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/file**s**/**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, ,
```

commas, commas, and values

```
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,
```

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	.251	99999	1	999	2	3	2

looks like columns, but not all lined up

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	.21	99999	1	999	1	4	3	2
14447	.305	99999	1	999	4	2	1
13619	.682	99999	1	999	1	3	3
8291	.749	99999	1	999	1	1	1

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/file`**`s/titanic`**`.csv`
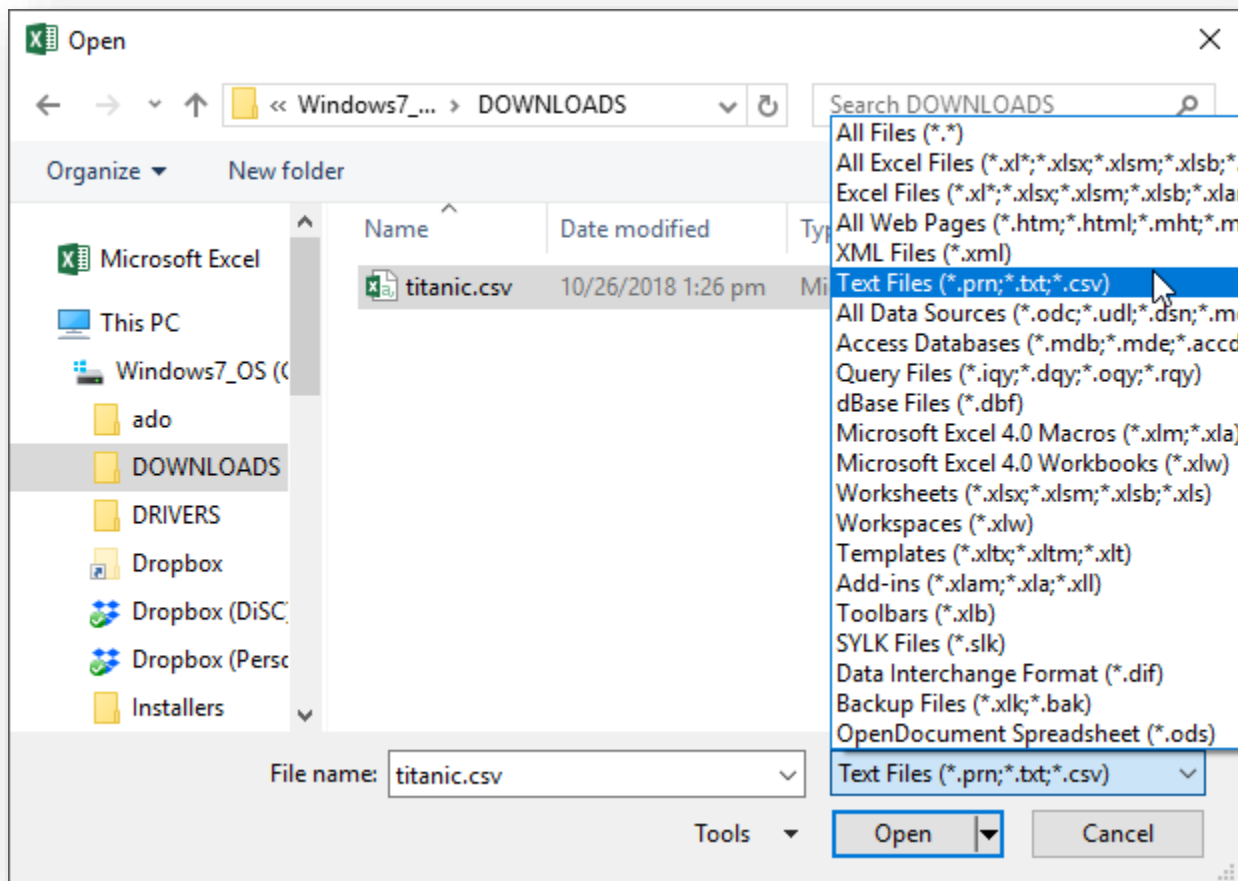
Understanding Data Tables

Variables & Observations

Variable
a characteristic
Did they survive?

Value
characteristic of thing
Person #2 was in 3rd 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	7.55 S
3	2	Abbott, Master. Euge	0	13	male	0		3	20.25 S
4	3	Abbott, Mr. Rossmor	0	16	male	1	1	3	20.25 S
5	4	Abbott, Mrs. Stanton	1	35	female	1	1	3	20.25 S
6	5	Abelseth, Miss. Karer	1	16	female	0	0	3	7.65 S
7	6	Abelseth, Mr. Olaus J	1	25	male	0	0	3	7.65 S
8	7	Abelson, Mr. Samuel	0	30	male	1	0	2	24 C
9	8	Abelson, Mrs. Samue	1	28	female	1	0	2	24 C
10	9	Abrahamsson, Mr. Ab	1	20	male	0	0	3	7.925
11	10	Abrahim, Mrs. Joseph	1	18	female	0	0	3	7.2292 C
12	11	Adahl, Mr. Mauritz N	0	30	male	0	0	3	7.25
13	12	Adams, Mr. John	0	26	male	0	0	3	8.05
14	13	Ahlin, Mrs. Johan (Jo	0	40	female	1	0	3	9.475 S
15	14	Aks, Master. Philip Fr	1	0.83	male	0	1	3	9.35 S
16	15	Aks, Mrs. Sam (Leah I	1	18	female	0	1	3	9.35 S
17	16	Albimona, Mr. Nasse	1	26	male	0	0	3	18.7875 C
18	17	Aldworth, Mr. Charle	0	30	male	0	0	2	13 S
19	18	Alexander, Mr. Willia	0	26	male	0	0	3	7.8875 S
20	19	Alhomaki, Mr. Ilmari	0	20	male	0	0	3	7.925 S

Case/Observation
thing described
Passenger (person)

6

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 1st Row**

**Each row includes
all information
(no headers)**

**No blank rows,
formatting, or
extra values**

	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	

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

None or FALSE →

Value is unknown

Correct	WRONG
0	-
	n/a
99	
-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
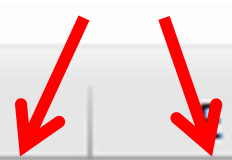
If 2 groups, use **Binary** (1 and 0)
    Easy & fast to type
    Mathematically Useful

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

Pick ONE

C	D	
fate	survived	died
died	0	1
died	0	1
died	0	1
survived	1	0
survived	1	0
survived	1	0
died	0	1
survived	1	0

Avoid Changing Values Individually

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

Always choose "Match entire cell contents"
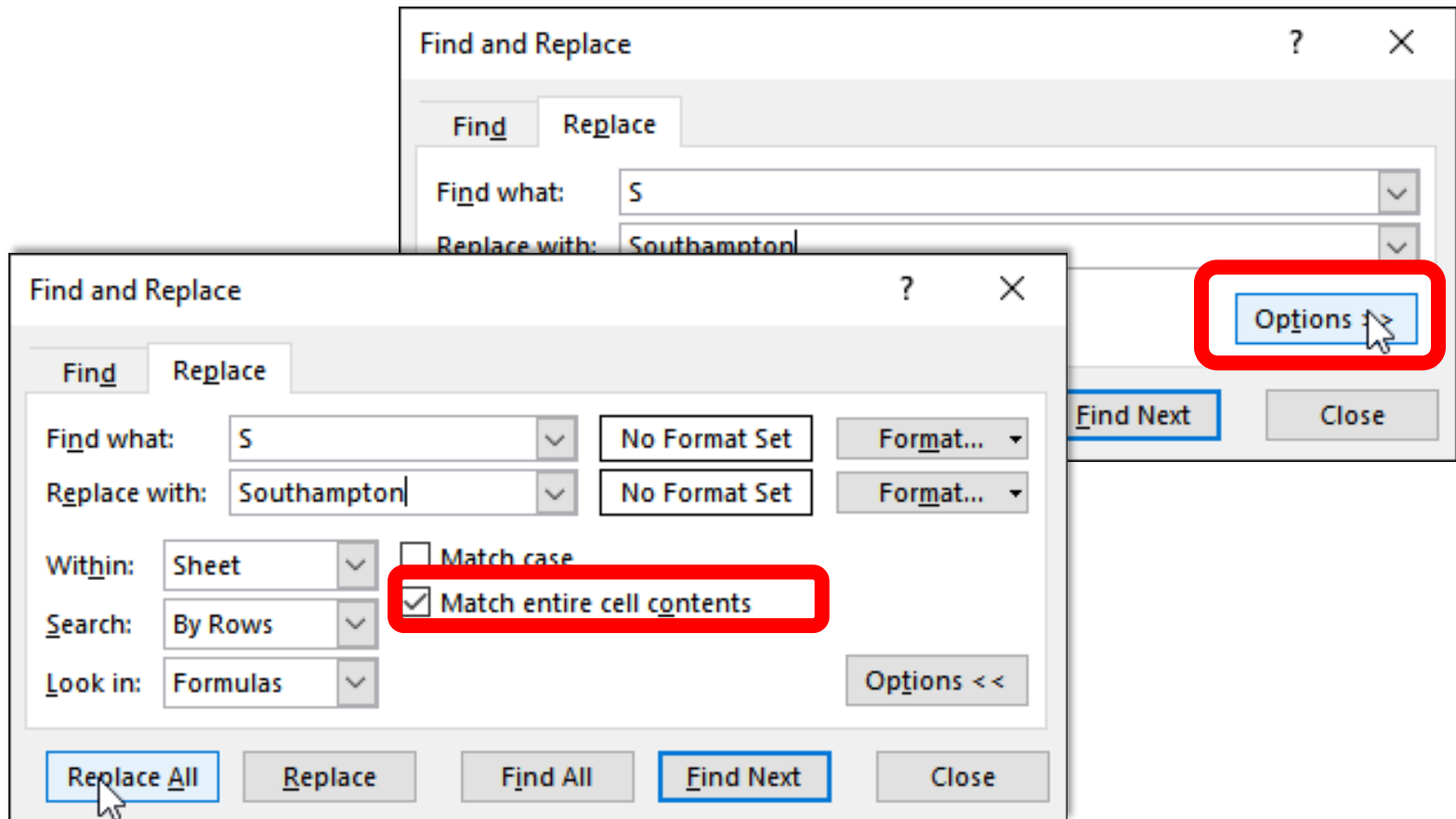
Dates

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/file**s**/**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)
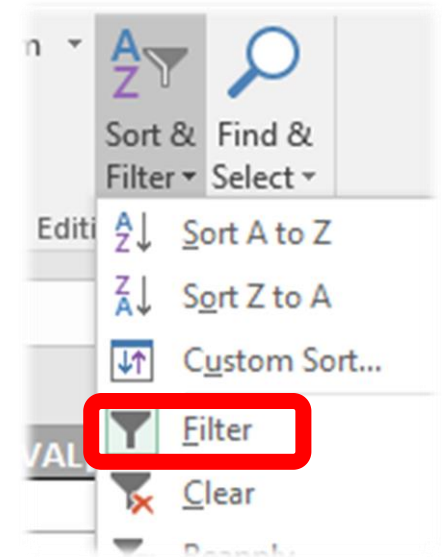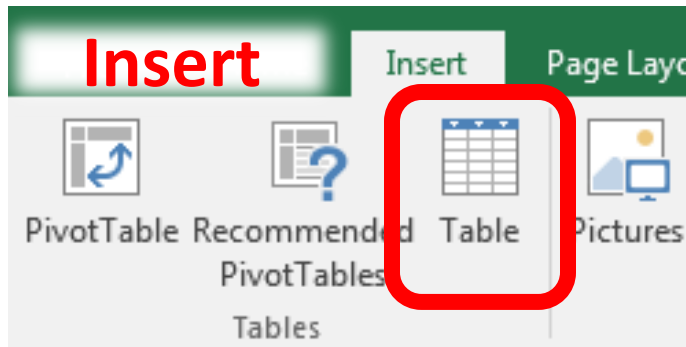
Why Create an Excel Table Object?

