

5. Data Visual Design

Above all else show the data.

Edward Tufte (a pioneer in data visualization, professor at Yale)

A Framework for Visualization

The framework of Tamara Munzner, a computer scientist at the University of British Columbia, breaks the design of a visualization into a sequence of questions (Munzner 2014):

- **Actions** of data visualization → users’ actions toward visualization
- **Target** of data visualization → which part of the data users focus on
- The choice of visualization **idioms** → which form best presents it
- **Marks and channels** → how to encode them visually
- **Expressiveness** vs **effectiveness** → key standards for good visualization
- **Elements** of a single graph → components in a graph
- User **experience** and **organization** → how people will use and understand the visualization
- Graphic **interaction** → how people can interact with the visualization

It represents a thinking flow: from understanding the goal of visualization, to selecting the right form, to ensuring clarity, usability, and interaction.

A Running Example

Throughout this lecture and lab, we will use a dataset on transit agencies. Most transit agencies, including the four largest ones in the Puget Sound region, King County Metro (KCM), Sound Transit (ST), Community Transit (CT), and Pierce Transit (PT), report their yearly operations to the Federal Transit Administration’s [National Transit Database](#) (NTD). Our subset has 89 rows, one for each agency, mode, type of service, and year from 2022 to 2025.

Attribute	Type	Example: Sound Transit, 2025
Agency	Categorical	Sound Transit
Mode	Categorical	Light rail, commuter rail, commuter bus, ...
Type of service	Categorical	Run by the agency, or by a contractor
Year	Ordinal	2025
Fare revenue	Quantitative	\$74.6 million
Operating expenses	Quantitative	\$704.7 million
Unlinked passenger trips	Quantitative	50.9 million boardings
Vehicle revenue hours	Quantitative	1.23 million hours of service

Table 5.1: The attributes of the Puget Sound transit data. Source: [National Transit Database](#).

Actions: Why Are They Looking?

Munzner describes three levels of actions for different user purposes: analyze (high level), search (middle level), and query (low level). **Much of the content in the**

following sections comes from Munzner (2014).

Analyze (high level). From the perspective of users, why are they looking at the visualization?

- **Consuming** a visualization rather than producing one
 - Discover: explore data
 - Present: show what is already known to others (explain)
 - Enjoy: casual and social
- **Producing** new information
 - Annotate: take notes and annotate key points
 - Record: record the discovery and analysis process
 - Derive: derive new variables and produce new charts

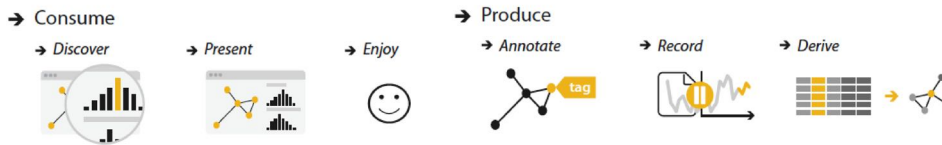


Figure 5.1: Actions at the analyze level.

Search (middle level). Users are trying to find a known or unknown item in the visualization. Because users attempt to search for something, interaction and navigation are important to help them find the right places.

- **Lookup:** for example, a word in a dictionary in alphabetical order
- **Locate:** for example, keys in your house, or a node in a network
- **Browse:** for example, books in a bookstore
- **Explore:** for example, finding a cool neighborhood in a new city

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

Figure 5.2: Actions at the search level.

Query (low level). Users are trying to find out about the detailed characteristics of items: identify, compare, and summarize. How much of the data matters:

- **One:** identify
- **Some:** compare
- **All:** summarize

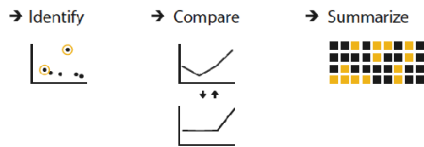


Figure 5.3: Actions at the query level.

A task usually combines one action from each level. In the web version of these notes, picking one in each column gives an example of that task on the Puget Sound transit data.

Analyze (high level)	Search (middle level)	Query (low level)
Discover: find something new	Lookup: know what and where	Identify: one item
Present: show what is known	Locate: know what, not where	Compare: some items
Enjoy: look out of curiosity	Browse: know where, not what	Summarize: all items
Annotate: add notes to a chart	Explore: know neither	
Record: save what was found		
Derive: compute new attributes		

Table 5.2: Munzner’s actions at three levels; the examples in the web version were generated with Claude Opus 5.5.

Targets: What Will Users Look At?

A target describes **which part of the data** users are examining, and this depends on the **data type**. Almost any dataset can be decomposed into three parts: the dataset as a whole (all data), its items, and its attributes.

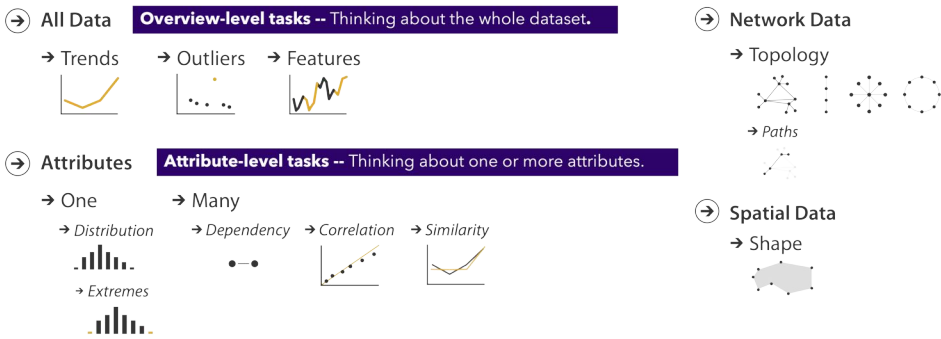


Figure 5.4: Targets: all data, attributes, network data, and spatial data.

Example 5.1. Figure 5.5 looks at the transit data as a whole: a trend over the years, the points that break a pattern, and a group that stands apart. Figure 5.6 looks at its attributes, first one alone and then two together.

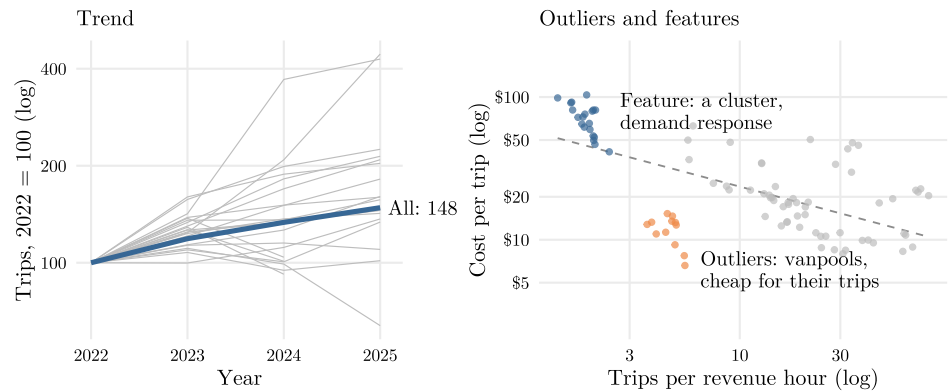


Figure 5.5: All data as the target, all 89 rows. Left: each service’s trips as an index (its 2022 = 100) as thin grey lines, all services together as the thick line. Right: cost per trip against trips per revenue hour, with the fitted log-log line. Source: [National Transit Database](#).

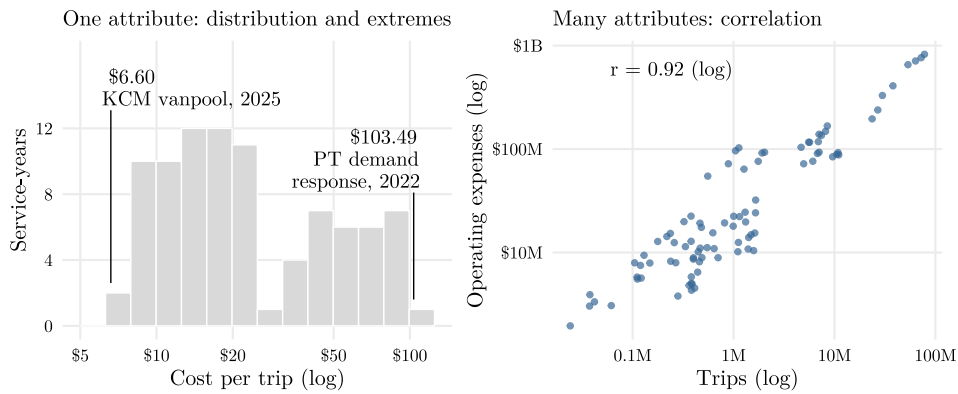


Figure 5.6: Attributes as the target, all 89 rows. Left: one attribute, cost per trip, its distribution and two extremes. Right: two attributes, trips and operating expenses, and their correlation. Source: [National Transit Database](#).

Idioms: How to Visualize?

The types of the variables narrow down the charts that fit; [From Data to Viz](#) offers a decision tree for choosing among them.

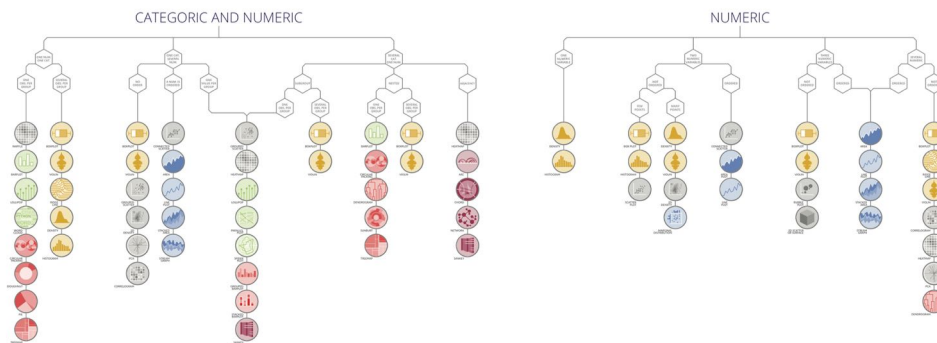


Figure 5.7: Example of choosing a chart from the types of the variables. Source: [From Data to Viz](#).

The idiom is chosen from the action and the target. Four families of idioms answer four questions:

- **Encode** asks how to convert data to visual elements, including positions of marks and their visual channels (more in the next section):
 - How to arrange the items (express, separate, order, align)?
 - How to map attributes to the other channels (size, value, texture, color, orientation, shape)?
- **Facet** (juxtapose, partition, superimpose): how to show multiple views, especially to compare?
- **Reduce** (filter, aggregate, embed): how to simplify complex data?
- **Manipulate** (change, select, navigate): how can users interact with the visualization?

