Interpreting visualizations

This lesson outlines how to read and interpret data visualizations.

Scope of data

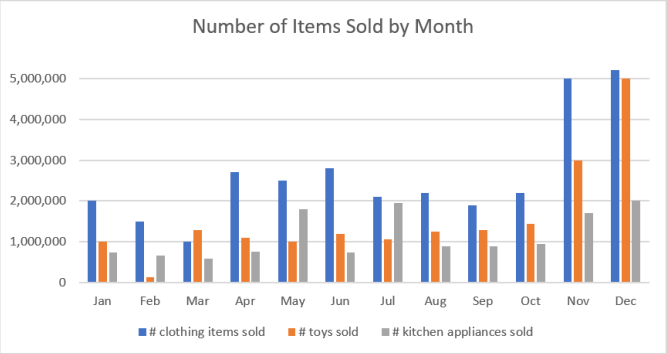
The extent that a dataset covers.

With the knowledge that you've gained in this course so far, you can now create impressive charts and powerful graphs. But all of this is useless if you don't know how to interpret these data visualizations and extract meaningful insights from them. This lesson outlines how you can read and interpret your visualizations, which is an essential first step before you go on to draw conclusions about the data and use your conclusions to identify important next steps.

Identifying the visualization type

When you're looking at a new visualization, the easiest thing to do first is identify the chart type. The chart type gives you a better idea about the information being presented, along with relationships in the data.

The visualization below uses a dataset containing a company's online shopping information for 2019.



This chart is a clustered bar chart. In general, bar charts are useful for comparing groups or categories. This chart compares the number of items sold across different months. Clustered bar charts add another visual comparison between additional categories. In this case, the additional categories being compared are clothing, toys, and kitchen appliances.

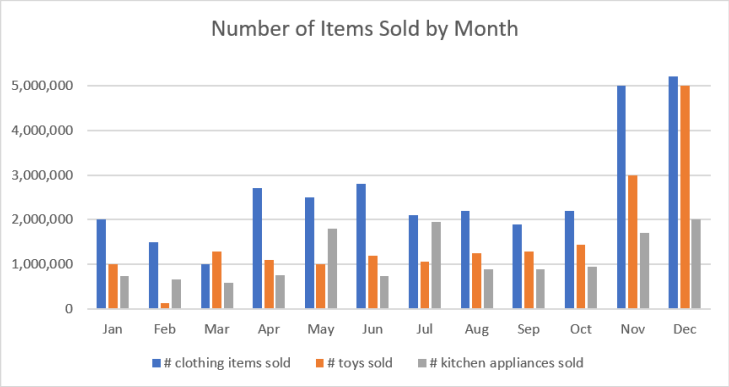
Identifying the scope of data

Context is important when reading a chart or graph. Before interpreting the visualization, first ask yourself what the scope of the data is. The scope of data is what the dataset covers. Just as you may identify the setting of a book, you can think about the scope of data as the "setting" of your dataset. Listed below are key questions to ask yourself when identifying the scope of your data.

- Region: Which geographic region does this dataset cover? For example, does it cover one small city or an entire country?

- Time period: What time period does this dataset cover? For example, does it cover a single day or multiple years?
- Size of dataset: Is the dataset large or small? For example, does it contain 40 records or 400,000 records?

In the clustered bar chart, shown again below, what can you gather about the scope of the data?