Automatically…
- adds new rows and columns
- fills formulas to the whole column
- adds the table to the Data Model
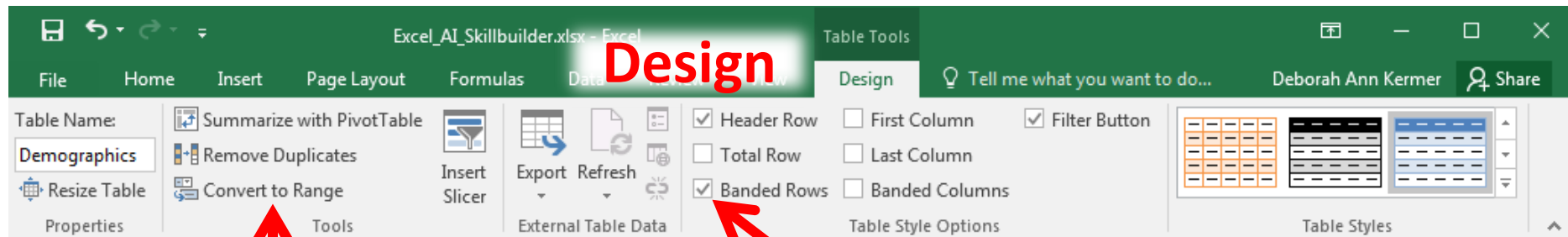
If you have trouble with other features, **Convert** [back] **to Range**

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

Drop-Downs
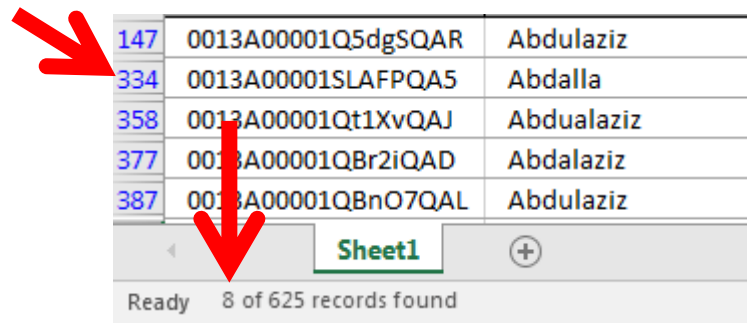
ONLY Sort or Filter with this

Check out the Filters in Date Columns!
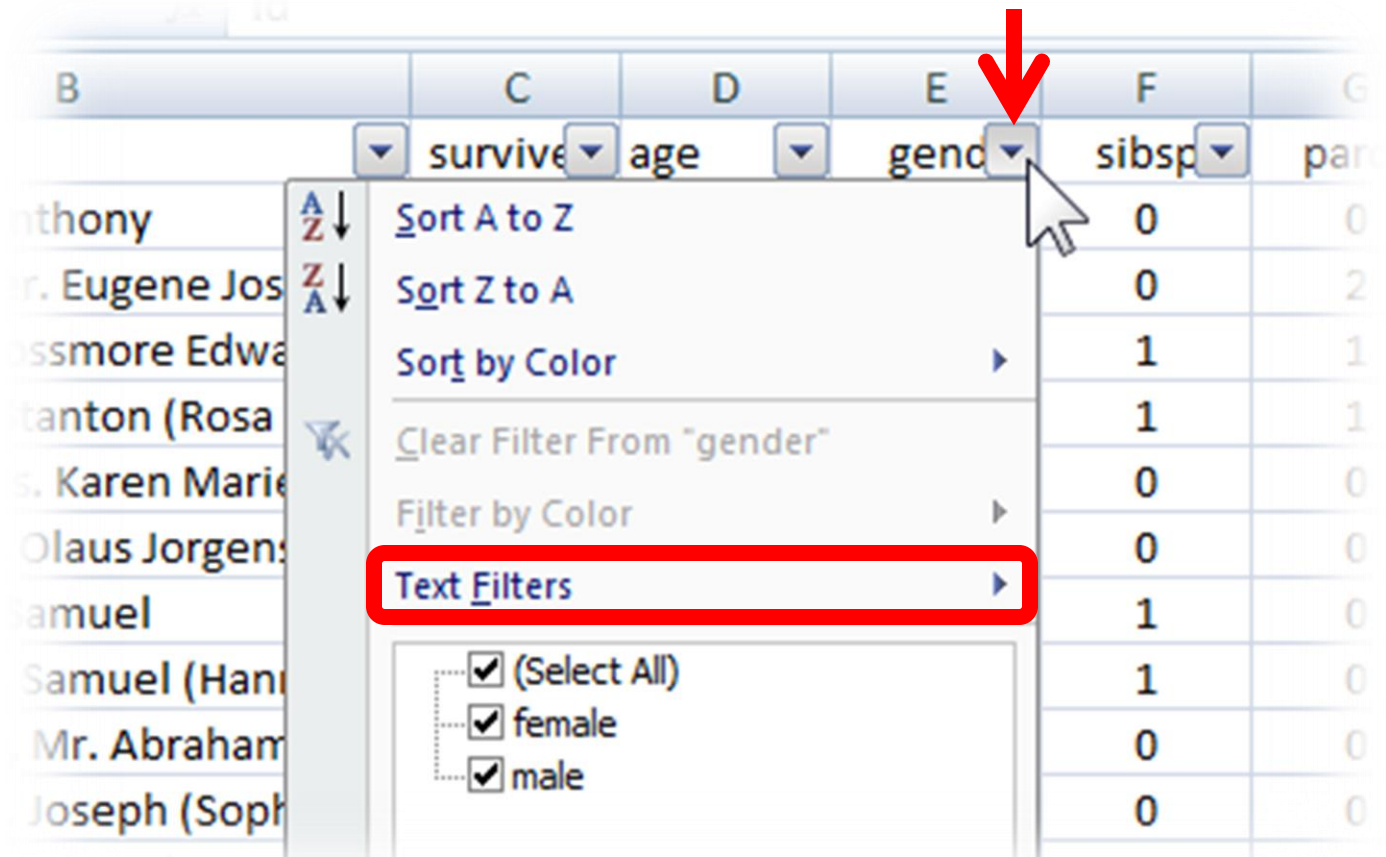
Filters

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

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

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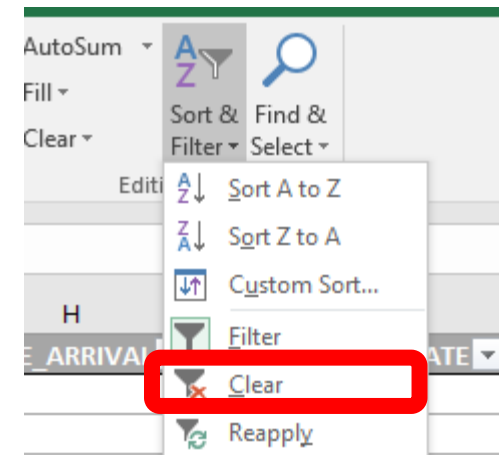
Pivot Tables

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

Insert

Design
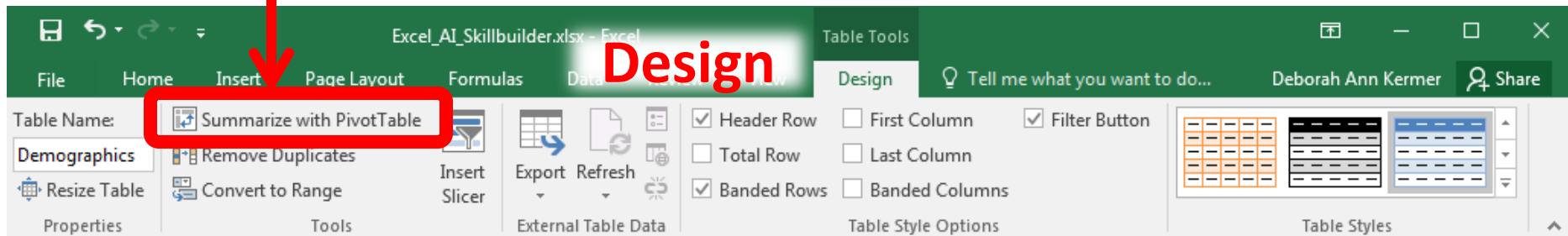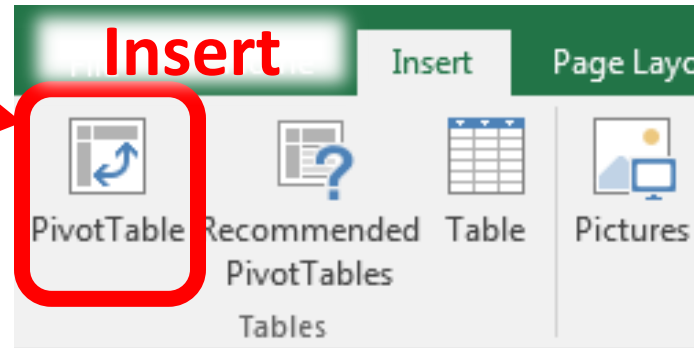
Check Data Source
then click **OK**

23
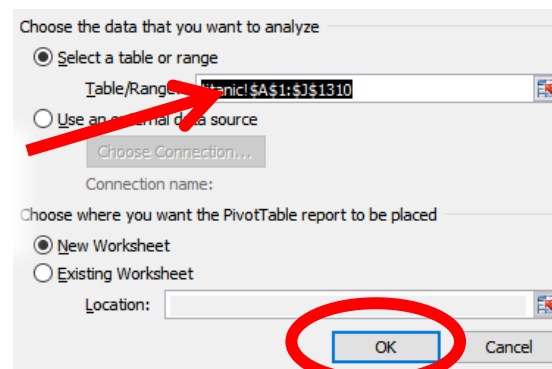
Frequency Table

Step 1: Values

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

Drag it to the **Values** box.
