MASTER OF COMPUTER APPLICATIONS

SEMESTER 1

DATA VISUALIZATION

Unit 4

Advanced Chart Types in Excel

Table of Contents

1. INTRODUCTION

In the ever-evolving landscape of data analysis and visualization, Microsoft Excel stands as a cornerstone tool for professionals and analysts across various industries. While many users are acquainted with its fundamental charting functions, this chapter acts as a gateway to a more intricate and powerful realm of data representation. Welcome to the world of advanced chart types, where we embark on a journey to explore the rich capabilities of histograms, Gantt charts, and heat maps.

Histograms offer a dynamic way to comprehend data distributions. They provide a visual summary of the frequency and distribution of data points, making it easier to identify patterns, peaks, and outliers within your dataset. Histograms are invaluable for tasks like understanding the distribution of test scores, product sales, or customer feedback ratings.

Gantt charts are indispensable for project management and scheduling. They allow you to visualize project timelines, task dependencies, and resource allocation. By using Gantt charts, you can efficiently plan and track project progress, ensuring that tasks are completed on time and within budget.

Heat maps, on the other hand, are instrumental in revealing patterns and trends within large datasets. They use color gradients to represent data values, making it effortless to spot correlations and variations. Heat maps are particularly useful in fields such as finance for visualizing stock price movements, in biology for gene expression analysis, and in marketing for customer behavior mapping.

As we navigate through the intricacies of these advanced charts, you'll acquire the knowledge and skills needed to unlock the full potential of these advanced chart types. You'll learn how to transform raw, often complex, data into actionable insights. By the conclusion of this chapter, you will be well-equipped to elevate your data visualization capabilities, enabling more informed decision-making and enhancing your proficiency in Excel's advanced charting techniques. It's an opportunity to gain a deeper understanding of your data and present it in a compelling, informative, and visually appealing manner. So, let's dive into the fascinating world of histograms, Gantt charts, and heat maps, and uncover the secrets that lie within these powerful visualization tools.

1.1 Learning Objectives

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

- *Identify the key components and elements of each chart type.*
- *Discuss the role of heat maps in identifying patterns and trends within datasets.*
- *Develop a Gantt chart for a project, including task dependencies and timelines.*
- *Analyse a histogram to draw conclusions about data characteristics, such as skewness or central tendency.*
- *Appraise the suitability of these chart types for various data analysis tasks and recommend the most appropriate choice based on specific objectives and datasets.*

2. HISTOGRAM

In statistics, a histogram serves as a graphical representation of data distribution. It is constructed using a series of adjacent rectangles, with each bar representing a specific type of data. Statistics, a branch of mathematics, finds applications across various domains. When numerical values repeat within statistical data, this repetition is referred to as "frequency" and can be organized into a table known as a "frequency distribution." Visual representations of frequency distributions can take the form of various graphs, with one of them being the histogram. In this article, we will delve into the concept of histograms, how to create them for given data, explore different types of histograms, and provide a detailed comparison between histograms and bar graphs.

A histogram is a graphical representation used to depict a grouped frequency distribution that involves continuous classes. It is essentially an area diagram comprising a series of rectangles, where the base of each rectangle aligns with the intervals between class boundaries, and the area of each rectangle is proportional to the frequency within the respective class. In this representation, the rectangles are all adjacent because the base covers the intervals between class boundaries, and the height of each rectangle is proportional to the frequency of its corresponding class. In simpler terms, a histogram is a diagram featuring rectangles, where the area of each rectangle corresponds to the frequency of a variable, and the width equals the class interval.

To create a histogram, the following steps should be followed:

1. Mark the class intervals on the X-axis and frequencies on the Y-axis.
2. Ensure that the scales on both axes are the same.
3. Make sure that the class intervals are non-overlapping.
4. Construct rectangles with their bases aligned to the class intervals and their heights representing the corresponding frequencies.
5. Place one rectangle over each class interval, with the class limits marked on the horizontal axis and the frequencies indicated on the vertical axis.
6. If the class intervals are of equal size, the height of each rectangle is proportional to the frequency of its respective class. If the class intervals vary in size, the area of each rectangle is proportional to the frequency of its corresponding class.

It's important to note that histograms differ slightly from regular graphs in that they do not have gaps between successive bars.

Histograms are employed in specific situations, and these conditions include:

1. The data being analyzed should consist of numerical values.
2. Histograms are utilized to assess the shape or pattern of data distribution.
3. They are employed to ascertain if there are variations or changes in a process over different time periods.
4. Histograms help in determining if there are distinctions in the output when multiple processes are involved.
5. They are instrumental in analyzing whether a given process aligns with customer requirements and expectations.

2.1 Difference Between Bar Graph and Histogram

A histogram is one of the most commonly used graphs to show the frequency distribution. As we know that the frequency distribution defines how often each different value occurs in the data set. The histogram looks more similar to the bar graph, but there is a difference between them. The list of differences between the bar graph and the histogram is given below

Histogram	Bar Graph
It is a two-dimensional figure	It is a one-dimensional figure
The frequency is shown by the area of each rectangle	The height shows the frequency and the width has no significance.
It shows rectangles touching each other	It consists of rectangles separated from each other with equal spaces.

The above differences can be observed from the below figures:

Fig. 1: Bar Graph (gap between bars) :

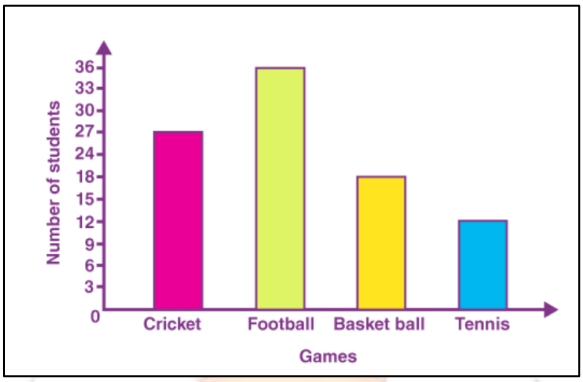
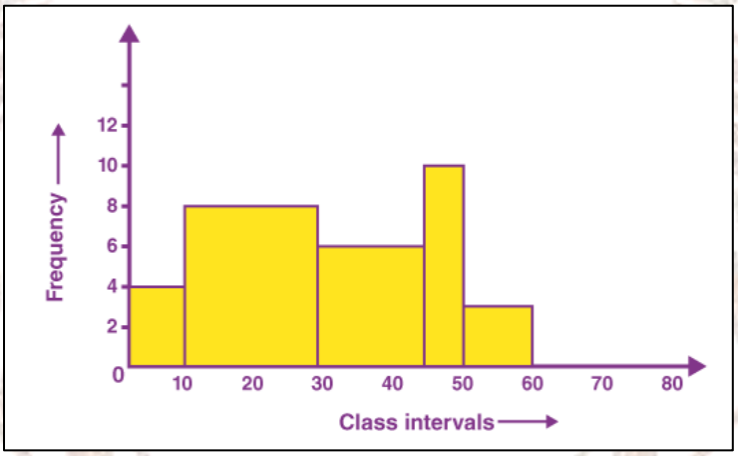


Fig. 2: Histogram: (No gap between bars)



Histograms come in various types, categorized based on the frequency distribution of the data they represent. Different types of data distributions, including normal, skewed, bimodal, multimodal, comb, edge peak, dog food, heart cut, and others, can be effectively depicted using histograms. The various types of histograms include:

1. Uniform Histogram
2. Symmetric Histogram
3. Bimodal Histogram
4. Probability Histogram

Uniform Histogram

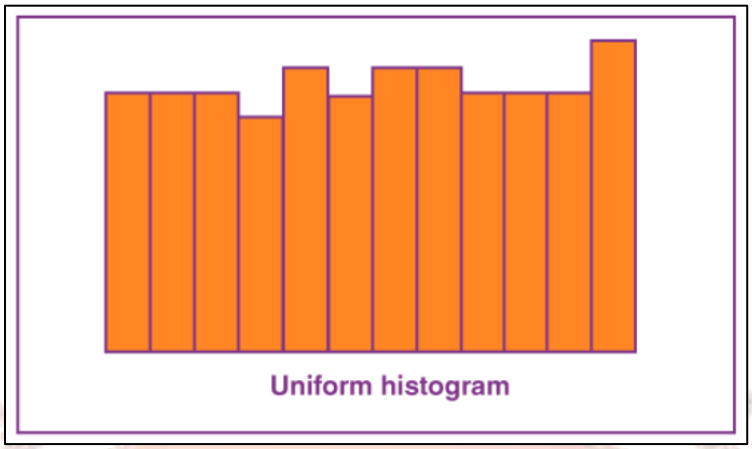


Fig 3: Uniform Histogram

A uniform distribution reveals that the number of classes is too small, and each class has the same number of elements. It may involve distribution that has several peaks.

Bimodal Histogram

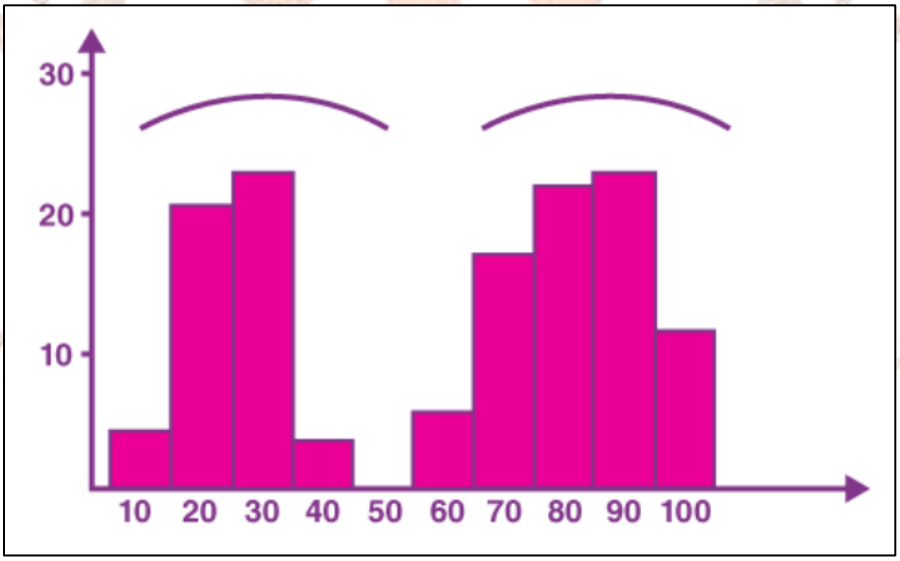


Fig 4: Bimodal Histogram

If a histogram has two peaks, it is said to be bimodal. Bimodality occurs when the data set has observations on two different kinds of individuals or combined groups if the centers of the two separate histograms are far enough to the variability in both the data sets.

Symmetric Histogram

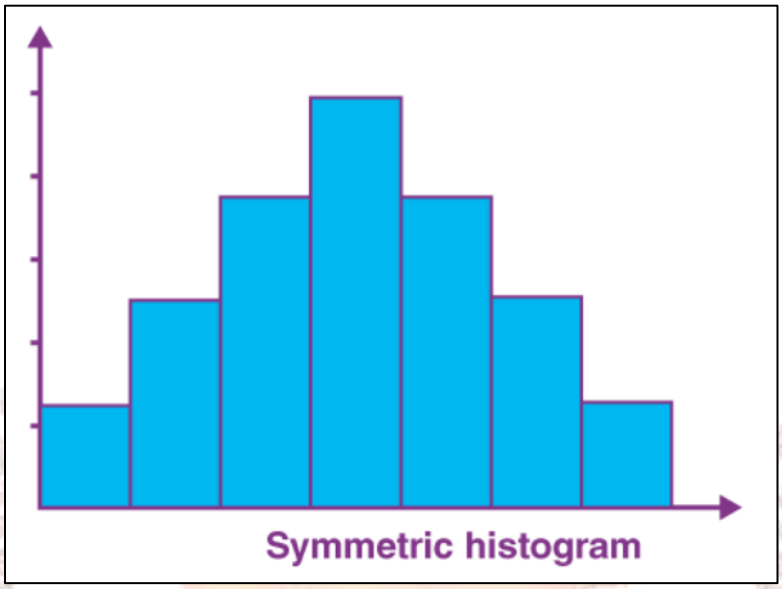


Fig. 5: Symmetric Histogram

A symmetric histogram is often referred to as a bell-shaped histogram. It earns this designation when a vertical line drawn through the centre of the histogram reveals that both sides are mirror images in terms of size and shape, rendering the histogram symmetric. In an ideal scenario, the right half of the diagram should be an exact replica of the left half. Histograms that do not exhibit this symmetrical quality are categorized as skewed.

Probability Histogram

Serves as a visual representation of a discrete probability distribution. It comprises rectangles centered on each x-value, with the area of each rectangle being directly proportional to the probability associated with that specific value. The creation of a probability histogram involves selecting the classes, and the probabilities of individual outcomes determine the heights of the histogram bars.

Histograms find various applications in the analysis of different types of data distributions:

1. Normal Distribution: The normal distribution, often represented as a bell curve, is characterized by a symmetrical pattern where data points are equally likely to appear on one side of the average as on the other. While other distributions may resemble normal distribution patterns, statistical calculations are employed to confirm the

normality of a distribution specific to a process. It's crucial to note that "normal" in this context signifies a distribution typical for a particular process, and deviations from this norm occur in various processes.

2. Skewed Distribution: Skewed distributions are asymmetrical, typically because a natural limit constrains outcomes on one side. The peak of the distribution is off-center in the direction of this limit, resulting in a long tail extending away from it. For example, a distribution of purity levels for a product would be skewed if purity cannot exceed 100%. These distributions can be classified as right-skewed or left-skewed based on the direction of the tail.
3. Multimodal Distribution: Also known as a plateau distribution, multimodal distributions combine several processes with normal distributions. The distribution's peak resembles a plateau due to the presence of multiple adjacent peaks.
4. Edge Peak Distribution: Edge peak distributions closely resemble normal distributions but exhibit a larger peak at one tail. Often, this is a result of histogram construction errors, with data aggregated into a category labeled "greater than."
5. Comb Distribution: Comb distributions feature alternating tall and short bars, typically arising from rounded-off data or incorrectly drawn histograms. For example, rounding temperatures to the nearest 0.2°C with a histogram bar width of 0.1°C would result in a comb-shaped distribution.
6. Truncated or Heart-Cut Distribution: This distribution closely resembles a normal distribution, but its tails are truncated or cut off. This can occur when a producer initially manufactures a product with a normal distribution and then relies on inspection to separate items falling within specification limits from those outside. The portion within specifications is referred to as the "heart cut."
7. Dog Food Distribution: This distribution is characterized by its absence of something significant. It typically clusters near the average. In cases where an end-user receives this distribution, others may be receiving heart-cut distributions, while the remaining end-users receive what's metaphorically referred to as "dog food." Even if an end-user's received items fall within specification limits, the items are divided into two clusters—one close to the upper specification limit and another close to the lower specification limit. This variation can cause problems in the end-user's processes.

2.2 Steps to create a Histogram:

Creation of Histogram using analysis tool pak:

To enable the Analysis ToolPak add-in in Microsoft Excel, follow these steps:

1. In Excel versions 2010 to 365, click on the "File" tab, then select "Options." For Excel 2007, click the Microsoft Office button and choose "Excel Options."
2. Within the Excel Options dialog box, click on "Add-Ins" in the left sidebar.
3. In the "Manage" box, select "Excel Add-ins," and click the "Go" button.

By following these steps, you can activate the Analysis Tool Pak add-in, which is a valuable tool for data analysis in Excel.

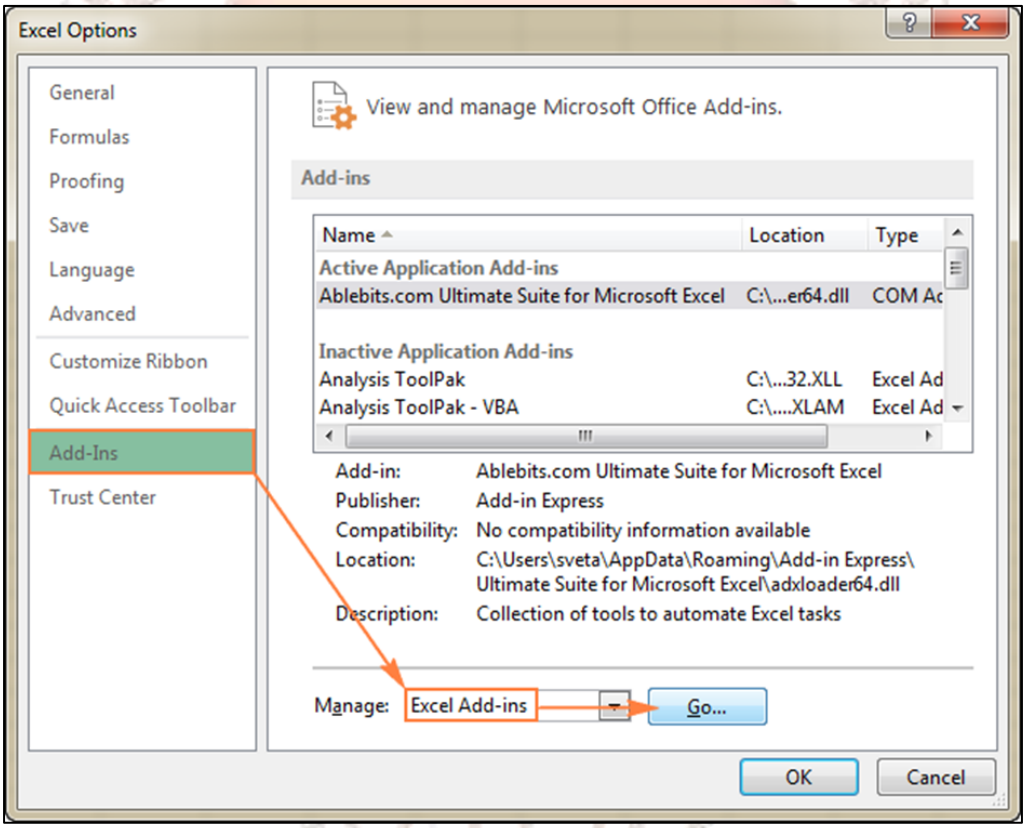


Fig. 6: Excel options

4. In the Add-Ins dialog box, tick the box next to Analysis ToolPak, and then click OK to close the dialog.

If Excel prompts you with a message stating that the Analysis ToolPak is not presently installed on your computer, select Yes to proceed with the installation.

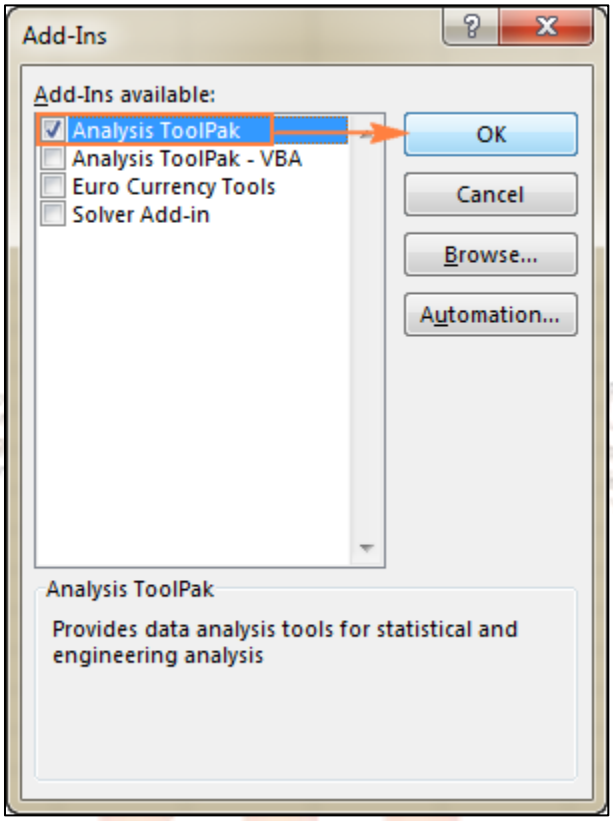


Fig. 7: Add-In Screen

Now that the Analysis ToolPak is integrated into Excel, you can find its commands in the Analysis group on the Data tab.

Specify excel histogram bit range:

Before creating a histogram chart, you need to perform one more step: define the bins in a separate column. Bins represent the intervals into which you want to categorize your source data, and they should be consecutive, non-overlapping, and typically of equal size.

Excel's Histogram tool places the input data values into the specified bins based on the following logic:

- A value is assigned to a particular bin if it falls within the range defined by that bin's lower and upper bounds.
- Any values in the input data that exceed the upper bound of the highest bin are grouped into the "More" category.
- If you don't specify a bin range, Excel will automatically create a set of evenly spaced bins covering the minimum to maximum values in your data.

Given these criteria, input the bin numbers in an ascending order into a separate column. Your Excel histogram bin range should be confined to the range of your input data.

As an example, suppose you have order numbers in column A and estimated delivery times in column B. To create an Excel histogram showing the number of deliveries within certain time frames (1-5 days, 6-10 days, 11-15 days, 16-20 days, and over 20 days), you would input the bin range from 5 to 20 with an increment of 5, as illustrated in the screenshot below:

	A	B	C	D
1	Order no.	Delivery (days)		Bins
2	1002369	4		5
3	1002382	24		10
4	1002378	13		15
5	1002372	6		20
6	1002346	19		
7	1002345	5		
8	1002347	14		
9	1002362	9		
10	1002358	5		
11	1002360	17		
12	1002376	14		
13	1002371	2		
14	1002380	13		
15	1002354	5		

Fig. 8: Dataset

Make a histogram using Excel's Analysis ToolPak:

With the Analysis ToolPak enabled and bins specified, perform the following steps to create a histogram in your Excel sheet:

1.On the *Data tab,* in the *Analysis* group, click the **Data Analysis** button.

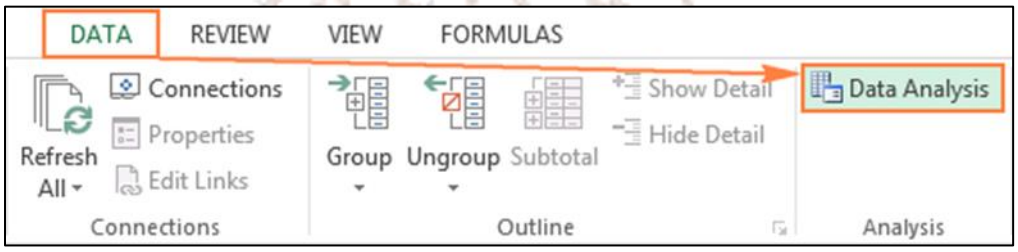


Fig. 9: Data Analysis Tab

2. In the *Data Analysis* dialog, select **Histogram** and click OK.

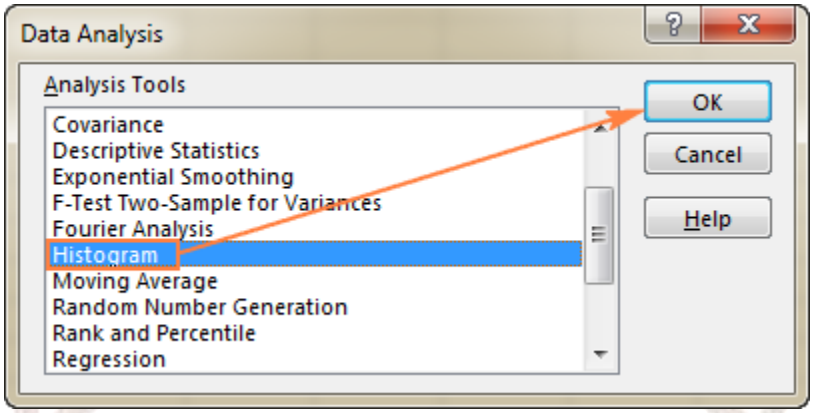


Fig. 10: Choose Histogram

3. In the *Histogram* dialog window, do the following:
   a. Specify the **Input range** and the **Bin range.**

To do this, you can place the cursor in the box, and then simply select the corresponding range on your worksheet using the mouse. Alternatively, you can click the *Collapse Dialog button* , select the range on the sheet, and then click the *Collapse Dialog button* again to return to the *Histogram* dialog box.

