UVA DS2003 Communicating with Data

Fall 2026

Infographics

Telling a story with a montage of figures

Sarah Groves

based on Claus O. Wilke, Fundamentals of Data Visualization, ch. 29: “Telling a story and making a point”

Part 1

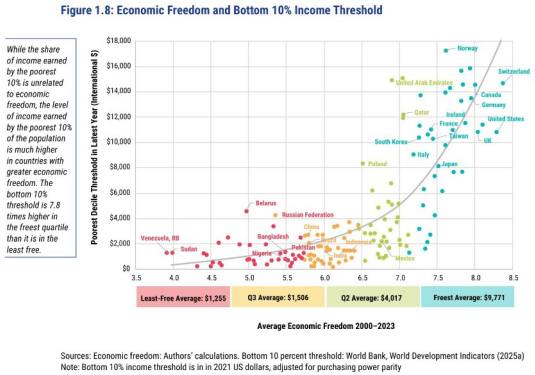
What is an infographic?

From annotated charts to free-form compositions

Infographics span a spectrum of compositions

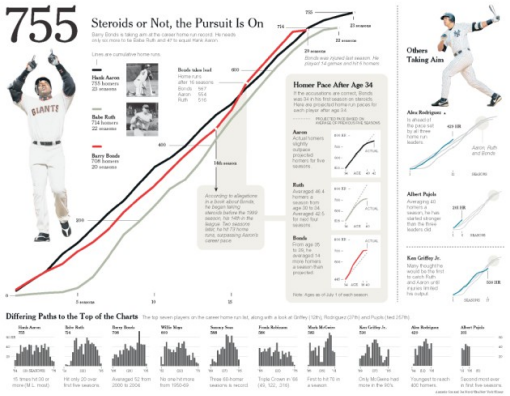
“A clear and visual representation of data and information.”

RESTRICTED



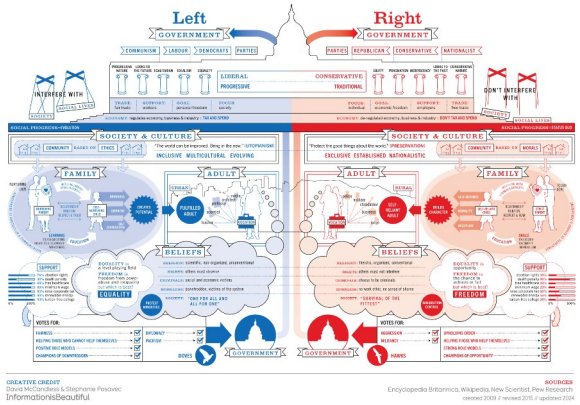
Elementary plots, labeled and annotated

BOTH



Plots combined with free-form elements

FREE-FORM

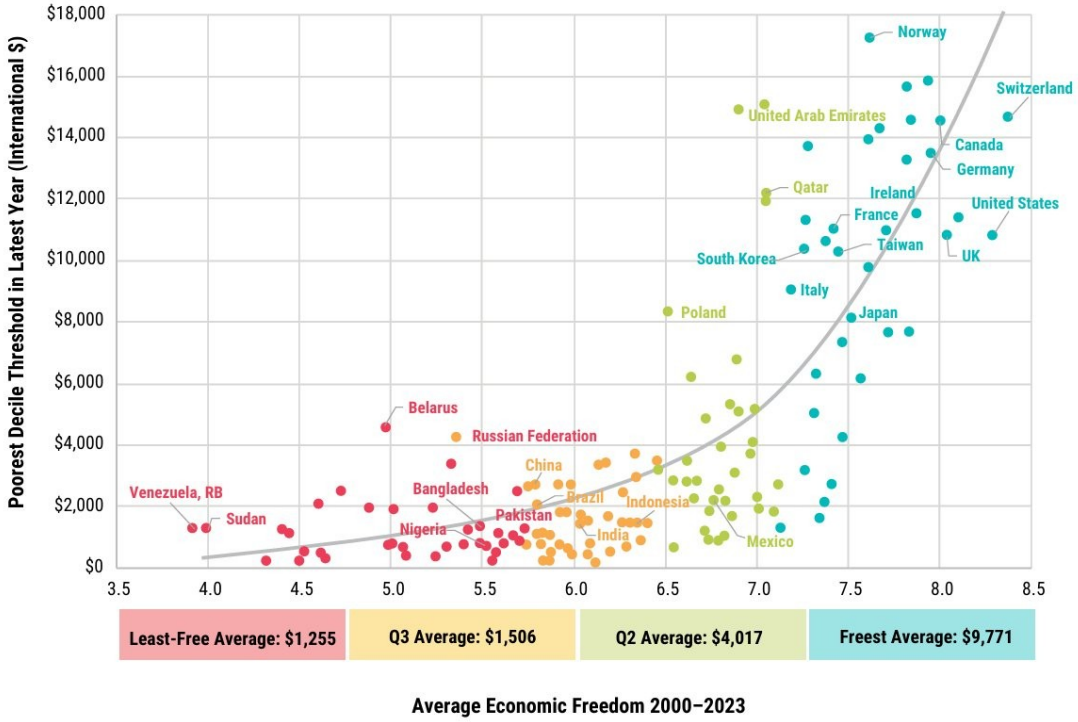


Free interpretations of data and information

A scatterplot, labeled to make a point

Figure 1.8: Economic Freedom and Bottom 10% Income Threshold

While the share of income earned by the poorest 10% is unrelated to economic freedom, the level of income earned by the poorest 10% of the population is much higher in countries with greater economic freedom. The bottom 10% threshold is 7.8 times higher in the freest quartile than it is in the least free.



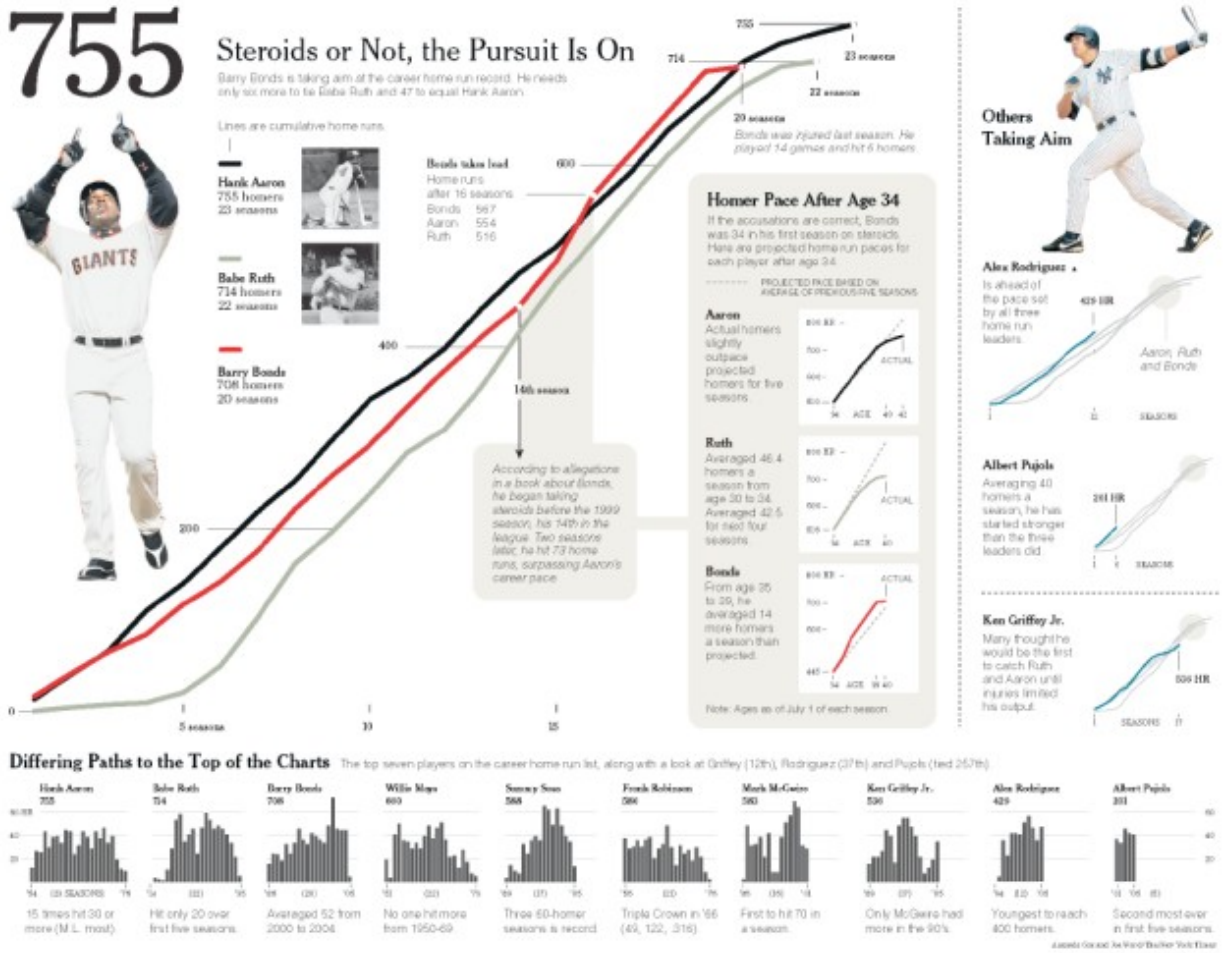
Sources: Economic freedom: Authors' calculations. Bottom 10 percent threshold: World Bank, World Development Indicators (2025a)
Note: Bottom 10% income threshold is in in 2021 US dollars, adjusted for purchasing power parity

RESTRICTED

A simple scatterplot is labeled and annotated to make a point.

The annotation does the work: the reader knows what to look for before reading the axes.

Plots and free-form elements together



BOTH

Combines plots with annotations in a visually elegant and compelling way.

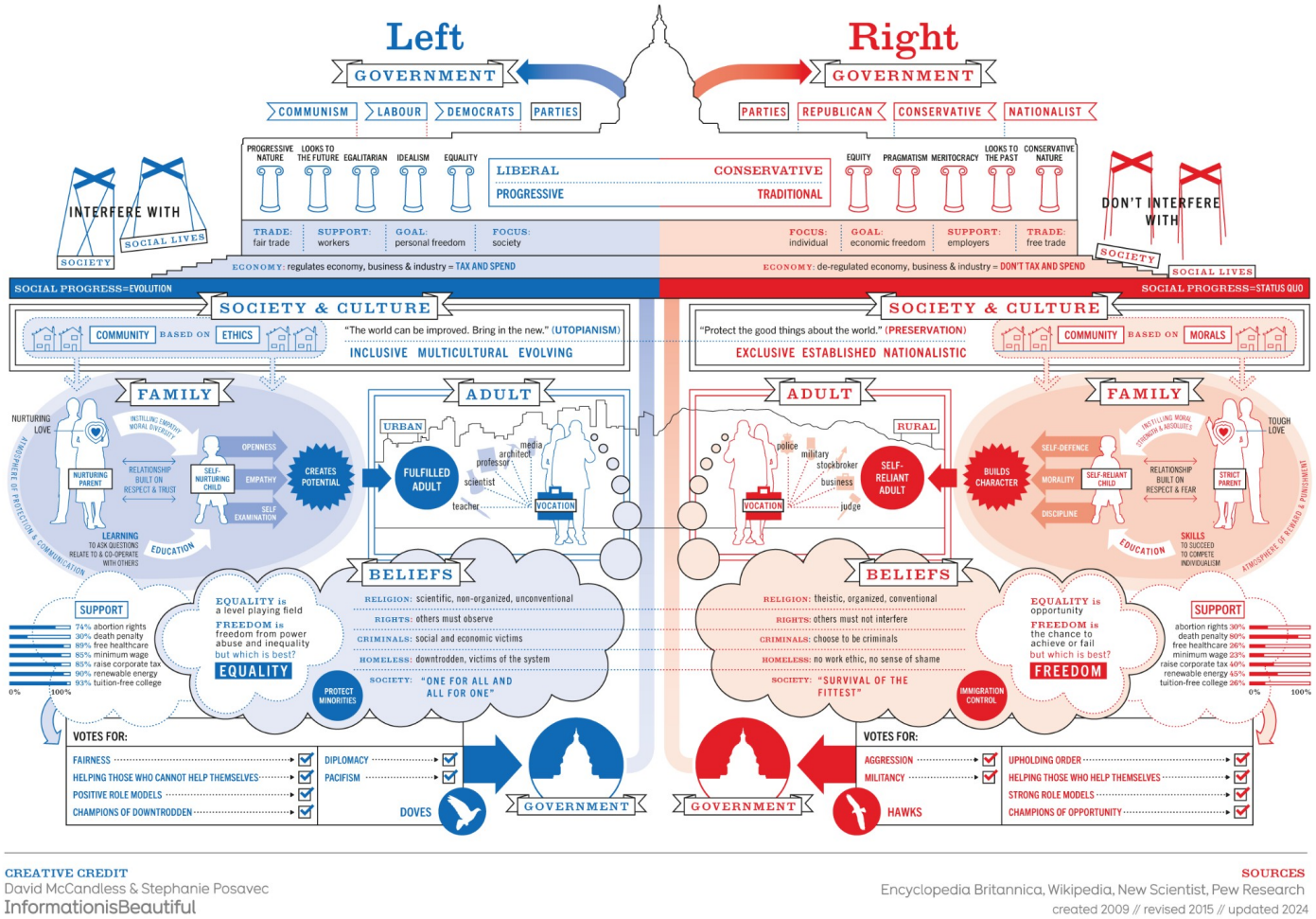
755 is Hank Aaron's career home-run record; the graphic tracks the players chasing it.

Two of Wilke's rules at work:

- **Raw data before derived quantities:** cumulative counts come first, the projected pace after.
- **Build up to complex figures:** one big chart first, then the small multiples.

