

Storytelling with Data

Module 3: Business writing, audience analysis, visual considerations

Scott Spencer

Faculty and Lecturer
Columbia University

Agenda

Upcoming deliverable

Objectives of business writing

Business writing

Example 250-word memo

Visual component of writing

Group work

Upcoming deliverable

Upcoming deliverable

For your chosen company and case study,
as an imagined member of the analytics team ...

**250-word
memo**

Write a memo to **CAO** about an opportunity to leverage analytics. Consider background context, problem, data, solution, and impact.

**750-word
proposal**

On approval of the memo, write a proposal to **CAO**, detailing the anticipated project.

Storyboard

Present project result in storyboard to **CMO**, using narrative forms, and with comparisons, metaphors and other storytelling concepts.

Infographic

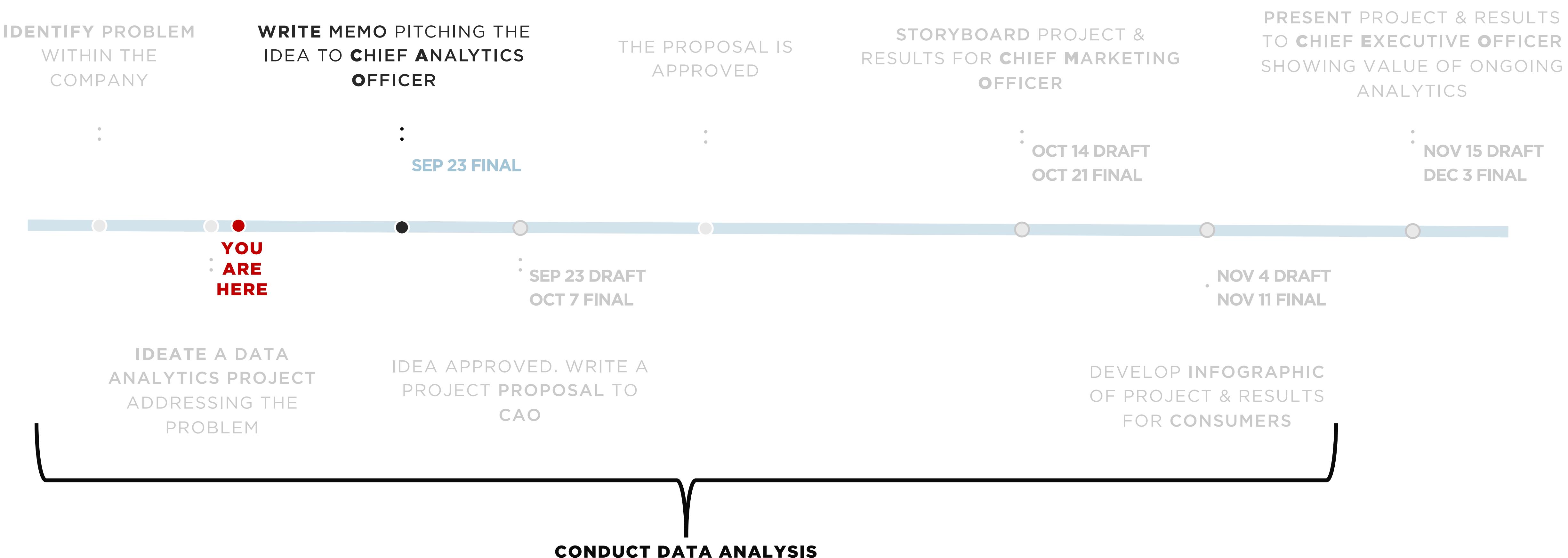
Recraft the results, telling the narrative through an infographic for the **public** or **potential consumers**, using data visualization with brand awareness.

Presentation

Construct and deliver a 4-5 minute persuasive presentation with up to 10 slides to the **CEO**, telling the story of the analytics project to convince them of further investment in analytics.

Upcoming deliverable

250-word memo — Write a memo to **CAO** about an opportunity to leverage analytics. Consider background context, problem, data, solution, and impact. Be sure you have data and a plan for analysis before pitching the idea in the memo.



Objectives of business writing

For **all** business
communications

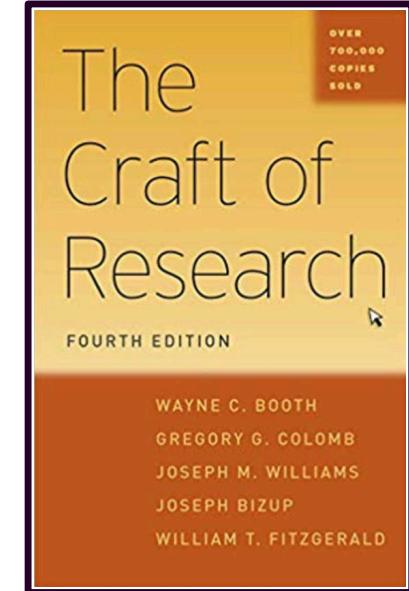
Step into your audiences' shoes to
get them to
pay attention to,
understand,
(be able to) act upon
a maximum of messages, given
constraints.

Who is your audience?

**What's your purpose
for communicating?**

tl;dr

Ideas on writing



Revising style: telling your story clearly

Booth and co-authors

All are university professors of English, and their book is first among Amazon's ranking of books in methodology and statistics.

