Storytelling With Data

Persuasive business writing: purpose, audience, and craft

Scott Spencer | Columbia University

# **Conceptual project timeline**



Scott Spencer / 🖓 https://github.com/ssp3nc3r 😰 scott.spencer@columbia.edu



## the point of (business) communication

#### Get our audience to



Doumont, Jean-luc

pay attention to, understand, (be able to) act upon a maximum of **messages**, given constraints.

#### Information versus messages

**INFORMATION** | A concentration of 175 µg per m<sup>3</sup> has been observed in urban areas.

MESSAGE | A concentration in urban areas (175 µg/m<sup>3</sup>) is unacceptably high.



Doumont, Jean-luc

Scott Spencer / 🗘 https://github.com/ssp3nc3r 😰 scott.spencer@columbia.edu

## Three laws of communication

Adapt to your audience

Maximize the signal-to-noise ratio

Use effective redundancy



Doumont, Jean-luc

Scott Spencer / 🖓 https://github.com/ssp3nc3r 😰 scott.spencer@columbia.edu

document structure

Scott Spencer / 🖓 https://github.com/ssp3nc3r 🛛 😰 scott.spencer@columbia.edu

#### document structure





Doumont, Jean-luc

Make the audience receptive to the topic of the communication

Once you have their attention, tell them your main message

Next, support this message: tell them how you got there

Last of all, present separately what fewer will want to know

Scott Spencer / 🗘 https://github.com/ssp3nc3r 😰 scott.spencer@columbia.edu

# document structure, story or narrative



Scott Spencer / 🖓 https://github.com/ssp3nc3r 🖄 scott.spencer@columbia.edu

## document structure, story — from Shakespeare to data science?!

data analytics

## business problem





Freytag, Gustav

Scott Spencer / 🗘 https://github.com/ssp3nc3r 😰 scott.spencer@columbia.edu

#### document structure, beginning a story

unexpected change

opening of an information gap



Storr, Will

Scott Spencer / 💭 https://github.com/ssp3nc3r 🛛 😰 scott.spencer@columbia.edu



#### document structure



Zinsser, William

# the *lead* and the *ending*

Scott Spencer / 🗘 https://github.com/ssp3nc3r 🛛 😰 scott.spencer@columbia.edu



sentence structure





#### sentence structure

# old



new

Scott Spencer / 🖓 https://github.com/ssp3nc3r 😰 scott.spencer@columbia.edu



Williams, Bizup, Fitzgerald

#### sentence structure, old before new

# **Section 17.3**, example 10(a)

Because the naming power of words was distrusted by Locke, he repeated himself often. Seventeenth-century theories of language, especially Wilkins's scheme for a universal language involving the creation of countless symbols for countless meanings, had centered on this naming power. A new era in the study of language that focused on the ambiguous relationship between sense and reference begins with Locke's distrust.



#### example 10(b)

Locke often repeated himself because he distrusted the naming power of words. This naming power had been central to seventeenth-century theories of language, especially Wilkins's scheme for a universal language involving the creation of countless symbols for countless meanings. Locke's distrust begins a new era in the study of language, one that focused on the ambiguous relationship between sense and reference.



Williams, Bizup, Fitzgerald

#### structure

#### Want to confuse? new before old

"One of the most compelling and challenging films of this year, you'll be gripped, enthralled and exhausted. A MODERN CLASSIC. UNFORGETTABLE!"  $\star \star \star \star \star \star$ - UNCUT MAGAZINE

"Dazzling, a masterpiece on a par with The Usual Suspects" - FHM

> **GUY PEARCE CARRIE-ANNE MOSS** JOE PANTOLIANO

www.pathe.co.uk/memento



N 📲 CHKISTIPHER NILAN

ÉMENTO

MEMEAITO

SOME MEMORIES ARE BEST FORGOTTEN.

SUMMIT PATHE!



communicating data analytics, discuss Citi Bike memo

Scott Spencer / 🖓 https://github.com/ssp3nc3r 🖄 scott.spencer@columbia.edu



# audience for the memo, chief analytics officer

# **CHIEF ANALYTICS OFFICER** | heads up a company's data analytics operations, transforming data into business value, and drives data-related business change.

