

UCB CALSTAPH EXCEL AND EPIDEMIOLOGY

Gail Sondermeyer, MPH
Infectious Diseases Branch
California Department of Public Health
gail.sondermeyer@cdph.ca.gov

Overview

- Introduction
- Functions
- Point and Click
- Excel & CDPH
- Excel & other programs

**See corresponding Excel sheet for examples*

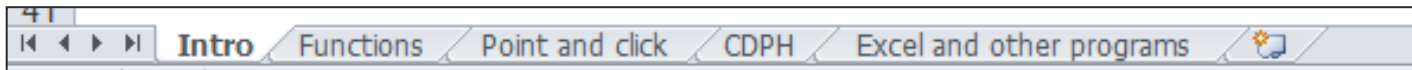
Overview

- **Introduction**
- Functions
- Point and Click
- Excel & CDPH
- Excel & other programs

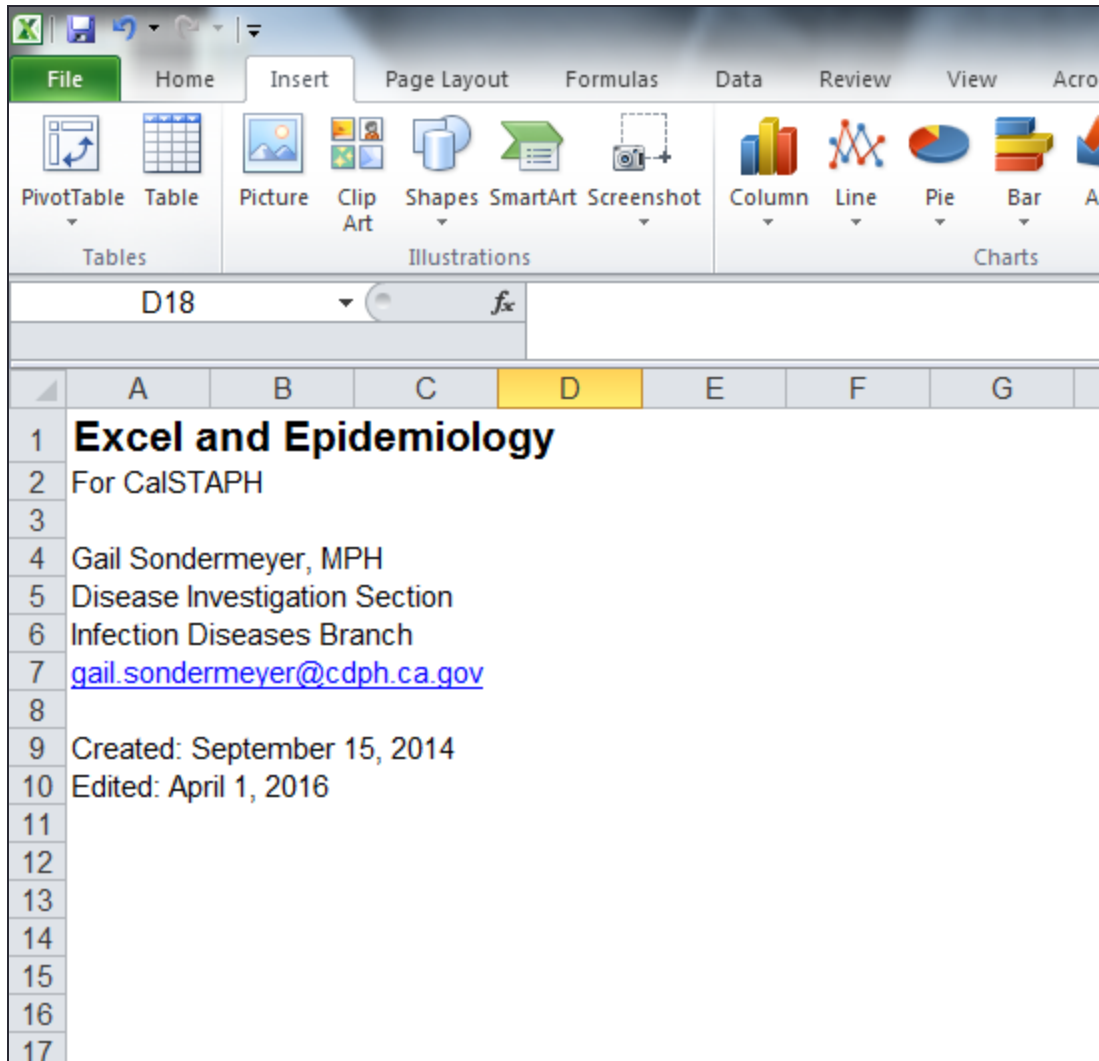
**See corresponding Excel sheet for examples*

Introduction

- Excel & epidemiology
 - Data cleaning
 - Analysis
 - Presentation
- Excel functionalities
 - Functions – type formulas into cells
 - Point and click
 - Other
 - Macros



Introduction



Overview

- Introduction
- **Functions**
- Point and Click
- Excel & CDPH
- Excel & other programs

**See corresponding Excel sheet for examples*

Functions

- How to use:
 - Type formulas into cells
=fxn(cell range)
 - Add cell ranges by typing in text or highlighting rows/columns
e.g. =sum(F2:F5)
- Useful for data analysis and cleaning
- See Excel sheet for examples

Excel and Epidemiology

Frequently used functions

Sample dataset:	Sex	Age group	Count	Date entered
	male	<60	10	1/1/2014
	Male	>60	23	5/6/2014
	FEMALE	<60		8/6/2011
	fEmale	>60	18	9/8/2012

Function name	Formula	Function	Sample results	Use
Sum	=sum(F2:F5)	Add a set of numbers	51	Analysis
Mean	=average(F2:F5)	Calculate mean of a set of numbers	17	Analysis
Median	=median(F2:F5)	Calculate median of a set of numbers	18	Analysis
Minimum	=min(F2:F5)	Find minimum in set of numbers	10	Analysis
Maximum	=max(F2:F5)	Find maximum in set of numbers	23	Analysis
Upper	=upper(D5)	Make text uppercase	FEMALE	Data cleaning
Lower	=lower(D4)	Make text lowercase	female	Data cleaning
Proper	=proper(D4)	Capitalizes first letter of each word	Female	Data cleaning
Left	=left(D2,1)	Extract characters from beginning of text	m	Data cleaning
Right	=right(E5,2)	Extract characters from end of text	60	Data cleaning
Mid	=mid(D4,2,2)	Extracts characters starting at specified point	EM	Data cleaning
Count	=count(D2:D5)	Counts number of observations	3	Data cleaning/analysis
Count if	=countif(D2:D5,"male")	Counts number of observations meeting certain criteria	2	Data cleaning/analysis
Count A	=counta(F2:F5)	Counts the number of observations that are not blank	3	Data cleaning/analysis
Count blank	=countblank(F2:F5)	Counts the number of observations that are blank	1	Data cleaning/analysis
Day	=day(G2)	Extracts day in a date	1	Data cleaning
Month	=month(G3)	Extracts month in a date	5	Data cleaning
Year	=year(G4)	Extracts year in a date	2011	Data cleaning
Concatenate	=I2&E2&J2	Combine cells	(<60)	Data cleaning/presentation
Exact	=exact(D2,D3)	Compares two strings to see if they are the same	FALSE	Data cleaning

