

# Summarizing and Displaying Data in Excel

Debby Kermer

University Libraries

Digital Scholarship Center (DiSC)

<https://etherpad.wikimedia.org/p/GMU-GradRecon-Excel>

# Agenda

1. Understanding Data Tables
2. Pivot Tables
3. Charts and Graphs

Datafile:

<http://dataservices.gmu.edu/files/titanic.csv>

# Delimited Data Files

## Same Data, Different Formats

### Comma Delimited

*Use this unless commas are in your data*

```
caseid,wght,zip,osmp,cnty,flp3,flp2,flp4
8451,0.327,99999,1,999,1,4,2,8,3,3,44,3,
10533,0.513,99999,1,999,4,3,3,8,3,3,44,3,
4262,1.305,99999,9,999,2,3,3, ,2,2,14, ,
15157,0.251,99999,1,999,2,3,3, ,2,2,14, ,
781,0.305,99999,1,999,4,3,3,8,3,3,44,3,
13619,0.682,99999,1,999,1,3,3,8,3,3,47,3,
8291,0.749,99999,1,999,1,1,1,3,2,2,36,2,
10737,0.696,99999,1,999,4,4,1,16,2,2,14,
674,2.465,99999,9,999,4,4,2, ,2,2,15, ,
2366,5.214,99999,9,999,2,4,2, ,3,3,10, ,
4305,1.044,99999,9,999,3,4,2, ,3,3,19, ,
16261,0.21,99999,1,999,1,4,3,2,2,2,23,5,
14447,0.305,99999,1,999,4,2,1,23,4,4,5,3,
13619,0.682,99999,1,999,1,3,3,8,3,3,47,3,
8291,0.749,99999,1,999,1,1,1,3,2,2,36,2,

```

commas, commas, and values

### Tab Delimited

*Easier to examine, if your data has no tabs*

```
caseid wght zip osmp cnty flp3
8451 .327 99999 1 999 1 4 2
10533 .513 99999 1 999 4 3 3
4262 1.305 99999 9 999 2 3 3
15157 0.251 99999 1 999 2 3 2
781 0.305 99999 1 999 4 3 3
13619 0.682 99999 1 999 1 3 3
8291 0.749 99999 1 999 1 1 1
674 2.465 99999 9 999 4 4 2
2366 5.214 99999 9 999 2 4 2
4305 1.044 99999 9 999 3 4 2
16261 0.21 99999 1 999 1 4 3 2
14447 0.305 99999 1 999 4 2 1

```

looks like columns, but not all lined up

### Spreadsheet

	A	B	C	D	E	
1	caseid	wght	zip	osmp	cnty	flp3
2	8451	0.327	99999	1	999	
3	10533	0.513	99999	1	999	
4	4262	1.305	99999	9	999	
5	15157	0.251	99999	1	999	
6						
7						
8						
9						
10						
11	674	2.465	99999	9	999	
12	2366	5.214	99999	9	999	
13	4305	1.044	99999	9	999	
14	16261	0.21	99999	1	999	
15	14447	0.305	99999	1	999	
16	13619	0.682	99999	1	999	
17	8291	0.749	99999	1	999	

proprietary format, can save as csv or tsv

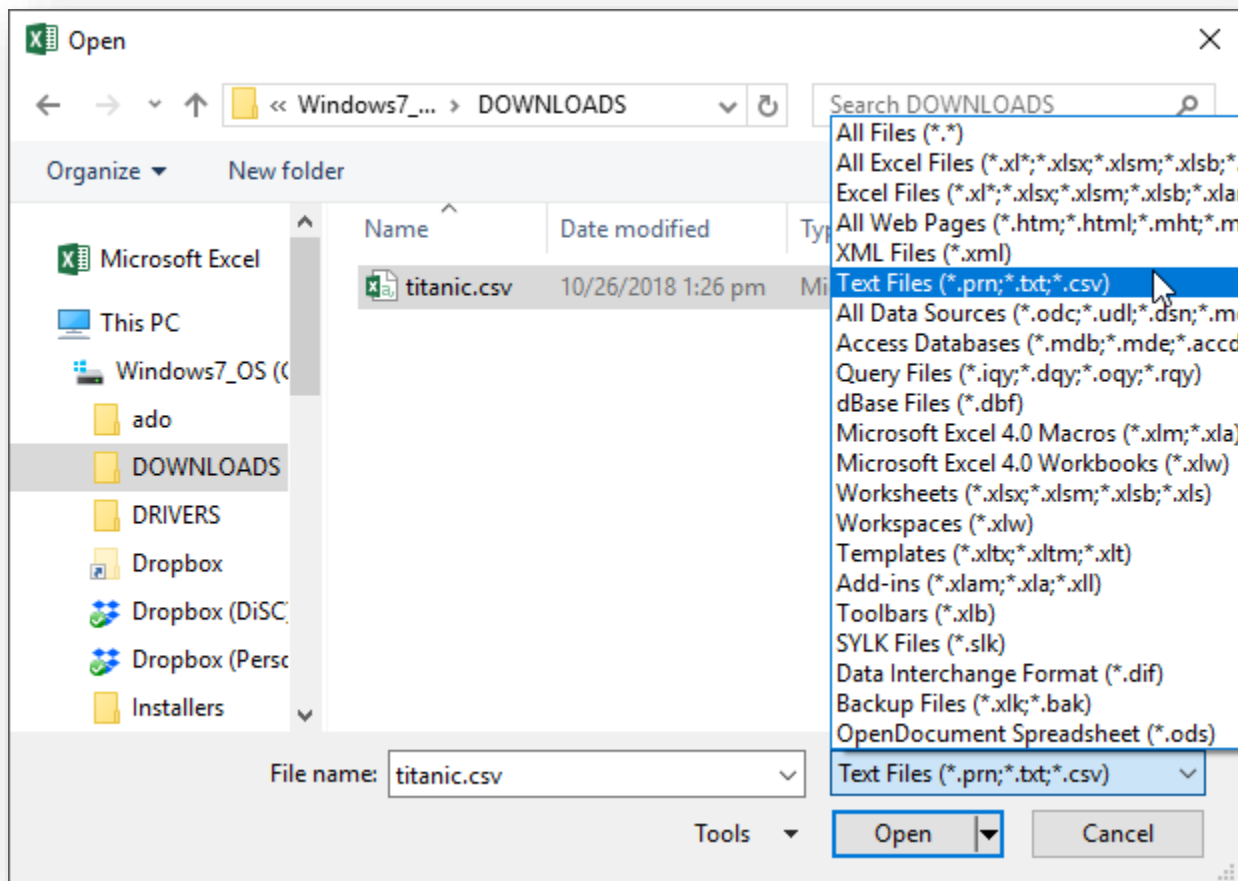
Datafile:

<http://dsc.gmu.edu/files/titanic.csv>

# Download and Open the CSV File

Datafile:

<http://dsc.gmu.edu/files/titanic.csv>



# Understanding Data Tables

# Variables & Observations

## Variable

*a characteristic*

**Did they survive?**

## Value

*characteristic of thing*

**Person #2 was in 3<sup>rd</sup> Class**

	A	B	C	D	E	F	G	H
1	id	name	survived	age	gender	sibsp	parch	pclass
2	1	Abbing, Mr. Anthony	0	42	male	0	0	3
3	2	Abbott, Master. Eugene	0	13	male	0	1	3
4	3	Abbott, Mr. Rossmore	0	16	male	1	1	3
5	4	Abbott, Mrs. Stanton	1	35	female	1	1	3
6	5	Abelseth, Miss. Karen	1	16	female	0	0	3
7	6	Abelseth, Mr. Olaus	1	25	male	0	0	3
8	7	Abelson, Mr. Samuel	0	30	male	1	0	2
9	8	Abelson, Mrs. Samuel	1	28	female	1	0	2
10	9	Abrahamsson, Mr. Ab	1	20	male	0	0	3
11	10	Abraham, Mrs. Joseph	1	18	female	0	0	3
12	11	Adahl, Mr. Mauritz N	0	30	male	0	0	3
13	12	Adams, Mr. John	0	26	male	0	0	3
14	13	Ahlin, Mrs. Johan (Jo	0	40	female	1	0	3
15	14	Aks, Master. Philip Fr	1	0.83	male	0	1	3
16	15	Aks, Mrs. Sam (Leah F	1	18	female	0	1	3
17	16	Albimona, Mr. Nasse	1	26	male	0	0	3
18	17	Aldworth, Mr. Charles	0	30	male	0	0	2
19	18	Alexander, Mr. William	0	26	male	0	0	3
20	19	Alhomaki, Mr. Humeri	0	20	male	0	0	3

**Case/Observation**

*thing described*

**Passenger (person)**

# Excel Table Requirements

Head each column with a short, **unique** label

All information should be represented by **values**

Do not use formatting AS data

NO empty **rows** or **columns** within the data

**Unique Names  
in the 1<sup>st</sup> Row**

	A	B	C	D	E	F	G	H	I	J	K
1	id	name	survived	age	gender	sibsp	parch	pclass	fare	embarked	
2	1	Abbing, Mr. Anthony	0	42	male	0	0	3	7.55	S	
3	2	Abbott, Master. Euger	0	13	male	0	2	3	20.25	S	
4				16	male	1	1	3	20.25	S	
5				35	female	1	1	3	20.25	S	
6				16	female	0	0	3	7.65	S	
7				25	male	0	0	3	7.65	S	
8				30	male	1	0	2	24	C	
9	8	Abelson, Mrs. Samuel	1	28	female	1	0	2	24	C	
10	9	Abrahamsson, Mr. Abr	1	20	male	0	0	3	7.925	S	

Each row includes  
all information  
(no headers)

No blank rows,  
formatting, or  
extra values

# Do not worry about it being "Pretty"

Store **original** values, keep original precision

Use formatting to round values and add commas

Have a **single** worksheet if the columns are the same.

Add **new columns** to store information or groupings



# Zero

If the value is **Zero**, **None**, or **False**, type a **0**

If it is unknown, use **blank**

To record reasons for this, use extreme, distinctive values

	Correct	WRONG
None or FALSE →	0	-
{ Value is unknown	99	n/a
	-1	??

**Why?** Formatting will not be saved in csv files. Non-numeric values could cause problems importing into statistical software.

# Groups / Attributes

Generally, use **words** (labels) to avoid confusion.

Be consistent!

**Numbers** may be useful for **Ordered** values

Grouped		Ordered	
GOOD	AVOID	GOOD	AVOID
Virginia	2	5	Very Satisfied
Maryland	3	2	Somewhat Unsatisfied
Virginia	2	3	Neutral
DC	1	4	Somewhat Satisfied
Maryland	3	1	Very Unsatisfied

# Using Binary

Pick **one** of the two groups and put **that** label at the top of the column (the variable name).

Assign values so that 1 represents that group.

**1** = Yes, Selected, True

**0** = No, Unselected, False

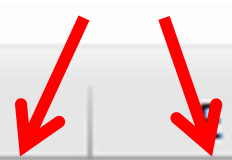
If 2 groups, use **Binary** (1 and 0)

Easy & fast to type

Mathematically Useful

- To **count** the 1's, use the **sum**
- To get the **proportion** or % of 1's, use the **average**

Pick ONE

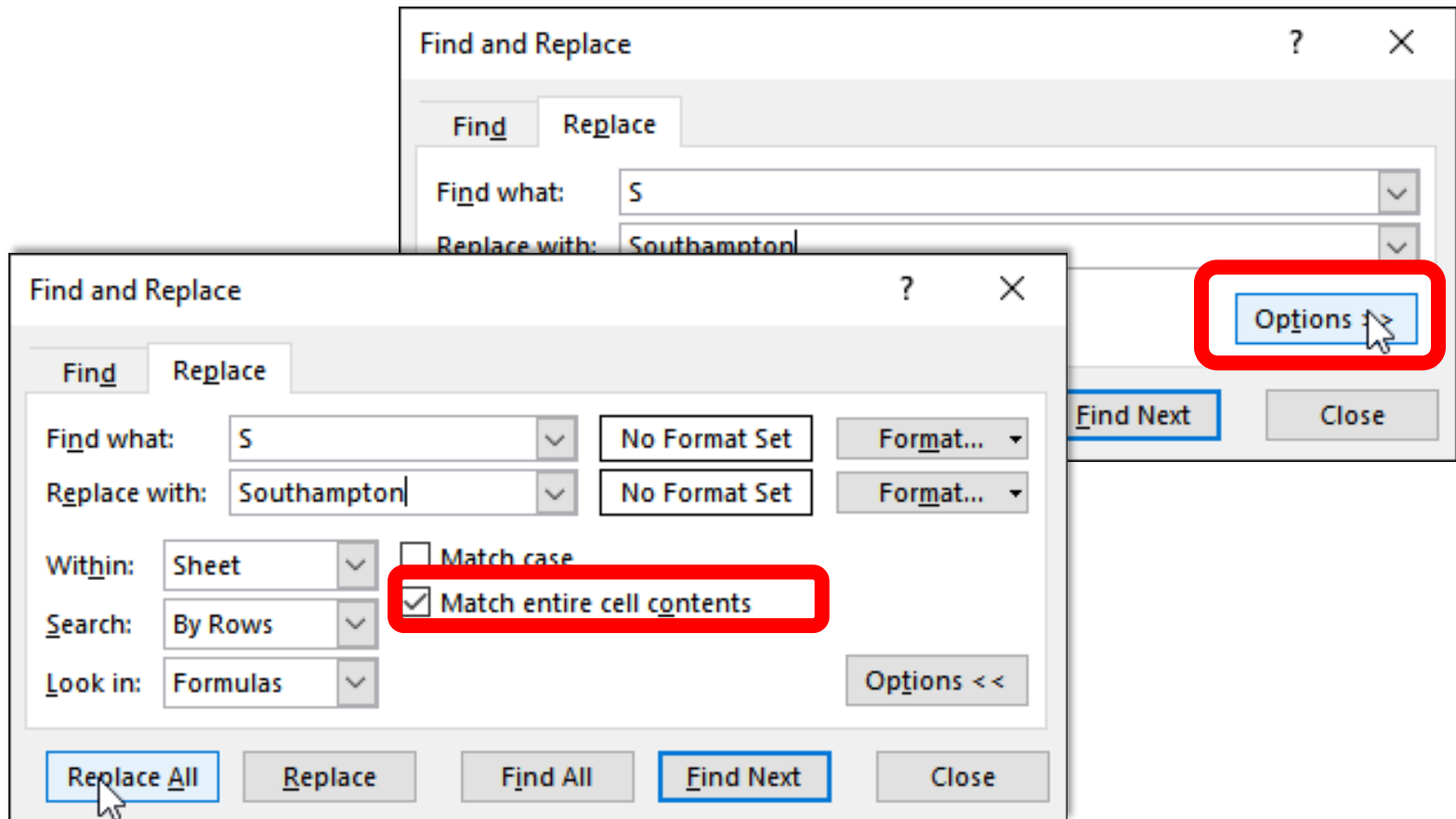


	C	D	E	
	fate	survived	died	age
	died	0	1	
	died	0	1	
	died	0	1	
	survived	1	0	
	survived	1	0	
	survived	1	0	
	died	0	1	
	survived	1	0	
	survived	1	0	

