MASTER OF COMPUTER APPLICATIONS

SEMESTER 1

DATA VISUALIZATION

Unit 5

Statistical Charts in Excel

Table of Contents

1. INTRODUCTION

The upcoming chapter is a comprehensive exploration of three indispensable data visualization tools: The Box and Whisker Plot, the Waterfall Chart, and the Area Chart. Each of these visualization techniques offers unique insights into diverse aspects of data analysis. The Box and Whisker Plot allows us to understand data distributions, detect outliers, and compare datasets effectively. The Waterfall Chart is a powerful tool for tracking financial data, showcasing incremental changes, and dissecting the cumulative impact of various factors. Lastly, the Area Chart provides an excellent means to visualize cumulative trends over time, making it invaluable for time-series data analysis. By delving into the intricacies of these visualization methods, you will gain a deeper understanding of how to present and interpret data effectively in a wide range of analytical scenarios.

1.1 Learning Objectives

By the end of this chapter you will be able to:

- *Describe the fundamental principles and components of Box and Whisker Plots, Waterfall Charts, and Area Charts, including how they represent data and when to use each type of visualization.*
- *Interpret and analyze Box and Whisker Plots, Waterfall Charts, and Area Charts to identify data distributions, trends, and patterns, and explain their significance in various real-world scenarios.*
- *Demonstrate the ability to create, customize, and apply Box and Whisker Plots, Waterfall Charts, and Area Charts using data visualization software or tools, effectively conveying information to different audiences.*
- *Evaluate the strengths and limitations of each visualization method in addressing specific data analysis tasks, such as outlier detection, trend identification, and cumulative data representation.*
- *Construct advanced data visualizations by combining elements of Box and Whisker Plots, Waterfall Charts, and Area Charts to solve complex analytical problems, and critically assess the appropriateness of the chosen visualization techniques for specific data sets and research objectives.*

2. BOX AND WHISKER PLOT

A box and whisker plot, also known simply as a box plot, is a method used to summarize data measured on an interval scale. This graphical representation is widely employed for data interpretation, particularly when seeking to understand data variation within a dataset. While histograms can be used to display data, box plots offer a distinct advantage by providing additional insights. Box plots excel in situations where multiple sets of data need to be presented within the same graph, making them a valuable tool for comparative analysis. Unlike histograms, which primarily focus on the distribution of a single dataset, box and whisker plots allow for the concurrent visualization of multiple datasets, making them an essential choice for data exploration and interpretation.

In essence, a box plot can be defined in terms of descriptive statistics concepts. It visually represents groups of numerical data through quartiles, often accompanied by lines extending from the boxes or whiskers to highlight variability beyond the lower and upper quartiles. This feature gives rise to the names "box-and-whisker plot" or "box-and-whisker diagram," with outliers typically depicted as individual data points. Box plots efficiently convey information about the degree of data variation or spread, complementing measures of central tendency. They are not only space-efficient but also offer a concise and insightful pictorial representation of data. By presenting the central tendency, spread, and overall range at a glance, box plots facilitate straightforward comparisons of distributions.

Box plots, or box and whisker plots, serve as valuable graphical representations for gaining insights into three critical aspects of data:

- distribution shape
- central value
- variability.

When constructing a box plot, a box is drawn to encompass the first quartile to the third quartile, with a vertical line inside the box representing the median. Additionally, small lines, known as whiskers, extend from each quartile towards the minimum and maximum values of the dataset. This visual presentation effectively provides a concise summary of the distribution's characteristics.

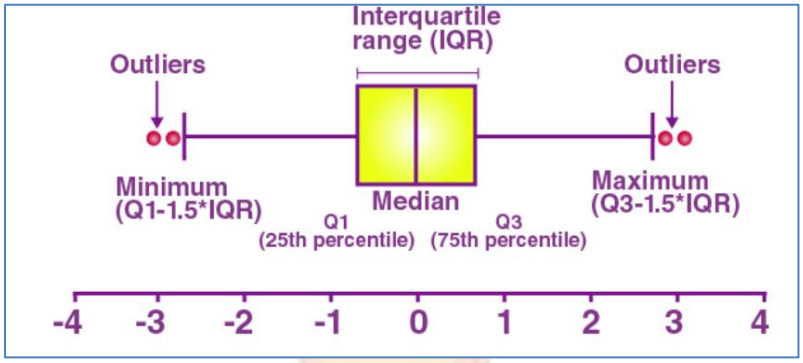


Fig 1: Box and Whisker Plot

A box and whisker plot is a graphical representation that summarizes data using a five-number summary, including one of the measures of central tendency. While it may not depict the distribution as precisely as a stem and leaf plot or histogram, its primary purpose is to indicate whether a distribution is skewed and to identify potential outliers or unusual observations in the dataset. Boxplots prove highly beneficial when dealing with large volumes of data sets or making comparisons between them. These plots offer an immediate understanding of central tendencies, spreads, and overall data ranges, facilitating easy comparisons. In essence, a box and whisker plot serves as a concise and descriptive tool for interpreting data presented on an interval scale, helping visualize data distribution and characteristics.

Elements of a Box and Whisker Plot

The elements required to construct a box and whisker plot outliers are given below.

1. Minimum value (Q0 or 0th percentile)
2. First quartile (Q1 or 25th percentile)
3. Median (Q2 or 50th percentile)
4. Third quartile (Q3 or 75th percentile)
5. Maximum value (Q4 or 100th percentile)
6. Interquartile range

The meaning of each of these elements is listed below.

1. The minimum value in the dataset, which is displayed at the far left end of the diagram.
2. The first quartile (Q1) at the left side, which is in between the minimum value and median.

3. The median value, represented by the line in the center of the box.
4. The third quartile (Q3) at the right side, which is in between the median and the maximum value.
5. The maximum value in the dataset, which is displayed at the far right end of the diagram.
6. Interquartile range (IQR) is the difference between upper and lower quartiles, i.e. Q3 and Q1.

Boxplot Distribution

The box plot distribution will explain how tightly the data is grouped, how the data is skewed, and also about the symmetry of data. As you can see in the figure given below

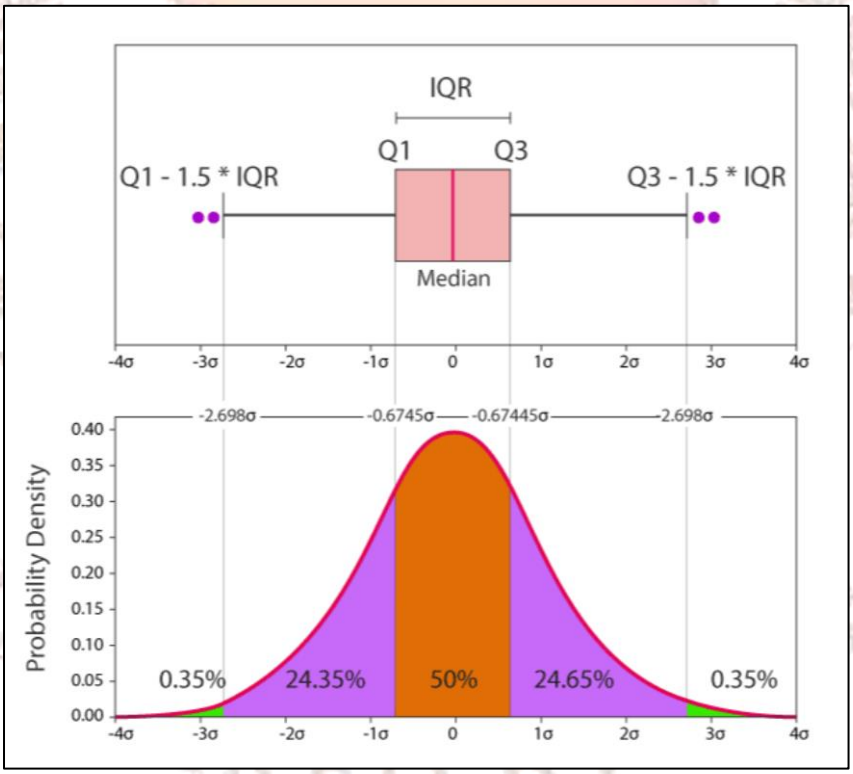


Fig 2: Box Plot Distribution

Positively Skewed: If the distance from the median to the maximum is greater than the distance from the median to the minimum, then the box plot is positively skewed.

Negatively Skewed: If the distance from the median to minimum is greater than the distance from the median to the maximum, then the box plot is negatively skewed.

Symmetric: The box plot is said to be symmetric if the median is equidistant from the maximum and minimum values.

Box Plot Chart

In a box and whisker plot:

- the ends of the box are the upper and lower quartiles so that the box crosses the interquartile range
- a vertical line inside the box marks the median
- the two lines outside the box are the whiskers extending to the highest and lowest observations.

Applications

It is used to know:

- The outliers and their values
- Symmetry of Data
- Tight grouping of data
- Data skewness – if, in which direction and how

How to Draw a Box and Whisker Plot?

The box and whiskers plot can be drawn using five simple steps. To draw a box and whisker diagram, we need to find:

Step 1: The smallest value in the data is called the minimum value.

Step 2: The value below the lower 25% of data contained, called the first quartile.

Step 3: Median value from the given set of data.

Step 4: The value above the lower 25% of data contained, called the third quartile.

Step 5: The largest value in the dataset is called maximum value.

Box and Whisker Plot Solved Example

Example: Draw the box plot for the given set of data: {3, 7, 8, 5, 12, 14, 21, 13, 18}.

Solution:

Firstly, write the given data in increasing order.

3, 5, 7, 8, 12, 13, 14, 18, 21

Range = Maximum value – Minimum value

Range = 21 – 3 = 18

Now, Median = center value of the given data

Median = 12

Now, we need to find the quartiles.

First quartile = Q1 = Median of data values present at the left side of Median

Q1 = Median of (3, 5, 7, 8)

Q1 = (5+7)/2 = 12/2 = 6

Third quartile = Q3 = Median of data values present at the right side of Median

Q3 = Median of (13, 14, 18, 21)

Q3 = (14+18)/2 = 32/2 = 16

Therefore, the interquartile range = Q3 – Q1 = 16 – 6 = 10

The five-number summary is given by:

Minimum, Q1, Median, Q3, Maximum

Hence, 3, 6, 12, 16, 21 is the five-number summary for the given data.

Now, we can draw the box and whisker plot, based on the five-number summary.

