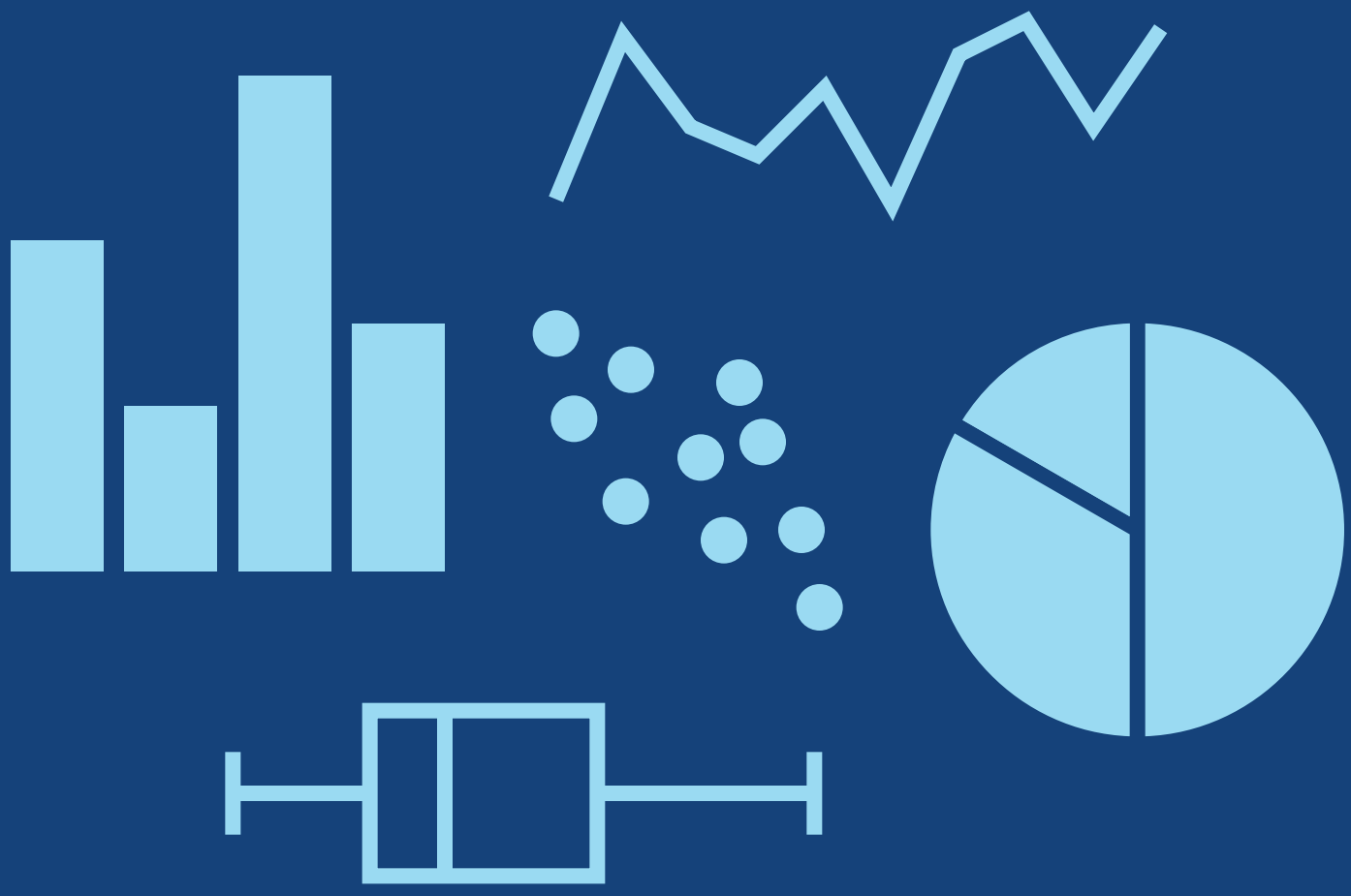
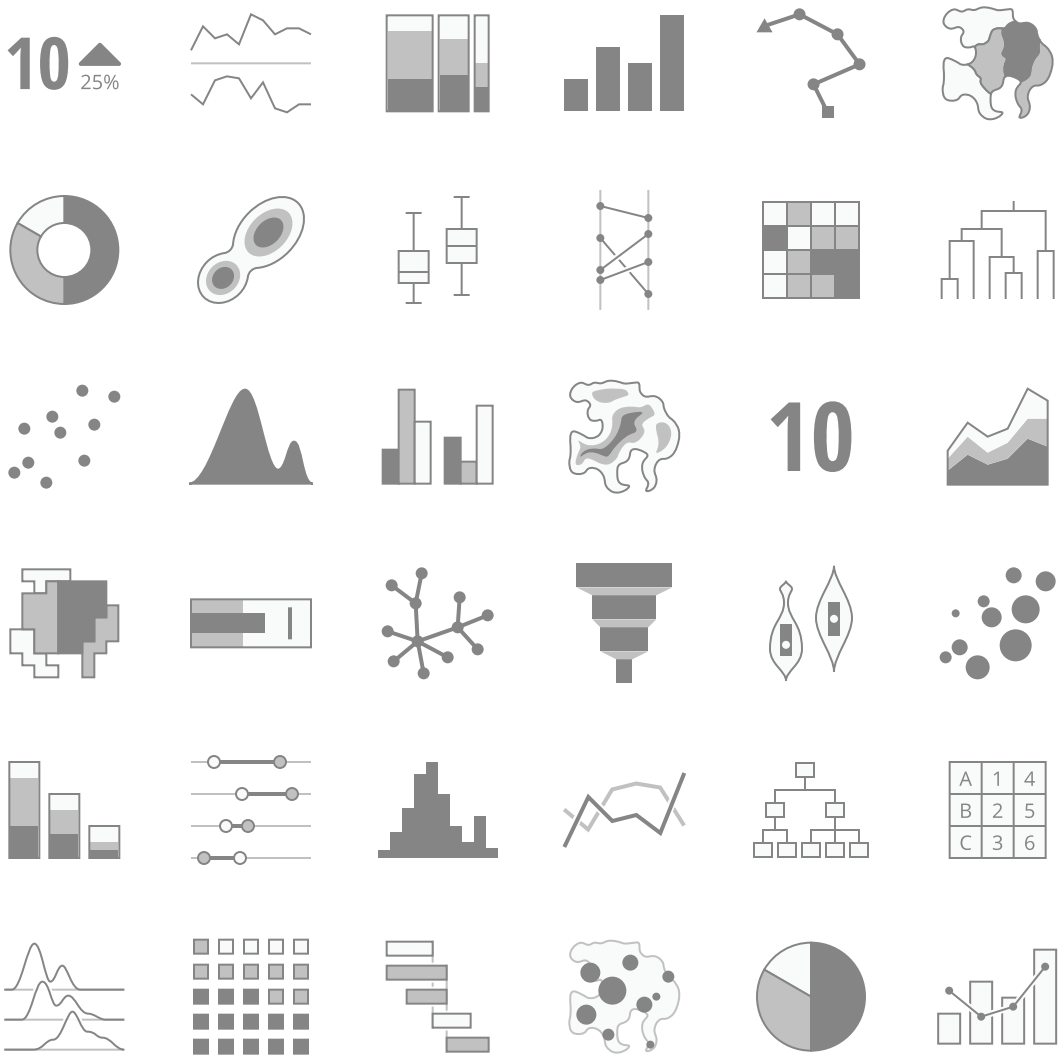
How to Choose the Right Data Visualization



