Let's get pivotal!

Pivot tables and how to use them

What is a pivot table?

A pivot table allows us to summarize the data grouped by one or more other variables.

Some common ways of summarizing:

- Count by group
- Sum by group
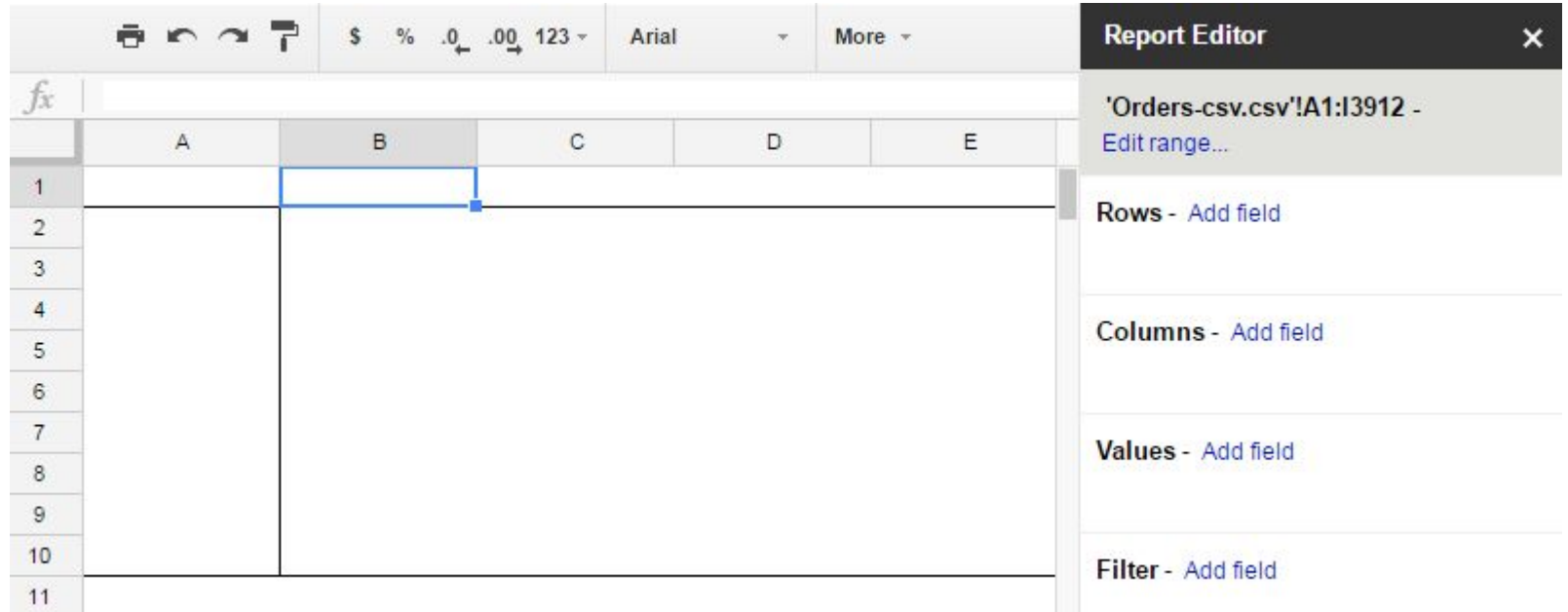- Average by group
- Maximum or minimum by group

Let's load the data

Use best practice to load the orders-csv file into Google sheets, including making a log.

The pivot table interface

Data -> Pivot table

You'll notice that a new tab opens with an interface that looks like this:

On the grid - grouping by one variable

Range: the values we're including (make sure you agree with google's choice)

Rows: Our grouping variable

Values: The variable we want summarized by group

Exercise: IL deadbeat parents

The orders file is a database of child support orders against parents in Illinois. It was used for a series of real-life stories about the state's lack of effort when it came to enforcing payment of child support.

One thing you may have noticed is that there can be multiple rows for the same person. This is because a single parent can have multiple orders, either for different children or because they have fallen behind on the same child more than once.

