MODULE 6 | VIDEO 1

Data stories

with The Pudding

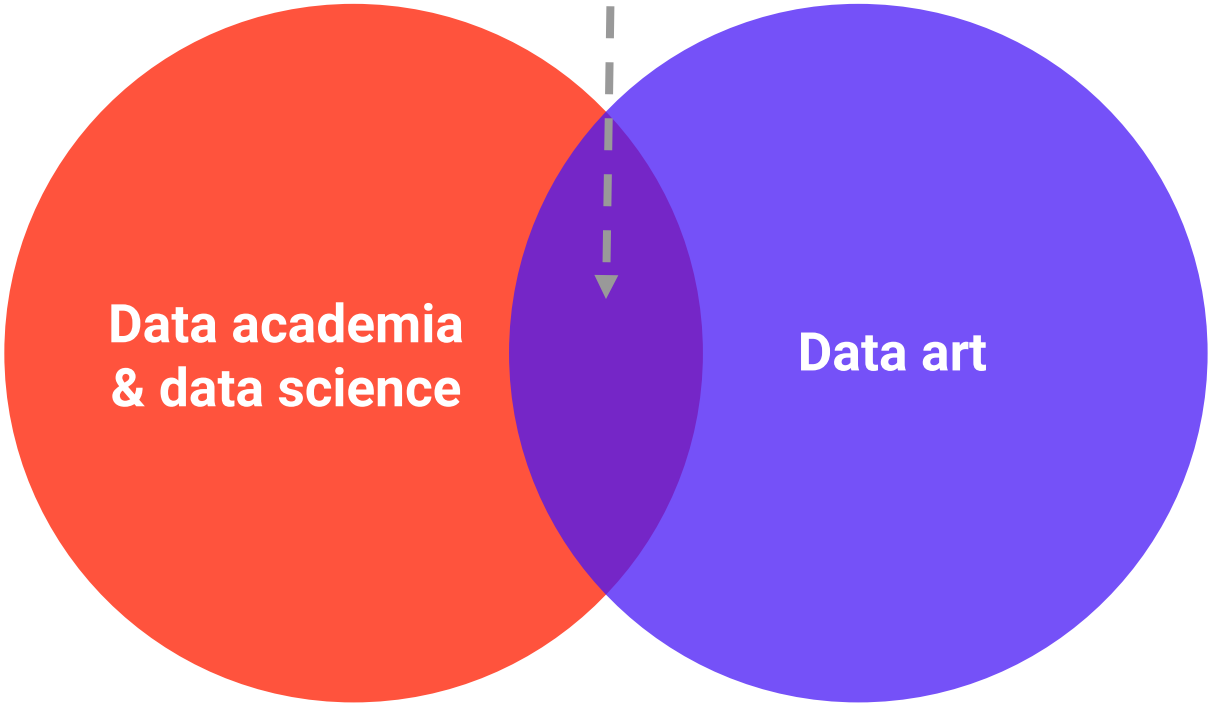
Data academia & data science

- dense
- complex
- static/non-interactive
- often not accompanied by an easy-to-follow narrative

Data art

- beautiful
- abstract
- compelling
- complex

Data storytelling



MODULE 6 | VIDEO 2

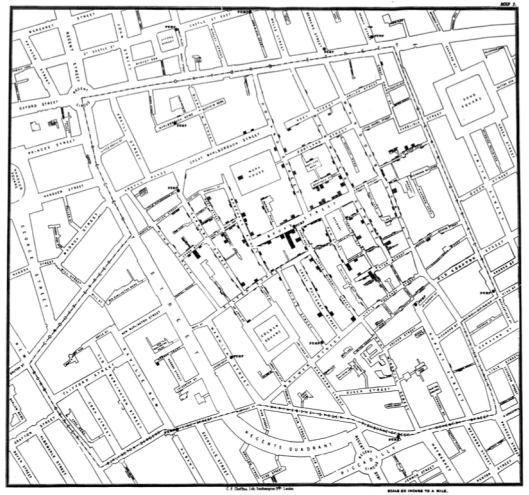
Data stories

with The Pudding

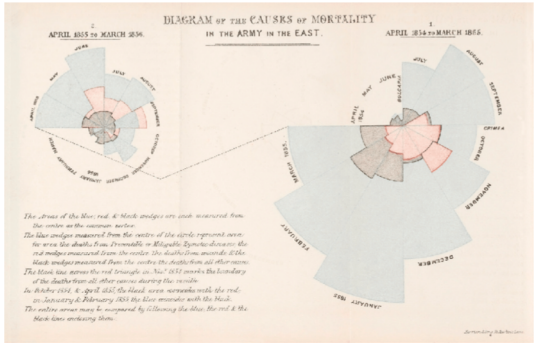
MODULE 6 | VIDEO 2

How did we get where we are today?

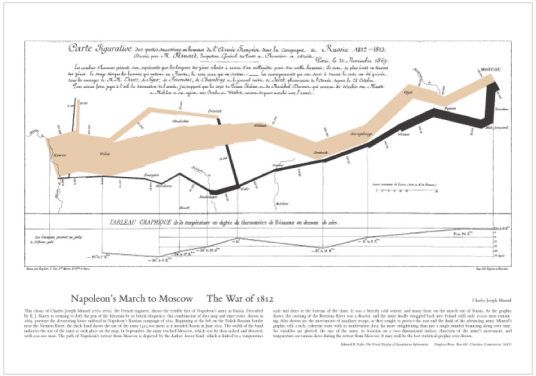
MODULE 6 | VIDEO 2



John Snow's Cholera map



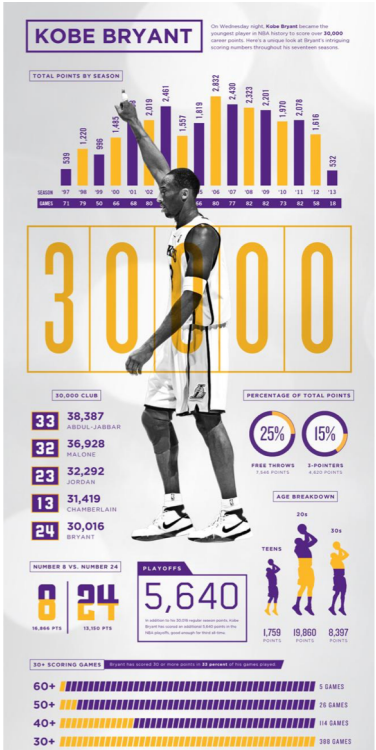
Florence Nightingale's rose chart



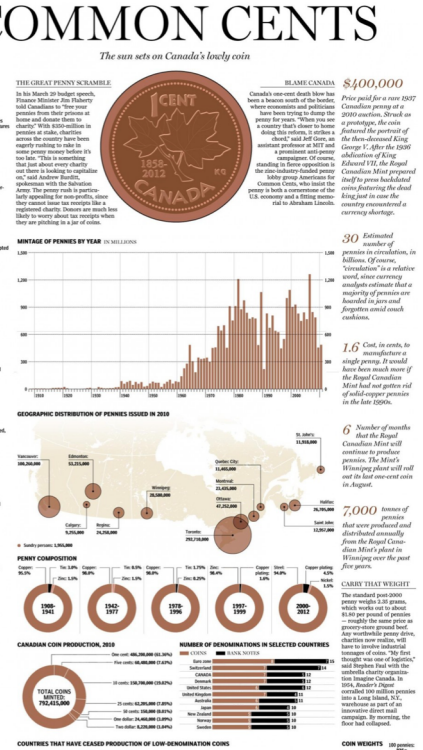
Charles Joseph Minard's march diagram

MODULE 6 | VIDEO 2

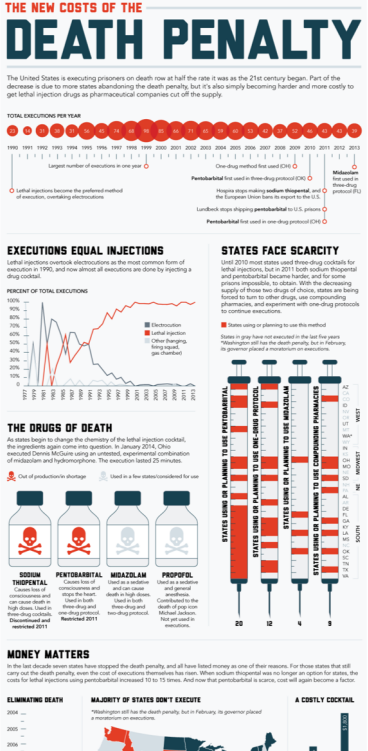
[J Alexander Diaz/Lakers.com]



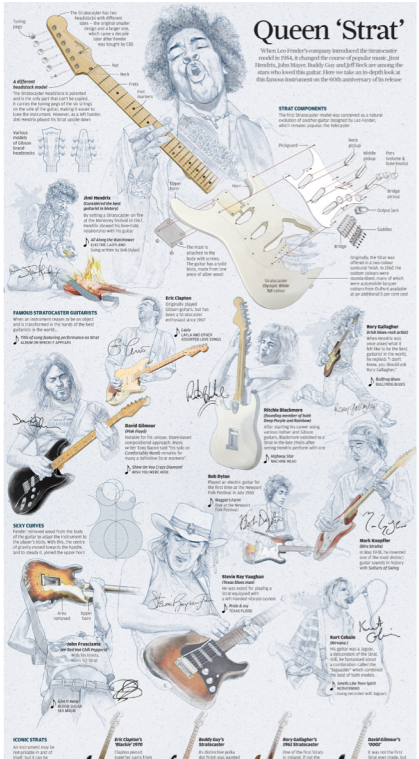
[Richard Johnson/The National Post]



