DATA ANALYSIS AND VISUALIZATION

INSTRUCTOR: UMME AMMARAH

DATA VISUALIZATION

WHAT IS DATA VISUALIZATION

- The graphical or visual representation of data.
- It helps to highlight the most useful insights from a dataset, making it easier to spot
  - Trends
  - Patterns
  - Outliers
  - Correlations

STORY TELLING

- When done well, data visualization tells a story.
- This storytelling aspect is crucial as it makes your data actionable.
- There's a huge difference between simply having lots of data versus actually understanding how to use it to drive actions and decisions

TYPES

- Exploratory visualization
  - To figure out what's in your data
  - Exploration takes place while you're still analyzing the data
- Explanatory visualization
  - To communicate what you've found
  - Explanation comes towards the end of the process when you're ready to share your findings

WHY?

- Data analytics allows us to make sense of (at least some of) that data. We need it in order to make smart decisions—and data visualization is a crucial part of that.

- Data visualization helps us to understand what certain data is telling us, presenting it in a way that's accessible to a range of audiences—not just data experts.

- It's how you bridge the gap between your expertise as a data analyst or data scientist, and those people who can use or act upon the insights you discover.

PRINCIPLES

Accurate

Prioritize data accuracy, clarity, and integrity, presenting information in a way that doesn't distort it.

Helpful

Help users navigate data with context and affordances that emphasize exploration and comparison.

Scalable

Adapt visualizations for different device sizes, while anticipating user needs on data depth, complexity, and modality.

CHARTS

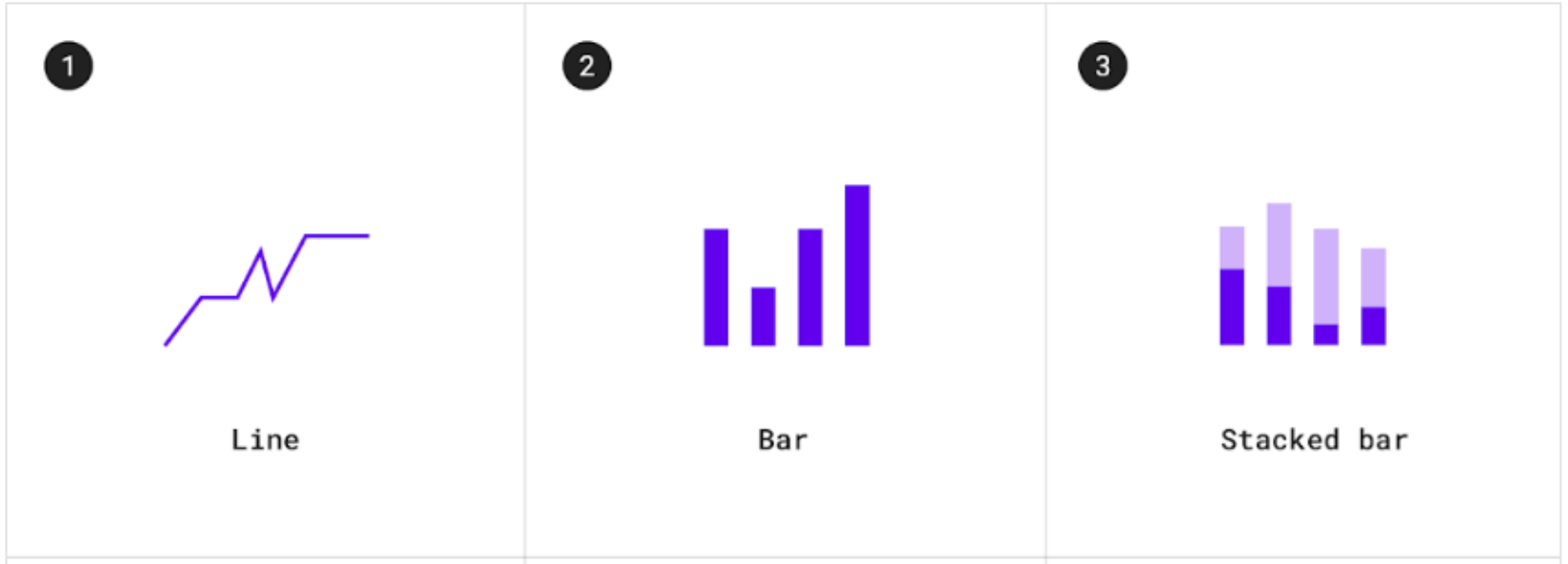
- Data visualization can be expressed in different forms. Charts are a common way of expressing data, as they depict different data varieties and allow data comparison.

- The type of chart you use depends primarily on two things:
  - the data you want to communicate, and
  - what you want to convey about that data.

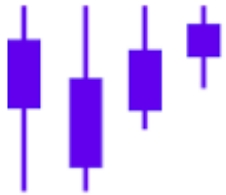
TYPES OF CHARTS

CHANGE OVER TIME

- Change over time charts show data over a period of time, such as trends or comparisons across multiple categories.
- Common use cases include:
  - Stock price performance
  - Health statistics
  - Chronologies



4



Candlestick

5

Area

6

Timeline

7

Horizon

8

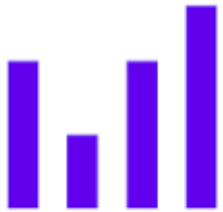
Waterfall

CATEGORY COMPARISON

- Category comparison charts compare data between multiple distinct categories.

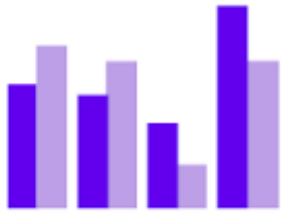
- Use cases include:
  - Income across different countries
  - Popular venue times
  - Team allocations

1



Bar

2



Grouped bar

3

Bubble

4

Parallel coordinates

5

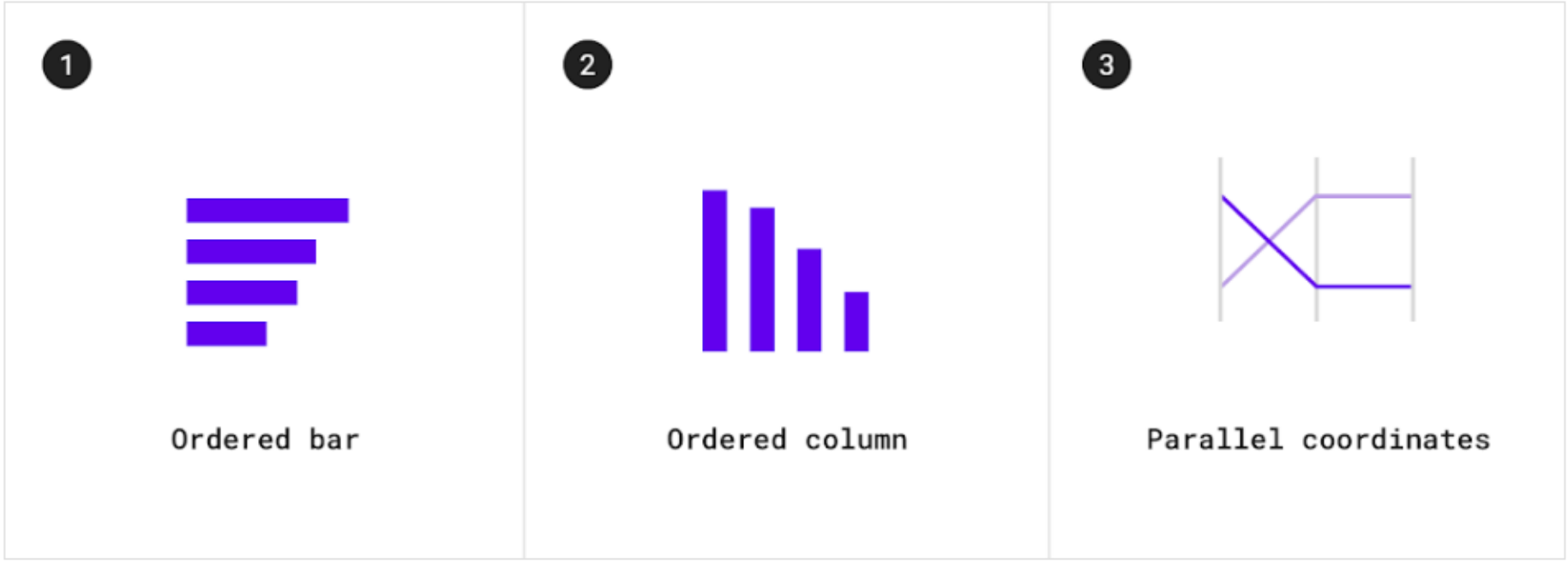
Multi-line

6

Bullet

RANKING

- Ranking charts show an item's position in an ordered list.
- Use cases include:
  - Election results
  - Performance statistics



PART-TO-WHOLE

- Part-to-whole charts show how partial elements add up to a total.

