Beginners Use of Excel for Analysis

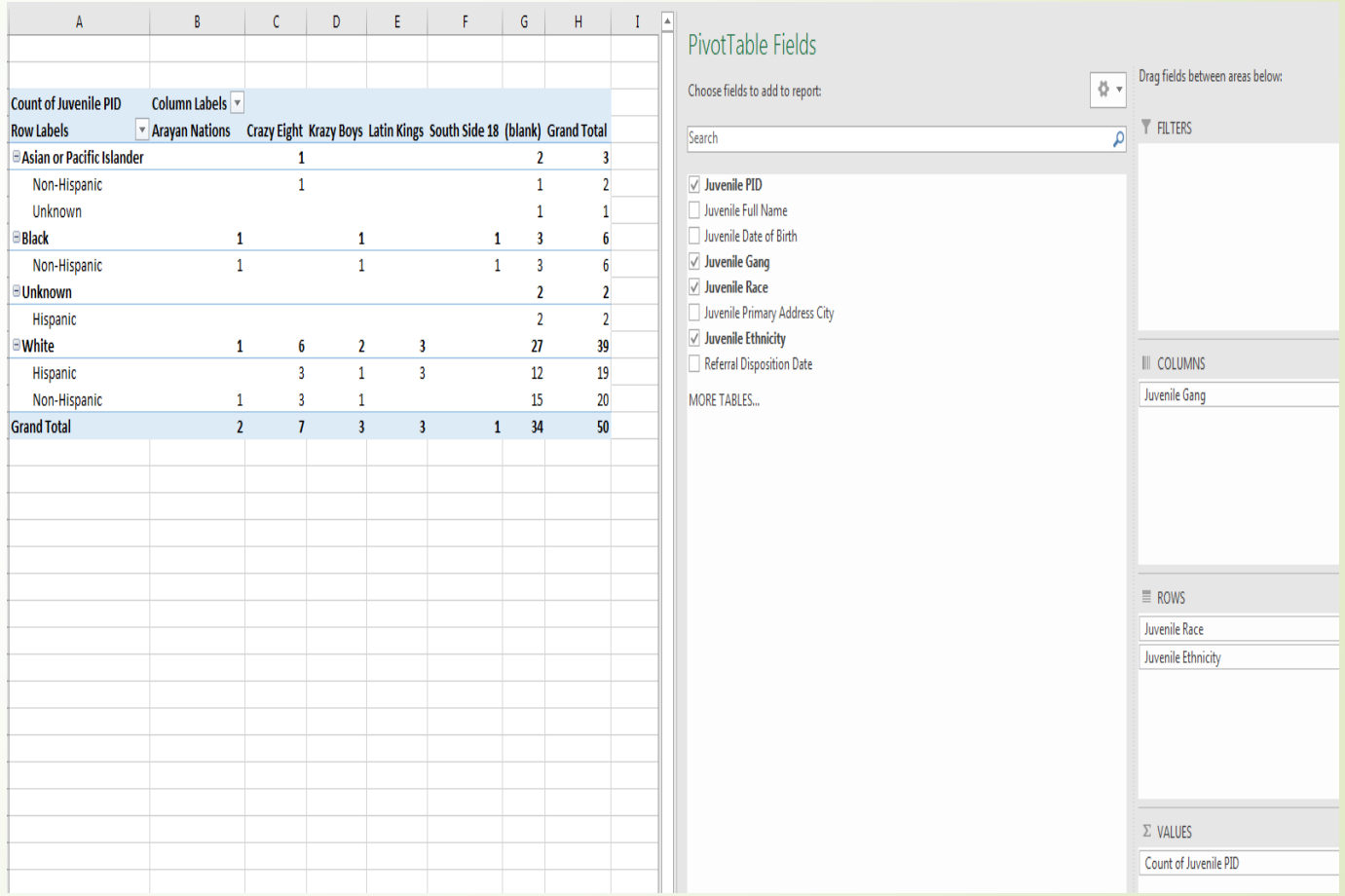
Overview

- Sometimes all you need to is summarize what you have, so we will start with output: summary tables (pivot tables) and graphs/charts. We will also talk about formatting data, sorting data, and what basic information Excel gives you just by highlighting.
- Sometimes the data needs to be processed before it is summarized. We will talk about copying/cutting/pasting, filtering, and functions.
- To show how some of this works together, we will work through a couple of exercises:
  - Build a worksheet to track student attendance in classes.
  - Build a worksheet to calculate Average Daily Population (ADP).
  - Calculate the recidivism rate (primarily the use of VLOOKUP function)
- If you have questions, please ask.

Summarizing Data

Pivot Table

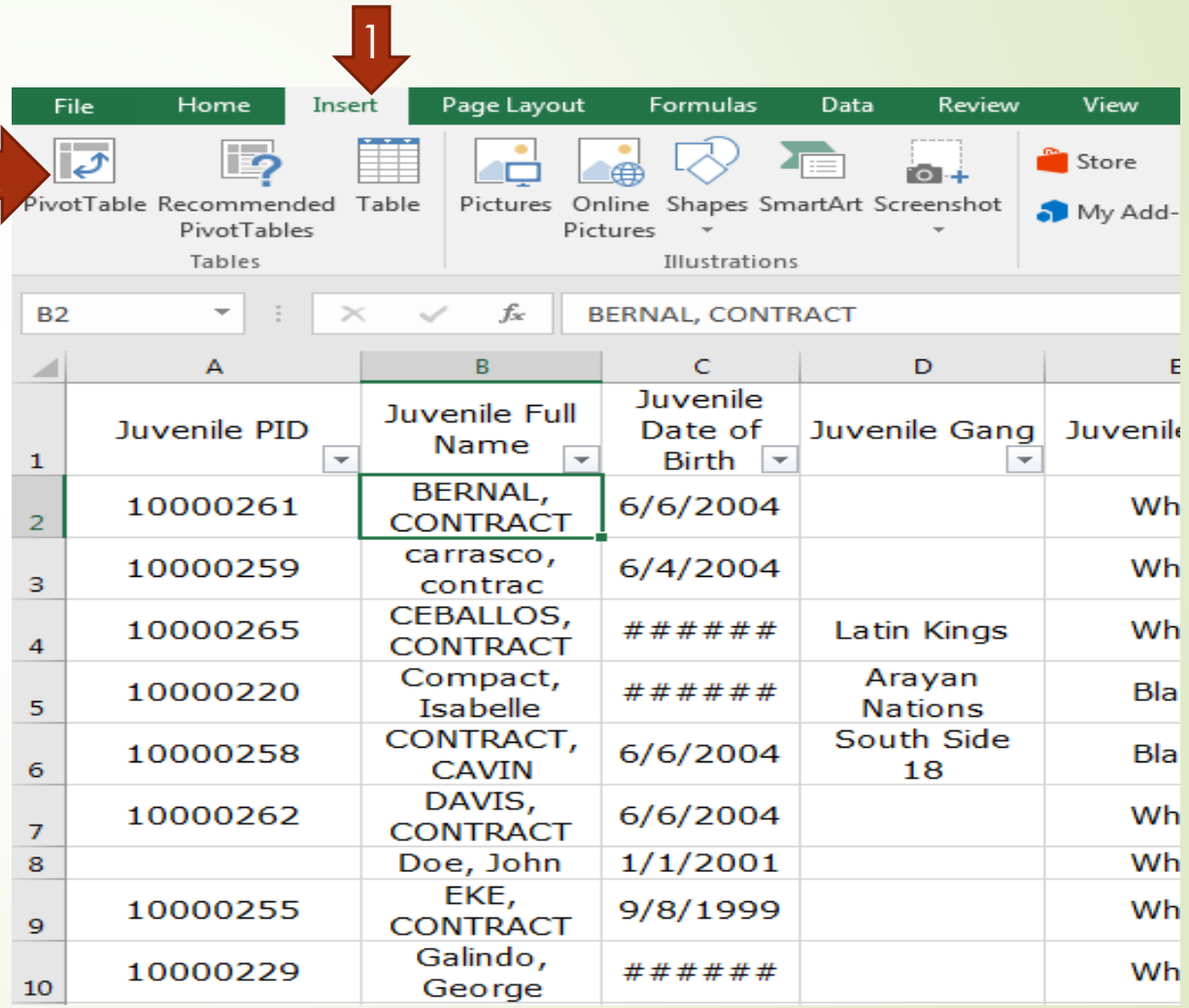
- Pivot tables are a way to get a frequency distribution (summary table) of your data.
- Here is an example of what one looks like.



Count of Juvenile PID	Column Labels						
Row Labels	Arayan Nations	Crazy Eight	Krazy Boys	Latin Kings	South Side 18	(blank)	Grand Total
Asian or Pacific Islander		1				2	3
Non-Hispanic		1				1	2
Unknown						1	1
Black	1		1		1	3	6
Non-Hispanic	1		1		1	3	6
Unknown						2	2
Hispanic						2	2
White	1	6	2	3		27	39
Hispanic		3	1	3		12	19
Non-Hispanic	1	3	1			15	20
Grand Total	2	7	3	3	1	34	50

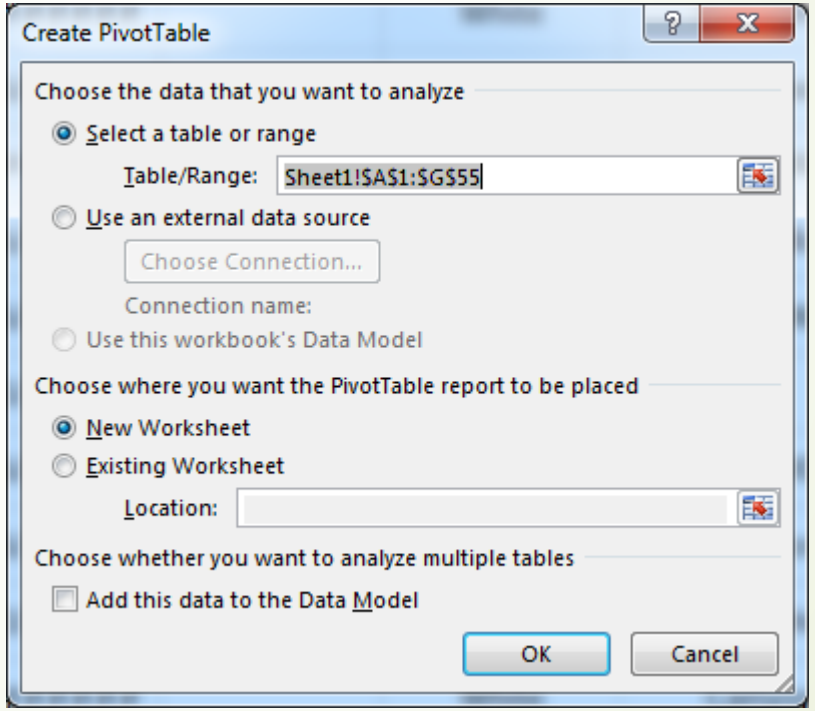
Insert Pivot Table Screenshot 1

- Click on one record that will be part of the pivot table
- Click on Insert on the top left part of the screen (next to Home)
- Pivot tables are in the far left.
- I would check to see if the recommended pivot table has what you want (or most of it). Otherwise click on pivot table.



