VISUALIZING DATA IN EXCEL

1 Pie, Column, and Histogram Charts

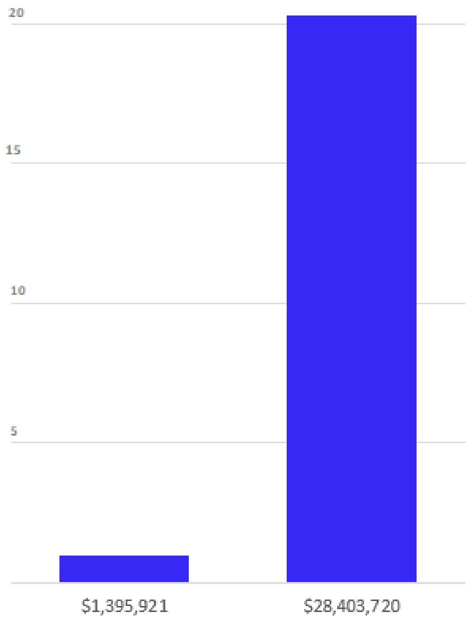
1.1 Introduction

In the previous course, we learned how to clean a dataset. But how do we explore the relationships within our freshly cleaned dataset? The answer is **data visualizations**. If it's true that a picture is worth a thousand words, then surely a data visualization is worth at least a million numbers!

We previously learned why it's important to clean our data before we analyze it. As we mentioned, the old saying "Garbage in, garbage out" applies to datasets and analysis. The same can also be said of data visualizations: messy data will only lead to messy data visualizations. Therefore, we can also use data visualizations as diagnostic tools to identify messy datasets. If the visualization doesn't look right, it could be a sign that the dataset could use some more cleaning.

Unfortunately, humans aren't very good at extracting meaningful insights by simply looking at a table of numbers. We are much better at finding patterns, recognizing relationships, and gaining insights when we see numbers represented visually.

As a simple example, consider how \$1395921 compares to \$2840372099. Is it obvious, or do you find yourself struggling a little to compare them? When most people see numbers like this, their brain simply says "Wow, those are big numbers," and that's the end of the conversation. It takes extra time and effort to appreciate the difference in their magnitude by counting how many digits each has and then comparing them.



However, even a simple data visualization like this one makes the comparison of these two values almost instantaneous. Looking at the visualization, it's much easier to see that the second value is approximately 20 times larger than the first.

Regardless of your business or career, data visualization skills are essential for communicating information in a way that is both effecient and universal. In this first lesson, we'll explore some basic chart types, and we'll learn how to create a frequency table using a built-in Excel function.

1.2 Bike-Sharing Data

In this lesson, we'll be using data about bike-sharing. Bike-sharing companies allow people to rent bikes for short trips. Someone can rent a bike from a station near their location, take a ride, and then return the bike to any other station.

The dataset we'll be using describes the daily activity of Capital Bikeshare (a bike-sharing company). Let's look at the first five rows of this cleaned dataset: **data.csv**

Each row describes the number of bike rentals for a single day — and the corresponding weather. The data spans a period of two years (January 1, 2011–December 31, 2012). Here are some of the columns we'll be focusing on:

- `dteday`: date in month-day-year format
- `weathersit`: `1` = Clear, `2` = Mist, `3` = Rain/Snow
- `temp`: normalized temperature in Celsius (1 is the maximum value: it's equivalent to 39°C; 0 is the minimum value: it's equivalent to -8°C)
- `atemp`: normalized feeling temperature in Celsius (1 is the maximum value: it's equivalent to 50°C; 0 is the minimum value: it's equivalent to -16°C)
- `casual`: the number of casual (non-registered) people who rented a bike that day
- `registered`: the number of registered people who rented a bike that day (registered people bought a membership)
- `cnt`: the total number of people who rented a bike that day (casual plus registered)

Researcher [Hadi Fanaee-T] collected the dataset, and you can download it from the [UCI Machine Learning Repository].

1.2.1 Instructions

We've provided a solutions file (`day_solution_file_1.xlsx`) in case you get stuck or would like to verify your answers while working on the exercises in this lesson.

1. Open the `day.csv` file in Excel.
    - You can find the file by selecting **File, Open, Browse** and navigating to the **Home Folder** under **This PC**.
    - Since this is a .CSV file, be sure you're viewing **All Files**; otherwise `day.csv` won't be displayed.
2. Save it as an Excel Workbook file: `day.xlsx`.
3. Explore the dataset to become familiar with its contents.

1.3 Frequency Tables

Our dataset describes bike-sharing activity for the Capital Bikeshare company in Washington, D.C. Let's assume the company has decided to expand to another American city. Weather patterns are going to play an important part in choosing the next location.

We're asked to create a data visualization that quickly summarizes the weather data about Washington D.C. and how often each weather type occurred over the two years of data — the decision-making team wants to use this as a benchmark for evaluating other cities.

A useful column in our dataset for this task is `weathersit`, which describes the daily weather type. This ordinal column contains three unique values: `1`, `2`, and `3`, which correspond to increasingly unpleasant biking weather. We want to know how many times each of these values appear in the column so we can create a visualization to quickly compare how often the different types of weather occur.

To do this, we will use the `COUNTIF` function in Excel. In general, we use `COUNTIF` like this:

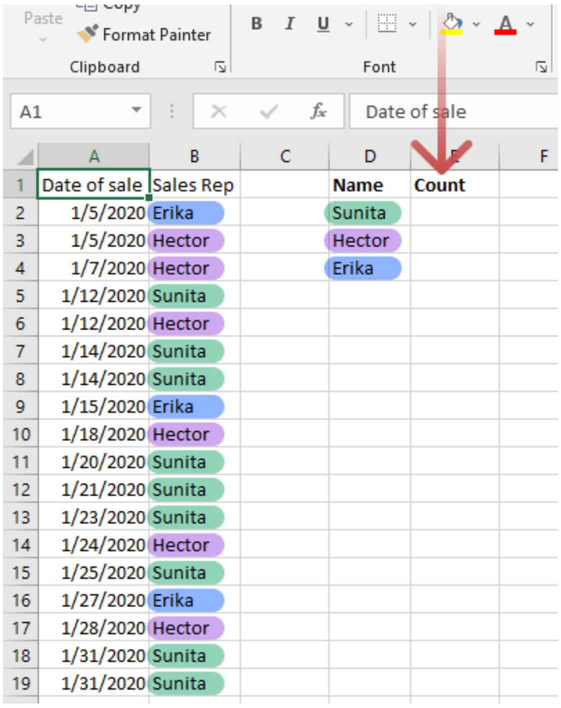
```
=COUNTIF(range, criteria)
```

ExplainCopy

We use it to count the number of cells in the `range` that meet a particular `criteria`. The criteria we can use to count the cells are quite flexible. We can use it to look for exact matches or partial matches using [wildcards], or we can even use comparison operators such as "<" or ">" on numerical data.

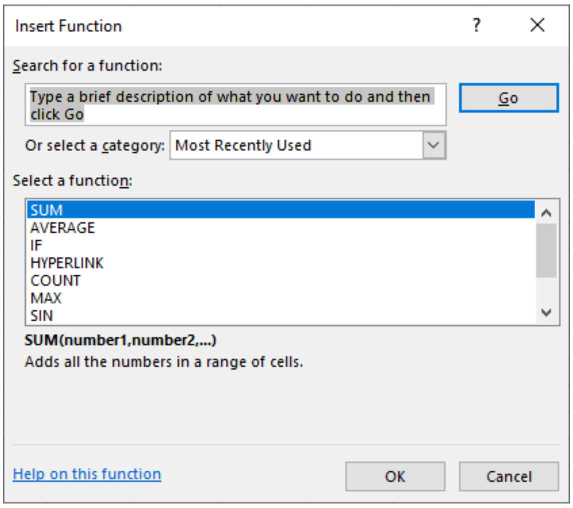
To look for an exact match, we can use a numerical value or a cell reference such as `10` or `C3`. For comparisions with numerical values, we can use something like: `">10"` or `"<=10"`. For comparisons with cell values, we need to use the `&` operator to [concatenate] (link together) the comparison operator and the cell reference, like so: `">" & C3` or `"<=" & C3`. This function is so flexible that we can use it on categorical or numerical data. Let's work on an example together before using it on our dataset.

Let's say we have a list of sales representatives and when they made their sales. We would like to know how many times each person made a sale. To keep our data organized, we create a table to the side of our main data to store our counts. This new table lists the unique names that appear in the list, and it has a column to store the counts for each, like this:



	A	B	C	D	E	F
1	Date of sale	Sales Rep		Name	Count	
2	1/5/2020	Erika		Sunita		
3	1/5/2020	Hector		Hector		
4	1/7/2020	Hector		Erika		
5	1/12/2020	Sunita				
6	1/12/2020	Hector				
7	1/14/2020	Sunita				
8	1/14/2020	Sunita				
9	1/15/2020	Erika				
10	1/18/2020	Hector				
11	1/20/2020	Sunita				
12	1/21/2020	Sunita				
13	1/23/2020	Sunita				
14	1/24/2020	Hector				
15	1/25/2020	Sunita				
16	1/27/2020	Erika				
17	1/28/2020	Hector				
18	1/31/2020	Sunita				
19	1/31/2020	Sunita				

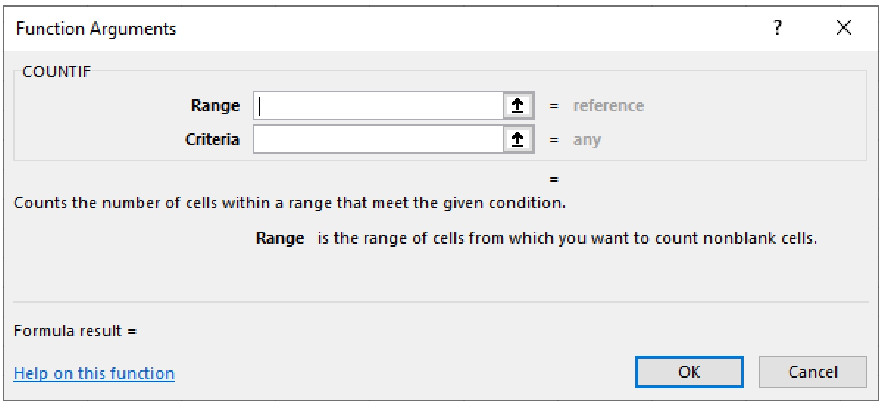
Next, we select cell E2. This cell will eventually display how many times *Sunita* appears in the list. Next, we click on fx (**Insert Function**). We can find this to the left of the **Formula Bar** above the spreadsheet. This will bring up the **Insert Function** dialog box, like this:



In the **Search for a function** section, type *countif*, and click **Go**. When the results are returned, select **COUNTIF** in the **Select a function** section, and click **OK**. While it's possible to use Excel functions directly in a cell by typing `=` followed by the name of the function, using f_x to insert a function offers many advantages:

- It inserts the `=` sign automatically for us; without it, our function would be treated as straight text.
- We avoid getting a `#NAME?` error by ensuring we use a recognized function (i.e., we avoid typos).
- We can find new functions we didn't know existed by searching by keyword.
- We can fill the arguments of the function interactively.
- We have easy access to the official Microsoft documentation for the function we select.

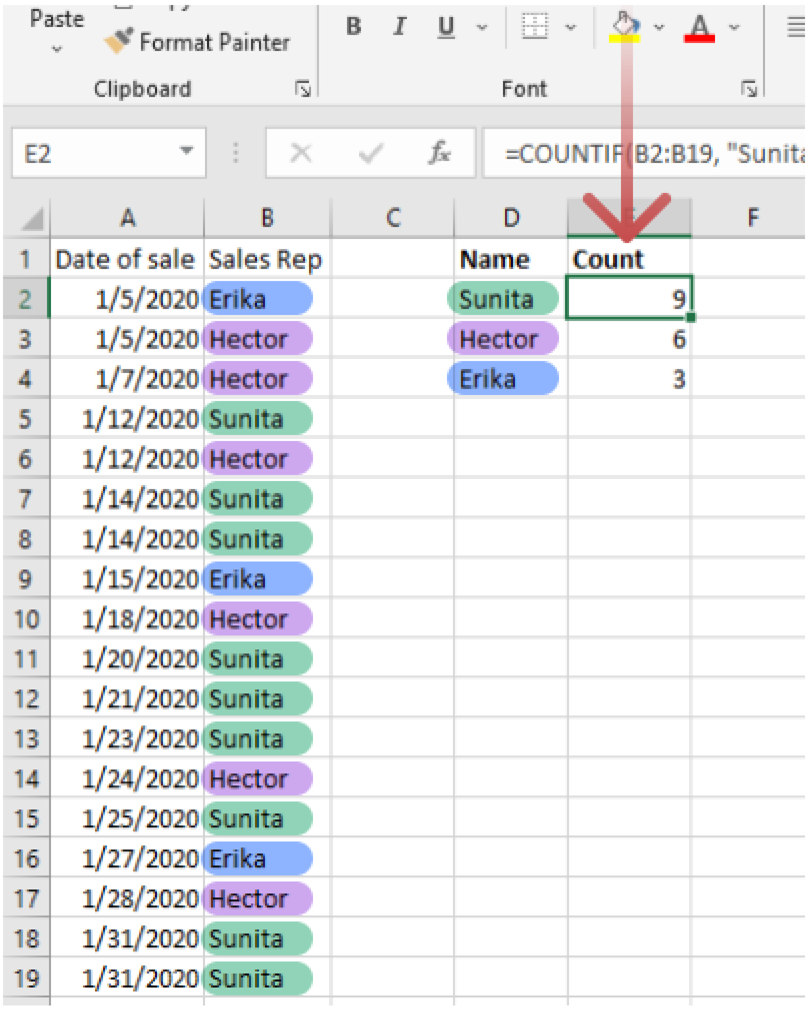
You should now see the **Function Arguments** dialog box for the `COUNTIF` function, like this:



The **Range** field specifies the range of cells to count, and the **Criteria** field specifies the criteria that control which cells to count. To fill in the **Range** field, we can manually type in the cell range or click the up arrow to select the range of cells using the mouse. For our example, we would select the cells B2 to B19 and press enter. Similarly, for the **Criteria** field, we can either type it in manually or use the up arrow to select a particular cell to define it. Let's type in the name *Sunita* and press **OK**.

