Storytelling with Data

Module 7: The storytelling process continued, and visual perception

Scott Spencer
Faculty and Lecturer
Columbia University

Agenda

Upcoming deliverable — *final* storyboard

Today’s objectives

Storytelling process continued

Group work on storyboards and project

Perception in visual narrative

Questions or suggestions?

Upcoming deliverable

Upcoming deliverable

In **storyboard** form — describe (1) your project, (2) preliminary results or insights so far, and (3) why those results are interesting for the marketing team. Use a distinct narrative arc (beginning, middle, and end), be clear and accessible for the **CMO**.

IDENTIFY PROBLEM WITHIN THE COMPANY

IDEATE A DATA ANALYTICS PROJECT ADDRESSING THE PROBLEM

WRITE MEMO PITCHING THE IDEA TO **CHIEF ANALYTICS OFFICER**

IDEA APPROVED. WRITE A PROJECT **PROPOSAL** TO **CAO**

THE PROPOSAL IS APPROVED

YOU ARE HERE

JUL 14 *FINAL*

STORYBOARD PROJECT & RESULTS FOR **CHIEF MARKETING OFFICER**

DEVELOP **INFOGRAPHIC** OF PROJECT & RESULTS FOR **CONSUMERS**

PRESENT PROJECT & RESULTS TO **CHIEF EXECUTIVE OFFICER** SHOWING VALUE OF ONGOING ANALYTICS

CONDUCT DATA ANALYSIS

Today's Objectives

Objectives

1. Explain importance of effectively framing a story
2. Group work on storyboards and projects
3. Consider audience perception of visual components of a story

The storytelling process, continued

See, Think, Design, Produce

Corum

A former student of Edward Tufte at Yale, Jonathan is science graphics editor at The New York Times and has won 28 awards from the Society for News Design and 18 medals from the international Malofiej competition, including Best of Show.

His projects for an audience of NYT readers …

13pt style.org

Corum, Jonathan. *See. Think. Design. Produce.* 28 Mar 2016. Web. http://style.org/stdp3/

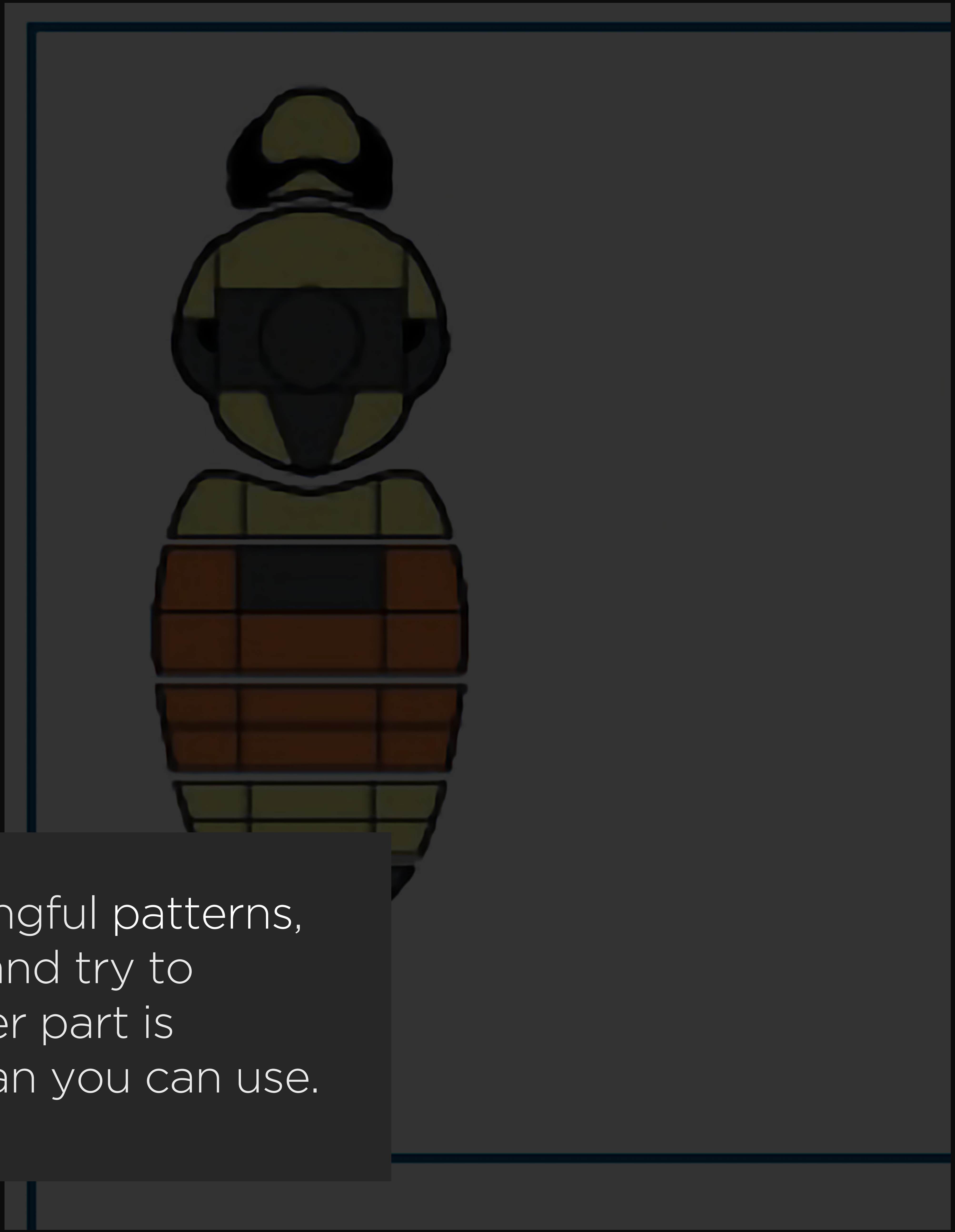
See, **Think**, **Design**, **Produce**

$\underbrace{\text{See, Think,}}_{\text{understand}}\ \underbrace{\text{Design, Produce}}_{\text{explain}}$

Corum, Jonathan. *See. Think. Design. Produce*. 28 Mar 2016. Web. http://style.org/stdp3/

See, Think, Design, Produce

$\underbrace{\text{See, Think,}}_{\text{understand}}\ \underbrace{\text{Design, Produce}}_{\text{explain}}$



Search for patterns by comparing

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

Corum, Jonathan. *See. Think. Design. Produce.* 28 Mar 2016. Web. http://style.org/stdp3/

Williams, Paul. *Bumble bees of North America: an identification guide.*2014. Princeton University Press.

See, Think, Design, Produce

understand — explain

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.

Corum, Jonathan. *See. Think. Design. Produce.* 28 Mar 2016. Web. http://style.org/stdp3/

See, Think, **Design**, Produce

understand

explain

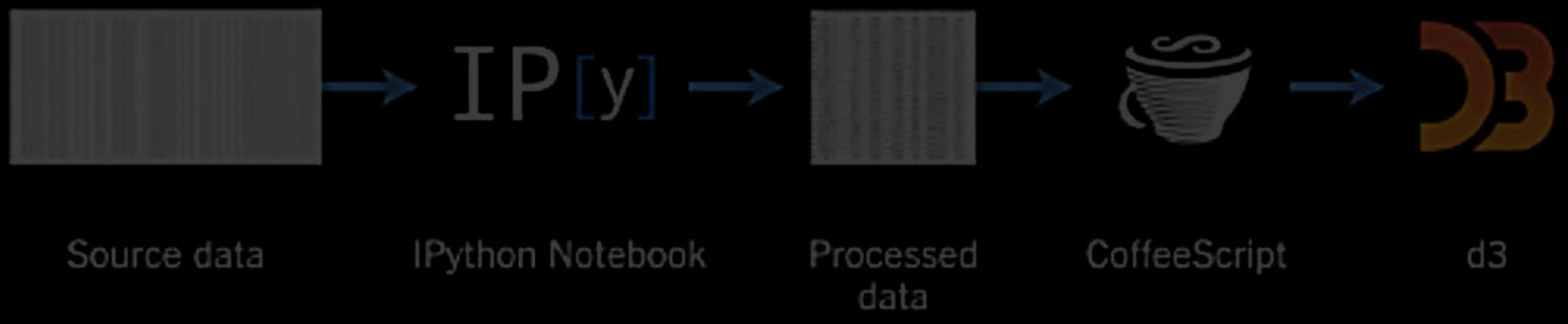
Design for someone else, show varying details

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**.

Corum, Jonathan. *See. Think. Design. Produce.* 28 Mar 2016. Web. http://style.org/stdp3/
Fairfield, Hannah. *Driving Safety, in Fits and Starts.*17 Sept 2012. Web. https://archive.nytimes.com/www.nytimes.com/interactive/2012/09/17/science/driving-safety-in-fits-and-starts.html

See, Think, Design, Produce

$\underbrace{\text{See, Think,}}_{\text{understand}}\ \underbrace{\text{Design, Produce}}_{\text{explain}}$



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**.

Corum, Jonathan. *See. Think. Design. Produce.* 28 Mar 2016. Web. http://style.org/stdp3/

Questions for discussion.

What did you find interesting or helpful in the way Jonathan described his process of visual storytelling?

Which of his examples were most vivid, most memorable, to you? What made them so?

Group work on storyboards and projects

Help your colleagues

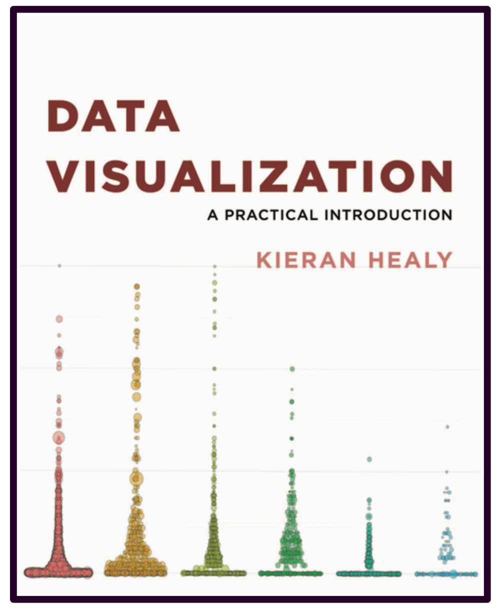
Pair up, take turns providing feedback. Look for:

narrative and messaging

consideration of audience

coherence in design, combining words and graphics

Audience (mis)perception of data graphics



Data Visualization: a practical introduction

Healy

A PhD graduate from Princeton, Kieran is associate professor of sociology at Duke University. His book has been described as "covering the 'why do' as well as the 'how to' of data visualization." — Andrew Gelman

Issues when visually encoding data

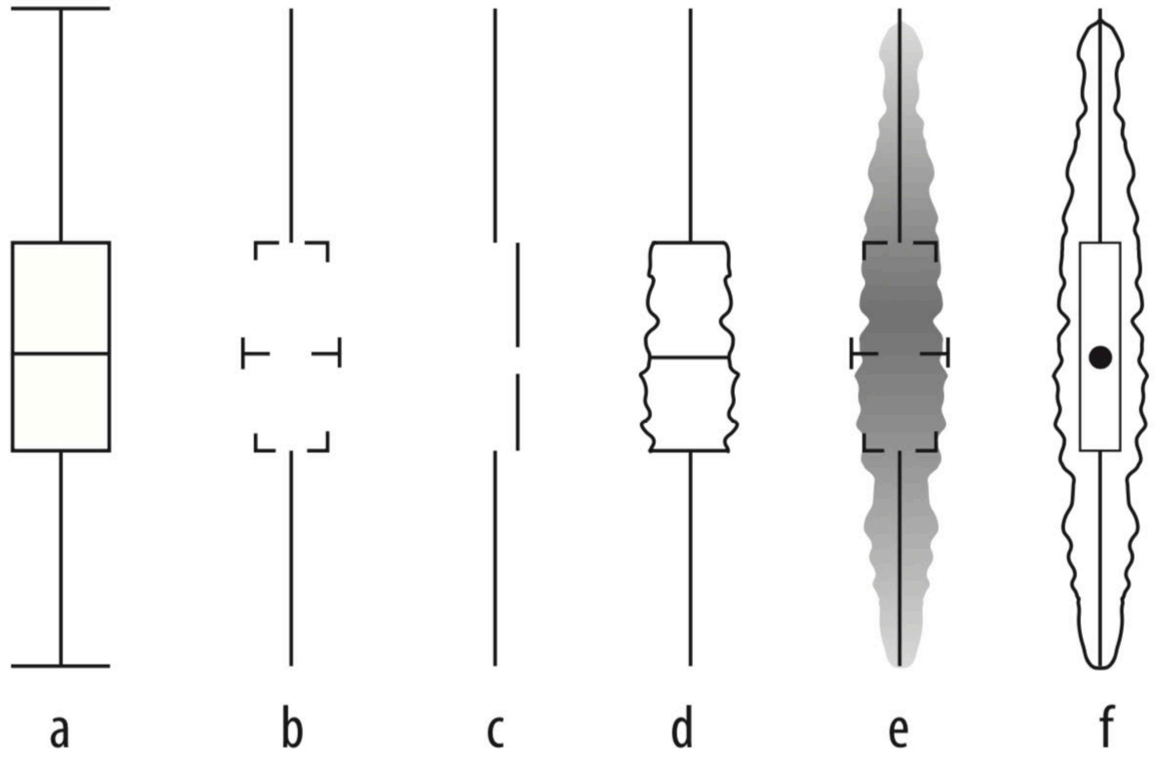
- **Aesthetic**
- Substantive
- Perceptual

Aesthetic

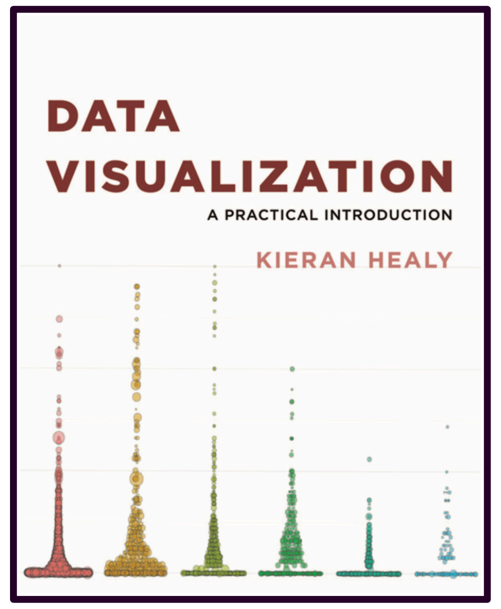
Consider whether every mark, color, and luminance on a statistical graph conveys information and supports messaging.

Memorable, but hard to read. "Junk chart"

Consider the audience when erasing non- or redundant-data ink



Healy, Kieran. *Data Visualization*. Princeton University Press, 2019. Print.



