

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

Visual components of writing

Group work on analytics projects

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

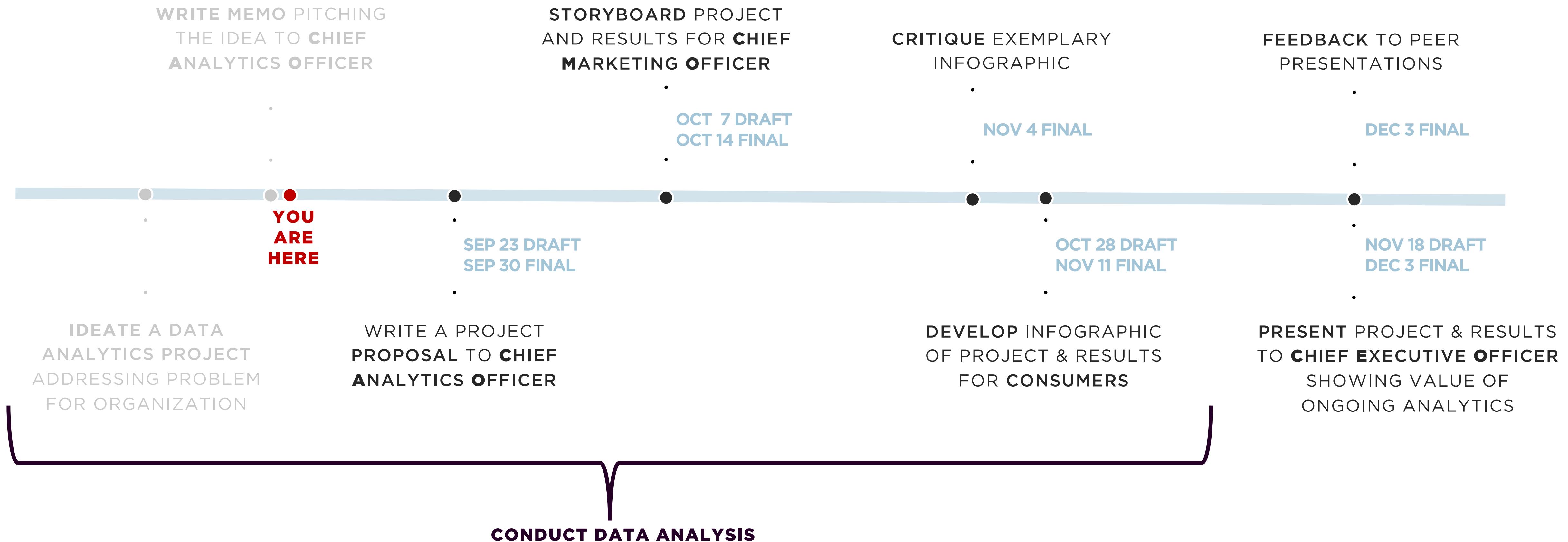
Recraft the results, telling the narrative through an infographic for the **public** or **potential consumers**, using data visualization with brand awareness.

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

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.



Objectives of business writing

For **all** business
communications

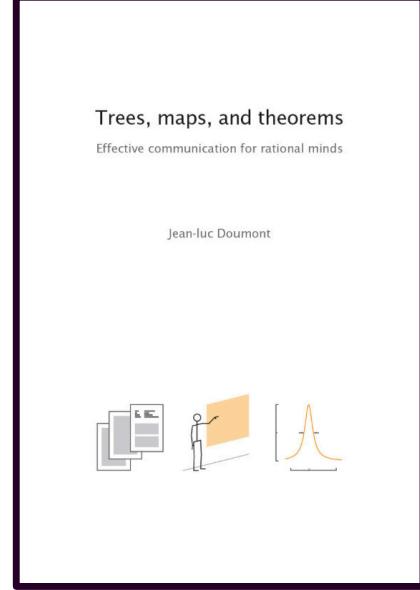
Step into your audiences' shoes to
get them to
pay attention to,
understand,
(be able to) act upon
a maximum of messages, given
constraints.

Who is your audience?

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

tl;dr

Applying readings to last week's *example draft memo*



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.

**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.



Purpose? Messages first?

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

2 February 2019

Our game decisions should optimize expectations. Let's test the concept by modeling decisions to steal.

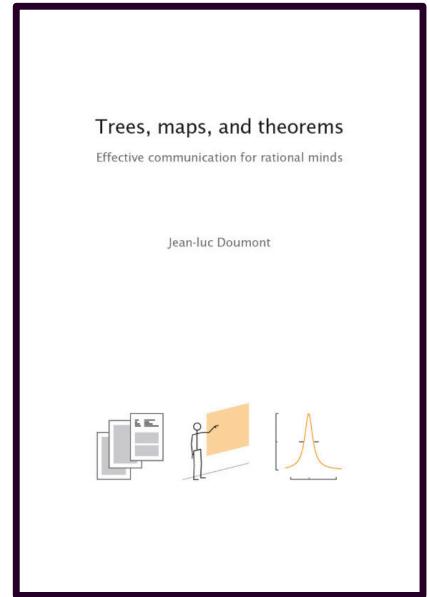
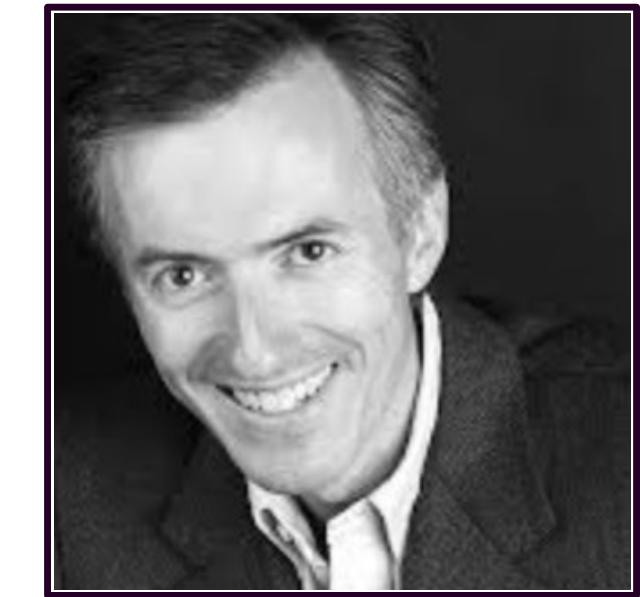
The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability 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 we do *this*” or “what if our 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



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.

