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

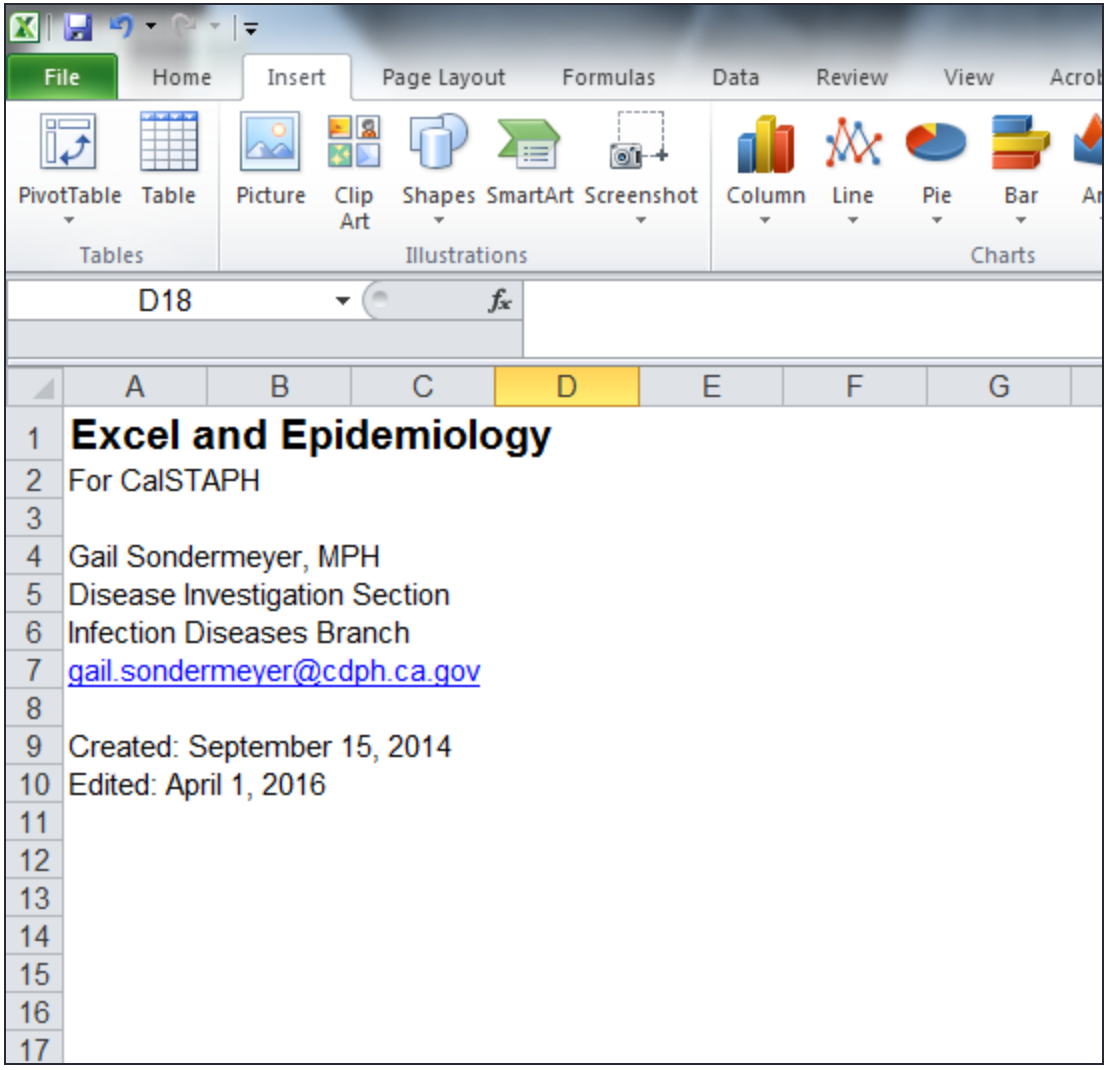
- **Introduction**
- Functions
- Point and Click
- Excel & CDPH
- Excel & other programs

**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



NOT APPROVED BY CA DEPT OF PUBLIC HEALTH

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

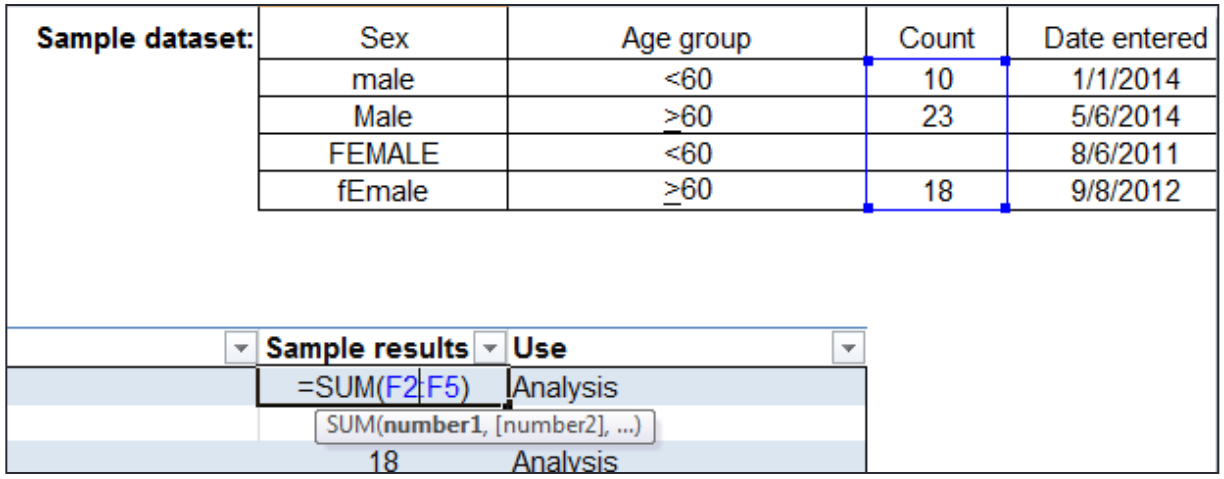
Function name	Formula	Function	Sample results	Use
Sum	=sum(F2:F5)	Add a set of numbers	51	Analysis
Mean	=average(F2:F5)	Calculate mean of a set of numbers	17	Analysis
Median	=median(F2:F5)	Calculate median of a set of numbers	18	Analysis
Minimum	=min(F2:F5)	Find minumum in set of numbers	10	Analysis
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
18	Analysis

SUM(**number1**, [number2], ...)

