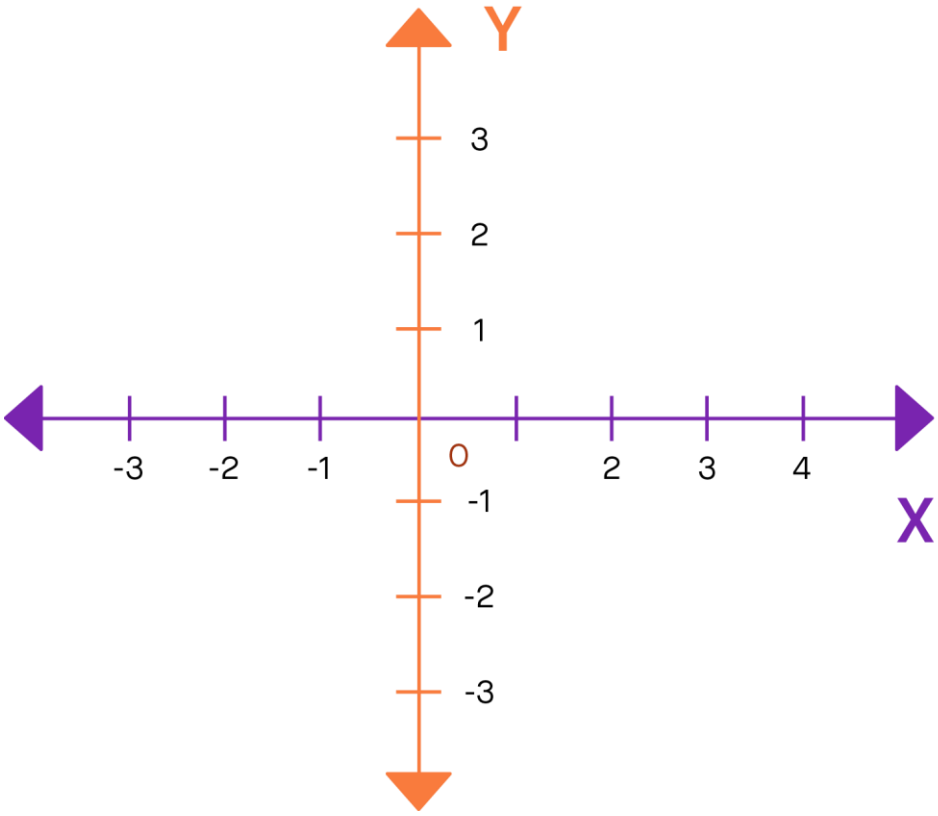
The coordinate plane

Before diving into the types of data visualizations, it's important to note that many visualizations use a coordinate plane as a foundation. A coordinate plane is a two-dimensional surface that you can use to plot points, lines, and curves. The coordinate plane is an important concept for data visualizations because it allows you to translate data into a graphical representation.

A coordinate plane has two axes: the x-axis and the y-axis. The x-axis is the horizontal axis, while the y-axis is the vertical axis, as shown below.



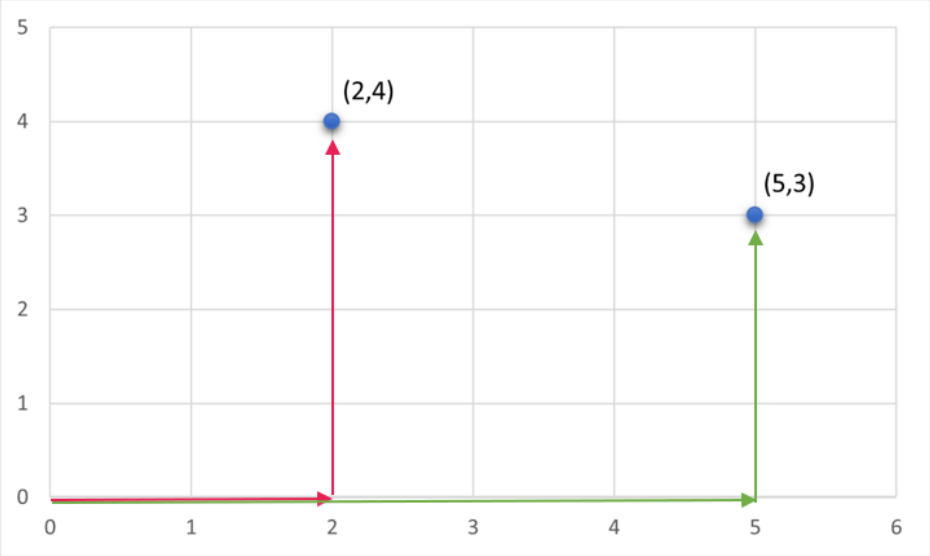
The center of the coordinate plane is where the x-axis and the y-axis intersect, or cross. At this point, the values for X and Y are zero. The x-axis has positive values to the right of the zero and negative values to the left of the zero. Similarly, starting from zero, the y-axis has positive values going upward and negative values going downward.

Points on a graph

Every point on a graph has its own "address," or coordinate pair, which is a combination of exactly one X coordinate and one Y coordinate, in the form (X,Y). To plot a point, use the following process:

1. Start at the center where the coordinates are (0,0).
2. Move the number of spaces specified by the X coordinate. For positive values, move to the right. For negative values, move to the left.
3. Move the number of spaces specified by the Y coordinate. Move upward for positive values and downward for negative values.
4. Draw the point at this specific (X,Y) location.

Below is an example showing two plotted points: one at (2,4) and one at (5,3).

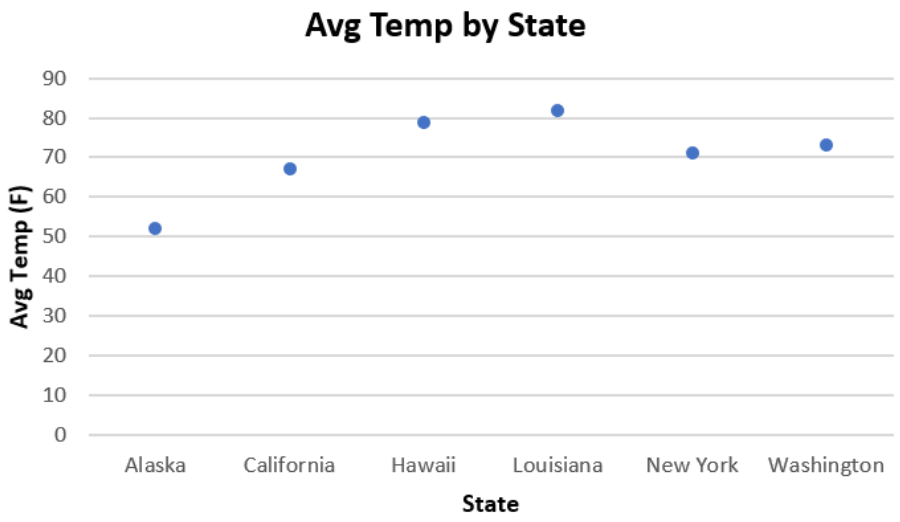


Now that you know how to plot a point, you're in a good position to be able to interpret graphs in later lessons.

