

Storytelling with Data

Module 6: The storytelling process

Scott Spencer

Faculty and Lecturer
Columbia University



Unanswered, or new, questions from discussion?

Agenda

Upcoming deliverable — **final** storyboard

Today's objectives

Stories, scope and process

Stories, audience and narratives

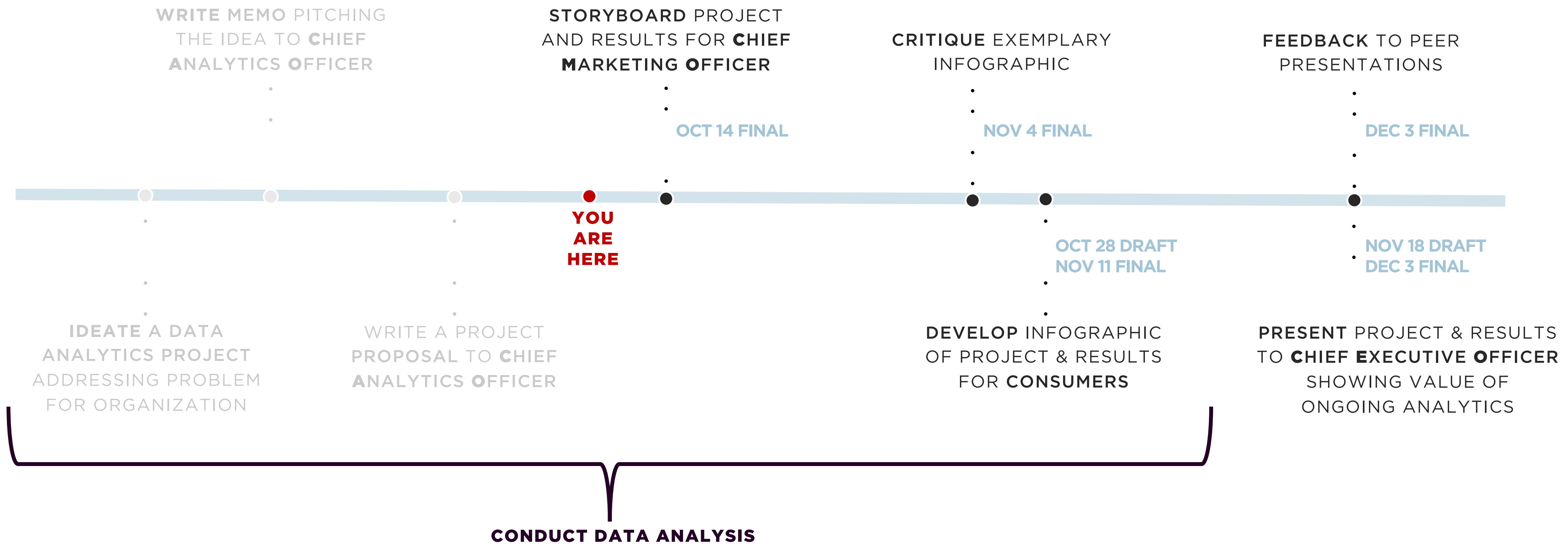
Stories, in words and pictures

Back to storyboards,
tools for testing narratives

Upcoming deliverable

Upcoming deliverable

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



Today's Objectives

Objectives

1

Build a story around an analytics project.

2

Craft a narrative to captivate, convince, and inspire an audience.

Stories – scope and process

Transforming data into visually shared stories

Lee & co-author

Authors Lee and Riche, who was also authors of Data-Driven Storytelling, work at Microsoft, advancing research on data stories.



Scope of a visual data story

A visual data stories are more than a visualization of data. They include a set of story pieces — that is, **specific facts backed up by data**.

They are **visualized** to support intended messages, include **annotations** (labels, pointers, text) and **narration** to highlight and emphasize the message and avoid ambiguity.

They are presented in a **meaningful order** or with a connection between them to support the author's high-level communication goal.

From data and visuals to visual story

To make a story, the data excerpts gathered in step one (e.g., raw data and charts) need to be **assembled into a storyline** that is interesting, illuminating, and compelling.

The **sequence** plays a critical role in a story; the same set of excerpts can have impact or can fall flat. A significant part of making a story is the process of constructing the storyline or plot.

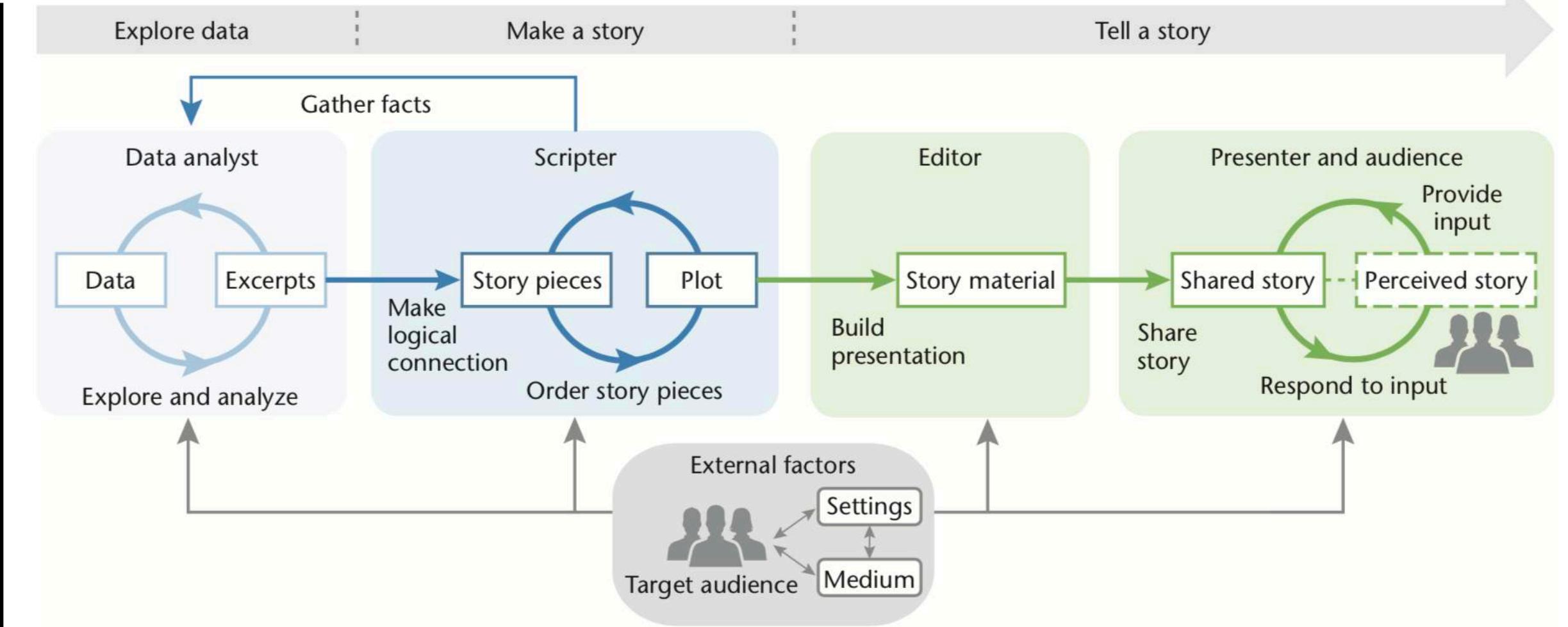
The **activities involved are ordering, establishing logical connections, developing flow, formulating a message, and creating the denouement**.



Transforming data into visually shared stories

Lee & co-author

Authors Lee and Riche, who was also authors of Data-Driven Storytelling, work at Microsoft, advancing research on data stories.



Open research questions in evaluating visual data stories remain

Understanding why a story is or is not effective may require an answer that goes beyond looking at story delivery and the story's reception by the audience.

Stories – why, what, how



Management is much more than a science

Martin & Golsby

Professor Martin is director of the Martin Prosperity Institute at the School of Management, University of Toronto. Golsby-Smith is founder of 2nd Road, a strategy and innovation firm, now part of Accenture Strategy.

Narratives frame hypotheses for what could be

To make decisions about **what could be**, managers should devise narratives about possible futures, applying the tools of metaphor, logic, and emotion.

Consider possibilities broader than data

The **absence of data does not preclude possibility**. If we are talking about new outcomes and behaviors, then there is no prior evidence. Consider not only what the data suggests but also what within the bounds of possibility could happen.

Metaphor helps compare unlinked ideas

The core engine of creative synthesis is “associative fluency”—the mental ability to **connect two concepts that are not usually linked** and to forge them into a new idea.

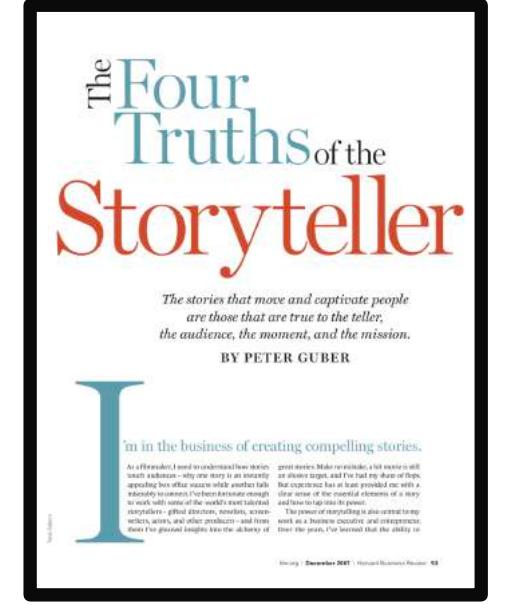
Clarify conditions, create new data

With new narratives—hypotheses—we must **hypothesize** what would have to be true for those narratives to happen and **validate** their hypotheses through prototyping.

The four truths of the storyteller

Guber

Peter is author of *Tell to Win*, has been awarded business person of the year by Los Angeles Business Journal, is owner of several pro sports teams and an entertainment company. He was top executive at Sony and Columbia Pictures.



Persuasive stories depend on four truths

Great storytelling does not conflict with truth. Stories that move people are true to the teller, audience, moment, and mission.

Teller. The teller must be congruent with the story, must share emotions felt, allowing the audience to identify with those emotions.

Audience. As a teller, we must take time to understand what our listeners know about, care about, and want to hear. And then, craft a story that resonates with what we learn. Test the story on others, skeptics. Include audience in the tale—/ becomes we.

Moment. Know the tale well enough to improvise, molding it to the moment's context.

Mission. The tale should capture the mission, and truth to the mission should win over any conflicts with truth to the audience.

Telling tales

Denning

Stephen is the former director of Knowledge Management at the World Bank, an author, and focuses on leadership strategy.



A story — narrative as sequence of events

Narrative patterns for business

Although good business arguments are developed through the use of numbers, they are typically approved on the basis of a story—that is, a narrative that links a set of events in some kind of causal sequence.

Stories can advance business objectives, e.g.:

Sparking action

Communicating who you are

Transmitting values

Fostering collaboration

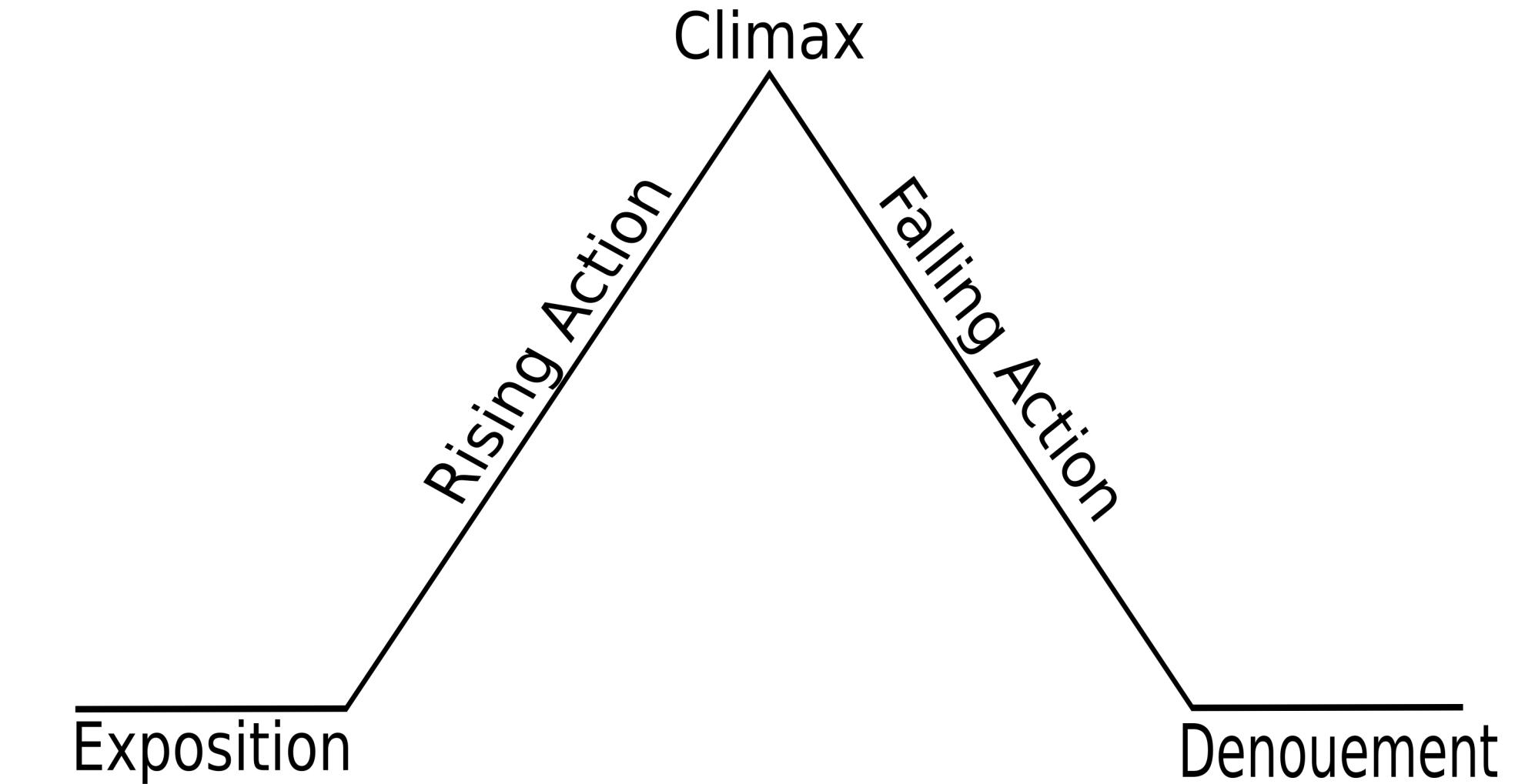
Taming the grapevine

Sharing knowledge

Leading people into the future

Stories – more about narrative and audience

One of many ways to narrate analytics projects



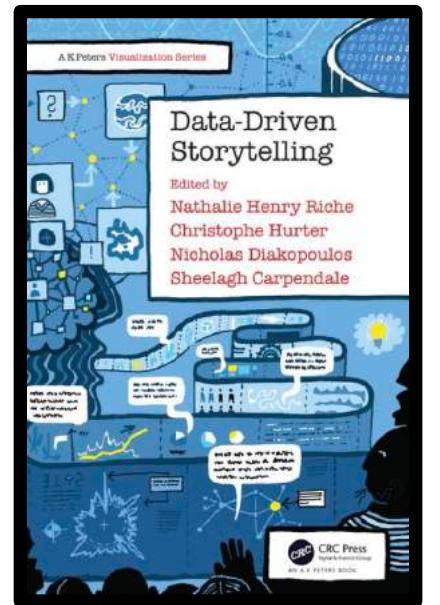
The primary elements of an applied analytics project are a well-articulated business problem, a data science solution, and a measurable outcome to produce value for the organization.

The analytics project may thus be conceptualized as a story arc, with a beginning (**problem**), middle (**analytics**), and end (**overcoming of the problem**), along with characters (**analysts, colleagues, clients**) who play important roles.

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.



Narrative in time

The narrative can be linear, e.g., representing the actual order of events (the story) **chronologically**; or **nonlinear**, e.g., **circular**, where the beginning and end mirror themselves.

Narrative using many other ways

A single story can be narrated in an endless number of ways. **Beyond the temporal ordering** of events, it can be told at a varying pace to create emphasis:

slowing down the narrative pace can give the audience time to think about what is being “told” and to increase the importance of this moment in the story; **speeding up** the narrative pace can produce an overwhelming effect of information onslaught and reduce the importance of moments in the story.

Arguments can be made **explicit** or kept **implicit**.

The narrative can emphasize **comparison** and **contrasting** points of view; it can convey abstract concepts through the use of **analogies**, making them more concrete; it can address and engage the audience directly with abrupt **questions**, while interrupting the flow of the story; or it can draw the audience in so that it becomes a part of the story, releasing it with a call-to-action.

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.



Narrative “Patterns”

A narrative pattern is a low-level narrative device that serves a specific intent. A pattern can be used individually or in combination with others to give form to a story.

