universidad cenfotec_ tecnologías digitales	**Universidad Cenfotec** ***Gestión de Datos para la Toma de Decisiones***

Oracle Data Visualization Desktop

Purpose

There are three sections in this tutorial:

1. Creating a Data Visualization with Sample Data
2. Adding Data Sources from Comma Separated Value Files
3. Creating a Data Flow

Background

Oracle Data Visualization makes rich, powerful visual analytics accessible to every business user. People at all levels of an organization can blend and analyze data in just a few clicks, effectively sifting through data clutter to quickly uncover and share hidden patterns and actionable insights.

You begin by creating a project in Oracle Data Visualization with sample data sources. Then you create visualization, modify the visualization by moving data elements into and out of the canvas, manually change the visualization type, and apply filters. You also learn how to navigate and adjust the canvas layout, and how to find, organize, and manage content.

What you need?

For this tutorial we need the following data files:

- Income.xlsx
- People.xlsx

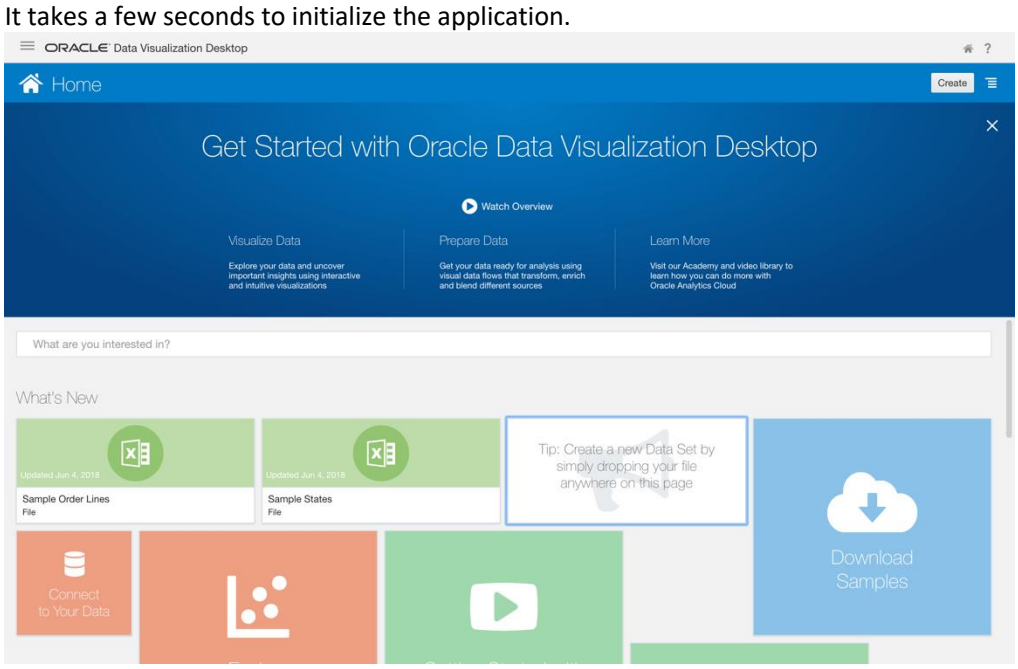
What to send as a lab result?

- Compress the following folder and send it.
  - C:\Users\<username>\AppData\Local\DVDesktop\service_instances\ssi\metadata\content\catalog\root\users\weblogic
- You can find this location browsing the C: unit and enter **People and Income Viz**.
- This step can be done until you finish the lab.

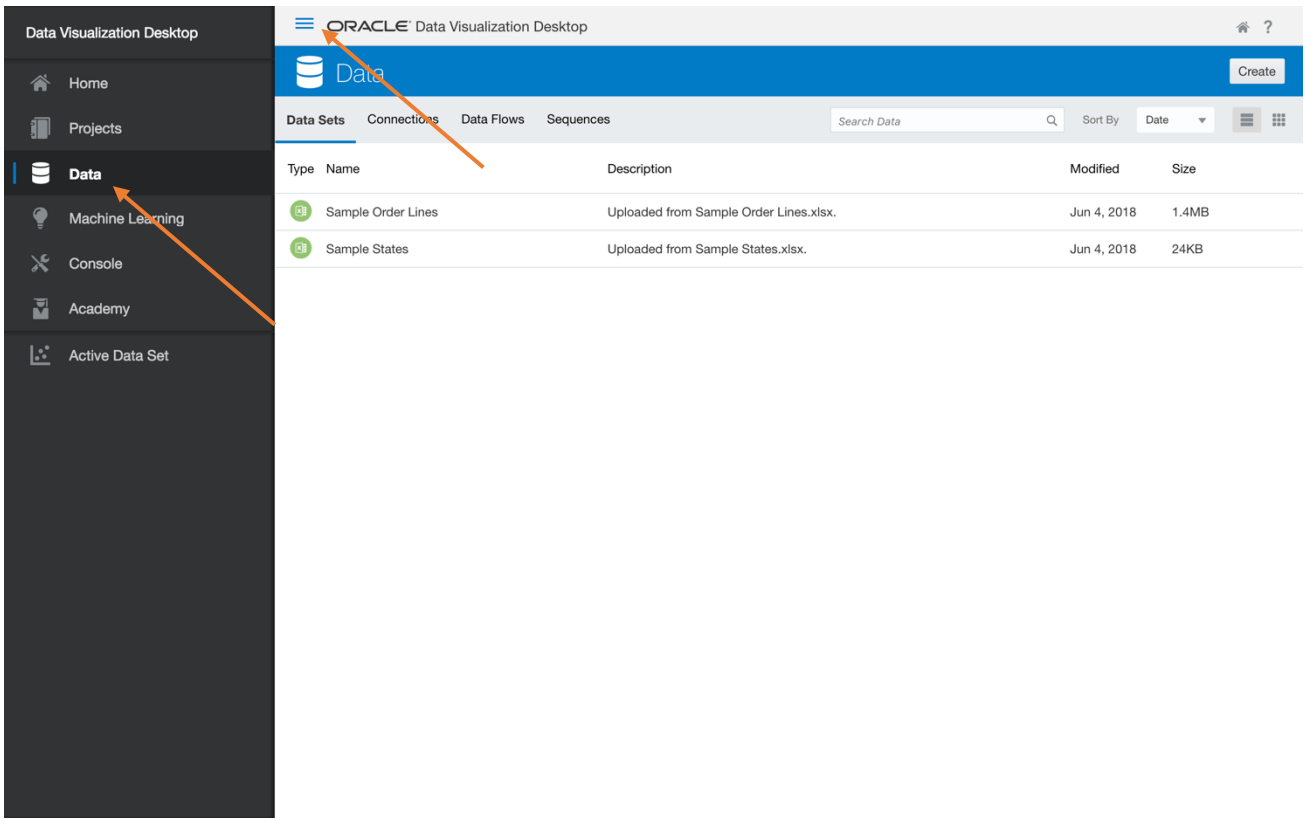
SECTION 1: Creating a Data Visualization with Sample Data

Double click the Oracle Data Visualization Desktop 12c icon to start the application.

It takes a few seconds to initialize the application.

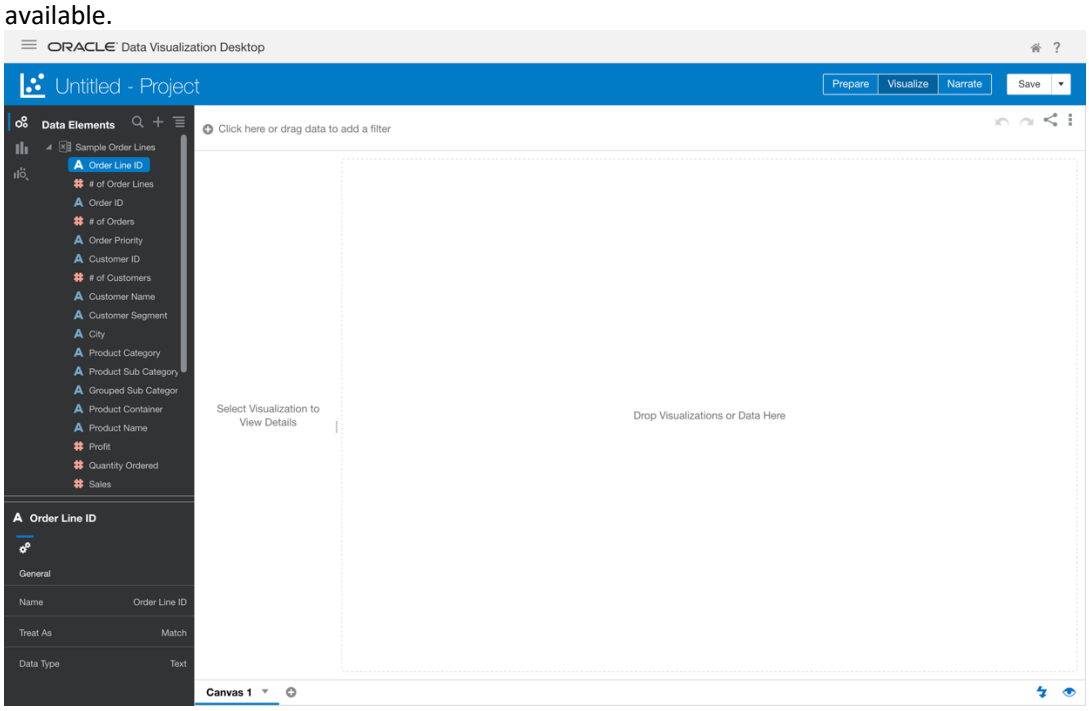


Before you can begin to explore data in a new project, you must select a data source for that information. Click **Navigator** > **Data** to display the existing data sources.

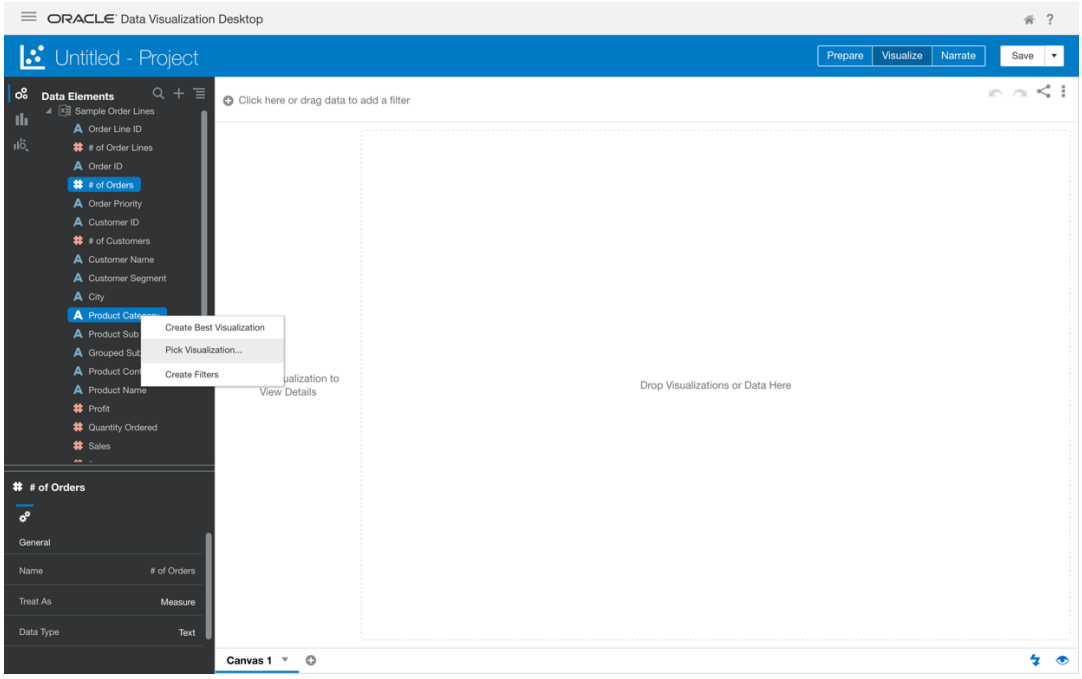


You can select Oracle Applications, databases, or uploaded data files as your data sources. We use the sample data source provided to discover the capabilities of Oracle Data Visualization Desktop in this step. The sample data set is based on Sales Orders data and contains meaningful measures and attributes.

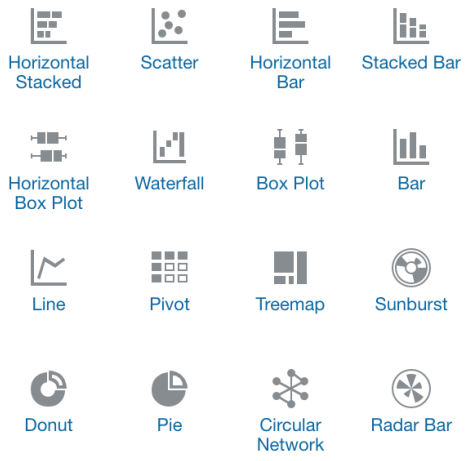
Click **Sample Order Lines**. It brings you into the Data Elements pane with all the data elements in the sample data available.



