|  |  |
| --- | --- |
|  |  |

|  |
| --- |
| Box and Whisker Features Document |
|  |

# Short Description

This visual displays data distribution into quartiles in a box with a thin line going through it (whiskers) indicating variability outside the upper and lower quartiles.

# Overview

Box and whisker charts are most commonly used in statistical analysis like comparing medical trial results or teachers' test scores et. al. It consists of two parts—the main body called the Box and the thin vertical lines coming out of the Box called Whiskers. The first quartile forms the bottom and the third quartile forms the top of the Box. The Whiskers connect the minimum and the maximum values to the Box. In addition to showing median, first and third quartile and maximum and minimum values, the Box and Whisker chart is also used to depict Mean, Standard Deviation and Quartile Deviation.

While other Power BI Box plots have the capability to display these statistics data, Box Plot by MAQ Software provides more details and clarity of the figures by providing more user controls and capabilities to suit the exact needs of the use case like ability to add parent axis and flip visual horizontally or vertically based on reporting requirement.

Other common features include:

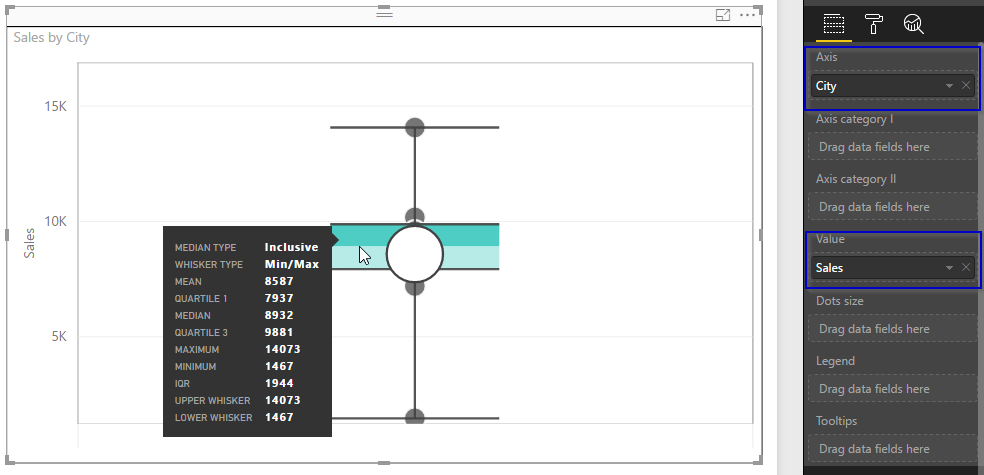
* Option to enable/disable - mean dot and median line
* Option to choose the type of quartile - inclusive/exclusive
* Option to choose type of whiskers - Min/Max, < 1.5 IQR, > 1.5 IQR or custom
* Option to choose shapes of mean and data points – circle, square, rectangle
* Option to hide/show all data points
* Option to hide/show outliers
* Option to control the opacity of boxes

Thus, this visual lets you visualize the complete picture of your statistical data comparison, in a glance.

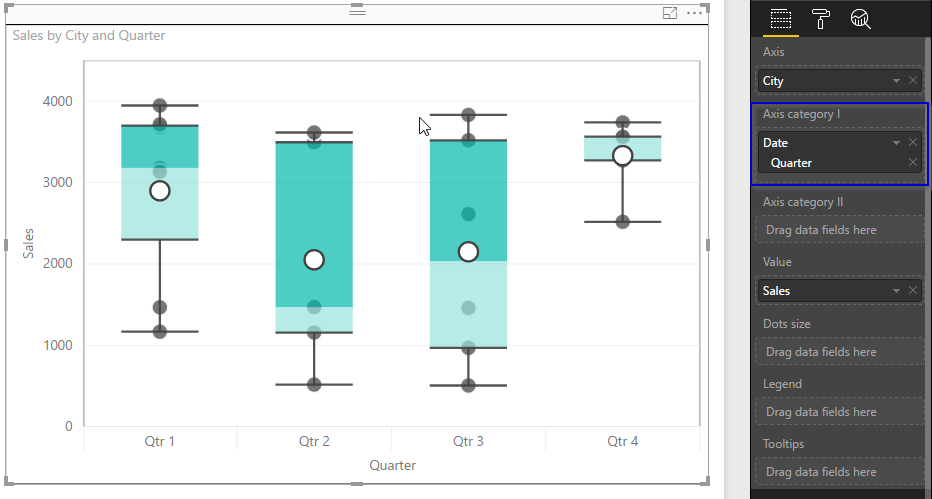
# Example

Assume that the CEO of a company wants to know total sales of all the cities of his company.

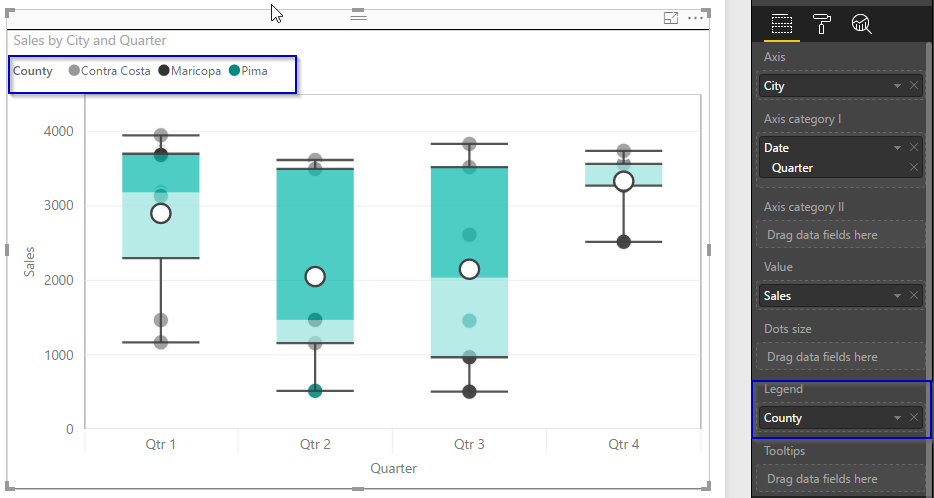
So, he will just enter ‘City’ in Axis field and ‘Sales’ in the value field and will get dots representing all the cities with their respective Total Sales. The box showing the quartiles, mean, median and whiskers, is plotted on top of the dots.



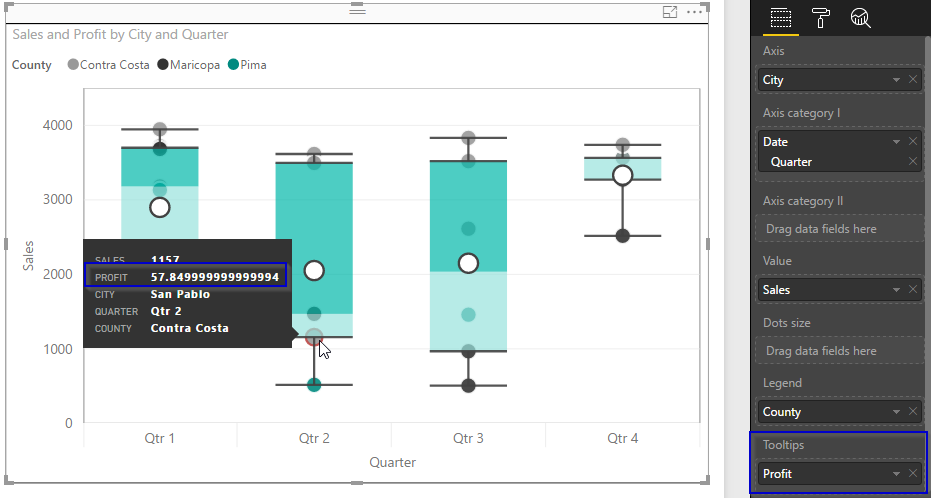
But from this, he cannot understand how much sales have been made in different quarters relative to each other. This can be found by dragging ‘Quarter’ in Axis Category I field.