- Use cases include:
  - Consolidated revenue of product categories
  - Budgets

1

Stacked bar

2

Pie

3

Donut

4

Stacked area

5

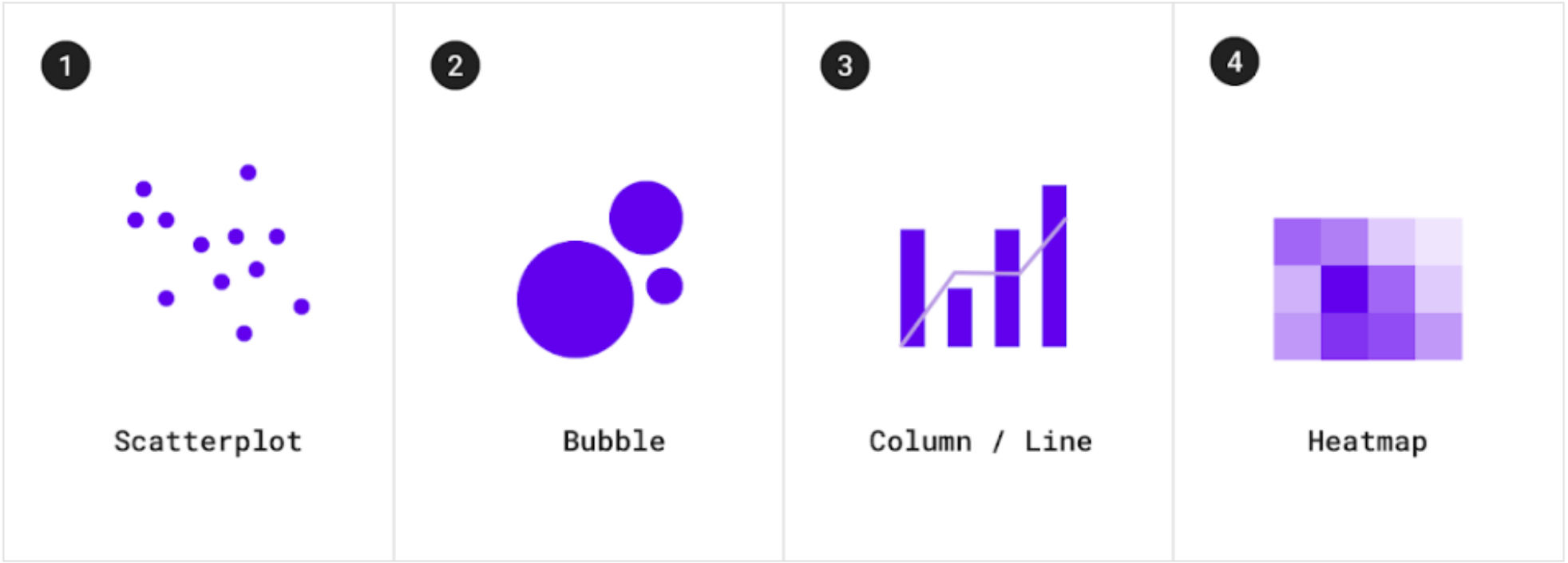
Treemap

6

Sunburst

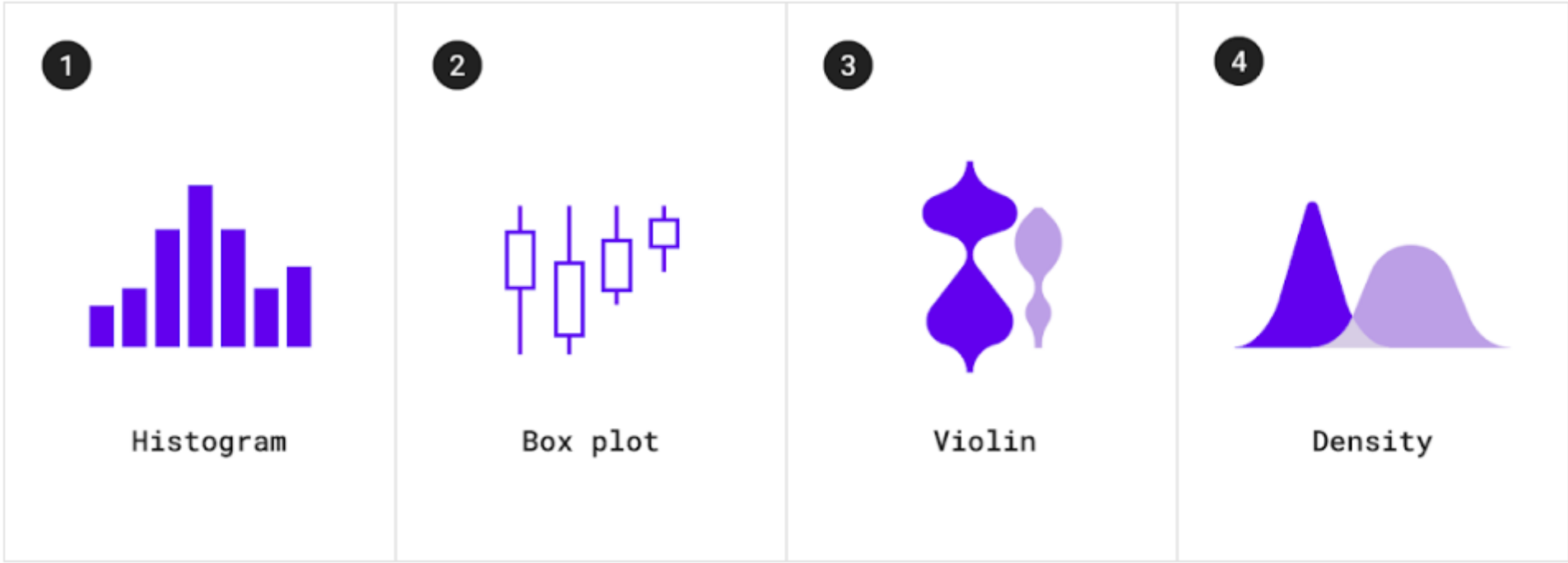
CORRELATION

- Correlation charts show correlation between two or more variables.
- Use cases include:
  - Income and life expectancy



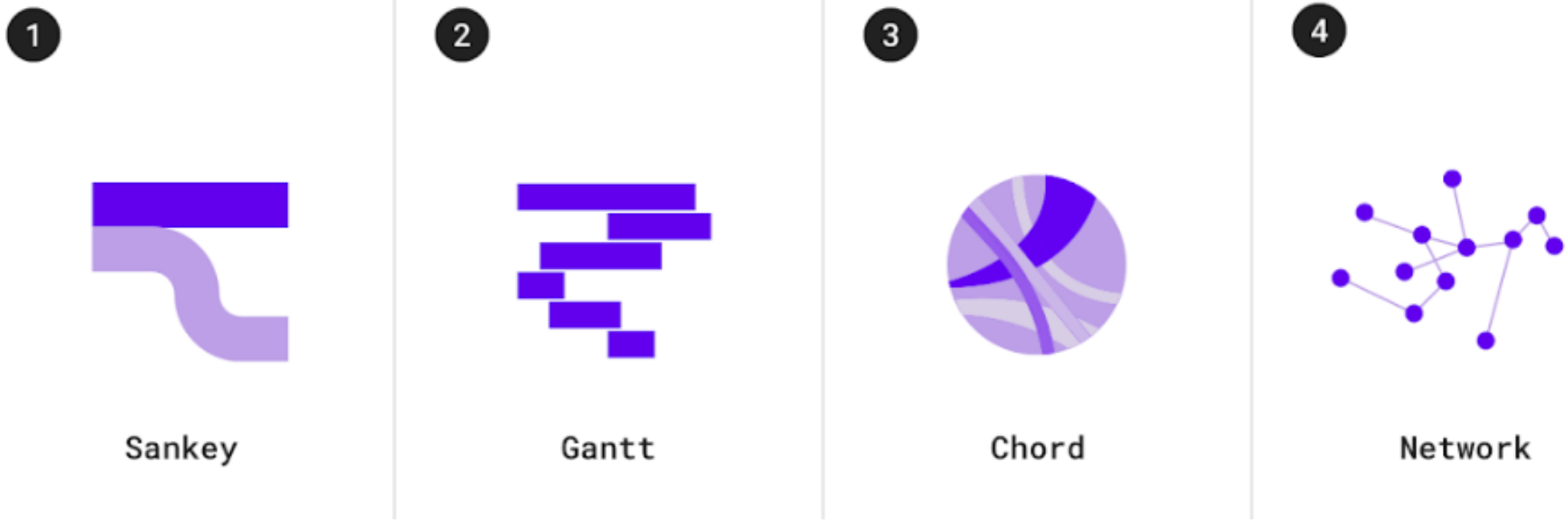
DISTRIBUTION

- Distribution charts show how often each values occur in a dataset.
- Use cases include:
  - Population distribution
  - Income distribution



FLOW

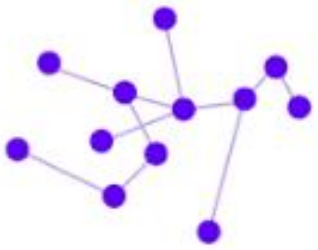
- Flow charts show movement of data between multiple states.
- Use cases include:
  - Fund transfers
  - Vote counts and election results



RELATIONSHIP

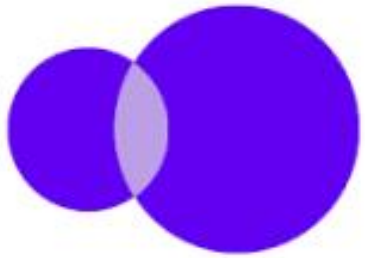
- Relationship charts show how multiple items relate to one other.
- Use cases include
  - Social networks
  - Word charts

1



Network

2



Venn diagram

3

Chord

4

Sunburst

SELECTING CHARTS

SHOWING CHANGE OVER TIME

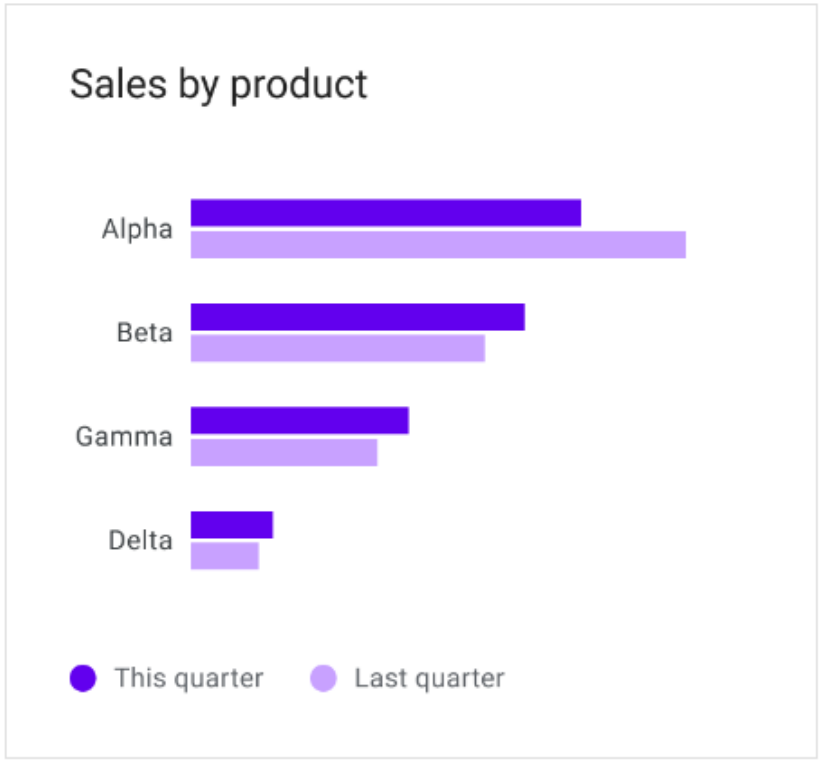
Type of chart	Usage	Baseline value *	Quantity of time series	Data type
Line chart	To express minor variations in data	Any value	Any time series (works well for charts with 8 or more time series)	Continuous
Bar chart	To express larger variations in data, how individual data points relate to a whole, comparisons, and ranking	Zero	4 or fewer	Discrete or categorical
Area chart	To summarize relationships between datasets, how individual data points relate to a whole	Zero (when there's more than one series)	8 or fewer	Continuous