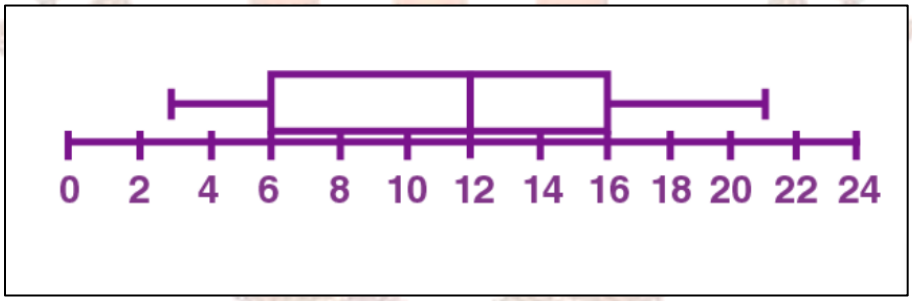


Fig 3: Output of Boix and Whisker Plot Solved Example

Creating a box and whisker plot in excel:

The data below has a list of temperatures recorded for a region

	A	B	C
1	**Months**	**Temperatures (in celcius)**	
2	January	10	
3	February	9	
4	March	5	
5	April	17	
6	May	15	
7	June	29	
8	July	13	
9	August	18	
10	September	11	
11	October	7	
12	November	12	
13	December	30	
14			

Fig 4: Data Set for Box and Whisker Plot

Let's create a box and a whisker plot out of it. To do that:

1. Select the data to be plotted (the numbers only)

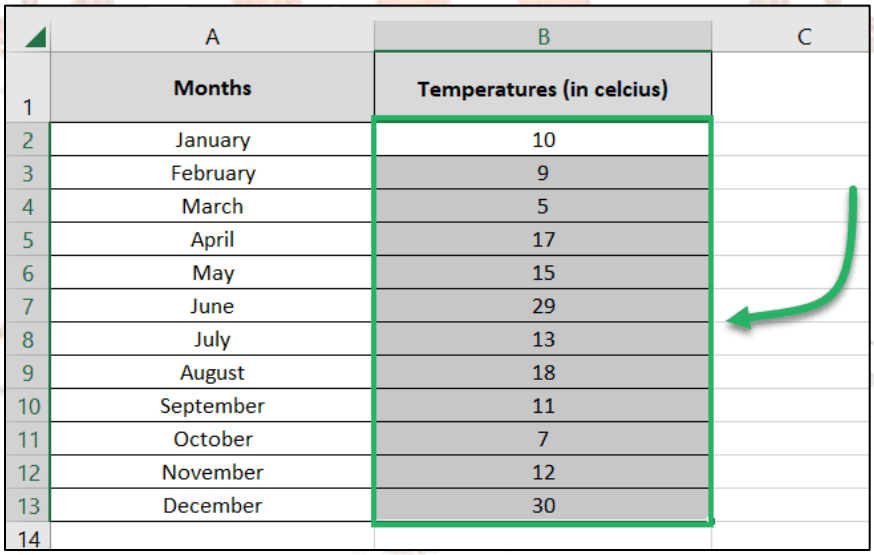


Fig 5: Selecting Data

2. Go to the **Insert tab > Charts.**

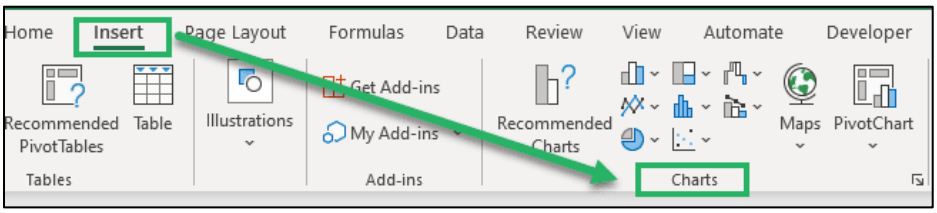


Fig 6: Insert tab

3. Click on the **Statistical Chart Icon > Box & Whisker Plot.**

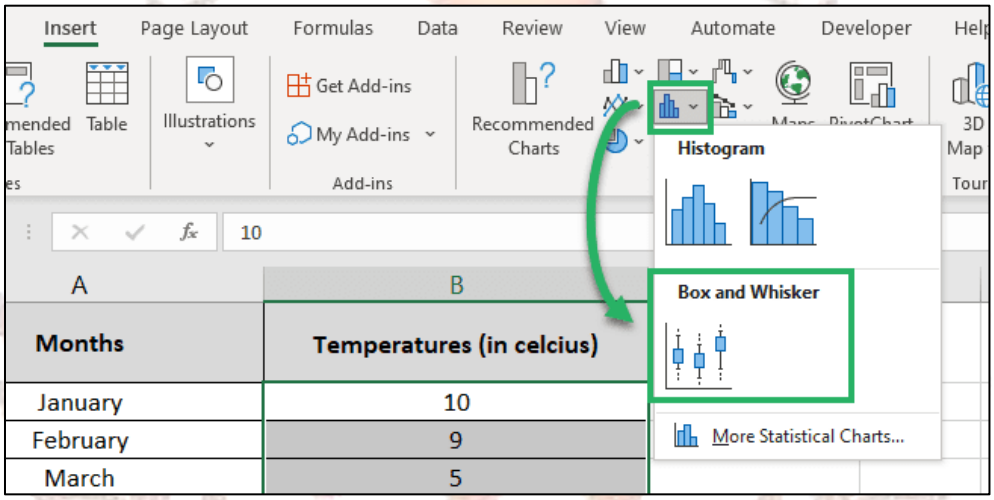


Fig 7: Select the Chart

And there you have a box and whisker chart created! Yes, creating it in Excel is only that simple

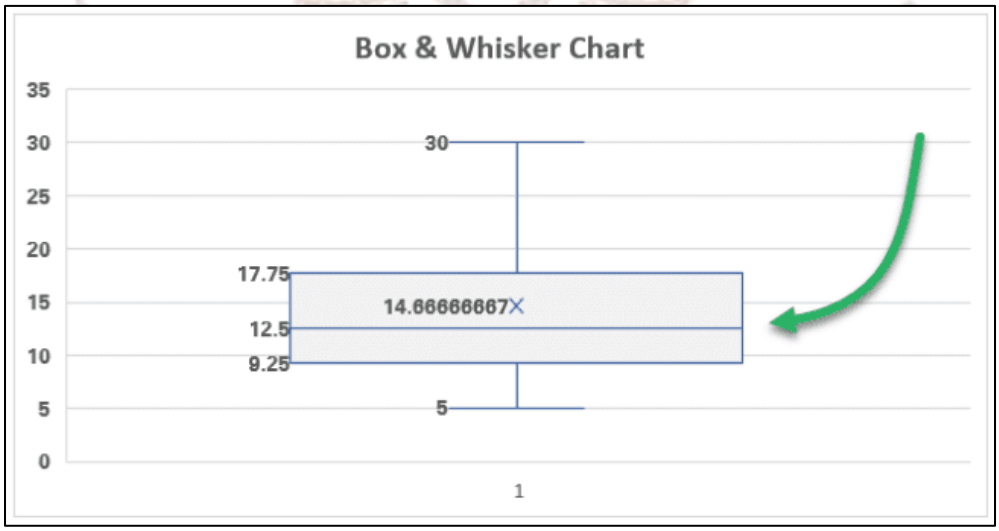


Fig 8: Output of Box-Whisker Plot

Explanation:

1. The whisker at the bottom shows the minimum value of our dataset (5).
2. Similarly, the whisker at the top shows the maximum value of our dataset (30).
3. The line in the center of the grey box shows the point where the Median for this dataset lies (at 12.5).
4. The Median makes the middle value of any dataset

You can quickly double-check the Median for your dataset by using the MEDIAN function as follows:

= MEDIAN (B2:B13)

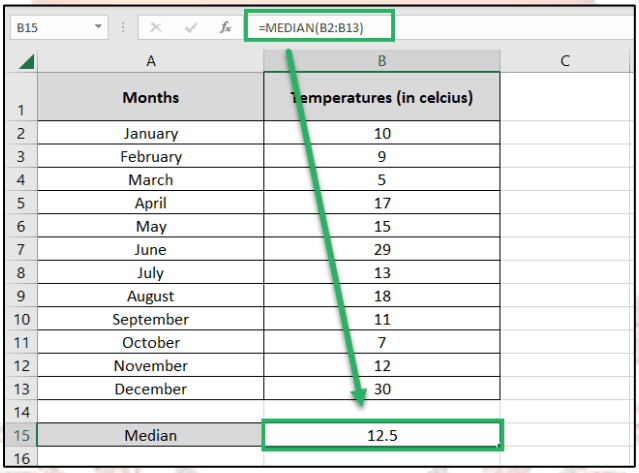


Fig 9: Output Explanation 1

Tip:

How to Calculate the Median:

1. Organize your data in ascending order.
2. Divide the data into two equal halves.
3. If the dataset contains an odd number of values, the median is the middle value.

For instance, with the dataset {1,3,5,7,9}, split it into {1,3} and {7,9}, and the middle value {5} is your median.

4. If the dataset has an even number of values, calculate the median by averaging the two central numbers.

For instance, with the dataset {1,3,5,7}, take the middle numbers 3 and 5, and average them: ((3+5) / 2 = 4). So, the median for this dataset is 4.

The "X" symbol in the center of the box represents the mean of the data, which in this case is 14.67.

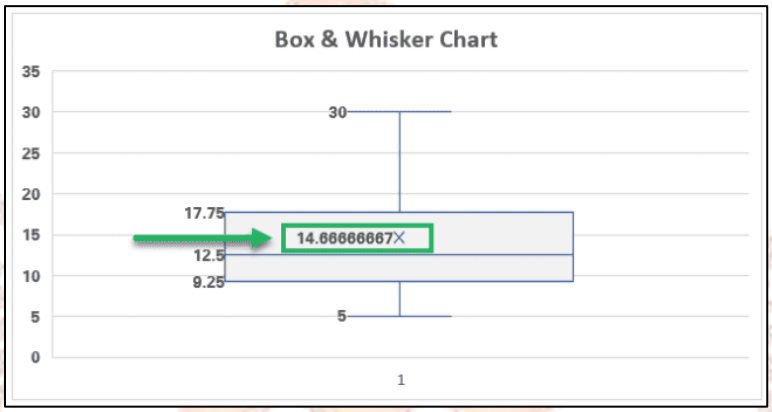


Fig 10: Output Explanation2

You can calculate the mean for any dataset by using the AVERAGE function in excel.

Add outliers to the box and whisker plot

Adding outliers to a box and whisker plot is the easiest.

But for this, we'll make a slight change to our dataset, which will cause the Box and the Whisker Plot to change too.

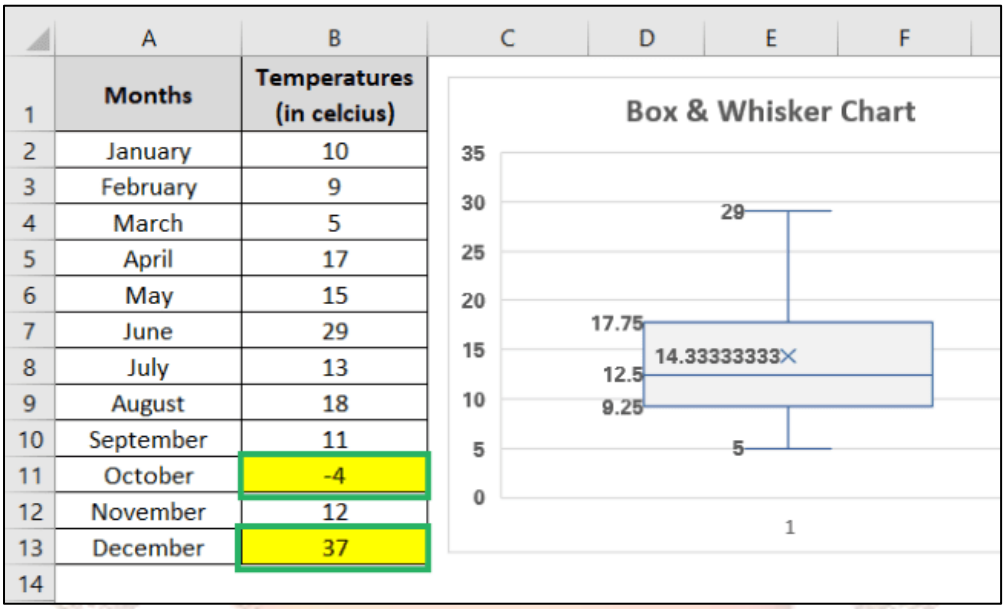


Fig 11: Updation in Data

No significant changes were made to the dataset, except for two values: one was changed from 30 to 37, and another from 7 to -4. Consequently, the top whisker of the chart now extends only up to 29 instead of 37 (the new maximum value of the dataset). The minimum value of the dataset has become -4, but the bottom whisker still extends only to 5. Why does this happen? We will explore the reason after understanding how to introduce an outlier. To add outliers to your box and whisker plot:

1. Select any part of the chart, such as the box or the whiskers.

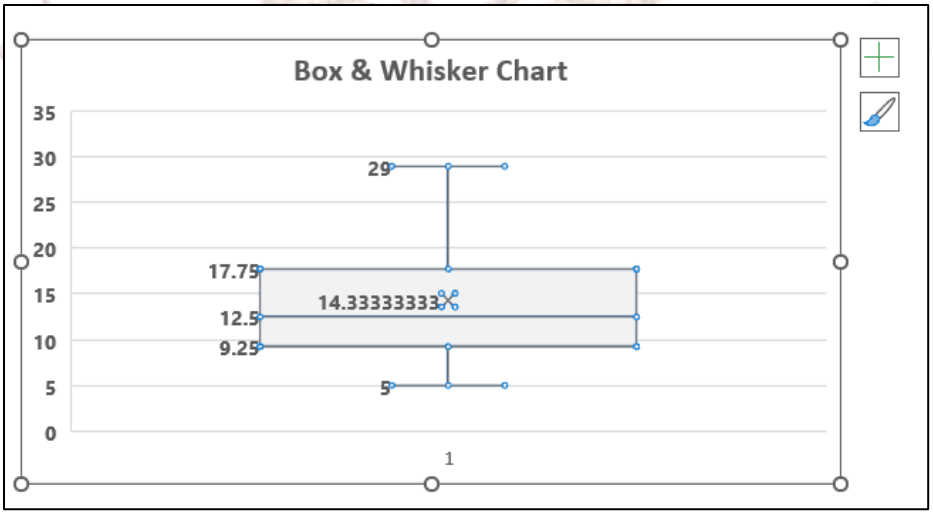


Fig 12: Select the Box or Whiskers on the Histogram

The Format data series pane will launch as shown below.