Independent and dependent variables

When creating data visualizations, you generally have both independent and dependent variables. Independent variables, which are fixed, are represented on one axis. Dependent values, which can change, are represented on the other axis.

To see what this looks like, check out the visualization below. States are on the x-axis and represent the independent variables, because states are stable from year to year. Average temperatures are on the y-axis and represent the dependent variables, because temperatures change from year to year.


Avg Temp by State
90
80
70
60
50
40
30
20
10
0
Avg Temp (F)
Alaska
California
Hawaii
Louisiana
New York
Washington
State

Bar charts

This lesson introduces vertical and horizontal bar charts, stacked and clustered bar charts, and waterfall charts. It also explores when to use each one.

Categorical data

Also called qualitative data, data that can be divided into groups or categories.

Bar chart

A chart that shows categorical data with rectangular bars whose heights (or lengths) represent numerical values.

Introduction

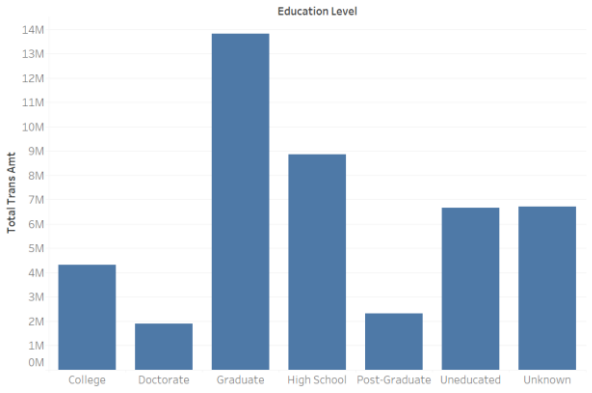
One of the most common data visualizations is the bar chart; bar charts are a great way to organize categorical data in a diagram format. Categorical data is data that can be divided into groups or categories. It is qualitative data that contains characteristics rather than numeric values.

A bar chart shows categorical data by using bars that represent numerical values. Bar charts are popular because they have picture-like qualities that make complex information simpler to interpret. As you'll explore in this lesson, there are several types of bar charts, and each shows a slightly different view of the data.

Vertical bar charts

Vertical bar charts show categorical data in groups along the x-axis. This type of chart uses bars rising vertically to represent the numerical values of each group. The heights of the bars correspond to the numerical values that they represent.

In the example below, education level categories are shown along the x-axis, and the total transaction amounts are shown along the y-axis. The heights of the bars correspond to the transaction amounts for each education level.

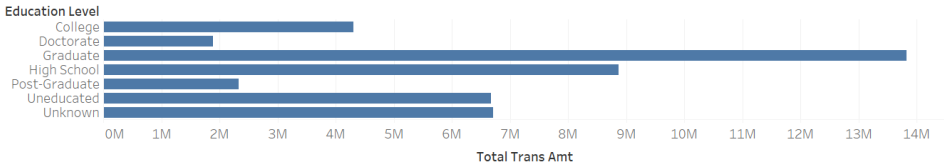


Vertical bar charts are used to compare several groups and their relative amounts, side by side. They are used when there's one qualitative (categorical) field and one quantitative field.

Horizontal bar charts

Horizontal bar charts show the same information as a vertical bar chart, except that the bars lay horizontally across rather than vertically up and down. More specifically, a horizontal bar chart has the categorical data in groups along the y-axis. It uses bars lying horizontally to represent the numerical values of each group. The lengths of the bars correspond to the amounts that they represent.

In the example chart below, the education level categories are shown on the y-axis, and the total transaction amounts are shown on the x-axis. Each education group has a bar whose length corresponds to the transaction amount.



Like vertical bar charts, horizontal bar charts are used to compare several groups and their relative amounts. They are used when there's one qualitative (categorical) field and one quantitative field. Horizontal bar charts are preferred over vertical bar charts when category labels are especially long and won't fit along the bottom of the chart. Some people also prefer horizontal bar charts when there are a lot of categories, because it's easier to scroll up and down to see them than it is to scroll left to right.

There are no strict rules for when to use vertical or horizontal bar charts, but the main goal is to create a visualization that is easy for your audience to follow and interpret.

For consistency, this course uses the term bar chart for all visualizations that use bars to represent values. However, in the real world, you may see some variation in terminology; horizontal bar charts and vertical bar charts are sometimes called bar charts and column charts, respectively.

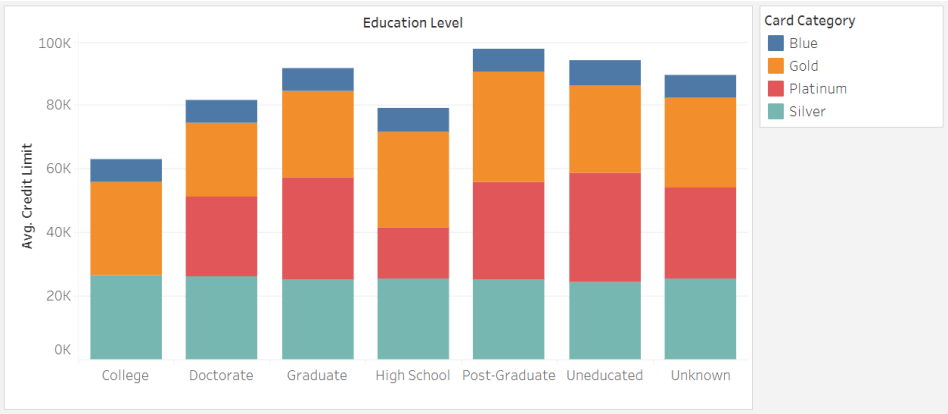
Stacked bar charts

A stacked bar chart shows categorical data in groups, with bars whose heights (or lengths) represent the numeric values of each group. But it also shows an additional category split up into subgroups, and it indicates how these subgroups make up the total values.

