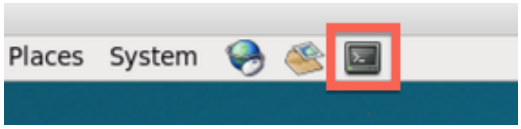
[Download a PDF of this hands-on tutorial]

ExploringtheRelationalDataModelofCSV.pdf

By the end of this activity, you will be able to:

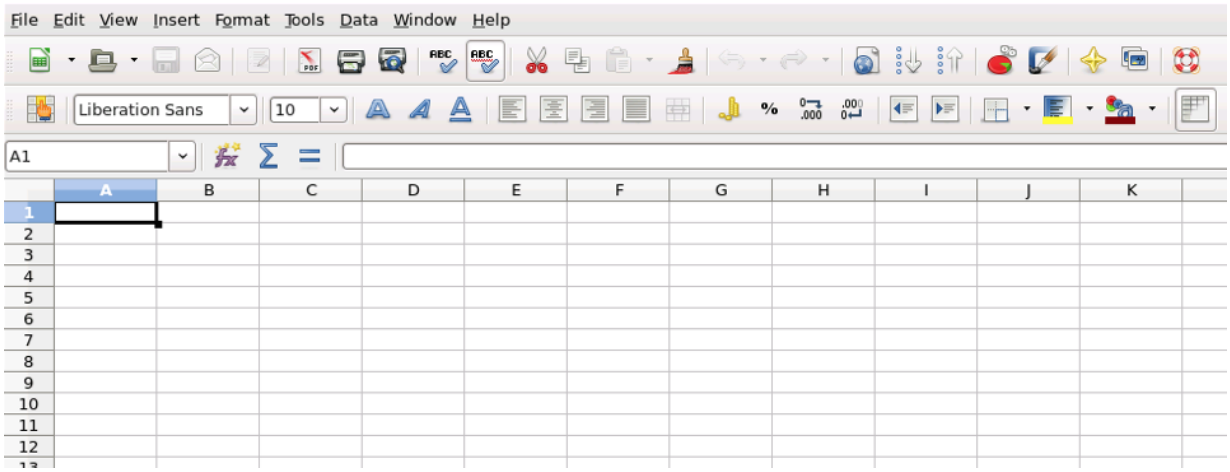
1. Filter rows in a spreadsheet
2. Perform aggregate operations such as average and sum

Step 1. Open a terminal shell. Open a terminal shell by clicking on the square black box on the top left of the screen.

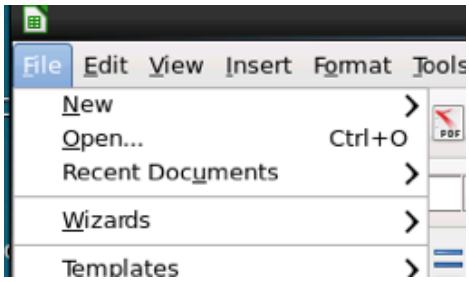


Step 2. Open spreadsheet application. Run *oocalc* to start the spreadsheet application.

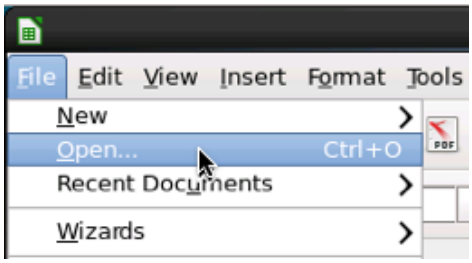
```
1  oocalc
```



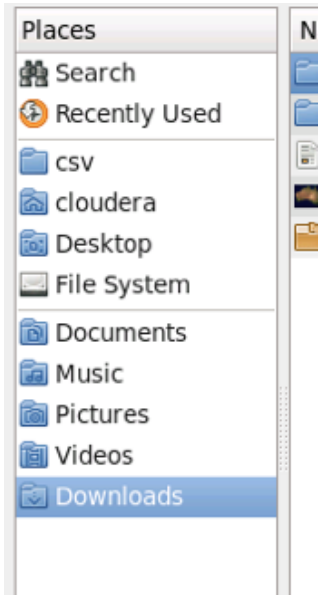
Step 3. Import CSV to spreadsheet. Let's import the CSV file to the spreadsheet by clicking on *File:*