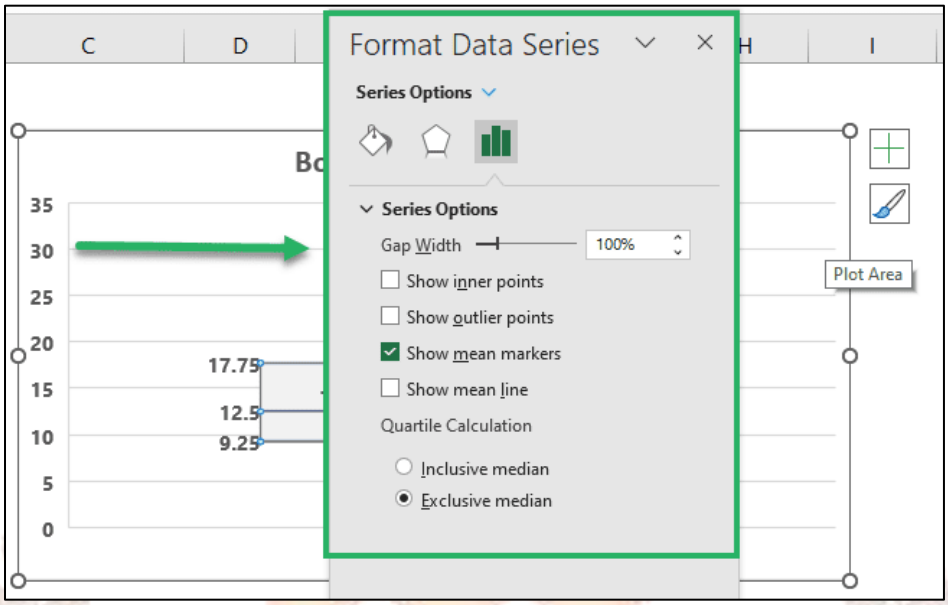


Fig 13: Format Data Series

2. Under **Series Options**, check the box for **"Show Outlier Points".**

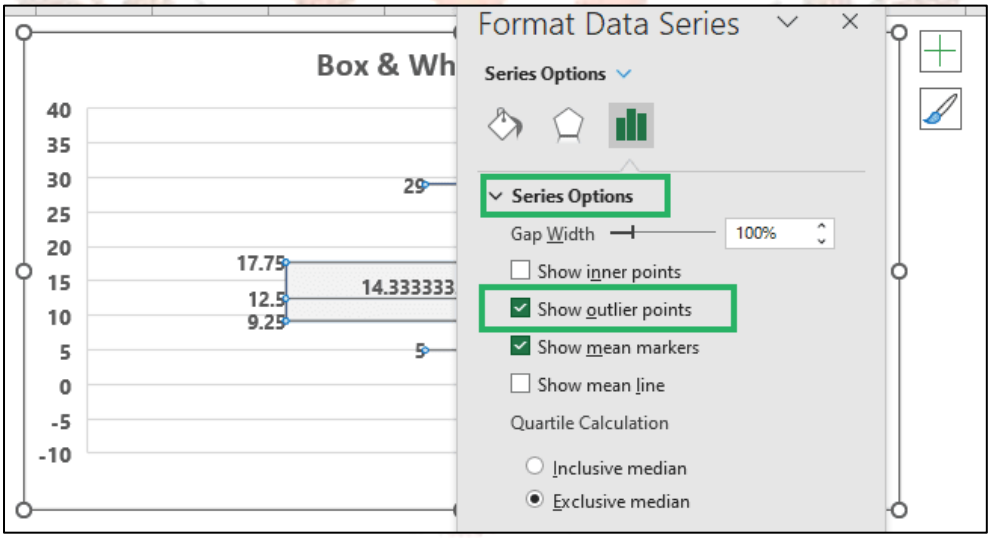


Fig 14: Select the Box on Outlier Points

There comes a blue dot on your chart that marks the highest (37) and the lowest value (-4) of our revised dataset. That's what we call an outlier.

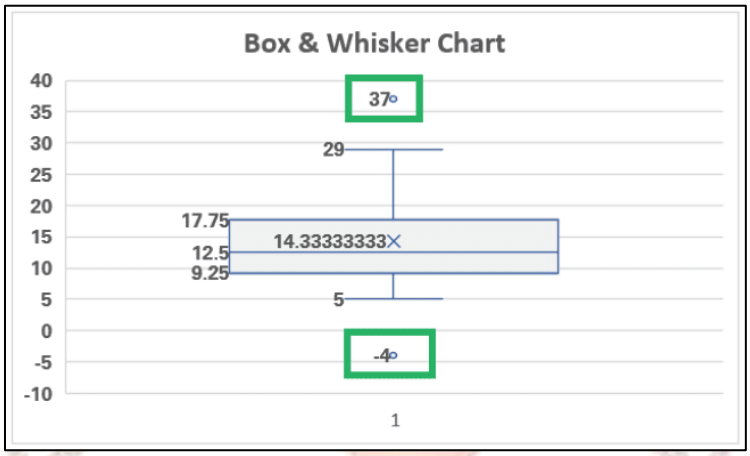


Fig 15: Updated Chart

Note: An outlier is an exceptional value that falls at either extreme end of a dataset, being significantly smaller or larger than the majority of the data points. Outliers have the potential to disrupt the overall distribution pattern of the data. Calculating outliers involves using the following formula:

- Lower Outlier = Quartile 1 - (IQR 1.5)
- Upper Outlier = Quartile 3 + (IQR 1.5)

To determine the interquartile range (IQR), you subtract the first quartile (Quartile 1) from the third quartile (Quartile 3):

- IQR = Quartile 3 - Quartile 1

The IQR rule states that if any data point in your dataset exceeds 1.5 times the IQR value, it is considered an outlier. Outliers are values that are either too small or too large and can significantly impact the dataset's overall distribution.

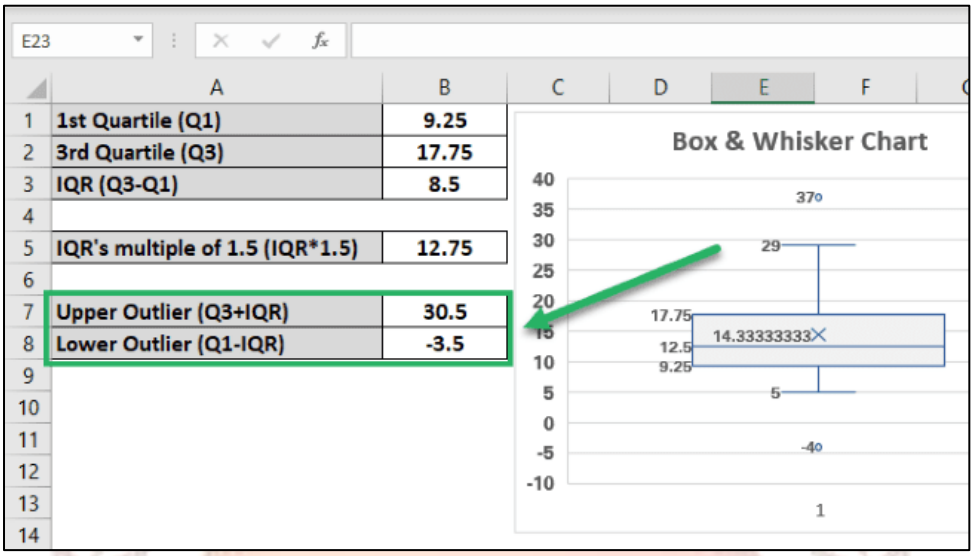


Fig 16: Outliers based on Data

Our outliers range from -3.5 to 30.5.

We have taken the numbers for the 1st and 3rd Quartile from the chart i.e. 17.75 and 9.25

3. WATER FALL CHART

A Waterfall Chart, a powerful graphical tool used to illustrate the cumulative effect of consecutive positive and negative variables on an initial starting point, offers a clear and structured representation of the step-by-step transitions and the complex interplay of factors contributing to a final outcome. It comprises vertical bars, with each bar representing a distinct data category, originating from an initial benchmark and extending either upwards (representing positive contributions) or downwards (indicating negative contributions) from this baseline. These bars' lengths visually indicate their impact on the overall change, culminating in the final value, reflecting the cumulative result of all contributing elements. Waterfall charts find extensive application in financial statements, budget variance analysis, and project cost breakdowns, providing a transparent visual exposition of individual components' influence on the outcome. The significance and adaptability of the Waterfall Chart, its functionality, guidelines for effective usage, and applications in various fields underscore its role as an invaluable data visualization tool.

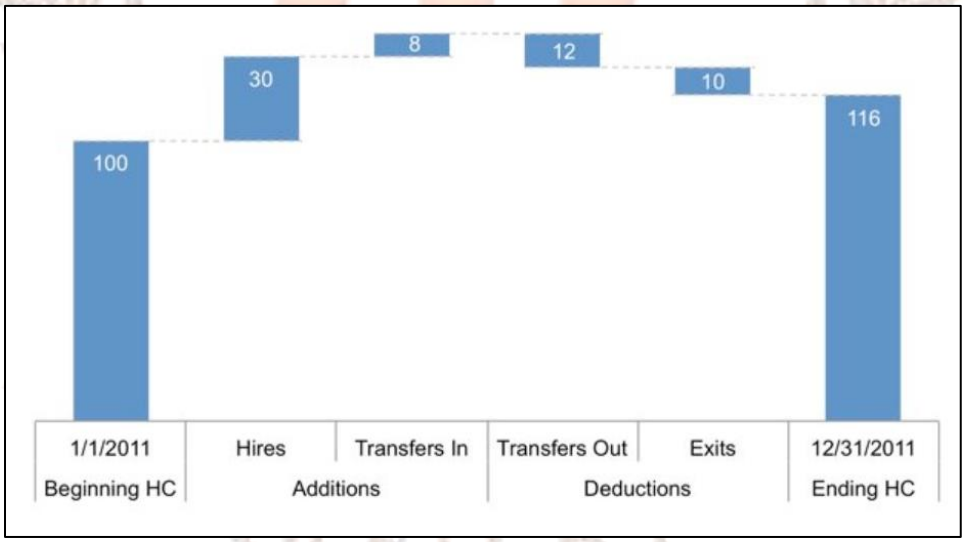


Fig 17: Waterfall Chart 1

A Waterfall Chart derives its name from its distinctive shape and structure. Typically, it begins with the first bar, originating from a baseline of zero, representing the initial quantity of the measured parameter. Subsequently, a sequence of smaller bars follows, appearing to levitate in space, often ascending to a peak before descending back towards the baseline.

