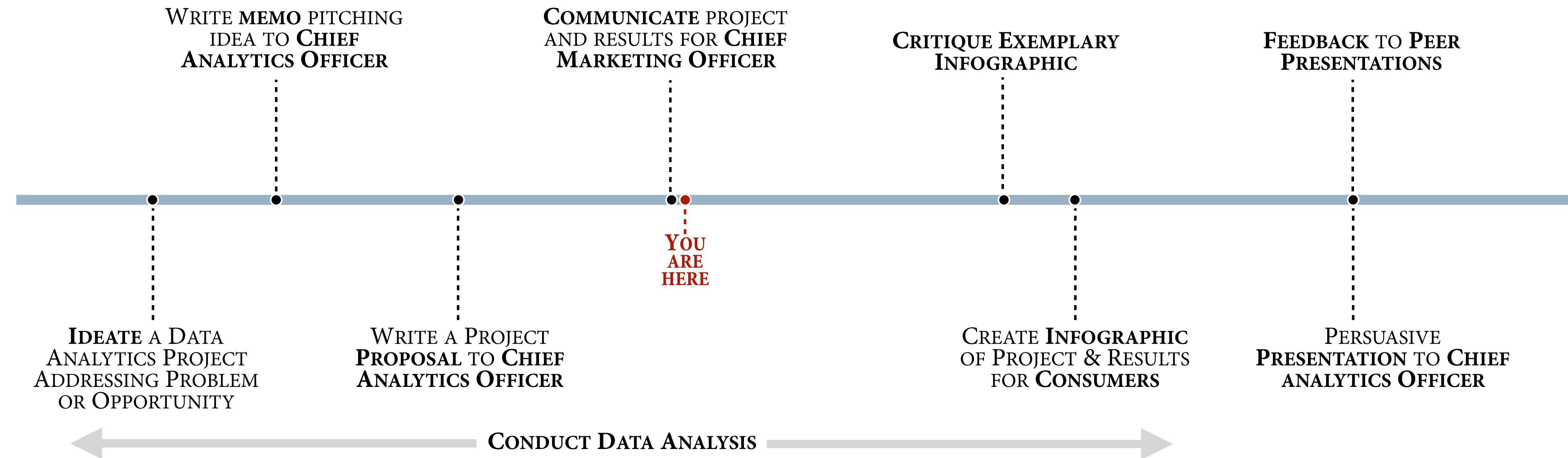


Storytelling With Data

# **The storytelling process, and designing data graphics**

# Conceptual project timeline



## **(more on) storytelling**

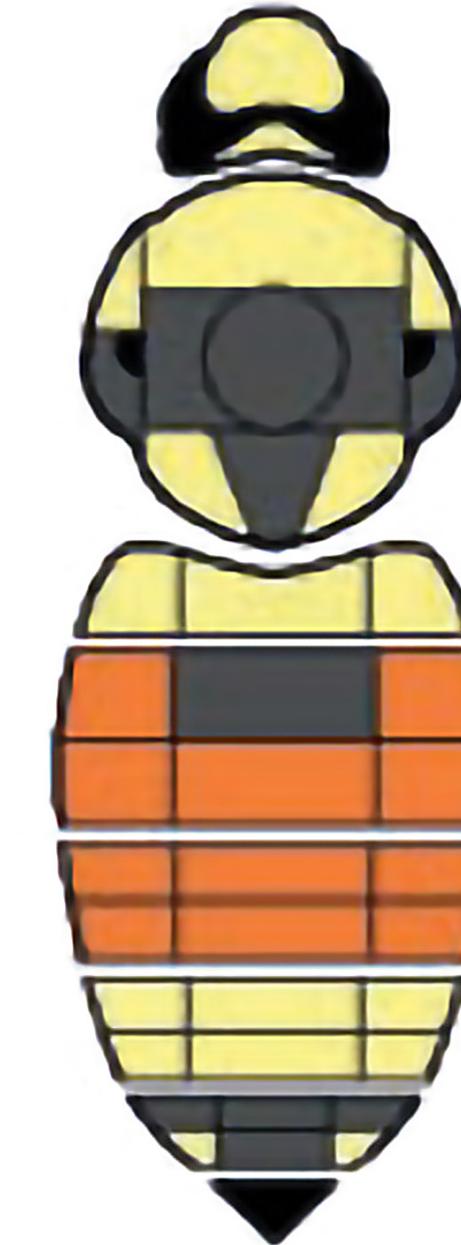
# See, Think, Design, Produce

understand      explain



---

Corum, Jonathan



### Search for patterns by comparing

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

### Sketch until your aha! moment

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

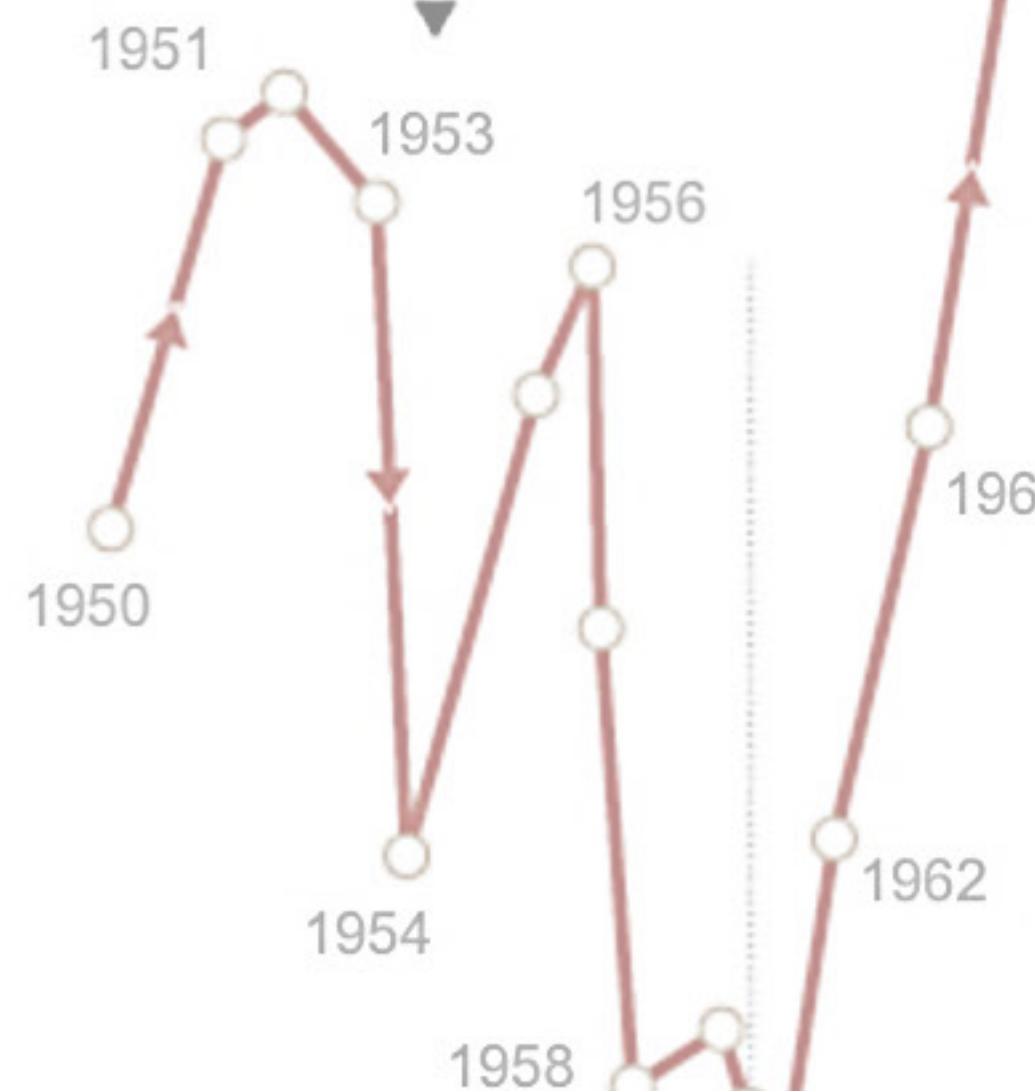
# See, Think, Design, Produce

understand

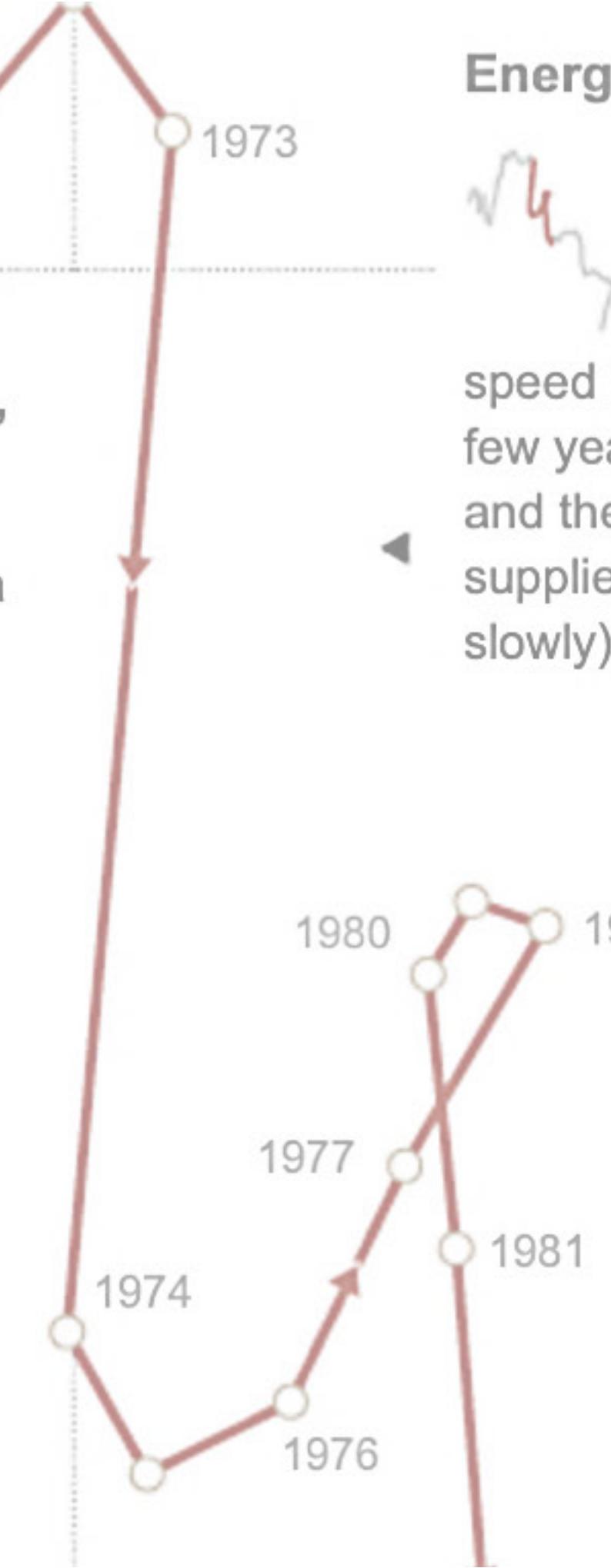
explain

25

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



▲ “Unsafe at Any Speed”  
In 1965, Ralph Nader publishes a best seller about auto companies’ resisting safety features. The government creates the first agency devoted to highway safety. Auto fatalities hit a plateau.



## Energy Crises

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

## Seat Belts and Sobriety

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

**Design for someone else, show varying details**

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

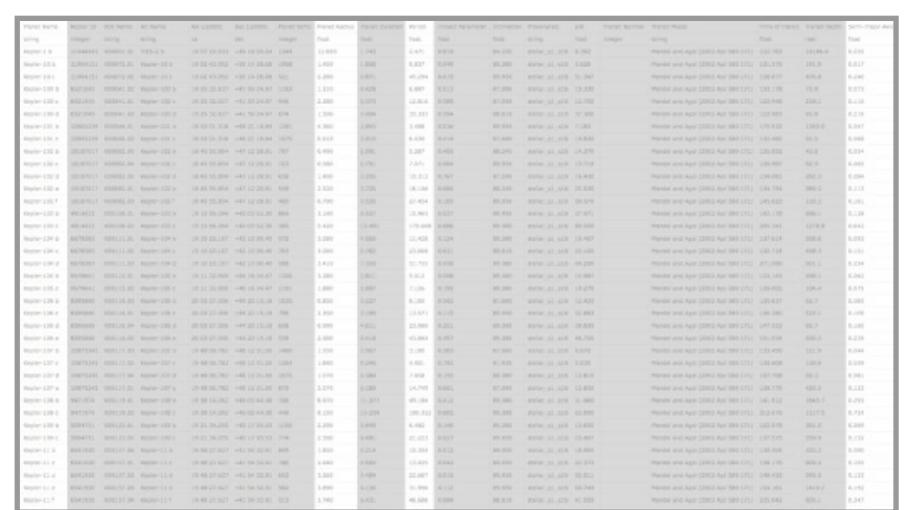
Auto fatalities per 100,000 people

# See, Think, Design, Produce

understand

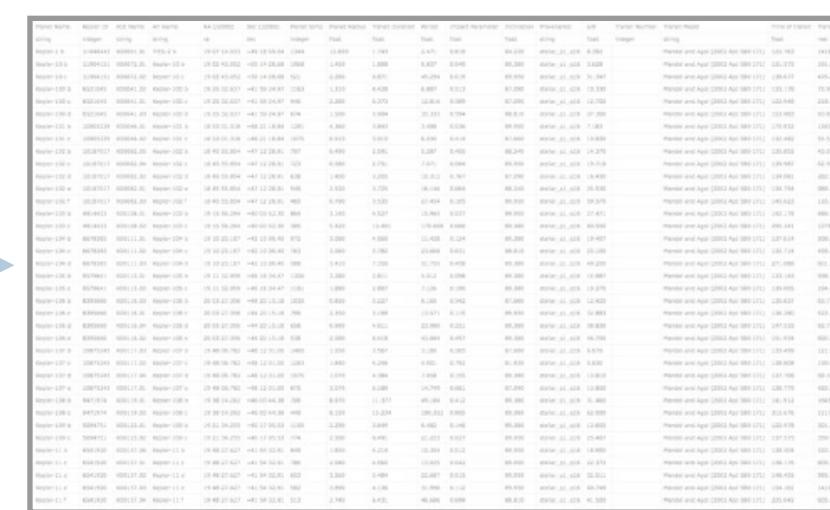
explain

*one of Corum's project process examples*



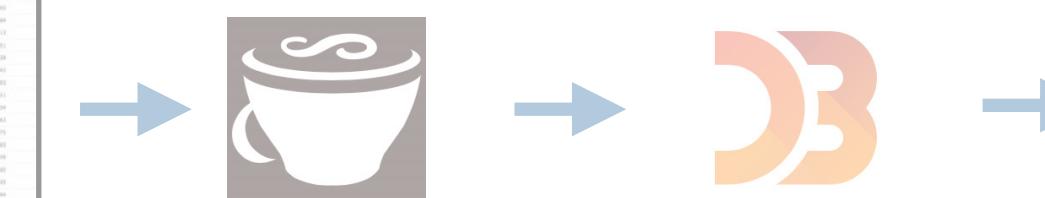
source data

→ IP[y] →



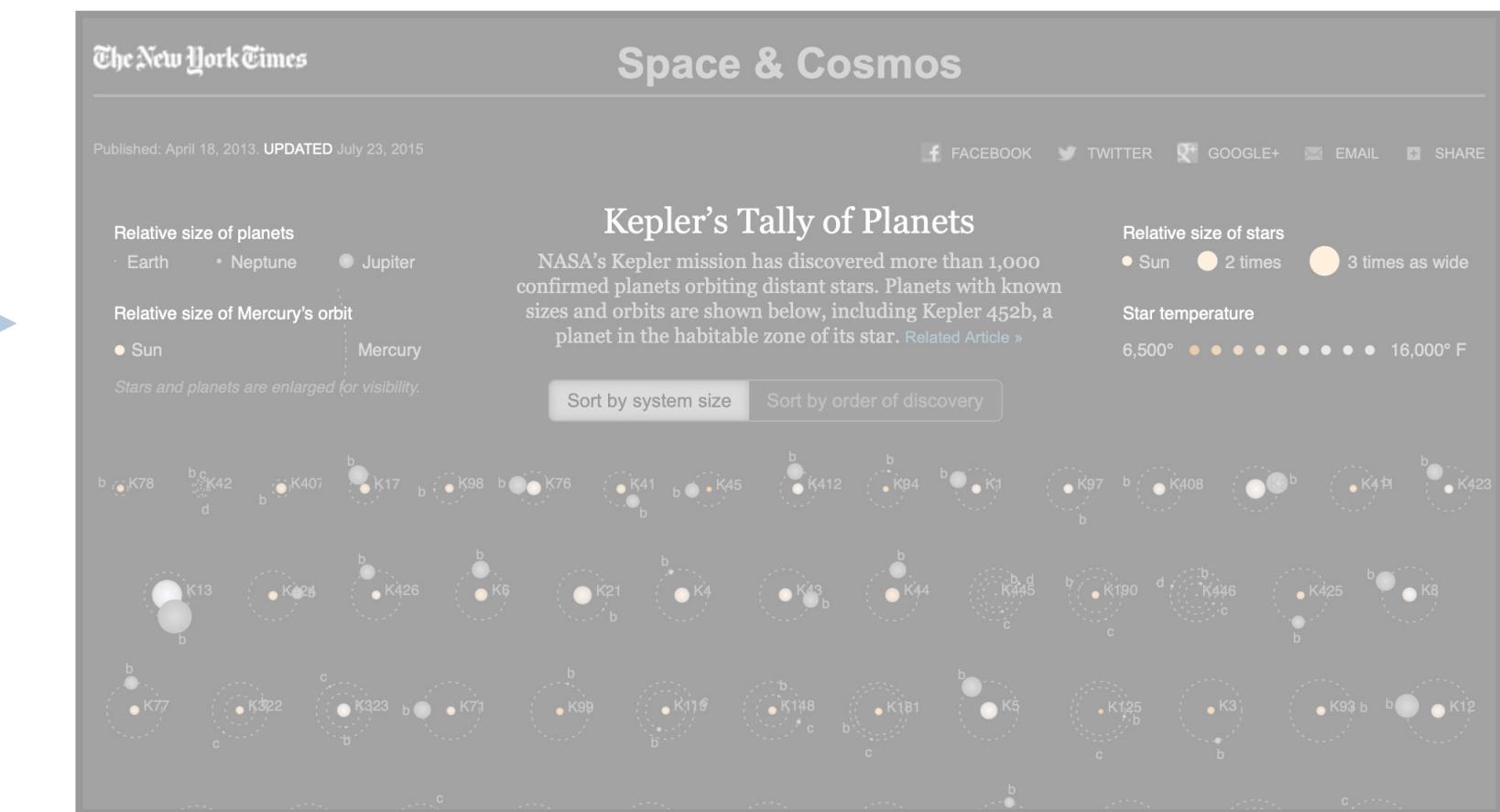
iPython  
notebook

processed  
data



coffee script

d3

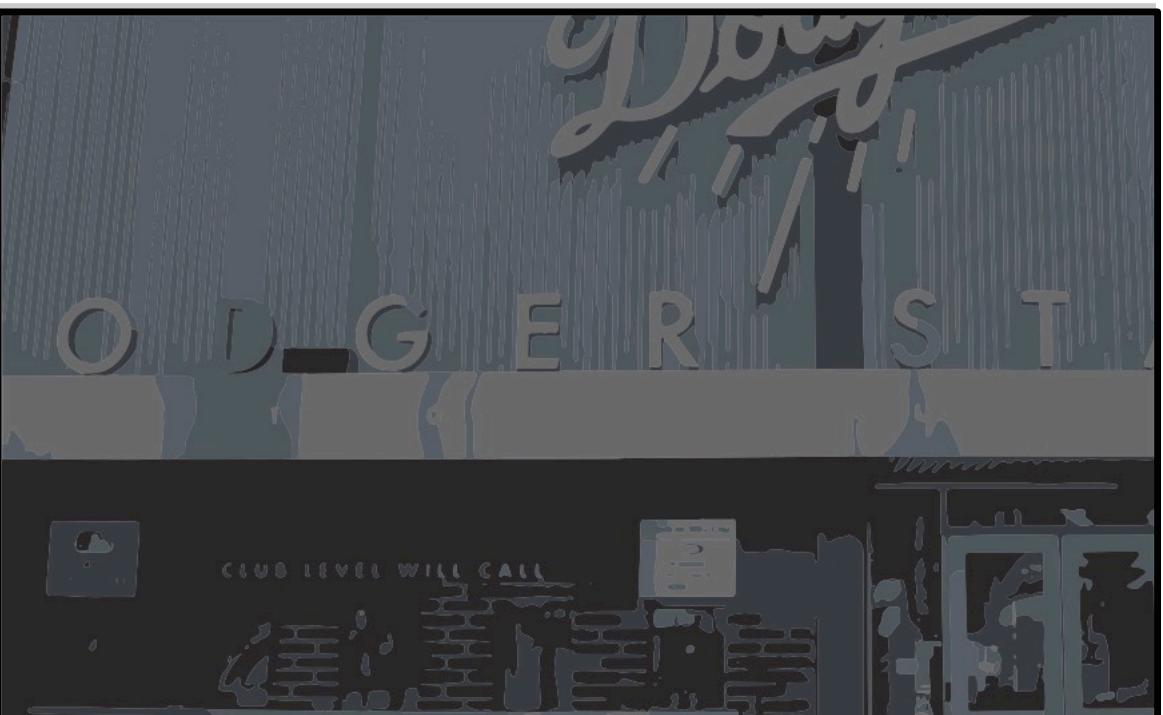


**Hone ideas  
within  
limitations**

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

## **storyboards — (another draft) example**

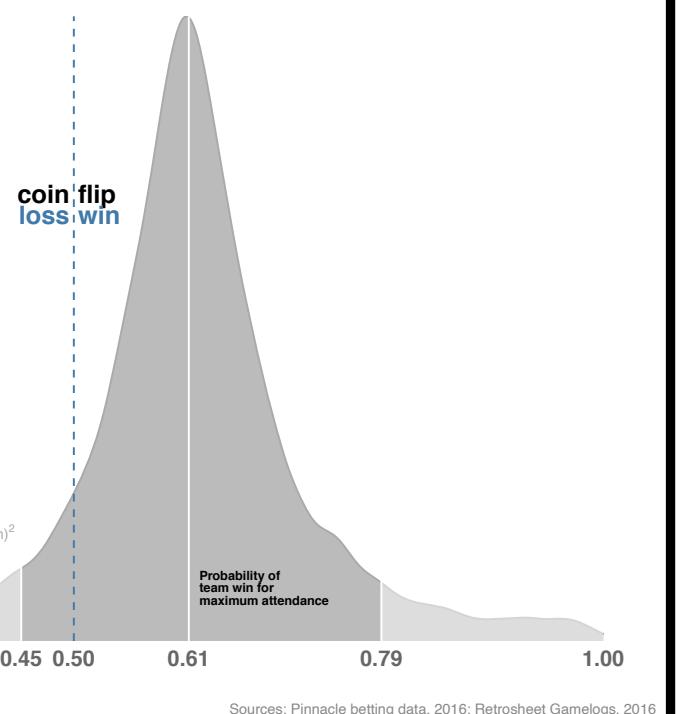
## Visual narrative



**“Uncertainty of outcome is necessary if the consumer is to be willing to pay admission to the game.”**

— Simon Rottenberg, 1956

More fans generally pay admission to our games when the chance of winning was near a median of 0.61. Fans want favorable odds without predicting the outcome.



**“What is the most exciting play in baseball?**

One in which the batter makes a perfect slide and **just beats the throw** right as a crowd is about to go absolutely nuts with joy, anger or amazement, depending on their rooting interests.

Which, now that I think about it, the play could be a putout if it unfolds exactly that way too. I don't care. I'm going for the **kinetic energy** and the execution of the whole deal, one way or another.

— Craig Calcaterra, Lead baseball writer, NBC Sports. August 2018



## 1

We dread an empty ticket booth. Our average game attendance hovers around mid-80 percent capacity — we have seats to fill, despite having winning seasons.

To draw more fans, consider the words of famed economist Simon Rotternberg, who said “uncertainty of outcome” is necessary for consumers to pay admission to ball games. While winning is important, it’s the not-knowing that creates excitement. Have we maximized winning with uncertainty of outcome?

## 2

Simon’s statement held true at our stadium, too, last year.

We modeled attendance as a function of the uncertainty of outcome, attempting to account for other factors like day of the week, time of day, and our cumulative fraction of wins.

We learned that attendance was associated with a win probability — as measured by betting markets — near 0.61.

Game uncertainty, of course, comes from play uncertainty.

## 3

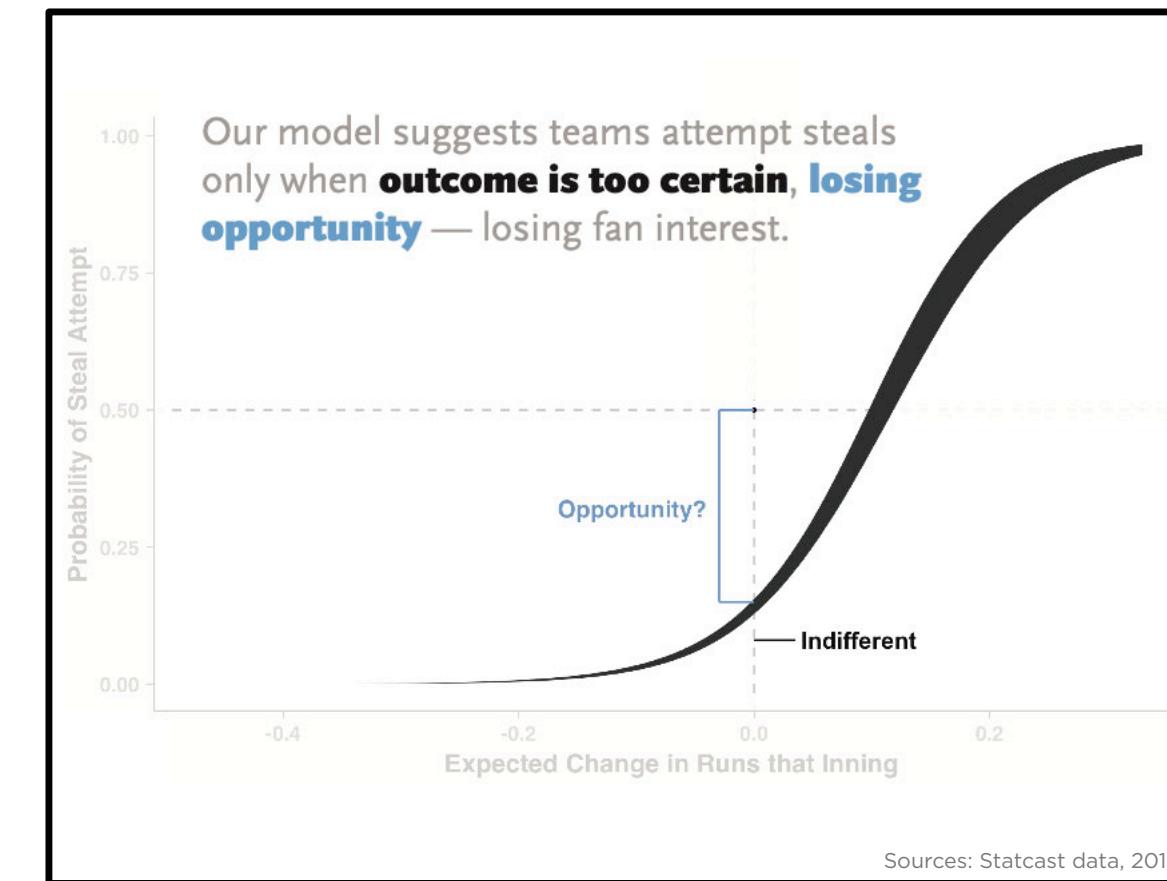
Let’s focus, then, on an event that lead baseball writer for NBC sports, Calcaterra, said is potentially the most exciting play in baseball. He says that the most exciting play may be one where the baserunner just beats the catcher’s throw when stealing a base.