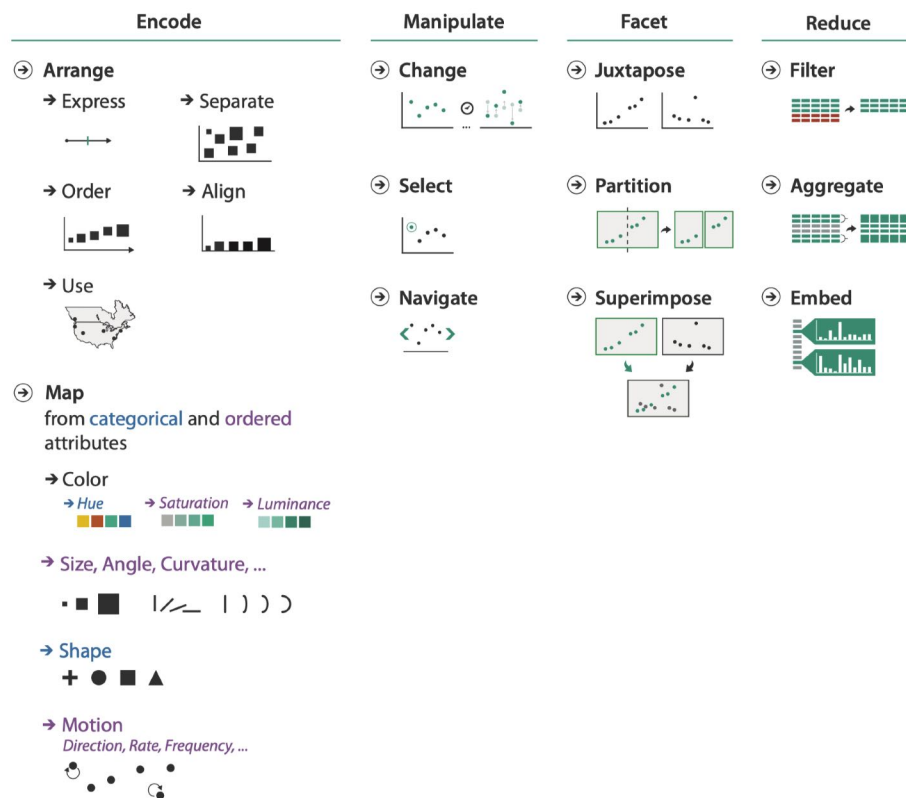


Figure 5.8: Encode: arrange (express, separate, order, align) and use.

Example 5.2. The following figure puts the facet, reduce, and manipulate idioms to work on the monthly ridership of the four transit agencies, 2015 to 2025.

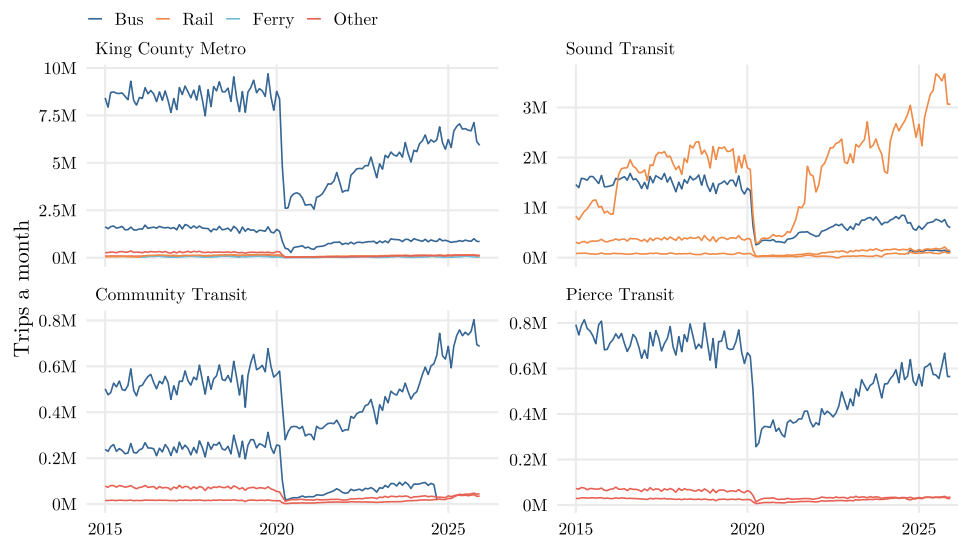


Figure 5.9: Monthly trips and revenue hours of the four agencies by mode, 2015-2025, under the idioms of this section. Source: [National Transit Database](#).

Marks and Channels

- **Marks** represent **items** or **links**.
- **Channels** (also called visual variables) control the **appearance** of marks based on **attributes**.

Marks as items are basic geometric elements of a chart: points (0D), lines (1D), and areas (2D). A 3D mark, volume, is rarely used.

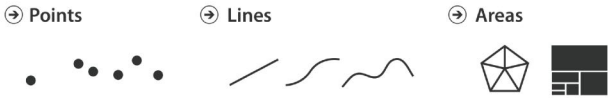


Figure 5.10: Marks as items: points, lines, and areas.

Marks as links show containment or connection:

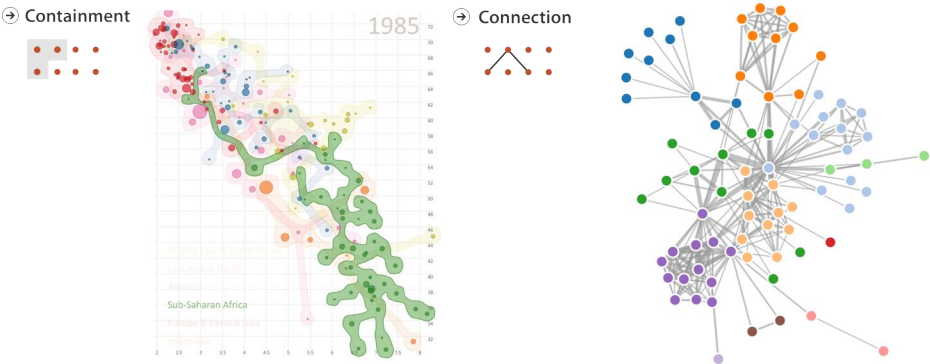


Figure 5.11: Marks as links. Left: containment, regions enclosing points of the same group. Source: [Bubble Sets](#), [vialab](#). Right: connection, a force-directed network. Source: [Observable](#).

Channels include position, color, size (length, area, volume), shape, and tilt.

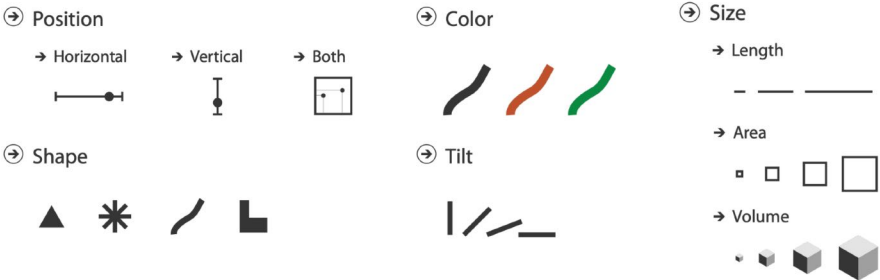


Figure 5.12: Channels: position, color, size, shape, and tilt.

Expressiveness: Matching Channel to Data Type

In Mackinlay’s terms, a visualization is **expressive** if it shows all the facts in the data, and only those facts. Channels differ in how accurately people read them, and the ranking depends on the type of data (Figure 5.13).

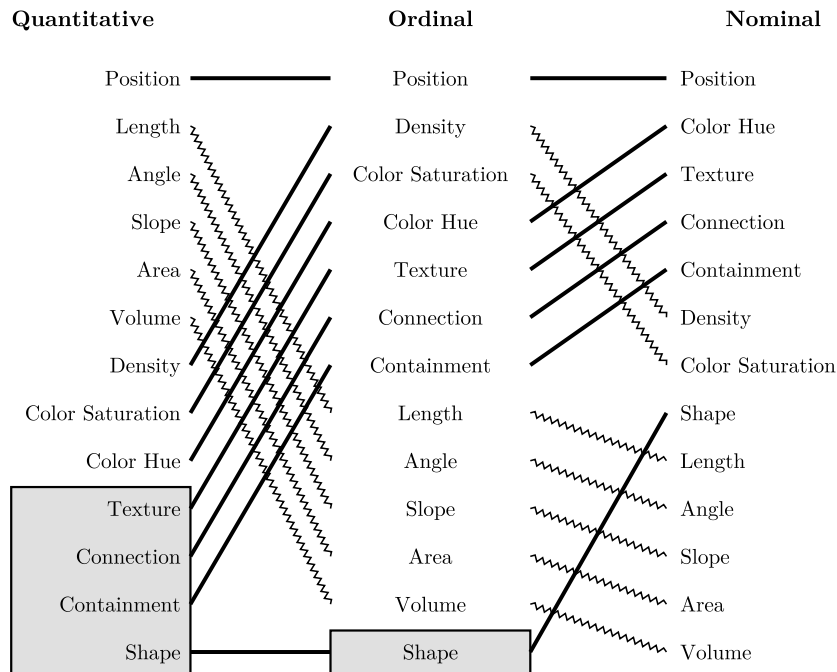


Figure 5.13: Ranking of perceptual tasks for quantitative, ordinal, and nominal data. The tasks in the gray boxes are not relevant to that type of data. Source: modified from Mackinlay (1986).

Example 5.3. Figure 5.14 adds two attributes to a scatter plot: the trips each service carried in the year, and its mode group. The two panels encode them differently.

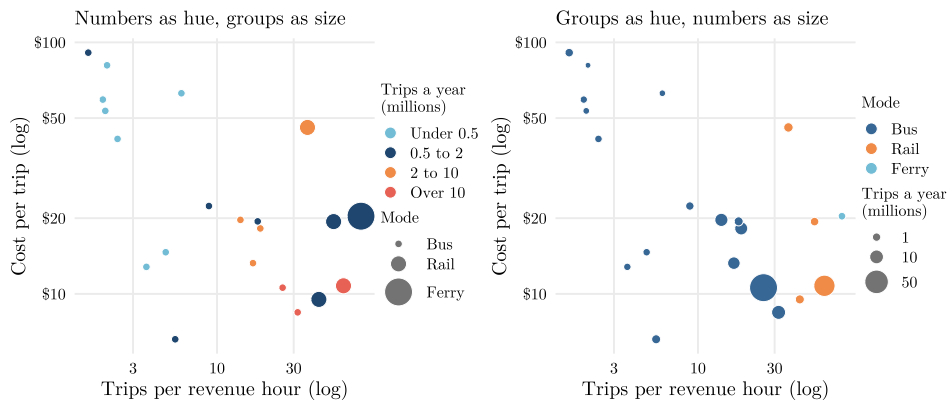


Figure 5.14: Cost per trip against trips per revenue hour in 2025, with trips a year and mode group. Source: National Transit Database.

Effectiveness: Choosing Channels by Importance

In short, **effectiveness** means giving the most accurate channel (at the top of the list) to the most important variable. When a weaker channel cannot be avoided, print the numbers on the marks.

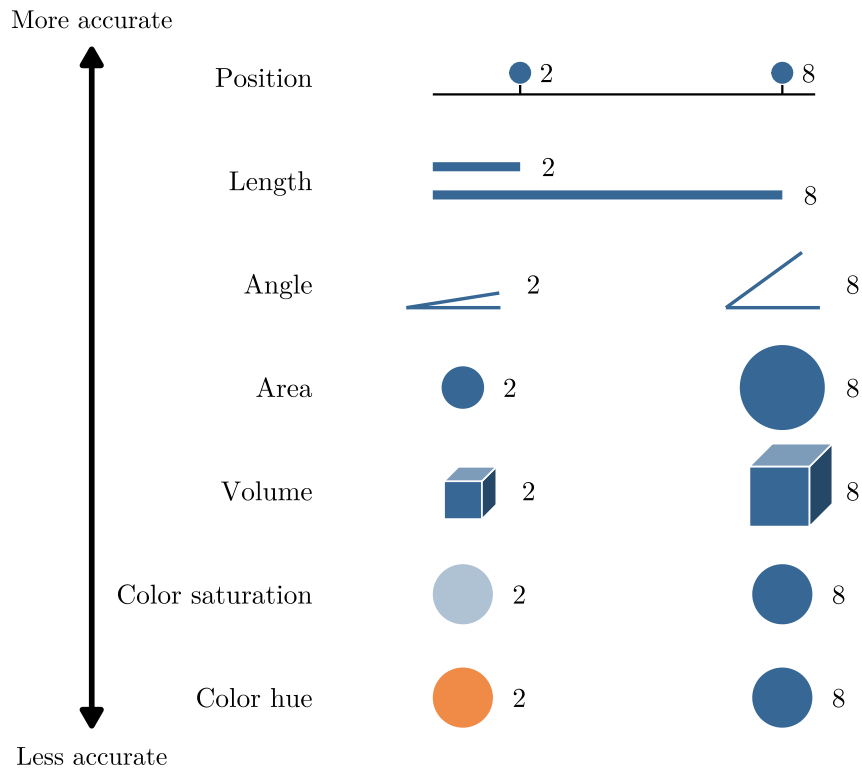


Figure 5.15: The values 2 and 8 drawn with each channel, from the most accurately read (top) to the least (bottom).