CHARTIO

How to Choose the Right Data Visualization

by Mike Yi



CHARTIO

Introduction

Data visualizations are a vital component of a data analysis, as they have the ability to efficiently summarize large amounts of data through a graphical format. There are many chart types available, each with their own strengths and use cases. One of the trickiest parts of the analysis process is choosing the right way to represent your data using one of these visualizations.

When deciding on a chart type, first think about the type of role the chart will serve. Common roles for data visualization include:

- showing change over time
- showing a part-to-whole composition
- depicting flows and processes
- looking at how data is distributed
- comparing values between groups
- observing relationships between variables
- looking at geographical data

Next, consider the types of data you want to plot. The type of chart you use will depend on if the data is categorical, numeric, or some combination of both. Certain visualizations can also be used for multiple purposes depending on these factors. This book is organized with this approach in mind, with one chapter for each visualization role, each with multiple chart types to cover common types of data and subtasks.

Note that this document should only serve as a general guideline: it is possible that breaking out of the standard modes will help you gain additional insights. Experiment with not just different chart types, but also how the variables are encoded in each chart. It's also good to keep in mind that you aren't limited to showing everything in just one plot. It is often better to keep each individual plot as simple and clear as possible, and instead use multiple plots to make comparisons, show trends, and demonstrate relationships between multiple variables.

