MASTER OF COMPUTER APPLICATIONS

SEMESTER 1

DATA VISUALIZATION

Unit 6

Advanced Plot types in excel

Table of Contents

1. INTRODUCTION

In Chapter 5, we dive into the world of data visualization techniques that offer both clarity and depth to your insights. Here, we explore the power of various visualization tools to enhance your understanding of data. Let's take a sneak peek at what this chapter has in store: Scatter plots are a visual storyteller's dream. We delve into this visualization technique, which allows you to uncover relationships, patterns, and anomalies in your data with just a glance. Discover how to craft scatter plots that transform data points into meaningful narratives. Sometimes, numbers alone don't do justice to your data. In this section, we introduce you to pictogram charts, a creative way to represent data using icons and symbols. Learn how to use visuals to convey complex information effectively. Timelines are an invaluable tool for tracking events, processes, and changes over time. We explore how to construct timelines that tell a compelling story, whether you're analyzing historical events, project progress, or trends. In Chapter 5, you'll acquire the skills to transform raw data into insightful visuals, making your analysis not only more accessible but also more impactful. Join us on this visual journey as we unlock the potential of data through scatter plots, pictogram charts, and timelines. Get ready to see your data in a whole new light!

1.1 Learning objectives

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

❖ Define what a Scatter Plot is and explain its primary purpose in data visualization.

❖ Identify the key components and terminology used in Scatter Plots, such as data points, axes, and correlation.

❖ Describe the significance of using Pictogram Charts as a visual representation of data and recognize situations where they are most effective.

❖ Explain the concept of a Timeline and discuss its relevance in illustrating chronological events, processes, or historical developments.

❖ Apply the principles of Scatter Plots to analyze and interpret a given dataset, making inferences about data relationships, trends, or outliers.

2. SCATTER PLOT

An Excel scatter plot, also known as an XY graph or scatter diagram, is a two-dimensional chart that illustrates the connection between two variables. This type of graph utilizes both the horizontal and vertical axes as numerical value scales to plot data points. Typically, the independent variable is placed on the x-axis, while the dependent variable is represented on the y-axis. The scatter plot presents data as points where the x and y axes intersect, combining them into single data points. The primary objective of a scatter plot is to visually demonstrate the strength of the relationship, or correlation, between the two variables. A stronger correlation is indicated by data points clustering more closely along a straight line.
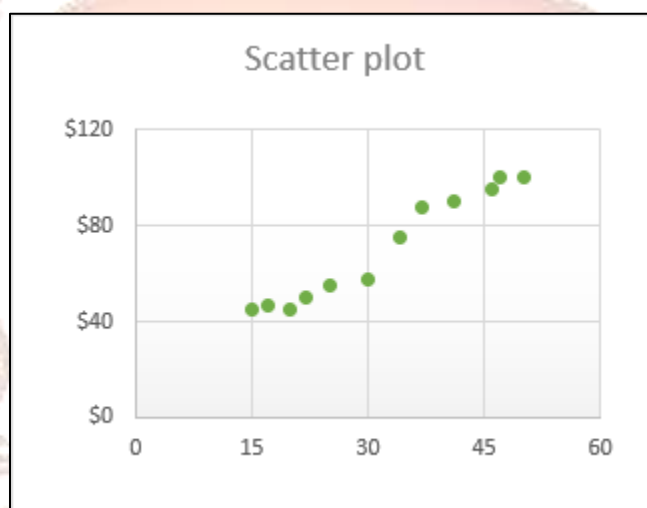

Fig 1: Scatter Plot

How to arrange data for a scatter chart

Excel offers a convenient selection of built-in chart templates, making the creation of a scatter diagram a straightforward task that requires just a few clicks. However, it's essential to structure your source data correctly before proceeding. As previously mentioned, a scatter graph visualizes the relationship between two quantitative variables, necessitating the input of two sets of numerical data into separate columns. To ensure user-friendliness, position the independent variable in the left column since this column will be plotted on the x-axis. On the other hand, place the dependent variable (the one influenced by the independent variable) in the right column, which will be plotted on the y-axis. Additionally, if your dependent variable is listed before the independent variable and you cannot change this in your worksheet, Excel allows you to directly swap the x and y axes within the chart for

convenience.

In our example, we are going to visualize the relationship between the advertising budget for a certain month (independent variable) and the number of items sold (dependent variable), so we arrange the data accordingly:

	Month	Advertising	Items sold
	Jan	$45	15
	Feb	$55	25
	Mar	$47	17
	Apr	$75	34
	May	$90	41
	Jun	$100	47
	Jul	$100	50
	Aug	$95	46
	Sep	$88	37
	Oct	$50	22
	Nov	$45	20
	Dec	$58	30

Fig 2: Dataset

How to create a scatter plot in Excel

When your source data is appropriately structured, creating a scatter plot in Excel involves these two simple steps:

1. Highlight two columns containing numeric data, ensuring you include the column headers. In our example, this corresponds to the range C1:D13. Avoid selecting any additional columns to prevent Excel from becoming confused.
2. Navigate to the Insert tab, located within the Charts group, and click on the Scatter chart icon. From the available options, select your preferred template. To insert a standard scatter graph, simply click on the first thumbnail.
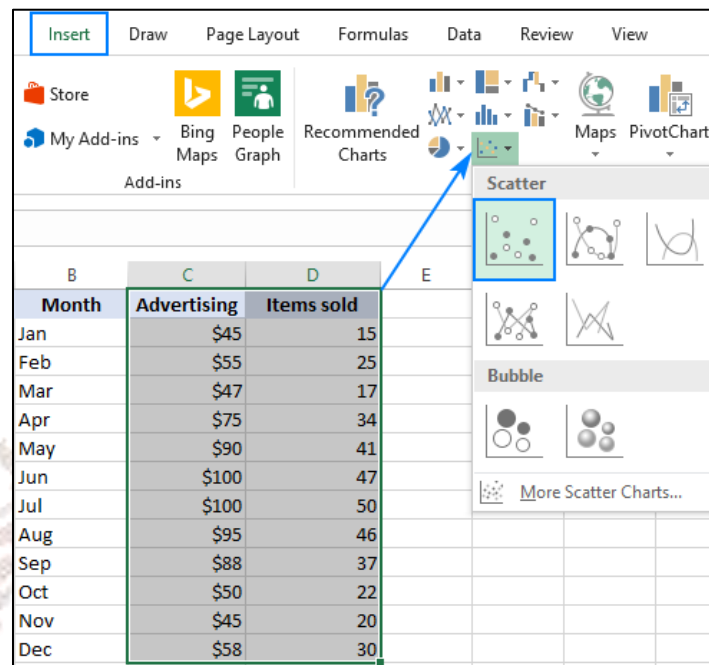
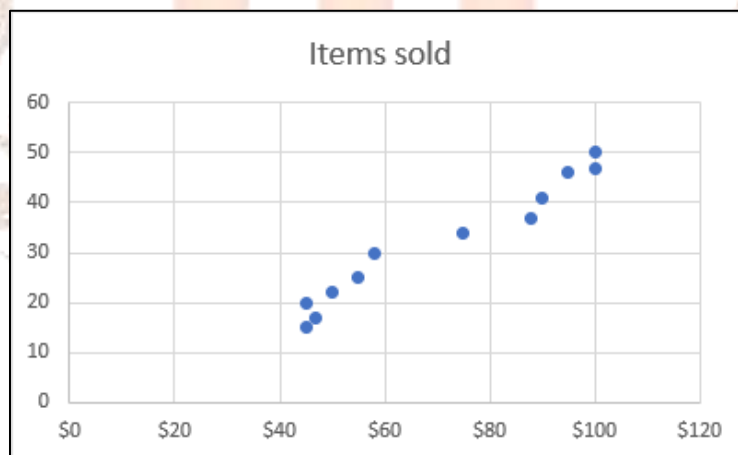**Fig 3:** Select Data and Chart Type

Fig 4: Output Chart

The scatter diagram will be immediately inserted in your worksheet:

Basically, you may consider the work done. Or, you can customize some elements of your graph to make it look more beautiful and to convey the correlation between the two variables clearer.

Scatter chart types

In addition to the standard scatter plot demonstrated in the previous example, there are several other templates at your disposal, including:

- Scatter plot with both smooth lines and markers
- Scatter plot with smooth lines only
- Scatter plot with both straight lines and markers
- Scatter plot with straight lines exclusively

Scatter with lines is best to be used when you have few data points. For example, here's how you can represent the data for the first four months by using the scatter graph with smooth lines and markers:
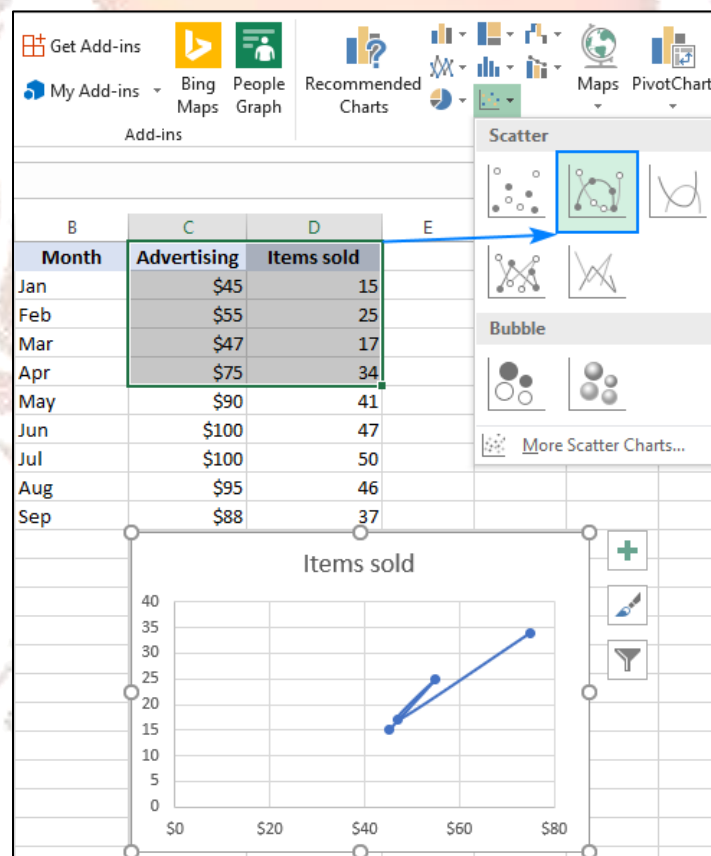
Fig 5: Output Chart: Scatter with Lines

Excel's XY plot templates offer an alternative way to depict the same relationships by plotting each variable individually. To achieve this, you need to choose three columns containing data: the first column should consist of text values (labels), while the other two columns should contain numerical data.

