

Making an

IMPACT

with **DATA VISUALIZATION**

Lorin Bruckner



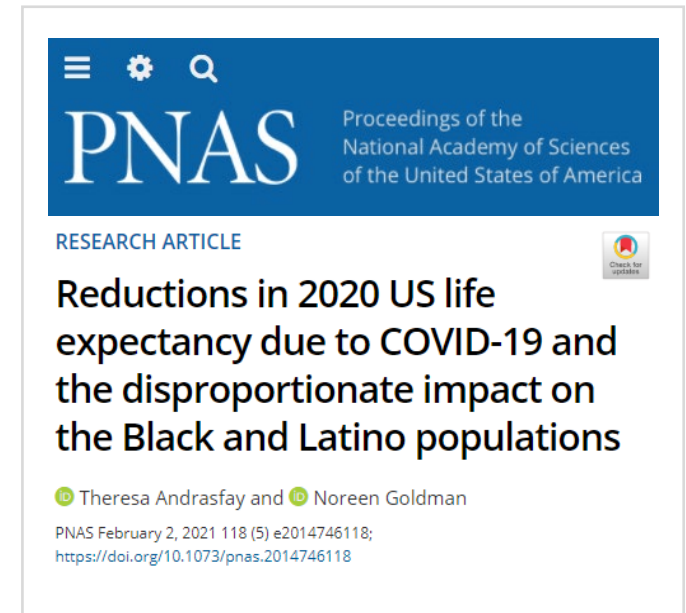
THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

Why Visualize Data?

Why Visualize Data?

Study: Change in Life Expectancy Due to COVID-19 & Disproportionate Impact on Black and Latino Populations

- Projects that **life expectancy** will be **reduced by 1.13%** for people living in the **United States**
- Estimated **reductions** for the **Black** and **Latino** populations are **3 to 4 times that for Whites**
- This means there was a **pattern** found in the data. How can we **describe** that pattern so it is easily understood?



Why Visualize Data?

1. Paragraph from the study

Our medium estimate indicates a reduction in US life expectancy at birth of 1.13 y to 77.48 y, lower than any year since 2003. We also project a 0.87-y reduction in life expectancy at age 65 y. The Black and Latino populations are estimated to experience declines in life expectancy at birth of 2.10 and 3.05 y, respectively, both of which are several times the 0.68-y reduction for Whites. These projections imply an increase of nearly 40% in the Black–White life expectancy gap, from 3.6 y to over 5 y, thereby eliminating progress made in reducing this differential since 2006. Latinos, who have consistently experienced lower mortality than Whites (a phenomenon known as the Latino or Hispanic paradox), would see their more than 3-y survival advantage reduced to less than 1 y.

Why Visualize Data?

2. Table describing the data

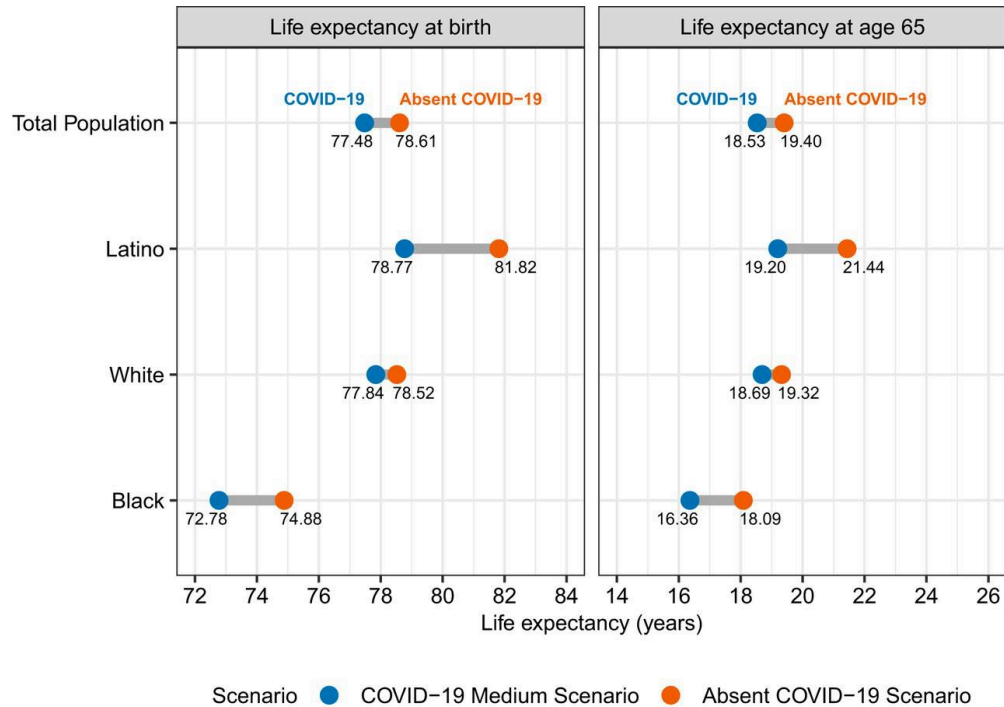
Table 1. [View popup](#)
Life expectancy projections for the United States in 2020 by race and ethnicity under different COVID-19 mortality scenarios

Race and ethnicity	Total population life expectancy		Non-Latino White life expectancy		Non-Latino Black life expectancy		Latino life expectancy	
	At birth (e_0)	At age 65 y (e_{65})	At birth (e_0)	At age 65 y (e_{65})	At birth (e_0)	At age 65 y (e_{65})	At birth (e_0)	At age 65 y (e_{65})
Absent COVID-19 (referent)	78.61	19.40	78.52	19.32	74.88	18.09	81.82	21.44
IHME current projection (medium scenario)	77.48	18.53	77.84	18.69	72.78	16.36	78.77	19.20
Difference (years)	-1.13	-0.87	-0.68	-0.63	-2.10	-1.73	-3.05	-2.24
IHME mandates easing scenario (higher mortality scenario)	77.39	18.46	77.79	18.64	72.62	16.23	78.54	19.03
Difference (years)	-1.22	-0.94	-0.73	-0.68	-2.26	-1.86	-3.28	-2.41
IHME universal masks scenario (lower mortality scenario)	77.63	18.65	77.93	18.78	73.05	16.58	79.16	19.48
Difference (years)	-0.98	-0.75	-0.59	-0.54	-1.83	-1.51	-2.66	-1.96

Source is authors' calculations of life expectancy projections in 2020. Difference refers to the difference between the projection and the absent COVID-19 scenario. Calculations are based on IHME projections that were updated October 9, 2020.

Why Visualize Data?

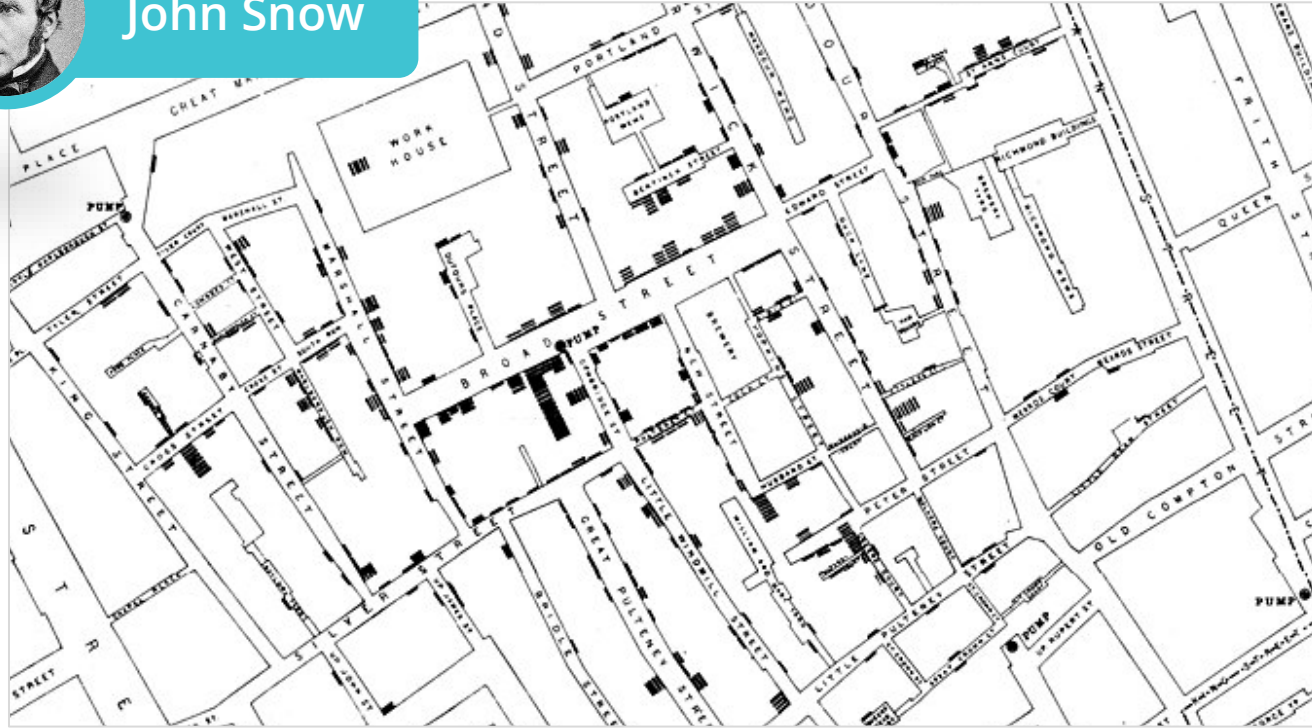
3. Visualization



Why Visualize Data?



John Snow



Source: Wikimedia Commons

Why Visualize Data?

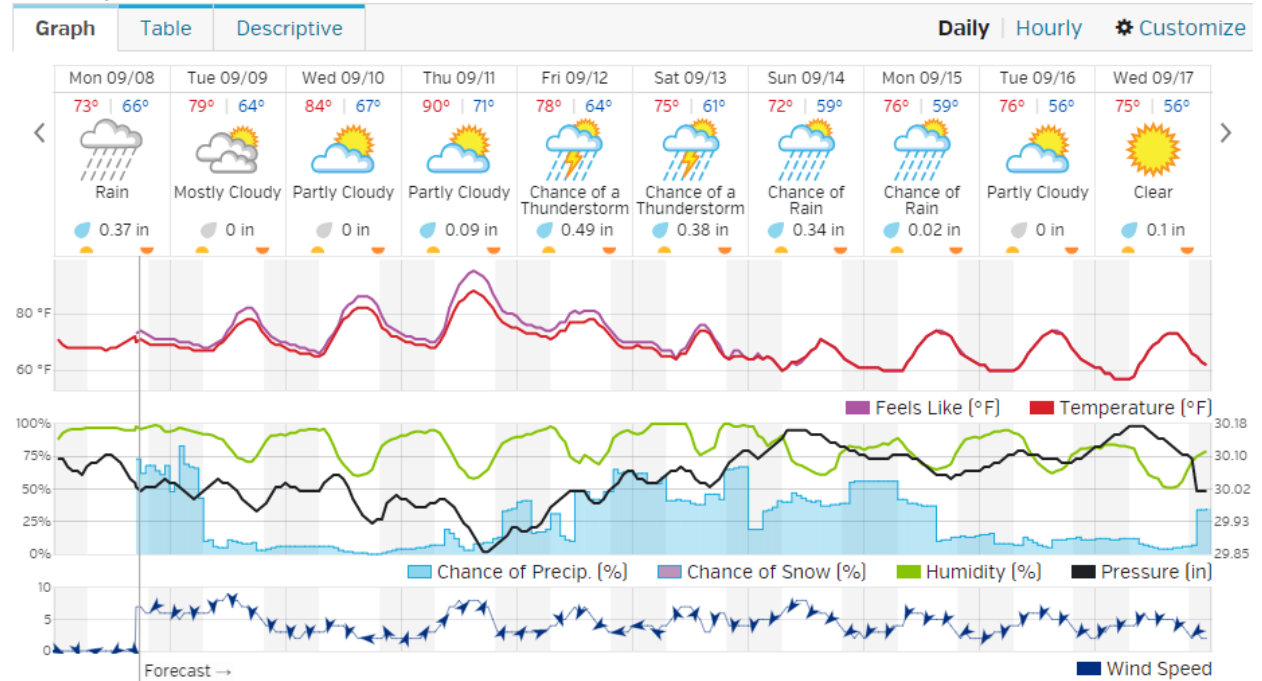
We use visualization to help us explore and understand **complex patterns** in data that cannot be as easily perceived through other means of communication.