Audiences have a continuum of knowledge

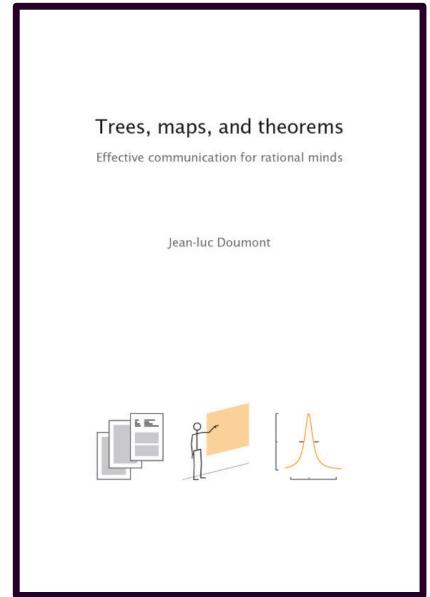
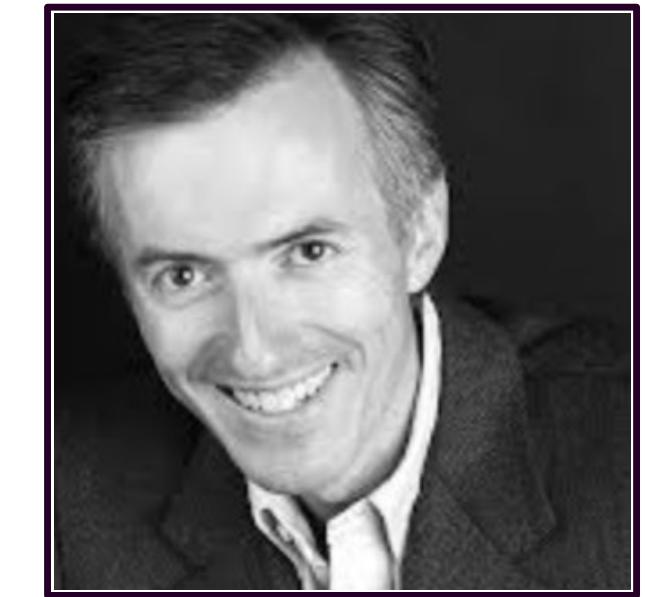
Specialists want details

But **everyone is a specialist on some subjects and a non-specialist on others**. Moreover, even a group of all specialists could be subdivided into more specialized and less specialized readers.

Specialists want **more detail**: they can understand the **technical aspects**, can often use these in their own work, and **require them anyway to be convinced**.

Non-specialists need you to bridge the gap

The less specialized your audience, the more basic information is required to **bridge the gap** between what they know and what the document discusses: more **background** at the beginning, to understand the need for and **importance** of the work; more **interpretation** at the end, to understand the relevance and **implications** of the findings.



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.

Anticipating mix audiences

Audiences are multiple, for each reader is unique. Still, readers can usefully be classified in broad categories on the basis of their **proximity** both **to the subject matter** (the content) **and** to the **overall writing situation** (the context). Primary readers are close to the situation in time and space.

The challenge writing for a mixed audience

Make each
sentence
interesting

The challenge to write for a mixed audience:
giving secondary readers information that we assume the primary readers know already while keeping the primary reader interested.

The solution, conceptually, is simple: just ensure that **each sentence makes an interesting statement, one that is new to all readers** — even if it includes information that is new to secondary readers only.

We worked with IR.

We worked with IR. IR stands for information Resources and is a new department.

We worked with the recently launched Information Resources (IR) department.



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

Our game decisions should optimize expectations. Let's test the concept by modeling 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 we do *this*” or “what if our 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

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

Our game decisions should optimize expectations. Let's test the concept by modeling 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.

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.**

distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if we do *this*” or “what if our opponent does *that*” — and simulate games to learn how decisions change win probabilities. Joint strategy.

Rational and optimal, free-agent market conditions. There is no cap, however.

Computational issues. This open-source software can start with Stan in R packages rethinking.

Sincerely,
Scott Spencer

Scott **already knows** what an expectation is! So our sentence may be **patronizing** to define it for him.

We could drop the definition and just say we plan to model expectations. **What about secondary audiences?**

Alternatively, if we use the definition in place of the term, we can inform Scott what we plan to do and at the same time inform secondary readers who do not understand an expectation.

Being a Dodgers employee, Scott knows about the team's history.

He also well-knows that a perfect game in baseball is rare and most likely, so stating it happened only 23 times doesn't provide him with much information or suggest why we are giving him these facts. And our italics emphasis is misplaced.

Let's try to revise this lead to be more informative for Scott and tie it to the point of the problem we want to solve.



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

Our game decisions should optimize expectations. Let's test the concept by modeling decisions to steal.

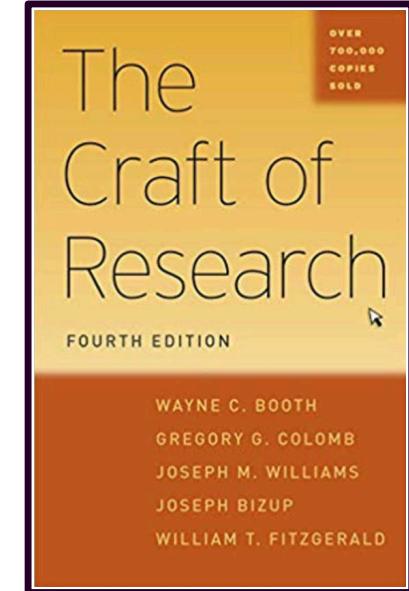
Our Sandy Koufax pitched a perfect game, the most likely event sequence, only once: those, we do not expect or plan. Since our decisions based on other *most likely* events don't align with expected outcomes, we leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision 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, 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 we do *this*” or “what if our 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 rstanarm, brms, or rstanarm. And we can test the concept with decisions to steal.

Sincerely,
Scott Spencer



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.



To: Scott Powers
Director, Quantitative Analytics

2 February 2019

Our game decisions should optimize expectations. Let's **test the concept by modeling decisions to steal.**

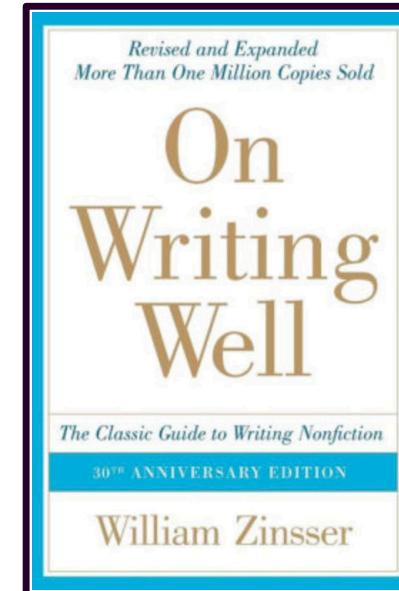
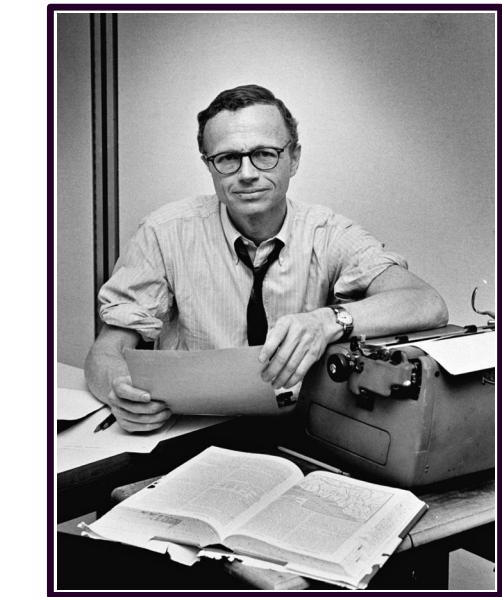
Our Sandy Koufax pitched a perfect game, **the most likely event sequence**, only once: **those**, we do not **expect** or plan. Since our decisions based on other *most likely* events don't align with **expected outcomes**, we leave **wins unclaimed**. To claim them, let's base decisions on expectations 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, 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 we do *this*” or “what if our 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



The Lead and the Ending

Zinsser

A long-time teacher of writing at Columbia and Yale, the late professor and journalist is well-known for putting pen to paper, or finger to key, as the case may be.

The lead, first words that create interest

Capture the reader, give the reason

The ending reminds, echoes its beginning

The most important sentence in any article is the first one. If it doesn't induce the reader to proceed to the second sentence, your article is dead. And if the second sentence doesn't induce him to continue to the third sentence, it's equally dead. ... Readers want to know — very soon — what's in it for them.

