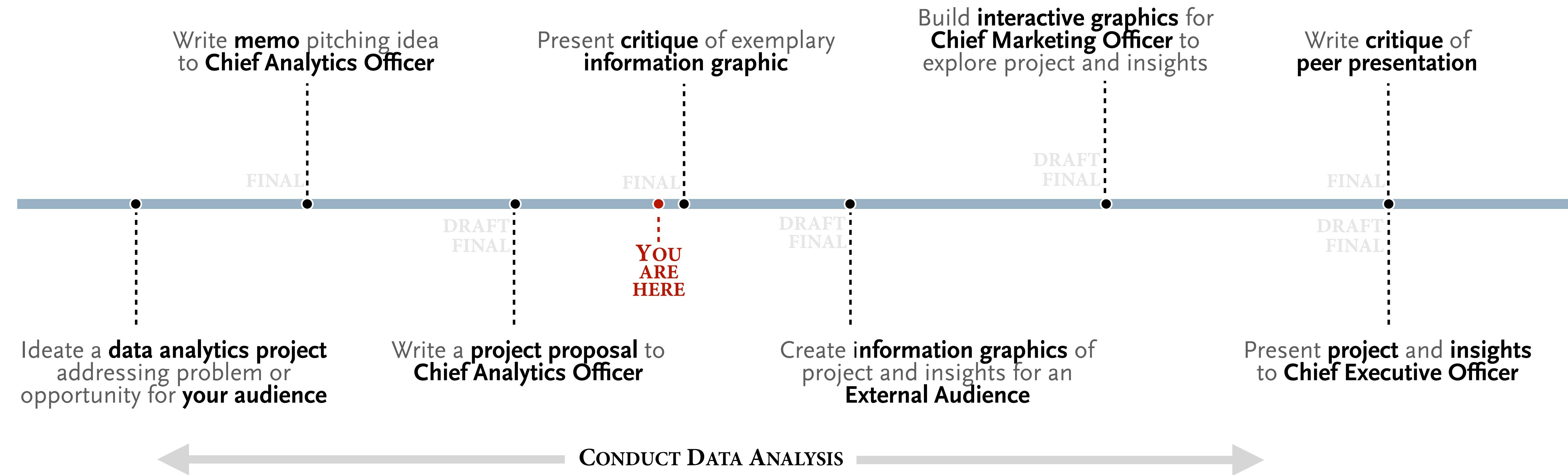


Storytelling with data

**08 | Design mini-review; critiquing data-driven, visual narratives;
encoding uncertainty, estimates, forecasts; pacing for attention**

course overview | main course deliverables

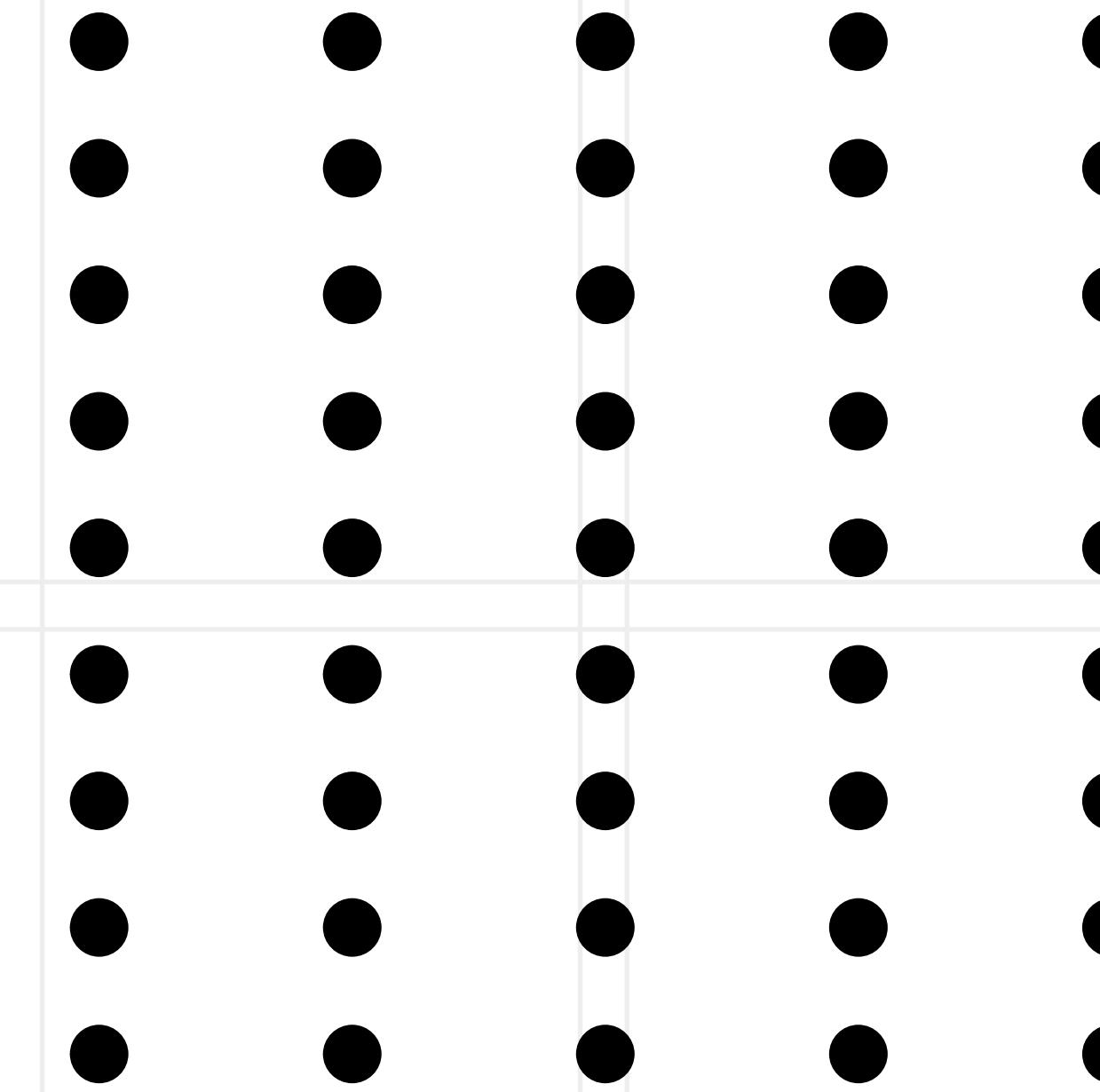
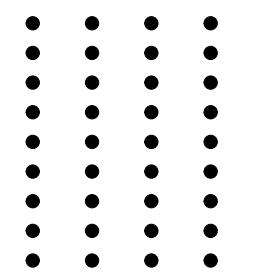


design mini-review

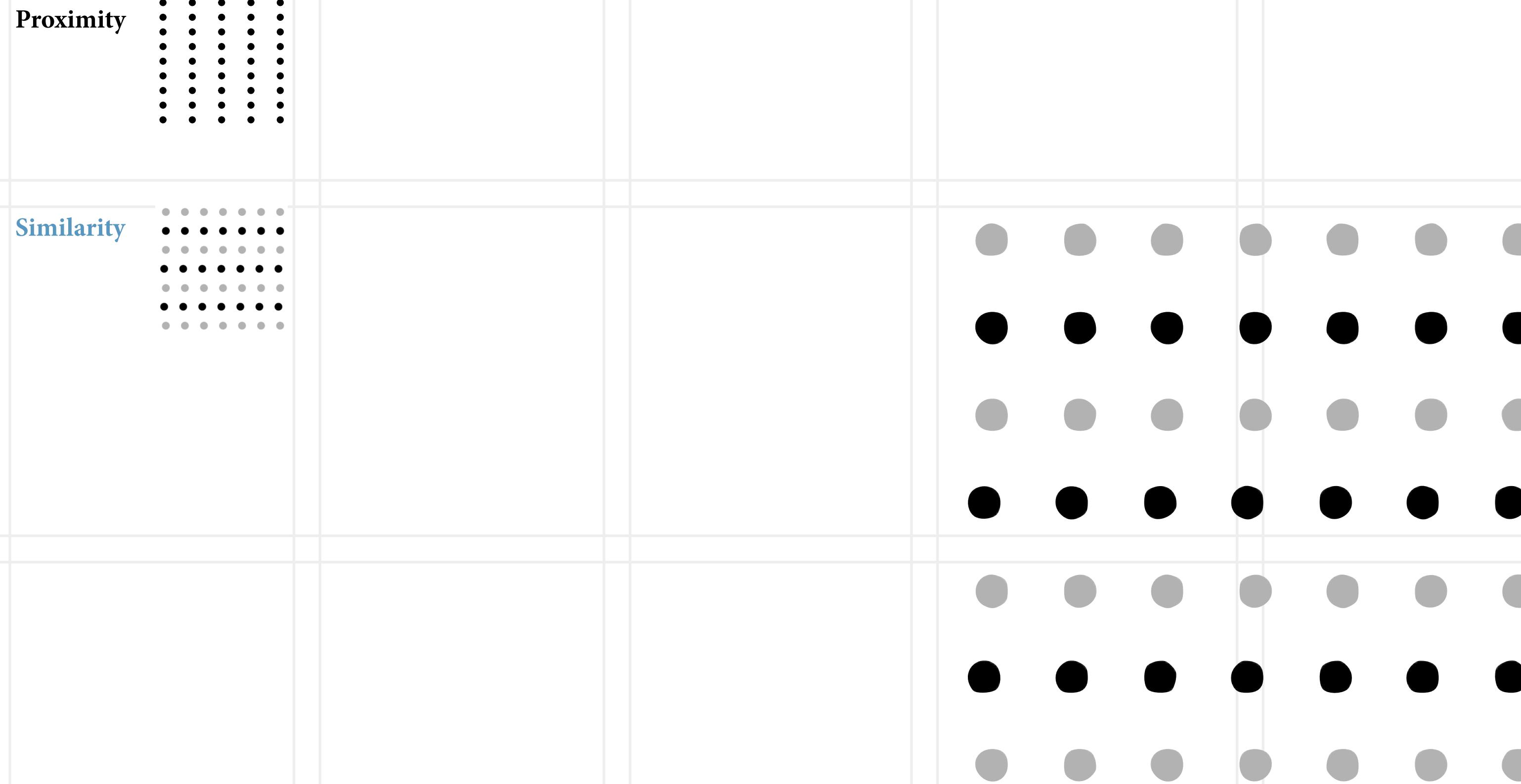
design mini-review | aligning and organizing information reduces cognitive load — *grids*

design mini-review | aligning and organizing information reduces cognitive load — *proximity*

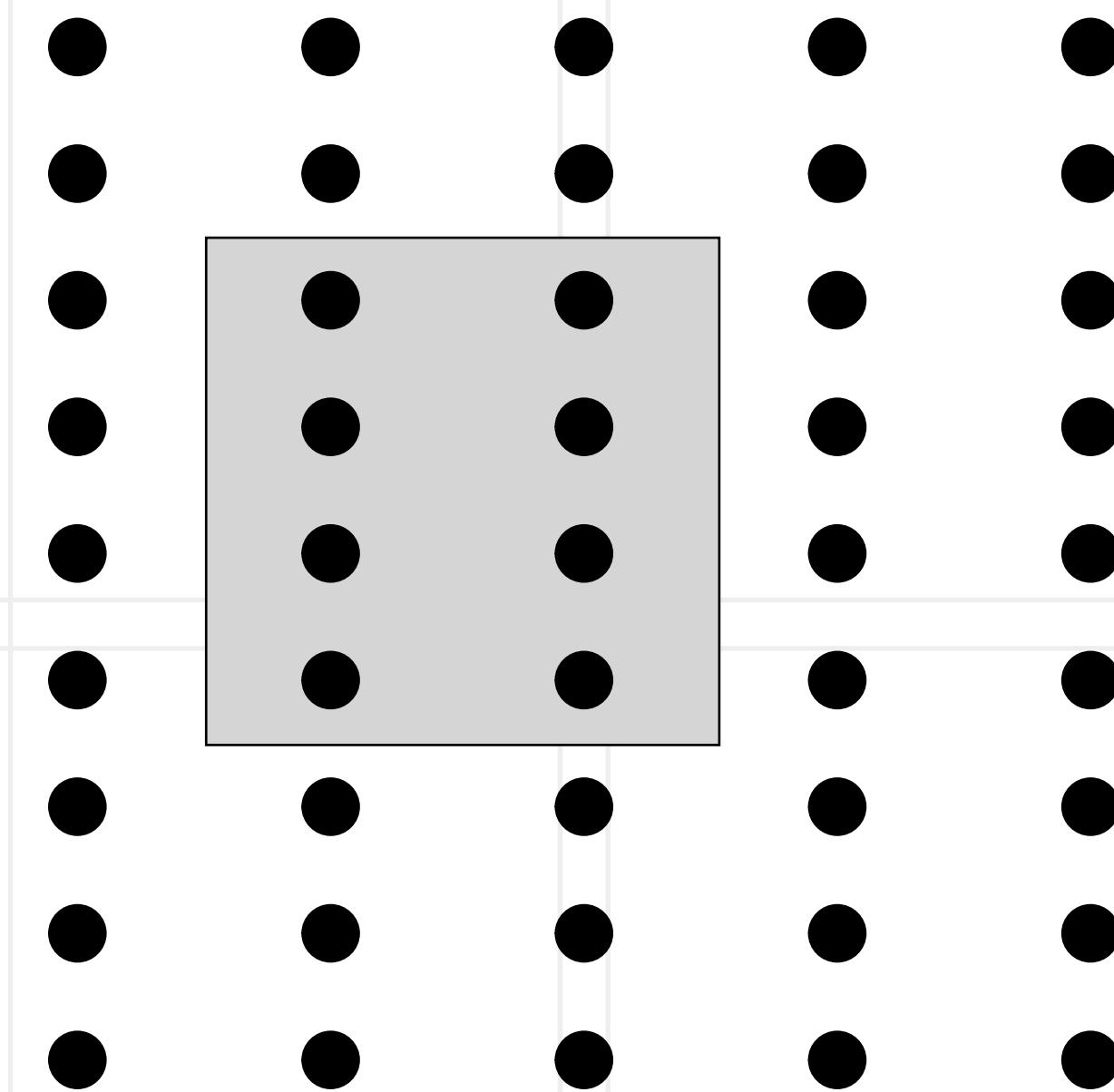
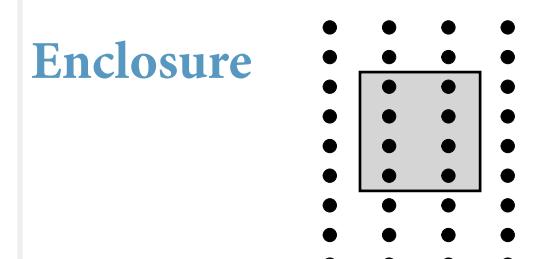
Proximity



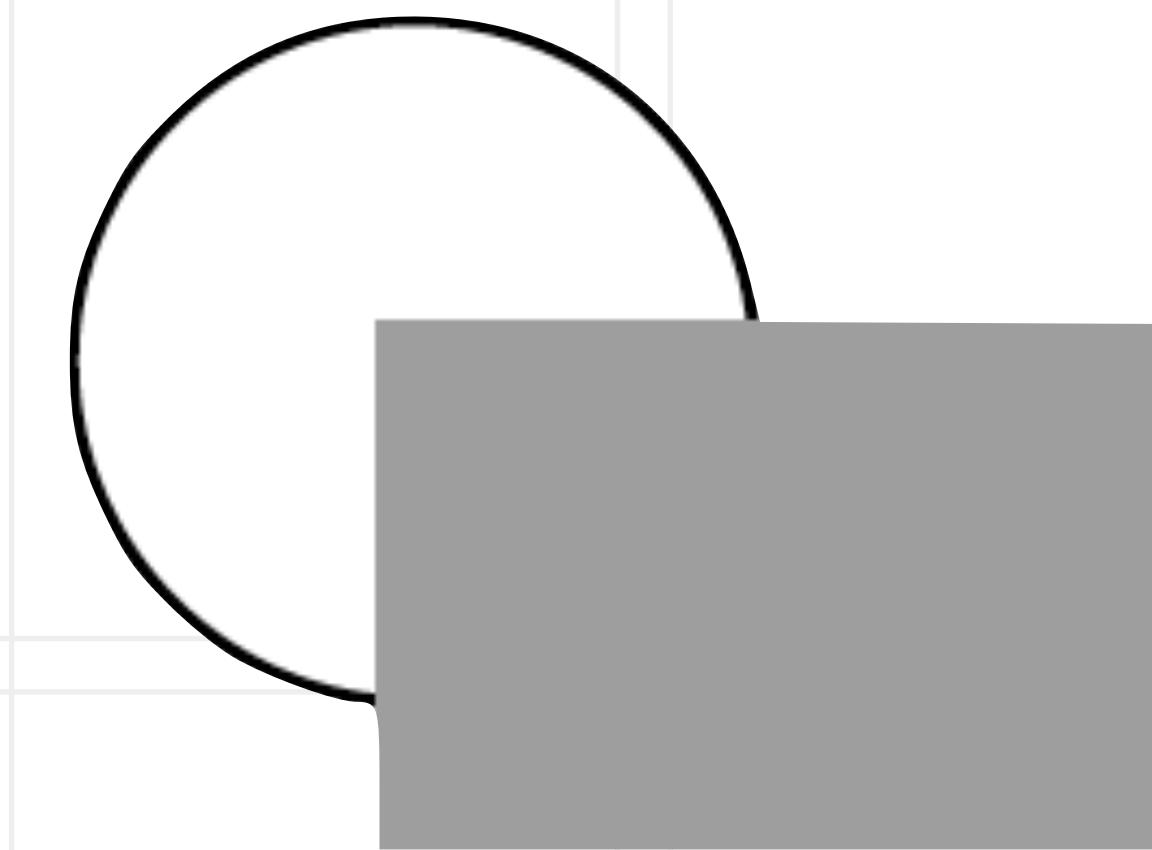
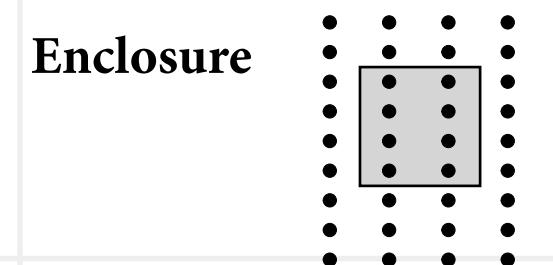
design mini-review | aligning and organizing information reduces cognitive load — *similarity*



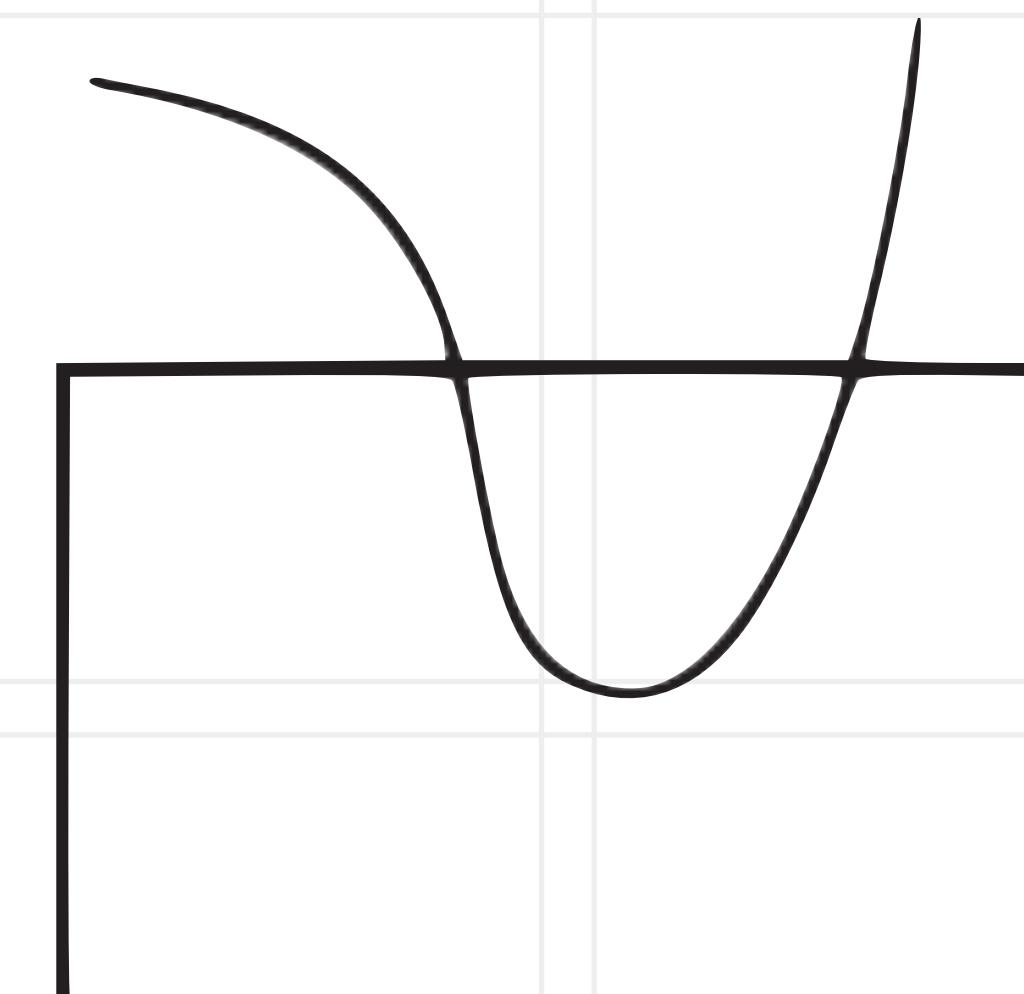
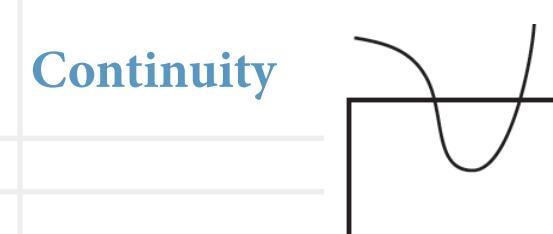
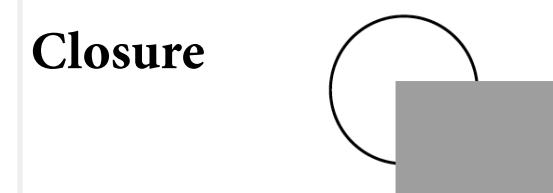
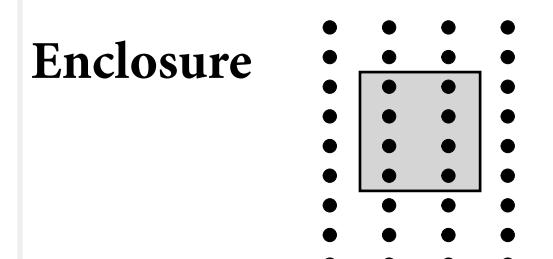
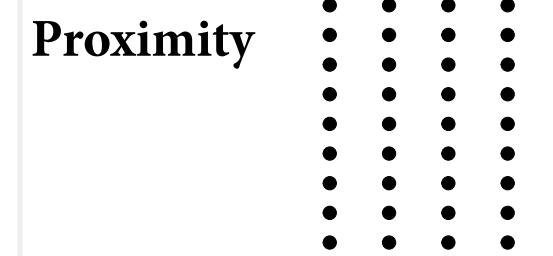
design mini-review | aligning and organizing information reduces cognitive load — *enclosure*



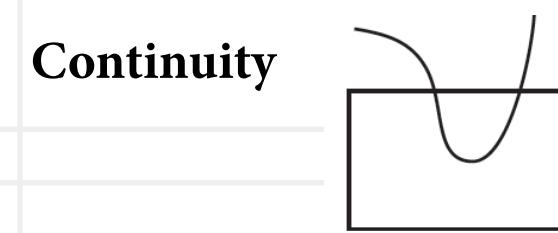
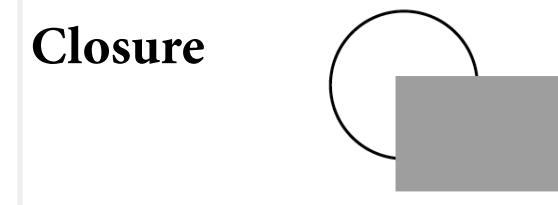
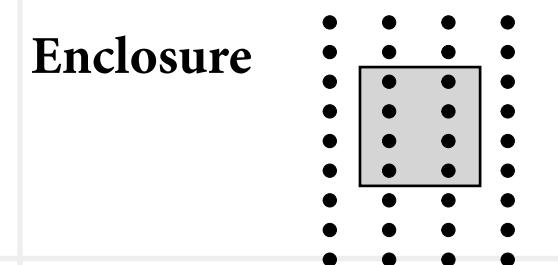
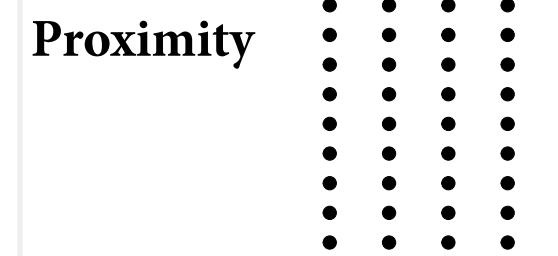
design mini-review | aligning and organizing information reduces cognitive load — *closure*



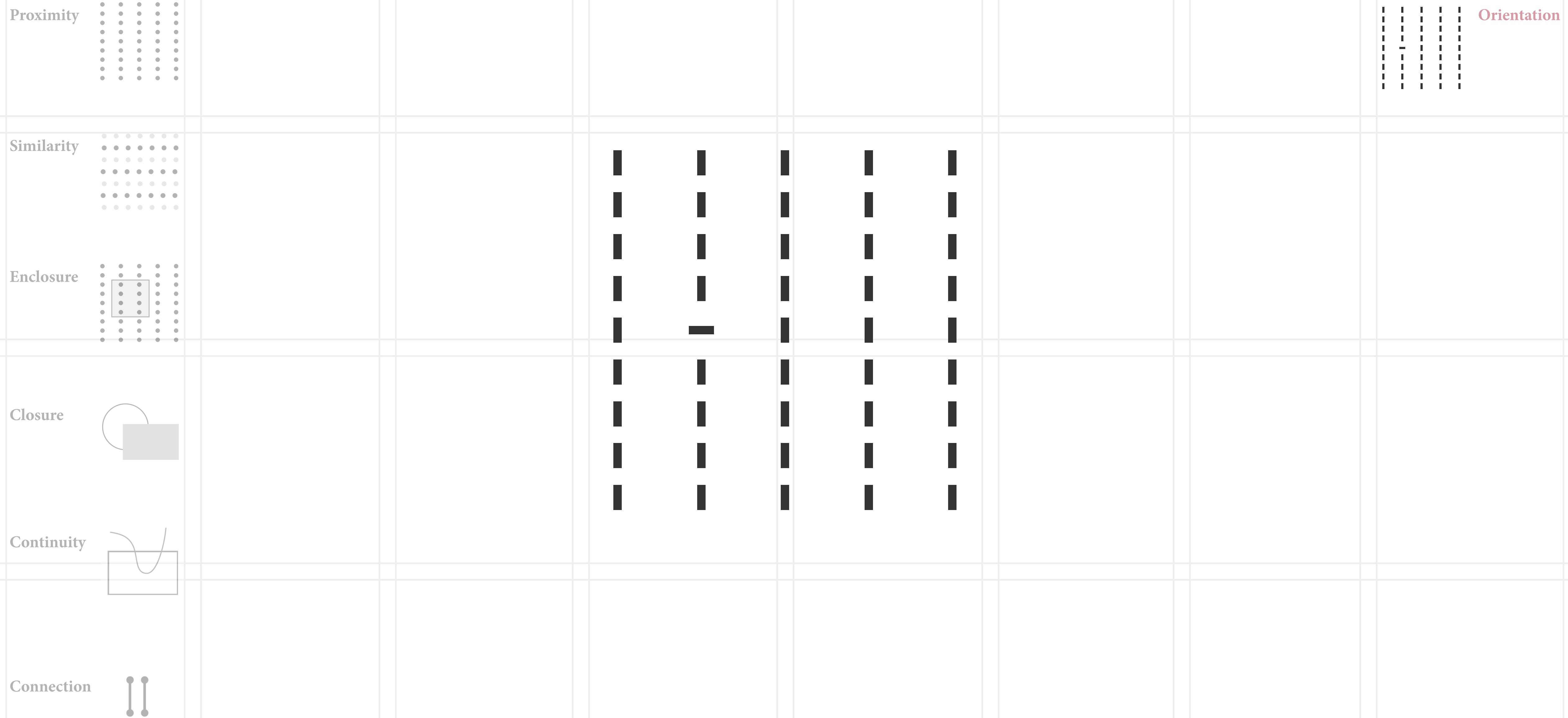
design mini-review | aligning and organizing information reduces cognitive load — *continuity*



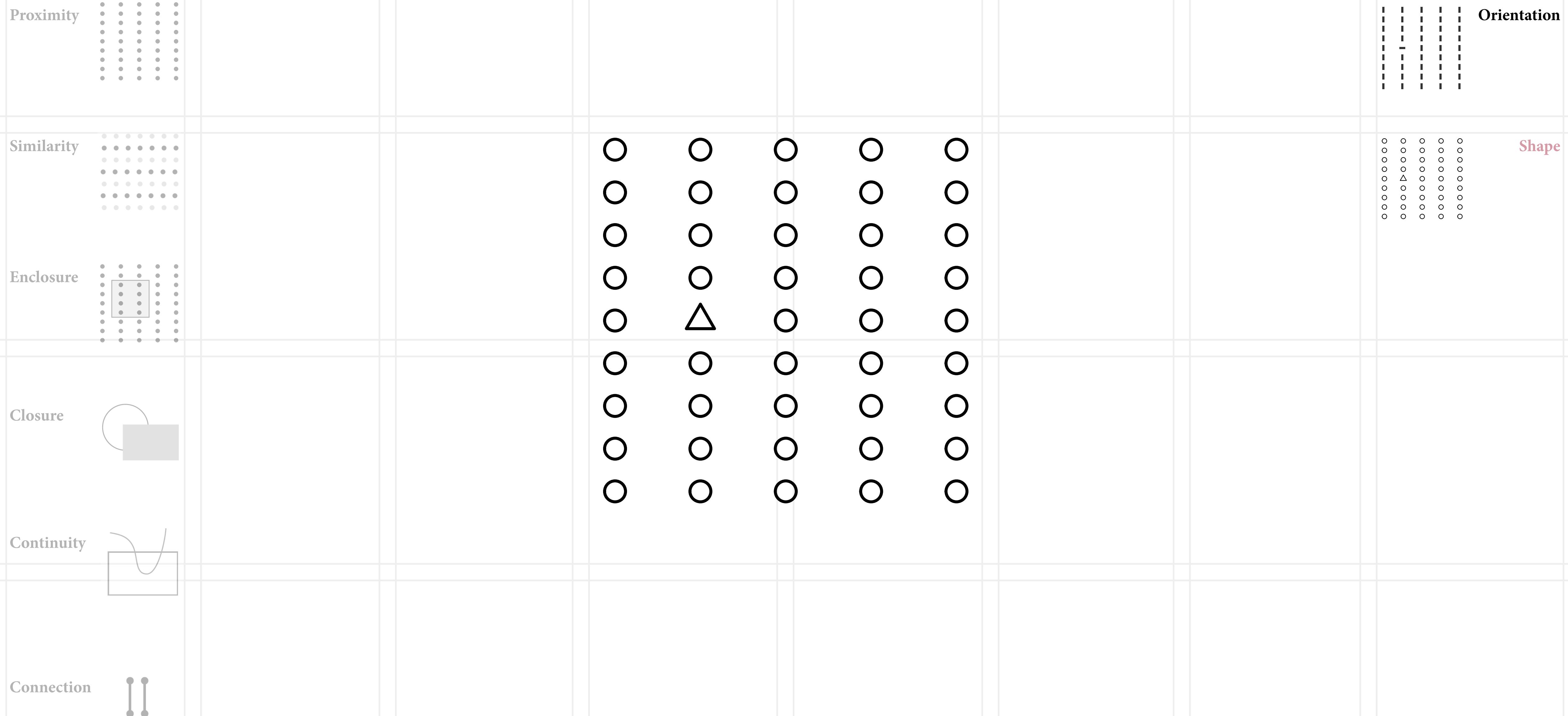
design mini-review | aligning and organizing information reduces cognitive load — *connection*