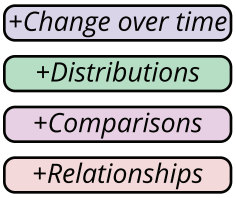
How this book is organized

This book is divided into chapters, one for each of the main categories for using a data visualization. Each chapter is headed by a short introduction, followed by a list of chart types falling in that category. Each chart type is accompanied by a short description and one or more icons. Below is a key for decoding these symbols:

- **BASIC:** Chart types with this icon represent typical or standard chart types. When you need to create a data visualization, try to see if one of these chart types works first, before deciding on an uncommon or advanced type.

- **UNCOMMON**: Chart types with this icon are slightly more unusual than the most common chart types. Use cases for these charts are more specialized than other chart types in that same category or more frequently seen in other roles.

- **ADVANCED**: Chart types with this icon are even more specialized in their roles. Make sure that the chart type is the best one for your use case before implementing it. Sometimes, these chart types will not be built into visualization software or libraries, and additional work will need to be done to put these types of chart together.



Connection icons: Some chart types appear in multiple chapters of the book, having either multiple use cases or use cases that straddle multiple roles. In these cases, you'll see a rounded rectangle with its entry noting the other chapters in which that chart type directly appears.

Chart types seen in boxes represent sub-topics within each visualization role; these will have more specialized and advanced use cases.

Table of contents

Raw numbers: just showing the data

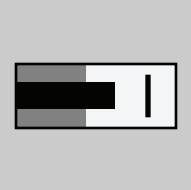
It is important to keep in mind that you don't always need to use a chart to depict your data. Sometimes, just showing the data as text is the most effective way of conveying information.

Single value chart ●

When you just have one number, it's best to just report it as-is. Plotting a single value graphically (such as with a bar or point) usually isn't meaningful if there aren't other values to compare it to.

Single value with indicator ●

An indicator compares the single value to a second number. This is often to compare a metric's value between the current period and the previous period.



Bullet chart ●

Chart type comparing a single value to another number, often a benchmark rather than another data point. The single value is shown with a bar's length, while comparison points are shown as shaded regions or a perpendicular line.

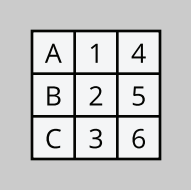


Table ●

Compares data points (rows) across multiple different attributes (columns). Usually sorted by an important or prominent attribute to improve utility.

Charts for showing change over time

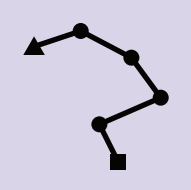
One of the most common applications for visualizing data is to see the change in numeric value for a feature or metric across time. These charts usually have time on the horizontal axis, moving from left to right, with the variable of interest's values on the vertical axis.

Line chart ● *+Comparisons*

Most common chart type for showing change over time. A point is plotted for each time period from left to right; each point's vertical position indicates the feature's value. Points are connected by line segments to emphasize progression across time.

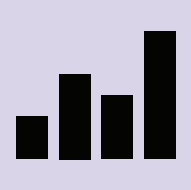
Sparkline ● *+Comparisons*

A miniature line chart with little to no labeling, designed to be placed alongside text or in tables. Provides a high-level overview without attracting too much attention. Can also be seen in a sparkbar form, or miniature bar chart (see below).



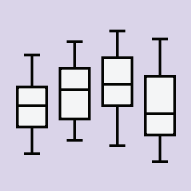
Connected scatter plot ◆ *+Relationships*

Shows change over time across two numeric variables (see scatter plot in *Relationships*). Line segments still connect points across time, but they may not consistently go from left to right like in a line chart.



Bar chart ● *+Distributions* *+Comparisons*

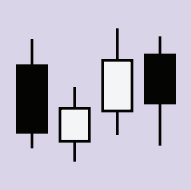
Each time period is associated with a bar; each bar's value is represented in its height above (or below) a zero-baseline. Works best when there aren't too many time periods to show.



Box plot +Distributions +Comparisons

Each time period is associated with a box and whiskers; each set of box and whiskers shows the range of the most common data values. Best when there are multiple recordings for each time period and a distribution of values needs to be plotted.

Tracking change over time is of key interest in the financial domain. One specialist chart developed for this field includes the following:



Candlestick chart

Looks like a box plot, but each box and whiskers encodes different statistics. The box ends indicate opening and closing prices, while color indicates the direction of change.