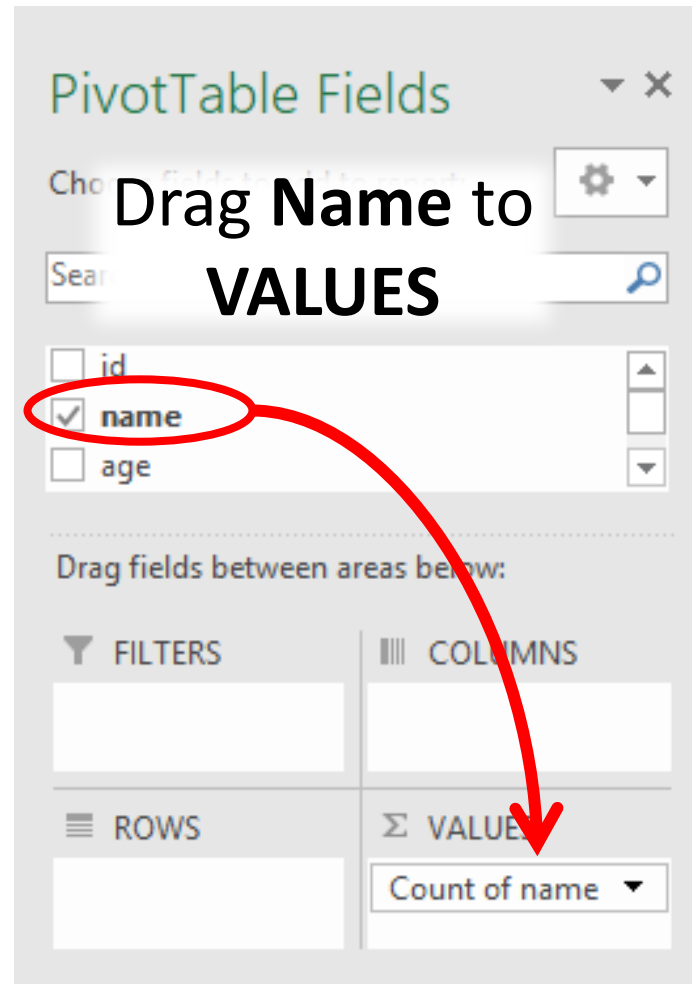
Always Check

Check your n (sample size)

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

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

Percentages

Right click the last column and choose
Show Values As … % of Column Total
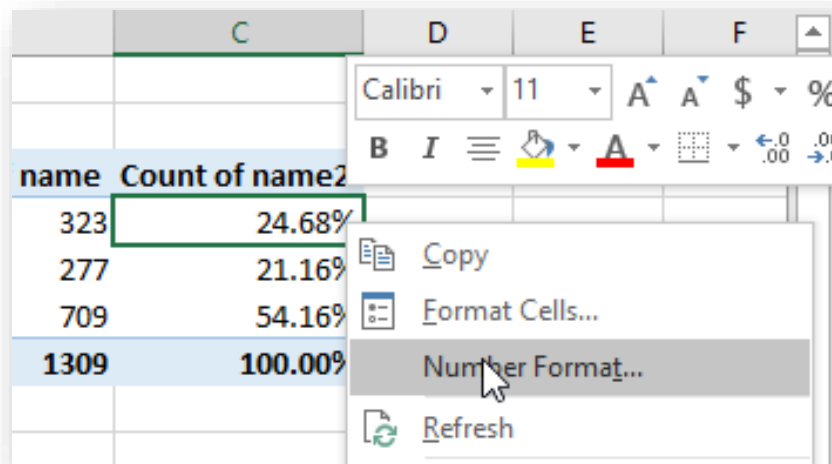
Number Format

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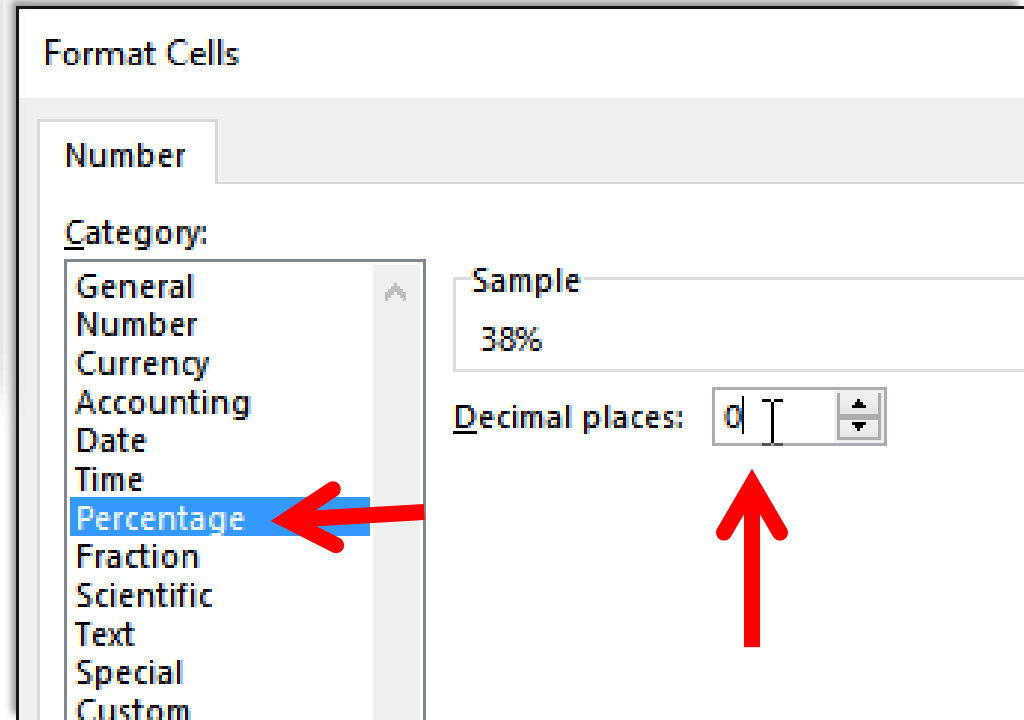
Labels

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			

You can center the labels
and reduce the column size.

