight	Activity	Run?	Smokes?	Male?
2	58	70	72	145	2	1	0	1
3	60	76	71	170	3	1	0	1
4	62	76	73.5	160	3	1	1	1
5	62	75	72	195	2	1	0	1
6	62	58	72	175	3	1	0	1
7	62	100	66	120	2	1	0	0
8	62	98	62.75	112	2	1	1	0
9	64	88	66	140	2	1	0	1
10	64	80	69	155	2	1	0	1
11	66	78	73	190	1	1	1	1
12	66	82	69	175	2	1	1	1
13	66	102	70	130	2	1	0	1
14	68	72	74	190	2	1	0	1
15	68	76	67	145	2	1	0	1
16	68	76	74	180	2	1	1	1
17	68	112	70	125	2	1	0	0
18	70	72	73	170	3	1	1	1

## 3c) From Sort-by, select RUN; Order from Small-to-Large

	A	B	C	D	E	F	G	H
1	Pulse1	Pulse2	Height	Weight	Activity	Run?	Smokes?	Male?
2	58	70	72	145	2	1	0	1
3	60	76	71	170	3	1	0	1
4	Sort dialog box							
5	Sort dialog box							
6	Sort dialog box							
7	Sort dialog box							
8	Sort dialog box							
9	Sort dialog box							
10	Sort dialog box							
11	Sort dialog box							
12	Sort dialog box							
13	Sort dialog box							
14	Sort dialog box							
15	68	76	67	145	2	1	0	1
16	68	76	74	180	2	1	1	1
17	68	112	70	125	2	1	0	0



# 3d) Select A1:B58. Insert XY Chart

Select data with Run=0

	A	B	C	D	E	F	G	H
48	84	84	66	130	2	0	0	0
49	84	80	65	118	1	0	0	0
50	86	84	67	150	3	0	0	0
51	87	84	63	95	3	0	0	0
52	88	84	73.5	155	2	0	0	1
53	88	74	65	135	2	0	1	0
54	90	88	67	140	2	0	1	1
55	90	90	68	145	1	0	0	1
56	90	92	64	125	1	0	1	0
57	92	94	69	150	2	0	1	1
58	94	92	62	131	2	0	1	0
59	58	70	72	145	2	1	0	1
76		71	170	3	1	0	1	
76		73.5	160	3	1	1	1	
75		72	195	2	1	0	1	
58		72	175	3	1	0	1	
100		66	120	2	1	0	0	
98		62.75	112	2	1	1	0	