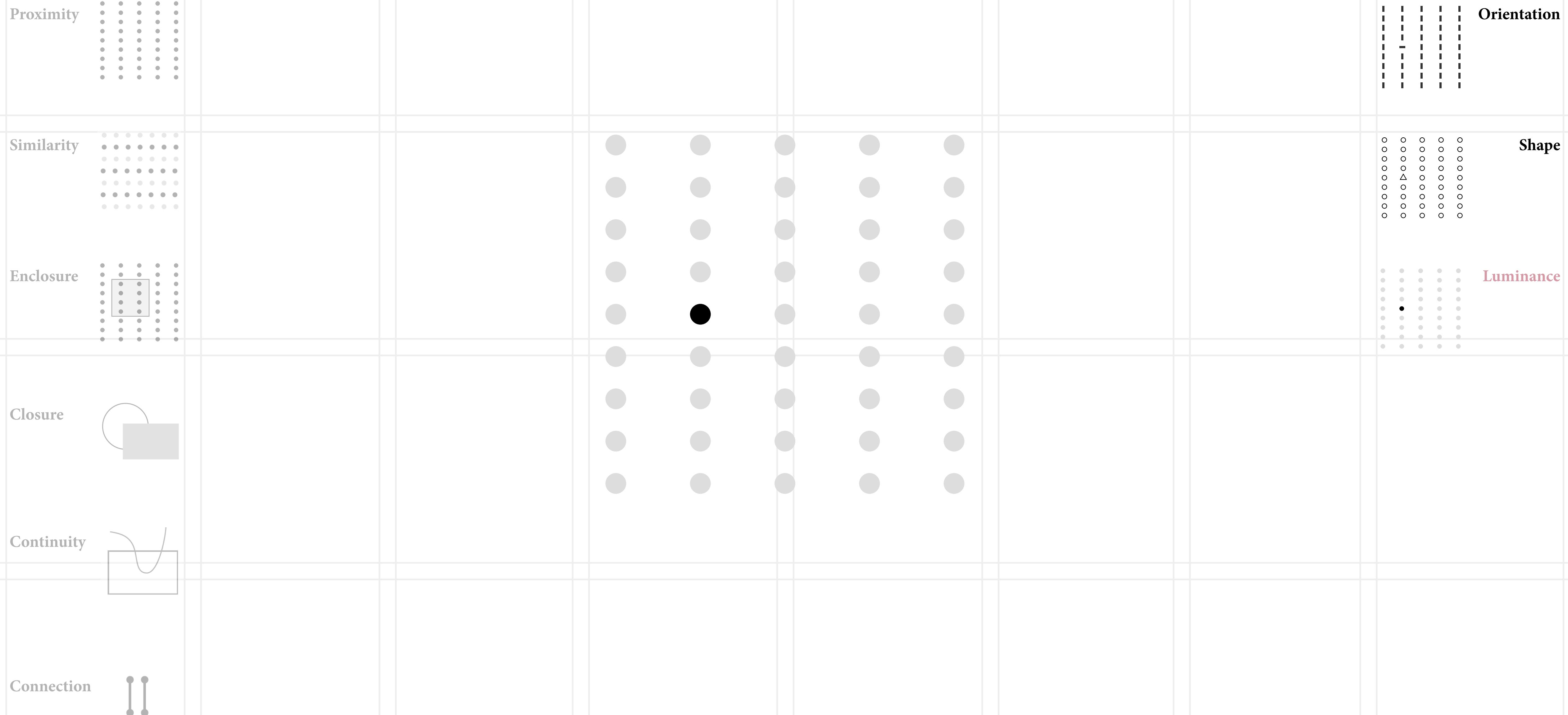
design mini-review | purposeful change of a visual channel can focus attention — *orientation*



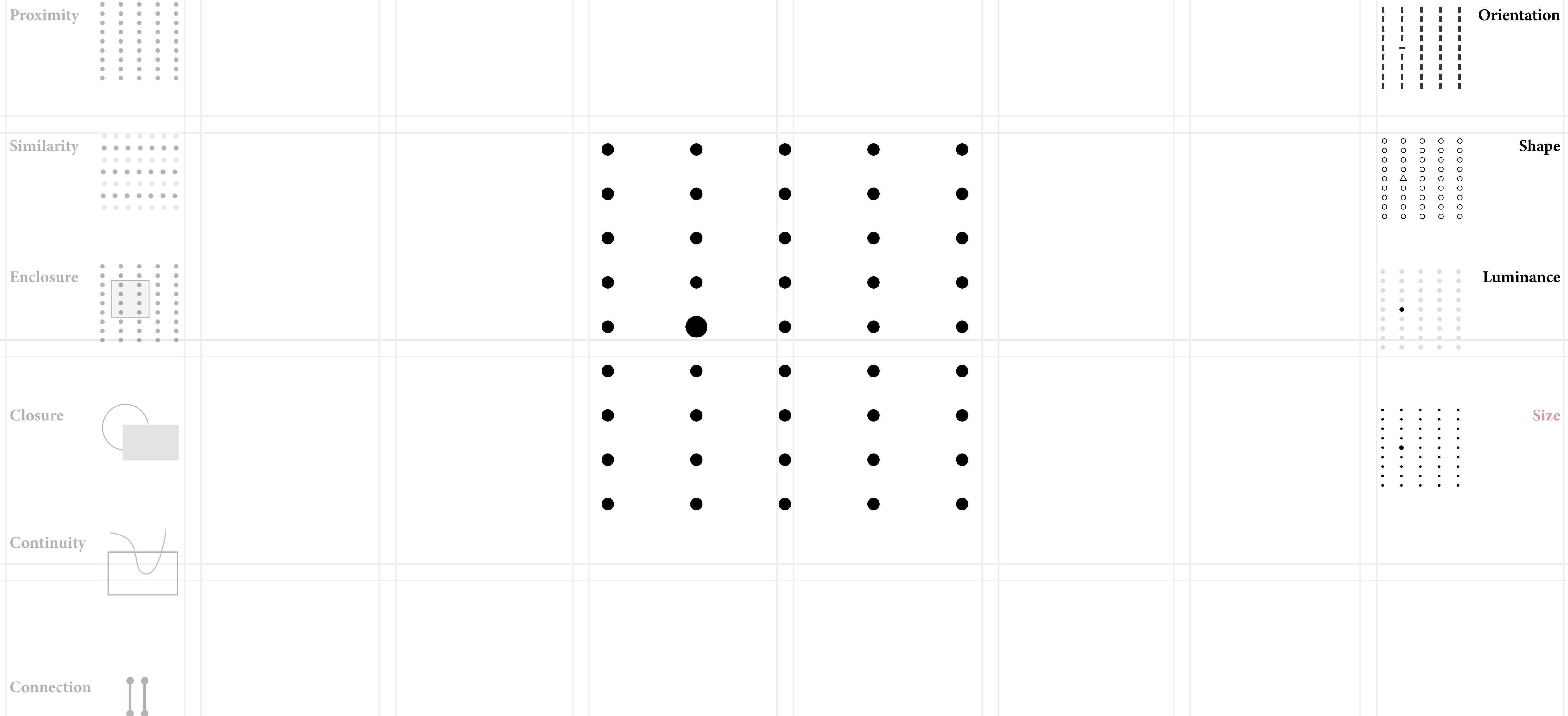
design mini-review | purposeful change of a visual channel can focus attention — *shape*



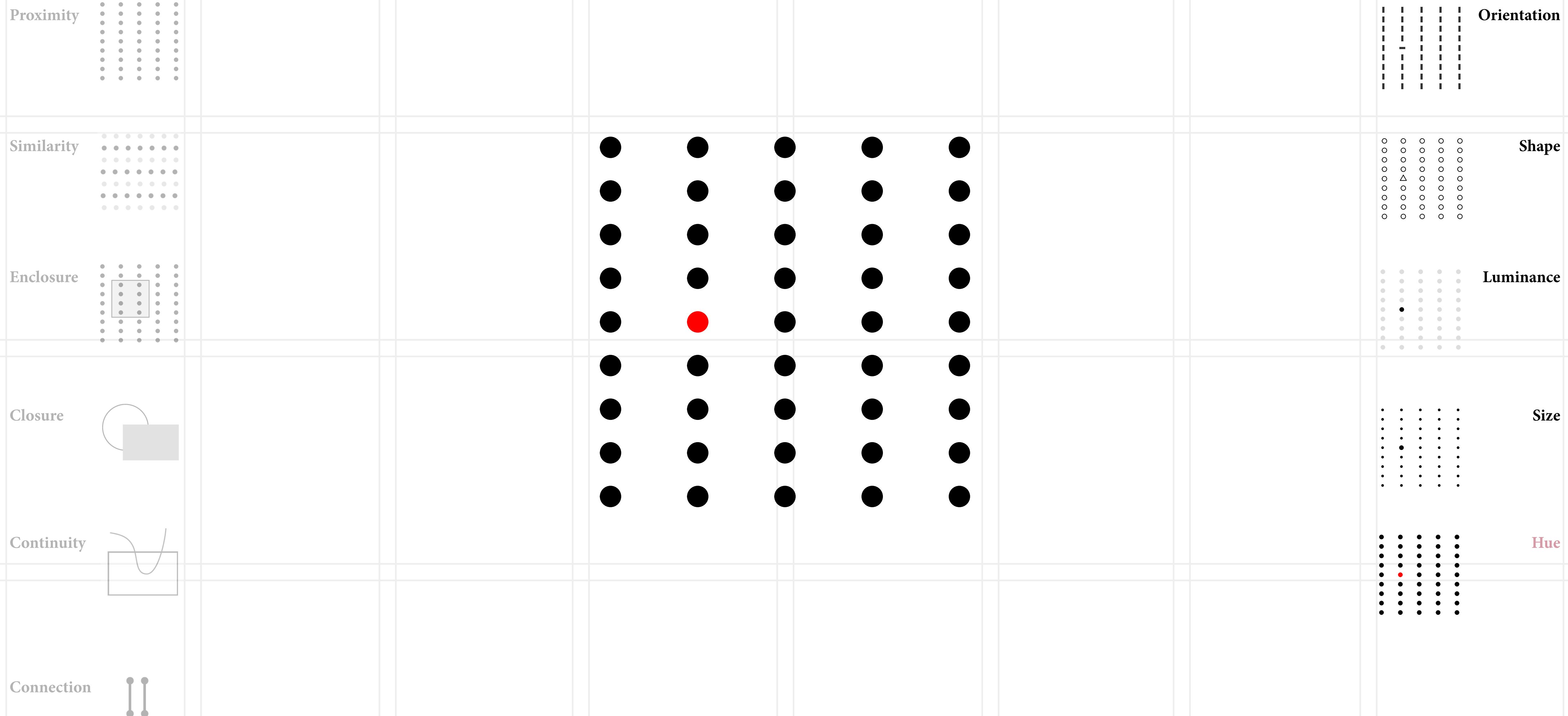
design mini-review | purposeful change of a visual channel can focus attention — *luminance*



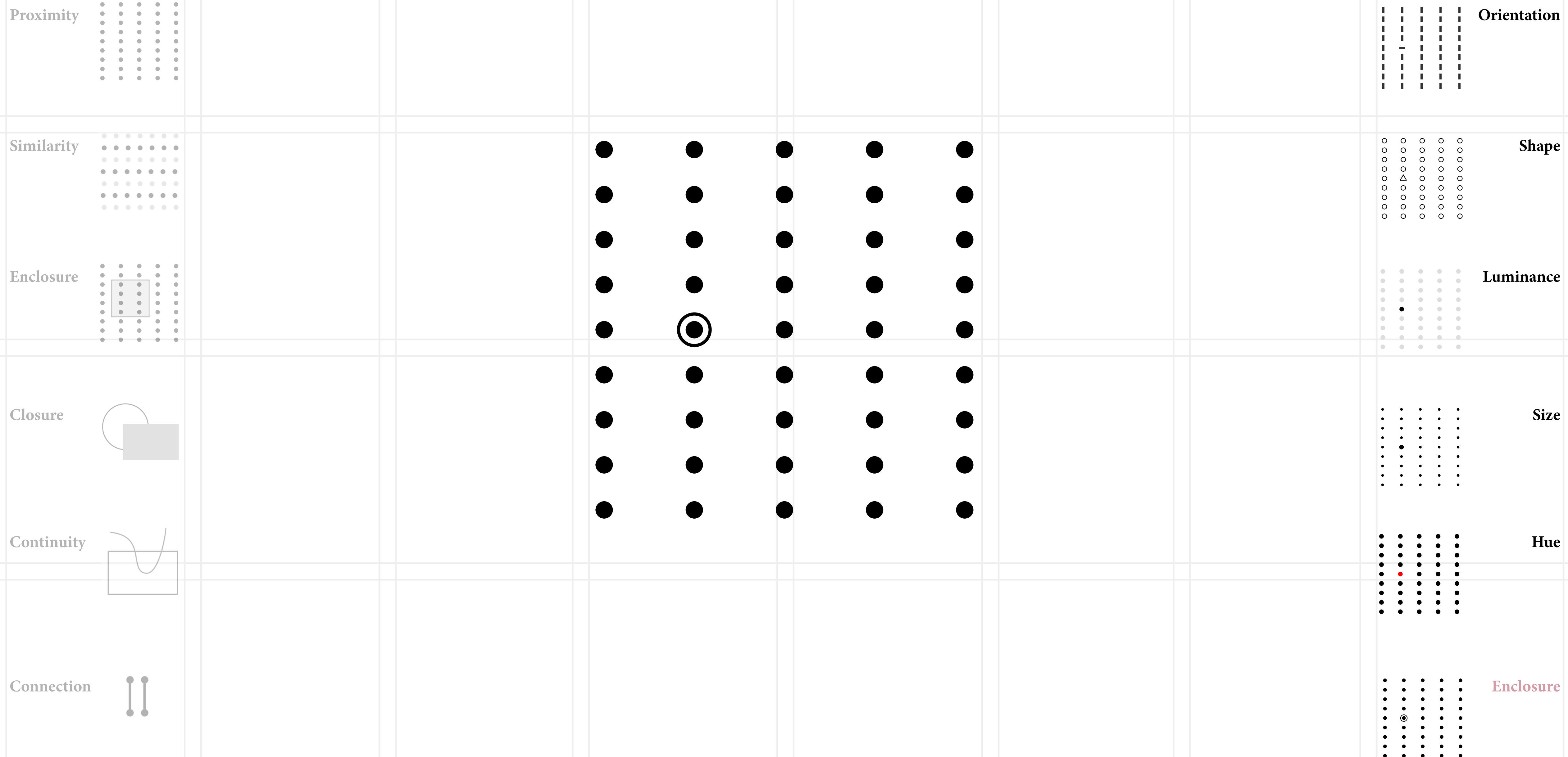
design mini-review | purposeful change of a visual channel can focus attention — *size*



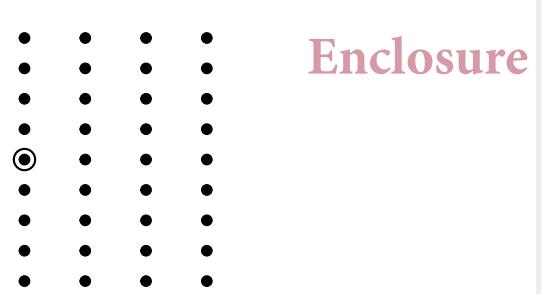
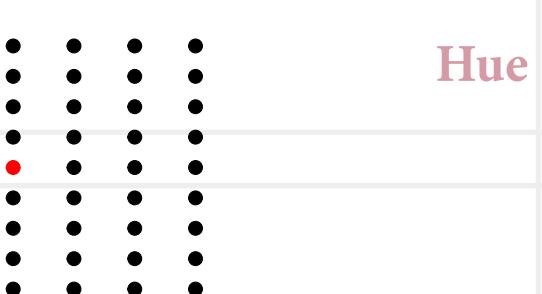
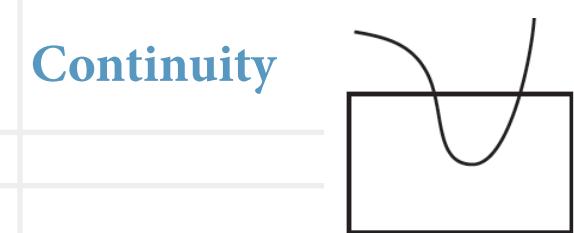
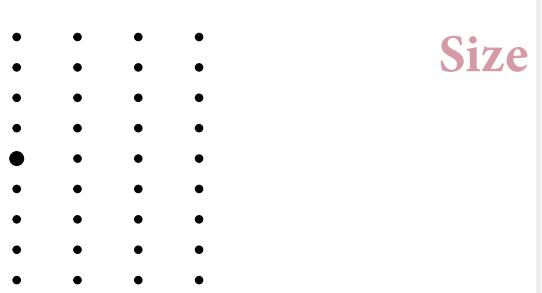
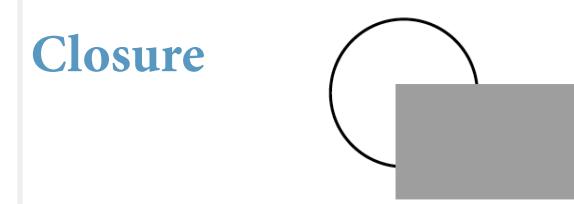
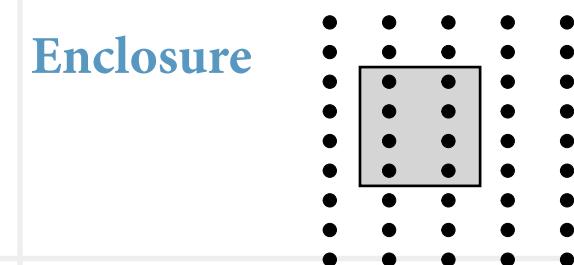
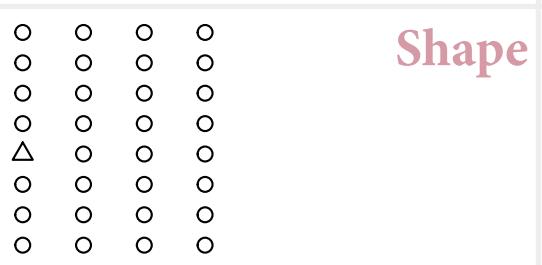
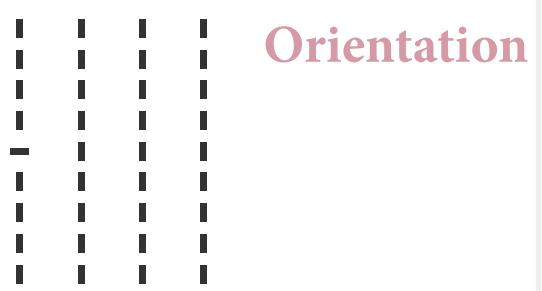
design mini-review | purposeful change of a visual channel can focus attention — *hue*



design mini-review | purposeful change of a visual channel can focus attention — *enclosure*

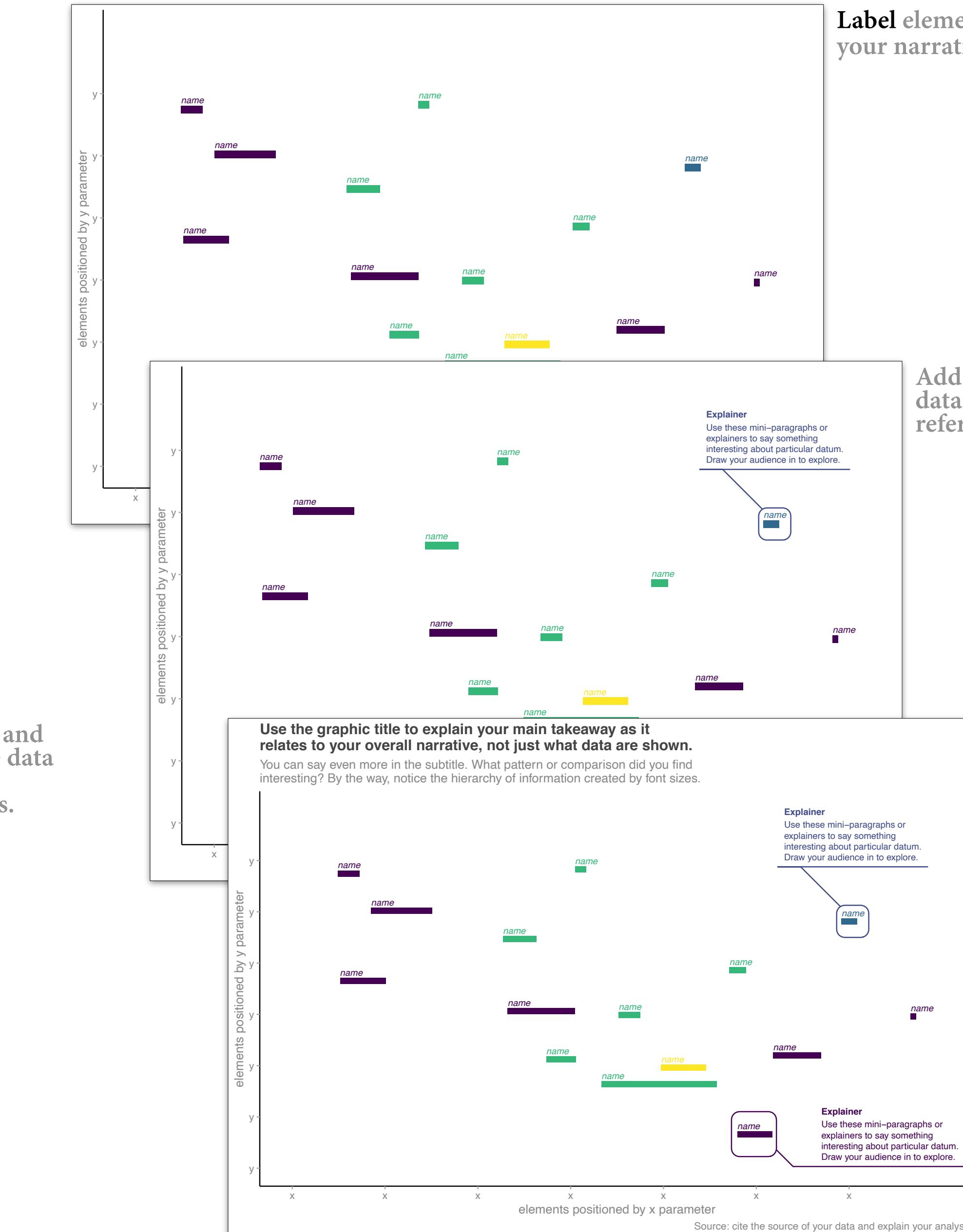
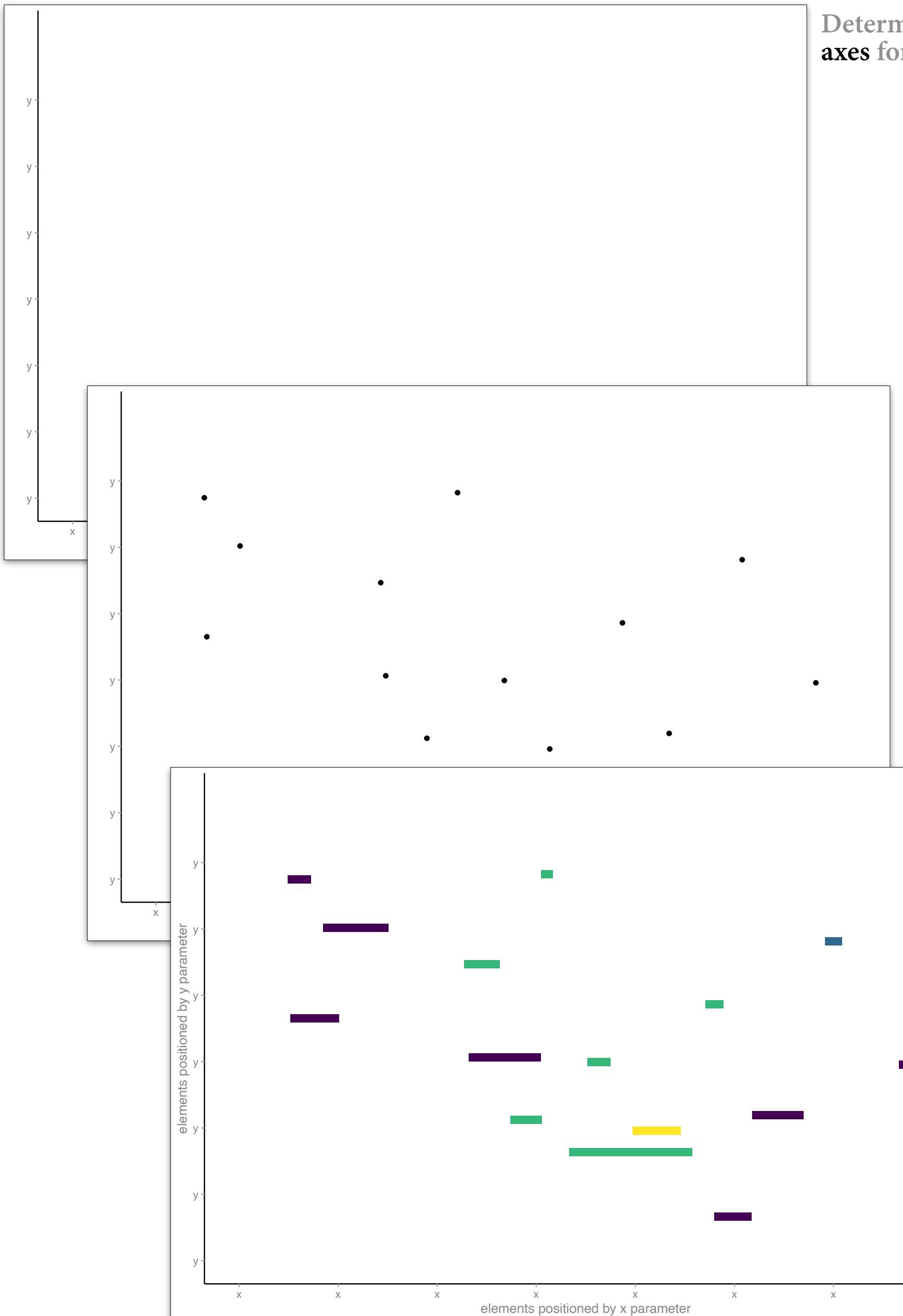


design mini-review

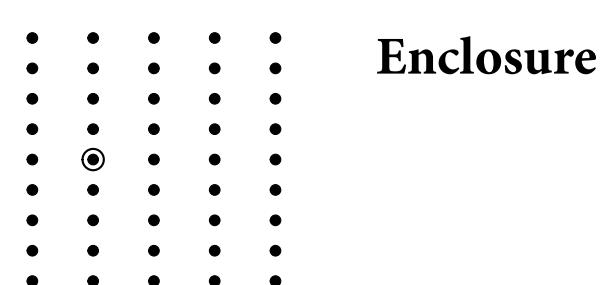
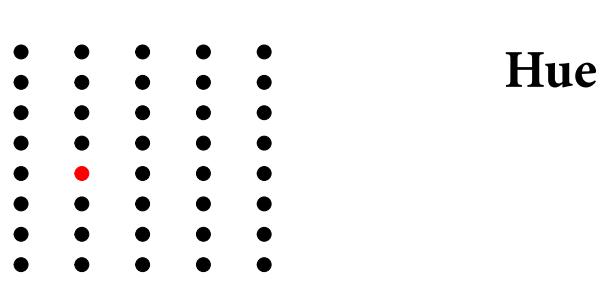
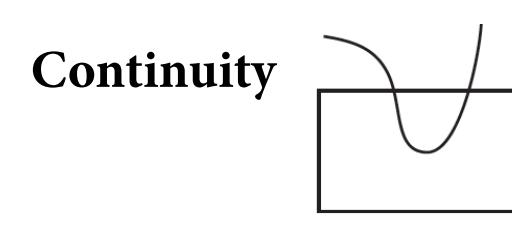
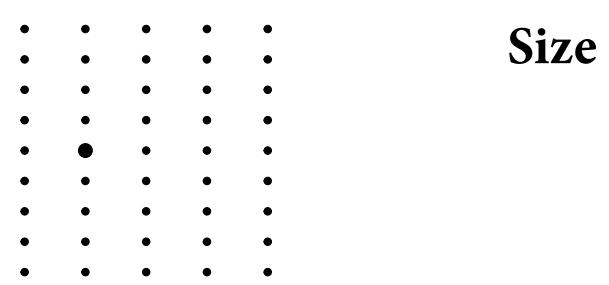
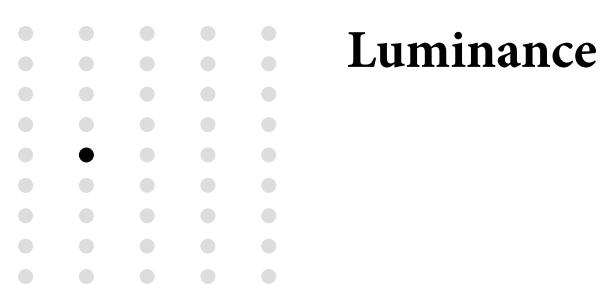
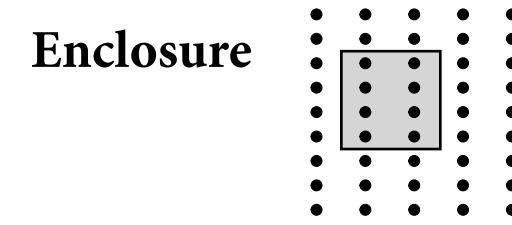
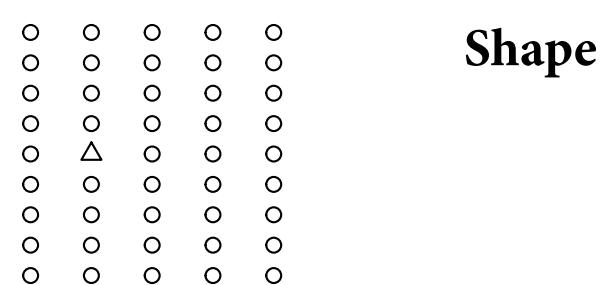
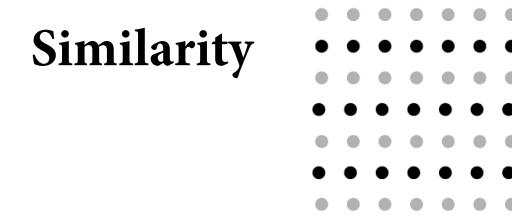
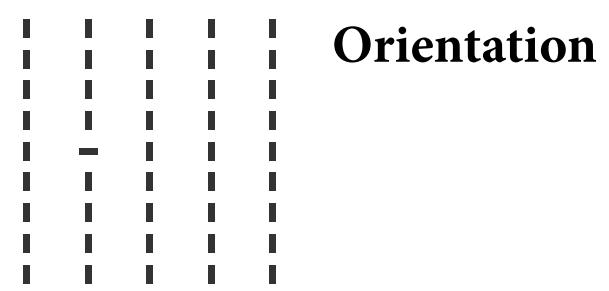
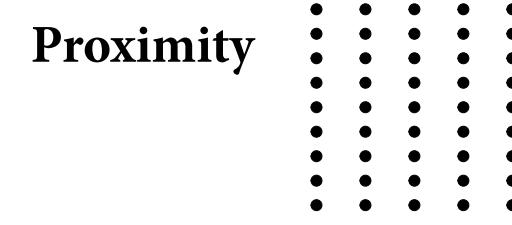