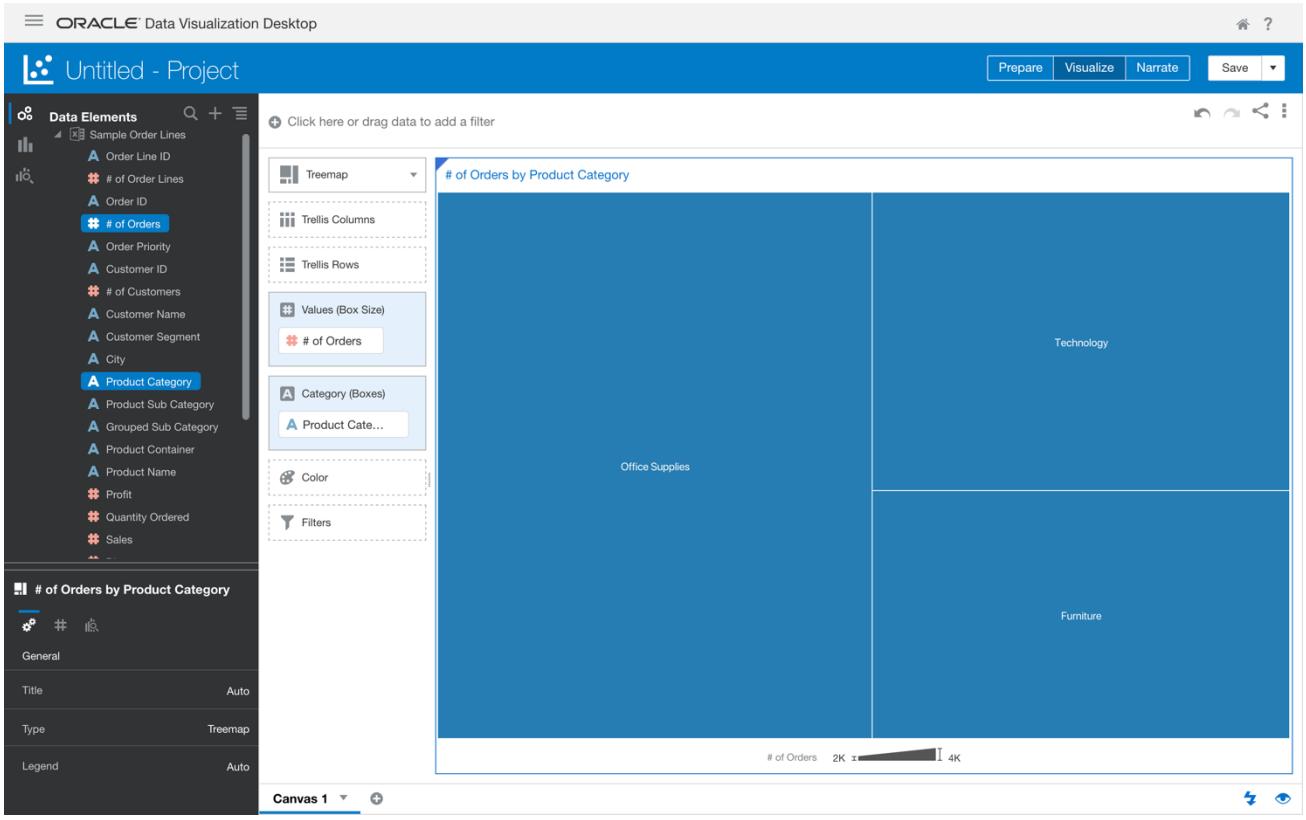
After you select the data sources for your project, you can begin to add data elements such as measures and attributes to visualizations. You can see the list of data elements on the left.



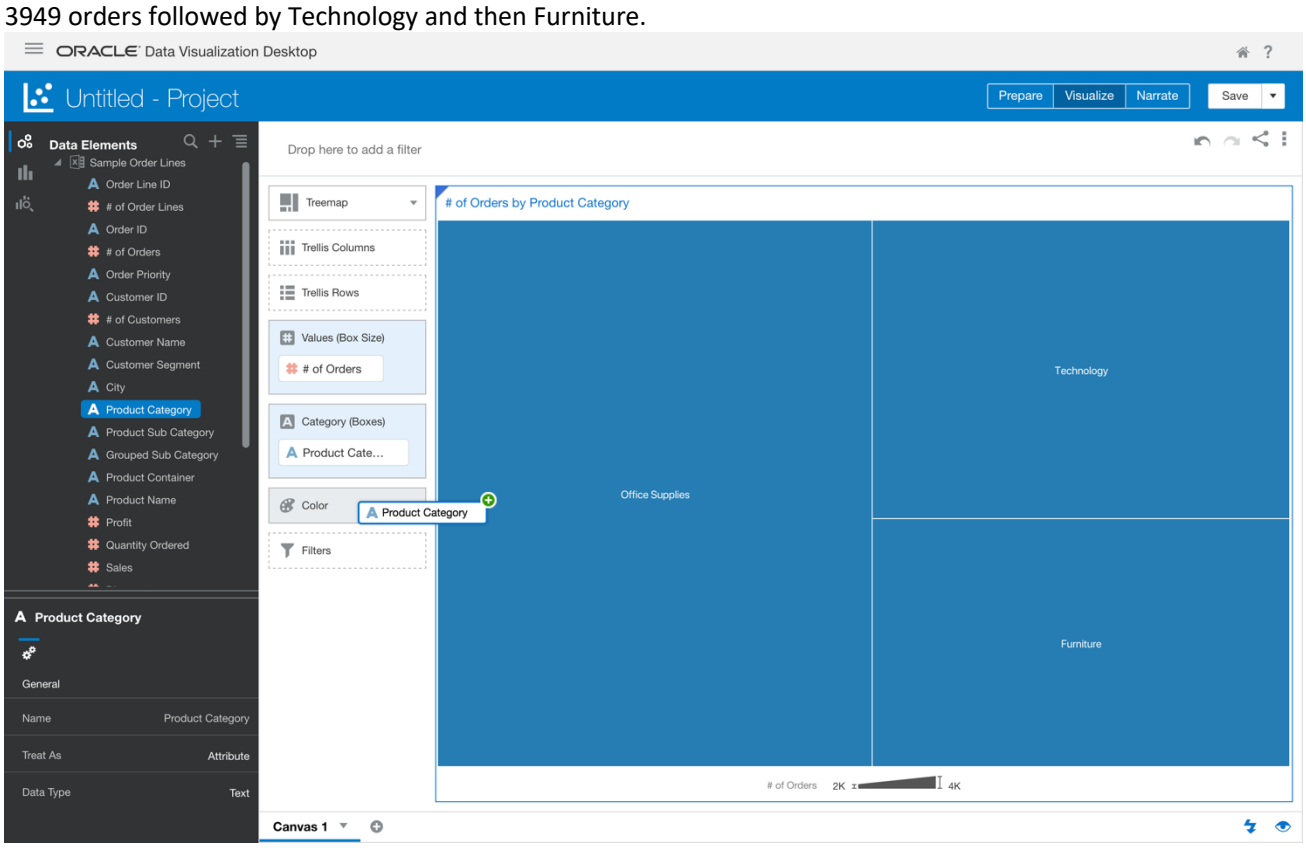
Select both **Product Category** and **# of Orders** and then right click to show the right-click menu. Select **Pick Visualization**... on the pop-up menu.



In the View Select dialog, select the **Treemap** visualization type.

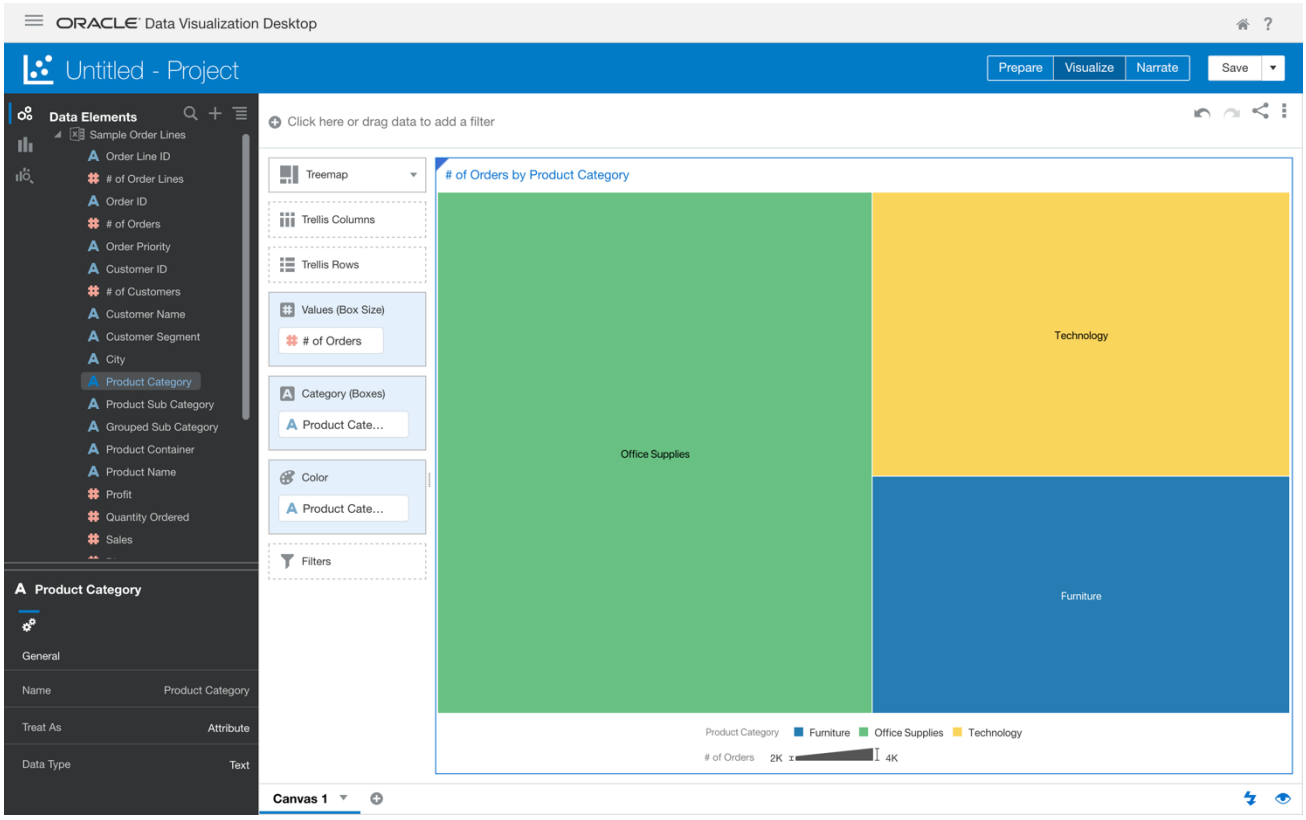


A tree map of # of Orders by Product Category is generated. You can see the largest one is Office Supplies with 3949 orders followed by Technology and then Furniture.

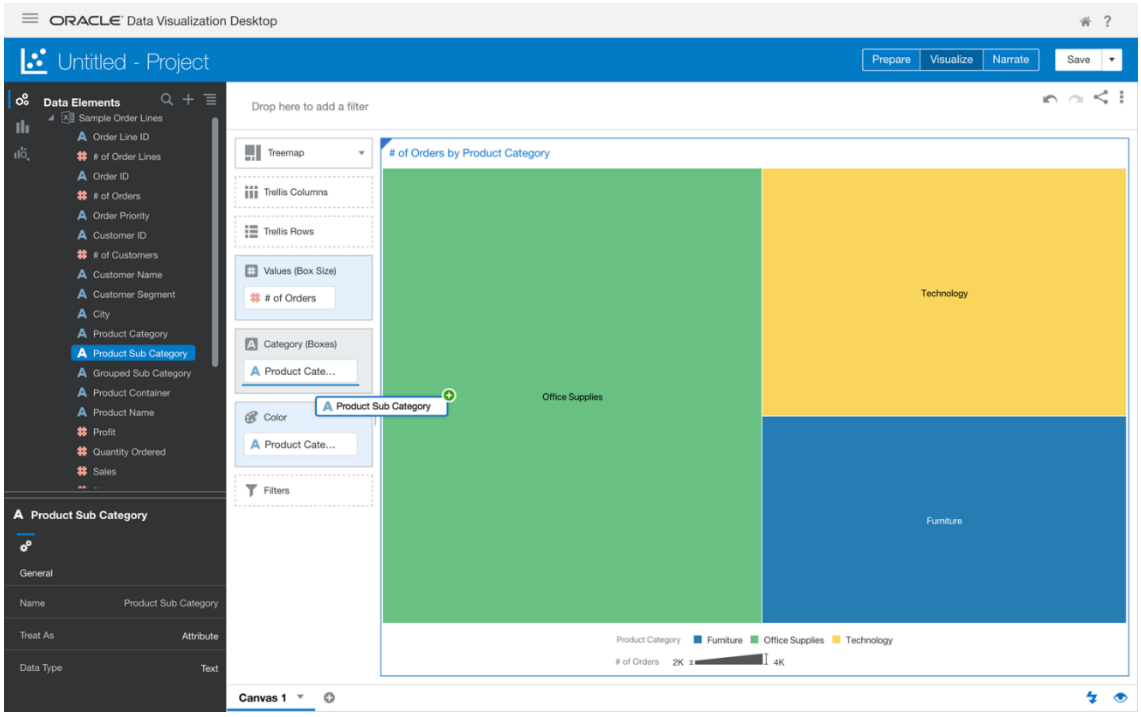


You can work with color to make visualizations more attractive, dynamic, and informative. The Visualize canvas has a Color drop target where you can put a measure column, attribute column, or set of attributes columns.