BAR AND PIE CHARTS

- Both bar charts and pie charts can be used to show proportion, which expresses a partial value in comparison to a total value.
  - Bar charts express quantities through a bar's length, using a common baseline
  - Pie charts express portions of a whole, using arcs or angles within a circle

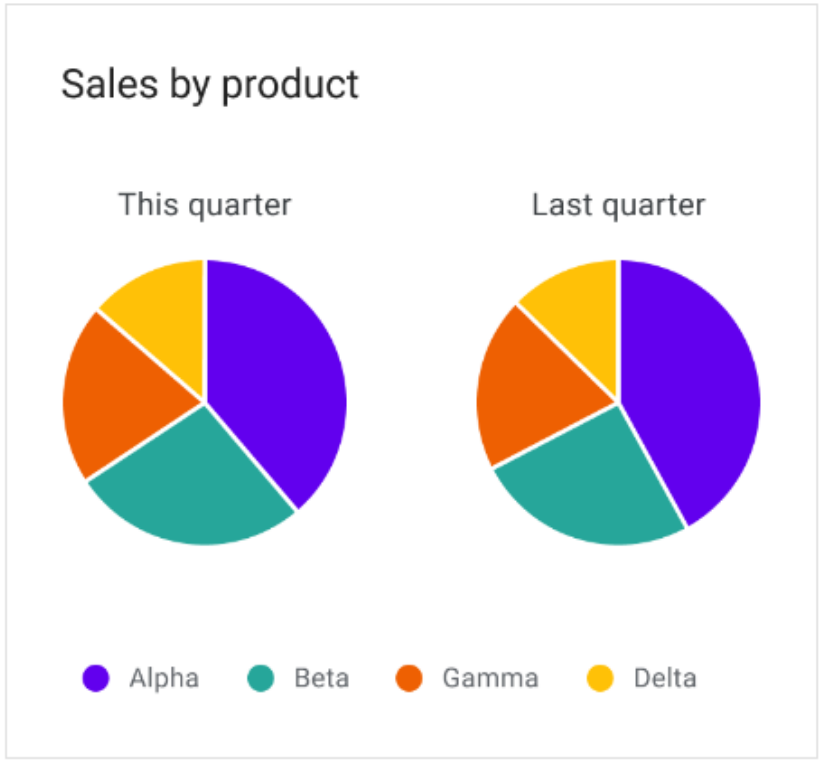
- Bar charts, line charts, and stacked area charts are more effective at showing change over time than pie charts. Because all three of these charts share the same baseline of possible values, it's easier to compare value differences based on bar length.

EXAMPLE



Do

Use bar charts to show changes over time or differences between categories.



Don't

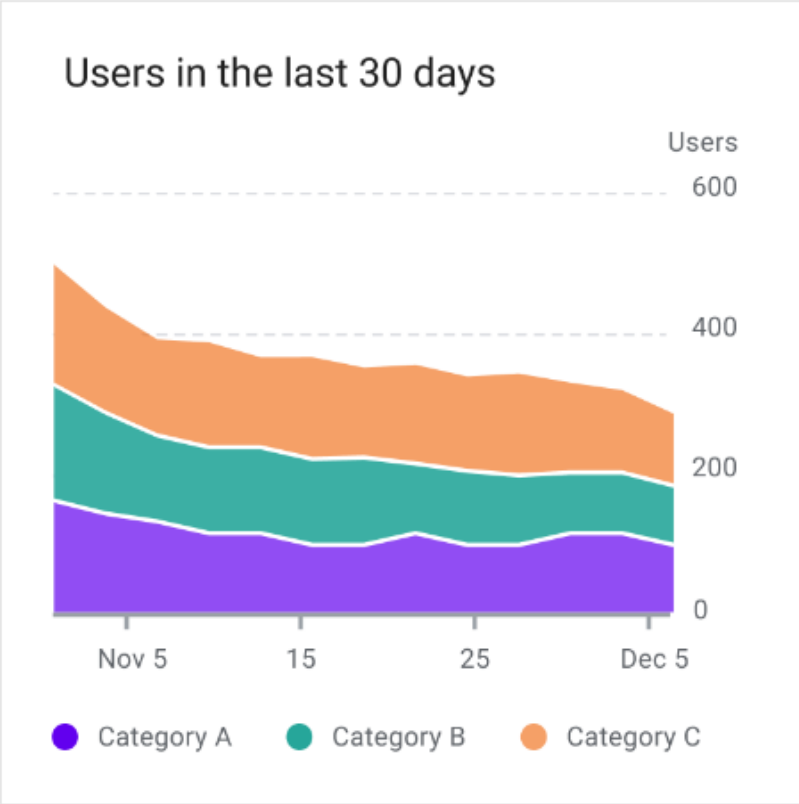
Don’t use multiple pie charts to show changes over time. It’s difficult to compare the difference in size across each slice of the pie.

AREA CHARTS

- Area charts come in several varieties, including stacked area charts and overlapped area charts:
  - Stacked area charts show multiple time series (over the same time period) stacked on top of one another
  - Overlapped area charts show multiple time series (over the same time period) overlapping one another

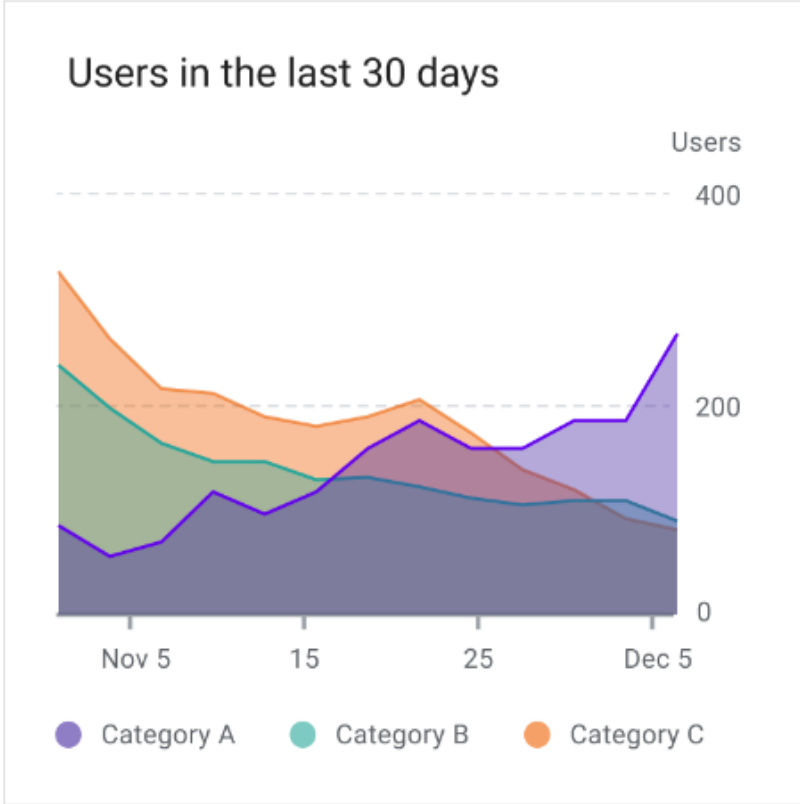
- Overlapping area charts are not recommended with more than two time series, as doing so can obscure the data. Instead, use a stacked area chart to compare multiple values over a time interval (with time represented on the horizontal axis).

EXAMPLE



Do

Use a stacked area chart to represent multiple time series and maintain a good level of legibility.



Don't

Don’t use overlapped area charts as it obscures data values and reduces readability.

STYLE

STYLE

- Data visualizations use custom styles and shapes to make data easier to understand at a glance, in ways that suit the user's needs and context.

- Charts can benefit from customizing the following:
  - Graphical elements
  - Typography
  - Iconography
  - Axes and labels
  - Legends and annotations

VISUAL ENCODING

- **Visual encoding** is the process of translating data into visual form. Unique graphical attributes can be applied to both quantitative data (such as temperature, price, or speed) and qualitative data (such as categories, flavors, or expressions).

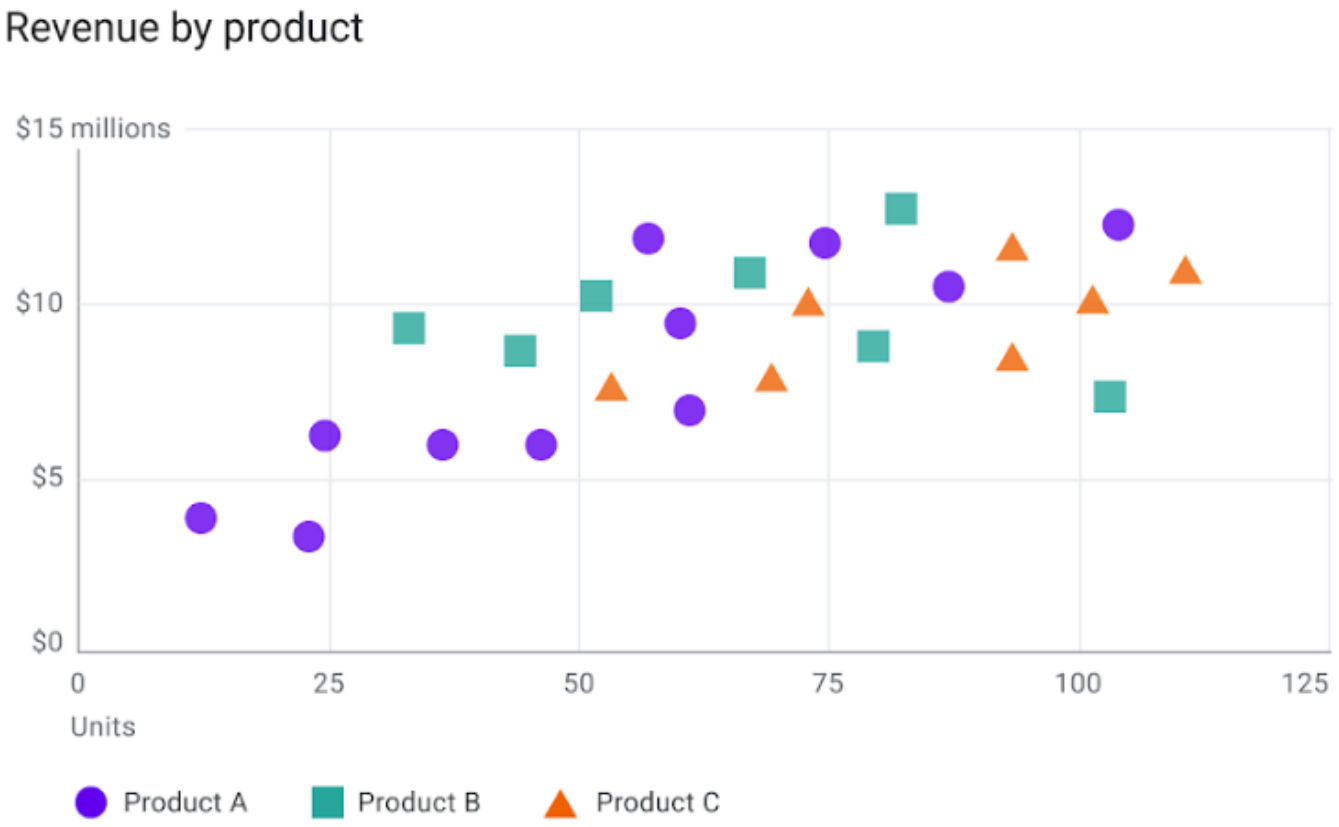
- These attributes include:
  - Shape
  - Color
  - Size
  - Area
  - Volume
  - Length
  - Angle
  - Position
  - Direction
  - Density

EXPRESSING MULTIPLE ATTRIBUTES

- Multiple visual treatments can be applied to more than one aspect of a data point. For example, a bar color can represent a category, while a bar's length can express a value (like population size).