grids, **Gestalt principles**, and
preattentive attributes may be combined

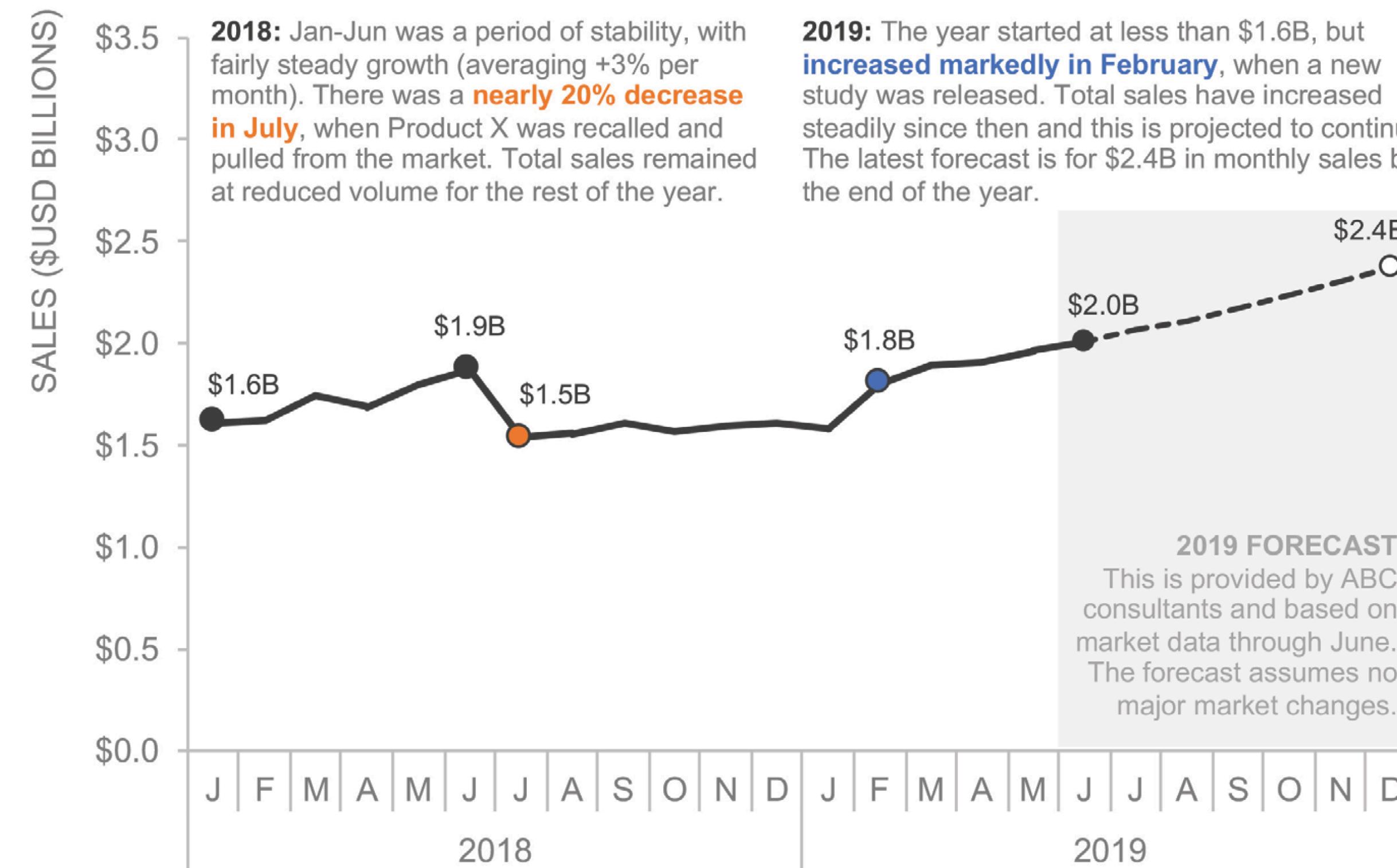
design mini-review | layer the graphic, encode visual channels, annotate, and make hierarchies clear



design mini-review | what Gestalt principles are used in this data graphic? How is attention focused?

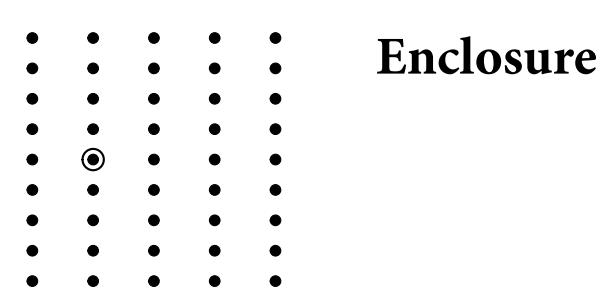
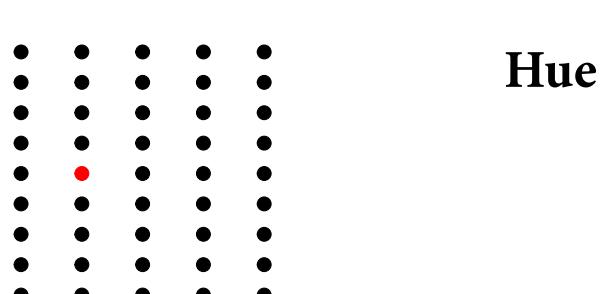
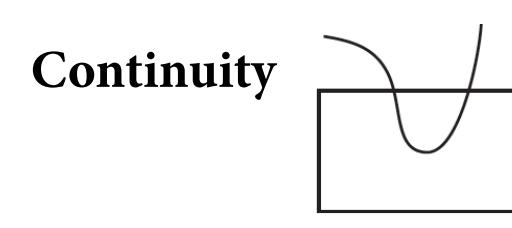
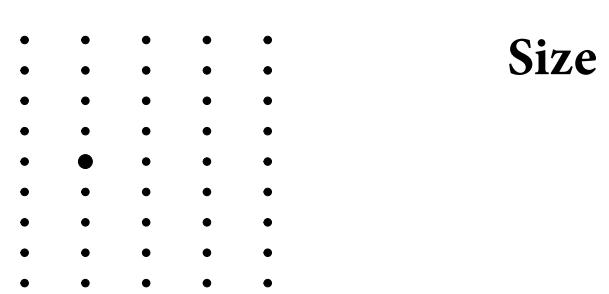
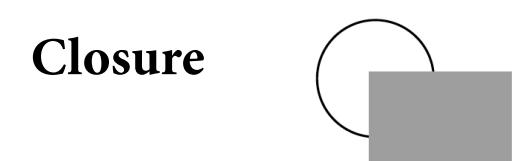
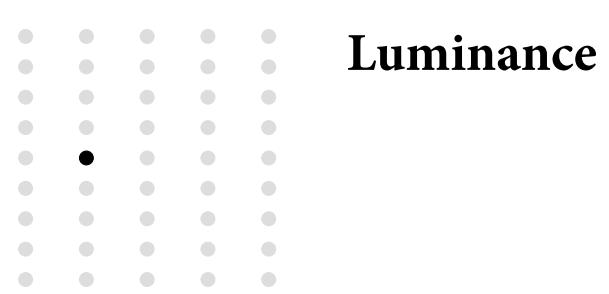
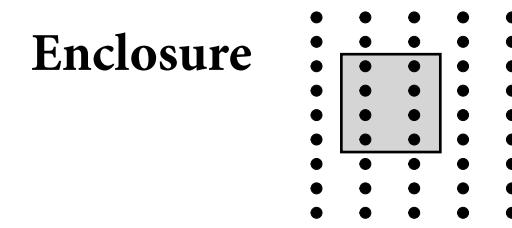
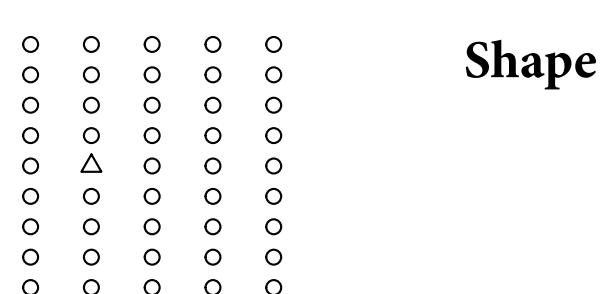
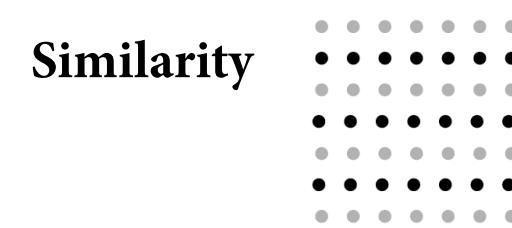
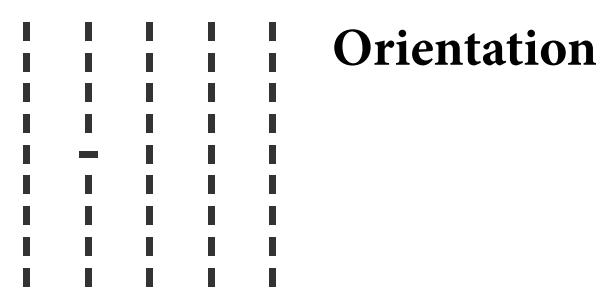
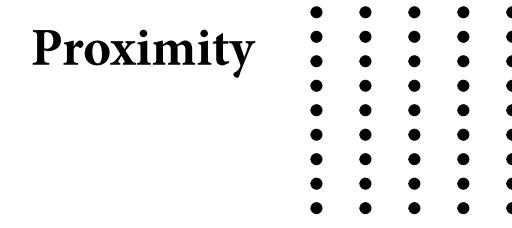


Market size over time



Example from: Knaflic, Cole Nussbaumer. *Storytelling with Data: Let's Practice!* Hoboken, New Jersey: John Wiley & Sons, Inc, 2019.

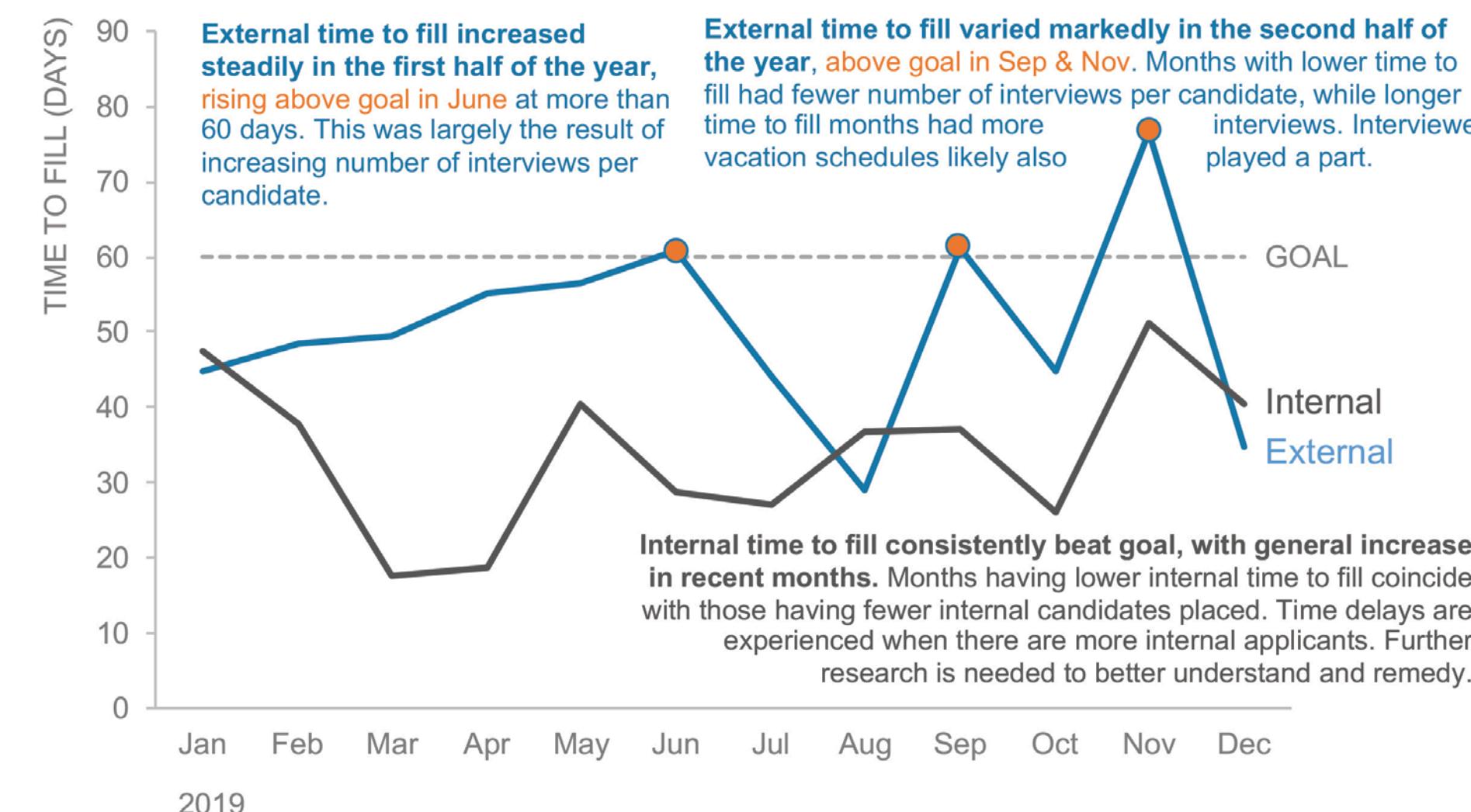
design mini-review | what Gestalt principles are used in this data graphic? How is attention focused?



Time to fill role discussion needed: where do we go from here?

Both **External** and **Internal** time to fill have varied in the past year. Understanding contributing factors—number of interviews, vacation schedules, and current internal transfer volume constraints—can help us better plan for the future.

Time to fill



LET'S DISCUSS: Should we put stricter guidelines around maximum number of interviews? How can we keep vacation schedules from impacting time to hire? What can we do to improve efficiency of internal transfer process in order to better handle higher volumes?

Example from: Knaflic, Cole Nussbaumer. *Storytelling with Data: Let's Practice!* Hoboken, New Jersey: John Wiley & Sons, Inc, 2019.

design mini-review | what Gestalt principles are used in this data graphic? How is attention focused?

Proximity

Orientation

Similarity

Shape

Enclosure

Luminance

Closure

Size

Continuity

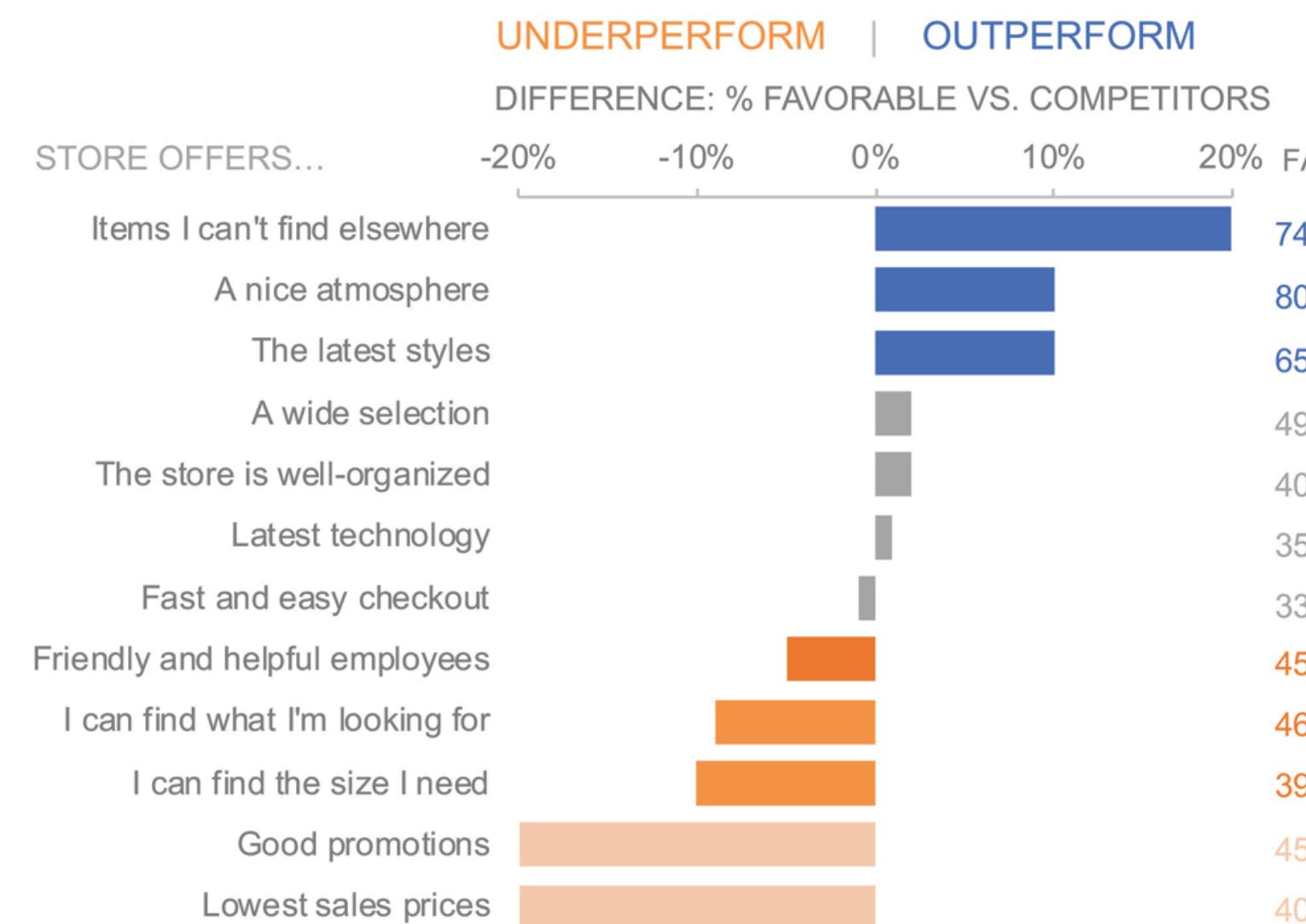
Hue

Connection

Enclosure

Action needed: invest in employee training

Back-to-school shopping: consumer sentiment



THE GOOD NEWS:
We're beating the competition when it comes to **the latest styles that people can't find elsewhere and store atmosphere.**

WE CAN IMPROVE:
We score low and lower than the competition in areas related to **helpful employees and customers being able to find what they are looking for.** We also score lower than the competition on promotions/sales, but don't recommend focusing here.

RECOMMENDATION:
Invest in employee training to improve customer experience.

Data Source: 2019 Back-to-School shopping survey (represents 21,862 survey responses).
Additional survey and methodology details available upon request. Reach out to Insights Team.

Example from: Knaflic, Cole Nussbaumer. *Storytelling with Data: Let's Practice!* Hoboken, New Jersey: John Wiley & Sons, Inc, 2019.

**a framework for critiquing
data-driven, visual narratives**

criticism for visuals, information graphics — *our* working definition

information graphic : *a data-driven, visual narrative*

criticism for data-driven, visual narratives, visualization criticism is critical thinking about data visualization

Establish the purpose of the critique

When reviewing someone else's document, center yourself on the **purpose that was agreed upon**, such as clarity, accuracy, or correctness. Should this purpose be multiple, **review one aspect at a time, focusing on content first**.

Offer alternative solutions

In your comments—help, don't judge. A critique must serve the goal. Simply pointing to problems is not enough. The critic must **state an alternative solution** in a way that is clear and complete enough to provide a basis for improvement.

Be objective, well-reasoned

Typos are usually more conspicuous than reasoning flaws, but also less important. Each statement should be **objective**, delivered in **neutral language**, and backed up by **theoretical reasoning or empirical evidence**.

Structure the review

First, provide a **global assessment**, to place further comments in proper perspective. As a rule, point out the **weaknesses**, to prompt improvements, but also the **strengths**, to increase the authors' willingness to revise the document and to learn.

criticism for data-driven visual narratives, using theory and experiment, identify issues *and* suggest solutions

Get Specific

Audience? | Does the information graphic seem designed to communicate with an *identified* or *particular* audience? If so, who?

Purpose? | Do you see a purpose? If so, is it trying to inform, entertain, or *persuade the audience to act*? Something else?

Encoding, decoding? | What data are encoded? How? Any issues of perception in decoding? Most important measures encoded with most accurately decoded *visual channels* and their *attributes*?

Comparison or change? | Does the information graphic show *comparisons* or *change*? Would other *context* help with *meaning*?

Narrative? | Does it use *messages*, stated first, within a narrative? If so, what structure? An *arc*? With *examples*? *Metaphors*?

Color, coherency? | Is color used? If so, for what purposes are its hue, chroma, or luminance used? How might other uses help?

Hierarchy, annotation? | Does it layer information as a hierarchy? If so, how does that hierarchy separate information? Are data encodings explained? If so, how?

Layering, layout? | How is the information organized? Can a grid, negative space, or Gestalt principles — *proximity, similarity, enclosure, closure, continuity, connection* — help simplify or focus attention?

Credibility, transparency? | Are data sources identified, explained? Limitations, issues, exceptions discussed?

Learning to see — let's critique

criticism for visuals, example — a very basic critique of Scarr's *Hazy days*.

Audience?

Published in a newspaper. An external, general audience. Primary audience are the population of readers of the South China Morning Post.

Purpose?

To inform or raise awareness through exploration. No explicit call-to-action of its audience.

Data encodings, decodings?

Scarr uses multiple visual channels. A heat map encodes each cell as an hour of the year. Day is aligned on the y-axis, hour is aligned on the x-axis, enabling comparisons across either. Luminosity represents air-pollution index. Wind direction is encoded alongside each day by orienting a line segment with an arrow end. But wind direction continually shifts. And data on each hour likely exists. Perhaps we could experiment with placing an oriented line segment inside each square as a layer, creating a vector field?

Comparison or change?

Encodings are arranged to allow overall comparisons of pollution by day or hour, and the graphic points to a few specific, interesting patterns.

Narrative?

No narrative is developed. Perhaps placing this information graphic into a historical context of pollution in Hong Kong, or in the context of people's lives, would help develop a narrative.

Color, coherency?

Only shades of gray encode data, and especially for encoding pollution on a heat map. Notice the gray palette also serves as a visual metaphor as we think of pollution as physically graying our otherwise blue skies. Would a blue-to-gray encoding strengthen the metaphor?

Hierarchy, layering, layout?

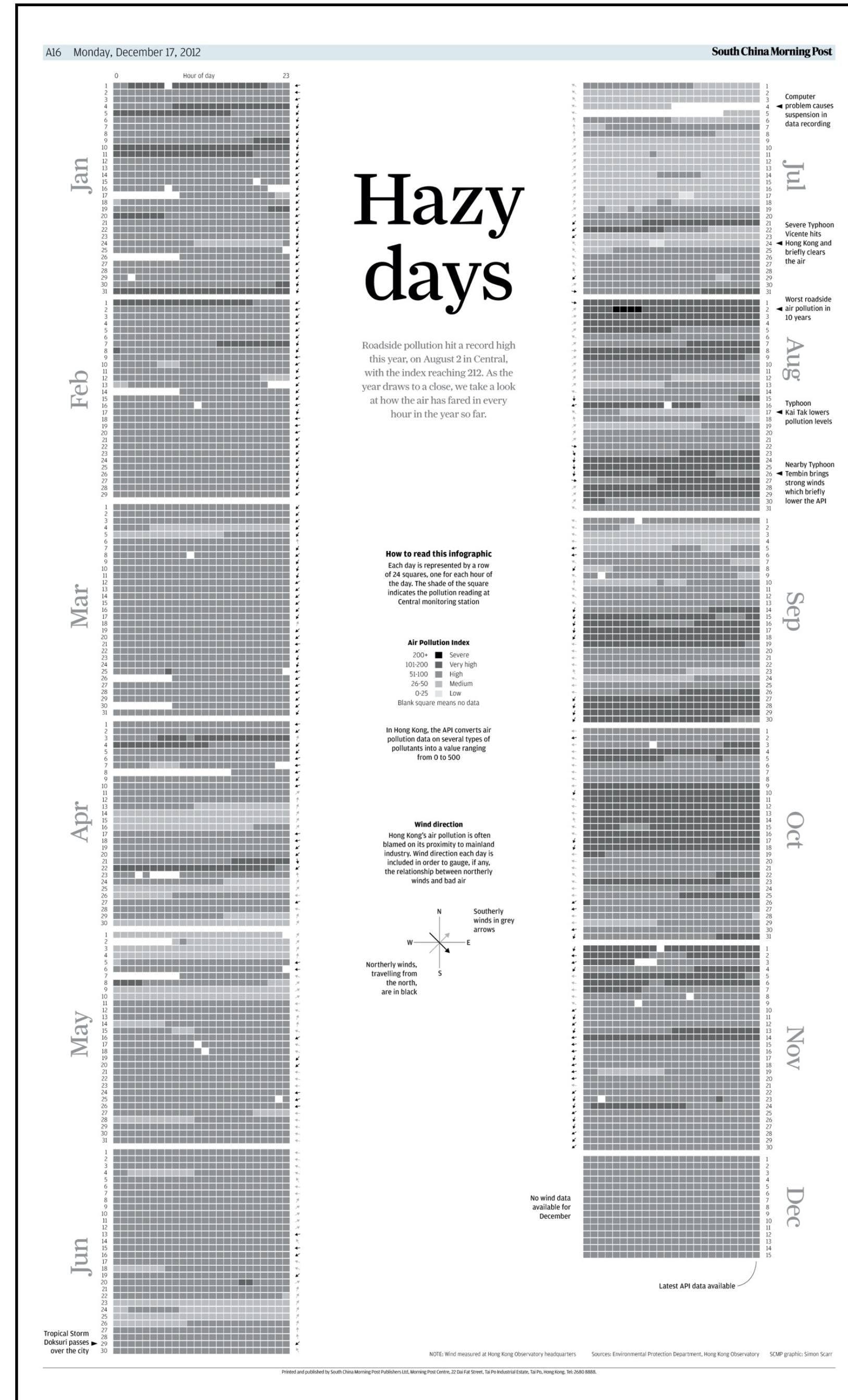
Scarr uses typography effectively — especially font sizing, bold, leading, and white space — to create a hierarchy that guides the audience's eyes through the graphic, starting at the title, *Hazy days*, and negative space plus the heat maps direct the audience's view towards the encoding explanations. Almost half of the graphic uses negative space, and carefully separates types of information to reduce cognitive load in understanding the information. Mini-explainers are paragraph-aligned towards the side it refers to.

Credibility, transparency?

Provides explicit citations to the underlying data — Environmental Protection Department, Hong Kong Observatory — and explains missing data encodings: "no wind data available..."

Scarr, Simon. "Hazy days" South China Morning Post, December 17, 2012, sec. Infographics. <https://multimedia.scmp.com/culture/article/SCMP-printed-graphics-memory/lonelyGraphics/201212A230.html>.

Overall assessment: Scarr's information graphic succeeds for its general audience and purpose, which is primarily to allow exploration of the data. Its use of white space is particularly helpful as an example to our own work. The graphic may become stronger with messaging in the title, rather than just description, encoding wind by the hour, and adding narrative from either a historical perspective, or to show its correlation with something about people's lives. Other purposes would need better narrative and a call to action.



criticism for data-driven, visual narratives, practicing critiques

Audience?

Purpose?

Data encodings, decodings?

Comparison or change?

Narrative?

Color, coherency?

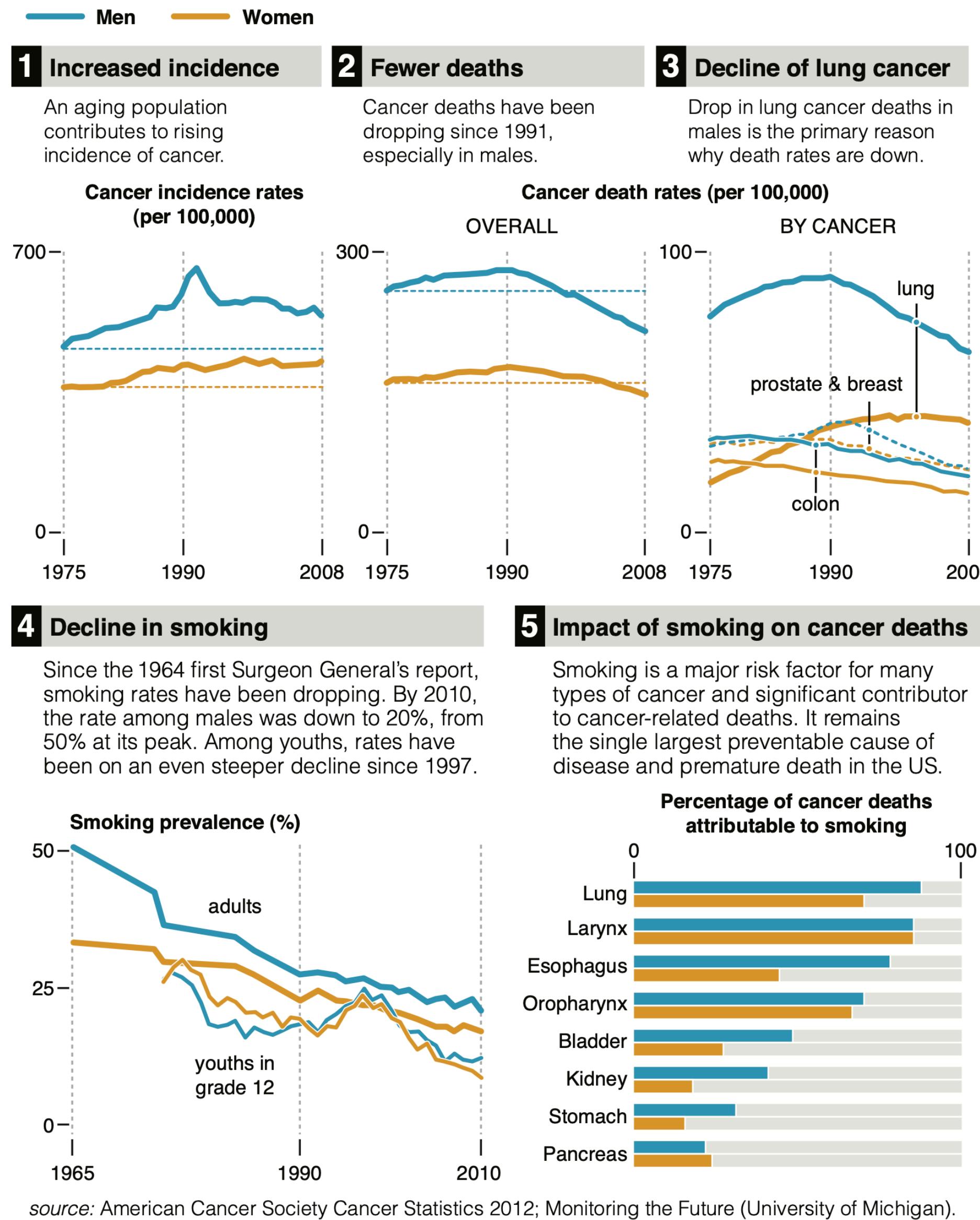
Hierarchy, layering, layout?