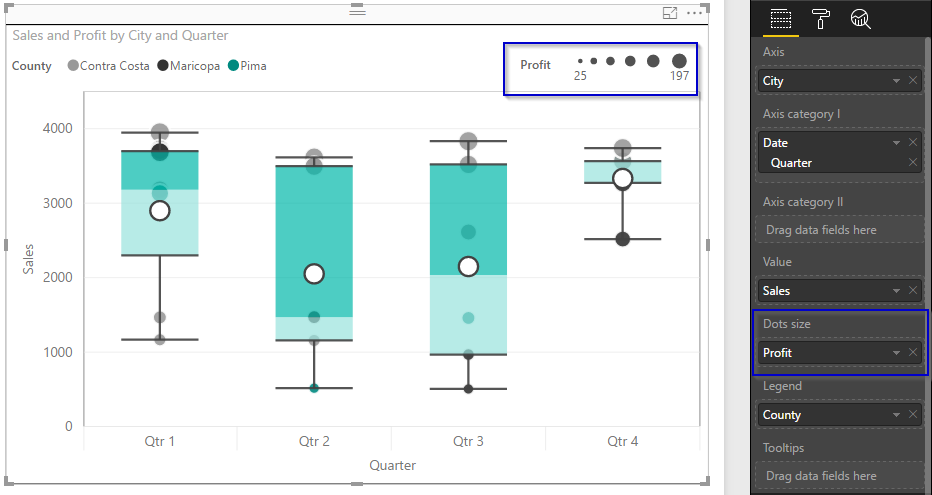
For even deeper analysis, he wants to know which counties have what sales. To know this, he selects ‘County’ in ‘Legend’ field.



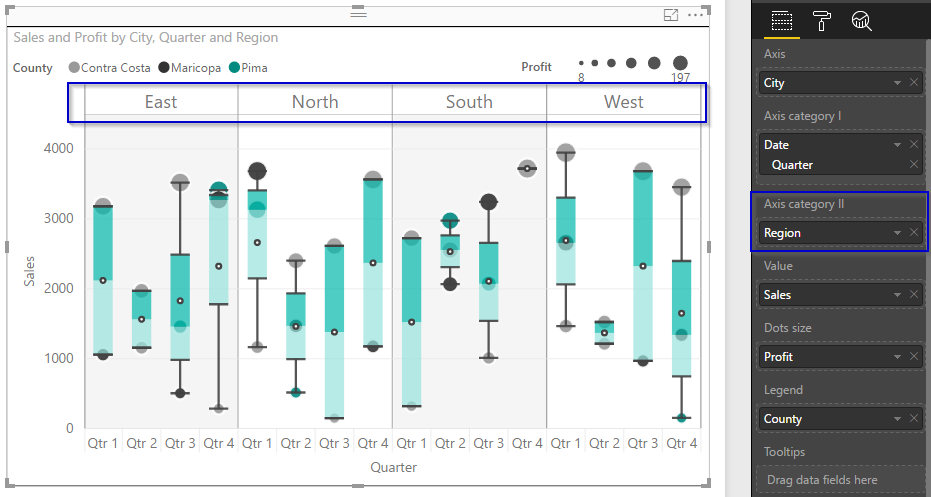
For analyzing every individual value which is not displayed from the graph, Tooltips can be added. Suppose now the CEO wants to know, what is the profit associated with every sales value, then ‘Profit’ can be added in tooltips.



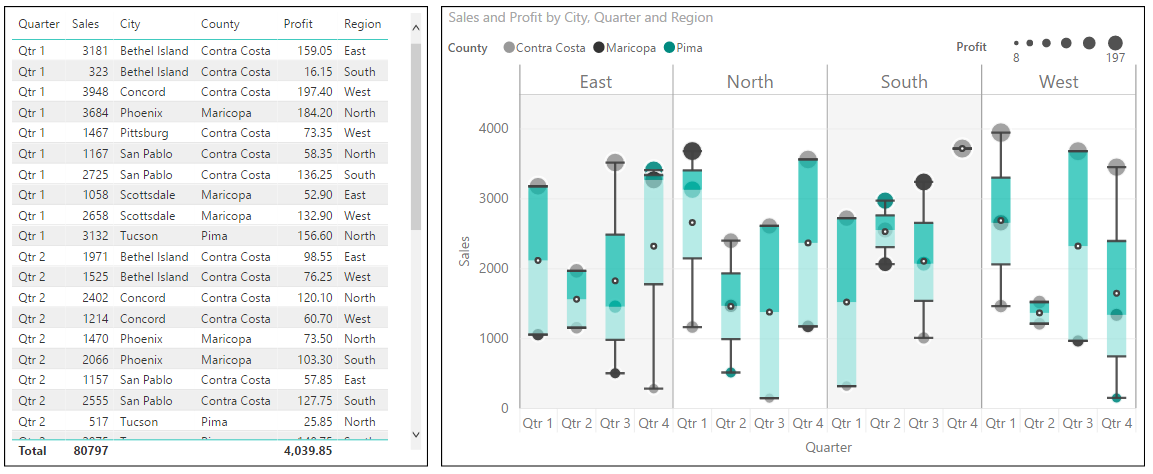
Also, if he wants to see Profit of sales associated with each city, then he can drag Profit value in Dots size field. Bigger the dot, higher is the Profit value.



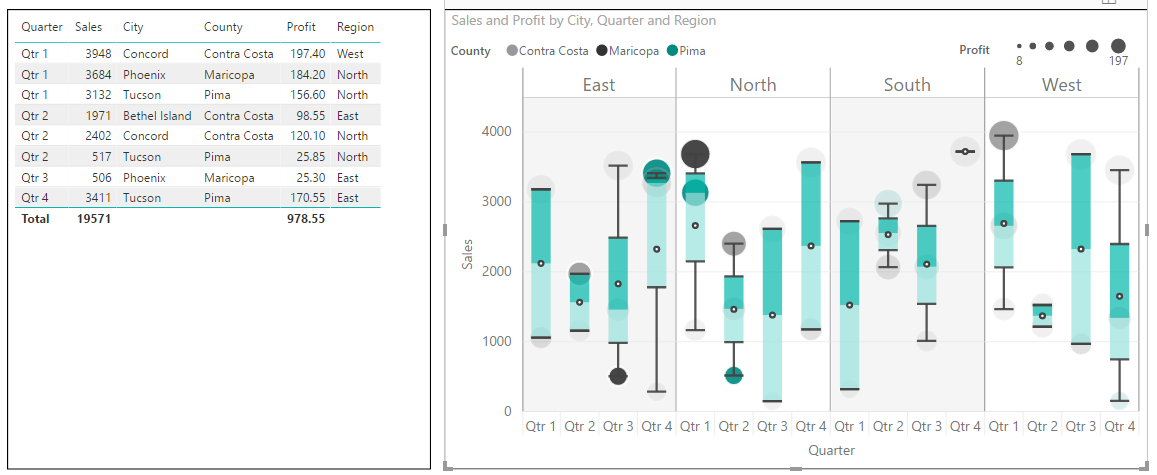
Now if in a single instant, data needs to classified based on Region and Quarter (another dimension or another insight), he can add Region in Axis Category II.



