Storytelling with data

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

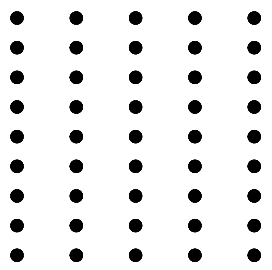
Scott Spencer | Columbia University

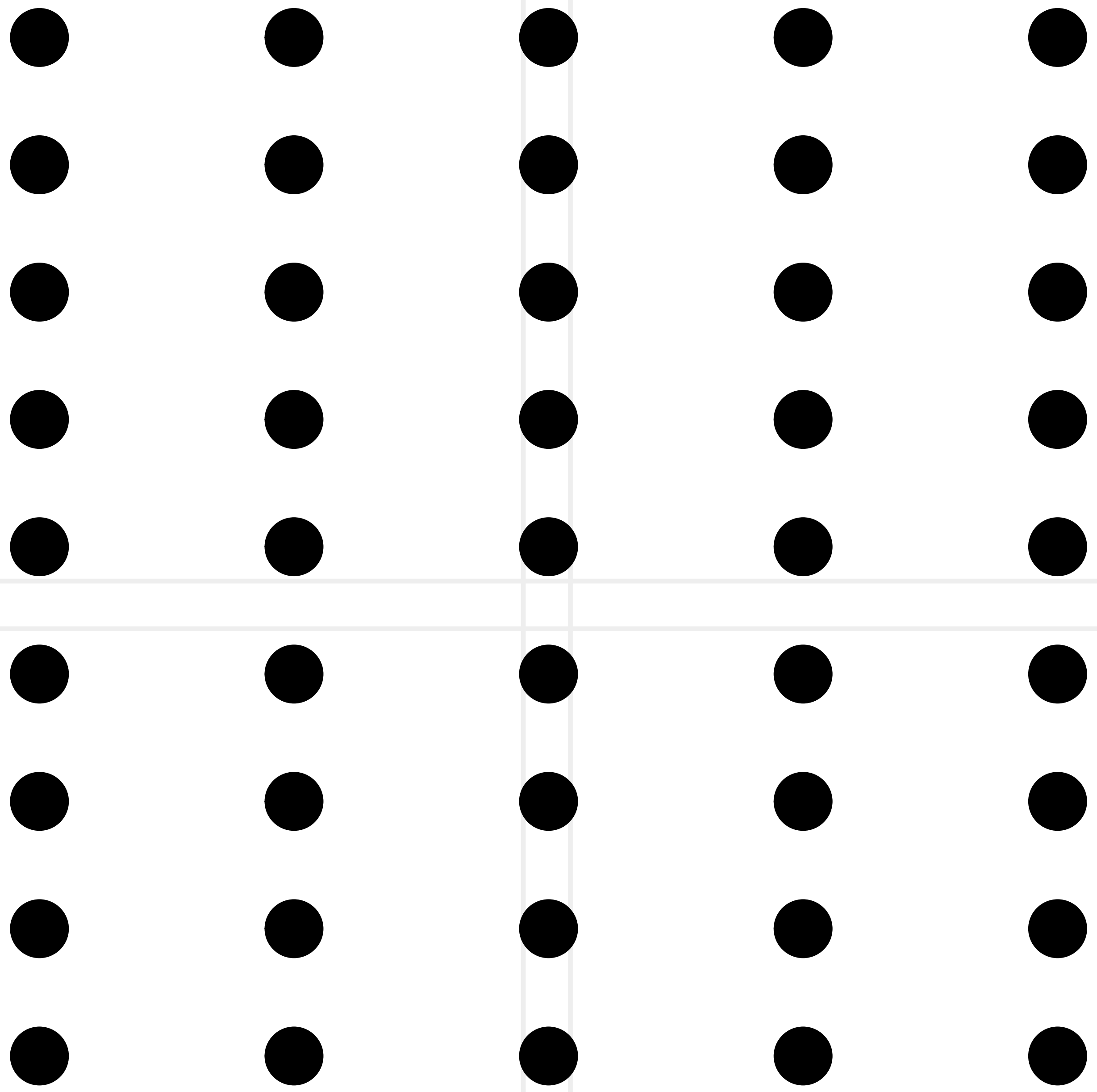

Write memo pitching idea to Chief Analytics Officer
Present critique of exemplary information graphic
Build interactive graphics for Chief Marketing Officer to explore project and insights
Write critique of peer presentation
FINAL
FINAL
DRAFT
FINAL
FINAL
DRAFT
FINAL
YOU ARE HERE
DRAFT
FINAL
DRAFT
FINAL
Ideate a data analytics project addressing problem or opportunity for your audience
Write a project proposal to Chief Analytics Officer
Create information graphics of project and insights for an External Audience
Present project and insights to Chief Executive Officer
CONDUCT DATA ANALYSIS

design mini-review

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

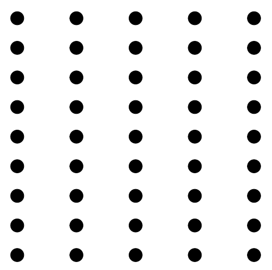
Proximity



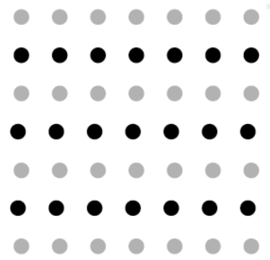


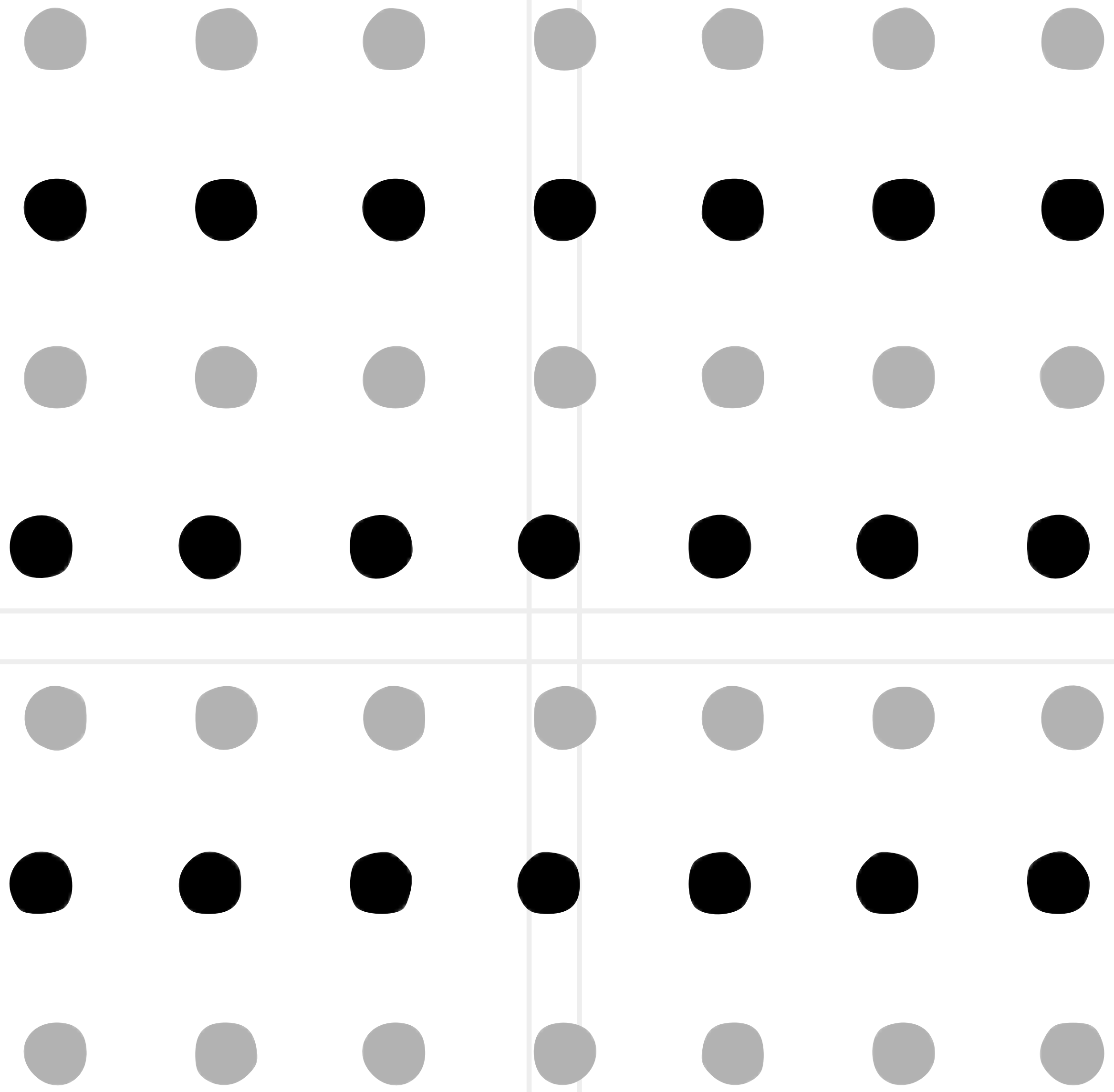
design mini-review | aligning and organizing information reduces cognitive load — *similarity*

Proximity



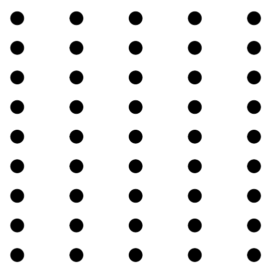
Similarity



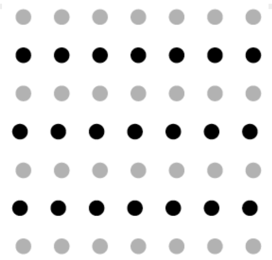


design mini-review | aligning and organizing information reduces cognitive load — *enclosure*

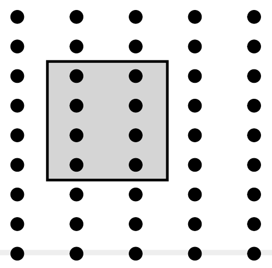
Proximity

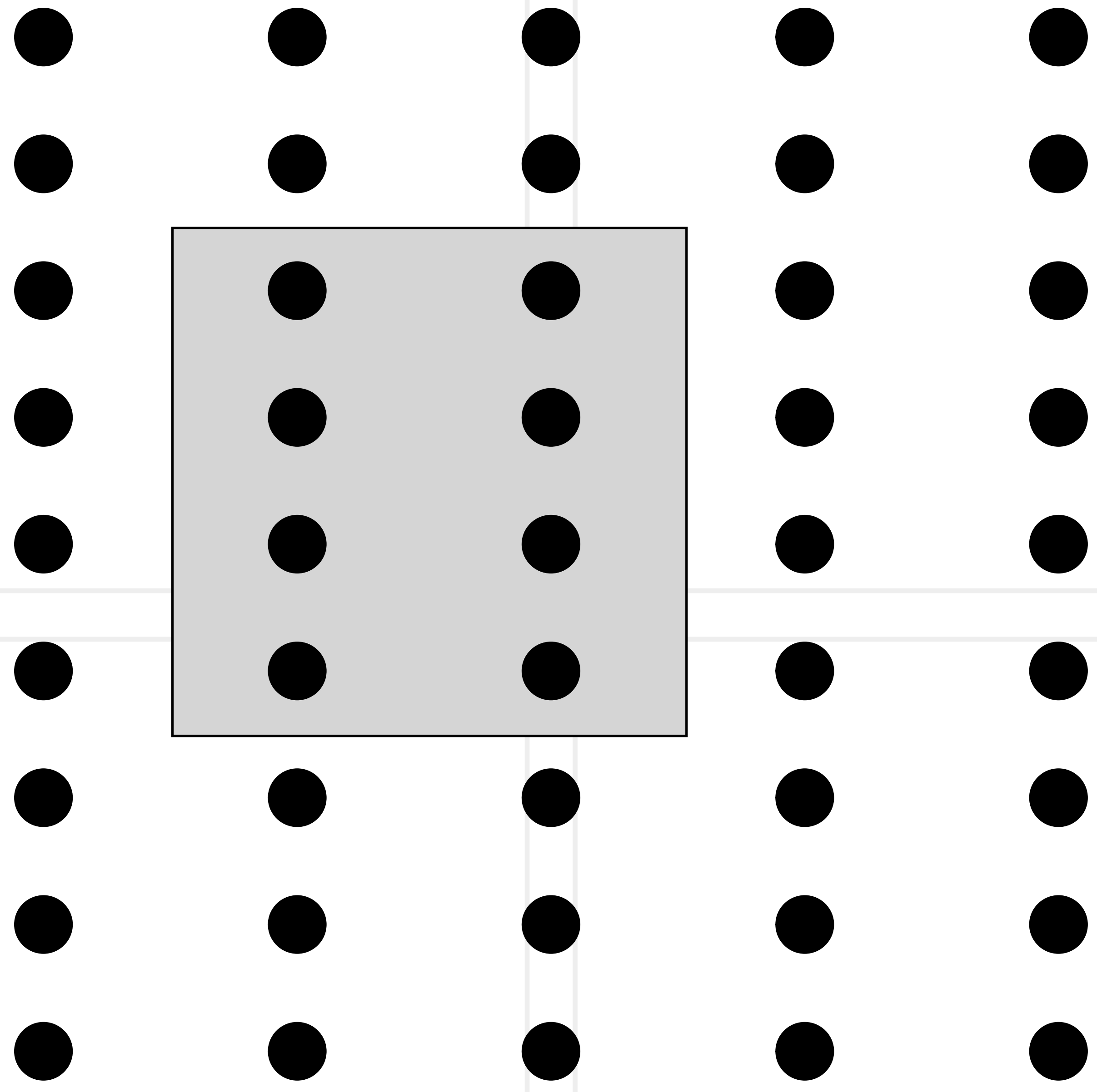


Similarity



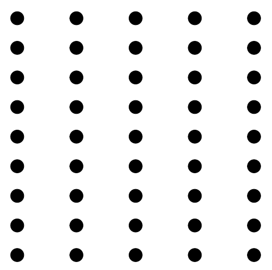
Enclosure



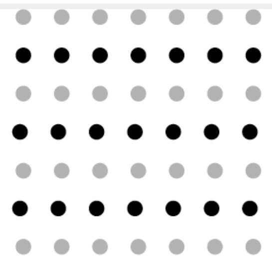


design mini-review | aligning and organizing information reduces cognitive load — *closure*

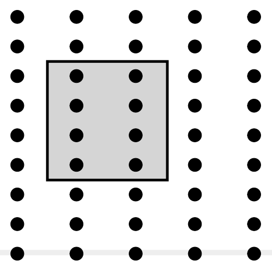
Proximity



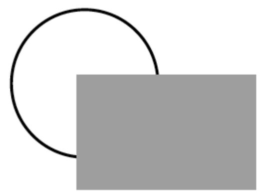
Similarity

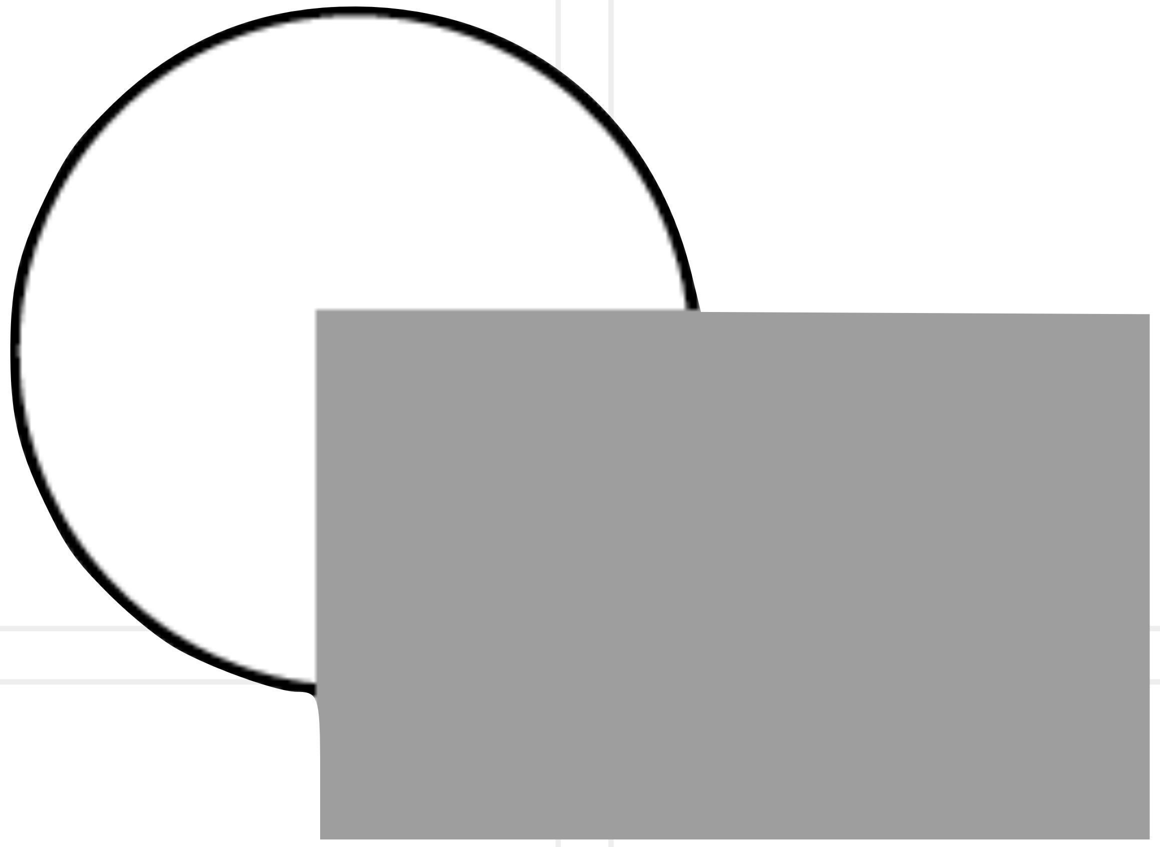


Enclosure



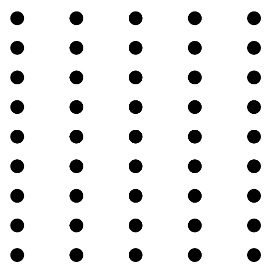
Closure



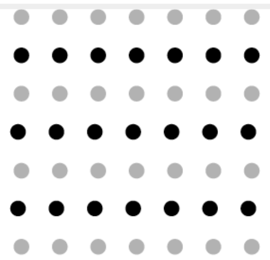


 aligning and organizing information reduces cognitive load — ***continuity***

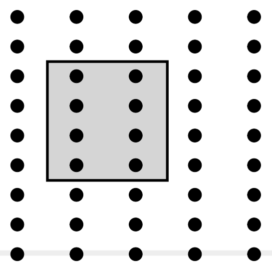
Proximity



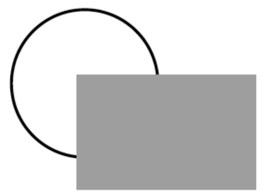
Similarity



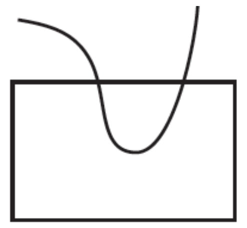
Enclosure

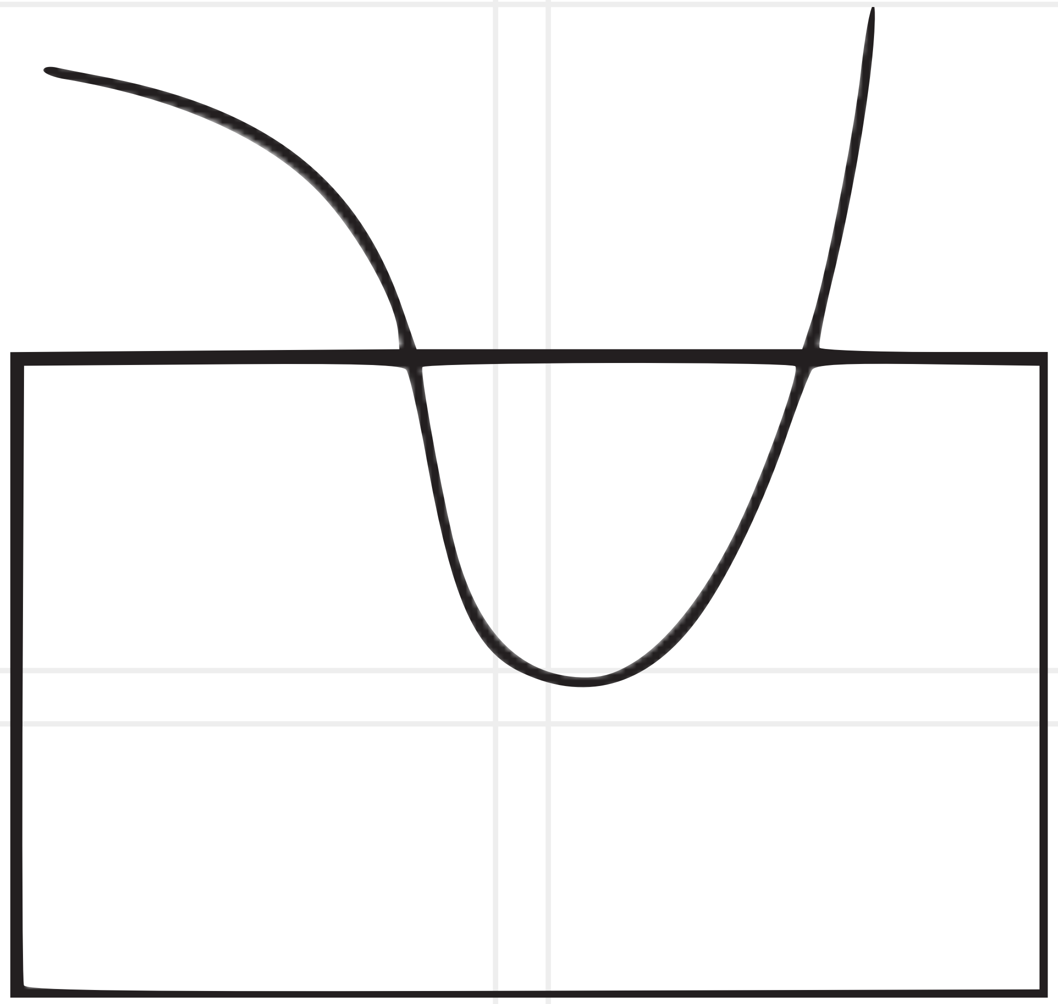


Closure



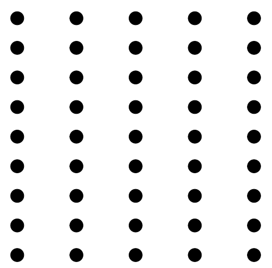
Continuity



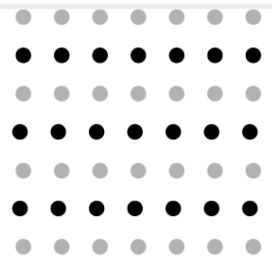


 | **aligning and organizing information reduces cognitive load — *connection***

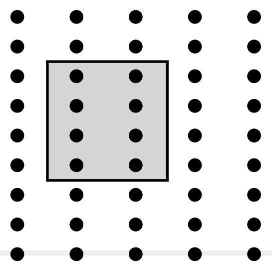
Proximity



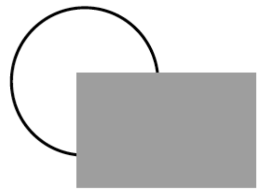
Similarity



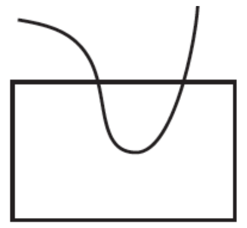
Enclosure



Closure

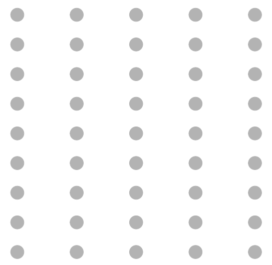


Continuity

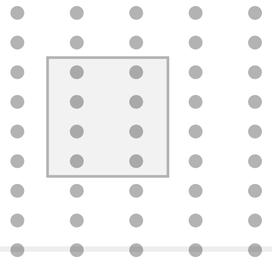


Connection

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

Proximity 

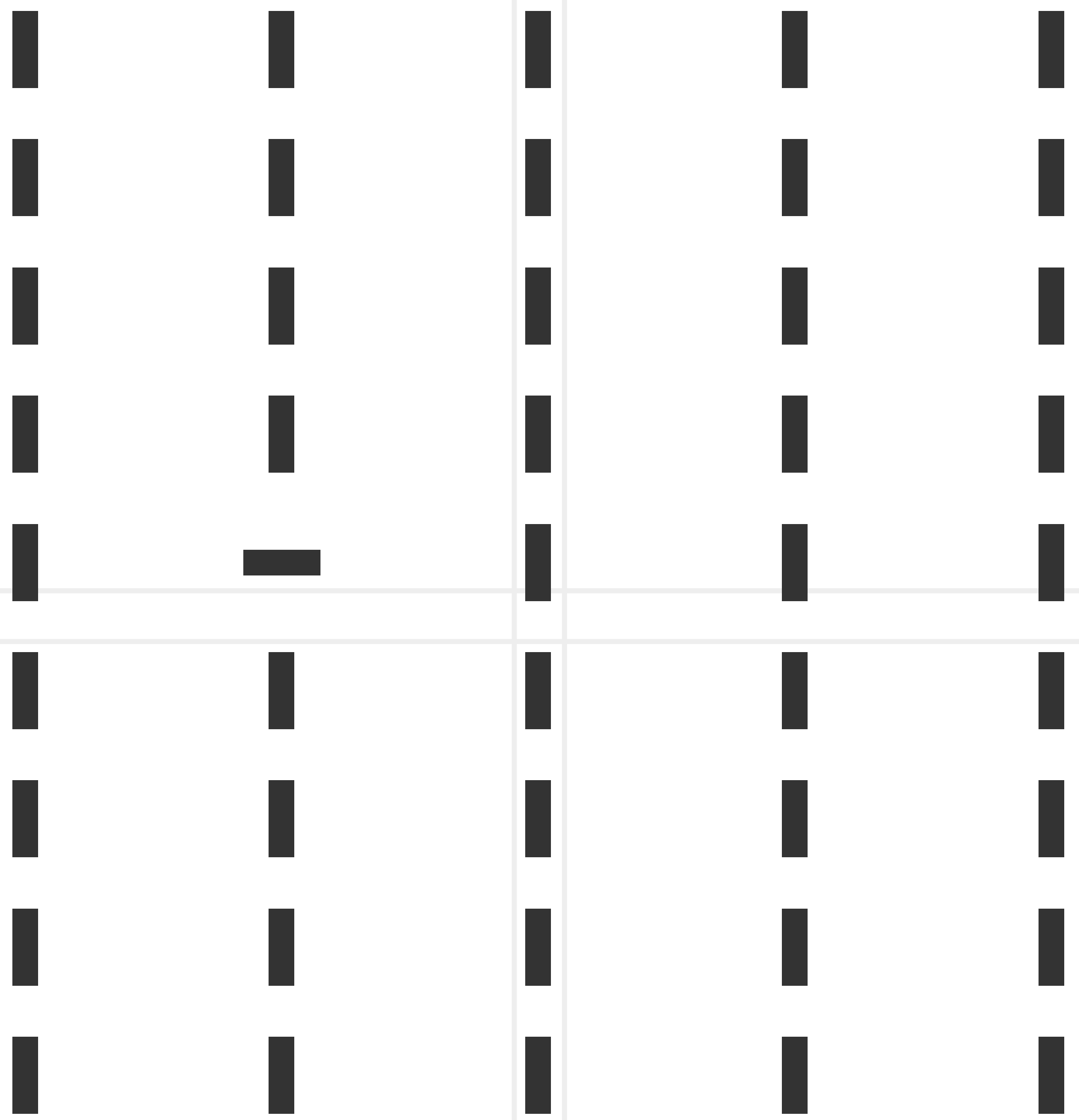
Similarity 

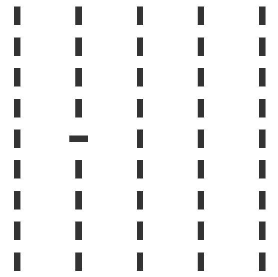
Enclosure 

Closure 

Continuity 

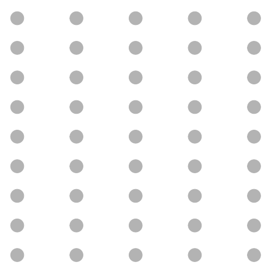
Connection 



 Orientation

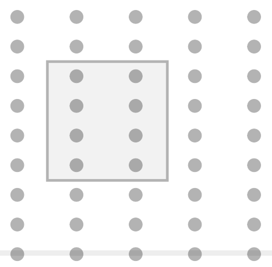
design mini-review | purposeful change of a visual channel can focus attention — *shape*