In our example, the blue dots represent advertising cost, and orange dots represent the items sold:
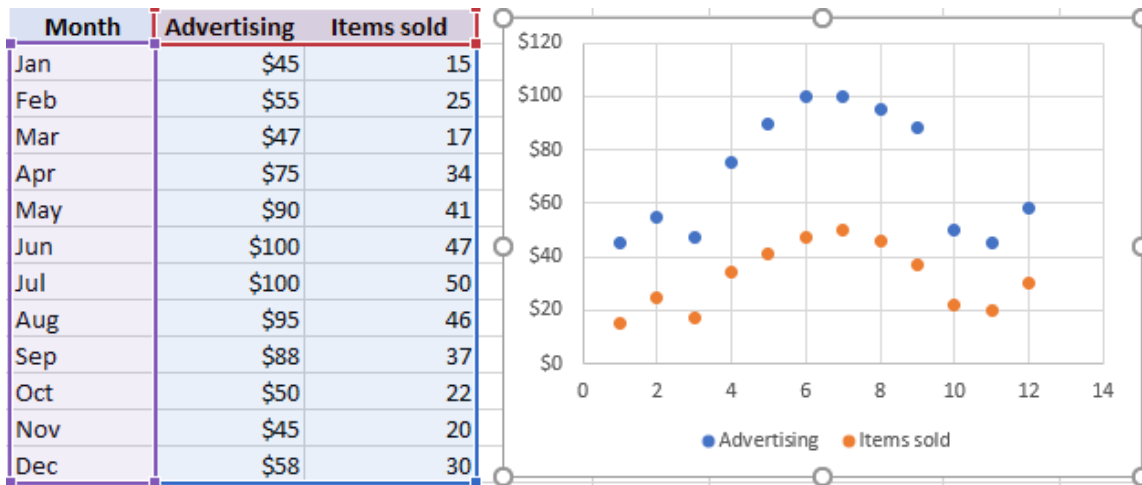
Month	Advertising	Items sold
Jan	$45	15
Feb	$55	25
Mar	$47	17
Apr	$75	34
May	$90	41
Jun	$100	47
Jul	$100	50
Aug	$95	46
Sep	$88	37
Oct	$50	22
Nov	$45	20
Dec	$58	30

Fig 6: Output Chart and Data Selection

To access a comprehensive selection of scatter plot types in one location, begin by selecting your data. Then, go to the ribbon and click on the Scatter (X, Y) icon. Next, click on "More Scatter Charts..." This action will open the Insert Chart dialog box, with the XY (Scatter) type pre-selected. At the top of the dialog box, you can easily switch between various templates to determine which one offers the most suitable visual representation for your data.
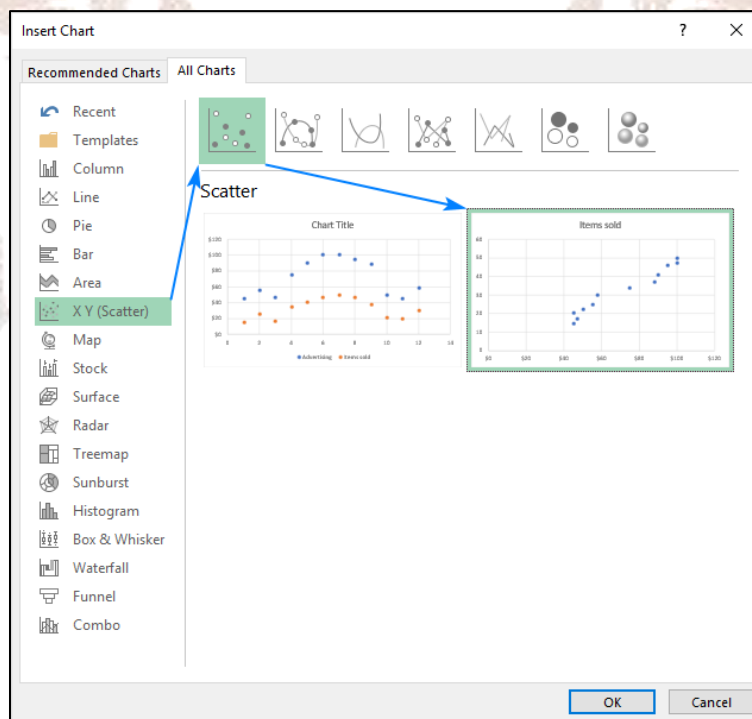
Fig 7: Alternate way of accessing chart

3D scatter plot

In contrast to a traditional XY scatter chart, a 3D scatter plot visualizes data points along three axes (x, y, and z) to illustrate the relationships among three variables, often referred to as an XYZ plot. Unfortunately, Excel, including the latest version Excel 2019, does not provide the capability to create 3D scatter plots. If your data analysis necessitates this chart type, you may want to explore alternative third-party tools like plot.ly, which can generate the kind of 3D scatter graph depicted in the screenshot below.
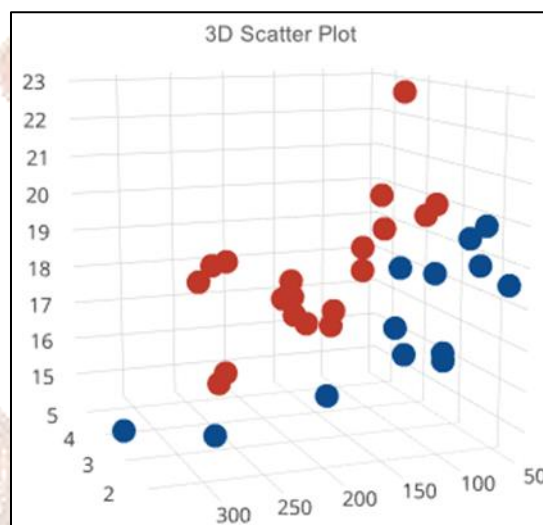
Fig 8: 3D Scatter Plot

Scatter graph and correlation

To correctly interpret a scatter plot, it's essential to grasp how the variables are interconnected. Generally, there are three types of correlations to consider:

1. Positive Correlation - This occurs when an increase in the x variable corresponds to an increase in the y variable. A prime example of a strong positive correlation is the more time students spend studying, the higher their grades tend to be.

2. Negative Correlation - In contrast, negative correlation arises when an increase in the x variable leads to a decrease in the y variable. An instance of this is seen in the negative relationship between class absences and grades; as the number of absences goes up, exam scores tend to decline.

3. No Correlation - When there's no discernible connection between the two variables, the data points scatter randomly across the entire chart area. For instance, there seems to be no correlation between students' height and their grades, as one doesn't significantly impact the other.
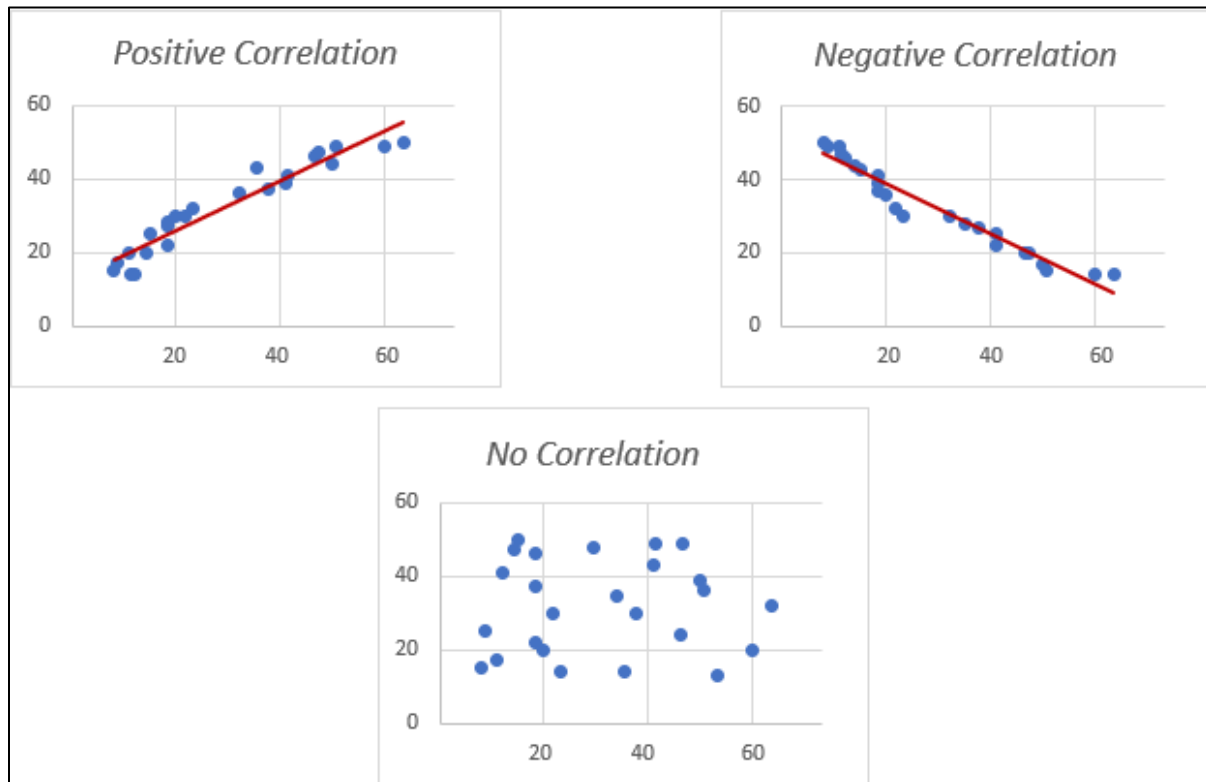

Fig 9: Correlation Analysis Using Scatter Plot

Customizing XY scatter plot in Excel

Similar to other chart types, Excel offers extensive customization options for a scatter plot. You have the flexibility to modify the chart title, incorporate axis titles, hide gridlines, select your preferred chart colors, and more.

Here, we'll focus on specific customizations tailored to scatter plots. One such customization involves adjusting the axis scale to reduce excess white space.

