

# Storytelling with Data

**Module 9: Frame the story – consumers and information graphics**

**Scott Spencer**  
Faculty and Lecturer  
Columbia University

# Agenda

Upcoming deliverable

Today's objectives

Considering uncertainty

Information graphics

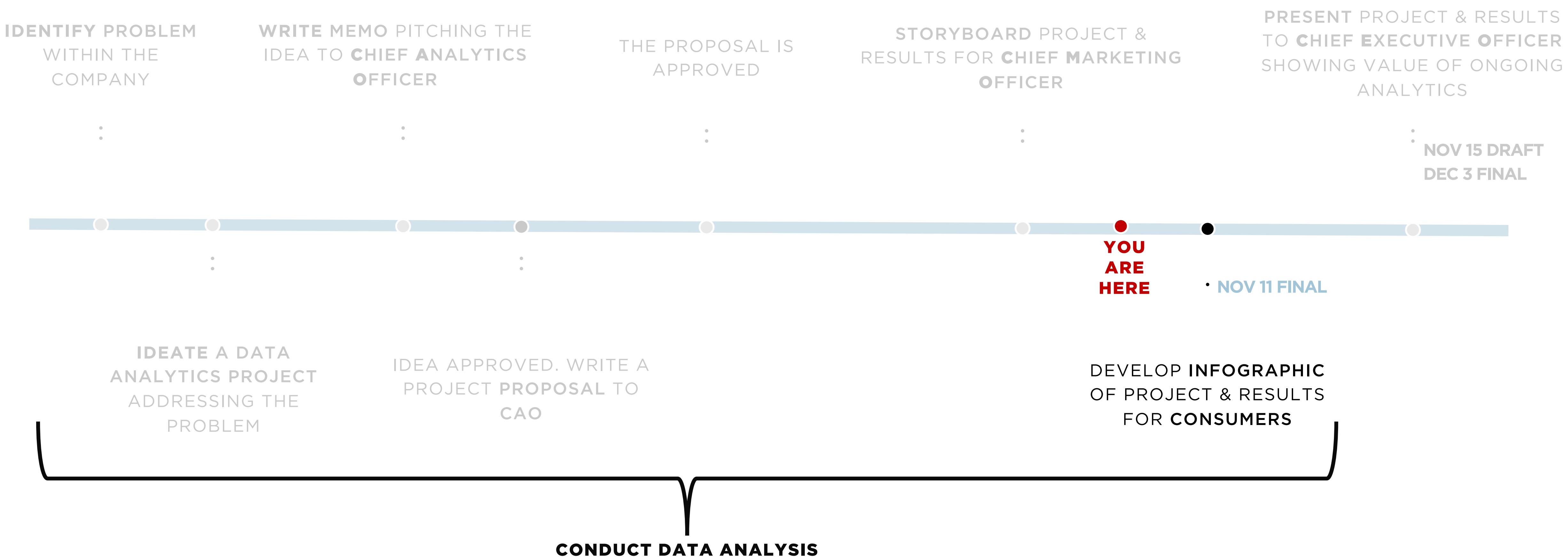
# Questions or suggestions?

---

# Upcoming deliverables

# Upcoming deliverable

**Information graphic** – reframe your story, this time building off the messages you built for the marketing team in order to craft an infographic that displays the results of the analytic work in a way that is accessible, engaging, and exciting for a **general or consumer audience**.



# Today's Objectives

# Objectives

1

Communicating uncertainty  
with data visualization

2

Organizing information

3

Use information graphics  
to tell a story

# Communicating uncertainty

# Where's my data? Evaluating Visualizations with Missing Data

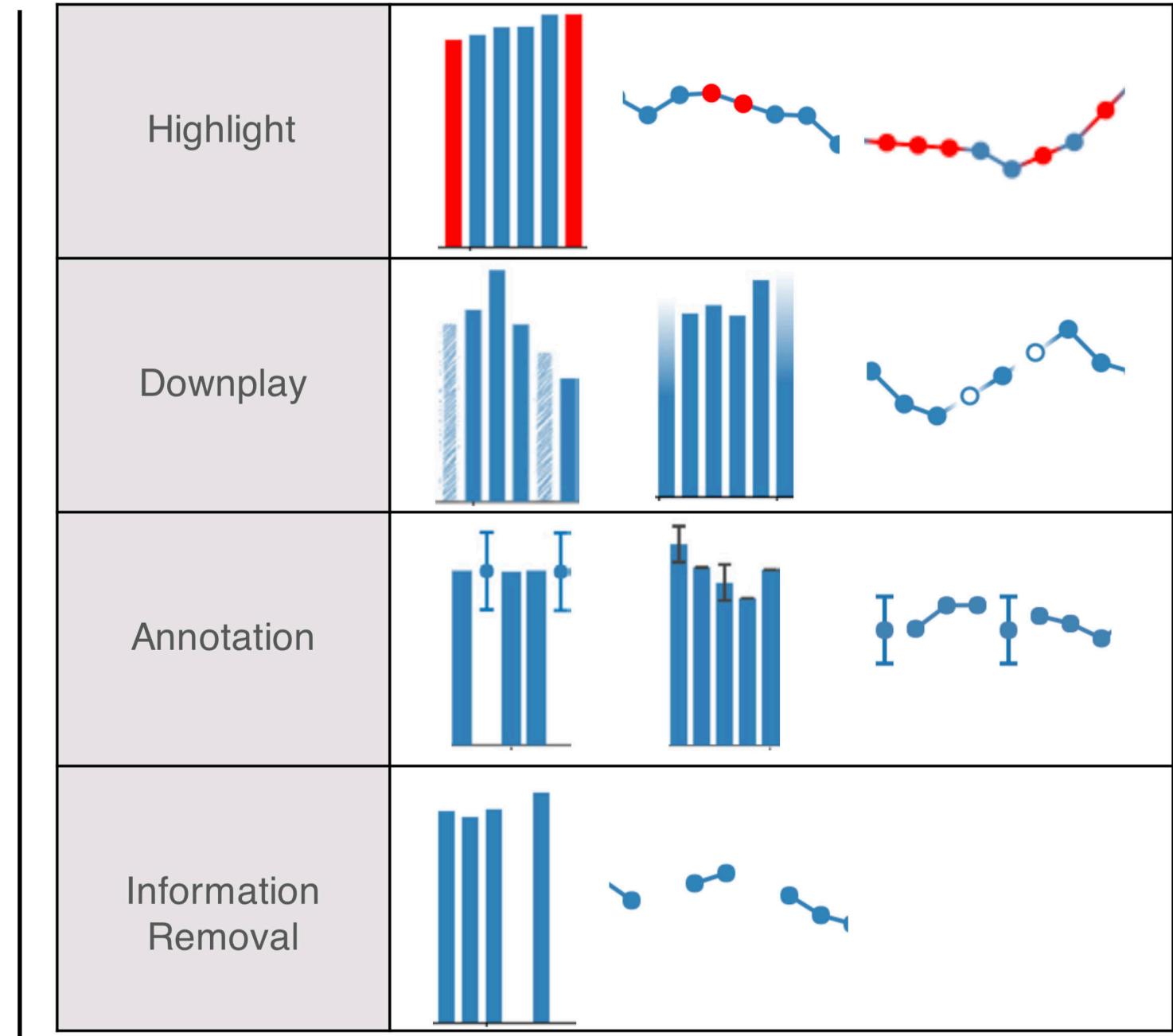
## *Song & Szafir*

The authors were academics at the University of Colorado Boulder.



Visual choices for missing values on perception of data quality

Choice of missing-data visualization alters perceptions of data quality



Perceived data quality and confidence generally degrade as the amount of missing data increases.

Data visualized by **highlighting missing values** tends to be seen as **higher quality than downplay or information removal**.

**Information removal** can significantly **degrade** perceptions of data quality, and confidence. These methods even lead to incorrect responses if missing values break the visual continuity of a visualization.

**Linear interpolation** leads to **higher perceptions** of quality and confidence in analysis.

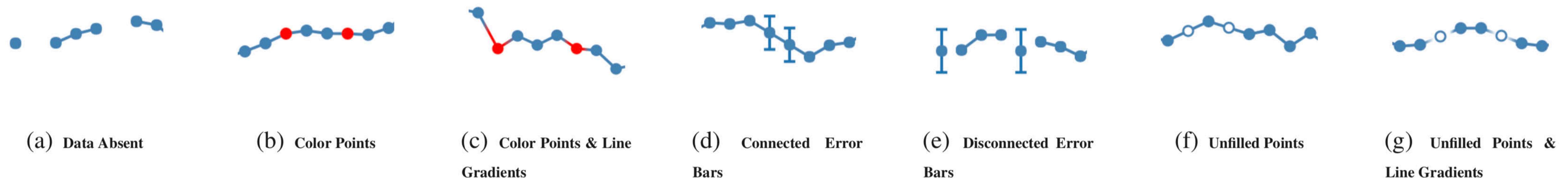


Fig. 4: We tested seven different methods for visualizing missing values in line graphs manipulating both point and line appearance: two highlighting missing values, two downplaying missing values, two annotating missing values, and one removing missing values. .

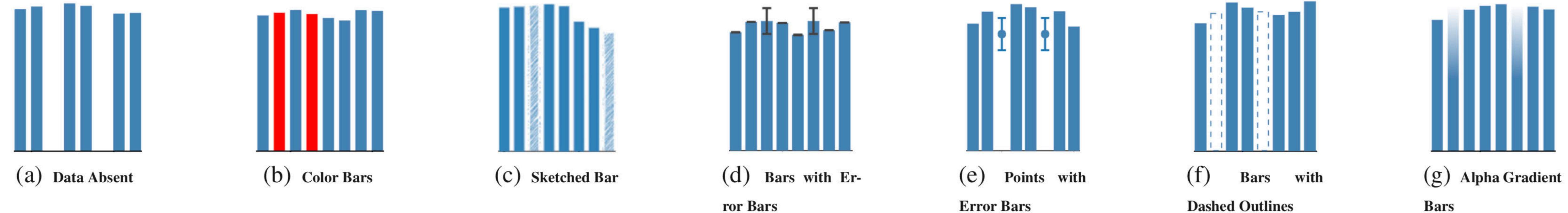
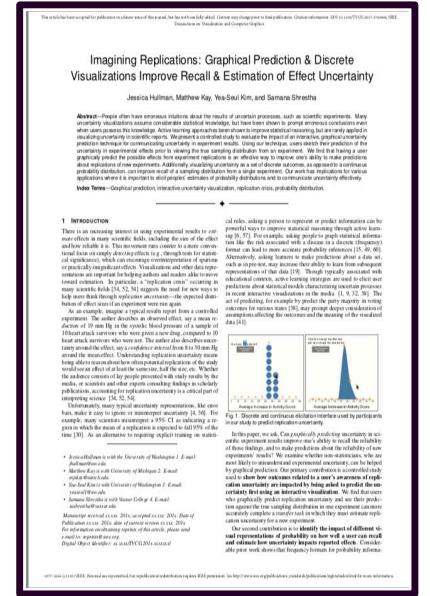


Fig. 6: We tested seven different methods for visualizing missing values in bar charts: one highlighting missing values, three downplaying missing values, two annotating missing values, and one removing missing values.

# Imagining Replications

## Hullman & co-authors

The authors are part of an academic group formally studying the perception of uncertainty, among other things.



## Sketching perception of uncertainty before seeing results improves accuracy

Crowd-sourced study that evaluated the impact of an interactive, graphical uncertainty prediction technique for communicating uncertainty in results. Users sketched their prediction of the uncertainty in experimental effects prior to viewing the true sampling distribution from an experiment.

## Visualizing uncertainty as set of discrete outcomes improved recall

Visualizing uncertainty as a set of discrete outcomes, as opposed to a continuous probability distribution, can improve recall of a sampling distribution from a single experiment.