Proximity



Similarity

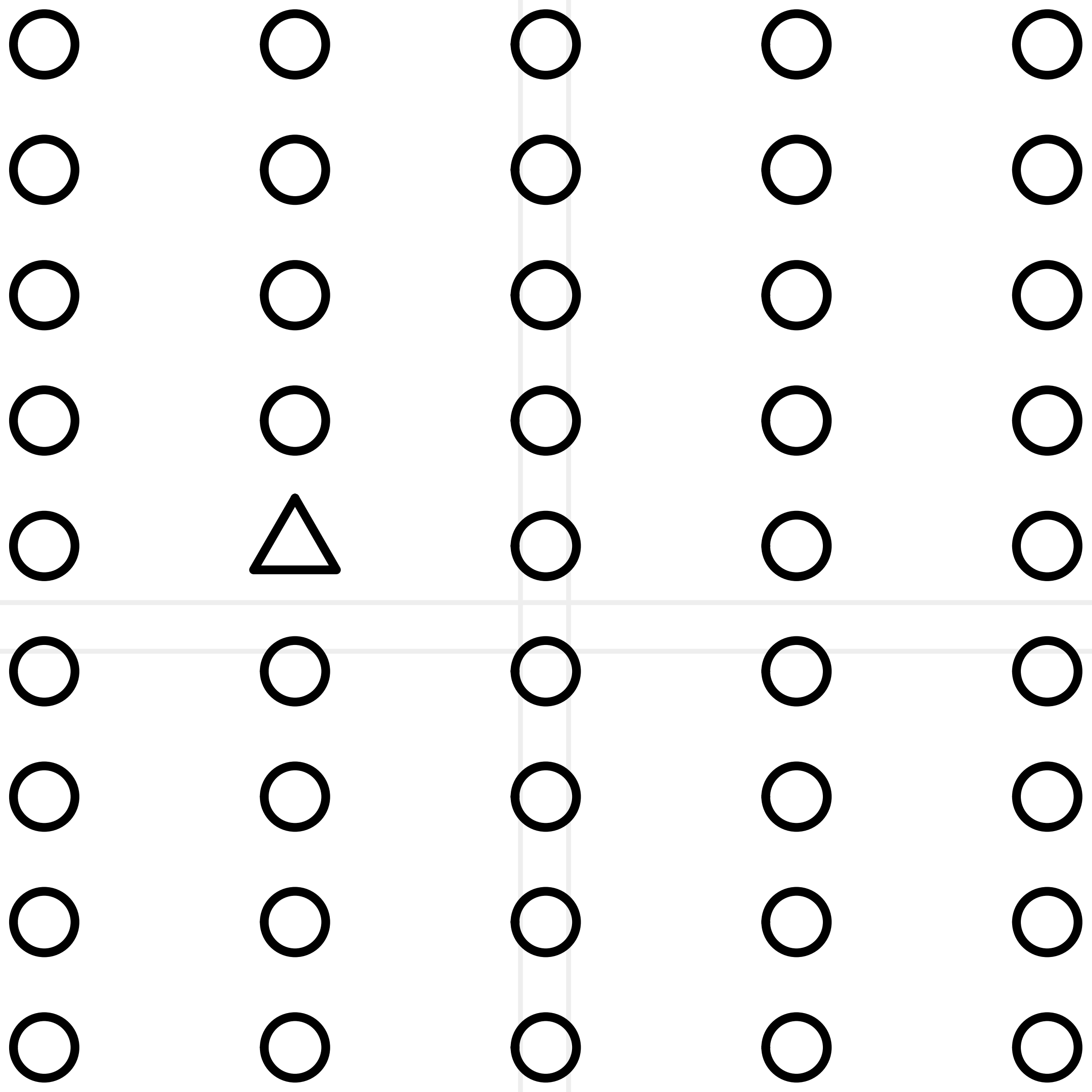
Enclosure

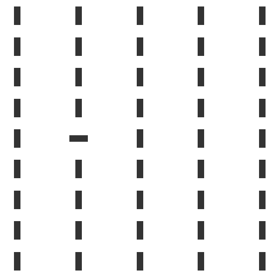


Closure

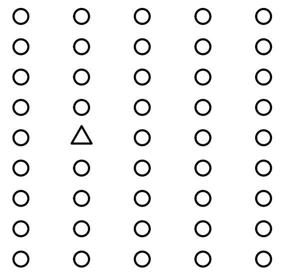
Continuity

Connection





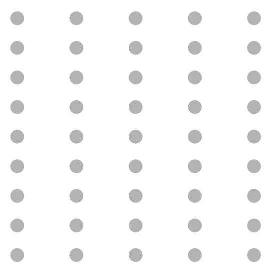
Orientation



Shape

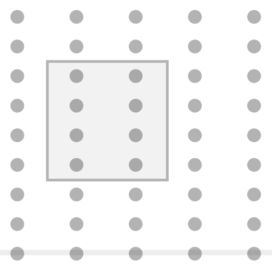
design mini-review | purposeful change of a visual channel can focus attention — *luminance*

Proximity



Similarity

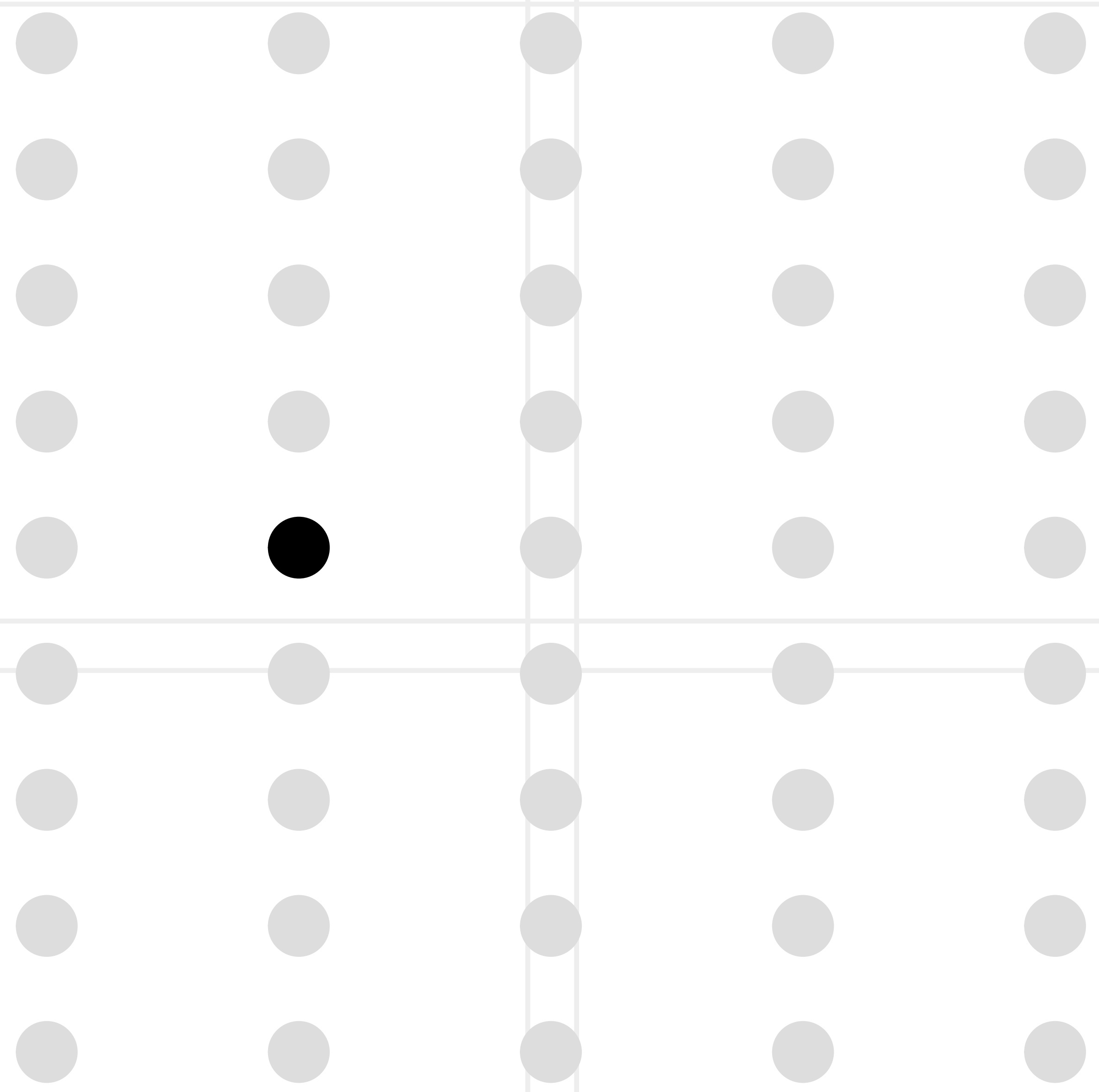
Enclosure

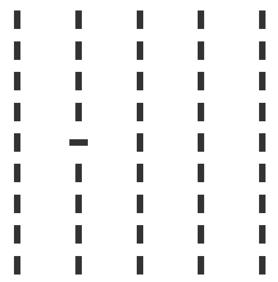


Closure

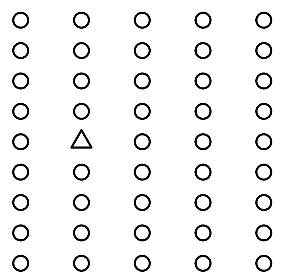
Continuity

Connection





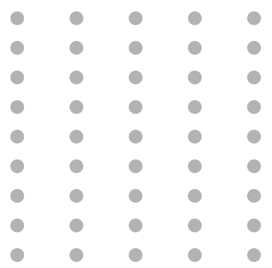
Orientation



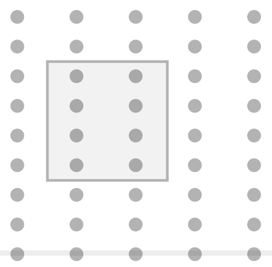
Shape

Luminance

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

Proximity


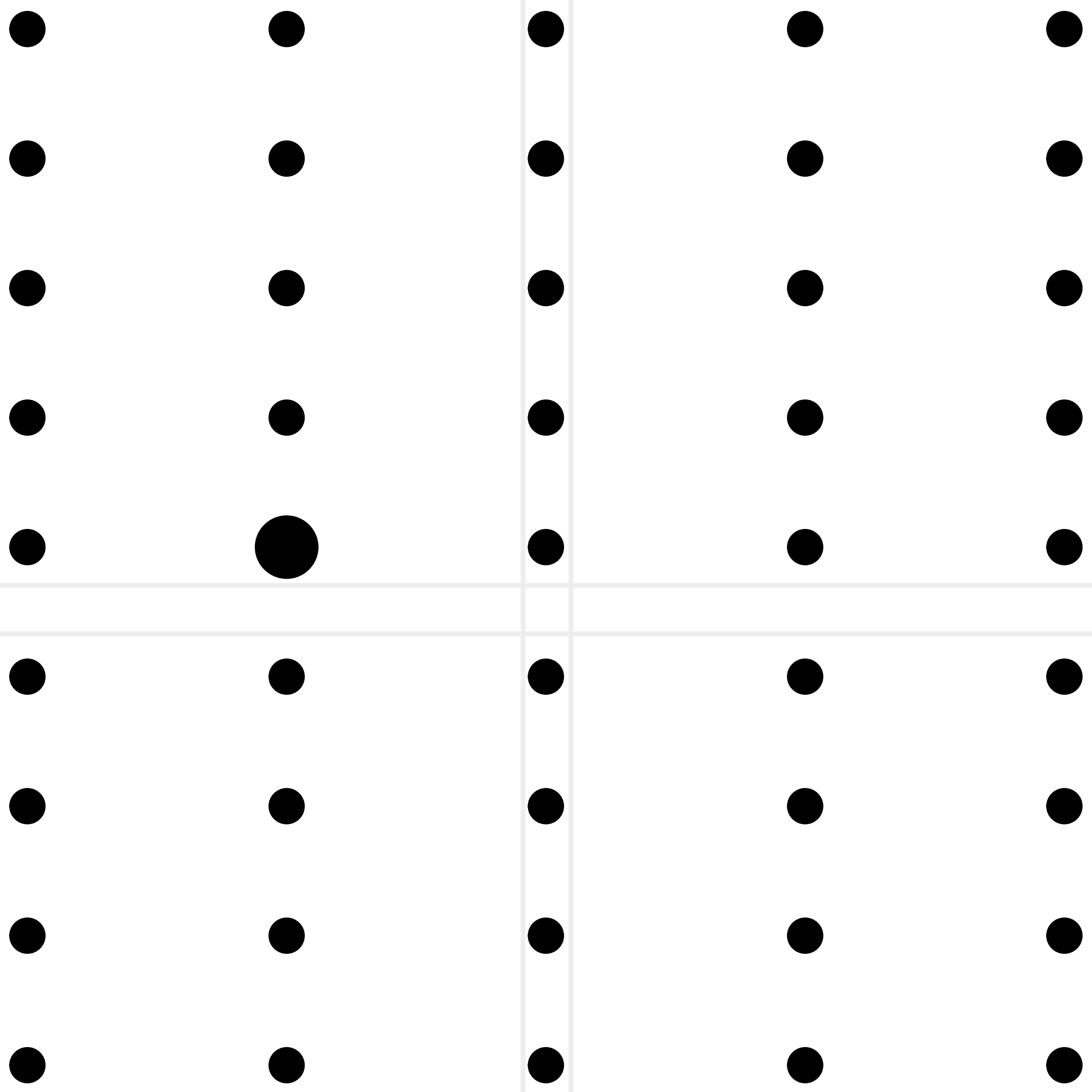
Similarity

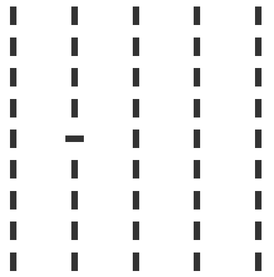

Enclosure


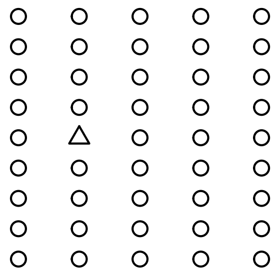
Closure


Continuity

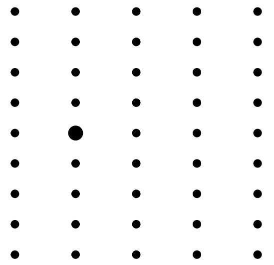

Connection


Orientation


Shape

Luminance


Size

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

Proximity

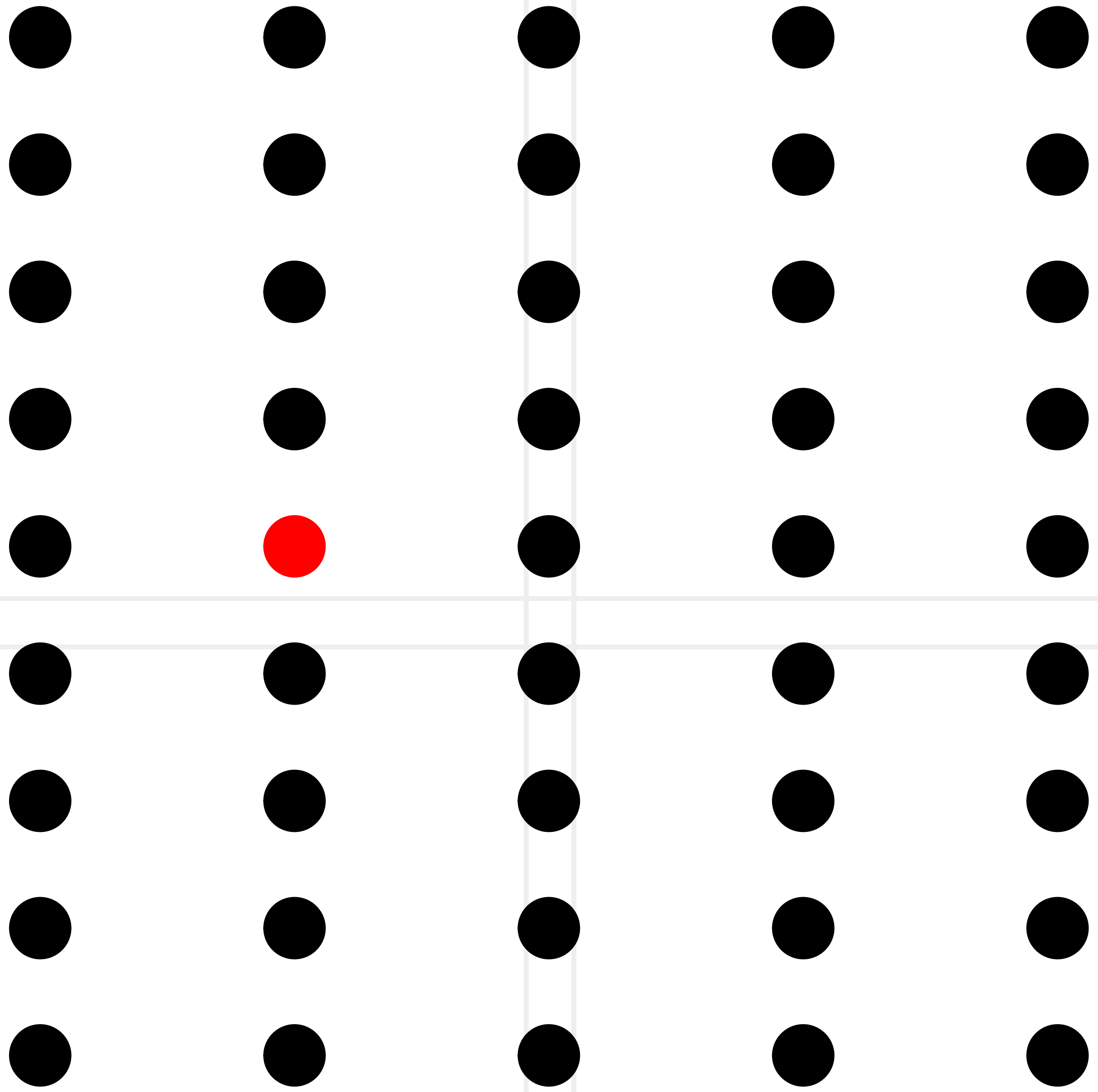
Similarity

Enclosure

Closure

Continuity

Connection



Orientation

Shape

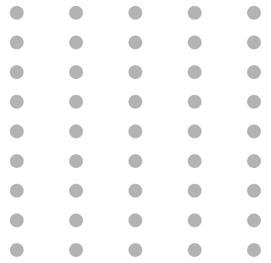
Luminance

Size

Hue

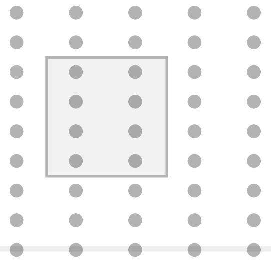
design mini-review | purposeful change of a visual channel can focus attention — *enclosure*