SHAPE

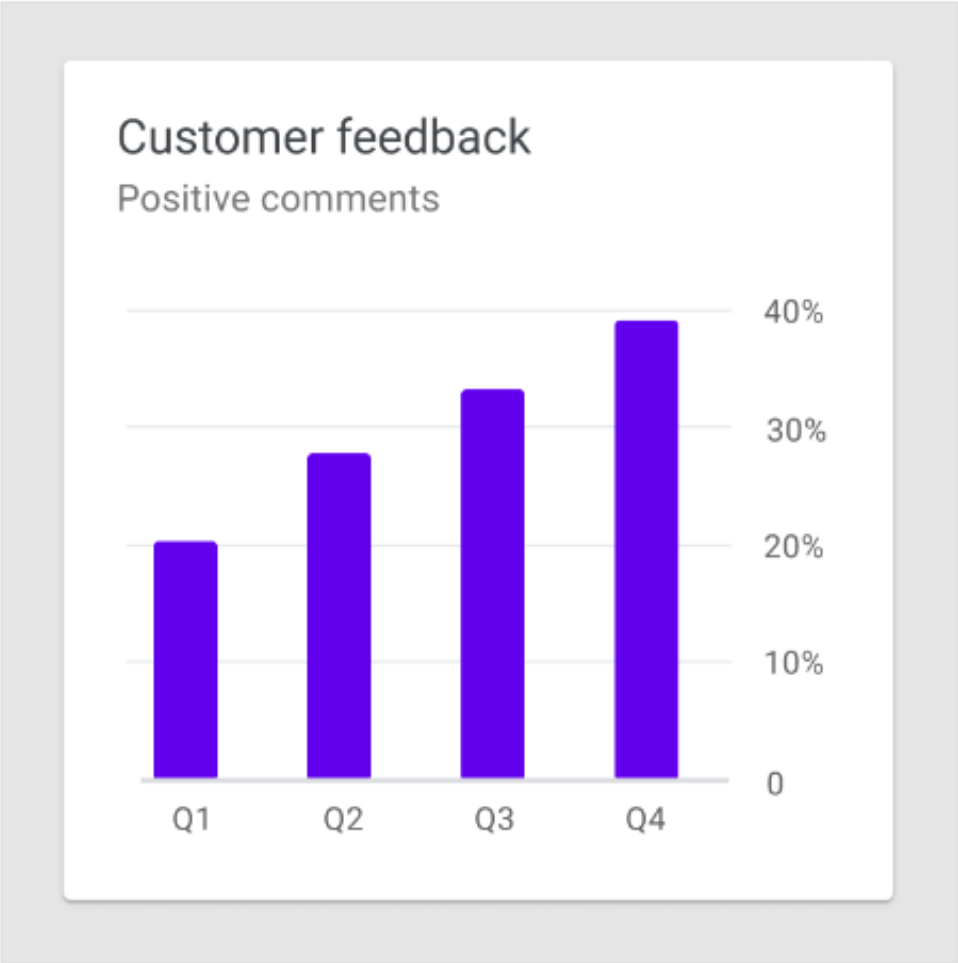
- Shape can be used to represent qualitative data. In this chart, each category is represented by a specific shape (circles, squares, and triangles), which makes it easy to compare data both within a specific range or against other categories.



LEVEL OF SHAPE DETAIL

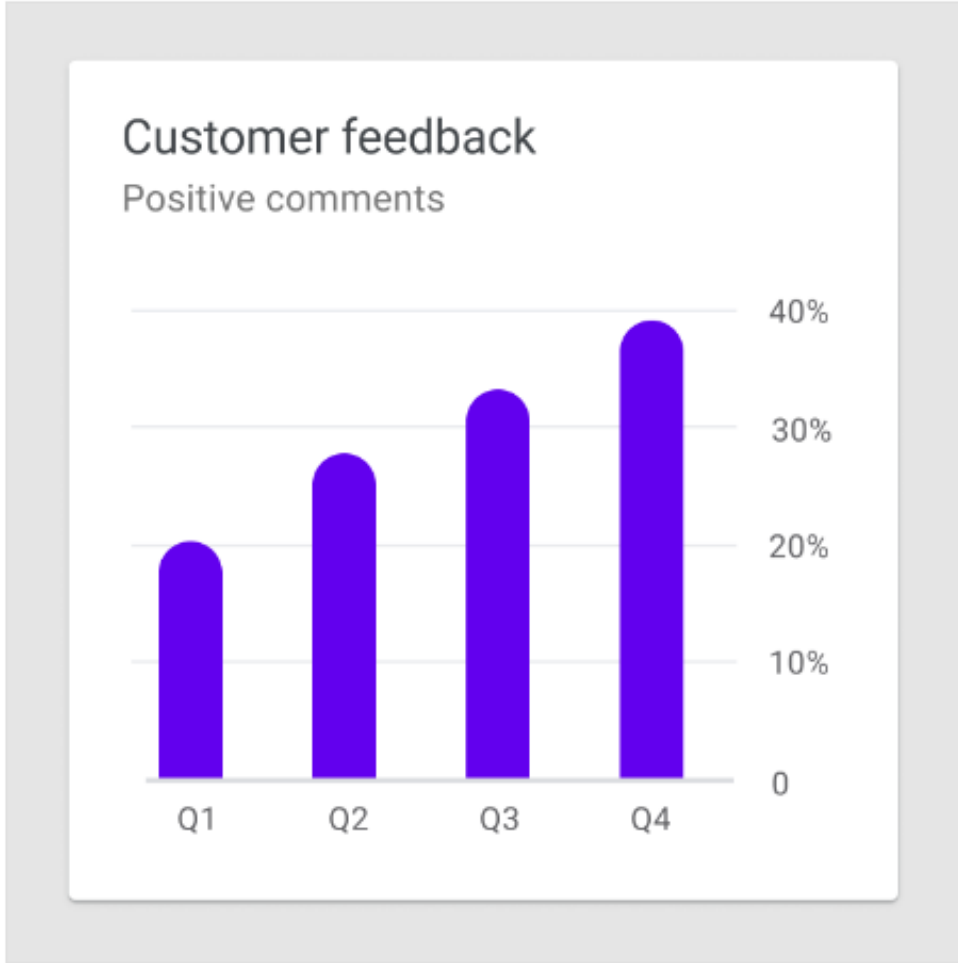
- Charts can use shapes to display data in a range of ways. A shape can be styled as playful and curvilinear, or precise and high-fidelity, among other ways in between.

- Charts can represent data at varying levels of precision. Data intended for close exploration should be represented by shapes that are suitable for interaction (in terms of touch target size and related affordances). Whereas data that's intended to express a general idea or trend can use shapes with less detail.



Do

The bars in this chart have subtle rounded corners, ensuring that the top of the bar precisely measures the bar's length.

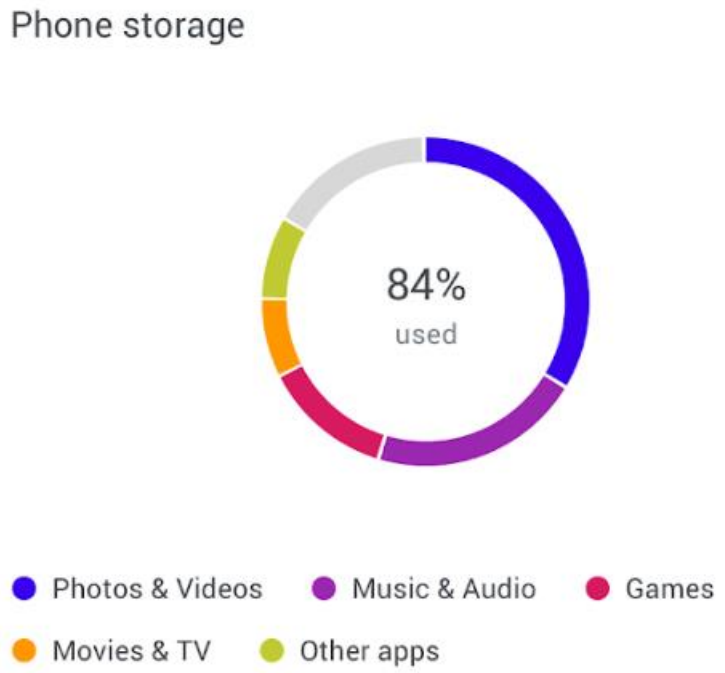


Don't

Don't use shapes that make it hard to read a chart, such as bars with imprecise top edges.