If your data points tend to cluster towards the top, bottom, left, or right edges of the graph, you might want to optimize the chart to eliminate unnecessary white space. To achieve this:

1. Right-click on the x-axis and select "Format Axis."

2. In the Format Axis pane, specify the desired Minimum and Maximum bounds to suit your data.

3. Additionally, you can alter the Major units to control the spacing between the gridlines for a cleaner presentation. The below screenshot shows the settings:
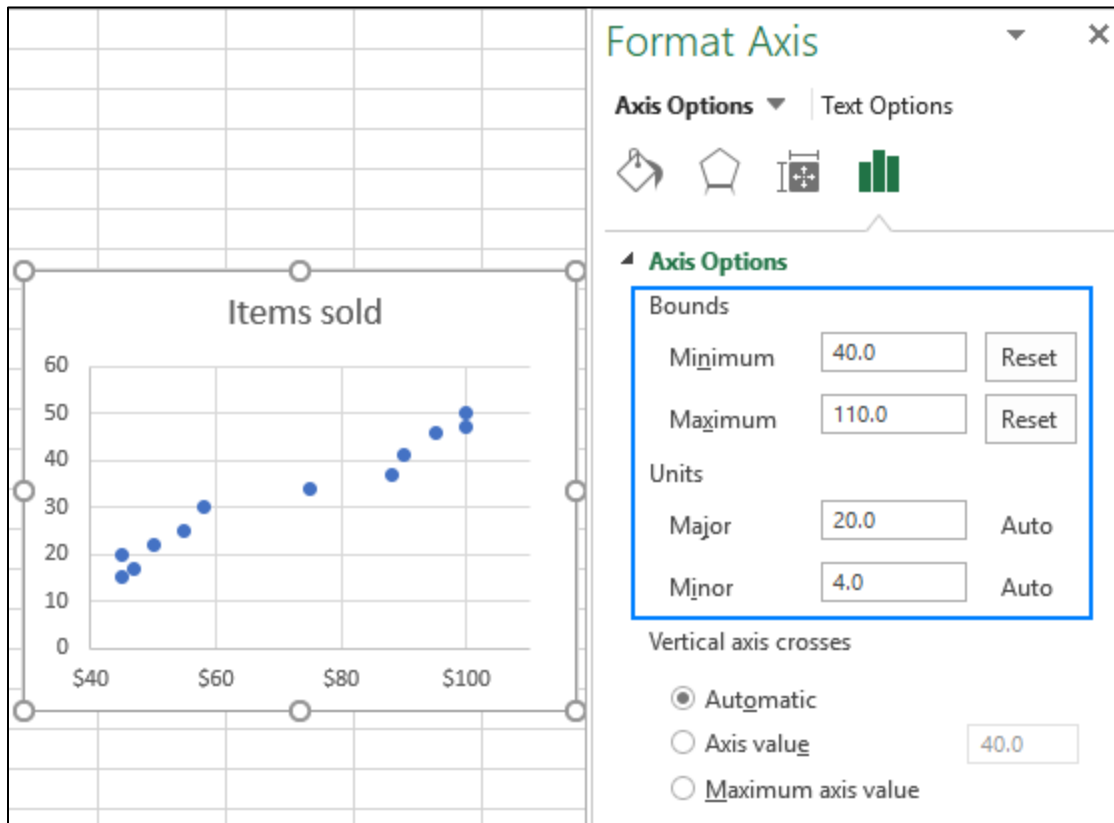

Fig 10: Customised Scatter Plot

To remove space between the data points and the top/bottom edges of the plot area, format the vertical y axis in a similar manner.

Add labels to scatter plot data points

When crafting a scatter plot with a limited number of data points, you might want to assign labels to these points by name to enhance the clarity of your visual representation. Here's a simple process to achieve this:

1. Start by selecting the scatter plot, and then locate the "Chart Elements" button.

2. Check the "Data Labels" box, and then click the small black arrow adjacent to it.

Subsequently, select "More Options…" to access additional customization settings for
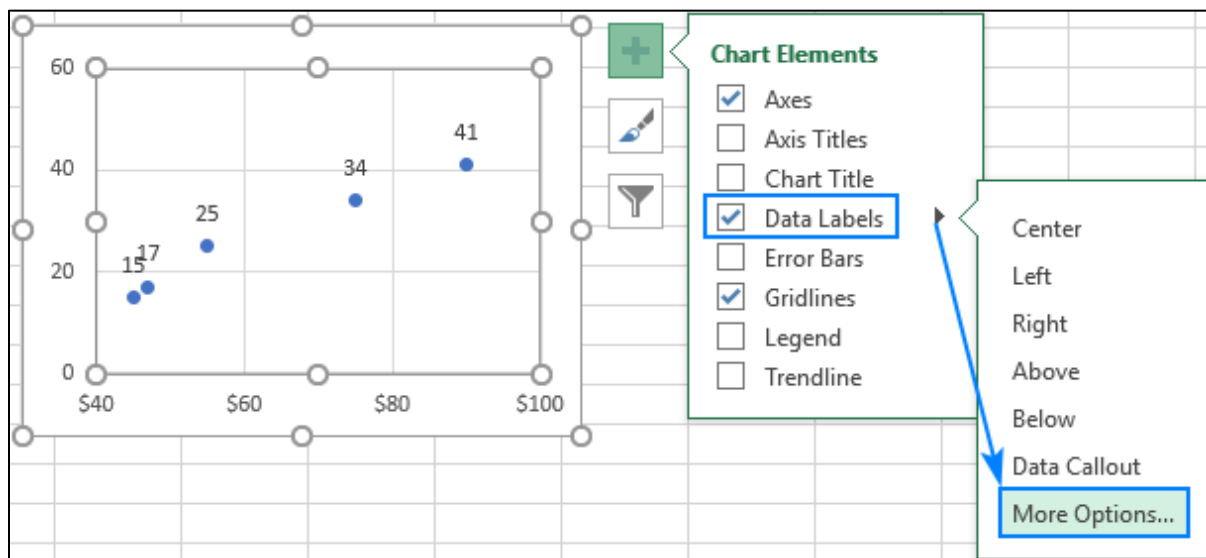
your data labels

Fig 11: Updating Data labels

3. Within the Format Data Labels pane, navigate to the Label Options tab (the final tab), and adjust your data labels using the following settings:
   – Mark the "Value From Cells" box and choose the range from which you wish to extract data labels (for instance, B2:B6 in our scenario).
   – If you prefer to display only the names and exclude the numeric values, uncheck the X Value and/or Y Value boxes to eliminate the numerical data from the labels.
   – Define the position of the labels, which, in our example, is set above the data points.
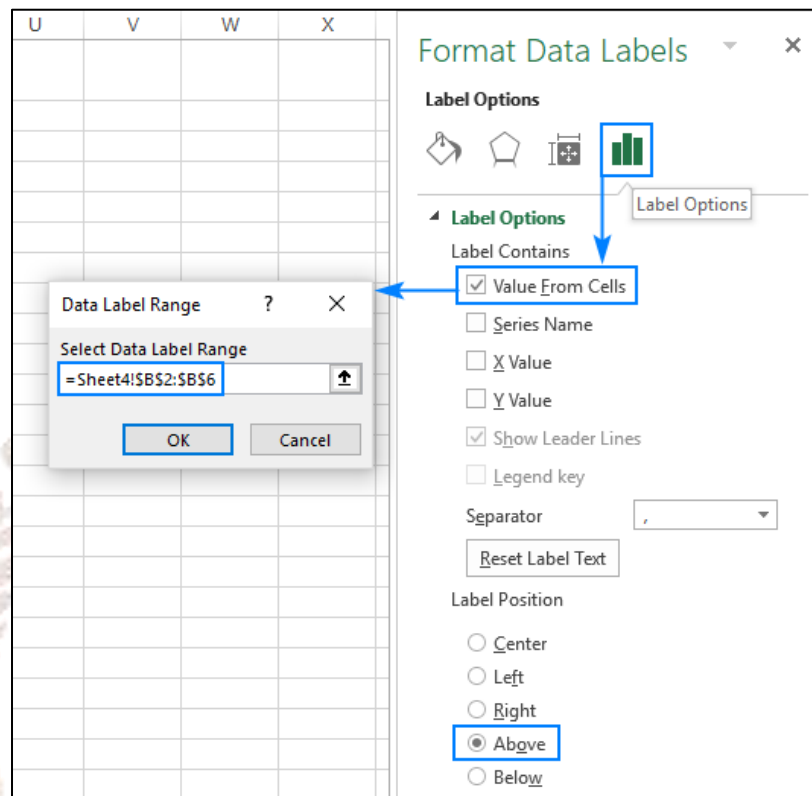
Fig 12: Formatting Data Labels

That's it! All data points in our Excel scatter plot are now labeled by name:
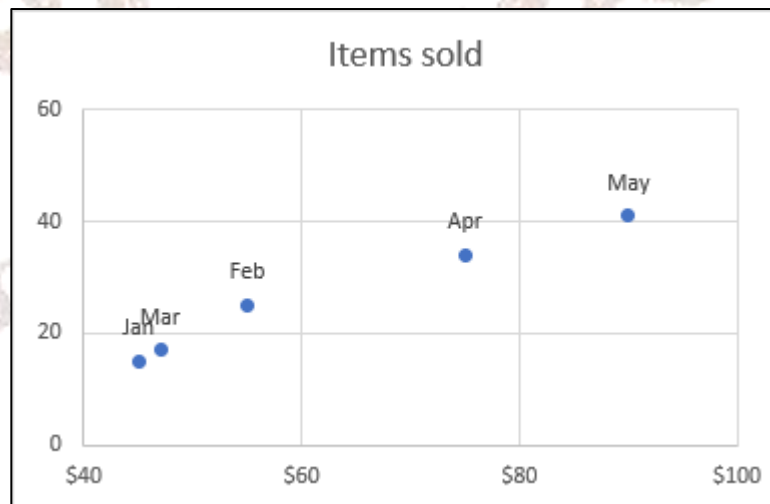
Fig 13: Output Chart 1

Tip: Addressing Label Overlaps

In situations where two or more data points are in close proximity, you might encounter label overlaps, similar to the scenario observed with the Jan and Mar labels in our scatter plot. To resolve this, select the labels and choose the one that's overlapping. Click on the overlapping label to select it exclusively. Then, position your mouse cursor over the selected label until it transforms into a four-sided arrow, allowing you to drag the label to your preferred location.

As the result, you will have a nice Excel scatter plot with perfectly legible labels:
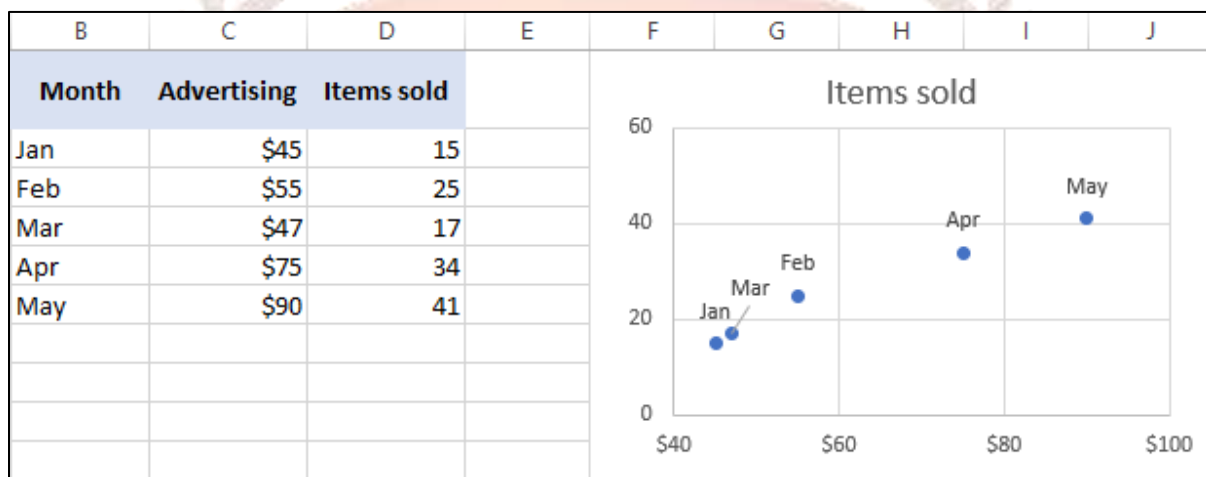
Fig 14: output Chart 2

Add a trendline and equation

To better visualize the relationship between the two variables, you can draw a trendline in your Excel scatter graph, also called a *line of best fit*.

