

Data Visualization

Vaibhav P. Vasani

Assistant Professor

Department of Computer Engineering

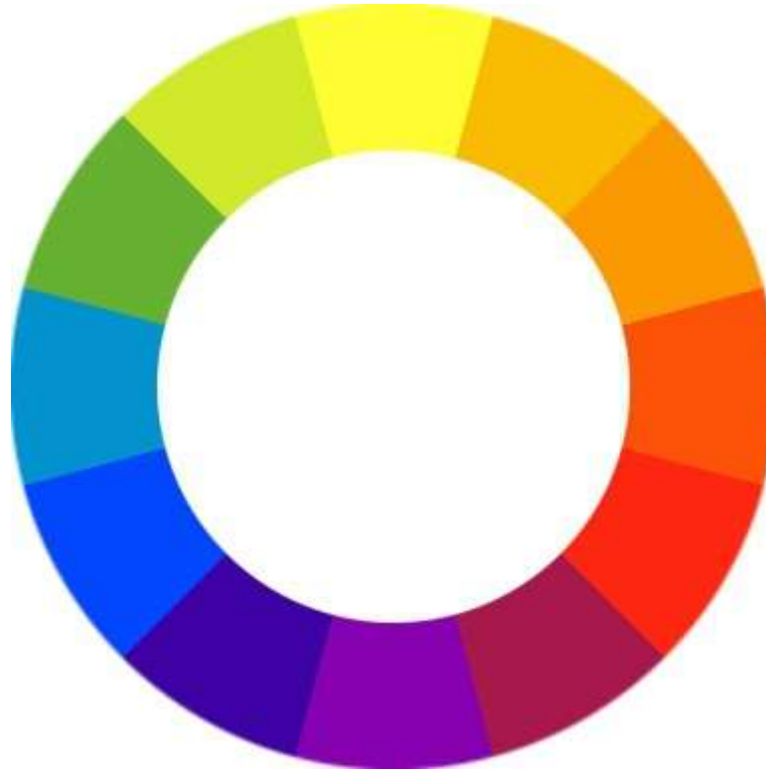
K. J. Somaiya College of Engineering

Somaiya Vidyavihar University

Using color effectively

- Color falls low on the perceptual hierarchy of visual cues, Still deployed to highlight particular elements of a chart, and sometimes to encode data values.
- Poor choice of color schemes is a problem that bedevils many news graphics, so it is worth taking some time to consider how to use color to maximum effect.
- It helps to think about colors in terms of the color wheel, which places colors that “harmonize” well together side by side, and arranges those that have strong visual contrast — blue and orange, for instance — at opposite sides of the circle:



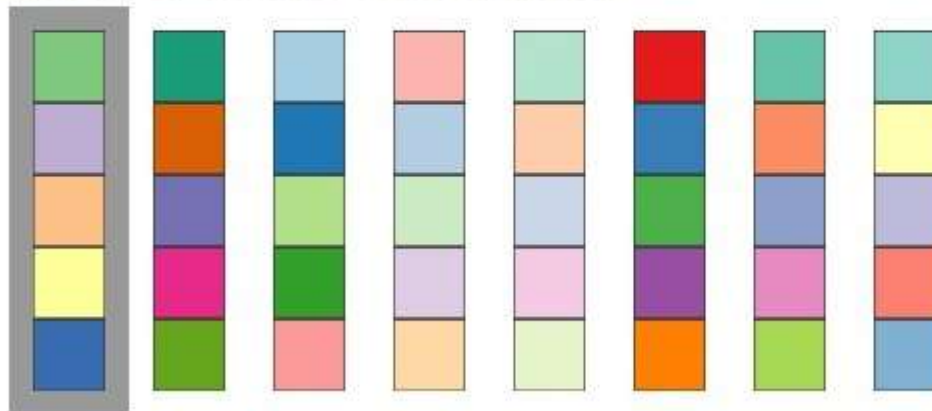


(Source: [Wikimedia Commons](#))

- When encoding data with color, take care to fit the color scheme to your data, and the story you're aiming to tell.
- Color is often used to encode the values of categorical data.
- Here you want to use “qualitative” color schemes, where the aim is to pick colors that will be maximally distinctive, as widely spread around the color wheel as possible:

☐ sequential ☐ diverging ☒ qualitative

Pick a color scheme:



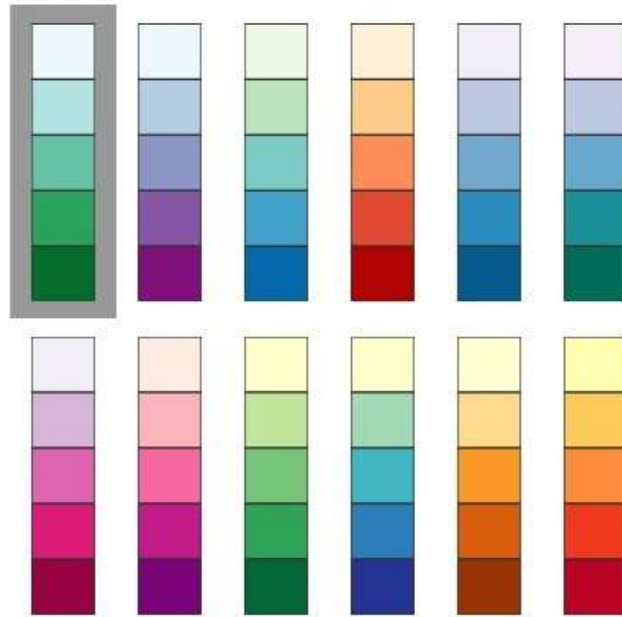
(Source: [ColorBrewer](https://colorbrewer2.org/))

- When using color to encode continuous data, it usually makes sense to use increasing intensity, or saturation of color to indicate larger values.
- These are called “sequential” color schemes:

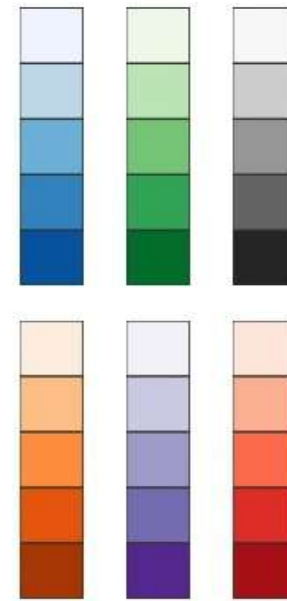
☒ sequential
 ☐ diverging
 ☐ qualitative

Pick a color scheme:

Multi-hue:



Single hue:

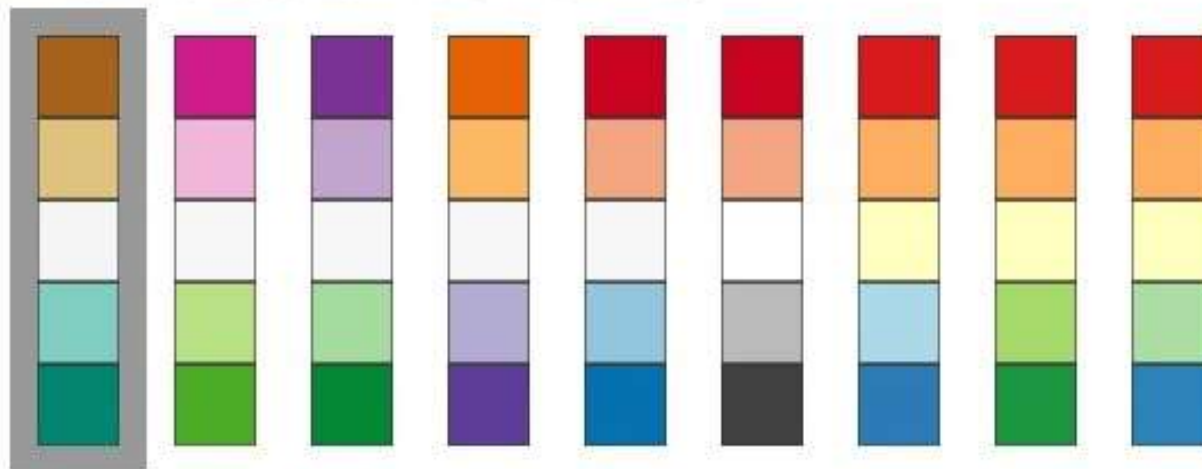


(Source: [ColorBrewer](https://colorbrewer2.org/))

- In some circumstances, you may have data that has positive and negative values, or which highlights deviation from a central value.
- Here, you should use a “diverging” color scheme, which will usually have two colors reasonably well separated on the color wheel as its end points, and cycle through a neutral color in the middle:

☐ sequential ☒ diverging ☐ qualitative

Pick a color scheme:



(Source: [ColorBrewer](https://colorbrewer2.org/))

- Choosing color schemes is a complex science and art
- Many visualization tools include suggested color palettes
- [ColorBrewer](#), Originally designed for maps, but useful for charts in general, these color schemes have been rigorously tested to be maximally informative.