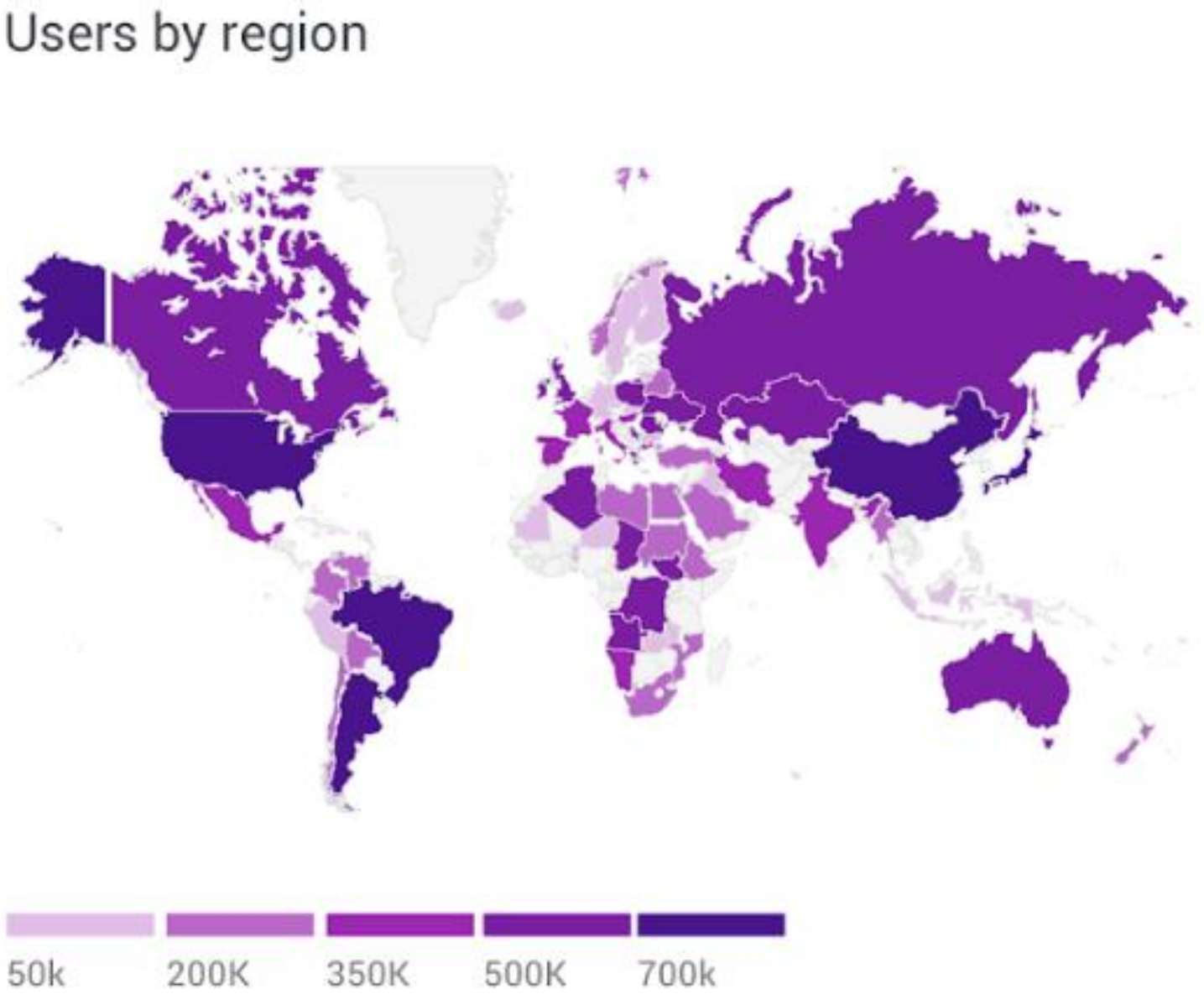
COLOR

- Color can be used to differentiate chart data in four primary ways:
  - Distinguishing categories from one another
  - Representing quantity
  - Highlighting specific data
  - Expressing meaning

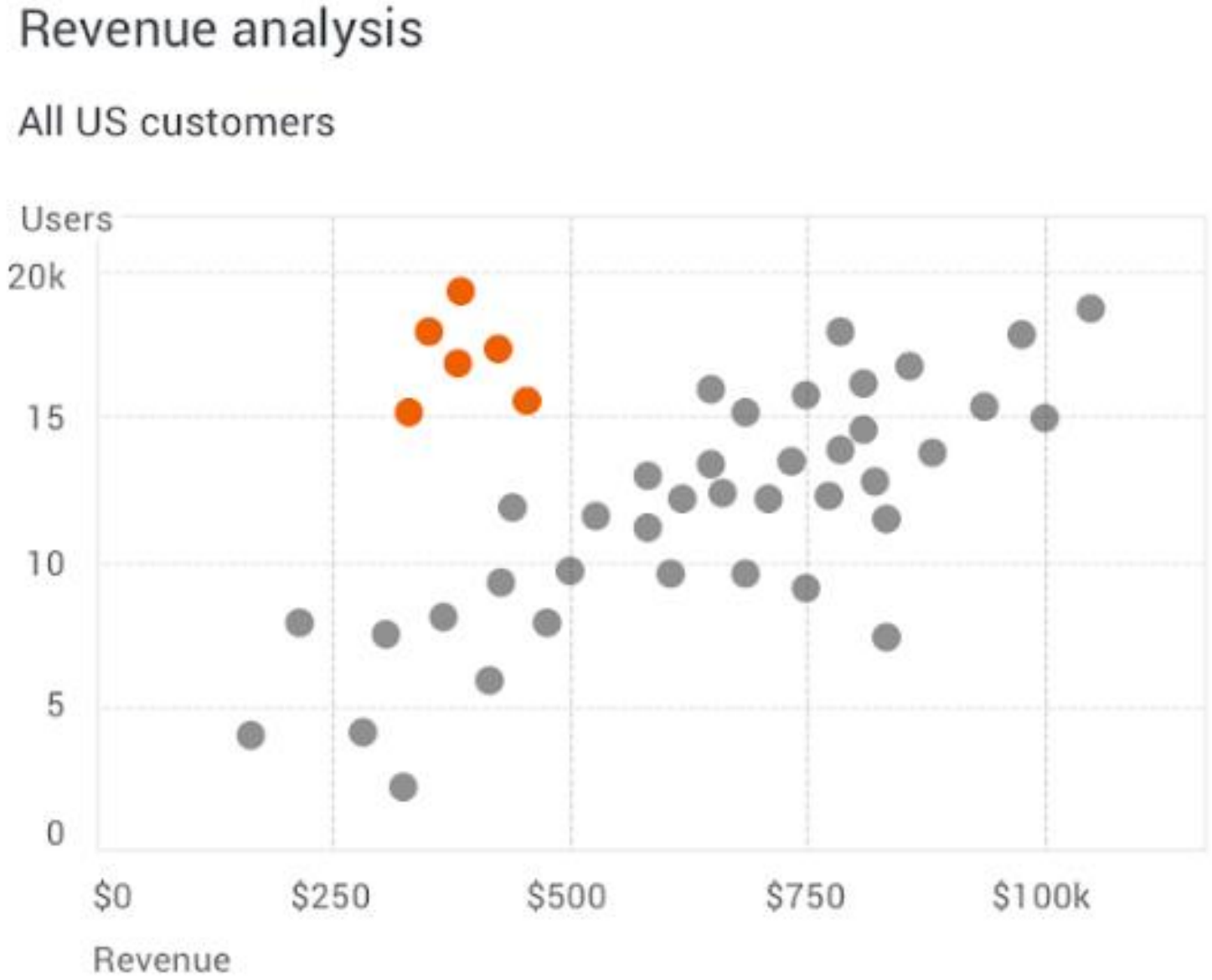


Color distinguishes categories

COLOR REPRESENTS QUANTITY



COLOR HIGHLIGHTS DATA



COLOR INDICATES MEANING

Do

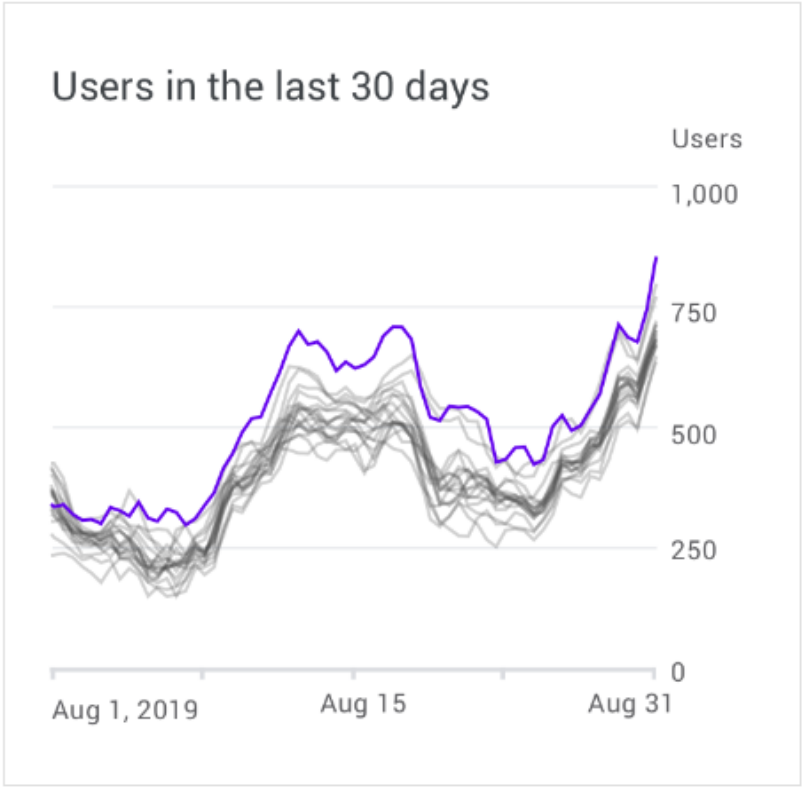
Reinforce the meaning of chart colors with other visual cues, like icons.

Don't

Don’t use color alone to indicate meaning.

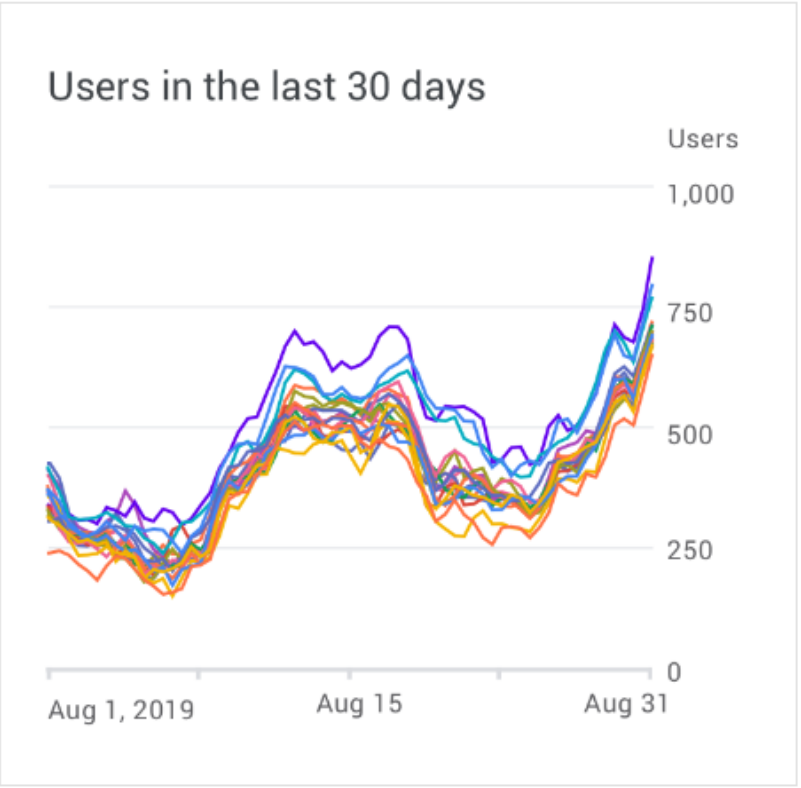
AREA OF FOCUS

- Color can highlight an area of focus, when it is used sparingly. It's not recommended to use a substantial amount of color highlights, as they can distract and hinder user focus.