If a table is included in the report along with the box and whisker visual, the page will look like below:

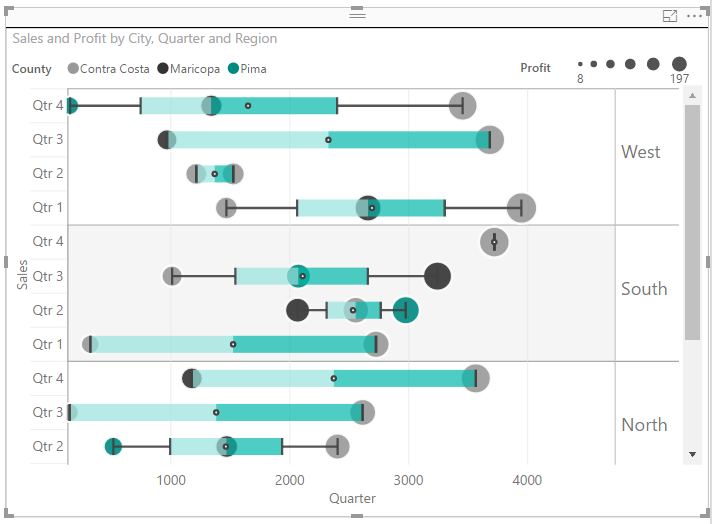
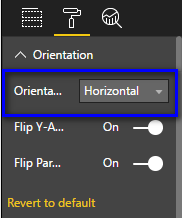


User can filter data either by clicking on bubble or on other visuals (in this case table visual)

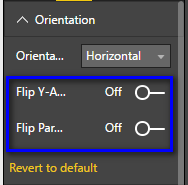


Now the visual can be formatted using different options in Format pane like:

1. Using orientation, the overall visual orientation can be changed from vertical to horizontal

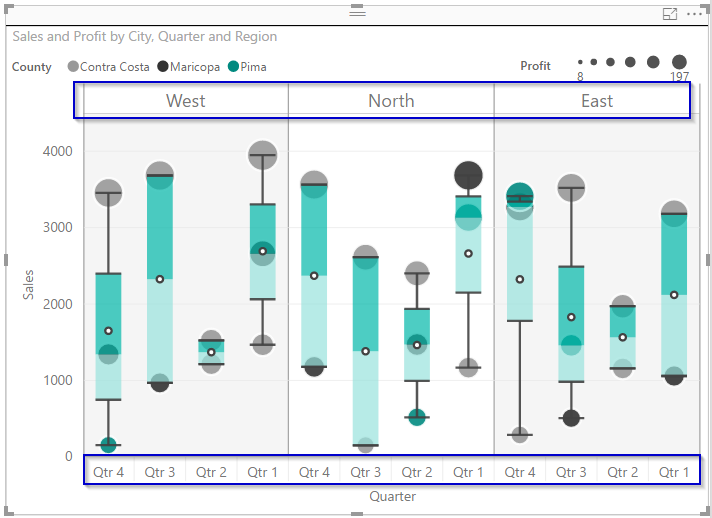
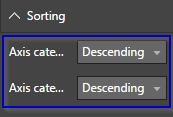


1. In ‘horizontal’ orientation, the orientation of Y-axis labels and Parent labels can be flipped.



1. Sorting

The visual can be sorted based on 'Axis category I' and 'Axis category II' fields, in ascending or descending order.



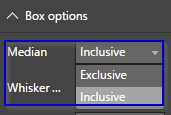
1. Box options:
   1. Median
      1. Inclusive

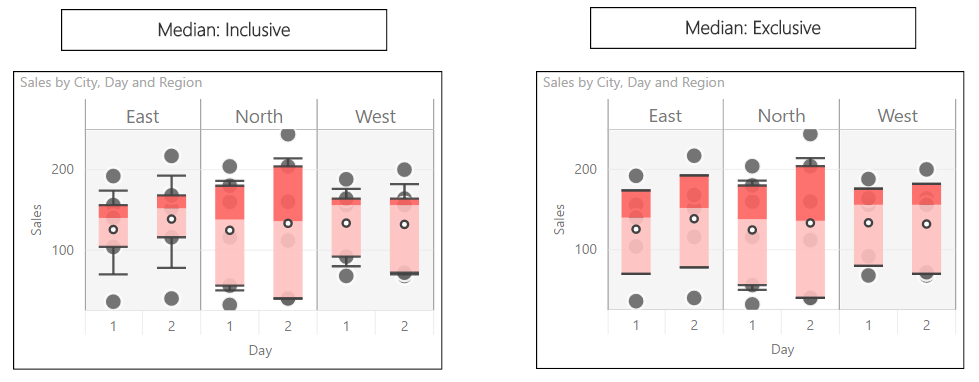
In this, while calculating the quartile values Q1 and Q3, the median is also included.

* + 1. Exclusive

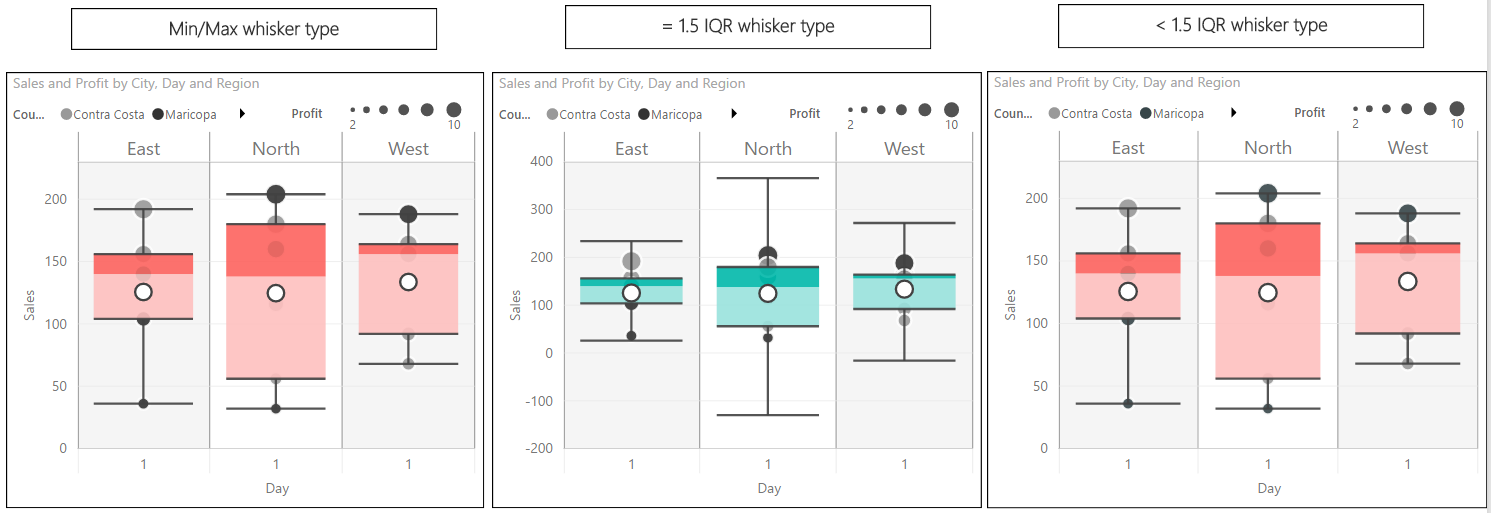
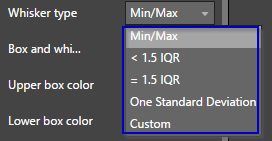
In this, while calculating the quartile values Q1 and Q3, the median is excluded.