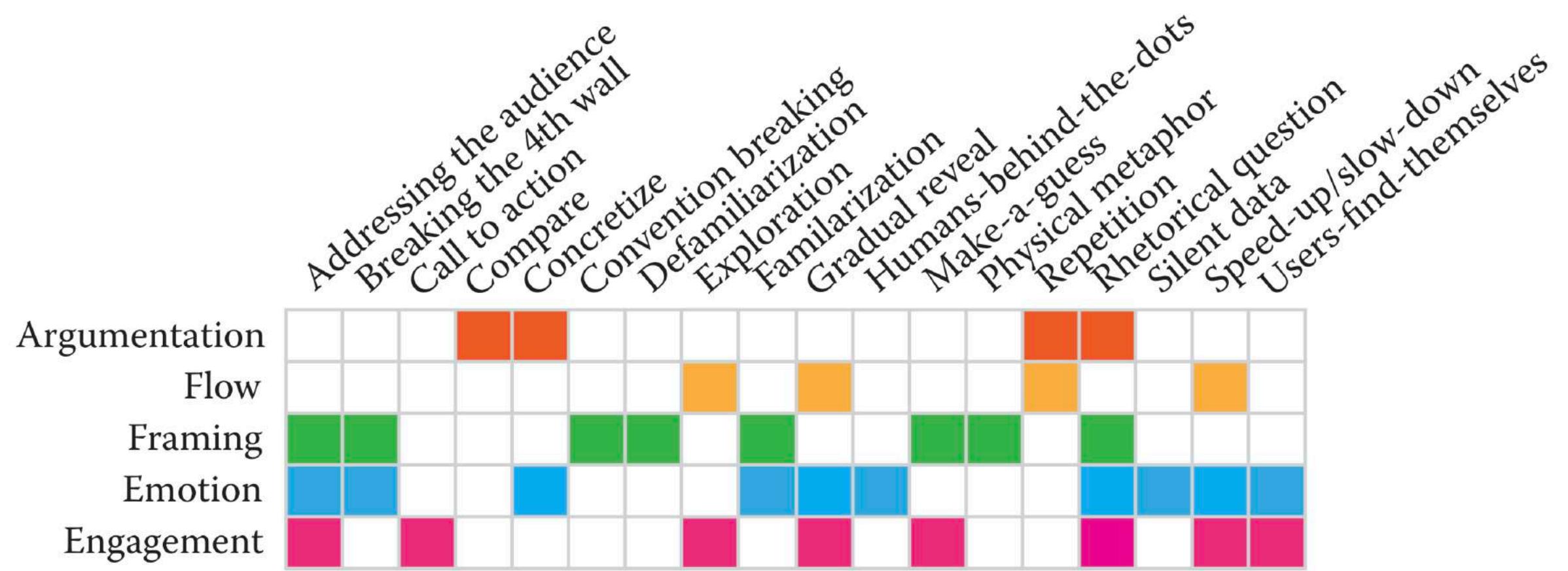
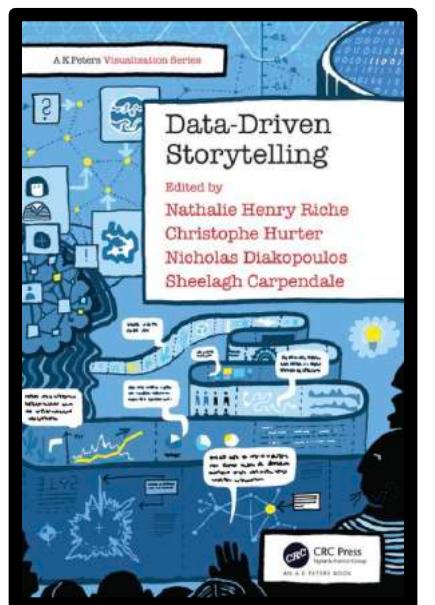


FIGURE 5.2 18 narrative patterns grouped into 5 major pattern groups.

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.



Patterns for argument

Argumentation is the action or process of reasoning systematically in support of an idea, action, or theory. We previously discussed its components (**ethos, pathos, logos**) and some methods (**compare, concretize, repetition**).

Patterns for flow

These are devices that help structure the **sequencing** of messages and arguments. They are essential to every story, as they set the **order, rhythm, and pace**, and help **buildup the climax** and **resolution**. Examples include **reveal, slowing down, speeding up**.

Recall the news article, *What is Really Warming the World?* by Bloomberg? It sets flow by gradual reveal.

Patterns for framing

Framing builds **the way facts and events in a story are perceived and understood** through narration. It feeds on the audience's expectations, but it can also play with those expectations and go against them to create surprise.



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.

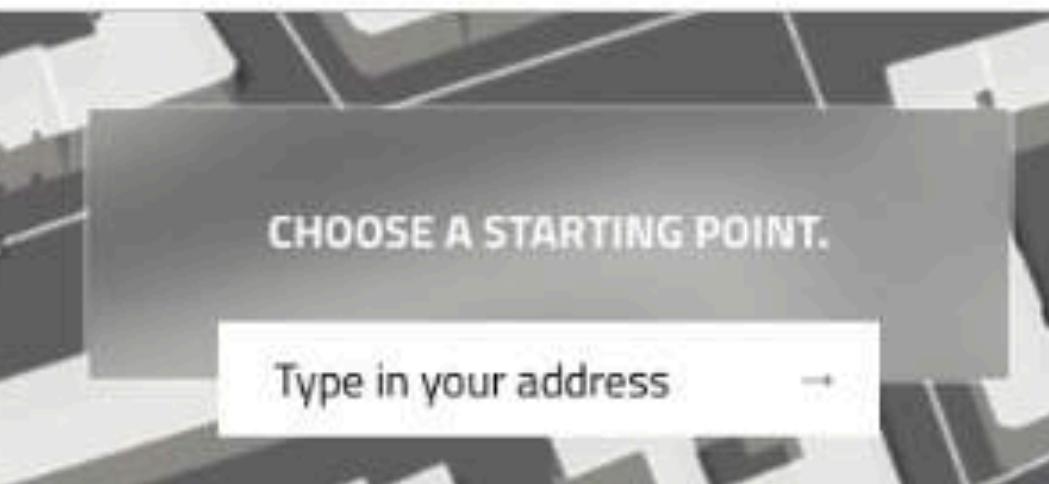
Patterns for empathy, emotion

These patterns are for resituating and reorienting the reader's perspective while **enhancing our ability to understand and share the feelings and experiences important in the story.**

Patterns for engagement

Engagement can be seen as **the feeling of being part of the story**, of **being connected** to it, and being in control over the interactions with the story's content.

Examples include **asking** the audience a question; their answering makes them part of the story — *do you agree?* Or asking them to **imagine themselves** in our characters' shoes.

Incorporating the audience**HOW**

Audience becomes part of the narrative.

WHY

Increase empathy, facilitate personal comparisons. Helps make the data and information more personally relevant. To allow reader to use herself as measuring device.

EXAMPLE

Ubisoft: Collapse

Repetition**HOW**

A phenomenon is re-presented multiple times with changes to the main variable/dimension (the data) or the time frame or the animation pattern (see the example) while other elements of the story don't change. Note: this could be like running multiple simulations with the data to see which is most effective (and in this case, the story would be more reader driven)

WHY

To show that the same phenomenon happens over and over again. It also strengthens the narrative through rhythm. Establishing a constant reference frame can help to emphasize change and differences.

EXAMPLE

Bloomberg: What's really warming the world?

Juxtaposition**HOW**

Showing two or more complementary visualizations (pictures) juxtaposed.

WHY

Compare cases or samples, allows to highlight contrast, difference, absence of difference, change over time. Allows for individual exploration of (minor) differences.

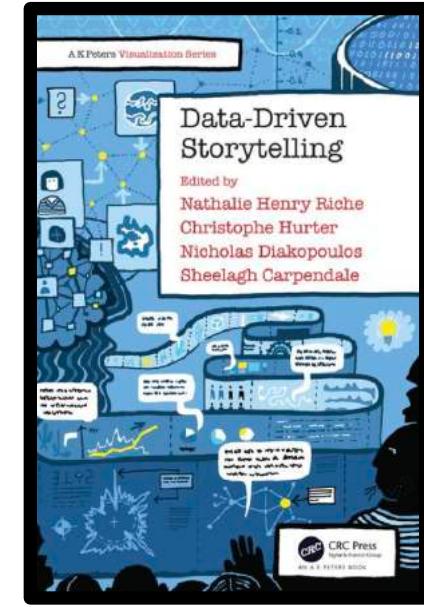
EXAMPLE

David McCandless: Billion Dollar Gram

Breaking the fourth wall**Humans behind the dots****Make a guess**

Draw your line on the chart below





Communicating data to an audience

Riche, co-editors

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

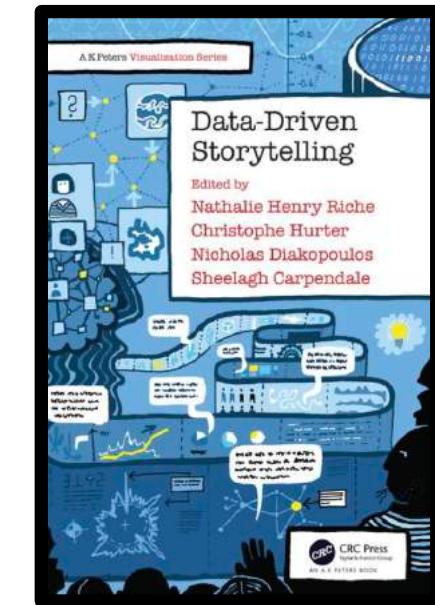
Characteristics of audience background knowledge

1. How **literate** is the audience in terms of data and visualization? And what can be done to increase it?
2. How knowledgeable is the audience in terms of jargon, domain **expertise**, and other background knowledge?
3. What are audience expectations about the **design** of visualization, such as style, tone, or the use of iconography?

Communicating data to an audience

Riche, co-editors

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



Characteristics of audience background knowledge

1. How **literate** is the audience in terms of data and visualization? And what can be done to increase it?

Support data literacy through annotations that *explain*

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

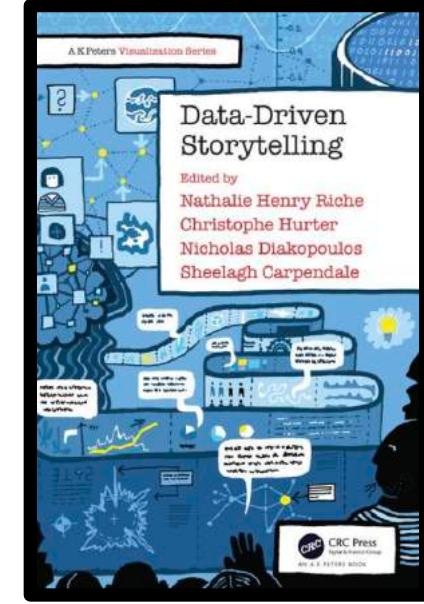
They can also **explain how to read the graph**, which helps readers unfamiliar with the graph—whether a simple line chart or an advanced technique like a treemap or scatterplot.

When done right, the annotation layer will not get in the way for experienced users.

Communicating data to an audience

Riche, co-editors

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



Characteristics of audience background knowledge

Research your audiences' likely baseline knowledge

Minimize jargon

Consider their viewpoints, biases

2. How knowledgeable is the audience in terms of jargon, domain **expertise**, and other background knowledge?

Data storytelling needs to take into account the **baseline of information and knowledge** that the intended audience is expected or assumed to have.

Niche audiences may be more comfortable with the jargon of their group, but designers may consider whether **removing jargon** would serve a larger audience just as easily while **making the content more broadly accessible**.

An individual in the audience brings their own viewpoints, backgrounds, and experience to each and every data-driven story they consume. Whether driven by their cultural background, social position, education, or other demographic characteristics, **readers carry with them their own unique set of knowledge and biases**.

Communicating data to an audience

Riche, co-editors

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



Characteristics of audience background knowledge

Design consistency aids understanding

3. What are audience expectations about the **design** of visualization, such as style, tone, or the use of iconography?

A **consistent design approach** (one might call it “visual branding”) can **help an audience relate to a producer’s work** and entice them to return.

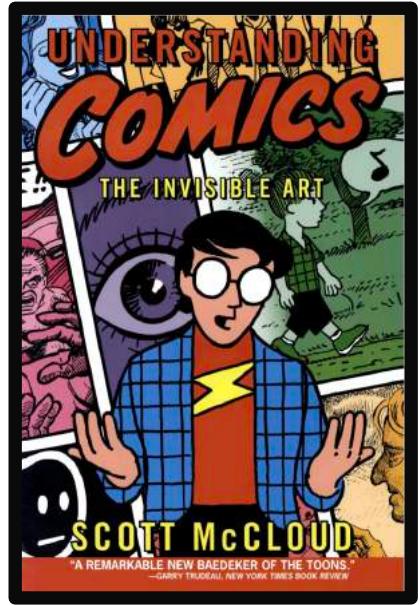
Consistency across color, font, layout, and type of interactivity, for example, can help the audience better identify what the producer is going to deliver, when, and for what level of expertise.

Conventions used in presentations can also reinforce good data hygiene and inform the audience as to the quality of a visualization such as when source and designer credits or methodologies are consistently incorporated into the story delivery.

Stories, in pictures and words

Edward Tufte, in *Visual Explanations* — his book about pictures of verbs to show causes and effects, explanations and narratives — cites to **comics** for understanding the idea of “**visual narrative**”.

Comics tell stories using
visual components, corresponding
written narratives, and **sequencing**.



Understanding Comics

McCloud

Scott provides the seminal reference, in the context of comics, on understanding the connection between words and visuals.

**Words and visuals
have shared meaning**

In comics, visuals and words complement one another, each contributing to a shared meaning.

Story in sentences

I CROSSED THE
STREET TO THE
CONVENIENCE STORE.
THE RAIN SOAKED
INTO MY BOOTS.

I FOUND THE
LAST PINT OF
CHOCOLATE
CHOCOLATE CHIP
IN THE FREEZER.

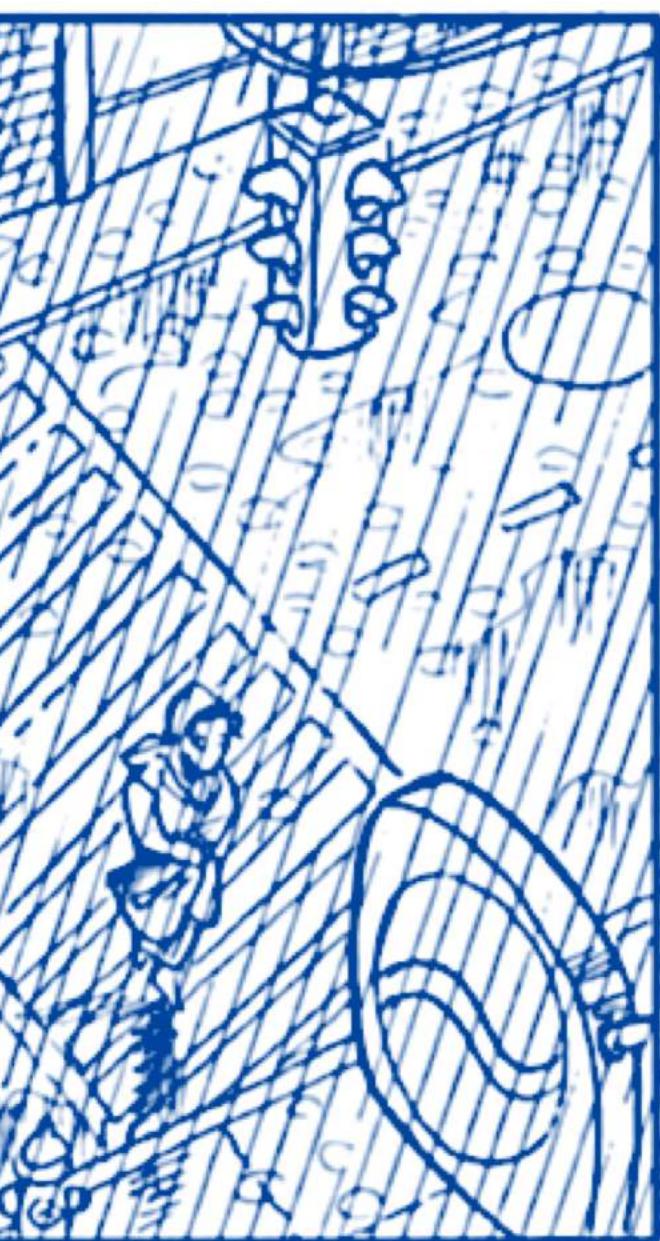
THE CLERK TRIED TO PICK ME UP. I SAID NO THANKS. HE GAVE ME THIS CREEPY LOOK...

I WENT BACK TO THE
APARTMENT--

--AND FINISHED IT ALL IN AN HOUR.