Do

Use a combination of color highlights and neutral colors to provide contrast and emphasis.



Caution

Many colors in a single chart can hinder focus.

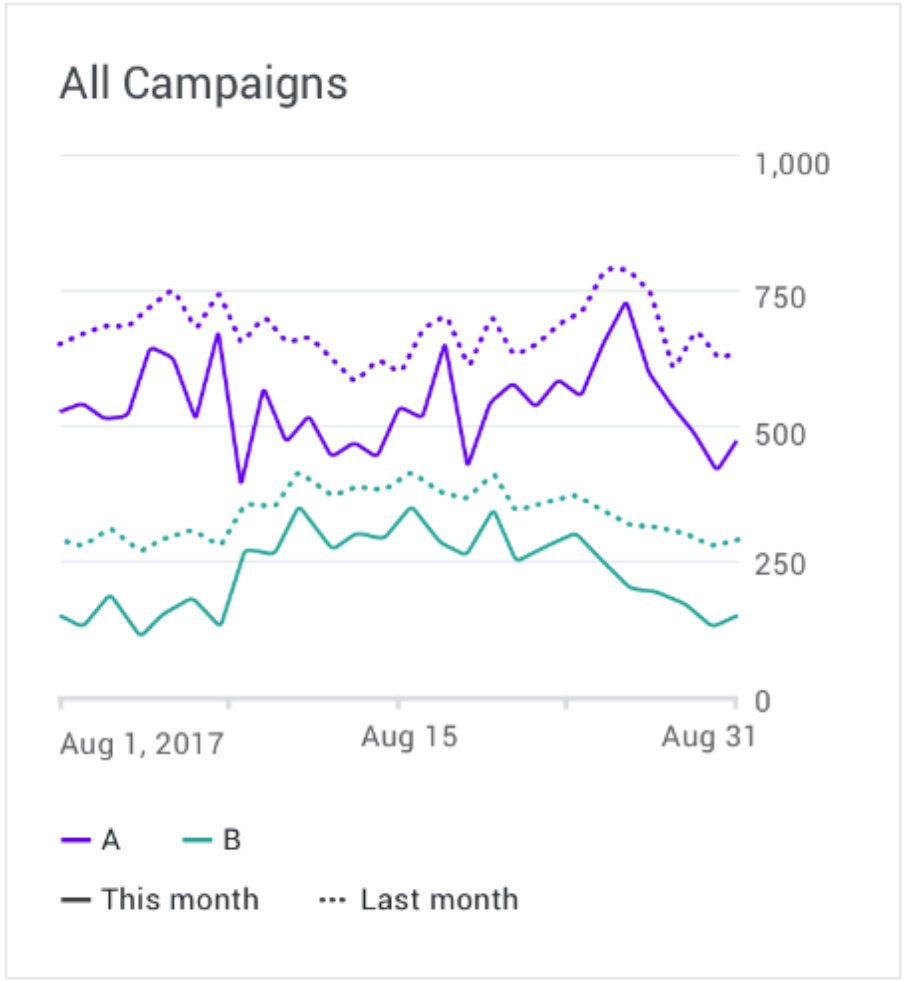
ACCESSIBILITY

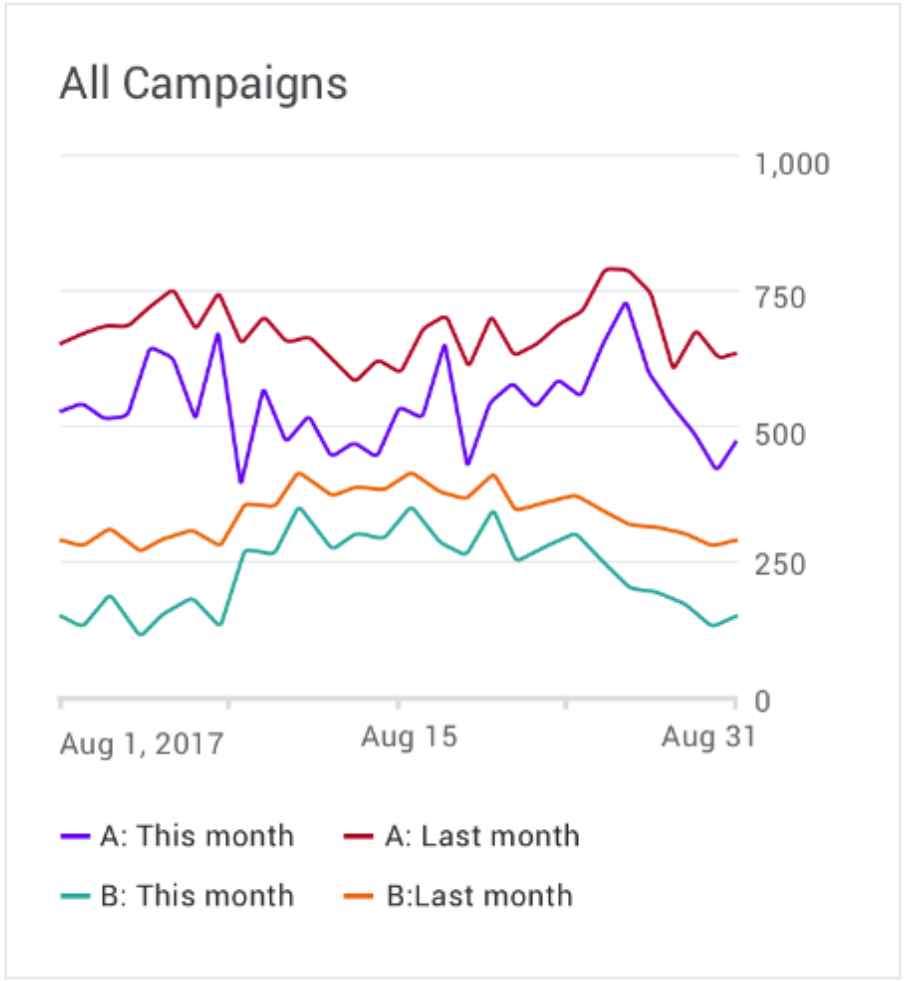
- To accommodate users who don't see color differences, you can use other methods to accentuate data, such as high-contrast shading, shape, or texture.

- Applying text labels to data also helps clarify its meaning, while eliminating the need for a legend.

LINE

- Chart lines can express qualities about data, such as hierarchy, highlights, and comparisons. Line styles can be styled in different ways, such as using dashes or varied opacities.
- Lines can be applied to specific elements, including:
  - Annotations
  - Forecasting elements
  - Comparative tools
  - Confidence intervals
  - Anomalies





Do

Vary a line's texture to represent different data types.

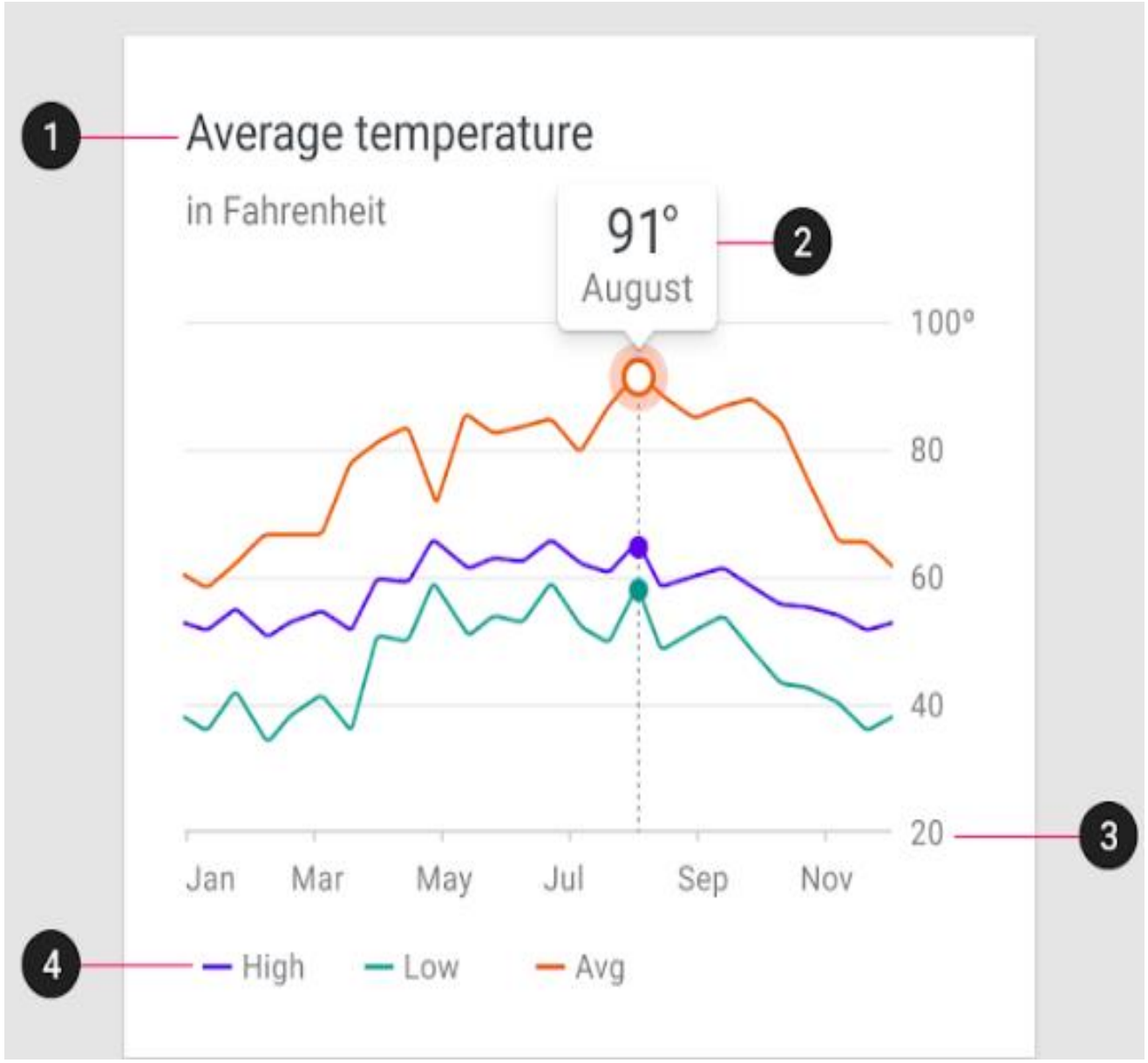
Don't

Don't use different colors to show periodical variation for the same data category.

