EY // Knack Training

Data Visualization

Using Excel for Compelling Charts

Overview

Excel Data Visualization is structured to challenge the participants with advanced charting features and design theory. We begin by focusing on the application of common design concepts to client deliverables, then build a variety of charts that demonstrate the effectiveness of those concepts. We also demonstrate advanced tips and tricks to encourage the student to explore and stretch their knowledge of Excel charting processes.

Learning objectives:

- Chart choosing and formatting
- Using design principles in formatting choices
- Creating non-standard Excel charts
- Overlapping data sets in charts for unconventional results
- Creating small multiple charts

Table of Contents

Chart Choosing

Because various chart types encode values as dots, circles, squares, lines, and other visual elements, they each have data *messages* that they are best suited to deliver. For more chart picking documentation, see the following resources:

- Depict Data Studio's Chart Chooser
- Berkeley School of Journalism's Chart Picking Guidelines
- Evergreen Data's Qualitative Chart Chooser

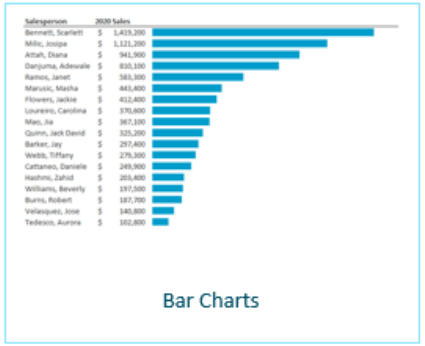
When choosing a chart, consider the **question** the data should be answering:

- How big are the values, relative to one another?
- What values are the biggest components of the whole?
- How are values changing over time?
- Is there a relationship between values that predicts a change?

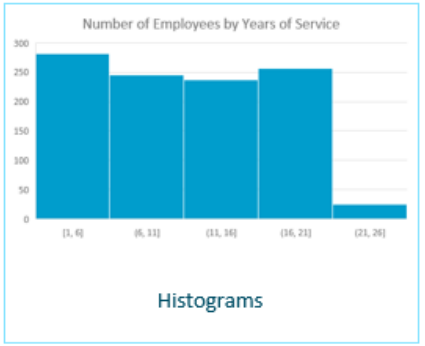
We have laid out this grid in the related spreadsheet, **Chart Choosing.xlsx**, if you would like to view each as an active Excel chart. Additionally, some of these charts are conventional Excel charts, and others are built through a custom manipulation of available Excel tools. This is incomplete, but hopefully it helps you make better decisions when starting your visualization.

How big are values, relative to one another? To answer this question, we need to see the size of the values encoded clearly. The *whole* is not as important as sizes of individual values. In these scenarios, consider the following chart types:

Column Charts

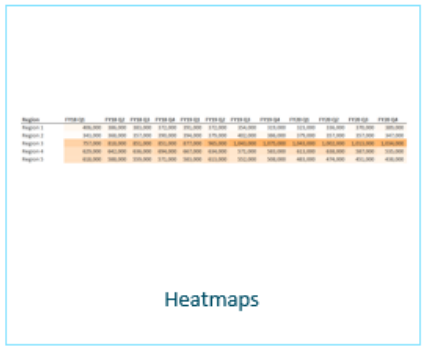


Bar Charts



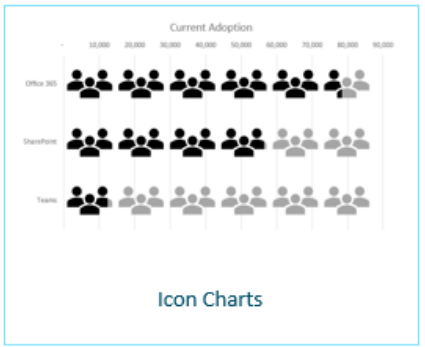
Histograms

Pareto Charts

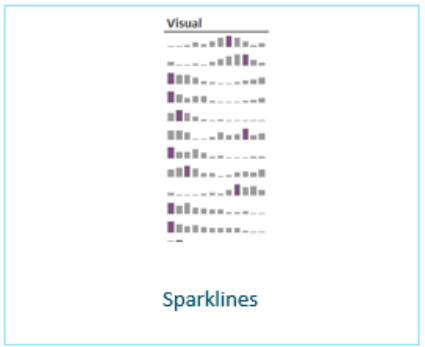


Heatmaps

Chloropleths



Icon Charts



Sparklines

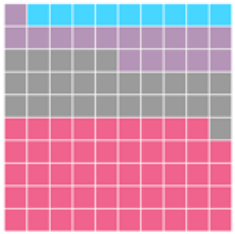
Clustered Column and Clustered Bar Charts are the most common and easily read chart types for this type of information. Further, notice that Histograms, Pareto Charts, Icon Charts, and Sparklines are each a different version of columns and bars. The reason these charts are best for this type of data is we only read a single dimension (down the bar or column) and often have a clear label for the amount being compared. This is much more easily read than a Pie Chart, which uses angles and arcs to encode the same information.

Additionally, color-coded charts like Heatmaps and Chloropleths (color-coded maps) encode the relative value in the darkness, intensity, or hue of a color. For this reason, we must consider readability for those with visual impairments.

What values are the biggest components of the whole? To answer this question, we must be able to see the values as *percentages*, rather than raw numbers. For this analysis, it is difficult to compare many values in a single chart, so simpler data sets are more appropriate to this kind of comparison.

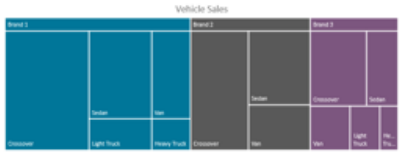
Fewer Components / Lower Complexity

Pie/Donut Charts



Waffle Charts

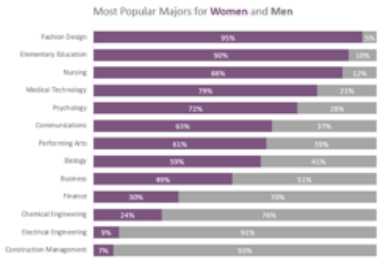
More Components / More Complexity



Treemaps

Sunbursts

100% Stacked Columns



100% Stacked Columns

One of the most common chart types here is a Pie Chart. Pie Charts are the best solution for components that are near 25, 50, or 75% of the whole, since those are the easiest angles and arcs for most people. However, creating a Waffle Chart – a 10x10 grid – allows us to more easily count up blocks reaching 100%. 9% is obviously different from 10% in this chart type.

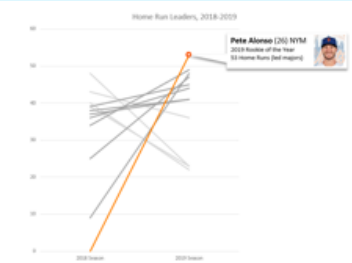
Stacked Column Charts and Stacked Bar Charts make it easy to compare multiple percentages.

Treemaps and Sunbursts – the new Hierarchy Charts in Excel – do a good job of giving us larger blocks, broken down into segments we can analyze more carefully.

How are values changing over time? When considering time-based data, *lines* are the most common and easily-read encoding for the data – because one point in time is physically connected to another by a line. Area charts are line charts with the area underneath filled with color – helpful for giving more weight to the values.

Change in Value

Line Charts

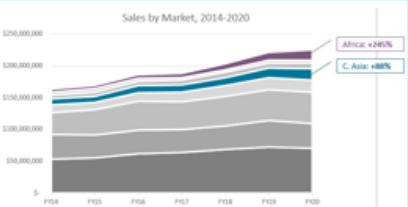


Slope Graphs

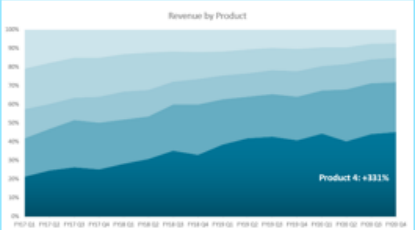
Cleveland Dot Plots

Bump Charts

Change in Composition

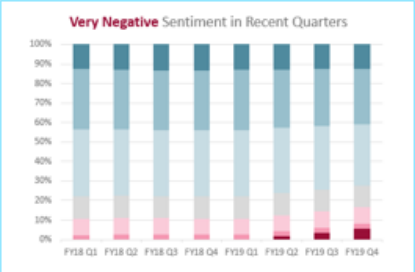


Stacked Area Charts



100% Stacked Area Charts

Stacked Column Charts

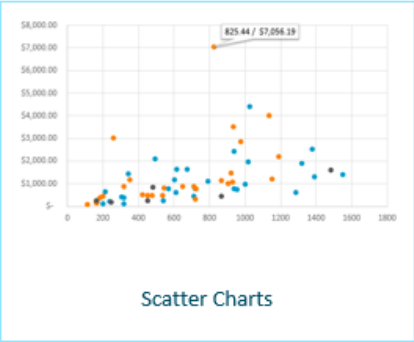


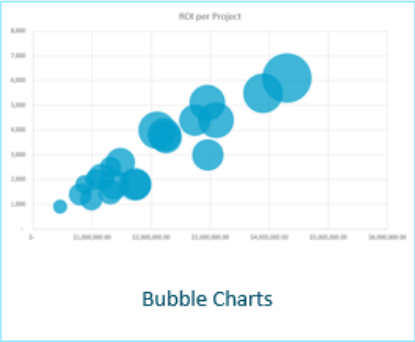
100% Stacked Column Charts

While Line Charts compare value changes, Area Charts compare composition changes – notice how easily you read the total sales going up in the Stacked Area Chart, for example. The relative sizes of individual slivers are more difficult to read. You can also used Stacked Columns for this analysis, but often the lost connection (without the line) makes them harder to compare.