Using our model, we asked whether we are maximizing base-stealing excitement. We considered running speed of the player on first base, catcher throw time to second base, the latent talents of the specific pitcher, catcher, and baserunner, and estimated the probability of successfully stealing second.

## Written narrative

## 4

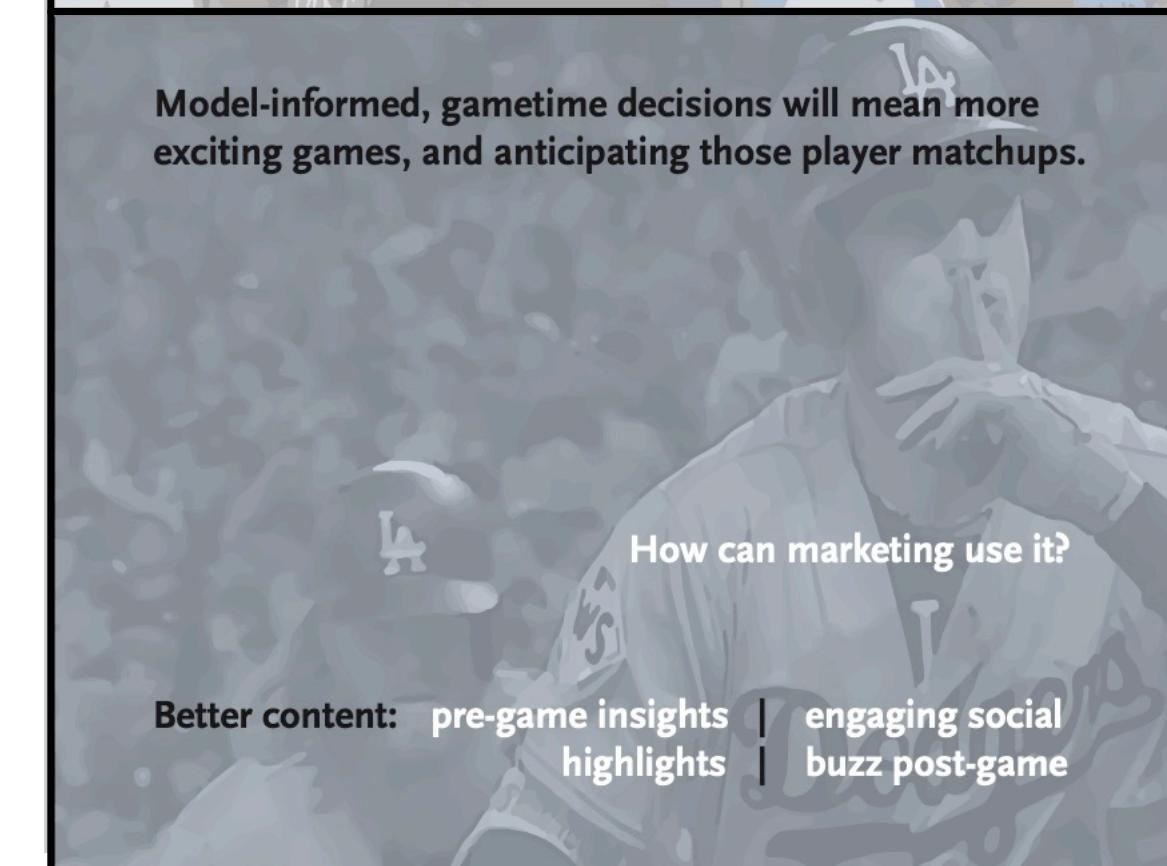
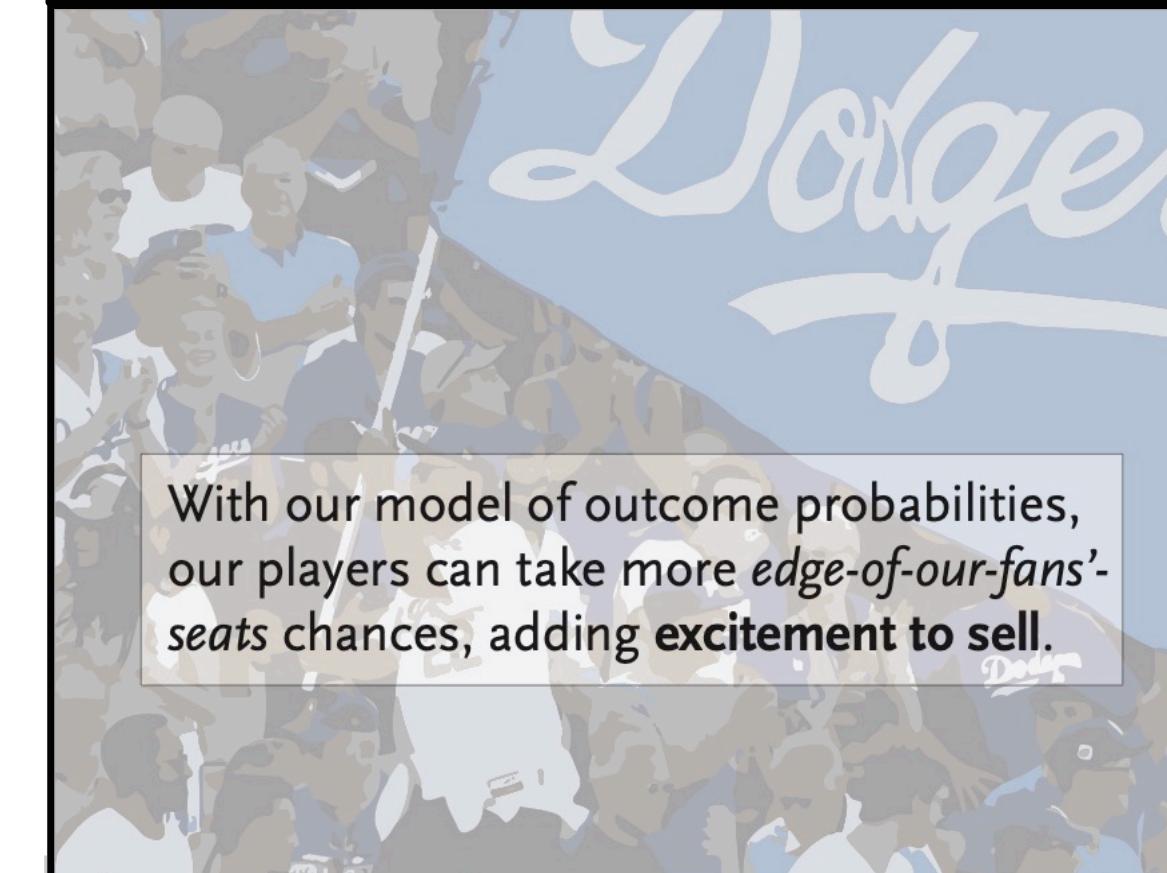
## Visual narrative



Sources: Statcast data, 2016.

## 5

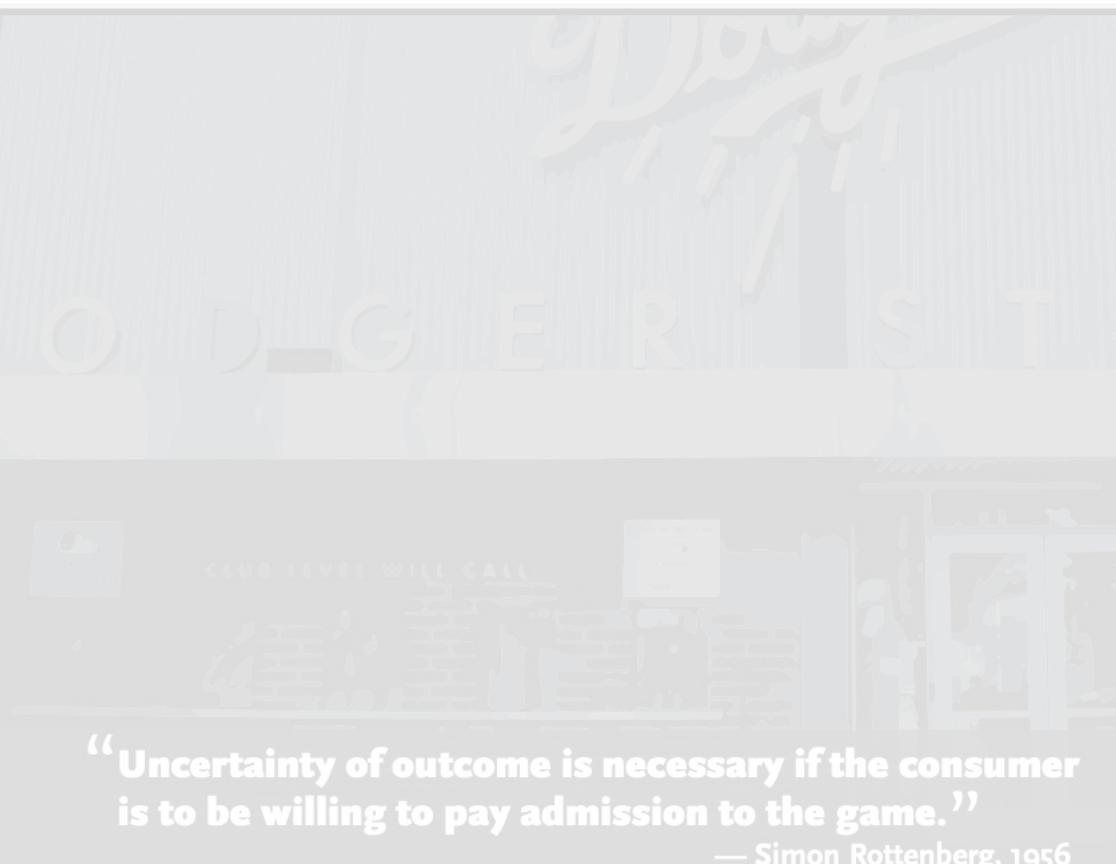
As we know, our wins come from runs. From a game outcome perspective — whether teams win — managers should be indifferent if a baserunner steals when the expected change in runs is zero. Our model suggests baserunners on all teams are too conservative. And us? We ranked 13th in uncertainty of outcome on base stealing.



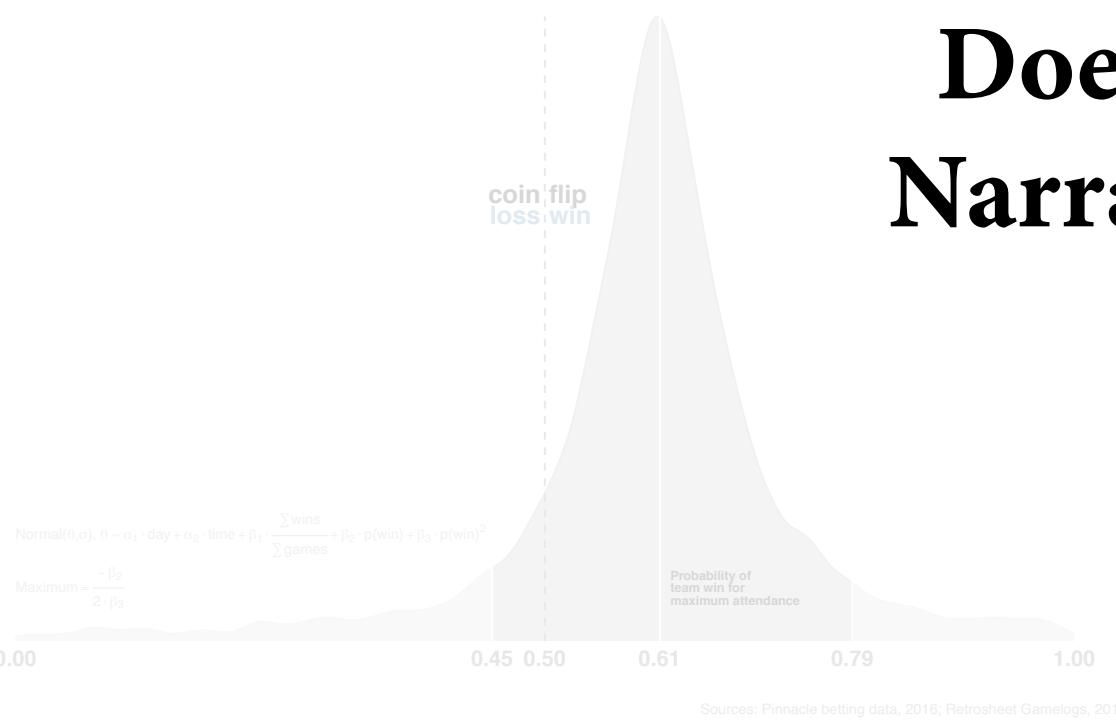
## 6

By accessing our model, marketing can make better content even before games start. While the model is proprietary, marketing will know ahead of time what matchups may be more interesting and can use our estimates to build anticipation. Along with pre-game insights, marketing can better engage with fans on social media, and stream better highlights. Marketing better content will add post-game buzz.

## Visual narrative



More fans generally pay admission to our games when the chance of winning was near a median of 0.61. Fans want favorable odds without predicting the outcome.

**“What is the most exciting play in baseball?**

One in which the batter makes a perfect slide and **just beats the throw** right as a crowd is about to go absolutely nuts with joy, anger or amazement, depending on their rooting interests.

Which, now that I think about it, the play could be a putout if it unfolds exactly that way too. I don't care. I'm going for the **kinetic energy** and the execution of the whole deal, one way or another.

— Craig Calcaterra, Lead baseball writer, NBC Sports. August 2018



## Written narrative

1

We dread an empty ticket booth. Our average game attendance hovers around mid-80 percent capacity — we have seats to fill, despite having winning seasons.

To draw more fans, consider the words of famed economist Simon Rottenberg, who said “uncertainty of outcome” is necessary for consumers to pay admission to ball games. While winning is important, it’s the not-knowing that creates excitement. Have we maximized winning with uncertainty of outcome?

2

## Does the storyboard show change or just trivia? Interesting patterns? Narrative and annotations use the language of our audience, the CMO? Storyboard help evaluate the combined narrative? Explain.

Simon’s statement held true at our stadium, too last year. We modeled attendance as a function of the uncertainty of

the week, time of day and our cumulative fraction of wins.

We learned that attendance was associated with a win probability — as measured by betting markets — near 0.62.

Game uncertainty, of course, comes from play uncertainty.

3

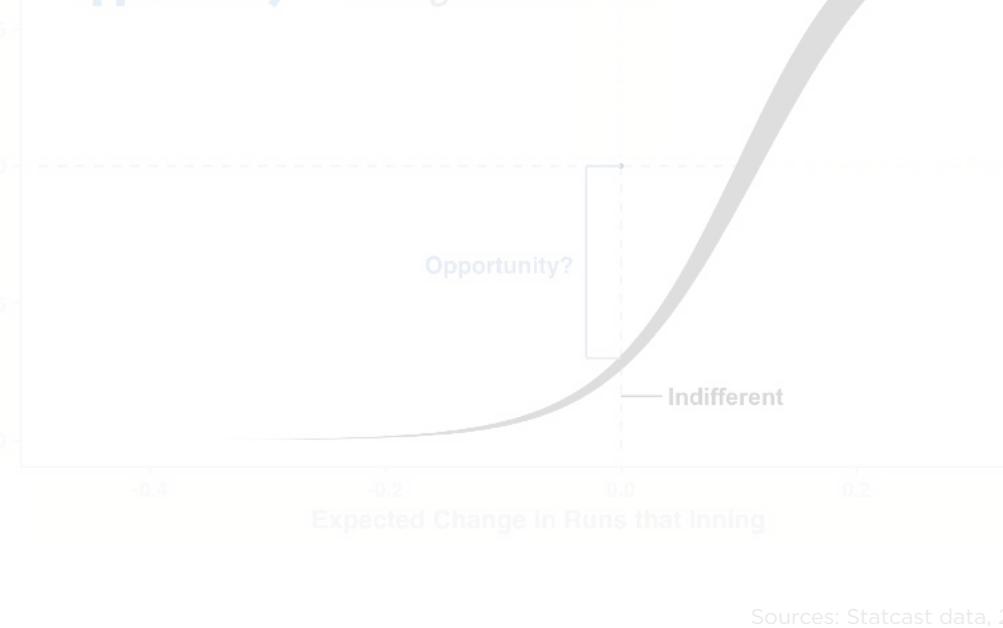
Let’s focus, then, on an event that lead baseball writer for NBC sports, Calcaterra, said is potentially the most exciting play in baseball. He says that the most exciting play may be one where the baserunner just beats the catcher’s throw when stealing a base.

Using our model, we asked whether we are maximizing base-stealing excitement. We considered running speed of the player on first base, catcher throw time to second base, the latent talents of the specific pitcher, catcher, and baserunner, and estimated the probability of successfully stealing second.

## Visual narrative

4

Our model suggests teams attempt steals only when **outcome is too certain, losing opportunity** — losing fan interest.



As we know, our wins come from runs. From a game outcome perspective — whether teams win — managers should be indifferent if a baserunner steals when the expected change in runs is zero. Our model suggests baserunners on all teams are too conservative. And us? We ranked 13th in uncertainty of outcome on base stealing.

5

With our model, our baserunners will know to steal more often, when the outcome is less certain, but the process still favors our team.

We still have room to take more risks — with *all plays*, not just base stealing. With this knowledge and marketing, we can help consumers experience more uncertainty in what happens, to enticing more fans into seats.

6

Model-informed, gametime decisions will mean more exciting games, and anticipating those player matchups.

How can marketing use it?  
Better content: pre-game insights highlights | engaging social buzz post-game

By accessing our model, marketing can make better content even before games start. While the model is proprietary, marketing will know ahead of time what matchups may be more interesting and can use our estimates to build anticipation. Along with pre-game insights, marketing can better engage with fans on social media, and stream better highlights. Marketing better content will add post-game buzz.

## Establishing *context* for your data

**Who What When Where Compare**

**Poor:** “There were 25 million deaths.”

**Better:** “During the fourteenth century, 25 million people died in Europe.”

**Best:** “When the Black Plague hit Europe in the latter half of the fourteenth century, it took the lives of 25 million people, young and old, city dwellers and those living in the countryside. The disease killed about one-quarter of Europe’s total population at the time (Mack, n.d.).”

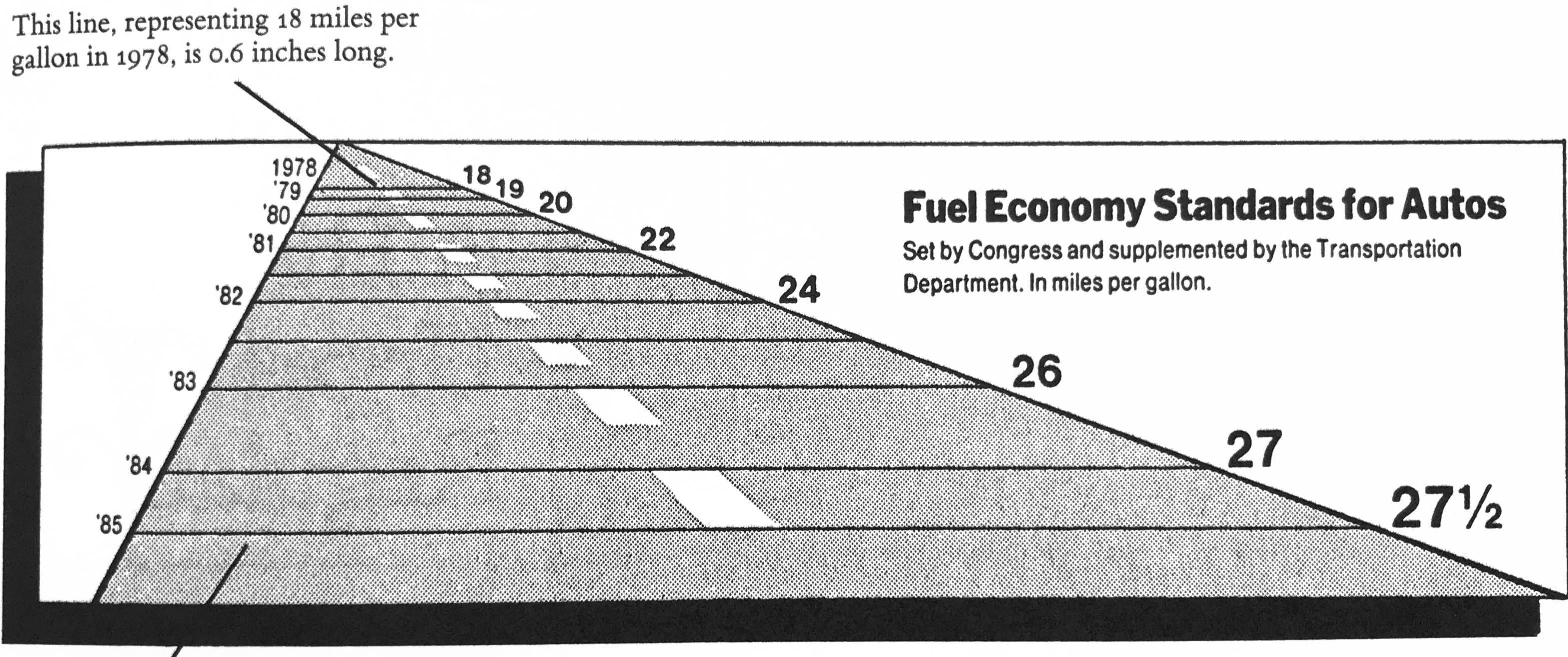
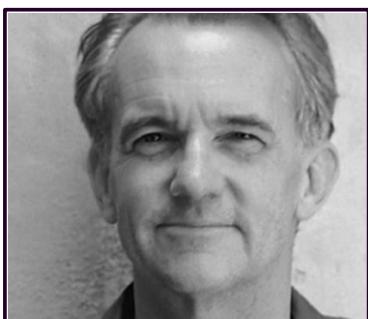
**(more on) data graphics**

## (un)helpful encodings

Fuel economy difference  
between 1985 and 1978

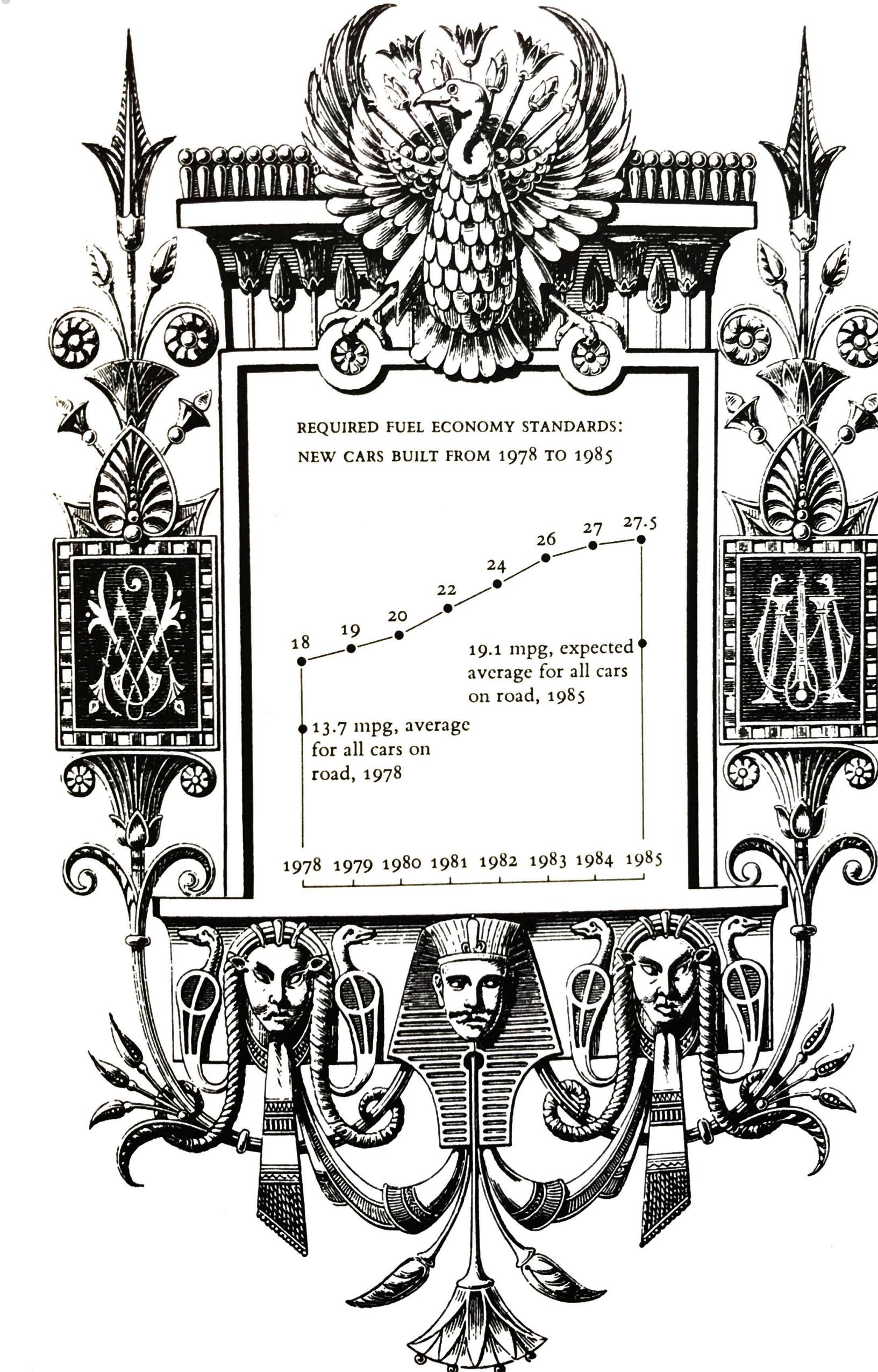
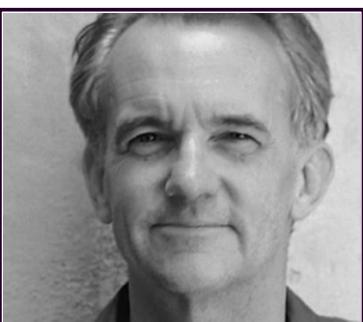
Numerical  
14.8 percent

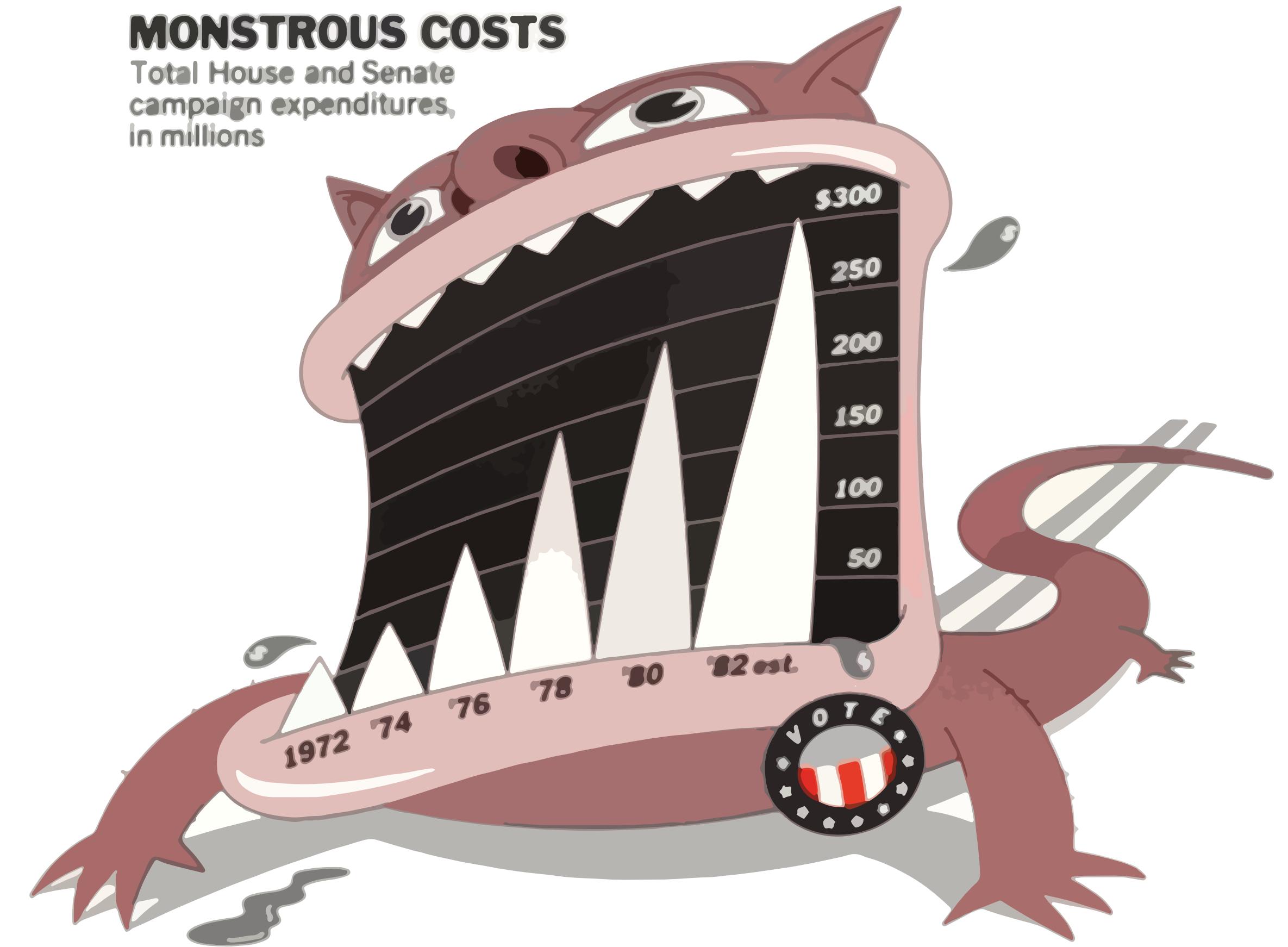
Graphical  
783 percent



New York Times, August 9, 1978, D-2.