Credibility, transparency?

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

WHERE THERE'S SMOKE—THERE'S CANCER

Cancer rates are up, but mortality is down. New diagnostics and treatments are responsible for part of this trend. But the greatest single contributing factor is the decline in smoking—rates are at their lowest level in 50 years.



criticism for data-driven, visual narratives, practicing critiques

Audience?

Purpose?

Data encodings, decodings?

Comparison or change?

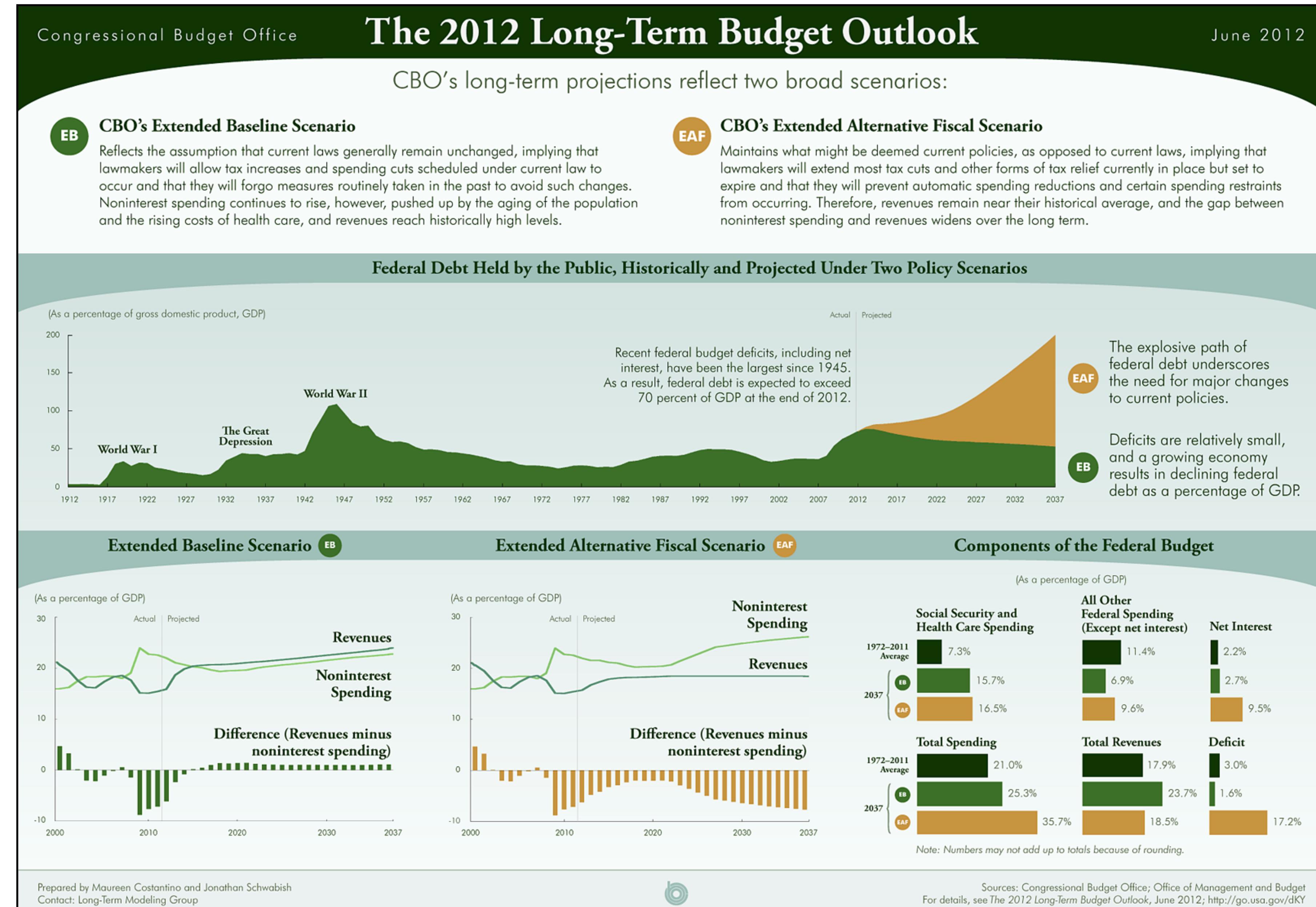
Narrative?

Color, coherency?

Hierarchy, layering, layout?

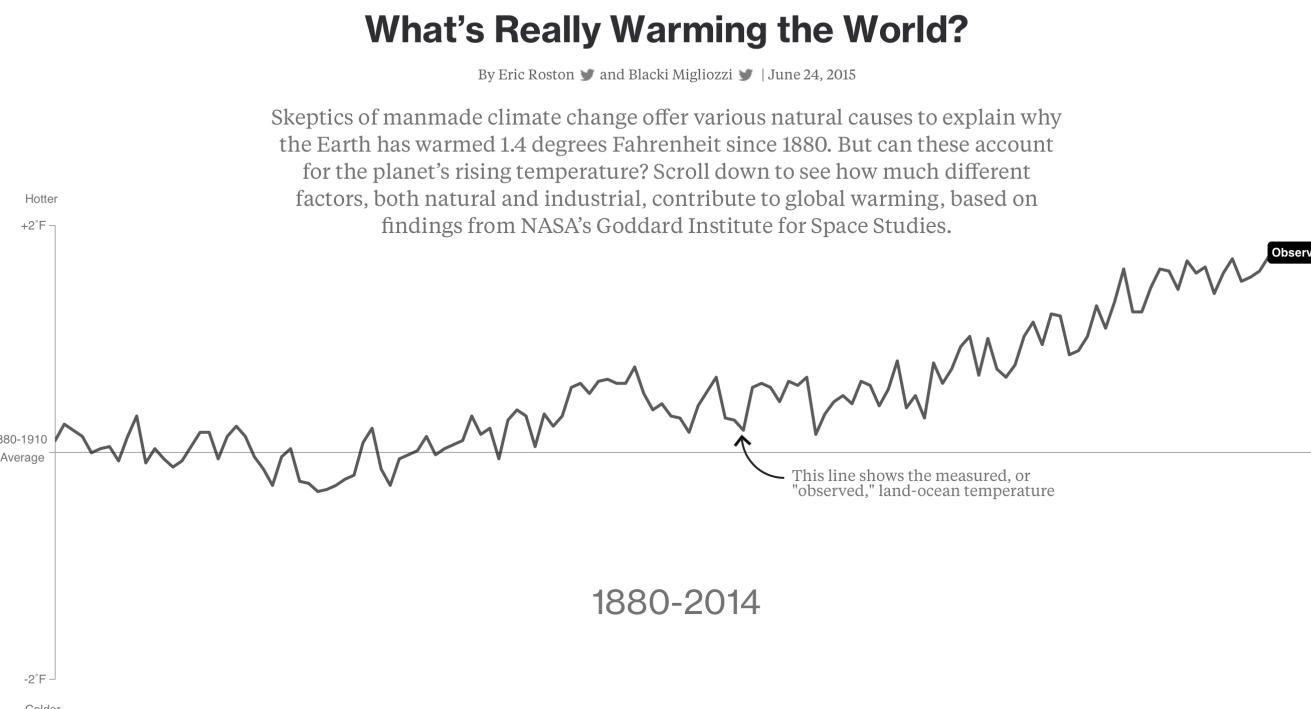
Credibility, transparency?

Schwabish, Jonathan, Maureen Costantino. "The 2012 Long-Term Budget Outlook: Infographic." Congressional Budget Office, June 5, 2012. <https://www.cbo.gov/publication/43289>.



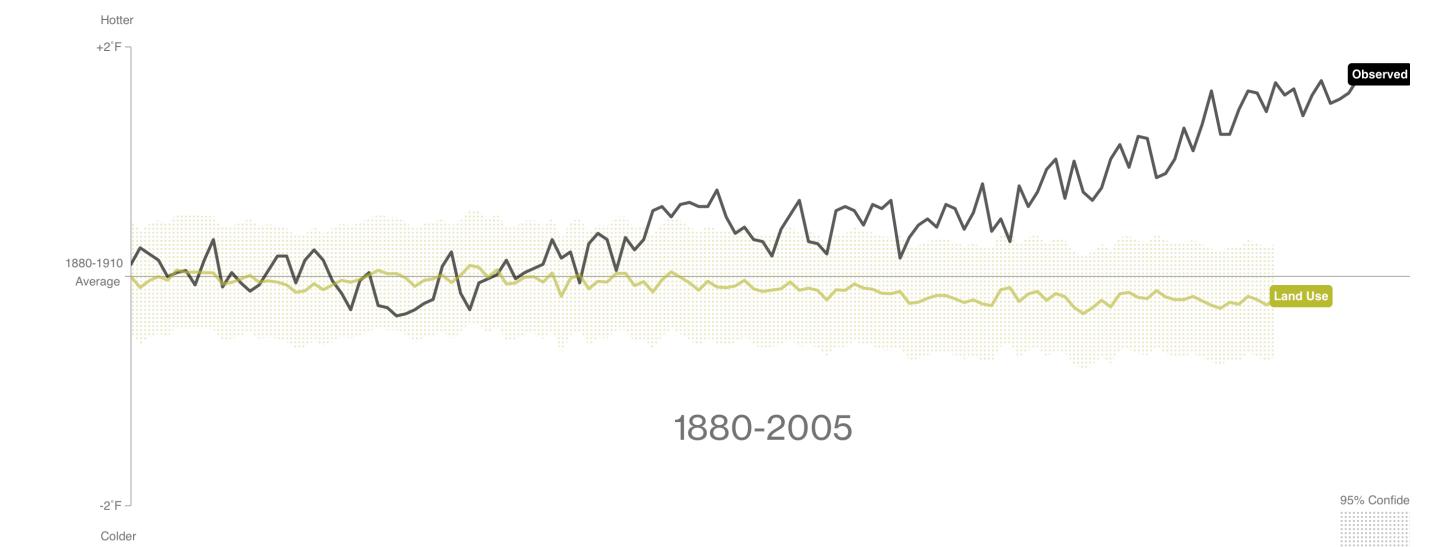
criticism for data-driven, visual narratives, practicing critiques

Audience?



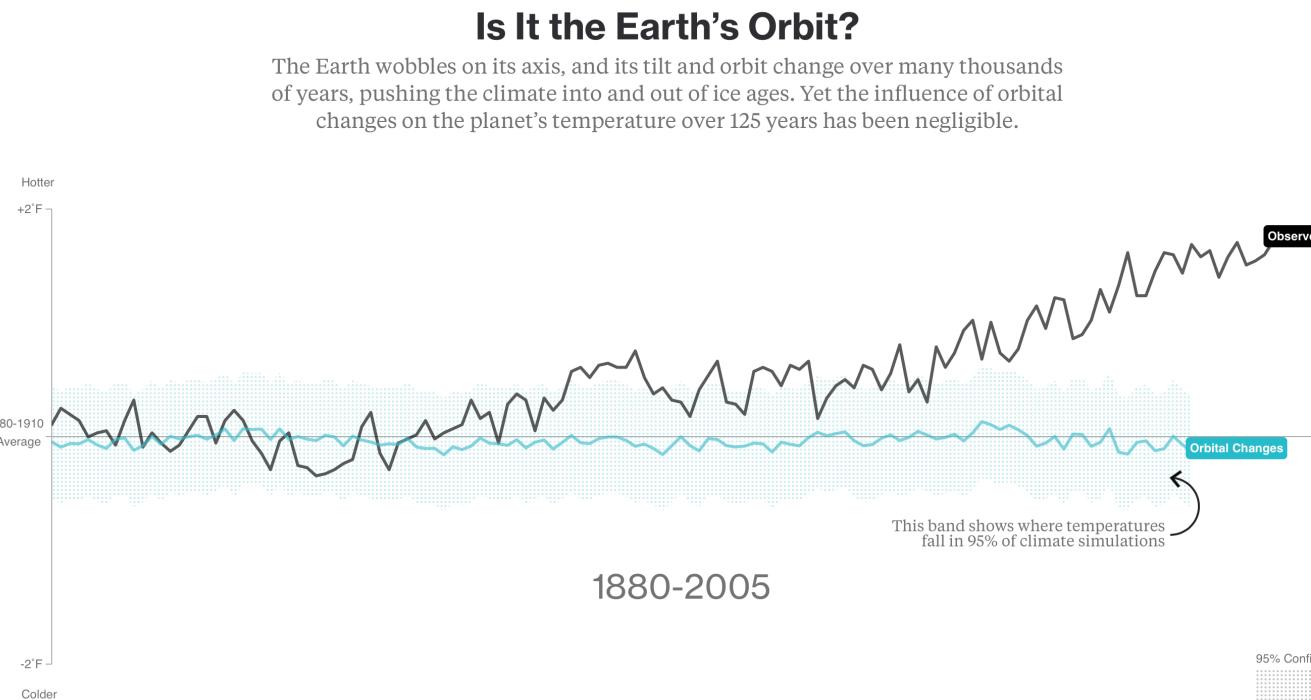
So If It's Not Nature, Is It Deforestation?

Humans have cut, plowed, and paved more than half the Earth's land surface. Dark forests are yielding to lighter patches, which reflect more sunlight—and have a slight cooling effect.



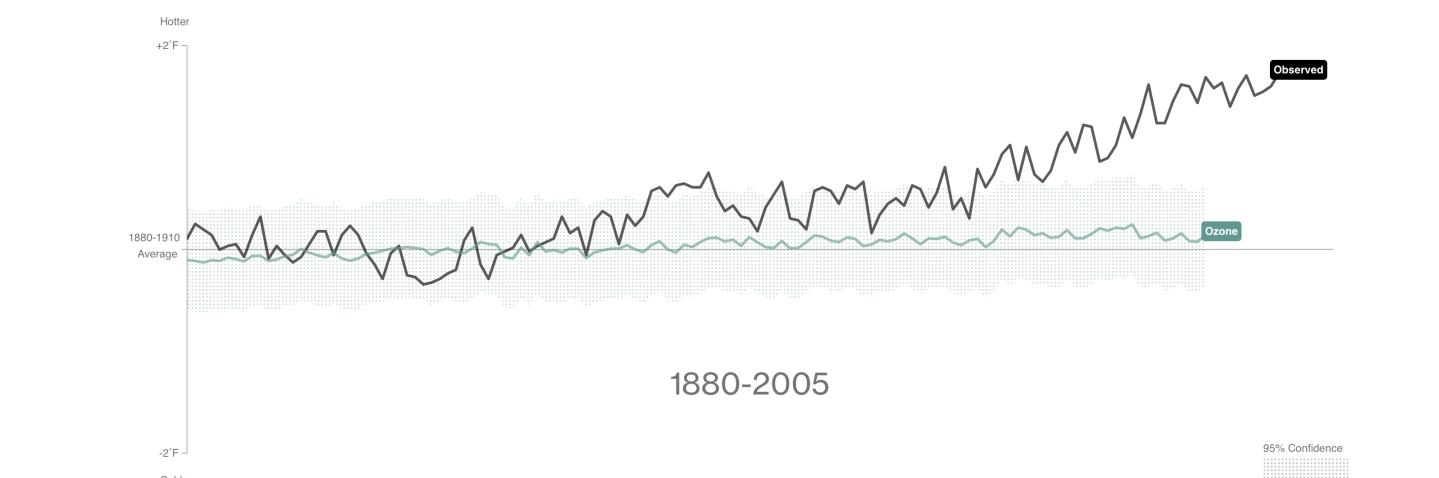
Purpose?

Data encodings, decodings?



Or Ozone Pollution?

Natural ozone high in the atmosphere blocks harmful sunlight and cools things slightly. Closer to Earth, ozone is created by pollution and traps heat, making the climate a little bit hotter. What's the overall effect? Not much.

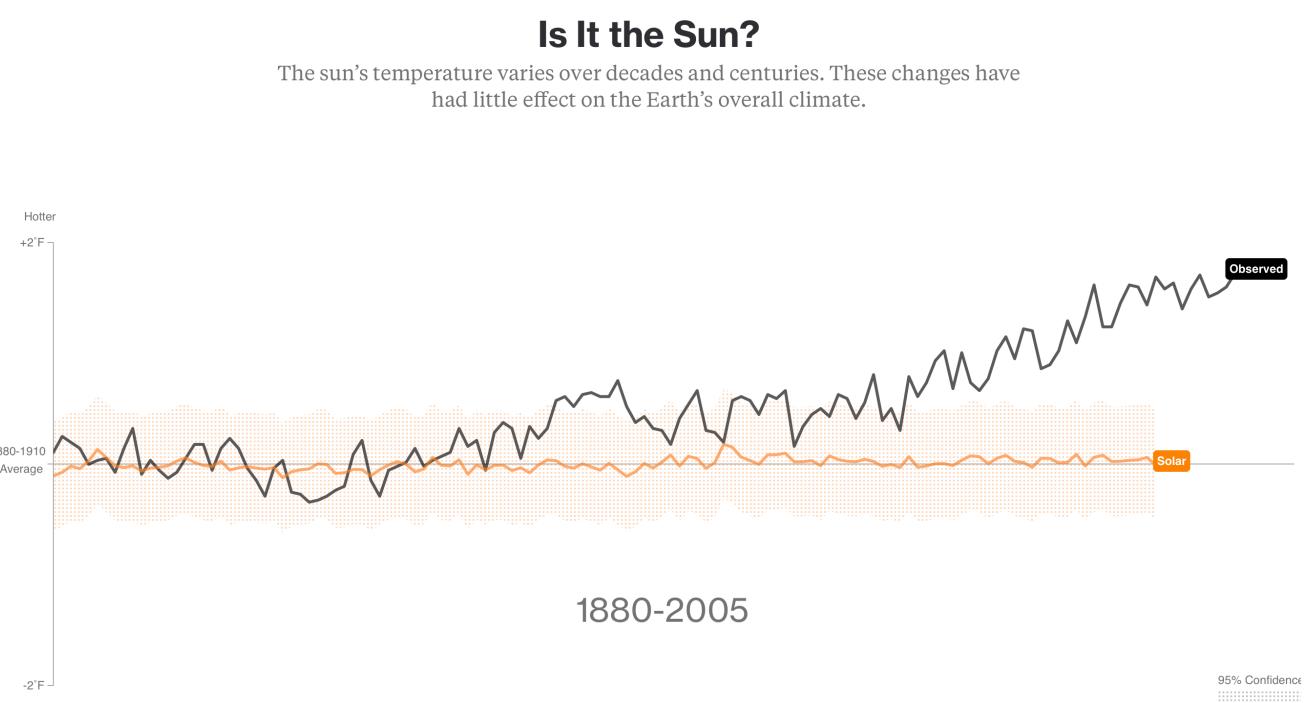


Comparison or change?

Narrative?

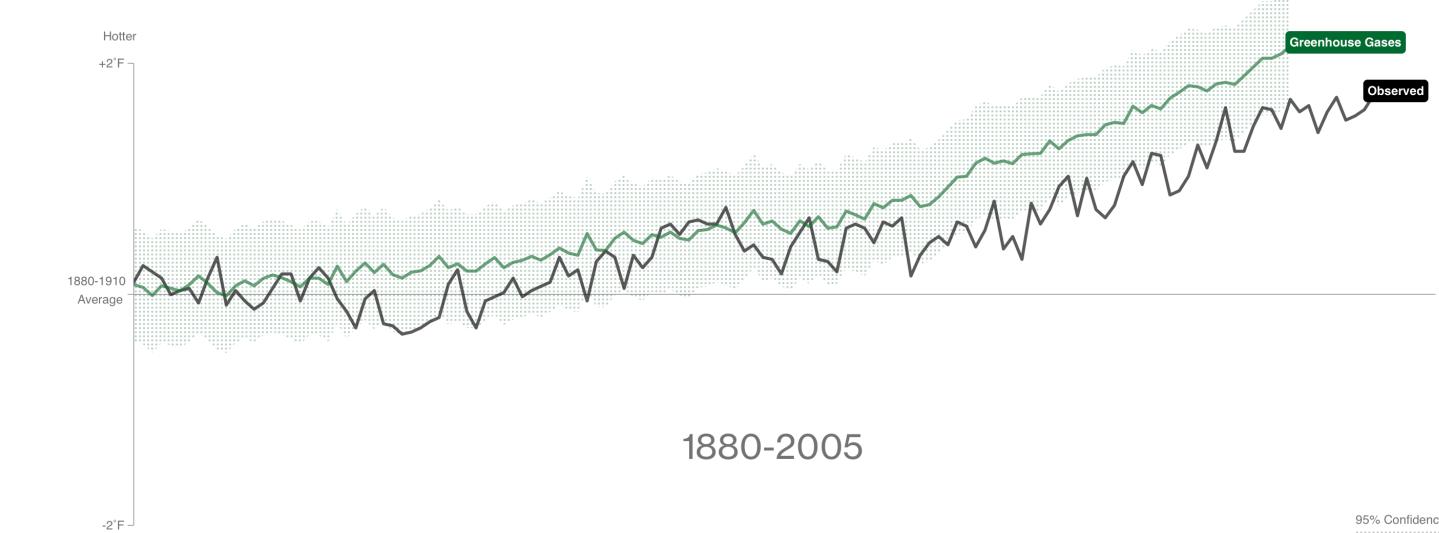
Color, coherency?

Hierarchy, layering, layout?



No, It Really Is Greenhouse Gases.

Atmospheric CO₂ levels are 40 percent higher than they were in 1750. The green line shows the influence of greenhouse gas emissions. It's no contest.



Credibility, transparency?

Roston, Eric, and Blacki Migliozzi. "What's Really Warming the World?" Bloomberg, June 24, 2015, Businessweek edition. <https://www.bloomberg.com/graphics/2015-whats-warming-the-world/>.

criticism for data-driven, visual narratives, practicing critiques

Audience?

Purpose?

Data encodings, decodings?

Comparison or change?

Narrative?

Color, coherency?

Hierarchy, layering, layout?

Credibility, transparency?

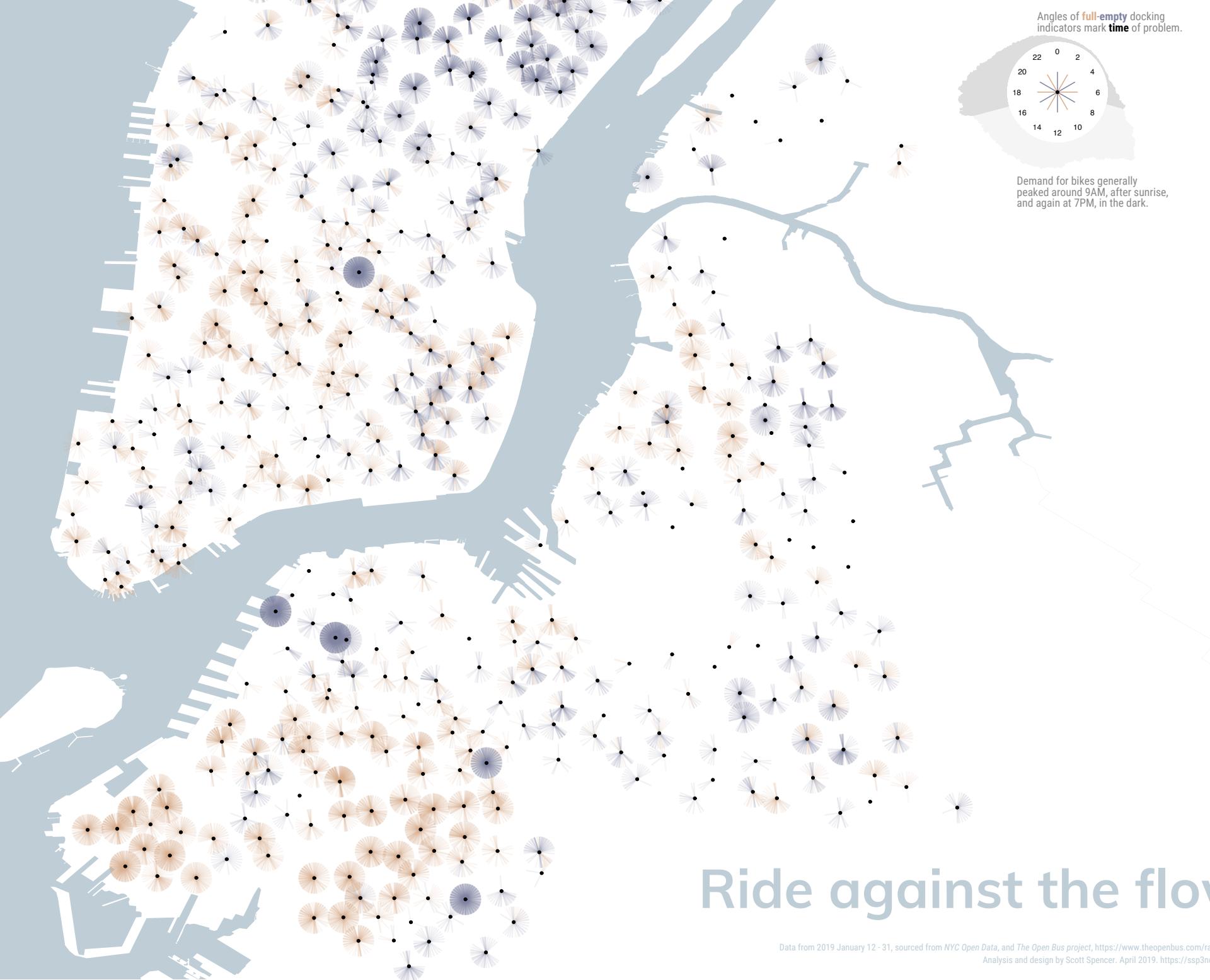
Spencer, Scott. *Ride Against the Flow*. 2019. Kantar IIB Awards. <https://www.informationisbeautifulawards.com/showcase/4367-ride-against-the-flow>.

For the past six springs, New Yorkers pedaled past colorful blossoms on their way to work, home, or just cruising.

Yet, some cruisers waded 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.



criticism for data-driven, visual narratives, practicing critiques — goals for *your* data-driven, visual narratives

Audience? An external, general audience. Categorically, who are this mixed audience?

Purpose? Decide on your purpose; be specific. *E.g.*, Advertising? Public relations? Investor interest? Get your audience's attention, help them understand, and be able to act on your message's purpose.

Data encodings, decodings? Encode your data, statistics, and modelling estimates using best practices we've discussed. Data encodings should directly support your main messages.

Comparison or change? Encode data to show comparisons or change, add data as context to impart meaning.

Narrative? Think about your narrative arc, and how change drives your narrative forward. Do you use explainers or labels and mini paragraphs on your data graphics to help your audience?

Color, coherency? Purposefully use color for encodings and linking data encodings to textual narrative.

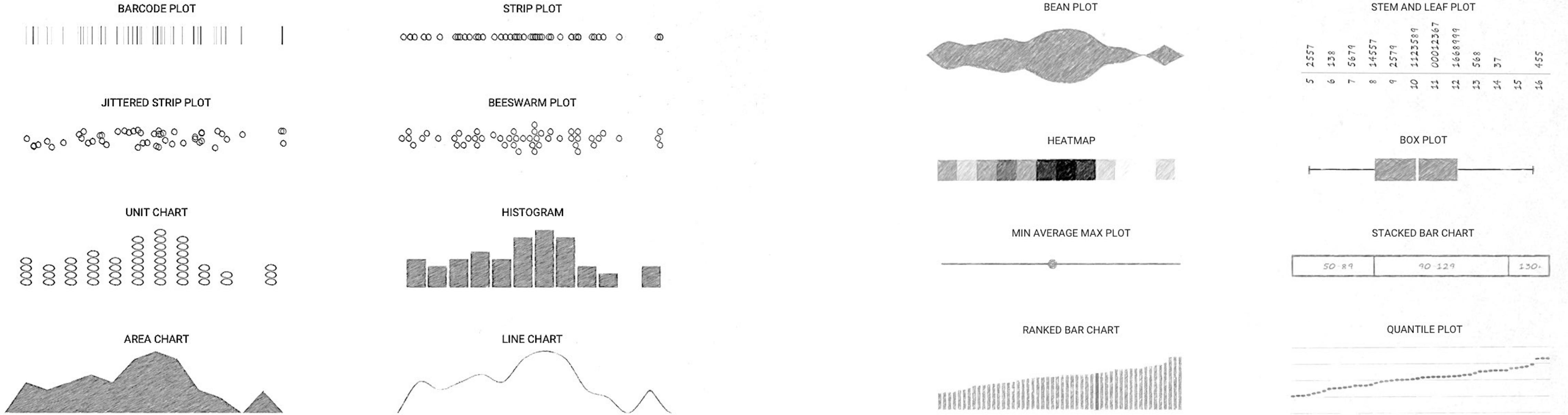
Hierarchy, layering, layout? Your titles, headers, mini-paragraphs, and text should use messages, not just information. Use best practices in typography (size, bold, color, spacing, etc) and grid alignment to focus your audience on your messages.

Credibility, transparency? Cite your sources, briefly mention any important elements of your analysis. Consider whether you need to explain any limitations or exceptions.

help your colleagues — group ideating

**encoding uncertainty,
estimates, forecasts**

Measurements are observed. Examples of common visual encodings for variation in measures ...



Cherdarchuk, Joey. "Visualizing Distributions." Business. Dark Horse Analytics (blog), November 8, 2016. <https://www.darkhorseanalytics.com/blog/visualizing-distributions-3>.