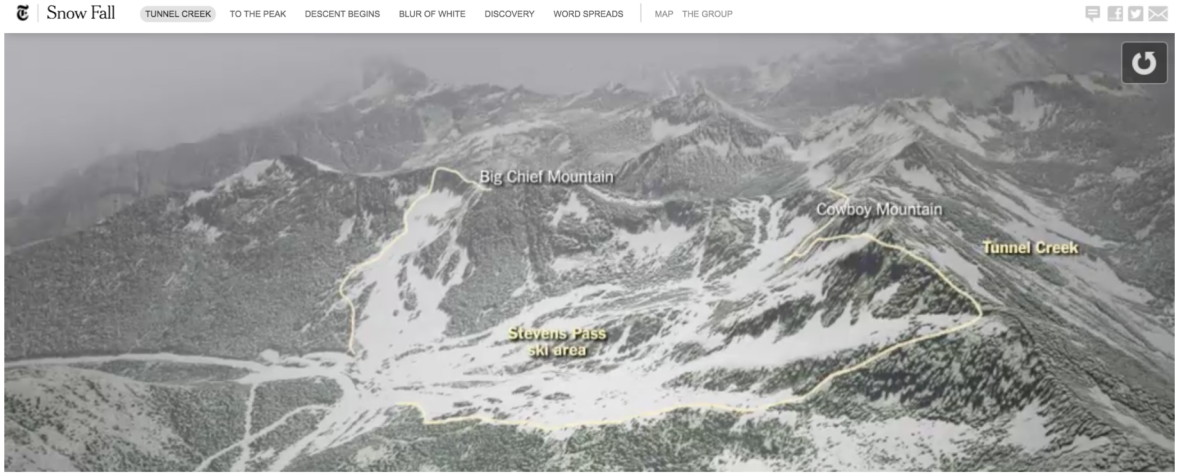
[Jan Diehm/HuffPost]



[Alberto Lucas López/SCMP]



MODULE 6 | VIDEO 2



It is a term with broad meaning. The name is derived from the Cascade Tunnel, originally a 2.6-mile railroad tube completed in 1900 that connected the east and west sides of the Cascades, a boon for the growth of Seattle and Puget Sound. The mountain pass that it burrowed beneath was named for the project's engineer, John Frank Stevens, who later helped build the Panama Canal.

In late February 1910, ceaseless snowstorms over several days marooned two passenger trains just outside the tunnel's west portal. Before the tracks could be cleared, the trains were buried by what still stands as the nation's deadliest avalanche. It killed 96 people.

Snowfall, The New York Times

MODULE 6 | VIDEO 2

Jan Diehm/The Guardian (unpublished)

Jan Diehm/The Guardian (unpublished)

MODULE 6 | VIDEO 2

What does a data story look like now?

Scrollytelling

- Twenty Years of the NBA Redrafted
  The Pudding

- Going Gray
  Reuters

- Here's How America Uses Its Land
  Bloomberg News

Microstories

- Data illustrations
  Mona Chalabi

- Dear Data
  Giorgia Lupi and Stefanie Posavec

- The Daily Chart
  The Economist

Non-traditional

- Apollo 11: As They Shot It
  The New York Times
- Concussion Protocol
  Josh Begley
- Borderline
  The Washington Post

Explainers

- Parable of the Polygons
  Vi Hart and Nicky Case

- How To: Tune a Guitar
  Matthew Conlen and Alex Kale

- Let’s Learn about Waveforms
  The Pudding

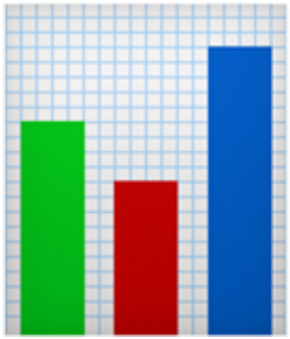
MODULE 6 | VIDEO 3

Data stories

with The Pudding

MODULE 6 | VIDEO 3

How do you find stories?



Question first

Data second

MODULE 6 | VIDEO 3

Where do your questions come from?

The bad news

The good news

MODULE 6 | VIDEO 3

How do you make sure your idea is interesting?

Sparks intense debate or discussion

Gently pitch the idea to others and get feedback

Trust your instincts

I'm a cool dad. That's my thang.

MODULE 6 | VIDEO 3

What types of stories make for successful data stories?

👍 Works

- DATA!
- Universal truths
- Visuals help communicate the story
- Question driven

👎 Doesn’t work

- Highly personal
- Stuff that “boils the ocean”
- Too obvious - “water is wet”

1. Question

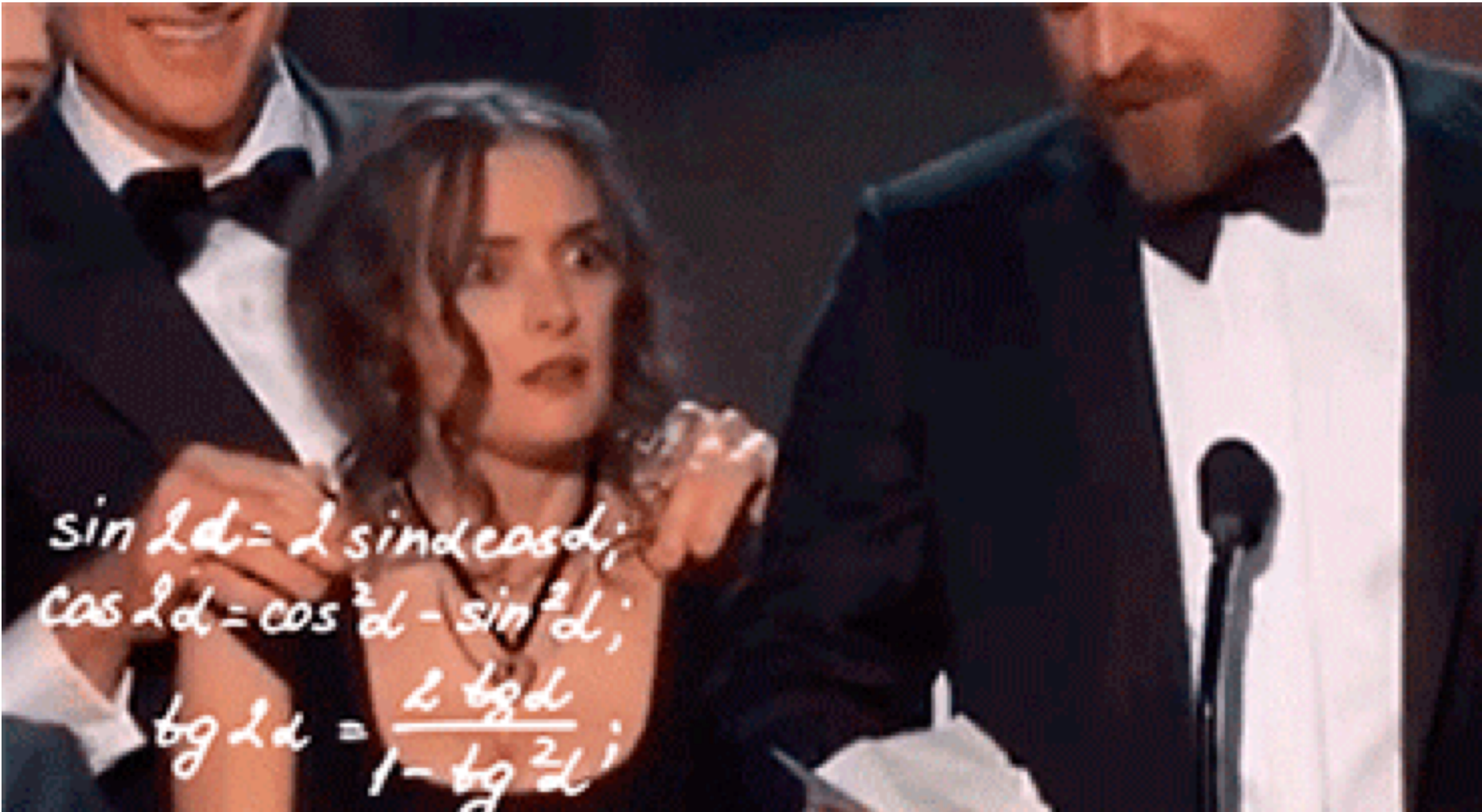
2. Data
3. Storyboard
4. Design
5. Development

MODULE 6 | VIDEO 3

1. Questions

2. Data

3. Storyboard
4. Design
5. Development


$\sin 2\alpha = 2\sin\alpha\cos\alpha;$
$\cos 2\alpha = \cos^2\alpha - \sin^2\alpha;$
$\operatorname{tg} 2\alpha = \frac{2\operatorname{tg}\alpha}{1-\operatorname{tg}^2\alpha}$