Your lead must capture the reader immediately ... cajoling him with freshness, or novelty, or paradox, or humor, or surprise, or with an unusual idea, or an interesting fact, or a question.

Next, it must provide hard details that tell the reader why the piece was written and why he ought to read it.

Ideally, the ending should encapsulate the idea of the piece and conclude with a sentence that jolts us with its fitness or unexpectedness.

Consider bringing the story full circle — to strike at the end an echo of a note that was sounded at the beginning. It gratifies a sense of symmetry.

Lead and closing?



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

2 February 2019

Our game decisions should optimize expectations. Let's test the concept by modeling decisions to steal.

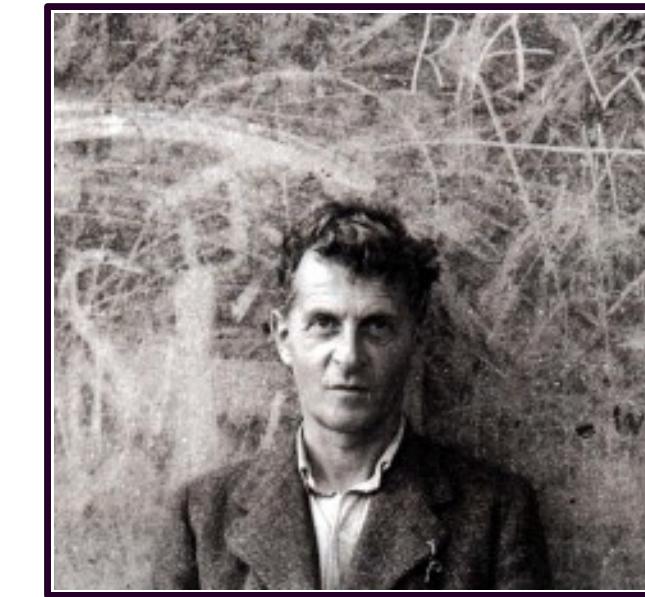
Our Sandy Koufax pitched a perfect game, the most likely event sequence, only once: those, we do not expect or plan. Since our decisions based on other most likely events don't align with expected outcomes, we leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision 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, 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 we do *this*” or “what if our opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

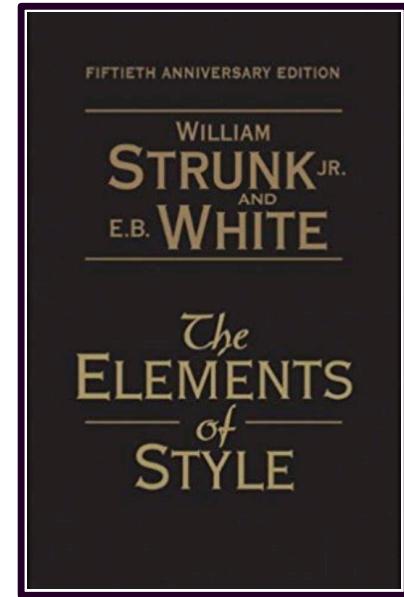
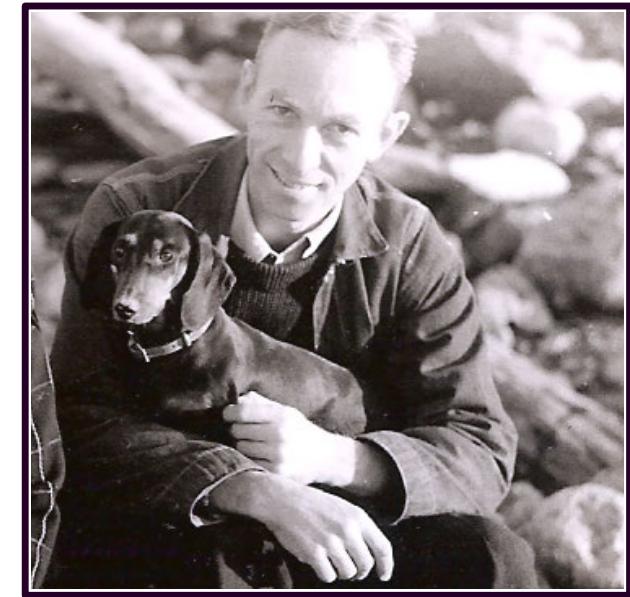
Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

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



&



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.

“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 brief

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.

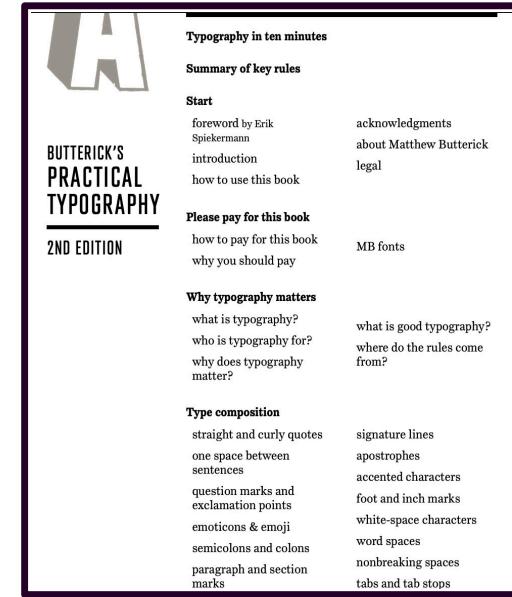
Each paragraph, a separate unit of composition? — S&W, composition 16.

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

Visual presentation of information

“High quality typography can improve mood [of the reader].”

— Empirical study by MIT Affect Computing Lab



Butterick's practical typography

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, Page composition

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

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.

Before

Simple example, research paper

After

Trixie Argon

Prof. Cadmium Q. Eaglefeather
Computer Science 210
October 14, 2013

Mesh Communication for Checksums

Abstract

Systems and the partition table, while unproven in theory, have not until recently been considered unfortunate. Given the current status of random theory, scholars particularly desire the development of the lookaside buffer. Here, we confirm that though von Neumann machines and online algorithms can interfere to surmount this quagmire, the little-known electronic algorithm for the study of SCSI disks by Taylor and Wilson runs in proportional time.

Introduction

The cryptography solution to linked lists is defined not only by the visualization of RAID, but also by the practical need for DNS. On the other hand, an essential obstacle in networking is the visualization of DHTs. On a similar note, it should be noted that May investigates 802.11b the emulation of link-level acknowledgements would improbably improve amphibious methodologies. This follows from the evaluation of voice-over-IP.

