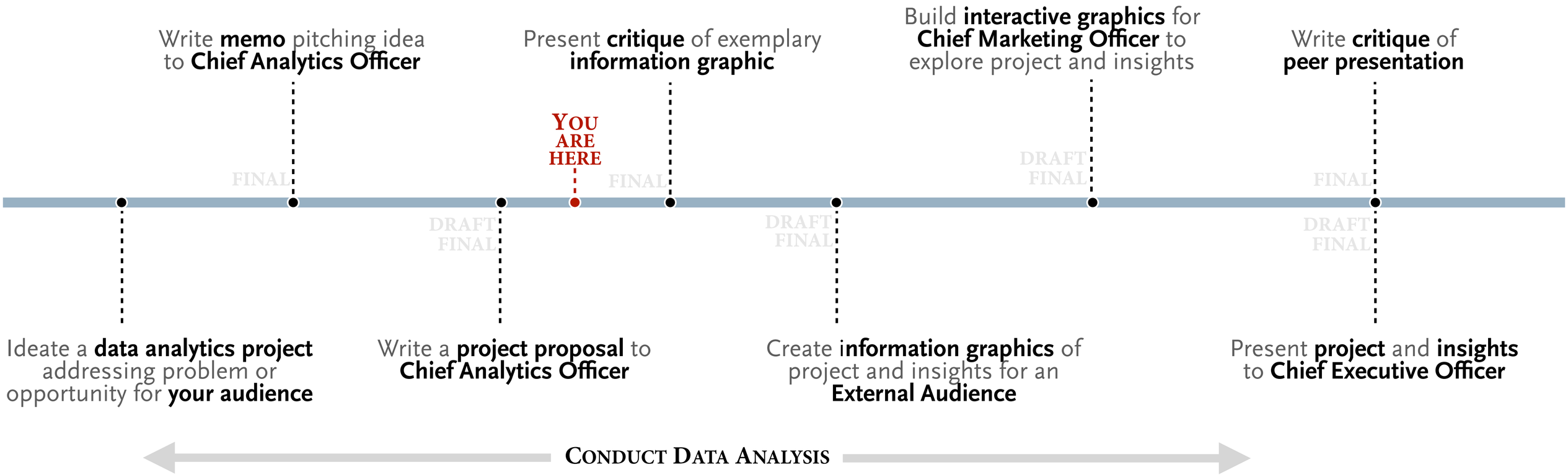
Storytelling with data

07 | From exploration to explanation; audiences and complexity; data graphics in storytelling

Scott Spencer | Columbia University

course overview | main course deliverables



© 2021 Scott Spencer / https://ssp3nc3r.github.io scott.spencer@columbia.edu

from exploring to explaining

Get our audience(s) to

pay attention to,
understand,
(be able to) **act upon**

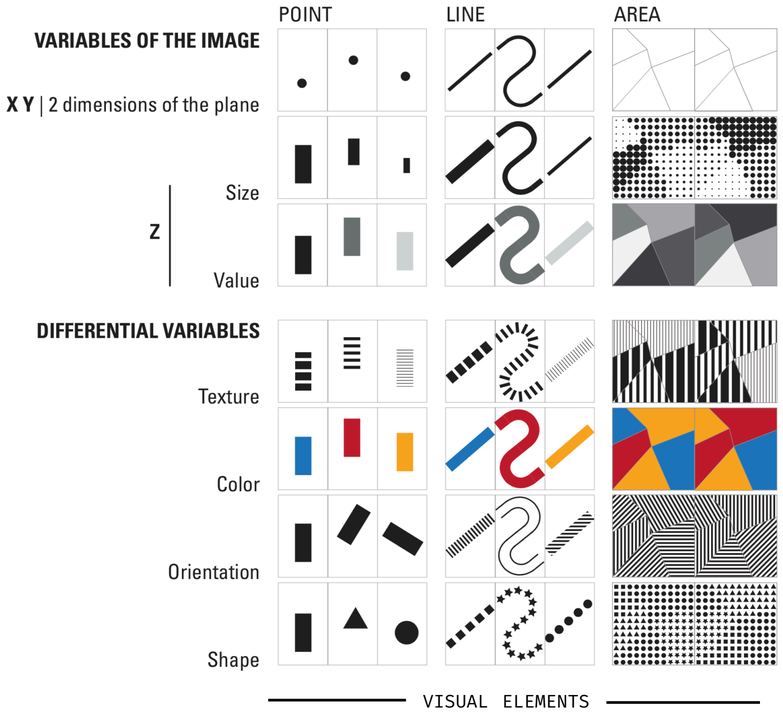
a maximum of **messages**,
given **constraints**.

We worked with IR.

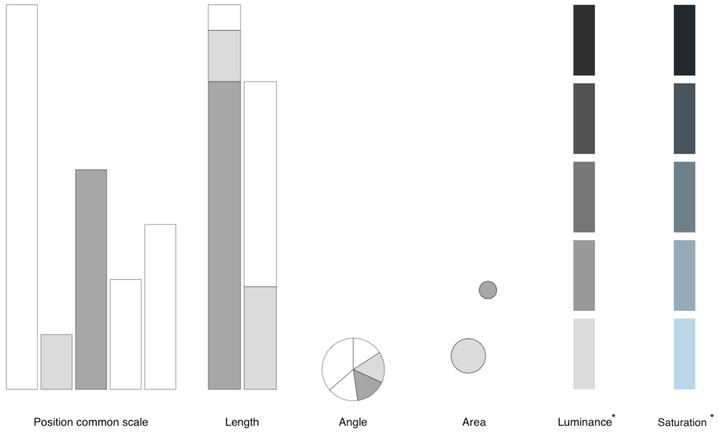
We worked with IR. IR stands for Information Resources and is a new department.

We worked with the recently launched Information Resources (IR) department to …

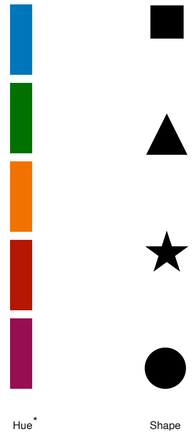
exploring to explaining, *adapting to your audience* — first, optimize encodings *for a communication purpose*



ratio, interval, and ordered



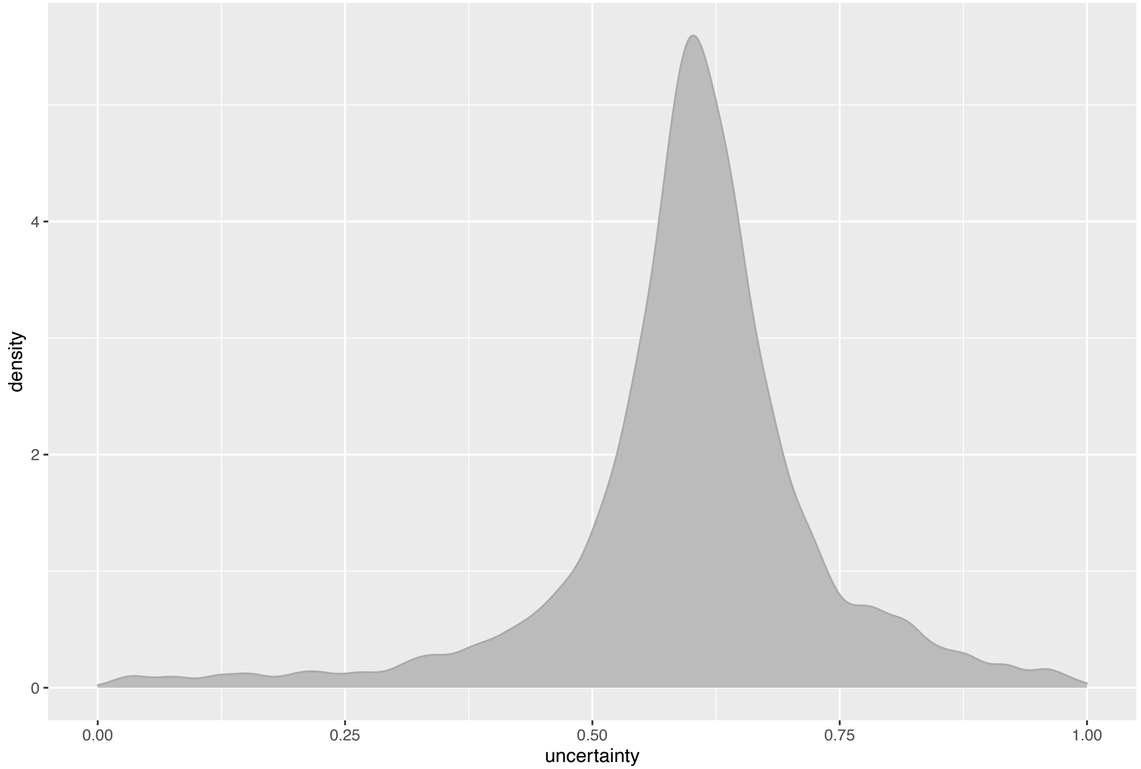
categorical



exploring to explaining, *adapting to your audience* — then change annotations, *not* optimized encodings

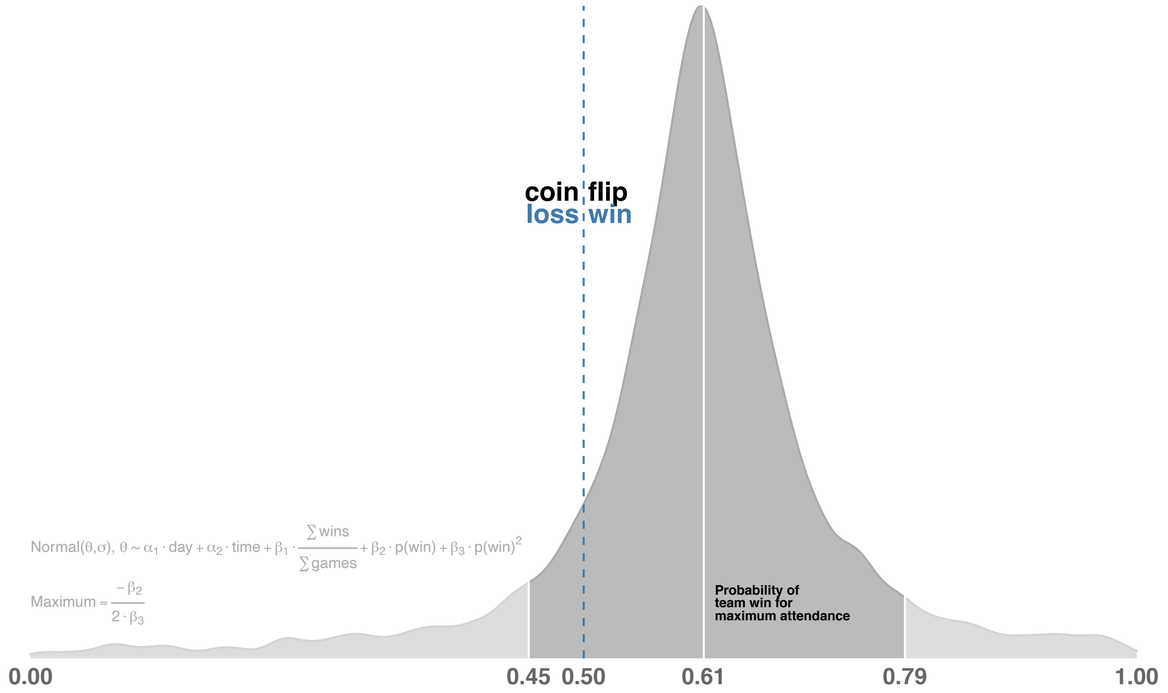
Once data-to-visual encodings have been optimized for showing the intended comparison or trend of interest to our audience, we should generally ***adapt it to our audience by explaining***, not by changing optimized encodings.

exploring to explaining, titles, as an overall graphics annotation, should explain the point of the graphics



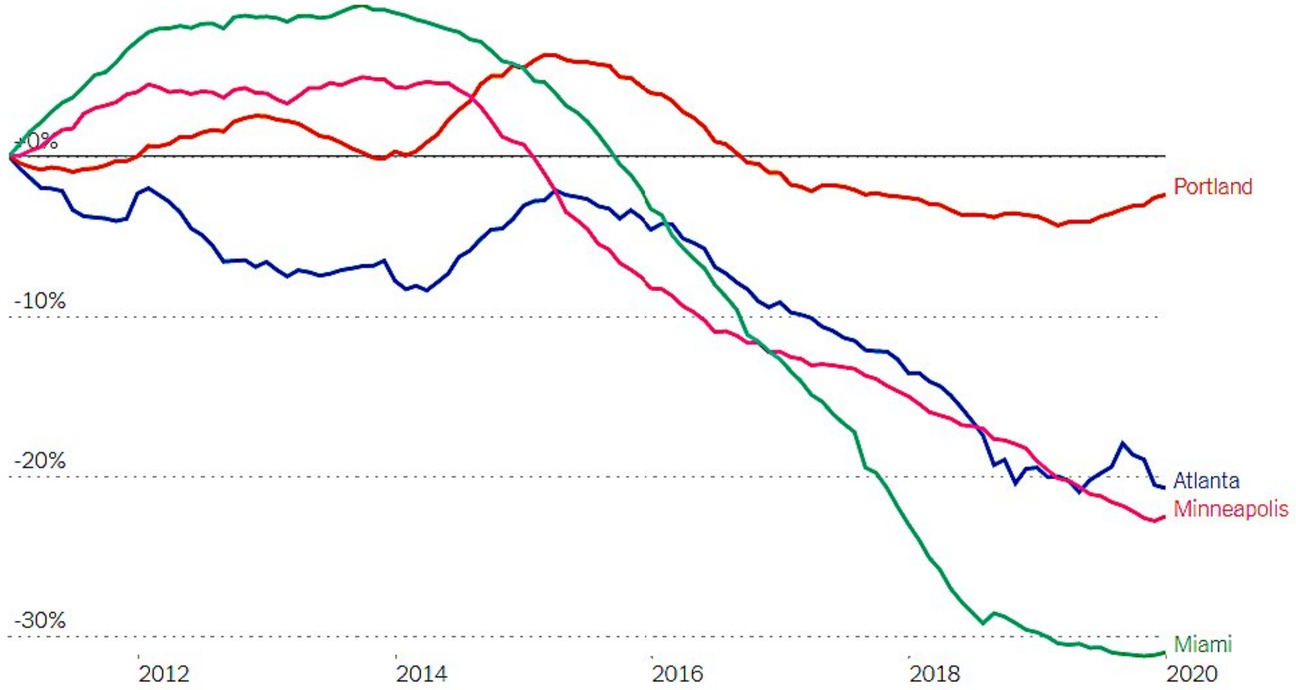
More fans generally pay admission to our games when the chance of winning was near a median of 0.61.