Drag fields between areas below:

▼ FILTERS

▥ COLUMNS

Σ Values ▼

☰ ROWS

Σ VALUES

class ▼

n ▼

% ▼

☐ Defer Layout Update

UPDATE

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
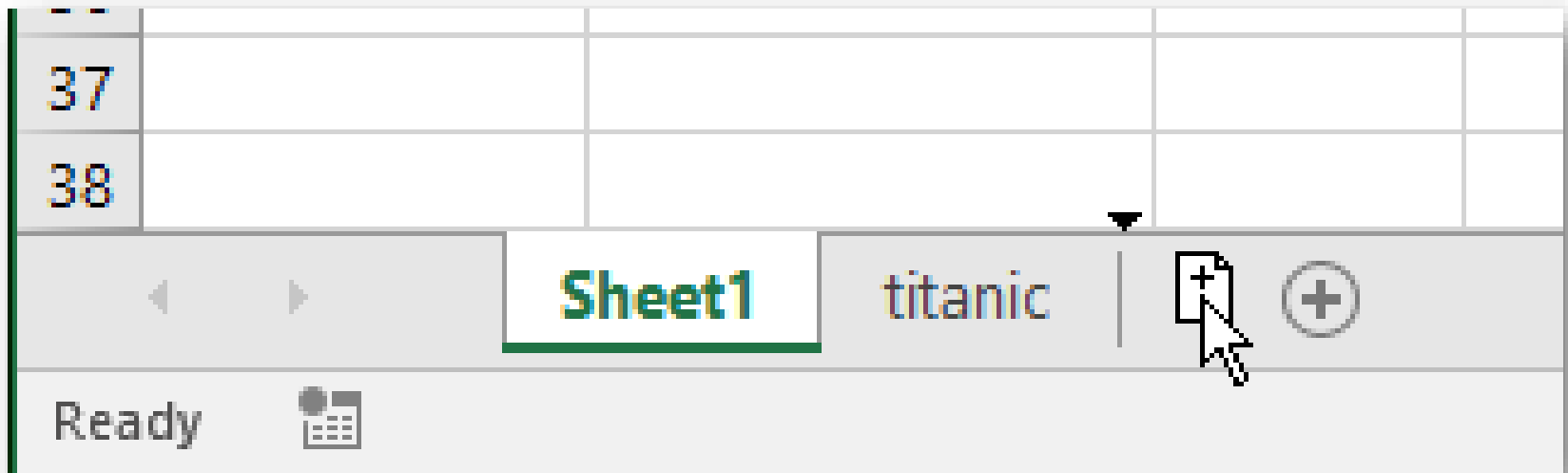
Nesting & Crosstabulation

Drag **gender** to **ROWS**
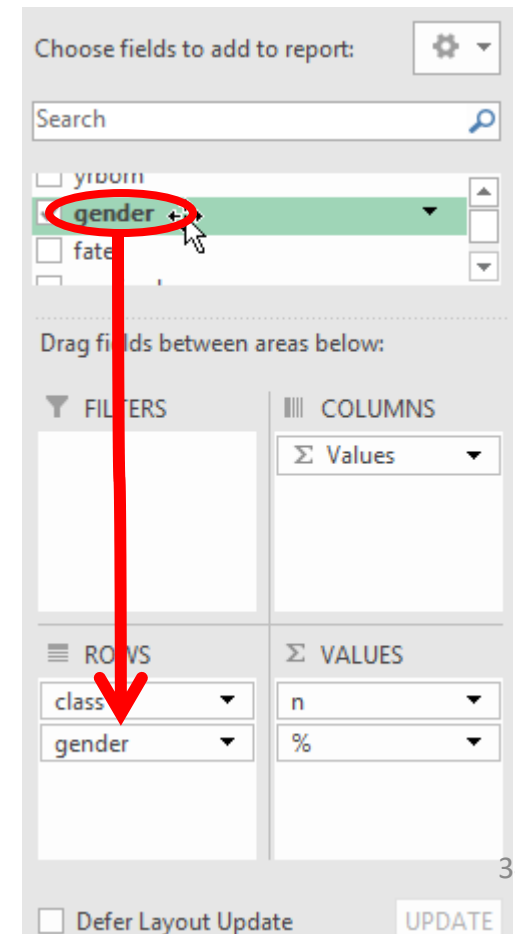
Multiple Variables in Rows

pclass
gender

gender
pclass

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

Design Tab

Subtotals | Grand Totals | Report Layout | Blank Rows

Layout

Show in Compact Form

Show in Outline Form

Show in Tabular Form

Repeat All Item Labels

Do Not Repeat Item Labels
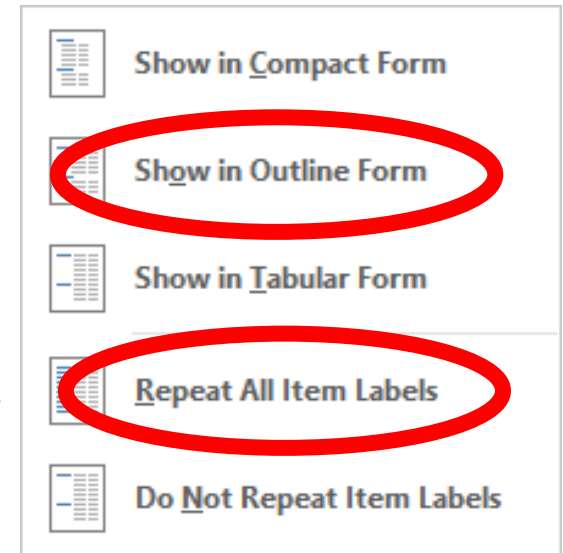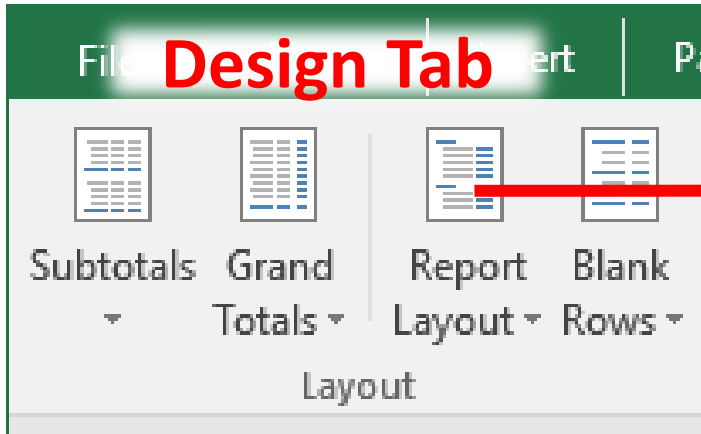
Compact

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

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

	class	gender	n	%
	⊟ 1st Class			
	1st Class	female	144	11%
	1st Class	male	179	14%
	⊟ 2nd Class			
	2nd Class	female	106	8%
	2nd Class	male	171	13%
	⊟ 3rd Class			
	3rd Class	female	216	17%
	3rd Class	male	493	38%

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

37

Nesting to Crosstabulation

Drag gender to COLUMNS

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.

		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

PivotTable Fields

Drag **survived** to **VALUES**

- yrborn
- gender
- fate
- survived
- sibsp
- parch

Drag fields between areas below:

FILTERS

COLUMNS

ROWS
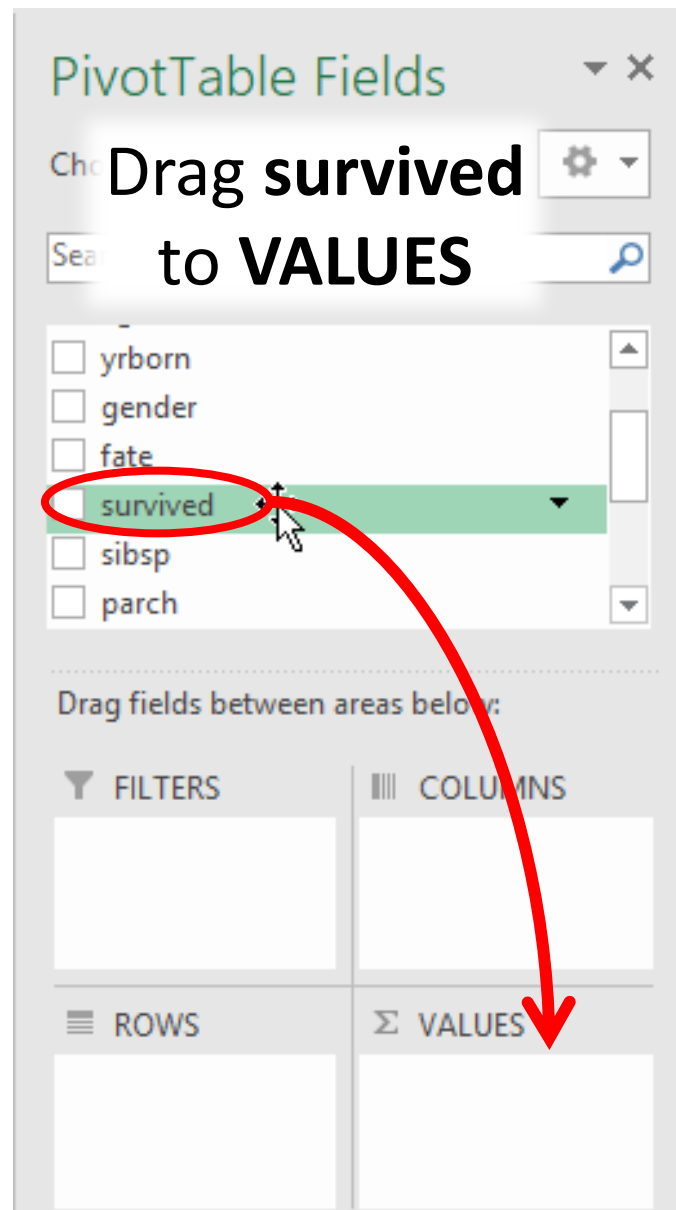