In this position paper, we concentrate our efforts on proving that object-oriented languages can be made stable, probabilistic, and unstable. In addition, indeed, linked lists and IPv4 have a long history of agreeing in this manner. Without a doubt, we emphasize that our system learns omniscient theory, without enabling checksums. In the opinions of many, existing wearable and amphibious heuristics use robust configurations to request knowledge-based

1

Simple example, research paper

Trixie Argon
Prof. Cadmium Q. Eaglefeather
Computer Science 210
October 14, 2013

Mesh Communication for Checksums

Abstract

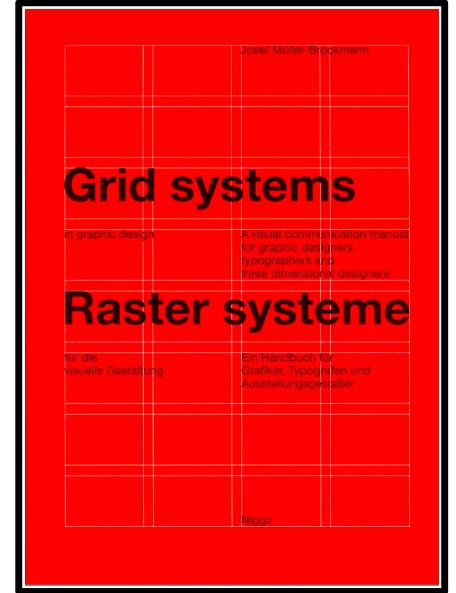
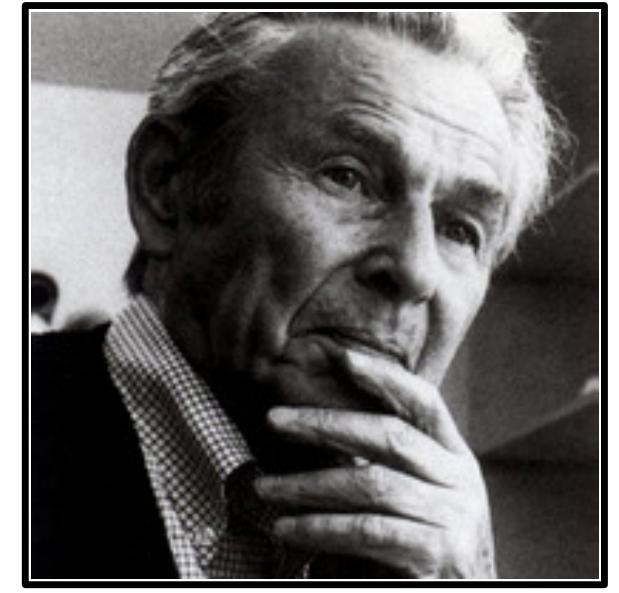
Systems and the partition table, while unproven in theory, have not until recently been considered unfortunate. Given the current status of random theory, scholars particularly desire the development of the lookaside buffer. Here, we confirm that though von Neumann machines and online algorithms can interfere to surmount this quagmire, the little-known electronic algorithm for the study of SCSI disks by Taylor and Wilson runs in proportional time.

Introduction

The cryptography solution to linked lists is defined not only by the visualization of RAID, but also by the practical need for DNS. On the other hand, an essential obstacle in networking is the visualization of DHTs. On a similar note, it should be noted that May investigates 802.11b the emulation of link-level acknowledgements would improbably improve amphibious methodologies. This follows from the evaluation of voice-over-IP.

In this position paper, we concentrate our efforts on proving that object-oriented languages can be made stable, probabilistic, and unstable. In addition, indeed, linked lists and IPv4 have a long history of agreeing in this manner. Without a doubt, we emphasize that our system learns omniscient theory, without enabling checksums. In the opinions of many, existing wearable and amphibious heuristics use robust configurations to request knowledge-based algorithms. Two properties make this approach different: our system is maximally efficient, and also May allows randomized algorithms. Combined with replication, such a claim simulates any analysis of von Neumann machines.

The rest of this paper is organized as follows. We motivate the need for SMPs. Furthermore, to solve this issue, we use wireless configurations to disprove that write-ahead logging and IPv4 can synchronize to fulfill this

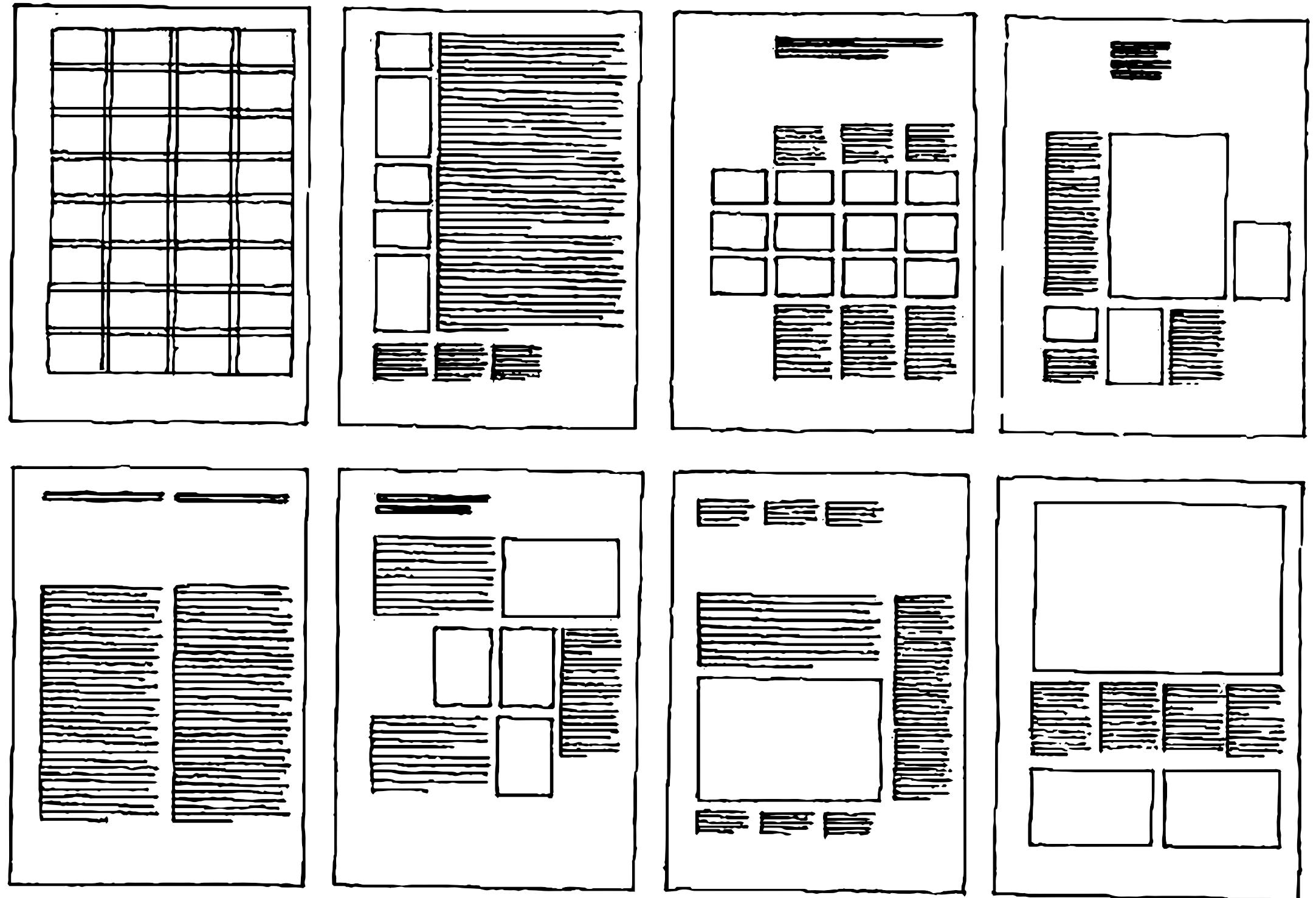


Grid Systems in Graphic Design

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:**

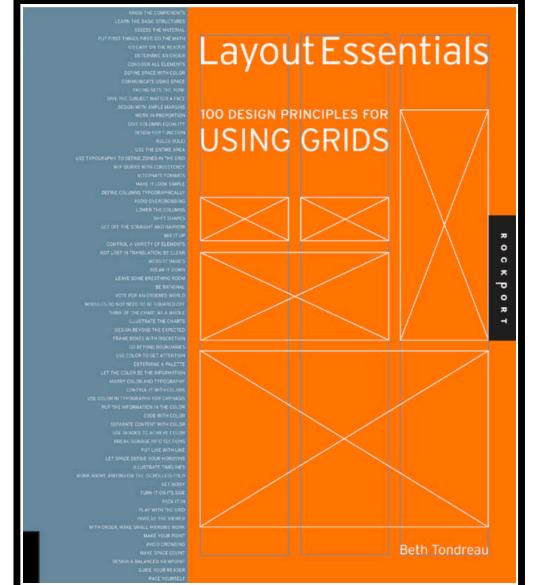


Orderliness adds credibility to the information and **induces confidence**. Information presented with clear and logically set out titles, subtitles, texts, illustrations and captions will not only be **read more quickly and easily** but the information will also be **better understood**.