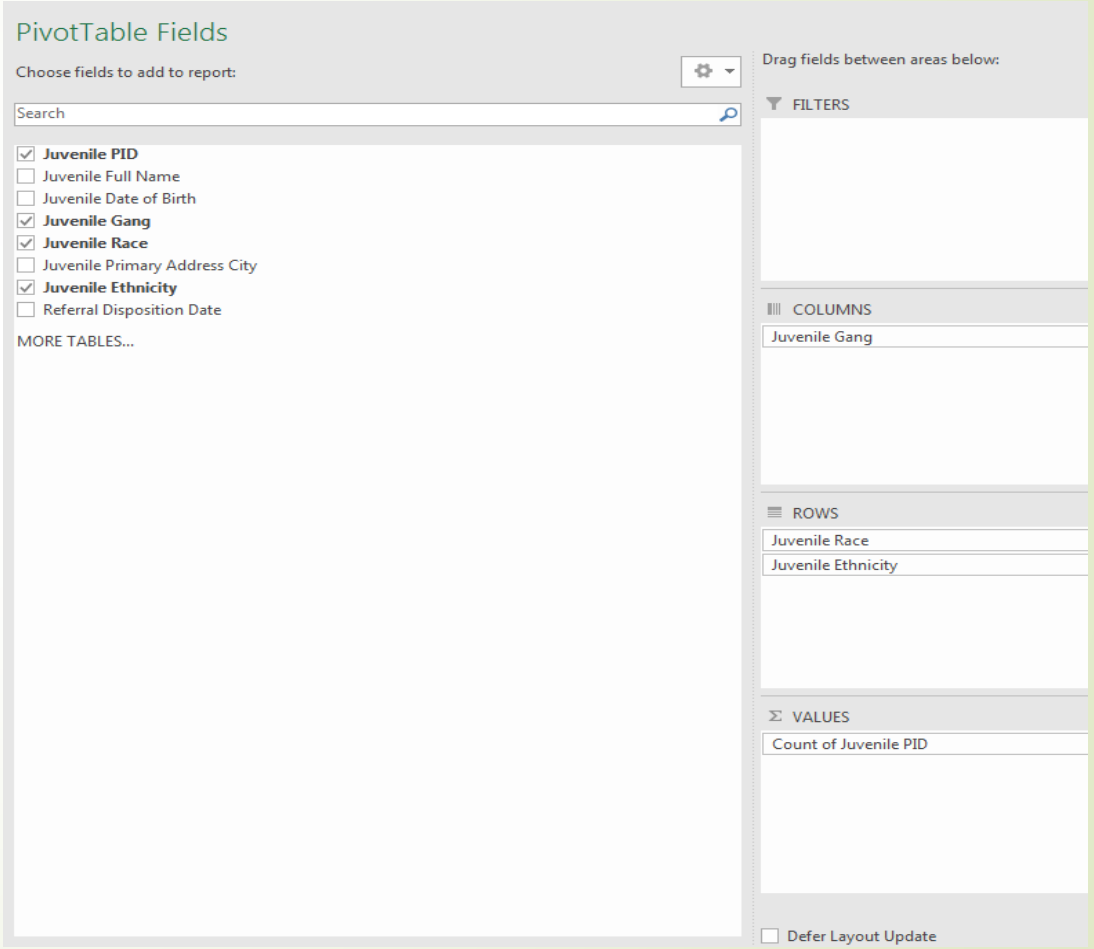
	A	B	C	D	E
1	Juvenile PID	Juvenile Full Name	Juvenile Date of Birth	Juvenile Gang	Juvenil
2	10000261	BERNAL, CONTRACT	6/6/2004		Wh
3	10000259	carrasco, contrac	6/4/2004		Wh
4	10000265	CEBALLOS, CONTRACT	######	Latin Kings	Wh
5	10000220	Compact, Isabelle	######	Arayan Nations	Bla
6	10000258	CONTRACT, CAVIN	6/6/2004	South Side 18	Bla
7	10000262	DAVIS, CONTRACT	6/6/2004		Wh
8		Doe, John	1/1/2001		Wh
9	10000255	EKE, CONTRACT	9/8/1999		Wh
10	10000229	Galindo, George	######		Wh

This window will come up in front of your data. You may want to change the range. Once you have the range you want, click OK.



Four boxes: so many choices

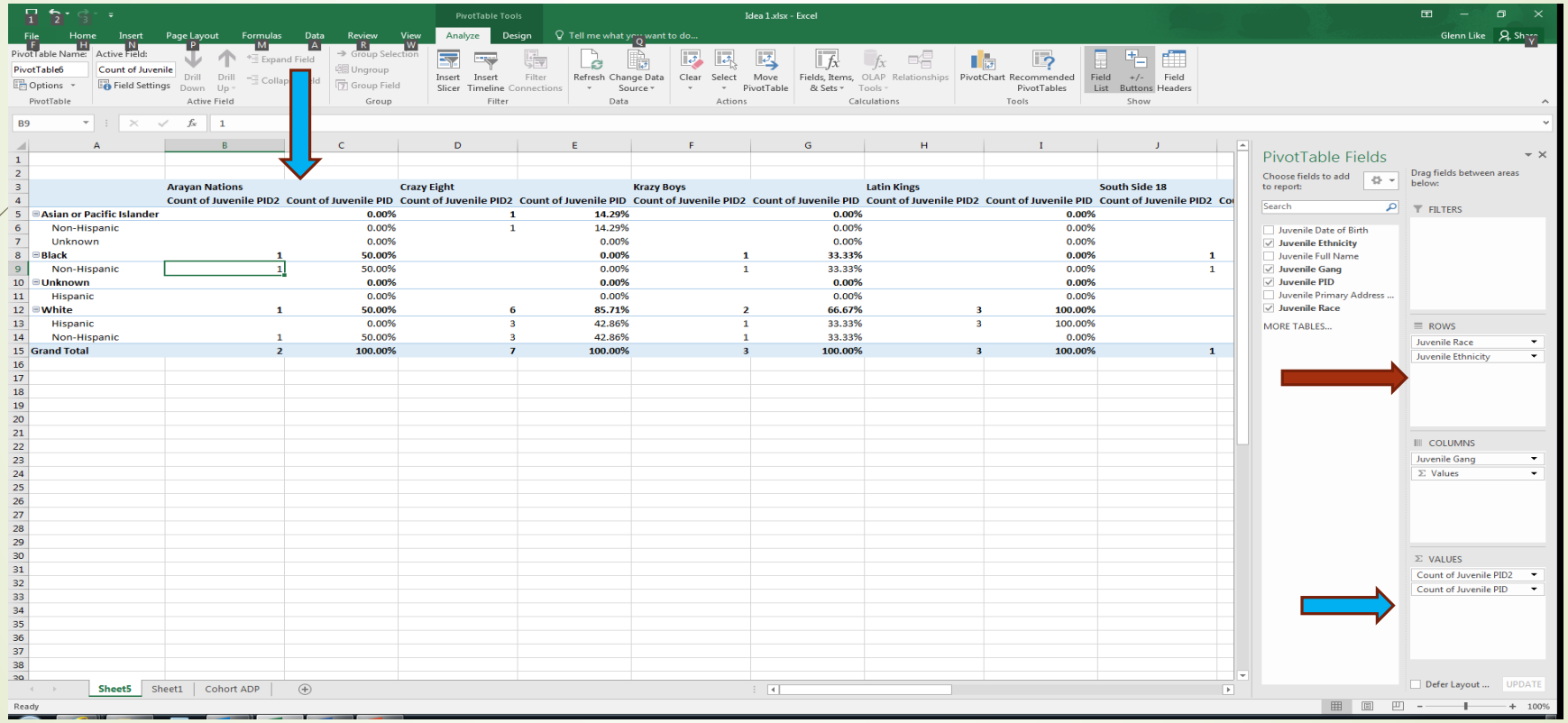
- A blank screen will come up with Pivot Table Fields on the right side.
- Dragging a variable to filter lets you decide what part of the variable not to show. You don't need to use the filter.
- You need at least one row or one column and one ∑. It is important to note that you won't actually see the variable in the ∑ unless you also include that variable as a row or column. Having a name or ID number (or anything that all records have) as the variable being counted is ideal.
- If you put more than one variable in row or column, the second (or subsequent) variable will be subvariables to the first. You can change the order by using the dropdown arrow next to the variable.



What if I want percentages?

- Numbers are nice, but a lot of times you need the percentage of each answer.
- Left click on one of the numbers in the pivot table that you would like to be a percent.
- One of the options will be Show Value As. Click on this.
- There are lots of options, but the main three are % of Grand Total (the percent of everything in the table), % of Column Total, and % of Row Total.
- If you want numbers and percentages in the same table, just drag the variable you used for ∑ into ∑ a second time. I recommend using the drop down to put the number first.

Example: Percent of Gang Membership by Race



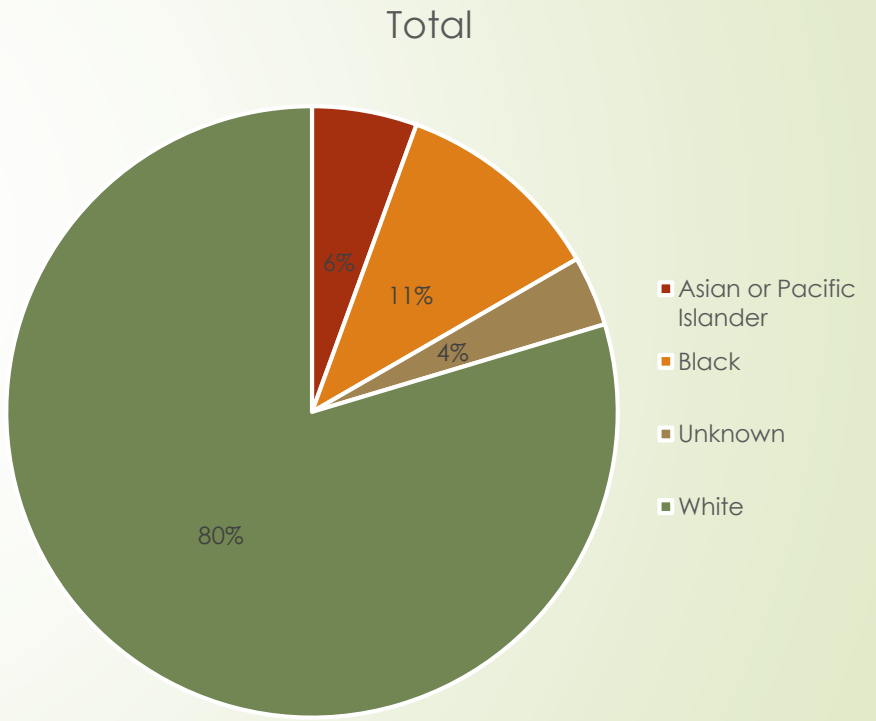
	Arayan Nations		Crazy Eight		Krazy Boys		Latin Kings		South Side 18	
	Count of Juvenile PID2	Count of Juvenile PID	Count of Juvenile PID2	Count of Juvenile PID	Count of Juvenile PID2	Count of Juvenile PID	Count of Juvenile PID2	Count of Juvenile PID	Count of Juvenile PID2	Co
Asian or Pacific Islander		**0.00%**	**1**	**14.29%**		**0.00%**		**0.00%**		
Non-Hispanic		0.00%	1	14.29%		0.00%		0.00%		
Unknown		0.00%		0.00%		0.00%		0.00%		
Black	**1**	**50.00%**		**0.00%**	**1**	**33.33%**		**0.00%**	**1**	
Non-Hispanic	1	50.00%		0.00%	1	33.33%		0.00%	1	
Unknown		**0.00%**		**0.00%**		**0.00%**		**0.00%**		
Hispanic		0.00%		0.00%		0.00%		0.00%		
White	**1**	**50.00%**	**6**	**85.71%**	**2**	**66.67%**	**3**	**100.00%**		
Hispanic		0.00%	3	42.86%	1	33.33%	3	100.00%		
Non-Hispanic	1	50.00%	3	42.86%	1	33.33%		0.00%		
Grand Total	**2**	**100.00%**	**7**	**100.00%**	**3**	**100.00%**	**3**	**100.00%**	**1**	

Charts

- When we were inserting the Pivot Table, you might have noticed that you could insert charts as well.
- When you create a chart, it is important to remember to try to give a clear and compelling picture for your reader: too much information is distracting, and too little (or too much white space) is often skipped over.
- Creating a chart or graph is similar to creating a pivot table:
  - Click on the data you want in the chart (don't worry if you don't get all of it or get too much, you can add or subtract it later)
  - Click on recommended chart or if you know the one you want, click on that one.
- There are lots of types of charts, but the main three are:

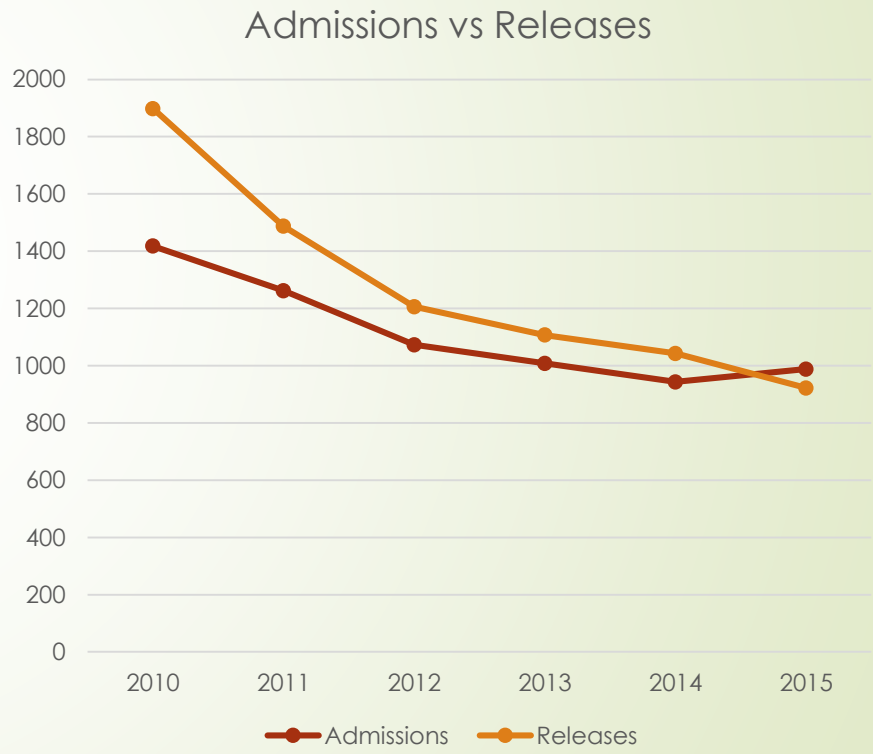
Charts-part II: Pie Charts

- Pie Chart—used to show the internal breakdown of something. The pie chart should always cover 100% of the thing. You can make as many slices as you want, but any slice that is less than 5% is hard to read, and it may be better to consolidate them if there is more than one.



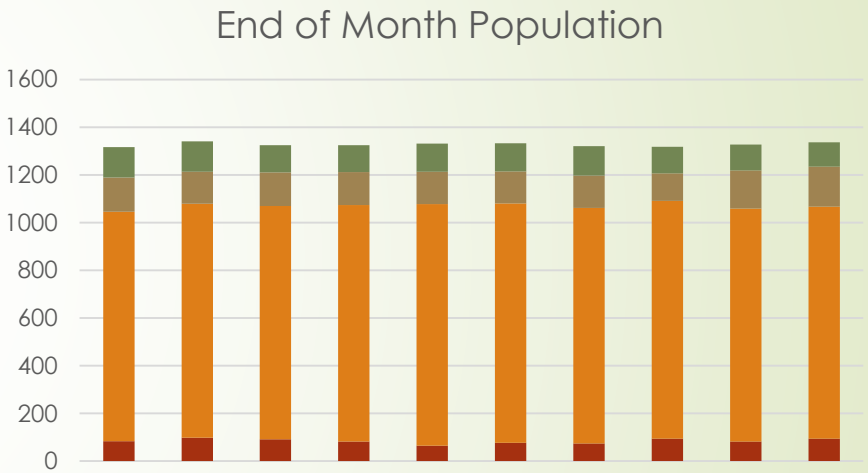
Charts part III: Line Charts

- Line charts are good for showing things that change over time. Excel has a trend line feature that you can use to project into the future (email me after class for more on that). Having too many lines can confuse your reader. I try to keep it at 2-5. If you are only tracking one thing, a bar graph (coming up next) is usually more visually interesting.



Charts part IV: Bar Graphs

- Bar graphs can be combination of pie charts and line graphs. You can convey how pieces of a whole change over time, but it is harder for the reader to interpret. I recommend writing a sentence telling the reader what you want them to know when you use a graph or chart, and I **strongly** recommend it for bar graphs that has stacked columns. You also might add a data table to this chart.

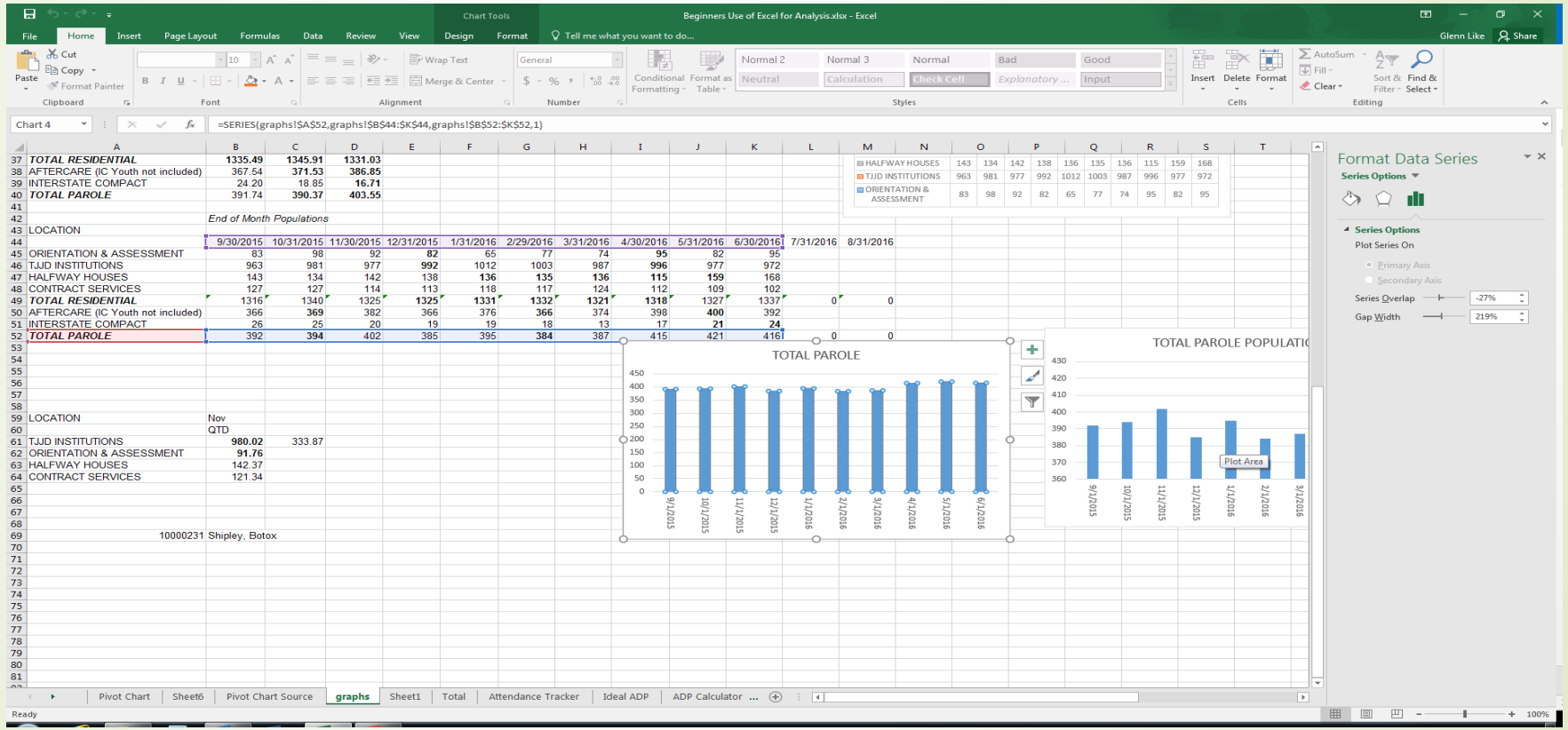


	9/30/2015	10/31/2015	11/30/2015	12/31/2015	1/31/2016	2/29/2016	3/31/2016	4/30/2016	5/31/2016	6/30/2016
CONTRACT SERVICES	127	127	114	113	118	117	124	112	109	102
HALFWAY HOUSES	143	134	142	138	136	135	136	115	159	168
TJJD INSTITUTIONS	963	981	977	992	1012	1003	987	996	977	972
ORIENTATION & ASSESSMENT	83	98	92	82	65	77	74	95	82	95

Customizing Charts

- When you click on a chart, two new tabs (Design and Format) will appear at the top of your worksheet.
- Design lets you change the style and type of the chart. It also has "Add Chart Elements" in the upper left hand corner. You can use this to add labels to the chart variables, add text to the side or under the chart, add a data table (usually better to use the "with legend key option" and delete the key the table automatically gives you), and add a trend line to the chart.
- Format lets you change the colors and fonts of individual elements of the table. On the more recent versions of Excel, clicking on part of the chart will open a tab on the right side of the worksheet that will give you a lot of the same options.

Example of Format screen that pops up when you click on part of the chart



Formatting Data

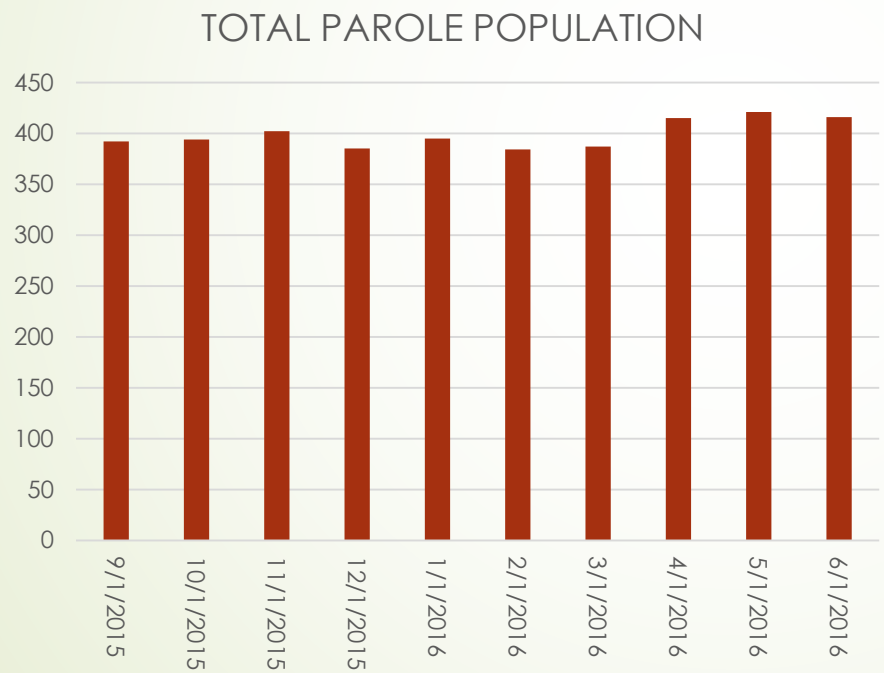
- Highlight the data that you want to format.
- If you are in the Home tab, look for the box that says Number between Alignment and Styles.
- There is a drop down box with various styles. Under the box, there are $, %, and ,--these are common numeric styles (currency, percentage, and numbers with commas). There are also two boxes with 0 and an arrow. These indicate how many spaces are shown behind the decimal place.
- If you have an identifier with leading zeros (like 0000012345), and you want to keep the zeros, you need to use Text in the dropdown.