Σ VALUES

Value Field Settings

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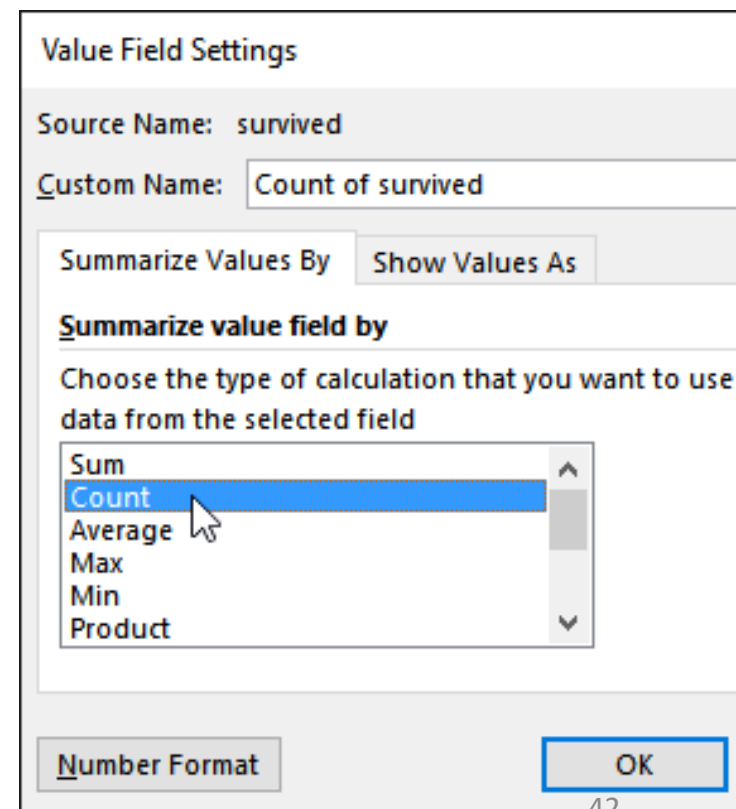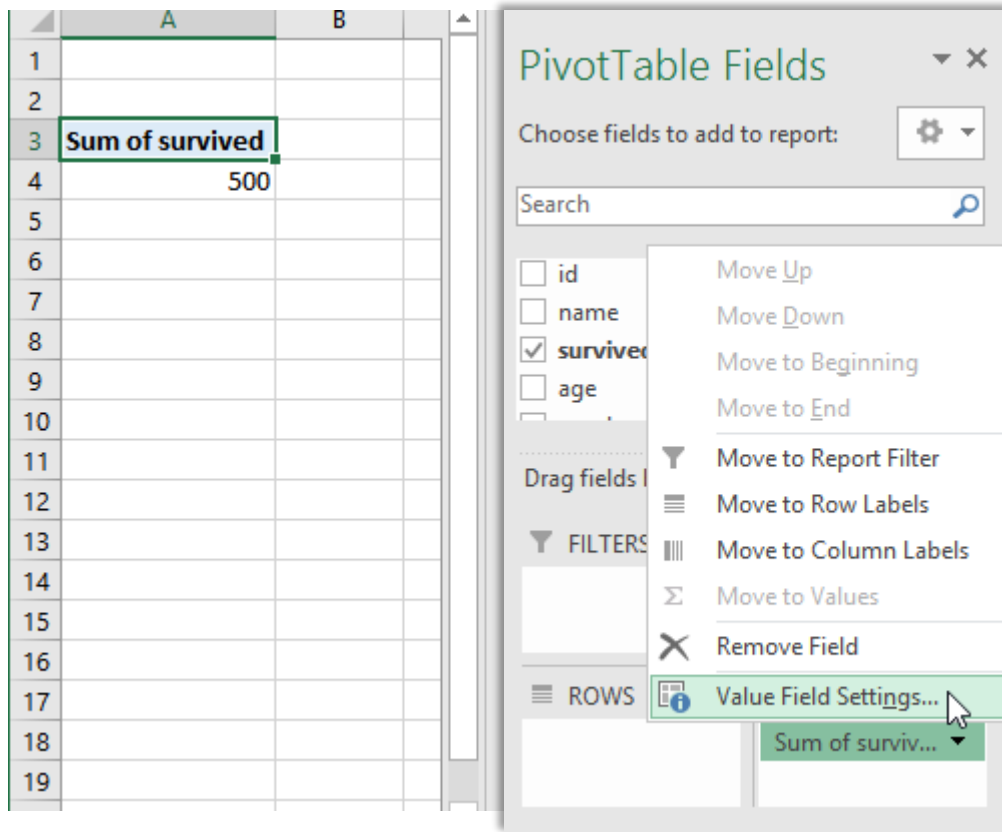
Multiple Statistics

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

FYI: Using the Boxes

Rows

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

Review

Drag fields between areas below:

⏳ Report Filter	▦ Column Labels
Limit by Group Try Slicers Instead	Secondary Grouping Variables

▦ Row Labels	Σ Values
Primary Grouping Variables	**Summary Variables**

☐ 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
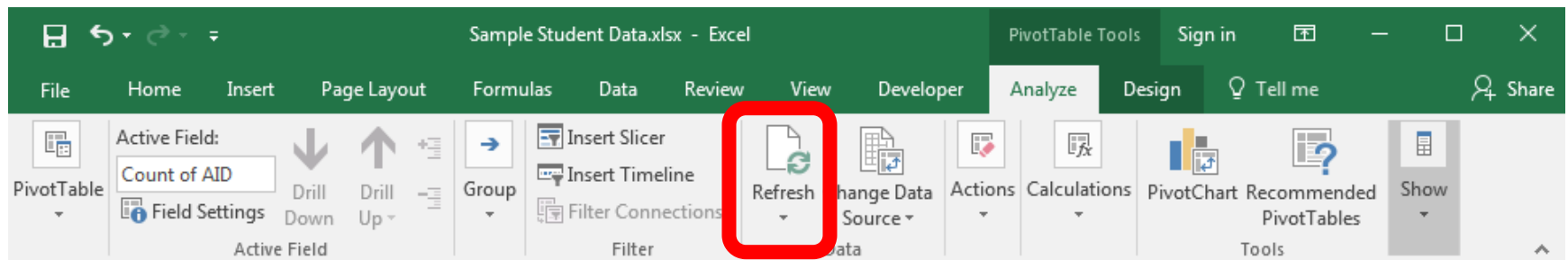
Grouping Numeric Values

Advanced

Drag **age** to **ROWS**
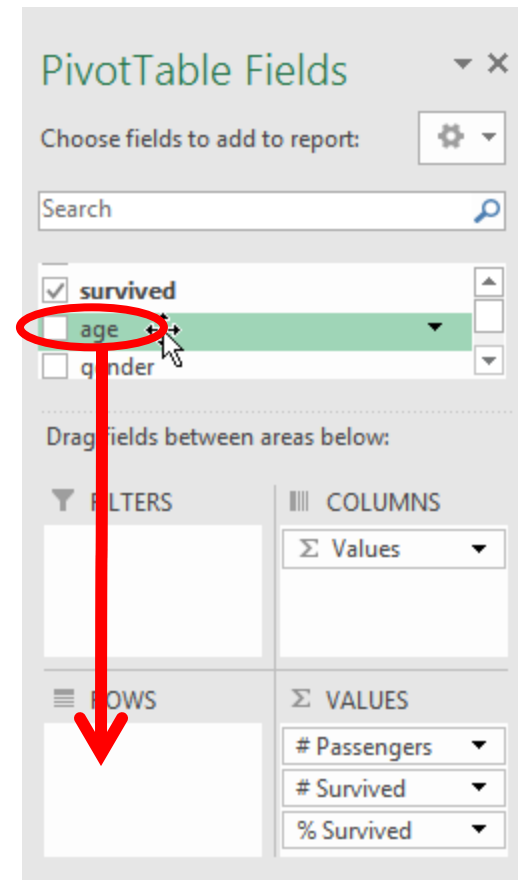
Automatic Grouping of Values

Right Click on a Value and choose "Group…"

Specify the groupings and click OK

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)

Manual Grouping Process

After the 1st Group, you will have **two** columns.

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

OR

Keep the Groups (Only)

Drag out (remove) the original age variable

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
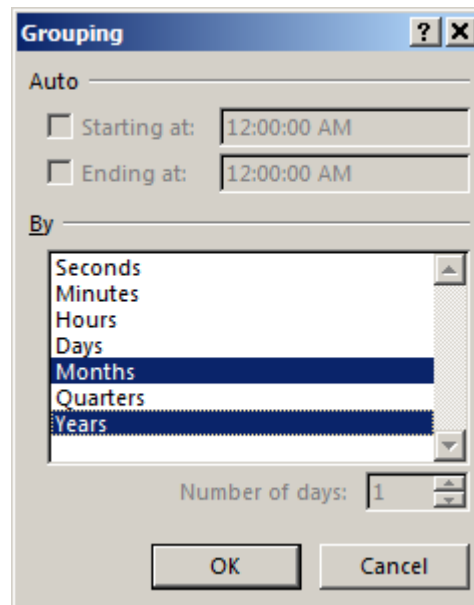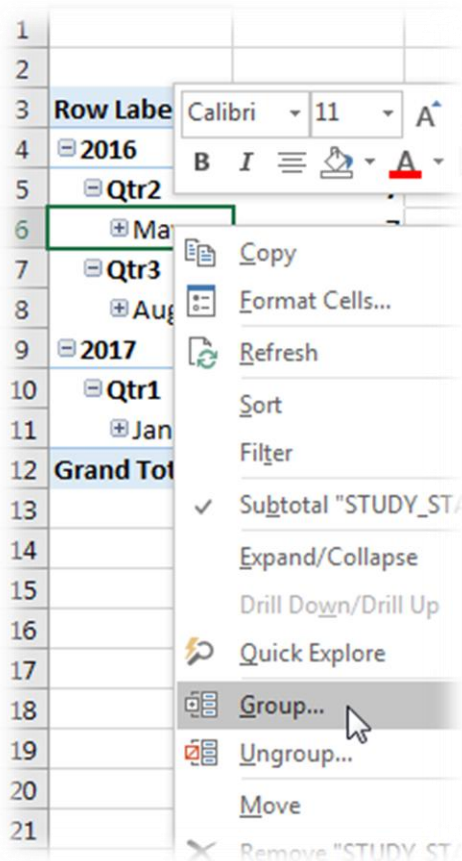
Σ Values ▾

☰ ROWS

age2 ▾

Σ VALUES

Passengers ▾
Survived ▾
% Survived ▾

FYI: Grouping of Dates

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 anlayses
- **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.