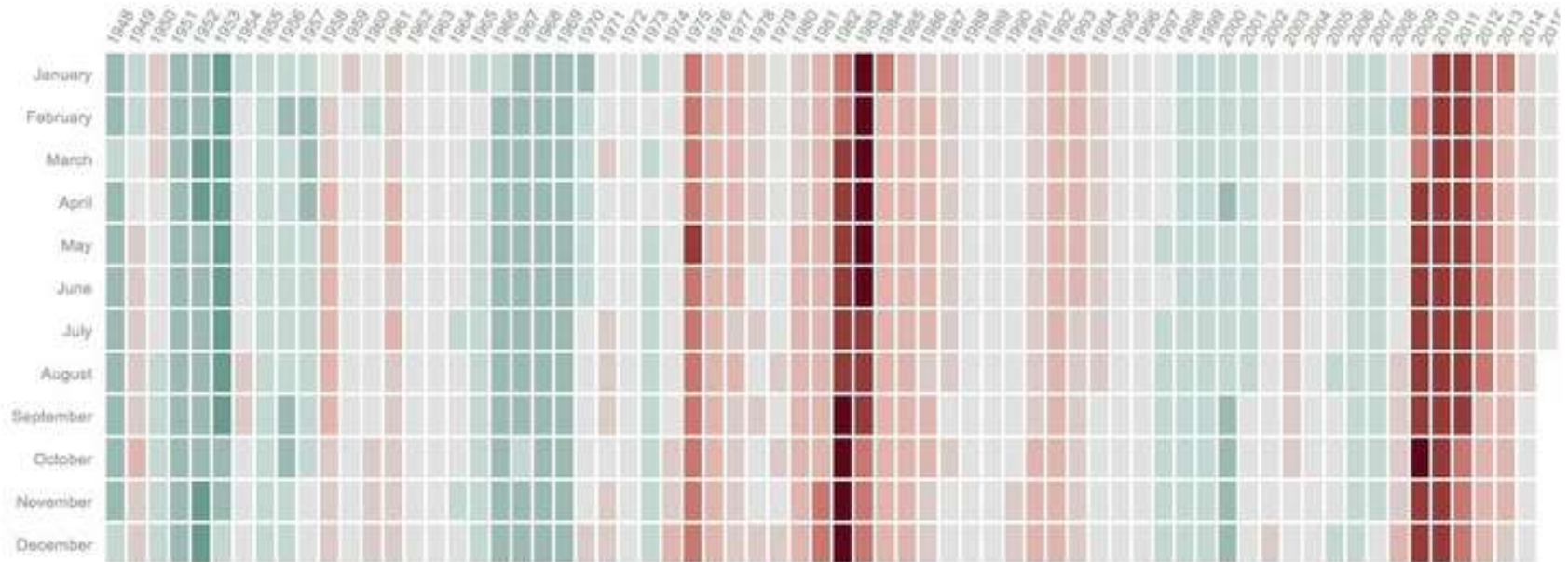


- It suggests can be displayed according to their values on three color “models”: HEX, RGB and CMYK.
 - **RGB** Three values, describing a color in terms of combinations of red, green, and blue light, with each scale ranging from 0 to 255; sometimes extended to RGB(A), where A is alpha, which encodes transparency. Example: `rgb(169, 104, 54)`.
 - **HEX** A six-figure “hexadecimal” encoding of RGB values, with each scale ranging from hex 00 (equivalent to 0) to hex ff (equivalent to 255); HEX values will be familiar if you have any experience with web design, as they are commonly used to denote color in HTML and CSS. Example: `#a96836`
 - **CMYK** Four values, describing a color in combinations of cyan, magenta, yellow and black, relevant to the combination of print inks. Example: `cmyk(0, 0.385, 0.68, 0.337)`
 - **HSL** Three values, describing a color in terms of hue, saturation and lightness (running from black, through the color in question, to white). Hue is the position on a blended version of the color wheel in degrees around the circle ranging from 0 to 360, where 0 is red. Saturation and lightness are given as percentages. Example: `hsl(26.1, 51.6%, 43.7%)`
 - **HSV/B** Similar to HSL, except that brightness (sometimes called value) replaces lightness, running from black to the color in question. `hsv(26.1, 68.07%, 66.25%)`
- [Colorizer](#) is one of several web apps for picking colors and converting values from one model to another.

- Custom color schemes can also work well, but experiment to see how different colors influence your story.
- The following graphic from *The Wall Street Journal*, for instance, uses an unusual pseudo-diverging scheme to encode data — the US unemployment rate — that would typically be represented using a sequential color scheme.
- It has the effect of strongly highlighting periods where the jobless rate rises to around 10%, which flow like rivers of blood through the graphic.

U.S. Unemployment: A Historical View

Track the national unemployment rate since 1948 – the first year in which the government provides data that can reliably be compared with the current rate. Numbers are seasonally adjusted.



Sources: Bureau of Labor Statistics; Current Population Survey. Updated: July 2, 2015

(Source: [The Wall Street Journal](#))