The **Formula Bar** displays the complete function as `=COUNTIF(B2:B19,"Sunita")`, and the resulting value in cell E2 shows that this name appears nine times in cells B2:B19. Notice how the name *Sunita* is in double quotation marks. This is a requirement for **Criteria** to work correctly. Without the quotation marks, the `COUNTIF` function would return a misleading result of zero. This is yet another reason to use **Insert Function** rather than manually typing in a function and its arguments; it automatically formats the argument for us!

To get the counts for *Hector* and *Erika* in cells E3 and E4 respectively, we would use the same **Range** (B2:B19), but we would use *Hector* and *Erika* for the **Criteria** fields. The final results look like this:



This tells us that in cells B2:B19, the following are true:

- *Sunita* appears nine times

- *Hector* appears six times
- *Erika* appears three times

The table we just built is called a **frequency table** because it tells us how frequently each unique value appears in our data. Frequency tables are great at consolidating categorical data for quick data visualizations, which we will create on the next screen. First, let's create a frequency table for `weathersit`.

1.3.1 Instructions

1. In your `day.xlsx` workbook, create a frequency table for the `weathersit` column:

   - In cells R1 and S1, enter the headings `Weather` and `Count`, respectively.
   - Under `Weather`, enter the labels `Clear`, `Mist`, and `Rain/Snow` in your frequency table to represent the **Criteria** `1`, `2`, and `3`.
   - Populate the `Count` column using the `COUNTIF` function as demonstrated in the **Learn** section above.

1.4 Pie Charts

On the previous screen, we created a frequency table that looks something like this:

R	S
Weather	**Count**
Clear	463
Mist	247
Rain/Snow	21

Our next goal is to represent this data visually using a **pie chart**.

What is a pie chart?

A pie chart is a chart that looks like a pie! Picture an actual pie where each slice represents one component of the data, and all slices combined equal the entire data. We use it to demonstrate the relationship between the pieces and the whole pie.

Since we want to understand the part-to-whole relationship in our weather data, a pie chart is an excellent choice. That said, there is great debate in the community over the use of pie charts. Many argue that they should never be used, while others feel they are acceptable under the right conditions. We leave it for you to decide.

In any case, pie charts are one of the most widely used charts in business, and they won't be going away anytime soon. They frequently demonstrate things like the following:

- Percentage of customer types
- Revenue from different sources
- Market share
- Expenses by department

When to use a pie chart

General guidelines:

- We want to compare parts to the whole.
- We only have two or three pieces of categorical data to compare.
- We want to understand how categories compare to each other (relative comparison).
- We aren't comparing changes over time.
- The underlying data represents 100% of the situation we're reporting; we can't leave out any pieces of the pie!

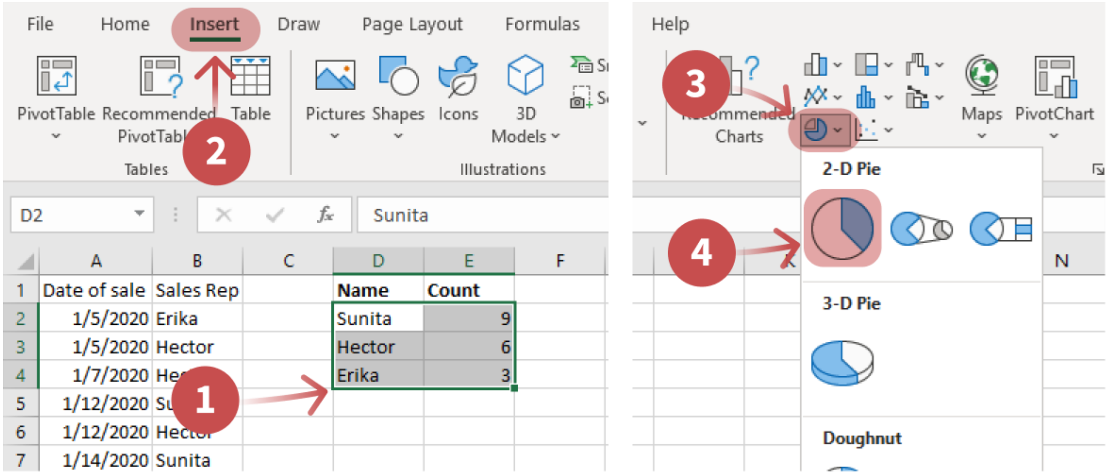
How to make a pie chart in Excel

As an example, we'll use the frequency table for the list of sales representatives on the previous screen to create a pie chart before creating one for our bike-sharing data.

Steps to create the pie chart:

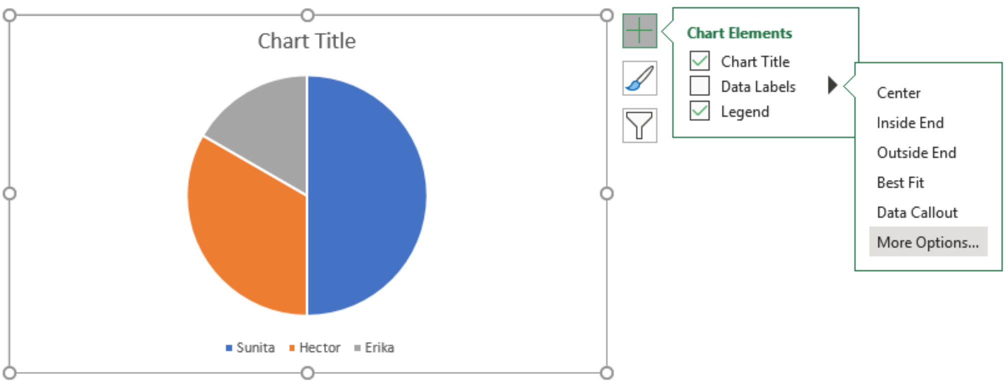
1. Select only the labels and their associated frequency counts from the frequency table.
2. Select the **Insert** tab at the top of the spreadsheet.
3. Select **Insert Pie or Doughnut Chart** from the **Charts** section.
4. Select **Pie** under **2-D Pie** (it's the first chart in the first row).

Here are the same steps visually:

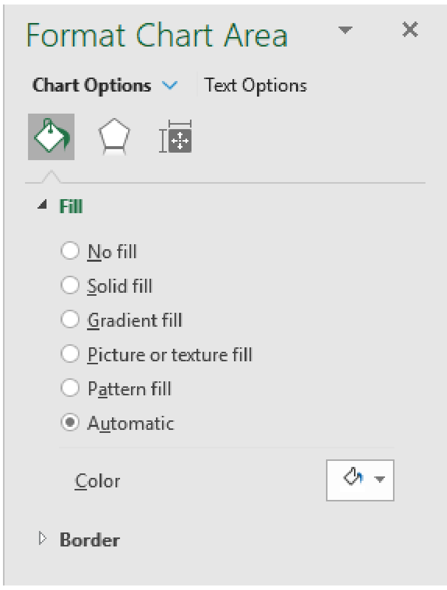


After selecting **Pie** under **2-D Pie**, the pie chart will appear. We can customize the chart by selecting it and clicking on the green plus sign in the top-right of the chart area. This gives us access to **Chart Elements**, where we can add or remove the elements that make up our particular chart type.

Different chart types will have different elements, but for a pie chart, **Chart Elements** gives us access to **Chart Title**, **Data Labels**, and **Legend**. Each of these also have **More Options**, which we access by clicking on the little black arrow to the right of the name for the element, like this:



We can also customize our chart by right-clicking on an element within the chart itself and selecting the option to format it from the list that appears (near the bottom). For example, we can format the **Chart Area** itself by right-clicking the chart and selecting **Format Chart Area**. This will display the options for **Format Chart Area** on the right side of the screen; it looks like this:



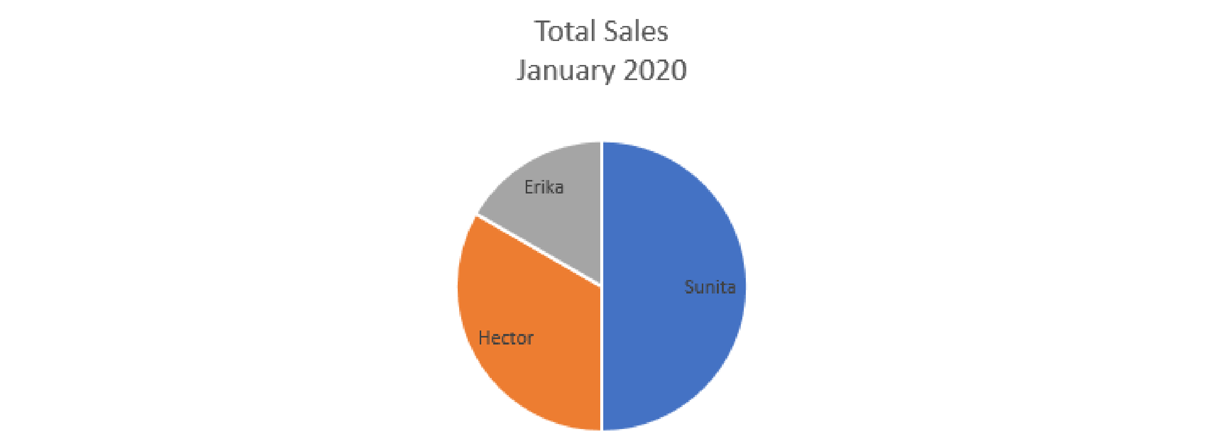
As you select different elements of the chart, the options in the formatting area will change to reflect the currently selected chart element.

We recommend experimenting with these options — there are more options than we can possibly cover here in an introductory course. Thankfully, many options are self-explanatory, and with practice, you'll quickly develop a sense of where to find the options you want.

Let's customize our pie chart by doing the following:

- Changing the chart title by selecting it and typing a new title
- Removing the chart border by right-clicking the chart and selecting **Format Chart Area**; on the **Fill & Line** tab, selecting **No Line** under **Border**
- Adding **Data Labels** using **Chart Elements**
- Displaying only the **Category Name** by formatting **Data Labels** and selecting **Category Name** and deselecting **Value**
- Removing the legend in **Chart Elements**

After making the above edits, here is the final result of the sales representatives pie chart:



1.4.1 Instructions

1. Create a pie chart for the frequency table we created on the previous screen for the `weathersit` column by following the example in the **Learn** section above.

2. Customize the chart to look similar to the example above:

    - Change the chart title to "Washington, D.C. Weather 2011–2012." You may want to split the title over two lines, depending on the size of your chart and how it looks.
    - Remove the chart **Border**.
    - Remove the **Legend**.
    - Add **Data Labels**, and format them so that we see only the **Category Name**.

1.5 Column Charts

While the pie chart we made for the decision-making team offered insight about the weather in Washington, D.C., it isn't very good for accurately comparing the differences between the categories. They have asked us to create another visualization that allows them to do that.

Another widely used chart in business, the **column chart**, would be an excellent choice here. These charts are useful when making comparisons between categories and showing the difference between them.

What is a column chart?

A column chart is a chart that represents categorical data using rectangular bars where their heights are proportional to the values that they represent. In other words, we encode the data by length in a column chart. Typically, the categories appear horizontally, while their associated values appear vertically.

Since our eyes are very good at comparing lengths when things are aligned, this chart is easy to interpret — one reason why column charts are so common in business. The opening screen of this lesson used a column chart to compare two very large numbers.

In Excel, when the bars are plotted vertically, we call it a **column chart**, and when they are plotted horizontally, we call it a **bar chart**. Aside from this obvious visual distinction, the charts are virtually the same. You may even see the terms used interchangeably. You may also see the terms "vertical bar chart" or "horizontal bar chart" used to describe them as well.

When to use a column chart

General guidelines:

- We want to compare different categories.
- We have 3 to 12 categories to compare.
- We want to show changes over time.
- We wish to find the largest, smallest, or most common value.
- The order of the categories isn't necessarily important.

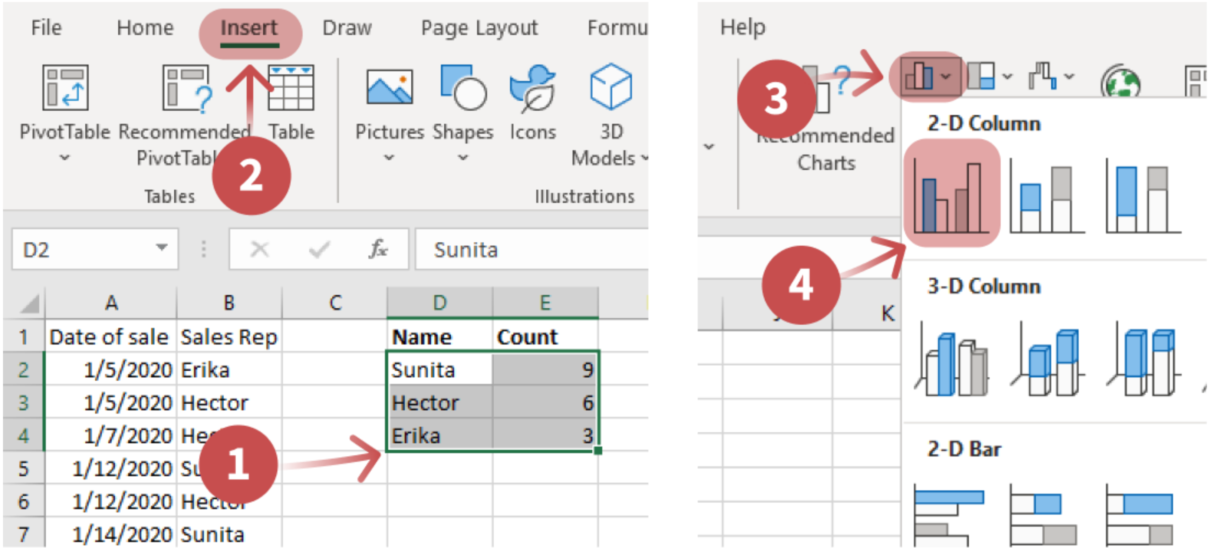
How to make a column chart in Excel

As an example, we will use the frequency table for the list of sales representatives on the previous screen to create a column chart before creating one for our bike-sharing data.