This progression leads up to the final bar, symbolizing the ending quantity of the chosen measure, and like the initial bar, commences from the zero baseline on the x-axis.

The Waterfall Chart proves valuable in unveiling the complexity often concealed within cumulative figures. For instance, in the context of human resources, observing that a department had 20 employees both this year and last year in January may seem unremarkable at first glance. However, a Waterfall Chart can visually represent each component contributing to the staffing change throughout the year, revealing that 15 individuals resigned during the year, necessitating the hiring of 10 new employees and the transfer of five from another department to maintain the workforce. Compared to traditional bar or line charts, the Waterfall Chart provides a more comprehensive and potentially eye-opening narrative.

How water fall charts work:

A Waterfall Chart operates with a unique approach to baseline positioning compared to traditional bar charts. While the general rule in data visualization is to start all bar charts at a common baseline of zero, this principle applies only to the first and last bars in a Waterfall Chart, representing the initial and final values, respectively (often referred to as the "before" and "after" values). The component bars between these two endpoints have different baselines, determined by the running total. In essence, the end of the preceding bar becomes the baseline for the next one, creating a visual effect resembling a staircase that ascends and descends, connecting two or more pillars.

This concept, though straightforward, can sometimes pose interpretation challenges for audiences due to the unusual appearance of bars seemingly suspended in space. Adding to the complexity is the variation in baseline positioning for these component bars, which can be at the top (for negative values) or at the bottom (for positive values).

Waterfall Charts can also encompass multiple time periods, further complicating their structure. For instance, a Waterfall Chart depicting changes over several quarters within a year may feature more than two "pillars," with one representing the initial quarter's total and additional pillars for each subsequent quarter. In the spaces between these pillars, distinct regions of component bars illustrate gains and losses for those periods. This

arrangement can result in the appearance of arches (when ordered by gains followed by losses), powerlines (if losses precede gains), or zig-zags (in the absence of a specific order).

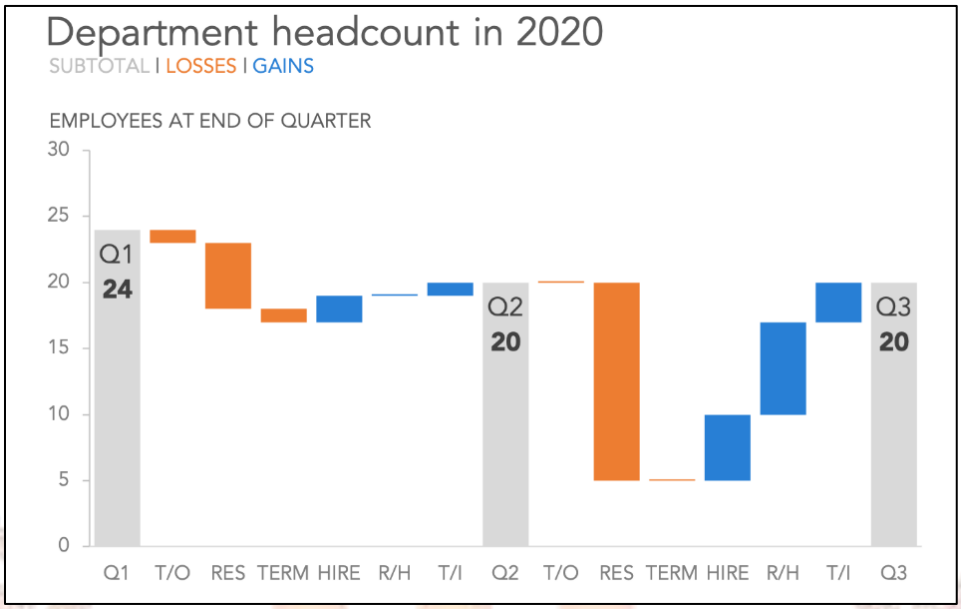


Fig 18: Waterfall Chart 2

Applications of Waterfall charts:

Waterfall Charts exhibit their versatility across various sectors and prove to be a valuable tool for illuminating trends and shifts in data. In the realm of human resources (HR), these charts are commonly employed to visually represent changes in hiring and workforce attrition rates. This visual representation equips HR professionals with a powerful tool to effectively track and communicate personnel changes over time, aiding strategic decision-making and resource allocation.

In the financial sector, Waterfall Charts play a crucial role in elucidating financial movements within specific timeframes. They succinctly present financial data, including credits, debits, gains, and losses, providing a clear depiction of the flow of financial resources. Financial analysts and industry experts rely on these charts to dissect complex financial data, simplifying the identification of key drivers behind profitability or areas of concern during the designated period.

Moreover, Waterfall Charts find utility in industries where real-time tracking of active accounts or subscriptions, along with associated revenue, is essential for core business

operations. These charts methodically illustrate the evolution of these accounts, shedding light on the factors contributing to revenue growth or decline. By continuously monitoring these changes, businesses can proactively adjust their strategies, optimizing performance and enhancing customer retention.

Drawbacks of water fall charts:

Waterfall charts have certain limitations that users should be aware of. Firstly, because these charts involve comparing the lengths of bars that do not share a common baseline, it can be challenging for viewers to accurately compare the specific sizes of growth or contraction between subcategories. When conveying the specific values of component pieces is crucial, it is advisable to directly label those pieces to enhance clarity.

Secondly, waterfall charts present the challenge of bars being read differently—bottom-up or top-down—depending on whether they represent gains or losses. This can be mitigated by grouping gains and losses together to maintain a regular arched or bowed shape in the chart. However, if subcategories are sorted in a way that intersperses gains and losses, it becomes more difficult to extract insights.

Another issue is related to the scale of values in waterfall charts. Sometimes, the "pillars" of these charts represent significantly larger numbers compared to the component pieces representing smaller changes. When the differences in values are substantial, it can be challenging to perceive changes accurately in the chart. Zooming in on the interesting part of the chart, i.e., the components of change, is not a viable solution because it negates the fundamental purpose of using a bar chart, which encodes quantitative value in the length of the bar, requiring it to start at zero.

To address this issue, one alternative approach is to use a bar of zero length as the starting point and illustrate the changes over time, concluding with a final bar that shows the net change from the prior period. The actual value of the starting amount can then be conveyed using text to ensure clarity in communication.

When creating waterfall charts, it's essential to follow several best practices to enhance their effectiveness as data visualization tools:

1. Ensure Full Columns at the Extremes: Start with a complete bar representing the initial value on the left and end with another complete bar for the final value on the right. These robust bookends provide viewers with a clear reference frame, enabling them to understand all the changes that occur in between, much like the starting and finishing lines in a race.
2. Incorporate Horizontal Lines for Clarity: To emphasize the interconnectedness of data points and help viewers comprehend spatial transitions, consider adding horizontal lines that connect the bars. These lines serve a dual purpose by highlighting relationships between figures and facilitating the visual tracking of changes, acting as bridges through the data landscape.
3. Strategically Use Colors: Color selection is crucial in data visualization. In waterfall charts, colors serve not only to enhance comprehension but also to prevent confusion. Two common approaches exist: using distinct colors for negative and positive values to create a clear visual contrast, or choosing colors for specific reasons, such as highlighting particular factors. Regardless of the approach, the color scheme should be a deliberate choice aimed at improving the clarity and accuracy of data presentation, avoiding potential misinterpretation.
4. Prioritize Clarity in Labeling: Clear labeling and headings are of paramount importance, especially in waterfall charts. Thoughtfully employ labels and headings as signposts to help guide the audience through the intricacies of the chart and enable them to grasp the conveyed information effortlessly.
5. Consider Scale Flexibility: While waterfall charts offer flexibility in scale compared to other chart types, it's crucial to understand the limits of this flexibility. Although starting the scale at zero is not mandatory, maintaining an even and non-misleading scale is imperative. For example, if the initial staff count was 120, and only 20 were hired, adjust the starting point of the scale to accurately reflect this change. This adjustment may involve using alternative colors or visual cues to clarify that the data represents total change, not incremental change. Such practices ensure that viewers can accurately interpret the chart, even when it deviates from traditional zero-based scales.

Creation of Waterfall charts using excel:

To enhance our comprehension of the concepts related to waterfall charts and their application, let's construct a straightforward sample table using sales figures as an illustration. I will use sales amounts to illustrate this concept. In the table provided below, we will track variations in sales over multiple months, observing instances of both positive and negative changes relative to the initial value.

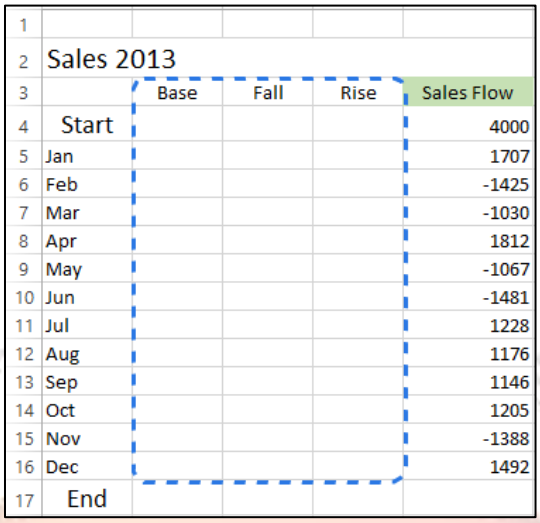
2	Sales 2013	
3		Sales Flow
4	Start	4000
5	Jan	1707
6	Feb	-1425
7	Mar	-1030
8	Apr	1812
9	May	-1067
10	Jun	-1481
11	Jul	1228
12	Aug	1176
13	Sep	1146
14	Oct	1205
15	Nov	-1388
16	Dec	1492

Fig 19: Dataset for Waterfall Chart

Creating an Excel bridge chart is an excellent method to visually represent the sales fluctuations over a span of twelve months. However, if you attempt to apply a Stacked Column chart template to these specific values, it won't resemble a waterfall chart in any way. Therefore, the initial step involves carefully reorganizing your data.

Step 1: Rearrange the data table

To begin, you should insert three additional columns into your Excel table, namely Base, Fall, and Rise. The Base column will contain a calculated amount that serves as the starting point for the Fall and Rise series within the chart. All the negative values from the Sales Flow column will be placed in the Fall column, while all the positive values will be situated in the Rise column. This restructuring of the data will facilitate the creation of a waterfall chart to effectively visualize these variations.



1					
2	Sales 2013				
3		Base	Fall	Rise	Sales Flow
4	Start				4000
5	Jan				1707
6	Feb				-1425
7	Mar				-1030
8	Apr				1812
9	May				-1067
10	Jun				-1481
11	Jul				1228
12	Aug				1176
13	Sep				1146
14	Oct				1205
15	Nov				-1388
16	Dec				1492
17	End				

Fig 20: Add helper Columns

Additionally, I've included an "End" row at the bottom of the list of months to calculate the total sales amount for the entire year. Now, let's proceed to the next step and populate these columns with the necessary values.

Step 2: Insert formulas

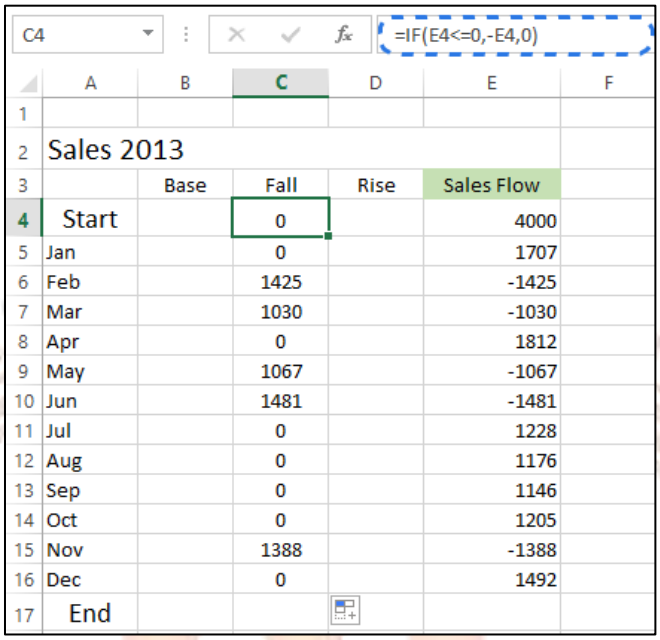
To complete the table efficiently, we'll input specific formulas into the initial cells of the corresponding columns and then extend them to adjacent cells using the fill handle.

