



## Creating a Box Plot Analog

Box plots are very useful for comparing sets of data and find outlying data. Tableau can create box plots quickly and across several different types of data. The following steps will help you make a useful box plot

### Getting Started

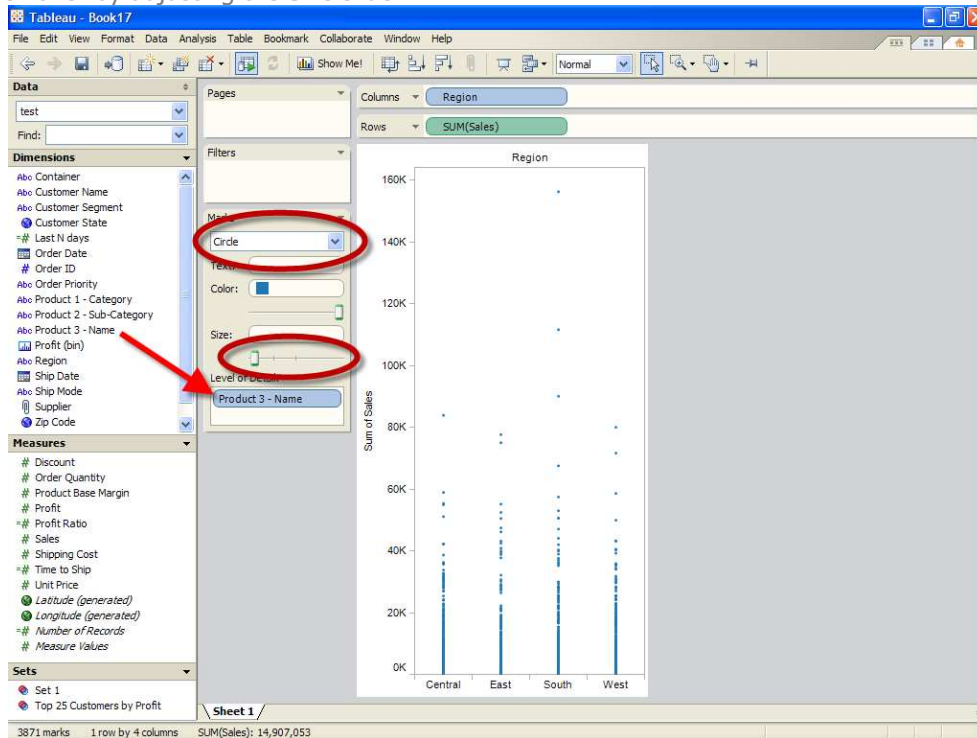
First, simply create a worksheet as you would normally in Tableau. In our example, connect to the Superstore example. Then put Region on the Columns shelf and Sales on the Rows shelf.

Now that we have the standard chart, we need to graph the Median and Quartiles. Tableau will calculate the Median and Quartiles lines for you based on the marks that are displayed in the pane. To create a box plot for the sales in each region, you will need to change the display all the sales values for the region, not the summarized sales value.

There are 2 options for this, first simply show all the data by unchecking 'Aggregate Measures' from the 'Analysis' menu. This may have unintended side effect of separating the data into too many variables. You may not want the box plot of all sales. Instead, you want a box plot per Customer.

The second option is to put a field into the Level of Detail shelf to calculate sales. In our example, put 'Customer Name' in the Level of Detail shelf. This shows a mark for each customer's sales in that region.

