Add Data Visualization and Discovery

by Mark Rittman

Use Oracle Business Intelligence Visual Analyzer Cloud Service to see more and see it more clearly.

"Upload, Model, Analyze, and Report" (Oracle Magazine, May/June 2015) showed how you can use Oracle Business Intelligence Cloud Service to share dashboards within a department by just uploading a spreadsheet and building the data model and analyses with simple web-based tools. This follow-up article focuses on the new Oracle Business Intelligence Visual Analyzer Cloud Service, which brings interactive data visualization and data discovery capabilities to Oracle's cloud business intelligence services.

In this example, I use one of the data sets provided by the DonorsChoose.org online charity (which connects donors with US public school projects seeking funding). Supported by Oprah Winfrey and Stephen Colbert, among others, DonorsChoose.org provides data that is often used in data science challenges. I will employ its Projects data set to provide a sample file of projects, schools, project donations, and other data available to potential donors.

For your convenience, you can download the Public Donations Data BICS Samples data set from Oracle Technology Network to your Oracle Business Intelligence Cloud Service environment so that you have the data and the model ready and can follow along with the steps presented in this article.

Exploring the DonorsChoose.org Data Set with Oracle Business Intelligence Visual Analyzer Cloud Service

Let's start by creating some initial data visualizations to help parse the data in the DonorsChoose.org Projects data set. At a high level, the metrics this article will pull from the data set include

• The total number of projects in the data set and how many of those projects are currently open for donation

• A breakdown of those open projects by their primary focus area and the type of resources they require—technology, books, and so on

• A breakdown of projects by the "poverty level" of the school requesting the money, defined by the DonorsChoose.org charity according to the percentage of students qualifying for a free or discounted lunch

If you haven't done so already, upload the data and snapshot file provided in the article download and log in to your Oracle Business Intelligence Cloud Service instance using the username, password, and identity domain provided by your Oracle Business Intelligence Cloud Service administrator. Then follow these steps to create your initial Oracle Business Intelligence Visual Analyzer Cloud Service project:

1. From the Oracle Business Intelligence Cloud Service home page, click the **Create a Visual Analyzer** Project link in the blue menu panel on the left side of the page. When prompted to select a datasource, click the **DONATIONS** datasource and click **OK** to proceed. The page will then be divided into two main sections: (1) a concertina-style menu panel on the left-hand side displaying the folders and columns you can add to the visualization, and (2) an empty visualization panel on the right-hand side awaiting your addition of columns, filters, and other settings.

2. Let's start by simply showing the number of projects in the data set. Click the **FACT** folder in the left-hand menu panel to display the list of measures within it. Then double-click the **# OF PROJECTS** column to add it to the data visualization, which will then display a horizontal stacked bar chart with a single bar for that measure.

3. This single measure value would probably be better displayed as a tile visualization, so click the down arrow next to the **Horizontal Stacked** label in the visualization settings panel (located between the column list and the visualization itself) and select **Tile**. The visualization will now switch to the Tile type, and the overall total for all projects will be displayed in a large font as a number rather than a single bar in a bar chart.

4. This number is actually the total number of projects, including those that are already completed. To display only the currently active projects, you filter that total by visualizing only those projects with a funding status of "live." To do this, click the **PROJECT** folder on the left-side panel and then right-click the **FUNDING_STATUS** column from the list displayed under that folder label. From the menu that is then displayed, select **Create Filter**, as shown in **Figure 1**.
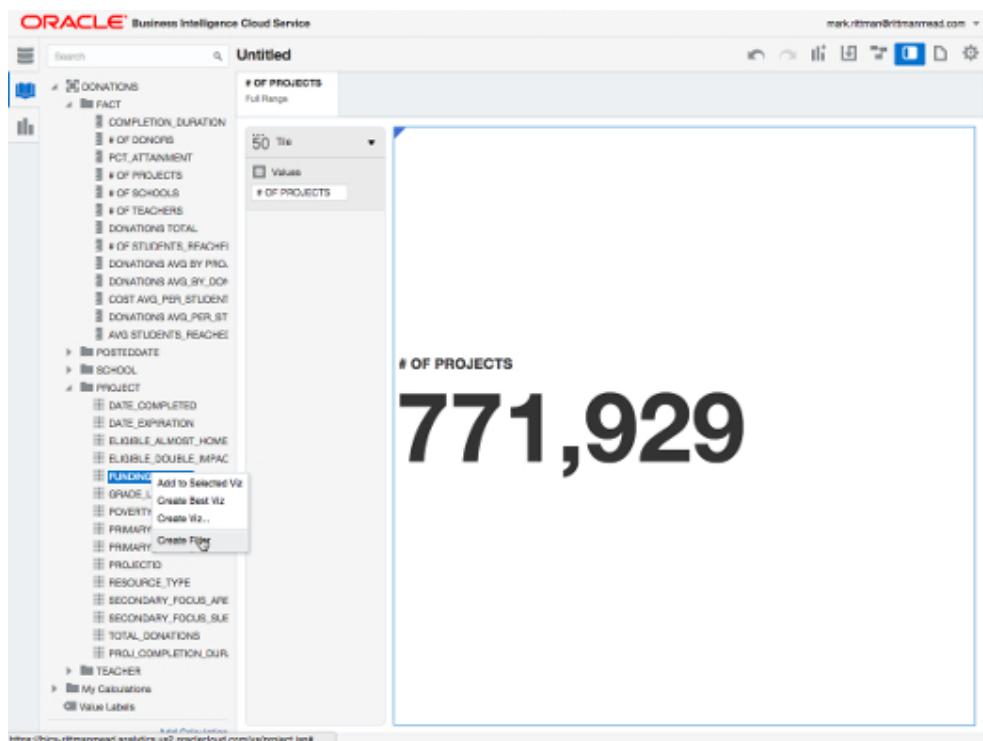