Tip. If you included column headers when selecting the input data and bin range, select the **Labels** check box.

   b. Select the **Output options.**

To place the histogram on the same sheet, click *Output Range*, and then enter the upper-left cell of the output table.

To paste the output table and histogram in a new sheet or a new workbook, select *New Worksheet Ply* or *New Workbook*, respectively.

Finally, choose any of the additional options:

i. To present data in the output table in descending order of frequency, select the **Pareto** (sorted histogram) box.
ii. To include a cumulative percentage line in your Excel histogram chart, select the **Cumulative Percentage** box.

iii. To create an embedded histogram chart, select the **Chart Output** box.

For this example, I've configured the following options:

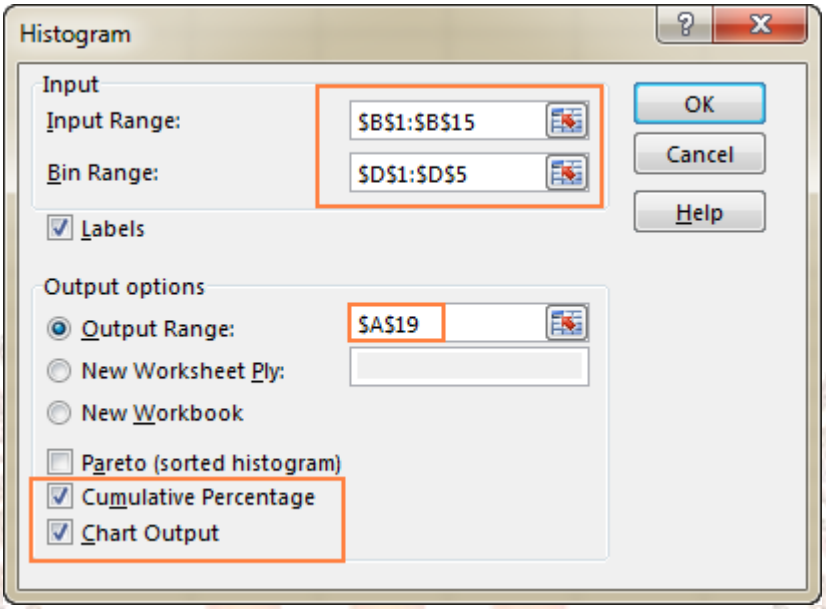


Fig. 11: Enter the Range

4. And now, click *OK*, and review the output table and histogram graph:

Bins	*Frequency*	*Cumulative %*
5	5	35.71%
10	2	50.00%
15	4	78.57%
20	2	92.86%
More	1	100.00%

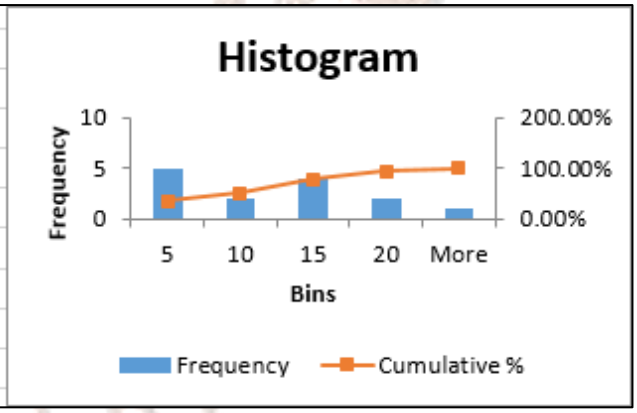


Fig. 12: Output View

Tip. To improve the histogram, you can replace the default Bins and Frequency with more meaningful axis titles, customize the chart legend, etc. Also, you can use the design, layout, and format options of the Chart Tools to change the display of the histogram, for example remove gaps between columns.

As you've just seen, it's very easy to make a histogram in Excel using the Analysis ToolPak. However, this method has a significant limitation - the embedded histogram chart is **static**, meaning that you will need to create a new histogram every time the input data is changed.

To make an automatically updatable histogram, you can either use Excel functions or build a PivotTable as demonstrated below.

How to make a histogram in Excel using formulas

Another way to create a histogram in Excel is using the FREQUENCY or COUNTIFS function. The biggest advantage of this approach is that you won't have to re-do your histogram with each change in the input data. Like a normal Excel chart, your histogram will update automatically as soon as you edit, add new or delete existing input values.

To begin with, arrange your source data in one column (column B in this example), and enter the bin numbers in another column (column D), like in the screenshot below:

	A	B	C	D
1	Order no.	Delivery (days)		Bins
2	1002369	4		5
3	1002382	20		10
4	1002378	13		15
5	1002372	6		20
6	1002346	25		25
7	1002345	5		30
8	1002347	27		35
9	1002362	9		
10	1002358	5		

Fig. 13: Dataset

Now, we will use a Frequency or Countifs formula to calculate how many values fall into the specified ranges (bins), and then, we will draw a histogram based on that summary data.

1. Creating a histogram using Excel's FREQUENCY function

The most obvious function to create a histogram in Excel is the FREQUENCY function that returns the number of values that fall within specific ranges, ignoring text values and blank cells.

The FREQUENCY function has the following syntax:

FREQUENCY (data_array, bins_array)

- Data_array - a set of values for which you want to count frequencies.
- Bins_array - an array of bins for grouping the values.

In this example, the data_array is B2:B40, bin array is D2:D8, so we get the following formula: **=FREQUENCY (B2:B40,D2:D8)**

Please keep in mind that FREQUENCY is a very specific function, so follow these rules to make it work right:

- An Excel Frequency formula should be entered as a **multi-cell array formula.** First, select a range of adjacent cells where you want to output the frequencies, then type the formula in the formula bar, and press Ctrl + Shift + Enter to complete it.
- It's recommended to enter one more Frequency formula than the number of bins. The extra cell is required to display the count of values above the highest bin. For the sake of clarity, you can label it *"More"* like in the following screenshot (but don't include that *"More"* cell in your bins_array!):

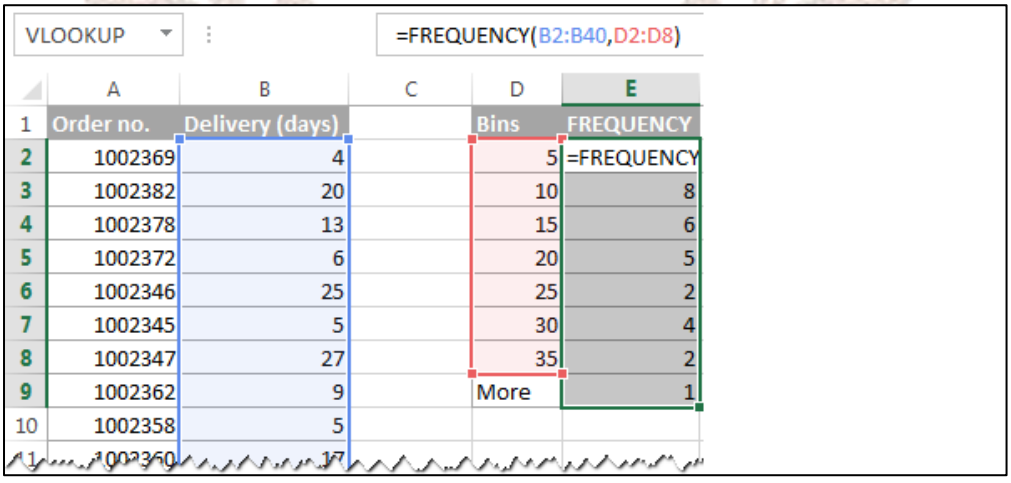


Fig. 14: Using Frequency Formula

Like the *Histogram* option of the Analysis ToolPak, the Excel FREQUENCY function returns values that are greater than a previous bin and less than or equal to a given bin. The last

Frequency formula (in cell E9) returns the number of values greater than the highest bin (i.e. the number of delivery days over 35).

To make things easier to understand, the following screenshot shows the bins (column D), corresponding intervals (column C), and computed frequencies (column E):

	C	D	E
	Interval	Bins	FREQUENCY
1	1-5	5	11
2	6-10	10	8
3	11-15	15	6
4	16-20	20	5
5	21-25	25	2
6	26-30	30	4
7	31-35	35	2
8	Over 35		1

Fig. 15: Final Output

Note. Because Excel FREQUENCY is an array function, you cannot edit, move, add or delete the individual cells containing the formula. If you decide to change the number of bins, you will have to delete the existing formula first, then add or delete the bins, select a new range of cells, and re-enter the formula.

2. Making a histogram using COUNTIFS function

Another function that can help you calculate frequency distributions to plot histogram in Excel is COUNTIFS. And in this case, you will need to use 3 different formulas:

- The formula for the first cell **- top bin** (F2 in the screenshot below):

=COUNTIFS($B$2:$B$40,"<="&$D2)

The formula counts how many values in column B a re less than the smallest bin in cell D2, i.e. returns the number of items delivered within 1-5 days.

- The formula for the last cell - **over the highest bin** (F9 in the screenshot below):

=COUNTIFS($B$2:$B$100,">"&$D8)

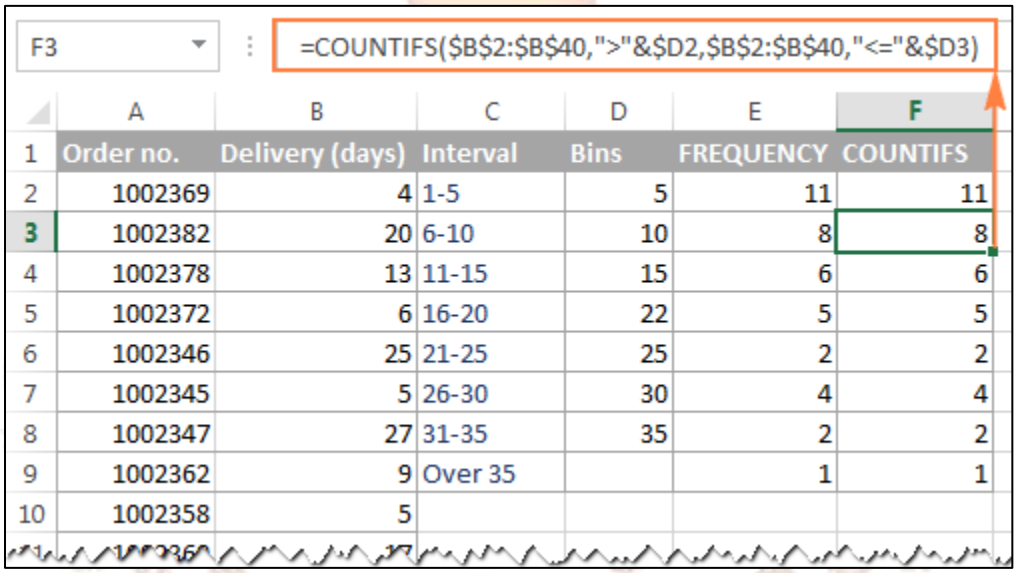
The formula counts how many values in column B are greater than the highest bin in D8.

- The formula for remaining bins (cells F3:F8 in the screenshot below):

=COUNTIFS($B$2:$B$40,">"&$D2,$B$2:$B$40,"<="&$D3)

The formula counts the number of values in column B that are greater than the bin in the above row and less than or equal to the bin in the same row.

As you see, the FREQUENCY and COUNTIFS functions return identical results:



	A	B	C	D	E	F
1	Order no.	Delivery (days)	Interval	Bins	FREQUENCY	COUNTIFS
2	1002369	4	1-5	5	11	11
3	1002382	20	6-10	10	8	8
4	1002378	13	11-15	15	6	6
5	1002372	6	16-20	22	5	5
6	1002346	25	21-25	25	2	2
7	1002345	5	26-30	30	4	4
8	1002347	27	31-35	35	2	2
9	1002362	9	Over 35		1	1
10	1002358	5				

Fig. 16: Use of COUNTIFS

"What is the reason of using three different formulas instead of one?" you may ask me. Basically, you get rid of the multi-cell array formula and can add and delete bins easily.

Tip. If you plan to add more input data rows in the future, you can supply a bigger range in your FREQUENCY or COUNTIFS formulas, and you won't have to change your formulas as you add more rows. In this example, the source data are in cells B2:B40. But you can supply the range B2:B100 or even B2:B1000, just in case :) For example:

=FREQUENCY(B2:B1000,D2:D8)

Make a histogram based on the summary data

Now that you have a list of frequency distributions computed with either FREQUENCY or COUNTIFS function, create a usual bar chart - select the frequencies, switch to the *Insert tab* and click the 2-D Column chart in the Charts group:

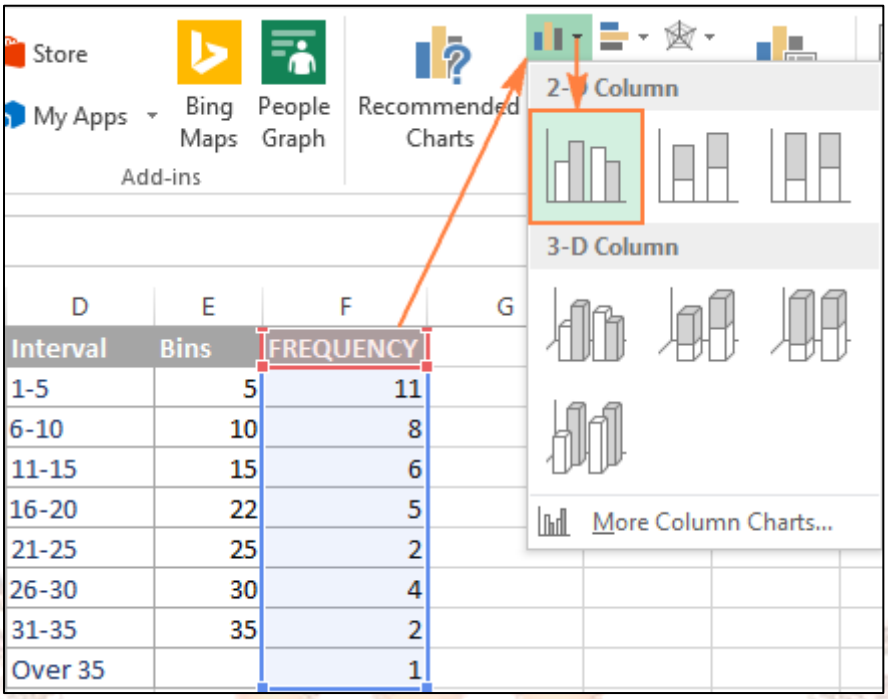


Fig. 17: Chart Selection

The bar graph will be immediately inserted in your sheet:

	A	B	C	D	E	F	G
1	Order no.	Delivery (days)		Interval	Bins	FREQUENCY	
2	1002369	4		1-5	5	11	
3	1002382	20		6-10	10	8	
4	1002378	13		11-15	15	6	
5	1002372	6		16-20	22	5	
6	1002346	25		21-25	25	2	
7	1002345	5		26-30	30	4	
8	1002347	27		31-35	35	2	
9	1002362	9		Over 35		1	
10	1002358	5					
11	1002360	17					
12	1002376	14					
13	1002371	2					
14	1002380	13					
15	1002354	5					
16	1002375	7					
17	1002368	17					
18	1002373	28					
19	1002381	23					
20	1002370	7					
21	1002366	26					
22	1002361	40					
23	1002352	3					
24	1002355	12					
25	1002356	7					

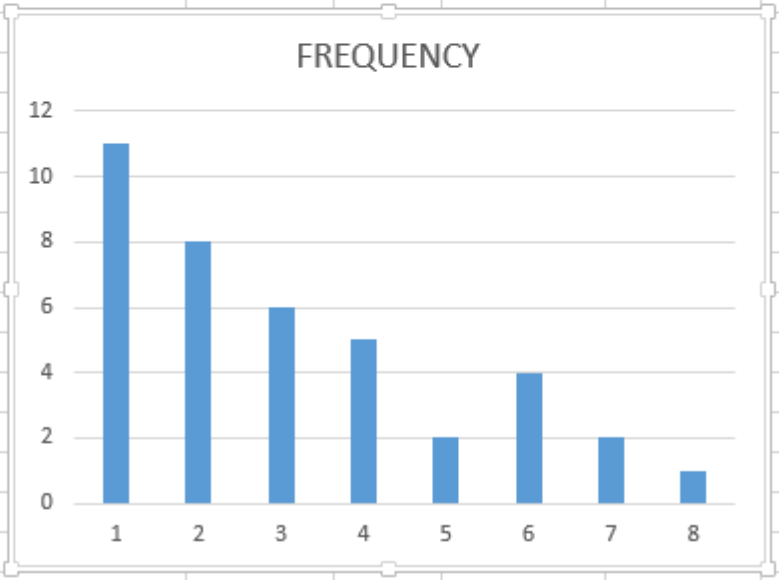


Fig. 18: Output

Generally speaking, you already have a histogram for your input data, though it definitely requires a few improvements. Most importantly, to make your Excel histogram easy to understand, you need to replace the default labels of the horizontal axis represented by serial numbers with your bin numbers or ranges.

The easiest way is to type the **ranges** in a column left to the column with the Frequency formula, select both columns - *Ranges and Frequencies* - and then create a bar chart. The ranges will be automatically used for the X axis labels, as shown in the below screenshot:

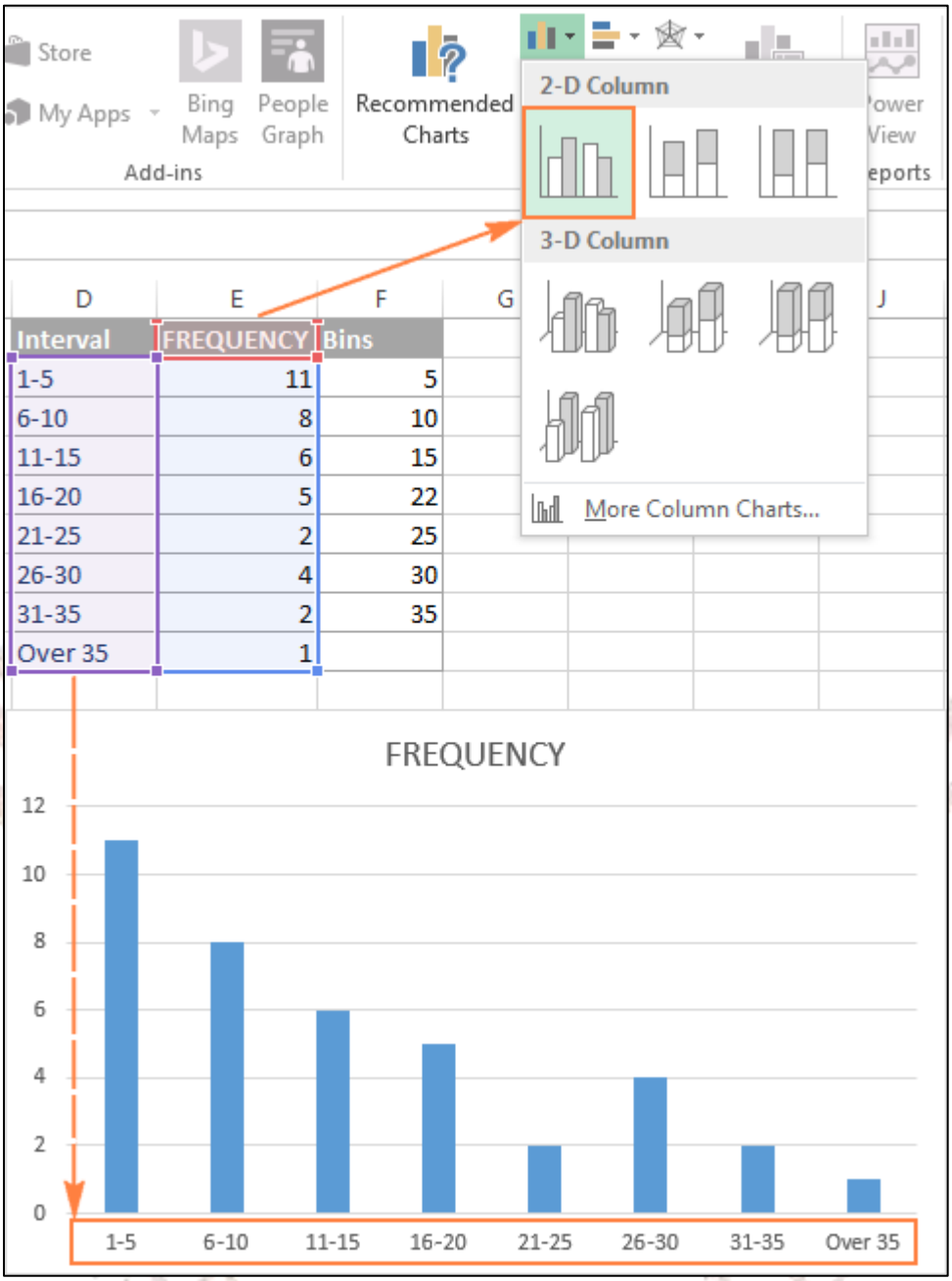


Fig. 19: Output

Tip. If Excel converts your intervals to dates (e.g. 1-5 can be automatically converted to 05-Jan), then type the intervals with a preceding apostrophe (') like '1-5. If you want the labels of your Excel histogram to display **bin numbers,** type them with preceding apostrophes too, e.g. '5, '10, etc. The apostrophe just converts numbers to text and is invisible in cells and on the histogram chart.

If there is no way you can type the desired histogram labels on your sheet, then you can enter them directly on the chart, independently of the worksheet data. The final part of this tutorial explains how to do this and shows a couple of other improvements that can be made to your Excel histogram.

How to make a histogram with a PivotChart

As you may have noticed in the two previous examples, the most time-consuming part of creating a histogram in Excel is calculating the number of items within each bin. Once the source data has been grouped, an Excel histogram chart is fairly easy to draw.

As you probably know, one of the fastest ways to automatically summarize data in Excel is a PivotTable. So, let's get to it and plot a histogram for the Delivery data (column B):

	A	B
1	Order no.	Delivery (days)
2	1002369	4
3	1002382	20
4	1002378	13
5	1002372	6
6	1002346	25
7	1002345	5
8	1002347	27

Fig. 20: Dataset

1. Create a pivot table

To create a pivot table, go to the *Insert tab* > *Tables* group, and click **PivotTable.** And then, move the Delivery field to the ROWS area, and the other field (*Order no*. in this example) to the VALUES area, as shown in the below screenshot.

