Storytelling with Data

Module 4: Principles of persuasion and brief proposals

Scott Spencer
Faculty and Lecturer
Columbia University

What we've discussed so far

Adapt to your audience

Understand data context
Choice of appropriate visual display
Eliminate clutter
Focus audience attention
Think like a designer
Tell a story

Knaflic's
Storytelling with data

Doumont's
Trees, Maps, Theorems

messages first, then details

step into their shoes!
CAO, CMO, CEO

Example **Jakarta** proposal
Improving traffic safety through video analysis

background > goals > problem > method > impact

Technical audience, not employee

Messages, *not just information*

beyond the minimum

Identifying events, user behaviors

Citi Bike,
example case studies
Measurements of events and behaviors

data analytics project proposal

Columbia University
The Writing Center

complexity last

Booth's
Revising style

subjects of verbs should be central characters

old before new

Technical audience, employee
Example 250-word memo
Dodgers, *game decisions should optimize expectations*

background > goals > problem > data > method > impact

be concise, every word tell

Strunk & White's
The Elements of style
overstatements diminish credibility

TL;DR

What problem is to be solved?
Is it important?
Does it have impact?
Do data play a role in solving the problem?
Are the right data available?
Is the organization ready to tackle the problem and take actions from insights?

Spencer's
Scoping a data analytics project

decisions > goals and actions > methods > data

ING's **General audience**
The Next Rembrandt

Getting to storytelling with data

Andrews, R J. *Info We Trust: How to Inspire the World with Data*. Wiley, 2019. Print. Drawing from page 6.

Agenda

Next deliverable — brief proposal

Today’s objectives

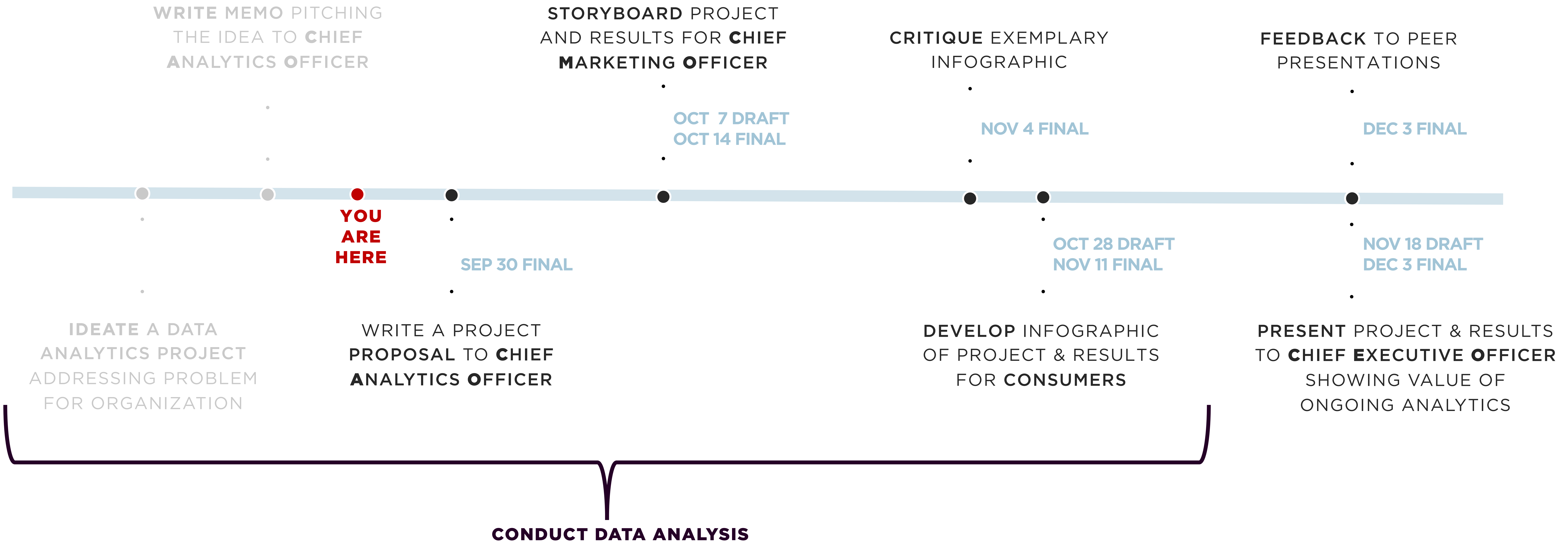
Perspectives on persuasion

Comparison, metaphor, patterns

Next deliverable

Upcoming deliverable

750-word brief proposal — Write a brief proposal to **CAO** detailing your proposed analytics project. Consider background context, problem, data, solution, and impact. At this point you should have data to start an analysis.



Example *draft* brief proposal

Example *draft* proposal. **Constraint—750 words or less in main body.**

Proposal for exploring game decisions informed by expectations of joint probability distributions

To: Scott Powers, Senior Baseball Analyst, Los Angeles Dodgers
From: Scott Spencer, Faculty and Lecturer, Columbia University

14 February 2019

Our game decisions based on current modeling do not maximize spend per win. We witnessed the mid-market Astros use analytics to overtake us in the 2017 World Series (Luhnow 2018ab). Our efforts also do not maximize expected wins. But we can. To do so, we need to jointly model probabilities of all game events and base decisions on *expectations* of those distributions. With adequate computing emerging, we can be first using the probabilistic programming language Stan and parallel processing. To demonstrate the concept, consider a probability model for decisions to steal second base, below, which suggests teams are too conservative, leaving wins unclaimed. This model allows us to ask, for example—*should Sanchez steal against Sabathia? Or against Pineda?*

1 Our current analyses do not optimize expected wins

Seven terabytes of uncompressed data generated per game overshadow the lack of situational data needed for decision-making that maximizes expected utility. Consider that pitchers, on average, only face10 percent of major league batters regardless of game state; the reverse is true, too. Or when deciding whether a base runner should attempt to steal against a specific pitcher and catcher in a state of play, say, we are lucky to have any data. Common analyses and heuristics for these situations are inadequate: they not only overfit the data (if any exist), but also offer no manner of estimating changes in probabilities for maximizing *expected* utility (winning the game).

Accurately quantifying probabilities, and changes thereof, in a given context enable us to answer counterfactuals, from which we can build strategies that maximize our objectives (Parmigiani 2002). This approach is possible at scale using Stan (Carpenter et al. 2017). It's time to jointly model probabilities of all events.

2 Modeling probabilities for steal success illustrates a broader benefit

To see the potential of implementing probability models, let's consider, again, the decision to steal bases, given a specific counterfactual:

> 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:

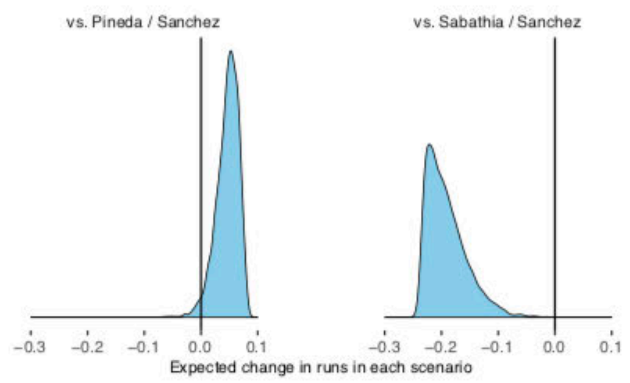


Figure 1. Of the two scenarios, Cain should only attempt to steal against the Sanchez–Pineda duo.

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:

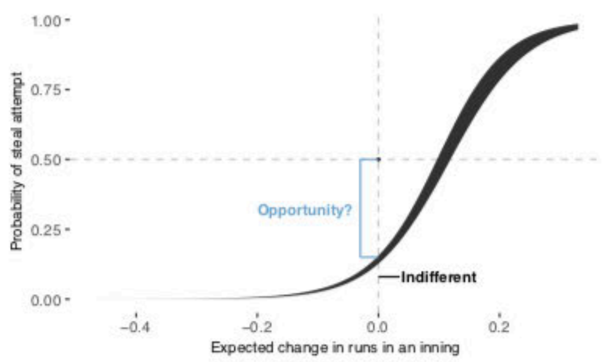


Figure 2. When the change in expected runs is zero, managers should be indifferent to attempted steals, saying go half the time.

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.

3 For value, compare an investment to free-agent costs

A fully-realized model will require significant effort from a team with deep experience in baseball, generative modeling, and Stan. To get the talent, we should compare cost to acquiring expected wins from free-agents. Each win above a *replacement-level* player costs about 10 million per year (Swartz 2017). As with free-agent value over replacement player, game-time decisions informed from more accurate probabilities should add wins over a season. The scope of what we can answer, moreover, goes beyond in-game strategy (player acquisitions, salary arbitration). More immediately, however, we can begin to implement this approach for specific events, with a scope closer to the example above, being mindful that information learnt are conditional upon unmodeled context.

4 For accuracy, compare model results to betting market odds

Measuring performance of a fully-realized model may seem tricky: *we only see the outcome of our decisions*. But we can, say, compare the accuracy of our estimates against the betting market where interested investors are trying to forecast game outcomes.

5 Conclusion

The mid-market Astros show teams can do more with information. Millions in additional revenue—and more wins—await discovery through a joint, probability model of all events from which we can maximize conditional expectations. Let's discuss how to draw the talent for a title worth our spend.

6 References

Carpenter, Bob, et. al. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 76 (1): 1–32.

Luhnow, Jeff. 2018a. "How the Houston Astros are winning through advanced analytics." McKinsey Quarterly 13 June 2018: 1–9.

———. 2018b. "A view from the front lines of baseball's data-analytics revolution." McKinsey Quarterly 5 July 2018: 1–8.

Parmigiani, G. 2002. "Decision Theory: Bayesian." In *International Encyclopedia of the Social Behavioral Sciences*, 3327–34.

Swartz, Matt. 2017. "The Recent History of Free-Agent Pricing." https://www.fangraphs.com/blogs/the-recent-history-of-free-agent-pricing/.

Readability Statistics

Counts	
Words	720
Characters	3,997
Paragraphs	16
Sentences	35
Averages	
Sentences per Paragraph	4.3
Words per Sentence	18.1
Characters per Word	5.3
Readability	
Flesch Reading Ease	33.2
Flesch-Kincaid Grade Level	13
Passive Sentences	0%

Spencer, Scott. *Proposal for Exploring Game Decisions Informed by Expectations of Joint Probability Distributions*. Columbia University, 2019. Print.

Discussion: the example draft

How does the example draft proposal structure **compare** with your grading rubric?

Does the draft proposal **follow logically** from the memo?

How do **structure and details** of the memo and proposal compare and differ?

Does the draft proposal rely upon **comparisons or examples** to explain ideas? If so, what kind(s)?

Do you see **applications** of communication concepts we've discussed so far?

Example *draft* proposal. **Messaging—We want messages first, not just information. Details follow.**

Proposal for exploring game decisions informed by expectations of joint probability distributions

To: Scott Powers, Senior Baseball Analyst, Los Angeles Dodgers
From: Scott Spencer, Faculty and Lecturer, Columbia University

14 February 2019

Our game decisions based on current modeling do not maximize spend per win. We witnessed the mid-market Astros use analytics to overtake us in the 2017 World Series (Luhnow 2018ab). Our efforts also do not maximize expected wins. But we can. To do so, we need to jointly model probabilities of all game events and base decisions on *expectations* of those distributions. With adequate computing emerging, we can be first using the probabilistic programming language Stan and parallel processing. To demonstrate the concept, consider a probability model for decisions to steal second base, below, which suggests teams are too conservative, leaving wins unclaimed. This model allows us to ask, for example—*should Sanchez steal against Sabathia? Or against Pineda?*

1 Our current analyses do not optimize expected wins

Seven terabytes of uncompressed data generated per game overshadow the lack of situational data needed for decision-making that maximizes expected utility. Consider that pitchers, on average, only face10 percent of major league batters regardless of game state; the reverse is true, too. Or when deciding whether a base runner should attempt to steal against a specific pitcher and catcher in a state of play, say, we are lucky to have any data. Common analyses and heuristics for these situations are inadequate: they not only overfit the data (if any exist), but also offer no manner of estimating changes in probabilities for maximizing *expected* utility (winning the game).

Accurately quantifying probabilities, and changes thereof, in a given context enable us to answer counterfactuals, from which we can build strategies that maximize our objectives (Parmigiani 2002). This approach is possible at scale using Stan (Carpenter et al. 2017). It's time to jointly model probabilities of all events.

2 Modeling probabilities for steal success illustrates a broader benefit

To see the potential of implementing probability models, let's consider, again, the decision to steal bases, given a specific counterfactual:

3 For value, compare an investment to free-agent costs

A fully-realized model will require significant effort from a team with deep experience in baseball, generative modeling, and Stan. To get the talent, we should compare cost to acquiring expected wins from free-agents. Each win above a *replacement-level* player costs about 10 million per year (Swartz 2017). As with free-agent value over replacement player, game-time decisions informed from more accurate probabilities should add wins over a season. The scope of what we can answer, moreover, goes beyond in-game strategy (player acquisitions, salary arbitration). More immediately, however, we can begin to implement this approach for specific events, with a scope closer to the example above, being mindful that information learnt are conditional upon unmodeled context.

4 For accuracy, compare model results to betting market odds

Measuring performance of a fully-realized model may seem tricky: *we only see the outcome of our decisions*. But we can, say, compare the accuracy of our estimates against the betting market where interested investors are trying to forecast game outcomes.

5 Conclusion

The mid-market Astros show teams can do more with information. Millions in additional revenue—and more wins—await discovery through a joint, probability model of all events from which we can maximize conditional expectations. Let's discuss how to draw the talent for a title worth our spend.

6 References

Carpenter, Bob, et. al. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 76 (1): 1–32.

Luhnow, Jeff. 2018a. "How the Houston Astros are winning through advanced analytics." McKinsey Quarterly 13 June 2018: 1–9.

———. 2018b. "A view from the front lines of baseball's data-analytics revolution." McKinsey Quarterly 5 July 2018: 1–8.

Parmigiani, G. 2002. "Decision Theory: Bayesian." In *International Encyclopedia of the Social Behavioral Sciences*, 3327–34.

Swartz, Matt. 2017. "The Recent History of Free-Agent Pricing." https://www.fangraphs.com/blogs/the-recent-history-of-free-agent-pricing/.

Doumont, Jean-Luc. *Trees, Maps, and Theorems*. Principiæ, 2009.

Spencer, Scott. *Proposal for Exploring Game Decisions Informed by Expectations of Joint Probability Distributions*. Columbia University, 2019. Print.

Example *draft* proposal. **Typography—Grid**: two columns with gutter. **Fonts** signal hierarchy.