Figure 1: Displaying a tile visualization

With the filter dialog box displayed, select live as the filter value and then click anywhere on the data visualization pane to show only the active projects.

5. Next let's see the number of projects subdivided by primary focus area. You can add this break-down as a second visualization within the same project by right-clicking anywhere on the tile data visualization and selecting **Copy**, right-clicking the same visualization again, and selecting **Paste**.

In the central visualization settings panel, change the visualization type from **Tile** to **Bar** so that a single vertical bar is redisplayed. Now double-click the **PRIMARY_FOCUS_AREA** column within the **PROJECT** folder on the left side of the screen to add it to the visualization settings. This will break the bar chart down by the focus area attribute; you should now see a vertical bar chart with **PRIMARY_FOCUS_AREA** as the chart's x axis and **# OF PROJECTS** as its y axis.

6. To further sub-divide these results, double-click the **POVERTY_LEVEL** column (also located in the **PROJECT** folder at the left). Then change the visualization type to Stacked Bar. Your project will now resemble **Figure 2**.
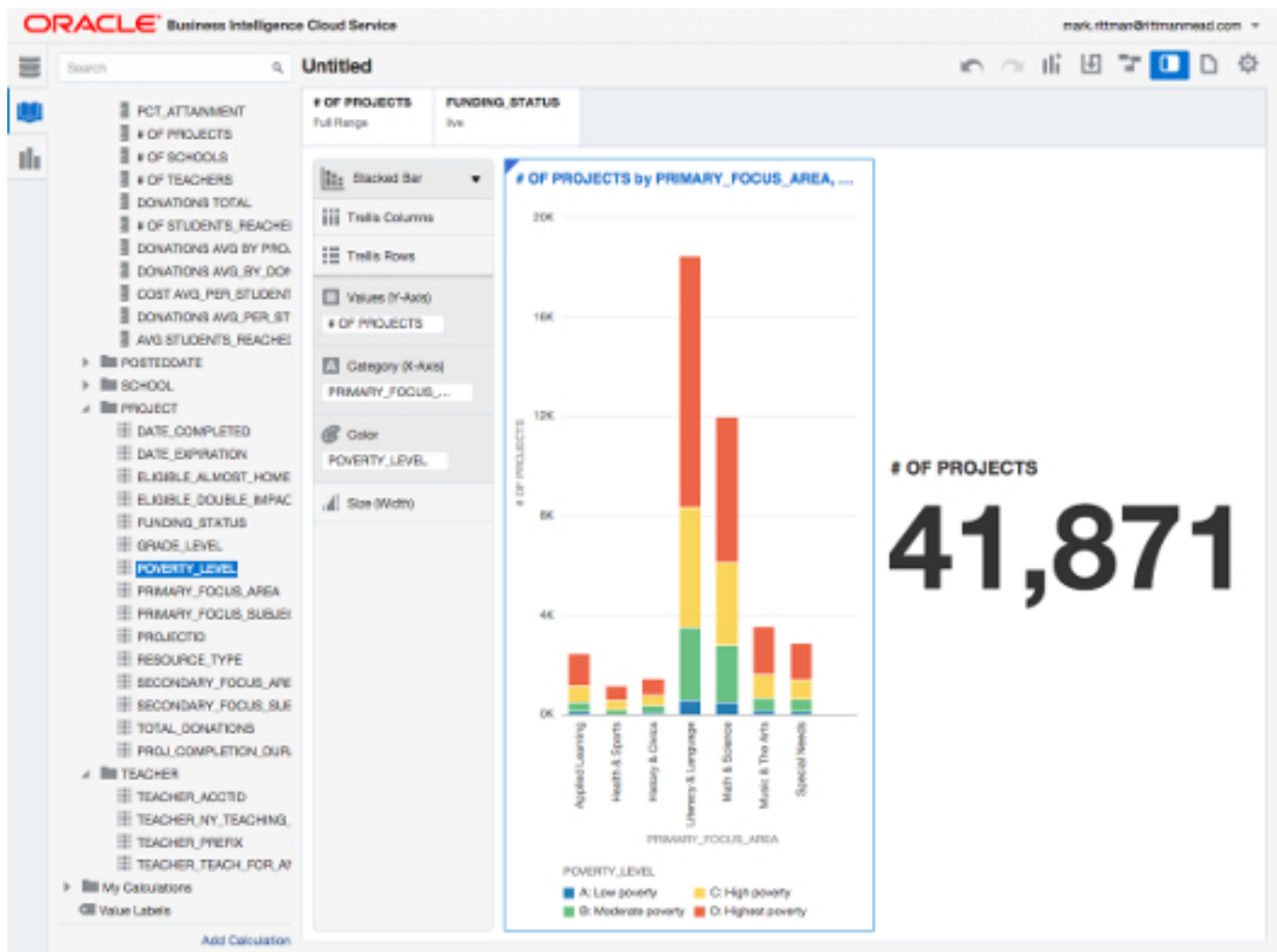