- 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	

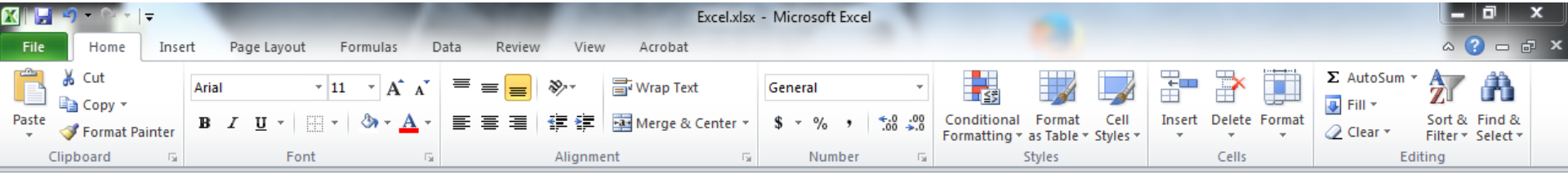
Overview

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

**See corresponding Excel sheet for examples*

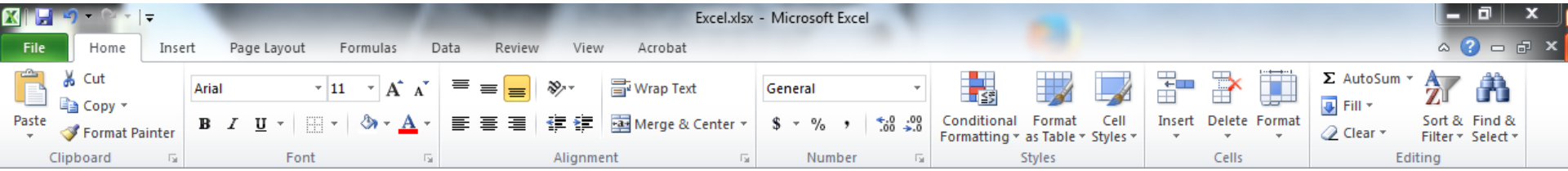
Point and click

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



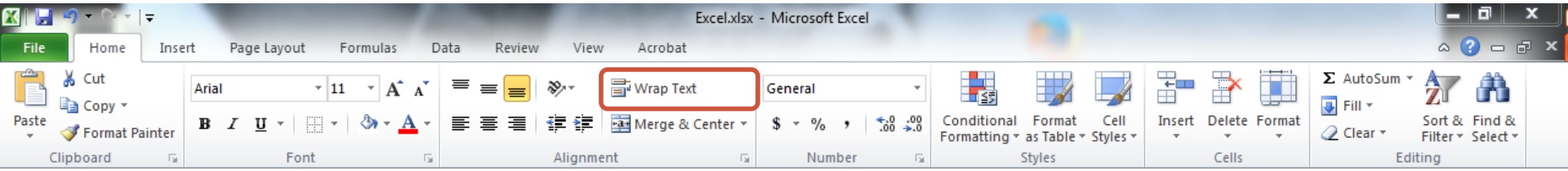
- Useful for data analysis and cleaning
- See Excel sheet for examples

Home



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

Home > Wrap Text

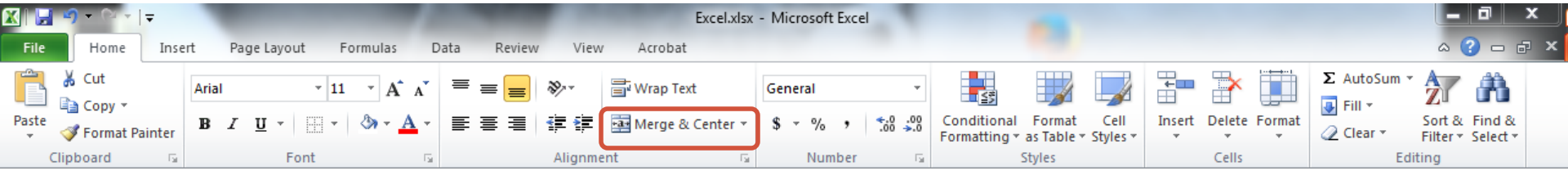


Sample datasets (all fake data):

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

NOT APPROVED BY CA DEPT OF PUBLIC HEALTH

Home > Merge & Center

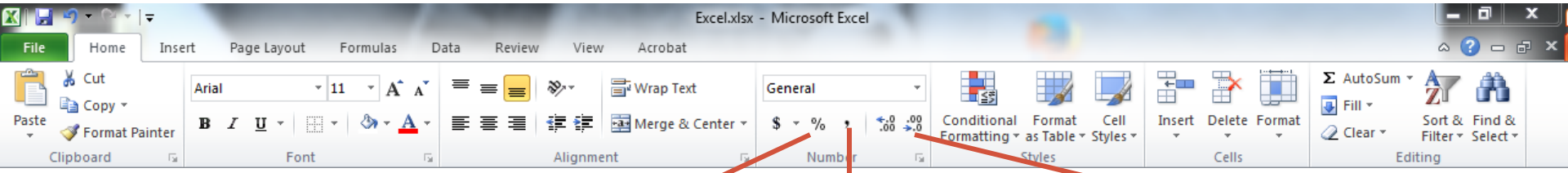


Sample datasets (all fake data):

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

NOT APPROVED BY CA DEPT OF PUBLIC HEALTH

Home > Number Formatting



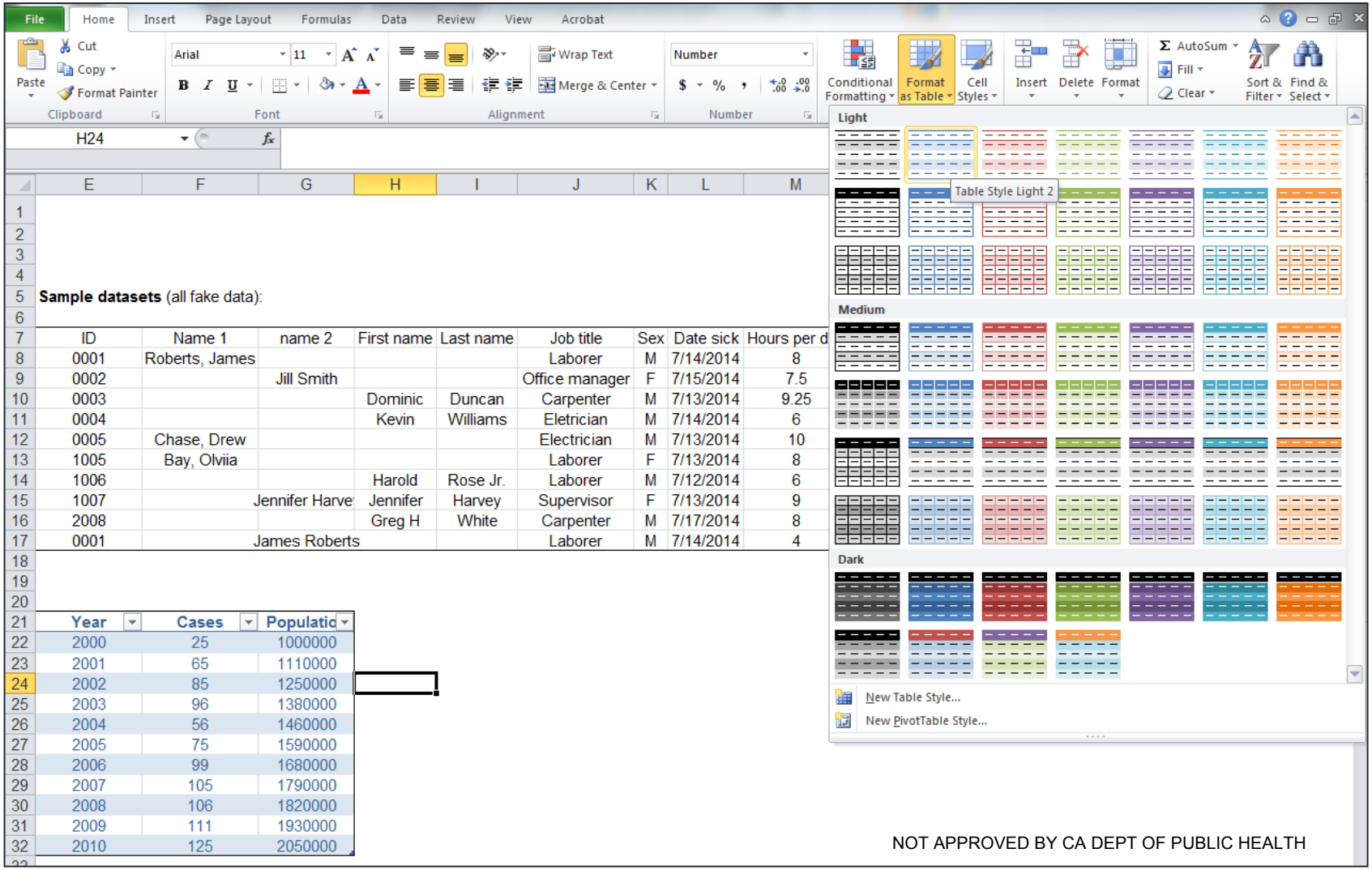
Year	Cases	Population
2000	25	100000000%
2001	65	1110000
2002	85	1250000
2003	96	1380000
2004	56	1460000
2005	75	1590000
2006	99	1680000
2007	105	1790000
2008	106	1820000
2009	111	1930000
2010	125	2050000