Note two unconventional Line Charts – the Slope Graph (nothing more than a two-point Line Chart) and the Bump Chart (a line chart for rankings, instead of absolute values). These are both created in Excel by repurposing the Line Chart. Finally, Dot Plots are useful for comparing the distance between two values between two points in time.

Is there a relationship between values that predicts a change? This question is lesser-known, but can often be the most important for a business. For this, we must encode multiple, related values to see the relationship between them.





The first two chart types – Scatter Charts and Bubble Charts – are related to one another. A Bubble Chart is simply a Scatter Chart with one additional piece of information (the size of the bubble). In these, we see the relationship between the X-dimension and the Y-dimension, and are often able to draw a line that describes the relationship between the values.

Pareto Charts encode a Column Chart that descends in value, but then superimposes a Line Chart representing the percentage of the whole. This helps us see how important each value is to the whole, while simultaneously seeing the individual values.

Finally, Radar Plots are used to describe multiple values simultaneously. The hope is that by seeing the values plotted together, the viewer can find a commonality. Above, we see three individuals who are similar because of similar radar shapes, and three others who are similar to one another, as well.

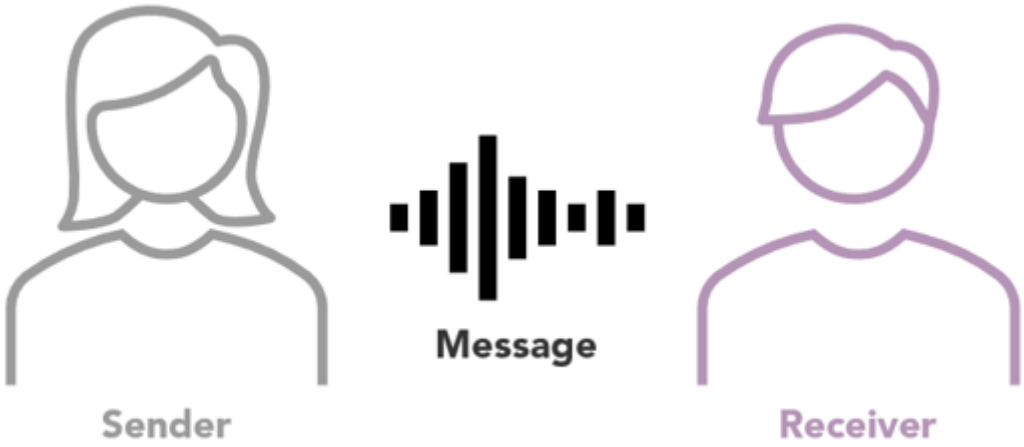
Design Theory in Data Visualization

Design is the conscious **choice** of colors, sizes, fonts, alignment, spacing, and other aesthetic qualities so our communication is more **effective**.

How can aesthetic choices make your communication more effective?

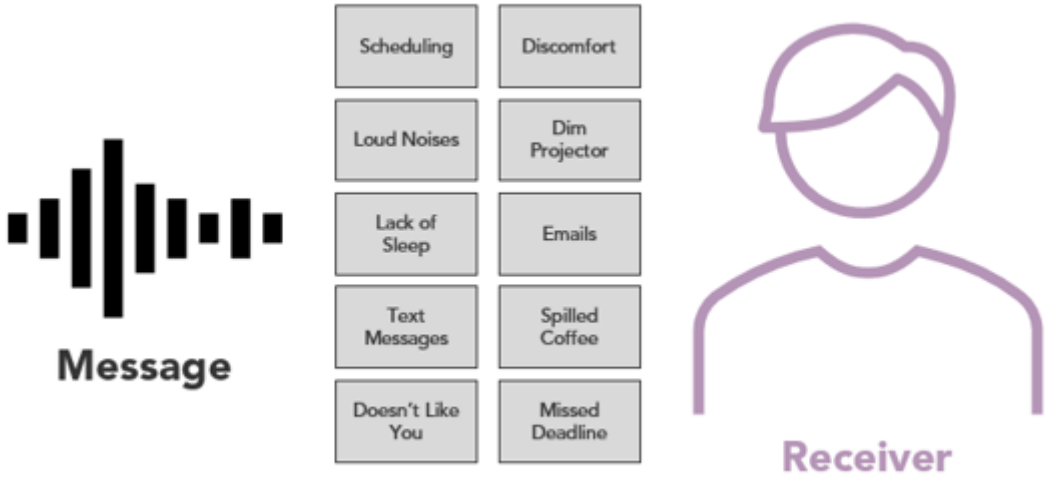
Core Design Principles

Signal vs. Noise: In communication theory, the **sender** has a **message** to deliver across a **channel** to the **receiver**.



The metaphor that is often used in communication theory comes from the world of radios and televisions: **signal** (message) vs. **noise** (everything else). In the same way that we change the radio station when there is too much static, our attention moves on when there is too much 'everything else' blocking the core of the message.

At this point, you may be able to connect the dots and assume that we'll talk about removing 'noise' from our charts, slides, etc. However, before we do that, there is an essential truth we must address. **Noise is everywhere.** Noise includes details like "the room is too hot," "my baby woke up at 3 am," and "I'm already late for my next meeting."

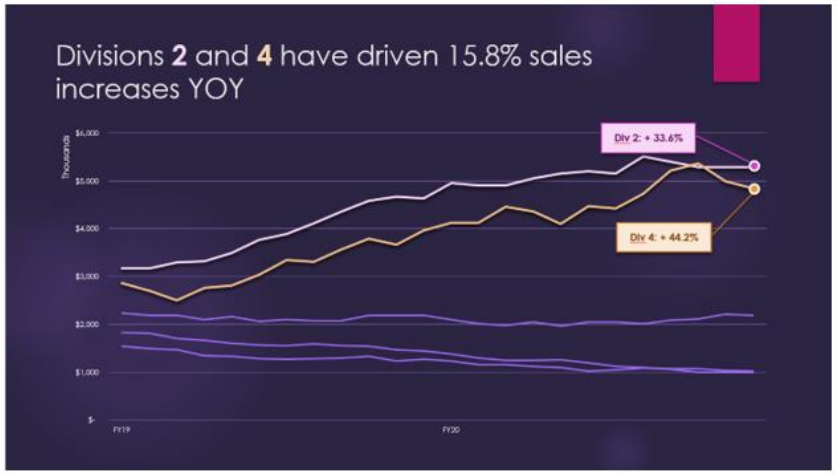


Because there is so much noise that the communicator has no control over – we can't give them back a missing three hours of sleep, for example – we *must* control every factor we can.

Simplifying & Emphasizing is the pair of tasks that make our visualizations full of message and devoid of noise. Let's take a look at what simplifying a slide and emphasizing the message looks like.

Visual Noise:

- Two titles on the page
- Five different colors in the lines of the chart
- Bright purple behind the chart makes the lines less legible
- Axis labels have large numbers and a *lot* of increments on the timeline – these labels are unreadable and complicated



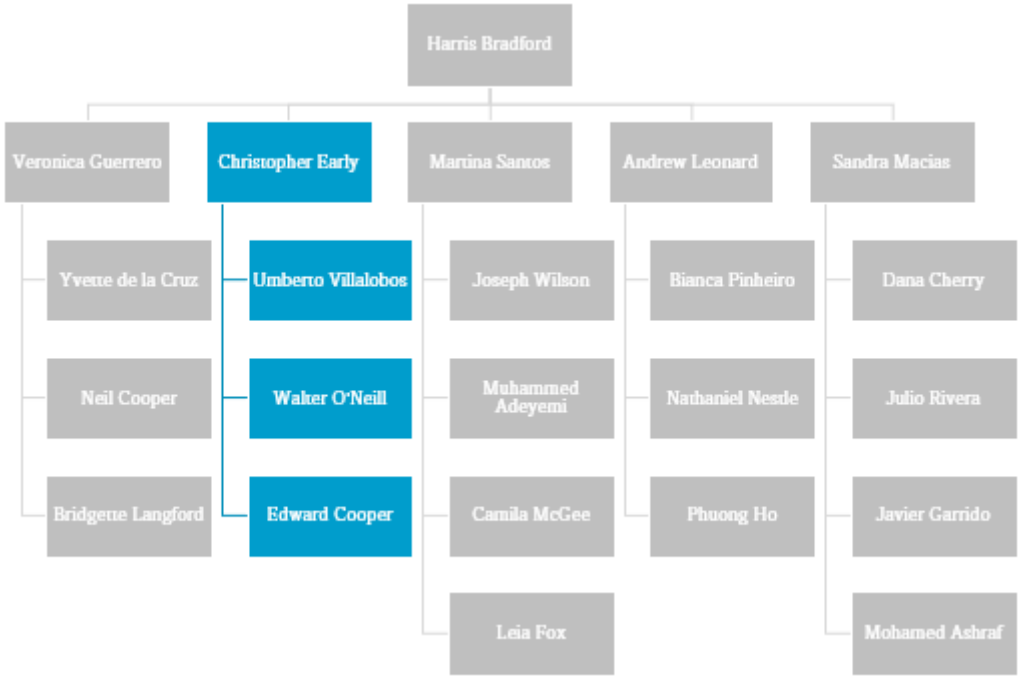
Removing noise:

- One title on the page
- Two contrasting colors for the focal lines, subtler purple for the others
- Dark purple background, making the lines more legible
- Axis labels have been simplified

Adding emphasis:

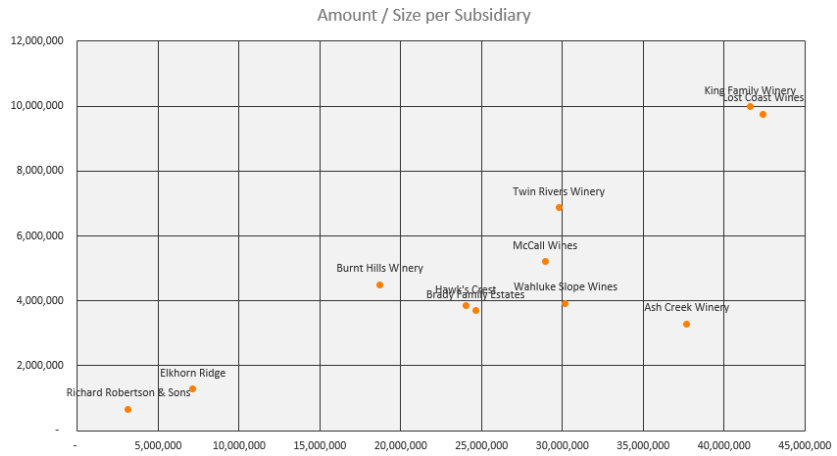
- Writing the message into the text makes the point of the slide clear
- Color-coding the lines, labels, and title connect all elements to the message
- Added labels call additional attention to the important lines on the chart

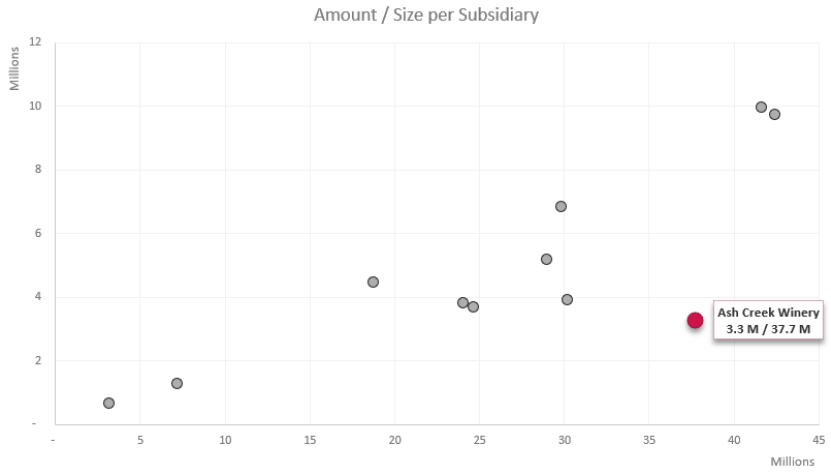
Repetition & Contrast: In our minds, the elements that are the same color are categorized together (repetition of the visual element, establishing a pattern), while the elements with drastically different color is categorized differently (it breaks the pattern). Without any additional text, the viewer understands they should look at the Christopher Early column of the organization chart.



A note about color: When considering which colors to choose for contrast, use a site designed for accessibility concerns, like https://www.color-blindness.com/coblis-color-blindness-simulator/. On this site, you'll see that choosing colors like red and green would not generate the contrast you're looking for, because the colors are too similar for those who have different varieties of color blindness.

These ideas: **simplification and emphasis**, partnered with **repetition and contrast**, make the point of the chart stand out very clearly. Look at these Scatter Plots side-by-side:





The chart on the left is more *complete* – it labels every point. However, the chart on the right is more *clear* – it uses simplification and repetition to inform the viewer that most of the points are irrelevant. Then, by using contrast to emphasize the most critical information, it becomes clear the viewer should focus on Ash Creek Winery.

Design guidance: eliminate unnecessary information from your visualizations, only including the elements that provide clear information to the viewer. This includes removing unnecessary colors that contribute to the visual noise of your communication. Then, once your chart is simple and repetitive, contrasting colors, alignments, sizes, or other visual elements clearly bring the focus to your most critical information.

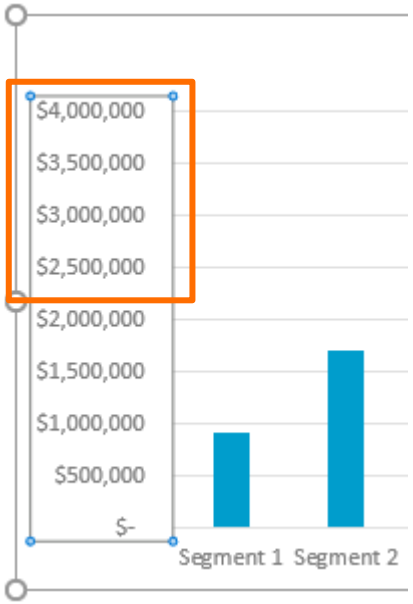
Custom Labeling

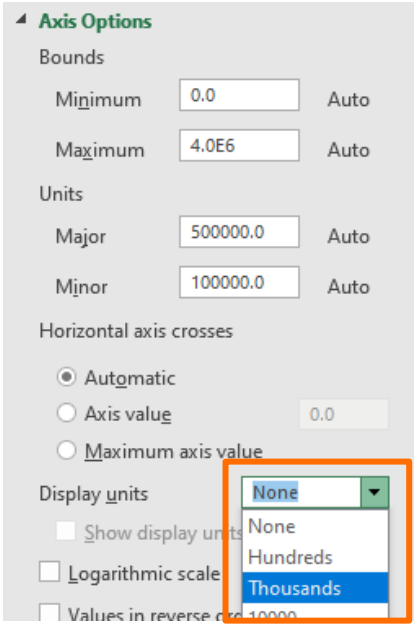
The text in your visualization provides critical information, so it should be clear. It also competes with the graphical elements, so it should not add visual complexity.

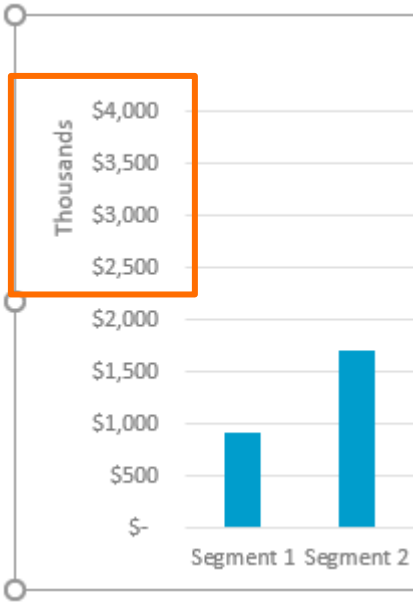
However, we should avoid manually-typed labels and titles whenever possible, to avoid errors when the underlying data changes.

Customizing Axis Labeling

To change the labeling of an axis, click the axis in question and use the keyboard shortcut **[CTRL] + [1]**. This opens the Formatting Pane, and allows you to make the following changes:

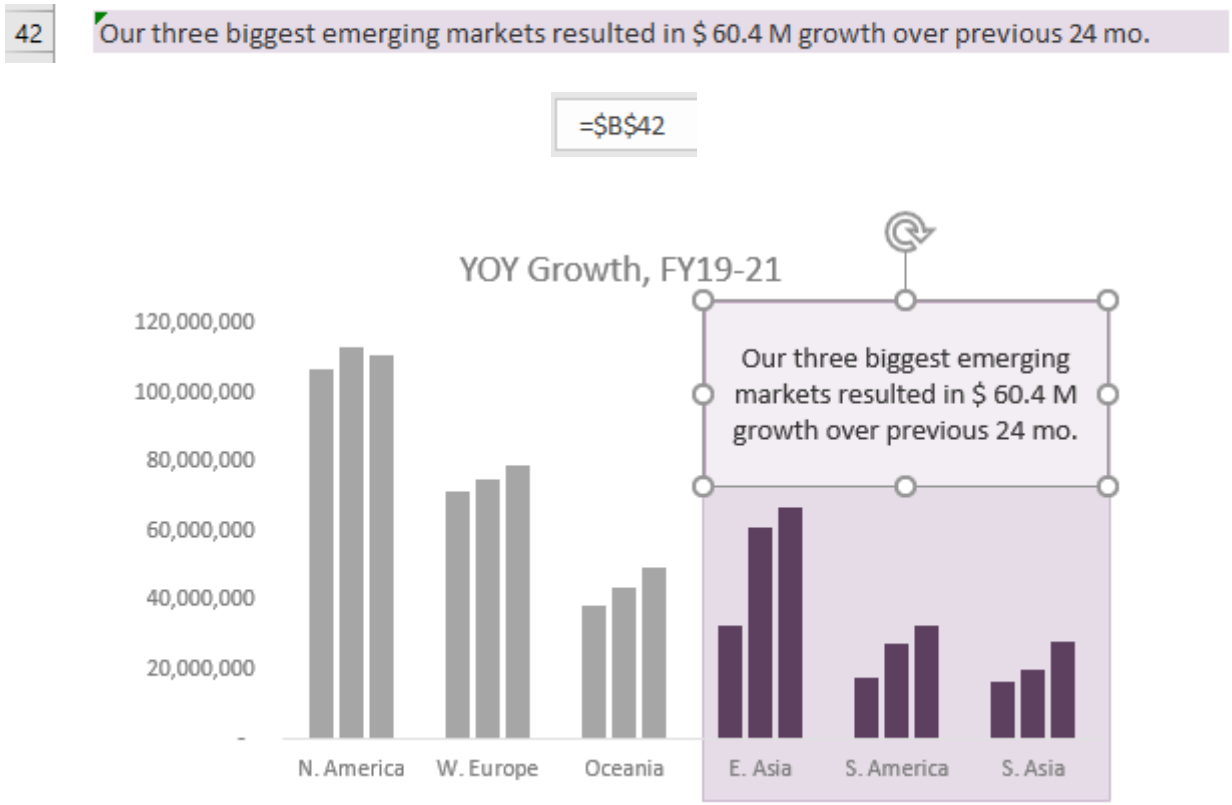






Inserting Cell Contents into Textboxes

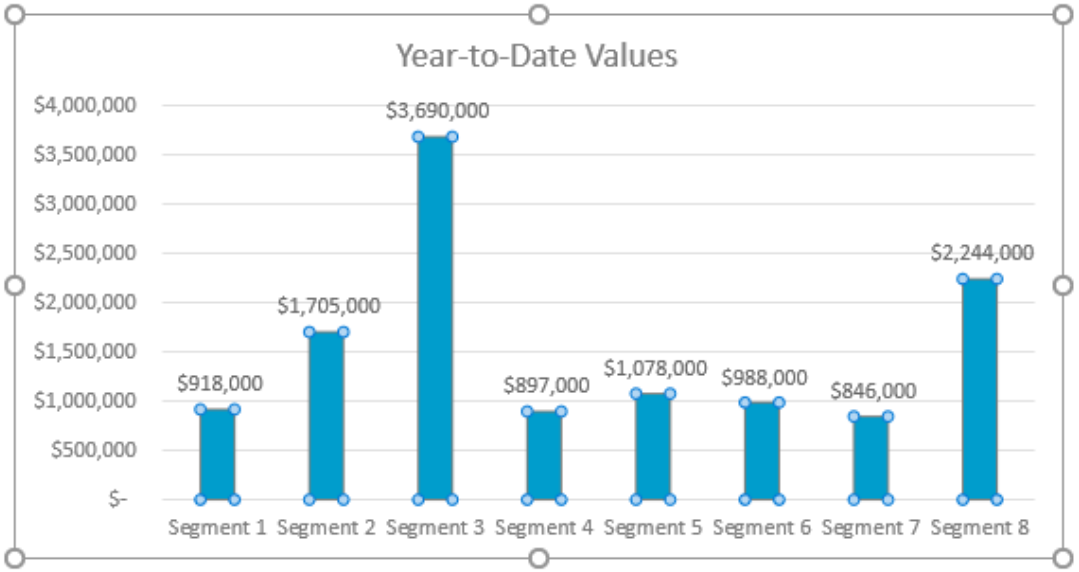
When you insert a textbox (or use a Chart Title), that textbox can be linked to the contents of a cell by clicking into the Formula Bar and referencing the cell (like **=B42**, for example):



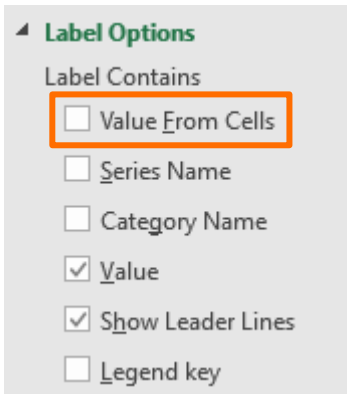
Using Additional Data for Labeling

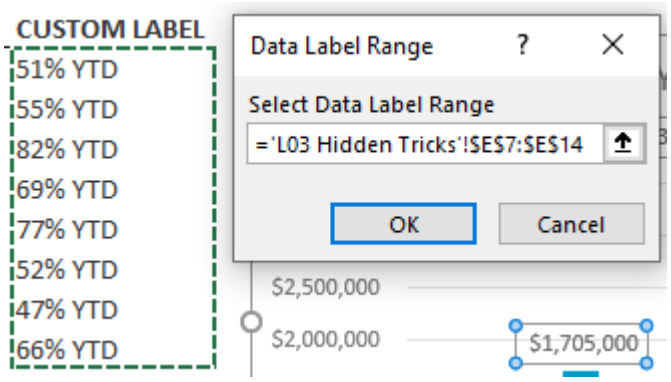
Typically, Data Labels are connected to the cells that are used in the chart – for example, the data labels on these columns are pulling the YTD value of **$918,000** for the first column:

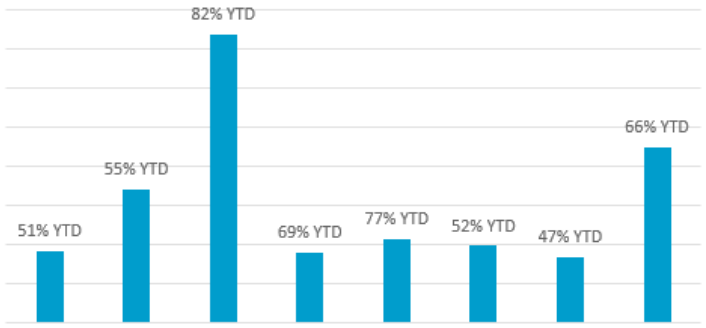
SEGMENT	QUOTA	YTD	CUSTOM LABEL
Segment 1	$ 1,800,000	$ 918,000	51% YTD
Segment 2	$ 3,100,000	$ 1,705,000	55% YTD
Segment 3	$ 4,500,000	$ 3,690,000	82% YTD
Segment 4	$ 1,300,000	$ 897,000	69% YTD
Segment 5	$ 1,400,000	$ 1,078,000	77% YTD
Segment 6	$ 1,900,000	$ 988,000	52% YTD
Segment 7	$ 1,800,000	$ 846,000	47% YTD
Segment 8	$ 3,400,000	$ 2,244,000	66% YTD



However, by clicking onto the labels themselves and using the keyboard shortcut **[CTRL] + [1]**, we see that we can connect the labels to a *different set of cells*.







Leveraging the TEXT Function

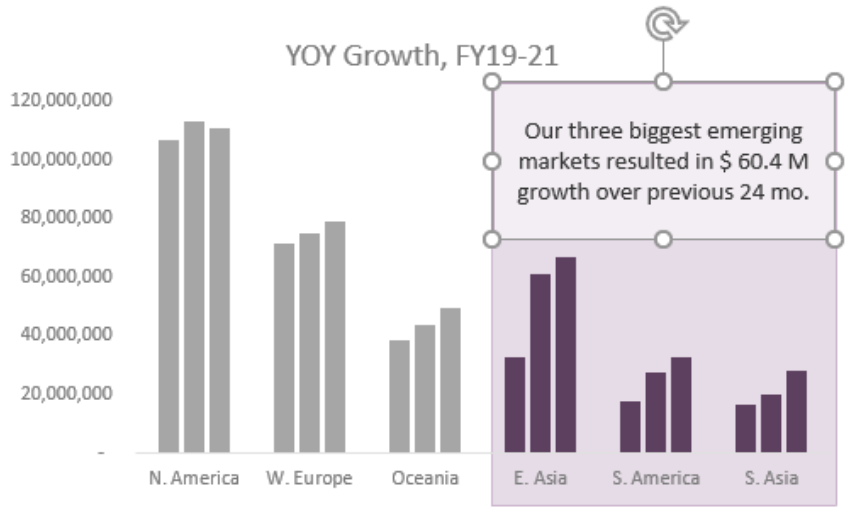
The TEXT function takes a numeric value and a custom number formatting code, and outputs the value with that formatting. To learn more about custom number formatting codes, please [see the Microsoft site](#).