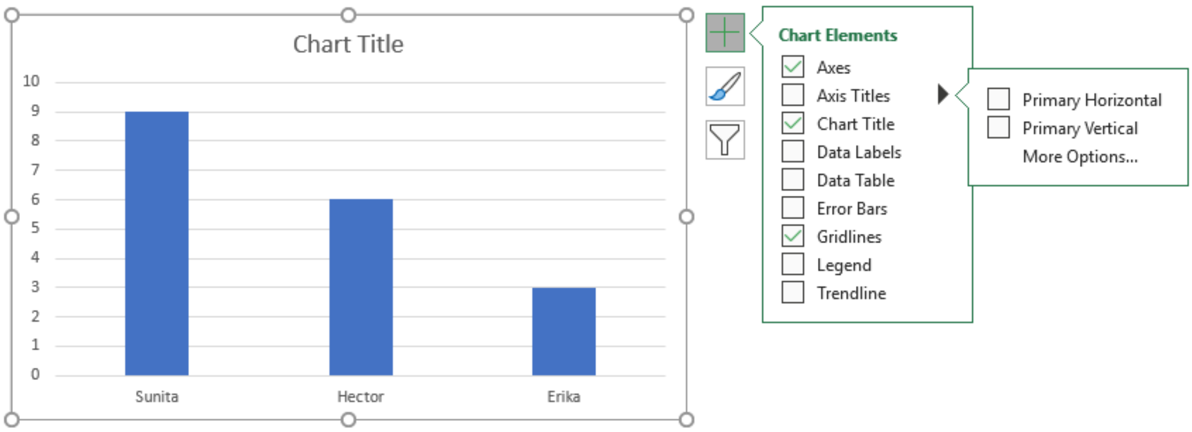
Steps to create the column chart:

1. Select only the labels and their associated frequency counts from the frequency table.
2. Select the **Insert** tab at the top of the spreadsheet.
3. Select **Insert Column or Bar Chart** from the **Charts** section (three multicolored bars).
4. Select **Clustered Column** under **2-D Column** (it's the first chart in the first row).

Here are the same steps visually:



After selecting **Clustered Column** under **2-D Column**, the column chart will appear. As we did with pie charts, we can customize the chart by selecting it and clicking on the green plus sign in the top-right of the chart area to access **Chart Elements**. As you can see, we have many more chart elements that we can add to a column chart compared to a pie chart.

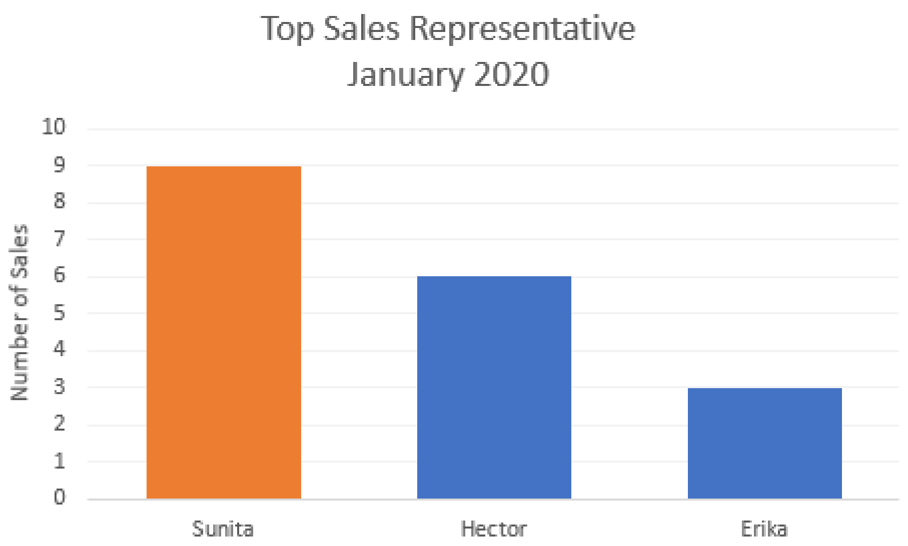


Let's customize our column chart by doing the following:

- Giving it a meaningful title: "Top Sales Representative January 2020," split across two lines
- Selecting **Primary Vertical** under **Axis Titles** in **Chart Elements** and changing the **Primary Vertical** title to "Number of Sales"
- Right-clicking a vertical bar in the chart and selecting **Format Data Series** to change the **Gap Width** (under **Series Options** tab) to 75% to reduce the white space between the bars
- Right-clicking on a horizontal gridline and selecting **Format Gridlines** to change the **Transparency** (under **Fill & Line** tab) to 50% so that the appearance of gridlines is softened, making sure that the **Color** is set to **Light Gray**

- Right-clicking the **Chart Area** and selecting **Format Chart Area** to change **Border** (under **Fill & Line** tab) to **None**
- Selecting only the bar for *Sunita*, right-clicking, and selecting **Format Data Point** to change the **Color** (under **Fill & Line** tab) to **Orange**

Here is the final result of the sales representatives column chart after making the above edits:



When creating a column chart, there should always be a gap between our data points because the data is categorical. Without the gap, it can make the data appear continuous, which is misleading.

1.5.1 Instructions

1. Create a column chart for the frequency table we created for the `weathersit` column by following the example in the **Learn** section above.

2. Customize the chart so it looks similar to the example above:

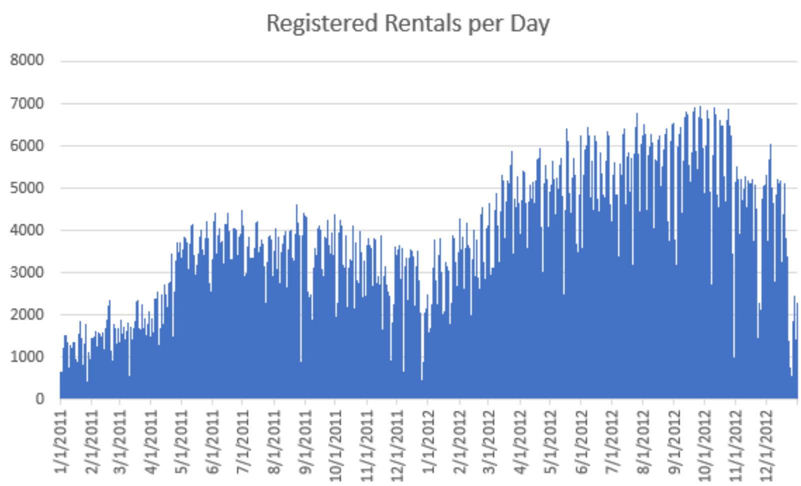
    - Change the chart title to "Washington, D.C. Weather 2011-2012". You may want to split the title over two lines depending on the size of your chart and how it looks.
    - Add **Axes Titles**:
        - vertical: "Number of Days"
        - remove the horizontal title
    - Change the **Gap Width** to 75%.
    - **Format Gridlines** to soften their appearance.
    - Remove the chart **Border**.

- Change the **Color** for *Clear* to orange.

1.6 Histogram Charts

The decision-making team is happy with our column chart on the frequency of weather types — they feel they are close to finding another city with similar weather! However, now they have shifted their focus to understanding how many bikes registered users rent in a day. Their goal is to make sure there are enough bikes available daily to these customers.

Our task is to create a visualization that shows the team how many bikes are rented daily by registered users. Recall that the `registered` column records the number of bike rentals registered users rent per day. If we were to create a column chart for the entire `registered` column, we would get something that looks like this:



That is chaotic and hard to interpret. Ideally, what we want is to create a frequency table for these values and plot them like we did on the previous screens. But imagine for a moment if we were to create a frequency table for this column, what would it look like? We would see that many values only occur once, and our resulting column chart would have many different categories along the horizontal axis with a height of one. That's not going to help the team!

What we really want to create is a **grouped frequency table**, where each group represents a range of continuous values rather than an individual categorical value. Luckily, this is exactly what a **histogram chart** does for us.

What is a histogram chart?

Although a histogram chart looks very similar to a column chart, there are some important differences. In a column chart, each vertical bar represents a category and its height is proportional to that category's value. In a histogram chart, we call each vertical bar a bin, and it represents a range of continuous values. The height of each bin is proportional to the frequency of the values that fall within the range of that bin. Lastly, column charts that display nominal data can support rearranging their bins without losing their meaning, whereas histogram charts lose their meaning if we rearrange their bins.

Column and histogram charts differ visually as well. A column chart has a visible gap between its bars to indicate that each category is separate from the next, whereas a histogram chart has no gaps between bins because its data is continuous and the end of one bin leads directly into the next.

When to use a histogram chart

General guidelines:

- We have continuous numerical data
- We want to know how frequently ranges of values occur in our data
- We have a large dataset we want to summarize quickly

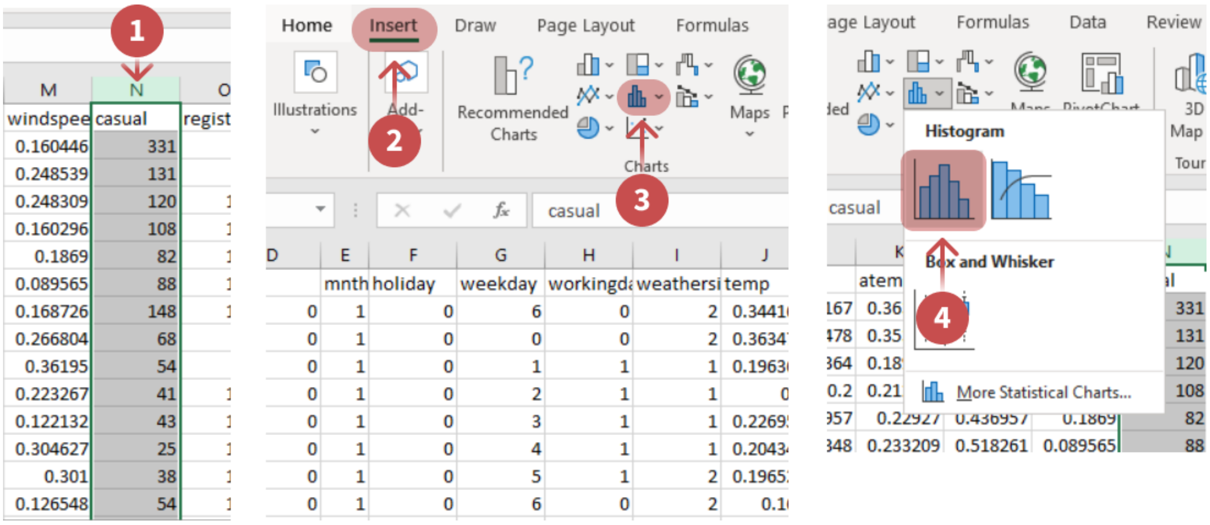
How to make a histogram chart in Excel

As an example, we'll use the `casual` column to create a histogram chart before creating the one for the `registered` column.

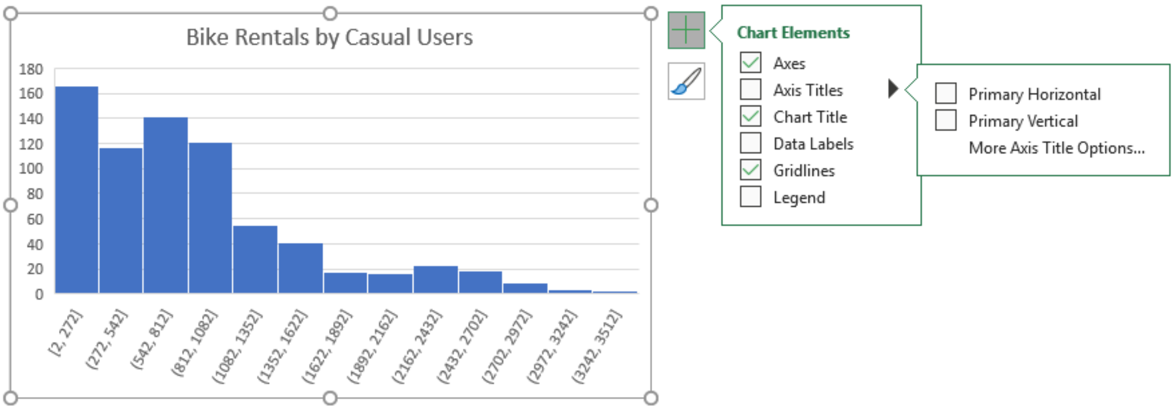
Steps to create the histogram chart:

1. Select the desired column.
2. Select the **Insert** tab at the top of the spreadsheet.
3. Select **Insert Statistic Chart** from the **Charts** section (four blue bars).
4. Select **Histogram** (it's the first chart in the first row).