New York Times, "Steroids or Not, the Pursuit Is On." Discussed in Segel & Heer, Narrative Visualization (2010), idl.uw.edu/papers/narrative

A montage that uses metaphor to organize



CREATIVE CREDIT
David McCandless & Stephanie Posavec
InformationisBeautiful

SOURCES
Encyclopedia Britannica, Wikipedia, New Scientist, Pew Research
created 2009 // revised 2015 // updated 2024

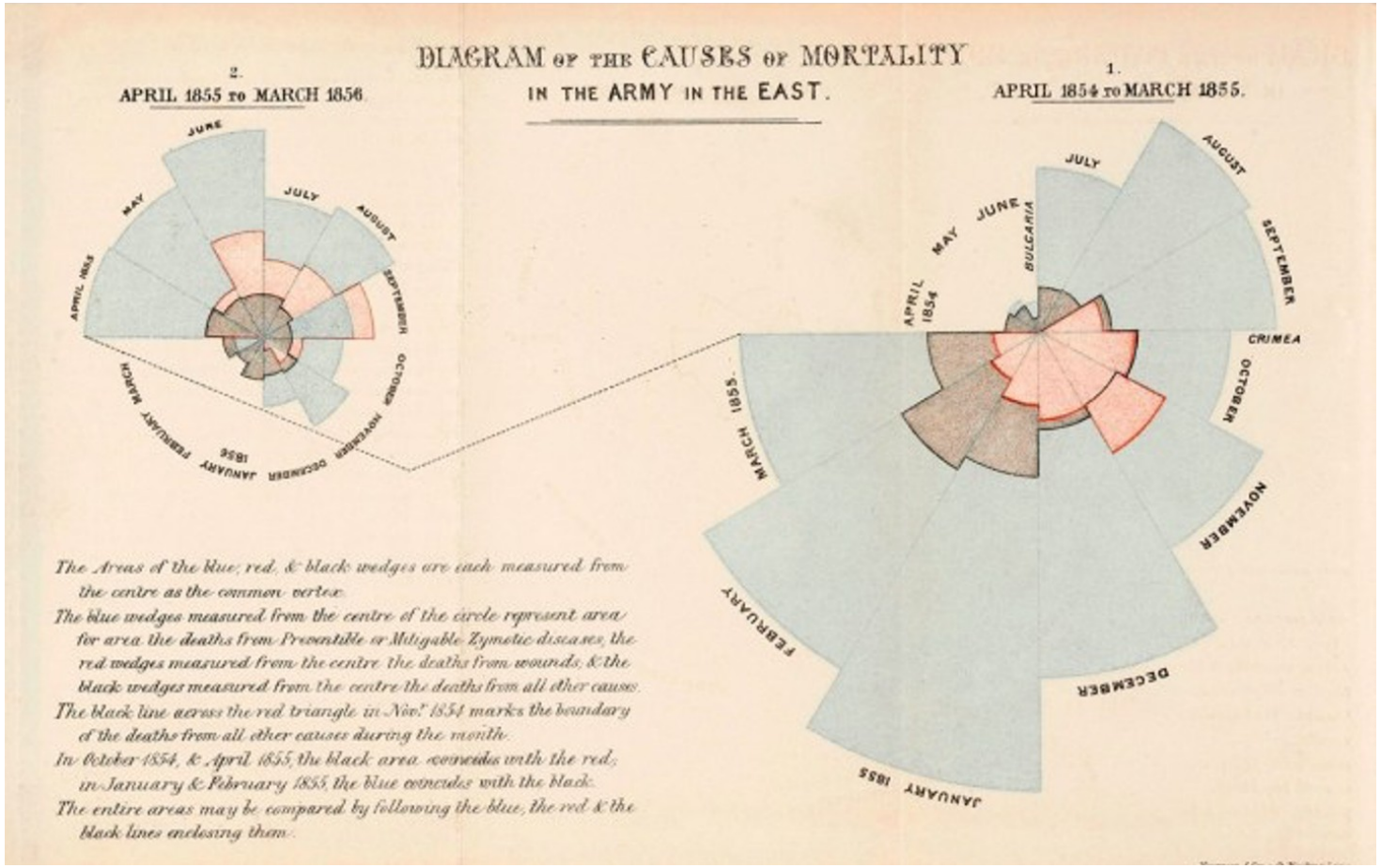
FREE-FORM

A montage of information and some data in one place.

It structures the information with metaphors of color and handedness: left and right.

David McCandless, “Left vs. Right US,” Information is Beautiful. informationisbeautiful.net/visualizations/left-vs-right-us

Infographics are not new



1858

Florence Nightingale's diagram of the causes of mortality in the army in the East.

She used polar area charts to convince the British Army to practice better hygiene: a data story with a call to action.

Florence Nightingale, "Diagram of the causes of mortality in the army in the East," 1858

Any visualization is a model.

We humans use models for perception, cognition, and reasoning because our limited brains are incapable of grasping reality in all its glorious complexity.

-Albert Cairo

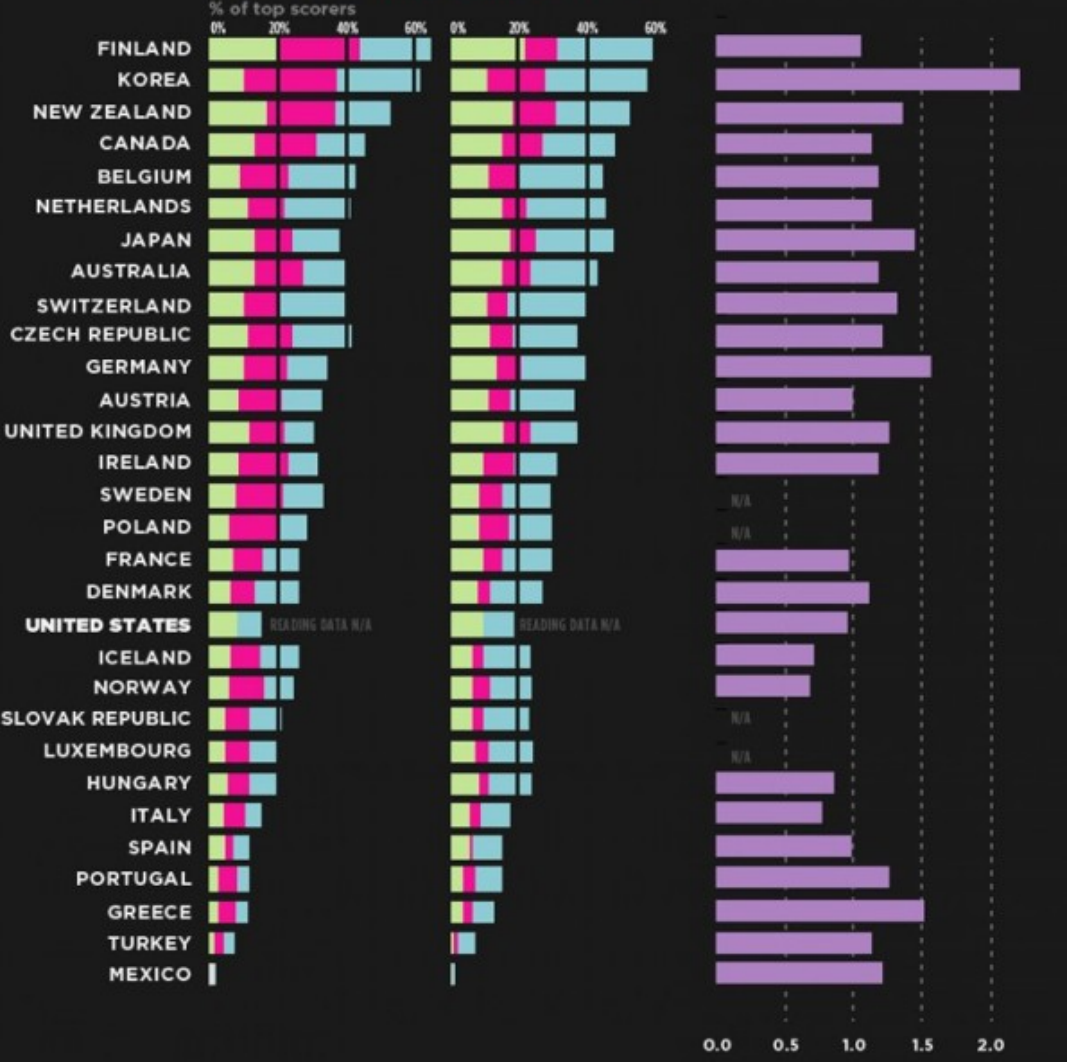
You must decide how to model the data so that the montage of your infographic or science visualization is *clear, accurate, functional,* and *beautiful.*

Every visualization is an argument for something. That something should be truth from the data.

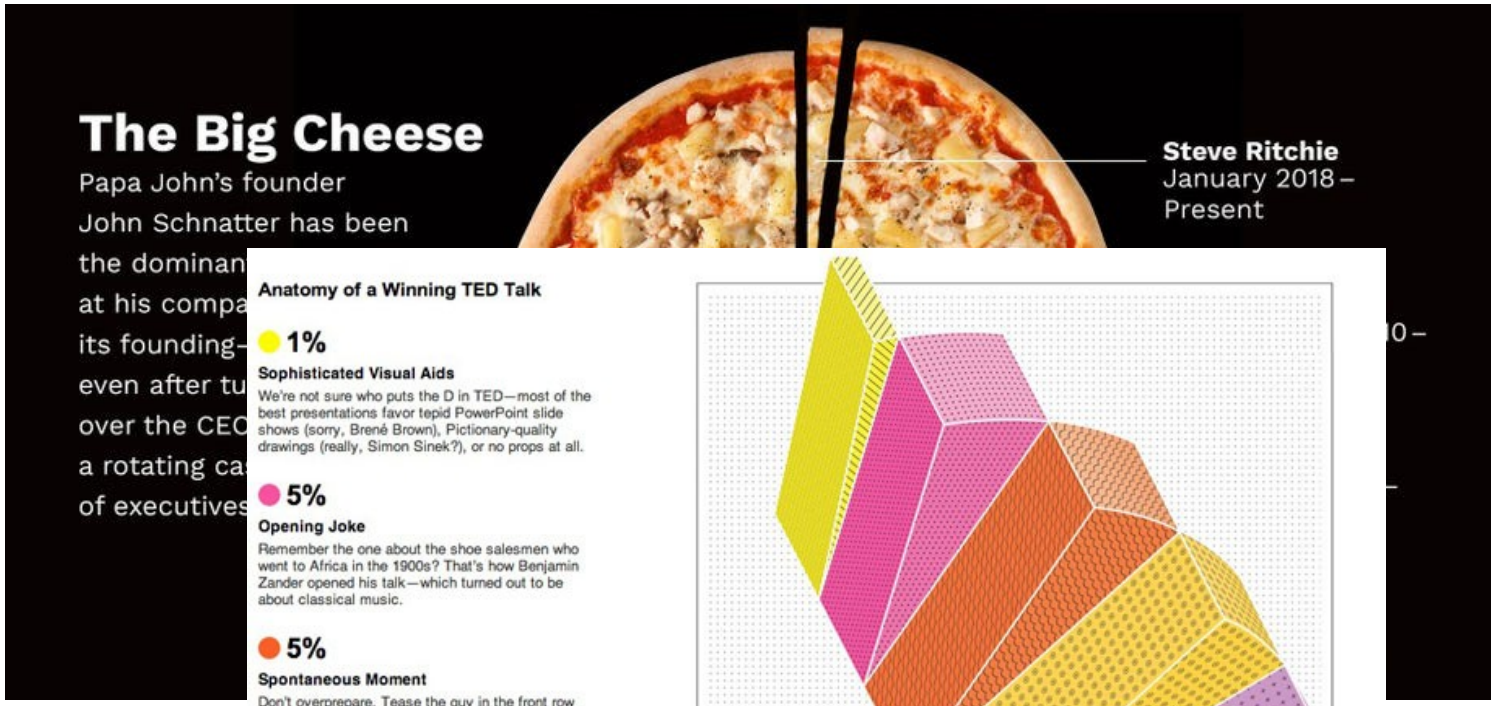
Three “mind bugs” that prevent us from interpreting data correctly

1. I detect interesting patterns, regardless of whether or not they are real. I’ll call this the **patternicity** bug.
2. I immediately come up with a coherent explanation for those patterns. This is the **storytelling** bug.
3. I start seeing all further information I receive, even the one that conflicts with my explanation, in a way that confirms it. I refuse to give my explanation up, no matter what. This is the **confirmation** bug.

DO BETTER PAID TEACHERS MAKE SMARTER STUDENTS?



Source:



Anatomy of a Winning TED Talk

1%
Sophisticated Visual Aids
We're not sure who puts the D in TED—most of the best presentations favor tepid PowerPoint slide shows (sorry, Brené Brown), Pictionary-quality drawings (really, Simon Sinek?), or no props at all.

5%
Opening Joke
Remember the one about the shoe salesmen who went to Africa in the 1900s? That's how Benjamin Zander opened his talk—which turned out to be about classical music.

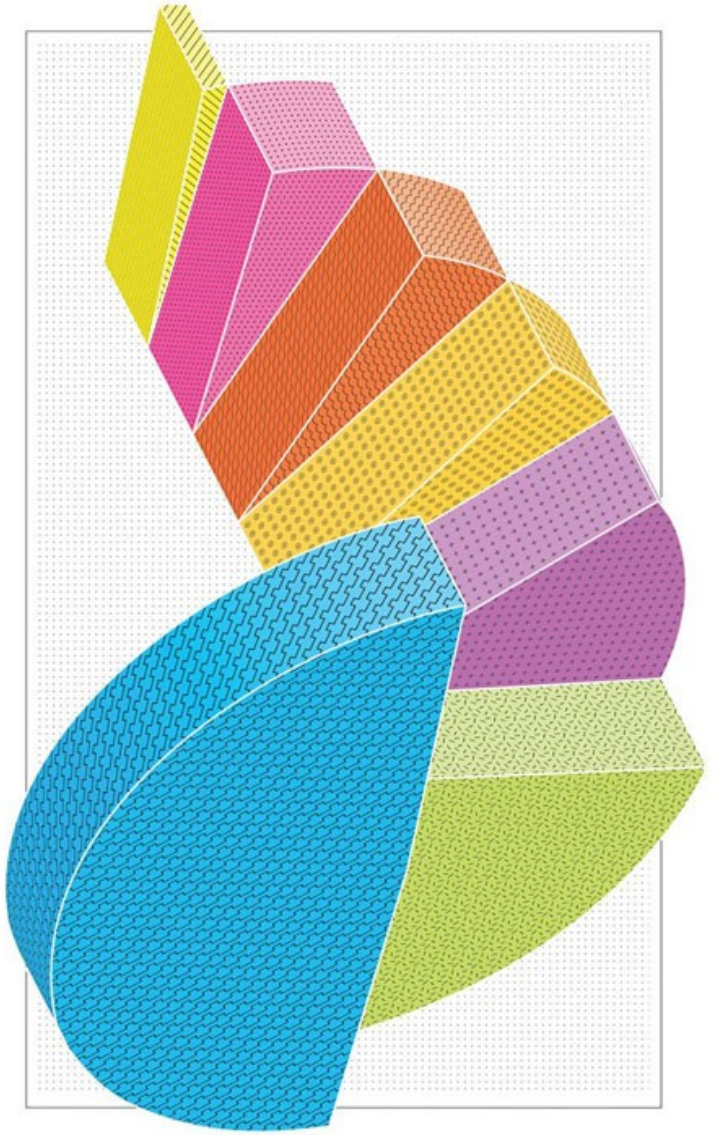
5%
Spontaneous Moment
Don't overprepare. Tease the guy in the front row ("You could light up a village with this guy's eyes"). Commend the stagehand who handles the human brain you brought.