Formatting Axes

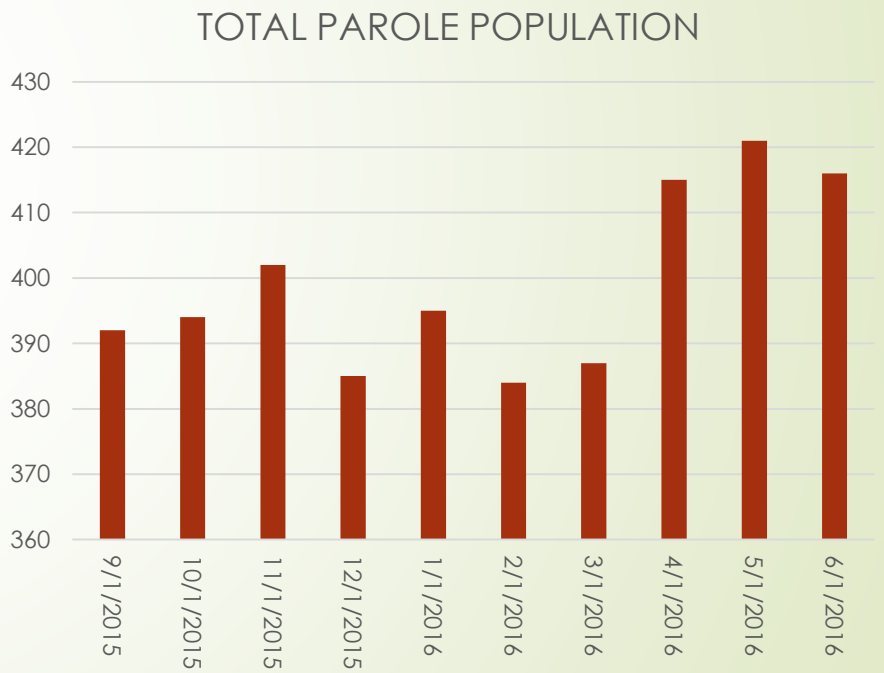
- Click on the axis of the graph.
- You will get a grey format screen on the right of the screen (newer Excel), or right click and pick format axis.
- A bunch of options will come up, but for the most part only 2 are important: **minimum**, which will indicate the smallest number shown on the graph (it is less common to change the maximum), and the **major** unit, which determines how far apart the columns or rows are.

The same graph with 2 axes

Axis set at 0



Axis Set at 360

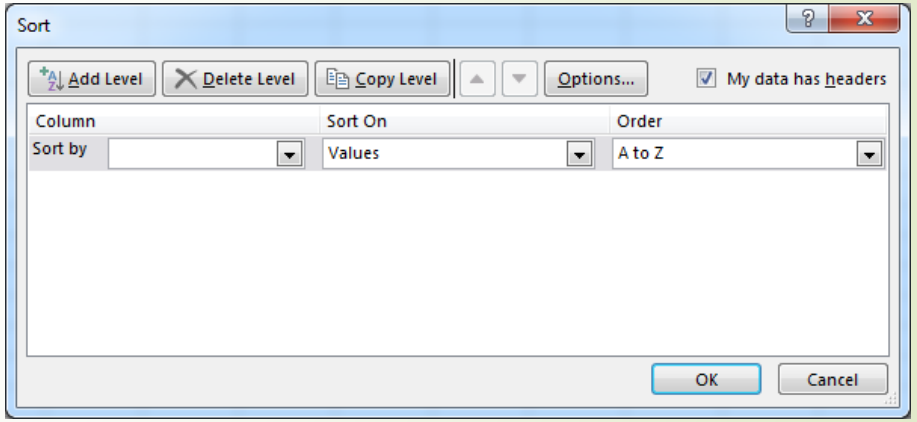


Sorting

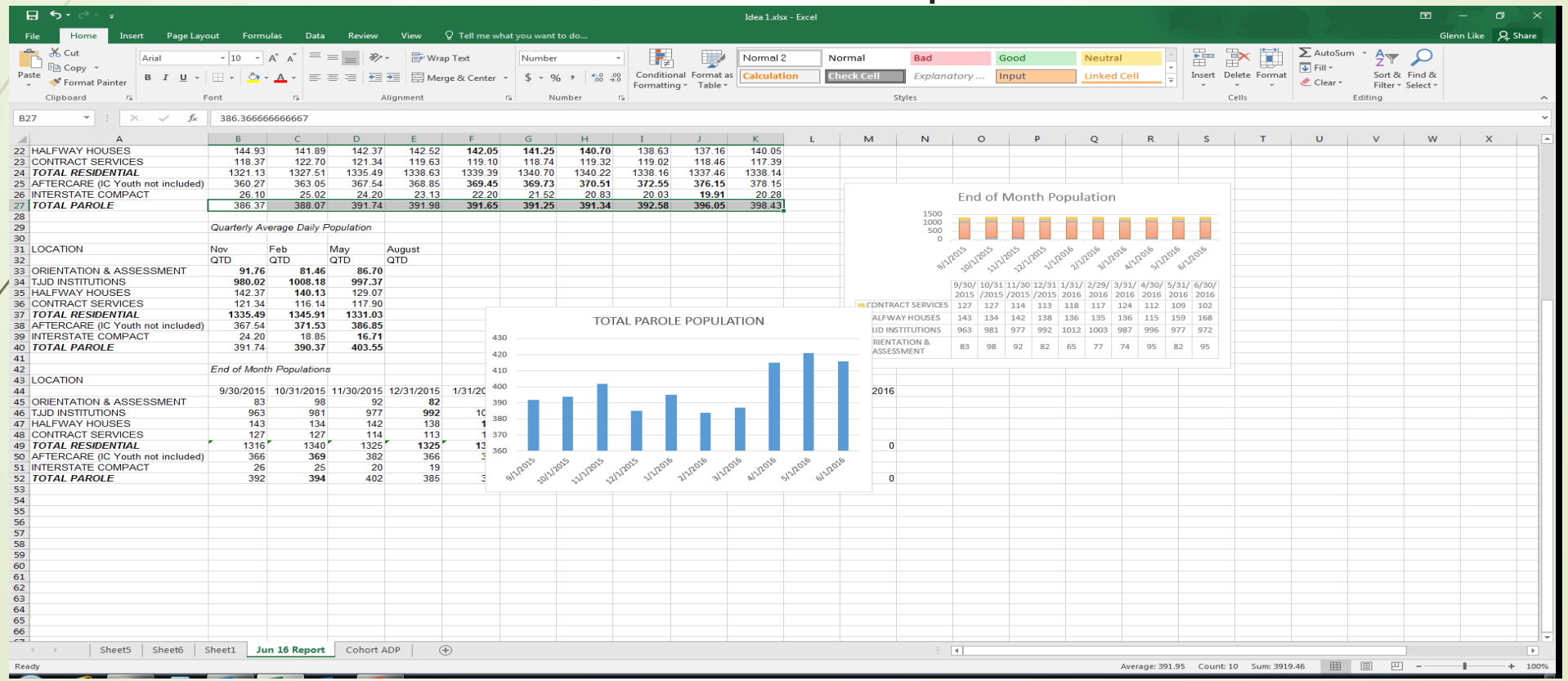
- Sometimes you want to change the order of things, particularly for graphs and tables.
- If you want things to be alphabetic, biggest to smallest, by date and time (or vice versa for any of those things), the easiest way is to sort.
- To sort:
  - Highlight everything you want sorted.
  - Click on Sort & Filter in the upper right hand corner of the worksheet.
  - A dropdown menu will appear with A to Z, Z to A, and custom sort. We will talk more about custom sort shortly.
- If you want to sort it in a different (manual) way, we will talk about that in processing data.

Custom Sorting

- Custom sort lets you sort by more than one thing at a time.
- Click on Add Level lets you add a second, third, etc. thing to sort by.
- Excel sorts everything by the top level first, and then by the second level inside the first.
- If you sorted by name and offense then it would show:
  - John Doe (first offense)
  - John Doe (second offense)



If you highlight a row or column of number, Excel gives you the average, count, and sum at the bottom left part of the screen.



Processing Data

Copying, Cutting, and Pasting

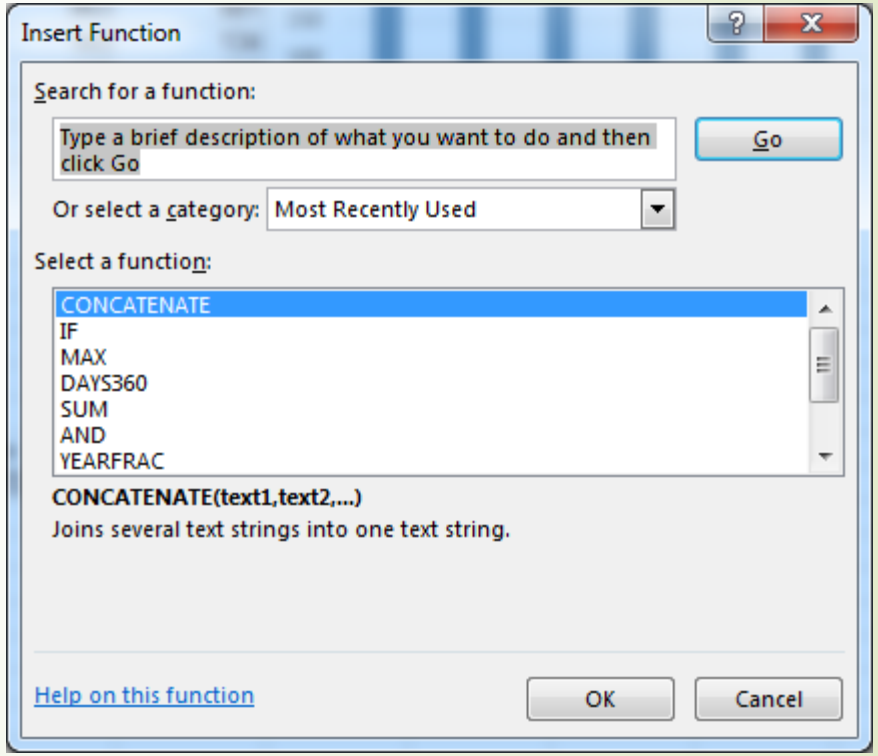
- The most common processing of data involves moving it around.
- To copy data, highlight the data you want to copy then click on crtl and C or click on copy in the upper left hand corner under Home. For the most part you don't want to copy data as a picture; that is usually only done if you are copying a table/graph/chart that you don't want the someone else to be able to change the exhibit.
- Cutting is just like copying, but the data disappears from where you took it. It uses crtl and V instead of C.
- To paste the data, you click on the cell where you want the data to start and click crtl P or click on paste in the upper left hand corner under Home. There are a lot of paste options, but the 2 most important are: the one with *fx* (which pastes the formula, more on that shortly)and the one with 123 at the bottom (which pastes the numbers, even if it comes from a formula).
- Remember that when you paste, you erase what was pasted over.

Functions part I

- Excel has a lot of functions. The set up for them is a little different from the operations strings in the criteria fields of JCMS, but once you get familiar with them, you should notice a lot of similarities.
- To activate a function, click a cell, and type =.
- Right above the A or B column, you will see a *fx* with an = in the really long bar to the right of it. Click on *fx*.

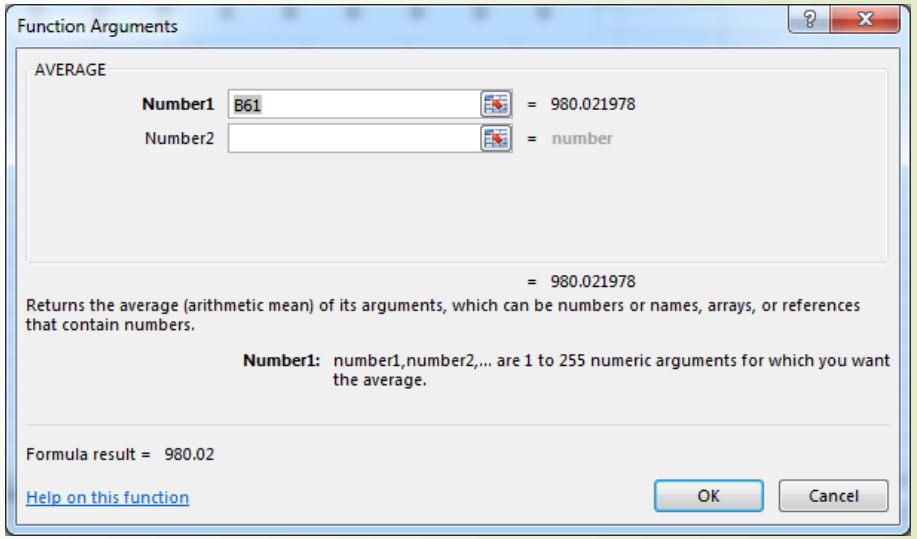
Functions part II

- A window looking like the one to the right of this appears.
- You can type in a description of what you are trying to do or look at the most recent functions.
- The best descriptions are similar to "find the biggest number" or "count the number of days between dates."
- The highlighted function gives you a description of what the function does and what you need to put into the function.