Fans want favorable odds without predicting the outcome.

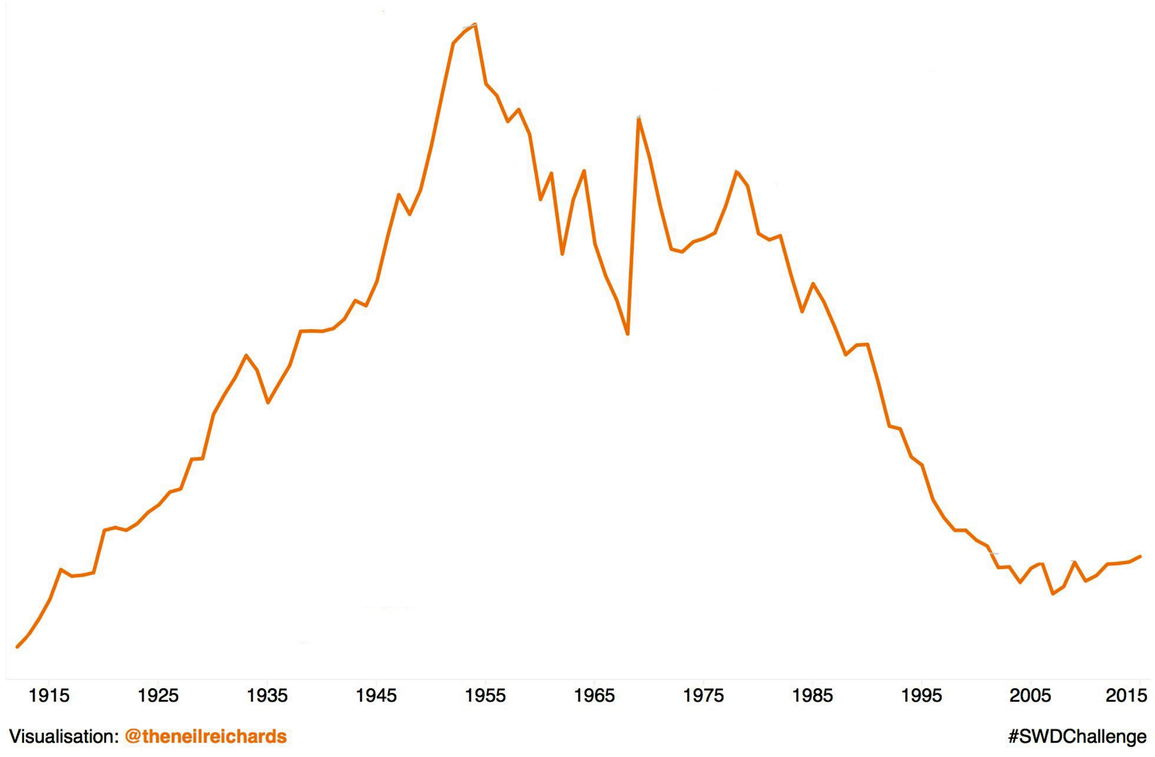


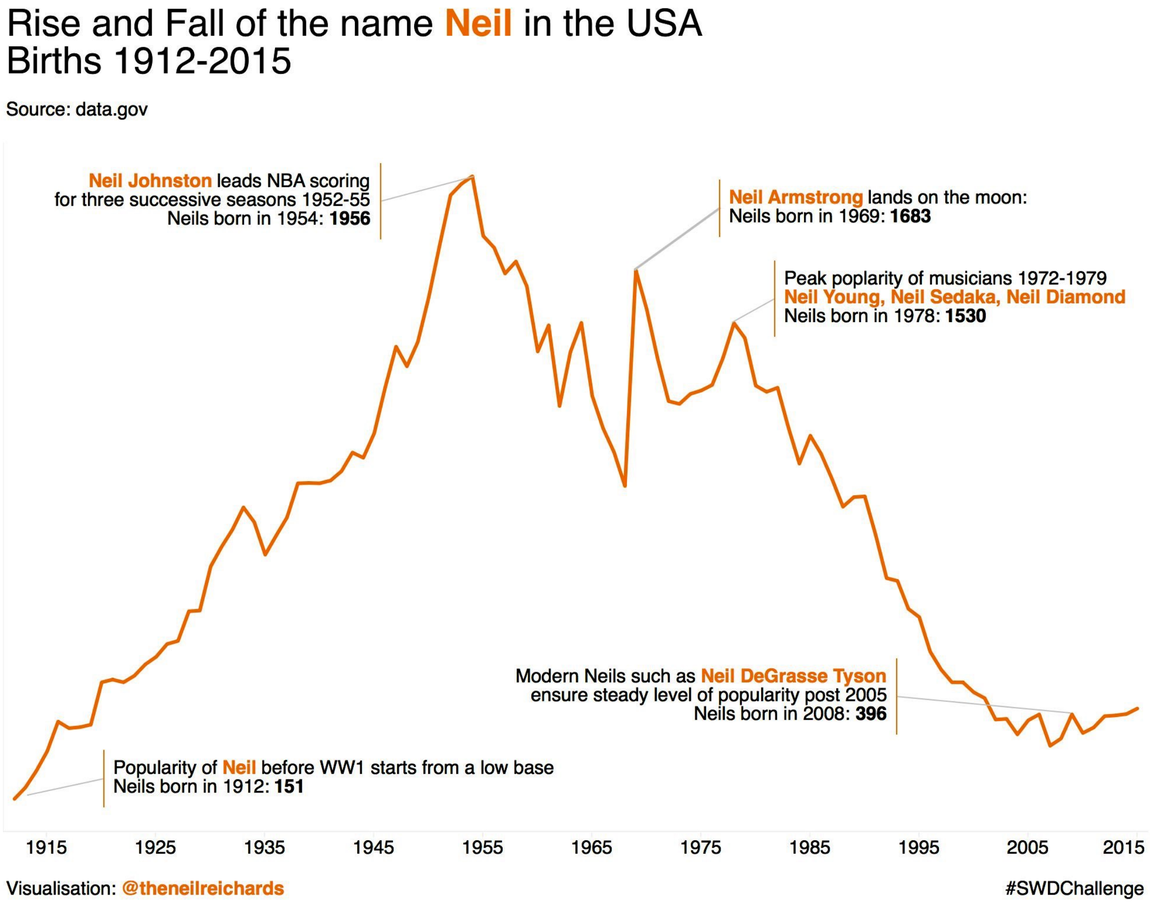
Sources: Pinnacle betting data, 2016; Retrosheet Gamelogs, 2016

exploring to explaining, replacing legends with direct data labeling reduces cognitive load



exploring to explaining, annotations go beyond labels, they can help *explain* and *compare* with *context*





exploring to explaining, annotations *as important as* data encodings

"The **annotation** layer is the *most important thing* we do ... otherwise it's a case of 'here it is, you go figure it out'.

— Amanda Cox, the Data Editor at the *New York Times*

"Although our primary focus on creating a visualization is the graphic elements—bars, points, or lines—the **text we include in and around our graphs** is *just as important*."

— Jonathan Schwabish, *Better Data Visualizations*

"**Annotations** are of *vital importance*. Often overlooked, **annotations** are one of the best ways to make a chart understandable to an audience. Underutilized in many data visualizations, **annotations** are the ideal way to highlight exactly those things that you, as the creator, want the audience to pay attention to."

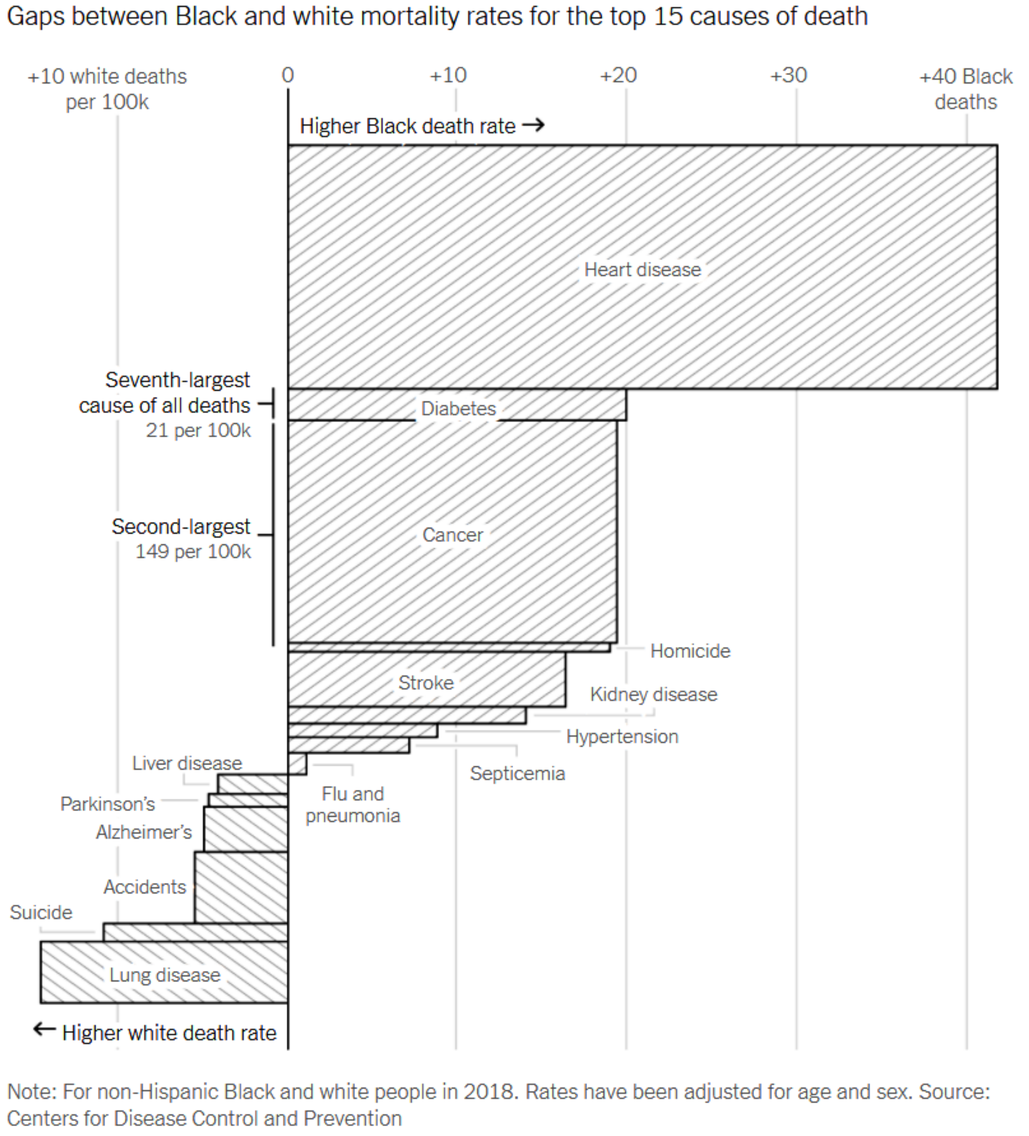
— Shirley Wu, *Data Sketches*

exploring to explaining, (focus) our audience on explained, visual encodings

remove clutter | start with gray

audiences understand graphic complexity — once you explain!

explain for audience, external or general audiences *can understand* complex graphics, *with guidance*