Figure 2: Adding a Stacked bar Chart Visualization

7. Let's create a final visualization for this project to see how the type of resource requested affects the total number of students reached across the various school poverty levels. You can display those totals by using "treemap" visualizations based on project focus areas to discern whether any focus areas particularly contribute to each total.

Start by Ctrl-selecting the **POVERTY_LEVEL, RESOURCE_TYPE**, and **PRIMARY_FOCUS_AREA** columns within the **PROJECT** folder and the **# OF STUDENTS_REACHED** column within the **FACT** folder, and then right-click and select **Create Viz**…. When prompted, select **Treemap** as the visualization type. Then, using the default settings, move the **POVERTY_LEVEL** column from the **Category (Boxes)** setting to the **Trellis Columns** area.

8. Arrange the visualizations within your project by dragging and dropping them around on the project canvas so that your final project looks like **Figure 3**.
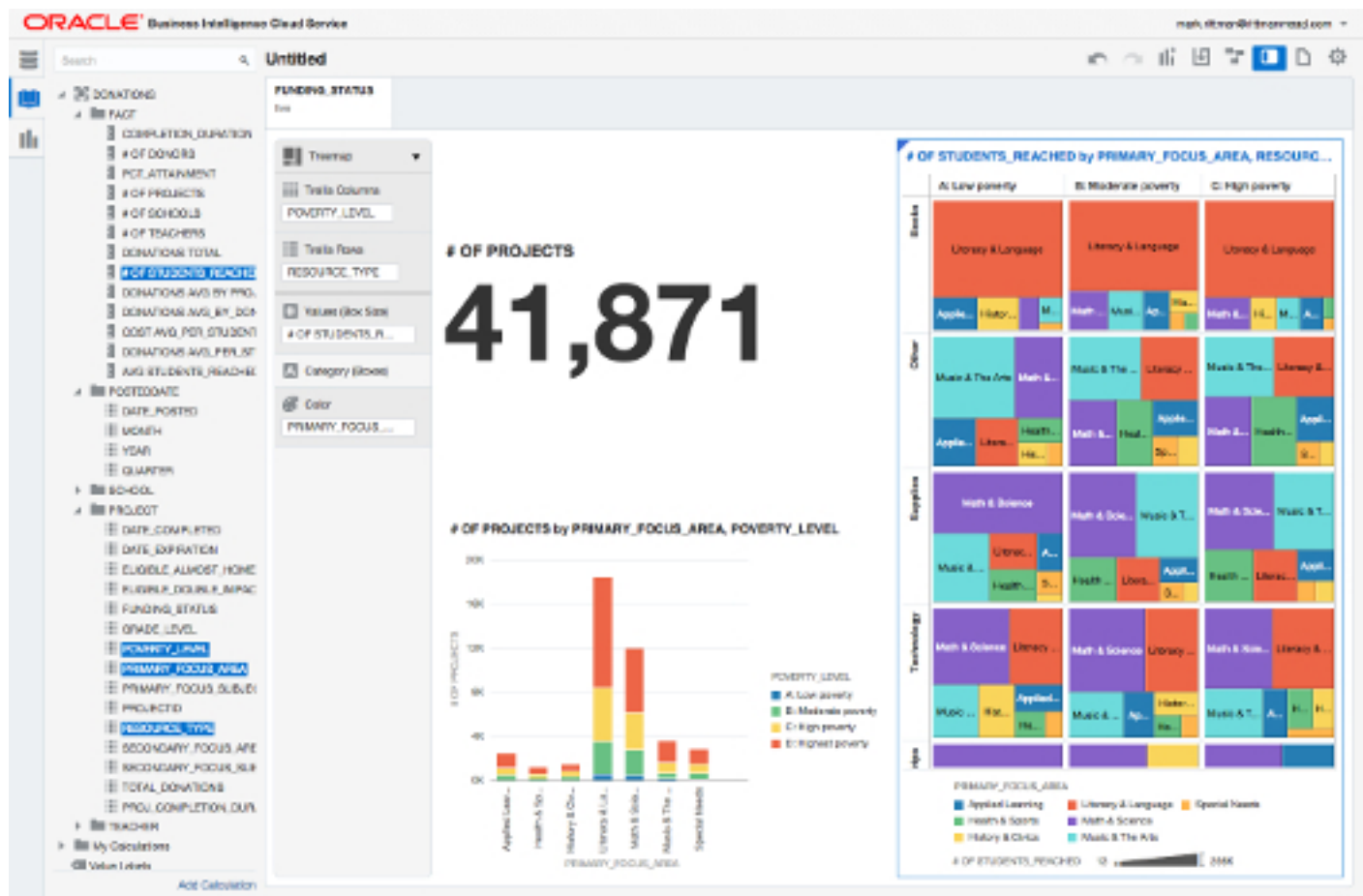