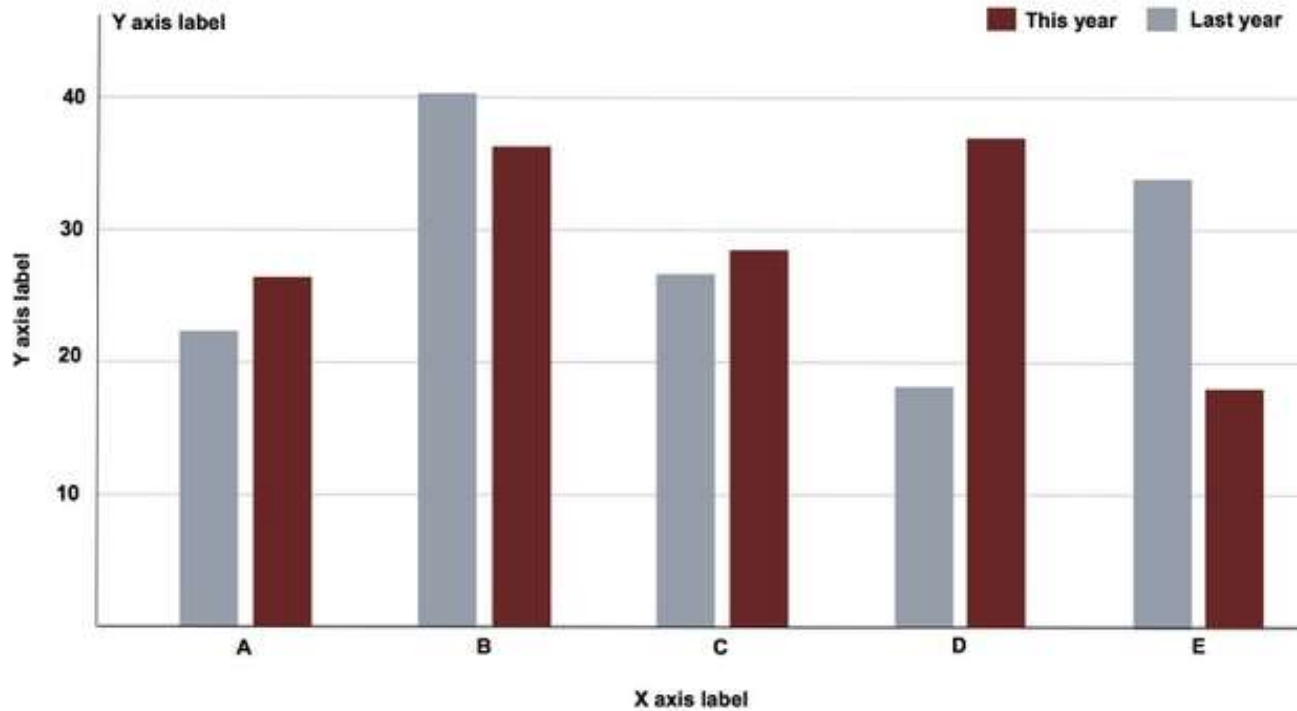


Using chart furniture, minimizing chart junk, highlighting the story

- **Title and subtitle** These provide context for the chart.
- **Coordinate system** For most charts, this is provided by the horizontal and vertical axes, giving a cartesian system defined by X and Y coordinates; for a pie chart it is provided by angles around a circle, called a polar coordinate system.
- **Scale** Labeled tick marks and grid lines can help your audience read data values.
- **Labels** : Label each axis. Other labels that may be necessary to explain the message of graphic.
- **Legend**: Color or shape to encode data, then legends are needed to explain encoding.
- **Source information** Usually given as a footnote. Don't forget this!

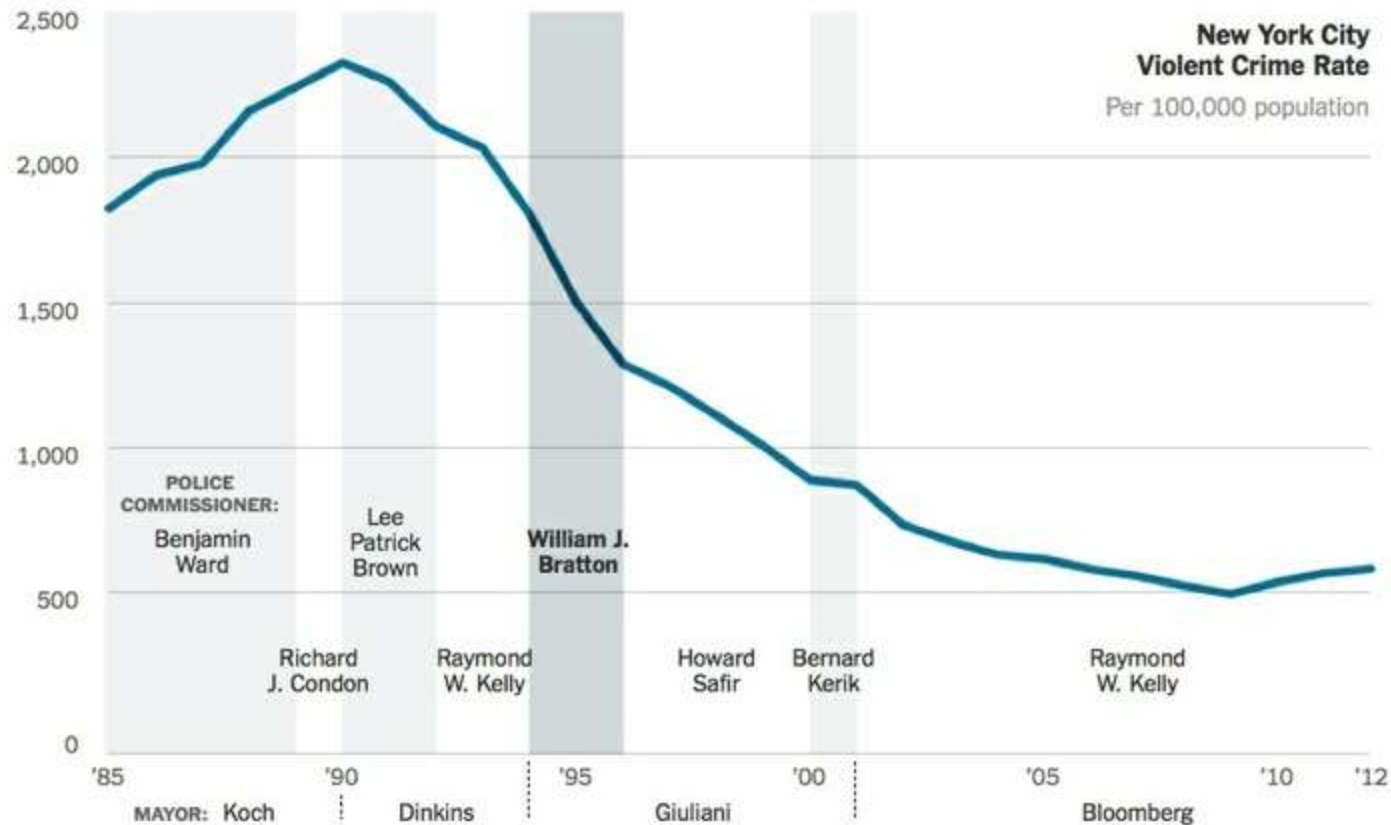
A title for the chart

And a subtitle, telling us some more about what it shows.



Source information

- Chart furniture can also be used to encode data, as in this example, which shows the terms of New York City's police commissioners and mayors with reference to the time scale on the X axis:

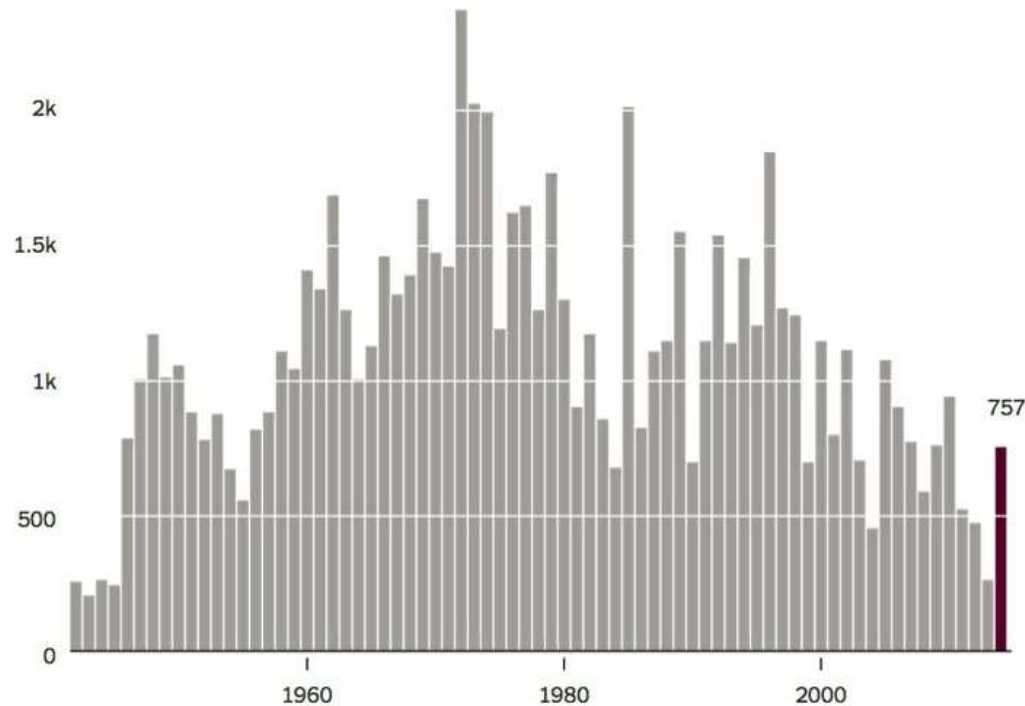


(Source: [The New York Times](#))

Here is a nice example of a graphic that minimizes chart junk, and maximizes data-ink. Notice how the Y axis doesn't need to be drawn, and the gridlines are an absence of ink, consisting of white lines passing through the columns:

Airline Fatalities Since 1947, Updated

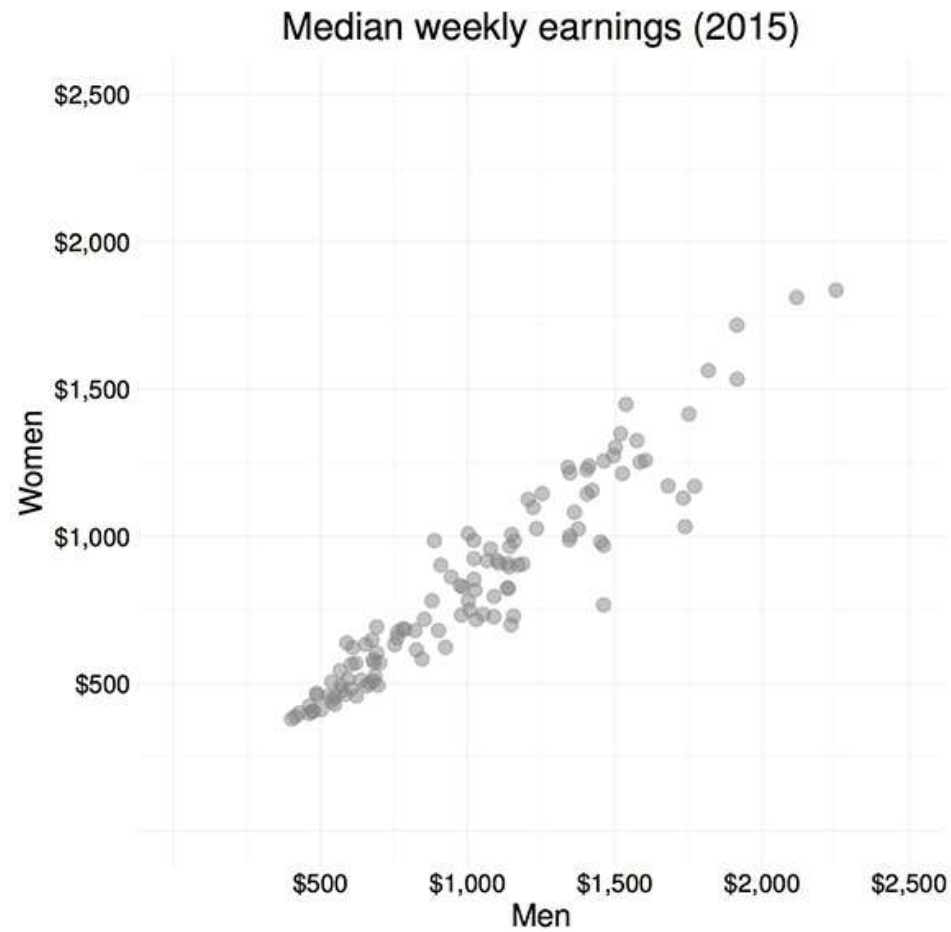
Worldwide airline deaths have fallen sharply in recent decades. This year will be an aberration.



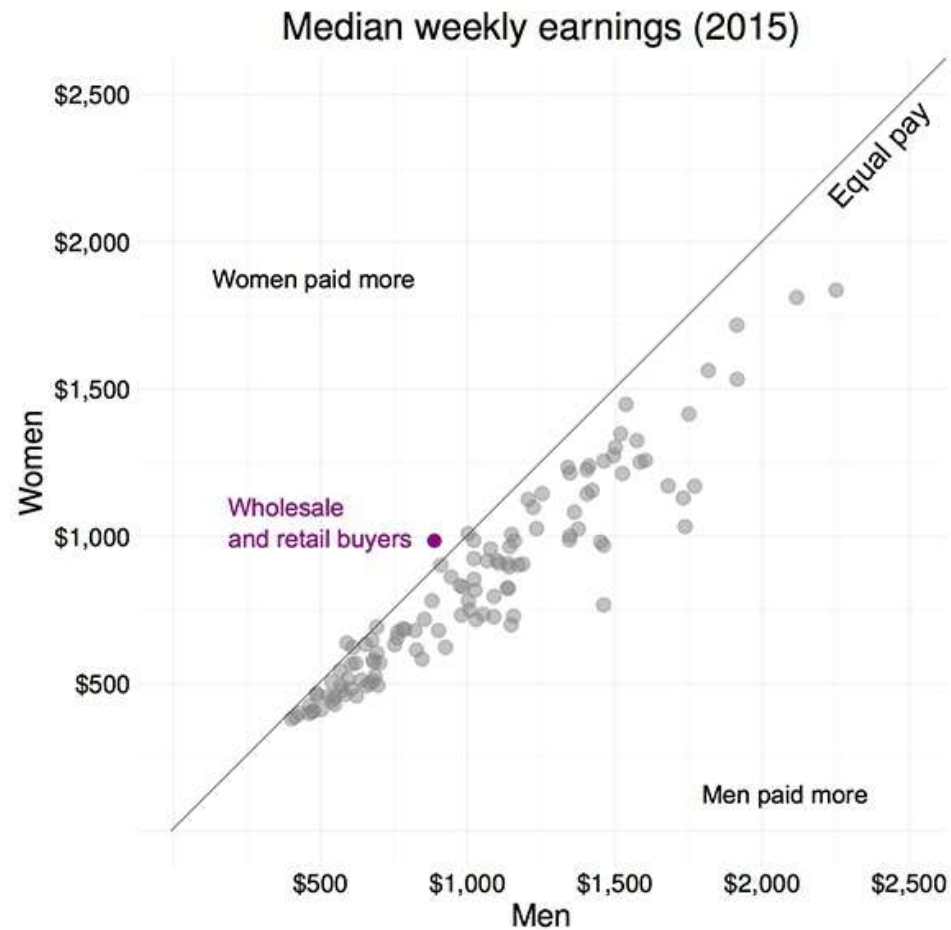
Source: Aviation Safety Network; news reports

Data for 2014 includes TransAsia and Air Algerie crashes. Fatality data is based on flights with 14 or more passengers and does not include corporate or military flights.

(Source: The Upshot, [The New York Times](#))



(Source: Peter Aldhous, from [Bureau of Labor Statistics](#) data)



(Source: Peter Aldhous, from [Bureau of Labor Statistics](#) data)



(Source: Peter Aldhous, from [Bureau of Labor Statistics](#) data)

Pitfalls to avoid

- If you ever decide to encode data using area, be sure to do so correctly.
- If one unit is a square with sides of length one, then the correct way to represent a value of four is a square with sides of length two ($2*2 = 4$), not a square with sides of length four ($4*4 = 16$).
- Mistakes are frequently made, however, when encoding data by the area of circles.
- In 2011, for instance, President Barack Obama's State of the Union Address for the first time included an “enhanced” online version with supporting data visualizations.
- This included the following chart, comparing US Gross Domestic Product to that of competing nations:



AMERICA'S ECONOMY

2010 GROSS DOMESTIC PRODUCT



(Source: [The 2011 State of the Union Address: Enhanced Version](#))

- Data-savvy bloggers were quick to point out that the data had been scaled by the radius of each circle, not its area. Because $\text{area} = \pi * \text{radius}^2$, you need to scale the circles by the square root of the radius to achieve the correct result, on the right:



AMERICA'S ECONOMY

2010 GROSS DOMESTIC PRODUCT



United States
\$14.6 TRILLION

China **\$5.7 TRILLION**

Japan **\$5.3 TRILLION**

Germany **\$3.3 TRILLION**

France **\$2.5 TRILLION**



Tweet This

AMERICA'S ECONOMY

2010 GROSS DOMESTIC PRODUCT



United States
\$14.6 TRILLION

China **\$5.7 TRILLION**

Japan **\$5.3 TRILLION**

Germany **\$3.3 TRILLION**

France **\$2.5 TRILLION**



Tweet This

(Source: [Fast Fedora](#) blog)

- Many software packages (Microsoft Excel is a notable culprit) allow users to create charts with 3-D effects.
- Some graphic designers produce customized charts with similar aesthetics.
- The problem is that that it is very hard to read the data values from 3-D representations, as this example illustrates:



BRIDGE OVER TROUBLED WATERS



In its third annual study of 160 countries, the Charity Aid Foundation measured giving behavior across three criteria - volunteering, helping strangers, and donating money. Although the results show that charitable giving - often a mirror of global economic patterns - has predictably declined since 2007, surprising findings went among the top 20.