Example 5.4. The figure shows where the riders are: one point for each agency and mode, its shade the number of trips in 2025. Can you still tell which variable matters most here?

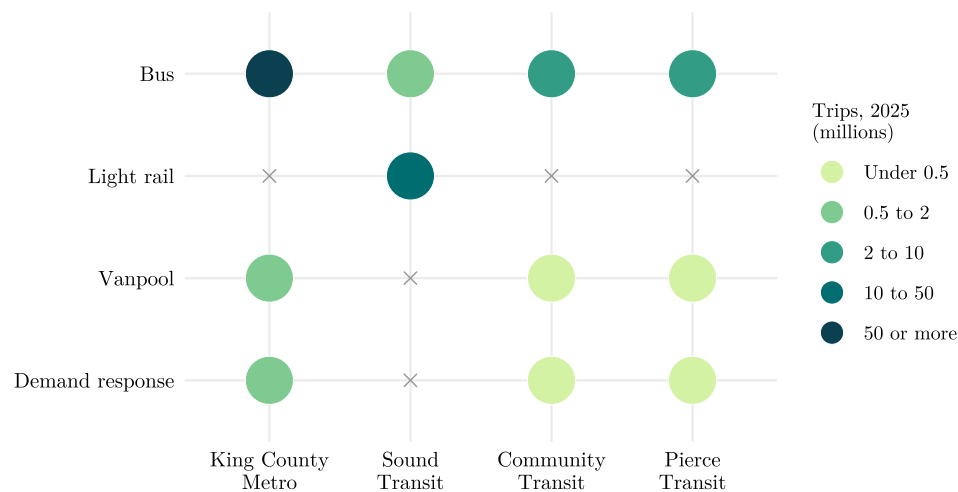


Figure 5.16: Trips in 2025 by agency and mode: one point for each pair, darker for more trips. Source: [National Transit Database](#).

Example 5.5. They Made a Terrible Visualization: The Alabama Department of Labor charted weekly unemployment insurance claims as packed bubbles. The numbers are encoded as area and the dates as color hue, two channels near the bottom of the ranking.

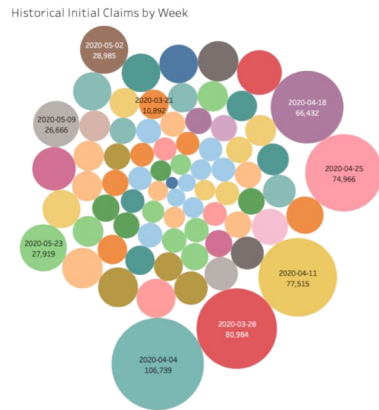


Figure 5.17: Weekly unemployment insurance claims as packed bubbles. Source: Alabama Department of Labor; [From Data to Viz.](#)

Pairing position with position (a line or area chart), position with length (a bar chart), or position with area (a bubble scatter plot) would read far better (Figure 5.18).

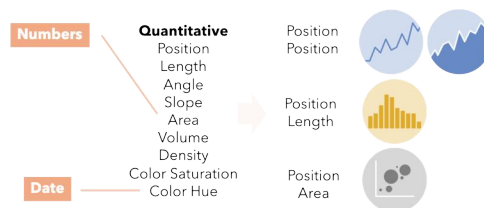


Figure 5.18: The channels of the claims chart and three better pairings. Source: [From Data to Viz.](#)

Stevens' Psychophysical Power Law OPTIONAL

Stevens' psychophysical power law explains part of the ranking: the perceived sensation S grows as a power of the physical intensity I , and $S = I^N$. So a channel with $N = 1$ is perceived in proportion, and one with $N < 1$ or $N > 1$ is under- or overestimated.

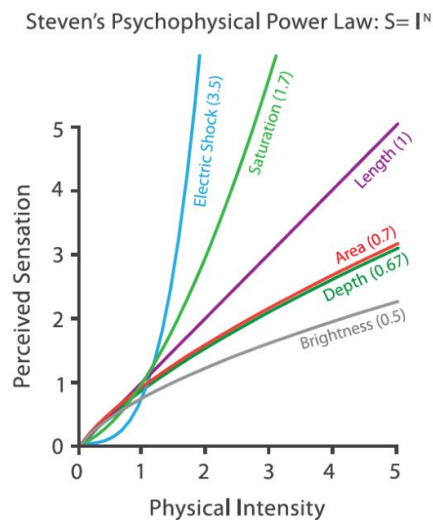


Figure 5.19: Perceived sensation against physical intensity: electric shock ($N = 3.5$) and saturation (1.7) are overestimated, length (1) is perceived in proportion, and area (0.7), depth (0.67), and brightness (0.5) are underestimated. Source: Munzner (2014).

Figure 5.20 draws three values in the ratio 1 : 2 : 5 with four of these channels.

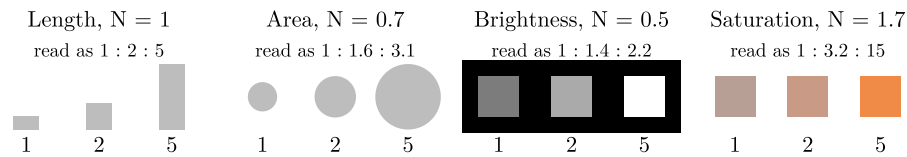


Figure 5.20: Three values in the ratio 1 : 2 : 5, drawn by length, area, brightness, and saturation, with the ratio Stevens' law predicts a reader perceives, v^N . Electric shock and depth are left out: neither can be drawn on a flat page.

Color

Color does three different jobs, and each needs its own kind of palette:

- A **sequential** palette runs from light to dark for an ordered variable from low to high.
- A **diverging** palette runs from one hue through a light neutral to another hue, for a variable with a meaningful midpoint, such as a change that can be negative or positive.
- A **qualitative** palette uses clearly different hues of similar lightness for categories with no order.

ColorBrewer, designed by Cynthia Brewer, offers tested palettes of all three kinds for maps.



Figure 5.21: Three kinds of palette: sequential, diverging, and qualitative.

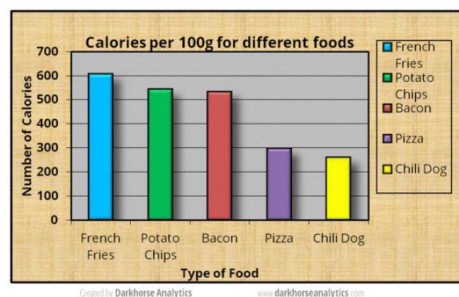
Data-Ink Ratio

Data-ink ratio = data ink / total ink used in the graphic

= the proportion of a graphic's ink devoted to the non-redundant display of data

= 1.0 - the proportion of the graphic that can be erased.

Remove backgrounds



Direct label

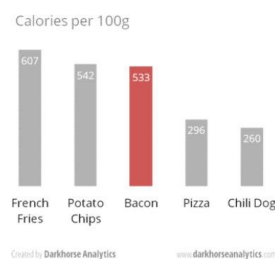


Figure 5.22: Removing the background and labeling the bars directly raises the data-ink ratio. Source: [Darkhorse Analytics](http://darkhorseanalytics.com).

Chartjunk is not always harmful. In some cases, it may even increase the memorability and engagement of a visualization.

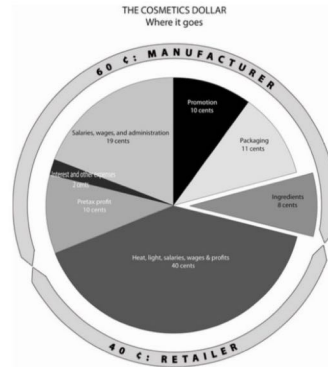
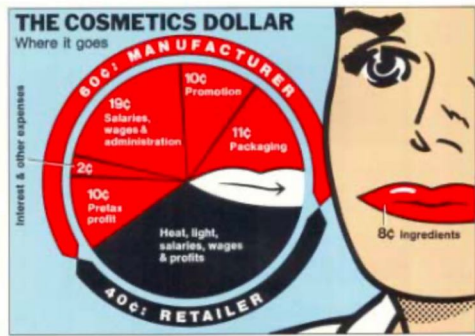


Figure 5.23: The same data as an embellished chart and as a plain pie chart. Source: Bateman et al. (2010).

Less is more: Tufte's rule is to erase ink that shows nothing: heavy grid lines, frames, shadows, three-dimensional effects, and legends that could be direct labels.

Figure 5.24: Remove to improve the data-ink ratio. Source: Darkhorse Analytics.

Remove
to improve
(the **data-ink** ratio)

Elements of a Single Graph

A single graph should have a title, a subtitle, encodings, annotations, axes, and a legend. A map should also include a scale bar, a north arrow, and its projection or coordinate system.

Washington State Monthly Apartment Rent Changes Since Covid-19
 Apartment Rent Compared to December 2019 by Quality Levels and Number of Bedroom. Data Source: RentHub, 2019-2022

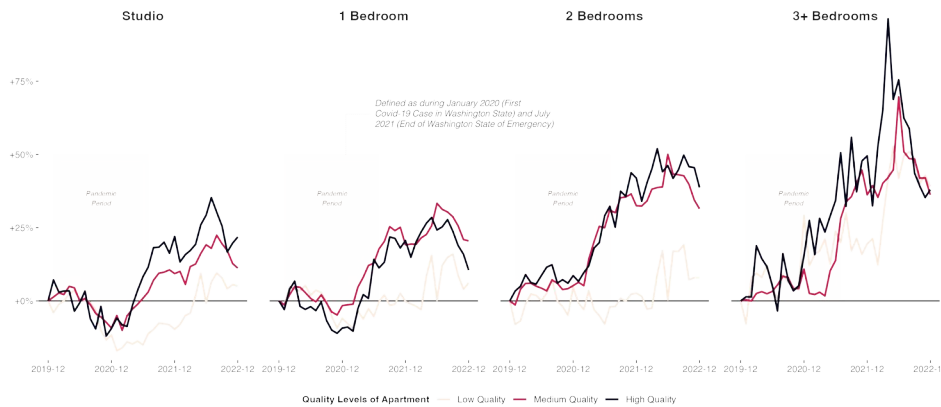


Figure 5.25: Washington State monthly apartment rent changes since COVID-19. In the HTML version, the element under the pointer is outlined and named below the chart. Data: RentHub, 2019-2022.