Proximity



Similarity

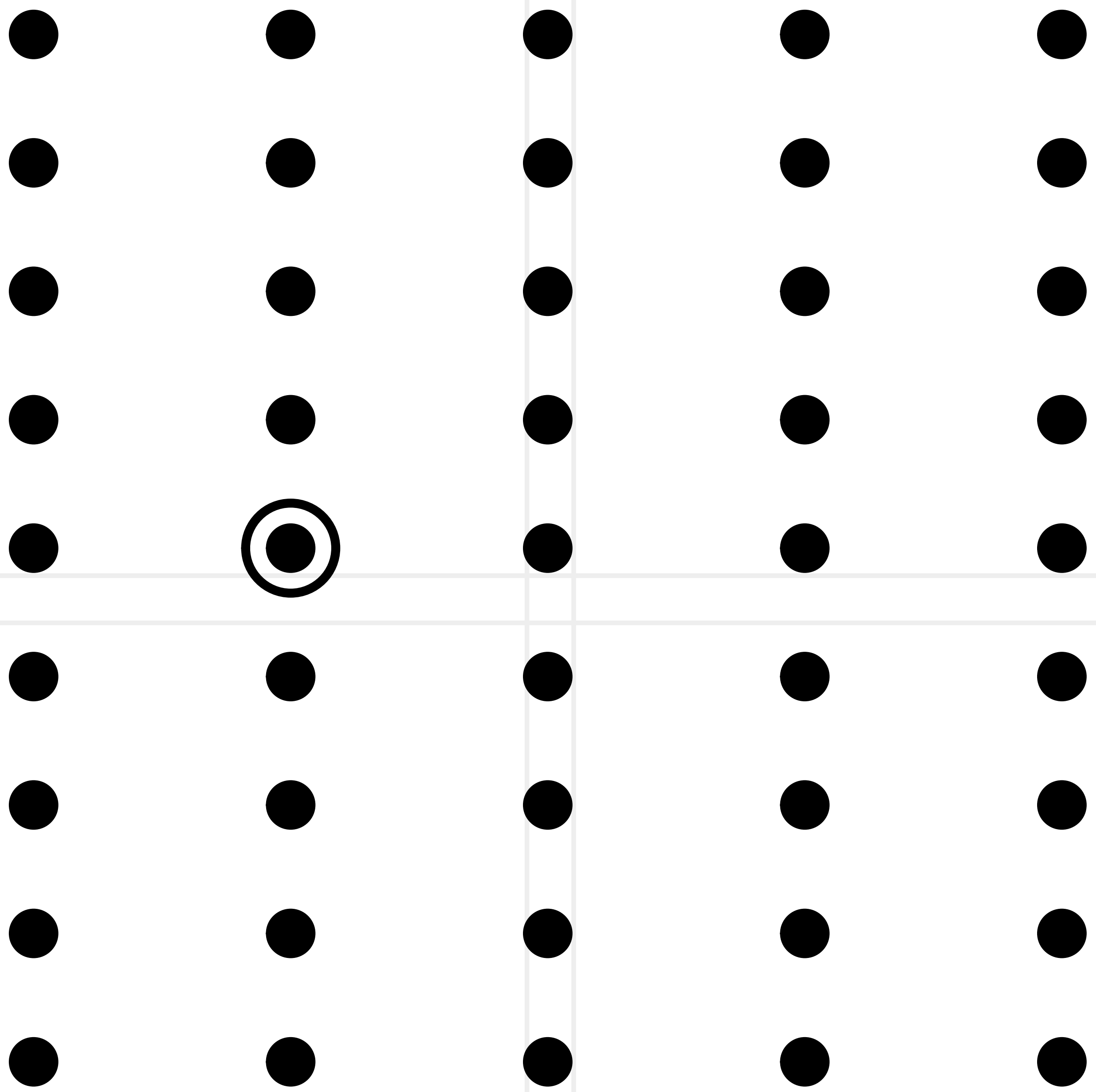
Enclosure

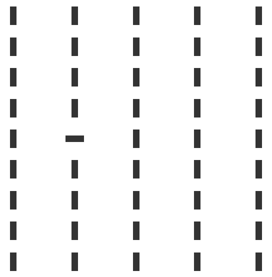


Closure

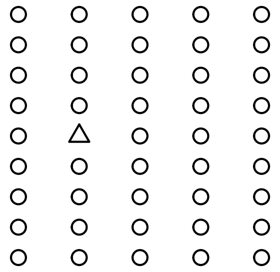
Continuity

Connection



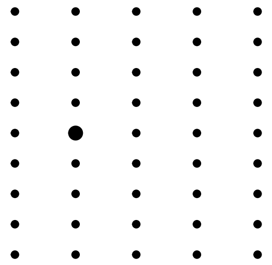


Orientation

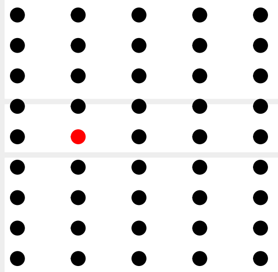


Shape

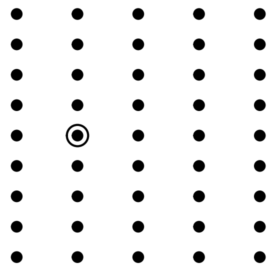
Luminance



Size



Hue



Enclosure

Proximity

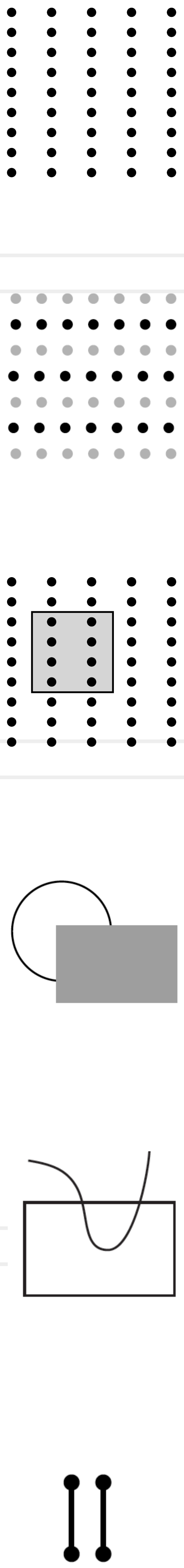
Similarity

Enclosure

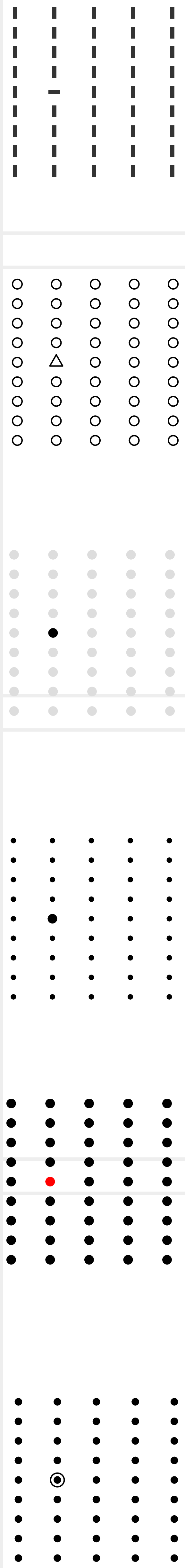
Closure

Continuity

Connection



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



Orientation

Shape

Luminance

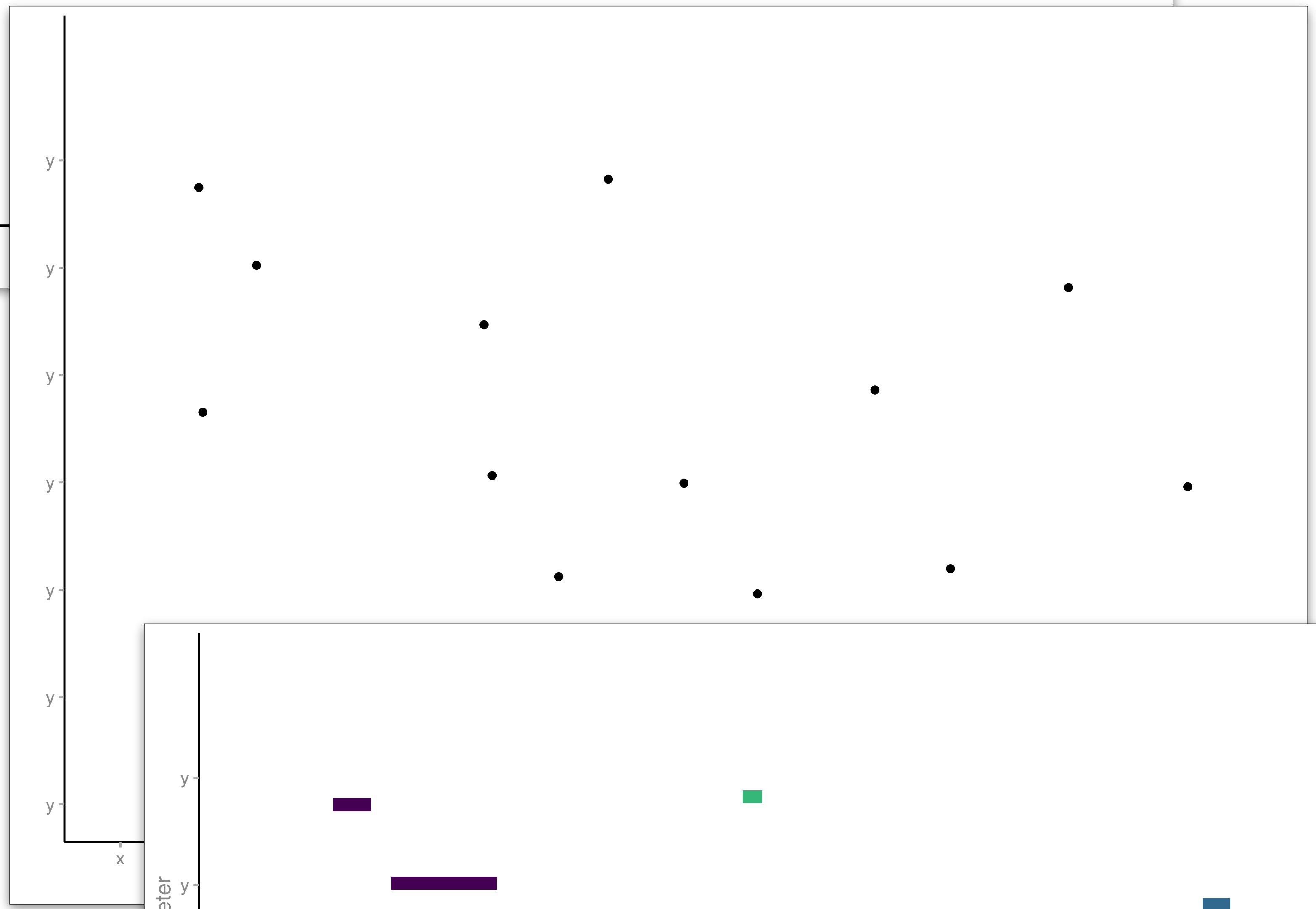
Size

Hue

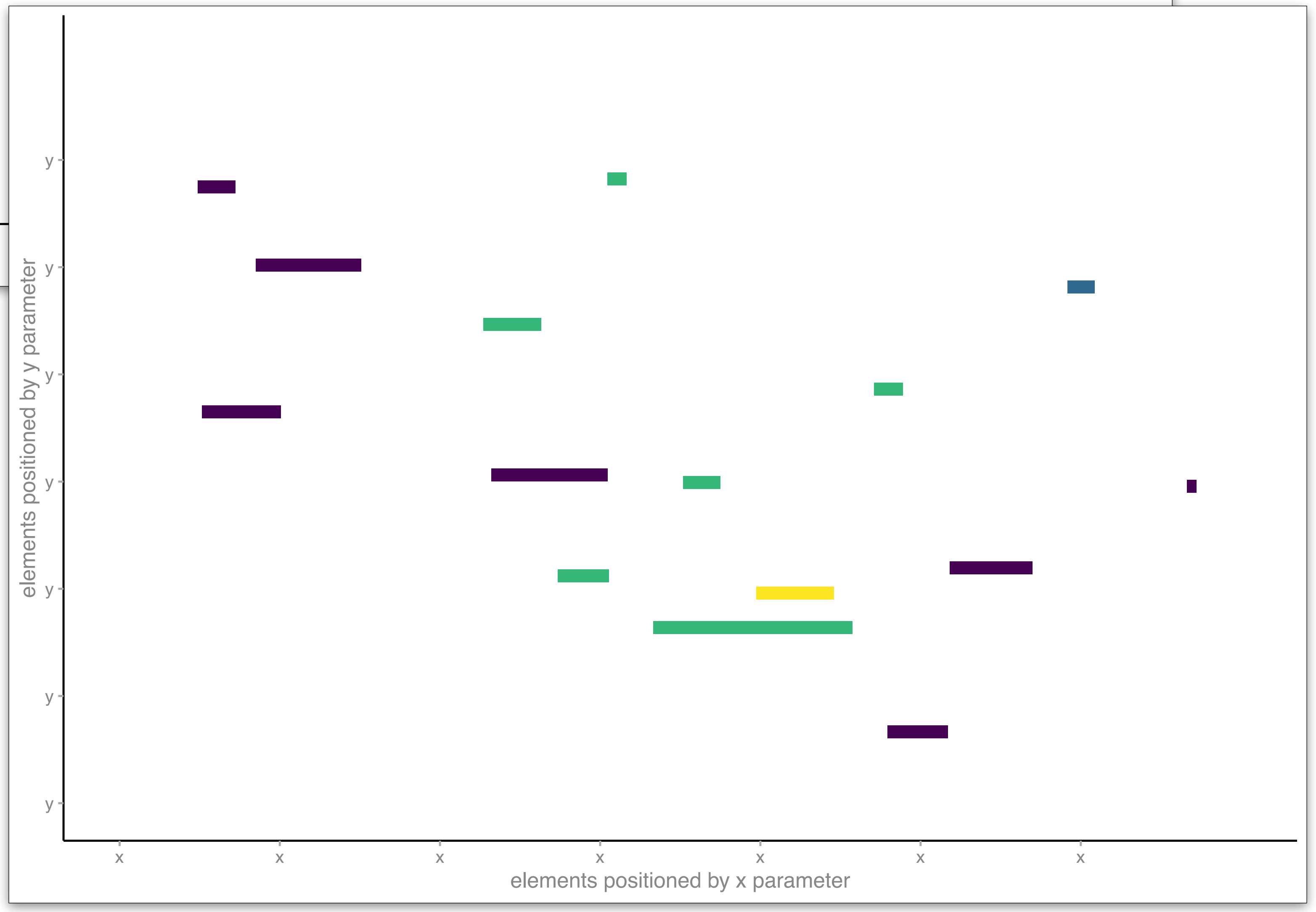
Enclosure

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

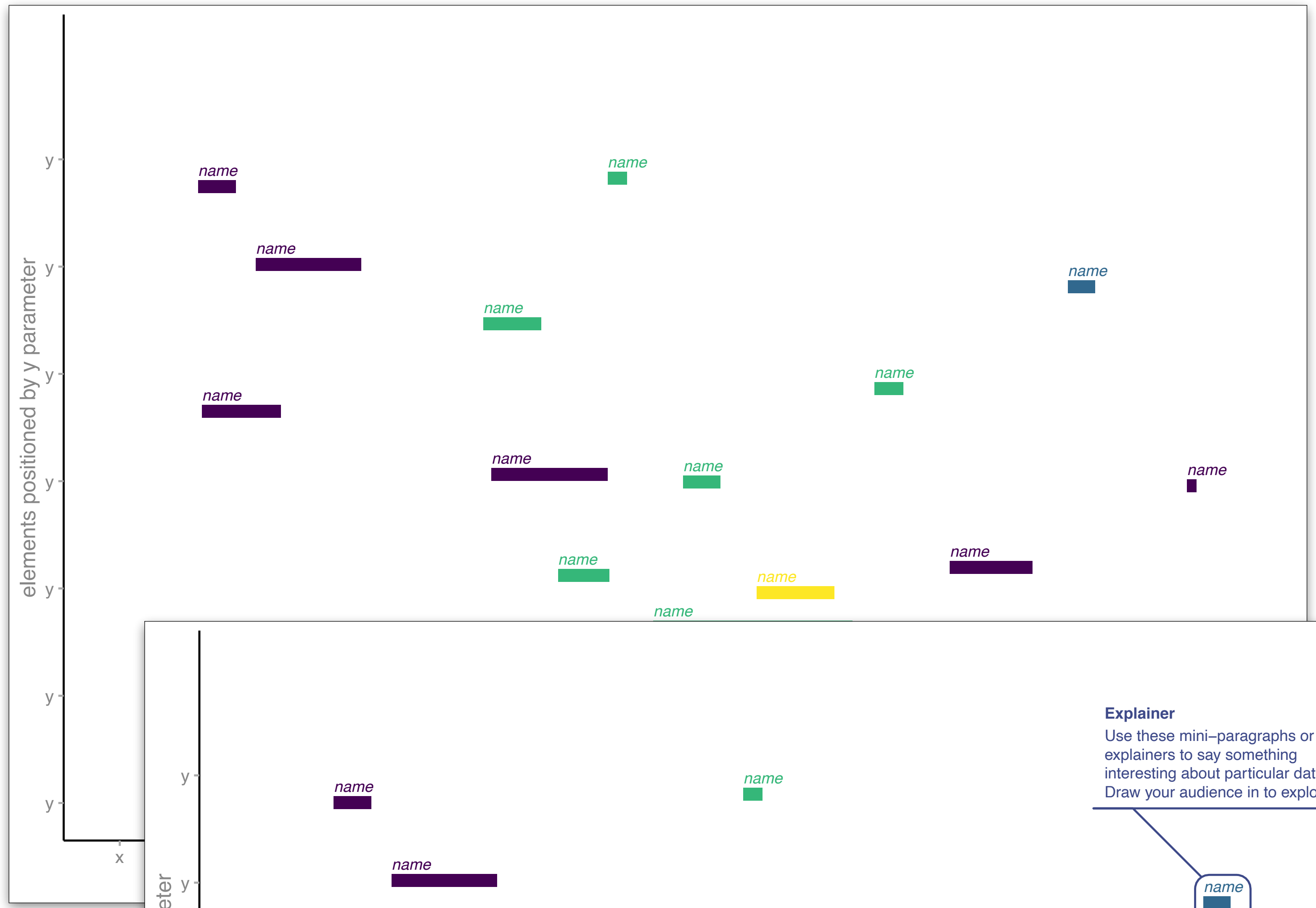
Determine appropriate **scaled axes** for the data architecture.



From data, **position** single elements within the chosen scaled axes.

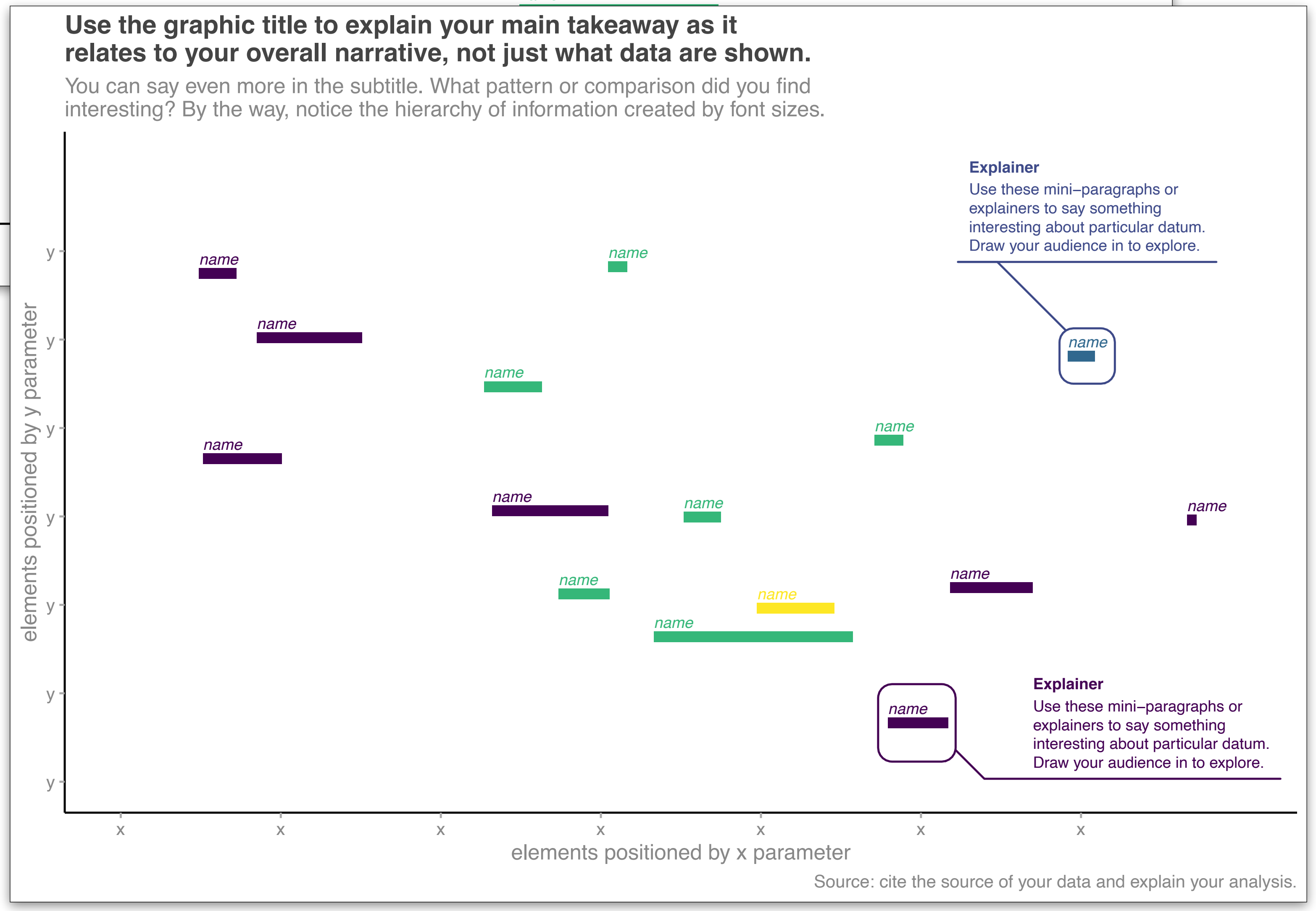


Use **visual channels** and **attributes** to encode data variables at their coordinate positions.



Label elements important to your narrative and audience.

Add **explainers** of interesting data, comparisons, and reference points.

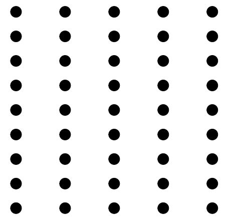


Finish the hierarchy with **insights** and **messages** in title, cite **sources**, other details, like **how to read** the graphic.

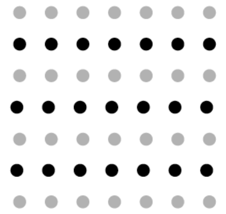
Lighten or color elements **gray** that add context but aren't the message focus.

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

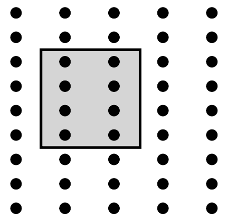
Proximity



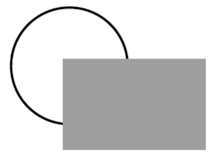
Similarity



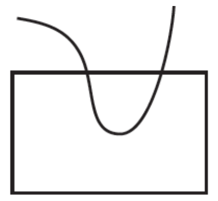
Enclosure



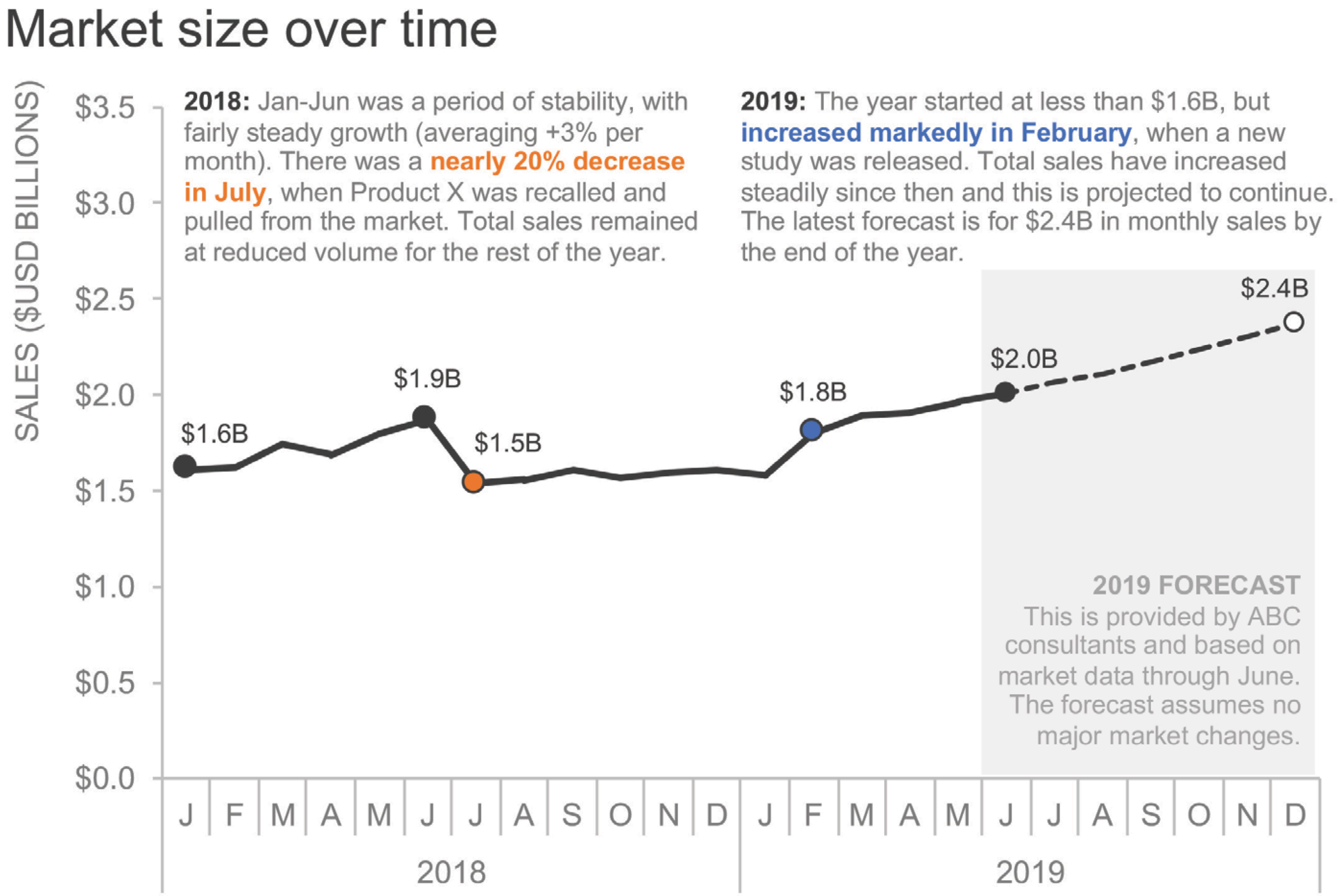
Closure



Continuity

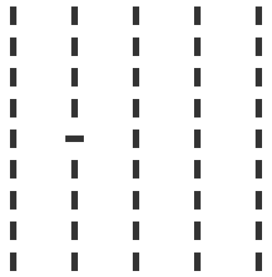


Connection

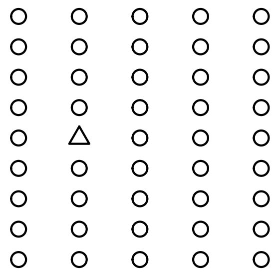


Example from: Knaflic, Cole Nussbaumer. *Storytelling with Data: Let's Practice!* Hoboken, New Jersey: John Wiley & Sons, Inc, 2019.

Orientation

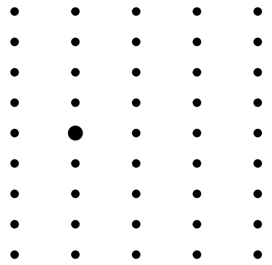


Shape

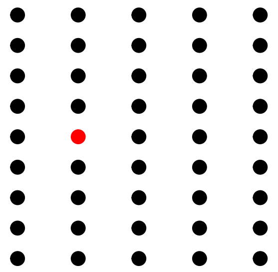


Luminance

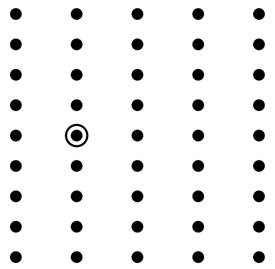
Size



Hue

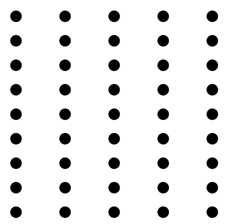


Enclosure

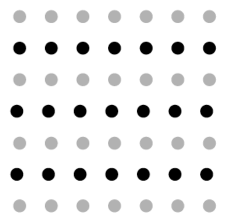


design mini-review | what Gestalt principles are used in this data graphic? How is attention focused?

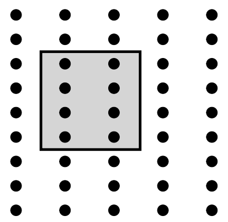
Proximity



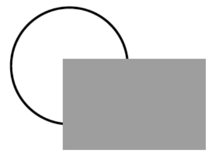
Similarity



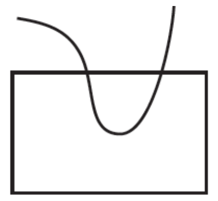
Enclosure



Closure



Continuity

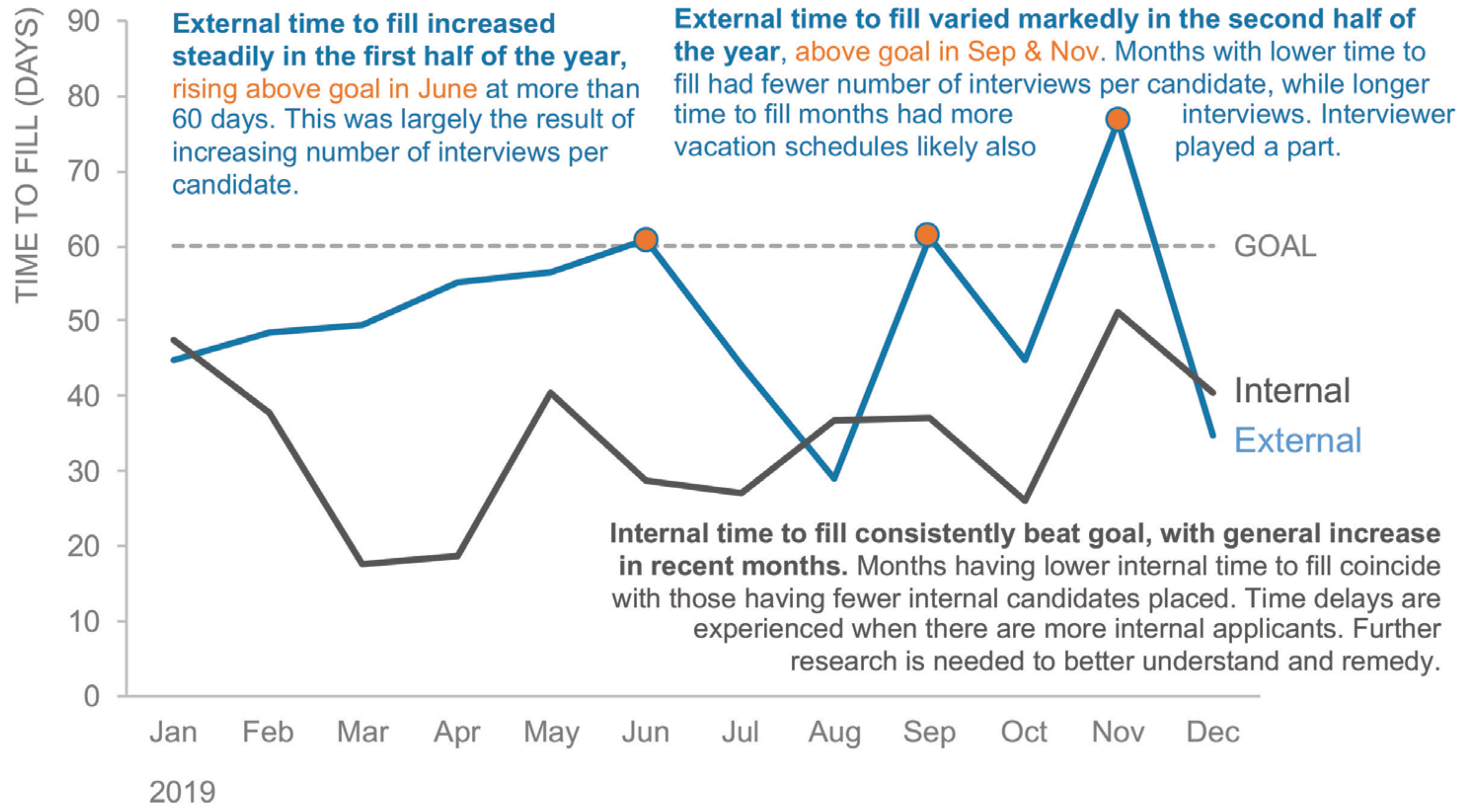


Connection

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

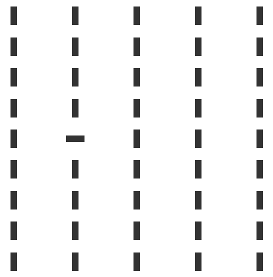
Both **External** and **Internal** time to fill have varied in the past year. Understanding contributing factors—number of interviews, vacation schedules, and current internal transfer volume constraints—can help us better plan for the future.

Time to fill

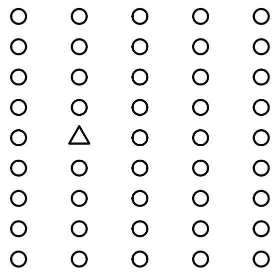


LET'S DISCUSS: Should we put stricter guidelines around maximum number of interviews? How can we keep vacation schedules from impacting time to hire? What can we do to improve efficiency of internal transfer process in order to better handle higher volumes?

Orientation

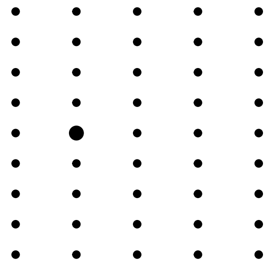


Shape

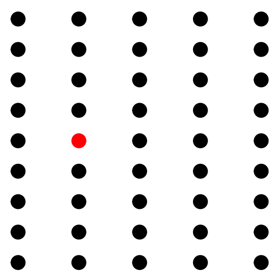


Luminance

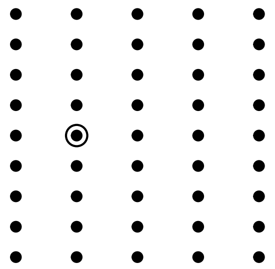
Size



Hue



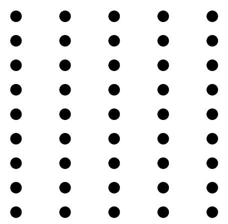
Enclosure



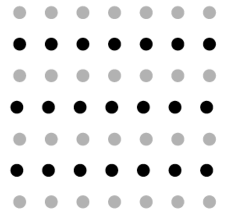
Example from: Knaflic, Cole Nussbaumer. *Storytelling with Data: Let's Practice!* Hoboken, New Jersey: John Wiley & Sons, Inc, 2019.