**Edit Series**

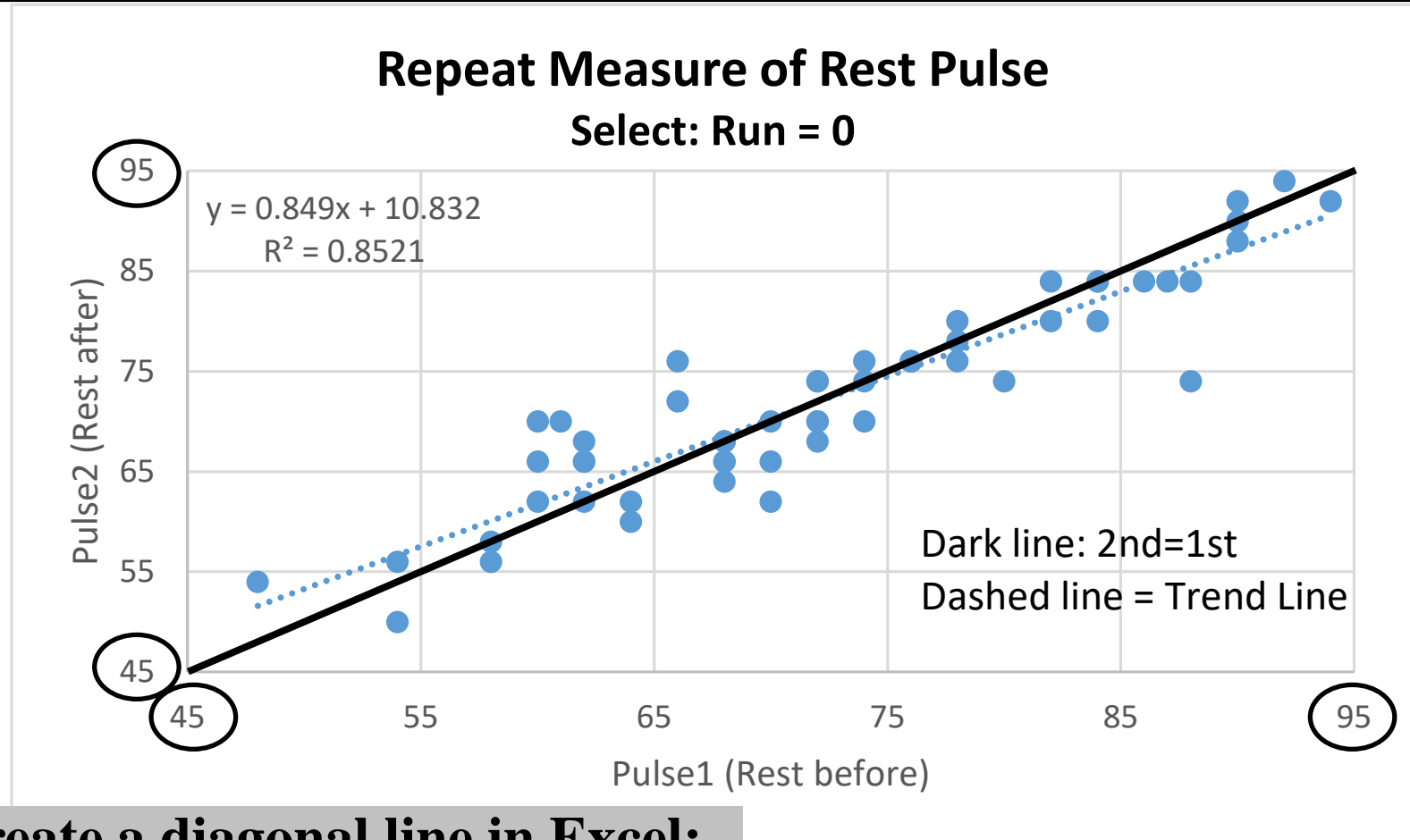
Series name:  = Pulse2

Series X values:  = 48, 54, 54, 58...

Series Y values:  = 54, 56, 50, 58...