THE MOST AFFLUENT COUNTRIES
AREN'T NECESSARILY THE MOST PHILANTHROPIC.
Only 6 of the countries with the world's top 20 GDP made it to the Charity Aid Foundation's top 20 list.



GLOBALITY, average participation in giving has fallen since 2007.



Source: Charity Aid Foundation World Giving Index 2012

AUSTRALIA IS THE MOST GENEROUS NATION IN THE WORLD.

The government's proactive effort to encourage philanthropy may be the reason the country has landed the highest score on average over the past five years.



IN EVERY COUNTRY OF THE TOP 10,
volunteering was the least common giving behavior.



THE TOP 9 COUNTRIES
HAVE WORLD-GIVING INDEX SCORES OF 50% OR MORE.
This means that at least half the population, on average, is taking part in at least 1 of the 3 giving behaviors on a regular basis.

KEY

- ▲ MONEY DONATIONS
- ▲ VOLUNTEERING
- ▲ HELPING STRANGERS
- ▲ TOP 10
- ▲ TOP 9
- ▲ ASIAN COUNTRIES
- ▲ AFFLUENT COUNTRIES

THERE IS DIVERSITY IN GENEROSITY

GEOGRAPHIC LOCATION
Every continent is represented.

POPULATION
300 MILLION PEOPLE
United States
1.3 MILLION PEOPLE
Timor and Togo

ECONOMIC STRENGTH
2ND LOWEST GDP PER CAPITA
Liberia
2ND HIGHEST GDP PER CAPITA
Qatar

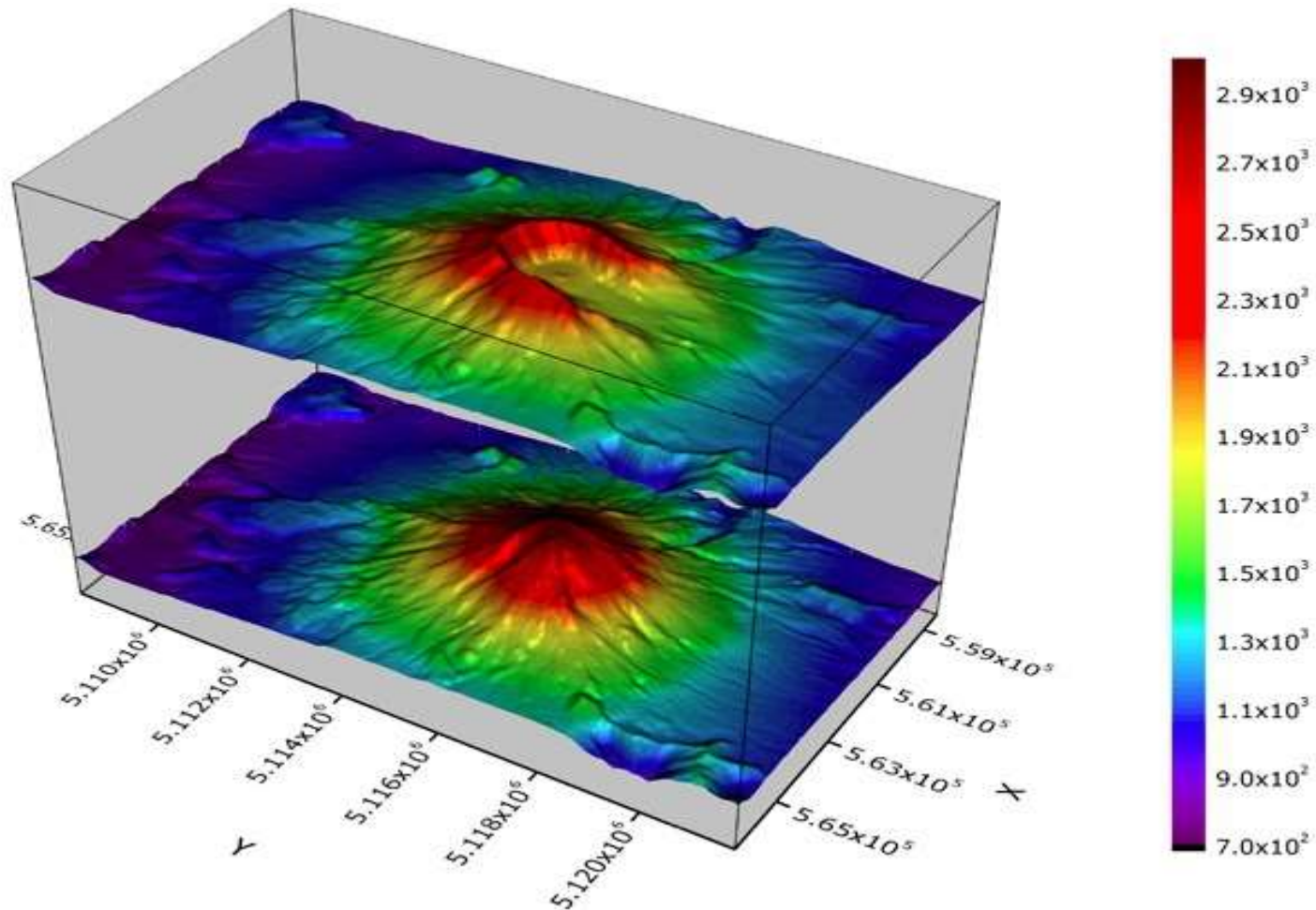


8 OUT OF THE TOP 20
COUNTRIES ARE FROM ASIA.

A COLLABORATION BETWEEN GOOD AND COUNTRY FIVE

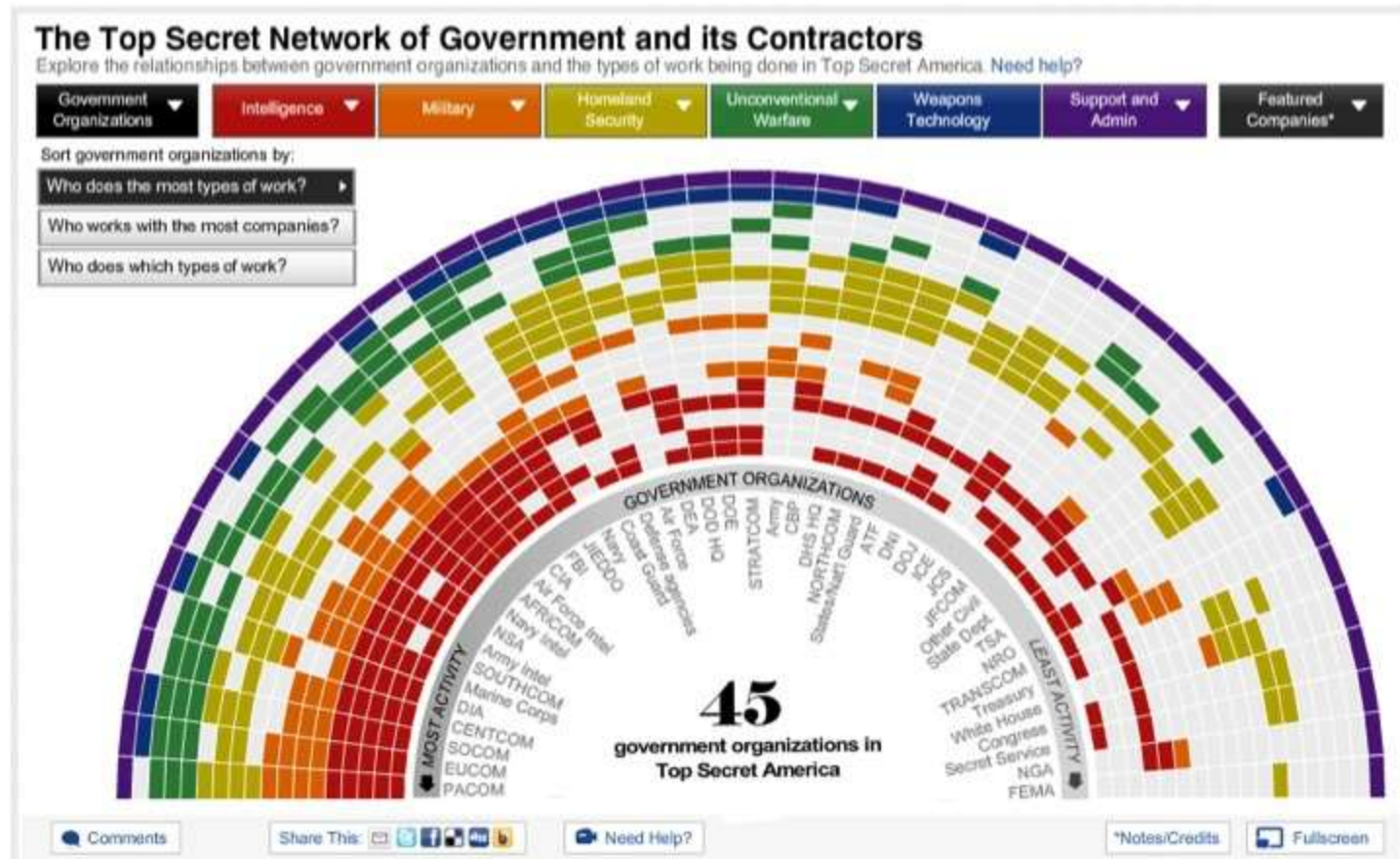
(Source: [Good](#))

