Understanding and Transforming Data with Tableau

EXERCISE PACKET

Contents

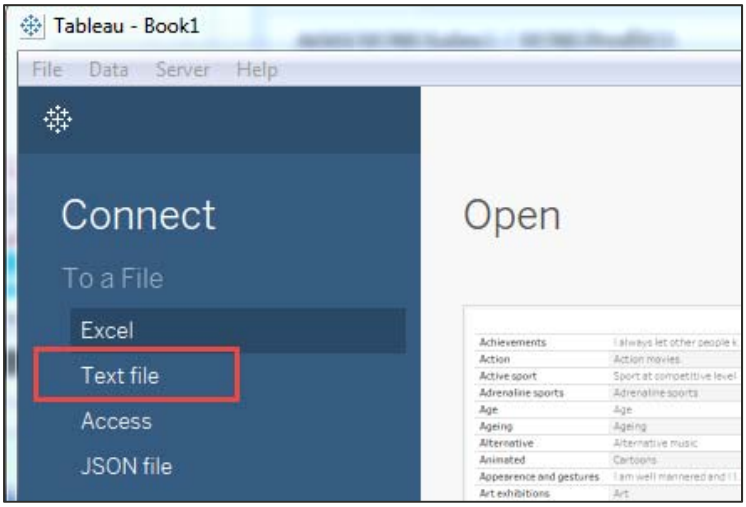
EXERCISE 1 – 10 MINUTES

Objective

1. Create data sources from the following data files in the exercise data sets folder:
   - bike_trips.csv
   - invoice_headers.csv
   - invoice_holds.csv
   - invoice_lines.csv
   - survey_data.csv
2. For each data source check the values of various fields and determine if the field needs to be adjusted in terms of dimension/measure, or discrete/continuous

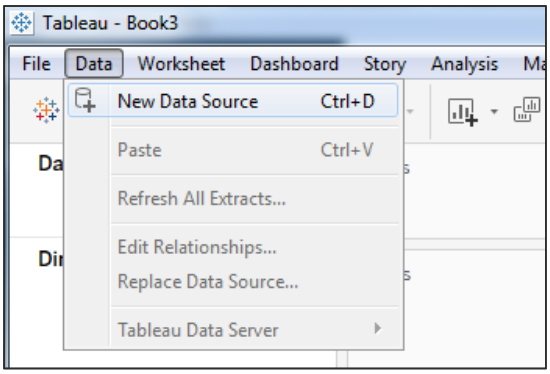
Creating a data source from a csv file:

1. In a new Tableau window, click on **Text file** on the right hand pane



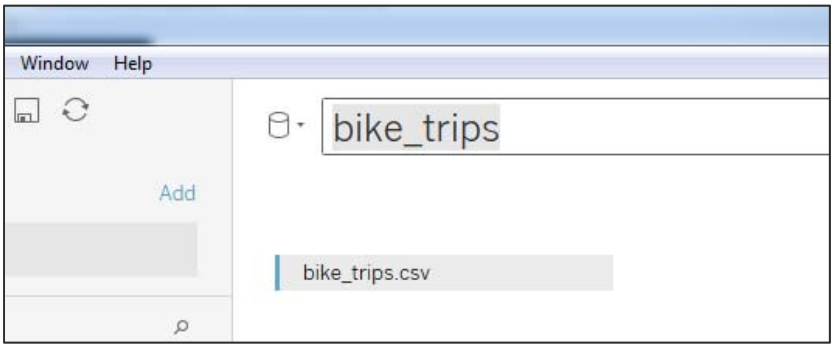
OR

In an existing Tableau file, click on the **Data** menu, select **New Data Source,** and then click **Text file**



2. Browse to and double click on the csv file

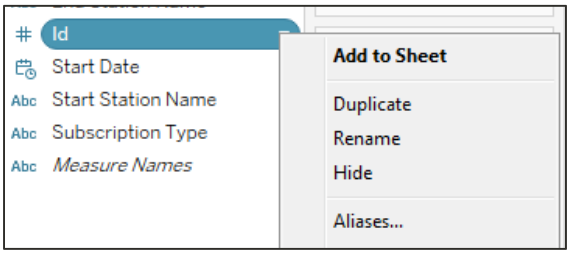
3. Optional: rename the csv file by clicking on the field below and typing a new name



