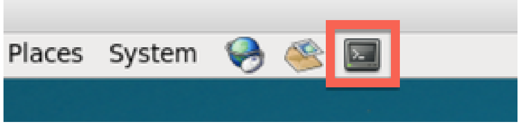
By the end of this activity, you will be able to:

1. Identify the key features in CSV data
2. Import CSV data to a spreadsheet and plot values

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

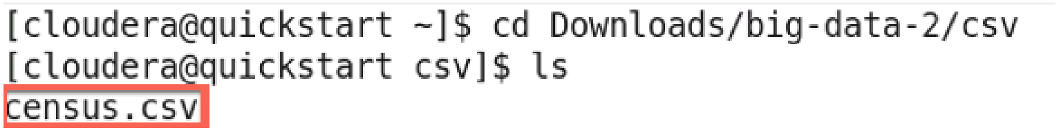


Run *cd Downloads/big-data-2/csv* to change into the directory containing the csv file. (This was downloaded in Week 1 https://www.coursera.org/learn/big-data-management/supplement/YVDPj/instructions-for-downloading-hands-on-datasets)

```
cd Downloads/big-data-2/csv
```

Step 2. **Look at CSV file.** The CSV file contains census data for the United States. Run *ls* to see the name of the CSV file.

```
ls
```

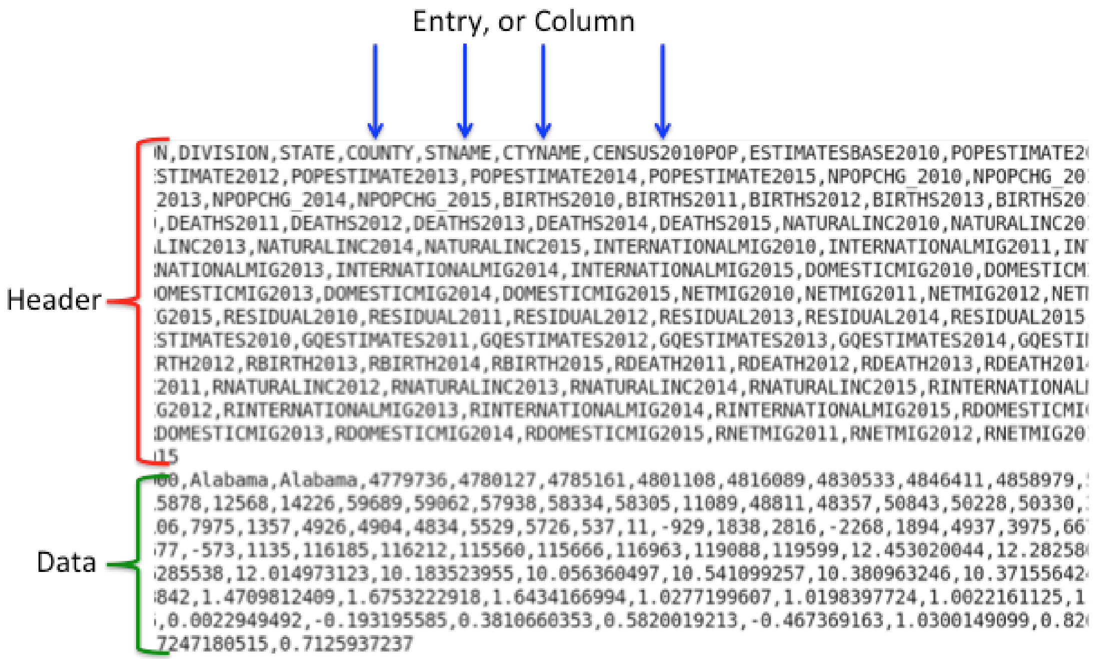


```
[cloudera@quickstart ~]$ cd Downloads/big-data-2/csv
[cloudera@quickstart csv]$ ls
census.csv
```

Run *more census.csv* to look at the contents of the CSV file.

```
more census.csv
```

The first line of the file is the head and the remaining lines are the data. Each entry in the file is separated by a comma.

