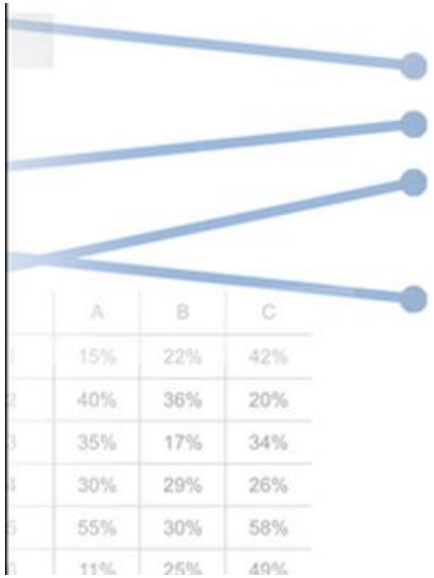
Lesson 10: Data Visualization Foundations & Design Principles

Lecturer: **Dr. Nguyen Tuan Long**

Email: ntlong@neu.edu.vn

Mobile: 0982 746 235

How To Design Charts And Graphs

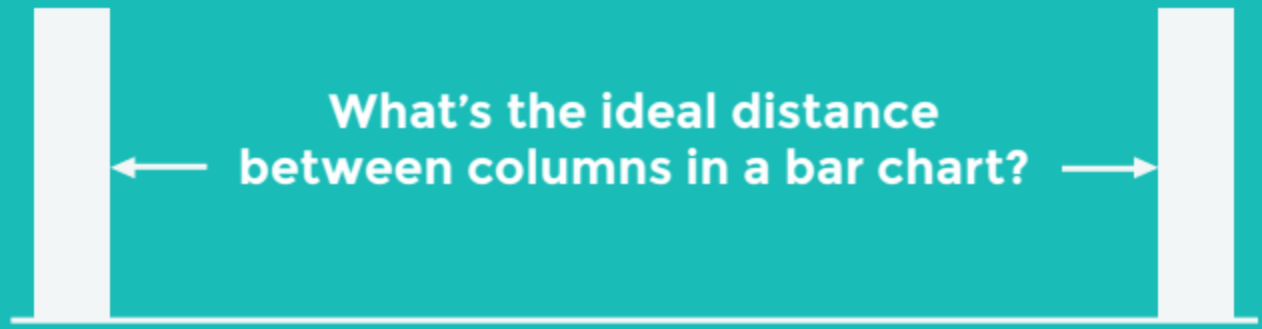


storytelling with data

Why Visualize?

Your data is only as good as your ability to understand and communicate it, which is why choosing the right visualization is essential.

If your data is misrepresented or presented ineffectively, key insights and understanding are lost, which hurts both your message and your reputation. The good news is that you don't need a PhD in statistics to crack the data visualization code. This guide will walk you through the most common charts and visualizations, help you choose the right presentation for your data, and give you practical design tips and tricks to make sure you avoid rookie mistakes. It's everything you need to help your data make a big impact.



You're about to find out.

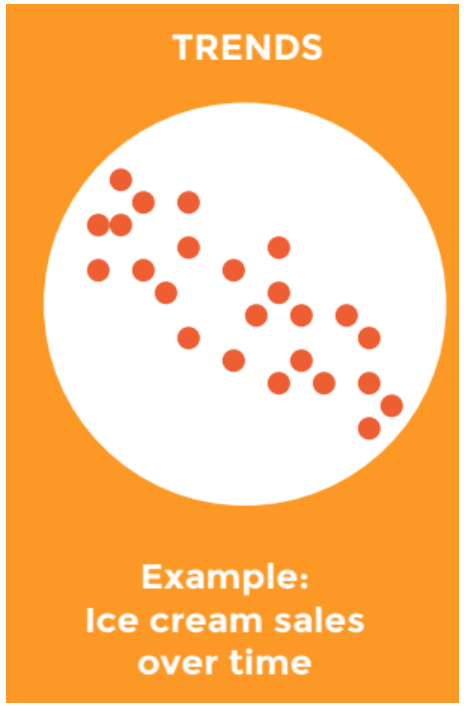
What is Your Purpose?

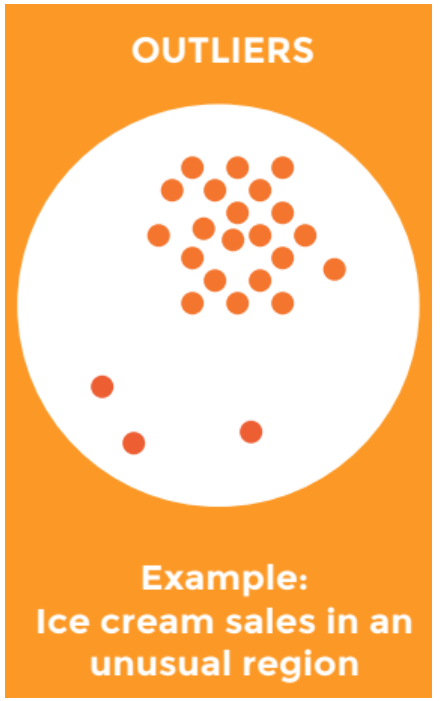
Exploratory vs. Explanatory Analysis

- **Exploratory:** This is when you are "hunting" for insights (Trends, Correlations...). You are doing it for **yourself**.
- **Explanatory:** This is when you *already know* the insight and are communicating it to others. You are doing it for your **audience**.

Finding The Story In Your Data

- Information can be visualized in a number of ways, each of which can provide a specific insight.
- When you start to work with your data, it's important to identify and understand the story you are trying to tell and the relationship you are looking to show.
- Knowing this information will help you select the proper visualization to best deliver your message.
- When analyzing data, search for patterns or interesting insights that can be a good starting place for finding your story, such as:





Know Your Data

- Before understanding visualizations, you must understand the types of data that can be visualized and their relationships to each other.
- Here are some of the most common you are likely to encounter.

DATA TYPES

QUANTITATIVE
Data that can be counted or measured; all values are numerical.

DISCRETE
Numerical data that has a finite number of possible values. Example: Number of employees in the office.

CONTINUOUS
Data that is measured and has a value within a range. Example: Rainfall in a year.

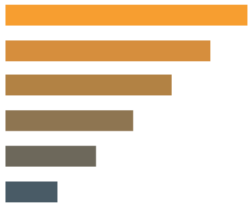
CATEGORICAL
Data that can be sorted according to group or category. Example: Types of products sold.

Data Relationships

NOMINAL COMPARISON
This is a simple comparison of the quantitative values of subcategories. Example: Number of visitors to various websites.

TIME-SERIES
This tracks changes in values of a consistent metric over time. Example: Monthly sales.

CORRELATION
This is data with two or more variables that may demonstrate a positive or negative correlation to each other. Example: Salaries according to education level.



RANKING
This shows how two or more values compare to each other in relative magnitude. Example: Historic weather patterns, ranked from the hottest months to the coldest.

DEVIATION
This examines how data points relate to each other, particularly how far any given data point differs from the mean. Example: Amusement park tickets sold on a rainy day vs. a regular day.

DISTRIBUTION
This shows data distribution, often around a central value. Example: Heights of players on a basketball team.

PART-TO-WHOLE RELATIONSHIPS
This shows a subset of data compared to the larger whole. Example: Percentage of customers purchasing specific products.

Now that you've got a handle on the most common data types and relationships you'll most likely have to work with, let's dive into the different ways you can visualize that data to get your point across.

