BI Concepts, Tools and Applications

Exercise #3: Enrollment in Distance Education Courses by State

In this exercise, You will learn more on the advanced features of the microstrategy tool and also learn on creating a table of contents.

INTRODUCTION

Distance learning is a major trend in education that gives people the flexibility to learn and pick up new skills at their own speed. In this Exercise, you will examine data on the value of online education in the United States. For the sake of this scenario, let's assume that your company made a significant investment in a distance learning program last year. Unfortunately, you did not have enough data to accurately determine the value of this new type of education. To remedy this, you decided to investigate the value of pursuing online education.
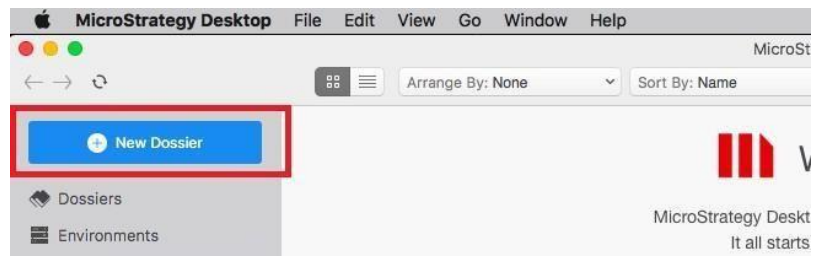
GETTING STARTED

Start by launching **MicroStrategy Desktop**. If you have not downloaded it yet, you can find the software at **www.microstrategy.com**

1. Double-click the **MicroStrategy Desktop** icon.

2. Click **New Dossier** on the Desktop home window.

microstrategy.com

The dossier authoring window opens. Before we get started building a dossier, let's briefly familiarize ourselves with the interface.
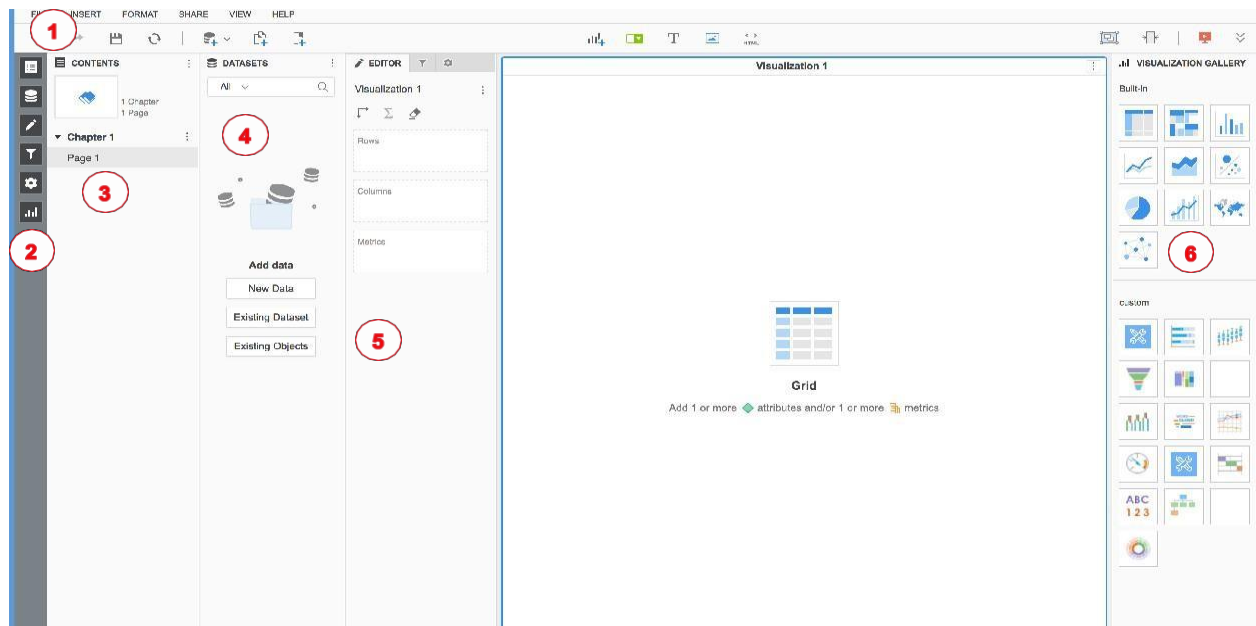
1. **Toolbar**: Provides controls to redo or undo an action, refresh the dossier, add dossier elements (including datasets, pages, chapters, visualizations, in-canvas filters, text, images, and links), change dossier formatting, and share your dossier. Click the **Responsive Preview** and **Responsive View Editor** icons to advantage of responsive design functionality to preview your dossier in mobile mode.

2. **Panel Control:** Click the icons on the left side of the interface to show or hide different panels, which are outlined below.

3. **Contents Panel:** This panel displays the dossier table of contents showing the chapters and pages of your dossier. Edit, add, and remove pages and chapters. In the table of contents, you can also add a thumbnail image for the dossier on this panel to help identify the dossier in Desktop. For example, you might add an image of a map for a dossier analyzing geographical performance. Click the editor icon in the Panel Control section to show or hide this panel.

4. **Dataset Panel:** The Datasets panel displays objects from each dataset in the dossier, organized by the dataset that they belong to. You can delete or add new datasets through this panel by clicking the Menu icon. To show or hide this panel, click the dataset icon in the Panel Control section.

5. The **Format**, **Filter**, and **Editor** panels are stacked next to the Dataset panel. Flip between these three panels to customize your visualizations and add filters to the dossier chapters.

   a. **Editor Panel:** To create your visualizations by adding data, use the Editor panel. Drag and drop attributes and metrics into the Rows, Columns, and Metrics boxes to see them appear in the visualization. Options in this panel depend on the selected visualization. Click the editor ✎ icon in the Panel Control section to show or hide this panel.

   b. **Filter Panel:** To apply filters to limit and specify the data visible on the dossier chapter, use the filter panel. You can use metrics, attributes, or visualizations to filter the data. Click the filter ▼ icon in the Panel Control section to show or hide this panel.

   c. **Format Panel:** To customize your visualizations, use this panel to change your visualization's font and the visualization container to fit your dossier needs. You can also customize visualizations by adding a legend, reference line, changing the shape of a graph, and so on. Click the format ⚙ icon in the Panel Control section to show or hide this panel.

6. **Visualization Gallery:** Use the Visualization Gallery to quickly select the visualization you want to use to display your data, or import a custom visualization with just a few clicks. Hover over an icon to see the visualization name and data requirements Click the visualization 📊 icon in the Panel Control section to show or hide this panel.

Download all the required files that are in the module, below the exercise document, for completing the exercise.

Overview

In this exercise, you will create a dossier and use data wrangling to prepare your data. To get started, import the dataset **Web Tables—Enrollment in Distance Education Courses**, **by State_ Fall 2012.xlxs** into MicroStrategy Desktop.

You will then use MicroStrategy's data wrangling functionality to prepare the data in the **Web Tables—Enrollment in Distance Education Courses by State_ Fall 2012.xlsx**. There are some discrepancies in the way the data in each column is presented. To do this, you will need to parse and edit the columns.
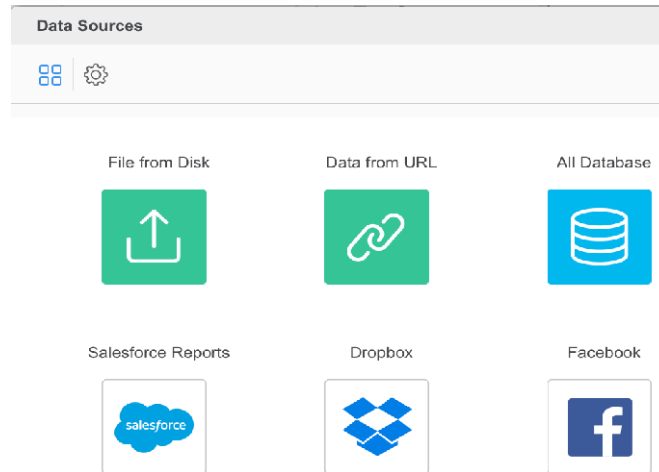
1. Click the **Save** icon on the MicroStrategy Desktop toolbar and save your dossier as **Online Education Data Wrangle** in your My Documents folder**.**