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

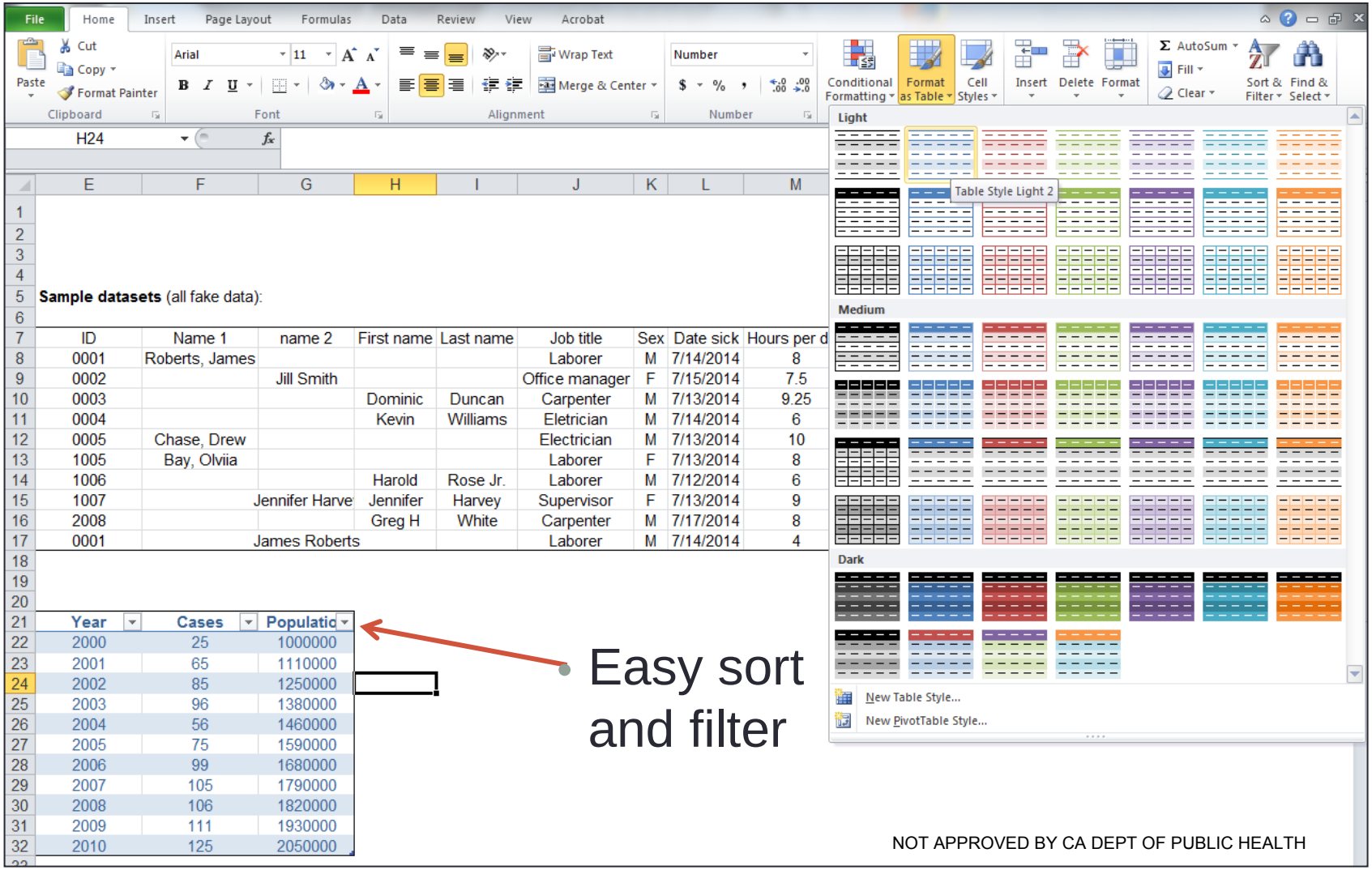
NOT APPROVED BY CA DEPT OF PUBLIC HEALTH

Home > Format as Table



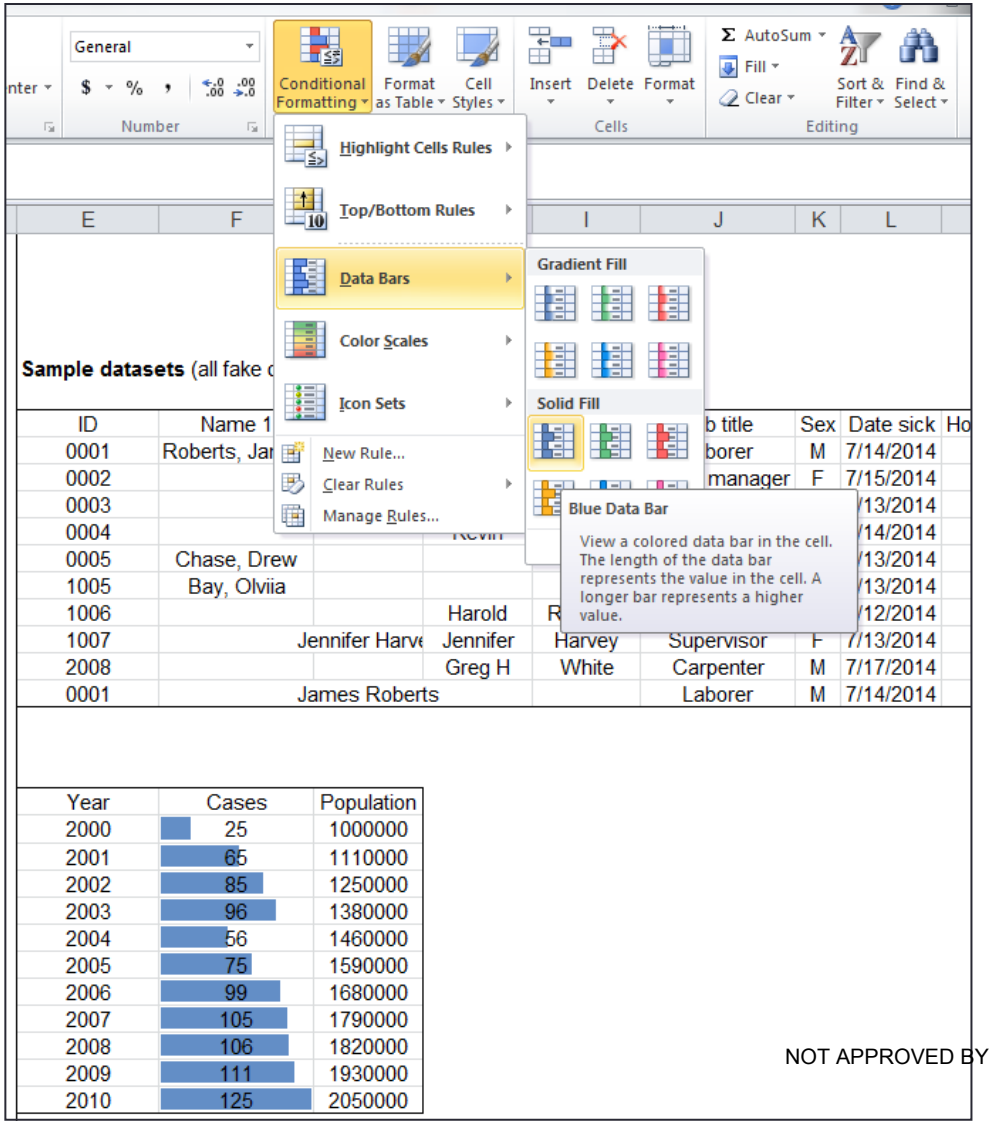
NOT APPROVED BY CA DEPT OF PUBLIC HEALTH