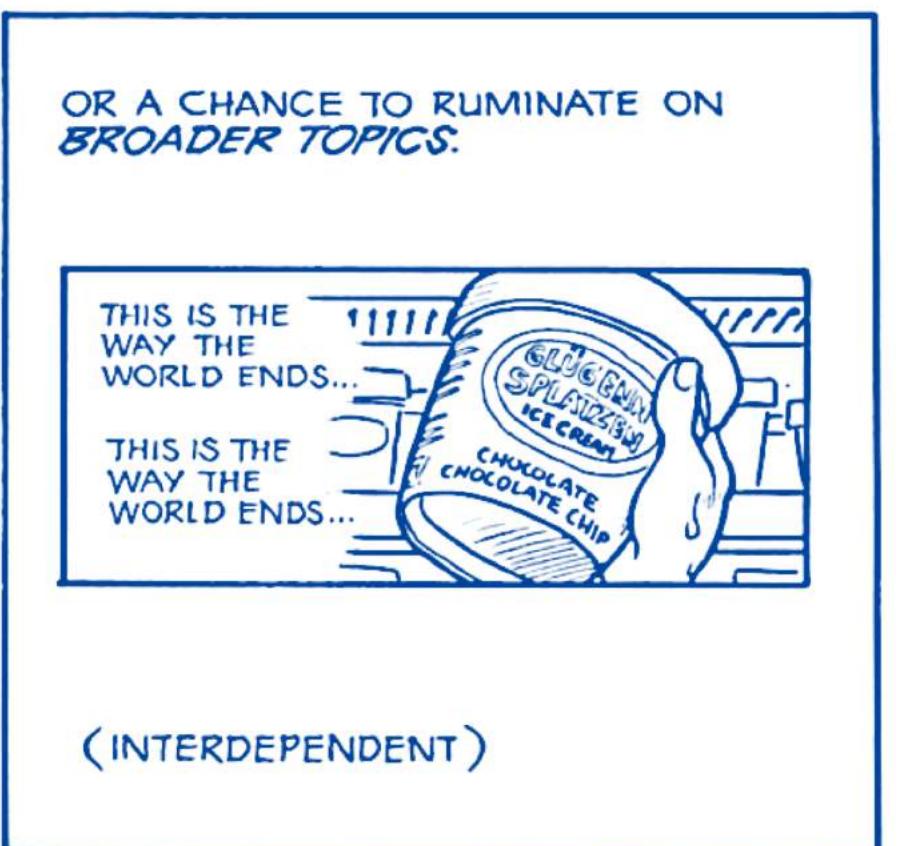
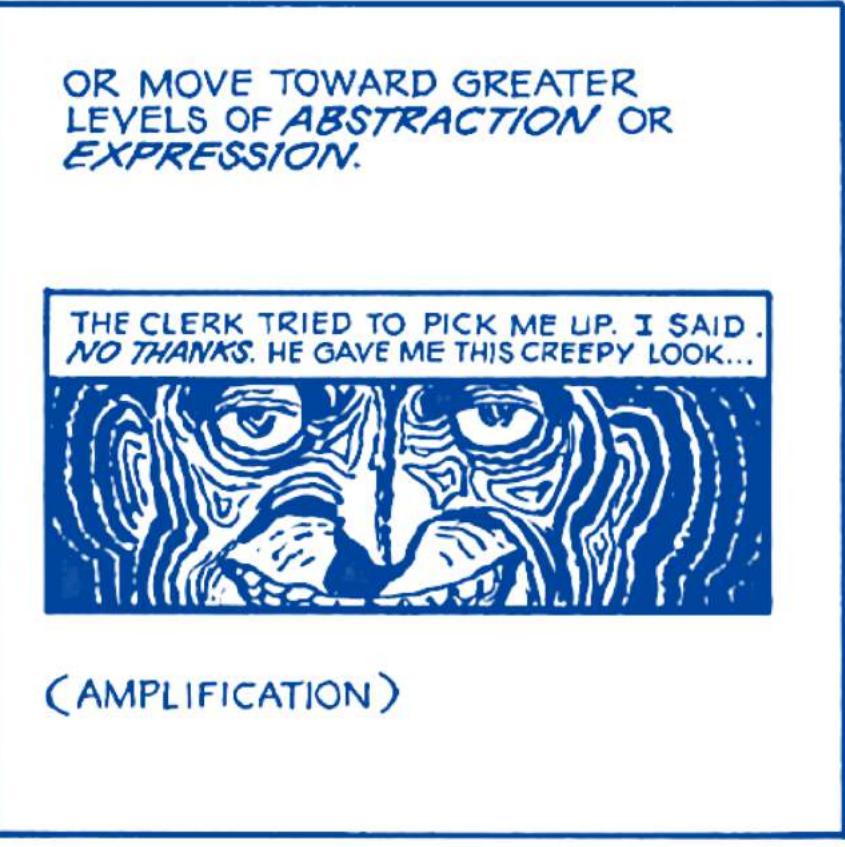
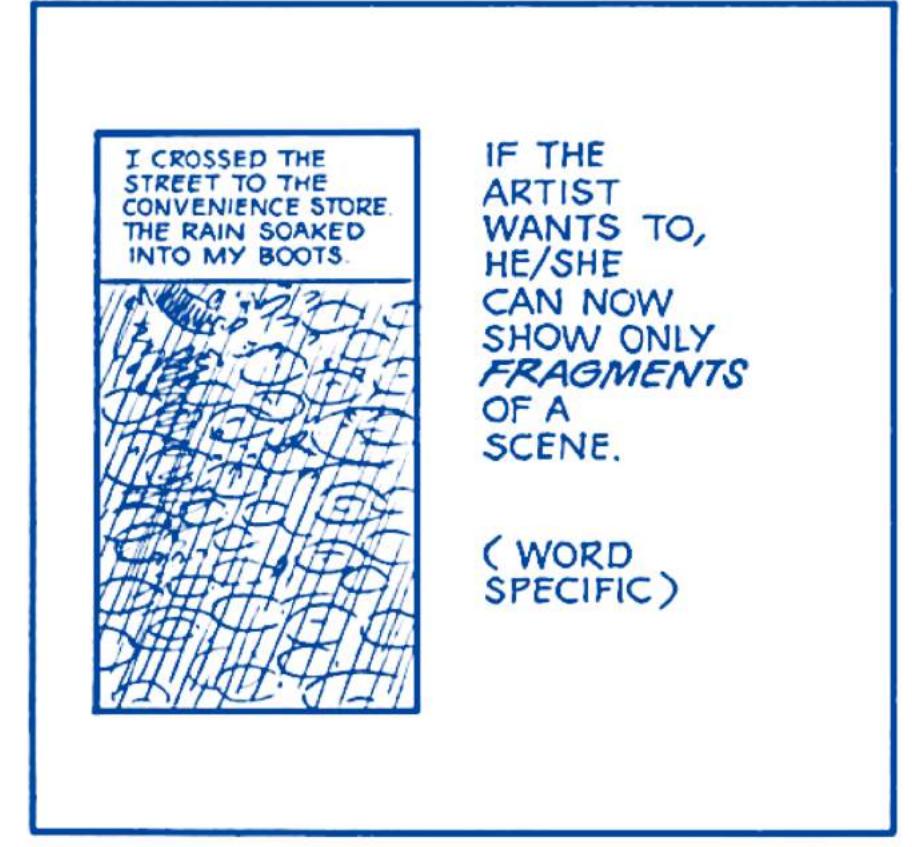
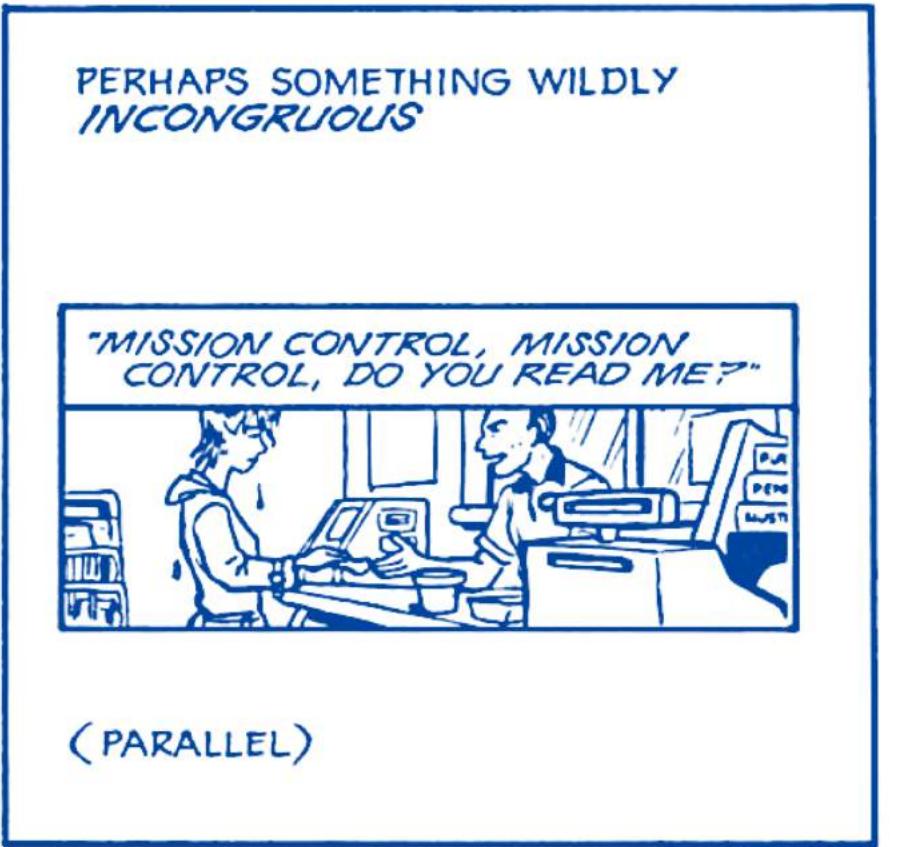
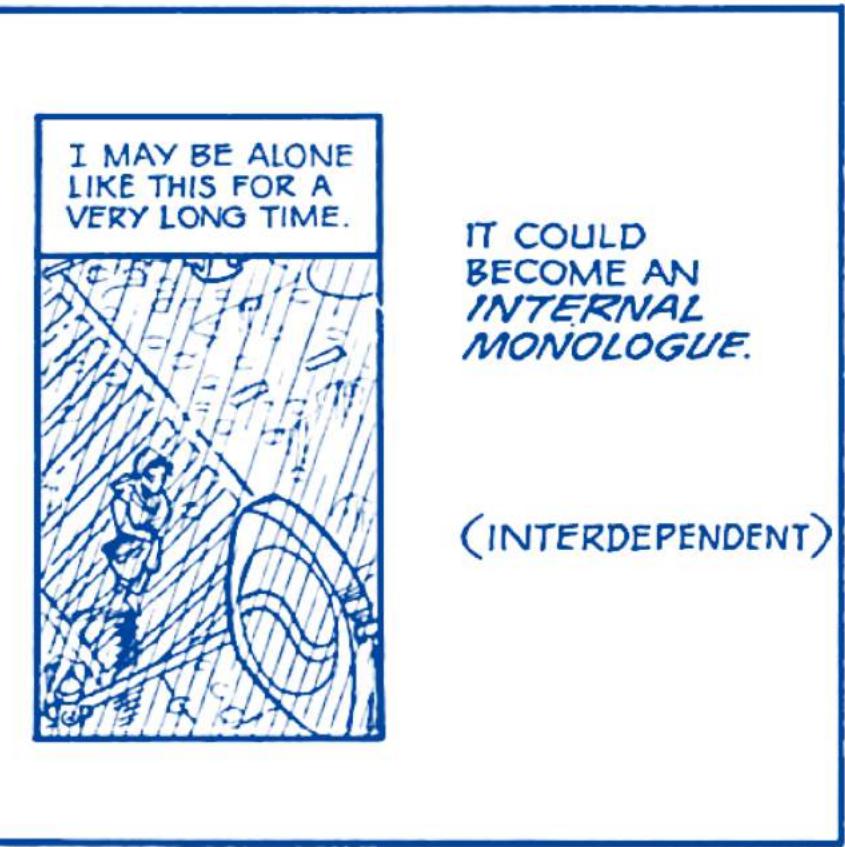
ALONE
AT LAST.

Story in pictures





Use words and images for different purposes



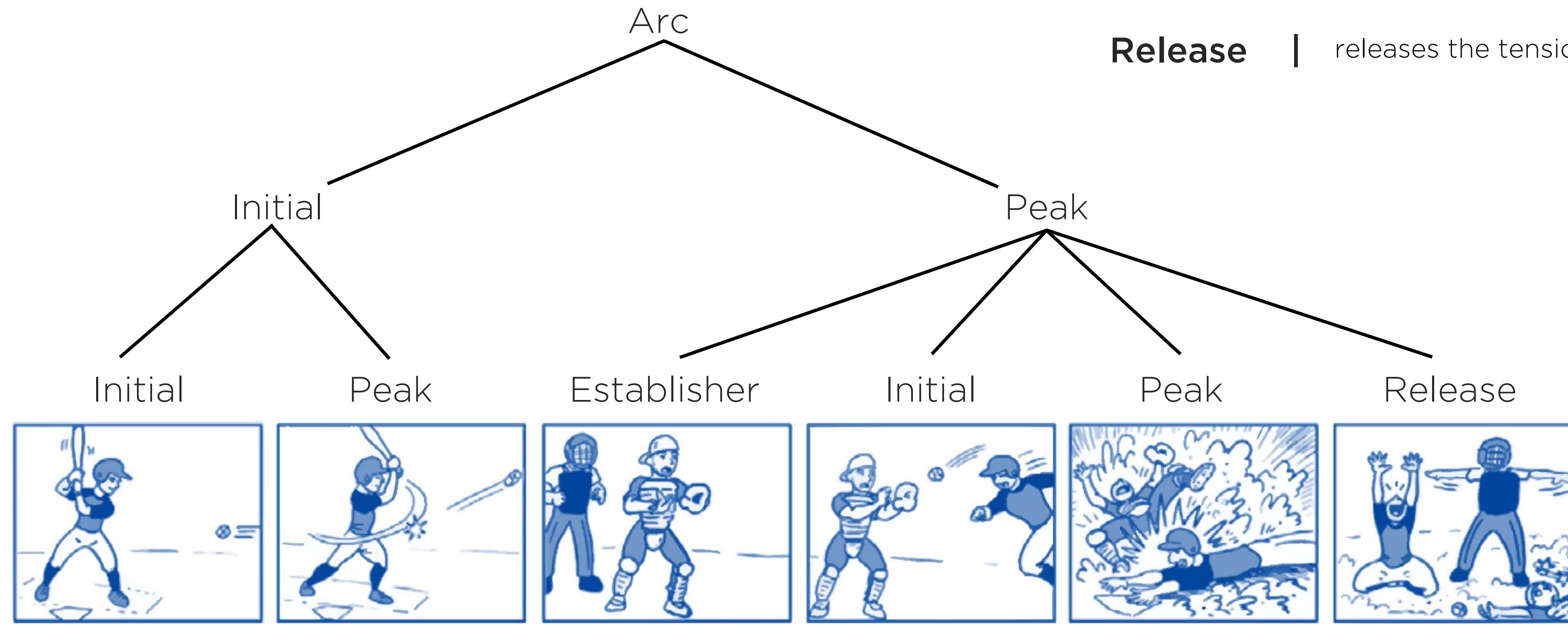
Like written narrative, sequential pictures can have grammar

Establisher | sets up an interaction without acting upon it

Initial | initiates the tension of the narrative arc

Peak | marks the height of narrative tension and point of maximal event structure

Release | releases the tension of the interaction



INTERDEPENDENT COMBINATIONS AREN'T ALWAYS AN EQUAL BALANCE THOUGH AND MAY FALL ANYWHERE ON A SCALE BETWEEN TYPES ONE AND TWO.

GENERALLY SPEAKING, THE MORE IS SAID WITH WORDS, THE MORE THE PICTURES CAN BE FREED TO GO EXPLORING AND VICE VERSA.

P

w



W

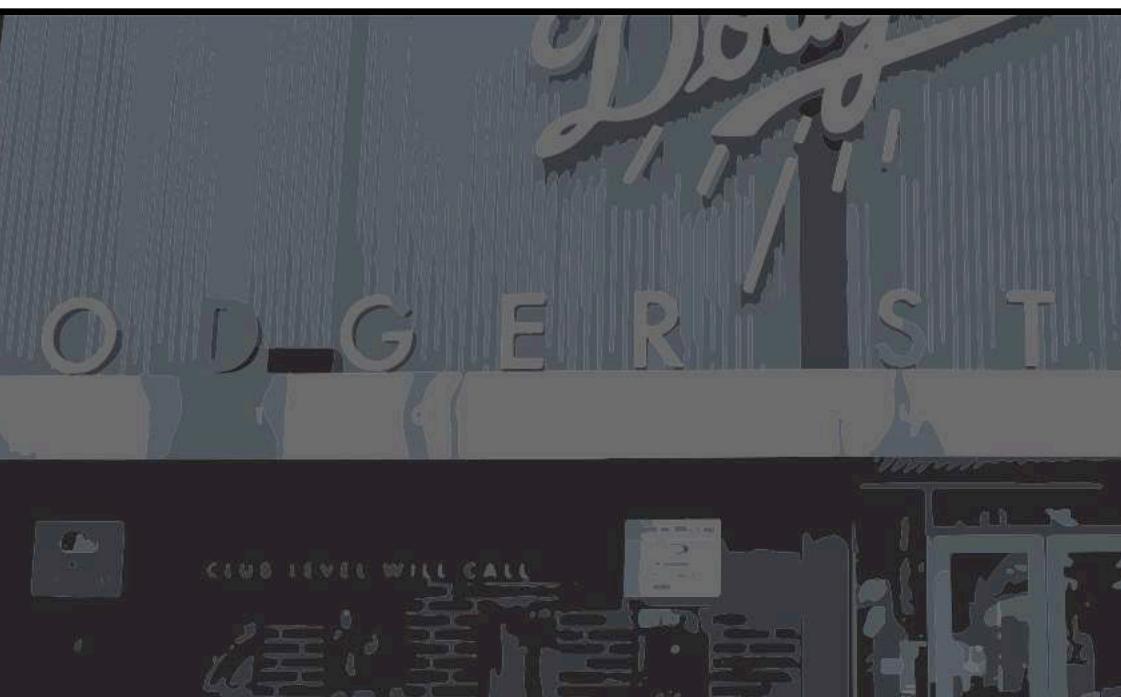
P

“The **balance** between text and visualization becomes an issue when **too much text** takes away from the data but **too little text** leaves the viewer confused and unable to see the connections.”

— Kosara & McKinley, visual analysis at Tableau

Back to storyboards as a tool for testing narrative Dodgers project

Visual narrative



“Uncertainty of outcome is necessary if the consumer is to be willing to pay admission to the game.”
— Simon Rotenberg, 1956

“What is the most exciting play in baseball?

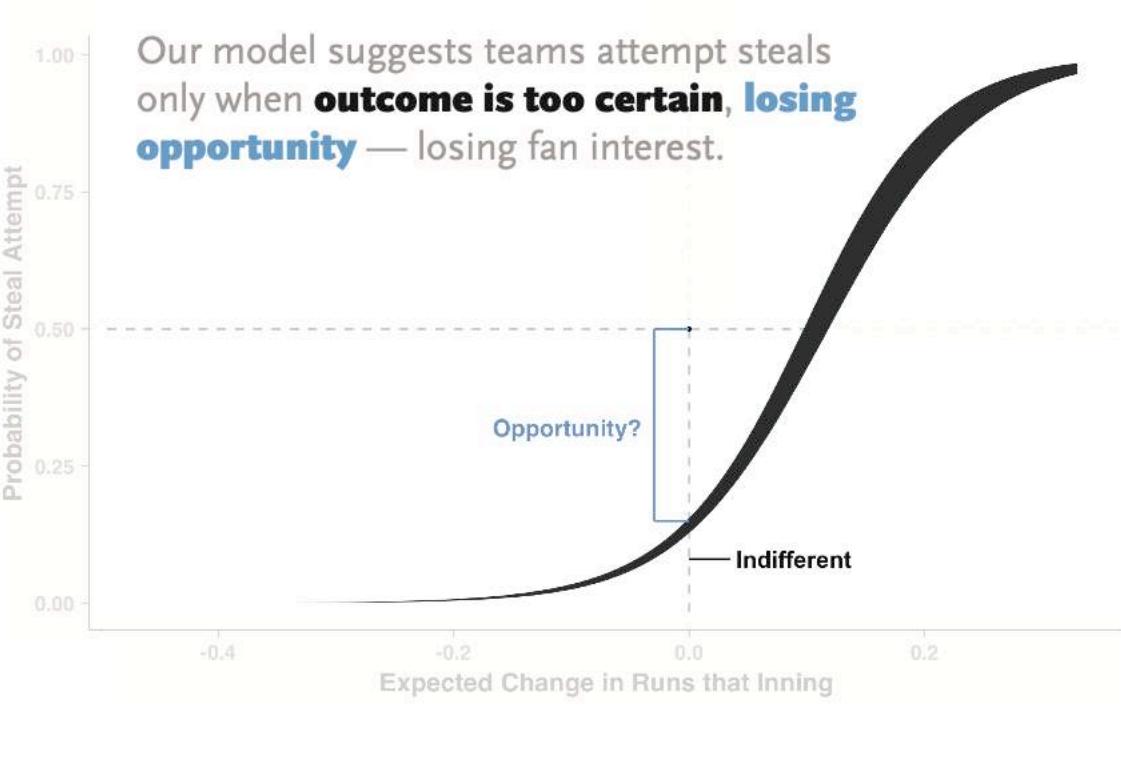
One in which the batter makes a perfect slide and **just beats the throw** right as a crowd is about to go absolutely nuts with joy, anger or amazement, depending on their rooting interests.