# Avoid Changing Values Individually

If needed, use **Find and Replace** (Ctrl-H)

Always choose "Match entire cell contents"



# Dates

What's First?		
Month (US)	4/10/2018	AVOID
Day (UK)	10/04/2018	AVOID
Year	2018-04-10	GOOD

Use Year-Month-Day (e.g., 2015-05-23)

Avoids misunderstandings *and* sorts properly

## Useful Keyboard Shortcuts

Current Date: Ctrl + ;

Current Time: Ctrl + Shift + ;

Double-check when importing into other software

Common source of error

Excel stores dates as the # of days since 1/1/1900

# Keyboard Shortcuts

To select entire rows, columns, or data tables

**Rows = Shift**



**Columns = Ctrl**



**Select = Space**



**Shift + Ctrl + Space → Select all data**

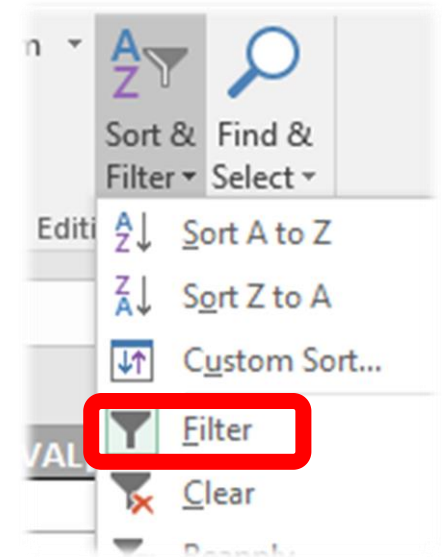
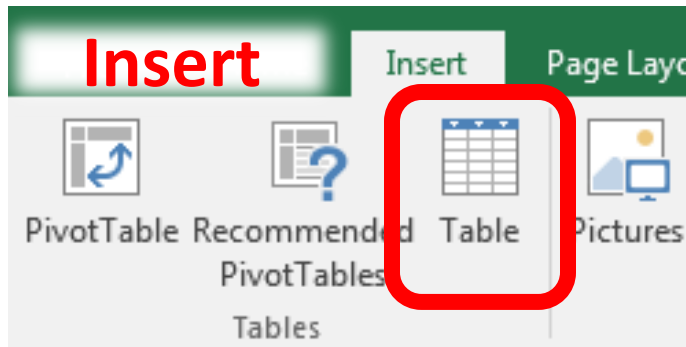
# Excel Table Objects

Datafile:

<http://dataservices.gmu.edu/files/titanic.csv>

# Make Excel Recognize your Table

1. Click in the Data
2. Click Table (Insert Menu)  
OR  
Click Filter under Sort & Filter (Home Menu)

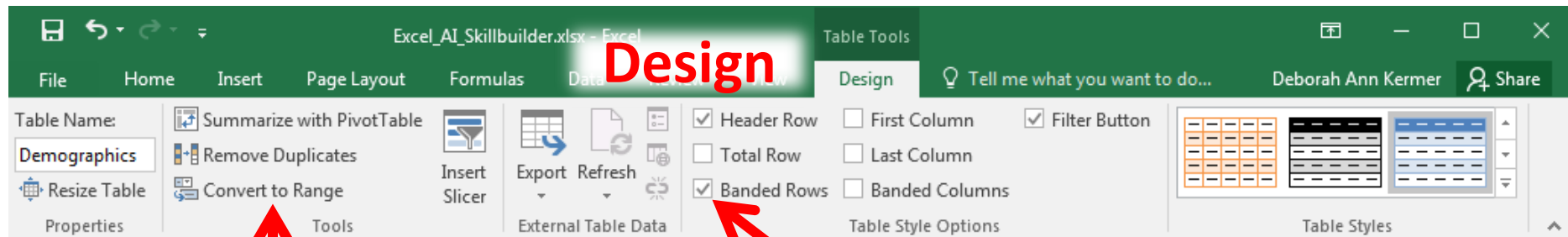




# Why Create an Excel Table Object?

Automatically...

- adds new rows and columns
- fills formulas to the whole column
- adds the table to the Data Model

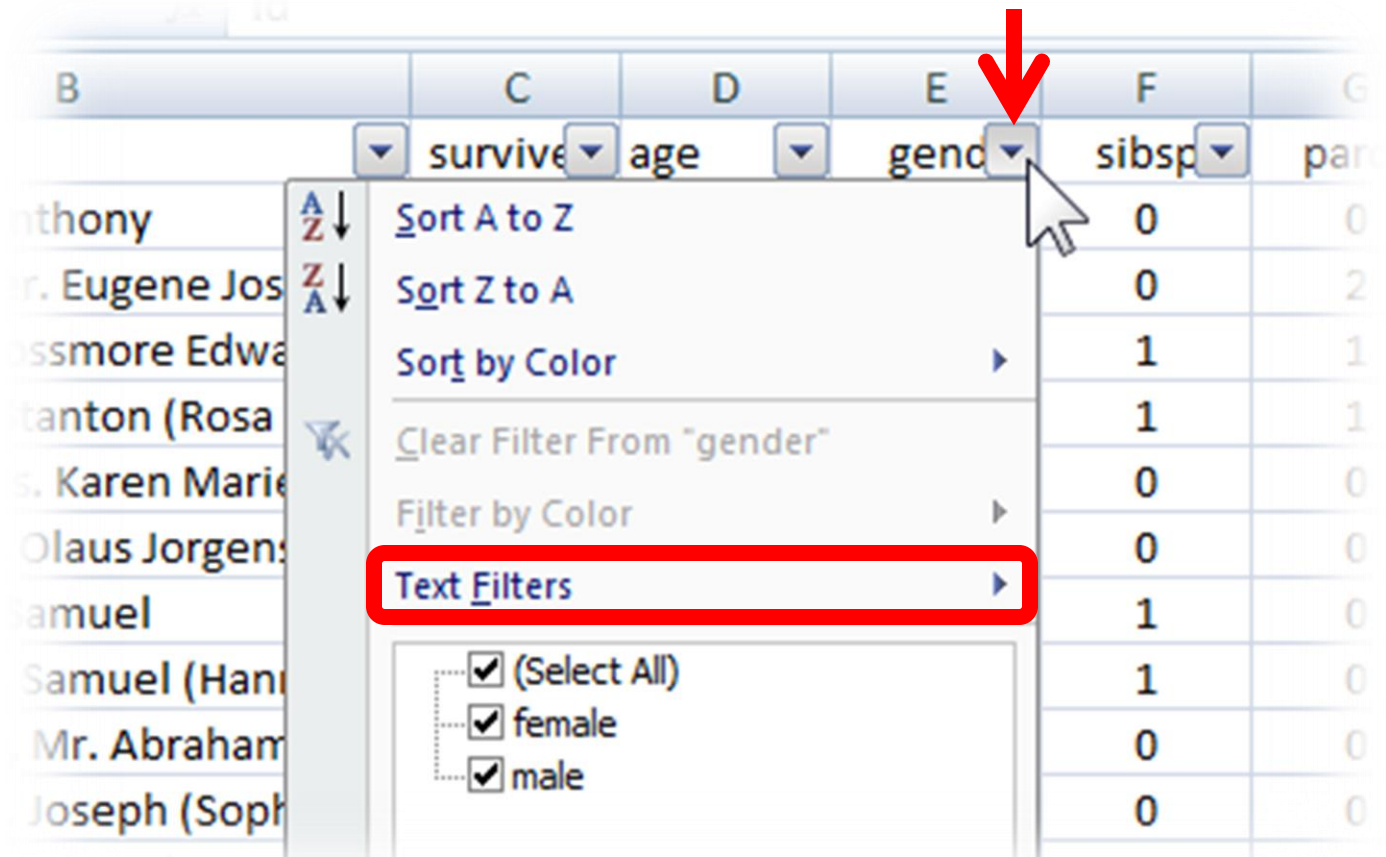


If you have trouble with other features, **Convert [back] to Range**

If you don't like the every-other-row highlighting, uncheck **Banded Rows**

# Drop-Downs

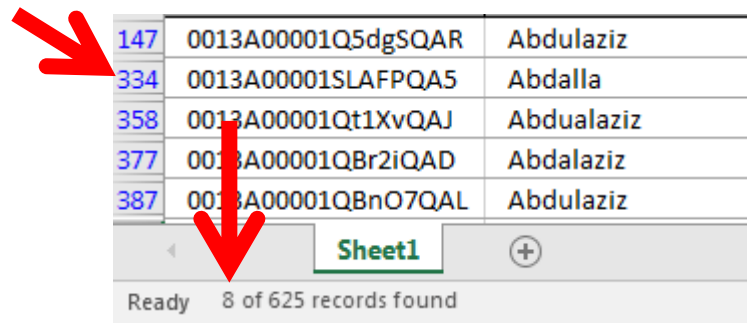
ONLY Sort or Filter with this  
Check out the Filters in Date Columns!



# Filters

When a filter is on, the numbers on the left are **blue**.

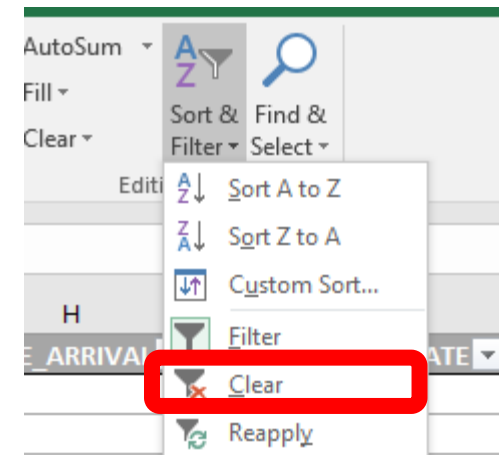
The status bar shows **how many** records were found.



147	0013A00001Q5dgSQAR	Abdulaziz
334	0013A00001SLAFPQA5	Abdalla
358	0013A00001Qt1XvQAJ	Abdualaziz
377	0013A00001QBr2iQAD	Abdalaziz
387	0013A00001QBnO7QAL	Abdulaziz

Ready 8 of 625 records found

If you forget which filter is on, you can **clear** all filters.