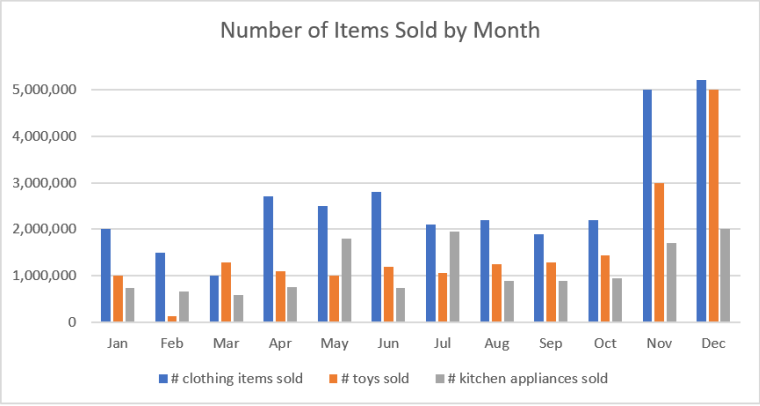
As you may have realized, the data used to create the chart covers the number of items sold each month of a given year.

Identifying the field types

Next, identify the field types presented in the visualization. The field types may be categorical, numeric, or even dates. Look at the field names and ask yourself what they represent. Do they represent a category of items? Do they represent numeric measurements? Do they represent dates, months, or years?

Similarly, identify the units used for each field. For example, if a field represents heights of individuals, are they measured in inches or centimeters? Understanding field types and units of measurement is important for correctly interpreting the chart.

In the clustered bar chart, shown again below, what can you identify about the fields?

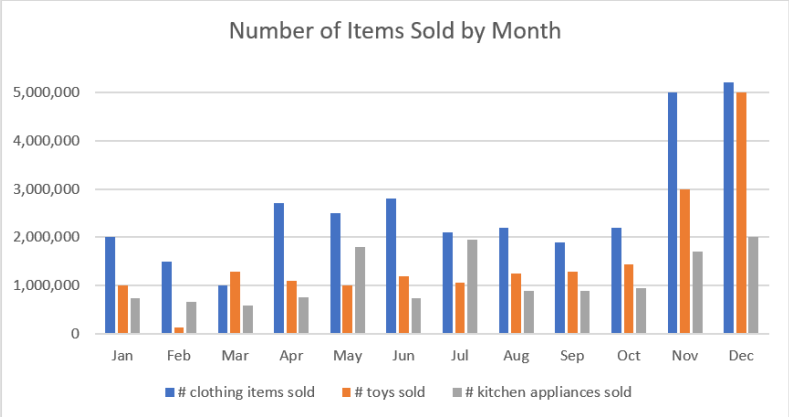


The field on the horizontal axis is categorical, showing months as well as clothing, toys, and kitchen appliances as the additional categories. The field on the vertical axis is numeric, showing the number of items.

Identifying relationships

Finally, identify any relationships that the chart is illustrating. When coming up with relationships, ask yourself what is being measured and how.

Does the chart have horizontal and vertical axes? If so, take note of which field is on the horizontal axis and which field is on the vertical axis. This tells you which fields are being compared and how. Or, if the chart has no axes, such as a pie chart, then how are the fields or groups arranged? Use your knowledge of the different chart types to help you pinpoint what is being measured and how.



This chart measures the number of items sold for each month of a given year. Items sold are shown separately for clothing, toys, and kitchen appliances.

Drawing conclusions

In this lesson, you will learn how to draw meaningful conclusions from visualizations.

Inference

A conclusion that is reached based on evidence and reasoning.

Association

A connection or relationship.

Causation

A cause-and-effect relationship.

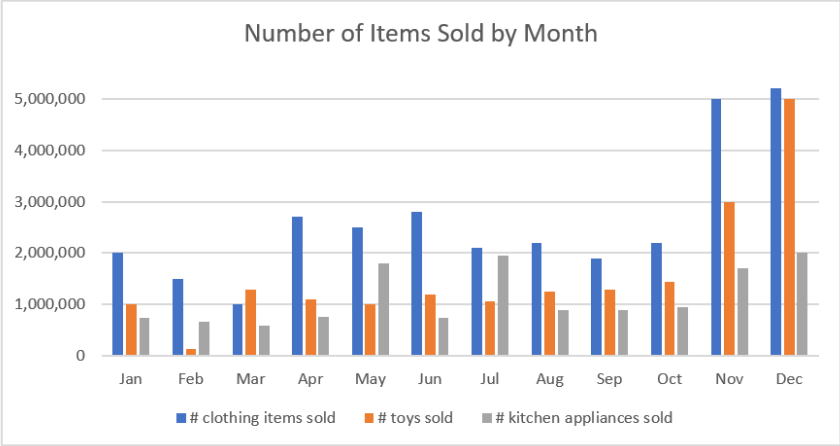
Identifying patterns, trends, and outliers