4. Click on a sheet at the bottom of the Tableau window to leave the screen

Checking the values of a specific field:

1. Highlight over a field and click the drop down arrow on the right



2. Select **Describe** at the bottom of the menu
3. Click the **Load** button on the pop window

Converting between dimension and measure

1. Highlight over a field and click the drop arrow on the right
2. Select **Convert to Measure** or **Convert to Dimension**

OR

Drag the field to either the Dimension or Measure left-hand panes in Tableau

Converting between discrete and continuous

1. Highlight over a field and click the drop arrow on the right
2. Select **Convert to Discrete** or **Convert to Continuous**
3. Notice the color of the column name changes between blue and green

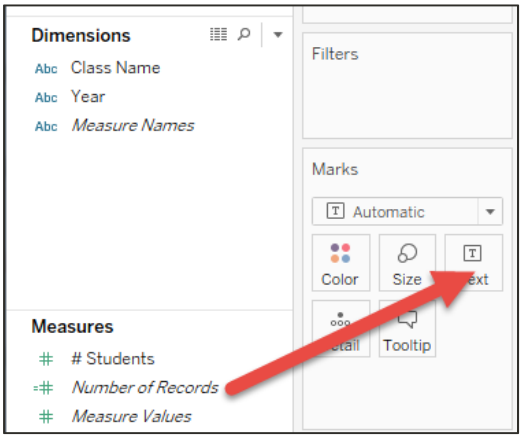
EXERCISE 2 – 15 MINUTES

Objectives

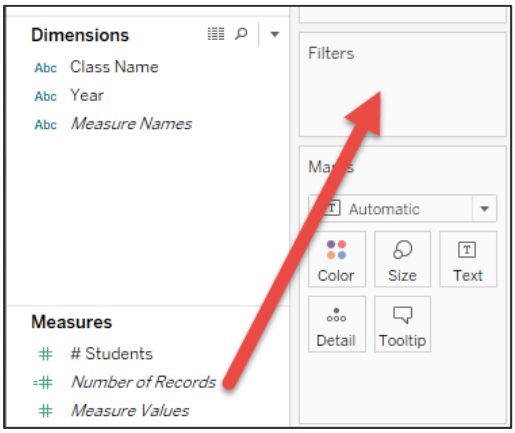
1. For each data source created in Exercise 1, find and verify the granularity
2. Create a copy of the survey data source and pivot on a set of columns. Determine what columns you should include or exclude in the pivot. Why?
3. Verify the granularity of the pivoted survey data source
4. Union the 5 individual question survey data files into one data source. What is the granularity before and after the union?

Verifying granularity

1. Drag the dimension(s) onto the **Row** shelf
2. Drag the **Number of Records** measure onto the **Text** self



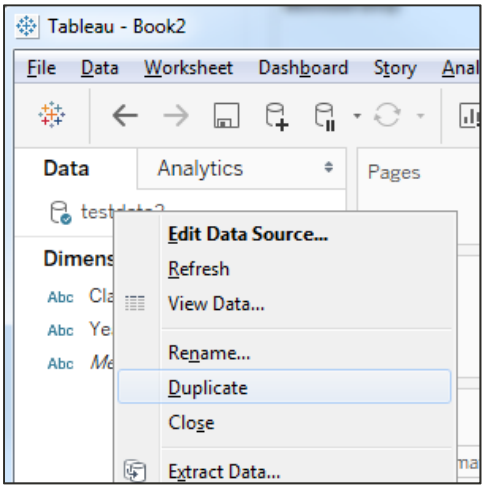
3. Drag the **Number of Records** measure again onto the **Filter** shelf



4. Double-click on **Sum**
5. Click on the **At least** box
6. Enter **2** in the number field provided and click **OK** button
7. If dimension(s) in the Row shelf does indeed define the granularity of the data set, then no rows should appear

