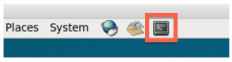
What is a Data Model?
Hands-On
Reading: Introduction to CSV Data
10 min
Video: Introduction to CSV Data
4 min
Different Kinds of Data Models (Part 1)
Video: What is a Relational Data Model?
10 min
Reading: Slides: What Is A Relational Data Model?
10 min
Video: What is a Semistructured Data Model?
6 min
Reading: Slides: What is a Semistructured Data Model?
10 min
Discussion Prompt: Let's discuss: Utilization of XML or JSON on the Internet
5 min
Hands-On

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)

```
1 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.

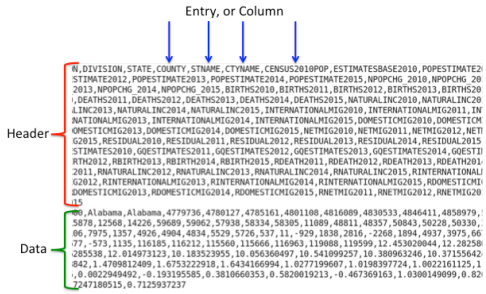
```
1 ls
```

```
[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.

```
1 more census.csv
```

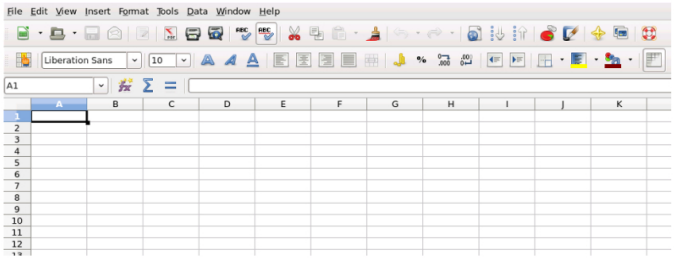
The first line of the file is the head and the remaining lines are the data. Each entry in the file is separated by a comma.



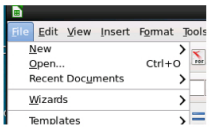
Hit the spacebar to scroll down, and *q* to quit more.

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

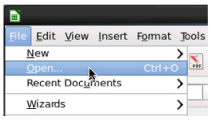
```
1 oocalc
```



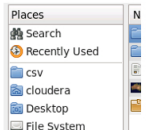
Step 4. **Import CSV to spreadsheet.** Let's import the CSV file to the spreadsheet by clicking on *File:*

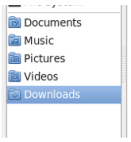


Next, click *Open:*

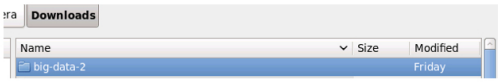


Next, click *Downloads* in the Places pane:

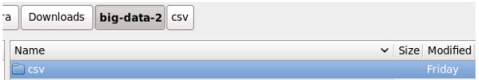




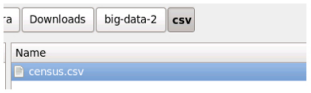
Next, double-click *big-data-2* in the file pane:



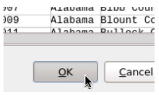
Next, double-click *csv*:



Next, double-click *census.csv*:



In the Text Import dialog, click *OK*:



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.

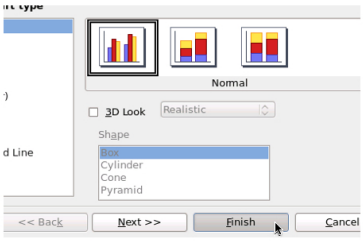
Scroll back to the top.

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.

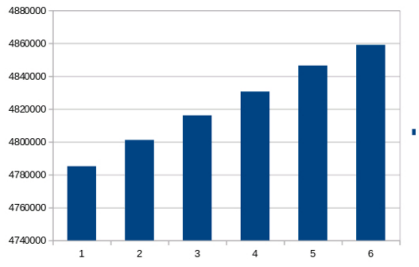
	J	K	L	M	N	O	
1	POPESTIMATE2010	POPESTIMATE2011	POPESTIMATE2012	POPESTIMATE2013	POPESTIMATE2014	POPESTIMATE2015	N
2	4785161	4801108	4816089	4830533	4846411	4858979	
3							

Click on the chart button:

Click *Finish* to display the chart:



The chart should be displayed in the spreadsheet:



Mark as completed