2. From the tool bar, click the **Add Data** icon in the tool bar, and select **New Data** to open the Connect to Your Data interface.
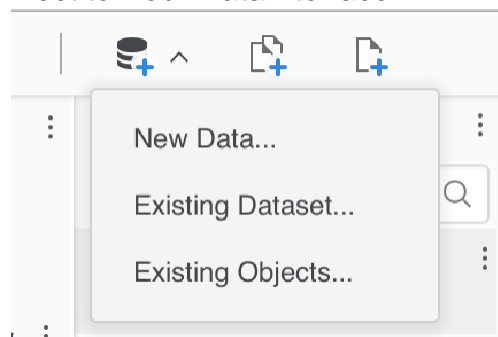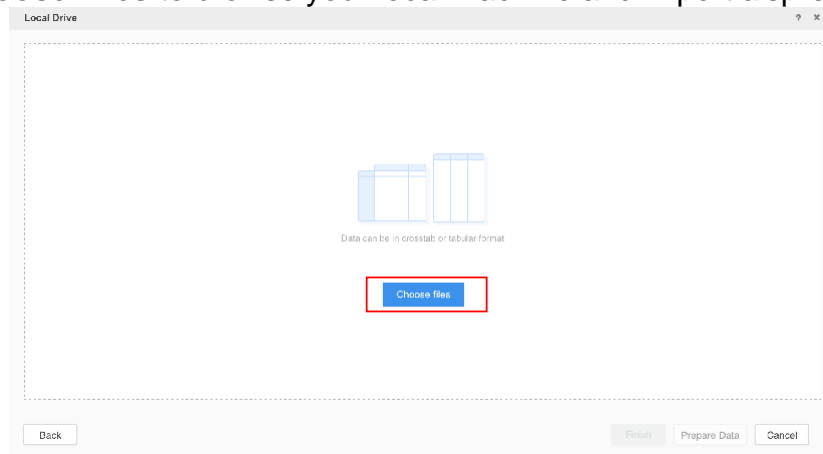

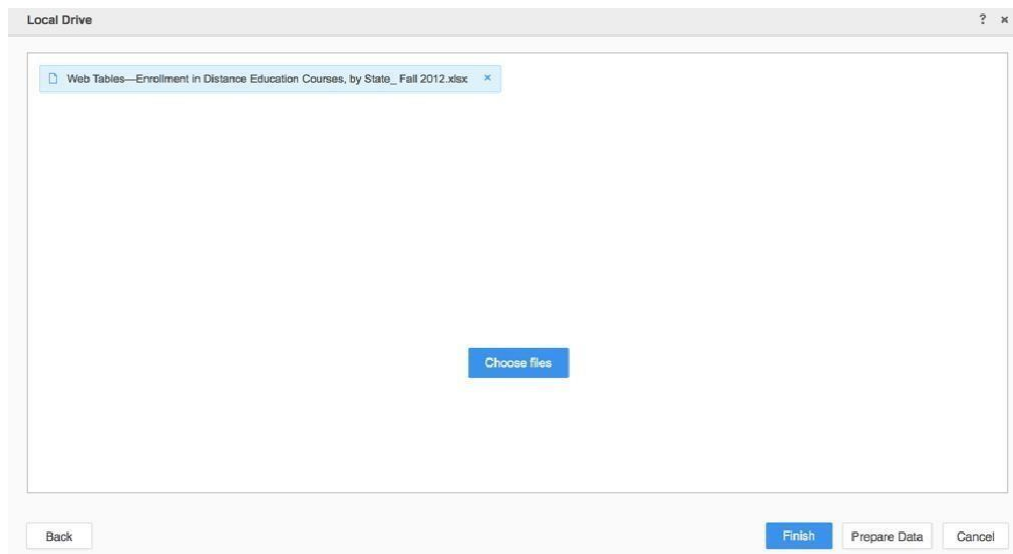3. For this Exercise, you will import data from an Excel file. Click **File from Disk**.

4. Click **Choose Files** to browse your local machine and import a spreadsheet.

5. Browse to the **Web Tables—Enrollment in Distance Education Courses by State_ Fall 2012** file and then click **Open**.

This file should be downloaded as mentioned at the start of the exercise.

TIP: *MicroStrategy allows users to import multiple sheets at the same time. This allows for easy uploading of entire Excel worksheets, without requiring you to delete sheets that aren't used for your visualization. Each sheet will automatically be treated as a different table within the dataset, allowing you to*

create joins between tables

6. Click **Prepare Data**.
7. Ensure that you select only the **Table 2** workbook in **Web Tables—Enrollment in Distance Education Courses by State_ Fall 2012.xlsx**.
8. Click **Select**.
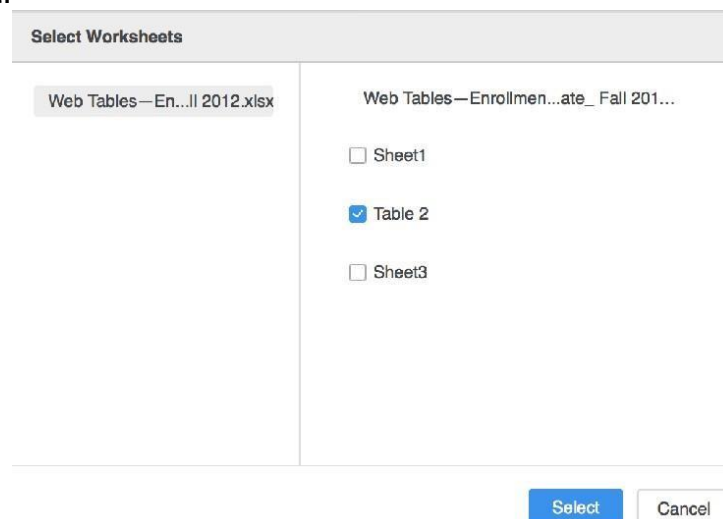
microstrategy.com 8

After the data loads, the Preview window opens, which contains information about the attributes and metrics included in the different datasets.

9. To view the columns in the dataset, click the **Web Tables—Enrollment in Distance Education Courses by State_ Fall 2012.xlsx** table.

10. Click the **down arrow** on the Web Tables—Enrollment in Distance Education Courses by State_ Fall 2012.xlsx table, and select **Parse**.

11. The Parse Your Data window opens. In the **Skip** box, type **4**, then click **Apply**.

Notice that all the rows are automatically moved up.

12. Click **Apply** to apply the changes and return to the Preview window. The metrics are now correctly named instead of being labeled as Column 1, Column 2, etc.

Wrangle your Data

13. Click the **Wrangle** icon. The Data Wrangling window opens.

TIP: *MicroStrategy automatically maps the imported rows and columns as attributes and metrics. Attributes provide a business model with context for reporting and analysis. Metrics represent business measures and key performance indicators (KPIs). If changes to any content or data types of the attributes or metrics are needed, they can be performed manually in the Preview window. Note that each attribute and metric is clearly marked. Attributes are represented by a blue icon* ◆*, while metrics are represented by an orange icon* ▤ *.*

Often, imported datasets contain errors, missing records, or incorrect formatting that can make analysis difficult. Rather than searching through each record in Excel and making individual modifications, we can make changes directly within MicroStrategy by using the **Data Wrangling** tool. The Preview window shows data for the first 50 rows in the selected dataset. It is designed to give you a general idea of the type of data contained within a specific dataset.

14. Select the **Regions** column.

15. You only want to see data by states, so you need to delete the information for the United States. In the **Select** Function dropdown menu, select **Delete Row**. In the **Where Cell Contains** text box, type **"United States"**.

16. Click **Apply**. Notice that this function removes the first row, deleting data for United States, while the State data remains.

17. You notice that the District of Columbia is separated into two rows. To combine them, select the **Regions** column again. From the **Select Function** dropdown menu, click **Text Selector**.

Delete Row
Remove Duplicate Cells
Remove Duplicate Rows
Delete Row included in selectors
Delete Row not included in selectors
Extract
Extract By Fixed Length
Extract Before Separator
Extract After Separator
Selector
Text Selector
Numeric Selector
Timeline Selector
Text Search
Find And Replace

Notice the Text Selector window appears below the data.

Regions (52)
Colorado (1)
Columbia (1)
Connecticut (1)
Delaware (1)
District of (1)
Florida (1)
Georgia (1)
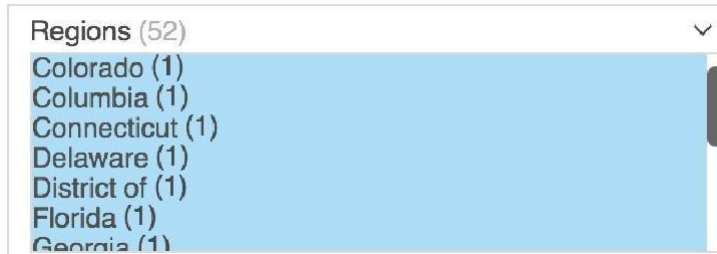
18. Click **District of** in the text selector window. A window will appear above the data.

19. Click **Delete Row Included in Selector**.
20. After you delete the **District of** row, in the Region window, hover over **Columbia**.
21. Click **Edit**, and replace **Columbia** with **District of Columbia.** Click **Apply to All.**

1 - 50 of 51 rows

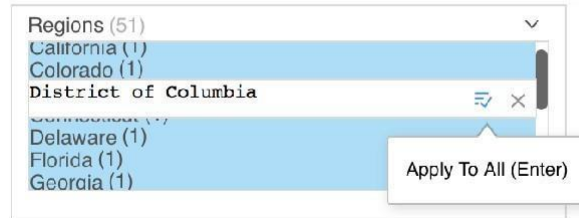As you saw, the row for District of did not contain any data, making it unnecessary for our analysis. On the other hand, the row Columbia contained all the data we needed, so we renamed it District of Columbia.