(more on) data graphics | *chart junk and unhelpful encodings?*

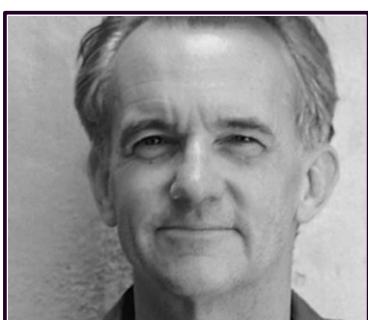




$$\text{data-ink ratio} = \frac{\text{data-ink}}{\text{total ink used to print the graphic}}$$

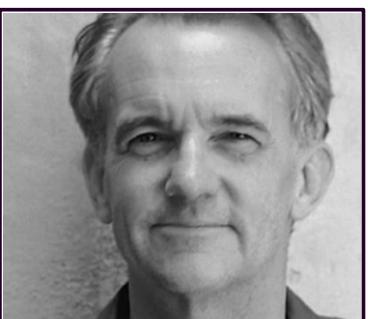
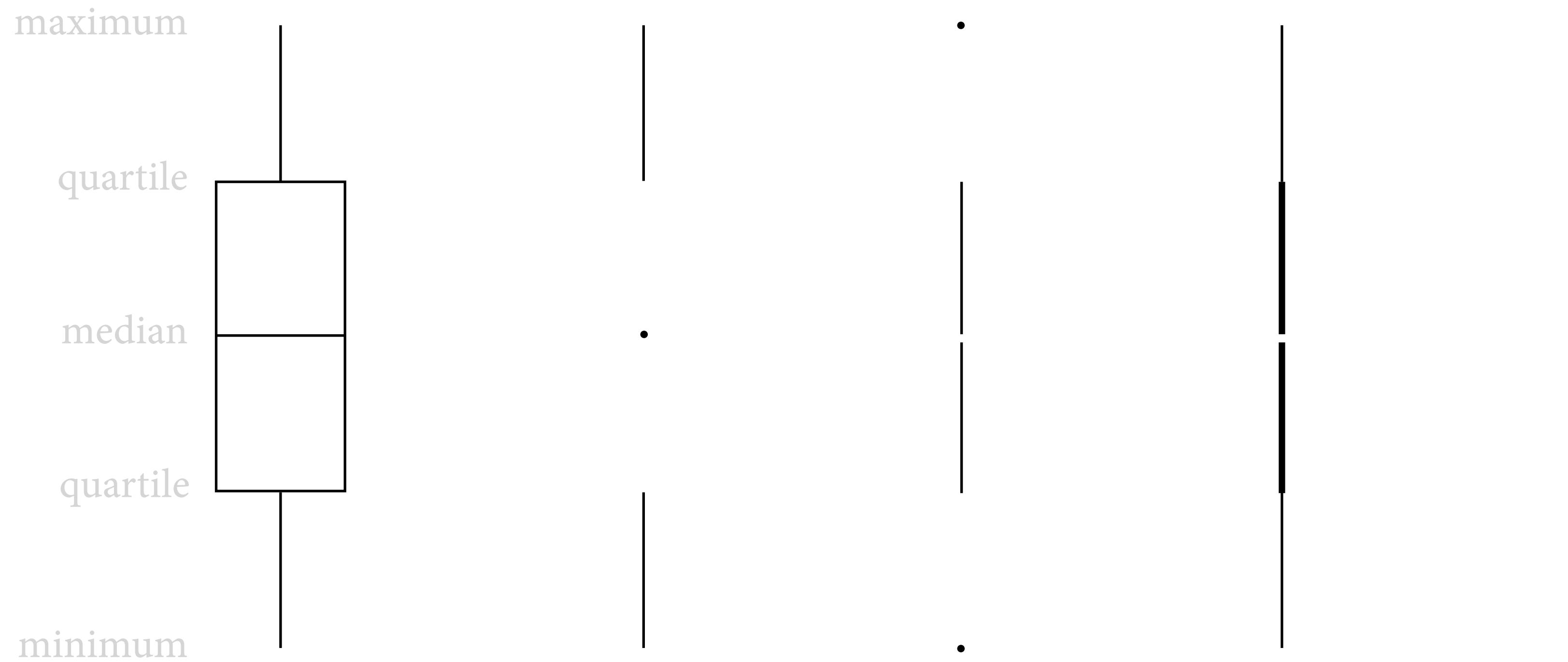
= proportion of a graphic's ink devoted to the non-redundant display of data-information

=  $1.0 - \text{proportion of a graphic that can be erased without loss of data-information}$



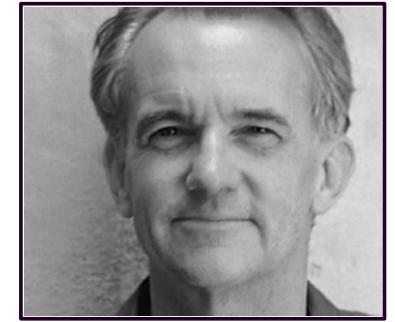
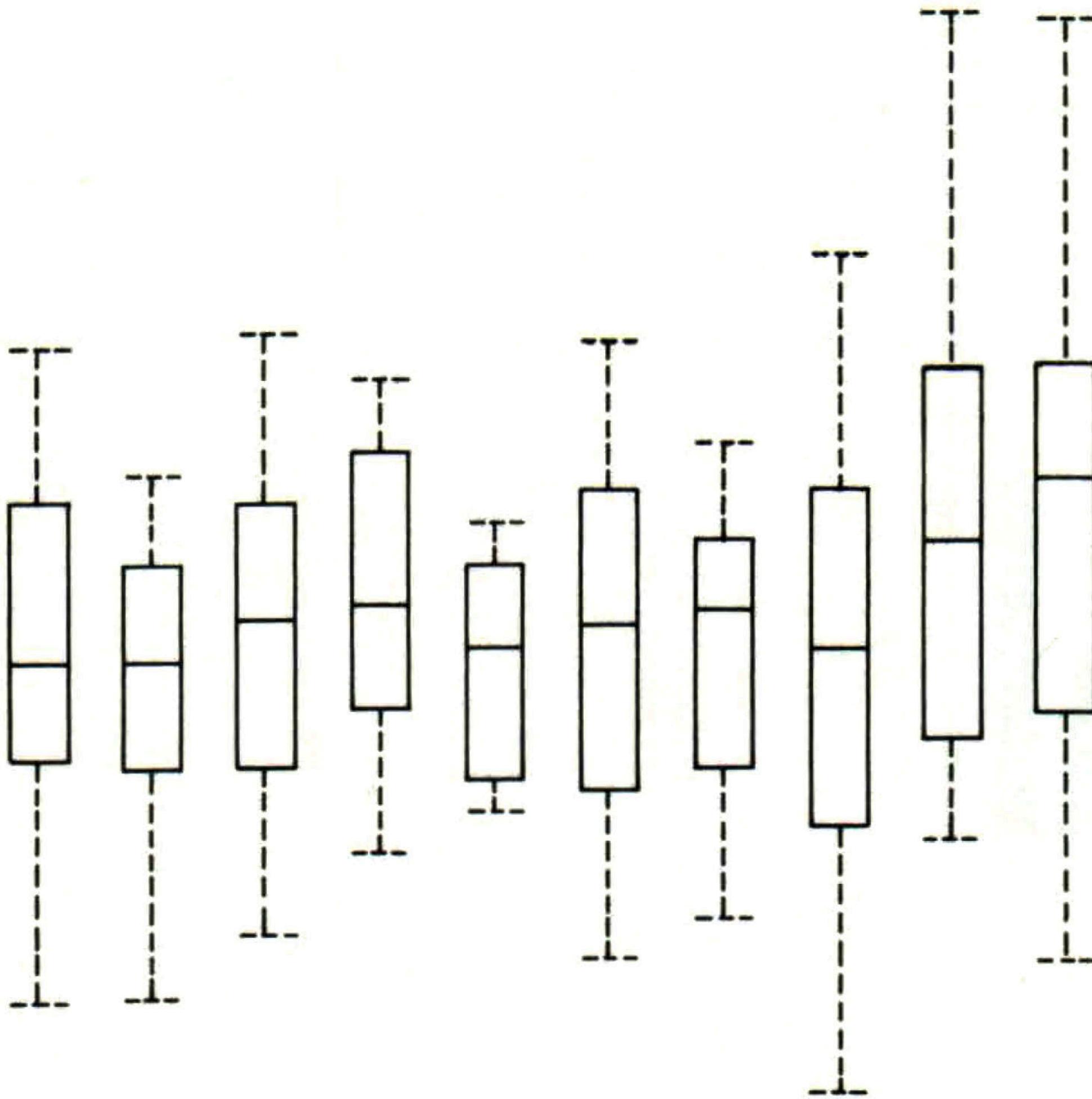
# Experimenting with data graphics

# Experimenting with data graphics | *redesigning the box plot*

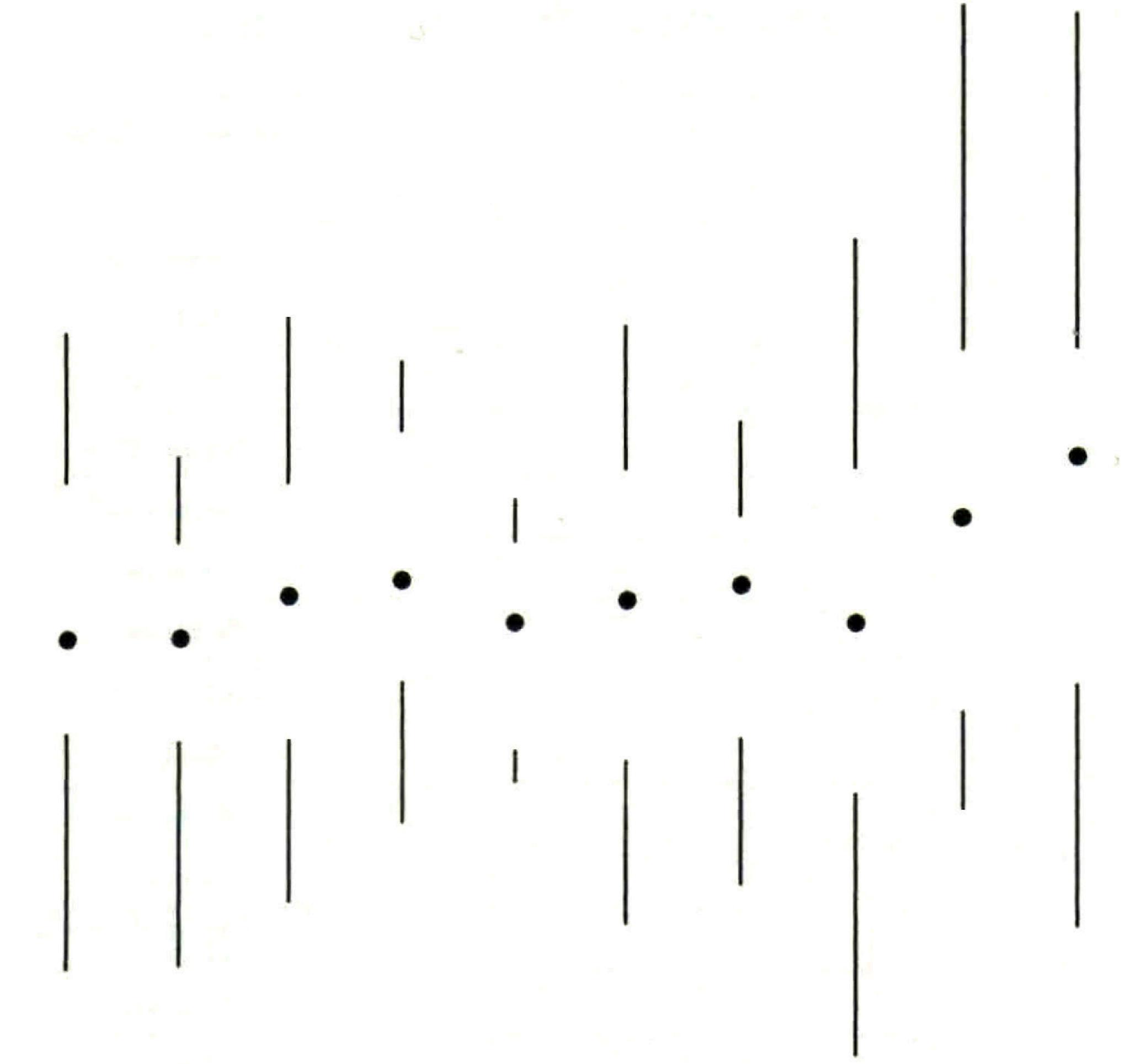
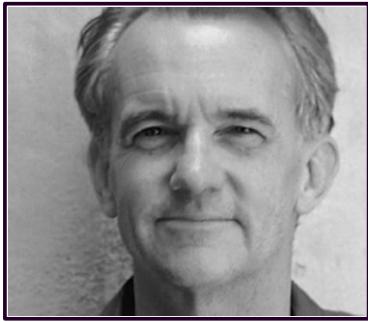


In these revisions of the box plot, . . . the best overall arrangement naturally also rests on statistical and aesthetic criteria — in other words, the procedure is one of *reasonable* data-ink maximizing.

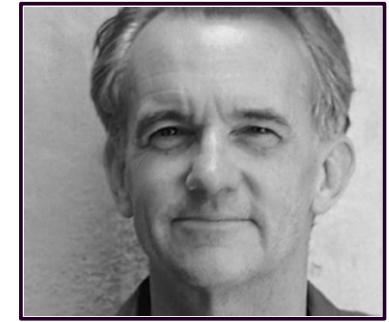
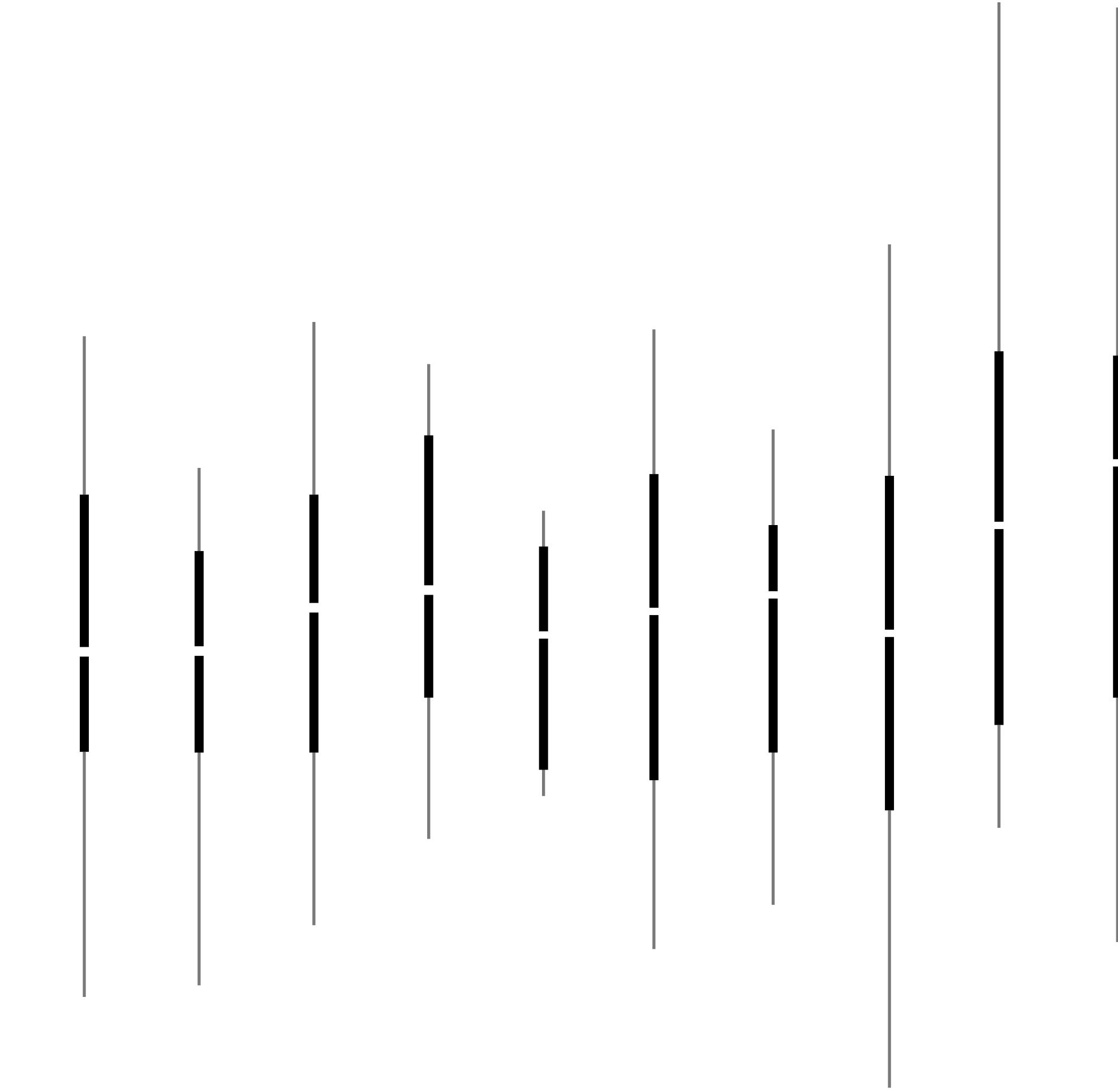
# Experimenting with data graphics | *redesigning the box plot*



# Experimenting with data graphics | *redesigning the box plot*



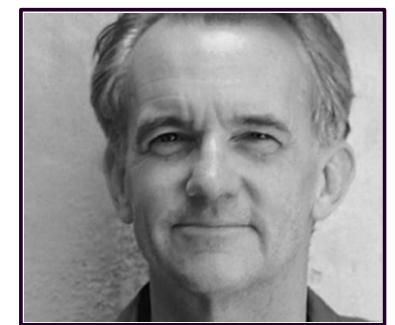
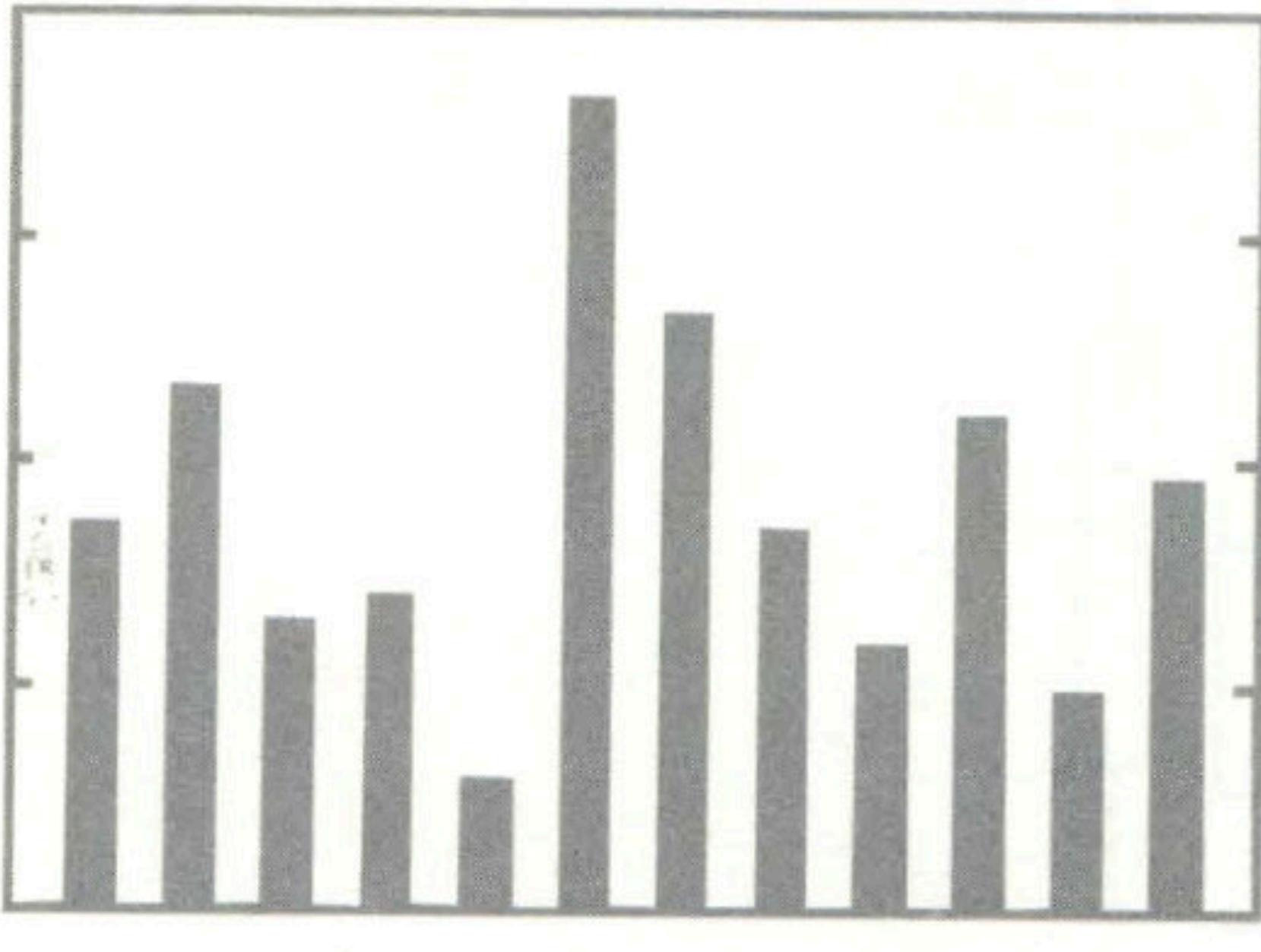
# Experimenting with data graphics | *redesigning the box plot*



---

Tufte, Edward

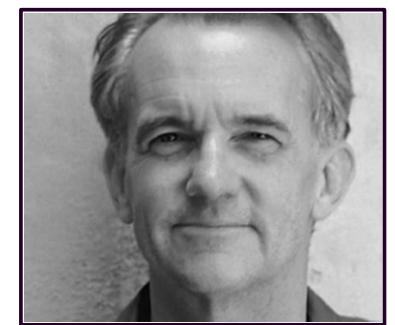
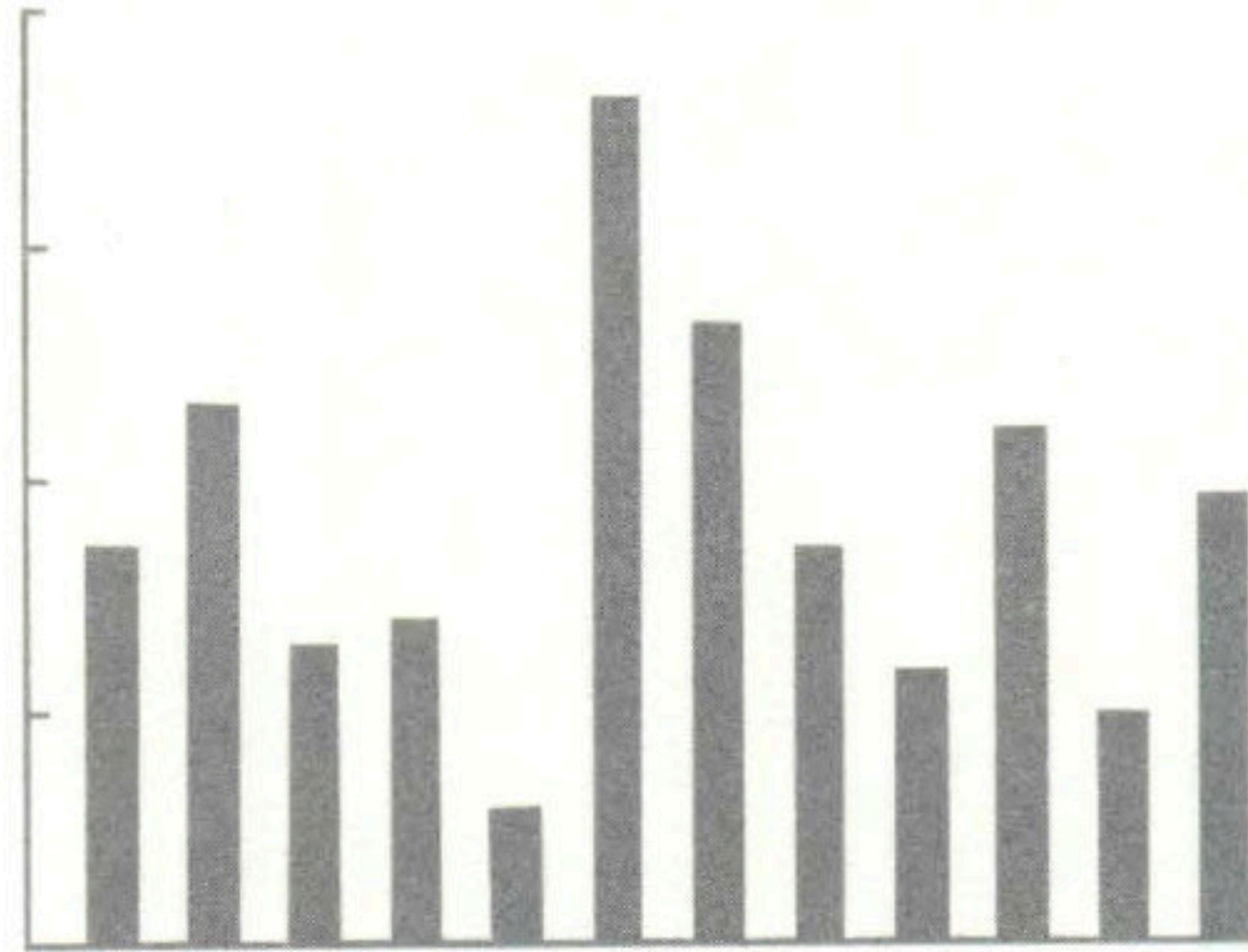
# Experimenting with data graphics | *redesigning the bar chart / histogram*