Here are the same steps visually:



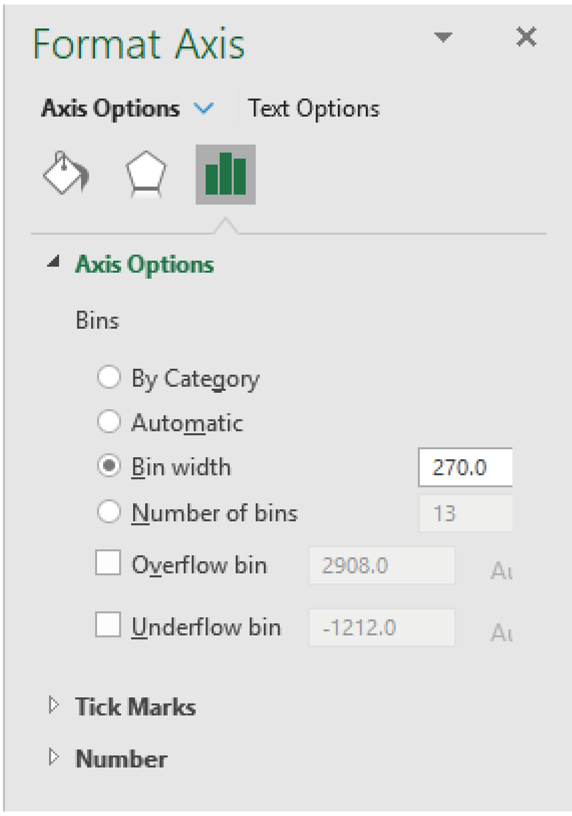
After selecting **Histogram**, the histogram chart will appear. As we did with previous charts, we can customize the chart by selecting it and clicking on the green plus sign in the top-right of the chart area to access **Chart Elements**.



This does look a lot like a column chart! But there are a few key differences. The horizontal axis is showing us that each bin represents a range. The first bin represents the range *[2, 272]*, while the second bin represents the range *(272, 542]*. Notice that these ranges use a mixture of square brackets *[ ]* and round parentheses *( )*. A square bracket indicates that the value is in the range while a round parenthesis indicates that all values greater than but not equal to the value are part of the range.

For example, the value *272* falls into the first bin because it has a square bracket next to *272* to indicate it is part of the first bin. *272* doesn't fall into the second bin because the second bin has a round parenthesis next to *272*, indicating that all values greater than but not equal to *272* are part of the second bin.

If we right-click the bin ranges along the horizontal axis and select **Format Axis**, we get the following options:

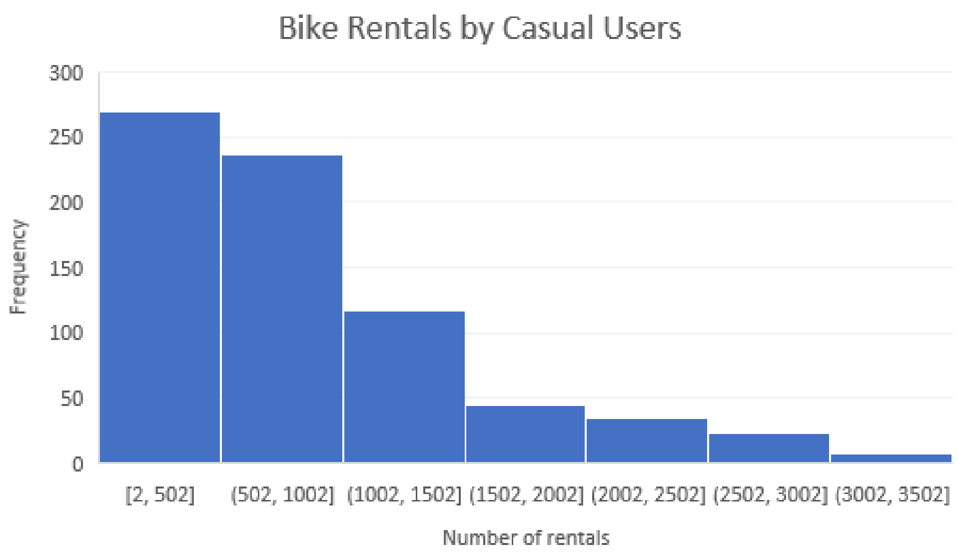


Since there is an inverse relationship between the width of each bin and the number of bins displayed, modifying either **Bin width** or **Number of bins** will affect the other. As the number of bins increases, the width of each bin decreases, and vice versa. Try experimenting with these options to see how it affects the chart.

Let's customize our histogram chart by doing the following:

- Changing the **Bin width** to 500 under **Axis Options**
- Adding **Axis Titles**:
    - horizontal: "Number of rentals"
    - vertical: "Frequency"
- Removing the chart border and softening the appearance of gridlines as we did on the previous screen.

Here is the final result of the casual users histogram chart after making the above edits:



If we hover the cursor over the first bin, we see that casual renters rented between 2 and 502 bikes 269 times. Hovering over the second bin, we see that they rented between 502 and 1002 bikes 236 times. By the time we get to the final bin, we see casual renters rented between 3002 and 3502 bikes only 7 times in two years. This tells us that on any given day, the majority of casual renters rent between 2 and 1002 bikes — and that they rarely rent more than 3002 bikes per day.

Let's use what we've just learned to create a histogram chart for `registered` users.

1.6.1 Instructions

1. Create a histogram chart for the `registered` users by following the example in the **Learn** section above.

2. Customize the chart to look similar to the example above:

    - Add a meaningful title to the chart.
    - Add appropriate axis titles.
    - Change **Bin width** to 1000.
    - Format the gridlines to soften their appearance.
    - Remove the chart border.

3. Add some text below the histogram explaining the pattern you see in how many bikes registered users rent per day.

2 Line, Scatter, and Combo Charts

2.1 Introduction

In the previous lesson, we learned how to create a few basic chart types and how to interpret them. Each of the chart types we learned to create used only one column of numerical data. In this lesson, we'll be creating chart types that use more than one.

This will allow us to discover relationships within our data and explain some of the patterns we see.

We'll continue to work with the dataset describing the daily activity of the Capital Bikeshare company based in Washington, D.C.

As a reminder, here are the first few rows of the dataset:

	instant	dteday	season	yr	mnth	holiday	weekday	workingday	weathersit	temp
	1	01-01-2011	1	0	1	0	6	0	2	0.344167

	instant	dteday	season	yr	mnth	holiday	weekday	workingday	weathersit	temp
	2	01-02-2011	1	0	1	0	0	0	2	0.363478
	3	01-03-2011	1	0	1	0	1	1	1	0.196364
	4	01-04-2011	1	0	1	0	2	1	1	0.200000
	5	01-05-2011	1	0	1	0	3	1	1	0.226957

Each row describes the number of bike rentals for a single day and the corresponding weather. The data spans a period of two years (January 1, 2011–December 31, 2012). Here are some of the columns we'll be focusing on:

- `dteday`: date in month-day-year format.
- `yr`: `0` = 2011, `1` = 2012
- `season`: `1` = Winter, `2` = Spring, `3` = Summer, `4` = Fall
- `weathersit`: `1` = Clear, `2` = Cloudy, `3` = Rain/Snow
- `temp`: normalized temperature in Celsius (1 is the maximum value: it's equivalent to 39°C; 0 is the minimum value: it's equivalent to -8°C).

- `atemp`: normalized feeling temperature in Celsius (1 is the maximum value: it's equivalent to 50°C; 0 is the minimum value: it's equivalent to -16°C).
- `casual`: the number of casual (non-registered) people who rented a bike that day.
- `registered`: the number of registered people who rented a bike that day (registered people bought a membership).
- `cnt`: the total number of people who rented a bike that day (casual plus registered).

Researcher [Hadi Fanaee-T] collected the dataset, and you can download it from the [UCI Machine Learning Repository].

Before we move on to our next chart type, let's go over some basic chart terminology on the next screen.

Next

2.2 Chart Terminology

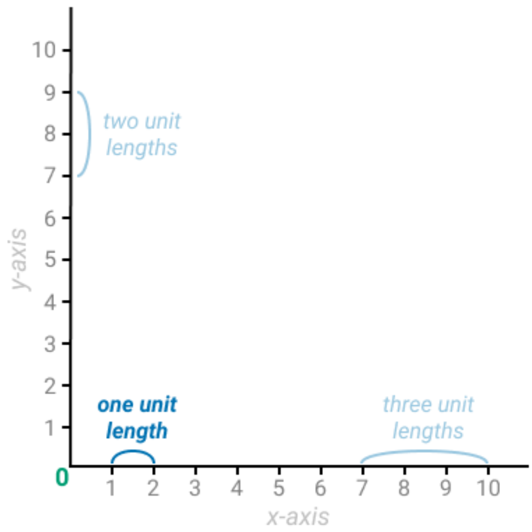
When reading about data visualizations, you may come across the terms **chart**, **graph**, or **plot**. While these all mean the same thing, Excel tends to use the term chart when referring to visualizations, so that's what we will use!

Many charts (but certainly not all) begin by drawing two lines at right angles to each other:

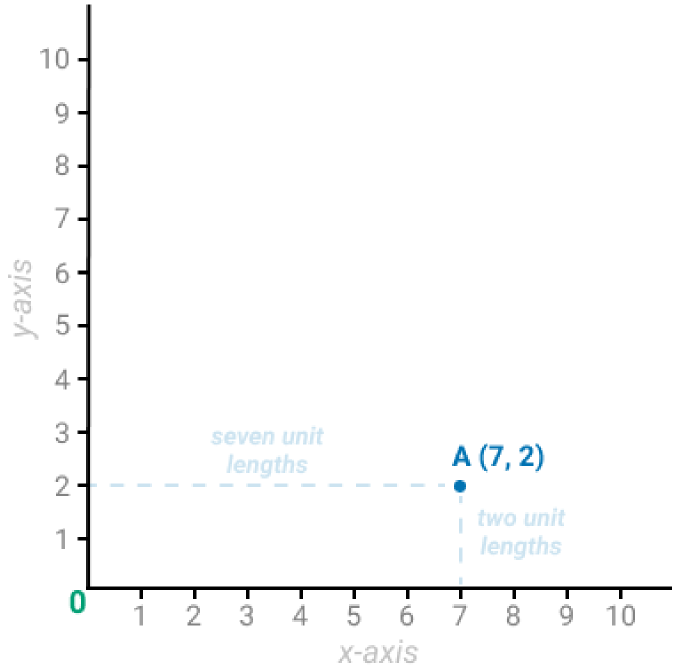
- Each line is called an **axis** — the horizontal line at the bottom is the **x-axis**, and the vertical line on the left is the **y-axis**.
- The plural form of axis is **axes**.
- The point where the two lines intersect is called the **origin**.
- The x-axis is often referred to as the **independent axis** while the y-axis is referred to as the **dependent axis**.

We can think of the data being represented along the x-axis as being the *cause* and the data along the y-axis as being the *effect*. We often have some control over the values along the independent axis, but we usually cannot control the values along the dependent axis. For example, if we wanted to create a chart showing the relationship between the `cost of an item` and the `number of items sold` at that price, we would make the `cost of an item` the independent values (x-axis) and the `number of items` `sold` the dependent values (y-axis). This is because the number of units that get sold *depends* on the cost of each unit. Also, we can control the price (independent), but we cannot control the number of units sold (dependent).

Each axis has length — below, we see both axes marked with numbers, which represent unit lengths.



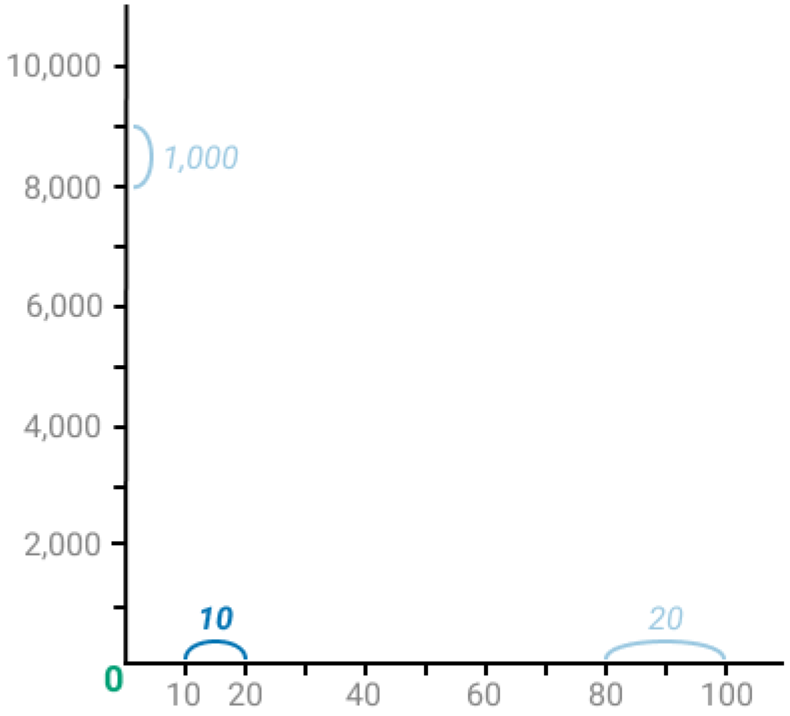
The length of the axes helps us locate any point on the chart. Point *A* on the chart below, for instance, is seven length units away from the y-axis and two units away from the x-axis.



The two numbers that represent the distances of a point from the x- and y-axis are called **coordinates**. Point A above has two coordinates: seven and two. Seven is the x-coordinate, and two is the y-coordinate.

The coordinates often appear in the form (x, y), with the x-coordinate first. So the coordinates of A are (7, 2). So, here's what we need to know about coordinates:

- The x-coordinate shows the distance in unit lengths relative to the y-axis.
- The y-coordinate shows the distance in unit lengths relative to the x-axis.
- The unit lengths of the x- and y-axes don't have to be the same. Below, we see the unit of length on the x-axis is 10, while on the y-axis it's 1,000 (note that we can also hide some of the numbers to make the chart look better).



Now, let's practice what we just learned.

2.3 Line Charts

The decision-making team at Capital Bikeshare has reached out to us again for some data visualization help. They would like to know how the total number of bike rentals changes over time. Ultimately, they want to see if there is a correlation between total rentals and the time of year. This information could help drive decisions like when to perform routine maintenance and when they should add more bikes and stations to meet higher demand.

None of the charts we have learned so far would be very good at displaying this relationship. We need a new type of chart that is good for showing changes over time using a large dataset. This is where a **line chart** truly shines!