Functions Part III

- When you click on the function, a window for the function will open.
- Fill in the data location requested. It is just like highlighting cells, but you have the green moving fence around it.
- Click OK.



Modifying Functions

- Before we sample some of the functions that Excel provides, we should look at how to modify them.
- You can add (+), subtract (-), multiply (*), and divide (/) the values of other cells in the cell you are creating the function in.
- You can also compare cell values to see if they are greater than (>), less than (<), greater than or equal (>=), less than or equal (<=), or not equal (<>) to other cells. This is particularly useful in conjunction with the IF function.

Commonly Used Functions

- Some of the functions that I use regularly and think would be useful to you include:
- Sum
- Average
- Count
- Min (short for minimum or smallest number in a set)
- Max (short for maximum or largest number in a set)
- These are all fairly straightforward to use. The next several functions benefit from a little explanation.

Functions: IF, COUNTIF, AND, and OR (part I)

- You use the IF function to create different results based on whether a set of criteria is true or false.
- The IF function looks like IF (A, output if A is true, output if A is false) where A is a set of criteria.
- The COUNTIF function looks like COUNTIF (range, A) where A is a set of criteria.
- You can put one function inside another one (we call this nesting), and the IF/COUNTIF functions are often both nests (particularly with the AND/OR functions)and nestlings.

Functions: IF, COUNTIF, AND, and OR (part II)

- The AND function looks like AND (A, B), and gives a result of true (1) only if both A and B are true, and gives a result of false (0) if either A or B are false.
- If variable A was the sex of the juvenile and B was the offense level, an IF statement could look like IF (AND (A="M", B="SJ"), "Transfer to Male State Jail Diversion", "Not Eligible for Male State Jail Diversion"). [When you are nesting functions is important to keep track of parenthesis, although Excel will try to fix it if you don't.]
- The OR function is like the AND function, but instead of requiring both A and B to be true to give a true result, OR only requires one of A or B to be true.

Functions: DAYS360

- The DAYS360 function gives you an **approximation** of the number of days between 2 dates.
- It treats every month as having 30 days and every year as having 360 days.

Examples

Class (or Program) Attendance Tracking Spreadsheet

Goals and Basic Set Up

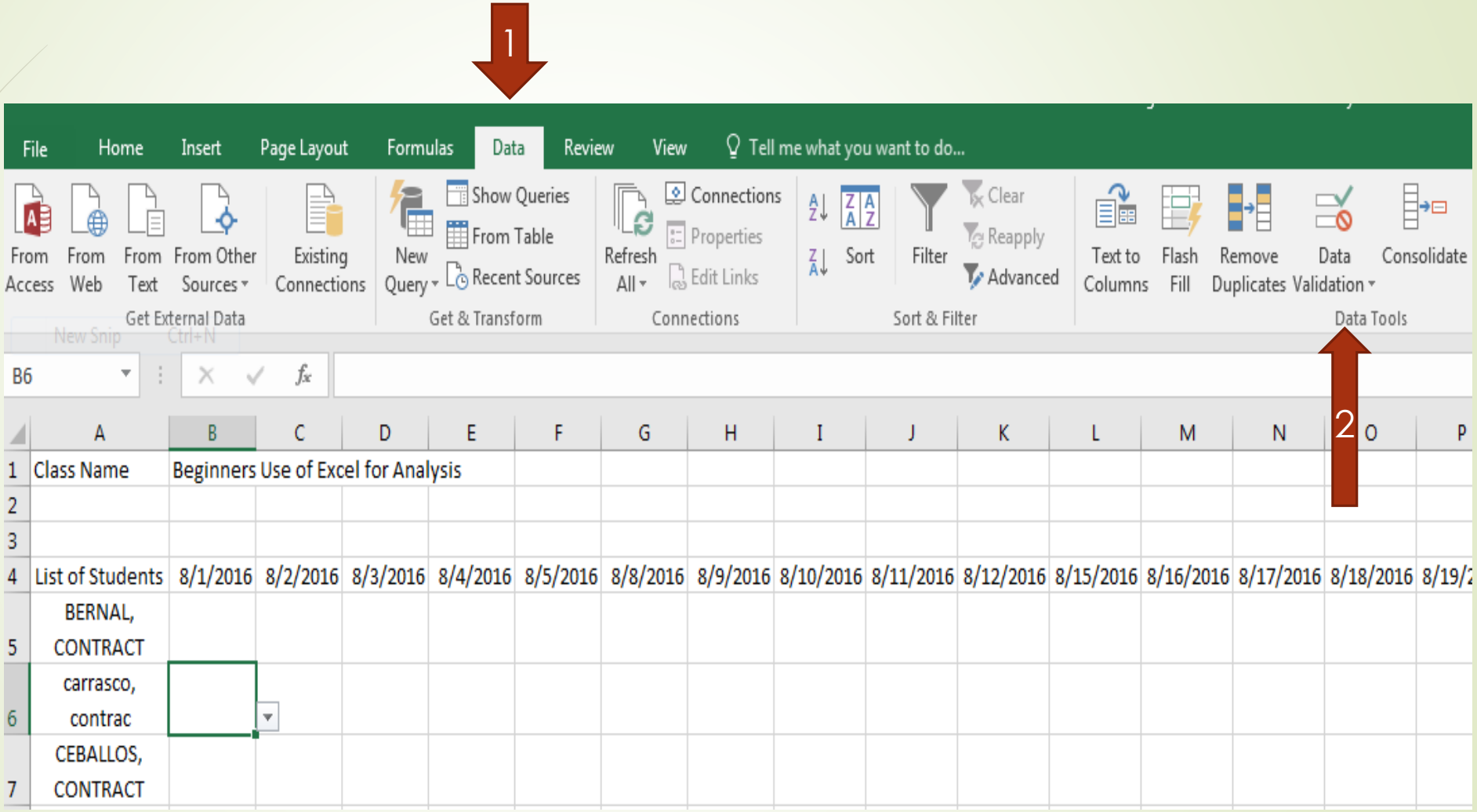
- Build a Drop Down Menu
  - This speeds up data entry
  - Reduces typos by reducing choices
- Make use of the COUNTIF function
- We will start by assuming you have a list of students and the dates that the class will be in session.

List of Students	8/1/2016	8/2/2016	8/3/2016	8/4/2016	8/5/2016	8/8/2016	8/9/2016	8/10/2016	8/11/2016	8/12/2016	8/15/2016	8/16/2016	8/17/2016	8/18/2016	8/19/2016	8/22/2016	8/23/2016	8/24/2016	8/25/2016	8/26/2016
BERNAL, CONTRACT																				
carrasco, contrac																				
CEBALLOS, CONTRACT																				
Compact, Isabelle																				
CONTRACT, CAVIN																				
DAVIS, CONTRACT																				
Doe, John																				
EKE, CONTRACT																				
Galindo, George																				
GARCIA, CONTRACT																				
GARCIA, MIDLAND																				
Harris, Samantha																				
HENDERSON, CONTRACT																				
Henson, Ashley																				
HERNANDEZ, CONTRACT																				
James, Jessica																				

Setting Up the Drop Down List

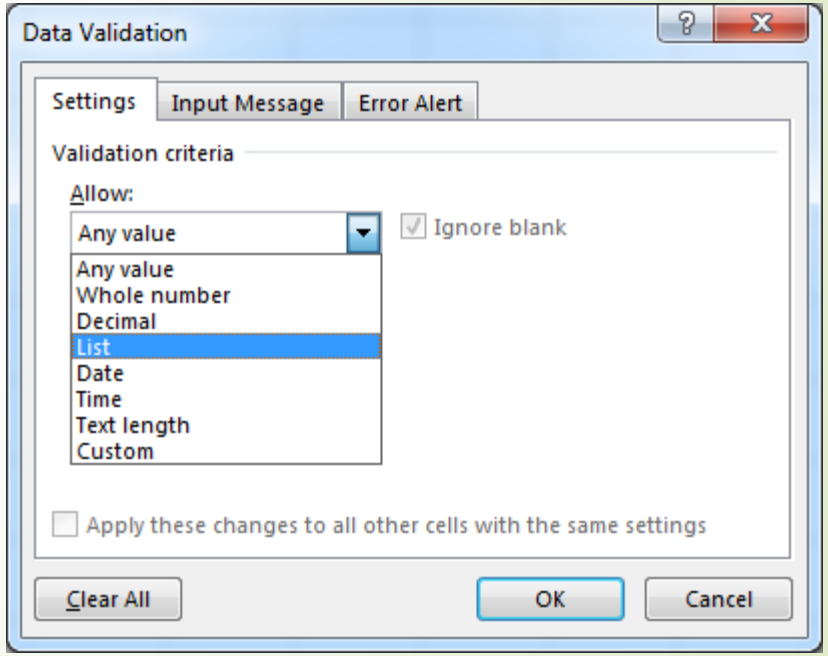
- Somewhere off to the side or on another sheet, you should type out a list of possible options to include in the drop down, one option per cell. Try to be complete.
- For example, we could have attended, excused absence, unexcused absence, class cancelled (approved), and class cancelled (unapproved).
- Click on the first cell under a date next to a student's name.
- At the top, by Home, Insert, Page Layout, and Formulas, you will see Data. Click on that.
- You will see Data Validation about midway on the screen (slightly to the right). Click on that.

Data and Validation



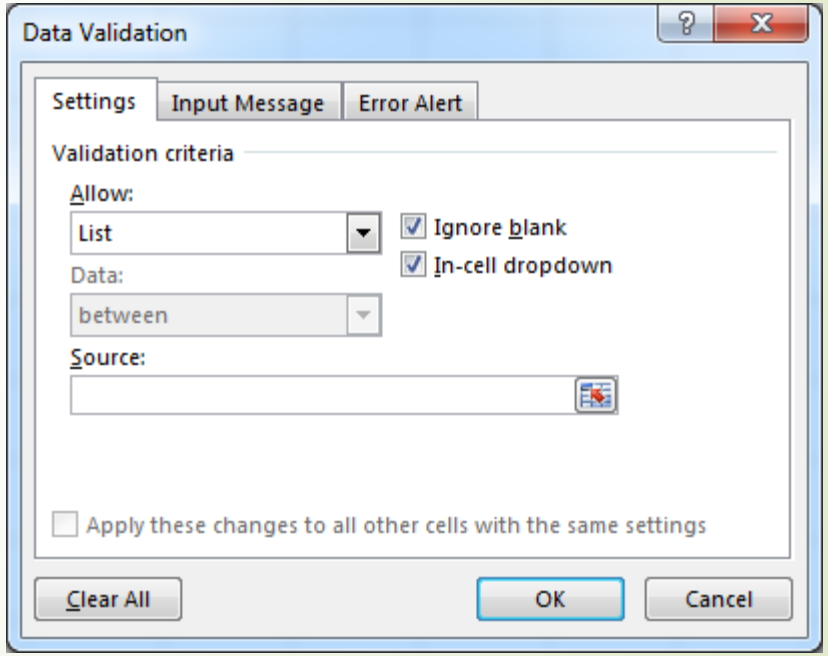
Data Validation part I

- Click on Data Validation and you will see this window.
- Click List.