TYPOGRAPHY

- Text can be used to label different chart elements, including:
  - Chart titles
  - Data labels
  - Axis labels
  - Legend

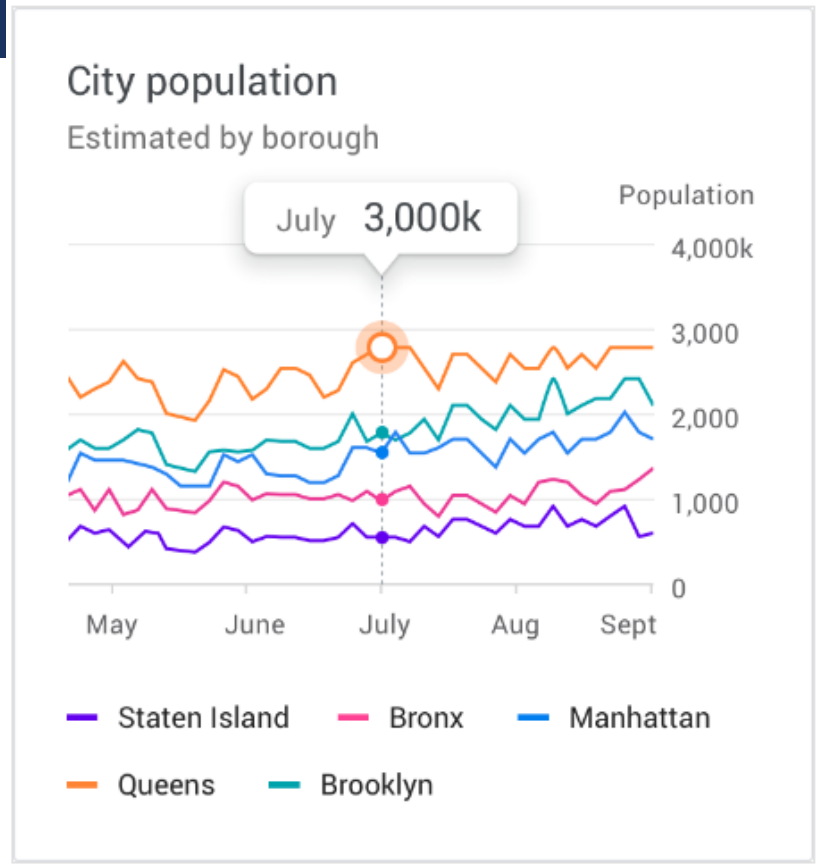
- The text with the highest level of hierarchy is usually the chart title, with axis labels and the legend having the lowest level of hierarchy.



Scale category
1. Chart title
Chart subtitle
2. Data label
Sub-label
3. Axis labels
4. Legend labels

TEXT WEIGHT

- Headings and varying font weights can communicate which content is more (or less) important than other content in the hierarchy. However, these treatments should be used sparingly, with a limited number of typographic styles.



Do

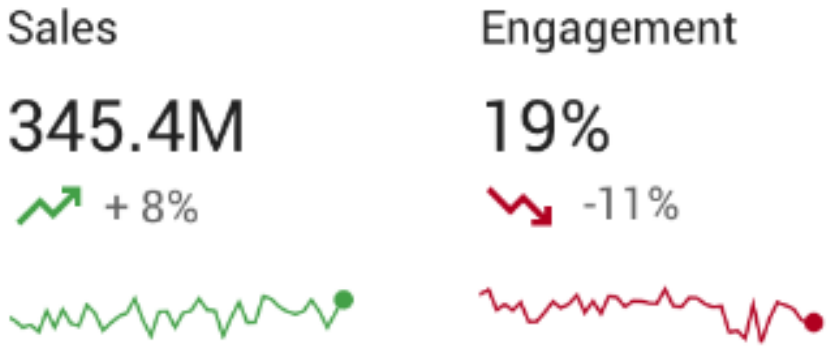
Bold used just for one or two key elements creates a balanced design.

Don't

Bold used on too many elements can make it harder to identify important elements.

ICONOGRAPHY

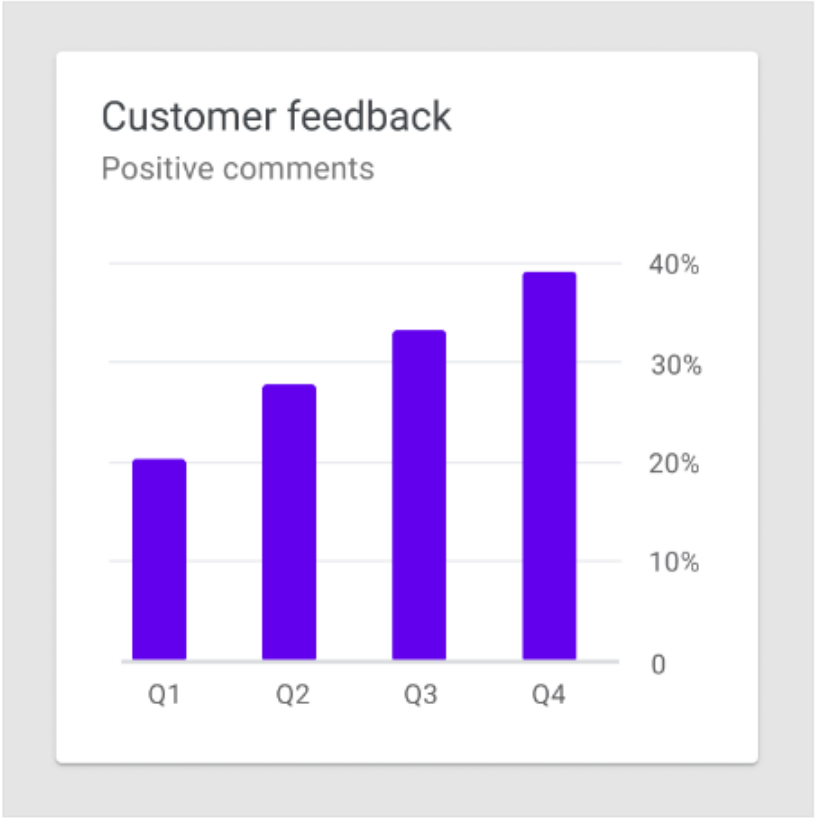
- Iconography can represent different types of data in a chart and improve a chart's overall usability.
- Iconography can be used for:
  - **Categorical data** to differentiate groups or categories
  - **States**, such as errors, no data, completed states, and danger
- When placing icons in a chart, it's recommended to use universally recognizable symbols, particularly when representing actions or states, such as: completed, error, and danger.



Icons supplement color in conveying meaning

LABELLED AXIS

A labelled axis, or multiple axes, indicates the scale and scope of the data displayed. For example, line charts display a range of values along both horizontal and vertical labelled axes.

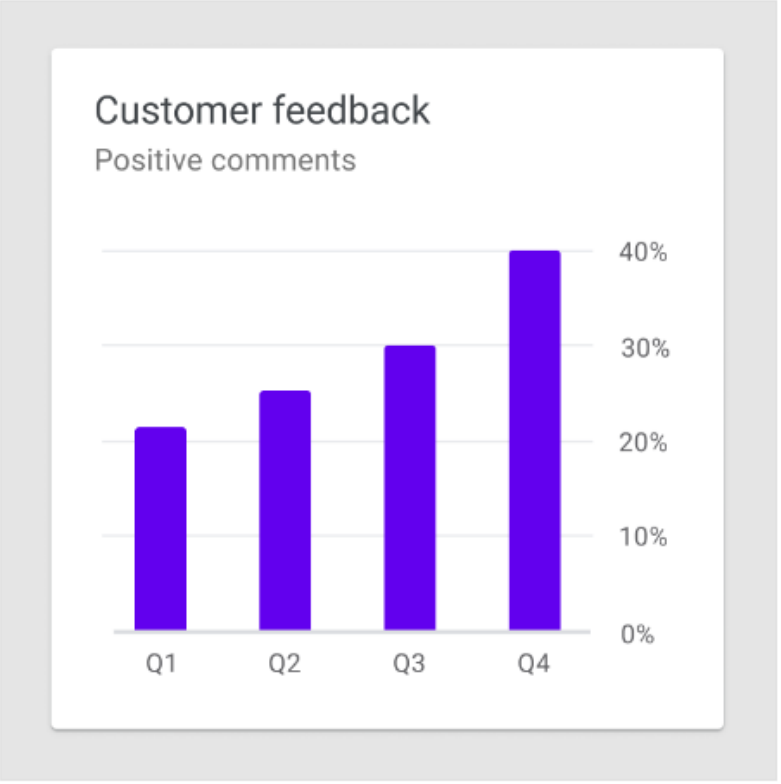


Bar charts should always start at the x-axis baseline value of zero.

Consider removing axes altogether to keep the data as the focal point. You can place values directly on their corresponding chart elements.

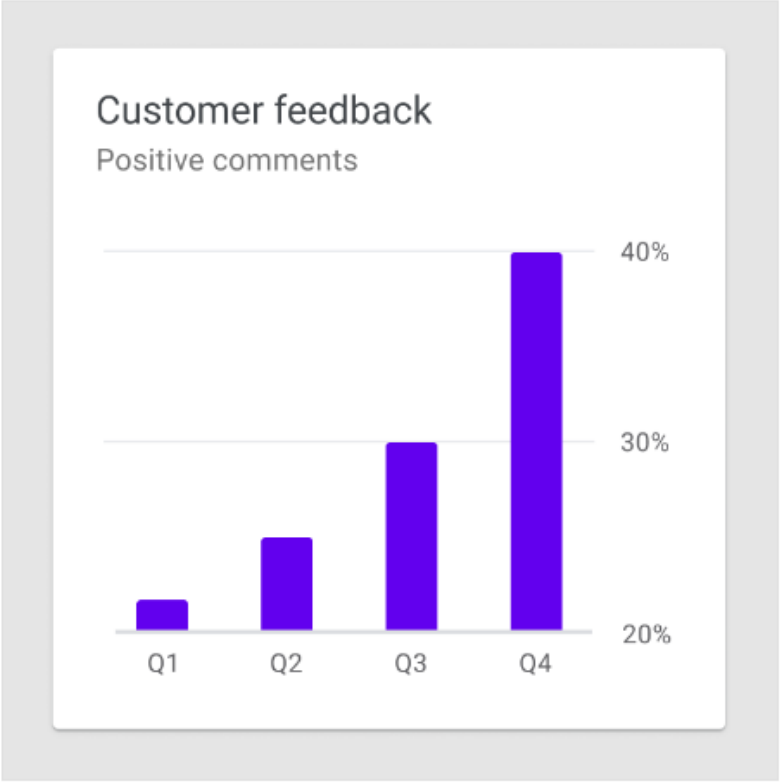
BAR CHART BASELINE

Bar charts should start at a baseline (the starting value on the y-axis) of zero. Starting at a baseline that isn't zero can cause the data to be perceived incorrectly.