Copying a data source

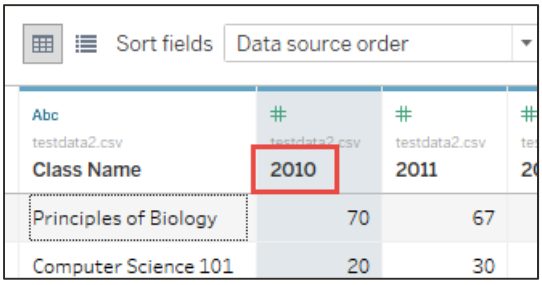
1. **Right click** on the data source in the **Data** pane to copy and select **Duplicate** in the menu



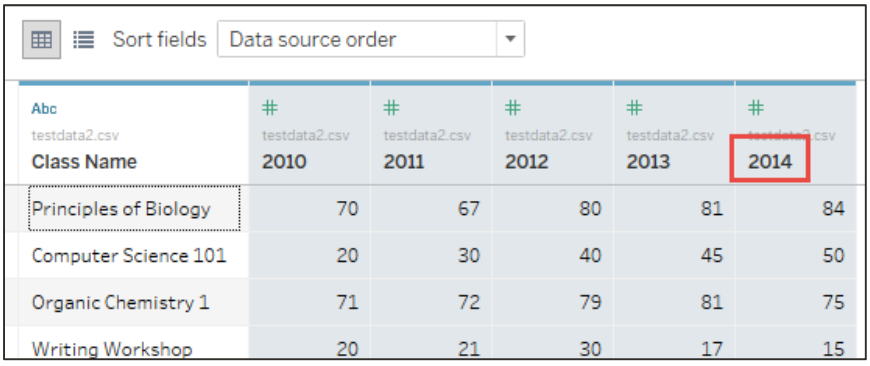
2. **Right click** on the copied data source and select **Rename** in the menu to specify a new name

Pivoting columns

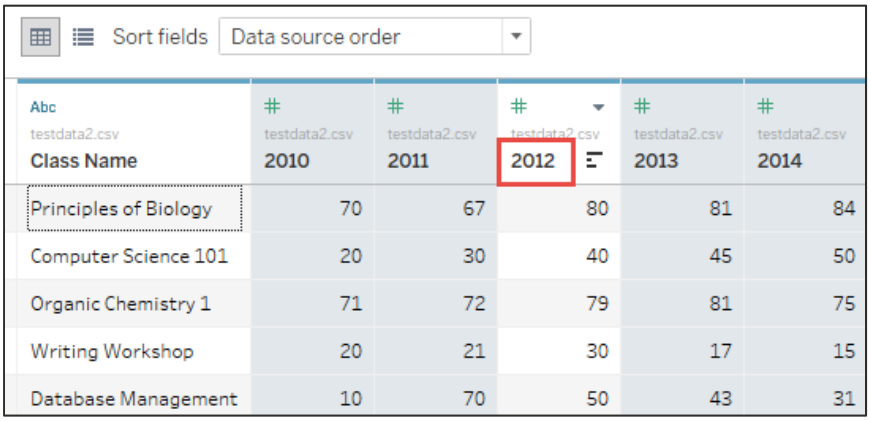
1. **Right click** on the data source in the **Data** pane and select **Edit Data Source** in the menu
2. In the edit data source view, highlight the first column to pivot by clicking on the column name



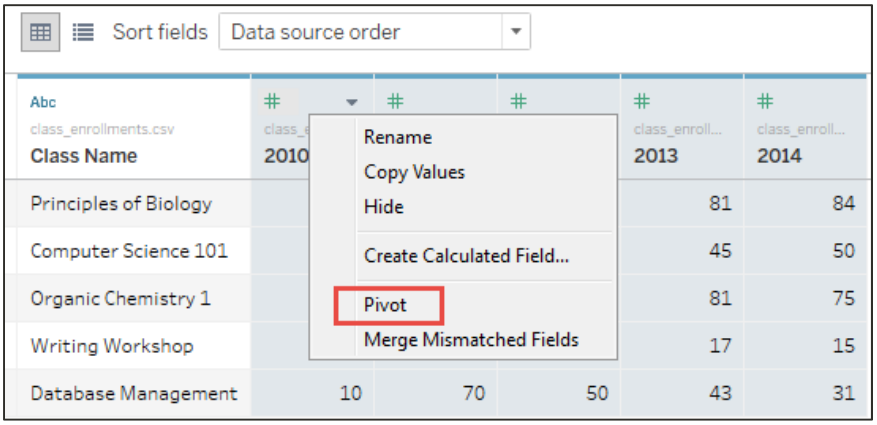
3. Hold **Shift** on the keyboard and click on name of the last column to pivot. This will highlight every column in between