---

Tufte, Edward

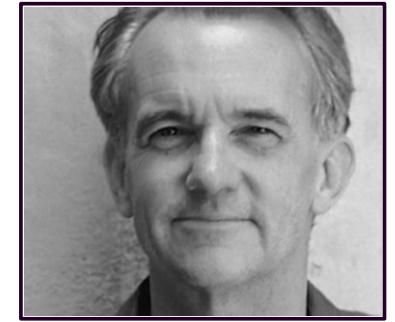
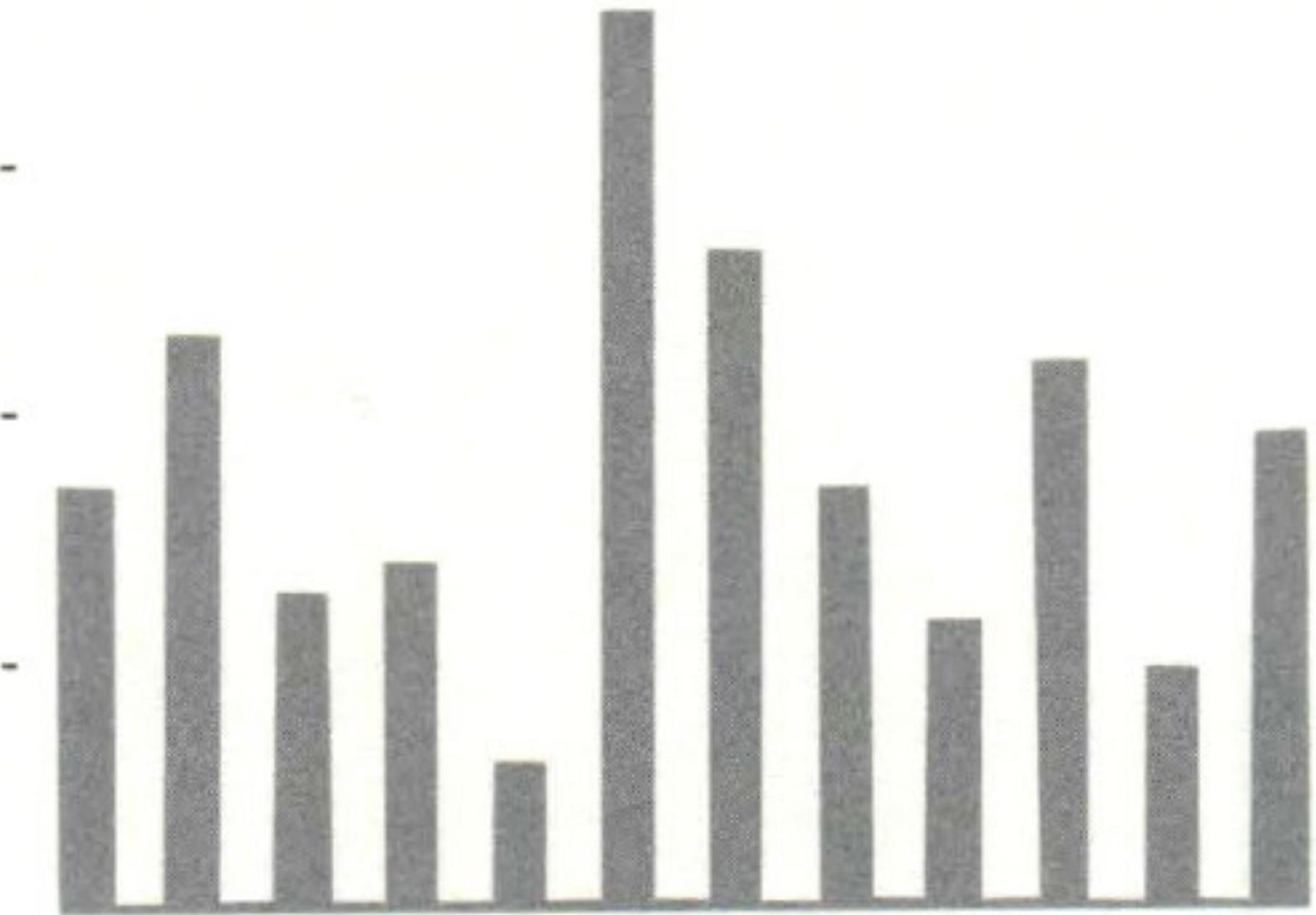
# Experimenting with data graphics | *redesigning the bar chart / histogram*



---

Tufte, Edward

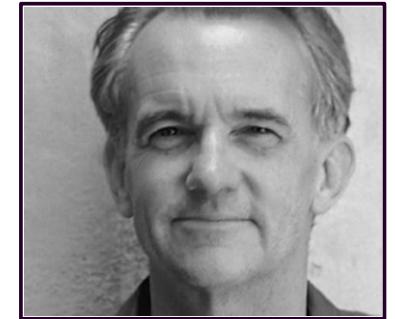
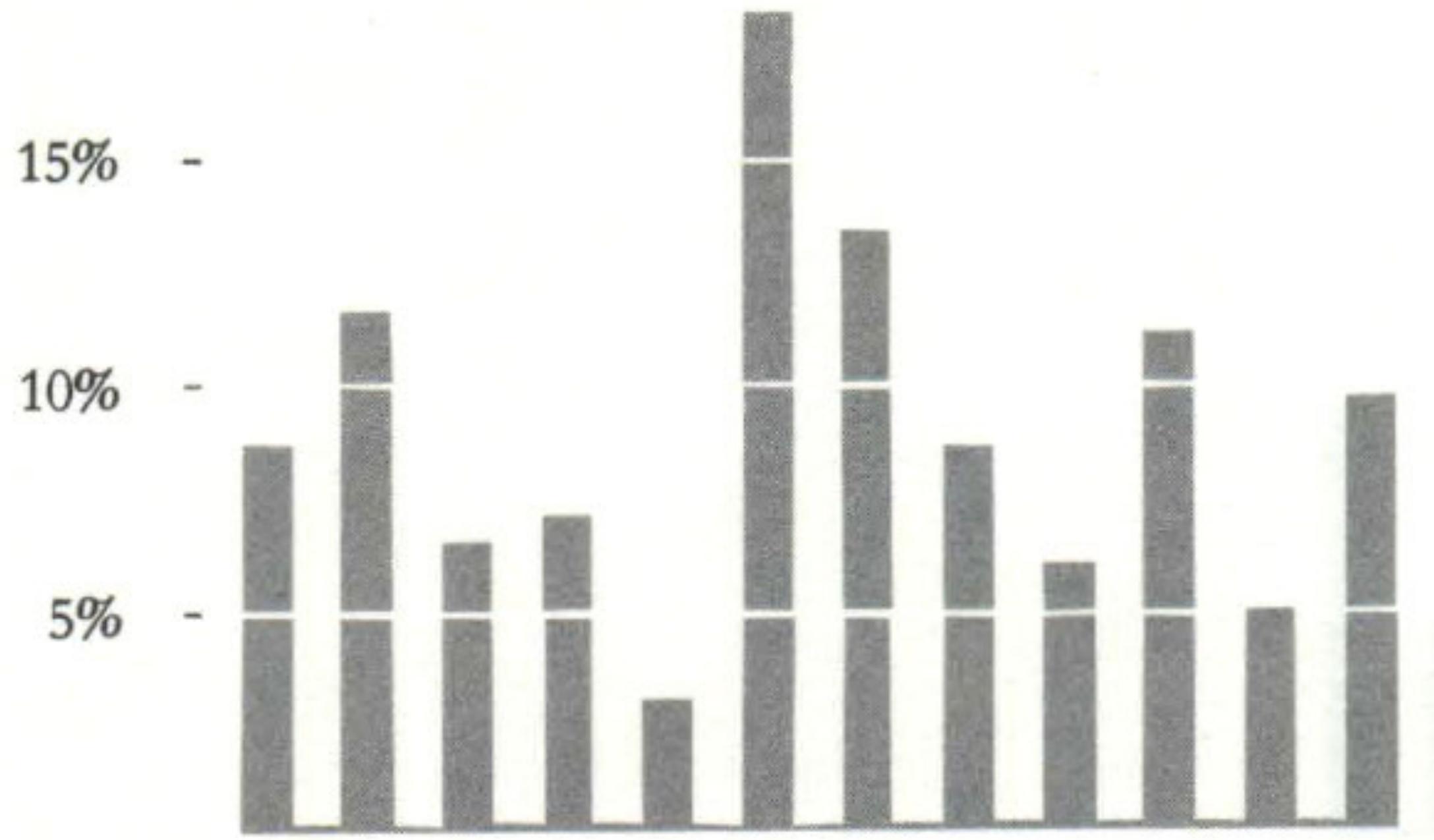
# Experimenting with data graphics | *redesigning the bar chart / histogram*



---

Tufte, Edward

# Experimenting with data graphics | *redesigning the bar chart / histogram*



---

Tufte, Edward

“

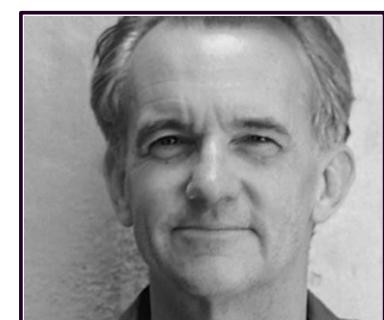
Maximizing data ink (within reason) is but a single dimension of a complex and multivariate design task.

The principle helps conduct  
**experiments in graphical design.**

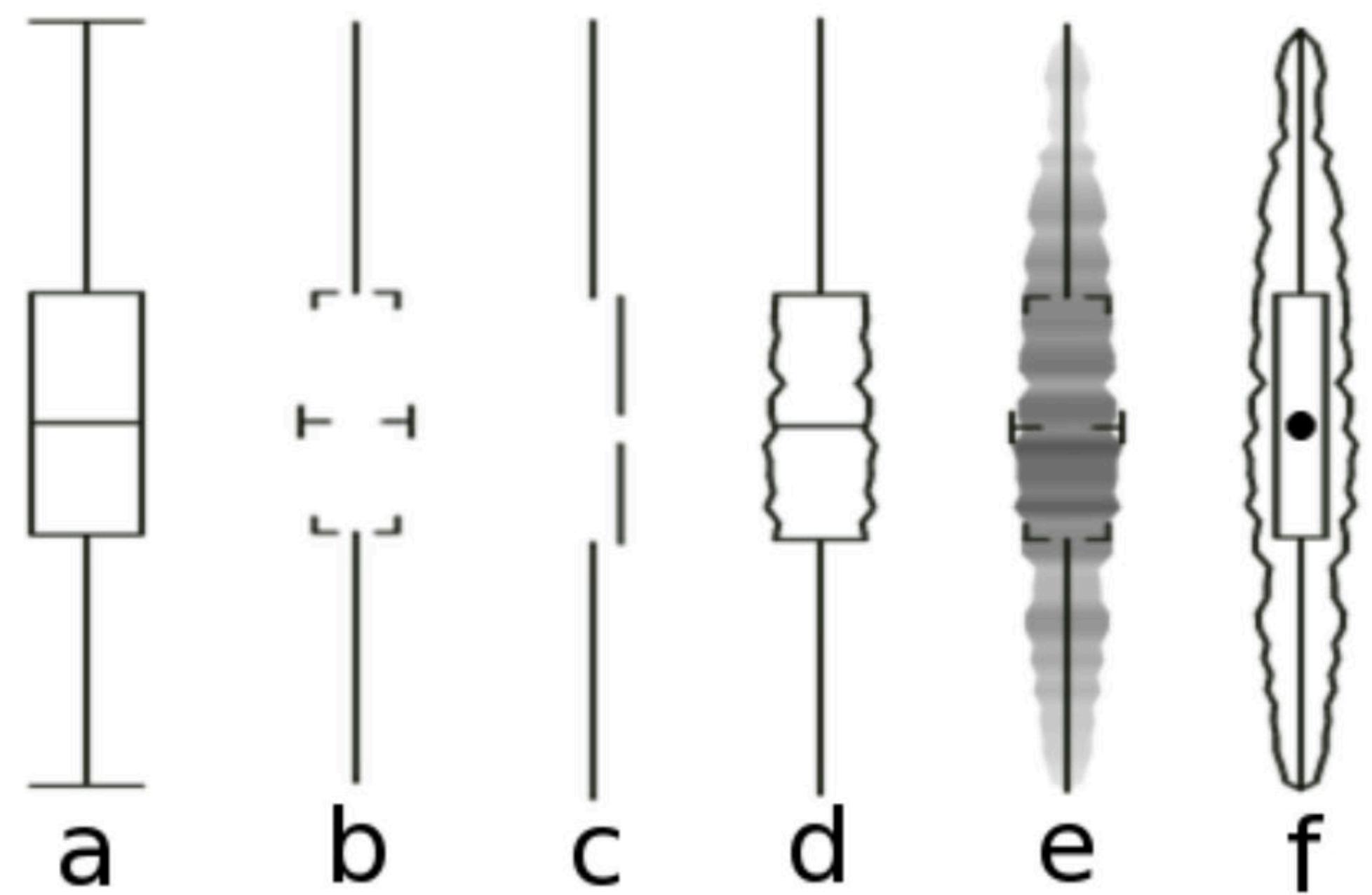
Some of those  
experiments will succeed.

There remain, however, many **other considerations** in the design of statistical graphics — not only of efficiency, but also of **complexity, structure, density, and even beauty.**

”



**How might we decide which  
data graphics from experiments to use?**



Anderson et al. (2011) found that, of the four kinds of boxplot shown in Figure 1.7, the minimalist version from Tufte's own work (option C) proved to be the most cognitively difficult for viewers to interpret.

Cues like labels and gridlines, together with some strictly superfluous embellishment of data points or other design elements, may often be an aid rather than an impediment to interpretation.



**Design is a search problem**



**Prototypes should  
emphasize speed over polish.**

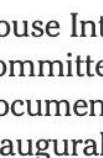
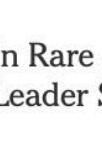
**Get fresh eyes frequently.  
Invite criticism.**

**Move from  
exploring to refining.**

# Experimenting with data graphics | *create, question, test*

≡ SECTIONS  HOME 🔎 SEARCH

The New York Times  

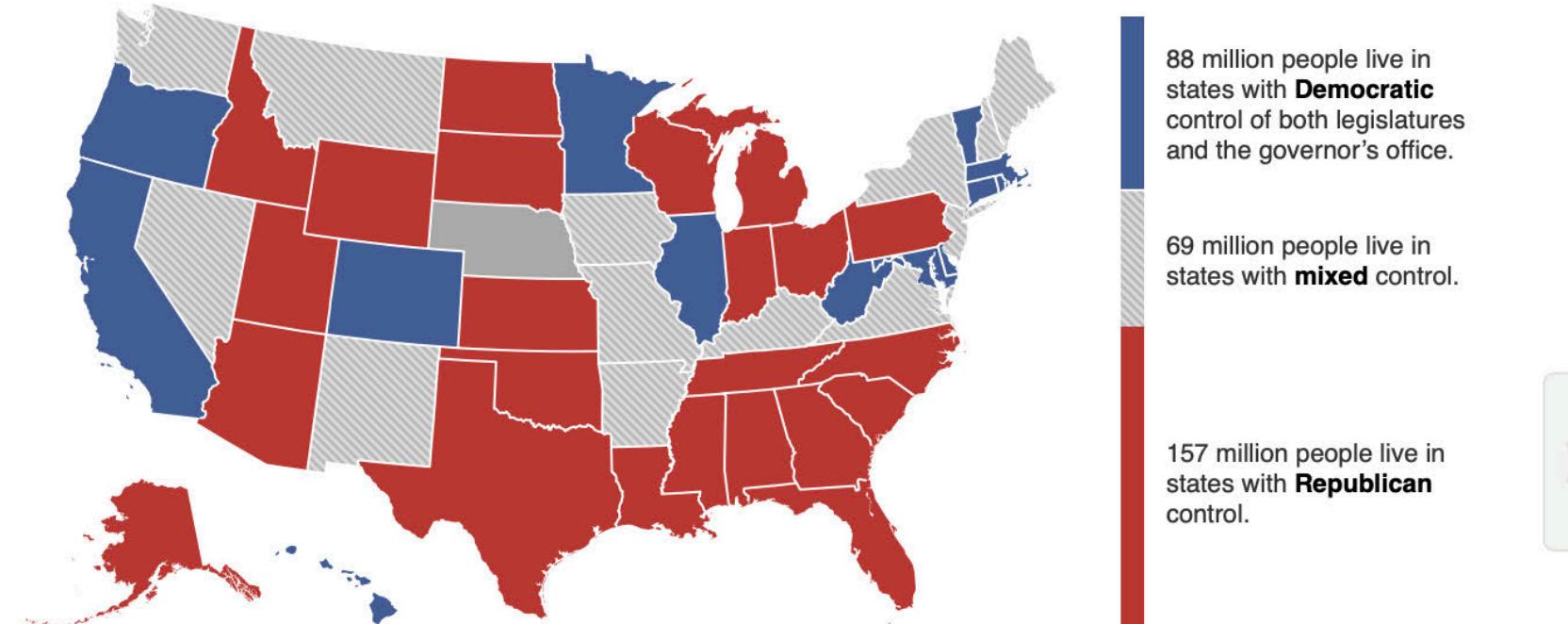
 Joe Biden, in Video, Says He Will Be 'More Mindful' of Personal Space  Senate Republicans Go 'Nuclear' to Speed Trump Confirmations  Subpoena for Mueller Report and Documents Approved by House Judiciary Committee  House Intelligence Committee Seeks Documents From Trump's Inaugural  In Rare S...

POLITICS | ONE-PARTY RULE  SHARE

## Taking the Battle to the States

By HAEYOUN PARK, JEREMY ASHKENAS and MIKE BOSTOCK JAN. 11, 2014

Republicans or Democrats have single-party control of both the legislature and the governor's office in 36 states, the most in six decades. Lawmakers in these states have been seeking to reshape government policy in recent years, from legalizing same-sex marriage to restricting labor unions. Some of these laws were passed after the rapid rise of single-party control in 2010; others have been in place for years. Below is a look at where states stand on some key issues.



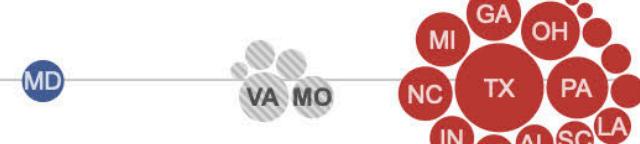
**Did not pass new restrictions in 2013**



**Abortion**

Twenty-two states — all but six controlled by Republicans — enacted 70 abortion restrictions in 2013, according to the Guttmacher Institute, a research group. Laws ranged from bans on abortions 20 weeks after fertilization to limitations on insurance coverage of abortions.

**Passed new restrictions in 2013**



**Expanded Medicaid**



**Medicaid**

A 2012 Supreme Court decision allowed states to decide whether to expand Medicaid to more low-income adults under the Affordable Care Act. All 13 Democratic states have expanded the program. Most Republican states did not. Pennsylvania, Indiana and Tennessee have not expanded it but may do so.

**Have not expanded at this time**



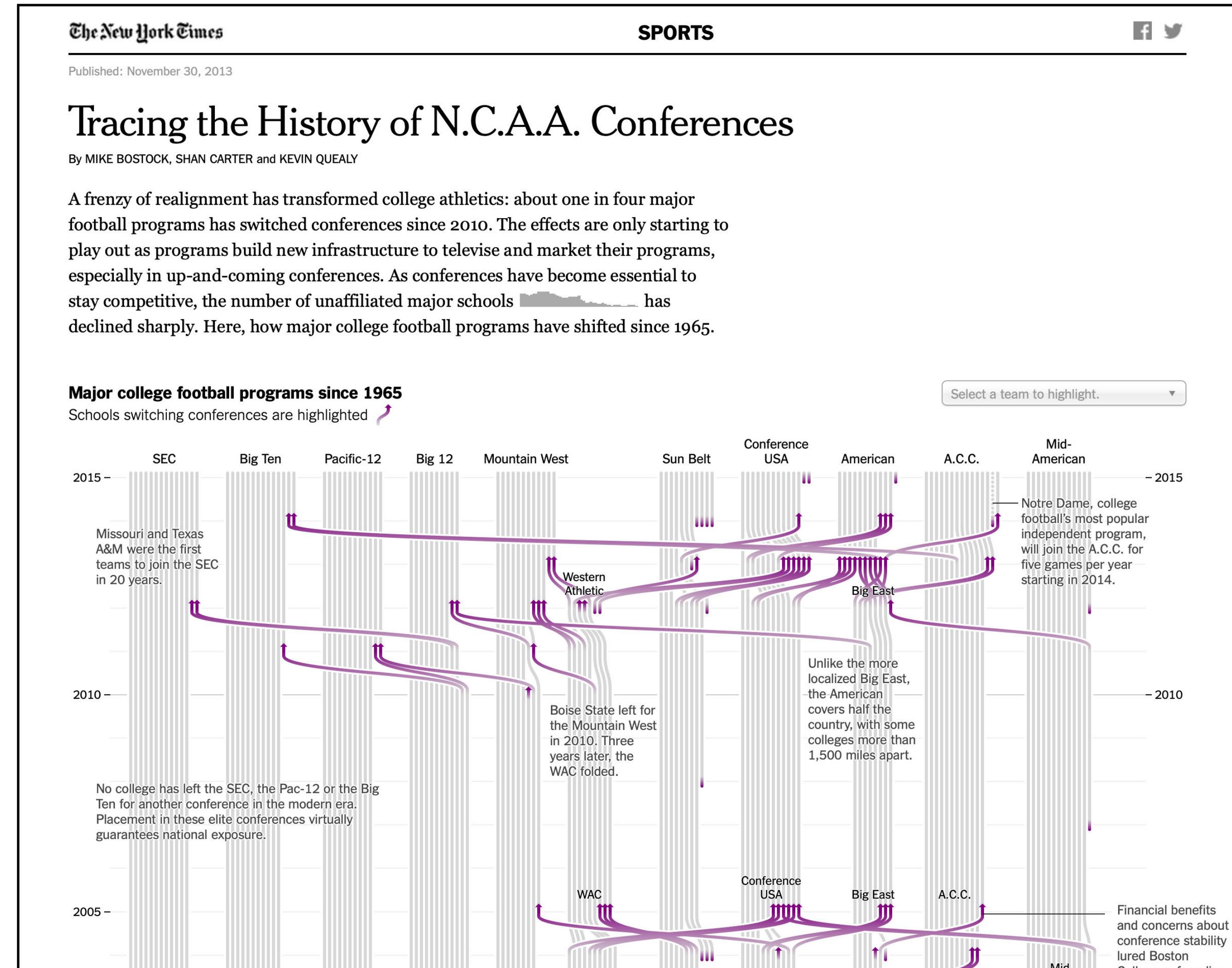
Bostock, Mike

Scott Spencer /  <https://github.com/ssp3nc3r>  [scott.spencer@columbia.edu](mailto:scott.spencer@columbia.edu)

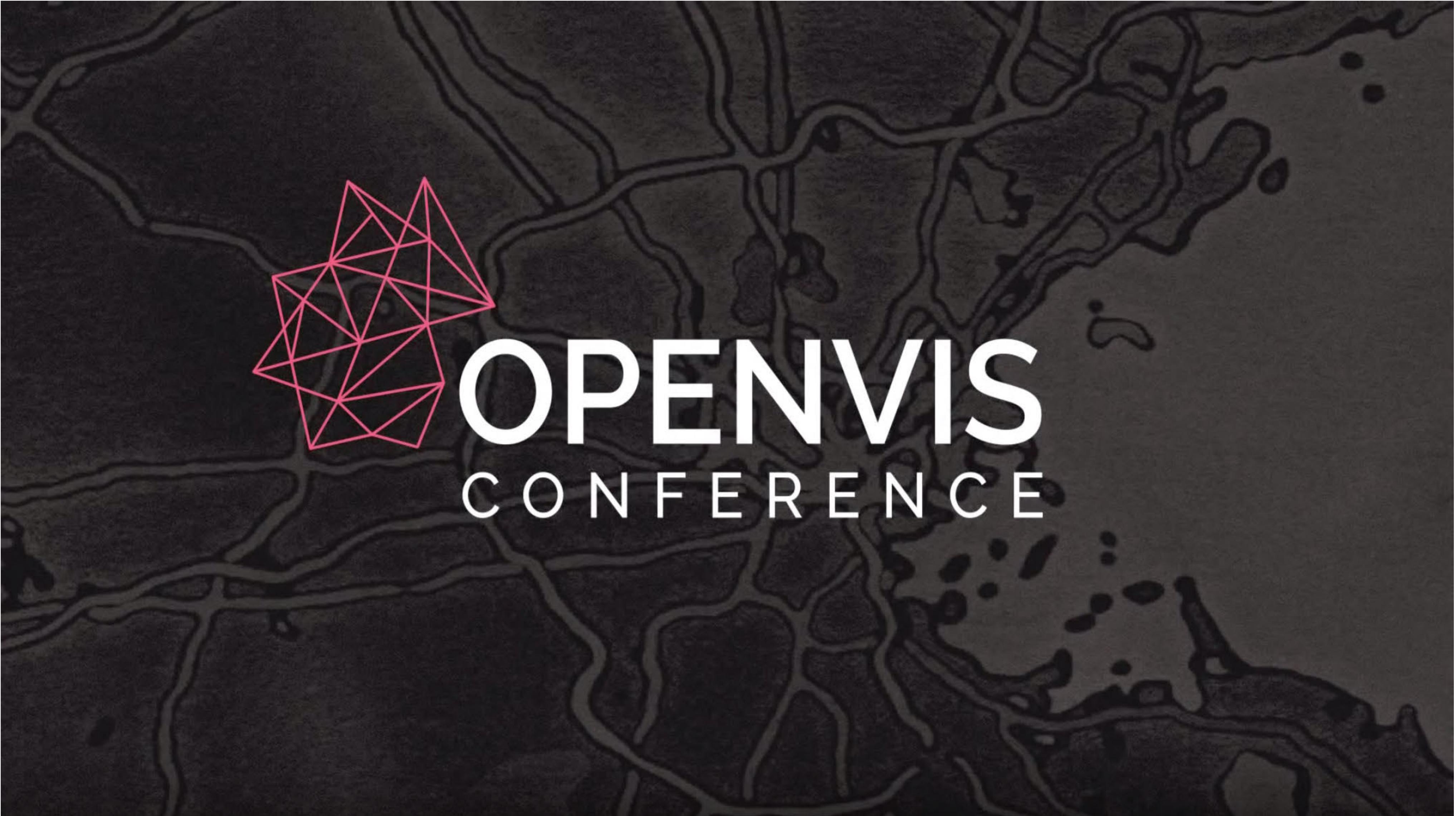
36



# Experimenting with data graphics | *create, question, test*



Bostock, Mike



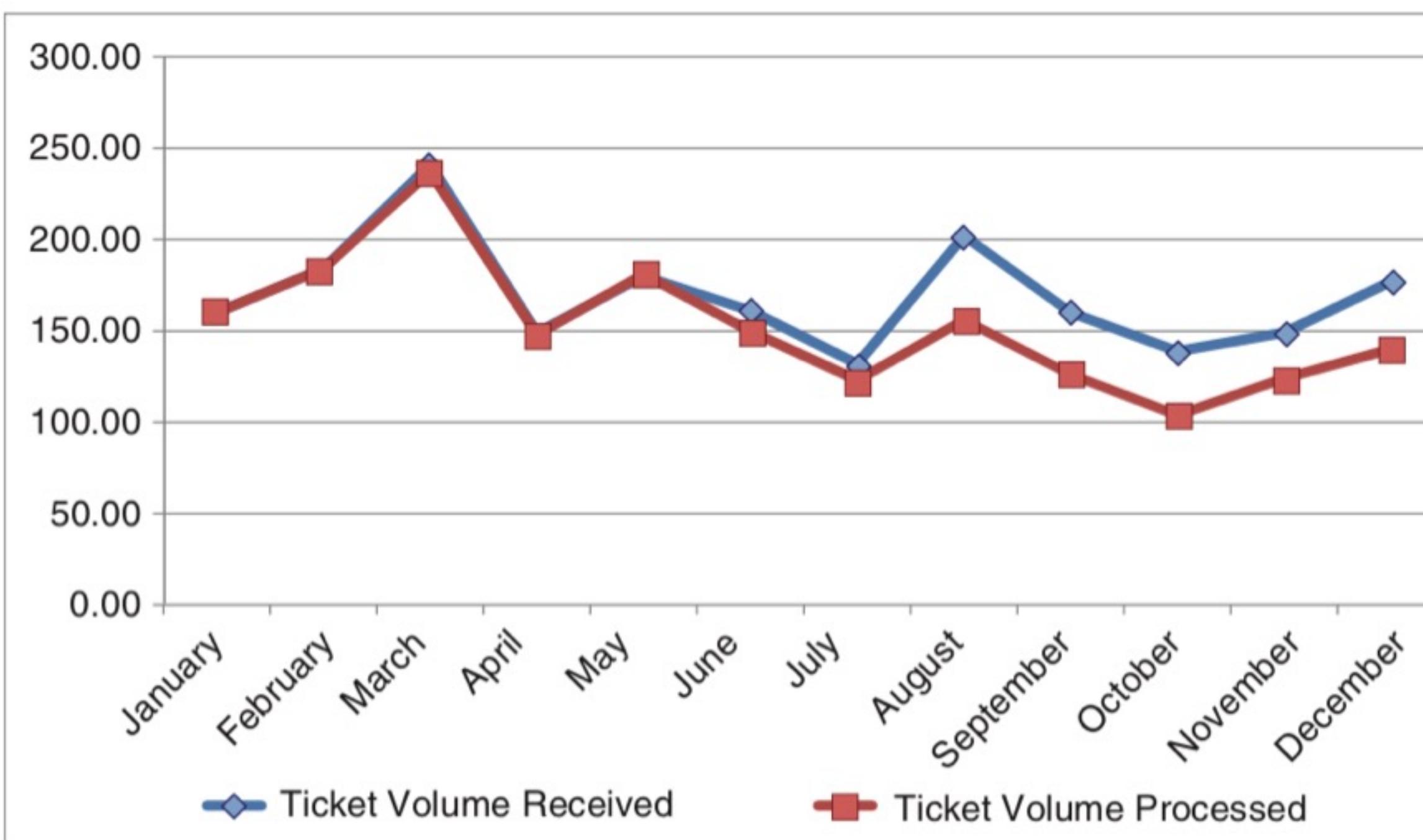
“

The ceramics teacher announced on opening day that he was dividing the class into two groups. All those on the left side of the studio, he said, would be graded solely on the quantity of work they produced, all those on the right solely on its quality. His procedure was simple: on the final day of class he would bring in his bathroom scales and weigh the work of the “quantity” group: fifty pounds of pots rated an “A”, forty pounds a “B”, and so on. Those being graded on “quality”, however, needed to produce only one pot —albeit a perfect one —to get an “A”.