R programming

Python

Node.js

QGIS

MySQL

Google Sheets/Excel

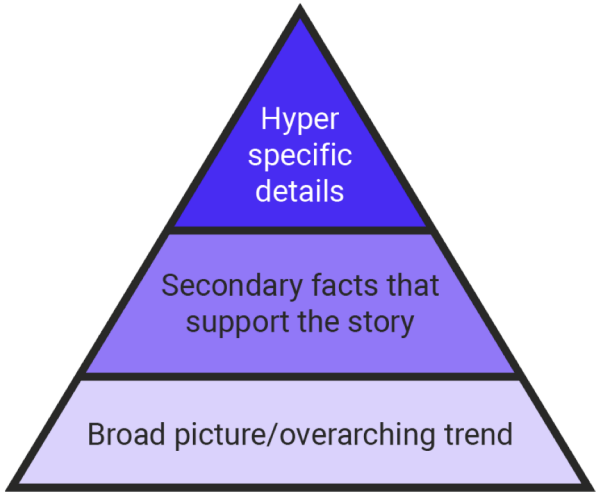
MODULE 6 | VIDEO 3

1. Questions
2. Data

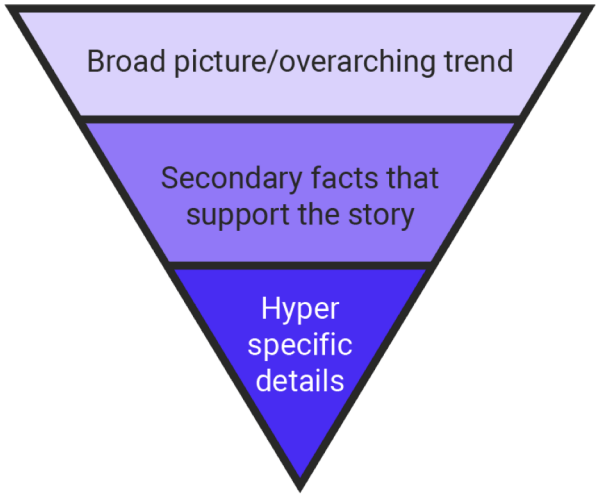
3. Storyboard

4. Design
5. Development

Small to big



Big to small



MODULE 6 | VIDEO 3

1. Questions
2. Data
3. Storyboard

4. Design

5. Development

Visual vocabulary

Designing with data

Deviation

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (positive/neutral/negative).

Example FT uses
Trade surplus/deficit, climate change

Diverging bar

A simple standard bar chart that can handle both negative and positive magnitude values.

Diverging stacked bar

Perfect for presenting survey results which involve sentiment (eg disagree/neutral/agree).

Spine

Splits a single value into two contrasting components (eg male/female).

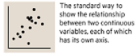
Surplus/deficit filled line

The shaded area of these charts allows a balance to be shown – either against a baseline or between two series.

Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationships you show them to be causal (ie. one causes the other).

Example FT uses
Inflation and unemployment, income and life expectancy

Scatterplot

The standard way to show the relationship between two continuous variables, each of which has its own axis.

Column + line timeline

A good way of showing the relationship between an amount (columns) and a rate (line).

Connected scatterplot

Usually used to show how the relationship between 2 variables has changed over time.

Bubble

Like a scatterplot, but adds additional detail by sizing the circles according to a third variable.

XY heatmap

A good way of showing the patterns between 2 categories of data, less effective at showing fine differences in amounts.

Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

Example FT uses
Wealth, deprivation, league tables, constituency election results

Ordered bar

Standard bar charts display the ranks of values much more easily when sorted into order.

Ordered column

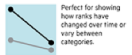
See above.

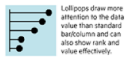
Ordered proportional symbol

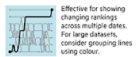
Use when there are big variations between values and/or seeing fine differences between data is not so important.

Dot strip plot

Dots placed in order on a strip are a space-efficient method of laying out ranks across multiple categories.

Slope

Perfect for showing how ranks have changed over time or vary between categories.

Lollipop

Lollipops draw more attention to the data value than standard bar/column and can also show rank and value effectively.

Bump

Effective for showing changing rankings across multiple dates. For large datasets, consider grouping lines using colour.

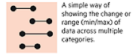
Distribution

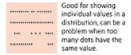
Show values in a dataset and how often they occur. The shape (or 'skew') of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.

Example FT uses
Income distribution, population (age/sex) distribution, revealing inequality

Histogram

The standard way to show a statistical distribution – keep the gaps between columns small to highlight the 'shape' of the data.

Dot plot

A simple way of showing the change or range (min/max) of data across multiple categories.

Dot strip plot

Good for showing individual values in a distribution, can be a problem when too many dots have the same value.

Barcode plot

Like dot strip plots, good for displaying all the data in a table, they work best when highlighting individual values.

Boxplot

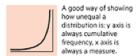
Summarise multiple distributions by showing the median (centre) and range of the data

Violin plot

Similar to a box plot but more effective with complex distributions (data that cannot be summarised with simple average).

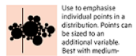
Population pyramid

A standard way for showing the age and sex breakdown of a population distribution; effectively, back to back histograms.

Cumulative curve

A good way of showing how unequal a distribution is; y axis is always cumulative frequency, x axis is always a measure.

Frequency polygons

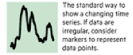
For displaying multiple distributions of data. Like a regular line chart, best limited to a maximum of 3 or 4 datasets.

Beeswarm

Use to emphasise individual points in a distribution. Points can be sized to an additional variable. Best with medium-sized datasets

Change over Time

Give emphasis to changing trends. These can be short (intra-day) movements or extended series traversing decades or centuries: Choosing the correct time period is important to provide suitable context for the reader.

Example FT uses
Share price movements, economic time series, sectoral changes in a market

Line

The standard way to show a changing time series. If data are irregular, consider markers to represent data points.

Column

Columns work well for showing change over time - but usually best with only one series of data at a time.

Column + line timeline

A good way of showing the relationship over time between an amount (columns) and a rate (line).

Slope

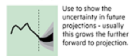
Good for showing changing data as long as the data can be simplified into 2 or 3 points without missing a key part of story.

Area chart

Use with care – these are good at showing changes to total, but seeing change in components can be very difficult.

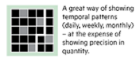
Candlestick

Usually focused on day-to-day activity, these charts show opening/closing and high/low points of each day.

Fan chart (projections)

Use to show the uncertainty in future projections – usually this grows the further forward to projection.

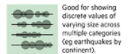
Connected scatterplot

A good way of showing changing data for two variables whenever there is a relatively clear pattern of progression.

Calendar heatmap

A great way of showing temporal patterns (daily, weekly, monthly) – at the expense of showing precision in quantity.

Priestley timeline

Great when date and duration are key elements of the story in the data.

Circle timeline

Good for showing discrete values of varying size across multiple categories (eg earthquakes by continent).

Magnitude

Show size comparisons. These can be relative (just being able to see larger/bigger) or absolute (need to see fine differences). Usually these show a 'counted' number (for example, barrels, dollars or people) rather than a calculated rate or per cent.

Example FT uses
Commodity production, market capitalisation, volumes in general

Column

The standard way to compare the size of things. Must always start at 0 on the axis.

Bar

See above. Good when the data are not time series and labels have long category names.

Paired column

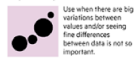
As per standard column but allows for multiple series. Can become tricky to read with more than 2 series.

Paired bar

See above.

Marimekko

A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.

Proportional symbol

Use when there are big variations between values and/or seeing fine differences between data is not so important.

Isotype (pictogram)

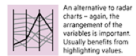
Excellent solution in some instances – use only with whole numbers (do not slice off an arm to represent a decimal).

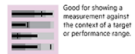
Lollipop

Lollipop charts draw more attention to the data value than standard bar/column – does not have to start at zero (but preferable).

Radar

A space-efficient way of showing value of multiple variables– but make sure they are organised in a way that makes sense to reader.

Parallel coordinates

An alternative to radar charts – again, the arrangement of the variables is important. Usually benefits from highlighting values.

Bullet

Good for showing a measurement against the context of a target or performance range.

Part-to-whole

Show how a single entity can be broken down into its component elements. If the reader's interest is solely in the size of the components, consider a magnitude-type chart instead.

Example FT uses
Fiscal budgets, company structures, national election results

Stacked column/bar

A simple way of showing part-to-whole relationships but can be difficult to read with more than a few components.

Marimekko

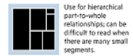
A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.

Pie

A common way of showing part-to-whole data – but be aware that it's difficult to accurately compare the size of the segments.

Donut

Similar to a pie chart – but the centre can be a good way of making space to include more information about the data (eg total).

Treemap

Use for hierarchical part-to-whole relationships; can be difficult to read when there are many small segments.

Voronoi

A way of turning points into areas – any point within each area is closer to the central point than any other centroid.

Arc

A hemicycle, often used for visualising parliamentary composition by number of seats.

Gridplot

Good for showing % information, they work best when used on whole numbers and work well in small multiple layout form.

Venn

Generally only used for schematic representation.