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

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



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



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

From a game outcome perspective — whether teams win — managers should be indifferent if a baserunner steals when the expected change in runs is zero. The model suggests baserunners are too conservative.

Written narrative

1

Our average game attendance hovers around mid-80 percent capacity — we have seats to fill, despite having multiple winning seasons.

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

2

We have started modeling the uncertainty of outcome for different events in the game.

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

3

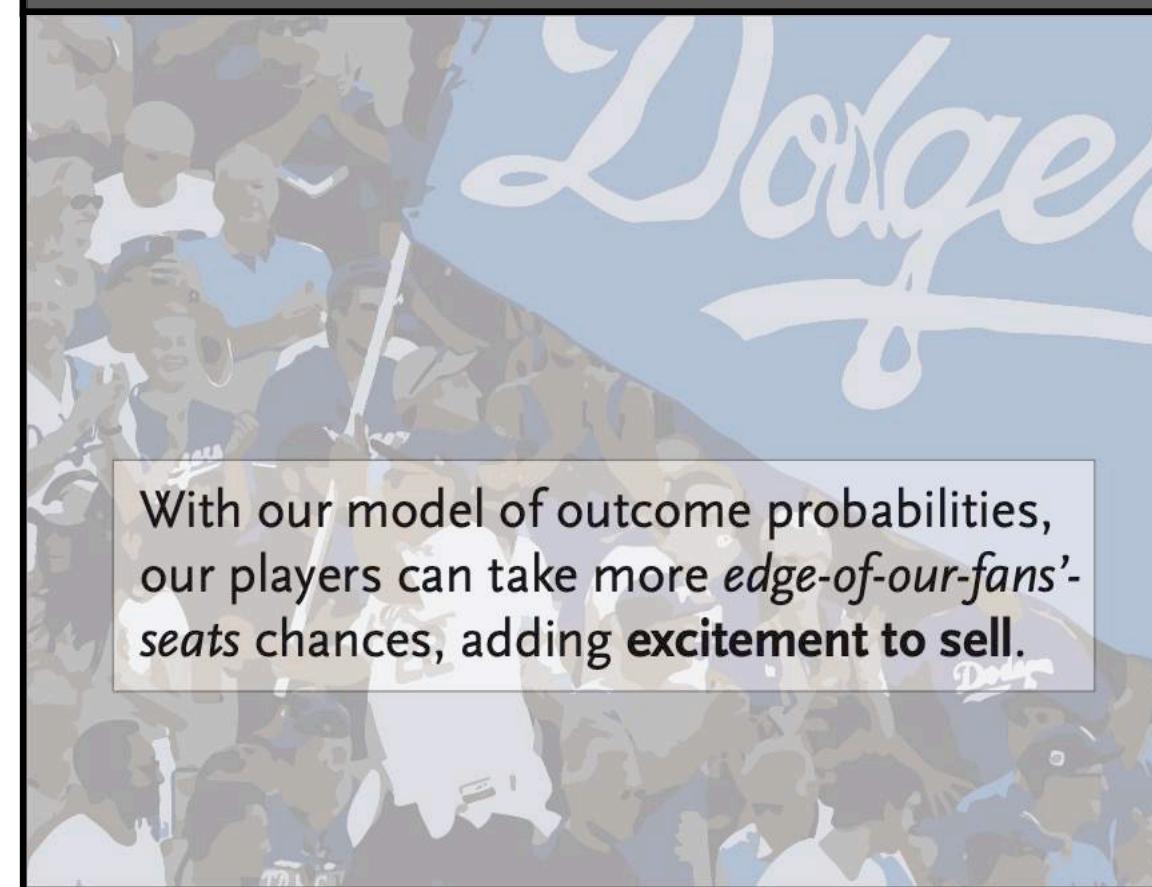
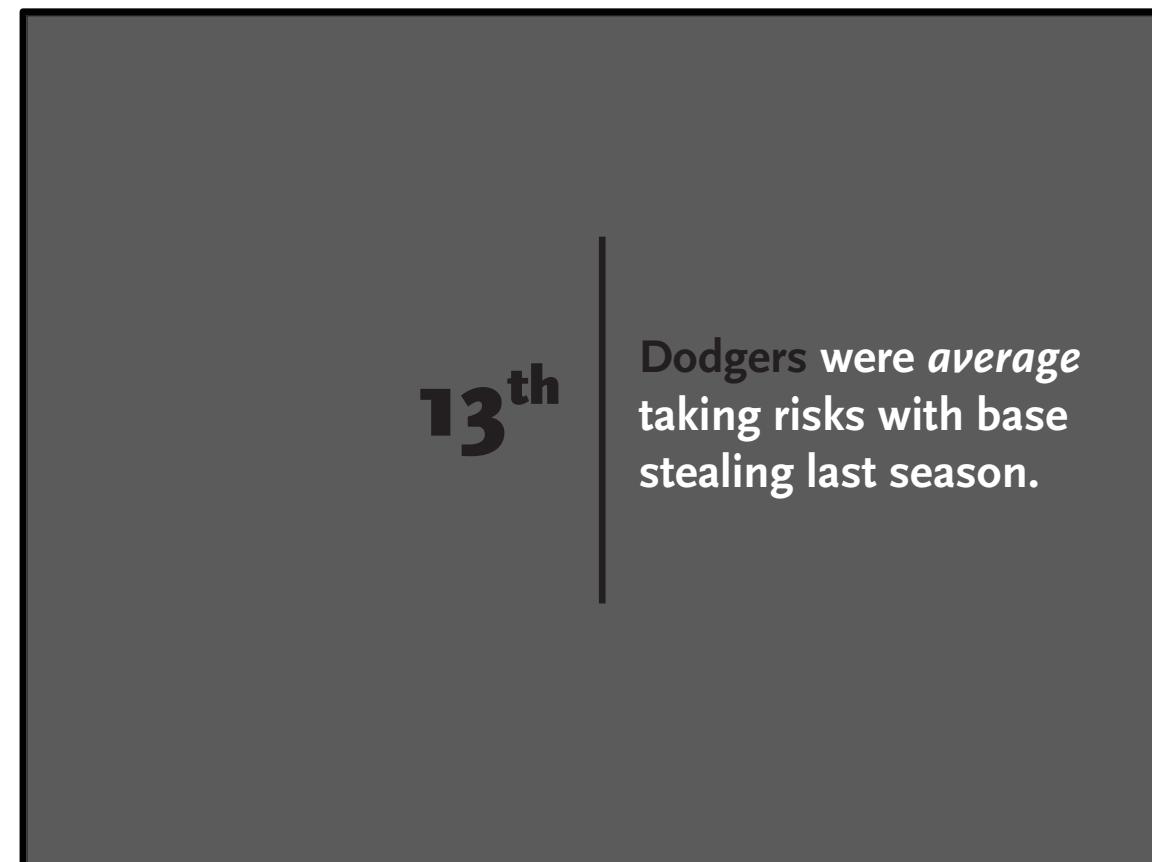
Using our model, we asked whether we are maximizing base-stealing excitement. We considered running speed of the player on first base, catcher throw time to second base, the latent talents of the specific pitcher, catcher, and baserunner, and estimated the probability of successfully stealing second.

From a game outcome perspective — whether teams win — managers should be indifferent if a baserunner steals when the expected change in runs is zero. The model suggests baserunners are too conservative.

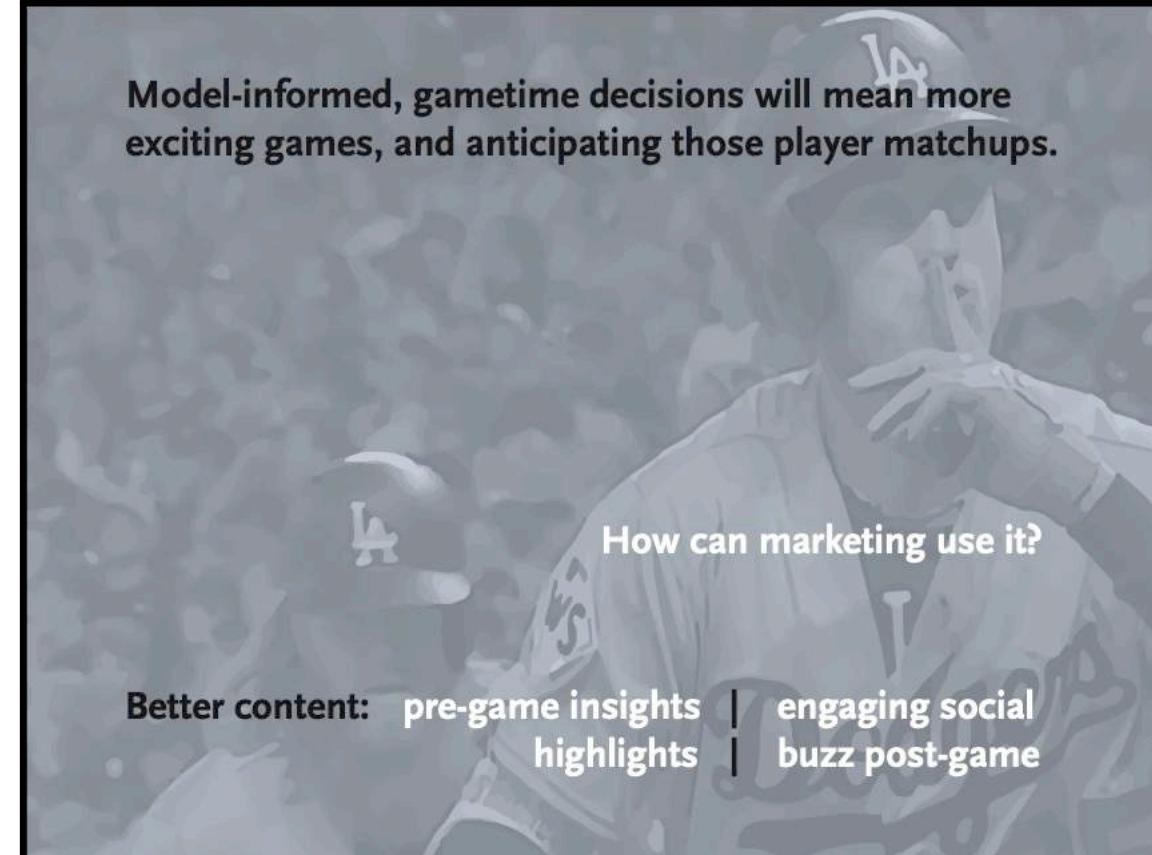
Visual narrative

4

Dodgers were average taking risks with base stealing last season.



With our model of outcome probabilities, our players can take more *edge-of-our-fans'-seats* chances, adding **excitement to sell**.



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

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

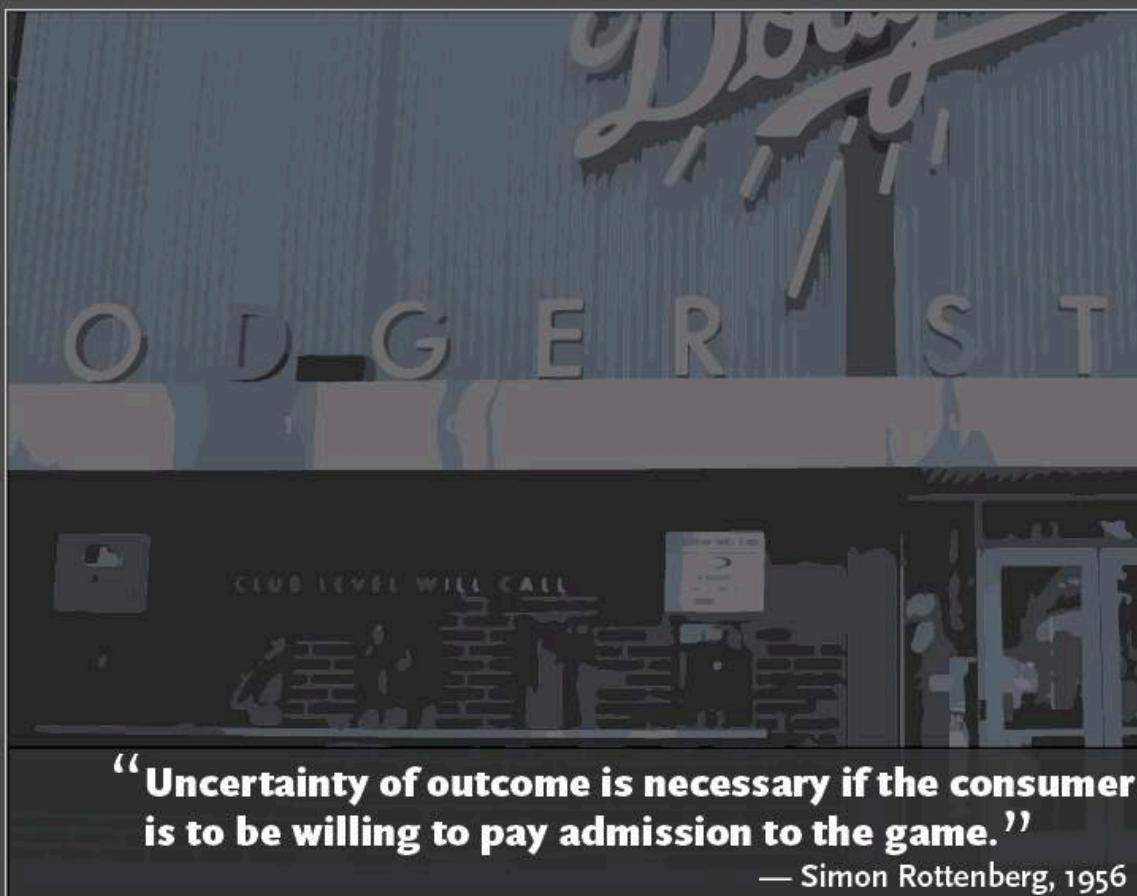
5

Having the uncertainty of each play outcome gives our managers a new tool to tell players to take more risks. The outcome can seem closer to a coin-flip, but in the long run, we can keep our decisions on the winning side. Our model-informed game will be more exciting, making it easier to market, sell tickets, generate buzz.