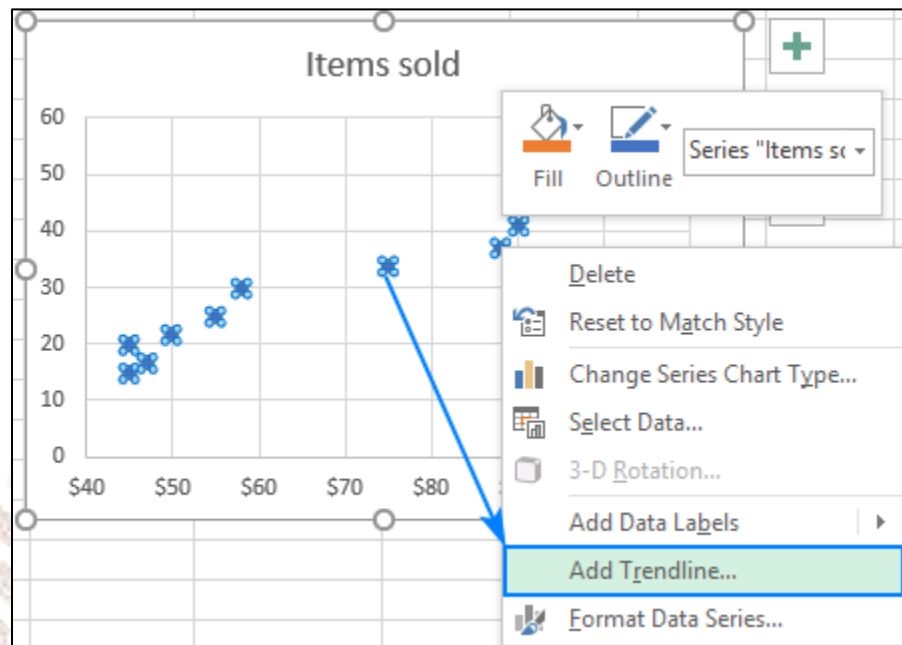To have it done, right click on any data point and choose **Add Trendline...** from the context menu.

Fig 15: Adding trendline

Excel will draw a line as close as possible to all data points so that there are as many points above the line as below.

Additionally, you can show the **equation for the trendline** that mathematically describes the relationship between the two variables. For this, check the *Display Equation on Chart* box on   the *Format Trendline* pane that should appear in the right part of your Excel window immediately after you've added a trendline. The result of these manipulations will look similar to this:
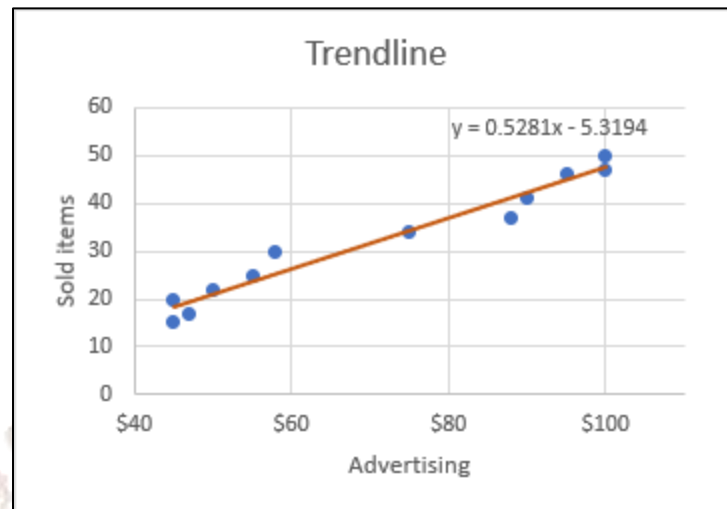
Fig 16: Output Chart with trendline

What's depicted in the screenshot above is commonly referred to as a linear regression chart, and you can find detailed instructions on how to create it here: "How to Make a Linear Regression Graph in Excel."

If you find that the scatter plot displays the independent variable on the vertical axis and the dependent variable on the horizontal axis, which is different from the standard arrangement, the easiest solution is to rearrange the source columns in your worksheet and then create the chart again.

However, if, for any reason, you cannot rearrange the columns in your worksheet, you have the option to directly swap the X and Y data series within the chart itself. Here's how to do it:Right-click any axis and click *Select Data…* in the context menu.
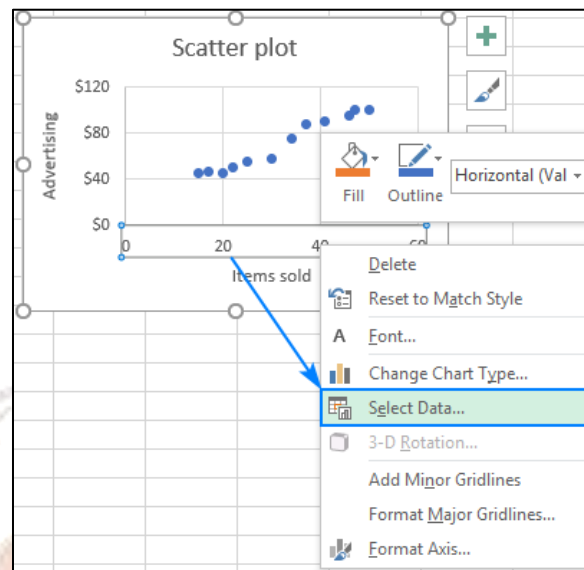
Fig 17: Data Updation

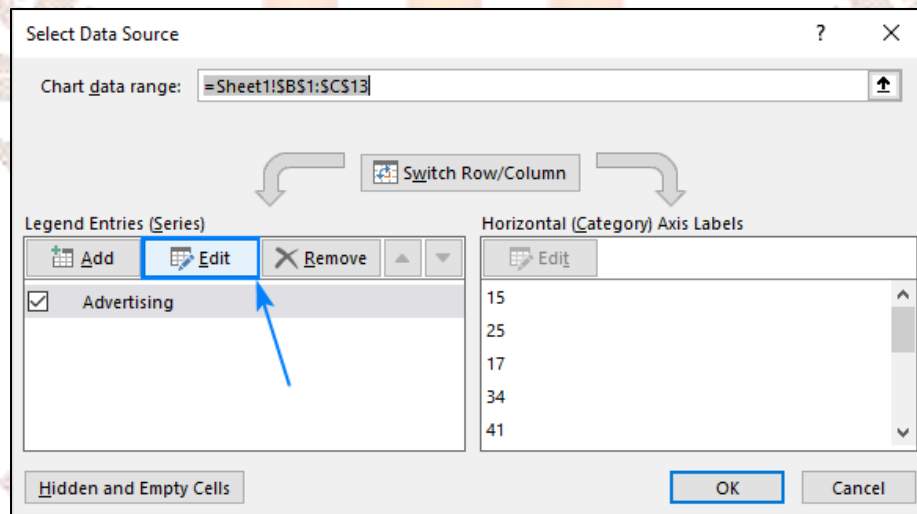1. In the *Select Data Source* dialog window, click the *Edit* button.

Fig 18: Data Swapping

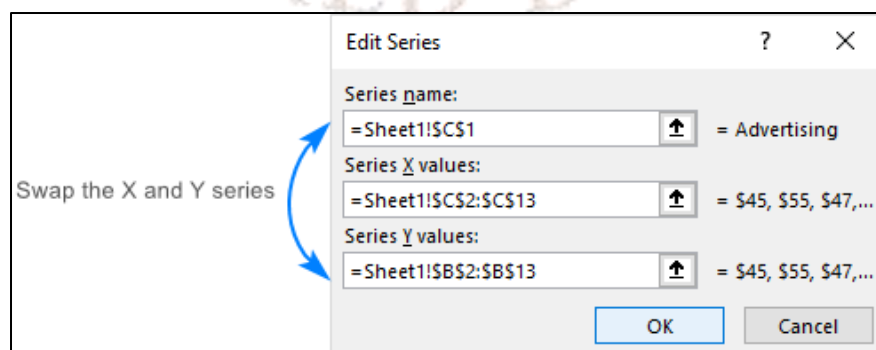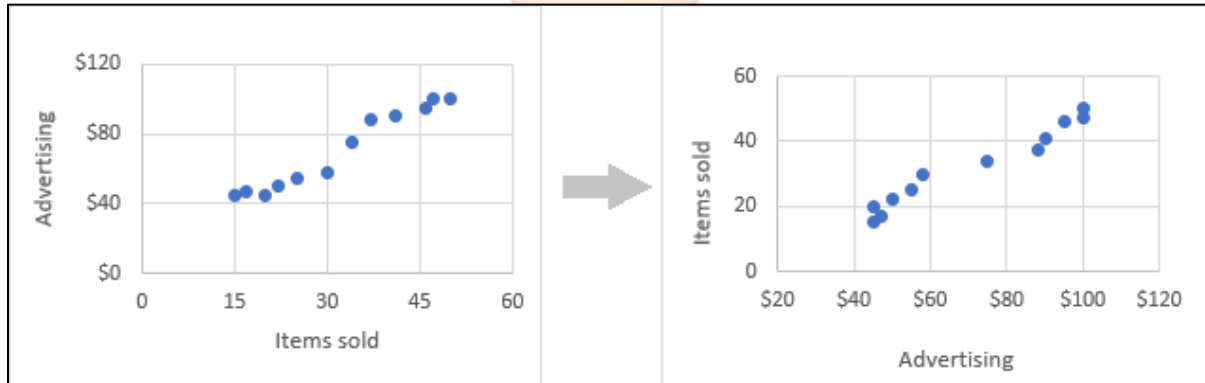2. Copy *Series X values* to the *Series Y values* box and vice versa.

Fig 19: Data After it is swapped

Tip. To safely edit the contents of the Series boxes, put the mouse pointer in the box, and press F2.

3. Click *OK* twice to close both windows.

As the result, your Excel scatter plot will undergo this transformation:

Fig 20: Updated Output Chart

3. PICTOGRAM CHART

A pictograph is a graphical representation consisting of pictorial symbols. Typically used in mathematics, it employs graphs with images or icons to convey specific quantities or numbers, such as people or books. It's alternatively referred to as a pictogram, pictogramme, pictorial chart, picture graph, or simply picto. Pictographs are commonly employed to depict statistical data, simplifying comprehension and analysis through visual means.

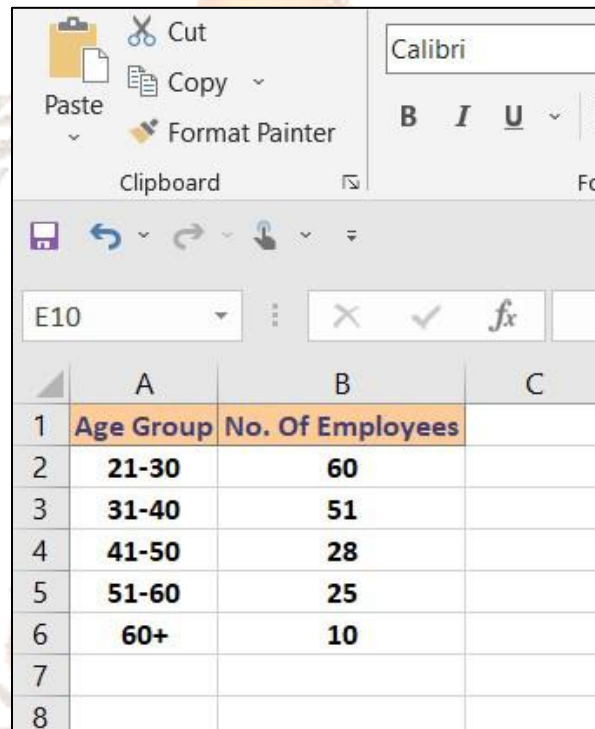**Advantages of using pictographs include:**
1. Facilitating the association of numbers and objects, especially for children.
2. Making statistical data visually appealing and easy to comprehend.
3. Enhancing the visual presentation of statistics.
4. Effectively representing large volumes of data.