After reading and orienting yourself to a visualization, look for patterns, trends, and outliers.

Patterns and trends highlight important behaviors.

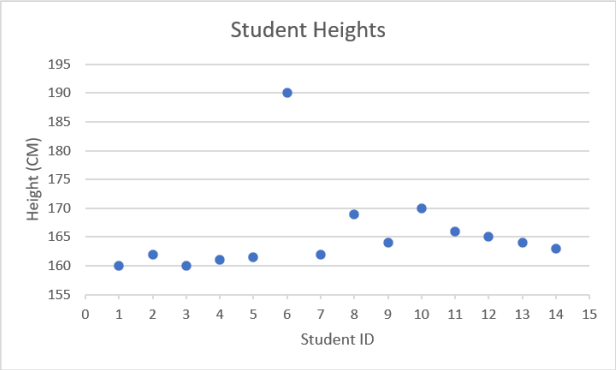
- Patterns to look for include **repetition** in graphs, as well as areas where a graph **plateaus**, or **flattens**.
- Repetition indicates a **cyclical pattern**, which can tell you **what to expect in the future**.
- Similarly, **plateaus** in graphs tell you **when things stop changing**, which can also be useful in understanding behavior.
- **Trends** to look for include **spikes** or **dips** in graphs, which indicate a **notable change** in behavior or differences between the categories shown.

The chart from the previous lesson is shown again below. Recall that this clustered bar chart shows the number of items sold for each month of 2019. Now, look at the chart again and see if you can identify any patterns or trends.



The standout trend in this visualization is that the number of items sold dramatically increased in November and December. Also, clothing is the category that sold the most in November and December, followed by toys and then kitchen appliances. Finally, this company sold more clothing items than toys or kitchen appliances, across all months.

Outliers are points of data that are far from the rest of the data. These can be problematic; they can make you mistakenly think that there's a large jump or dip in the data, even when it's simply a data point that occurred by chance or isn't representative of the rest of the data. For example, the graph below shows student heights for 16-year-old females. There's one outlier, student #6, who's much taller than the rest of the students. You won't want to factor this height into a calculation for an average height, because it will make the overall average height much larger than the real average height of 16-year-old females. So, when you see a clear outlier, simply ignore it.



Making inferences

After observing how the data behaves in a visualization, you can apply those observations to make inferences. An inference is a conclusion that is reached based on evidence (the observed patterns and trends) and reasoning.

Suppose there are three friends named Saira, Ariel, and Maria. You're given the following facts about them:

- Saira is taller than Ariel.
- Maria is taller than Saira.

From these two facts, what inference can be made? In other words, what else can you gather about the relative heights of the three friends?

Because Saira is taller than Ariel, but Maria is even taller than Saira, it can be inferred that Maria must also be taller than Ariel. You're able to draw this conclusion based on the evidence (the two initial statements given) and reasoning. In this same way, you can use evidence and reasoning to make inferences based on visualizations.

Going back to the clustered bar chart above, what inference can you make based on the trends that you've identified so far? Because the trend of items sold clearly increases in November and December, for all categories, one reasonable inference you can make is that there's some seasonality to consumer shopping behavior.

Identifying causation versus association

Be careful about making cause-and-effect assumptions, though! It is important to distinguish between causation and association. Association means that there is some relationship between things, but causation means that one directly causes another.