For example, the stacked bar chart below shows the average credit limit by education level, but also by card category within each education level. To interpret the credit limits for the education levels, look at the total height of each bar. For example, post-graduates have a higher average credit limit compared to other education levels; you can tell that because the entire bar for post-graduates is higher than any other bar. And within the post-graduate bar, you can see that most are in the gold card category, followed by platinum, followed by silver, followed by blue.

Then, to interpret the credit limits for card categories, look at the height of each of the colored stacks. You can tell that gold and platinum cardholders have the highest average credit limit because, for most

education levels, the orange and red stacks have taller heights than the blue or green stacks. The order of the colored stacks doesn't matter—the fact that blue is on top doesn't mean anything special about blue cards. What you can tell, though, it that the blue stacks are thinner than any other color—which indicates that this category has the lowest average credit limit.



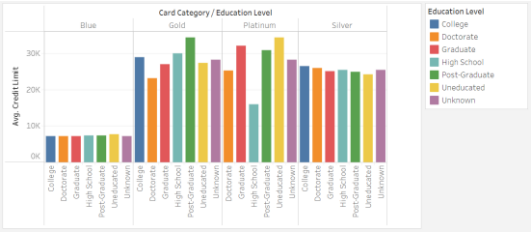
As you can see, stacked bar charts are used to show comparisons between groups, just like regular bar charts. But unlike regular bar charts, they also show how an additional category contributes to the totals.

Stacked bar charts are usually displayed vertically, but they can be displayed horizontally too. Select the orientation based on how you want to show your data.

Clustered bar charts

A clustered bar chart shows the same information as a stacked bar chart, except that the additional categorical variable data is laid out side by side, clustered rather than stacked. Also, clustered bar charts are read like regular bar charts, focusing on only the heights of the bars.

For example, the chart below shows the average credit limit by education level, and it also shows how average credit limits compare between the different card categories. Gold and platinum cardholders have the highest average credit limits, while blue cardholders have the lowest. Also, within the gold card category, post-graduates have the highest average credit limit.



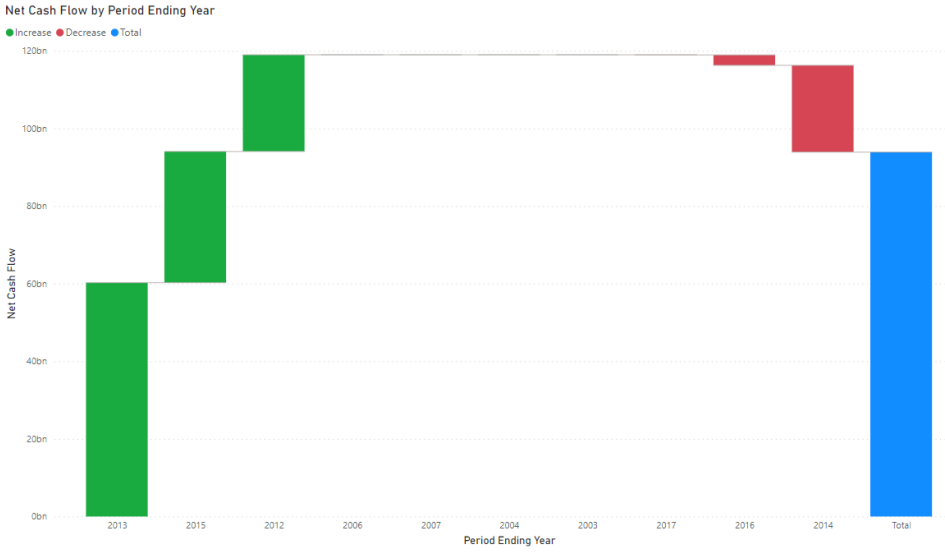
Like stacked bar charts, clustered bar charts are used to show comparisons between groups. However, clustered bar charts also show comparisons between the subgroups of the additional categorical field.

Clustered bar charts are typically displayed as vertical charts, but they can be displayed horizontally too. Select the orientation based on how you want to show your data.

Waterfall charts

A waterfall chart is a special type of bar chart that shows the cumulative effect of a sequence of values. In this type of chart, a positive value is shown as a rise and a negative value is shown as a drop, mimicking the look of a waterfall. Waterfall charts are useful because they present each step leading up to a cumulative total amount.

The waterfall chart below shows net cash flows by year, with positive values indicated in green and negative values in red. When you look at this chart, you can immediately tell that the cash flows were positive until 2016, and then they started to decrease. Both the positive and negative cash flows make up the cumulative cash-flow amount of $95 billion, shown by the blue bar.



So, waterfall charts are particularly good at explaining how numeric values develop, because they show how increases and decreases contribute to an overall total.

Scatter plots and graphs

This lesson explains when to use scatter plots, line graphs, and area graphs.

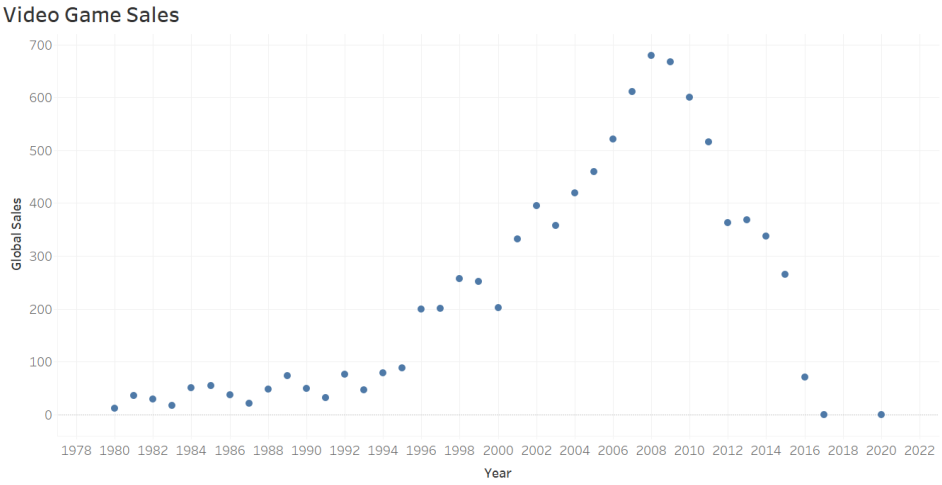
There are several visualizations that are created by plotting points on a graph: scatter plots, line graphs, and area graphs. These visualizations can help you identify trends.

Scatter plots

A scatter plot is the most basic graph. It's created by plotting points on an XY coordinate system, where one variable (or field) is on the x-axis and the other variable (or field) is on the y-axis.