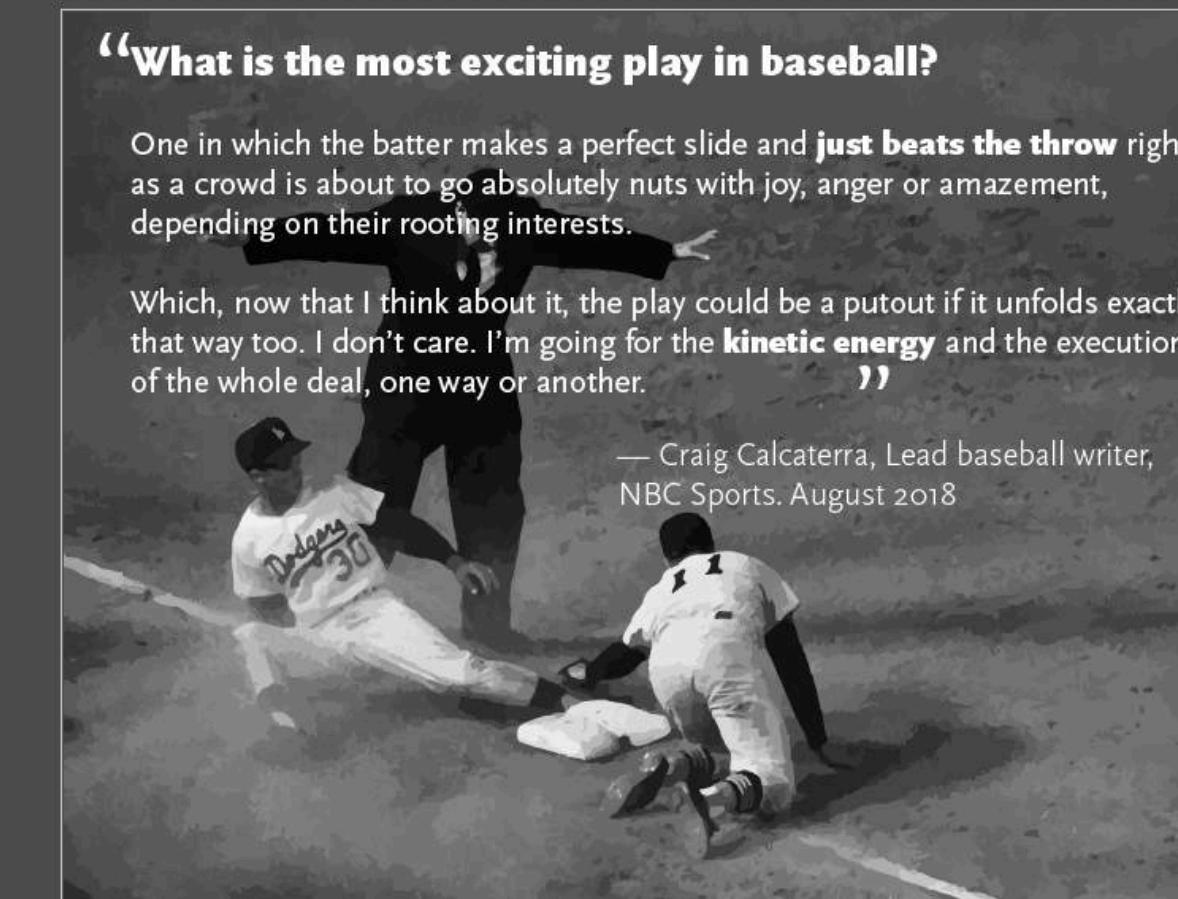
6

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

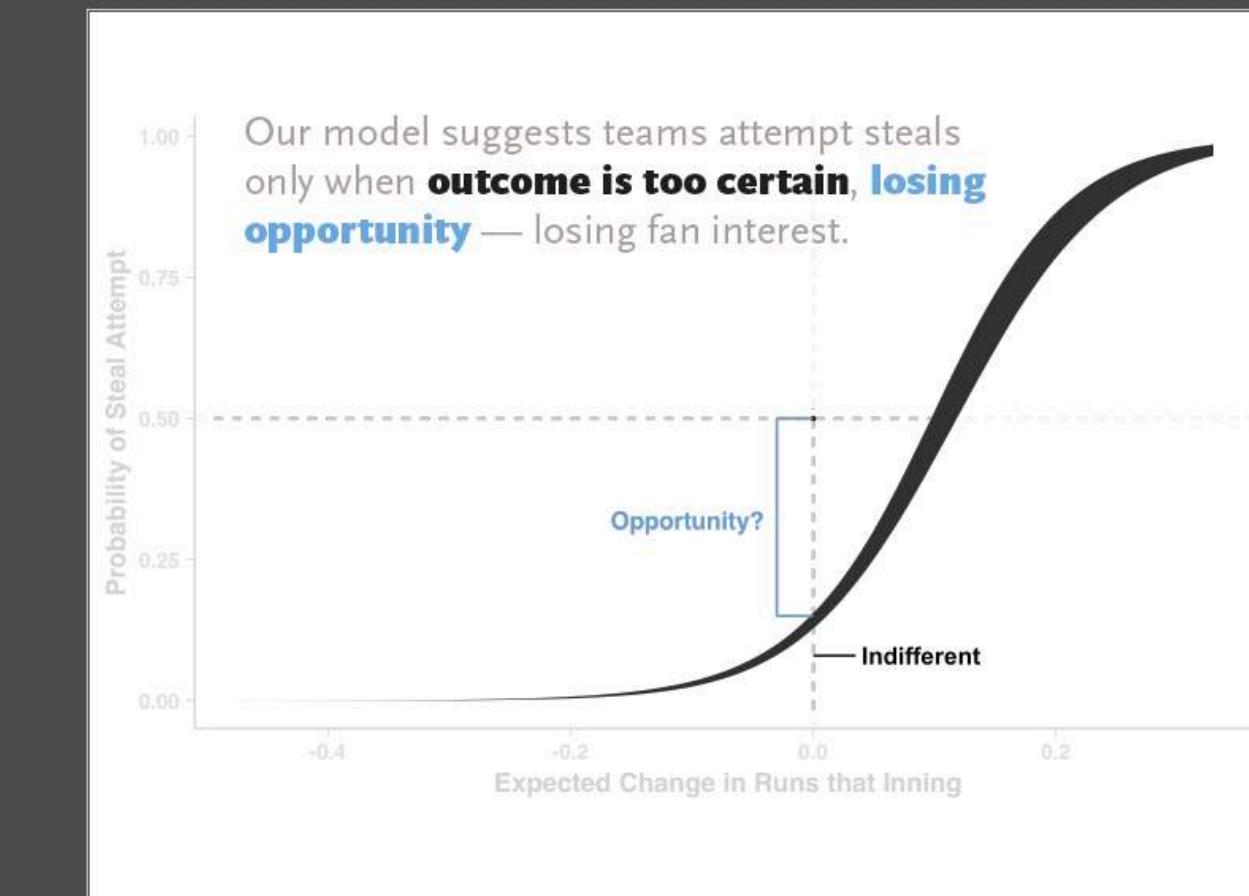
Storyboarding something relevant, with insight, from our preliminary results



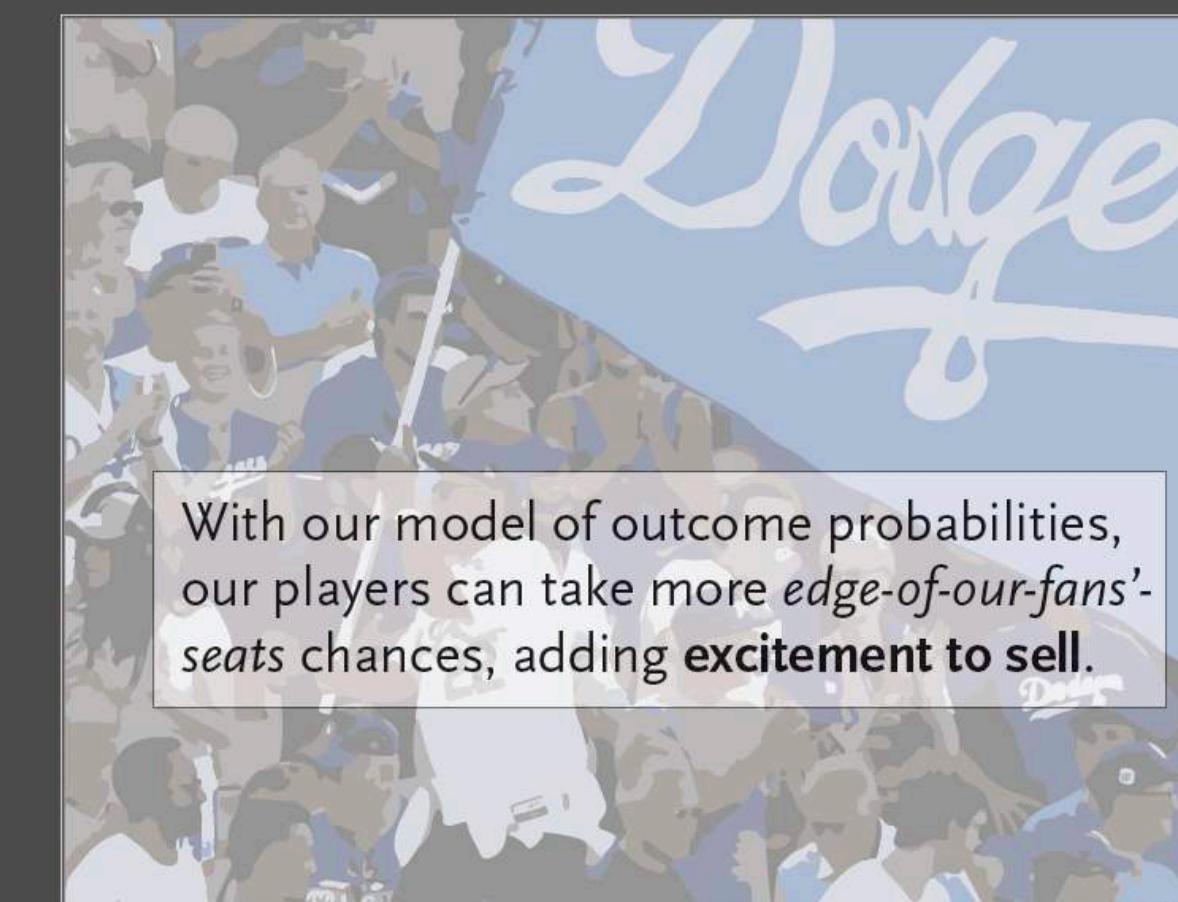
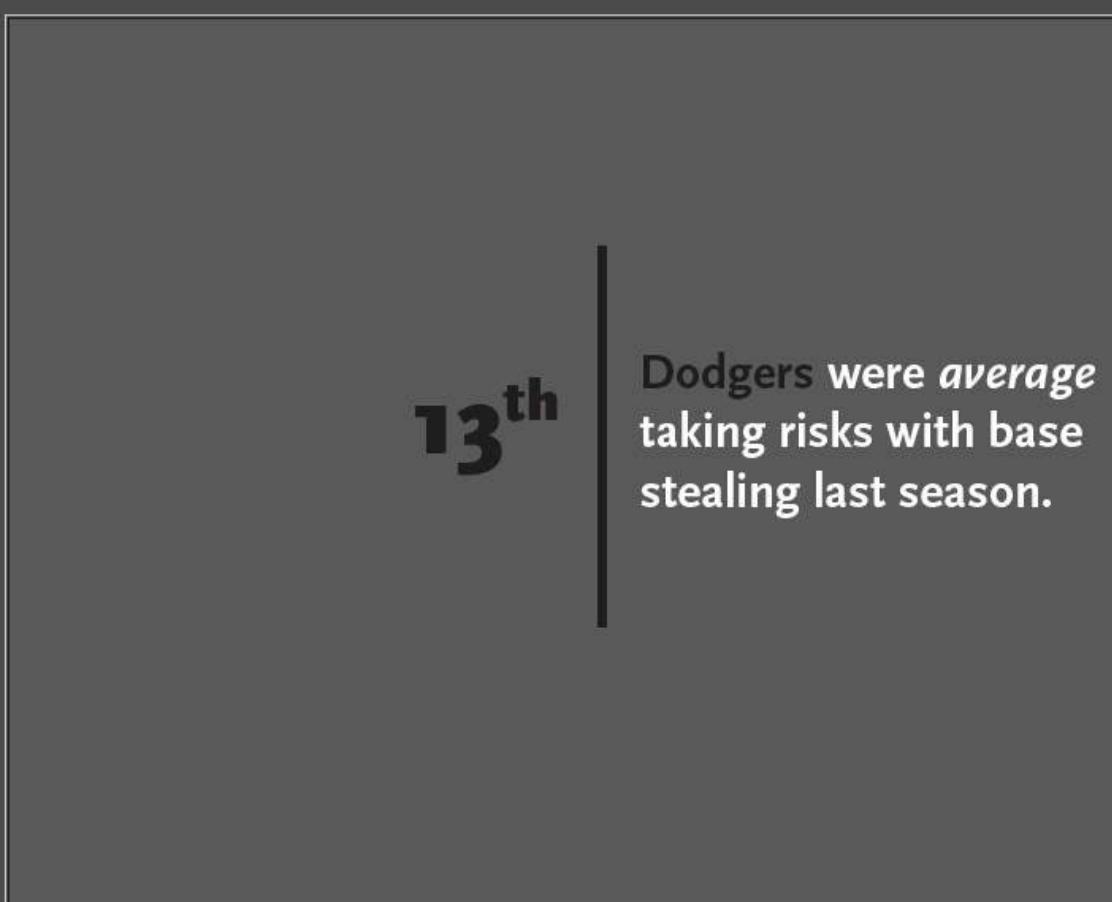
1



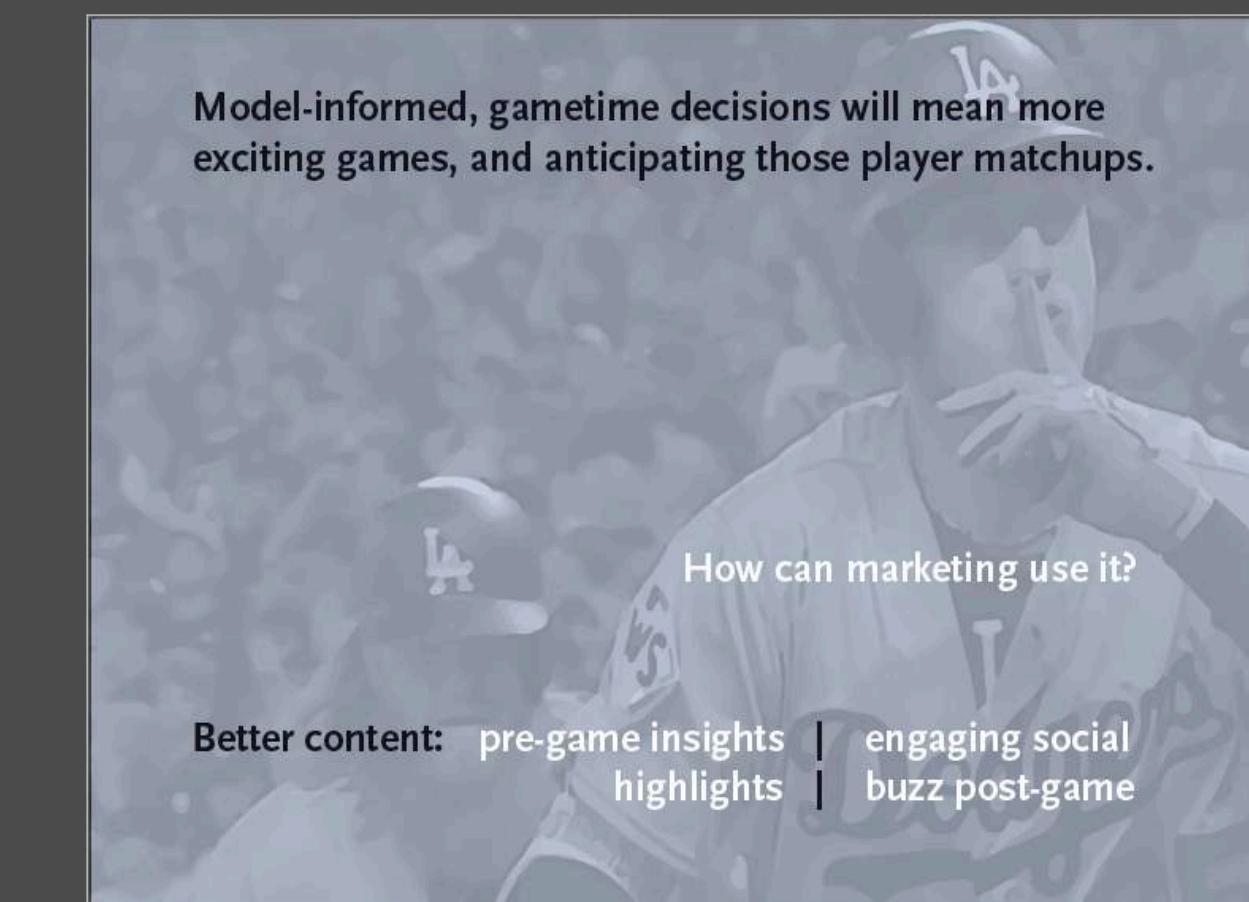
2



3



4



5

Did we show any change? Consider Duarte's tool.

What is

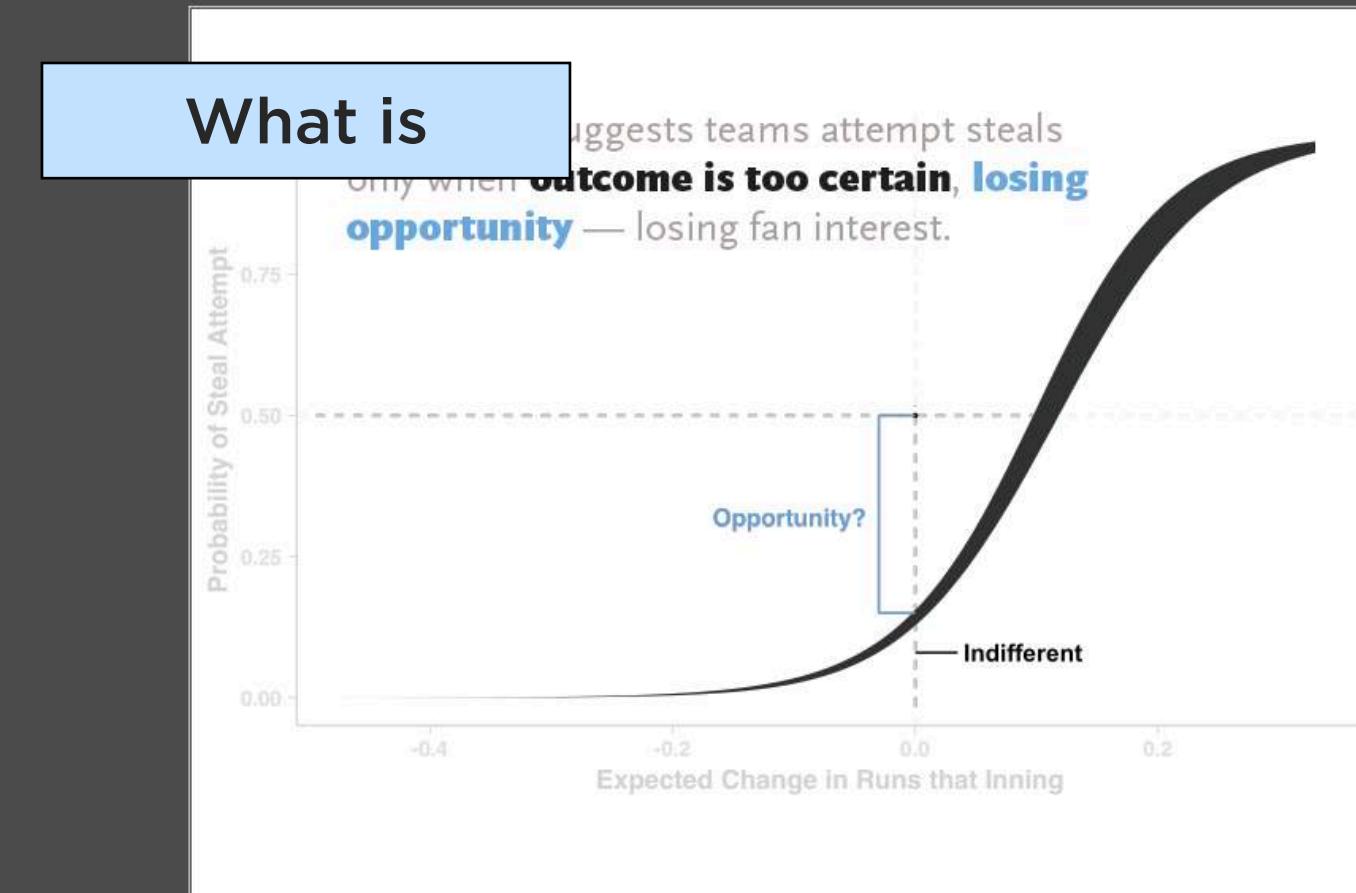
“Uncertainty of outcome is necessary if the consumer is to be willing to pay admission to the game.”
— Simon Rottenberg, 1956

1

“What is the most exciting play in baseball?”
What is? What could be?

Which, now that I think about it, the play could be a putout if it unfolds exactly that way too. I don't care. I'm going for the **kinetic energy** and the execution of the whole deal, one way or another.
— Craig Calcaterra, Lead baseball writer, NBC Sports. August 2018

2



3

What is

13th

Dodgers were *average* taking risks with base stealing last season.

