How to choose the right data visualization

POSTED BY: MIKE YI, MEL RESTORI

Data visualizations are a vital component of a data analysis, as they have the capability of summarizing large amounts of data efficiently in a graphical format. There are [many chart types available](), each with its 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.

In this article, we will approach the task of choosing a data visualization based on the type of task that you want to perform.

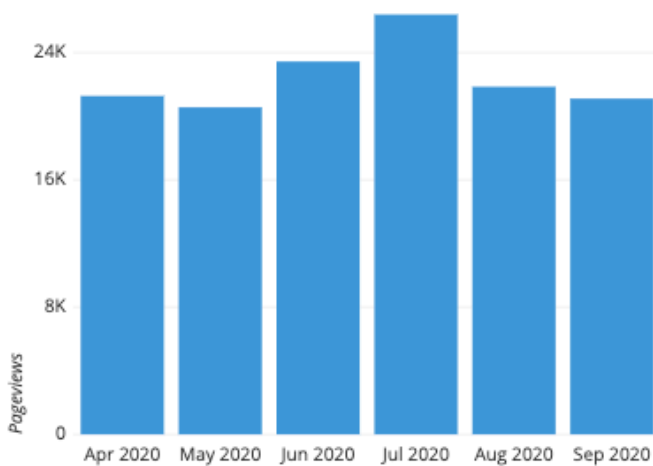Common roles for data visualization include:

- showing change over time

- showing a part-to-whole composition

- looking at how data is distributed

- comparing values between groups

- observing relationships between variables

- looking at geographical data

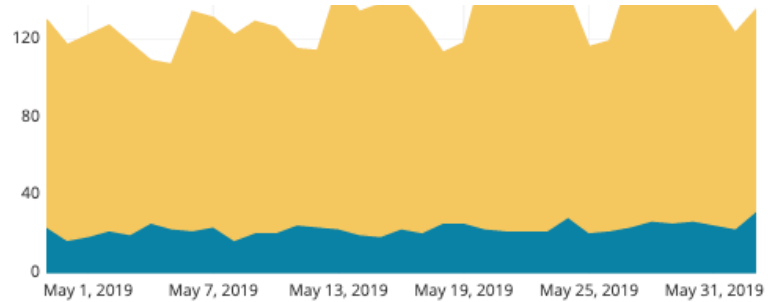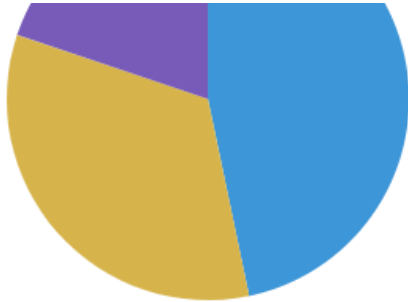encoding these values:

- [Bar charts](#) encode value by the heights of bars from a baseline.

- [Line charts](#) encode value by the vertical positions of points connected by line segments. This is useful when a baseline is not meaningful, or if the number of bars would be overwhelming to plot.

- A [box plot](#) can be useful when a distribution of values need to be plotted for each time period; each set of box and whiskers can show where the most common data values lie.

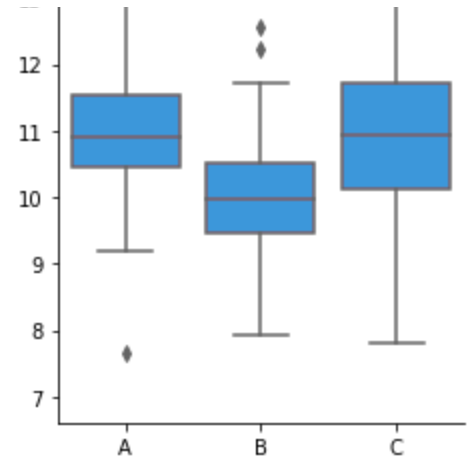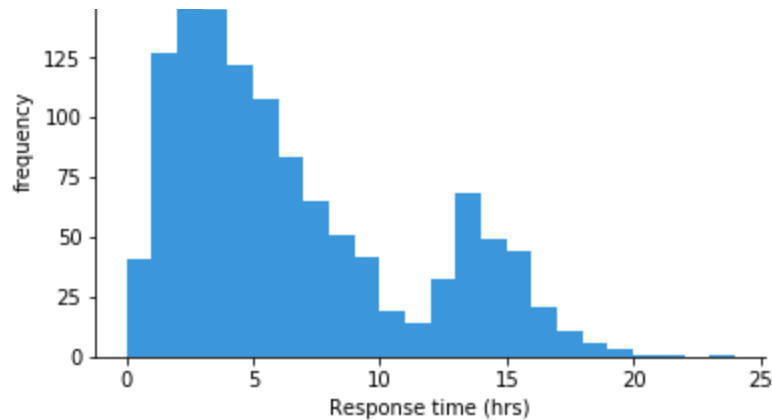- There are a number of specialist chart types for the financial domain, like the candlestick chart or Kagi chart.