Home > Format as Table



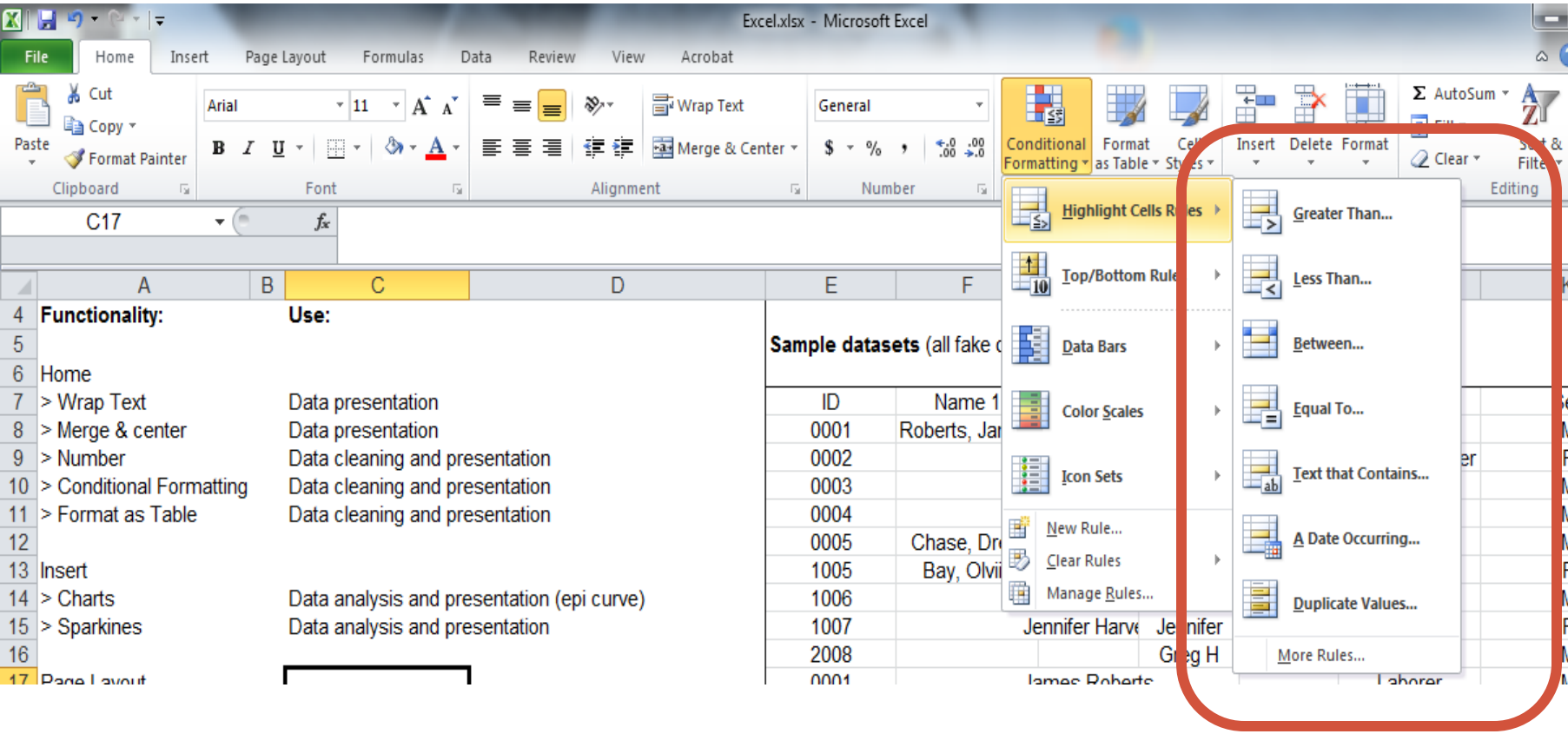
NOT APPROVED BY CA DEPT OF PUBLIC HEALTH

Home > Conditional Formatting



NOT APPROVED BY CA DEPT OF PUBLIC HEALTH

Home > Conditional Formatting

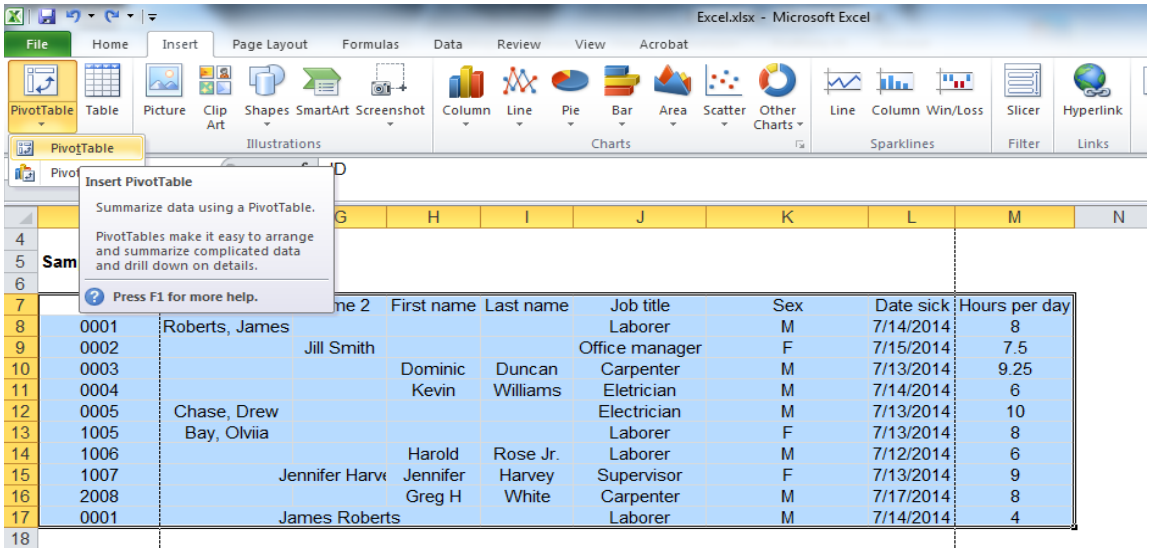


- If Format as Table then use Conditional Formatting – can sort by color

Insert

- PivotTable
- Charts
- Sparklines

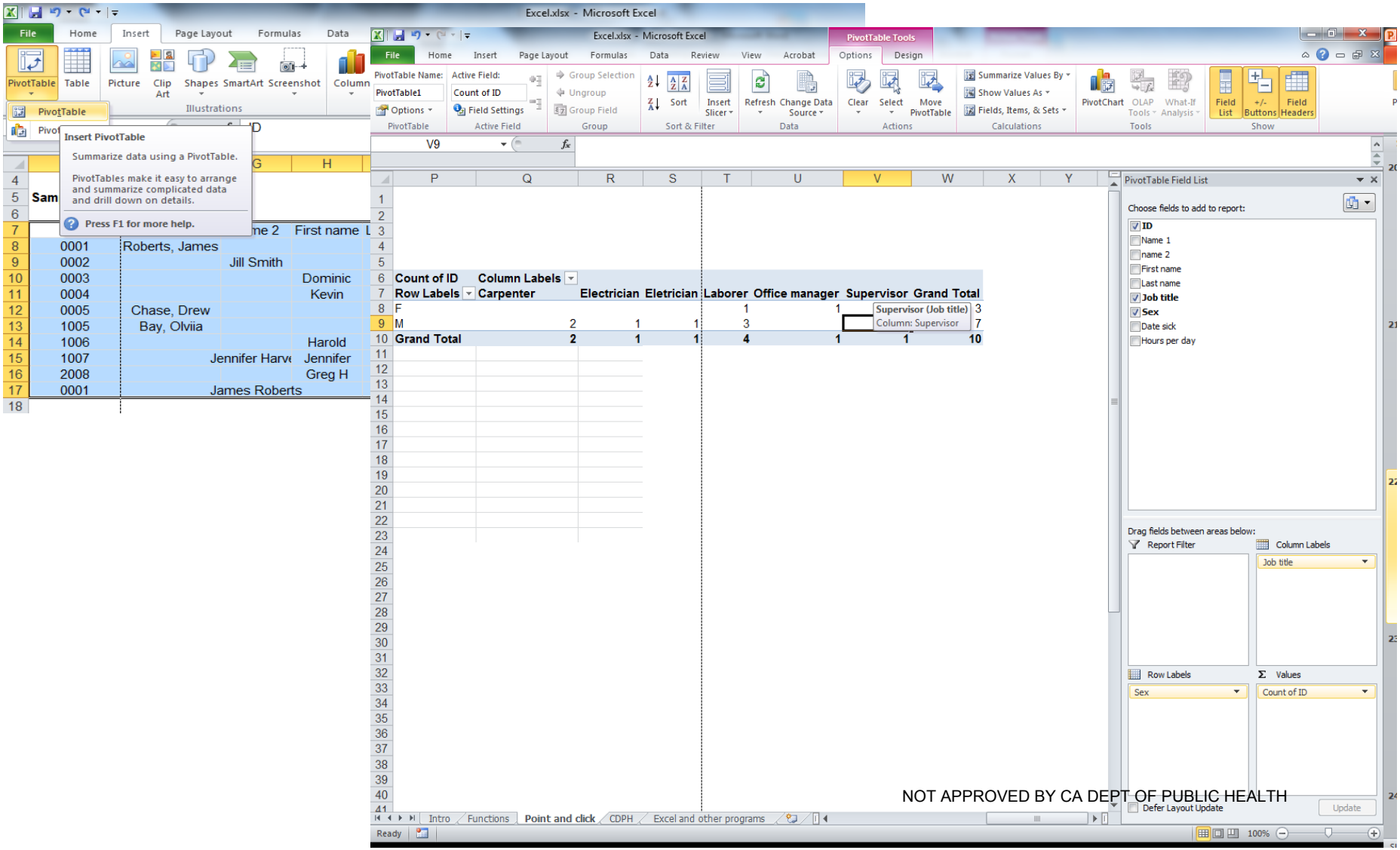
Insert > PivotTable



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	Eletrician	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