Note the ways this can be used for labeling:

```
=TEXT(D7/C7,"0%")&" YTD"
```

CUSTOM LABEL
51% YTD
55% YTD
82% YTD
69% YTD
77% YTD
52% YTD
47% YTD
66% YTD

```
="Our three biggest emerging markets resulted in "&TEXT(SUM(E30:E32)-SUM(C30:C32),"$ 0.0,,")&" M growth over previous 24 mo."
```



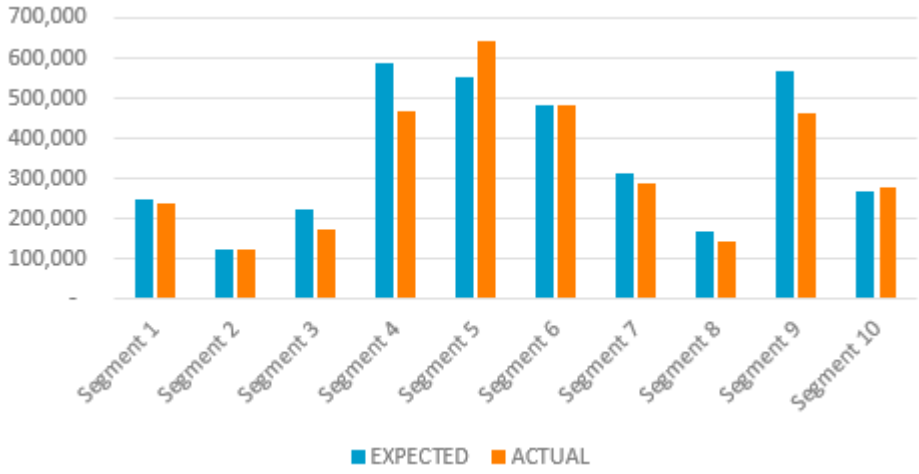
Overlapping Data

In an Excel chart, it is often possible to overlap one data element on top of another to create a useful visual effect. Here, we'll look at two examples:

Overlapping a Column Chart to 'Fill Up'

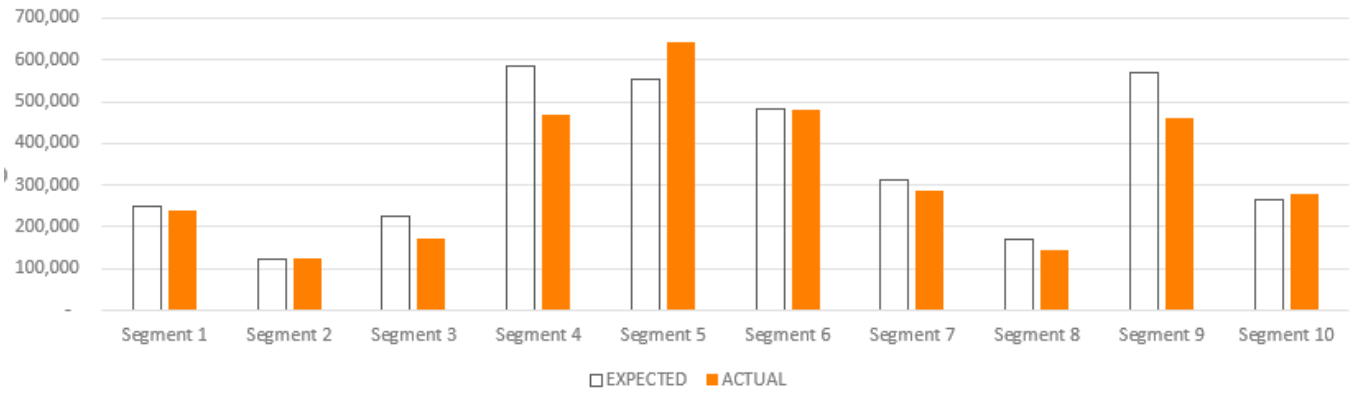
Here, we'll take two values – the Expected Value and the Actual Value. If the main message of our chart is to compare how many of these values met or exceeded the expectations, it will be important to visually compare the two numbers easily. Our default Column Chart would look like this:

SEGMENT	EXPECTED	ACTUAL
Segment 1	250,500	240,500
Segment 2	123,900	125,200
Segment 3	225,900	174,000
Segment 4	585,200	468,200
Segment 5	552,700	641,200
Segment 6	481,600	481,600
Segment 7	312,600	287,600
Segment 8	170,100	142,900
Segment 9	569,500	461,300
Segment 10	266,800	280,200

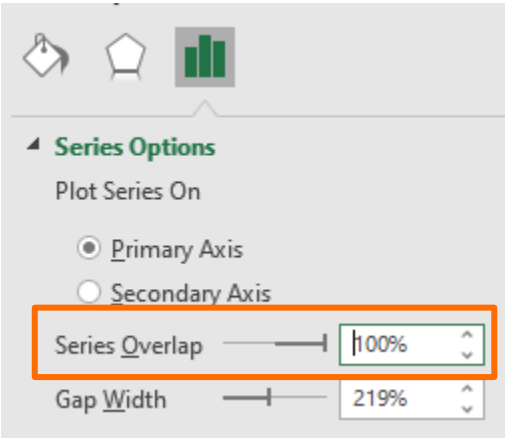


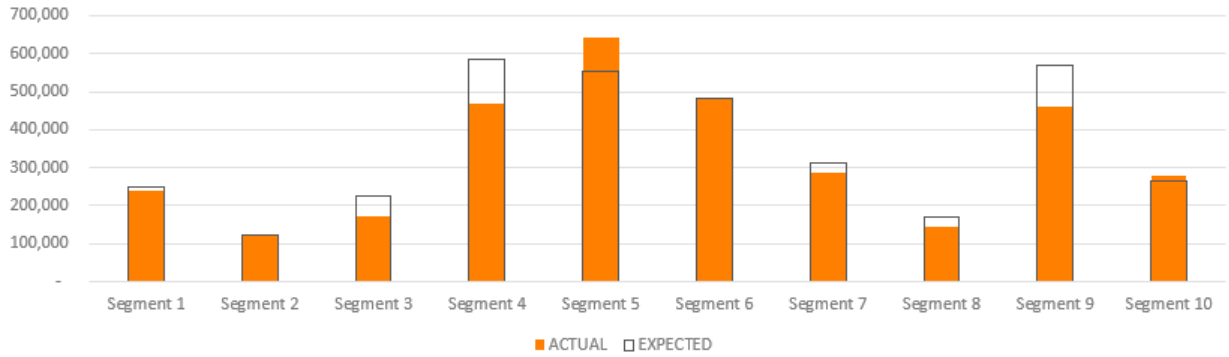
How would we improve this chart? Well, the viewer must currently move their eye back-and-forth in each Expected-Actual pairing, to compare the heights of the columns. It is very likely the viewer will *forget* which is the Expected and which is the Actual, making the experience frustrating. We would like to make the comparison easier, and make it clear which number is the Expected vs. Actual.

Create an Expected Amount that needs to be 'filled' by changing the column formatting to be more hollow:



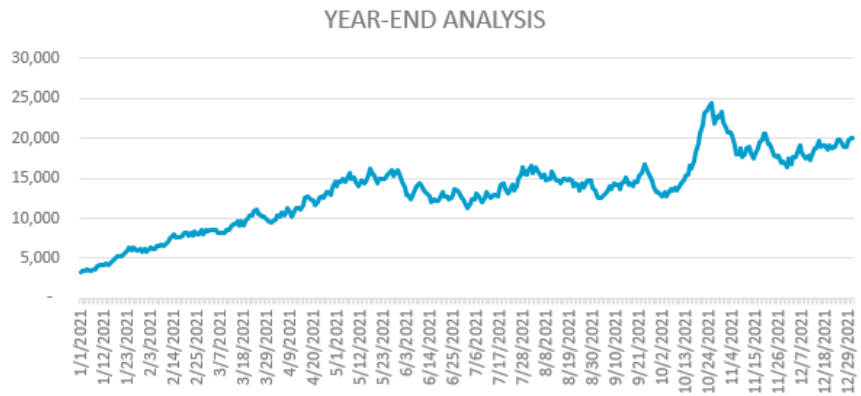
Now, use the keyboard shortcut **[CTRL] + [1]** to format that column further, and change the **Series Overlap** value to **100%.**





Now it is very clear when an Actual value under- or over-performs the Expected value.

Overlapping a Line Chart with Lines and Areas can bring emphasis to different elements. So, this line chart can become:



This is a flexible highlighting of a segment of the line, bringing the viewer's eye to an important insight. To do this, first create a line chart. Then, determine which portion of the line chart you'd like to highlight. In this example, we are deciding it dynamically with the IF function:

DATE	AMOUNT	HIGHLIGHTED
1/1/2021	3,200	#N/A
1/2/2021	3,400	#N/A

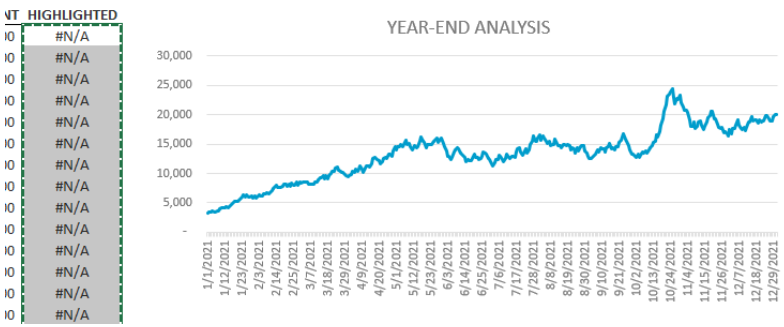
```
=IF(AND(R4>=$X$21,R4<=$X$22),S4,NA())
```

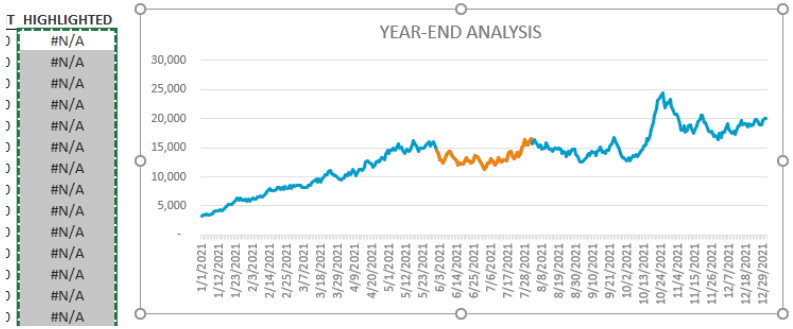
Begin Range:	6/1/2021
End Range:	8/1/2021

So here, we've compared the date **1/1/2021** with **6/1/2021** and **8/1/2021**. Because the date in question is not within the range, we've decided **not** to highlight it. However, note the essential element in this trick: the **NA()** function.