Scatter plot: A graph showing plotted points on an XY coordinate system

The scatter plot below shows Global Sales on the y-axis and Year on the x-axis. Each point on the graph is a combination of one Year and one Global Sales amount.



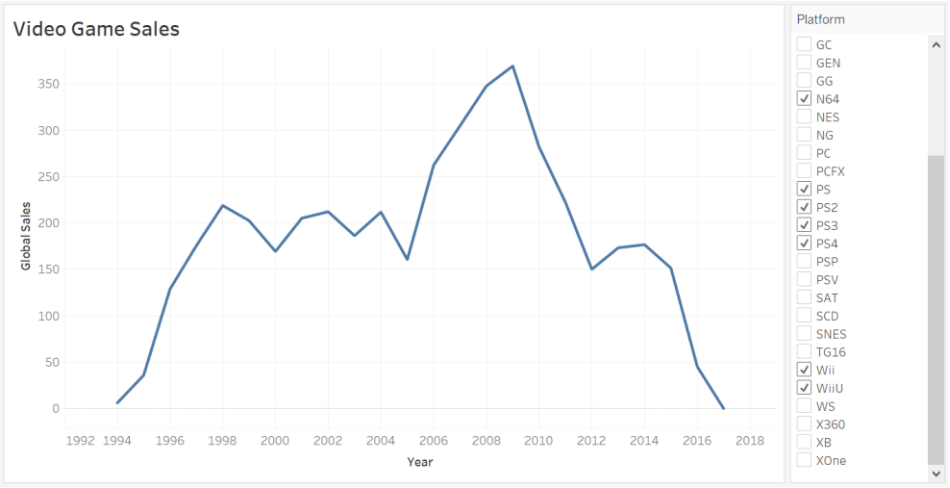
Scatter plots are used to see the general behavior of the data. They make it easier to investigate potential trends because patterns can be seen in the plot. For example, the scatter plot above shows that video game sales started to increase dramatically in 1996 and peaked in 2008.

Line graphs

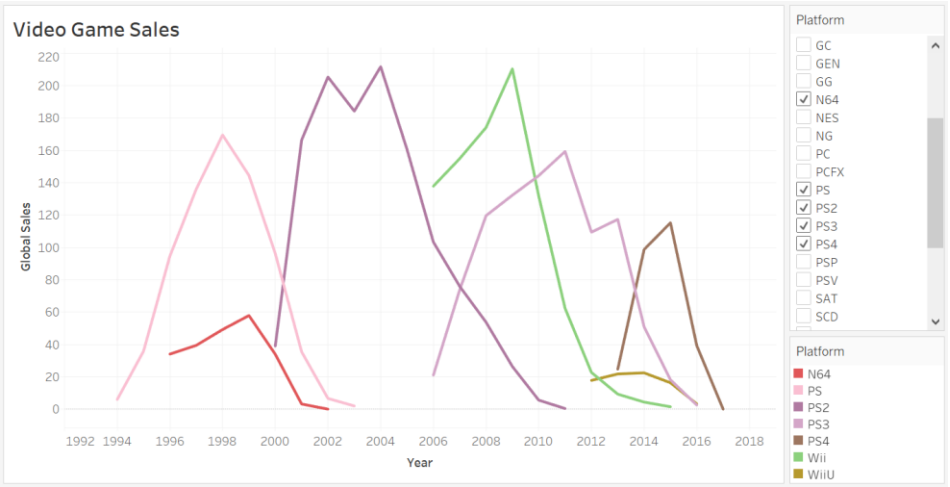
A line graph shows plotted points on an XY coordinate system, which are then connected by straight lines. In other words, a line graph is just a scatter plot with a straight line connecting the dots. The use of lines makes it easier to see the shape of the graph and identify patterns.

Line graph: A graph showing plotted points on an XY coordinate system, connected by straight lines

Line graphs are typically used to show how things trend over time. The line graph below shows how video game sales trended over time for the platforms N64, PS, PS2, PS3, PS4, Wii, and WiiU.



Stacked line graphs show multiple line graphs together in the same visualization. Stacked line graphs are useful when you want to compare trends over time for individual groups. The example below shows the same information from the line graph above, except the data is split out by video game platform. Each colored line represents the global sales over time for a different platform.

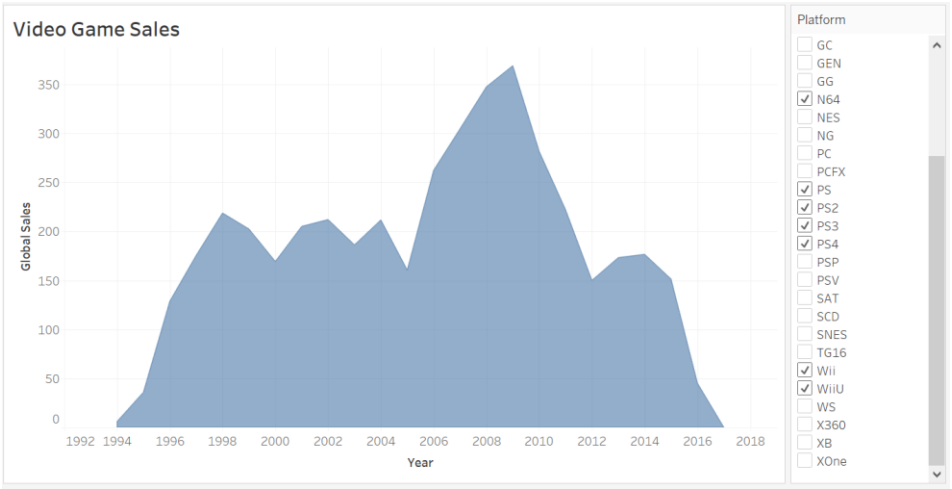


Area graphs

An area graph is the same thing as a line graph, except that the area between the x-axis and the lines is filled in.