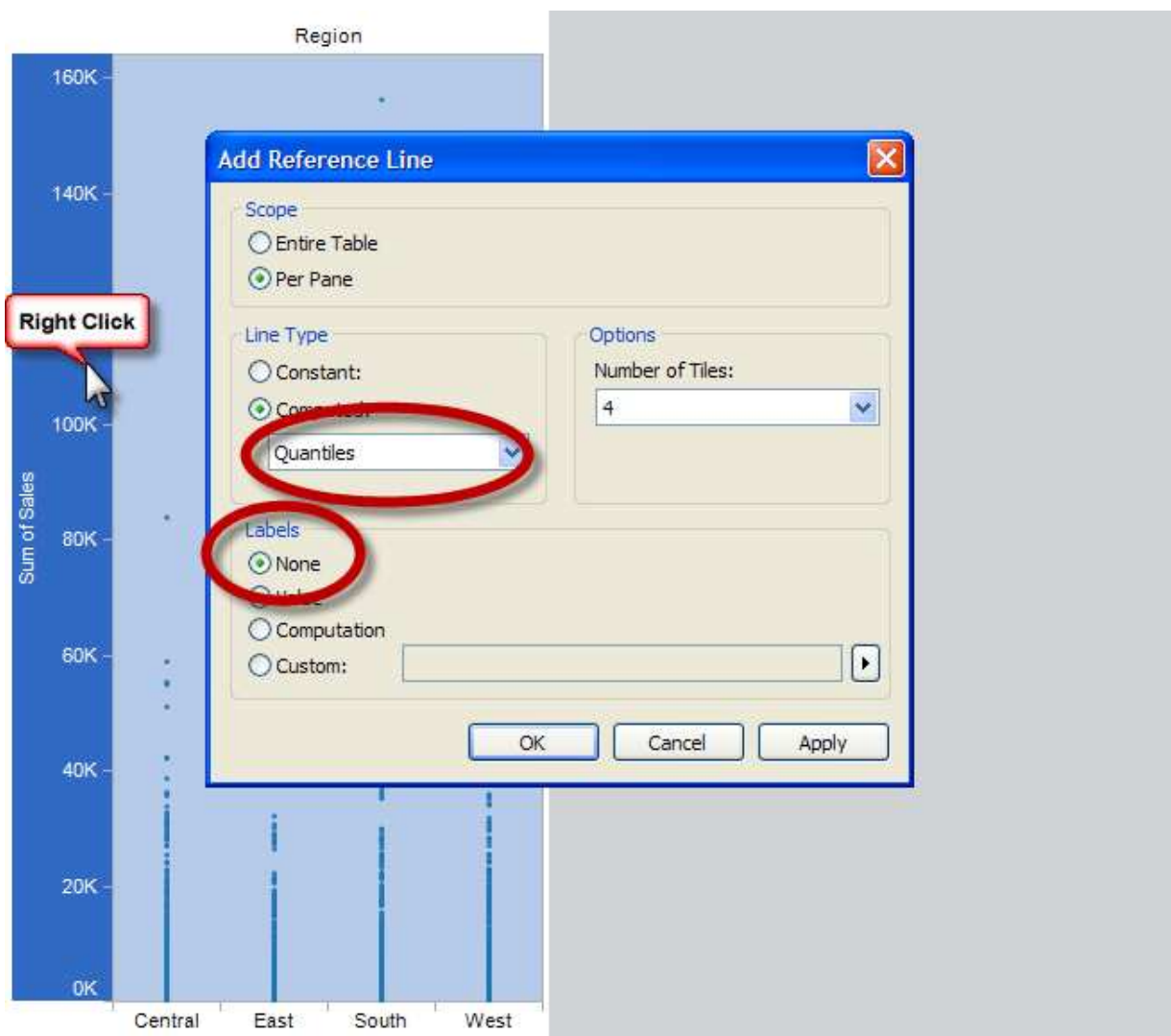
To clean up the view a bit, change the mark type from **Automatic** to **Circle** and make the circles smaller by adjusting the Size slider.



## Adding the box

Now that you have the desired marks, you can create the Median and Quantiles. These are added with Reference lines.