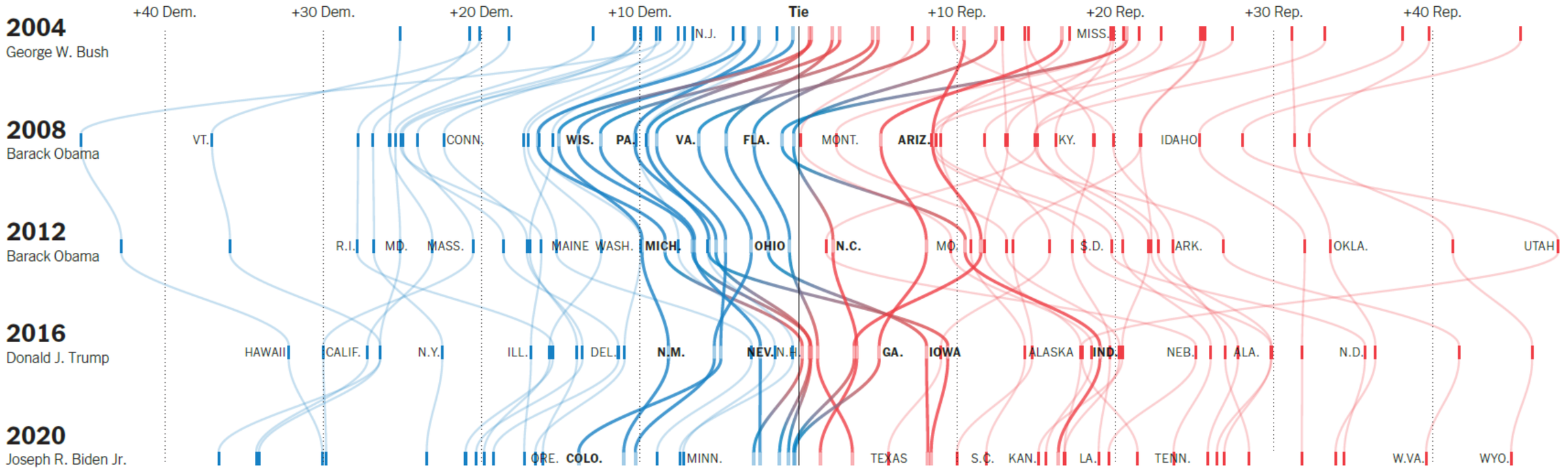


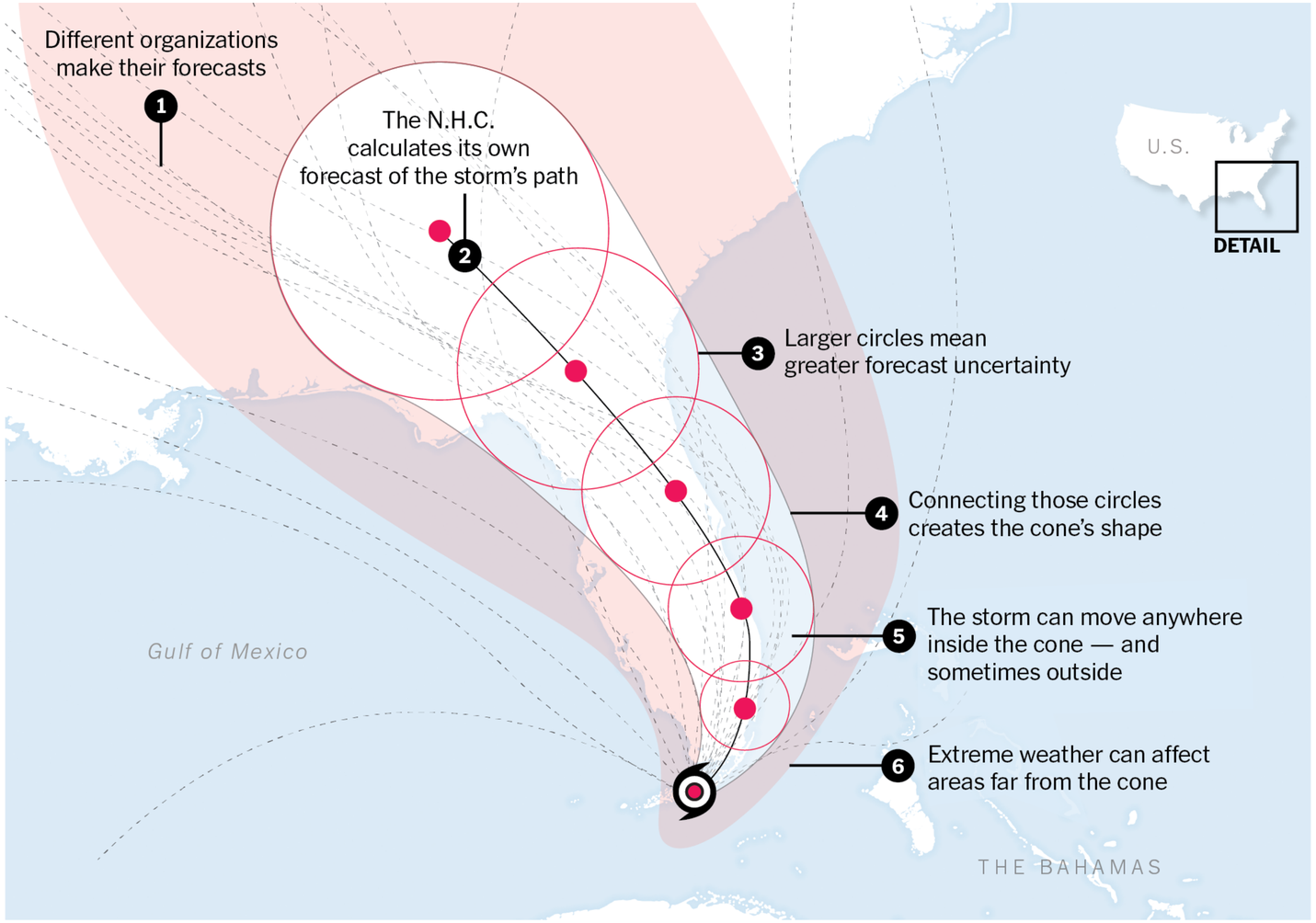
with guidance

Wezerek, Gus. "Racism's Hidden Toll." *The New York Times*, August 11, 2020, sec. Opinion. https://www.nytimes.com/interactive/2020/08/11/opinion/us-coronavirus-black-mortality.html.

explain for audience, external or general audiences *can understand* complex graphics, *with guidance*

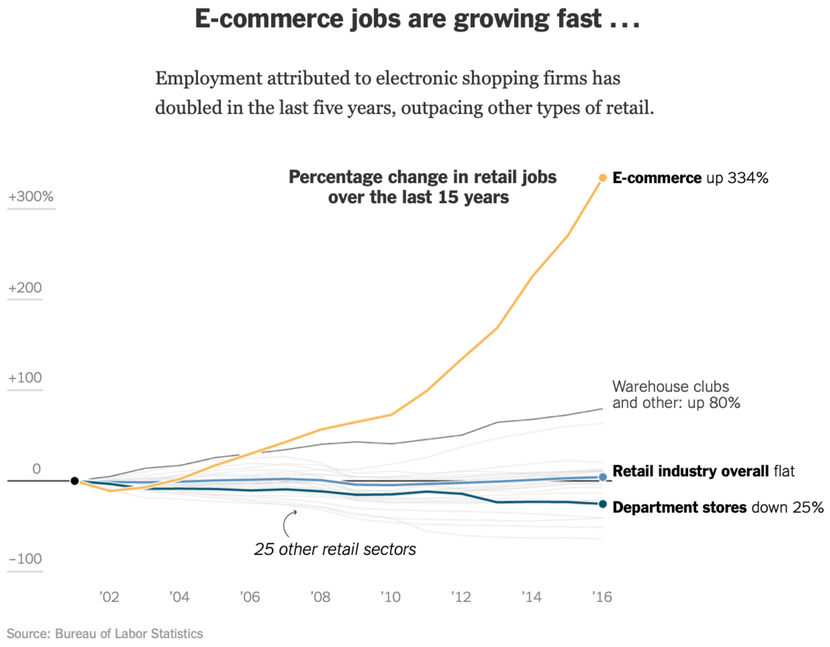
> ***with guidance***
>
> Lu, Denise, and Karen Yourish. "How Did Trump Do in Counties That Backed Him in 2016?" *The New York Times*, November 11, 2020, sec. Politics. https://www.nytimes.com/interactive/2020/11/09/us/politics/2016-election-trump-counties.html .

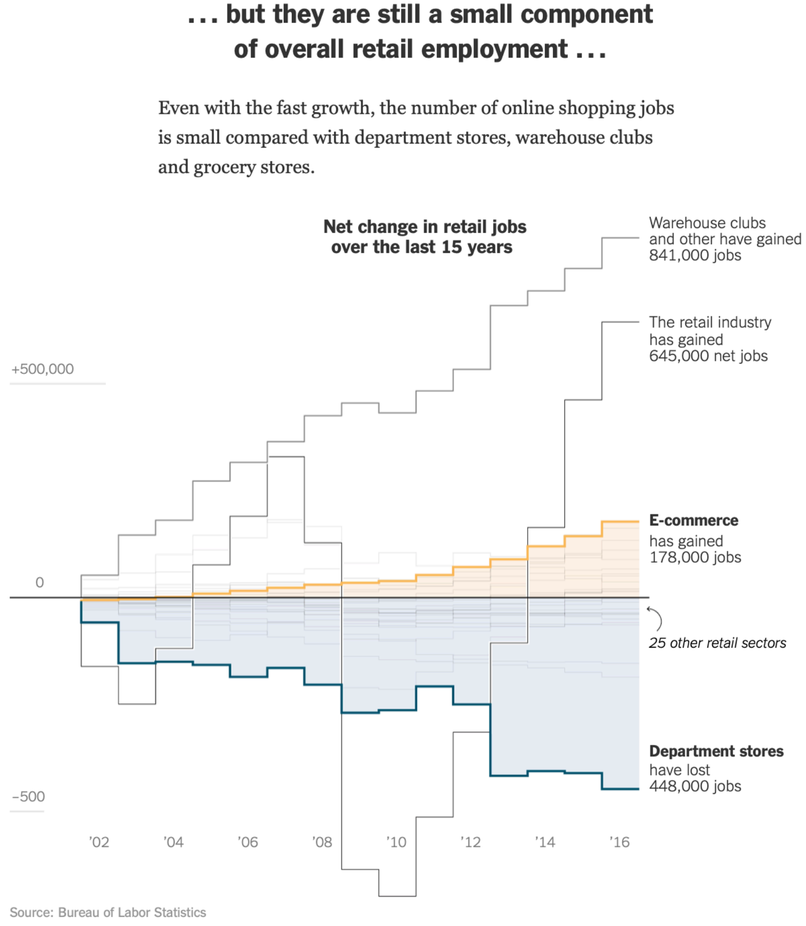




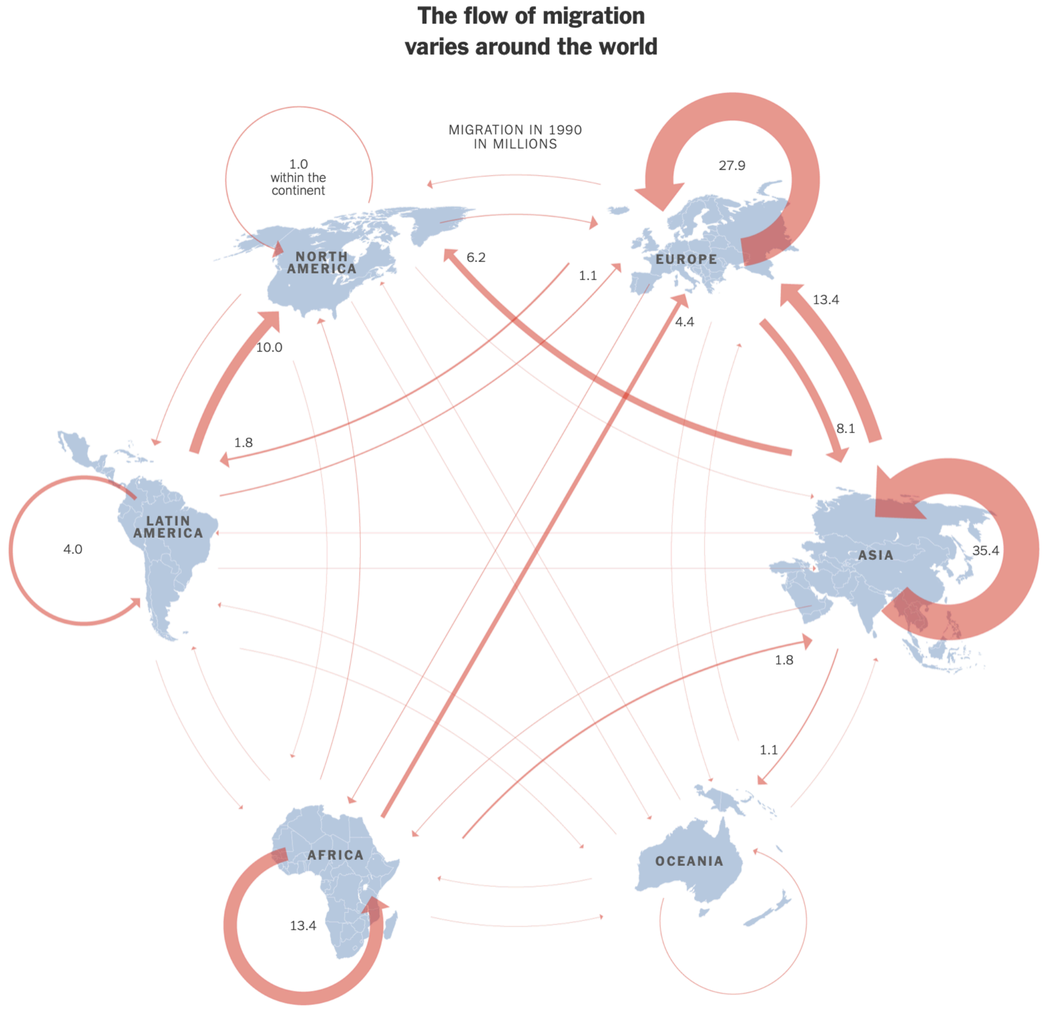
> ***with guidance***
>
> Cairo, Alberto, and Tala Schlossberg. "Those Hurricane Maps Don't Mean What You Think They Mean." *The New York Times*, March 10, 2019, sec. Opinion. https://www.nytimes.com/interactive/2019/08/29/opinion/hurricane-dorian-forecast-map.html.

explain for audience, external or general audiences *can understand* complex graphics, *with guidance*





> ***with guidance***
>
> Bebeloff, Robert, and Karl Russell. "How the Growth of E-Commerce Is Shifting Retail Jobs." *The New York Times*, July 6, 2017, sec. Business. https://www.nytimes.com/interactive/2017/07/06/business/ecommerce-retail-jobs.html.



Largest diaspora populations in 1990 in millions

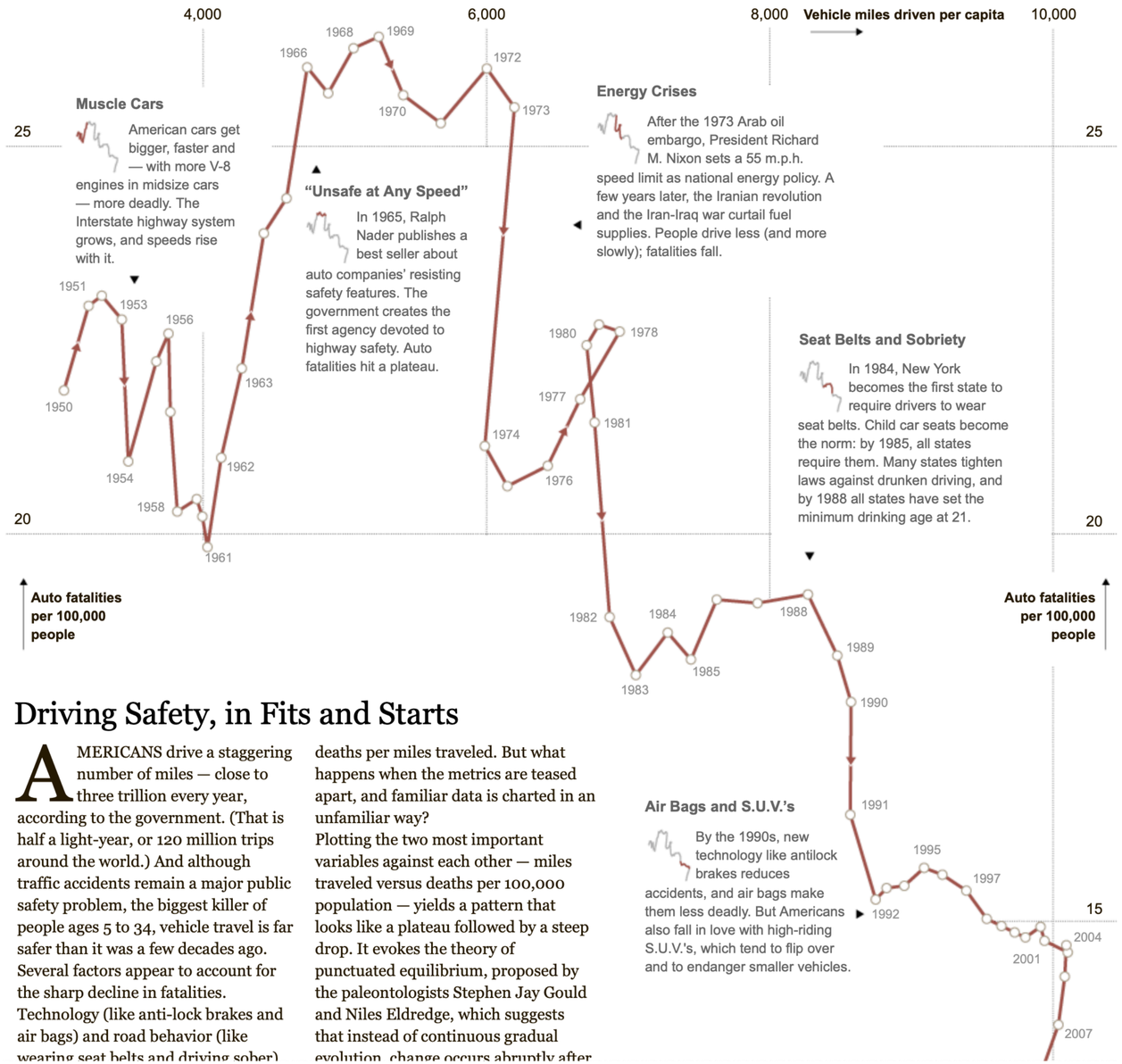
1. Russia	12.7	6. Mexico	4.4	11. Germany	3.3	16. Portugal	1.9	21. Indonesia	1.6	26. Iraq	1.5
2. Afghanistan	6.7	7. China	4.2	12. Kazakhstan	3.0	17. Palestine	1.8	22. Azerbaijan	1.6	27. Spain	1.4
3. India	6.7	8. Britain	3.8	13. Turkey	2.5	18. Belarus	1.8	23. South Korea	1.6	28. Uzbekistan	1.4
4. Ukraine	5.5	9. Italy	3.4	14. Mozambique	2.2	19. United States	1.7	24. Morocco	1.6	29. Egypt	1.3
5. Bangladesh	5.5	10. Pakistan	3.3	15. Philippines	2.0	20. Ethiopia	1.7	25. Poland	1.5	30. El Salvador	1.2