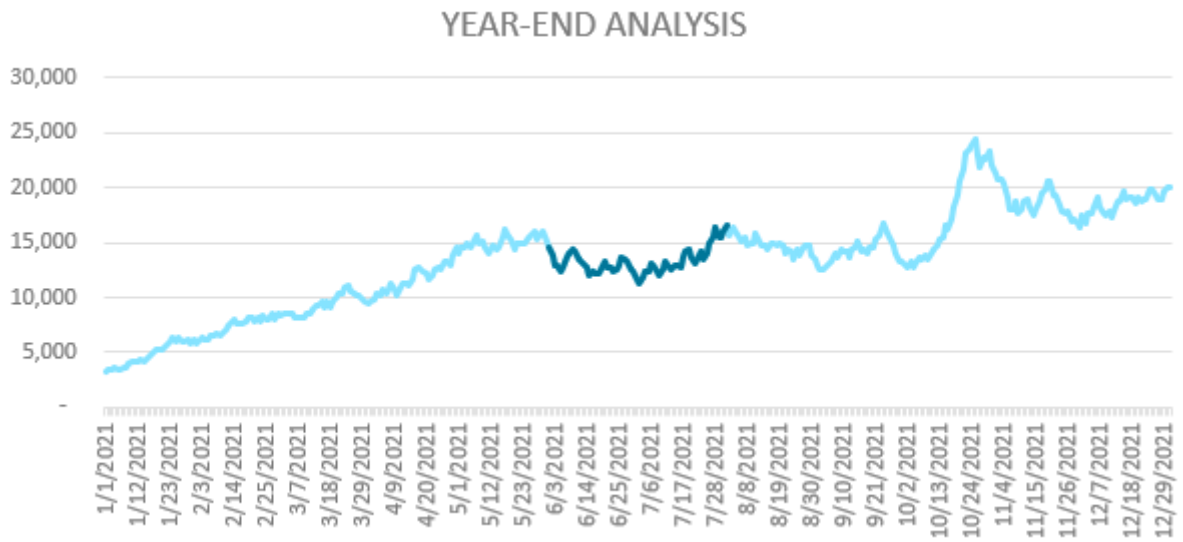
In a highlighted line or area chart, we want a *segment* of a line chart to do the highlighting. In the areas where there is no line, we do *not* use the value **0** because that would result in a line at 0. Instead, we use NA to generate an **#N/A** error message that is excluded from the chart.

Including the highlight in the chart is as easy as selecting the data in question, using **[CTRL] + [C]** to copy the data, then selecting the chart and using **[CTRL] + [V]** to paste it in:





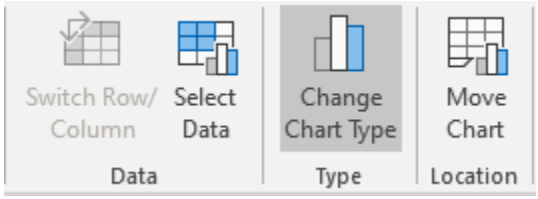
You can now select and format the larger line and the line segment in any way you like to highlight the segment.



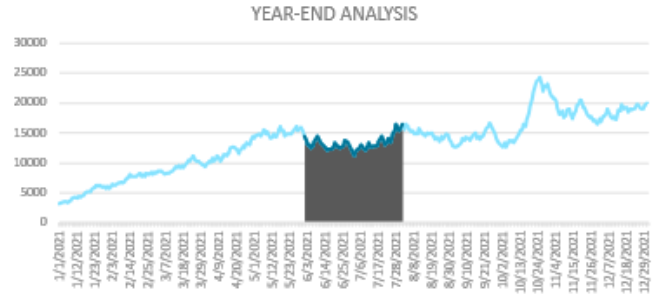
Additionally, you can highlight under the line segment to emphasize it even more. Copy-and-paste the same data set into the chart a second time, and note that you now have two line segments:

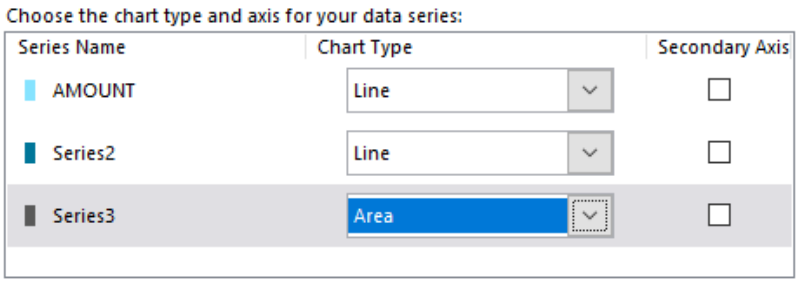
This has a dark blue line *and* a grey line superimposed on top of the bright blue line.

Now, click **Design Tab > Change Chart Type**:



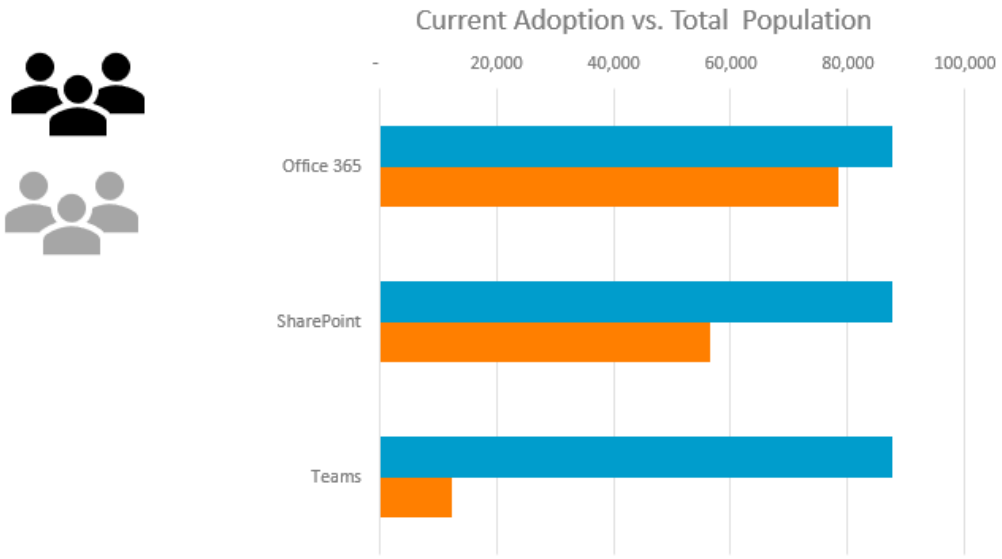
Choose the **Combo Chart** Type, and set the first two series to **Line Chart**, the third to **Area Chart**:





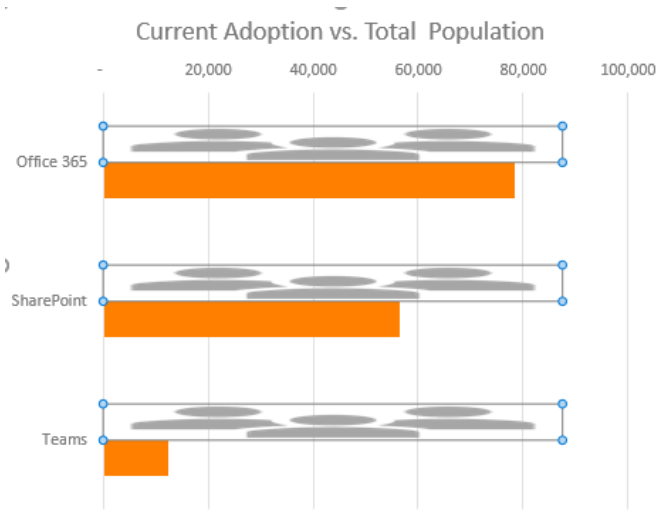
Icon Charts

The **Fill** property of any chart element (pies, bars, columns, bubbles, etc.) can be set to **Picture or texture**, which means you can insert logos, photos, or icons into your charts. A simple example of this is inserting icons of people into charts that represent people:

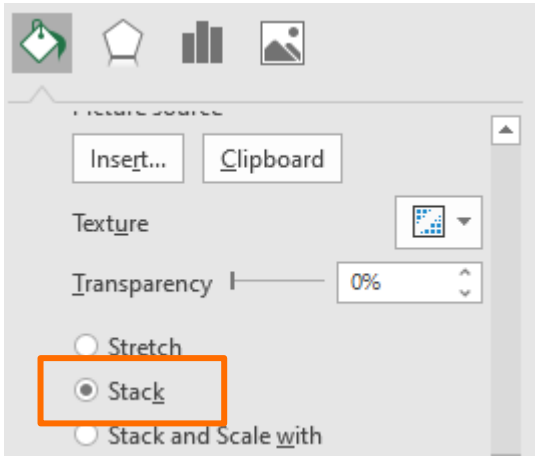


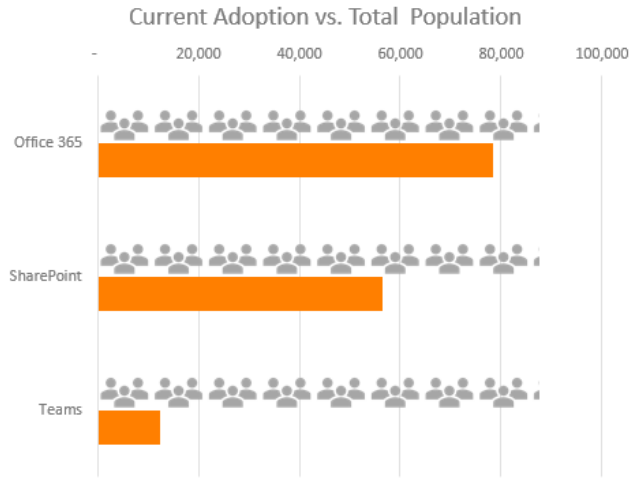
In this example, we have inserted icons using Office 365's icon gallery. If you are using an older version of Excel, you can also find icons from any online service, etc. Note that here, we have color-coded the icons so that the **Total Population** will be grey, and the **Current Adoption** will be black.

First, **copy** the icon you want. Then, click the bar to be filled with this icon. Use the shortcut **[CTRL] + [V]** to paste this icon into the bars:

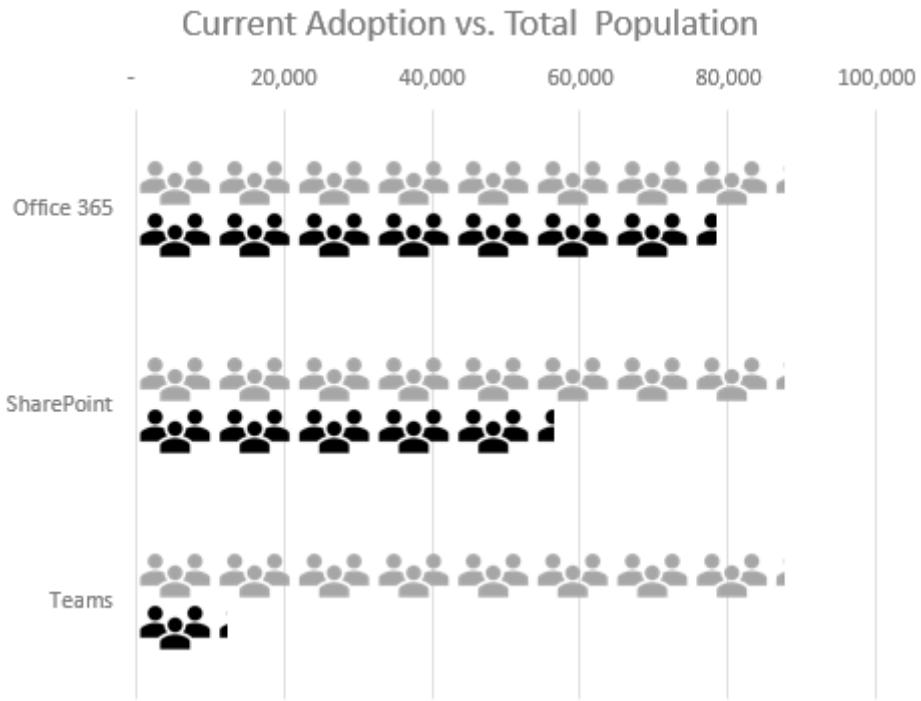


The grey icon is being **stretched**, so use the shortcut **[CTRL] + [1]** to change the formatting. Note the fill setting for **Stack** – change to that selection:

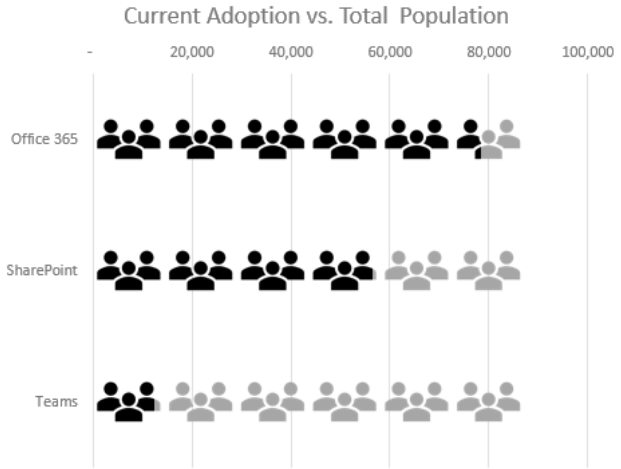




Now, in this example, we can repeat the process for the black icons and the **Current Adoption** columns:



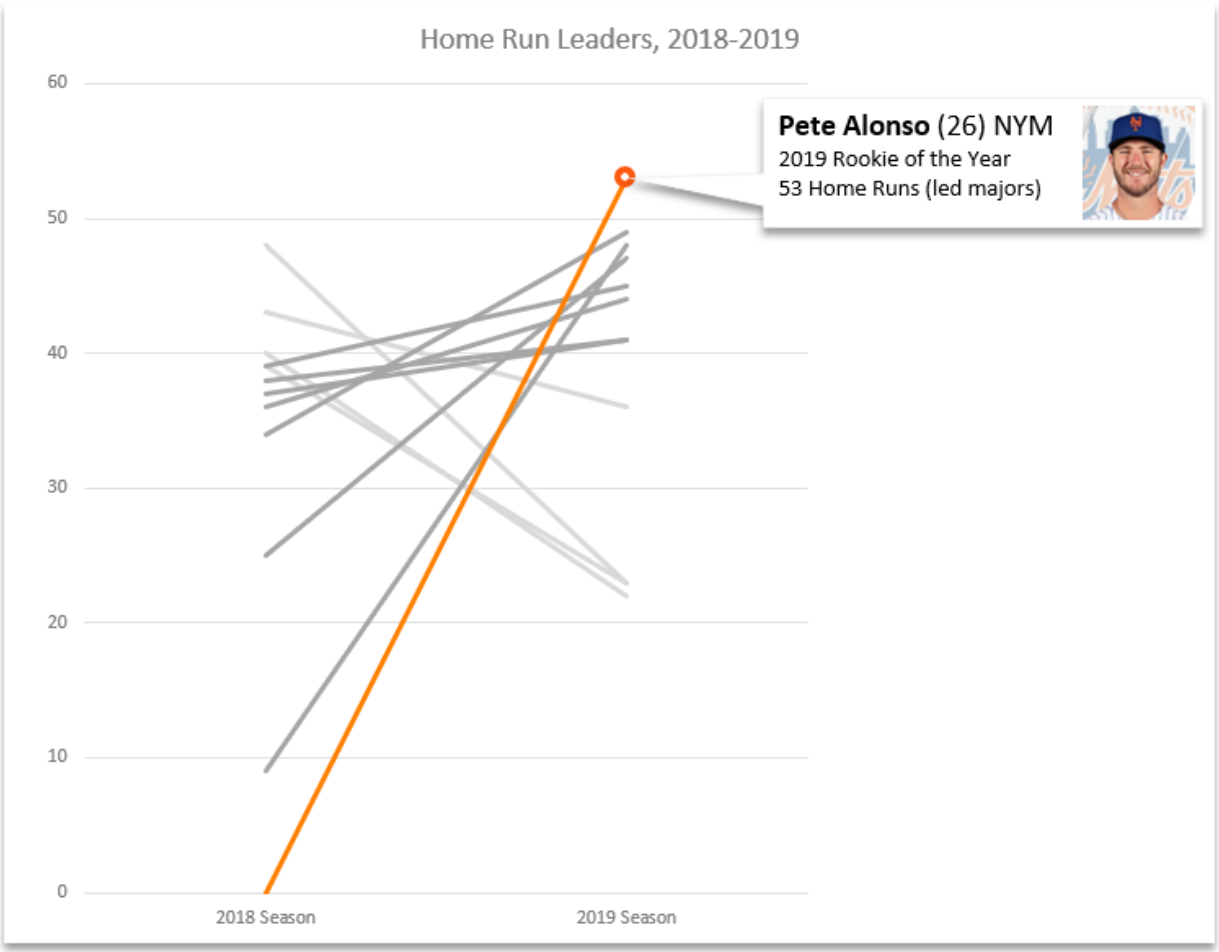
...and finally, use the setting for **Series overlap** set to **100%** and you will get this 'fill up' effect:



Slope Graphs

A Slope Graph is simply a Line Chart with two dates. Often, when visualizing two data points, chart creators will avoid a Line Chart and instead insert two Pie Charts, or put the values into a Column Chart. Notice how clearly this chart illustrates the change between 2018 and 2019:

Player	2018 Season	2019 Season
Joey Gallo	40	22
Khris Davis	48	23
Jose Ramirez	39	23
J.D. Martinez	43	36
Nelson Cruz	37	41
Nolan Arenado	38	41
Christian Yelich	36	44
Mike Trout	39	45
Cody Bellinger	25	47
Jorge Soler	9	48
Eugenio Suarez	34	49
Pete Alonso	0	53



Waffle Charts

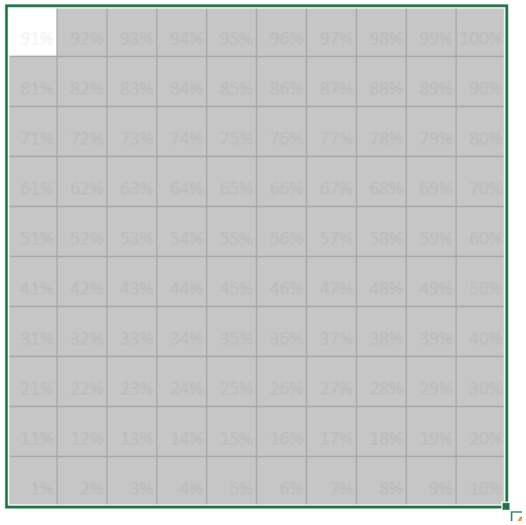
A Waffle Chart uses a 10x10 grid of boxes to illustrate percentages. Often, being able to read the number of boxes (each representing a percentage) makes the information clearer and more direct. A Waffle Chart is not a conventional Excel chart type, but rather a construction using standard cells and conditional formatting.

To create a Waffle Chart in Excel, first put percentages 1-100% into a 10x10 grid of cells, and resize the cells to be square:

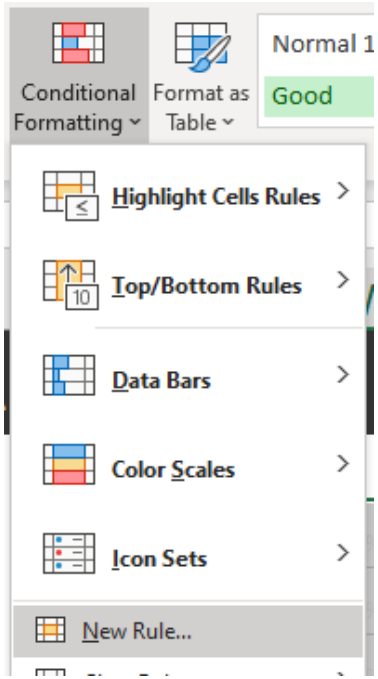
Before moving forward, make sure to have one or more cells with percentages to compare in the waffle chart:

49%
75%
91%
100%

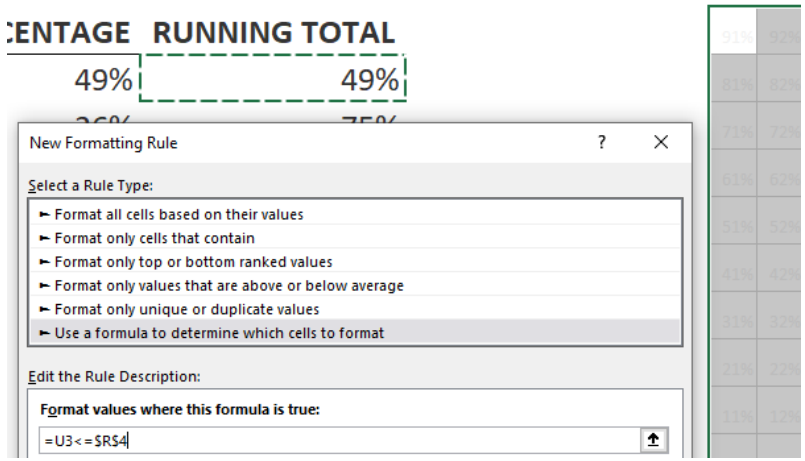
Next, select all the cells, starting with the top-left cell. If you've done this successfully, the top-left cell should be white, and the rest grey.



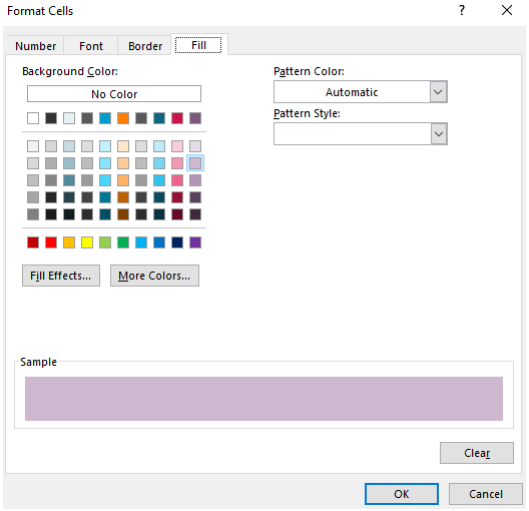
Now, click **Home Tab > Conditional Formatting > New Rule...**

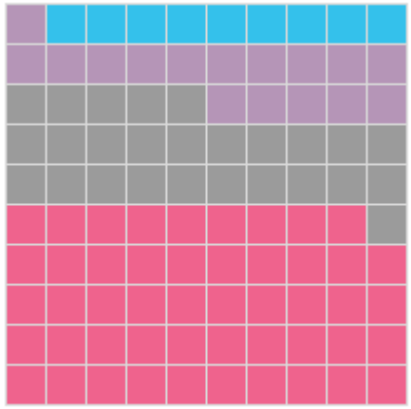


At this point, select the option for **Use a formula to determine which cells to format**, and enter a function that compares the values in this grid to the values you'd like to visualize. Make sure that you reference your top-left cell with **no dollar signs** to lock the reference – this will copy the same formatting across all the cells.



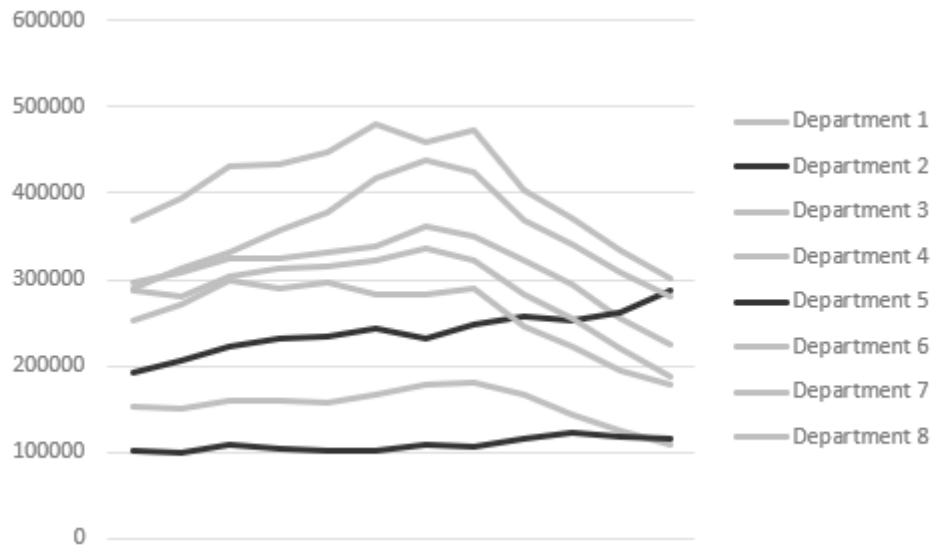
In this example, you are imposing a rule that judges **U3** and all the other cells in the grid against **49%** in cell **R4**. Click the **Format...** Button to choose the appropriate coloring for these cells. Continue with more rules until all the cells are filled appropriately.





Small Multiples

Small multiples are not a chart *type*; rather, it is an approach to visualization that considers whether the many, competing stories in your chart should be separated into many, smaller charts. Say, for example, you had a complex line chart like this:



Note what the chart creator is attempting to show – Dept 2 and Dept 5 are highlighted differently from the other departments, and we see that those two lines are trending *up* while the others are trending down.

Instead of illustrating all these values together in a single chart, we can plot individual series in separate charts. Then, by using grouping, labeling, color-coding, and other categorization techniques, you can describe the insight the viewer should be seeing.

Appendix

Chart Choosing Sites

- https://depictdatastudio.com/charts/
- https://multimedia.journalism.berkeley.edu/media/upload/tutorials/dataviz-basics/chartpicker.pdf
- https://stephanieevergreen.com/qualitative-chart-chooser-3/

Contrast, Color Blindness, and Accessibility

- https://contrastchecker.com
- https://webaim.org/resources/contrastchecker/
- https://coolors.co/contrast-checker
- http://colorsafe.co/
- https://www.color-blindness.com/coblis-color-blindness-simulator/

Design and Color

- https://www.amazon.com/Non-Designers-Design-Book-4th/dp/0133966151
- https://color.adobe.com