4

What could be

With our model of outcome probabilities, our players can take more *edge-of-our-fans'-seats* chances, adding **excitement to sell.**

5

What could be

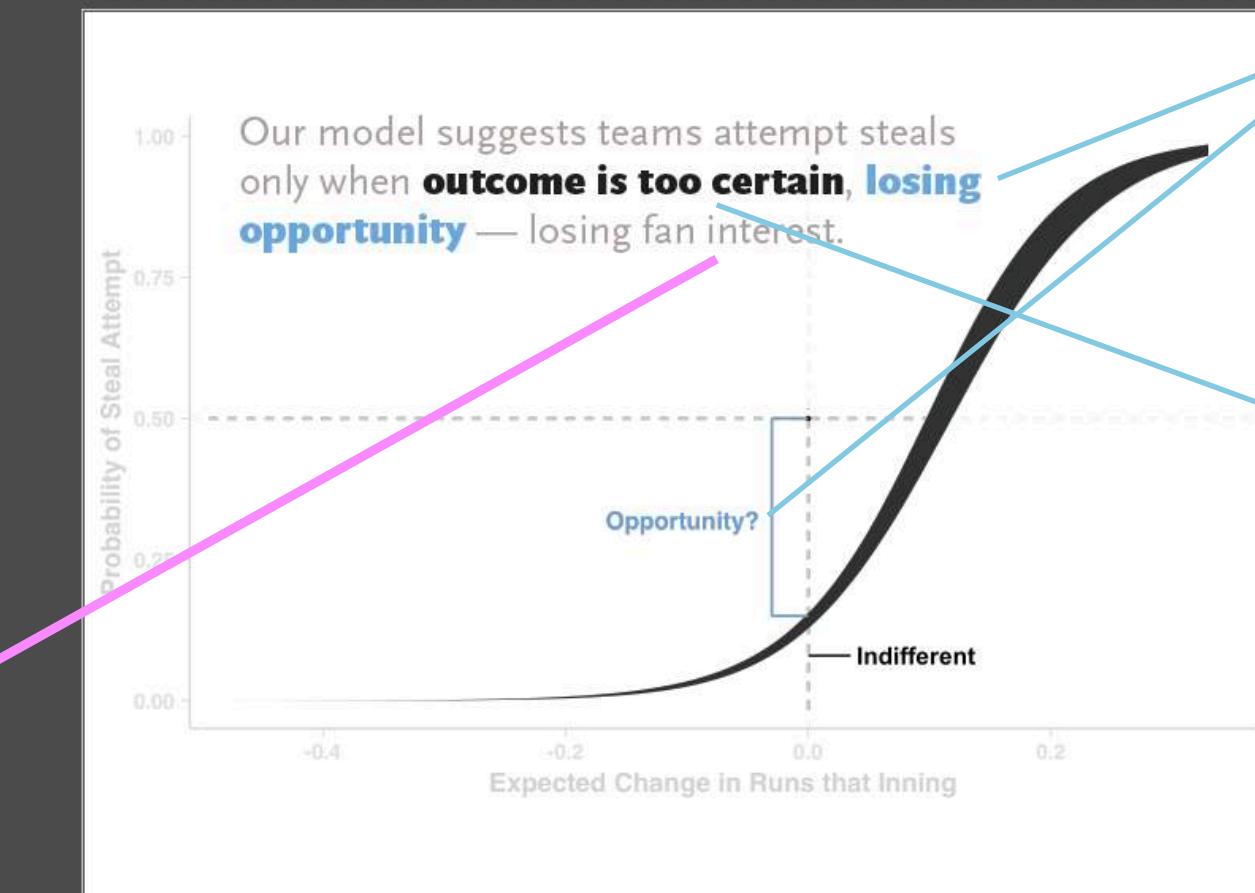
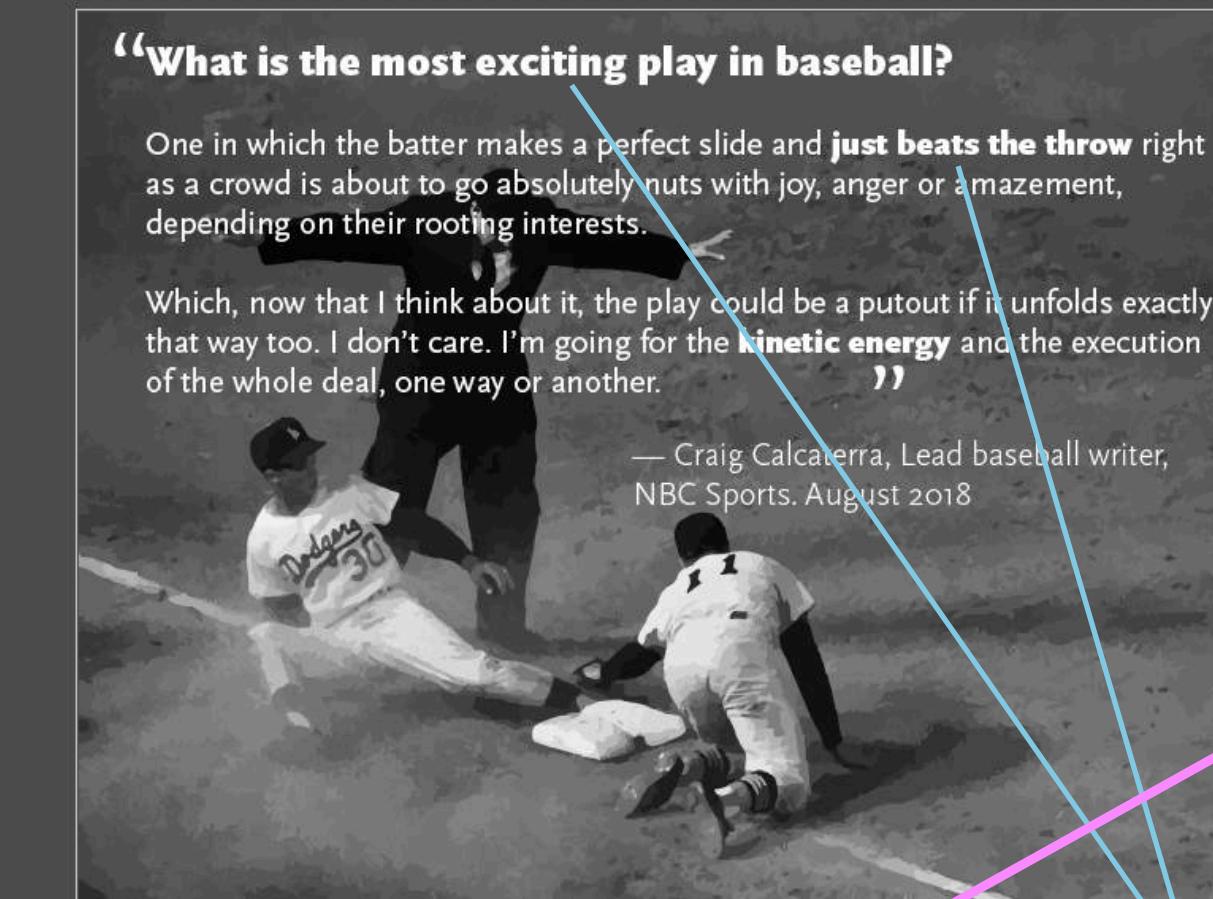
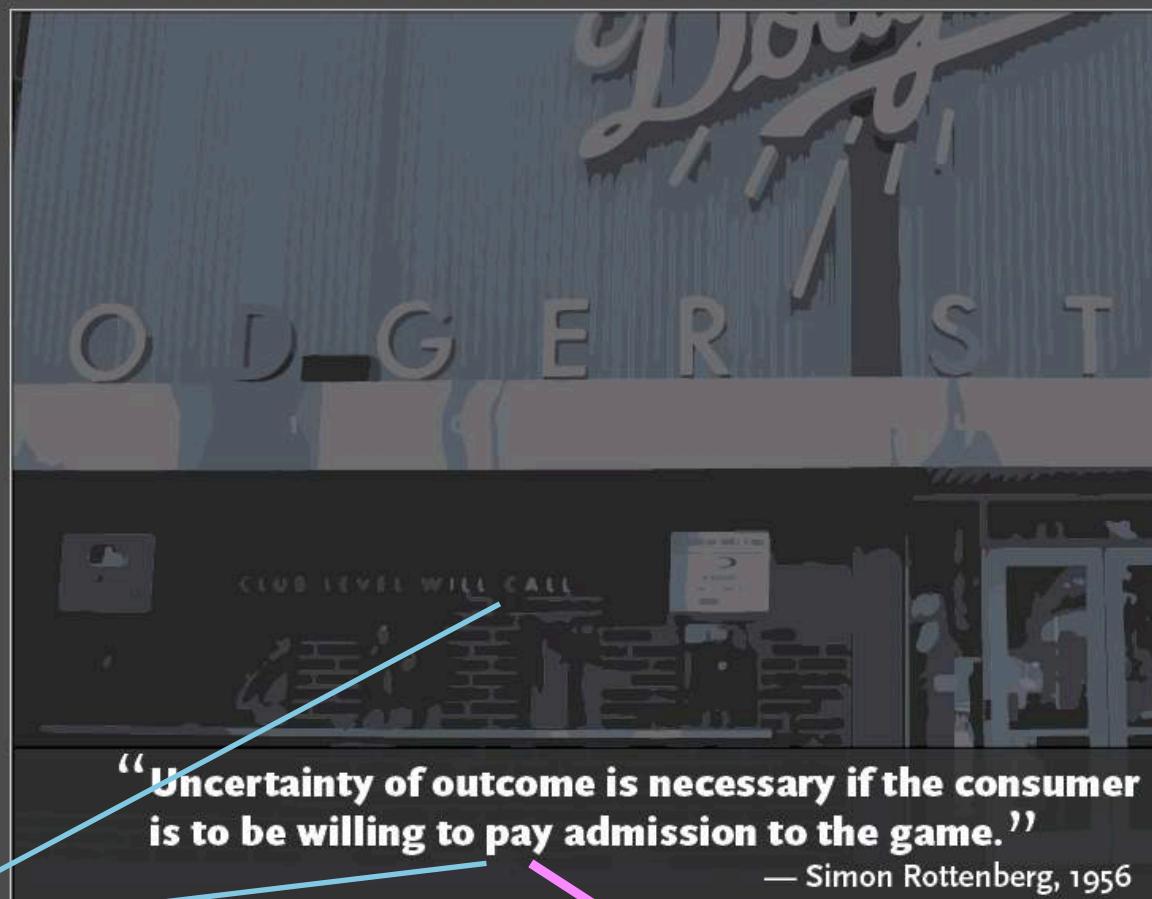
, gametime decisions will mean more and anticipating those player matchups.

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

6

One approach to structuring this example. Are there other ways to arrange a narrative?

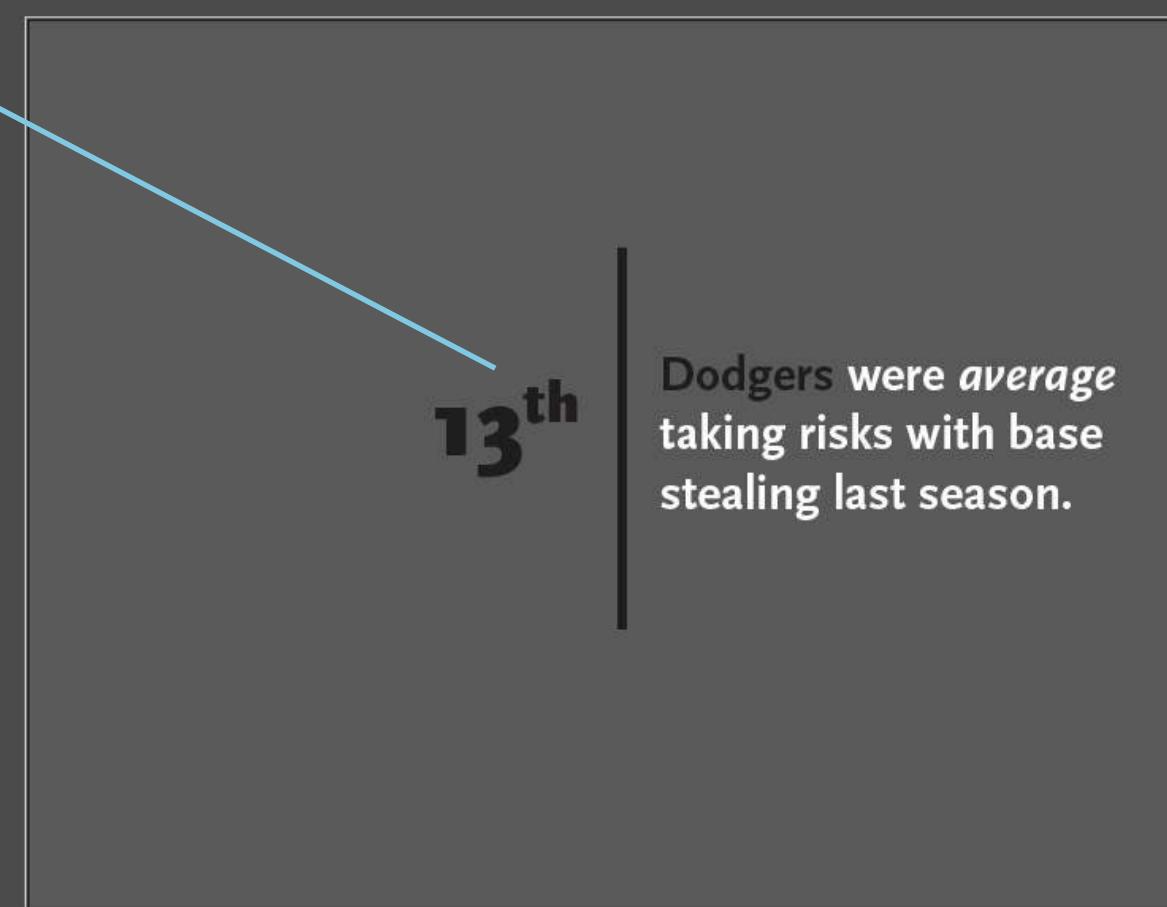
Illustration congruent with words



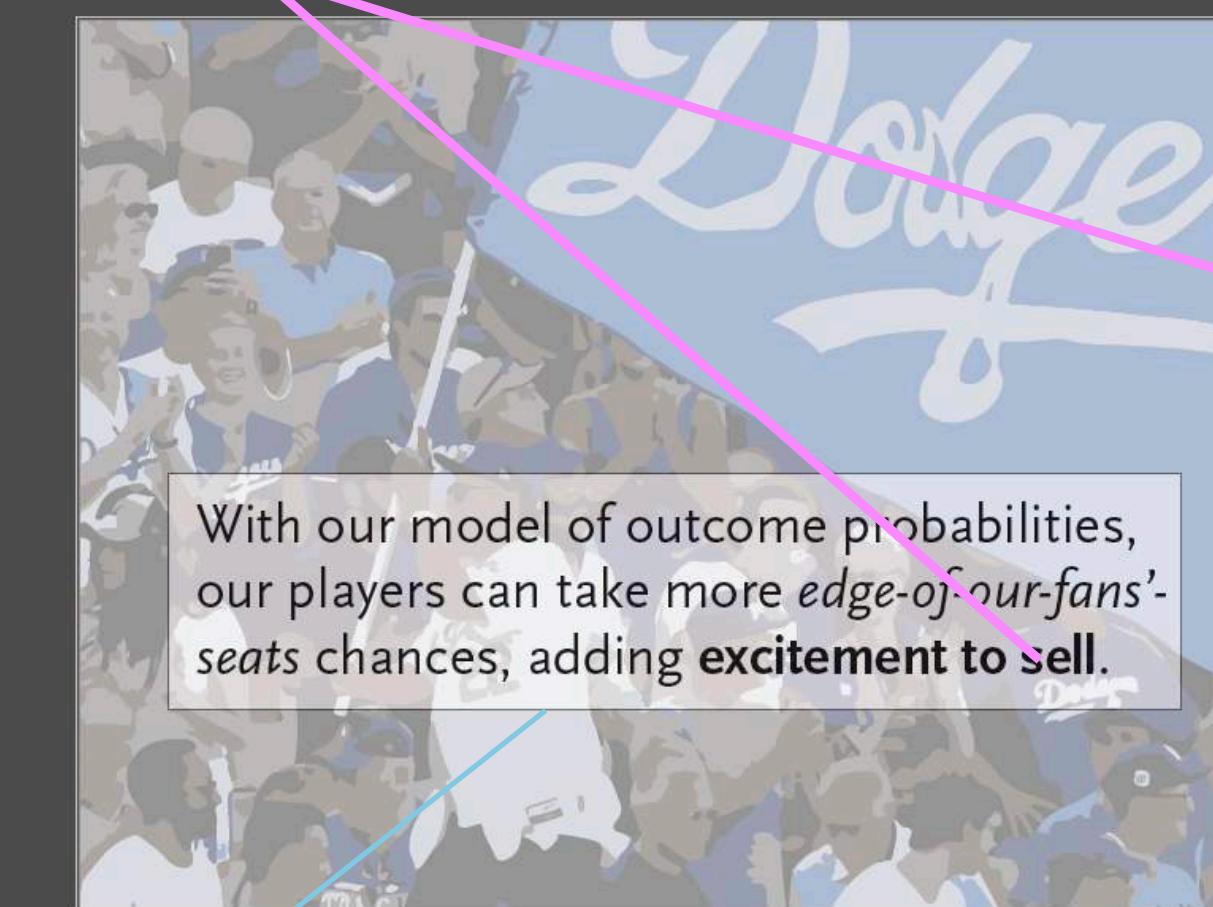
Color relates words to data display

Rising action or initial (problem)

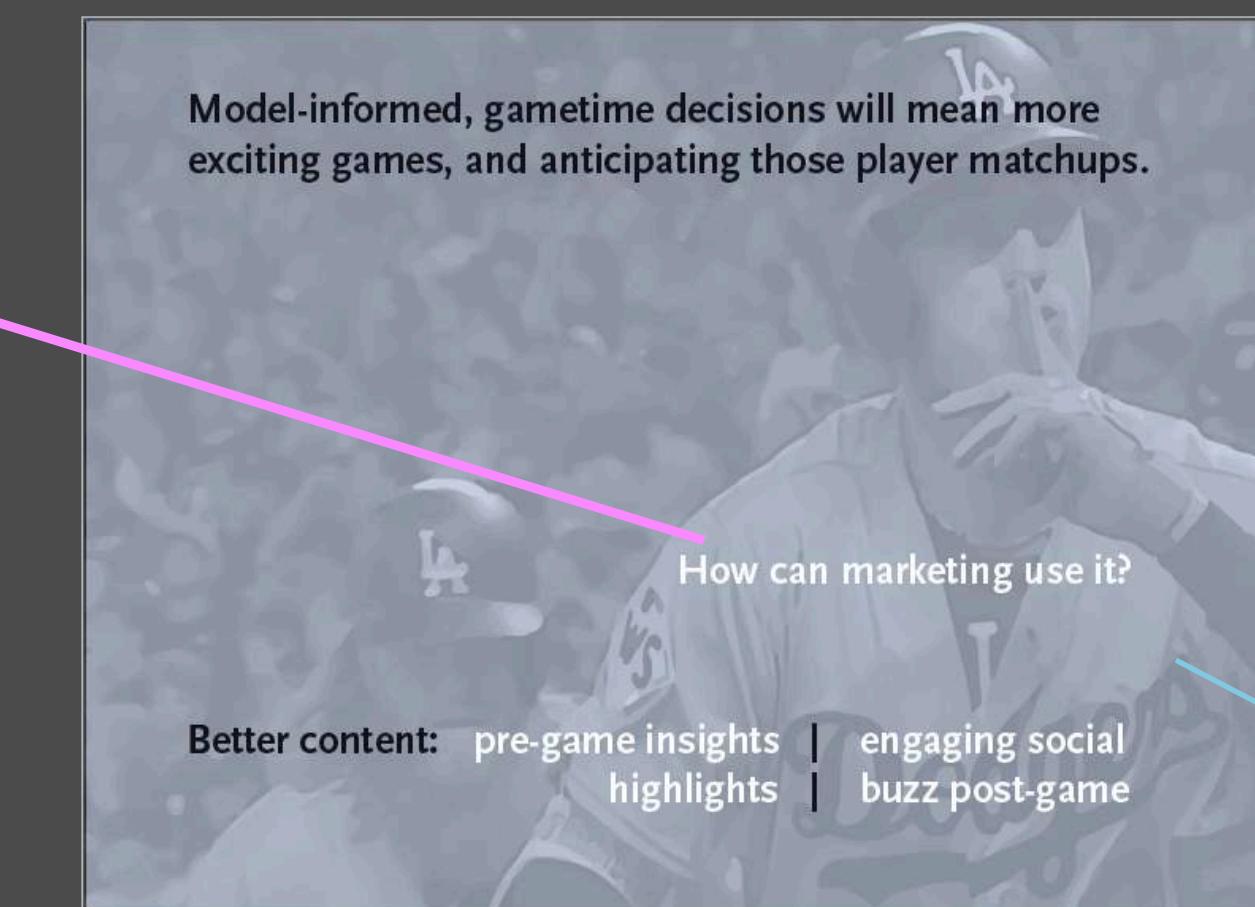
Climax or peak (analytics)



Falling action or release (solution)



5



6

Denouement or call to action

Imagining a *draft storyboard* for *The Next Rembrandt*

Visual narrative

1



Written narrative

We should market ourselves as data-driven, forward-thinking, and creative.