A few writing principles

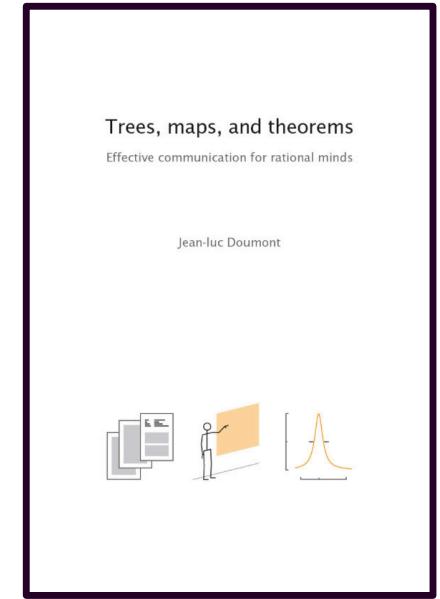
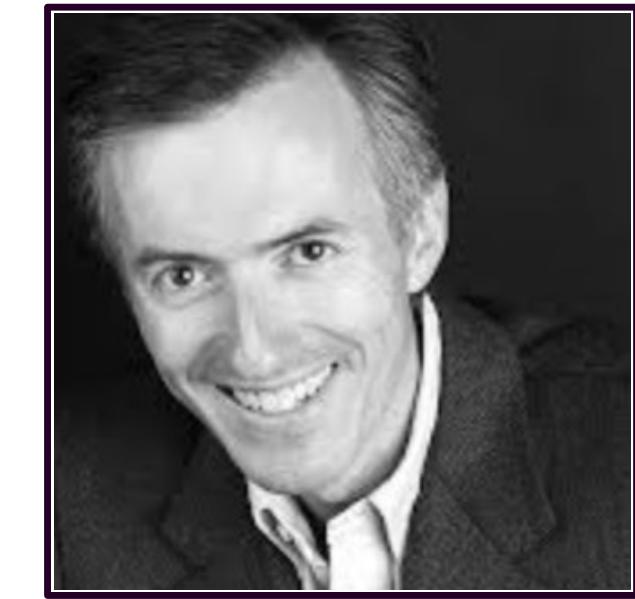
Express crucial actions in verbs.

Make your central characters the subjects of those verbs; keep those subjects short, concrete, and specific.

Old before new — “readers follow a story most easily if they can begin each sentence with a character or idea that is familiar to them, either because it was already mentioned or because it comes from the context.”

Complexity last, particularly important when:

introducing a new technical term,
presenting a long or complex unit of
information, introducing a concept to be
developed in what follows.



Trees, maps, theorems

Doumont

An engineer from the Louvain School of Engineering and PhD in applied physics from Stanford University, Jean-luc Doumont wrote this book to help engineers, scientists, and managers with business communication.

Wear shoes of your audience

Put yourselves in the shoes of the audience, anticipating their situation, their needs, their expectations.

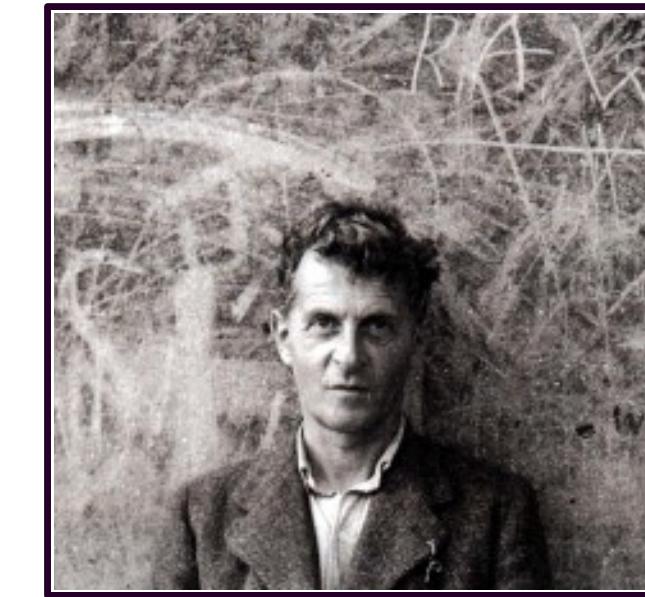
Structure the story along their line of reasoning, recognizing the constraints they might bring: their familiarity with the topic, their mastery of the language, the time they can free for us.

Messages, not just information

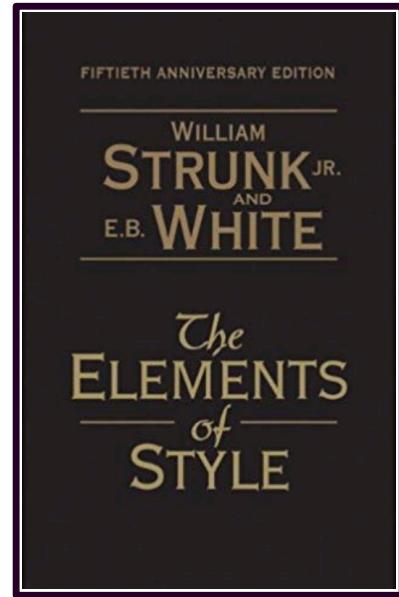
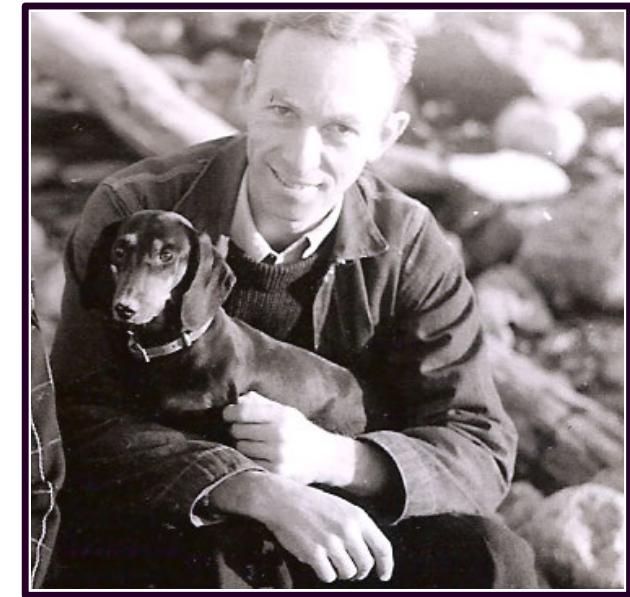
A message differs from raw information in that it interprets the information for a specific audience and for a specific purpose.

Organization, messages first

Get the audience interested. Once you have their attention, tell your main message. Last, support this message, tell how you got there.



&



The elements of style

Strunk & White

William Strunk Jr. was an English professor at Cornell University; E. B. White was his student. White wrote for *The New Yorker* — perhaps the best edited magazine — for sixty years, and won a Pulitzer for his writing.

Why read S&W, by Richard Ford

S&W doesn't really teach you how to write, it just tantalizingly reminds you that there's an orderly way to go about it, that clarity's ever your ideal, but — really — it's all going to be up to you.