Data Visualization: a practical introduction

Healy

A PhD graduate from Princeton, Kieran is associate professor of sociology at Duke University. His book has been described as “covering the ‘why do’ as well as the ‘how to’ of data visualization.” — Andrew Gelman

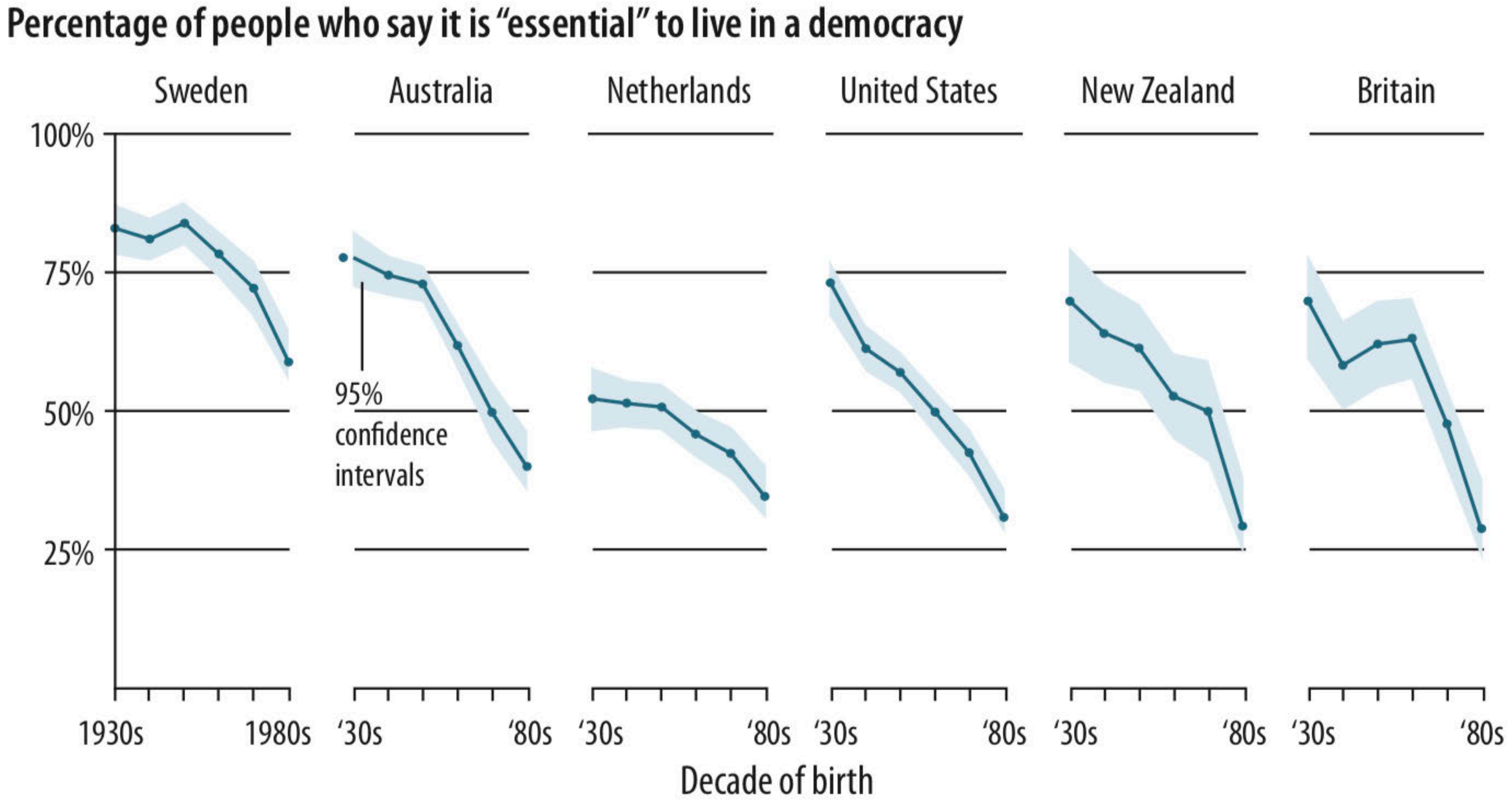
Issues when visually encoding data

Aesthetic
Substantive
Perceptual

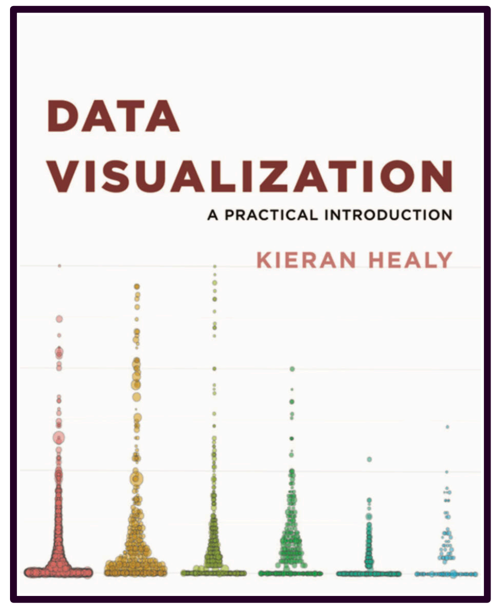
Substantive

Consider whether your data honestly and fairly represent your message.

The original chart below graphed the relative change in response of respondents who selected 10 on a 10 point scale.



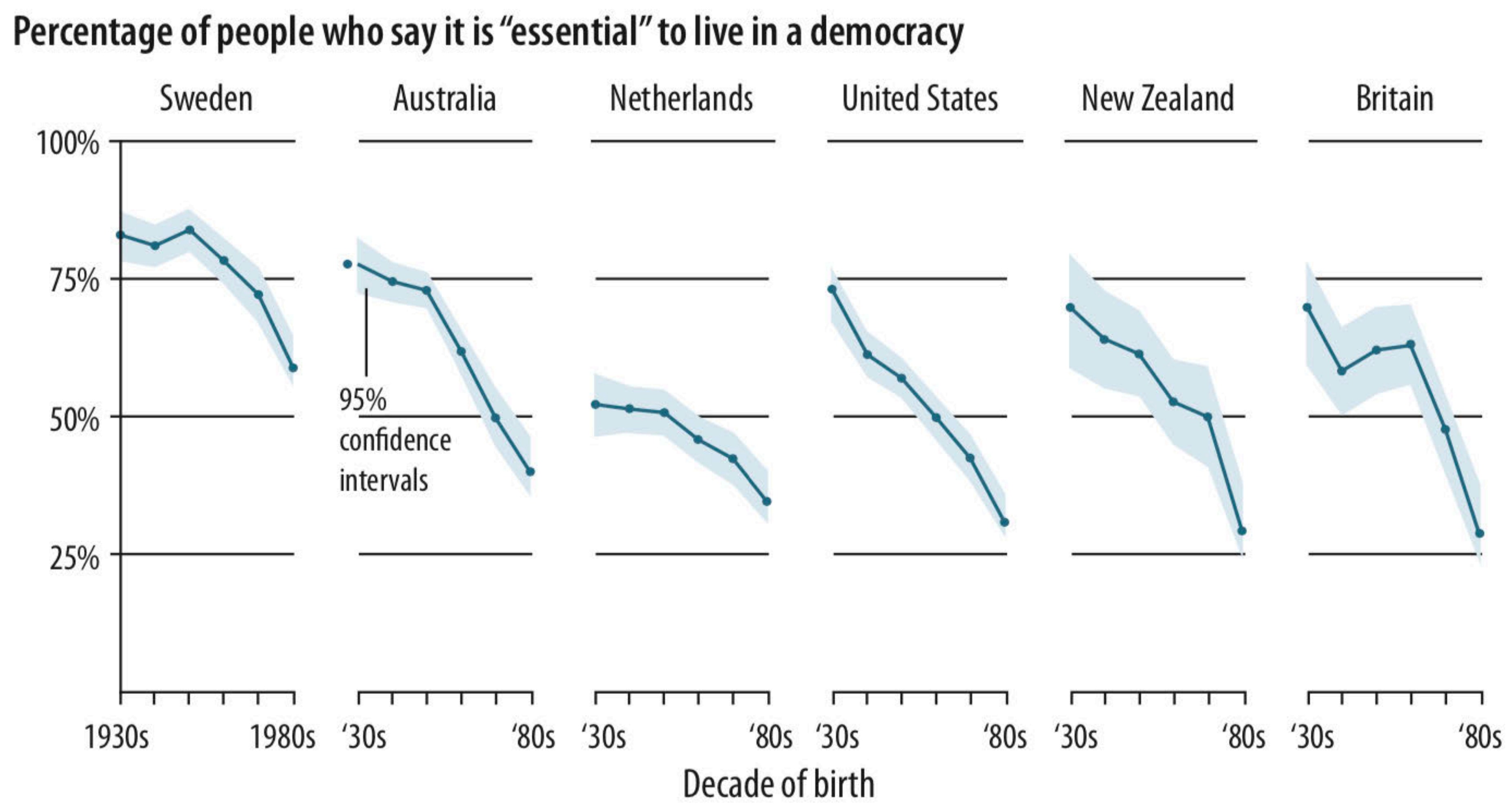
Healy, Kieran. *Data Visualization*. Princeton University Press, 2019. Print.

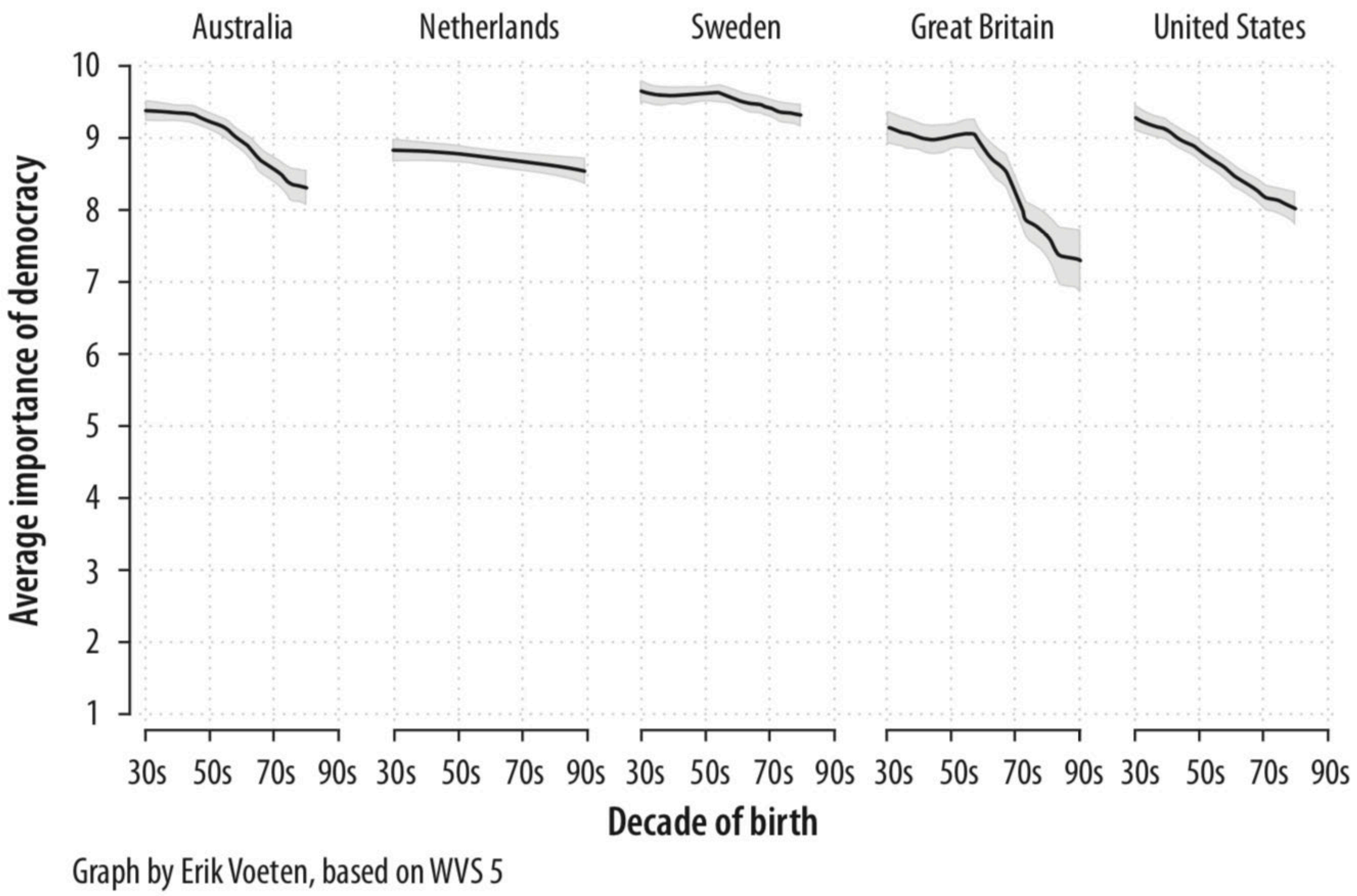


Data Visualization: a practical introduction

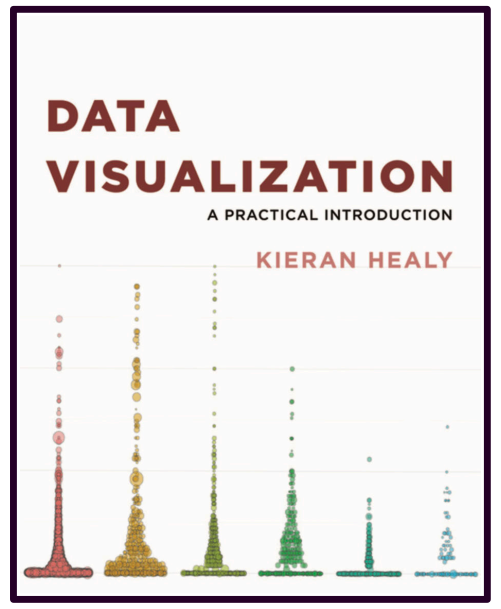
Healy

A PhD graduate from Princeton, Kieran is associate professor of sociology at Duke University. His book has been described as "covering the 'why do' as well as the 'how to' of data visualization." — Andrew Gelman





Healy, Kieran. *Data Visualization*. Princeton University Press, 2019. Print.



Data Visualization: a practical introduction

Healy

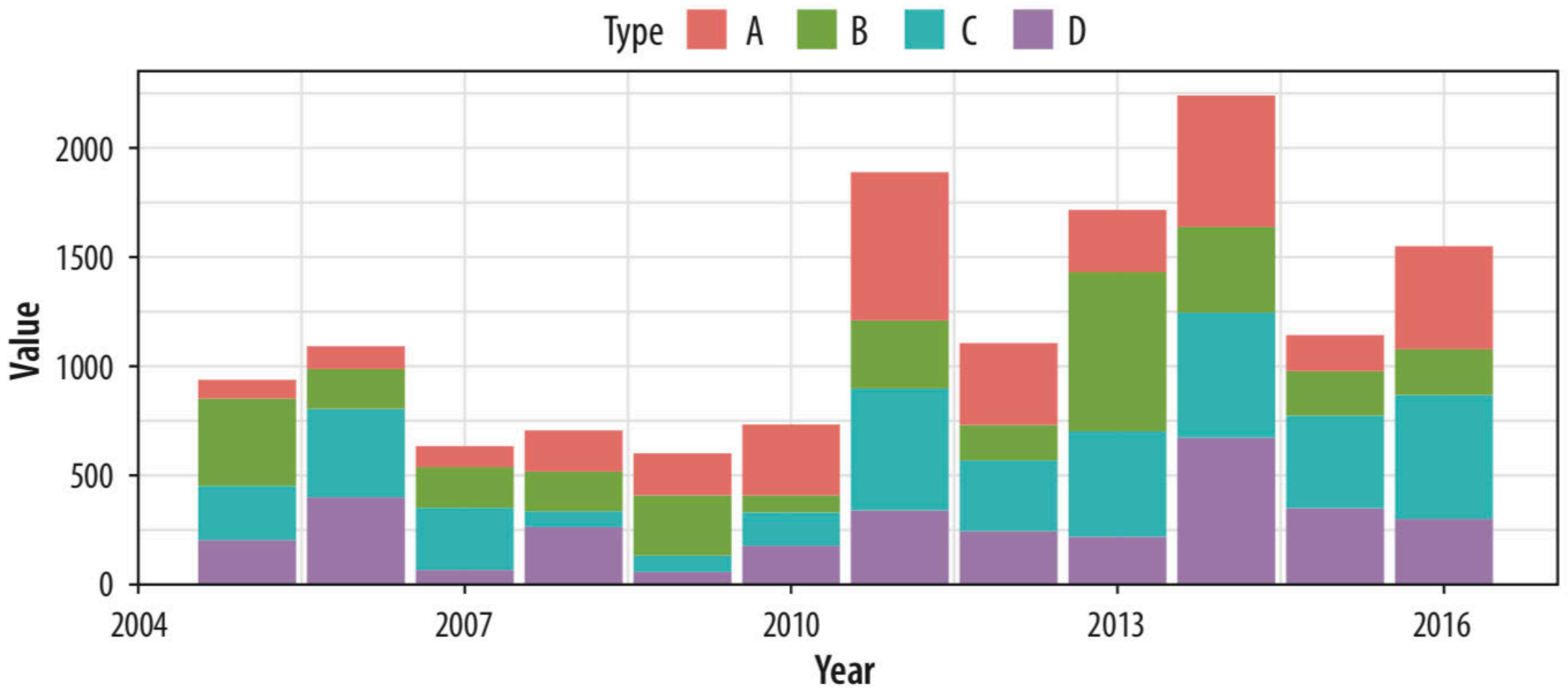
A PhD graduate from Princeton, Kieran is associate professor of sociology at Duke University. His book has been described as "covering the 'why do' as well as the 'how to' of data visualization." — Andrew Gelman