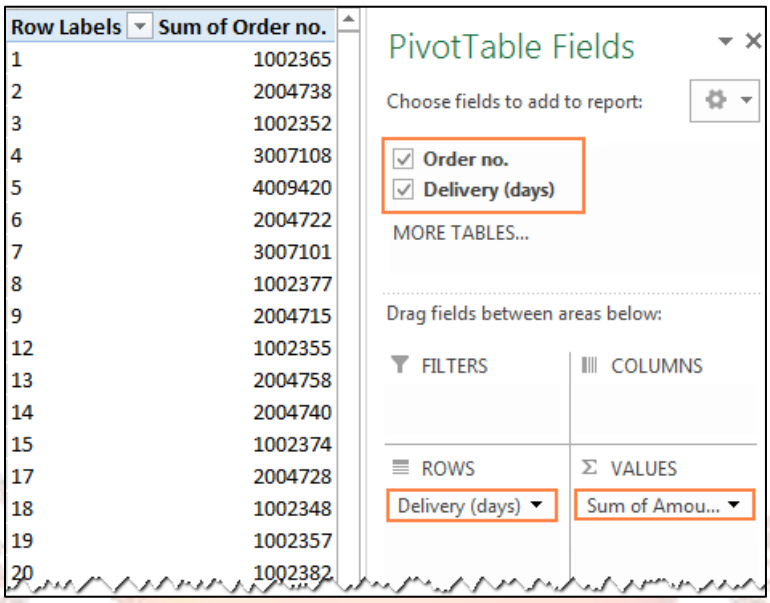


Fig. 21: Pivot Table

2. Summarize values by Count

By default, numeric fields in a PivotTable are summed, and so is our *Order numbers* column, which makes absolutely no sense :) Anyway, because for a histogram we need a count rather than sum, right-click any order number cell, and select *Summarize Values By* > **Count.**

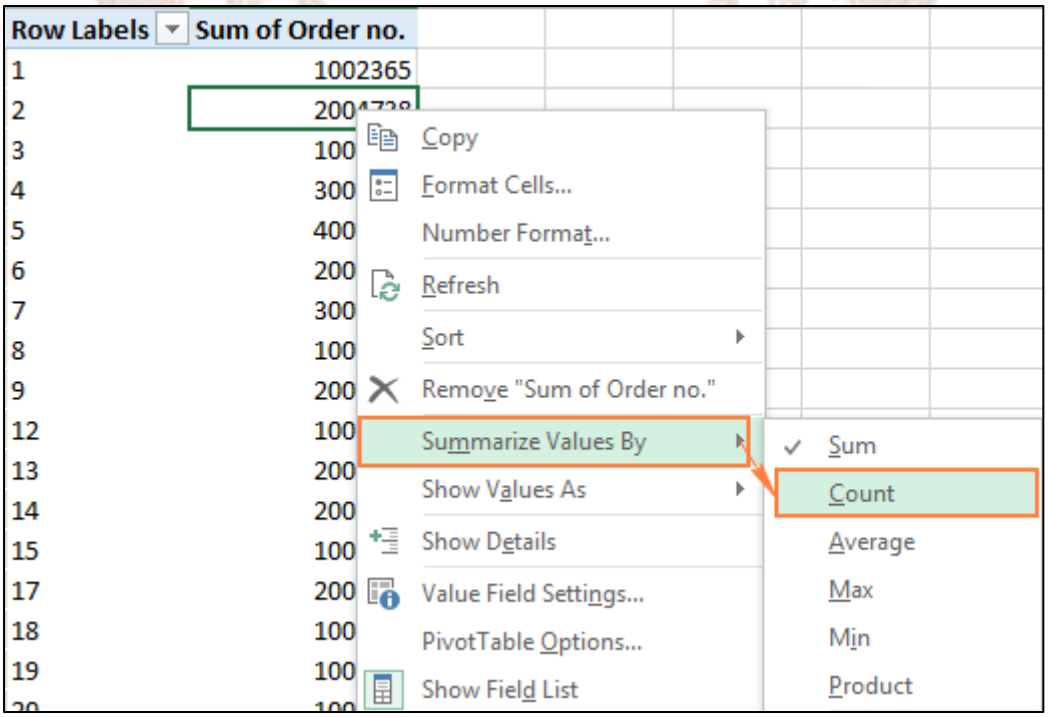


Fig. 22: Filtering in Pivot Table

Now, your updated PivotTable should look like this:

Row Labels	Count of Order no.
1	1
2	2
3	1
4	3
5	4
6	2
7	3
8	1
9	2

Fig 23: Output of Pivot Table

3. Create the intervals (bins)

The next step is to create the intervals, or bins. For this, we will be using the Grouping option. Right-click any cell under Row Labels in your pivot table, and select **Group…**

In the *Grouping* dialog box, specify the starting and ending values (usually Excel enters the minimum and maximum value automatically based on your data), and type the desired increment (interval length) in the By box.

In this example, the minimum delivery time is 1 day, maximum - 40 days, and the increment is set to 5 days:

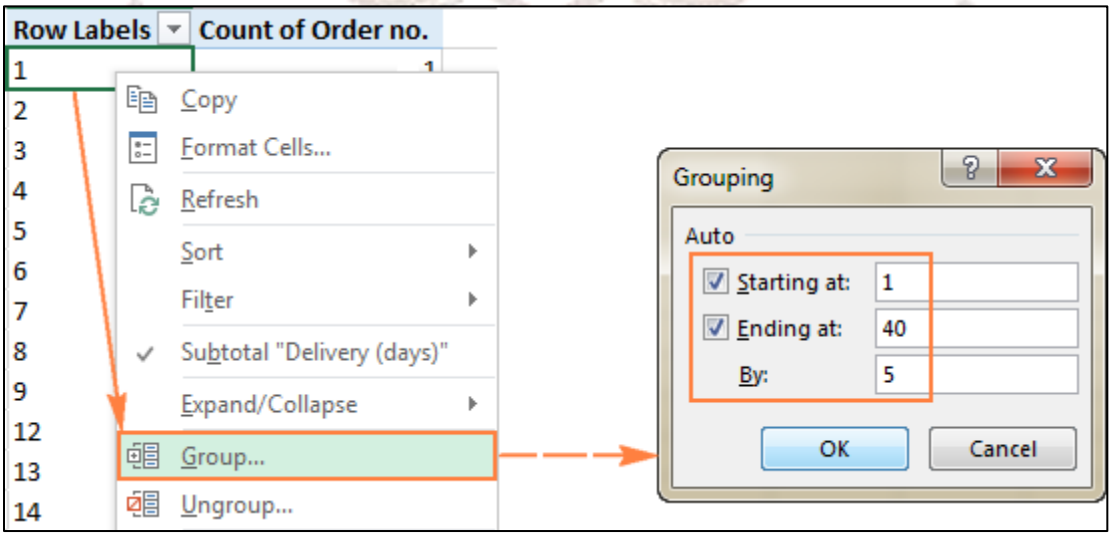


Fig 24: Setting up Range and Interval

Click OK, and your pivot table will display the intervals as specified:

Row Labels	Count of Order no.
1-5	11
6-10	8
11-15	6
16-20	5
21-25	2
26-30	4
31-35	2
36-40	1
Grand Total	39

Fig 25: Output after setting Range and Interval

4. Plot a histogram

One final step is left - draw a histogram. To do this, simply create a column pivot chart by clicking the **PivotChart** on the *Analyze* tab in *PivotTable* Tools group:

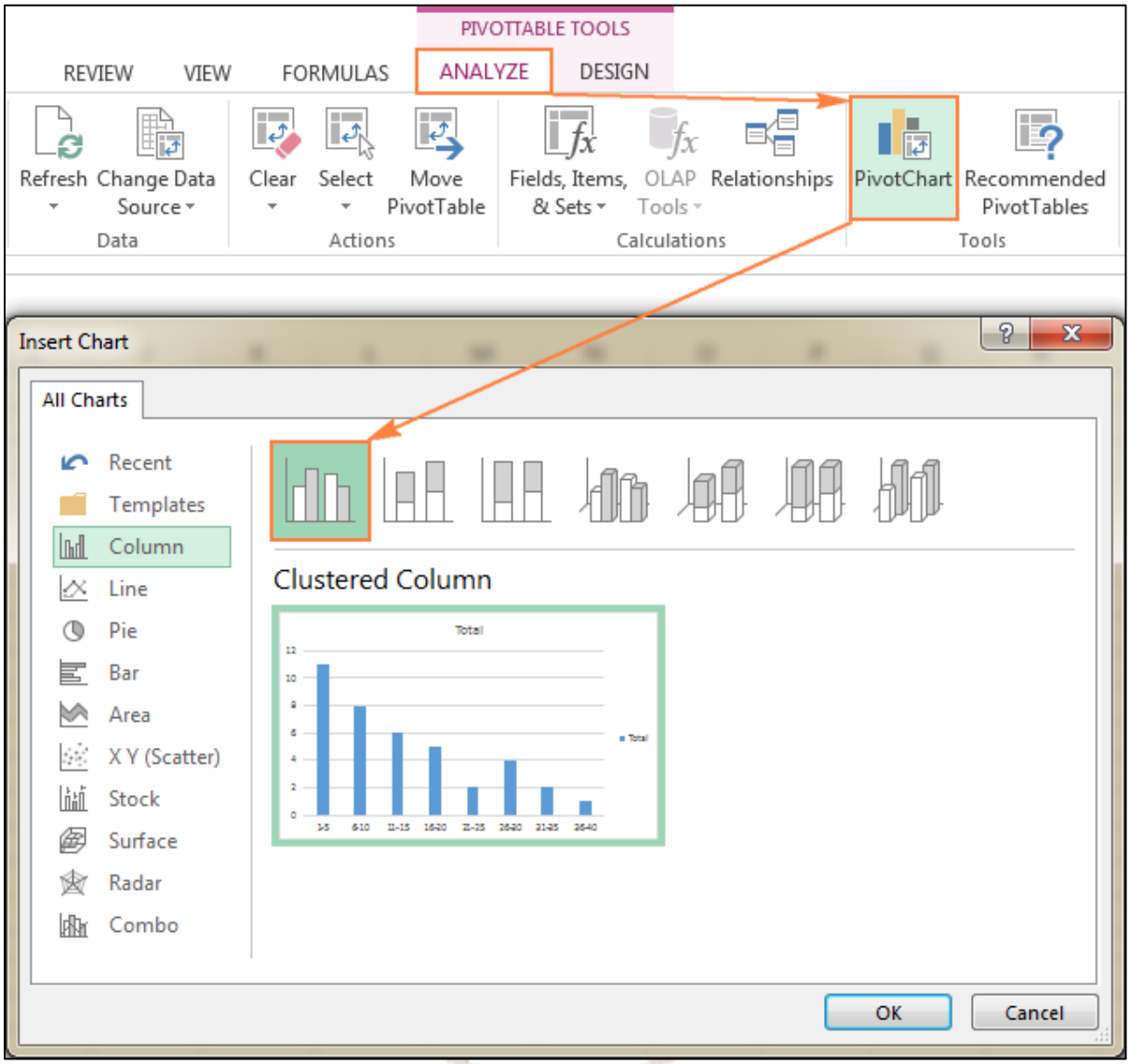


Fig 26: Selecting Pivot Chart

And the default column PivotChart will appear in your sheet straight away:

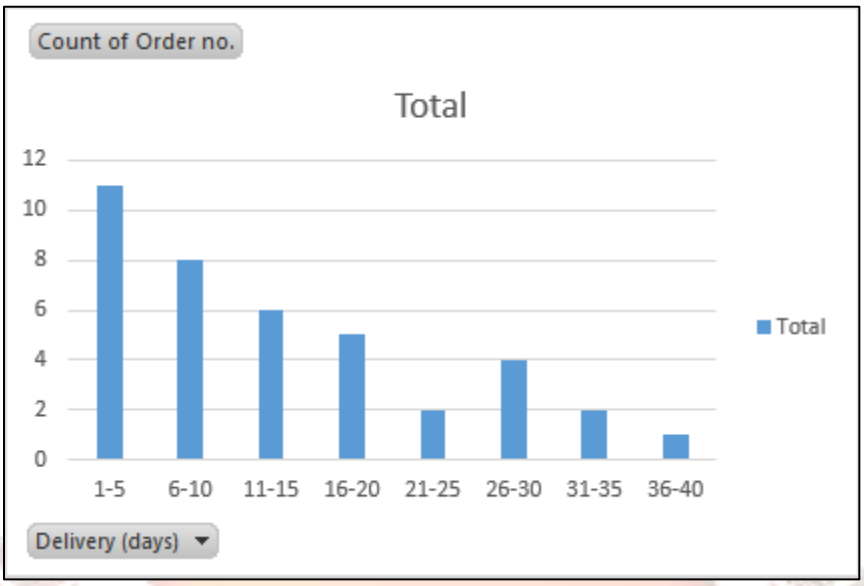


Fig 27. Output of the Pivot Chart

And now, polish up your histogram with a couple of finishing touches:

- Delete the legend by clicking the *Chart Elements* button  and removing the tick from the *Legend Or*, select the legend on the histogram and press the *Delete key* on your keyboard.
- Replace the default *Total* title with something more meaningful.
- Optionally, choose another chart style in the *Chart Styles* group on the *PivotChart Tools > Design tab.*
- Remove the chart buttons by clicking **Field Buttons** on the *PivotChart Tools > Analyze tab,* in the *Show/Hide group*:

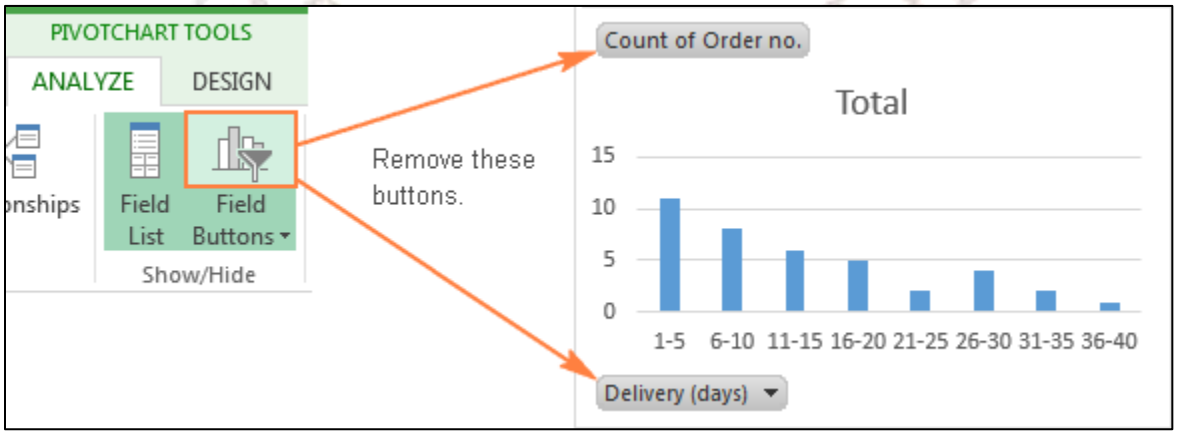


Fig 28: Updating the Field Buttons

2.3 Best Practices for Using a Histogram

Use a zero valued baseline:

An important aspect of histograms is that they must be plotted with a zero-valued baseline. Since the frequency of data in each bin is implied by the height of each bar, changing the baseline or introducing a gap in the scale will skew the perception of the distribution of data.

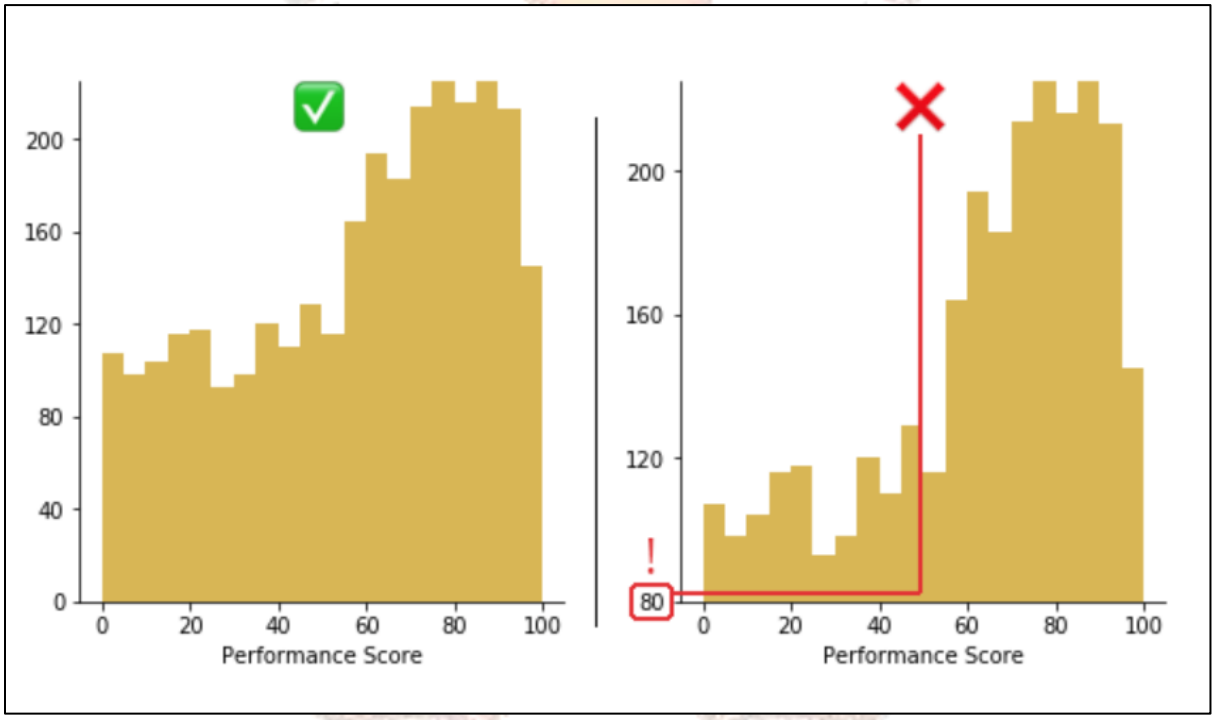


Fig 29: Comparison between a Zero Valued and Non-Zero Valued Histogram

Trimming 80 points from the vertical axis makes the distribution of performance scores look much better than they actually are.

Choose appropriate number of bins:

Selecting an appropriate number of bins is a crucial step in creating a histogram. While most histogram-generating tools offer default binning algorithms, it's often necessary to experiment with binning parameters to find the configuration that best suits your data. Wikipedia provides guidelines on choosing the right number of bins and their sizes, but the ideal choice often involves a combination of domain expertise and experimentation tailored to your specific objectives.

The bin size choice is inversely related to the number of bins. Larger bin sizes result in fewer bins needed to cover the entire data range, while smaller bin sizes require more bins. It's essential to invest time in testing various bin sizes to observe how the data distribution appears in each configuration and select the one that most accurately represents the data. Having too many bins can make the distribution appear noisy and challenging to interpret, whereas too few bins can lead to a lack of detail, hindering the identification of meaningful patterns in the data.

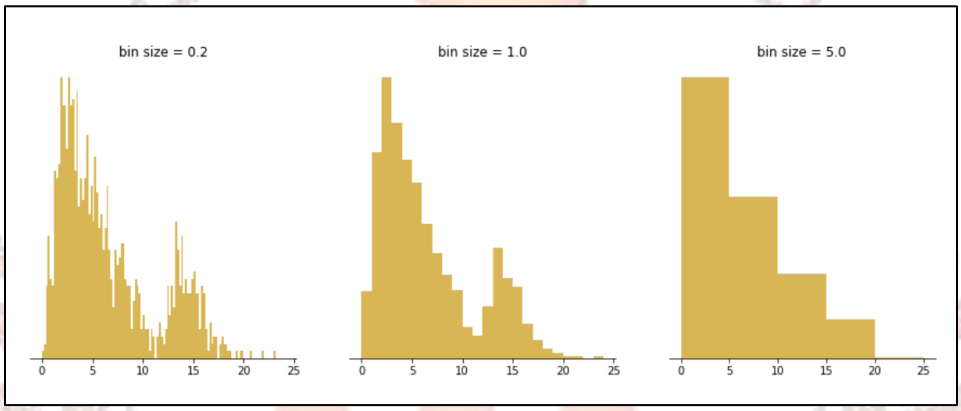


Fig 30: Comparison of Histograms with Different Bin Sizes

The left panel's bins are too small, implying a lot of spurious peaks and troughs. The right panel's bins are too large, hiding any indication of the second peak.

Choose interpretable bin boundaries:

It is advisable to position tick marks and labels at the bin boundaries to provide clear indications of where each bar's limits lie. While labels for every bar are not necessary, spacing them between a few bars helps readers track the values effectively. Additionally, using labels with a limited number of significant figures enhances readability.As a rule of thumb, bin sizes of 1, 2, 2.5, 4, or 5 (which evenly divide numbers like 5, 10, and 20) or their powers of ten are considered suitable starting points. Conversely, bin sizes such as 3, 7, or 9 may pose readability challenges and should be avoided unless the context justifies their use.

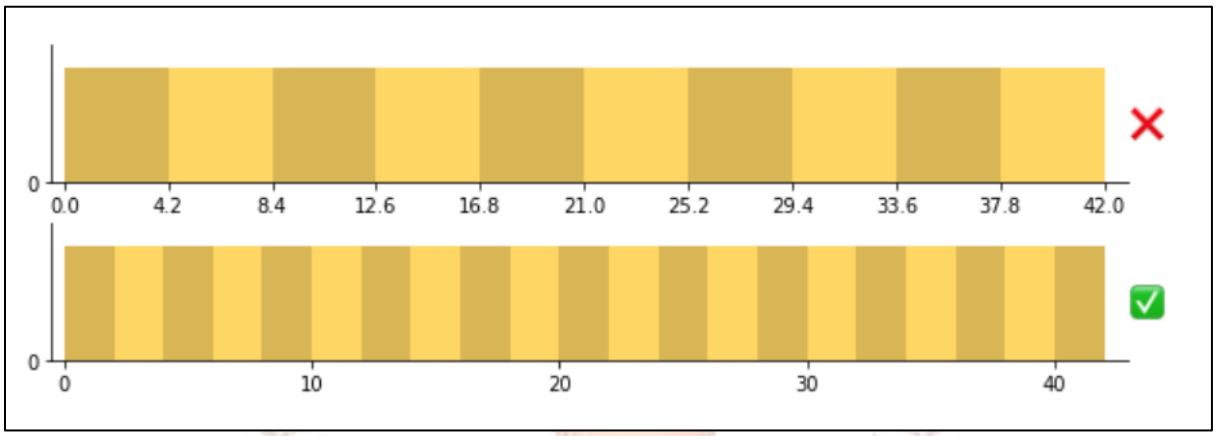


Fig 31: Different Usage of Bin Sizes 1

Top: carelessly splitting the data into ten bins from min to max can end up with some very odd bin divisions. Bottom: fewer tick marks are needed when the bin size is easy to follow.