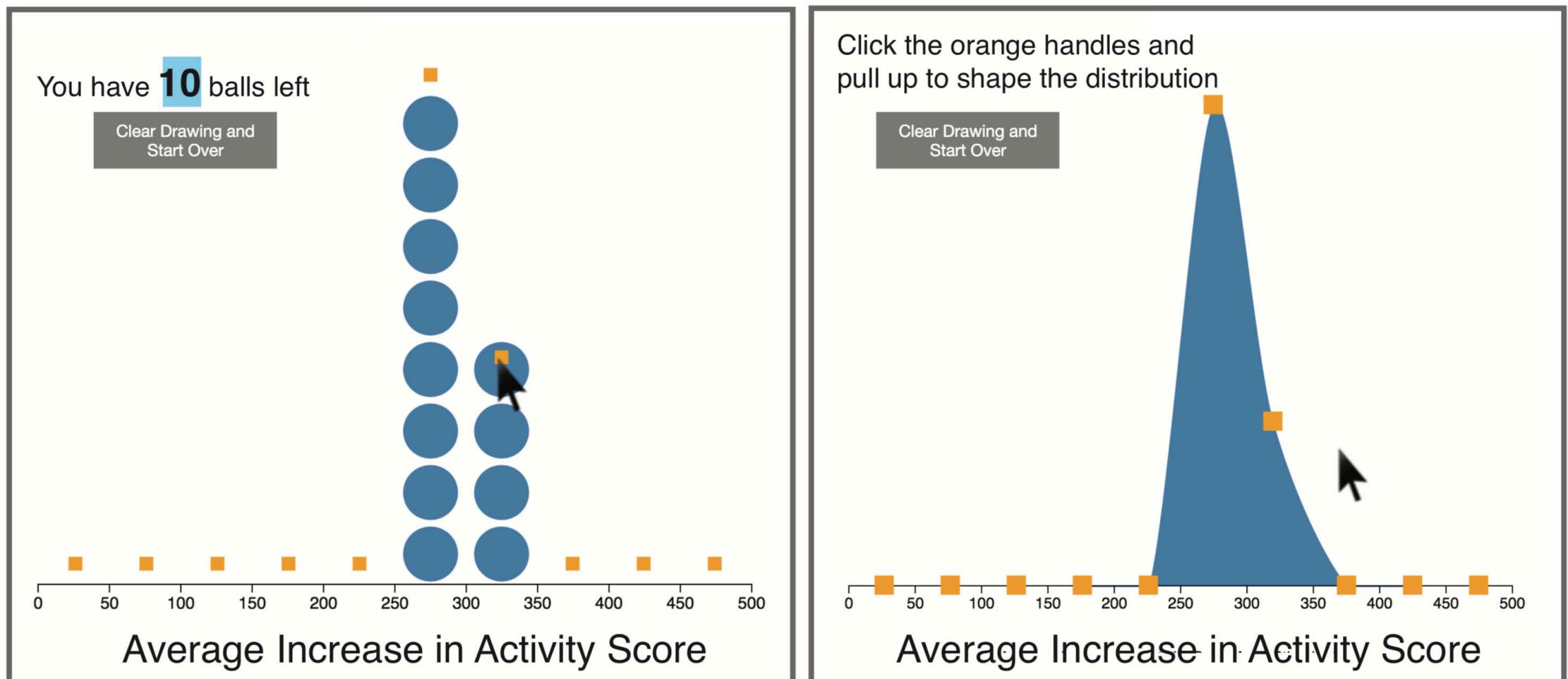


Fig. 1. Discrete and continuous elicitation interface used by participants in our study to predict replication uncertainty.

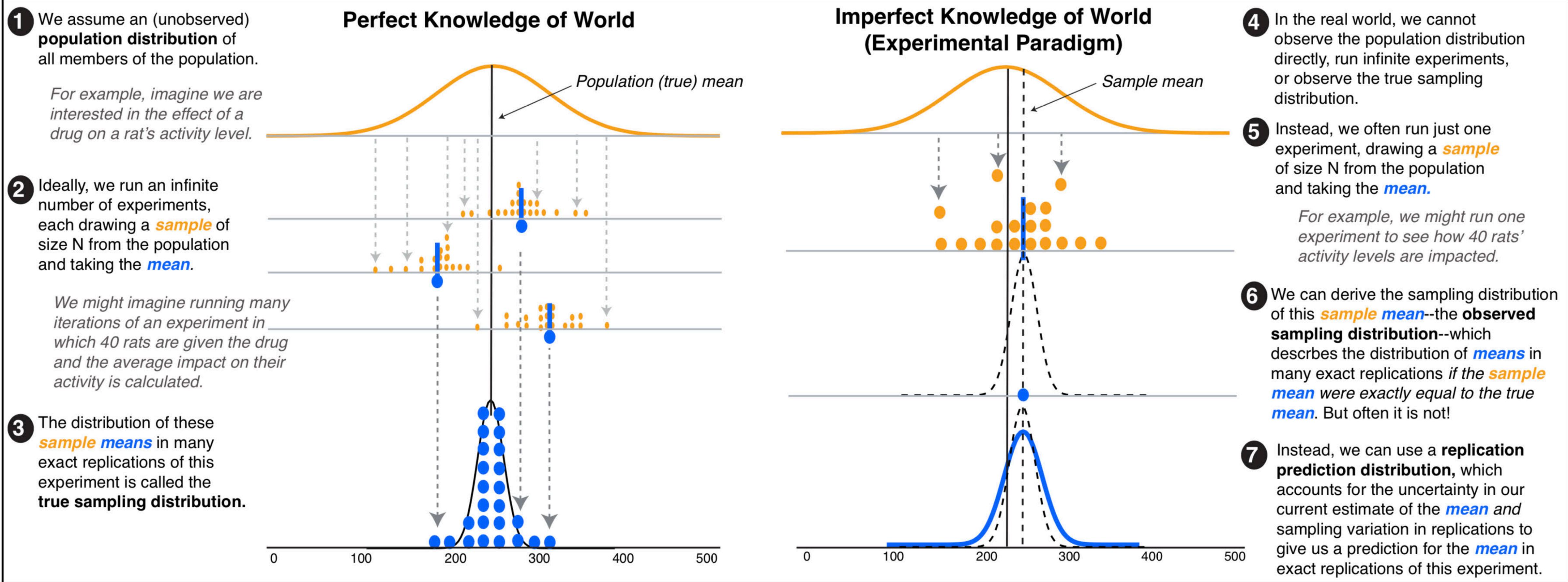


Fig. 2. A depiction of distributions relevant to replication uncertainty, including those based on perfect knowledge of the world (left) and those derived from samples obtained in experimentation (right).

# Information graphics

## Too broad, generic

**infographic** n. a visual image such as a chart or diagram used to represent information or data in an easily understandable form.



All

Images

News

Videos

Books

More

Settings

Tools



education



timeline



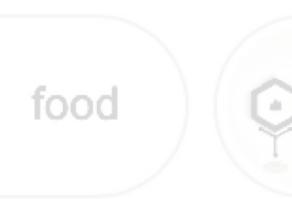
social media



health



food



design



water



technology



business



process

# Want bad examples? Just google.



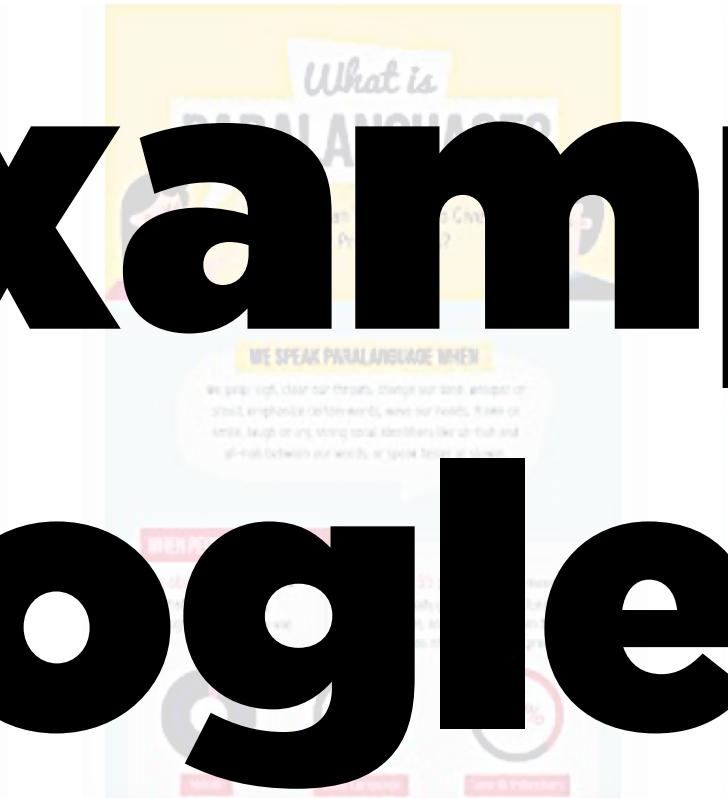
What is an Infographic? Infograp...  
venngage.com



Reports & Charts Online ...  
visme.co



What is an Infographic? Infograp...  
venngage.com



Free Online Infographic Mak...  
canva.com



Reports & Charts Online ...  
visme.co



Free Online Infographic Mak...  
canva.com



Inf...  
her...



Creating Beautiful Infographic...  
blog.hubspot.com



for patients—than ever before. ...  
advancecardiohealth.org



The 4 Most Common Infographic ...  
columnfivemedia.com



venngage.com



edgeforscholars.org  
visme.co



A ru...  
pro...





# 2017 PITCHER HEATMAPS

by Jacob Olsufka

## 2017 PITCHER HEATMAPS

**How to read this heatmap visual:** each column is a start, each row is an inning

### Runs per game

The bars across the top shows a pitcher's trend of total runs allowed over the season.  
A dot represents a quality start (>= 6 innings, <= 3 ER).

### Depth into game

Follow the depth of the bars across the bottom to see the trend of how far into a game a pitcher goes.

### Runs per inning

See which innings pitchers gave up the most total runs with the bars to the right.

### Individual innings

The heatmap shows each inning during a pitcher's season, and when they gave up their runs colored by intensity.