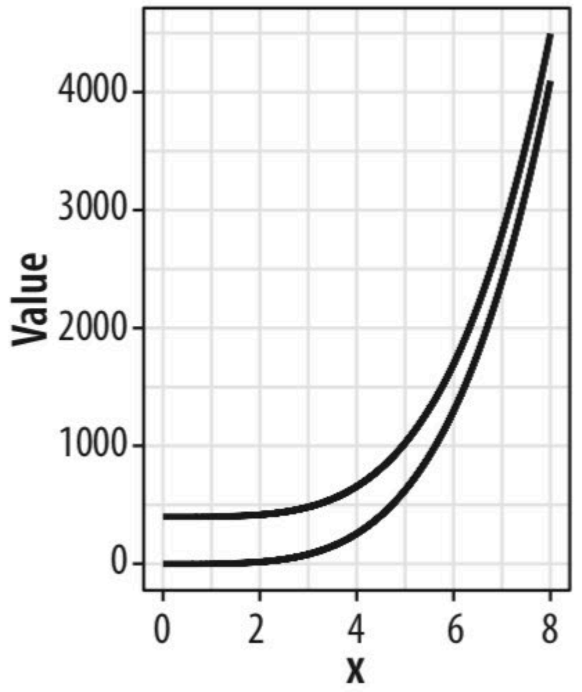
Issues when visually encoding data

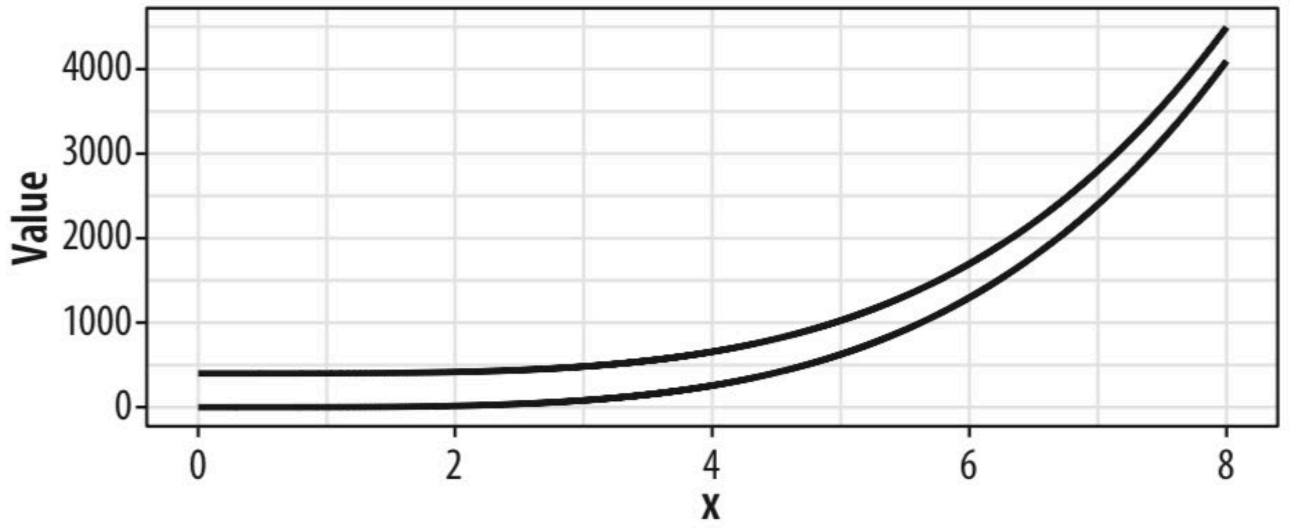
Aesthetic
Substantive
Perceptual

Perceptual

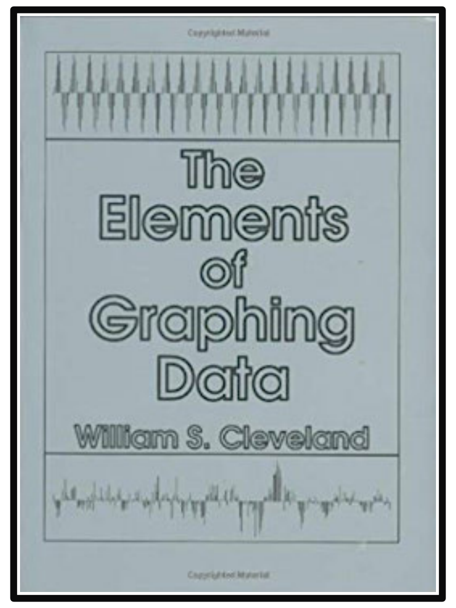
Even with a reasonably-high data-ink ratio, you must choose an encoding that most naturally guides the audience to understand and compare the data. The graphs below have perceptual issues.







Healy, Kieran. *Data Visualization*. Princeton University Press, 2019. Print.



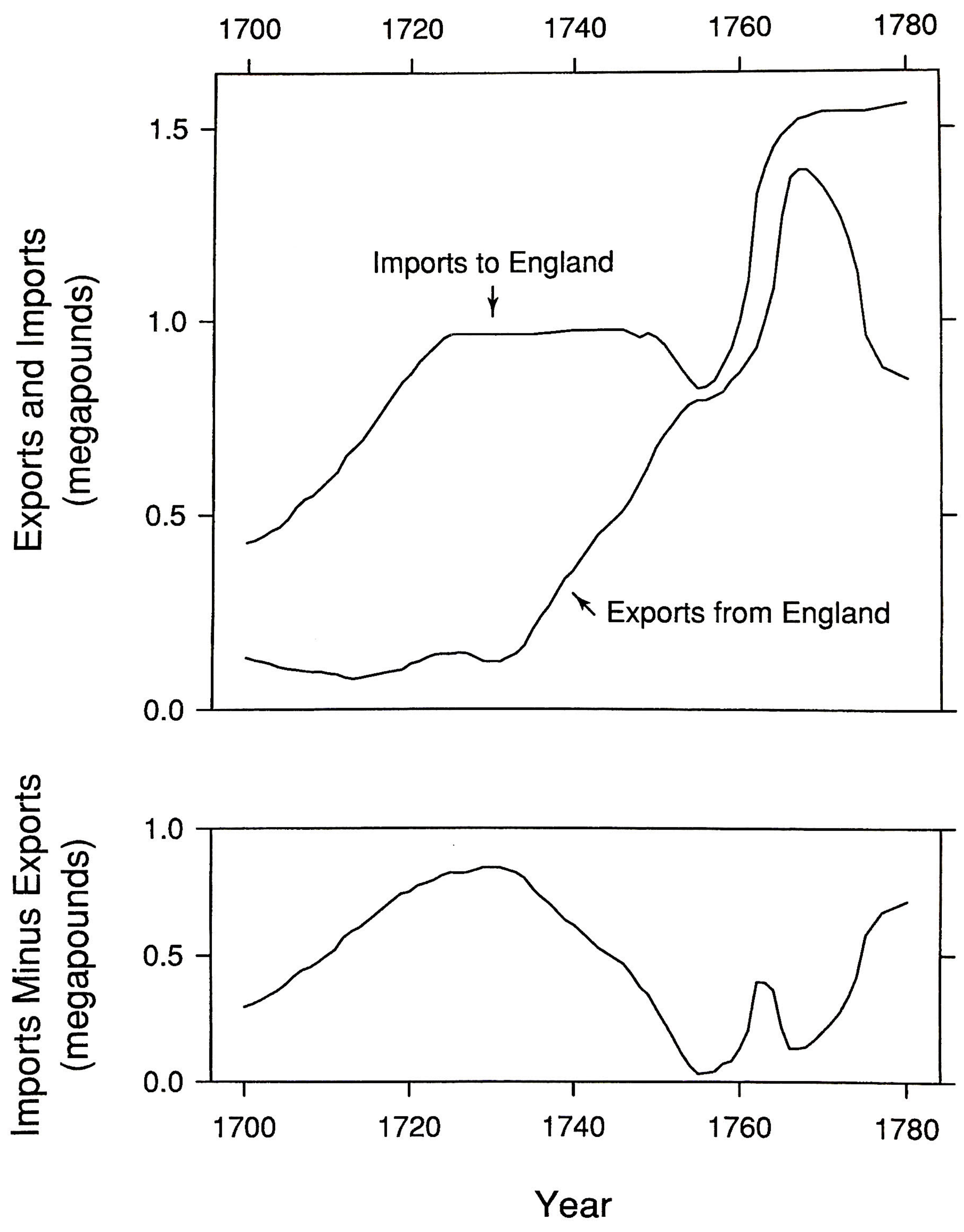
The Elements of Graphing Data

Cleveland

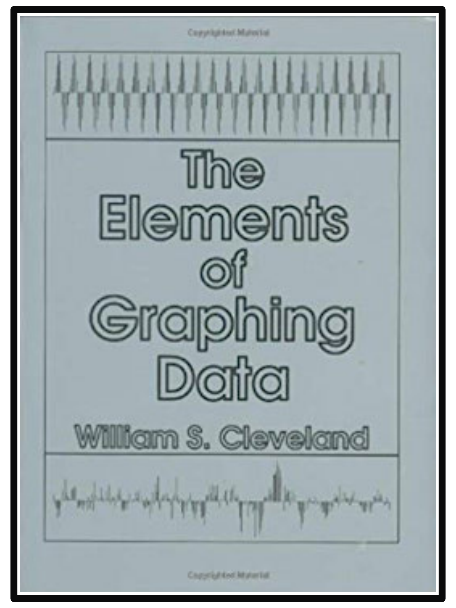
A graduate of Princeton, William is a computer scientist and Professor of Statistics and Professor of Computer Science at Purdue University, known for his work on data visualization, particularly on nonparametric regression and local regression.

Superposed curves have a decoding problem

Decoding differences between two lines or curves on a graph can be inaccurate because we naturally compare the shortest distance between the lines instead of the vertical distance between the lines.



Cleveland, William S. *The Elements of Graphing Data*. Wadsworth, 1985. Print.



The Elements of Graphing Data

Cleveland

A graduate of Princeton, William is a computer scientist and Professor of Statistics and Professor of Computer Science at Purdue University, known for his work on data visualization, particularly on nonparametric regression and local regression.

We need a fixed percentage change in something to detect a difference

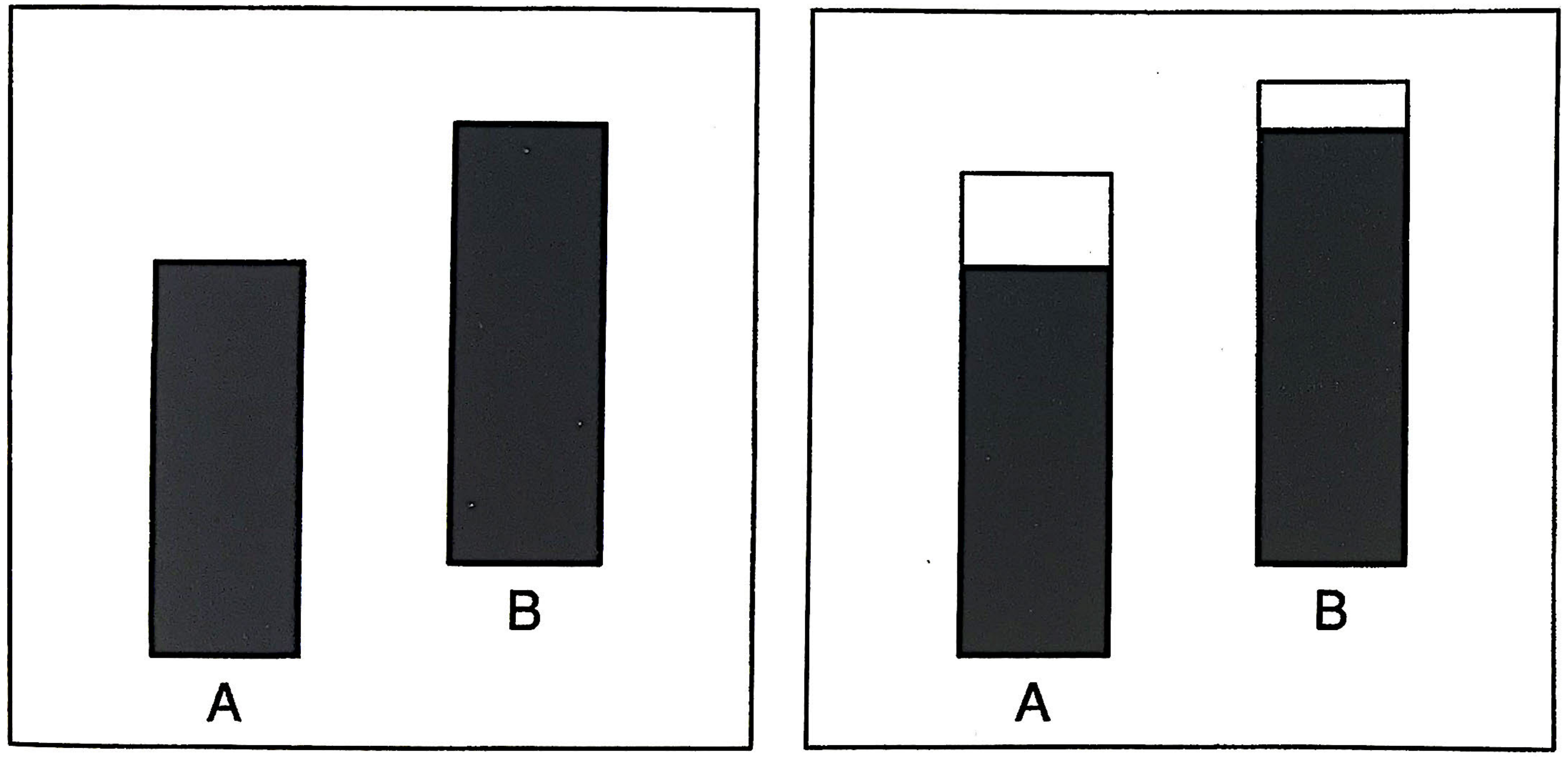
Our visual perception of differences can be stated mathematically, as Weber's Law.

If x is the magnitude of a physical attribute, say, length of a line segment, and $w_p(x)$ is a positive number such that a line of length

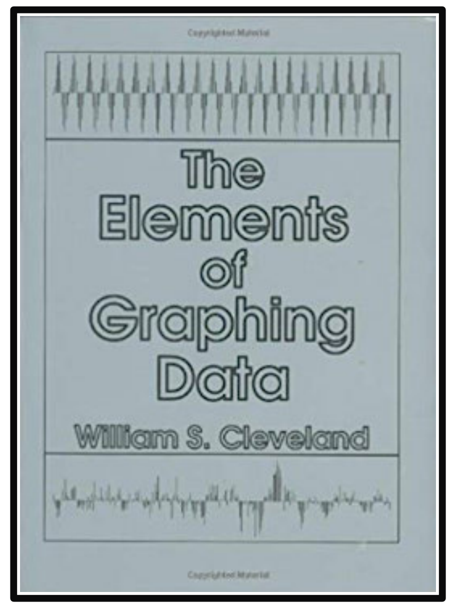
$$x + w_p(x)$$

is discriminated with probability p to be longer than the line of length x then,

$$w_p(x) = k_p x$$



Cleveland, William S. *The Elements of Graphing Data*. Wadsworth, 1985. Print.



The Elements of Graphing Data

Cleveland

A graduate of Princeton, William is a computer scientist and Professor of Statistics and Professor of Computer Science at Purdue University, known for his work on data visualization, particularly on nonparametric regression and local regression.

Consider (only) using reference grids in aid of Weber's Law

Reference grids can help us perceive the differences between two measurements, but they can also distract. Consider only using reference lines that guide your audience to the relevant point of your message.

And a little goes a long way: for any needed reference lines, consider making them, say, very faint gray or otherwise distinguish them from the data.

How did we do in the example proposal for the Dodgers?