**Mimic examples,
be concise,
don't overstate**

Leading by example, this tiny book provides dos and don'ts with examples of each. Re-read.

Heed their warnings:

Vigorous writing is concise. A sentence should contain no unnecessary words, a paragraph no unnecessary sentences, for the same reason that a drawing should have no unnecessary lines and a machine no unnecessary parts. This requires not that the writer make all his sentences short, or avoid all detail and treat subjects only in outline, but that **every word tell**.

A single overstatement, wherever or however it occurs, diminishes the whole, and a carefree superlative has the power to destroy, for readers, the object of your enthusiasm.

Example structure of a 250-word memo



A perfect game

Perfect Game

Los Angeles Dodgers

Pitcher Sandy Koufax

Statcast data

(Attempting to) steal a base

Salary cap

Baseball

Statistics, probability, computing

Models

Mode, maximum likelihood

Probability distributions

Joint distributions

R language and packages

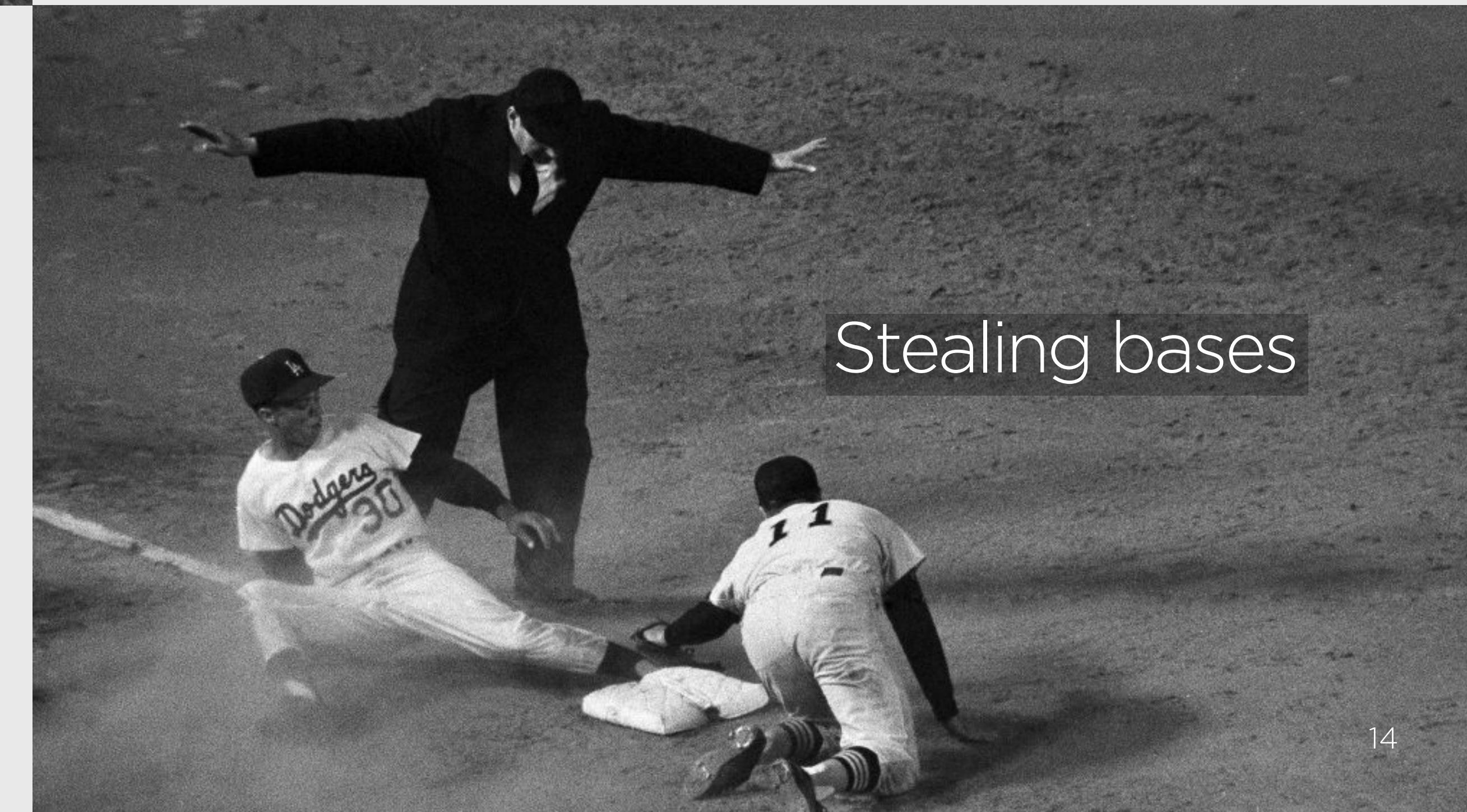
Inferences

Mean, expectations

Decision theory

Counterfactuals

Simulations



Stealing bases



To: **Scott Powers**
Director, Quantitative Analytics

2 February 2019

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint model of all events works best, but we can start small, with say, decisions to steal second base.

After defining our objective (e.g. optimize expected runs) we will, from Statcast data, compute expectations: weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do *this*” or “what if my opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations. This open-source software is free but teaching its applications will require time. To shorten our learning curve, we can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages *rethinking*, *brms*, or *rstanarm*. And we can test the concept with decisions to steal.

Sincerely,
Scott Spencer

Questions for Discussion

Who was the audience?

Was a few minutes enough time to understand?
Did you notice structure that guides the audience?

How do the audiences differ between this memo,
the *Jakarta proposal*, and *The Next Rembrandt*?

How do the structures and details of the
communications seem to depend on the audience?

Might the audience reply, tl;dr?



To: **Scott Powers**
Director, Quantitative Analytics

Audience background

Director of Quantitative Analytics
Ph.D. Statistics from Stanford University
Some publications use machine learning
Knows R programming
An employee, knows history of Dodgers

2 February 2019

Message first, context

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

Main body word count, 250

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint strategy of all events works best, but we can start small, with say, decisions to steal second base.