1. Begin by selecting cell C4 in the Fall column and enter the following formula: =IF(E4<=0, -E4, 0)

This formula determines that if the value in cell E4 is less than or equal to zero, it will be displayed as a positive number, and if it's a positive number, it will be shown as zero.

Note. If you want all the values in a waterfall chart lie above zero, you need to enter the minus sign **(-)** before the second cell reference **(E4)** in the formula. And two minuses will make a plus.

1. Copy the formula down to the end of the table.



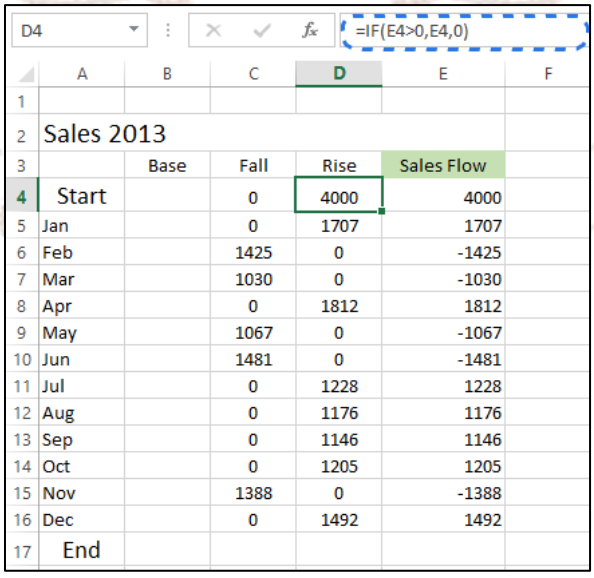
	A	B	C	D	E	F
1						
2	Sales 2013					
3		Base	Fall	Rise	Sales Flow	
4	Start		0		4000	
5	Jan		0		1707	
6	Feb		1425		-1425	
7	Mar		1030		-1030	
8	Apr		0		1812	
9	May		1067		-1067	
10	Jun		1481		-1481	
11	Jul		0		1228	
12	Aug		0		1176	
13	Sep		0		1146	
14	Oct		0		1205	
15	Nov		1388		-1388	
16	Dec		0		1492	
17	End					

Fig 21: Use formula to update values in Helper Column Fall

2. Click on cell **D4** and type in **=IF(E4>0, E4,0).**

It means if the value in cell E4 is greater than zero, all the positive numbers will be displayed as positive and the negative ones as zero.

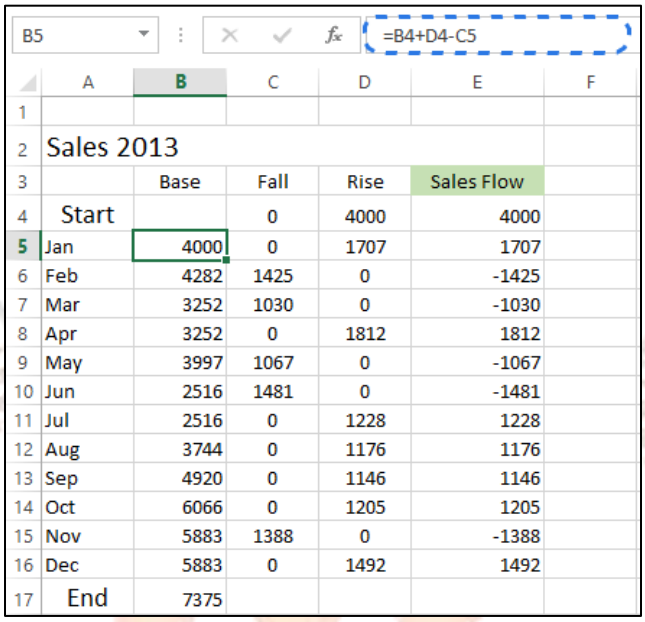
3. Use the fill handle to copy this formula down the column.



	A	B	C	D	E	F
1						
2	Sales 2013					
3		Base	Fall	Rise	Sales Flow	
4	Start		0	4000	4000	
5	Jan		0	1707	1707	
6	Feb		1425	0	-1425	
7	Mar		1030	0	-1030	
8	Apr		0	1812	1812	
9	May		1067	0	-1067	
10	Jun		1481	0	-1481	
11	Jul		0	1228	1228	
12	Aug		0	1176	1176	
13	Sep		0	1146	1146	
14	Oct		0	1205	1205	
15	Nov		1388	0	-1388	
16	Dec		0	1492	1492	
17	End					

Fig 22: Use formula to update values in Helper Column Rise

4. Insert the last formula =**B4+D4-C5** in cell **B5** and copy it down; include the End row.



B5 =B4+D4-C5

	A	B	C	D	E	F
1						
2	Sales 2013					
3		Base	Fall	Rise	Sales Flow	
4	Start		0	4000	4000	
5	Jan	4000	0	1707	1707	
6	Feb	4282	1425	0	-1425	
7	Mar	3252	1030	0	-1030	
8	Apr	3252	0	1812	1812	
9	May	3997	1067	0	-1067	
10	Jun	2516	1481	0	-1481	
11	Jul	2516	0	1228	1228	
12	Aug	3744	0	1176	1176	
13	Sep	4920	0	1146	1146	
14	Oct	6066	0	1205	1205	
15	Nov	5883	1388	0	-1388	
16	Dec	5883	0	1492	1492	
17	End	7375				

Fig 23: Use formula to update values in Helper Column Base

This formula calculates base values that will prop up the rises and falls to the appropriate height.

Step 3. Create a standard Stacked Column chart

Now your data are well-organized and you are ready to build the chart itself.

1. Select your data including the column and row headers, exclude the Sales Flow column.
2. Go to the Charts group on the INSERT tab.
3. Click on the Insert Column Chart icon and choose Stacked Column from the drop-down list.

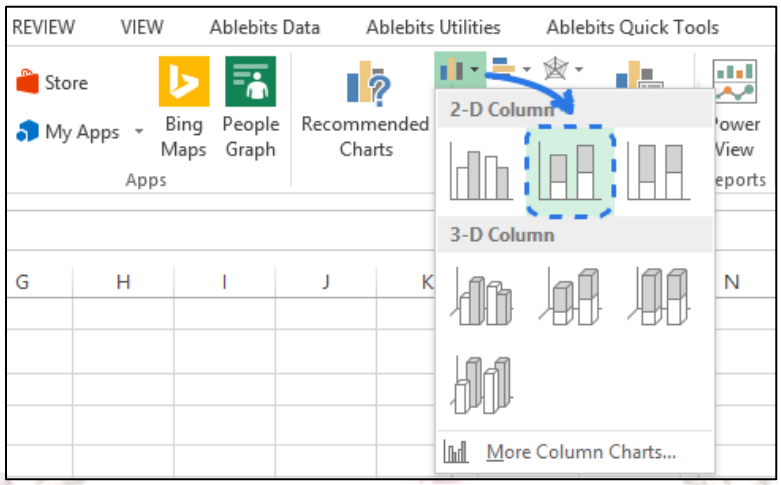


Fig 24: Select the chart

The graph appears in the worksheet, but it hardly looks like a waterfall chart. Take the next step and turn the stacked column graph into Excel bridge chart.

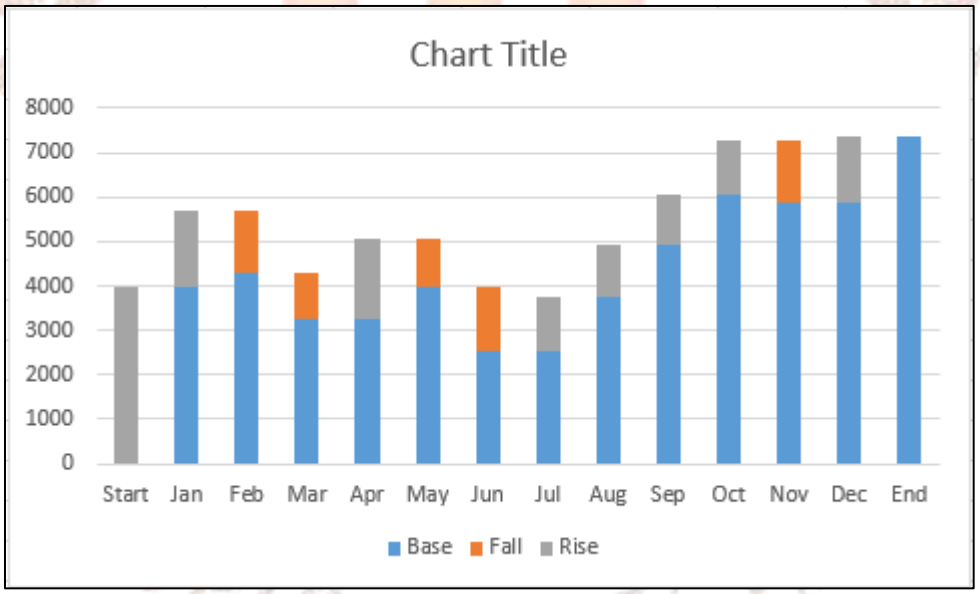


Fig 25: Chart Output

Step 4. Transform the column graph into a waterfall chart

Now you just need to make the Base series invisible to get a waterfall chart from a stacked column.

4. Click on the Base series to select them, right-click and choose the *Format Data Series…* option from the context menu.

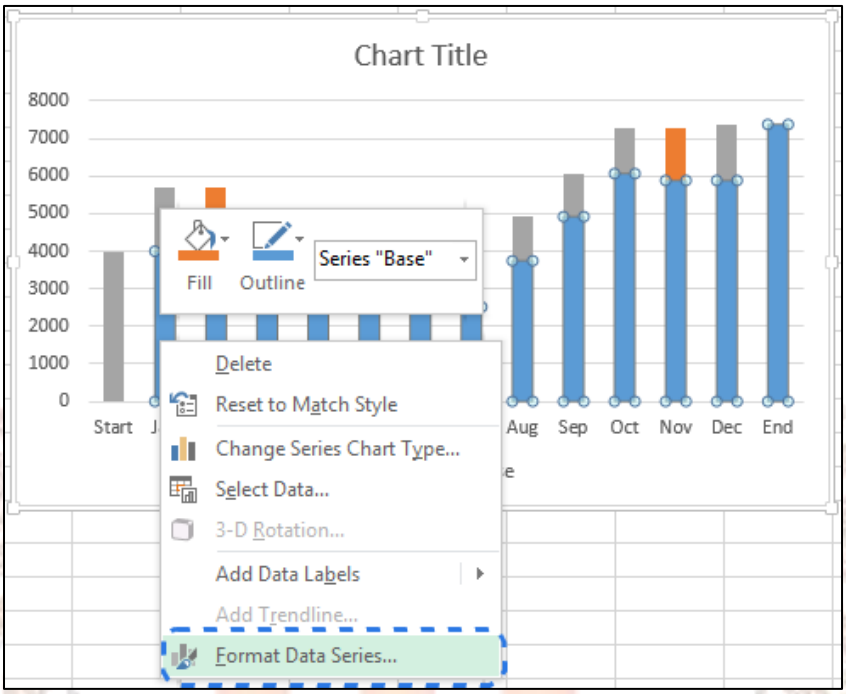


Fig 26: Format Data Series

The *Format Data Series* pane immediately appears to the right of your worksheet in Excel 2013 / 2016.