Charts for showing part-to-whole composition

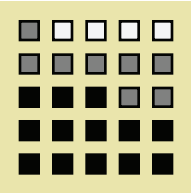
Sometimes, we need to know not just a total, but the components that comprise that total. While other charts like a standard bar chart can be used to compare the values of the components, the following charts put the part-to-whole decomposition at the forefront.

Pie chart 

The whole is represented by a filled circle. Parts are proportional slices from that circle, one for each categorical group. Best with five or fewer slices with distinct proportions.

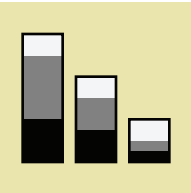
Doughnut chart 

A pie chart with a hole in the center. This central area can be used to show a relevant single numeric value. Sometimes used as an aesthetic alternative to a standard progress bar (see stacked bar chart below).



Waffle chart / grid plot 

Squares laid out in a (typically) 10 x 10 grid; each square represents one percent of the whole. Squares are colored based on categorical group size.



Stacked bar chart 

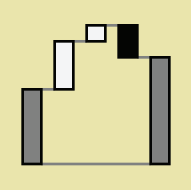
A bar chart (see *Change over time* or *Distributions*) where each bar has been divided into multiple sub-bars to show a part-to-whole breakdown. A single stacked bar can be used as an alternative to the pie or doughnut chart; people tend to make more precise judgments of length over area or angle.

Stacked area chart ●

A line chart (see *Change over time*) where shaded regions are added under the line to divide the total into sub-group values.

Stream graph ◆

Modified version of the stacked area chart where areas are stacked around a central axis. Highlights relative changes instead of exact values.



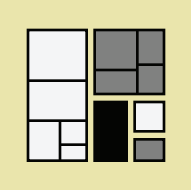
Waterfall chart ◆

Augments a change over time with a part-to-whole decomposition. Bars on the ends depict values at two time points, and lengths of intermediate floating bars' show the decomposition of the change between points.

Certain part-to-whole compositions follow a hierarchical form. In these cases, each part can be divided into finer parts on lower levels. Here are a couple of more specialized chart types for visualizing this type of data:

Mosaic plot / Marimekko chart ■

Can be thought of as a stacked bar divided on both axes. A box is divided on one axis based on one categorical variable, then each sub-box is divided in the other axis based on a second categorical variable.

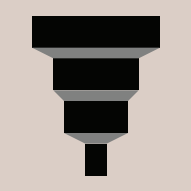


Treemap ◆

Can be thought of as a more generalized Marimekko plot. Sub-boxes do not need to have a consistent cut direction at a particular hierarchy level, and there can be more than two levels of hierarchy.

Charts for depicting flows and processes

A more specialized use for charts related to decomposition of a whole is the tracking of the flow of amounts across a multi-stage process. At their most advanced, these charts can efficiently show how multiple inputs are transformed into multiple outputs.

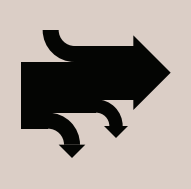


Funnel chart ■

Seen in business contexts, showing how people encounter a product and eventually become users or customers. One bar is plotted for each stage, whose lengths reflect the number of users. Connecting regions emphasize connections in stages and give the chart type's namesake shape.

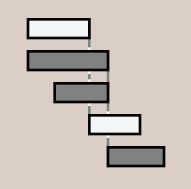
Parallel sets chart ◆

Multiple part-to-whole divisions on different dimensions are depicted as parallel stacked bars. Connecting regions show how different subgroups relate to one another between dimensions.



Sankey diagram ◆

The width of the colored region shows the relative volume at each part of a process. Allows for multiple sources of inputs and outputs to be visualized.



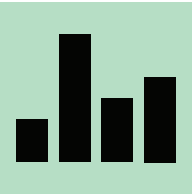
Gantt chart ■

Used for project scheduling, breaking them down into individual tasks. Each task is associated with a bar, providing a timeline for when each task should begin and end.

Charts for looking at how data is distributed

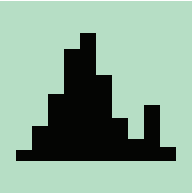
One important use for visualizations is to show how data points' values are distributed. This is particularly useful during the exploration process, when trying to build an understanding of the properties of data features.

Note: Charts for visualizing data distributions across two or more variables are covered in the Relationships chapter.



Bar chart ● *+Change over time* *+Comparisons*

Used when a variable is qualitative or takes discrete values. The height of each bar indicates the amount of each categorical group.