Figure 3: Three Visualizations

You can save this project to the catalog by clicking the **Save Project** button at the top right of the page, selecting **Save As** from the menu, and typing in a project name. Click **OK** to save the project and close the dialog box.

Using Oracle Business Intelligence Visual Analyzer Cloud Service for Data Discovery

Now let's use Oracle Business Intelligence Visual Analyzer Cloud Service for data discovery. Using the same DonorsChoose.org data, let's identify a project category where a donation will make the greatest impact in terms of reaching the highest average number of students among projects that don't yet have many donors. You can explore the data set to see which projects' primary focus areas are most likely to meet this requirement, and you can also identify projects that will benefit from funding bonuses and matching to make the overall donation go further.

1. Start by returning to the Oracle Business Intelligence Cloud Service home page. Click the menu item link to **Create a Visual Analyzer Project** again, and then select the **DONATIONS** datasource when prompted.

2. Now Ctrl-select the **# OF DONORS** and **AVG STUDENTS_REACHED** columns from the **FACT** folder and the **PRIMARY_FOCUS_AREA** attribute from the **PROJECT** folder. When you have two measures and a single dimension, a scatter diagram works best for showing the correlation between the measures for that dimension's attribute values. Right-click and select **Create Viz…**, and then select **Scatter** from the set of visualization types.

3. To filter for just those projects open for donation, right-click the **FUNDING_STATUS** column within the **PROJECT** folder. Select **Create Filter…**, and then select **live** as the filter value.

4. Now, if you hover your cursor over the various points in the scatter visualization, you can view the average number of students reached and the number of donations for each primary focus area—with project types that attract more donors located farther to the right along the x axis and projects reaching more students higher up the y axis.