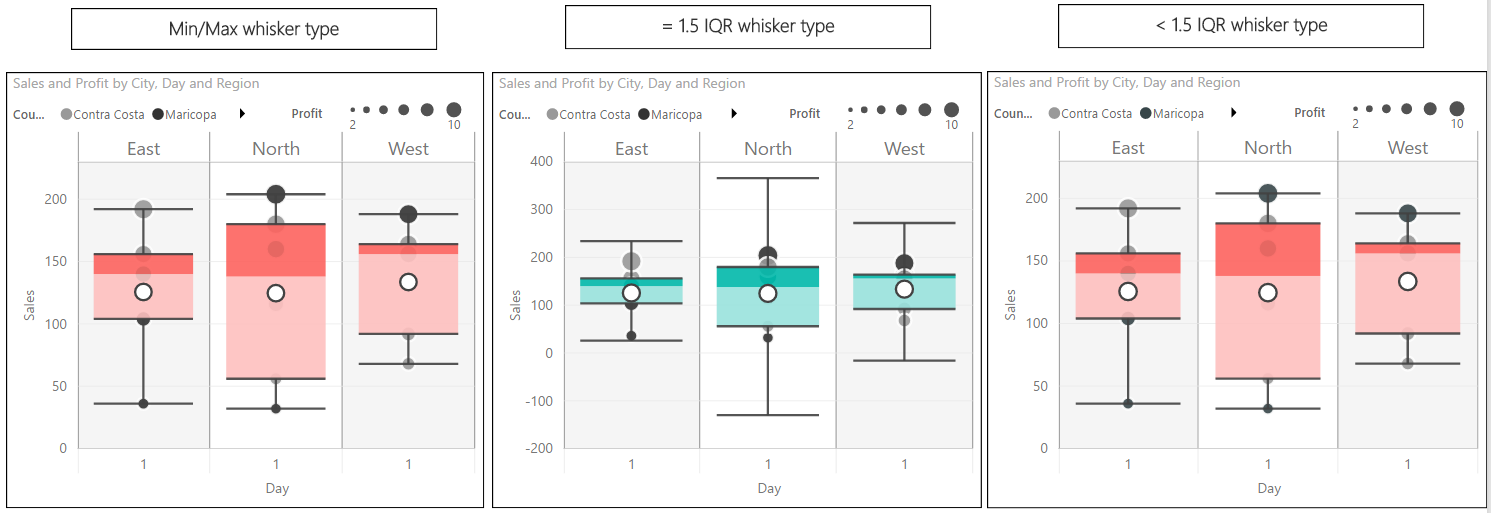
Both these options will give the same quartile values if the number of data points is even, but a different result for an odd number of data points since the median value will be included in both the upper and the lower half of the dataset.

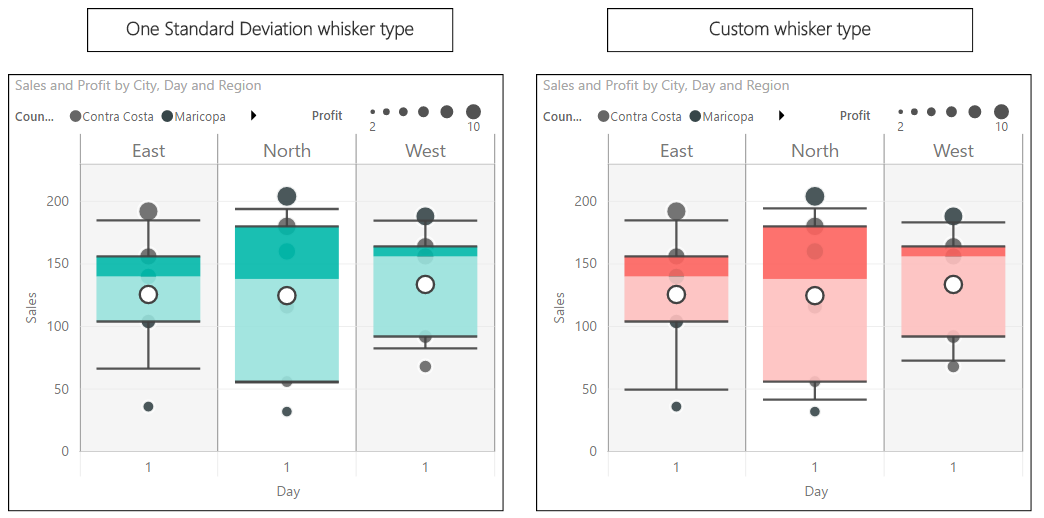




* 1. Whisker type
     1. Min/Max
     2. < 1.5 IQR
     3. = 1.5 IQR
     4. One Standard Deviation
     5. Custom