Waterfall

Can be useful for showing part-to-whole relationships where some of the components are negative.

Spatial

Aside from locator maps only used when precise locations or geographical patterns in data are more important to the reader than anything else.

Example FT uses
Population density, natural resource locations, natural disaster risk/impact, catchment areas, variation in election results

Basic choropleth (rate/ratio)

The standard approach for putting data on a map – should always be rates rather than totals and use a sensible base geography.

Proportional symbol (count/magnitude)

Use for totals rather than rates – be wary that small differences in data will be hard to see.

Flow map

For showing unambiguous movement across a map.

Contour map

For showing areas of equal value on a map. Can use deviation colour schemes for showing +/- values

Equalised cartogram

Converting each unit on a map to a regular and equally-sized shape – good for representing voting regions with equal value.

Scaled cartogram (value)

Stretching and shrinking a map so that each area is sized according to a particular value.

Dot density

Used to show the location of individual events/locations – make sure to annotate any patterns the reader should see.

Heat map

Grid-based data values mapped with an intensity colour scale. As choropleth map – but not snapped to an admin/political unit.

Flow

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations.

Example FT uses
Movement of funds, trade, migrants, lawsuits, information; relationship graphs.

Sankey

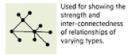
Shows changes in flows from one condition to at least one other, good for tracing the eventual outcome of a complex process.

Waterfall

Designed to show the sequencing of data through a flow process, typically budgets. Can include +/- components.

Chord

A complex but powerful diagram which can illustrate 2-way flows (and net winner) in a matrix.

Network

Used for showing the strength and inter-connectedness of relationships of varying types.

FT

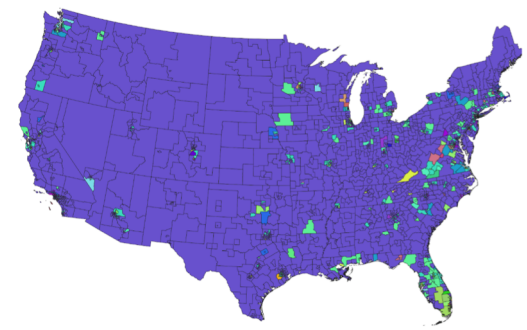
© Financial Times 2016-2019
Licensed under the terms of the MIT license.
For a copy, see opensource.org/licenses/MIT

MODULE 6 | VIDEO 3

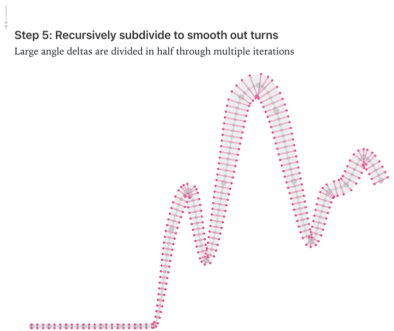
	A	B	C	D	E	F	G	H
1		1950	1960	1970	1980	1990	2000	2010
3	mlb	0.00%	0.00%	0.00%	0.00%	0.00%	0.05%	0.14%
4	mls					0.21%	0.47%	1.58%
5	nba	0.00%	0.21%	0.15%	0.00%	0.44%	0.87%	2.07%
6	nfl	0.00%	0.00%	0.11%	0.08%	0.19%	0.39%	1.04%
7	nhl	0.00%	0.00%	0.00%	0.00%	0.08%	0.15%	0.49%
8	nwls							0.55%
9	wnba					5.53%	4.75%	5.24%

Google sheets

Flourish



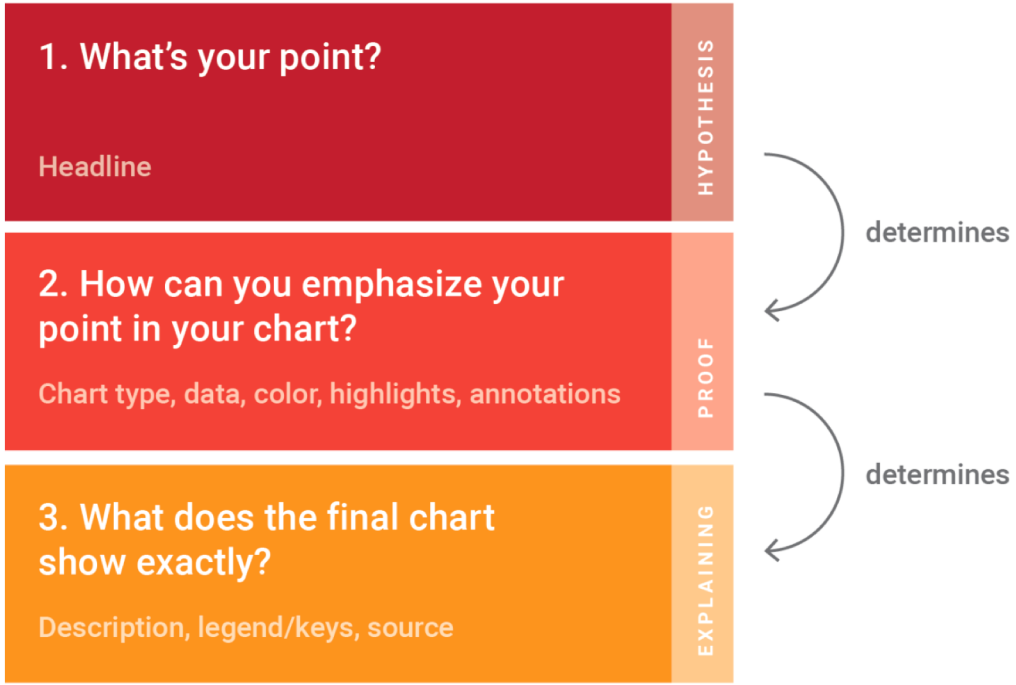
QGIS



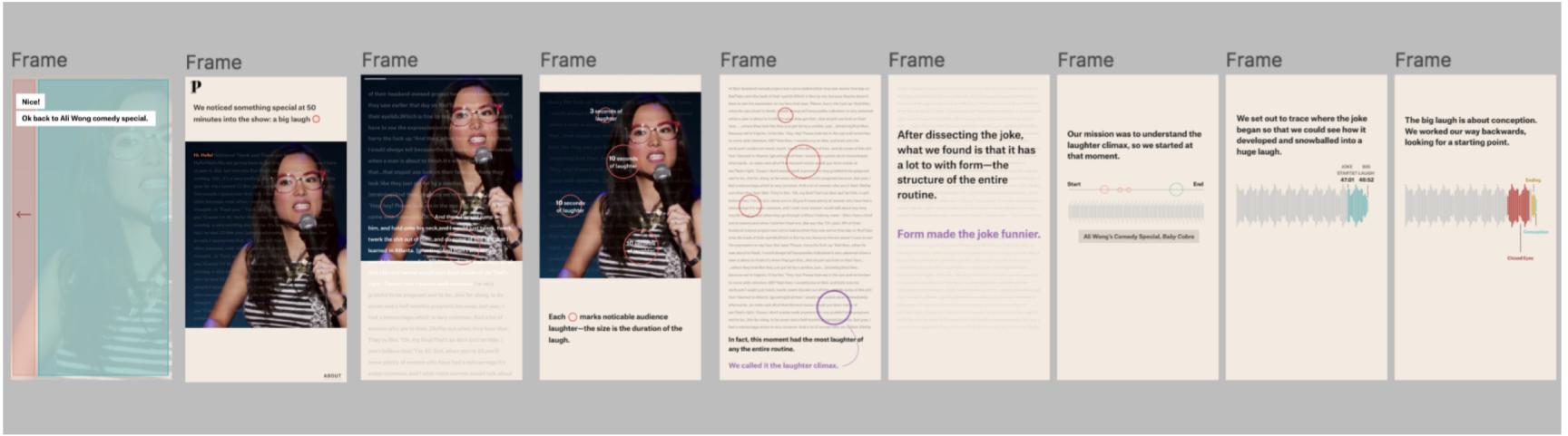
Observable

3 questions for creating a chart

Chartable blog, Datawrapper



MODULE 6 | VIDEO 3



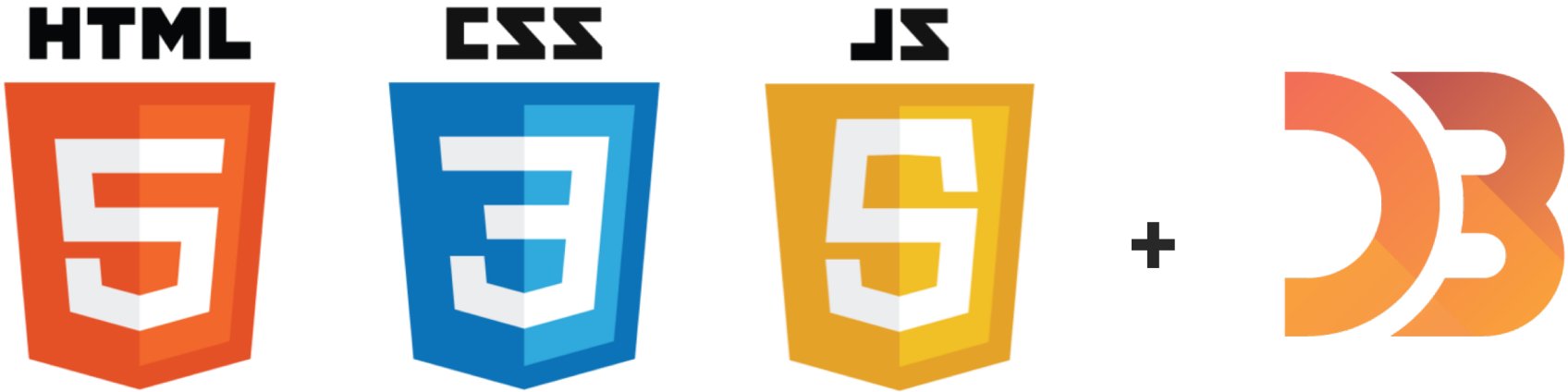
VORTEX
PULL
24 HR DOOR SERVICE
LOCKS • GLASS
800
69-VORTEX
(800) 698-6783

I HATE DOORS!
Pull
#newgirl

MODULE 6 | VIDEO 3

1. Questions
2. Data
3. Storyboard
4. Design

5. Development


HTML
CSS
JS

MODULE 6 | VIDEO 4

Data stories

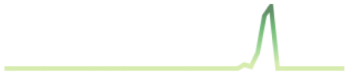
with The Pudding

MODULE 6 | VIDEO 4

Spend the majority of time crafting the story.

