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

- Import CSV files into Splunk.
- Query, filter, and plot data.
- Perform statistical calculations.

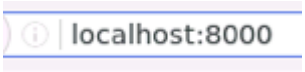
NOTE: Steps 4 and 5 below contain examples using the 'sort' command which are not covered in the video lecture but which will be covered in the accompanying quiz.

The Census CSV data used in this activity can be downloaded here:

census.zip

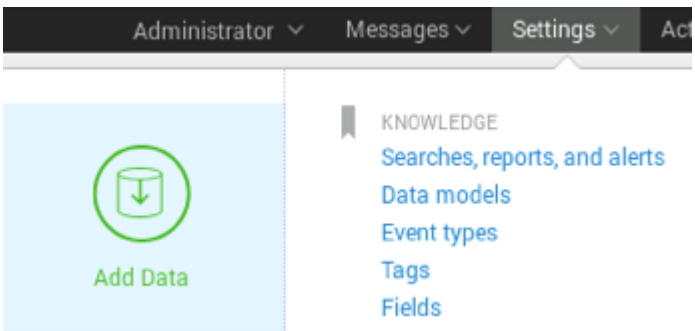
After downloading, unzip the file.

Step 1. **Login to Splunk.** Open a web browser and navigate to *localhost:8000*:

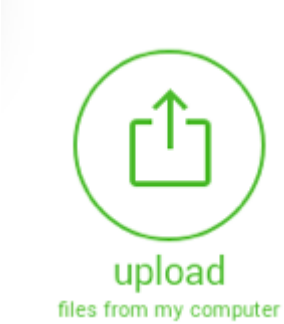


Next, login to Splunk by enter *admin* and the default password *changeme*:

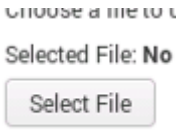
Step 2. **Import census data**. Let's import the census data CSV file to Splunk. First, click on *Settings* in the top right, then click on *Add Data*:



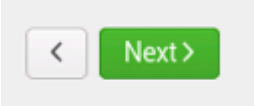
Next, click on *Upload*:



Click on *Select File:*

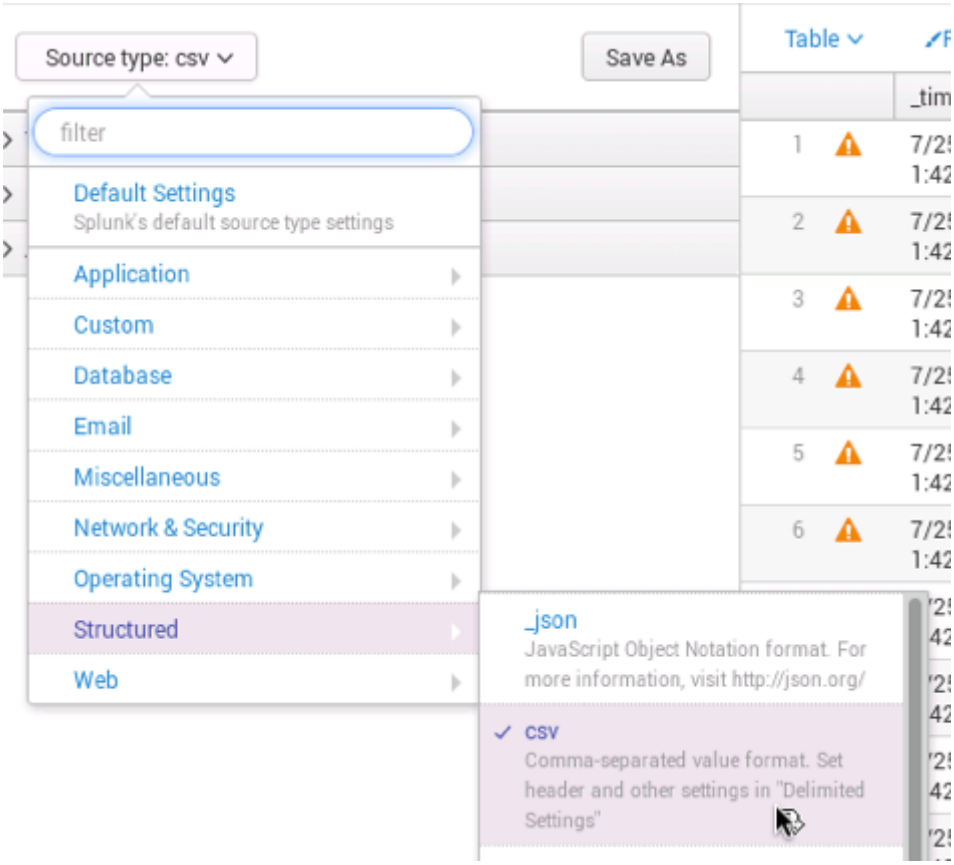


Navigate to *census.csv,* and select it. Then click *Next>*:



On the left, make sure the *Source type* is *csv*:

If the *Source type* is not *csv*, click on *Source type,* go down to *Structured*, and select *csv*:



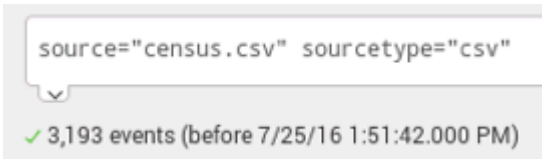
The table on the right is a preview of the CSV data. It shows the values for different fields:

	_time	BIRTHS2010	BIRTHS2011	BIRTHS2012	BIRTHS2013	BIRTHS2014	BIRTHS2015	CENSUS2010POP
1	7/25/16 1:42:30.000 PM	14226	59689	59062	57938	58334	58305	4779736
2	7/25/16 1:42:30.000 PM	151	636	615	574	623	600	54571
3	7/25/16 1:42:30.000 PM	517	2187	2092	2160	2186	2240	182265

Next, click on *Next,* click on *Review*, and then click *Submit.* Splunk will say the file is successfully uploaded:

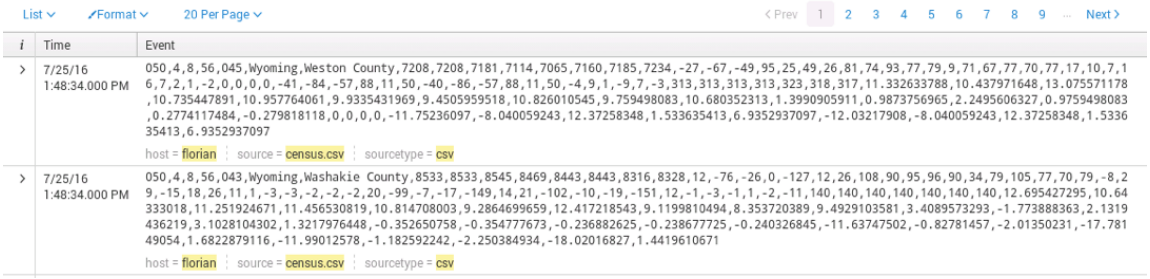
Step 3. **View data**. Click on *Start Searching*:

Splunk will enter a default query in the search box to show the data we just imported:



This query shows all the data from the *census.csv* file and whose data type is CSV. In general, we can use *source=* to query from different file names, and *sourcetype=* to query from different formats.

The table shows the results matching this query:



Step 4. **Filtering for specific values**. We can filter the results by looking for a field with a specific value. For example, we can find the entries where the state is California:

```
STNAME="California"
```

The field *STNAME* contains the name of the state, and the above query only shows the results where the state is California. We can use an *OR* to search for multiple values on the same field:

```
STNAME="California" OR STNAME="Alaska"
```

We can search multiple fields with specific values by adding them to query. For example, let's search for state name California and 2010 population greater than one million people:

```
STNAME="California" CENSUS2010POP > 1000000
```

We can filter our results to just show a single column. For example, let's just show the county names of the previous query:

```
STNAME="California" CENSUS2010POP > 1000000 | table CTYNAME
```

The | (pipe) is the syntax for sending the results from one query to the next, and the *table* command creates a table using only the specified column name(s).

We can sort the results based on any of the fields, such as population, and order them in either ascending or descending order. The image below shows an example of a search for all items with a population greater than 100000, sorts the results in

descending order, and creates a table containing the population and state name. [To sort in ascending order you would replace "desc" with "asc"].