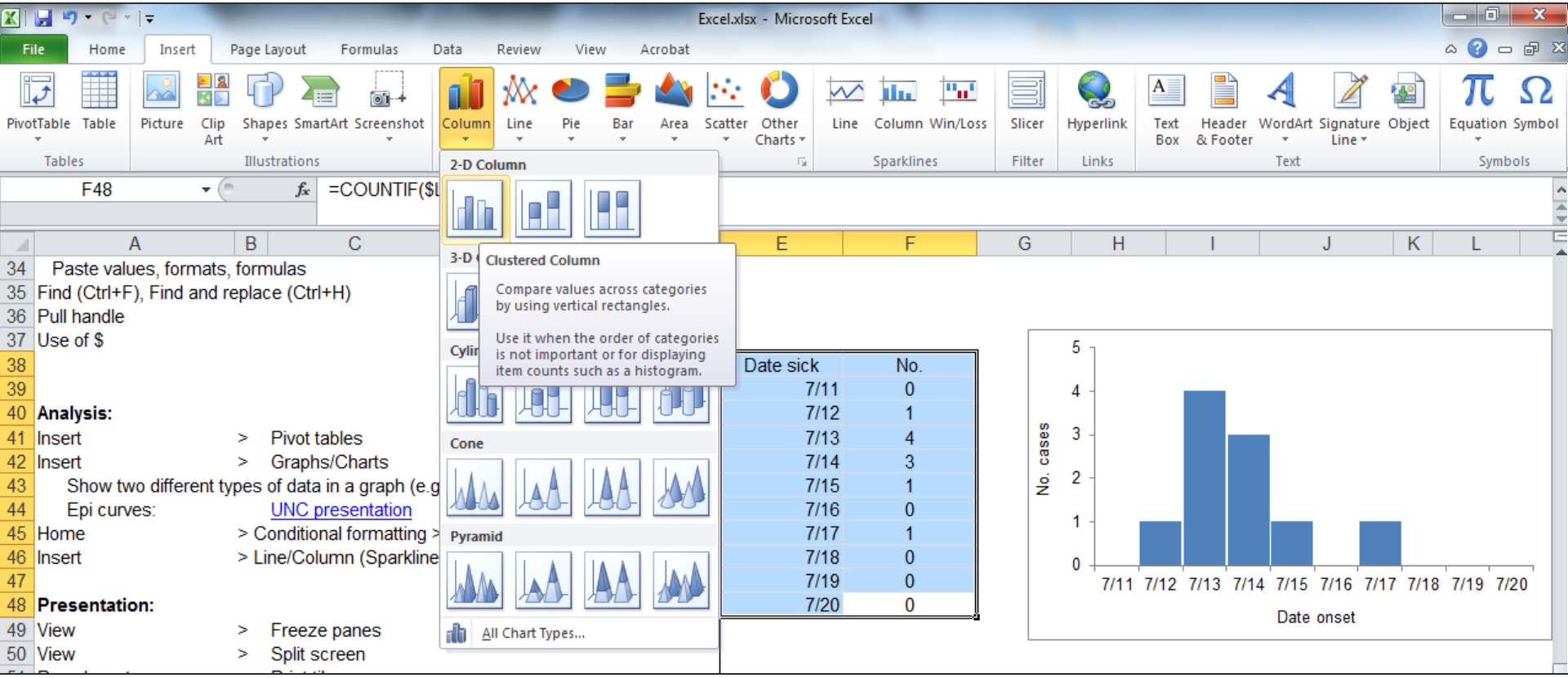
NOT APPROVED BY CA DEPT OF PUBLIC HEALTH

Insert > PivotTable



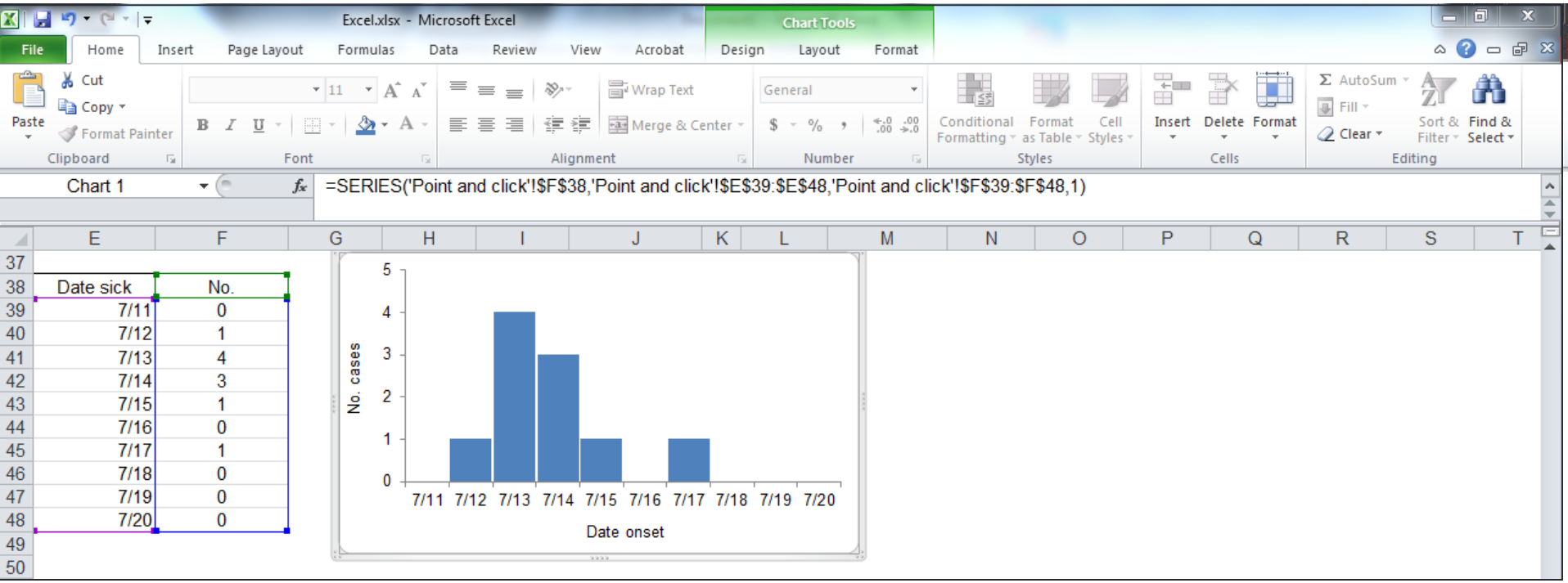
NOT APPROVED BY CA DEPT OF PUBLIC HEALTH