... but estimates are *not* observed measures — *they are modeled from measures* — be clear about distinguishing them with your encodings and annotations.

encoding uncertainty, estimates, forecasts, distinguishing measurements from estimates — examples

In a game against New York Yankees, should Milwaukee Brewers's Lorenzo Cain attempt to steal second base with no one else on base and two outs before the seventh inning, against Gary Sanchez as catcher and Michael Pineda as pitcher? What if against Sanchez and CC Sabathia as pitcher?

More specifically, how can we know the *expectation* that Cain's attempt in each situation increases the probability of expected runs that inning and by how much? Using Stan, I've coded a generative model that along with play outcomes considers various information (runner foot-speed, catcher pop-time) and player characteristics, like pitcher handedness. With the model, we have an answer that also shows the uncertainty. Given 2017 data, this model suggests Cain should steal against Pineda, not Sabathia:

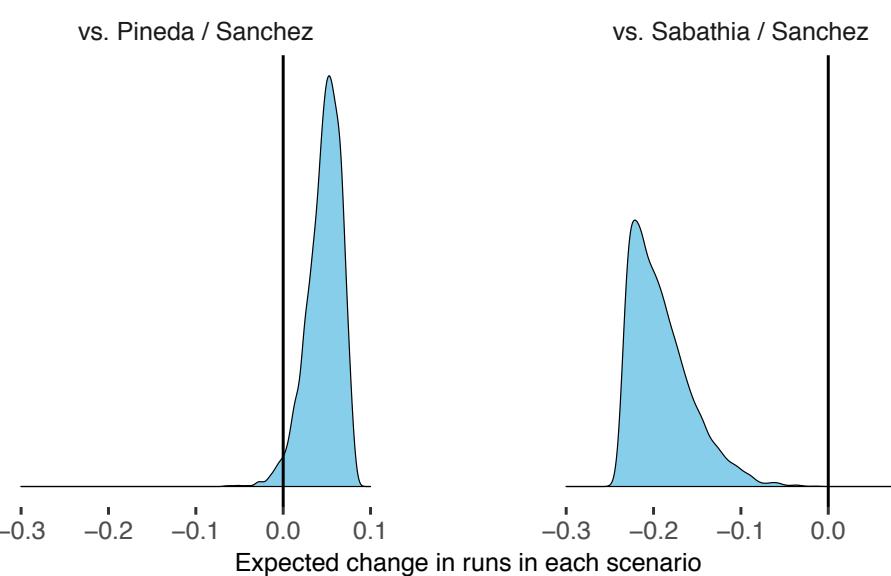


Figure 1. Of the two scenarios, Cain should only attempt to steal against the Sanchez–Pineda duo.

Notably, we get these expectations without multiple trials of either scenario. More generally, this model suggests that on average team managers are too conservative, leaving runs unrealized:

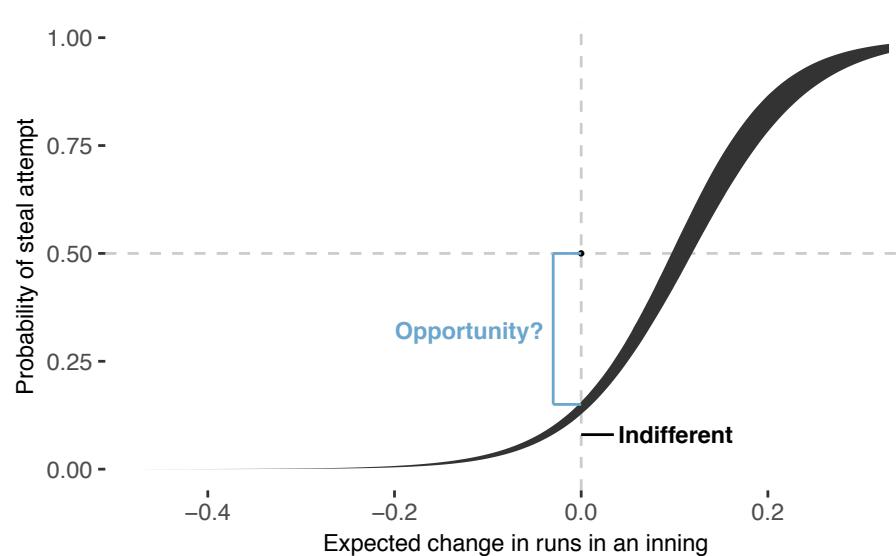


Figure 2. When the change in expected runs is zero, managers should be indifferent to attempted steals, saying go half the time.

The **black band** represents the range of variation across managers' decisions. At the intersection of **indifference**, managers tend to say steal only **10 percent** of the time, leaving opportunity.

The above is but one example of a more general approach that weighs probabilities of all possible outcomes to maximize expected utility. With broad implementation—jointly modeling the conditional probabilities of all relevant events—we can optimize decisions.

encoding uncertainty, estimates, forecasts, distinguishing measurements from estimates — examples

In a game against New York Yankees, should Milwaukee Brewers's Lorenzo Cain attempt to steal second base with no one else on base and two outs before the seventh inning, against Gary Sanchez as catcher and Michael Pineda as pitcher? What if against Sanchez and CC Sabathia as pitcher?

More specifically, how can we know the *expectation* that Cain's attempt in each situation increases the probability of *expected* runs that inning and by how much? Using Stan, I've coded a *generative model* that along with play outcomes considers various information (runner foot-speed, catcher pop-time) and player characteristics, like pitcher handedness. With the model, we have an answer that also shows the uncertainty. Given 2017 data, this model suggests Cain should steal against Pineda, not Sabathia:

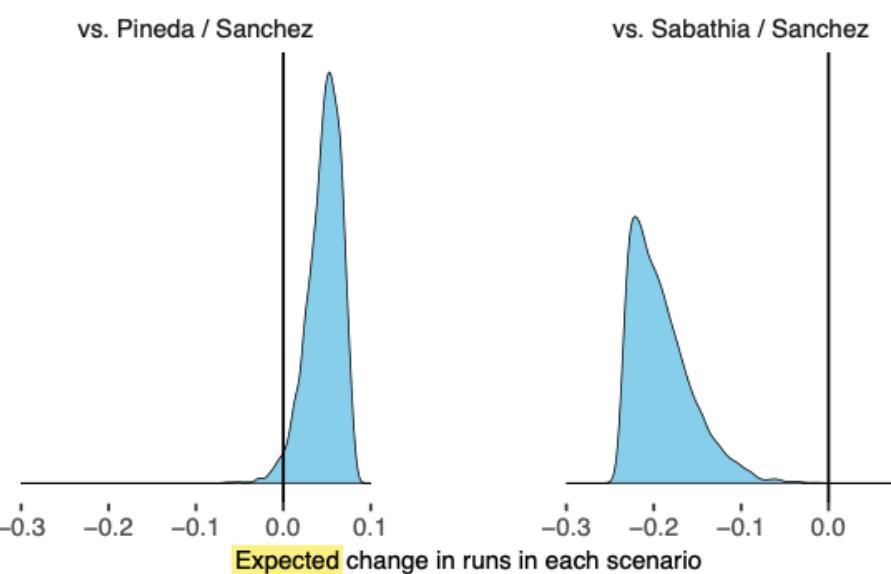


Figure 1. Of the two scenarios, Cain should only attempt to steal against the Sanchez–Pineda duo.

Notably, we get these *expectations* without multiple trials of either scenario. More generally, this model suggests that on average team managers are too conservative, leaving runs unrealized:



Figure 2. When the change in *expected* runs is zero, managers should be indifferent to attempted steals, saying go half the time.

The **black band** represents the range of variation across managers' decisions. At the intersection of *indifference*, managers tend to say steal only 10 percent of the time, leaving opportunity.

The above is but one example of a more general approach that weighs probabilities of all possible outcomes to maximize *expected utility*. With broad implementation—jointly modeling the conditional probabilities of all relevant events—we can optimize decisions.

encoding uncertainty, estimates, forecasts, distinguishing measurements from estimates — examples

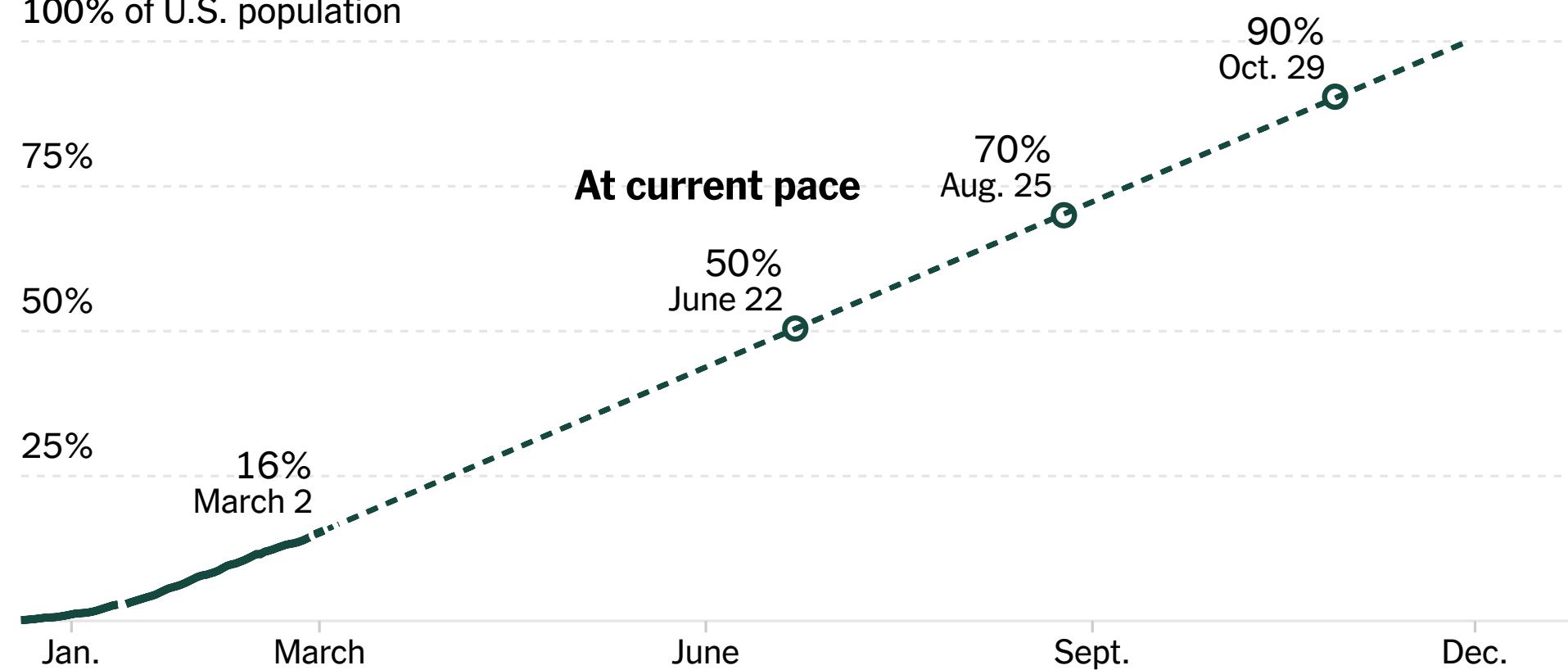
The projection below only shows the share of the total population with at least one shot based on the current rate of vaccination, but it provides a rough indication of when the virus's spread could begin to stall.

When a given share of the U.S. population might be at least partially vaccinated

The current vaccination rate is based on average daily increase in first doses administered over the past week.

Average daily first doses in last 7 days: 1,030,068

100% of U.S. population



Source: Centers for Disease Control and Prevention | Note: Data from Dec. 20 to Jan. 12 are for all doses administered. Data for Jan. 13 is unavailable. Projections could change if additional vaccines are authorized.

If the country maintains its current pace of administering first doses, about half of the total population would be at least partially vaccinated around late June, and nearly all around late October, assuming supply pledges are met and vaccines are eventually available to children.

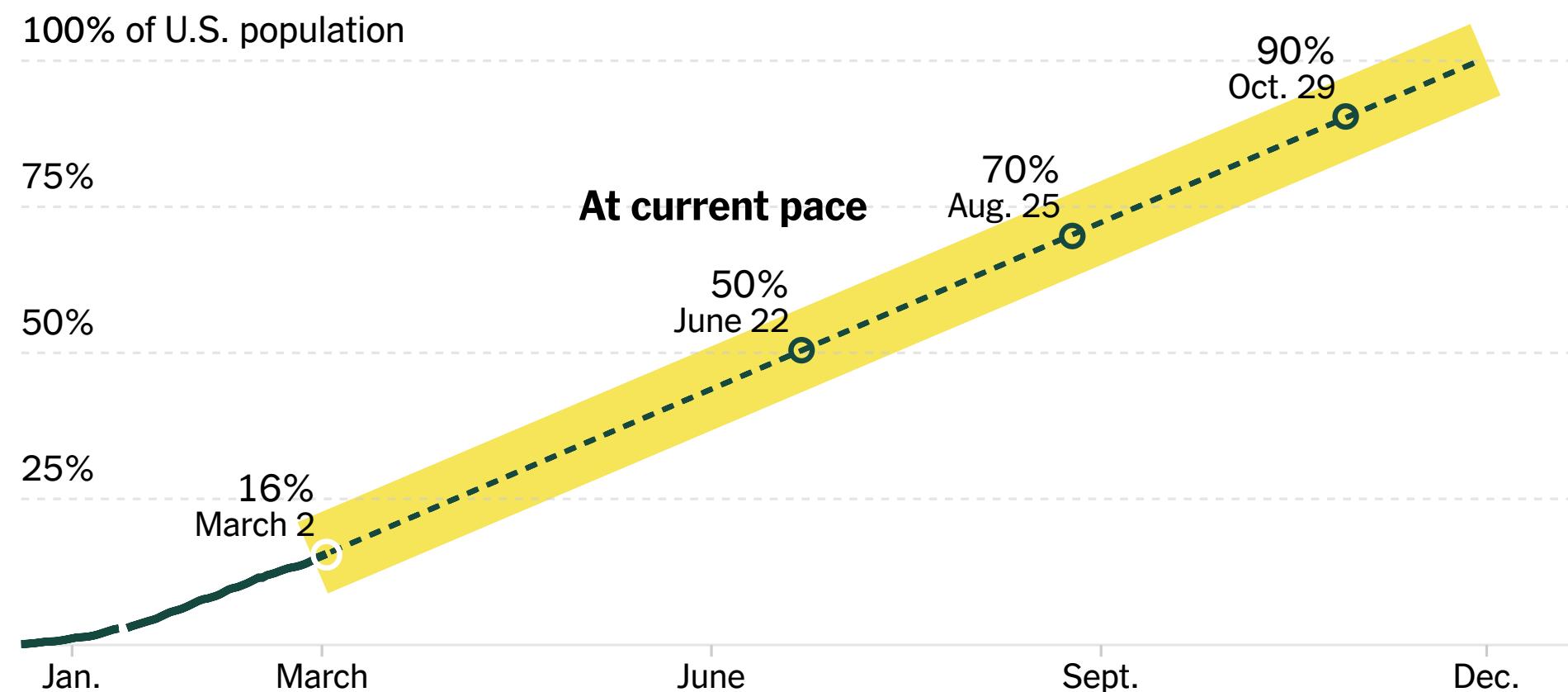
encoding uncertainty, estimates, forecasts, distinguishing measurements from estimates — examples

The projection below only shows the share of the total population with at least one shot based on the current rate of vaccination, but it provides a rough indication of when the virus's spread could begin to stall.

When a given share of the U.S. population might be at least partially vaccinated

The current vaccination rate is based on average daily increase in first doses administered over the past week.

Average daily first doses in last 7 days: 1,030,068



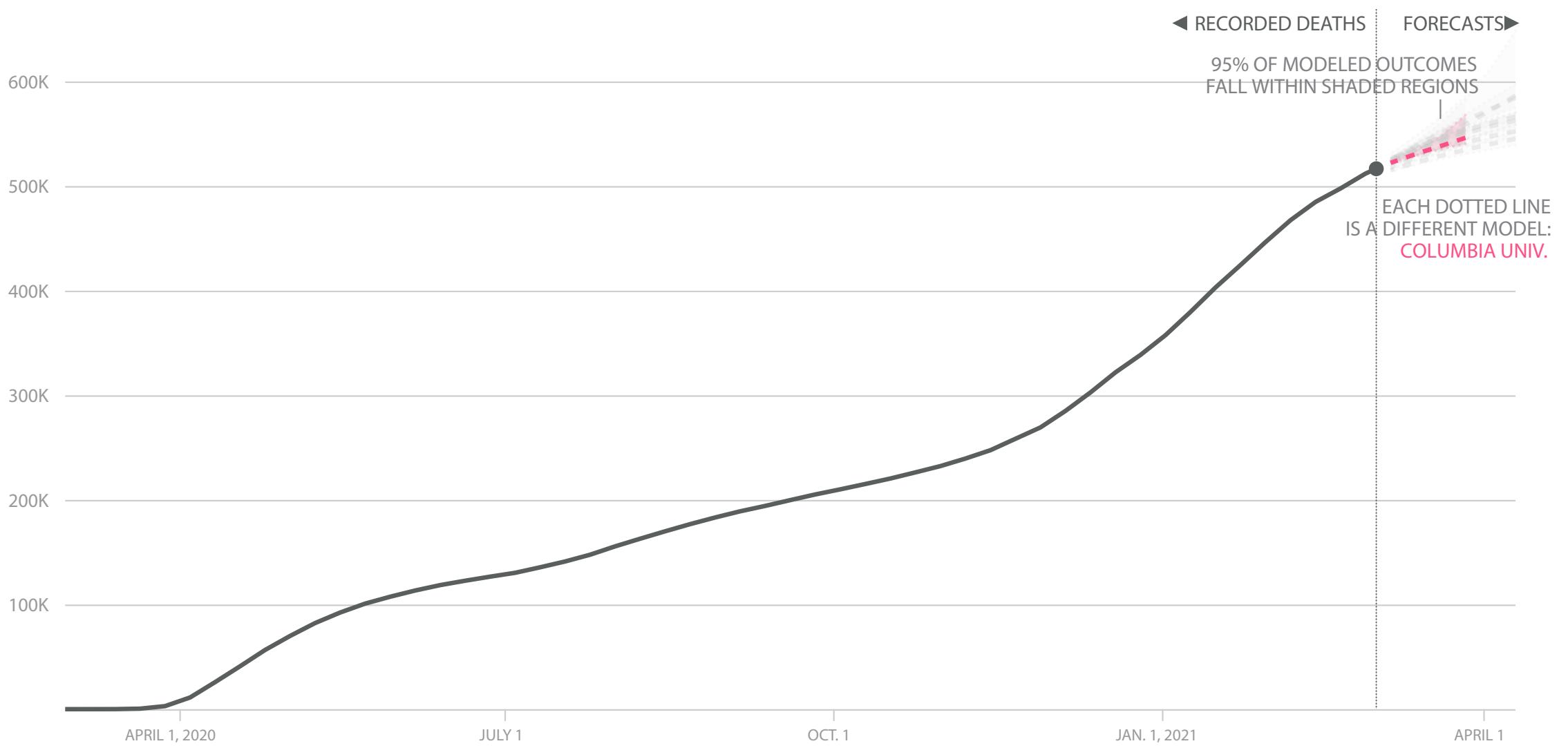
Source: Centers for Disease Control and Prevention | Note: Data from Dec. 20 to Jan. 12 are for all doses administered. Data for Jan. 13 is unavailable. Projections could change if additional vaccines are authorized.

If the country maintains its current pace of administering first doses, about half of the total population would be at least partially vaccinated around late June, and nearly all around late October, assuming supply pledges are met and vaccines are eventually available to children.

encoding uncertainty, estimates, forecasts, distinguishing measurements from estimates — examples

Models predicting the potential spread of the COVID-19 pandemic have become a fixture of American life. Yet each model tells a different story about the loss of life to come, making it hard to know which one is “right.” But COVID-19 models aren’t made to be unquestioned oracles. They’re not trying to tell us one precise future, but rather the range of possibilities given the facts on the ground.

One of their more sober tasks is predicting the number of Americans who will die due to COVID-19. FiveThirtyEight — with the help of data compiled by the [COVID-19 Forecast Hub](#) — has assembled 11 models published by scientists to illustrate possible trajectories of the pandemic’s death toll. In doing so, we hope to make them more accessible, as well as highlight how the assumptions underlying the models can lead to vastly different estimates. Here are the models’ U.S. fatality projections for the coming weeks.



Forecasts like these are useful because they help us understand the most likely outcomes as well as best- and worst-case possibilities — and they can help policymakers make decisions that can lead us closer to those best-case outcomes.

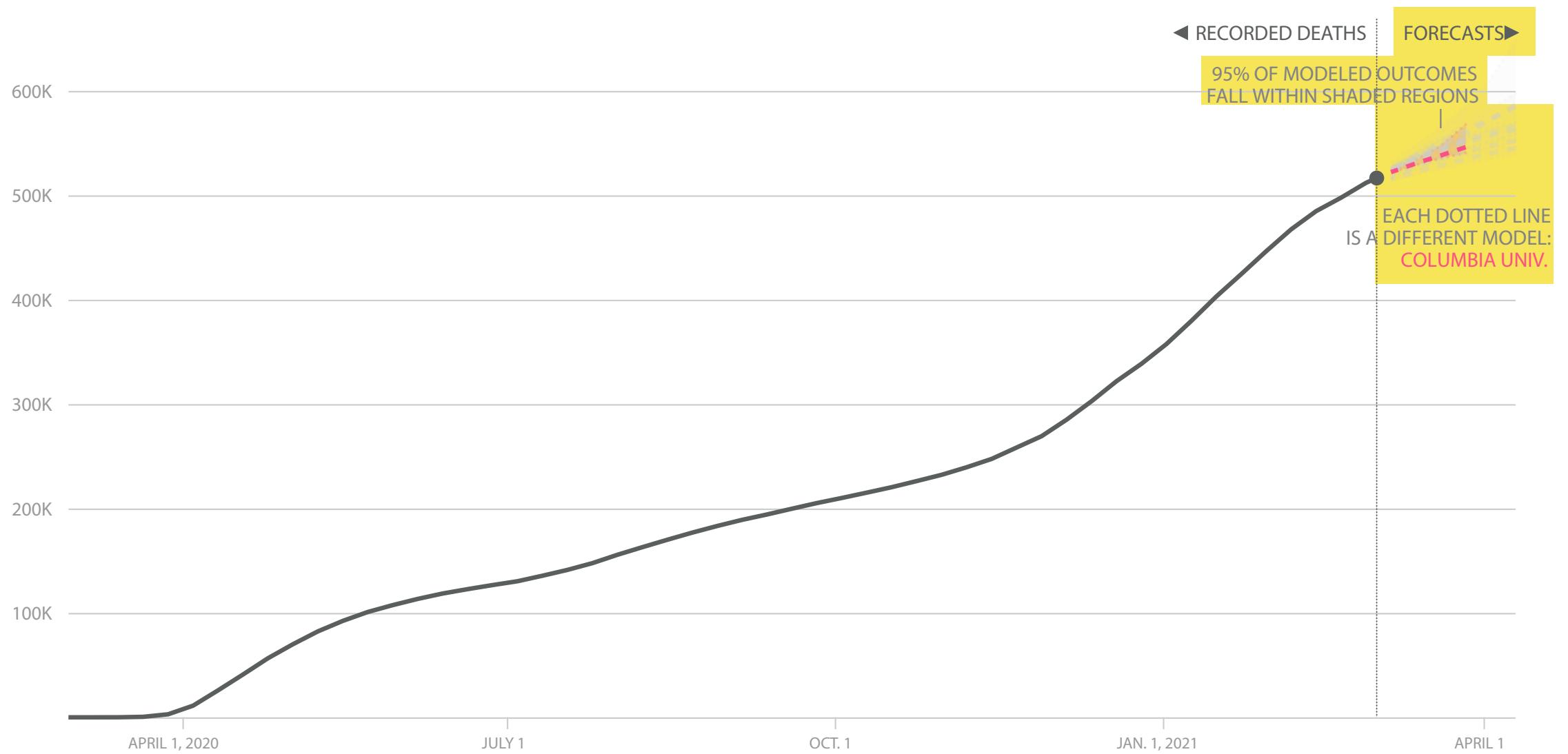
And looking at multiple models is better than looking at just one because it’s difficult to know which model will match reality the closest. Even when models disagree, understanding why they are different can give us valuable insight.

Best, Ryan, and Jay Boice. “Where The Latest COVID-19 Models Think We’re Headed — And Why They Disagree.” News. FiveThirtyEight, March 2, 2021. <https://projects.fivethirtyeight.com/covid-forecasts/>.

encoding uncertainty, estimates, forecasts, distinguishing measurements from estimates — examples

Models predicting the potential spread of the COVID-19 pandemic have become a fixture of American life. Yet each model tells a different story about the loss of life to come, making it hard to know which one is “right.” But COVID-19 models aren’t made to be unquestioned oracles. They’re not trying to tell us one precise future, but rather the range of possibilities given the facts on the ground.

One of their more sober tasks is predicting the number of Americans who will die due to COVID-19. FiveThirtyEight — with the help of data compiled by the [COVID-19 Forecast Hub](#) — has assembled 11 models published by scientists to illustrate possible trajectories of the pandemic’s death toll. In doing so, we hope to make them more accessible, as well as highlight how the assumptions underlying the models can lead to vastly different estimates. Here are the models’ U.S. fatality projections for the coming weeks.



Forecasts like these are useful because they help us understand the most likely outcomes as well as best- and worst-case possibilities — and they can help policymakers make decisions that can lead us closer to those best-case outcomes.

And looking at multiple models is better than looking at just one because it's difficult to know which model will match reality the closest. Even when models disagree, understanding why they are different can give us valuable insight.

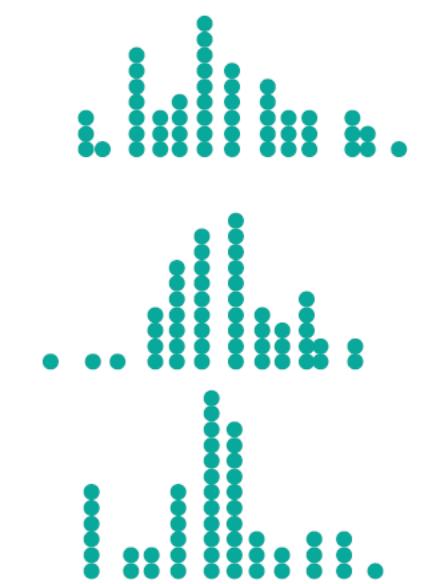
Best, Ryan, and Jay Boice. “Where The Latest COVID-19 Models Think We’re Headed — And Why They Disagree.” News. FiveThirtyEight, March 2, 2021. <https://projects.fivethirtyeight.com/covid-forecasts/>.

encoding uncertainty, estimates, forecasts, discretizing distributions to improve decisions — quantile dot plots

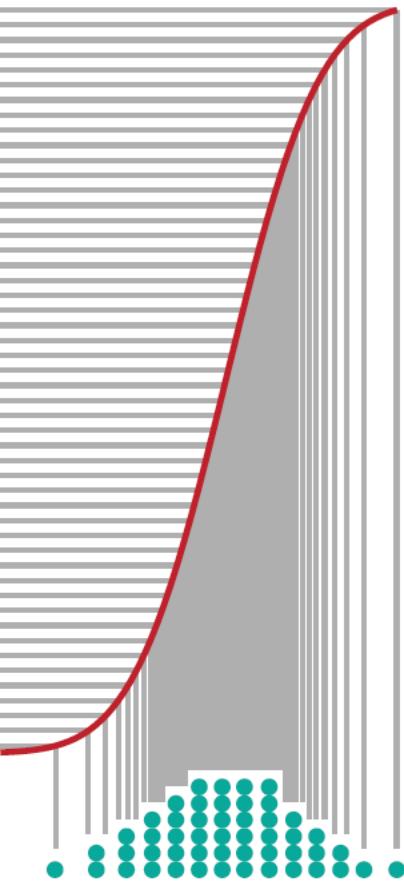
Probability density of Normal distribution



To generate a discrete plot of this distribution, we could try taking **random draws** from it. However, **this approach is noisy**: it may be very different from one instance to the next.



Instead, we use the **quantile function (inverse CDF)** of the distribution to generate "draws" from evenly-spaced quantiles.



We plot the quantile "draws" using a Wilkinsonian dotplot, yielding what we call a **quantile dotplot**: a consistent discrete representation of a probability distribution.