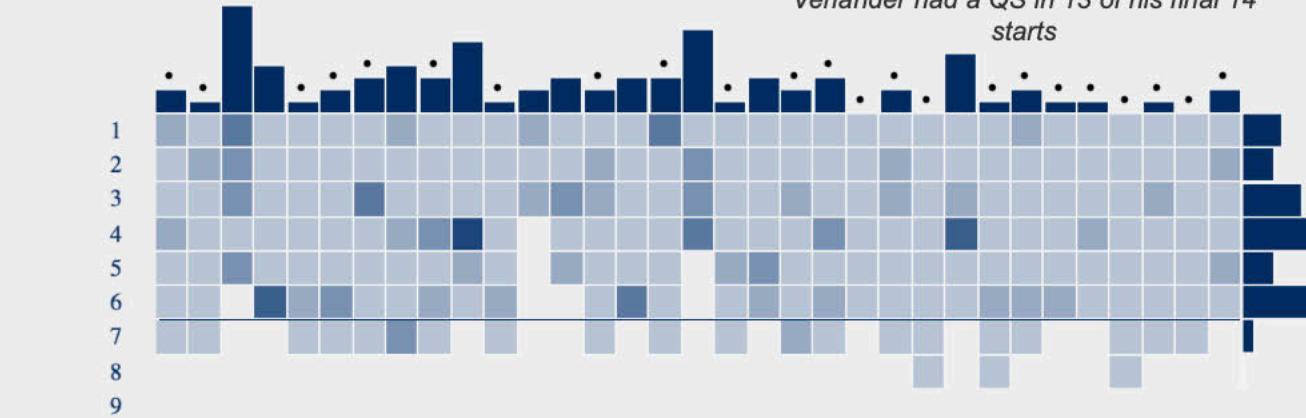
INDICATES LEAGUE  
LEADER



**JUSTIN VERLANDER**

15-8 3.36 ERA 1.17 WHIP 219 SO 23 QS

Verlander had a QS in 13 of his final 14 starts



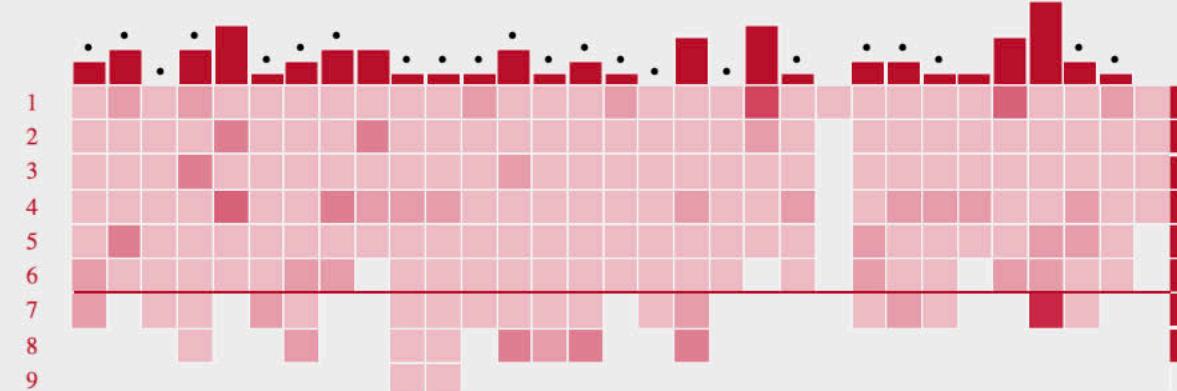
## THE CY YOUNGS



**MAX SCHERZER**

16-6 2.51 ERA 0.90 WHIP 268 SO 22 QS

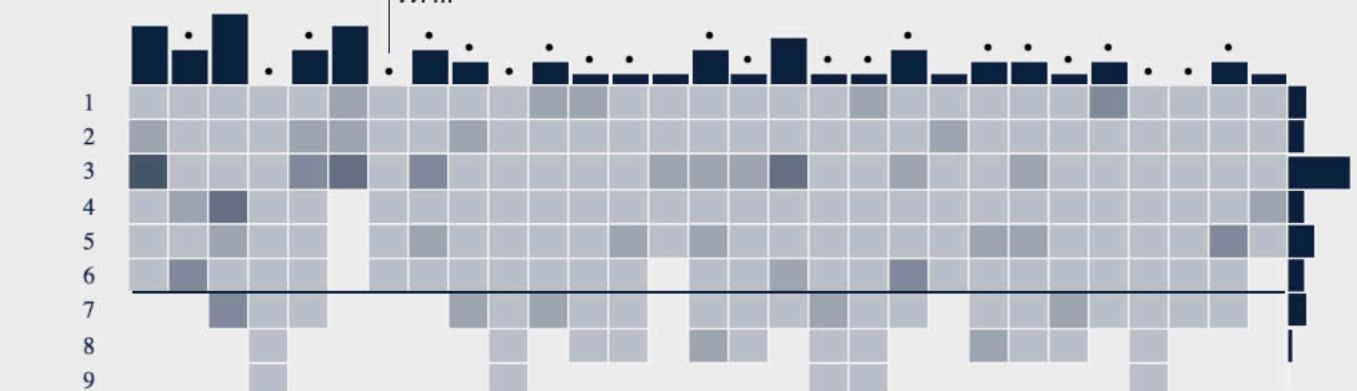
Scherzer led the league with a .178 batting avg against



**COREY KLUBER**

18-4 2.25 ERA 0.87 WHIP 265 SO 22 QS

From June on, Kluber had a 1.62 ERA and 0.76 WHIP



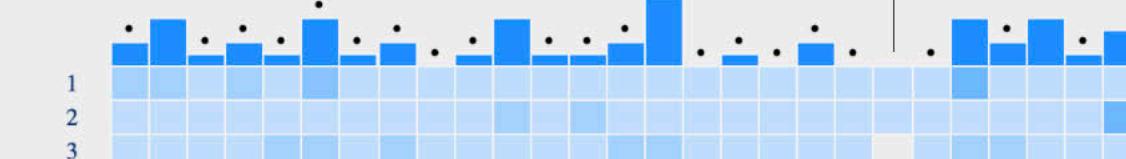
## THE 'MAYBE COULD HAVE WON' THE CY YOUNGS



**CLAYTON KERSHAW**

18-4 2.31 ERA 0.95 WHIP 202 SO 20 QS

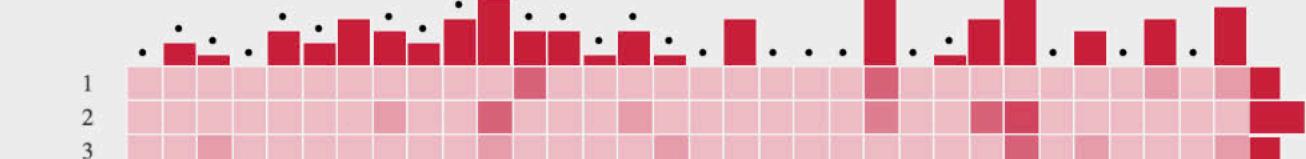
Kershaw left with a back injury



**CHRIS SALE**

17-8 2.90 ERA 0.97 WHIP 308 SO 23 QS

Sale had the most Ks in the AL since 1999



**Without a narrative,  
it's just trivia, list of facts:**



# Data-Driven Storytelling

*Riche, co-editors*

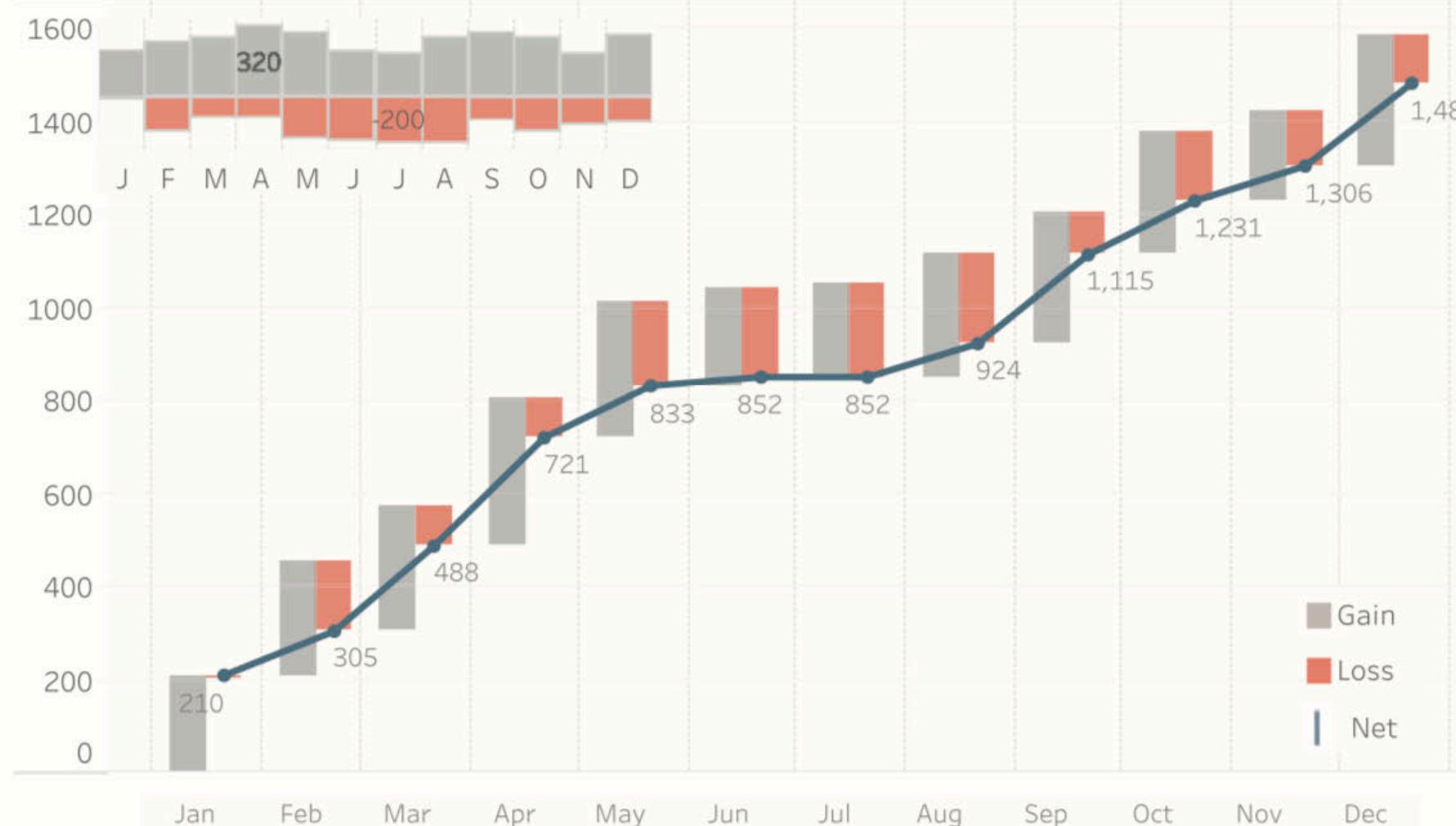
The editors are researchers and professors with focuses on human-computer interaction and information visualization.

“ We differentiated ... “**list of fact**” infographics from the infographics medium as a whole, and chose to **exclude them** because this specific submedium **lacks authorial narrative**. ”

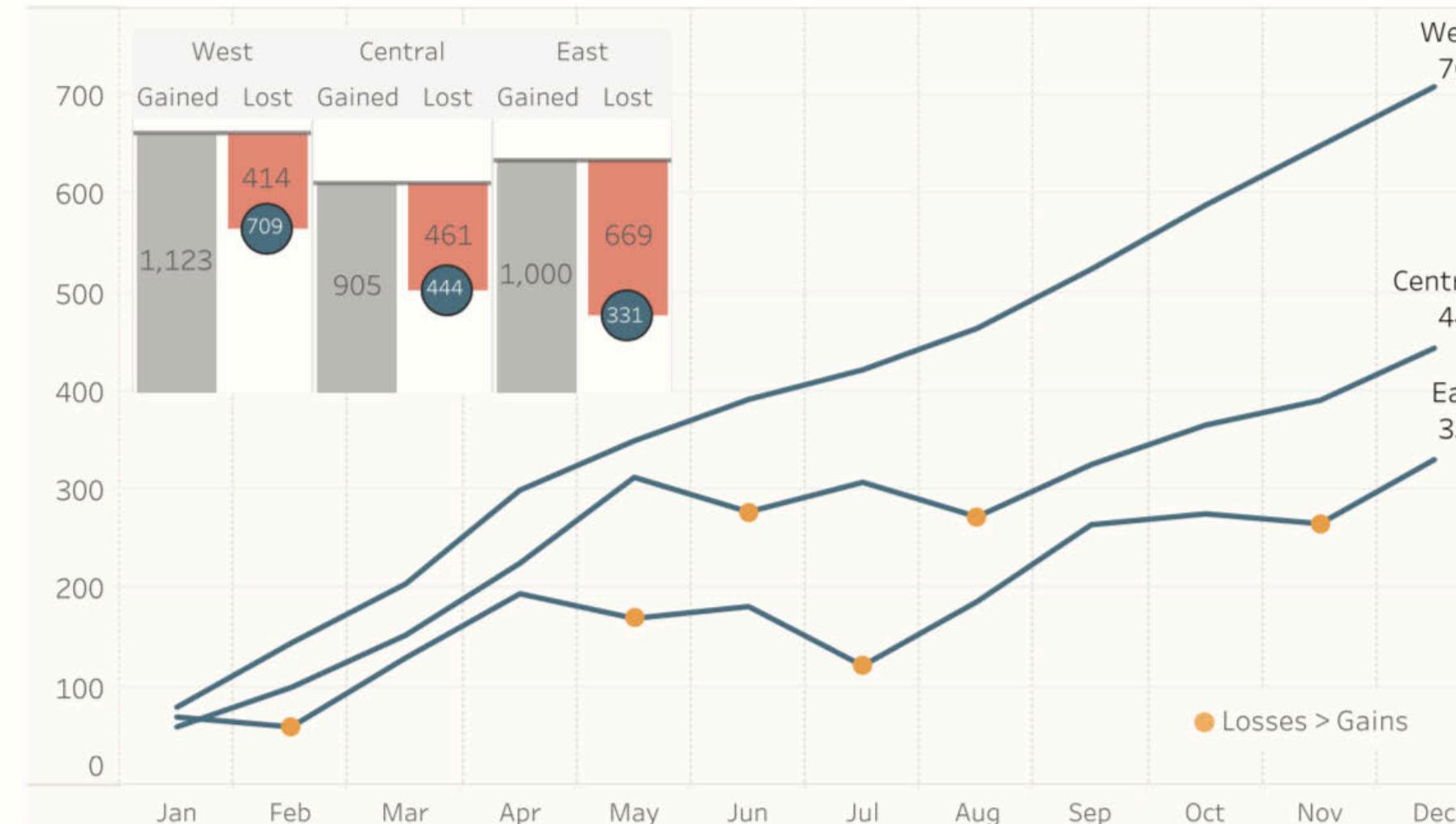
**Dashboards, while sometimes a part of an infographic, may lack narrative or story on their own.**

## Subscriber Churn Analysis

Subscriber activity - All



Net subscriber activity by division



Details

		Gained	Lost	Net	Running total
West	January	80	0	80	80
	February	80	-15	65	145
	March	90	-30	60	205
	April	120	-25	95	300
	May	100	-50	50	350
	June	119	-77	42	392
	July	75	-45	30	422
	August	119	-77	42	464
	September	90	-30	60	524
	October	80	-15	65	589
	November	80	-20	60	649
	December	90	-30	60	709
	Total	1,123	-414	709	
Central	January	60	0	60	60
	February	85	-45	40	100
	March	80	-27	53	153
	April	90	-17	73	226
	May	120	-33	87	313
	June	45	-80	-35	278
	July	75	-45	30	308
	August	45	-80	-35	273
	September	80	-27	53	326
	October	85	-45	40	366
	November	60	-35	25	391
	December	80	-27	53	444
	Total	905	-461	444	
East	January	70	0	70	70
	February	80	-90	-10	60
	March	100	-30	70	130
	April	110	-45	65	195
	May	70	-95	-25	170
	June	45	-33	12	182
	July	50	-110	-60	122
	August	99	-34	65	187
	September	112	-34	78	265
	October	99	-88	11	276
	November	55	-65	-10	266
	December	110	-45	65	331
	Total	1,000	-669	331	
<b>Grand Total</b>					
<b>3,028      -1,544      1,484</b>					

# Malofiej—Infographics World Summit

“The Pulitzer Prizes  
of Infographics”

## Interviews of Malofiej speakers on infographics



**Gregor Aisch**

Infographics is an abbreviated form of “information graphics”. It seems to mean a lot of different things to different people. I rarely use the term.