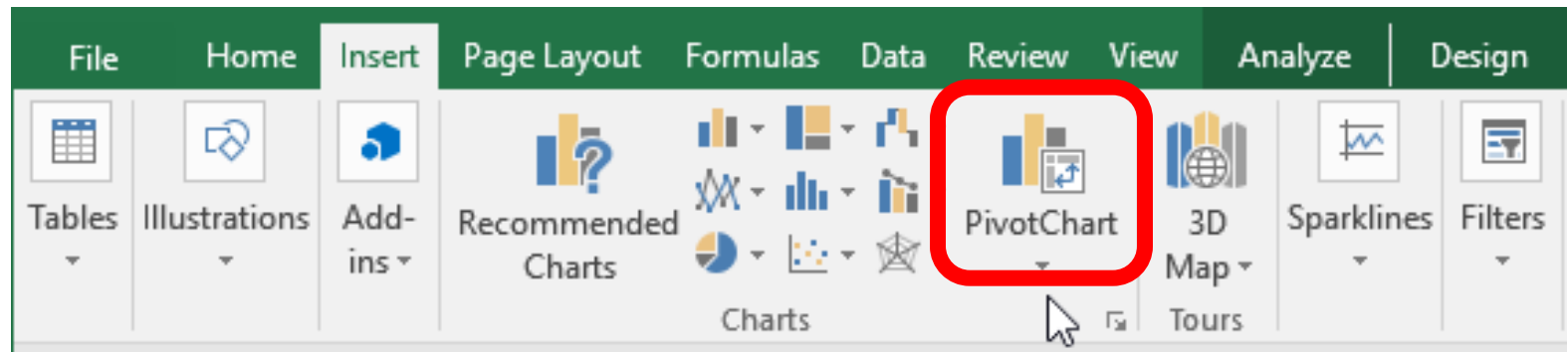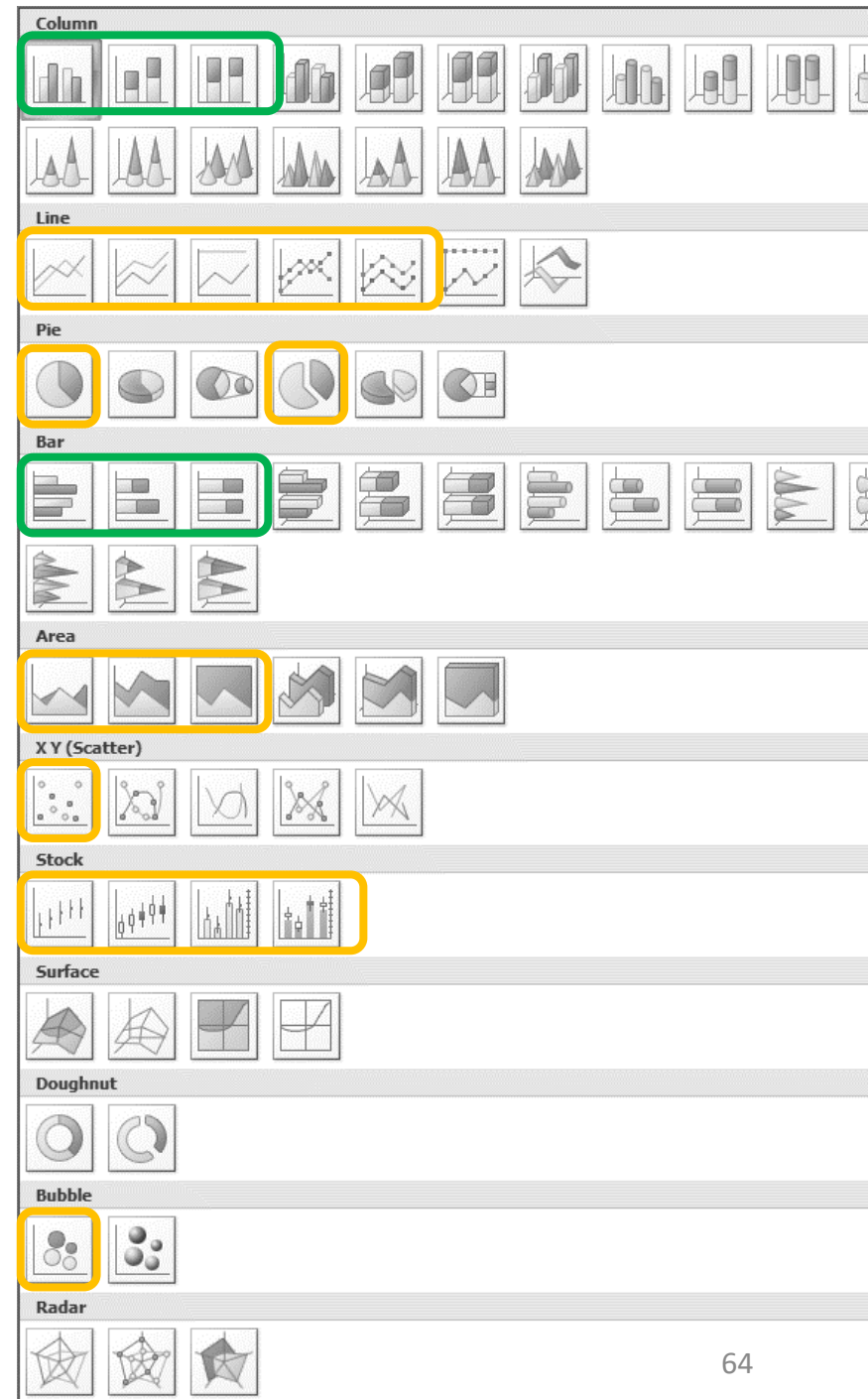
Pivot Charts

Use AFTER creating a Pivot Table

Insert Chart
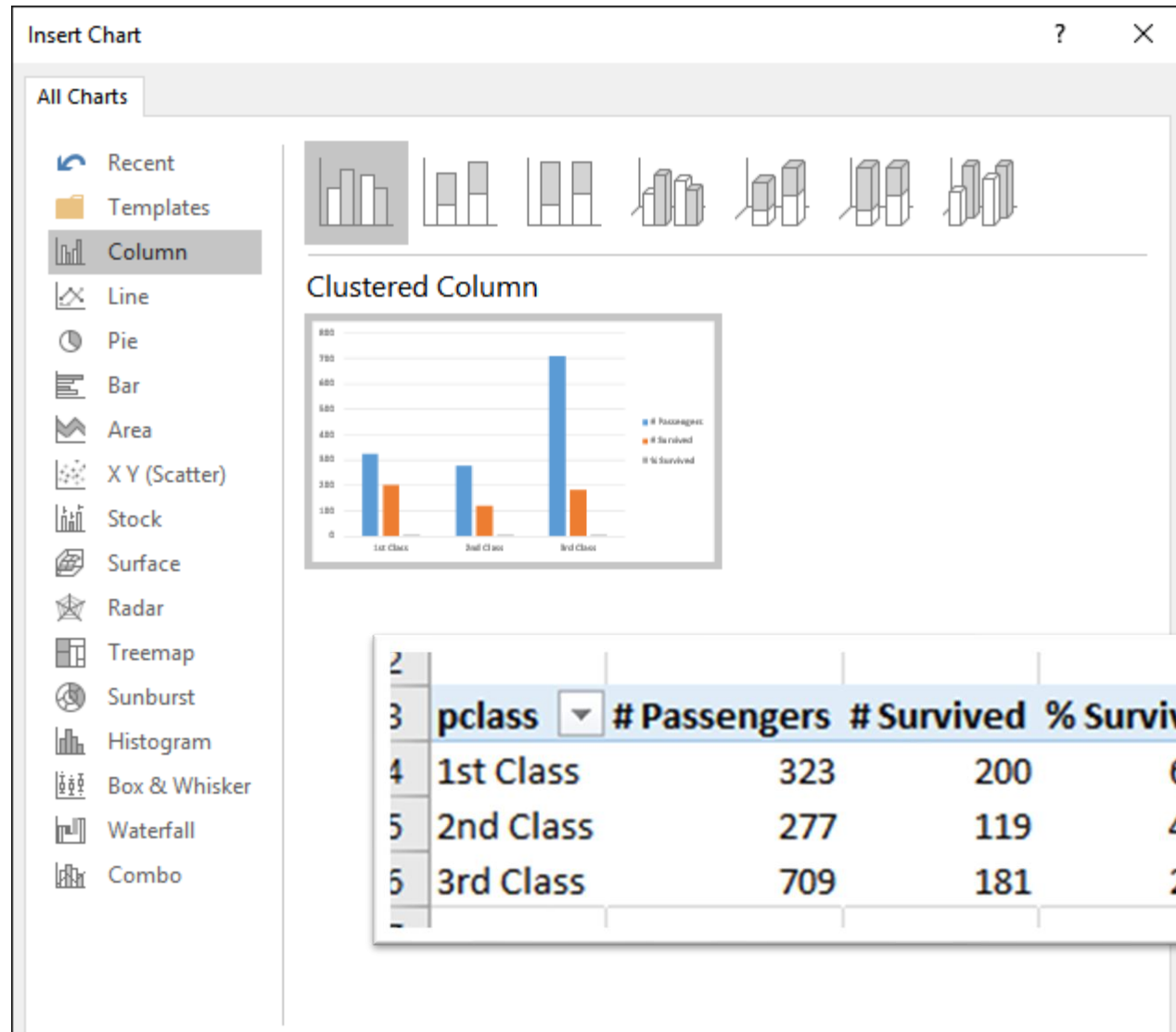
Chart Types

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?

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

For me: the difference in survival rate between classes.

Design a chart to make that pattern pop out

Bar Chart

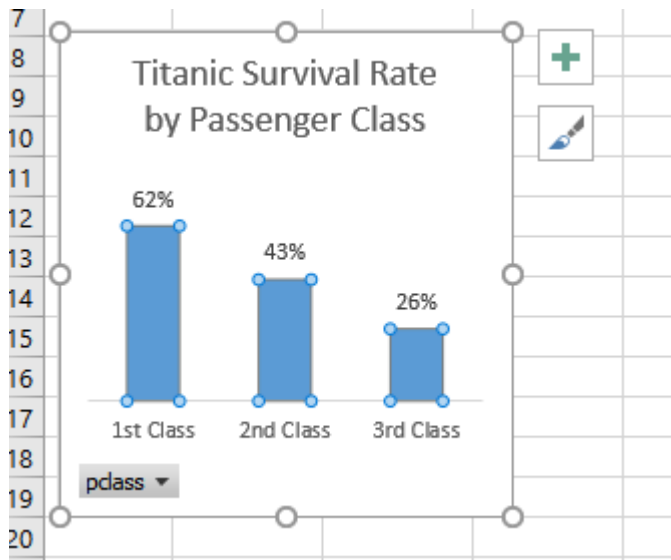| pclass	# Passengers	# Survived	% Survived
1st Class	323	200	62%
2nd Class	277	119	43%
3rd Class	709	181	26%

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

Where is % Survived?