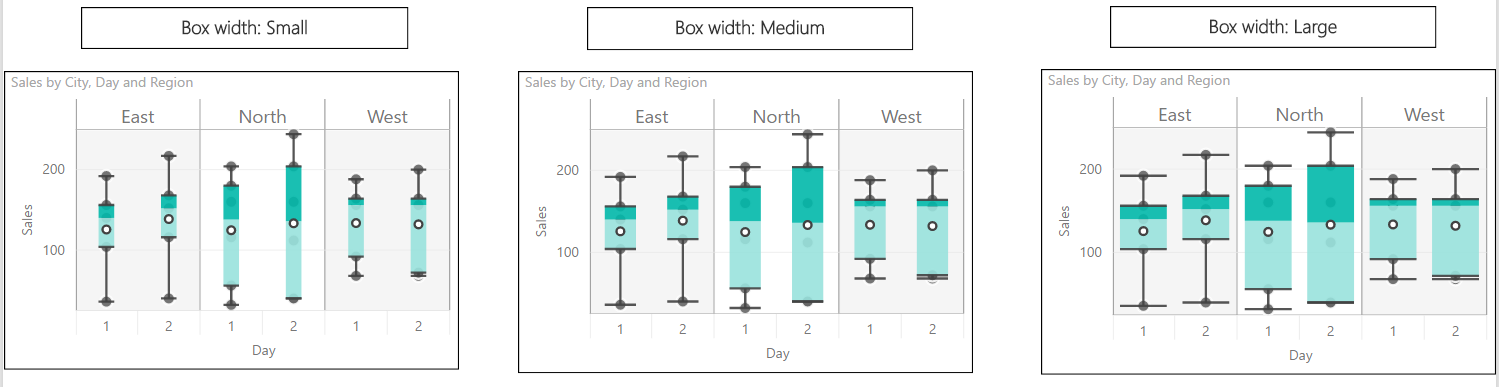




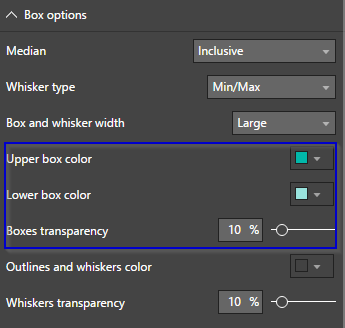


* 1. Box and Whisker width

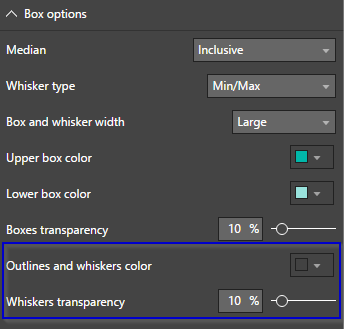




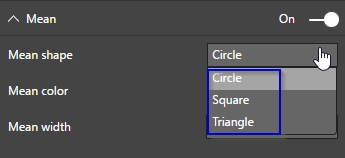
* 1. Box colors and transparency

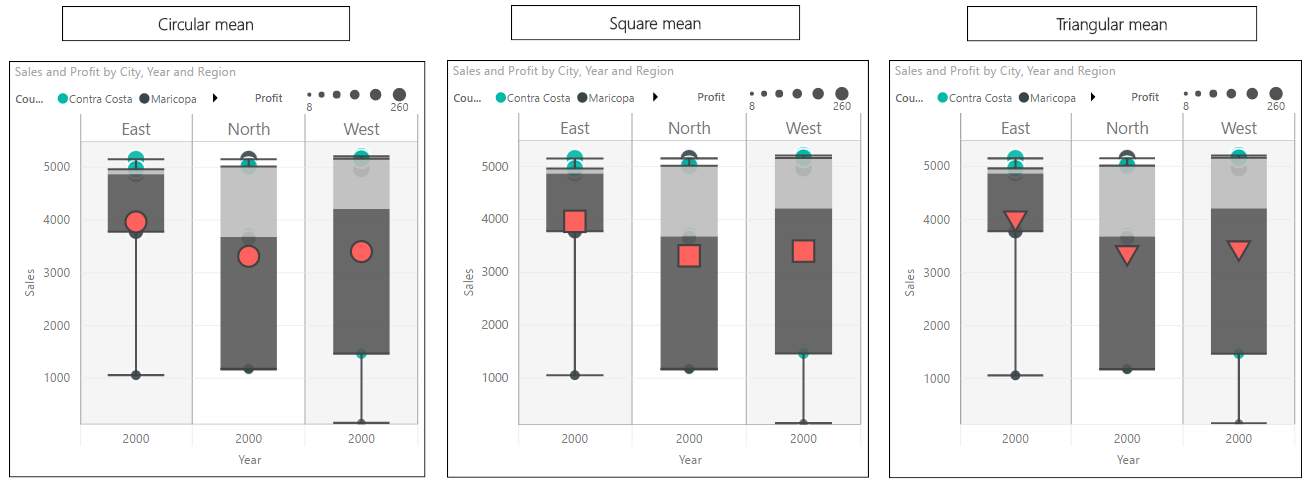


* 1. Outlines and whiskers color and transparency

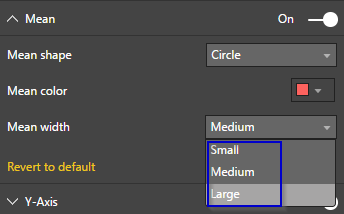


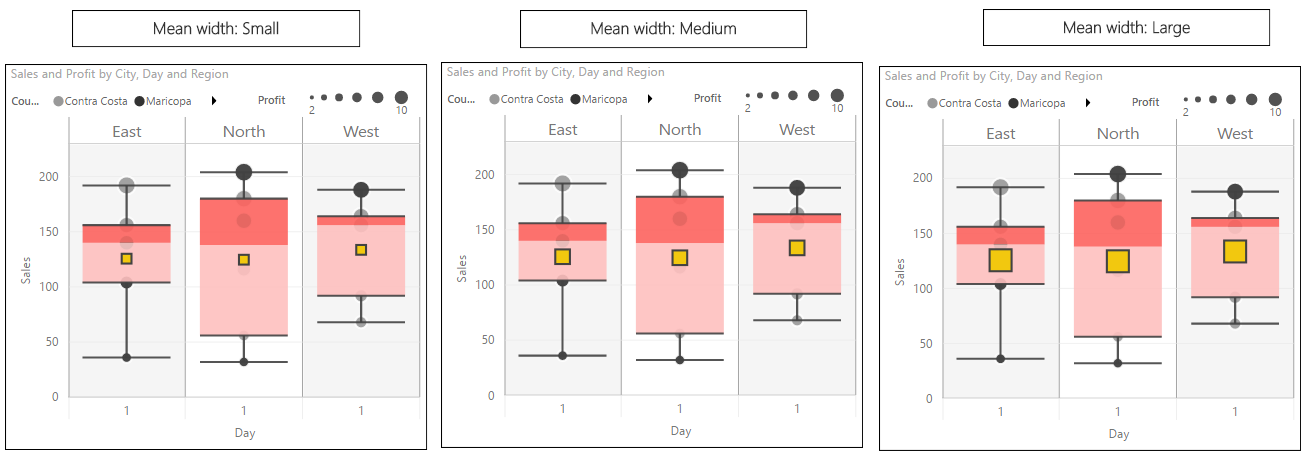
1. Formatting options for ‘Mean’:
   1. Mean shape