**Federica Fragapane**

A visual translation of data and information: a language to communicate topics, contents and **stories** to people.



**Laris Karklis**

Infographics is . . . using visuals **to tell a story**.



**Nadieh Bremer**

**Infographics ... combine graphical elements**, such as a drawn portion of an animal, human, map, etc. with small mini **data visualizations** (a small bar chart for example) and **annotations** **to tell a story**.

# We want information graphics to ...

---

**Tell a complete story** where the purpose is to inform, entertain or persuade the audience. It should:

simple, focused messages

new, surprising information

credible data sources

visually coherent, integrated

use comparisons for context, meaning

principles of information design, organized



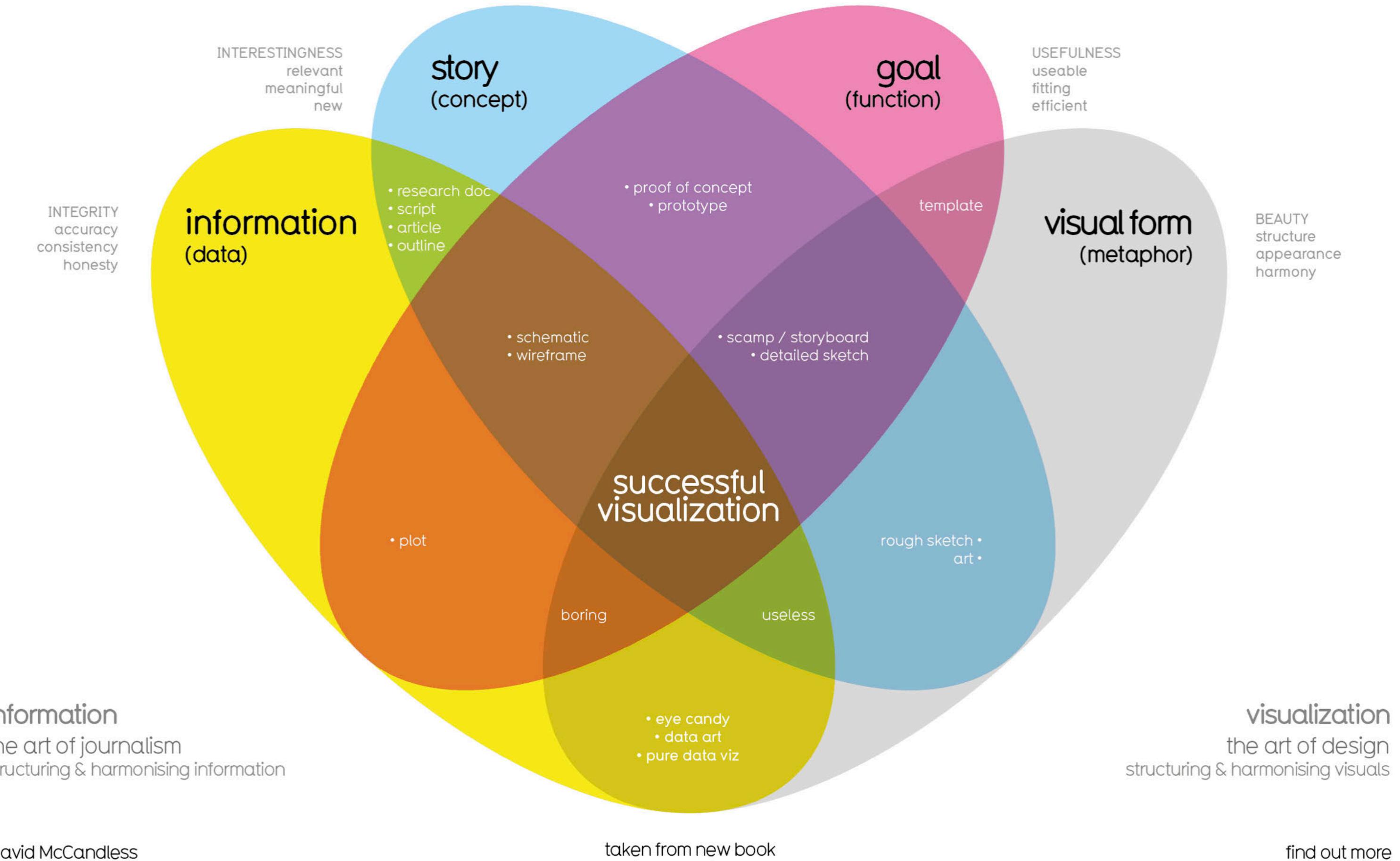
# Information is Beautiful

# McCandless

Founder of Information is Beautiful, David is a writer, designer, creative director and artist.

# What Makes a Good Visualization?

## explicit (implicit)



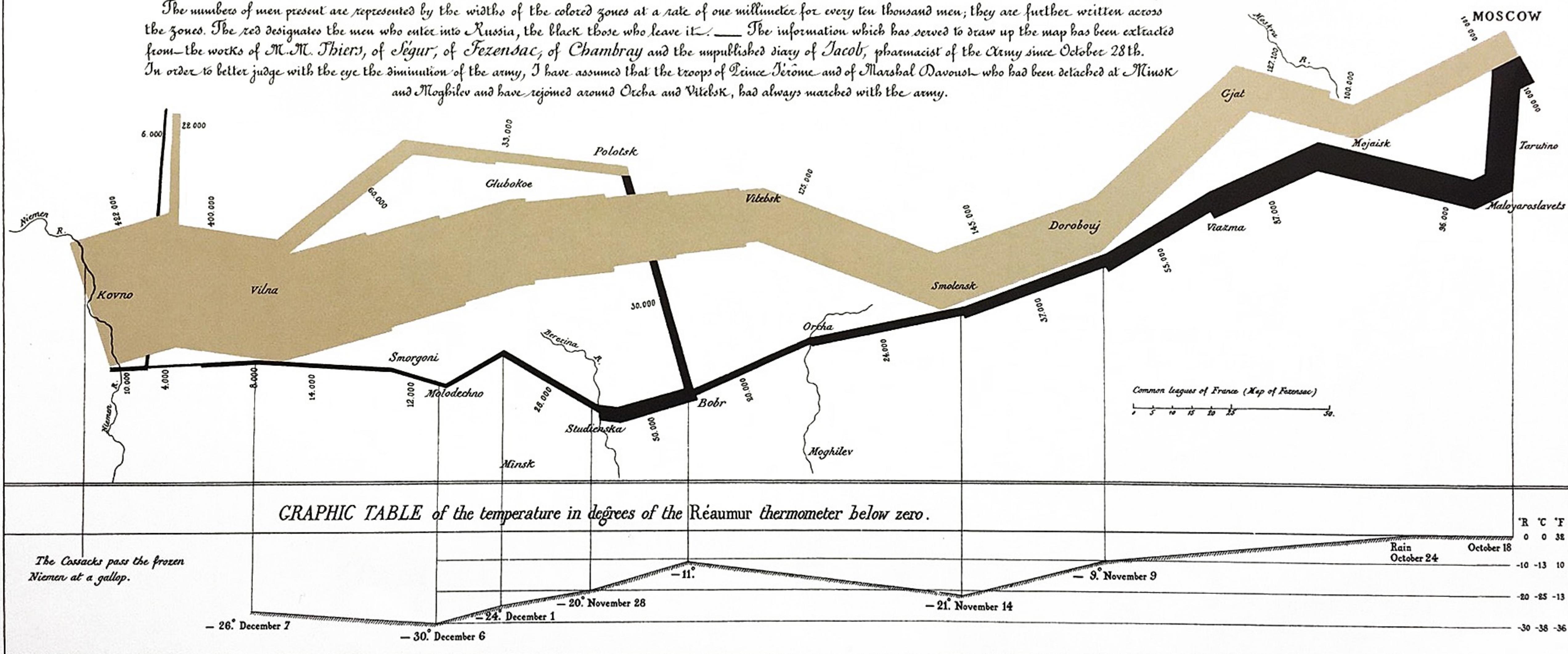
# **Information graphics, examples for discussion**

# Figurative Map of the successive losses in men of the French Army in the Russian Campaign 1812 ~1813.

Drawn up by M. Minard, Inspector General of Bridges and Roads in retirement.

Paris, November 20, 1869.

The numbers of men present are represented by the widths of the colored zones at a rate of one millimeter for every ten thousand men; they are further written across the zones. The red designates the men who enter into Russia, the black those who leave it. — The information which has served to draw up the map has been extracted from the works of M.M. Thiers, of Séguir, of Fezensac, of Chambray and the unpublished diary of Jacob, pharmacist of the Army since October 28th. In order to better judge with the eye the diminution of the army, I have assumed that the troops of Prince Jérôme and of Marshal Davout who had been detached at Minsk and Moghilev and have rejoined around Orsha and Vitebsk, had always marched with the army.



**Table 20.1** Napoleon's March Data

lonc	latc	city	lont	temp	date	lonp	latp	survivors	direction	group
24.0	55.0	Kowno	37.6	0	Oct 18	24.0	54.9	340,000	A	I
25.3	54.7	Wilna	36.0	0	Oct 24	24.5	55.0	340,000	A	I
26.4	54.4	Smorgoni	33.2	-9	Nov 9	25.5	54.5	340,000	A	I
26.8	54.3	Molodexno	32.0	-21	Nov 14	26.0	54.7	320,000	A	I
27.7	55.2	Gloubokoe	29.2	-11		27.0	54.8	300,000	A	I
27.6	53.9	Minsk	28.5	-20	Nov 28	28.0	54.9	280,000	A	I
28.5	54.3	Studienska	27.2	-24	Dec 1	28.5	55.0	240,000	A	I
28.7	55.5	Polotzk	26.7	-30	Dec 6	29.0	55.1	210,000	A	I
29.2	54.4	Bobr	25.3	-26	Dec 7	30.0	55.2	180,000	A	I
30.2	55.3	Witebsk				30.3	55.3	175,000	A	I
30.4	54.5	Orscha				32.0	54.8	145,000	A	I
30.4	53.9	Mohilow				33.2	54.9	140,000	A	I
32.0	54.8	Smolensk				34.4	55.5	127,100	A	I
33.2	54.9	Dorogobouge				35.5	55.4	100,000	A	I
34.3	55.2	Wixma				36.0	55.5	100,000	A	I
34.4	55.5	Chjat				37.6	55.8	100,000	A	I
36.0	55.5	Mojaisk				37.7	55.7	100,000	R	I
37.6	55.8	Moscou				37.5	55.7	98,000	R	I
36.6	55.3	Tarantino				37.0	55.0	97,000	R	I
36.5	55.0	Malo-jarosewli				36.8	55.0	96,000	R	I
						35.4	55.3	87,000	R	I
						34.3	55.2	55,000	R	I
						33.3	54.8	37,000	R	I
						32.0	54.6	24,000	R	I
						30.4	54.4	20,000	R	I
						29.2	54.3	20,000	R	I
						28.5	54.2	20,000	R	I
						28.3	54.3	20,000	R	I
						27.5	54.5	20,000	R	I
						26.8	54.3	12,000	R	I
						26.4	54.4	14,000	R	I
						25.0	54.4	8,000	R	I
						24.4	54.4	4,000	R	I
						24.2	54.4	4,000	R	I
						24.1	54.4	4,000	R	I
						24.0	55.1	60,000	A	II
						24.5	55.2	60,000	A	II
						25.5	54.7	60,000	A	II
						26.6	55.7	40,000	A	II
						27.4	55.6	33,000	A	II
						28.7	55.5	33,000	A	II
						28.7	55.5	33,000	R	II
						29.2	54.2	30,000	R	II
						28.5	54.1	30,000	R	II
						28.3	54.2	28,000	R	II
						24.0	55.2	22,000	A	III
						24.5	55.3	22,000	A	III
						24.6	55.8	6,000	A	III
						24.6	55.8	6,000	R	III
						24.2	54.4	6,000	R	III
						24.1	54.4	6,000	R	III

# Winner, Information is Beautiful Award

TASS



TASS



1812

## WHEN NAPOLEON VENTURED EAST

HOW THE 1812 PATRIOTIC WAR TURNED NAPOLEON'S GRAND ARMY INTO A HANDFUL OF SURVIVORS

PYC ENG



## Nobels no degrees

This visualization explores Nobel Prizes and graduate qualifications from 1901 to 1912, by analysing the age of recipients at the time prizes were awarded, average age evolution through time and among categories, graduation grades, main university affiliations and the principal hometowns of the graduates.