Other basic operations:

Addition	+
Subtraction	-
Multiplication	*
Division	/

Functions

- Double click cells with a formula to see what cell ranges are used in the calculation

Sample dataset:	Sex	Age group	Count	Date entered
	male	<60	10	1/1/2014
	Male	>60	23	5/6/2014
	FEMALE	<60		8/6/2011
	fEmale	>60	18	9/8/2012

Sample results	Use
=SUM(F2:F5)	Analysis
SUM(number1, [number2], ...)	
18	Analysis

- Move or adjust blue box as needed (pull handle)
- Hit ESC to get out of formula without editing

Functions

- Drag formula across or down to repeat for corresponding cells

=SUM(G2:G5)					
	D	E	F	G	H
Sample dataset:	Sex	Age group	Count A	Count B	Date entered
	male	<60	10	1	1/1/2014
	Male	>60	23	14	5/6/2014
	FEMALE	<60		13	8/6/2011
	fEmale	≥60	18	1	9/8/2012
			51	29	

- To hold a formula or piece of a formula constant add an \$

=SUM(\$F2:\$F5)					
	D	E	F	G	H
Sample dataset:	Sex	Age group	Count A	Count B	Date entered
	male	<60	10	1	1/1/2014
	Male	>60	23	14	5/6/2014
	FEMALE	<60		13	8/6/2011
	fEmale	≥60	18	1	9/8/2012
			51	51	

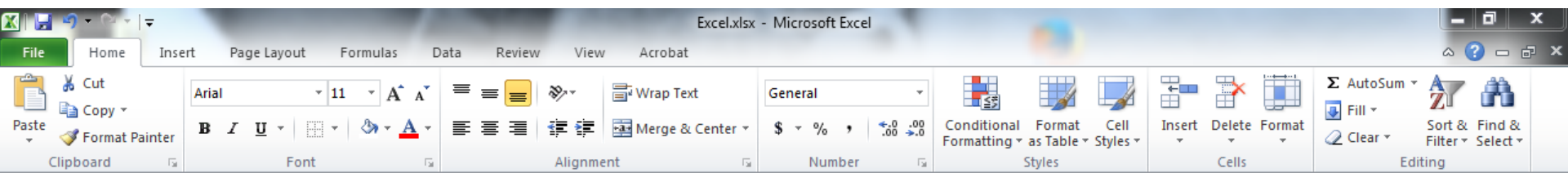
Overview

- Introduction
- Functions
- **Point and Click**
- Excel & CDPH
- Excel & other programs

**See corresponding Excel sheet for examples*

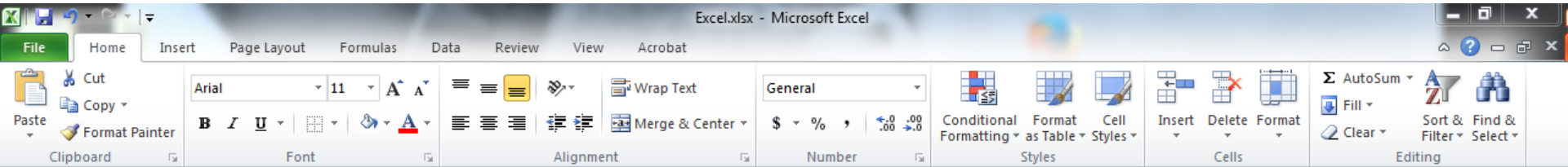
Point and click

- How to use:
 - Use tool bar at the top of the page



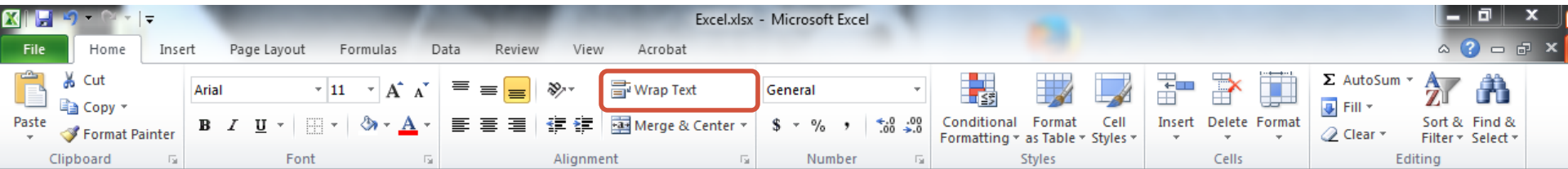
- Useful for data analysis and cleaning
- See Excel sheet for examples

Home



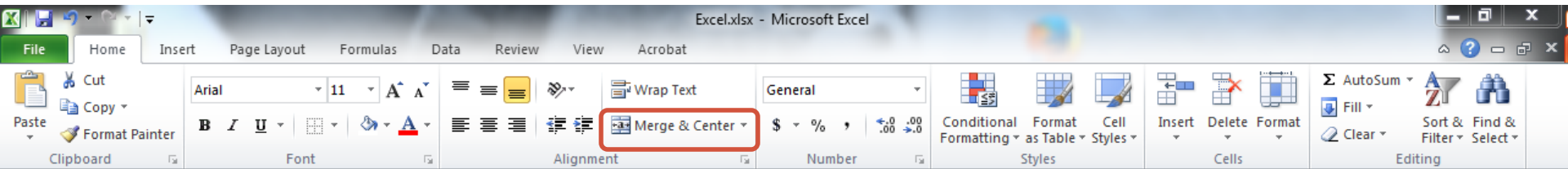
- Wrap Text
- Merge & Center
- Number formatting
- Conditional Formatting
- Format as Table

Home > Wrap Text



Sample datasets (all fake data):			
ID	Name 1	name 2	F
0001	Roberts, James		
0002		Jill Smith	
0003			
0004			
0005	Chase, Drew		
1005	Bay, Olviia		
1006			
1007		Jennifer Harvey	
2008			
0001		James Roberts	