Proposal for exploring game decisions informed by expectations of joint probability distributions

To: Scott Powers, Senior Baseball Analyst, Los Angeles Dodgers
From: Scott Spencer, Faculty and Lecturer, Columbia University

14 February 2019

Our game decisions based on current modeling do not maximize spend per win. We witnessed the mid-market Astros use analytics to overtake us in the 2017 World Series (Luhnow 2018ab). Our efforts also do not maximize expected wins. But we can. To do so, we need to jointly model probabilities of all game events and base decisions on *expectations* of those distributions. With adequate computing emerging, we can be first using the probabilistic programming language Stan and parallel processing. To demonstrate the concept, consider a probability model for decisions to steal second base, below, which suggests teams are too conservative, leaving wins unclaimed. This model allows us to ask, for example—*should Sanchez steal against Sabathia? Or against Pineda?*

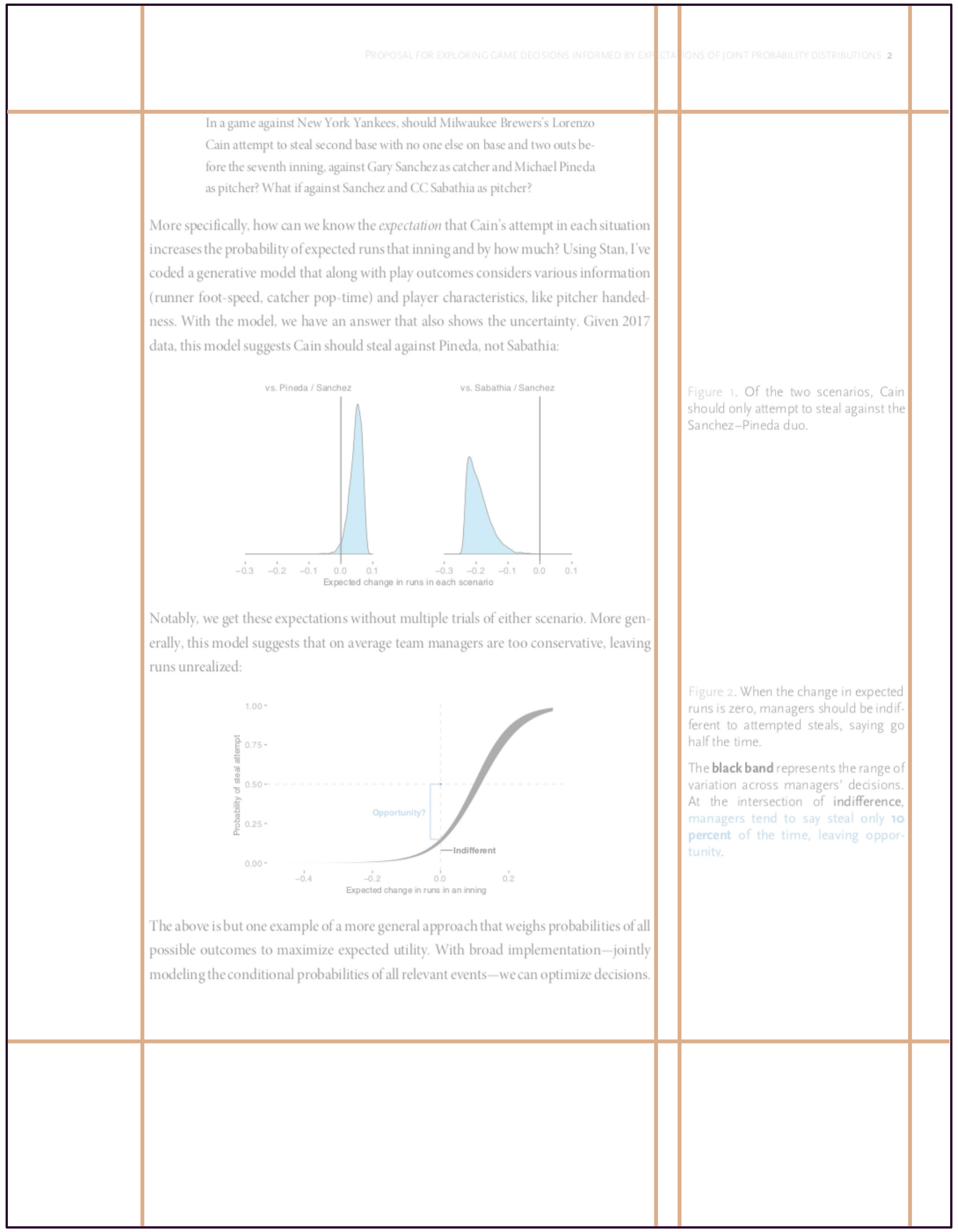
1 Our current analyses do not optimize expected wins

Seven terabytes of uncompressed data generated per game overshadow the lack of situational data needed for decision-making that maximizes expected utility. Consider that pitchers, on average, only face10 percent of major league batters regardless of game state; the reverse is true, too. Or when deciding whether a base runner should attempt to steal against a specific pitcher and catcher in a state of play, say, we are lucky to have any data. Common analyses and heuristics for these situations are inadequate: they not only overfit the data (if any exist), but also offer no manner of estimating changes in probabilities for maximizing *expected* utility (winning the game).

Accurately quantifying probabilities, and changes thereof, in a given context enable us to answer counterfactuals, from which we can build strategies that maximize our objectives (Parmigiani 2002). This approach is possible at scale using Stan (Carpenter et al. 2017). It's time to jointly model probabilities of all events.

2 Modeling probabilities for steal success illustrates a broader benefit

To see the potential of implementing probability models, let's consider, again, the decision to steal bases, given a specific counterfactual:



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.

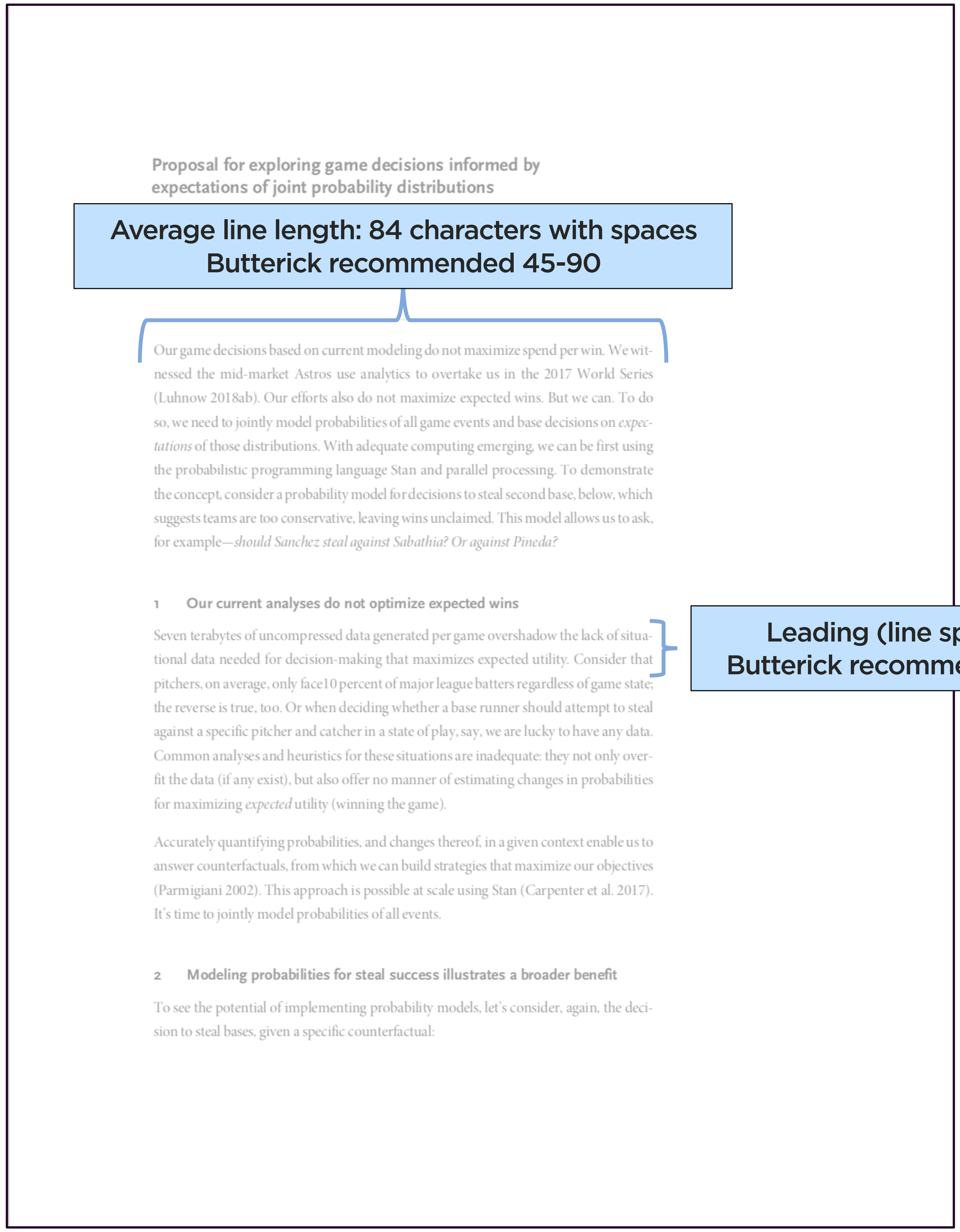
Müller-Brockmann, Josef. *Grid Systems in Graphic Design*. ARTHUR NIGGLI LTD., 1996. Print.

Tondreau, Beth. *Layout Essentials*. Rockport, 2008. Print.

Samara, Timothy. *Making and Breaking the Grid*. Second. Rockport, 2017. Print.

Spencer, Scott. *Proposal for Exploring Game Decisions Informed by Expectations of Joint Probability Distributions*. Columbia University, 2019. Print.

Example *draft* proposal. **Typography—Layout. Consider guidelines that ease readability.**



Proposal for exploring game decisions informed by expectations of joint probability distributions

Our game decisions based on current modeling do not maximize spend per win. We witnessed the mid-market Astros use analytics to overtake us in the 2017 World Series (Luhnow 2018ab). Our efforts also do not maximize expected wins. But we can. To do so, we need to jointly model probabilities of all game events and base decisions on *expectations* of those distributions. With adequate computing emerging, we can be first using the probabilistic programming language Stan and parallel processing. To demonstrate the concept, consider a probability model for decisions to steal second base, below, which suggests teams are too conservative, leaving wins unclaimed. This model allows us to ask, for example—*should Sanchez steal against Sabathia? Or against Pineda?*

1 Our current analyses do not optimize expected wins

Seven terabytes of uncompressed data generated per game overshadow the lack of situational data needed for decision-making that maximizes expected utility. Consider that pitchers, on average, only face10 percent of major league batters regardless of game state; the reverse is true, too. Or when deciding whether a base runner should attempt to steal against a specific pitcher and catcher in a state of play, say, we are lucky to have any data. Common analyses and heuristics for these situations are inadequate: they not only overfit the data (if any exist), but also offer no manner of estimating changes in probabilities for maximizing *expected* utility (winning the game).

Accurately quantifying probabilities, and changes thereof, in a given context enable us to answer counterfactuals, from which we can build strategies that maximize our objectives (Parmigiani 2002). This approach is possible at scale using Stan (Carpenter et al. 2017). It's time to jointly model probabilities of all events.

2 Modeling probabilities for steal success illustrates a broader benefit

To see the potential of implementing probability models, let's consider, again, the decision to steal bases, given a specific counterfactual:

Average line length: 84 characters with spaces
Butterick recommended 45-90

Leading (line spacing): 145% of font size
Butterick recommended: 120-145% of font size

Butterick, Matthew. *Butterick's Practical Typography*. practicaltypography.com. N.p., 2018. Web. 8 Sept. 2018.

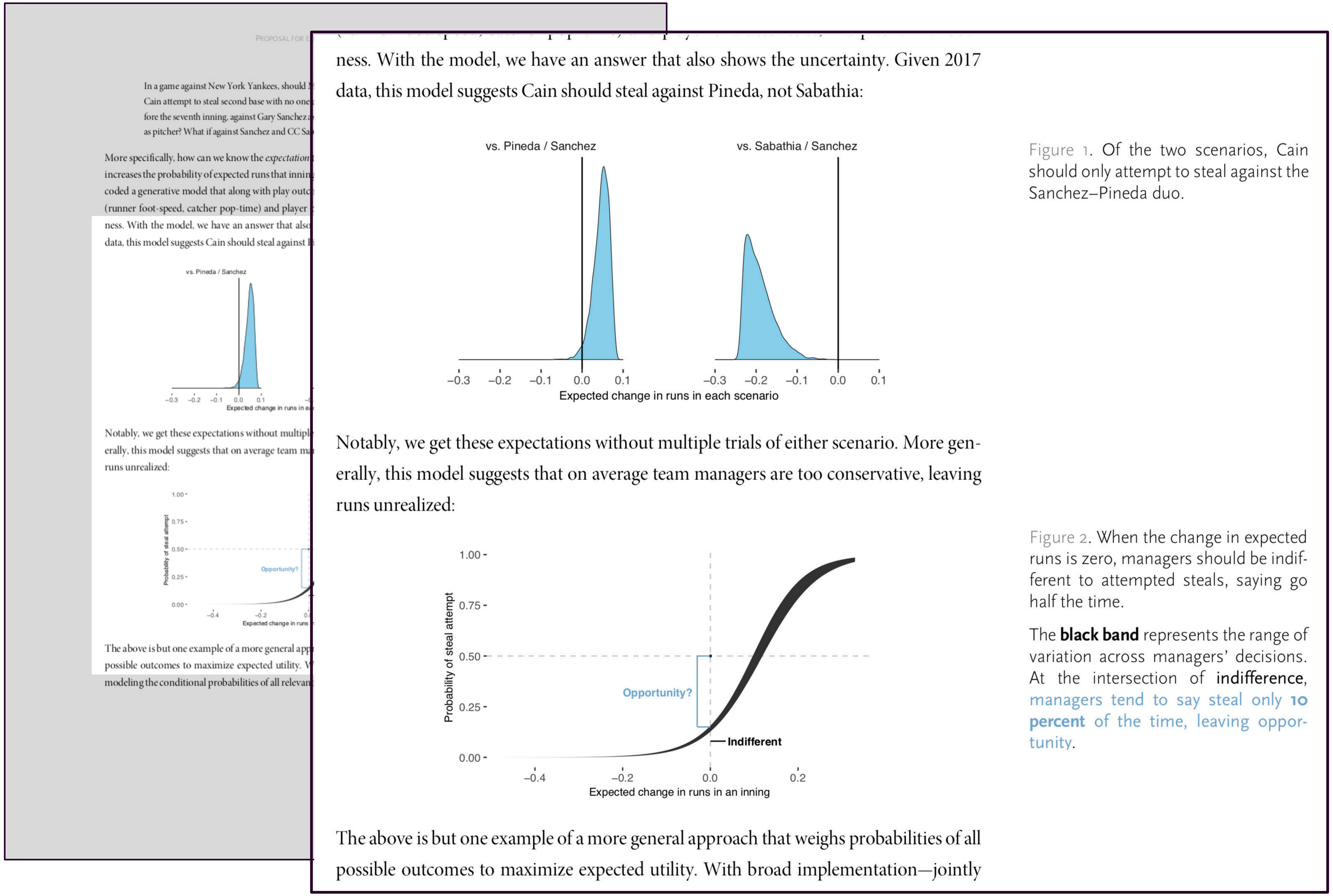
See also:

Bringhurst, Robert. *The Elements of Typographic Style*. Fourth. Hartley & Marks, 2012. Print.

Rutter, Richard. *Web Typography*. Ampersand Type, 2017. Print.

Spencer, Scott. *Proposal for Exploring Game Decisions Informed by Expectations of Joint Probability Distributions*. Columbia University, 2019. Print.

Example *draft* proposal. **Graphics as paragraphs; annotation, linking words to data in graphics.**



ness. 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

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

Kay, Matthew. *Figures*. www.mjskay.com. Aug. 2015. Web. 28 Mar. 2019.

Riche, Nathalie Henry et al. Ch. 9, *Communicating Data to an Audience*, in *Data-Driven Storytelling*. CRC Press, 2018.

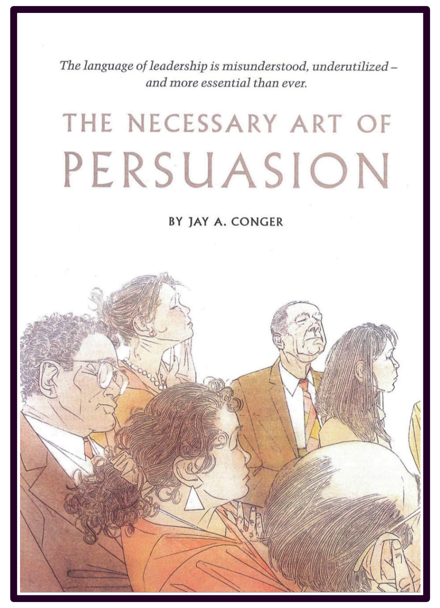
Spencer, Scott. *Proposal for Exploring Game Decisions Informed by Expectations of Joint Probability Distributions*. Columbia University, 2019. Print.

Today's Objectives

Objectives

1. Explain the role of persuasion in getting buy-in for analytics projects.
2. Explain the role of persuasion in implementing analytic insights.
3. Employ tools and techniques taught in class to persuade technical and non-technical audiences.