5%
Statement of Utter Certainty
People come for answers—give 'em what they want, as Shawn Achor did: "By training your brain … we can reverse the formula for happiness and success."

12%
Snappy Refrain
The TED equivalent of "I have a dream." Example: "People don't buy what you do; they buy why you do it." Repeat 7x.

23%
Personal Failure
Be relatable.We want to know about that nervous breakdown. Or at least the time you didn't fit in at summer camp.

49%
Contrarian Thesis
Wait a sec—we should be playing *more* videogames? The more choices we have, the worse off we are? TED is where conventional wisdom goes to die.



Evaluating the UK search marketing landscape: Exposing SEO CTRs by industry and who clicks on PPC



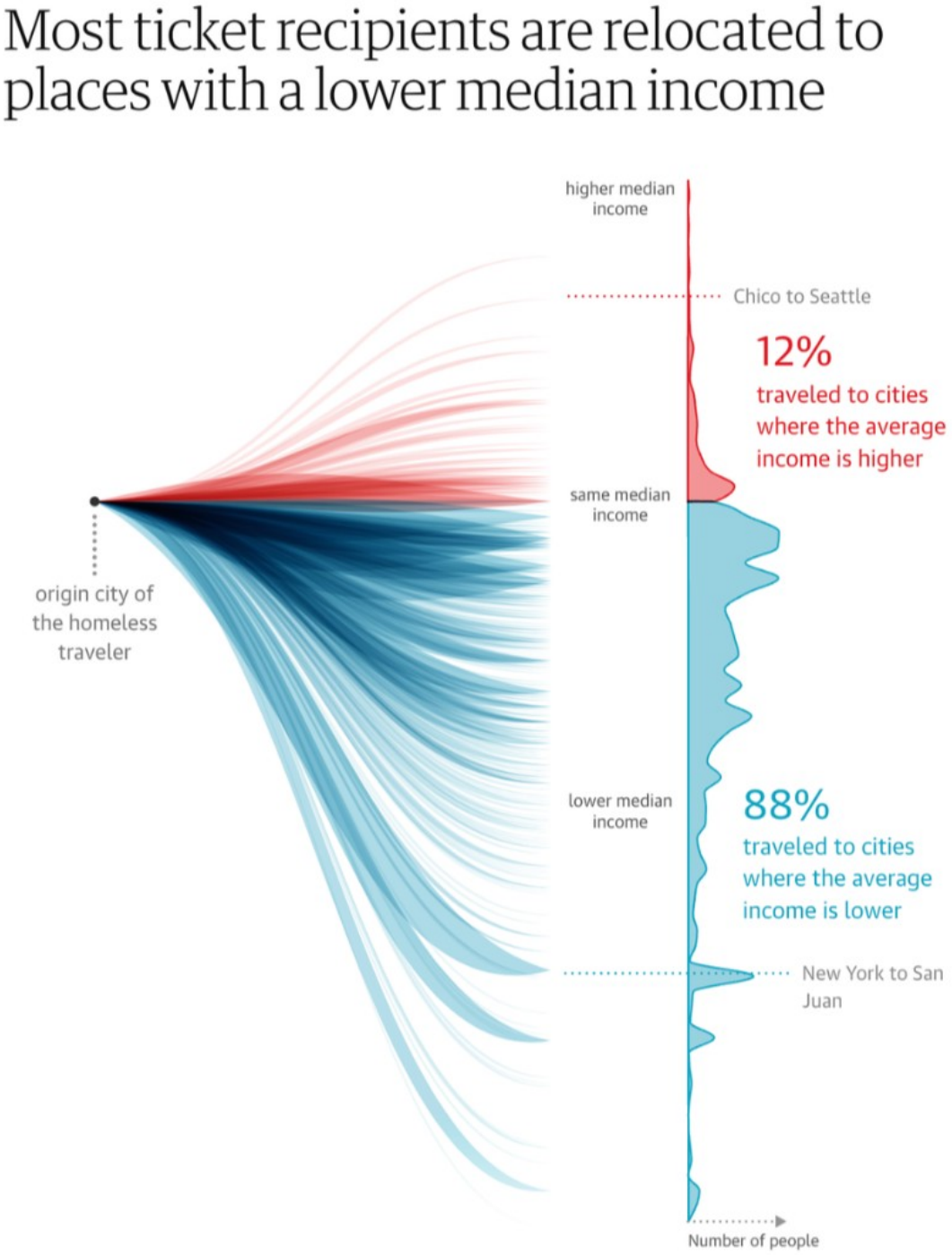
This is the first time that this information has been made available to the UK market.
The data covers all of June 2011.

This data was provided by Nielsen for GroupM.

28 million people searched during this time **with 1.4bn** search queries.

1.4bn search queries **X** average length of 2.84 words **=** 3.98bn Words

90 Encyclopaedia Britannicas or **WV1.6** English Language Wikipedias

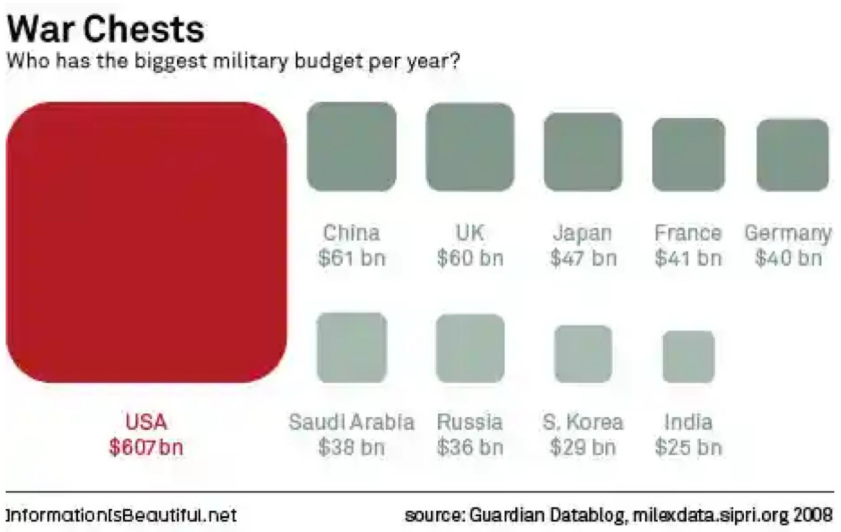


BOTH

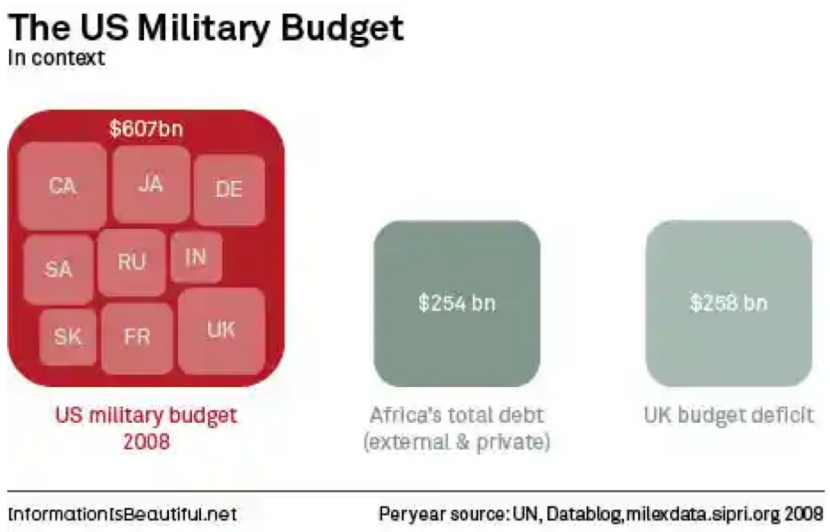
Where does your eye go first, and what in the design sends it there?

From Alvarado's slides (source not listed there; likely The Guardian, "Bussed out," 2017)

RAW DOLLARS



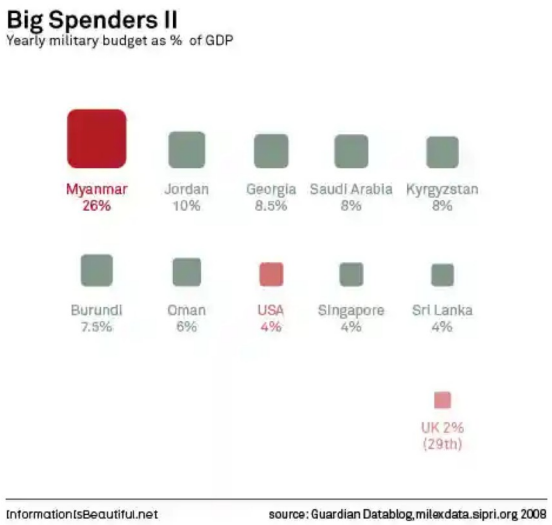
IN CONTEXT



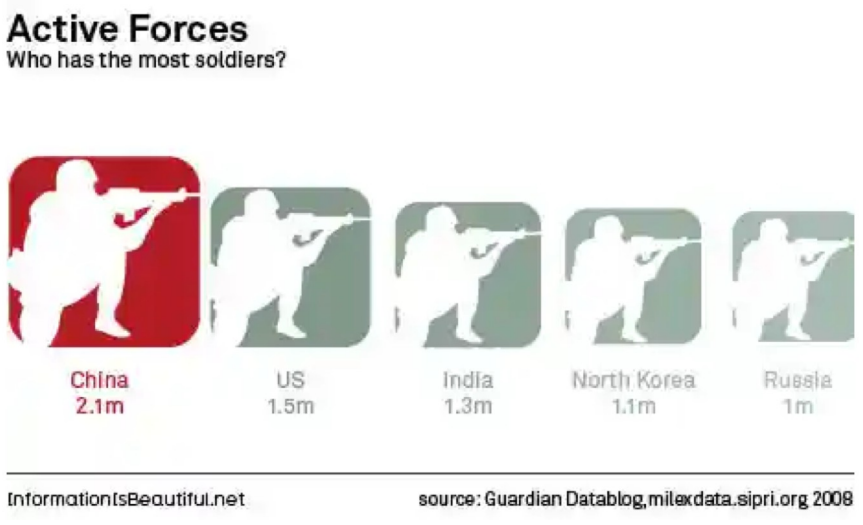
REFRAMING

Which framing would you use to argue the US spends too much? Too little?

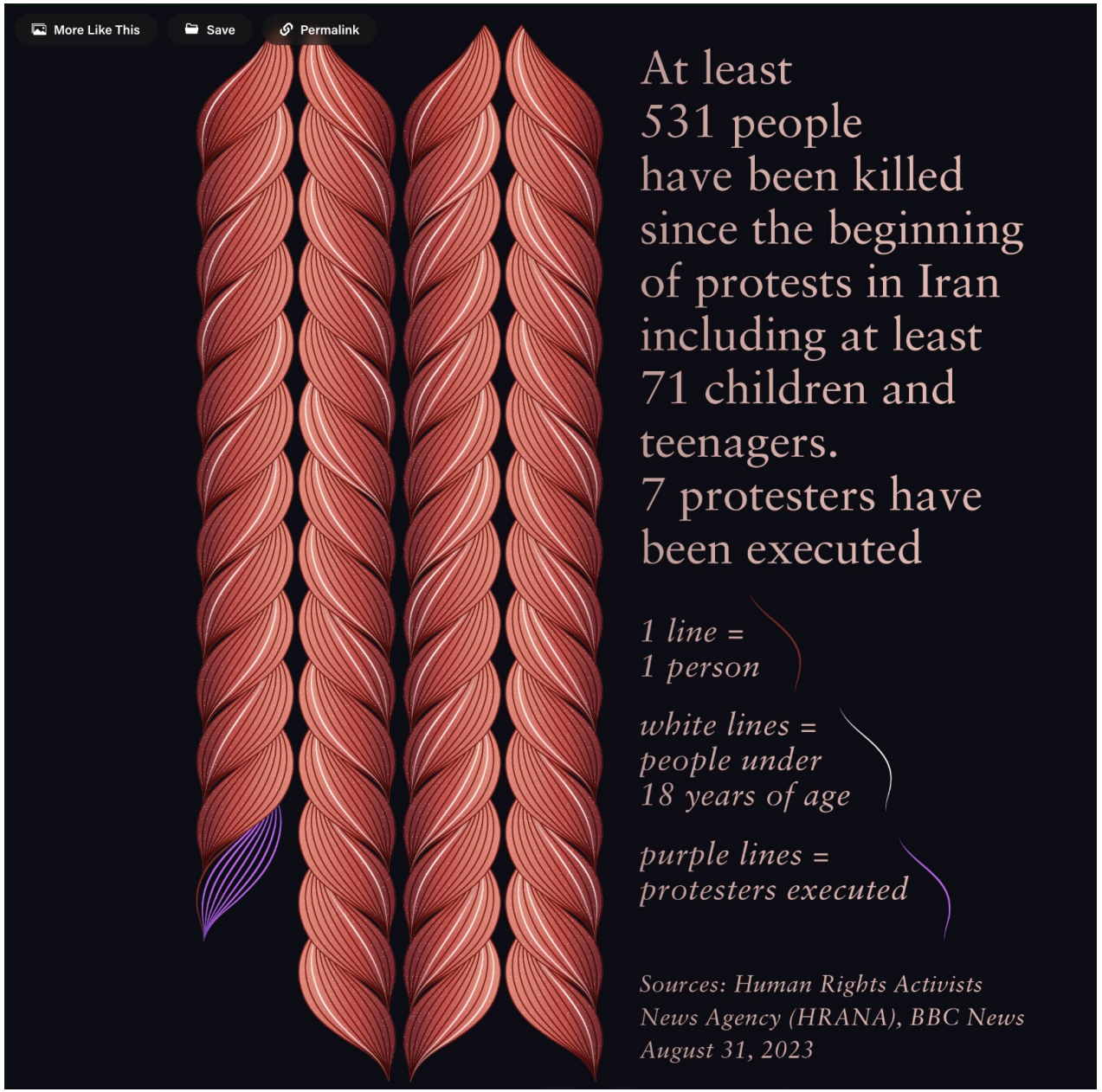
SHARE OF GDP



TOTAL SOLDIERS



David McCandless, Information is Beautiful (data: Guardian Datablog, milexdata.sipri.org, 2008)



FREE-FORM

Would a bar chart of the same numbers hit as hard? Why or why not?