Exercise: IL deadbeat parents

So then a reasonable question is how many orders does each unique person in the database have, and who has the most?
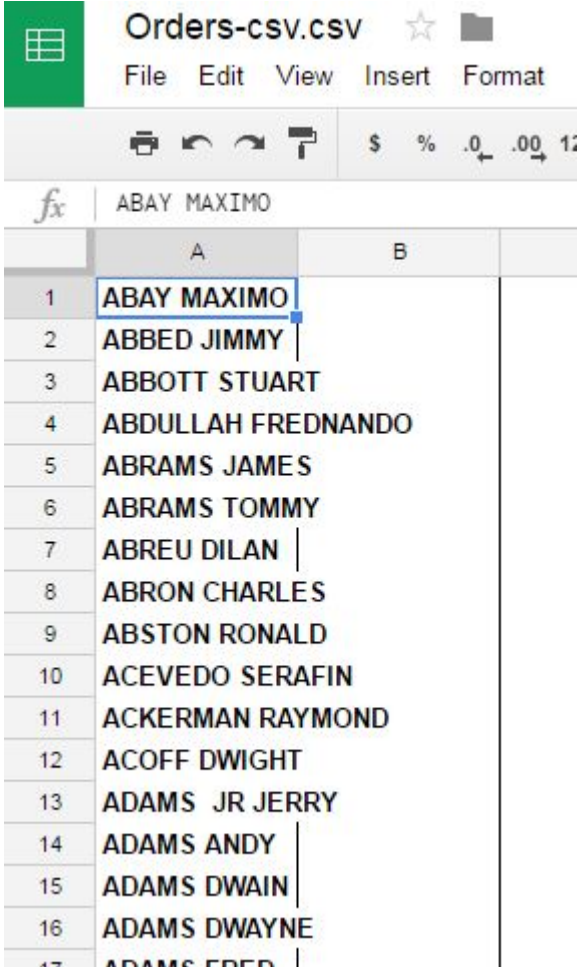
A pivot table can help with this.

Exercise: IL deadbeat parents

Since our grouping variable is 'NAME', let's add that field to Rows. We can do that in the upper right corner of our screen.

Exercise: IL deadbeat parents

You'll see that now
each name that
appears in the 'NAME'
field now appears
exactly once in our
pivot table.

Exercise: IL deadbeat parents

Now we want to know how many records correspond to each name in our data.

This is where our unique id comes in handy. Let's get a count of unique ids that correspond to each name.
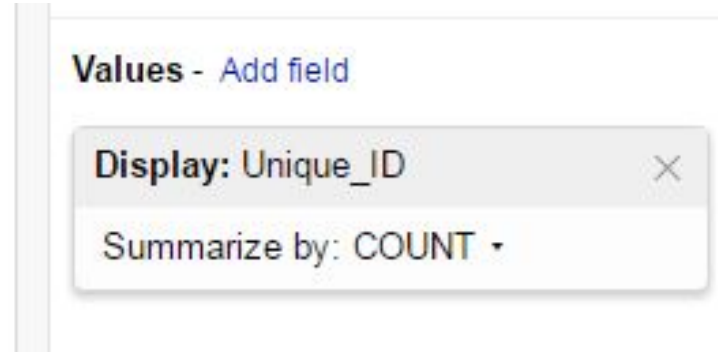
Exercise: IL deadbeat parents

Now we have a count of the number of records that correspond with each name in our data.

Yay!

But wait -- what would make this table more useful?

(Hint: the table is currently in alphabetical order. Is that what we want?)

Name	Count
ABAY MAXIMO	1
ABBED JIMMY	1
ABBOTT STUART	1
ABDULLAH FREDNANDO	1
ABRAMS JAMES	1
ABRAMS TOMMY	1
ABREU DILAN	1
ABRON CHARLES	2
ABSTON RONALD	1
ACEVEDO SERAFIN	1
ACKERMAN RAYMOND	1
ACOFF DWIGHT	1
ADAMS JR JERRY	1
ADAMS ANDY	1
ADAMS DWAIN	1
ADAMS DWAYNE	1
ADAMS FRED	1
ADAMS GLENN	1

Exercise: IL deadbeat parents

Now we have a count of the number of records that correspond with each name in our data.

Yay!

But wait -- what would make this table more useful?

(Hint: the table is currently in alphabetical order. Is that what we want?)

ABAY MAXIMO	1
ABBED JIMMY	1
ABBOTT STUART	1
ABDULLAH FREDNANDO	1
ABRAMS JAMES	1
ABRAMS TOMMY	1
ABREU DILAN	1
ABRON CHARLES	2
ABSTON RONALD	1
ACEVEDO SERAFIN	1
ACKERMAN RAYMOND	1
ACOFF DWIGHT	1
ADAMS JR JERRY	1
ADAMS ANDY	1
ADAMS DWAIN	1
ADAMS DWAYNE	1
ADAMS FRED	1
ADAMS GLENN	1

Exercise: IL deadbeat parents

That's right!
It would be more useful if this table were sorted by our COUNT of unique_id column that is in our pivot table.