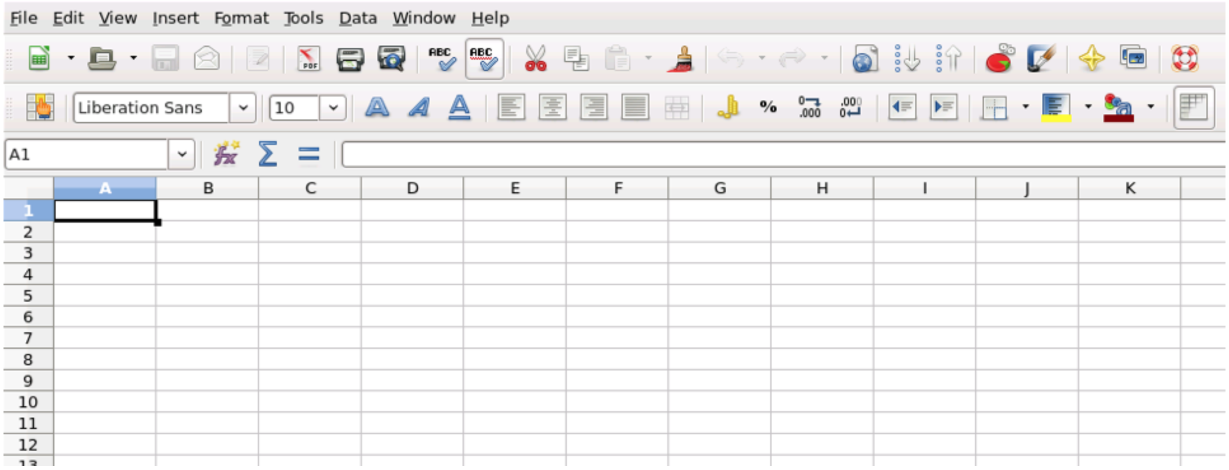


```
N,DIVISION,STATE,COUNTY,STNAME,CTYNAME,CENSUS2010POP,ESTIMATESBASE2010,POPESTIMATE2
STIMATE2012,POPESTIMATE2013,POPESTIMATE2014,POPESTIMATE2015,NPOPCHG_2010,NPOPCHG_20
2013,NPOPCHG_2014,NPOPCHG_2015,BIRTHS2010,BIRTHS2011,BIRTHS2012,BIRTHS2013,BIRTHS20
,DEATHS2011,DEATHS2012,DEATHS2013,DEATHS2014,DEATHS2015,NATURALINC2010,NATURALINC20
LINC2013,NATURALINC2014,NATURALINC2015,INTERNATIONALMIG2010,INTERNATIONALMIG2011,IN
NATIONALMIG2013,INTERNATIONALMIG2014,INTERNATIONALMIG2015,DOMESTICMIG2010,DOMESTICM
OMESTICMIG2013,DOMESTICMIG2014,DOMESTICMIG2015,NETMIG2010,NETMIG2011,NETMIG2012,NET
G2015,RESIDUAL2010,RESIDUAL2011,RESIDUAL2012,RESIDUAL2013,RESIDUAL2014,RESIDUAL2015
STIMATES2010,GQESTIMATES2011,GQESTIMATES2012,GQESTIMATES2013,GQESTIMATES2014,GQESTI
RTH2012,RBIRTH2013,RBIRTH2014,RBIRTH2015,RDEATH2011,RDEATH2012,RDEATH2013,RDEATH201
2011,RNATURALINC2012,RNATURALINC2013,RNATURALINC2014,RNATURALINC2015,RINTERNATIONAL
G2012,RINTERNATIONALMIG2013,RINTERNATIONALMIG2014,RINTERNATIONALMIG2015,RDOMESTICMI
DOMESTICMIG2013,RDOMESTICMIG2014,RDOMESTICMIG2015,RNETMIG2011,RNETMIG2012,RNETMIG20
15
00,Alabama,Alabama,4779736,4780127,4785161,4801108,4816089,4830533,4846411,4858979,
5878,12568,14226,59689,59062,57938,58334,58305,11089,48811,48357,50843,50228,50330,
06,7975,1357,4926,4904,4834,5529,5726,537,11,-929,1838,2816,-2268,1894,4937,3975,66
77,-573,1135,116185,116212,115560,115666,116963,119088,119599,12.453020044,12.28258
285538,12.014973123,10.183523955,10.056360497,10.541099257,10.380963246,10.37155642
842,1.4709812409,1.6753222918,1.6434166994,1.0277199607,1.0198397724,1.0022161125,1
,0.0022949492,-0.193195585,0.3810660353,0.5820019213,-0.467369163,1.0300149099,0.82
7247180515,0.7125937237
```

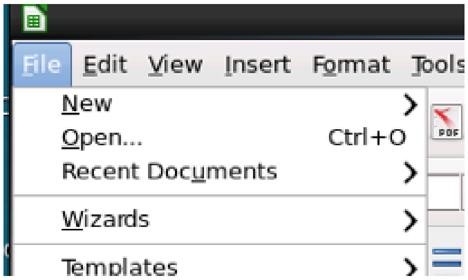
Hit the spacebar to scroll down, and *q* to quit more.