5. Click on the *Fill & Line* icon.
6. Select *No fill in the Fill section* and *No line* in the *Border* section.

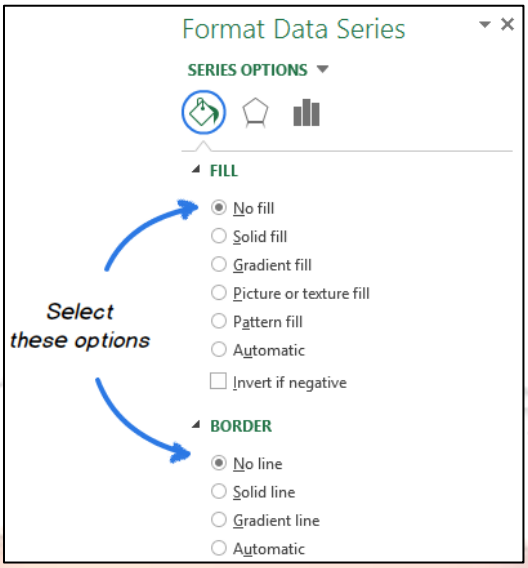


Fig 27: Update the format

When the blue columns become invisible, just delete Base from the chart legend to completely hide all the traces of the Base series.

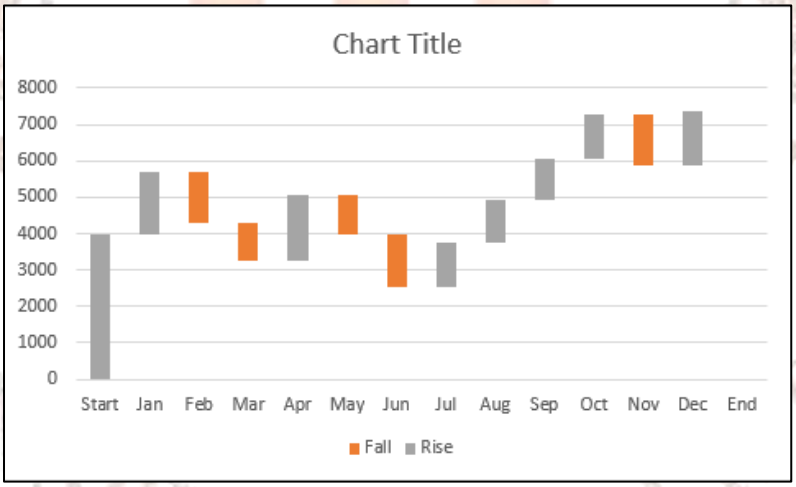


Fig 28: Updated Chart

Step 5. Format Excel bridge chart

Let's finish up with a little formatting. First we will make the flying bricks brighter and highlight the Start and End values in the chart:

1. Select the Fall series in the chart and go to the *FORMAT* tab under *CHART TOOLS.*
2. Click on *Shape Fill* in the *Shape Styles* group.

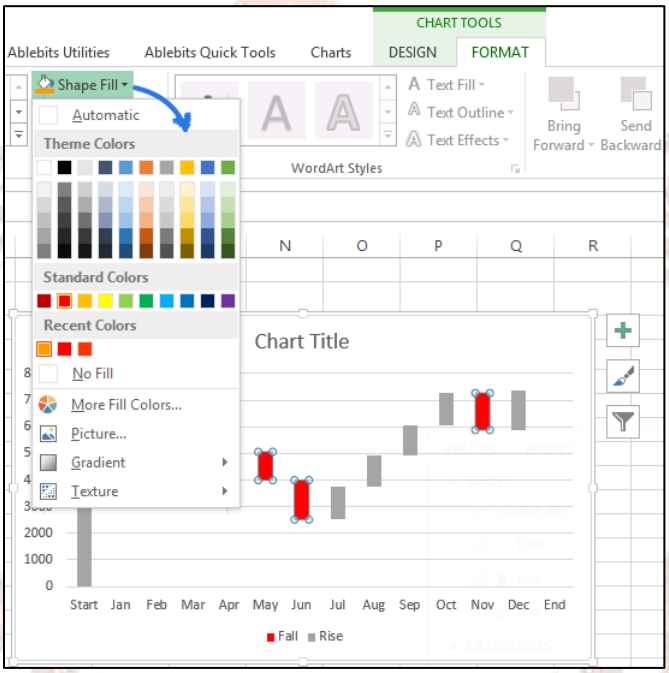


Fig 29: Select colour

3. Pick the color you want in the drop-down menu.

Here you can also experiment with the column outline or add special effects to them. Just use the *Shape Outline* and *Shape Effects* options on the *FORMAT* tab to make changes.

Then you should do the same trick with the Rise series. As for the Start and End columns, you need to color-code them individually, but they should be of the same color.

When you are done, the chart should look like the one below:

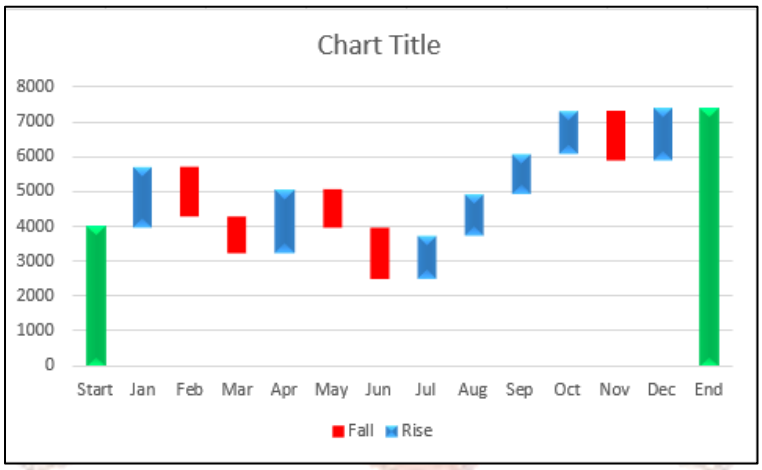


Fig 30: Updated Chart

Note. Alternatively, you can change the color and outline of the columns in the chart by opening the Format Data Series pane or choosing the *Fill* or *Outline* options in the right-click menu.

Then you can remove excess white spaces between the columns to make them stand closer to one another:

4. Double-click on one of the chart columns to bring up the **Format Data Series** pane.
5. Change the Gap Width to something smaller, like **15%.** Close the pane.

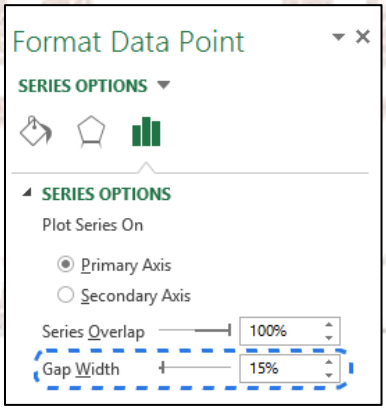


Fig 31: Update Gap Width

Now the holes in the bridge chart are patched.

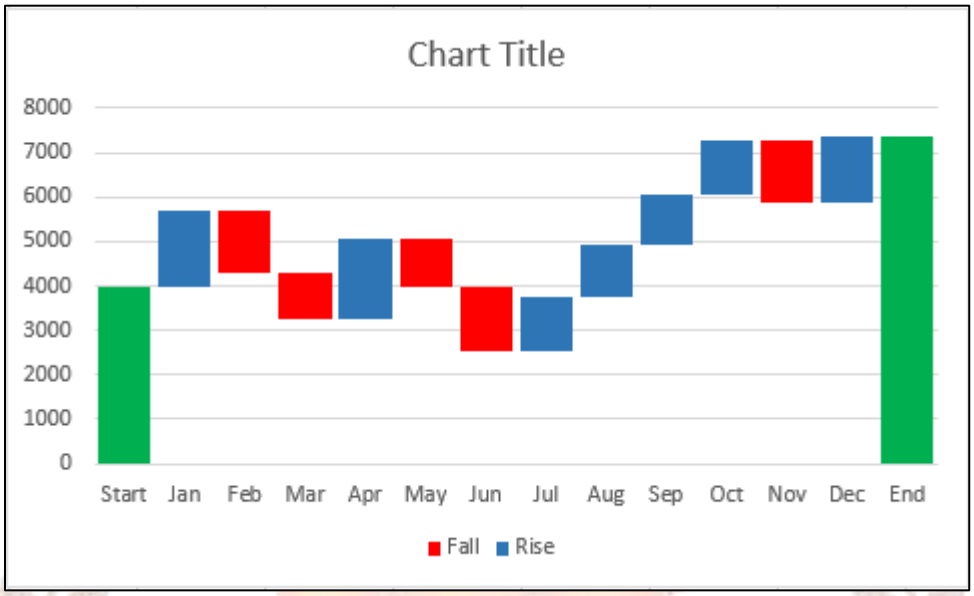


Fig 32: Updated Chart

When you look at the waterfall chart above, some of the flying bricks seem to be of the same size. However, when you refer to the data table, you'll see that the represented values are different. For more accurate analysis its recommended to add data labels to the columns.

6. Select the series that you want to label.
7. Right-click and choose the *Add Data Labels* option from the context menu.

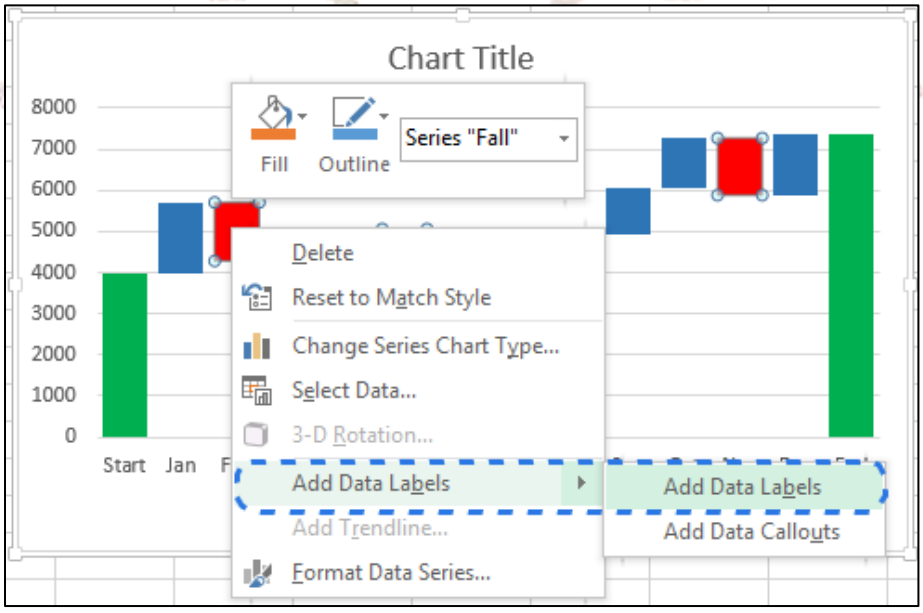


Fig 33: Add Data Labels

Repeat the process for the other series. You can also adjust the label position, the text font and color to make the numbers more readable.

Note. If there is an apparent difference in column size and the specifics aren't important, you can omit the data labels, but then you should add a Y-axis for better data interpretation.

When you are done with labelling the columns, just get rid of unnecessary elements such as zero values and the legend. You can also change the default chart title to something more descriptive.

The waterfall chart is now prepared! It presents a deep contrast to the frequently utilized chart formats and is exceptionally easy to understand.