Perspectives on persuasion



Necessary art of persuasion

Conger

Conger is an executive educator, coach, and program designer who teaches leadership to companies and individuals.

Do we find any of his ideas implemented in our example draft memo? Would more help?

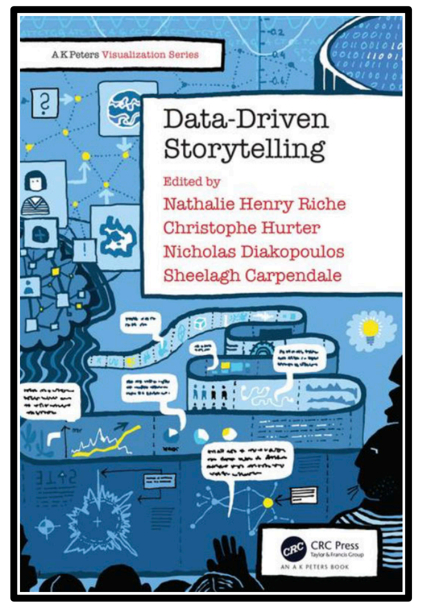
Establish credibility | First assess your credibility—*your knowledge about the strategy, product, or change proposed*—by **self reflection** and **asking others**. **Fill in gaps**: gain knowledge; cite outside sources; demonstrate the proposal by starting smaller.

Find common ground | Study the issues with colleagues; ***think through their arguments, evidence, and perspectives***. Address or include them, making your proposal something shared.

Combine evidence with story, metaphor | Numerical evidence should be **supplemented with** "examples, stories, metaphors, and analogies" to enliven your proposal. This is particularly helpful when presenting **comparable situations** to the one under discussion.

Connect emotionally | Understand **how your audience feels** on the issues, and recognize—even **share**—them. Empathize.

Conger, Jay. "The Necessary Art of Persuasion." Harvard Business Review (1998): 1–15. Print.



Narrative Design Patterns for Data-Driven Storytelling

Riche, co-editors

The editors are researchers and professors with focuses on human-computer interaction and information visualization.

Classical devices of rhetoric

The classical devices of rhetoric involve **logos** (reason, word), **ethos** (character, ideal), and **pathos** (experience, emotion).

Rhetoric in data-driven stories aim for truth, connect

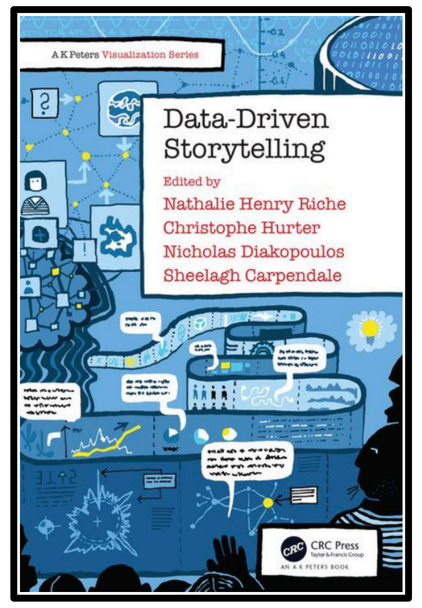
Though we believe the ultimate **goal of data-driven storytelling is to communicate truth** (most closely to logos), there are traces of both **pathos, and ethos in every story, which help connect the narrator with the audience**.

Patterns for argumentation

Argumentation is the action or process of reasoning systematically in support of an idea, action, or theory.

Patterns for argumentation serve the intent of persuading and convincing audiences.

Riche, Nathalie Henry et al. Chp. 5, *Data-Driven Storytelling*. CRC Press, 2018. Print.



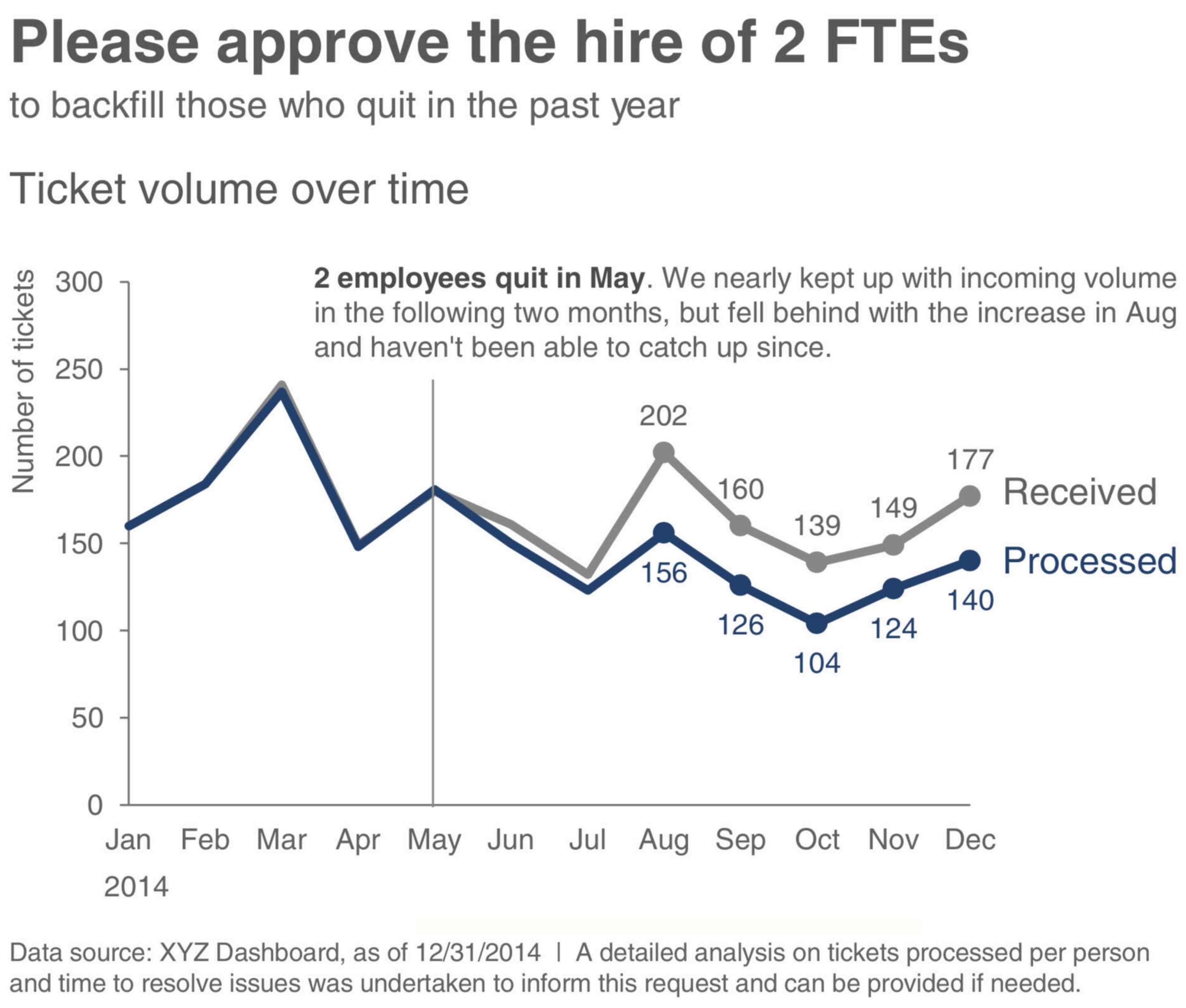
Narrative Design Patterns for Data-Driven Storytelling

Riche, co-editors

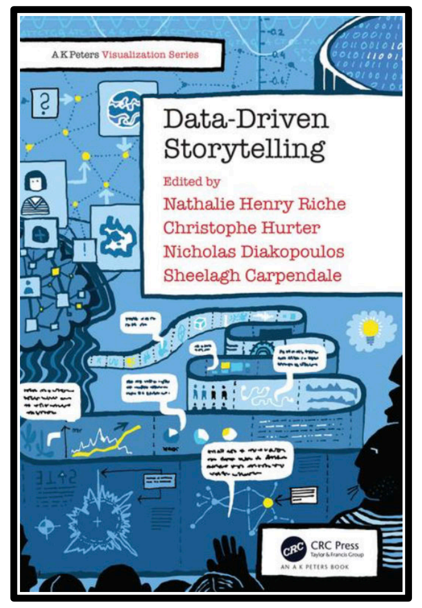
The editors are researchers and professors with focuses on human-computer interaction and information visualization.

Compare

Comparison allows the narrator to make the point about equality of both data sets, to explicitly highlight differences and similarities, or to give reasons for their difference.



Riche, Nathalie Henry et al. Chp. 5, *Data-Driven Storytelling*. CRC Press, 2018. Print.
Knaflic, Cole Nussbaumer. *Storytelling with Data*. Wiley, 2015. Print.



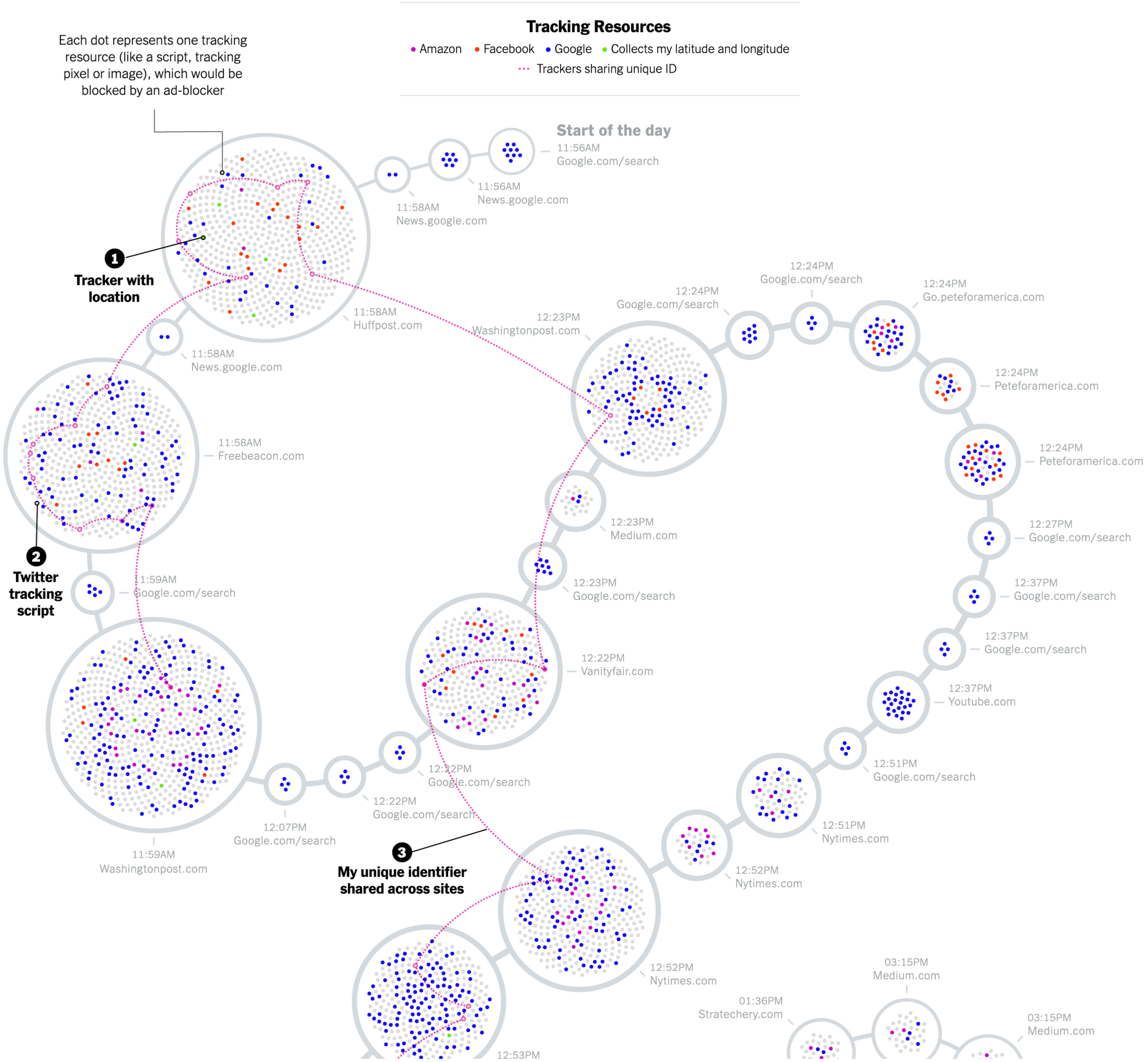
Narrative Design Patterns for Data-Driven Storytelling

Riche, co-editors

The editors are researchers and professors with focuses on human-computer interaction and information visualization.

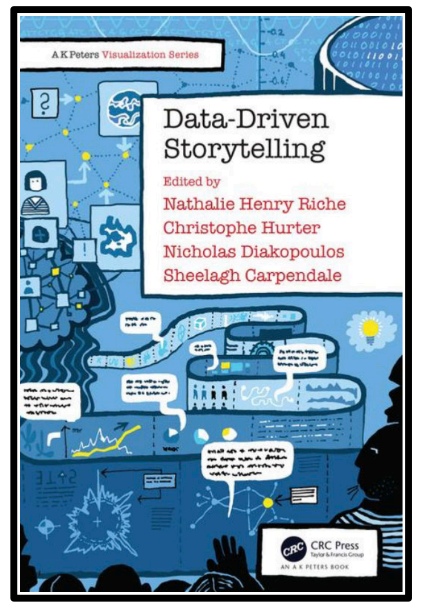
Concretize

Shows abstract concepts with concrete objects. Concretization usually implies that each data point is represented by an individual visual object (e.g., a point or shape), making them less abstract than aggregated statistics.



Riche, Nathalie Henry et al. Chp. 5, *Data-Driven Storytelling*. CRC Press, 2018. Print.

Manjoo, Farhad. Graphics by Bremer, Nadieh. *I Visited 47 Sites. Hundreds of trackers Followed Me.* New York Times. 2019 April 23. Web. https://www.nytimes.com/interactive/2019/08/23/opinion/data-internet-privacy-tracking.html



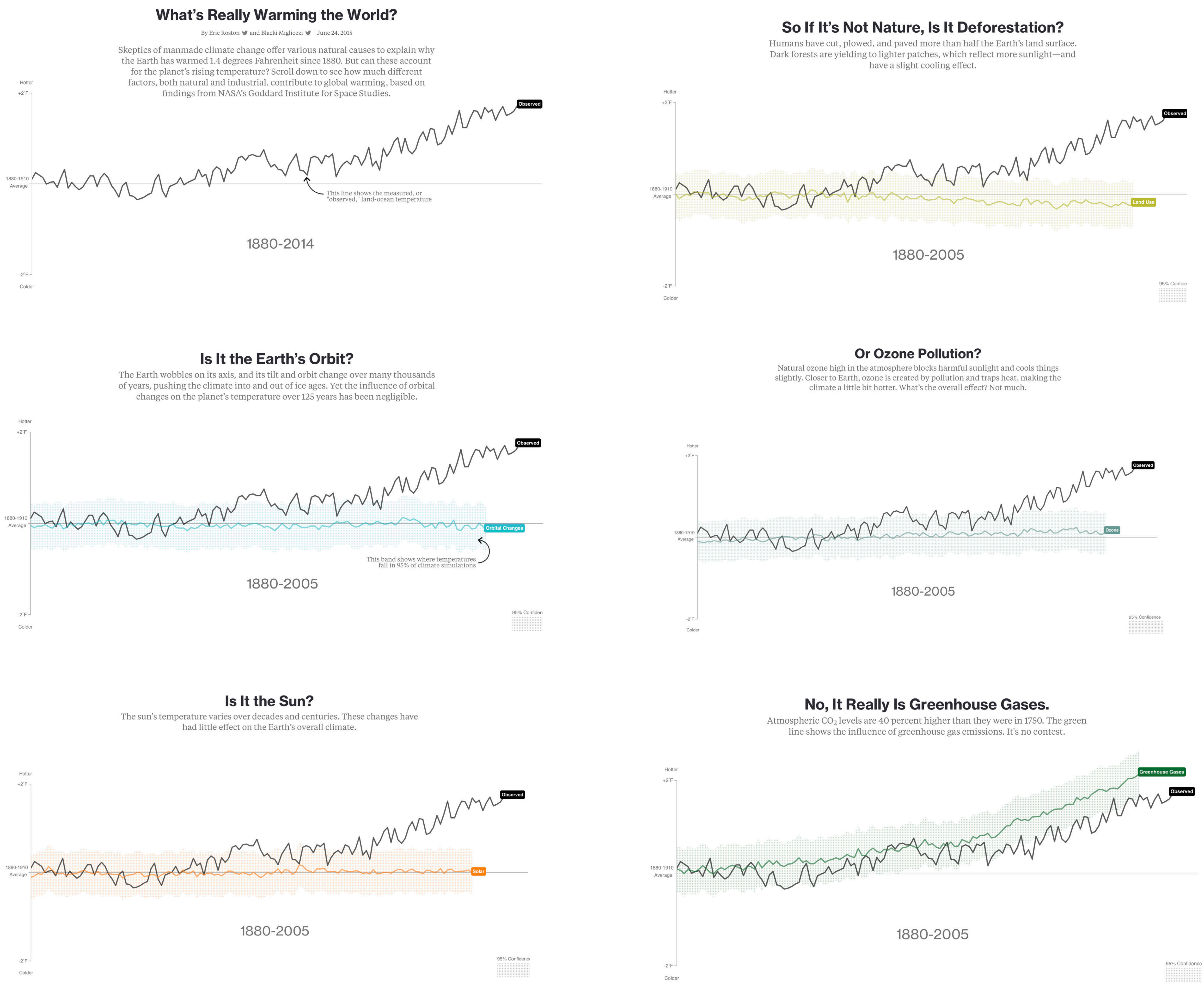
Narrative Design Patterns for Data-Driven Storytelling

Riche, co-editors

The editors are researchers and professors with focuses on human-computer interaction and information visualization.