In this instance, we'll utilize Excel to create a pictograph illustrating the number of employees categorized by age groups.

Step By Step Implementation to Create A Pictograph In Excel

Step 1: Create A Dataset

In this step, we will create a dataset for our graph. We will use two columns with the titles "**Age Group**" and "**Number Of Employees**" and insert some random data.
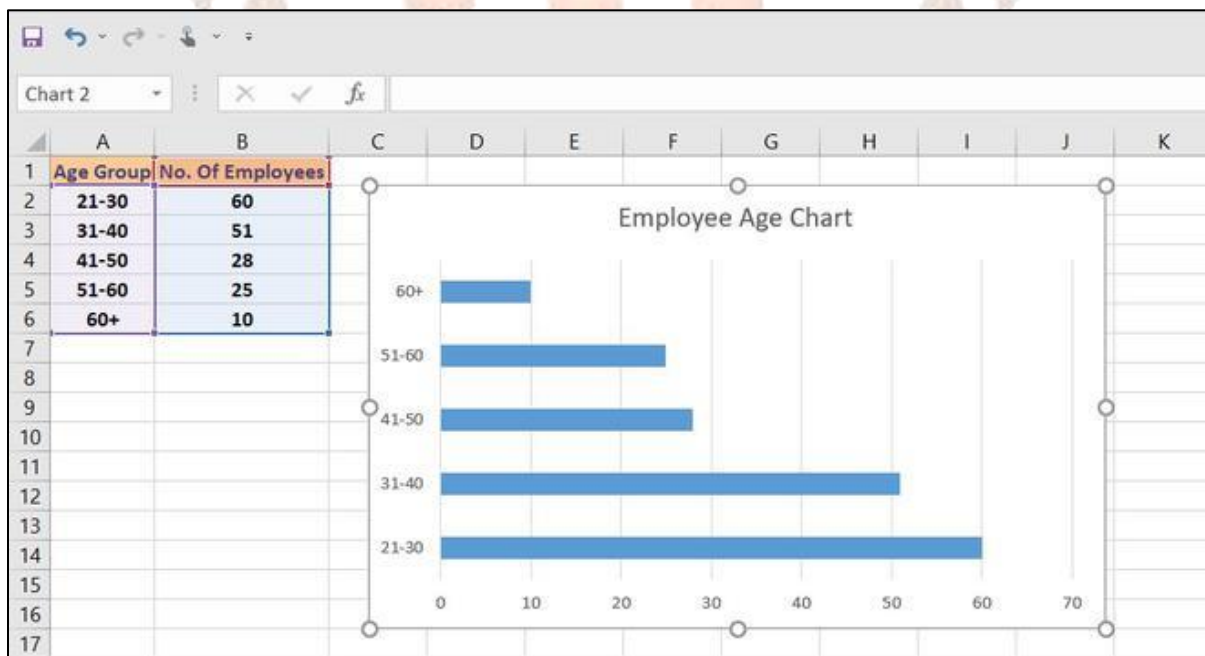

Fig 21: Dataset

Step 2: Insert Chart

In this step, we will create our chart using the above dataset. In order to insert a chart, we need to select out data and then go to,

Insert > Charts > Insert Bar Chart.
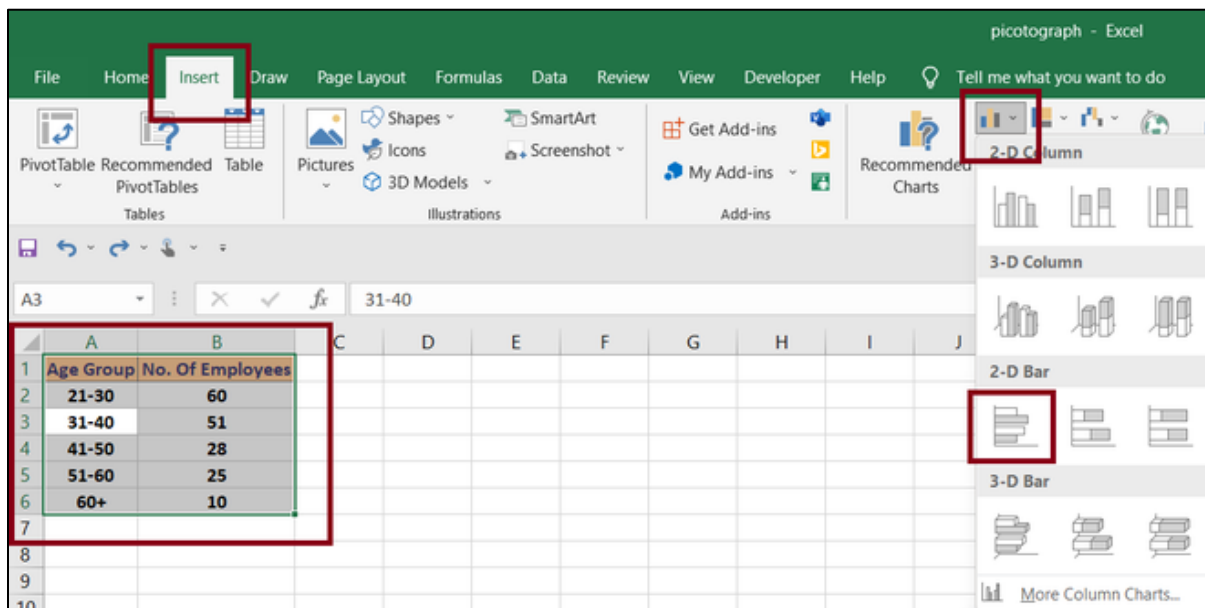
Step 3: Format Data Series

In this step, we will format our data series. i.e., the data bars. For this,

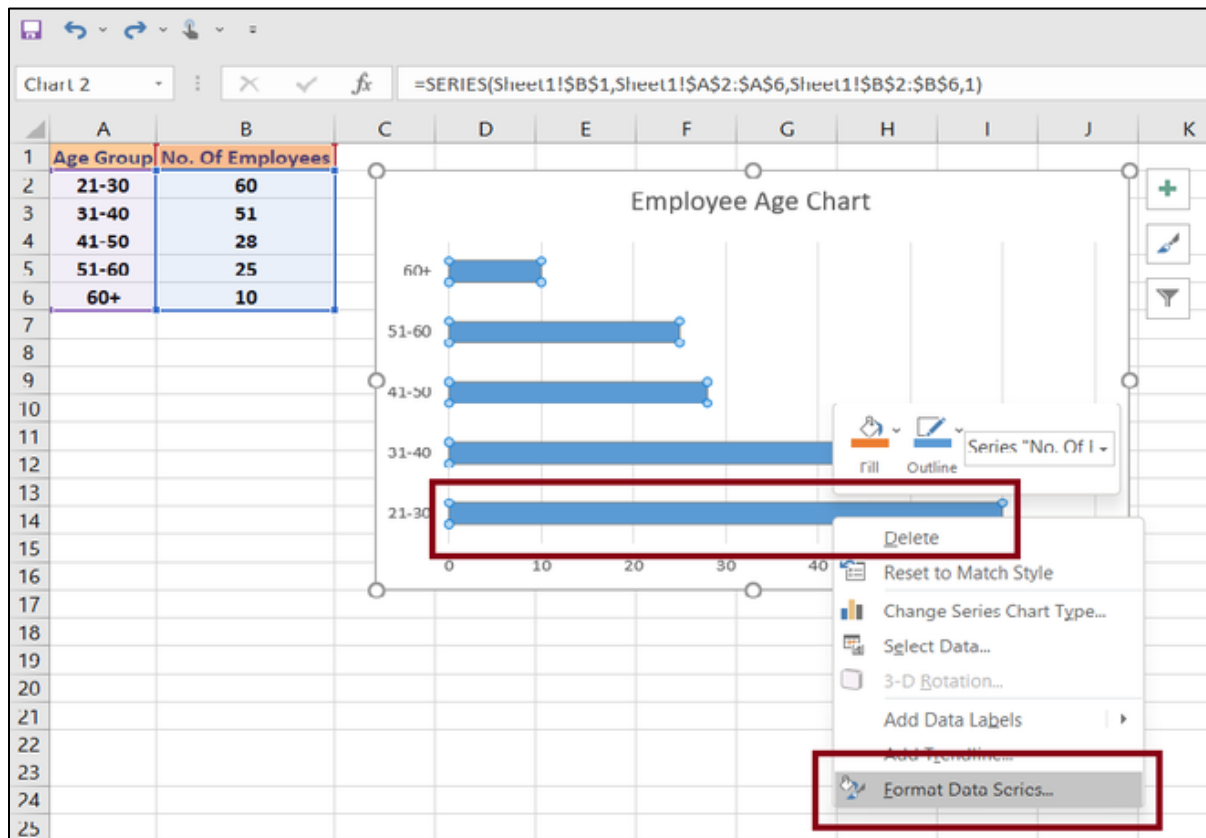***Select Bar > Right-Click > Format Data Series***.

Fig 24: Format Data Series

Upon selecting "Format Data Series," Excel will open a side pane called "Format Data Series," allowing us to customize the appearance of our data bars to suit our specific needs.