—Zetlin, Minda. "What Is a Chief Analytics Officer? The Exec Who Turns Data into Decisions." CIO, November 2, 2017.

Scott Spencer / 🖓 https://github.com/ssp3nc3r 😰 scott.spencer@columbia.edu



Minda Zetlin

#### communicating data analytics, memo for discussion

# To inform rebalancing, let's explore docking and bike availability in the context of subway and weather information.

We should explore station and ride data in the context of subway and weather information to gain insight for "rebalancing," what our Dani Simmons explains is "one of the biggest challenges of any bike share system, especially in ... New York where residents don't all work a traditional 9-5 schedule, and though there is a Central Business District, it's a huge one and people work in a variety of other neighborhoods as well."

A rebalancing study (Saldarriaga, 2013) by Columbia University Center for Spatial Research previously identified trends in bike usage using heatmaps. As those visualizations did not combine dimensions of space and time, which will be helpful to uncover trends in bike and station availability by neighborhood throughout a day, we can begin our analysis there.

NYC OpenData and The Open Bus Project provide published date, time, station ID, and ride instances for all our docking stations and bikes since we began service. To begin our project, we can visually explore the intersection of trends in both time and location with this data to understand problematic neighborhoods and, even, individual stations, using up-to-date information.

Then, we will build upon the initial work, exploring causal factors such as the availability of alternative transportation (e.g., subway stations near docking stations) and weather. Both of which, we have available data that can be joined using timestamps.

The project aligns with our goals to, in Simmons's words, "be innovative in how we meet this challenge." Let's draft a detailed proposal.

Sincerely, Scott Spencer

Scott Spencer / 💭 https://github.com/ssp3nc3r 🛛 😰 scott.spencer@columbia.edu



Saldarriaga, Juan Francisco. CitiBike Rebalancing Study. Spatial Information Design Lab, Columbia University, 2013. http://spatialinformationdesignlab.org/projects/citibike-rebalancing-study.

#### communicating data analytics, messages first?



# To inform rebalancing, let's explore docking and bike availability in the context of subway and weather information.

We should explore station and ride data in the context of subway and weather information to gain insight for "rebalancing," what our Dani Simmons explains is "one of the biggest challenges of any bike share system, especially in ... New York where residents don't all work a traditional 9-5 schedule, and though there is a Central Business District, it's a huge one and people work in a variety of other neighborhoods as well."

A rebalancing study (Saldarriaga, 2013) by Columbia University Center for Spatial Research previously identified trends in bike usage using heatmaps. As those visualizations did not combine dimensions of space and time, which will be helpful to uncover trends in bike and station availability by neighborhood throughout a day, we can begin our analysis there.

NYC OpenData and The Open Bus Project provide published date, time, station ID, and ride instances for all our docking stations and bikes since we began service. To begin our project, we can visually explore the intersection of trends in both time and location with this data to understand problematic neighborhoods and, even, individual stations, using up-to-date information.

Then, we will build upon the initial work, exploring causal factors such as the availability of alternative transportation (e.g., subway stations near docking stations) and weather. Both of which, we have available data that can be joined using timestamps.

The project aligns with our goals to, in Simmons's words, "be innovative in how we meet this challenge." Let's draft a detailed proposal.



Saldarriaga, Juan Francisco. *CitiBike Rebalancing Study*. Spatial Information Design Lab, Columbia University, 2013. http://spatialinformationdesignlab.org/projects/cit-ibike-rebalancing-study.

#### communicating data analytics, narrative structure?

**Unexpected change?** 

**Open an information gap?** 



#### The *lead* and the *ending*?

# To inform rebalancing, let's explore docking and bike availability in the context of subway and weather information.

We should explore station and ride data in the context of subway and weather information to gain insight for "rebalancing," what our Dani Simmons explains is "one of the biggest challenges of any bike share system, especially in ... New York where residents don't all work a traditional 9-5 schedule, and though there is a Central Business District, it's a huge one and people work in a variety of other neighborhoods as well."