User Experience and Organization

User Experience

What motivates us at a deep level are the things that satisfy basic psychological needs:

- **Autonomy:** structure beats no structure; make it easy to navigate.
- **Relatedness:** no need to be too creative; keep content relevant to the audience.
- **Competence:** select information and use guidance to make the visualization intuitive and easy to understand. It is not an IQ test!

Unity

Unity is the application of methods that ensure that elements in the design (color, font, shape) appear to “go together” (Figure 5.26).

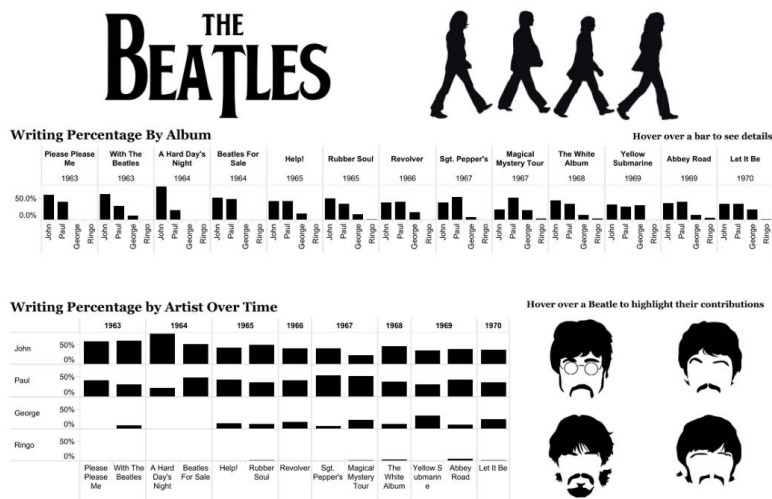


Figure 5.26: A dashboard on Beatles albums, all in black and white. Author: Mike Moore. Source: Tableau Public.

Hierarchy

Hierarchy is the application of design methods to indicate importance and “flow” within the visual, through size and placement.

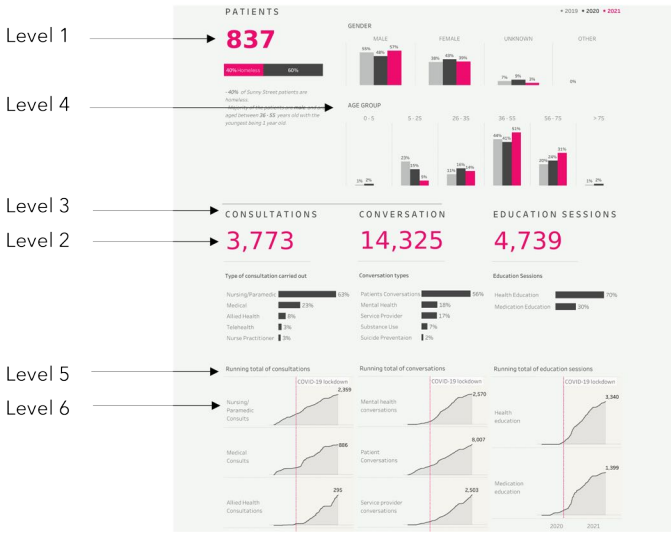


Figure 5.27: A dashboard with six levels. Source: Tableau Public.

Size

The larger element is read first.

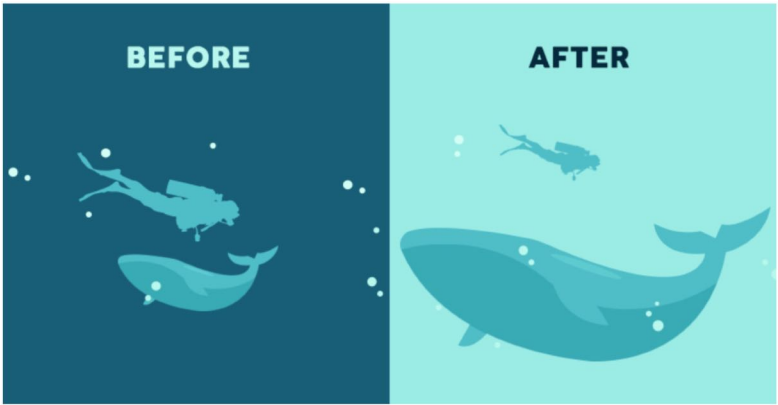


Figure 5.28: Size, before and after. Source: Visme.

Color Contrast

A contrasting color picks one element out of many.

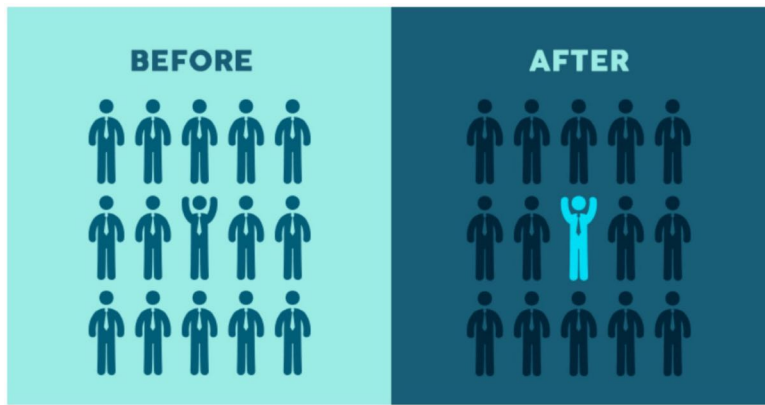


Figure 5.29: Color contrast, before and after. Source: [Visme](#).

Using Font to Organize Design

The size and weight of the type set the order in which text is read.



Figure 5.30: A business card, before and after: the name set large and bold above the contact details. Source: [Visme](#).

Balance and Alignment

Balance and alignment are used to create harmonious visuals that do not distract from the message being communicated.



Figure 5.31: Three calendar heat maps of crime in Chicago. Author: George Goczynski. Source: [Tableau Public](#).

Grouping and Spacing

Grouping and spacing can be used to associate similar elements and provide a narrative or visual flow within the visualization.

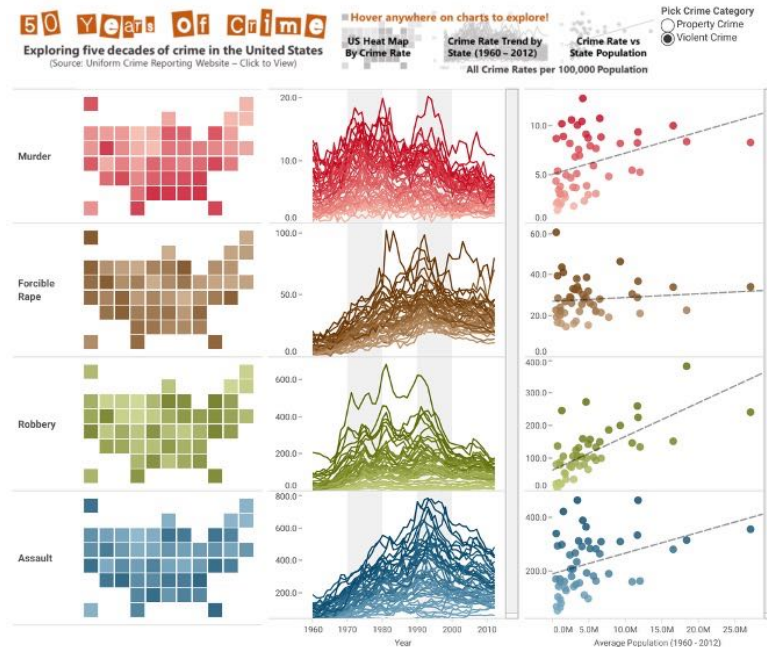


Figure 5.32: Fifty years of crime in the United States, one row per type of crime. Author: Shine Pulikathara. Source: Tableau Public.

Consume and Practice

The end of the principles is the start of practice: looking at many good visualizations and making your own matter more than any list of rules.

- **FlowingData:** <https://flowingdata.com/learning/>
- **From Data to Viz:** <https://www.data-to-viz.com/>
- **Data Viz Project:** <https://datavizproject.com/>
- **Information is Beautiful:** <https://informationisbeautiful.net/>

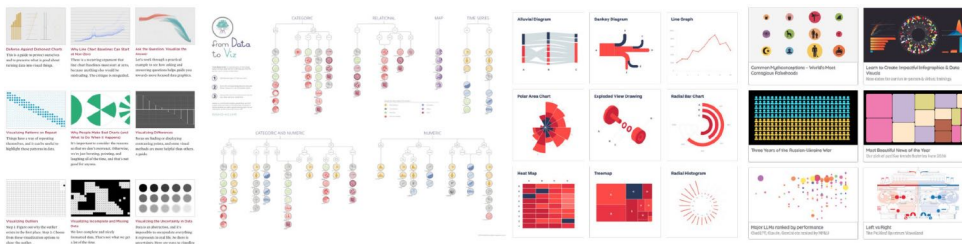


Figure 5.33: Four collections of data visualization examples and guides.

Critique and Deceptive Visualization

Visualization is a way to communicate. People can also use it to **lie, distort the truth, or mislead**, intentionally or unintentionally. What makes bad figures bad?

- **Aesthetic:** tacky, tasteless, or hodgepodge design.
- **Substantive:** the way the data are used is subpar, or the data themselves are flawed; a wrong interpretation.

- **Perceptual:** good data and aesthetics, but the graph is confusing or misleading.

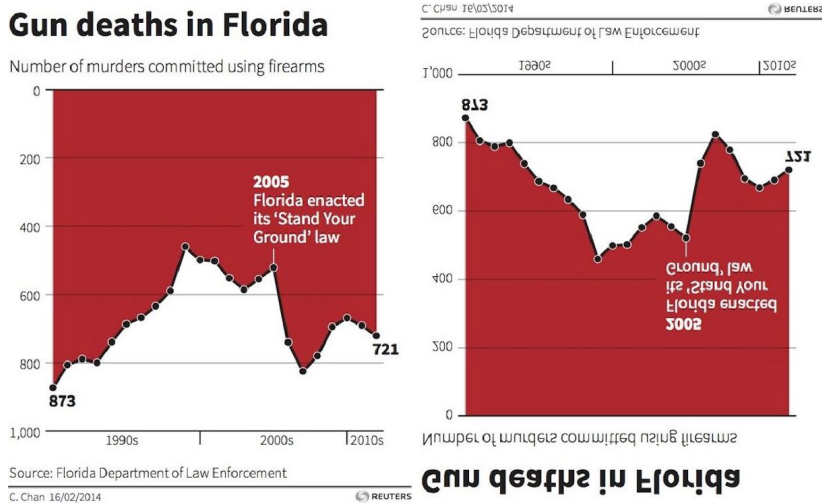


Figure 5.34: Gun deaths in Florida with an inverted y-axis (left) and the same chart flipped (right). Source: [The Society Pages](#).

Iraq's Two Stories

The same data can tell opposite stories through design choices: the title, the color, and the orientation of the bars. First, the title tells readers how to read the chart: three words, “Iraq’s Bloody Toll,” set the author’s agenda. Second, the bars hang downward. Finally, they are red, so the chart looks like blood dripping down the page.