From the bar chart shown at the beginning of this lesson, you can say that the number of items sold and the holiday season are associated, but you can't necessarily claim that the holidays caused more items to be sold. This is because there may be other factors involved. For example, perhaps this increase is because people shop online more during colder months because they're indoors more. Or maybe the increase is a response to an aggressive marketing approach during November and December. Look for trends, but don't automatically assume causation.

Finally, the amount of data that you have is crucial to how confident you can be in your inferences. This is why the previous lesson stressed the importance of the scope of data. For example, you can make an inference based on one year of data. But if you instead have data for the past 10 years, you would be much more confident in the patterns and trends that you see. If the one year of data in the chart above covers a year when consumer behavior is wildly different from other years (like many companies experienced in 2020 due to COVID-19), then any inferences you make may only be specific to that one year. They wouldn't apply to general consumer behavior. In short, the more data you have, the more accurate your inferences are.

Recommending next steps

Now that you know how to interpret visualizations and draw conclusions from them, this lesson will show you how to come up with useful recommendations or next steps.

The ultimate goal of any analysis—including data visualization—is to make smarter, data-driven decisions. To put it another way, the goal is to recommend next steps based on your conclusions. **Charts and graphs may convey a lot of useful information, but it's what you do with that information that's important.** If you can take what you learn and use it to increase productivity or identify weak areas of a business, then your analysis becomes a powerful tool.

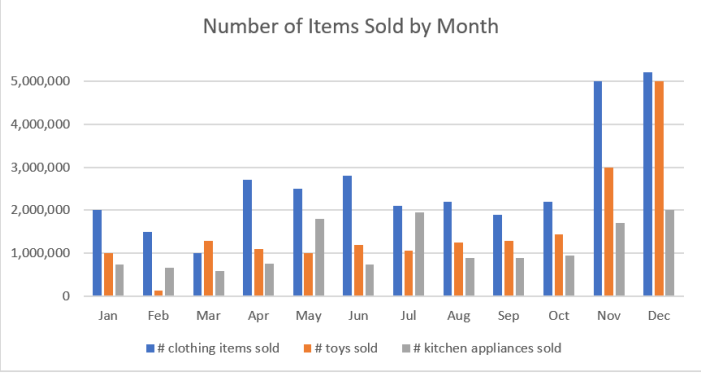
At work, you often need to be able to identify and share recommendations with others at your company. No matter what industry or career you're in, your employer will likely expect you to suggest ways that you, your team, or others at your company can make improvements and identify weaknesses in your work. You have a big advantage if you can base your recommendations on objective data rather than a personal hunch. And data visualizations can help.

Identifying the problem

Whenever you turn data into a data visualization, there's usually a reason behind it. If you collect and track data and then turn it into a line graph, it's likely because you're on a mission to identify or solve a problem. Identifying a potential problem can help you figure out which steps you should take next.

Think of charts and graphs as tools that help you identify a potential problem to be resolved or to investigate further. Whenever you encounter a visualization, ask yourself this: why was this data collected?

Recall the clustered bar chart from the previous two lessons, shown again below. Why was this data collected?



The data was probably collected to understand and improve the number of sales made by the company. It may bring clarity to online consumer shopping behavior. Similarly, this data may help identify weaknesses in the company's marketing efforts throughout the year.

Be creative and try to think of a few possible reasons why a dataset may have been collected. This process helps you use your insights effectively.

Coming up with recommendations

Once you identify a reasonable problem, what lessons, recommendations, or warnings can you extract from your observations? Remember that providing a thorough analysis includes being able to recommend next steps.

For the chart illustrating online shopping behavior, what lessons, recommendations, or warnings can you think of?

One lesson taught by the visualization is that there is some seasonality to the number of items sold for this company. One recommendation is to now analyze the company's marketing expenses and efforts to figure out if greater marketing resources should be used at certain times of the year. Another recommendation is to investigate if there's a difference in the behavior of online sales versus in-person sales.