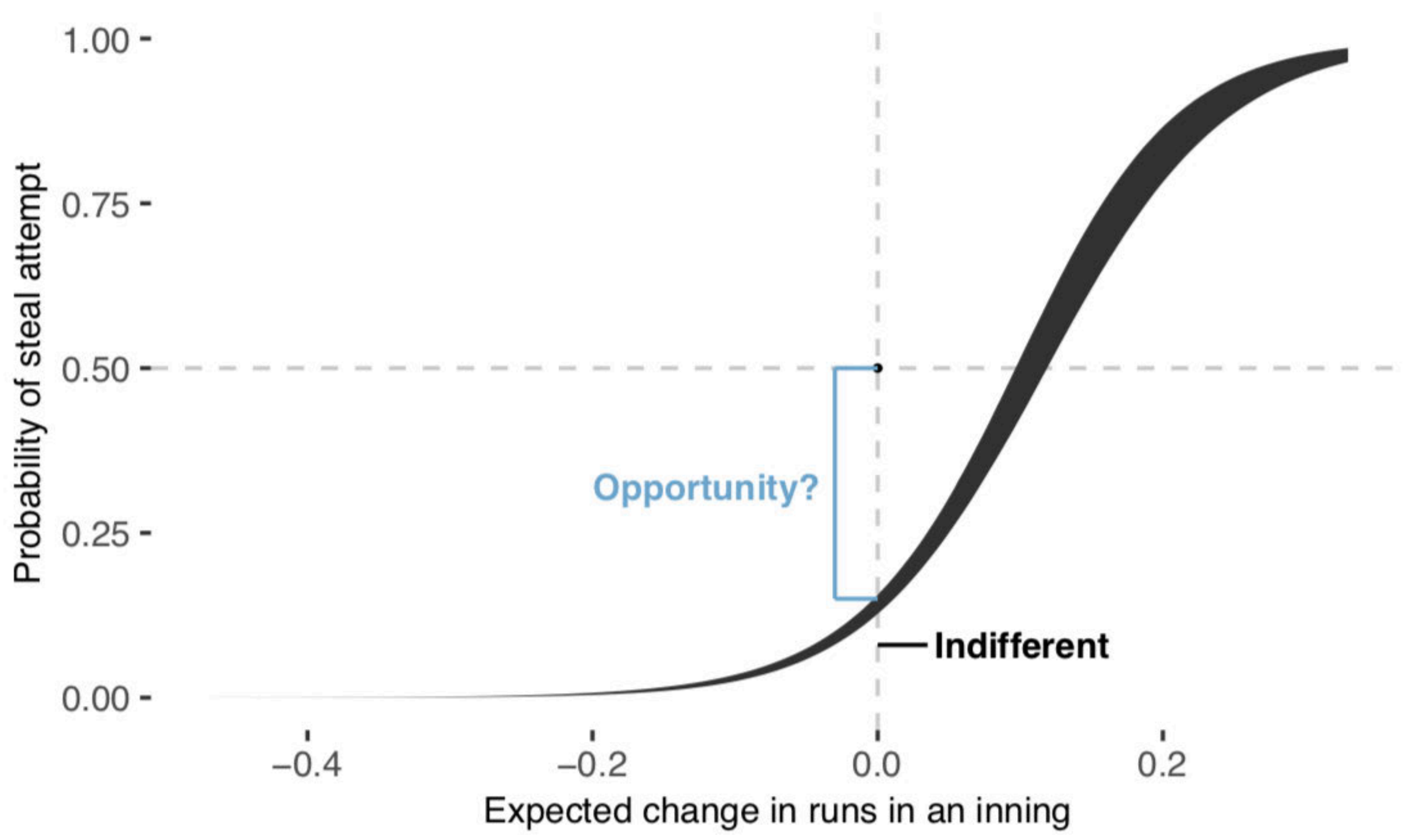
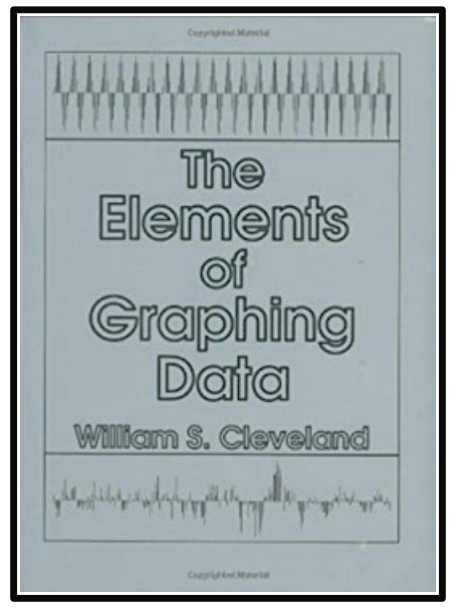


Figure 2. When the change in expected runs is zero, managers should be indifferent to attempted steals, saying go half the time.

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.

Cleveland, William S. *The Elements of Graphing Data*. Wadsworth, 1985. Print.



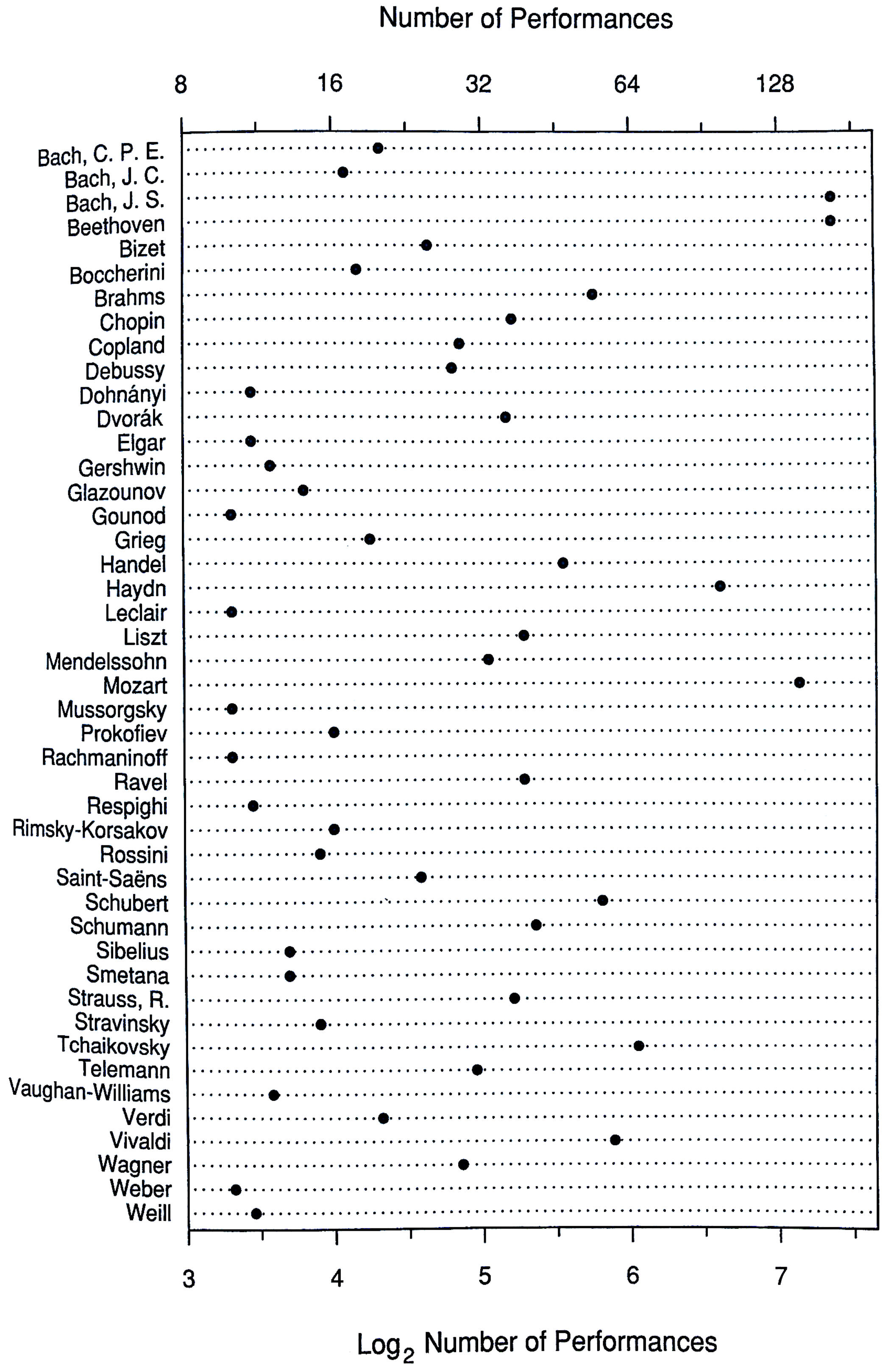
The Elements of Graphing Data

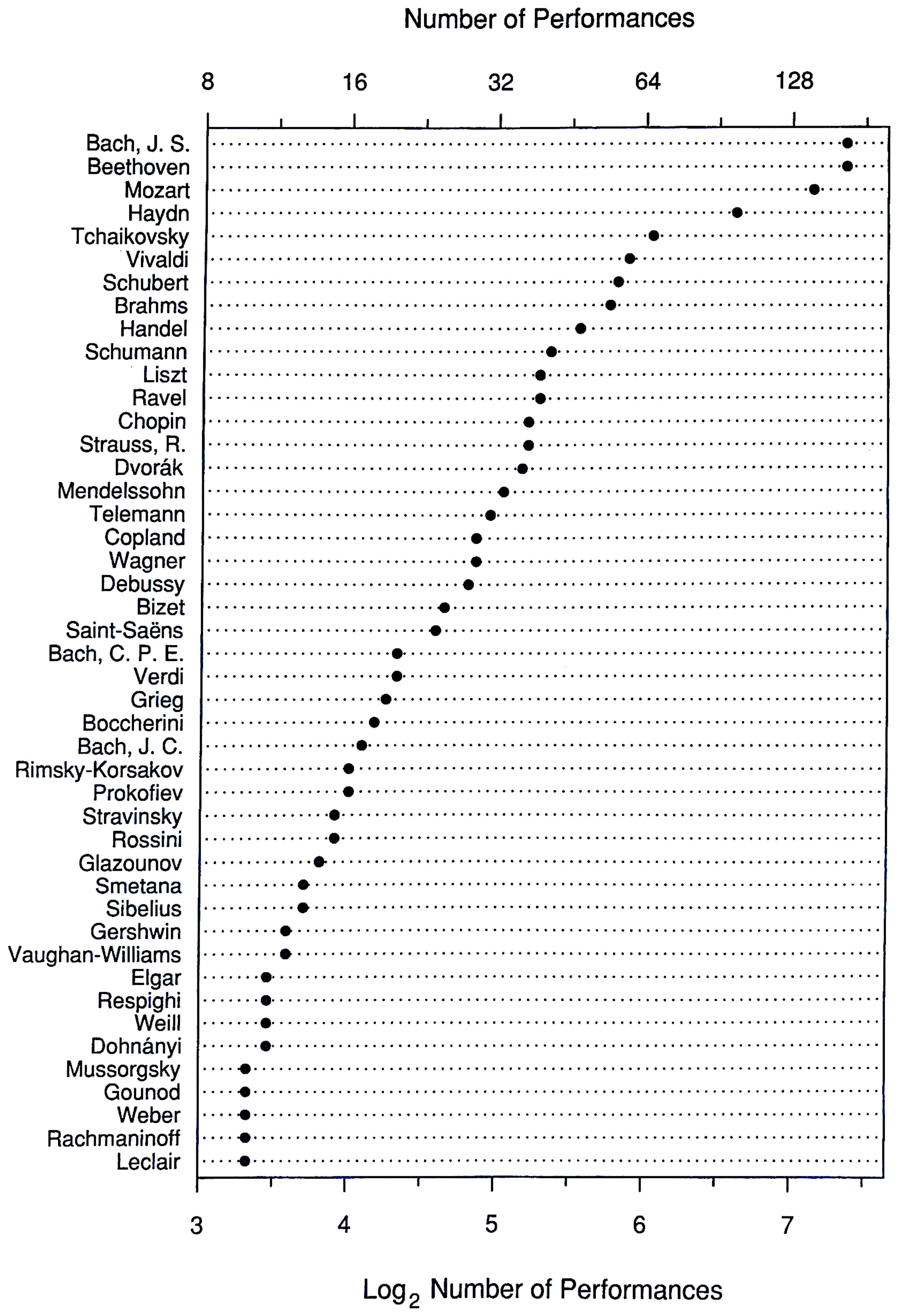
Cleveland

A graduate of Princeton, William is a computer scientist and Professor of Statistics and Professor of Computer Science at Purdue University, known for his work on data visualization, particularly on nonparametric regression and local regression.

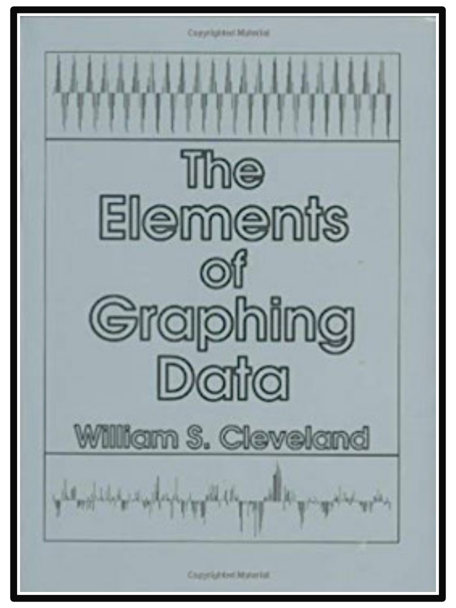
Ordering for categorical variables substantially affects our visual decoding

Ordering from, say, smallest to largest enhances our visual decoding of the distribution of values along the measurement scale.





Cleveland, William S. *The Elements of Graphing Data*. Wadsworth, 1985. Print.



The Elements of Graphing Data

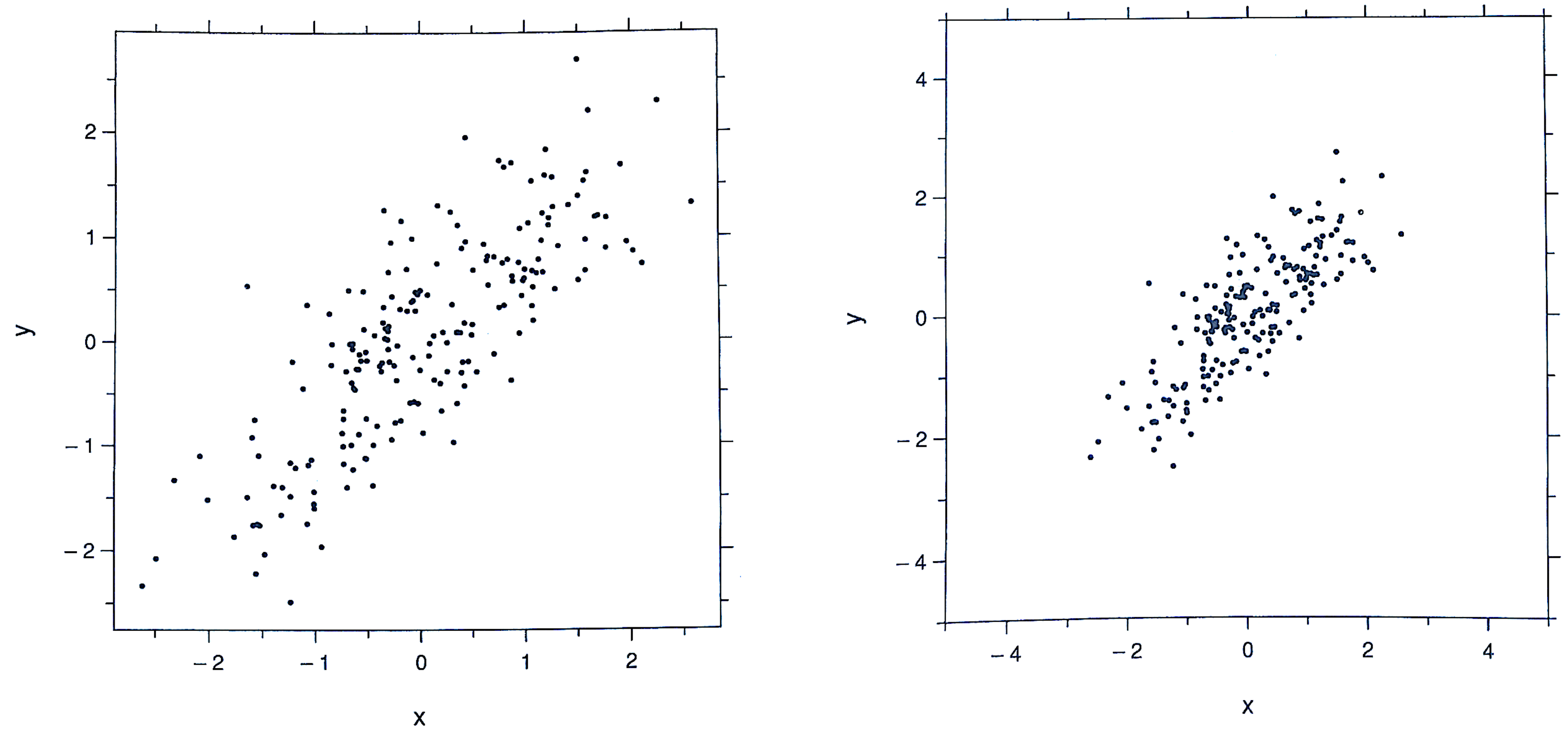
Cleveland

A graduate of Princeton, William is a computer scientist and Professor of Statistics and Professor of Computer Science at Purdue University, known for his work on data visualization, particularly on nonparametric regression and local regression.

Perceived correlation depends on ratio of plot area to data area

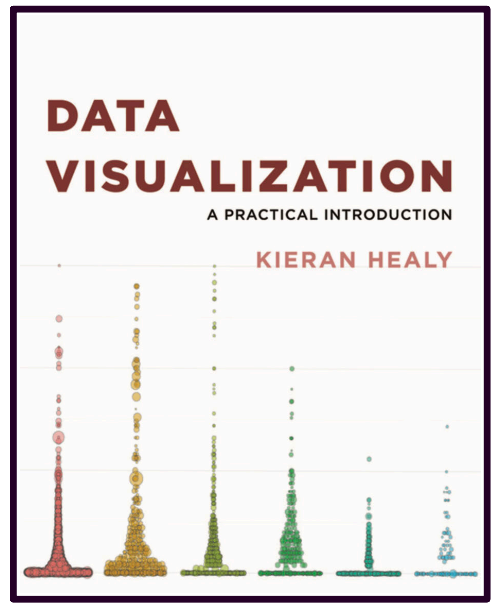
Our estimation of correlation is affected by the area of the data rectangle divided by the area of the scale-line rectangle.