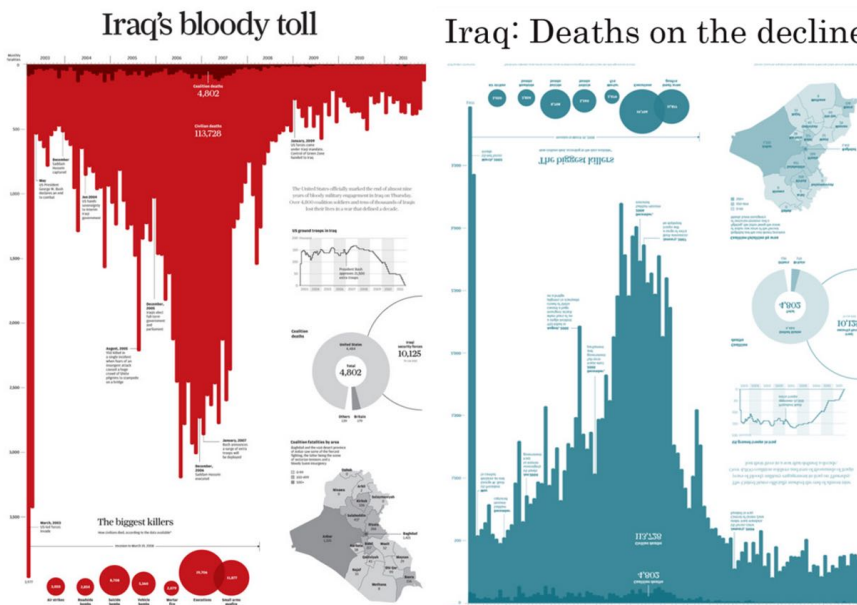


Figure 5.35: The same data as “Iraq’s Bloody Toll” and as “Iraq: Deaths on the Decline.” Source: [InfoWorld](#).

Truncated Axis

A bar chart encodes values as length. If the value axis starts above zero, every bar is cut short by the same amount, so small differences look large. **Note: the following sections are largely based on Nathan Yau’s website, [Flowing-](#)**

Data.

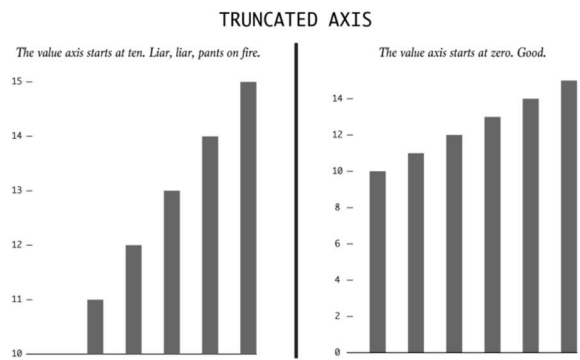


Figure 5.36: The same bars with the value axis starting at ten and at zero. Source: [FlowingData](#).

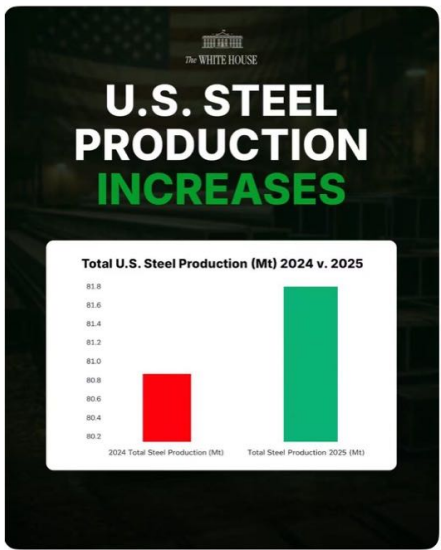


Figure 5.37: A bar chart of U.S. steel production in 2024 and 2025, posted with the line “American steel is BACK.” Source: The White House on X (2025).

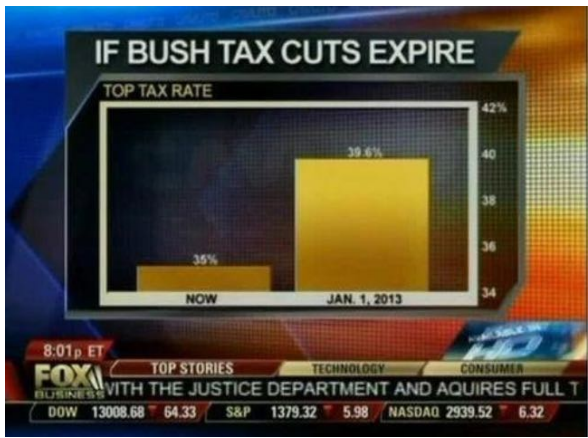


Figure 5.38: Fox News, 2012: the top tax rate rising from 35% to 39.6%, drawn on a value axis that starts at 34%, so the second bar looks about five times as tall. Source: [FlowingData](#).

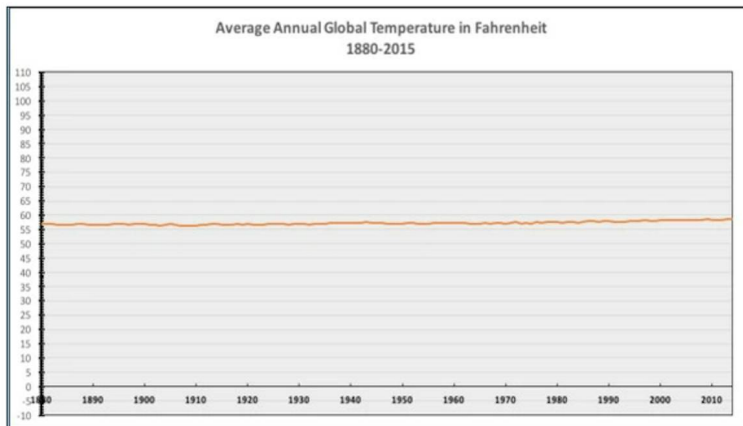


Figure 5.39: The opposite mistake. National Review, 2015: global average temperature, 1880-2015, on an axis from about -10 to 110 degrees Fahrenheit, which flattens the warming. Source: [The New Republic](#).

Dual Axes

With two value axes, each series can be stretched or squeezed on its own scale, so two unrelated trends can be made to move together. The chart then invites the reader to see a correlation, or even a cause, that may not be there.

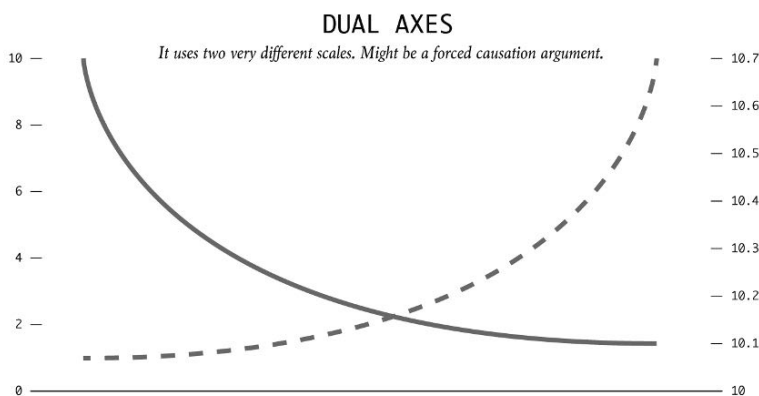


Figure 5.40: Two series on two different scales. Source: [FlowingData](#).

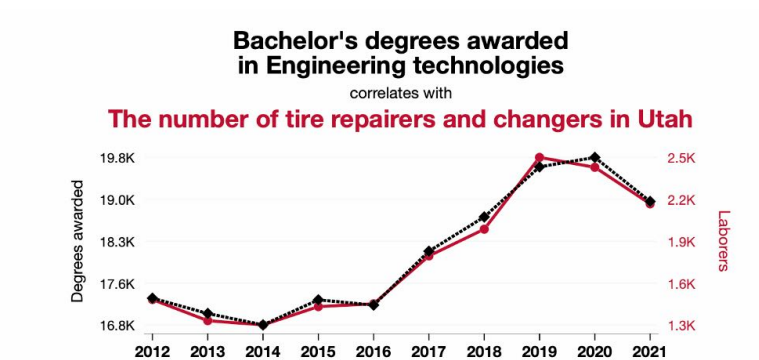


Figure 5.41: Two unrelated series made to track each other. Source: [Spurious Correlations](#).

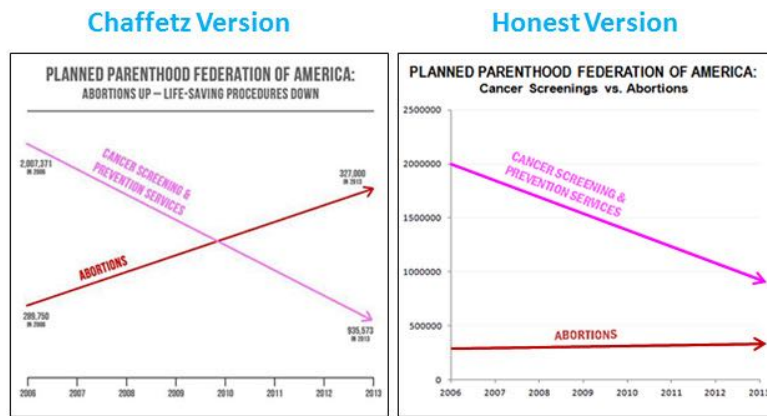


Figure 5.42: A chart shown at a 2015 congressional hearing (left) has no y-axis, so 327,000 abortions appear to overtake about 935,000 cancer screenings; the same numbers on one scale (right). Source: [Mother Jones](#).

Parts That Add Up to More Than the Whole

A chart of the parts of a whole, such as a pie, promises that the parts sum to 100 percent. When they sum to more, the chart is wrong.

IT DOES NOT ADD UP

*The parts add up to more than the whole, which is 100%.
For my next trick, I will turn this rabbit into a big bag of money.*

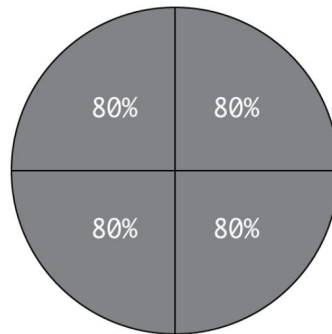


Figure 5.43: A pie chart of four wedges, each labeled 80 percent. Source: [FlowingData](#).

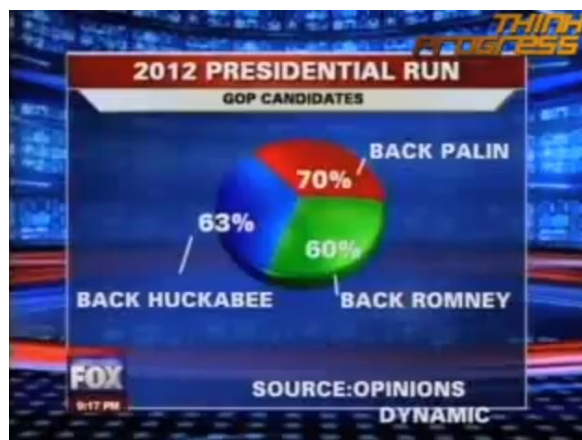


Figure 5.44: Fox News Chicago, 2009: favorability ratings of three possible 2012 candidates drawn as a pie whose slices add up to 193%. Source: [FlowingData](#).