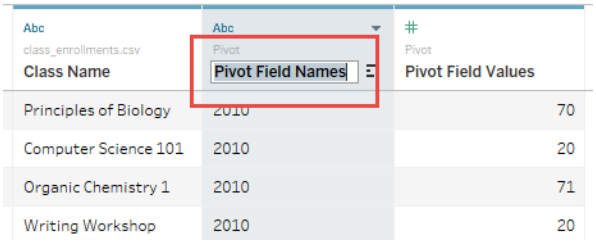
4. To un-highlight specific columns, click on name of the column while holding the **Ctrl** key



5. **Right click** over the highlighted columns and select **Pivot** in the menu

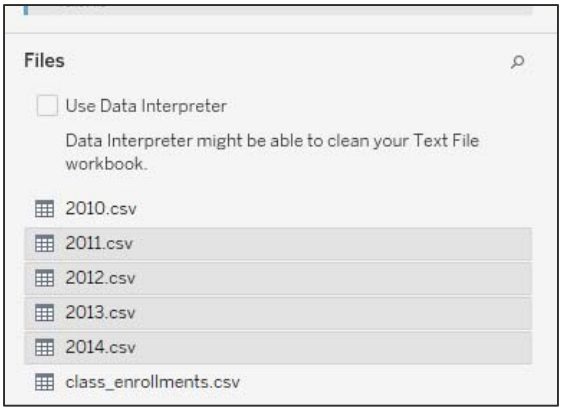


6. **Double click** on each of the two new columns (Pivot Field Names, Pivot Field Values) and provide a more appropriate column name

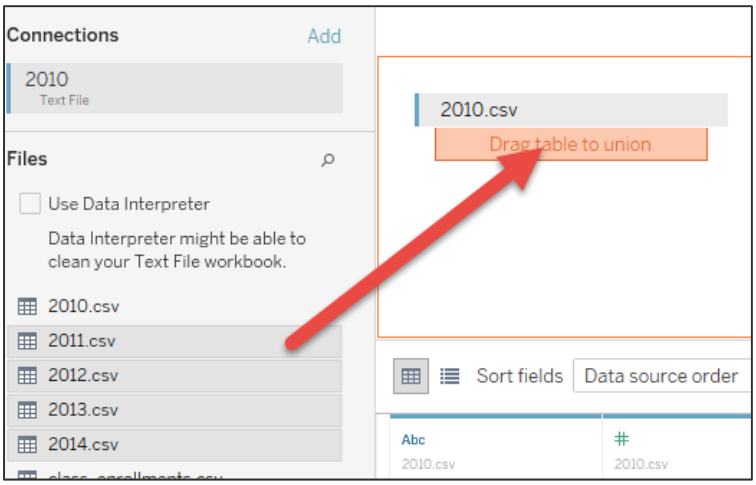


Union tables

1. Follow steps 1-3 in Create a data source from csv file
2. Highlight the remaining tables to union by holding down the **Ctrl** key and clicking on the name



3. Drag the highlighted tables underneath the table on the canvas and release when you see the text “Drag table to union” appear