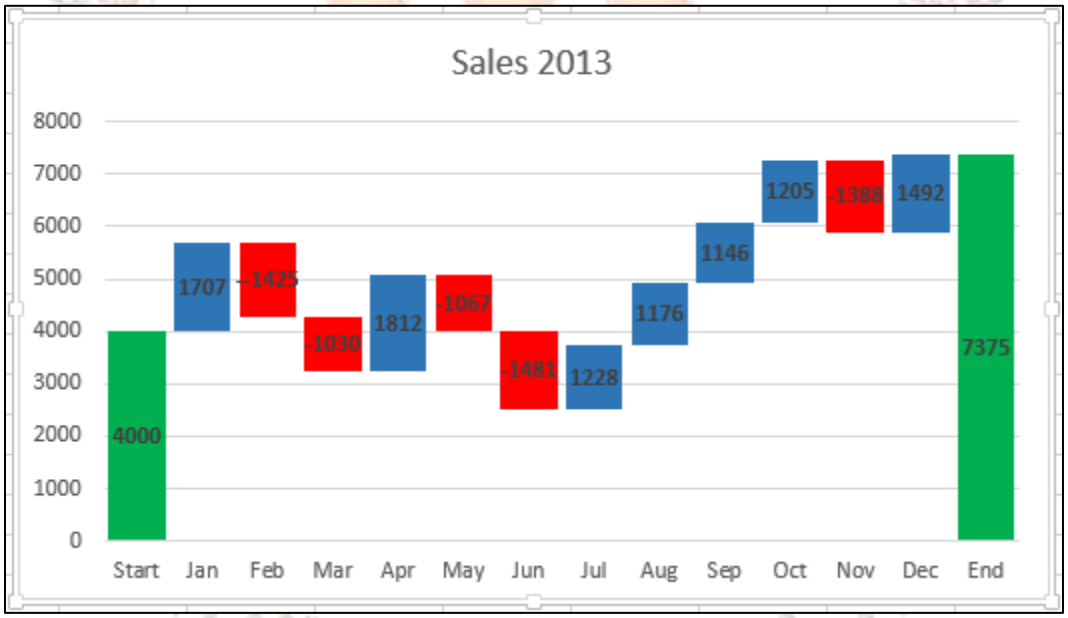


Fig 34: Updated Chart

Add-ins for creating waterfall charts

As you can see, it is not difficult at all to manually create a waterfall chart in Excel 2016 - 2010. But if you don't want to mess with rearranging your data, you can use special add-ins that can do all the work for you.

4. AREA CHART

An area chart in Excel is a type of data visualization that represents data series as filled areas, creating a visual depiction of their cumulative values over a continuous axis, typically time. It is similar to a line chart, but the area beneath each line is shaded or coloured, which makes it easier to observe the total or relative contribution of each data series to the whole, especially when comparing multiple data series. Area charts are commonly used to display trends and patterns over time, showing how different categories or variables contribute to the overall picture.

The Area Chart in Excel is best-suited in displaying patterns and trends of key data points.

Types of Area Chart

There are three types of Area Chart in Excel Let's check them out.

- **Standard Area Chart**

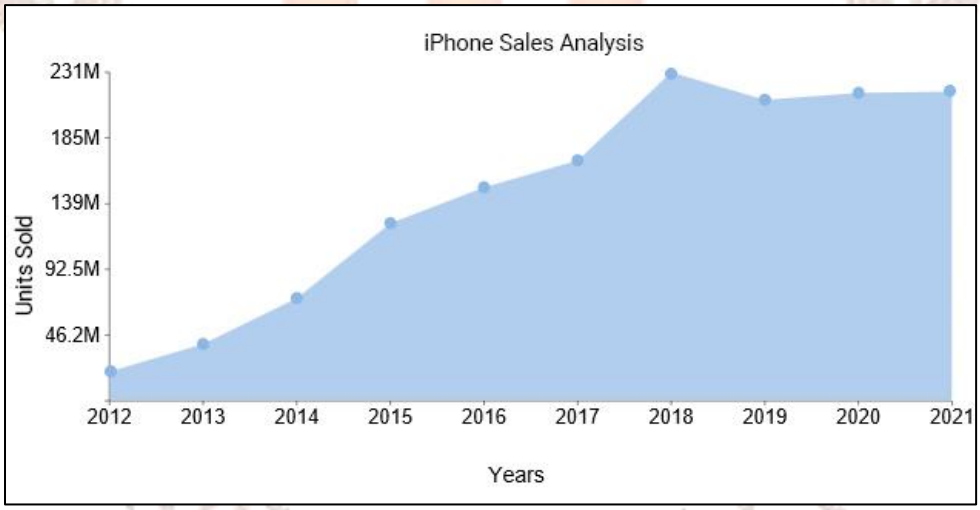


Fig 35: Area Chart 1

This variant of Area Chart in Excel is best suited in displaying changes of key variables over time. The variable depicted by the y-axis is the metric. On the other hand, the main dimension in the chart is time. For instance, the chart above displays the growth trend of units sold in the last 9 years.

- **Stacked Area Chart**

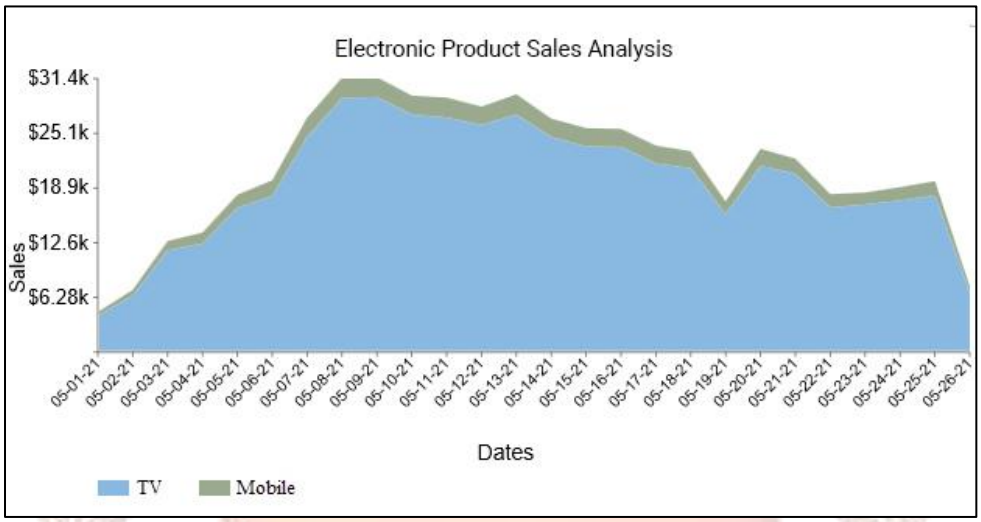


Fig 36: Area Chart 2

Visualization Source: Chart Expo

A Stacked Area Chart is a visualization design you can use to display trends and patterns of two or more variables in data. Besides, it shows relationships between variables over time.

Each line in the chart is shaded from its peak to a common baseline. The top line corresponds to the aggregate value when summing across all groups.

You can use the variant of Stacked Area Chart in Excel to:

- *Track trends and patterns over time*
- *Display part to whole relationships between key data point*

- **Area Line Chart**

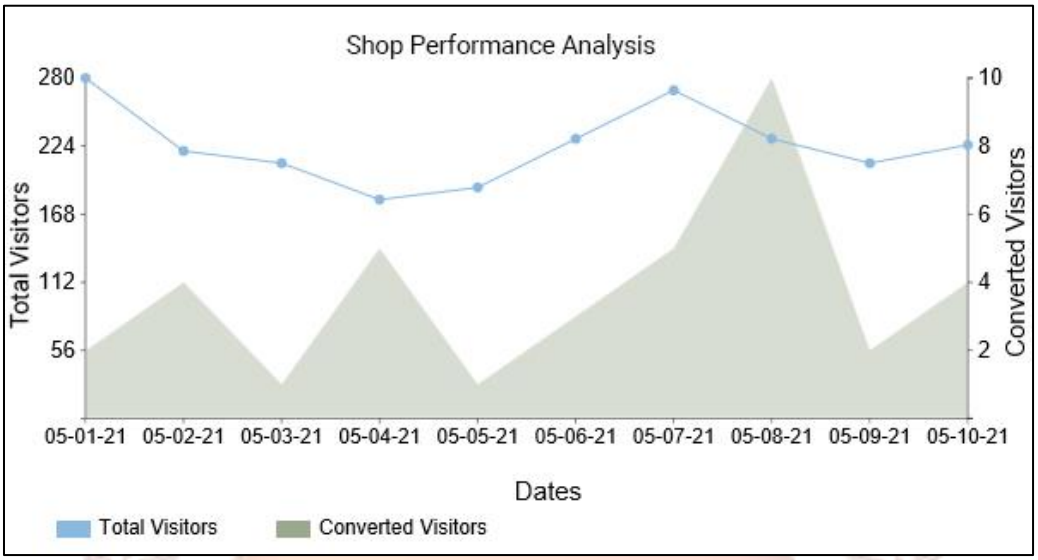


Fig 37: Area Line Chart

You can use the Area Line Chart to plot two different data series.

This chart achieves the task (above) using a secondary Y-axis on the right-hand side. More so, you can use the Area Line Chart to display the relationship between two variables. In other words, this chart is convenient in uncovering hidden insights into the relationship between two or more measures with different scales and measures.

Create area chart using excel:

Year	Units Sale
2012	20730000
2013	39990000
2014	72290000
2015	125050000
2016	150260000
2017	169220000
2018	231220000
2019	211880000
2020	216760000
2021	217720000

Fig 38: Dataset

We'll use Area Chart in Excel to visualize the table below.

- Open your Excel and paste the table above.
- Open the worksheet and click the Insert menu button.
- From Insert menu click the My Apps button to access the Chart Expo add-in.

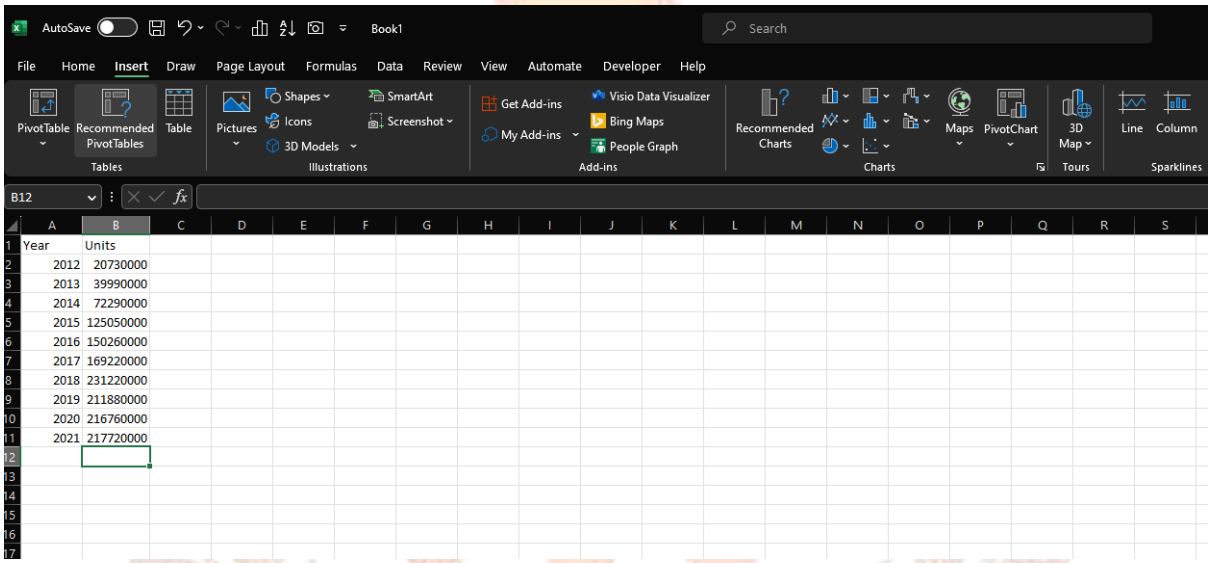


Fig 39: Insert tab

- Select **Chart Expo** and click the **Insert** button to get started with Chart Expo.