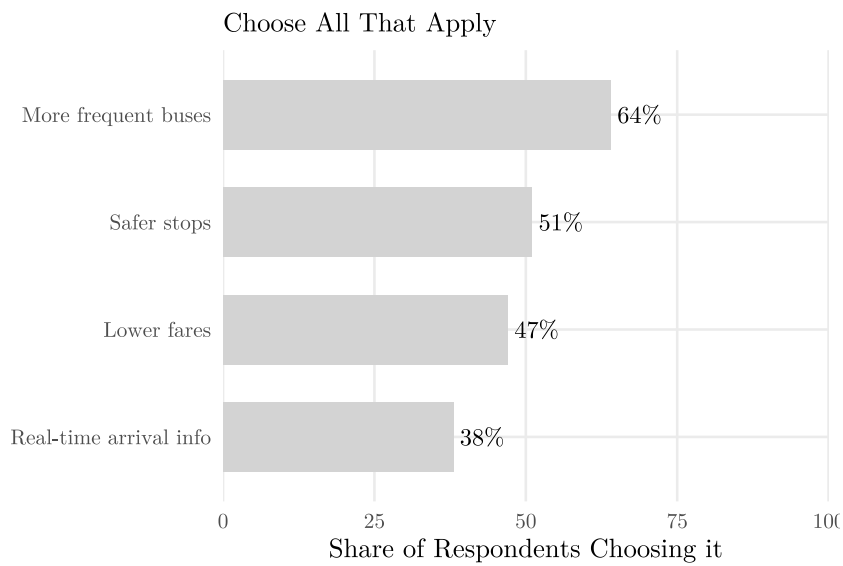


Figure 5.45: Answers to a hypothetical “choose all that apply” rider survey add up to 200 percent; bars show each share without pretending the answers are parts of one whole.

Seeing Only in Absolutes

Everything is relative. A town with two robberies is not more dangerous than a town with one if it has 1,000 times the population of the other. Rates and percentages usually say more than absolute counts and totals.

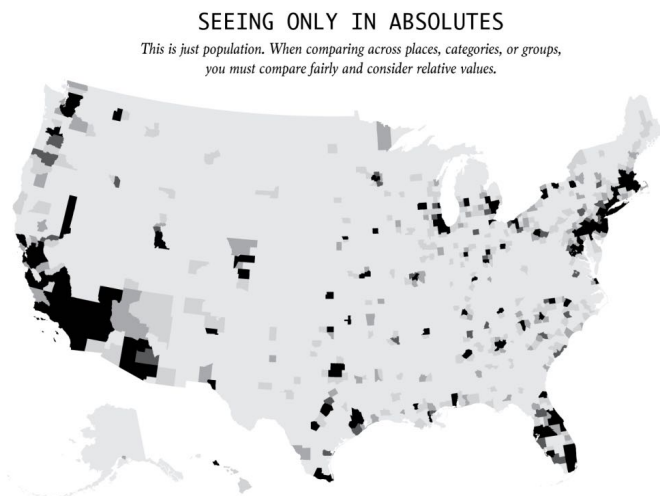


Figure 5.46: A county map of counts that mostly shows population. Source: [FlowingData](#).

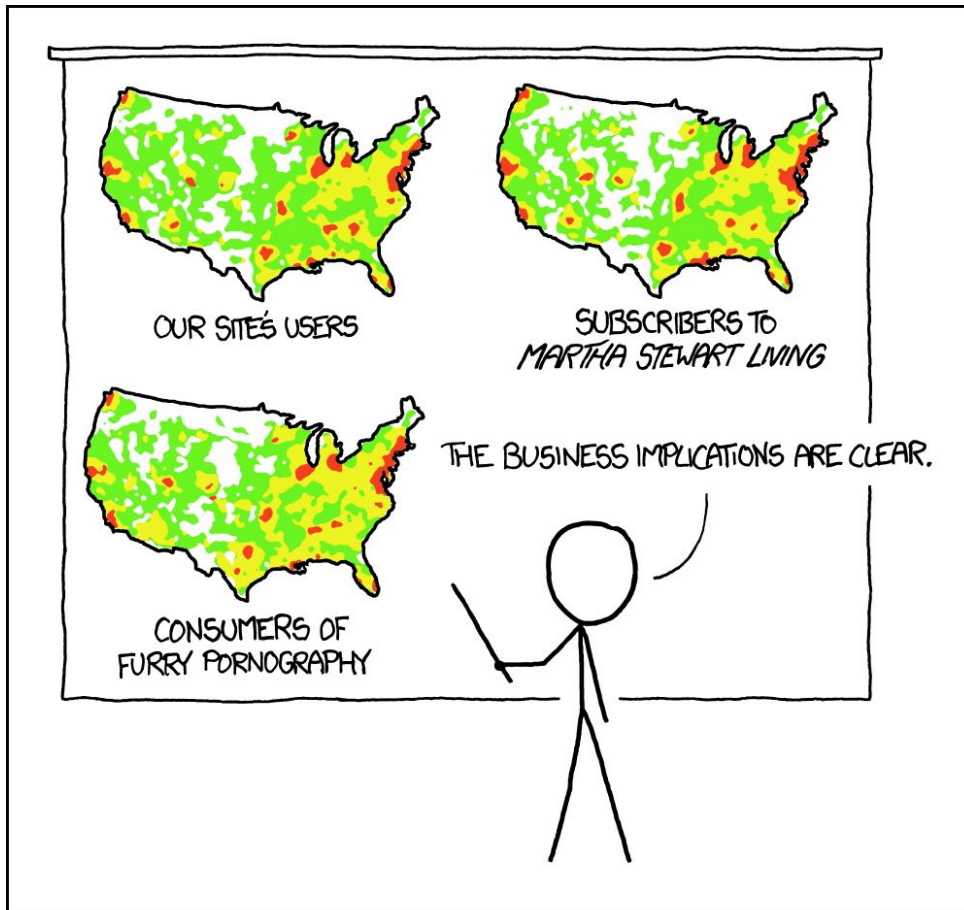


Figure 5.47: xkcd #1138, "Heatmap." Source: [xkcd](#), CC BY-NC 2.5.

PET PEEVE #208:
GEOGRAPHIC PROFILE MAPS WHICH ARE
BASICALLY JUST POPULATION MAPS

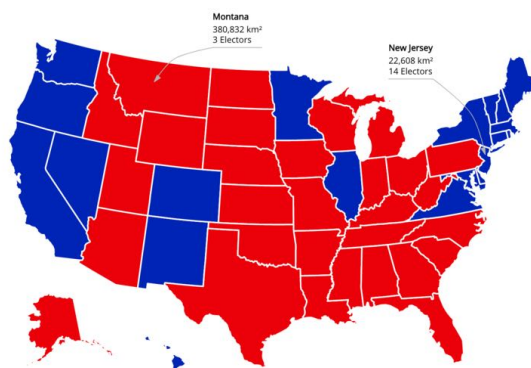


Figure 5.48: The 2016 presidential election colored by state: Montana (3 electoral votes) covers far more of the map than New Jersey (14). Source: Karim Douieb, "Land doesn't vote, people do" (2019), via [Lead Stories](#).

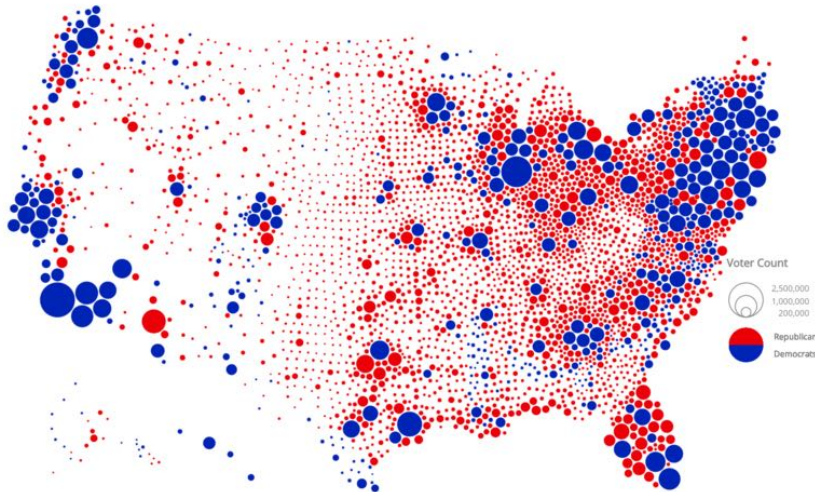


Figure 5.49: The same election as circles sized by each county's votes. Source: Karim Douieb, "Land doesn't vote, people do" (2019), via [Lead Stories](#).

Limited Scope

A short window of time can be picked to fit almost any story. Compare against the longer history, what usually happens, and a proper baseline.

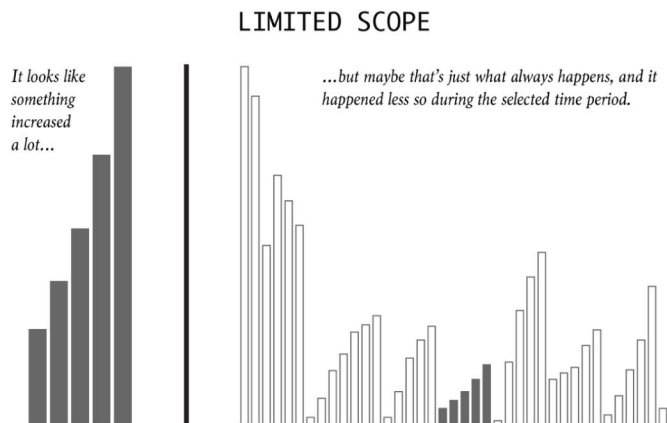


Figure 5.50: A rise over five periods (left) that is small next to the usual peaks of the longer series (right). Source: [FlowingData](#).

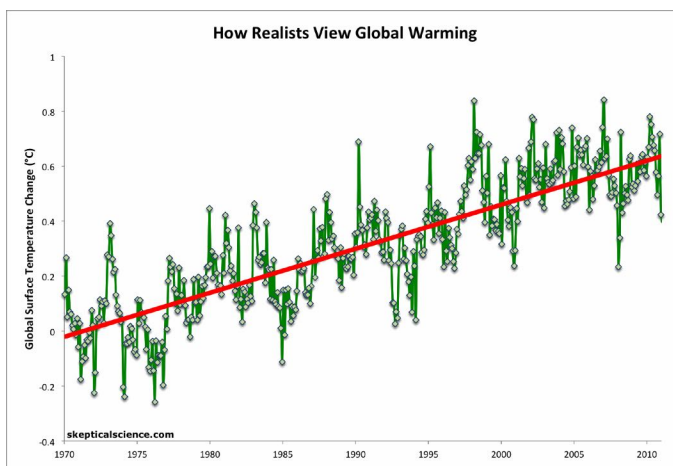


Figure 5.51: Short stretches of the global temperature record, each flat or falling, against the long-run rise; the last frame of an animation. Source: [Skeptical Science](#), CC BY 3.0.

Area Sized by One Dimension

If area is the encoding, size by area. When a square or a circle is sized by its side or radius instead, its area grows with the square of the value, so differences look far larger than they are.

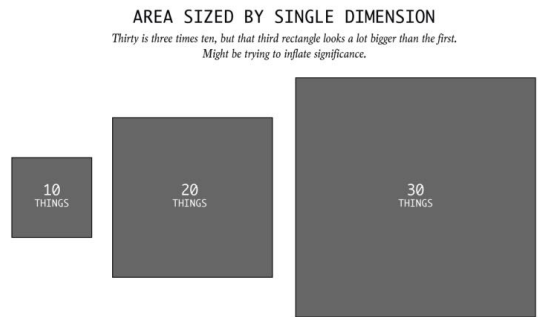


Figure 5.52: Squares for 10, 20, and 30 sized by side length. Source: [Flowing-Data](#).