Insert > Chart



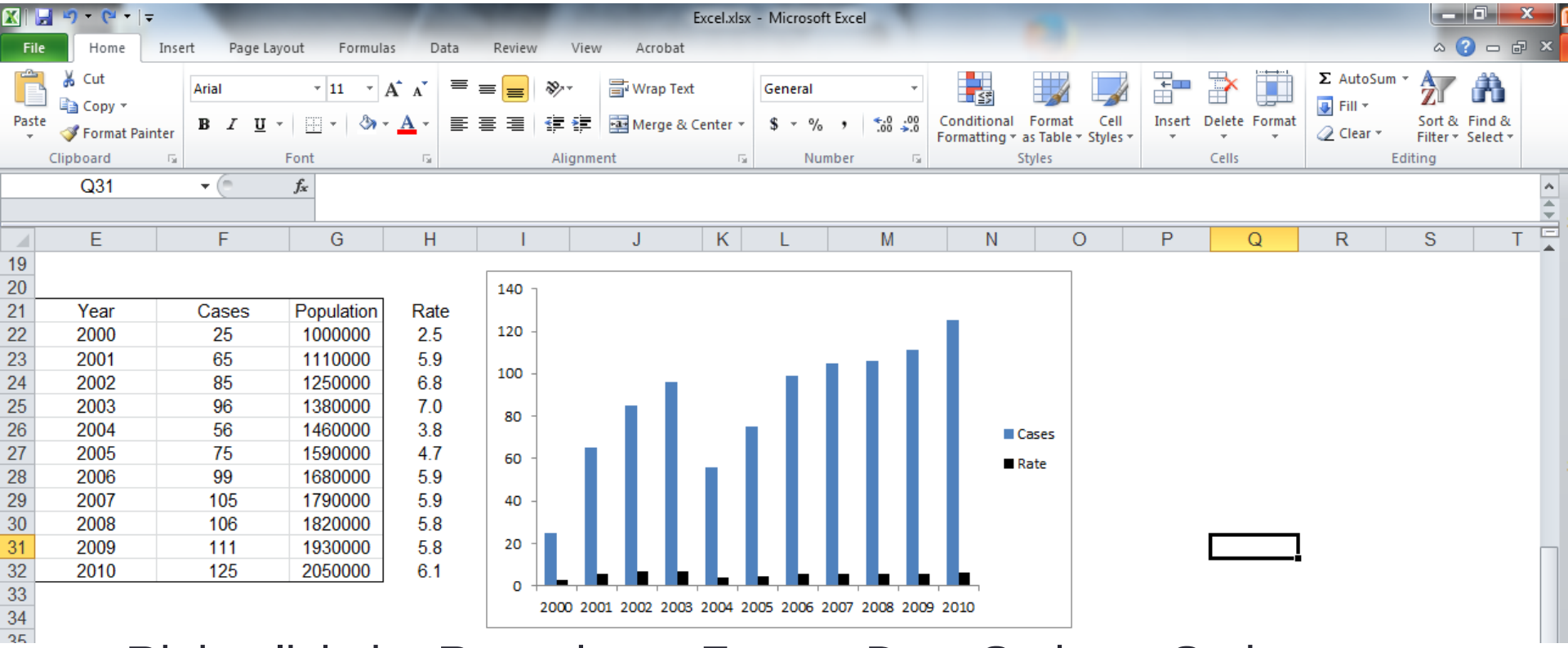
NOT APPROVED BY CA DEPT OF PUBLIC HEALTH