Select **Product Category** on the left, drag and drop it into the Color drop target.

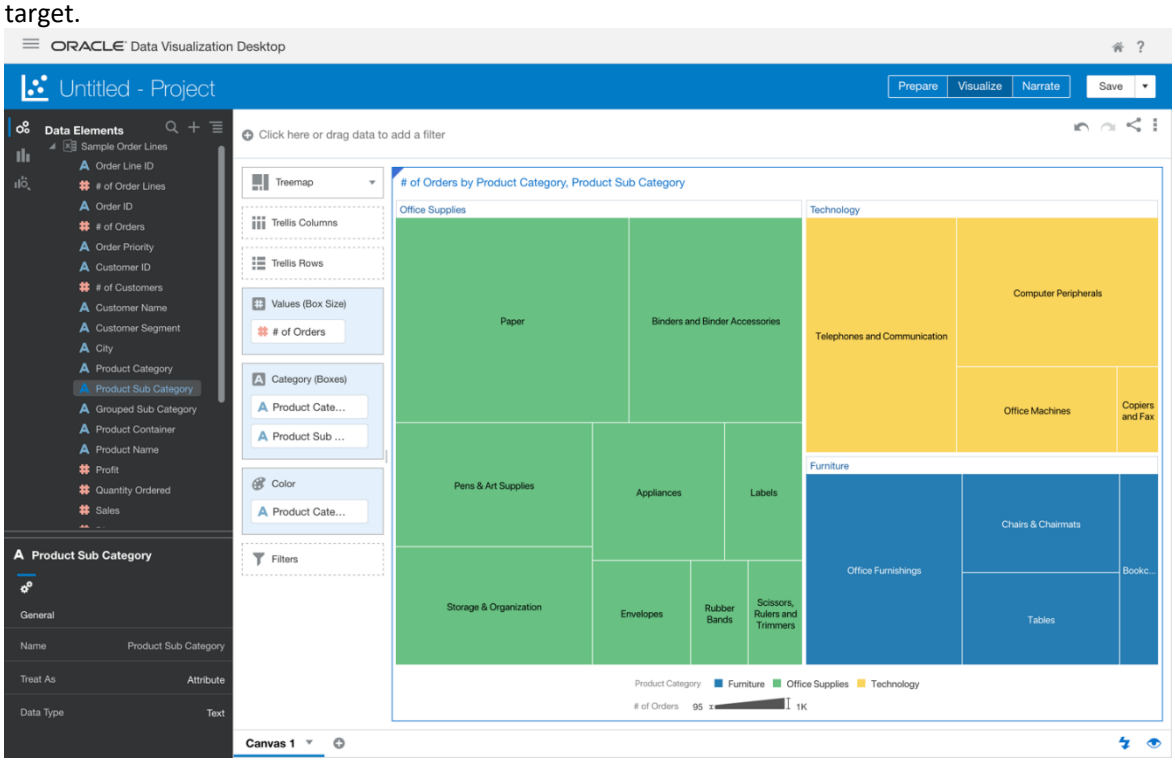


The blocks in the tree map are colored in different colors. It is easier to distinguish the differences among the product categories.

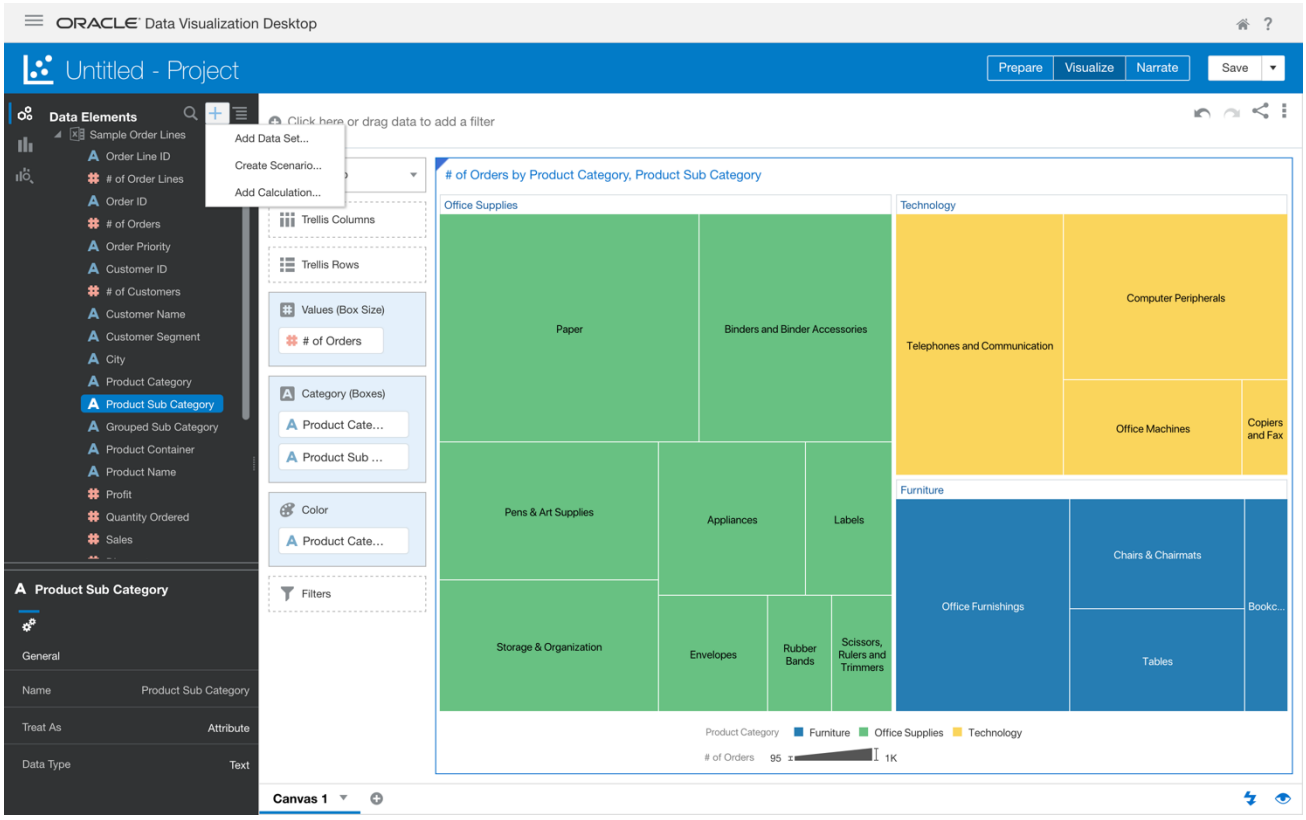


You might want to add Product Sub Category information to the tree map as well. It might help you understand the sales better.

Select **Product Sub Category** on the left, drag and drop it **under** the **Product Category** in the **Category** drop target.

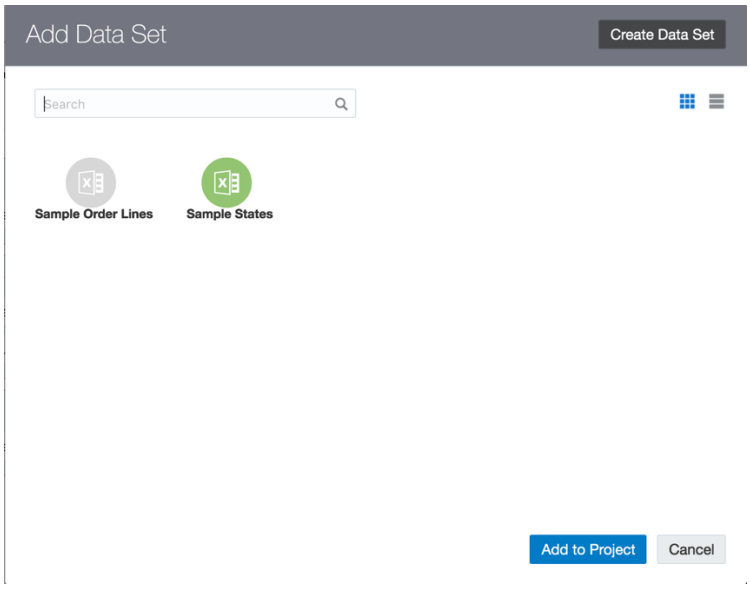


You can see there is a sub tree map in each product category block for you to drill down into each sub product category.

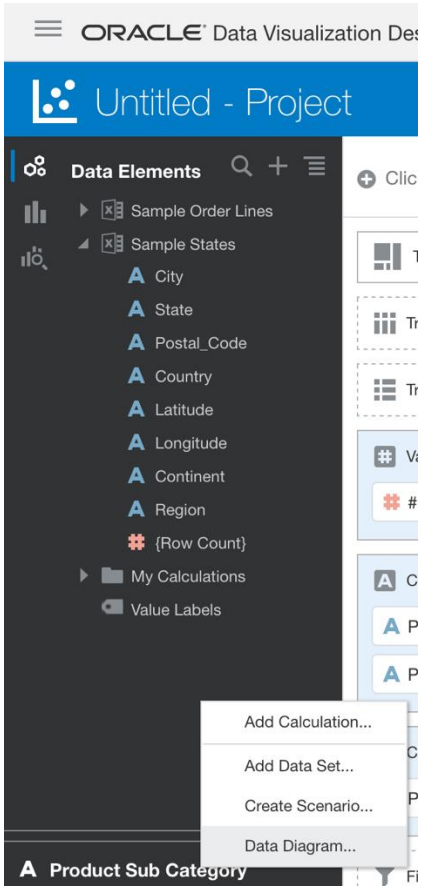


Now let's add a new visualization – a map of sales by states.

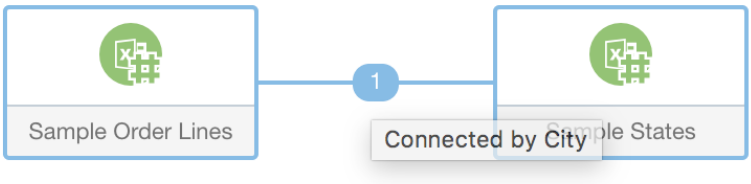
Click in the + Button on the pane on the left, select Add Data Set in the pop-up menu.



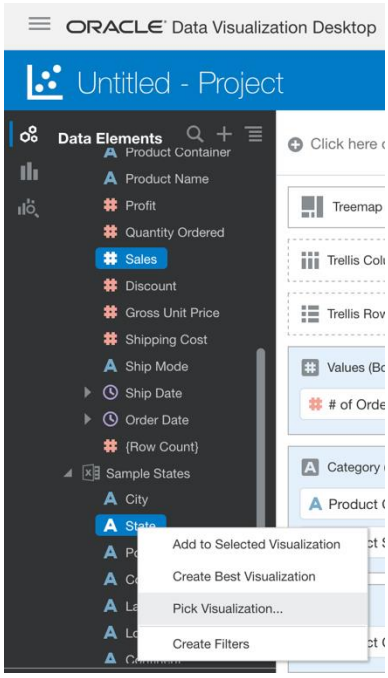
Select **Sample States**, click **Add to Project** in the Add Data Source dialog.



You can see the elements in the Sample States data source are expended in the Data Elements pane. Right click in the black area of Data Elements pane, select **Data Diagram**… in the pop-up menu.