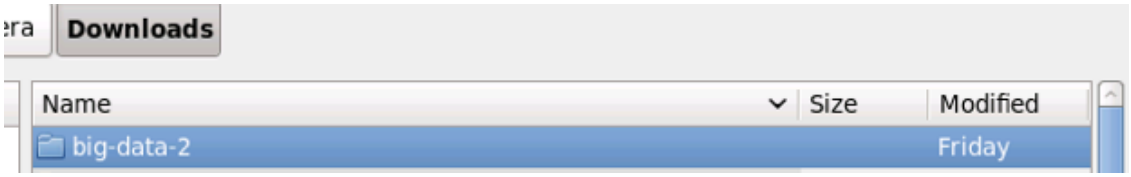
Next, click *Open:*



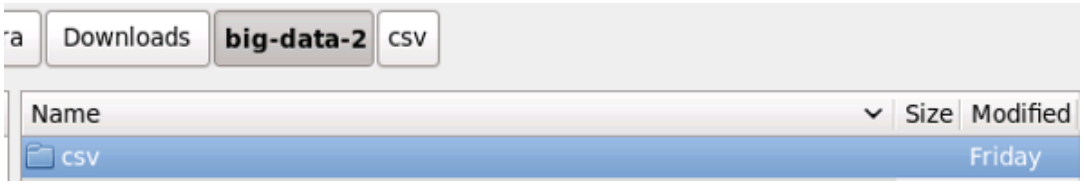
Next, click *Downloads* in the Places pane:



Next, double-click *big-data-2* in the file pane:

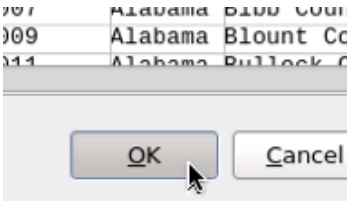


Next, double-click *csv*:



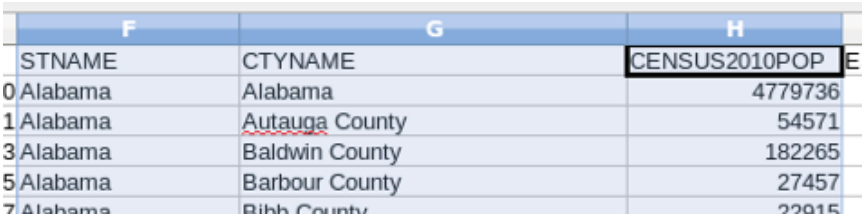
Next, double-click *census.csv*:

In the Text Import dialog, click *OK:*



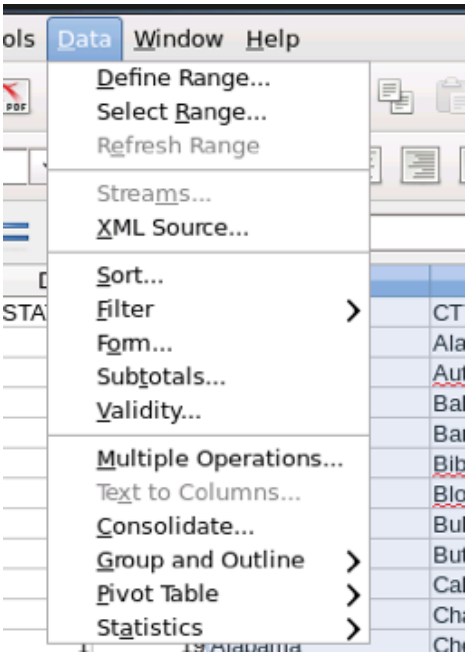
The CSV data is now loaded into the spreadsheet.

Step 4. Filter rows in the data. We can perform a select operation on the data to show only the counties in the state of California containing more than one million people. To perform the select operation, first select the column F, which contains the state name, through column H, which contains the 2010 population:

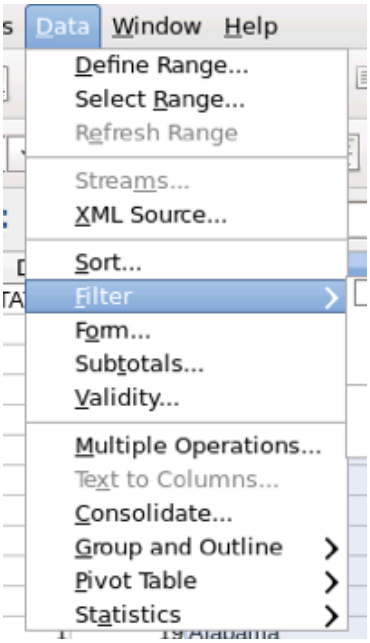


	F	G	H	
	STNAME	CTYNAME	CENSUS2010POP	E
0	Alabama	Alabama	4779736	
1	Alabama	Autauga County	54571	
3	Alabama	Baldwin County	182265	
5	Alabama	Barbour County	27457	

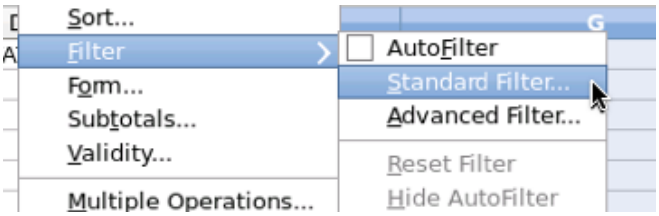
Next, click *Data*:



Next, select *Filter*:



Next, click *Standard Filter*:

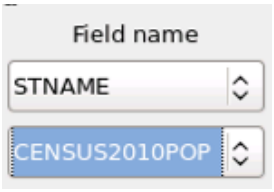


We can create the filter by first changing the Field name to *STNAME*:

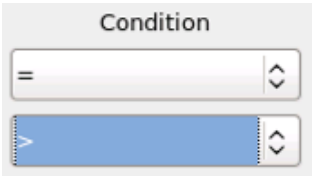
Next, enter *California* in the Value:

Next, change the Operator to *AND*:

Next, change the Field name in the second row to *CENSUS2010POP*:



Next, change the Condition in the second row to *>*:



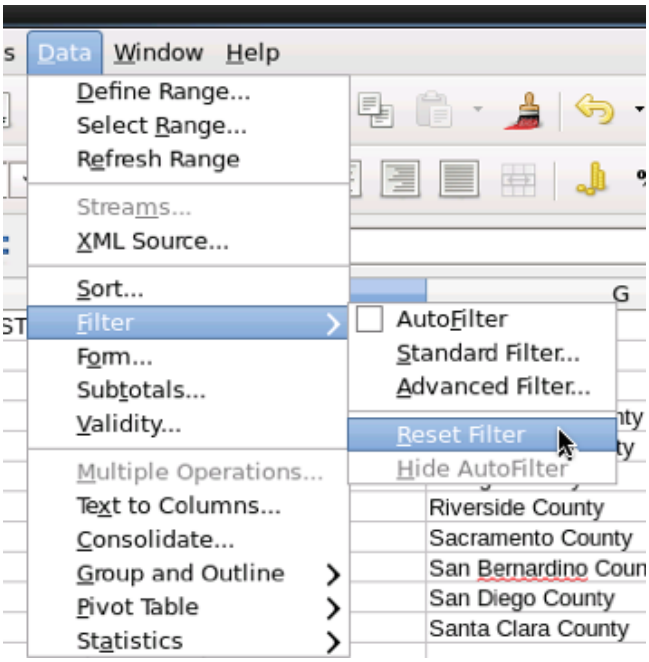
Next, set the Value in the second row to *1000000*:

The dialog should look like this when you're done:

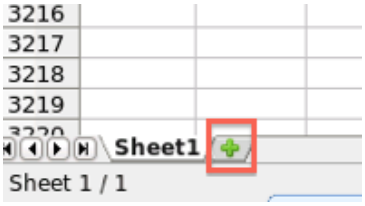
Click *OK* to perform the operation. The spreadsheet will look like the following, showing only the counties in California with a population above one million people:

	F	G	H	
	STNAME	CTYNAME	CENSUS2010POP	ES
0	California	California	37253956	
1	California	Alameda County	1510271	
3	California	Contra Costa County	1049025	
7	California	Los Angeles County	9818605	
9	California	Orange County	3010232	
5	California	Riverside County	2189641	
7	California	Sacramento County	1418788	
1	California	San Bernardino County	2035210	
3	California	San Diego County	3095313	
5	California	Santa Clara County	1781642	

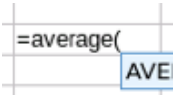
Step 5. Reset the filter. The select operation we performed in the previous steps hides the rows in the data that do not match the criteria we entered in the filter dialog. Let's reset the filter to show all the data. Click on *Data*, *Filter*, and then *Reset Filter*:



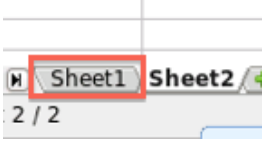
Step 6. Add a new sheet. We can perform aggregate operations on the data in the spreadsheet. Let's display the results of these calculations in a different sheet since Sheet 1 is full of the census data. Create a new sheet by clicking on the green plus button at the bottom left:



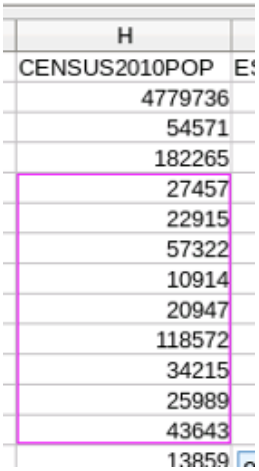
Step 7. Calculate average. Let's calculate the average of several counties in Alabama. First, click on a cell in Sheet 2, and enter *=average(*



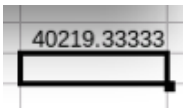
Next, click on Sheet 1:



Select rows 5-13 in column H:

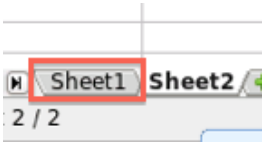


Press the *enter* key to compute the average:

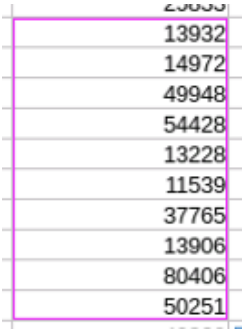


Step 8. Calculate sum. Now let's calculate the sum. Click on an empty cell in Sheet 2 and enter *=sum(*

Click on Sheet 1:



Select rows 16-25 in column H:



Press the *enter* key to compute the sum:

Mark as completed