- A good rule of thumb for data visualization is that trying to represent three dimensions on a two dimensional printed or web page is almost always one dimension too many, except in unusual circumstances, such as these representations of Mount St. Helens in Washington State, before and after its 1980 eruption:



(Source: [OriginLab](#))

- Sometimes even leading media outlets lose their way. See if you can make sense of this interactive graphic on clandestine US government agencies and their contractors:



(Source: [The Washington Post](#))

Be true to the ‘feel’ of the data

- Think about what the data represents in the real world, and use chart forms, visual encodings and color schemes that allow the audience’s senses to get close to what the data means — note again the “rivers of blood” running through *The Wall Street Journal*’s unemployment chart, which suggest human suffering.



Published: February 26, 2010

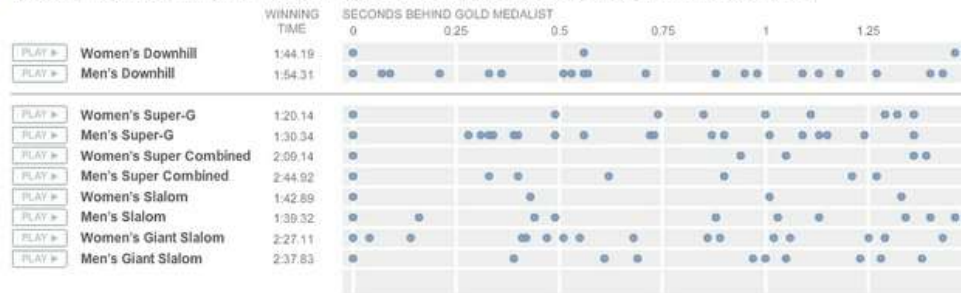
Fractions of a Second: An Olympic Musical

At the Olympics, the blink of an eye can be all that separates the gold medalist from the 10th-place finisher. In some events, this is obvious. But in others, with athletes racing one by one, the closeness of the race is harder to perceive. Listen to the differences below.



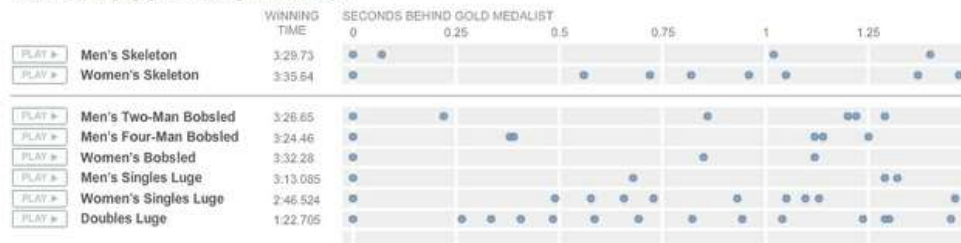
Alpine skiing

The women's downhill course was extremely tiring, and, because it was more challenging than the men's course, it ended up separating the skiers by much larger margins. This pattern appears in the two speed events: the downhill and the super-G.



Skeleton, Bobsled and Luge

In percentage terms, the men's skeleton had one of the tightest finishes in Vancouver, with only .07 of a second separating the top two finishers across a three-and-a-half-minute run. But the difficult track produced speeds higher than expected, and many sliding events had relatively large gaps between gold and silver.



(Source: [The New York Times](#))

Break the story down into scenes

- Many stories have a step-by-step narrative, and different charts may tell different parts of the story. So think about communicating such stories through a series of graphics. This is another good reason to experiment with different chart types when exploring a new dataset. Here is a nice example of this approach, examining demographic change in Brazil:

Primeiro Plano

Diagrama

A NOTÍCIA EM PERSPECTIVA

A oportunidade demográfica

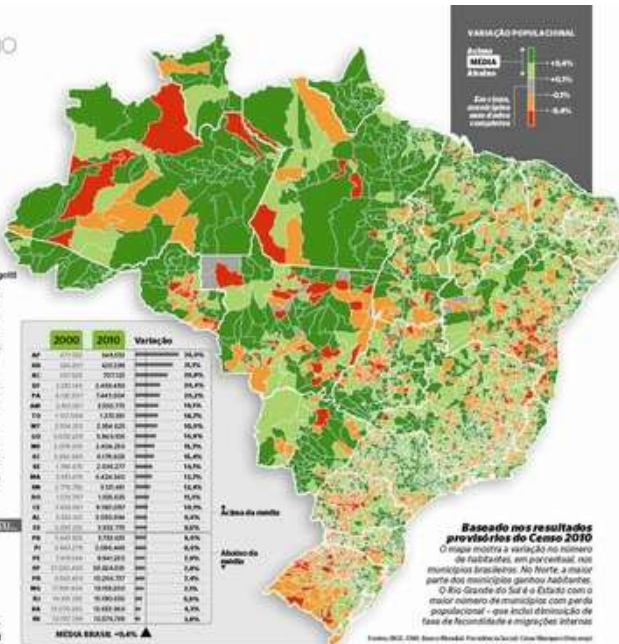
Como o Brasil pode aproveitar hoje um futuro com menos filhos por casal, menos jovens e mais idosos.

Alberto Cairo, Francisco Lima e Marco Vergotti

OS DADOS PREVISÓRIOS DO CENSO DE 2010 já permitem estimar um futuro da evolução da população no Brasil. Os números finais deverão ficar um pouco diferentes, mas as conclusões permanecerão válidas. A população cresceu cerca de 10% em dez anos, mas a fecundidade está abaixo de 2,1 filhos por mulher (limite para que a população continue constante). Índice de uma queda no crescimento a partir de 2030, segundo César Marques, pesquisador da Unicamp. Um país com menos jovens, mais gente madura – e mais idosos – tem necessidades prementes: educação de qualidade, mercado de trabalho desenvolvido, meio ambiente protegido e sistema previdenciário sustentável.

1. A POPULAÇÃO BRASILEIRA CRESCERÁ

Os dados do Censo do IBGE indicam por enquanto um crescimento populacional de 3,4% no Brasil entre 2000 e 2010. O valor para 2050 poderá variar em cerca de 5 milhões de pessoas nos resultados finais, segundo César Marques.



Good practice for interactives

- Nowadays the primary publication medium for many news graphics is the web or apps on mobile platforms, rather than print, which opens up many possibilities for interactivity.
- Greatly enhances ability to tell a story, but it also creates new possibilities to confuse and distract your audience — think of this as interactive chart junk.
- A good general approach for interactive graphics is to provide an overview first, and then allow the interested user to zoom or filter to dig deeper into the data.
- In such cases, the starting state for an interactive should tell a clear story: If users have to make an effort to dig into a graphic to get anything from it, few are likely to do so. Indeed, assume that much of your audience will spend only a short time interacting with the data. [“How Different Groups Spend Their Day”](#) from *The New York Times* is a good example of this approach.

Good practice for interactives

- Similarly, don't hide labels or information essential to understanding the graphic in tooltips that are accessed only on clicks or hovers.
- This is where to put more detailed information for users who have sufficient interest to explore further.
- Make the controls for an interactive obvious — play buttons should look like play buttons, for instance. You can include a few words of explanation, but only a very few: as far as possible, how to use the interactive should be intuitive, and built into its design.