Federica Fragapane, “Iran protests,” Behance (data: HRANA, BBC News, as of Aug. 31, 2023)

Think | Pair | Share

Think Of the infographics we just saw, which one will you still remember tomorrow? Which one would you be able to explain? Which one did you gain insight from?

Pair Was it restricted, free-form, or both? What made it stick?

Share Memorable because of the data, the picture, or the story?

Part 2

Let’s talk more about stories

Tension, resolution, and three standard formats

Without a story, the audience makes one up

Any figure you show creates a story in the viewer's head. Best case, it is close to yours. but often it is one of these:

"This is boring."

"The author is wrong."

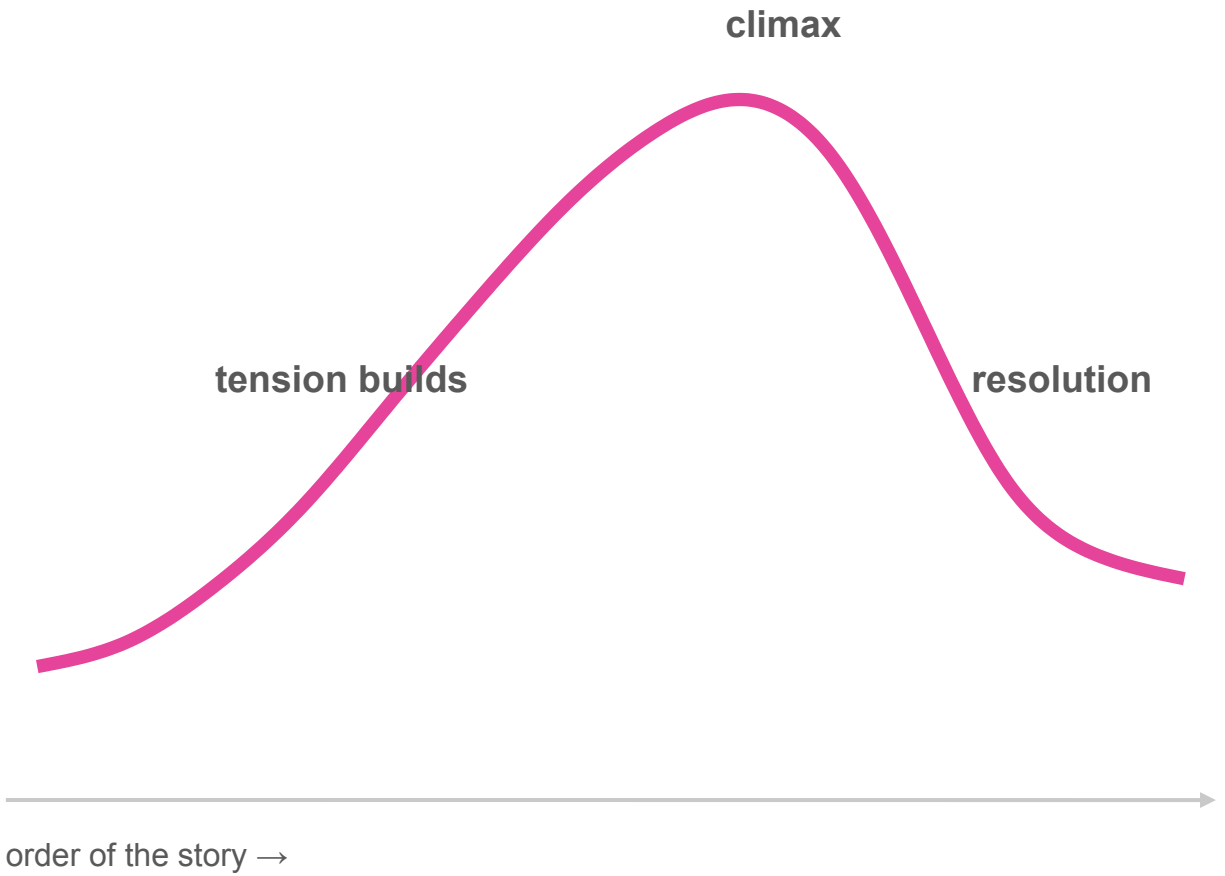
"The author doesn't know what they're doing."

Wilke's goal was to use facts and logical reasoning to get the audience interested and excited– you should have the same goal!

A story builds tension, then resolves it

Observations, facts, or events, presented in a specific order to create an emotional reaction.

The flow from tension to resolution is the **story arc**. Every good story has a clear one.



Opening → Challenge → Action → Resolution

Wilke's example: the physicist Stephen Hawking. Every sentence is true, and it is still a compelling story.

OPENING

Meet Stephen Hawking, a young theoretical physicist starting his PhD.

CHALLENGE

At 21 he is diagnosed with motor neuron disease and given two years to live.

ACTION

He refuses to accept it and pours all his energy into science.

RESOLUTION

He lives to 76 and becomes one of the most influential physicists of his time.

Today's question: what if each of these boxes were a figure?

Three standard storytelling formats

Format	How it moves	Good for
Opening–Challenge–Action–Resolution	Builds tension slowly, resolves at the end	Presentations, interactive notebooks, infographics
Lead–Development (Barry Bonds story)	Gives the main point first, then the details	News, social media, executive summaries
Action–Background–Development–Climax–Ending	Starts in the middle of the action; holds back the ending	Pulling in an audience early

Hawking as a lead: “Stephen Hawking outlived his doctors’ prognosis by 53 years and did his most influential work while severely disabled.”

Think | Pair | Share

Think Your favorite movie in one sentence: what's the challenge, and what's the resolution?

Pair Now tell it as a lead: give away the ending first. Better or worse?

Share Which version would you use in your interactive notebook? Which in your infographic/science visualization?

Part 3

Mapping story parts to figures

Each figure plays one role in the arc

One static figure rarely tells a whole story

OPENING	CHALLENGE	ACTION	RESOLUTION
figure?	*figure?*	*figure?*	*figure?*

- A figure can show the opening, the challenge, the action, or the resolution. It rarely shows all four.
- A complete story usually needs several figures in sequence
- One figure can carry a whole arc, but only if it holds the challenge and the resolution at once. That is a lead.

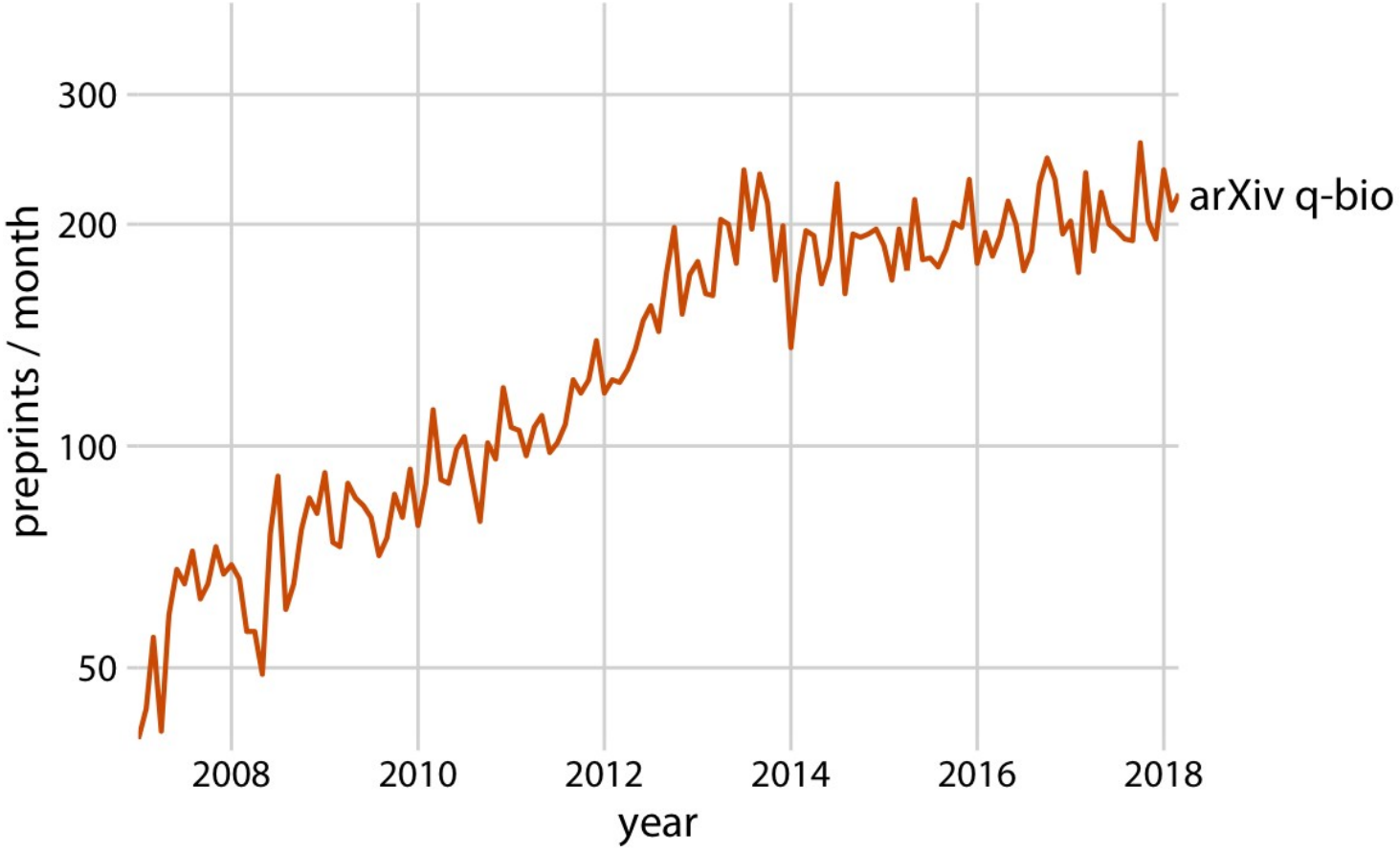
Opening:

Scientists have been **sharing manuscript drafts** for as long as scientific manuscripts have existed

In the early 90s, physicists realized the internet should store manuscript drafts in a **central repository**

They invented the **preprint server**, arXiv, where scientists can upload, download, and search

Challenge: q-bio growth suddenly stalls



CHALLENGE

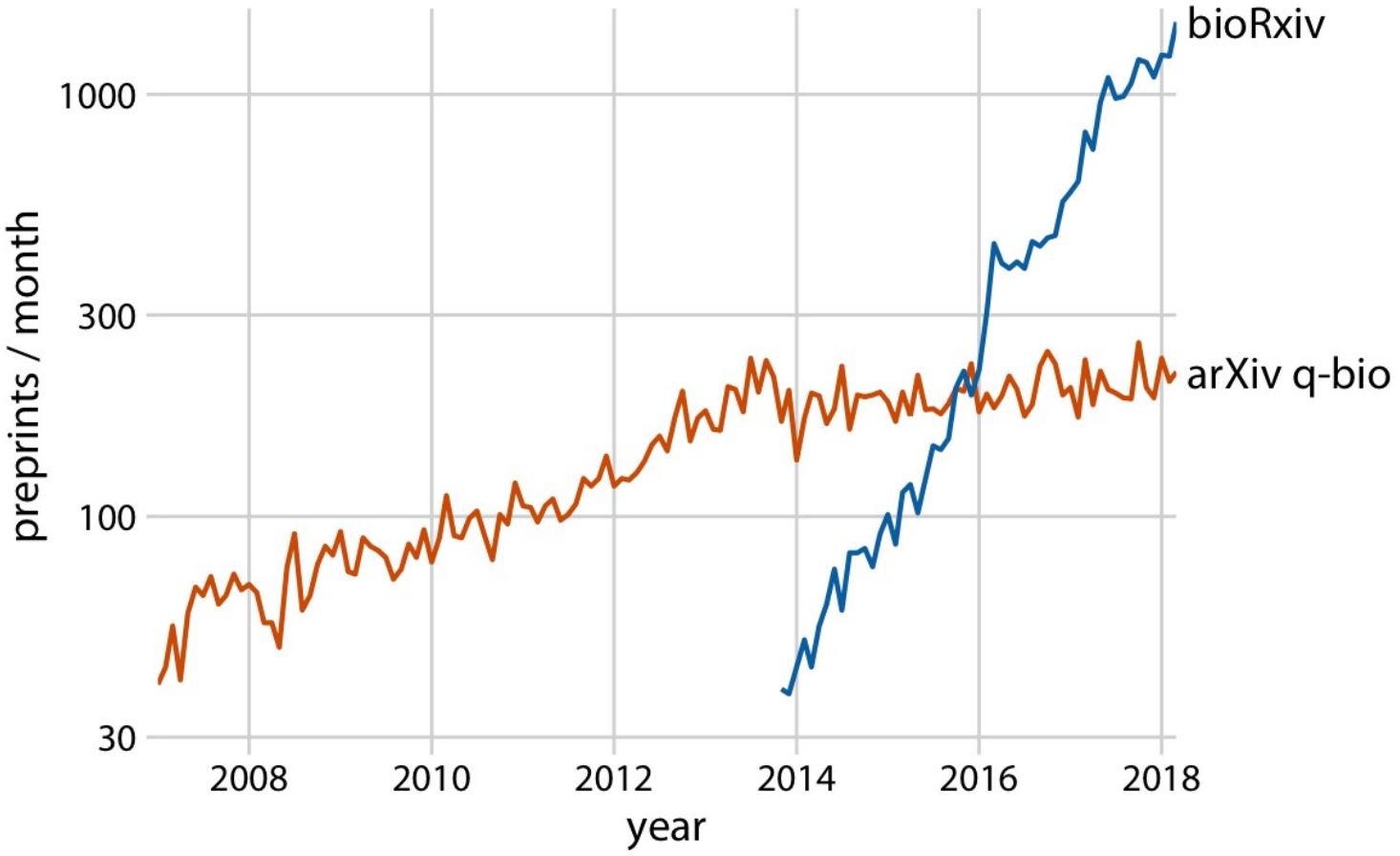
Monthly submissions to arXiv's quantitative biology section grew exponentially from 2007 to late 2013.