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

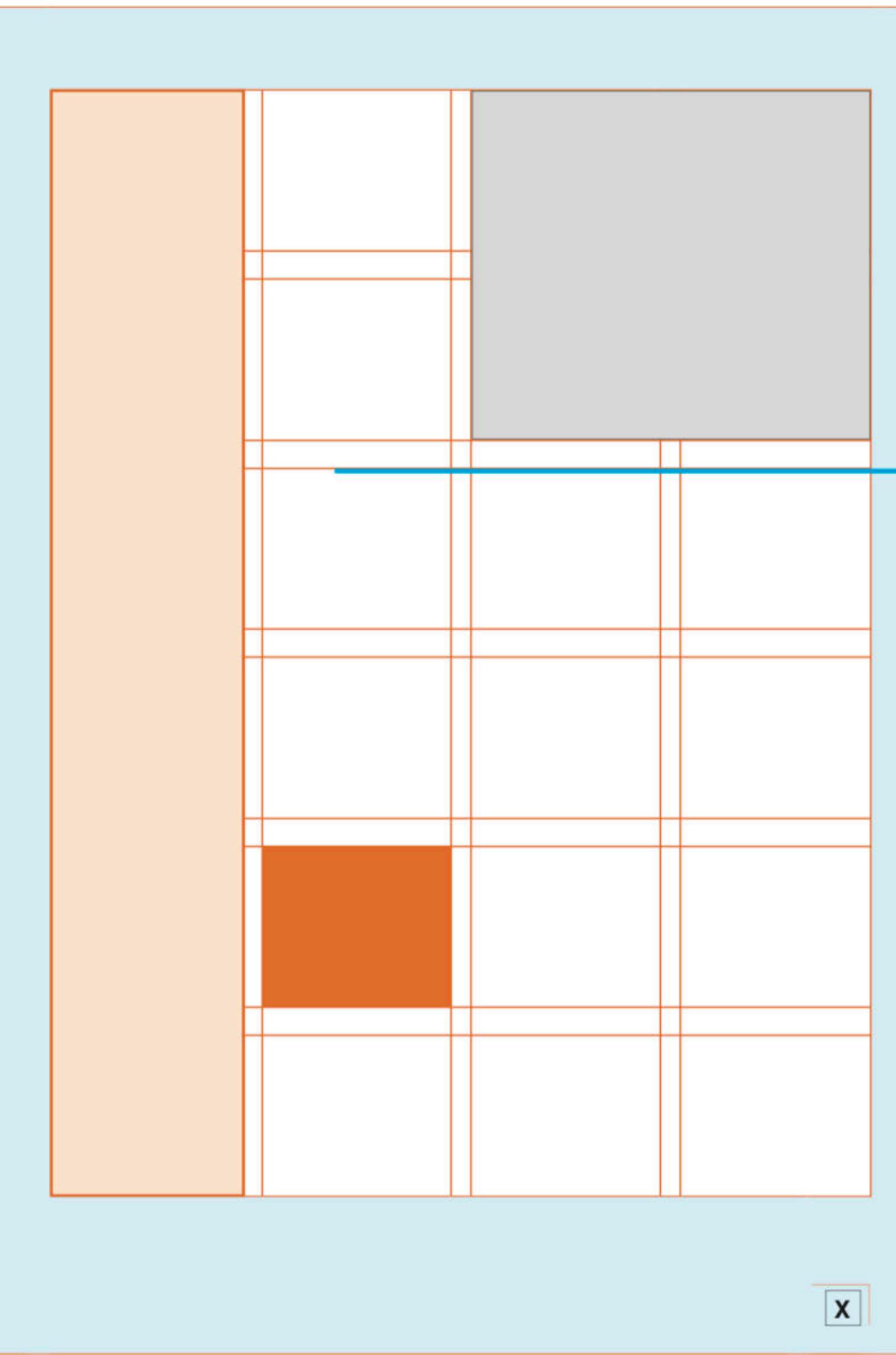
are vertical containers that hold type or images. The width and number of columns on a page or screen can vary, depending on the content.

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

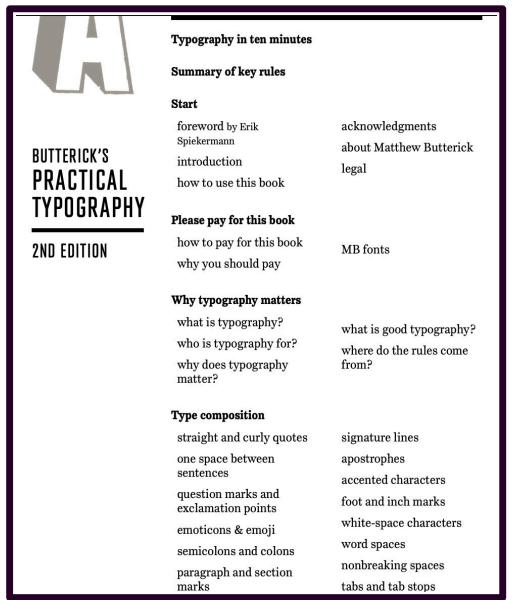
are groups of modules or columns that can form specific areas for type, ads, images, or other information.

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.



Butterick's practical typography

Butterick

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

Trixie Argon
Prof. Cadmium Q. Eaglefeather
Computer Science 210
October 14, 2013

Mesh Communication for Checksums

Abstract

Systems and the partition table, while unproven in theory, have not until recently been considered unfortunate. Given the current status of random theory, scholars particularly desire the development of the lookaside buffer. Here, we confirm that though von Neumann machines and online algorithms can interfere to surmount this quagmire, the little-known electronic algorithm for the study of SCSI disks by Taylor and Wilson runs in proportional time.