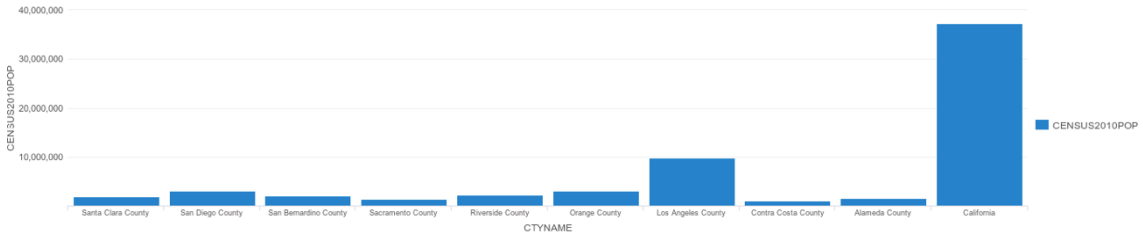
```
CENSUS2010POP > 100000 | sort CENSUS2010POP desc | table CENSUS2010POP,STNAME
```

Instead of using "desc" you can use a dash before the sorting field, e.g. "...| sort -CENSUS2010POP | table ..." for the above query.

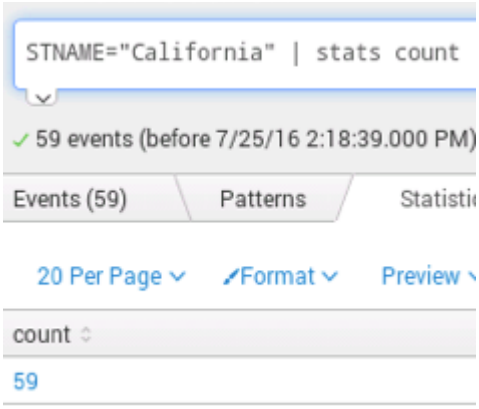
We can view plots of search results by clicking on the *Visualization* tab. For example, if we use our last query and add the 2010 population value to the table:

```
STNAME="California" CENSUS2010POP > 1000000 | table CTYNAME, CENSUS2010POP
```

We can click on the *Visualization* tab to see a chart of the results:

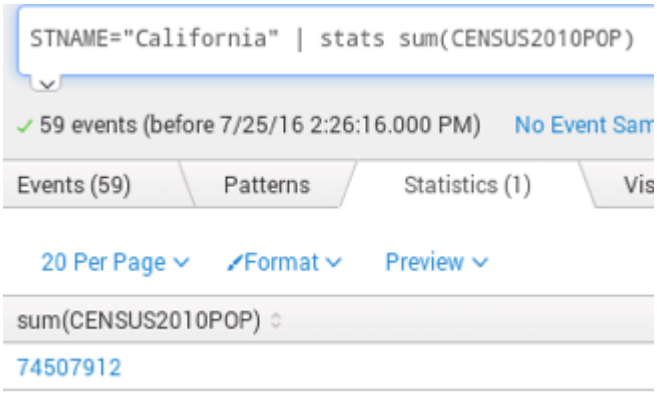


Step 5. **Perform statistical calculations**. The Splunk *stats* command is used to perform statistical calculations on the data. Let's count the results where the state is California:

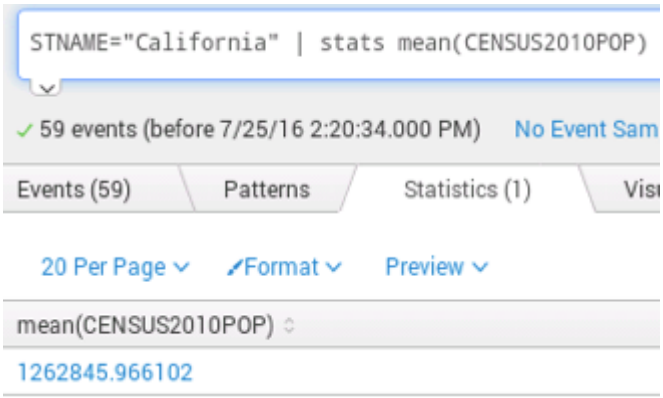


We can sort based on the count by adding "| sort count" to the above query. This would sort in ascending order. if we want to sort in descending order we would use " | sort -count".

Next, let's compute the total 2010 population for California:



Finally, let's compute the average 2010 population for California:



Marcar como concluído