# Pivot Tables

Summarizing Information  
*Across Rows*

# Why Pivot Tables

- You have **columns** that define **groups**
  - e.g., multiple rows have the same value
- You wish to summarize ***within*** each group
  - Count
  - Percentage
  - Sum
  - Average

# Measurement Types

## **Categorical Data**

- Frequency (N)
- Percent (%)
- Mode
- Bar Chart

## **Numeric Data**

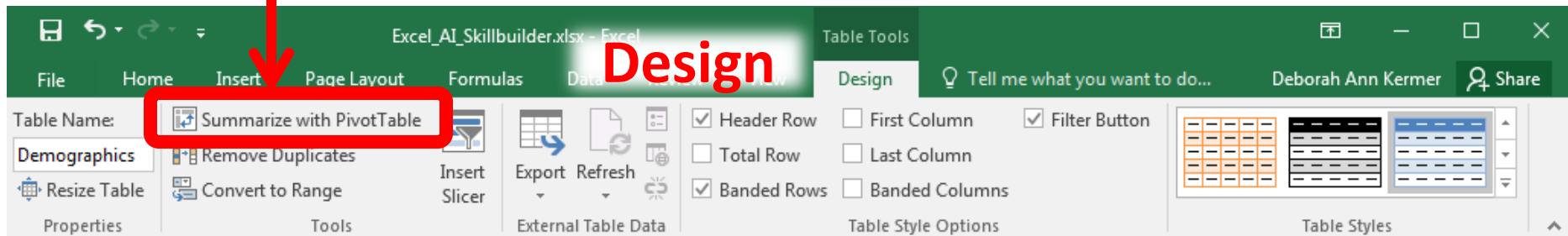
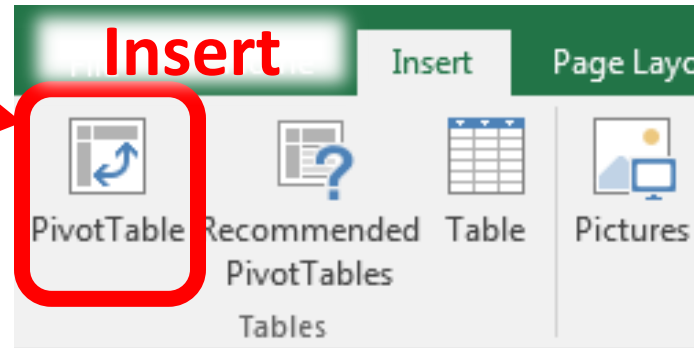
- Center
- Standard Deviation (sd)
- Mean or Median
- Histogram

# Create a Pivot Table

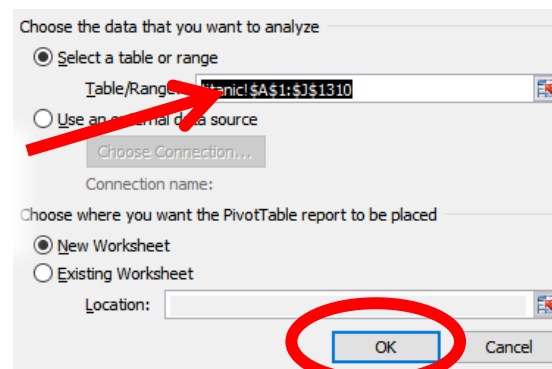
Insert **PivotTable**

OR

**Summarize with PivotTable**



Check Data Source  
then click **OK**



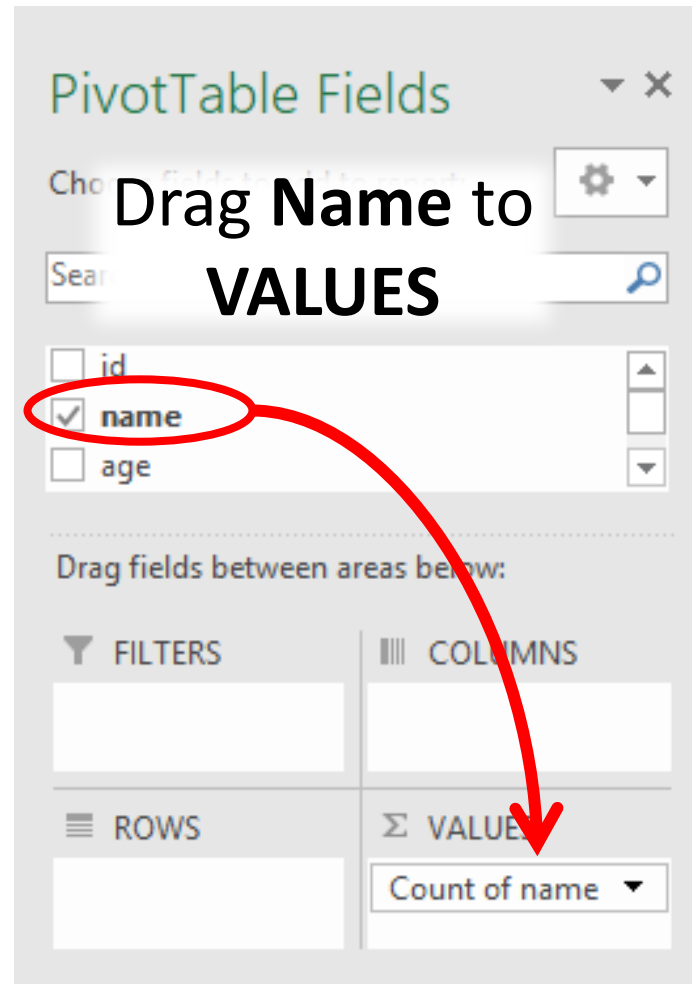
# Frequency Table



# Step 1: Values

Choose a **non-numeric** value that has **no missing values** like a Name or ID.

Drag it to the **Values** box.





# Always Check


Check your n (sample size)




The image shows an Excel spreadsheet with a PivotTable and the PivotTable Fields task pane. The PivotTable is located in cells A3:B4 and shows the count of names. The value 1309 is circled in red. The PivotTable Fields task pane is on the right, showing the fields 'id', 'name', and 'age'. The 'name' field is checked and added to the VALUES area. The task pane also shows the FILTERS, COLUMNS, and ROWS areas.

	A	B
1		
2		
3	Count of name	
4	1309	
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		

**PivotTable Fields**


Choose fields to add to report:  

Search 

☐ id   
☒ **name**   
☐ age 

Drag fields between areas below:

FILTERS	COLUMNS

ROWS	VALUES
	Count of name 

# Step 2: Rows

Drag a **Grouping Variable** to **Rows**

1. Use the filter box to find variables of interest

ex. Type “**class**”

2. Drag **Class** to the Rows area,  
or click the checkbox

	A	B
1		
2		
3	Row Labels	Count of name
4	1st Class	323
5	2nd Class	277
6	3rd Class	709
7	Grand Total	1309
8		
9		

The screenshot shows the 'PivotTable Fields' task pane. In the 'Choose fields to add to report:' section, the search box contains 'clas'. Below it, the list of fields includes 'pclass' and 'class'. The 'class' field is selected, indicated by a red circle and a red arrow pointing to the 'ROWS' area. The 'ROWS' area currently contains 'class'. The 'VALUES' area contains 'Count of name'. At the bottom, there is a 'Defer Layout Update' checkbox and an 'UPDATE' button.

# Reminder

These slides are available online at :

<https://infoguides.gmu.edu/spreadsheets/summarize>

# More Statistics

Clear the filter and drag Name to the Values box again

	A	B	C	D
1				
2				
3	Row Labels	Count of name	Count of name2	
4	1st Class	323	323	
5	2nd Class	277	277	
6	3rd Class	709	709	
7	Grand Total	1309	1309	
8				
9				

The screenshot shows the 'PivotTable Fields' task pane. A red circle highlights the search icon in the top right. Another red circle highlights the 'name' field in the list of available fields. A red arrow points from the 'name' field to the 'Count of name' entry in the 'VALUES' box.

PivotTable Fields

Choose fields to add to report: [Settings icon]

Search [Search icon]

☐ id

☒ name

☐ age

Drag fields between areas below:

**FILTERS**

**COLUMNS**

$\Sigma$  Values

**ROWS**

class

**VALUES**

Count of name

Count of name2

☐ Defer Layout Update [UPDATE]

# Percentages

Right click the last column and choose  
**Show Values As ... % of Column Total**

The screenshot shows an Excel PivotTable with the following data:

Row Labels	Count of name	Count of name2
1st Class	323	2
2nd Class	277	2
3rd Class	709	5
<b>Grand Total</b>	<b>1309</b>	<b>10</b>

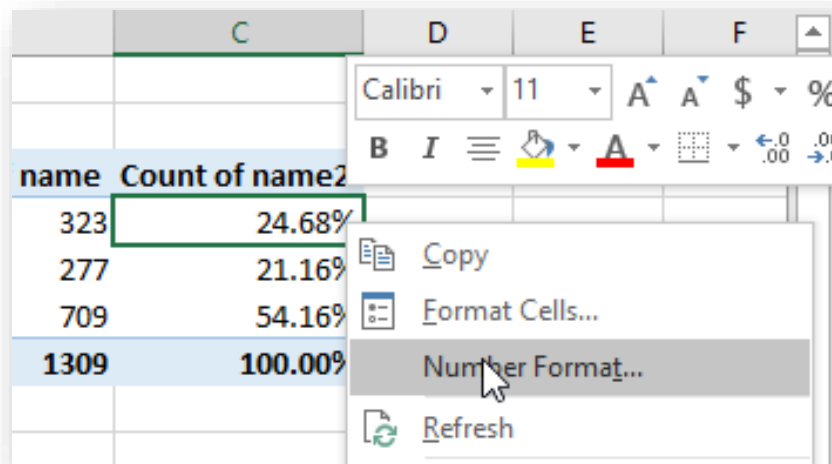
A right-click context menu is open over the 'Count of name2' column. The 'Show Values As' option is selected, which has opened a secondary menu. In this secondary menu, the option '% of Column Total' is highlighted with a mouse cursor.

The secondary menu options are:

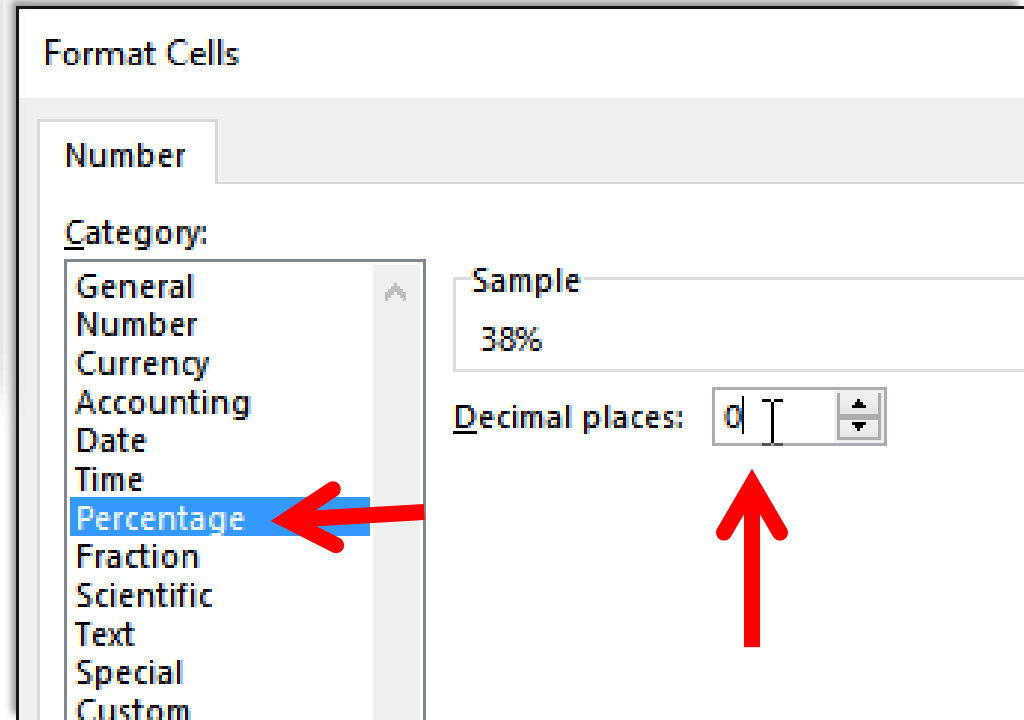
- No Calculation
- % of Grand Total
- % of Column Total** (selected)
- % of Row Total
- % Of...
- % of Parent Row Total
- % of Parent Column Total

# Number Format

Right Click the Percent Column and choose **Number Format**  
Change it to **Percentage** with **0** Decimal places and click **OK**



name	Count of name2
323	24.68%
277	21.16%
709	54.16%
1309	100.00%



# Labels

Click in a cell to change the label

**n** = number of people

**%** = percent

	A	B	C
1			
2			
3	Row Labels ▼	n	%
4	1st Class	323	25%
5	2nd Class	277	21%
6	3rd Class	709	54%
7	Grand Total	1309	100%
8			
9			

Drag fields between areas below:

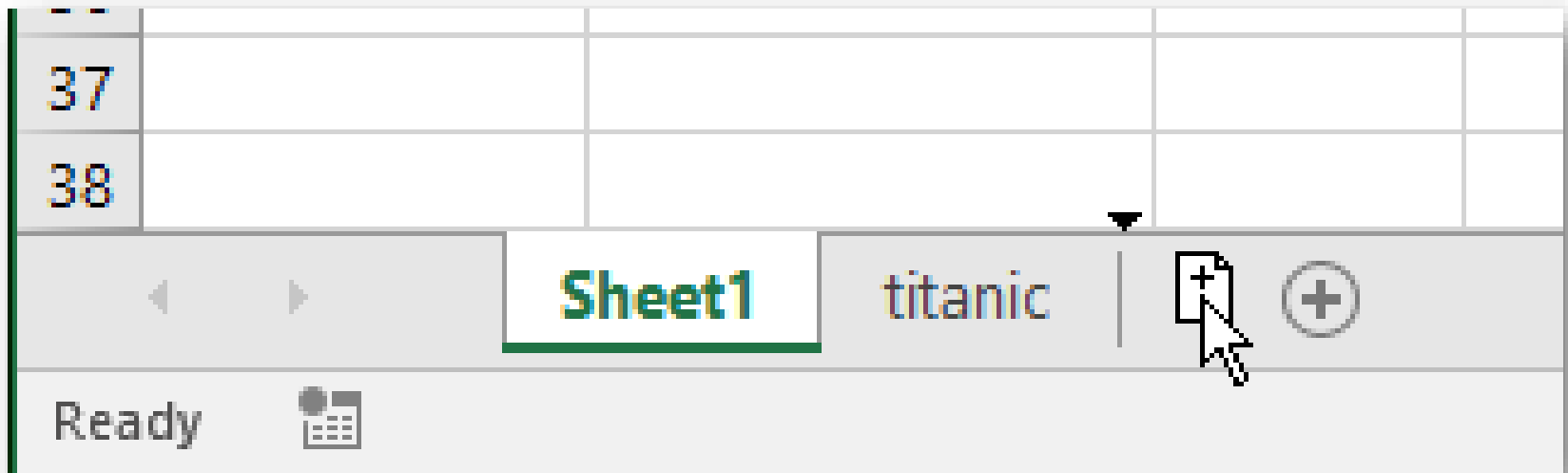
<b>▼ FILTERS</b>	<b>    COLUMNS</b>
	<b>Σ Values</b> ▼
<b>≡ ROWS</b>	<b>Σ VALUES</b>
class ▼	n ▼
	% ▼
<input type="checkbox"/> Defer Layout Update	UPDATE

You can center the labels  
and reduce the column size.