Well, came grading time and a curious fact emerged: **the works of highest quality were all produced by the group being graded for quantity.** It seems that while the "quantity" group was busily churning out piles of work—and learning from their mistakes —the “quality” group had sat theorizing about perfection, and in the end had little more to show for their efforts than grandiose theories and a pile of dead clay.

”

## **From exploring to explaining** *examples for discussion*



## Please approve the hire of 2 FTEs

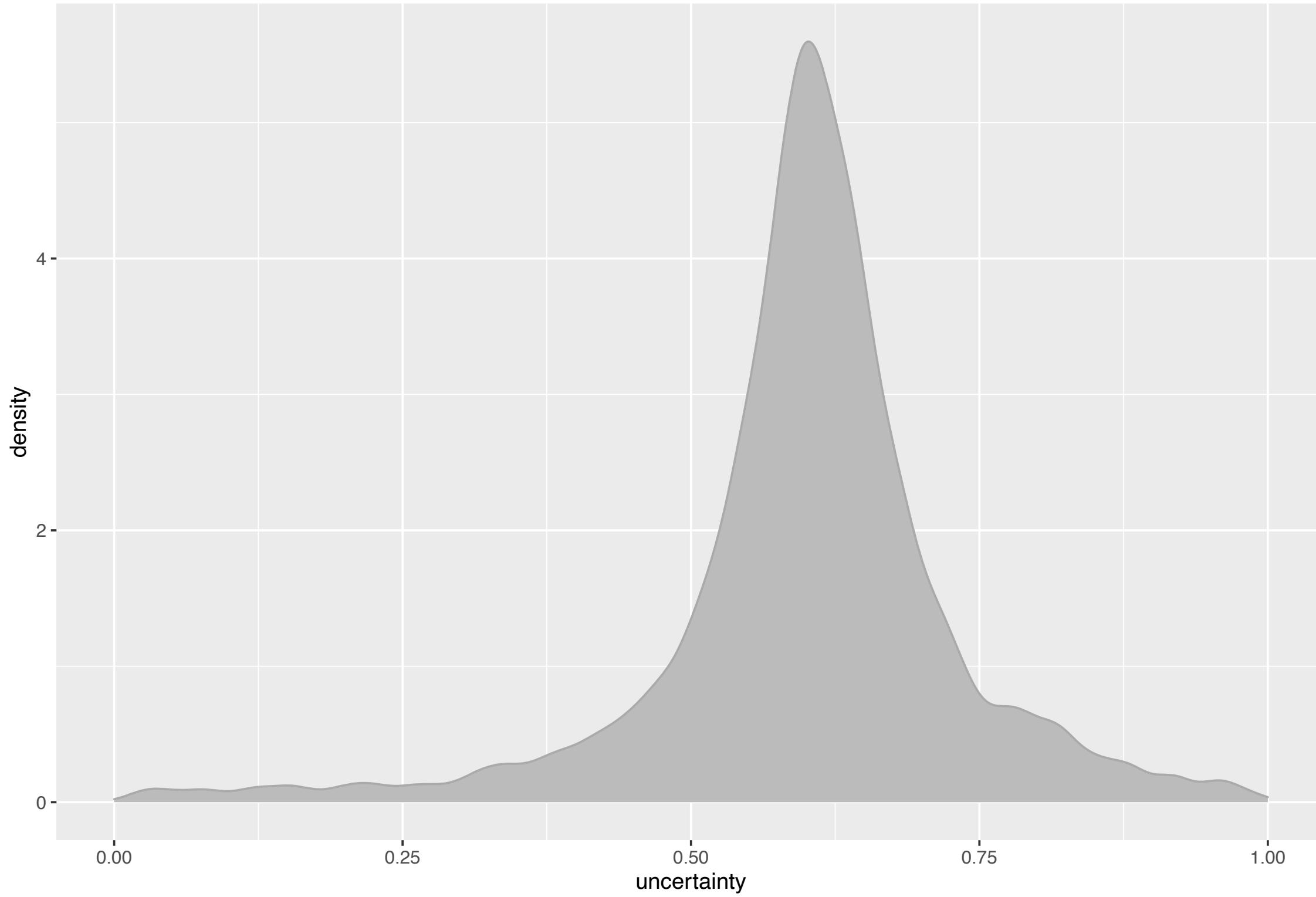
to backfill those who quit in the past year

Ticket volume over time



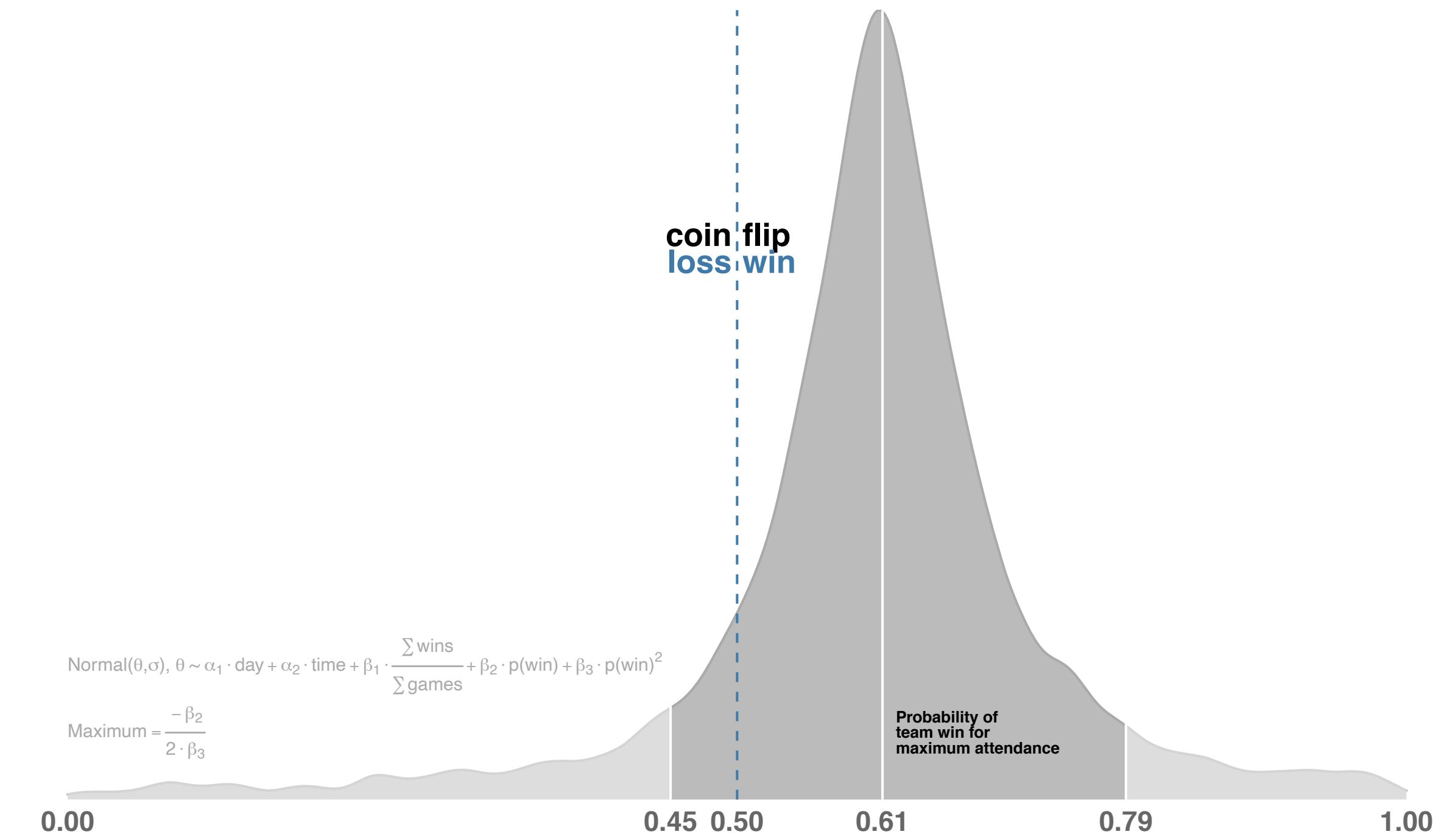
Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

# From exploring to explaining | Dodgers's example



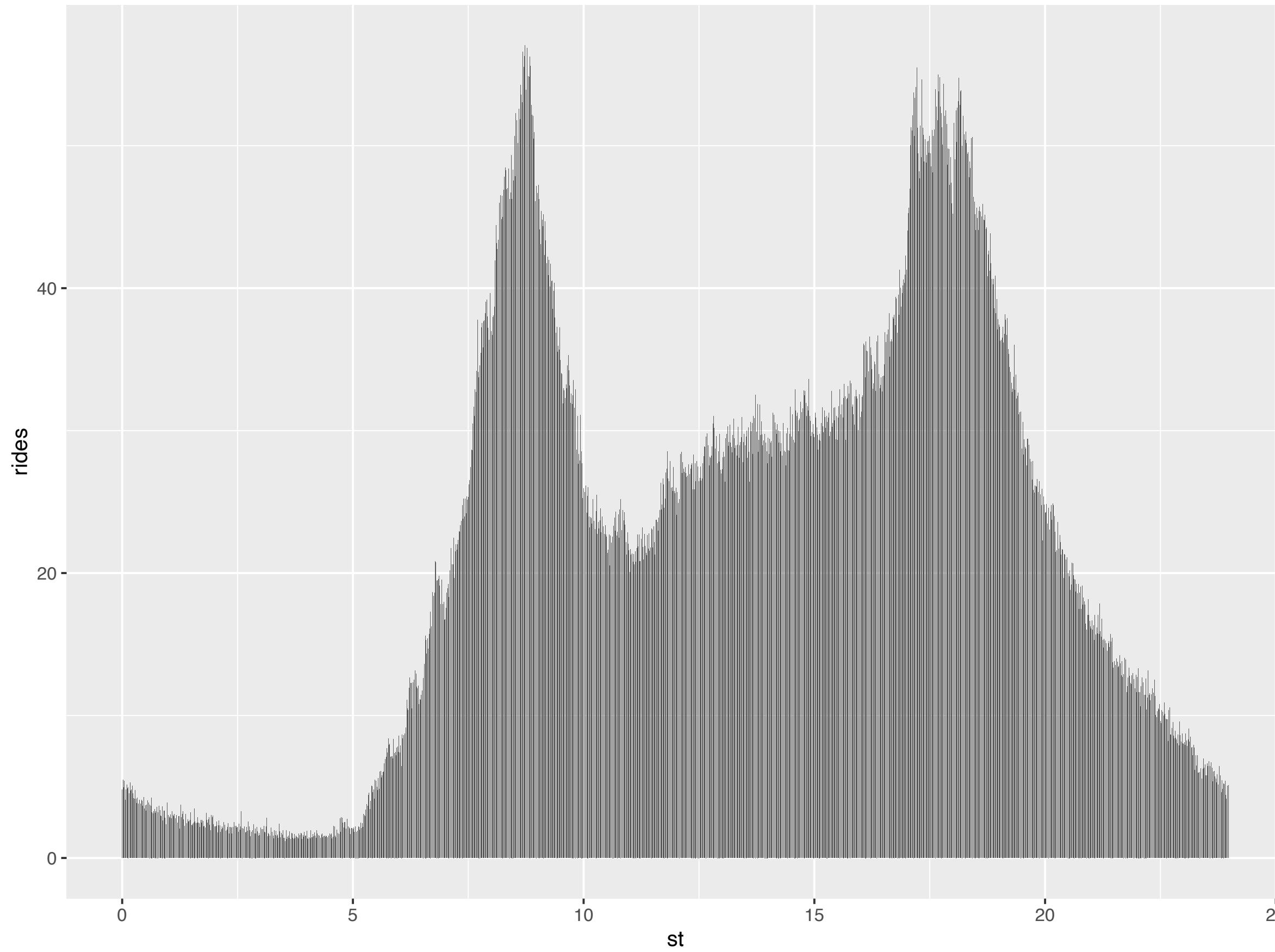
More fans generally pay admission to our games when the chance of winning was near a median of 0.61.

Fans want favorable odds without predicting the outcome.

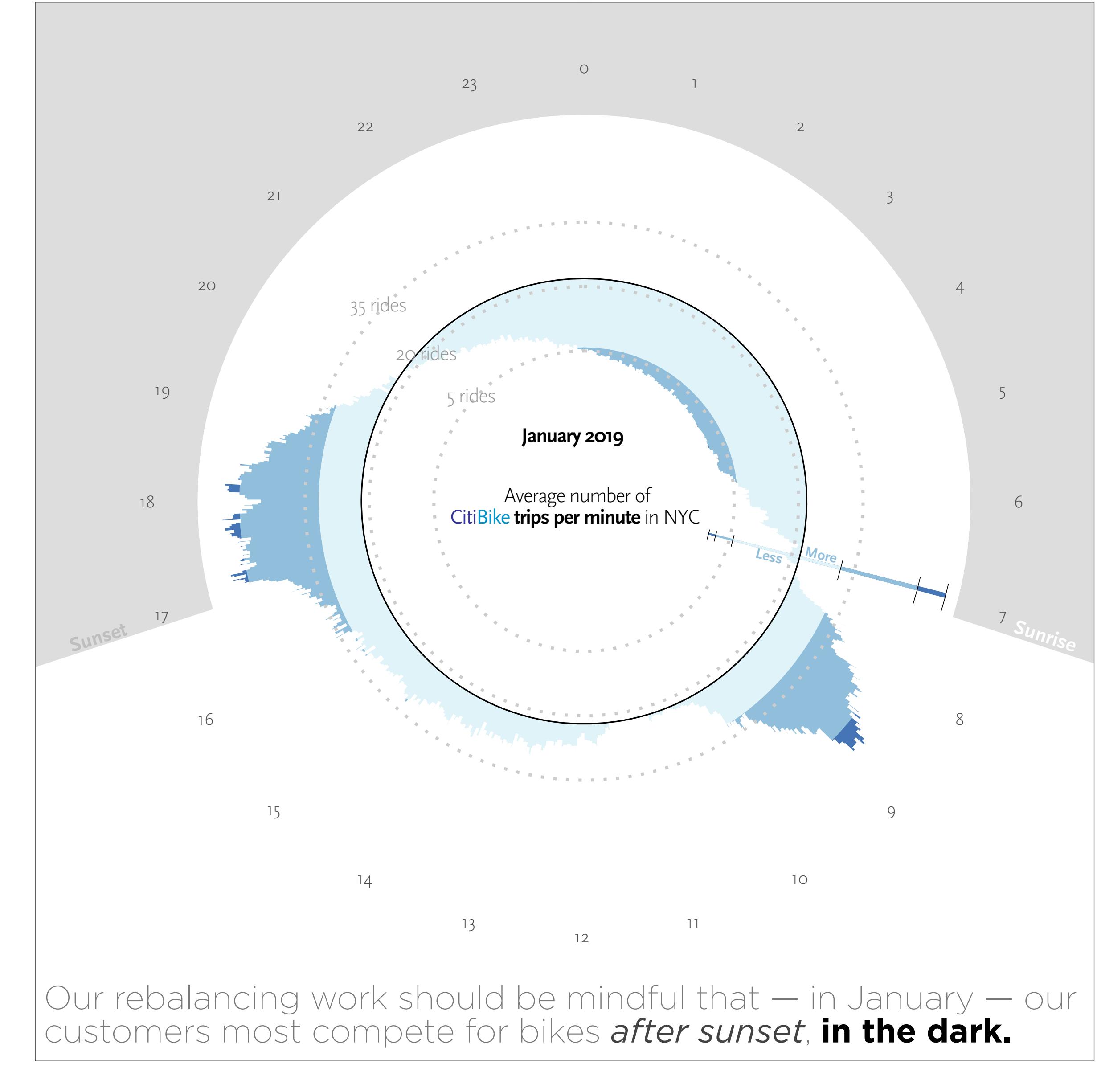
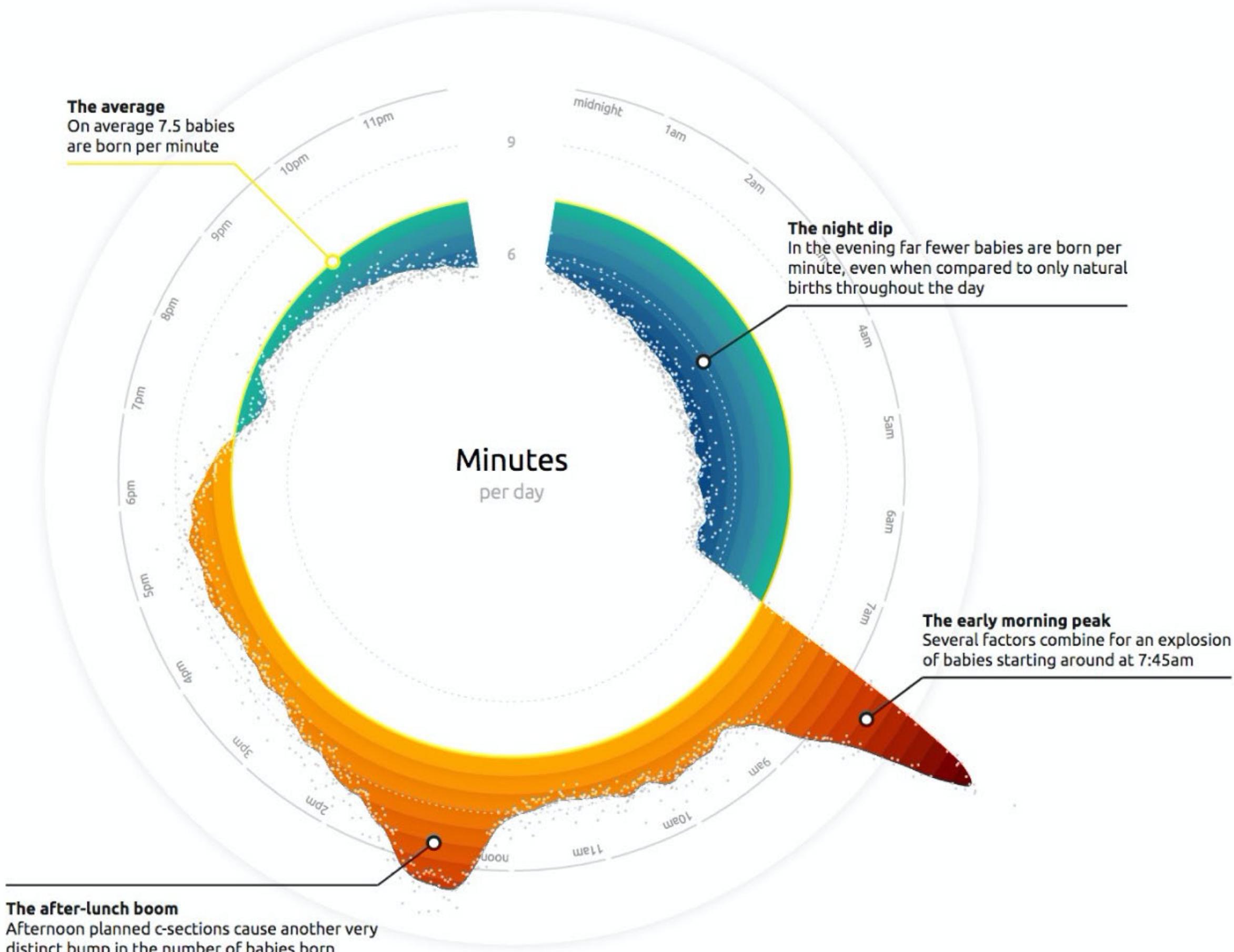


Sources: Pinnacle betting data, 2016; Retrosheet GameLogs, 2016

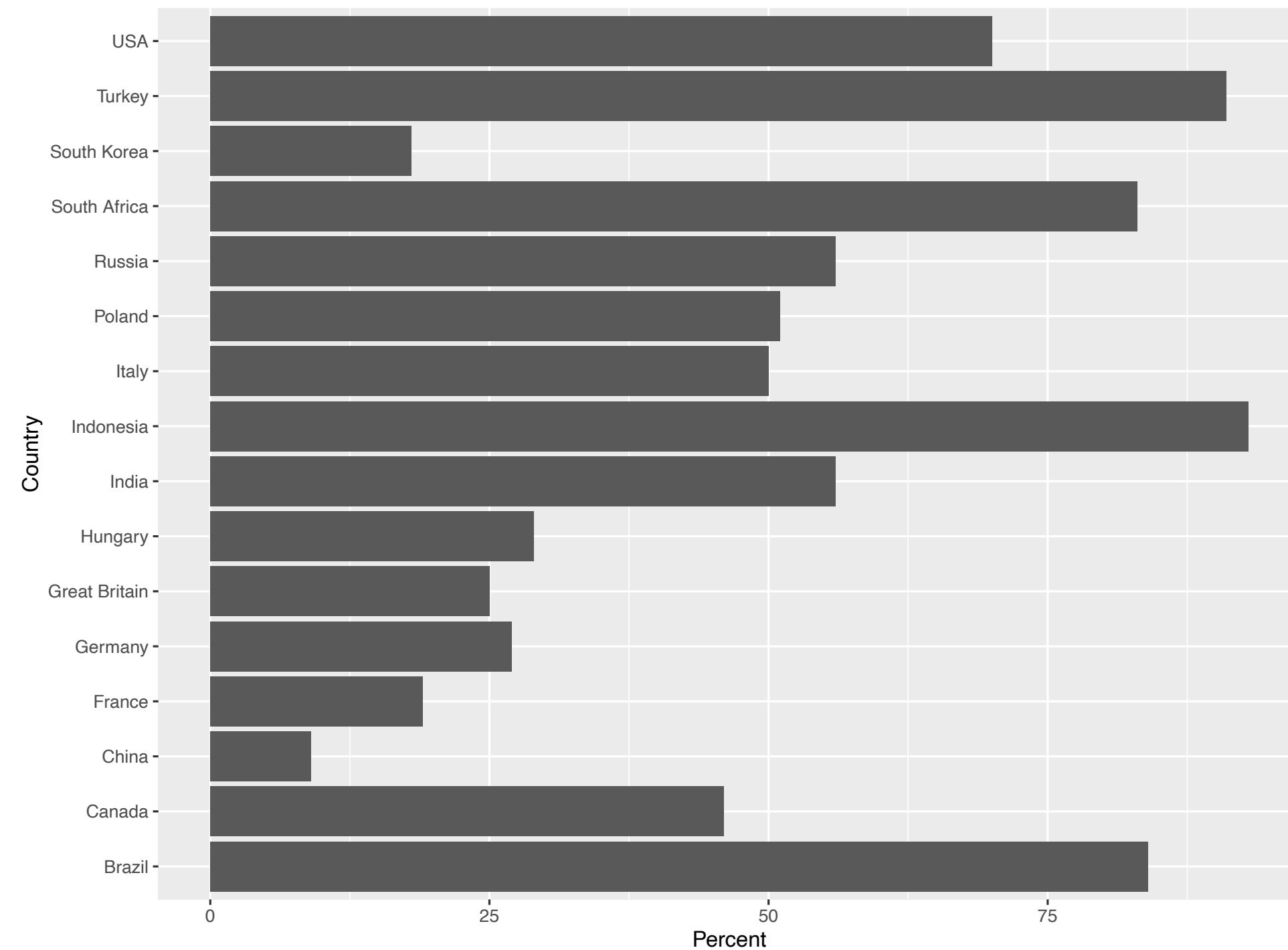
# From exploring to explaining | Citibike example



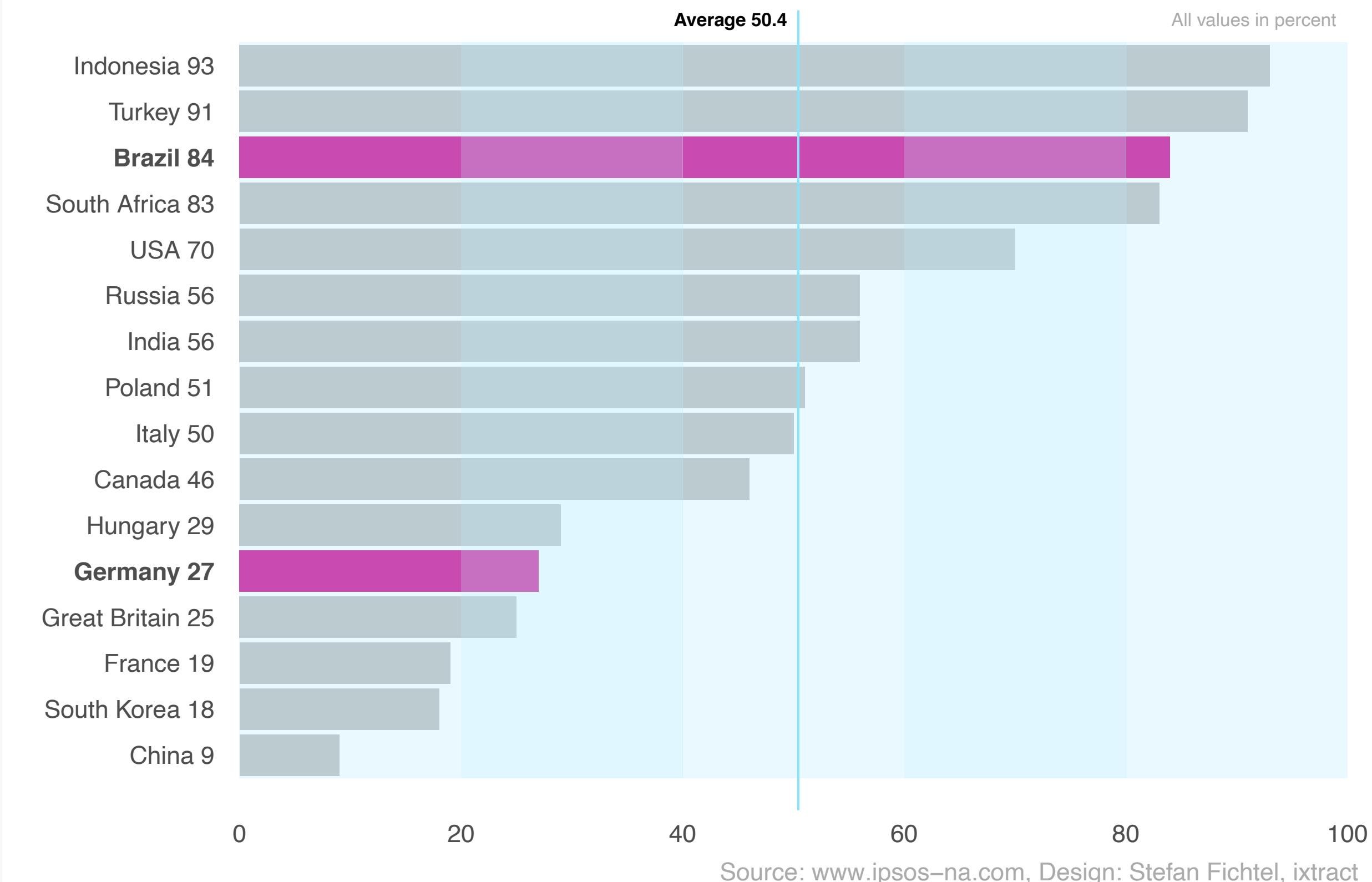
# From exploring to explaining | Citibike example