How does Data Visualization Work?

How Does Data Visualization Work?

- In modern visualizations, The **complexity and variety** of graphic symbols often rivals that of the data itself.
- **Interactivity** adds yet another dimension.
- How are we able to **perceive and synthesize** so much information?

10-Day Weather Forecast



Source: Weather Underground

Source: Weather Underground BestForecast

How Does Data Visualization Work?

Preattentive Processing

- The eye is drawn to certain **features and patterns** that “stand out” or **pop**
- Information is **subconsciously obtained** from our environment **before** being attentively processed (or not)
- **Pattern recognition**

How many 7's do you see?

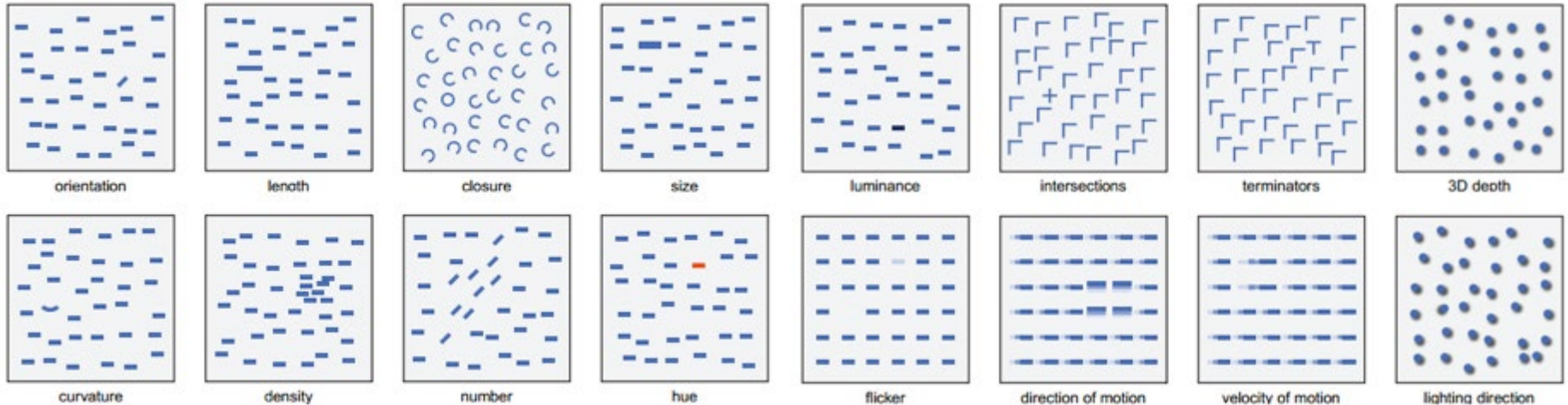
98187926383749819794961389746139
44978732184987621617995462132549
89796531859129939549719819295198
197354687929

98187926383749819794961389746139
44978732184987621617995462132549
89796531859129939549719819295198
197354687929

98187926383749819794961389746139
44978732184987621617995462132549
89796531859129939549719819295198
197354687929

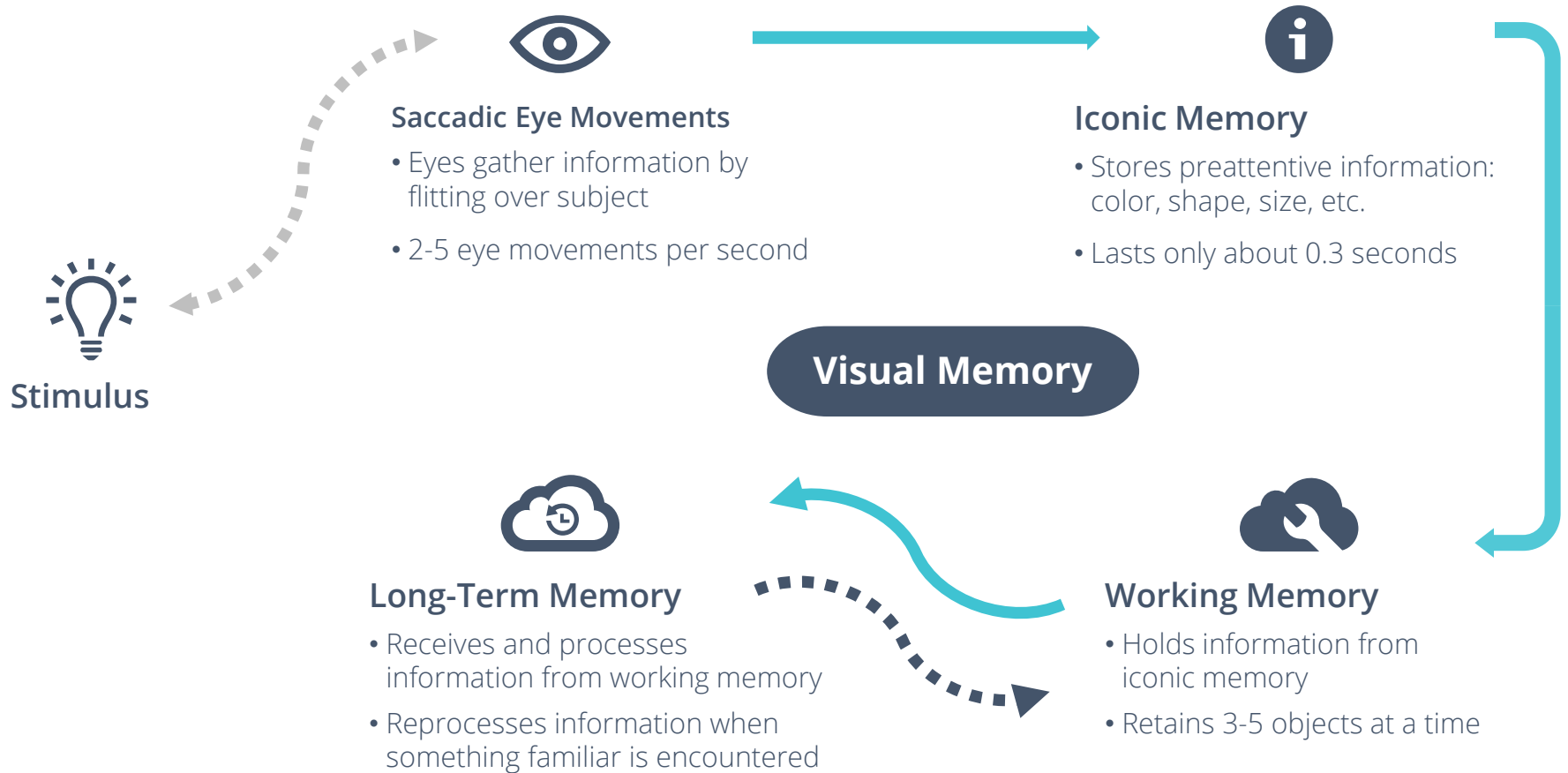
How Does Data Visualization Work?

Visual Features That Affect Preattentive Processing



Source: Healey, C. G., & Enns, J. T. (2012). Attention and Visual Memory in Visualization and Computer Graphics. *IEEE Trans. Visual. Comput. Graphics IEEE Transactions on Visualization and Computer Graphics*, 18(7), 1170-1188. doi:10.1109/tvcg.2011.127

How Does Data Visualization Work?

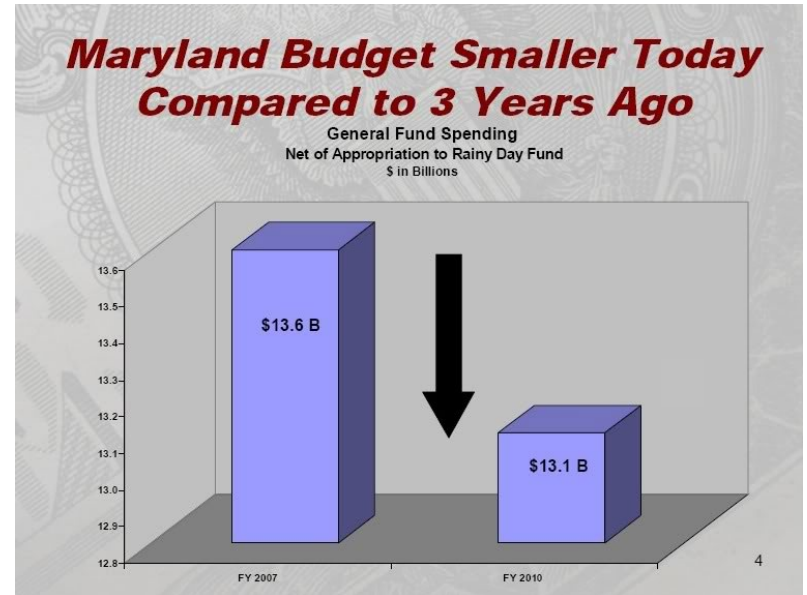


Selecting Visualizations

Selecting Visualizations

Step 1: Is a Visualization Necessary?

- Are you illustrating **complex patterns** and/or **large quantities** of data?
- Are you answering a **question**, making an **argument**, or telling a **story**?
- Will a visualization be **more informative** than a simple **table** or **text**?



Source:

<https://web.archive.org/web/20091119151328/http://www.governor.maryland.gov/documents/November09BPW.pdf>

Selecting Visualizations

Step 2: Who is Your Audience?

- How much does your audience know about the **research subject**?
- How much does your audience know about **data analysis**?
- What are the **norms and expectations** in this field?



Selecting Visualizations

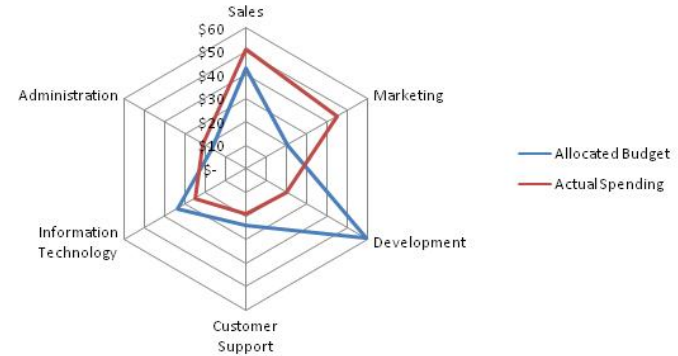
Step 3: What do you want to show your audience?