Readability Statistics	
Counts	After defining our objective (e.g. optimize expected runs) we will, from Statcast data, compute expectations:
Words	weight every event that could happen by its probability and accumulate these probability distributions. Joint
Characters	distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do <i>this</i> ” or “what if my
Paragraphs	opponent does <i>that</i> ” — and simulate games to learn how decisions change win probability. It enables optimal
Sentences	strategy.
Averages	
Sentences per Paragraph	Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the
Words per Sentence	free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent.
Characters per Word	There is no cap, however, on investing in rational decision processes.
Readability	
Flesch Reading Ease	Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations.
Flesch-Kincaid Grade Level	This open-source software is free but teaching its applications will require time. To shorten our learning curve, we
Passive Sentences	can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages rethinking, brms, or rstanarm. And we can test the concept with decisions to steal.

Sincerely,
Scott Spencer



To: **Scott Powers**
Director, Quantitative Analytics

2 February 2019

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

Context, orient the audience

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint model of all events works best, but we can start small, with say, decisions to steal second base.

After defining our objective (e.g. optimize expected runs) we will, from Statcast data, compute expectations: weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do *this*” or “what if my opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations. This open-source software is free but teaching its applications will require time. To shorten our learning curve, we can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages rethinking, brms, or rstanarm. And we can test the concept with decisions to steal.

Sincerely,
Scott Spencer



To: **Scott Powers**
Director, Quantitative Analytics

2 February 2019

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint model of all events works best, but we can start small, with say, decisions to steal second base.

Goal, action problem

After defining our objective (e.g. optimize expected runs) we will, from Statcast data, compute expectations: weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do *this*” or “what if my opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations. This open-source software is free but teaching its applications will require time. To shorten our learning curve, we can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages rethinking, brms, or rstanarm. And we can test the concept with decisions to steal.

Sincerely,
Scott Spencer



To: **Scott Powers**
Director, Quantitative Analytics

2 February 2019

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

Proposed solution

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint model of all events works best, but we can start small, with say, decisions to steal second base.

After defining our objective (e.g. optimize expected runs) we will, from Statcast data, compute expectations: weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do *this*” or “what if my opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations. This open-source software is free but teaching its applications will require time. To shorten our learning curve, we can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages rethinking, brms, or rstanarm. And we can test the concept with decisions to steal.

Sincerely,
Scott Spencer



To: **Scott Powers**
Director, Quantitative Analytics

2 February 2019

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint model of all events works best, but we can start small, with say, decisions to steal second base.

After defining our objective (e.g. optimize expected runs) we will, from Statcast data, compute expectations: weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do *this*” or “what if my opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

Objective, data, methods

Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations. This open-source software is free but teaching its applications will require time. To shorten our learning curve, we can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages rethinking, brms, or rstanarm. And we can test the concept with decisions to steal.

Sincerely,
Scott Spencer



To: **Scott Powers**
Director, Quantitative Analytics

2 February 2019

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint model of all events works best, but we can start small, with say, decisions to steal second base.

After defining our objective (e.g. optimize expected runs) we will, from Statcast data, compute expectations: weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do *this*” or “what if my opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations. This open-source software is free but teaching its applications will require time. To shorten our learning curve, we can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages rethinking, brms, or rstanarm. And we can test the concept with decisions to steal.

**Eventual
benefit**

Sincerely,
Scott Spencer



To: **Scott Powers**
Director, Quantitative Analytics

2 February 2019

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint model of all events works best, but we can start small, with say, decisions to steal second base.

After defining our objective (e.g. optimize expected runs) we will, from Statcast data, compute expectations: weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do *this*” or “what if my opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

**Benefit,
comparison,
financial**

Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations. This open-source software is free but teaching its applications will require time. To shorten our learning curve, we can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages rethinking, brms, or rstanarm. And we can test the concept with decisions to steal.

Sincerely,
Scott Spencer



To: **Scott Powers**
Director, Quantitative Analytics

2 February 2019

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint model of all events works best, but we can start small, with say, decisions to steal second base.

After defining our objective (e.g. optimize expected runs) we will, from Statcast data, compute expectations: weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do *this*” or “what if my opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations. This open-source software is free but teaching its applications will require time. To shorten our learning curve, we can start with Stan interfaces that use familiar syntax (like `lme4`) but return joint probability distributions: R packages `rethinking`, `brms`, or `rstanarm`. And we can test the concept with decisions to steal.

**Limitations,
short-term
solutions**

Sincerely,
Scott Spencer



To: **Scott Powers**
Director, Quantitative Analytics

2 February 2019

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint model of all events works best, but we can start small, with say, decisions to steal second base.

After defining our objective (e.g. optimize expected runs) we will, from Statcast data, compute expectations: weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do *this*” or “what if my opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations. This open-source software is free but teaching its applications will require time. To shorten our learning curve, we can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages rethinking, brms, or rstanarm. And we can test the concept with decisions to steal.

**Reminder
of proposal**

Sincerely,
Scott Spencer



To: Scott Powers
Director, Quantitative Analytics

Audience background

Senior Analyst
Ph.D. Statistics from Stanford University
Some publications use machine learning
Knows R programming
An employee, knows history of Dodgers

2 February 2019

Message first, context

Goal, action problem

Objective, data, methods

Limitations, short-term solutions

Context, orient the audience

Proposed solution

Eventual benefit

Benefit, comparison, financial

Reminder of proposal

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint model of all events works best, but we can start small, with say, decisions to steal second base.

After defining our objective (e.g. optimize expected runs) we will, from Statcast data, compute expectations: weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do *this*” or “what if my opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations. This open-source software is free but teaching its applications will require time. To shorten our learning curve, we can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages rethinking, brms, or rstanarm. And we can test the concept with decisions to steal.

Sincerely,
Scott Spencer

“When we read prose, we hear it... it’s variable sound. It’s sound with — pauses. With *emphasis*. With, well, you know, a certain rhythm.” — Richard Goodman

“If you **start your project early**, you’ll have time to **let your revised draft cool**. What seems good one day often looks different the next.” — Wayne Booth

Revise

“We write a first draft for ourselves; the drafts thereafter increasingly **for the reader.**” — Joseph Williams

Questions: tools for revising your memo

Sentence syntax, old before new? — Booth, chapter 17.

Needless words omitted? Every word tell? — S&W, composition 17.