Note: Immigration flows from unknown origins not shown. | Source: United Nations Department of Economic and Social Affairs, Population Division

with guidance

Porter, Eduardo, and Karl Russell. "Migrants Are on the Rise Around the World, and Myths About Them Are Shaping Attitudes." *The New York Times*, June 20, 2018, sec. Economy. https://www.nytimes.com/interactive/2018/06/20/business/economy/immigration-economic-impact.html.

explain for audience, external or general audiences *can understand* complex graphics, *with guidance*



Driving Safety, in Fits and Starts

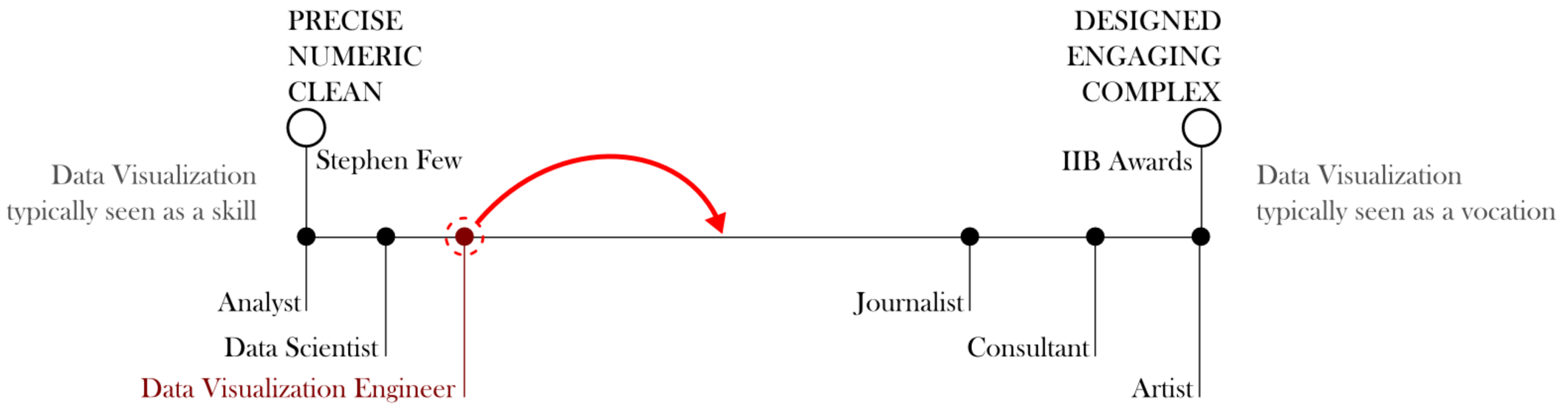
AMERICANS drive a staggering number of miles — close to three trillion every year, according to the government. (That is half a light-year, or 120 million trips around the world.) And although traffic accidents remain a major public safety problem, the biggest killer of people ages 5 to 34, vehicle travel is far safer than it was a few decades ago. Several factors appear to account for the sharp decline in fatalities. Technology (like anti-lock brakes and air bags) and road behavior (like wearing seat belts and driving sober) deaths per miles traveled. But what happens when the metrics are teased apart, and familiar data is charted in an unfamiliar way?

Plotting the two most important variables against each other — miles traveled versus deaths per 100,000 population — yields a pattern that looks like a plateau followed by a steep drop. It evokes the theory of punctuated equilibrium, proposed by the paleontologists Stephen Jay Gould and Niles Eldredge, which suggests that instead of continuous gradual evolution, change occurs abruptly after

> ***with guidance***
>
> Fairfield, Hannah. "Driving Safety, in Fits and Starts." *The New York Times*, September 12, 2012, sec. Science. https://archive.nytimes.com/www.nytimes.com/interactive/2012/09/17/science/driving-safety-in-fits-and-starts.html.

> ***empirical study on audience engagement***
>
> Haroz, Steve, Robert Kosara, and Steven L. Franconeri. "The Connected Scatterplot for Presenting Paired Time Series." *IEEE Transactions on Visualization and Computer Graphics* 22, no. 9 (September 1, 2016): 2174–86. https://doi.org/10.1109/TVCG.2015.2502587.

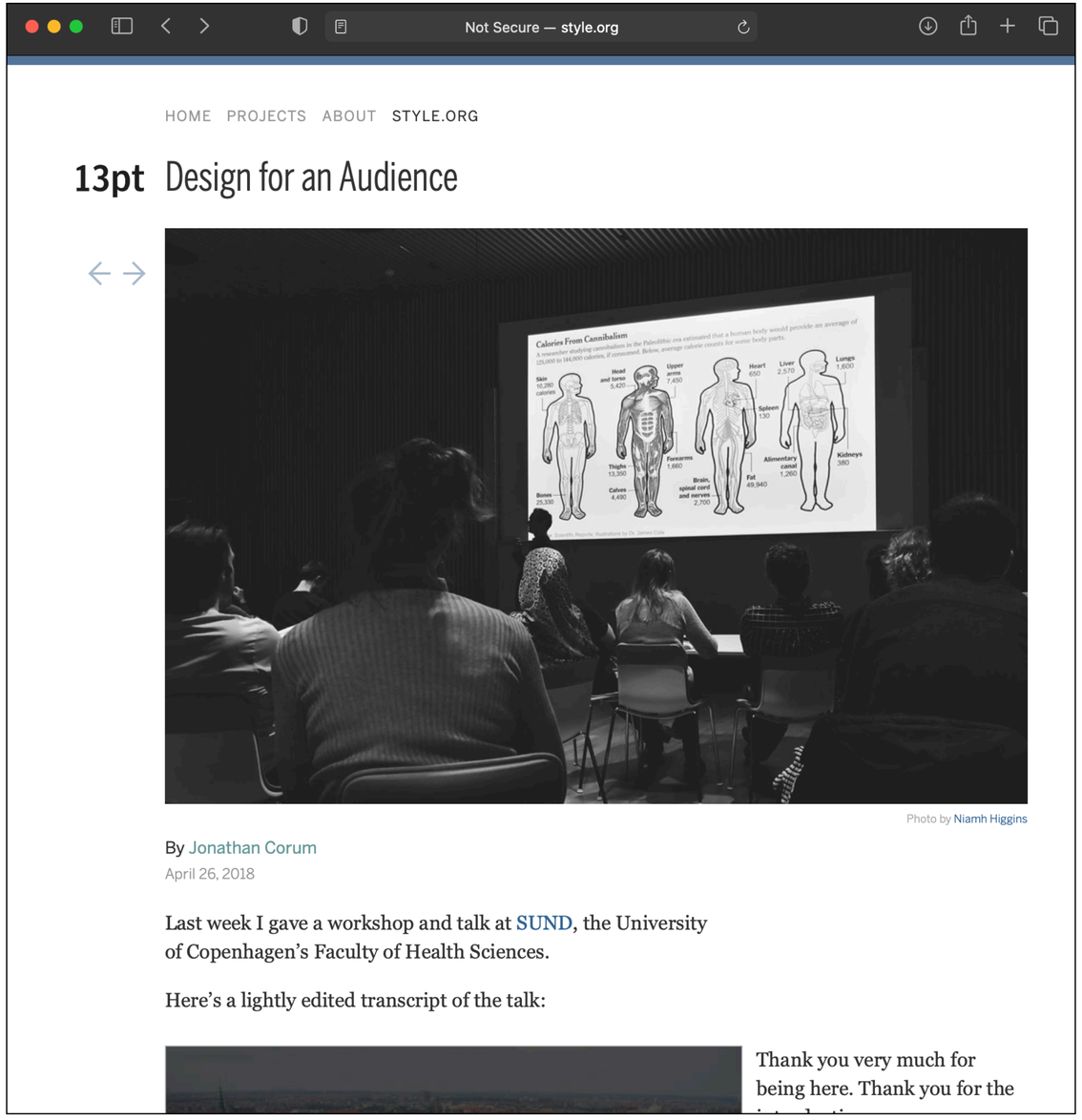
explain for audience, be open to new approaches for visual communication when designing for an audience



— Meeks, Elijah. "*If Data Visualization Is So Hot, Why Are People Leaving?*" Blog. Medium, March 21, 2017. https://medium.com/@Elijah_Meeks/why-people-leave-their-data-viz-jobs-be1a7ab5dddc.

(re)design for your audience

redesigns, advice from Jonathan Corum, Science Editor at the New York Times



Find the visual idea

Translate

Tell a visual story

Focus attention, don't scatter it

Show the content, not the frame

Show the content, not the table

Be consistent

What can you remove?

Reference the real world

Connect images and data

Explain why

Provide context

Build a sequence

Show movement and change

Encourage visual comparisons

More labels, fewer legends

Annotate

redesigns, example — original graphic within government publication explaining part of US economy

Bureau of Economic Analysis
U.S. DEPARTMENT OF COMMERCE

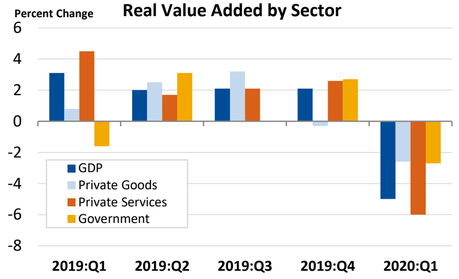
Monday, July 6, 2020
Contact: Jeannine Aversa, (301) 278-9003

Gross Domestic Product by Industry: First Quarter 2020

Accommodation and food services; finance and insurance; and health care and social assistance industries were the leading contributors to the 5.0 percent (annual rate) decrease in gross domestic product (GDP) in the first quarter of 2020.

All sectors of the U.S. economy contributed to the decrease, led by a decline in private services-producing industries. The decline in first-quarter GDP reflected the response to the spread of COVID-19, as governments issued "stay-at-home" orders in March. This led to rapid changes in production, as businesses and schools switched to remote work or canceled operations, and consumers and businesses canceled, restricted, or redirected their spending. For more information, see "Federal Recovery Programs and BEA Statistics: COVID-19 and Recovery" on the BEA website.

Real GDP and Real Value Added by Sector

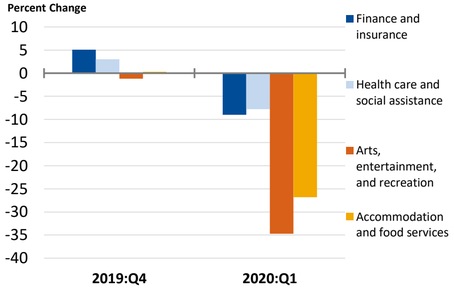


U.S. Bureau of Economic Analysis Seasonally adjusted annual rates

Overall, 17 of 22 industry groups contributed to the first-quarter decline in real GDP. Of the five industry groups that offset the decline in the first-quarter real GDP, agriculture, forestry, fishing, and hunting was the largest contributor, increasing 15.5 percent.

For accommodation and food services, real value added—a measure of an industry's contribution to GDP—decreased 26.8 percent, primarily reflecting a decrease in food services and drinking places.

Real Value Added by Selected Industries



U.S. Bureau of Economic Analysis Seasonally adjusted annual rates

Finance and insurance decreased 9.0 percent, primarily due to a decrease in insurance carriers and related activities.

Health care and social assistance decreased 7.8 percent, primarily reflecting decreases in ambulatory health care services and in hospitals.

Arts, entertainment and recreation decreased 34.7 percent, primarily reflecting a decrease in performing arts, spectator sports, museums, and related activities.

BEA statistics—including GDP, personal income, the balance of payments, foreign direct investment, the input-output accounts, and economic data for states, local areas, and industries—are available at www.bea.gov. E-mail alerts are also available.

© 2021 Scott Spencer / https://ssp3nc3r.github.io scott.spencer@columbia.edu

redesigns, example — what's the point of this graphic? Do encodings intuitively show the point? Let's redesign!