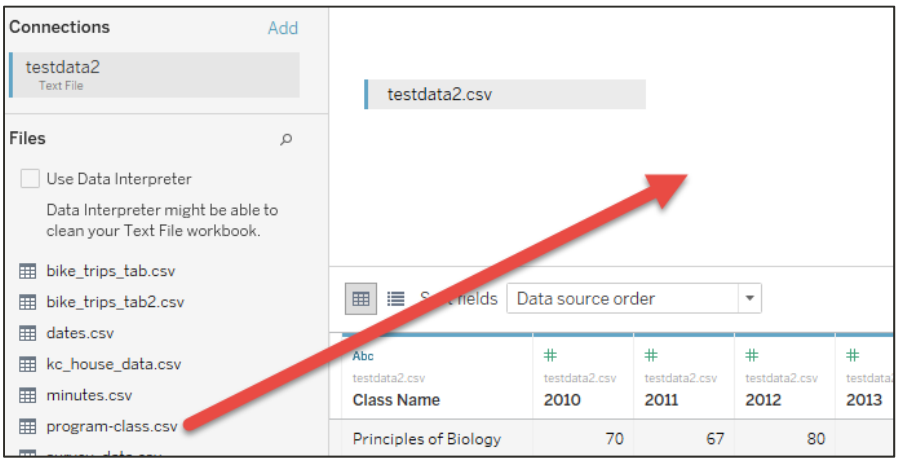
Exercise 3 – 20 MINUTES

Objectives

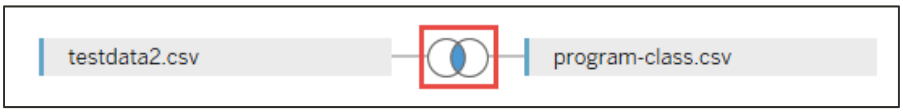
1. Make 3 copies of the invoice headers data source and name it the following:
   - header-lines
   - header-holds
   - header-lines-holds
2. For each for the 3 data sources, join in the appropriate tables. Note that the invoice holds table should be a left join to invoice headers.
3. Find and verify the granularity of each data source
4. Copy the bike trips data source and force a Cartesian product join against the datetime table. Create a calculated field to filter the entire data source so that you can answer the question: how many bikes are in-use at every minute of the day?

Joining tables

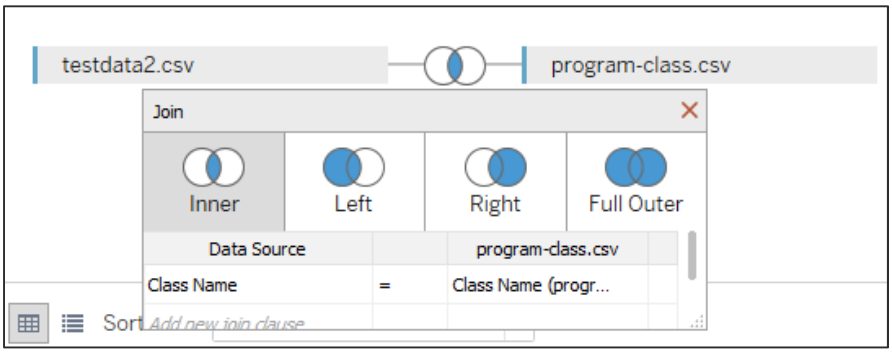
1. In the edit data source screen, drag the table to join from the left-hand pane into a blank part of the canvas area on the right



2. Tableau will try to default a join base on the similar column names. To verify or make changes to the join, click on the **Venn diagram** between the two tables to open the join configuration pop-up box