Showing the same data, the left panel displays a 1 to 1 ratio, while in the right panel displays the data rectangle as much smaller than the scale-line rectangle.



Cleveland, William S. *The Elements of Graphing Data*. Wadsworth, 1985. Print.

(In)accuracy in decoding visuals

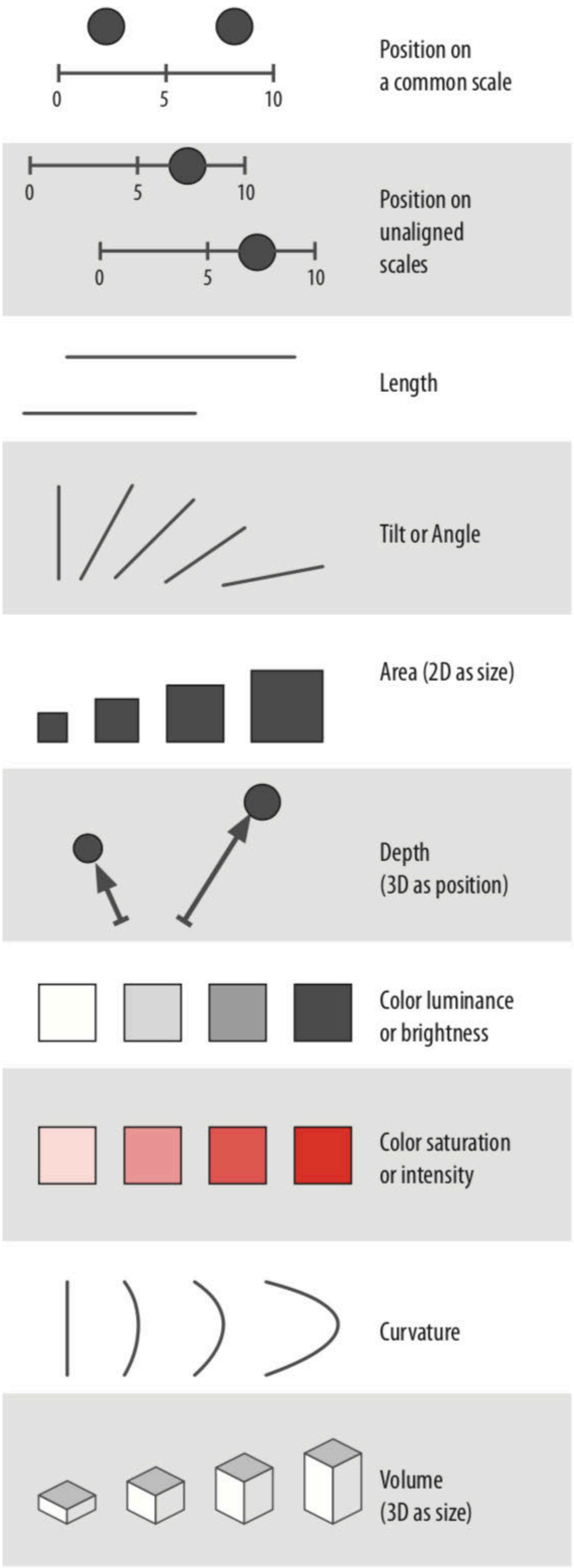


Data Visualization: a practical introduction

Healy

A PhD graduate from Princeton, Kieran is associate professor of sociology at Duke University. His book has been described as "covering the 'why do' as well as the 'how to' of data visualization." — Andrew Gelman

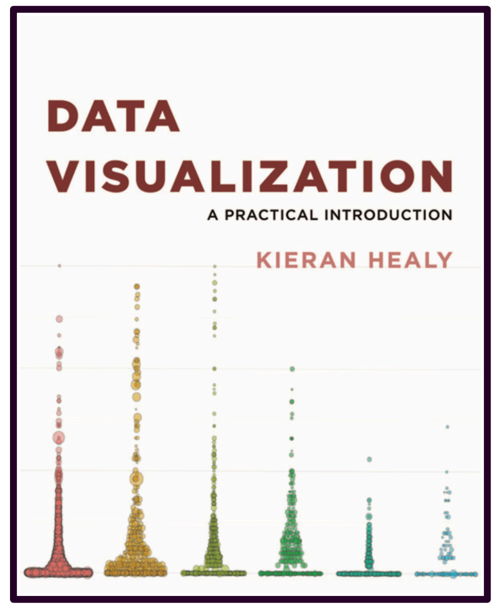
Most accurate



Least accurate

Healy, Kieran. *Data Visualization*. Princeton University Press, 2019. Print.

Focusing visual attention



Data Visualization: a practical introduction

Healy

A PhD graduate from Princeton, Kieran is associate professor of sociology at Duke University. His book has been described as "covering the 'why do' as well as the 'how to' of data visualization." — Andrew Gelman

Help your audience with Gestault principles

Our eyes automatically search for (grouping), difference and change.

Proximity: Things that are spatially near to one another seem to be related.

Similarity: Things that look alike seem to be related.

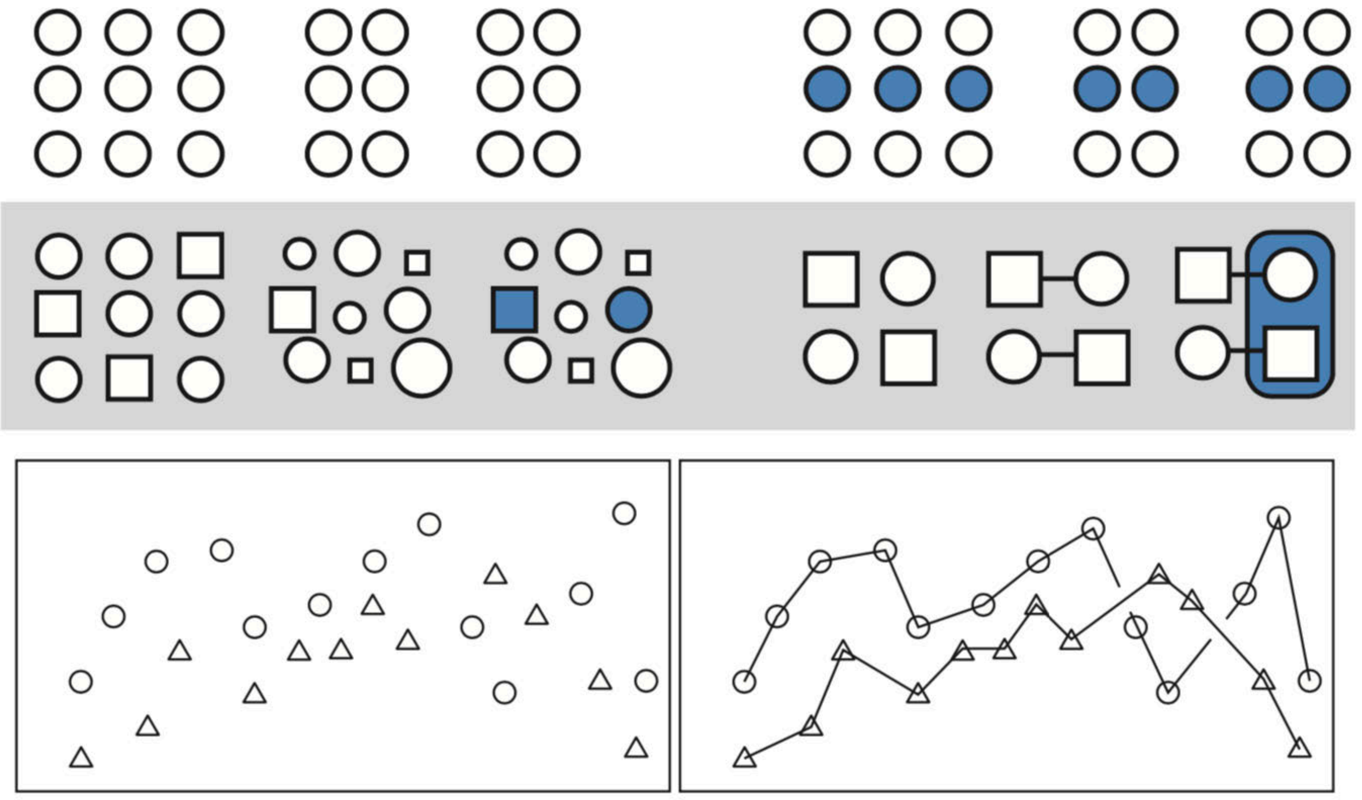
Connection: Things that are visually tied to one another seem to be related.

Continuity: Partially hidden objects are completed into familiar shapes.

Closure: Incomplete shapes are perceived as complete.

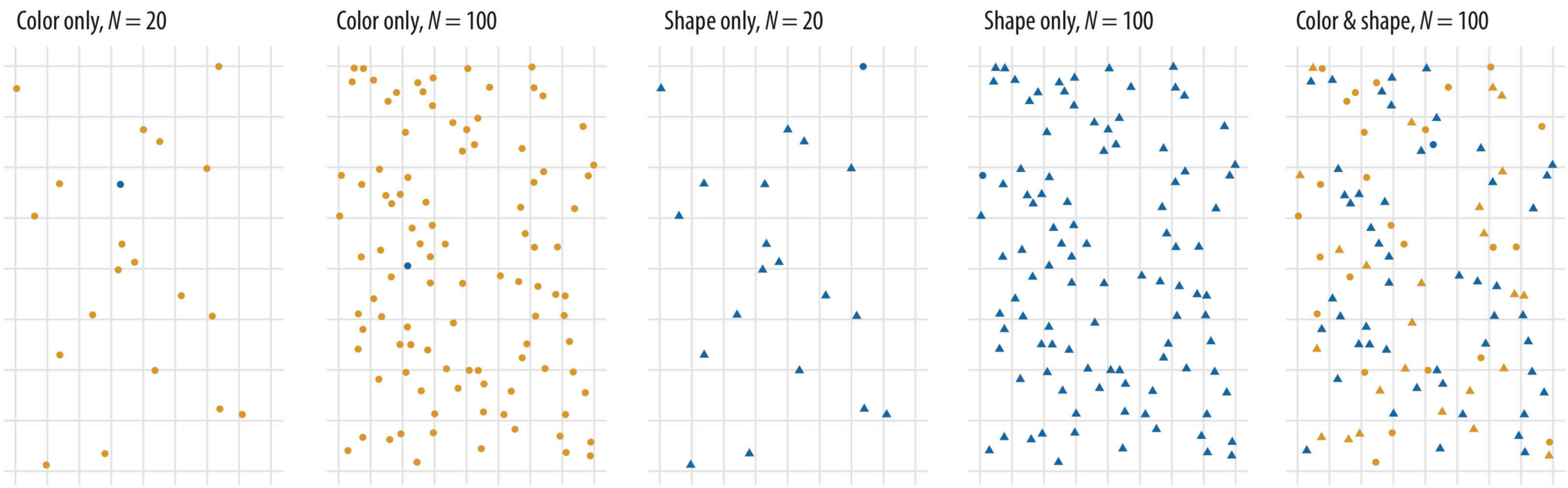
Figure and ground: Visual elements are taken to be either in the foreground or in the background.

Common fate: Elements sharing a direction of movement are perceived as a unit.



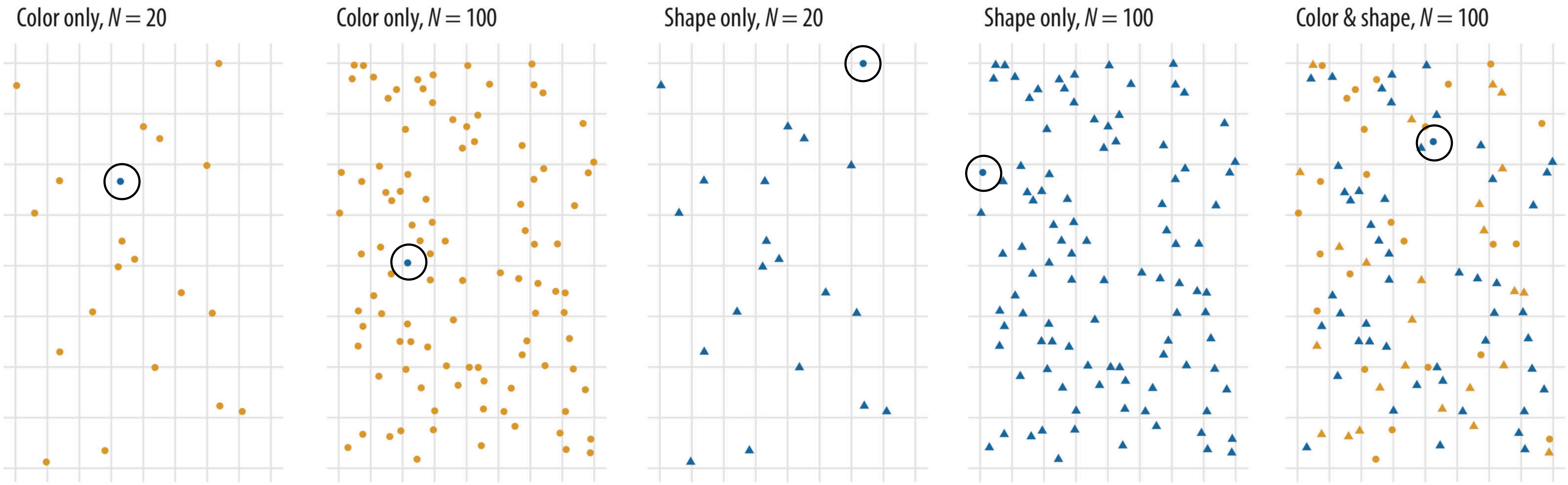
Healy, Kieran. *Data Visualization*. Princeton University Press, 2019. Print.

Consider Gestault principles when trying—as an audience—to find the blue circle.



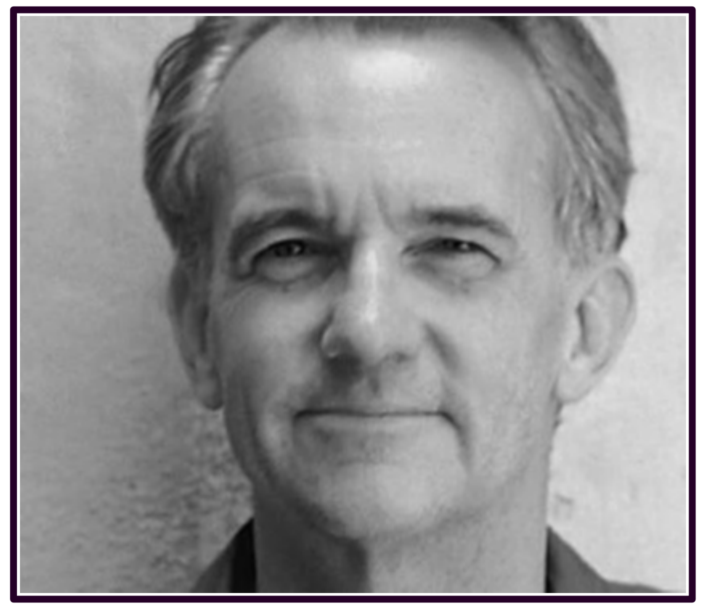
Healy, Kieran. *Data Visualization*. Princeton University Press, 2019. Print.

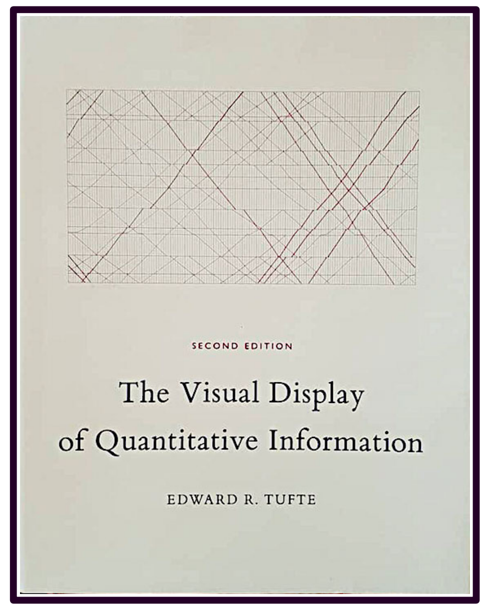
For these, another choice (enclosure) is more effective, which draw the eye instantly.