22. Click **Apply** to apply your changes and return to the preview window.

< Page [1] / 2 > »

[Apply] [Cancel]

23. Click **Finish**.

	All Students FT Enroll...		All Students Enrolled P...		All Students PT Enroll...		All Students Not E
	16.0	57,893		18.6		203,764	
	16.4	5,817		17.2		22,489	
	18.2	122,339		16.4		265,028	
	9.2	34,586		19.1		129,982	
	8.3	284,307		10.6		2,237,580	

[Finish] [Cancel]

TIP: *Because the data wrangling tool stores every action in a script, you can revert any of your changes. To do this, look at the History Script window in the top right of the Data Wrangling interface. You can select individual changes to revert using the undo/redo buttons, revert the entire dataset, or simply change the last action you performed.*

Create a table of contents

Now that you have imported your data, you are ready to put together a compelling story for your meeting. Dossiers allow you to build a modern and interactive analytical book of your business. Each dossier consists of individual pages arranged in easy-to-navigate chapters. On each page, you can tell compelling stories about your data with powerful visualizations such as graphs and maps. To provide a guided experience for end users, dossiers include a Table of Contents that help you logically organize content into chapters and pages.

Create a Table of Contents with the following structure:

- The first chapter should be called **State Education**. It should have a single page labelled **Online Education per State**.
- The second chapter should be called **Institution**. It should have one page called **Enrollment per Institution**.

    When finished with the Table of Contents, it will look like the following image:
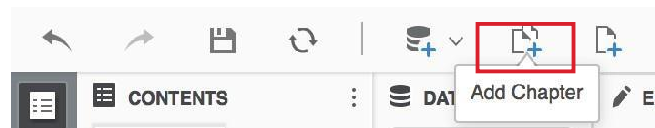
Detailed Instructions

24. Right-click **Chapter 1** in the Table of Contents and rename it **State Education**.
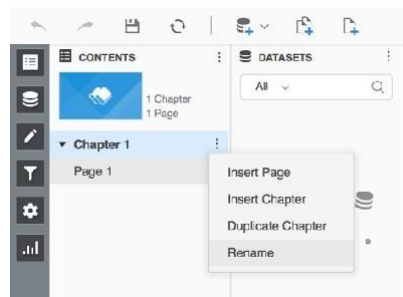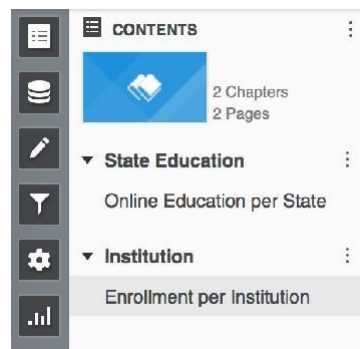

25. Right-click **Page 1** and rename it **Online Education per State**.

26. Click the **Insert chapter** icon on the toolbar to insert a new chapter into your dossier – notice that adding a new chapter also adds a new page in that chapter.

27. Name the new chapter **Institution** and the new page to **Enrollment per Institution**.

28. **Save** your dossier.

   • Select the **Online Education per State** page in the Table of Contents.

   In the Visualization Gallery, click the **Map** icon. Make sure you select the Map icon – when you hover over the correct icon Map will display, not Geospatial Services.

29. In the **Web Tables—Enrollment in Distance Education Courses, by State_ Fall 2012.xlsx** dataset, right-click on the **Regions** attribute, navigate to **Define Geography**, and select **State** from the dropdown menu.
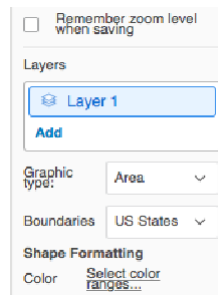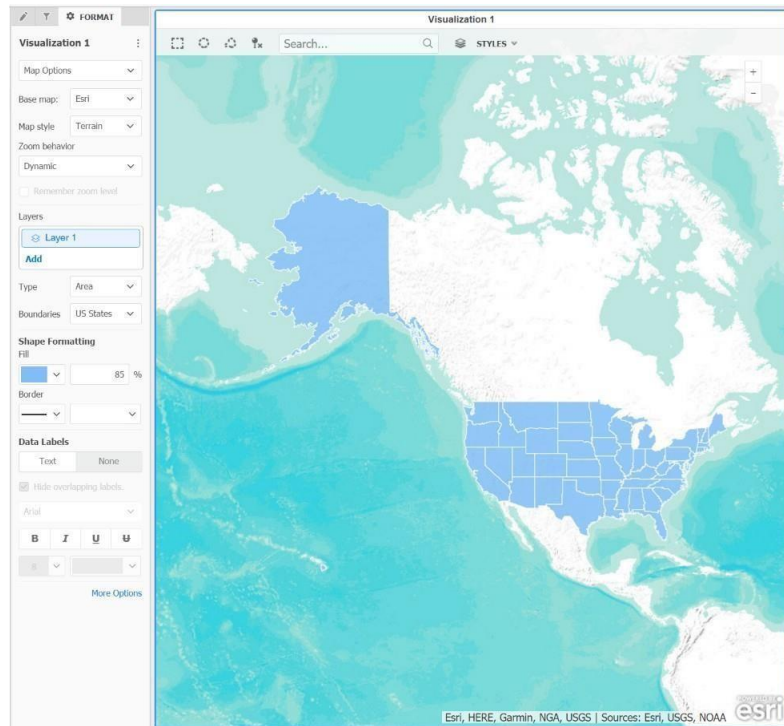
microstrategy.com

30. Click **OK**.
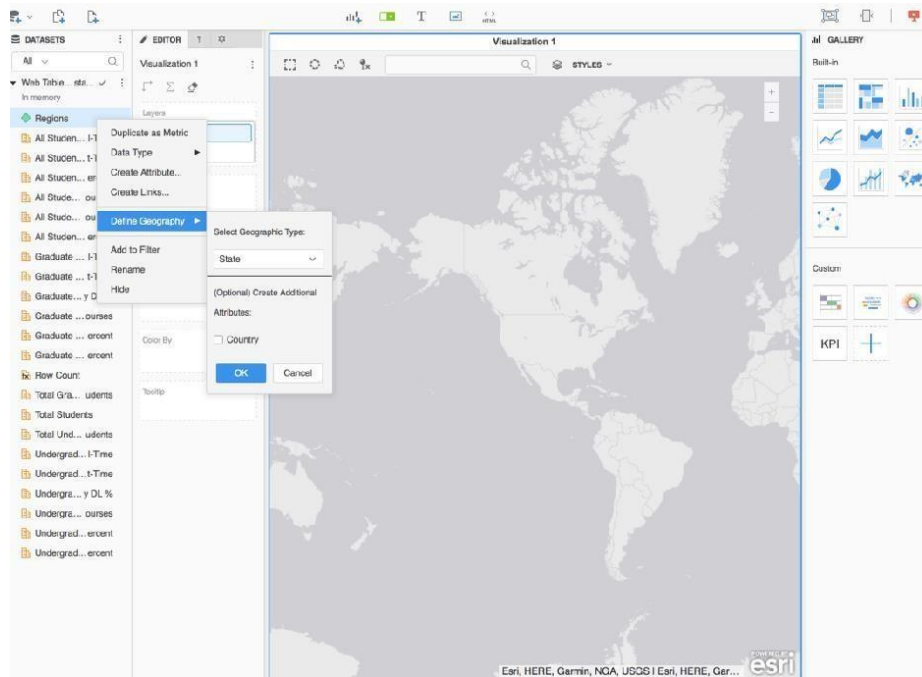
31. Drag the **Regions** attribute from the dataset to the **Geo Attribute** drop zone located in the **Editor Panel.**

32. Open the **Format Panel**.

33. In the **Graphic type** menu, select **Area.**

34. In the **Map style** dropdown menu, select **Terrain.** Your map should match the image below.

microstrategy.com   16

35. Navigate back to the **Editor Panel**.

36. Drag the **Undergrad Students FT Enrollment Percent** metric from the dataset to the **Color By** drop zone to color the states by enrollment percentage. The darker the state, the higher the enrollment.

37. Since the data type does not present the numbers as percent, we will need to reformat these values. Right-click the metric in the **Color By** drop zone and select **Calculation.**

38. Divide the selected metric by 100, then click **OK.**
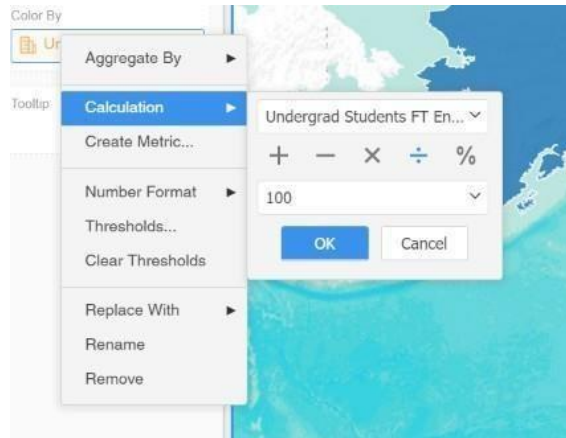
39. Right-click the calculated metric and rename it **Undergrad Students FT**

Enrollment Percent.