Step 3. **Open spreadsheet application.** Run *oocalc* to start the spreadsheet application.

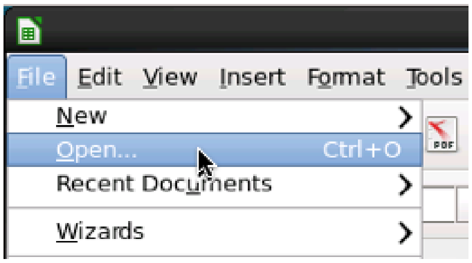
```
oocalc
```



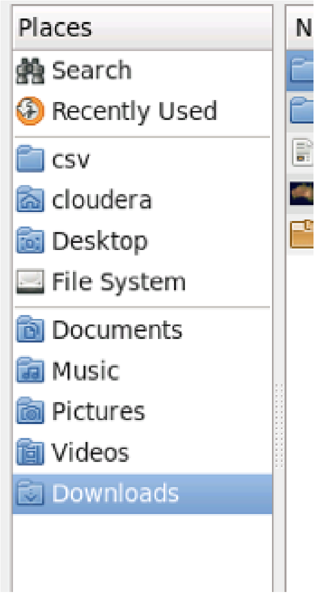
Step 4. **Import CSV to spreadsheet.** Let's import the CSV file to the spreadsheet by clicking on *File:*



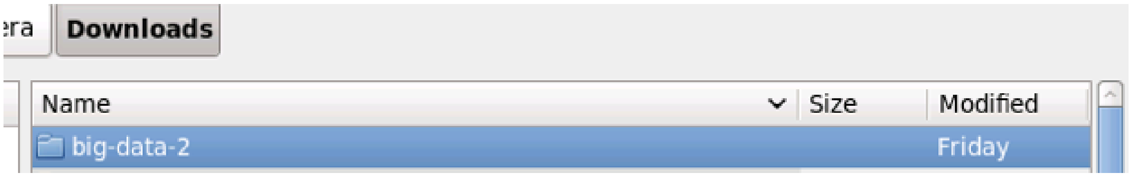
Next, click *Open:*



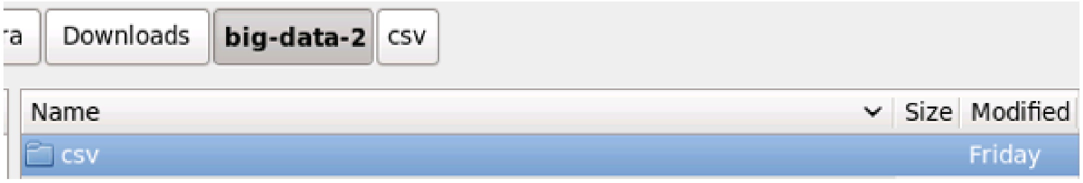
Next, click *Downloads* in the Places pane:



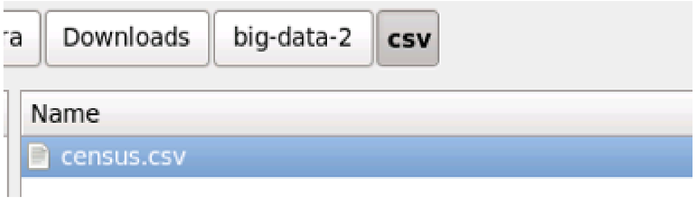
Next, double-click *big-data-2* in the file pane:



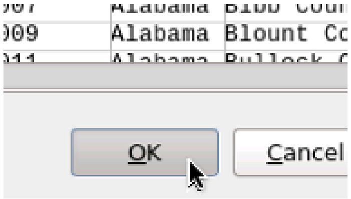
Next, double-click *csv*:



Next, double-click *census.csv*:

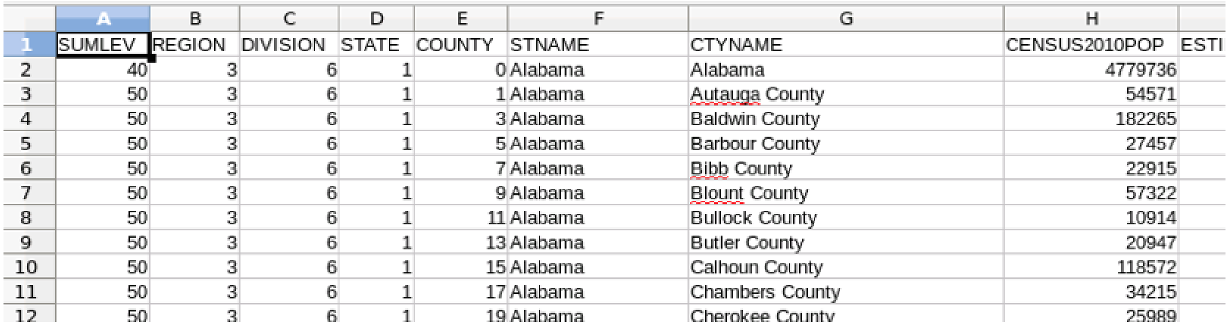


In the Text Import dialog, click *OK:*



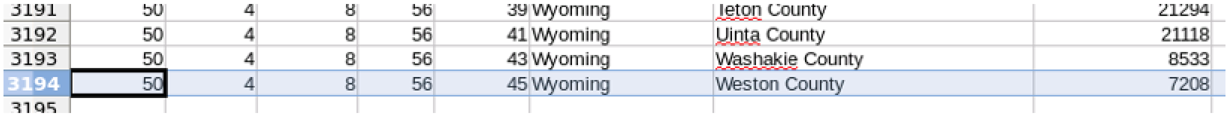
```
907    Alabama Bibb Coun
909    Alabama Blount Co
911    Alabama Bullock C
```

The CSV data is now loaded into the spreadsheet.



	A	B	C	D	E	F	G	H	
1	SUMLEV	REGION	DIVISION	STATE	COUNTY	STNAME	CTYNAME	CENSUS2010POP	ESTI
2	40	3	6	1	0	Alabama	Alabama	4779736	
3	50	3	6	1	1	Alabama	Autauga County	54571	
4	50	3	6	1	3	Alabama	Baldwin County	182265	
5	50	3	6	1	5	Alabama	Barbour County	27457	
6	50	3	6	1	7	Alabama	Bibb County	22915	
7	50	3	6	1	9	Alabama	Blount County	57322	
8	50	3	6	1	11	Alabama	Bullock County	10914	
9	50	3	6	1	13	Alabama	Butler County	20947	
10	50	3	6	1	15	Alabama	Calhoun County	118572	
11	50	3	6	1	17	Alabama	Chambers County	34215	
12	50	3	6	1	19	Alabama	Cherokee County	25989	

Step 5. **See size of CSV.** Scroll to the bottom of the spreadsheet to see the size of the CSV file.

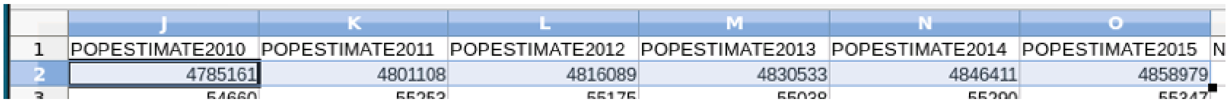


3191	50	4	8	56	39	Wyoming	Teton County	21294
3192	50	4	8	56	41	Wyoming	Uinta County	21118
3193	50	4	8	56	43	Wyoming	Washakie County	8533
3194	50	4	8	56	45	Wyoming	Weston County	7208
3195								

There are 3194 rows. If the CSV file had millions or more rows, then we could not import it into a spreadsheet. In this case, we would need a Big Data system such as Hadoop to analyze the data.

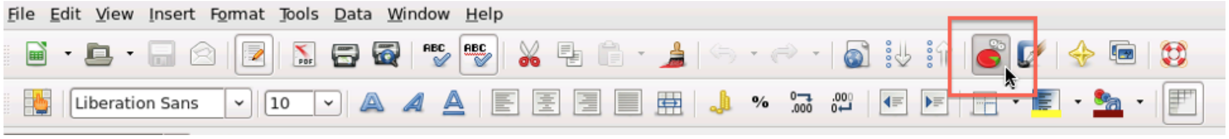
Scroll back to the top.