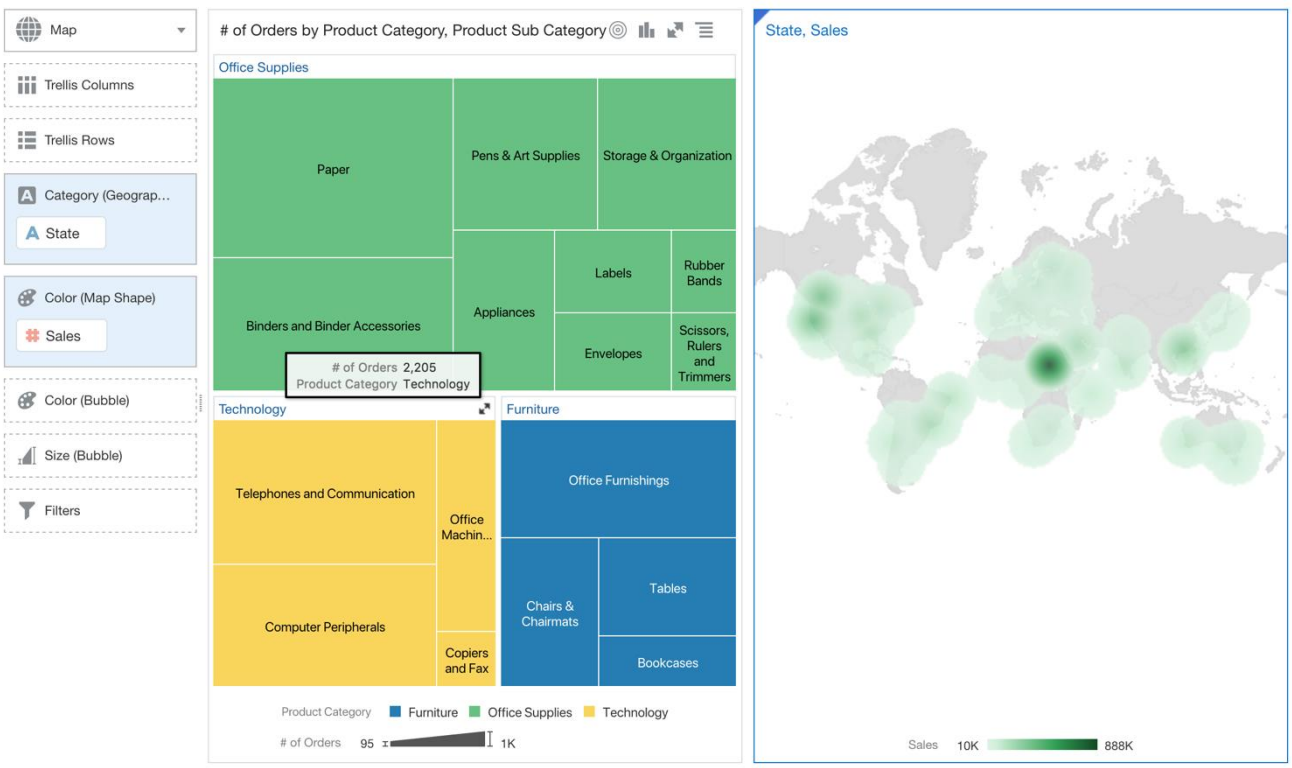


You can see that the Sample Order Lines and the Sample States are connected by City, which is an element in common in both of the data sources.

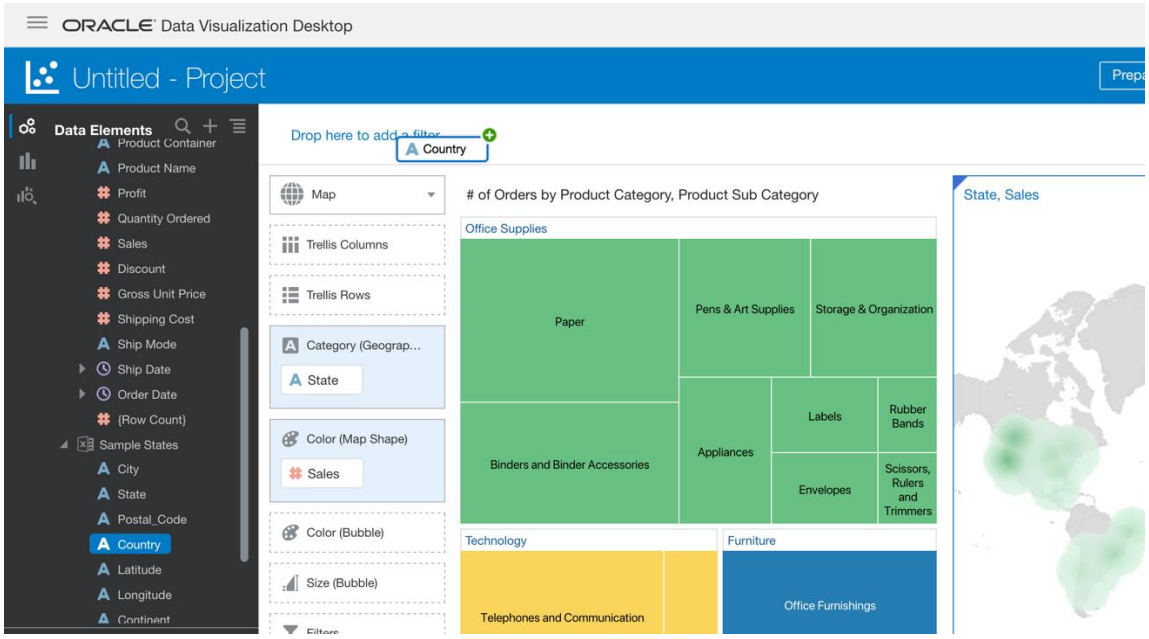


Select both **Sales** in the Sample Order Line and **State** in the Sample States, right click and select **Pick Visualization...** on the pop-up menu.

In the View Select dialog, select the **Map** visualization type.



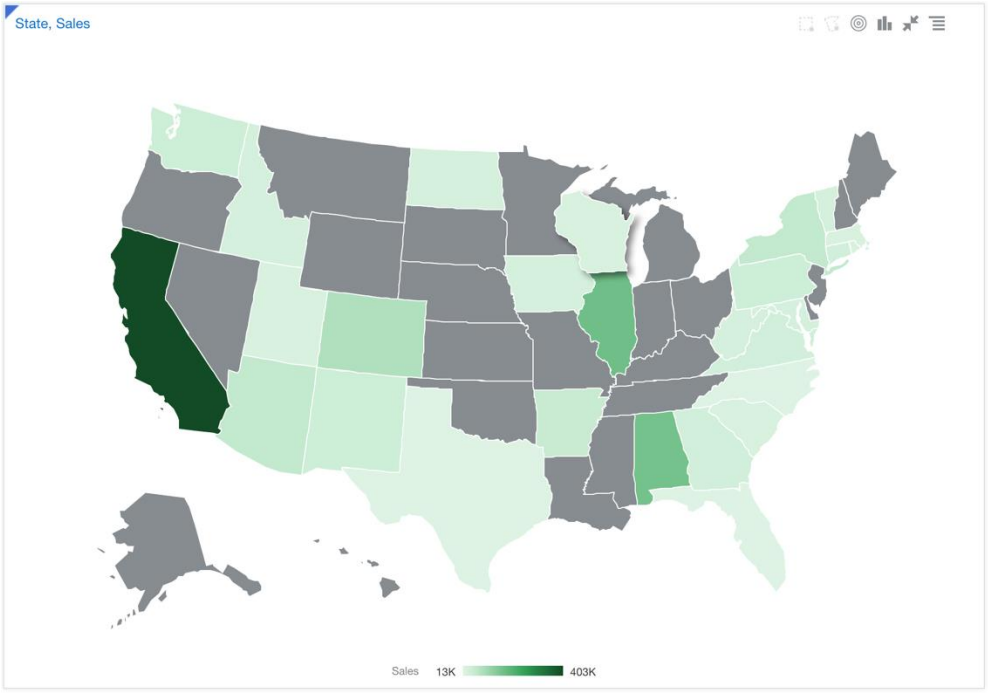
A map of sales by states is added to the right of tree map.



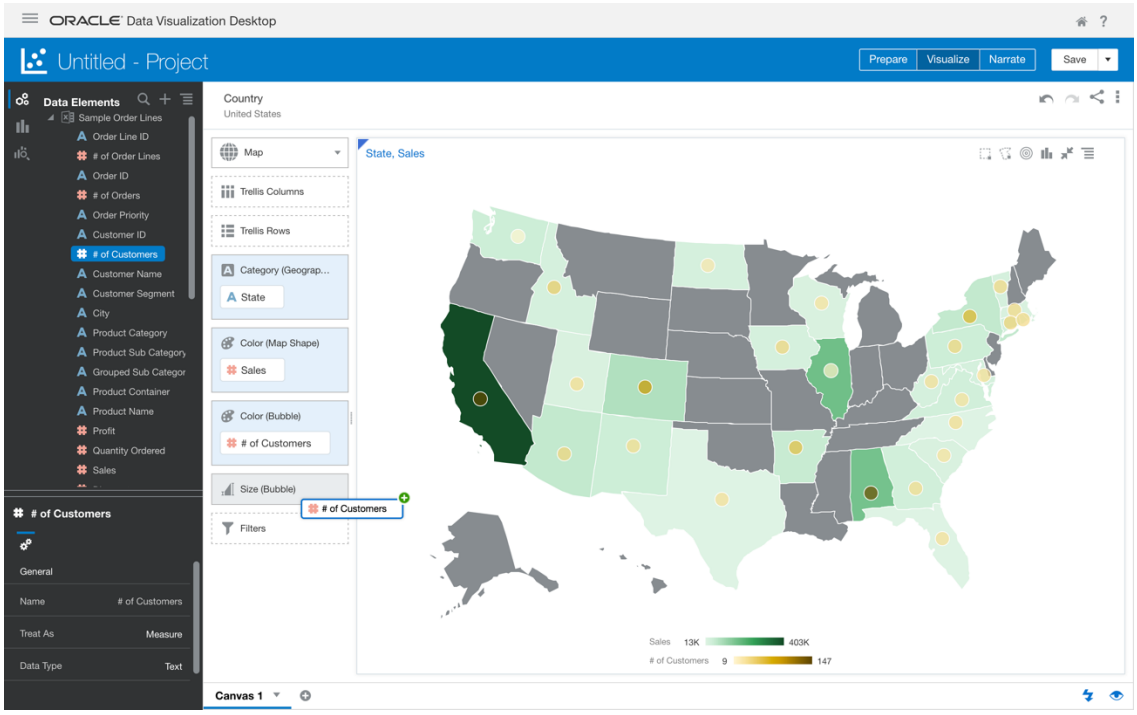
To make the map more interactive, we are going to do a couple of things. First, select **Country** in the Sample States on the left and drag and drop it to the filter bar area on the top.

In the pop-up country list, select **United States**.

The map is focused on United States area. Click the **Maximize the View** icon on the top-right corner of the map.



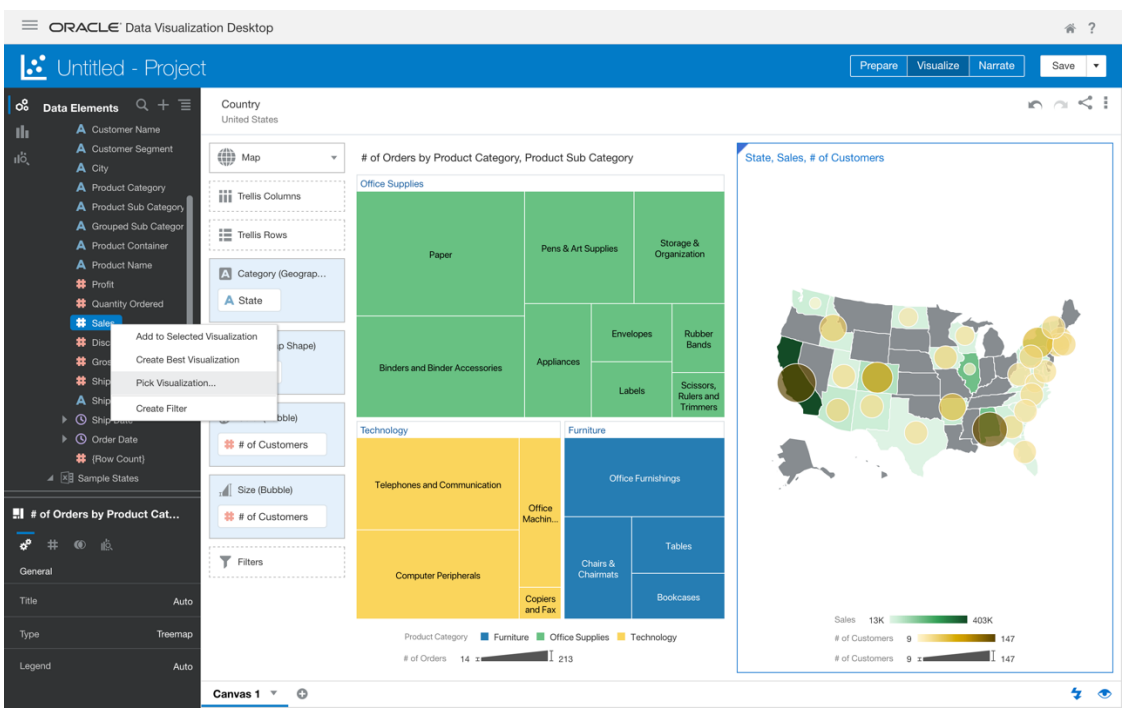
You can get a better view of sales by state in the United States. The revenue in the states range from 13K to 403K, colored from light green to deep green. Obviously, California has the largest revenue. The grey color indicates where there is no revenue.



Select **# of Customers** on the left and drag and drop it to the **Color (Bubble)** drop target and **Size (Bubble)** drop target respectively.