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

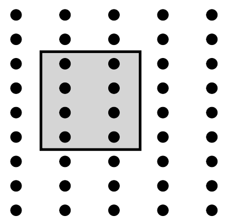
Proximity



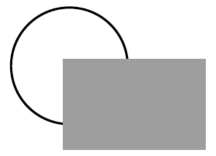
Similarity



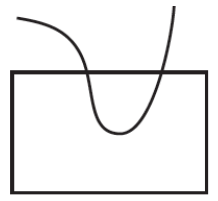
Enclosure



Closure



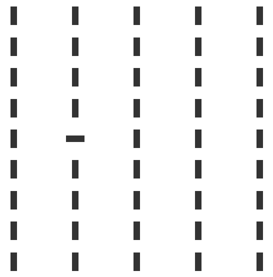
Continuity



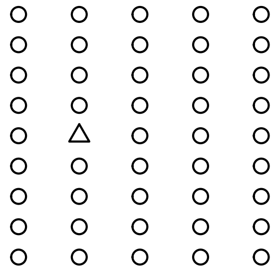
Connection

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

Orientation

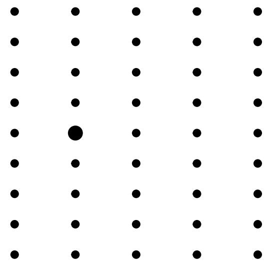


Shape

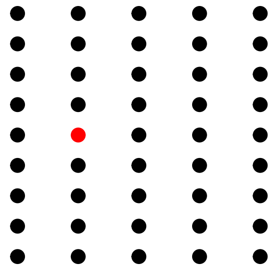


Luminance

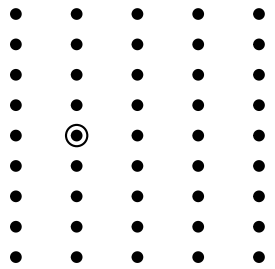
Size



Hue



Enclosure



a framework for critiquing data-driven, visual narratives

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

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.

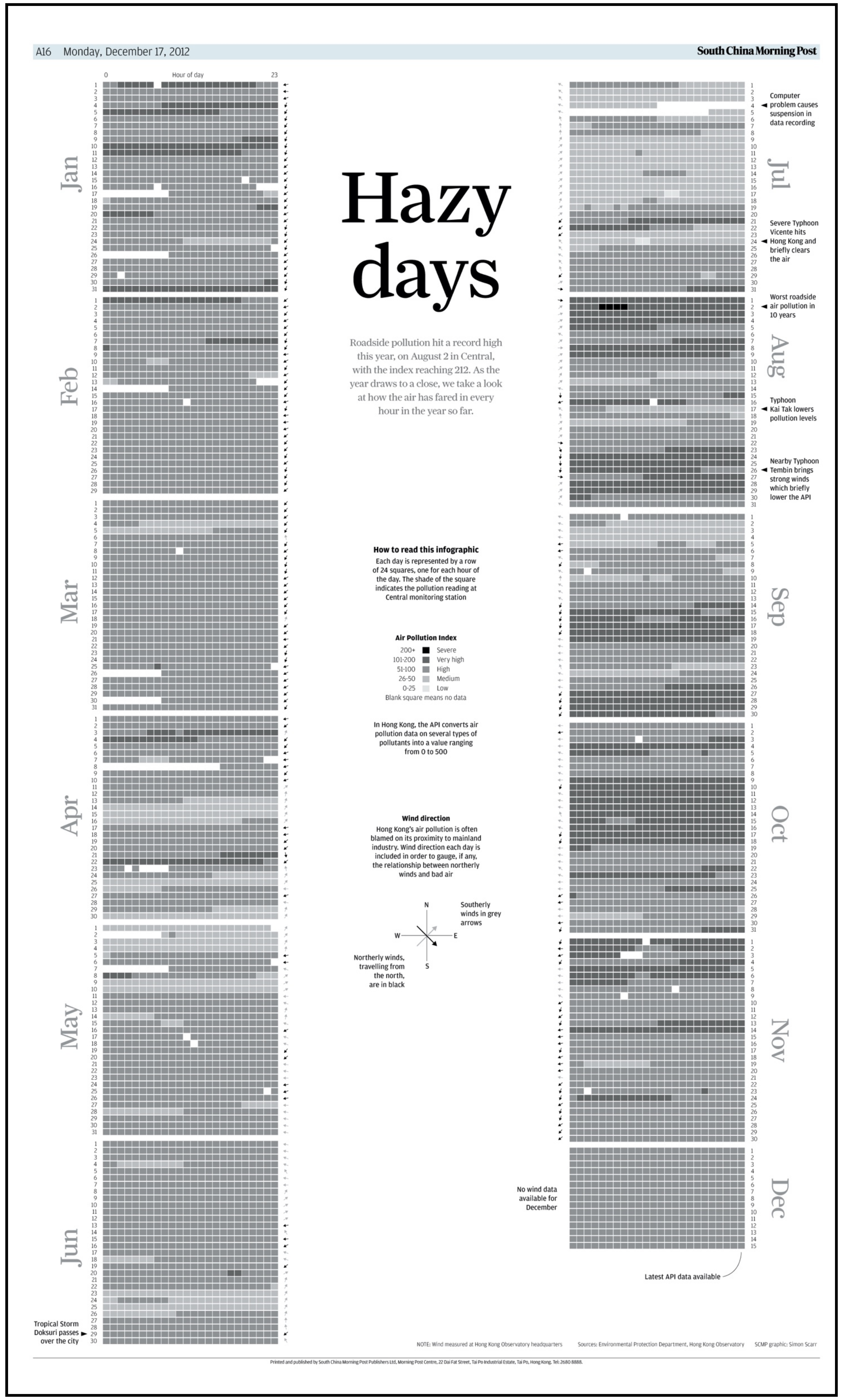
Color, coherency? Only shades of gray encode data, and especially for encoding pollution on a heat map. Notice the gray palette also serves as a visual metaphor as we think of pollution as physically graying our otherwise blue skies. Would a blue-to-gray encoding strengthen the metaphor?

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.



Audience?

Purpose?

Data encodings, decodings?

Comparison or change?

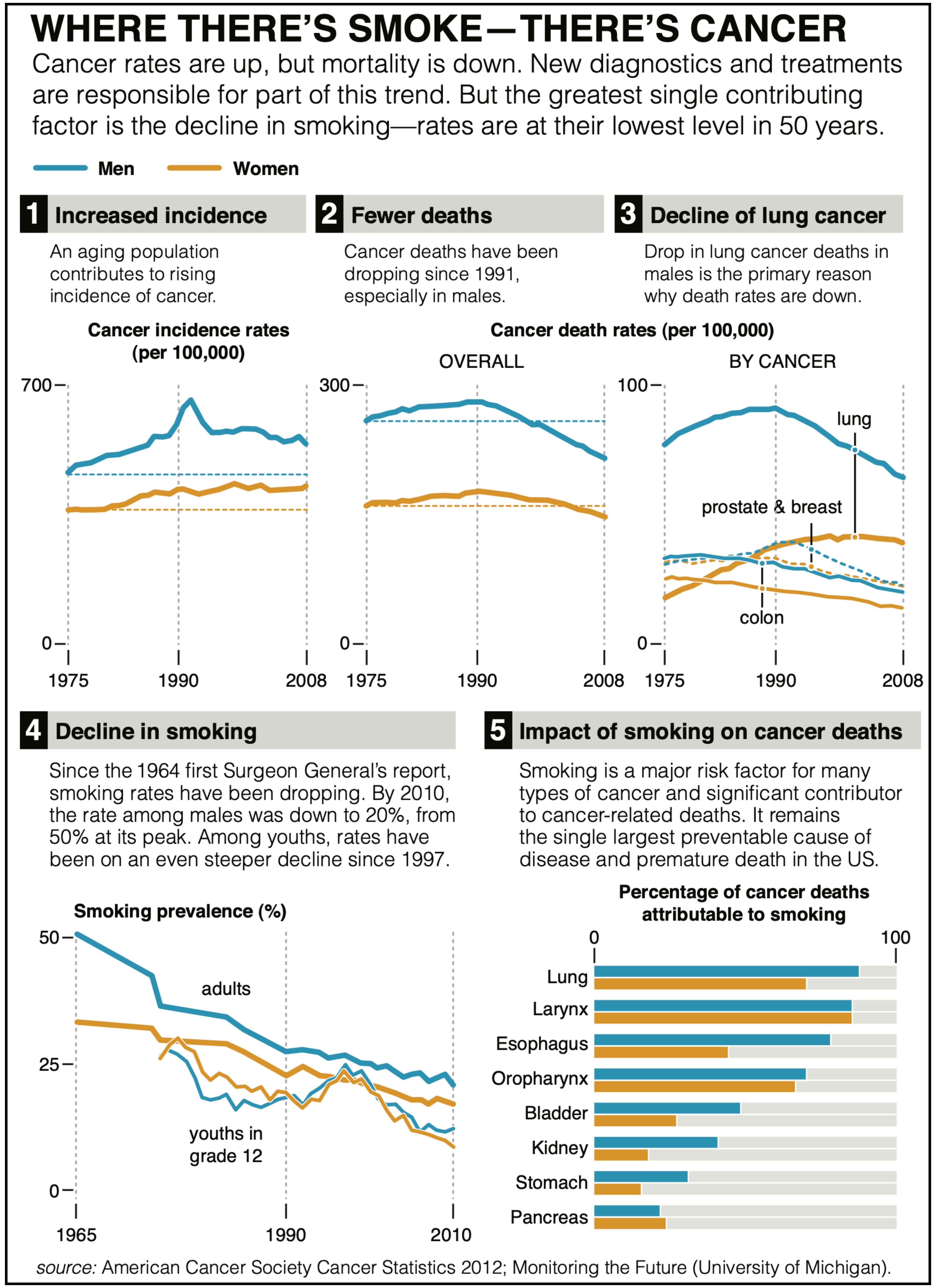
Narrative?

Color, coherency?

Hierarchy, layering, layout?

Credibility, transparency?

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



Audience?

Purpose?

Data encodings, decodings?

Comparison or change?

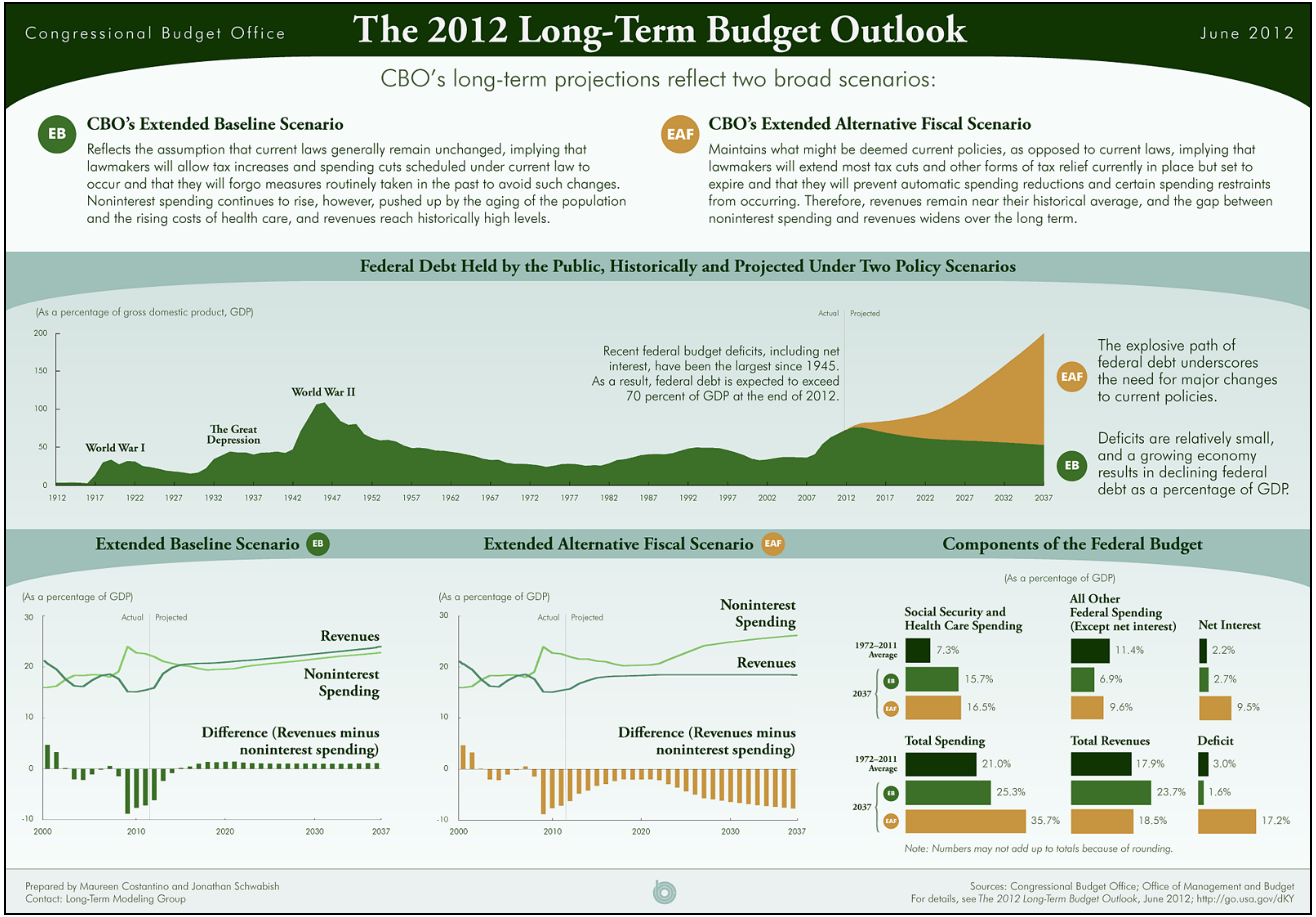
Narrative?

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.



Audience?

Purpose?

Data encodings, decodings?

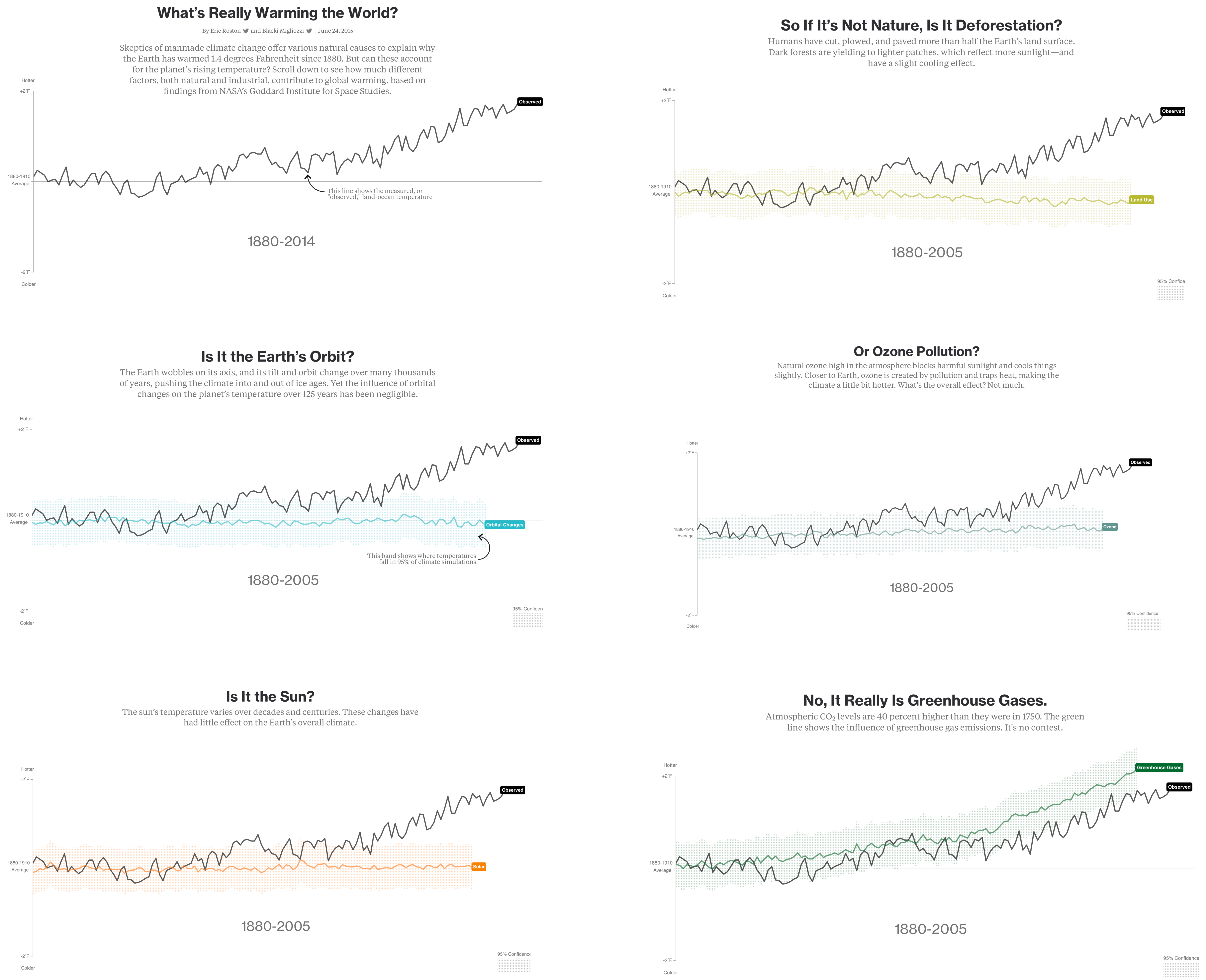
Comparison or change?

Narrative?

Color, coherency?

Hierarchy, layering, layout?

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/.

Audience?

Purpose?

Data encodings, decodings?

Comparison or change?

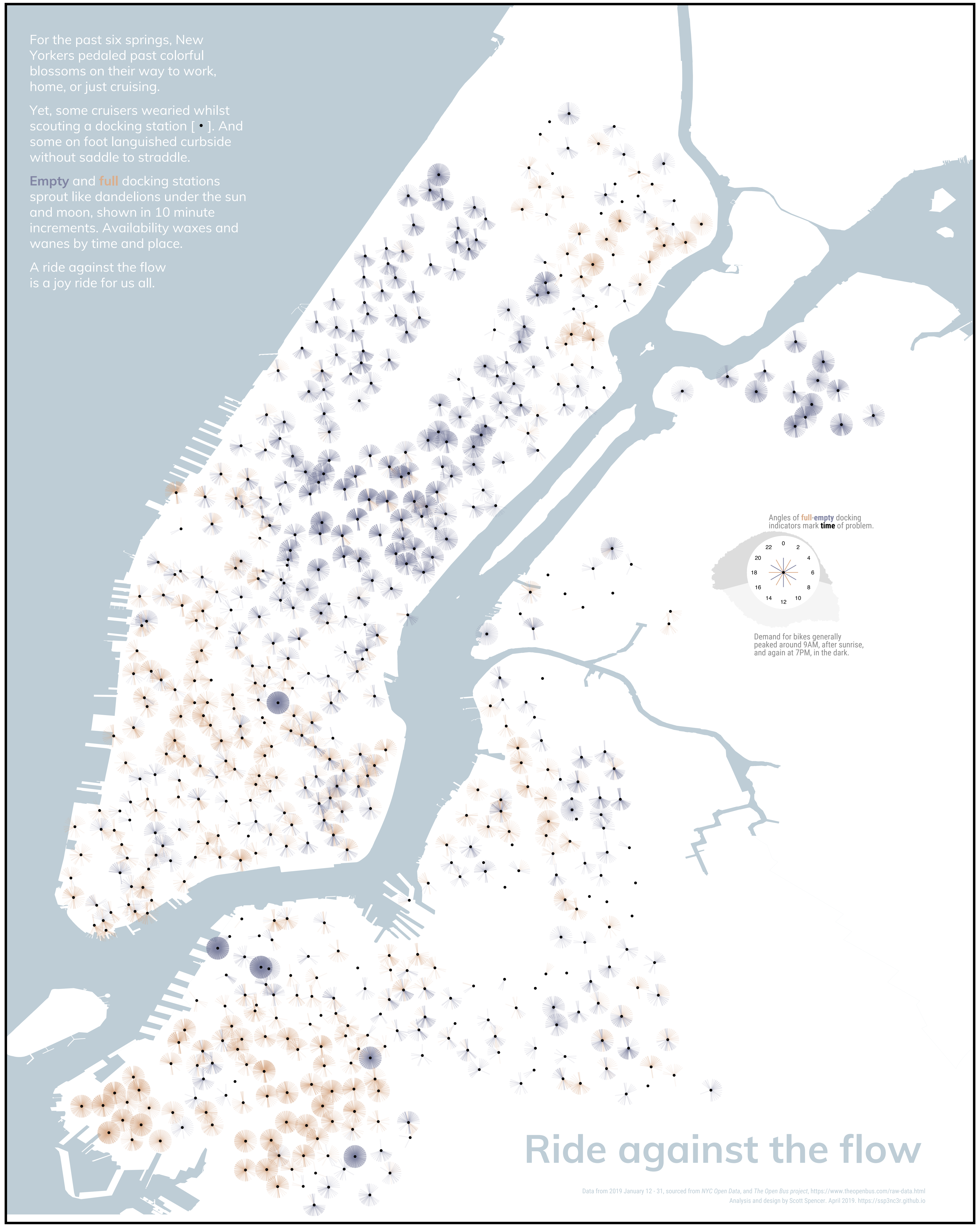
Narrative?

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.



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.

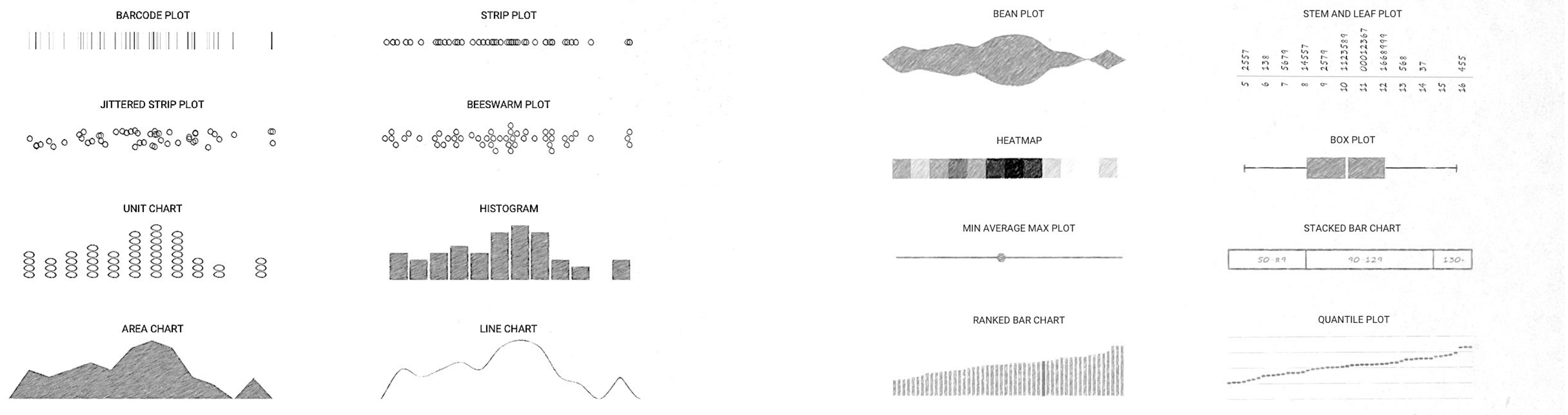
Hierarchy, layering, layout? Your titles, headers, mini-paragraphs, and text should use messages, not just information. Use best practices in typography (size, bold, color, spacing, etc) and grid alignment to focus your audience on your messages.

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.

> 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:

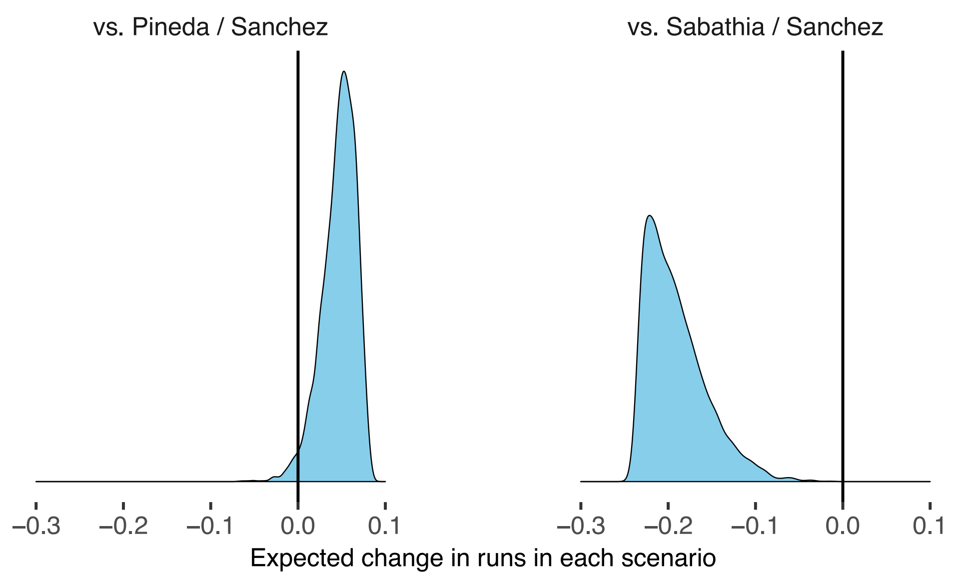


Figure 1. Of the two scenarios, Cain should only attempt to steal against the Sanchez–Pineda duo.

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:

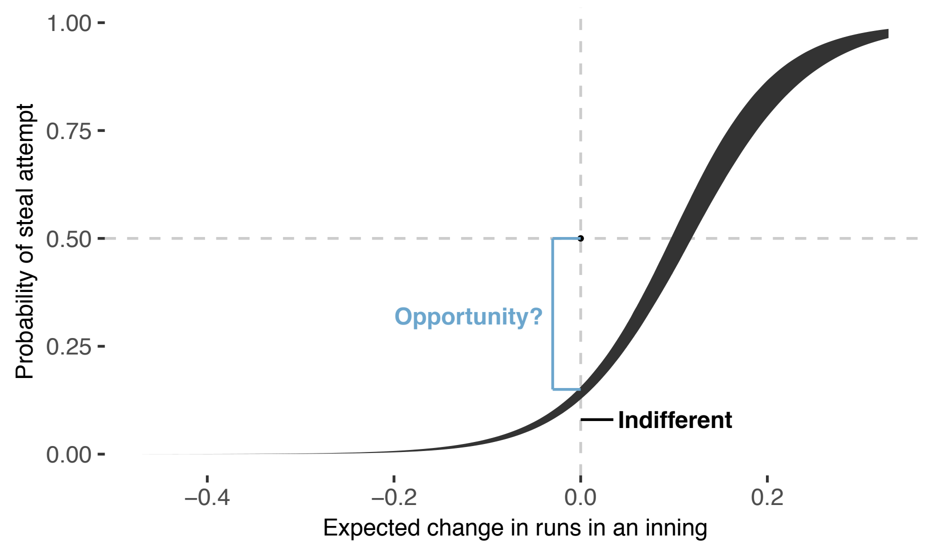


Figure 2. When the change in expected runs is zero, managers should be indifferent to attempted steals, saying go half the time.

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.

Spencer, Scott. Proposal to Scott Powers. "*Proposal for Exploring Game Decisions Informed by Expectations of Joint Probability Distributions.*" February 14, 2019.

> 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:

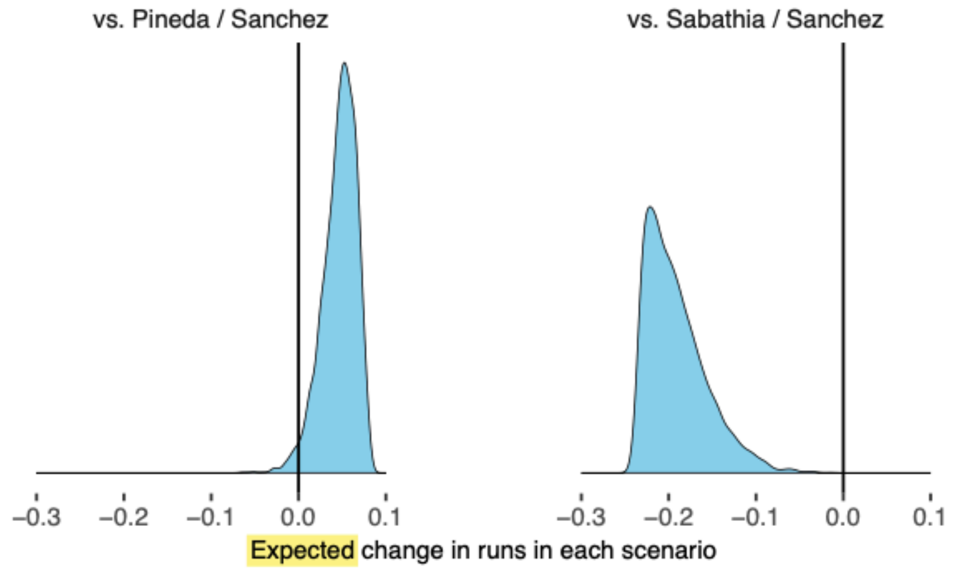


Figure 1. Of the two scenarios, Cain should only attempt to steal against the Sanchez–Pineda duo.