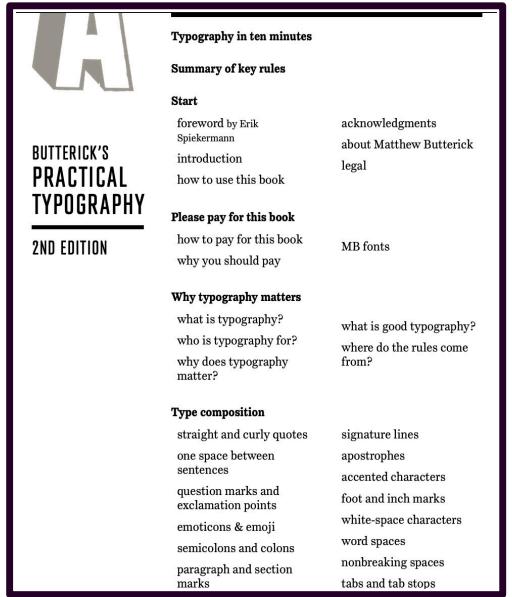
Overstatements? — S&W, style 7.

Statements, in positive form? — S&W, composition 15.

Each paragraph, a separate unit of composition? — S&W, composition 16.

Definite, specific, concrete language? — S&W, composition 13.

Visual presentation of information



Butterick's practical typography

Matthew Butterick

The author earned a visual-studies degree from Harvard, and a law degree from UCLA. He is a writer, typographer, programmer, and lawyer.

Conserve limited reader attention

Typography is the visual component of the written word.

“Typography is for the **benefit of the reader.**”

“Most readers are looking for reasons to stop reading. . . . Readers have other demands on their time. . . . The goal of most professional writing is persuasion, and attention is a prerequisite for persuasion. **Good typography can help your reader devote less attention to the mechanics of reading and more attention to your message.**”

Body text
(very basic guidelines)

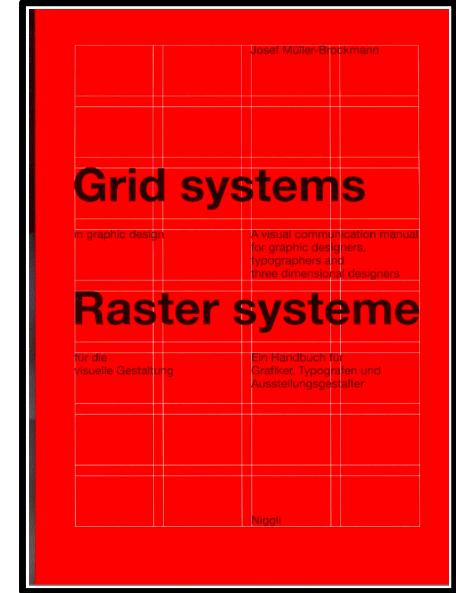
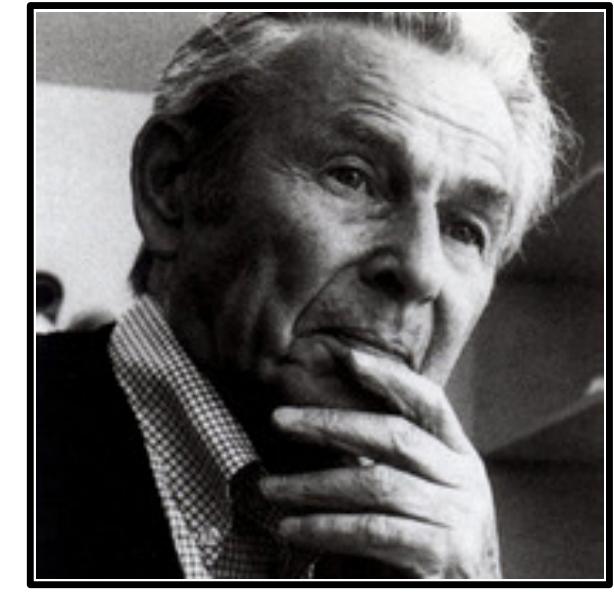
Point size: 10-12 (print), 15-25 pixels (web)
Line spacing: 120-145% of the point size
Line length: 45-90 characters per line
Fonts: *see his recommendations*

Layout

Grids are helpful when they encourage consistency. They make it easier to relate elements on the page to existing ones.

Color

Nothing draws the eye more powerfully than a contrast between light and dark colors. This is why a **bold** font creates more emphasis than an *italic* font.