Guide To Chart Types

GUIDE TO CHART TYPES

In this section, we'll cover the uses, variations, and best practices for some of the most common data visualizations:

BAR CHART

PIE CHART

LINE CHART

AREA CHART

SCATTER PLOT

BUBBLE CHART

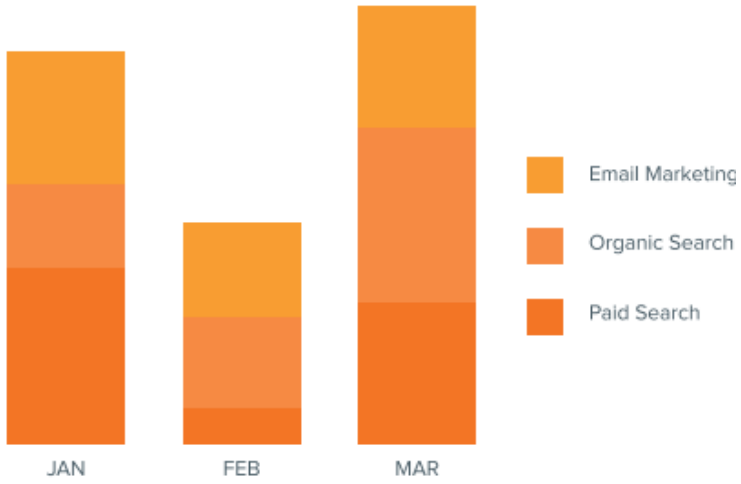
HEAT MAP

Bar Chart

Bar charts are very versatile. They are best used to show change over time, compare different categories, or compare parts of a whole.

VARIATIONS OF BAR CHARTS

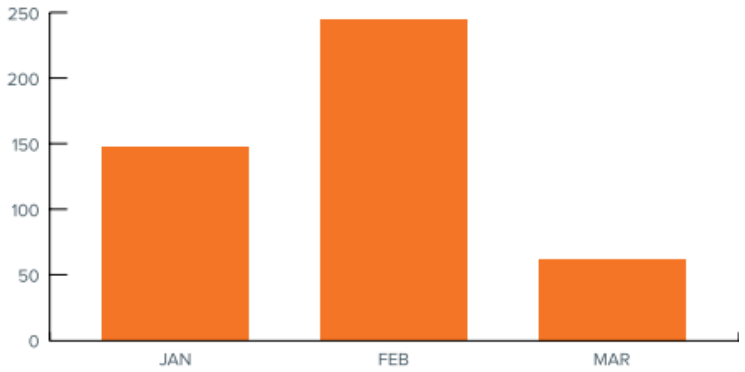
MONTHLY TRAFFIC, BY SOURCE



STACKED

Best used when there is a need to compare multiple part-to-whole relationships. These can use discrete or continuous data, oriented either vertically or horizontally.

PAGE VIEWS, BY MONTH



VERTICAL (COLUMN CHART)

Best used for chronological data (time-series should always run left to right), or when visualizing negative values below the x-axis.

CONTENT PUBLISHED, BY CATEGORY

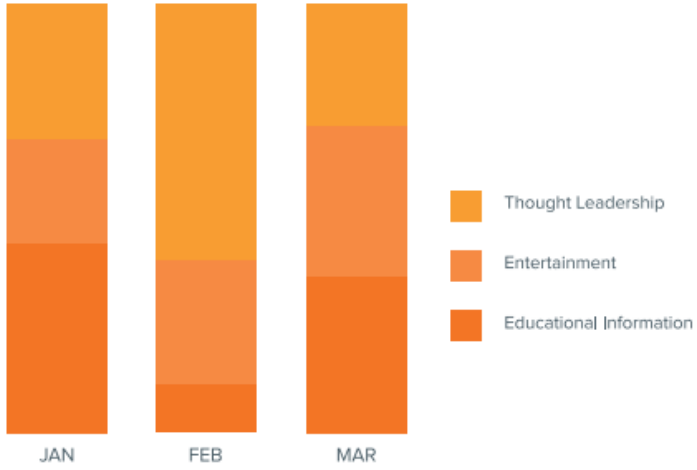
HORIZONTAL

Best used for data with long category labels.

PERCENTAGE OF CONTENT PUBLISHED, BY MONTH

100% STACKED

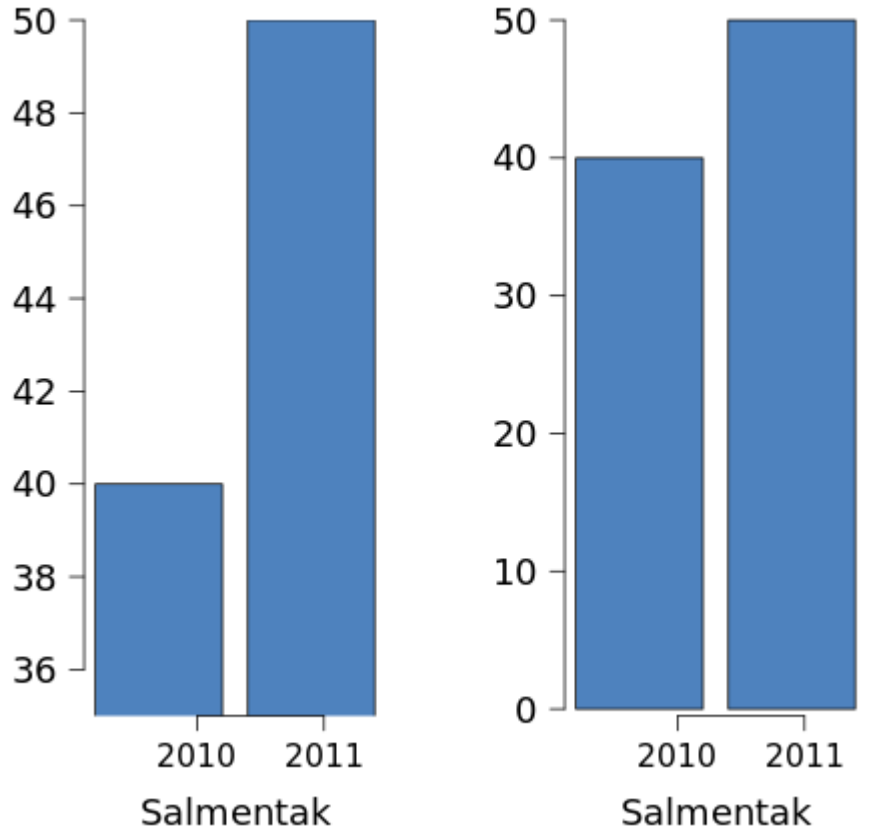
Best used when the total value of each category is unimportant and percentage distribution of subcategories is the primary message.



Why Bar Charts Work?

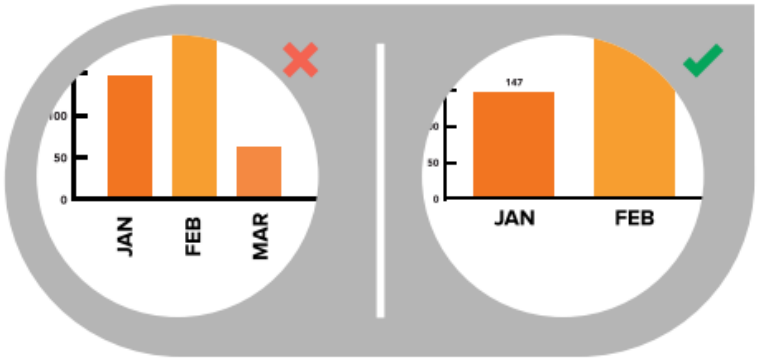
Humans are Great at Reading "Length"

- **Theory:** Bar charts are effective because our brains are excellent at quickly and accurately comparing **lengths**.
- **The Rule:** The Y-axis (quantitative axis) **MUST** start at zero.