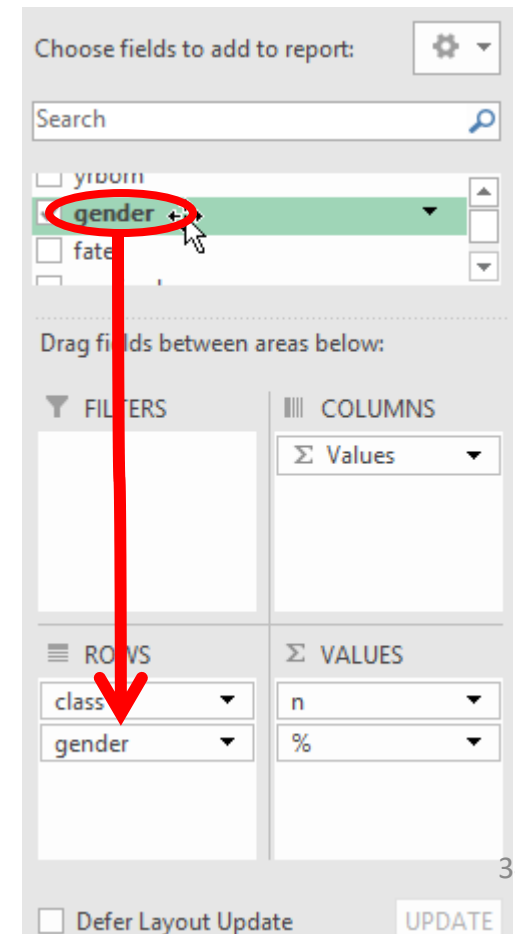
# Copying Worksheets

1. **Press** and **Hold** the Control (Win) or Option (Mac) key
2. **Click** and **Drag** the Sheet1 tab to the right and release



Drag **gender**  
to **ROWS**

Nesting & Crosstabulation



# Multiple Variables in Rows

pclass  
gender

FILTERS		COLUMNS	
		$\Sigma$ Values	
ROWS		$\Sigma$ VALUES	
class		n	
gender		%	

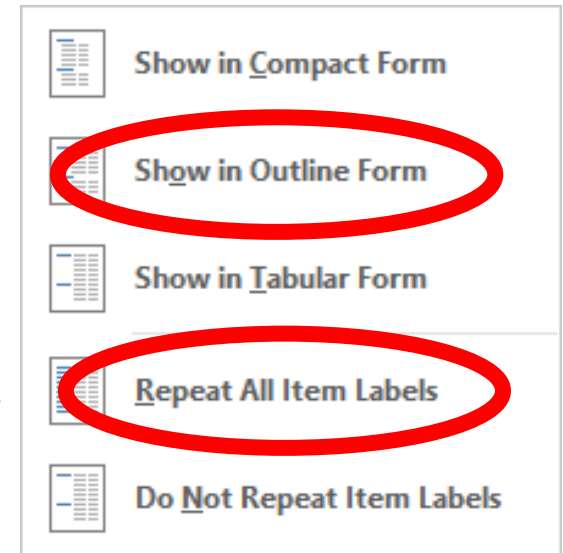
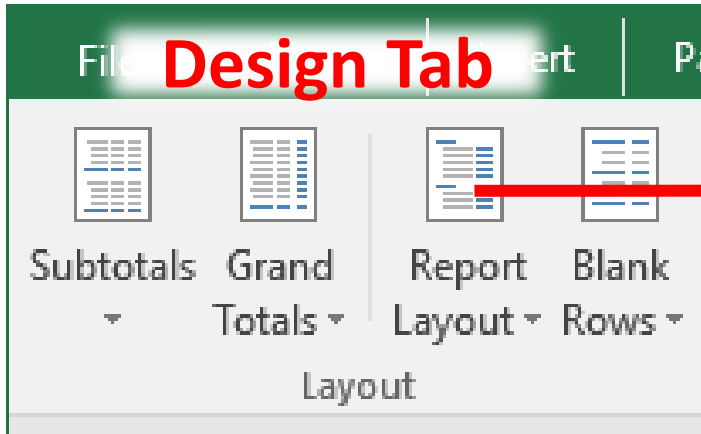
gender  
pclass

FILTERS		COLUMNS	
		$\Sigma$ Values	
ROWS		$\Sigma$ VALUES	
gender		n	
class		%	

Row Labels	n	%
1st Class	323	25%
female	144	11%
male	179	14%
2nd Class	277	21%
female	106	8%
male	171	13%
3rd Class	709	54%
female	216	17%
male	493	38%
Grand Total	1309	100%

Row Labels	n	%
female	466	36%
1st Class	144	11%
2nd Class	106	8%
3rd Class	216	17%
male	843	64%
1st Class	179	14%
2nd Class	171	13%
3rd Class	493	38%
Grand Total	1309	100%

# Report Layouts



## Compact

Row Labels	n	%
1st Class	323	25%
female	144	11%
male	179	14%
2nd Class	277	21%
female	106	8%
male	171	13%
3rd Class	709	54%
female	216	17%
male	493	38%
Grand Total	1309	100%

## Outline

class	gender	n	%
1st Class		323	25%
	female	144	11%
	male	179	14%
2nd Class		277	21%
	female	106	8%
	male	171	13%
3rd Class		709	54%
	female	216	17%
	male	493	38%
Grand Total		1309	100%

## Tabular

class	gender	n	%
1st Class	female	144	11%
	male	179	14%
1st Class Total		323	25%
2nd Class	female	106	8%
	male	171	13%
2nd Class Total		277	21%
3rd Class	female	216	17%
	male	493	38%
3rd Class Total		709	54%
Grand Total		1309	100%

# Totals

	A	B	C	D
1				
2				
3	class	gender	n	%
4	1st Class		323	25%
5	1st Class	female	144	11%
6	1st Class	male	179	14%
7	2nd Class		277	21%
8	2nd Class	female	106	8%
9	2nd Class	male	171	13%
10	3rd Class		709	54%
11	3rd Class	female	216	17%
12	3rd Class	male	493	38%
13	Grand Total		1309	100%
14				
15				

	A	B	C	D
1				
2				
3	class	gender	n	%
4	1st Class			
5	1st Class	female	144	11%
6	1st Class	male	179	14%
7	2nd Class			
8	2nd Class	female	106	8%
9	2nd Class	male	171	13%
10	3rd Class			
11	3rd Class	female	216	17%
12	3rd Class	male	493	38%
13				

## Grand Totals



Off for Rows and Columns



On for Rows and Columns



On for Rows Only



On for Columns Only

## Subtotals



Do Not Show Subtotals



Show all Subtotals at Bottom of Group



Show all Subtotals at Top of Group


# Nesting to Crosstabulation

Drag gender to COLUMNS

	A	B	C	D	E
1					
2					
3	class	gender	n	%	
4	1st Class		323	25%	
5	1st Class	female	144	11%	
6	1st Class	male	179	14%	
7	2nd Class		277	21%	
8	2nd Class	female	106	8%	
9	2nd Class	male	171	13%	
10	3rd Class		709	54%	

	A	B	C	D	E
11	3r				
12	3r 1				
13	Gra 2				
14	3	gender	Values		
15	4	female	female	male	male
	5	class	n	%	n
	6	1st Class	144	31%	179
	7	2nd Class	106	23%	171
	8	3rd Class	216	46%	493
	9	Grand Total	466	100%	843
	10				
	11				
	12				

	A	B	C	D	E
1					
2					
3	Values		gender		
4	n		%		
5	class	female	male	female	male
6	1st Class	144	179	31%	21%
7	2nd Class	106	171	23%	20%
8	3rd Class	216	493	46%	58%
		466	843	100%	100%

Choose fields to add to report: 

Search:

ge  
rborn  
gender  
ate  
urvived  
ibsp

Fields between areas below:

**FILTERS**

**COLUMNS**

Σ Values

**ROWS**

class  
gender

Σ VALUES

n  
%

☐ Defer Layout Update UPDATE

# Reading Percentages

Percentages equalize the group sizes. Since 100%'s are equal, you can compare parallel percents to see if they are equal, too. If they are not equal, there is a relationship between the variables.

3	gender ▾ Values						
4		female	female	male	male	Total n	Total %
5	pclass ▾	n	%	n	%		
6	1st Class	144	45%	179	55%	323	100%
7	2nd Class	106	38%	171	62%	277	100%
8	3rd Class	216	30%	493	70%	709	100%
9	Grand Total	466	36%	843	64%	1309	100%