What is a line chart?

A line chart also shows the relationship between independent and dependent values of continuous data. It displays information as a sequence of points called **markers** that are connected by straight line segments. The coordinates of each marker come from creating an (x, y) coordinate pair from the independent and dependent values of data. The line segments show how the values along the dependent axis change as we move across the independent axis from left to right while observing the slopes of the lines moving up or down. Since this chart typically highlights changes rather than the individual markers themselves, Excel doesn't show the markers by default.

A line chart is a popular choice when looking to find changes over time. In a line chart, time is always placed on the independent axis because, by its very nature, time doesn't depend on anything! Also, the independent axis should be what "causes" the results along the dependent axis, and time definitely causes change. For this reason, we often refer to a line chart as a **time series** when time is involved.

When to use a line chart

General guidelines:

- We have a large dataset with continuous data we want to compare.
- We want to see changes over time.
- We want to find a trend or pattern in our data.

How to make a line chart in Excel

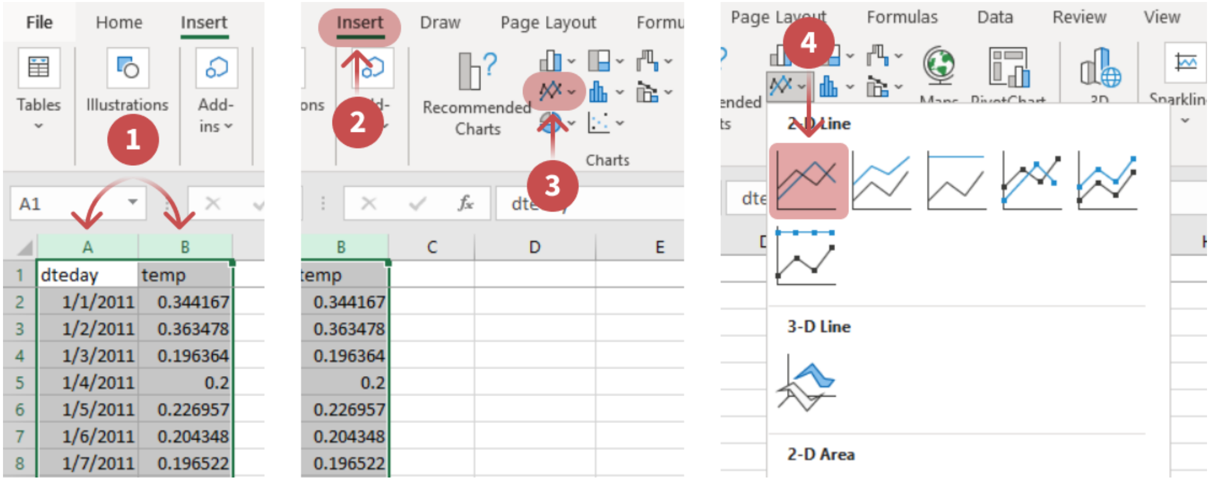
As an example, let's create a line chart for the `dteday` and `temp` columns before we move on to creating the visualization for the decision-making team.

First, let's copy the `dteday` and `temp` columns to a new worksheet so we're only working with the data we need to create our desired visualization. This will give our visualization more room on the worksheet. This also helps protect our original data from accidental changes. If we happen to make a mistake somewhere, we can always just copy the original data over again.

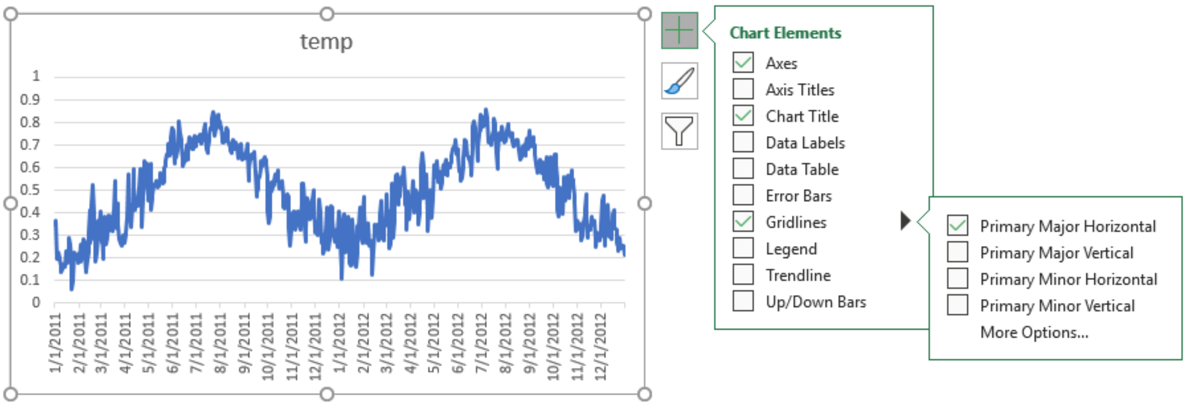
Steps to create the column chart:

1. Select the desired columns.
2. Select the **Insert** tab at the top of the spreadsheet.
3. Select **Insert Line or Area Chart** from the Charts section (two criss-crossing lines).
4. Select **Line** under **2-D Line** (it's the first chart in the first row).

Here are the same steps presented visually:



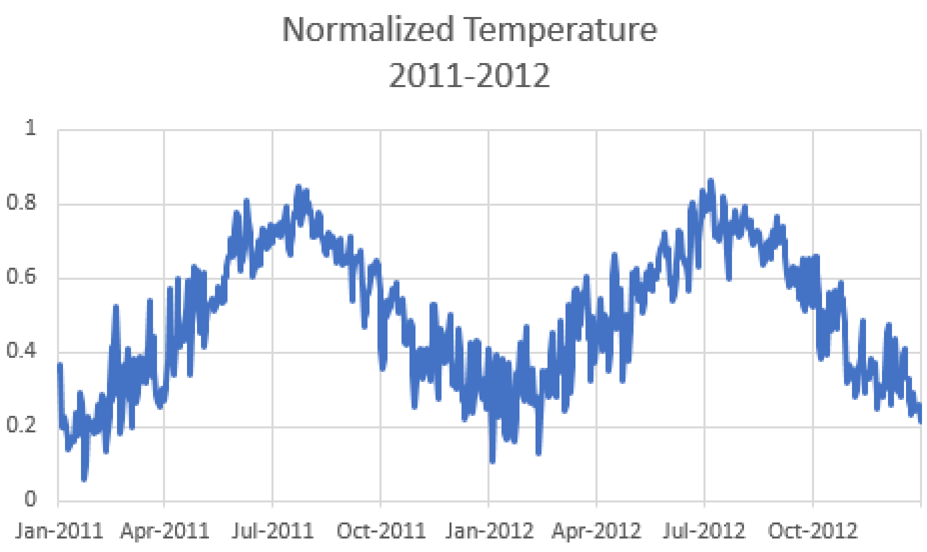
After selecting **Line**, the line chart will be created. As we did with previous charts, we can customize the chart by selecting it and clicking on the green plus sign in the top-right of the chart area to access **Chart Elements**.



Let's customize our line chart by doing the following:

- Giving it a meaningful title: "Normalized Temperature 2011-2012" split across two lines
- Right-clicking the horizontal axis and selecting **Format Axis** on the **Axis Options** tab; under **Number**, **Add** the **Format Code** `mmm-yyyy` so the dates appear in a three-letter month and four-digit year format. For more help on using **Format Code**, see the official Microsoft documentation here.
- Decreasing the number of dates shown along the horizontal axis by selecting **Axis Options** on the **Axis Options** tab, under **Units**, changing **Major** to `3 Months`
- Adding major vertical gridlines by opening **Chart Elements** and selecting **Primary Major Vertical** under **Gridlines**
- Decreasing the number of temperatures shown on the vertical axis by formatting the axis, selecting **Axis Options** on the **Axis Options** tab, under **Units**, changing **Major** to `0.2`
- Removing the chart border.

Here's the final result of the `dteday` and `temp` line chart after making the above edits:



We interpret this chart by reading it from left to right, and we watch to see if the line segments go up or down. As we can see by the jagged up-and-down line segments, there's a lot of variation in the temperature from day to day or week to week. However, looking at the overall trend in the chart, we can see there is a clear pattern that repeats each year. Here is the information about the daily temperature:

- It's lowest around January.
- It increases from January to about July.
- It's highest around July.
- It decreases from July to about January.

We call all of these similarities **seasonal trends**. In time series data, we sometimes see specific patterns occurring regularly at specific intervals of time — we call this **seasonality**.

Identifying seasonality can be useful for businesses:

- They can plan marketing campaigns at the right time.
- They don't need to panic needlessly when the sales are decreasing as a result of seasonality.
- They can hire extra employees right before the period of high activity begins.

2.3.1 Instructions

1. Open the file `day.xlsx` by selecting **File, Open, Browse** and navigating to the **Home Folder** under **This PC**.

2. Create a line chart for `dteday` and `cnt` by following the example in the **Learn** section above.

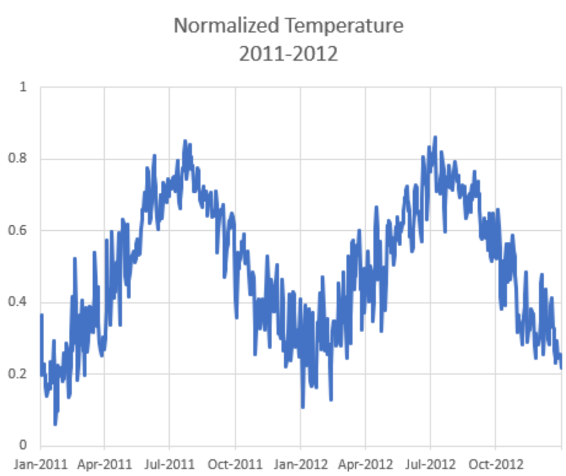
3. Customize the chart to look similar to the example above:

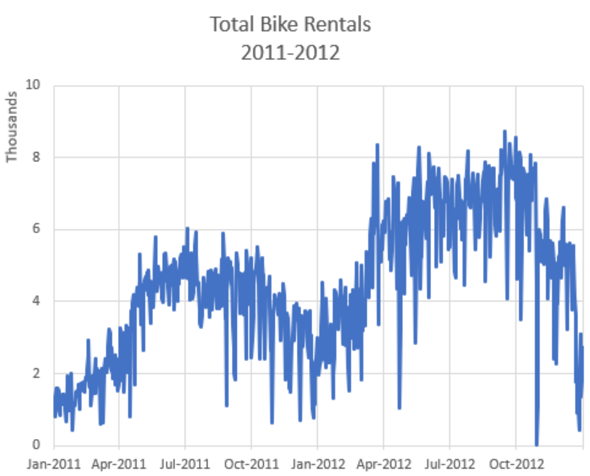
- Give the chart a meaningful title.
- Format the horizontal axis so that the dates appear in `mmm-yyyy` format.
- Format the horizontal axis to show dates every `3` months.
- Add **Primary Major Vertical** gridlines.
- Format the vertical axis so that **Display units** under **Axis Options** is set to `Thousands` and **Show display units label on chart** is enabled.
- Format the vertical axis so the number of bike rentals uses **Major** of `2000` under **Units** on the **Axis Options** tab.
- Remove the chart **Border**.

4. Below the chart, add a description explaining any pattern you observe and whether bike rentals show any seasonality.

2.4 Scatter charts

On the previous screen, we learned that both temperature and bike rentals show a similar seasonal pattern each year. This could indicate that the seasonality of bike rentals might correlate to temperature variations. The similarity in how the air temperature and the number of rented bikes changed over 2011 and 2012 supports this hypothesis.





We see that when the temperature increases, the number of bike rentals tends to increase too. When the temperature decreases, the number of rentals tends to decrease. Although the relationship isn't perfect, the values in the two charts follow similar patterns in how they change.

Above, we charted `cnt` and `temp` against `dteday`. To visualize the relationship between `cnt` and `temp`, it's better if we chart one against the other and leave out the `dteday` column. We can do this with a **scatter chart**.

What is a scatter chart

A scatter chart (also called a scatterplot) is similar to a line chart in that it shows the relationship between independent and dependent values of numerical data. However, unlike line charts where each x value only has one corresponding y value, scatter charts can have more than one y value for each x value.

Scatter charts focus on the markers (individual points on the chart) to display the relationship without connecting them with line segments like line charts do. The markers represent the individual values and, when charted, can potentially reveal a pattern. If the pattern forms a line, we say there is a **linear correlation** between our independent and dependent values. The closer the markers are to making a straight line, the stronger the correlation. The further away the markers are from forming a line, the weaker the correlation.

When to use a scatter chart

General guidelines:

- We want to see if there is a relationship between two columns of numerical data
- We have a lot of data to compare.
- We are looking for abnormal values that don't follow the overall pattern in our data.
- We want to explore data and find relationships

How to create a scatter chart in Excel

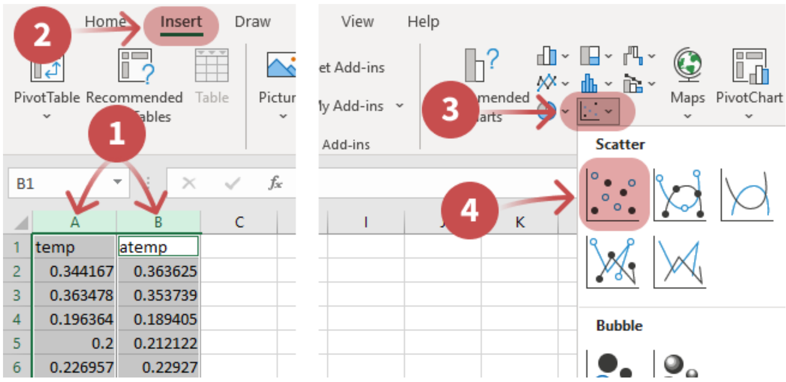
Let's create a scatter chart for two columns we believe will show a relationship: `temp` and `atemp`. Recall that the `temp` column records the actual air temperature, while the `atemp` column records the feeling air temperature (sometimes referred to as apparent temperature, real feel, or feels like).