Repetition

Repetition can increase a message's importance and memorability, and can help tie together different arguments about a given data set. Repetition can be employed as a means to search for an answer in the data.



Riche, Nathalie Henry et al. Chp. 5, *Data-Driven Storytelling*. CRC Press, 2018. Print.
Rosten, Eric. *What's Really Warming the World?* Bloomberg. 2015 June 24. Web. https://www.bloomberg.com/graphics/2015-whats-warming-the-world/

Statistical persuasion


IN PRACTICE

Ethics in statistical practice and communication
Five recommendations

Ethics in Statistical Practice and Communication

Gelman

Professor of Statistics and Political Science at Columbia University, he is known widely for his work in Bayesian statistics, and has authored several textbooks, including Teaching Statistics, and Bayesian Data Analysis.

Why statistics?

Consider this paradox: **statistics is the science of uncertainty and variation, but data-based claims in the scientific literature tend to be stated deterministically** (e.g. “We have discovered … the effect of X on Y is … hypothesis H is rejected”).

Is statistical communication about **exploration and discovery of the unexpected**, or is it about **making a persuasive, data-based case to back up an argument**?

Gelman, Andrew. *Ethics in Statistical Practice and Communication: Five Recommendations* Significance 15.5 (2018): 40–43. Web.


IN PRACTICE

Ethics in statistical practice and communication
Five recommendations

Ethics in Statistical Practice and Communication

Gelman

Professor of Statistics and Political Science at Columbia University, he is known widely for his work in Bayesian statistics, and has authored several textbooks, including Teaching Statistics, and Bayesian Data Analysis.

Exploring and persuading

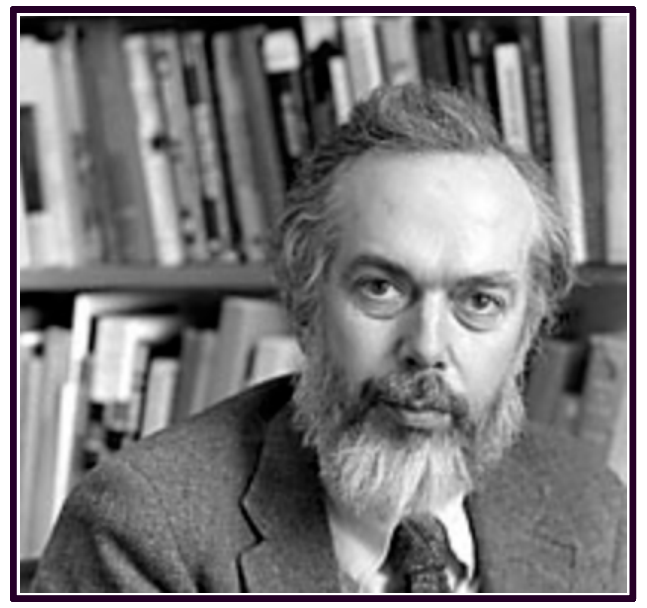
The answer to this question is necessarily **each at different times, and sometimes both at the same time**.

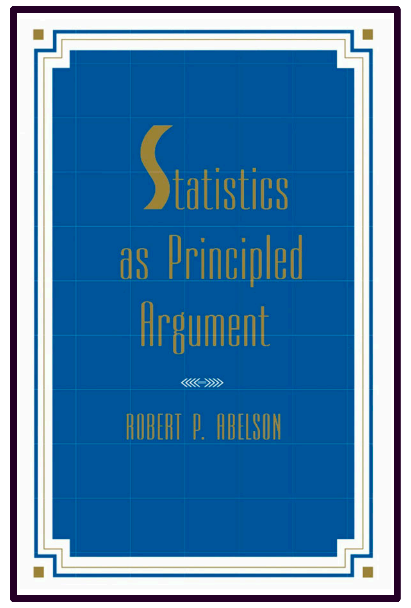
Just as you write in part in order to figure out what you are trying to say, so you do statistics not just to learn from data but also to learn what you can learn from data, and to decide how to gather future data to help resolve key uncertainties.

Traditional advice on statistics and ethics focuses on professional integrity, accountability, and responsibility to collaborators and research subjects.

All these are important, but when considering ethics, statisticians **must also wrestle with fundamental dilemmas regarding the analysis and communication of uncertainty and variation**.

Gelman, Andrew. *Ethics in Statistical Practice and Communication: Five Recommendations* Significance 15.5 (2018): 40–43. Web.





Statistics as principled argument

Abelson

Educated at MIT and Princeton, the late professor of psychology and political science taught at Yale 42 years, consulted for NBC, and was an analyst for three presidential campaigns.

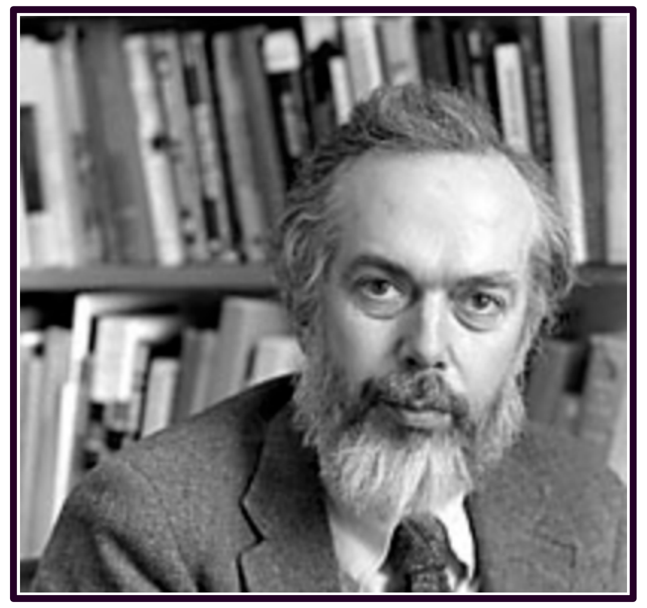
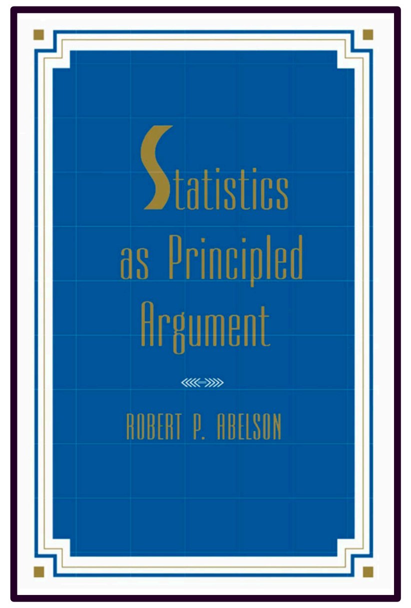
The purpose of statistics is persuasion

The purpose of statistics is to organize a useful argument from quantitative evidence, using a form of principled rhetoric … that conveys an interesting and credible point.

To make statistical arguments, it helps to wear different hats

His "image of the ideal statistician, already conceived as a good (but honest!) lawyer and a good storyteller, also includes the virtues of a good detective."

Abelson, Robert P. *Statistics as Principled Argument*. Psychology Press, 1995. Print.

Statistics as principled argument

Abelson

Educated at MIT and Princeton, the late professor of psychology and political science taught at Yale 42 years, consulted for NBC, and was an analyst for three presidential campaigns.

Comparison gives meaning

"**The idea of *comparison* is crucial**. To make a point that is at all meaningful, statistical presentations must refer to differences between observation and expectation, or differences among observations."

"**The average life expectancy of famous orchestral conductors is 73.4 years**."

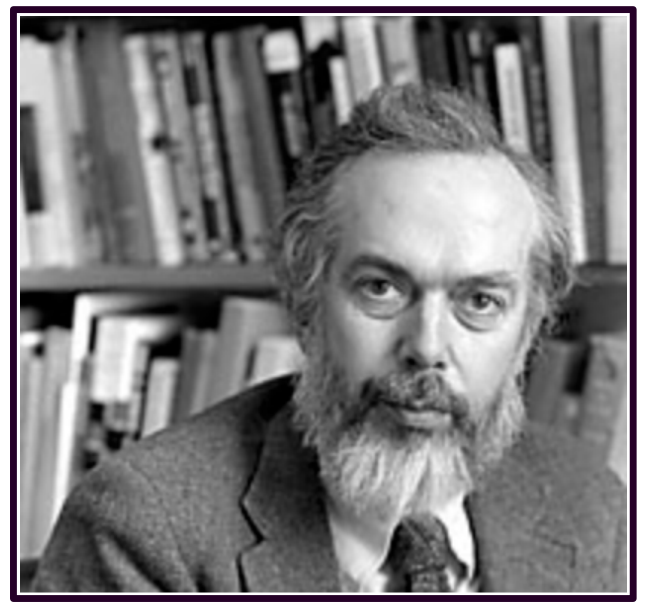
Why is this important?
How unusual is this?

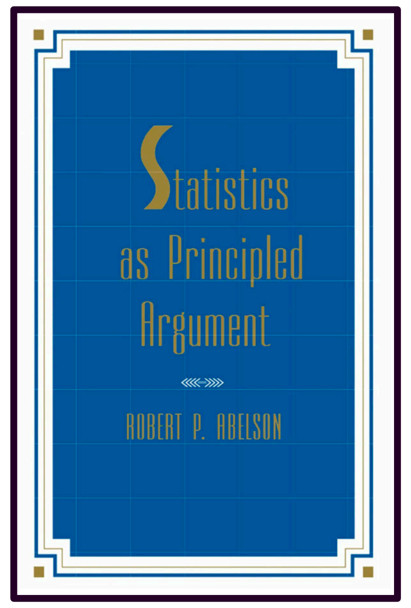
Consider standards of comparison

Should we compare with orchestra *players*? With *non-famous* conductors, with the *public*? With other *males* in the United States, whose average life expectancy was 68.5 at the time of the study?

With other males who have already reached the age of 32, the average age of appointment to a first conducting post, almost all of whom are male? This group's average life expectancy was 72.0.

Abelson, Robert P. *Statistics as Principled Argument*. Psychology Press, 1995. Print.





Statistics as principled argument

Abelson

Educated at MIT and Princeton, the late professor of psychology and political science taught at Yale 42 years, consulted for NBC, and was an analyst for three presidential campaigns.

Elements of statistical persuasion

Several properties of data, and its analysis and presentation, govern its persuasive force.

Magnitude of effects

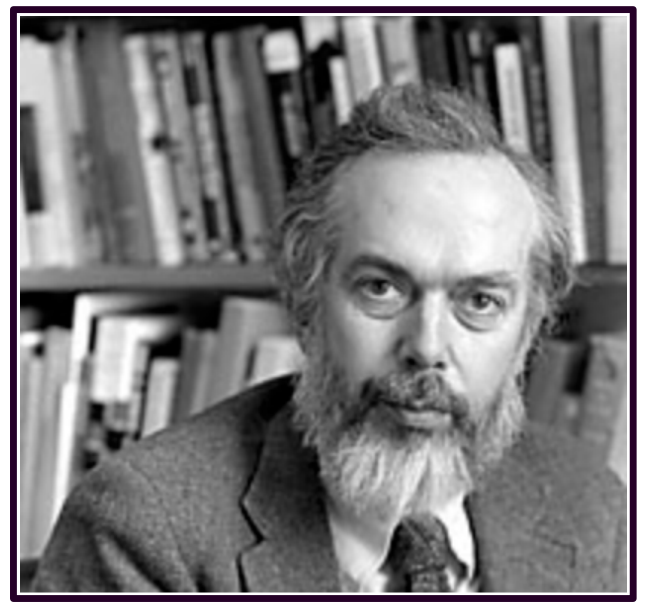
Articulation of results

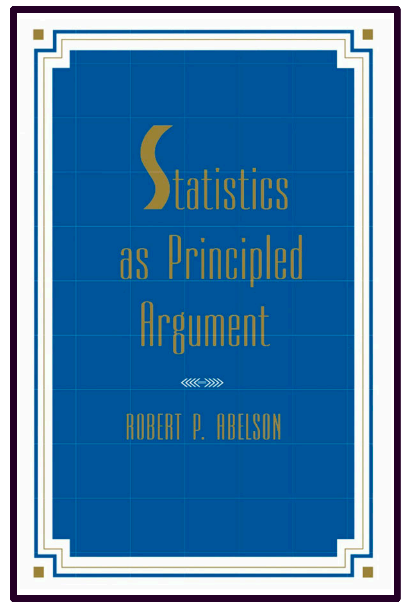
Generality of effects

Interestingness of argument

Credibility of argument

Abelson, Robert P. *Statistics as Principled Argument*. Psychology Press, 1995. Print.





Statistics as principled argument

Abelson

Educated at MIT and Princeton, the late professor of psychology and political science taught at Yale 42 years, consulted for NBC, and was an analyst for three presidential campaigns.

M | **Magnitude of effects**. The strength of a statistical argument is enhanced in accord with the quantitative magnitude of support for its qualitative claim. Consider describing effect sizes like the difference between means, not dichotomous tests.

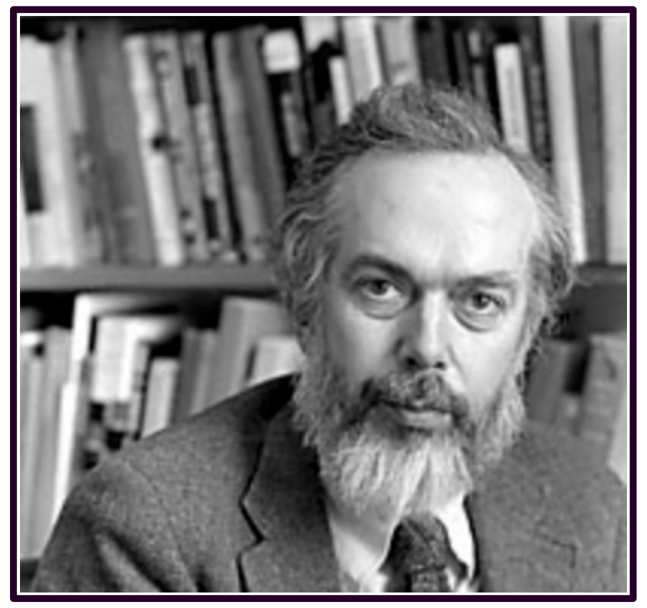
A | **Articulation of results**. The degree of comprehensible detail in which conclusions are phrased. This is a form of specificity. We want to honestly describe and frame our results to maximize clarity (minimizing exceptions or limitations to the result) and parsimony (focusing on consistent, connected claims).