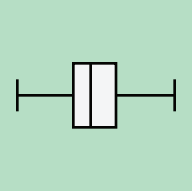


Histogram ●

Similar to a bar chart, but used when a variable takes continuous numeric values. The variable's numeric range is divided into bins for aggregating counts. Bars are plotted flush against each other to emphasize the variable's continuous nature.

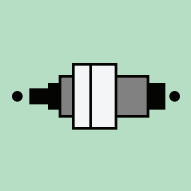
Density curve ●

An alternative to the histogram when a variable takes numeric values. Each data point contributes a small amount of local area; the areas are summed across all points to form the full curve.



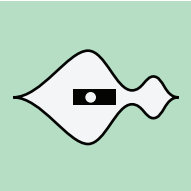
Box plot ■ *+Change over time* *+Comparisons*

A box and whiskers shows the range of the most common data values. The ends of the box outline the central 50% of the data. More often used to compare distributions between groups rather than as an overall summary.



Letter-value plot ◆ *+Comparisons*

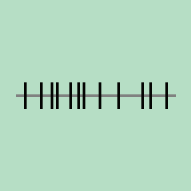
Extends the box plot's marking of quartiles with additional boxes that denote eighths, sixteenths, and smaller quantiles. Best when there are lots of data available to make estimates stable.



Violin plot ■ *+Comparisons*

Combines a density curve plotted on a center line with a box plot as a statistical summary. More often used to compare distributions between groups rather than as an overall summary.

The violin plot usually includes a box plot to provide statistical detail to the density curve. The internal box plot may sometimes be excluded, or another type of linear distribution chart can also be used instead. All of the below are best with few or a moderate number of data points; with many data points, a summary like the box plot is best.

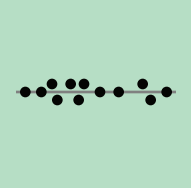


Rug plot ■

All data points are plotted as tick marks on a straight line with value corresponding precisely with position.

Strip plot ■

Like a rug plot, but with dots instead of tick marks. Sometimes plotted with points randomly jittered up or down to reduce overlapping.



Swarm plot ◆

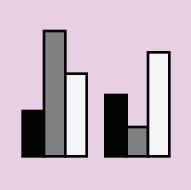
Like a strip plot, but deliberate shifting is performed to prevent overlapping. Some horizontal jitter may be needed in order to keep the dot swarm compact.

Charts for comparing values between groups

A very common application for data visualization is to compare values between distinct groups. This is frequently combined with other roles for data visualization, like showing change over time, or looking at how data is distributed. As a result, this is the largest category of chart types.

Bar chart *+Change over time* *+Distributions*

Most basic way of comparing numeric values between groups or categories. Each group is assigned a bar; each bar's value is represented in its height above (or below) a zero-baseline.

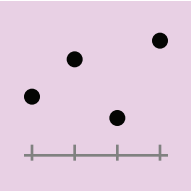


Grouped bar chart *+Relationships*

Extends a bar chart to compare data across two categorical variables. Each bar corresponds to an intersection of variable levels: categories for one variable are indicated by the bar cluster positions, while the second variable is indicated by bar color or position within each cluster.

Lollipop chart

Replaces the bars of a bar chart with lines and dots. Useful for when there are a lot of groups or categories to plot.



Dot plot

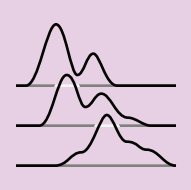
Replaces the bars of a bar chart with just dots. Since value is indicated by position instead of length, the dot plot can be good when a zero baseline is not useful.

Line chart +Change over time

Each line in a line chart shows how values (vertical position) change across time (horizontal). One line is plotted for each group to be compared. Best when there are five or fewer groups to plot.

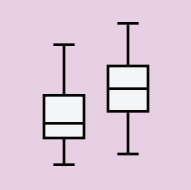
Sparkline +Change over time

Smaller line charts typically with little to no labeling. Designed to show a high-level overview inline with text or tables, but also useful when there are many groups to plot.



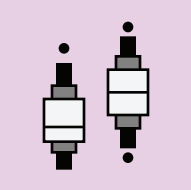
Ridgeline

A series of line charts or density curves (see *Distributions*) with partially offset axes used to compare distributions between groups. Best when there are distinct patterns across groups.