Again, before we begin, let's copy the required columns to a new worksheet so that we're only working with the data we need to create our desired visualization.

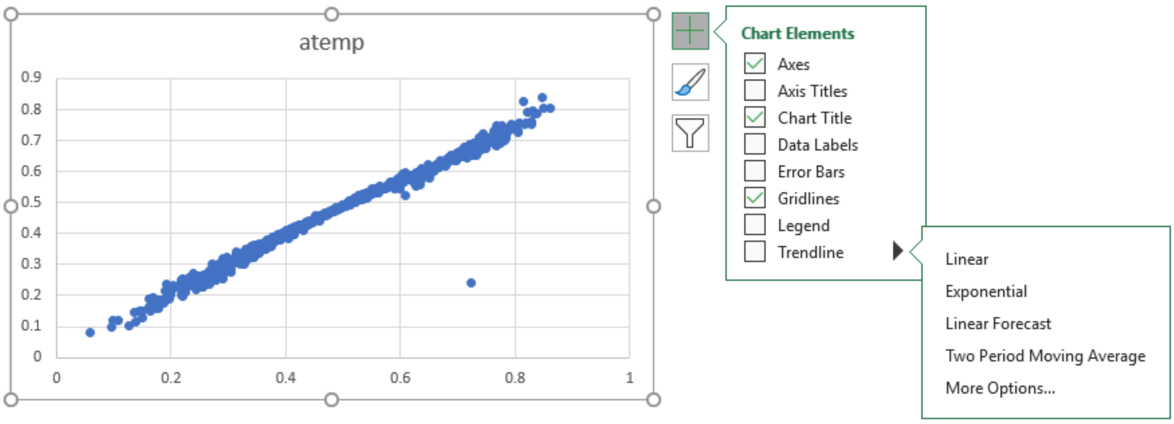
Steps to create the scatter chart:

1. Select the desired columns, making sure the independent column is to the left of the dependent column.
2. Select the **Insert** tab at the top of the spreadsheet.
3. Select **Insert Scatter (X,Y) or Bubble Chart** from the **Charts** section (x/y-axis with black and blue markers).
4. Select **Scatter** under **Scatter** (it's the first chart in the first row).

Here are the same steps represented visually:



After selecting **Scatter**, the scatter chart will be created. As we did with previous charts, we can customize the chart by selecting it and clicking on the green plus sign in the top-right of the chart area to access **Chart Elements**.

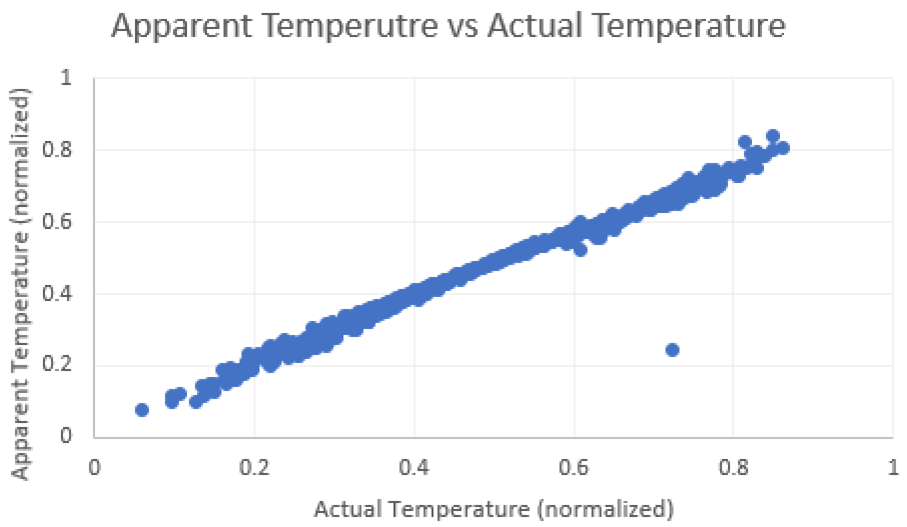


Let's customize our scatter chart by doing the following:

- Giving it a meaningful title: "Apparent Temperature vs Actual Temperature"
- Adding **Axis Titles**:
    - "Actual Temperature (normalized)" for the independent axis
    - "Apparent Temperature (normalized)" for the dependent axis
- Changing **Major** to `0.2` for the vertical axis

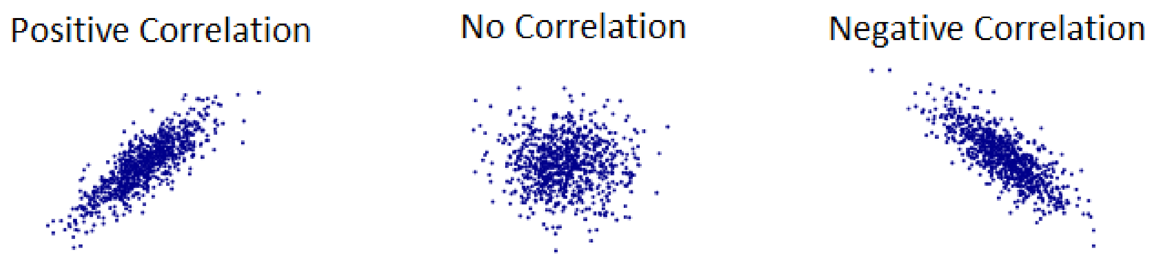
- Softening the appearance of **Gridlines**
- Removing the chart **Border**

Here is the final result of the `atemp` and `temp` scatter chart after making the above edits:



Looking at this chart, there is clearly a relationship between `atemp` and `temp`! Because the markers form an almost perfectly straight line, we say these two columns show a **linear correlation**. We can also say it is a **positive correlation** because as `temp` increases, `atemp` increases proportionally. This also means that as `temp` decreases, `atemp` decreases proportionally.

On the other hand, a **negative correlation** would show dependent values decreasing as independent values were increasing. Also, a negative correlation would show dependent values increasing as independent values were decreasing. Below are examples of positive correlation, no correlation, and negative correlation.



Returning to our chart above, you may be wondering about that marker in the lower right of the chart that doesn't seem to fit with the rest of our data. When we see a marker that is an abnormal distance away from the rest, we call it an **outlier**. While there are statistical methods for calculating outliers precisely, such methods are beyond the scope of this course and will have to wait for a dedicated course on statistics. That said, the general definition works well enough for our purposes here.

Outliers can be a sign that there's a problem with our data (like a typo), and we should investigate them before continuing with any kind of analysis or visualization. In this way, we can use scatter charts to identify data that needs cleaning. Recall from our previous course on data cleaning that we can correct outliers by replacing them with another value (like an average), deleting them from the dataset, or verifying the data with a reliable source.

2.4.1 Instructions

1. Create a scatter chart for `temp` (independent axis) and `cnt` (dependent axis) by following the example in the **Learn** section above.

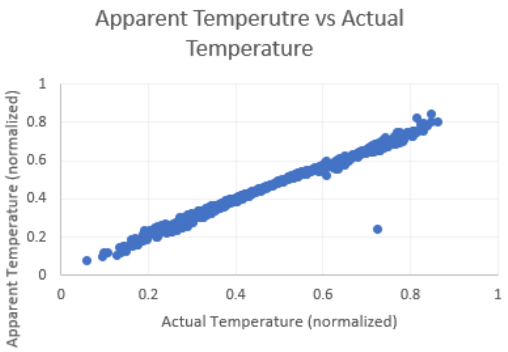
2. Customize the chart to look similar to the example above:

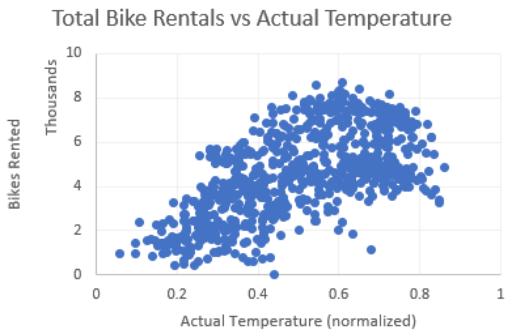
    - Give the chart a meaningful title.
    - Add **Axis Titles**.
    - Change **Major** to `0.2` for the horizontal axis.
    - Change **Major** to `2000` for the vertical axis.
    - Set **Display units** to `Thousands` and make sure **Show display units label on chart** is enabled for the vertical axis.
    - Soften the appearance of **Gridlines**.
    - Remove the chart **Border**.

3. Below the chart, add a description explaining the relationship between `temp` and `cnt`.

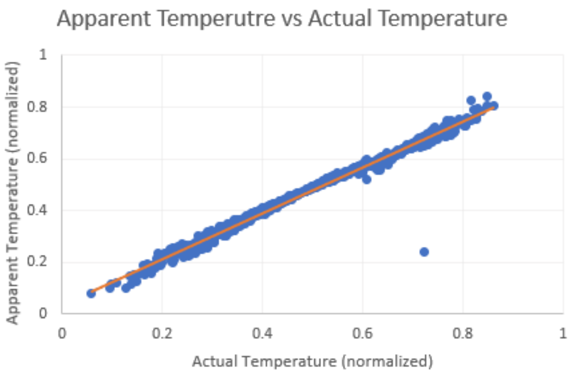
2.5 Measuring Correlation

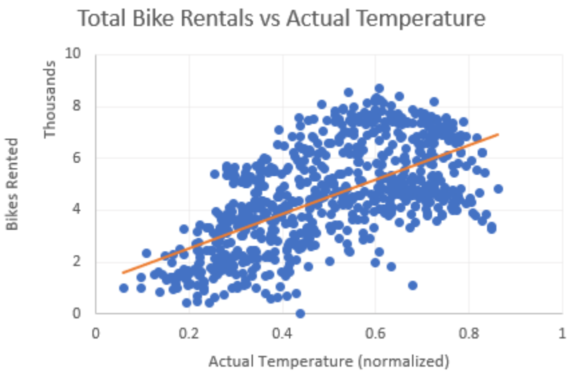
On the previous screen, we saw that a correlation exists between `temp` and `atemp` as well as between `temp` and `cnt`. However, their scatter charts didn't exactly look the same:





We can see that the first one has a much stronger correlation than the second, but it's not entirely clear by how much. How do we compare these two correlations? The first thing we can do is add a **Trendline** to our scatter charts using **Chart Elements**. Adding a trendline will automatically insert a line that best fits our data as if it had a perfectly linear relationship. This is why a trendline is sometimes called a **line of best fit**. The trendlines shown below have been formatted for easier viewing:



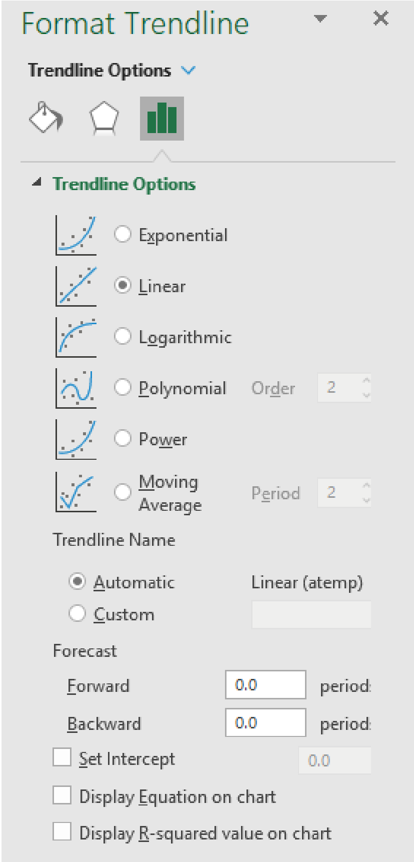


Looking at how far each marker is away from the trendline gives us an idea of how strong or weak the correlation is. The closer all the markers are to the trendline, the stronger the correlation. Conversely, the further away they are, the weaker the correlation.

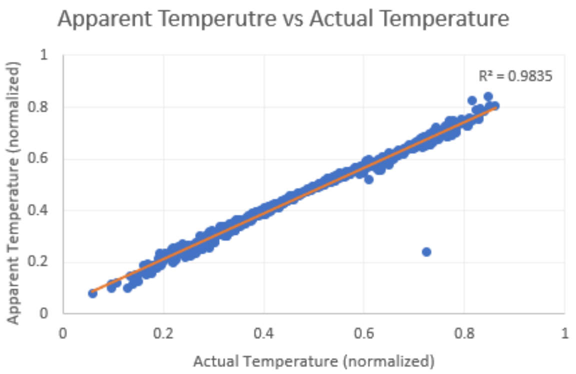
But how can we measure it? This is where a special statistical measurement called the coefficient of determination, denoted as **R-squared**, can help us. It's a measurement used to explain how much variability in the values of one column can be explained by its relationship to another related column.

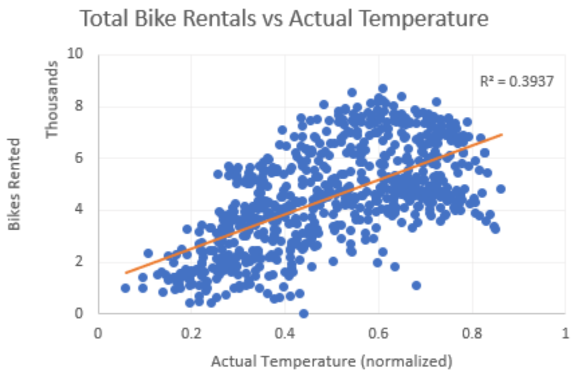
R-squared measures the strength of the relationship between the independent and the dependent columns on a convenient 0 to 1 scale, where 1 is a perfect fit. In fact, it gives us the percentage of variation in the dependent values that can be explained by the variation in the independent values. For example, an R-squared value of 1 means that 100% of the variation in the dependent values can be explained by the variation in the independent values.