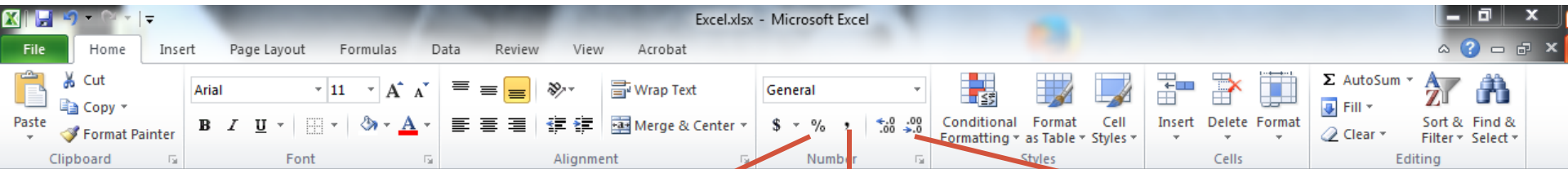
Home > Merge & Center



Sample datasets (all fake data):

ID	Name 1	name 2	Fir
0001	Roberts, James		
0002		Jill Smith	
0003			D
0004			
0005	Chase, Drew		
1005	Bay, Olviia		
1006			I
1007	Jennifer Harvey		J
2008			(
0001		James Roberts	

Home > Number Formatting



Year	Cases	Population
2000	25	100000000%
2001	65	1110000
2002	85	1250000
2003	96	1380000
2004	56	1460000
2005	75	1590000
2006	99	1680000
2007	105	1790000
2008	106	1820000
2009	111	1930000
2010	125	2050000

Year	Cases	Population
2000	25	1,000,000.00
2001	65	1110000
2002	85	1250000
2003	96	1380000
2004	56	1460000
2005	75	1590000
2006	99	1680000
2007	105	1790000
2008	106	1820000
2009	111	1930000
2010	125	2050000

Year	Cases	Population
2000	25	1,000,000
2001	65	1110000
2002	85	1250000
2003	96	1380000
2004	56	1460000
2005	75	1590000
2006	99	1680000
2007	105	1790000
2008	106	1820000
2009	111	1930000
2010	125	2050000

Home > Format as Table

The screenshot shows the Microsoft Excel interface with the 'Format as Table' ribbon selected. The ribbon displays various table styles categorized into 'Light', 'Medium', and 'Dark'. The 'Table Style Light 2' is currently selected. The main worksheet area contains two tables. The first table, titled 'Sample datasets (all fake data):', has columns for ID, Name 1, name 2, First name, Last name, Job title, Sex, Date sick, and Hours per d. The second table, starting at row 21, has columns for Year, Cases, and Populatio. A small black box is visible next to the 'Cases' column header in the second table.

ID	Name 1	name 2	First name	Last name	Job title	Sex	Date sick	Hours per d
0001	Roberts, James				Laborer	M	7/14/2014	8
0002		Jill Smith			Office manager	F	7/15/2014	7.5
0003			Dominic	Duncan	Carpenter	M	7/13/2014	9.25
0004			Kevin	Williams	Electrician	M	7/14/2014	6
0005	Chase, Drew				Electrician	M	7/13/2014	10
1005	Bay, Olviia				Laborer	F	7/13/2014	8
1006			Harold	Rose Jr.	Laborer	M	7/12/2014	6
1007		Jennifer Harve	Jennifer	Harvey	Supervisor	F	7/13/2014	9
2008			Greg H	White	Carpenter	M	7/17/2014	8
0001		James Roberts			Laborer	M	7/14/2014	4

Year	Cases	Populatio
2000	25	1000000
2001	65	1110000
2002	85	1250000
2003	96	1380000
2004	56	1460000
2005	75	1590000
2006	99	1680000
2007	105	1790000
2008	106	1820000
2009	111	1930000
2010	125	2050000

Home > Format as Table

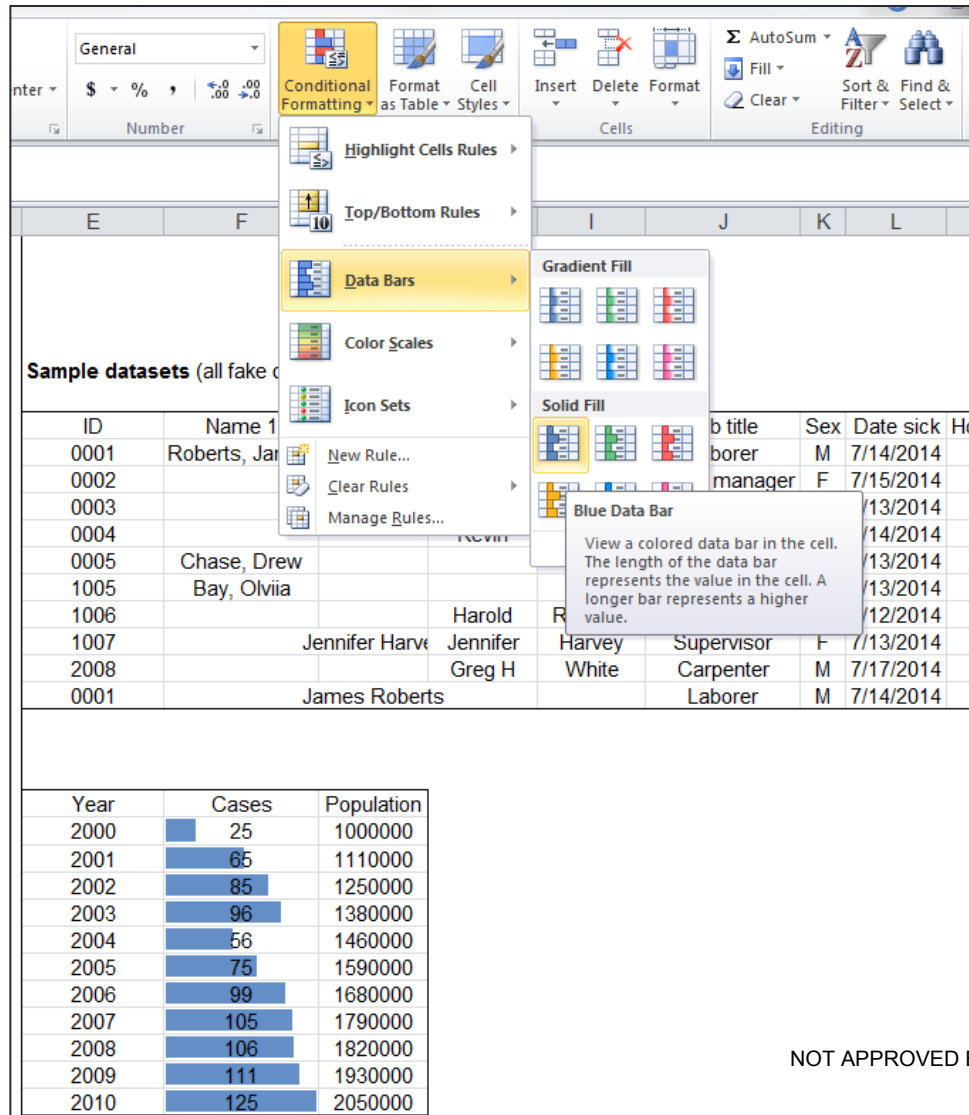
The screenshot shows the Microsoft Excel interface with the 'Format as Table' ribbon selected. The ribbon displays various table styles categorized into 'Light', 'Medium', and 'Dark'. A red arrow points from the text 'Easy sort and filter' to the 'Format as Table' button on the ribbon. Below the ribbon, a data table is visible with columns for Year, Cases, and Population. The table is formatted with a blue header and alternating row colors. The 'Format as Table' ribbon also shows options for 'Conditional Formatting', 'Cell Styles', 'Insert', 'Delete', 'Format', 'AutoSum', 'Fill', 'Clear', 'Sort & Filter', and 'Find & Select'.