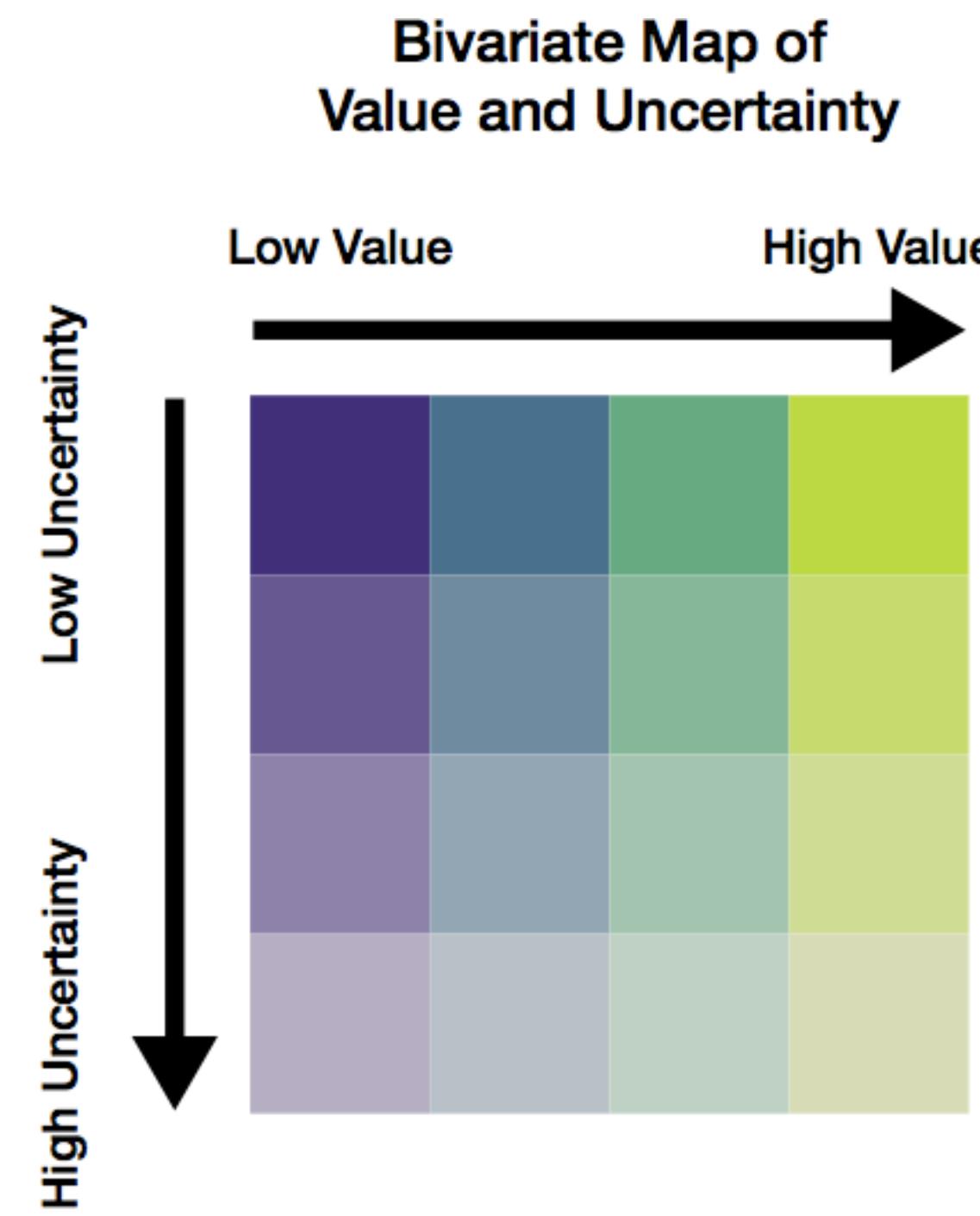
By using quantiles we facilitate interval estimation from frequencies: e.g., knowing there are 50 dots here, if we are willing to miss our bus **3/50** times, we can count **3 dots** from the left to get a one-sided **94% ($1 - 3/50$) prediction interval** corresponding to that risk tolerance.

encoding uncertainty, estimates, forecasts, using color to encode uncertainty — value suppressing uncertainty palettes



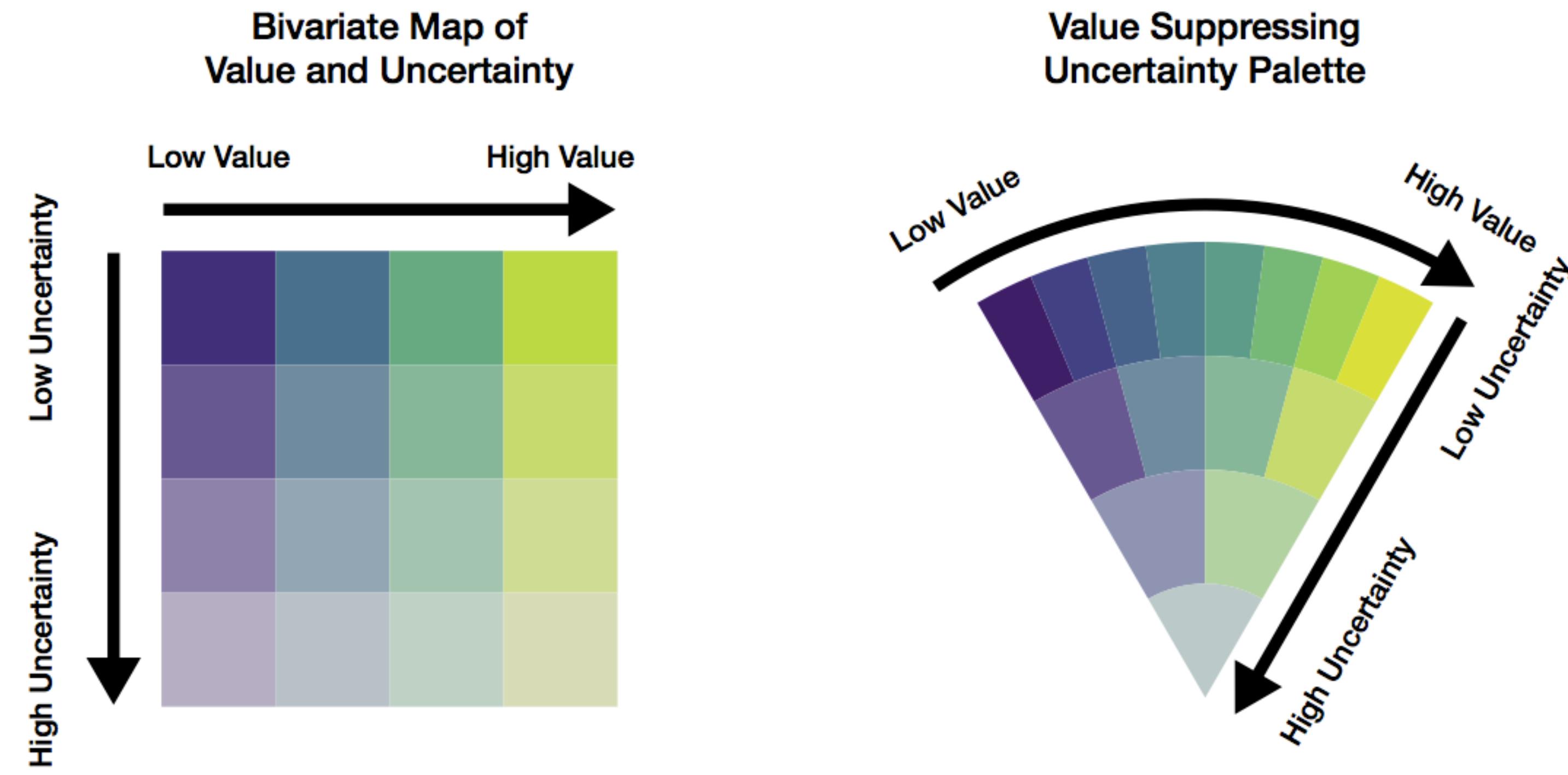
Correll, Michael, Dominik Moritz, and Jeffrey Heer. "Value-Suppressing Uncertainty Palettes." In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–11. Montreal QC, Canada: ACM Press, 2018.

encoding uncertainty, estimates, forecasts, using color to encode uncertainty — value suppressing uncertainty palettes



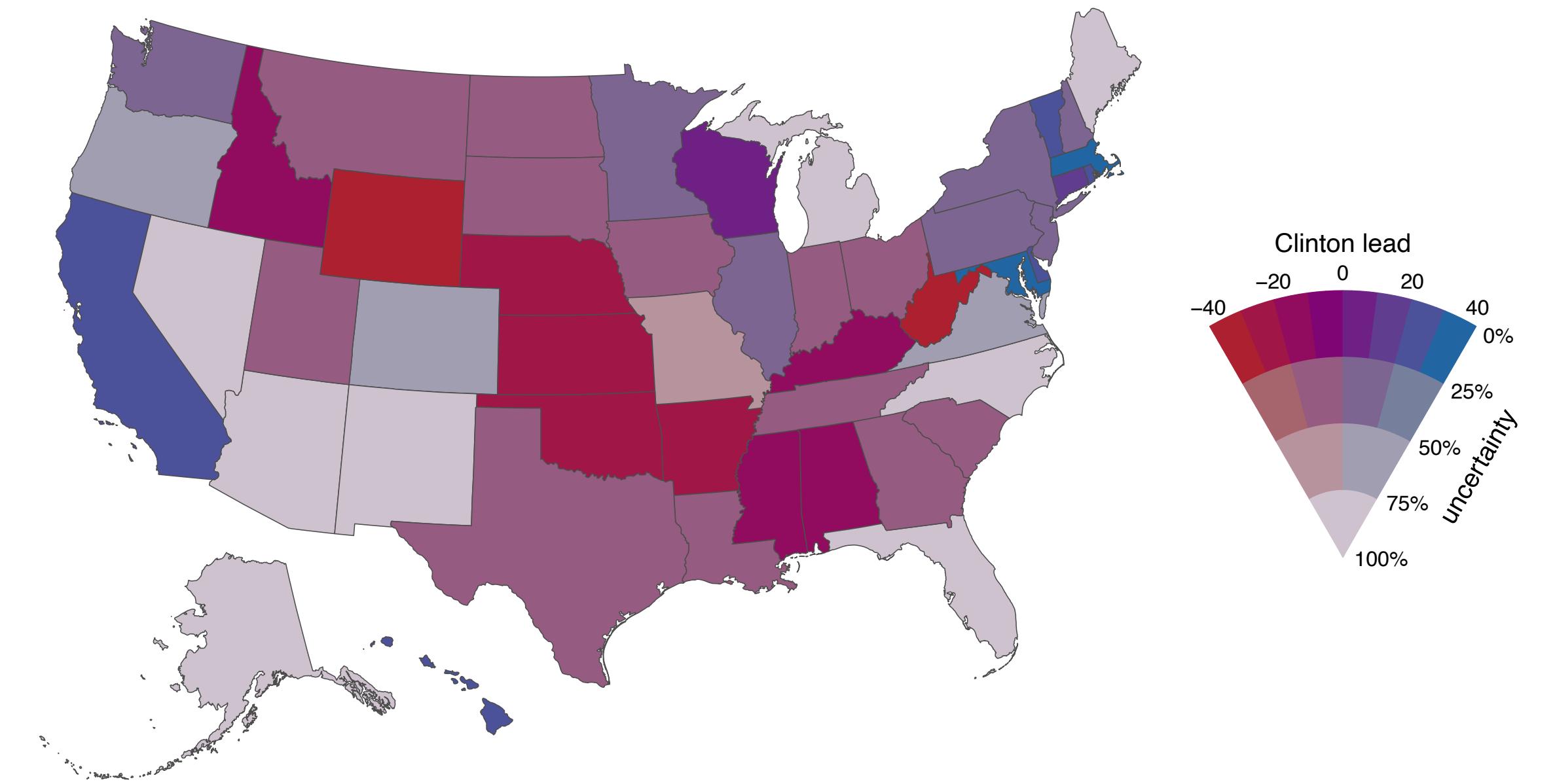
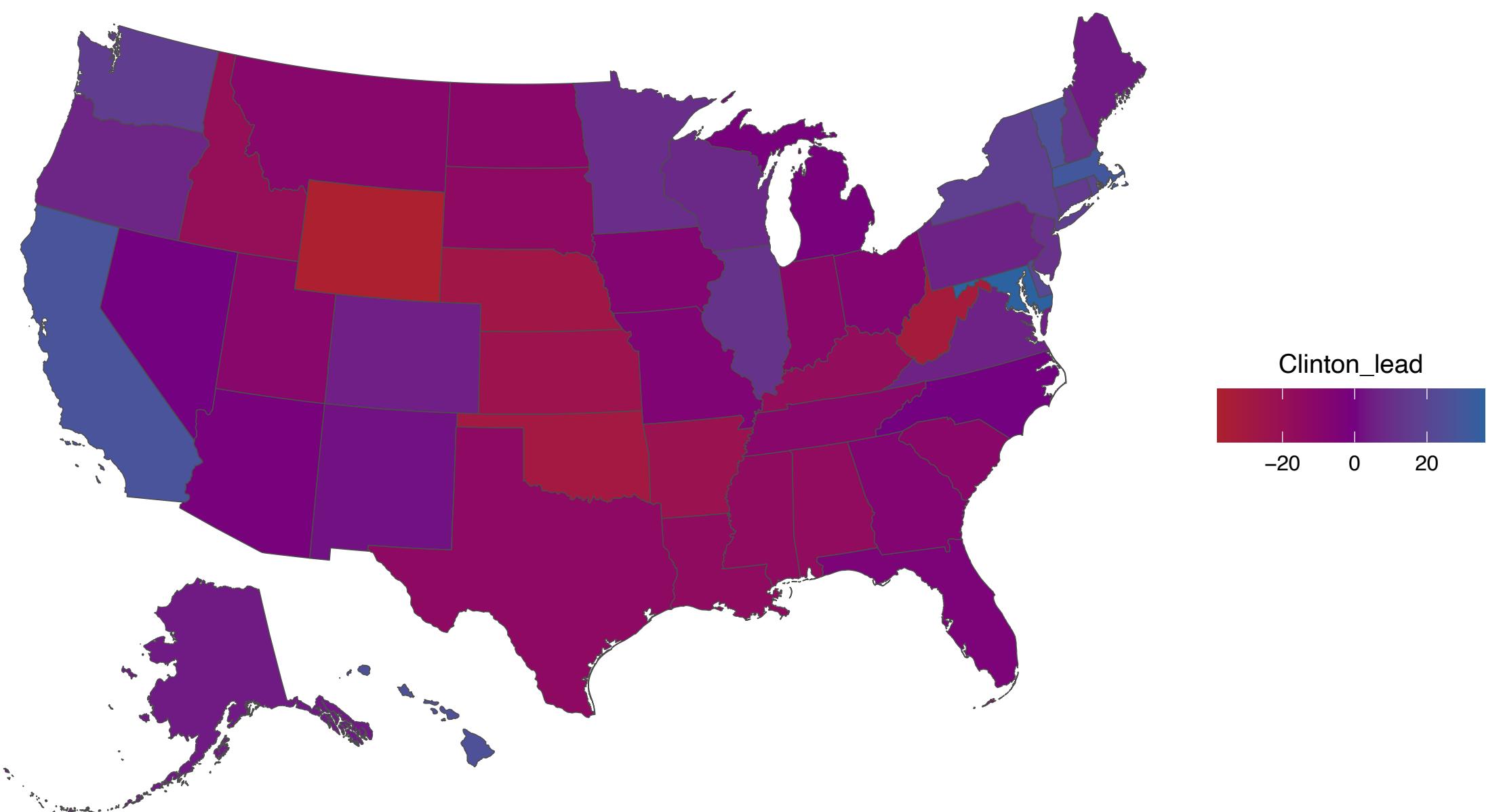
Correll, Michael, Dominik Moritz, and Jeffrey Heer. "Value-Suppressing Uncertainty Palettes." In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–11. Montreal QC, Canada: ACM Press, 2018.

encoding uncertainty, estimates, forecasts, using color to encode uncertainty — value suppressing uncertainty palettes



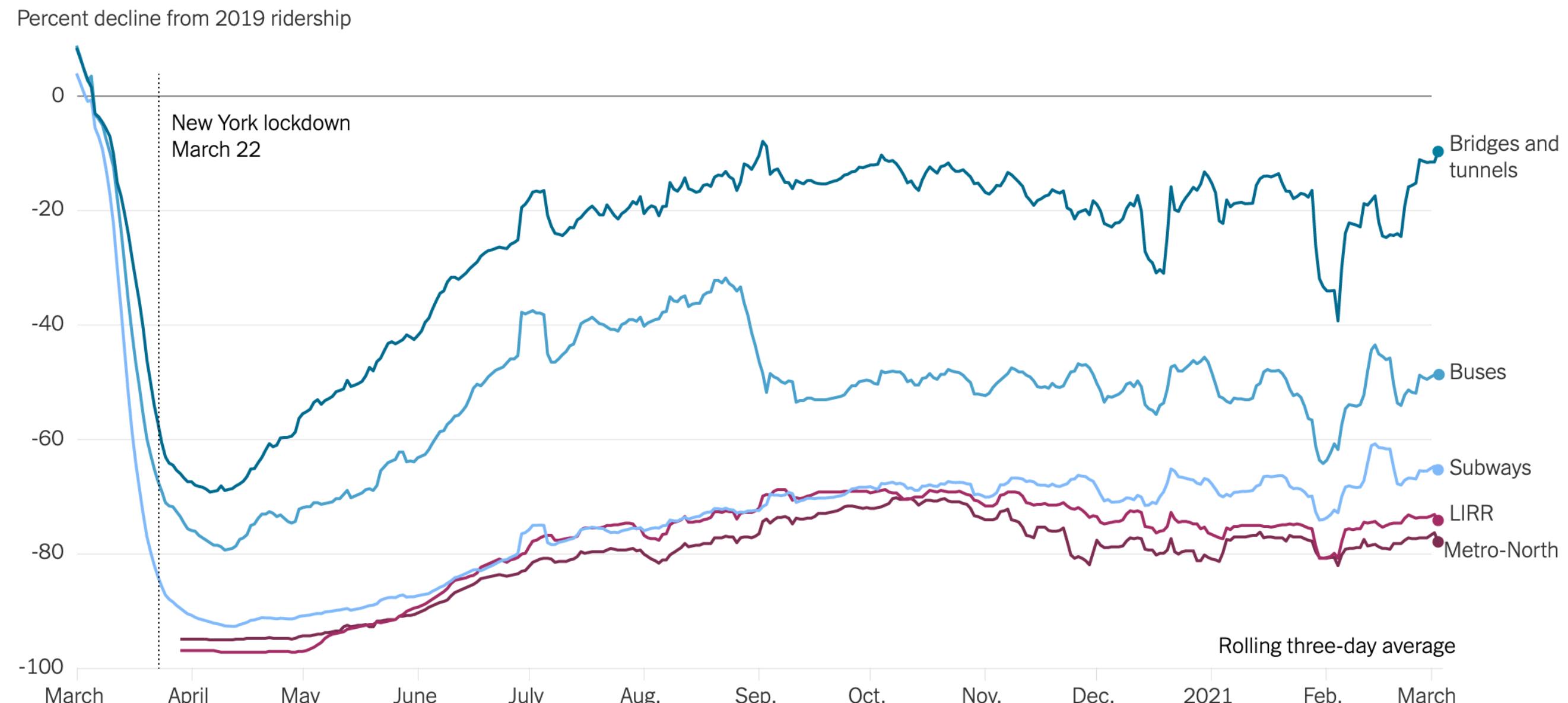
Correll, Michael, Dominik Moritz, and Jeffrey Heer. "Value-Suppressing Uncertainty Palettes." In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–11. Montreal QC, Canada: ACM Press, 2018.

encoding uncertainty, estimates, forecasts, using color to encode uncertainty — value suppressing uncertainty palettes

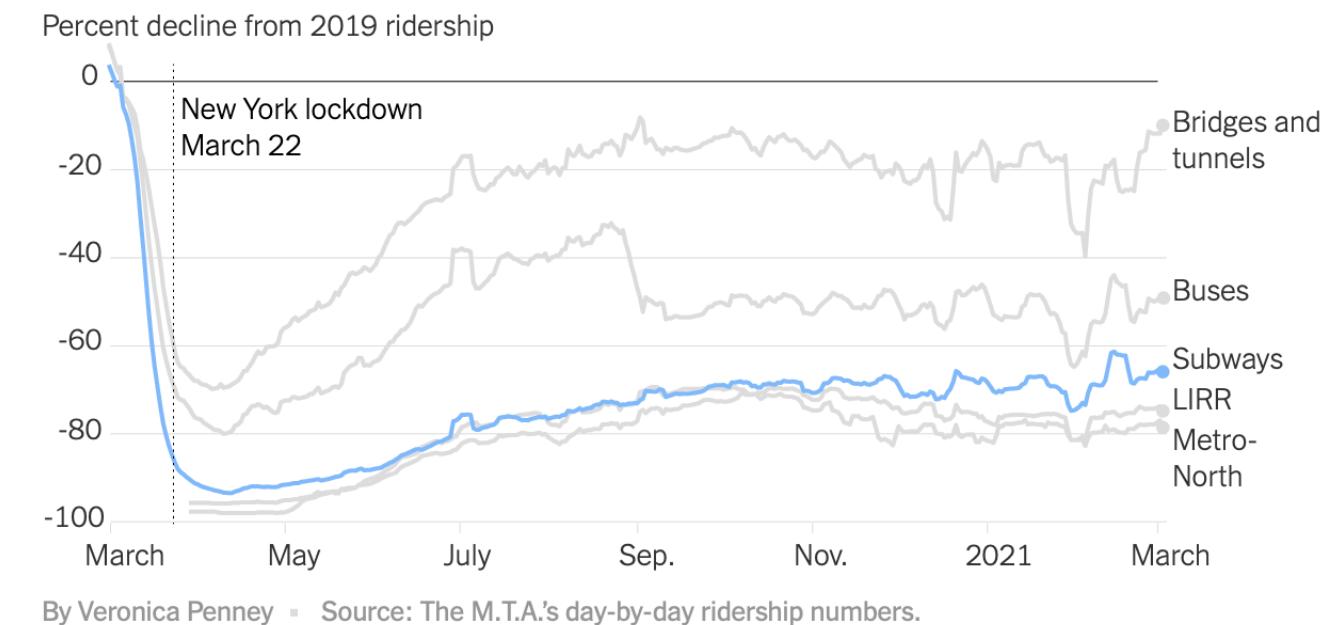


pacing for attention

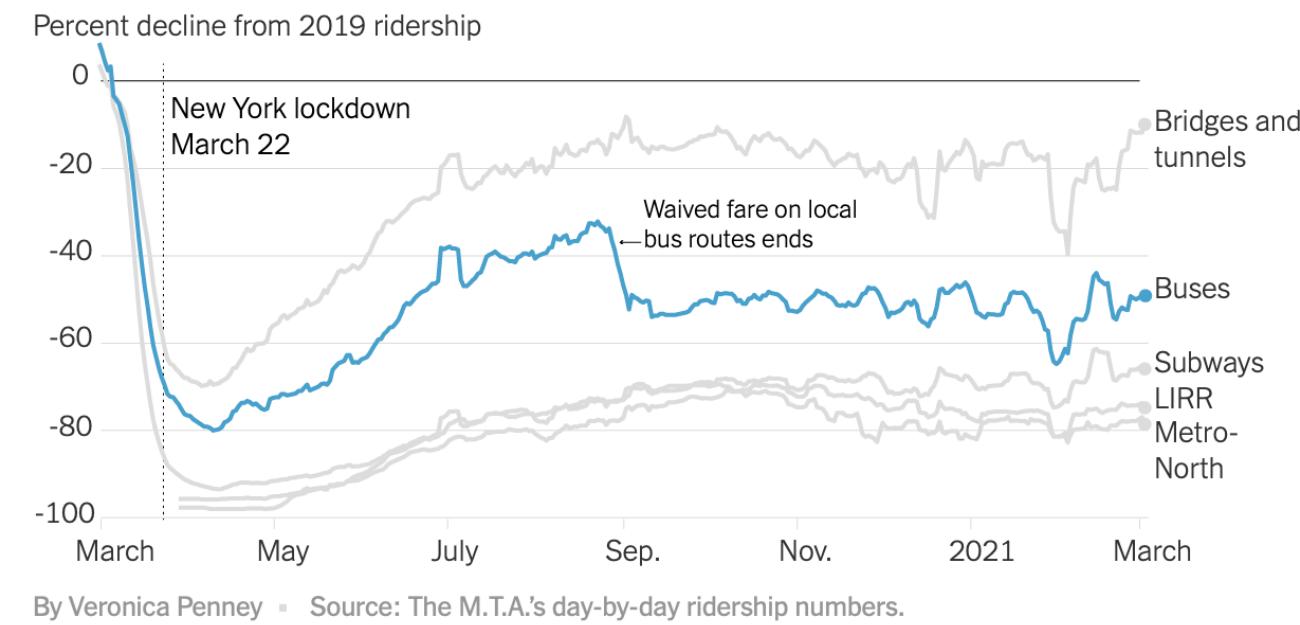
pacing for attention, you can focus on consecutive layers of a sequence with *multiples*



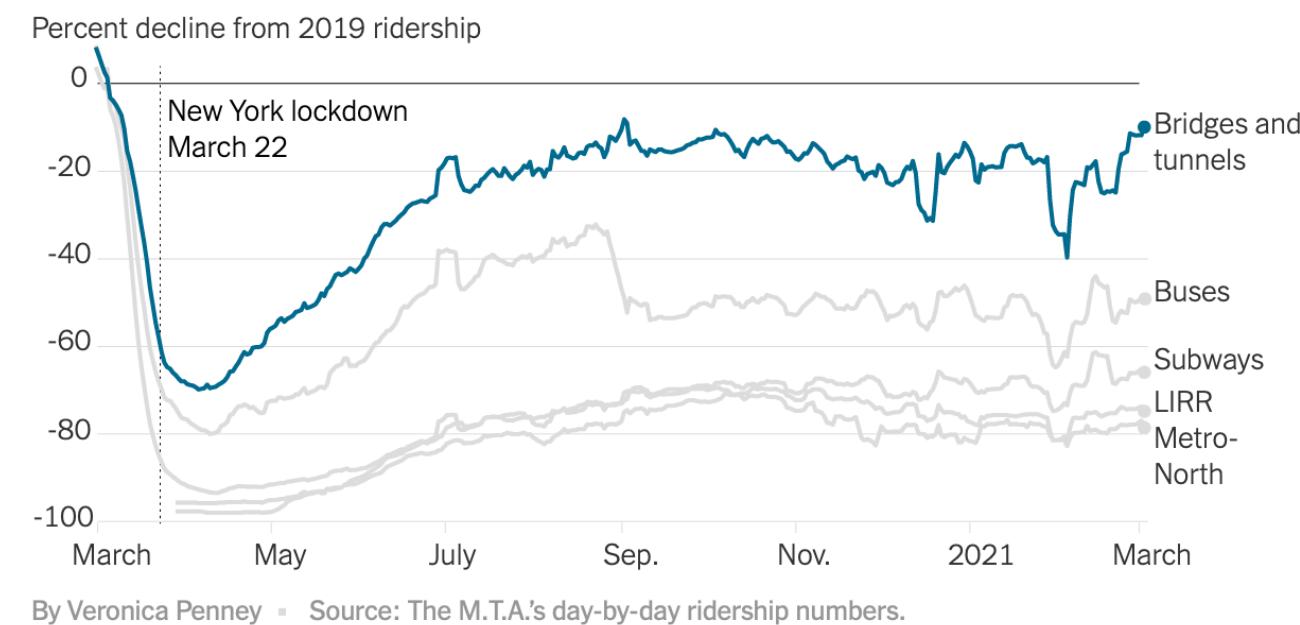
Subway Ridership Is Slow to Recover



The Pandemic Cut Bus Ridership by Half



Car Travel Is Near Pre-Pandemic Levels



Penney, Veronica. "How Coronavirus Has Changed New York City Transit, in One Chart" New York Times, March 8, 2021, Climate sec. <https://www.nytimes.com/interactive/2021/03/08/climate/nyc-transit-covid.html>.

pacing for attention, you can also focus on consecutive layers of a graphic with *animation* — a grammar of animated graphics



A Grammar of Animated Graphics

gganimate 1.0.5.9000 [Home](#) Getting Started Reference Talks News [Search...](#) [GitHub](#)

Build up a plot, layer by layer

Source: R/transition-layers.R

This transition gradually adds layers to the plot in the order they have been defined. By default prior layers are kept for the remainder of the animation, but they can also be set to be removed as the next layer enters.

```
transition_layers(  
  layer_length = 1,  
  transition_length = 1,  
  keep_layers = TRUE,  
  from_blank = TRUE,  
  layer_order = NULL,  
  layer_names = NULL  
)
```

Contents

- Arguments
- Label variables
- Object permanence
- See also
- Examples

Arguments

layer_length The proportional time to pause at each layer before a new one enters

transition_length The proportional time to use for the entrance of a new layer

keep_layers Either an integer indicating for how many following layers the layers should stay on screen or a logical. In the case of the latter, `TRUE` will mean keep the layer for the remainder of the animation (equivalent to setting it to `Inf`) and `FALSE` will mean to transition the layer out as the next layer enters.

from_blank Should the first layer transition in or be present on the onset of the animation

layer_order An alternative order the layers should appear in (default to using the stacking order). All other arguments that references the layers index in some way refers to this order.

layer_names A character vector of names for each layers, to be used when interpreting label literals

Label variables

`transition_layers` makes the following variables available for string literal interpretation, in addition to the general ones provided by `animate()`:

- **transitioning** is a boolean indicating whether the frame is part of the transitioning phase
- **previous_layer** The name of the last layer the animation was showing
- **closest_layer** The name of the layer the animation is closest to showing
- **next_layer** The name of the next layer the animation will show
- **nlayers** The total number of layers

Object permanence

`transition_layer` does not link rows across data to the same graphic element, so elements will be defined uniquely by each row and the enter and exit of the layer it belongs to.

Pedersen, Thomas Lin, and David Robinson. “Gganimate: A Grammar of Animated Graphics.” Manual, 2021. <https://gganimate.com>.

resources

References

Spencer, Scott. Sec. 2.3 In *Data in Wonderland*. 2021. https://ssp3nc3r.github.io/data_in_wonderland.

Brath, Richard, and Ebad Banissi. "Evaluation of Visualization by Critiques." In *Proceedings of the Beyond Time and Errors on Novel Evaluation Methods for Visualization - BELIV '16*, 19–26. Baltimore, MD, USA: ACM Press, 2016. <https://doi.org/10.1145/2993901.2993904>.

Chu, Tony. 2016. "Animation, Pacing, and Exposition." OpenVis Conf 2016. <https://www.youtube.com/watch?v=Z4tB6qyxHJA>.

Correll, Michael, Dominik Moritz, and Jeffrey Heer. 2018. "Value-Suppressing Uncertainty Palettes." In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–11. Montreal QC, Canada: ACM Press. <https://dl.acm.org.ezproxy.cul.columbia.edu/doi/10.1145/3173574.3174216>.

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

Kay, Matthew. 2020. *ggdist: Visualizations of Distributions and Uncertainty*. R package version 2.4.0, <https://mjskay.github.io/ggdist/>.

Kosara, Robert. "Visualization Criticism - The Missing Link Between Information Visualization and Art." In *2007 11th International Conference Information Visualization (IV '07)*, 631–36. Zurich, Switzerland: IEEE, 2007. <https://doi.org/10.1109/IV.2007.130>.

Lupi, Giorgia. "The Architecture of a Data Visualization: Multilayered Storytelling through 'Info-Spatial' Compositions," February 2015. <https://medium.com/accurate-studio/the-architecture-of-a-data-visualization-470b807799b4>.

Pedersen, Thomas Lin, and David Robinson. 2021. *Gganimate: A Grammar of Animated Graphics*. Manual, <https://gganimate.com>.

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

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

supplemental material

criticism for visuals, practicing critiques

Audience?

Purpose?

Data encodings, decodings?

Comparison or change?

Narrative?

Color, coherency?

Hierarchy, layering, layout?

Credibility, transparency?