### How to read it?

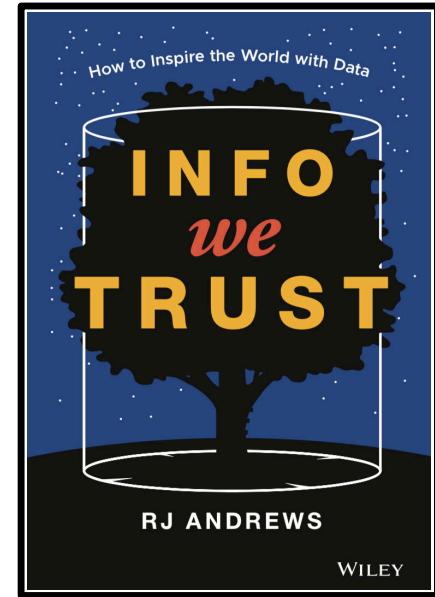
Each dot represents a Nobel laureate, each recipient is positioned according to the year the prize was awarded (x axis) and age of the person at the time of the award (y axis).



# Winner, Information is Beautiful Award

## Lupi, Fragapane

The co-authors work together at Lupi's design firm, whom we've previously met when considering this visualization.

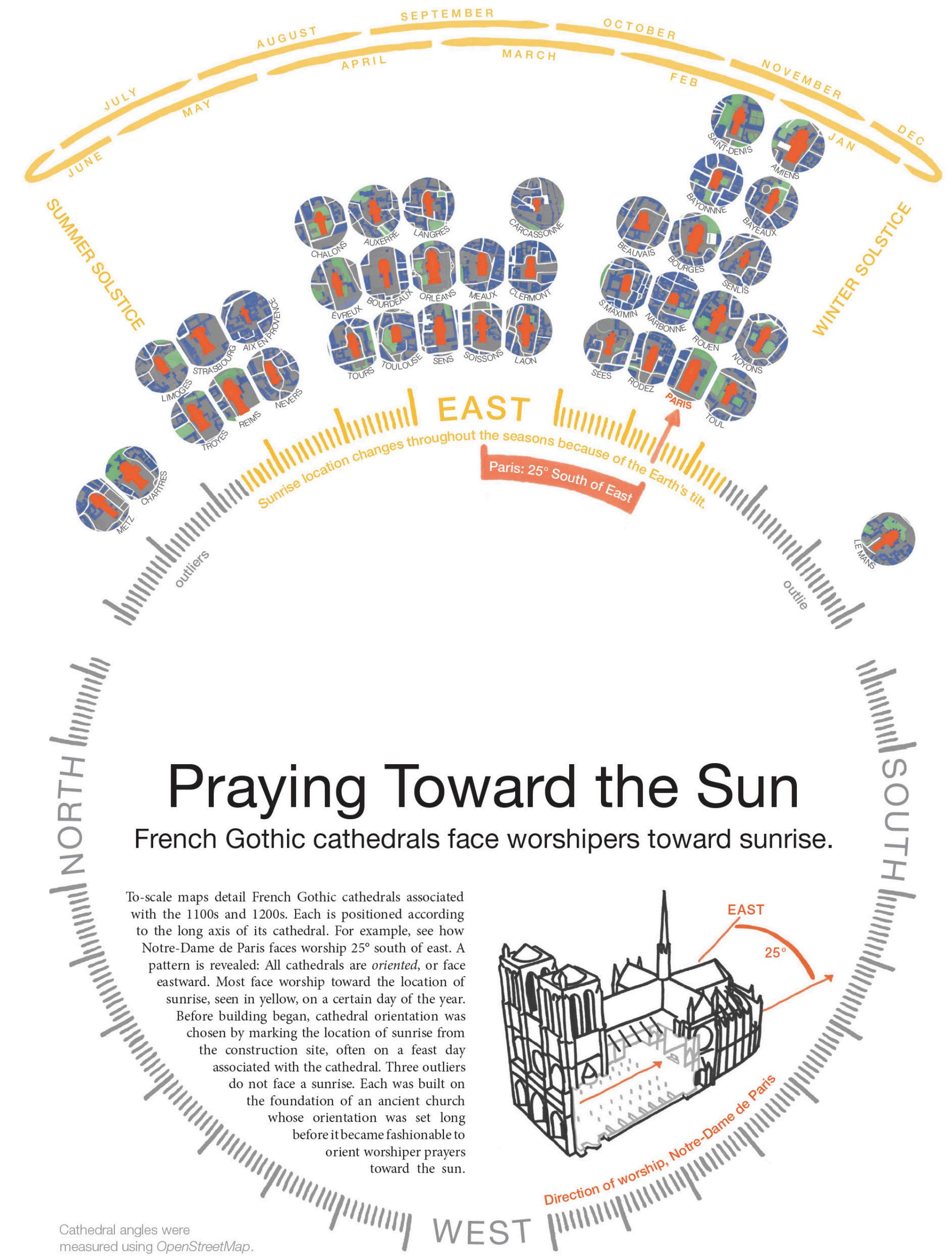


# Info We Trust

## How to inspire the world with data

# Andrews

He is a data storyteller. His book is an adventure exploring how to inspire the world with data. RJ is the creator of [www.infowetrust.com](http://www.infowetrust.com), where he makes available some of his data stories.



## The Cost of Mobile Ads on 50 News Websites

By GREGOR AISCH, WILSON ANDREWS and JOSH KELLER OCT. 1, 2015

Ad blockers, which Apple first allowed on the iPhone in September, promise to conserve data and make websites load faster. But how much of your mobile data comes from advertising? We measured the mix of **advertising** and **editorial** on the mobile home pages of the top 50 news websites – including ours – and found that **more than half of all data came from ads** and other content filtered by ad blockers. Not all of the news websites were equal. [RELATED ARTICLE](#)

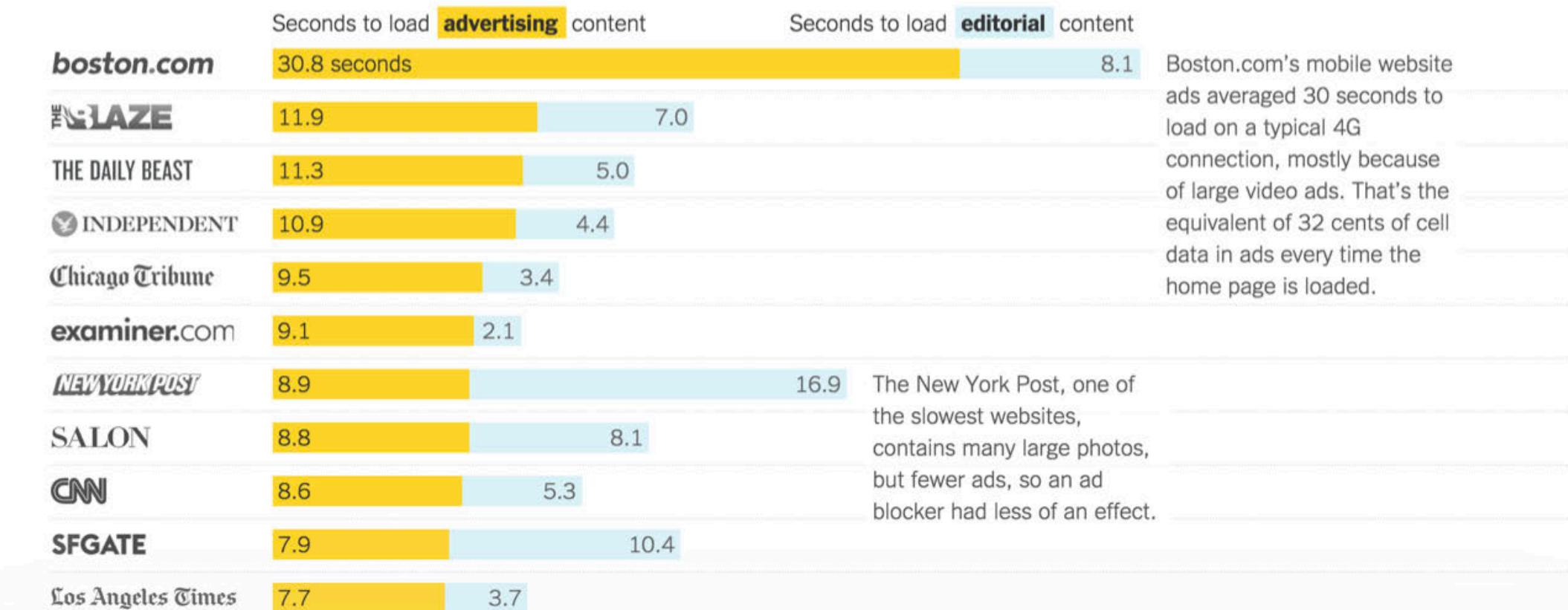
Estimated load time on a 4G LTE network	Data usage to load mobile home page	Cost per page on a typical data plan
--	--	---



# Graphics portfolio, NYT Infographic

# Aisch

Gregor was graphics editor at the NY Times, and recipient of infographics awards at Malofiej.



Boston.com's mobile website ads averaged 30 seconds to load on a typical 4G connection, mostly because of large video ads. That's the equivalent of 32 cents of cell data in ads every time the home page is loaded.

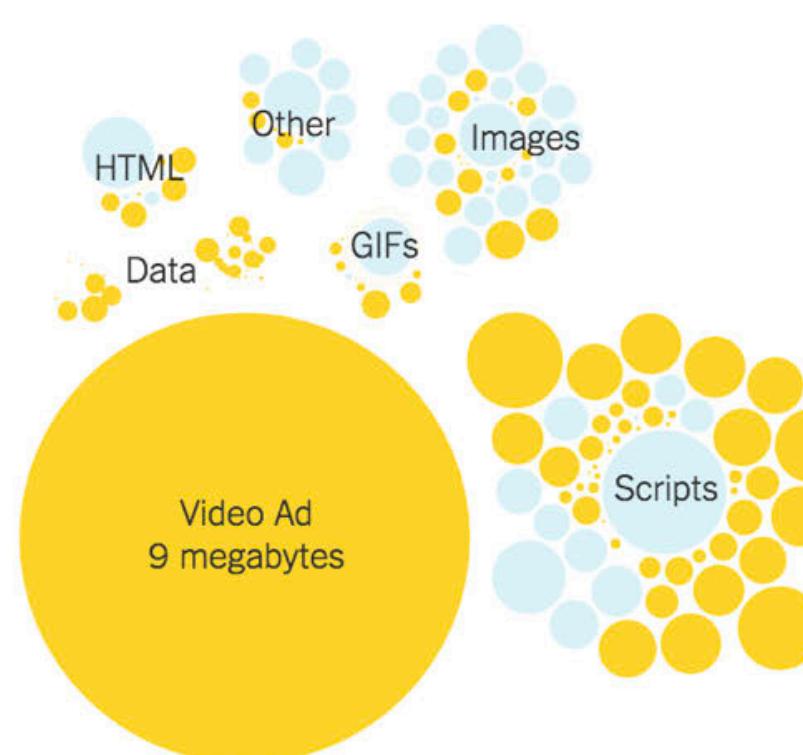
The New York Post, one of the slowest websites, contains many large photos, but fewer ads, so an ad blocker had less of an effect.

### boston.com

Here are all the files that made up the Boston.com data during one visit, including one large video ad and many script files used by ad networks. With an ad blocker, those files were gone.