MODULE 6 | VIDEO 4

hyphenated-names-viz

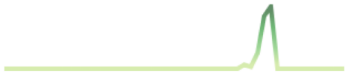


JavaScript MIT 0 0 0 0 Updated on May 10

hyphenated-names

HTML MIT 0 2 1 0 Updated on May 10

hyphenated-names-viz



JavaScript MIT 0 0 0 0 Updated on May 10

hyphenated-names

HTML MIT 0 2 1 0 Updated on May 10

hyphenated-story

Story MIT 0 2 1 0 Updated on May 10

MODULE 6 | VIDEO 4

It’s equally important to show as tell.

Show

+

Tell

=

Powerful connection

MODULE 6 | VIDEO 4

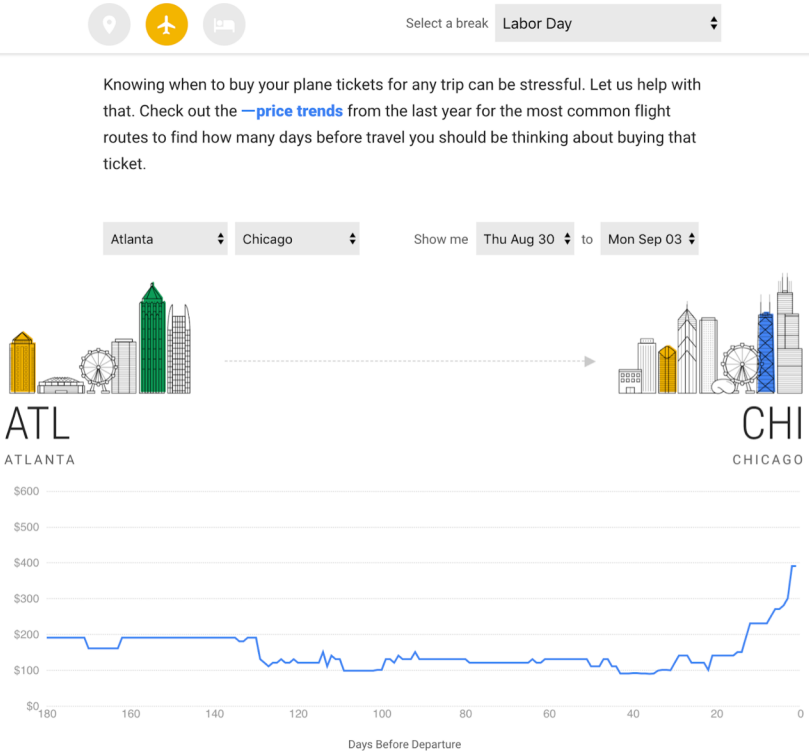
Know how your audience interacts with your piece.

- Although a lot of readers may never interact with your piece, the most loyal will
- They allow readers to dig deep and explore the full dataset
- They can pull back the curtain, making data more transparent and trustworthy
- And they can add whimsy and personal connections to your story

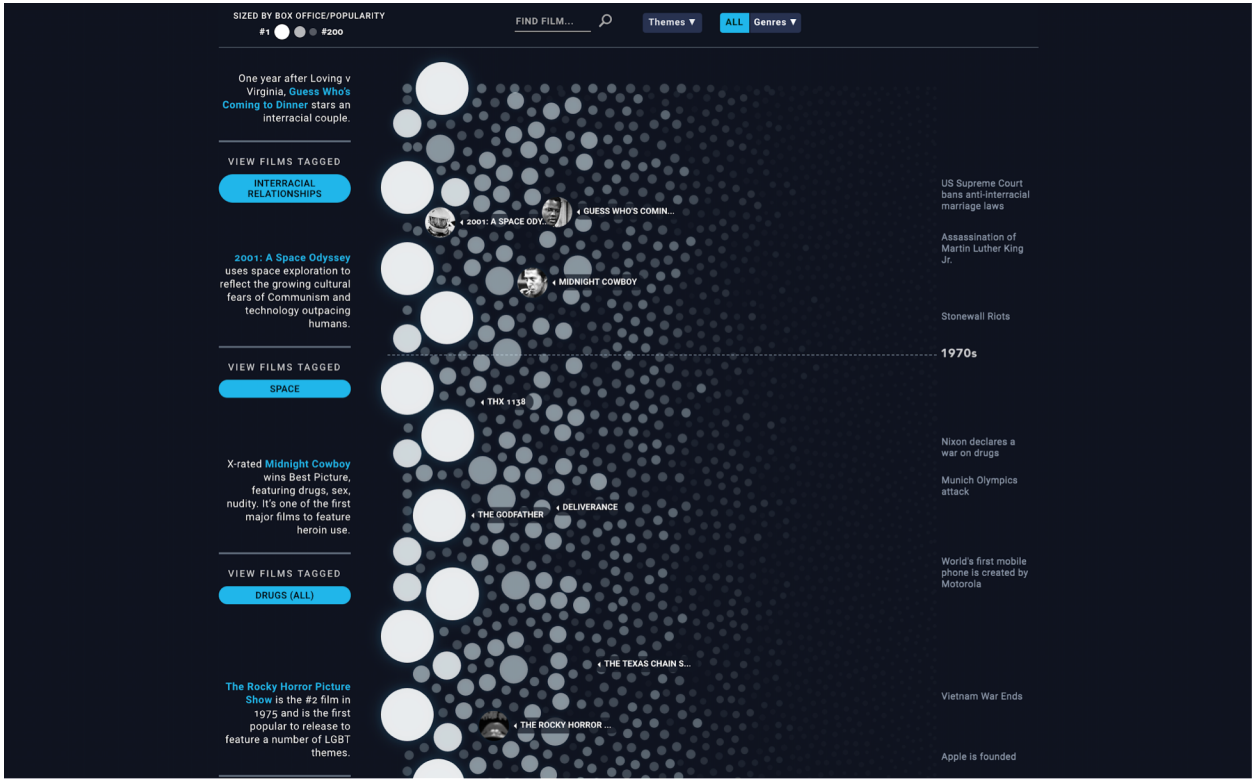
MODULE 6 | VIDEO 4

Embrace the ‘littles of visualization design.’