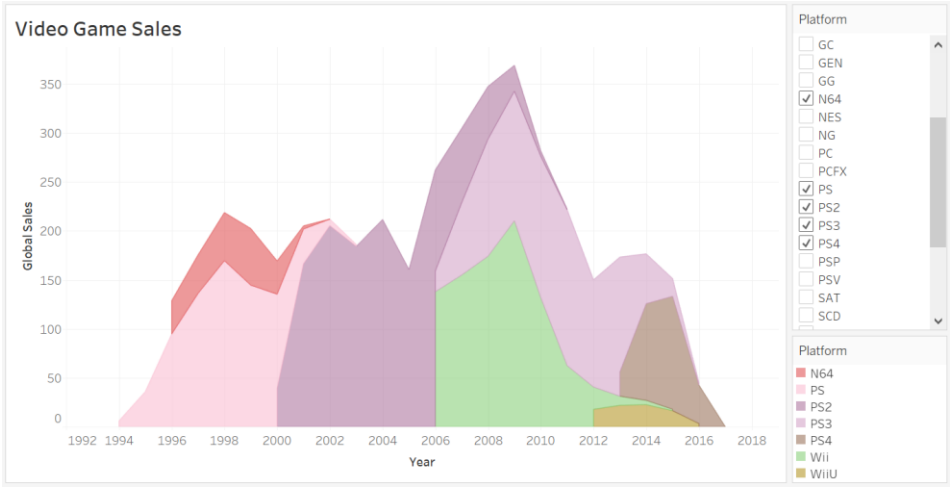
Area graph: A line graph in which the space between the lines and the x-axis is filled in

The example below is an area graph that shows global sales over time for N64, PS, PS2, PS3, PS4, Wii, and Wii U. Notice that this graph has the same shape as the single line graph from above. A single area chart shows the same information as a single line graph, and it can be used in the same scenarios.



A stacked area graph is a useful way to compare parts of a whole. While a stacked line graph is used to see how individual groups trend over time, a stacked area graph shows how individual groups contribute to a total. The filled-in areas show how each group contributes to the total.

For example, the stacked area graph below shows the same information as the earlier stacked line graph. The difference is that these filled-in areas give you a better idea of how much of the total sales each platform makes up.



Conclusion

This lesson introduced scatter plots, line graphs, and area graphs. It also described how each of these visualizations is used.

Pie and doughnut charts, heat maps, and tables

This lesson introduces pie and doughnut charts, heat maps, and tables. It also explains when to use each one.

Proportion

The relation of a part to another part or to the whole.

Introduction

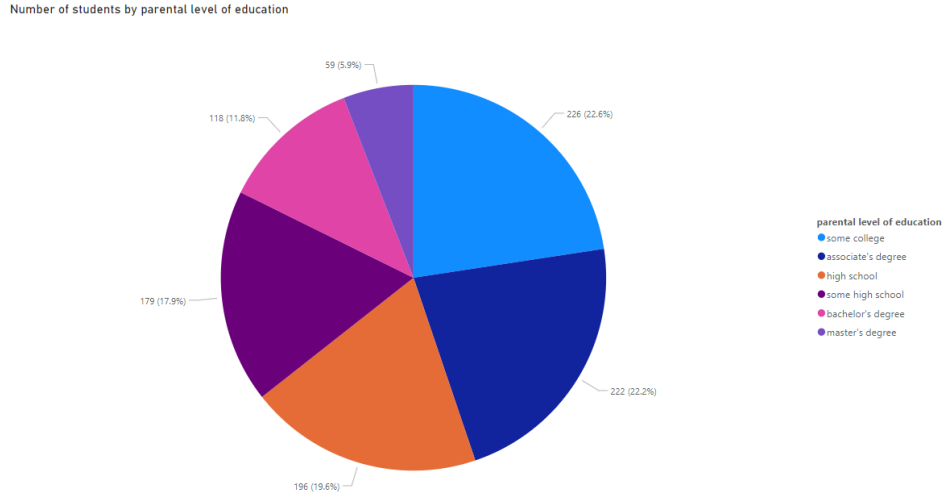
There are several visualizations that don't require plotting points on a coordinate plane but are still effective at summarizing data. You'll focus on those visualizations in this lesson.

Pie charts

A pie chart is a circular chart that resembles a pie. Each slice, or section, of the pie represents a portion of the whole, and its size represents the proportion of the whole that it makes up. So, a pie chart is most commonly used to show how smaller components make up a larger group.

Pie chart: A circular chart that resembles a pie divided into slices, where each slice's size represents a proportion of the whole.

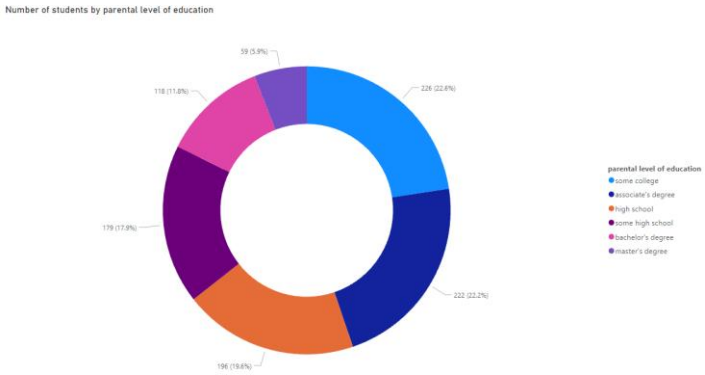
For example, the pie chart below shows the number of students in each parental education group. Showing this breakdown is more useful to the viewer than just seeing the total number of students because this gives additional information about the students.



Pie charts are valuable ways to add context to any analysis. They're also quick to read and easy to interpret. If the pie chart above didn't have any numbers on it, you'd still be able to tell that most students come from parents who have some college, an associate degree, or a high school diploma, because these sections are the largest.

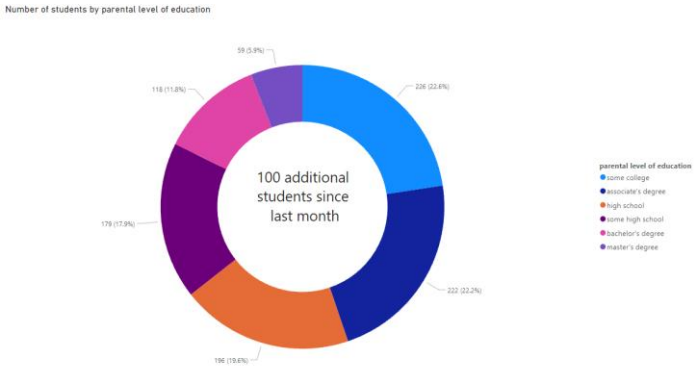
Doughnut charts

A doughnut chart is the same thing as a pie chart, but with the center removed so that it resembles a doughnut. The doughnut chart below shows the exact same information as the previous pie chart; the only difference is that the middle of the pie is cut out.



Although the information shown is the same, doughnut charts are preferred over pie charts when there are lots of slices, or sections, to show. This is because pie charts that have too many slices can quickly become distorted or hard to read in the center of the pie. Doughnut charts tend to look less cluttered.