# From exploring to explaining | adapted from *Rahlf example*



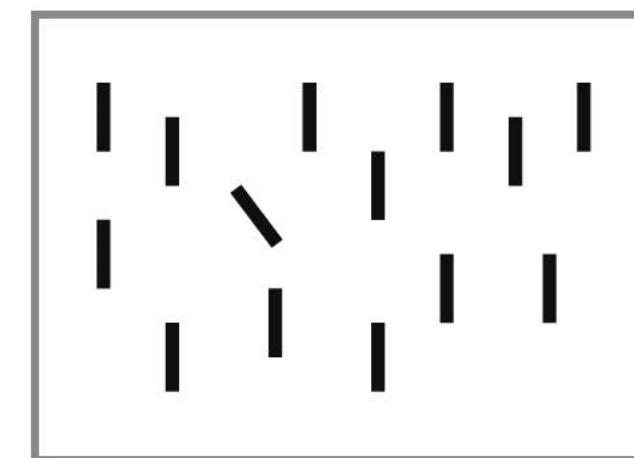
'I definitely believe in God or a Supreme Being'  
was said in 2010 in:



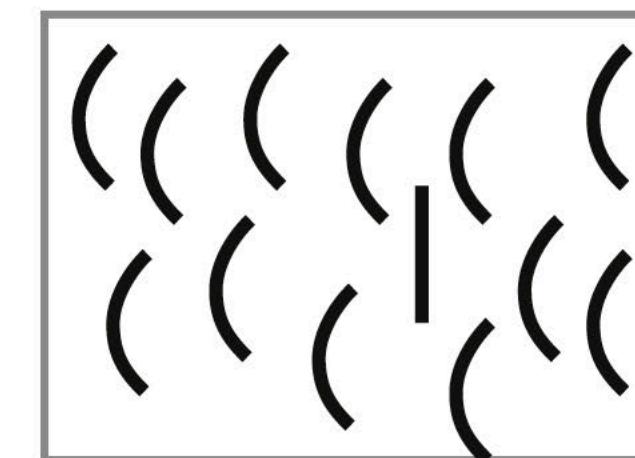
# focusing visual attention

# focusing visual attention, Gestalt principles | *pre-attentive attributes*

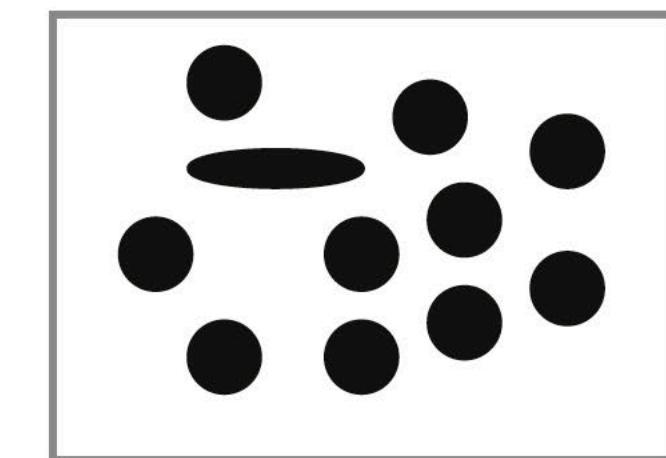
Orientation



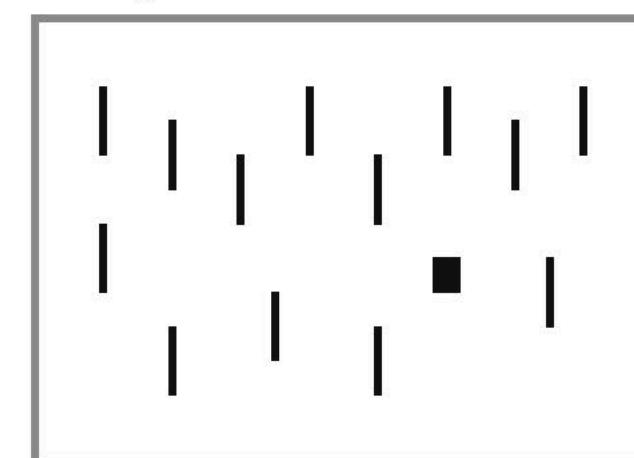
Curved straight



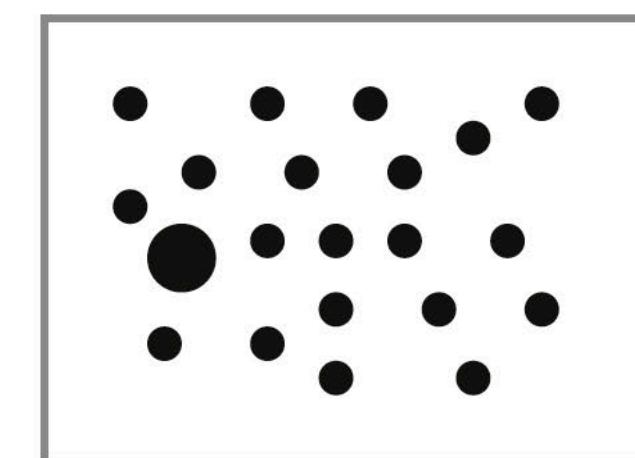
Shape



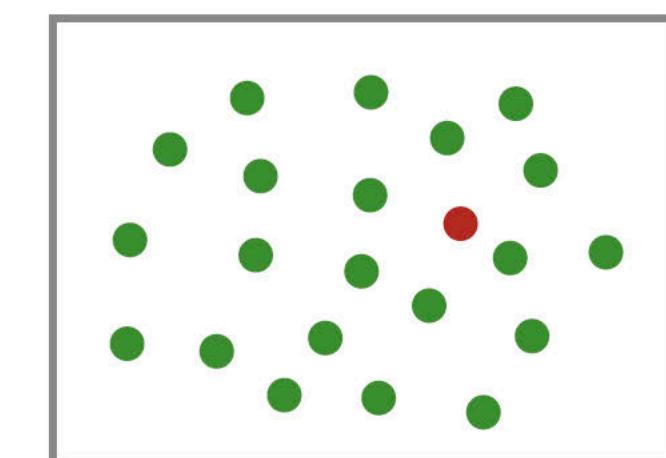
Shape



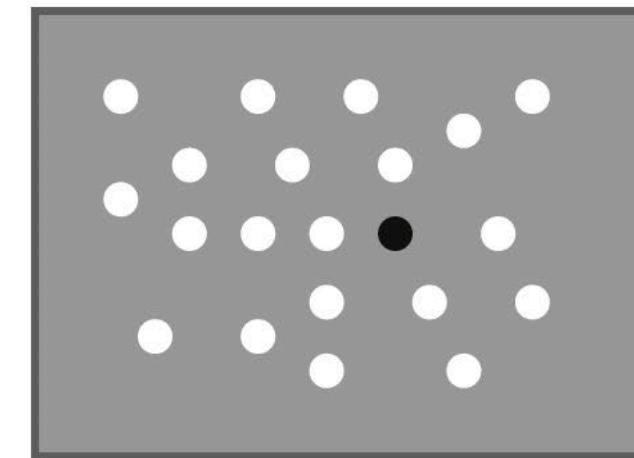
Size



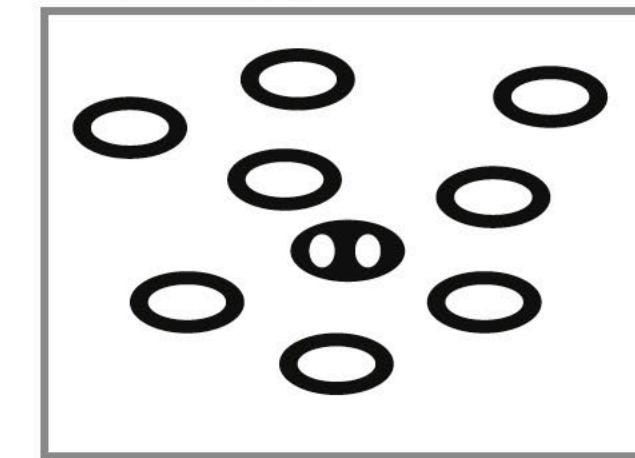
Color



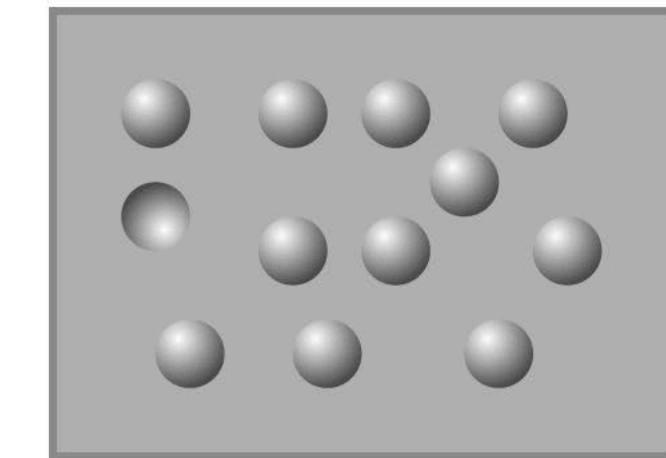
Light/dark



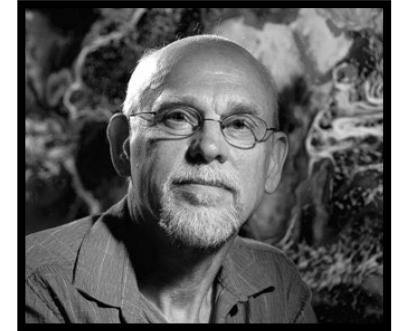
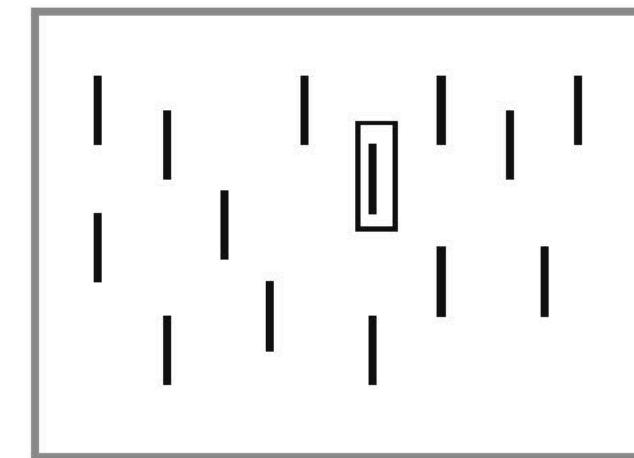
Topology (or count)



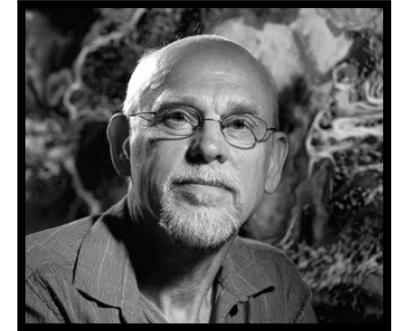
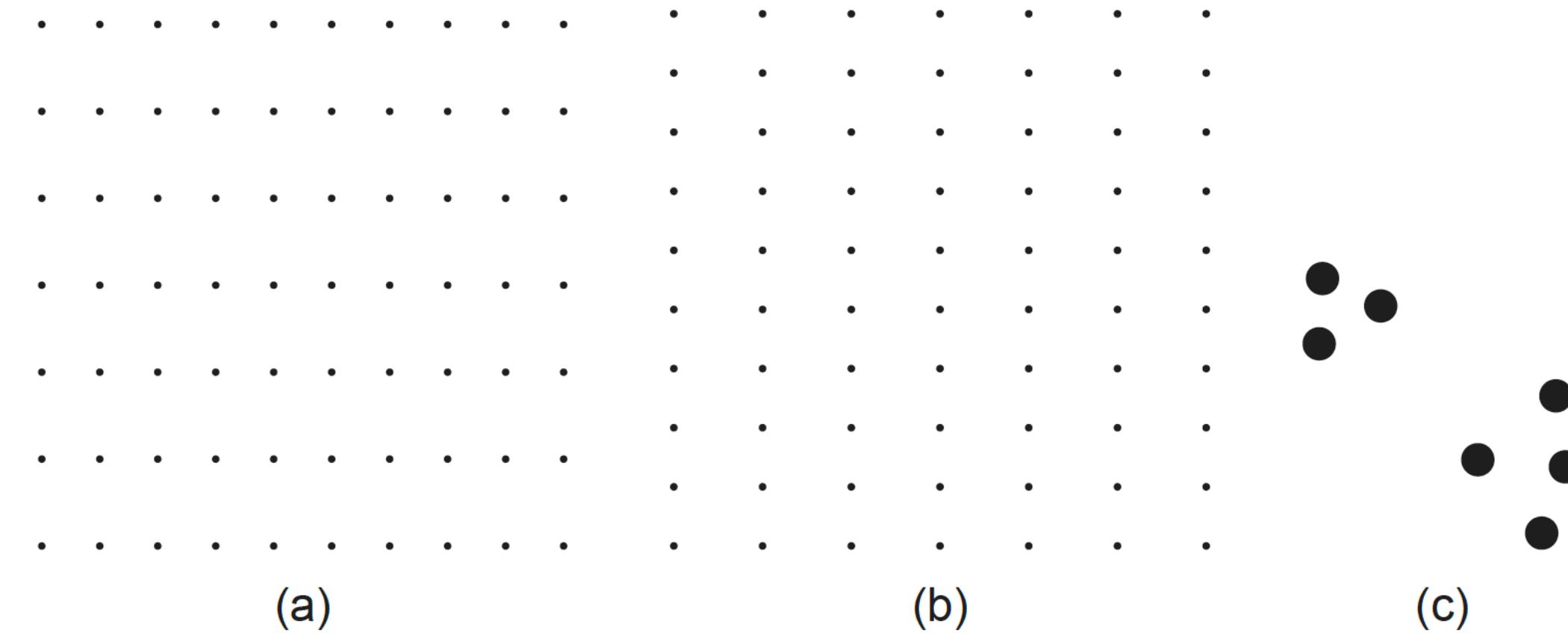
Convex/concave



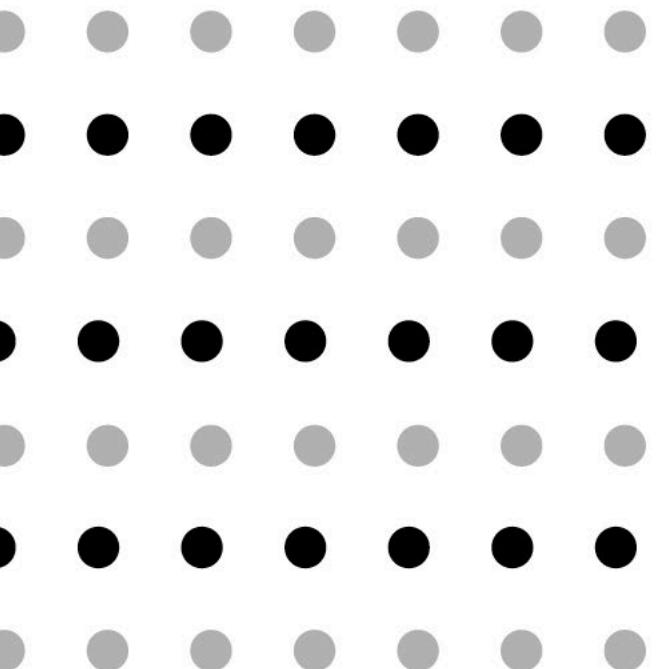
Addition



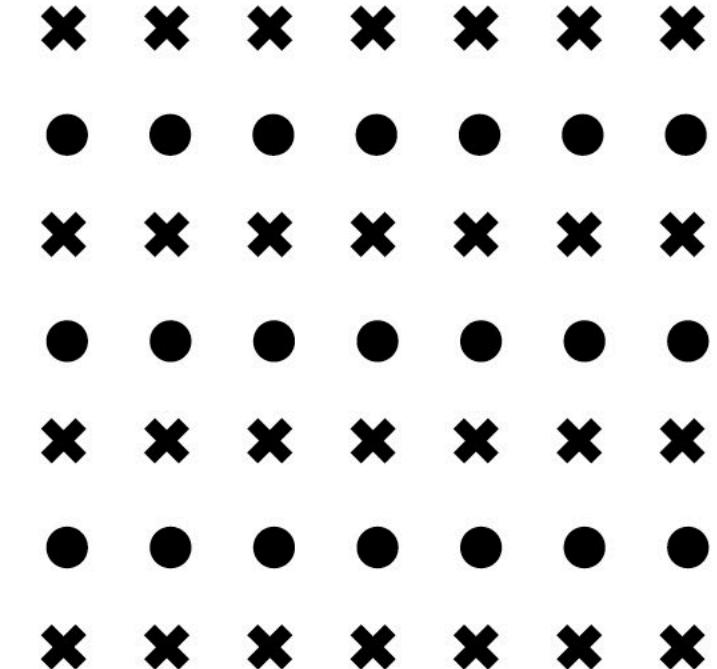
# focusing visual attention, Gestalt principles | *proximity*



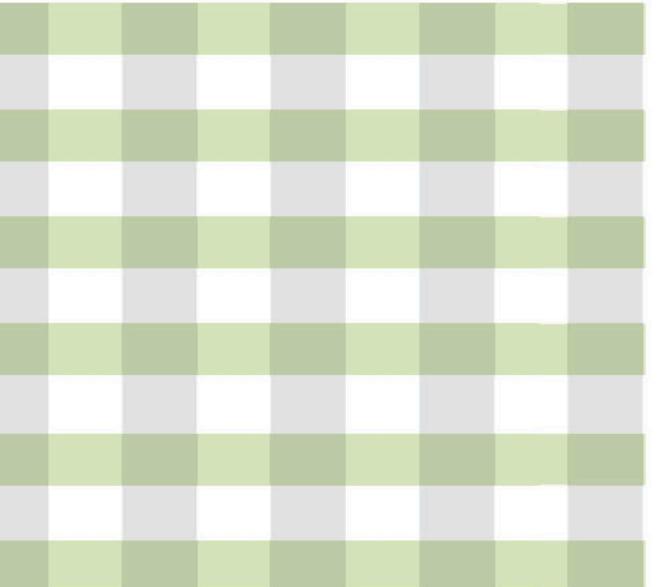
# focusing visual attention, Gestalt principles | *similarity*



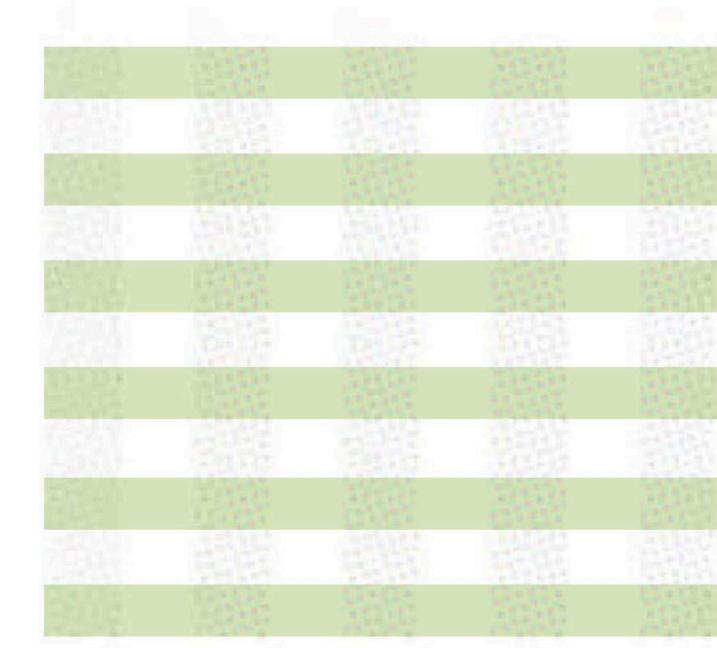
(a)



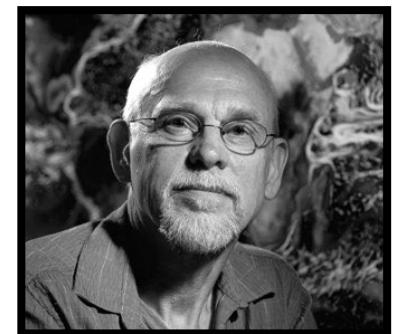
(b)



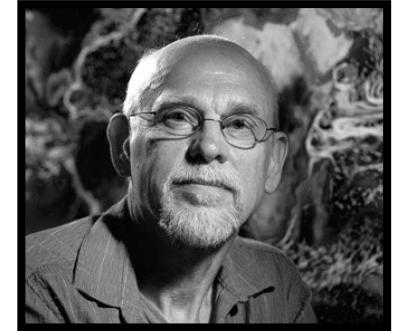
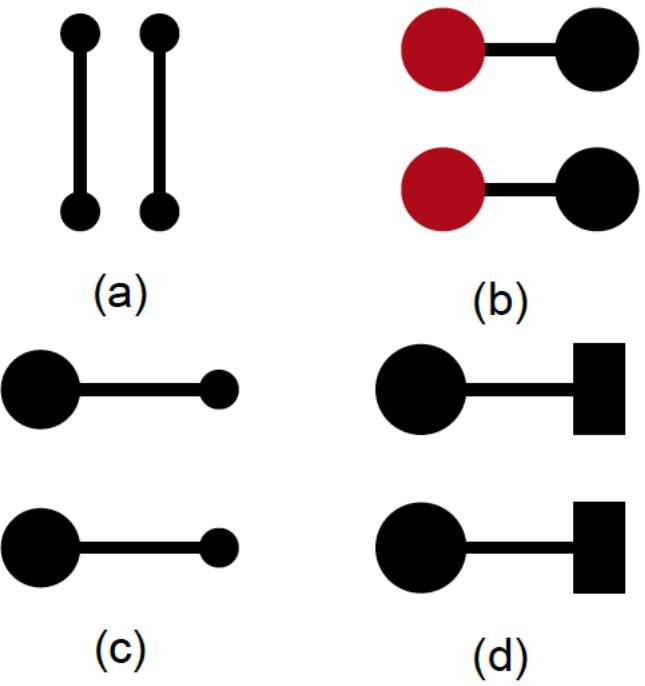
(c)



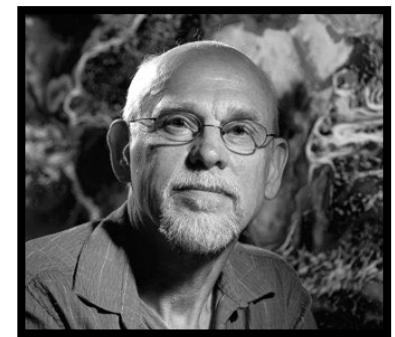
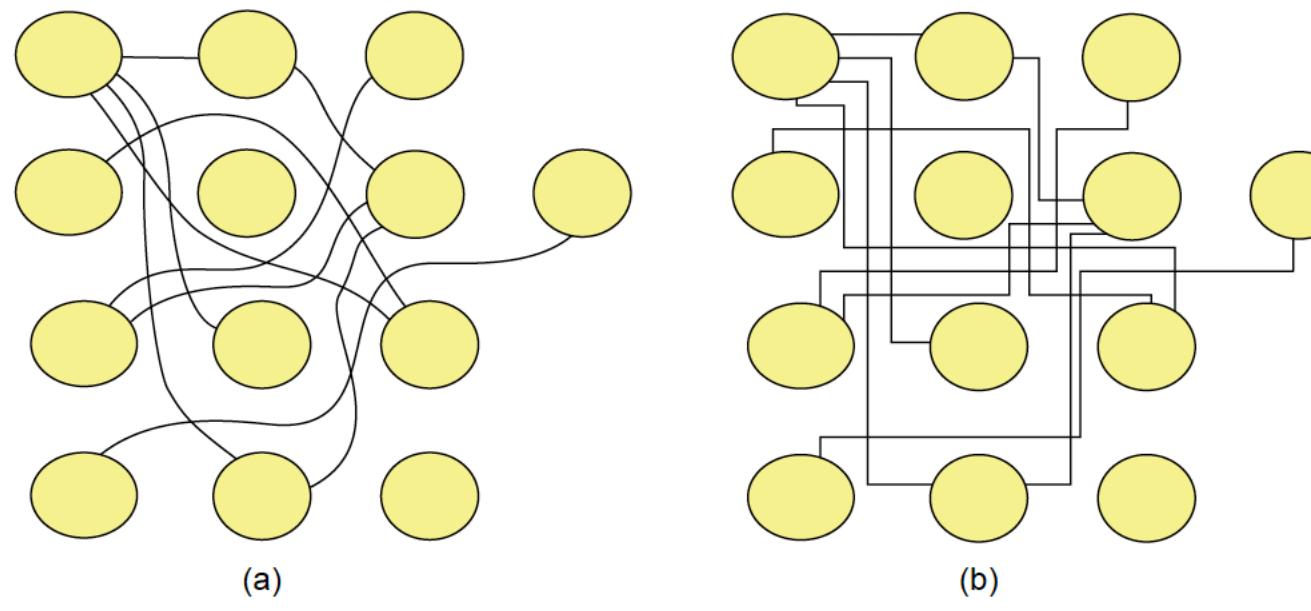
(d)



# focusing visual attention, Gestalt principles | *connectedness*



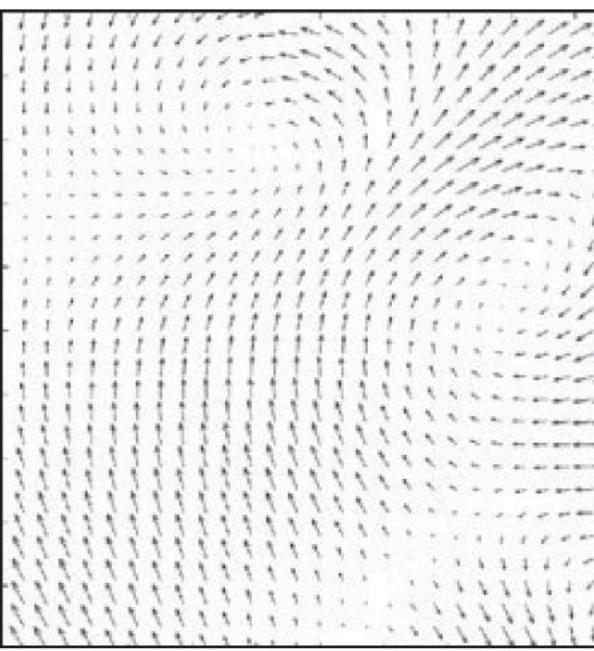
# focusing visual attention, Gestalt principles | *connectedness*