Healy, Kieran. *Data Visualization*. Princeton University Press, 2019. Print.

From default visual to visual narrative





The Visual Display of Quantitative Information

Tufte

Hailed "The Leonardo da Vinci of data" by the New York Times. He is professor emeritus of Political Science, Statistics, and Computer Science at Yale University.

Simplicity of design, complexity of data

Graphical excellence is often found in simplicity of design and complexity of data.

Words and pictures belong together

Viewers need the help that words can provide. **Words on graphics are data-ink**, making effective use of the space freed up by erasing redundant and non-data-ink.

Note, the **size of type** on and around graphics can be quite **small**, since the phrases and sentences are usually not too long.

The principle of data/text integration is: data **graphics are paragraphs about data** and should be treated as such.

Proportion and scale: the shape of graphics

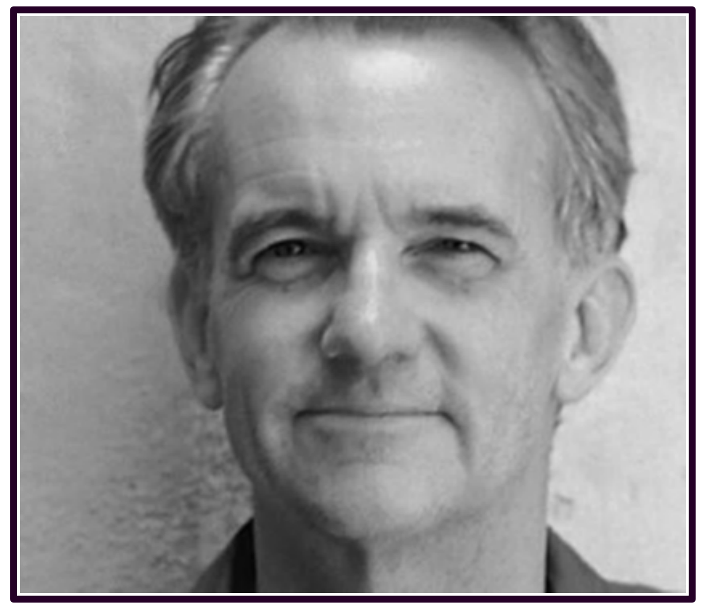
Our eye is naturally practiced in detecting **deviations from the horizon**.

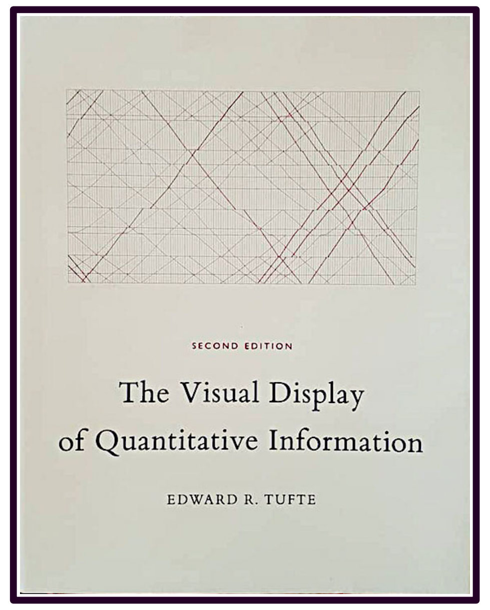
Horizontally shaped plots tend to make it **easier to directly label** and explain the data.

Tradition places the effect vertically, the cause horizontally.

The empirically studied **Golden Rectangle**, a 1.0 x 1.618 ratio is aesthetically pleasing.

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





The Visual Display of Quantitative Information

Tufte

Hailed "The Leonardo da Vinci of data" by the New York Times. He is professor emeritus of Political Science, Statistics, and Computer Science at Yale University.

Data-ink | The non-erasable core of a graphic, the non-redundant ink arranged in response to variation in the numbers represented. Annotations are part of the data-ink.

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

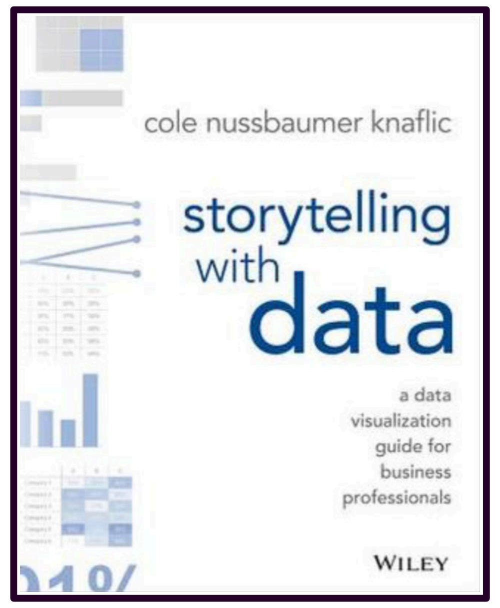
= **proportion of a graphic's ink devoted to the non-redundant display of data-information**

= **1.0 – proportion of a graphic that can be erased without loss of data-information**

Maximize within reason | Maximize the data-ink ratio, within reason. Erase non-data-ink, within reason. Erase redundant data-ink, within reason.

Size graphics for legibility | As the quantity of data increases, data measures must shrink. . . . The way to increase data density other than by enlarging the data matrix is to reduce the area of a graphic. **Graphics can be shrunk way down**.

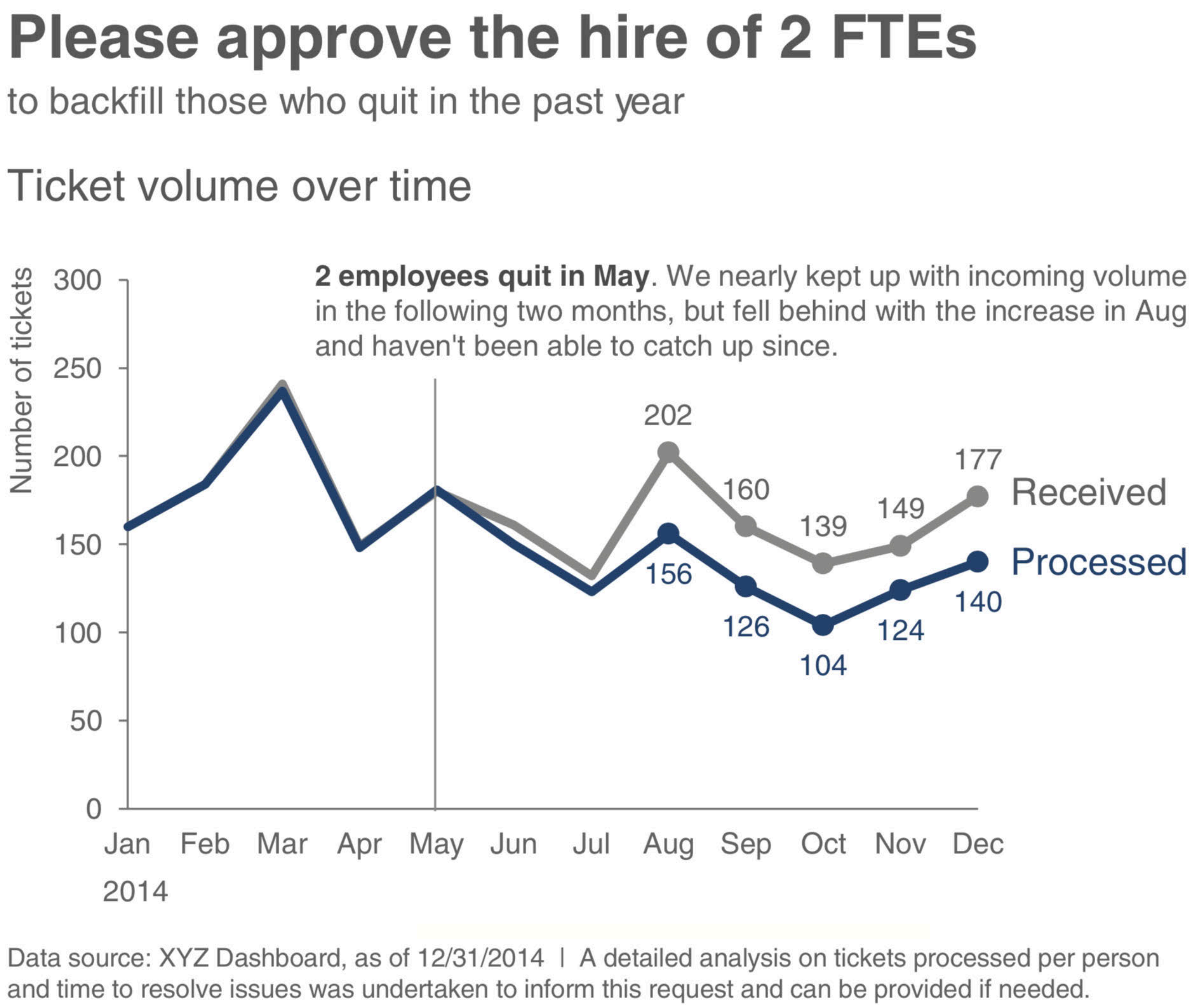
Tufte, Edward R. *The Visual Display of Quantitative Information*. Second. Graphics Press, 2001. Print.



Storytelling with data

Knaflic

The author is a consultant focused on visual displays. Her experience arose from human resources in Google where she applied theory learned as a student of Yale's Edward Tufte.



Knaflic, Cole Nussbaumer. *Storytelling with Data*. Wiley, 2015. Print.

Discussion: what differences do you see? What advice has she applied?

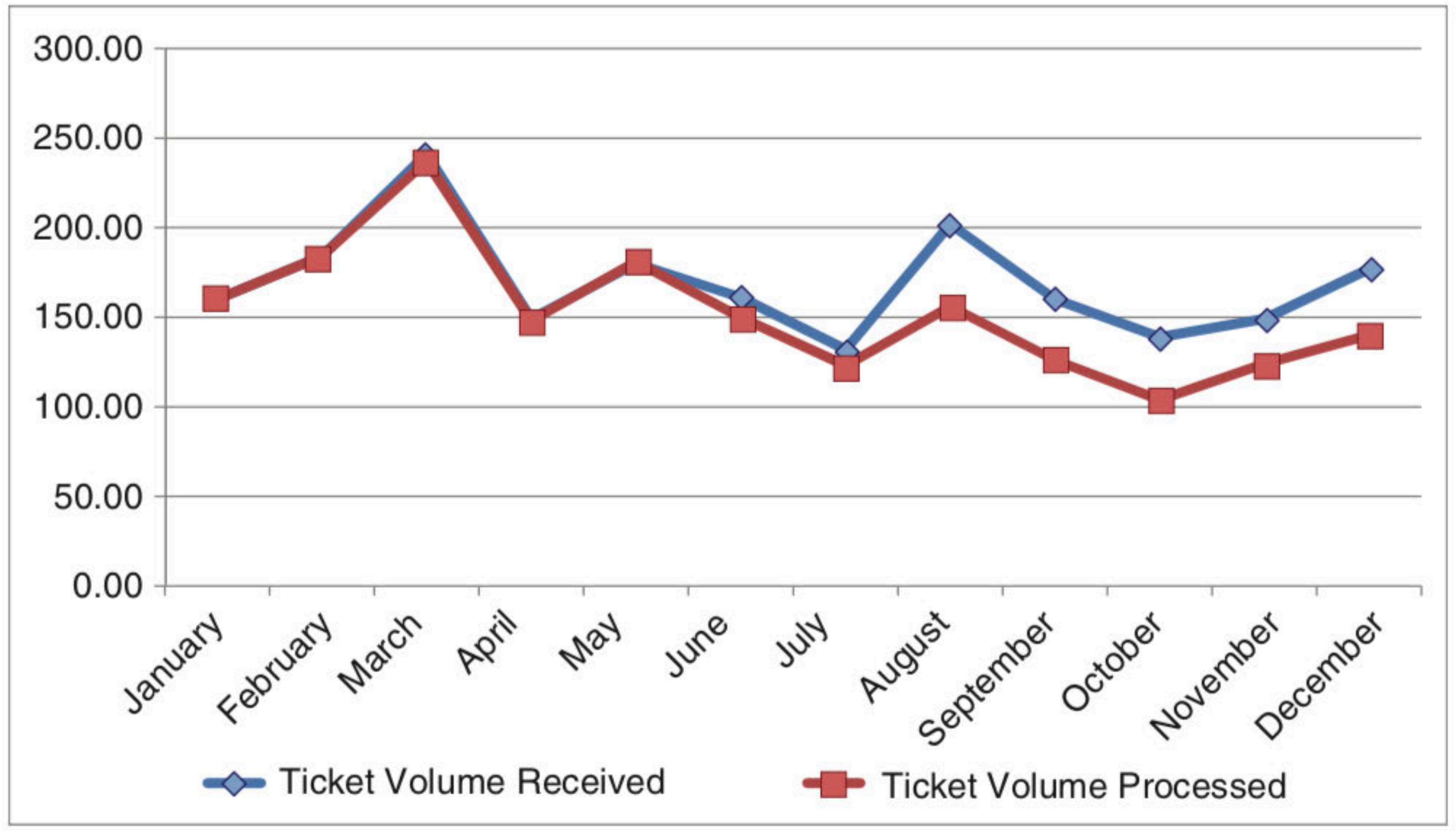
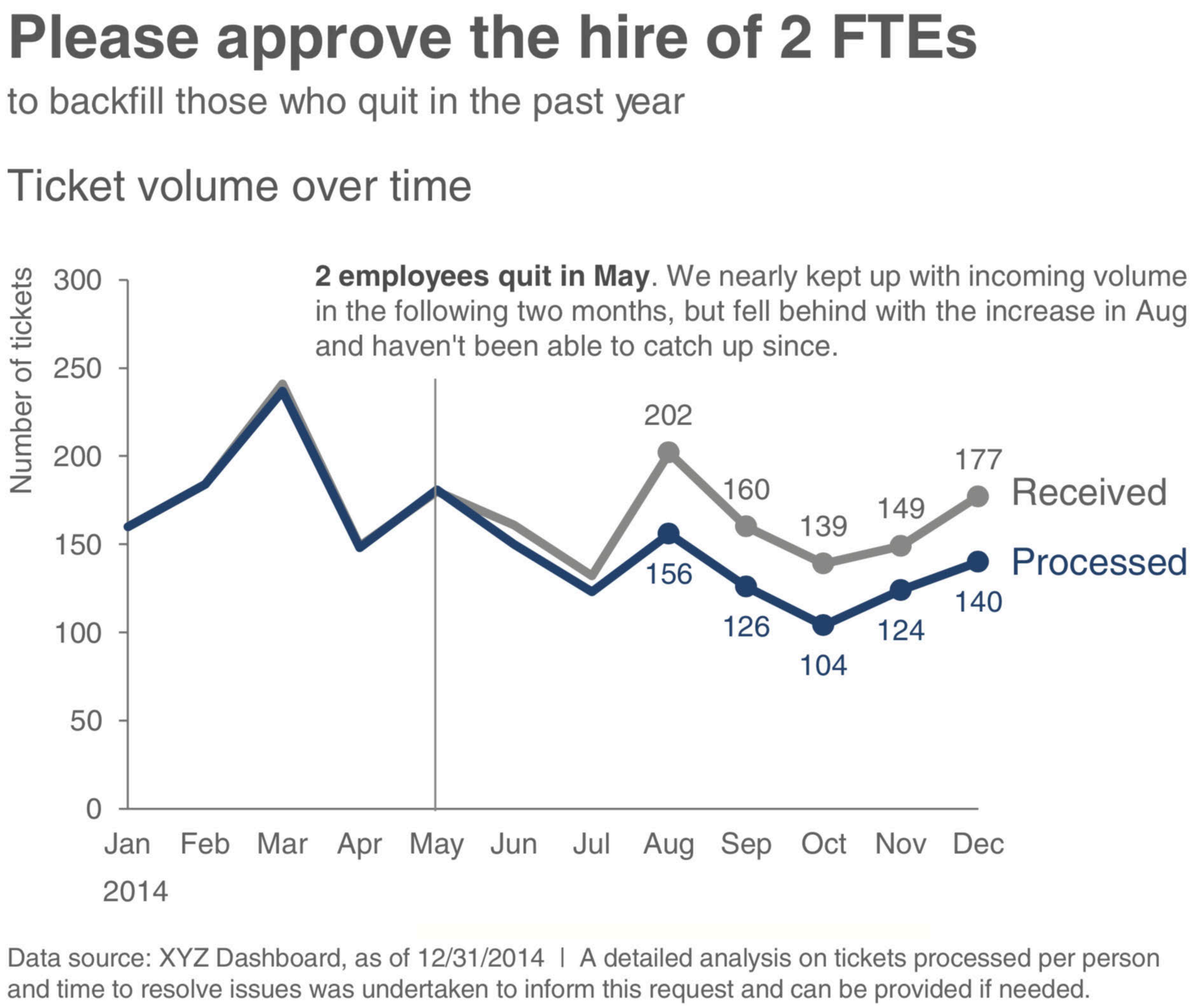


FIGURE 3.17 Original graph



Knaflic, Cole Nussbaumer. *Storytelling with Data*. Wiley, 2015. Print.