Orientation assistance or mini maps

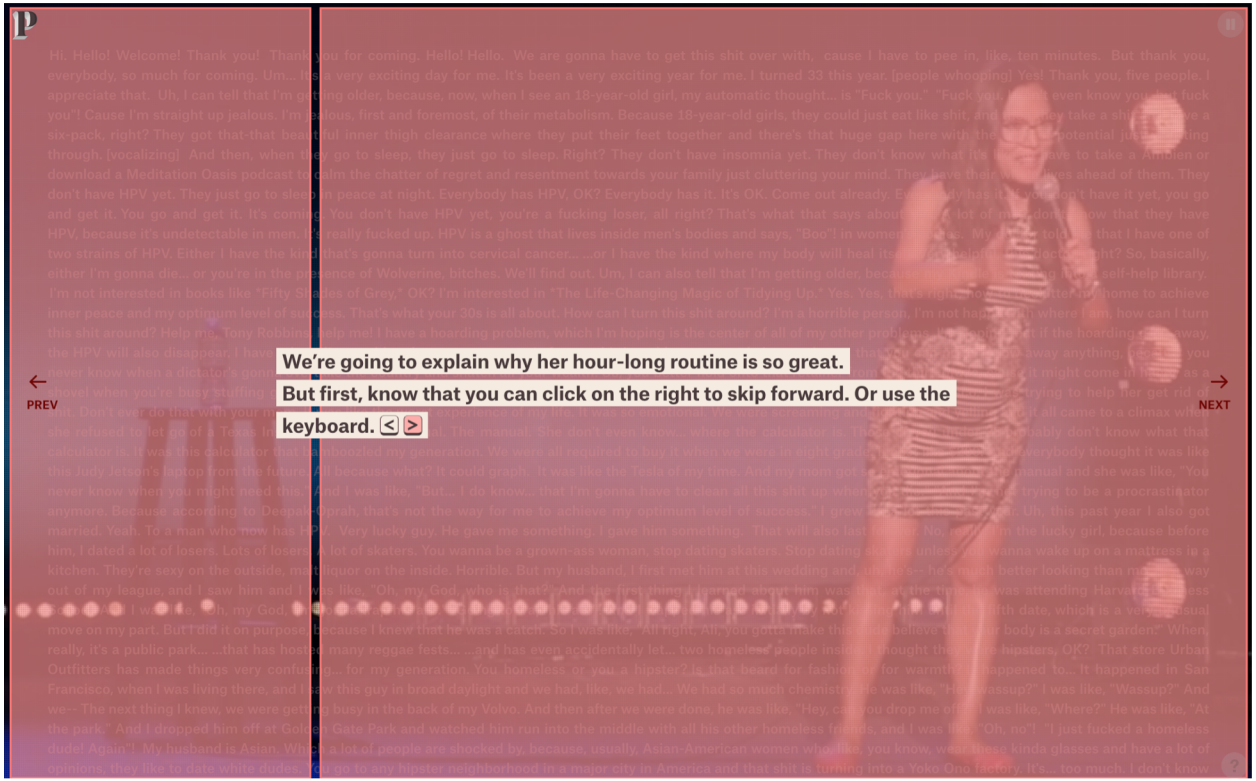


Integrated graphics into text



Fading small values

MODULE 6 | VIDEO 4

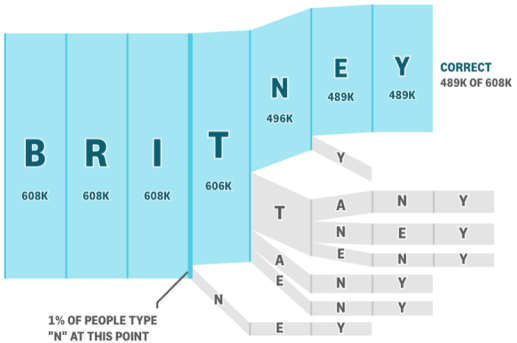


Seamless help

Reader guides

HOW IT WORKS

This chart shows over 600k attempts of people trying to search for **Britney Spears.** Here are the 8 most common permutations of her first name, Britney. People nail the first three letters —then things get interesting.



YOUR TURN

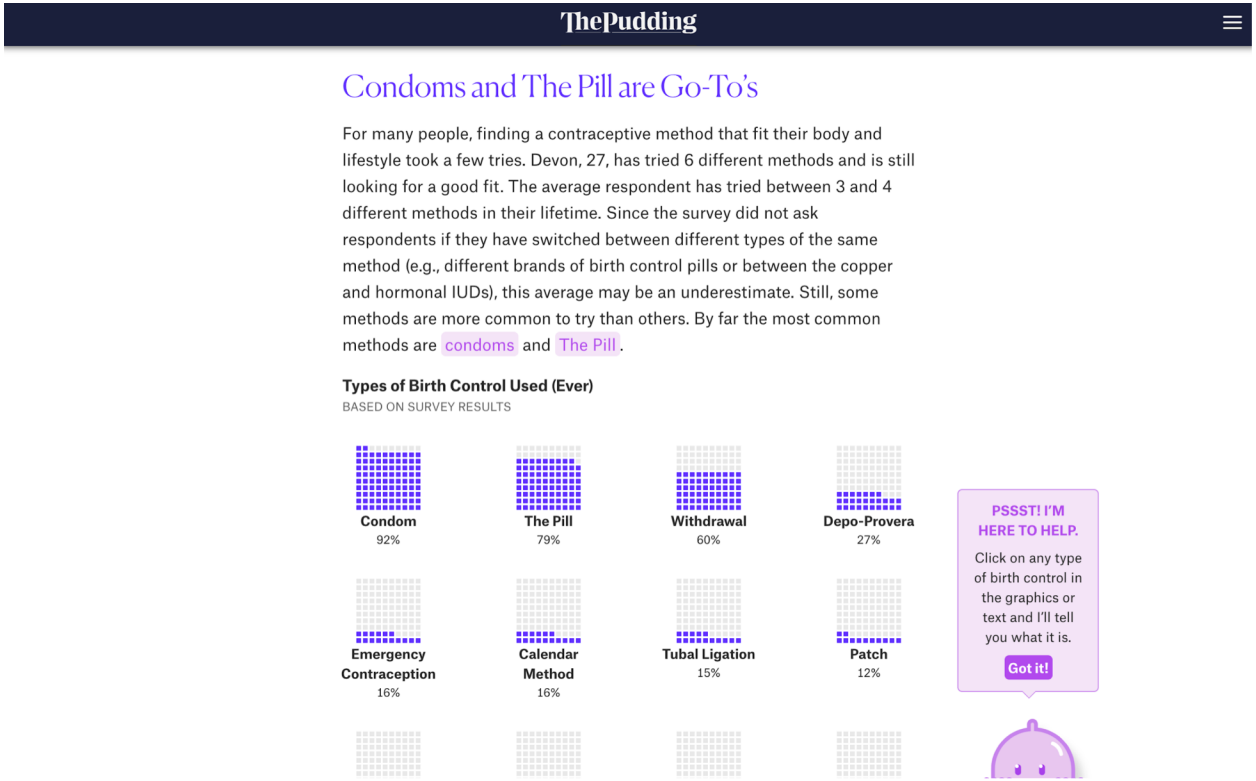
Below we'll put you through a series of tests. You can skip ahead if you just want to see the results, but be warned, once you do there is no turning back!

MODULE 6 | VIDEO 4

Have fun!

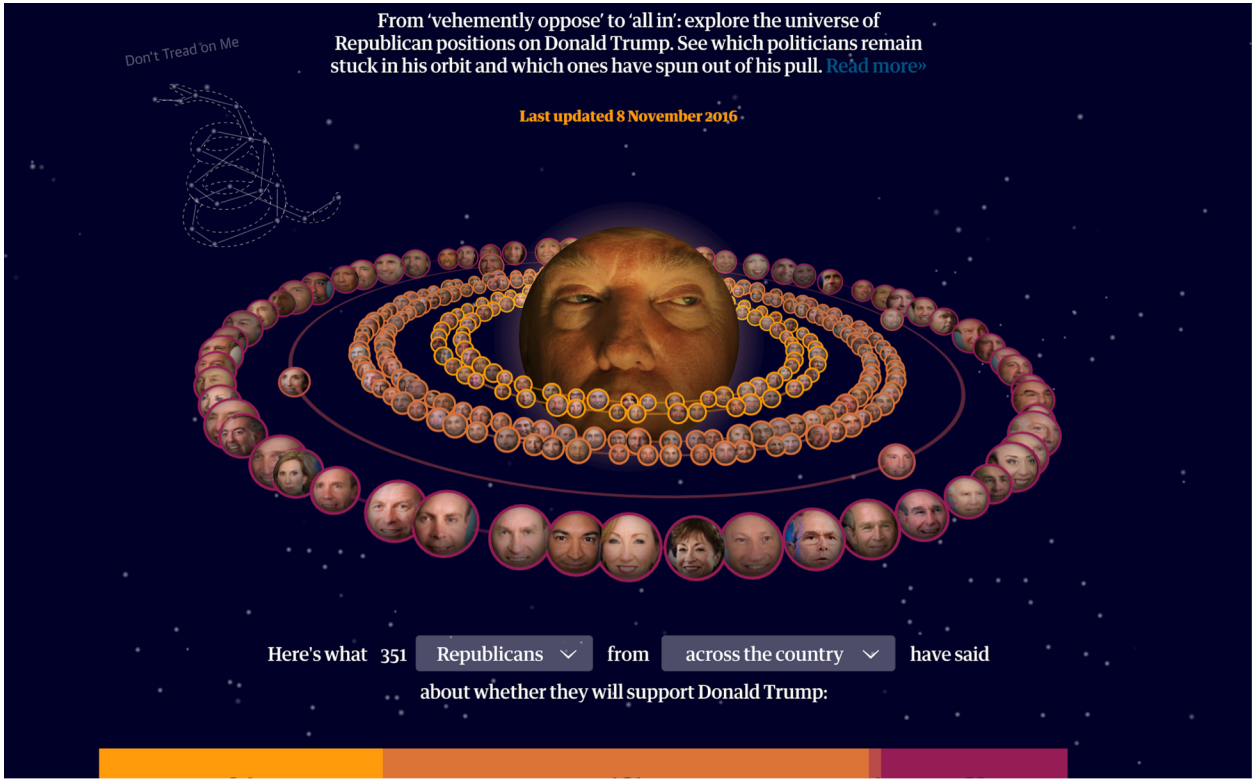
play-by-play data reveals that defensive three seconds is the NBA's least consistently called rule.