A rebalancing study (Saldarriaga, 2013) by Columbia University Center for Spatial Research previously identified trends in bike usage using heatmaps. As those visualizations did not combine dimensions of space and time, which will be helpful to uncover trends in bike and station availability by neighborhood throughout a day, we can begin our analysis there.

NYC OpenData and The Open Bus Project provide published date, time, station ID, and ride instances for all our docking stations and bikes since we began service. To begin our project, we can visually explore the intersection of trends in both time and location with this data to understand problematic neighborhoods and, even, individual stations, using up-to-date information.

Then, we will build upon the initial work, exploring causal factors such as the availability of alternative transportation (e.g., subway stations near docking stations) and weather. Both of which, we have available data that can be joined using timestamps.

The project aligns with our goals to, in Simmons's words, "be innovative in how we meet this challenge." Let's draft a detailed proposal.





Saldarriaga, Juan Francisco. *CitiBike Rebalancing Study*. Spatial Information Design Lab, Columbia University, 2013. http://spatialinformationdesignlab.org/projects/cit-ibike-rebalancing-study.

Readability Statistics	
Counts	
Words	246
Characters	1,312
Paragraphs	5
Sentences	9
Averages	
Sentences per Paragraph	1.8
Words per Sentence	27.3
Characters per Word	5.1
Readability	
Flesch Reading Ease	33.6
Flesch-Kincaid Grade Level	15.3
Passive Sentences	11.1%

#### audience

message first, context solution, high-level goal, action, problem

background, context

data sources and types solution, first stage

solution, second stage

link project back to goals

author

reference information

CitiBike То Director of Analytics

# To inform rebalancing, let's explore docking and bike availability in the context of subway and weather information.

- We should explore station and ride data in the context of subway and weather information to gain insight for "rebalancing," what our Dani Simmons explains is "one of the biggest challenges of any bike share system, especially in ... New York where residents don't all work a traditional 9-5 schedule, and though there is a Central Business District, it's a huge one and people work in a variety of other neighborhoods as well."
- A rebalancing study (Saldarriaga, 2013) by Columbia University Center for Spatial Research previously identified trends in bike usage using heatmaps. As those visualizations did not combine dimensions of space and time, which will be helpful to uncover trends in bike and station availability by neighborhood throughout a day, we can begin our analysis there.
- NYC OpenData and The Open Bus Project provide published date, time, station ID, and ride instances for all our docking stations and bikes since we began service. To begin our project, we can visually explore the intersection of trends in both time and location with this data to understand problematic neighborhoods and, even, individual stations, using up-to-date information.
- Then, we will build upon the initial work, exploring causal factors such as the availability of alternative transportation (e.g., subway stations near docking stations) and weather. Both of which, we have available data that can be joined using timestamps.
- The project aligns with our goals to, in Simmons's words, "be innovative in how we meet this challenge." Let's draft a detailed proposal.

Sincerely, Scott Spencer

Scott Spencer / 💭 https://github.com/ssp3nc3r 🛛 😰 scott.spencer@columbia.edu



Saldarriaga, Juan Francisco. CitiBike Rebalancing Study. Spatial Information Design Lab, Columbia University, 2013. http://spatialinformationdesignlab.org/projects/citibike-rebalancing-study.

communicating data analytics, old before new?



# To inform rebalancing, let's explore docking and bike availability in the context of subway and weather information.

We should explore station and ride data in the context of subway and weather information to gain insight for "rebalancing," what our Dani Simmons explains is "one of the biggest challenges of any bike share system, especially in ... New York where residents don't all work a traditional 9-5 schedule, and though there is a Central Business District, it's a huge one and people work in a variety of other neighborhoods as well."

A rebalancing study (Saldarriaga, 2013) by Columbia University Center for Spatial Research previously identified trends in bike usage using heatmaps. As those visualizations did not combine dimensions of space and time, which will be helpful to uncover trends in bike and station availability by neighborhood throughout a day, we can begin our analysis there.