Then growth stopped.

What happened in late 2013?

Figure 29.1: Claus O. Wilke, Fundamentals of Data Visualization (O'Reilly, 2019), ch. 29. clauswilke.com/dataviz

Resolution: bioRxiv took over the growth



RESOLUTION

bioRxiv, a biology-only preprint server from a publisher biologists trusted, launched in November 2013.

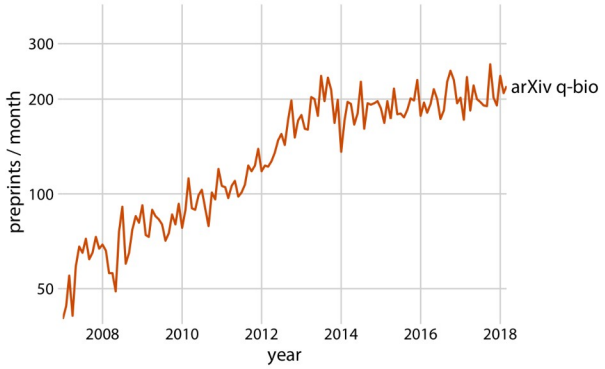
Its submissions grew exponentially from the start, exactly when q-bio flattened.

The new series is the **action**; the matching timing is the resolution.

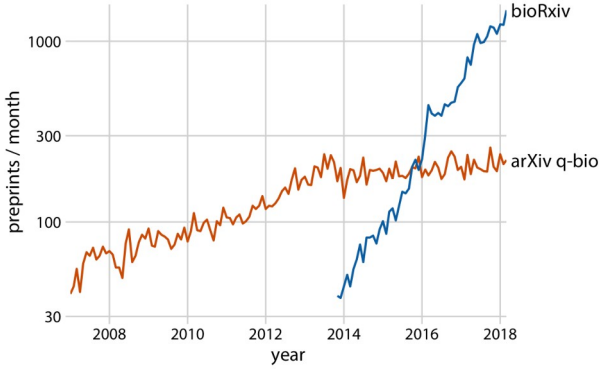
Figure 29.2: Claus O. Wilke, Fundamentals of Data Visualization (O'Reilly, 2019), ch. 29. clauswilke.com/dataviz

Spread information over multiple figures for an infographic

CHALLENGE

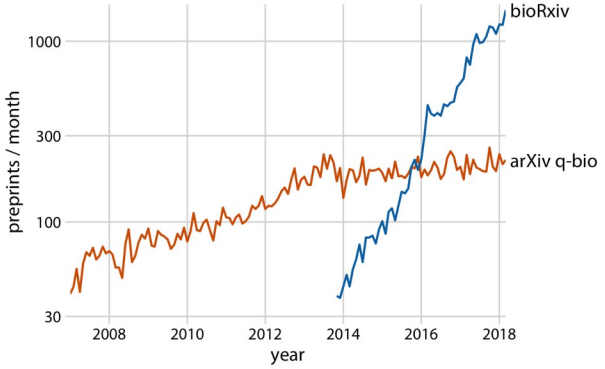


RESOLUTION



In an infographic: two plots means the audience sits with the question before you answer it

LEAD



In a single plot: one figure tells the whole arc

Figure 29.1 and 29.2: Claus O. Wilke, Fundamentals of Data Visualization (O'Reilly, 2019), ch. 29. clauswilke.com/dataviz

What each figure has to do

Story part	Question the figure answers	The figure's job	Often looks like
Opening	What are we looking at?	Set context; show close-to-raw data	Simple time series, map, one bar chart
Challenge	What is surprising?	Make the tension visible: a break, gap, or outlier	Annotated trend break; actual vs. expected
Action	What changed, or what caused it?	Bring in the new variable or event	Added series, before/after, highlight
Resolution	So what?	Show the outcome as one clear comparison	Ranked bars, % change, one big number

You will use this in the activity. Not every story needs all four parts; a lead can be a single resolution figure.

Think | Pair | Share

Think A chart shows a store's sales growing for years, then suddenly dropping. Is that the challenge or the resolution?

Pair What figure would you show next to resolve it?

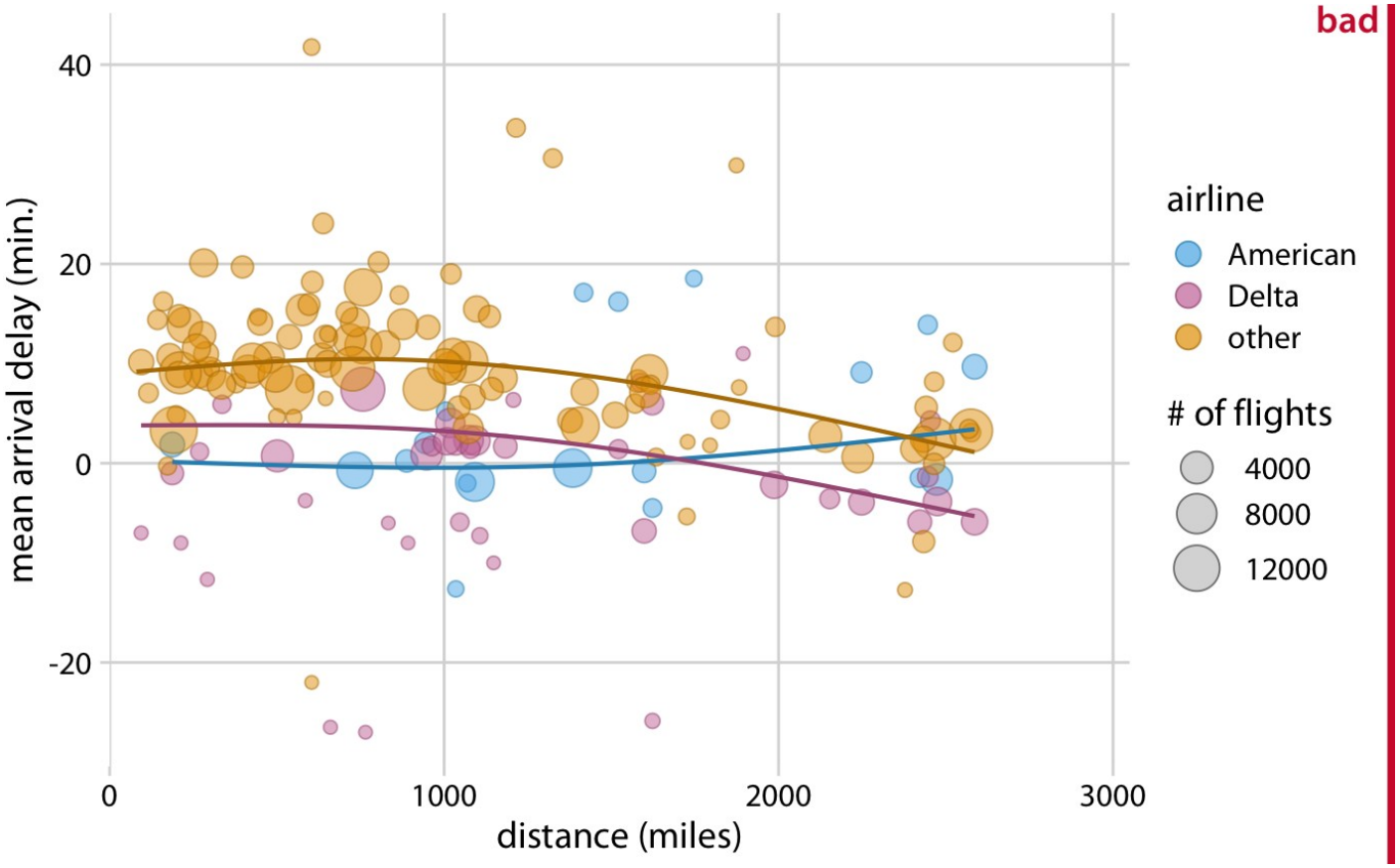
Share Could one figure tell the whole story?

Part 4

Rules for figures in a story

Five habits that keep the audience with you

1. Make a figure for the generals



TOO MUCH

Wilke's Army program managers: a general should see the point of each figure at once.

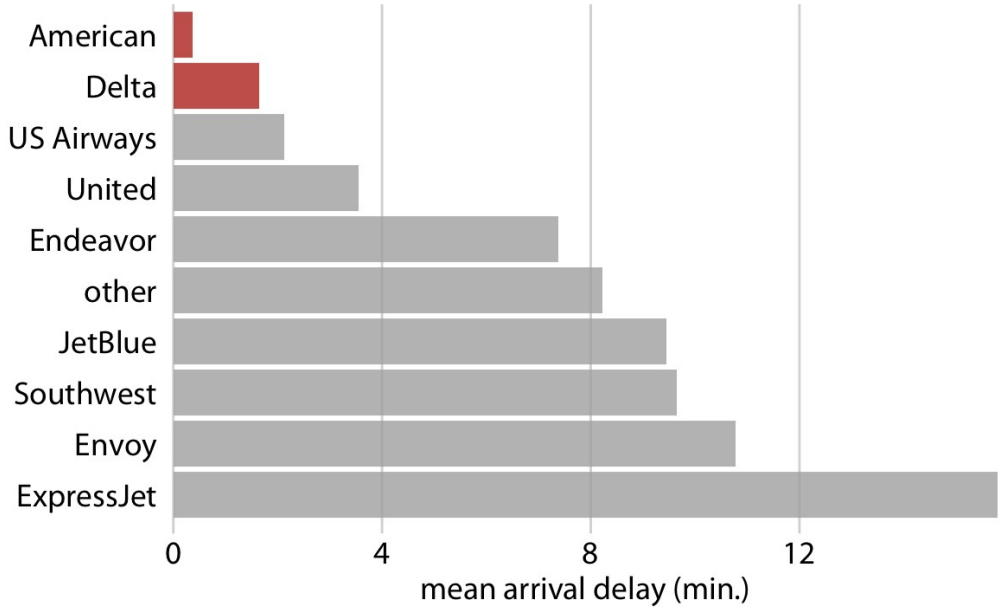
Generals aren't slow. They're busy. So are reviewers, editors, and your boss.

Never assume your audience can rapidly process complex displays.

Figure 29.3: Claus O. Wilke, Fundamentals of Data Visualization (O'Reilly, 2019), ch. 29. clauswilke.com/dataviz Labeled "bad" in the book.

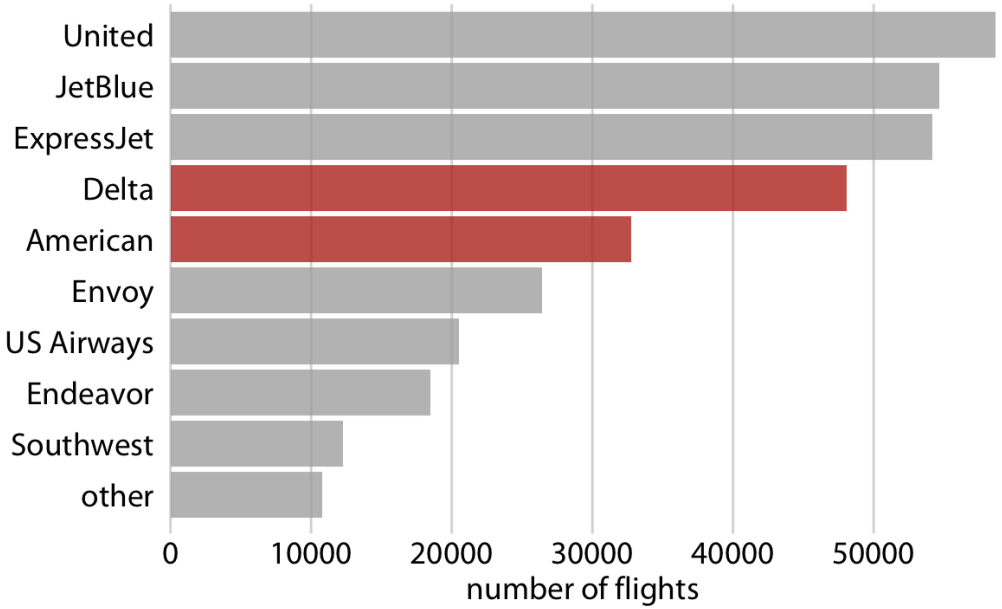
Show the point then answer the next question

THE POINT



Delta and American have the lowest delays.