Using the cursor and clicking to select, you can draw a selection box around the projects in the top left of the visualization (high average number of students reached, low number of donors) to target where a donation that meets your criteria will make the most impact, as shown in **Figure 4**.
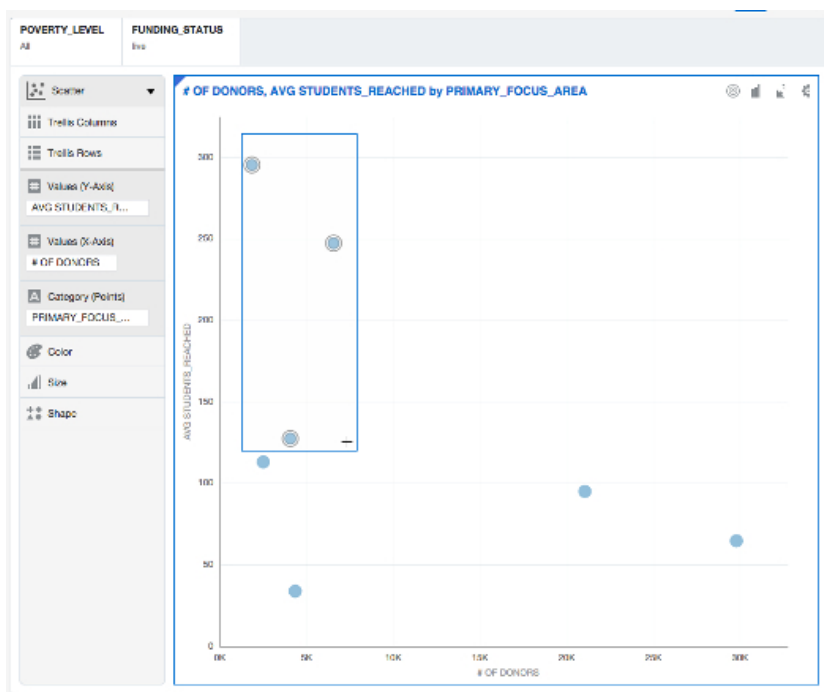
If you now right-click within that selection and choose **Keep Selected** from the menu that appears, as shown in **Figure 5**, all other points on the scatter chart will be removed. The chart layout will then automatically adjust to display those points that remain across the full height and width of the visualization.
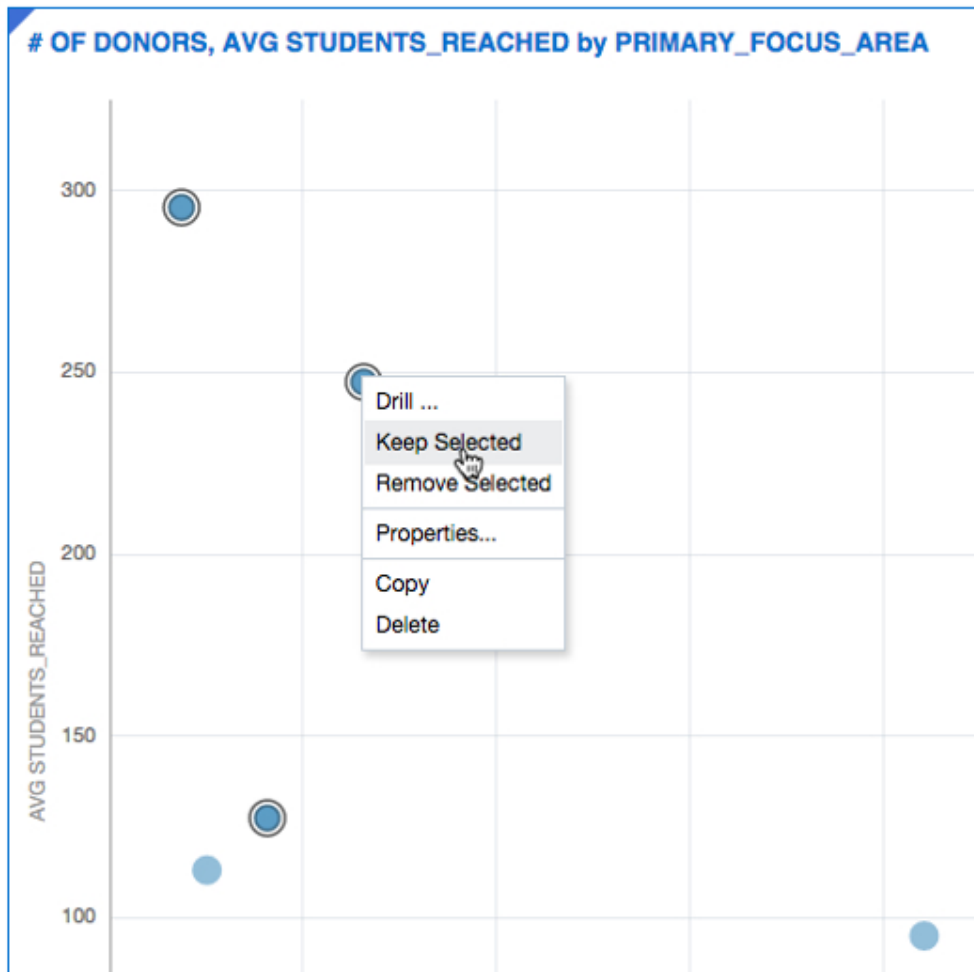

Figure 5: Using the Menu to Keep Only the Selected Points

5. Our scatter visualization now has just three points displayed within it—for the Health and Sports, Music and the Arts, and Applied Learning primary focus areas. Because you also want to target projects that attract a double-impact match (where foundations or corporations agree to match any donation amount given for special qualifying projects), right-click the **ELIGIBLE_DOUBLE_IMPACT_MATCH** column within the **PROJECTS** folder on the left-hand side of the page, select Create Filter, and then choose t (for true) from the two options displayed. This adds another filter and retains only those projects meeting the match requirement in the visualization.