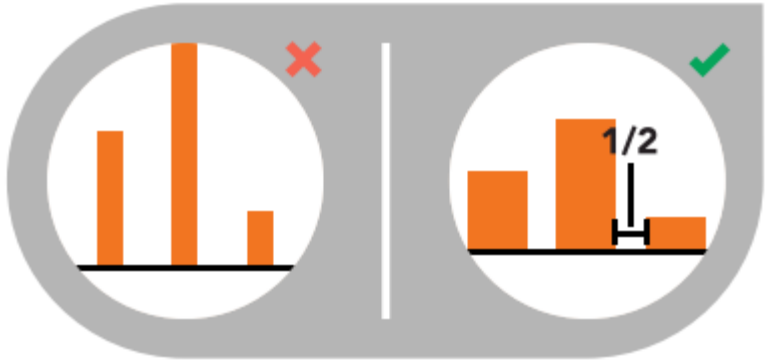


Bar Chart

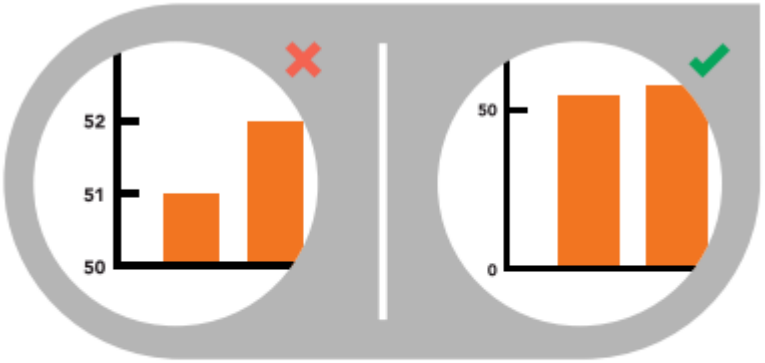
DESIGN BEST PRACTICES



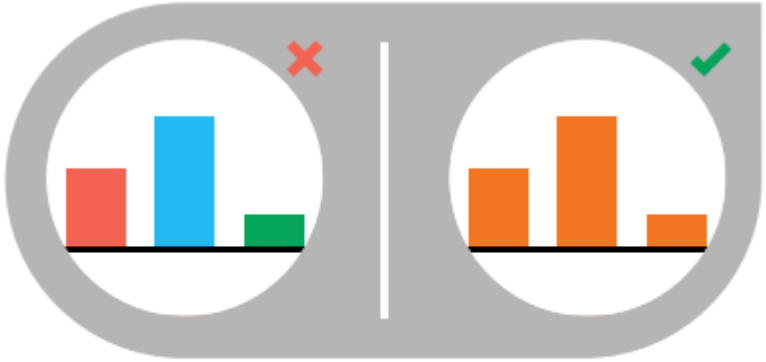
USE HORIZONTAL LABELS
Avoid steep diagonal or vertical type, as it can be difficult to read.



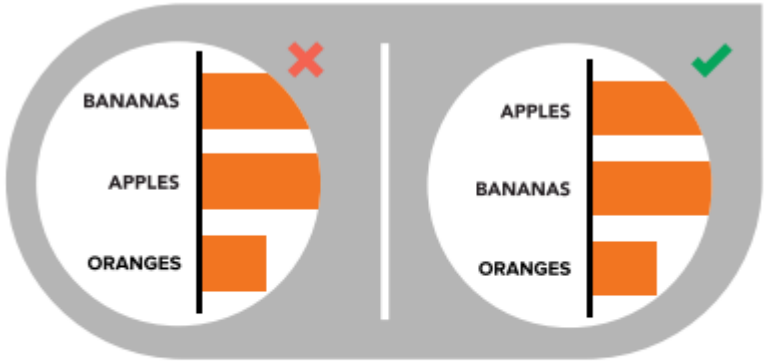
SPACE BARS APPROPRIATELY
Space between bars should be ½ bar width.



START THE Y-AXIS VALUE AT 0
Starting at a value above zero truncates the bars and doesn't accurately reflect the full value.



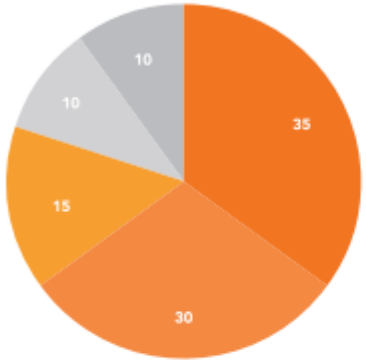
USE CONSISTENT COLORS
Use one color for bar charts. You may use an accent color to highlight a significant data point.



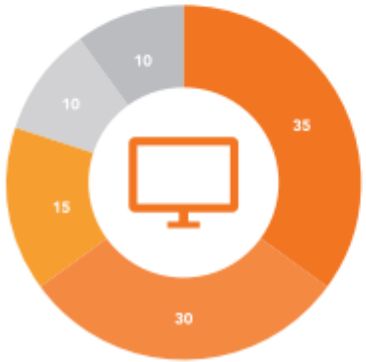
ORDER DATA APPROPRIATELY
Order categories alphabetically, sequentially, or by value.

Pie Chart

VARIATIONS OF PIE CHARTS



STANDARD
Used to show part-to-whole relationships.



DONUT
Stylistic variation that enables the inclusion of a total value or design element in the center.

A Warning on Pie Charts!

Use Pie Charts with Caution

- **The Problem:** Humans are terrible at accurately comparing **angles** and **areas**.
- **Example (Image):** "Can you quickly tell which slice is biggest? It's difficult."
- **The Alternatives :**
  - **Horizontal Bar Chart (Sorted):** Always easier for comparison.
  - **When to use:** Only for very few categories (max 3-4) or simple fractions (like 25%, 50%).
- **NEVER USE:** 3D Pie Charts (they severely distort perception).

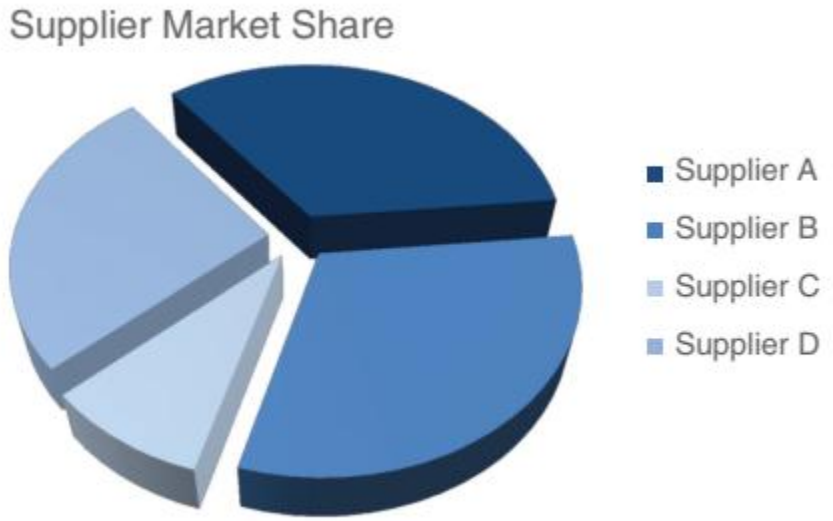
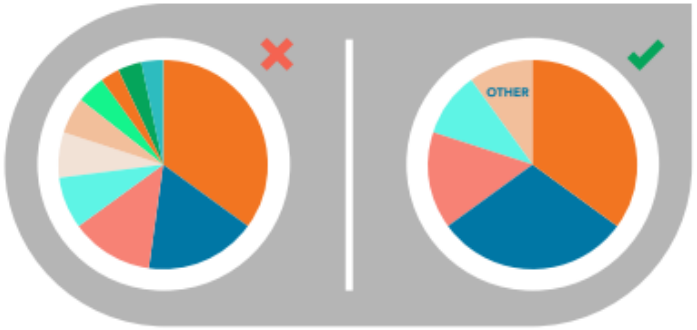


FIGURE 2.21 Pie chart

Pie Chart

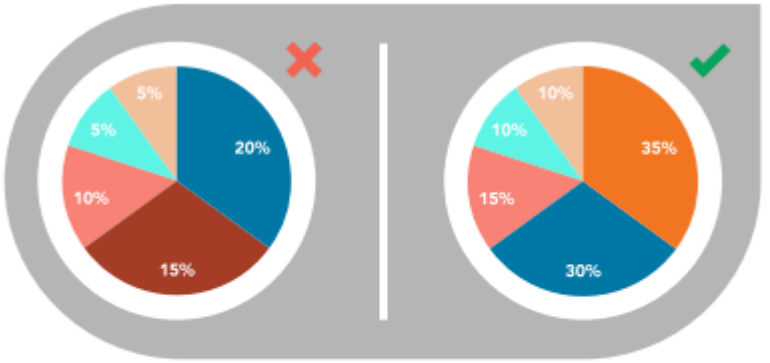


VISUALIZE NO MORE THAN 5 CATEGORIES PER CHART

It is difficult to differentiate between small values; depicting too many slices decreases the impact of the visualization. If needed, you can group smaller values into an “other” or “miscellaneous” category, but make sure it does not hide interesting or significant information.