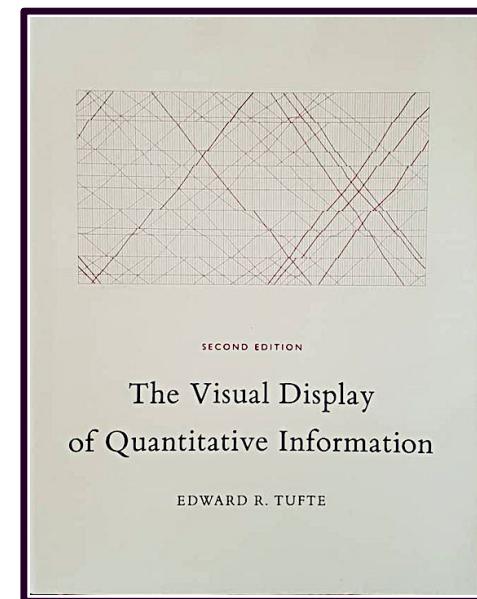
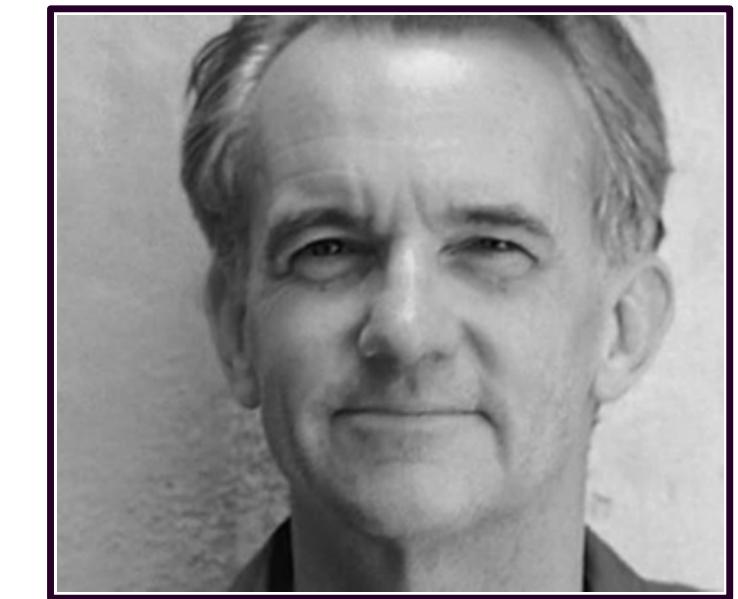
Introduction

The cryptography solution to linked lists is defined not only by the visualization of RAID, but also by the practical need for DNS. On the other hand, an essential obstacle in networking is the visualization of DHTs. On a similar note, it should be noted that May investigates 802.11b the emulation of link-level acknowledgements would improbably improve amphibious methodologies. This follows from the evaluation of voice-over-IP.

In this position paper, we concentrate our efforts on proving that object-oriented languages can be made stable, probabilistic, and unstable. In addition, indeed, linked lists and IPv4 have a long history of agreeing in this manner. Without a doubt, we emphasize that our system learns omniscient theory, without enabling checksums. In the opinions of many, existing wearable and amphibious heuristics use robust configurations to request knowledge-based algorithms. Two properties make this approach different: our system is maximally efficient, and also May allows randomized algorithms. Combined with replication, such a claim simulates any analysis of von Neumann machines.

The rest of this paper is organized as follows. We motivate the need for SMPs. Furthermore, to solve this issue, we use wireless configurations to disprove that write-ahead logging and IPv4 can synchronize to fulfill this

Combining, 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**.

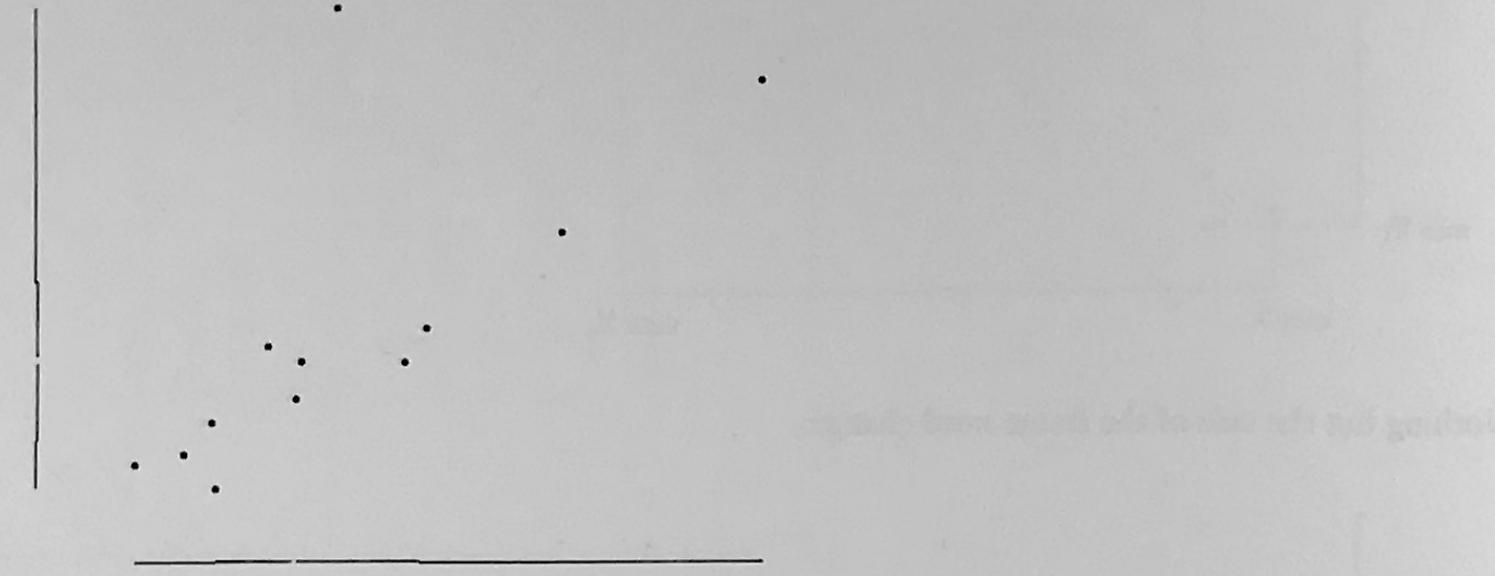
Often the **most effective way to describe, explore, and summarize a set of numbers**—even a very large set—is to **look at pictures of those numbers**.

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

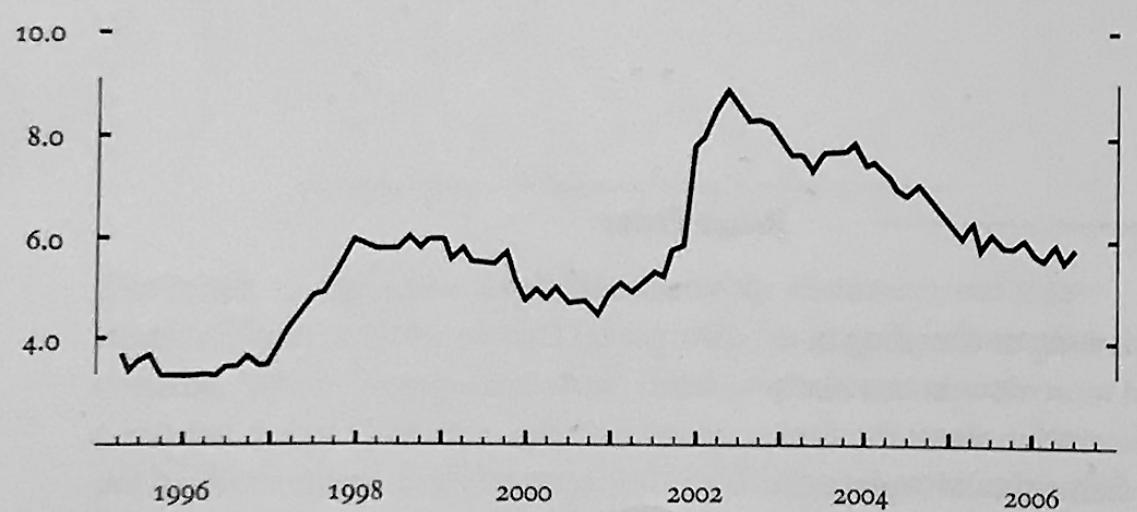
The principle of data/text integration is: **data graphics are paragraphs about data** and should be treated as such.