6. To further examine these remaining projects and break them down by the resource type requested, drag the RESOURCE_TYPE column from the PROJECTS folder and drop it onto the visualization's Color setting. Then, drag and drop the # OF PROJECTS column onto the scatter visualization's Size setting area. The points will now be color-coded and sized larger or smaller to indicate the number of projects that each represents in total. Your visualization should now resemble **Figure 6**.
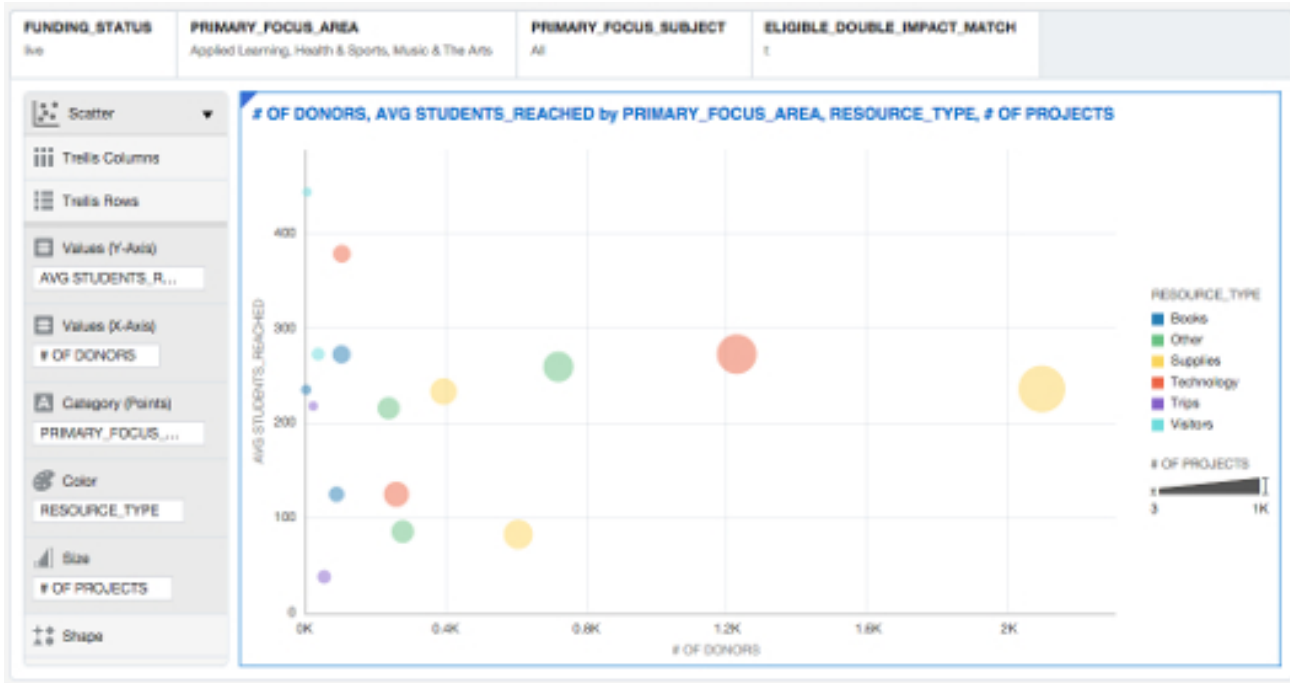
Figure 6: Breaking Down the Data Set By Additional Column Attributes