I WANT TO...	SOLUTION
Compare values	Bar/Column Chart
Show relationships between variables	Scatter Plot
Show how something changes over time	Line Chart
Demonstrate how proportions compare to the whole	Pie Chart; Stacked Bar Chart
Compare the values between geographic/spatial areas	Maps

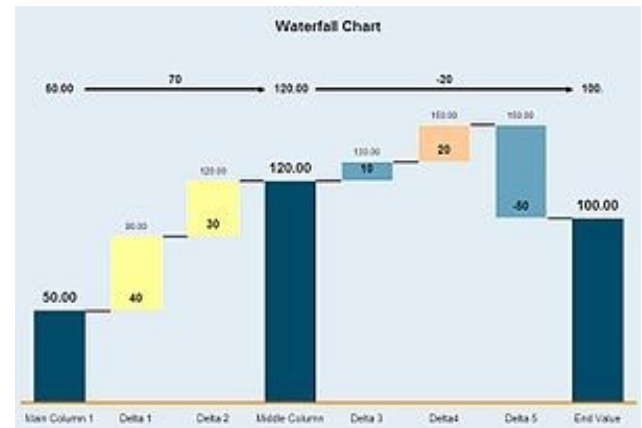
Selecting Visualizations

What about other charts?

- There are **many, many types of charts** to choose from.
- Some require **specialized knowledge** to interpret correctly.
- Some are simply **misleading**.
- **Always keep your audience in mind when making a decision.**



Source: Wikimedia Commons



Source: Wikimedia Commons

Selecting Visualizations

SCENARIO 1

You are a **journalist for a national paper**. You are writing an article about **crime in the U.S.**

Some believe that crime is on the rise, while others argue that it has diminished over the years.

The data tells a different story depending on the type of crime, geographic area and time period that is being studied.

SCENARIO 2

You are a **social media developer for an ad agency**. You provide one of your clients with a **weekly report on brand activity**.

Mentions of your client's brand on Twitter have decreased by 32% since the previous week.

Sales have remained unchanged.

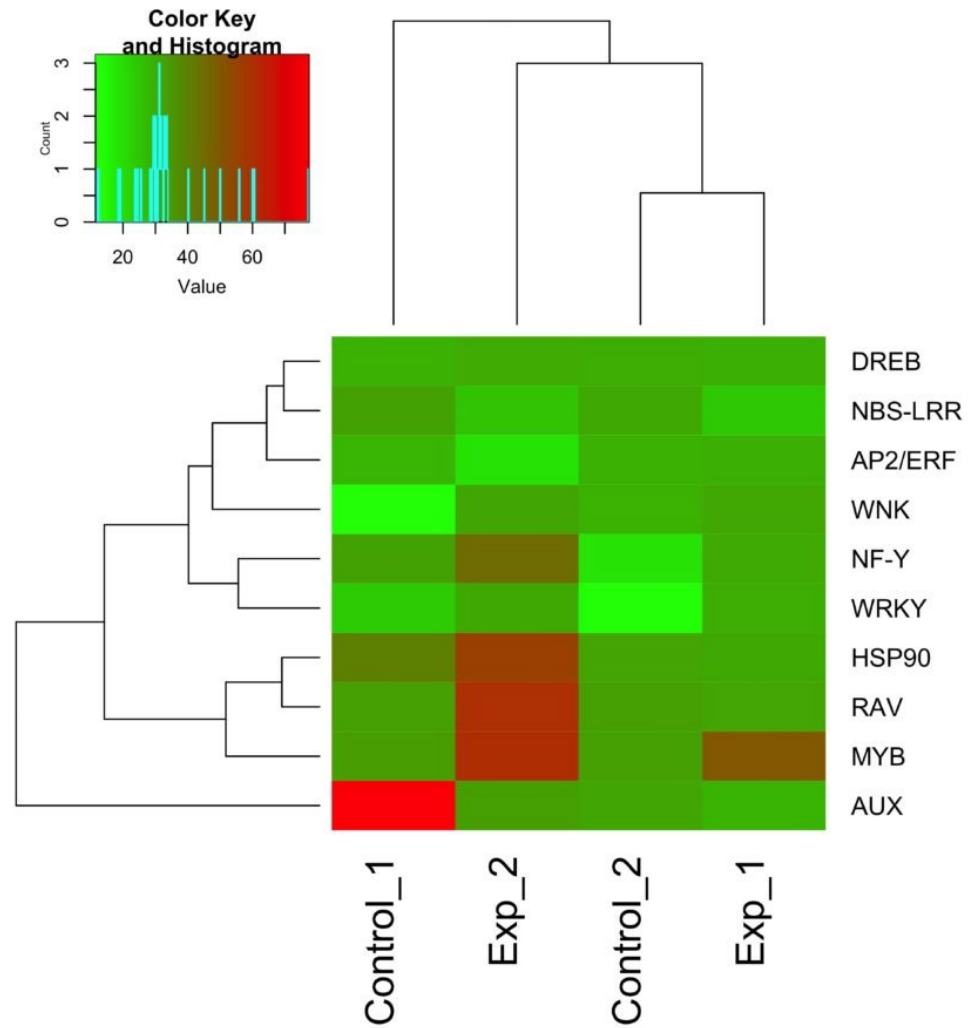
SCENARIO 3

You are a **researcher studying Parkinson's disease**.

You are publishing the results of a **study that uses microarrays** to measure gene expression levels in mice.

Your dataset includes over 9,000 genes.

1. Should a visualization be used in this scenario?
2. Given the audience and the data, what is the best type of visualization to use?

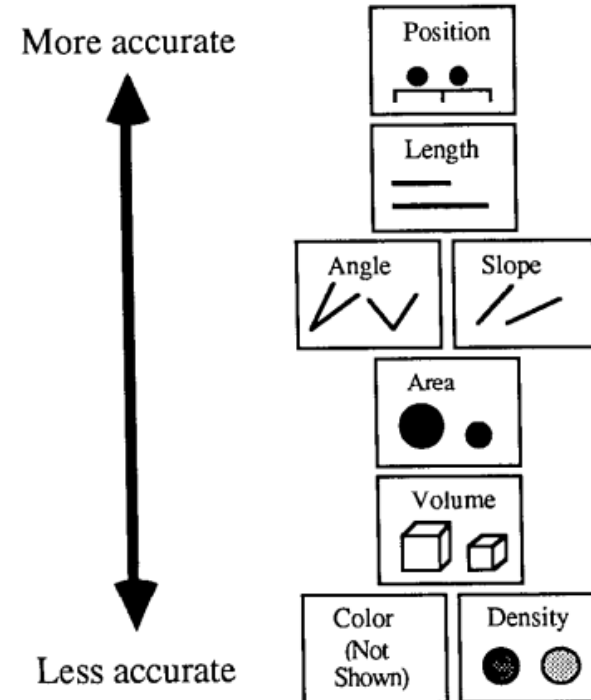


Accuracy

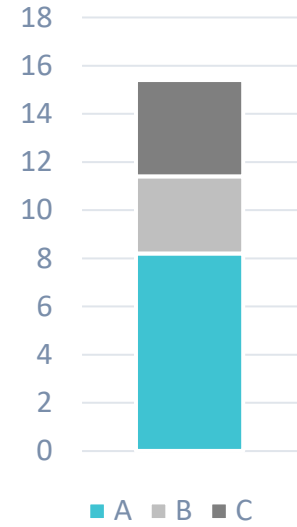
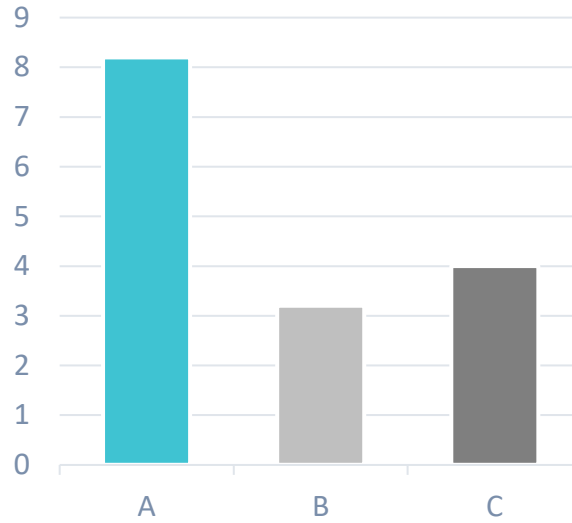
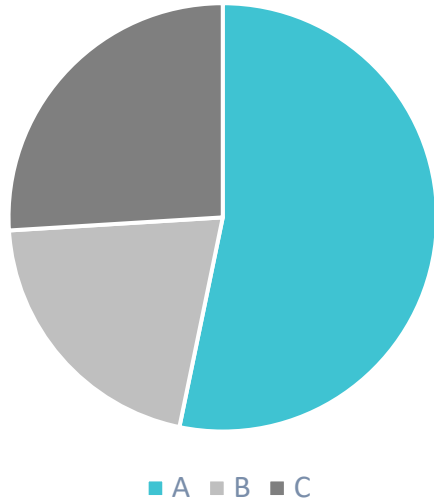
Accuracy

Graphical Perception

- 1984 experiments by McGill and Cleveland **rank how accurately people assess graphic depictions of data**
- The rankings are useful but **not uncompromising**
- Context and **audience** should also be considered

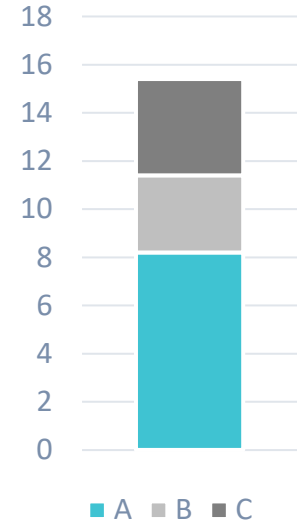
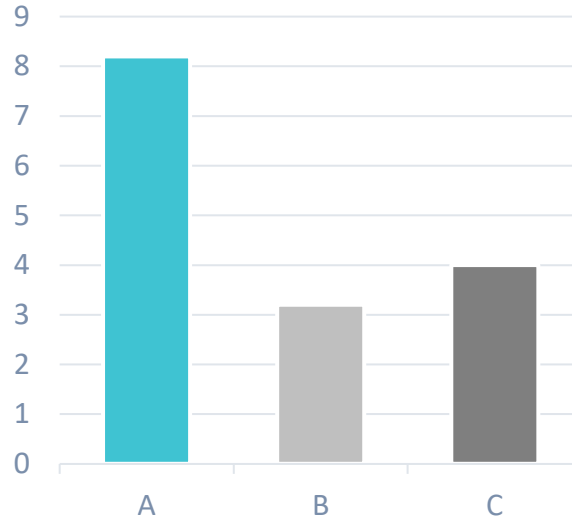
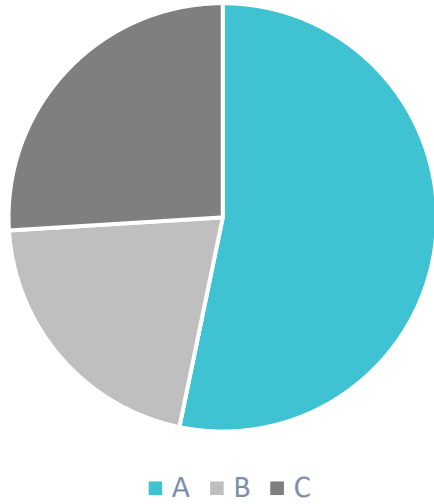