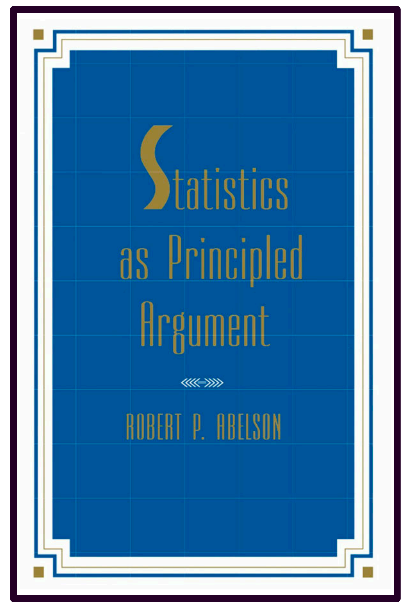
G | **Generality of effects**. This is the breadth of applicability of the conclusions. Over what context can the results be replicated?

I | **Interestingness of argument**. For a statistical story to be theoretically interesting, it must have the potential, through empirical analysis, to change what people believe about an important issue.

C | **Credibility of argument**. Refers to the believability of a research claim, and requires both methodological soundness and theoretical coherence.

Abelson, Robert P. *Statistics as Principled Argument*. Psychology Press, 1995. Print.





Statistics as principled argument

Abelson

Educated at MIT and Princeton, the late professor of psychology and political science taught at Yale 42 years, consulted for NBC, and was an analyst for three presidential campaigns.

MAGIC

Magnitude of effects. The strength of a statistical argument is enhanced in accord with the quantitative magnitude of support for its qualitative claim. **Consider describing effect sizes like the difference between means, not dichotomous tests.**

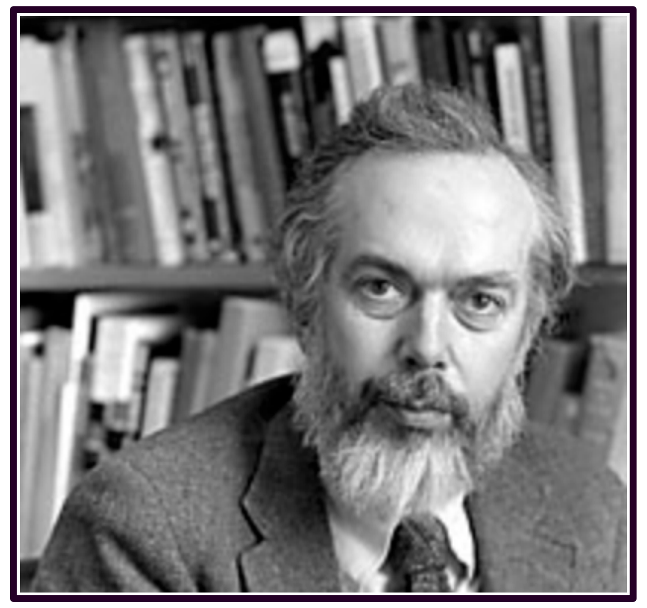
p-values say little, can mislead

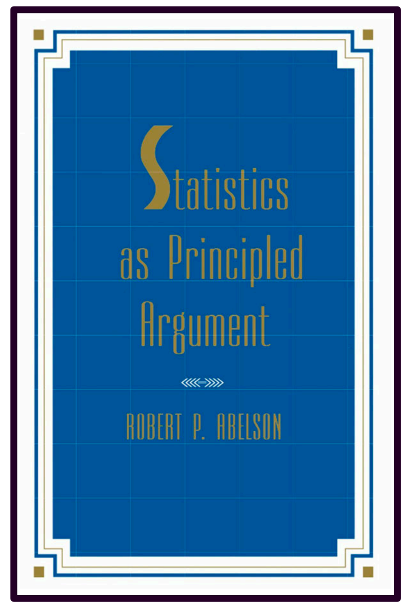
The information yield from null hypothesis tests is ordinarily quite modest, because **all one carries away is a possibly misleading accept-reject decision**.

```
> y <- rnorm(n = 100000, mean = 0, sd = 1)
> x <- rnorm(n = 100000, mean = 0, sd = 1)
> model_fit <- lm(y ~ x)
> summary(model_fit)

Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-4.6381 -0.6755  0.0064  0.6705  4.0234

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.001911   0.003157   0.606   0.5448
x           -0.008707   0.003149  -2.765   0.0057 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9982 on 99998 degrees of freedom
Multiple R-squared:  7.643e-05,	Adjusted R-squared:  6.643e-05
F-statistic: 7.644 on 1 and 99998 DF,  p-value: 0.005699
```

p-value < 0.01
Stars, significant !?

Abelson, Robert P. *Statistics as Principled Argument*. Psychology Press, 1995. Print.





Statistics as principled argument

Abelson

Educated at MIT and Princeton, the late professor of psychology and political science taught at Yale 42 years, consulted for NBC, and was an analyst for three presidential campaigns.

A **p-value** less than 0.01 is **not**:

Having observed the data, the probability that the null hypothesis is true is less than one in a hundred.

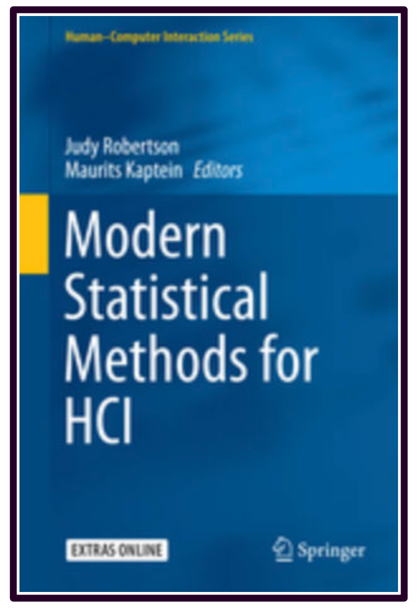
Instead, it means: $P(D \mid H)$

If it were true that there were no systematic difference between the means in the populations from which the samples came, **then** the probability that the observed means would have been as different as they were, or more different, is less than one in a hundred.

This being strong grounds for doubting the viability of the null hypothesis, the null hypothesis is rejected.

$$P(H \mid D) = \frac{P(D \mid H)\, P(H)}{P(D \mid H)P(H) + P(D \mid \neg H)P(\neg H)}$$

Abelson, Robert P. *Statistics as Principled Argument*. Psychology Press, 1995. Print.



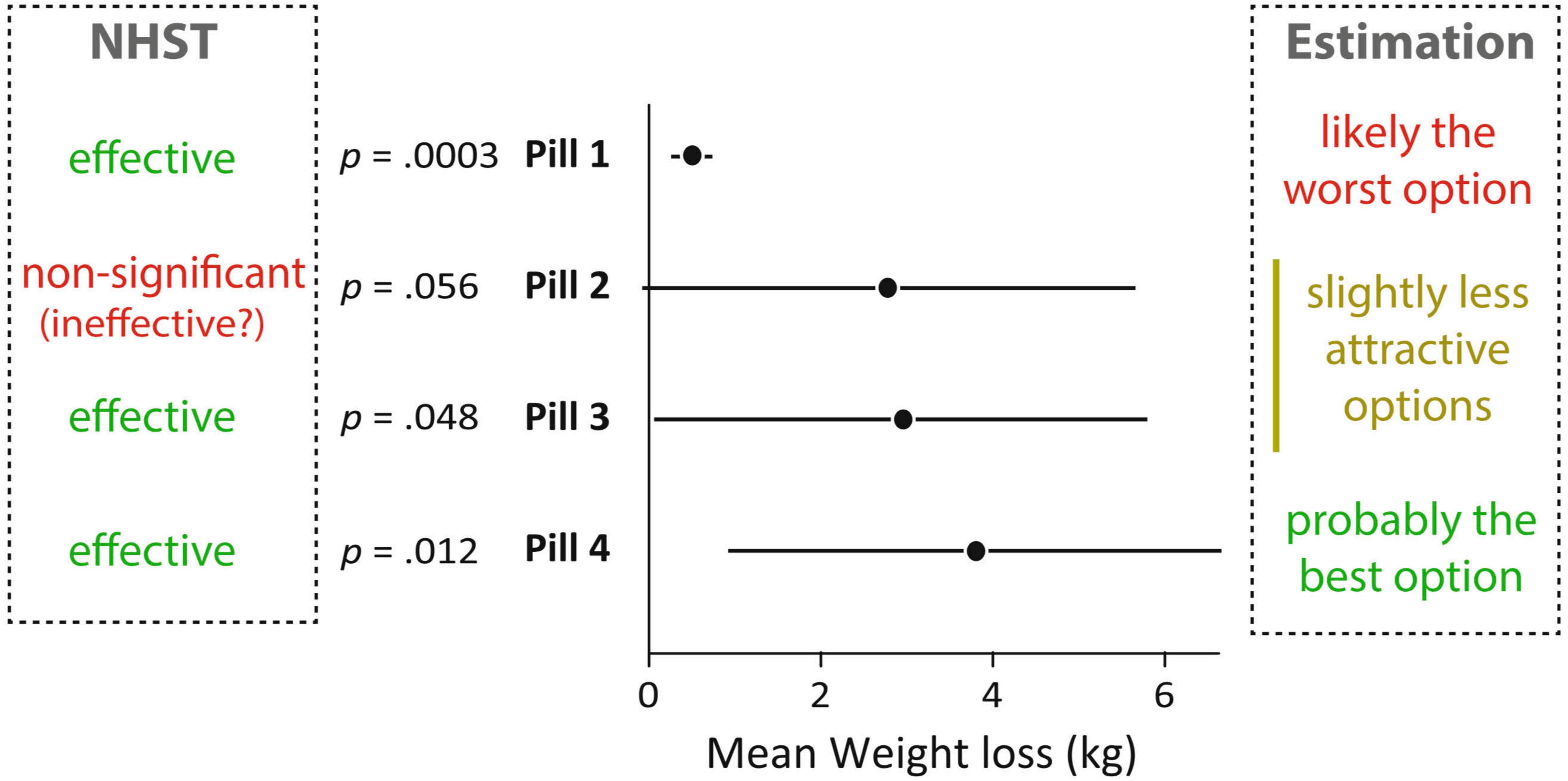
Ch. 13, Fair statistical communication in HCI

Dragicevic

He is a researcher, focusing on psychology of data visualization for judgment and decision making, and on transparent statistical communication.

He agrees with Abelson*, decisions are informed by comparing effect sizes and intervals

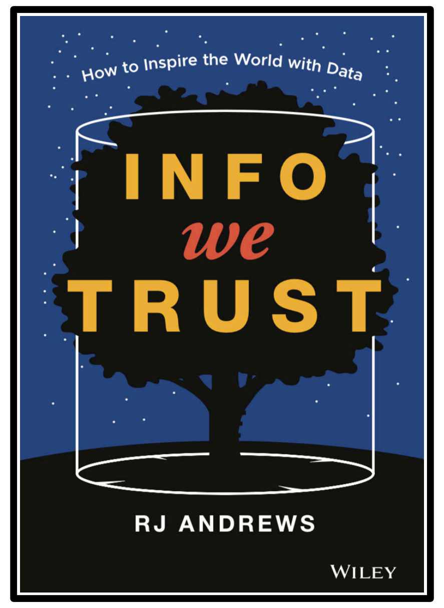
Whether exploring or confirming analyses, show results using an *estimation* approach — use graphs to show **effect sizes** and **interval estimates**, and offer **nuanced interpretations** of results. **Avoid the pitfalls of dichotomous tests and p-values**.



> **The notion of binary significance testing is a terrible idea for those who want to achieve fair statistical communication.**

Dragicevic, Pierre. "Fair Statistical Communication in HCI." *Modern Statistical Methods for HCI*. Ed. Judy Robertson and Maurits Kaptein. Springer International Publishing, 2016. 291–330. Print.

*The two are in agreement on this point, but disagree in that Pierre believes p-values should never be used while Abelson believes they should be de-emphasized but reported with everything else.



Info We Trust
How to inspire the world with data
Andrews

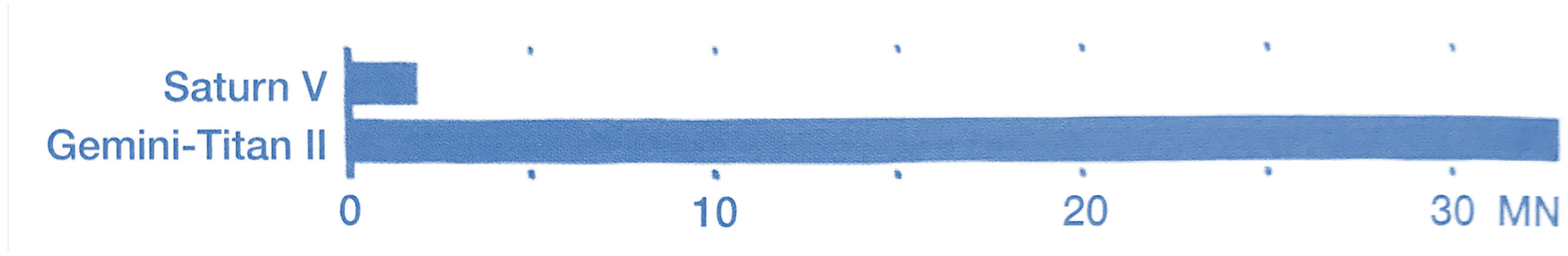
He is a data storyteller. He book is an adventure exploring how to inspire the world with data. RJ is the creator of www.infowetrust.com, where he makes available some of his data stories.

Language for comparing quantities

In language describing quantities, we have two main ways to compare. One form is additive or subtractive. The other is multiplicative. We perceive or process these comparisons differently.

> “The Apollo program crew had **one more** astronaut than Project Gemini. Apollo’s Saturn V rocket had about **seventeen times more** thrust than the Gemini-Titan II.”

“Seventeen times more”
“1,700 percent more”
“33 versus 1.9”



Additive comparisons

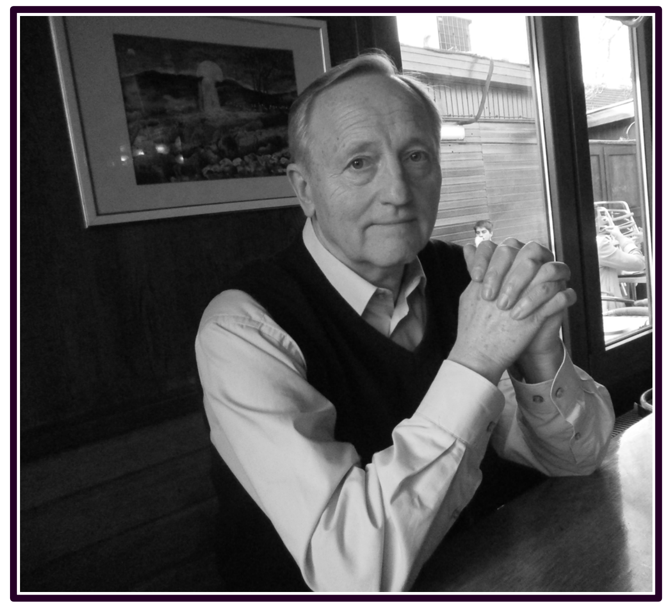
Add and subtract comparisons are easier for people to understand, especially with small numbers.