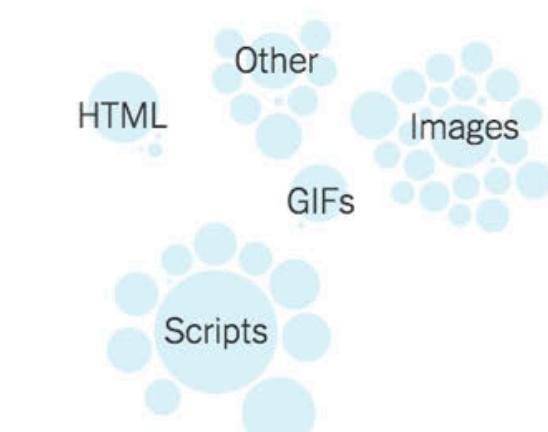
#### Without ad blocker

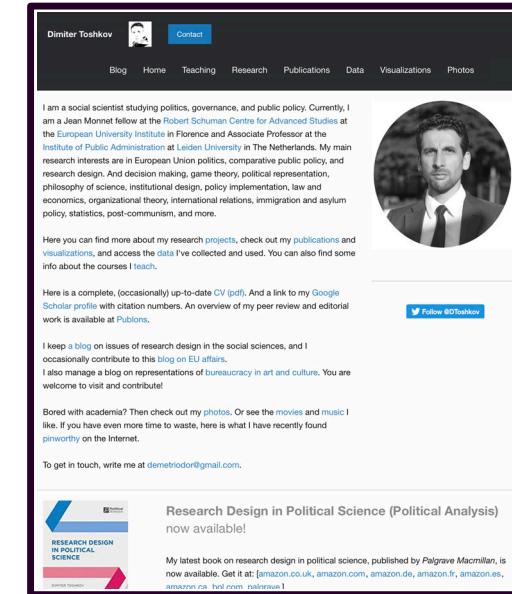
389 files, 16.3 megabytes, 33 seconds



#### With ad blocker

52 files, 3.5 megabytes, 7 seconds



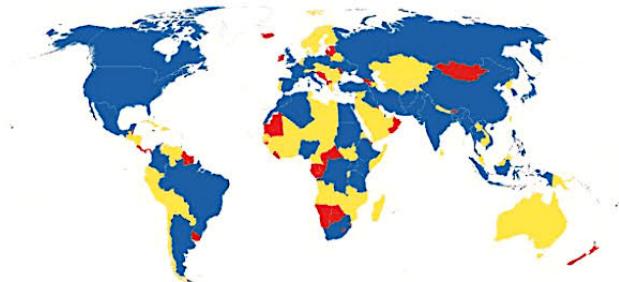
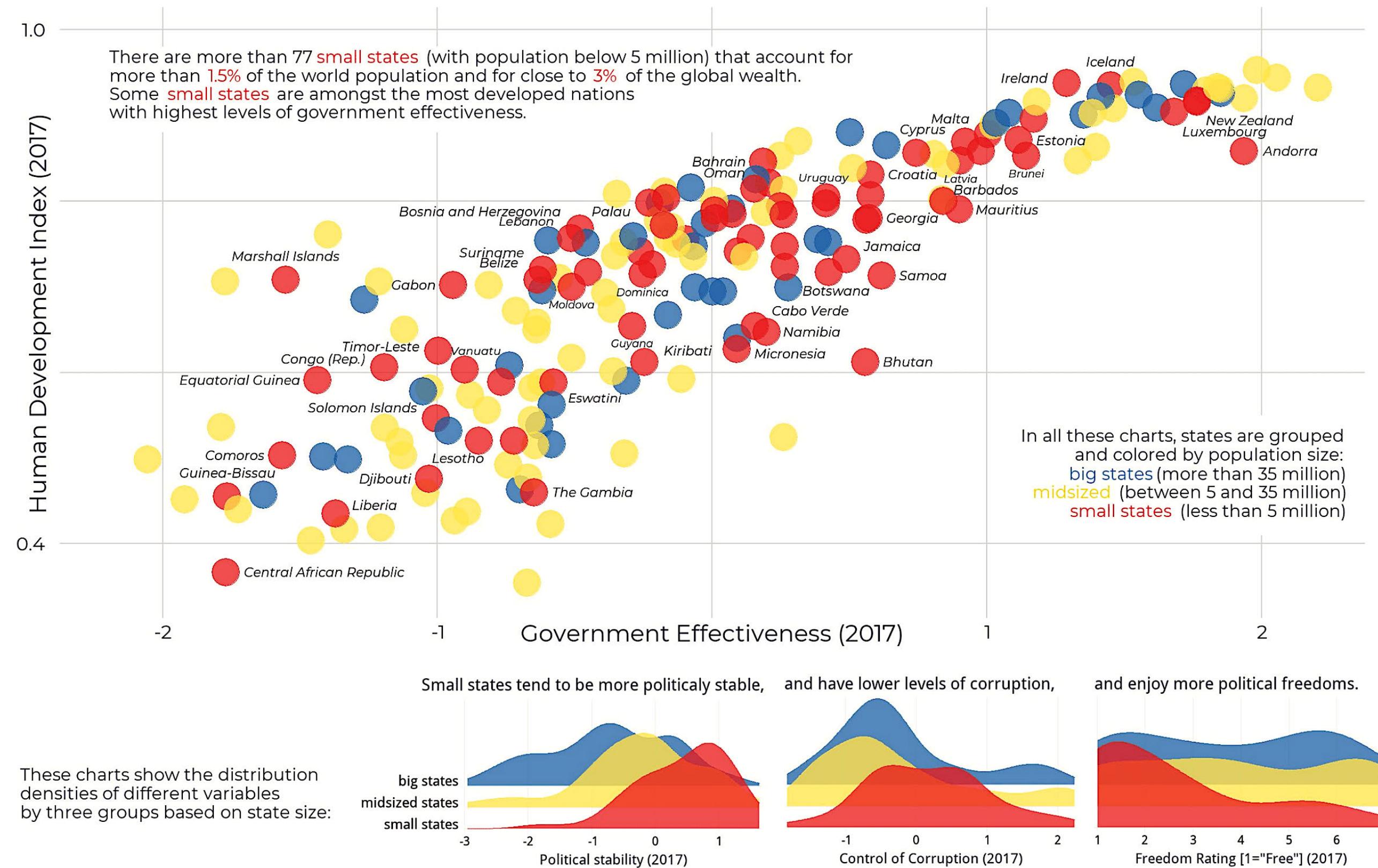


# Winner, Information is Beautiful Award

Toshkov

Associate Professor at the Institute of Public Administration, won an award for this infographic, which was made using R.

## Small States Can Be Big Players in Development and Good Governance



But small states have improved less

and have even lost some ground.

الجائزـةـ العـالـمـيـةـ لـفـنـ عـرـضـ لـلـبـلـيـانـاتـ WORL~D DATA VISUALIZATION PRIZE

CBO

Even the Congressional Budget Office relies on infographics to convey large tomes of information in summary form.



## Summary of long term budget report

Congressional Budget Office

# The 2012 Long-Term Budget Outlook

June 2012

CBO's long-term projections reflect two broad scenarios:

EB

### CBO's Extended Baseline Scenario

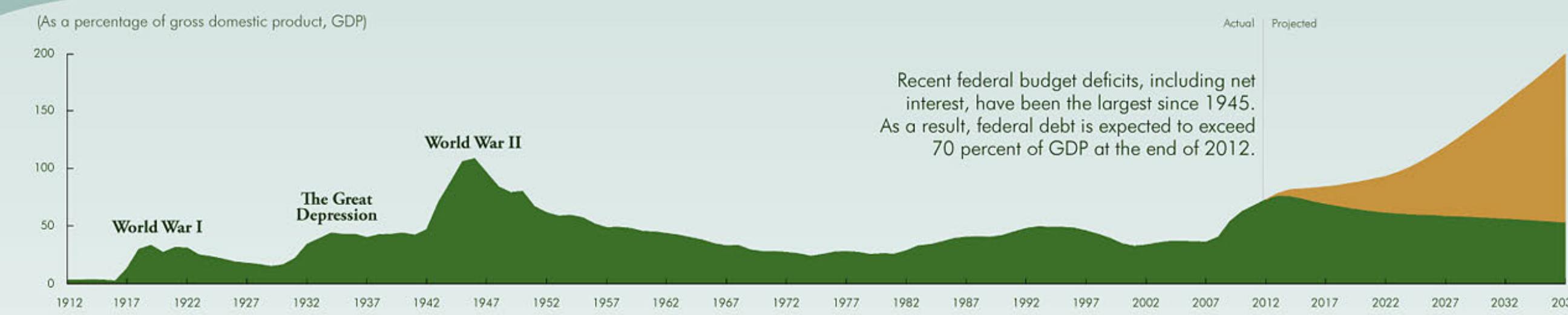
Reflects the assumption that current laws generally remain unchanged, implying that lawmakers will allow tax increases and spending cuts scheduled under current law to occur and that they will forgo measures routinely taken in the past to avoid such changes. Noninterest spending continues to rise, however, pushed up by the aging of the population and the rising costs of health care, and revenues reach historically high levels.

EAF

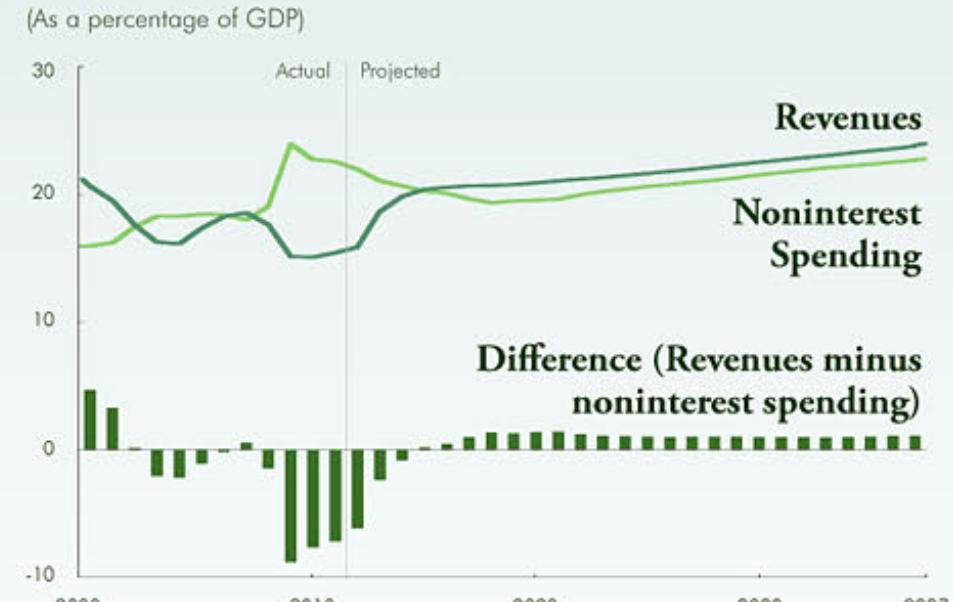
### CBO's Extended Alternative Fiscal Scenario

Maintains what might be deemed current policies, as opposed to current laws, implying that lawmakers will extend most tax cuts and other forms of tax relief currently in place but set to expire and that they will prevent automatic spending reductions and certain spending restraints from occurring. Therefore, revenues remain near their historical average, and the gap between noninterest spending and revenues widens over the long term.