Pivot Table

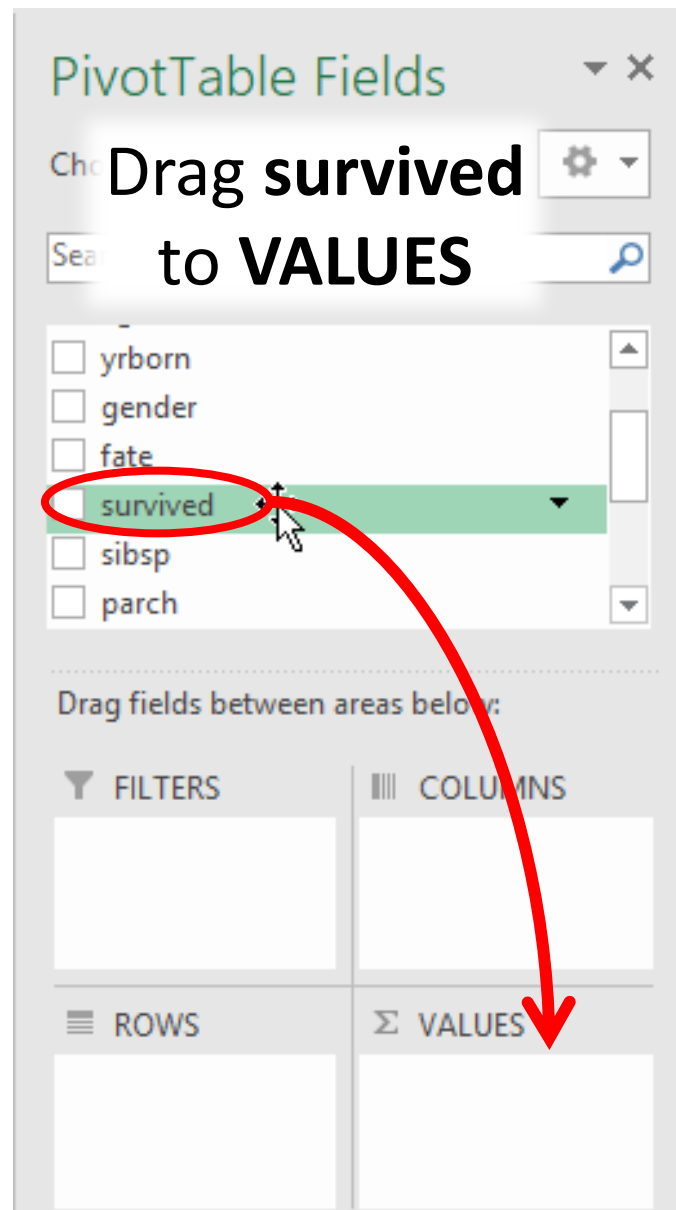


# Create another Pivot Table

Go back to your data tab

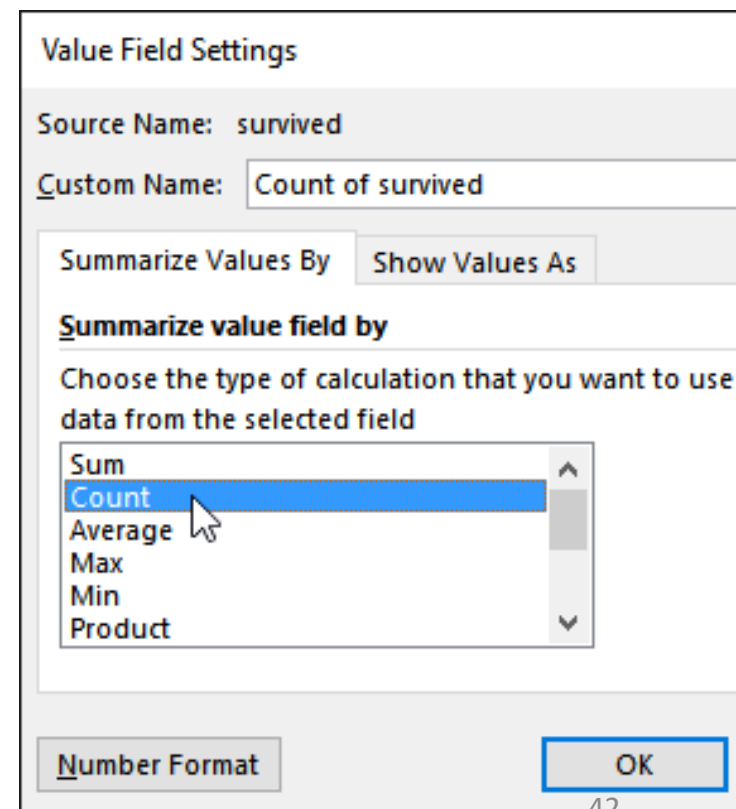
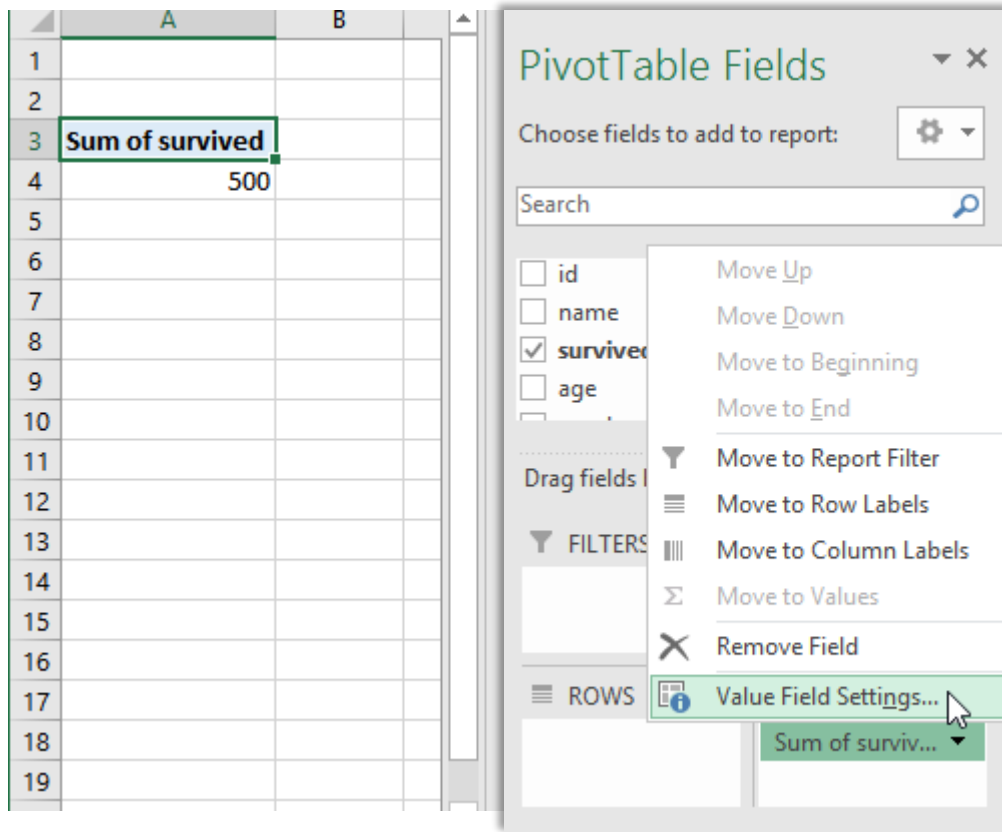
Click inside your data

Click Pivot Table



# Value Field Settings

Click survived in the Values box choose **Value Field Settings**,  
or Right Click and Summarize Values By **Count**



# Multiple Statistics

	A	B	C
1			
2			
3	Count of survived	Sum of survived	Average of survived2
4	1309	500	0.38197097
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			

2. For the 3<sup>rd</sup>,  
Summarize Values By  
the **Average**

Value Field Settings

Source Name: survived

Custom Name: Average of survived2

Summarize Values By: **Average**

Show Values As: [dropdown]

Summarize value field by

Choose the type of calculation that you want to use to summarize data from the selected field

Sum  
Count  
**Average**  
Max  
Min  
Product

Number Format

OK Cancel

1. Drag **survived**  
to **VALUES**  
two **more** times

PivotTable Fields

☐ name

☒ **survived**

Drag fields between areas below:

**FILTERS**

**COLUMNS**

**ROWS**

**VALUES**

Count of survi...  
Sum of surviv...  
Average of su... 43

# Labeling

3	Count of survived	Sum of survived	Average of survived2
4	1309	500	0.38197097
5			

How many cells

How many 1's

Proportion of 1's

3	# Passengers	# Survived	% Survived
4	1309	500	0.38197097
5			

*Just click in the cell and type to change the label*

# Change the % Number Format

The screenshot shows an Excel spreadsheet with columns labeled 'B', 'C', 'D', 'E', 'F', and 'G'. The first row of data has headers: 'ers', '# Survived', and '% Survive'. The second row contains the values '309', '500', and '38%'. A right-click context menu is open over the '38%' cell, with 'Number Format...' selected. The 'Format Cells' dialog box is open, showing the 'Number' category. The 'Category' list on the left includes General, Number, Currency, Accounting, Date, Time, Percentage (selected), Fraction, Scientific, Text, Special, and Custom. The 'Sample' field displays '38%'. The 'Decimal places' field is set to '0'. A descriptive text at the bottom of the dialog states: 'Percentage formats multiply the cell value by 100 and displays the result with a percent symbol.' The 'OK' and 'Cancel' buttons are at the bottom right.

ers	# Survived	% Survive
309	500	38%

**Format Cells**

**Number**

**Category:**

- General
- Number
- Currency
- Accounting
- Date
- Time
- Percentage**
- Fraction
- Scientific
- Text
- Special
- Custom

**Sample**

38%

**Decimal places:** 0

Percentage formats multiply the cell value by 100 and displays the result with a percent symbol.

OK Cancel

# FYI: Using the Boxes

Values	
# Passengers	1309
# Survived	500
% Survived	38%

# Passengers	# Survived	% Survived
1309	500	38%

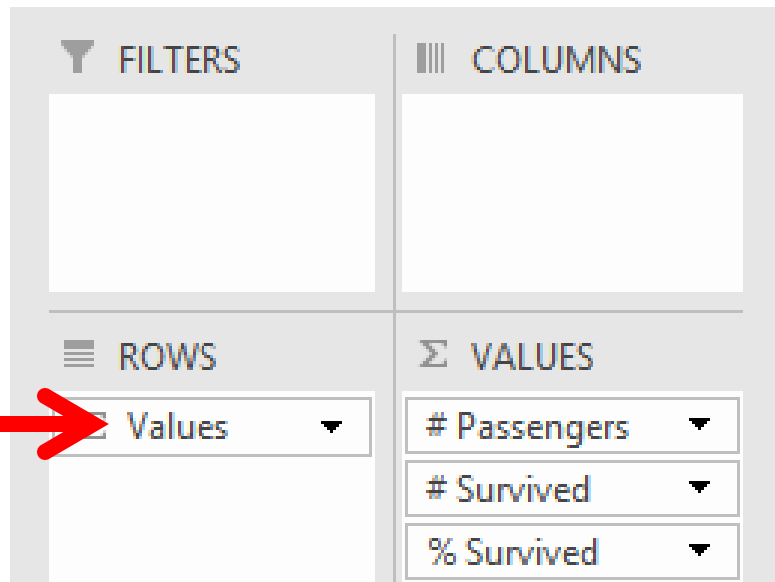


Tableau interface showing the 'VALUES' box in the ROWS shelf. A red arrow points to the 'VALUES' box.

**FILTERS**

**COLUMNS**

**ROWS**

**VALUES**

- # Passengers
- # Survived
- % Survived

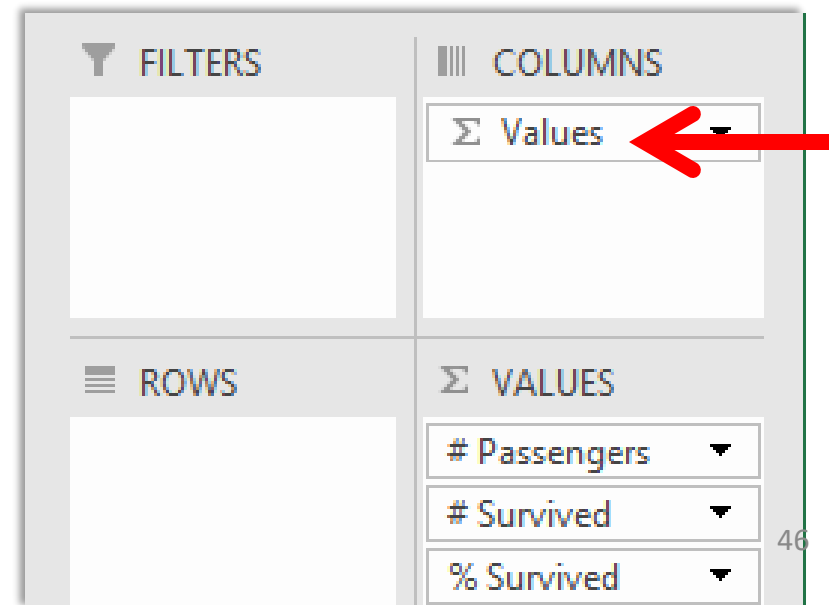


Tableau interface showing the 'VALUES' box in the COLUMNS shelf. A red arrow points to the 'VALUES' box.

**FILTERS**

**COLUMNS**

**ROWS**

**VALUES**

- # Passengers
- # Survived
- % Survived

# Rows

	A	B	C	D	E
1					
2					
3	Row Labels	# Passengers	# Survived	% Survived	
4	1st Class	323	200	62%	
5	2nd Class	277	119	43%	
6	3rd Class	709	181	26%	
7	Grand Total	1309	500	38%	
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					

Drag **class** to  
**ROWS** box

### PivotTable Fields

Choose fields to add to report:

Search

- ☐ pclass
- ☒ class
- ☐ fare
- ☐ embarked

MORE TABLES

Drag fields between areas below:




FILTERS	COLUMNS
	<div>Σ Values </div>

ROWS	Σ VALUES
class	# Passengers
	# Survived
	% Survived

☐ Defer Layout Update UPDATE

# Review

Drag fields between areas below:

 Report Filter	 Column Labels
Limit by Group Try Slicers Instead	Secondary Grouping Variables
 Row Labels	$\Sigma$ Values
Primary Grouping Variables	<b>Summary Variables</b>

☐ **Defer Layout Update** Update

## Numbers - Pivot Table

- Measurement
  - Amount
  - Count
  - Binary (0/1)
- *Numbers* default to **Sum**

## IDs - Crosstabulation

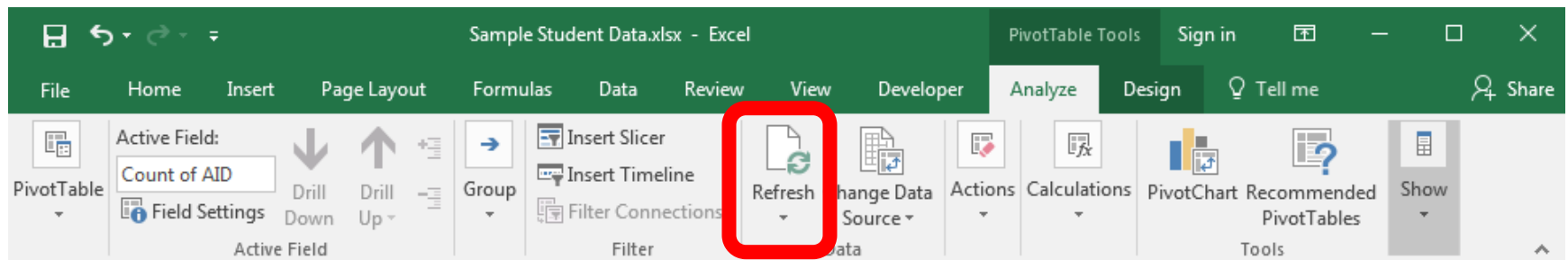
- Text (or change to count)
  - No missing values
- *Text* defaults to **Count**



# Adding New Rows or Columns

For Excel Table Objects, new rows and columns **next** to existing data are **automatically included** in the table.

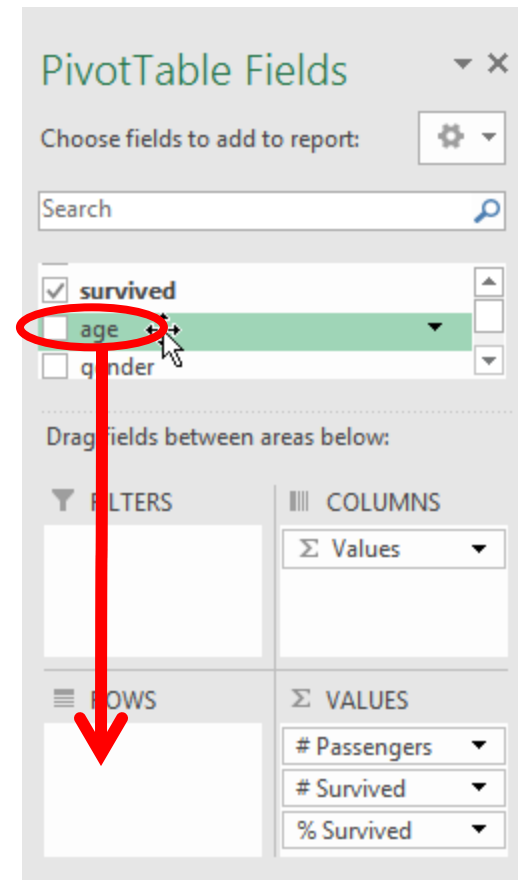
To update your pivot table after making changes, click Refresh



# Grouping Numeric Values

Advanced

Drag **age** to **ROWS**



# Automatic Grouping of Values

Right Click on a Value and choose "Group..."