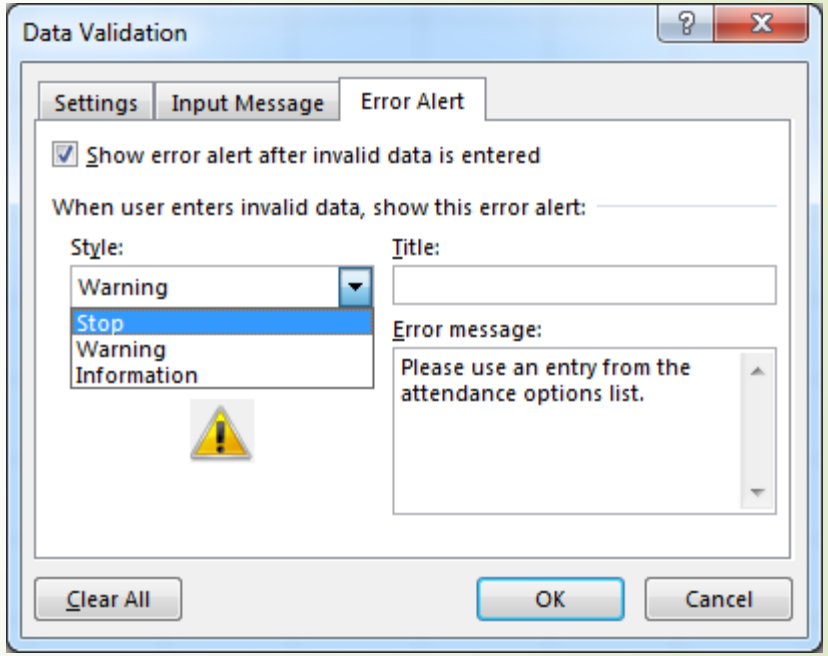
Data Validation part II

- The screen will change to look like this. Click on the box below Source:
- Then you want to go to your list of options for the drop down.
- If you want to restrict the user from being able to enter something not on the list, click Error Alert, otherwise click OK.



Data Validation part III the Error Alert

- If you want to restrict (or strongly encourage) users to using the options on your list, go to Error Alert.
- Stop will not allow them to enter anything not on the list.
- Warning will make a pop up window appear that the user click "Yes" to make a nonlist entry.
- Information just lets them know they made a nonlist entry.
- Excel has default messages, but you can add your own.

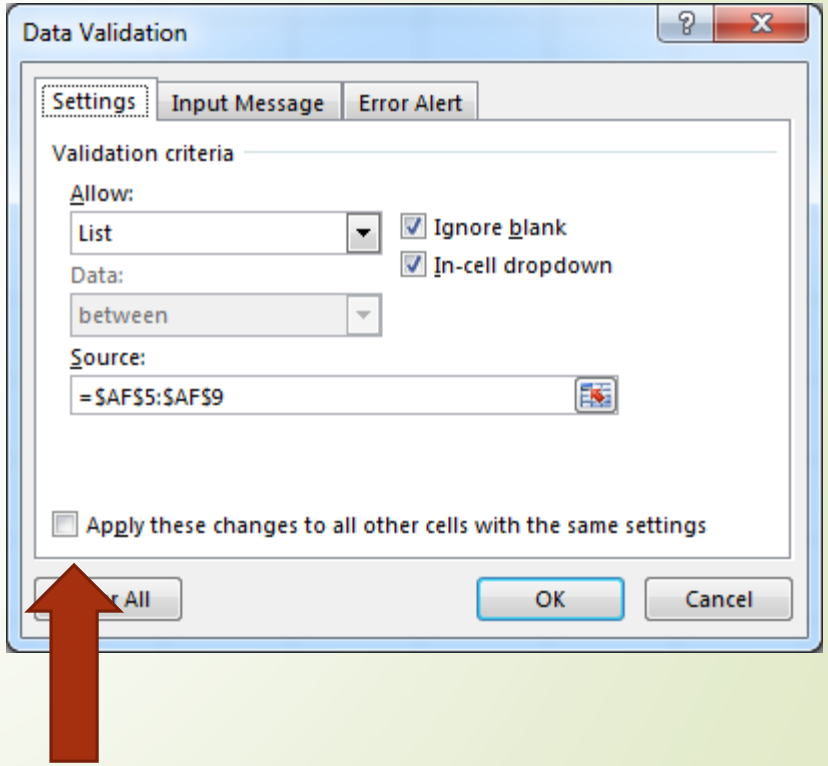


Drop Down Finale

- When you click on the cell under the first date next to the first name, you will see a little grey box with an upside down triangle inside it. Clicking on that little box will give you the list of things in the drop down.
- You can copy that cell and paste it on all the other cells under the dates next to the names, and all those cells will have the drop down.

What if I Need to Change the List?

- Make the change in the list.
- Click on one of the cells with the drop down menu.
- Get back to data validation.
- Make your changes.
- Click on the little box in the bottom so you don't have to copy and paste again.



COUNTIF and billing

- As long we have come this far, we might as well take care of figuring out the cost too.
- Let us say that the provider gets $25 per student per day if the student attended, $20 if they didn't unless the class was cancelled, but the cancellation was not authorized by you (in which case they get $0).
- We can create 4 variables to get the cost.
  - The first is Cost Attended, which has a formula 25*COUNTIF(B5:U5,"Attended").
  - The second is Days Unpaid, which has a formula COUNTIF(B5:U5,"Class Cancelled (unapproved)")
  - The third is Partial Payment. In our example, there are 20 possible days, so the cost will be $20* (20-the days we paid $25-the days we paid $0) or 20*(20-COUNTIF(B5:U5,"Attended")-X5)
  - And finally a total of Cost Attended plus Partial Payment.

Average Daily Population Calculation

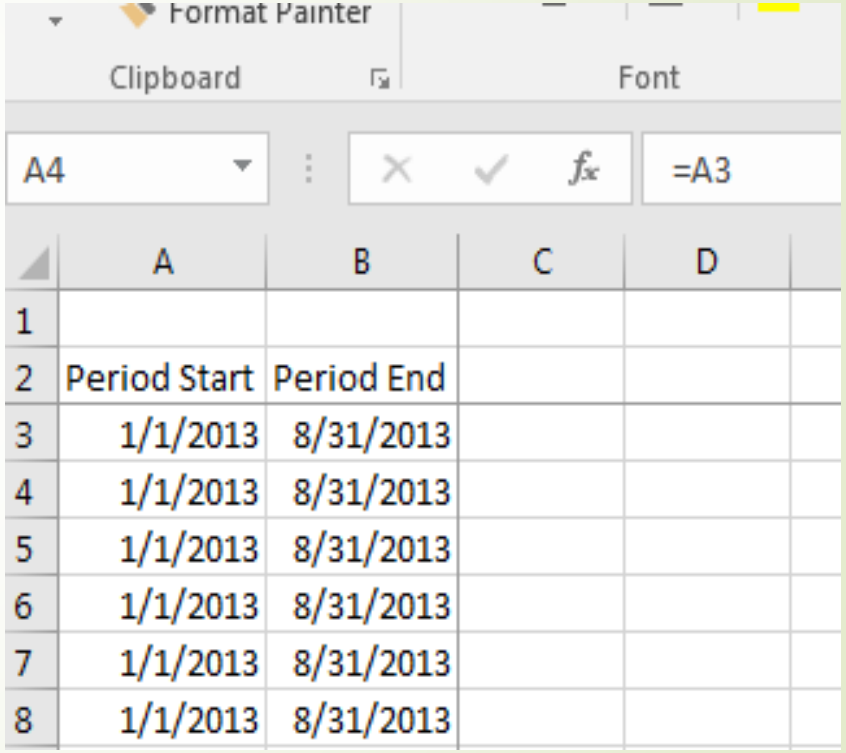
Ideal way to calculate ADP

- Ideally you would have a list of population by day, and you would just take the average (either with the formula or by highlighting the numbers).
- If that isn't available, but you have a list of everyone in custody (and when they entered and exited custody), you can still calculate ADP in 6 steps.

DAY	Population
1/2/2014	1316
1/3/2014	1319
1/4/2014	1319
1/5/2014	1317
1/6/2014	1309
1/7/2014	1310
1/8/2014	1310
1/9/2014	1314
1/10/2014	1315
1/11/2014	1315
1/12/2014	1311
1/13/2014	1314
1/14/2014	1313
1/15/2014	1314
1/16/2014	1314
1/17/2014	1320
/18/2014	1320
1/19/2014	1310
1/20/2014	1309
1/21/2014	1309
1/22/2014	1307
1/23/2014	1303
1/24/2014	1303
1/25/2014	1309
1/26/2014	1307
1/27/2014	1309
1/28/2014	1306
1/29/2014	1306
1/30/2014	1305
1/31/2014	1311
Average	1311.47

Step 1: Establish the Period Start and End dates

- Type Period Start and Period End in two cells.
- Format the columns Period Start and Period End as dates:
  - Click on the letters above Period Start and Period End
  - Go to Number (between Alignment and Styles at the top of the screen)
  - Select a date from the drop down menu
- I don't want to have to retype the Period Start and Period End for everyone in custody, so we will create a function to duplicate it for every record:
  - On the second empty cell under Period Start, enter a =.
  - Then click on the first empty cell under Period Start.
  - Drag the function down for as many rows as you expect juveniles.
  - Drag it across to Period End.



	A	B	C	D
1				
2	Period Start	Period End		
3	1/1/2013	8/31/2013		
4	1/1/2013	8/31/2013		
5	1/1/2013	8/31/2013		
6	1/1/2013	8/31/2013		
7	1/1/2013	8/31/2013		
8	1/1/2013	8/31/2013		

Step II

- If you know your period start and end dates, I would enter them. This also helps you see if it is working correctly. For this example the period is between January 1 and August 31, 2013.
- Next to Period Start and Period End, enter Start Date and End Date. These are the actual starting and ending dates. If the juvenile is still in custody, you can leave End Date empty.
- Format their columns as dates.
- Enter (or copy) one set of Start Date and End Date per juvenile who was in custody. It doesn't matter if they were in before or after the period start and end dates (step III takes care of that). Just to keep things from being **Do not include any juveniles who weren't in custody during the period** (more complicated versions of this process could take care of that, but this is an intro class).

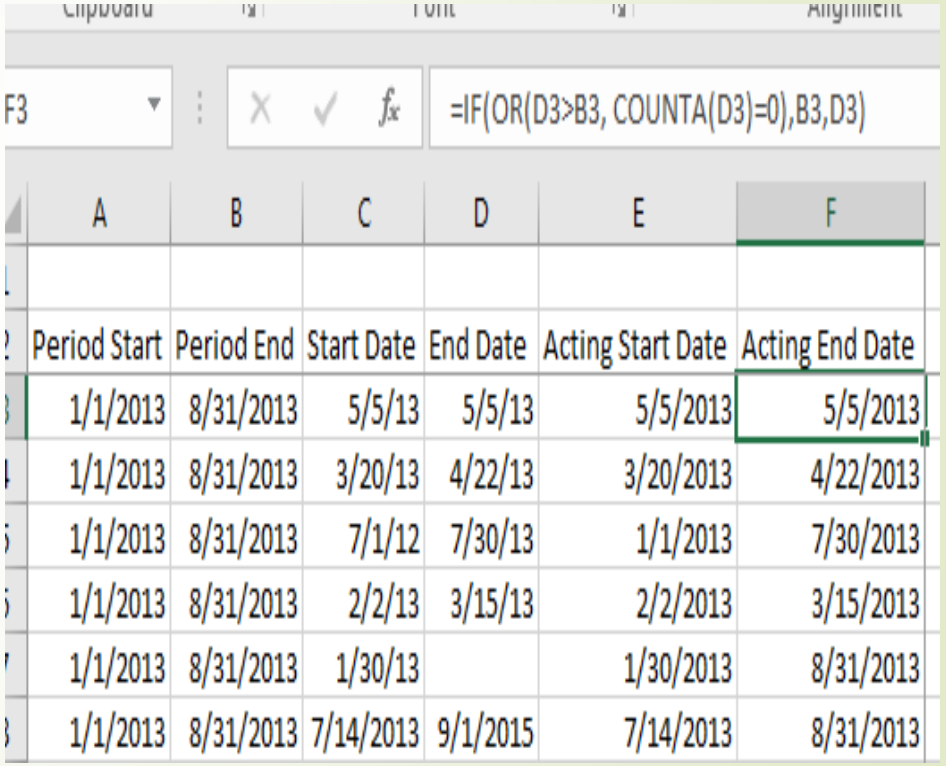
Period Start	Period End	Start Date	End Date
1/1/2013	8/31/2013	5/5/13	5/5/13
1/1/2013	8/31/2013	3/20/13	4/22/13
1/1/2013	8/31/2013	7/1/12	7/30/13
1/1/2013	8/31/2013	2/2/13	3/15/13
1/1/2013	8/31/2013	1/30/13	
1/1/2013	8/31/2013	7/14/2013	9/1/2015

Step IIIa

- We only want to include information that took place during the period of interest. We can use Excel to cut off the dates. Next to Start Date and End Date, type Acting Start Date and Acting End Date.
- Format these columns as dates.
- We are going to use the IF function. Under Acting Start Date, enter a =, then pick IF in functions.
- If the start date is before the period started, we want to use the period start date instead, so in the IF function, put C3<A3,A3,C3
- If the end date is after the period ended, we want to use the period end date instead. But we also have to consider what if the juvenile hasn't left yet. This becomes complicated.