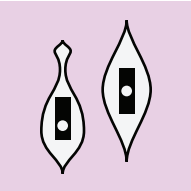
Box plot +Change over time +Distributions

Compares a statistical summary of numeric values between groups. A set of box and whiskers depicting the range of the most common data values (see *Distributions*) is assigned to each group or category.



Letter-value plot +Distributions

Used in a similar way as the box plot, but a letter-value plot (see *Distributions*) is assigned to each group instead. Best used when there are lots of data in each group so that statistical estimates are stable.



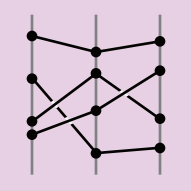
Violin plot +Distributions

Compares distributions between groups. A violin assembly of density curve and box plot (see *Distributions*) is assigned to each group or category.

One sub-category of comparison charts comes from the comparison of values between groups for multiple attributes.

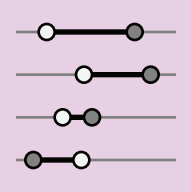
Slope chart

Specialized type of line chart. Two parallel lines indicate different times, with vertical position indicating value. One line segment is drawn between the two times for each data point. Useful for when there are many data points; line slopes provide a quick indicator for direction of change for each one.



Parallel coordinates plot

Extension of the slope plot for multiple dimensions. Each vertical line now indicates a different variable; each may have its own scale. Useful for observing patterns and relationships in the data. When there are only two variables, a scatter plot (see *Relationships*) is often easier to read.



Dumbbell plot

Used to compare two data points across multiple variables. Similar to parallel coordinates, each data point has a value plotted on each line. In contrast, line segments connect points *within* each variable, emphasizing the difference in value. Can be used as an alternative to the slope chart to show change between two time periods for multiple groups.

In certain cases, you might be interested in just the ranking between groups without needing to see the actual values.

Bump chart

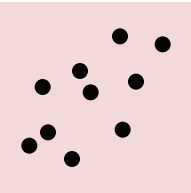
Modified version of a line chart where vertical position corresponds to rank rather than value. This change allows it to support more categories than a standard line chart.

Grouped bar chart

Normally, grouped bar charts will plot the bars within each group in a consistent order. However, they can instead be sorted by value within each group to emphasize ranking, at the cost of making it more difficult to find each sub-category.

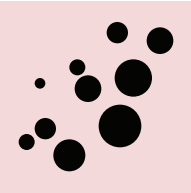
Charts for observing relationships between variables

One task that shows up in data exploration is understanding the relationship between data features. The chart types below can be used to plot two or more variables against each other to observe trends and patterns between them.



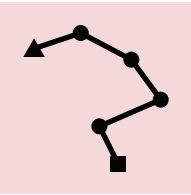
Scatter plot ●

Standard chart type for showing relationships between two numeric variables. Each point's position on the horizontal and vertical axes indicate value on the associated variable.



Bubble chart ●

Scatter plot with point size dictated by a third numeric variable. Scatter plots can be extended in other ways: point shapes can encode a categorical variable, and color can be used to indicate either categorical or numeric data. It is best to keep a scatter plot to a maximum of three variables to maintain understandability.

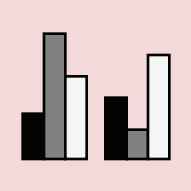


Connected scatter plot ◆

When a third variable represents time, points in a scatter plot can be connected with line segments to show progression in values across time.

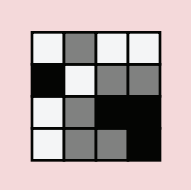
Dual-axis bar-line plot ◆

A bar-line plot shares a horizontal axis (typically time) across two chart types: the bar chart and line chart. Useful for when the variables plotted with each chart type are related, but are on different numeric scales.



Grouped bar chart ● +Comparisons

Extension of bar chart (see *Comparisons* or *Distributions*) to two categorical variables. Bar clusters are associated with levels of one variable, while color or position in each cluster indicates levels of the second variable. The length of each bar at the corresponding intersection of levels indicates a value for that group, like data frequency or a summary of a third numeric variable.

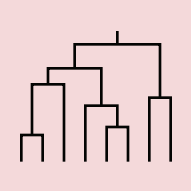


Heatmap ●

Extension of bar charts and histograms (see *Distributions*) to two variables, each of which can be categorical or numeric. Each axis represents groups or bins of values for one of the variables, forming a grid. Cell colors indicate data frequency or a summary of a third variable for each intersection of axis variables.