Fig 25: Output Screen
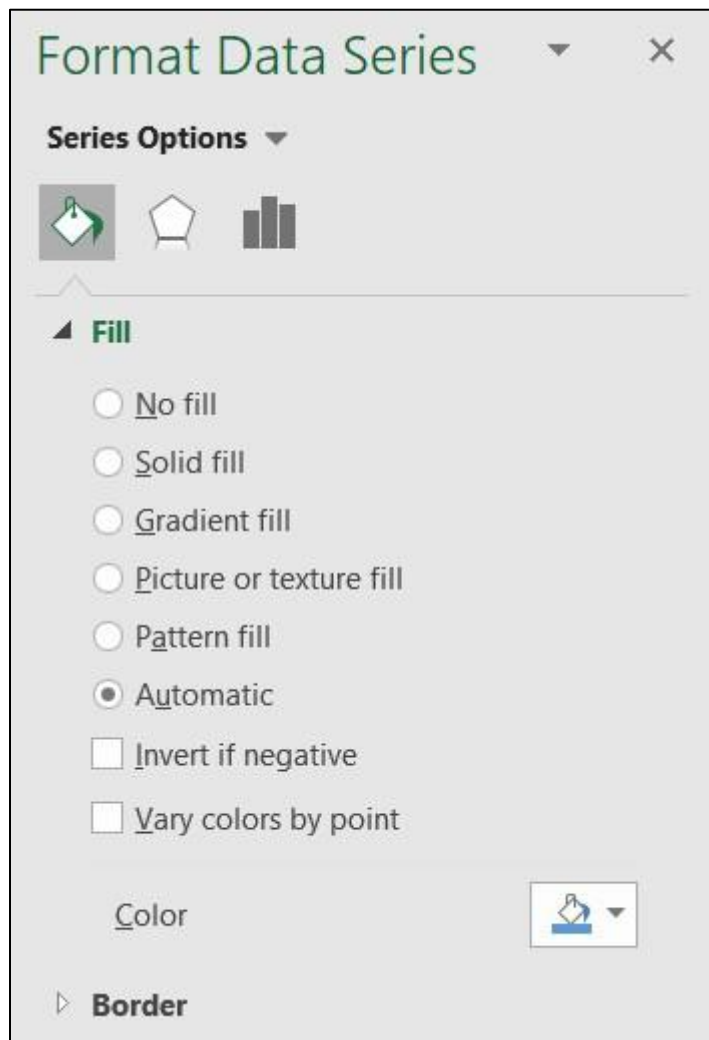
Step 4: Insert Picture In Data Bars

In this step, we will insert pictures in our data bars of the graph. For this, we need to go to,

Fill > Picture or texture fill.
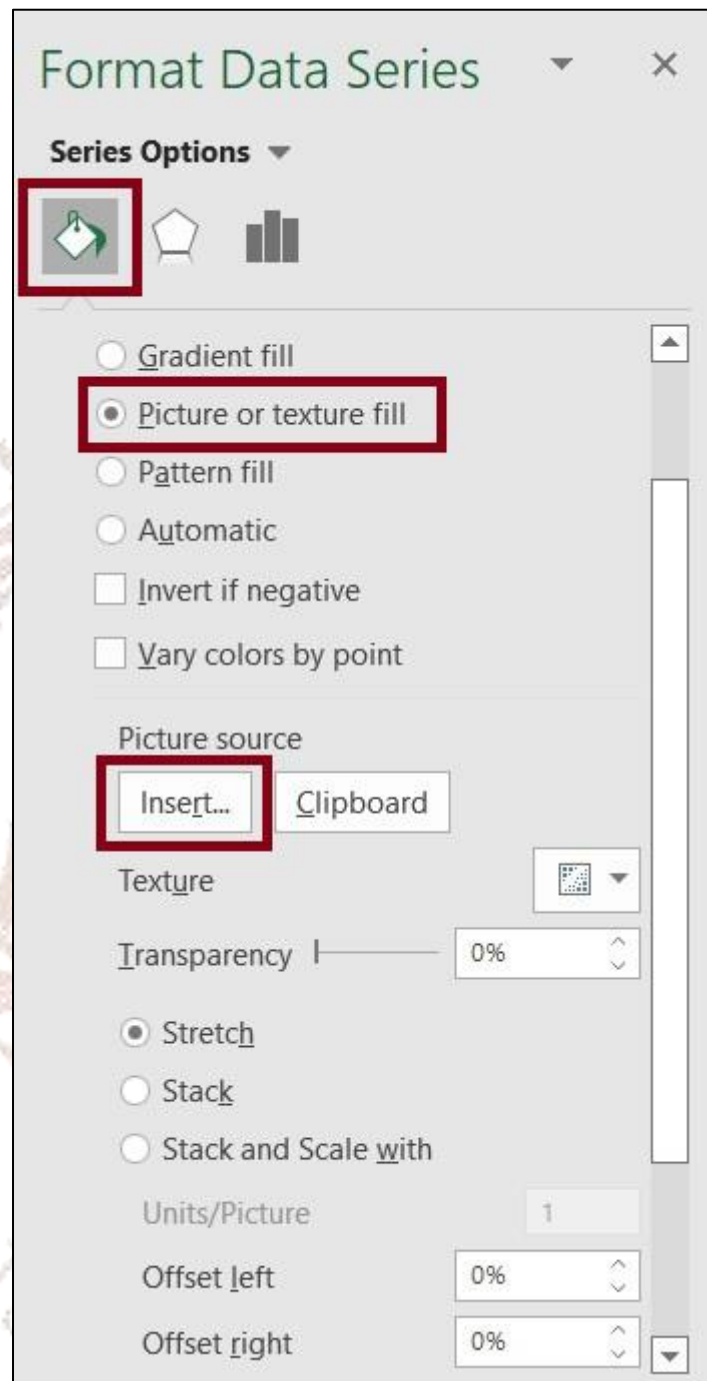
Fig 26: Update the selection to Picture Fill

Once we click on the Insert button, the excel will ask us to insert an icon(Here, we are using default icons provided by excel. You can use your own icons or picture).
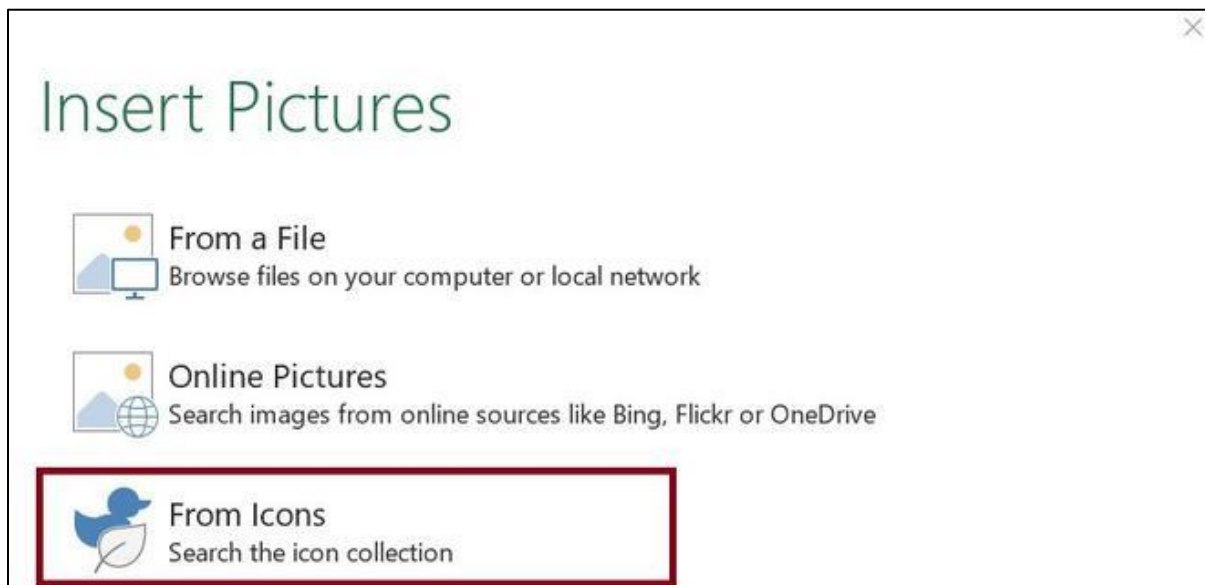
Fig 27: Select Icon

Here, we are using Employee icons to fill in the data bars.
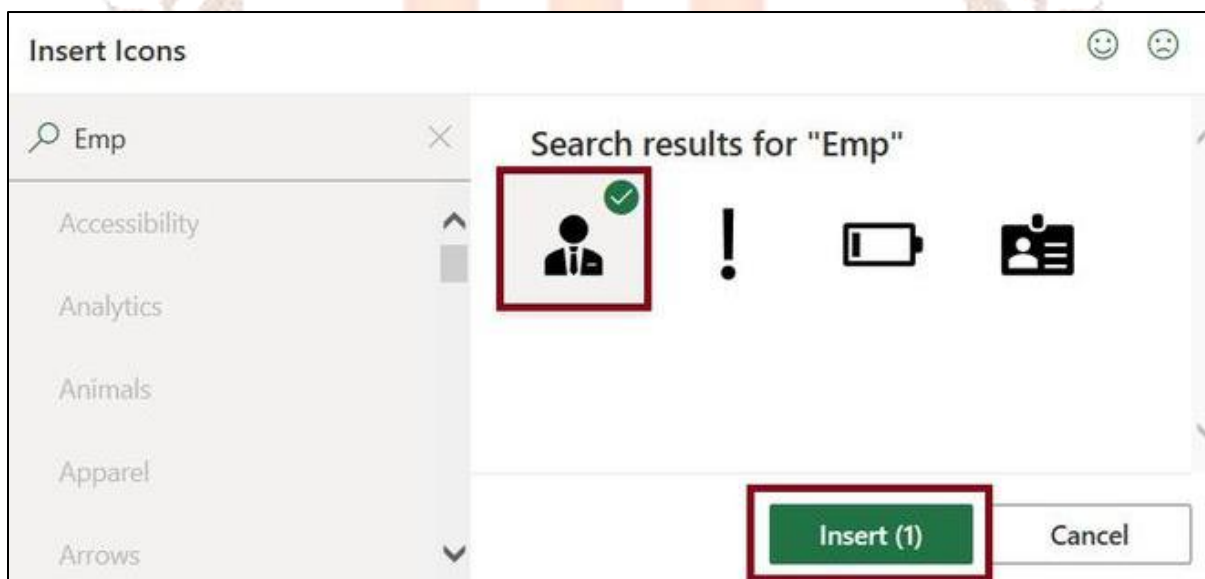
Fig 28: Select the Icon of Choice

Once we Insert the employee icon from the default icons provided by excel, we need to use some validations regarding the inserted picture or icons.

Fig 29: Select how the Icons will be represented on the Chart

1. **Stretch:** This will stretch one single image incomplete data bar.

2. **Stack:** This will represent the stacked images in the data bar.

3. **Stack and Scale with:** This will show the images depending on the data count.