Step IIIb

- In addition to IF, there are a couple of functions to consider:
  - CountA counts how many cells aren't empty. So if CountA (D3)=0, then D3 is empty.
  - OR (A, B) gives us a value of true if either A or B are true.
- Remember IF is really IF (A, value if A is true, value if A is false).
- Put an IF function under Acting End Date with the following:
  - OR(D3>B3, COUNTA(D3)=0),B3,D3



	A	B	C	D	E	F
	Period Start	Period End	Start Date	End Date	Acting Start Date	Acting End Date
	1/1/2013	8/31/2013	5/5/13	5/5/13	5/5/2013	5/5/2013
	1/1/2013	8/31/2013	3/20/13	4/22/13	3/20/2013	4/22/2013
	1/1/2013	8/31/2013	7/1/12	7/30/13	1/1/2013	7/30/2013
	1/1/2013	8/31/2013	2/2/13	3/15/13	2/2/2013	3/15/2013
	1/1/2013	8/31/2013	1/30/13		1/30/2013	8/31/2013
	1/1/2013	8/31/2013	7/14/2013	9/1/2015	7/14/2013	8/31/2013

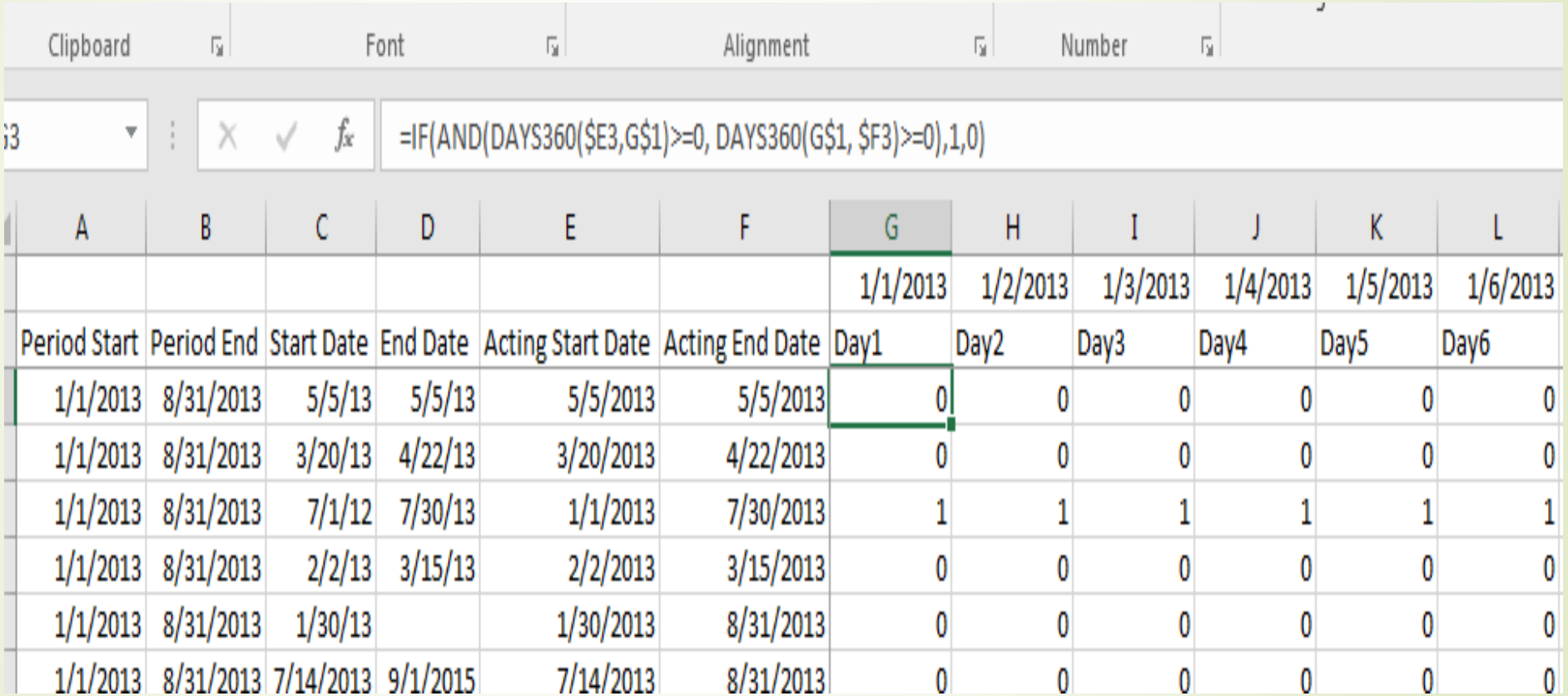
Step IV

- We have completed the most complicated part.
- In order to get the ADP, we need to distribute these stays into days.
- Next to Acting End Date, type Day1.
- Drag this as far you are interested in. I generally go out 366 days (for Leap Year).
- I like to see the dates, so go to the row above Day1, type = and set it equal to Period Start Date.
- Then at the cell above Day2, type =, then set it equal to the cell above Day1+1. Drag that to the end.
- Optional: Click on the first Acting End Date. Go to View at the top (to the right of Home), click Freeze Panes, Freeze Panes. This will make it so you always see everything above and/or to the left of the first Acting End Date.

Step V

- Step V is the second most complicated part. You need to know 3 things:
- The function Days360 gives the duration between 2 dates (if all the months had 30 days).
- If you put a $ in front of cell name in a formula, it will keep the same place if you drag the formula left or right.
- If you put a $ in between the letter and the number of a cell name in a formula, it will keep the same place if you drag the formula up or down.
- Under Day1, type=, then activate the IF function and put AND(DAYS360($E3,G$1)>=0, DAYS360(G$1, $F3)>=0),1,0 in it.
- Drag the function down to the last entry and then across to the last day.
- You may have to check on the day after the period ending day, as you might get some false positives, but since you shouldn't be doing the average of that date, it shouldn't matter.

Step V Illustrated



Clipboard | Font | Alignment | Number

=IF(AND(DAYS360($E3,G$1)>=0, DAYS360(G$1, $F3)>=0),1,0)

	A	B	C	D	E	F	G	H	I	J	K	L
							1/1/2013	1/2/2013	1/3/2013	1/4/2013	1/5/2013	1/6/2013
	Period Start	Period End	Start Date	End Date	Acting Start Date	Acting End Date	Day1	Day2	Day3	Day4	Day5	Day6
	1/1/2013	8/31/2013	5/5/13	5/5/13	5/5/2013	5/5/2013	0	0	0	0	0	0
	1/1/2013	8/31/2013	3/20/13	4/22/13	3/20/2013	4/22/2013	0	0	0	0	0	0
	1/1/2013	8/31/2013	7/1/12	7/30/13	1/1/2013	7/30/2013	1	1	1	1	1	1
	1/1/2013	8/31/2013	2/2/13	3/15/13	2/2/2013	3/15/2013	0	0	0	0	0	0
	1/1/2013	8/31/2013	1/30/13		1/30/2013	8/31/2013	0	0	0	0	0	0
	1/1/2013	8/31/2013	7/14/2013	9/1/2015	7/14/2013	8/31/2013	0	0	0	0	0	0

Step VI

- A few cells under the last Acting End Date, you can type Daily Population.
- Under Day1, type = then get the Sum function. It should automatically put the range from the first to last juvenile in there, but double check to make sure.
- Drag the sum function to the last day.
- You can either highlight the sums to get the average or create an average function in an empty cell. Either way, make sure to only go as far as the period end date.

Step VI Illustrated

G13 =AVERAGE(G11:IO11)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1							1/1/2013	1/2/2013	1/3/2013	1/4/2013	1/5/2013	1/6/2013	1/7/2013
2	Period Start	Period End	Start Date	End Date	Acting Start Date	Acting End Date	Day1	Day2	Day3	Day4	Day5	Day6	Day7
3	1/1/2013	8/31/2013	5/5/13	5/5/13	5/5/2013	5/5/2013	0	0	0	0	0	0	0
4	1/1/2013	8/31/2013	3/20/13	4/22/13	3/20/2013	4/22/2013	0	0	0	0	0	0	0
5	1/1/2013	8/31/2013	7/1/12	7/30/13	1/1/2013	7/30/2013	1	1	1	1	1	1	1
6	1/1/2013	8/31/2013	2/2/13	3/15/13	2/2/2013	3/15/2013	0	0	0	0	0	0	0
7	1/1/2013	8/31/2013	1/30/13		1/30/2013	8/31/2013	0	0	0	0	0	0	0
8	1/1/2013	8/31/2013	7/14/2013	9/1/2015	7/14/2013	8/31/2013	0	0	0	0	0	0	0
9													
10													
11						Daily Population	1	1	1	1	1	1	1
12													
13						Average	2.27						

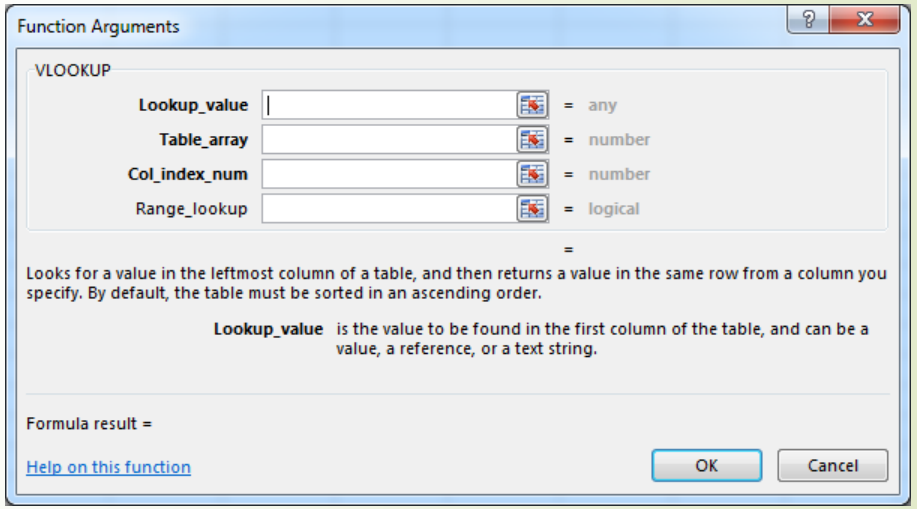
Recidivism

Overview

- Assumption: you have two Excel worksheets, one with youths who were disposed for supervision that you are trying to find out if they recidivated (we will call that the supervision worksheet) and one file with people who had an event that would count as recidivism and the date of the event (we will call this the referral worksheet). PID number is in column A in the supervision worksheet.
- Goal: to use the VLOOKUP function to get information from the referral worksheet to the supervision worksheet if there is a match, hopefully in a way that let's you feel confident in using it for other things.
- VLOOKUP is short for vertical lookup. This means it searches down a column for a match (but pulls from the row the match is made from). There is also a horizontal lookup (HLOOKUP) that does the opposite (searches across a row for a match, but pulls from the column the row is in).
- Just a reminder, you can always talk to someone on the JCMS team to make sure you are getting the right files (and if you have any questions on how to calculate recidivism).