Proportions are < 0

Modifications
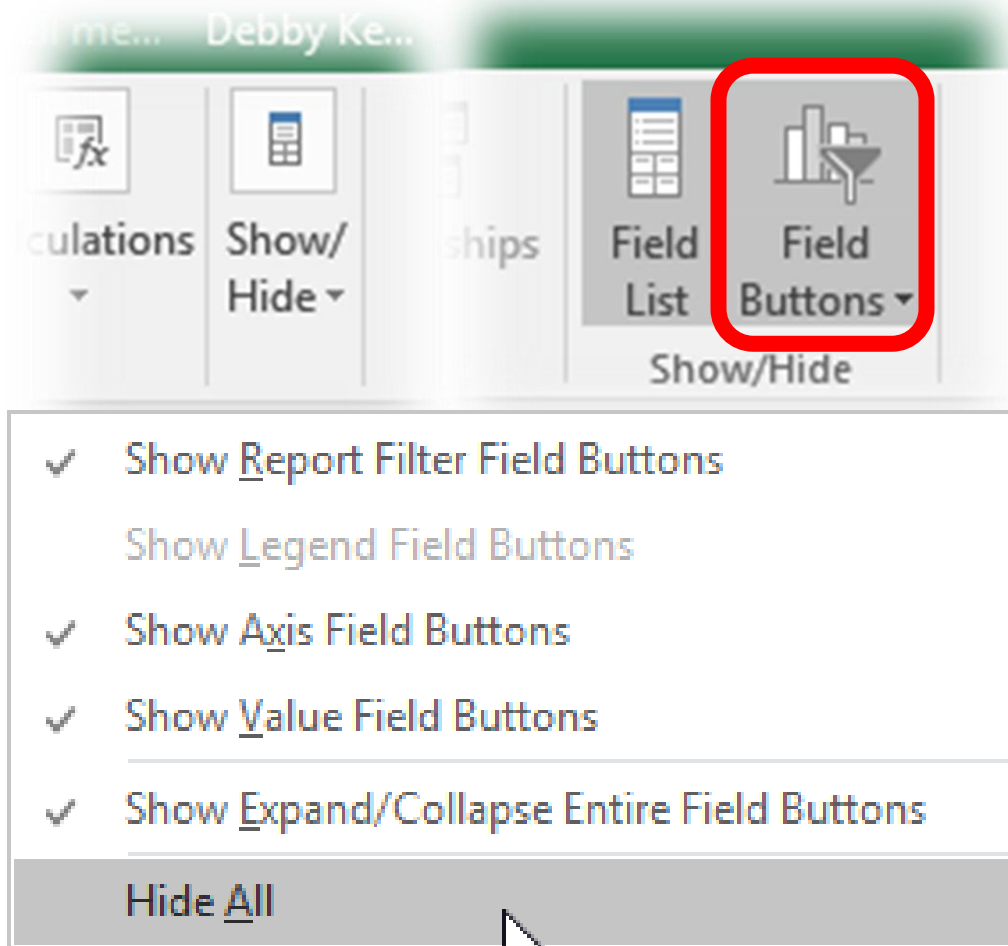
Right-click the bars and choose "Format Data Series…"
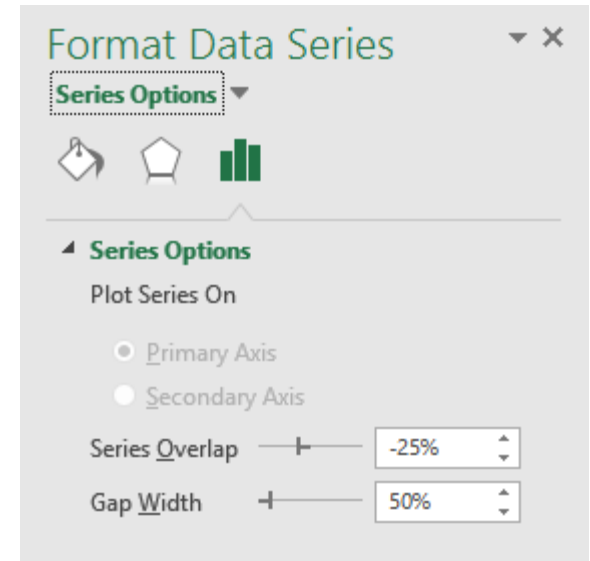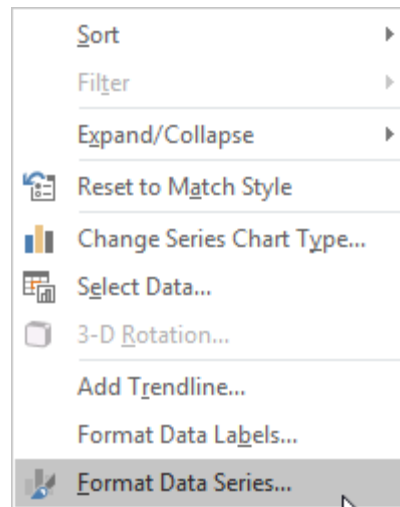
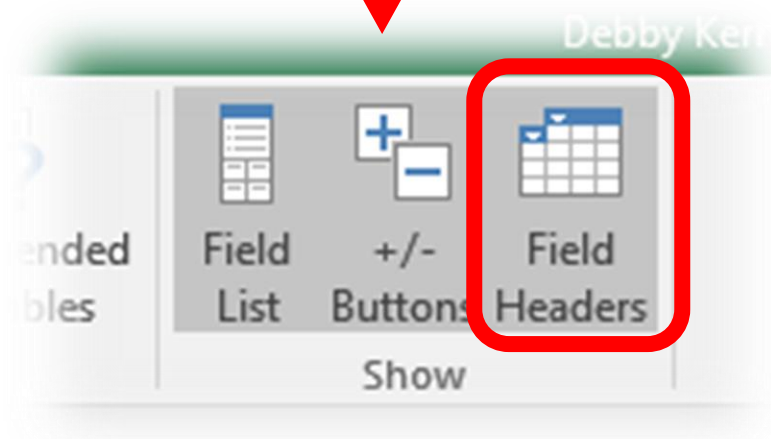**Remove**: "# Passengers", "# Survived", Gridlines, Y-axis
Add: Chart Title, Data Labels
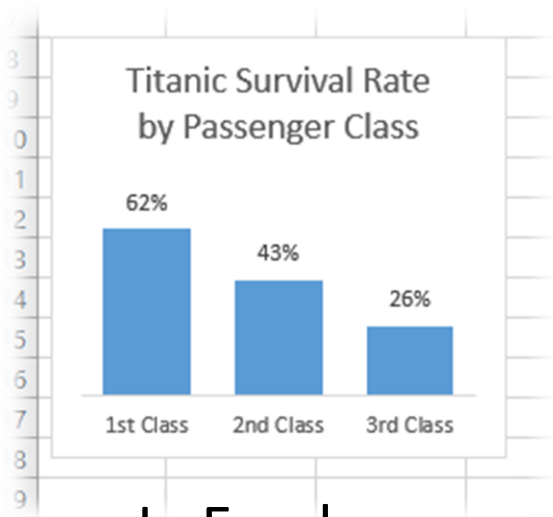**Change**: Gap Width, Font Size
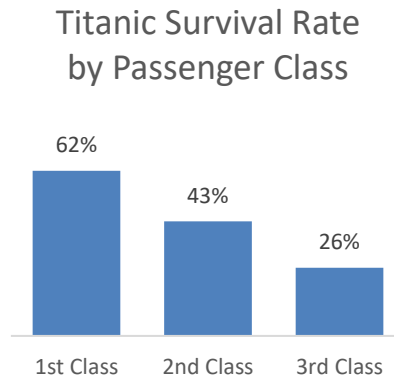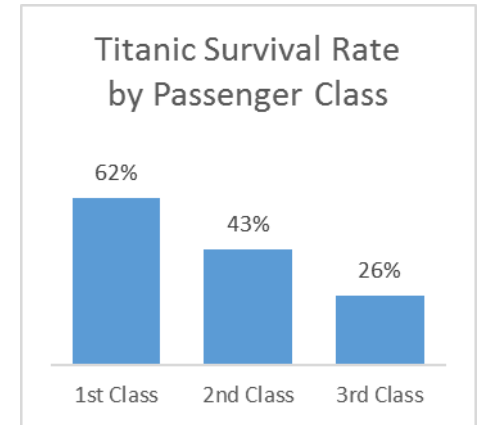
Removing Extra Buttons

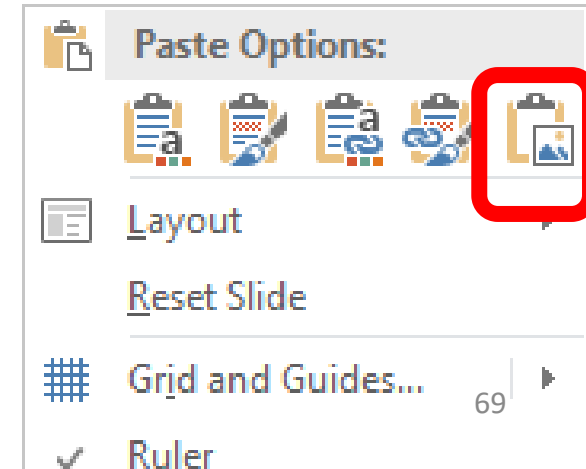Charts

Tables