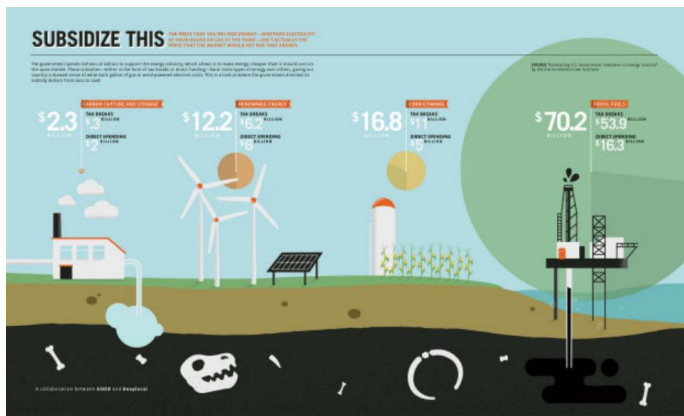


Figure 5.53: U.S. energy subsidies drawn as circles of very different sizes. Source: [FlowingData](#).

Rubber-Band Scales

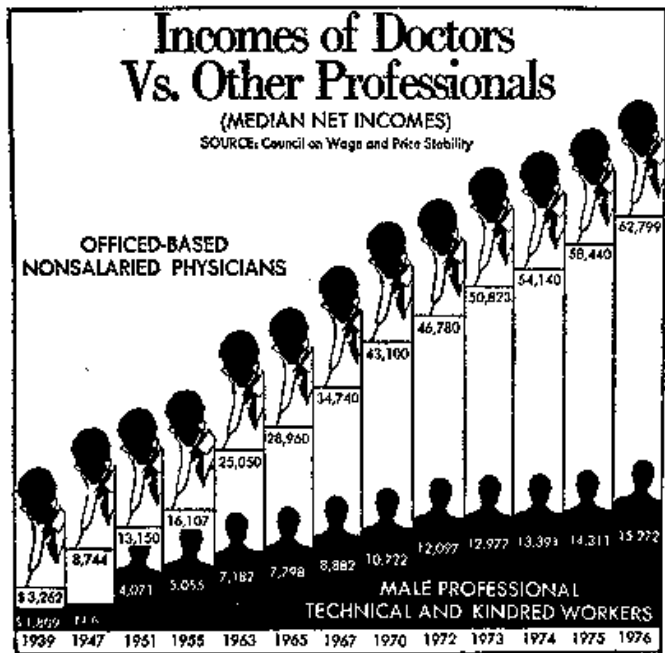


Figure 5.54: Washington Post: incomes of doctors and other professionals, 1939-1976, with the years unevenly spaced along the x-axis. Source: Wainer (1997), via [datavis.ca](#).

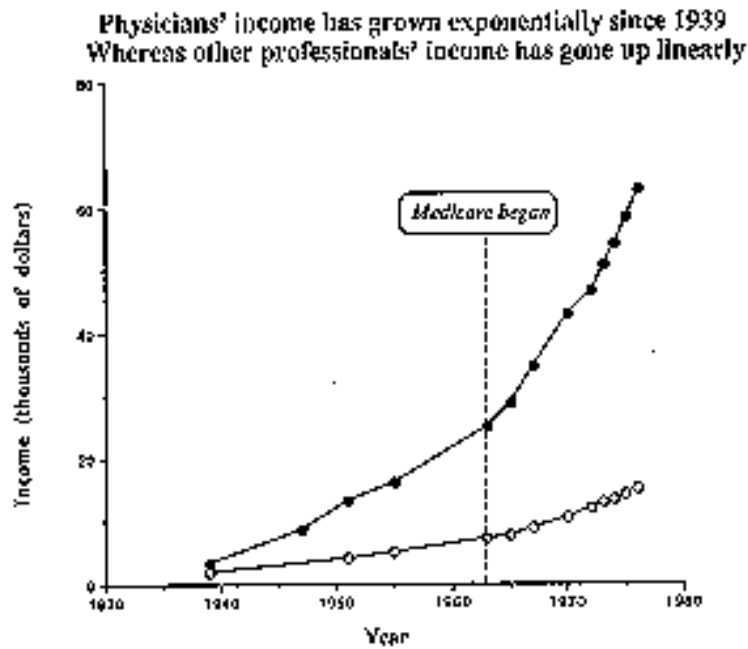


Figure 5.55: The same data with the years evenly spaced: doctors' incomes grew far faster than the other professionals'. Source: Wainer (1997), via datavis.ca.

Time Out of Order

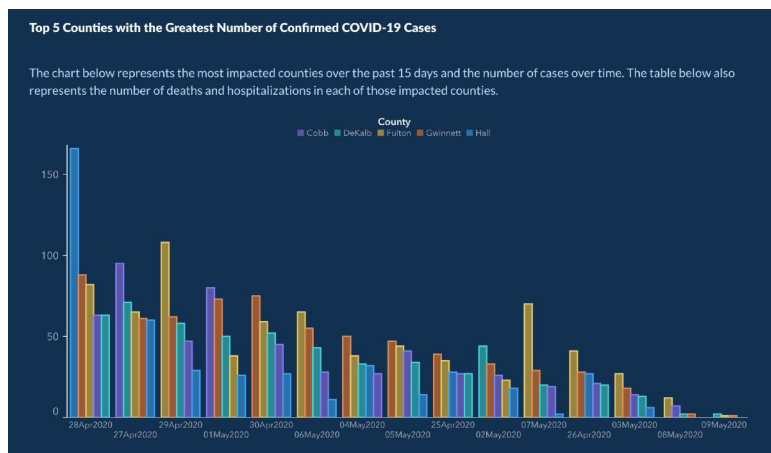


Figure 5.56: Georgia Department of Public Health, May 2020: cases in the five hardest-hit counties with the dates sorted by case count rather than by date, so cases appear to fall. Source: [This Is True](https://www.dph.ga.gov).

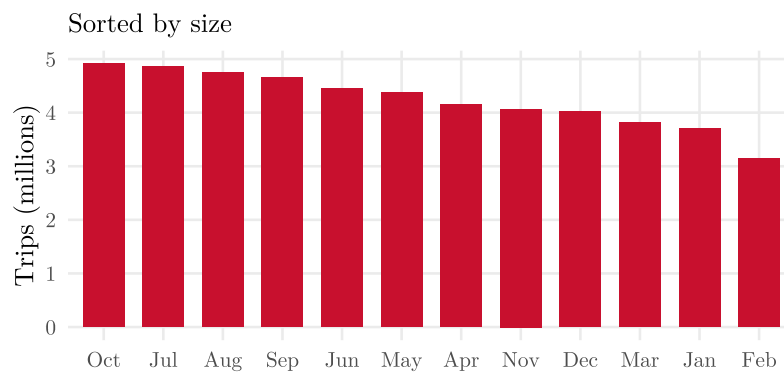


Figure 5.57: Misleading: Sound Transit's trips in each month of 2025, sorted from most to fewest, read as a steady decline. Source: [National Transit Database](https://www.soundtransit.org).

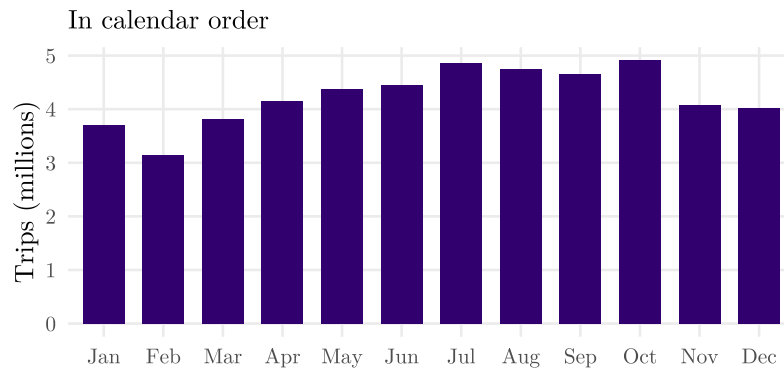


Figure 5.58: Honest: the same months in calendar order. Source: [National Transit Database](#).

Cumulative Totals

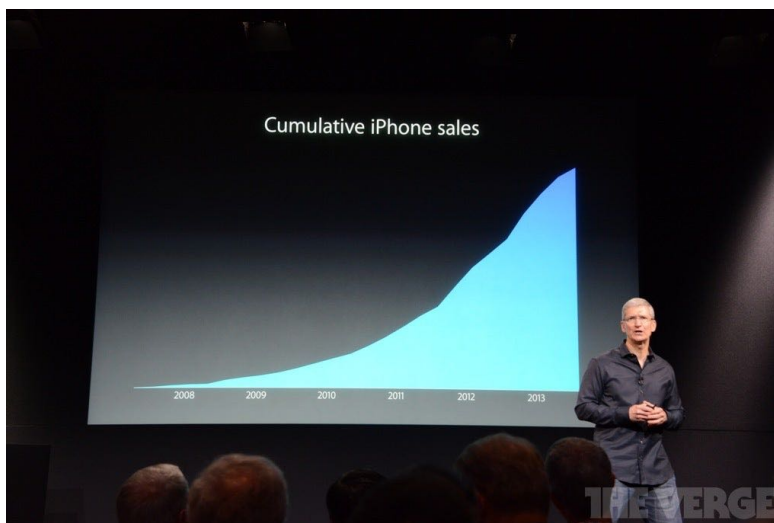


Figure 5.59: Apple, 2013: cumulative iPhone sales, a curve that can only rise, shown without a y-axis. Source: [The Verge](#), via [Quartz](#).

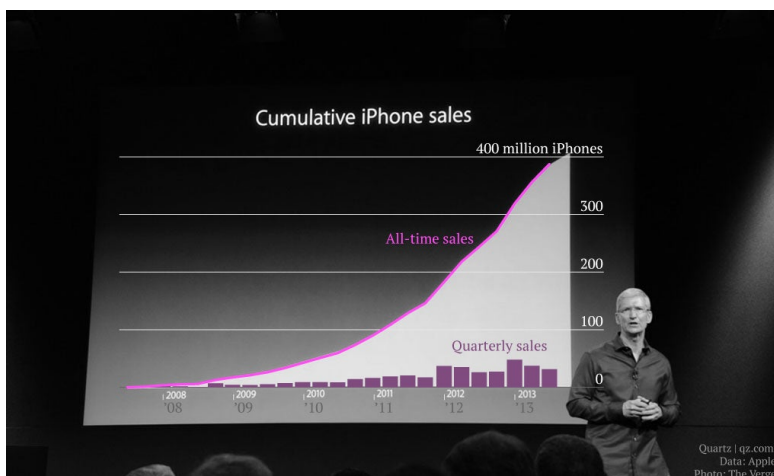


Figure 5.60: The same curve with quarterly sales added: sales had fallen for three quarters. Source: [Quartz](#).

Figure 1: COVID-19 Related Deaths in United States Between February 15th and April 18th in a linear scale (left panel) and in a log scale (right panel). Source: www.worldometers.info

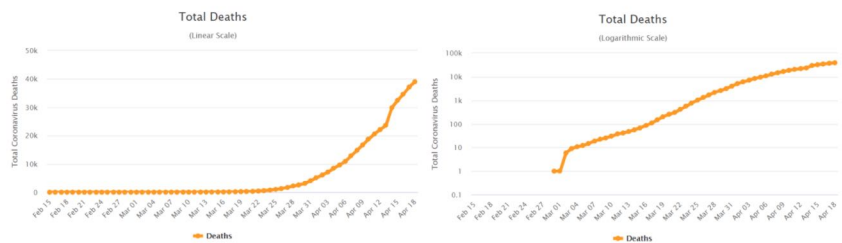


Figure 5.61: COVID-19 deaths in the United States, February 15 to April 18, 2020, on a linear scale (left) and a log scale (right). In an experiment, readers shown the log version predicted the trend less accurately. Source: Romano et al. (2020).