Accuracy



1. Which chart looks better to you?

Accuracy



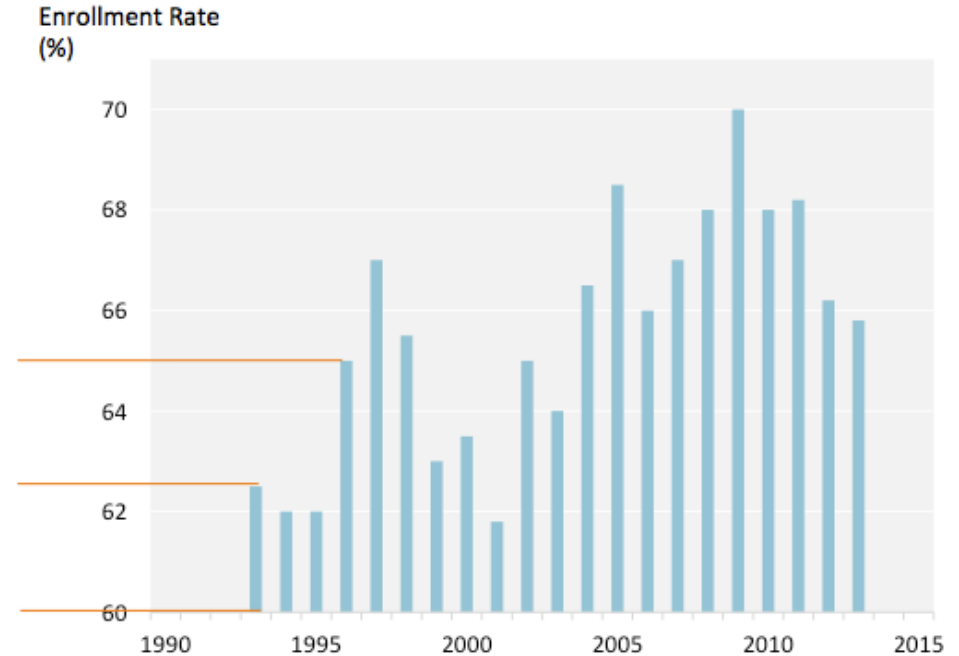
1. Which chart looks better to you?
2. Which chart makes it easier to judge the difference between B and C?

Best Practices

Best Practices

When using column/bar charts, always start the scale at 0.

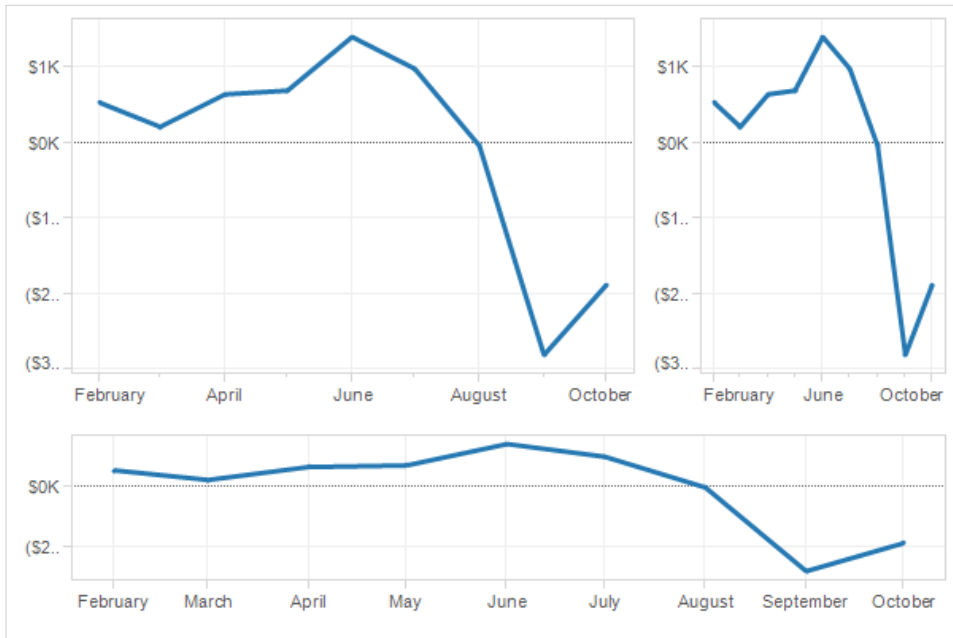
- The column for 1996 **appears to be twice the height** of the column for 1993.
- Yet, the axis labels tell us **the difference between the two years is only 2.5%** (65% versus 62.5%).
- This is a **common distortion tactic**.



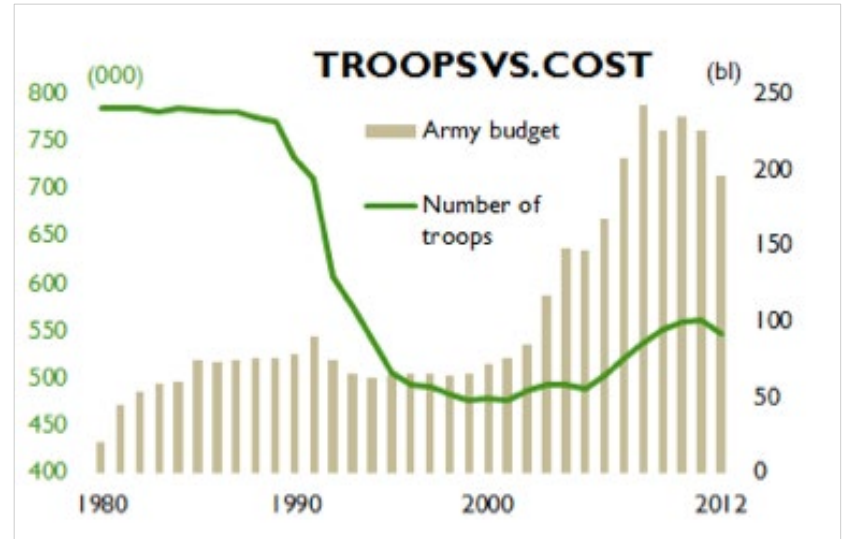
Source: http://junkcharts.typepad.com/junk_charts/2014/04/when-to-use-the-start-at-zero-rule-.html

Best Practices

Make sure scale is consistent and honest.



SAME DATA WITH 3 DIFFERENT SCALES



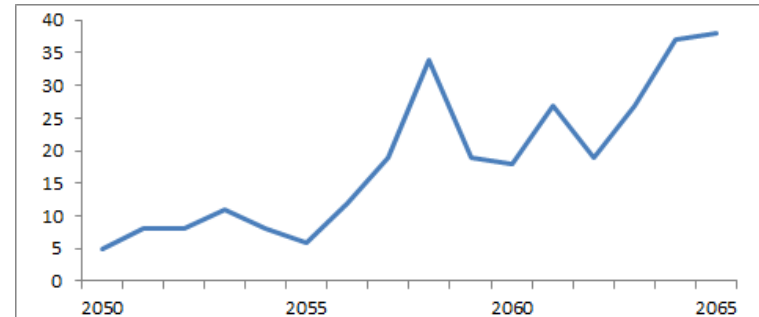
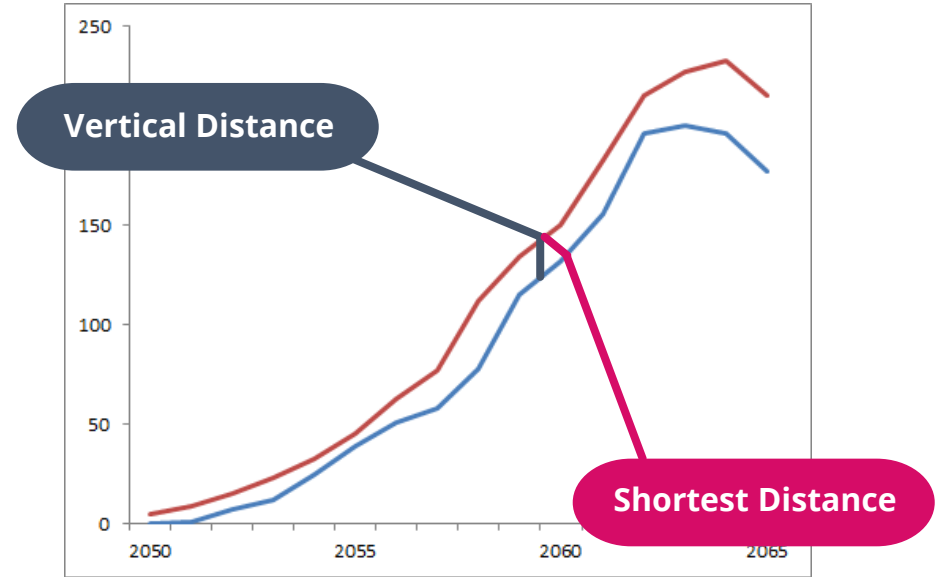
Source: <http://www.excelcharts.com/blog/redraw-troops-vs-cost-time-magazine/>

DIFFERENT SCALES IN ONE GRAPH

Best Practices

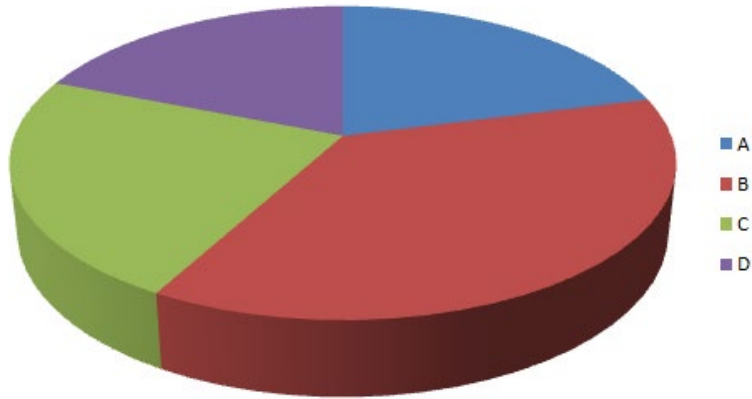
With line graphs, consider using a separate chart to demonstrate differences.

- Human perception defaults to the **shortest distance** between two lines rather than the **vertical distance**.
- A chart of the **difference alone** can be more **accurate** and **informative**.