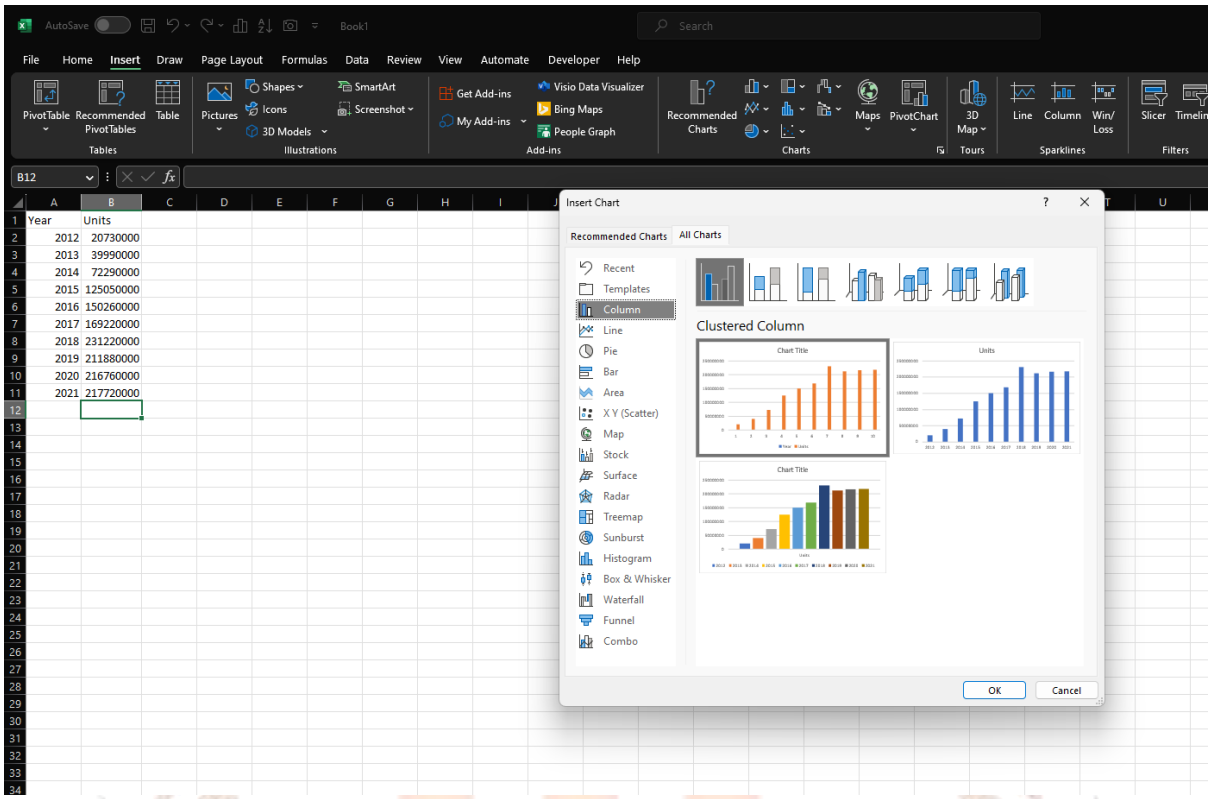


Fig 40: Select Chart Expo

- Once ChartExpo is loaded, you will see a list of charts.

Fig 37: Resultant Chart

- Look for "Area Chart."
- Once the Chart pops up, click on its icon to get started, as shown below.

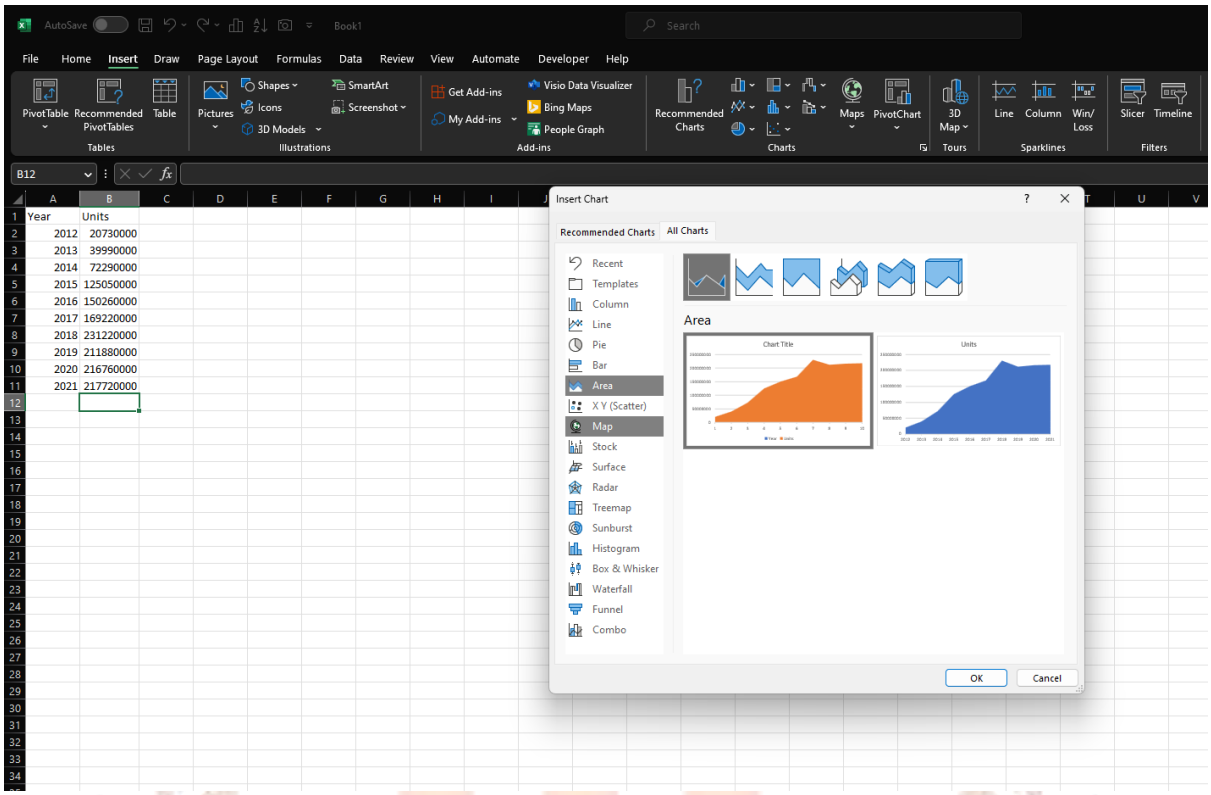


Fig 41: Select Area Chart

- Select the sheet holding your data and click the **Create Chart From Selection** button.

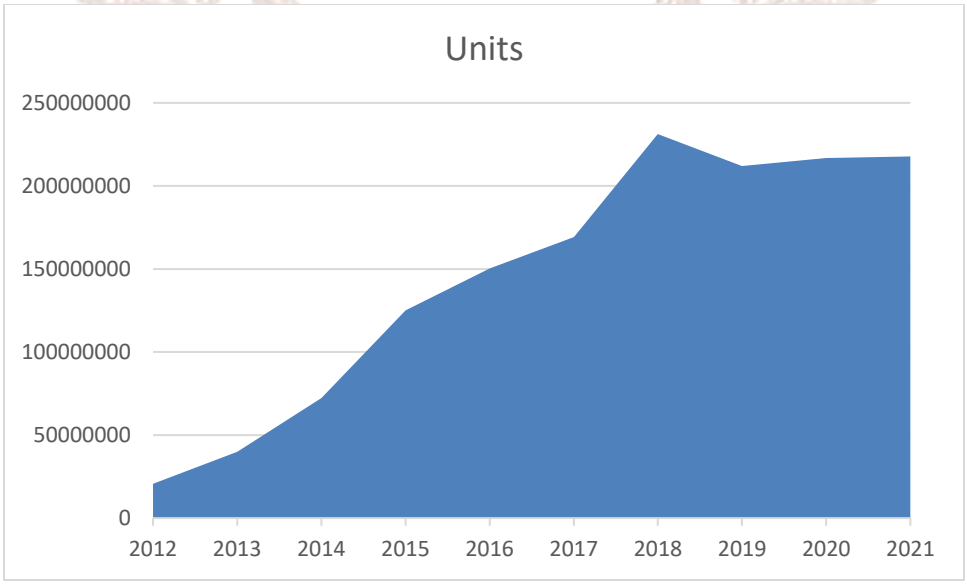


Fig 42: Output Chart

Insights

- Between 2012 and 2018, the number of units sold was on a growth trend.
- The 6-year growth peaked in 2018 before taking a nosedive.
- There was no trend between 2019 and 2020. In other words, the quantity sold did not grow or decline.

5. SUMMARY

The chapter on data visualization techniques encompasses three distinct chart types: the box and whisker plot, area chart, and waterfall chart. The box and whisker plot provides a concise and informative way to summarize data distribution, central values, and variability. The area chart, on the other hand, presents data series as filled areas, making it ideal for showcasing cumulative trends and contributions of different variables over time. Finally, the waterfall chart is a powerful tool for illustrating the sequential impact of positive and negative variables on an initial value, shedding light on complex data transitions. These three chart types offer diverse ways to visualize and interpret data in various fields and scenarios, enhancing the understanding of data trends and patterns.

6. QUESTIONS

Seft Assessment Questions:

SELF-ASSESSMENT QUESTIONS – 1

1. What type of data visualization is a box and whisker plot?
2. In a box and whisker plot, what does the box represent?
3. How does an area chart differ from a line chart?
4. What is the primary use of an area chart?
5. In a waterfall chart, what do the bars represent?
6. What is the purpose of using a waterfall chart in data visualization?
7. What is the key advantage of using a box and whisker plot for data analysis?
8. When might you choose to use an area chart to represent data?
9. How is the central tendency of data represented in a box and whisker plot?
10. In a waterfall chart, how are positive and negative values typically depicted?

Terminal Questions

1. Explain the components of a box and whisker plot and their significance in summarizing a dataset's distribution. Provide a step-by-step example of how to create and interpret a box and whisker plot using a sample dataset.
2. Describe the main characteristics and use cases of area charts in data visualization. Provide an example of a situation where an area chart is more suitable than other chart types, and explain why.
3. Discuss the purpose and benefits of using a waterfall chart in representing financial data or cumulative effects. Provide a step-by-step guide on how to create a waterfall chart in Excel and explain the significance of its different components.
4. Compare and contrast the box and whisker plot, area chart, and waterfall chart in terms of their primary functions, strengths, and limitations. Explain when each chart type should be used and under what circumstances they might not be appropriate.

5. Identify and describe real-world scenarios or industries where box and whisker plots, area charts, and waterfall charts are commonly used for data analysis. Explain how these charts add value in those contexts and what insights they help reveal.
6. Discuss the best practices for creating effective box and whisker plots, area charts, and waterfall charts. Highlight common challenges or misconceptions that users may encounter when working with these chart types and suggest strategies to overcome them.

7. ANSWERS

Self-Assessment Questions

1. A box and whisker plot is a type of data visualization used to display the distribution of a dataset.
2. In a box and whisker plot, the box represents the interquartile range (IQR) of the data, which contains the middle 50% of the values. It spans from the first quartile (Q1) to the third quartile (Q3).
3. An area chart differs from a line chart in that it fills the area beneath the line, providing a visual representation of cumulative data.
4. The primary use of an area chart is to show the cumulative values or trends of different data series over time or categories.
5. In a waterfall chart, the bars represent the sequential positive and negative changes in a dataset, often used to depict financial data or cumulative effects.
6. The purpose of using a waterfall chart in data visualization is to illustrate how multiple sequential factors contribute to a final outcome, revealing the cumulative impact of each factor.
7. The key advantage of using a box and whisker plot for data analysis is that it provides a concise summary of the dataset's distribution, including information about central tendency and variability.
8. An area chart might be chosen to represent data when you want to emphasize the cumulative trend or total values over time or categories.
9. The central tendency of data is represented in a box and whisker plot by the median, which is the line inside the box.

10. In a waterfall chart, positive values are typically depicted as bars ascending from the baseline, while negative values are shown as bars descending from the baseline.

Terminal Questions

1. Refer Section 3
2. Refer Section 5
3. Refer Section 4
4. Refer Section 3
5. Refer Section 4
6. Refer Section 5