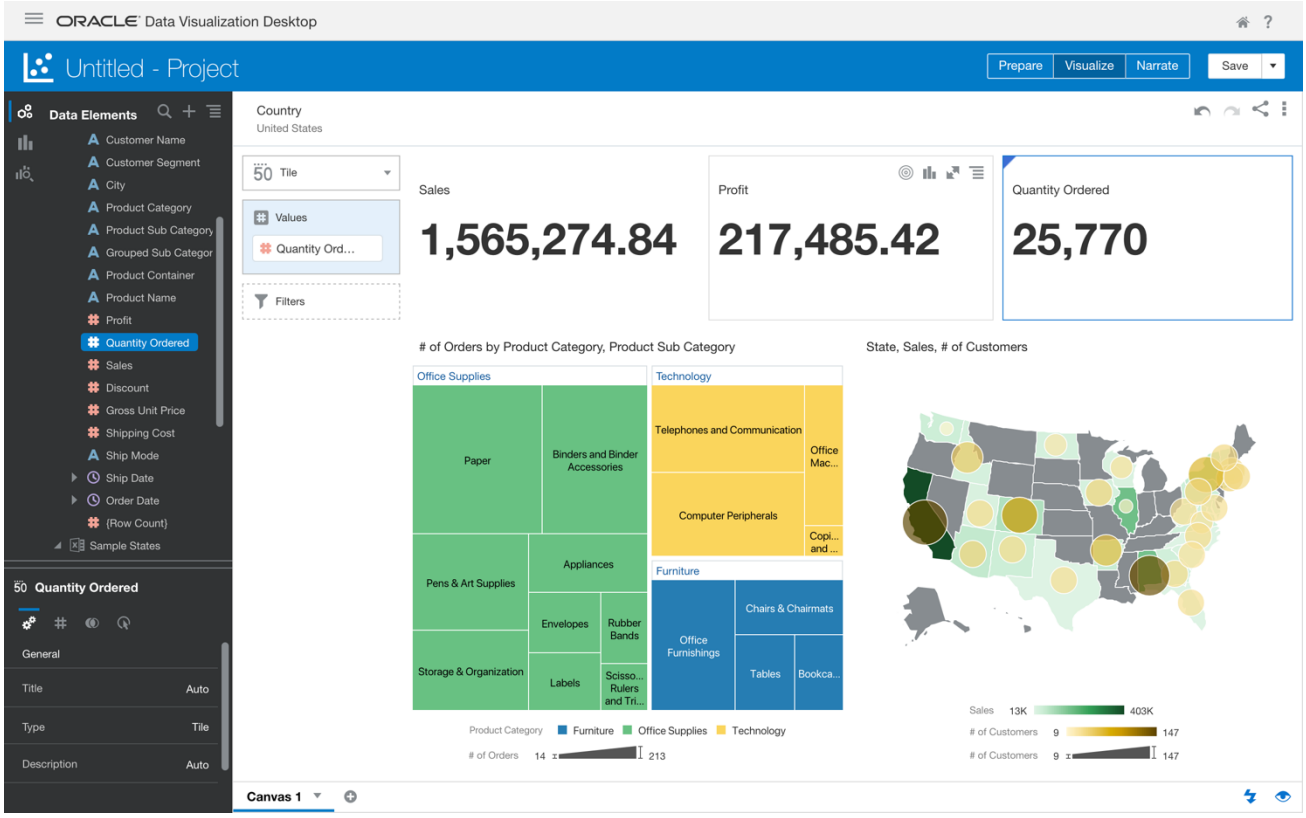
The more customers a state has, the deeper color and larger bubble it has.

Click the **Close Maximized Visualization** icon in the top-right corner of the map.



Select **Sales** in the Sample Order Line list on the left, right click it and select **Pick Visualization...** on the pop-up menu.

In the View Select dialog, select the **Tile** visualization type.

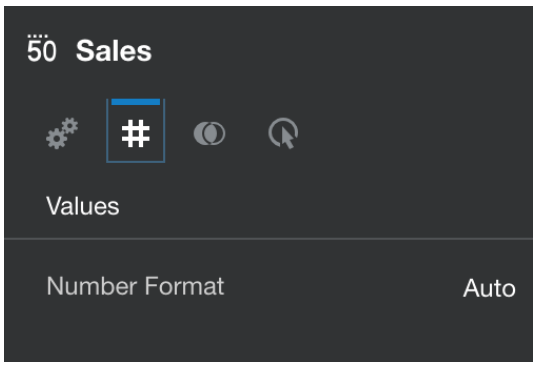


Drag and drop the Profit tile to the top after it is displayed at the bottom.

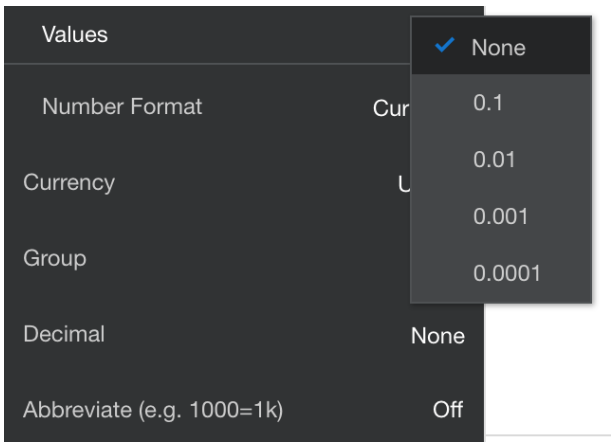
Create the **Profit** tile and **Quantity Ordered** tile respectively in the same way.

You can adjust the positions of the tiles by drag and drop.

Click the **Sales** tile and select in the Properties panel the **Values** tab.

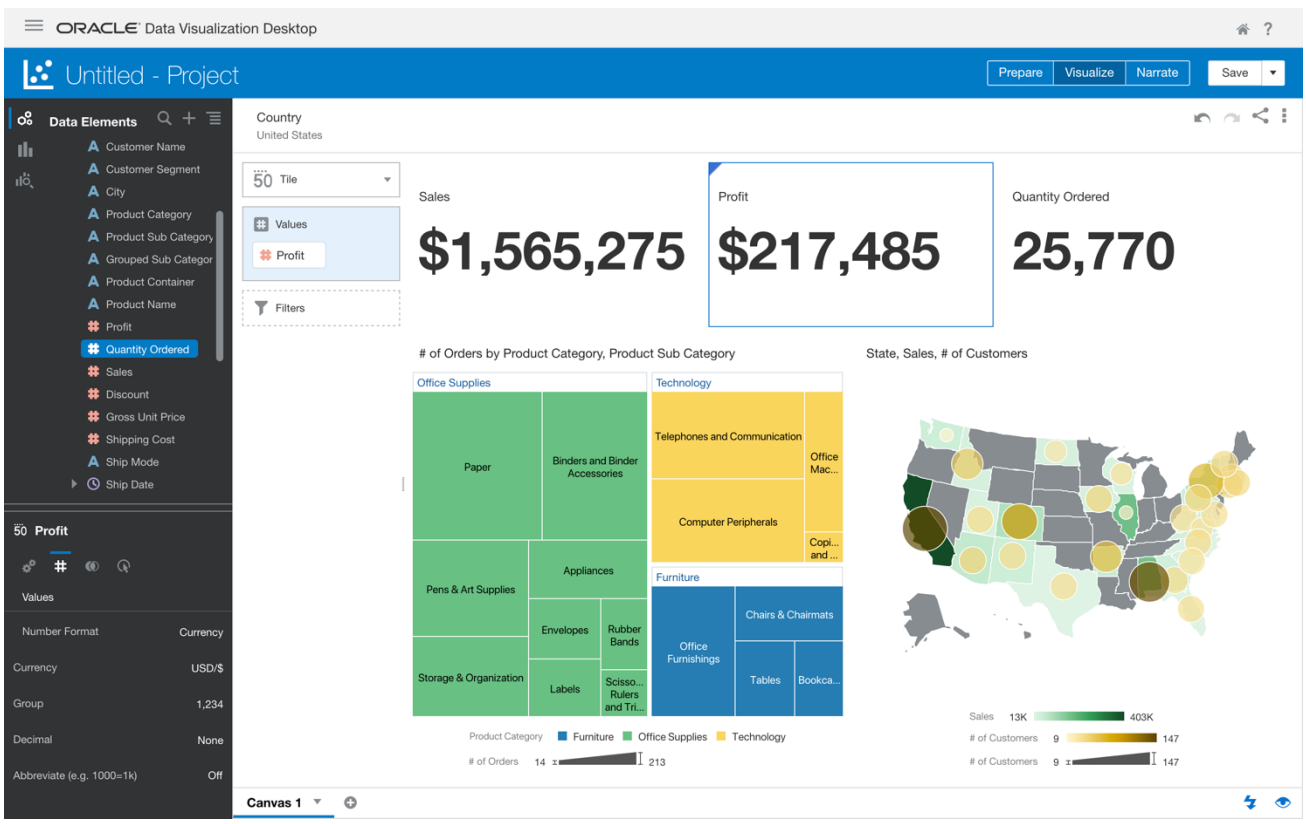


On the values tab, click **Auto** in the Number format line, and then select **Currency**. The property items change accordingly after the Currency item is selected.



Click **0.01** at the right of Decimal, select **None** in the pop-up menu.

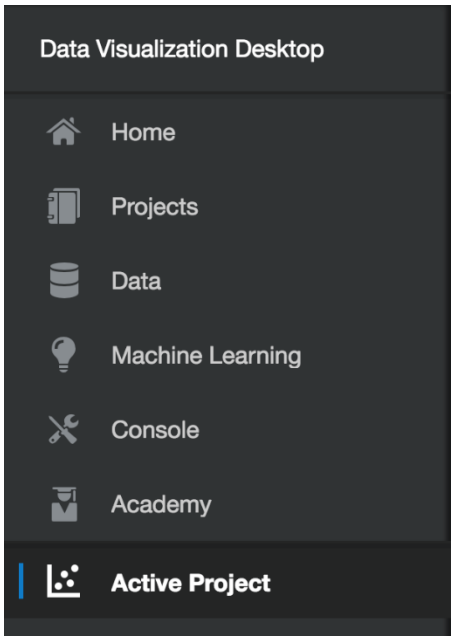
Change the properties of Profit tile in the same way.



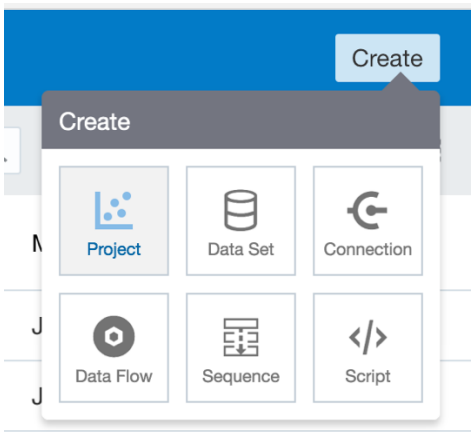
Now you get a nice visualization. You can save and share. You can see the tiles of Sales, Profits, and Quantity Ordered on the top. Down below you can see the tree map by product category and sub product category. You can also see the geographic distribution of sales and # of customers by states US.

Adding Data Sources from Comma Separated Value files

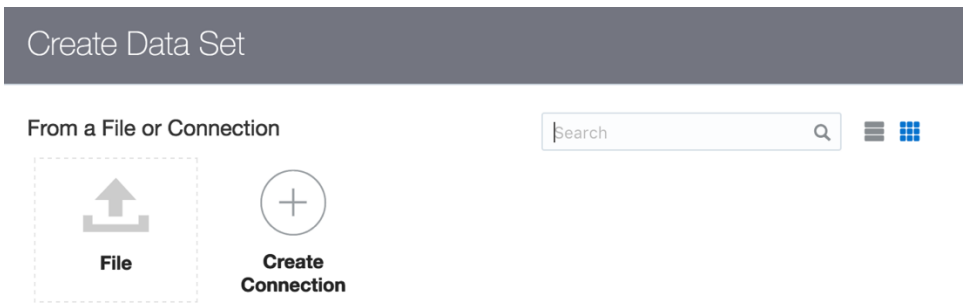
Open the Navigator menu and select Data to create a new Data Set.



Click on button **Create** at the upper right area of the window.

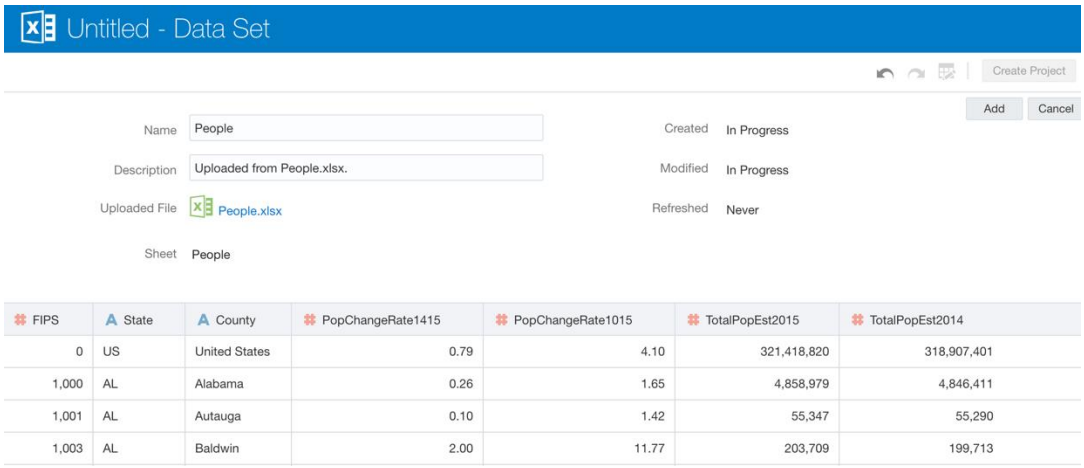


Select **Data Set** as the new component to add.