1. Right click on the 'Sum of Sales' axis and choose **Add Reference Line**.
2. Select **Quantiles** > 4 Tiles
3. Change the label to **None**

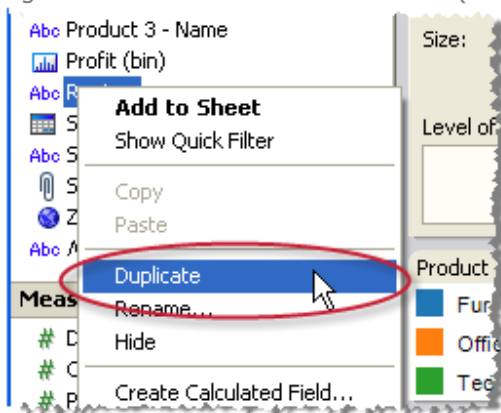


## Adjusting the Scope of the Box

Viewing all the regions together is a nice start – and often times a useful view. For example, this is good for indicating which region has the most outliers.

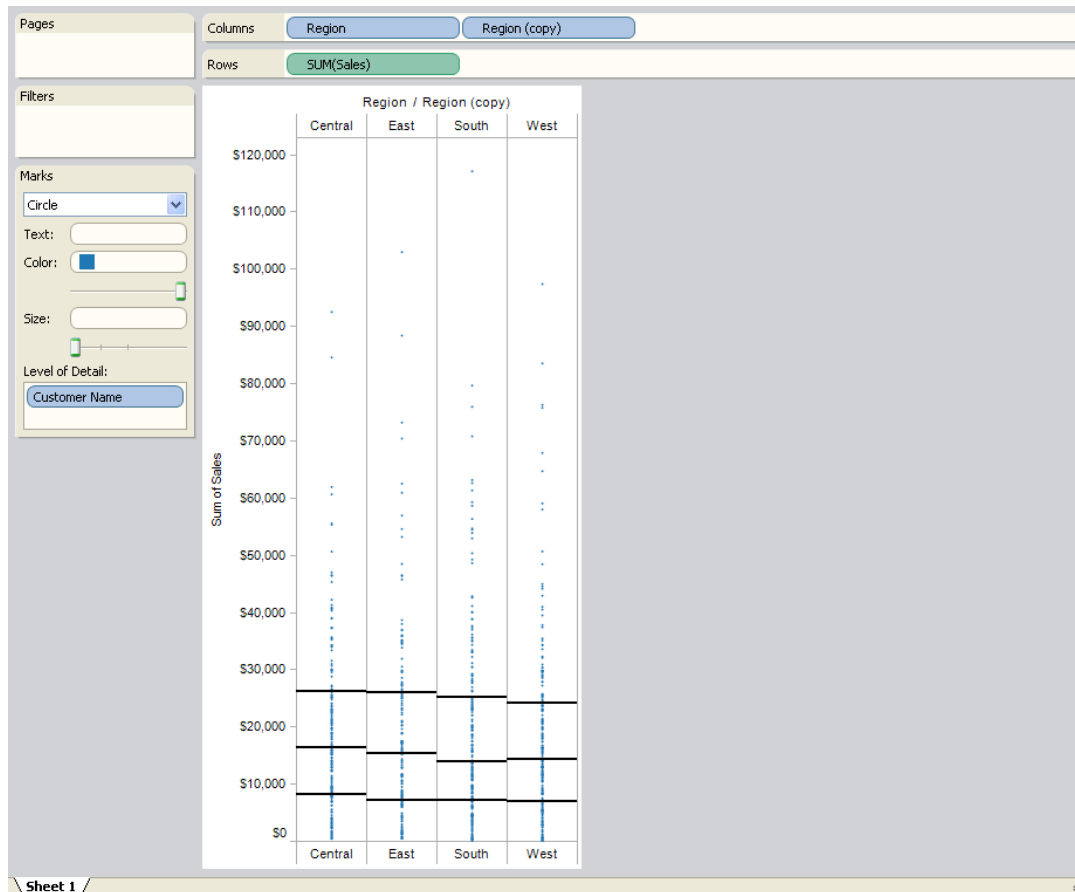
If we want a box for each region, we have one more step – we need to create more ‘panes’ so that the reference lines are calculated on a smaller set of data, not the entire view.

1. Right-click the field in the Data window (in this case Region) and select Duplicate.



2. Drag the duplicated field to the Columns shelf





Now you can add other fields to the view to see what products are impacting sales outliers, or select a cluster of customers and create a set so you can further analyze them.