You can highlight metric data in a visualization by displaying the data in a different color or replacing the data with an image. For example, a Grid visualization displays revenue data for different geographic regions. You can use thresholds to display revenue values less than $500,000 in red. Thresholds can make analyzing large amounts of data easier because different data colors are quickly identified.

40. You want to change the threshold colors for the states. Right-click the **Undergrad Students FT Enrollment Percent** metric and select **Threshold**.

41. In the Threshold window, select **Red-Orange-Green** from the **Color** dropdown. Click **OK**. Now, if a state is green it has a higher enrollment percentage, while red states have a lower enrollment percentage.
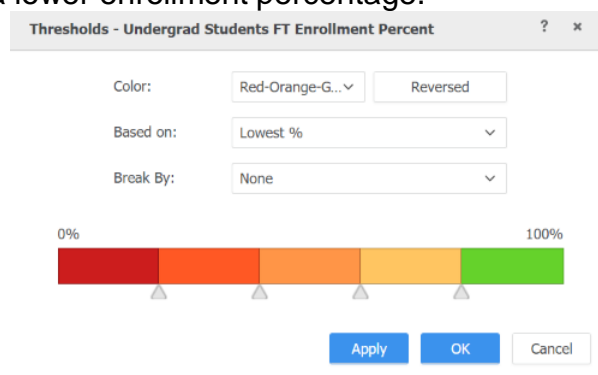

**42.** Add a new visualization by clicking the **Insert Visualization** icon at the top of the window.

43. Click the **title bar** of the newly created visualization and drag it underneath the map visualization – when dragging the visualization, you will notice a blue horizontal line -- this line indicates the location of the visualization.
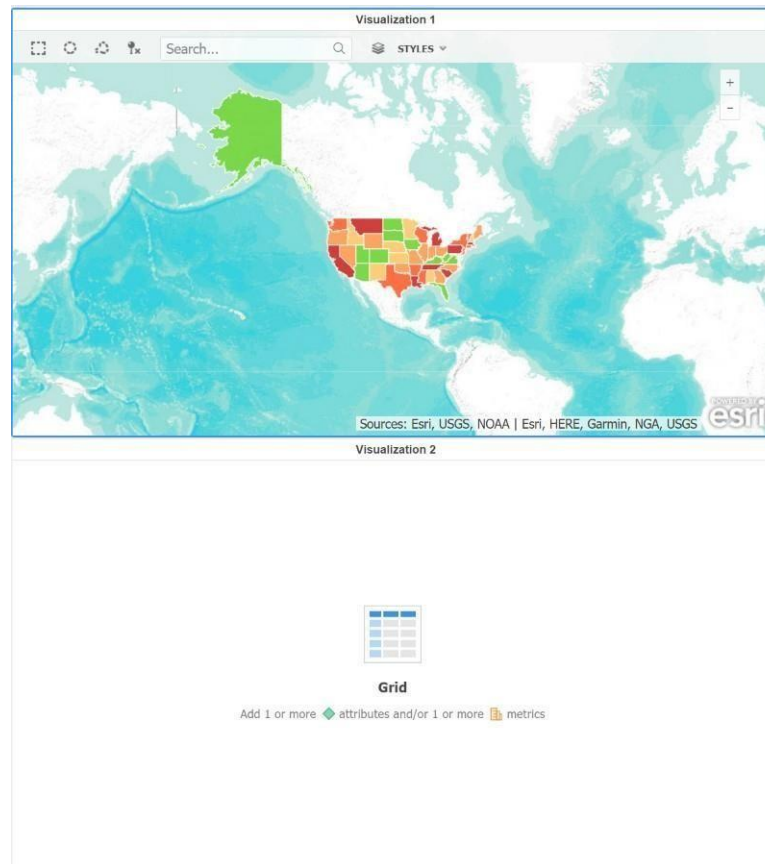
44. From the Visualization Gallery, select the **Bar Chart** icon.

45. Drag and drop **Undergrad Enrolled Full-Time** and **Graduate Enrolled Full-Time** metrics from the dataset into the **Vertical** drop zone.

46. Drag the **Regions** attribute to the **Horizontal** drop zone.

47. Move the **Metric Names** (located in the vertical drop zone) label to the **Break By** drop zone. Now there is one chart, with the bars stacked by the graduate and undergraduate attributes

48. Add the **Metric Names** label to the **Color By** drop zone. Now each bar is colored by the attributes to visually differentiate the stacked bars.
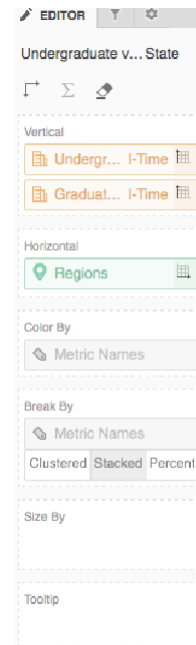
49. Right-click the bar chart horizontal gridlines, and select **Hide** from the available options.
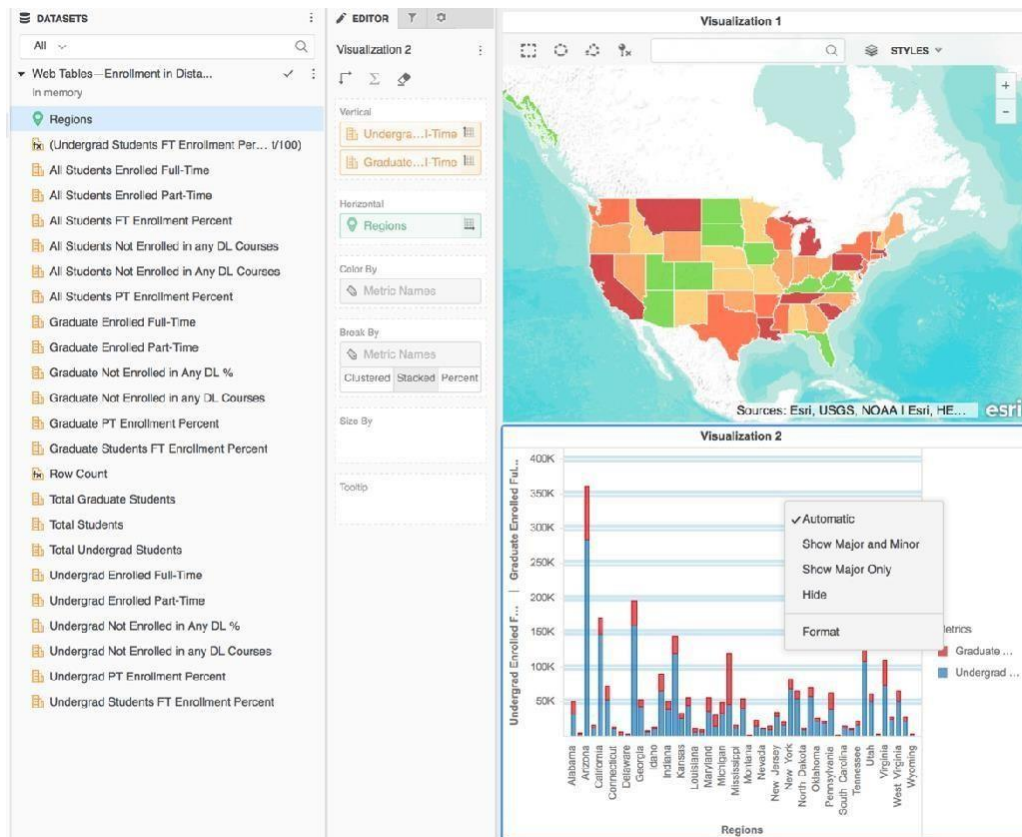
50. Double-click the title of **Visualization 1**, type **Percent of Undergraduate Students per State**, and press **Enter**.