- The pie chart and cousin donut chart represent the whole with a circle, divided by slices into parts.

- A stacked bar chart modifies a bar chart by dividing each bar into multiple sub-bars, showing a part-to-whole composition within each primary bar.

- Similarly, a stacked area chart modifies the line chart by using shading under the line to divide the total into sub-group values.

- A host of other more intricate chart types have also been developed to show hierarchical relationships. These include the Marimekko plot and treemap.

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.

- [Bar charts](#) are used when a variable is qualitative and takes a number of discrete values.

- A [histogram](#) is used when a variable is quantitative, taking numeric values.

- Alternatively, a **density curve** can be used in place of a histogram, as a smoothed estimate of the underlying distribution.

- A [violin plot](#) compares numeric value distributions between groups by plotting a density curve for each group.

- The [box plot](#) is another way of comparing distributions between groups, but with a summary of statistics rather than an estimated distributional shape.

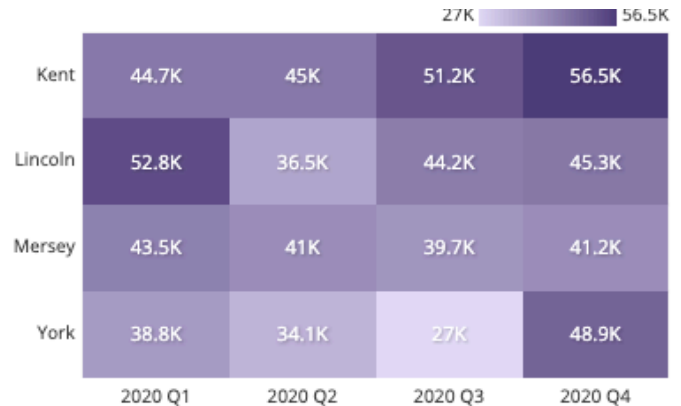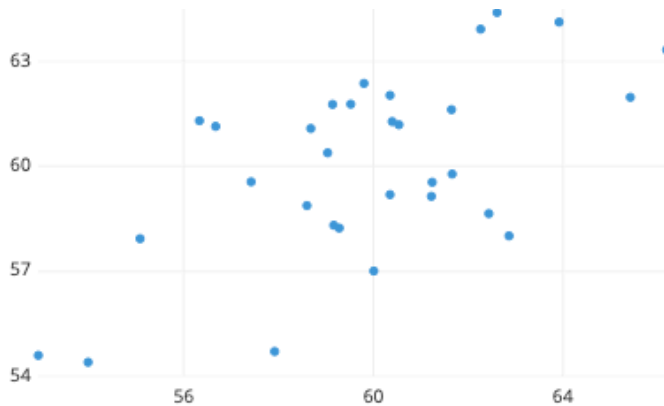## Charts for comparing values between groups

Another very common application for a data visualization is to compare values between

This site uses cookies to improve your browsing experience and perform analytics and research. We may also disclose information about your use of our site (including through the use of third-party cookies) to third parties for targeted advertising and other purposes described in our **Privacy Policy**. Depending on your location, you may have the right to opt out of certain disclosures of data for targeted advertising. To opt-out or change your preferences, click on "Manage Preferences" and follow the instructions. You may also opt out of sharing data for targeted marketing via the **Global Privacy Control**.
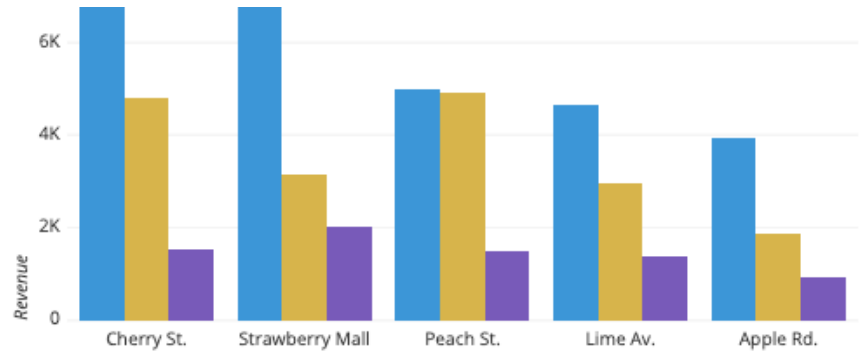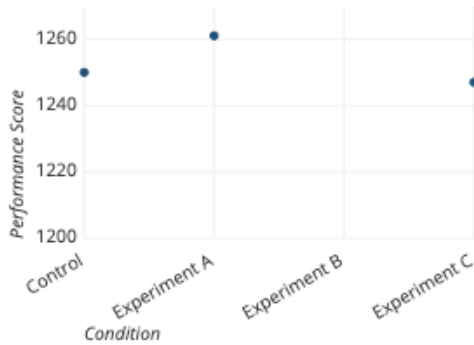
- A [bar chart](#) compares values between groups by assigning a bar to each group.