Set the 'Sort by' to the field we want. Then set the 'Order.' We want the biggest values first, so we set it to Descending.
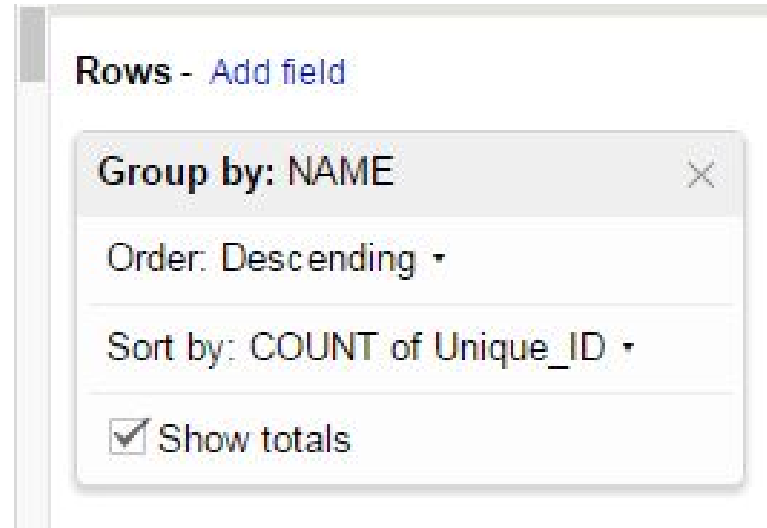
Exercise: IL deadbeat parents

But COUNT is just one way to summarize data by group.

Let's find the SUM of the NUM_KIDS field.

Now let's find the AVERAGE of the DEBT field.

We can also create a custom calculated field, like the square root of the number of kids divided by the total debt per NAME. This isn't really useful here, but could be good to keep in mind for the future.

Exercise: IL deadbeat parents

But COUNT is just one way to summarize data by group.

Let's find the SUM of the NUM_KIDS field.

Now let's find the AVERAGE of the DEBT field.

We can also create a custom calculated field, like the square root of the number of kids divided by the total debt per NAME. This isn't really useful here, but could be good to keep in mind for the future.

But wait ...

Right now we are just grouping by name. Is that specific enough if we are looking at the entire state of Illinois?

Let's go back and use the filter button to find Michael Williams, who led the rankings for number of kids.

NAME ▼	STREET	CITY	STATE	ZIP	DOB
WILLIAMS MICHAEL	10154 S STATE ST	W CHICAGO	IL	60628	12/27/1955
WILLIAMS MICHAEL	1118 E HYDE PARK BLVD	W CHICAGO	IL	60615	2/18/1949
WILLIAMS MICHAEL	5721 S LOOMIS	W CHICAGO	IL		6/2/1957
WILLIAMS MICHAEL	1130 N MAYFIELD	W CHICAGO	IL	60651	7/30/1959
WILLIAMS MICHAEL	6152 S INDIANA AV	W CHICAGO	IL	60637	5/17/1964
WILLIAMS MICHAEL	1501 PRAIRIEVIEW	RANTOUL	IL	61866	9/14/1962
WILLIAMS MICHAEL	4640 SHERIDAN#1005	W CHICAGO	IL	60640	11/19/1961
WILLIAMS MICHAEL	1551 S DRAKE	W CHICAGO	IL	60623	7/19/1964
WILLIAMS MICHAEL	328 E 117TH STREET	CHICAGO	IL	60628	4/4/1952
WILLIAMS MICHAEL	7157 S EUCLID	W CHICAGO	IL	60649	5/23/1957
WILLIAMS MICHAEL	413 S 4TH ST	ARCOLA	IL	61910	9/25/1958
WILLIAMS MICHAEL	1436 MARENGO	FOREST PARK	IL	60130	7/25/1959

We can group by more than one variable

What is another identifying characteristic that might be useful?

Let's add birthdate as another grouping characteristic.

We still want to sort by number of kids, and if we uncheck 'Show totals' it will get rid of those pesky rows
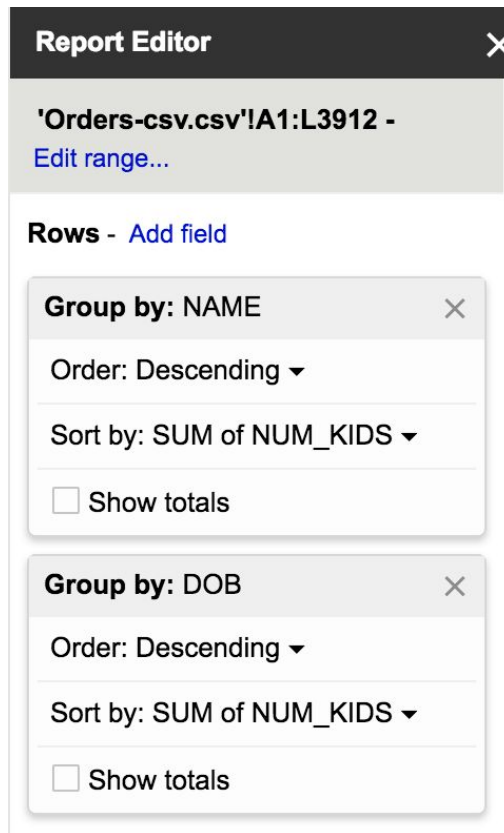
We can group by more than one variable

But wait - it isn't sorting globally from largest to smallest number of kids anymore. So how do we know who has the most overall?