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:

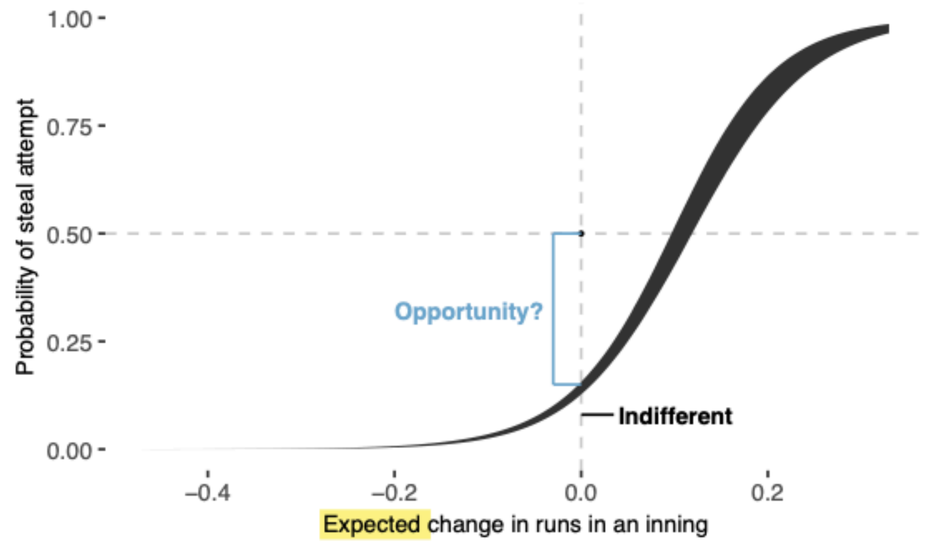


Figure 2. When the change in expected runs is zero, managers should be indifferent to attempted steals, saying go half the time.

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.

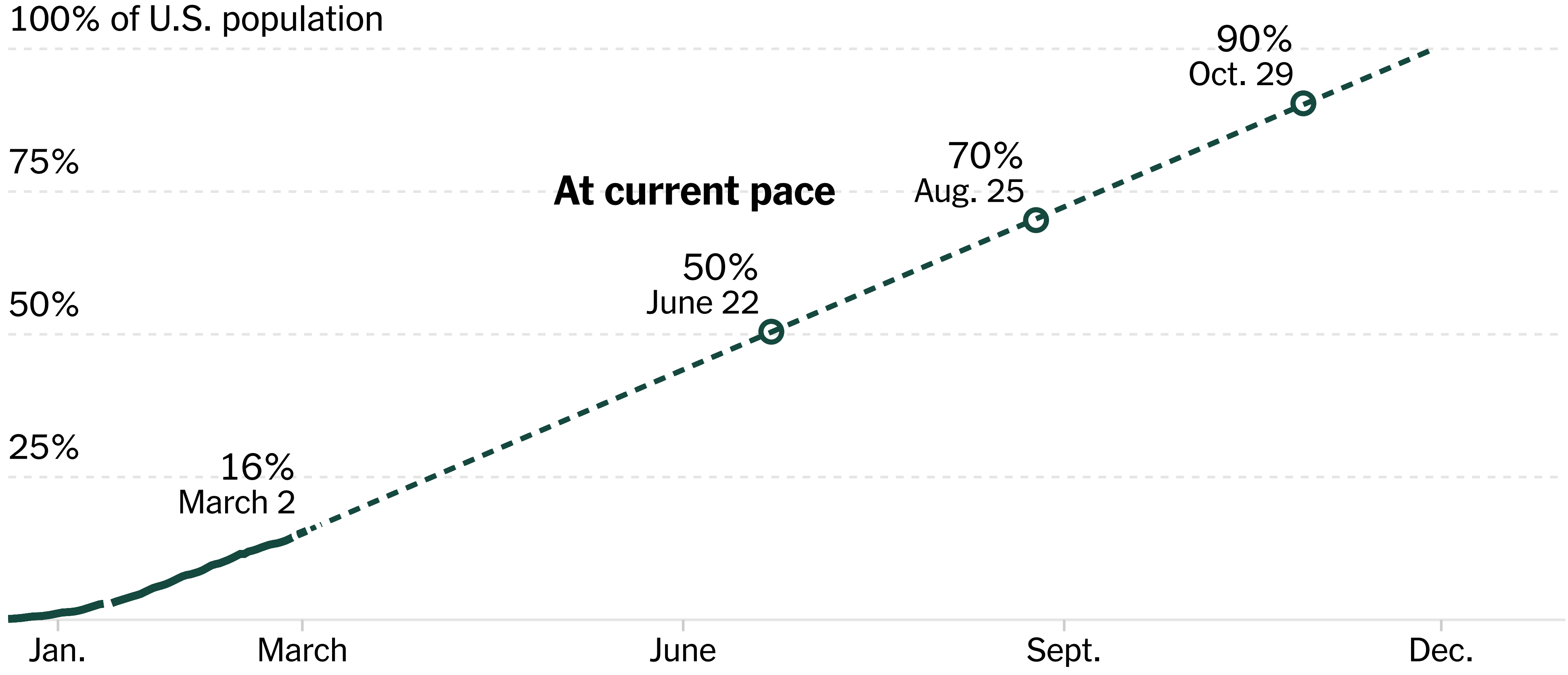
Spencer, Scott. Proposal to Scott Powers. "*Proposal for Exploring Game Decisions Informed by Expectations of Joint Probability Distributions.*" February 14, 2019.

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.

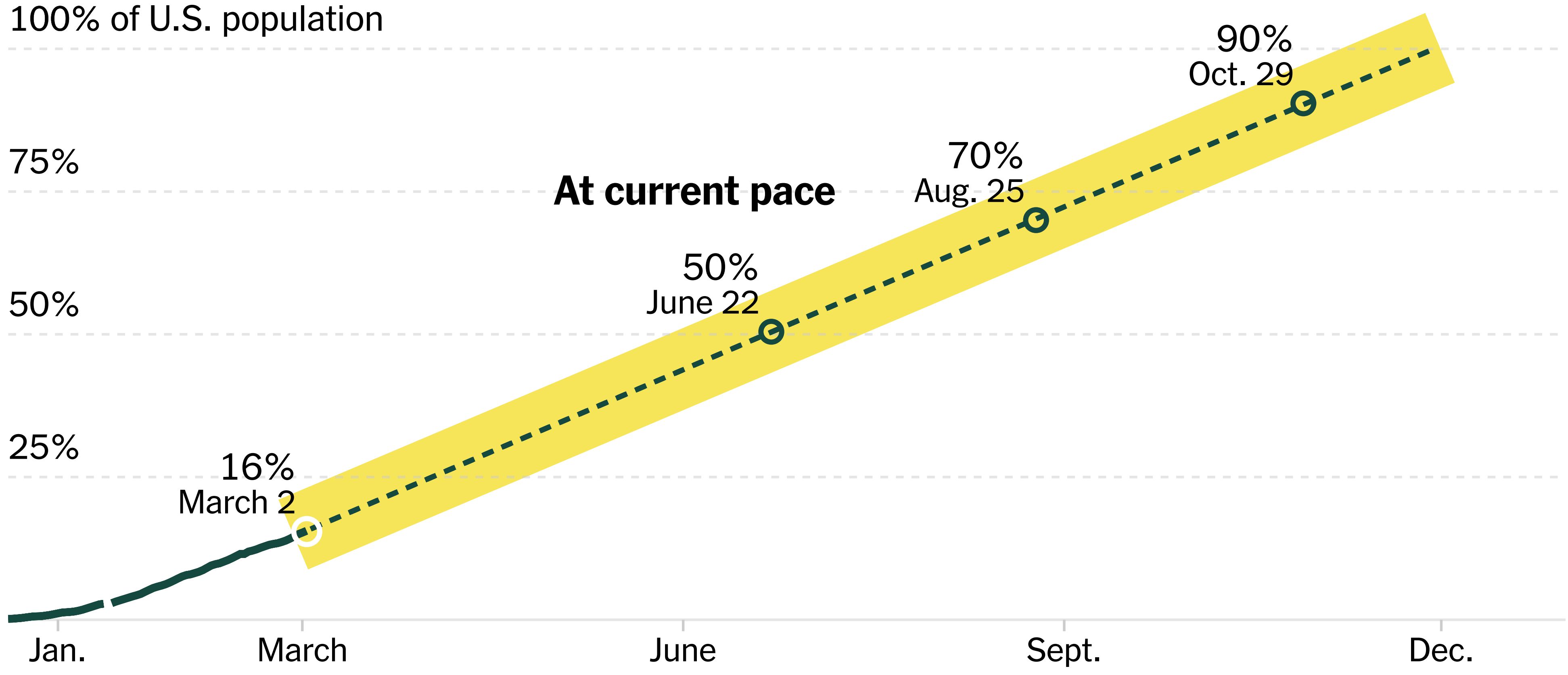
New York Times. "*Covid-19 Vaccine Rollout: State by State*." News, March 1, 2021. https://www.nytimes.com/interactive/2020/us/covid-19-vaccine-doses.html.

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



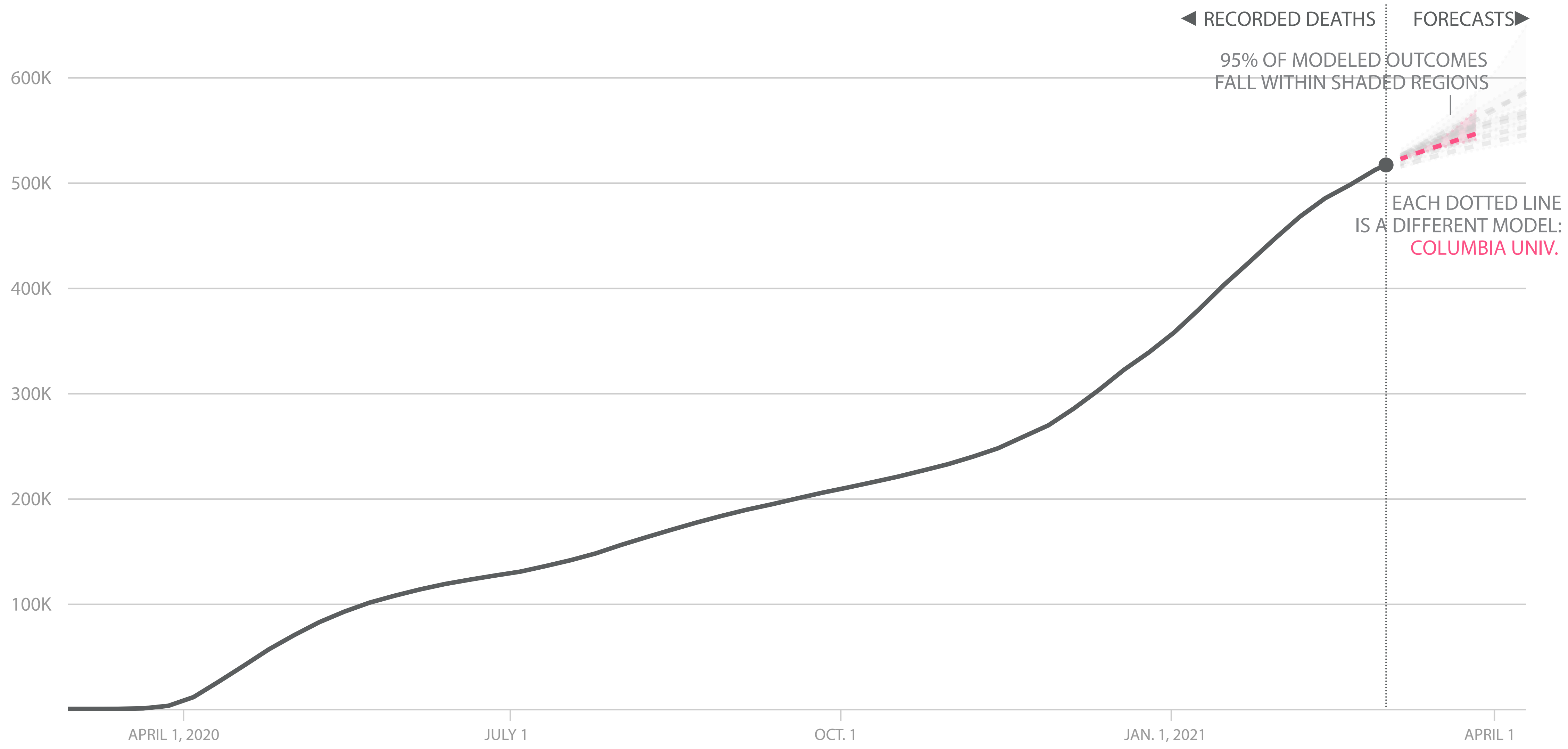
Source: Centers for Disease Control and Prevention | Note: Data from Dec. 20 to Jan. 12 are for all doses administered. Data for Jan. 13 is unavailable. Projections could change if additional vaccines are authorized.

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.

New York Times. "*Covid-19 Vaccine Rollout: State by State*." News, March 1, 2021. https://www.nytimes.com/interactive/2020/us/covid-19-vaccine-doses.html.

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.

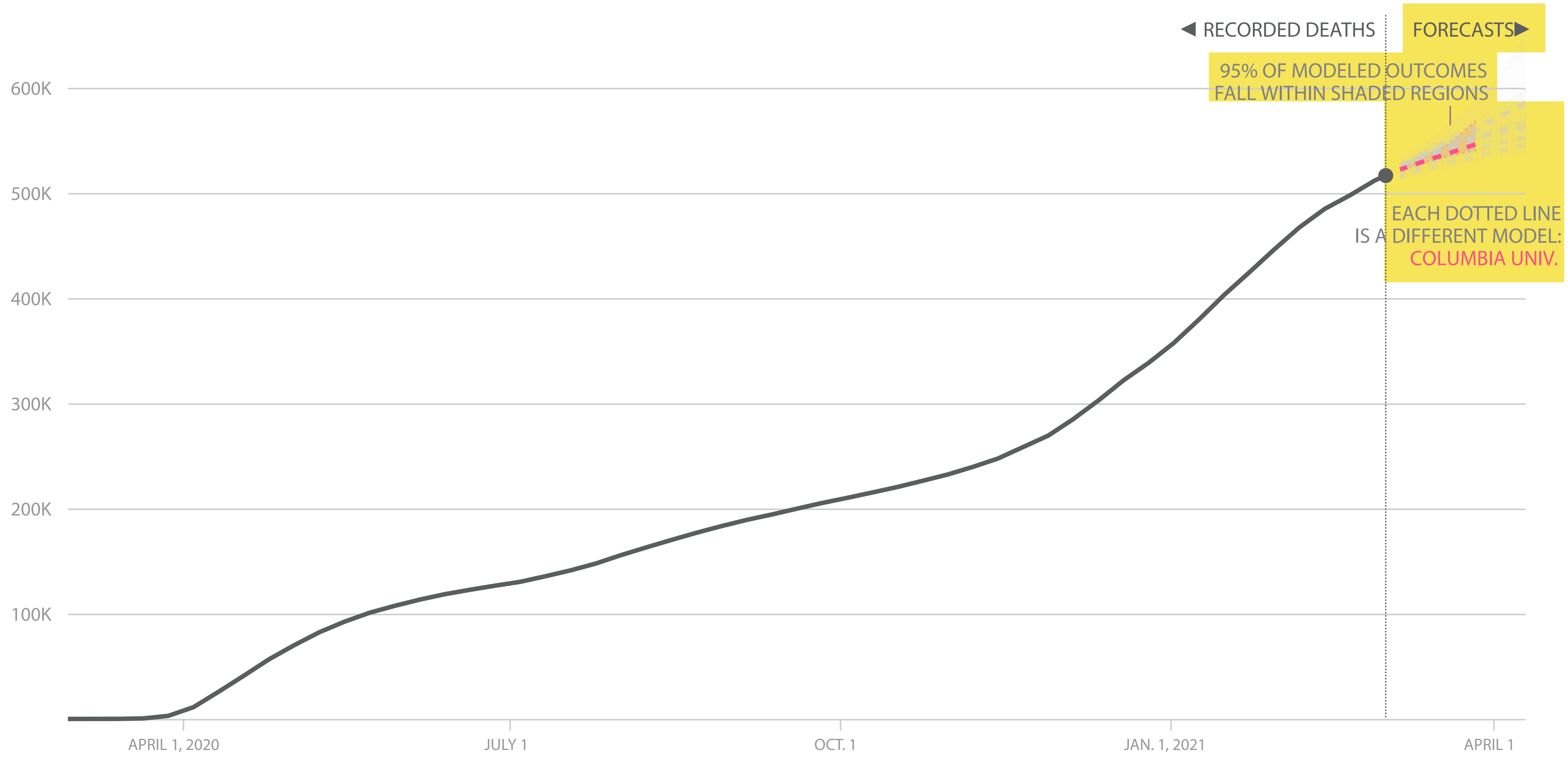
And looking at multiple models is better than looking at just one because it's difficult to know which model will match reality the closest. Even when models disagree, understanding why they are dierent can give us valuable insight.

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.

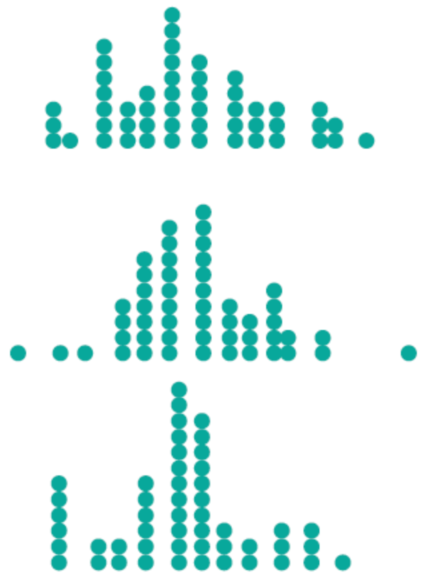
And looking at multiple models is better than looking at just one because it's difficult to know which model will match reality the closest. Even when models disagree, understanding why they are dierent can give us valuable insight.

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/.

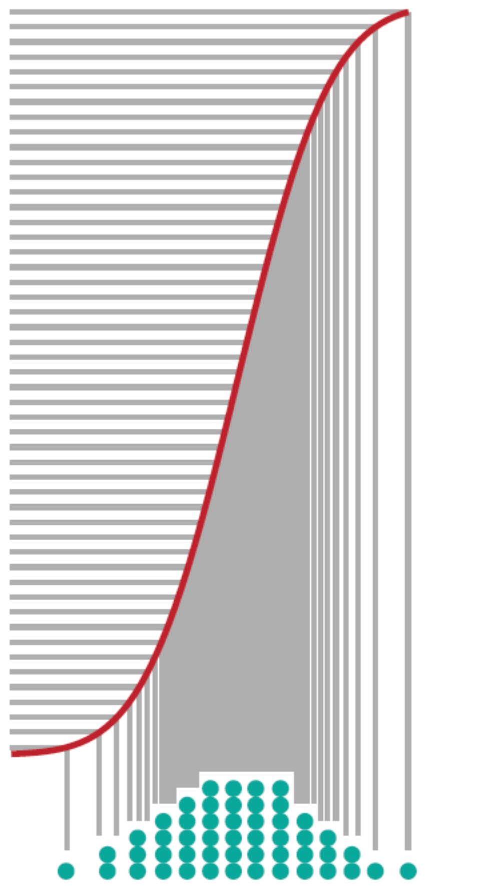
 discretizing distributions to improve decisions — quantile dot plots

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.

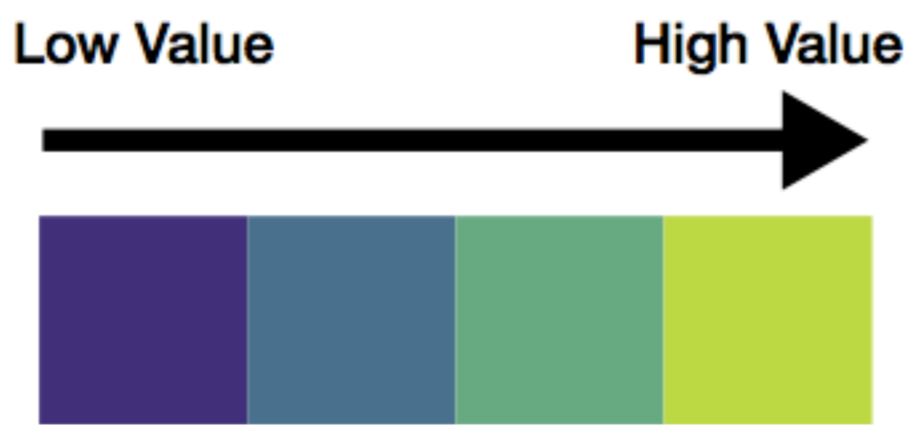


We plot the quantile “draws” using a Wilkinsonian dotplot, yielding what we call a **quantile dotplot**: a consistent discrete representation of a probability distribution.

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.

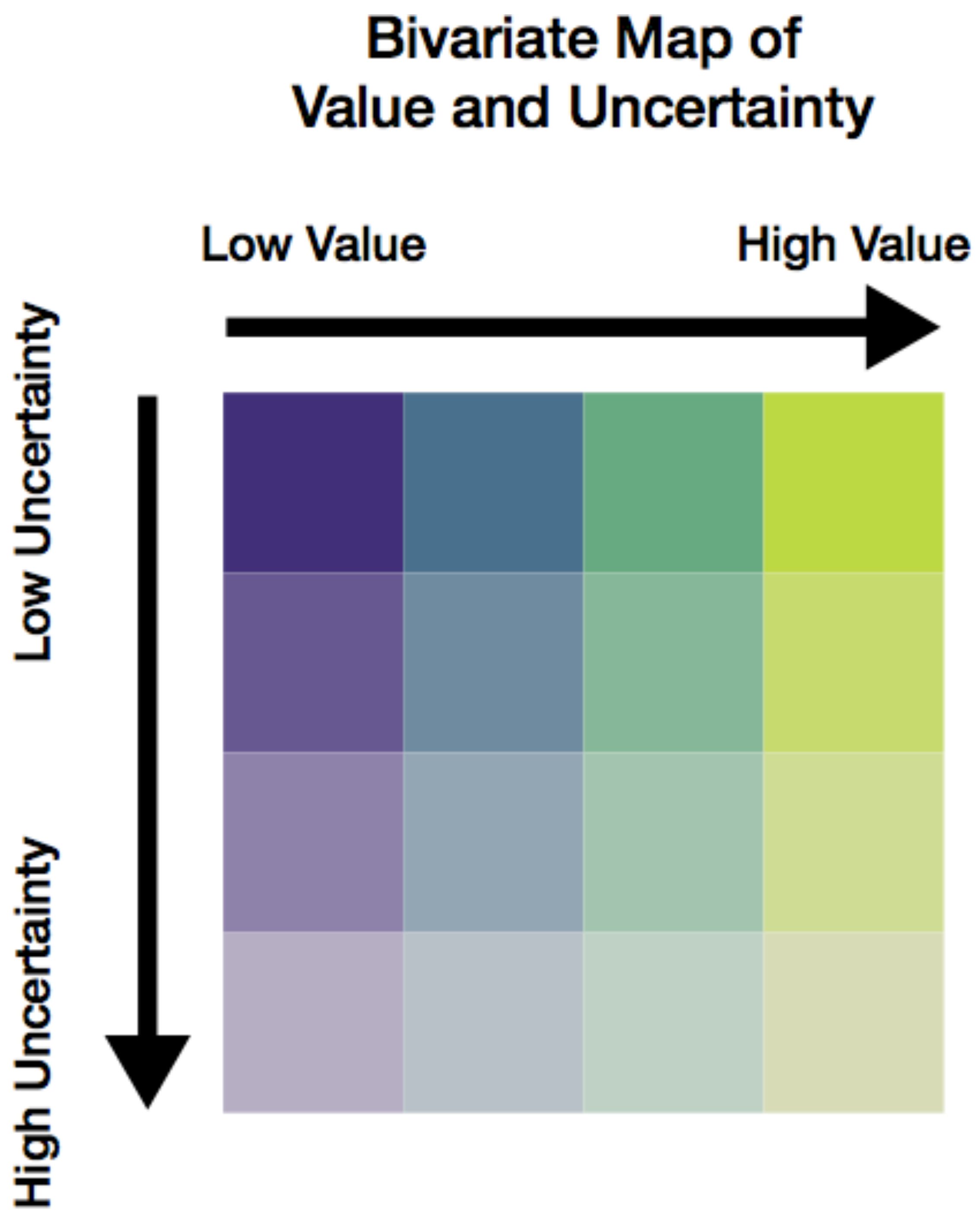
Fernandes, M., Walls, L., Munson, S., Hullman, J., & Kay, M. (2018). *Uncertainty Displays Using Quantile Dotplots or CDFs Improve Transit Decision-Making.*Â Conference on Human Factors in Computing Systems - CHI '18. doi:Â 10.1145/3173574.3173718

 using color to encode uncertainty — value suppressing uncertainty palettes



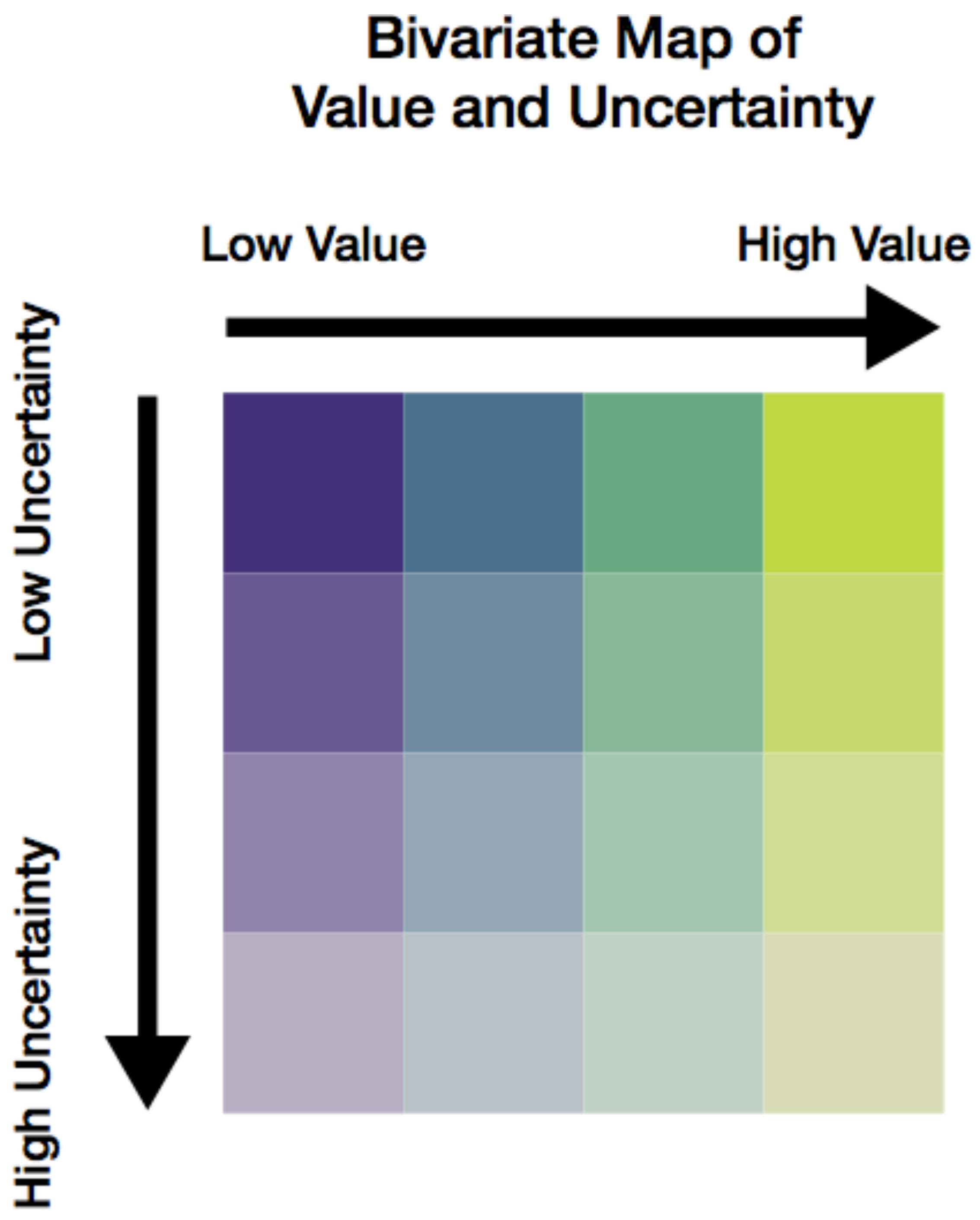
Correll, Michael, Dominik Moritz, and Jeffrey Heer. "Value-Suppressing Uncertainty Palettes." In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–11. Montreal QC, Canada: ACM Press, 2018.