From default visual to visual narrative — erase clutter, add narrative annotation …

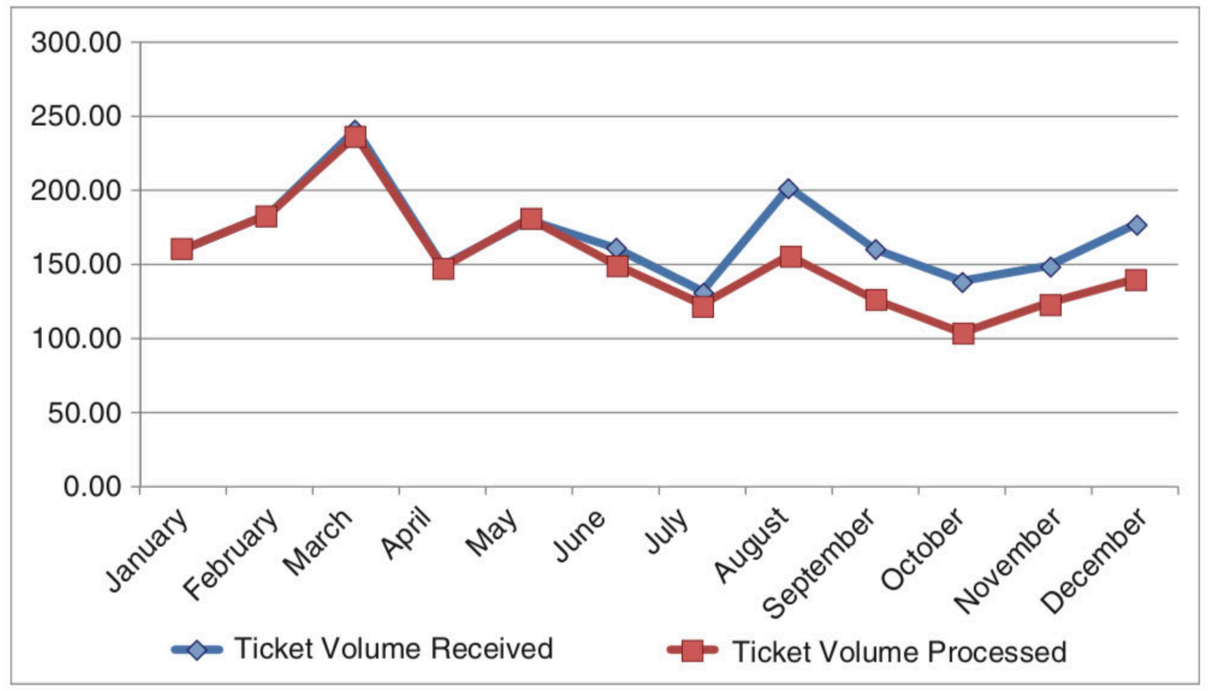


FIGURE 3.17 Original graph

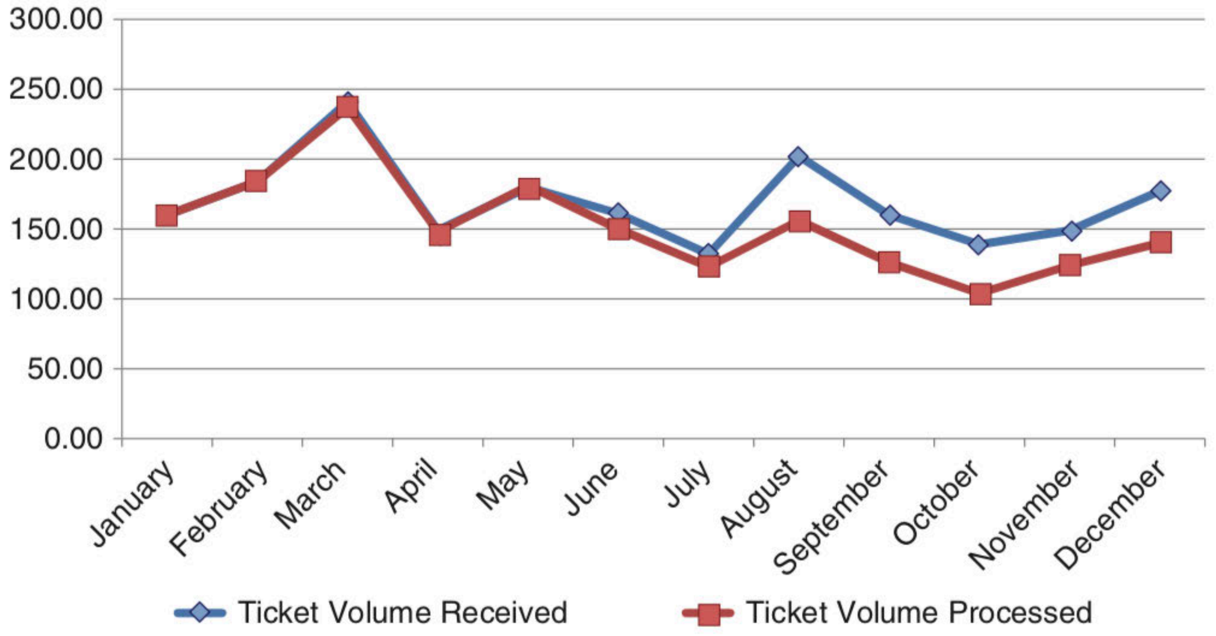


FIGURE 3.18 Remove chart border

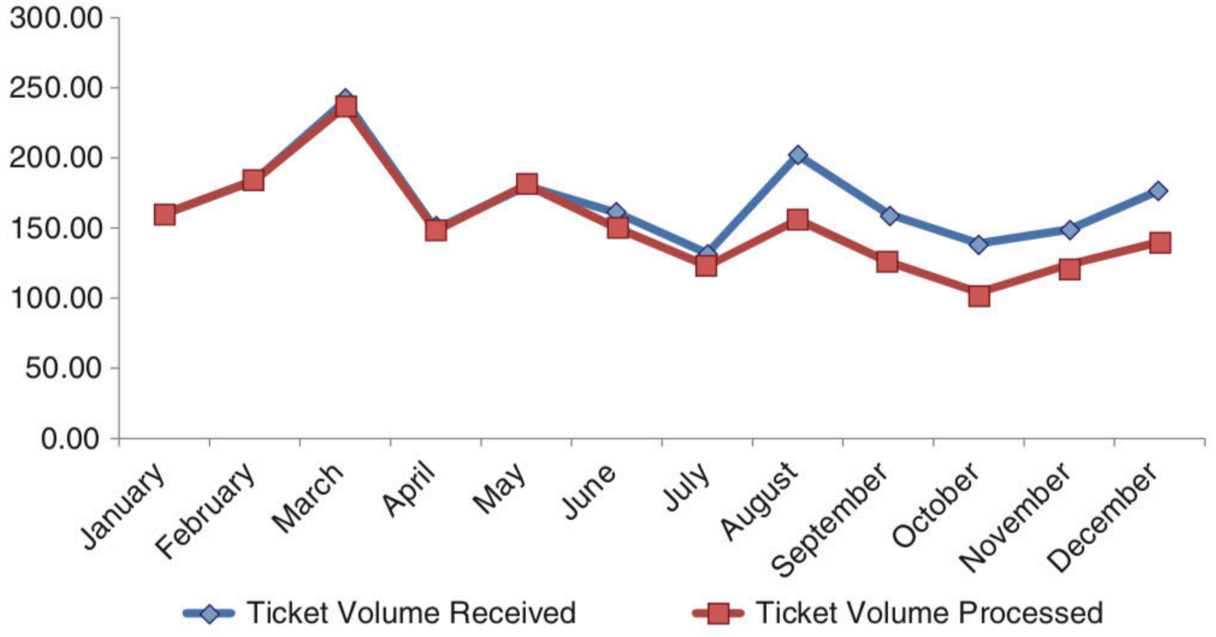


FIGURE 3.19 Remove gridlines

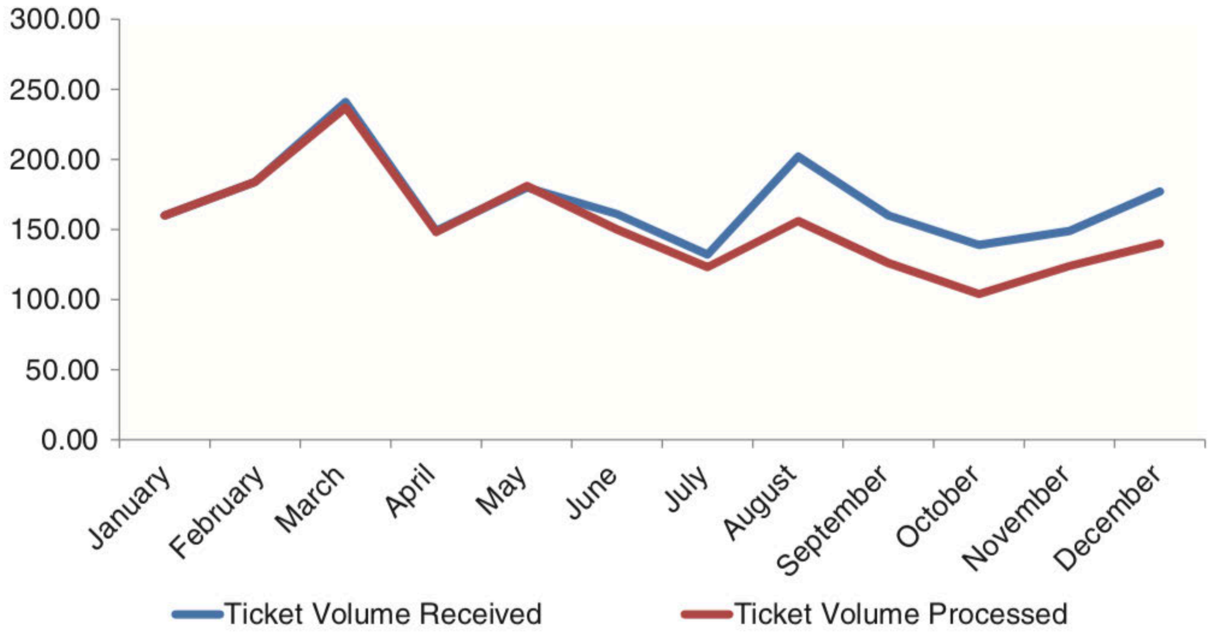


FIGURE 3.20 Remove data markers

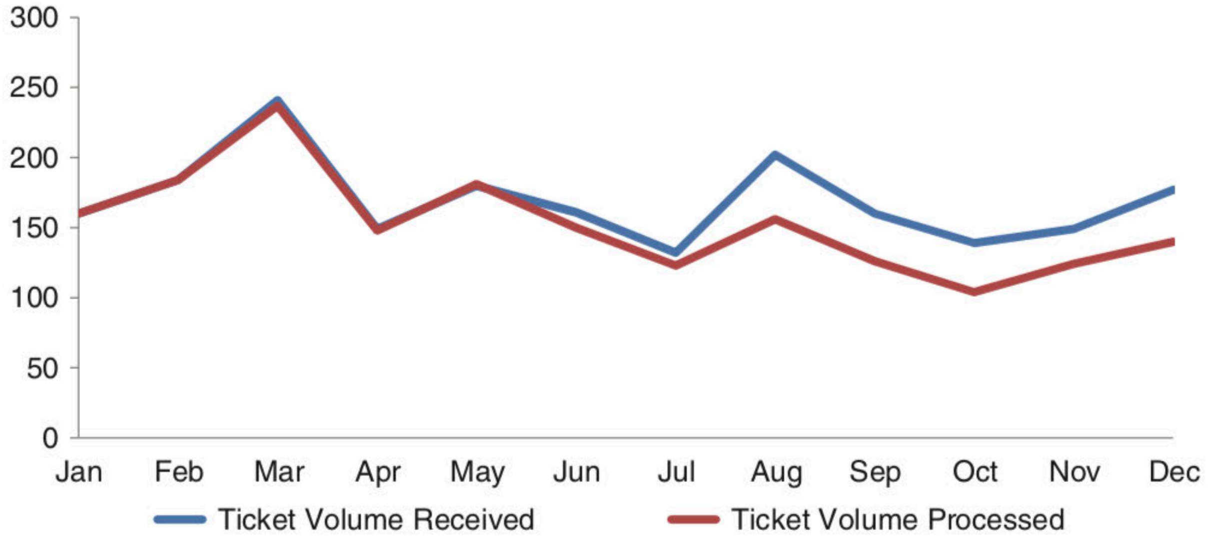


FIGURE 3.21 Clean up axis labels

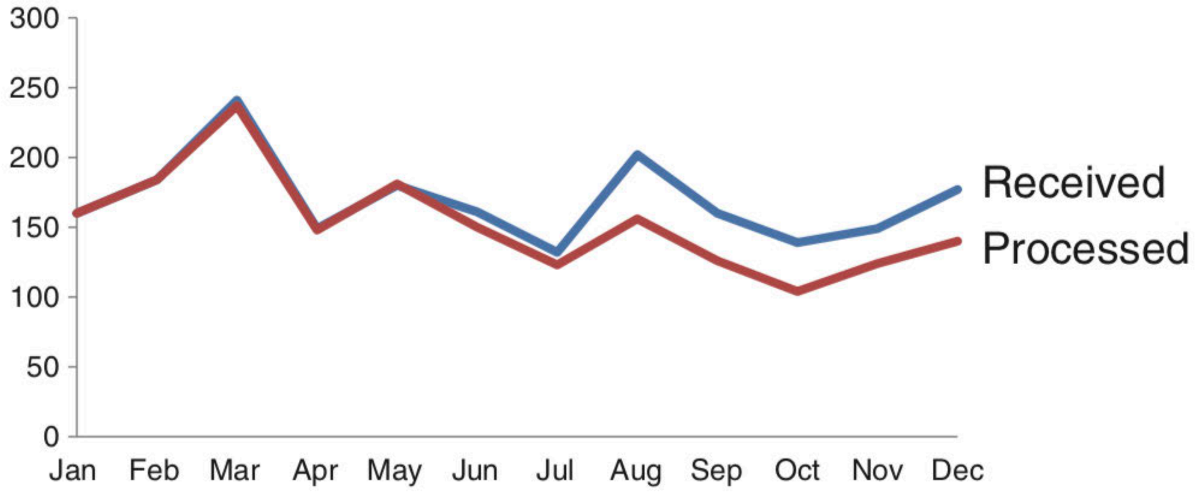


FIGURE 3.22 Label data directly

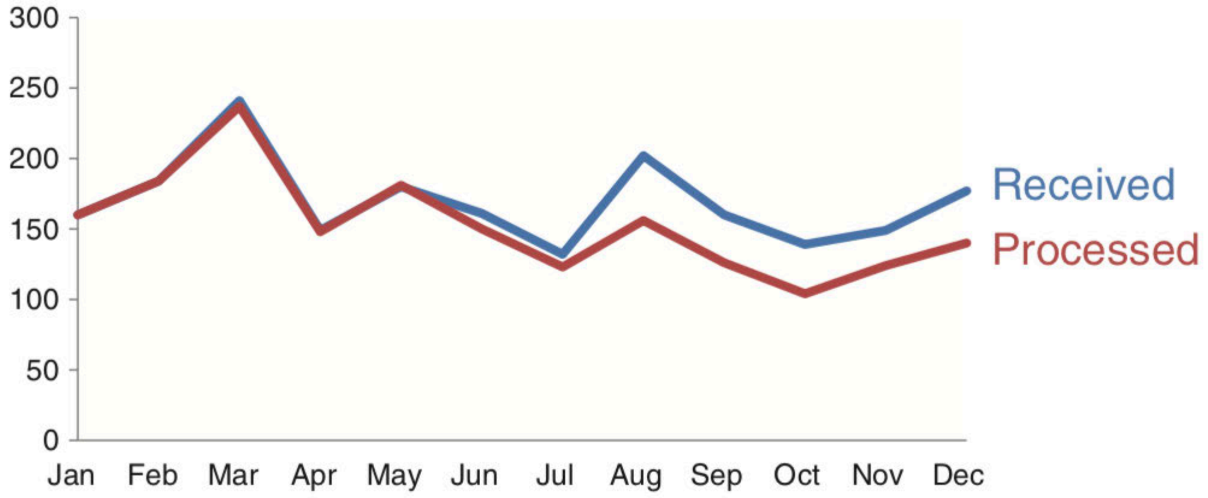


FIGURE 3.23 Leverage consistent color

FIGURE 4.11 First, push everything to the background

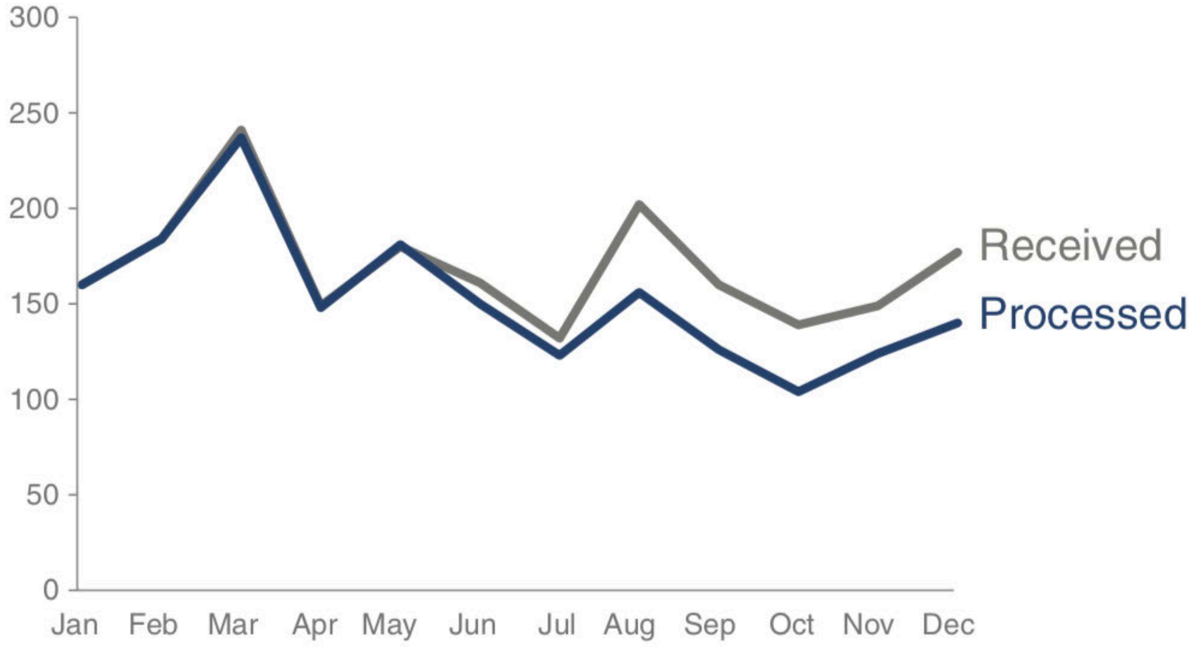


FIGURE 4.12 Make the data stand out

Knaflic, Cole Nussbaumer. *Storytelling with Data.* Wiley, 2015. Print.

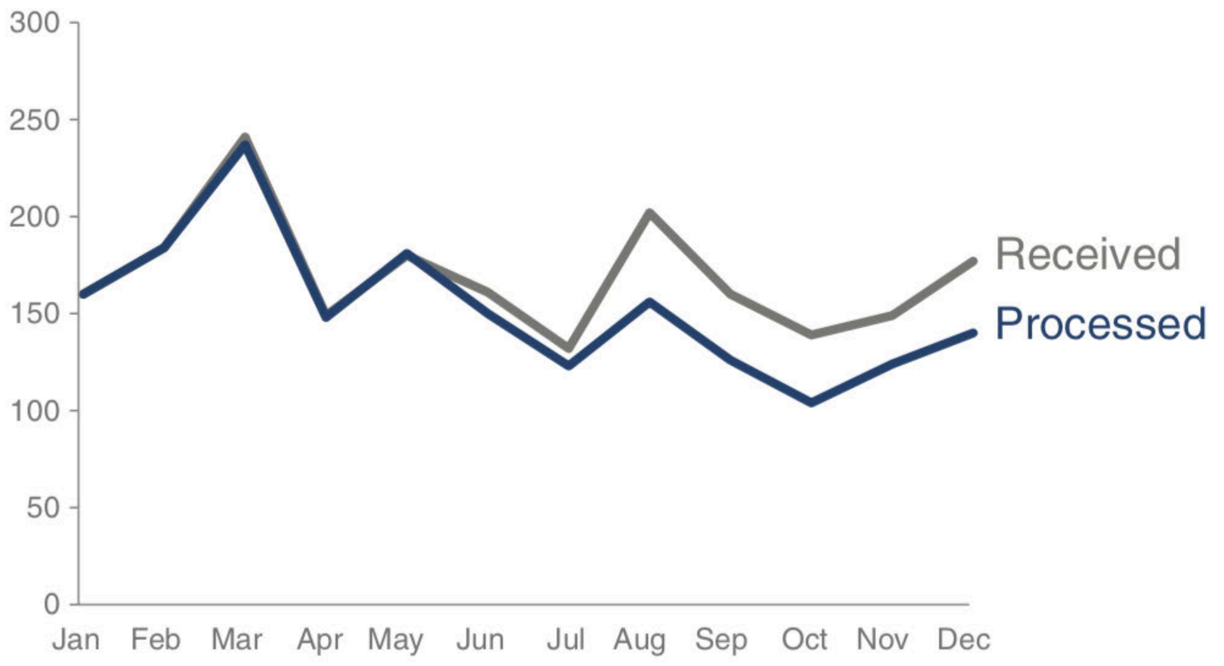


FIGURE 4.12 Make the data stand out

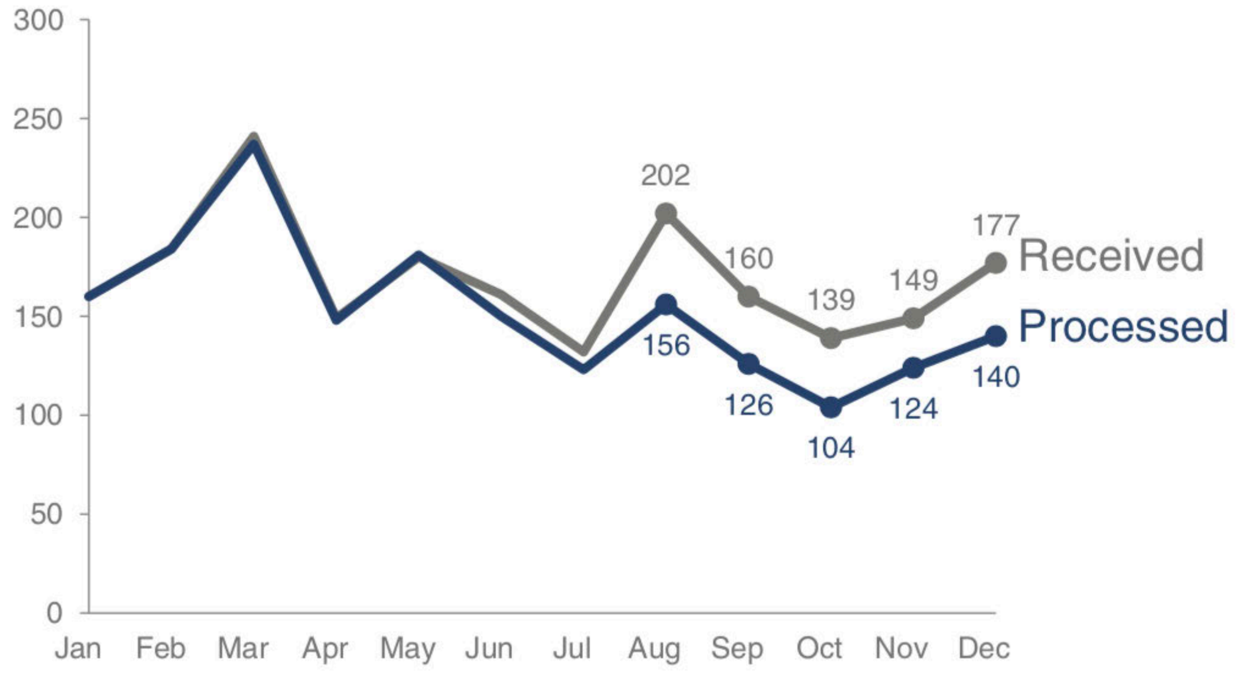


FIGURE 4.14 Data labels used sparingly help draw attention

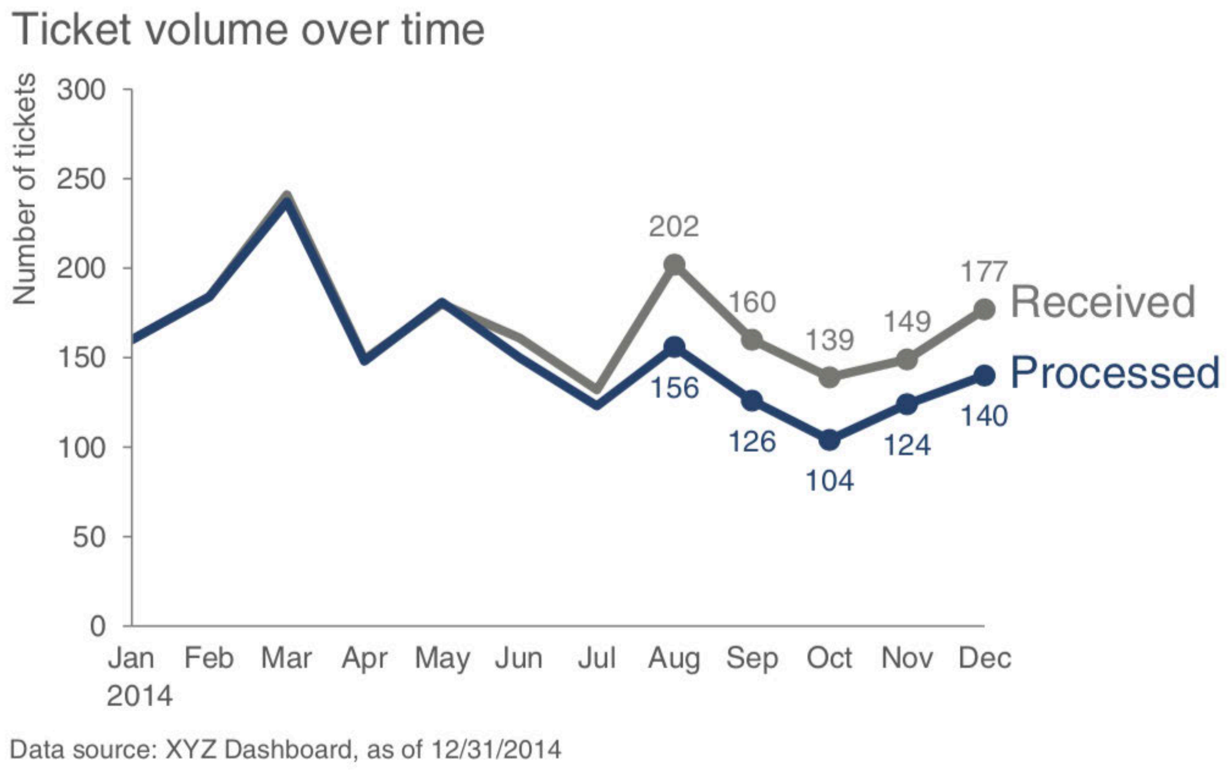


FIGURE 5.9 Use words to make the graph accessible

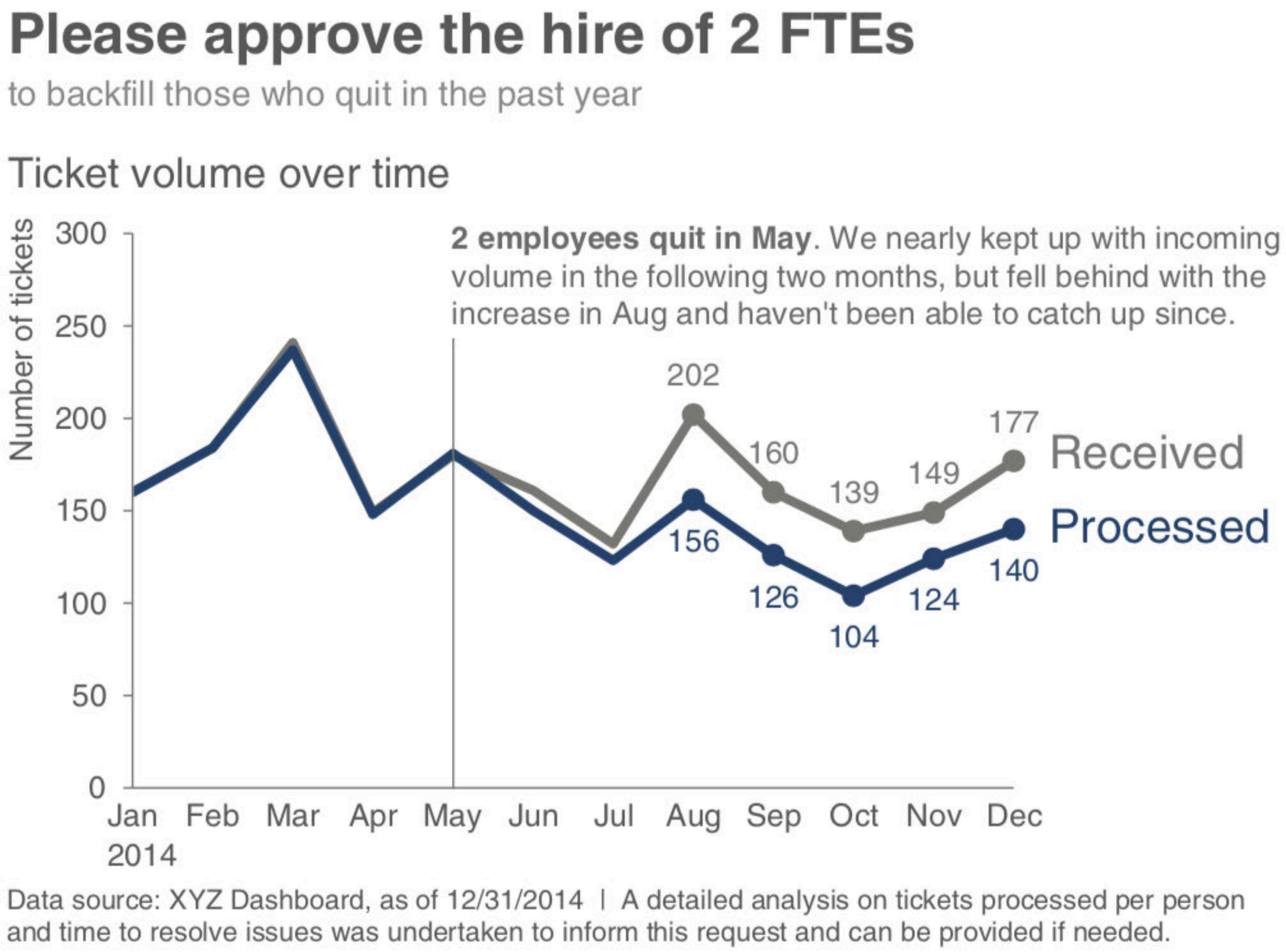


FIGURE 5.10 Add action title and annotation

Knaflic, Cole Nussbaumer. *Storytelling with Data*. Wiley, 2015. Print.

Between now and next class

For Next Week, Module 8:

Agenda next week

Next deliverable, *final* storyboard