If you now hover your cursor over the point in the scatter visualization that represents the highest average number of students reached and one of the lowest numbers of project donors, you can see that this represents 60 projects with Health & Sports as the primary focus area and Technology as the resource type, as shown in **Figure 7**.
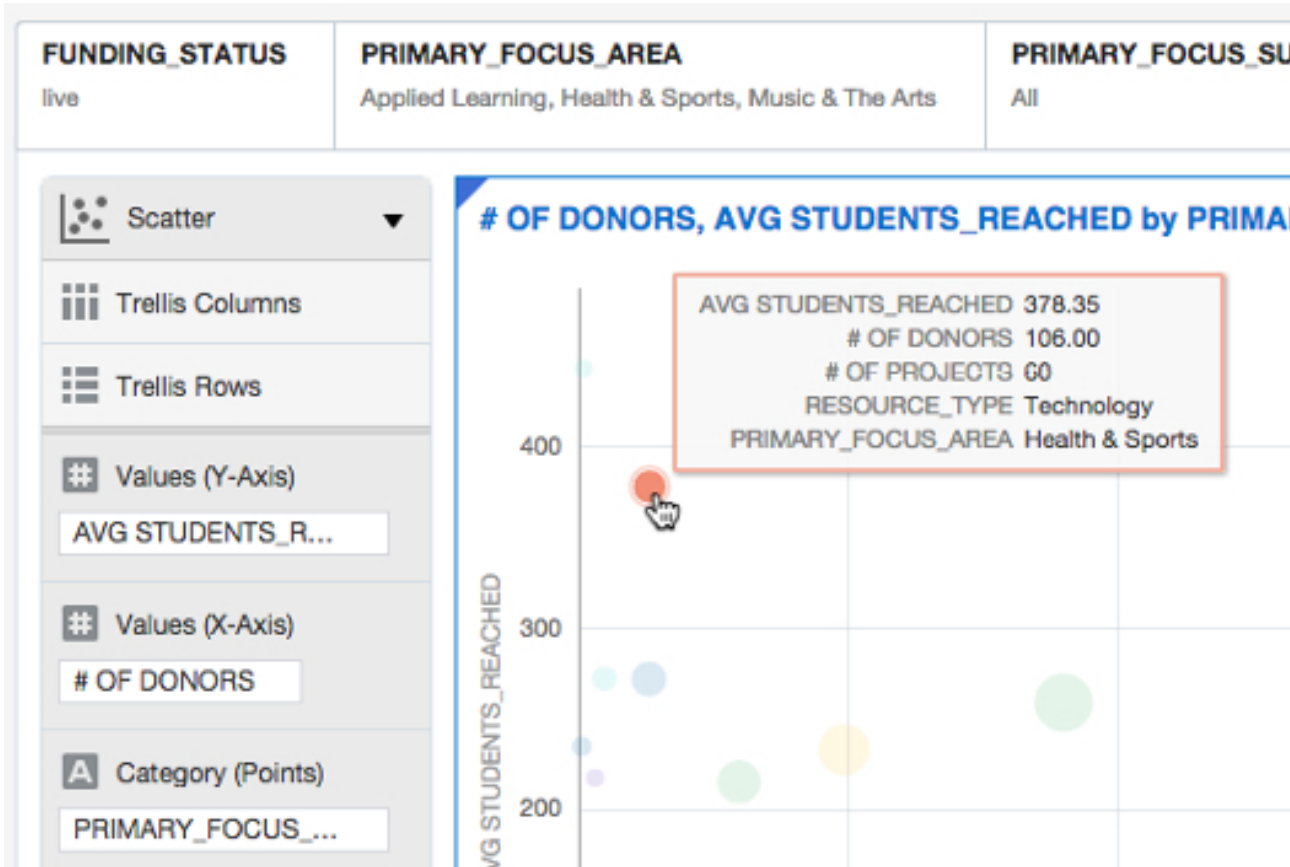

Figure 7: Using the Tooltip to Display Column Attribute Values for a Point

Finally, to locate where these projects are running, let's turn the visualization into a map and plot the schools running these projects on it.

7. To create the map, first change the visualization type from **Scatter** to **Map** by clicking the down arrow at the top of the left column (next to the **Scatter** setting) and selecting **Map**. Then right-click the **PRIMARY_FOCUS_AREA** column that has been automatically designated as the **Category (Geography)** selection. Select **Create Filter**, choose **Health & Sports** as the filter value, and then remove **PRIMARY_FOCUS_AREA** as the **Category (Geography)** selection by hovering your cursor over it and clicking the **X** (Remove Column) button. Next, repeat this step for the **RESOURCE_TYPE** column—which has also been automatically selected as the new **Color (Map Shape)** setting— filtering **Technology** as the only resource type and then removing **RESOURCE_TYPE** from that setting by clicking the **X** button next to it.

8. Then, to actually plot the schools onto this map visualization, drag and drop the **SCHOOL_LATI-TUDE** column onto the **Category (Geography)** area where **PRIMARY_FOCUS_AREA** used to be. Then drag and drop the **SCHOOL_LONGITUDE** column just under it, so that both columns are used to provide the required latitude and longitude map settings. Finally, drag and drop the **# OF PROJ-ECTS** column from the **FACT** folder onto the map visualization's **Size (Bubble)** setting, so that bubbles representing schools are sized according to the number of projects each school is running, and drag and drop the **DONATIONS TOTAL** column from the same folder onto the **Color (Bubble)** setting, to help delineate already well-funded schools and projects.