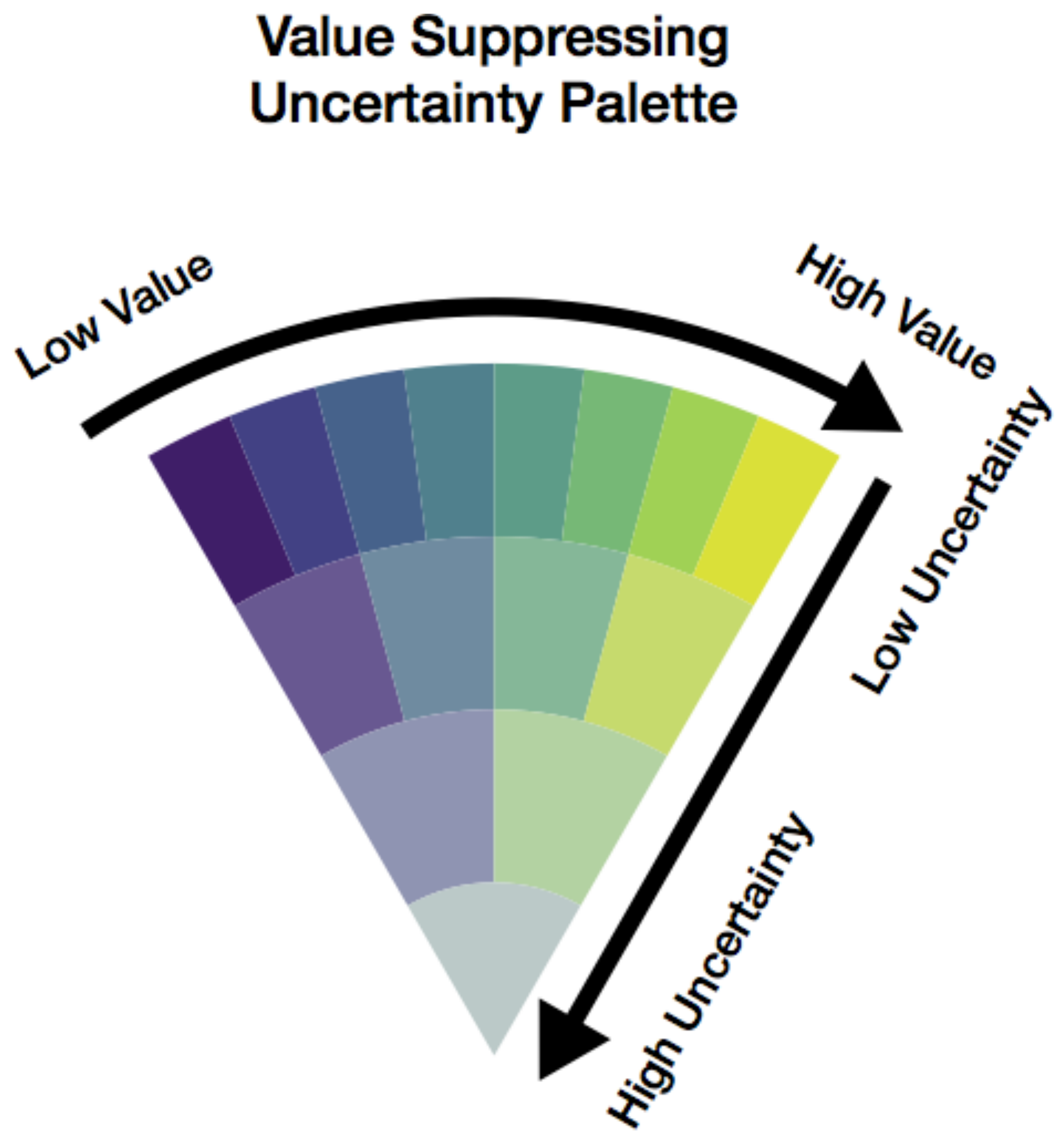
 using color to encode uncertainty — value suppressing uncertainty palettes



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.

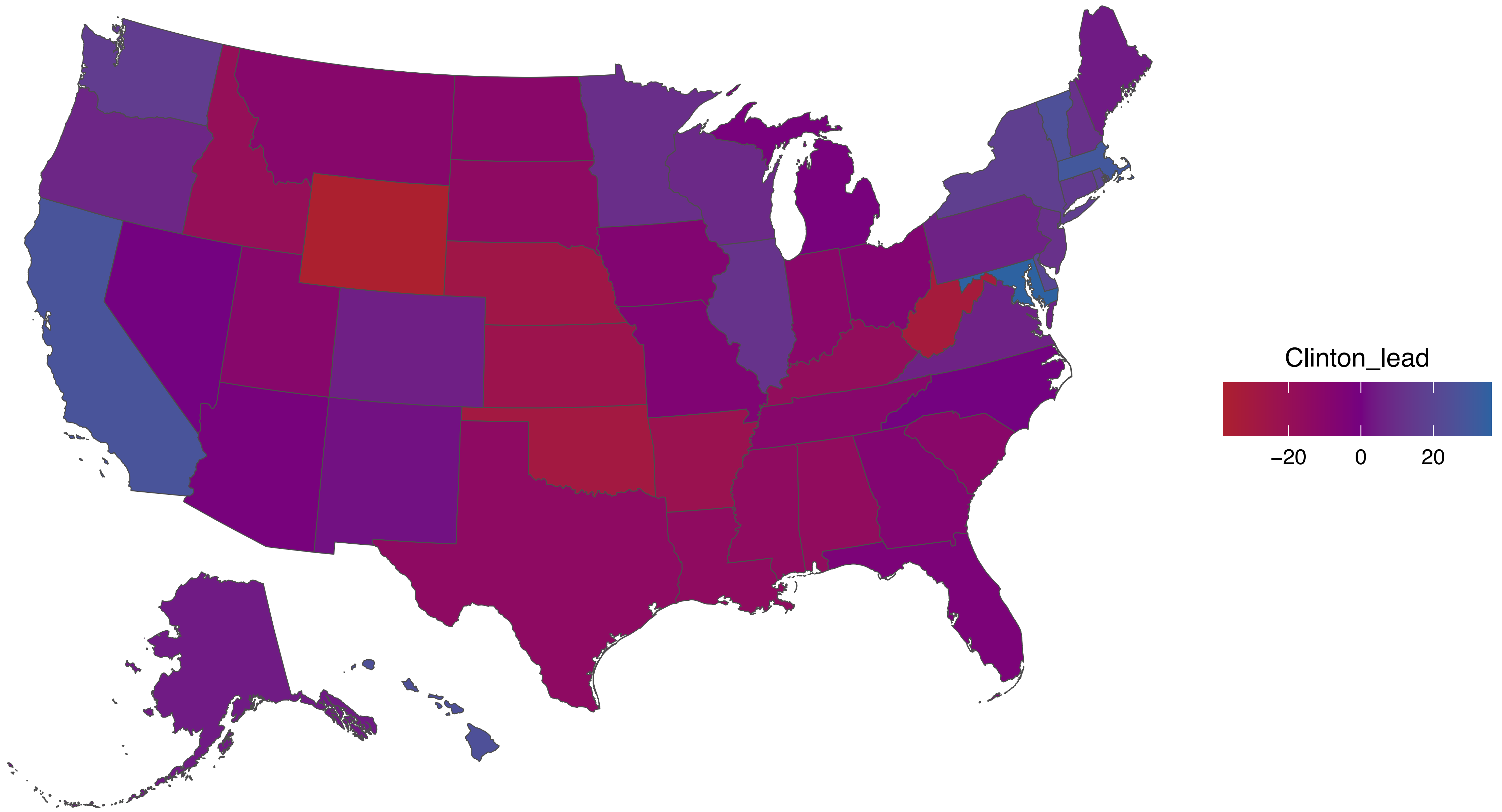
 using color to encode uncertainty — value suppressing uncertainty palettes

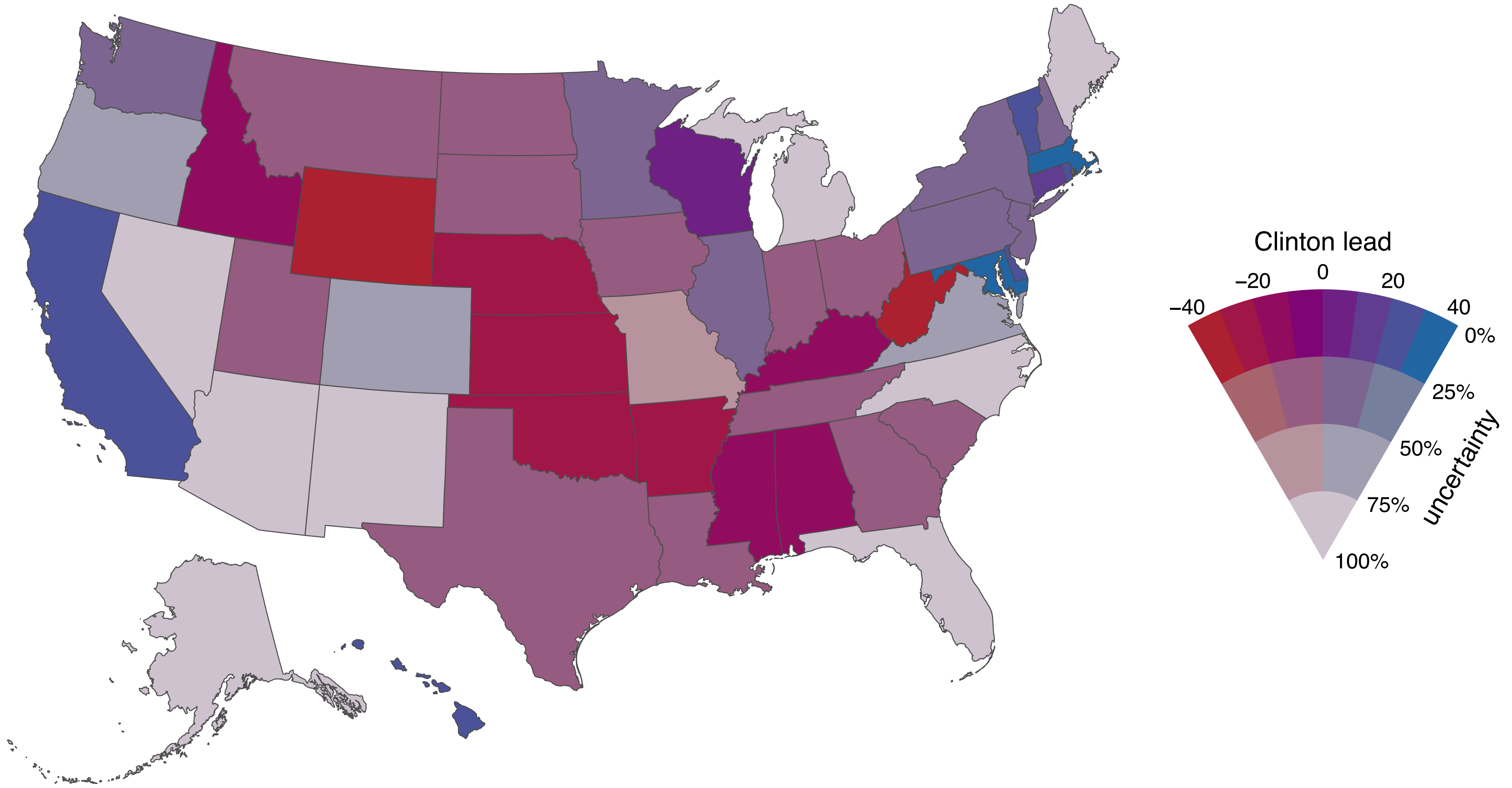




Correll, Michael, Dominik Moritz, and Jeffrey Heer. "Value-Suppressing Uncertainty Palettes." In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–11. Montreal QC, Canada: ACM Press, 2018.

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

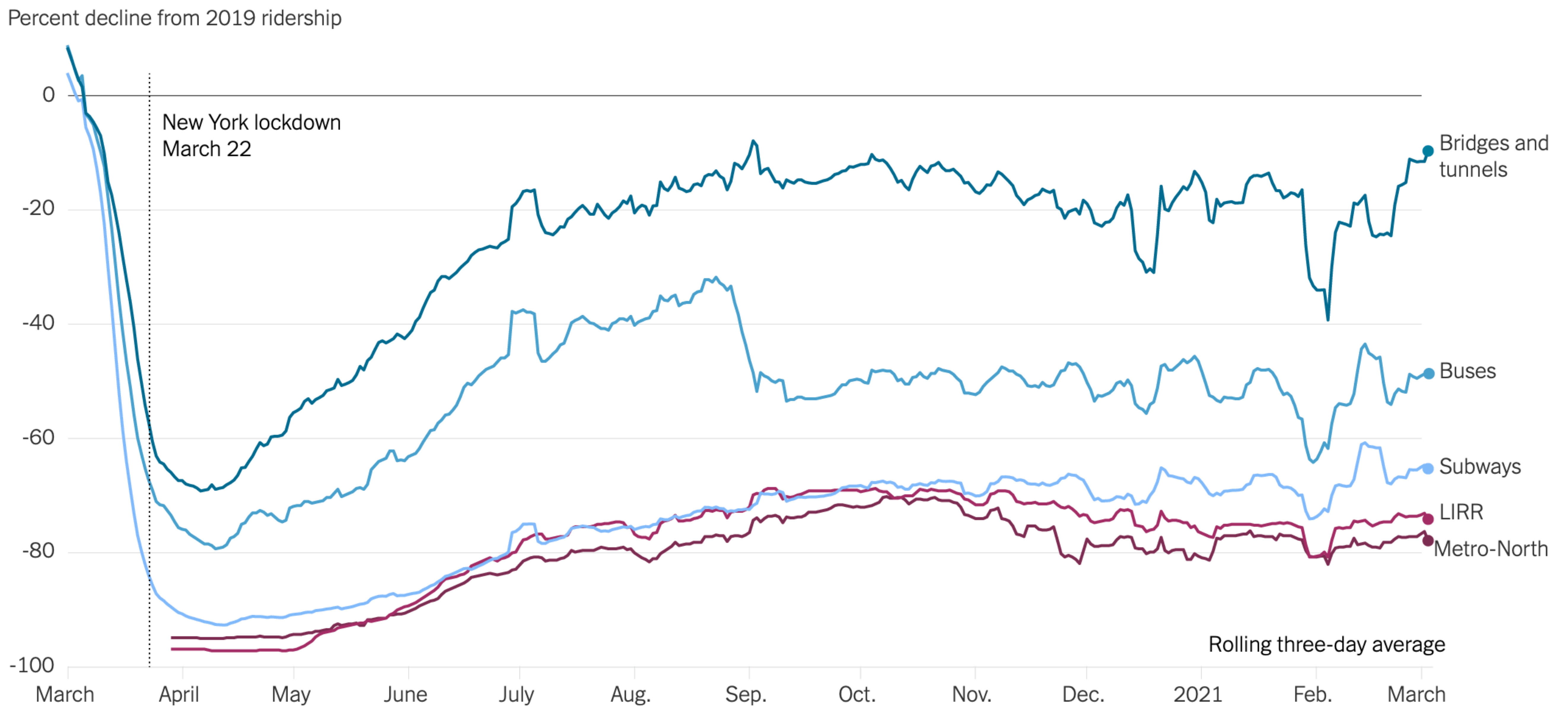




Claus O. Wilke (2021). *multiscales: Multivariate scales for 'ggplot2'*. R package version 0.1.0. https://github.com/clauswilke/multiscales

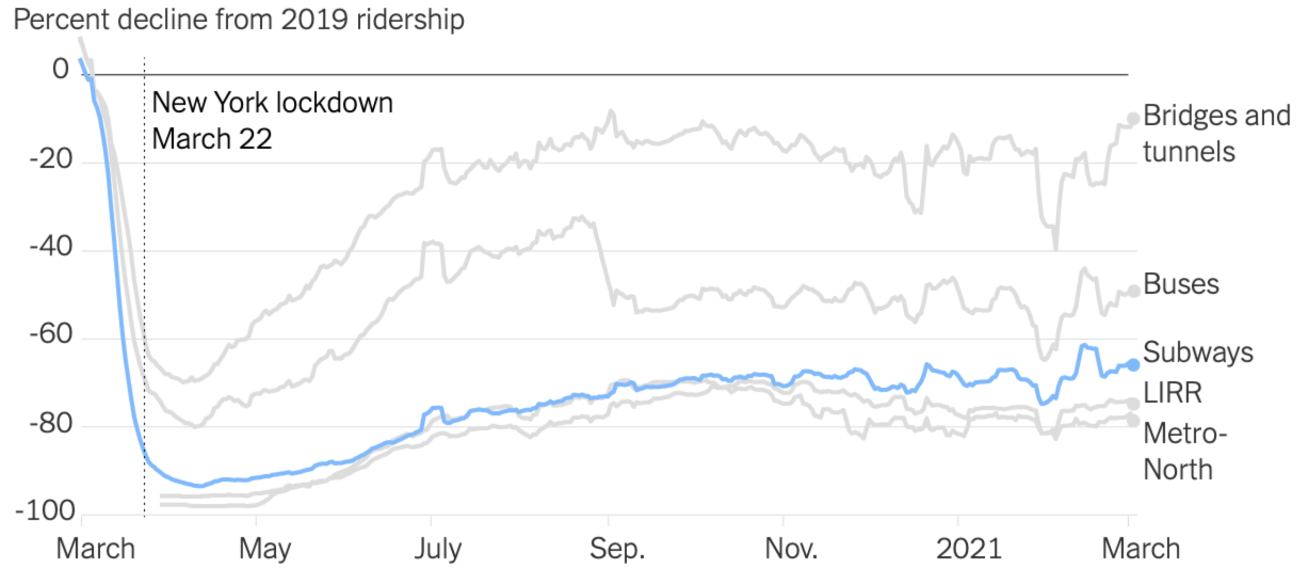
pacing for attention

pacing for attention, you can focus on consecutive layers of a graphic *spatially (multiples)*



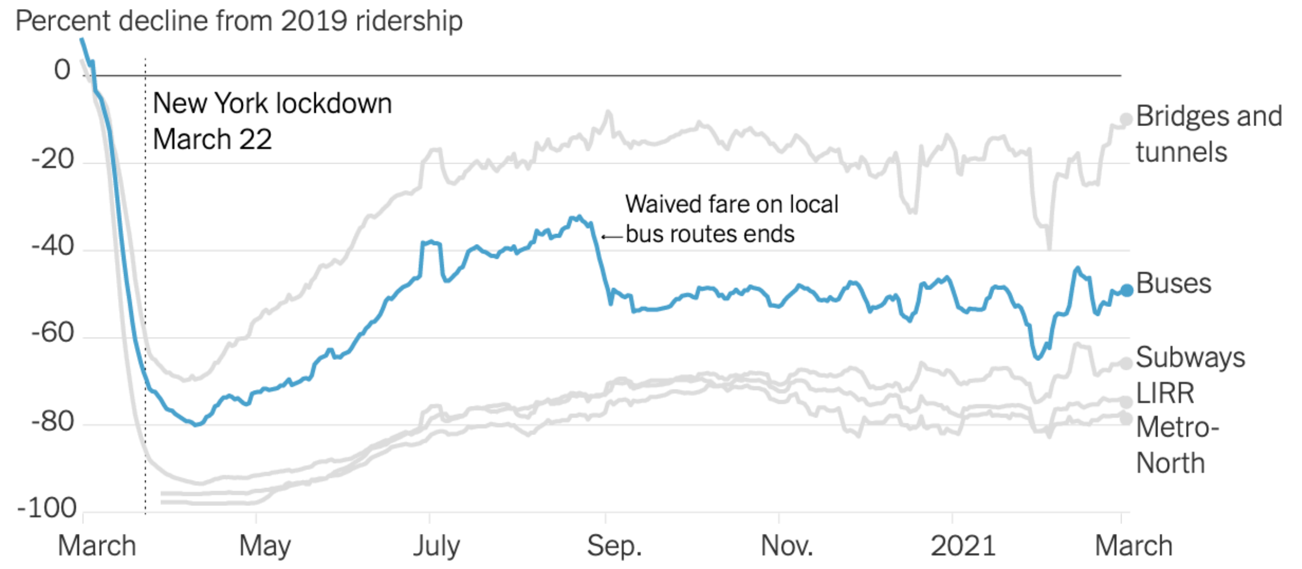
By Veronica Penney • Source: The M.T.A.'s day-by-day ridership numbers. | Note: Percent change is calculated as a comparison to the preceding-year equivalent day, with the exception of the commuter rail systems, which are compared to the 2019 monthly weekday/Saturday/Sunday average.

Subway Ridership Is Slow to Recover



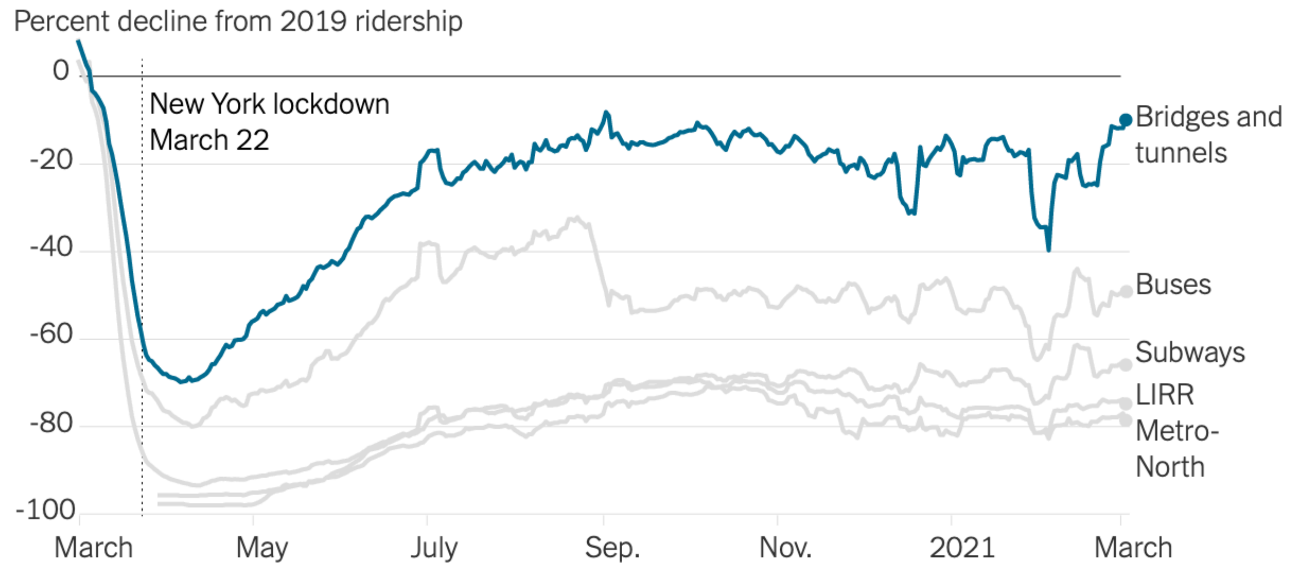
By Veronica Penney • Source: The M.T.A.'s day-by-day ridership numbers.

The Pandemic Cut Bus Ridership by Half



By Veronica Penney • Source: The M.T.A.'s day-by-day ridership numbers.

Car Travel Is Near Pre-Pandemic Levels



By Veronica Penney • Source: The M.T.A.'s day-by-day ridership numbers.

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 *temporally* — a grammar of animated graphics

gganimate 1.0.5.9000 | Getting Started | Reference | Talks | News | Search...

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 later, `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.

A Grammar of Animated Graphics

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/accurat-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?

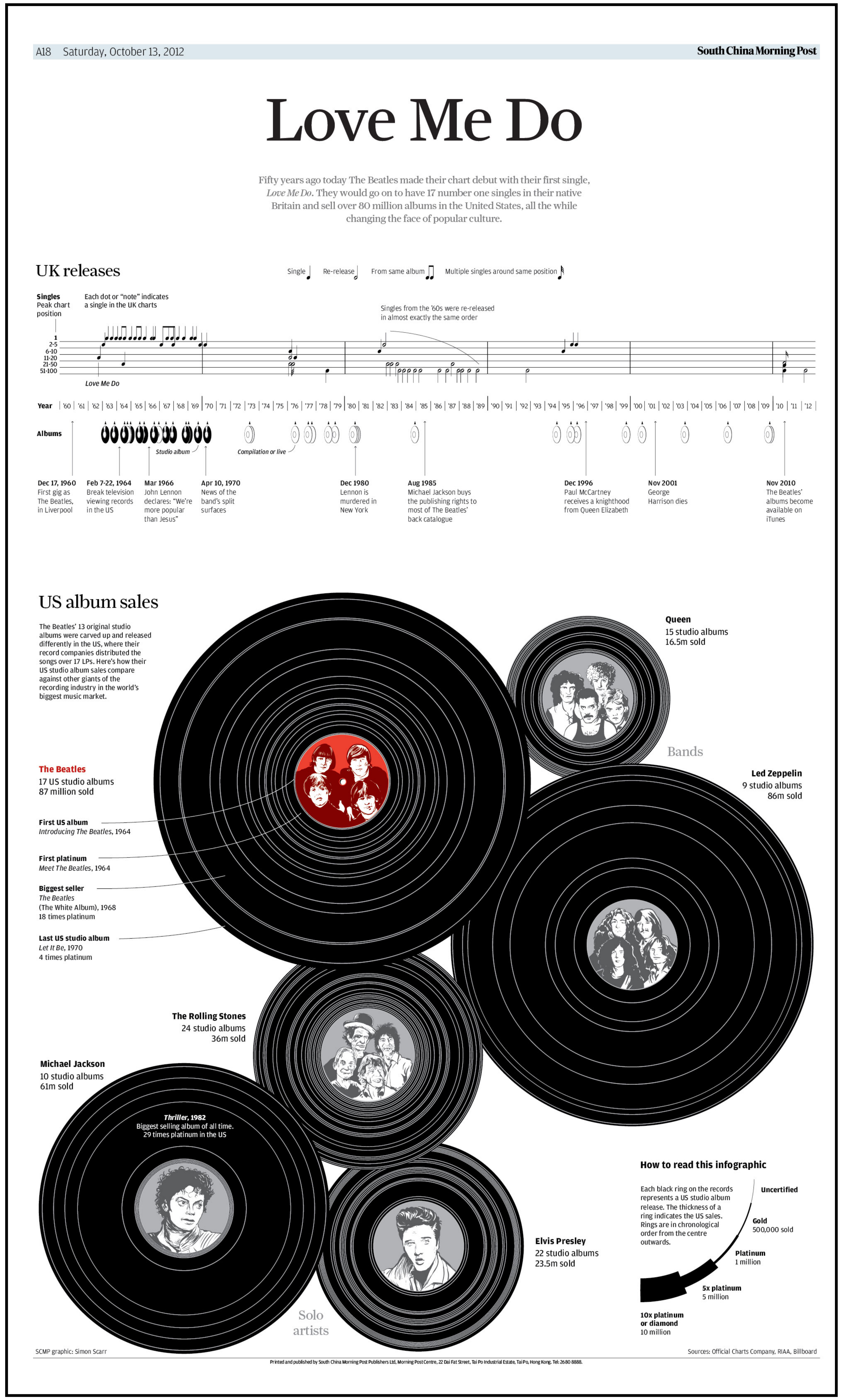
Narrative?