A small word of caution: make sure you consider the types of values that your variable of interest takes. In the case of a fractional bin size like 2.5, this can be a problem if your variable only takes integer values. A bin running from 0 to 2.5 has opportunity to collect three different values (0, 1, 2) but the following bin from 2.5 to 5 can only collect two different values (3, 4 – 5 will fall into the following bin). This means that your histogram can look unnaturally "bumpy" simply due to the number of values that each bin could possibly take.

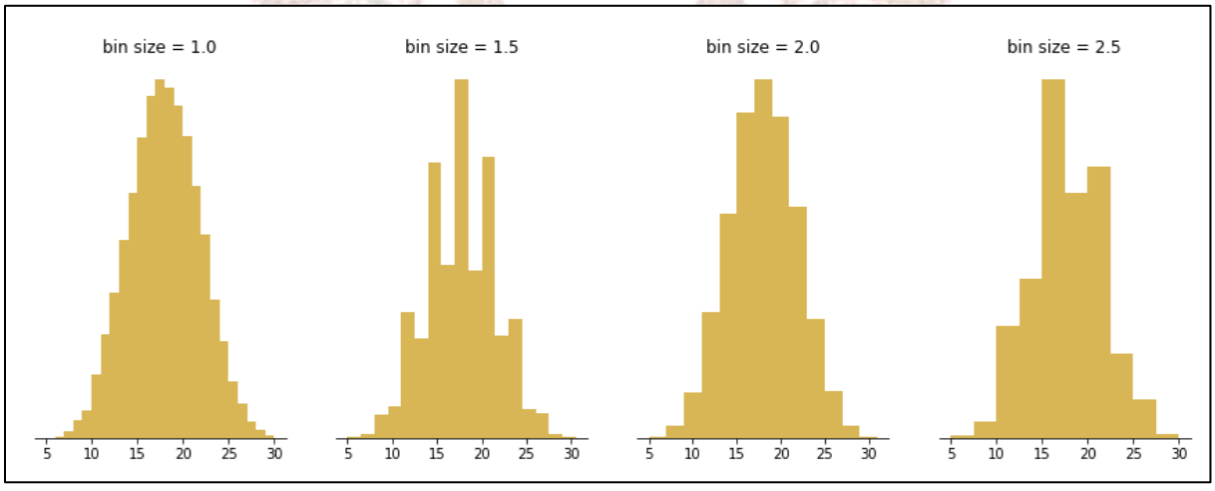


Fig 32: Different Usage of Bin Sizes 2

The figure above visualizes the distribution of outcomes when summing the result of five die rolls, repeated 20 000 times. The expected bell shape looks spiky or lopsided when bin sizes that capture different amounts of integer outcomes are chosen.

2.4 Common Misuses

Measured variable is not continuous numeric:

As mentioned earlier, histograms are designed to represent the frequency distribution of continuous numeric variables. When dealing with variables that don't adhere to this characteristic, such as categorical variables like user type (e.g., guest, user) or location, it is more appropriate to use a bar chart. However, there are certain types of variables that can be somewhat ambiguous in classification: those with discrete numeric values and those involving time-based values. Variables with discrete numeric values (e.g., integers 1, 2, 3, etc.) can be displayed using either a bar chart or a histogram, depending on the specific context. Histograms become more suitable when there is a considerable number of distinct values to represent. When the numeric value range is extensive, the discrete nature of the values often becomes less significant, and grouping them into continuous intervals becomes more appropriate. A critical consideration is ensuring that the numeric values genuinely represent actual measurements. If the numbers are codes for a categorical or loosely-ordered variable, then it's an indication that a bar chart should be used. For instance, if survey responses are on a scale from 1 to 5, representing values from "strongly disagree" to "strongly agree," it's more appropriate to visualize the frequency distribution using a bar chart. This is because the differences between individual values may not be consistent, and we can't assume that the meaningful difference between a 1 and 2 ("strongly disagree" to "disagree") is the same as the difference between a 2 and 3 ("disagree" to "neither agree nor disagree").

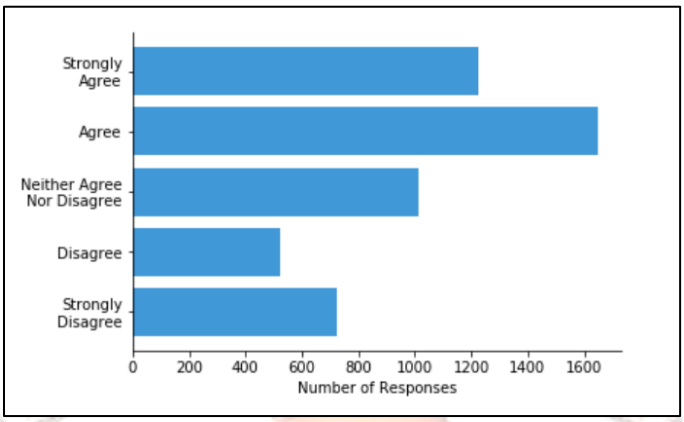


Fig. 33: Example of Misuse of Histogram 1

A more complex scenario arises when we're dealing with a time-based variable. If the values represent relative time periods (e.g., 30 seconds, 20 minutes), then grouping them by time intervals for a histogram is a logical choice. However, when values correspond to absolute timestamps (e.g., January 10, 12:15), the situation becomes less clear. With absolute times, new data points are typically assigned to newly-created bins, rather than falling within existing bins. Additionally, certain natural grouping options, such as by month or quarter, can result in slightly uneven bin sizes. Due to these factors, it's not uncommon to observe the use of different chart types like bar charts or line charts in such cases.

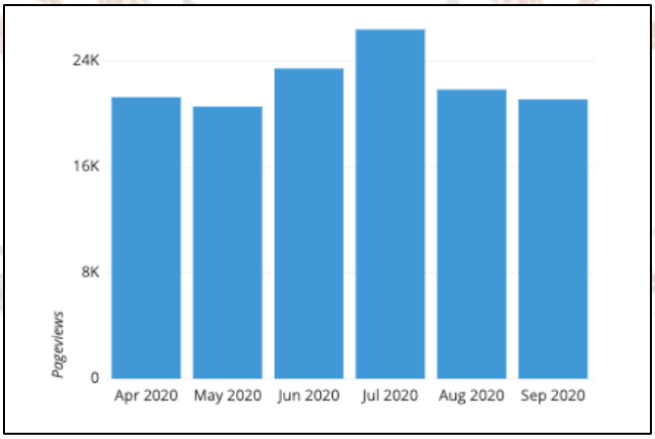


Fig. 34: Example of Misuse of Histogram 2

Using unequal bin size:

While the previous examples have demonstrated histograms using bins of equal width, it's important to note that this is not a strict requirement from a technical standpoint. In

situations where data is sparse, such as a long tail of data points, you might consider using wider bins to cover that space. However, creating a histogram with bins of unequal size, although not a technical mistake, introduces significant changes in the histogram's construction and can complicate interpretation. The key technical point regarding histograms is that the total area of the bars represents the whole dataset, and the area of each bar corresponds to the proportion of the dataset contained within each bin. When bin sizes are consistent, measuring both bar area and height is equivalent. However, in a histogram with varying bin sizes, the height can no longer represent the total frequency of occurrences accurately. Doing so would distort the perception of how many points are in each bin, as increasing a bin's size would make it appear larger. In the middle plot of the figure below, bins from 5-6, 6-7, and 7-10 appear to contain more points than they actually do.

Instead, the vertical axis should represent the frequency density per unit of bin size. For instance, in the right pane of the figure below, the bin from 2-2.5 has a height of about 0.32. When multiplied by the bin width (0.5), we can estimate that approximately 16% of the data falls within that bin. The heights of wider bins have been scaled down compared to the central pane, and you'll notice that the overall shape looks similar to the original histogram with equal bin sizes. However, density can be a challenging concept to grasp, and presenting such a plot to individuals unfamiliar with this concept may lead to difficulties in interpretation.

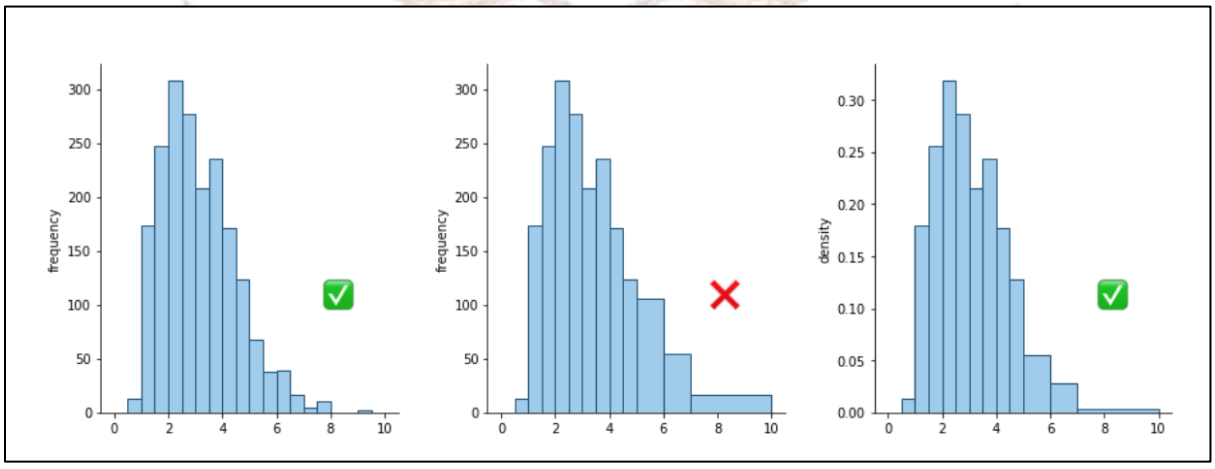


Fig. 35: Comparison of Unequal Bin Size for Histogram

Left: histogram with equal-sized bins; Center: histogram with unequal bins but improper vertical axis units; Right: histogram with unequal bins with density heights

Therefore, it is generally advisable to use completely equal bin sizes. While this may result in empty bins and some increased noise in ranges with sparse data, it significantly enhances the interpretability of your histogram. Alternatively, if the nature of the variable suggests uneven bin sizes, it may be more suitable to opt for a bar chart instead.

3. GANTT CHART

What Is a Gantt Chart?

A Gantt chart is a frequently employed visual representation of a project schedule. It serves as a type of bar chart that illustrates the commencement and conclusion dates of various project elements, including resources, planning activities, and interdependencies.

This charting method was originally devised by Henry Gantt (1861-1919), an American mechanical engineer, and is widely used in project management.

Key Takeaways:

- Gantt charts are valuable tools for scheduling, managing, and monitoring specific tasks and resources within a project.
- They consist of a task list and bars that depict the progress of each task.
- The horizontal bars, which vary in length, symbolize the project timeline and can convey information about task sequences, durations, as well as the start and end dates for each task.
- Gantt charts are the most commonly employed charts in the field of project management.
- They find application in various industries, including heavy construction projects like dam and bridge construction, highway development, as well as in software development and the creation of other products and services.

Understanding Gantt Charts:

Gantt charts are an indispensable tool for project planning and involve delineating the sequence of tasks required for project completion. Typically, these charts are displayed as horizontal bar charts. The length of each horizontal bar signifies the project's timeline,

encompassing task sequences, durations, and the commencement and conclusion dates for individual tasks. Furthermore, the extent of completion for each task is visually represented within these horizontal bars. A Gantt chart plays a crucial role in the scheduling, management, and tracking of particular tasks and resources within a project. This chart provides a visual representation of the project's timeline, encompassing both planned and accomplished work spanning a specific duration. It serves as a valuable communication tool for project managers to convey the progress and completion status of individual tasks within the project, thereby contributing to the project's overall adherence to its schedule. Conventionally, it is considered a standard instrument that fosters consistent communication within the engineering and project management sectors.

Benefits of Gantt Chart: Gantt charts offer several advantages in project management. They enable the identification of tasks that can run concurrently and those that are dependent on others for start or completion. This helps in uncovering potential bottlenecks and recognizing any tasks that might have been omitted from the project timeline. The chart also illustrates factors like task slack time, which represents extra time for task completion that won't affect the project's schedule, as well as noncritical activities that could be delayed without impacting the project and critical activities that must be completed as scheduled. Gantt charts are versatile tools suitable for managing projects of various sizes and types, ranging from large-scale infrastructure projects like dams and bridges to software development and other technological endeavors. Utilizing project management software such as Microsoft Visio, Project, SharePoint, Excel, or specialized applications like Gantto or Matchware can facilitate the creation and use of Gantt charts.

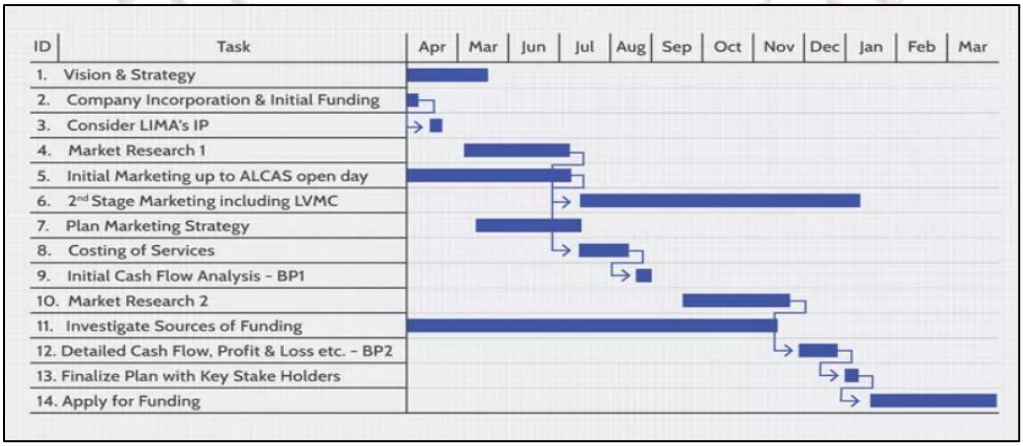


Fig. 36: Gantt Chart

Example of a Gantt Chart

Company A is embarking on a project to install new software on a large investment bank's servers. To effectively manage this project, Company A opts to create a Gantt chart, providing a visual representation of the project's different phases. These project tasks, including research, software selection, testing, and installation, are depicted as vertical lines on the Gantt chart. Notably, selecting the software serves as a pivotal milestone in the project. The entire project spans 40 days, with each task requiring 10 days for completion and a sequential dependency. An essential activity is the testing of the software in both development and test environments. The Gantt chart portrays task start and end dates, duration, and milestones as horizontal bars, complete with the percentage of work accomplished for each task. Leveraging the Gantt chart, Company A can effectively outline the project's timeline to achieve full completion within the agreed-upon 40 days.

Uses and Components of a Gantt chart:

Gantt charts serve as valuable tools to visualize concurrent tasks and projects within an organization, tracking their progress. Management employs them for planning and scheduling, ensuring optimal resource allocation and prioritizing projects based on their importance. Henry Gantt, a social scientist and management consultant, developed these charts in the early 1910s as part of scientific management, aiming to enhance corporate productivity and task monitoring.

Components:

A typical Gantt chart comprises three key components: tasks or activities along the y-axis, progress stages or milestones along the x-axis (either at the chart's top or bottom), and horizontal progress bars representing the current status of each task. In essence, a Gantt chart provides a visual representation of a project's timeline, showcasing the start and end dates of its components, aiding in project organization.

Let's dissect the fundamental elements of a Gantt chart to grasp their roles within a project plan:

1. Task List: Positioned vertically along the left side of the Gantt chart, this section delineates project tasks. It can be structured hierarchically, with the potential for grouping and sub-grouping.
2. Timeline: Extending horizontally across the top of the Gantt chart, the timeline displays time units such as months, weeks, days, and years.
3. Dateline: A vertical line within the Gantt chart that serves to highlight the present date.
4. Bars: Situated on the right side of the Gantt chart, these horizontal markers symbolize tasks. They convey essential information like progress, duration, and task start and end dates.
5. Milestones: Recognizable as yellow diamond symbols, milestones prominently denote significant project events, dates, decisions, and deliverables.
6. Dependencies: Visualized as light gray lines, dependencies connect tasks that must be completed in a specific sequence.
7. Progress: This section illustrates the advancement of work, often indicated by a percentage of completion and/or shading within the bars.
8. Resource Assignment: This element signifies the individuals or teams responsible for executing each task.

Who uses Gantt charts?

Gantt charts are used by project managers, business owners, team leaders, and executives in many different industries across various departments. Here are just a few types of companies and teams that use Gantt charts to plan, schedule, and execute their projects:

- Construction
- Consulting agencies
- Engineering firms
- Event planning
- Human resources
- Marketing teams
- Manufacturing
- Software developmentTop of Form

How to create Gantt Chart in Excel :

Unfortunately, Microsoft Excel does not provide a pre-made Gantt chart template. Nevertheless, you can efficiently generate a Gantt chart in Excel by utilizing its bar graph feature along with some formatting. Follow these steps meticulously, and you can create a basic Gantt chart in less than three minutes. This Gantt chart demonstration is based on Excel 2010, but the process remains consistent across Excel versions from 2013 to Excel 365.

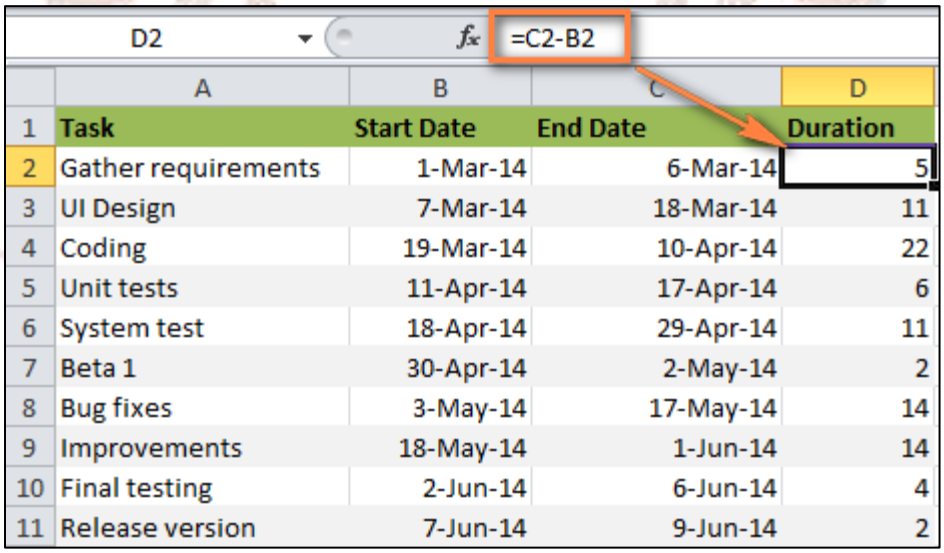
1. Create a project table

You start by entering your project's data in an Excel spreadsheet. List each task is a separate row and structure your project plan by including the *Start date, End date* and *Duration,* i.e. the number of days required to complete the tasks.

Tip. Only the *Start date* and *Duration* columns are necessary for creating an Excel Gantt chart. If you have *Start Dates* and *End Dates,* you can use one of these simple formulas to calculate *Duration,* whichever makes more sense for you:

Duration = End Date - Start Date

Duration = End date - Start date + 1



	A	B	C	D
1	Task	Start Date	End Date	Duration
2	Gather requirements	1-Mar-14	6-Mar-14	5
3	UI Design	7-Mar-14	18-Mar-14	11
4	Coding	19-Mar-14	10-Apr-14	22
5	Unit tests	11-Apr-14	17-Apr-14	6
6	System test	18-Apr-14	29-Apr-14	11
7	Beta 1	30-Apr-14	2-May-14	2
8	Bug fixes	3-May-14	17-May-14	14
9	Improvements	18-May-14	1-Jun-14	14
10	Final testing	2-Jun-14	6-Jun-14	4
11	Release version	7-Jun-14	9-Jun-14	2

Fig. 37: Creating Helper Columns for Visualising Data

2. Make a standard Excel Bar chart based on Start date

You begin making your Gantt chart in Excel by setting up a usual *Stacked Bar chart.*

- Select a range of your **Start Dates** with the column header, it's B1:B11 in our case. Be sure to select only the cells with data, and not the entire column.
- Switch to the *Insert tab > Charts* group and click **Bar.**
- Under the *2-D Bar section*, click **Stacked Bar.**

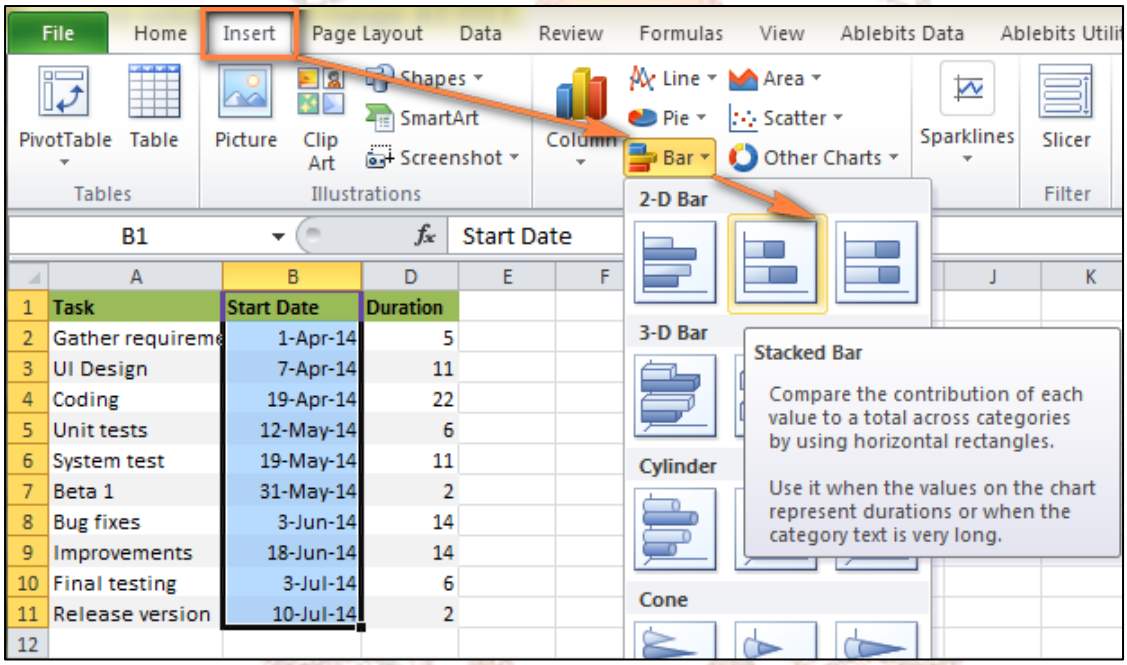


Fig 38: Choosing the Chart Type

As a result, you will have the following Stacked bar added to your worksheet:

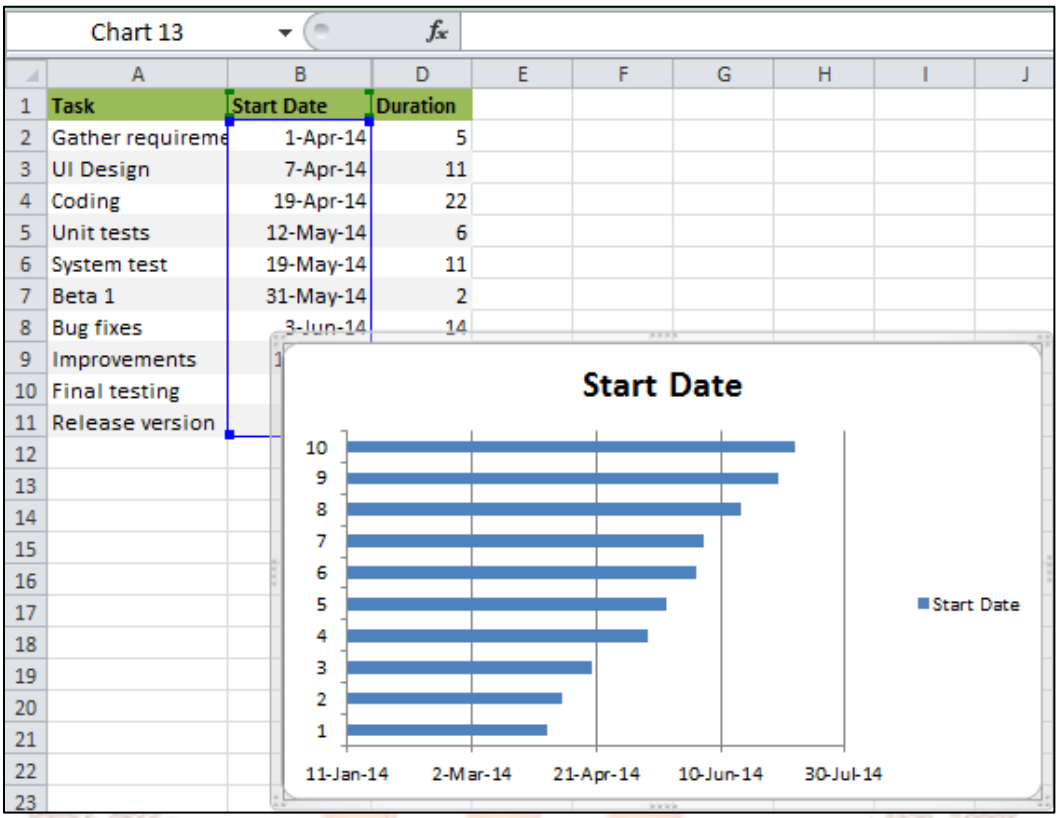


Fig 39: Output of the Bar Chart

Note. Some other Gantt Chart tutorials you can find on the web recommend creating an empty bar chart first and then populating it with data as explained in the next step. But I think the above approach is better because Microsoft Excel will add one data series to the chart automatically, and in this way save you some time.