Do

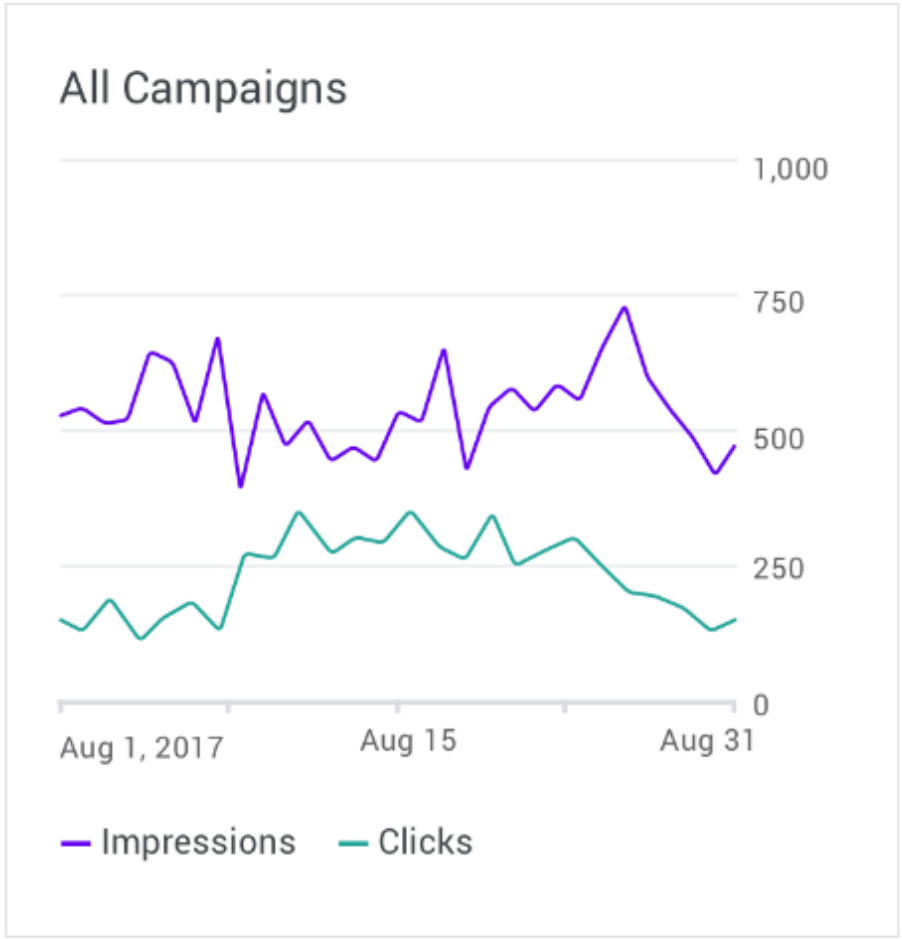
A bar chart starting at the zero baseline



Don't

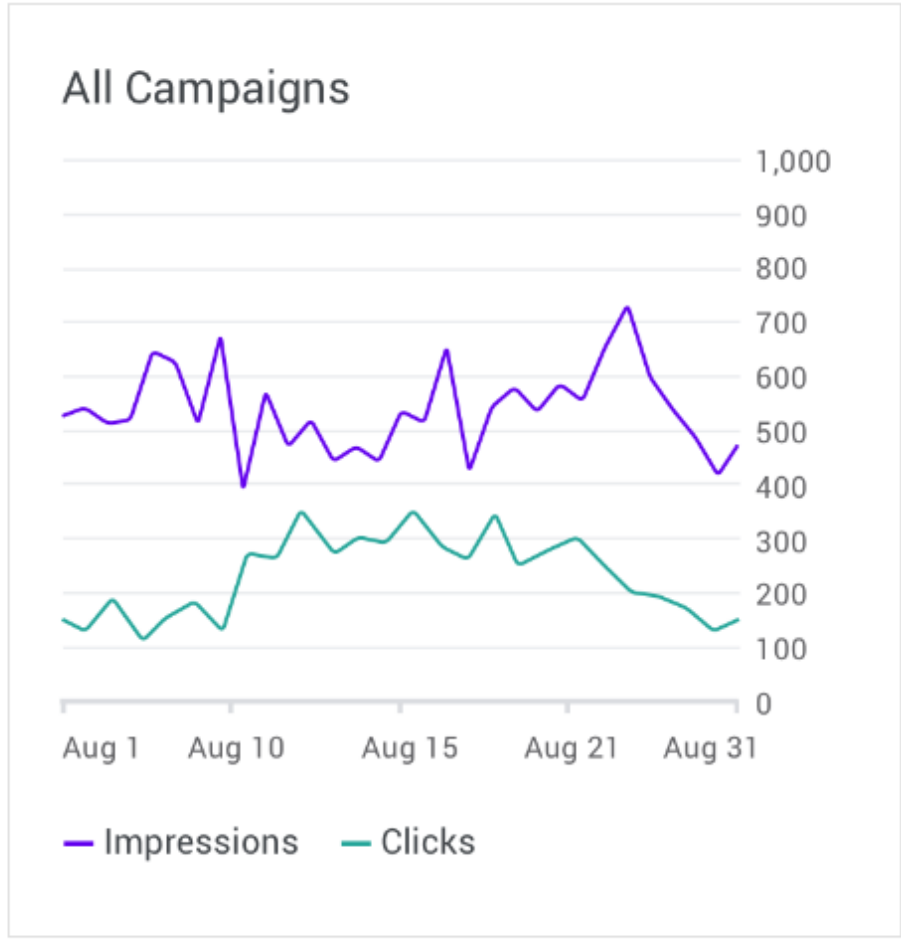
Don't start the baseline at values other than zero. This baseline starts at 20%, making the bar differences look more dramatic.

AXIS LABELS



Do

Support legibility by using a balanced number of axis labels.



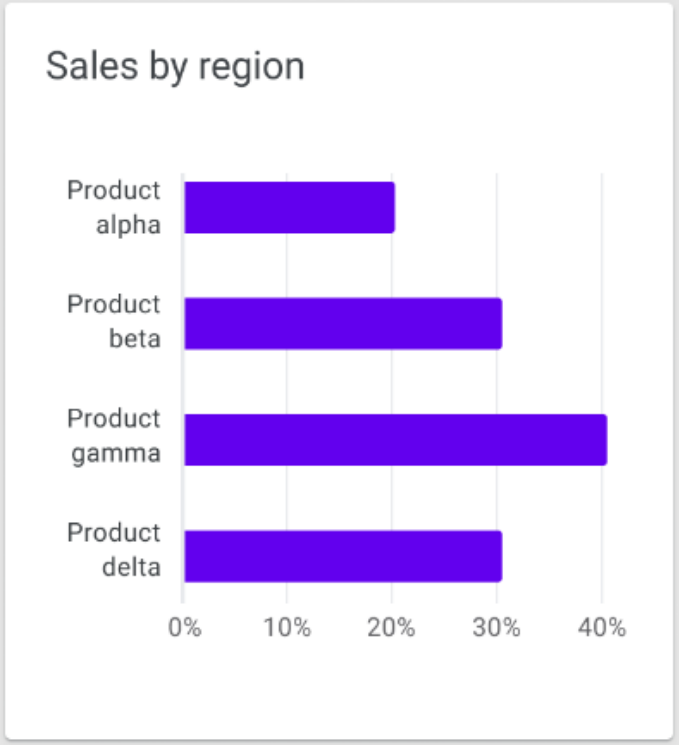
Don't

Don't overload the chart with numerous axis labels.

- Label usage should reflect the most important data insights in a chart. Axis labels should be used as needed, and in consistent ways across a UI. Their presence should not inhibit reading the chart.

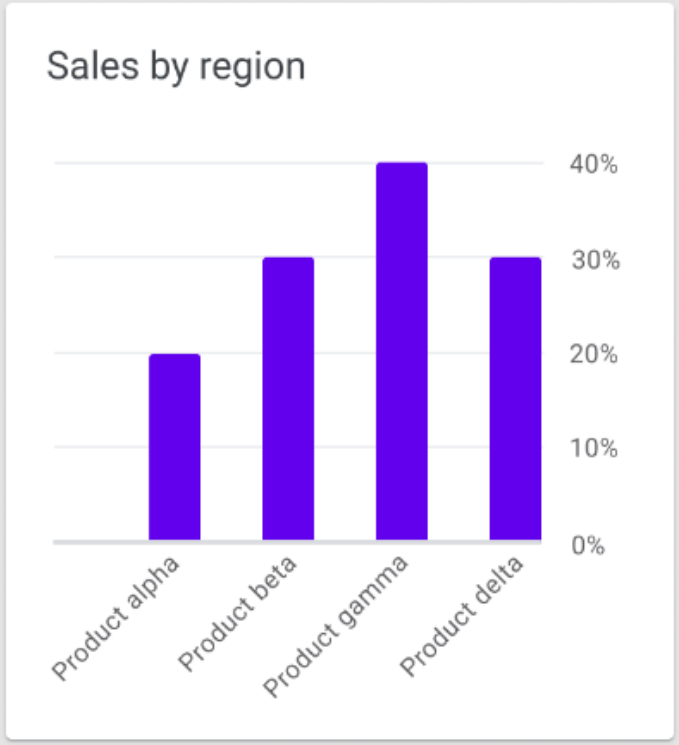
TEXT ORIENTATION

- Text labels should be placed horizontally on the chart so that they are easy to read.
- Text labels should not:
  - Be rotated
  - Stacked vertically



Do

Orient text horizontally on bar charts, rotating the bars if needed to make space.



Caution

Don't rotate bar labels, as it makes them difficult to read.

LEGENDS AND ANNOTATIONS

- Legends and annotations describe a chart's information. Annotations should highlight data points, data outliers, and any noteworthy content.