2-d density curve ◆

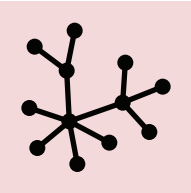
Extension of density curves (see *Distributions*) to two numeric variables. Colors are mapped to values like in a heatmap, but applied smoothly across the plotted area rather than in discrete bins. Somewhat confusingly, this chart is sometimes also known as a heatmap.



Dendrogram ◆

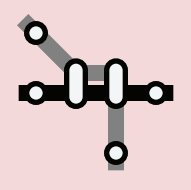
Specialized chart type to show similarity between data points. The lower the branch connecting two data points is, the more similar they are. Sometimes plotted with an accompanying heatmap to depict the underlying data.

Sometimes the form of a relationship is that of a network of connections. A mathematical graph consisting of nodes connected by edges is a basic form, but other chart types exist for showing this type of data.



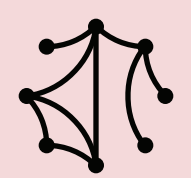
Network diagram ◆

Points (nodes or vertices) represent individual entities. Lines (edges) connect entities with a particular relationship. Line thickness may be used to encode value. Vertex positions do not necessarily have any inherent meaning, and may simply be placed just to make connections as clear as possible.



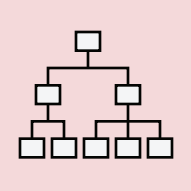
Transit map ◆

Practical application of network diagrams for train and subway systems. Frequently, these take a fair level of abstraction, emphasizing connections between stations rather than their actual geographical locations.



Chord diagram ◆

Like a standard network diagram, but vertices are arranged in a circle.



Tree diagram ◆

A network diagram organized to show hierarchical relationships. The direction of each edge corresponds to a relationship between the connected nodes, such as parent-child or senior-junior relationships.

Charts for looking at geographical data

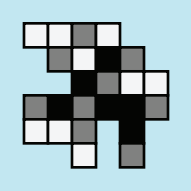
Sometimes, data includes geographical information like latitude and longitude or regions like country or state. While plotting this data might simply extend an existing visualization onto a map background (such as those in the previous chapter for depicting relationships), there are some chart types that specifically take the mapping domain into account.

Scatter map ●

Scatter plot built on top of a geographical map, using geographic coordinates as point positions.

Bubble map ●

Bubble chart built on top of a geographic map, where point size is an indicator of value. Can also be used to group together points in a scatter map if they are too dense.

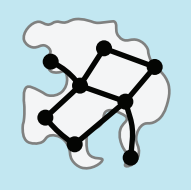


2-d histogram ●

Heatmaps can be built on top of geographic areas. Sometimes seen with a hexagon-shaped grid rather than a rectangular grid. May distort the geography on its edges.

Isopleth / contour map ◆

2-d density curve built on top of a geographic map.

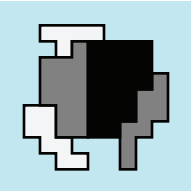


Connection map ◆

Network information and flows built on top of a geographic map.

Choropleth ●

Similar to a heatmap, but colors are assigned to geopolitical regions rather than an arbitrary grid. Values are often in the form of rates or ratios to avoid distortion due to population density.



Cartogram ◆

Geopolitical regions sized by value. This necessarily requires distortion in shapes and topology.

Appendix A: Essential charts for data analysis

This guide covers dozens of chart types, and many more exist for even more specialized use cases. It can sometimes be daunting to figure out which chart will work best for the data at hand.

To help with the chart choosing process, the next page contains a full-page graphic featuring eighteen common chart types for data analysis. Most visualizations for dashboards and reports will be served well one of these chart types. Feel free to print the graphic out and use it as a quick reference for any time you need to visualize your data.

When using the chart picker, don't forget to keep in mind three points:

1. What kind of **role** or analysis will the chart perform?
2. What **types of data** do I have – categorical or numeric – and how many variables am I going to plot?
3. After creating the chart, does it **convey useful information**? If not, try a different chart type or a different way of encoding variables. You may need an uncommon or advanced chart type not found in the diagram, or need to use more than one chart.