Year	Cases	Population
2000	25	1000000
2001	65	1110000
2002	85	1250000
2003	96	1380000
2004	56	1460000
2005	75	1590000
2006	99	1680000
2007	105	1790000
2008	106	1820000
2009	111	1930000
2010	125	2050000

Easy sort and filter

NOT APPROVED BY CA DEPT OF PUBLIC HEALTH

Home > Conditional Formatting



The screenshot shows the Microsoft Excel ribbon with the 'Conditional Formatting' menu open. The menu options include 'Highlight Cells Rules', 'Top/Bottom Rules', 'Data Bars', 'Color Scales', 'Icon Sets', 'New Rule...', 'Clear Rules', and 'Manage Rules...'. The 'Data Bars' option is highlighted, and a tooltip is visible: 'View a colored data bar in the cell. The length of the data bar represents the value in the cell. A longer bar represents a higher value.'

Below the ribbon, there is a table of sample datasets. The table has columns for ID, Name, Job title, Sex, and Date sick. The data is as follows:

ID	Name	Job title	Sex	Date sick
0001	Roberts, Jar	Laborer	M	7/14/2014
0002		manager	F	7/15/2014
0003				7/13/2014
0004				7/14/2014
0005	Chase, Drew			7/13/2014
1005	Bay, Olivia			7/13/2014
1006		Harold		7/12/2014
1007	Jennifer Harv	Jennifer	Supervisor	F 7/13/2014
2008		Greg H	Carpenter	M 7/17/2014
0001	James Roberts	Laborer	M	7/14/2014

Below the sample datasets, there is a table showing data for the years 2000 to 2010, with columns for Year, Cases, and Population. The data is as follows:

Year	Cases	Population
2000	25	1000000
2001	65	1110000
2002	85	1250000
2003	96	1380000
2004	56	1460000
2005	75	1590000
2006	99	1680000
2007	105	1790000
2008	106	1820000
2009	111	1930000
2010	125	2050000

Home > Conditional Formatting

The screenshot shows the Microsoft Excel interface with the 'Home' tab selected. The 'Conditional Formatting' button in the 'Styles' group is highlighted with a red rectangle. The 'Conditional Formatting' dropdown menu is open, showing options like 'Highlight Cells Rules', 'Top/Bottom Rules', 'Data Bars', 'Color Scales', 'Icon Sets', 'New Rule...', 'Clear Rules', and 'Manage Rules...'. The 'Highlight Cells Rules' option is selected, and its submenu is open, showing rules like 'Greater Than...', 'Less Than...', 'Between...', 'Equal To...', 'Text that Contains...', 'A Date Occurring...', and 'Duplicate Values...'. The background shows a spreadsheet with columns A through F and rows 4 through 17. The spreadsheet contains data about 'Sample datasets' and 'Functionality'.

Functionality:	Use:	ID	Name
Home	Data presentation	0001	Roberts, Jan
> Wrap Text	Data presentation	0002	
> Merge & center	Data cleaning and presentation	0003	
> Number	Data cleaning and presentation	0004	
> Conditional Formatting	Data cleaning and presentation	0005	Chase, Dr
> Format as Table	Data cleaning and presentation	1005	Bay, Oliv
Insert		1006	
> Charts	Data analysis and presentation (epi curve)	1007	
> Sparklines	Data analysis and presentation	2008	
Page Layout		0001	James Roberts

- If Format as Table then use Conditional Formatting – can sort by color

Insert



- PivotTable
- Charts
- Sparklines

Insert > PivotTable

The screenshot shows the Microsoft Excel interface with the 'Insert' tab selected. The 'PivotTable' option is highlighted in the ribbon. A tooltip for 'Insert PivotTable' is visible, stating: 'Summarize data using a PivotTable. PivotTables make it easy to arrange and summarize complicated data and drill down on details. Press F1 for more help.'

The data table below is a list of employees with their IDs, names, job titles, and dates.

Line	First name	Last name	Job title	Sex	Date sick	Hours per day
0001	Roberts, James		Laborer	M	7/14/2014	8
0002	Jill Smith		Office manager	F	7/15/2014	7.5
0003	Dominic	Duncan	Carpenter	M	7/13/2014	9.25
0004	Kevin	Williams	Electrician	M	7/14/2014	6
0005	Chase, Drew		Electrician	M	7/13/2014	10
1005	Bay, Olviia		Laborer	F	7/13/2014	8
1006	Harold	Rose Jr.	Laborer	M	7/12/2014	6
1007	Jennifer Harv	Jennifer	Supervisor	F	7/13/2014	9
2008	Greg H	White	Carpenter	M	7/17/2014	8
0001	James Roberts		Laborer	M	7/14/2014	4

Insert > PivotTable

Excel.xlsx - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View Acrobat

PivotTable Tools: Options Design

Options Design

Summarize Values By: Show Values As Fields, Items, & Sets

PivotChart OLAP Tools What-If Analysis

Field List Buttons Field Headers

Insert PivotTable

Summarize data using a PivotTable. PivotTables make it easy to arrange and summarize complicated data and drill down on details. Press F1 for more help.