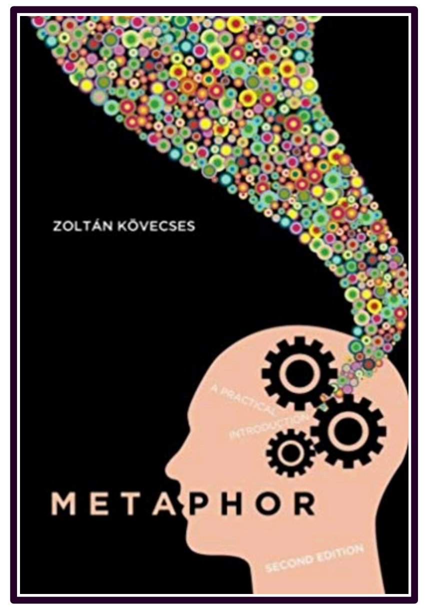
Multiplicative comparisons

Relative to additive comparisons, multiplying or dividing are more difficult. This includes comparisons expressed as ratios: a few times more, a few times less. People generally try to interpret multiplying operations through pooling, or repeat addition.

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

Comparison through metaphor, simile, analogy





Metaphor: a practical introduction

Kövecses

He is professor of linguistics at Eötvös Loránd University, Budapest. He researches language and conceptualization of emotions, cross-cultural variation in metaphor, and the issue of the relationship between language, mind, and culture.

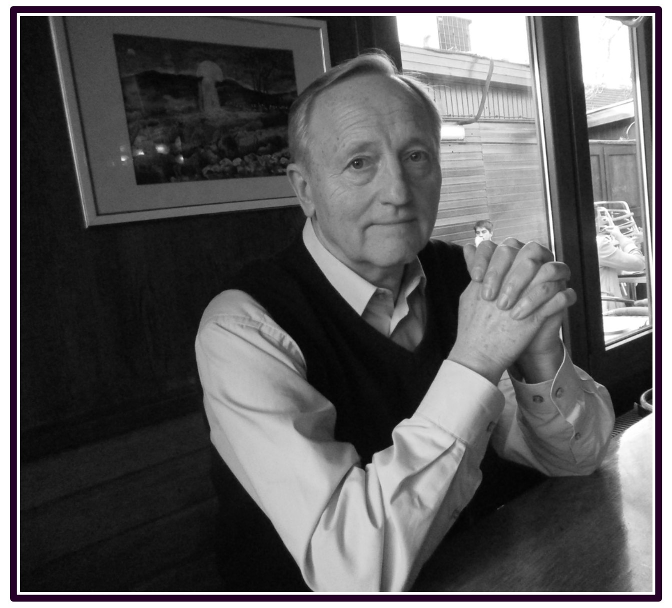
Mapping | **Source Domain > Target Domain**

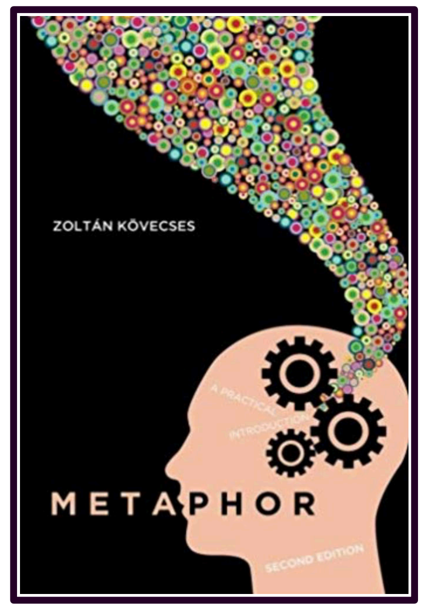
Target domains | The abstract concepts we need help explaining, ideas we need to make important, or the multiple ideas we need to link.

Common source domains |

- Human body
- Animals
- Plants
- Buildings and constructions
- Machines and tools
- Games and Sport
- Money
- Cooking and food
- Heat and cold
- Light and darkness
- Movement and direction

Kövecses, Z.(2010). *Metaphor: A Practical Introduction*. Oxford University Press, second edition.





Metaphor: a practical introduction

Kövecses

He is professor of linguistics at Eötvös Loránd University, Budapest. He researches language and conceptualization of emotions, cross-cultural variation in metaphor, and the issue of the relationship between language, mind, and culture.

Metaphor adds to persuasiveness by **reforming abstract concepts into something more familiar** to our senses, **signaling** particular aspects of **importance**, **memorializing the concept**, or providing **coherence throughout a writing**.

Kövecses, Z.(2010). *Metaphor: A Practical Introduction*. Oxford University Press, second edition.

The Next Rembrandt

ING, et al.

Three examples from their data narrative

To bring him back, we distilled the artistic DNA from his work and used it to create The Next Rembrandt.

* * *

To create new artwork using data from Rembrandt's paintings, we had to maximize the data pool from which to pull information.

* * *

We created a height map using two different algorithms that found texture patterns of canvas surfaces and layers of paint. That information was transformed into height data, allowing us to mimic the brushstrokes used by Rembrandt.

ING et al. "*The Next Rembrandt.*" 5 Apr. 2016. Web. https://www.nextrembrandt.com

The Next Rembrandt

ING, et al.

Three examples from their data narrative

To bring him back, we distilled the **artistic DNA** from his work and used it to create The Next Rembrandt.

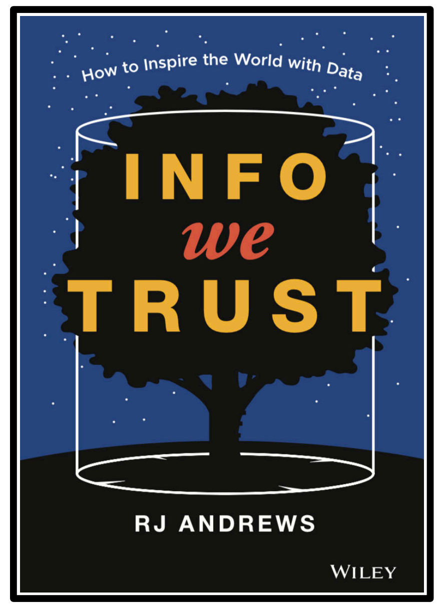
* * *

To create new artwork using data from Rembrandt's paintings, we had to maximize the **data pool** from which to pull information.

* * *

We created a **height map** using two different algorithms that found texture patterns of canvas surfaces and layers of paint. That information was transformed into height data, allowing us to mimic the brushstrokes used by Rembrandt.

ING et al. "*The Next Rembrandt*." 5 Apr. 2016. Web. https://www.nextrembrandt.com



Info We Trust

Andrews

He is a data storyteller. He book is an adventure exploring how to inspire the world with data. RJ is the creator of www.infowetrust.com, where he makes available some of his data stories.

Setting up the metaphor:

data to Information is a record to music

How do we think about the albums we love? A lonely microphone in a smoky recording studio? A needle's press into hot wax? A rotating can of magnetic tape? A button that clicks before the first note drops? No!

The mechanical ephemera of music's recording, storage, and playback may cue nostalgia, but they are not where the magic lies. **The magic is in the music. The magic is in the information** that the apparatuses capture, preserve, and make accessible. **It is the same with all information**.

Refers back:

When you envision data, do not get stuck in encoding and storage. Instead, **try to see the music**.

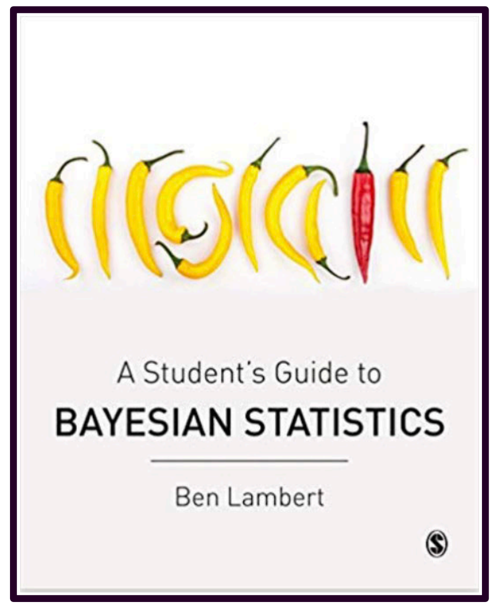
Refers back:

Looking at tables of any substantial size is a little like looking at the grooves of a record with a magnifying glass. **You can see the data but you will not hear the music**.

Refers back:

Then, we can **see data for what it is, whispers from a past world waiting for its music to be heard** again.

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



A Student's Guide to Bayesian Statistics

Lambert

Ben is a researcher at Imperial College London, has worked in applied statistical inference for about a decade, formerly at the University of Oxford, and is the author of online lectures on econometrics and statistics.

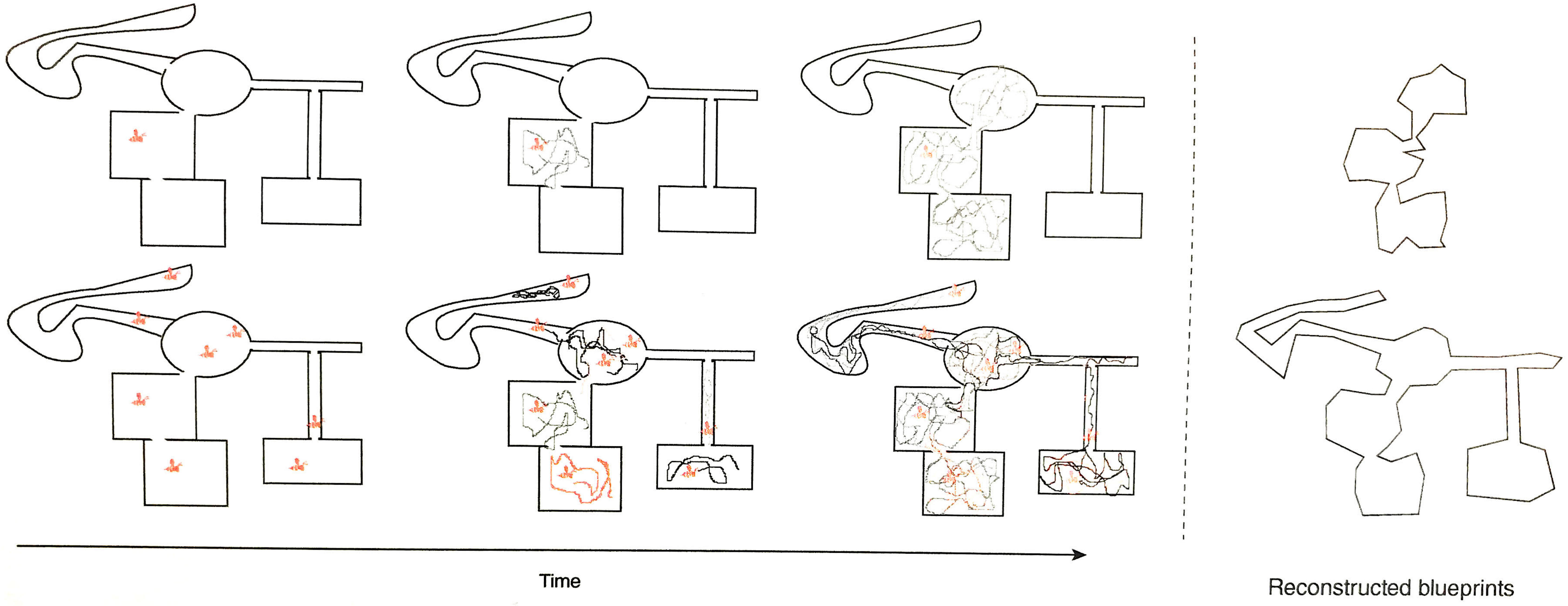


Figure 13.17 Using Bob's bees to reproduce blueprints of a house. The three left columns show the effect of using a single bee (top) and multiple bees (bottom) to randomly traverse the house. Right: blueprints reconstructed from the paths of the bees.

"The importance of using multiple bees (chains) to judge MCMC convergence."

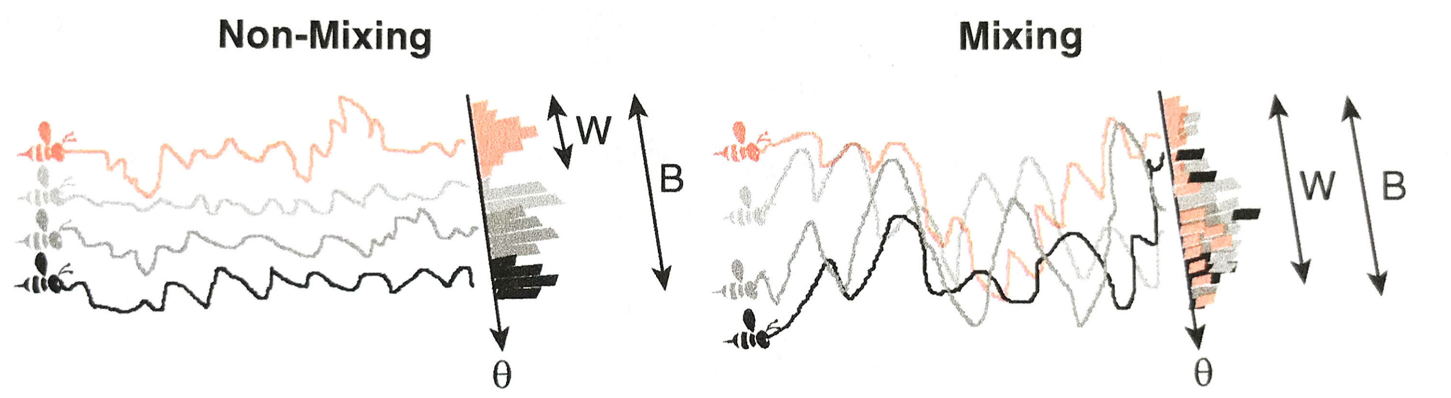
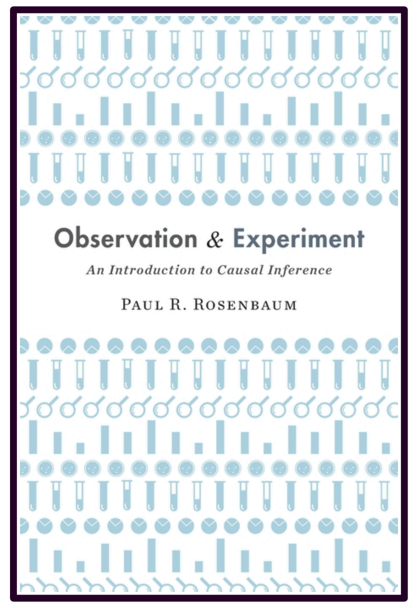


Figure 13.19 The within-path variation (*W*) versus the between-path variation (*B*) for the case of non-mixing (left) and mixing (right) paths.

Ben Lambert. *A Student's Guide to Bayesian Statistics*. Sage Publications, 2018. Print.



Observation & Experiment

Rosenbaum

He is Professor of Statistics at the Wharton School and a Senior Fellow of the Leonard Davis Institute of Health Economics, University of Pennsylvania. His book epitomizes the idea that "the most important ideas in statistics can be clearly explained in plain English, with little or no math."

Setting up the analogy:

Traveling one of two roads to distinguish covariate from outcome

The distinction between covariates and outcomes is basic in causal inference. Failure to distinguish covariates and outcomes quickly leads to confusion, errors, and false conclusions. … Perhaps the distinction between covariate and outcome is most vivid, most palpable, in **Robert Frost's poem** "**The Road Not Taken**" (1916):

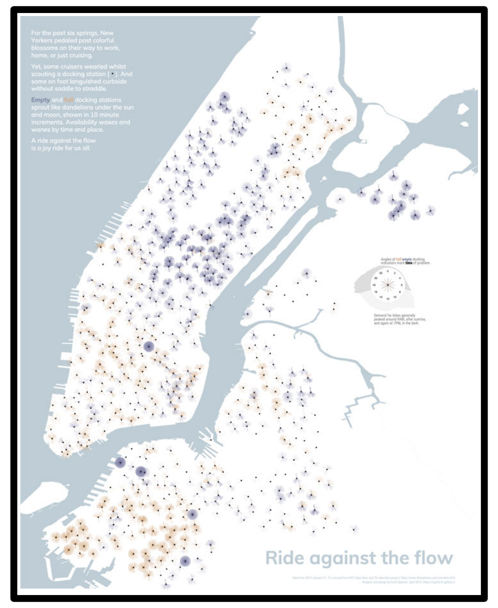
> **Two roads diverged in a yellow wood**
> And sorry I **could not travel both**
> And be one traveler, long I stood
> And looked down one as far as I could
> To where it bent in the undergrowth

Frost creates the mood attending a decision, one whose full consequences we cannot see or anticipate: "Knowing how way leads on to way," we will not see the road not taken. As it was for **Frost in a yellow wood**, so it is for a patient at risk of death in the ProCESS Trial, and so it will be in every causal question.