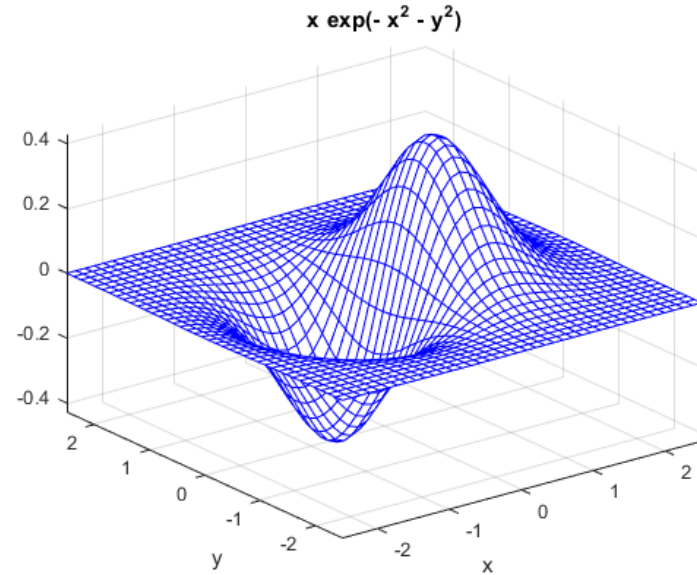


Best Practices

Use “3D” charts for a reason.
(there is almost never a reason)



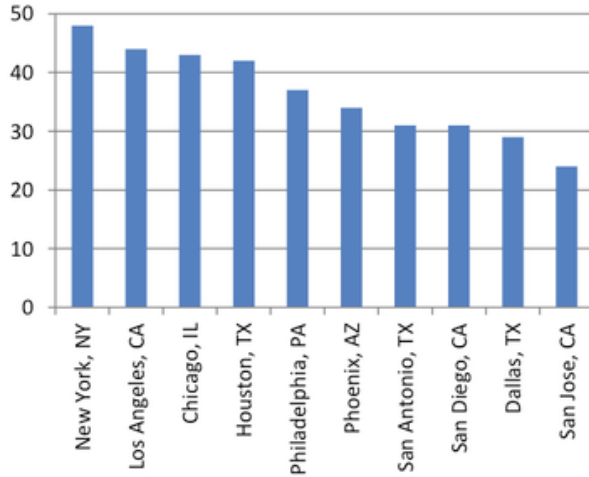
JUST DON'T



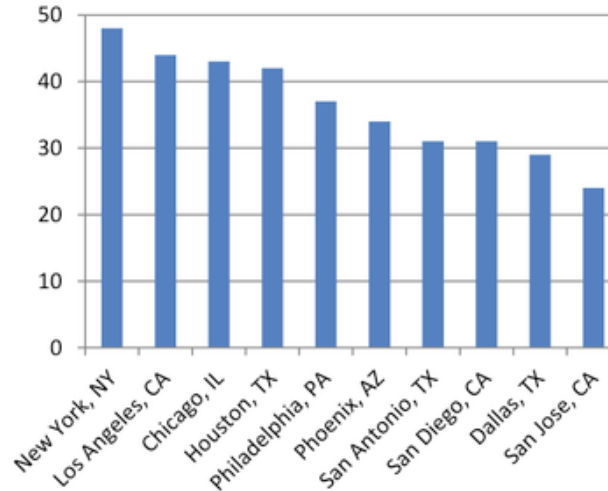
OK

Best Practices

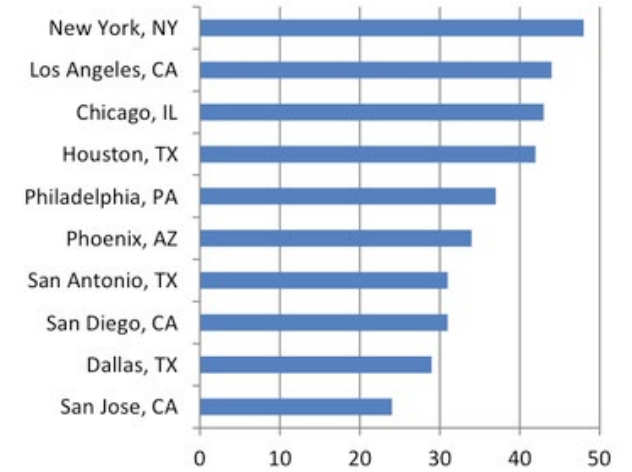
Pay attention to text orientation.



BAD



BETTER



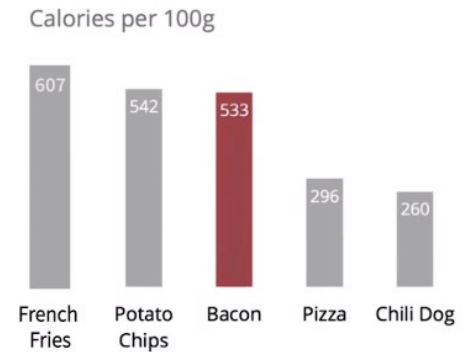
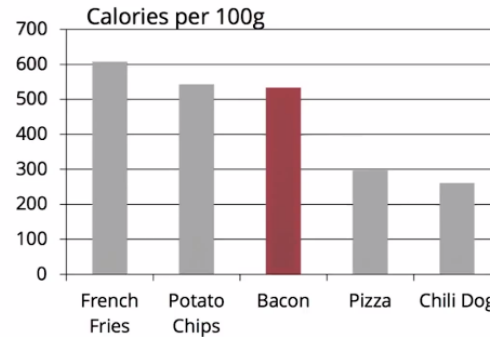
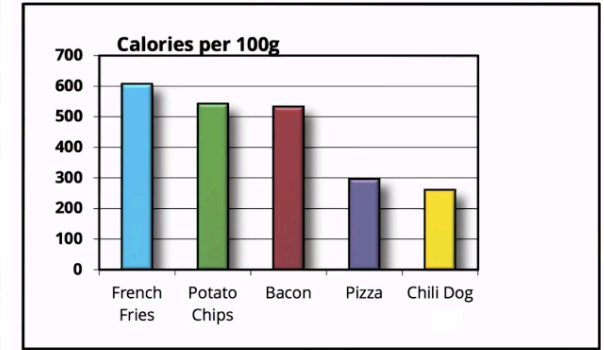
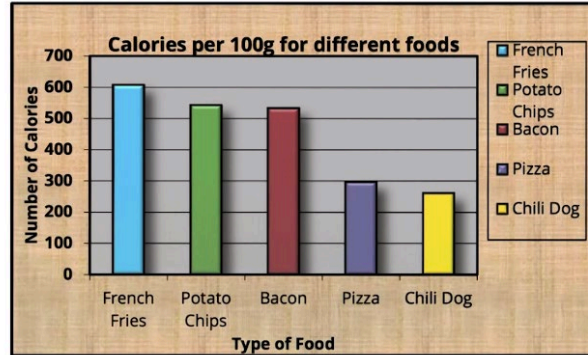
BEST!

Best Practices

Watch your data-to-ink ratio:

$$\frac{\text{Ink used to directly encode data values}}{\text{Total Ink used in the visualization}}$$

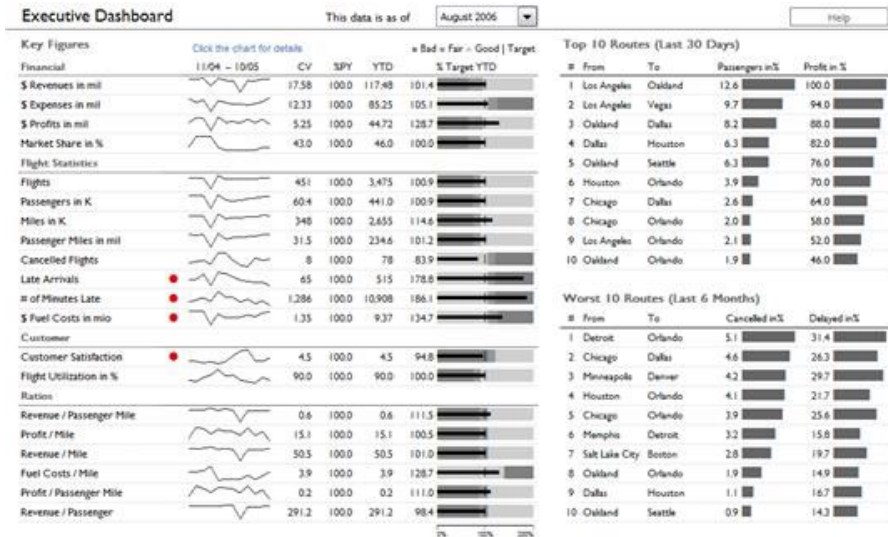
- Only use graphical elements that are **necessary** for the chart to be **easily read** by your audience.
- Once you've completed a visualization, **check to see if there's anything that can be removed**.



Best Practices

Classic Trick of the Trade: The Squint Test

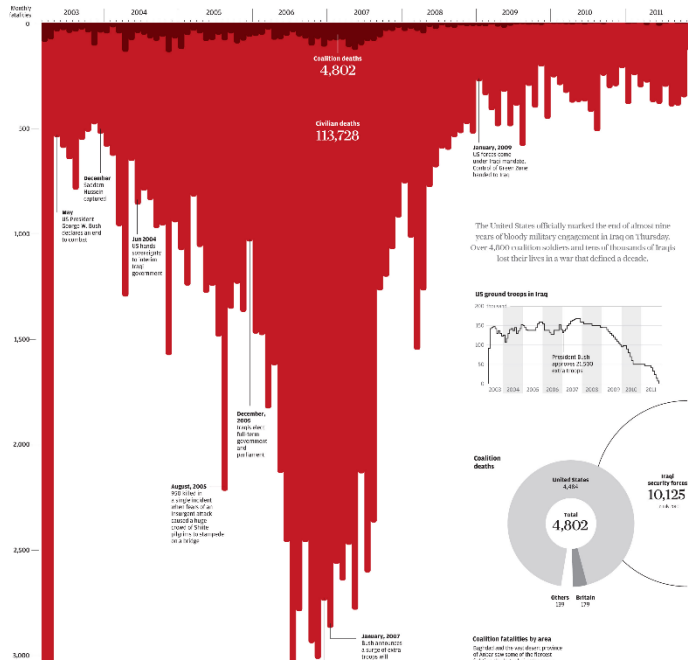
Which elements “pop out” and catch your eye?
Are these the elements you want to draw attention to?



Best Practices

Creativity is good, but clarity is key.