Lance Stevenson blowing data points

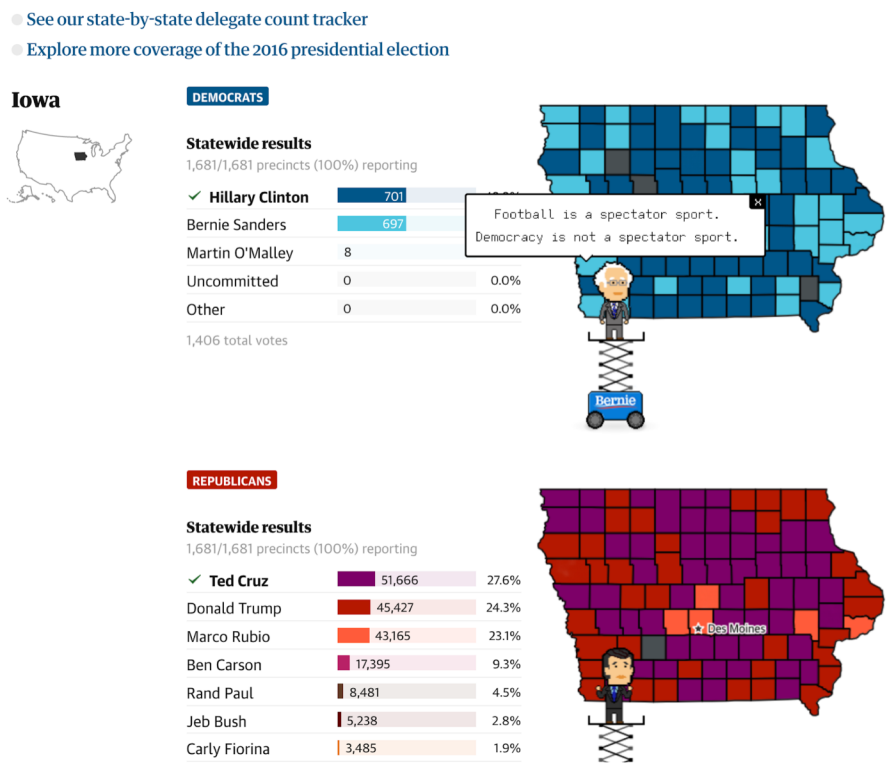


Condom clippy assistance

Progressive scrolling sad emoji



Republican themed constellations



8-bit painting candidates

MODULE 6 | VIDEO 5

Data stories

with The Pudding

Pockets

- Starts with a question
- Big to small narrative
- Scrollytelling
- No traditional charts
- Whimsical design elements

Hyphenated last names

- Starts with a question
- Big to small narrative
- No traditional charts
- Animation

Colorism in Vogue

- Starts with a question
- Scrollytelling (x2)
- Animation

MODULE 6 | VIDEO 5

Why is experimentation important?

- If no one broke the rules, data viz would be boring
- Keeps the readers (and us!) interested
- Doesn’t limit the storytelling by the delivery
- Experiments add to the whole industry’s toolkit

MODULE 6 | VIDEO 5

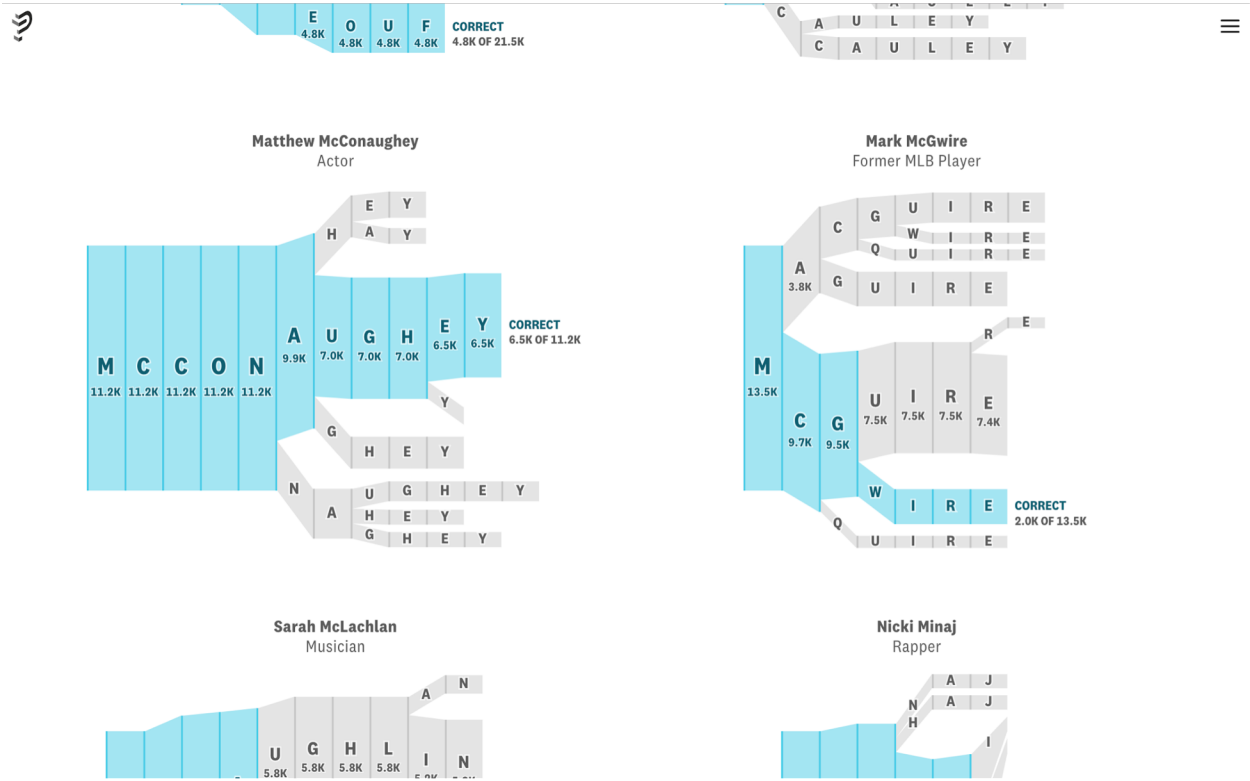
How do you know when to break the rules?

👍 When to break

- There’s a better way to tell the story
- If the story is more light hearted
- For a connection with the reader

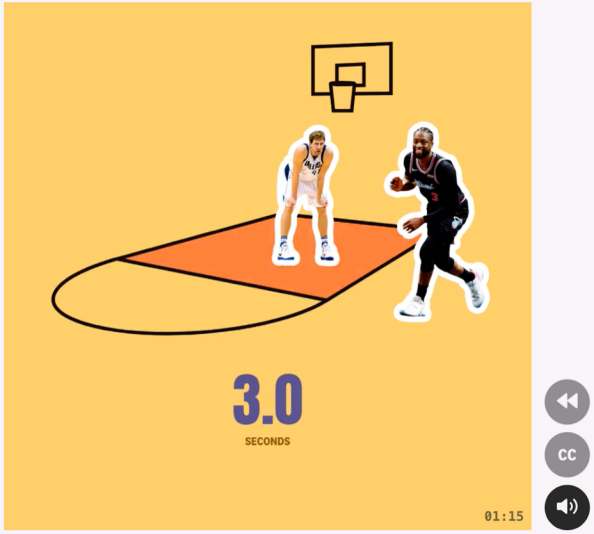
👎 When to follow

- You’re at risk of deceiving the reader
- There are already heavily ingrained patterns



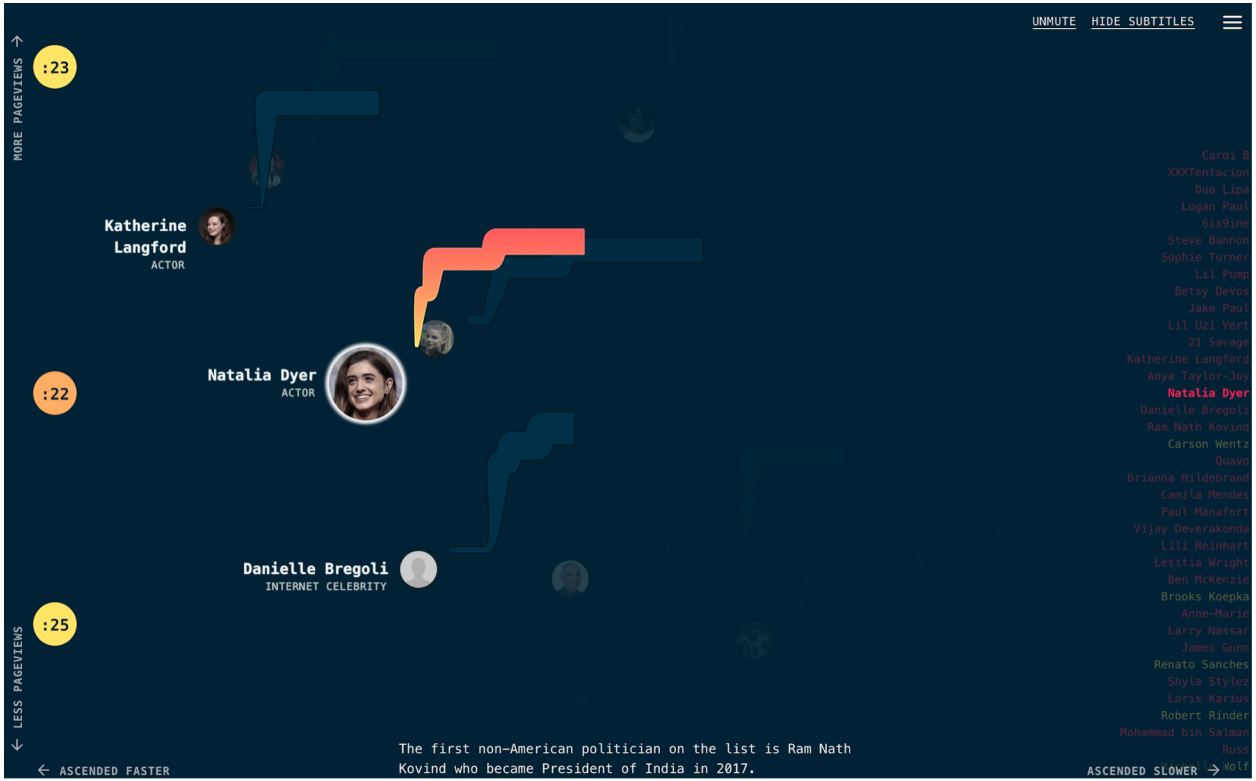
Building sankey-ish diagrams

play-by-play data reveals that defensive three seconds is the NBA's least consistently called rule.