THE FOLLOW-UP



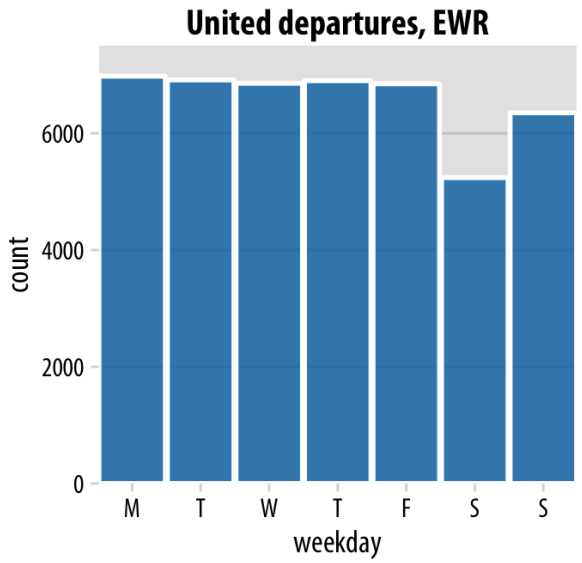
…and they are big NYC carriers, not tiny ones.

Both drop distance. That's fine: don't show dimensions that are tangential to your story.

Figure 29.4 and 29.5: Claus O. Wilke, Fundamentals of Data Visualization (O'Reilly, 2019), ch. 29. clauswilke.com/dataviz

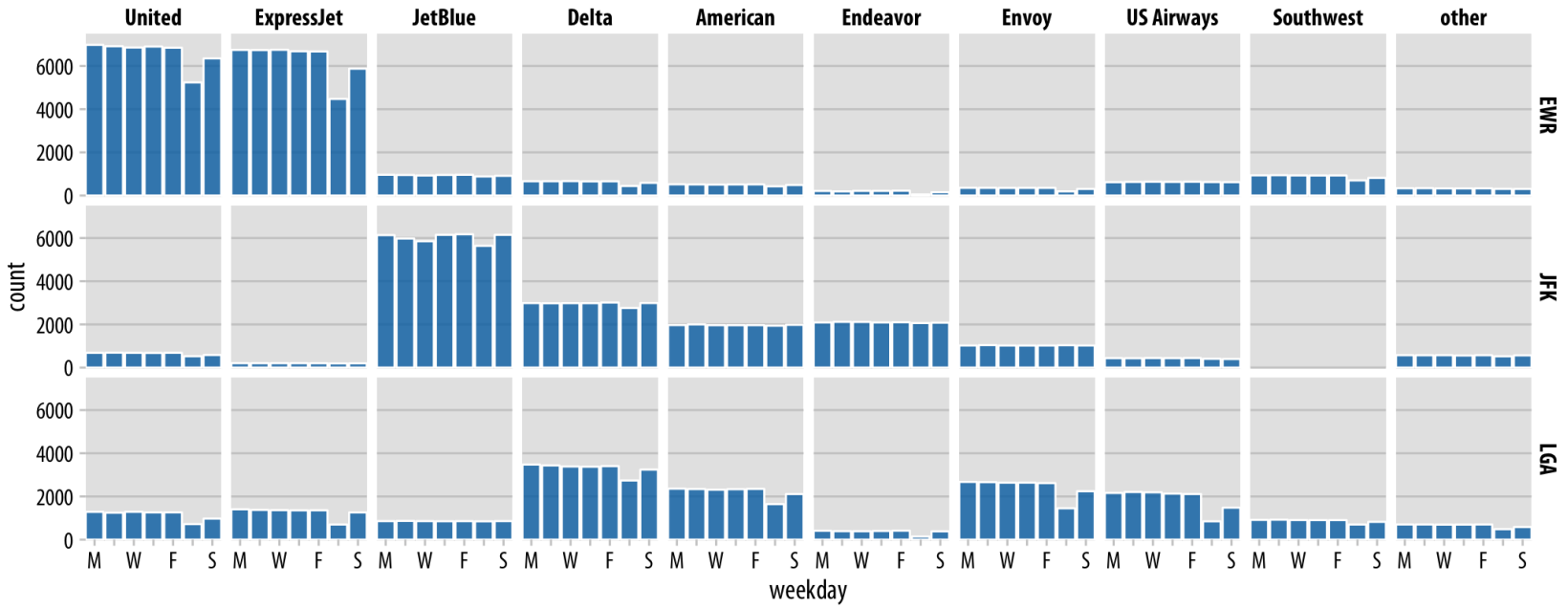
2. Build up to complex figures

FIRST



United departures from Newark, by weekday

THEN



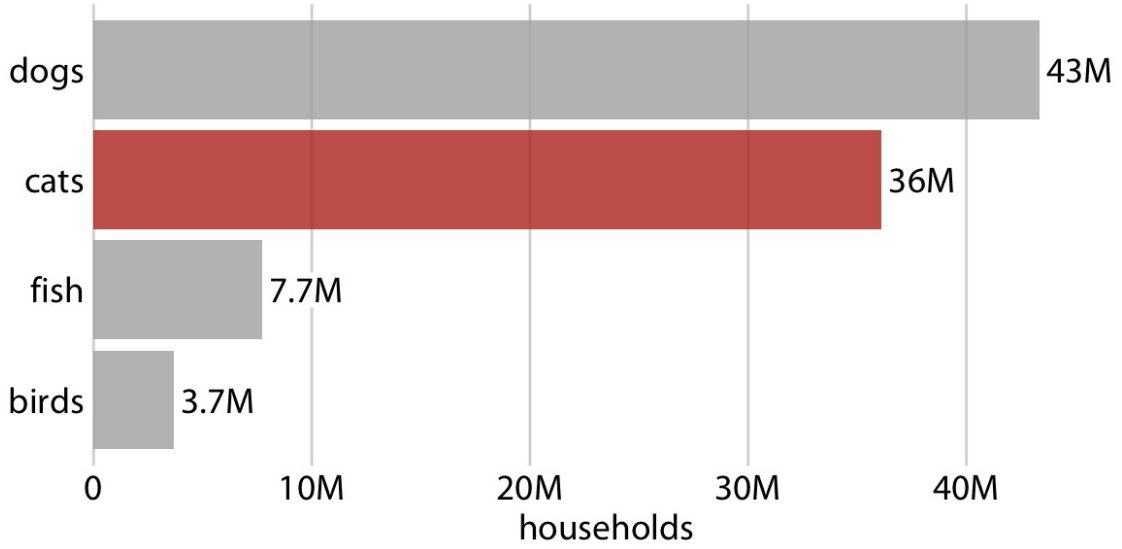
Every airline × airport, same panel design

Never jump straight to a complex figure. Show one panel of a small-multiples grid first; the full grid is then easy to read.

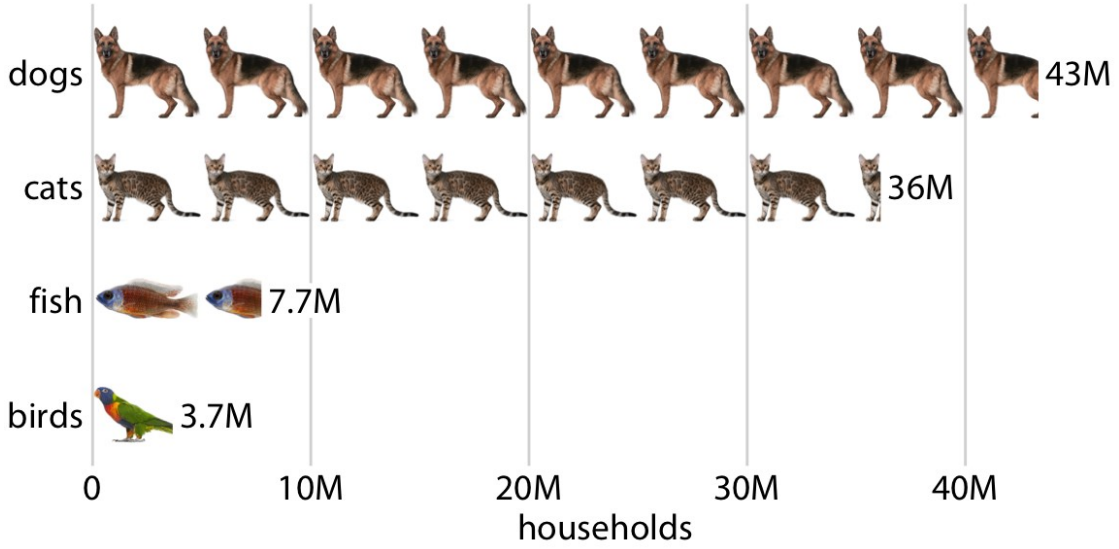
Figure 29.6 and 29.7: Claus O. Wilke, Fundamentals of Data Visualization (O'Reilly, 2019), ch. 29. clauswilke.com/dataviz

3. Make figures clear and memorable

CLEAR BUT GENERIC



CLEAR AND MEMORABLE

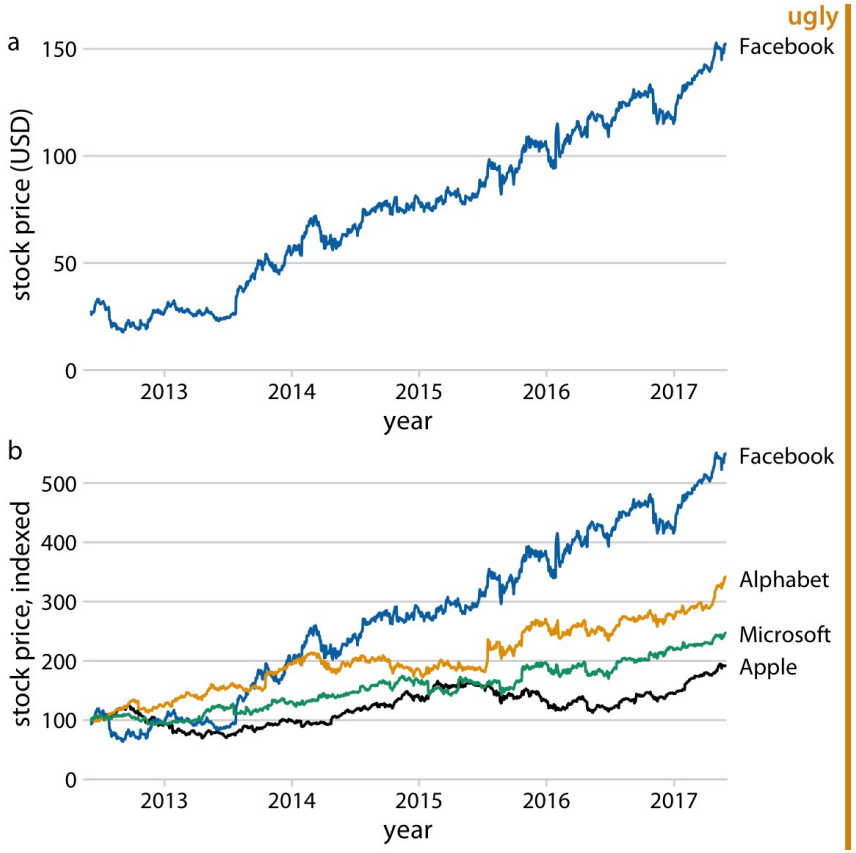


- An isotype chart is still a bar chart: each whole animal = 5 million households.
- Images should be the data, not decorate the axes. Decorative images distract (Haroz et al., 2015).

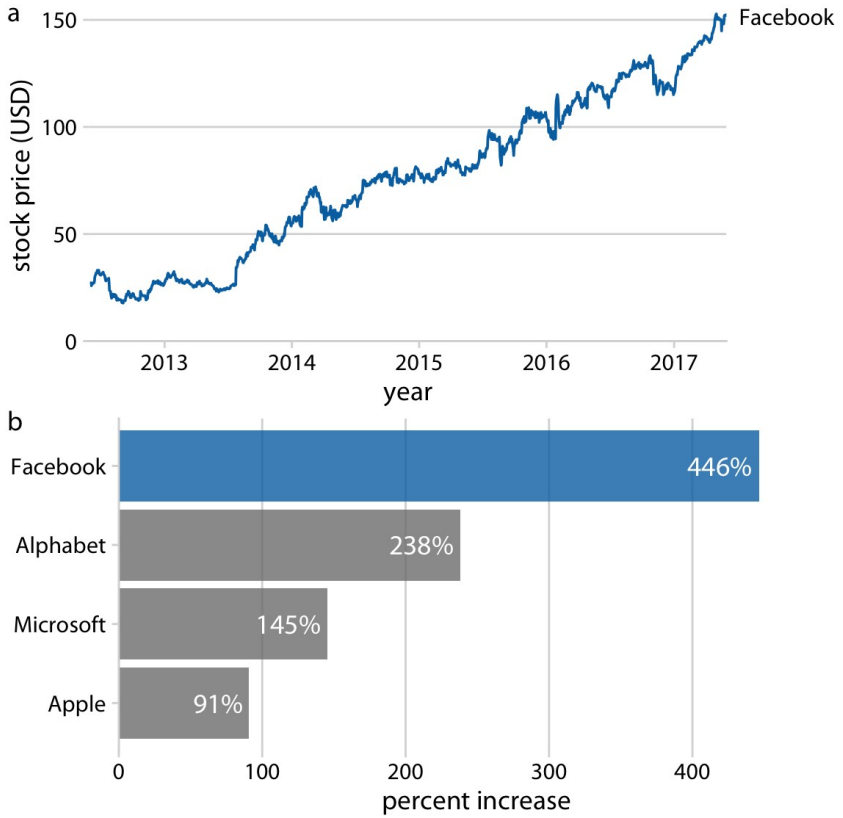
Figure 29.8 and 29.9: Claus O. Wilke, Fundamentals of Data Visualization (O'Reilly, 2019), ch. 29. clauswilke.com/dataviz Data: AVMA, 2012.

4. Be consistent, but don’t repeat yourself

REPETITIVE



DISTINCT



Same visual language (colors, fonts), different chart types for different analyses.

Panel (b) only needs to say “Facebook grew more”: a bar of % increase does that.

Raw data first, derived quantities later.

Figure 29.11 and 29.12: Claus O. Wilke, Fundamentals of Data Visualization (O'Reilly, 2019), ch. 29. clauswilke.com/dataviz Data: Yahoo Finance.

5. Check the uncertainty before you claim it

- Before you report a result, ask how sure you are: sample size, variation, missing data, confidence intervals.
- Your infographic doesn’t need error bars or CIs, but your main conclusion has to survive them.
- If a difference is inside the noise, it isn’t your headline.
- Your plots must not mislead. (More on deceptive plots later in the semester.)