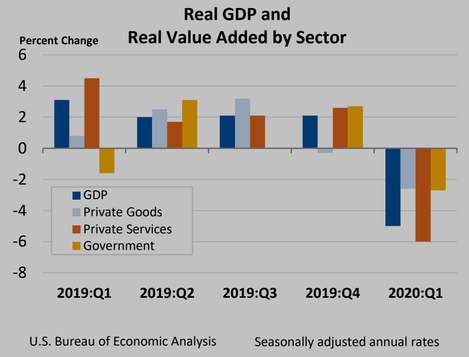
bea
Bureau of Economic Analysis
U.S. DEPARTMENT OF COMMERCE

Monday, July 6, 2020
Contact: Jeannine Aversa, (301) 278-9003

Gross Domestic Product by Industry: First Quarter 2020

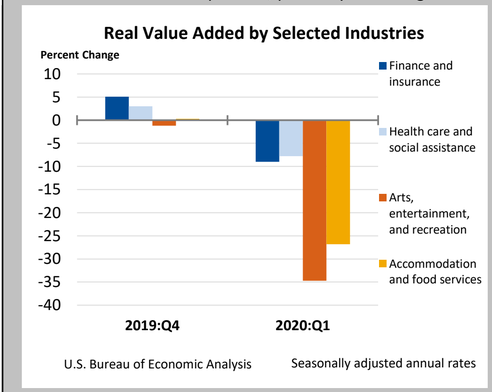
Accommodation and food services; finance and insurance; and health care and social assistance industries were the leading contributors to the 5.0 percent (annual rate) decrease in gross domestic product (GDP) in the first quarter of 2020.

All sectors of the U.S. economy contributed to the decrease, led by a decline in private services-producing industries. The decline in first-quarter GDP reflected the response to the spread of COVID-19, as governments issued "stay-at-home" orders in March. This led to rapid changes in production, as businesses and schools switched to remote work or canceled operations, and consumers and businesses canceled, restricted, or redirected their spending. For more information, see "Federal Recovery Programs and BEA Statistics: COVID-19 and Recovery" on the BEA website.



Overall, 17 of 22 industry groups contributed to the first-quarter decline in real GDP. Of the five industry groups that offset the decline in the first-quarter real GDP, agriculture, forestry, fishing, and hunting was the largest contributor, increasing 15.5 percent.

For accommodation and food services, real value added—a measure of an industry's contribution to GDP—decreased 26.8 percent, primarily reflecting a decrease in food services and drinking places.

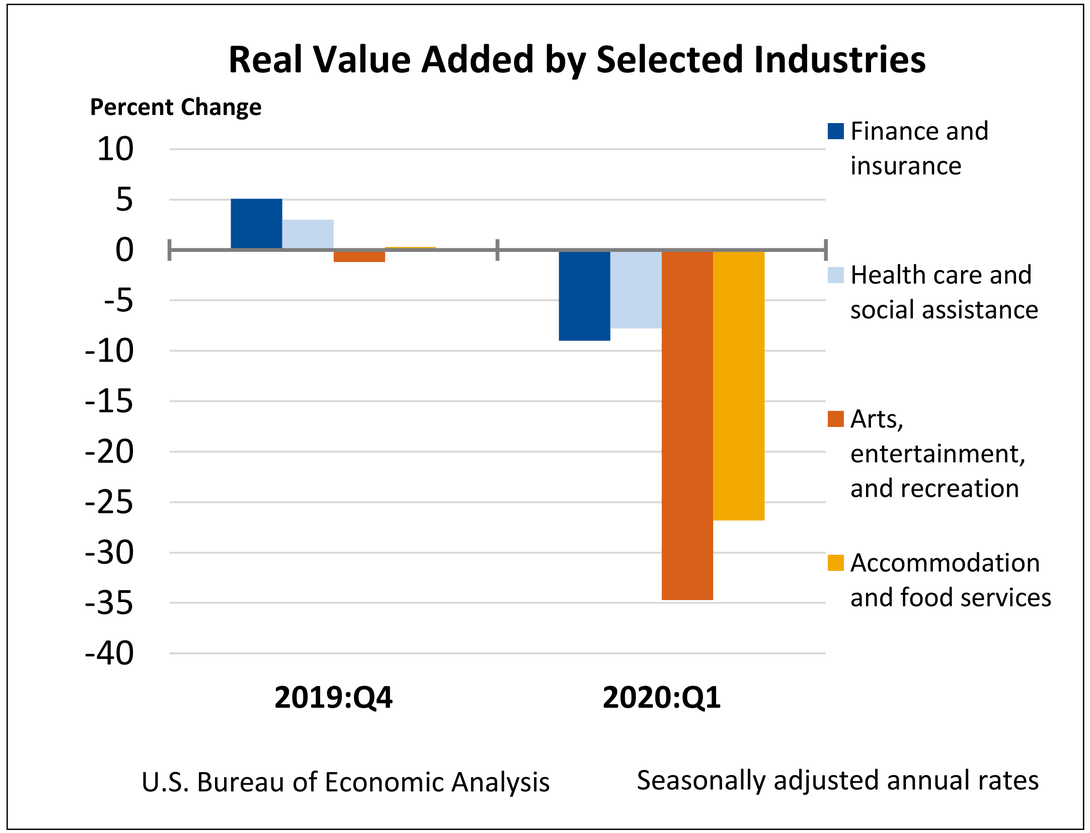


Finance and insurance decreased 9.0 percent, primarily due to a decrease in insurance carriers and related activities.

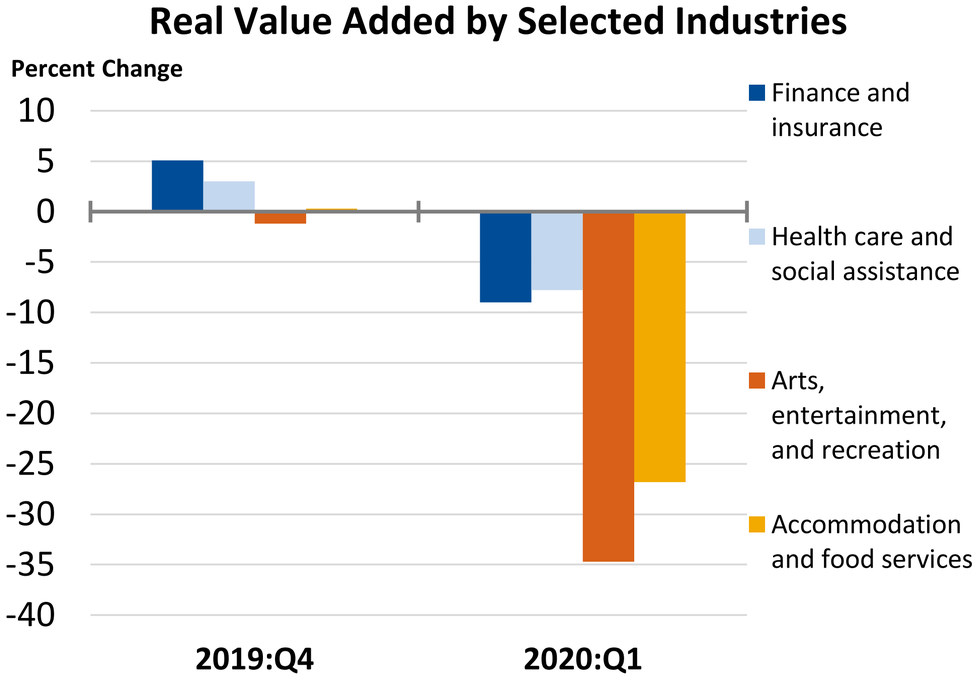
Health care and social assistance decreased 7.8 percent, primarily reflecting decreases in ambulatory health care services and in hospitals.

Arts, entertainment and recreation decreased 34.7 percent, primarily reflecting a decrease in performing arts, spectator sports, museums, and related activities.

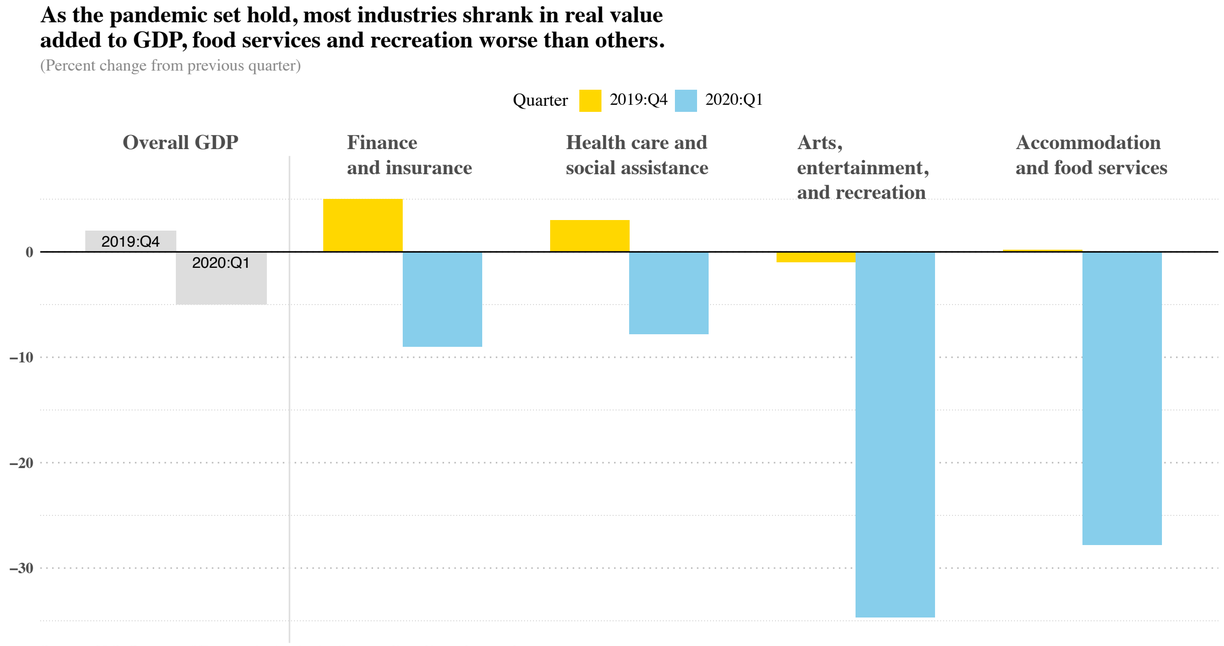
BEA statistics—including GDP, personal income, the balance of payments, foreign direct investment, the input-output accounts, and economic data for states, local areas, and industries—are available at www.bea.gov. E-mail alerts are also available.



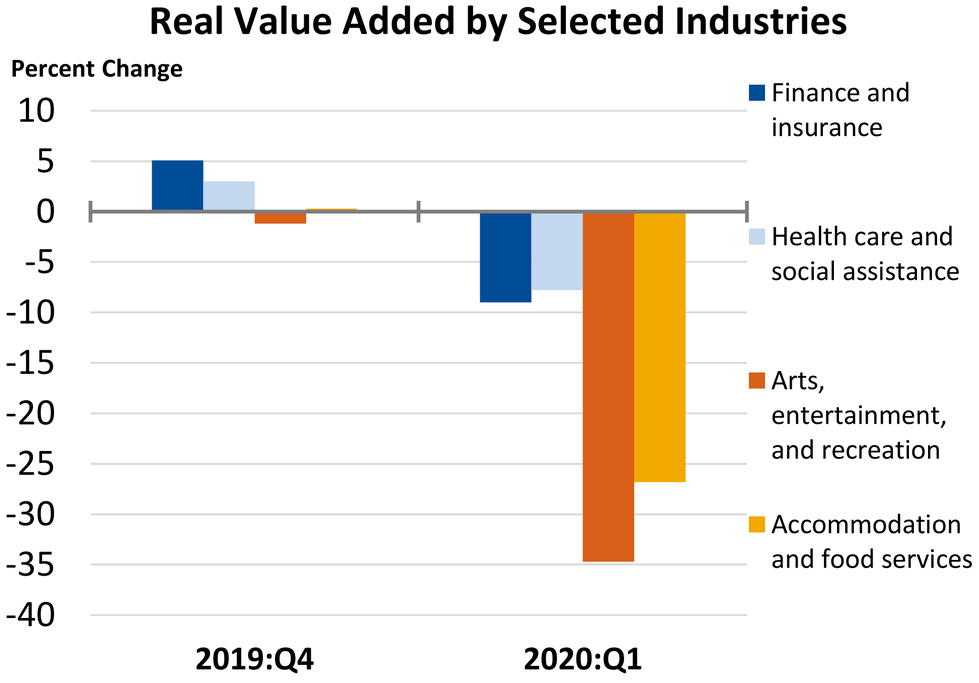
redesigns, example — first possible redesign. Does this redesign more intuitively convey a point?



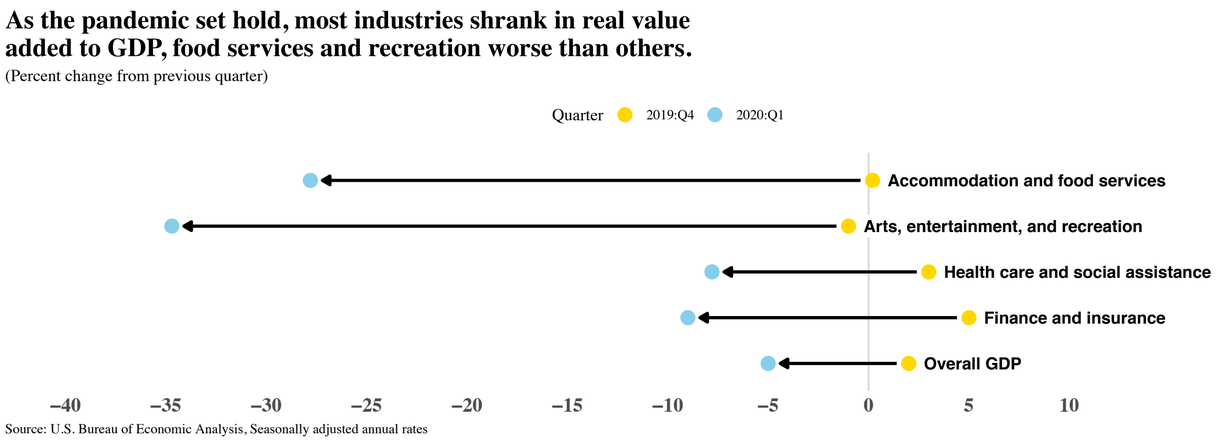
U.S. Bureau of Economic Analysis Seasonally adjusted annual rates



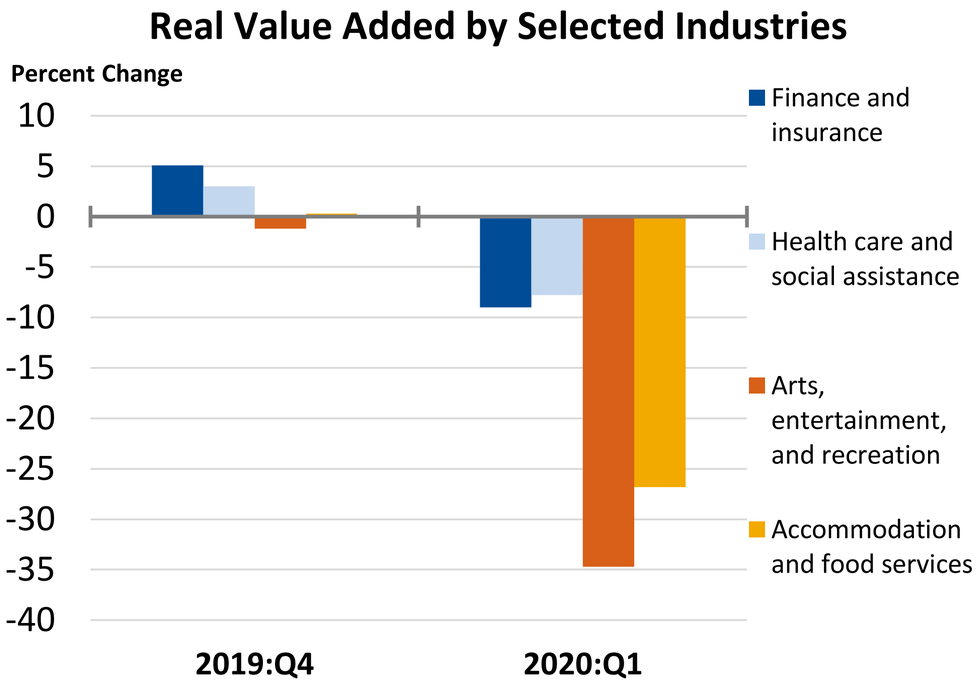
redesigns, example — second possible redesign. Does this redesign more intuitively convey a point?