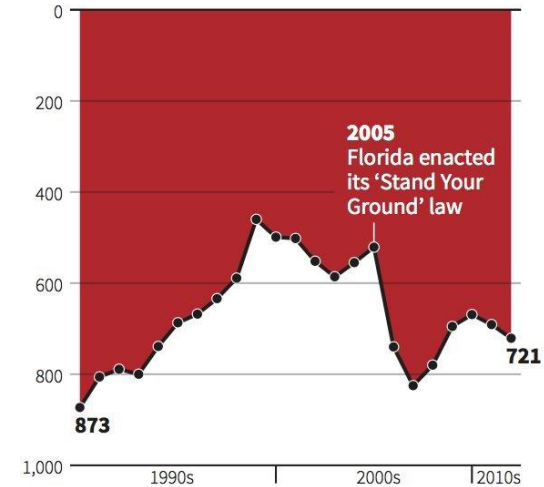
Iraq's bloody toll



Source: <http://www.scmp.com/infographics/article/1284683/iraqs-bloody-toll>

Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

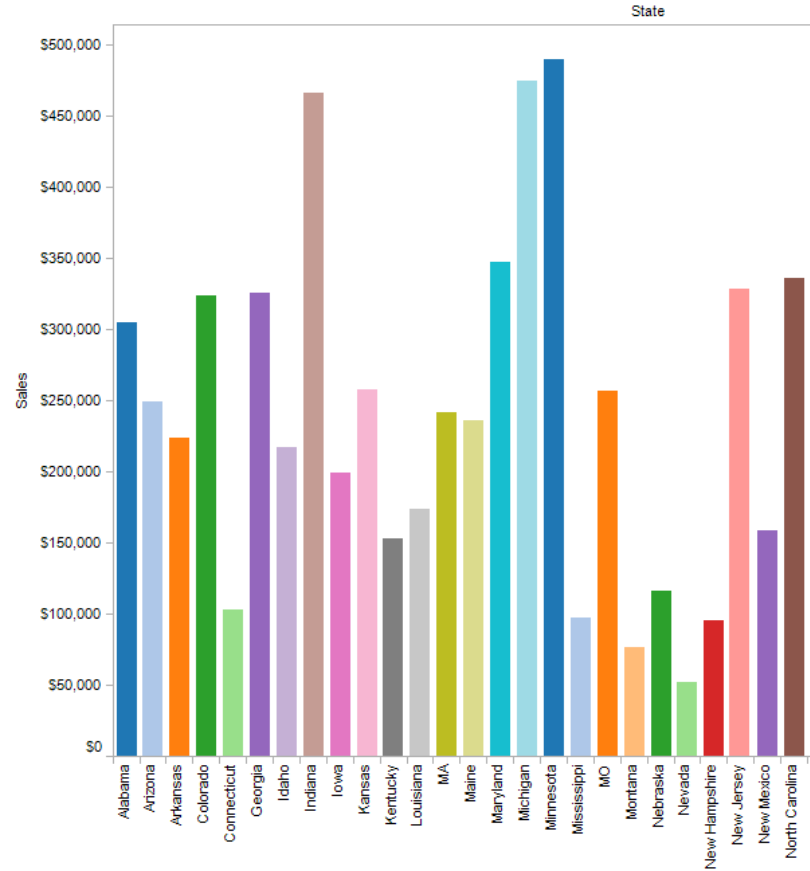
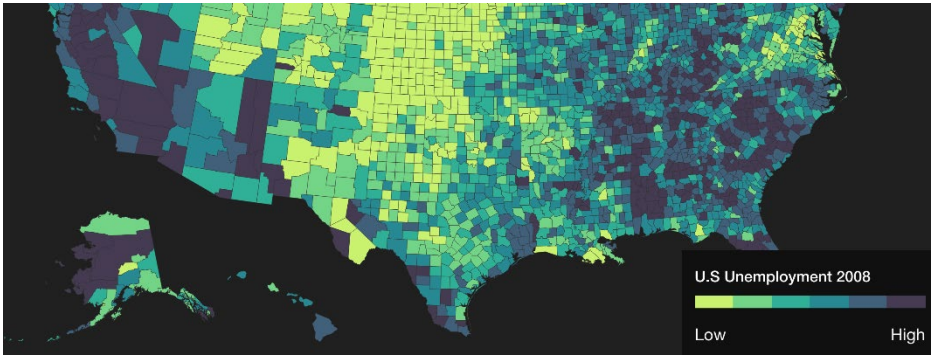
Source: <http://www.businessinsider.com/gun-deaths-in-florida-increased-with-stand-your-ground-2014-2>

Working With Color

Color

Color isn't always necessary.

- Many visualization tools add color **by default**.
- Often a **label** on its own is enough.
- Color can be useful to **distinguish groups or intervals**.



Color

Many people are color blind.

- Most people affected by **color blindness** have diminished ability to differentiate **red and green**.
- **Don't** use color schemes that **involve both**.



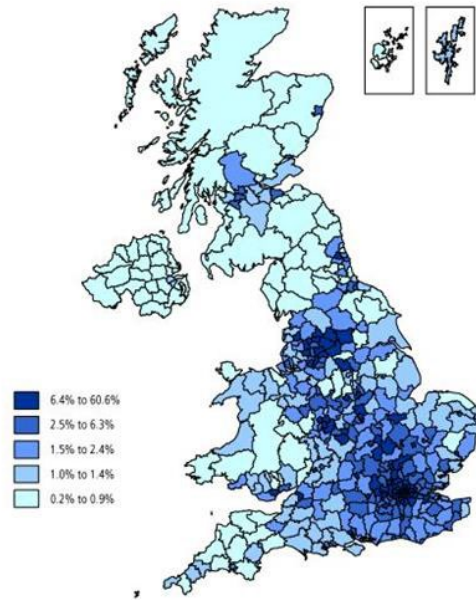
- Vary the **lightness/darkness/saturation** of colors as well. Check by printing or viewing in **grayscale**.



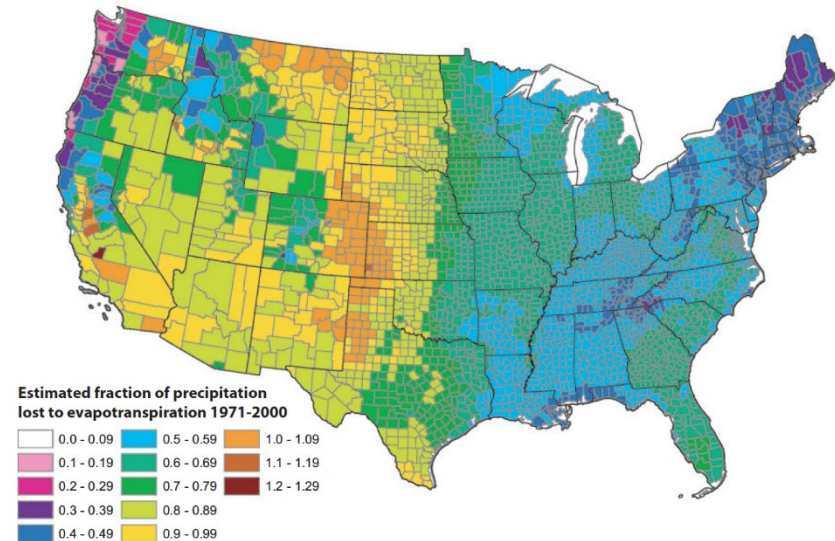
Source: <http://24ways.org/2012/colour-accessibility/>

Color

Natural scales are more accurately interpreted than “rainbow” scales.



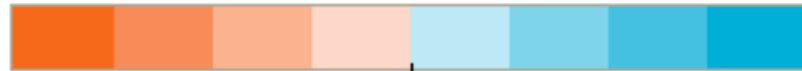
Source: <http://spatial.ly>



Source: Borland, D., & Ji, R. T. (2007). Rainbow Color Map (Still) Considered Harmful. *IEEE Comput. Graph. Appl. IEEE Computer Graphics and Applications*, 27(2), 14-17. doi:10.1109/mcg.2007.323435

Color

Examples of Natural Scales



Orange → Blue



Dark → Light

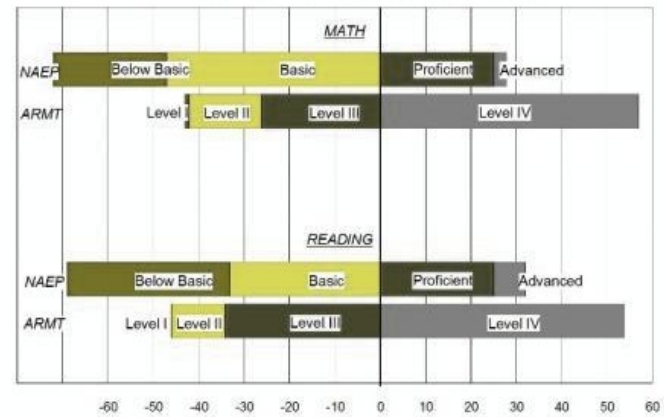
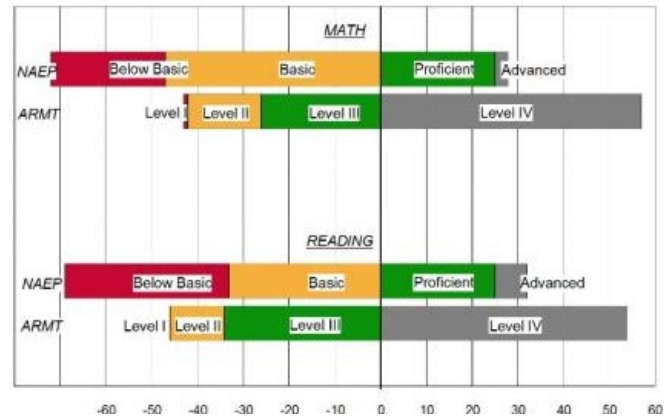


Low Saturation → High Saturation

Color

Resources for Working With Color

- **iwanthue** (medialab.github.io/iwanthue/) creates a color palette from colorblind-friendly presets and examines the accessibility of different color pairings
- **Coblis** (color-blindness.com/coblis-color-blindness-simulator) allows you to upload images and displays how they will appear to someone who is colorblind.

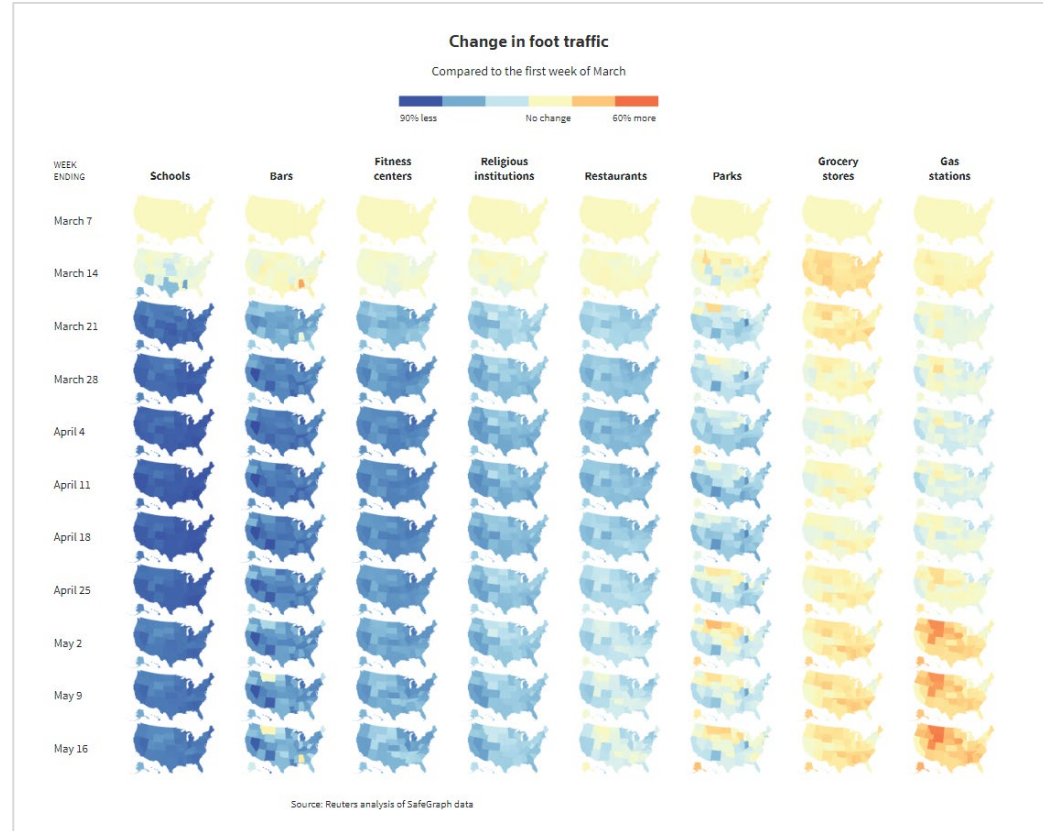


Small Multiples

Small Multiples

Small Multiples are a good way to handle lots of data.