Specify the groupings and click OK

The screenshot illustrates the process of automatically grouping values in a PivotTable. A right-click context menu is open over the 'age' field, with the 'Group...' option selected. The 'Grouping' dialog box is displayed, showing the 'Auto' section with the following settings:

- ☐ Starting at: 0
- ☐ Ending at: 98
- By: 10

The 'OK' button is highlighted, indicating the next step is to confirm the grouping. The background shows the 'PivotTable Fields' task pane and a list of age ranges from 0-10 to >100.

age
0-10
10-20
20-30
30-40
40-50
50-60
60-70
70-80
80-90
>100
<b>Grand Total</b>

# FYI: Manual Grouping

1. Select the values to group
2. Right Click, choose "Group" like before
3. Type the Group Name/Label

A new variable  
appears (age2)

The figure consists of three side-by-side screenshots of a data table, illustrating the effect of different grouping methods on the 'age' variable.

**Left Screenshot:** Shows the original data with 'age' values ranging from 0 to 5. The 'age' column is highlighted with a red box.

	age	Pass
3		
4	0.17	
5	0.33	
6	0.42	
7	0.67	
8	0.75	
9	0.83	
10	0.92	
11	1	
12	2	
13	3	
14	4	
15	5	
16	6	
17	7	

**Middle Screenshot:** Shows the data grouped by 'age' into 'Group1'. The 'Group1' column is highlighted with a red box.

	age2	age
3		
4	Group1	
5	Group1	
6	Group1	
7	Group1	
8	Group1	
9	Group1	
10	Group1	
11	Group1	
12	Group1	
13	Group1	
14	Group1	
15	Group1	
16	5	
17	5	

**Right Screenshot:** Shows the data grouped by 'age2' into 'Age 0-4'. The 'Age 0-4' column is highlighted with a red box.

	age2	age
3		
4	Age 0-4	
5	Age 0-4	0.33
6	Age 0-4	0.42
7	Age 0-4	0.67
8	Age 0-4	0.75
9	Age 0-4	0.83
10	Age 0-4	0.92
11	Age 0-4	1
12	Age 0-4	2
13	Age 0-4	3
14	Age 0-4	4
15	Age 0-4	5
16	5	
17	5	

# Manual Grouping Process

After the 1<sup>st</sup> Group, you will have **two** columns.

**Ignore** the groupings and select values in either column to group.

	A	B	C
16	5		
17	5	5	
18	6		
19	6	6	
20	7		
21	7	7	
22	8		
23	8	8	
24	9		
25	9	9	
26	10		
27	10	10	
28	11		
29	11	11	
30	11.5		
31	11.5	11.5	
32	12		
33	12	12	
34	13		
35	13	13	
36	14		

OR

	A	B	C
16	5		
17	5	5	
18	6		
19	6	6	
20	7		
21	7	7	
22	8		
23	8	8	
24	9		
25	9	9	
26	10		
27	10	10	
28	11		
29	11	11	
30	11.5		
31	11.5	11.5	
32	12		
33	12	12	
34	13		
35	13	13	
36	14		

# Keep the Groups (Only)

Drag out (remove) the original age variable

	A	B	C	D	E	F	G
1							
2							
3	age2	age	# Passengers	# Survived	% Survived		
4	Age 0-4		51	33	65%		
5	Age 5-12		43	21	49%		
6	Age 13-17						
7	Age 13-1	13	5	3	60%		
8	Age 13-1	14	8	4	50%		
9	Age 13-1	14.5	2	0	0%		
10	Age 13-1	15	6	5	83%		
11	Age 13-1	16	19	8	42%		
12	Age 13-1	17	20	7	35%		
13	Age 18-24						
14	Age 18-2	18	39	14	36%		
15	Age 18-2	18.5	3	0	0%		
16	Age 18-2	19	29	11	38%		
17	Age 18-2	20	23	8	35%		
18	Age 18-2	20.5	1	0	0%		
19	Age 18-2	21	41	11	27%		
20	Age 18-2	22	43	20	47%		
21	Age 18-2	22.5	1	0	0%		

### PivotTable Fields

Choose fields to add to report:

Search

☒ survived

☒ age

☐ gender

Drag fields between areas below:

FILTERS

ROWS

age2

age

COLUMNS

Values

VALUES

# Passengers

# Survived

% Survived

# Manual Grouping: Final Product

	A	B	C	D	E	F	G
1							
2							
3	age2	# Passengers	# Survived	% Survived			
4	Age 0-4	51	33	65%			
5	Age 5-12	43	21	49%			
6	Age 13-17	60	27	45%			
7	Age 18-24	255	96	38%			
8	Age 25-34	292	110	38%			
9	Age 35-44	169	64	38%			
10	Age 45-54	109	53	49%			
11	Age 55+	67	23	34%			
12	No Age	263	73	28%			
13							
14							
15							
16							
17							
18							
19							
20							
21							

## PivotTable Fields

Choose fields to add to report:

Search

☐ id

☒ name

☒ survived

Drag fields between areas below:

**FILTERS**

**COLUMNS**

Σ Values

**ROWS**

age2

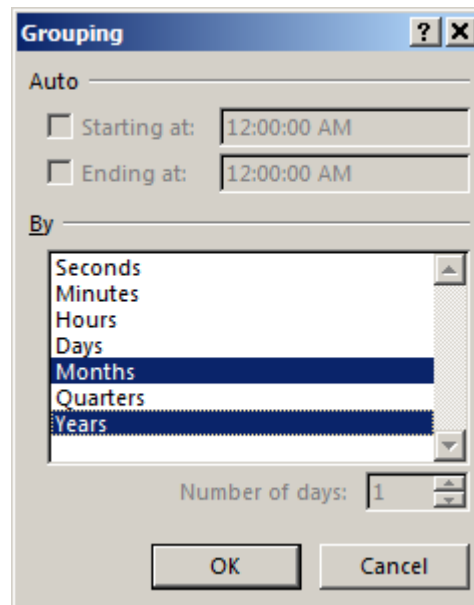
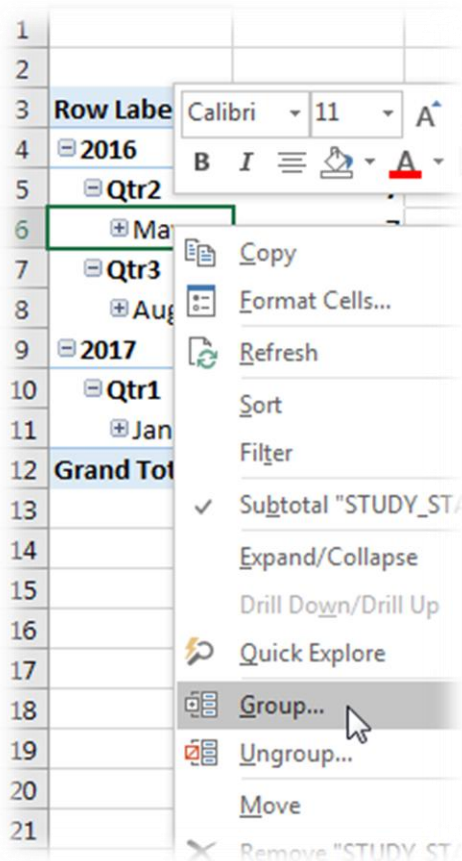
**VALUES**

# Passengers

# Survived

% Survived

# FYI: Grouping of Dates



2		
3	Row Labels	Count of AID
4	2016	312
5	May	7
6	Aug	305
7	2017	313
8	Jan	313
9	Grand Total	625



Excel has its place

# Good Uses of Excel

- When someone gives you an Excel file
- Data Entry
- For general summaries of tabular data
- If you only need to create a simple graph once

# Excel for Data Management

Do not make changes to data "by hand" or using functions.

- **Power Query** in Excel
- **Power BI**
  - Power Query through a separate data-focused software
- **OpenRefine**
  - Free, Open Source, Cross-Platform
  - Can do anything except Merge tables
  - Good for examining text in tables and frequencies

# Excel for Statistics

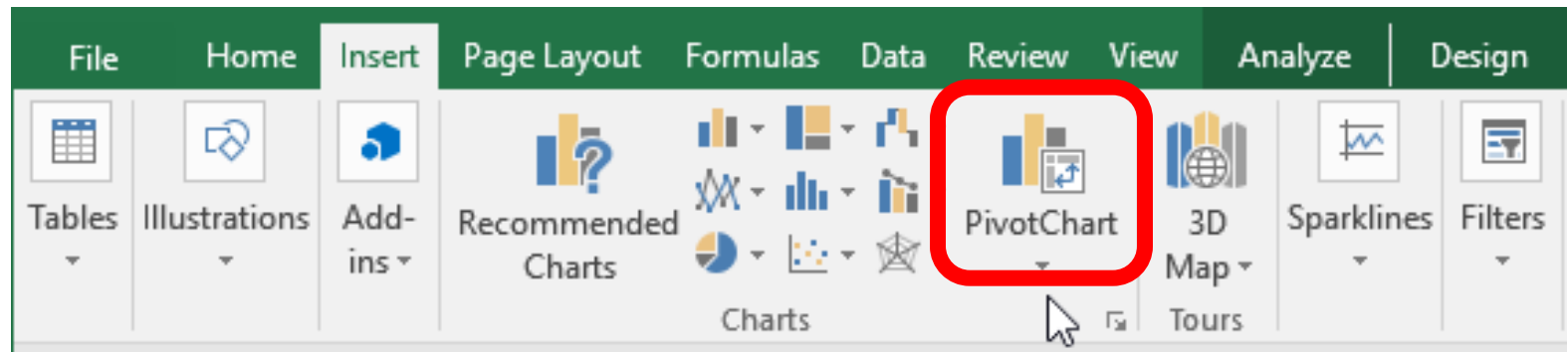
- Do NOT use Excel for statistical testing

Free, Open Source, Cross-Platform Alternatives

- **Jamovi** – Rapidly developing companion to R
- **PSPP** – SPSS look-alike with all the basics
- **JASP** – Long-standing favorite with many analyses
- **R** – Statistical Language

# Keep in Mind

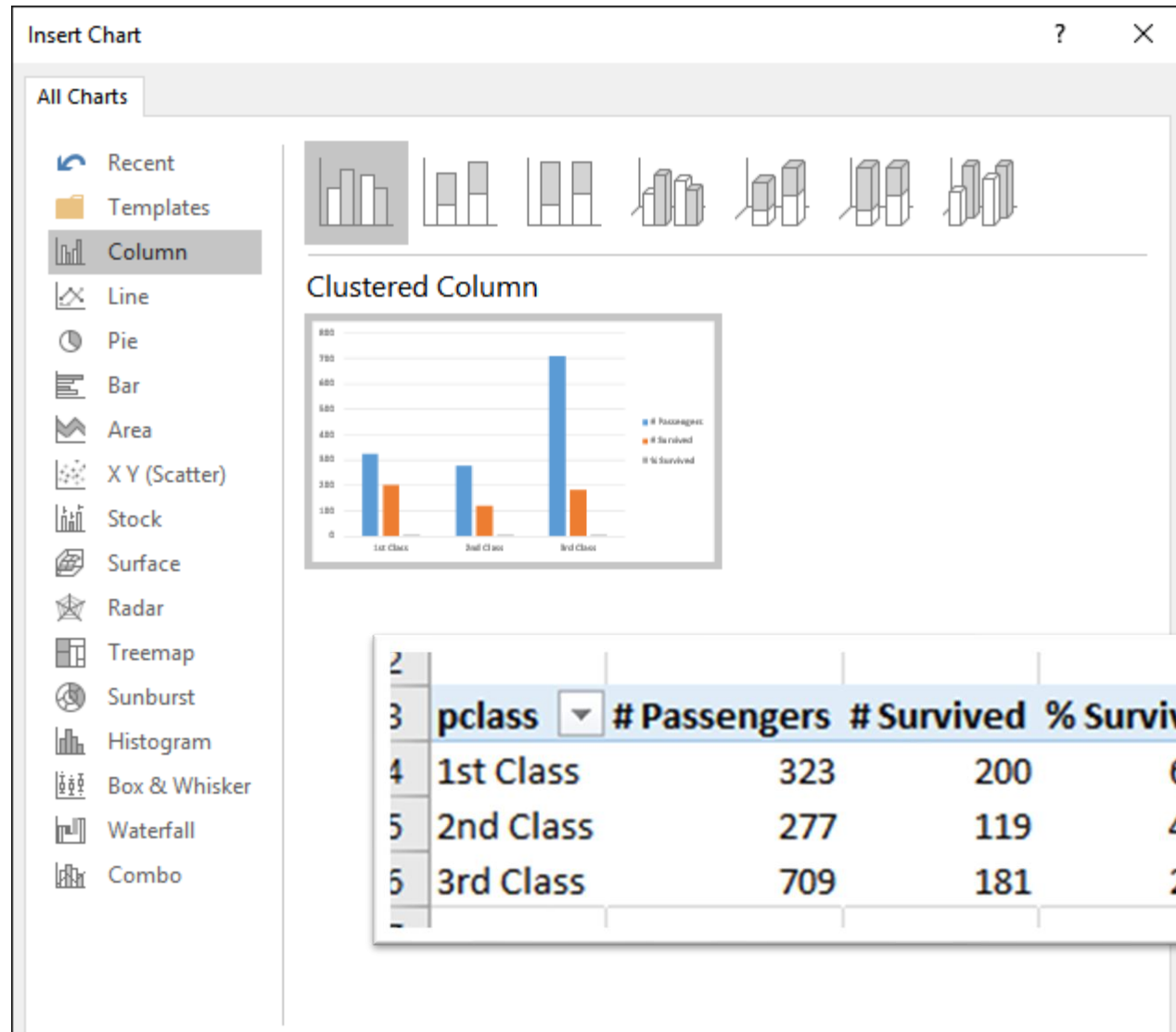
- All statistical software will **open files** from other software
  - All software will open .csv , .tsv, and .txt files
- Files for statistical software will have **metadata**
  - Metadata is useful
  - Qualtrics offers data in SPSS format
- **Excel messes with data**
  - If you open a file, save a separate copy just for Excel.