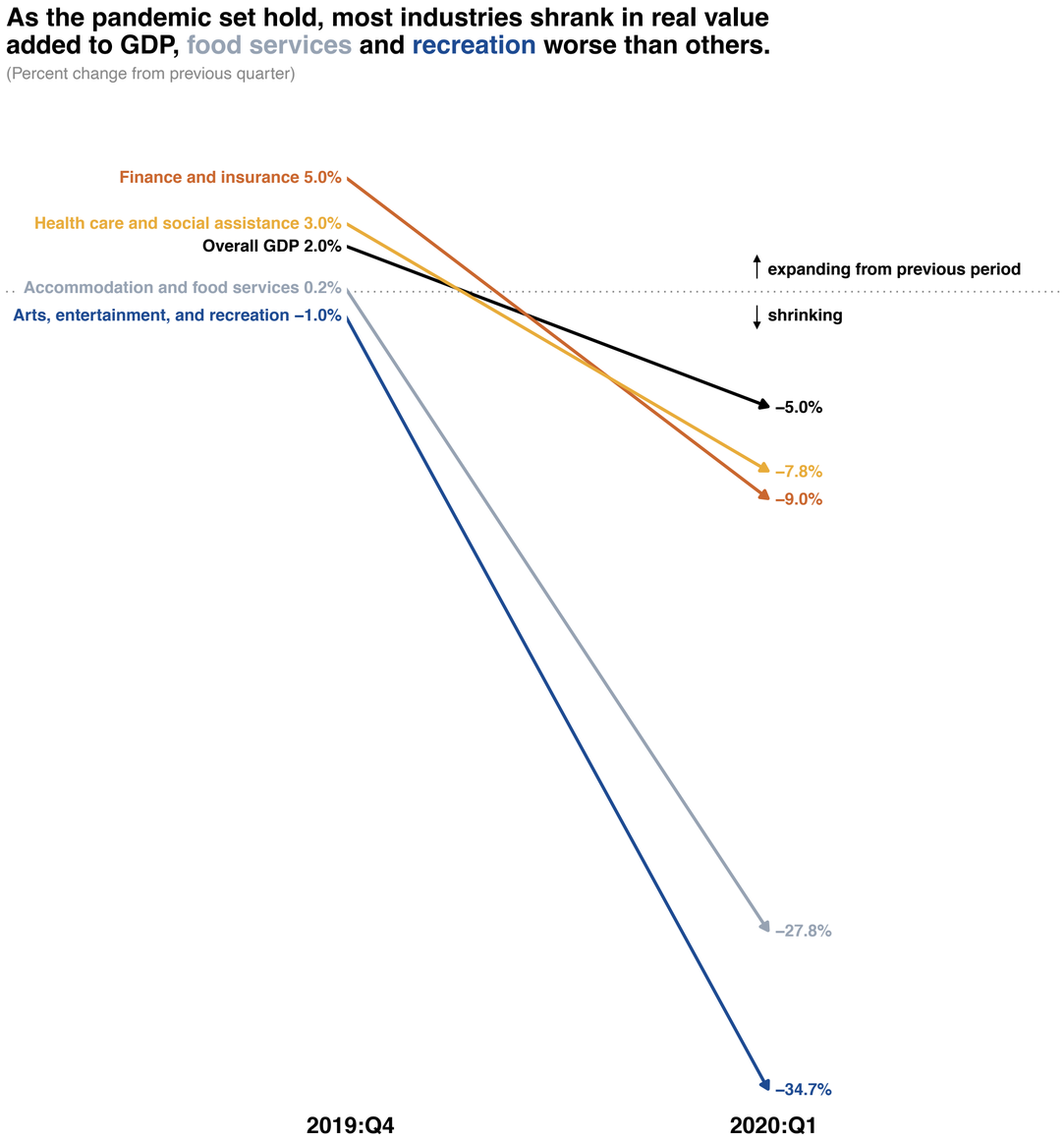
U.S. Bureau of Economic Analysis Seasonally adjusted annual rates



redesigns, example — third possible redesign. Does this redesign more intuitively convey a point?



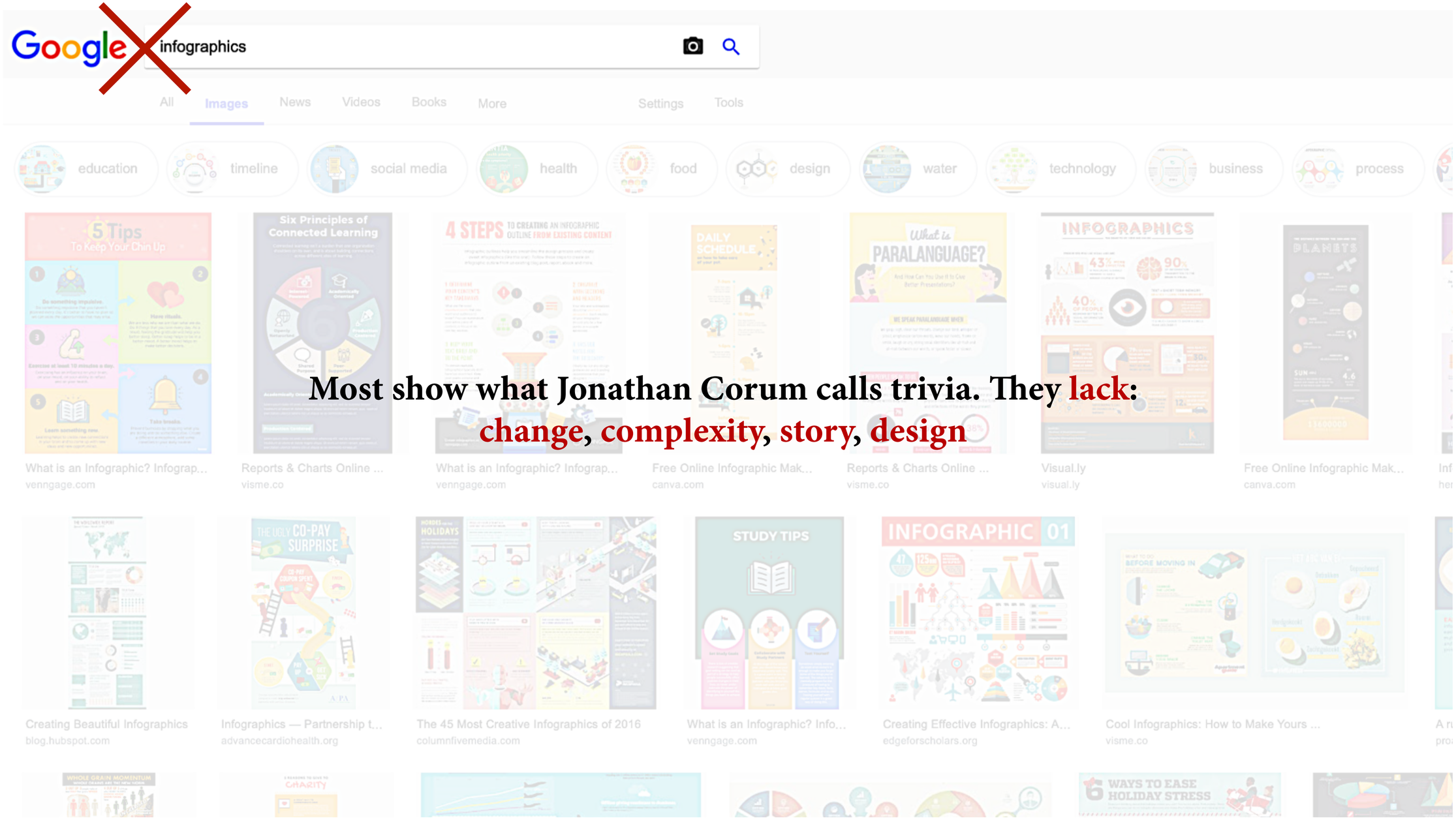
U.S. Bureau of Economic Analysis Seasonally adjusted annual rates



Source: U.S. Bureau of Economic Analysis, Seasonally adjusted annual rates

data graphics in storytelling

data graphics in storytelling, **information graphics — don't generically google this to learn!**



Most show what Jonathan Corum calls trivia. They lack:
change, complexity, story, design

INFOGRAPHIC n. a visual image such as a chart or diagram used to represent information or data in an easily understandable form.

Gregor Aisch

Infographics is an abbreviated form of "information graphics". It seems to mean a lot of different things to different people. I rarely use the term.

Federica Fragapane

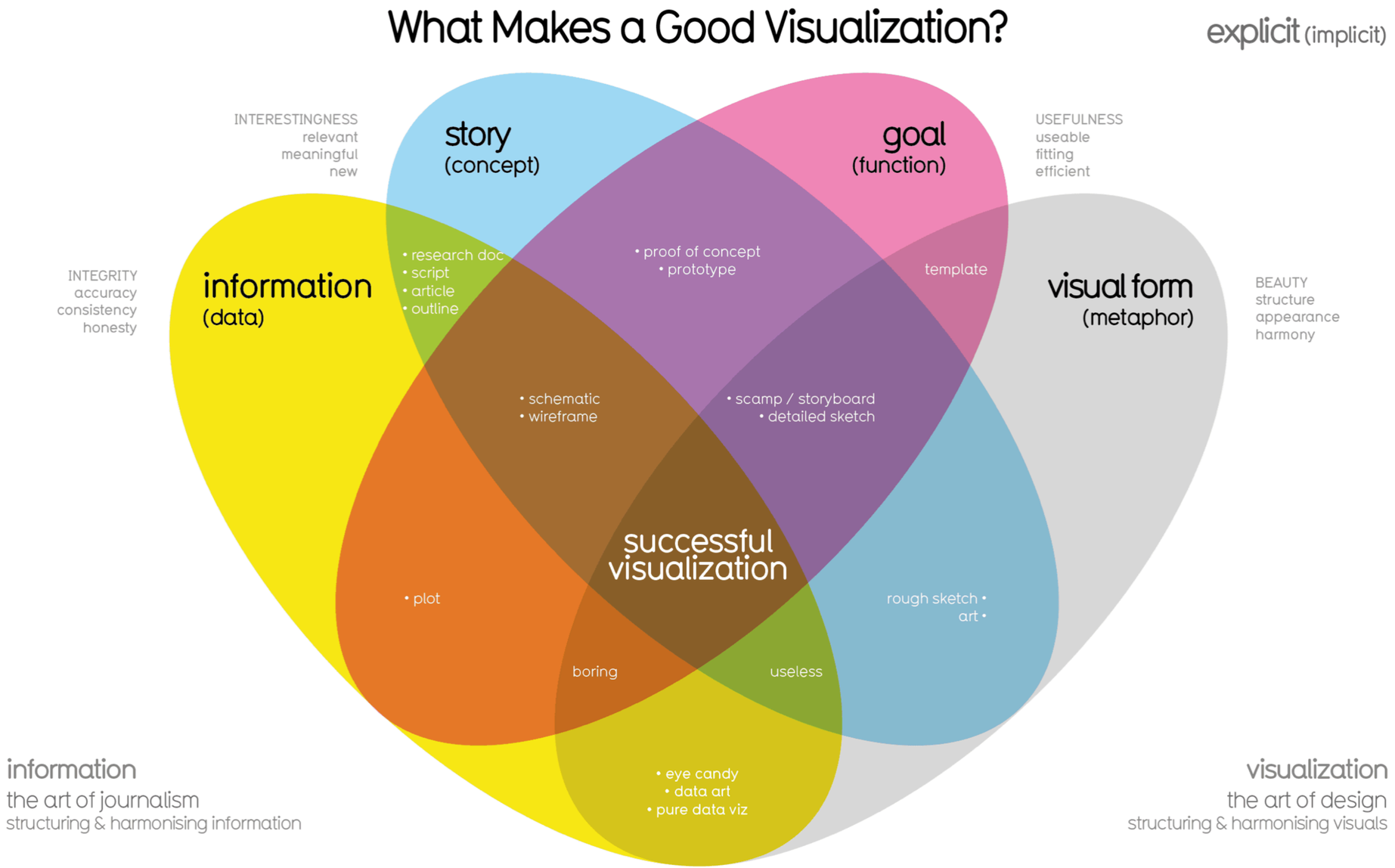
A visual translation of data and information: a language to communicate topics, contents and stories to people.