Holds up

Claim: Friday the 13th has 5.1% fewer births than other Fridays.

Check: Single Fridays swing about ±7%, but 23 of the 25 Friday the 13ths came in below average.

Doesn’t hold up

Claim: Library Science is the worst-paid major ($22,000).

Check: That median comes from 2 survey respondents.

Most importantly…

Simplify,
simplify,
simplify!

Simplicity is about subtracting the obvious and adding the meaningful.

Think | Pair | Share

Think Find one infographic around SDS

Pair What's the main story? What story format does it use?
Does it follow or break the rules we just learned?

Share Take a picture of the infographic. Report back to the class.

Activity

Sketch a data story

Paper and pens only. No code.

Sketches are…

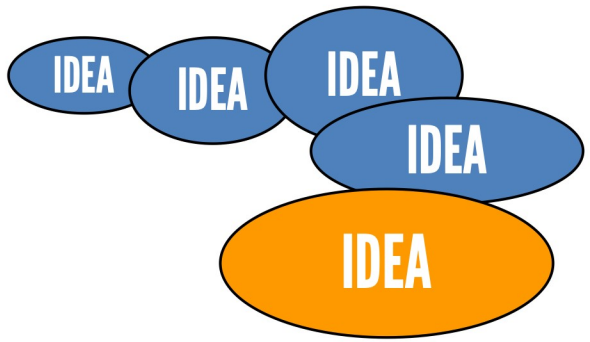
- Quick, freehand drawings
- Can include labels or captions
- Don’t need to be pretty
- **Goal**:
  - For communication
  - For brainstorming

GOAL:
try to communicate ideas with as few lines (as little “ink”) as possible!

Why Sketch?

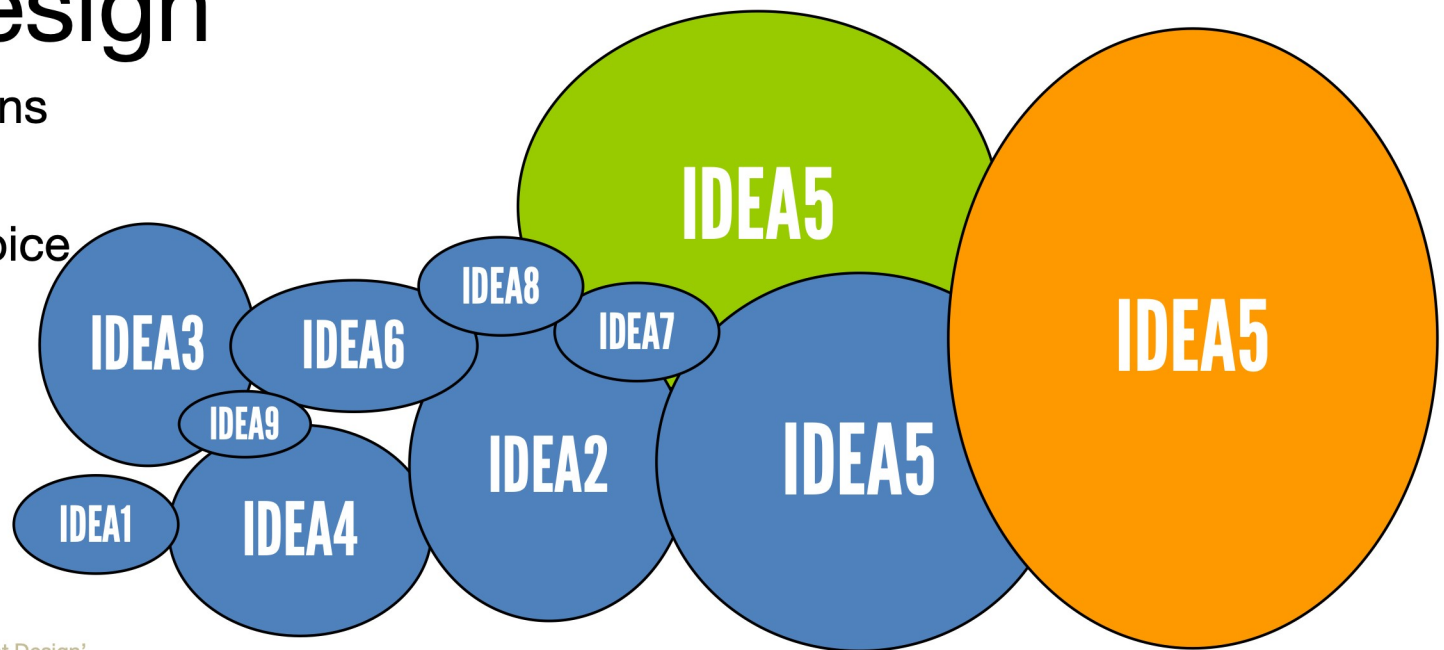
getting the design right

- generate an idea
- iterate and develop it



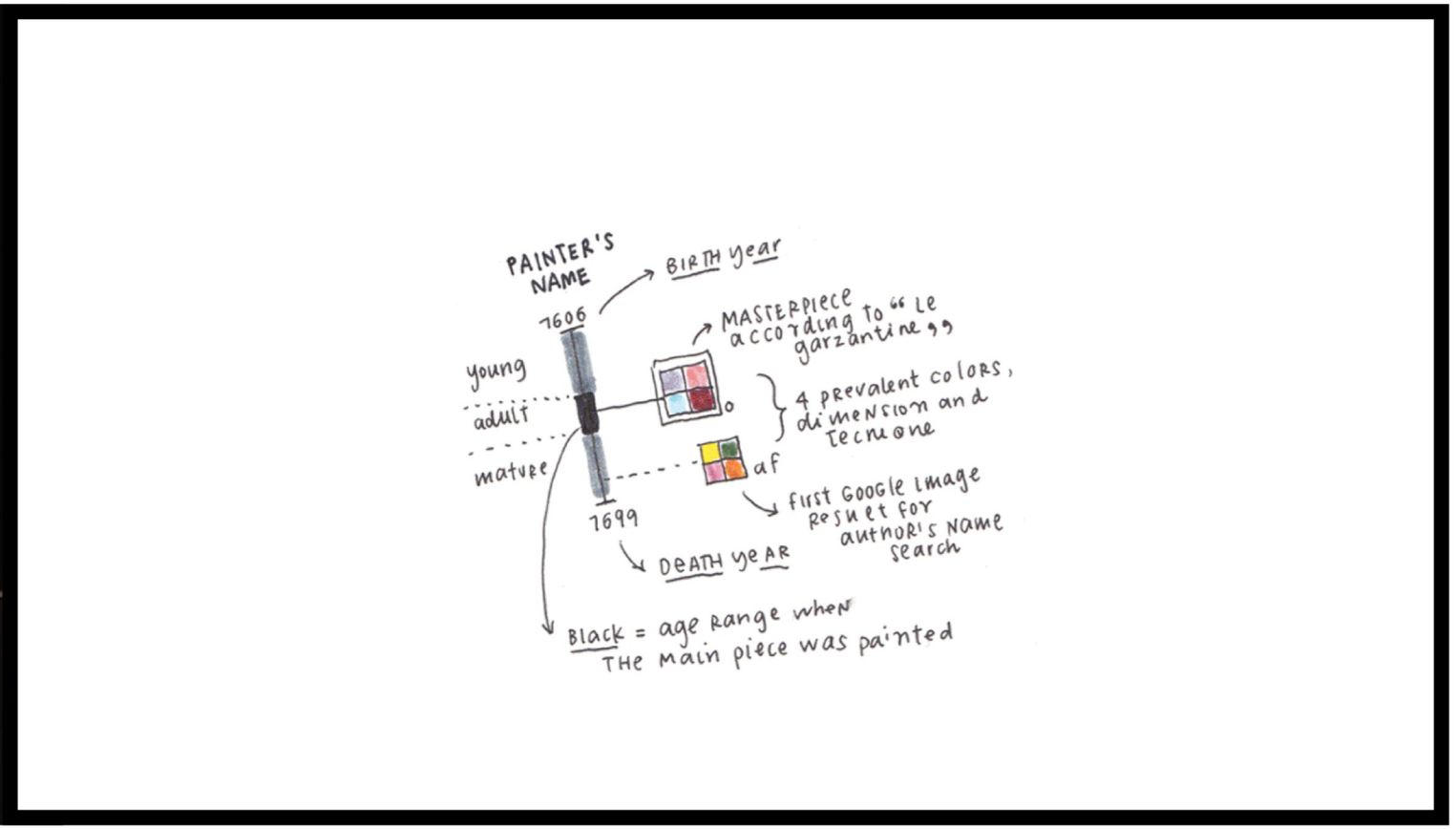
getting the right design

- generate many ideas and variations
- reflect and choose
- *then* iterate and develop your choice



Bill Buxton coined the expression 'Getting the Design Right *vs.* Getting the Right Design'





Giorgia Lupi - The Shape of My Thoughts

Watch the full talk: https://vimeo.com/channels/eyeo2014/109954123


BIG TITLE
GUEST NIGHTS BOOKED
LISTINGS ON AIRBNB
3-4 HOST INCOME FACTOIDS GO HERE
HOSTS


GUEST NIGHTS BOOKED
5 MILLION NIGHTS BOOKED
500%
2011
19,000+
192
UK 748%
FRANCE 425%
ITALY 948%
SPAIN 718%
JUST LAST YEAR...
RIGHT NOW...
AIRBNB TRAVELER NETWORK
—NETWORK EFFECTS—
2012
AND THE CYCLE BEGINS AGAIN


Longest consecutive reservation
91 nights
Properties Listed
Positive reviews left
Nights Stayed
THEN
2008
2009
2010
2011
Notable event
Notable event
Quote
Quote
Quote
Building Culture


A SNAPSHOT from FEBRUARY 2012
airbnb
GUEST NIGHTS BOOKED
5M NIGHTS BOOKED!
2011
25%
33%
42%
U.S. Guest
Non-U.S. Guest
LISTINGS ON AIRBNB

AIRBNB

BRINGING THE WORLD TOGETHER

1

10M
8M
6M
4M
2M

NIGHTS BOOKED JUNE

5M JUST 5 MONTHS AGO!

JUNE 08 JUNE 09 JUNE 10 JUNE 11 JUNE 12

10 MILLION GUEST NIGHTS BOOKED

2 A NIGHT ON AIRBNB IS BOOKED...

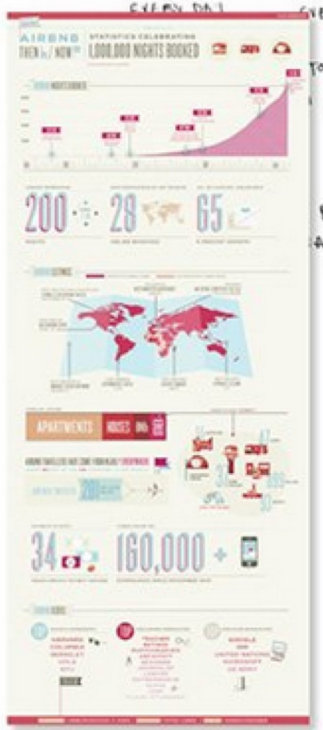
2008: EVERY DAY
2009: EVERY FIVE MINUTES
2010: EVERY MINUTE
2011: EVERY 10 SECONDS
2012: EVERY 2 SECONDS

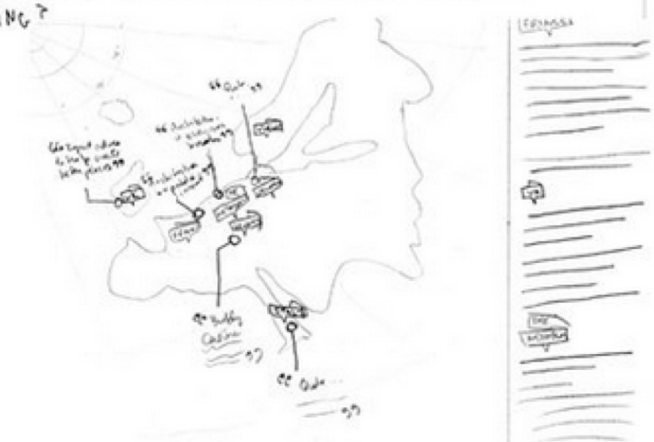
TOOK YOU TO READ THIS, SOMEONE BOOKED A NIGHT'S AIRBNB. IN FACT, IT WAS THIS GUY:

HE'S GOING TO _____ (CITY BASED ON COUNTRY)

READ THIS, 38,000 PEOPLE ARE STAYING ON AIRBNB.

AYING?





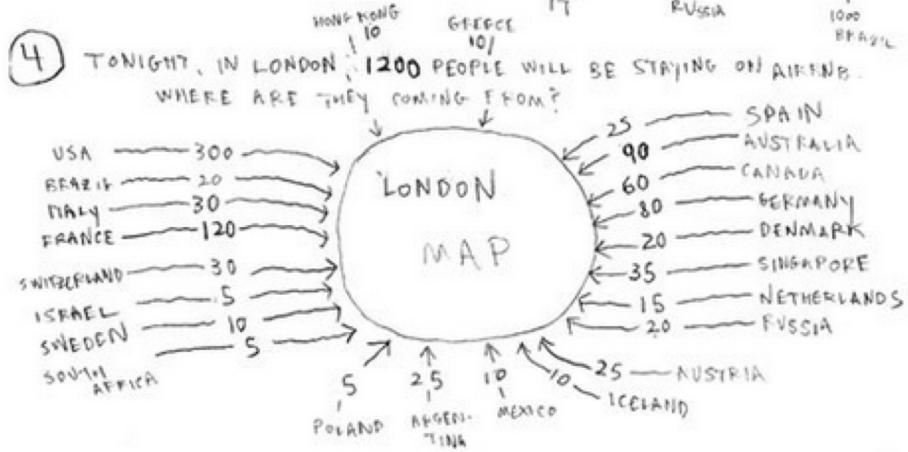
WHERE ARE THEY FROM?