PivotTable Name: PivotTable1

Active Field: Count of ID

Field Settings Active Field

Group Selection Ungroup Group Field Group

Sort Sort & Filter

Insert Slicer Refresh Change Data Source Data

Clear Select Move PivotTable Actions

Choose fields to add to report:

- ☒ ID
- ☐ Name 1
- ☐ Name 2
- ☐ First name
- ☐ Last name
- ☒ Job title
- ☒ Sex
- ☐ Date sick
- ☐ Hours per day

Drag fields between areas below:

Report Filter Column Labels

Job title

Row Labels Values

Sex Count of ID

Update

Defer Layout Update

Count of ID	Column Labels	Electrician	Eletrician	Laborer	Office manager	Supervisor	Grand Total
2	Carpenter						
1							
3							
7							
2							
1							
1							
4							
1							
1							
10							

NOT APPROVED BY CA DEPT. OF PUBLIC HEALTH

Insert > Chart

Excel.xlsx - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View Acrobat

PivotTable Table Picture Clip Art Shapes SmartArt Screenshot Column Line Pie Bar Area Scatter Other Charts Line Column Win/Loss Slicer Hyperlink Text Box Header & Footer WordArt Signature Line Object Equation Symbol

Tables Illustrations Sparklines Filter Links Text

F48 =COUNTIF(\$

34 Paste values, formats, formulas
 35 Find (Ctrl+F), Find and replace (Ctrl+H)
 36 Pull handle
 37 Use of \$
 38
 39
 40 **Analysis:**
 41 Insert > Pivot tables
 42 Insert > Graphs/Charts
 43 Show two different types of data in a graph (e.g.
 44 Epi curves: [UNC presentation](#)
 45 Home > Conditional formatting >
 46 Insert > Line/Column (Sparkline)
 47
 48 **Presentation:**
 49 View > Freeze panes
 50 View > Split screen

2-D Column

Clustered Column

Compare values across categories by using vertical rectangles.

Use it when the order of categories is not important or for displaying item counts such as a histogram.

Date sick	No.
7/11	0
7/12	1
7/13	4
7/14	3
7/15	1
7/16	0
7/17	1
7/18	0
7/19	0
7/20	0

3-D

Cylindrical

Cone

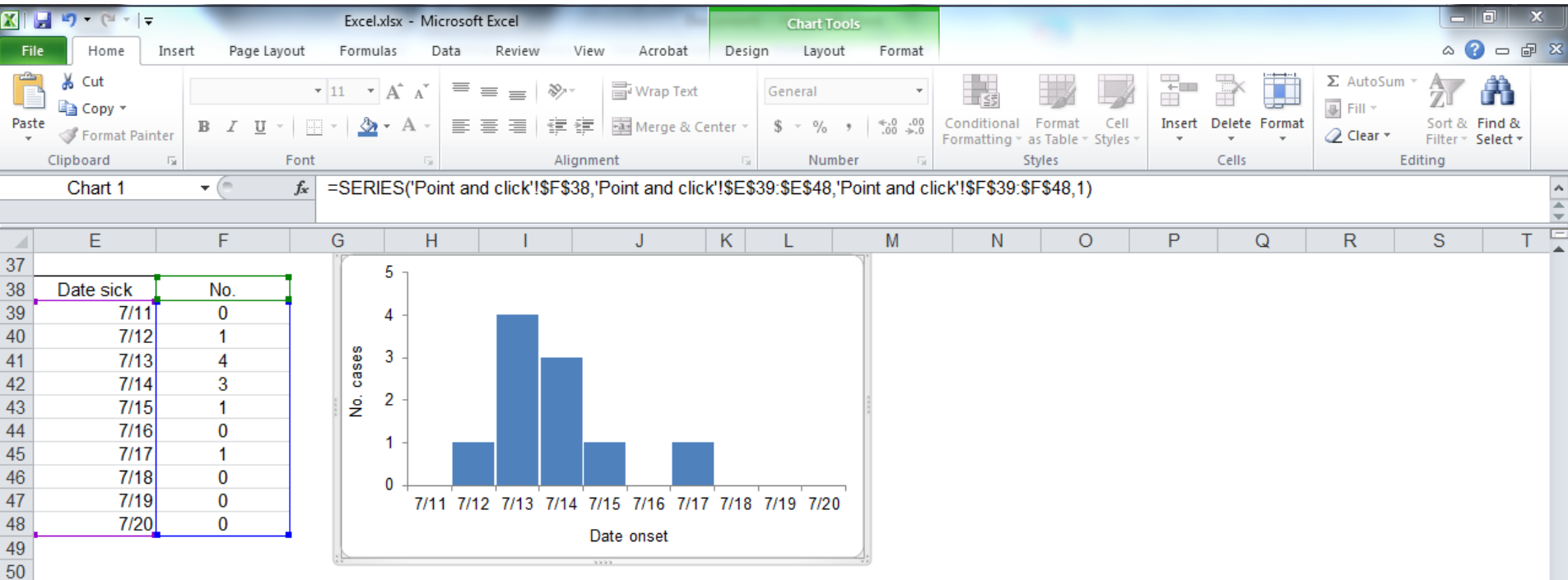
Pyramid

All Chart Types...