Map Projections

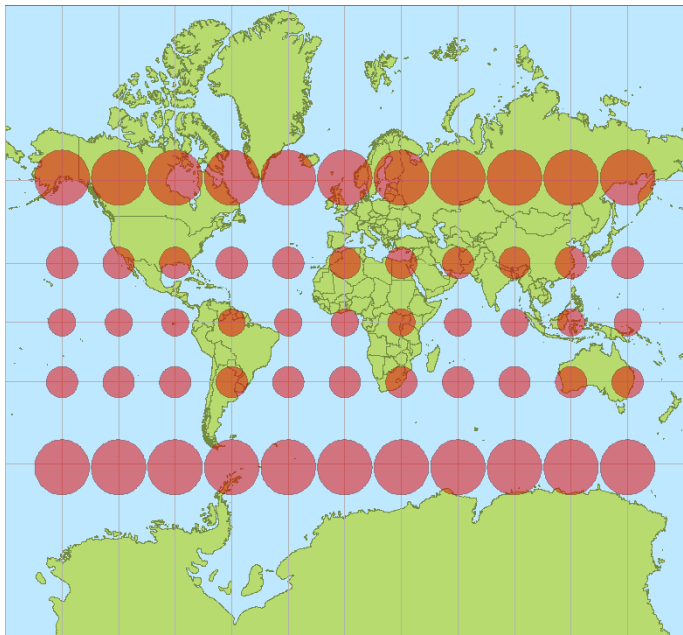


Figure 5.62: The Mercator projection with Tissot's indicatrices: each circle covers the same area on the globe, so their growth toward the poles is distortion. Source: Stefan Kuehn, [Wikimedia Commons](https://commons.wikimedia.org/wiki/File:Mercator_projection_Tissot_indicatrices.png), CC BY-SA 3.0.

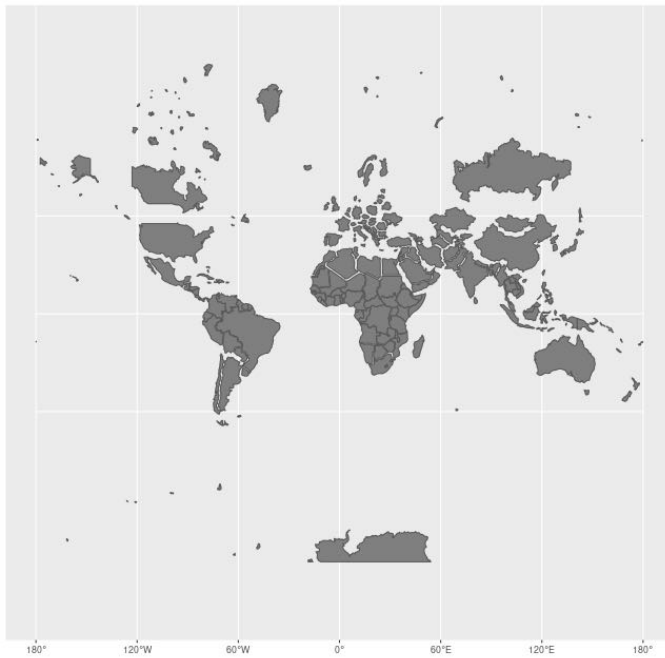


Figure 5.63: Countries drawn at their true relative size inside their Mercator outlines; a frame from an animation. Source: Jakub Nowosad, [Wikimedia Commons](#), CC BY-SA 4.0.

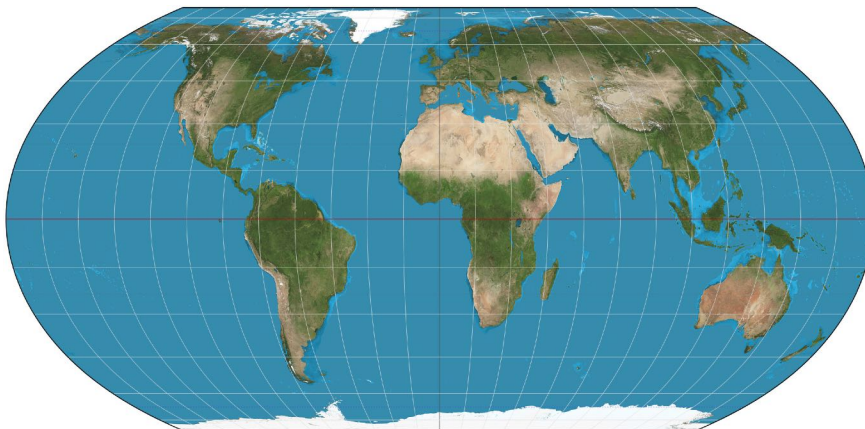


Figure 5.64: The Equal Earth projection keeps every area in true proportion. Source: Strebe, [Wikimedia Commons](#), CC BY-SA 4.0.

Aggregation Hides the Groups

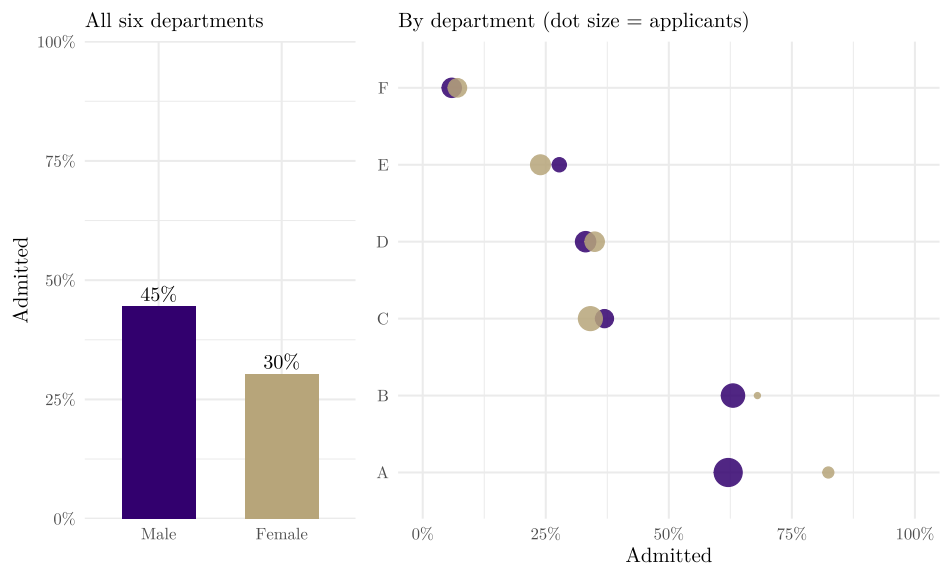


Figure 5.65: Berkeley graduate admissions, 1973, six largest departments. Overall, 45% of men and 30% of women were admitted, yet women were admitted at a higher rate in four of the six departments: most women applied to departments that admit few applicants. Data: R's UCBA admissions (Bickel et al. 1975).

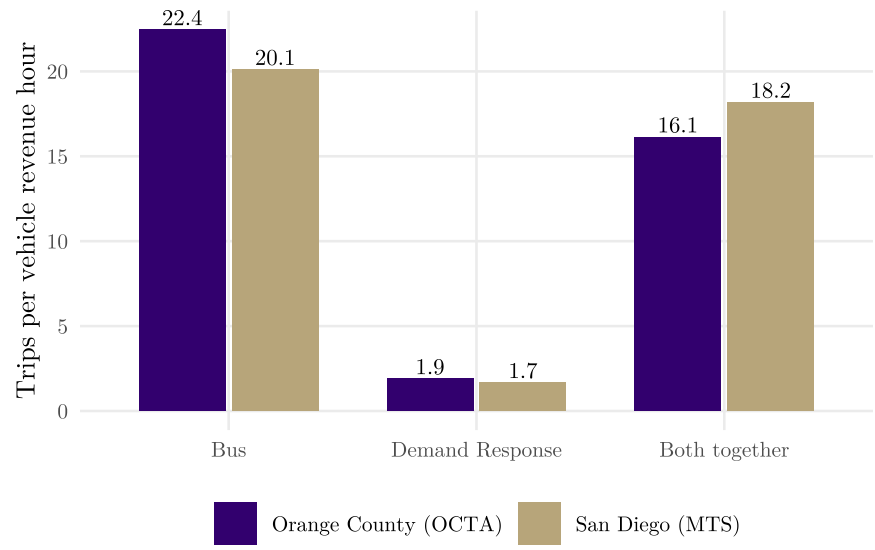


Figure 5.66: Orange County (OCTA) carries more riders per service hour than San Diego (MTS) in both bus and demand response, yet fewer in the two together: 31% of OCTA's hours go to demand response, against 10% of MTS's. Bus and demand response only, 2025. Source: National Transit Database.

Rainbow Color Scales

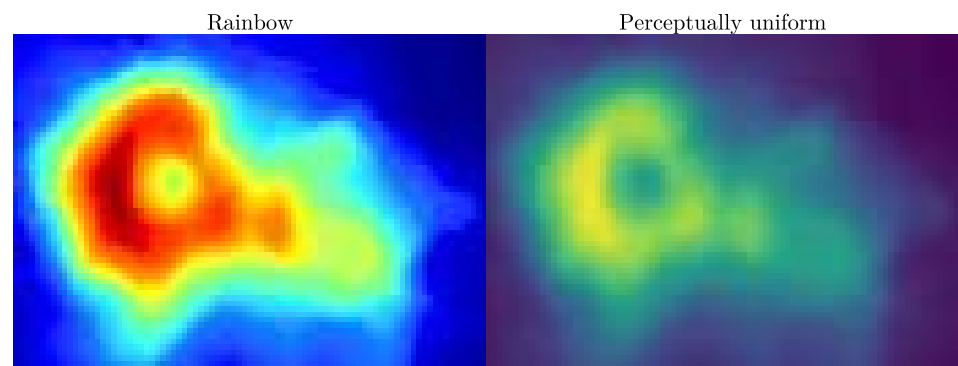


Figure 5.67: The same elevation data (R's volcano) with a rainbow palette (left), which draws bright bands where nothing changes sharply, and a perceptually uniform palette (right) (Borland and Taylor 2007).

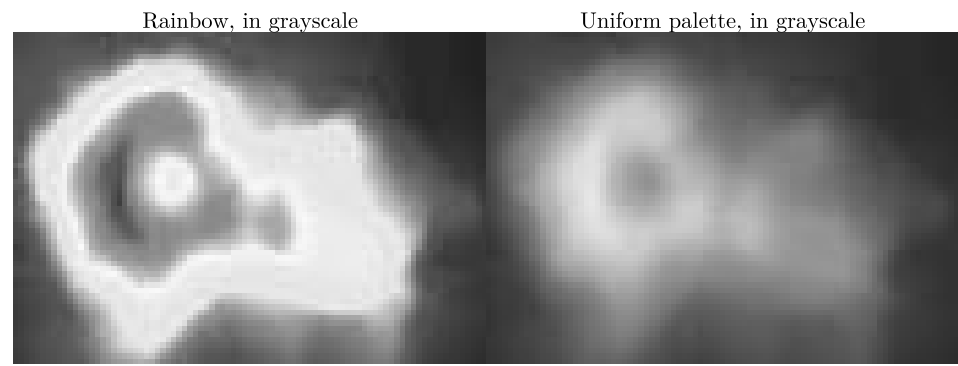


Figure 5.68: Printed in grayscale, the rainbow version of the volcano loses its order, while the uniform palette keeps it.