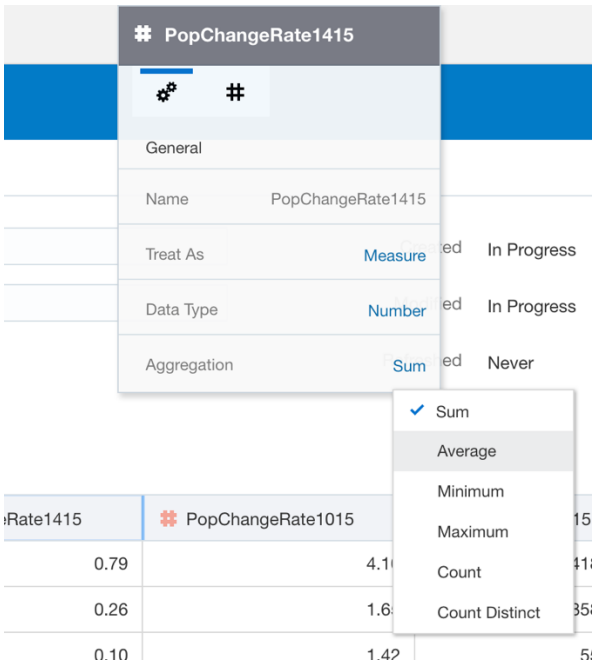
In the Create New Data Source dialog, click **File**.

In the Select File dialog, select the **People.xlsx** file.



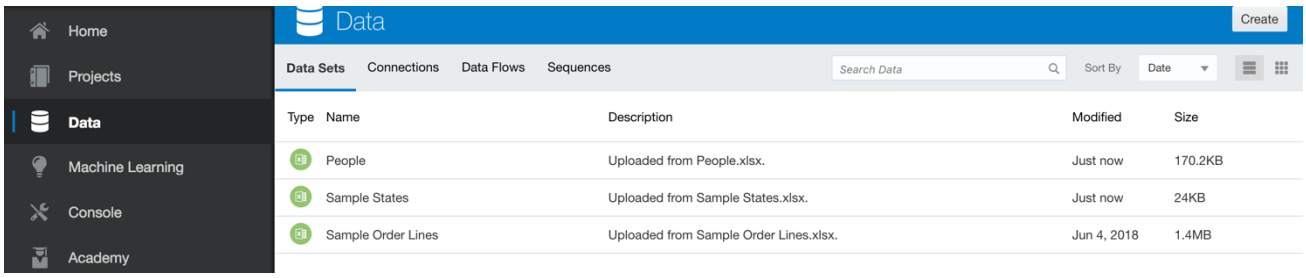
FIPS	State	County	PopChangeRate1415	PopChangeRate1015	TotalPopEst2015	TotalPopEst2014
0	US	United States	0.79	4.10	321,418,820	318,907,401
1,000	AL	Alabama	0.26	1.65	4,858,979	4,846,411
1,001	AL	Autauga	0.10	1.42	55,347	55,290
1,003	AL	Baldwin	2.00	11.77	203,709	199,713

The column names and data from People.csv are listed. Note, only part of the records in the data source is displayed.

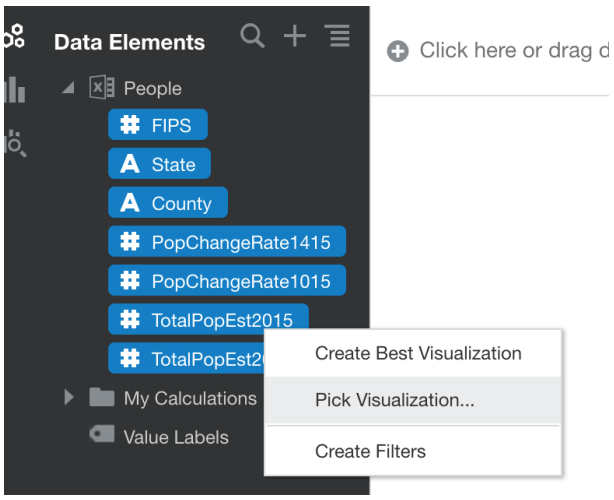


Click the **gear icon** alongside PopChangeRate1415, click **Sum** and select **Average** as the aggregation type instead in the pop-up menu.

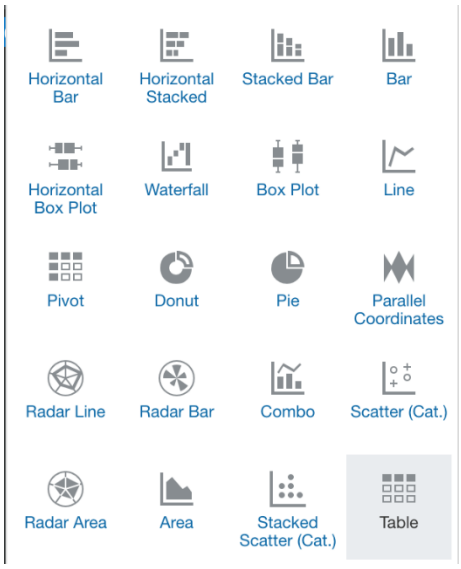
Do the same thing for **PopChangeRate1015** column. Click **Add**.



The People data source is created as shown above. Click **People** to view its elements.



Select all of the data elements in the People data source, right click them, and click **Pick Visualization...** in the pop-up menu.

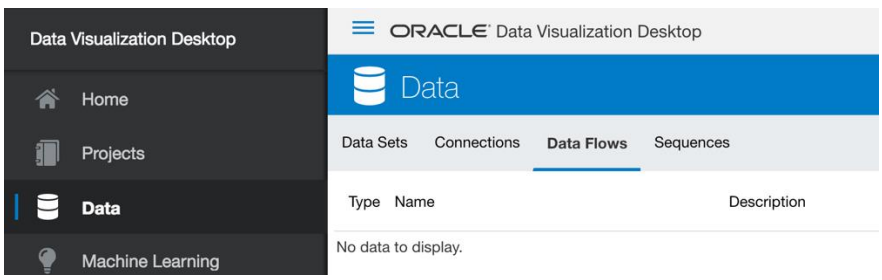


The columns and rows of data in the People data source are listed in table.

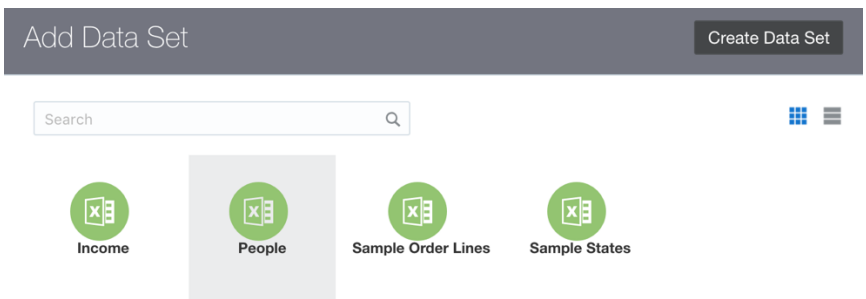
Create the **Income** data source from **Income.csv** file in the same way as we create the People data source above.

Creating a Data Flow

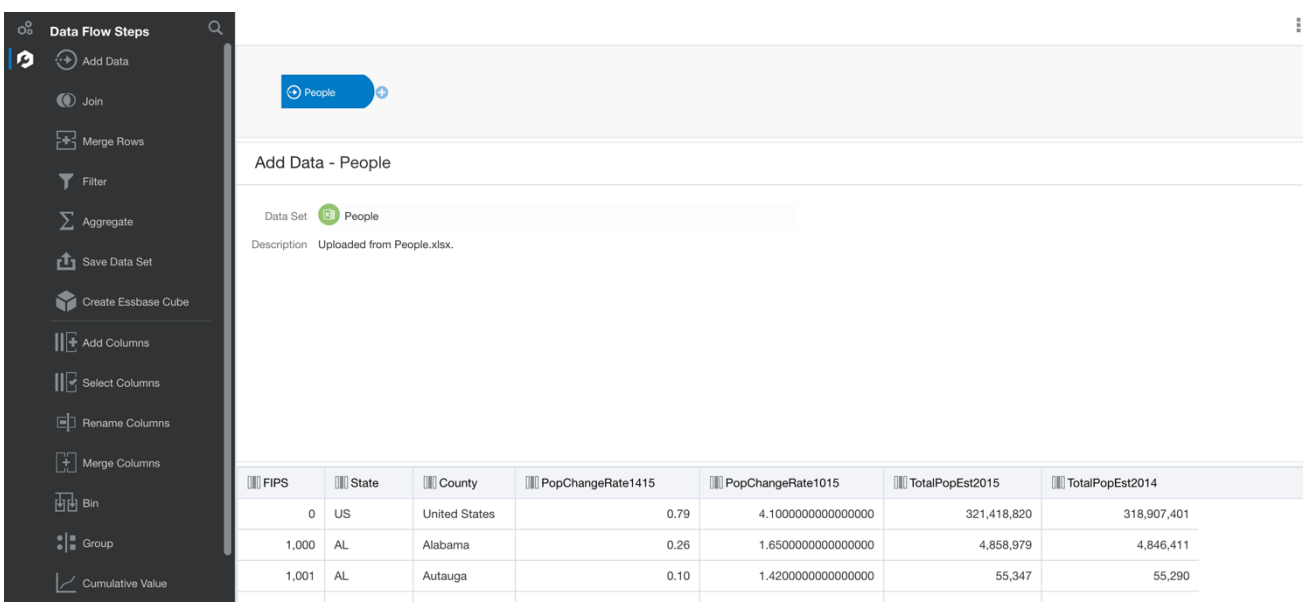
In this section we create a data flow from the People and Income data sources created in the previous section. Data flows are a way to produce a curated data source that you can use to easily and efficiently create meaningful visualizations. You can create a data flow from one or more data sources.



Return to the Home page, and click **Data** on and click **Data Flows** on the menu. Click on the **Create** button on the top right area, and select Data Flow in the pop up menu.



In the Add Data Source dialog, select the **People** data source that we want to base the data flow on. You can select only one data source in this dialog; if needed, you can add additional data sources later. Click **Add Data**.

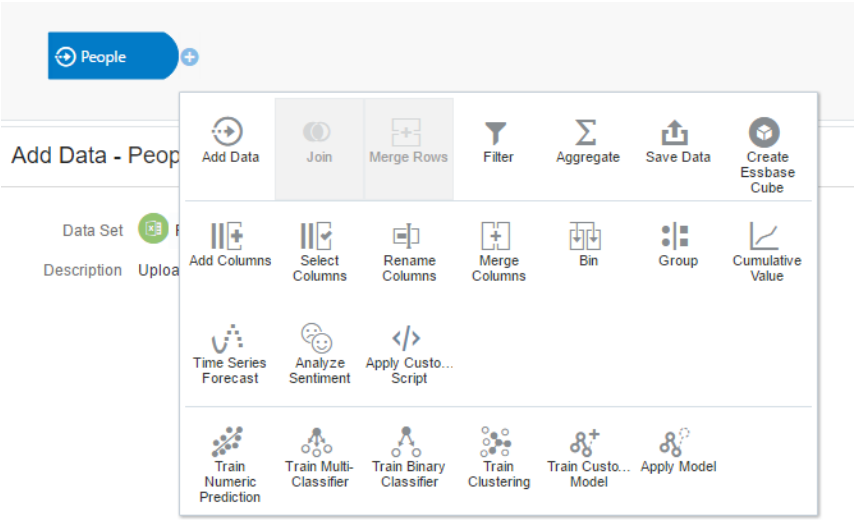


The Data Flow editor is displayed and the columns from the selected data source are displayed in the Data Elements pane. The data source **People** is displayed as the first step in the workflow diagram pane.

Click the save data flow icon in the top left corner and select **Save**. Although you have not completed the data flow, it's always a good habit of saving it in the time.

In the Save Data Flow dialog, input **People and Income DF** as the name of the data flow you are working on. Click **OK.**

In the Data Flow editor, go to workflow diagram pane and right click the **People** icon. Select **Add Data** to add the Income data set.



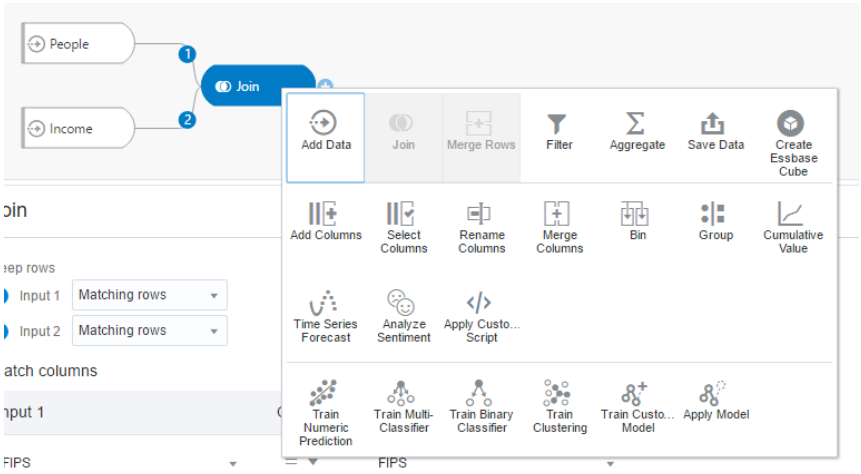
In the Details pane, you can see the two data sources are joined by the match column FIPS.