Doughnut charts are also preferred over pie charts when a key takeaway needs to be shared with the audience. The empty space in the center of the doughnut chart draws the viewer's eyes in, so it's a good place to highlight information. This can be seen in the doughnut chart below. The information in the colored sections is the same as in the previous doughnut chart, but now the center of the doughnut shows a key takeaway: 100 additional students have joined since the previous month. The viewer sees this first, and then they can look at the rest of the doughnut to see the breakdown of students in each category.



Heat maps

A heat map uses variation in color to show relative amounts. For example, a heat map may use darker shades of green to represent larger values and lighter shades of green to represent smaller values. Heat maps are used to show a snapshot of categories and their relative values.

Heat map: A type of chart that indicates a variable's magnitude by color variation such as hue or intensity.

The heat map below shows the average math score by parental level of education. The legend to the right of the heat map indicates that shades of blue are used to represent the relative average math scores. The darker the blue, the higher the average math score. This heat map reveals that students whose parents have a bachelor's degree or master's degree scored the highest, because these sections have the darkest shade of blue. As you can see, even though the heat map doesn't contain specific values, you can still quickly get a sense of the data. So, heat maps are used to give the viewer a general sense of the data, without the need for additional details or numbers.

Tables

Tables are the most basic type of visualization because they show actual data or summarized data in an organized diagram. Tables are structured using columns and rows, and this structure can help with keeping track of what you're looking at. Tables can be small or large and can hold actual data values or calculated values. The difference between tables and other visualizations is that tables contain only numbers or text; they have no additional illustrations. Consider using a table when you have lots of numbers or text that you want to present in a simple, organized way.

The table below shows average math scores for each parental level of education. The overall information is similar to what's shown in the heat map above, with one major exception: this table shows actual average math scores for each group, while the heat map only shows relative math scores between the groups. So, although tables can be less interesting to look at, they include more detail and actual values than other data visualizations.

Parental Level of Education	Average math score
associate's degree	67.88
bachelor's degree	69.39
high school	62.14
master's degree	69.75
some college	67.13
some high school	63.50
Grand Total	**66.09**

From the table, you can clearly see that students whose parents finished their master's degrees have the highest average math score, while students whose parents only finished high school received the lowest average math score.

Conclusion

This lesson introduces pie and doughnut charts, heat maps, and tables.

Dashboards

This lesson introduces dashboards and discusses best practices for creating them.

Introduction

As you now know, there are many data visualizations to choose from. In fact, this course focuses on only the most popular visualizations, but there are even more out there. Although one visualization is able to convey important information about your data, sometimes you need multiple visualizations to tell the full story. A dashboard is a useful way to show multiple visualizations together.

Dashboard: A visual display of text and graphics that summarizes multiple key findings, arranged on a single sheet, slide, or computer screen to tell a story.

Dashboards may display visualizations that include the same data in multiple different chart types. For example, a dashboard may have one visualization showing a line graph of product sales over time and another visualization showing a bar chart of the number of each type of product sold.

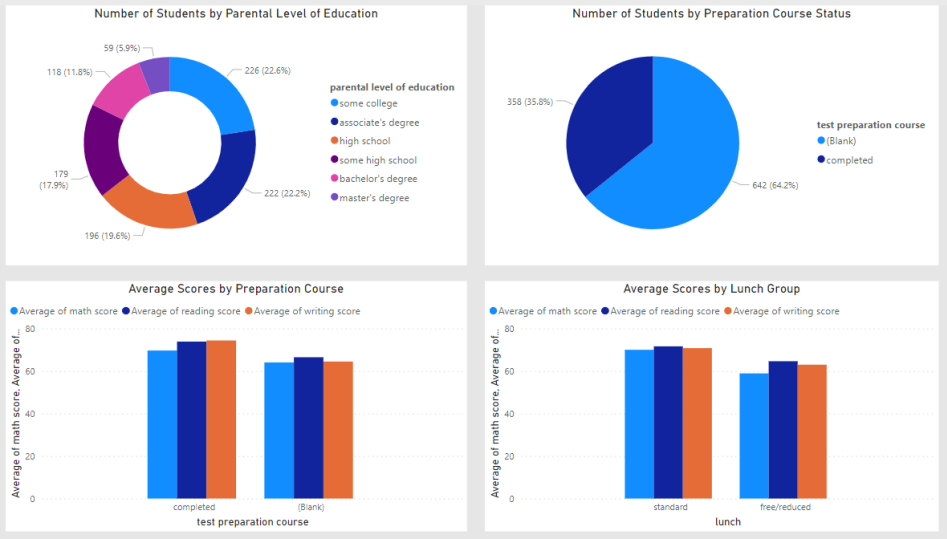
Dashboards can also be used to present different pieces of information on a single topic. For example, say that you want to create a dashboard about diversity and inclusion at your company. You could include a visualization of the demographics of your employees, a visualization of the demographics of your customers, a visualization of employee survey results of satisfaction over time, and a visualization of how you're progressing towards different diversity and inclusion initiatives at your company.

Commented [NA1]: Imagine keeping track of survey scores with holidays, policy implementation, etc.

Sometimes, you may create and use a dashboard only once. Other times, you may save a dashboard on a server so that it continuously shows up-to-date values, and you can return to it regularly to track your progress. Whatever the use may be, the goal of a dashboard is to tell some kind of story using engaging, clear, and organized visualizations.

Example of an effective dashboard

The following is an example of a four-visualization dashboard showing a summary of student performance and demographics.



From this dashboard, the viewer learns about the student population in this dataset, as well as how different factors affect student performance. For example, the top two visualizations tell you that most students have parents whose education levels are either an associate's degree, some college, or high school, and that only one-third of students completed a test preparation course. The bottom two visualizations tell you that students who completed the test preparation course scored higher than students who didn't, and that students with standard lunch scored higher than students on free/reduced lunch.

All four visualizations show different information, but together, they give you an understanding of the student population and the factors that may contribute to scoring high on exams.