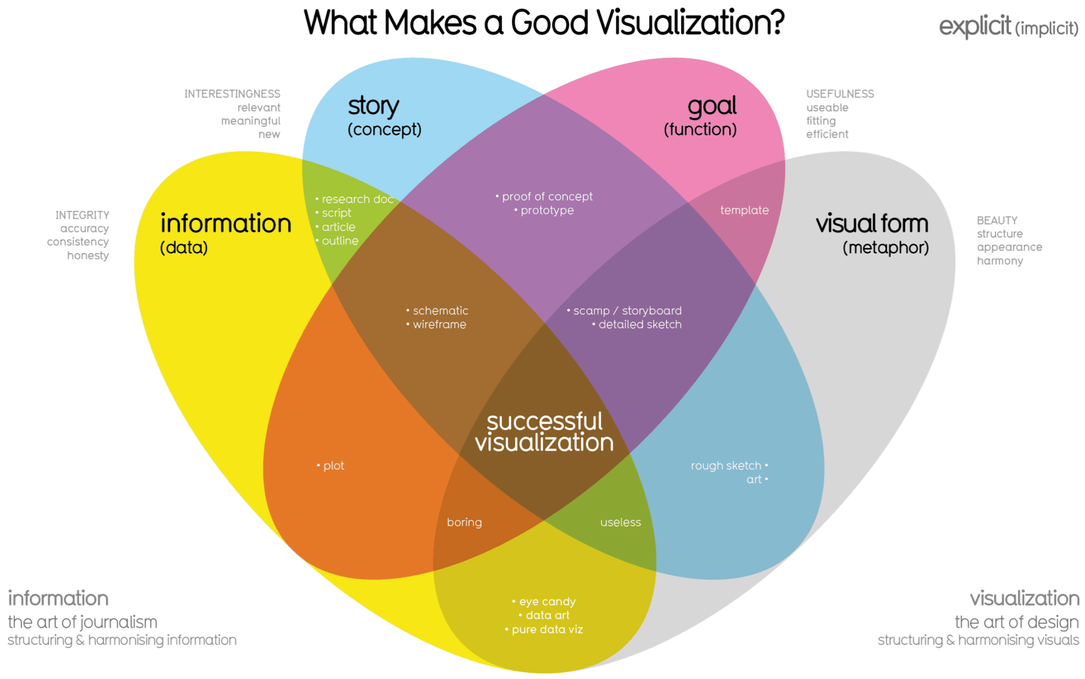
Laris Karklis

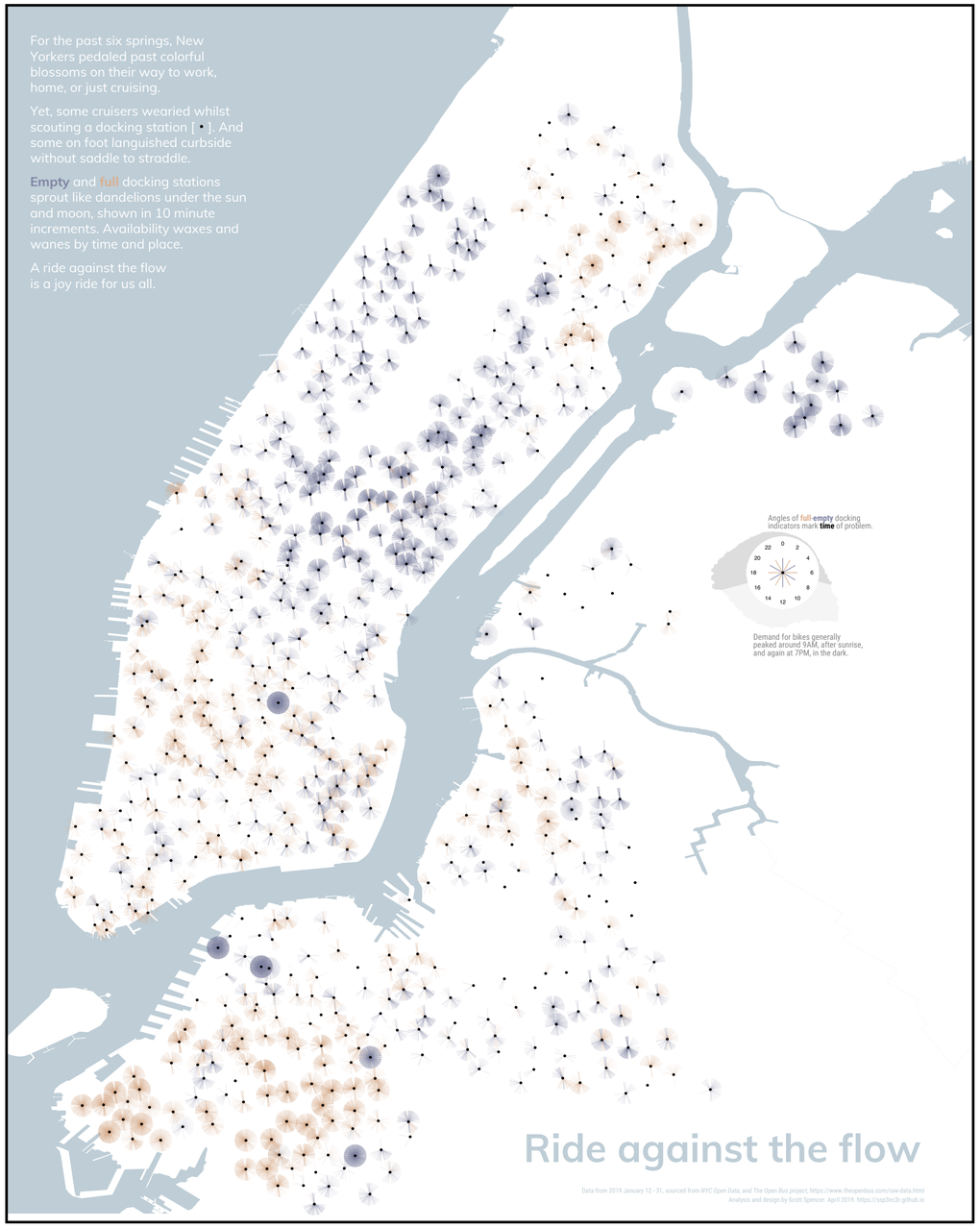
Infographics is . . . using visuals to tell a story.

Nadieh Bremer

Infographics … combine graphical elements, such as a drawn portion of an animal, human, map, etc. with small mini data visualizations (a small bar chart for example) and annotations to tell a story.


What Makes a Good Visualization?
explicit (implicit)
INTERESTINGNESS
relevant
meaningful
new
story
(concept)
goal
(function)
USEFULNESS
useable
fitting
efficient
INTEGRITY
accuracy
consistency
honesty
information
(data)
visual form
(metaphor)
BEAUTY
structure
appearance
harmony
• research doc
• script
• article
• outline
• proof of concept
• prototype
template
• schematic
• wireframe
• scamp / storyboard
• detailed sketch
successful
visualization
• plot
rough sketch •
art •
boring
useless
• eye candy
• data art
• pure data viz
information
the art of journalism
structuring & harmonising information
visualization
the art of design
structuring & harmonising visuals


What Makes a Good Visualization?
explicit (implicit)
INTERESTINGNESS
relevant
meaningful
new
story
(concept)
goal
(function)
USEFULNESS
useable
fitting
efficient
INTEGRITY
accuracy
consistency
honesty
information
(data)
visual form
(metaphor)
BEAUTY
structure
appearance
harmony
• research doc
• script
• article
• outline
• proof of concept
• prototype
template
• schematic
• wireframe
• scamp / storyboard
• detailed sketch
successful
visualization
• plot
rough sketch •
art •
boring
useless
• eye candy
• data art
• pure data viz
information
the art of journalism
structuring & harmonising information
visualization
the art of design
structuring & harmonising visuals


For the past six springs, New Yorkers pedaled past colorful blossoms on their way to work, home, or just cruising.
Yet, some cruisers wearied whilst scouting a docking station [•]. And some on foot languished curbside without saddle to straddle.
Empty and full docking stations sprout like dandelions under the sun and moon, shown in 10 minute increments. Availability waxes and wanes by time and place.
A ride against the flow is a joy ride for us all.
Angles of full-empty docking indicators mark time of problem.
Demand for bikes generally peaked around 9AM, after sunrise, and again at 7PM, in the dark.
Ride against the flow

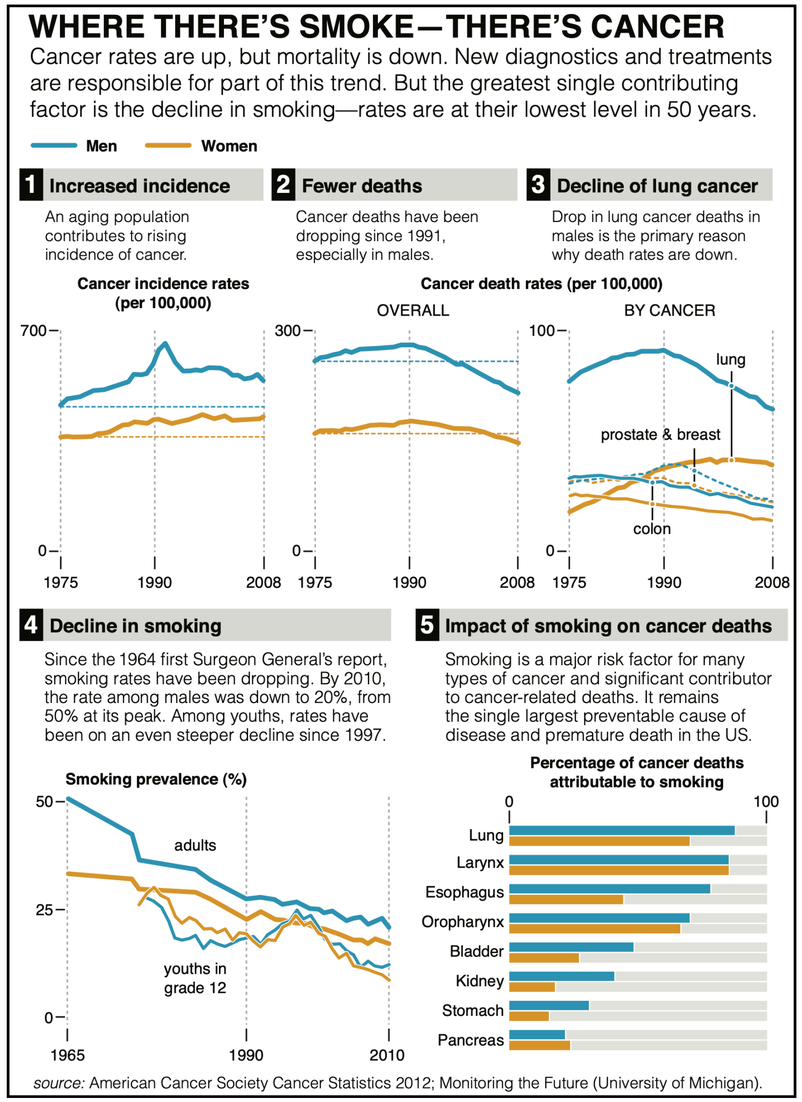
data graphics in storytelling, information graphics — are stories appropriate to show scientific results?

A story **adds meaning and clarity** to complex statistics.

Maintain focus … by leaving out detail that does not advance the plot. Distinguish necessary detail from minutiae; do not give in to the desire to show all your hard-won data. Provide sufficient support for your story, but stick to the plot.

Cairo & Krzywinski — "yes"

Use multiple panels to establish flow, and use colloquial language when addressing a general audience. . . . **Always be accurate**, but balance qualitative and quantitative expositions. An **occasional tangent . . . adds texture** to the presentation **without diluting the message**.



Use of graphics storytelling often result in a **distorted** and **unrepresentative** display of data. Great storytellers **embellish** and **conceal** information to evoke a response in their audience.

Katz — "no"

Inconvenient **truths are swept away**, and **marginalities are spun** to make a point more spectacular. A storyteller would plot the data in the way most persuasive rather than most informative or representative.

resources

References

Spencer, Scott. Sec. 2.2.1, 2.2.4 In *Data in Wonderland.* 2021. https://ssp3nc3r.github.io/data_in_wonderland.

Bertini, Enrico, and Moritz Stefaner. "Visual Storytelling w/ Alberto Cairo and Robert Kosara." In *Data Stories: A podcast on data visualization with Enrico Bertini and Moritz Stefaner*. Podcast, April 16, 2014. https://datastori.es/data-stories-35-visual-storytelling-w-alberto-cairo-and-robert-kosara/.

Bremer, Nadieh. *Data Sketches A Journey of Imagination, Exploration, and Beautiful Data Visualizations*. Milton, UNITED KINGDOM: A K Peters/CRC Press, 2021.

Corum, Jonathan. "See, Think, Design, Produce 3." 13pt Information Design, March 28, 2016. http://style.org/stdp3/.

———. "Design for an audience" 13pt Information Design, April 26, 2018. http://style.org/ku/.

Doumont, Jean-Luc. "Fundamentals." In *Trees, Maps, and Theorems. Effective Communication for Rational Minds*. Principiæ, 2009.

Koponen, Juuso, and Jonatan Hildén. *Data Visualization Handbook*. First. Finland: Aalto Art Books, 2019.

McCandless, David. "What Makes a Good Visualization?" *Information is Beautiful*. Accessed April 2, 2020. https://informationisbeautiful.net/visualizations/what-makes-a-good-data-visualization/.

Meeks, Elijah. "If Data Visualization Is So Hot, Why Are People Leaving?" Blog. Medium, March 21, 2017. https://medium.com/@Elijah_Meeks/why-people-leave-their-data-viz-jobs-be1a7ab5dddc.

Katz, Yarden. "Against Storytelling of Scientific Results." Nature Publishing Group 10, no. 11 (November 2013): 1045–1045.

Krzywinski, Martin, and Alberto Cairo. "Storytelling." Nature Publishing Group 10, no. 8 (August 2013): 687–687.

Krzywinski, Martin, and Alberto Cairo. "Reply to: Against Storytelling of Scientific Results." Nature Publishing Group 10, no. 11 (November 2013): 1046–1046.

Schwabish, Jonathan A. *Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks*. New York: Columbia University Press, 2021.

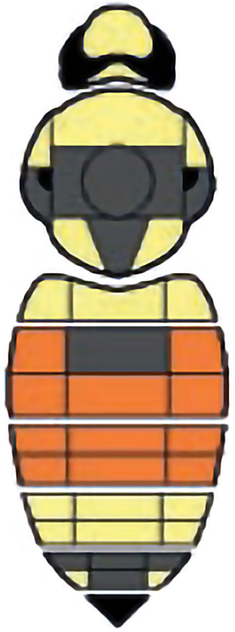
Spencer, Scott. *Ride Against the Flow*. 2019. Computer graphics. https://www.informationisbeautifulawards.com/showcase/4367-ride-against-the-flow.