Step 6. **Create chart.** Let's create a chart of the estimated population of the state of Alabama. Row 2 contains the data for Alabama. Select cells in row 2 and columns J through O to get the estimated population for 2010 through 2015.

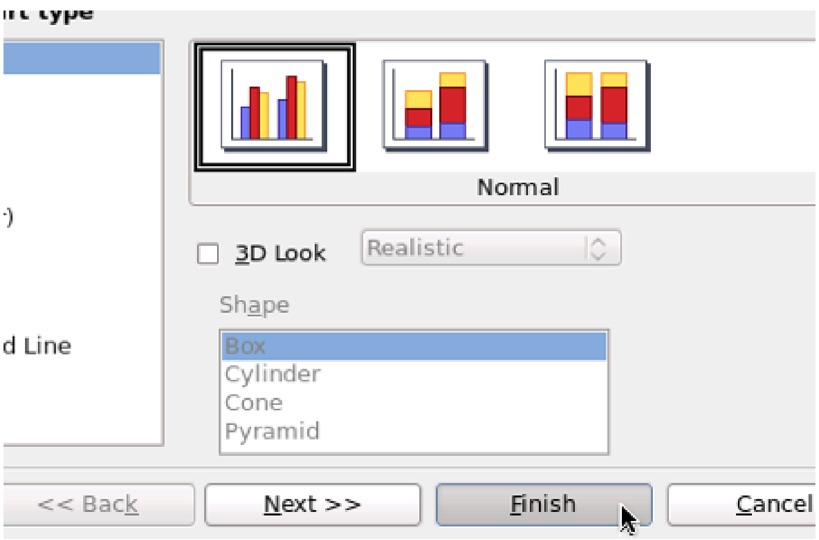


	J	K	L	M	N	O	
1	POPESTIMATE2010	POPESTIMATE2011	POPESTIMATE2012	POPESTIMATE2013	POPESTIMATE2014	POPESTIMATE2015	N
2	4785161	4801108	4816089	4830533	4846411	4858979	
3	54660	55253	55175	55038	55290	55347	

Click on the chart button:



Click *Finish* to display the chart:

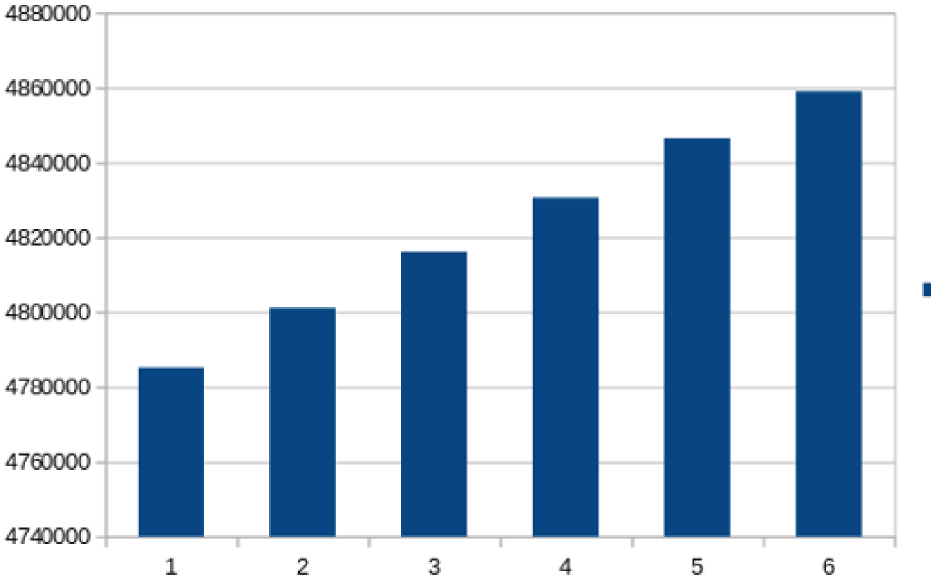


☐ 3D Look

Shape

- Box
- Cylinder
- Cone
- Pyramid

The chart should be displayed in the spreadsheet:



Mark as completed