3. Add Duration data to the chart

Now you need to add one more series to your Excel Gantt chart-to-be.

1. Right-click anywhere within the chart area and choose **Select Data** from the context menu.

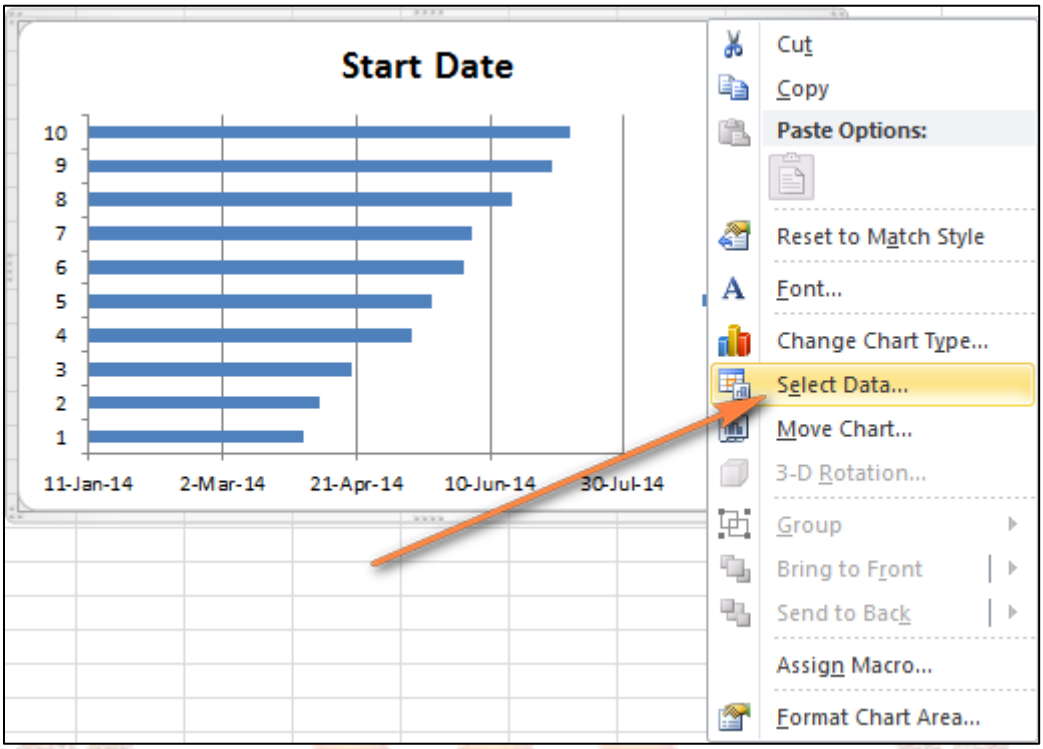


Fig 40: Updating the Data Selection

The **Select Data Source** window will open. As you can see in the screenshot below, *Start Date* is already added under **Legend Entries (Series).** And you need to add D*uration* there as well.

2. Click the **Add** button to select more data (Duration) you want to plot in the Gantt chart.

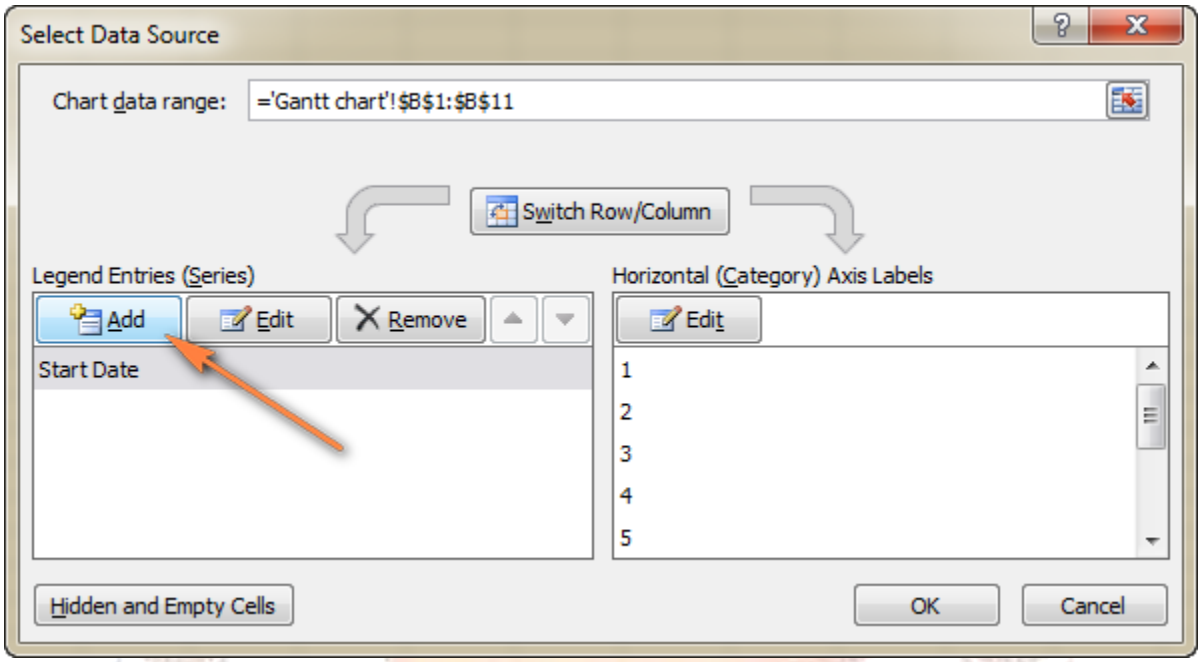


Fig 41: Window to Select Data

3. The *Edit Series* window opens and you do the following:
   - In the "Series name" field, input "Duration" or a name of your preference. Alternatively, you can click on a column header in your spreadsheet, and the clicked header will automatically become the Series name for your Gantt chart.
   - Click the range selection icon next to the **Series Values** field.

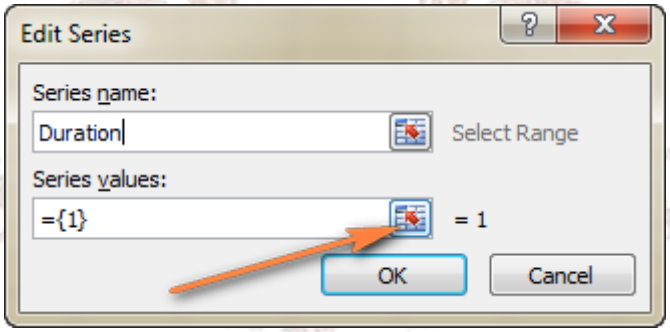


Fig 42: Select Data Series

4. A small "Edit Series" window will pop up. To choose your project's duration data, click on the first duration cell (D2 in our example) and drag the mouse down to the final duration cell (D11). Ensure that you haven't accidentally included the header or any empty cells.

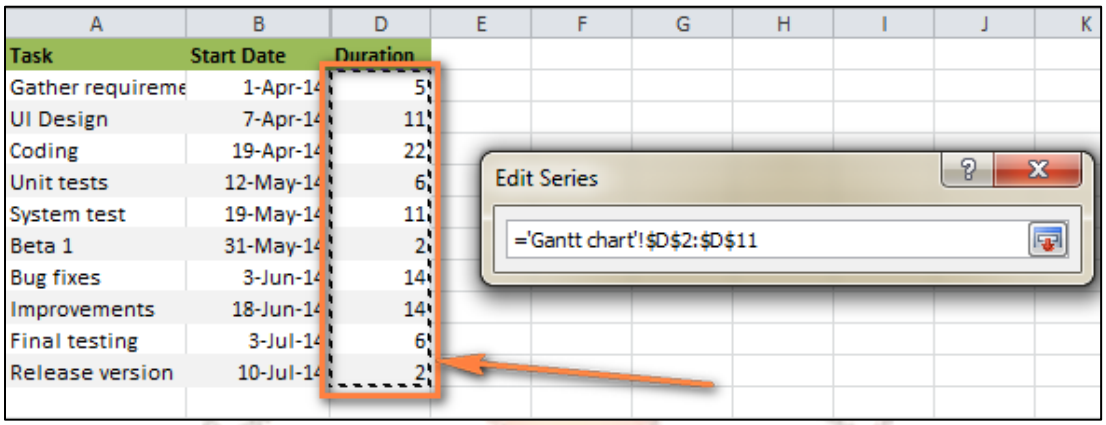


Fig. 43: Select Data

5. Click the Collapse Dialog icon to exit this small window. This will bring you back to the previous *Edit Series* window with *Series name* and *Series values* filled in, where you click *OK.*

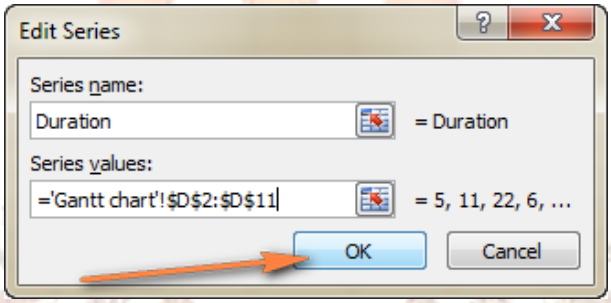


Fig. 44: Finalising the Selected Data

6. Now you are back at the *Select Data Source* window with both *Start Date* and *Duration* added under **Legend Entries (Series).** Simply click *OK* for the Duration data to be added to your Excel chart.

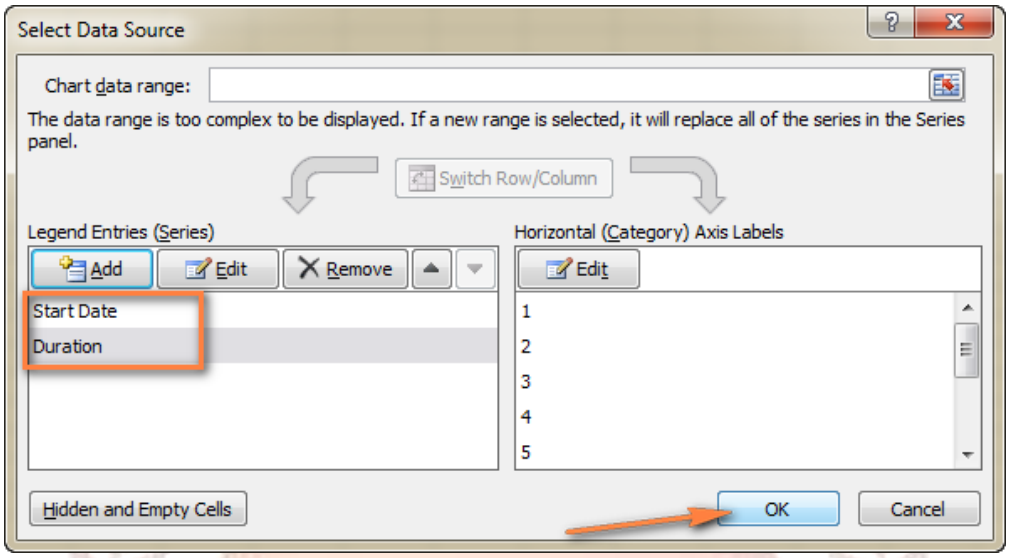


Fig 45: Select Okay Button to finalise the Data Selection

The resulting bar chart should look similar to this:

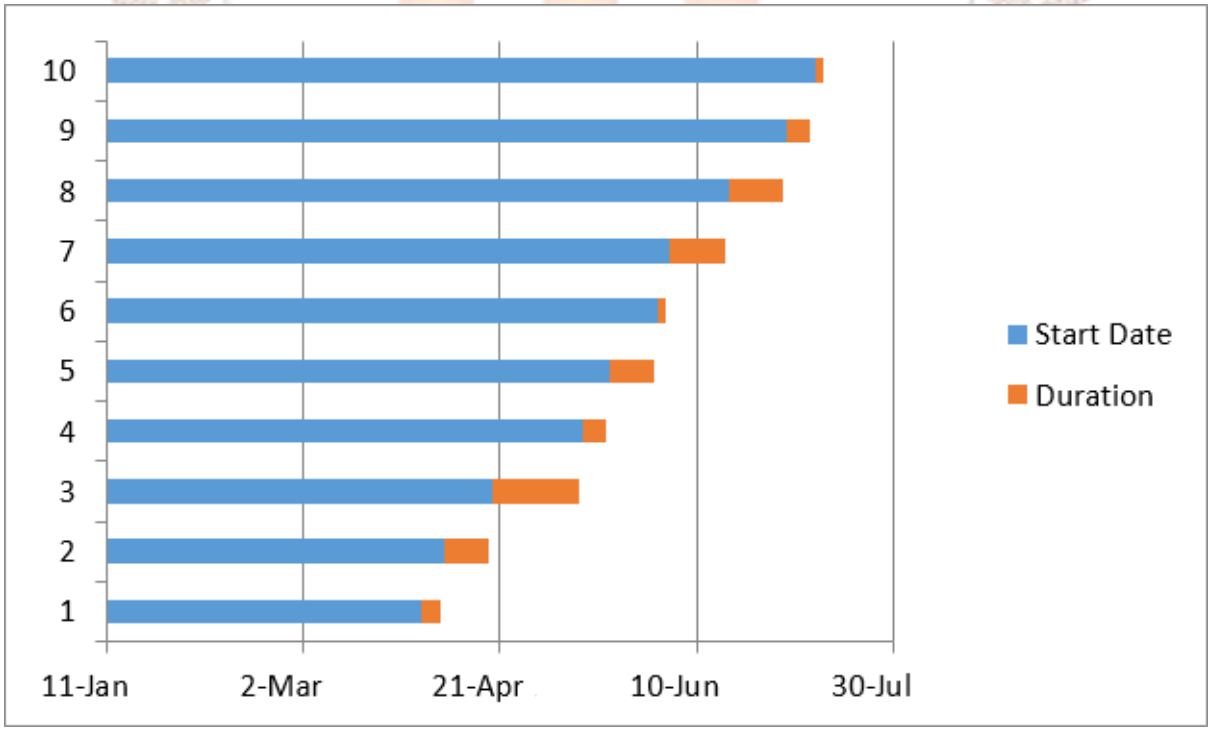


Fig 46: Output of updated chart

7. Add task descriptions to the Gantt chart

Now you need to replace the days on the left side of the chart with the list of tasks.

8. Right-click anywhere within the chart plot area (the area with blue and orange bars) and click **Select Data** to bring up the *Select Data Source* window again.
9. Make sure the **Start Date** is selected on the left pane and click the **Edit** button on the right pane, under *Horizontal (Category)* Axis Labels.

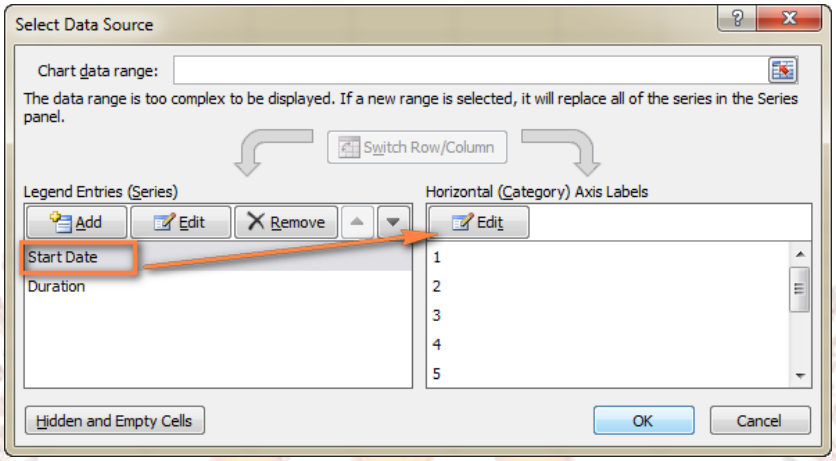


Fig 47: Swap the values

10. A small *Axis Label* window opens and you select your tasks in the same fashion as you selected Durations in the previous step - click the range selection icon , then click on the first task in your table and drag the mouse down to the last task. Remember, the column header should not be included. When done, exit the window by clicking on the range selection icon again.

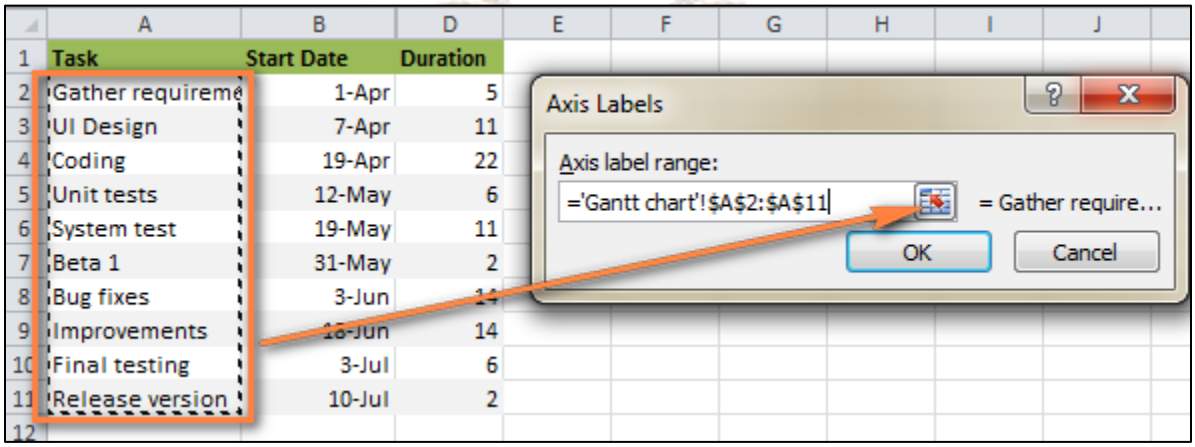


Fig. 48: Swap Date Values with Task Values

11. Click *OK* twice to close the open windows.
12. Remove the chart labels block by right-clicking it and selecting *Delete* from the context menu.

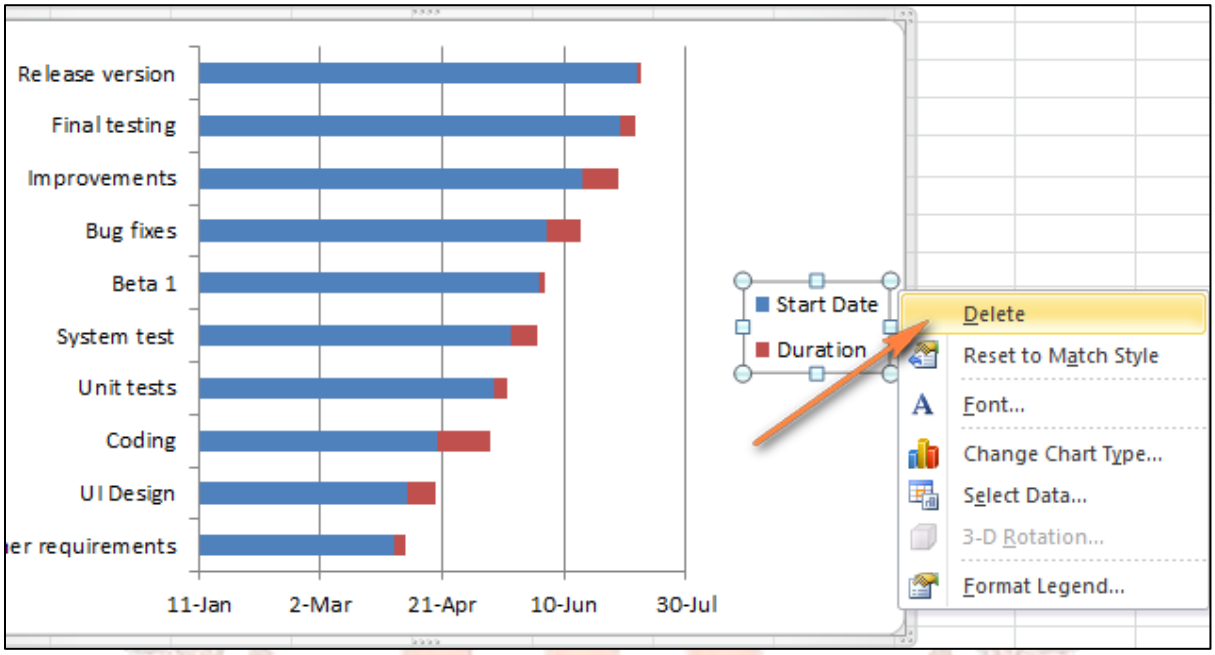


Fig. 49: Remove Chart Labels

At this point your Gantt chart should have task descriptions on the left side and look something like this:

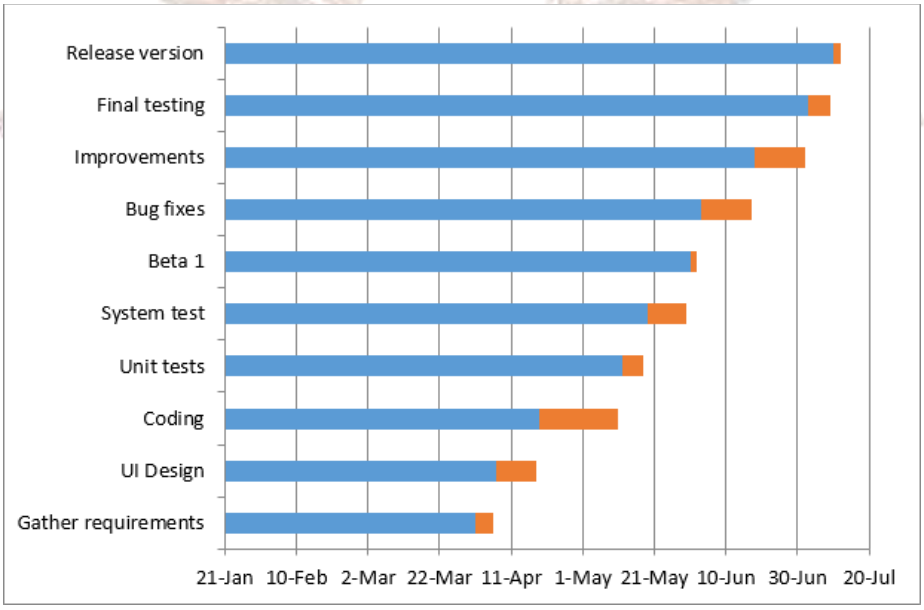


Fig. 50: Output after Removing Chart Labels and settings done above

13. Transform the bar graph into the Excel Gantt chart

What you have now is still a stacked bar chart. You have to add the proper formatting to make it look more like a Gantt chart. Our goal is to remove the blue bars so that only the orange parts representing the project's tasks will be visible. In technical terms, we won't really delete the blue bars, but rather make them transparent and therefore invisible.