Insert > Chart



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

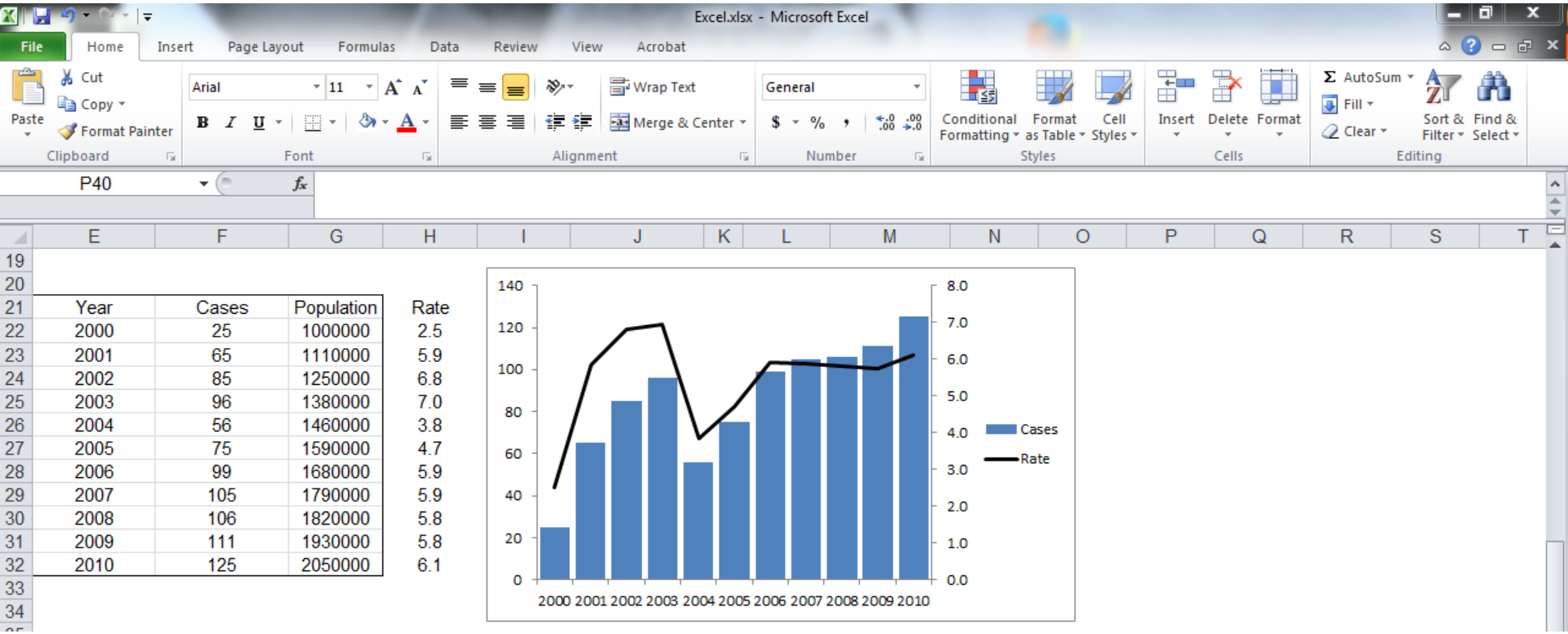
Insert > Numbers and Rates



- 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

NOT APPROVED BY CA DEPT OF PUBLIC HEALTH

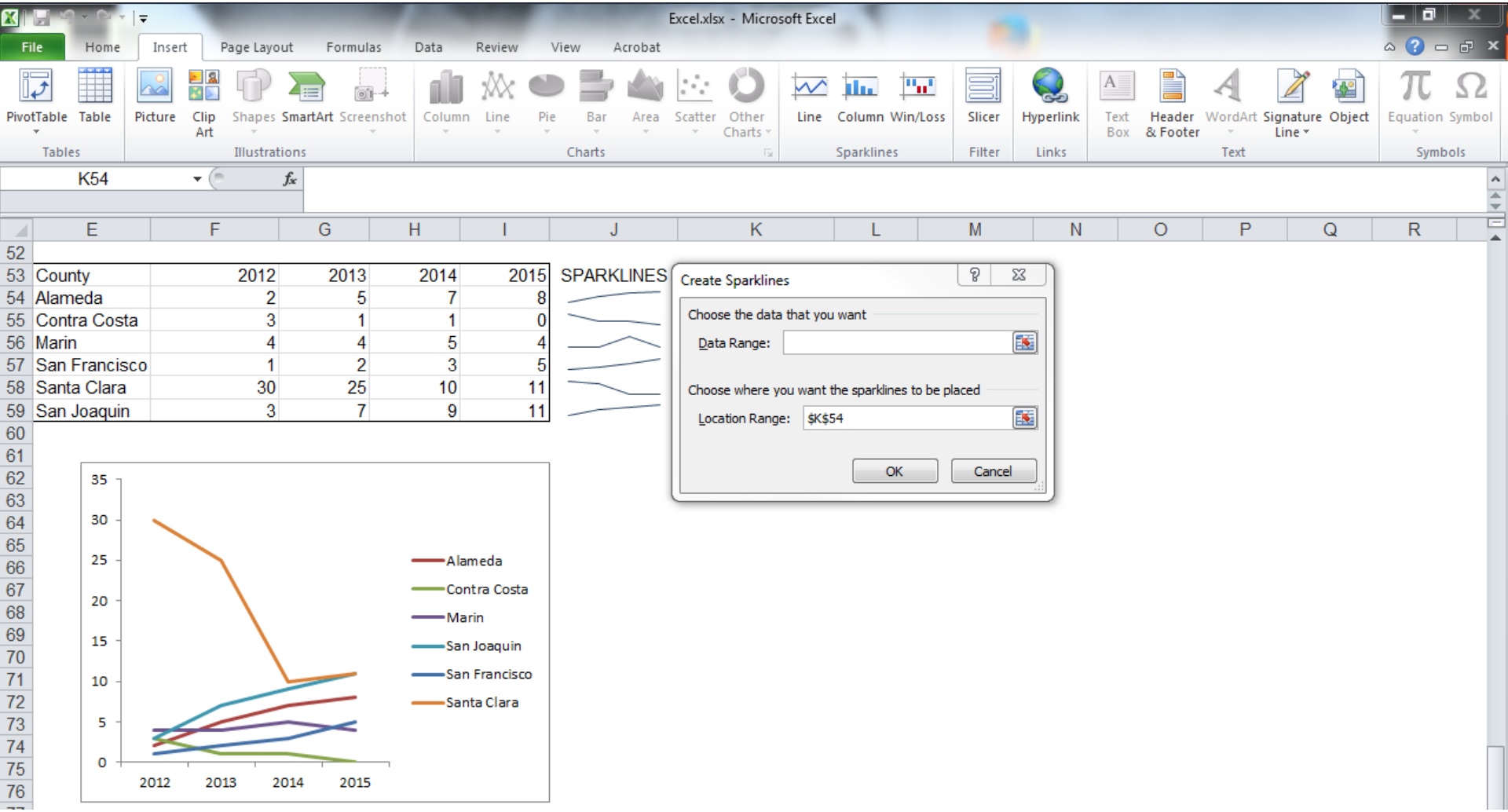
Insert > Numbers and Rates



- 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

NOT APPROVED BY CA DEPT OF PUBLIC HEALTH

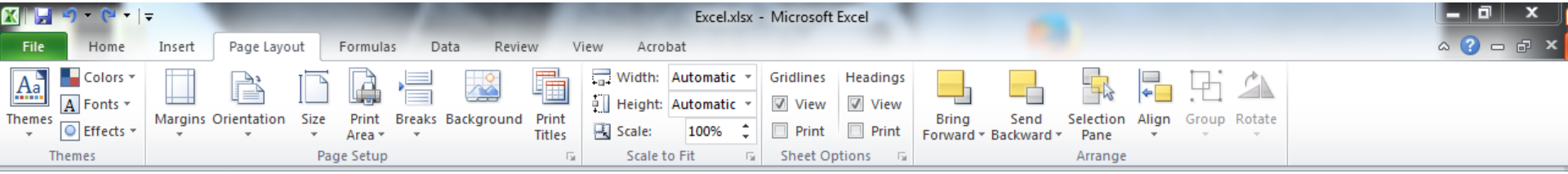
Insert > Sparklines



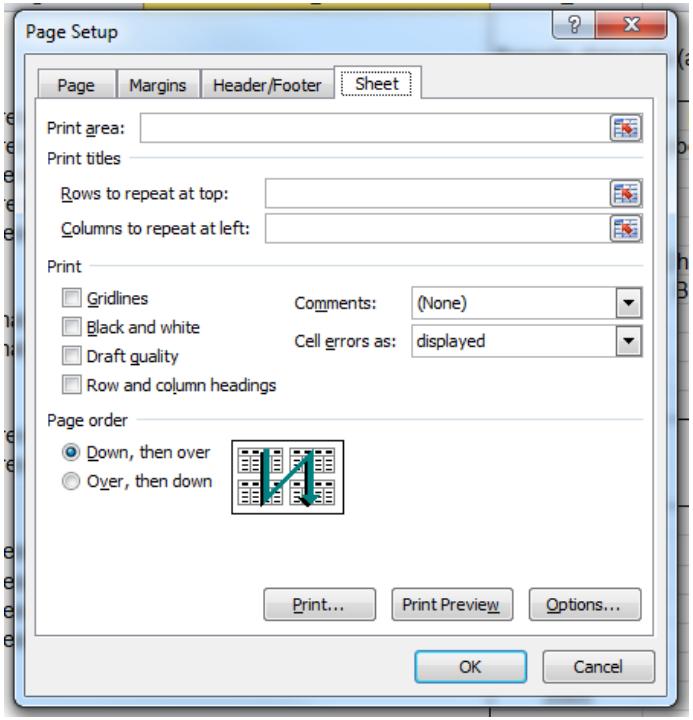
County	2012	2013	2014	2015
Alameda	2	5	7	8
Contra Costa	3	1	1	0
Marin	4	4	5	4
San Francisco	1	2	3	5
Santa Clara	30	25	10	11
San Joaquin	3	7	9	11

NOT APPROVED BY CA DEPT OF PUBLIC HEALTH

Page Layout

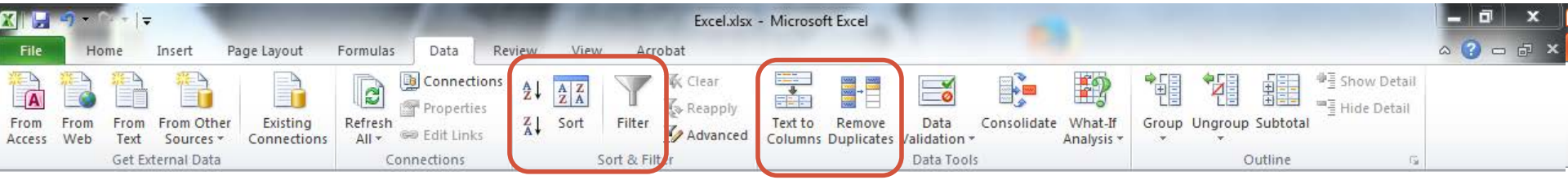


- Print Titles
- Breaks



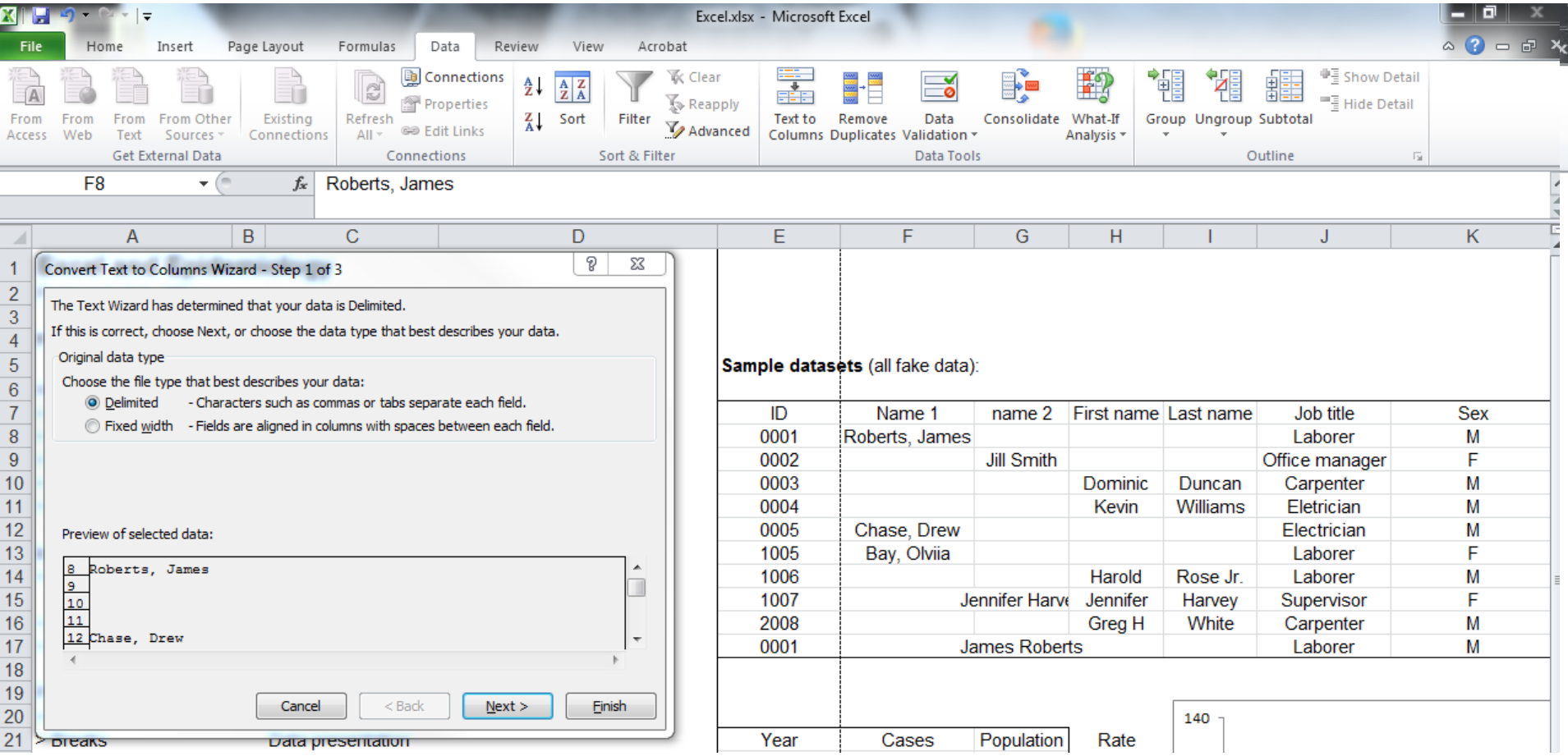
NOT APPROVED BY CA DEPT OF PUBLIC HEALTH