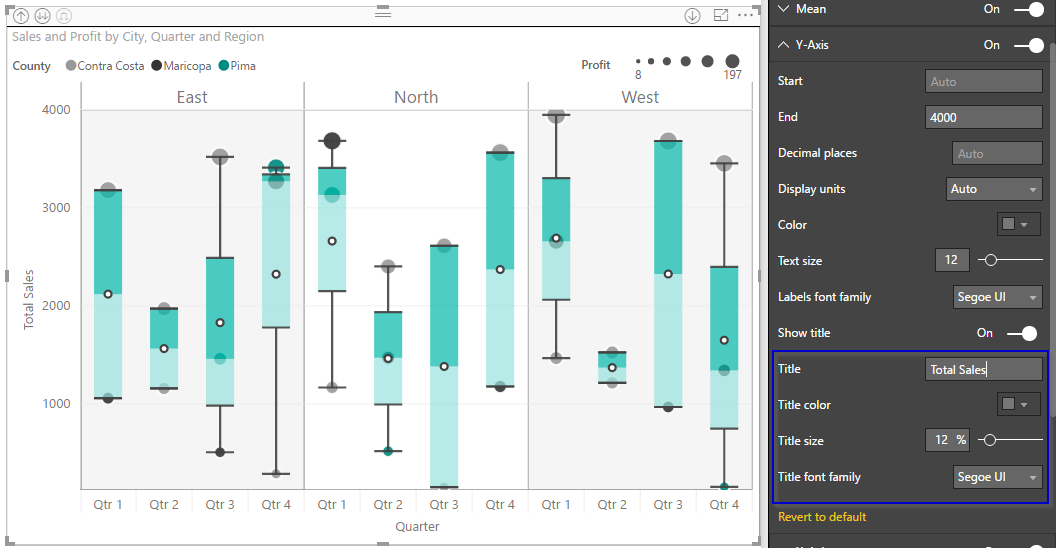


* 1. Mean color and width

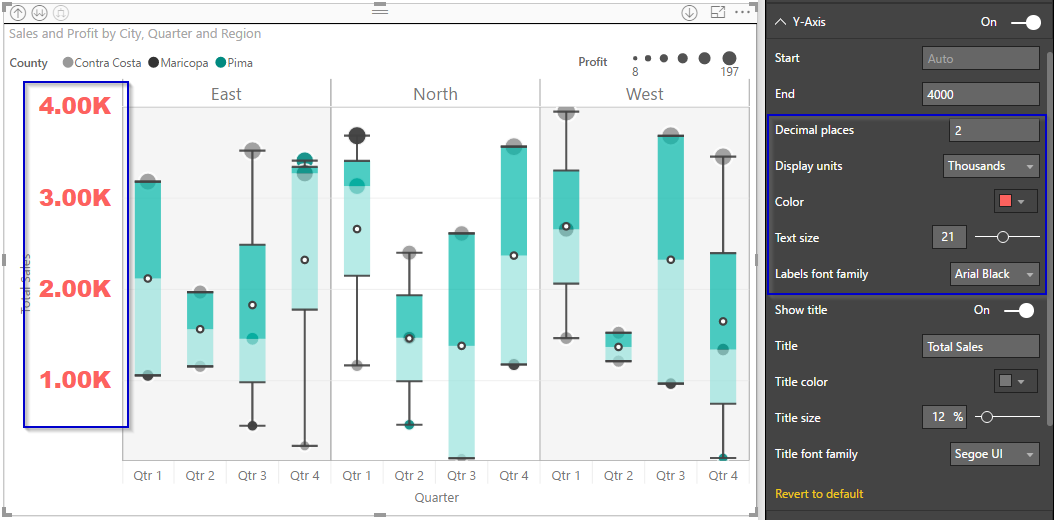




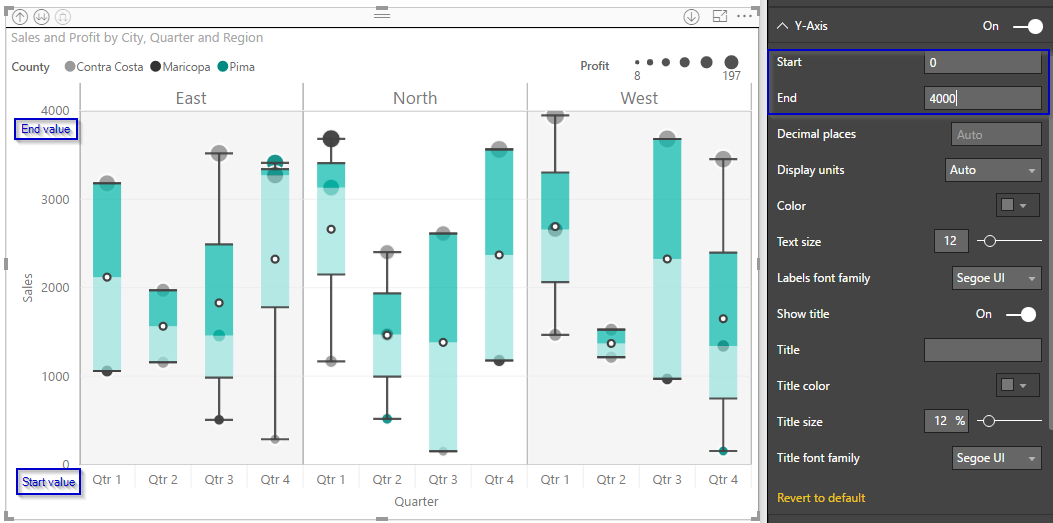
1. Formatting option for Y-Axis:
   1. The title, title text, title color, title font size and font family can be updated.



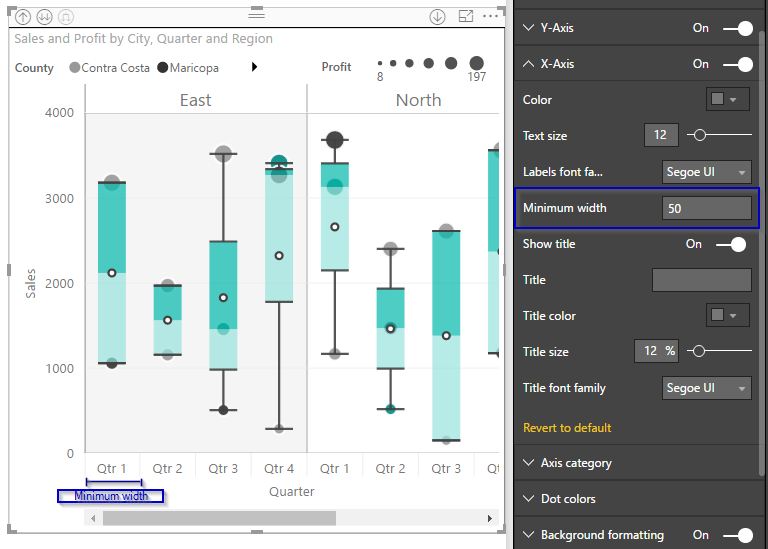
* 1. Display units, decimal points, color and font size of labels can be updated.



* 1. The Y-axis start and end values can be changed, so no magnitude comparison can be done, without having to start the y-axis from 0.



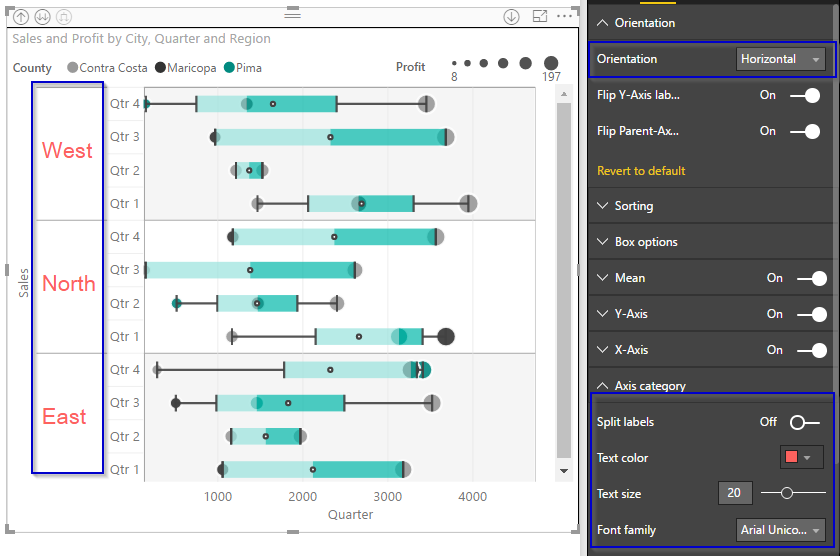
1. Similar options are available for X-axis. Also, the minimum width of each X-axis label can be configured. After updating the width if the labels cannot be accommodated, a horizontal scrollbar appears.



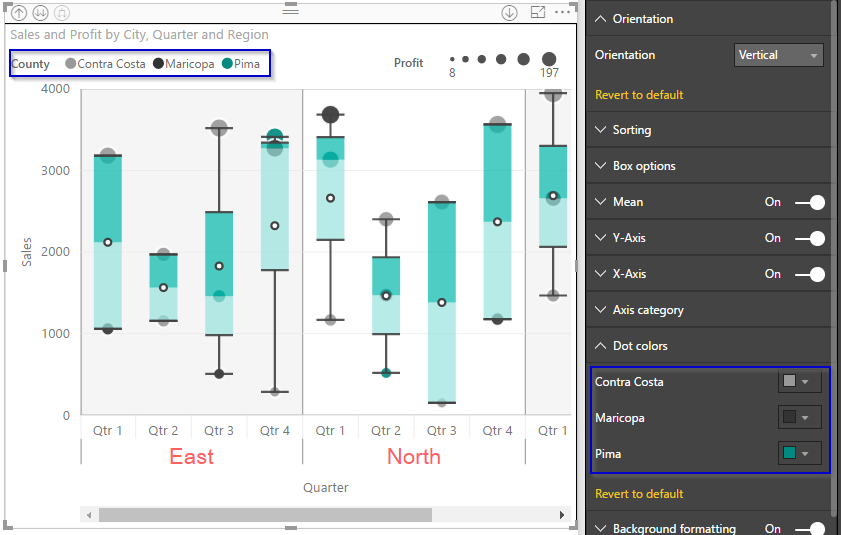
1. Using ‘Axis category’ settings, Axis category II labels can be flipped from top to bottom. The color, font size and font-family also can be updated.



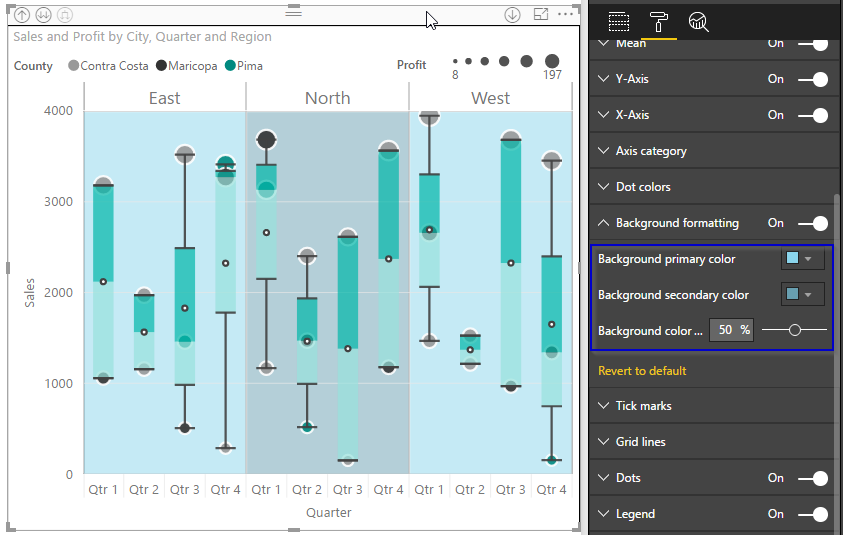
* 1. In case of ‘Horizontal’ orientation, Axis category II labels can be flipped from right to left.