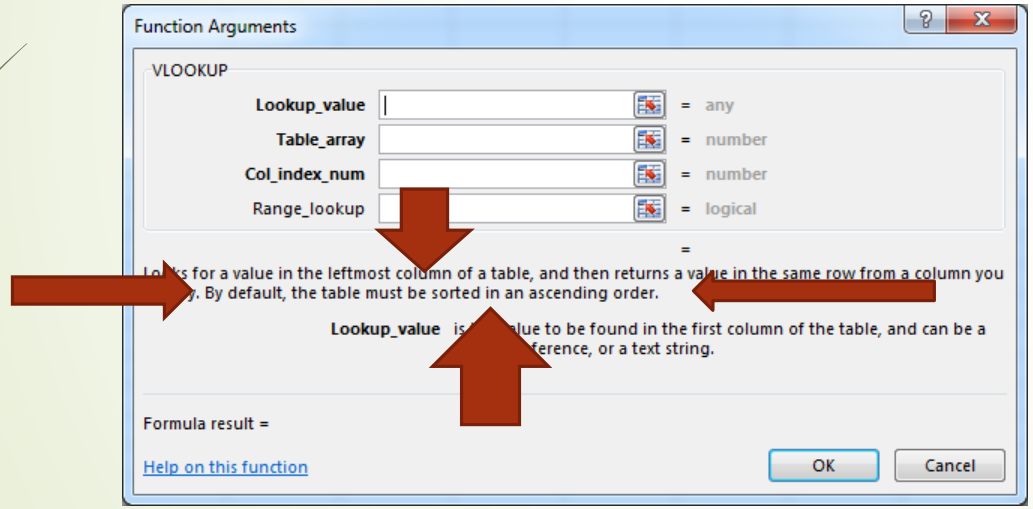
VLOOKUP Components

- The **lookup_value** is the very first thing you want to match in your supervision sheet.
- The **table_array** is the referral worksheet, specifically the column you want to match with, the column(s) you want to get data from, and everything in between.
- The **col_index_num** is the **NUMBER** of columns to the right of the column you are matching that contains the information you want to match. Remember this is not the column name.
- **Range_lookup** is really a question: Are you comfortable with matching on an approximation or do you need an exact match? If you are good with an approximation, type TRUE, if you want an exact match, type FALSE. In this example you will type FALSE.



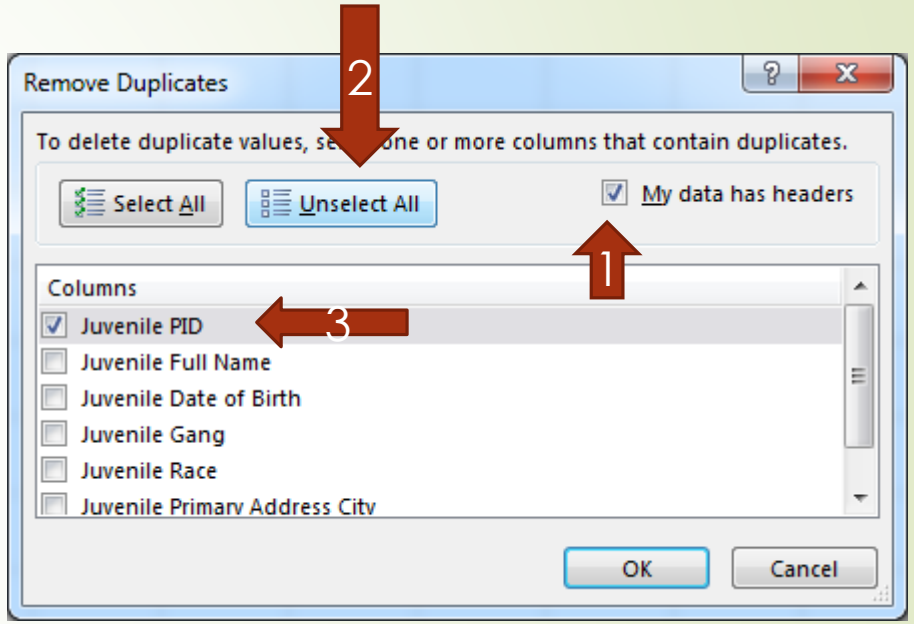
Don't Forget to Remind them of the Very Important Thing

- Sort the column in the referral file that you are matching on in ascending order before you start the VLOOKUP function in the supervision file.



Let's Calculate Some Recidivism: The Referral Worksheet

- In the referral worksheet, sort by the PID number and Disposition Date. We are going to match on PID number, but we want the first match.
- Click any cell inside the referral worksheet that has a value in it. Then click on Data at the top. Right next to Data Validation, there is a button for Remove Duplicates. Click on the button.
- This window will pop up. Check My data has headers, click Unselect All, and check PID.
- Go to the supervision worksheet.

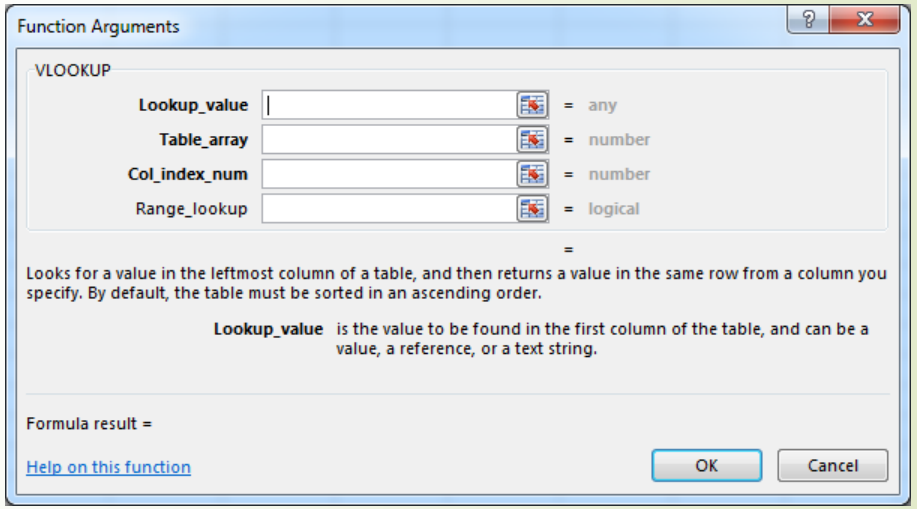


Let's Calculate Some Recidivism: The Supervision Worksheet

- Optional: I would make a copy of the supervision worksheet in case you delete too much. If you don't make a copy, I suggest you highlight the PID numbers (except the column header) and get the count of how many PID's you have. You will need that later.
- Excel files pulled from other programs (like JCMS) tend to have titles and dates on the top. These could cause problems. Better to be same and delete everything above the column headers (like PID). I would highlight the rows, and go to Cells on the upper right hand side, and click delete. Ditto any columns in front of PID (if there is anything in those columns you want to keep, cut and paste them somewhere else).
- You can't recidivate if you didn't have a recidivating event occur on some date. So our first VLOOKUP will be to get the date this occurred (if it did for the youth). This will answer if and when. So on the first empty column, type "Referral Disposition Date" in the same row as PID. In the next cell down, type a = and get the VLOOKUP function.

VLOOKUP: What to Enter

- The **lookup_value** should be the cell name for the first cell that has a PID number in it (which should be in the same row that you are creating the function).
- The **table_array** is the referral file—at least everything from the PID number to the date.
- The **col_index_num** is the number of columns the date is from the PID number in the referral file. This treats the PID column as distance 1, so it is always the number of columns plus 1.
- **Range_lookup**: type FALSE.



VLOOKUP: What the formula looks like

- In the bar next to the *fx* at the top, the formula should look like: VLOOKUP(A2,'[FileName.xlsx]SheetName'!DataRange,#columnin File2adding to FIle1,false).
- If your referral worksheet was called Referral, and the Referral Disposition Date was the next column over from the PID (and they are columns A and B respectively), and there were 100 records under the column header, then the formula would look like:
- VLOOKUP(A2,'Referral'!A2:B101,2,false).
- If your referral file is in a different spreadsheet, which we will call Recidivism County A, then the formula looks like:
- VLOOKUP(A2,'[Recidivism County A.xlsx] Referral'!A2:B101,2,false).

Fixing the Date

- If the column next to your recidivism date was a date, your recidivism date probably has a date format. If not, reformat the column.
- Optional but highly recommended: I would copy or write down the PID # and go to the referral sheet and double check that you got the right date. If you didn't get a date at all (but get a #N/A), then your youth didn't recidivate (victory!), but I would check the referral sheet anyway, and then drag the VLOOKUP function until I found a youth with a referral disposition date and check that.
- Drag the function down to the row with the last PID number you are checking.

#N/A #N/A #N/A #N/A, #N/A #N/A #N/A #N/A, Hey, Hey Goodbye-Part I

- With luck, you have a lot of #N/A's. On one hand, that is good (less recidivism), but they complicate what follows so we are going to get rid of the records that have them.
- Click on any cell with a value in it. Go to Sort and Filter (top right) and click on it and click filter (next to a picture of a funnel). Little grey boxes with upside down triangles will appear next to the column headers.
- Click on the little grey box next to Referral Disposition Date.
- Click on the box next to Select All. This will make all the check marks vanish.
- Click on the box next to #N/A. This will put a little check mark there and a black box inside the box next to Select All.
- Click OK.
- You will now see a table with just the records with a #N/A in them.
- Highlight all of the records except the column header.

#N/A #N/A #N/A #N/A, #N/A #N/A #N/A #N/A, Hey, Hey Goodbye-Part II

- Right click. You will see a pop up screen. One option will be to delete rows.
- Click that.
- It will ask if you are sure. Click "yes."
- Click OK.
- Now you have a nearly empty screen. Click on the little box next to Referral Disposition Date.
- Click Select All. A bunch of check marks should appear.
- Click OK.

Actually Calculating Recidivism-part I

- If you work in the justice system, you know that going a full year after supervision ends is big deal. So if anyone gets arrested after a year, we aren't going to count them as recidivating.
- Go back to the referral sheet. There should be a Referral Date and a Referral Disposition Date. In the first empty column, type DateDif (short for date difference) in the column header row.
- In the next cell down, type = and get the DAYS360 formula. You want to go from Referral Disposition Date (when they where released from supervision) to Referral Date (when they got referred again). Drag the formula down until you run out of rows with data in them.
- Next to DateDif, type Recidivism.

Actually Calculating Recidivism-part II

- For many of you, the first DateDif will be in cell E2.
- We will use the IF and the AND function to flag the rearrests within range. Remember that AND (A, B) is only true if both A and B are true.
- In F2, type = and pick the formula IF, then put in the boxes for IF: in the top box: AND(E2>0, E2<=360), in the true box: Y, and the false box: N. Remember we are using 360, because this formula treats 360 as a year. Click OK, then drag the formula until you hit the row with the last PID number in it.
- Add a filter like we did with for the #N/A's earlier, but this time get rid of the N's. When the N's are gone, go to the filter and select all again. Your referral sheet should only have entries with Y's in the Recidivism column.

Actually Calculating Recidivism-part III

- Changing the referral sheet is nice and all, but your interest is in the supervision sheet, which is the people from your jurisdiction. We need to bring that information to the supervision sheet. If you think that involves VLOOKUP, you would be right.
- On the first open column, type "Recidivate."
- You are going to use VLOOKUP like before, but the **table_array** will expand to cover the Recidivism column in the referral sheet, and the **col_index_num** will specify the number of columns from PID to Recidivism.
- You can use the COUNTIF function to find out how many PID's had a "Y" in Recidivate. You can divide (/) that count by the number of PID's you originally had in the supervision sheet. If you do this in an Excel cell, format the cell as a % and you will have your recidivism rate.

VLOOKUP: What the formula looks like

- As we did last time, in the bar next to the *fx* at the top, the formula should look like: VLOOKUP(A2,'[FileName.xlsx]SheetName'!DataRange,#columnin File2adding to FIle1,false).
- If your referral worksheet was called Referral, and the Recidivism column was three columns over from the PID (and they are columns A and D respectively), and there were 100 records under the column header, then the formula would look like:
- VLOOKUP(A2,'Referral'!A2:B101,4,false).
- If your referral file is in a different spreadsheet, which we will call Recidivism County A, then the formula looks like:
- VLOOKUP(A2,'[Recidivism County A.xlsx] Referral'!A2:B101,4,false).

If you have any questions, feel free to contact me at

Glenn Like
(512)-490-7767
glenn.like@tjjd.texas.gov