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.



criticism for visuals, **practicing critiques**

Audience?

Purpose?

Data encodings, decodings?

Comparison or change?

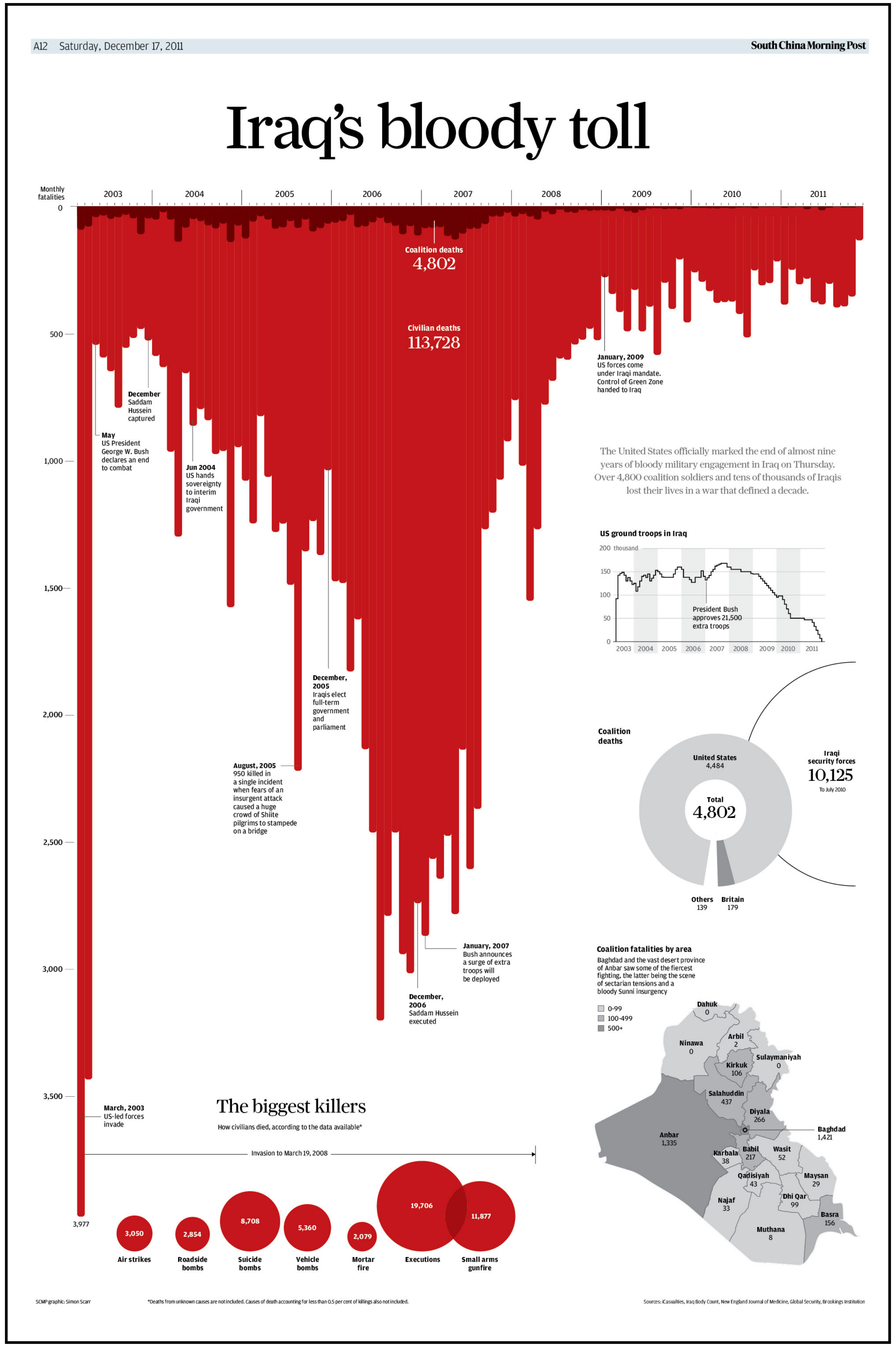
Narrative?

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.



criticism for visuals, **practicing critiques**

Audience?

Purpose?

Data encodings, decodings?

Comparison or change?

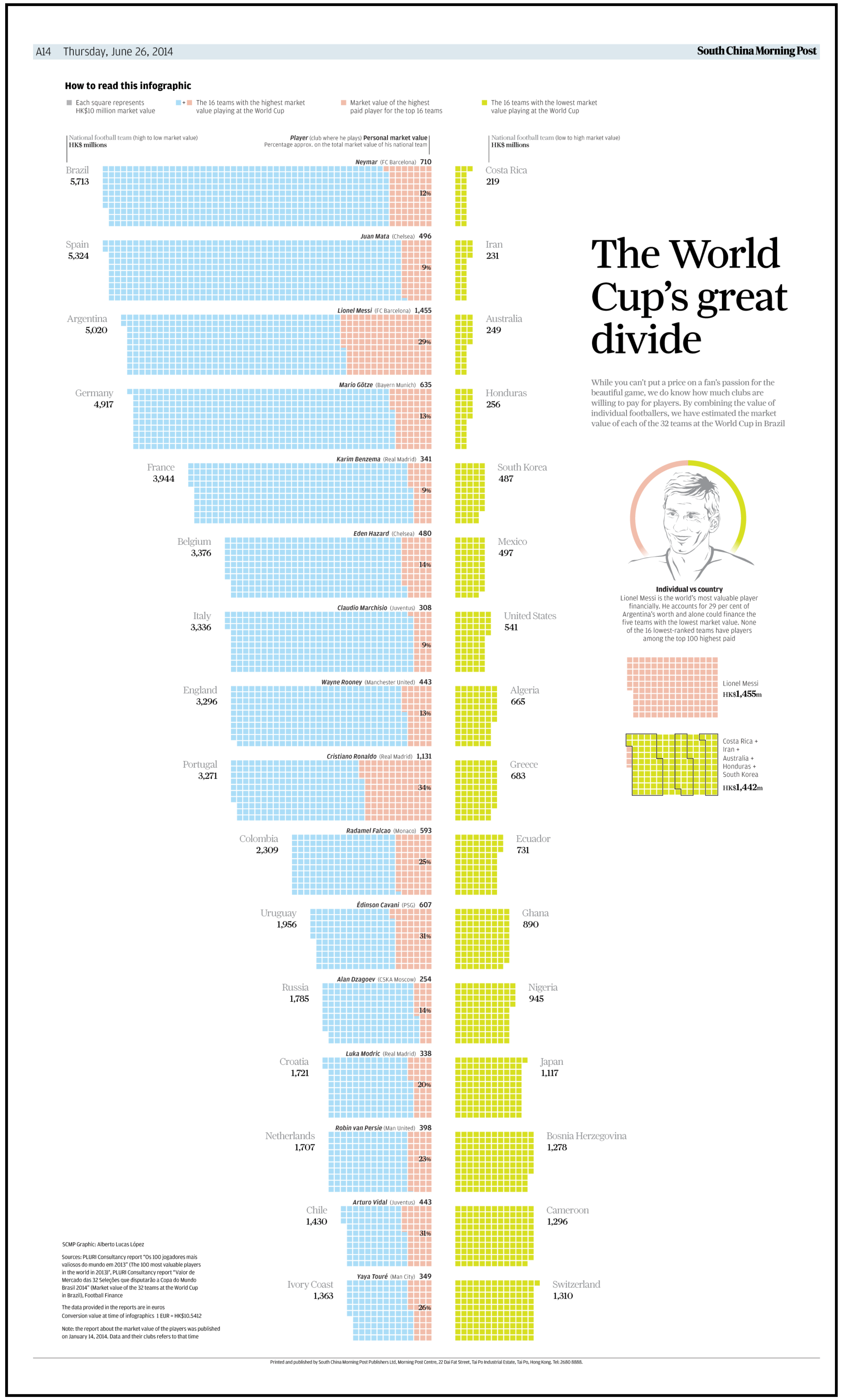
Narrative?

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.



Audience?

Purpose?

Data encodings, decodings?

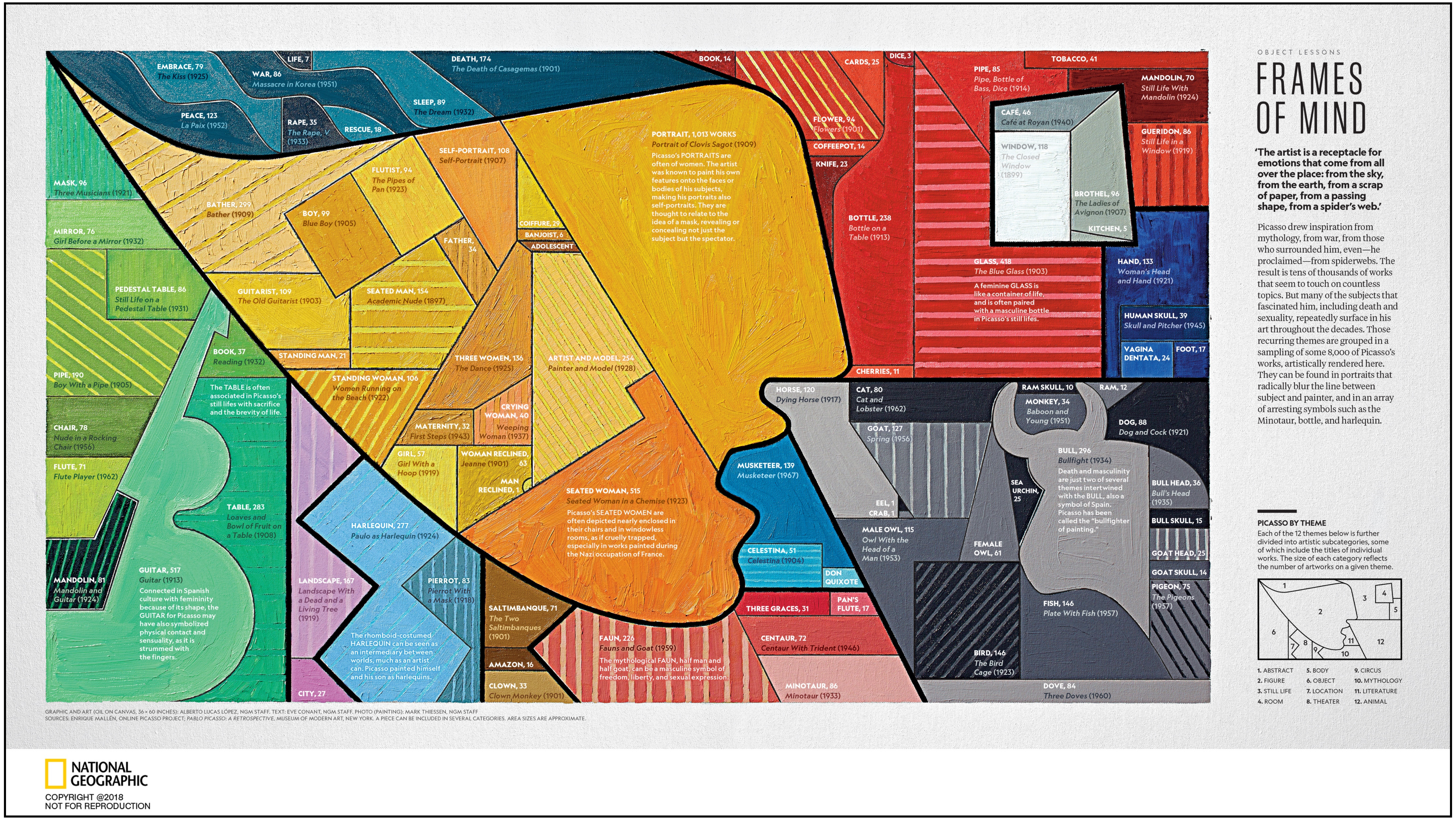
Comparison or change?

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+.

Audience?

Purpose?

Data encodings, decodings?

Comparison or change?

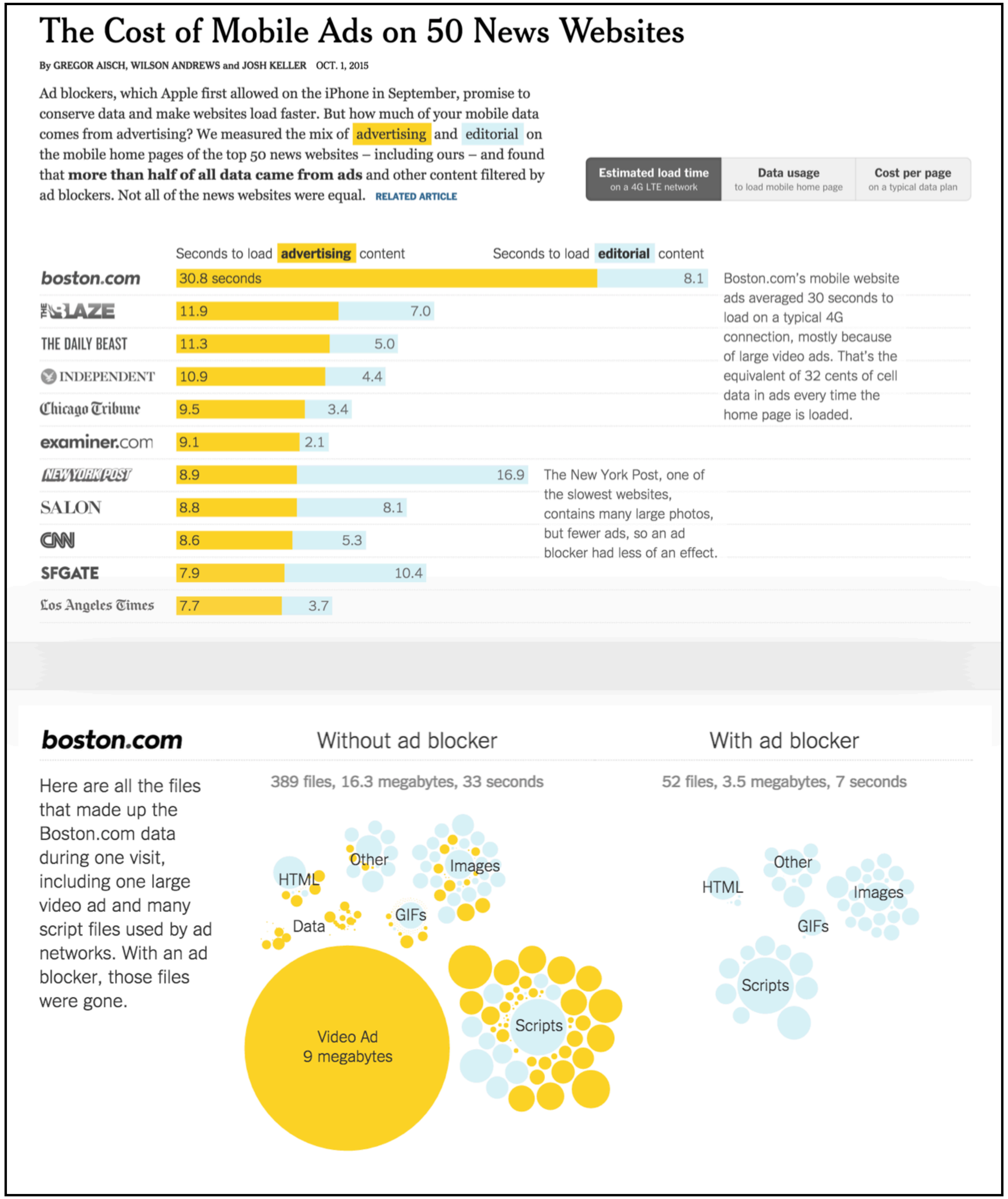
Narrative?

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.



Audience?

Purpose?

Data encodings, decodings?

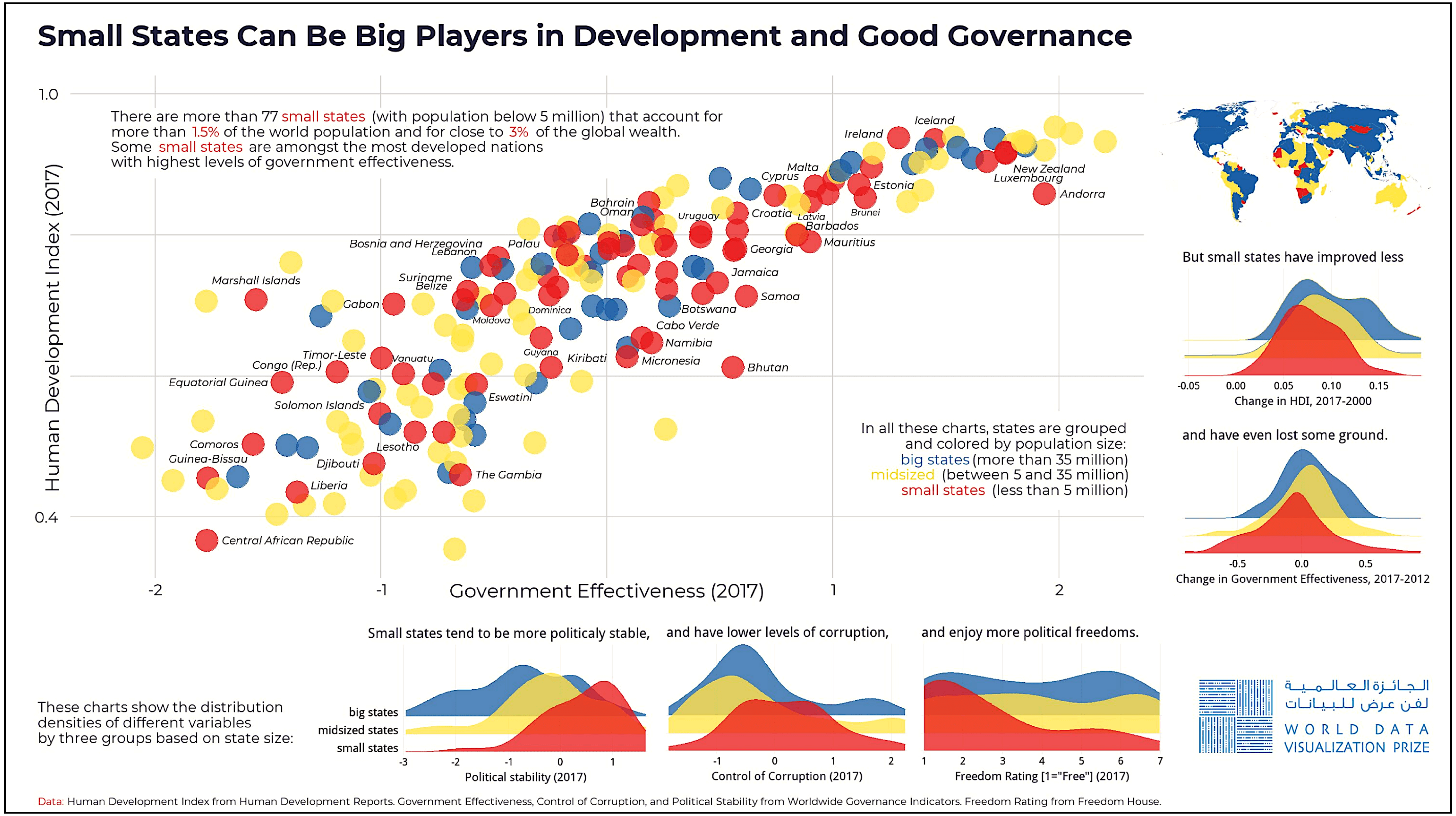
Comparison or change?

Narrative?

Color, coherency?

Hierarchy, layering, layout?

Credibility, transparency?



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

Audience?

Purpose?

Data encodings, decodings?

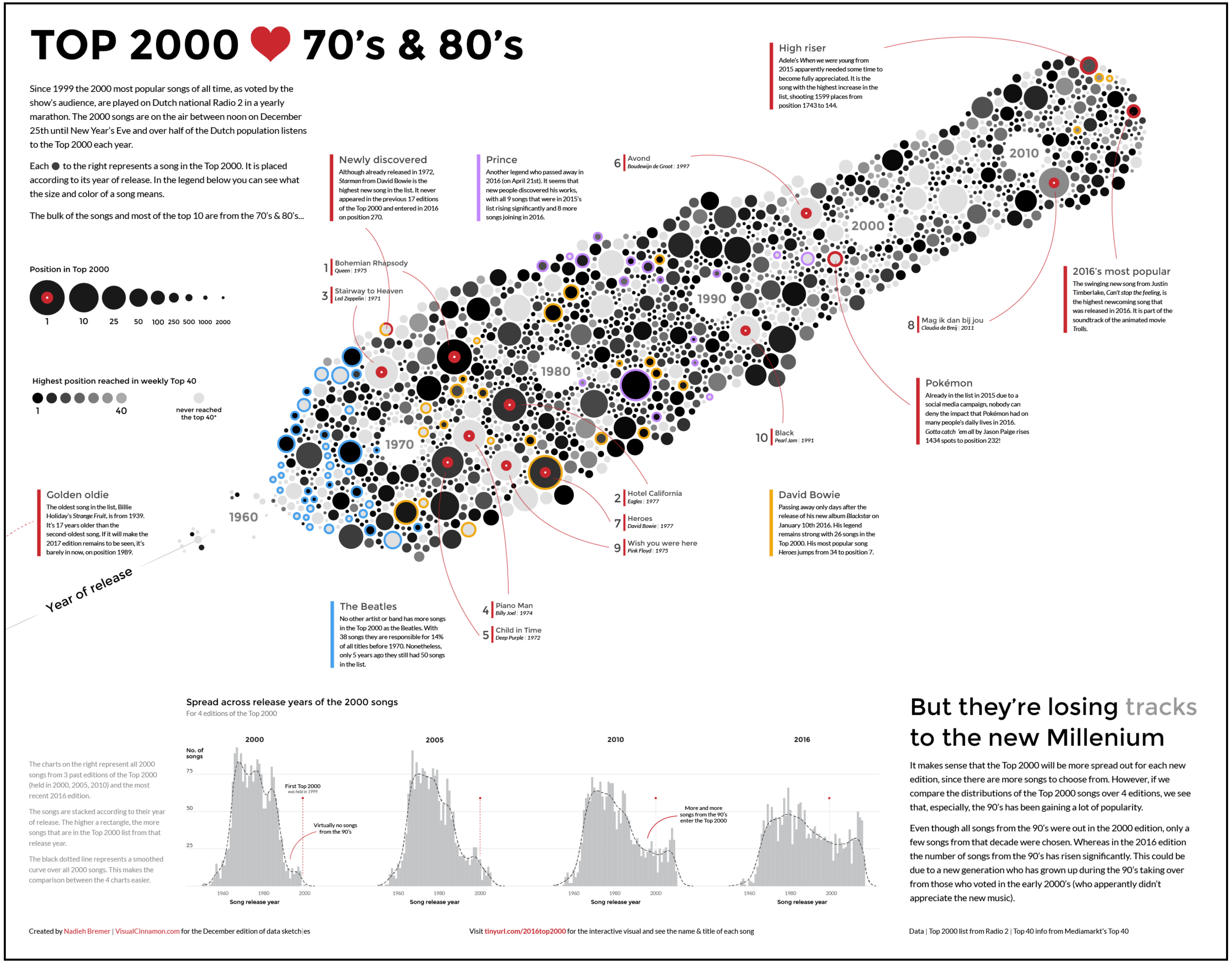
Comparison or change?

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.

Audience?

Purpose?

Data encodings, decodings?

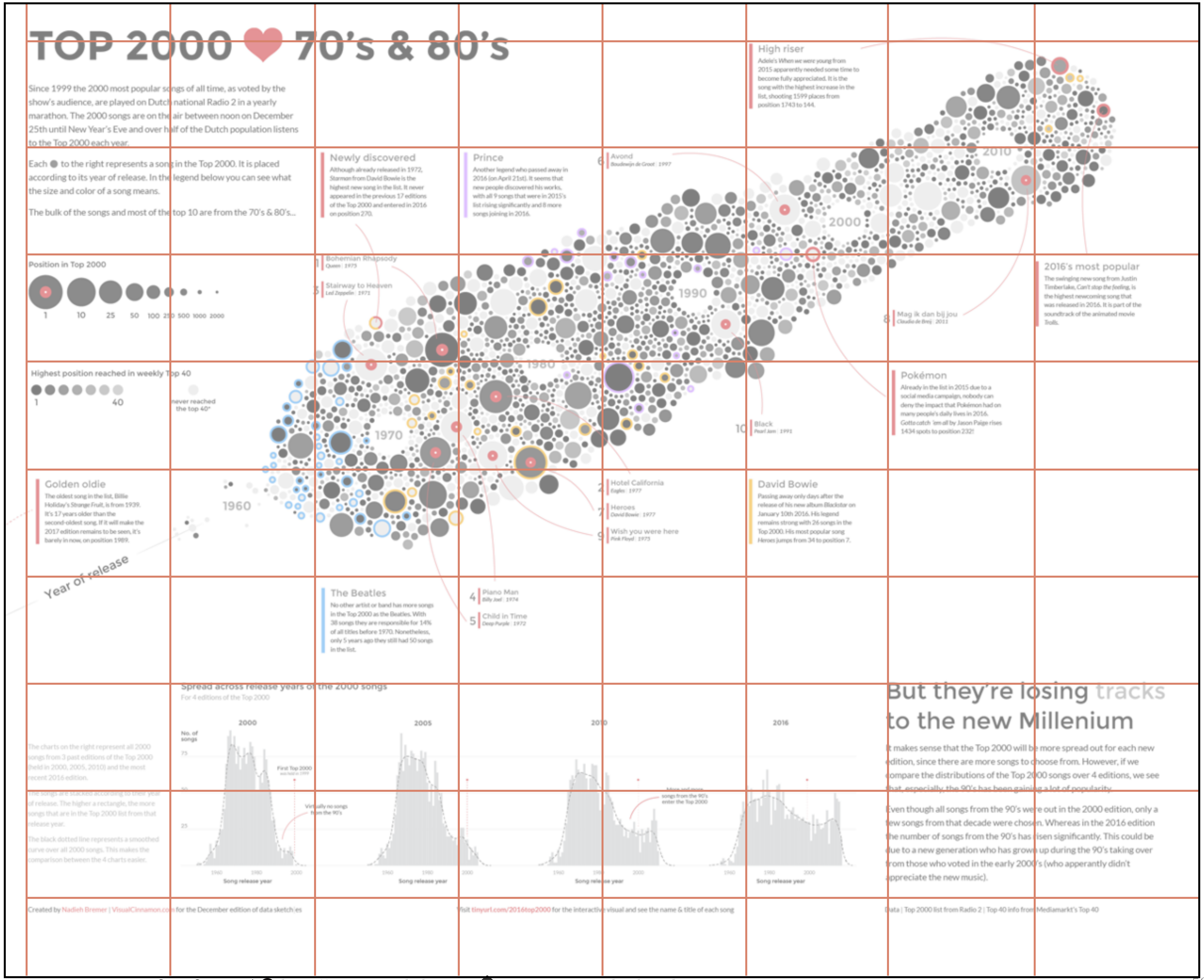
Comparison or change?

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.

Audience?

Purpose?

Data encodings, decodings?

Comparison or change?