To demonstrate this, we created algorithms to turn paintings into data, analyze it, and create an entirely new painting perhaps indistinguishable from the master, Rembrandt. This one. It's new, not a copy.

Curious about how we researched, developed, and created something entirely new, and can touch, from data?

2



3



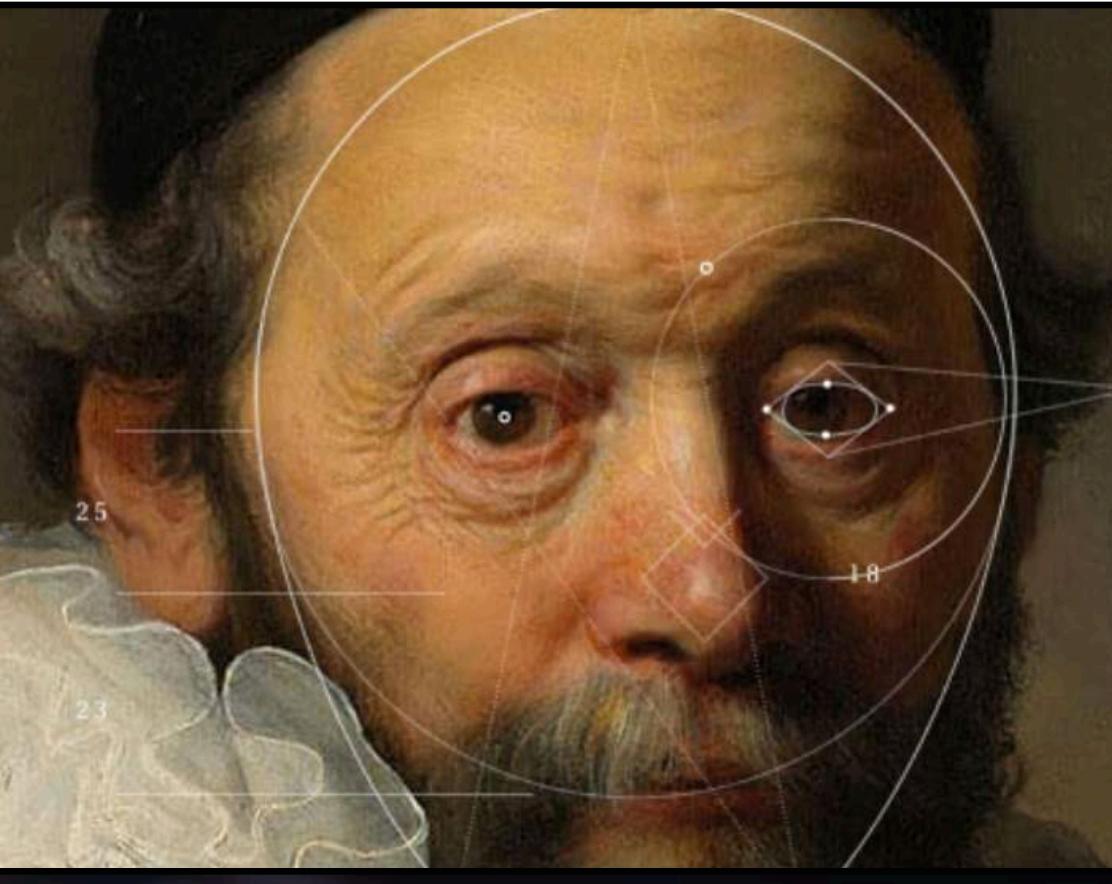
First we gathered a extensive pool of data. Perhaps to some, a large collection of paintings are not data. But we partnered with museums to collect many of Rembrandt's works.

Our analysis of these works — these data — helped us understand how to paint like Rembrandt.

With these paintings, we coded and trained an algorithm to see what we can see through our own eyes, to identify the subject of each painting, to learn the demographics of this master painter's focus.

Visual narrative

4



5



6

That wasn't enough. Then, our algorithm identified the features of his subjects, like the shape of the face and eyes.

To make a painting like Rembrandt, though, we would need it to look and feel like his originals. Paintings have texture and depth. We measured the actual depth of his paint strokes on each of the works we collected, turning messy, real-world information into structured data for our algorithm.

After transforming the data generated by our algorithm, we were able to create the new painting by feeding it back into a 3D printer.

We apply this same dedication using technology to inform our customers. We can do even more with data for our clients. Let's make them curious about how we can use data to create solutions for them.

1

2

3

Visual narrative



1

Written narrative

We should market ourselves as data-driven, forward-thinking, and creative.

To demonstrate this, we created algorithms to turn paintings into data, analyze it, and create an entirely new painting perhaps indistinguishable from the master, Rembrandt. This one. It's new, not a copy.

Curious about how we researched, developed, and created something entirely new, and can touch, from data?



2

First we gathered a extensive pool of data. Perhaps to some, a large collection of paintings are not data. But we partnered with museums to collect many of Rembrandt's works.

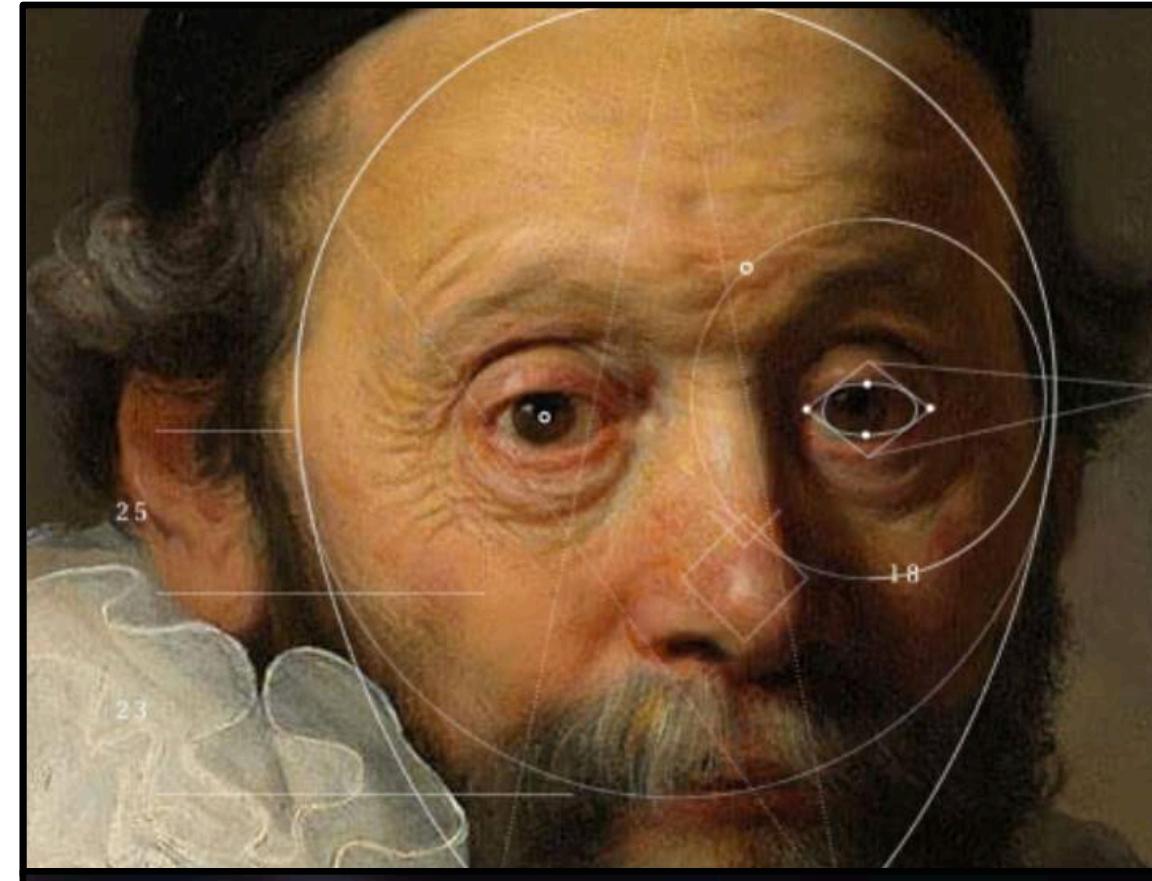
Our analysis of these works — these data — helped us understand how to paint like Rembrandt.



3

With these paintings, we coded and trained an algorithm to see what we can see through our own eyes, to identify the subject of each painting, to learn the demographics of this master painter's focus.

Visual narrative



4



5



6

That wasn't enough. Then, our algorithm identified the features of his subjects, like the shape of the face and eyes.

To make a painting like Rembrandt, though, we would need it to look and feel like his originals. Paintings have texture and depth. We measured the actual depth of his paint strokes on each of the works we collected, turning messy, real-world information into structured data for our algorithm.

After transforming the data generated by our algorithm, we were able to create the new painting by feeding it back into a 3D printer.

We apply this same dedication using technology to inform our customers. We can do even more with data for our clients. Let's make them curious about how we can use data to create solutions for them.

Storyboarding tips

1

Focus on the outcome of your persuasion. What are you trying to persuade your audience to think or do differently?

2

Include storytelling tools in your storyboard. In the beginning, middle or end of your narrative, include one of the tactics described in class to help you connect emotionally with your audience.

3

Combine images, graphs, and headlines to tell your story. Make it easy for the reader to follow your story by using less copy and more instantly “telegraph” images, graphs and headlines.

4

Include your written narrative. The slide images, data, and copy may not be clear enough to tell your story; tell us your written narrative (e.g., what you would say) beside the visual.

5

Use real data and, as needed, strong supporting images. Do the analysis required to tell your story, show the key data. If your story depends on the images, then you should absolutely find them to make a bigger impact.

“It’s hard to have a story if nothing moves or nothing changes.”

- Jonathan Corum: the science graphics editor at The New York Times and founder of 13pt LLC, an information design studio.

Group discussion, creating a storyboard for Citi Bike

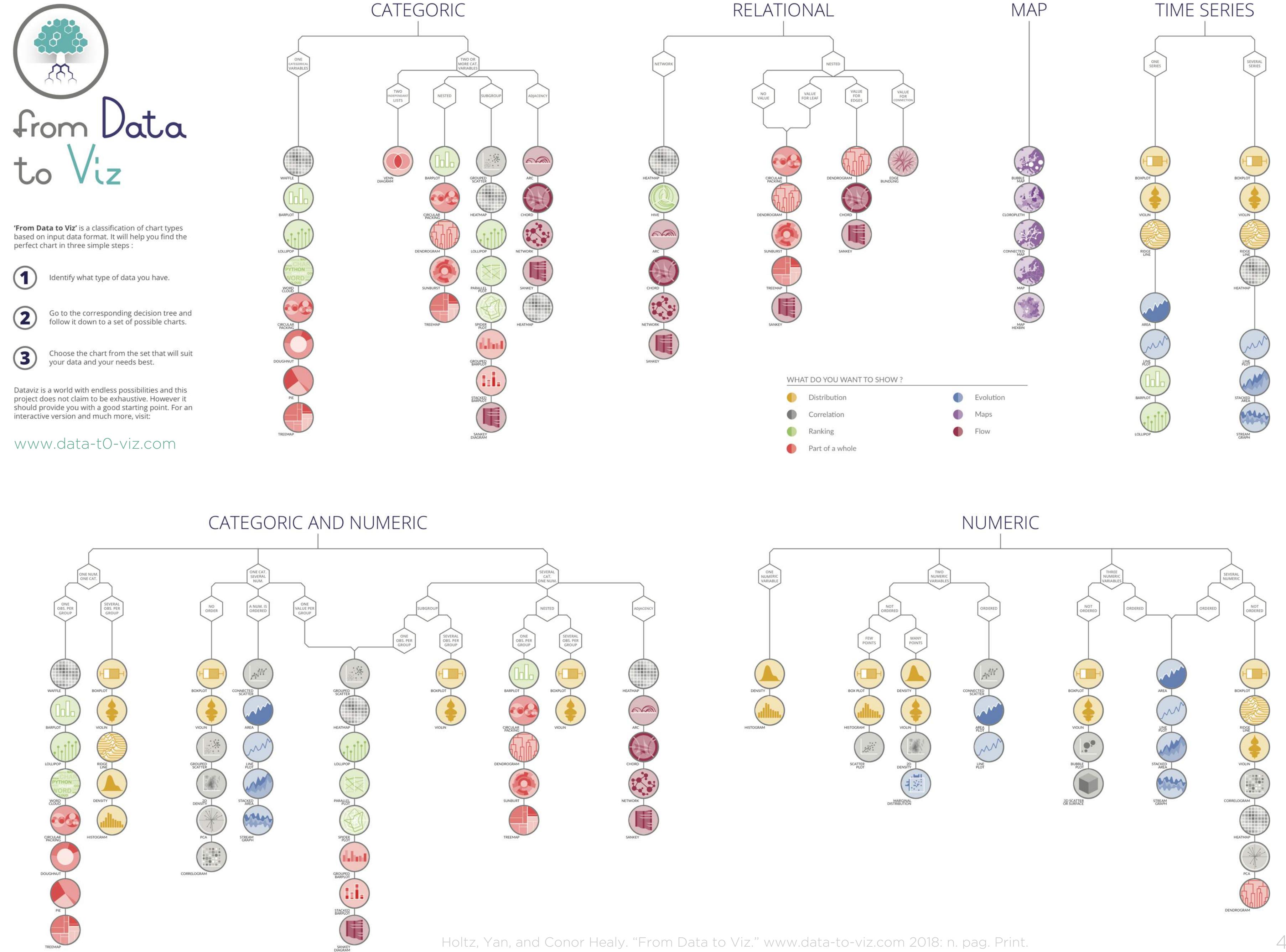


A resource for making data visuals for your storyboard



from Data to Viz, Winner, Information is Beautiful Award

Holtz & Healy



Conveying comparison and change, examples in data visuals

Example of comparison in Malofiej award-winning data story.