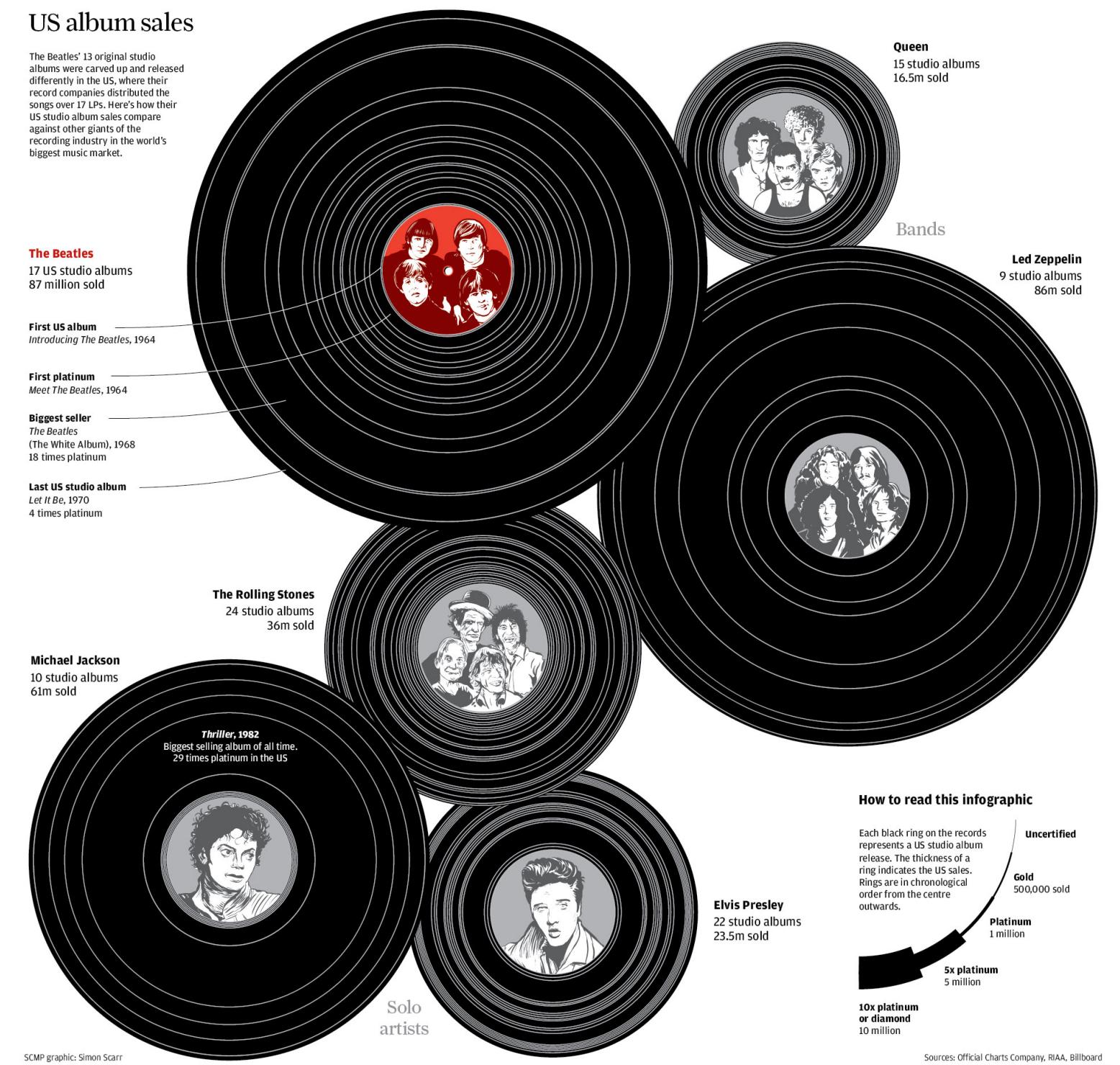
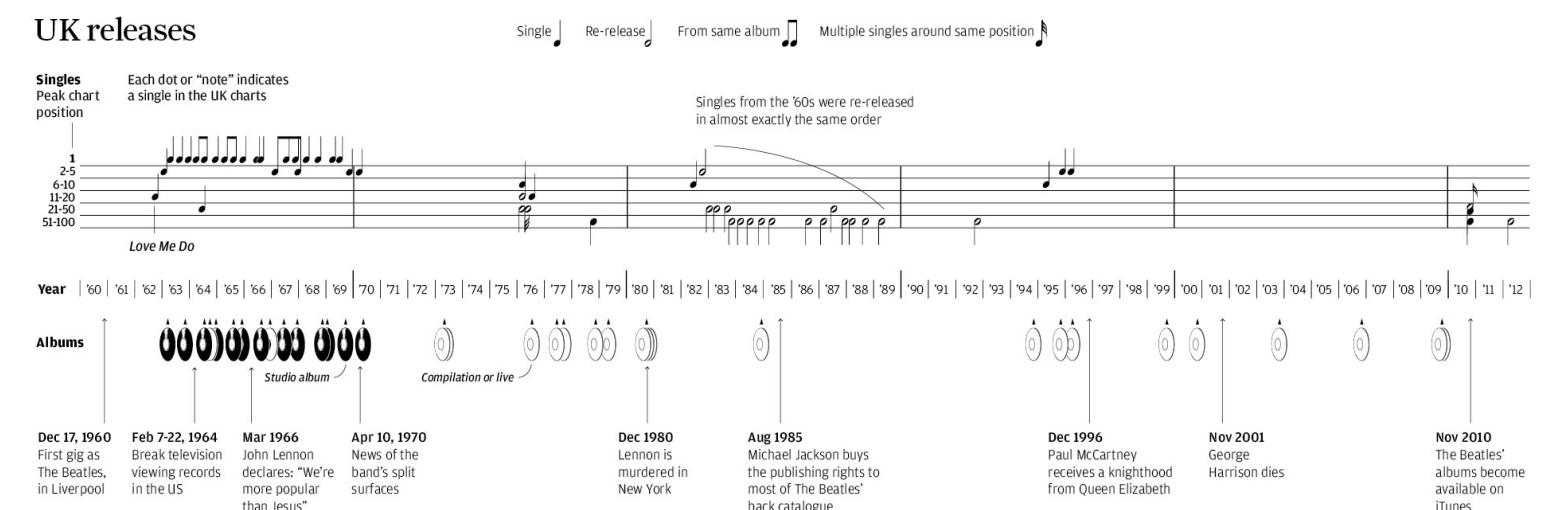
Scarr, Simon. "Love Me Do." South China Morning Post, October 13, 2012, sec. Infographics. <https://multimedia scmp.com/culture/article/SCMP-printed-graphics-memory/lonelyGraphics/201210A114.html>.

A18 Saturday, October 13, 2012

South China Morning Post

Love Me Do

Fifty years ago today The Beatles made their chart debut with their first single, *Love Me Do*. They would go on to have 17 number one singles in their native Britain and sell over 80 million albums in the United States, all the while changing the face of popular culture.



criticism for visuals, practicing critiques

Audience?

Purpose?

Data encodings, decodings?

Comparison or change?

Narrative?

Color, coherency?

Hierarchy, layering, layout?

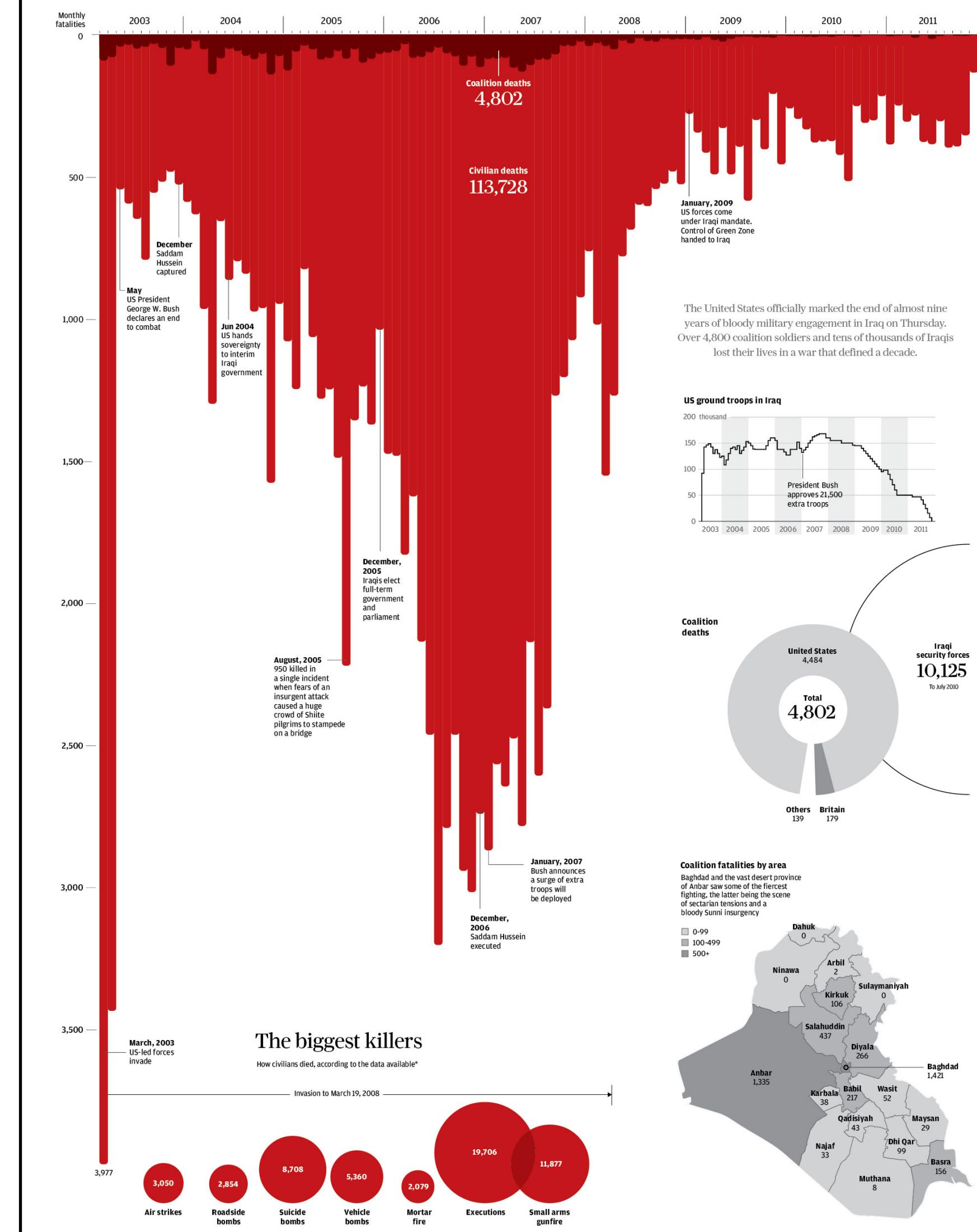
Credibility, transparency?

Scarr, Simon. "Iraq's bloody toll." South China Morning Post, December 17, 2011, sec. Infographics. <https://multimedia scmp.com/culture/article/SCMP-printed-graphics-memory/lonelyGraphics/201112A131.html>.

A12 Saturday, December 17, 2011

South China Morning Post

Iraq's bloody toll



criticism for visuals, practicing critiques

Audience?

Purpose?

Data encodings, decodings?

Comparison or change?

Narrative?

Color, coherency?

Hierarchy, layering, layout?

Credibility, transparency?

López, Alberto Lucas. "Infographic: The World Cup's Great Divide." South China Morning Post, June 26, 2014, sec. Infographics. <https://www.scmp.com/infographics/article/1540818/world-cups-great-divide>.

A14 Thursday, June 26, 2014

South China Morning Post

How to read this infographic

Each square represents HK\$10 million market value. The 16 teams with the highest market value playing at the World Cup. Market value of the highest paid player for the top 16 teams.

The 16 teams with the lowest market value playing at the World Cup.

Player (club where he plays) Personal market value

Percentage approx. on the total market value of his national team

National football team (low to high market value)

HK\$ millions

Brazil 5,713

Neymar (FC Barcelona) 710

12%

Spain 5,324

Juan Mata (Chelsea) 496

9%

Argentina 5,020

Lionel Messi (FC Barcelona) 1,455

29%

Germany 4,917

Mario Götze (Bayern Munich) 635

13%

France 3,944

Karim Benzema (Real Madrid) 341

9%

Belgium 3,376

Eden Hazard (Chelsea) 480

14%

Italy 3,336

Claudio Marchisio (Juventus) 308

9%

England 3,296

Wayne Rooney (Manchester United) 443

13%

Portugal 3,271

Cristiano Ronaldo (Real Madrid) 1,131

34%

Colombia 2,309

Radamel Falcao (Monaco) 593

25%

Uruguay 1,956

Edinson Cavani (PSG) 607

31%

Russia 1,785

Alan Dzagoev (CSKA Moscow) 254

14%

Croatia 1,721

Luka Modrić (Real Madrid) 338

20%

Netherlands 1,707

Robin van Persie (Man United) 398

23%

Chile 1,430

Arturo Vidal (Juventus) 443

31%

Ivory Coast 1,363

Yaya Touré (Man City) 349

26%

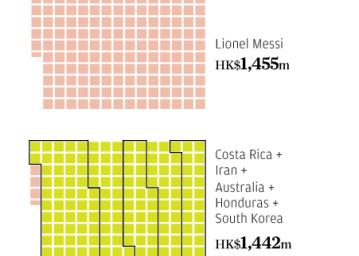
Costa Rica + Iran + Australia + Honduras + South Korea HK\$1,442m

The World Cup's great divide

While you can't put a price on a fan's passion for the beautiful game, we do know how much clubs are willing to pay for players. By combining the value of individual footballers, we have estimated the market value of each of the 32 teams at the World Cup in Brazil.



Individual vs country
Lionel Messi is the world's most valuable player. Financially, the market value of the team of Argentina's worth alone could finance the five teams with the lowest market value. None of the 16 lowest-ranked teams have players among the top 100 highest paid.



SCMP graphic: Alberto Lucas López

Sources: PLURB Consultancy report "Os 100 jogadores mais valiosos do mundo em 2013" (The 100 most valuable players in the world in 2013), PLURB Consultancy report "Valor de Mercado das 32 Seleções que disputarão a Copa do Mundo Brasil 2014" (Market value of the 32 teams at the World Cup in Brazil), Football Finance.

The data provided in the reports are in euro.
Conversion value at time of infographics: 1 EUR = HK\$10.5412

Note: the report about the market value of the players was published on January 14, 2014. Data and their clubs refers to that time.

Printed and published by South China Morning Post Publishers Ltd, Morning Post Centre, 22 Des Voeux Street, Tai Po Industrial Estate, Tai Po, Hong Kong. Tel: 2660 8888.

criticism for visuals, practicing critiques

Audience?

Purpose?

Data encodings, decodings?

Comparison or change?

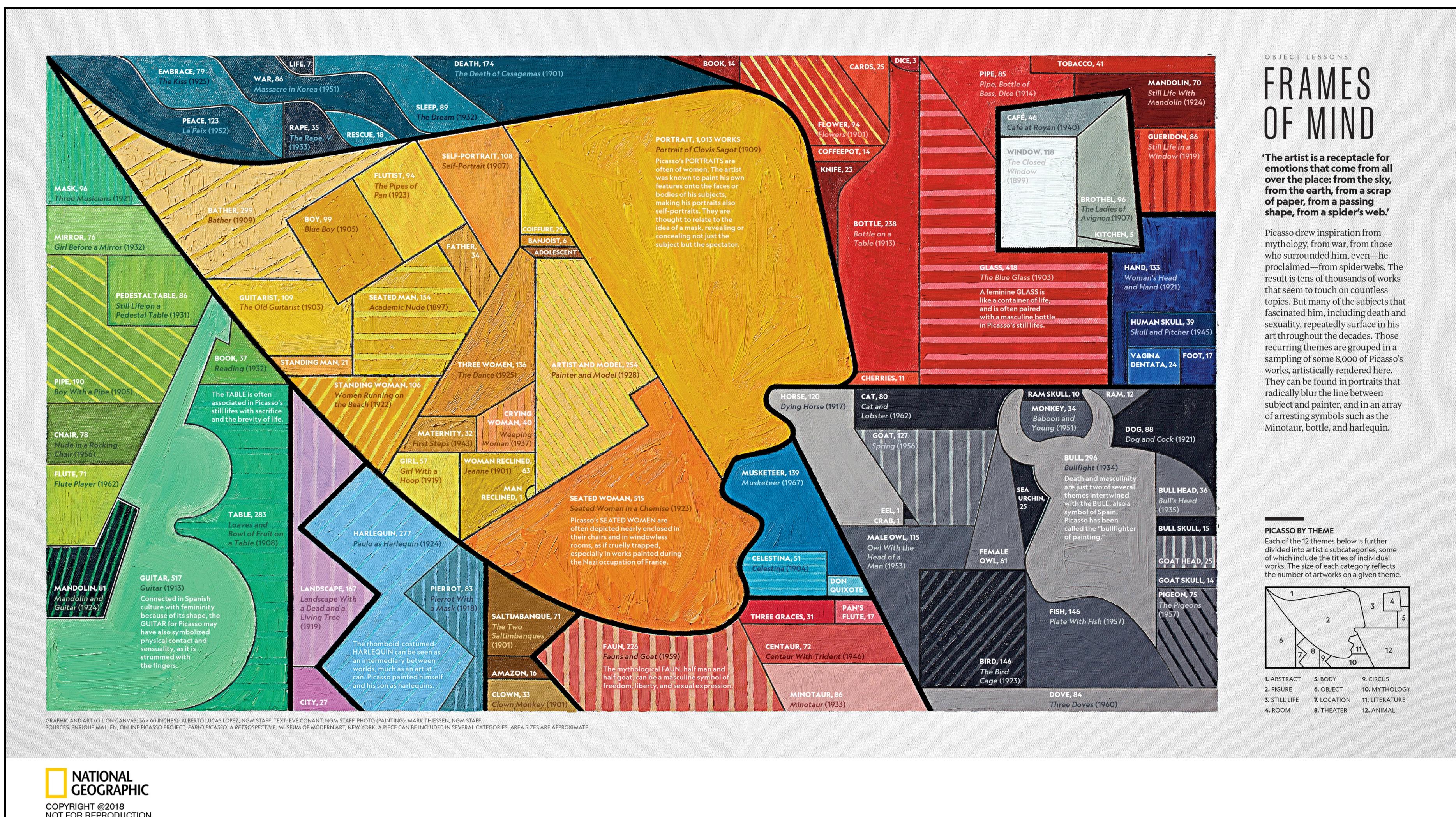
Narrative?

Color, coherency?

Hierarchy, layering, layout?

Credibility, transparency?

Kalb, Claudia, Paolo Woods, and Gabriele Galimberti.
"Intense, Provocative, Disturbing, Captivating, Genius, Picasso." National Geographic Magazine, May 2018.
National Geographic Archive 1995+.



criticism for visuals, practicing critiques

Audience?

Purpose?

Data encodings, decodings?

Comparison or change?

Narrative?

Color, coherency?

Hierarchy, layering, layout?

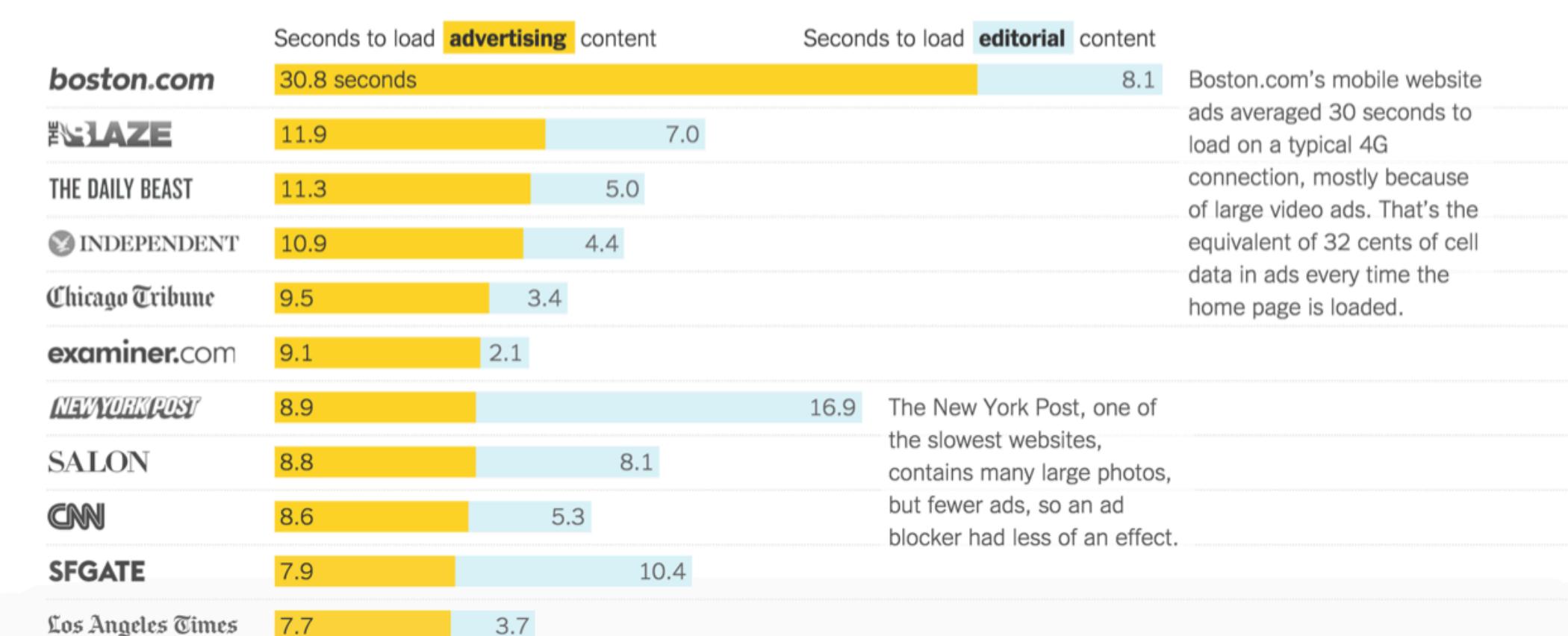
Credibility, transparency?

Aisch, Gregor, Wilson Andrews, and Josh Keller. "The Cost of Mobile Ads on 50 News Websites." New York Times, October 1, 2015, Online edition, sec. Business. <https://www.nytimes.com/interactive/2015/10/01/business/cost-of-mobile-ads.html>.

The Cost of Mobile Ads on 50 News Websites

By GREGOR AISCH, WILSON ANDREWS and JOSH KELLER OCT. 1, 2015

Ad blockers, which Apple first allowed on the iPhone in September, promise to conserve data and make websites load faster. But how much of your mobile data comes from advertising? We measured the mix of advertising and editorial on the mobile home pages of the top 50 news websites – including ours – and found that **more than half of all data came from ads** and other content filtered by ad blockers. Not all of the news websites were equal. [RELATED ARTICLE](#)

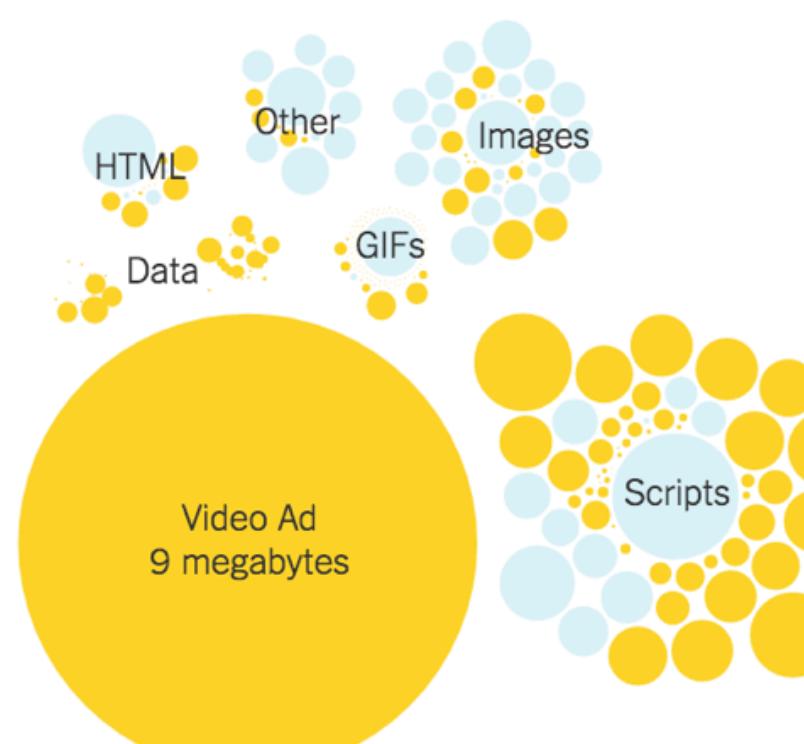


boston.com

Here are all the files that made up the Boston.com data during one visit, including one large video ad and many script files used by ad networks. With an ad blocker, those files were gone.

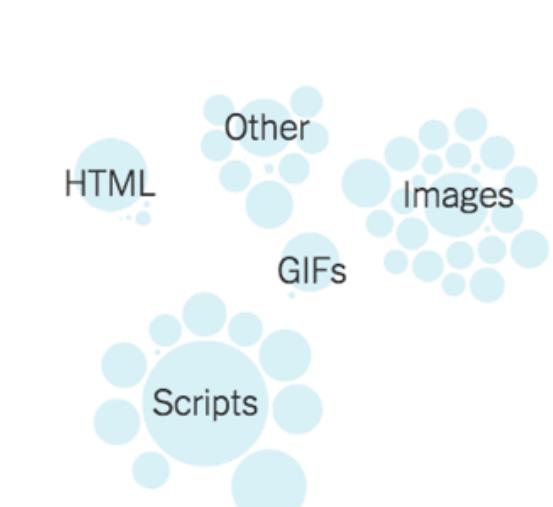
Without ad blocker

389 files, 16.3 megabytes, 33 seconds



With ad blocker

52 files, 3.5 megabytes, 7 seconds



criticism for visuals, practicing critiques

Audience?

Purpose?

Data encodings, decodings?

Comparison or change?

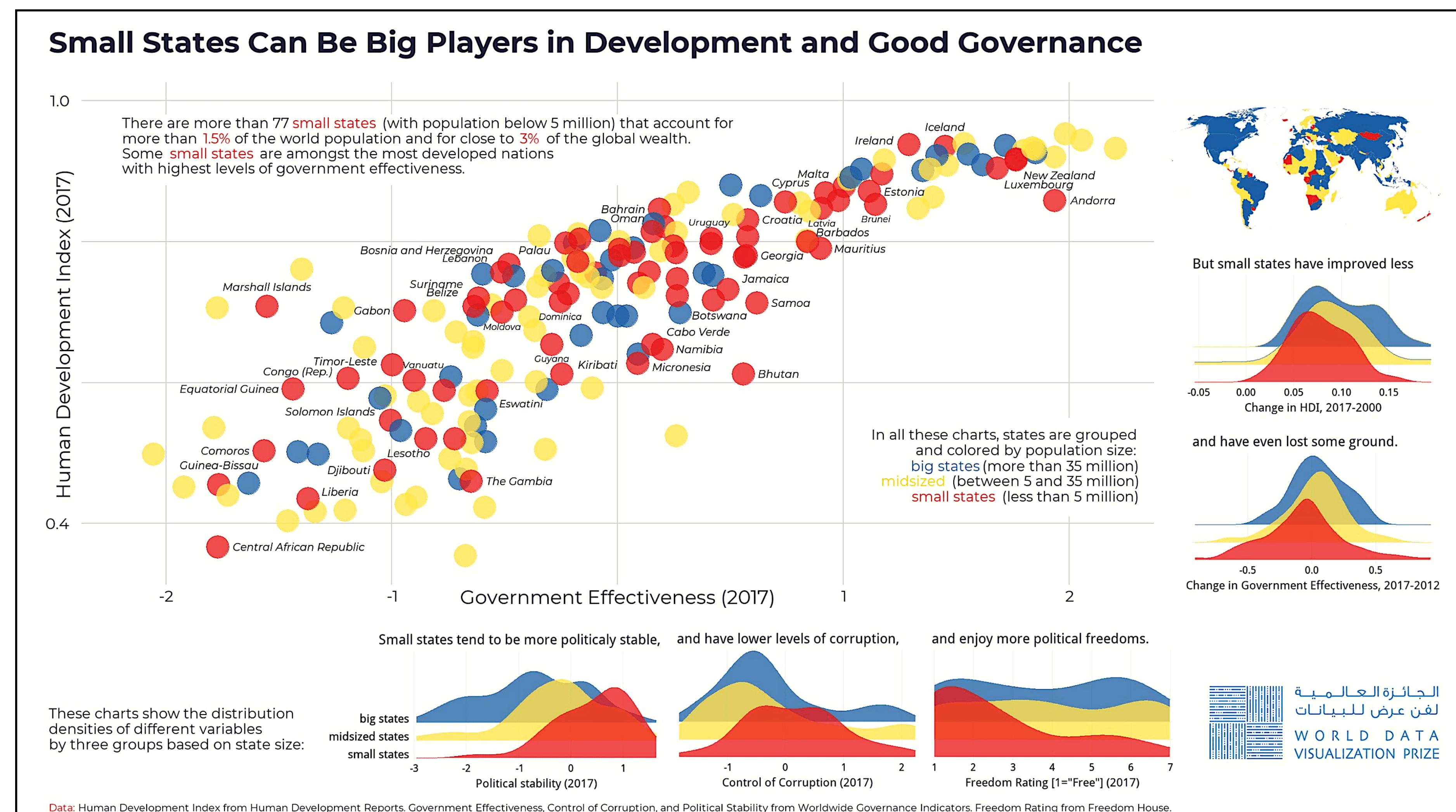
Narrative?

Color, coherency?

Hierarchy, layering, layout?

Credibility, transparency?

Dimitar Toshkov. "World Data Visualization Prize: Small States Can Be Big Players in Development and Good Governance." Personal. Dimitar Toshkov, 2019. http://www.dimitar.eu/Visualizations_files/WDVP.html.



criticism for visuals, practicing critiques

Audience?

Purpose?

Data encodings, decodings?

Comparison or change?

Narrative?

Color, coherency?

Hierarchy, layering, layout?

Credibility, transparency?

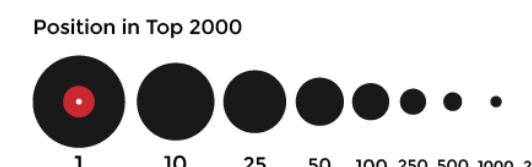
Bremer, Nadieh. "The Top 2000 Loves the 70s & 80s." Personal. Visual Cinnamon, December 2016. <https://www.visualcinnamon.com/portfolio/top2000>.

TOP 2000 ❤️ 70's & 80's

Since 1999 the 2000 most popular songs of all time, as voted by the show's audience, are played on Dutch national Radio 2 in a yearly marathon. The 2000 songs are on the air between noon on December 25th until New Year's Eve and over half of the Dutch population listens to the Top 2000 each year.

Each ● to the right represents a song in the Top 2000. It is placed according to its year of release. In the legend below you can see what the size and color of a song means.

The bulk of the songs and most of the top 10 are from the 70's & 80's...