# 3e) XY plot: Pulse2 vs Pulse1

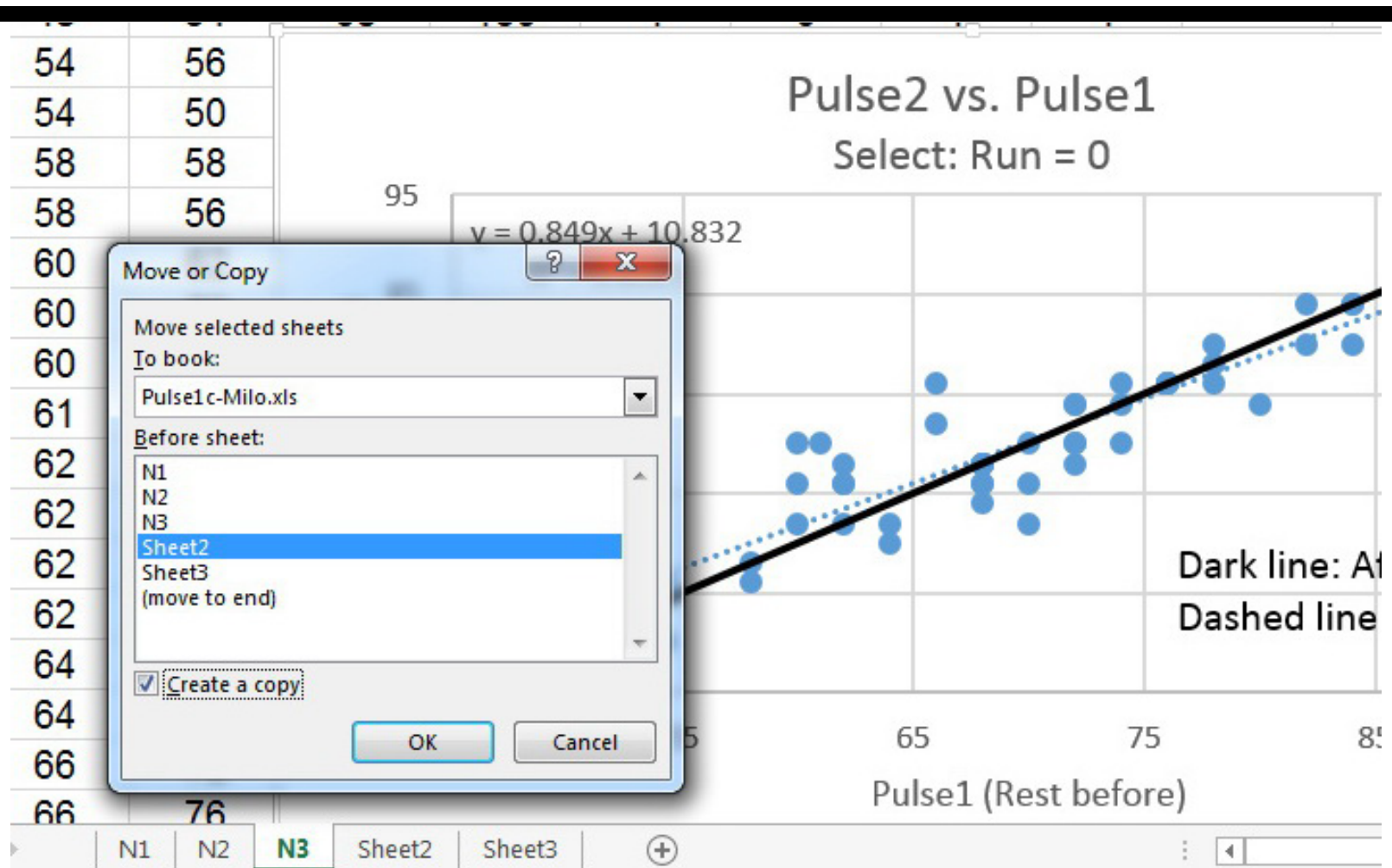
## Add "No Change" line: 45,45 - 95,95



To create a diagonal line in Excel:

Insert/ Shapes OR Chart Tools/Format/Insert Shapes: Select Line.

# 4a) Right-Mouse N3 Tab; Copy N3 Rename as N4. Delete graph



Use new worksheet. Otherwise new sort will change previous graphs.

## 4b) Select all data; Sort by Run; Sort Values: Large to Small

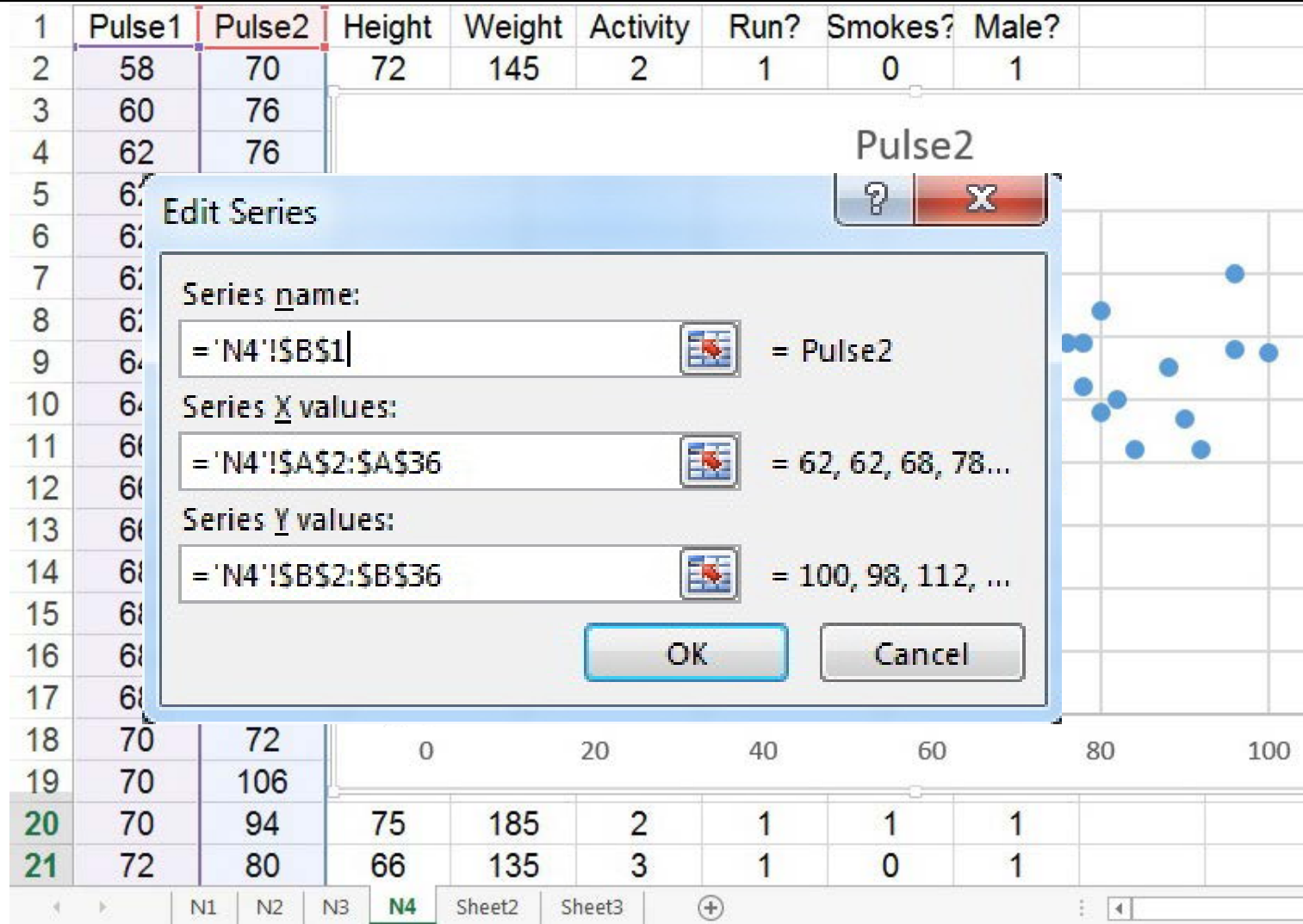
1	Pulse1	Pulse2	Height	Weight	Activity	Run?	Smokes?	Male?
2	48	54	68	150	1	0	1	1
3	54	56	69	145	2	0	1	1
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15	64	62	75	160	3	0	0	1