NYC OpenData and The Open Bus Project provide published date, time, station ID, and ride instances for all our docking stations and bikes since we began service. To begin our project, we can visually explore the intersection of trends in both time and location with this data to understand problematic neighborhoods and, even, individual stations, using up-to-date information.

Then, we will build upon the initial work, exploring causal factors such as the availability of alternative transportation (e.g., subway stations near docking stations) and weather. Both of which, we have available data that can be joined using timestamps.

The project aligns with our goals to, in Simmons's words, "be innovative in how we meet this challenge." Let's draft a detailed proposal.



Saldarriaga, Juan Francisco. *CitiBike Rebalancing Study*. Spatial Information Design Lab, Columbia University, 2013. http://spatialinformationdesignlab.org/projects/cit-ibike-rebalancing-study.

communicating data analytics, discuss Dodgers memo

Scott Spencer / 🗘 https://github.com/ssp3nc3r 🛛 😰 scott.spencer@columbia.edu



perfect game

- Los Angeles Dodgers
- pitcher Sandy Koufax

# baseball

Statcast data

(attempting to) steal a base

salary cap



#### models

inferences

mode, maximum likelihood

mean, expectations

# statistics, probability, computing

probability distributions

decision theory

joint distributions

counterfactuals



**R** language and packages

Scott Powers Director of quantitative analytics PhD Statistics, Stanford





#### communicating data analytics, memo for discussion

# 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. Perfect games aside, we can test the concept with decisions to steal.



#### communicating data analytics, messages first?



# 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. Perfect games aside, we can test the concept with decisions to steal.



#### communicating data analytics, narrative structure?

**Unexpected change?** 

**Open an information gap?** 



#### The *lead* and the *ending*?

# 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. Perfect games aside, we can test the concept with decisions to steal.



Readability Statistics	
Counts	
Words	250
Characters	1,357
Paragraphs	4
Sentences	14
Averages	
Sentences per Paragraph	3.5
Words per Sentence	17.8
Characters per Word	5.2
Readability	
Flesch Reading Ease	43.5
Flesch-Kincaid Grade Level	11.6
Passive Sentences	0%

#### audience

Our game decisions should optimize expectations. Let's test the concept by modeling decisions to steal. message first, context orient the audience 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, goal, action, problem we leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint solution proposed 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 objectives, data, method Statcast data, weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all benefit (eventual) 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.

benefit, comparison, financial

limitations, solution (short-term)

reminder of proposal

author

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. Perfect games aside, we can test the concept with decisions to steal.



communicating data analytics, old before new?



# 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. Perfect games aside, we can test the concept with decisions to steal.



# To inform rebalancing, let's explore docking and bike availability in the context of subway and weather information.

We should explore station and ride data in the context of subway and weather information to gain insight for "rebalancing," what our Dani Simmons explains is "one of the biggest challenges of any bike share system, especially in ... New York where residents don't all work a traditional 9-5 schedule, and though there is a Central Business District, it's a huge one and people work in a variety of other neighborhoods as well."

A rebalancing study (Saldarriaga, 2013) by Columbia University Center for Spatial Research previously identified trends in bike usage using heatmaps. As those visualizations did not combine dimensions of space and time, which will be helpful to uncover trends in bike and station availability by neighborhood throughout a day, we can begin our analysis there.

NYC OpenData and The Open Bus Project provide published date, time, station ID, and ride instances for all our docking stations and bikes since we began service. To begin our project, we can visually explore the intersection of trends in both time and location with this data to understand problematic neighborhoods and, even, individual stations, using up-to-date information.

Then, we will build upon the initial work, exploring causal factors such as the availability of alternative transportation (e.g., subway stations near docking stations) and weather. Both of which, we have available data that can be joined using timestamps.

The project aligns with our goals to, in Simmons's words, "be innovative in how we meet this challenge." Let's draft a detailed proposal.

Sincerely, Scott Spencer

# 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. Perfect games aside, we can test the concept with decisions to steal.





Saldarriaga, Juan Francisco. *CitiBike Rebalancing Study*. Spatial Information Design Lab, Columbia University, 2013. http://spatialinformationdesignlab.org/projects/cit-ibike-rebalancing-study.