38,000 STAYING ON AIRBNB RIGHT NOW

15,000 USA
2000 FRANCE
1000 ES
1000 IT
UK 3000
GERM 2500
1000 RUSSIA
CANADA 2500
1000 BRAZIL
2500 AUSTRALIA

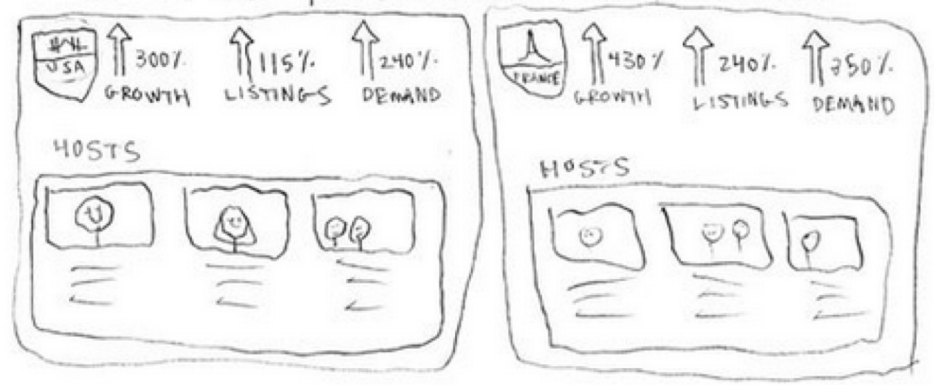
4 TONIGHT, IN LONDON, 1200 PEOPLE WILL BE STAYING ON AIRBNB.

WHERE ARE THEY COMING FROM?



5 INTERNATIONAL GROWTH -

IN THE LAST YEAR, THE AIRBNB COMMUNITY HAS GROWN...



OTHER COUNTRIES

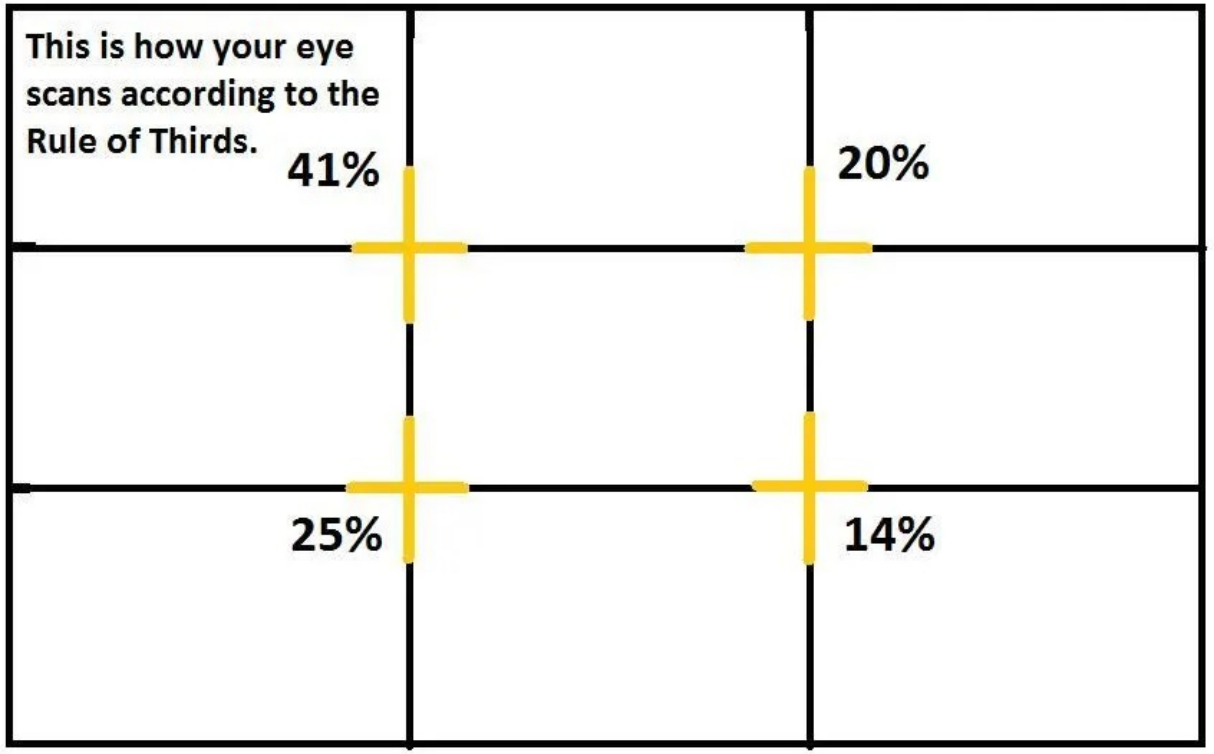
Lay out panels so the eye knows where to go

Most infographics are built from simple plot types: bar, line, pie.

Often organized like a document: read left to right, top to bottom.

Sometimes organized like a painting: one central focus with content around it.

Put the most important elements on the lines and intersections of a rule-of-thirds grid.



Where the eye lands on a rule-of-thirds grid

Rule-of-thirds image: “The rule of thirds in UX design,” Design Bootcamp (Medium)

In your group

1. **Read your story card**	Who is the audience? What do they need to believe or do?	**3 min**
2. **Write a one-sentence lead**	The single point a general should get in five seconds.	**3 min**
3. **Pick a format and map the parts**	Assign each story part to a panel. Name the chart type.	**7 min**
4. **Sketch the infographic by hand**	3–6 panels. Titles that state the point. Rough is fine.	**15 min**
5. **Swap and critique**	Use the checklist. Then revise one panel.	**10 min**

Story 1: Four stars for everything

- Audience: moviegoers deciding what to see this weekend
- Data: fandango_score_comparison.csv: 146 rows × 22 columns, films from 2015 (FiveThirtyEight)
- One row is one film with 30+ fan reviews on Fandango, with its scores on Fandango, Rotten Tomatoes, Metacritic, and IMDb.

First five rows (7 of 22 columns)

FILM	RottenTomatoes	Metacritic	IMDB	Fandango_Stars	Fandango_Ratingvalue	Fandango_votes
Avengers: Age of Ultron (2015)	74	66	7.8	5	4.5	14846
Cinderella (2015)	85	67	7.1	5	4.5	12640
Ant-Man (2015)	80	64	7.8	5	4.5	12055
Do You Believe? (2015)	18	22	5.4	5	4.5	1793
Hot Tub Time Machine 2 (2015)	14	29	5.1	3.5	3	1021

Story 1: what the data shows

- The stars Fandango showed were higher than its own average rating for 130 of 146 films, and never lower
- It rounded up by as much as half a star: 13 films gained a full half star
- No film got fewer than 3 stars
- Average score out of 5: Fandango 4.09 stars; IMDb users 3.37; Rotten Tomatoes critics 3.04; Metacritic critics 2.94
- 73% of films got 4 stars or more on Fandango, against 7% on IMDb (scaled to 5)
- “Do You Believe?” got 5 stars on Fandango and 18% from Rotten Tomatoes critics
- “Paul Blart: Mall Cop 2” got 3.5 stars on Fandango and 5% from critics
- Context: Fandango makes its money selling movie tickets
- Uncertainty: fan vote counts range from 35 to 34,846 per film

Story 2: Birthdays are scheduled

- Audience: hospital administrators planning maternity ward staffing
- Data: US_births_2000-2014_SSA.csv: 5,479 rows × 5 columns, every day from 2000 to 2014 (Social Security Administration, via FiveThirtyEight)
- One row is one day in the U.S. and how many babies were born. day_of_week: 1 = Monday, 7 = Sunday.

First five rows (all 5 columns)

year	month	date_of_month	day_of_week	births
2000	1	1	6	9083
2000	1	2	7	8006
2000	1	3	1	11363
2000	1	4	2	13032
2000	1	5	3	12558

Story 2: what the data shows

- Weekends have 37% fewer births than weekdays (8,040 vs. 12,675 a day)
- Tuesday is the busiest day (13,122 on average); Sunday the quietest (7,518)
- Christmas Day has 41% fewer births than a typical day of the same weekday; New Year’s Day 28% fewer; July 4 22% fewer
- Halloween has 11% fewer births; Valentine’s Day has 3% more
- Friday the 13th has 5.1% fewer births than other Fridays of the same year
- Mid-September is busiest: September 12, 19, and 20 have the highest averages
- The busiest single day was 9/9/09: 16,081 births
- Births per year peaked at 4.38 million in 2007 and fell 9% by 2013
- Uncertainty: single Fridays swing about ±7%, but 23 of the 25 Friday the 13ths were below average

Story 3: Majors with more women pay less

- Audience: first-year students choosing a major
- Data: recent-grads.csv: 173 rows × 21 columns, American Community Survey 2010–2012 (via FiveThirtyEight)
- One row is one college major, with jobs and yearly earnings for its graduates under 28.

First five rows (8 of 21 columns)

Major	Major_category	Total	ShareWomen	Sample_size	Median	P25th	P75th
PETROLEUM ENGINEERING	Engineering	2339	0.120564344	36	110000	95000	125000
MINING AND MINERAL ENGINEERING	Engineering	756	0.101851852	7	75000	55000	90000
METALLURGICAL ENGINEERING	Engineering	856	0.153037383	3	73000	50000	105000
NAVAL ARCHITECTURE AND MARINE ENGINEERING	Engineering	1258	0.107313196	16	70000	43000	80000
CHEMICAL ENGINEERING	Engineering	32260	0.341630502	289	65000	50000	75000

Story 3: what the data shows

- Across 172 majors, the more women in a major, the lower its median pay (correlation −0.62)
- The 10 best-paid majors average 23% women; the 10 worst-paid average 79%
- Majors that are 75%+ women: median pay $32,400; majors with 25% or fewer women: $50,000
- Highest: Petroleum Engineering, $110,000, 12% women. Lowest: Library Science, $22,000, 88% women
- A typical engineering major pays $57,000 (24% women); education $32,750 (81% women)
- Pay varies within a major: the middle half of Computer Science grads earn $39,000–$70,000
- Uncertainty: 32 majors have fewer than 30 survey respondents; Library Science has 2
- Caveat: this compares majors, not men and women in the same job, and it doesn’t show why

Story 4: Smaller budgets, same returns

- Audience: studio executives deciding which films to fund
- Data: bechdel_movies.csv: 1,794 rows × 15 columns, films from 1970 to 2013 (BechdelTest.com and The-Numbers.com, via FiveThirtyEight)
- One row is one film, whether it passes the Bechdel test (two women who talk to each other about something besides a man), and its budget and box office. Columns ending in _2013$ are adjusted for inflation.

First five rows (7 of 15 columns)

year	title	clean_test	binary	budget_2013$	domgross_2013$	intgross_2013$
2013	21 & Over	notalk	FAIL	13000000	25682380	42195766
2012	Dredd 3D	ok	PASS	45658735	13611086	41467257
2013	12 Years a Slave	notalk	FAIL	20000000	53107035	158607035
2013	2 Guns	notalk	FAIL	61000000	75612460	132493015
2013	42	men	FAIL	40000000	95020213	95020213

Story 4: what the data shows

- 45% of films pass (803 of 1,794)
- Why films fail: in 514 the women never talk to each other, in 194 they only talk about men, and 141 have fewer than two women
- Passing films get smaller budgets: median $31.5M vs. $44.0M for films that fail (2013 dollars)
- Only 31% of $100M+ films pass, compared with 52% of films under $20M
- Per dollar spent they earn about the same: median worldwide gross is 2.7× the budget for passing films, 2.6× for failing ones
- The pass rate rose from 26% in the 1970s to 49% in the 2000s
- Users disputed 425 of the ratings (the “-disagree” tags in the test column)
- Watch for errors: Shaft (1971) is listed with a $305M budget, likely the 2000 remake’s numbers
- Uncertainty: returns vary hugely (the middle half runs from about 1.3× to 5×), so “earn the same per dollar” holds up but “earn more” doesn’t

Story 5: Past crashes don’t predict future crashes

- Audience: travelers nervous about choosing an airline
- Data: airline-safety.csv: 56 rows × 8 columns, 1985–1999 and 2000–2014 (Aviation Safety Network, via FiveThirtyEight)
- One row is one airline, with its incidents, fatal accidents, and deaths in two 15-year periods. An asterisk means regional subsidiaries are included.

First five rows (all 8 columns)

airline	avail_seat_km_per_week	incidents_85_99	fatal_accidents_85_99	fatalities_85_99	incidents_00_14	fatal_accidents_00_14	fatalities_00_14
Aer Lingus	320906734	2	0	0	0	0	0
Aeroflot*	1197672318	76	14	128	6	1	88
Aerolineas Argentinas	385803648	6	0	0	1	0	0
Aeromexico*	596871813	3	1	64	5	0	0
Air Canada	1865253802	2	0	0	2	0	0

Story 5: what the data shows

- Flying got safer: fatal accidents fell from 122 to 37 (−70%), and deaths from 6,295 to 3,109
- 32 of 56 airlines had no fatal accidents from 2000 to 2014
- An airline’s incidents in 1985–99 somewhat predict its incidents in 2000–14 (correlation 0.4)
- Its deaths don’t predict later deaths at all (correlation 0.05)
- Japan Airlines (520 deaths) and Korean Air (425) were among the worst in 1985–99; neither had any deaths in 2000–14
- 5 airlines with no fatal accidents in 1985–99 had at least one later, including Alaska and SAS
- Malaysia Airlines had the most deaths in 2000–14: 537, from 2 crashes
- Bigger airlines fly more, so compare per seat-kilometer, not raw counts
- Uncertainty: fatal crashes are rare, so one or two crashes swing an airline’s record

Swap and critique

The generals
Can someone get your lead in five seconds?

Roles
Does each panel play one clear role in the arc?

Order
Does raw data come before derived numbers?

Variety
Do different analyses use different chart types?

Focus
Is anything on the page tangential to the story? Cut it.

Honesty
Does the story claim more than the data shows?

Takeaways

Give them the story	or the audience will invent one.
One figure, one role	A set of figures makes the arc; a single figure can only be a lead.
Simplify, then build up	Make it for the generals; show one panel before the grid.
Memorable and clear	Use images that are the data. Vary chart types, keep the visual language.
Raw first, derived later	And 3–6 figures per story.

Reading: Wilke, Fundamentals of Data Visualization, ch. 29, clauswilke.com/dataviz/telling-a-story.html