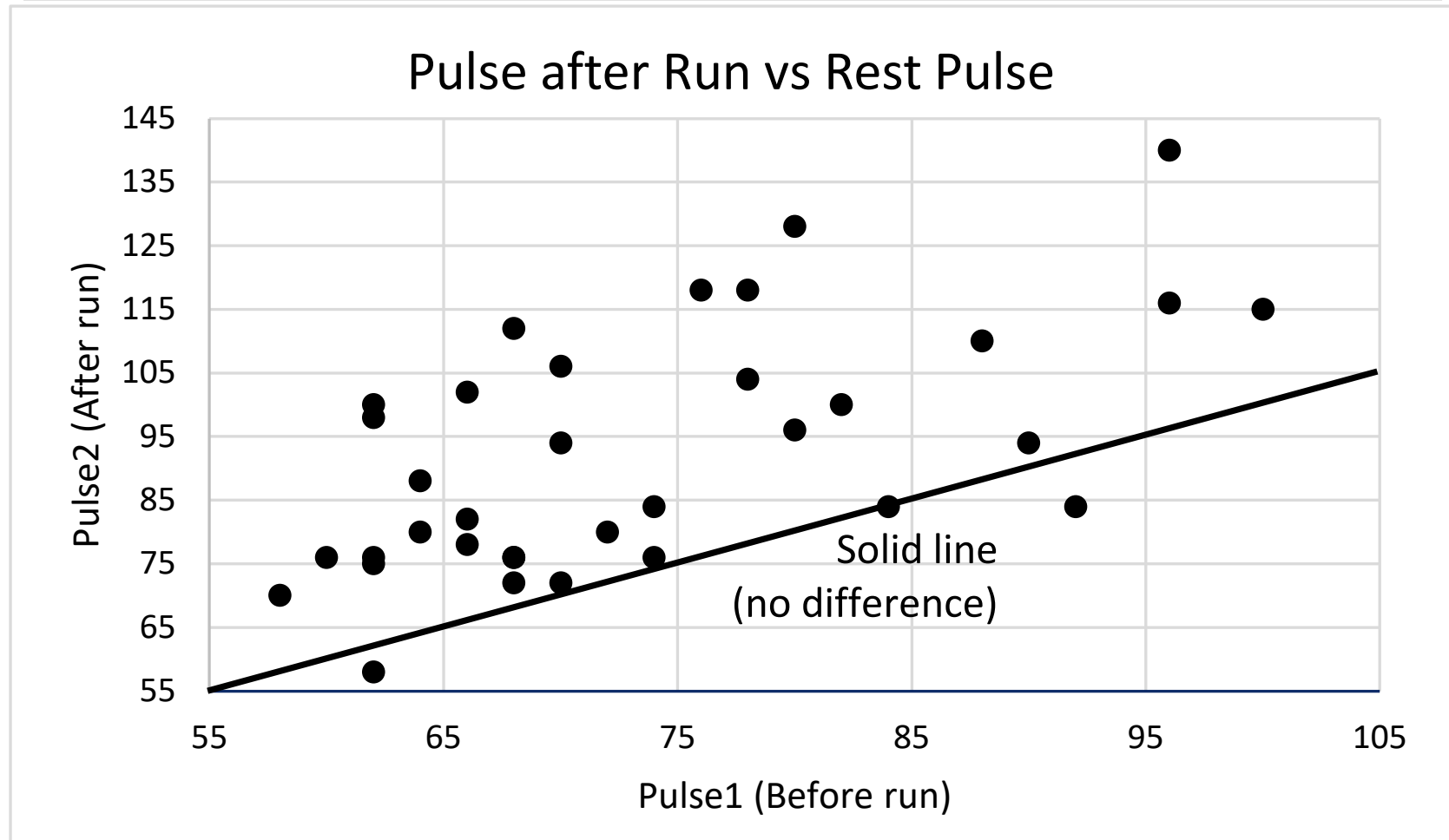
Sort dialog box configuration:

- My data has headers:
- Sort by: Run?
- Sort On: Values
- Order: Largest to Smallest