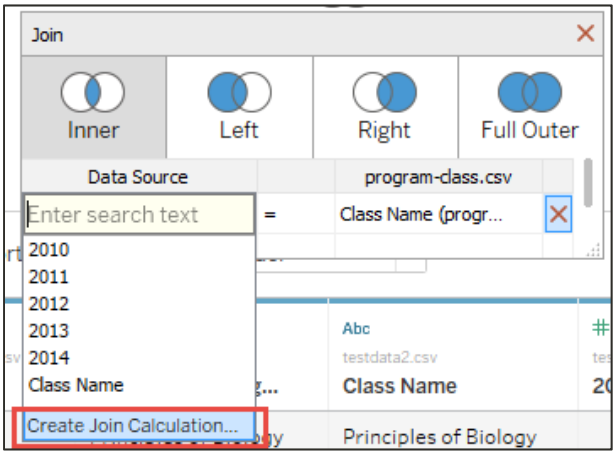


3. Select the join type (Inner, Left, etc.) and select the name of columns to join in each table in the two fields below

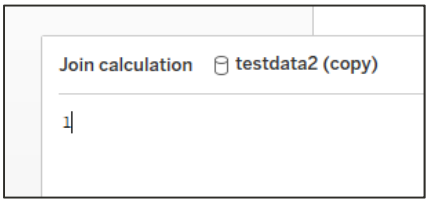


Forcing a Cartesian product

1. In the join configure pop-up box, select **Create Join Calculation** in the column name drop-down list



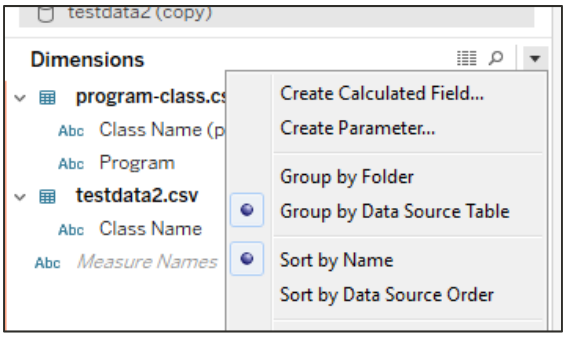
2. In the Join calculation pop-up window, erase the contents and enter the value **1** (this can be any value you choose). Click the **OK** button.



3. Repeat steps 1-2 for the right side column. Make sure you enter the same value in the Join calculation window.

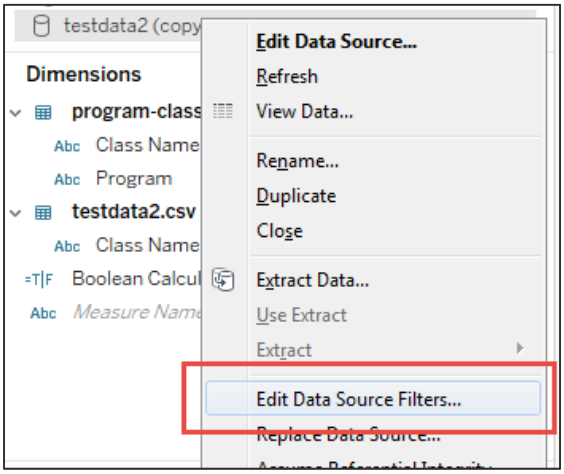
Creating a data source filter

1. In a Tableau worksheet screen, click on the drop-down arrow in the upper right corner of the **Dimensions** pane and select **Create Calculated Field** in the menu



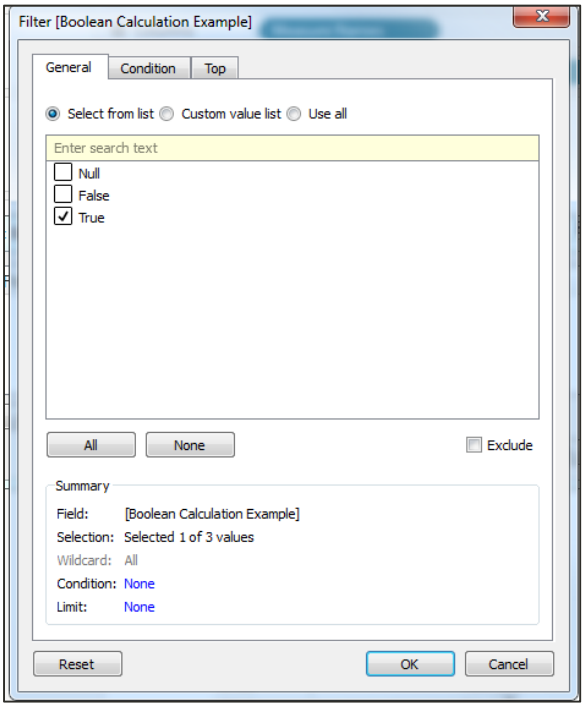
2. Type a new name for the calculated field. In the box below, you can either drag columns into it or start typing the name of the column and it will start to auto populate
3. The **equal =, greater than >, and less than <** signs are comparison operators that can be used to compare the values of either two different columns or a column and a static value
4. **AND** / **OR** keywords can be used to string multiple comparison making it more / less restrictive, respectively