51. Repeat the process for **Visualization 2**, renaming it **Undergraduate vs. Graduate Students per State.**

**52.** To view the dossier, click the **Presentation Mode** icon on the toolbar.
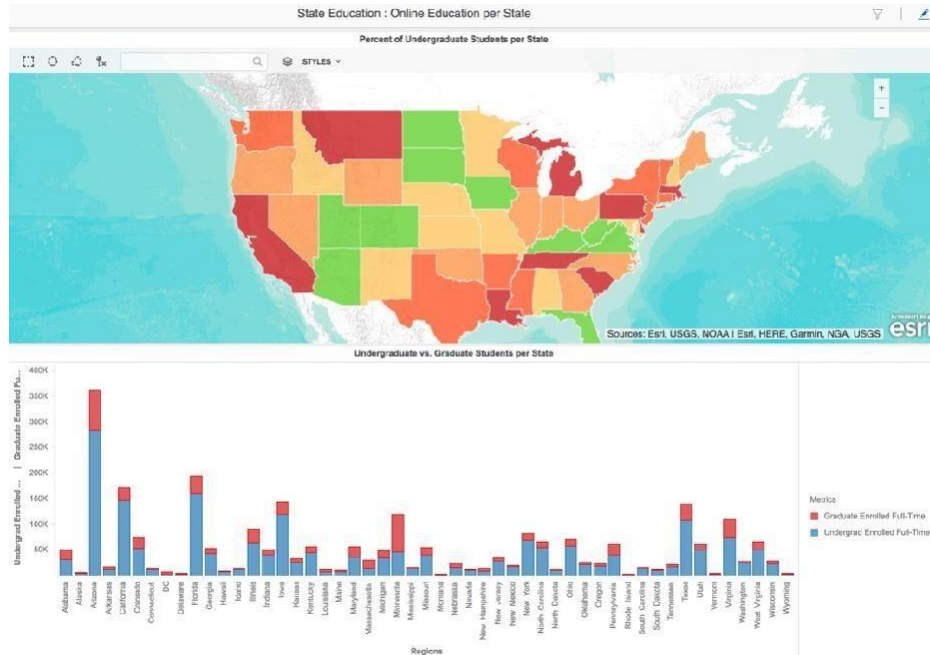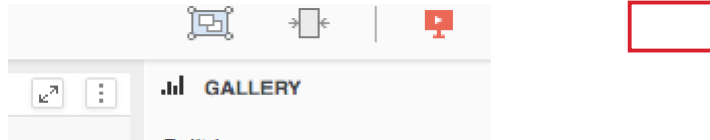
53. To return to edit mode, click the **Edit** icon.

microstrategy.com   21
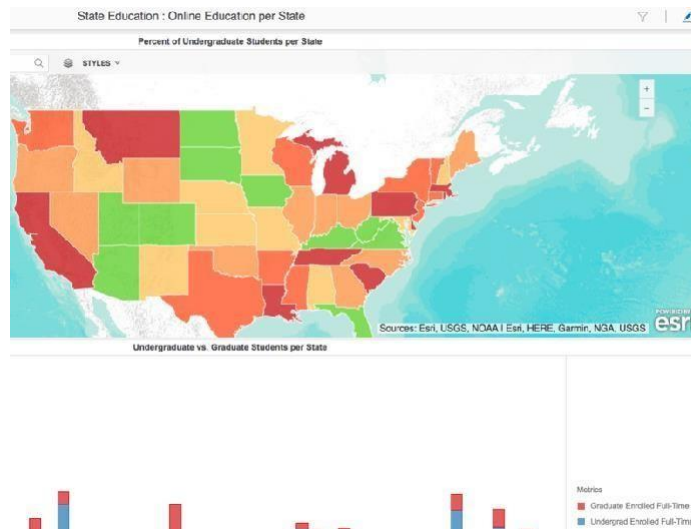
We've now visualized the breakdown of distance education in the United States by State. Next, let's evaluate the type of institution and location that have the largest enrollment of online education in the US. In this exercise, you will learn how to bring JSON files into MicroStrategy Desktop.

Import Institutions.json dataset

54. From the menu bar, click the **Add Data** 🗄️ icon and select **New Data** toopen the Data Source window.
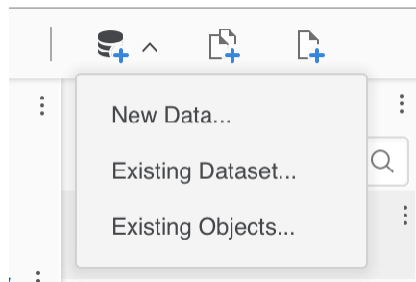
55. For this exercise, you will import data from a JSON file. Select the **File fromDisk** option.

56. Click **Choose Files** to browse your local machine.

57. Browse to **Institutions.json**, then click **Open.**

    This file should be downloaded as mentioned at the start of the exercise

58. *Click* **Prepare Data***.*

    In the preview window, you notice this JSON file holds attributes and metrics, like the previous Excel file. When you right-click **Percent DL Enrollment** and select **Change Data Type,** you will notice this metric is a **Double** data type, while **Total Enrollment** and **Distance Students** are Integers.
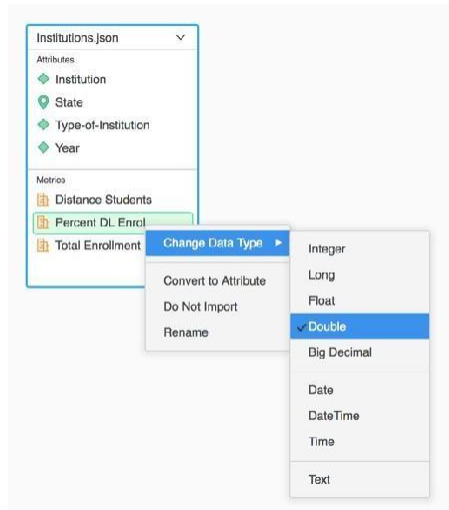
59. Click the **arrow** in the Institutions.json table. Select **Define Relationships.**

60. Define the relationship as below. In each **State**, there are many **Institutions** (one-to-many), and there is one **Type-of-Institution** per each **Institution** (one-to-one).

Define Relationships   ?   x

+ Add New   Convert All ⌄ | Delete All

	Parent Attribute	Child Attribute	Relationship
☰	State ⌄	Institution ⌄	One to Many ⌄
♟	Institution ⌄	Type-of-Institution ⌄	One to One ⌄

Save   Cancel

61. Click **Save** in the Define Relationships window.

62. Click **Finish** in the Preview window.

Now let's add new visualizations in the **Enrollment per Institution** Page.

63. Select the **Enrollment per Institution** page from the Table of Contents.

64. Select the **Bubble Chart** from the Visualization Gallery.

65. From the Institutions.json dataset, drag the **Distance Students** metric to the **Vertical** drop zone. The higher the bubble, the higher the number of distance students.

66. Drag the **Total Enrollment** metric to the **Horizontal** drop zone. Now, the more to the right the bubble, the higher the total enrollment.

67. Drag and drop the **Type-of-Institution** attribute to the **Color By** drop zone. Now, the bubbles are colored and broken by institution type.

68. Next, drop the **Year** and **State** attributes in the **Break By** drop zone (the Type-of-Institution attribute should already be there). Now each bubble represents a year and state, colored by institution type.
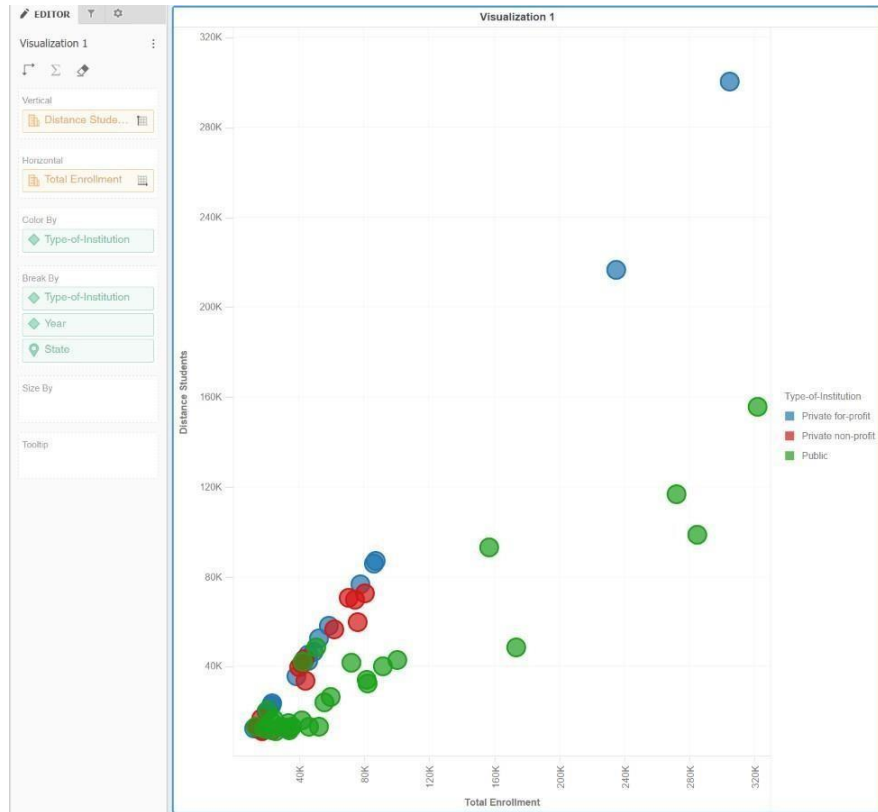
69. You want to see average enrollment percentage for each bubble. First, create the average enrollment metric. Right-click the **Percent DL Enrollment** metric in the Datasets panel and point to **Aggregate by**, then select **Average**.
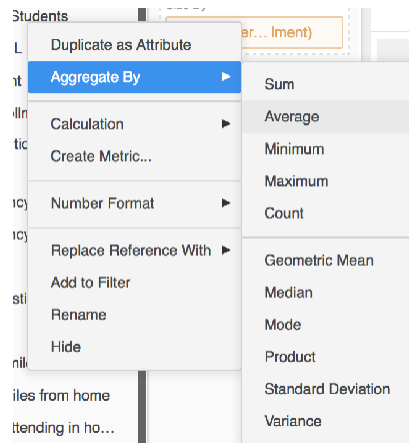
70. Right-click the newly created metric, select **Number Format**, and change from Fixed to **Percentage**.
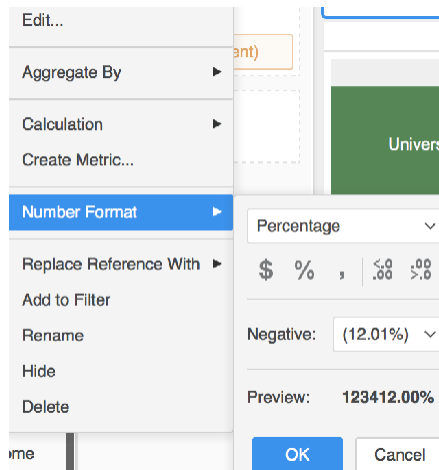
71. Add the **Avg (Percent DL Enrollment)** metric to the **Size By** drop zone. The bubbles are now sized by average distance learning enrollment percentage.
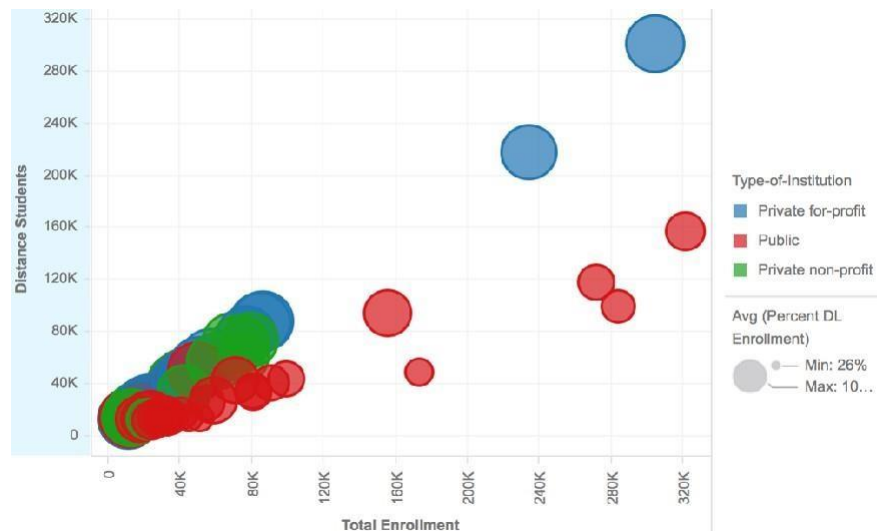
72. You want to add a visualization that shows distance students by institution, year, and percentage of distance students enrolled. Insert a new visualization and drag it beneath the existing bubble chart.