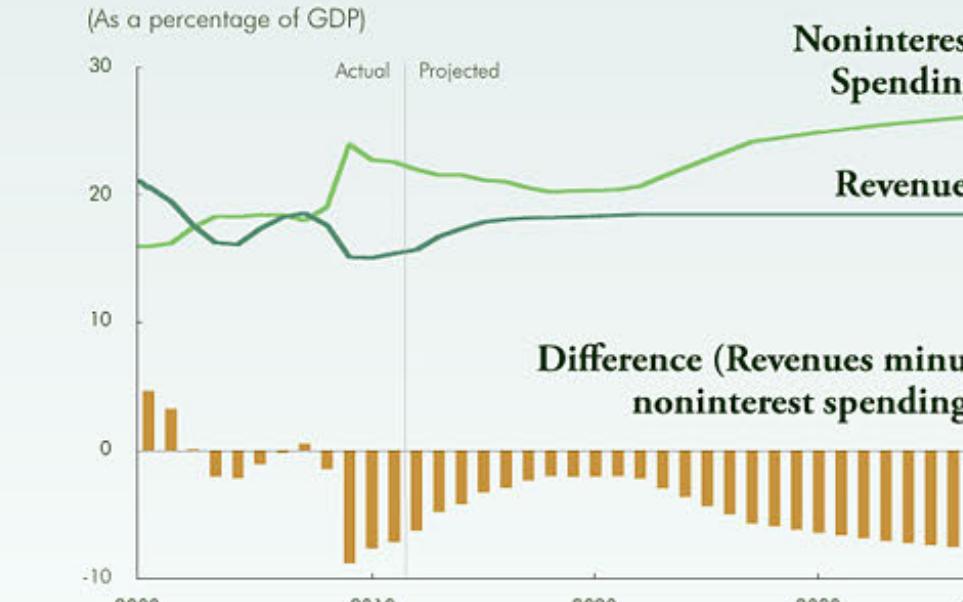
#### Federal Debt Held by the Public, Historically and Projected Under Two Policy Scenarios



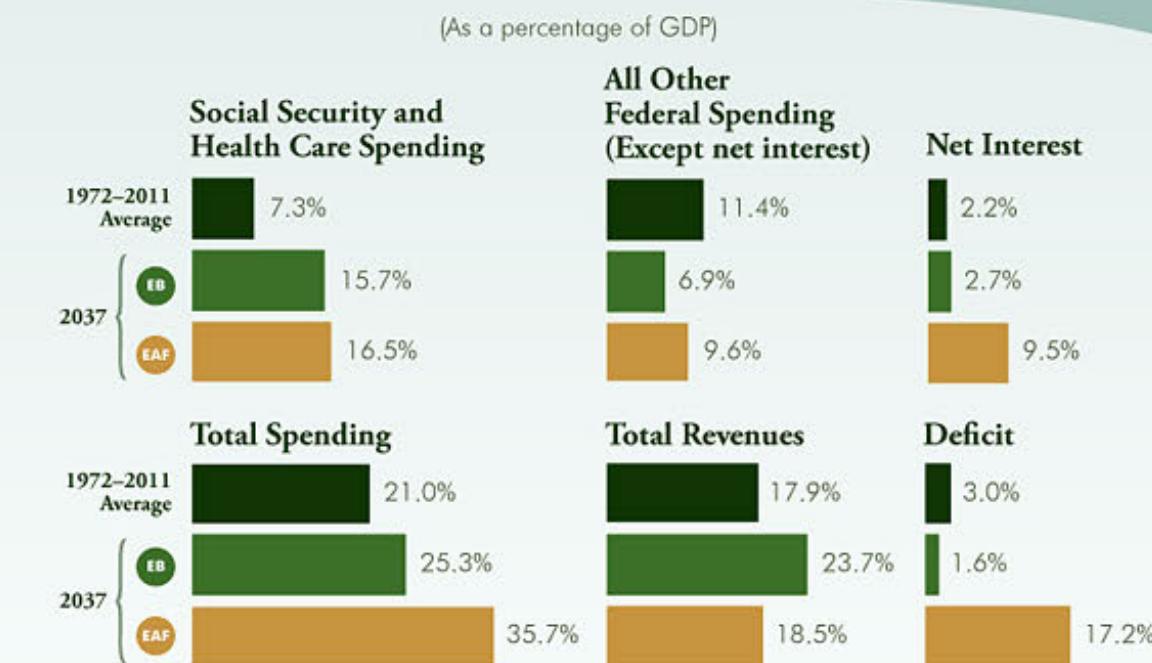
#### Extended Baseline Scenario EB



#### Extended Alternative Fiscal Scenario EAF



#### Components of the Federal Budget



Prepared by Maureen Costantino and Jonathan Schwabish  
Contact: Long-Term Modeling Group



Sources: Congressional Budget Office; Office of Management and Budget  
For details, see *The 2012 Long-Term Budget Outlook*, June 2012; <http://go.usa.gov/dKY>



# Winner, Information is Beautiful Award

*Bremmer*

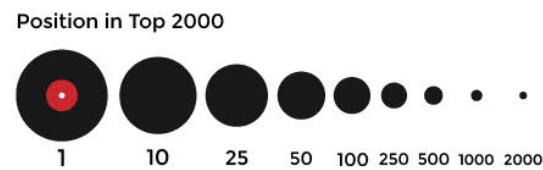
Previously an astronomer and analyst at a consulting company, Nadieh is a powerhouse freelance data visualization designer, and winner of numerous awards.

## TOP 2000 ❤️ 70's & 80's

Since 1999 the 2000 most popular songs of all time, as voted by the show's audience, are played on Dutch national Radio 2 in a yearly marathon. The 2000 songs are on the air between noon on December 25th until New Year's Eve and over half of the Dutch population listens to the Top 2000 each year.

Each ● to the right represents a song in the Top 2000. It is placed according to its year of release. In the legend below you can see what the size and color of a song means.

The bulk of the songs and most of the top 10 are from the 70's & 80's...



Highest position reached in weekly Top 40



Golden oldie

The oldest song in the list, Billie Holiday's *Strange Fruit*, is from 1939. It's 17 years older than the second-oldest song. If it will make the 2017 edition remains to be seen, it's barely in now, on position 189.

Year of release

Spread across release years of the 2000 songs  
For 4 editions of the Top 2000

The charts on the right represent all 2000 songs from 3 past editions of the Top 2000 (held in 2000, 2005, 2010) and the most recent 2016 edition.

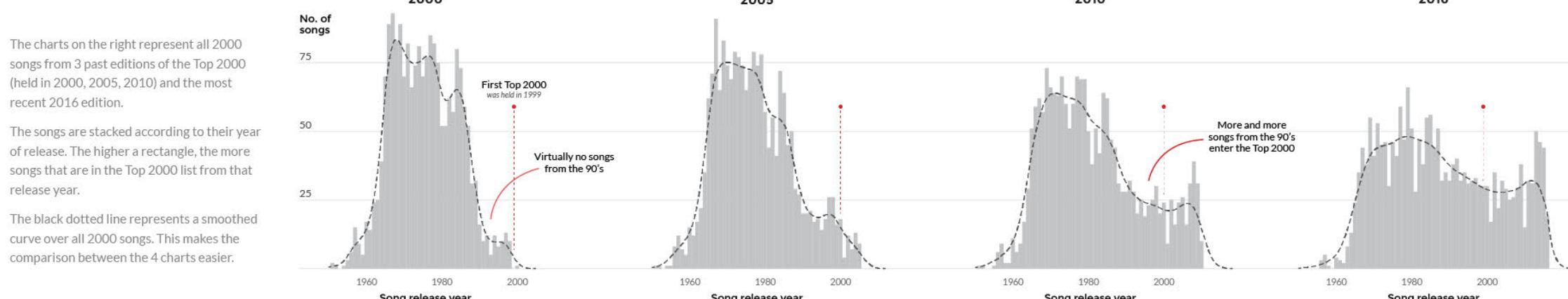
The songs are stacked according to their year of release. The higher a rectangle, the more songs that are in the Top 2000 list from that release year.

The black dotted line represents a smoothed curve over all 2000 songs. This makes the comparison between the 4 charts easier.

Created by Nadieh Bremer | VisualCinnamon.com for the December edition of data sketch|es

Visit [tinyurl.com/2016top2000](http://tinyurl.com/2016top2000) for the interactive visual and see the name & title of each song

Data | Top 2000 list from Radio 2 | Top 40 info from Mediamarkt's Top 40



But they're losing tracks to the new Millennium

It makes sense that the Top 2000 will be more spread out for each new edition, since there are more songs to choose from. However, if we compare the distributions of the Top 2000 songs over 4 editions, we see that, especially, the 90's has been gaining a lot of popularity.

Even though all songs from the 90's were out in the 2000 edition, only a few songs from that decade were chosen. Whereas in the 2016 edition the number of songs from the 90's has risen significantly. This could be due to a new generation who has grown up during the 90's taking over from those who voted in the early 2000's (who apparently didn't appreciate the new music).

# CitiBike example, infographic

# CITI BIKE HOURLY BALANCE

IMBALANCE HOTSPOTS - A.M. PEAK HOUR - AVG. WEEKDAY - OCTOBER 2013

● OUTLIER STATIONS (BALANCED STATIONS WITHIN AN IMBALANCE HOTSPOT)