Step 5: Output

In this step, we will see the complete pictograph of our dataset.
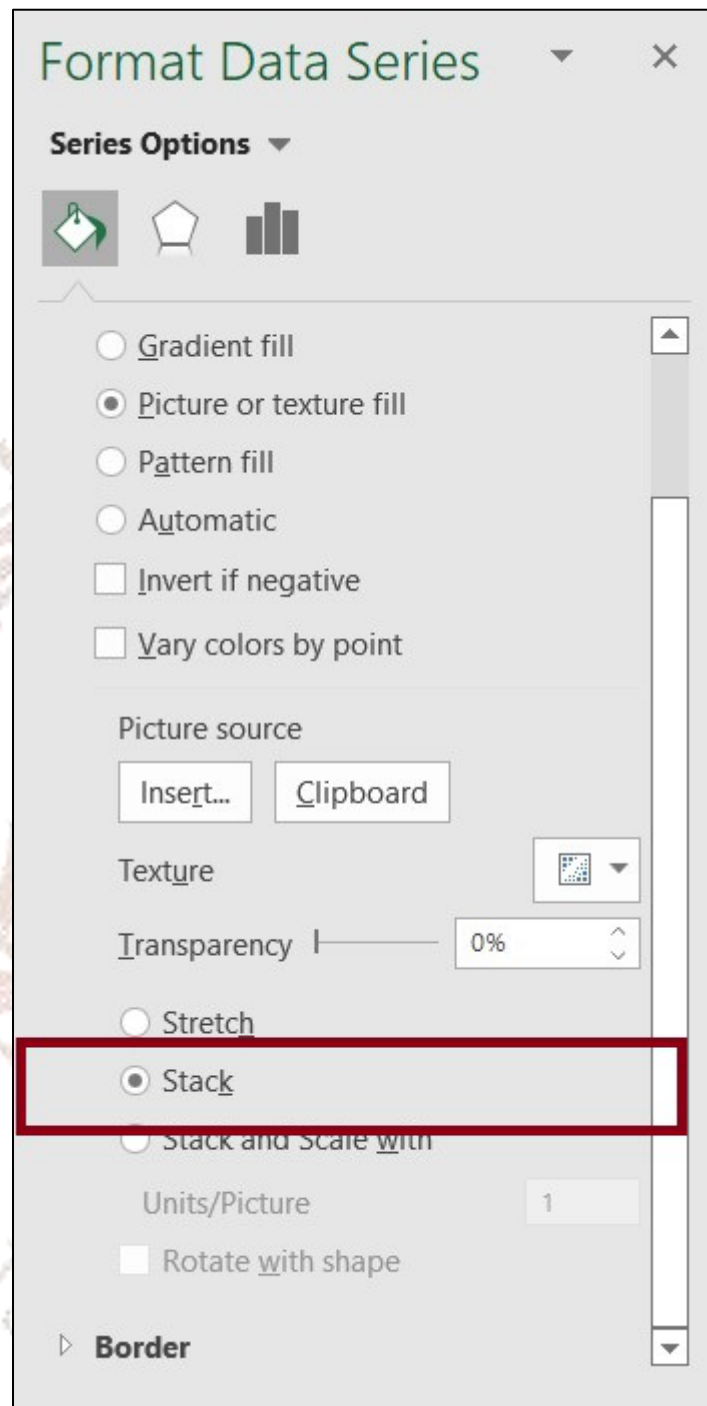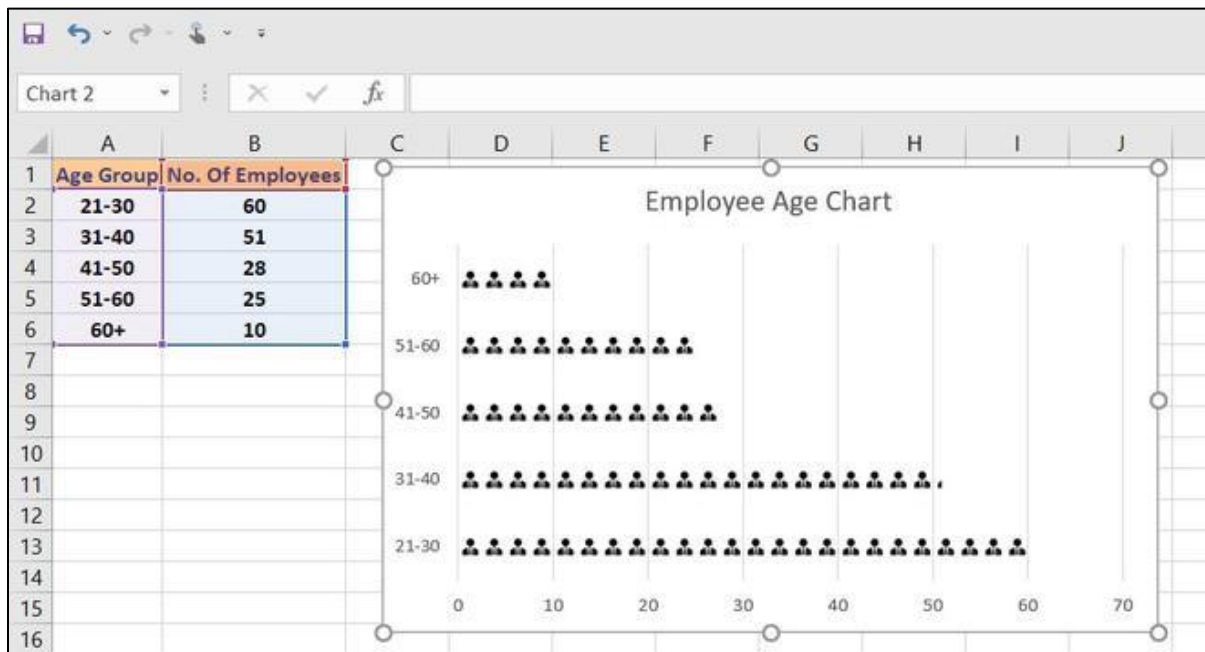

Fig 30: Output Chart

4. TIMELINE

A timeline chart, often simply referred to as a timeline, is a graphical representation of events or activities displayed chronologically along a linear axis, typically from left to right. Timelines are used to visualize the progression of events over time and provide a clear, chronological overview of historical, project-related, or other sequential data.

In a timeline chart:

1. Events or activities are represented as points or bars along the timeline, with their positions indicating their chronological order.
2. Time is typically displayed on the horizontal axis, allowing viewers to understand when each event occurred.
3. Timelines can vary in scale, representing short-term or long-term durations, depending on the context.
4. They are frequently used in historical research, project management, education, and various other fields to help people understand and analyze the sequence and duration of events or activities.

Timeline charts come in various forms, from simple linear timelines to more complex Gantt charts, which display events or tasks along a timeline and can include additional information such as task dependencies and durations.

Creating a Timeline chart:

To begin, create a task list to determine the content you want to display on the timeline. If your goal is to showcase milestones, consider seeking an Excel timeline template tailored for milestone data input. Alternatively, if you aim to illustrate various aspects of a specific project's timeline, look for an Excel project timeline template. These templates offer greater customization with additional fields, providing more detailed information about tasks and their durations.

Microsoft also provides Excel timeline templates that offer a comprehensive view of conference planning timelines. Unlike Gantt chart-based timelines, these Excel timelines require manual input of your data into predefined template fields. These templates are flexible, allowing you to modify names and add field as needed.

1. To locate a Microsoft Excel timeline template, open Excel, enter "Timeline" into the search box, and press Enter. Please note that this template was found using the latest version of Excel on Windows 8.
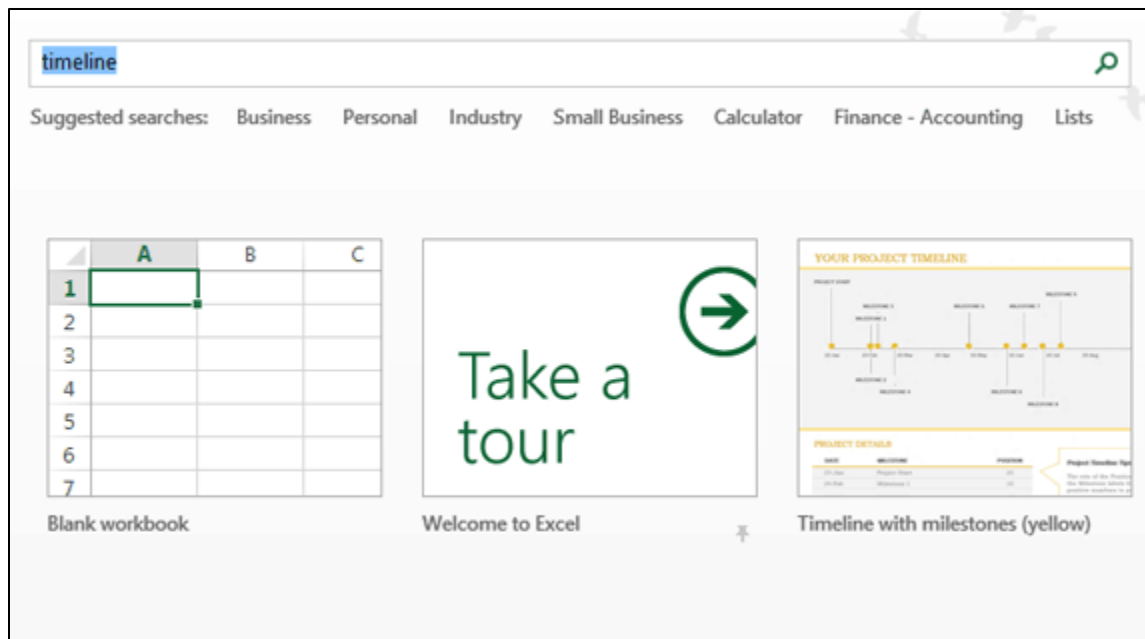

Fig 31: Search for Timeline Template Chart

2. Double-click on the Excel Project Timeline template to open the spreadsheet.
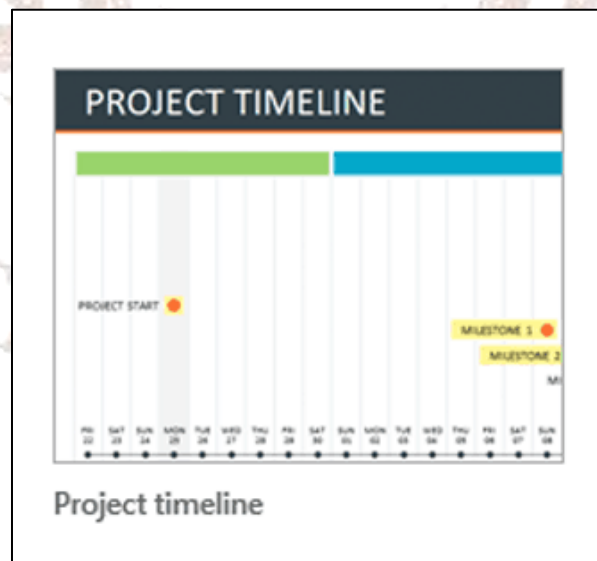

Fig 32: Select the Timeline template

Add Your Information to the Timeline in Excel

Once you open the template, you'll find an Excel spreadsheet with pre-filled information in the fields. Please note that this information is merely a placeholder. Towards the top of the template, there's a timeline displayed. As you scroll down, you'll encounter a preformatted chart where you can input details related to your conference planning along with their respective due dates. Utilizing an Excel project timeline template offers the advantage of pre-set formatting, requiring only customization on your part.