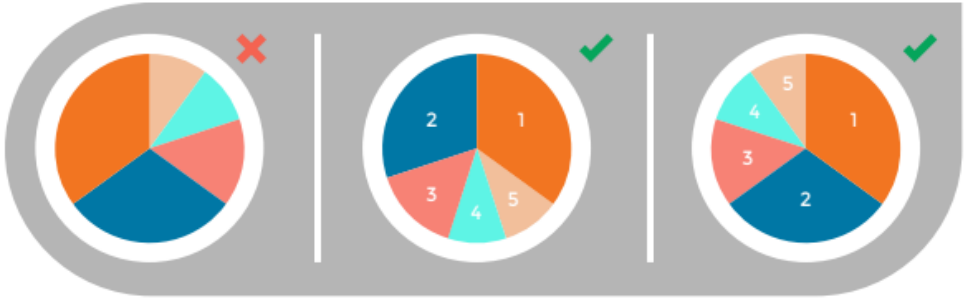
DON’T USE MULTIPLE PIE CHARTS FOR COMPARISON

Slice sizes are very difficult to compare side-by-side. Use a stacked bar chart instead.



MAKE SURE ALL DATA ADDS UP TO 100%

Verify that values total 100% and that pie slices are sized proportionate to their corresponding value.



ORDER SLICES CORRECTLY

There are two ways to order sections, both of which are meant to aid comprehension:

OPTION 1

Place the largest section at 12 o’clock, going clockwise. Place the second largest section at 12 o’clock, going counterclockwise. The remaining sections can be placed below, continuing counterclockwise.

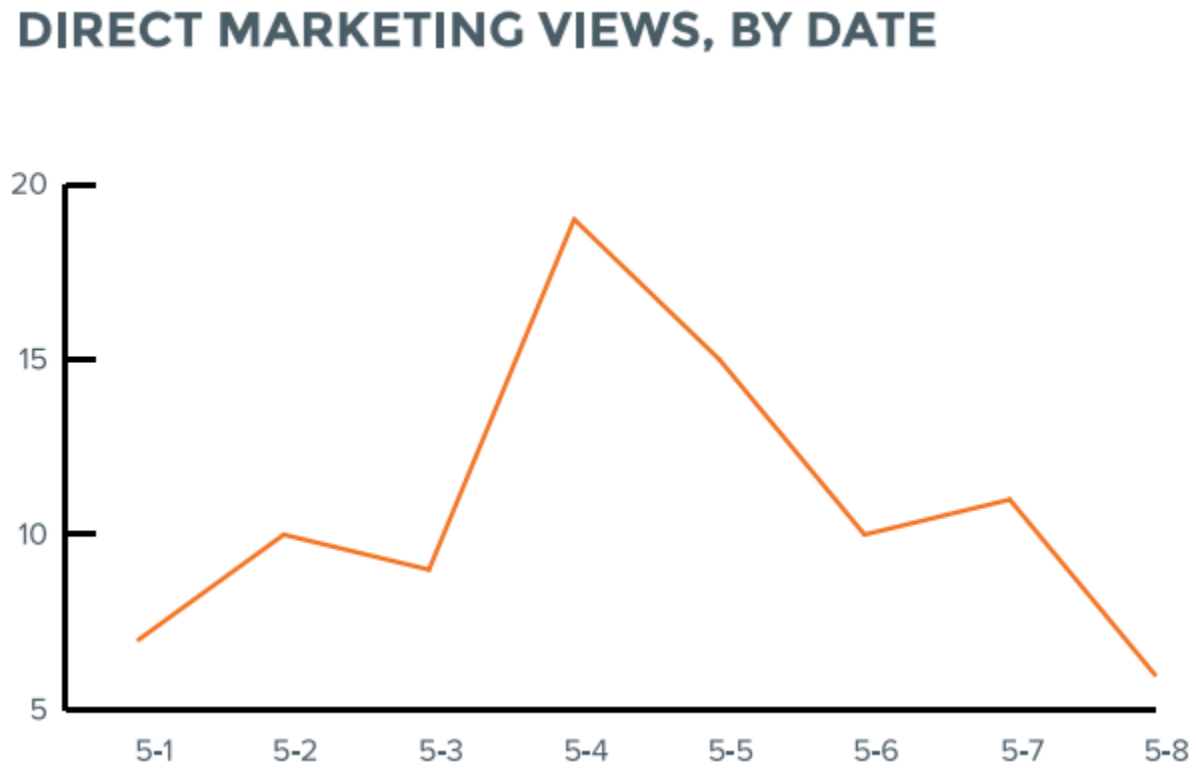
OPTION 2

Start the largest section at 12 o’clock, going clockwise. Place remaining sections in descending order, going clockwise.

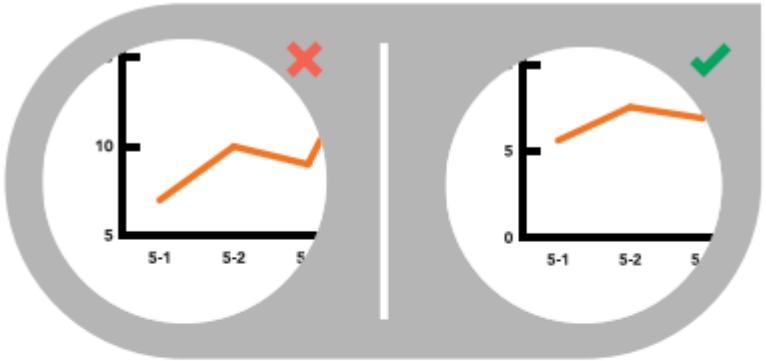
PIE CHART

DESIGN BEST PRACTICES

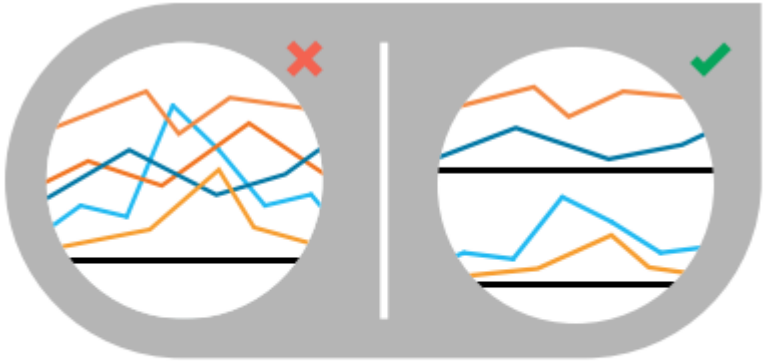
Line Chart



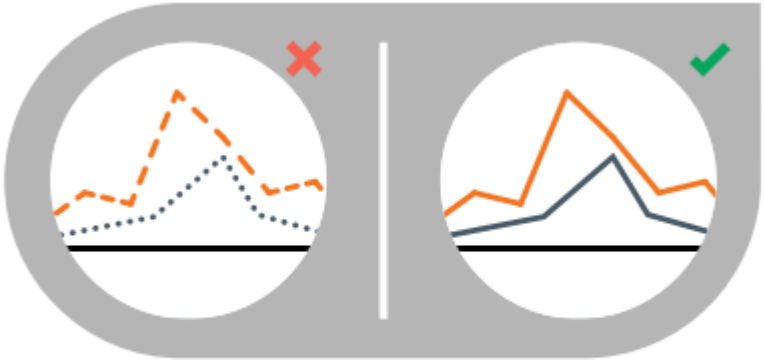
Line Chart



INCLUDE A ZERO BASELINE IF POSSIBLE
Although a line chart does not have to start at a zero baseline, it should be included if possible. If relatively small fluctuations in data are meaningful (e.g., in stock market data), you may truncate the scale to showcase these variances.