Dashboard best practices

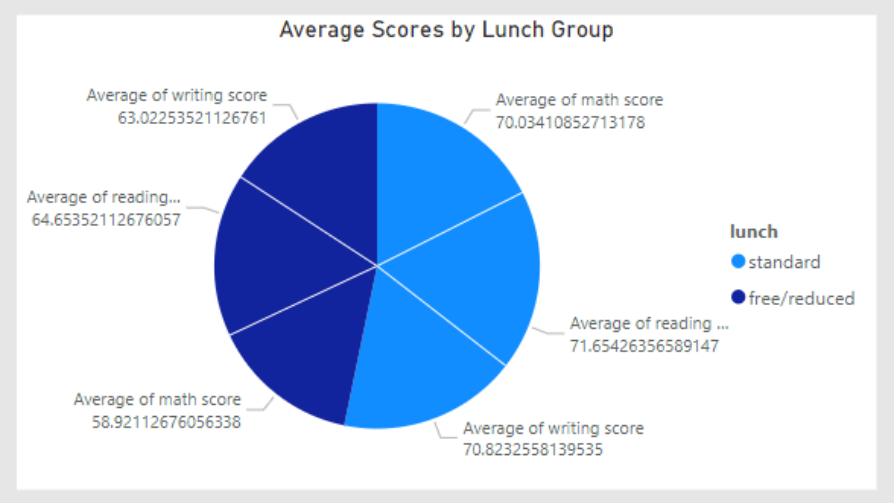
There are no strict rules for creating dashboards, but there are best practices that can help you create a powerful dashboard. You'll explore these next.

Choose appropriate visualizations

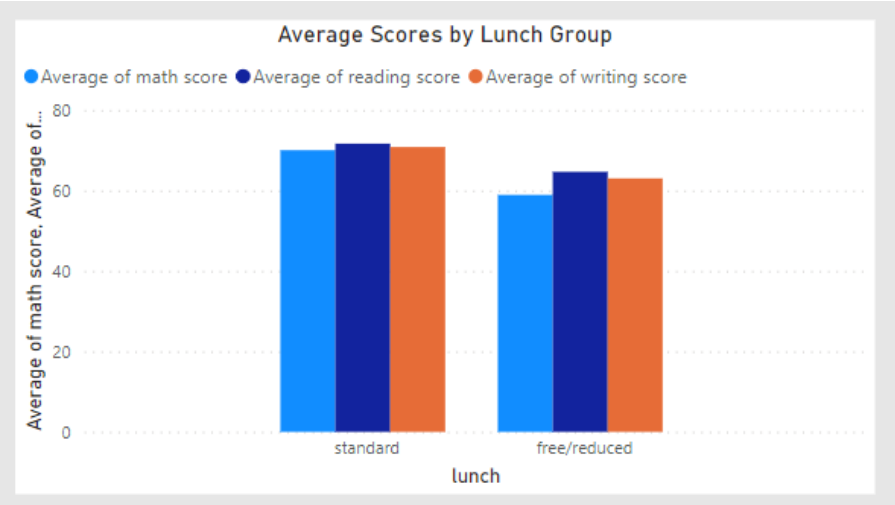
If you create 10 data visualizations during an analysis, that doesn't mean that you should use all 10 in your dashboard. **Dashboard visualizations should highlight valuable information and work together to tell a story.** You can include as many or as few visualizations as you want, but generally aim for 4-7 visualizations per dashboard.

Recall that each visualization type has its own strengths and weaknesses. Use what you've learned about each chart type to select the most appropriate one for what you want to show. For example, a bar chart is particularly useful for showing comparisons between groups, while a pie chart is better at summarizing characteristics in data. So, if you want to show a comparison, opt for a bar chart rather than a pie chart.

You can see this difference below. The pie chart below tries to compare the average scores for each subject, by lunch group. It's challenging to extract information from this visualization.



Now, look at the bar graph below. It contains the same information but is much easier to read. Visualizations shouldn't require the viewer to do much work to understand the information being presented.



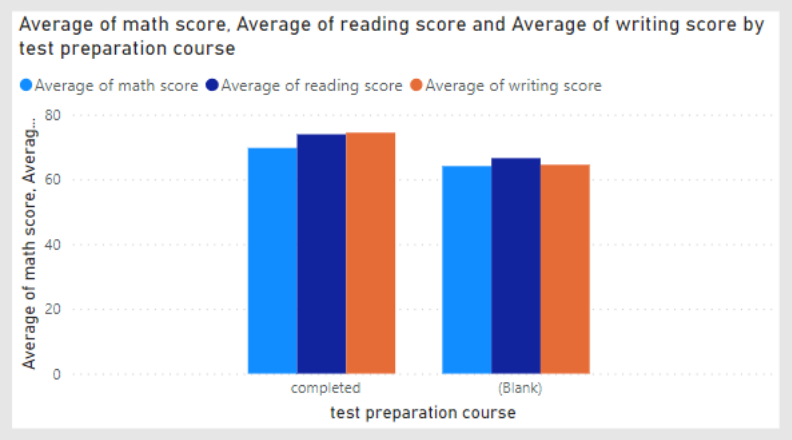
When building a dashboard, keep in mind the story that you're trying to share with your audience. Don't include a visualization just because you can; adding unnecessary visualizations that don't add anything

valuable is a waste of space and may confuse your audience. Select visualizations that help you deliver your message clearly.

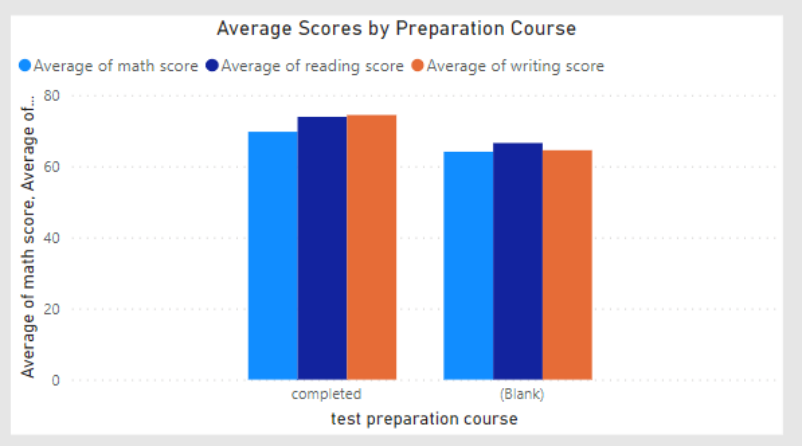
Include titles

Because a dashboard contains multiple visualizations, it's important to include titles for each visualization. Titles should be descriptive without being too long; you don't want to clutter the space with unnecessary text. Always include a title for every visualization that you create, regardless of whether you're including it in a dashboard or presenting it individually.

The first visualization below shows a title that is descriptive but is far too long. This title takes a while to read and clutters up the space above the chart.