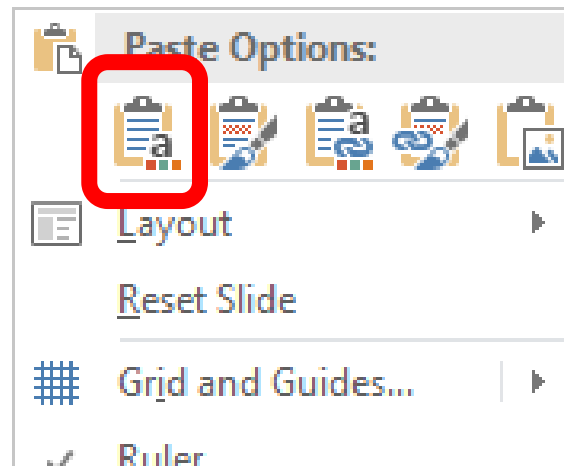# Copy-Paste to Documents
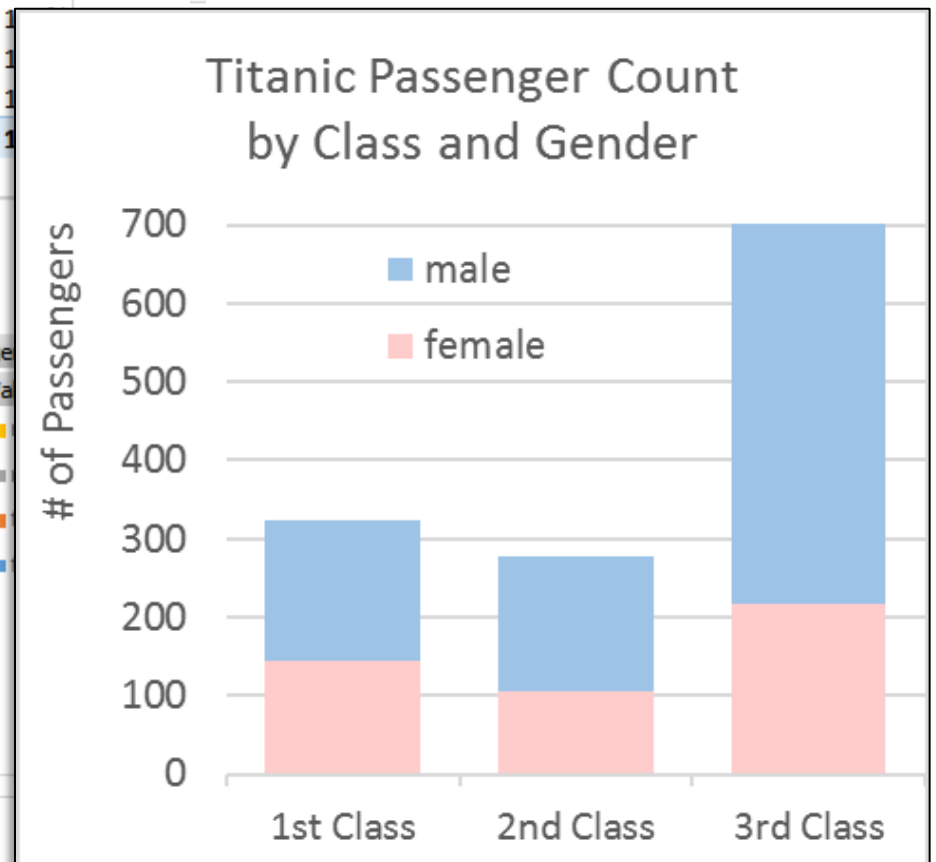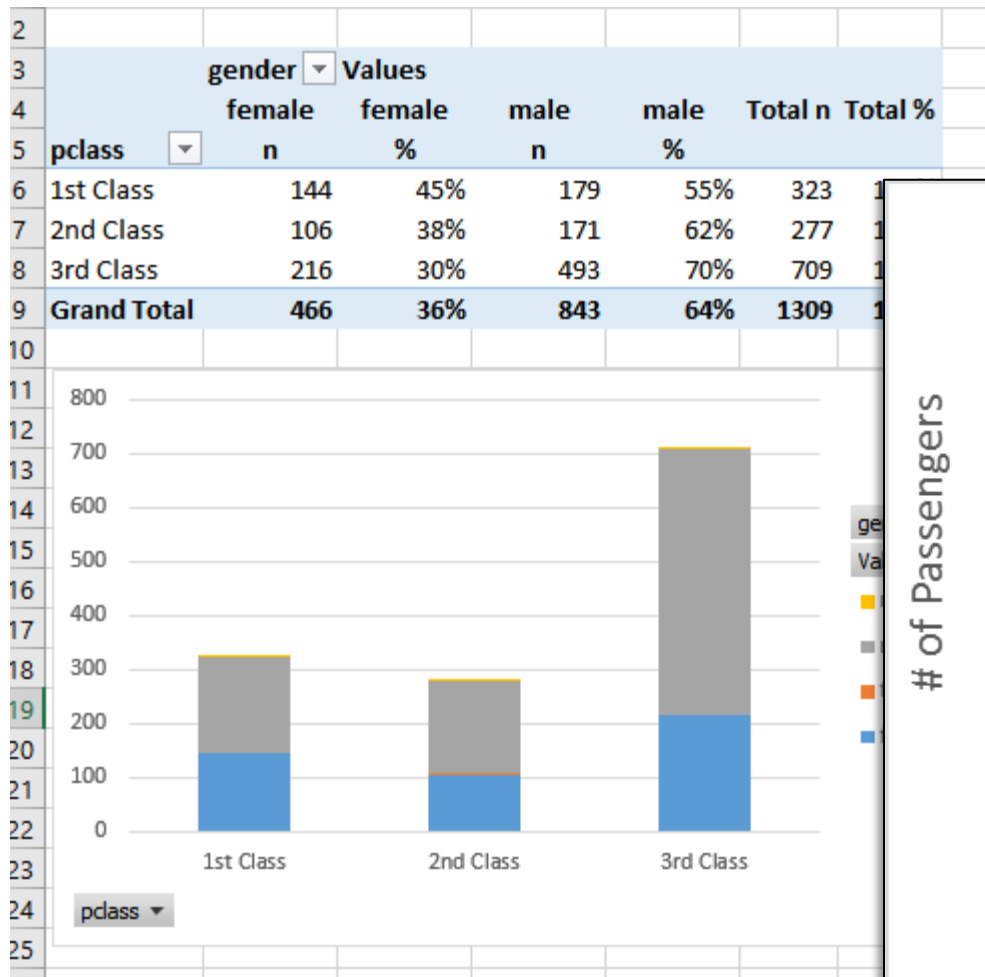

In Excel

Use Destination Theme and Embed Workbook

Picture

DEFAULT
for Ctrl-V

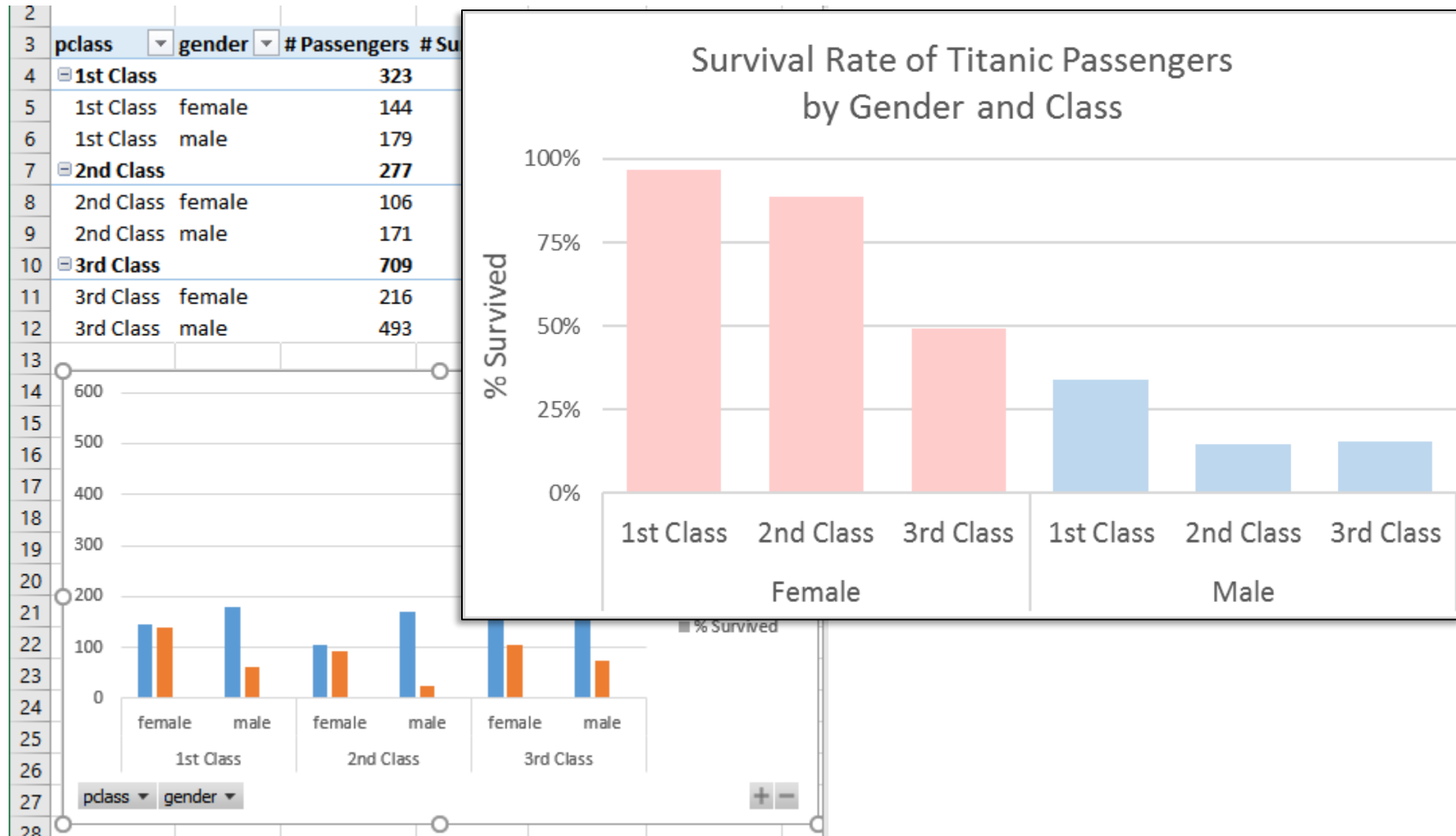69

Stacked Bar Chart

What's interesting here?

pclass	gender female n	female %	male n	male %	Total n	Total %
1st Class	144	45%	179	55%	323	
2nd Class	106	38%	171	62%	277	
3rd Class	216	30%	493	70%	709	
Grand Total	466	36%	843	64%	1309	

Titanic Passenger Count by Class and Gender

Clustered Bar Chart

Area Chart