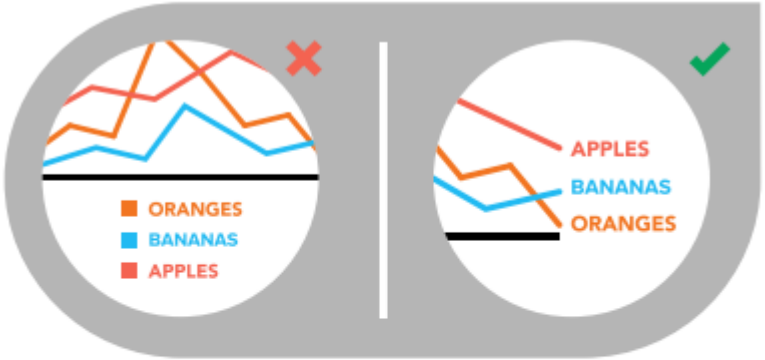
DON'T PLOT MORE THAN 4 LINES
If you need to display more, break them out into separate charts for better comparison.



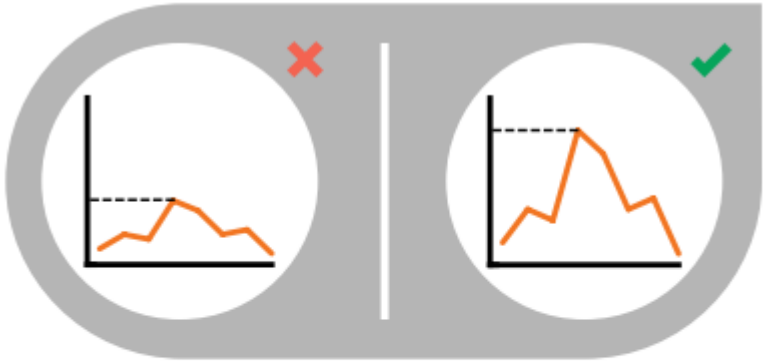
USE SOLID LINES ONLY
Dashed and dotted lines can be distracting.

LINE CHART

DESIGN BEST PRACTICES



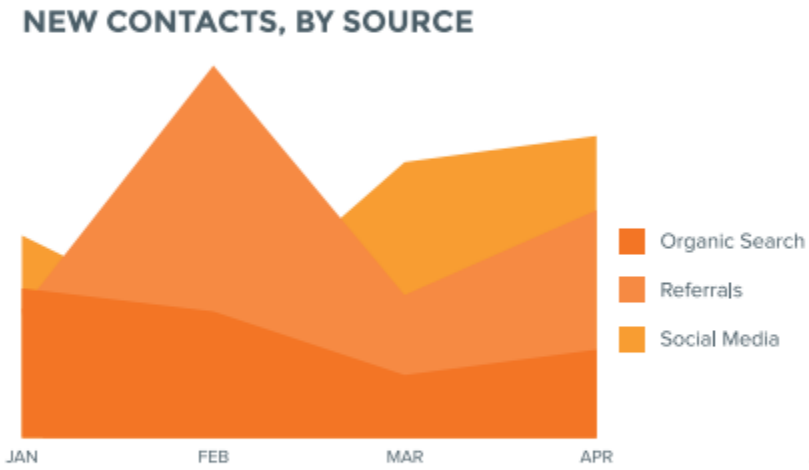
LABEL THE LINES DIRECTLY
This lets readers quickly identify lines and corresponding labels instead of referencing a legend.



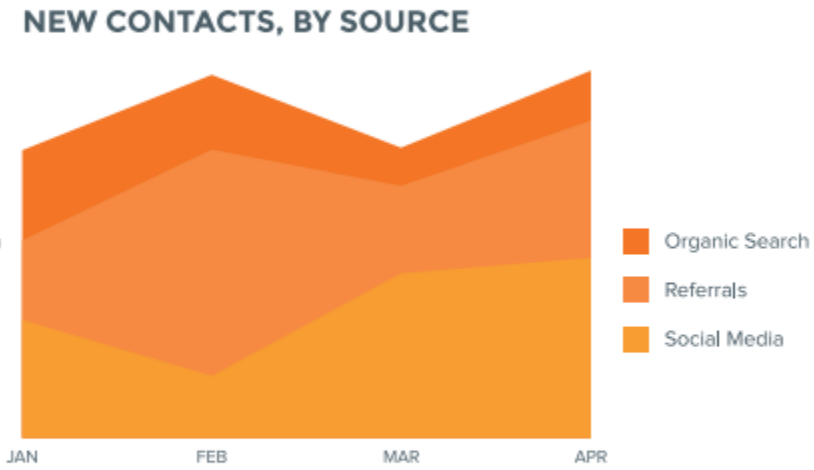
USE THE RIGHT HEIGHT
Plot all data points so that the line chart takes up approximately two-thirds of the y-axis' total scale.

Area Chart

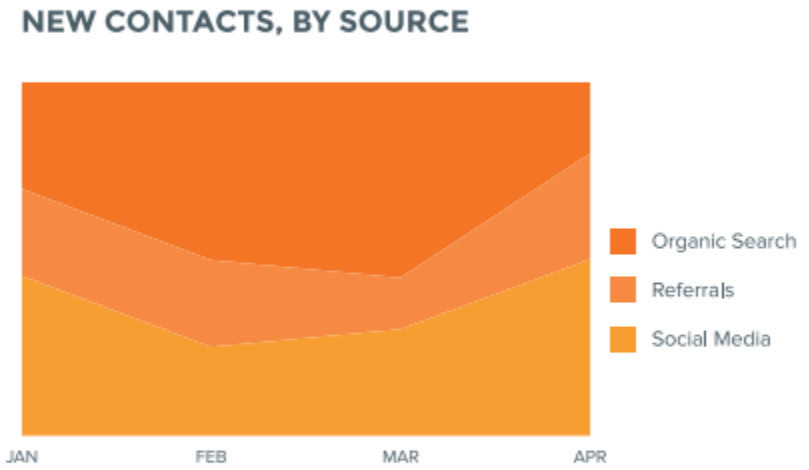
VARIATIONS OF AREA CHARTS



AREA CHART
Best used to show or compare a quantitative progression over time.



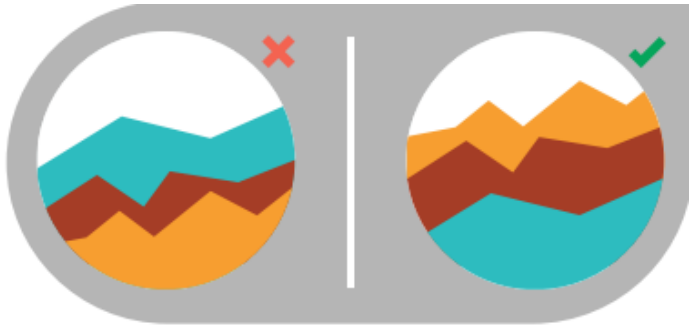
STACKED AREA
Best used to visualize part-to-whole relationships, helping show how each category contributes to the cumulative total.



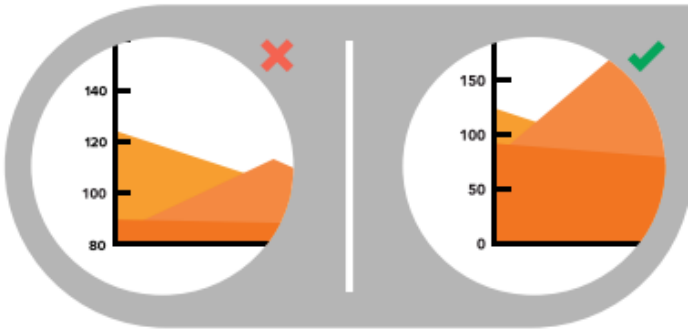
100% STACKED AREA
Best used to show distribution of categories as part of a whole, where the cumulative total is unimportant.