- The interactivity of the web also facilitates a scene-by-scene narrative — a device employed frequently by The New York Times' graphics team in recent years.
- At New Scientist, used this approach for this interactive, exploring the likely number of Earth-like planets in our Galaxy:

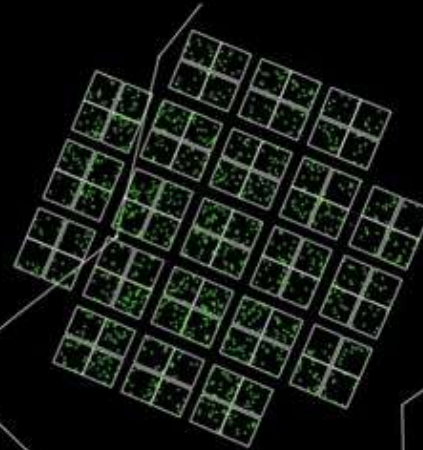
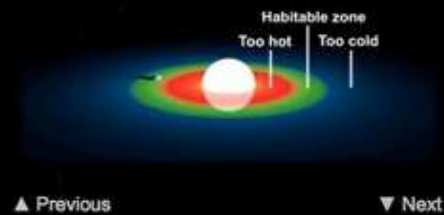


In the Goldilocks zone?

The remaining planets are the right size, but not all of them will be ripe for life, with life-sustaining liquid water.

We can't determine a planet's composition from Kepler's data, but we can see whether a planet seems to orbit in the habitable zone around its parent star.

1696 planets



(Source: [New Scientist](#))

‘Mobile-first’ may change your approach

- Increasingly, news content is being viewed on mobile devices with small screens
- At the most basic level, this means making graphics “responsive,” so that their size adjusts to screen size. But there is more to effective design for mobile than this.
- We have already discussed the value of small multiples, which can be made to reflow for different screen sizes.



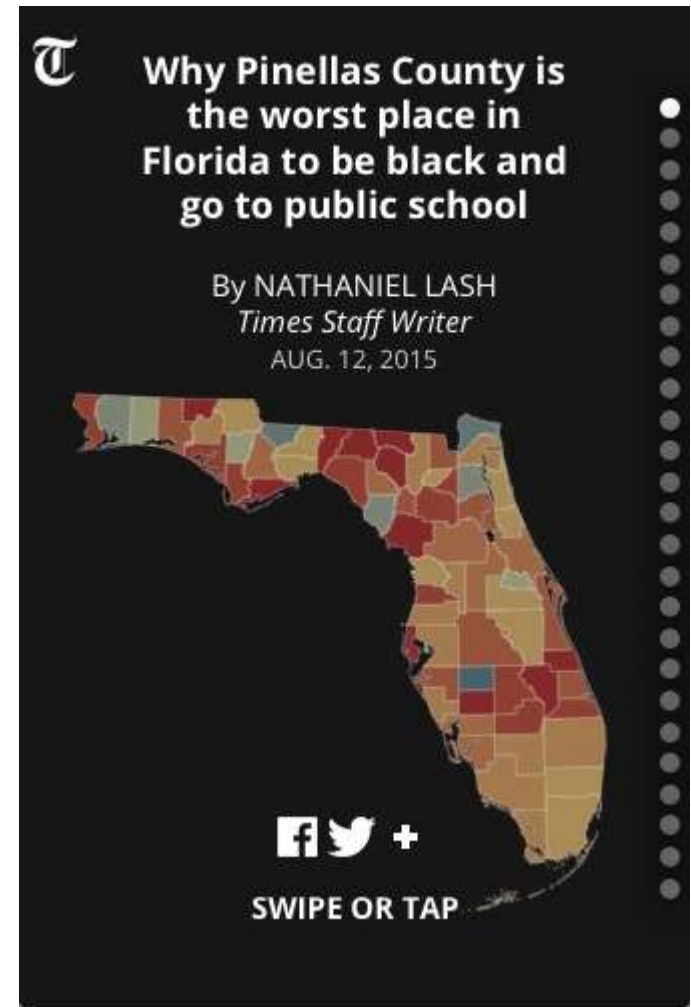
- [This interactive](#), exploring spending on incarceration by block in Chicago, is a nice example of organizing and displaying the same material differently for different screen sizes. Open it up on your laptop then reduce the size of your browser window to see how it behaves.





(Source: [DataMade](#))

- Again, a step-by-step narrative can be a useful device in overcoming the limitations of a small screen. This interactive, exploring school segregation by race in Florida, is a good example of this approach:



Be careful with animation

- Animation in interactives can be very effective. But remember the goal of staying true to the ‘feel’ of the data. Animated images evolve over time, so animation can be particularly useful to encode data that changes over time. But again you need to think about what the human brain is able to perceive. Research has shown that people have trouble tracking more than about four points at a time. Try playing Gapminder World without the energetic audio commentary of Hans Rosling’s “200 Countries” video, and see whether the story told by the data is clear.
- Animated transitions between different states of a graphic can be pleasing. But overdo it, and you’re into the realm of annoying Powerpoint presentations with items zooming into slides with distracting animation effects. It’s also possible for elegant animated transitions to “steal the show” from the story told by the data, which arguably is the case for this exploration by The New York Times of President Obama’s 2013 budget request to Congress:



UPDATED February 12, 2012.

TWITTER LINKEDIN SHARE

Four Ways to Slice Obama's 2013 Budget Proposal

Explore every nook and cranny of President Obama's federal budget proposal.

All Spending

Types of Spending

Changes

Department Totals

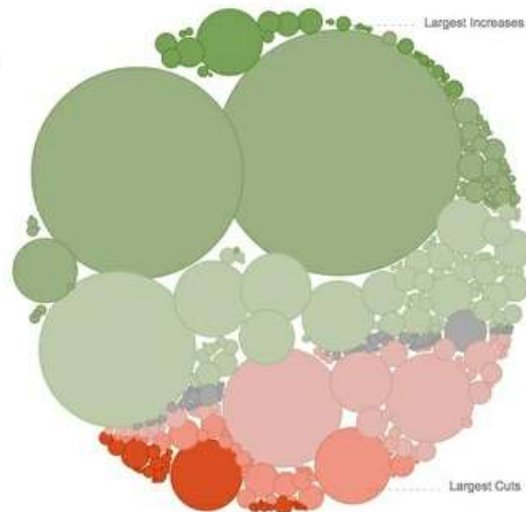
How \$3.7 Trillion Is Spent

Mr. Obama's budget proposal includes \$3.7 trillion in spending in 2013, and forecasts a \$901 billion deficit.

Circles are sized according to the proposed spending.



Color shows amount of cut or increase from 2012.



The proposal forecasts a \$901 billion deficit.

Chart shows \$3.7 trillion authorized to be spent in 2013. (Total spending is estimated to be \$3.8 trillion, including funds authorized in other years). Negative budget authority, which results from fees or other collections, is shown only on the department totals tab, but is included in other totals.

By SHAN CARTER | Send Feedback

Office of Management and Budget

(Source: [The New York Times](#))

Sketch and experiment to find the story

- There are many ways of visualizing the same data. Effective graphics and interactives do not usually emerge fully formed.
- They usually arise through sketching and experimentation.
- As you sketch and experiment with data, use the framework suggested by the chart selector thought-starter to prioritize different chart types, and always keep the perceptual hierarchy of visual cues at the front of your mind.
- Remember the mantra: Design for the human brain!
- Also, show your experiments to friends and colleagues.
- If people are confused or don't see the story, you may need to try a different approach.