Your final Oracle Business Intelligence Visual Analyzer Cloud Service map visualization should now look like **Figure 8**, showing the location of each school running projects that fit your criteria.
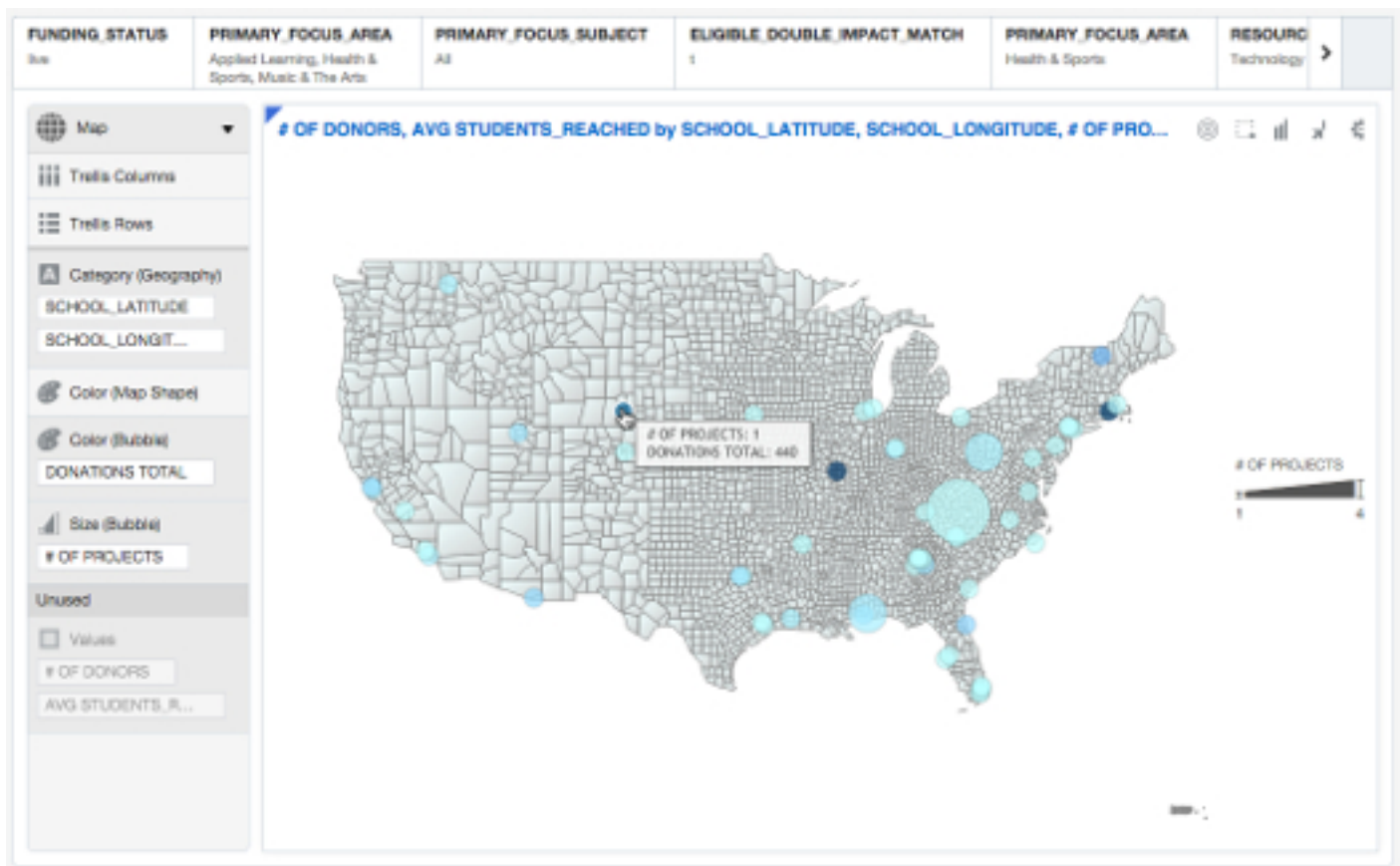

Figure 8: Mapping Eligible Schools and Showing Project Count and Donation Totals

Conclusion

Oracle Business Intelligence Visual Analyzer Cloud Service adds data discovery and visualization capabilities to the reporting and dashboard capabilities of Oracle Business Intelligence Cloud Service. Using a single web page that enables you to easily drag and drop data items onto a canvas, quickly switch between visualization types, and manipulate your view of the data with simple graphical tools, Oracle Business Intelligence Visual Analyzer Cloud Service is the ideal tool for exploring and analyzing enterprise data sets—all within Oracle's easy-to-use cloud environment.