You can analyze a data source on its own, or you can analyze two or more data sources together, depending on what the source contains. If you use multiple sources together, then at least one match column must exist in each source. The requirements for matching are:

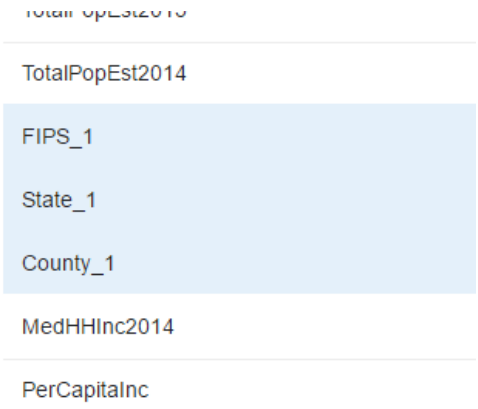
- The sources contain common values (for example, Customer ID or Product ID)
- The match must be of the same data type (for example, number with number, data with date, or text with text).

The joined data appears on the preview pane. Notice that the columns of FIPS, State, and County exist in both data sources, so the duplicate columns are renamed as FIPS_1, State_1, County_1 automatically.

In the workflow diagram pane, right click the Join icon. Select Add Step.



From the Add Step window, click **Select Columns**.



In the Step Details pane, select the duplicate columns FIPS_1, State_1, and County_1. Click **Remove selected**.

The duplicate columns are removed to the left. The **Preview** pane, shows the updated dataset.

Now there is no duplicate column in the join data set.

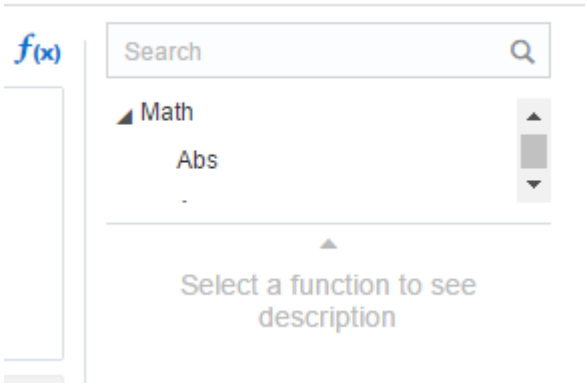
However, if you look into the top five rows, you can find that:

- In the first row, the United States should be a country rather than a county. The FIPS is 0.
- In the second row, Alabama should be a state rather than a county in the United States. The FIPS is 1000.
- Autauga, Baldwin, and Barbour and so on are counties in Alabama, the FIPS' are 1001, 1003, and 1005 and so on respectively.

The country, state, and county level rows are mixed together in this data source. We need to remove the rows of the United States and the states from the table, and leave only the county rows in the table for further analysis and visualization.

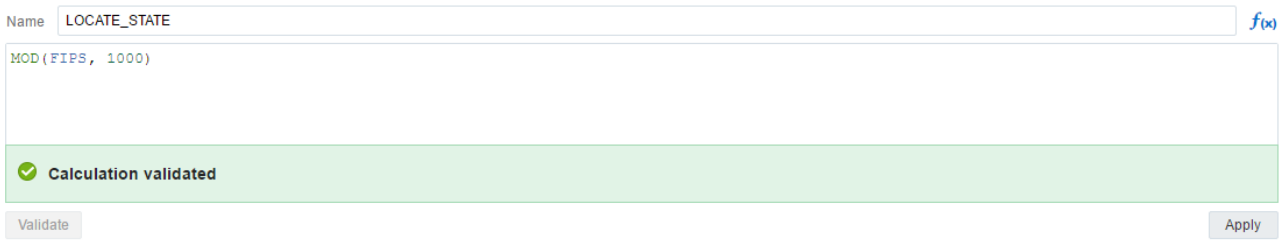
In the workflow diagram pane, right click the **Select Column** icon. Select **Add step**.

From the Add Step window, click **Add Columns**.

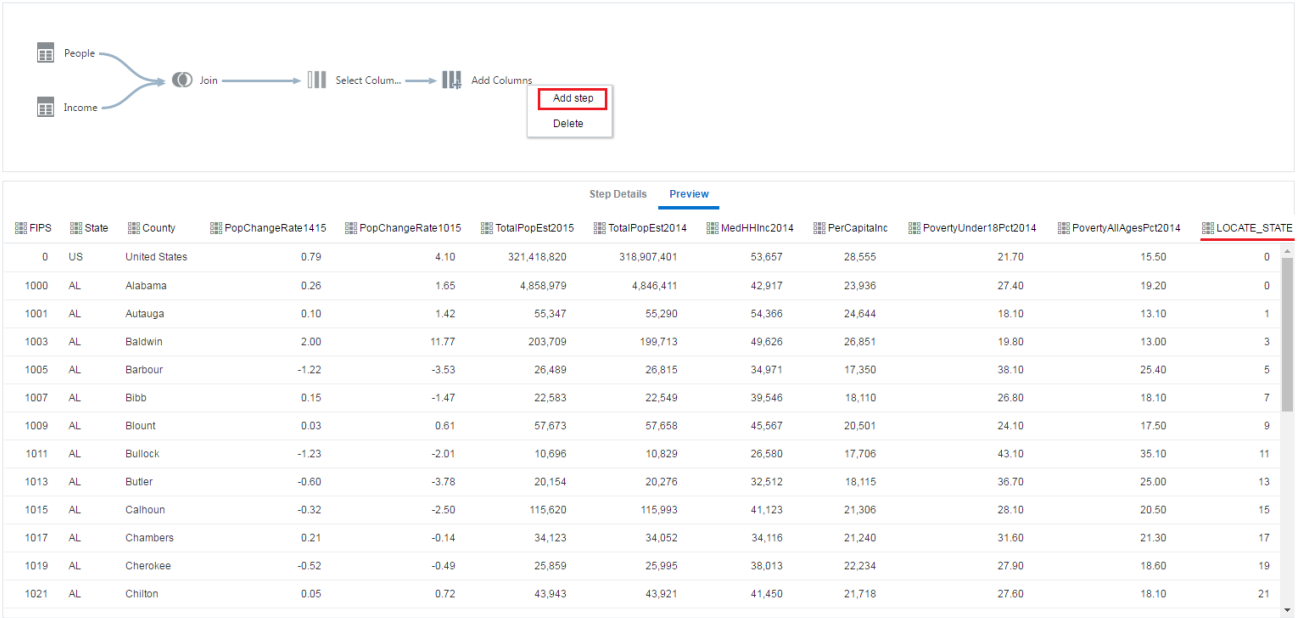


In the Step Details pane, click **Math** to expand the list of math functions.

Scroll down the list and double click the **Mod** function.



The MOD function is selected in the edit box. Input **FIPS** as the first parameter and **1000** as the second parameter for the MOD function. Input **LOCATE_STATE** as the new column name. Click **Validate** and **Apply**. Then check the **Preview** pane.

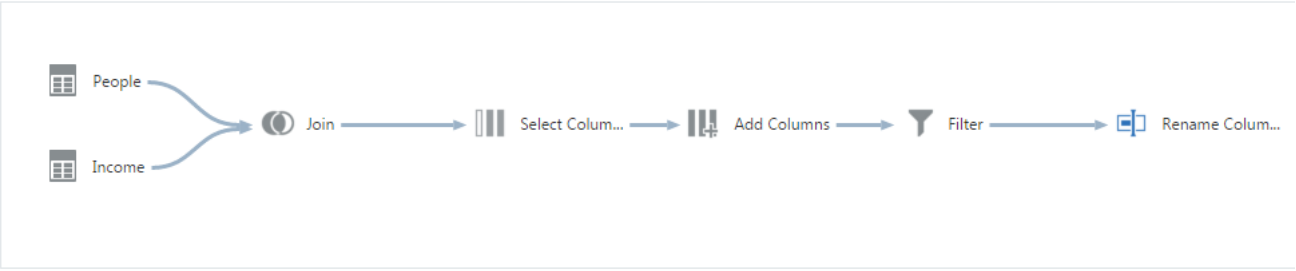


FIPS	State	County	PopChangeRate1415	PopChangeRate1015	TotalPopEst2015	TotalPopEst2014	MedHHInc2014	PerCapitaInc	PovertyUnder18Pct2014	PovertyAllAgesPct2014	LOCATE_STATE
0	US	United States	0.79	4.10	321,418,820	318,907,401	53,657	28,555	21.70	15.50	0
1000	AL	Alabama	0.26	1.65	4,858,979	4,846,411	42,917	23,936	27.40	19.20	0
1001	AL	Autauga	0.10	1.42	55,347	55,290	54,366	24,644	18.10	13.10	1
1003	AL	Baldwin	2.00	11.77	203,709	199,713	49,626	26,851	19.80	13.00	3
1005	AL	Barbour	-1.22	-3.53	26,489	26,815	34,971	17,350	38.10	25.40	5
1007	AL	Bibb	0.15	-1.47	22,583	22,549	39,546	18,110	26.80	18.10	7
1009	AL	Blount	0.03	0.61	57,673	57,658	45,567	20,501	24.10	17.50	9
1011	AL	Bullock	-1.23	-2.01	10,696	10,829	26,580	17,706	43.10	35.10	11
1013	AL	Butler	-0.60	-3.78	20,154	20,276	32,512	18,115	36.70	25.00	13
1015	AL	Calhoun	-0.32	-2.50	115,620	115,993	41,123	21,306	28.10	20.50	15
1017	AL	Chambers	0.21	-0.14	34,123	34,052	34,116	21,240	31.60	21.30	17
1019	AL	Cherokee	-0.52	-0.49	25,859	25,995	38,013	22,234	27.90	18.60	19
1021	AL	Chilton	0.05	0.72	43,943	43,921	41,450	21,718	27.60	18.10	21

You can see that only the rows with non-zero values of LOCATE_STATE is kept in the table for futher analysis.

In the workflow diagram pane, right click the Filter icon. Select Add step.

From the Add Step window, click Rename Columns.



Step Details | Preview

Rename Columns

Source	Name
FIPS	FIPS
State	State
County	County
PopChangeRate1415	Pop Change Rate 1415
PopChangeRate1015	Pop Change Rate 1015
TotalPopEst2015	Total Pop Est 2015
TotalPopEst2014	Total Pop Est 2014
MedHHInc2014	Med HH Inc 2014
PerCapitaInc	Per Capita Inc 2014
PovertyUnder18Pct2014	% Poverty Under 18 2014
PovertyAllAgesPct2014	% Poverty All Ages 2014
LOCATE_STATE	LOCATE_STATE

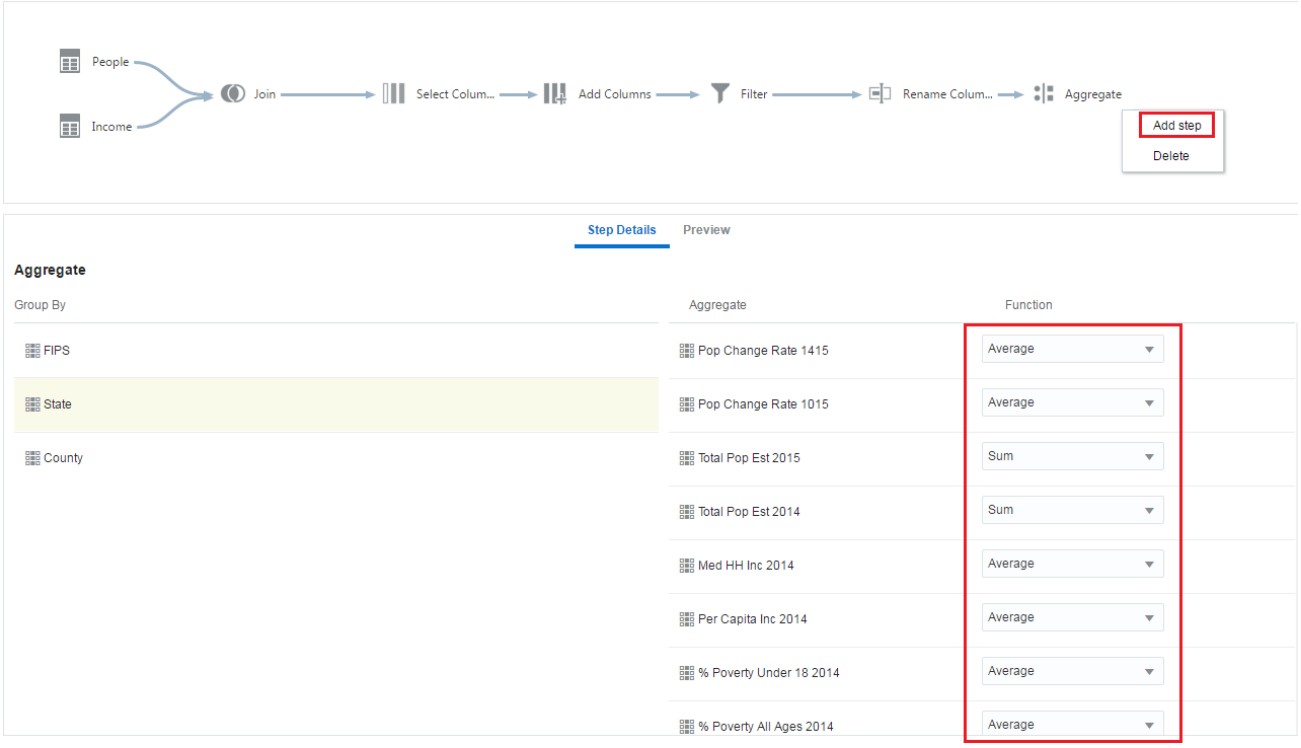
In the Step details pane, rename the following columns:

- PopChangeRate1415 -> Pop Change Rate 1415
- PopChangeRate1015 -> Pop Change Rate 1015
- TotalPopEst2015 -> Total Pop Est 2015
- TotalPopEst2014 -> Total Pop Est 2014
- MedHHInc2014 -> Med HH Inc 2014
- PerCapitaInc -> Per Capita Inc 2014
- PovertyUnder18Pct2014 -> % Poverty Under 18 2014
- PovertyAllAgesPct2014 -> % Poverty All Ages 2014

You can see the renamed columns in the table.