Refers back:

If this were true, then the two potential outcomes would be equal for each patient. If this were true in **Frost's poem**, the two paths would always end in the same place.

Rosenbaum, P. (2017). *Observation and Experiment: An Introduction to Causal Inference*. Harvard University Press.



Longlisted, Information is Beautiful Awards

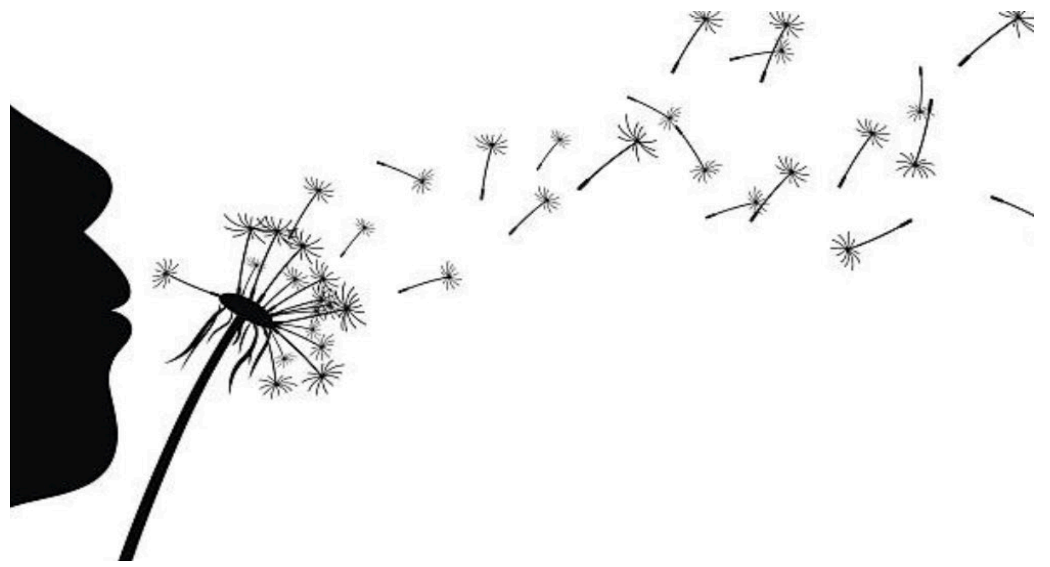
Spencer

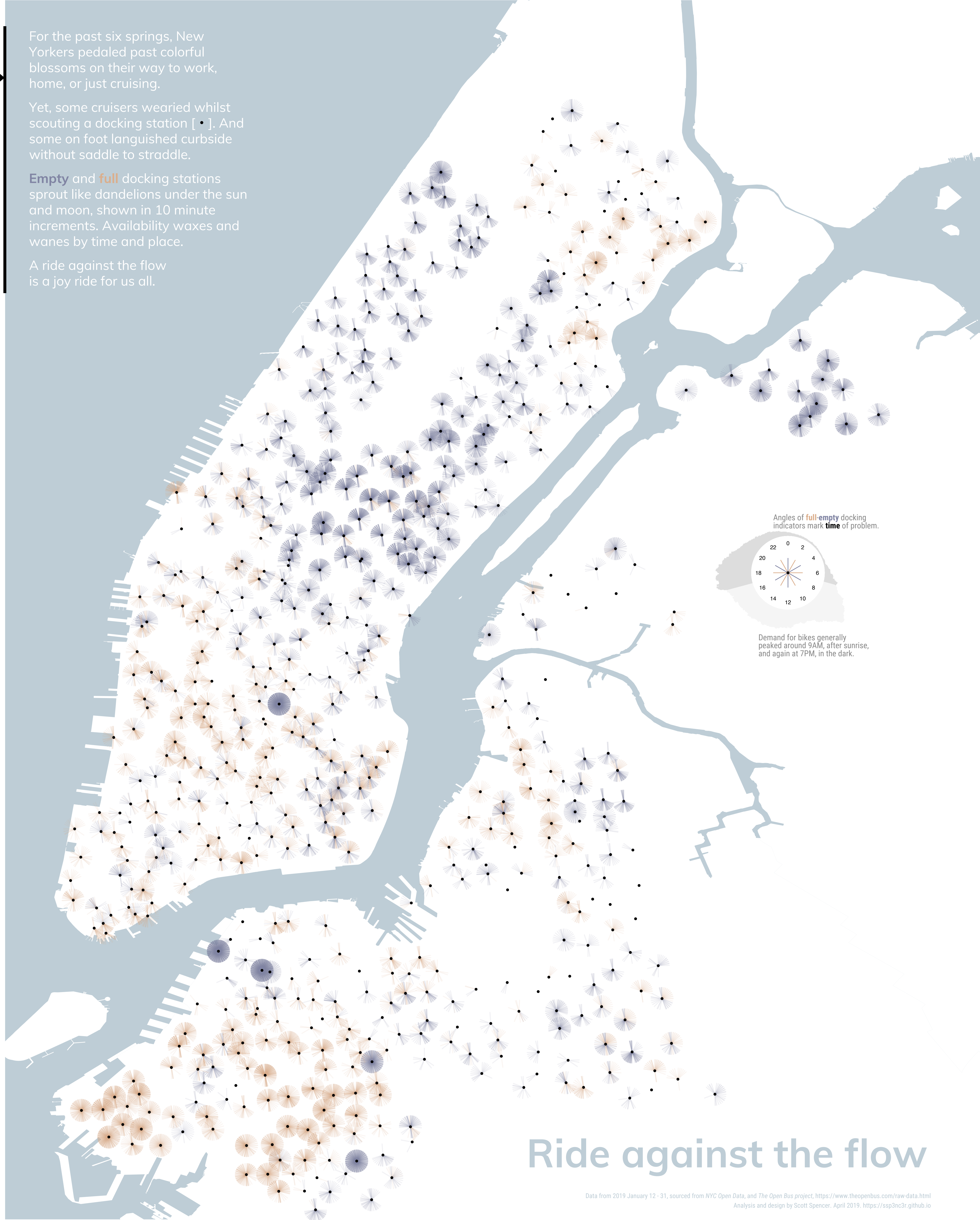
For the past six springs, New Yorkers pedaled past colorful blossoms on their way to work, home, or just cruising.

Yet, some cruisers wearied whilst scouting a docking station [•]. And some on foot languished curbside without saddle to straddle.

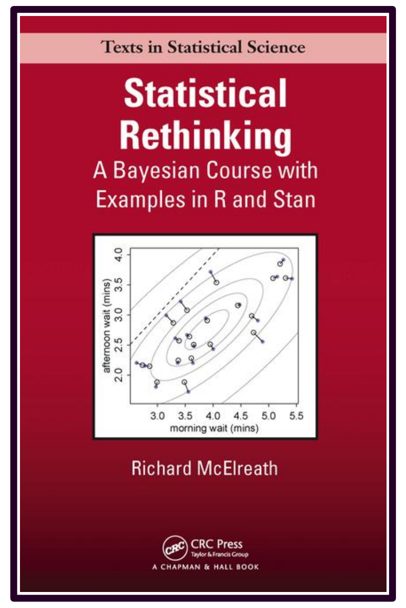
Empty and full docking stations sprout like dandelions under the sun and moon, shown in 10 minute increments. Availability waxes and wanes by time and place.

A ride against the flow is a joy ride for us all.





Spencer, S. *Ride Against the Flow*. Information is Beautiful Awards 2019. Showcase. https://www.informationisbeautifulawards.com/showcase/4367-ride-against-the-flow



Statistical Rethinking

McElreath

He is the director of the Department of Human Behavior, Ecology, and Culture at the Max Planck Institute for Evolutionary Anthropology. He is also a professor in the Department of Anthropology at the University of California, Davis.

Setting up the metaphor:

Mythical clay robot (golem) as a statistical model

A **golem (goh-lem) is a clay robot** from Jewish folklore, constructed from dust and fire and water. It is brought to life by inscribing *emet*, Hebrew for "truth," on its brow. Animated by truth, but lacking free will, a golem always does exactly what it is told. … However, its obedience also brings danger, as careless instructions or unexpected events can turn a golem against its makers. Its abundance of power is matched by its lack of wisdom.

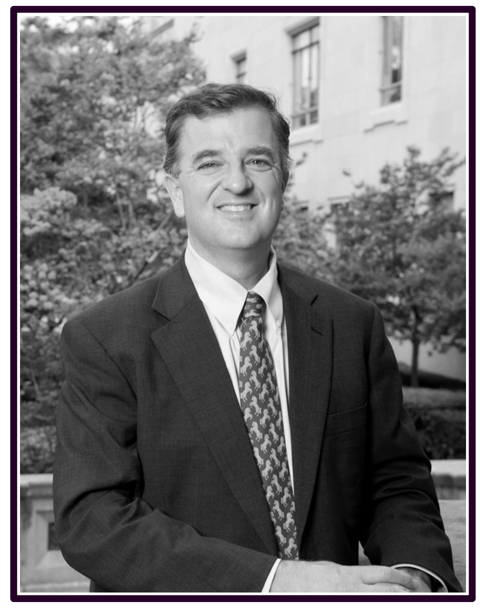
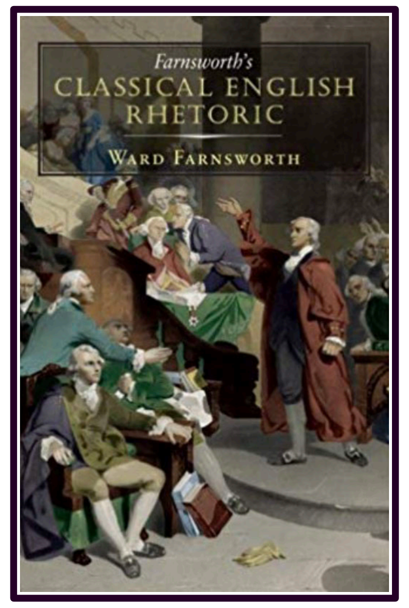
Using secret techniques from the Kabbalah, Rabbi Judah was able to build a golem, animate it with "truth," and order it to defend the Jewish people of Prague. Ultimately Judah was forced to destroy the golem, as its combination of extraordinary power with clumsiness eventually led to innocent deaths.

Describes models as golems, then refers back 100 times in book

Our golems rarely have physical form, but they too are often made of clay, living in silicon as computer code. These golems are scientific models. Just like a golem, scientific models are neither true nor false. Rather they are constructs engineered for some purpose. These constructs are incredibly powerful, dutifully conducting their programmed calculations.

McElreath, Richard. *Statistical Rethinking*. Second. Second. 2019. Print.

Patterns that compare, organize, grab attention

Classical English Rhetoric

Farnsworth

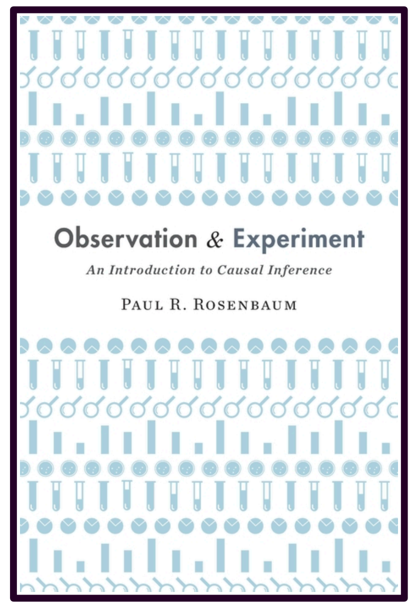
He is dean and professor of the University of Texas School of Law. Before teaching, he graduated from University of Chicago Law School, clerked for Supreme Court Justice Kennedy, and served as advisor to an international tribunal in the Hague.

Use patterns to compare, grab attention, add emphasis

We can use patterns to “make the words they arrange more emphatic or memorable or otherwise effective.” These patterns can be the most effective and efficient ways to show comparisons and contrasts.

Farnsworth, W. (2011). *Farnsworth’s Classical English Rhetoric*. David R. Godine Publisher.

Example: Reversal of structure, repetition at the end



Observation & Experiment

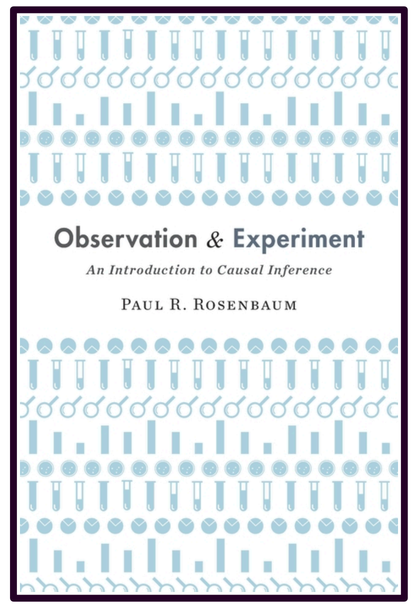
Rosenbaum

He is Professor of Statistics at the Wharton School and a Senior Fellow of the Leonard Davis Institute of Health Economics, University of Pennsylvania. His book epitomizes the idea that "the most important ideas in statistics can be clearly explained in plain English, with little or no math."

> A covariate is a quantity determined prior to treatment assignment. In the Pro-CESS Trial, the age of the patient at the time of admission to the emergency room was a covariate. The gender of the patient was a covariate. Whether the patient was admitted from a nursing home was a covariate.

* Example from Rosenbaum, P. (2017). *Observation and Experiment: An Introduction to Causal Inference.* Harvard University Press.

Example: Reversal of structure, repetition at the end



Observation & Experiment

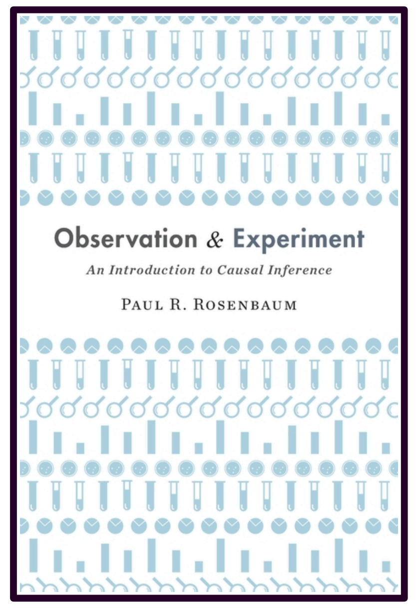
Rosenbaum

He is Professor of Statistics at the Wharton School and a Senior Fellow of the Leonard Davis Institute of Health Economics, University of Pennsylvania. His book epitomizes the idea that "the most important ideas in statistics can be clearly explained in plain English, with little or no math."

> "**A covariate is** a quantity determined prior to treatment assignment. In the Pro-CESS Trial, the age of the patient at the time of admission to the emergency room **was a covariate**. The gender of the patient **was a covariate**. Whether the patient was admitted from a nursing home **was a covariate**.

* Example from Rosenbaum, P. (2017). *Observation and Experiment: An Introduction to Causal Inference*. Harvard University Press.