#### Exercise: draft first paragraph for your project

In the last couple of weeks, we discussed scoping an analytics project, and we have researched potential data analytics projects we find interesting, along with data we may analyze to create insights. In the next several minutes, draft a first paragraph of the memo for your data analytics project. Consider the *concepts we've discussed as tools* for structuring your paragraph and sentences.



# Exercise: Revise a colleague's memo

Exchange your drafts, and offer a few revision suggestions by applying the concepts we have covered.

#### Get specific!

Scott Spencer / 💭 https://github.com/ssp3nc3r 😰 scott.spencer@columbia.edu

# visual organization

Scott Spencer / 🖓 https://github.com/ssp3nc3r 🛛 🗟 scott.spencer@columbia.edu





Butterick, Matthew

Scott Spencer / 🖓 https://github.com/ssp3nc3r 🛛 🙊 scott.spencer@columbia.edu



#### visual organization, very basic guidelines for optimizing text readability

Body text Point size

print 10-12 pt web 15-25 px



Butterick, Matthew

To <b>CitiBike</b> 2019 February : Director of Analytics <b>To inform rebalancing, let's explore docking and bike availability</b> <b>in the context of subway and weather information.</b> We should explore station and ride data in the context of subway an weather information to gain insight for "rebalancing," what our Dani Sim mons explains is "one of the biggest challenges of any bike share systen especially in New York where residents don't all work a traditional 9- schedule, and though there is a Central Business District, it's a huge one an people work in a variety of other neighborhoods as well." A rebalancing study (Saldarriaga, 2013) by Columbia University Center for Spatial Research previously identified trends in bike usage using heatmap As those visualizations did not combine dimensions of space and tim- which will be helpful to uncover trends in bike and station availability b neighborhood throughout a day, we can begin our analysis there. NYC OpenData and The Open Bus Project provide published date, tim- station ID, and ride instances for all our docking stations and bikes since w	2 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4	⊐ Line height 120 - 145 % of point size
can be joined using timestamps. The project aligns with our goals to, in Simmons's words, "be innovative i how we meet this challenge." Let's draft a detailed proposal. Sincerely, Scott Spencer  Saldarriaga, Juan Francisco. <i>CitiBike Rebalancing Study</i> . Spatial Information Design Lab, Columbia University, 2013. http://spatialinformationdesignlab.org/projects/cit- ibike-rebalancing-study.	n k	
← Line length 45 - 90 characters →		

#### References

**Booth**, Wayne C, Gregory G Columb, Joseph M Williams, Joseph Bizup, and William T Fitzgerald. The Craft of Research. Fourth. University of Chicago Press, 2016.

Butterick, Matthew. "Butterick's Practical Typography," 2018. https:// practicaltypography.com/.

**Freytag**, Gustav. Freytag's Technique of the Drama: An Exposition of Dramatic Composition and Art. Translated by Elias MacEwan. Third. Chicago: Scott, Foresman and Co., 1900.

Shakespeare, William, Wilbur L. Cross, and Tucker Brooke. The Yale Shakespeare: The *Complete Works*. New York: Barnes and Noble, 1993.

Spencer, Scott. Memo to Scott Powers. "Our Game Decisions Should Optimize" Expectations; Let's Test the Concept by Modeling Decisions to Steal." Memo, February 2, 2019.

———. Memo to Director of Analytics Citi Bike. "*To Inform Rebalancing, Let's Explore Bike* and Docking Availability in the Context of Subway and Weather Information." Memo, February 2, 2019.

Storr, Will. Science of Storytelling. New York, NY: Abrams Books, 2020.

**Zetlin**, Minda. "What Is a Chief Analytics Officer? The Exec Who Turns Data into Decisions." CIO, November 2, 2017.

Zinsser, William. On Writing Well. Sixth. The Classic Guide to Writing Nonfiction. Harper Resource, 2001.

Scott Spencer / 💭 https://github.com/ssp3nc3r 🛛 😰 scott.spencer@columbia.edu