While the calculation of R-squared is best left to a dedicated course on statistics, we can easily have Excel calculate it for us after adding a trendline. Right-clicking the trendline and selecting **Format Trendline** gives us the following options:



Ensuring **Linear** has been selected under **Trendline Options**, we enable **Display R-squared value on chart** at the bottom. Here are the results of enabling this option for the above two charts:

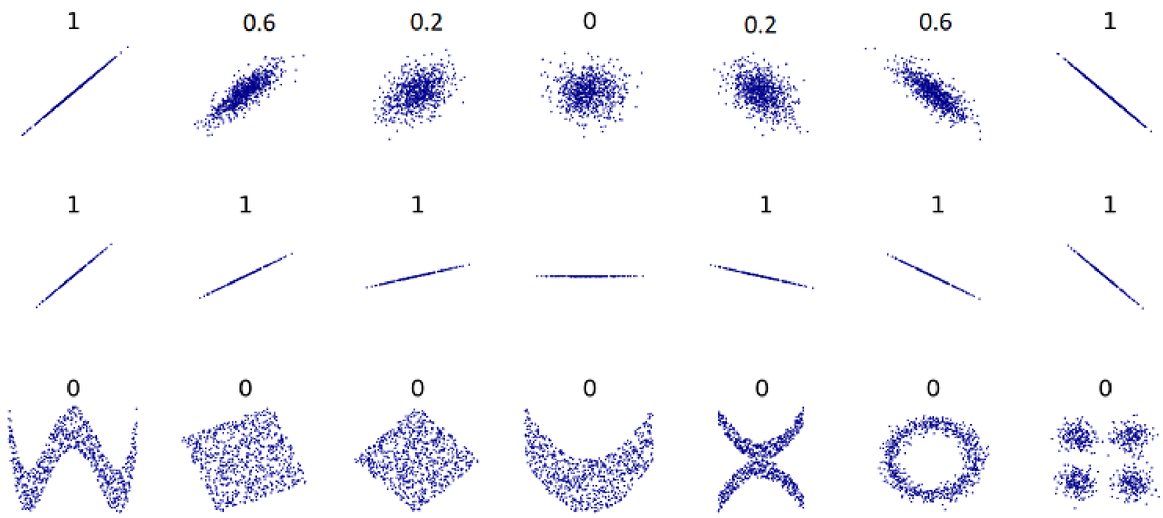




For the chart on the left, the R-squared value of `0.9835` tells us that a little over 98% of the changes in `atemp` can be explained by the changes in `temp`. Whereas for the chart on the right, the R-squared value of `0.3937` tells us that a little under 40% of the changes in `cnt` can be explained by the changes in `temp`.

What is a good R-squared value?

What qualifies as a good R-squared value will depend on the situation. In some fields, such as the social sciences, even a relatively low R-squared value like 0.5 could be considered relatively strong. In other fields, the standards can be much higher, such as 0.9 or greater. In finance for example, an R-squared value above 0.7 would generally be considered a high level of correlation, whereas a value below 0.4 would be considered a low correlation. These will all depend on context. That said, here are some examples of scatter charts and their associated R-squared values:



In the first two rows, the left side of the diagram shows positive correlations and the right side shows negative correlations. Notice how the same R-squared value appears on both sides. This is a reminder that R-squared is a measure of the correlation but not whether the correlation is positive or negative. The third row shows us that even though our data can produce interesting patterns on a scatter chart, we can still end up with an R-squared value of 0. In other words, visually interesting patterns don't always translate into something that's analytically useful.

Lastly, the middle chart in the middle row doesn't have an R-squared value despite the markers appearing to form a perfectly straight line. R-squared is **undefined** when the scatter chart produces a horizontal line. This is a direct result of how R-squared is calculated; in this case, we would end up trying to divide by zero, which is, by definition, undefined. This makes sense since R-squared measures how much variation in the dependent values is explained by the variation in the independent values. When we have a horizontal line, our dependent values don't vary; they are the same for all independent values. Therefore it doesn't make sense to have a defined R-squared value in these situations.

Correlation vs. Causation

When we find a strong correlation between two columns, one common mistake is to assume that the columns must also have a relationship of causality — one is the cause, and the other is the effect.

For instance, let's say we found a strong positive correlation between the number of bike rentals and ice cream sales. When the number of bike rentals goes up, the number of ice cream sales tends to go up as well. When bike rentals decrease, ice cream sales tend to decrease too.

The positive correlation doesn't imply that ice cream sales are causing bike rentals to increase or decrease (or vice versa). Most likely, both ice cream sales and bike rentals are related to temperature variations in a similar way, and this is why they change together similarly.

However, correlation can suggest causality. If two columns are correlated, it might be because of a cause-and-effect relationship. To establish a relationship of causality, we often need to perform a rigorous experiment.

The overarching point is that proving causality requires more than just correlation. We can't say that X is the cause of Y simply because columns X and Y are strongly correlated. In other words, correlation does not imply causation.

2.5.1 Instructions

1. Copy the `hum`, `windspeed`, and `cnt` columns to a new worksheet.
2. Create two separate scatter charts:
    - `hum` (independent values) and `cnt` (dependent values)
    - `windspeed` (independent values) and `cnt` (dependent values)
3. Give each chart a meaningful title.
4. For each chart, add a **Trendline** using **Chart Elements**.
5. Format the trendlines:
    - On the **Fill & Line** tab, change **Color** to a shade of orange.
    - On the **Fill & Line** tab, change **Dash Type** to use a **Solid** line (first option).
    - On the **Trendline Options** tab, enable **Display R-squared value on chart**. Using your mouse, select and move the R-squared value to the top-right of the chart area.

2.6 Summarizing Data

Over the previous screens, we discovered a correlation between `temp` and `cnt`. While both the line chart and the scatter chart showed this relationship, they each contain a lot of data that takes a little extra time to read and interpret. This is quite common for charts that we create while exploring our data for patterns and relationships.

But now that we have found this relationship, the decision-making team at Capital Bikeshare would like us to produce a chart they can show their investors that clearly conveys this relationship without presenting too much information. They are extremely busy, and they want to understand the data at a glance.

But before we can introduce our final chart type that will allow us to accomplish this, we need to organize our data into groups. Since we know the number of bike rentals is seasonal, we will first organize our data by season and year before putting it into a chart.

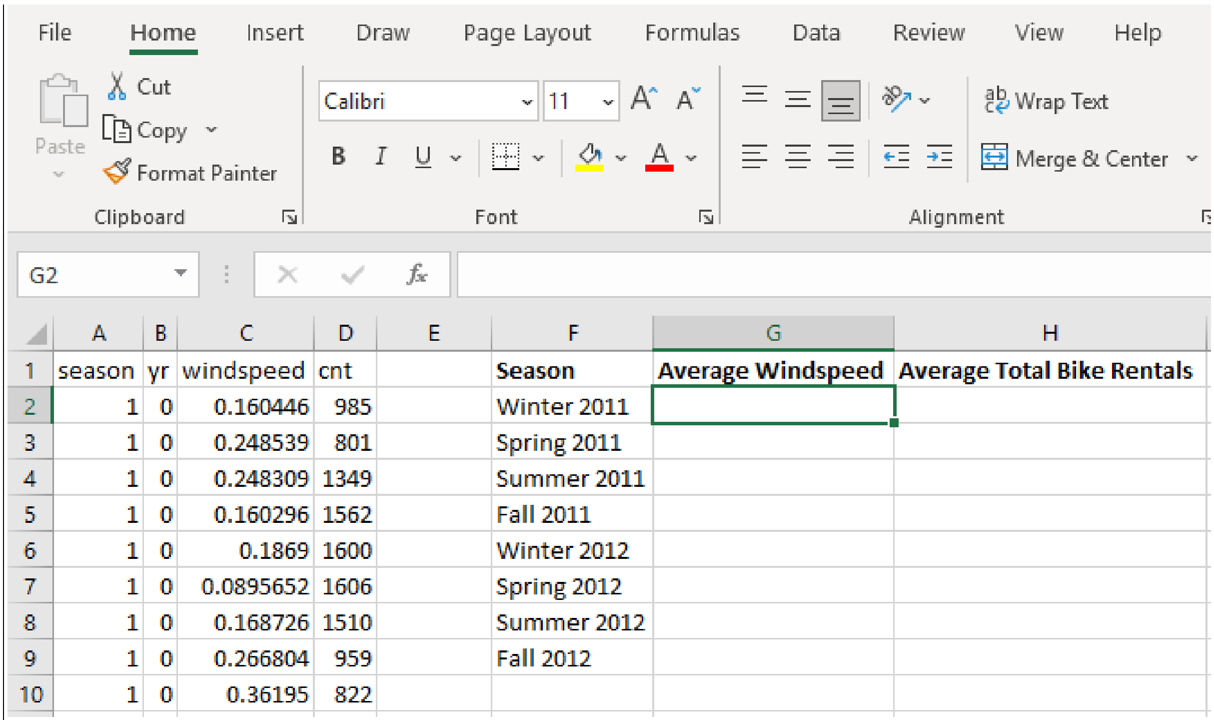
To do this, we will use the `AVERAGEIFS` function in Excel. We use `AVERAGEIFS` like this:

```
=AVERAGEIFS(average_range, criteria_range1, criteria1, [criteria_range2, criteria2], ...)
```

ExplainCopy

It's similar to the `COUNTIF` function that we learned in the first lesson, except `AVERAGEIFS` uses multiple criteria to determine which values it should be acting upon.

First, let's go through an example together by calculating the averages of `windspeed` and `cnt` based on `season` and `yr`. Since there are four unique values in `season` and two unique values in `yr`, we will have a total of 8 averages for each column. To stay organized, we copy all the necessary columns to a new worksheet and create a table to store our averages, like this:



	A	B	C	D	E	F	G	H
1	season	yr	windspeed	cnt		**Season**	**Average Windspeed**	**Average Total Bike Rentals**
2	1	0	0.160446	985		Winter 2011		
3	1	0	0.248539	801		Spring 2011		
4	1	0	0.248309	1349		Summer 2011		
5	1	0	0.160296	1562		Fall 2011		
6	1	0	0.1869	1600		Winter 2012		
7	1	0	0.0895652	1606		Spring 2012		
8	1	0	0.168726	1510		Summer 2012		
9	1	0	0.266804	959		Fall 2012		
10	1	0	0.36195	822				

Steps to create averages using `AVERAGEIFS`:

1. Select cell G2. This cell will eventually show the average `windspeed` for *Winter 2011*.
2. Use fx (**Insert Function**) to search for and then select the *AVERAGEIFS* function.
3. The dialog box for defining **Function Arguments** will appear with two fields:
   - For **Average_range**, select the `windspeed` column (column C).
   - For **Criteria_range1**, select the `season` column (column A).
4. Once the above is filled out, a new field **Criteria1** will appear, which we set to `1` (referring to the dataset documentation, the `season` column uses `1` to represent *Winter*, `2` for *Spring*, `3` for *Summer*, and `4` for *Fall*).
5. Another field, **Criteria_range2**, will automatically be added, and we select the `yr` column (column B) for this argument.
6. Finally, we set **Criteria2** to `0`, which represents *2011*, and press *OK*. (`1` represents *2012*)

Our new table will look like this:

G2 | =AVERAGEIFS(C:C,A:A,1,B:B,0)

	A	B	C	D	E	F	G	H
1	season	yr	windspeed	cnt		**Season**	**Average Windspeed**	**Average Total Bike Rentals**
2	1	0	0.160446	985		Winter 2011	0.214441257	
3	1	0	0.248539	801		Spring 2011		
4	1	0	0.248309	1349		Summer 2011		
5	1	0	0.160296	1562		Fall 2011		
6	1	0	0.1869	1600		Winter 2012		
7	1	0	0.0895652	1606		Spring 2012		
8	1	0	0.168726	1510		Summer 2012		
9	1	0	0.266804	959		Fall 2012		

By copying and pasting the formula we just created in cell G2, we can fill out the rest of the *Average Windspeed* colum by editing the values for `criteria1` and `criteria2` for each row accordingly. Once this column has been filled out correctly, we can copy the entire *Average Windspeed* column over to the *Average Total Bike Rentals* column and set the **Average_range** argument for each row to the `cnt` column (column D). Here is the final result:

H9 | =AVERAGEIFS(D:D,A:A,4,B:B,1)

	A	B	C	D	E	F	G	H
1	season	yr	windspeed	cnt		**Season**	**Average Windspeed**	**Average Total Bike Rentals**
2	1	0	0.160446	985		Winter 2011	0.214441257	1666.666667
3	1	0	0.248539	801		Spring 2011	0.205777647	3775.173913
4	1	0	0.248309	1349		Summer 2011	0.177307932	4464.361702
5	1	0	0.160296	1562		Fall 2011	0.168132129	3664.460674
6	1	0	0.1869	1600		Winter 2012	0.214631837	3531.296703
7	1	0	0.0895652	1606		Spring 2012	0.201236614	6209.48913
8	1	0	0.168726	1510		Summer 2012	0.16688211	6824.244681
9	1	0	0.266804	959		Fall 2012	0.175857022	5791.865169

2.6.1 Instructions

1. Create a table of averages similar to the example in the **Learn** section above for the `temp` and `cnt` columns using the same `season` and `yr` criteria.

    - Copy all required columns to a new worksheet before creating the table of averages.

2.7 Combo Charts

Now that we have organized our data, it's time to create a data visualization for the investors of Capital Bikeshare. They have concerns about the fluctuations in bike rentals throughout the year. Our goal is to put their minds at ease by showing that these ups and downs are nothing more than seasonal trends due to yearly temperature changes. To show this, we will use a combination of a column chart and a line chart by creating what's called a **combo chart**. From the previous screen, here is the data we will be using for our final chart:

F	G	H
Season	**Average Temperature**	**Average Total Bike Rentals**
Winter 2011	0.27352846	1666.666667
Spring 2011	0.534252946	3775.173913
Summer 2011	0.701173638	4464.361702
Fall 2011	0.426444404	3664.460674
Winter 2012	0.321700495	3531.296703
Spring 2012	0.55455737	6209.48913
Summer 2012	0.711444979	6824.244681
Fall 2012	0.419367674	5791.865169

What is a combo chart?