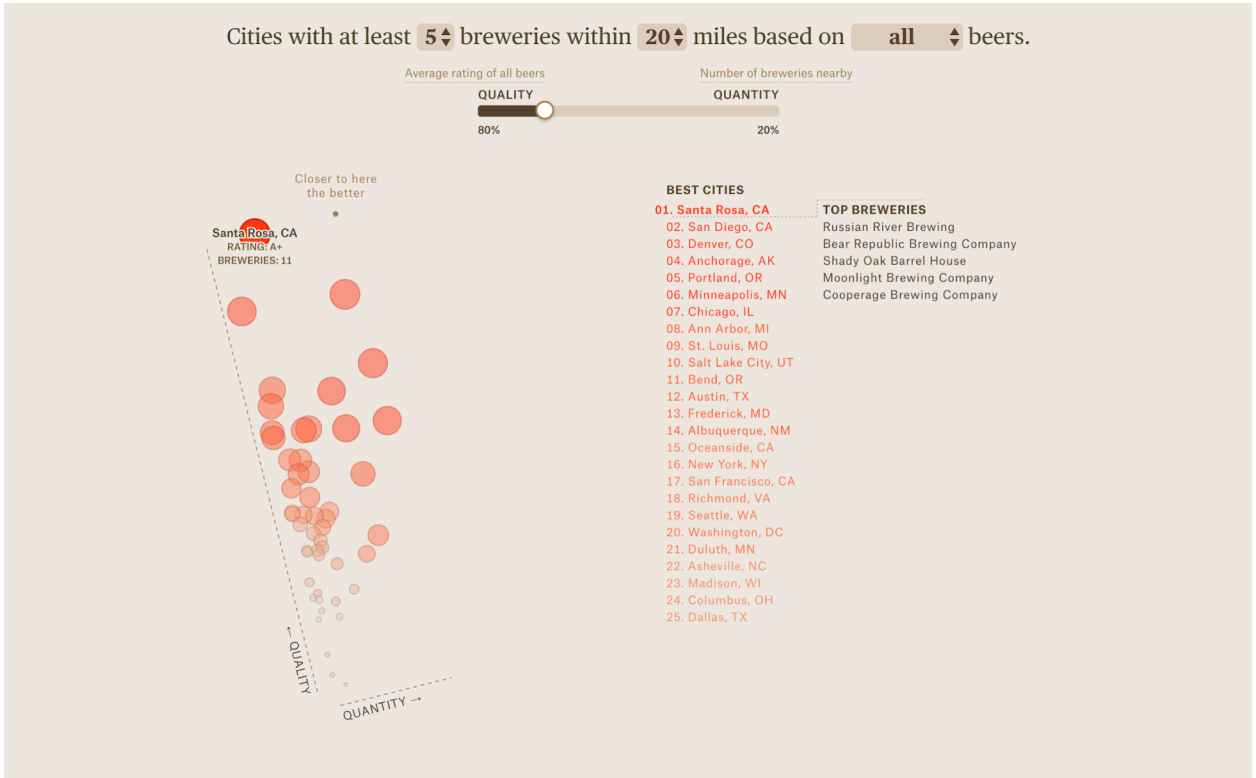


Rendering code to video and building an off-platform experience

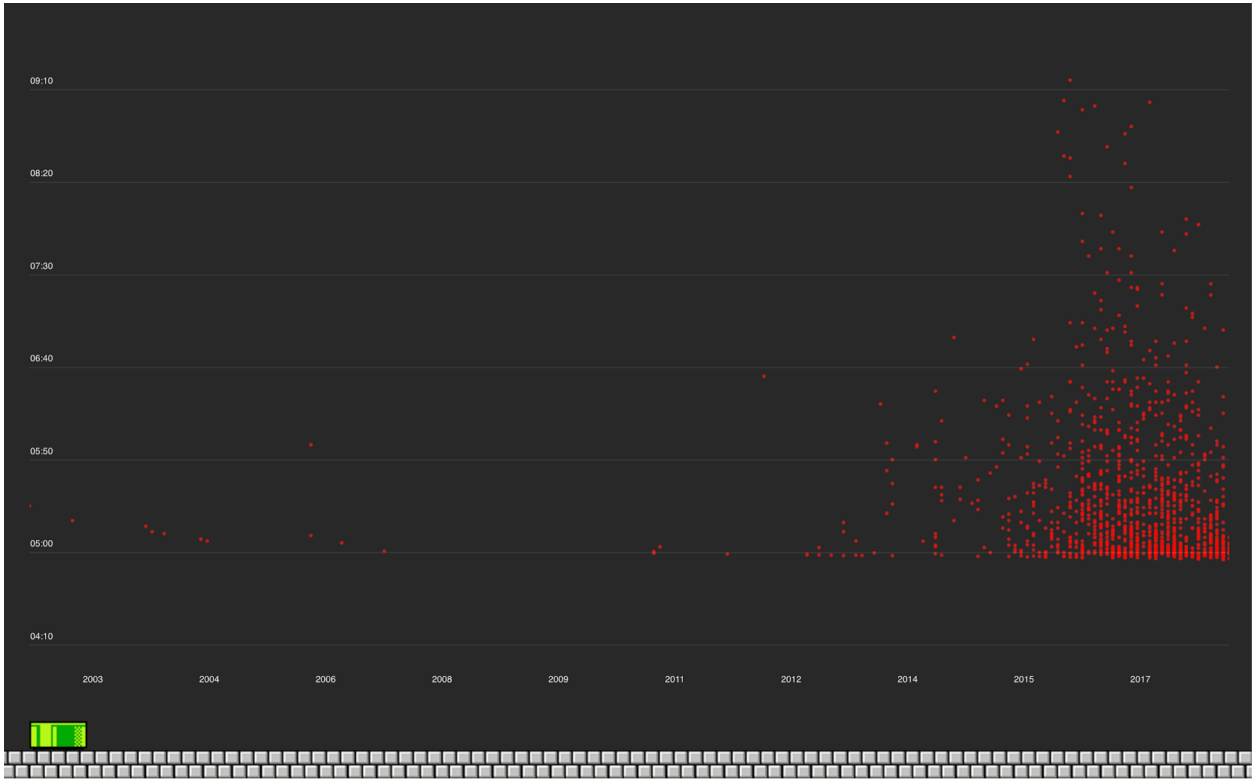
MODULE 6 | VIDEO 5



Using the entire page as a chart, variable thickness band charts, and audio narration triggers



Pivoting scatterplot


09:10
08:20
07:30
06:40
05:50
05:00
04:10
2003
2004
2006
2008
2009
2011
2012
2014
2015
2017

Speedruns

Political fonts

Florida man

Florida man arrested, accused of shoving woman to get to egg rolls
Florida man dons Spider-Man mask during liquor store burglary
Naked Florida man starts house fire after baking cookies on George Foreman grill
Florida man breaks into store, flips off security camera, deputies say
Throwing food, sex for chicken nuggets and other area Florida man and woman arrests
Thrown pancake batter leads to battery charges for Florida man
Oh, Florida Man. Playing policeman never ends well for you
Florida man charged in naked rake attack
Hot sauce saves Florida man after car crashes into Taco Bell
Shirtless Florida man travels to Myrtle Beach to head bang during Hurricane Florence
Florida Man claiming people were "eating his brains" leads police on insane golf course chase
Florida man who attacked McDonald's worker over straw sentenced to jail
Predictably, Florida Man Blows Off Part of His Hand in July 4th Fireworks Accident
FL man blames dog for drunk driving stop
Florida man arrested for throwing alligator through drive-thru window
Police: Florida man wrecks liquor shop, blames "caterpillar"
Florida man makes beer run with alligator in hand, wildlife officials investigating
Florida man caught trying to steal ice cream truck

That’s a wrap!

- Don’t forget to participate in the discussion forums and quizzes
- Also be sure to check out the Facebook group for this course
- If you have any questions for me post them in the “Questions for the instructors forum”
- Post questions related to the course platform in the “Questions about the course platform”

www.pudding.cool

- Twitter: @puddingviz
- Instagram: the.pudding