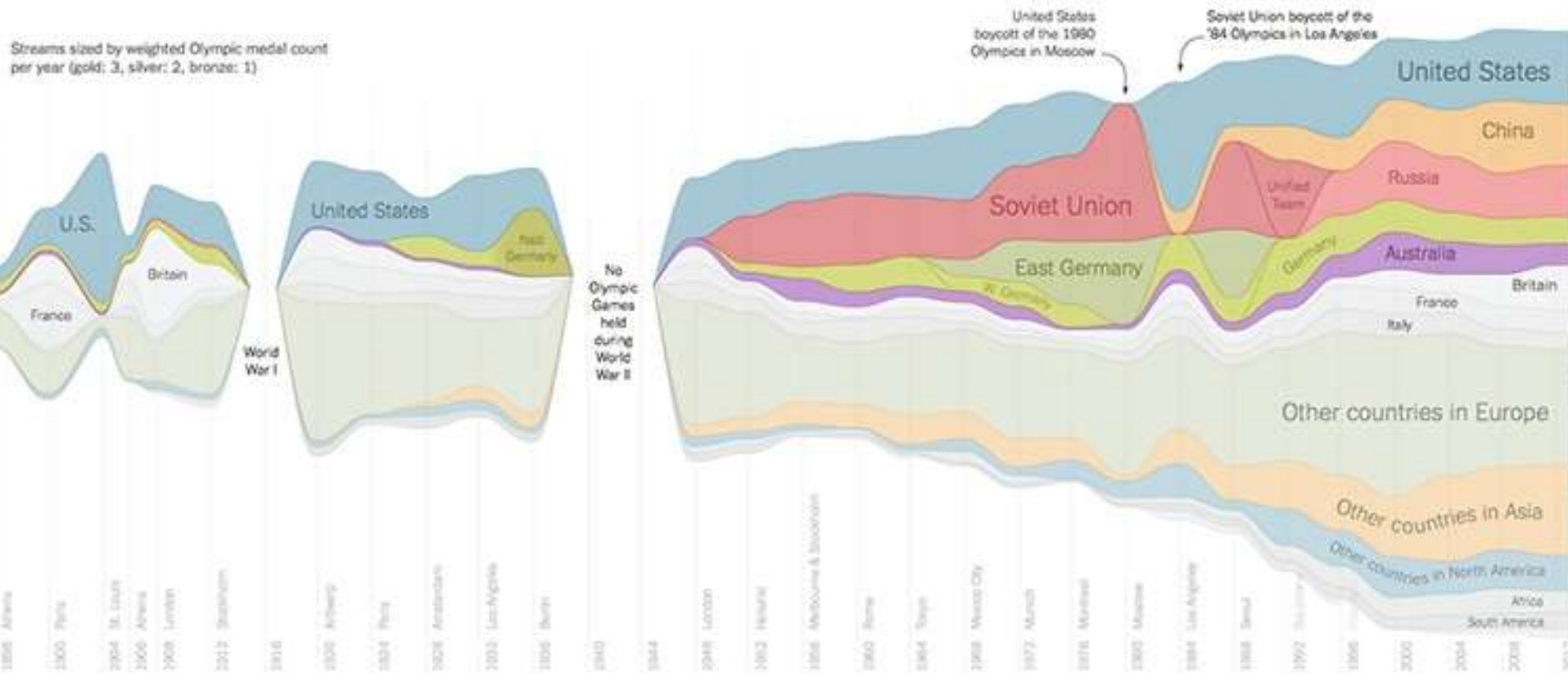
Learn from the experts

- Over the coming weeks and beyond, make a habit of looking for innovative graphics, especially those employing unusual chart forms, that communicate the story from data in an effective way.
- Work out how they use visual cues to encode data. Here are a couple of examples from *The New York Times* to get you started.



A Visual History of Which Countries Have Dominated the Summer Olympics

By GREGOR AISCH and LARRY BUCHANAN AUG. 8, 2006



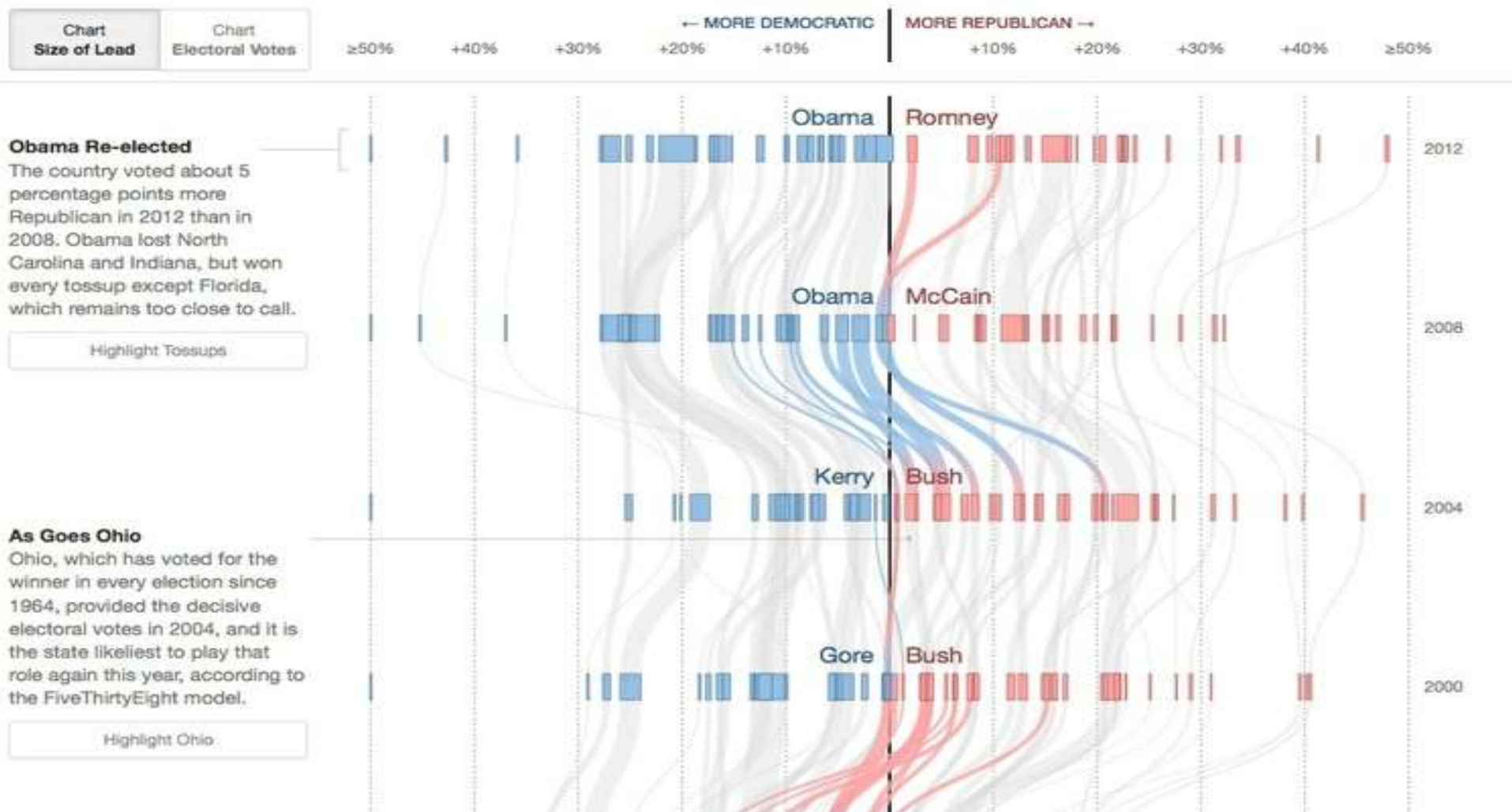
(Source: [The New York Times](#))

Over the Decades, How States Have Shifted

Recent elections have placed a heavy emphasis on “swing states” — Ohio, Florida and the other competitive states. Yet in the past, many more states shifted between the Democratic and Republican parties. A look at how the states stacked up in the 2012 election and how they have shifted over past elections.

Each box represents a state sized by number of electoral votes.

Each curve shows how much it shifted left or right between elections.

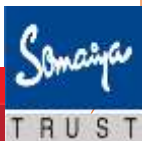


(Source: [The New York Times](#))



SOMAIYA
VIDYAVIHAR UNIVERSITY

K J Somaiya College of Engineering



Further Reading

- Amazon.com: Functional Art, The: An introduction to information graphics and visualization (Voices That Matter): 9780321834737: Cairo, Alberto: Books
- **Data Points: Visualization That Means Something 1st Edition**

Question ?



SOMAIYA
VIDYAVIHAR UNIVERSITY

K J Somaiya College of Engineering