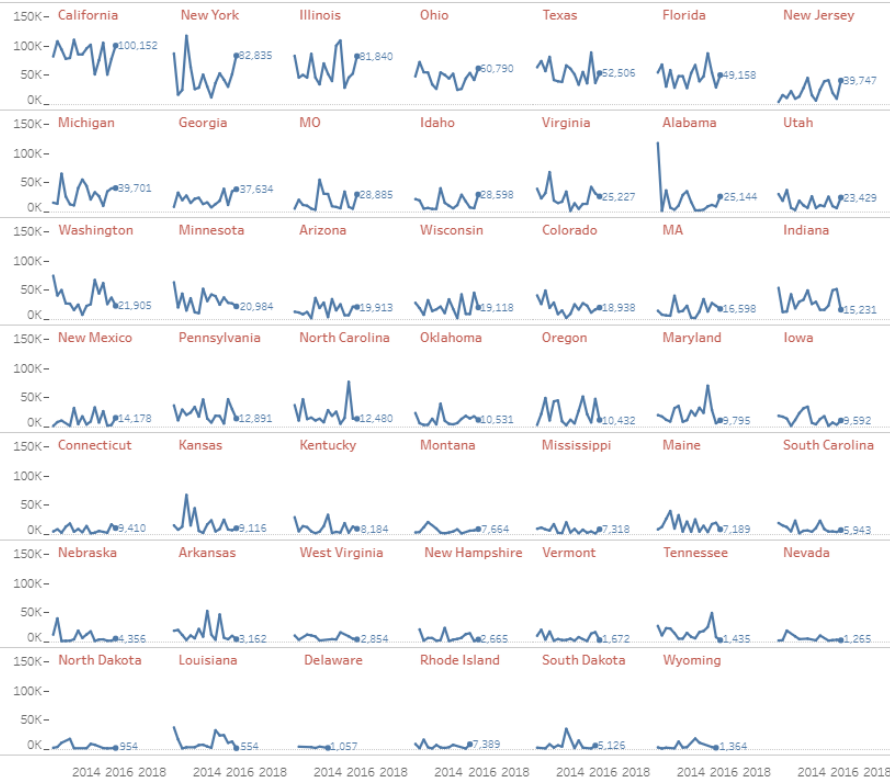
- **Series of charts** using the **same scale on both axes** so they can easily be compared
- Charts are often **arranged in a grid** according to **dimensions** in the data
- Often provide an **alternative to color**



Small Multiples

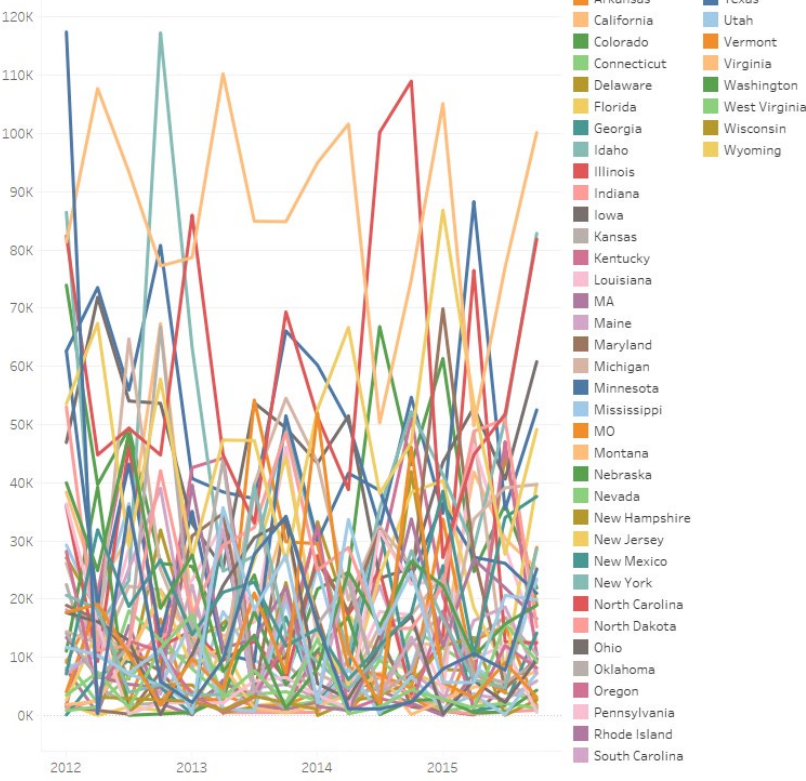
Quarterly Sales Trend by State

States ordered from highest to lowest based on sales from the most recent month.



vs.

Quarterly Sales Trend by State

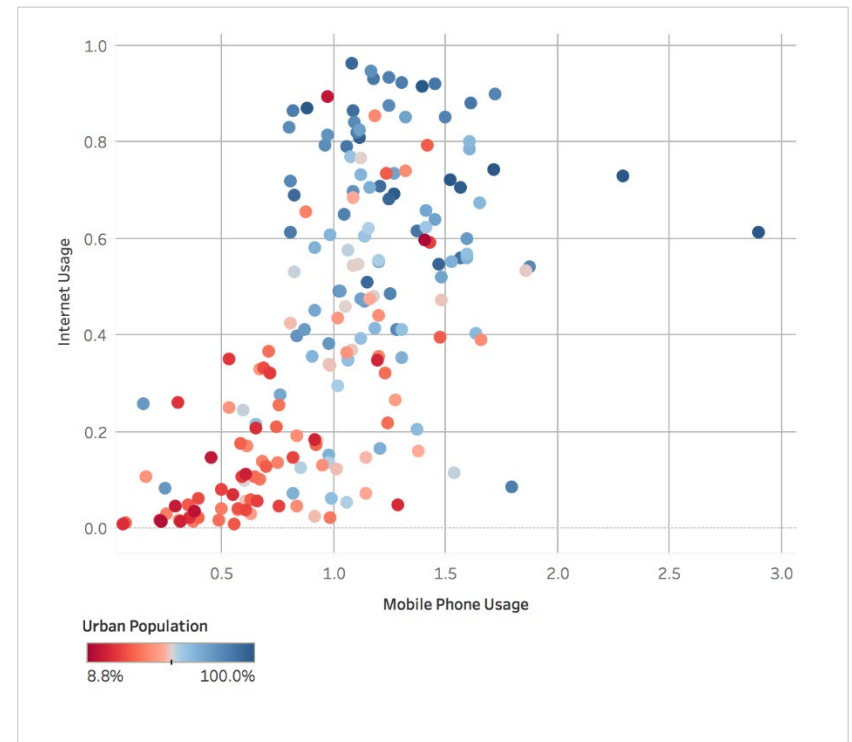


Small Multiples

Scatterplot Matrix

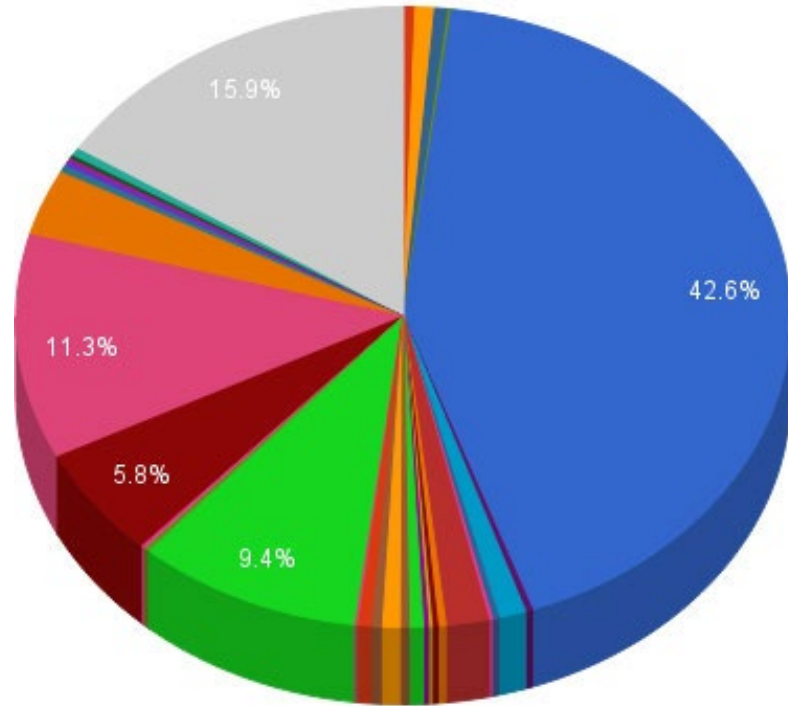


vs.



Improving Visualizations

Which incredible Canadian woman should be featured on the \$20 bill?



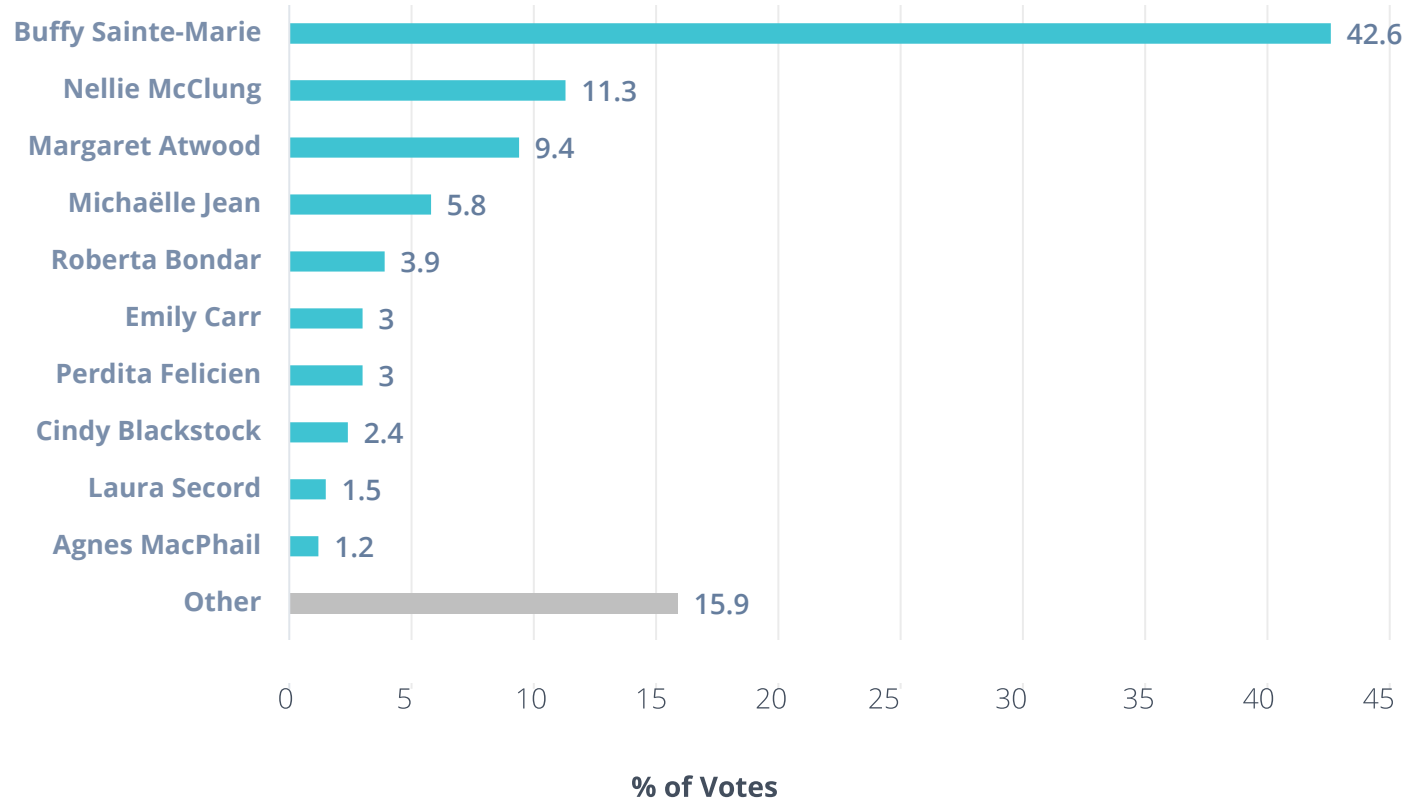
- Adrienne Clarkson
- Agnes MacPhail
- Alice Munro
- Anne Murray
- Buffy Sainte-Marie
- Celine Dion
- Cindy Blackstock
- Clara Hughes
- Elizabeth May
- Emily Carr
- Emily Murphy
- Emily Stowe
- Harriet Tubman
- Jennie Trout
- Joni Mitchell
- Kim Campbell
- Laura Secord
- Louise Arbour
- Madeleine Parent
- Margaret Atwood
- Margaret Laurence
- Maud Menten
- Michaëlle Jean
- Nellie McClung
- Pauline Johnson
- Perdita Felicien
- Roberta Bondar
- Rosemary Brown
- Tantoo Cardinal
- The Founders Of Idle No More
- Thérèse Casgrain
- Viola Desmond
- Other