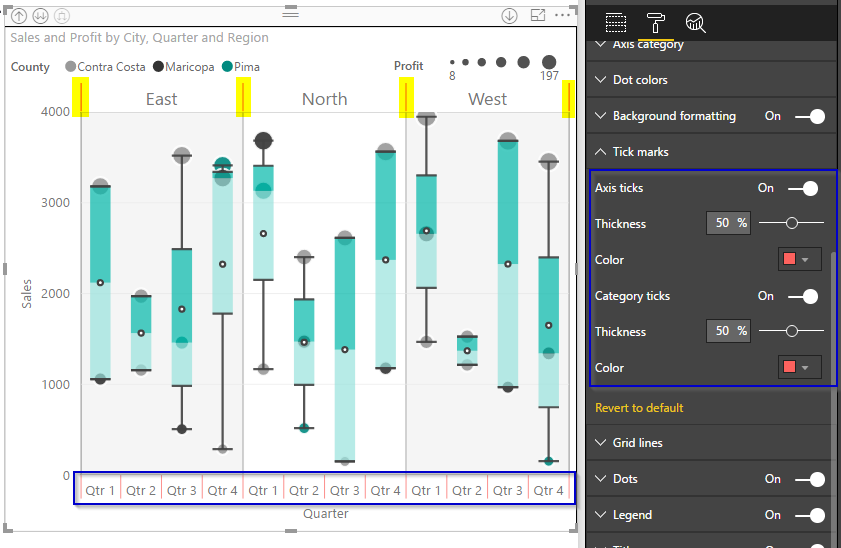
1. Dot Colors
   1. When legend is present the color of the various legend items and hence dots, can be modified using ‘Dot Colors’

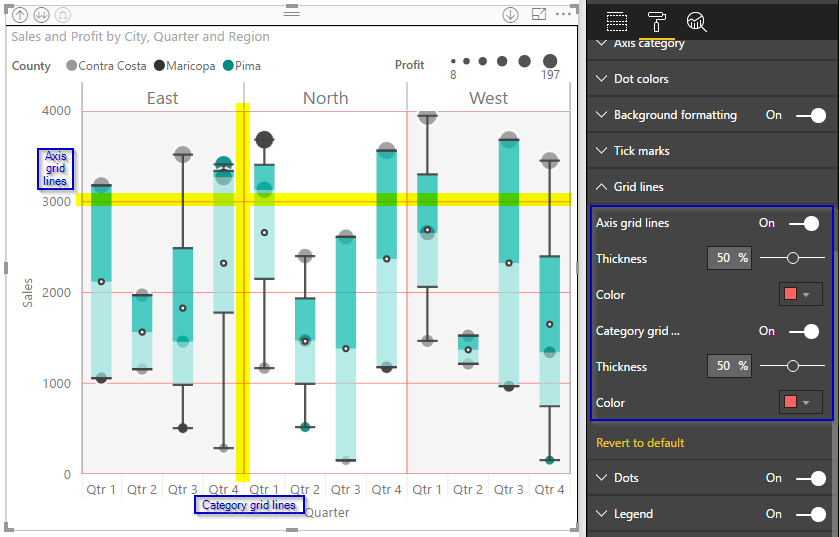


1. The primary and secondary colors of the graph can be modified along with the transparency.

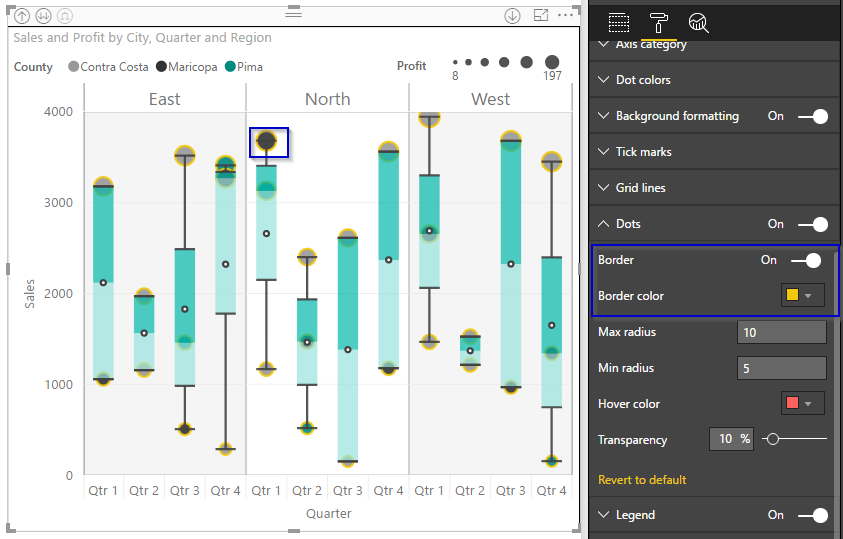


1. The axis and category tick marks can be shown or hidden. The color, thickness of axis and category tick marks can be updated. Similar options are available for grid lines as well.

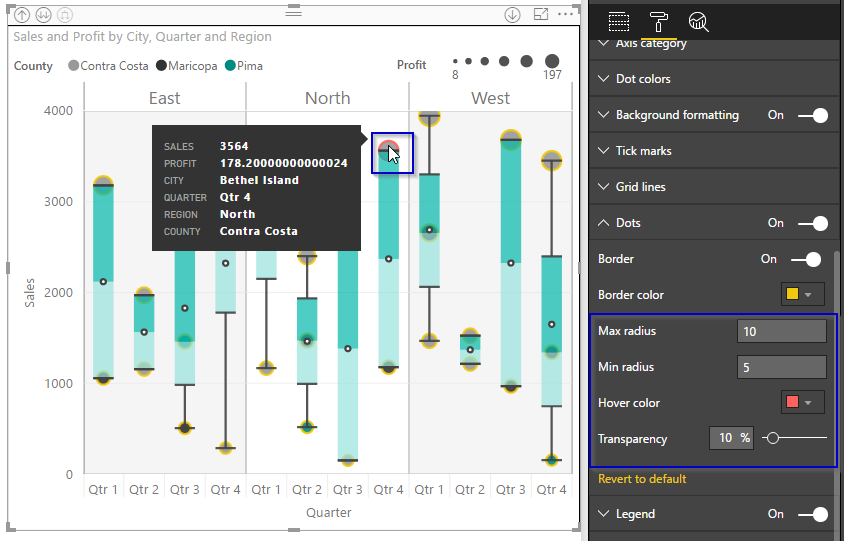




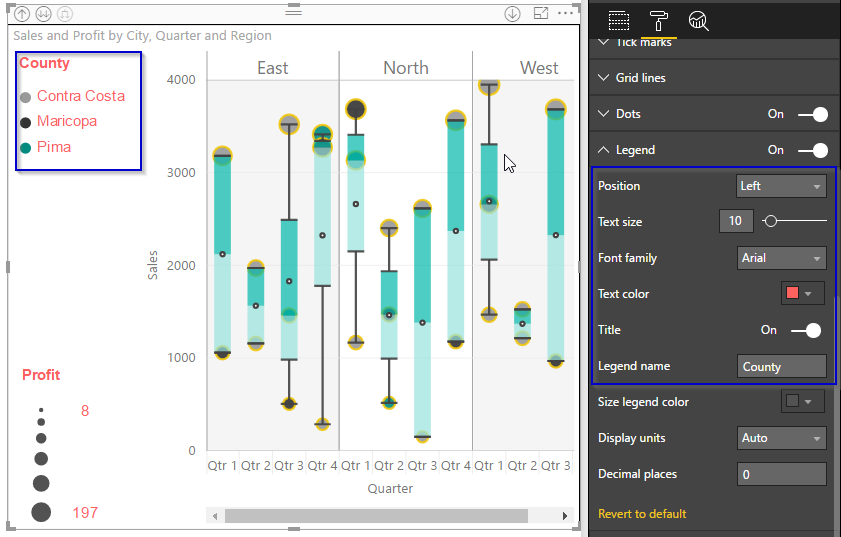
1. ‘Dots’ settings
   1. The border for dots can be enabled/disabled and its color can also be changed.