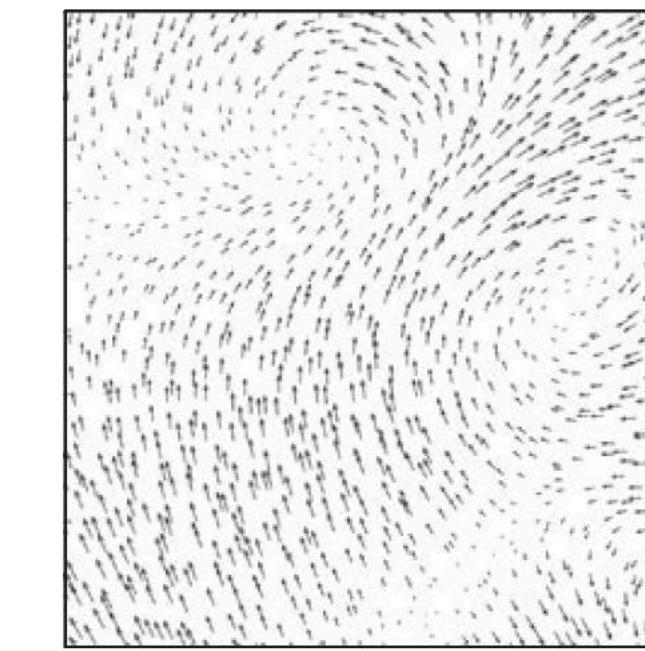
---

Ware, Colin

# focusing visual attention, Gestalt principles | *orientation, direction, magnitude*



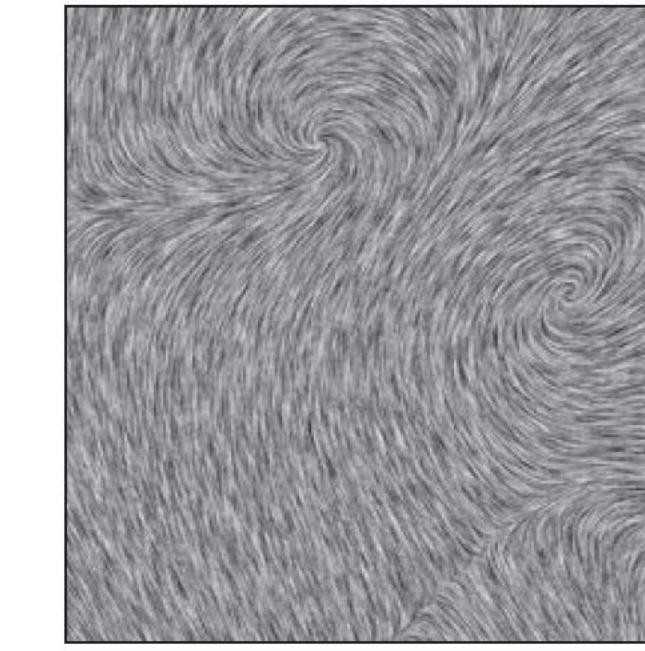
(a)



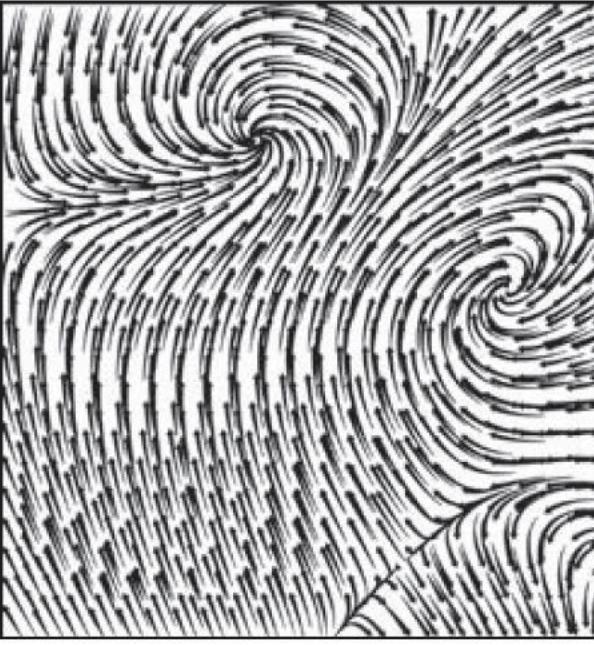
(b)



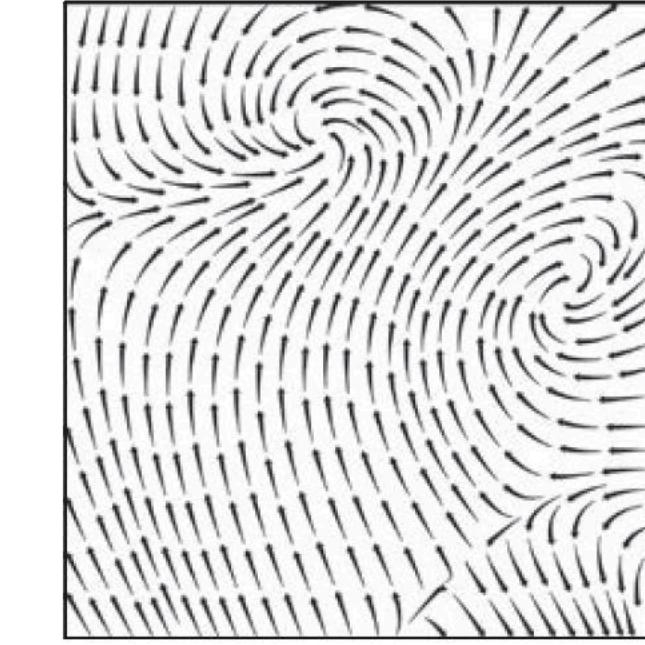
(c)



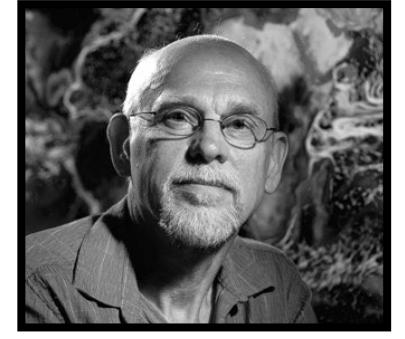
(d)



(e)



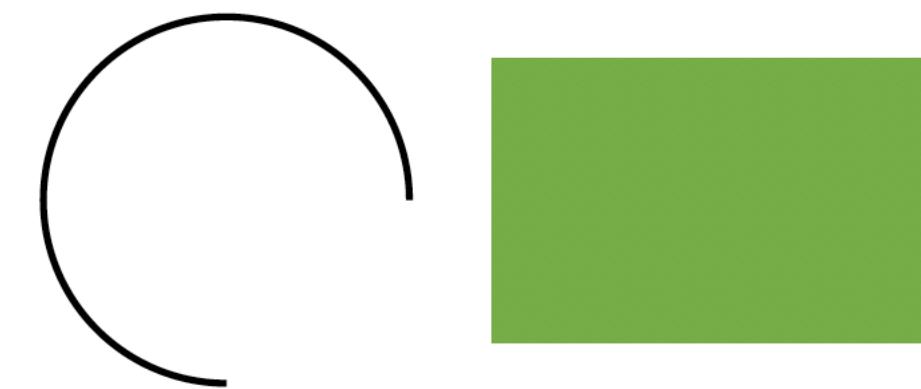
(f)



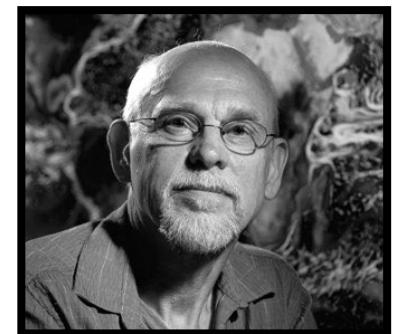
# focusing visual attention, Gestalt principles | closure



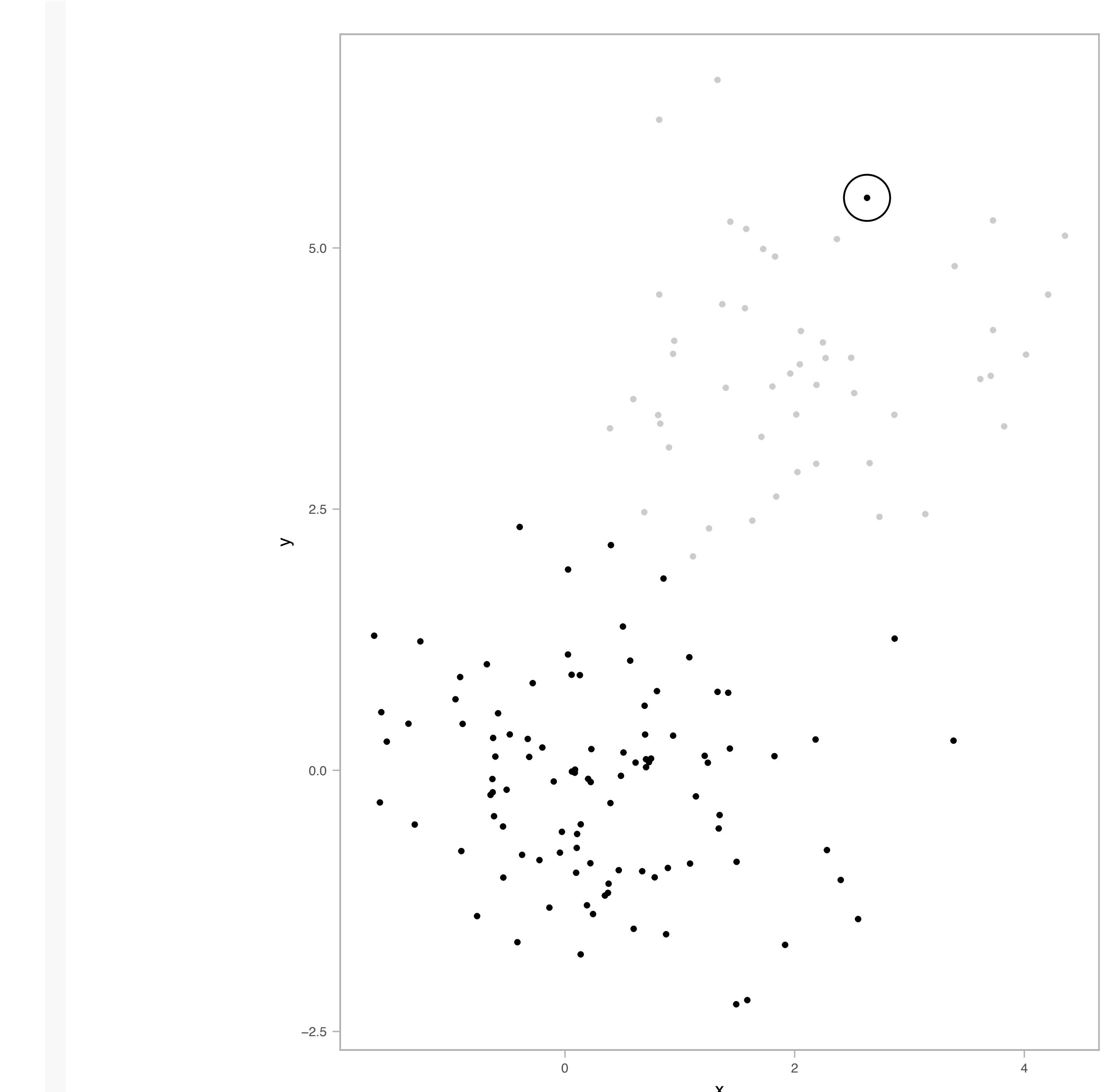
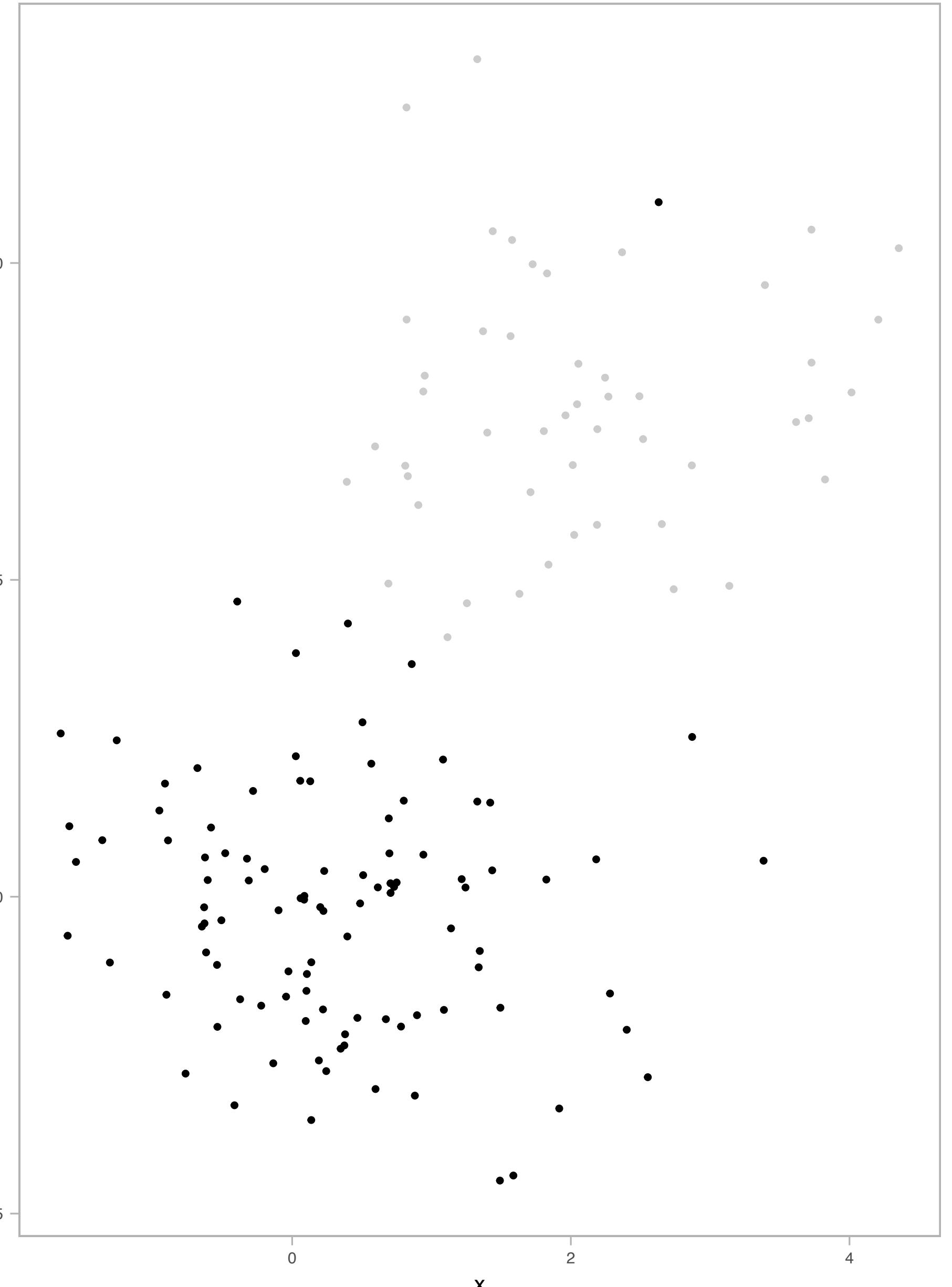
(a)



(b)