14. Click on any **blue bar** in your Gantt chart to select them all, right-click and choose **Format Data Series** from the context menu.

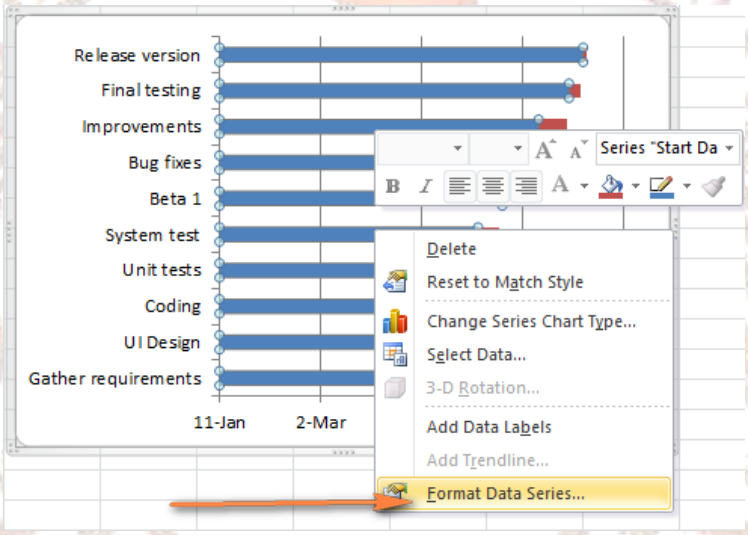


Fig 51: Format Data Series

15. The *Format Data Series* window will show up and you do the following:
    - Switch to the *Fill tab* and select **No Fill.**
    - Go to the *Border Color* tab and select **No Line.**

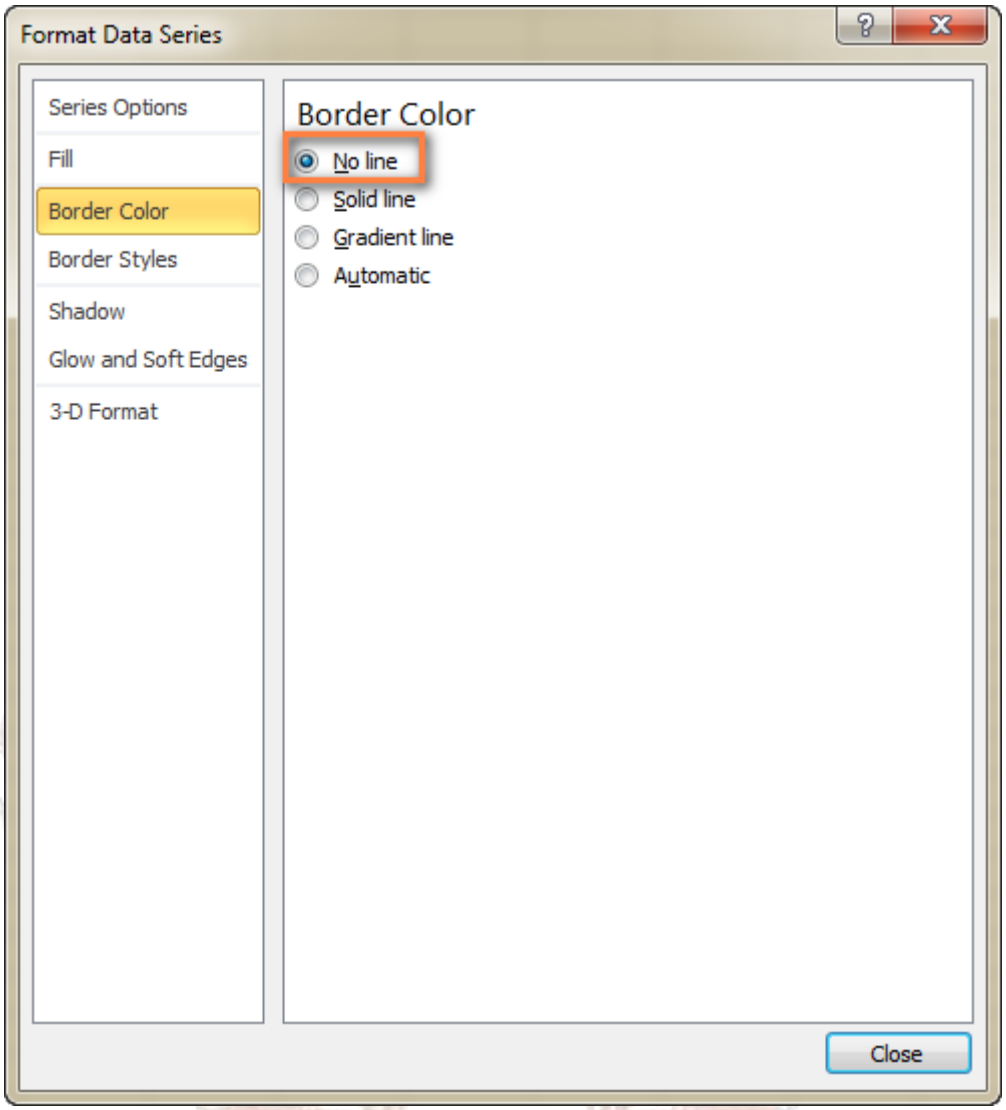


Fig 52: Formatting Data 1

Note. You do not need to close the dialog because you will use it again in the next step.

16. As you have probably noticed, the tasks on your Excel Gantt chart are listed in **reverse order.** And now we are going to fix this. Click on the list of tasks in the left-hand part of your Gantt chart to select them. This will display the *Format Axis* dialog for you. Select the **Categories in reverse order option** under *Axis Options* and then click the *Close* button to save all the changes.

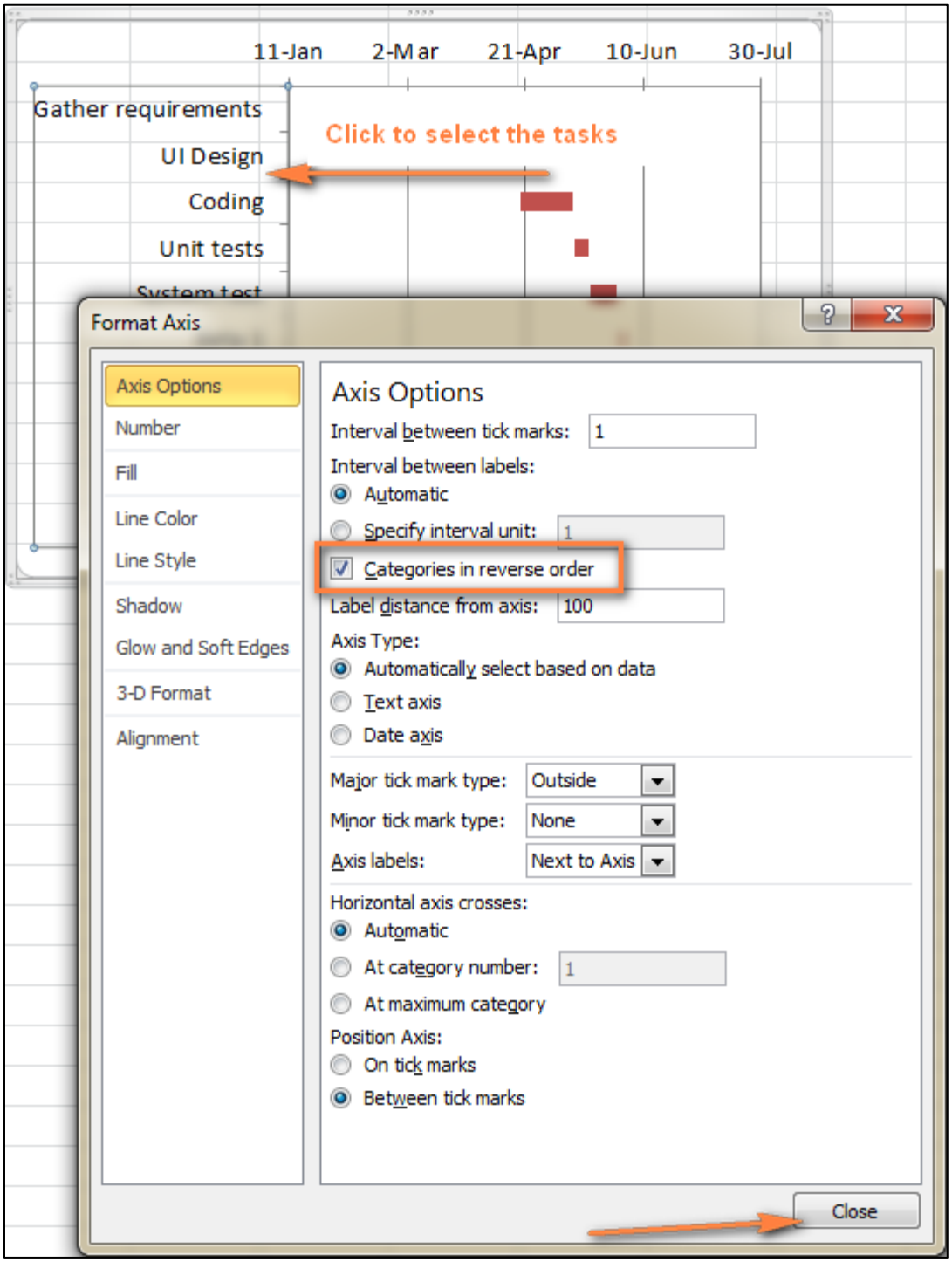


Fig 53: Formatting Data 2

The results of the changes you have just made are:

- Your tasks are arranged in a proper order on a Gantt chart.
- Date markers are moved from the bottom to the top of the graph.

Your Excel chart is starting to look like a normal Gantt chart, isn't it? For example, my Gantt diagram looks like this now:

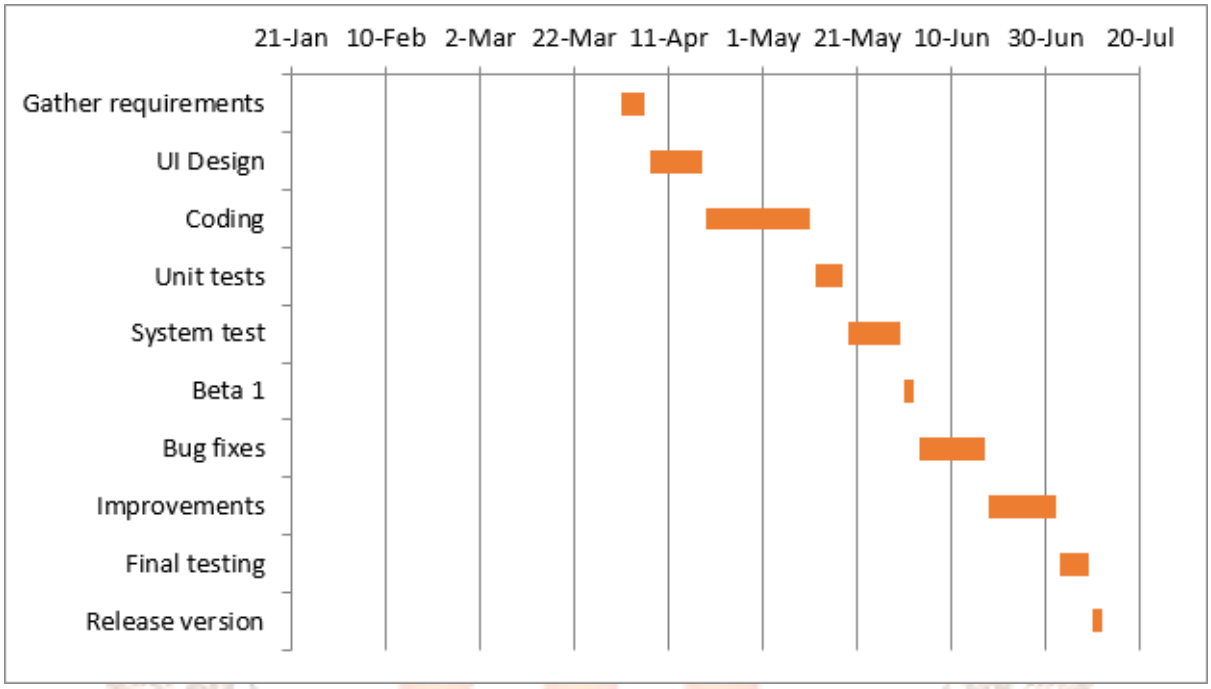


Fig 54: Output of Formatting

17. Improve the design of your Excel Gantt chart

Though your Excel Gantt chart is beginning to take shape, you can add a few more finishing touches to make it really stylish.

18. **Remove the empty space on the left side of the Gantt chart.** As you may recall, initially, the blue bars representing the start dates were positioned at the beginning of your Excel Gantt chart. To eliminate the empty white space and bring your tasks closer to the left vertical axis, follow these steps:
19. Right-click on the first Start Date entry in your data table.
20. Choose "Format Cells" and then select "General."
21. Take note of the number displayed, which represents the date numerically (e.g., 41730 in my case). Excel stores dates as numbers based on the count of days since 1-Jan-1900.
22. Click "Cancel" because you don't need to make any actual changes here.

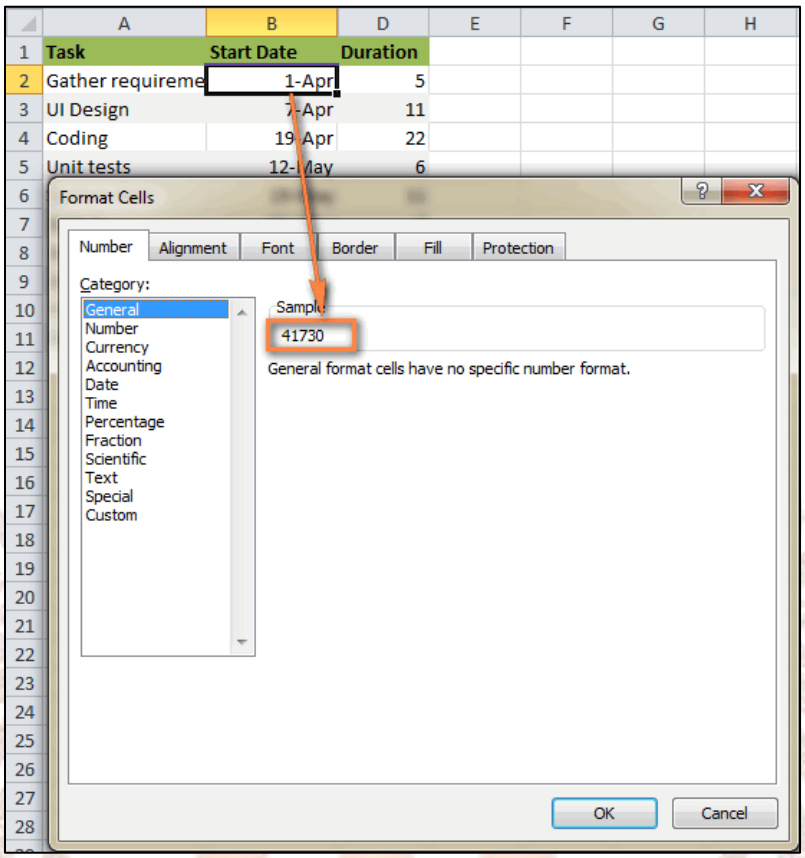


Fig 55: Updating Data Type

23. Click on any date above the task bars in your Gantt chart. One click will select all the dates, you right click them and choose **Format Axis** from the context menu.

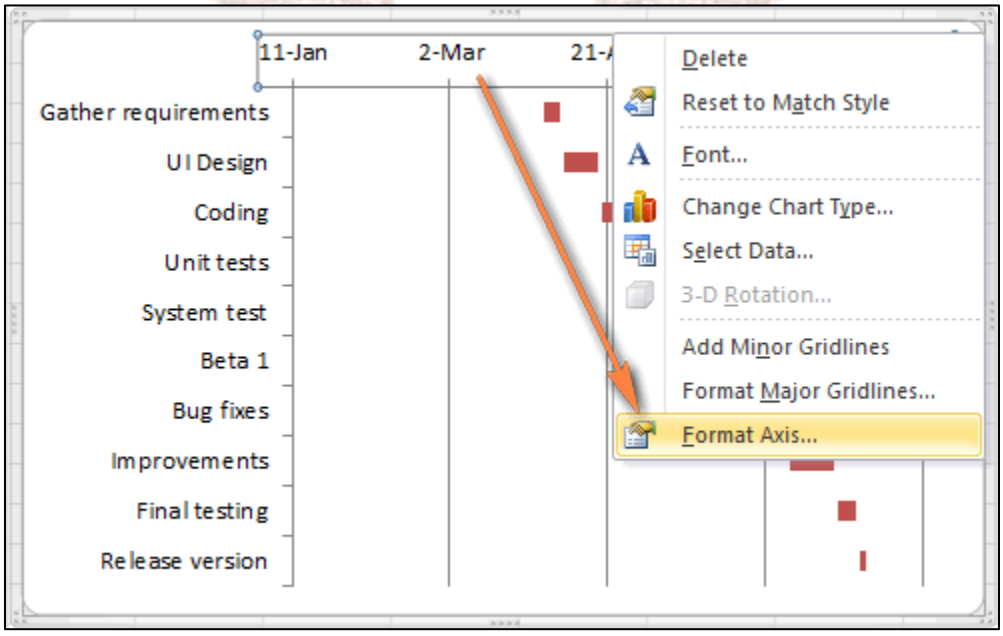


Fig 56: Format Axis

24. Under *Axis Options*, change **Minimum** to **Fixed** and type the number you recorded in the previous **step.**
25. **Adjust the number of dates on your Gantt chart.** In the same *Format Axis* window that you used in the previous step, change **Major unit** and **Minor unit** to Fixed too, and then add the numbers you want for the date intervals. Typically, the shorter your project's timeframe is, the smaller numbers you use. For example, if you want to show every other date, enter 2 in the *Major unit.* You can see the settings in the screenshot below.

Note. In Excel 365, Excel 2021 - 2013, there are no *Auto* and *Fixed* radio buttons, so you simply type the number in the box.

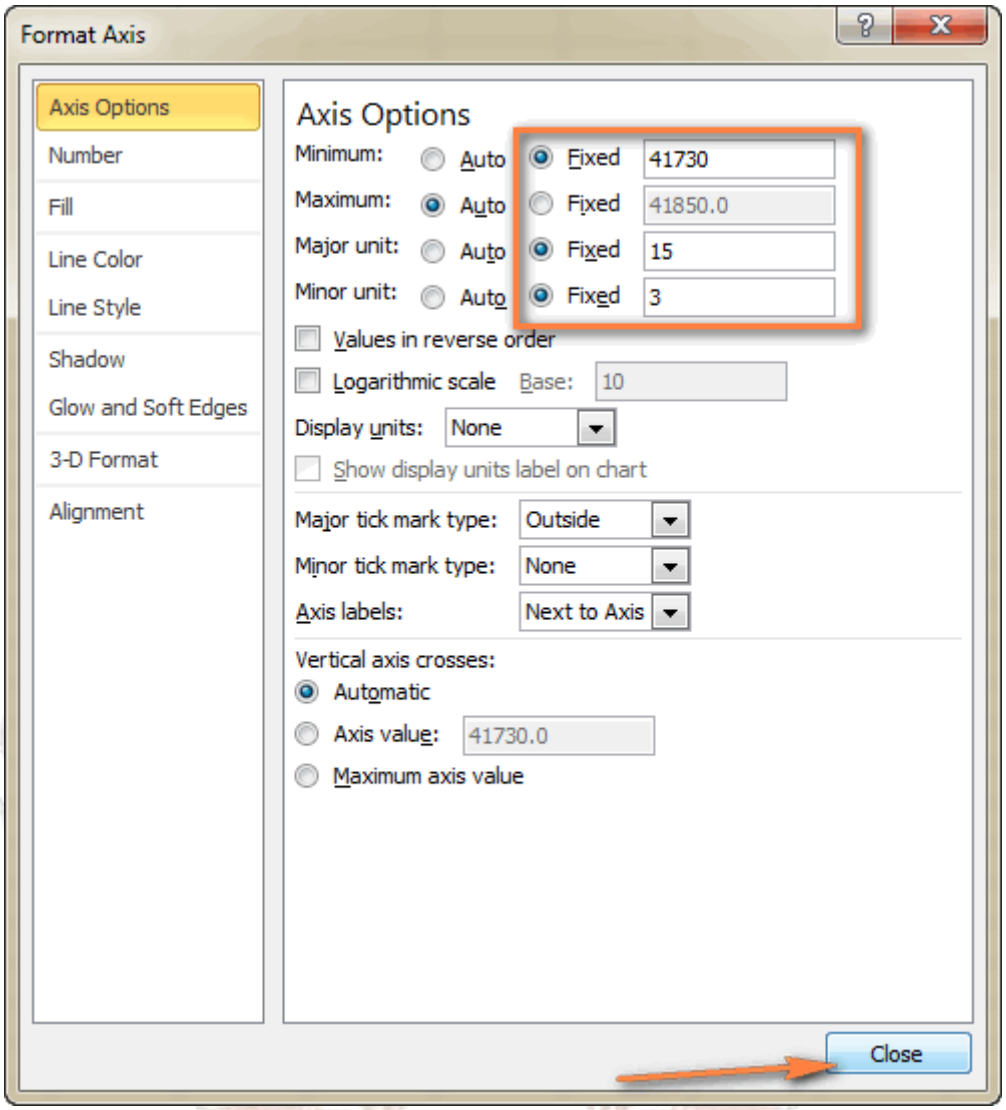


Fig 57: Updating Axis Format

Tip. You can play with different settings until you get the result that works best for you. Don't be afraid to do something wrong because you can always revert to the default settings by switching back to Auto in Excel 2010 and 2007, or click Reset in Excel 2013 and later.