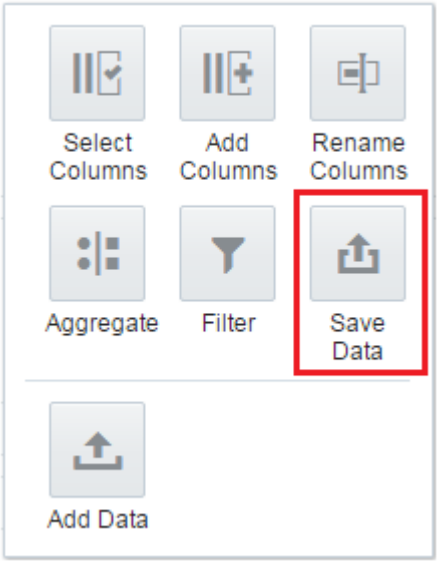
In the workflow diagram pane, right click the **Rename Columns** icon. Select **Add step**.

From the Add Step window, click **Aggregate**.



The aggregate function for each data element should be shown as that in the screenshot. If not, change the functions to Average or Sum accordingly. The function is **Sum** for **Total Pop Est 2015** and **Total Pop Est 2014**, and Average for the rest of data elements.

In the workflow diagram pane, right click the **Aggregate** icon. Select **Add Step**.

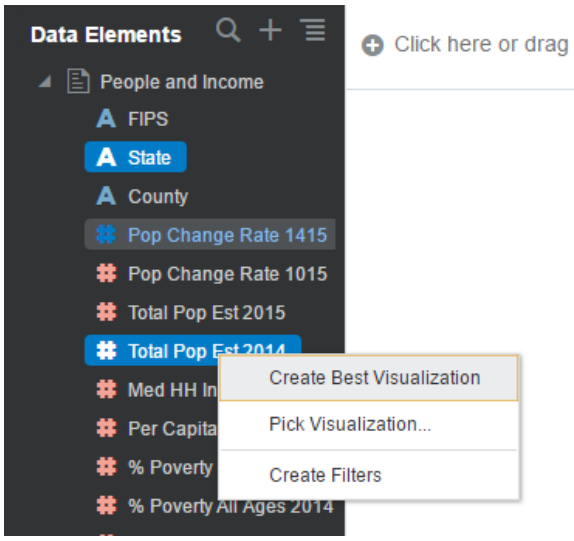


From the Add Step window, click **Save Data**.

In the Step Details pane, input **People and Income** as the output data source name.

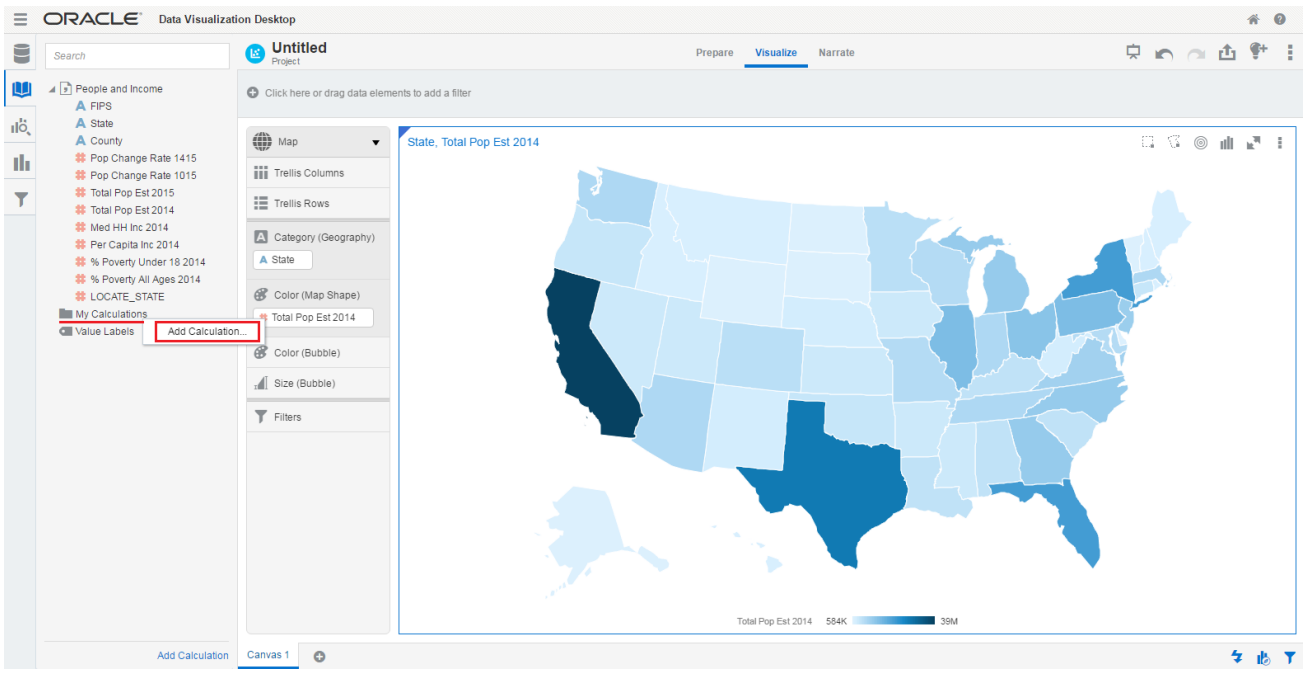
Click the **Execute Data Flow** button at the top right corner to execute the data flow.

Return to the Home page, and click **Data Sources** on the top and click the newly created **People and Income** data source.



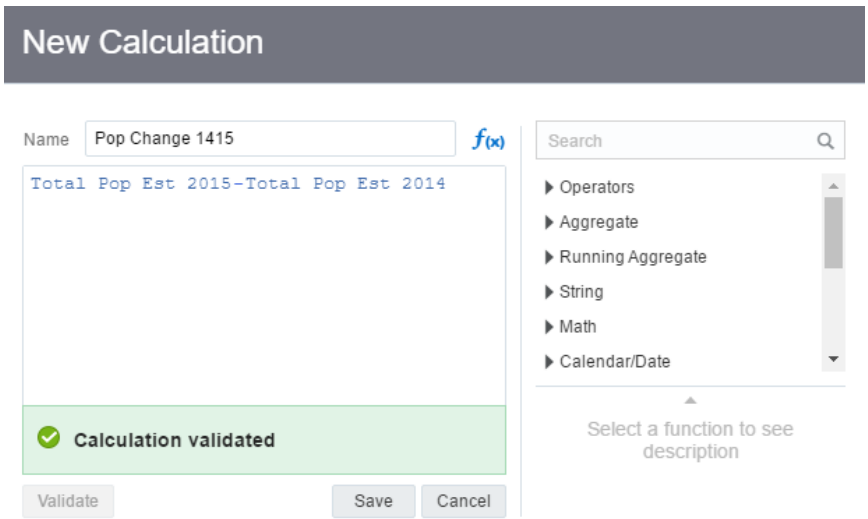
In the data element list, select both **State** and **Total Pop Est 2014**, right click them, and select **Pick Visualization**… on the pop-up menu.

In the View Select dialog, select **Map** visualization type.



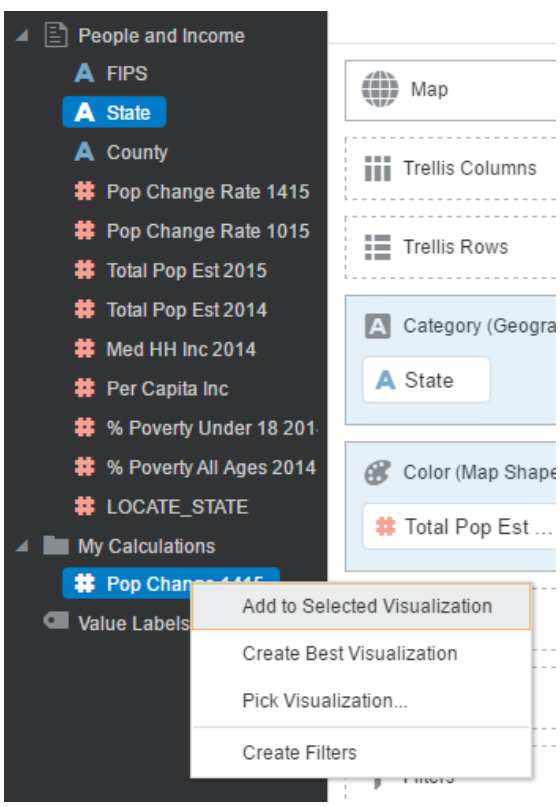
You can see that states are colored in blue. The darker the color is, the higher value of Total Prop Est 2014 is.

Right Click My Calculations on the left, select Add Calculation.

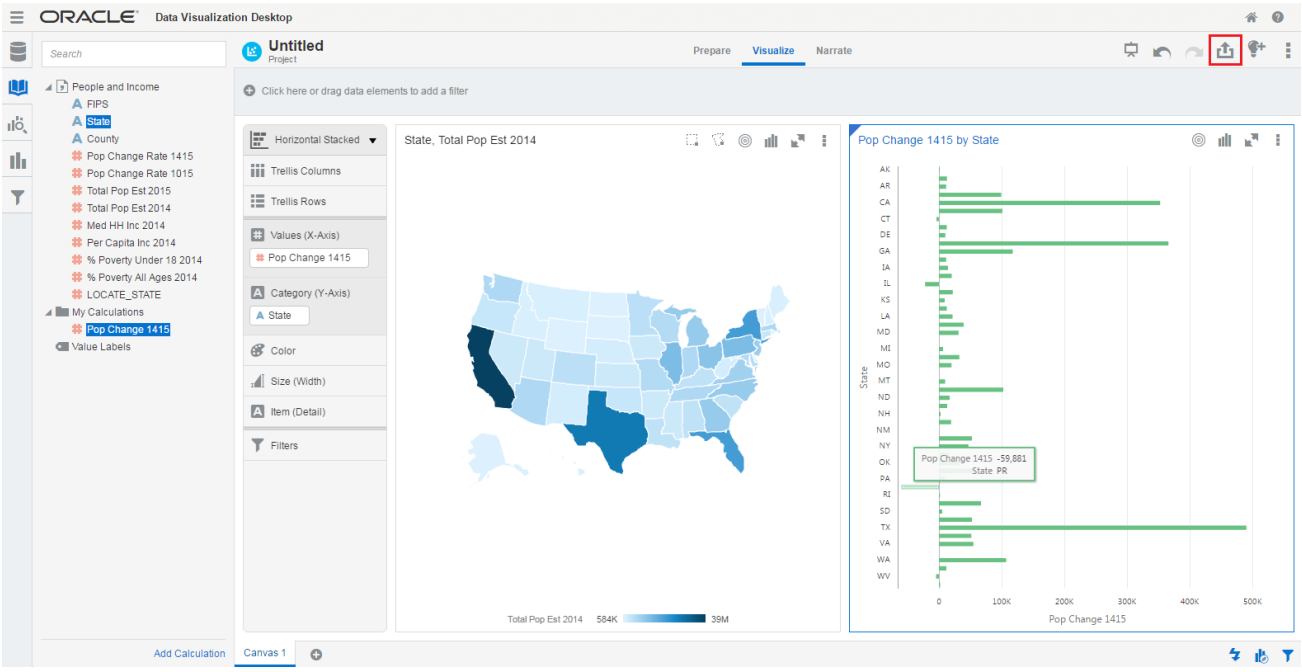


Drag and drop **Total Pop Est 2015** and **Total Pop Est 2014** from the data elements list to the edit box in the New Calculation dialog, add a minus (-) symbol between them. Name it as **Pop Change 1415**.

Click **Validate**. Click **Save**.



The newly created Pop Change 1415 calculation is listed under My Calculations folder on the left. Select both **State** and **Pop Change 1415**, right click them, and select **Create Best Visualization** in the pop-up menu.



Oracle Visualization allows even nontechnical users to select a few data elements and generate what the software thinks is the best visualization for the relationships between those items. It helps business users uncover patterns and understand their data better. In this example, it creates a bar chart for the estimated population change between year 2105 and 2014. From the bar chart, we can easily see the states with big positive or negative changes.

Click the **Save icon** in the top right corner, save the project as **People and Income Viz**.