The filter function can help us here.

We can group by more than one variable

The filter function on google pivot tables is just a smidge buggy, so before we start let's go ahead and insert a blank row above our pivot table.

Then let's select Column C and click the filter button.

Because filtering groups the values in the column for us, we can scroll down and see that the maximum number of children per unique name/birthdate combination is 7.

Let's filter so that just the sevens are shown.

We can group by more than one variable

There's our answer folks.
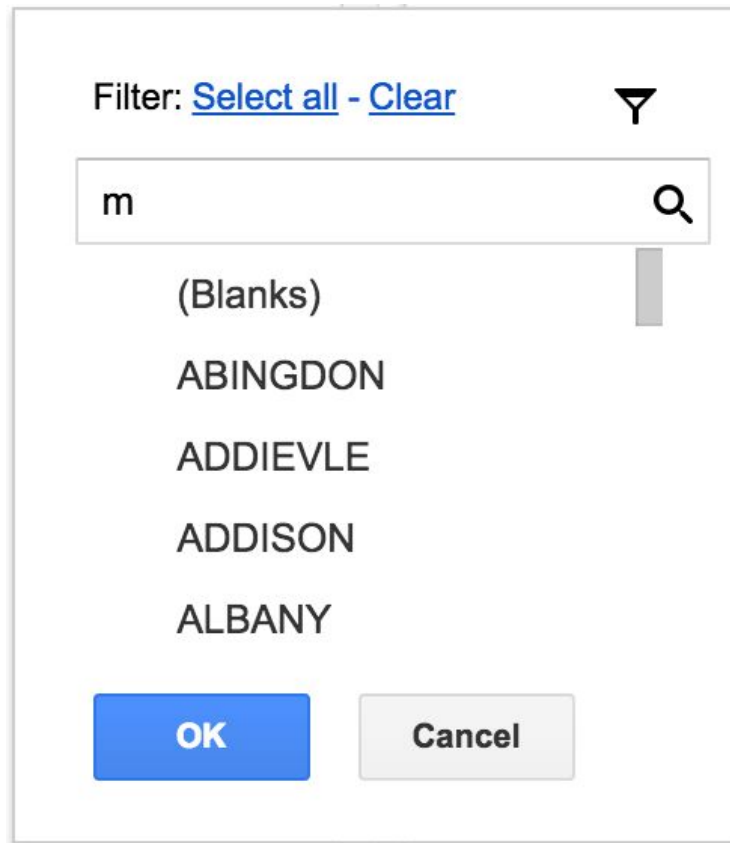
⊟ **TURNER BILLEY**	9/29/1941	7
⊟ **LINNERTZ EARL**	4/29/1934	7
⊟ **DAVIS COOPER**	10/13/1951	7

We can also do a cross-tab

Let's try removing our filter and birthday and adding 'CITY' as a column header.

Oops - we'll have to filter to fall within google's cell size limit for pivot tables.

Let's choose only cities that have the letter 'm' in them.

We can also do a cross-tab

This is what we get. Do you understand what google did here?

This example is not very meaningful, but crosstabs can be very useful.

	CARRIER MLS	EAST MOLINE	MACHESNEY P.	MACOMB	MADISON	MAHOMET	MANHATTAN	MANITO
WHEELER DENSIL							5	
NORIN RANDY								
KENNEY CHARLES								
VOLLBRACHT CARL								
SCHWESKA JOHN								
SADOWSKI JERRY								
RASMUSSEN JEFFREY			3					
MURPHY PATRICK								
HUNTER HENRY								
HOEFT FREDERICK								
DREFCHINSKI MICHAEL								
CRABB DONALD				3				
YARD MICHAEL								
WILLIAMS CHARLES								
WELLSBY NORMAN								
WARDINGLEY BENJAMIN								
TURNBOUGH WILLIAM					2			

One last thing

When you make a pivot table, it remains tethered to the underlying data you used to create it.

On the upside, this means that if you change your data, the pivot table will update to reflect the change.

On the downside, this means that there are a lot of limitations on the editing you can do in pivot mode.

To get around this, you can:

Copy the table, then go into a new tab and do:

Edit -> Paste special -> paste values only

Then you have an untethered table that you can edit as you please. But BEWARE - this table will not update if you change the underlying data.