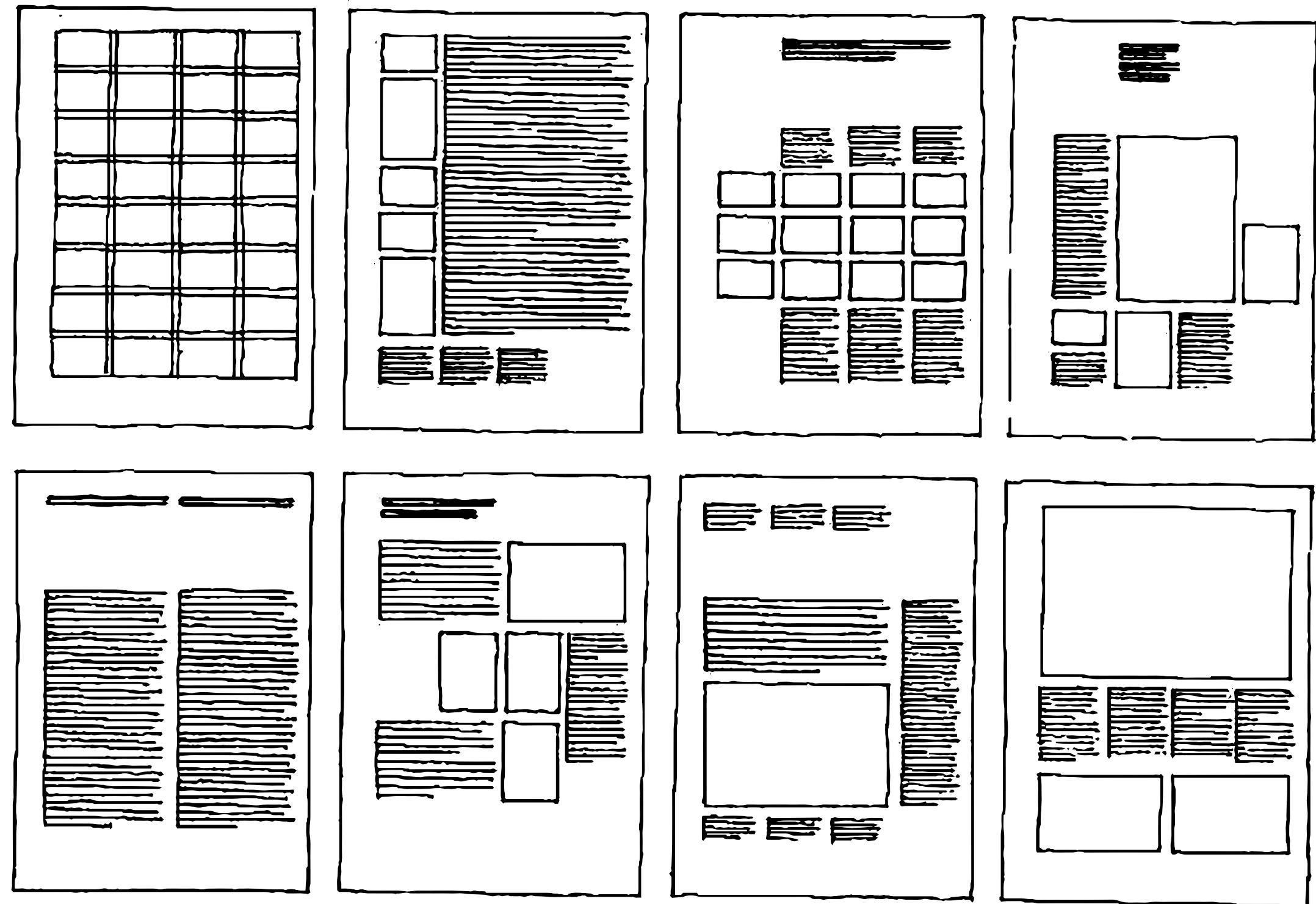


Grid Systems in Graphic Design

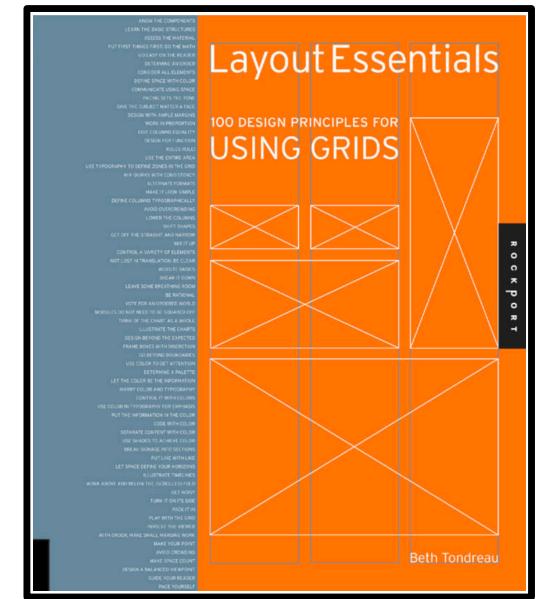
Müller-Brockmann

His book, an in-depth analysis of layout in design, is seminal and remains influential among theory of communication through visual design.

Arranging surfaces and spaces into a grid creates conformity among texts, images and diagrams. The size of each implies its importance. Reducing elements in a grid suggests planning, intelligibility, clarity, and orderliness of design. **One grid allows many creative ways to show relationships:**



Orderliness adds credibility to the information and **induces confidence**. Information presented with clear and logically set out titles, subtitles, texts, illustrations and captions will not only be **read more quickly and easily** but the information will also be **better understood**.



Layout Essentials

Tondreau

Before founding a design firm, Tondreau was Design Director at Viking / Penguin publishing company.

Her book on layout essentials helps readers consider information organization.

The main components of a grid are margins, markers, columns, flowlines, spatial zones, and modules.

COLUMNS

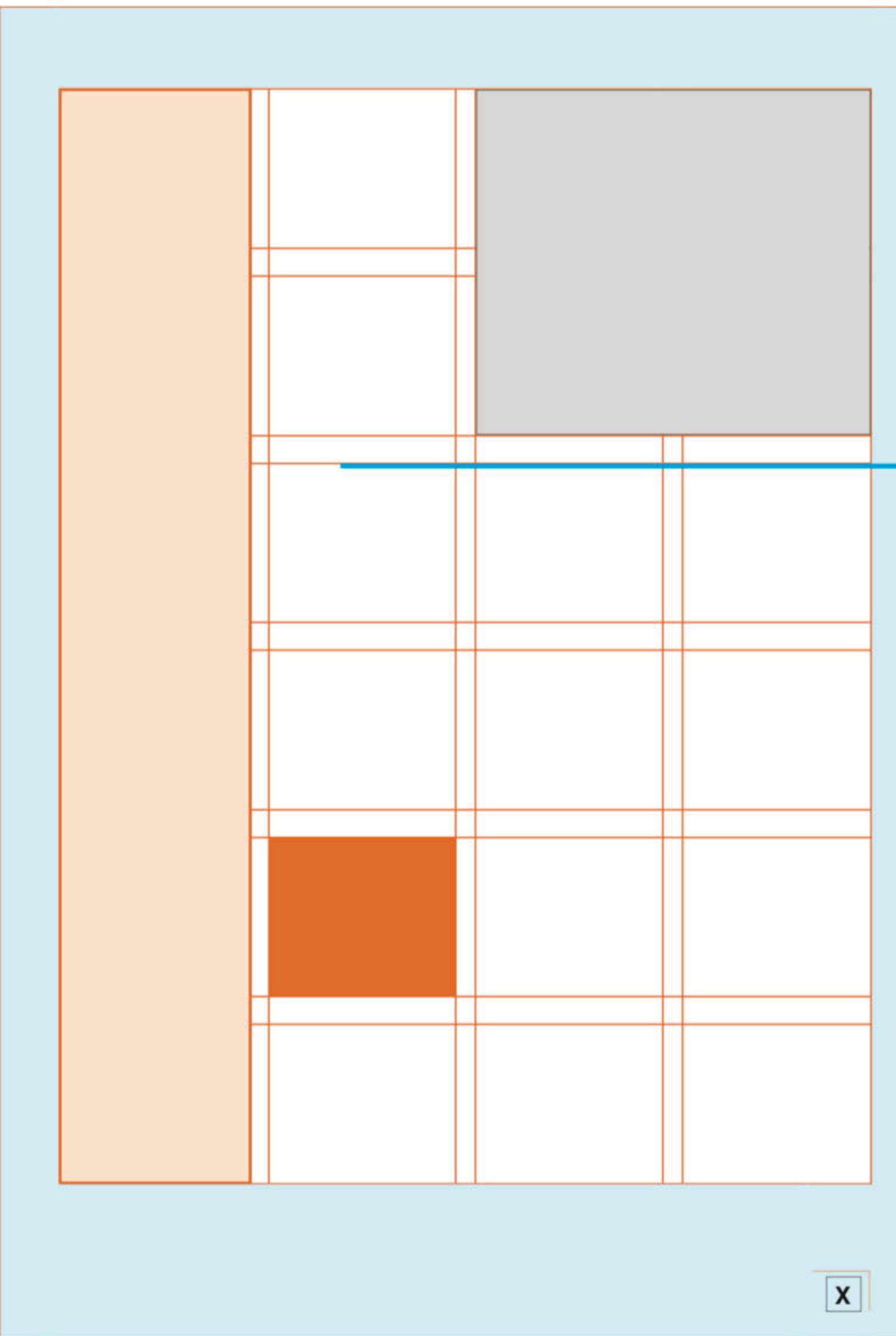
are vertical containers that hold type or images. The width and number of columns on a page or screen can vary, depending on the content.