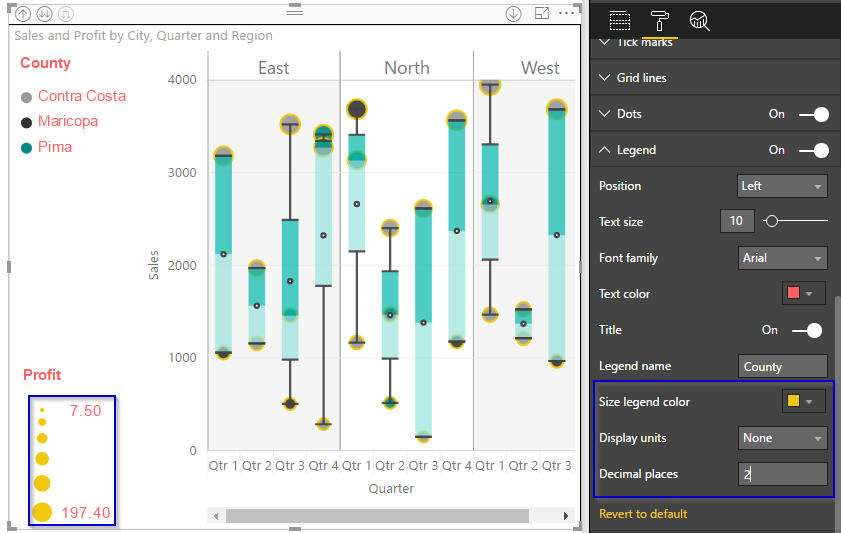
* 1. The radius range, hover color and transparency of the dots can also be updated.



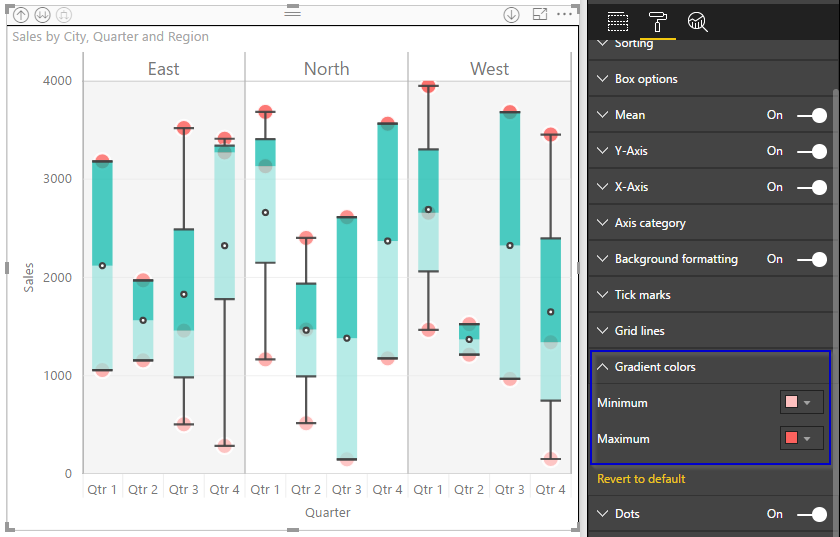
1. ‘Legend’ settings
   1. Legend positions, font color, font size, font-family and title text can be updated.



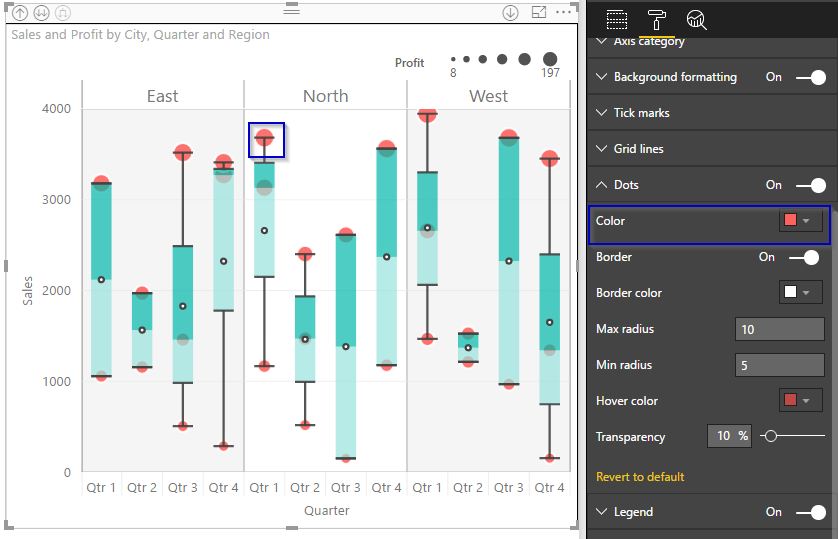
* 1. The color, display units and decimal places of size legends can be updated.



1. If a measure is selected instead of a category in ‘Legend’ field, legend disappears and Gradient colors option will appear. The color of each dot will be a different gradient ranging from minimum to maximum color provided in the ‘Gradient colors’ option.



1. If the legend is absent, ‘Dot Colors’ option disappears and the colors of the dots can be set from ‘Dots’ formatting option.



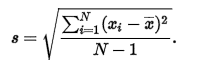
# Calculations used

1. **Order** data in numerical order before calculation
2. **Median** = (n+1)/2th element; //if n is odd

**Median** = ((n/2th element) + (n/2 + 1th element)) / 2; //if n is even

1. Inter Quartile Range(IQR) = Q3 - Q1
2. Find median of whole data (**Q2**), divides data into two halves
3. Find medians of these two halves, (**Q1<Q3**), divides the data into quarters
4. **Whisker Choices**:
   1. Min/Max – Whiskers on minimum and maximum value
   2. = 1.5 IQR – Whiskers on Q3 + 1.5 IQR and Q1 – 1.5 IQR
   3. < 1.5 IQR – Whisker on dots <= Q3 + 1.5 IQR and >= Q1 – 1.5 IQR
   4. One Standard Deviation
      1. Upper whisker – Max(Mean + sigma, Q3)
      2. Lower whisker – Min(Mean – sigma, Q1)

Sigma,



* 1. Custom percentile calculation, example below for Pth percentile:
     1. The first step is to compute the rank (R) of the Pth percentile. This is done using the following formula:

R = P/100 x (N + 1)

where P is the desired percentile and N is the number of values.

* + 1. If R is an integer, the Pth percentile is the number with rank R. When R is not an integer, we compute the Pth percentile by interpolation as follows:

1. Define IR as the integer portion of R (the number to the left of the decimal point).
2. Define FR as the fractional portion of R.
3. Find the scores with Rank IR and with Rank IR + 1.
4. Interpolate by multiplying the difference between the scores by FR and add the result to the lower score.