Wilke, C. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. First edition. Sebastopol, CA: O'Reilly Media, 2019.

supplemental material

See, Think, Design, Produce

understand explain



Search for patterns by comparing

Visualization is not counting. Search for meaningful patterns, try to understand patterns, visualize patterns and try to explain them. Part of this is comparing. Another part is finding what's possible. Look at more ideas than you can use. Finally, practice — a lot!

Sketch until your aha! moment

Finding a clear thought through visualization can begin with sketching, on either paper or screen. Sketching is visual problem solving, not a commitment. It's much easier to begin with an ugly sketch and make it prettier as you work on design.

© 2021 Scott Spencer / https://ssp3nc3r.github.io scott.spencer@columbia.edu

— with more V-8 engines in midsize cars — more deadly. The Interstate highway system grows, and speeds rise with it.

"Unsafe at Any Speed"

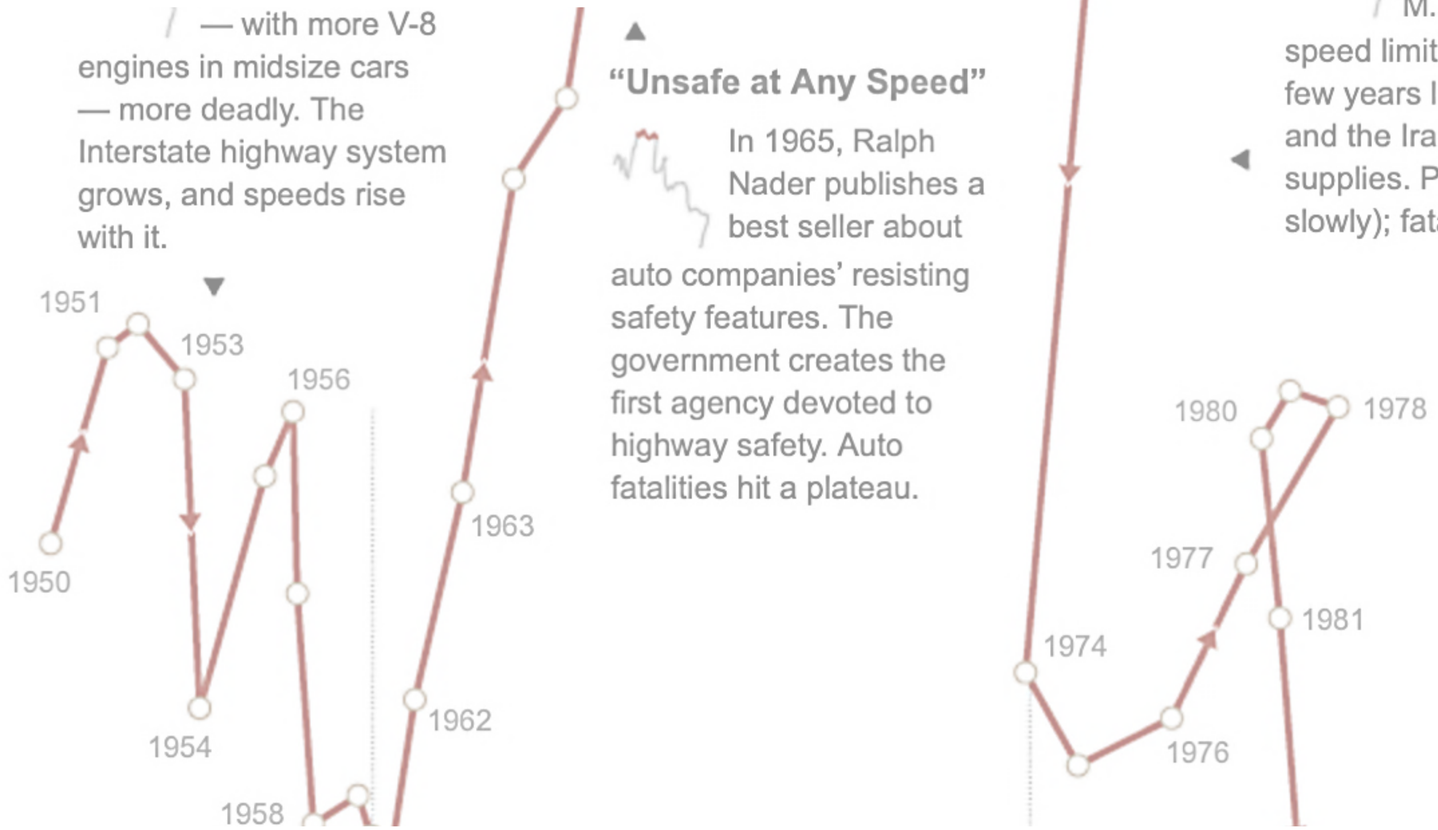
In 1965, Ralph Nader publishes a best seller about auto companies' resisting safety features. The government creates the first agency devoted to highway safety. Auto fatalities hit a plateau.

Energy Crises

After the 1973 Arab oil embargo, President Richard M. Nixon sets a 55 m.p.h. speed limit as national energy policy. A few years later, the Iranian revolution and the Iran-Iraq war curtail fuel supplies. People drive less (and more slowly); fatalities fall.

Seat Belts and Sobriety

In 1984, New York becomes the first state to require drivers to wear seat belts. Child car seats become the norm: by 1985, all states require them. Many states tighten laws against drunken driving, and by 1988 all states have set the drinking age at 21.

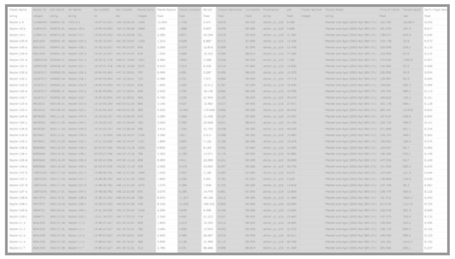


Design for someone else, show varying details

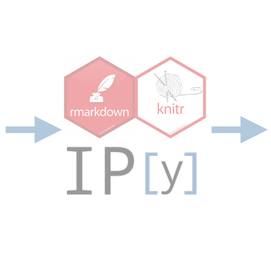
E.T. said "Good design is clear thinking made visible." The goal of design is to elegantly show your clear thought. Try to use a range of scales, or viewpoints, in what you show. Very important — show change, not trivia! Annotate.

© 2021 Scott Spencer / https://ssp3nc3r.github.io scottspencer@columbia.edu

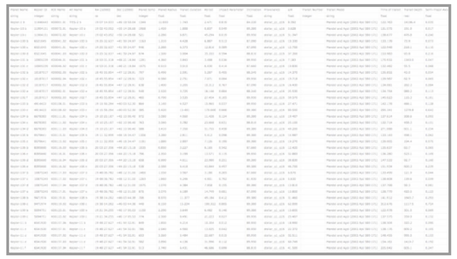
one of Corum's project process examples



source data



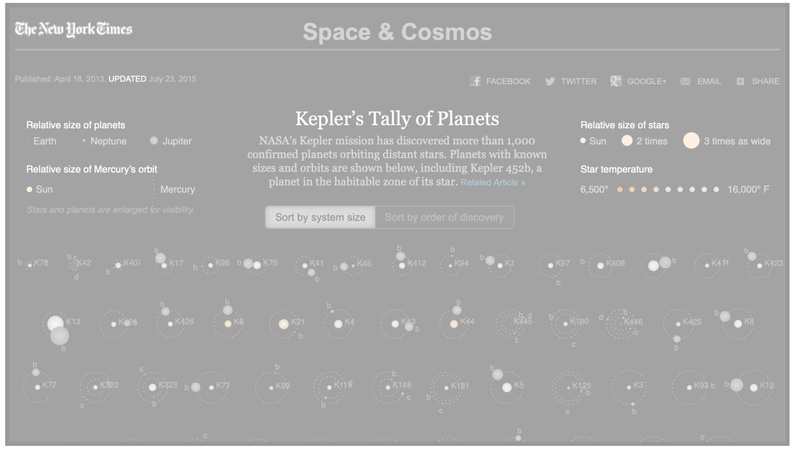
R markdown or iPython notebook



processed data

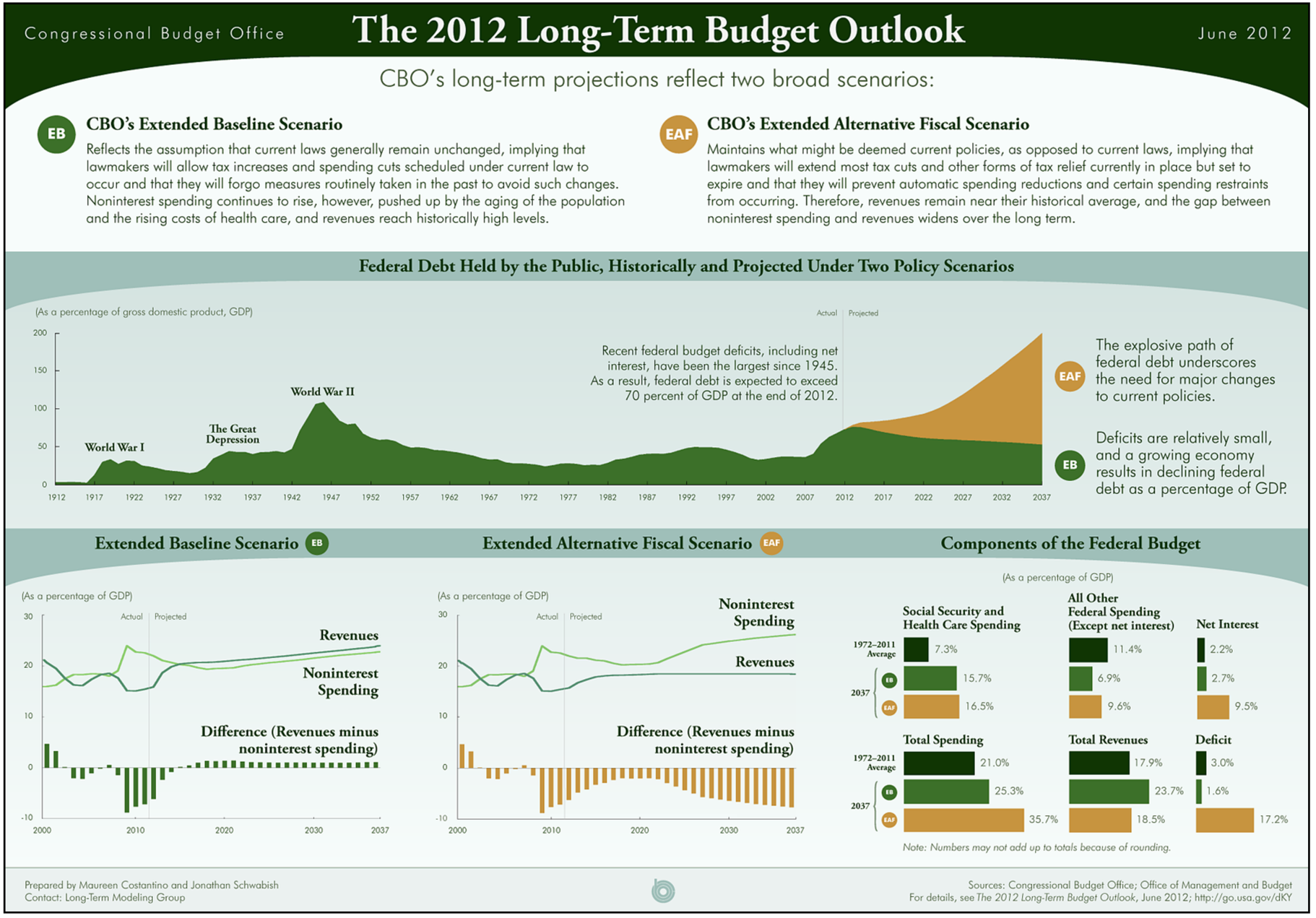
coffee script

d3



Hone ideas within limitations

Embrace limitations; use them to hone your ideas. Understand every step—leave nothing to magic—in your production. Design is cumulative decision making. Remember what it is like to not understand.


Congressional Budget Office
The 2012 Long-Term Budget Outlook
June 2012
CBO's long-term projections reflect two broad scenarios:
EB
CBO's Extended Baseline Scenario
Reflects the assumption that current laws generally remain unchanged, implying that lawmakers will allow tax increases and spending cuts scheduled under current law to occur and that they will forgo measures routinely taken in the past to avoid such changes. Noninterest spending continues to rise, however, pushed up by the aging of the population and the rising costs of health care, and revenues reach historically high levels.
EAF
CBO's Extended Alternative Fiscal Scenario
Maintains what might be deemed current policies, as opposed to current laws, implying that lawmakers will extend most tax cuts and other forms of tax relief currently in place but set to expire and that they will prevent automatic spending reductions and certain spending restraints from occurring. Therefore, revenues remain near their historical average, and the gap between noninterest spending and revenues widens over the long term.
Federal Debt Held by the Public, Historically and Projected Under Two Policy Scenarios
(As a percentage of gross domestic product, GDP)
Actual
Projected
Recent federal budget deficits, including net interest, have been the largest since 1945. As a result, federal debt is expected to exceed 70 percent of GDP at the end of 2012.
World War I
The Great Depression
World War II
EAF
The explosive path of federal debt underscores the need for major changes to current policies.
EB
Deficits are relatively small, and a growing economy results in declining federal debt as a percentage of GDP.
Extended Baseline Scenario EB
Extended Alternative Fiscal Scenario EAF
Components of the Federal Budget
(As a percentage of GDP)
Revenues
Noninterest Spending
Difference (Revenues minus noninterest spending)
Social Security and Health Care Spending
All Other Federal Spending (Except net interest)
Net Interest
1972–2011 Average
2037
7.3%
15.7%
16.5%
11.4%
6.9%
9.6%
2.2%
2.7%
9.5%
Total Spending
Total Revenues
Deficit
21.0%
25.3%
35.7%
17.9%
23.7%
18.5%
3.0%
1.6%
17.2%
Note: Numbers may not add up to totals because of rounding.
Prepared by Maureen Costantino and Jonathan Schwabish
Contact: Long-Term Modeling Group
Sources: Congressional Budget Office; Office of Management and Budget
For details, see The 2012 Long-Term Budget Outlook, June 2012; http://go.usa.gov/dKY