1. To begin, click on the "Project Timeline" field located at 1C in the spreadsheet and enter the name of your conference.
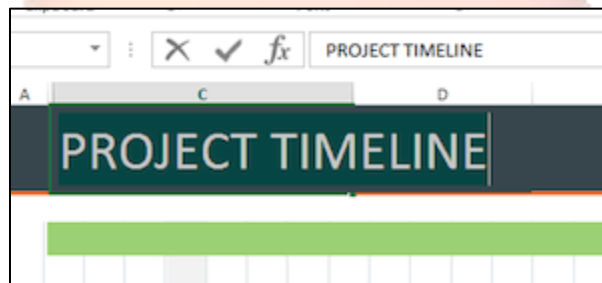
Fig 33: Update the title of the chart

2. Scroll down in the spreadsheet and enter a start date.
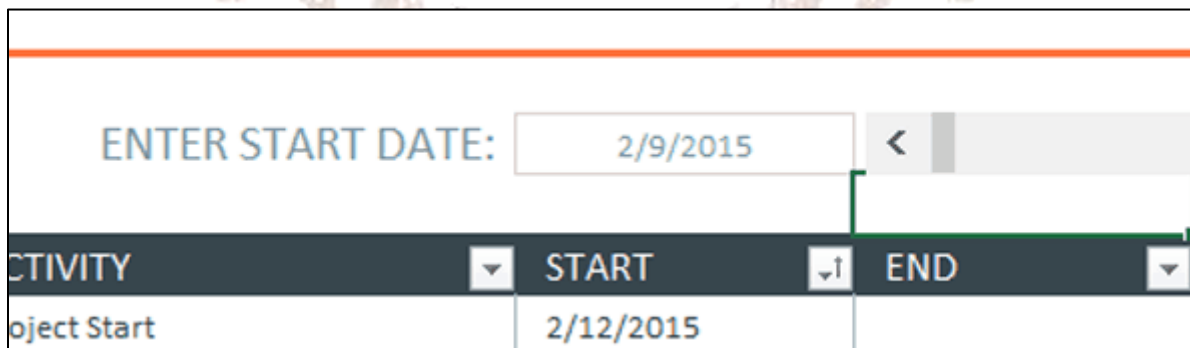
Fig 34: Enter the start date

Given that you're in the process of planning a conference, it's essential to select the kick-off date for your planning activities. Please be aware that there's an existing formula that automatically sets the start date as the day you initiated the use of the event planning template. If you prefer to use a different date, click on the respective cell, remove the formula,

and input your desired date. You'll observe that the preconfigured start and end dates will adjust accordingly.

3. Next, input the initial significant task that needs to be completed. To add tasks to the "Activity" column, simply double-click on the field labeled "Milestone."


Fig 35: Update Activity Column

To proceed, press the Tab key to move to the corresponding "Start" field, where you should enter the date when you intend to begin your research on potential conference venues. Once done, press the Tab key once more to access the "End" field, where you should input the date by which you aim to have selected the venue.

Fig 36: Update Relevant columns

5. Repeat steps 3 and 4 to complete the remainder of the chart.

Customize the Excel Timeline

After you've input all the important conference milestones into the chart, you have the flexibility to modify the timeline's appearance to your liking. This includes altering how the timeline data is presented and adding color for visual clarity.

If your conference planning timeline spans more than a month (which is quite likely), you can access additional data on the timeline by clicking the arrows located in the gray bar adjacent to the "Start date" box. This action allows you to navigate through the Excel timeline and view more information as needed.
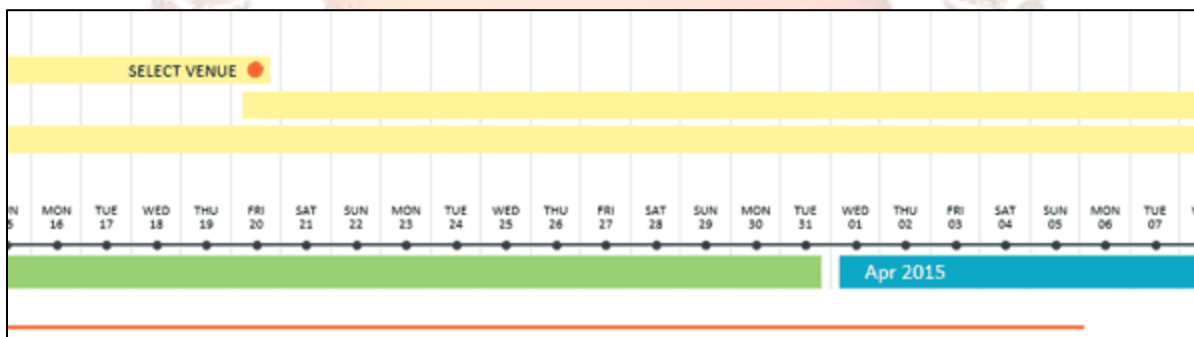
Fig 37: Timeline Chart

1. To change the overall chart presentation, click on the chart and then click on the box with a paintbrush icon.
2. A pop-up box will emerge, presenting various styles for the timeline chart. Hover your cursor over these formats to preview how they will appear on the timeline. If you find one that suits your preference, simply click on it. The timeline will then be modified to adopt the chosen style.

To modify the color palette of the Excel timeline, follow these steps:
1. Select the chart.
2. Click on the paintbrush icon and then select "Color" at the top of the displayed pop-up box.

3. Hover your cursor over the various timeline colors to preview how they will look on the timeline. If you find a color scheme that appeals to you, click on it. This action will update the timeline to incorporate the selected style.
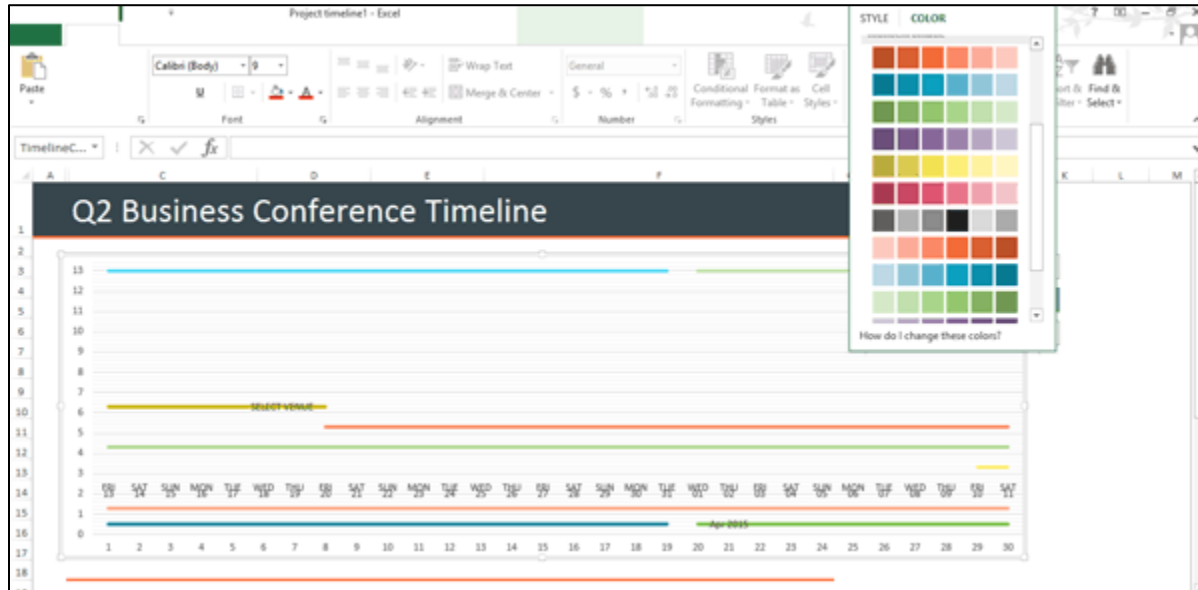
Fig 38: Update Colour Pattern

While this timeline template serves as an excellent tool for sharing essential information with stakeholders and executives, offering them a broad overview of the conference's necessary tasks, it does have limitations. Notably, it lacks elements such as a budget breakdown and the ability to track the timeliness of task completion or assign responsibility for each task. If you require a more comprehensive conference planning timeline with these features, you might want to explore the option of creating a Gantt chart in Excel.

5. SUMMARY

The chapter encompasses a diverse set of data visualization tools, including timeline charts, scatter plots, and pictogram charts, each serving distinct purposes in representing data. Timeline charts offer a chronological view of events or tasks over time, aiding project management, historical analysis, and event planning. Scatter plots provide a two-dimensional display of data points to reveal relationships and correlations between variables. Pictogram charts employ pictorial symbols to convey quantitative information effectively. Together, these visualization techniques offer a comprehensive toolkit for understanding, analyzing, and presenting data across various domains and scenarios, from project management and historical research to data-driven storytelling and educational contexts.

6. QUESTIONS

SELF-ASSESSMENT QUESTIONS – 1

1. What is the primary purpose of a timeline chart in Excel?
2. How do you create a timeline chart in Excel?
3. What type of data is typically represented using a scatter plot in Excel?
4. How can you determine the correlation between two variables in a scatter plot?
5. What are the advantages of using a pictogram chart in Excel for data representation?
6. Describe the key characteristics of a timeline chart in Excel.
7. How can you customize the appearance of data points in a scatter plot?
8. In Excel, what is the significance of the x-axis and y-axis in a scatter plot?
9. What does the size of pictorial symbols in a pictogram chart represent?
10. What are the steps to switch the X and Y axes in a scatter chart in Excel?

Terminal Questions

1. Explain the step-by-step process of creating a timeline chart in Excel. Include details on how to input data, customize the chart's appearance, and add labels for better clarity. Provide an example scenario where a timeline chart would be particularly useful in project management or data presentation.

2. Discuss the advantages and disadvantages of using a timeline chart over other visualizations, such as Gantt charts or bar graphs, when displaying chronological data. Provide specific examples of situations where a timeline chart excels in conveying information effectively.

3. Define and illustrate the concept of correlation in the context of a scatter plot. Explain the differences between positive, negative, and zero correlations, using graphical representations and real-life examples to aid your explanation. Additionally, discuss the implications of strong versus weak correlations.

4. When dealing with a dataset for which you intend to create a scatter plot, what considerations should you keep in mind regarding data preprocessing and data cleaning? Describe potential challenges related to outliers and missing values and propose strategies for addressing these issues to ensure the accuracy of your scatter plot.

5. Provide a comprehensive guide to creating a pictogram chart in Excel, covering the selection of appropriate symbols, resizing and arranging them, and assigning numerical values. Explain how the use of pictogram charts enhances data visualization and improves the audience's understanding of complex data sets, giving examples.

6. Compare and contrast pictogram charts with bar graphs when representing categorical data in Excel. Highlight the strengths and weaknesses of each visualization type, and justify why you might choose one over the other for different types of data presentations. Provide real-world scenarios as examples.

7. ANSWERS

Self-Assessment Questions

1. The primary purpose of a timeline chart in Excel is to visualize chronological data and events over time.

2. To create a timeline chart in Excel, you can use various methods, including inserting a SmartArt graphic or manually creating a table and customizing it.

3. Scatter plots in Excel typically represent data with two quantitative variables, showing how they relate to each other.

4. To determine the correlation between two variables in a scatter plot, you can observe the pattern of data points. The tighter the points cluster along a line, the higher the correlation.

5. The advantages of using a pictogram chart in Excel include making data more visually appealing, aiding in understanding, and representing quantities using images.

6. Key characteristics of a timeline chart in Excel include the representation of time-based data, events, or tasks along a horizontal axis, often with labels and markers.

7. You can customize the appearance of data points in a scatter plot by changing their size, shape, color, or adding data labels.

8. In Excel, the x-axis represents the independent variable, while the y-axis represents the dependent variable in a scatter plot.

9. The size of pictorial symbols in a pictogram chart typically represents the quantity or frequency of data points they represent.

10. To switch the X and Y axes in a scatter chart in Excel, you can rearrange the data source columns in your worksheet or use Excel's chart formatting options.

Terminal Questions

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