Area Chart

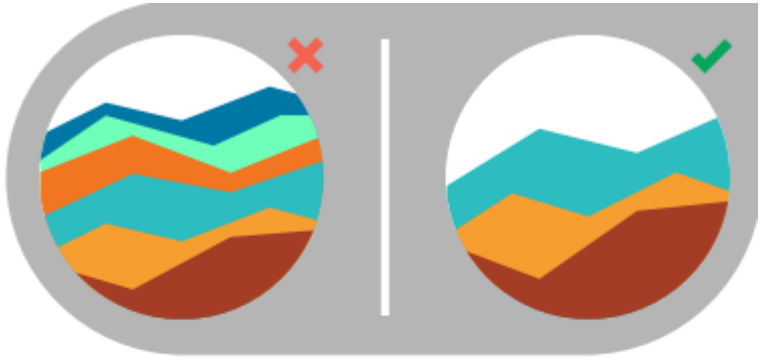
DESIGN BEST PRACTICES



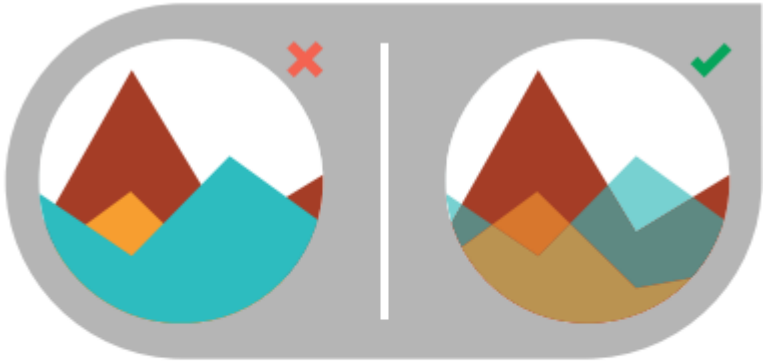
MAKE IT EASY TO READ
In stacked area charts, arrange data to position categories with highly variable data on the top of the chart and low variability on the bottom.



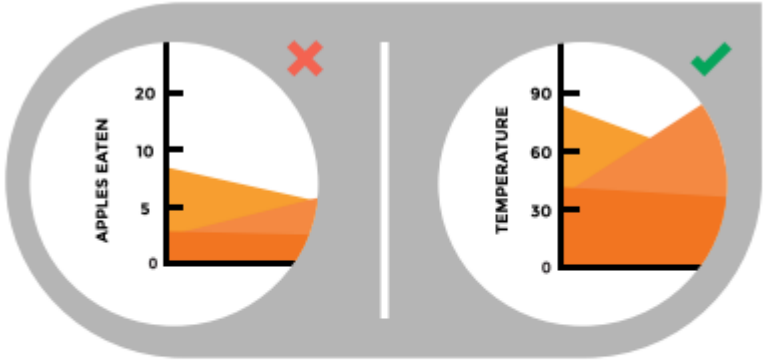
START Y-AXIS VALUE AT 0
Starting the axis above zero truncates the visualization of values.



DON'T DISPLAY MORE THAN 4 DATA CATEGORIES
Too many will result in a cluttered visual that is difficult to decipher.



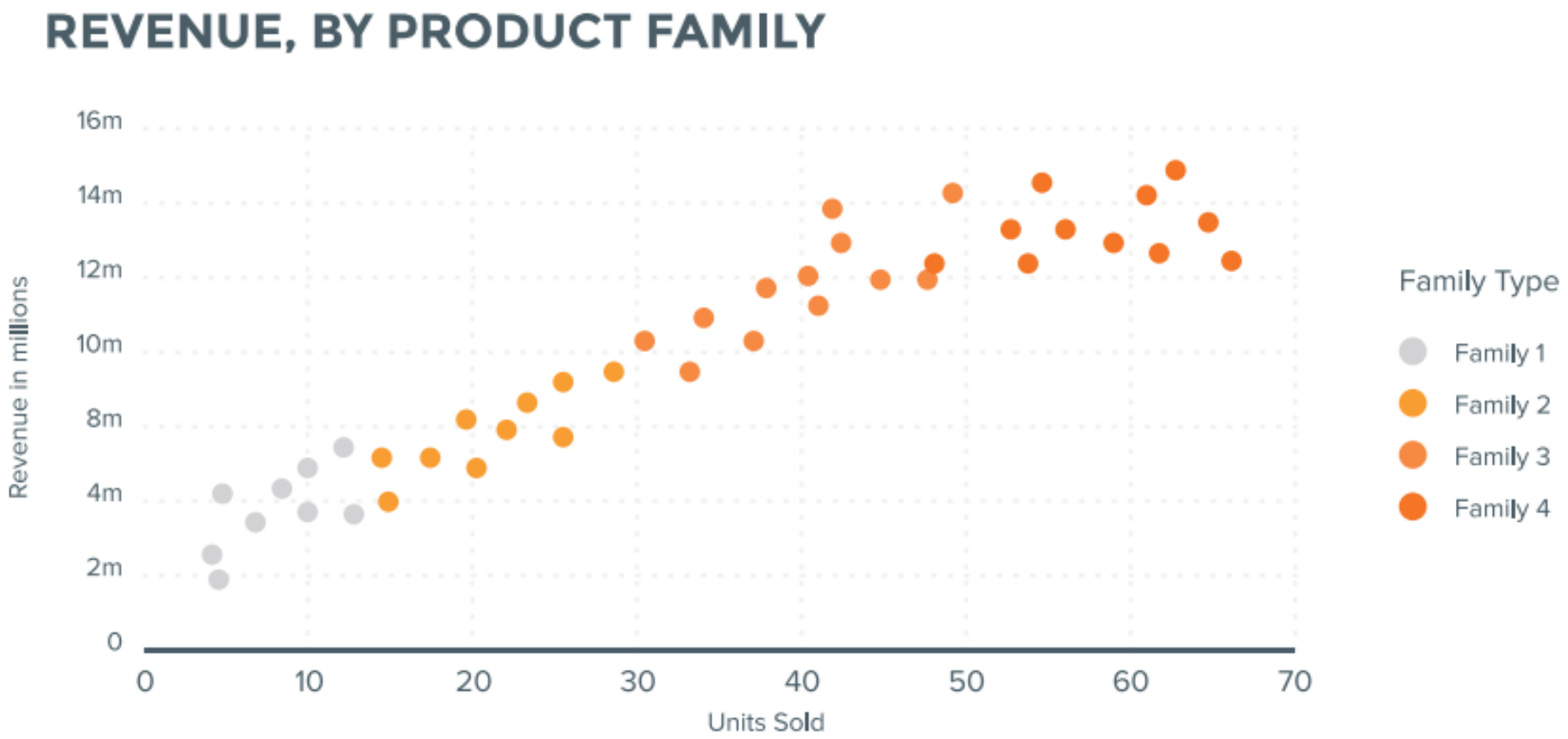
USE TRANSPARENT COLORS
In standard area charts, ensure data isn't obscured in the background by ordering thoughtfully and using transparency.



DON'T USE AREA CHARTS TO DISPLAY DISCRETE DATA
The connected lines imply intermediate values, which only exist with continuous data.

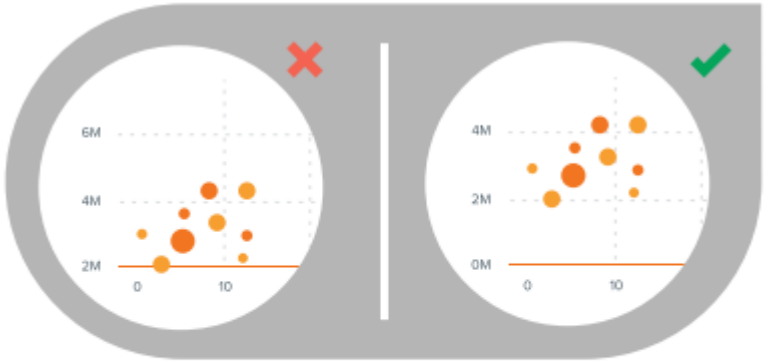
Scatter Plot

Scatter plots show the relationship between items based on two sets of variables. They are best used to show correlation in a large amount of data.