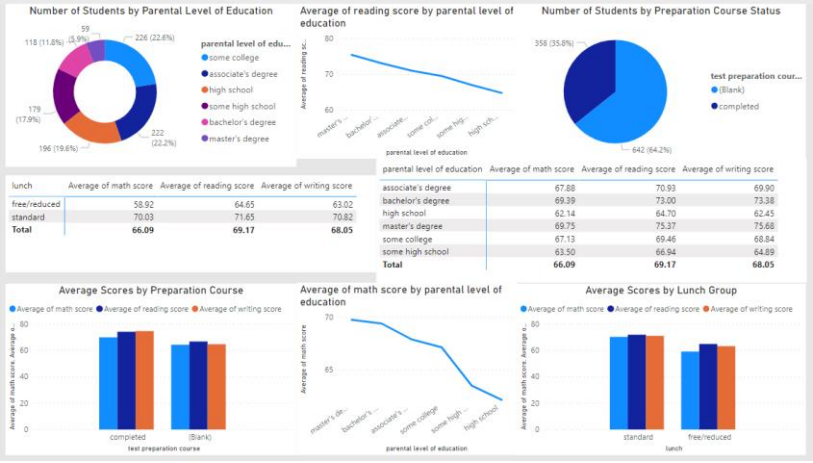
This second visualization has a title that is still descriptive but is more concise and takes up much less space. A viewer can read the title quickly and understand what the chart shows.



Avoid overcrowding

While creating a dashboard, it can be tempting to dump a bunch of information on it. But try to avoid cluttering your dashboard with unnecessary information. Keeping a clean dashboard makes it easier to visually navigate around and identify key information. Visualizations on dashboards are presented altogether, not one at a time. So, too much information at once can be overwhelming to the viewer and may not convey your message properly.

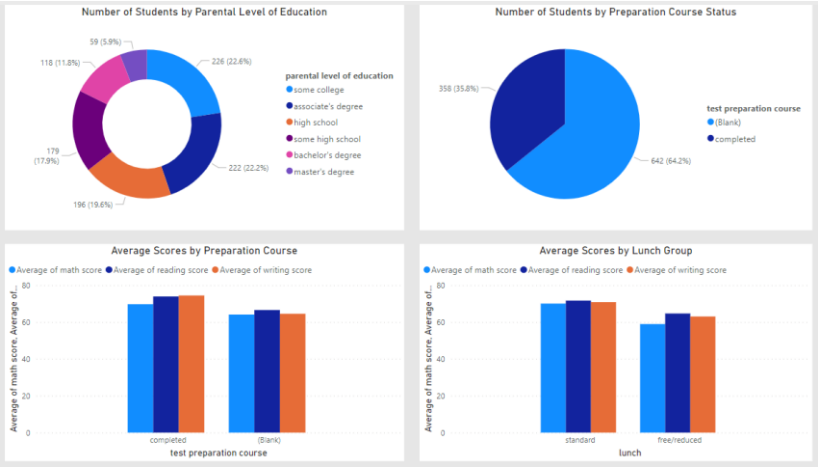
In the first dashboard below, there are too many visualizations. It's busy and cluttered, which makes it hard to process the information quickly or know where to look.



lunch	Average of math score	Average of reading score	Average of writing score
free/reduced	58.92	64.65	63.02
standard	70.03	71.65	70.82
Total	**66.09**	**69.17**	**68.05**

parental level of education	Average of math score	Average of reading score	Average of writing score
associate's degree	67.88	70.93	69.90
bachelor's degree	69.39	73.00	73.38
high school	62.14	64.70	62.45
master's degree	69.75	75.37	75.68
some college	67.13	69.46	68.84
some high school	63.50	66.94	64.89
Total	**66.09**	**69.17**	**68.05**

But the dashboard below shows only four visualizations, and they're clean and organized. It's much easier to digest the information in this dashboard.

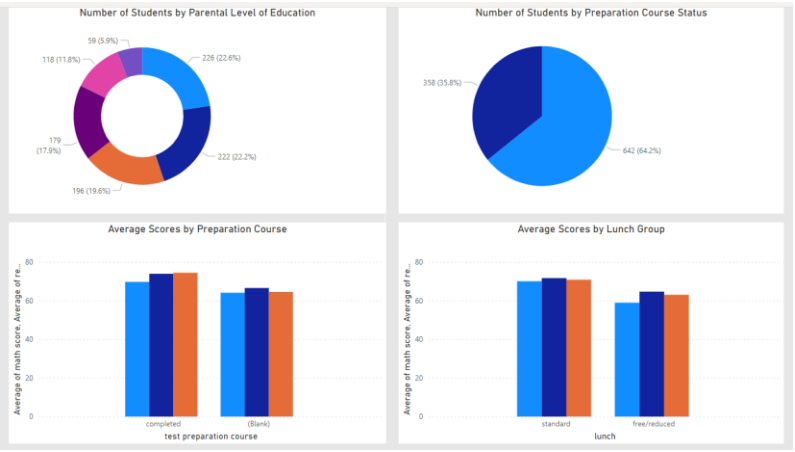


Include filters and legends

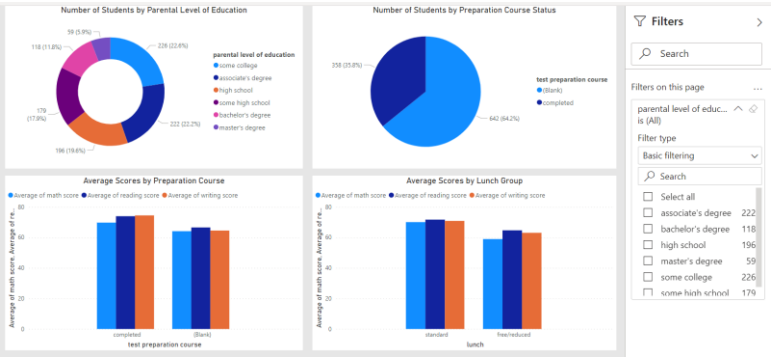
Make sure to include legends in your dashboard if your visualizations use colors. You want the information to be as easy as possible for your audience to interpret. Most visualization tools include legends by default, which makes this easy to remember.

In addition, if you're presenting a dashboard live to an audience, include filters that may be of interest to them. For example, if you have a line graph showing total global sales for five companies together, it's a good idea to add a filter on company names. This way, if your audience is interested in seeing global sales for individual companies, you can immediately filter it for them.

The dashboard below doesn't have legends or filters for the viewer to see. It's impossible to know what each color in each chart represents, which makes this dashboard challenging to interpret.



But the following dashboard shows legends so that the viewer can interpret the charts. The viewer can also focus on a specific parental level of education using the filter.



Conclusion

This lesson introduced dashboards and described dashboard best practices.