73. Click the **Heat Map** icon in the Visualization Gallery**.**
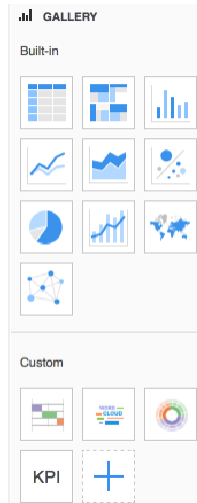
74. Place the **Year** and **Institution** attributes inside the **Grouping** drop zone to break up the heat map by year and institution.

75. Drag the **Distance Students** metric to the **Size By** drop zone. The larger the square, the higher the number of distance students enrolled in that institution. Notice that Distance Students is automatically added to the Color By drop zone as well.

76. In the **Color By** drop zone, replace Distance Students with the **Percent DL Enrollment** metric. Now, each square is colored by percentage of distance students enrolled in the institution.

77. Right-click the **Percent DL Enrollment** metric inside the **Color By** box and click **Thresholds.**

78. Select **Green** from the **Color** drop down.

79. Ensure **Lowest** is selected in the Based on drop down.

80. Change the bands to **25**, **50**, **75**, and **90** units. Click **OK**. Now, the darker green the square, the higher percentage of distance students enrolled.
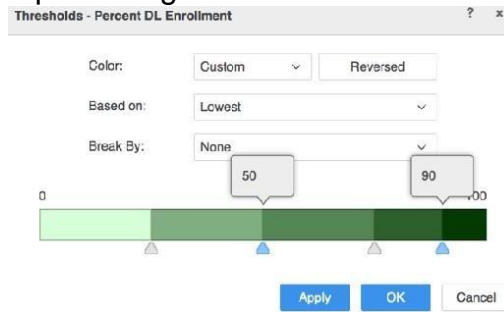
81. Right-click **Percent DL Enrollment** again in the **Color By** drop zone, and change the **Number Format** to **Percentage**.
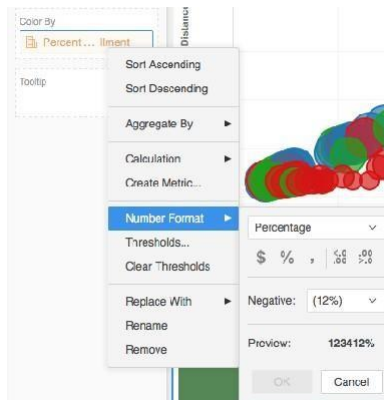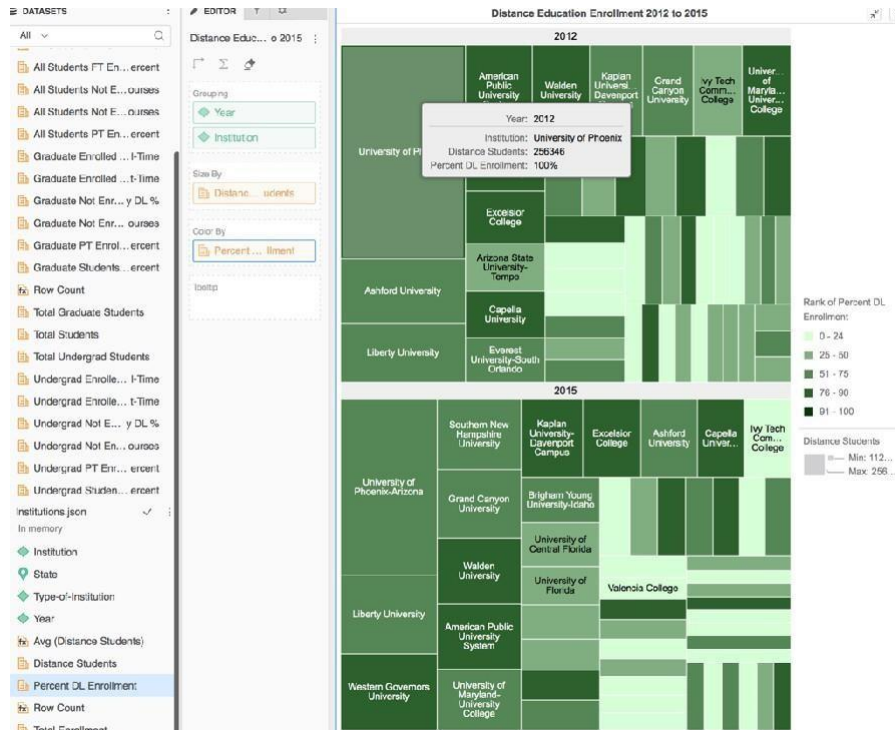

Your visualization should now look like the image below:
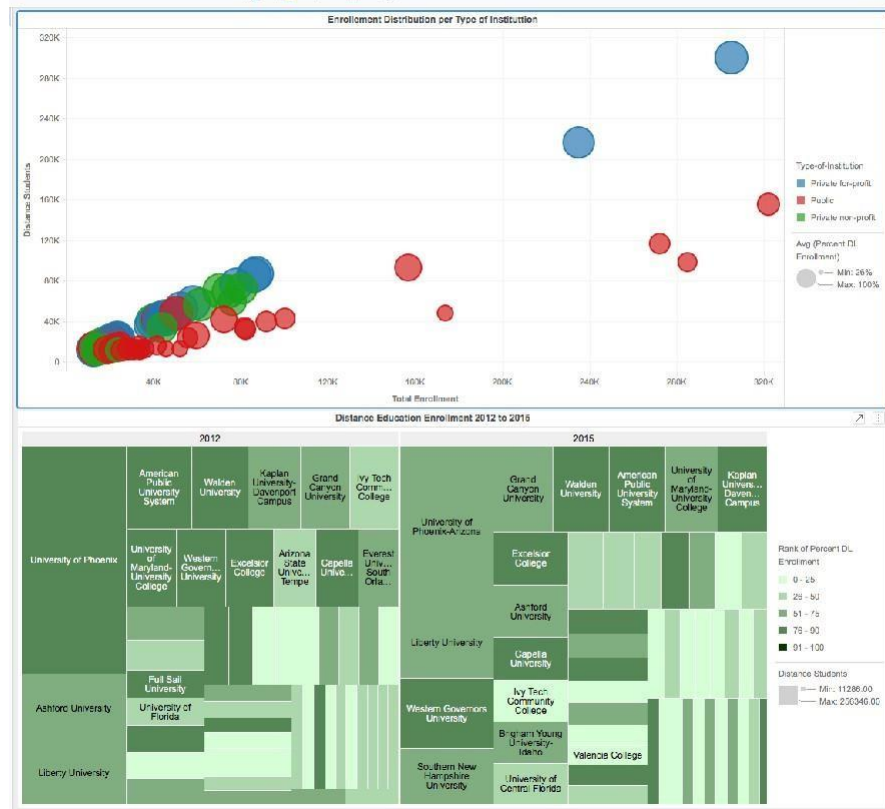
82. Expand your heat map by clicking the **Expand** icon over the title bar.