No. cases

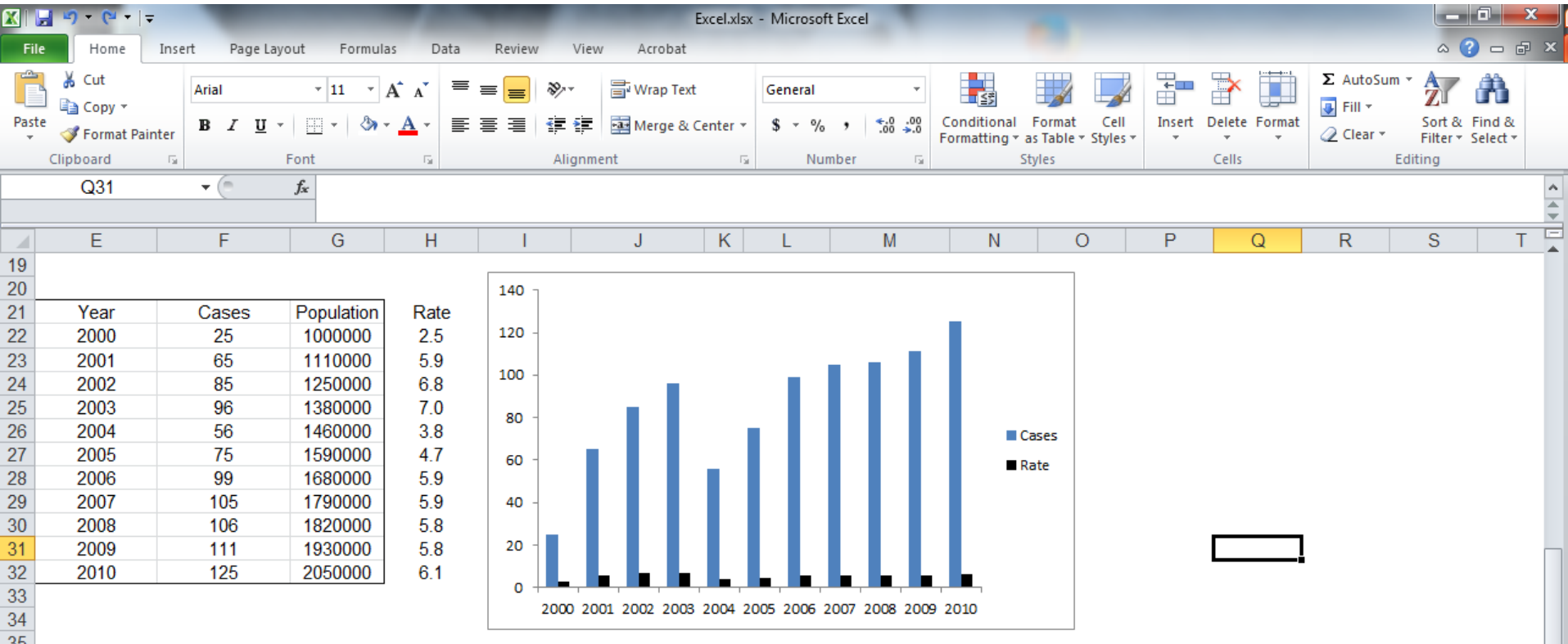
Date onset

Insert > Chart



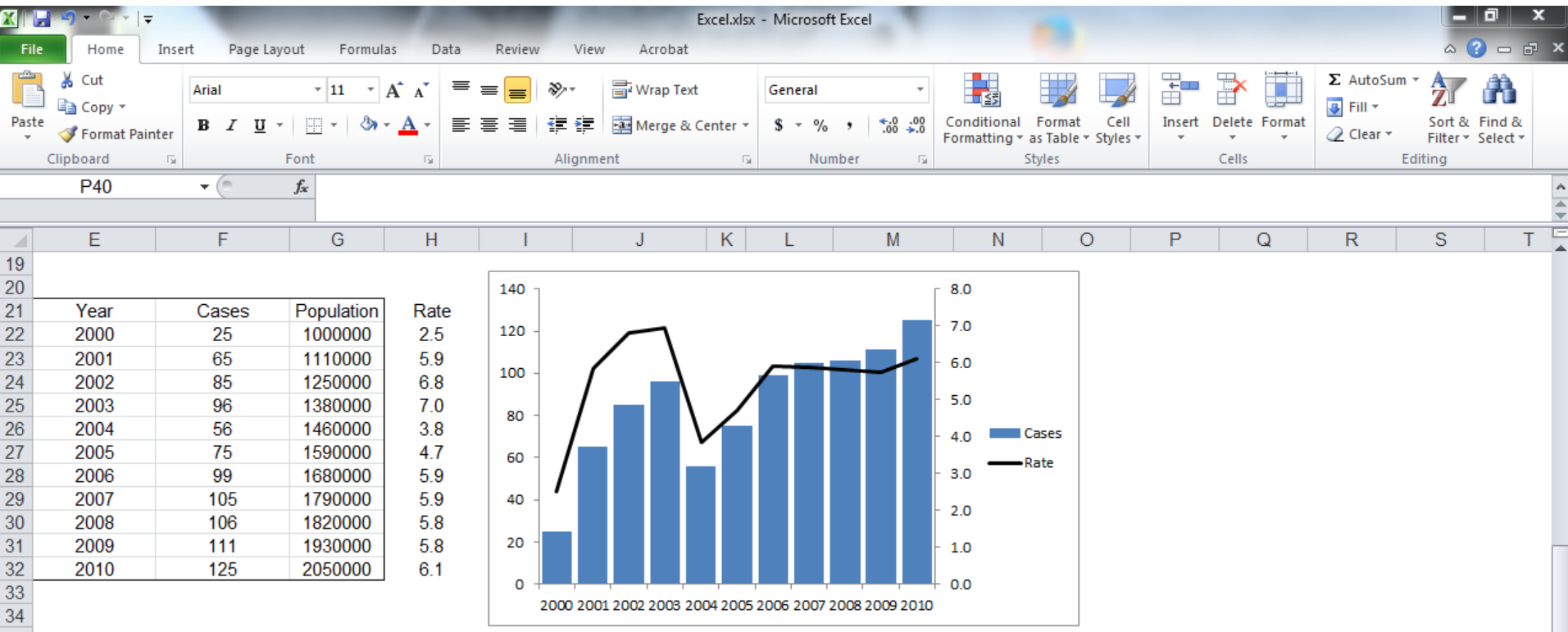
- Click bars/lines to see range of data shown
- Move boxes like with functions

Insert > Numbers and Rates



- Right click the Rates bar > Format Data Series > Series Options > Plot Series on Secondary Axis
- Right click the new Rates bar > Change Series Chart Type > Line

Insert > Numbers and Rates



- Right click the Rates bar > Format Data Series > Series Options > Plot Series on Secondary Axis
- Right click the new Rates bar > Change Series Chart Type > Line

Insert > Sparklines

Excel.xlsx - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View Acrobat

PivotTable Table Picture Clip Art Shapes SmartArt Screenshot Column Line Pie Bar Area Scatter Other Charts Line Column Win/Loss Slicer Hyperlink Text Box Header & Footer WordArt Signature Line Object Equation Symbol

Tables Illustrations Charts Sparklines Filter Links Text

K54

County	2012	2013	2014	2015
Alameda	2	5	7	8
Contra Costa	3	1	1	0
Marin	4	4	5	4
San Francisco	1	2	3	5
Santa Clara	30	25	10	11
San Joaquin	3	7	9	11

SPARKLINES

Create Sparklines

Choose the data that you want

Data Range:

Choose where you want the sparklines to be placed

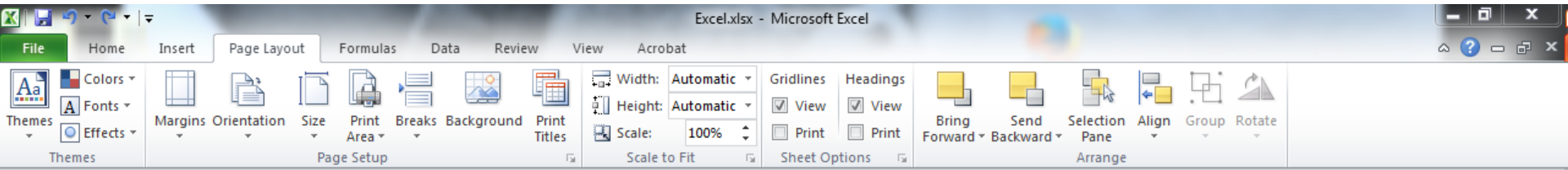
Location Range:

OK Cancel

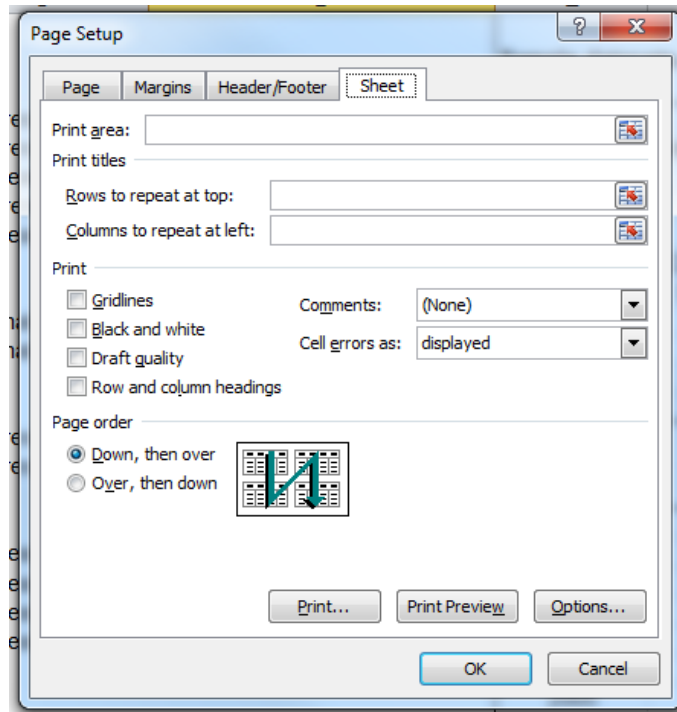
Legend:

- Alameda
- Contra Costa
- Marin
- San Joaquin
- San Francisco
- Santa Clara

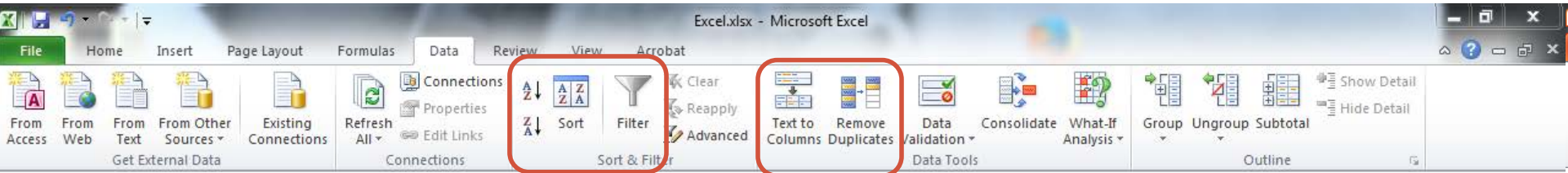
Page Layout



- Print Titles
- Breaks



Data



- Sort
- Filter
- Remove Duplicates
- Text to Columns

Data > Text to Columns

Excel.xlsx - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View Acrobat

From Access From Web From Text From Other Sources Get External Data Existing Connections Refresh All Edit Links Connections Sort Filter Clear Reapply Advanced Text to Columns Remove Duplicates Data Validation Consolidate What-If Analysis Group Ungroup Subtotal Show Detail Hide Detail Outline

F8 Roberts, James

Convert Text to Columns Wizard - Step 1 of 3

The Text Wizard has determined that your data is Delimited.
If this is correct, choose Next, or choose the data type that best describes your data.

Original data type

Choose the file type that best describes your data:

☒ Delimited - Characters such as commas or tabs separate each field.

☐ Fixed width - Fields are aligned in columns with spaces between each field.

Preview of selected data:

8	Roberts, James
9	
10	
11	
12	Chase, Drew

Cancel < Back Next > Finish

Sample datasets (all fake data):

ID	Name 1	name 2	First name	Last name	Job title	Sex
0001	Roberts, James				Laborer	M
0002		Jill Smith			Office manager	F
0003			Dominic	Duncan	Carpenter	M
0004			Kevin	Williams	Electrician	M
0005	Chase, Drew				Electrician	M
1005	Bay, Olivia				Laborer	F
1006			Harold	Rose Jr.	Laborer	M
1007		Jennifer Harve	Jennifer	Harvey	Supervisor	F
2008			Greg H	White	Carpenter	M
0001		James Roberts			Laborer	M