- A **dot plot** can be used similarly, except with value indicated by point positions instead of bar lengths. This is like a line chart with the line segments removed, eliminating the 'connection' between sequential points. Also like a line chart, a dot plot is useful when including a vertical baseline would not be meaningful.

- A [line chart](#) can be used to compare values between groups across time by plotting one line per group.

- A [grouped bar chart](#) allows for comparison of data across two different grouping variables by plotting multiple bars at each location, not just one.

- [Violin plots](#) and [box plots](#) are used to compare data distributions between groups.

- A [funnel chart](#) is a specialist chart for showing how quantities move through a process, like tracking how many visitors get from being shown an ad to eventually making a purchase.

- **Bullet charts** are another specialist chart for comparing a true value to one or more benchmarks.

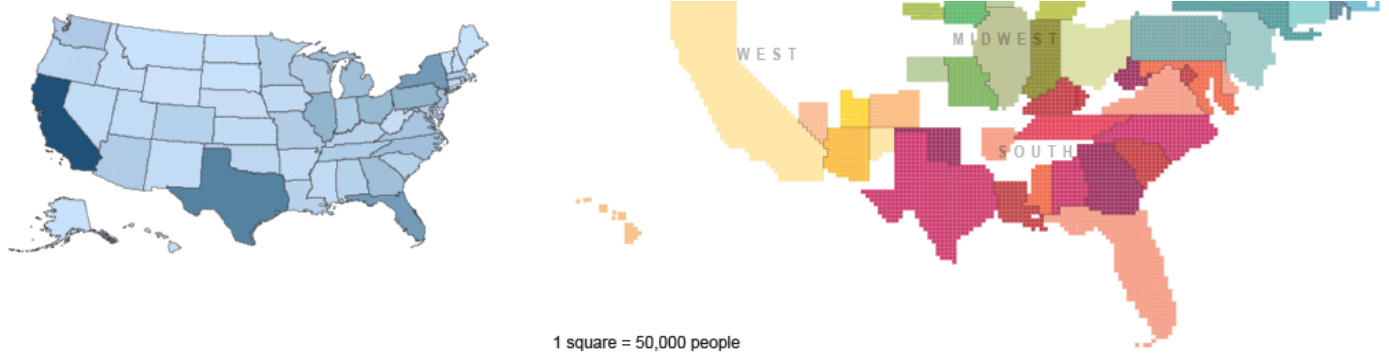- One sub-category of charts comes from the comparison of values between groups for

- The scatter plot is the standard way of showing the relationship between two variables.

- Scatter plots can also be expanded to additional variables by adding color, shape, or size to each point as indicators, as in a bubble chart.

- When a third variable represents time, points in a scatter plot can be connected with line segments, generating a **connected scatter plot**.

- Another alternative for a temporal third-variable is a **dual-axis plot**, such as plotting a line chart and bar chart with a shared horizontal axis.

- When one or both variables being compared are not numeric, a heatmap can show the relationship between groups. Heatmaps can also be used for purely numeric data, like in a 2-d histogram or 2-d density curve.

Charts for looking at geographical data

1 square = 50,000 people

Right: Cartogram of US Population from census.gov

- A **choropleth** is like a heatmap that colors in geopolitical regions rather than a strict grid.

- **Cartograms** take a different approach by using the size of each region to encode value. This approach necessitates some distortion in shapes and topology.

Closing thoughts

Choosing the right chart for the job depends on the kinds of variables that you are looking at and what you want to get out of them. The above is only 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. Often it is 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.

ATLASSIAN

Company

Careers

Events

Blogs

Investor Relations

Atlassian Foundation

Contact us

PRODUCTS

Rovo

Jira

Jira Align

Jira Service Management

Confluence

Loom

Trello

ATLASSIAN

Atlassian Community

Knowledge base

Marketplace

My account

Create support ticket ➔

Partners

Training & certification

Documentation

Developer resources

Enterprise services

See all resources ➔