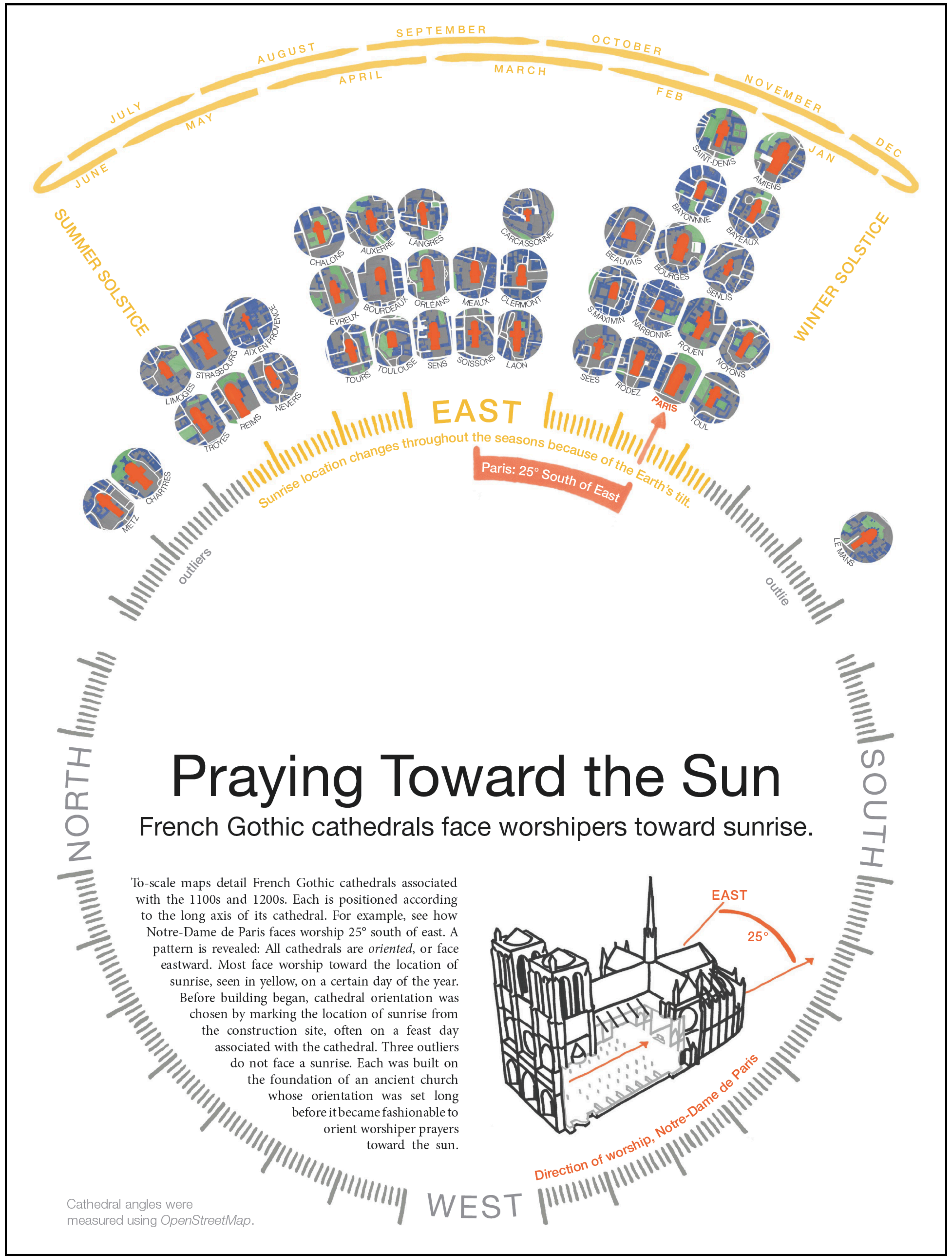
Narrative?

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?

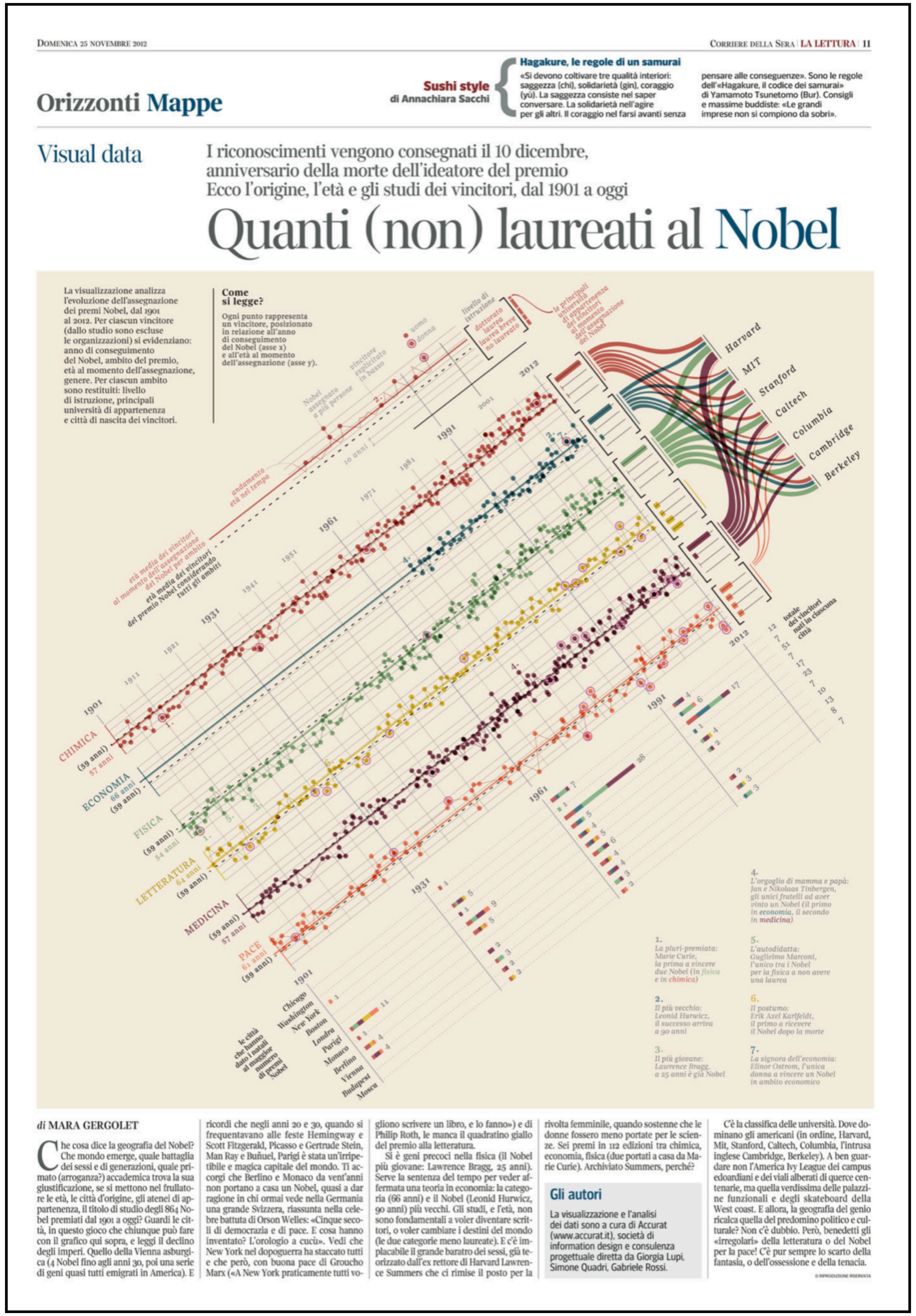
Comparison or change?

Narrative?

Color, coherency?

Hierarchy, layering, layout?

Credibility, transparency?


Domenica 25 Novembre 2012

Corriere della Sera | La Lettura | 11

Orizzonti Mappe

Sushi style

di Annachiara Sacchi

Hagakure, le regole di un samurai

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

Gli autori

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

Audience?

Purpose?

Data encodings, decodings?

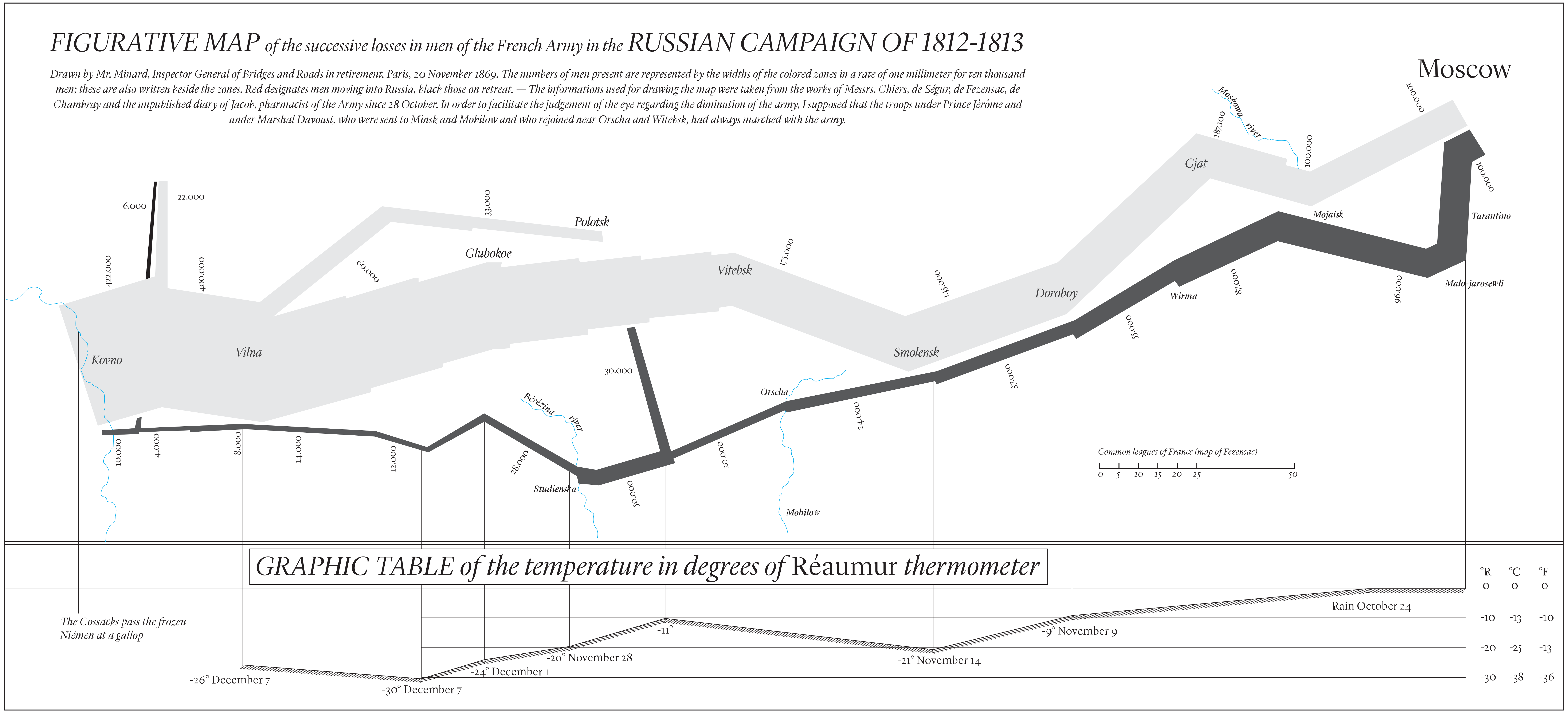
Comparison or change?

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.

Audience?

Purpose?

Data encodings, decodings?

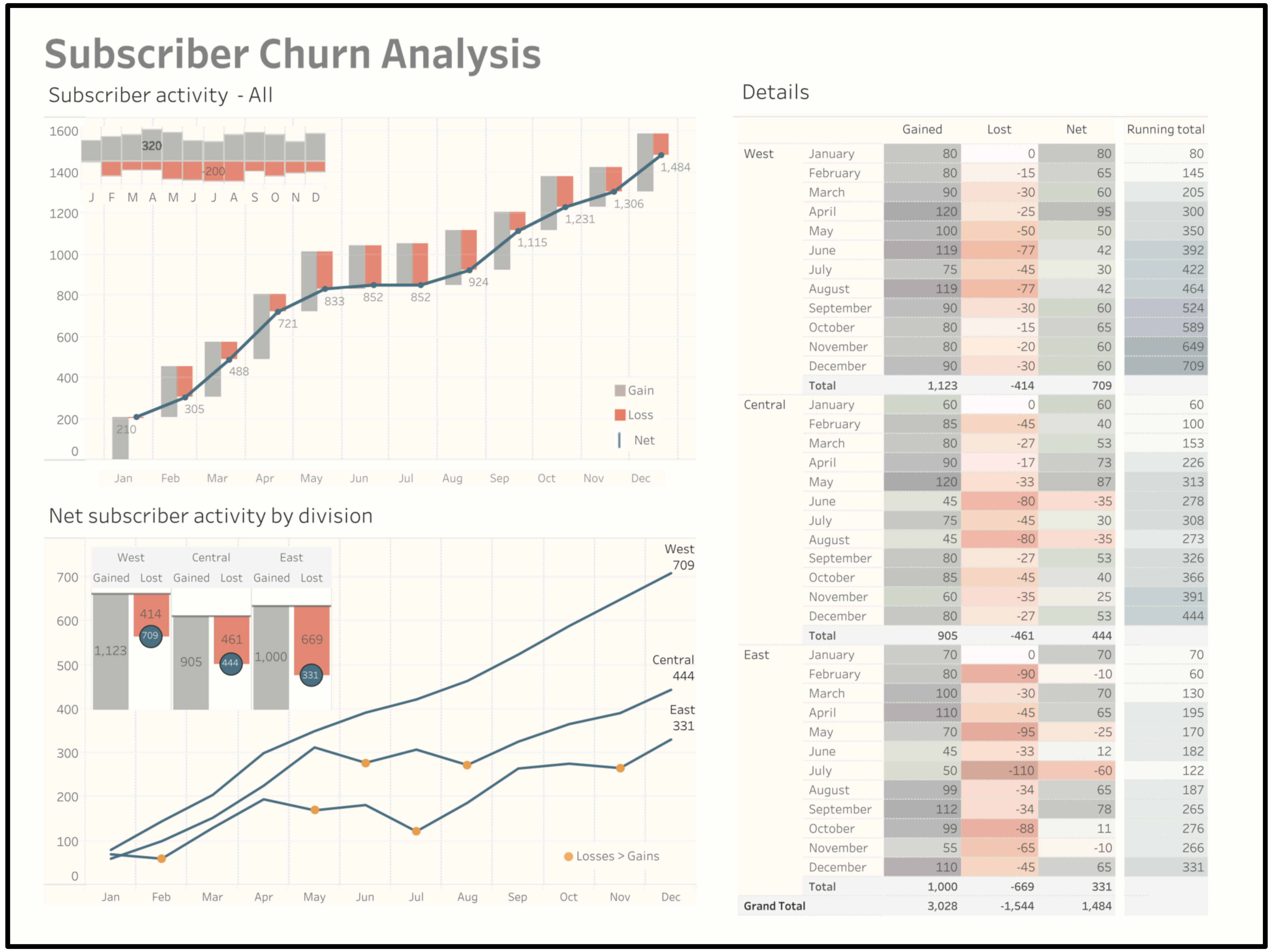
Comparison or change?

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?

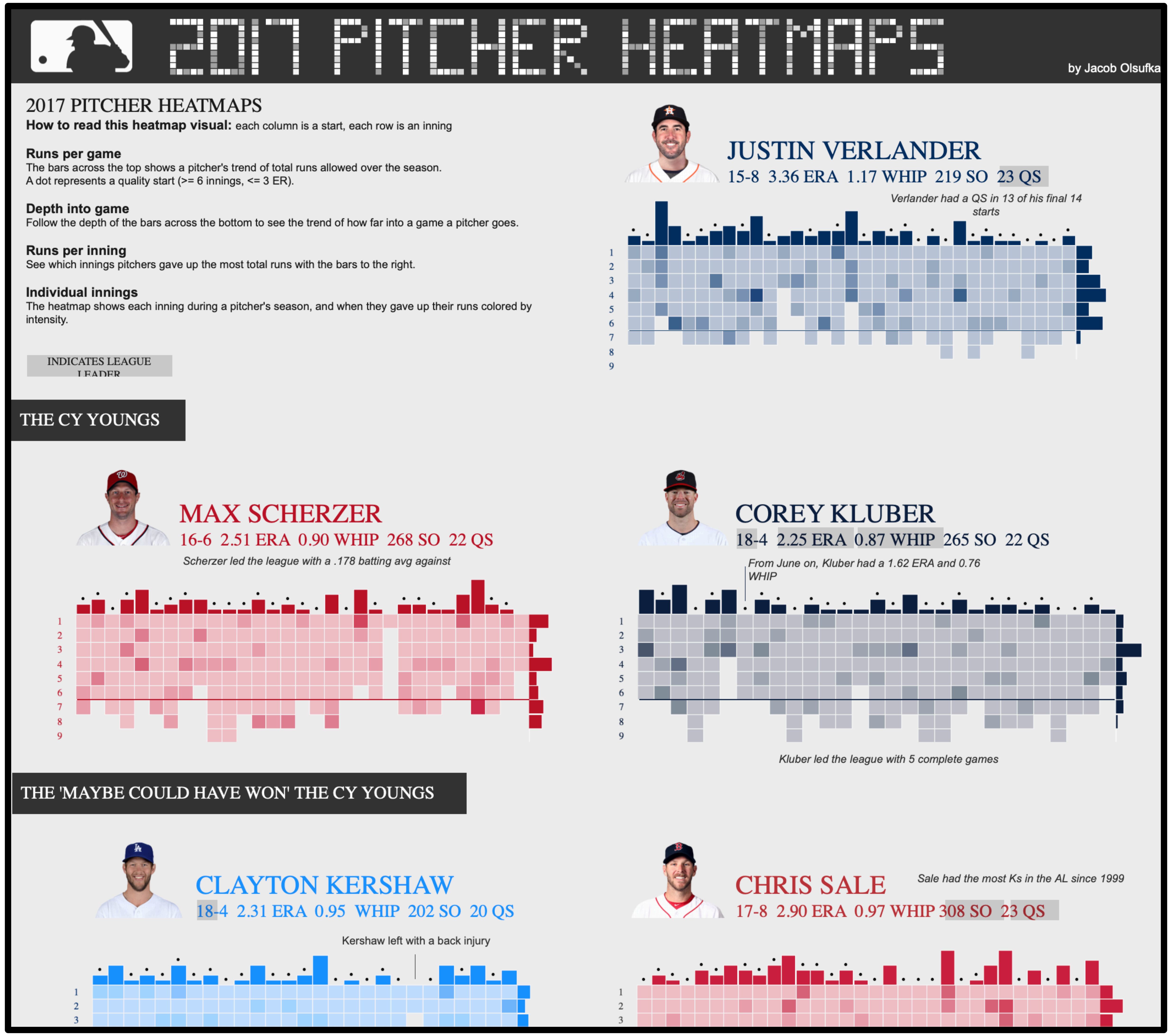
Narrative?

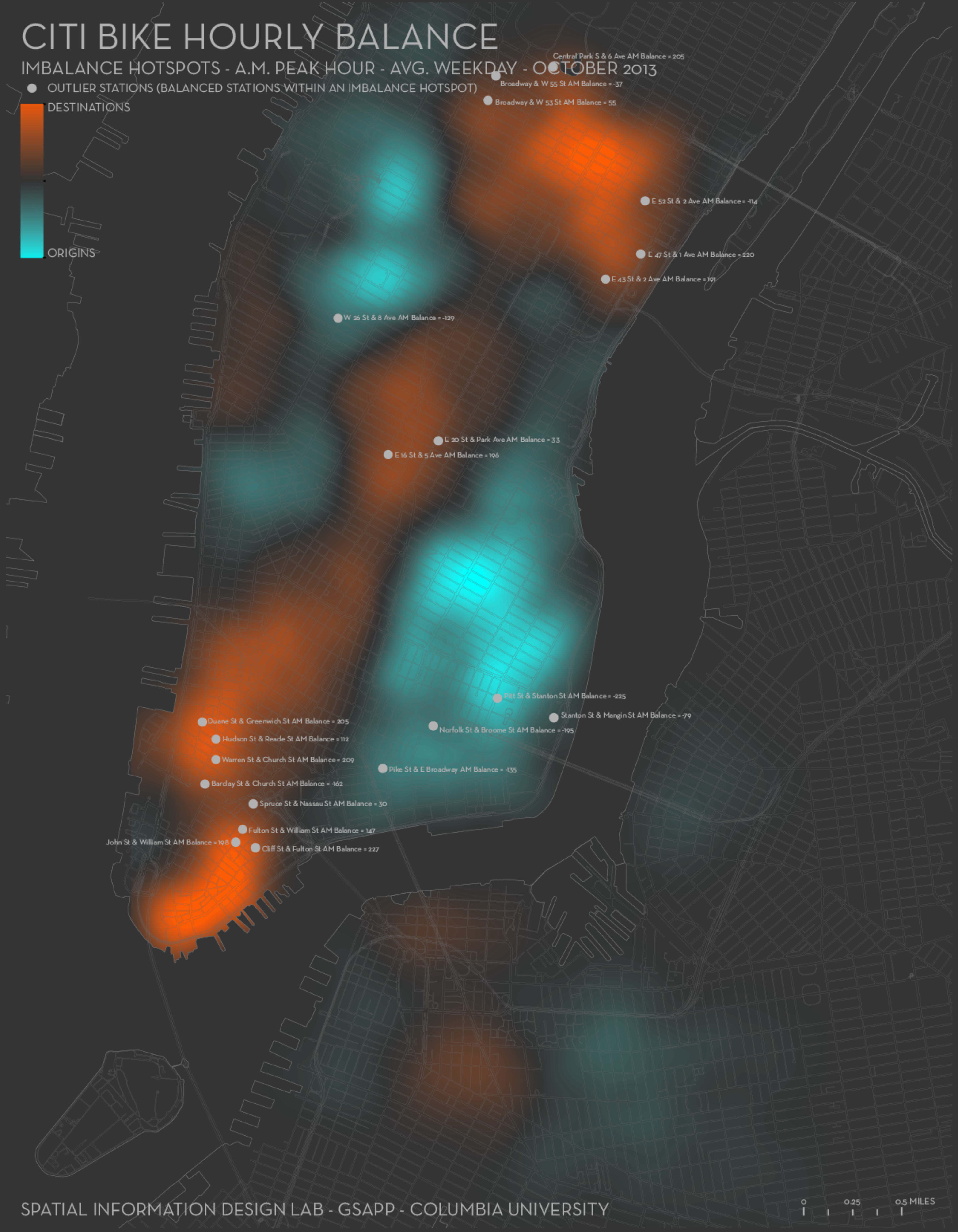
Color, coherency?

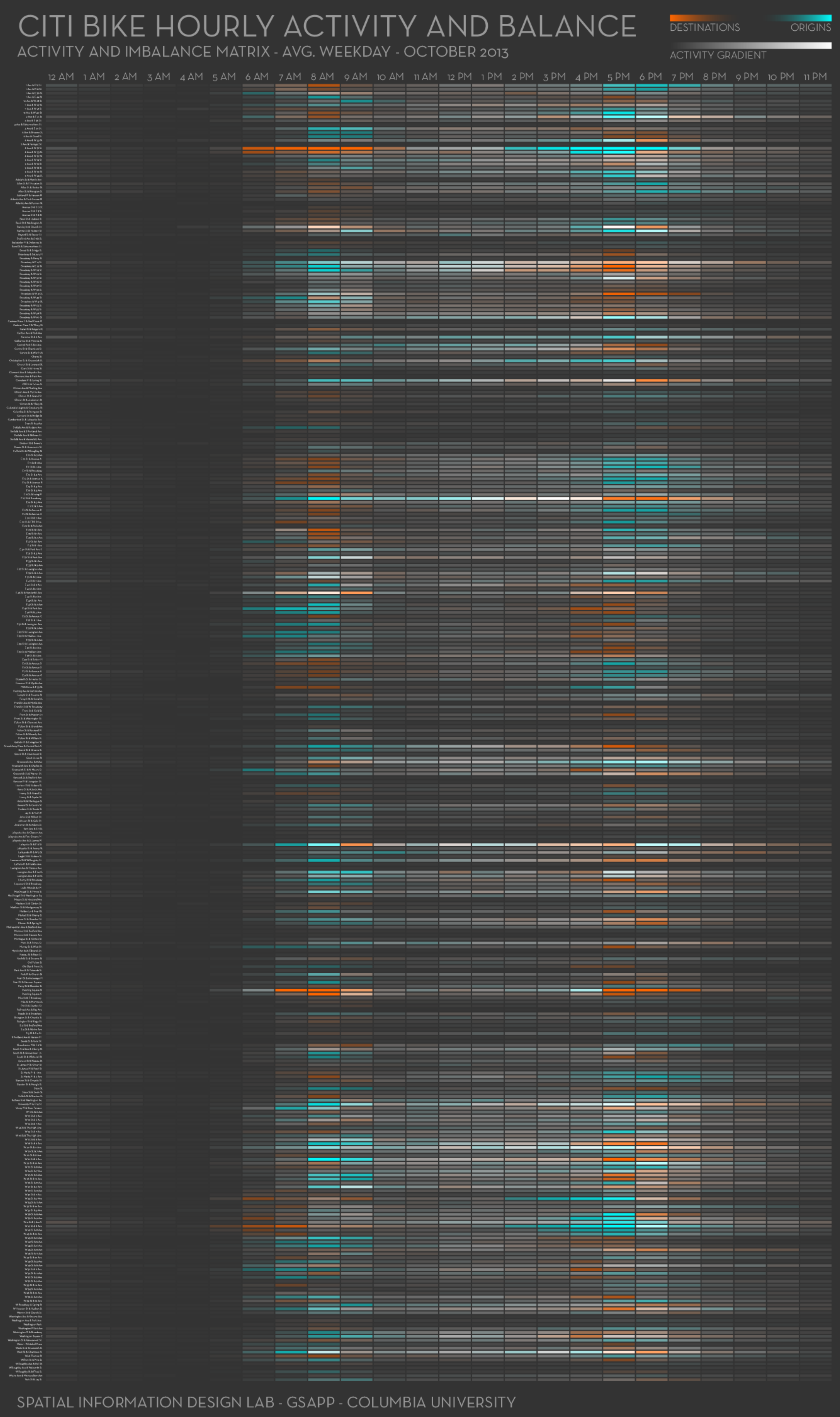
Hierarchy, layering, layout?

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




CITI BIKE HOURLY BALANCE
IMBALANCE HOTSPOTS - A.M. PEAK HOUR - AVG. WEEKDAY - OCTOBER 2013
OUTLIER STATIONS (BALANCED STATIONS WITHIN AN IMBALANCE HOTSPOT)
DESTINATIONS
ORIGINS
Central Park S & 6 Ave AM Balance = 205
Broadway & W 55 St AM Balance = -37
Broadway & W 53 St AM Balance = 55
E 52 St & 2 Ave AM Balance = -114
E 47 St & 1 Ave AM Balance = 220
E 43 St & 2 Ave AM Balance = 191
W 26 St & 8 Ave AM Balance = -129
E 20 St & Park Ave AM Balance = 33
E 16 St & 5 Ave AM Balance = 196
Pitt St & Stanton St AM Balance = -225
Stanton St & Mangin St AM Balance = -79
Norfolk St & Broome St AM Balance = -195
Duane St & Greenwich St AM Balance = 205
Hudson St & Reade St AM Balance = 112
Warren St & Church St AM Balance = 209
Pike St & E Broadway AM Balance = -135
Barclay St & Church St AM Balance = -162
Spruce St & Nassau St AM Balance = 30
Fulton St & William St AM Balance = 147
John St & William St AM Balance = 198
Cliff St & Fulton St AM Balance = 227
SPATIAL INFORMATION DESIGN LAB - GSAPP - COLUMBIA UNIVERSITY
0
0.25
0.5 MILES


CITI BIKE HOURLY ACTIVITY AND BALANCE
ACTIVITY AND IMBALANCE MATRIX - AVG. WEEKDAY - OCTOBER 2013
DESTINATIONS
ORIGINS
ACTIVITY GRADIENT
12 AM
1 AM
2 AM
3 AM
4 AM
5 AM
6 AM
7 AM
8 AM
9 AM
10 AM
11 AM
12 PM
1 PM
2 PM
3 PM
4 PM
5 PM
6 PM
7 PM
8 PM
9 PM
10 PM
11 PM
SPATIAL INFORMATION DESIGN LAB - GSAPP - COLUMBIA UNIVERSITY