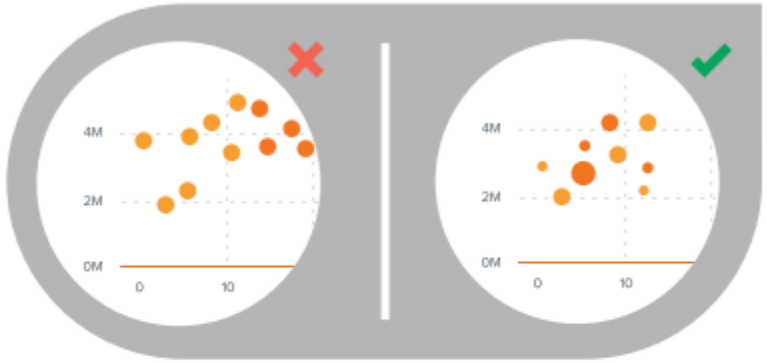


Scatter Plot

DESIGN BEST PRACTICES

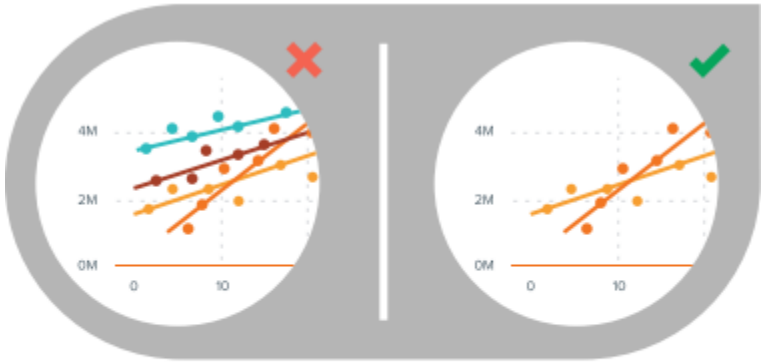


START Y-AXIS VALUE AT 0
Starting the axis above zero truncates the visualization of values.



INCLUDE MORE VARIABLES
Use size and dot color to encode additional data variables.

USE TREND LINES
These help draw correlation between the variables to show trends.



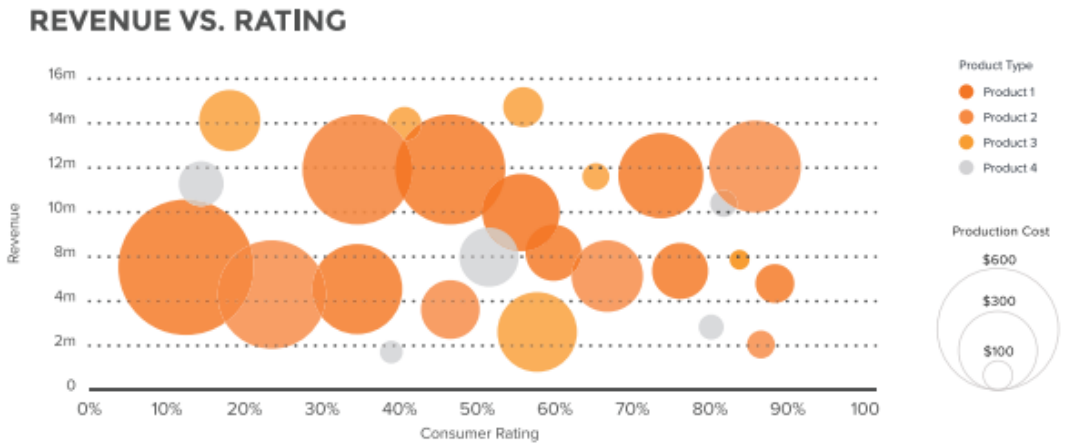
DON'T COMPARE MORE THAN 2 TREND LINES
Too many lines make data difficult to interpret.

Bubble Chart

BUBBLE CHART

Bubble charts are good for displaying nominal comparisons or ranking relationships.

VARIATIONS OF BUBBLE CHARTS



BUBBLE PLOT

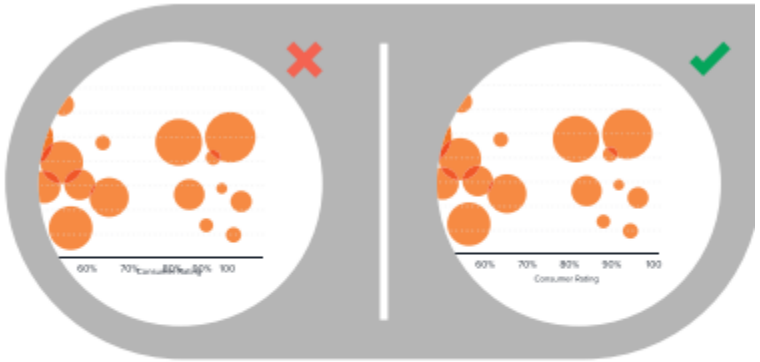
This is a scatter plot with bubbles, best used to display an additional variable.

BUBBLE MAP

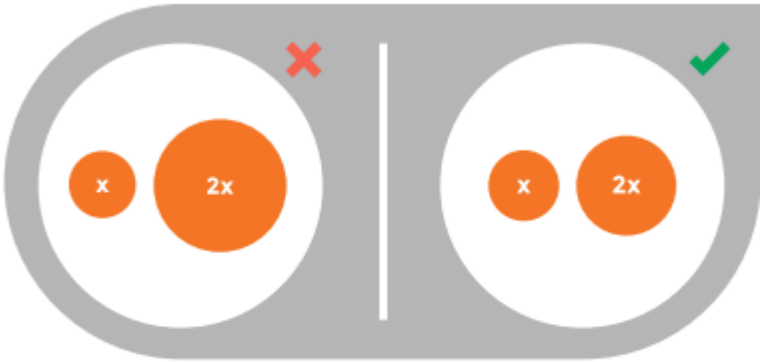
Best used for visualizing values for specific geographic regions.

Bubble Chart

DESIGN BEST PRACTICES



MAKE SURE LABELS ARE VISIBLE
All labels should be unobstructed and easily identified with the corresponding bubble.

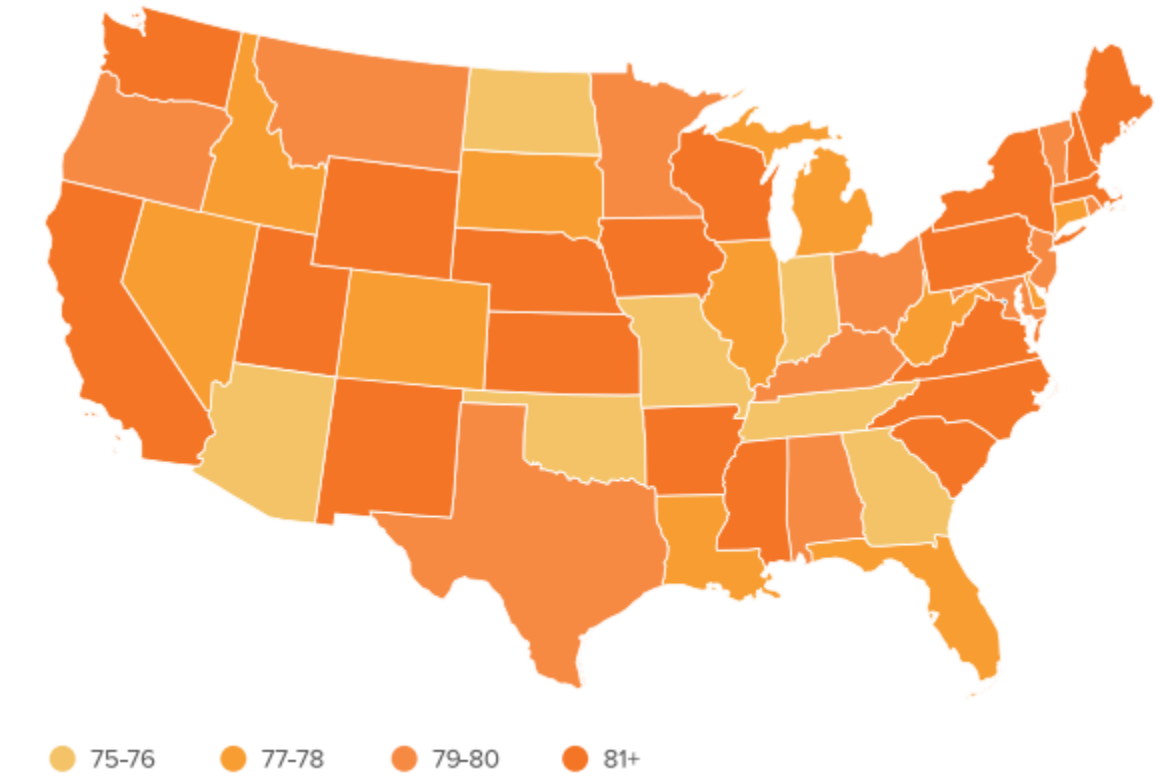


SIZE BUBBLES APPROPRIATELY
Bubbles should be scaled according to area, not diameter.