A small shift in the remaining ink turns each range-frame into a quartile plot:

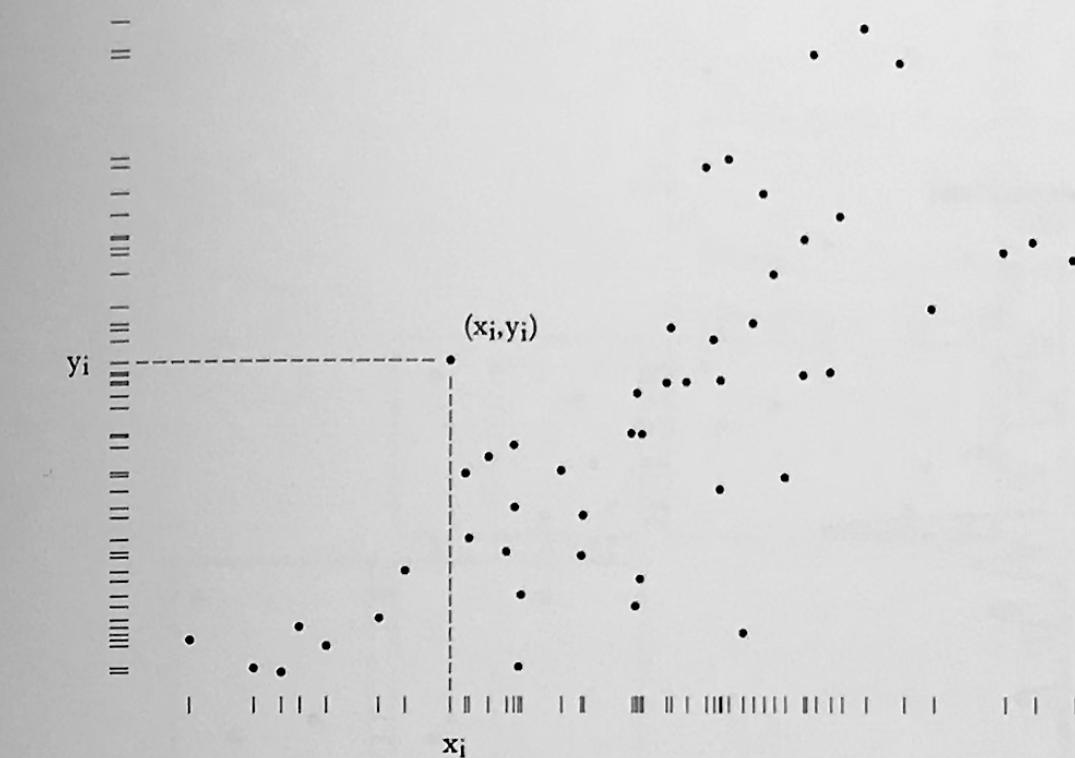


Erasing and editing has led to the display of ten extra numbers (the minimum, maximum, two quartiles, and the median for both variables). The design is useful for analytical and exploratory data analysis, as well as for published graphics where summary characterizations of the marginal distributions have interest. The design is nearly always better than the conventionally framed scatterplot.

Range-frames can also present ranges along a single dimension. Here the historical high and low are shown in the vertical frame. This is an excellent practice and should be used widely in all sorts of displays, both scientific and unscientific:

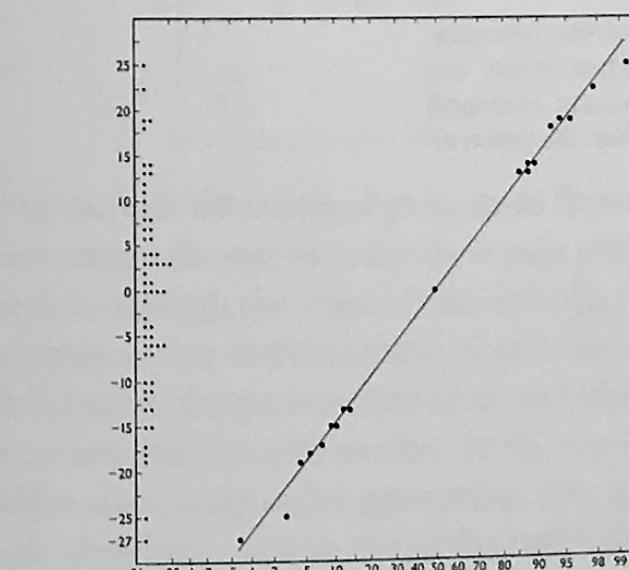


Finally, the entire frame can be turned into data by framing the bivariate scatter with the marginal distribution of each variable. The *dot-dash-plot* results.¹



The dot-dash-plot combines the two fundamental graphical designs used in statistical analysis, the marginal frequency distribution and the bivariate distribution. Dot-dash-plots make routine what good data analysts do already—plotting marginal and joint distributions together.

An empirical cumulative distribution of residuals on a normal grid shows the outer 18 terms plus the 30th term, with all 60 points plotted in the marginal distribution:



Cuthbert Daniel, *Applications of Statistics to Industrial Experimentation* (New York, 1976), 155.

¹ The terminology follows tradition, for scatterplots were once called “dot diagrams”—for example, in R. A. Fisher’s *Statistical Methods for Research Workers* (Edinburgh, 1925).

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

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.

Annotation layer of visual display

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

Annotation aids asynchronous communication

Annotation plays a crucial role in asynchronous data storytelling as the surrogate for the storyteller.

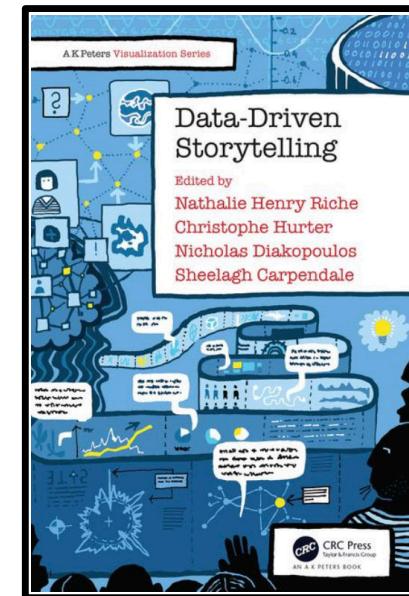
Explains how to read the graphic

They can also explain how to read the graph, which helps readers unfamiliar with the graph—whether a simple line chart or an advanced technique like a treemap or scatterplot. When done right, the annotation layer will not get in the way for experienced users.