As the name suggests, a combo chart is simply a combination of charts. They allow us to display different types of data in different ways on the same chart. We can use a combination of column, bar, scatter, line, or even pie charts to create a combo chart. That said, not all combinations will be effective, so we should take care when choosing chart types to combine. For example, a pie chart and a bar chart do not make for a good combination, whereas a column chart and a line chart are a very popular combination. This is the combination we will use.

When we combine two or more columns of data into a combo chart, they will be charted along the same independent axis. However, their dependent values can be charted along separate dependent axes if necessary. This gives us the option to compare trends between columns that are measured differently. If this seems confusing, don't worry: we'll work through an example together below to clarify.

When to use a combo chart

General guidelines:

- We have multiple columns that share the same independent values.
- We want to compare multiple columns with different value ranges.
- We want to show the correlation between two columns in one chart.
- We want to show whether one column of values meets a target defined by another column.

How to create a combo chart in Excel

Let's create a couple charts using the following data to show two common use cases for combo charts:

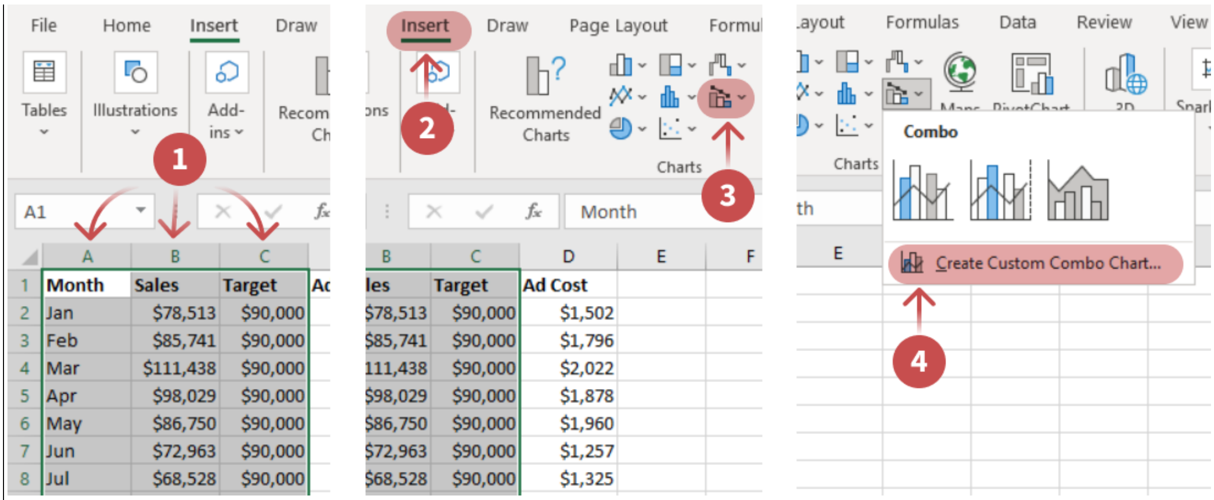
	A	B	C	D	E
1	**Month**	**Sales**	**Target**	**Ad Cost**	
2	Jan	$78,513	$90,000	$1,502	
3	Feb	$85,741	$90,000	$1,796	
4	Mar	$111,438	$90,000	$2,022	
5	Apr	$98,029	$90,000	$1,878	
6	May	$86,750	$90,000	$1,960	
7	Jun	$72,963	$90,000	$1,257	
8	Jul	$68,528	$90,000	$1,325	
9	Aug	$81,617	$90,000	$1,702	
10	Sep	$92,007	$90,000	$1,976	
11	Oct	$106,894	$90,000	$2,098	
12	Nov	$94,355	$90,000	$1,809	
13	Dec	$80,456	$90,000	$1,696	
14					

First we'll create a combo chart to compare the dependent *Sales* and *Target* columns along the common independent *Month* column. Next, we'll create a combo chart to compare the dependent *Sales* and *Ad Cost* columns along the common independent *Month* column.

Steps to create a combo chart:

1. Select the desired data, making sure the independent values are to the left of the dependent values.
2. Select the **Insert** tab at the top of the spreadsheet.
3. Select **Insert Combo Chart** from the **Charts** section (a column chart with a line above it).
4. Select **Create Custom Combo Chart** at the bottom.

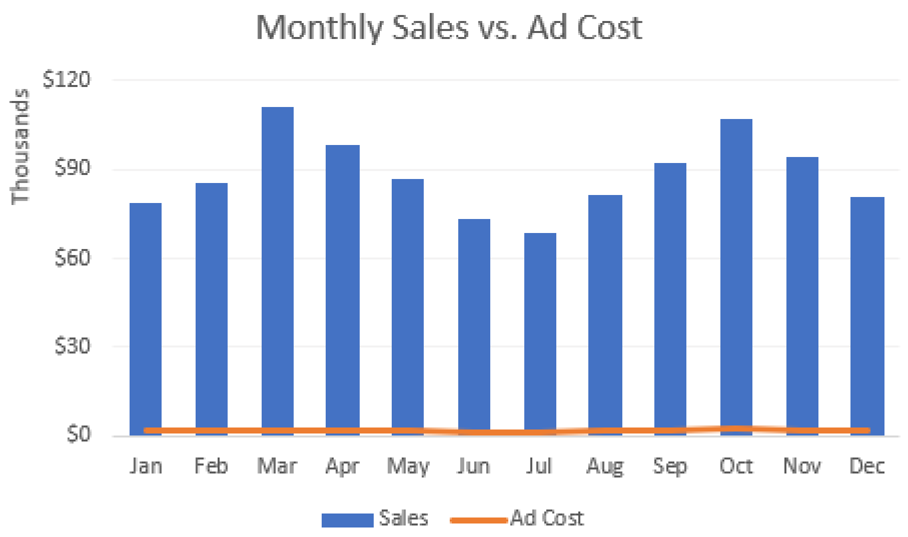
Here are the same steps represented visually:



After selecting **Create Custom Combo Chart**, the **Insert Chart** window will appear. After making sure *Sales* is set to *Clustered Column* and *Target* is set to *Line*, we click *OK* to create the combo chart. As always, we can customize our chart by formatting individual elements and/or by using **Chart Elements**. After changing the title and applying formatting as we did with previous charts, here is our first combo chart comparing *Sales* to *Target*:

By combining the data into one chart, we can quickly see which months met their target and which ones didn't. As you can see, this chart is very easy to interpret, so it's a great choice when presenting data to others.

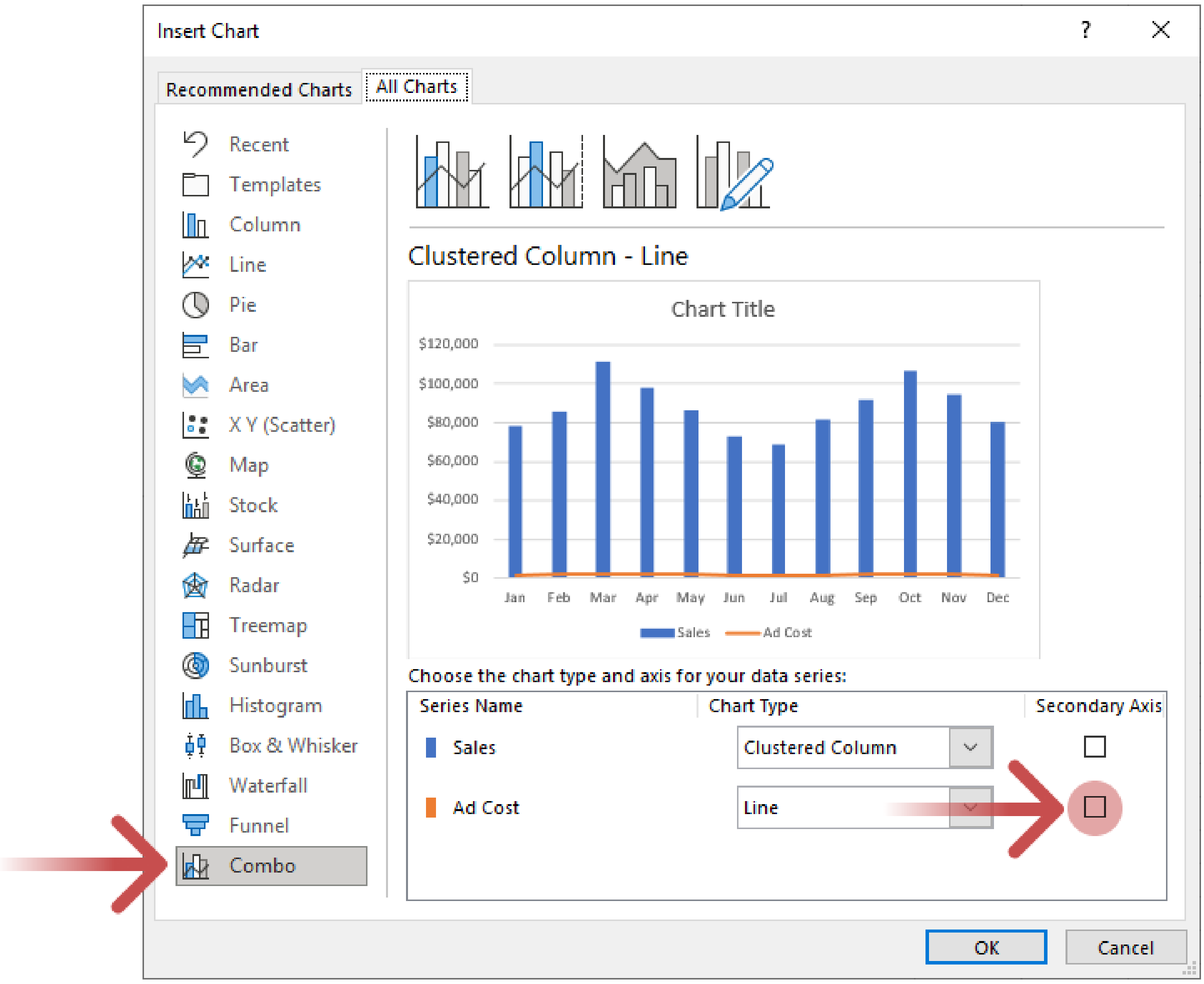
Following the same steps as above for the *Sales* and *Ad Cost* columns, we get the following chart:



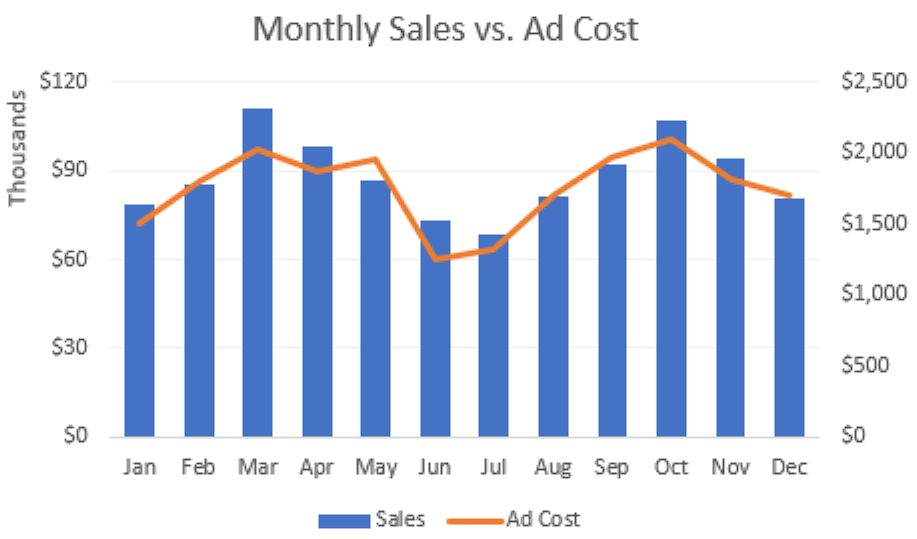
That doesn't look right! What happened? Since the values in the *Ad Cost* column are so much smaller than the values in the *Sales* column (they are on a different scale), when they are charted along the same dependent axis, we can barely see any difference month to month for *Ad Cost*.

To fix this, we need separate dependent axes for *Sales* and *Ad Cost*. In other words, we need to add a **Secondary Axis** to the chart. This means we will have two dependent axes in one chart; one on the left and one on the right.

To enable this feature, we can simply select **Secondary Axis** when the **Insert Chart** window appears during the creation of the chart. Or, if the combo chart has already been created, we can enable a **Secondary Axis** by right-clicking the chart and selecting **Change Chart Type**. In either case, we get a window that looks something like this:



After we enable a **Secondary Axis** for *Ad Cost*, we get our desired result:



That looks a lot better! We can now see that *Sales* and *Ad Cost* usually follow a similar pattern month to month. As one increases/decreases, so does the other.

Now let's practice!

2.7.1 Instructions

1. Create a combo chart for the table of averages you created on the previous screen.
    - Select the entire table, and insert a custom combo chart.
    - Use **Clustered Column** for *Average Temperature*.
    - Use **Line** for *Average Total Bike Rentals*.
    - Enable a **Secondary Axis** for *Average Temperature*.
2. Customize the chart to look similar to the example above:
    - Give the chart a meaningful title.
    - Change **Display units** to **Thousands** for the left vertical axis.
    - Change **Gap Width** to **100%** for *Average Temperature*.
    - Soften the appearance of **Gridlines**.
    - Remove the chart **Border**.
3. Below the chart, add a description explaining the relationship between *Average Temperature* and *Average Total Bike Rentals*. Is there an observable difference between 2011 and 2012?