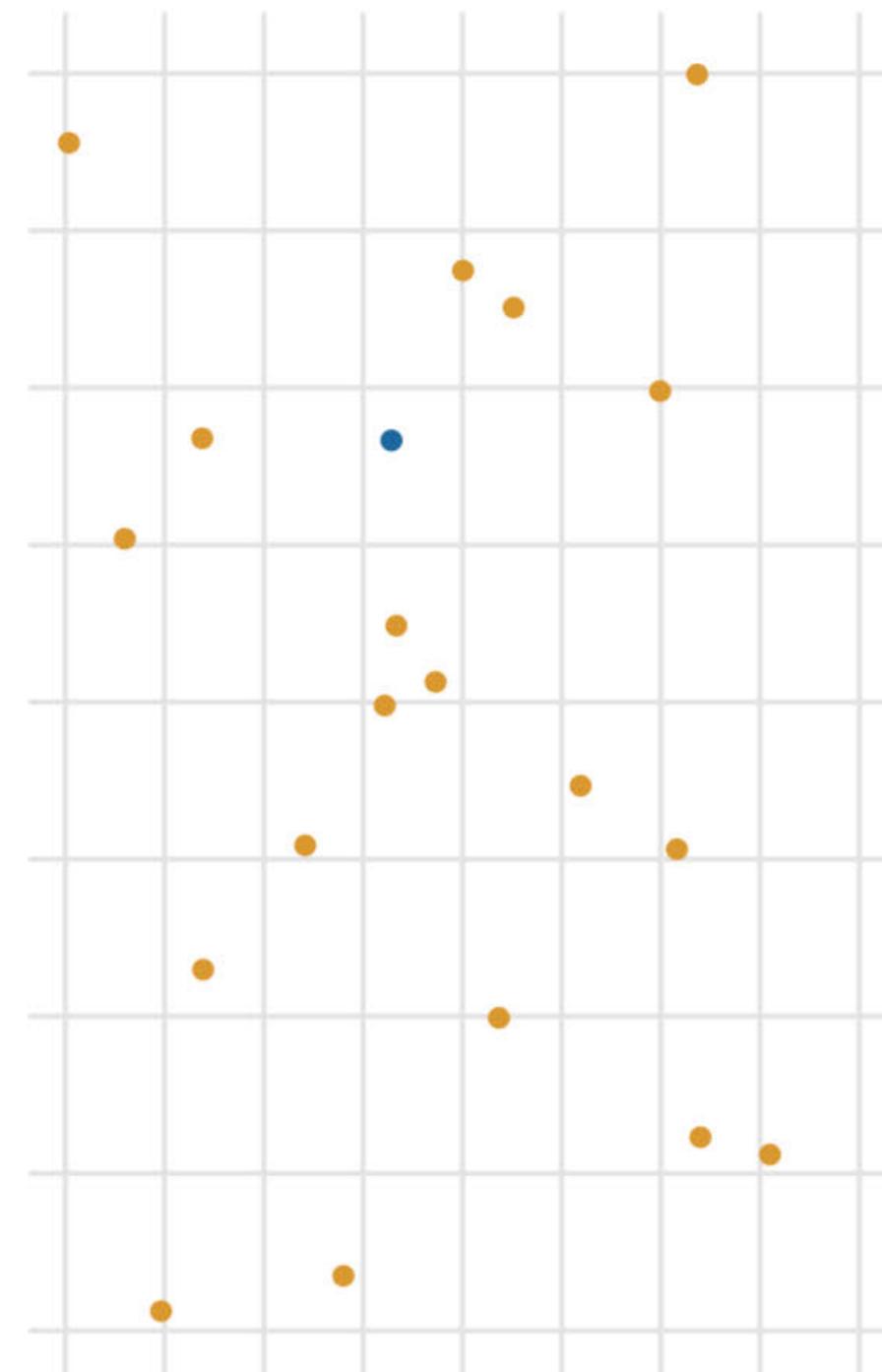


# focusing visual attention | *Gestalt principles*

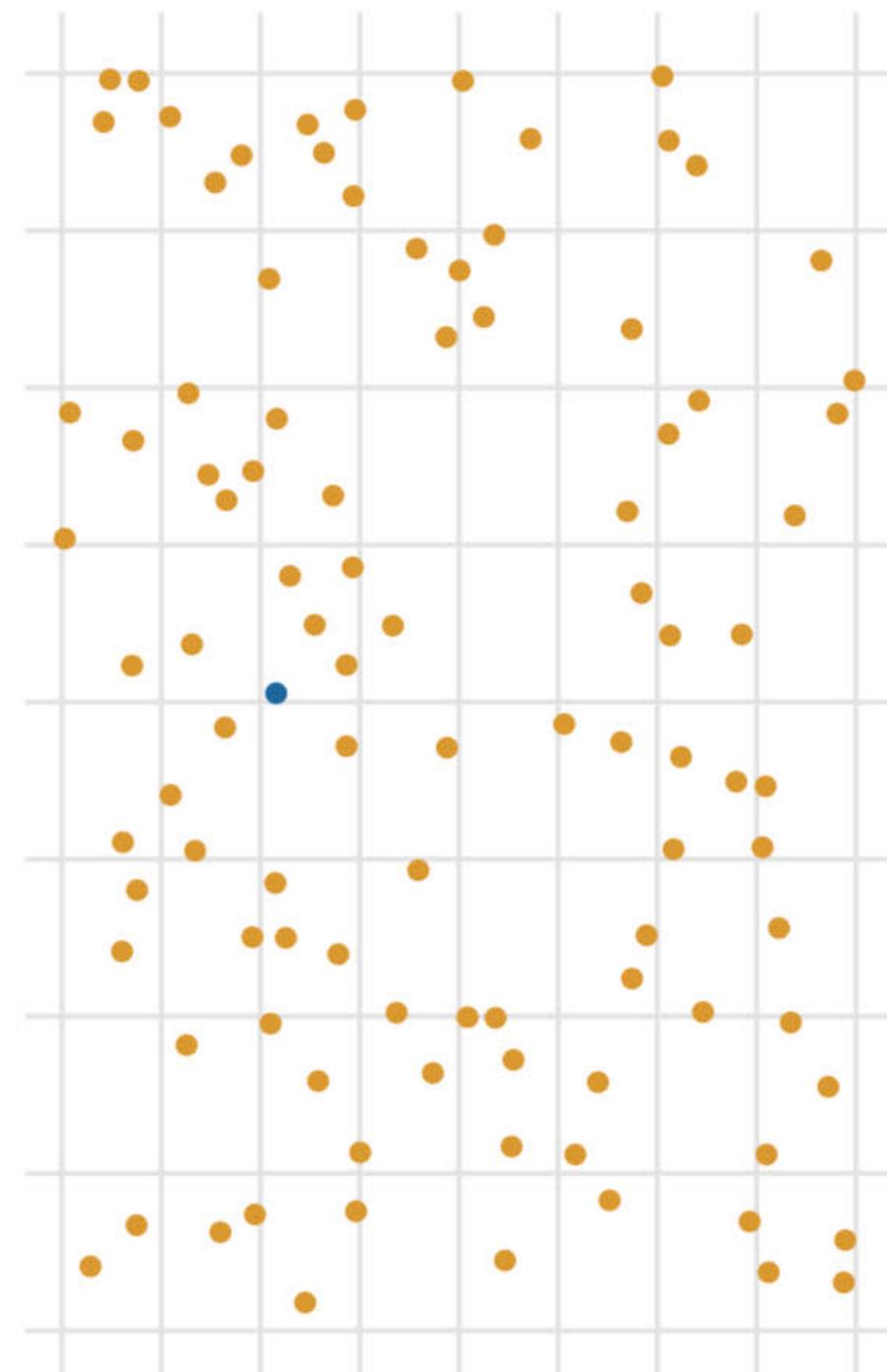


# focusing visual attention | *Gestalt principles*

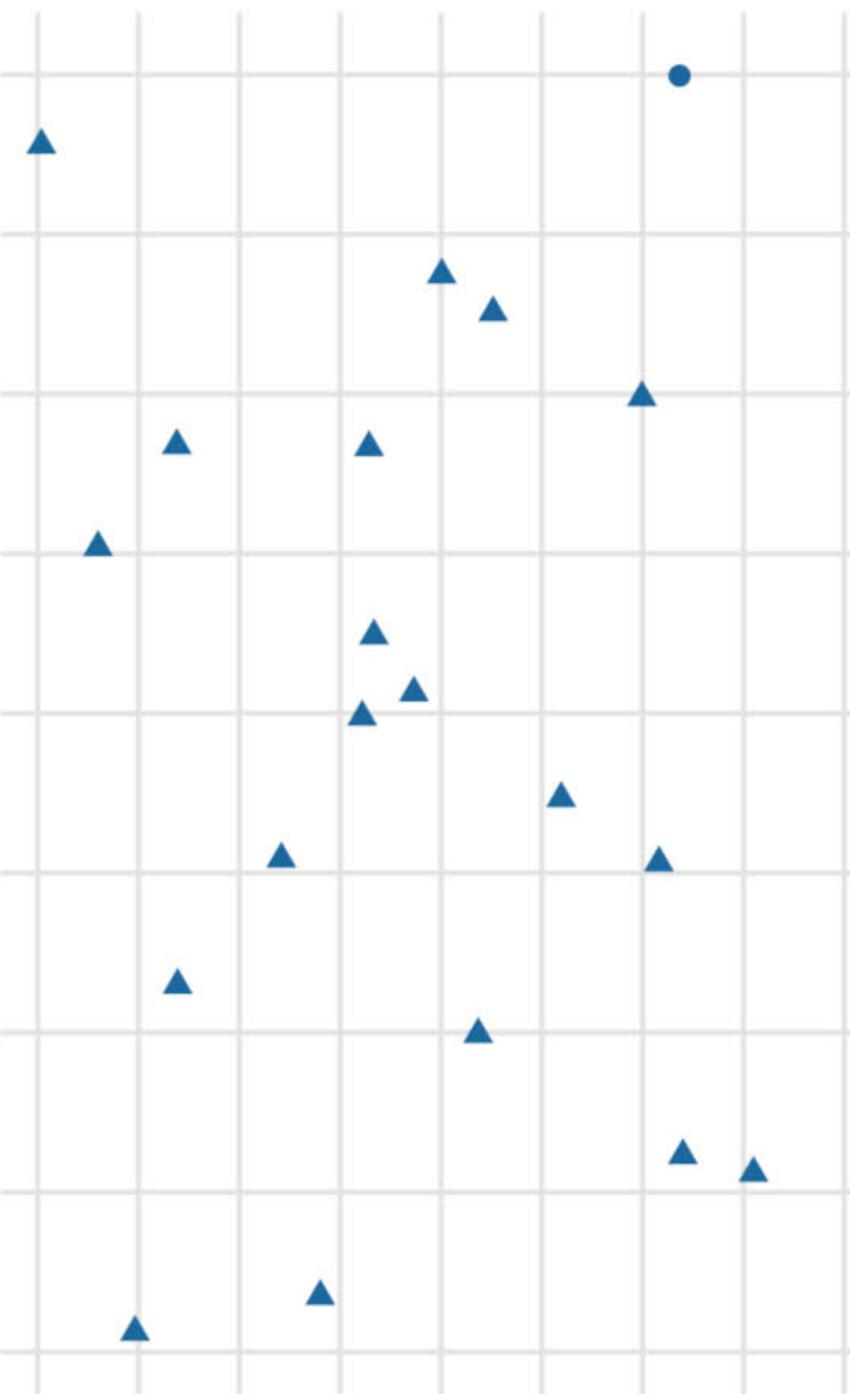
Color only,  $N = 20$



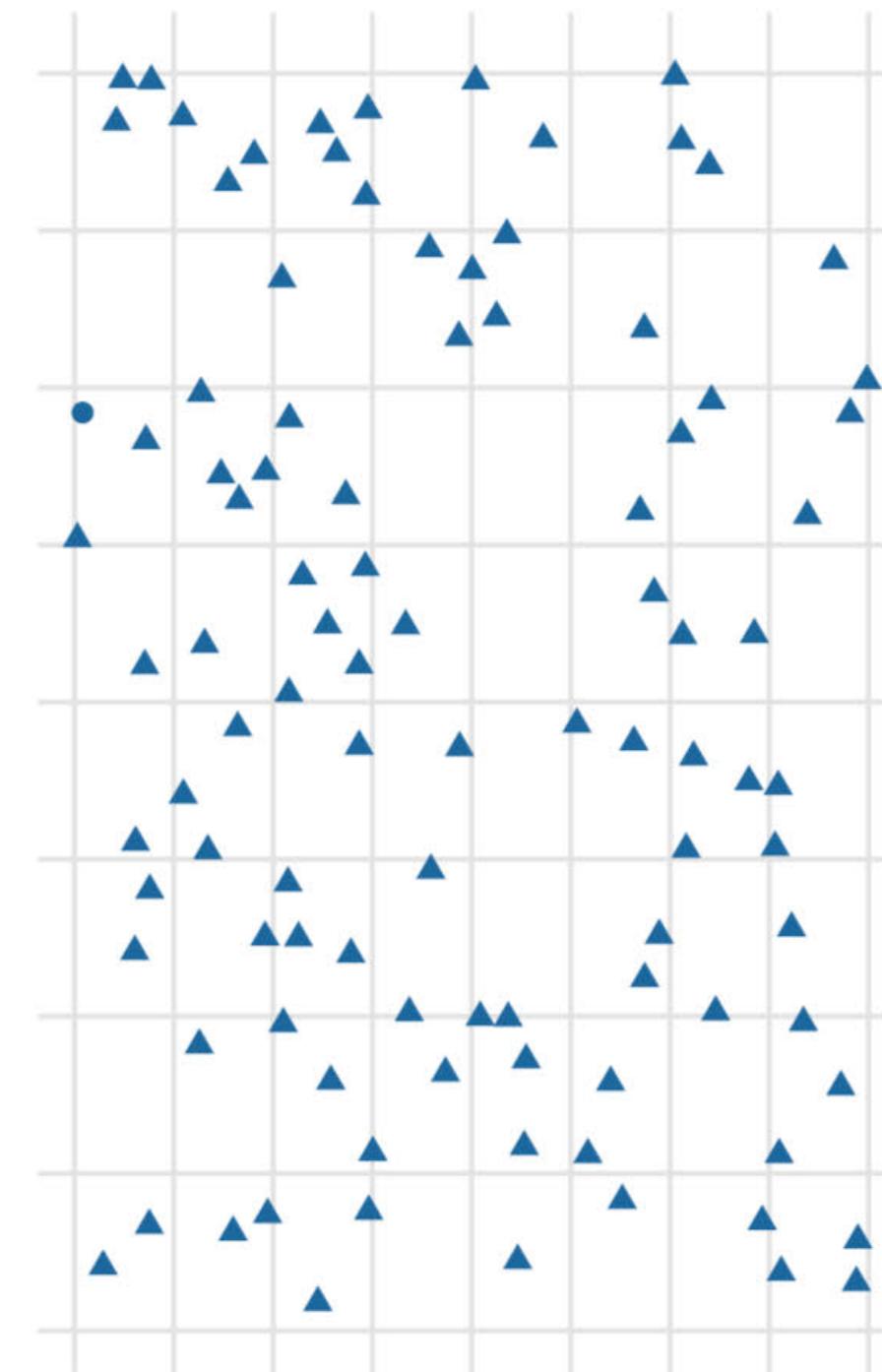
Color only,  $N = 100$



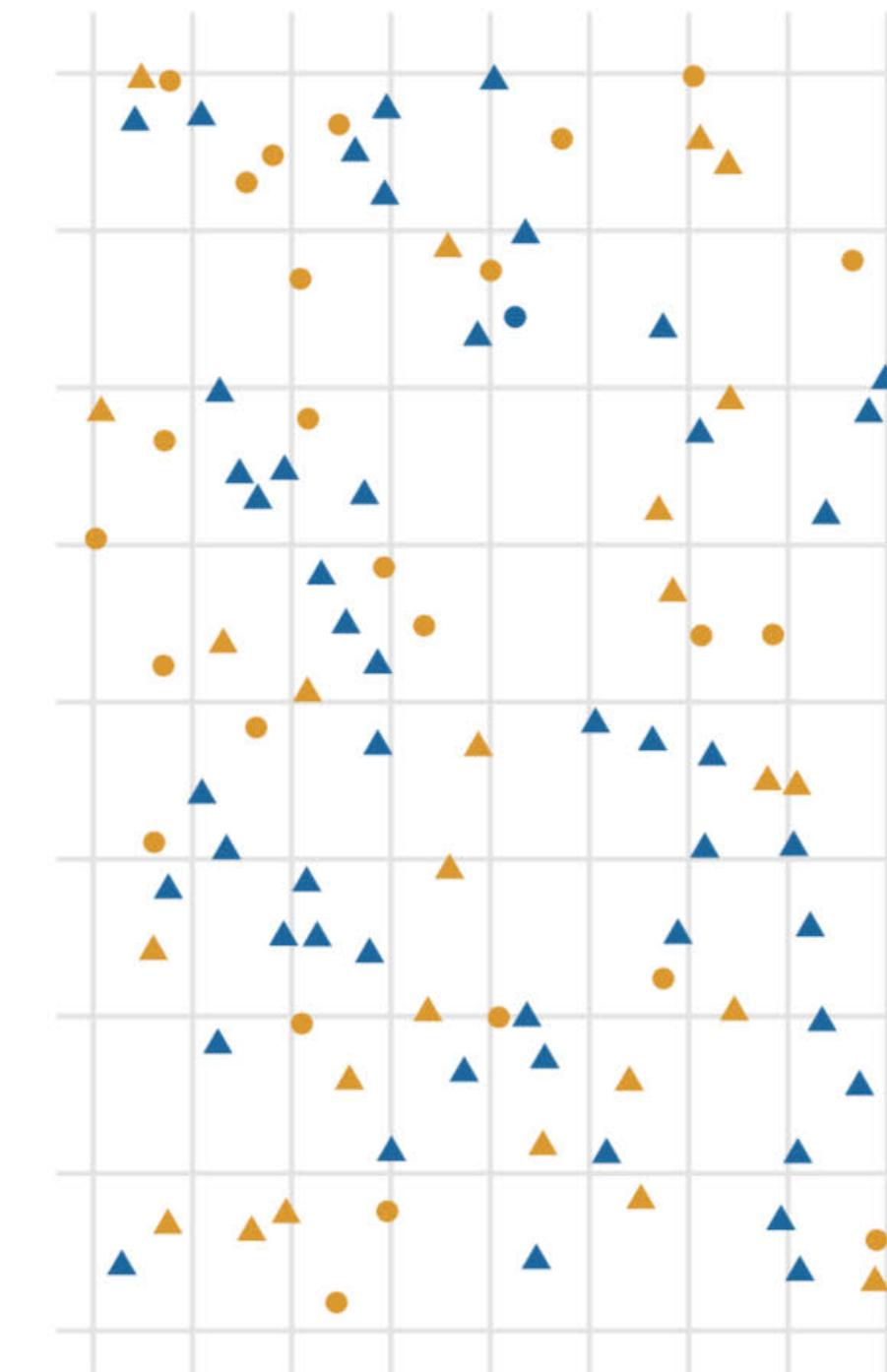
Shape only,  $N = 20$



Shape only,  $N = 100$

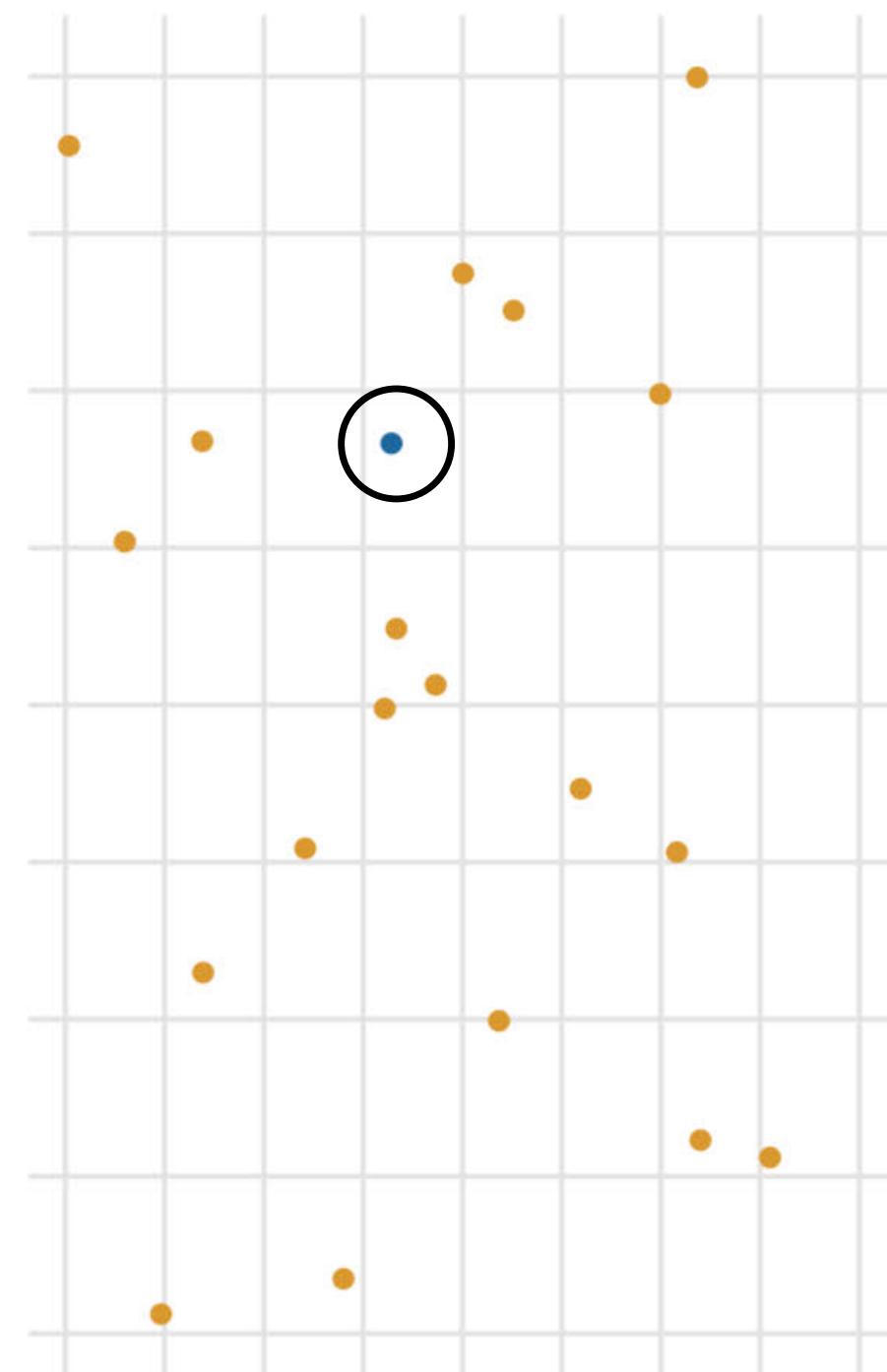


Color & shape,  $N = 100$

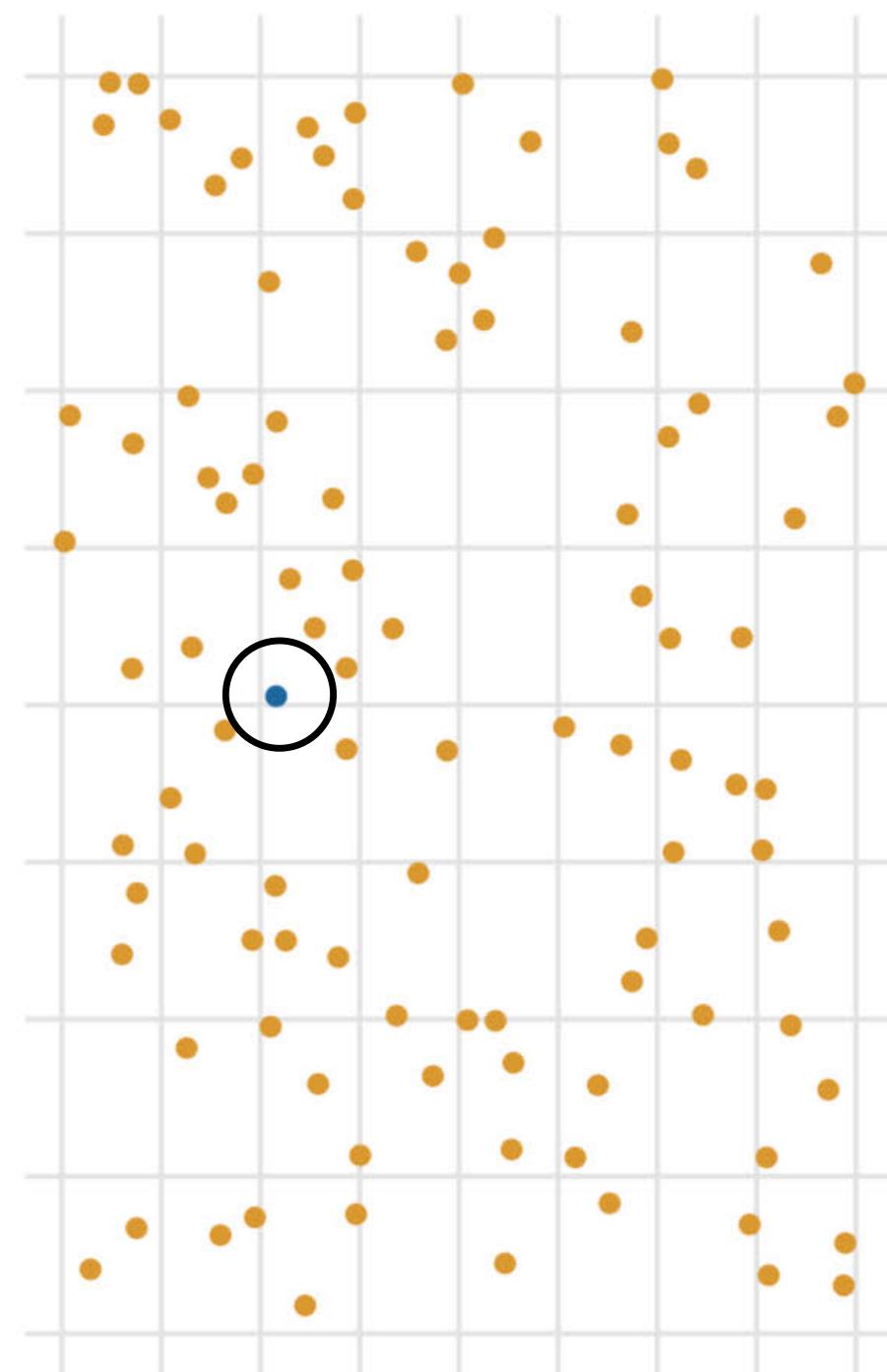


# focusing visual attention | *Gestalt principles*

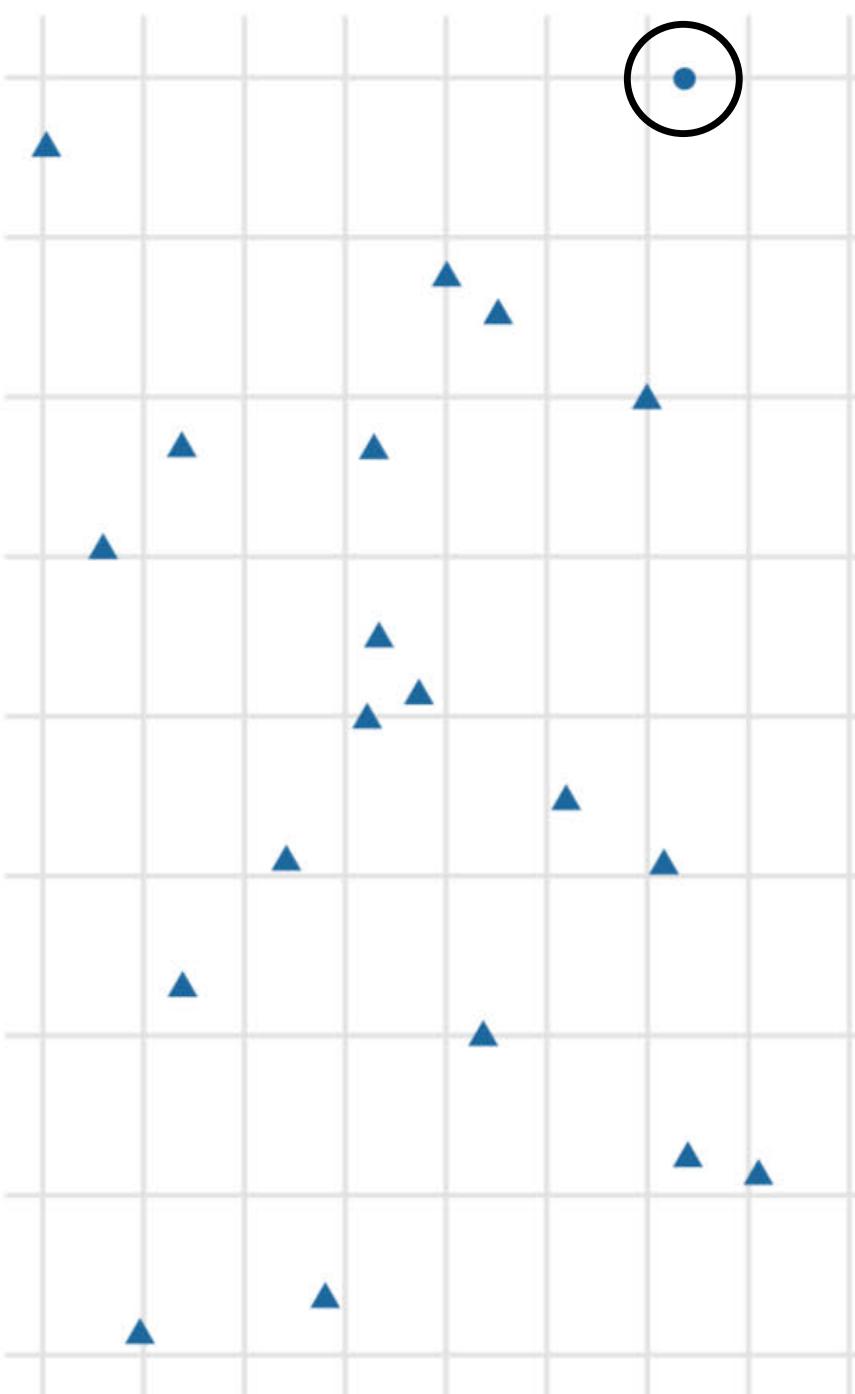
Color only,  $N = 20$



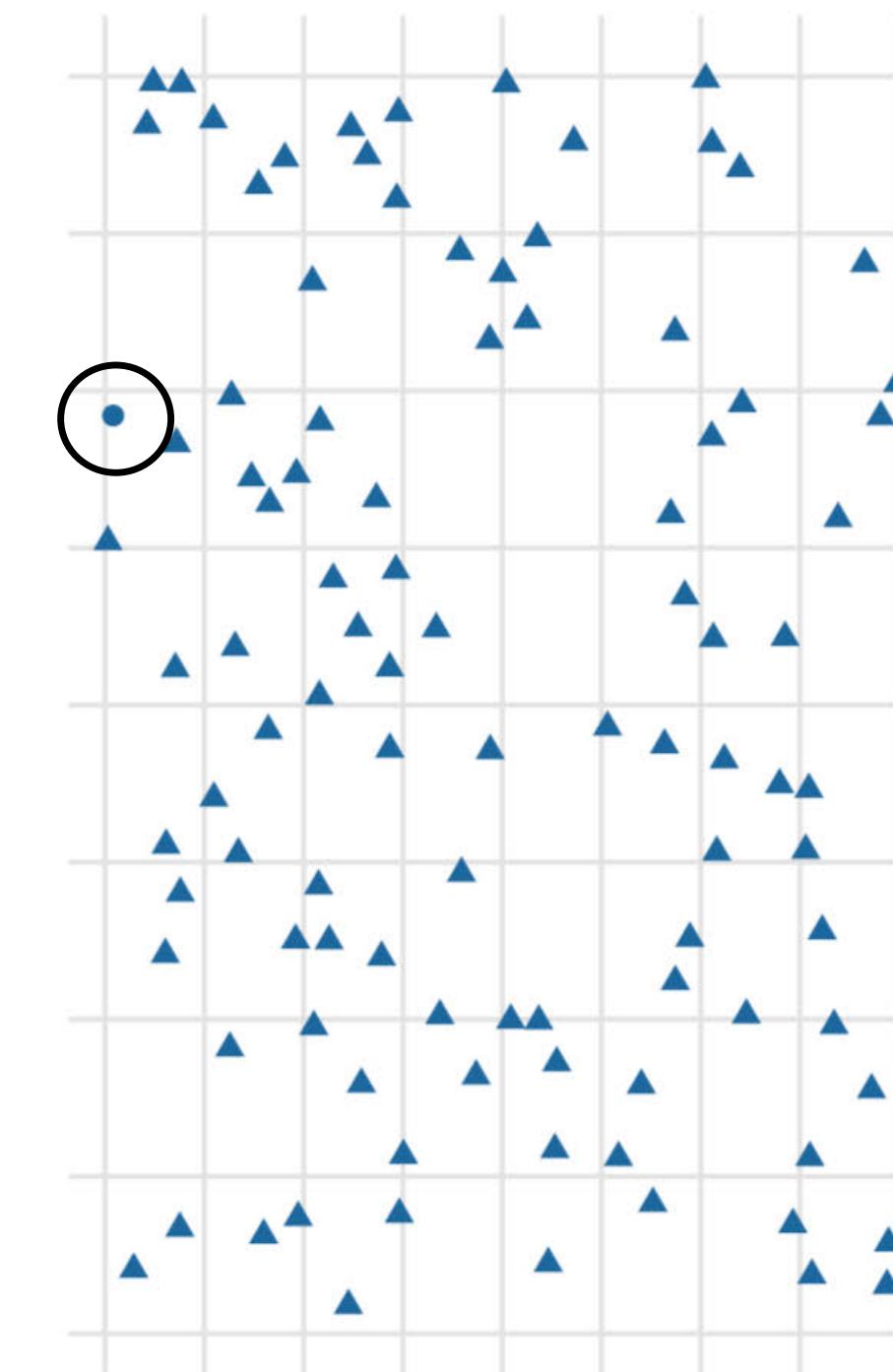
Color only,  $N = 100$



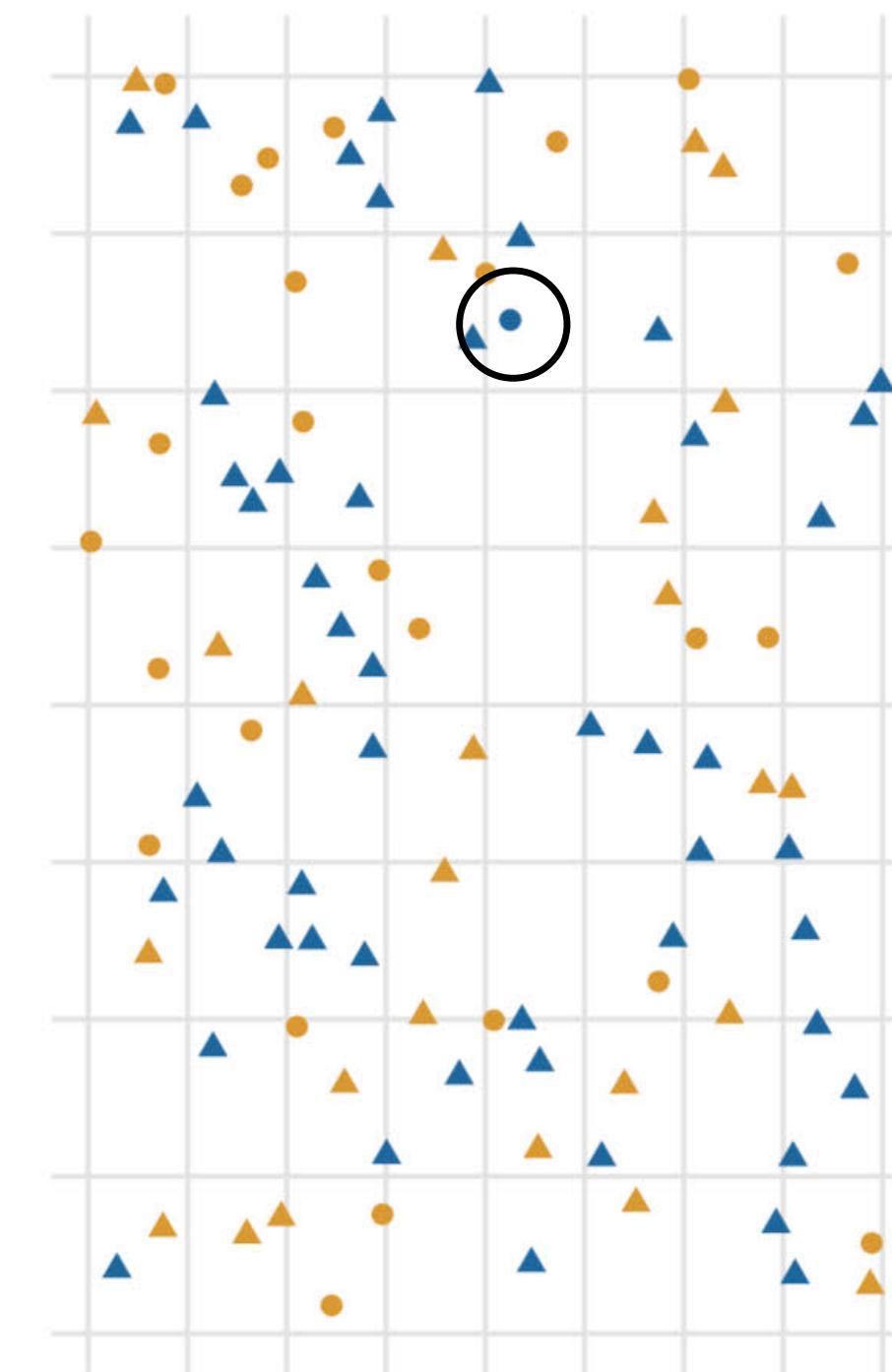
Shape only,  $N = 20$



Shape only,  $N = 100$



Color & shape,  $N = 100$



# References

**Anderson**, E. W., K. C. Potter, L. E. Matzen, J. F. Shepherd, G. A. Preston, and C. T. Silva. “A User Study of Visualization Effectiveness Using EEG and Cognitive Load.” *Computer Graphics Forum* 30, no. 3 (June 2011): 791–800.

**Armstrong**, Zan, and Nadieh Bremer. “Why Are so Many Babies Born around 8:00 A.M.?” *Scientific American* (blog), June 20, 2017. <https://blogs.scientificamerican.com/sa-visual/why-are-so-many-babies-born-around-8-00-a-m/>.

**Bayles**, David, and Ted Orland. *Art & Fear. Observations on the Perils (and Rewards) of Artmaking*. The Image Continuum, 1993.

**Bostock**, Michael. “Design Is a Search Problem.” presented at the OpenViz Conference, Boston, MA, 2014. <http://www.openvisconf.com/2014/>.

**Bostock**, Michael, Shan Carter, and Kevin Quealy. “Tracing the History of N.C.A.A. Conferences.” *New York Times*. November 30, 2013, Online edition, sec. Sports. <http://www.nytimes.com/newsgraphics/2013/11/30/football-conferences/index.html>.

**Bremer**, Nadieh. “Creating the Scientific American ‘Baby Spike’ Visual.” Portfolio. Visual Cinnamon (blog), October 31, 2017. <https://www.visualcinnamon.com/2017/10/creating-baby-births-visual>.

**Bremer**, Nadieh. “The Baby Spike.” Portfolio. Visual Cinnamon, April 2017. <https://www.visualcinnamon.com/portfolio/baby-spike>.

**Corum**, Jonathan. “See, Think, Design, Produce 3.” 13pt Information Design, March 28, 2016. <http://style.org/stdp3/>.

**Healy**, Kieran. *Data Visualization*. Princeton University Press, 2018.

**Knaflic**, Cole Nussbaumer. *Storytelling with Data. A Data Visualization Guide for Business Professionals*. Wiley, 2015.

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

**Miller**, Jane E., ed. *The Chicago Guide to Writing about Numbers*. Second edition. Chicago; London: The University of Chicago Press, 2015.

**Park**, Haeyoun, Jeremy Ashkenas, and Michael Bostock. “Taking the Battle to the States.” *New York Times*. January 11, 2014, Online edition, sec. Politics. <https://www.nytimes.com/interactive/2014/01/11/us/politics/who-controls-the-states-and-where-they-stand.html>.

**Rahlf**, Thomas. *Data Visualization with R — 111 Examples*. S.l.: Springer Nature, 2019.

**Tufte**, Edward R. *The Visual Display of Quantitative Information*. Second. Graphics Press, 2001.

**Ware**, Colin. *Information Visualization: Perception for Design*. Fourth. Philadelphia: Elsevier, Inc, 2020.

save