Example: Repetition at the start, parallel structure



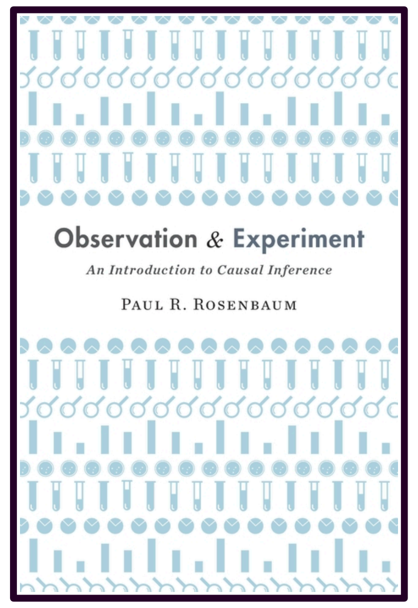
Observation & Experiment

Rosenbaum

He is Professor of Statistics at the Wharton School and a Senior Fellow of the Leonard Davis Institute of Health Economics, University of Pennsylvania. His book epitomizes the idea that “the most important ideas in statistics can be clearly explained in plain English, with little or no math.”

> One might hope that panel (a) of Figure 7.3 is analogous to a simple randomized experiment in which one child in each of 33 matched pairs was picked at random for exposure. One might hope that panel (b) of Figure 7.3 is analogous to a different simple randomized experiment in which levels of exposure were assigned to pairs at random. One might hope that panels (a) and (b) are jointly analogous to a randomized experiment in which both randomizations were done, within and among pairs. All three of these hopes may fail to be realized: there might be bias in treatment assignment within pairs or bias in assignment of levels of exposure to pairs.

* Example from Rosenbaum, P. (2017). *Observation and Experiment: An Introduction to Causal Inference.* Harvard University Press.



Observation & Experiment

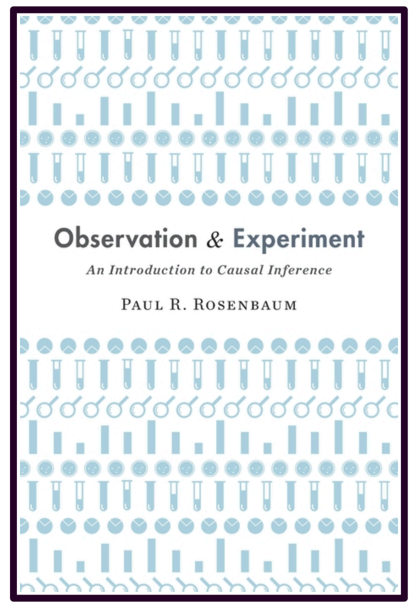
Rosenbaum

He is Professor of Statistics at the Wharton School and a Senior Fellow of the Leonard Davis Institute of Health Economics, University of Pennsylvania. His book epitomizes the idea that "the most important ideas in statistics can be clearly explained in plain English, with little or no math."

Example: Repetition at the start, parallel structure

> **One might hope** that panel (a) of Figure 7.3 is analogous to a simple randomized experiment in which one child in each of 33 matched pairs was picked at random for exposure. **One might hope** that panel (b) of Figure 7.3 is analogous to a different simple randomized experiment in which levels of exposure were assigned to pairs at random. **One might hope** that panels (a) and (b) are jointly analogous to a randomized experiment in which both randomizations were done, within and among pairs. **All three of these hopes may fail** to be realized: there might be bias in treatment assignment within pairs or bias in assignment of levels of exposure to pairs.

* Example from Rosenbaum, P. (2017). *Observation and Experiment: An Introduction to Causal Inference*. Harvard University Press.



Observation & Experiment

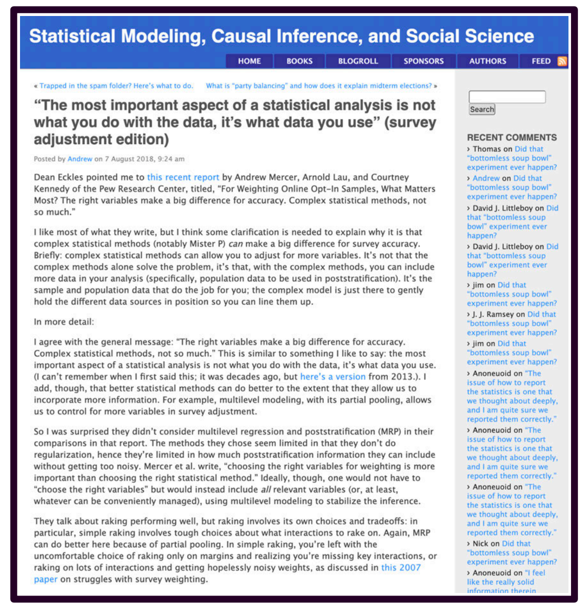
Rosenbaum

He is Professor of Statistics at the Wharton School and a Senior Fellow of the Leonard Davis Institute of Health Economics, University of Pennsylvania. His book epitomizes the idea that "the most important ideas in statistics can be clearly explained in plain English, with little or no math."

Example: Asking questions and answering them

> Where did Fisher's null distribution come from? From the coin in Fisher's hand.

* Example from Rosenbaum, P. (2017). *Observation and Experiment: An Introduction to Causal Inference.* Harvard University Press.



Statistical Modeling, Causal Inference, and Social Science

HOME BOOKS BLOGROLL SPONSORS AUTHORS FEED

« Trapped in the spam folder? Here's what to do. What is "party balancing" and how does it explain midterm elections? »

"The most important aspect of a statistical analysis is not what you do with the data, it's what data you use" (survey adjustment edition)

Posted by Andrew on 7 August 2018, 9:24 am

Dean Eckles pointed me to this recent report by Andrew Mercer, Arnold Lau, and Courtney Kennedy of the Pew Research Center, titled, "For Weighting Online Opt-In Samples, What Matters Most? The right variables make a big difference for accuracy. Complex statistical methods, not so much."

I like most of what they write, but I think some clarification is needed to explain why it is that complex statistical methods (notably Mister P) *can* make a big difference for survey accuracy. Briefly: complex statistical methods can allow you to adjust for more variables. It's not that the complex methods alone solve the problem, it's that, with the complex methods, you can include more data in your analysis (specifically, population data to be used in poststratification). It's the sample and population data that do the job for you; the complex model is just there to gently hold the different data sources in position so you can line them up.

In more detail:

I agree with the general message: "The right variables make a big difference for accuracy. Complex statistical methods, not so much." This is similar to something I like to say: the most important aspect of a statistical analysis is not what you do with the data, it's what data you use. (I can't remember when I first said this; it was decades ago, but here's a version from 2013.). I add, though, that better statistical methods can do better to the extent that they allow us to incorporate more information. For example, multilevel modeling, with its partial pooling, allows us to control for more variables in survey adjustment.

So I was surprised they didn't consider multilevel regression and poststratification (MRP) in their comparisons in that report. The methods they chose seem limited in that they don't do regularization, hence they're limited in how much poststratification information they can include without getting too noisy. Mercer et al. write, "choosing the right variables for weighting is more important than choosing the right statistical method." Ideally, though, one would not have to "choose the right variables" but would instead include *all* relevant variables (or, at least, whatever can be conveniently managed), using multilevel modeling to stabilize the inference.

They talk about raking performing well, but raking involves its own choices and tradeoffs: in particular, simple raking involves tough choices about what interactions to rake on. Again, MRP can do better here because of partial pooling. In simple raking, you're left with the uncomfortable choice of raking only on margins and realizing you're missing key interactions, or raking on lots of interactions and getting hopelessly noisy weights, as discussed in this 2007 paper on struggles with survey weighting.

Search

RECENT COMMENTS

- Thomas on Did that "bottomless soup bowl" experiment ever happen?
- Andrew on Did that "bottomless soup bowl" experiment ever happen?
- David J. Littleboy on Did that "bottomless soup bowl" experiment ever happen?
- David J. Littleboy on Did that "bottomless soup bowl" experiment ever happen?
- jim on Did that "bottomless soup bowl" experiment ever happen?
- J. J. Ramsey on Did that "bottomless soup bowl" experiment ever happen?
- jim on Did that "bottomless soup bowl" experiment ever happen?
- Anoneuoid on "The issue of how to report the statistics is one that we thought about deeply, and I am quite sure we reported them correctly."
- Anoneuoid on "The issue of how to report the statistics is one that we thought about deeply, and I am quite sure we reported them correctly."
- Anoneuoid on "The issue of how to report the statistics is one that we thought about deeply, and I am quite sure we reported them correctly."
- Nick on Did that "bottomless soup bowl" experiment ever happen?
- Anoneuoid on "I feel like the really solid information therein

Statistical Modeling, Causal Inference, and Social Science

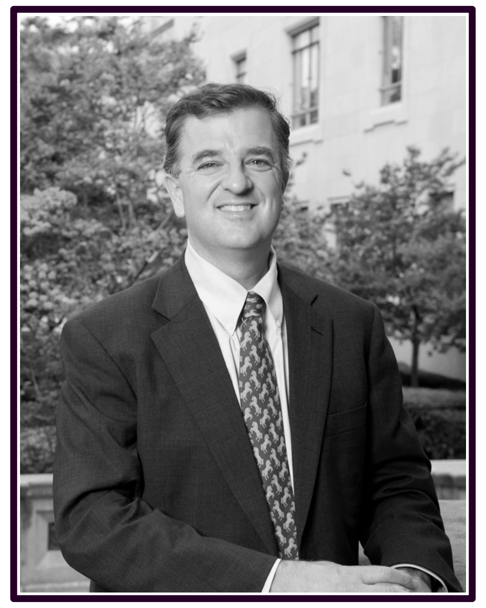
Gelman

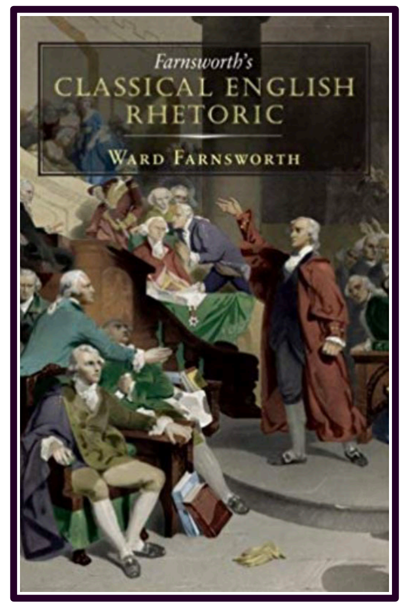
Professor of Statistics and Political Science at Columbia University, he is known widely for his work in Bayesian statistics, and has authored several textbooks, including Teaching Statistics, and Bayesian Data Analysis.

Example: Inversion of words

> The most important aspect of a statistical analysis is not what **you** do with the ***data***, it's what ***data*** **you** use.

Gelman, Andrew. Blog post: https://statmodeling.stat.columbia.edu/2018/08/07/important-aspect-statistical-analysis-not-data-data-use-survey-adjustment-edition/





Classical English Rhetoric

Farnsworth

He is dean and professor of the University of Texas School of Law. Before teaching, he graduated from University of Chicago Law School, clerked for Supreme Court Justice Kennedy, and served as advisor to an international tribunal in the Hague.

Repetition of words & phrases

- simple repetition (*epizeuxis, epimone*)
- repetition at the start (*anaphora*)
- repetition at the end (*epistrophe*)
- repetition at the start and end (*symploce*)
- repeating the ending at the beginning (*anadiplosis*)
- repetition of the root (*polyptoton*)

Structural matters

- parallel structure (*isocolon*)
- reversal of structure (*chiasmus*)
- inversion of words (*anastrophe*)
- leaving out words (*ellipsis*)

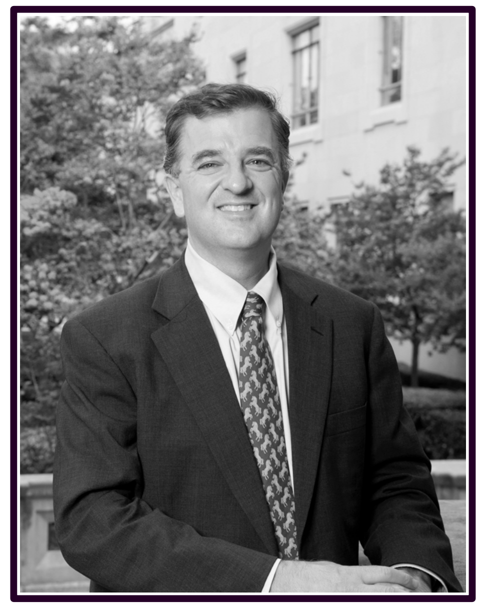
Dramatic devices

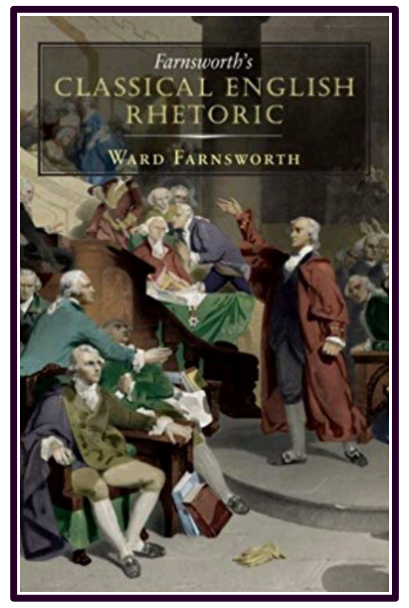
- saying things by not saying them (*præteritio*)
- correcting oneself (*metanoia*)
- rhetorical uses of the negative (*litotes*)
- rhetorical questions (*erotema*)
- asking questions and answering them (*hypophora*)
- anticipating objections and meeting them (*prolepsis*)

† Extensive instruction in Farnsworth, W. (2011). *Farnsworth's Classical English Rhetoric*. David R. Godine Publisher.

How unexpected patterns work

Unexpected word placement calls attention to them, creates emphasis by coming earlier than expected or violating the reader's expectations. Note that, to violate expectations necessarily means reserving a technique like inversion for just the point to be made, lest the reader come to expect it — **more is less, less is more**. Secondly, it can create an attractive rhythm. Thirdly, when the words that bring full meaning come later, it can add suspense, and finish more climactic.





Classical English Rhetoric

Farnsworth

He is dean and professor of the University of Texas School of Law. Before teaching, he graduated from University of Chicago Law School, clerked for Supreme Court Justice Kennedy, and served as advisor to an international tribunal in the Hague.

Immersion precedes implementation

Seeing just a few examples invites direct imitation of them, which tends to be clumsy. Immersion in many examples allows them to do their work by way of a subtler process of influence, with a gentler and happier effect on the resulting style.

Farnsworth, W. (2011). *Farnsworth's Classical English Rhetoric*. David R. Godine Publisher.

Wrapping up

For Next Week, Module 5:

Agenda next week

Next deliverable, **final** 750-word (or less) proposal
(More) audience analysis

The minimum

Kahneman, Daniel, Dan Lovallo, and Olivier Sibony. *Before You Make That Big Decision …* Harvard Business Review 89.6 (2011): 50–60. Print.

Read to understand common limitations, biases, and approaches to reasoning and making decisions amid uncertainty.

Dragicevic, Pierre. "Fair Statistical Communication in HCI." *Modern Statistical Methods for HCI*. Springer International Publishing, 2016. 1–40.

Read to consider what may be important in communicating statistical analysis. Also, consider the graphical displays integrated into the writing.

Carr, David J. *What Value Do You Create? Marketing's 3 Types of Value*. 28 Jan. 2019, *A Map of Modern Brand Building*. 7 Nov. 2016. Medium. Web. https://medium.com/@djc1805

Consider the potential relevance of your analytics project for a marketing executive, and how your messaging may change.

Carr, David J. *Data is the new oil: dirty, misunderstood, polluting the world & pulled from all the wrong places*. 7 Jan. 2018. Medium. Web. https://medium.com/@djc1805

Consider David's background, and how a marketing executive thinks about, and communicates, data,

Craft this course for you,

About that metaphor

We just discussed metaphor, and began to consider its value. Do you recall a personal experience where metaphor helped you understand a concept? Alternatively, how might the idea of metaphor apply to data-driven visuals?

Overcoming obstacles?

Have you been able to identify available data and begin preliminary or exploratory analysis? If so, please share how you overcame any obstacle you faced in getting started? If not, what obstacle is slowing you down and what's your strategy to overcome it?

See you next week!

Columbia University. Kees Verdake, “Tight Rope Walker,” 1973-79, Revson Plaza. Illustration by Spencer.