DON'T USE ODD SHAPES
Avoid adding too much detail or using shapes that are not entirely circular; this can lead to inaccuracies.

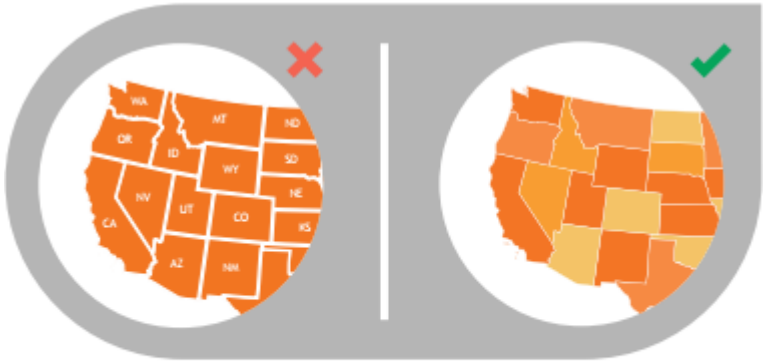
Heat Map

Heat maps display categorical data, using intensity of color to represent values of geographic areas or data tables.

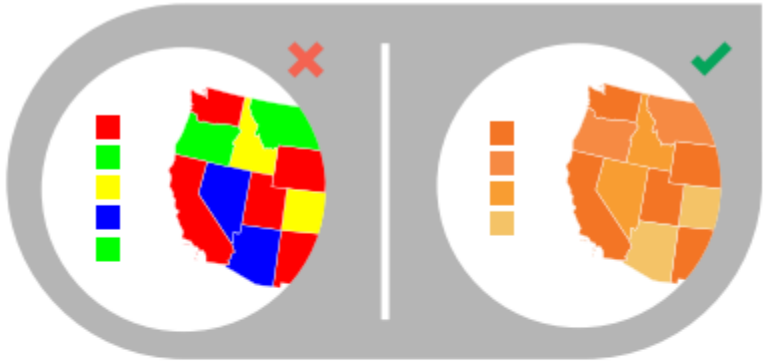


Head Map

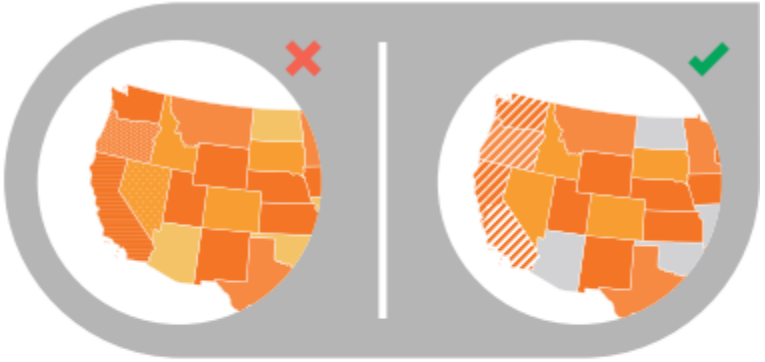
DESIGN BEST PRACTICES



USE A SIMPLE MAP OUTLINE
These lines are meant to frame the data, not distract.



SELECT COLORS APPROPRIATELY
Some colors stand out more than others, giving unnecessary weight to that data. Instead, use a single color with varying shade or a spectrum between two analogous colors to show intensity. Also remember to intuitively code color intensity according to values.



USE PATTERNS SPARINGLY
A pattern overlay that indicates a second variable is acceptable, but using multiple is overwhelming and distracting.

CHOOSE APPROPRIATE DATA RANGES
Select 3-5 numerical ranges that enable fairly even distribution of data between them. Use +/- signs to extend high and low ranges.

DOS AND DON'TS

10 DATA DESIGN DOS AND DON'TS

Designing your data doesn't have to be overwhelming. With a basic understanding of how different data sets should be visualized, along with a few fundamental design tips and best practices, you can create more accurate, more effective data visualizations. Follow these 10 tips to ensure your design does your data justice.

1 | DO USE ONE COLOR TO REPRESENT EACH CATEGORY.

2 | DO ORDER DATA SETS USING LOGICAL HEIRARCHY.

3 | DO USE CALLOUTS TO HIGHLIGHT IMPORTANT OR INTERESTING INFORMATION.

4 | DO VISUALIZE DATA IN A WAY THAT IS EASY FOR READERS TO COMPARE VALUES.

5 | DO USE ICONS TO ENHANCE COMPREHENSION AND REDUCE UNNECESSARY LABELING.

6 | DON'T USE HIGH CONTRAST COLOR COMBINATIONS SUCH AS RED/GREEN OR BLUE/YELLOW.

7 | DON'T USE 3D CHARTS. THEY CAN SKEW PERCEPTION OF THE VISUALIZATION.

8 | DON'T ADD CHART JUNK. UNNECESSARY ILLUSTRATIONS, DROP SHADOWS, OR ORNAMENTATIONS DISTRACT FROM THE DATA.

9 | DON'T USE MORE THAN 6 COLORS IN A SINGLE LAYOUT.

10 | DON'T USE DISTRACTING FONTS OR ELEMENTS (SUCH AS BOLD, ITALIC, OR UNDERLINED TEXT).

Theory 1: "Clutter is your enemy!"

Clutter = Cognitive Load

- Every unnecessary element on a chart (borders, shadows, backgrounds, heavy gridlines) is **Clutter**.
- Clutter increases **"Cognitive Load,"** forcing your audience's brain to work harder just to understand the visual.
- **Goal:** Maximize the "Data-to-Ink Ratio."

FIGURE 3.8 The graph still appears complete without the border and background shading

Declutter Using Gestalt Principles

Leverage How Our Brains Automatically Group Information

Theory: The **Gestalt Principles** explain how our brains find patterns. Use them to organize your visuals:

- **Proximity:** Objects close together are seen as a group. (Use to group bars).
- **Similarity:** Objects of the same color/shape are seen as a group. (Use to encode categories).
- **Enclosure:** Objects within a border/shaded area are seen as a group.

FIGURE 3.1 Gestalt principle of proximity

FIGURE 3.3 Gestalt principle of similarity

FIGURE 3.5 Gestalt principle of enclosure

Theory 2: Preattentive Attributes

Strategically Direct Your Audience's Attention

- **Preattentive Attributes** are visual cues your brain processes in milliseconds *before* you even consciously "look" (e.g., Color, Size, Position).
- They are your most powerful tool to **focus your audience's attention** where you want it.

Example using Color:

- Instead of 10 competing colors (clutter), use **grey** for 9 context and **one bold color** (e.g., blue) for the single line (the story) you are talking about.
- This is the **"Gray-out" Technique**—a key storytelling tool.

FIGURE 4.2 Count the 3s example

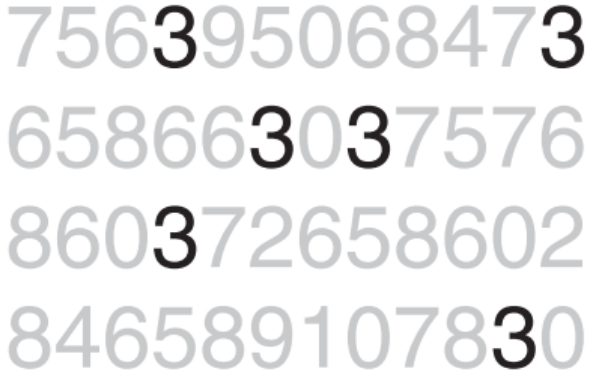


FIGURE 4.3 Count the 3s example with preattentive attributes

Summary

Key Principles

1. **Purpose:** Always start with your goal: **Exploratory** (for you) vs. **Explanatory** (for your audience)?
2. **Chart Choice:** Choose the right chart for the job (Compare -> Bar, Trend -> Line). Avoid pie charts.
3. **Declutter:** Remove all "Clutter" (borders, shadows, etc.) to reduce Cognitive Load.
4. **Focus:** Use Preattentive Attributes (like **Color**) strategically to guide your audience's attention to the main message.

Q&A