26. **Remove excess white space between the bars.** Compacting the task bars will make your Gantt graph look even better.
    - Click any of the orange bars to get them all selected, right click and select **Format Data Series.**
    - In the Format Data Series dialog, set **Separated** to **100%** and **Gap Width** to **0%** (or close to 0%).

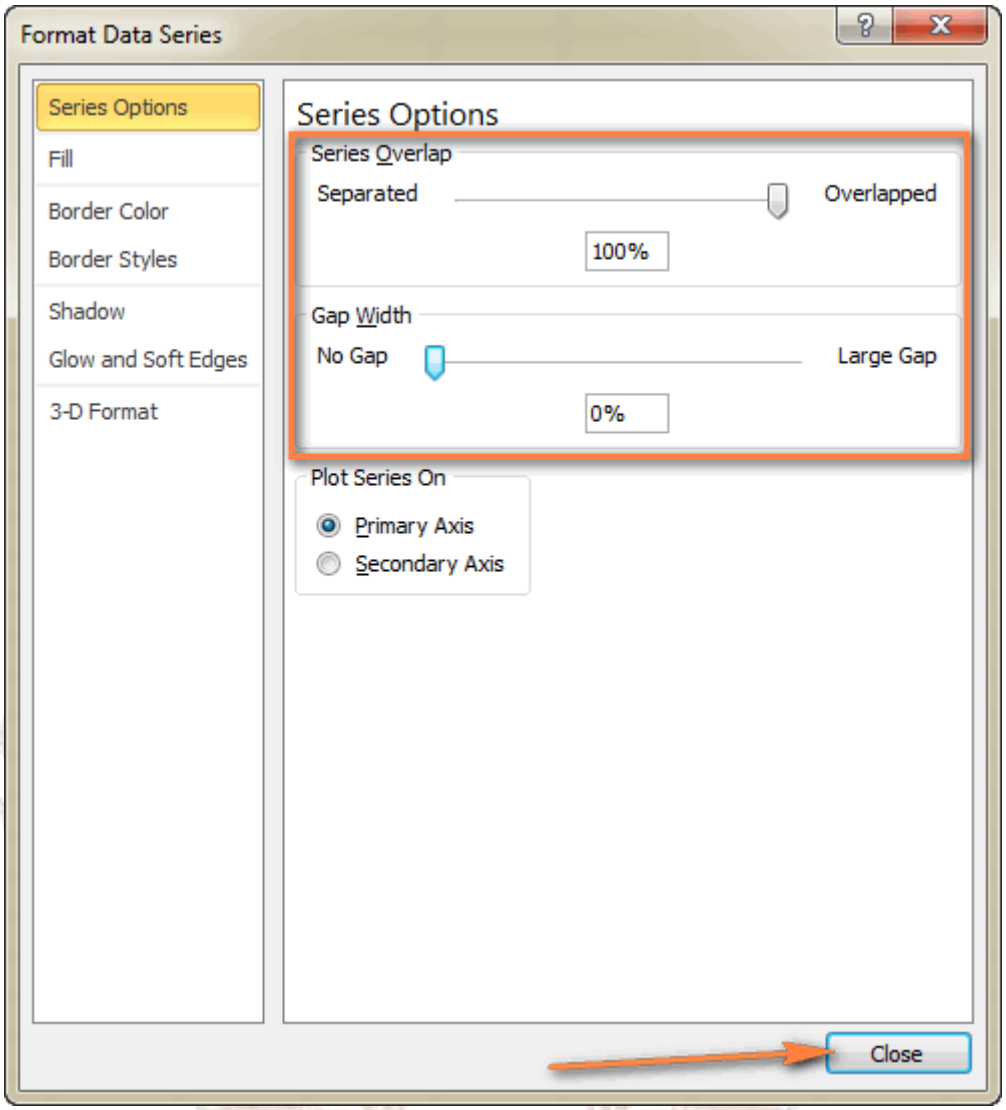


Fig 58: Update Series Formating

And here is the result of our efforts - a simple but nice-looking Excel Gantt chart:

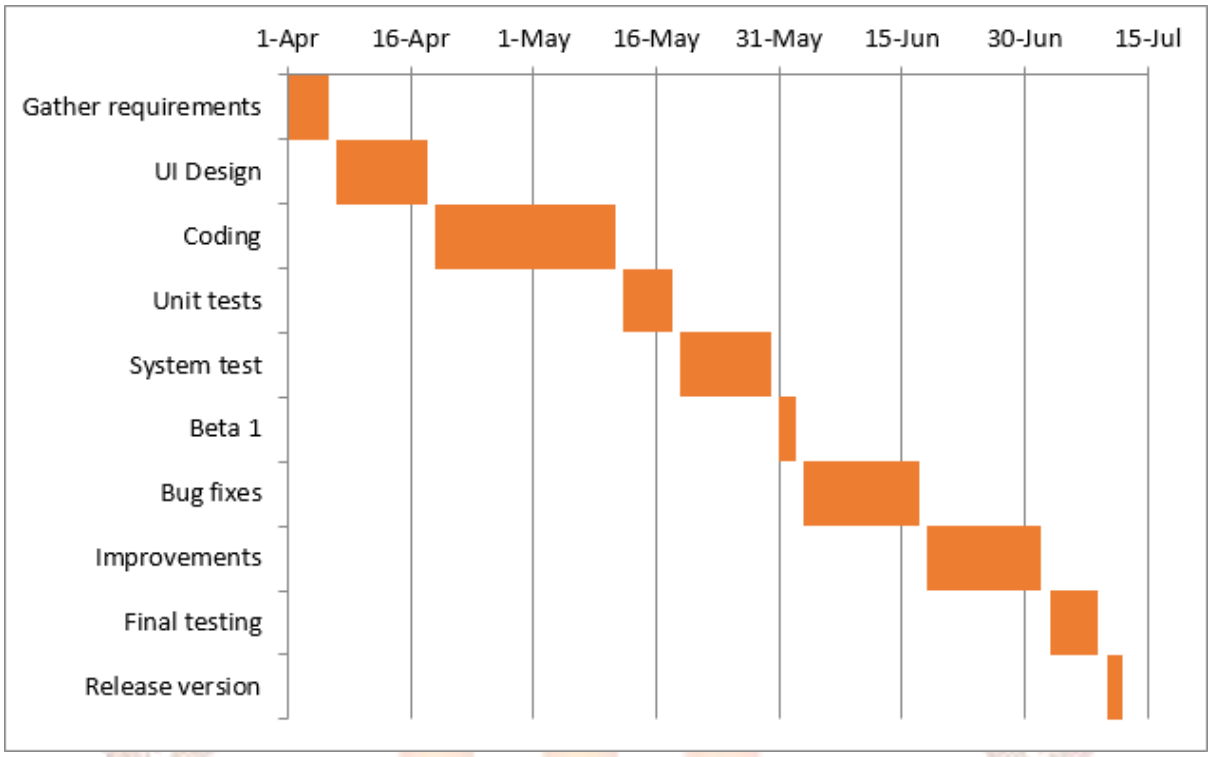


Fig 59: Output Gantt Chart

Remember, though your Excel chart simulates a Gantt diagram very closely, it still keeps the main features of a standard Excel chart:

- Your Excel Gantt chart will resize when you add or remove tasks.
- You can change a Start date or Duration; the chart will reflect the changes and adjust automatically.
- You can save your Excel Gantt chart as an image or convert to HTML and publish online.

Tips: You can design your Excel Gant chart in different ways by changing the fill colors, border colors, shadow and even applying the 3-D format. All these options are available in the **Format Data Series** window (right-click the bars in the chart area and select *Format Data Series* from the context menu).

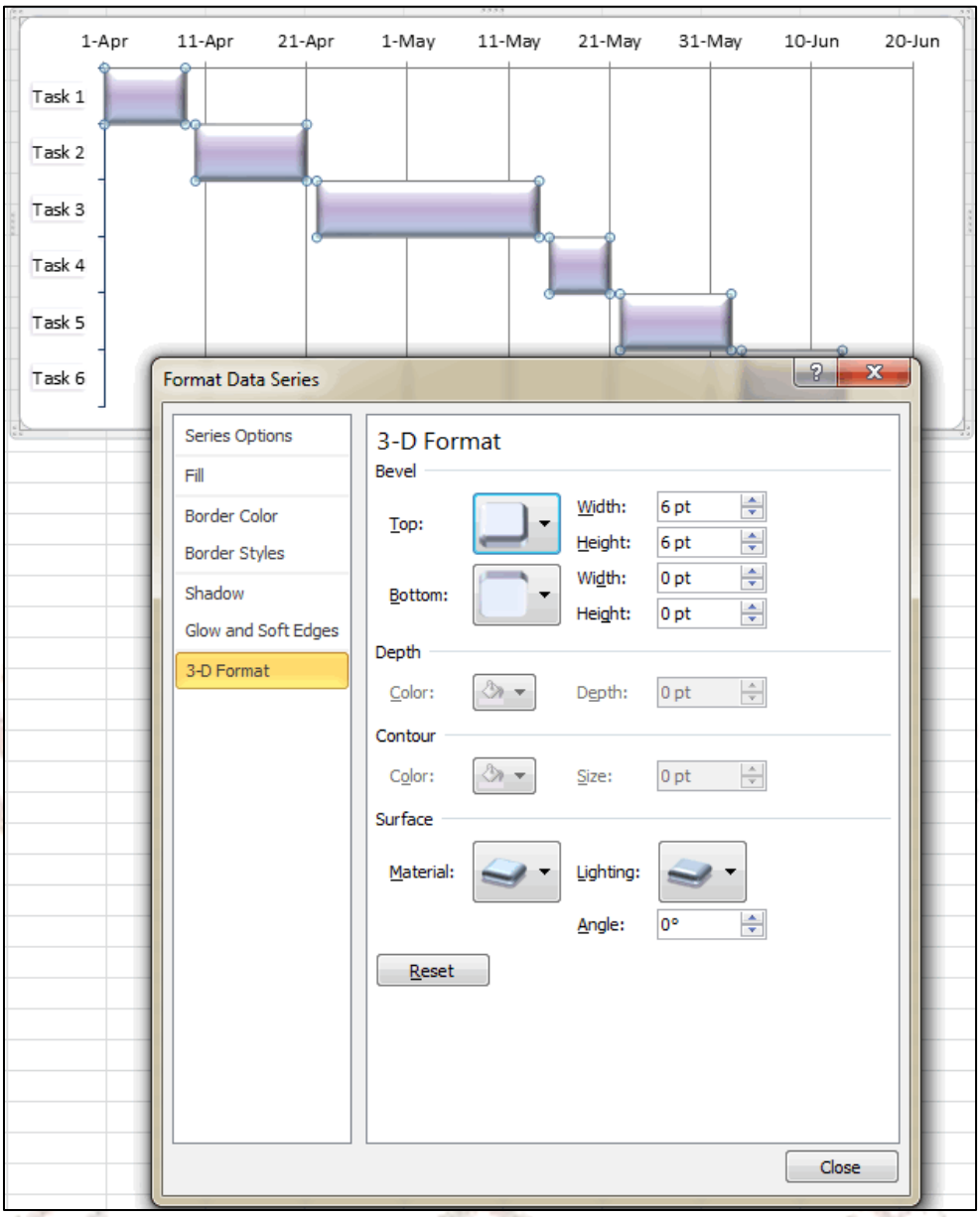


Fig 60: Saving Gantt Chart as template

- When you have created an awesome design, it might be a good idea to save your Excel Gantt chart as a template for future use. To do this, click the chart, switch to the *Design* tab on the ribbon and click **Save as Template.**

4. HEAT MAPS

A heat map, also known as a heatmap, is a visual representation of numerical data where varying values are presented using different colors. Typically, heatmaps use a color scheme that transitions from warm to cool tones, representing data as hot and cold spots.

In comparison to standard analytical reports, heatmaps offer a more intuitive way to visualize and analyze complex data. They are widely utilized by scientists, analysts, and marketers for initial data analysis and the identification of common patterns.

Here are a few common examples:

1. Air Temperature Heat Map: This type of heatmap is used to visualize air temperature data within a specific region. It allows for a quick assessment of temperature variations across the area.
2. Geographical Heat Map: Geographical heatmaps display numerical data over a geographic area using different shades of color. They provide insights into how data values vary across different geographical regions.
3. Risk Management Heat Map: Risk management heatmaps visually represent various risks and their potential impacts in a concise and accessible manner. They help stakeholders understand and prioritize risks.

In Excel, a heat map is created to represent individual cells using various color codes based on their respective values. For instance, in the heatmap shown below, you can easily identify the wettest regions (highlighted in green) and the driest regions (highlighted in red) over different decades at a glance:

	A	B	C	D	E	F	G	H	I
1	**Heat Map in Excel**								
2	*Average precipitation by decade, inch*								
3	**City**	**1940**	**1950**	**1960**	**1970**	**1980**	**1990**	**2000**	**2010**
4	Portland	38.37	41.21	40.82	46.54	45.70	45.71	50.33	48.95
5	Burlington	33.38	34.26	31.04	35.76	36.61	35.74	37.96	39.60
6	Albany	32.94	36.13	30.90	40.10	37.44	38.23	42.26	41.86
7	Philadelphia	41.40	40.84	38.06	45.36	40.85	39.91	43.74	48.49
8	Washington	41.95	39.56	36.47	40.99	38.40	38.65	42.09	43.21
9	Lynchburg	41.38	38.89	34.70	46.15	41.86	41.82	40.92	42.73
10	Savannah	51.36	49.06	51.16	48.92	47.63	52.29	43.91	47.82
11	Raleigh	46.31	43.09	39.59	42.62	42.08	44.43	43.41	49.73
12	Nashville	46.31	46.85	45.68	53.49	42.76	47.96	50.97	52.57
13	Atlanta	49.63	44.40	51.08	50.39	50.84	49.37	48.85	51.47
14	Pensacola	68.18	61.37	61.40	61.79	63.32	68.14	64.31	71.99
15	Birmingham	53.25	49.24	56.65	57.59	49.55	54.83	56.73	56.58
16	Miami	60.14	59.58	59.65	53.93	54.72	67.46	63.57	69.38
17	Buffalo	35.08	38.03	34.06	40.41	41.30	39.94	40.08	42.15
18	Pittsburgh	38.00	37.05	33.86	38.41	38.32	36.77	39.37	42.97
19	Detroit	32.21	31.25	30.72	31.49	34.64	32.65	33.05	37.00
20	Columbus	37.41	35.01	36.67	39.33	38.28	37.89	41.63	44.49
21	Louisville	43.83	40.43	42.99	47.23	43.00	47.23	50.09	53.69
22	Madison	30.23	31.55	29.13	31.68	31.78	35.42	36.19	39.78
23	Chicago	34.35	33.79	35.26	37.32	38.63	34.41	37.62	41.78
24	Indianapolis	39.07	39.16	37.94	40.24	41.70	40.91	44.61	45.32

Fig 61: Heat map

4.1 How to create a heat map in Excel

1. If you were considering the manual task of individually coloring each cell based on its value, it's advisable to abandon that approach as it would be both time-consuming and unnecessary. Firstly, it would require significant effort to assign the appropriate color shade for each value based on its rank. Secondly, you'd need to repeat the color-coding process whenever the values change. Excel's conditional formatting feature offers an efficient solution to overcome these challenges.
2. To create a heat map in Excel, we will utilize conditional formatting with color scales. Below are the steps to follow: Select your dataset. In our case, it's B3:M5.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	*Air Temperature (°F)*												
2		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
3	Avg.	40.2	41.8	47.4	56.3	65	71.3	78.5	79.5	73.6	64.3	54.2	44.4
4	Min.	35.6	36.5	41.1	49.6	58.8	66.5	73.8	74.8	69.2	59.7	49.5	40
5	Max.	45.8	48.4	55	64.3	72.3	77.2	84.2	85.3	78.9	69.8	59.8	49.7

Fig. 62: Selecting Data

3. Navigate to the Home tab, located in the Styles group, and click on Conditional Formatting, then select Color Scales. From the available options, choose the color scale

that best suits your needs. Excel will provide a live preview of how the selected color scale will appear in your dataset as you hover your mouse over it.

For this example, we've chosen *Red - Yellow - Green* color scale:

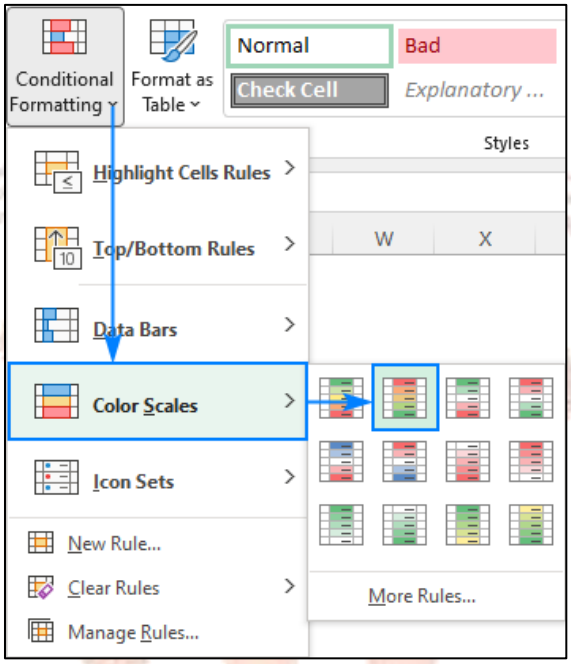


Fig 63: Conditional Formatting

In the result, you will have the high values highlighted in red, middle in yellow, and low in green. The colors will adjust automatically when the cell values change.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		*Air Temperature (°F)*											
2		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
3	Avg.	40.2	41.8	47.4	56.3	65	71.3	78.5	79.5	73.6	64.3	54.2	44.4
4	Min.	35.6	36.5	41.1	49.6	58.8	66.5	73.8	74.8	69.2	59.7	49.5	40
5	Max.	45.8	48.4	55	64.3	72.3	77.2	84.2	85.3	78.9	69.8	59.8	49.7

Fig. 64: Output Heat Map

Tip. For the conditional formatting rule to apply to new data automatically, you can convert your data range to a fully-functional Excel table.

4.2 Make a Heatmap with a Custom Color Scale

When applying a preset color scale, it depicts the lowest, middle and highest values in the predefined colors (green, yellow and red in our case). All the remaining values get different shades of the three main colors.

1. If you wish to highlight cells based on whether they are lower or higher than a specific number, regardless of their values, you can create a custom color scale instead of using the built-in options. Here's how to do it:

On the *Home tab*, in the *Styles* group, click *Conditional Formatting* > *Color Scales* > *More* Rules.

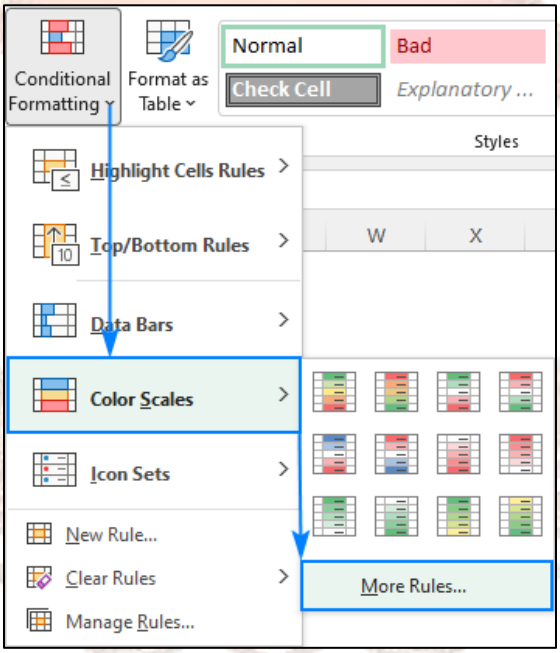


Fig 65: Conditional Fomratting

2. In the *New Formatting Rule* dialog box, do the following:
   - Pick **3-Color scale** from the *Format Style drop* down list.
   - For *Minimum* and/or *Maximum* value, choose **Number** in the *Type* drop down, and enter the desired values in the corresponding boxes.
   - For *Midpoint,* you can set either *Number or Percentile* (normally, 50%).
   - Assign a color to each of the three values.

For this example, we've configured the following settings:

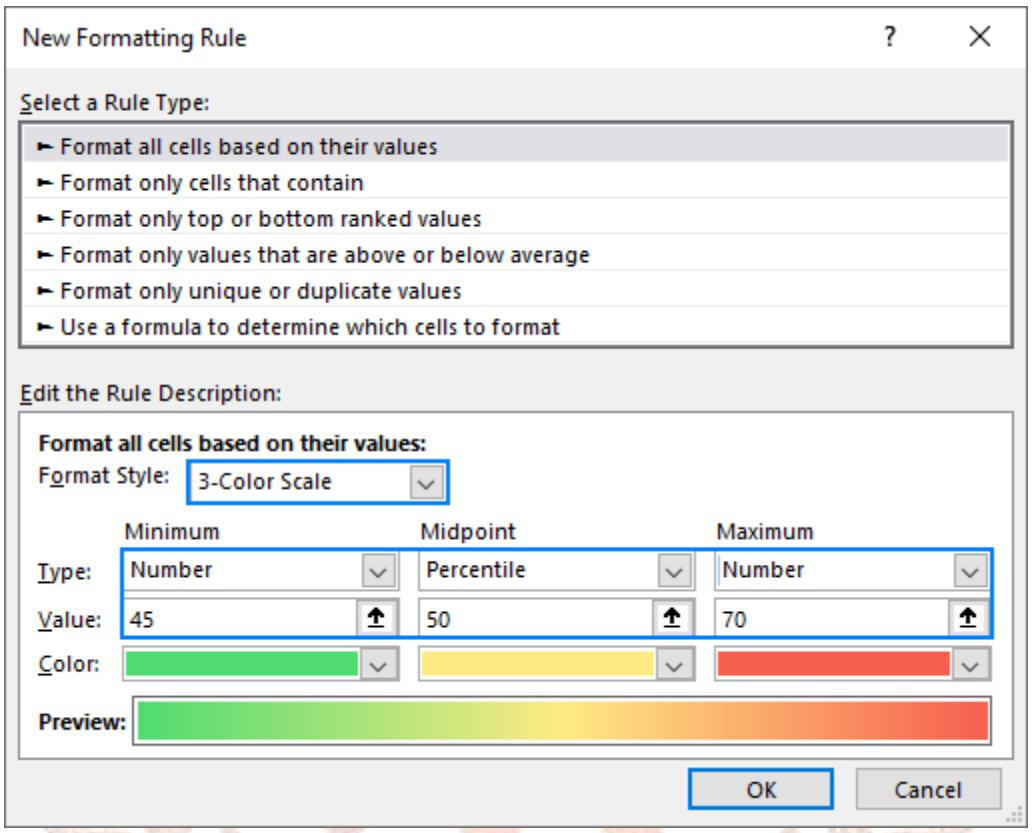


Fig 66: Updating Colour Scales

In this custom heatmap, all the temperatures below 45 °F are highlighted in the same shade of green and all the temperatures above 70 °F in the same shade of red:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	*Air Temperature (°F)*												
2		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
3	Avg.	40.2	41.8	47.4	56.3	65	71.3	78.5	79.5	73.6	64.3	54.2	44.4
4	Min.	35.6	36.5	41.1	49.6	58.8	66.5	73.8	74.8	69.2	59.7	49.5	40
5	Max.	45.8	48.4	55	64.3	72.3	77.2	84.2	85.3	78.9	69.8	59.8	49.7

Fig 67: Updated Heat Map

Create a heat map in Excel without numbers

The heat map you generate in Excel relies on the actual cell values, and deleting these values would result in the loss of the heat map. If you want to hide the cell values without deleting

them from the sheet, you can achieve this using custom number formatting. Here's how to do it step by step:

1. Select the heat map.
2. Press Ctrl + 1 to open the *Format Cells* dialog.
3. On the Number tab, under *Category,* select *Custom.*
4. In the *Type* box, type 3 semicolons (;;;).
5. Click OK to apply the custom number format.