Data



- Sort
- Filter
- Remove Duplicates
- Text to Columns

Data > Text to Columns



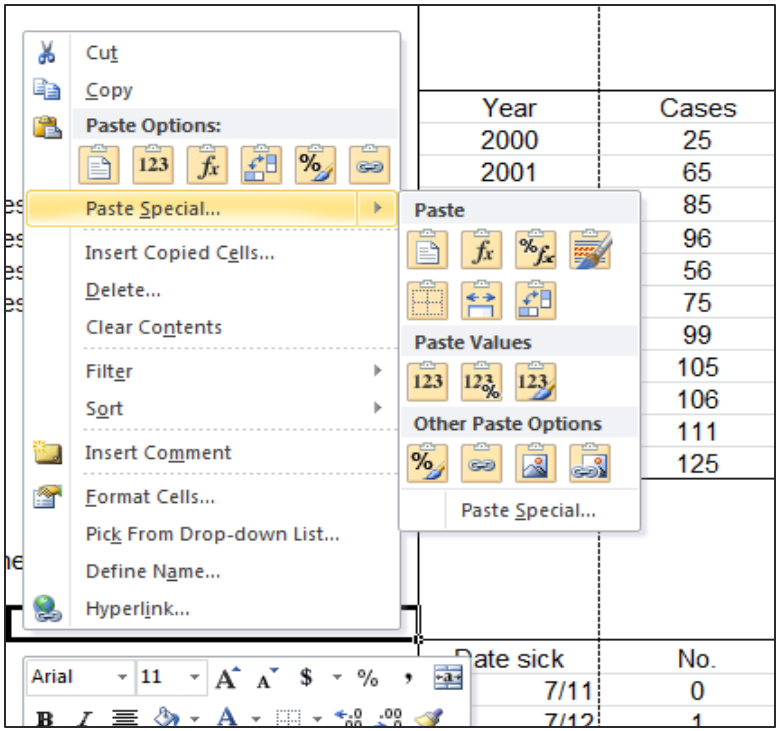
NOT APPROVED BY CA DEPT OF PUBLIC HEALTH

View

- Zoom to Selection
- Freeze Panes
- Split Screen

Other notes:

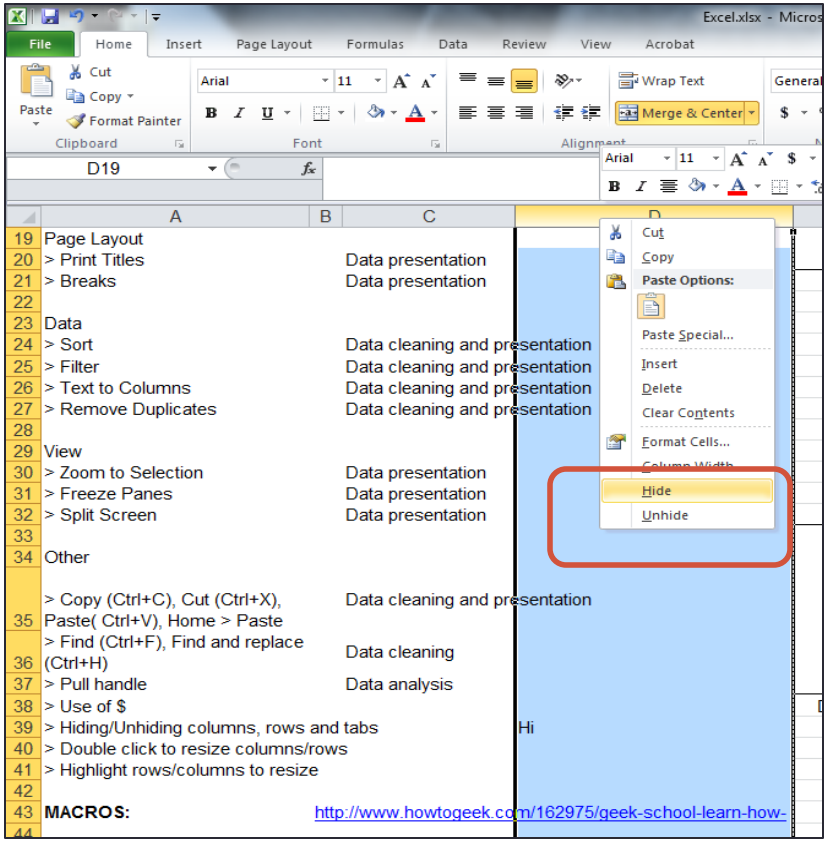
- Copy, Cut & Paste
  - Copy (Ctrl+C), Cut (Ctrl+X), Paste (Ctrl+V)
  - Right click > Paste
    - Paste text, formula, transpose, format, Paste Special
- Find & Replace
  - Find (Ctrl+F), Replace (Ctrl+H)
- Ctrl+Shift+Arrows
  - Use to quickly highlight rows and/or columns in table
- Ctrl+A
  - Use to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



NOT APPROVED BY CA DEPT OF PUBLIC HEALTH

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.

NOT APPROVED BY CA DEPT OF PUBLIC HEALTH

Overview

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

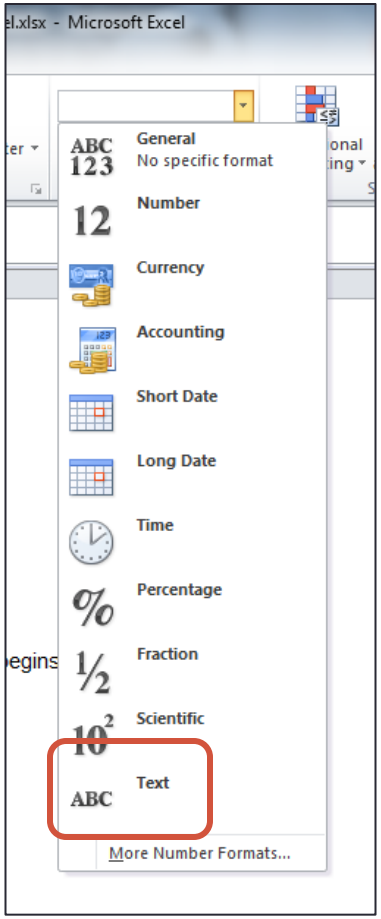
**See corresponding Excel sheet for examples*

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

  May change soon to:

  gail.cooksey@cdph.ca.gov