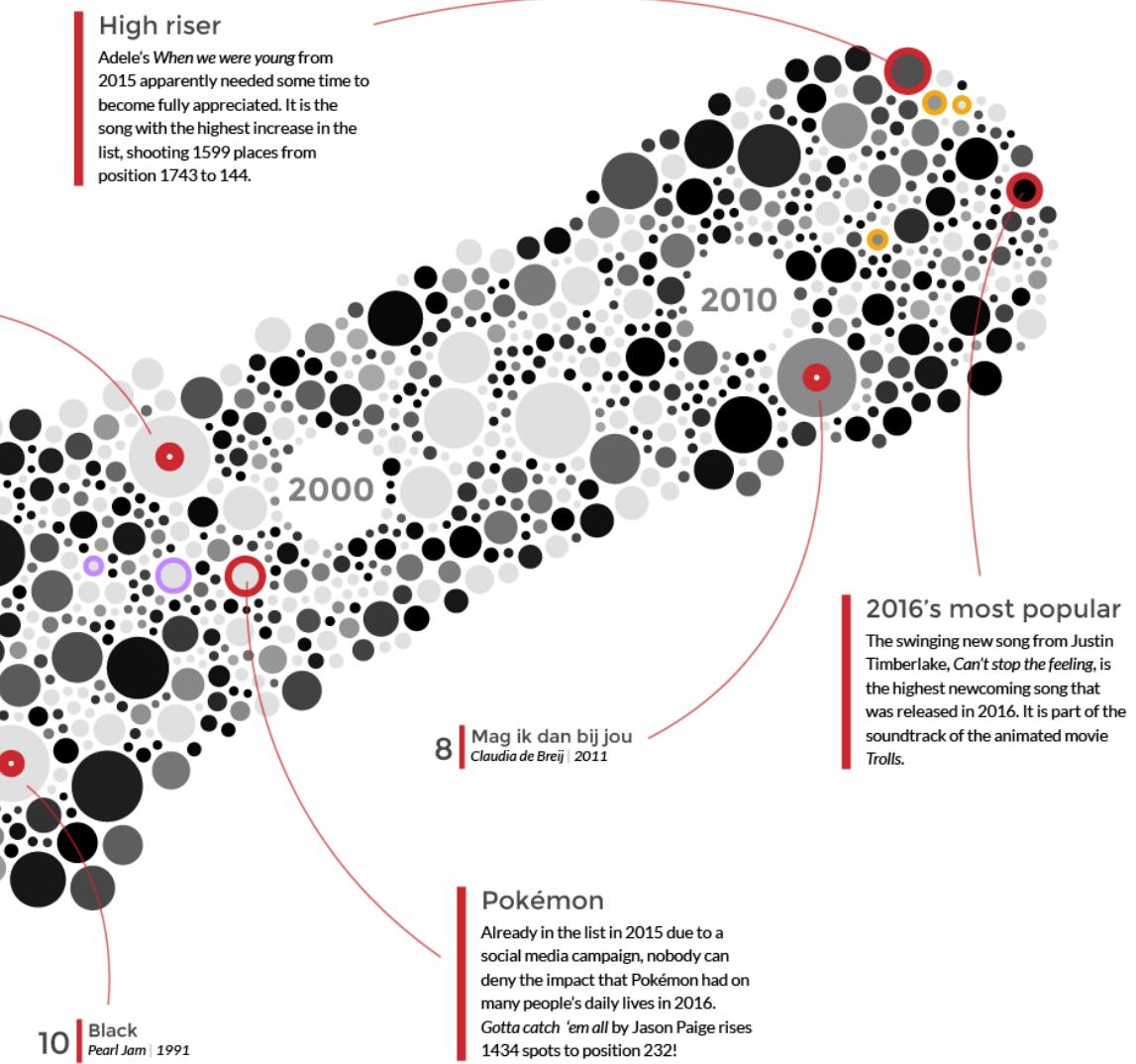
The charts on the right represent all 2000 songs from 3 past editions of the Top 2000 (held in 2000, 2005, 2010) and the most recent 2016 edition.

The songs are stacked according to their year of release. The higher a rectangle, the more songs that are in the Top 2000 list from that release year.

The black dotted line represents a smoothed curve over all 2000 songs. This makes the comparison between the 4 charts easier.

Created by Nadieh Bremer | VisualCinnamon.com for the December edition of data sketches

Visit tinyurl.com/2016top2000 for the interactive visual and see the name & title of each song



6

Avond

Boudewijn de Groot | 1997

2000

1990

1980

1970

2010

10

Black

Pearl Jam | 1991

8

Mag ik dan bij jou

Claudia de Breij | 2011

2016's most popular

The swinging new song from Justin Timberlake, 'Can't stop the feeling', is the highest newcomer song that was released in 2016. It is part of the soundtrack of the animated movie Trolls.

1

Bohemian Rhapsody

Queen | 1975

3

Stairway to Heaven

Led Zeppelin | 1971

2

Hotel California

Eagles | 1977

7

Heroes

David Bowie | 1977

9

Wish you were here

Pink Floyd | 1975

4

Piano Man

Billy Joel | 1974

5

Child in Time

Deep Purple | 1972

The Beatles

No other artist or band has more songs in the Top 2000 as the Beatles. With 38 songs they are responsible for 14% of all titles before 1970. Nonetheless, only 5 years ago they still had 50 songs in the list.

2000

2005

2010

2016

But they're losing tracks
to the new Millennium

It makes sense that the Top 2000 will be more spread out for each new edition, since there are more songs to choose from. However, if we compare the distributions of the Top 2000 songs over 4 editions, we see that, especially, the 90's has been gaining a lot of popularity.

Even though all songs from the 90's were out in the 2000 edition, only a few songs from that decade were chosen. Whereas in the 2016 edition the number of songs from the 90's has risen significantly. This could be due to a new generation who has grown up during the 90's taking over from those who voted in the early 2000's (who apparently didn't appreciate the new music).

Data | Top 2000 list from Radio 2 | Top 40 info from Mediamarkt's Top 40

criticism for visuals, practicing critiques

Audience?

Purpose?

Data encodings, decodings?

Comparison or change?

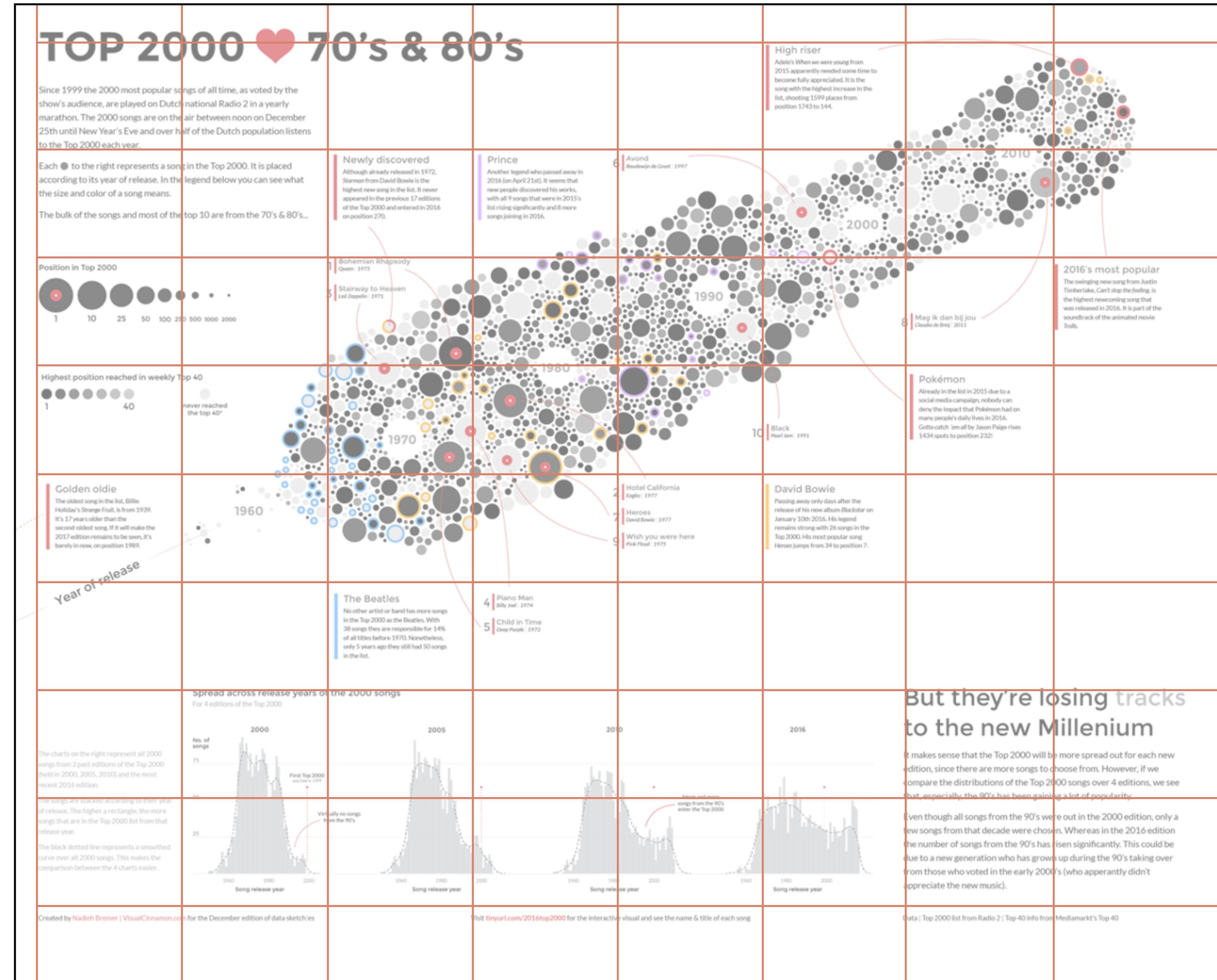
Narrative?

Color, coherency?

Hierarchy, layering, layout?

Credibility, transparency?

Bremer, Nadieh. "The Top 2000 Loves the 70s & 80s." Personal. Visual Cinnamon, December 2016. <https://www.visualcinnamon.com/portfolio/top2000>.



criticism for visuals, practicing critiques

Audience?

Purpose?

Data encodings, decodings?

Comparison or change?

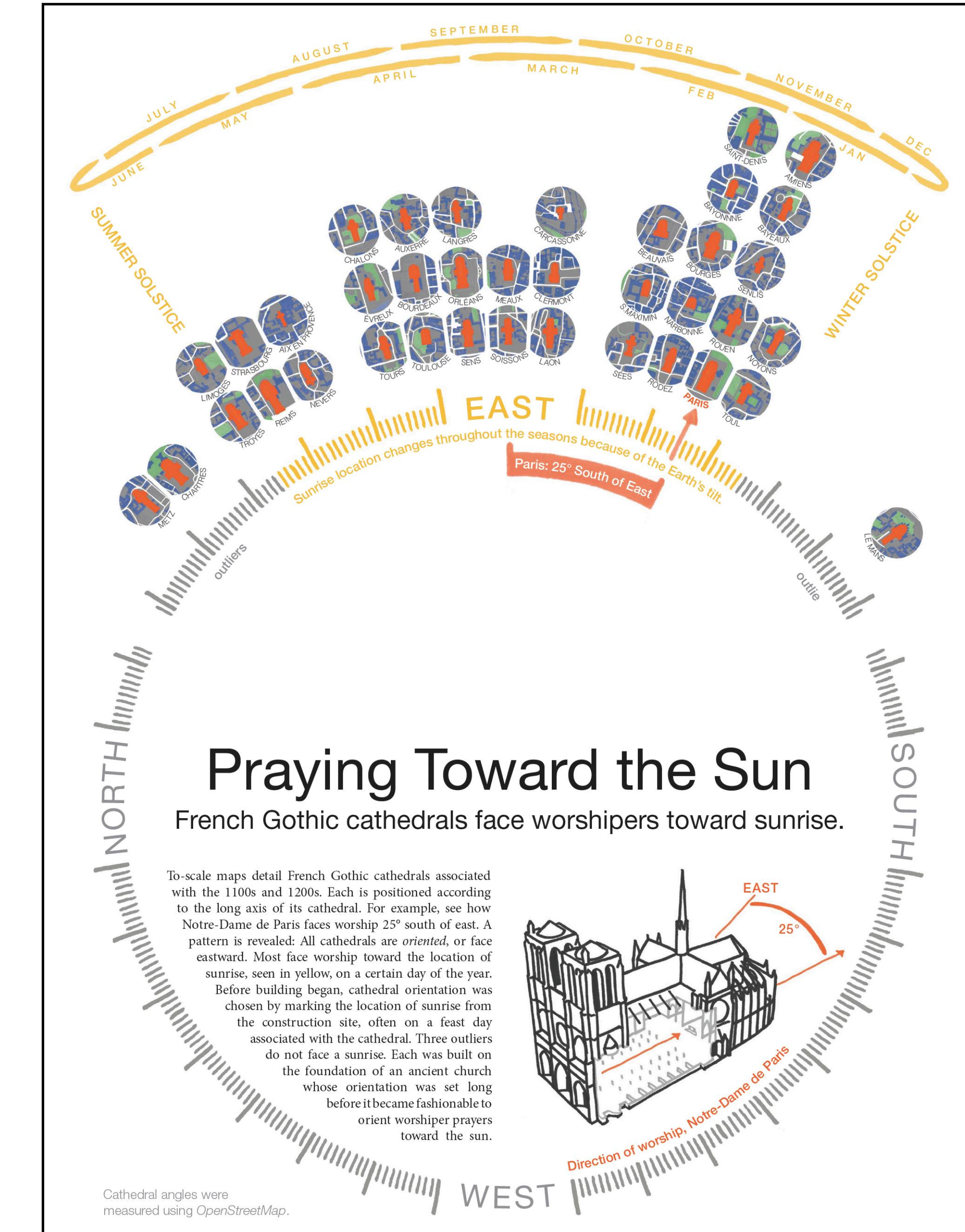
Narrative?

Color, coherency?

Hierarchy, layering, layout?

Credibility, transparency?

Andrews, R J. *Info We Trust: How to Inspire the World with Data*. Wiley, 2019.



criticism for visuals, practicing critiques

Audience?

Purpose?

Data encodings, decodings?

Comparison or change?

Narrative?

Color, coherency?

Hierarchy, layering, layout?

Credibility, transparency?

Lupi, Giorgia. "Visual Data - La Lettura," 2016. <http://giorgialupi.com/lalettura>

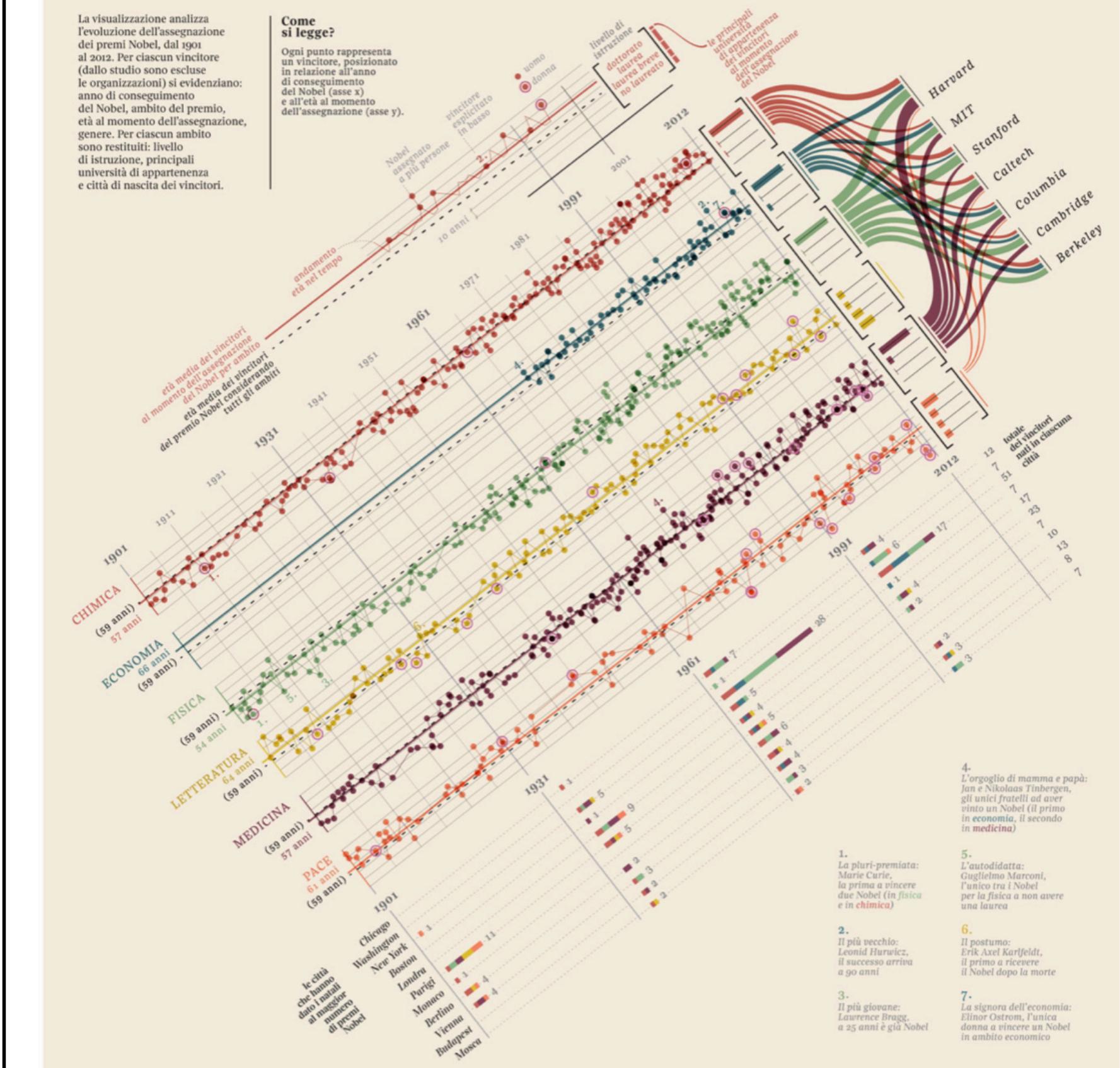
DOMENICA 25 NOVEMBRE 2012

Orizzonti Mappe

Visual data

I riconoscimenti vengono consegnati il 10 dicembre, anniversario della morte dell'ideatore del premio. Ecco l'origine, l'età e gli studi dei vincitori, dal 1901 a oggi.

Quanti (non) laureati al Nobel



di MARA GERGOLET

Che cosa dice la geografia del Nobel? Che mondo emerge quale battaglia dei sessi e di generazioni, quale primitivo e magica capitale del mondo. Ti accorgi che Berlino e Monaco da vent'anni non portano a casa un Nobel, quasi a distagione in chi ormai vede nella Germania una grande Svizzera, riassunta nella celebre battuta di Orson Welles: «In queste secoli, sono fondamentali a voler diventare scrittori, o voler cambiare i destini del mondo (le due categorie meno laureate). E c'è impossibile scrivere un libro, e lo fanno» e di Philip Roth, le manca il quadratino giallo del premio alla letteratura.

Sai premi in 112 edizioni tra chimica, economia, fisica (due portati a casa da Marie Curie). Archivio Summers, perché?

gli donne portano meno portate per le scienze. Si premi in 112 edizioni tra chimica, economia, fisica (due portati a casa da Marie Curie). Archivio Summers, perché?

C'è la classifica delle università. Dove dominano gli americani (in ordine, Harvard, MIT, Stanford, Caltech, Columbia, l'intrusiva Cambridge, Berkley). A ben guardare non l'America Ivy League dei campus edoardiani e dei viali bluferi di querce centenarie, ma quella verdisissima delle palazzine funzionali e degli skateboard della West coast. E allora, la geografia del genio ricorda quello del predominio politico e culturale? Non c'è dubbio. Però, benedetti gli «irregolari» della letteratura o del Nobel per la pace! C'è pur sempre lo scarto della fantasia, o dell'ossessione e della tenacia.

Gli autori
La visualizzazione e l'analisi dei dati sono a cura di Accurat (www.accurat.it), società di information design e consulenza progettuale diretta da Giorgia Lupi, Simone Quadrì, Gabriele Rossi.

© RIPRODUZIONE RISERVATA

criticism for visuals, practicing critiques

Audience?

Purpose?

Data encodings, decodings?

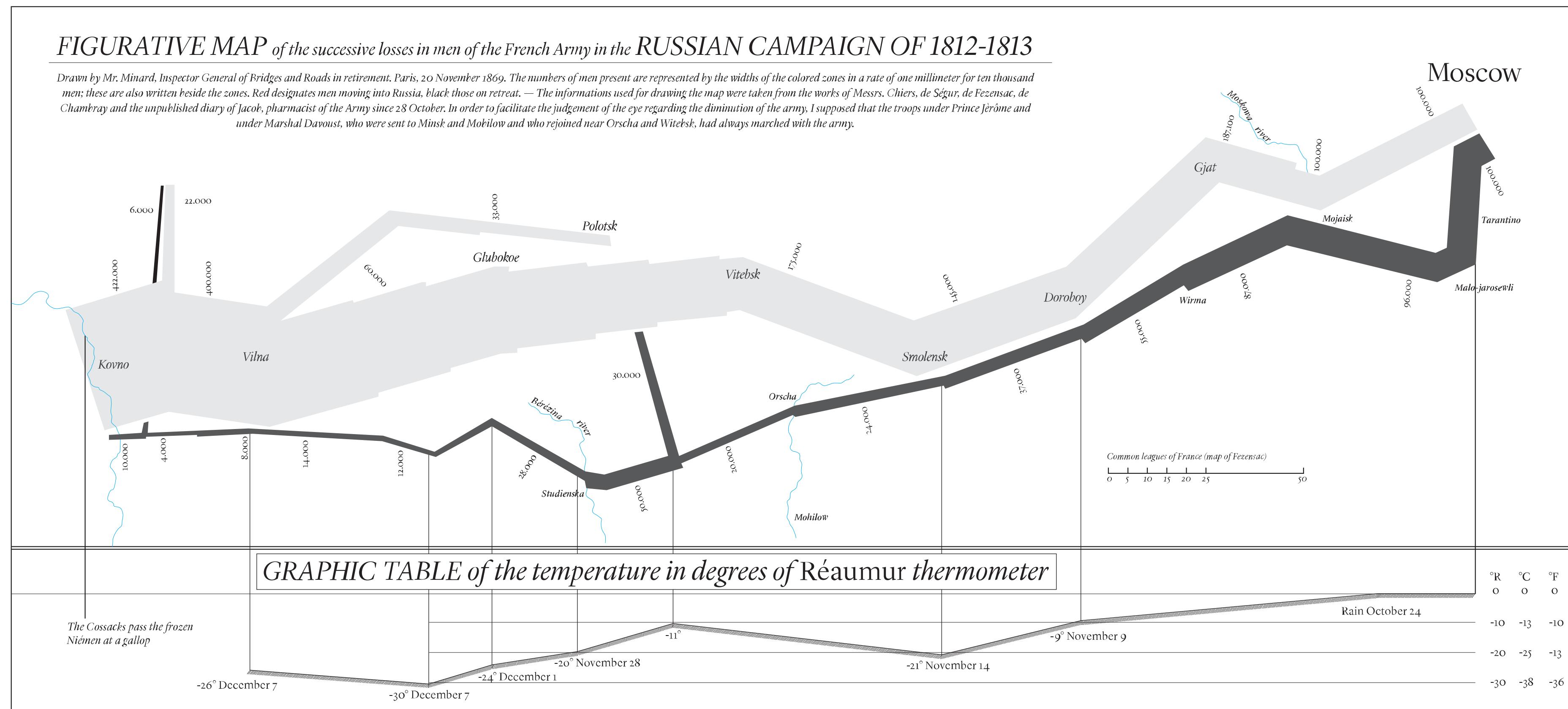
Comparison or change?

Narrative?

Color, coherency?

Hierarchy, layering, layout?

Credibility, transparency?



Commons, Wikimedia. "Redrawing of Minard's Napoleon Map," 2018. https://commons.wikimedia.org/wiki/File:Redrawing_of_Minard%27s_Napoleon_map.svg.

criticism for visuals, practicing critiques

Audience?

Purpose?

Data encodings, decodings?

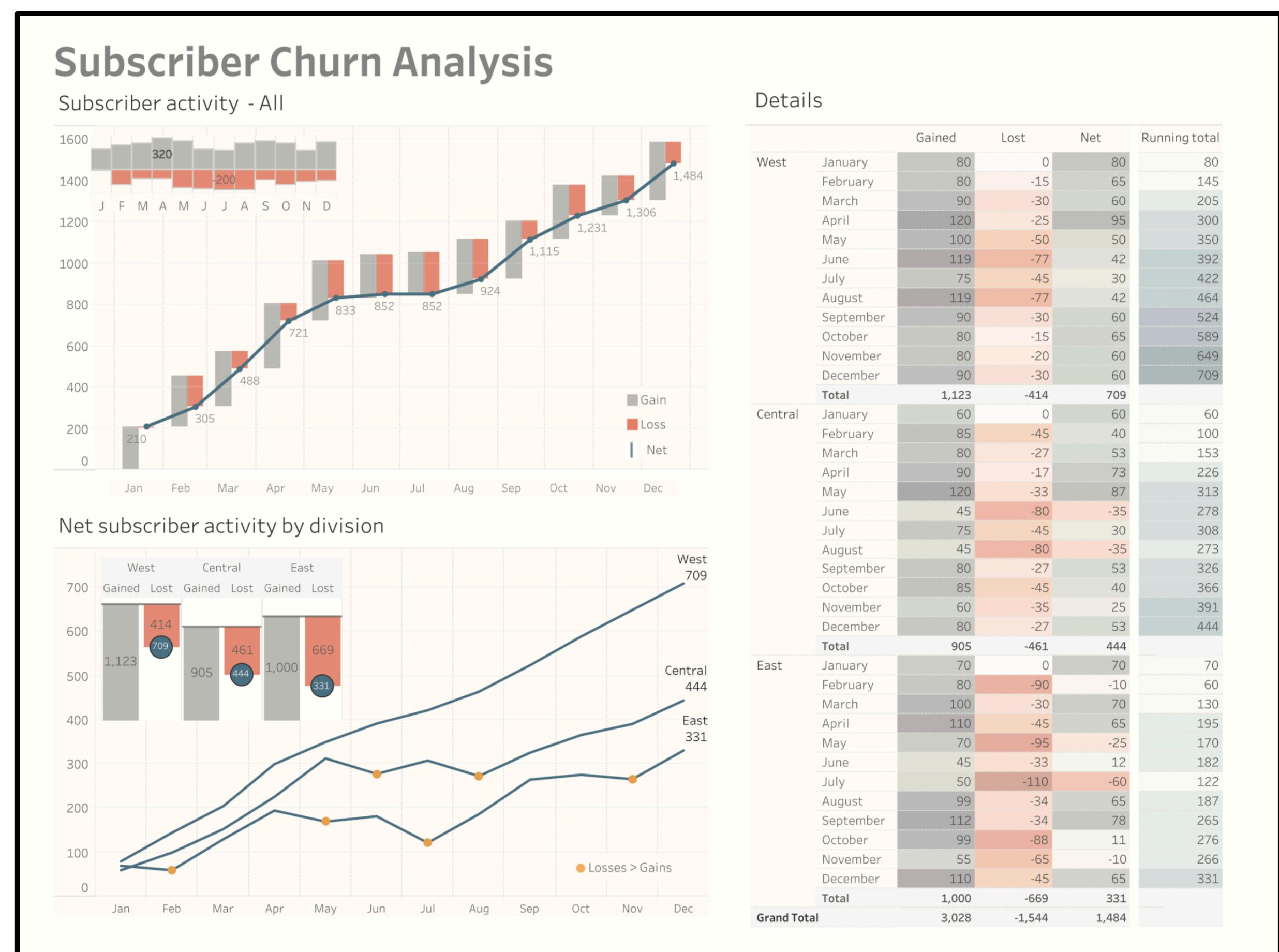
Comparison or change?

Narrative?

Color, coherency?

Hierarchy, layering, layout?

Credibility, transparency?



Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. *The Big Book of Dashboards: Visualizing Your Data Using Real-World Business Scenarios*. Hoboken, New Jersey: Wiley, 2017.

criticism for visuals, practicing critiques

Audience?

Purpose?

Data encodings, decodings?

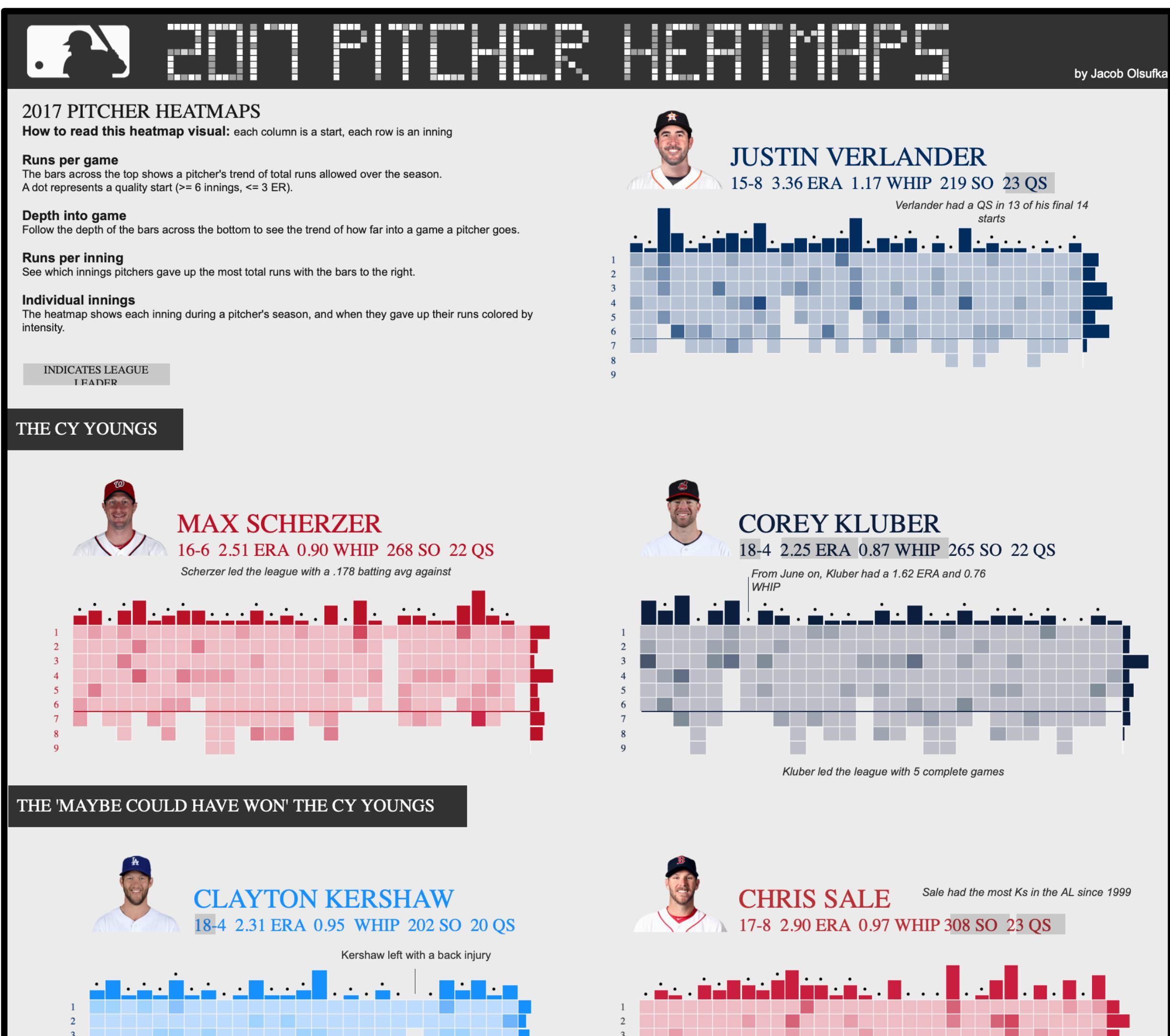
Comparison or change?

Narrative?

Color, coherency?

Hierarchy, layering, layout?

Credibility, transparency?



Olsufka, Jacob, 2017 MLB Pitcher Heatmaps. *Tableau Public*. https://public.tableau.com/views/2017MLBPitcherHeatmaps/MLB?:language=en&:display_count=y&:origin=viz_share_link