Year	Cases	Population	Rate
			140

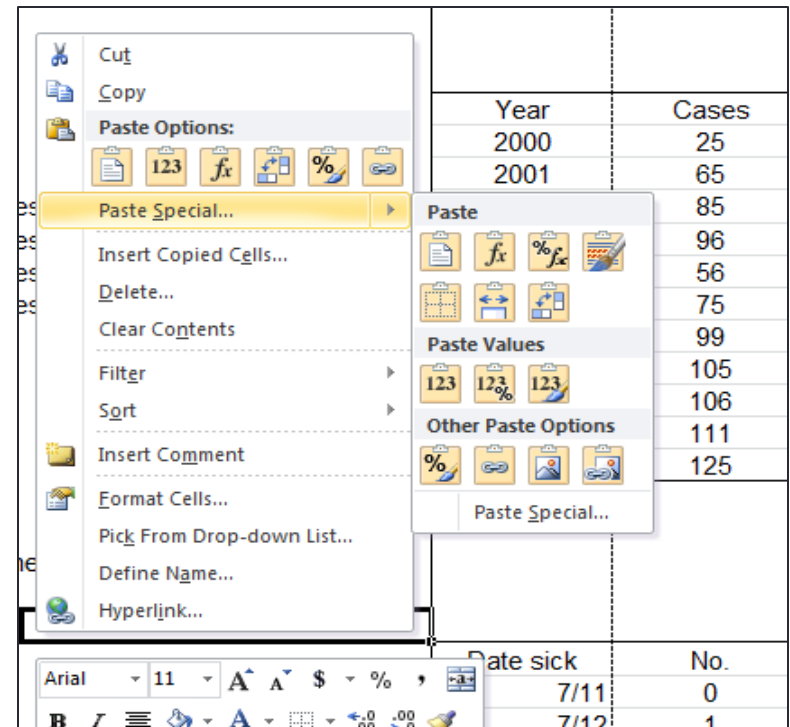
View



- Zoom to Selection
- Freeze Panes
- Split Screen

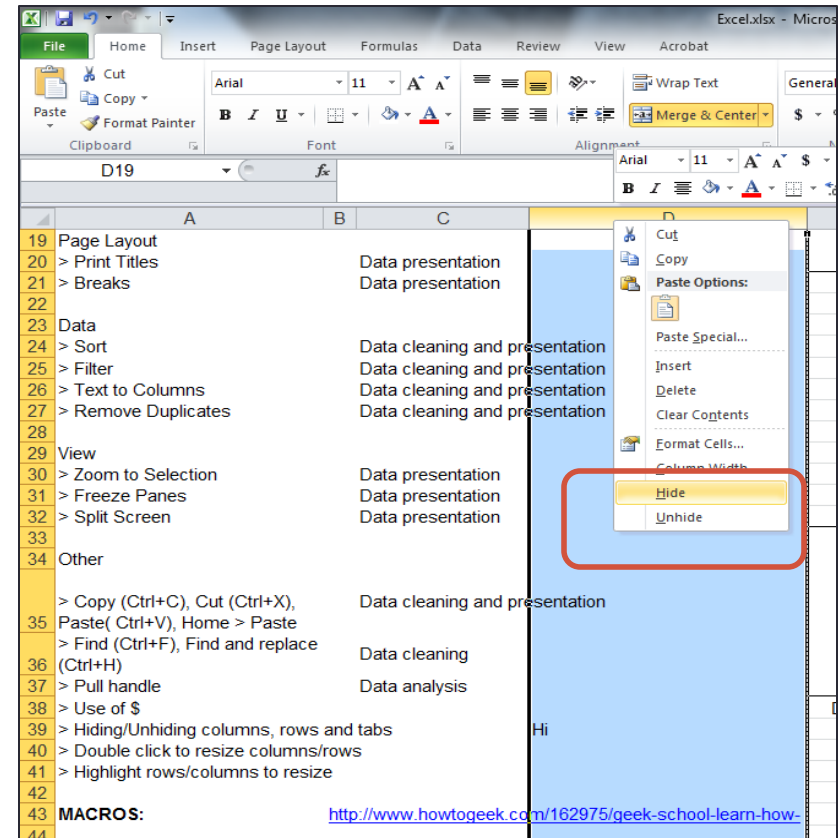
Other notes:

- Copy, Cut & Paste
 - Copy (Ctrl+C), Cut (Ctrl+X), Paste (Ctrl+V)
 - Right click > Paste
 - Paste text, formula, transpose, format, Paste Special
- Find & Replace
 - Find (Ctrl+F), Replace (Ctrl+H)
- Ctrl+Shift+Arrows
 - Use to quickly highlight rows and/or columns in table
- Ctrl+A
 - Use to quickly highlight entire table



Other notes:

- Hiding/Unhiding
 - Right click column, row, or tab
- Double click to resize columns or rows
- Highlight rows or columns to resize similarly
- Macros: <http://www.howtogeek.com/162975/geek-school-learn-how-to-use-excel-macros-to-automate-tedious-tasks/>



Overview

- Introduction
- Functions
- Point and Click
- **Excel & CDPH**
- Excel & other programs

**See corresponding Excel sheet for examples*

Excel and CDPH

Excel and Epidemiology

Resources made available and used by CDPH staff

<http://www.cdph.ca.gov/pubsforms/forms/Pages/CD-Report-Forms.aspx>

Case-Control Study	Excel (1/12)	Table for calculating odds ratios, 95% confidence intervals, and chi-square p-values Food-Specific Attack Rate Table
Cohort Study	Excel (1/12)	Table for calculating risk ratios, 95% confidence intervals, and chi-square p-values

<https://public.health.oregon.gov/DiseasesConditions/CommunicableDisease/LocalHealthDepartments/Gastroenteritis/Pages/Outbreak-Investigation-Tools.aspx#binomial>

Binomial Probability Worksheet

- What if 5 of 9 cases report consumption of alfalfa sprouts? Is that "significant"?
- This Excel worksheet allows you to calculate the binomial probability of getting X or more yes answers to a given exposure question if you know or can guesstimate the background rate.

Overview

- Introduction
- Functions
- Point and Click
- Excel & CDPH
- **Excel & other programs**

**See corresponding Excel sheet for examples*

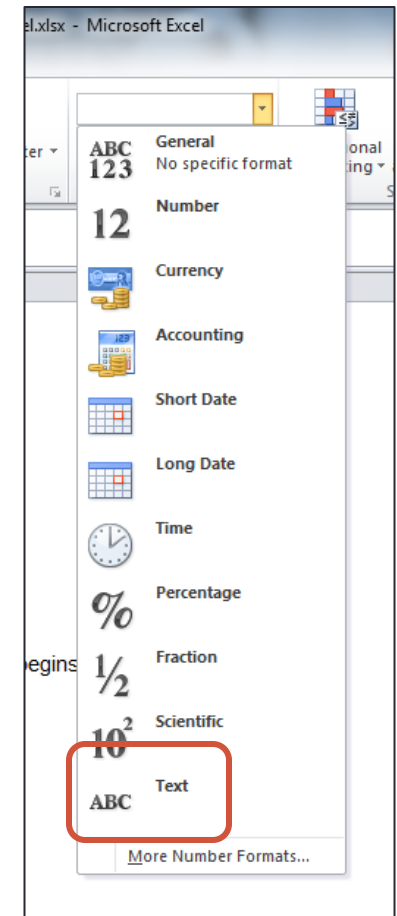
Excel and other programs

- General tips:
 - *Excel not regarded as reliable/powerful analysis tool*
 - Can become corrupted (e.g. “ghost”)
 - Whenever you import/export Excel files:
 - Check leading zeros (especially in IDs)
 - Check date formats
 - Check missing values
 - Check for truncated data
 - There is a limit to how much data Excel can manage
 - Make sure data element names are valid if exporting (e.g. SAS no spaces, begins with letter, etc.)

Excel and other programs

- SAS

- Free SAS programming 1 tutorial includes using SAS with Excel
- Export SAS results as HTML and you can copy and paste to Excel
 - Tools > Options > Results > Create HTML
- May want to highlight entire Excel sheet beforehand and change to Text



Excel and other programs

- SAS Concatenation
 - Using formula with & to concatenate
 - Helpful for quickly writing SAS code to recode data elements

County ID #1	County ID #2				SAS Code	SAS Code
1	11	if county=	then newcounty=	;	=K12&I12&L12&J12&M12	if county=1 then newcounty=11;
2	22	if county=	then newcounty=	;	=K12&I12&L12&J12&M13	if county=2 then newcounty=22;
3	33	if county=	then newcounty=	;	=K12&I12&L12&J12&M14	if county=3 then newcounty=33;
4	44	if county=	then newcounty=	;	=K12&I12&L12&J12&M15	if county=4 then newcounty=44;
5	55	if county=	then newcounty=	;	=K12&I12&L12&J12&M16	if county=5 then newcounty=55;
6	66	if county=	then newcounty=	;	=K12&I12&L12&J12&M17	if county=6 then newcounty=66;
7	77	if county=	then newcounty=	;	=K12&I12&L12&J12&M18	if county=7 then newcounty=77;
8	88	if county=	then newcounty=	;	=K12&I12&L12&J12&M19	if county=8 then newcounty=88;
9	99	if county=	then newcounty=	;	=K12&I12&L12&J12&M20	if county=9 then newcounty=99;

Thanks! Questions?

- Feel free to contact me at:
[*gail.sondermeyer@cdph.ca.gov*](mailto:gail.sondermeyer@cdph.ca.gov)

May change soon to:

[*gail.cooksey@cdph.ca.gov*](mailto:gail.cooksey@cdph.ca.gov)