DESTINATIONS

ORIGINS

Spatial Information Design Lab - GSAPP - Columbia University

0 0.25 0.5 MILES

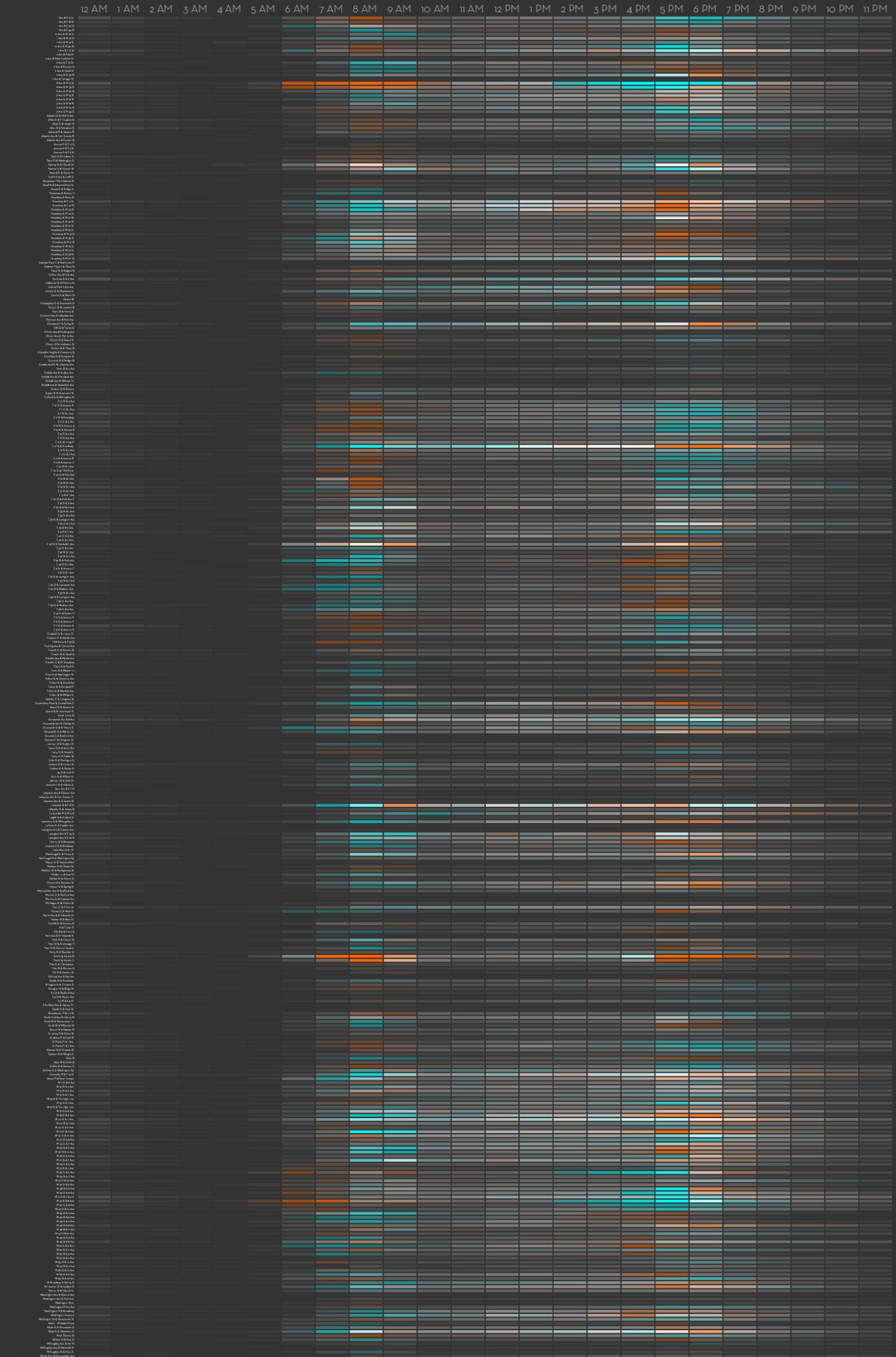
# CITI BIKE HOURLY ACTIVITY AND BALANCE

ACTIVITY AND IMBALANCE MATRIX - AVG. WEEKDAY - OCTOBER 2013

DESTINATIONS

ORIGINS

ACTIVITY GRADIENT



Spatial Information Design Lab - GSAPP - Columbia University



# CitiBike example, Ride against the flow

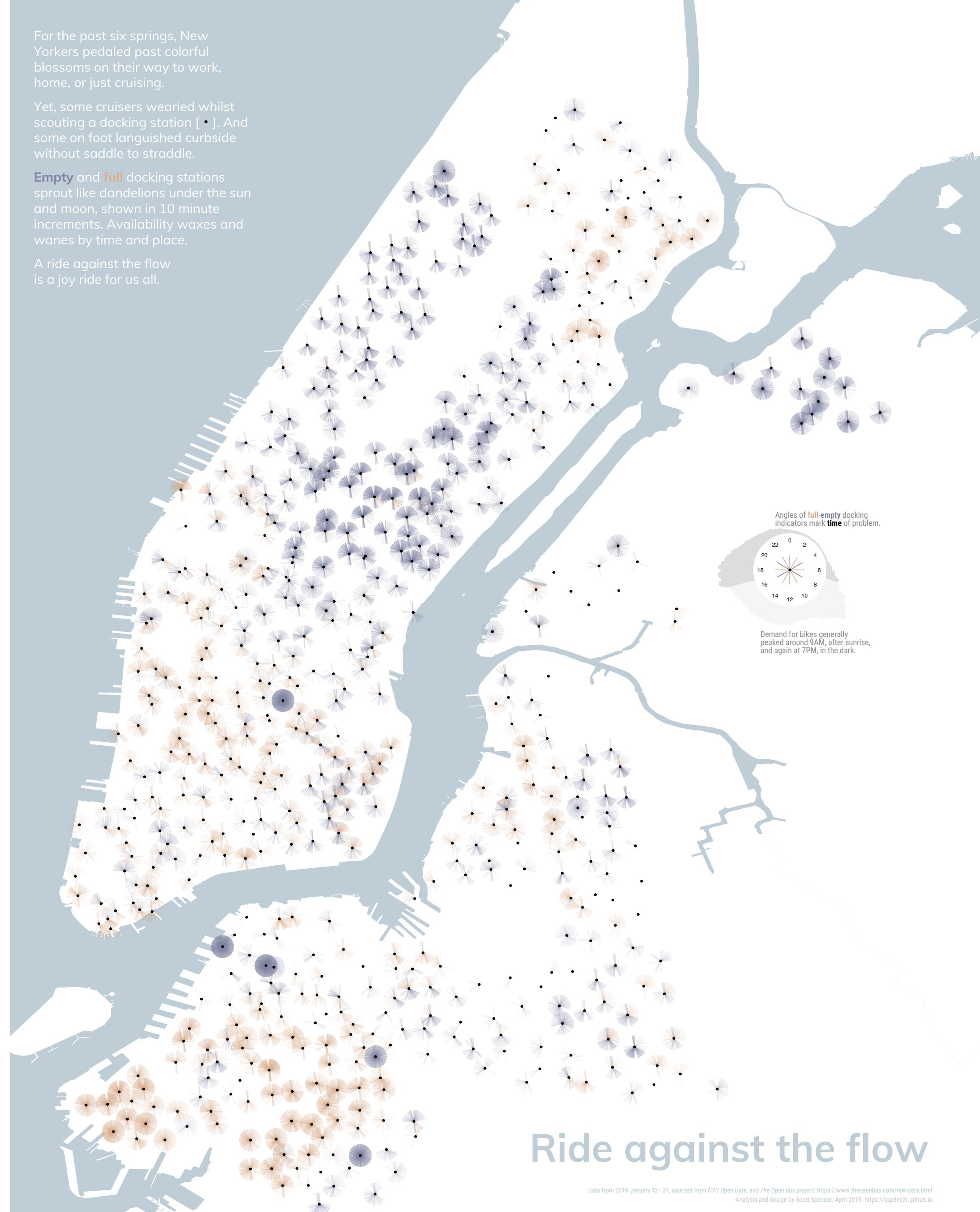
*Spencer*

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

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

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

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



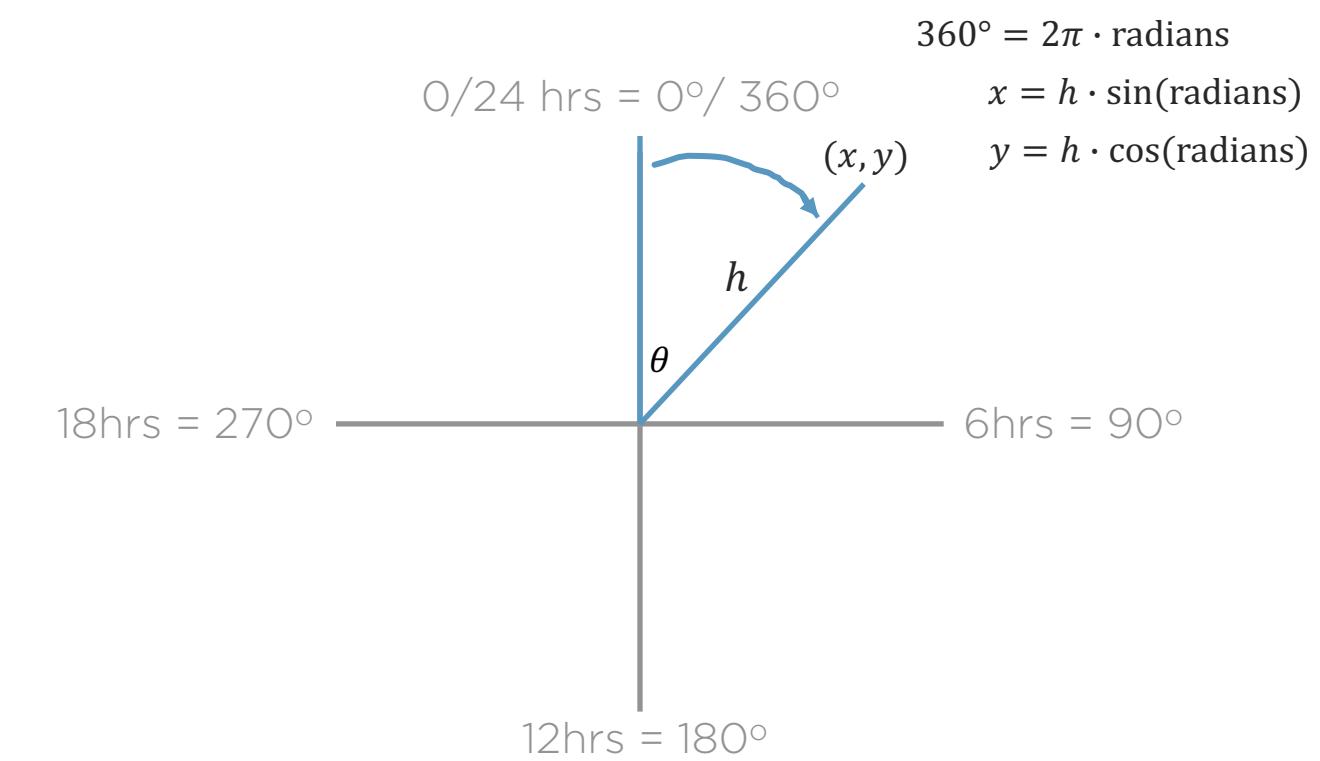
For the past six springs, New Yorkers pedaled past colorful blossoms on their way to work, home, or just cruising.

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

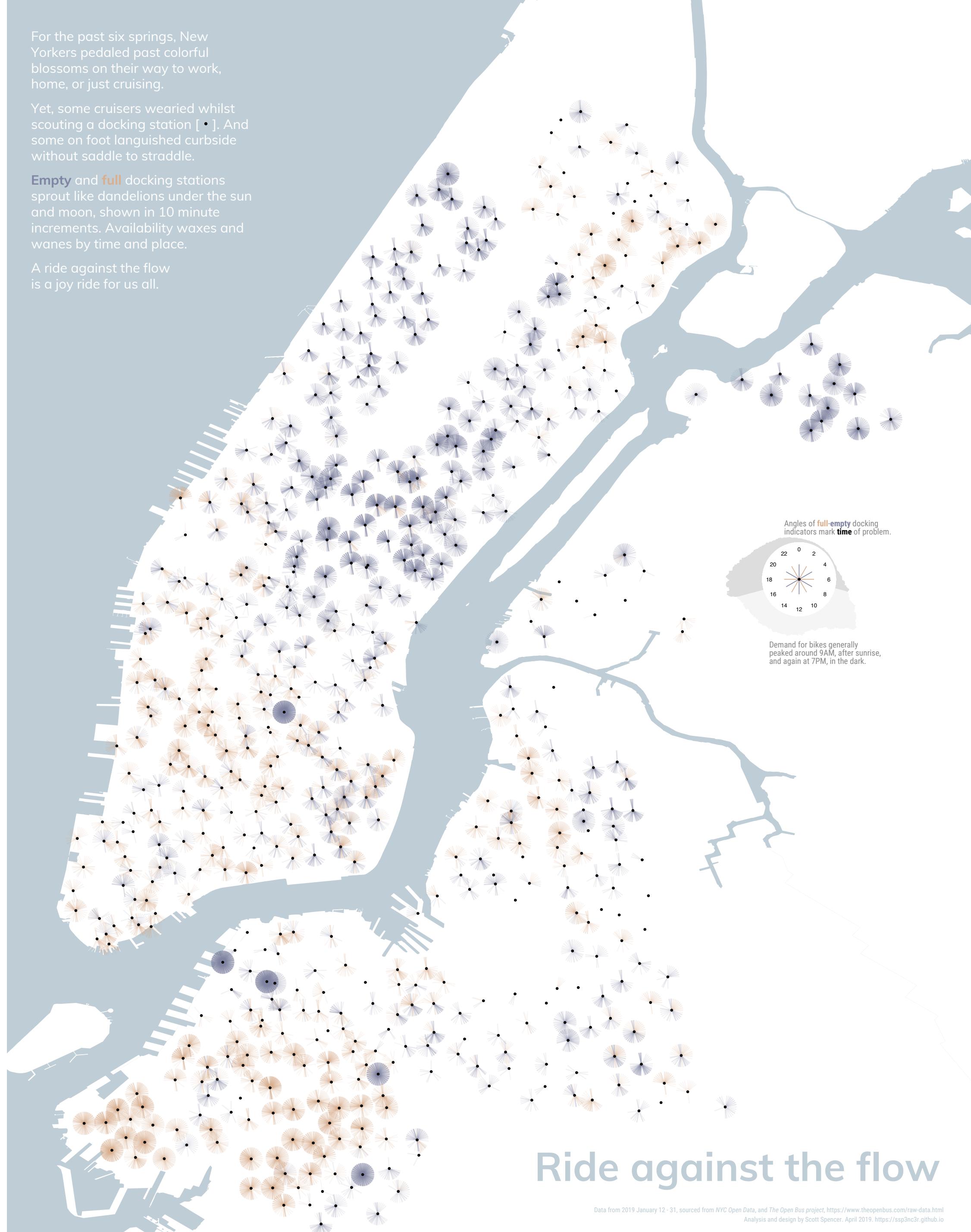
Empty and full docking stations sprout like dandelions under the sun and moon, shown in 10 minute increments. Availability waxes and wanes by time and place.

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

### Maths to create data encodings at each dock station



**Opacity is the lack of transparency.** Opacity was lowered so that a single unavailable bike or full station would not be very noticeable but several markings at that time (on different days) would increase opacity, making the marking brighter and thus increase its visual importance.



Basic math can help when making custom graphics. If you'd like a refresher on basic algebra, geometry, trigonometry: [Simmons, George F. Precalculus Mathematics in a Nutshell](#). Barnes & Noble Books, 1987. Print.

# Let's look ahead

# For Next Week, Module 10:

## Agenda next week

### The minimum

Next deliverable, *draft* information graphic

More on information graphics

Communicating uncertainty

Kay, Matthew et al. *When (Ish) Is My Bus? User-Centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems*. New York, New York, USA: ACM Press, 2016. 5092–5103. Web.

Consider how test subjects responded to varying visual representations of uncertainty, and any implications for your projects.

Fischhoff, Baruch. *Communicating Uncertainty: Fulfilling the Duty to Inform*. Issues in Science and Technology 28.4 (2012): 63–70. Print.

Consider the author's view on issues with communicating uncertainty, and his suggestions for addressing these issues.

Wainer, Howard. *The Most Dangerous Equation*, in Chapter 1, *Picturing the Uncertain World*. Princeton University Press, 2009. Print.

Consider the issues he raises of when variation is misunderstood, and the exemplary implications.

Wacharamanotham, Chat et al. “Special Interest Group on Transparent Statistics Guidelines..” CHI Extended Abstracts (2018): n. pag. Print and web. <http://transparentstatistics.org>

# Feedback

## Looking back?

Of the material covered so far, what material, if any, would you like further review?

## Group work

Next week we will work as groups on your information graphic. How would you like feedback from your colleagues structured to best help you? What questions should they help give you perspective?

**See you  
next week!**