Theory and best practices for the visual components of your analytics project, continued

The minimum

Hullman, Jessica et al. *Imagining Replications: Graphical Prediction & Discrete Visualizations Improve Recall & Estimation of Effect Uncertainty*. IEEE Transactions on Visualization and Computer Graphics 24.1 (2017): 446–456. Web.

How do audiences perceive uncertainty? Consider approaches to communicating uncertainty important to your messages about your projects.

Song, Hayeong, and Danielle Albers Szafir. *Where's My Data? Evaluating Visualizations with Missing Data*. IEEE Transactions on Visualization and Computer Graphics 25.1 (2018): 914–924. Web.

How should might we reason, explore, visualize and communicate the implications of missing data?

Lupi, Giorgia. *The New Aesthetic of Data Narrative* in Chapter 3 of Bihanic, David. New Challenges for Data Design. Springer, 2015. Print.

Read to get a sense of how Lupi thinks through making a graphic.

Feedback

Coding?

Was the example walk through of code to create the CitiBike visual helpful? Conceptually, what would you like code tutorials to cover?

Uncertain?

What types of uncertainty have you identified in your project, and how might that be important for our different audiences?

Visualize it

Consider three ways you might visually represent the uncertainty you identified.

See you next week!

Spencer, Scott. Columbia C, Harlem River. 11 Mar 2019.. Illustration.