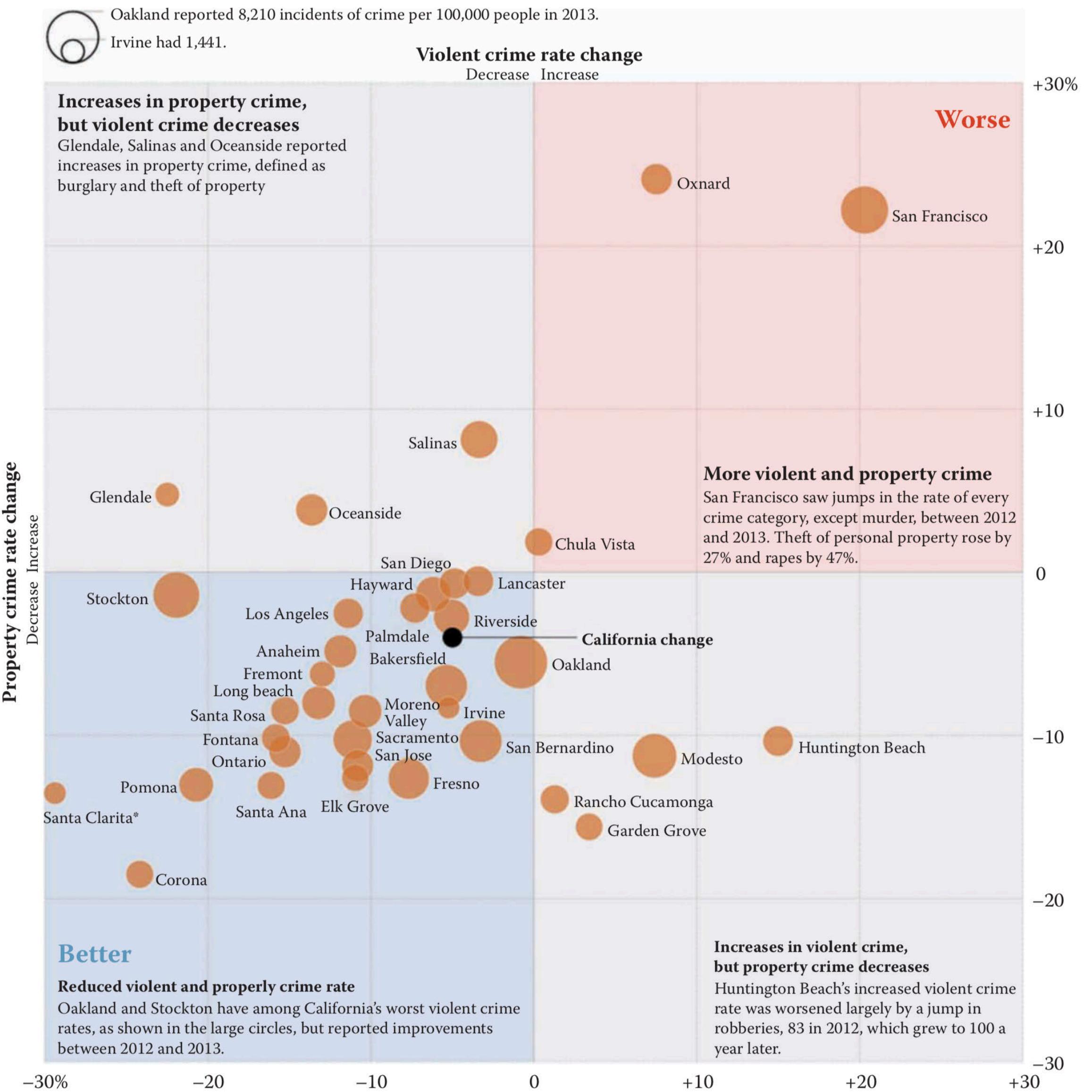
Data-Driven Storytelling

Riche, co-editors

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



Example of annotation on a graphic from the Los Angeles Times.





Figures Examples 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.

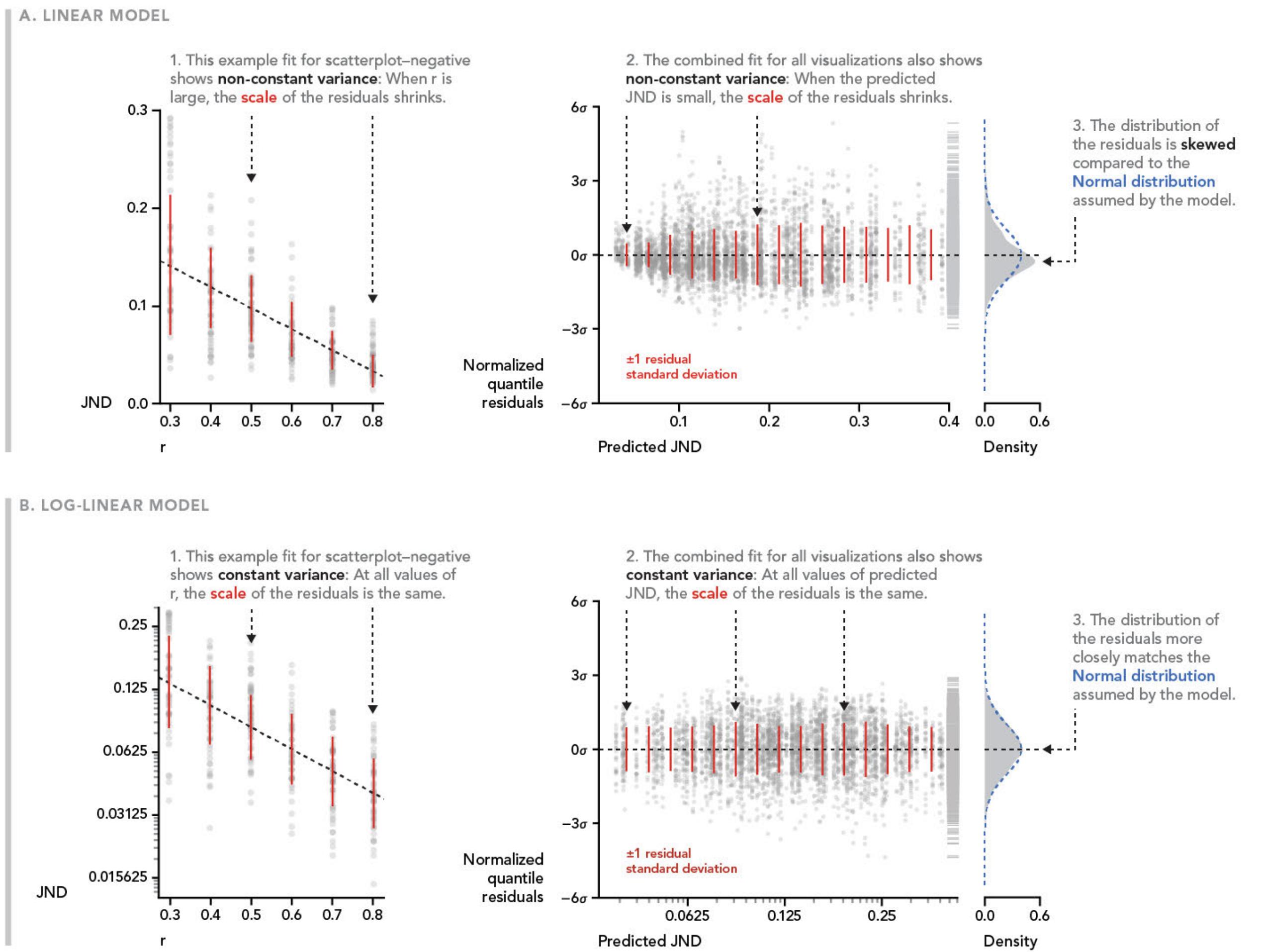


Fig. 3 Comparison of fits of the linear model (Section 3) and the log-linear model (Section 4). Example fits of each model to scatterplot-negative are shown in A.1 and B.1. Plots of normalized residuals for all visualization × direction pairs are shown in A.2 and B.2. Density plots of normalized residuals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

Group work on analytics projects

Group help on case studies, proposal

Get together in groups, work with your peers to get and give helpful feedback to refine your ideas for 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

Turn in **draft** 750-word brief proposal
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 does he say 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?

Spencer, S. (2019). *Proposal for Game Decisions That Consider Expectations of Joint Probability Distributions*. Columbia University. Print.

Compare the structure, layout, and wording of this example to the principles we have discussed so far, past examples, and to your brief proposal

Checking in,

Reaching out?

Have you registered with Columbia's Writing Center? Scheduled a consultation with them or 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!