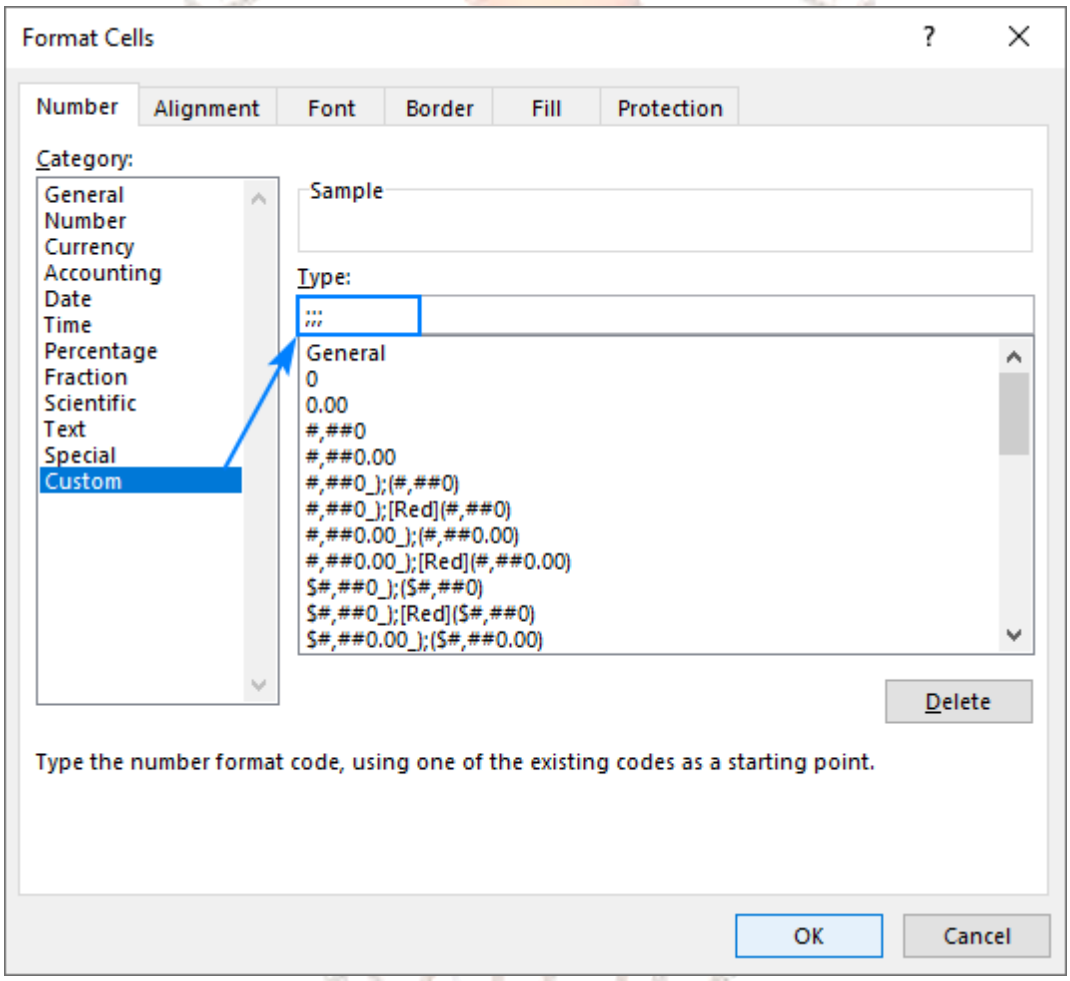


Fig 68: Create Custom Data Format

That's it! Now, your Excel heat map displays only the color-codes without numbers:

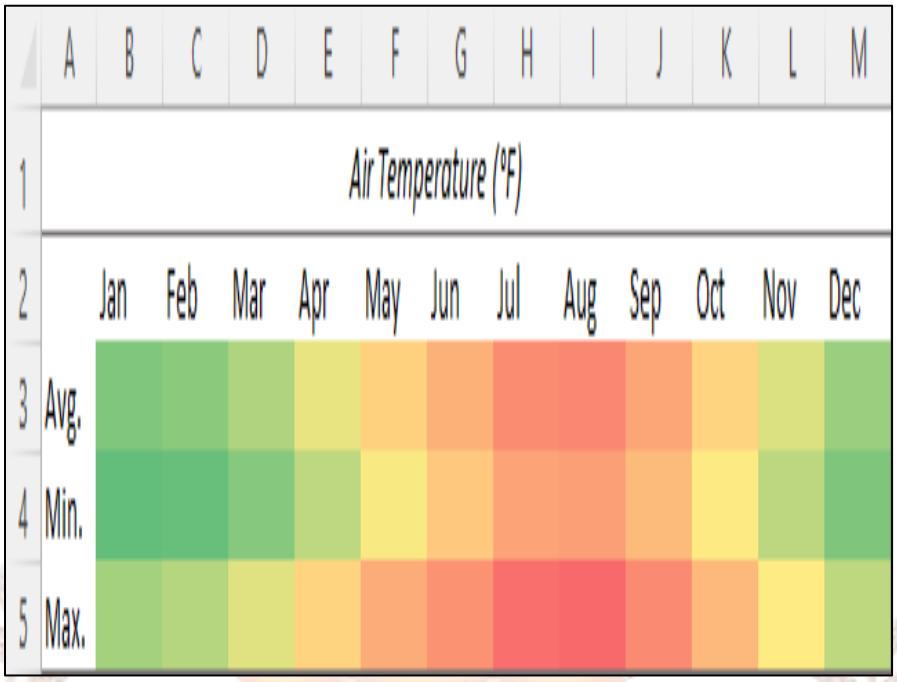


Fig 69: Output

4.3 Excel Heat Map with Square Cells

Another improvement you can make to your heatmap is perfectly square cells. Below is the fastest way to do this without any scripts or VBA codes:

1. **Align column headers vertically.** To prevent column headers from getting cut off, change their alignment to vertical. This can be done with the help of the *Orientation* button on the *Home tab,* in the *Alignment group:*

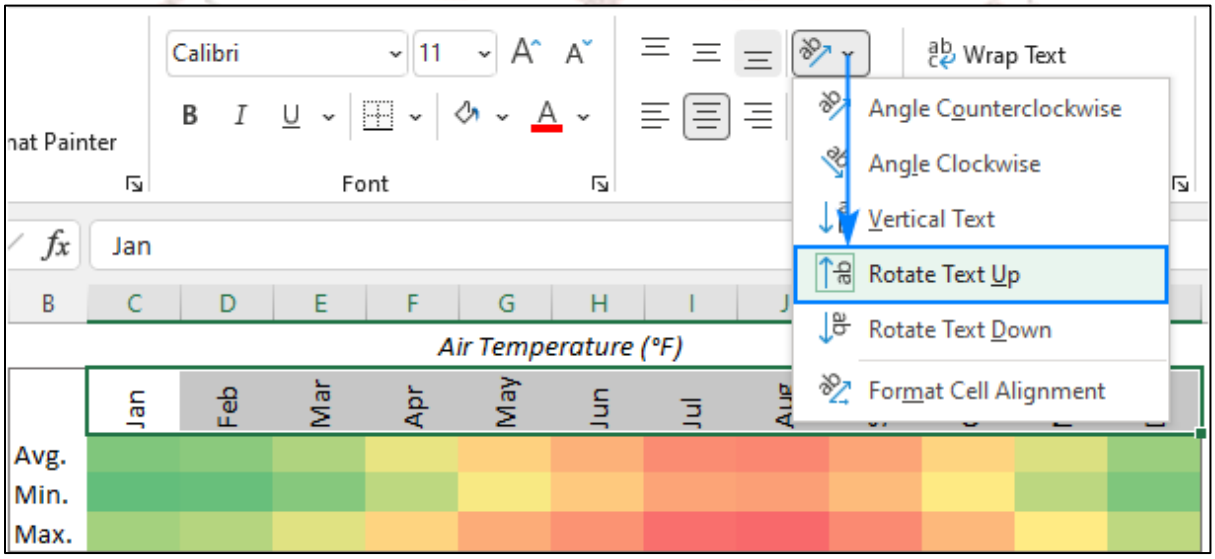


Fig 70: Updating Text Rotation

2. **Set column width.** Select all the columns and drag any column header's edge to make it wider or narrower. As you do this, a tooltip will appear showing an exact **pixel count** - remember this number.

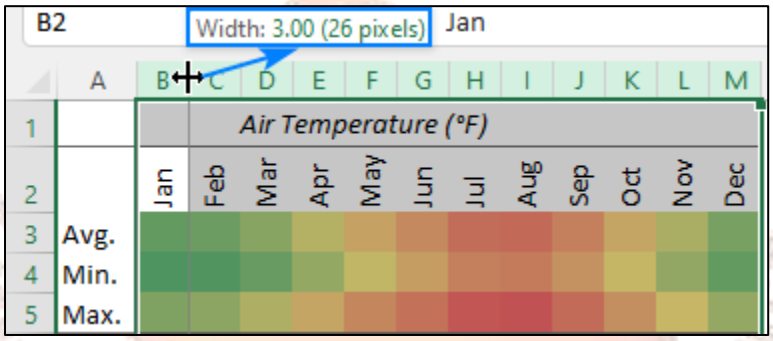


Fig 71: Adjusting Width of Columns

3. **Set row height.** Select all the rows and drag any row header's edge to the same pixel value as columns (26 pixels in our case).

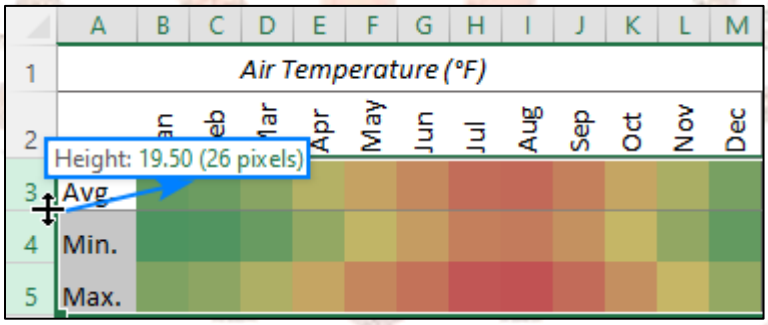


Fig 72: Adjusting Row Heights

Done! All the cells of your hat map are now square shaped:

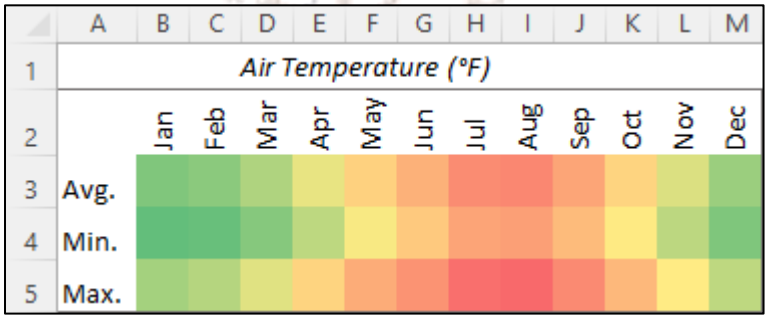


Fig 73: Final Output

How to make a heat map in Excel PivotTable

In essence, the process of creating a heatmap in a pivot table is similar to doing so in a regular data range, which involves utilizing conditional formatting with a color scale. However, there's an important consideration: when new data gets added to the source table, the conditional formatting won't be automatically applied to that new data.

For example, we've added Lui's sales to the source table, refreshed the PivotTable, and see that Lui's numbers are still outside the heat map:

	A	B	C	D	E
1	Sum of Sales	Product			
2	Reseller	Apples	Bananas	Cherries	Oranges
3	John	$180	$400	$250	$120
4	Mike	$120	$200	$580	$400
5	Pete	$290	$180	$330	$120
6	Sally	$250	$250	$380	$200
7	Lui			$240	$120
8	Grand Total	$840	$1,030	$1,780	$960

Heat map is not extended to new data

Fig 74: Heat map on a pivot table

4.4 How to Make Pivottable Heat Map Dynamic

To ensure that an Excel pivot table heatmap automatically includes new entries, follow these steps:

1. Select any cell within your existing heatmap.
2. Go to the Home tab, then navigate to the Styles group, and click on Conditional Formatting > Manage Rules...
3. In the Conditional Formatting Rules Manager, choose the rule you want to modify and click the Edit Rule button.
4. In the Edit Formatting Rule dialog box, under the "Apply Rule To" section, select the third option, which, for instance, might say: "All cells showing 'Sum of Sales' values for 'Reseller' and 'Product'."
5. Click OK in both dialog boxes to save your changes.

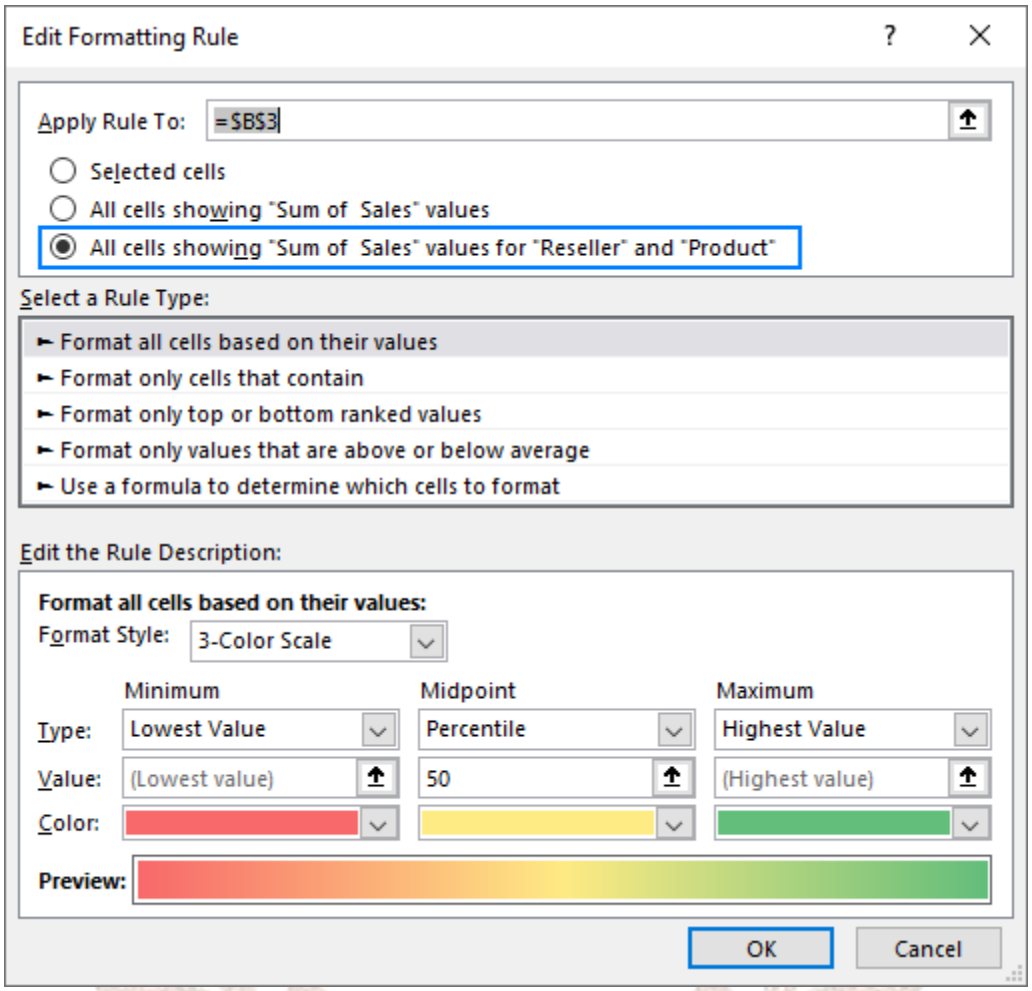


Fig 75: Updating Colour Scales

Now, your heat map is dynamic and will update automatically as you add new information in the back end. Just remember to refresh your PivotTable :)

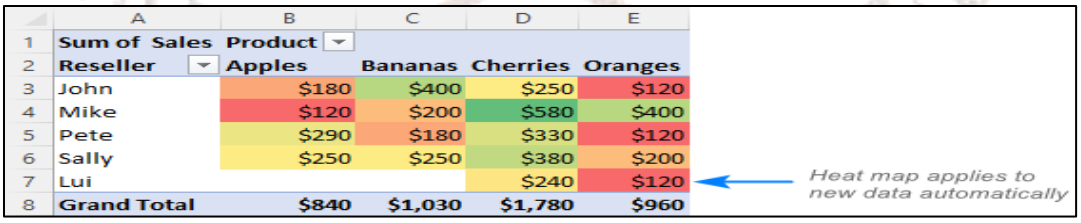


Fig 76: Updated Rule applies to new cells added

4.5 Heat Maps Offer Several Advantages for Data Visualization and Analysis

1. Simplicity: Heat maps simplify complex data by representing it visually. They condense large datasets into an easily understandable format.
2. Pattern Recognition: They make it easy to spot patterns, trends, and outliers in data, even for non-technical users.
3. Efficiency: Heat maps allow for quick insights into data without the need for extensive numerical analysis. This efficiency is especially valuable when dealing with big data.
4. User-Friendly: They are intuitive and user-friendly, making them accessible to a wide range of audiences.
5. Comparison: Heat maps enable easy comparison between different data points or categories.
6. Focus on Key Information: They can highlight specific areas of interest within the data, drawing attention to critical information.
7. Spatial Data Representation: Geographical heat maps are particularly useful for representing data on maps, such as regional sales data or population distributions.
8. Customization: Heat maps can be customized with different color scales and legends to suit specific needs.
9. Interactivity: In digital formats, heat maps can be interactive, allowing users to zoom in, filter, or explore data in more detail.
10. Decision-Making: They aid in decision-making processes by presenting data in a visually compelling way, helping stakeholders make informed choices.
11. Communication: Heat maps are excellent tools for conveying information to a broad audience, making complex data more accessible in presentations and reports.
12. Real-Time Data: They can be updated in real-time, allowing for dynamic monitoring of changing data.

Overall, heat maps are powerful tools for data exploration, analysis, and communication, making them valuable in various fields, including business, science, and research.

5. SUMMARY

The chapter delves into three crucial data visualization tools: Gantt charts, histograms, and heat maps. Gantt charts, originally devised by Henry Gantt, are essential in project management, aiding in task scheduling and monitoring project progress. Histograms are presented as graphical representations of data distribution, with various types discussed, from symmetric to probability histograms. Heat maps, on the other hand, are introduced as powerful tools for visualizing numeric data, employing color gradients to depict values. The chapter also provides practical guidance on creating these visualizations, underlining their significance in diverse fields, from project management to data analysis, and their role in simplifying complex data patterns.

6. QUESTIONS

Self-Assessment Questions:

SELF-ASSESSMENT QUESTIONS – 1

1. What is the primary purpose of a histogram?
2. How do you determine the number of bins in a histogram?
3. Explain the concept of frequency density in a histogram.
4. Who is credited with inventing the Gantt chart?
5. What is the key information conveyed by a Gantt chart?
6. How are dependencies between tasks typically represented in a Gantt chart?
7. How does conditional formatting in Excel facilitate the creation of heat maps?
8. What type of data is typically visualized using a heat map?
9. Explain the significance of color gradients in a heat map.
10. What advantage do heat maps offer when compared to traditional numeric tables?

Terminal Questions:

1. Explain the fundamental purpose of a histogram in statistics. Provide a step-by-step process for creating a histogram, including how to determine the number of bins and their boundaries. Illustrate your explanation with an example.
2. Discuss the significance of Gantt charts in project management. Describe the key components of a Gantt chart and explain how they help project managers plan and monitor projects effectively. Provide an example of a project and create a simplified Gantt chart to represent its timeline.
3. Define what a heat map is and describe its primary applications in data visualization. Explain the concept of conditional formatting in creating heat maps using tools like Microsoft Excel. Provide a real-world scenario and demonstrate how to create a heat map to visualize the data effectively.
4. Compare and contrast histograms with other types of data visualization, such as line charts and bar charts. Highlight the situations in which histograms are more suitable for representing data distribution and patterns. Provide examples to support your comparisons.
5. Describe the challenges and limitations associated with using Gantt charts in project management. Discuss how Gantt charts handle complex projects with multiple dependencies and constraints. Offer strategies for overcoming these challenges and maximizing the benefits of Gantt charts.
6. Explain the role of color selection and gradient scales in creating informative and visually effective heat maps. Discuss the psychological aspects of color perception and how they impact data interpretation. Provide guidelines for choosing appropriate color schemes for different types of data and objectives.

7. ANSWERS

Self-Assessment Question

1. The primary purpose of a histogram is to represent the distribution of data by using bars or bins to display the frequency of values within specified intervals or ranges.
2. The number of bins in a histogram can be determined using various methods, including the square root rule, Sturges' formula, or Freedman-Diaconis' rule.
3. Frequency density in a histogram refers to the ratio of the frequency of data points in a bin to the width of that bin.
4. Henry Gantt is credited with inventing the Gantt chart.
5. A Gantt chart conveys information about tasks or activities in a project, including their start and end dates, duration, dependencies, and progress.
6. Dependencies between tasks in a Gantt chart are typically represented by linking arrows or lines connecting dependent tasks.
7. Conditional formatting in Excel facilitates the creation of heat maps by allowing users to apply color scales based on cell values.
8. Heat maps are typically used to visualize data that has both row and column dimensions, such as data matrices, to identify patterns and variations.
9. Color gradients in a heat map signify the magnitude or intensity of values, with colors transitioning from cool (e.g., blue) for lower values to warm (e.g., red) for higher values.
10. Heat maps offer an advantage over traditional numeric tables because they make it easier to identify trends, outliers, and relationships in complex data sets through visual patterns and color differentiation.

Terminal question

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