5. After the calculated field is created, **right click** on the data source you wish to filter and select Edit Data Source Filters in the menu



6. Click the **Add** button and double click the calculated field in the list

7. Select the value(s) to filter for and click the **OK** button at the bottom



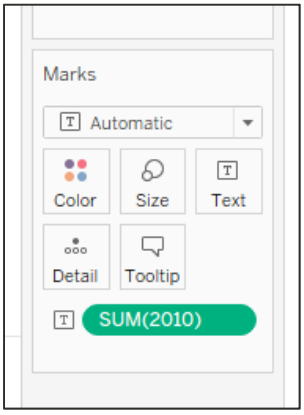
Exercise 4 – 20 MINUTES

Objectives

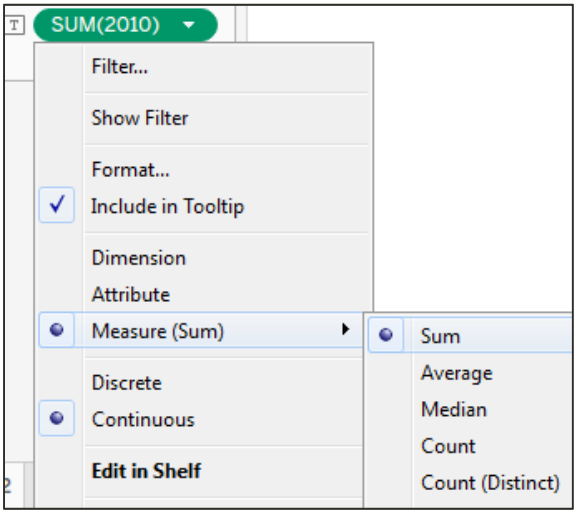
1. Bring a measure (or dimension) from any data source and switch among all the aggregation functions
2. Calculate the average invoice amount using the invoice headers data source
3. Derive that same average amount using the header-lines data source by creating an LoD calculation
4. In the header-holds data source, create a mutually exclusive Boolean LoD calculation to distinguish between all invoices with a '**MAX SHIP AMOUNT'** hold code and invoices that do not have this code. Keep in mind that an invoice can have multiple hold codes.

Switching aggregations on a measure

1. In a Tableau worksheet, drag a measure to the **Text** shelf. The measure should appear as a pill in the **Marks** card

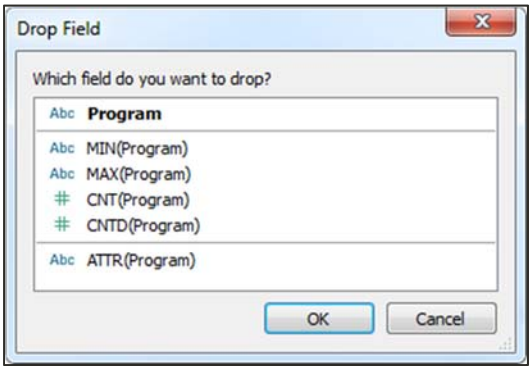


2. Mouse over this pill and click the drop down arrow on the right. Mouse over to **Measure (Sum)** in the menu and select another aggregation function in the sub-menu.



Switching aggregations on a dimension

1. The above steps also works for dimensions. However, since a dimension can have many values, dragging it onto the Text shelf in the worksheet will cause Tableau to render every distinct value in the visualization. This can take a while if the data set is large.
2. Alternatively, instead of left-click dragging, you can **right-click** drag the dimension to the **Text** shelf
3. Tableau will ask you to select an option first before trying to render the visualization



4. Select an aggregation function from the window and click **OK**

Creating a LoD calculation

1. Follows steps 1 and 2 in Exercise 2 – Creating a data source filter
2. The LoD calculation needs to follow the syntax below