# Pivot Charts

Use AFTER creating a Pivot Table

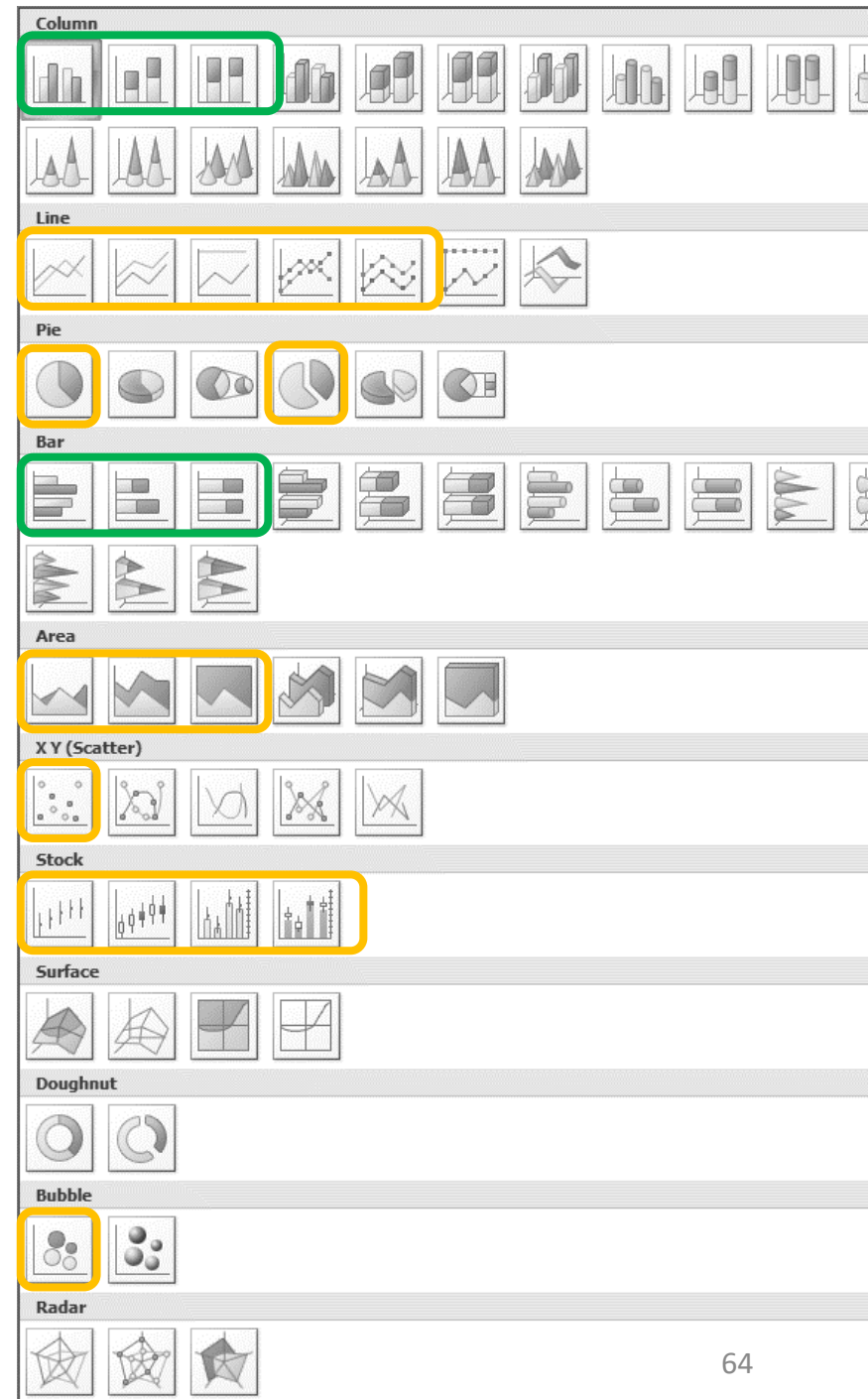
# Insert Chart



# Chart Types

Most Excel chart styles are **NOT** best practices

- Comparison
- Distribution
- Composition
- Trend
- Relationship





# Find a Pattern, ***THEN*** Make a Chart

What is the most **interesting** story from this data?

2				
3	pclass ▼	# Passengers	# Survived	% Survived
4	1st Class	323	200	62%
5	2nd Class	277	119	43%
6	3rd Class	709	181	26%
7				

*For me:* the difference in survival rate between classes.

Design a chart to make that pattern pop out

# Bar Chart

pclass	# Passengers	# Survived	% Survived
1st Class	323	200	62%
2nd Class	277	119	43%
3rd Class	709	181	26%

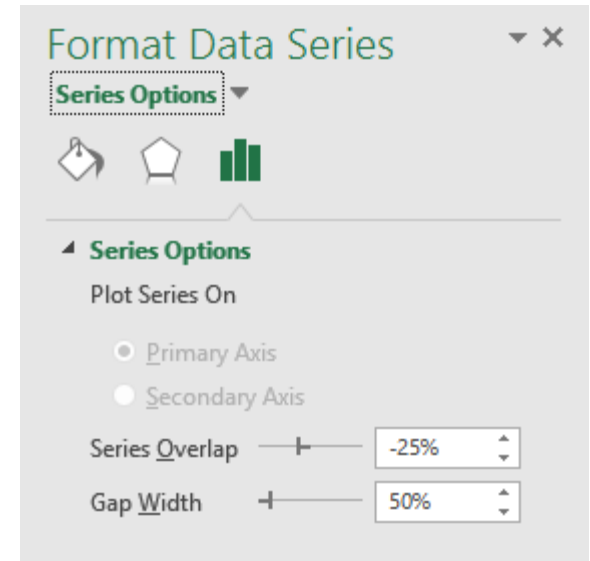
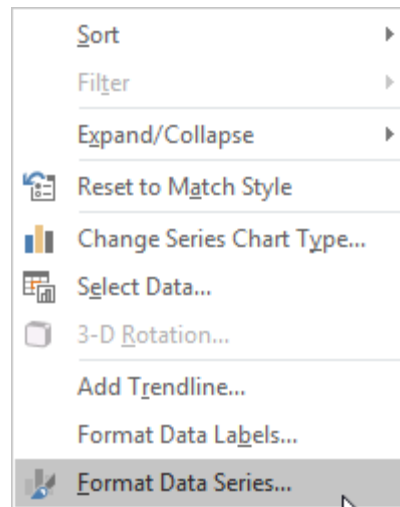
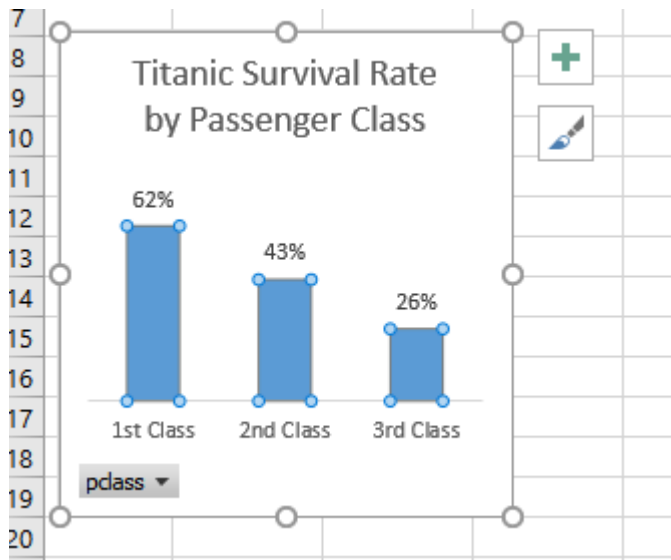
Does this make the identified pattern **obvious**?



Where is  
% Survived?  
Proportions  
are < 0

# Modifications

Right-click the bars and choose "Format Data Series..."