## 4c) Select Pulse1 & 2 columns, Rows 2:36. Insert XY Chart



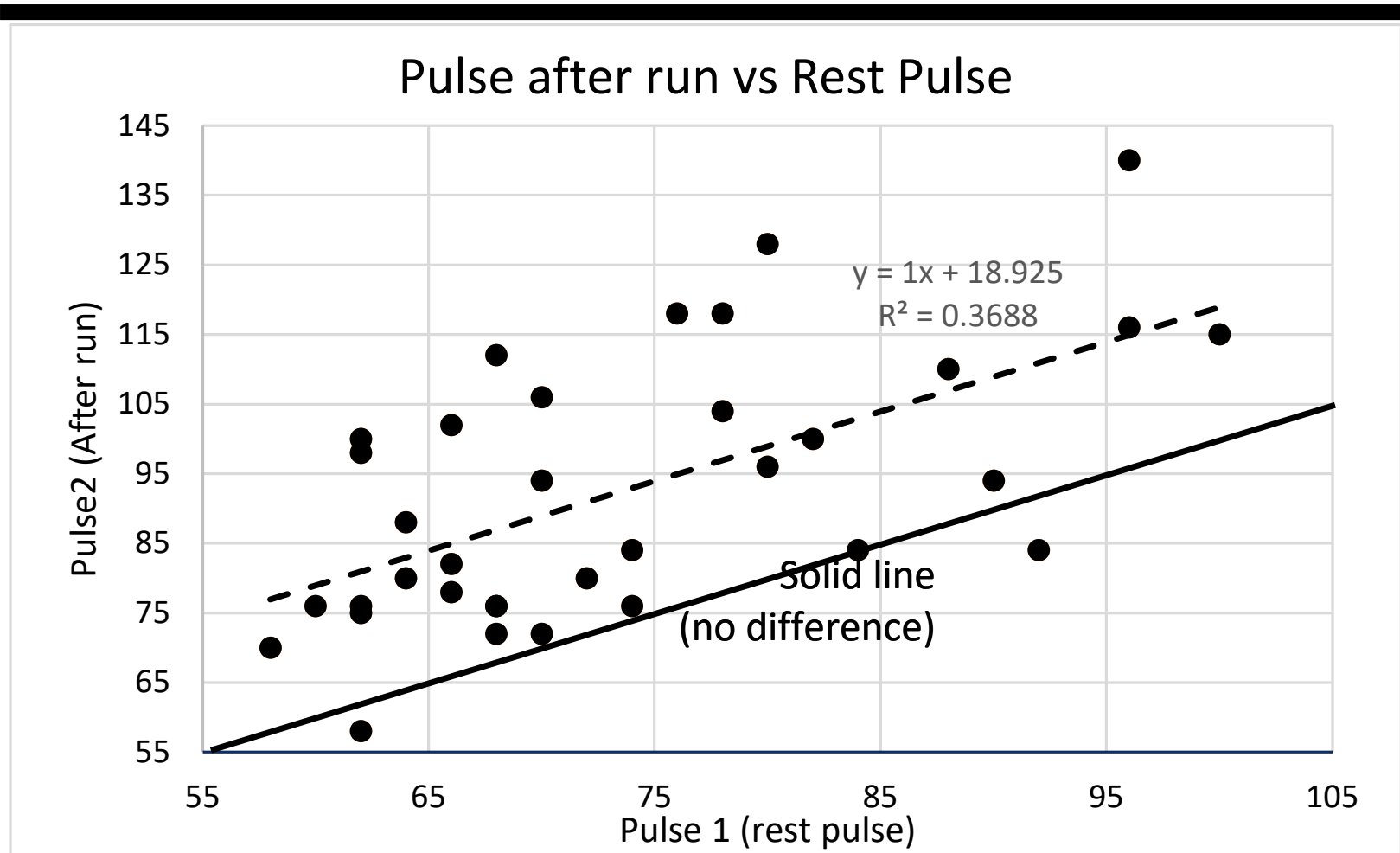
## 4d) Insert no-change line



**Excel: Insert/Illustrations/Shapes. Select Line**

# 4e) Set Intercept = 18.925

## Create trend-line and equation



**Note: This slide is slightly different from the previous slide.**

## Summary of N3 and N4

---

N3: Test-retest of rest pulse has slope of 0.849.  
A perfect test-retest would have a slope of 1.00.  
Explanation: Regression to the mean.

R-squared = 85%. The original resting pulse explains 85% of variability in the retest of pulse.

N4: Running in place increases pulse by 18.9 bpm over rest pulse when model has unit slope.

R-squared = 37%. Rest pulse explains 37% of the variation in pulse after running in place.



## Two Common Problems

---

1. Including column heading in data selection. You may include top row when you select data. Normally Excel excludes this so the data starts in row 2. Sometimes Excel does not exclude row 1 and this really messes up your graph! Solution: Select Data and edit the X and Y data ranges manually.
2. Getting X (horizontal axis) and Y (vertical axis) reversed. There is no simple way to fix this. You need to right mouse the data, select “select data” and Edit the data in question. Manually make the changes necessary without making things worse...

## Two More Problems

---

1. Sometimes your data just gets all mixed up. The most likely cause: you sorted just a part of the data instead of sorting the entire data set. Solution. Copy the data from a previous page. Paste it on the desired page and then repeat the sort.
2. Sometimes when you sort the data, a prior graph goes bad. The simplest explanation is that both graphs are 'pointing' at the same data. Solution: Make sure each graph has its own unique data on its own unique page. You may need to delete and recreate graphs too.