MODULES

are individual divisions separated by consistent space, providing a repeating, ordered grid. Combining modules can create columns and rows of varying sizes.

MARGINS

are buffer zones. They represent the amount of space between the trim size, including gutter, and the page content. Margins can also house secondary information, such as notes and captions.



SPATIAL ZONES

are groups of modules or columns that can form specific areas for type, ads, images, or other information.

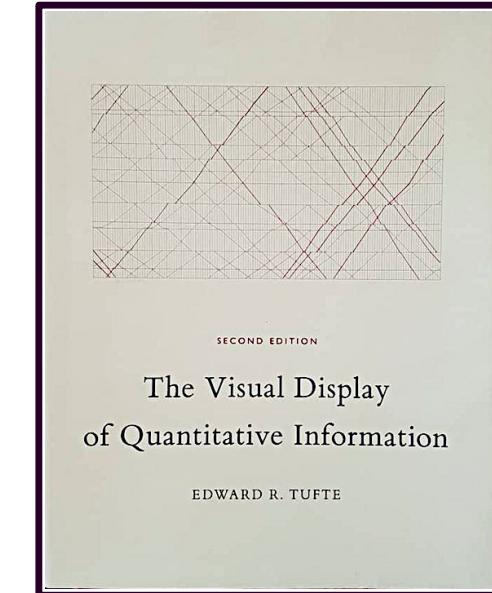
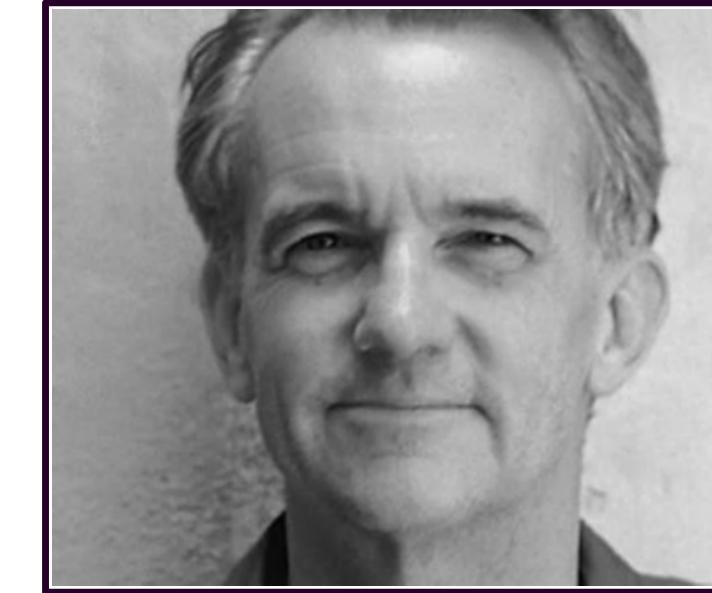
FLOWLINES

are alignments that break space into horizontal bands. Not actual lines, flowlines are a method for using space and elements to guide a reader across a page.

MARKERS

help a reader navigate a document. Indicating placement for material that appears in the same location, markers include page numbers, running heads and feet (headers and footers), and icons.

Visually linking words with graphics



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.

Graphics help us reason with data

At their best, **graphics are instruments for reasoning about quantitative information**.

Often the **most effective way to describe, explore, and summarize a set of numbers**—even a very large set—is to **look at pictures of those numbers**.

Furthermore, of all methods for analyzing and communicating statistical information, well-designed **data graphics are usually the simplest and at the same time the most powerful**.

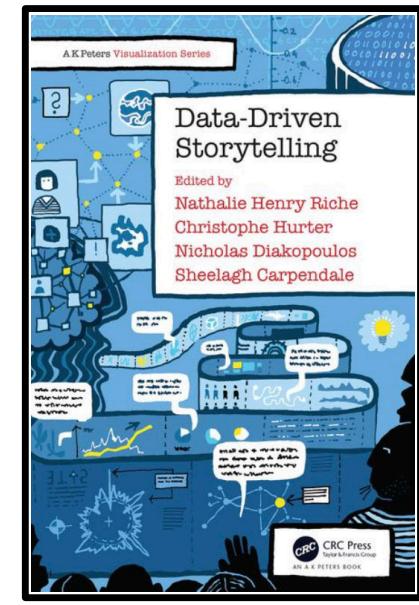
Use words and pictures together

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

Data-Driven Storytelling

Riche, co-editors

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



Link between narrative and visual

Annotation layer of visual display

Visual data comparisons: patterns for persuasion

The link between the narrative and the visualization **helps the reader discern what item in the visualization the author is referencing in the text.**

Create links with annotation, color, luminosity, or lines.