The expanded heat map should look like the image below:
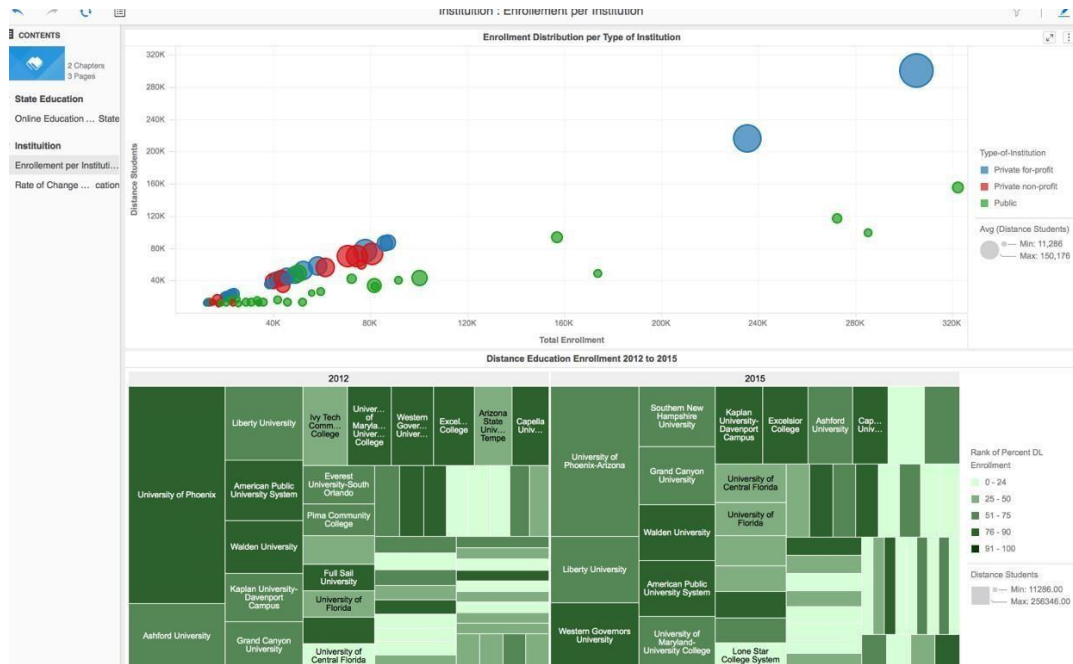
83. Restore the Heat Map visualization back to the previous size.

84. Rename **Visualization 1** to **Enrollment Distribution per Type of Institution**.

85. Rename **Visualization 2** to **Distance Education Enrollment 2012 to 2015**.

86. To view your dossier in Presentation Mode, click the **Presentation Mode** icon.

87. Click the **Edit** icon to return to edit mode.

Next you want to evaluate the type of students who opt for online education. In this section, you will determine the income level and work schedule of Americans studying online by creating two new charts. You will also explore how to handle multiple tables as one single dataset.

88. From the menu bar, click the **Add Data** icon and select **New Data**.
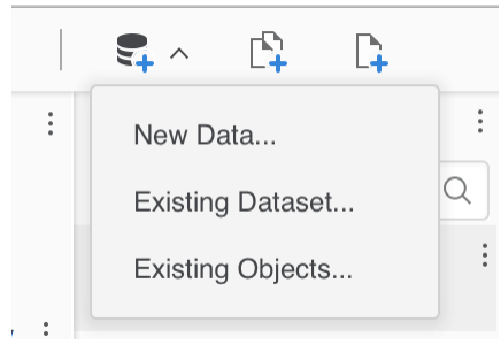
89. Select **File from Disk**.

90. Click **Choose Files** to browse your local machine and import a spreadsheet.
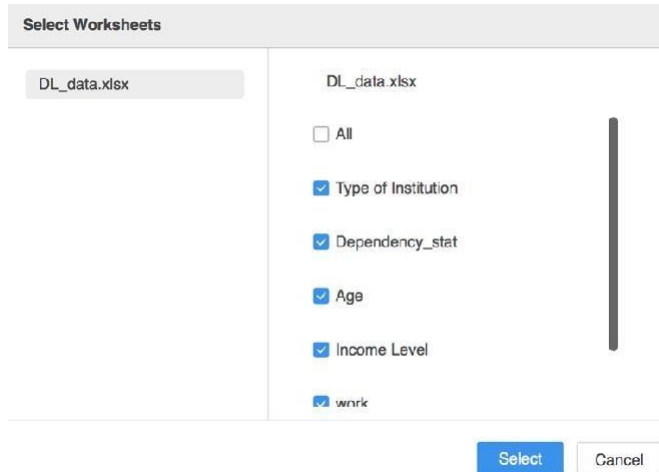
91. Select **DL_data.xlsx** and click **Open.**

   This file should be downloaded as mentioned at the start of the exercise

92. Click **Prepare Data**.

93. Clear the **All** option, and select the remaining worksheets.

94. Click **Select.**

After the data loads, the Preview window opens, which contains information about the attributes and metrics included in the different datasets.

In the preview window, you see six tables. Each table corresponds to the worksheet you selected in the previous window. The third table (**Age**), contains a meaningless attribute called **Column 7**.

95. Right-click **Column 7** and select **Do Not Import**.
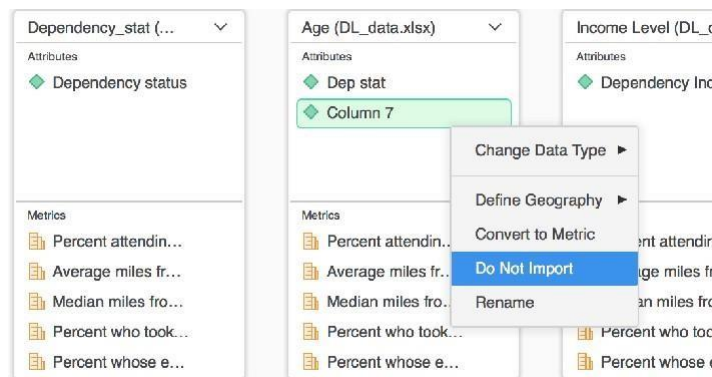
You may notice there is a **Dependency status** attribute in the **Dependency_stat** table and a **Dep stat** attribute in the **Age** table. You will now map these two columns together since we know that they are the same attribute with different names.

96. Right-click the **Dep stat** attribute in the **Age** table, rename it **Dependency status**, and press **Enter**.

97. A notification window will appear asking you whether you want to map this attribute to Dependency status. Click **Map?** to map the attributes.

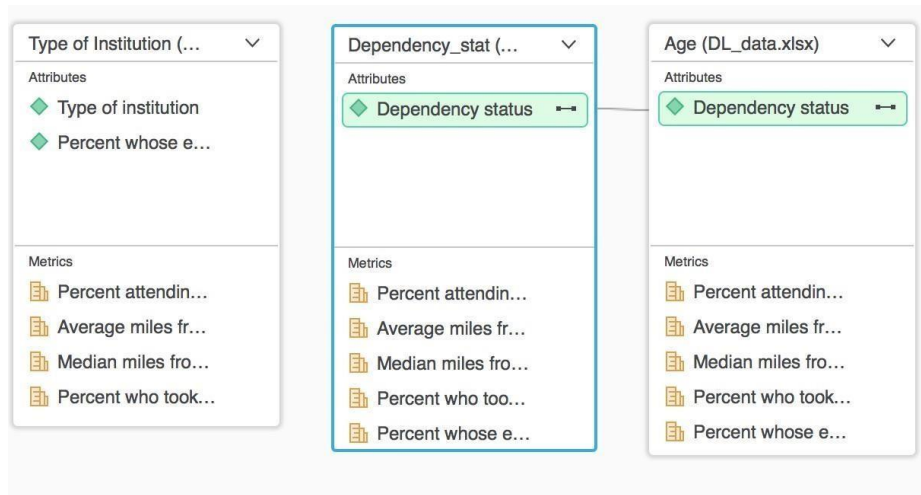A small link icon ⊢⊣ appears next to **Dependency status.** It indicates that these two attributes are now linked.


The first table, **Type of Institution**, contains an erroneous attribute called **Percent whose entire program was distance education**. This column needs to be converted to a metric.