<http://www.cbc.ca/radio/dnto/just-get-the-thing-done-and-let-them-howl-nellie-mcclung-1.3418924/which-incredible-canadian-woman-should-be-featured-on-the-20-bill-1.3419070>

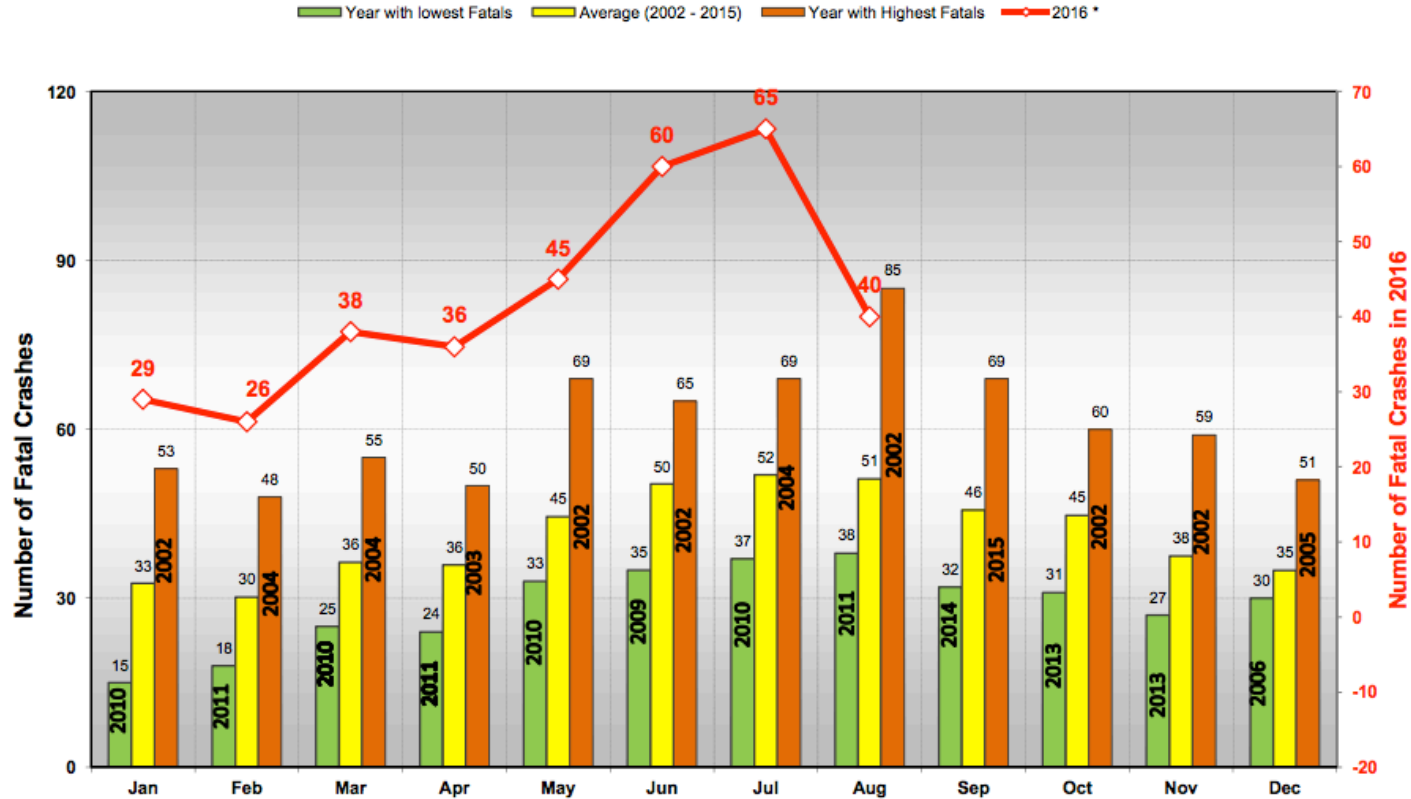
How can this visualization be improved?

Which Incredible Canadian Woman Should Be Featured On The \$20 Bill?

Poll Results



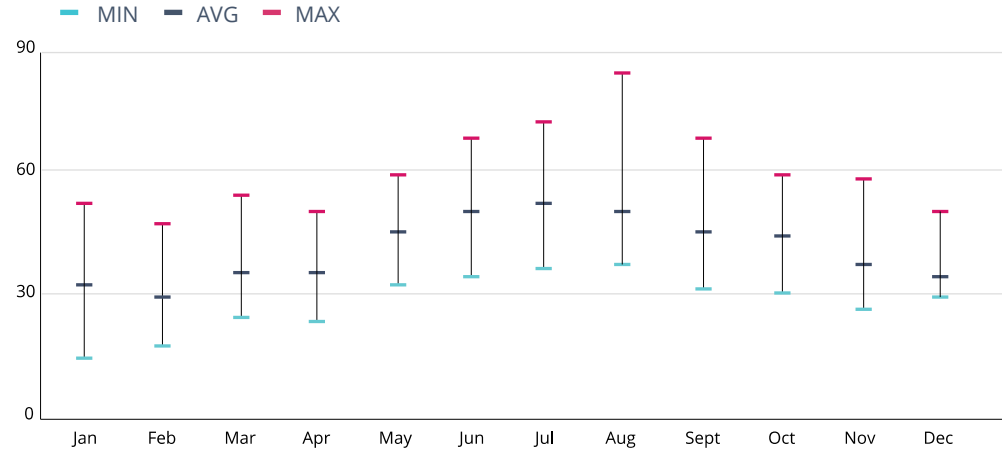
Colorado Fatal Crash by Month Since 2002



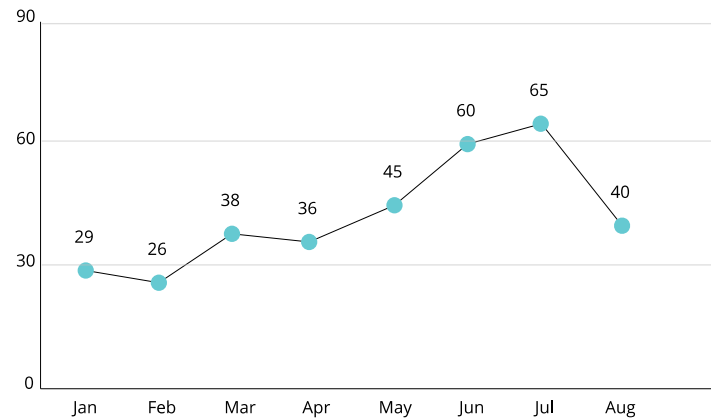
Source: https://www.codot.gov/library/traffic/safety-crash-data/fatal-crash-data-city-county/historical_fatalis.pdf/at_download/file

How can this visualization be improved?

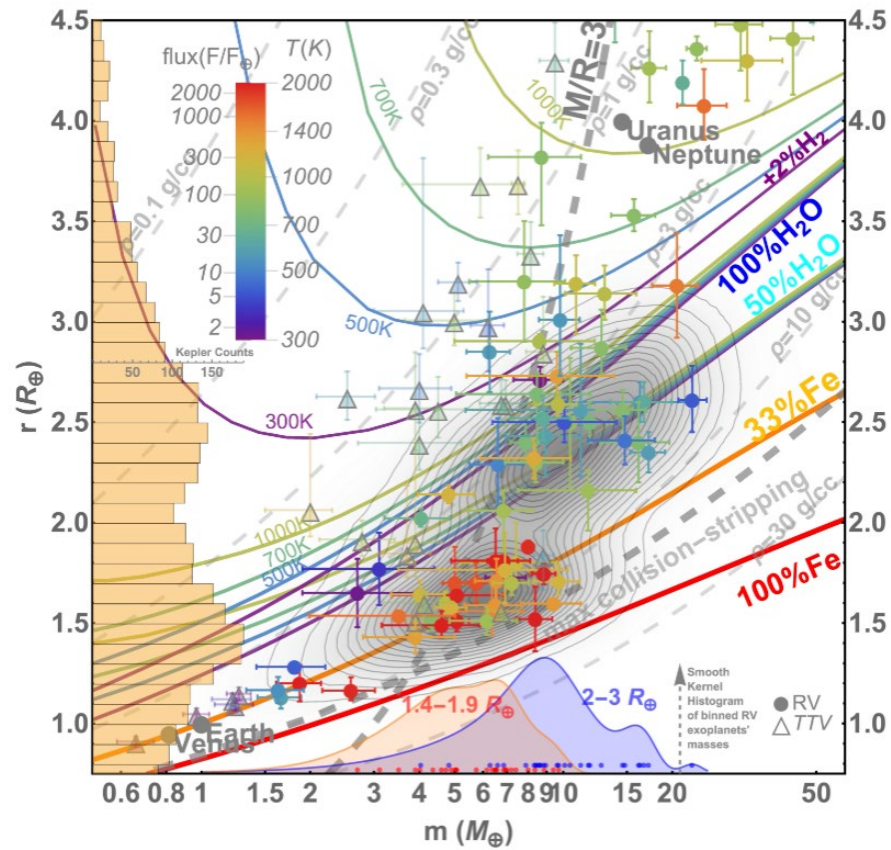
Colorado Fatal Crashes by Month Since 2002



Colorado Fatal Crashes in 2016



	LOWEST YEAR	HIGHEST YEAR
JAN	2002	2010
FEB	2004	2011
MAR	2004	2010
APR	2003	2011
MAY	2002	2010
JUN	2002	2009
JUL	2004	2010
AUG	2002	2011
SEP	2015	2014
OCT	2002	2013
NOV	2002	2013
DEC	2005	2006



Source: <https://www.pnas.org/content/116/20/9723>

How can this visualization be improved?

Resources

Resources – Tools

go.unc.edu/VizTools

Resources – Reading

Books

Tufte, Edward

- The Visual Display of Quantitative Information (2001)
- Visual Explanations (1997)
- Envisioning Information (1990)

Cleveland, William S

- Visualizing Data (1993)
- The Elements of Graphing Data (1985)

BLOGS

Flowing Data

flowingdata.com

Information is Beautiful

informationisbeautiful.net

Junk Charts

junkcharts.typepad.com

Learn More

library.unc.edu/data/davis-hub-events



QUESTIONS?