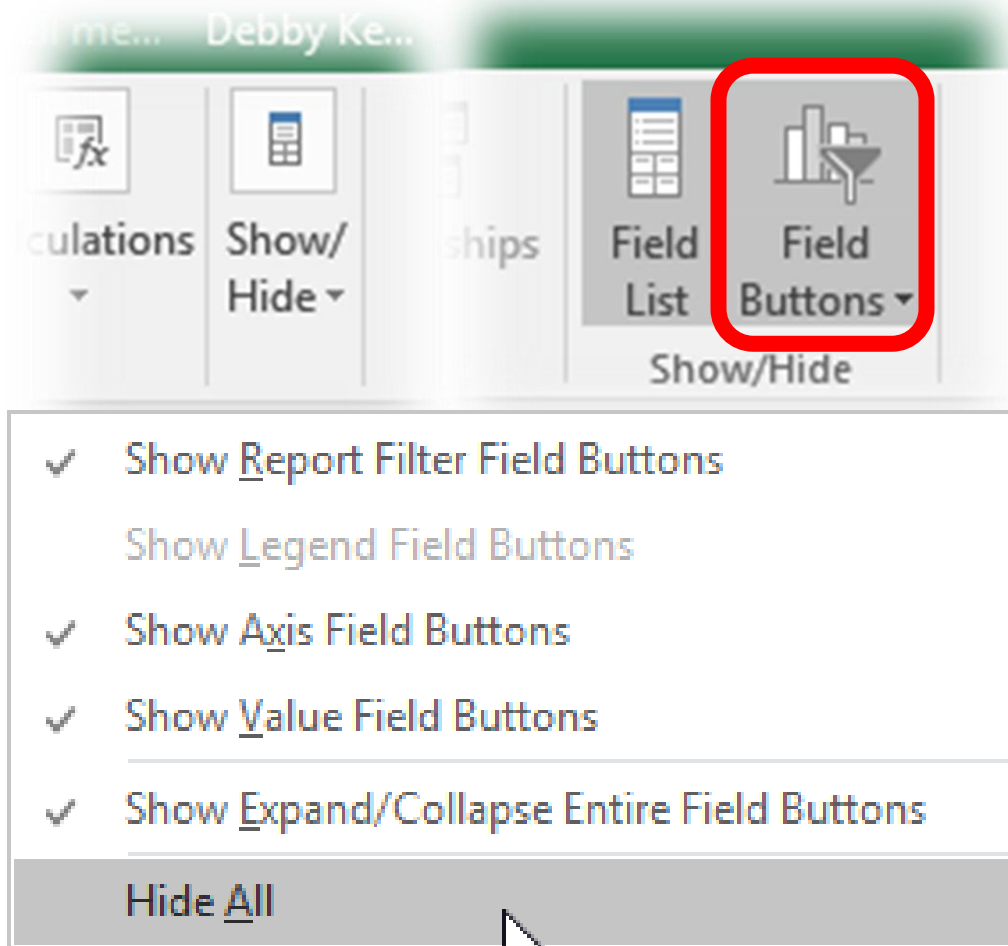


**Remove:** "# Passengers", "# Survived", Gridlines, Y-axis

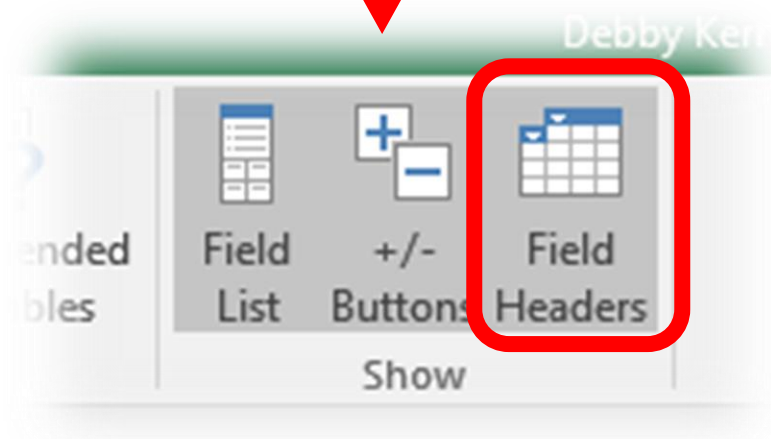
**Add:** Chart Title, Data Labels

**Change:** Gap Width, Font Size

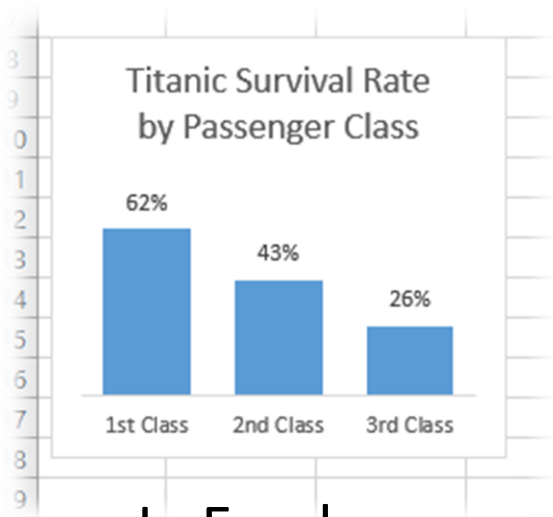
# Removing Extra Buttons



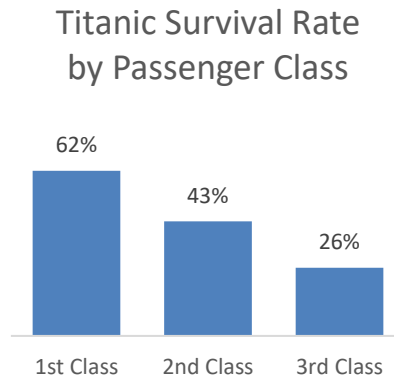
Charts  
Tables



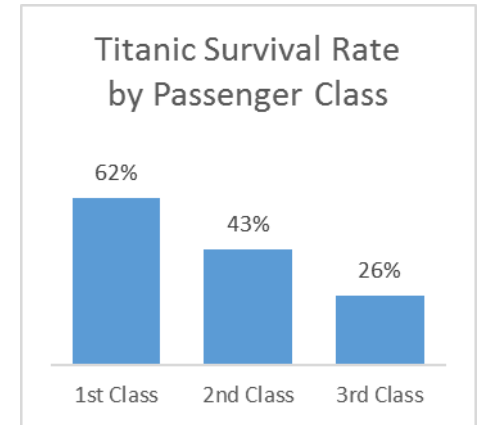
# Copy-Paste to Documents



In Excel

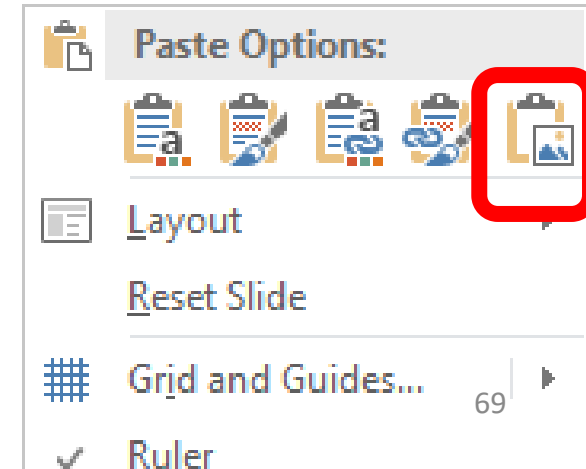
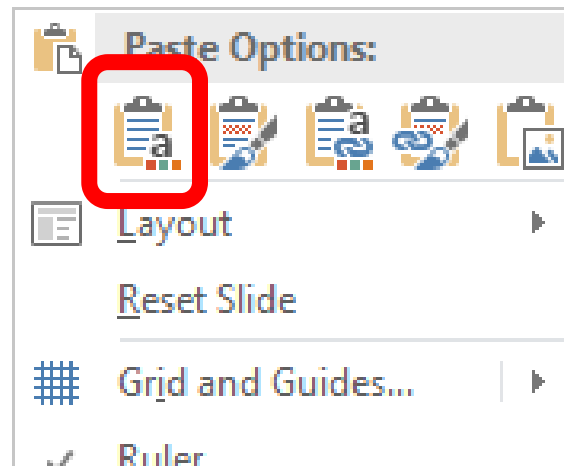


Use Destination Theme and Embed Workbook



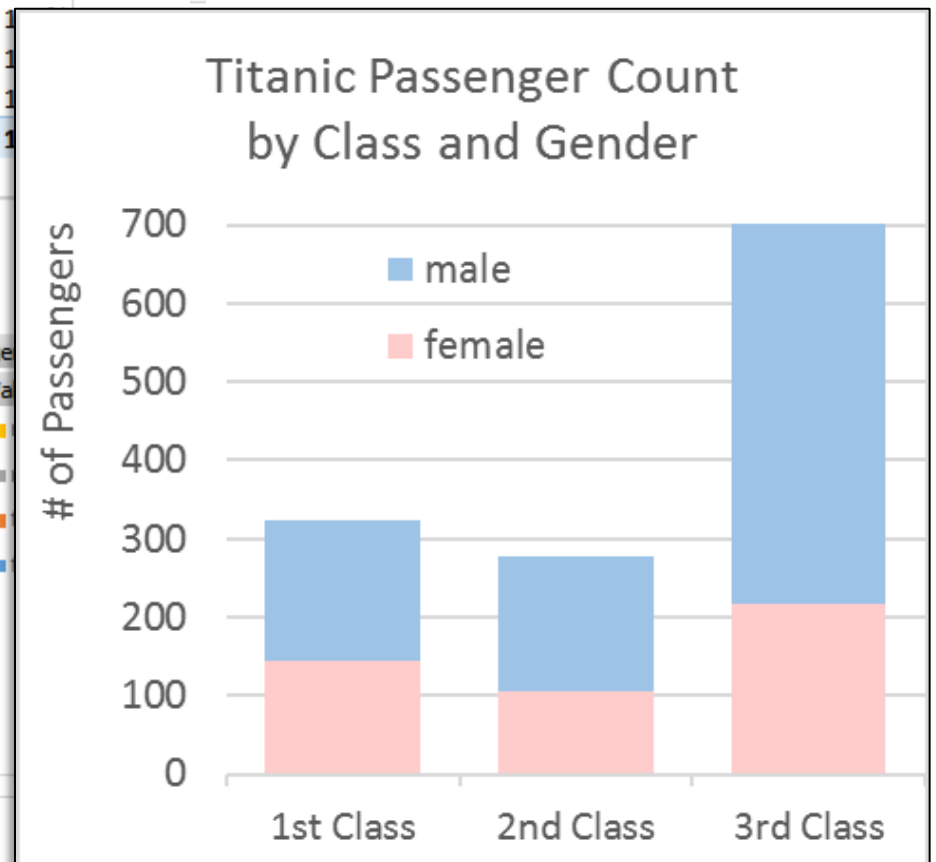
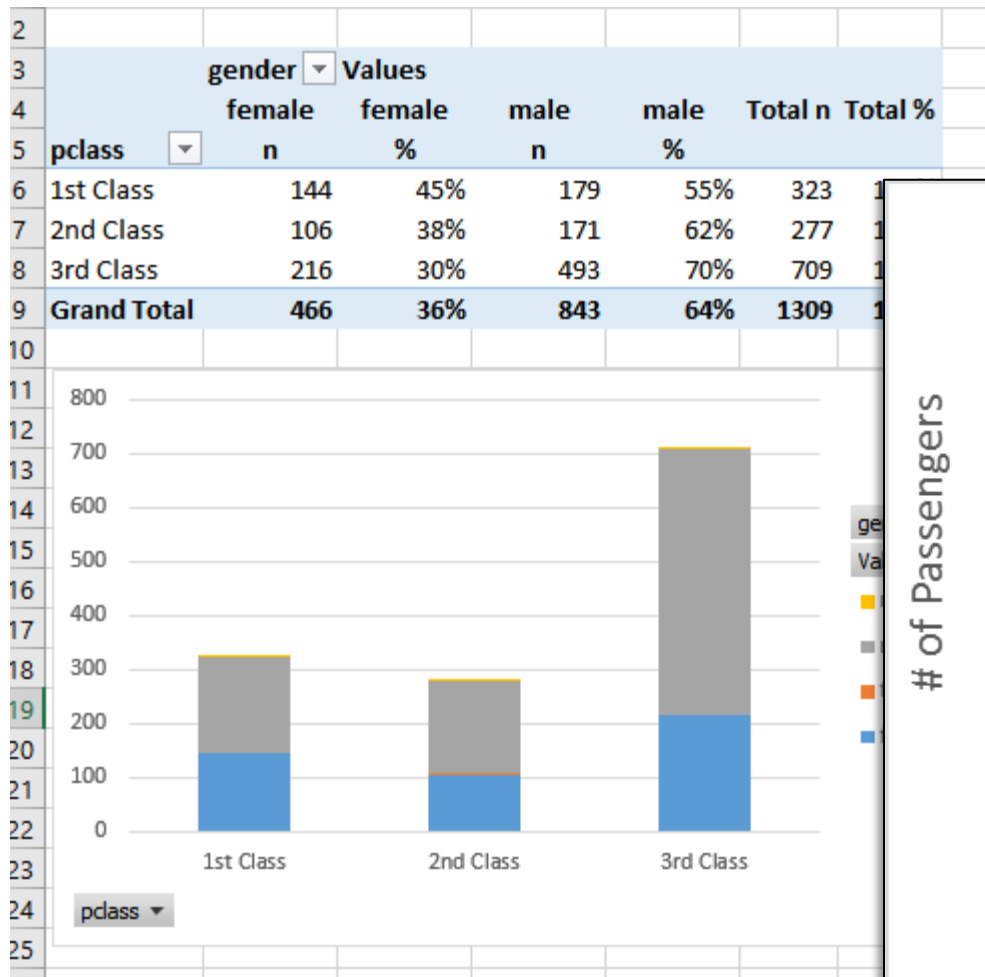
Picture

**DEFAULT**  
for Ctrl-V

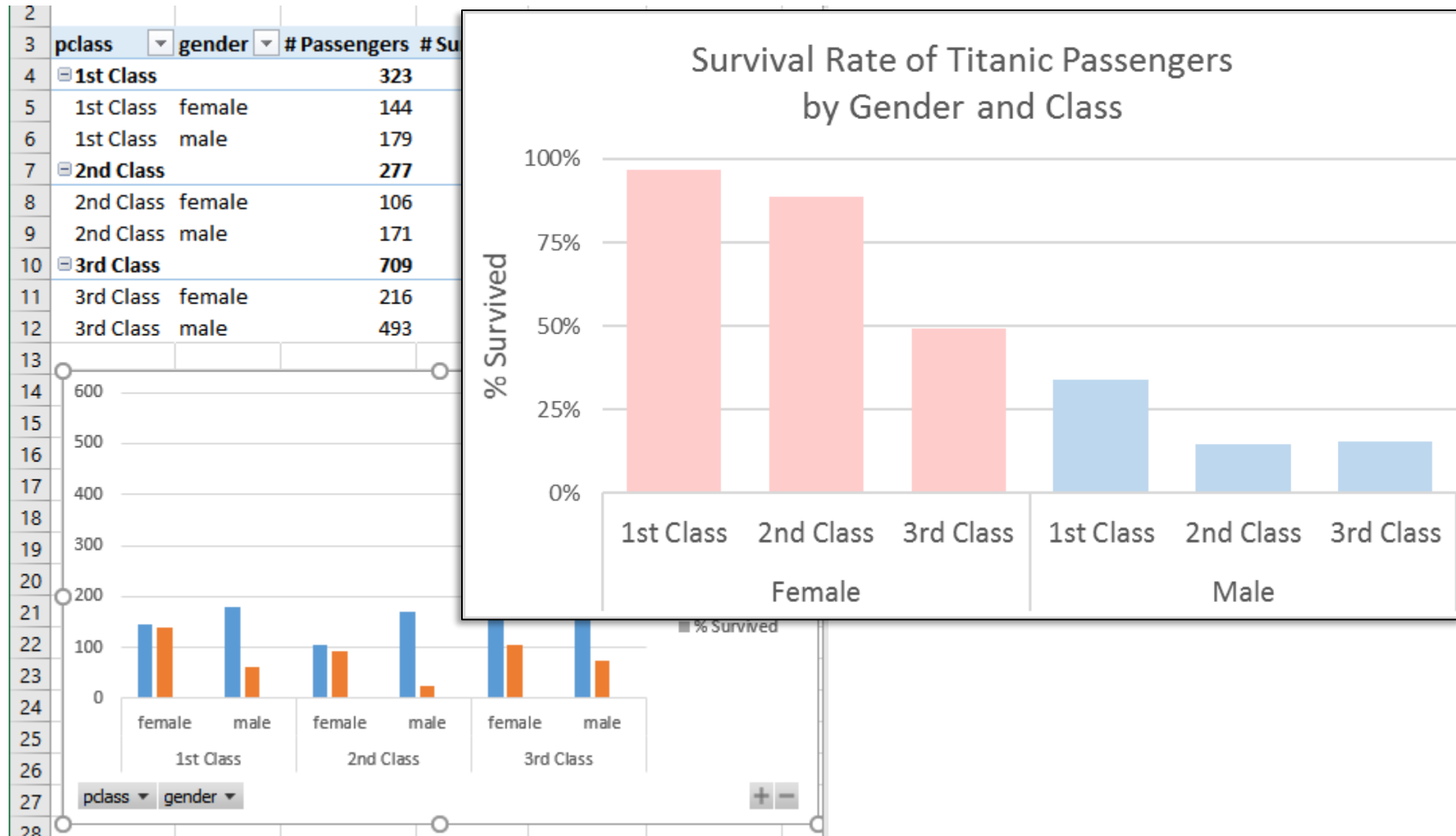


# Stacked Bar Chart

What's interesting here?



# Clustered Bar Chart



# Area Chart

age2	# Passengers	# Survived	% Survived
0-4	51	33	65%
5-12	43	21	49%
13-18	60	27	45%
19-24	255	96	38%
25-34	292	110	38%
35-44	169	64	38%
45-54	109	53	49%
55+	67	23	34%