Essential Charts for Data Analysis

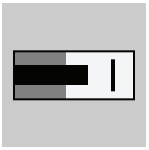
Raw Number

Single Value Chart

Show a **raw singular value**

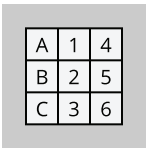
Single Value w/ Indicator

Comparison of a **single value** against a **previous value**



Bullet Chart

Comparison of a **single value** against a **benchmark value**



Table

Show raw values for **multiple data points** on **multiple variables**

Change over Time

Line Chart

Change over time for a **numeric** variable or to **compare 1-5 groups**

Sparkline

Miniature **line charts** to **compare many groups**

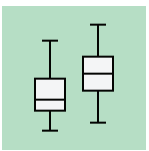
Distribution

Bar Chart

Comparison or **distribution** by a **single categorical** variable

Histogram

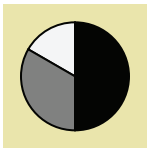
Distribution by a binned **single numeric** variable



Box Plot

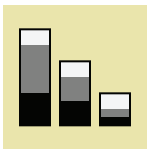
Compare distribution summaries across a **categorical** variable

Part-to-Whole



Pie Chart

Part-to-whole breakdown by a **single categorical** variable



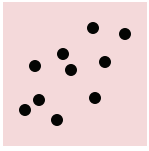
Stacked Bar Chart

Bar chart with additional **part-to-whole** breakdown

Stacked Area Chart

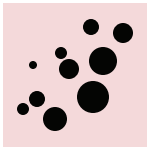
Line chart with additional **part-to-whole** breakdown

Relationship



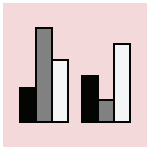
Scatter Plot

Relationship between **two numeric** variables



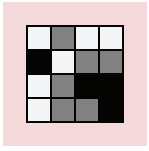
Bubble Chart

Relationship between **three numeric** variables



Grouped Bar Chart

Comparison or **distribution** by **two categorical** variables



Heatmap

Distribution by **two binned** variables (**categorical or numeric**)

Geospatial

Bubble Map

Bubble chart built on top of a **geographic map**

Choropleth

Comparison between **geopolitical regions** by **color**

Appendix B: Charts that should be used judiciously

There are a few chart types excluded from the guide that probably wouldn't be considered too rare or specialized. Chart types like the ones in this section have been excluded since they are less efficient than other, more common chart types, or have flaws that make them more difficult to understand. Only use these charts when you have a unique or specific point that would benefit from an alternative representation.

Pictogram / Isotype

Used to compare values between groups and other places a bar chart might be used. Each icon represents a specific quantity; values are usually rounded to the nearest whole number of icons. Thus, this loses some precision compared to the more common bar chart.

Circular / radial bar chart

A bar chart, but with bars plotted in concentric arcs. However, this distorts each group's value, since it is not clear whether values are indicated by bar angles or arc lengths. It is better to just stick with a standard bar chart.

Radar / spider plot

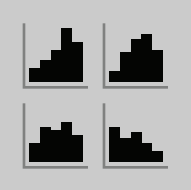
Used to compare values between data points on multiple attributes. Each attribute is a spoke, with value indicated by distance from the center. One polygon is traced for each data point. However, people often perceive value based on polygon area, which depends on the order of attributes. It is usually better to use a parallel coordinates plot or multiple grouped bar charts to avoid this distortion.

Appendix C: Additional ways to visualize data

There are many charting techniques that go beyond just choosing the right chart type and data encodings. Here are a few common techniques that can make your data easier to read and interpret.

Horizontal vs. vertical orientation

Certain chart types for performing comparisons, like the bar chart or box plot, can be plotted vertically or horizontally. The horizontal orientation can be useful when the groups have long names.



Small multiples / faceting

Rather than plotting multiple groups or categories on a single axis, just create one plot for each group. This can make it easier to distinguish between groups, especially when there are many of them (e.g. line chart vs. sparkline). You may want to order the facets by some data feature, like overall size.

Chart compositions / dashboards

Collections of charts, statistics, and tables are used to quickly convey key information to users at an organization. Grouping related elements together and arranging them from most important at the top and least important at the bottom can help viewers gain insights from the data.

About Chartio

Founded in 2010, Chartio is a cloud-based data exploration solution for all.

Featuring a revolutionary drag-and-drop interface and hundreds of data source integrations, Chartio empowers anyone at a company to easily connect, transform, and visualize the metrics that matter most to them – no coding required. With Chartio, users can create powerful interactive dashboards, collaborate with teammates, and share insights with just a few clicks. And Chartio comes with expert customer support and extensive documentation, ensuring that everyone in your company has the tools they need to successfully work with data and get informed.

CHARTIO

Learn how to quickly understand your business data at chartio.com, and learn more about how to work with data at chartio.com/learn/.