Annotations add explanations and descriptions to introduce the graph's context, which is important for almost any audience.

For comparison, the narrator presents multiple data sets, and draws conclusions. Visually, it can be made through side-by-side presentation of graphics, or changes of a single graphic over time.

Comparison can

show equality of both data sets, highlight differences and similarities, or give reasons for their difference.



Figures Example linking words to graphics

Kay

He is an Assistant Professor of Information at UMSI, and works in human-computer interaction information visualization, and Communicating uncertainty.

Use color to link information

Color is used to directly and implicitly refer to relevant parts of the visualization within annotation text, making potentially complex references clear and succinct—without the need for more explicit legends or additional annotation. Thus, the narrative flows linearly.

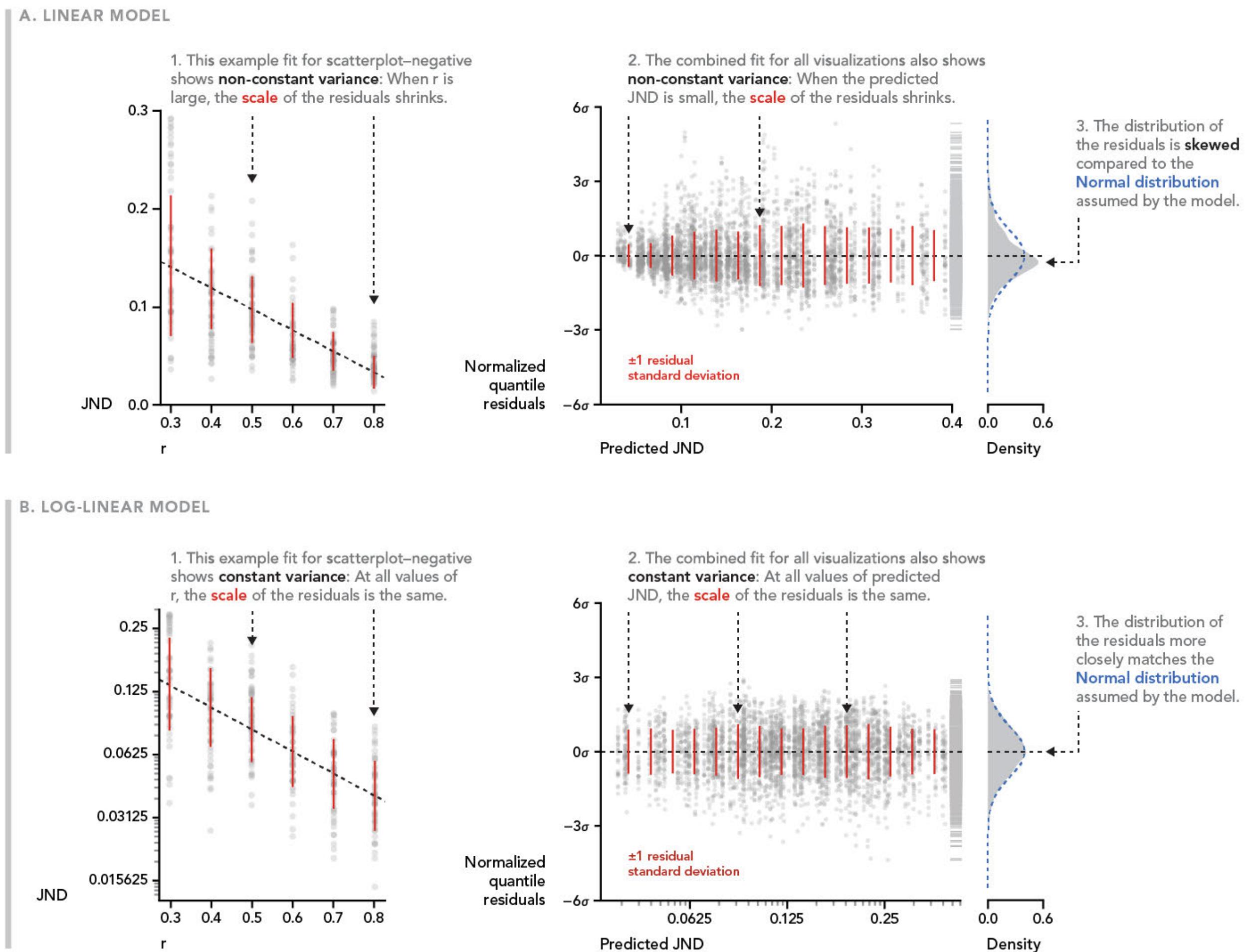


Fig. 3 Comparison of fits of the linear model (Section 3) and the log-linear model (Section 4). Example fits of each model to scatterplot-negative are shown in A.1 and B.1. Plots of normalized residuals for all visualization \times direction pairs are shown in A.2 and B.2. Density plots of normalized residuals with comparison to the standard normal distribution are shown in A.3 and B.3.

Group work

Group help on case studies, memo

Get together in groups, work with your peers to get and give helpful feedback to refine your ideas for **your choice of entity**, your **data analytics project**, potential **sources of data**, and **writing it up**.

Then, let's share some of these ideas together.

For Next Week, Module 4:

Agenda next week

The minimum

250-word memo due by 16th of June
Principles of Persuasion and Brief Proposals

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

Read to understand his framework for using statistics as persuasive communication.

Also:

What's his ideal statistician?

What does he mean by *MAGIC*?

What are his thoughts about *comparisons*?

Conger J.A. (1998, May-June). *The necessary art of persuasion*. Harvard Business Review, 84-97.

What steps, in his view, must be considered?
What examples of successes and failures in these steps have you witnessed? Has he categorically omitted anything you consider important in his generalization of persuasion?

Checking in,

Reaching out?

Have you registered with Columbia's Writing Center? Scheduled a consultation with our Research Data Services?

**Coding?
Exploring data?**

What coding help do you need to gather, clean or visually explore your data? Be as specific as you can and we can setup a one-on-one.

**See you
next week!**