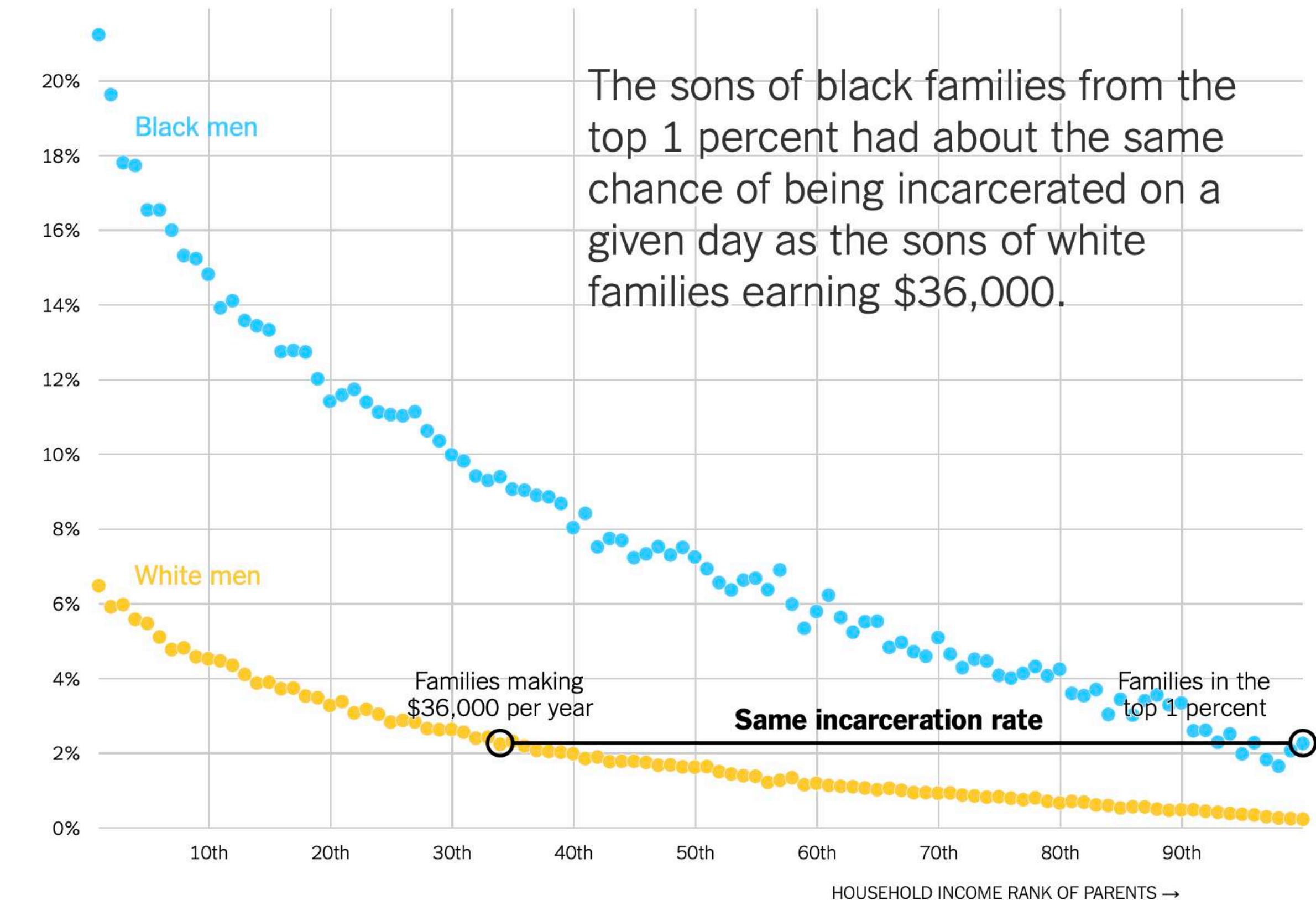


Extensive Data Shows Punishing Reach of Racism for Black Boys

Badger, co-authors

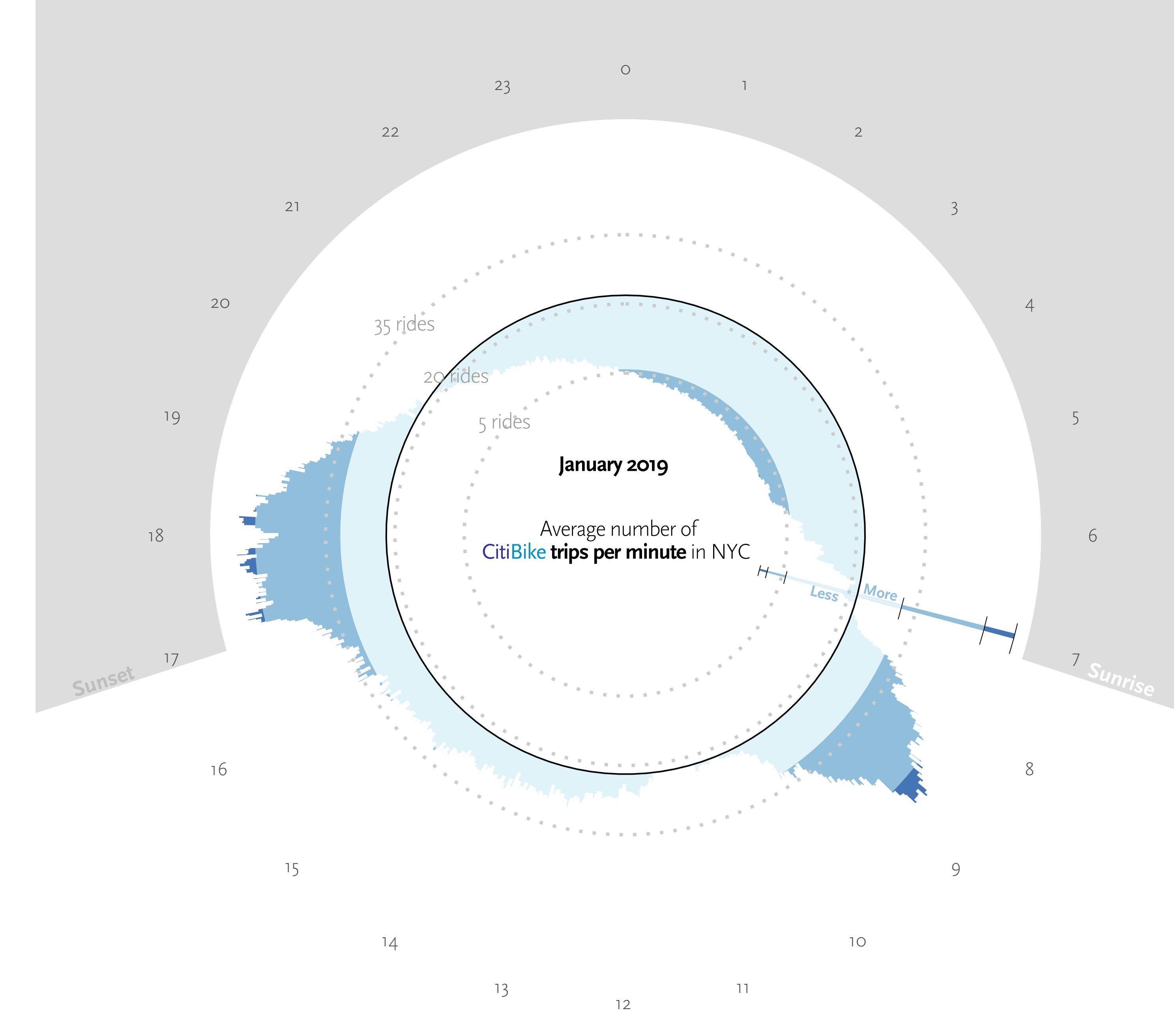
Emily writes about cities and urban policy for The Upshot from the Washington bureau. She's particularly interested in housing, transportation and inequality — and how they're all connected. She joined the Times in 2016 from The Washington Post.

Share of the men incarcerated on April 1, 2010



Includes men who were ages 27 to 32 in 2010.

Revisiting CitiBike, showing temporal changes in activity



Between now and next class

For Next Week, Module 7:

Agenda next week

For online discussion

Behind the visual curtain

Find a non-moving data visualization that is also a story—one you really like—but that you don’t know how you could make it. Share the visual with us and explain what part you don’t know how you could make, but would like to.

Learning with practice

Aside from your assigned projects, briefly describe one of your attempts in the past week to use any of the techniques we have discussed in the course.

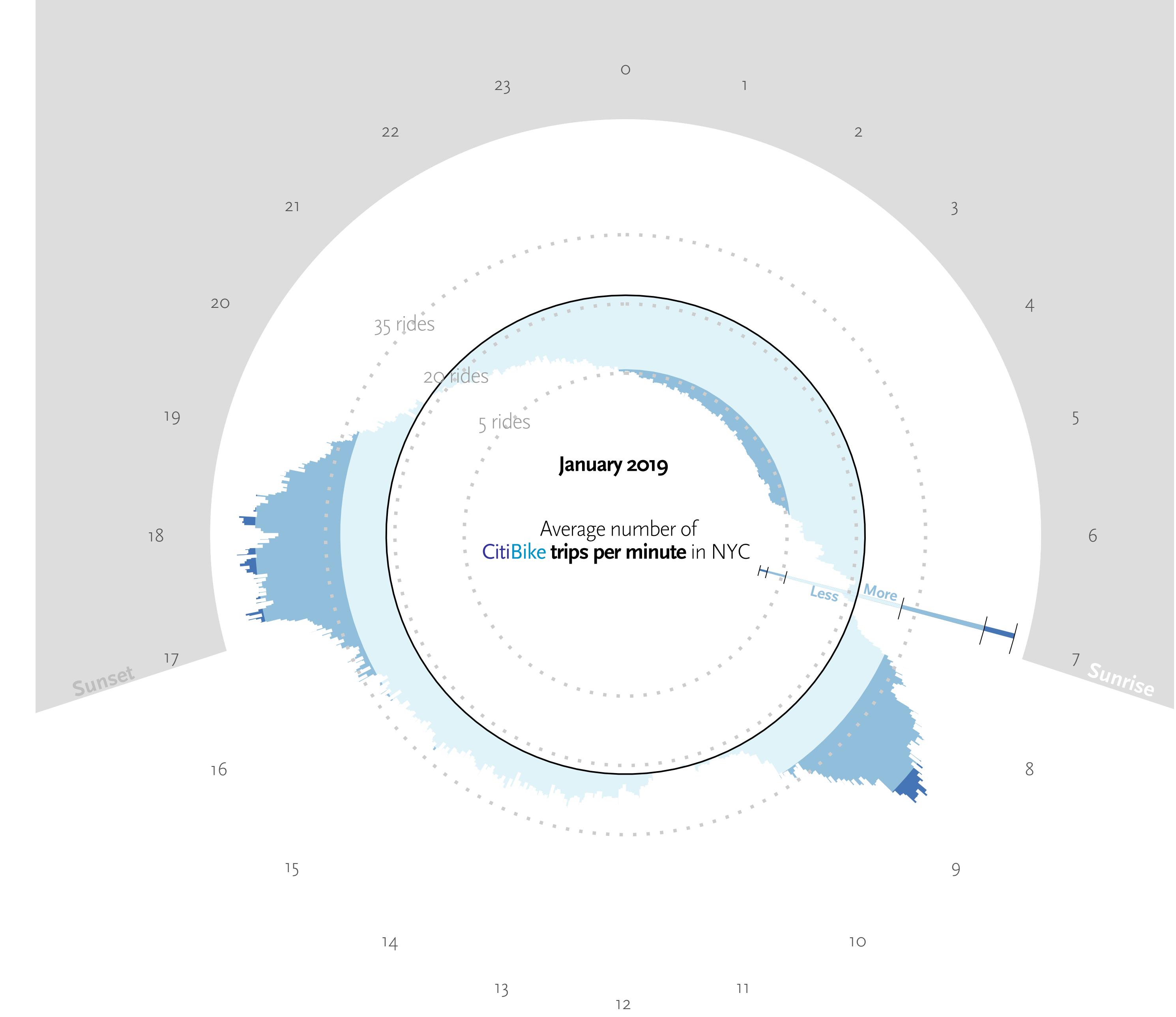
See you
next week!



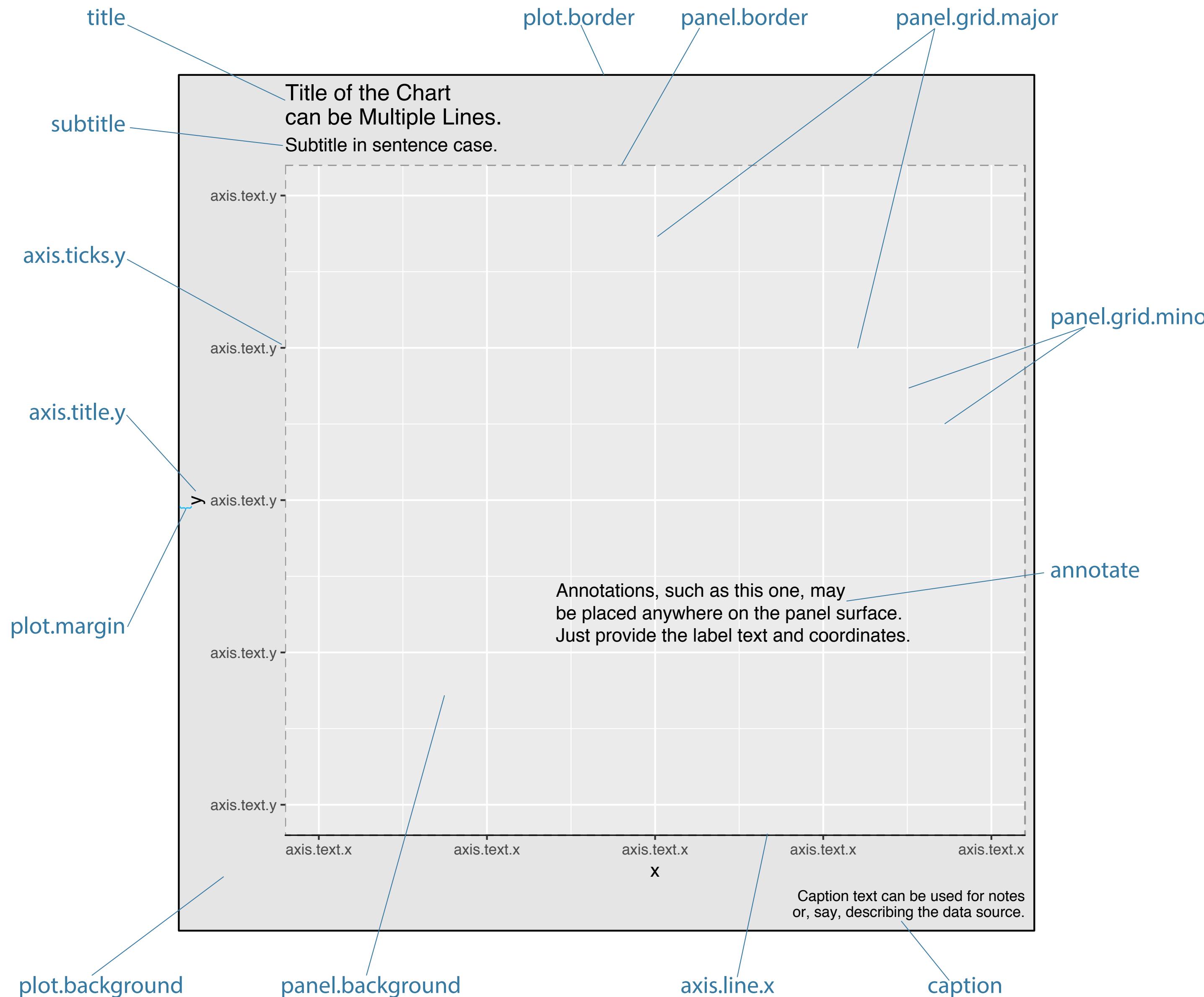
Creating graphics, a short, partial intro

Todo: Maybe delete this
section for now, is bonus
material if have time.
Otherwise it goes in lecture
with graphics.

**Think about
graphics as layers:
here, a white
layer partly masks
the band of blues**



First, think visually in describing the graphic. Then create the layers.



```
# load grammar of graphics  
library(ggplot2)
```

```
p <-
```

```
# functions for data ink
```

```
ggplot(data = <data>,  
       mapping = aes(<aesthetic> = <variable>,  
                     <aesthetic> = <variable>,  
                     <...> = <...>) +  
       geom_<type>(<...>) +  
       scale_<mapping>_<type>(<...>) +  
       coord_<type>(<...>) +  
       facet_<type>(<...>) +  
       <...> +
```

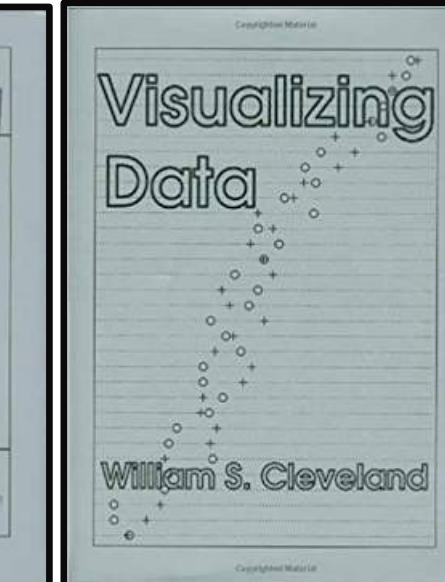
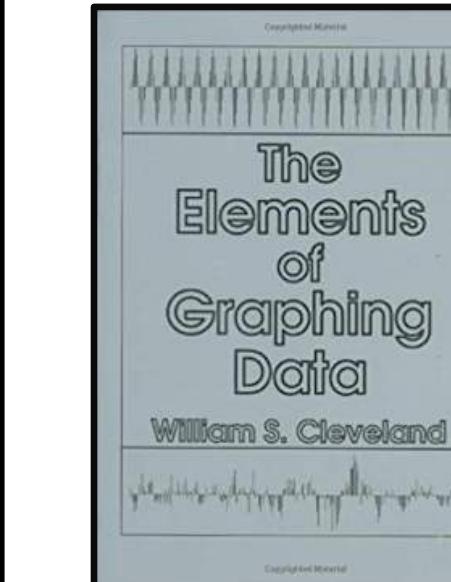
```
# functions for non-data ink
```

```
labs(<...>) +  
theme(<...> = <...>) +  
annotate(<...>) +  
<...>
```

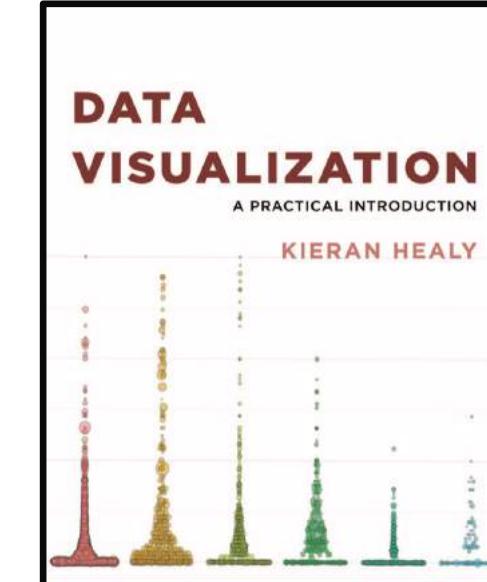
```
element_blank()  
element_line(<...> = <...>)  
element_rect(<...> = <...>)  
element_text(<...> = <...>)
```

Learning & References

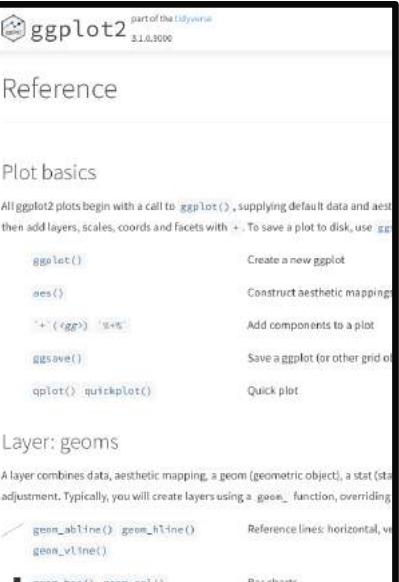
---- visual relationships in data ----



Implementation in R



ggplot reference



socviz.co

ggplot2.tidyverse.org