98. Right-click the **Percent whose entire program was distance education** attribute column, and select **Convert to Metric**.
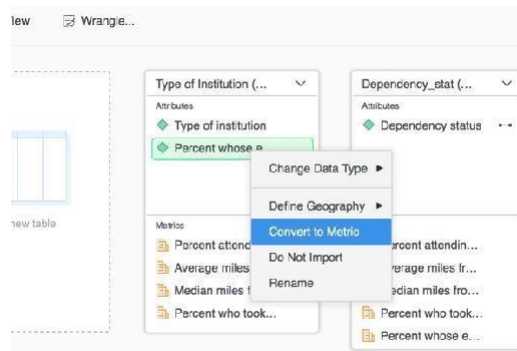
As soon as you convert this column to a metric, it is renamed **Percent whose entire program was distance education(2).**

99. Right-click the new metric and select **Change Data Type** to **Double**.
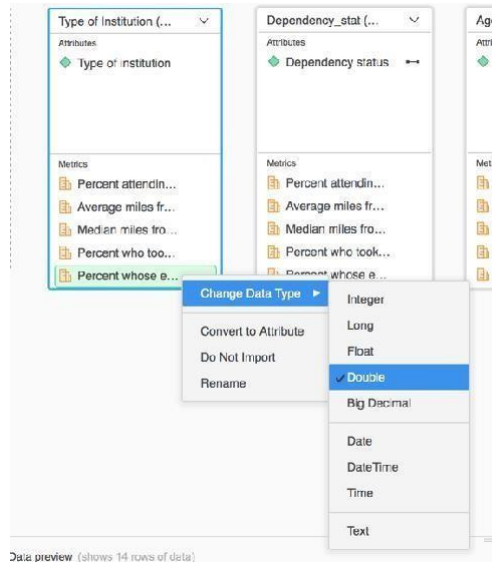
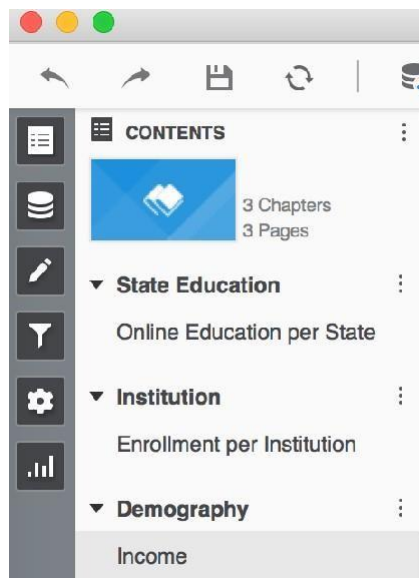**100.** Right-click the metric again and select **Rename**.
101. Change the name to **Percent whose entire program was distance education**.
102. Click **Finish** to return to the dossier authoring window.
103. **Save** your dossier.

Add a new chapter with visualizations

**104.** Create a new chapter called **Demography** with a page named **Income**.

105. Select the **Combo Chart** visualization.

106. From the **DL_data** dataset, place the **Dependency Income level** attribute in the **Vertical** drop zone.

107. Drag and drop the metrics **Percent who took any distance education course in 2011-12** and **Percent whose entire program was distance education** into the **Horizontal** drop zone.

108. Move the **Metric Names** label to the **Break By** and **Color By** drop zones to create one chart with the bars colored by each metric.
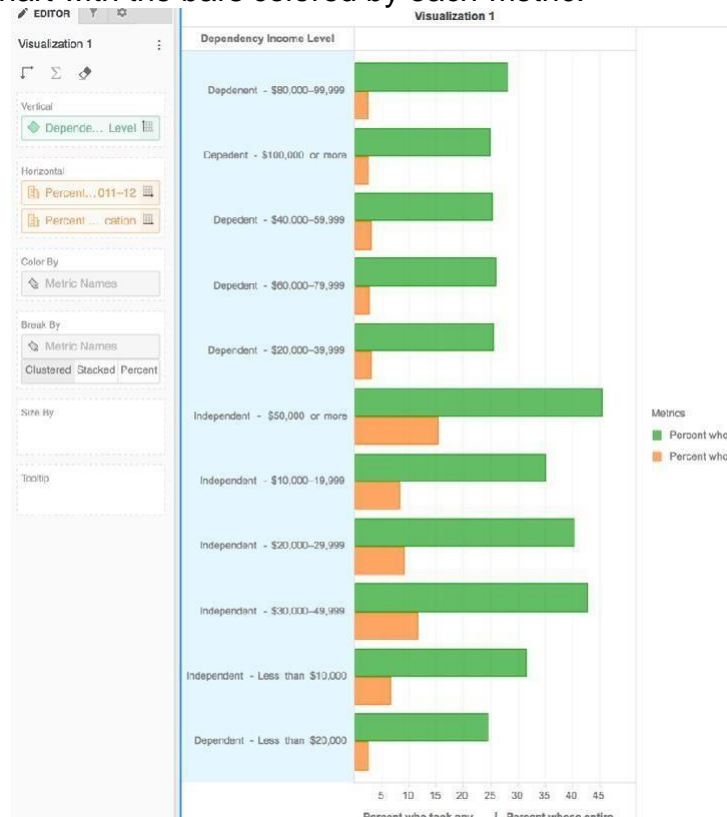
109. Right-click the **Percent who took any distance education course in 2011-12** metric in the **Horizontal** drop zone. Select **Change Shape**, then choose **Tick.**
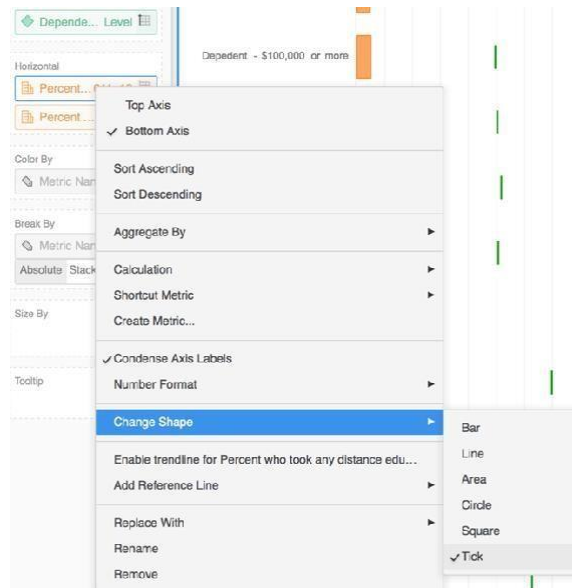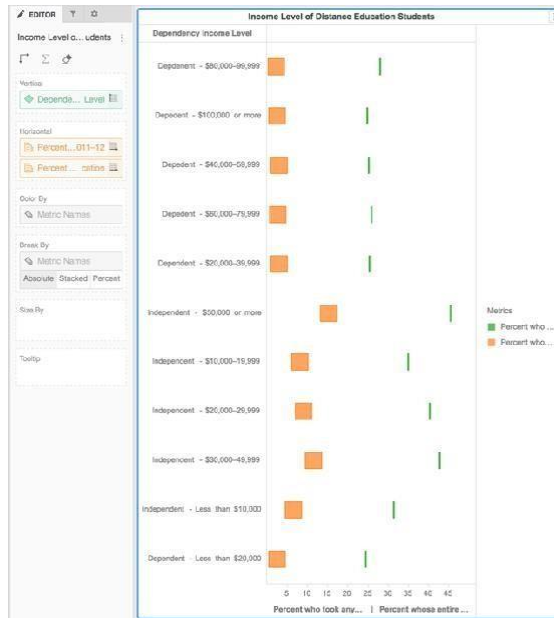
110. Right-click the visualization's vertical **grid lines**, and select **Hide**.
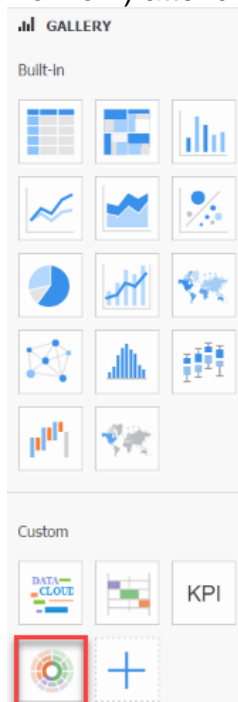

111. Right-click the **Percent whose entire program was distance education** metric. Select **Change Shape**, then **Square**.

112. Rename the visualization to **Income Level of Distance Education Students**.

113. Add a **new visualization** to the **Income** page.

114. Select the **Sequences Sunburst** Chart from the Visualization Gallery. The Sequences Sunburst represents the relative frequencies of a series of events. In this case, your visualization will show how often different types of working students (full time, part time, no work) attended each type of institution.

115. From the **DL_data** dataset, place the **Work** and **Type of institution** attributes in the **Attribute** drop zone.

116. Drag and drop the **Percent whose entire program was distance education** metric to the **Metrics** drop zone.

117. Rename the visualization to **Working Professionals Pursing Online Education Entirely**.

118. Move the Sequence Sunburst visualization under the first visualization.

119. Save your dossier.

Your Income page should now look like the image below: