

First, set up an account

Datawrapper.de

Then, go to:
[github.com/
paulbradshaw/viz](https://github.com/paulbradshaw/viz)

AREA
ENSION

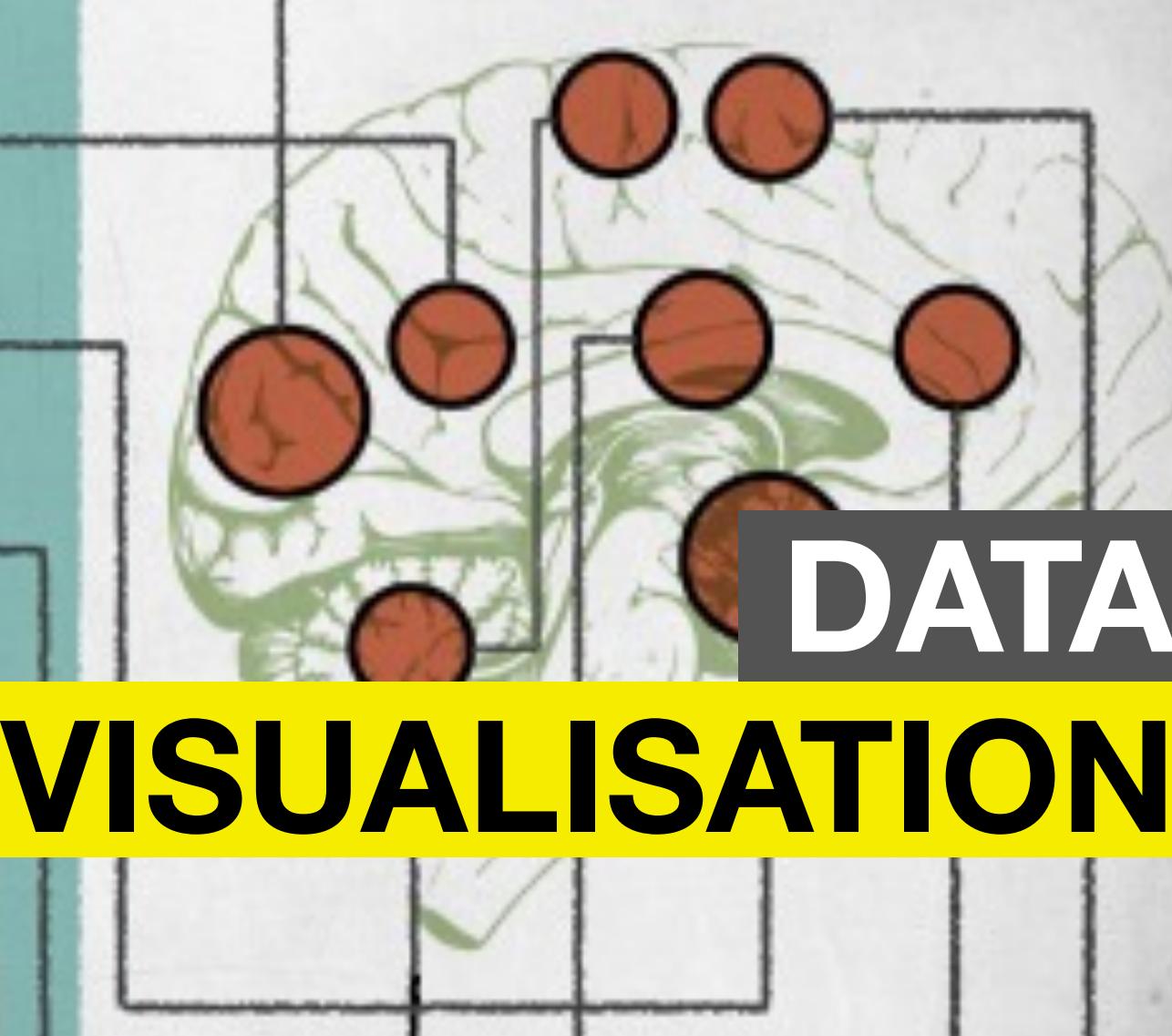
ORTEX

RTEX

A
ING

EX

TIEY



DATA VISUALISATION

@PaulBradshaw, Online Journalism Blog, Birmingham City University
Author: Scraping for Journalists, Finding Stories in Spreadsheets, Data
Journalism Heist, Online Journalism Handbook

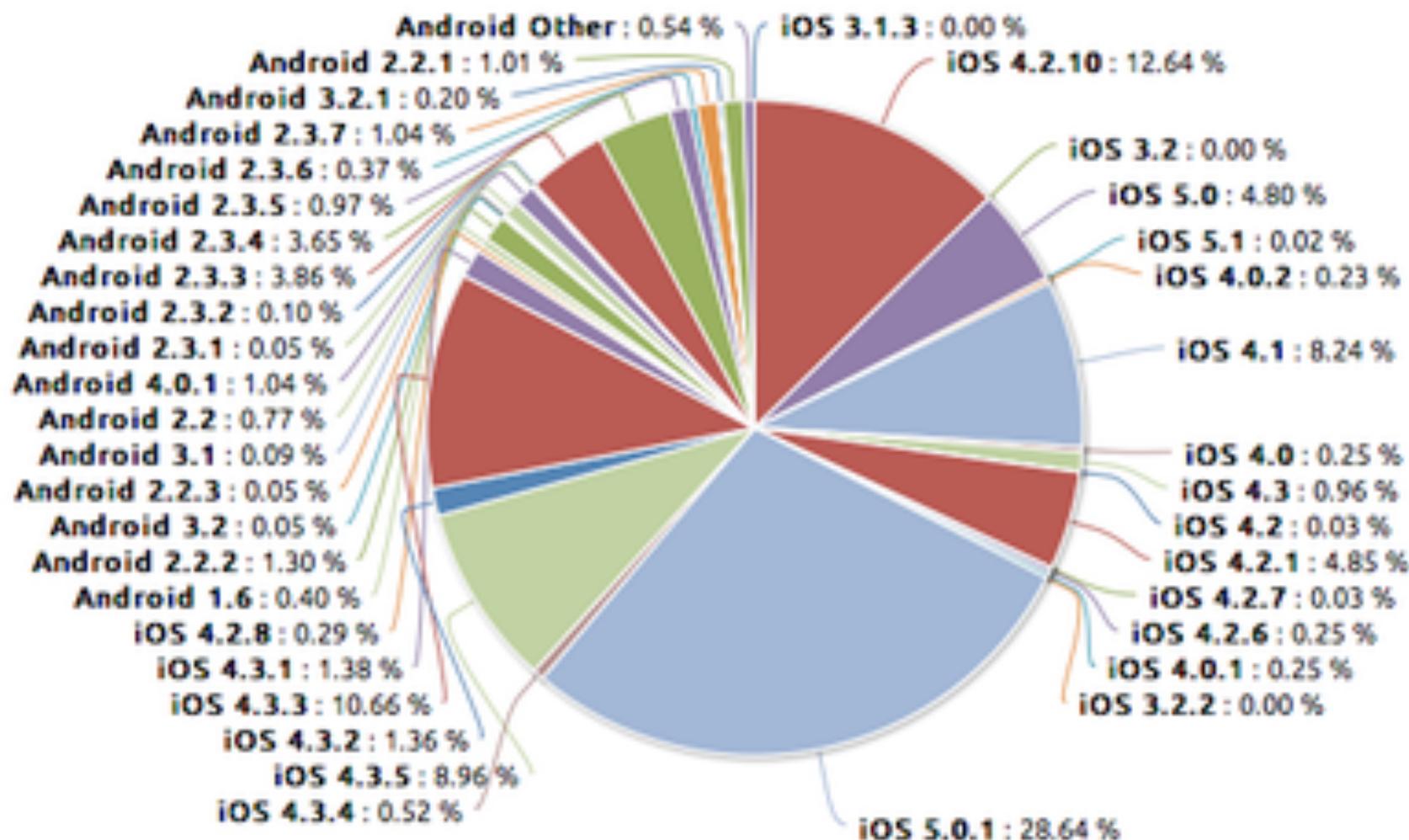
How and what

How to choose the right chart

Colour, tips (and why you
should never use a 3D pie
chart)

What tools to use to tell
stories visually

Crashes by OS Version Normalized (12/1 - 12/15)



<http://www.bgr.com/2012/02/03/ios-apps-crash-more-than-android-apps-study-shows/>

DILMA (2015)

10,67%

14%

38 MIL PONTOS

533 PONTOS

-3,8%

INFLAÇÃO

JUROS

IBOVESPA

RISCO-PAÍS

PIB

BOLSONARO (2019)

3%

5%

108 MIL PONTOS

117 PONTOS

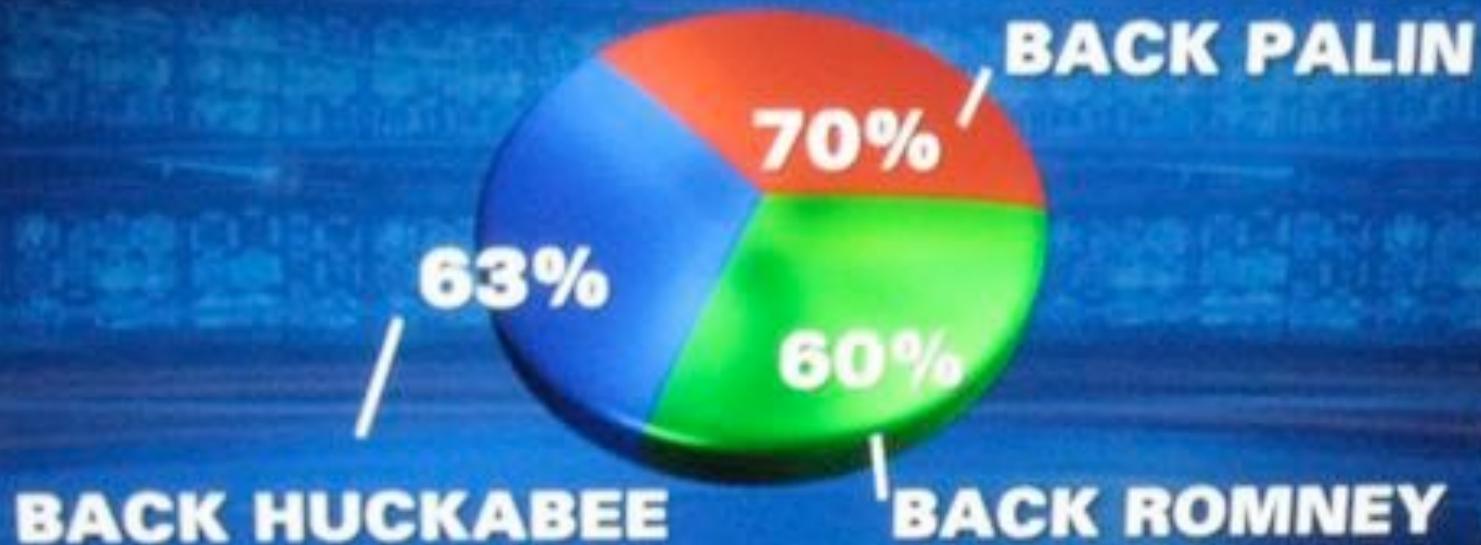
+0,8%

Fonte: IBGE / Banco Central / CDS



2012 PRESIDENTIAL RUN

GOP CANDIDATES



SOURCE: OPINIONS

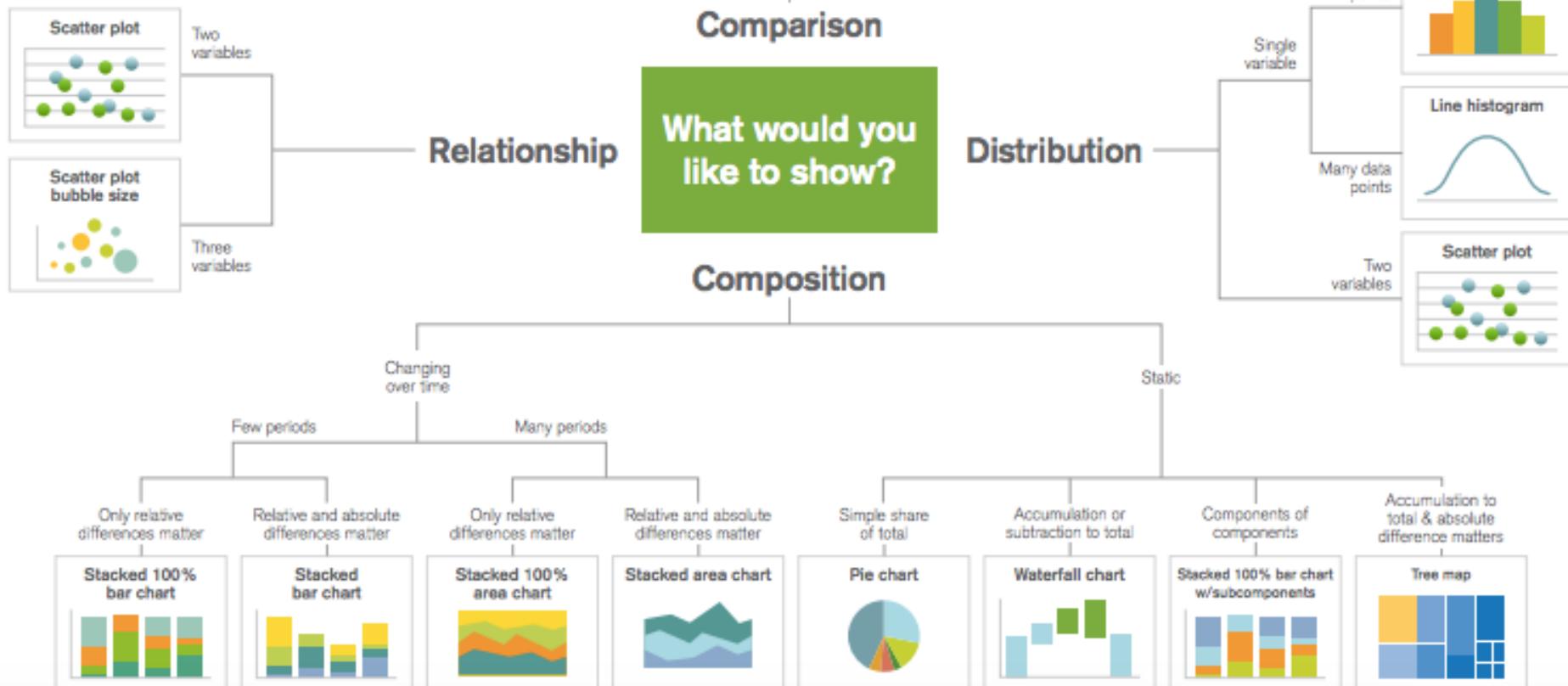
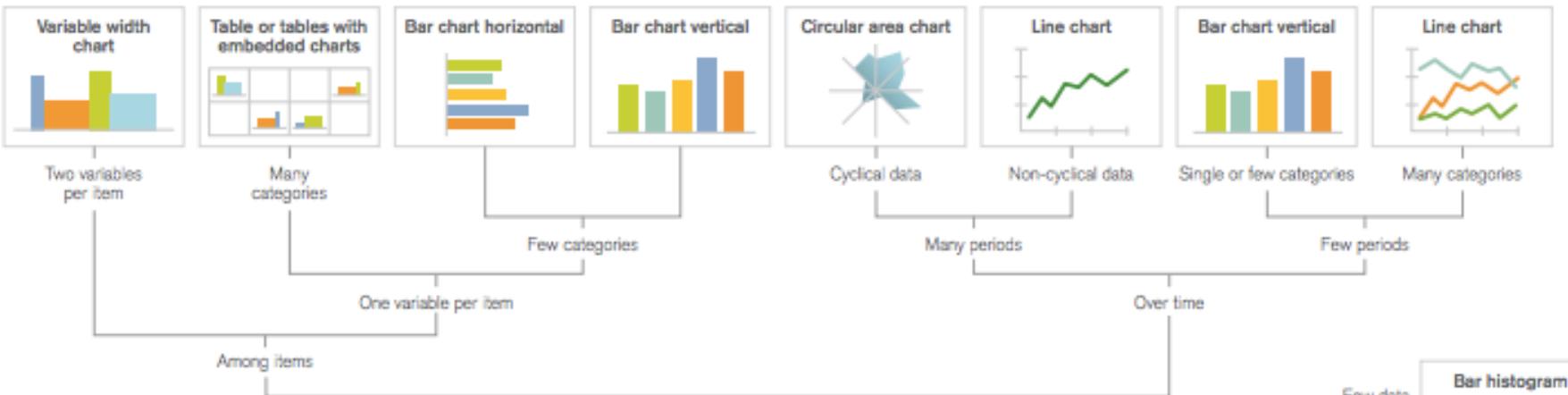
DYNAMIC

FOX

47°

“We find that providing participants with graphical information **significantly decreases false and unsupported factual beliefs.**” Crucially, they show that data presented in graphs and illustrations does a better job of fighting misperceptions than the same information presented in text form.

What is your story?



or <http://chartchooser.juiceanalytics.com/>

Deviation

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

Example FT uses
Trade surplus/deficit, climate change

Correlation

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

Example FT uses
Inflation & unemployment, income & life expectancy

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

Distribution

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

Example FT uses
Income distribution, population
(age/gender) distribution

Change over Time

Give emphasis to changing trends. These can be short (intra-day) movements or extended series (over months or years). Choosing the correct time period is key to providing suitable context for the reader.

Example FT uses
Share price movements, economic time series

Magnitude

Show size comparisons. These can be relative (just being able to see larger/bigger) or absolute (need to see the raw numbers). It's good to have a 'counted' number (for example, barrels, customers etc.) rather than a calculated rate or per cent.

Example FT uses
Commodity production, market capitalisation

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 using magnitude-type charts instead.

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

Spatial

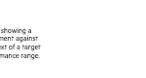
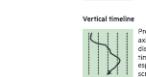
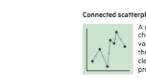
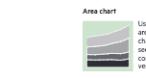
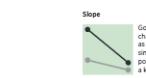
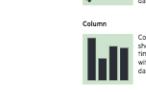
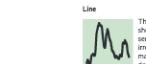
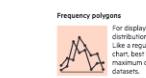
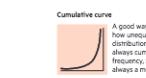
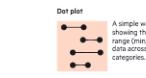
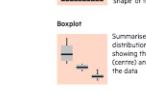
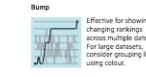
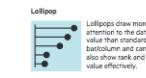
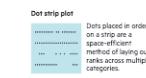
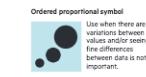
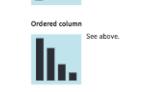
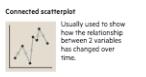
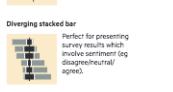
A aside from location or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations.

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

Flow

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

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



Visual vocabulary

Designing with data

There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

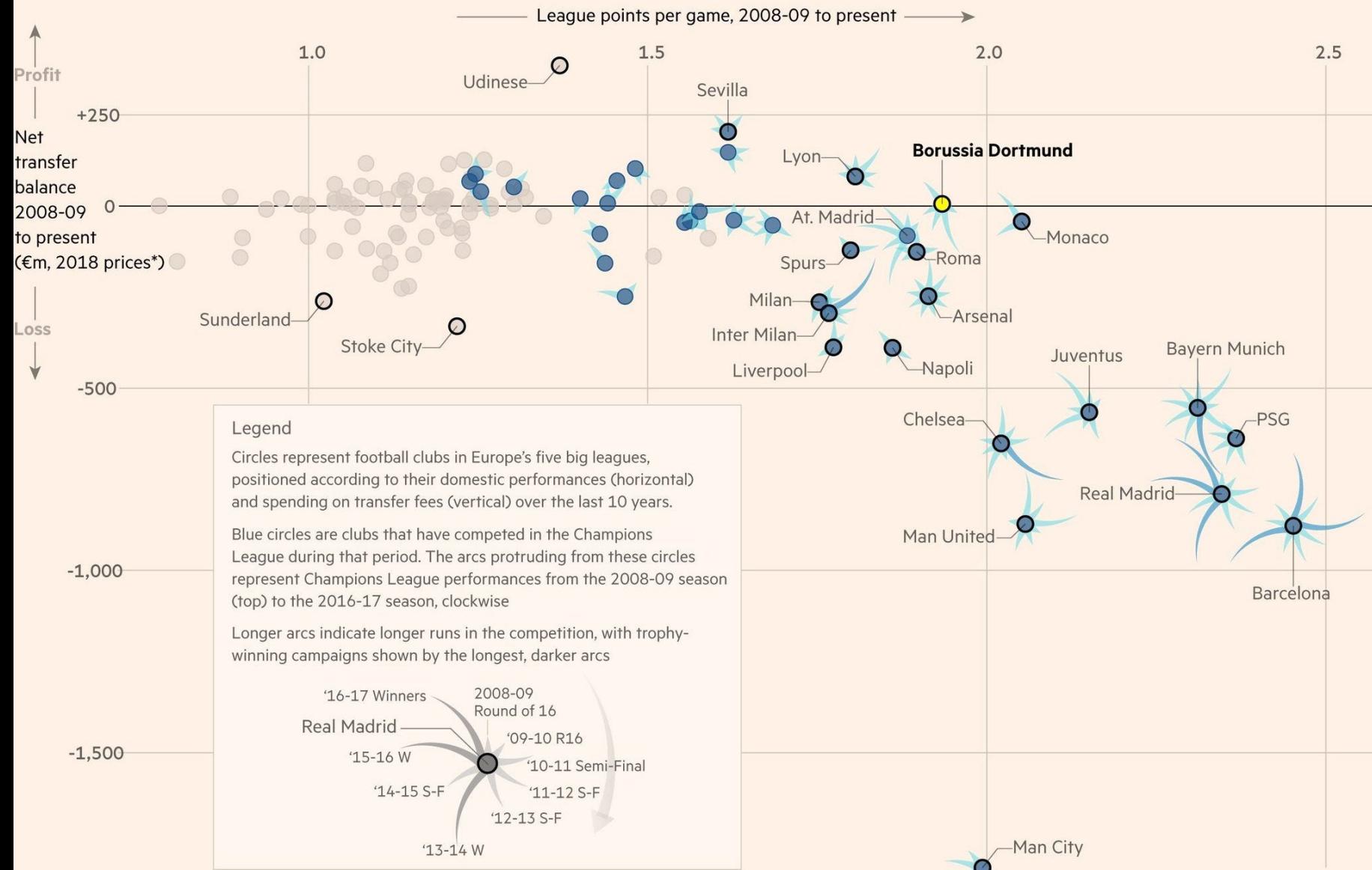
ft.com/vocabulary

FT graphic: Alan Smith; Chris Campbell; Ben Bell; Li Huaiwu; Graham French; Billy Oberberg; Paul McCullagh; Martin Sibley
Inspired by the Graphic Contrarium by Jon Schmidbauer and Svenneth Blawie

FT

Is it an **explanatory**
story — or an
exploratory story?

Dortmund's policy of buying younger, cheaper players — often selling them on for large profits — means they achieve on-pitch success at much lower net outlay than other big clubs



*Adjusted for football transfer inflation based on growth in club revenues

Source: Transfermarkt, Deloitte Football Money League

Analysis and graphic: John Burn-Murdoch / @jburnmurdoch

© FT

Tools... vs CMS? Social?

Datawrapper.de

Infogr.am

Flourish

Piktochart

Plot.ly

Highcharts

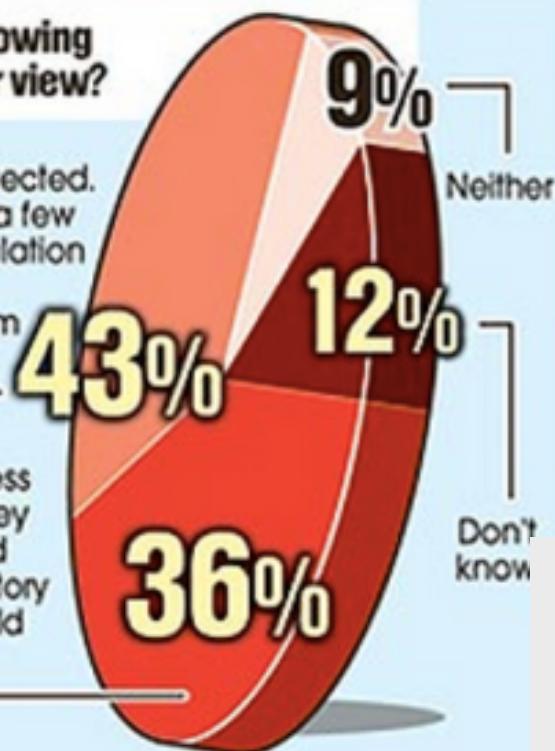


Xaquín G.V. ✅
@xocasgv

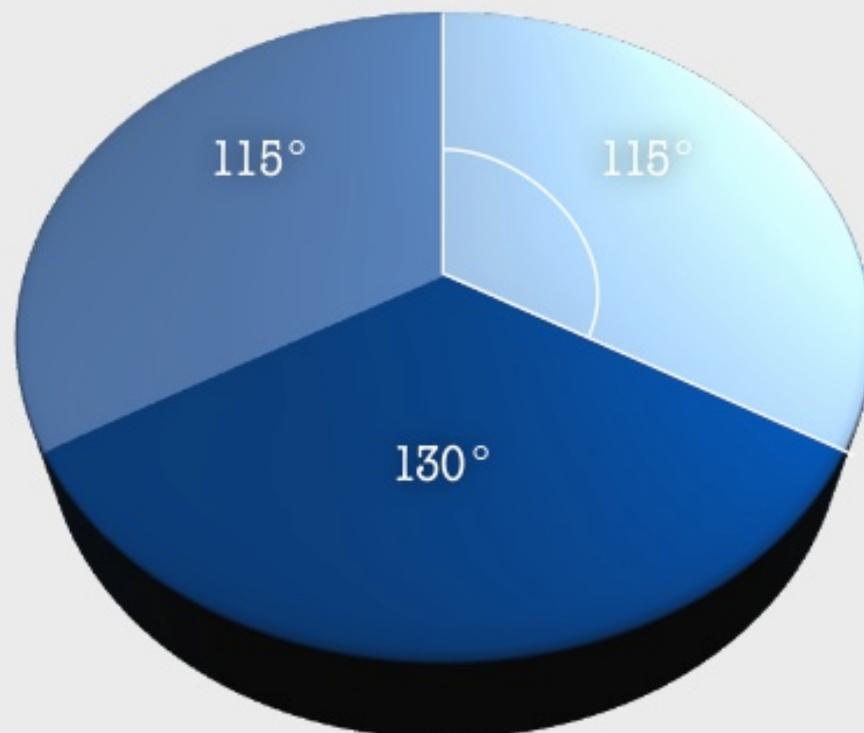
I had to do it #DistractedBoyfriend

Which of the following best reflects your view?

A free Press must be protected. Whatever the failings of a few journalists, statutory regulation by politicians would risk damaging press freedom



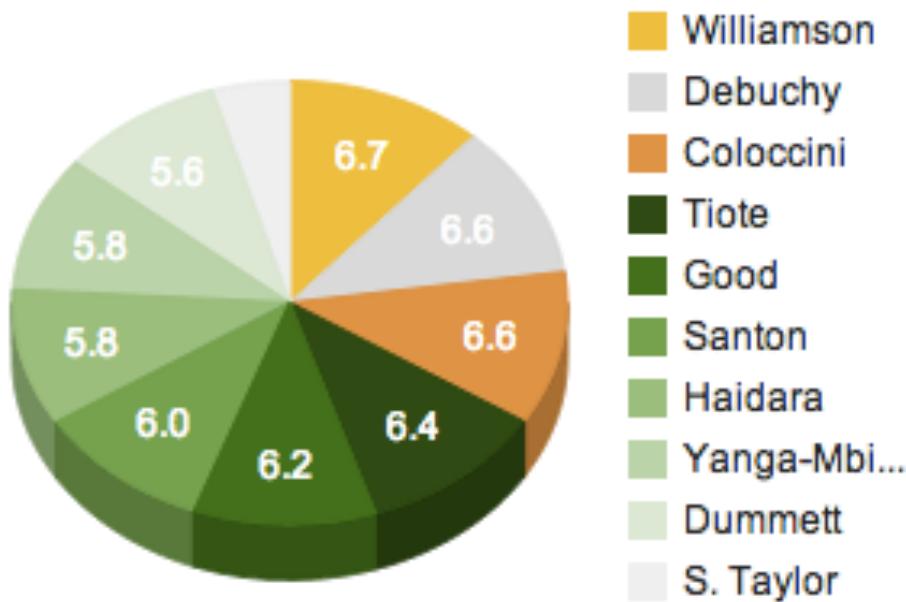
The behaviour of our press has gone too far and they can no longer be trusted to set up their own regulatory system. Parliament should introduce proper legal regulation



Newcastle United player stats: The season so far

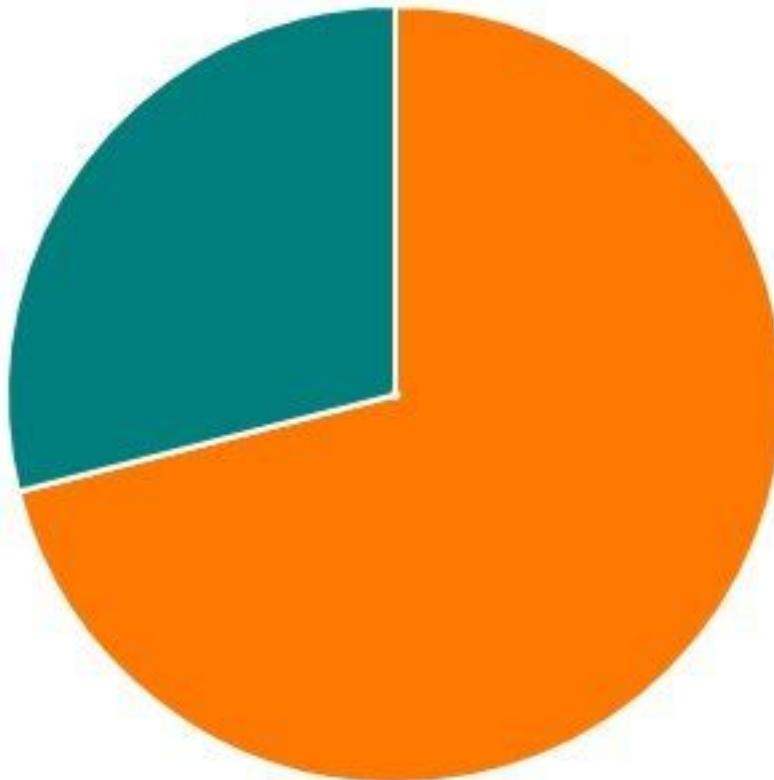
See how each of the Newcastle United FC players stack up against each other, as rated by fans

Defenders



Compensation for historical NHS failings (2012-17)

Maternity payouts
All other payouts



Source: NHS Resolution

BBC

Data structure for pie:

Category	Number
One part	481.0
Another part	551.1
A third part	527.4

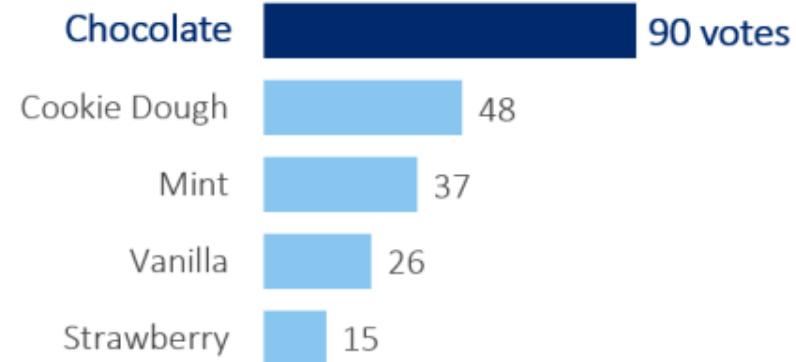
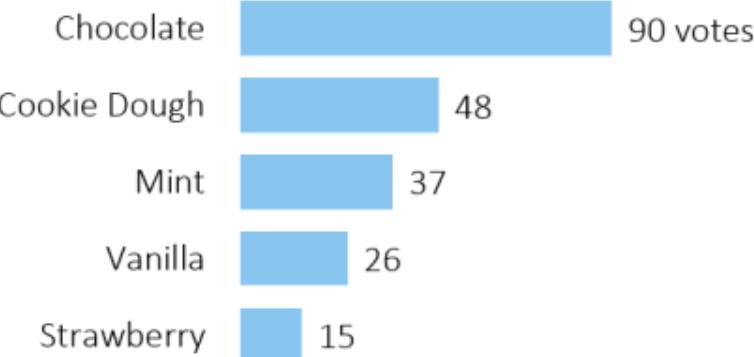
Category	Percent
One part	0.31
Another part	0.35
A third part	0.34

Colour.

Ice cream flavor preferences based on
2014 survey of elementary school
students (n=216)

or

Chocolate was most popular flavor
among elementary students surveyed



Source: 2014 survey of elementary school
students (n=216)

Roberto Rodriguez - Adwords Account

0.71%

Click-through rate (CTR)



Facebook Benchmarking #1

Total likes

COCA-COLA	67.3M	FANTA	9.78M
REDBULL	38.5M	MOUNTAINDEW	7.98M
PEPSI	17M	BURNENERGY	1.67M
DRIPEPPER	15.4M	SCHWEPPES	2.61K
SPRITE	14.5M		

Twitter Benchmarking #1

Mentions

COCACOLA	32.1K	BURN	1,767
REDBULL	20.2K	SPRITE	1,039
PEPSI	9,927	PEPSIMAX	705
DRIPEPPER	3,924	FANTAFUN	85
MTN_DEW	2,018		

Roberto Rodriguez - Adwords Account

\$0.11

Cost-per-click (CPC)



Welovvrol - Fan Page

8

New post published
-66.67%

57

People talking ab...
-56.49%

Welovvrol - Twitter

0.5533%

Engagement rate
-51.6598%

Roberto Rodriguez - Adwords Account

\$0.79

Cost-per-thousand impressions (CPM)



Welovvrol - Fan Page

7,544

Reach
-44.73%

4,620

Reach(viral)
-29.82%

Welovvrol - Twitter

109

Updates timeline
-0.91%

150

Mentions
+7.1%

Roberto Rodriguez - Adwords Account

Devices

COMPUTERS : 1



Audience Social Media #1

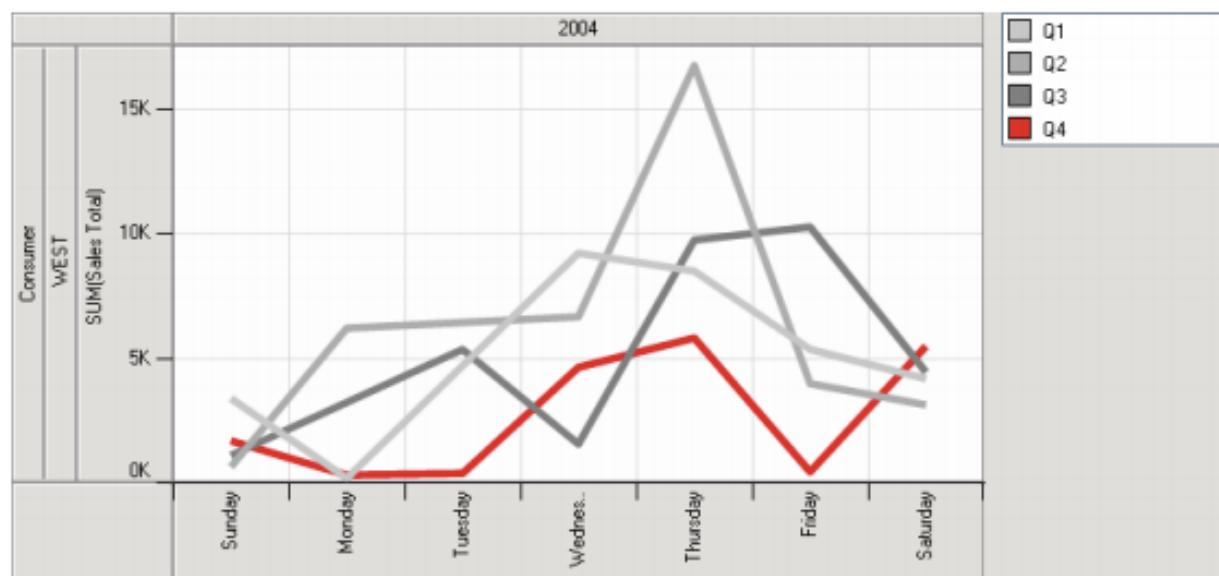
1.834

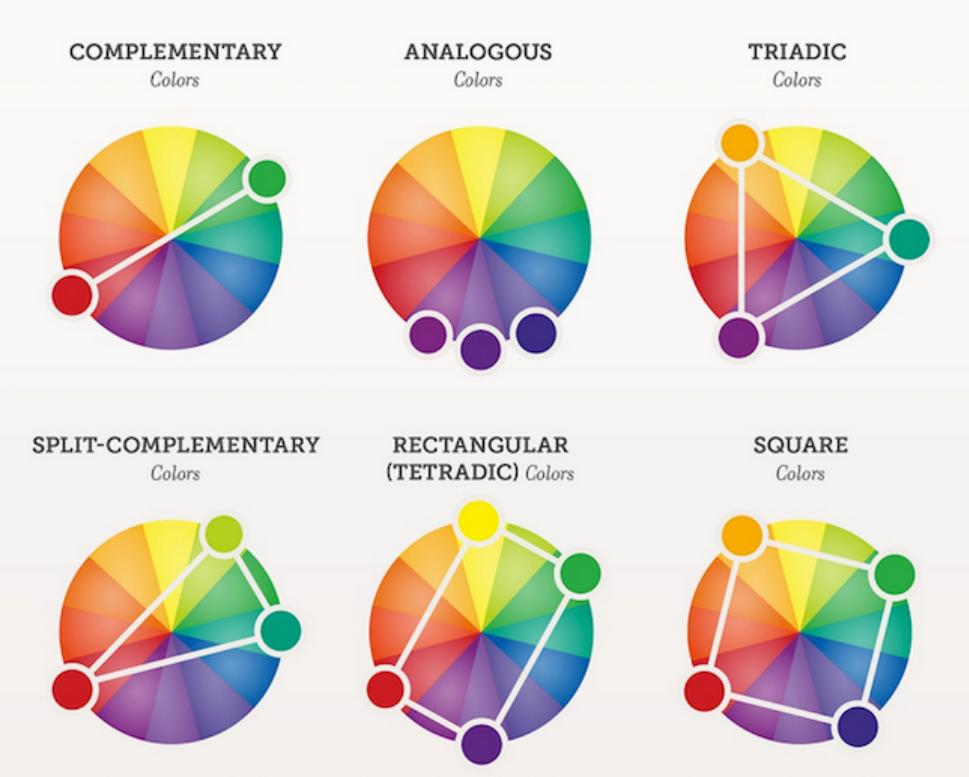
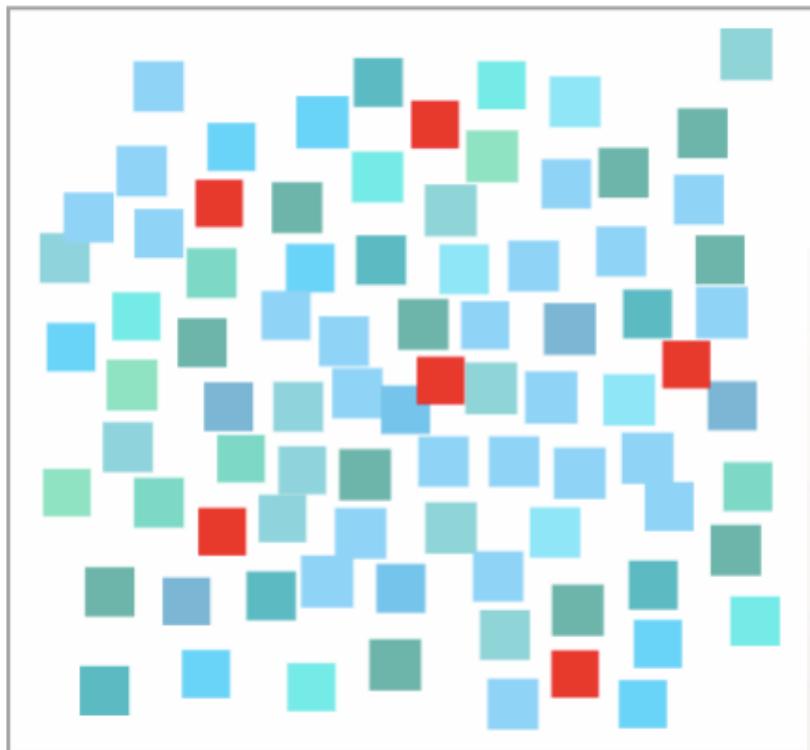
Audien

Seomoz Benchmarking #1 ⓘ

Links

WWW.PEPSI.COM	599K	WWW.DRPEPPER.COM	40.8K
WWW.COCA-COLA.COM	362K	BURN-ENERGY.TUMBLR...	4,323





number of data classes on your map

3 [learn more >](#)

the nature of your data

sequential [learn more >](#)

theme: BuGn

pick

multihue single hue

(optional) only show schemes that are

colorblind safe print friendly

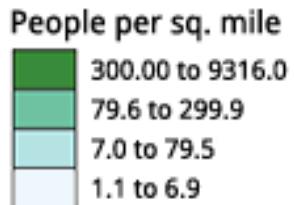
photocopy-able [learn more >](#)

pick a color system

0xE5F5F9 RGB CMYK HEB
0x99D8C9 [adjust map context](#)

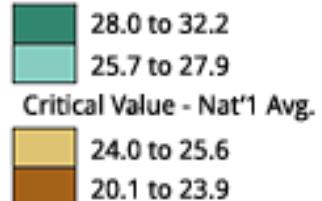


1. Sequential schemes are suited to ordered data that progress from low to high. Lightness steps dominate the look of these schemes, with light colors for low data values to dark colors for high data values.



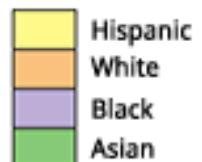
2. Diverging schemes put equal emphasis on mid-range critical values and extremes at both ends of the data range. The critical class or break in the middle of the legend is emphasized with light colors and low and high extremes are emphasized with dark colors that have contrasting hues.
[Learn more »](#)

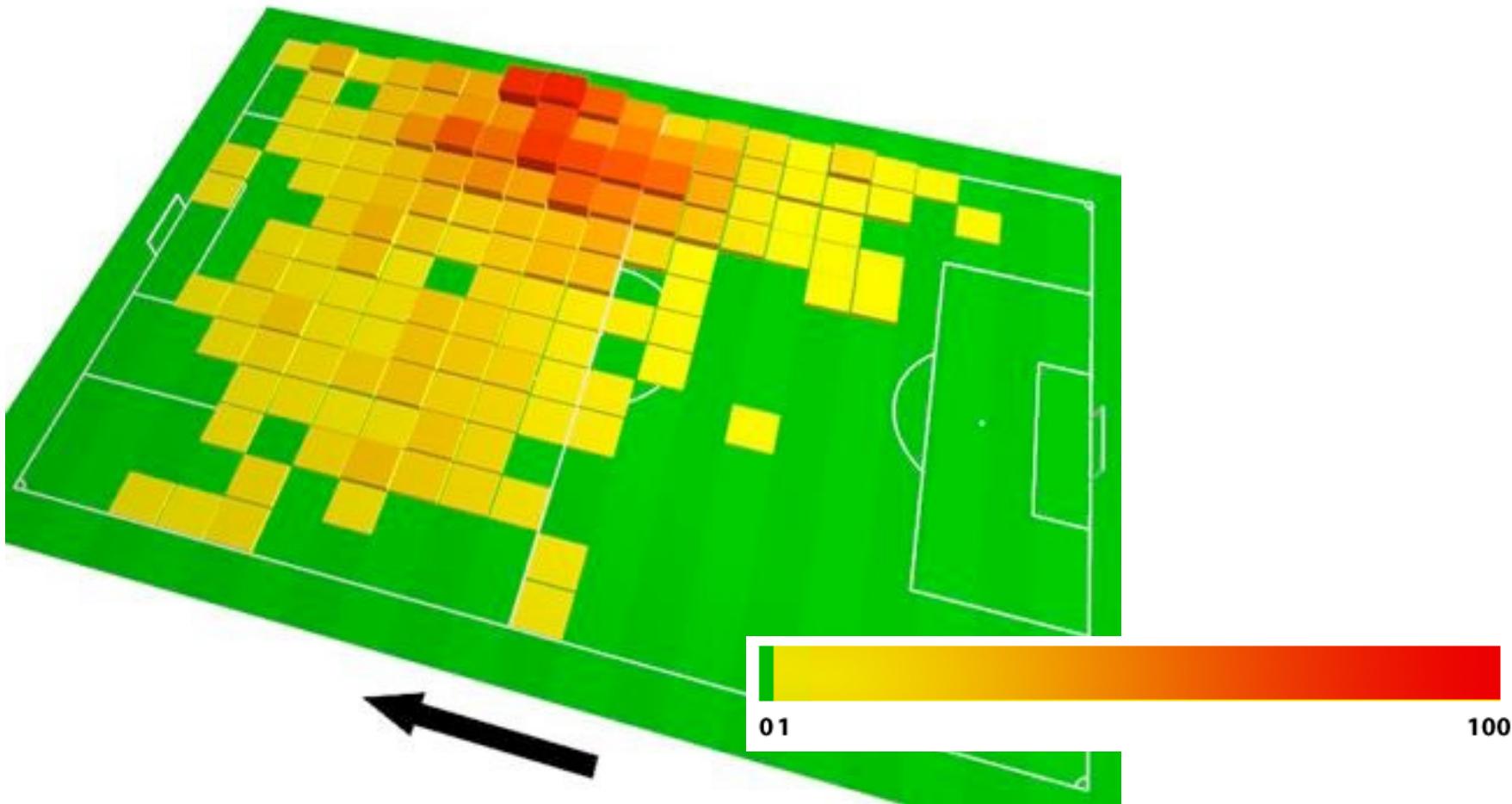
Percent of population under 18 by state



3. Qualitative schemes do not imply magnitude differences between legend classes, and hues are used to create the primary visual differences between classes. Qualitative schemes are best suited to representing nominal or categorical data.
[Learn more »](#)

Race or ethnicity



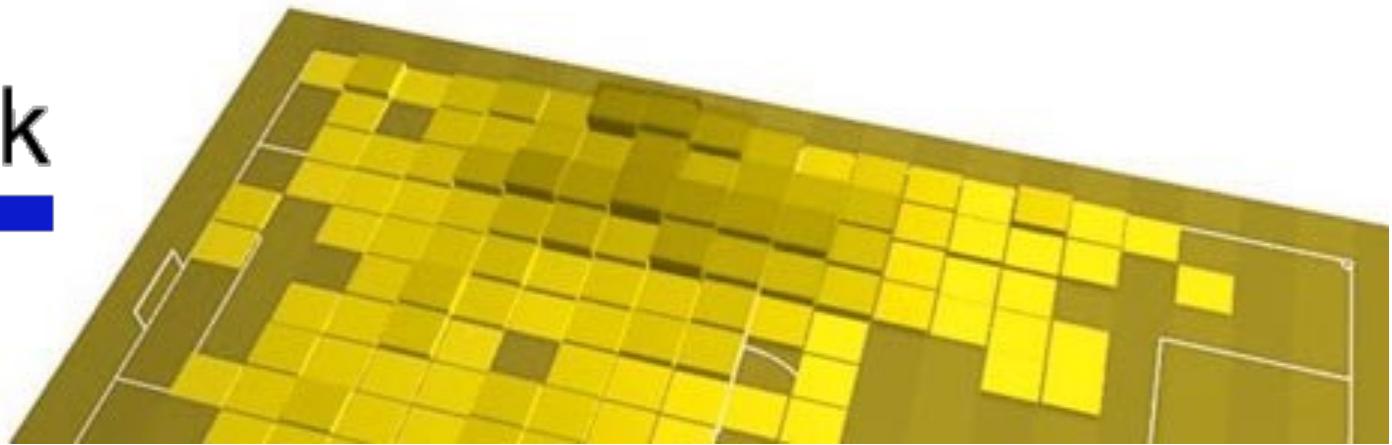




- Men who **can** read this chart
- Men who **cannot** read this chart

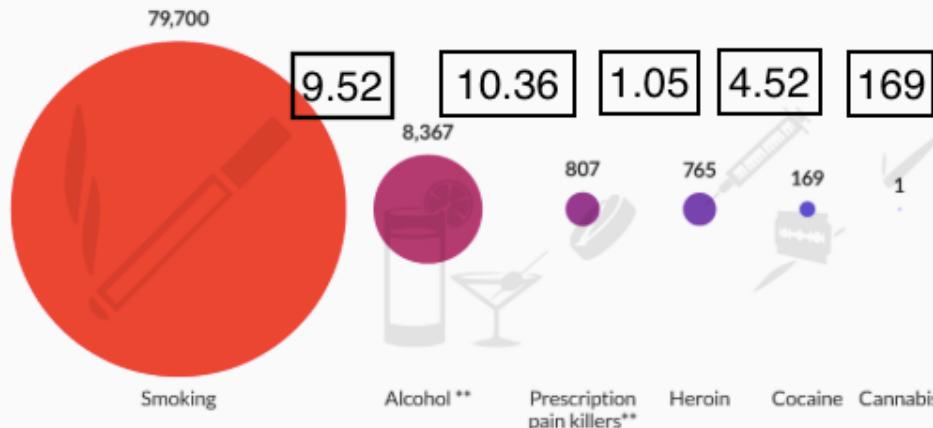


Vischeck



These drugs are all deadlier than cannabis

Annual drug-related deaths in England and Wales*



*latest available year

**UK - Alcohol-related deaths do not include indirect causes like traffic accidents and homicide

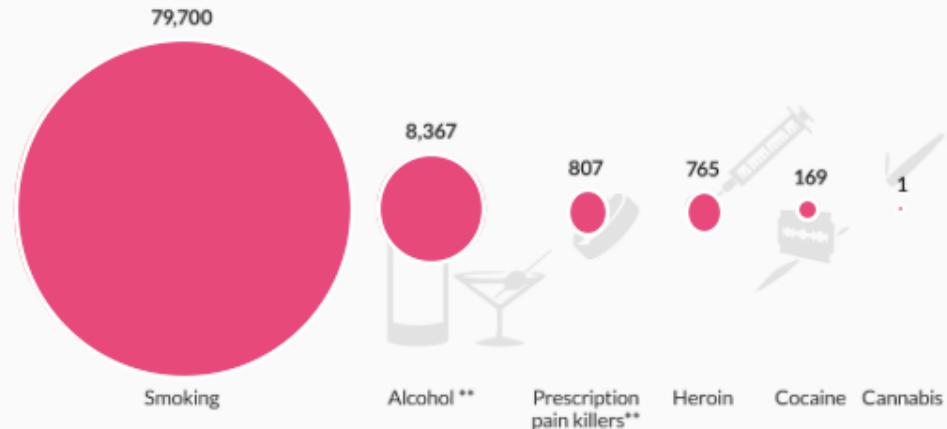


Sources: ONS, Ash, World Drug Report

i100 from The INDEPENDENT

These drugs are all deadlier than cannabis

Annual drug-related deaths in England and Wales*



*latest available year

**UK - Alcohol-related deaths do not include indirect causes like traffic accidents and homicide



Sources: ONS, Ash, World Drug Report

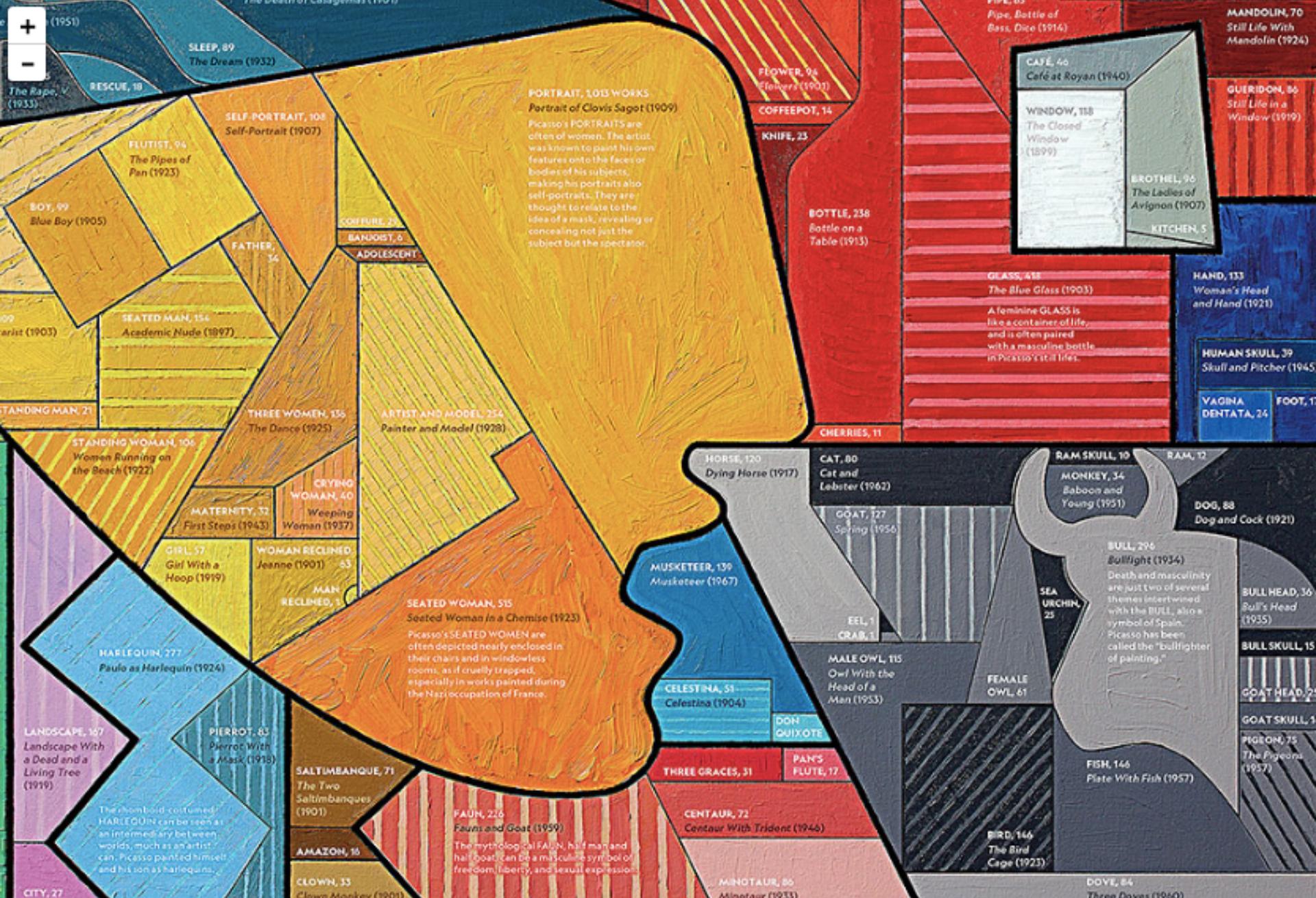
i100 from The INDEPENDENT

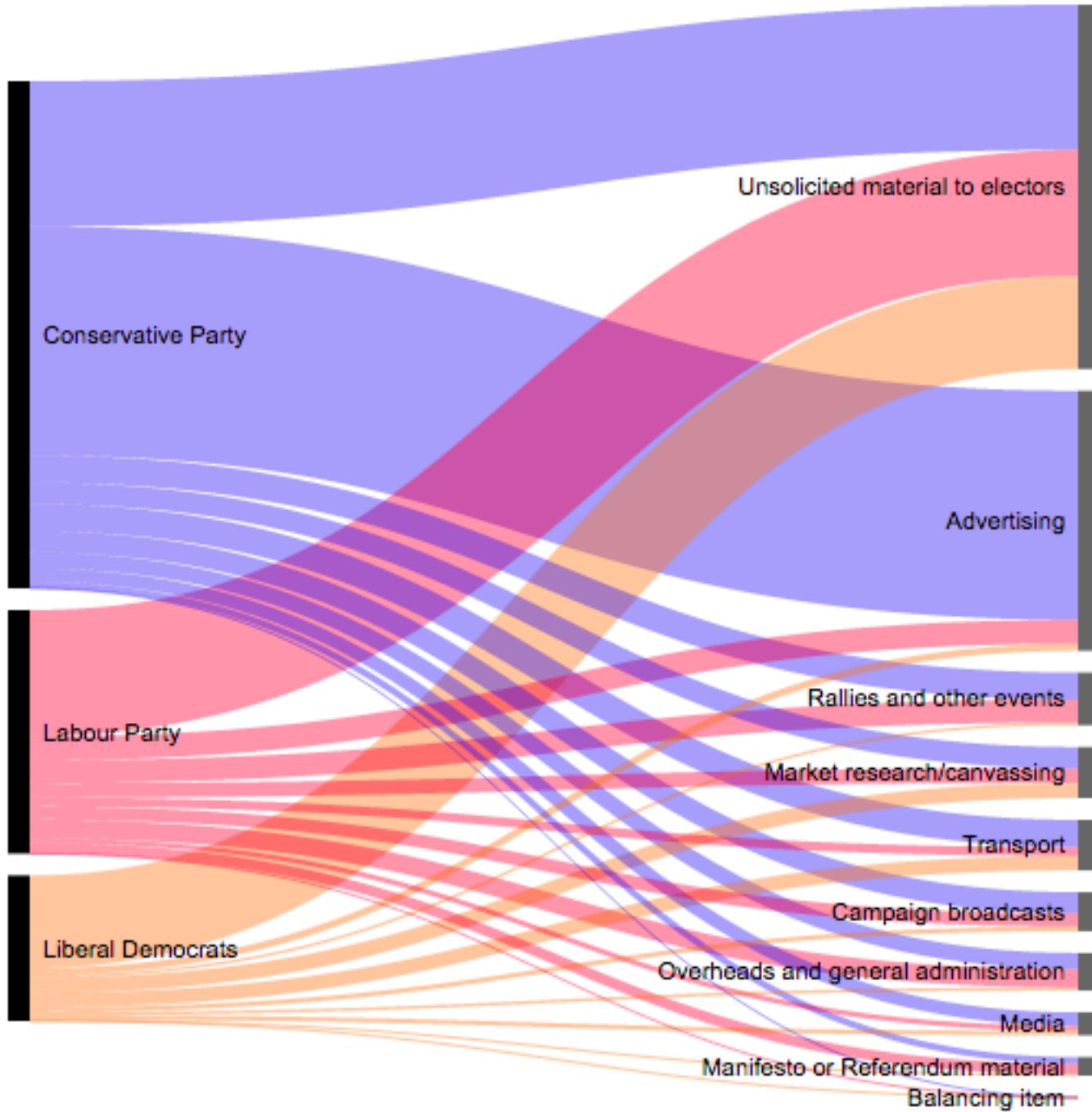
statista

More charts.

Distribution of seats by party and region, Great Britain, May 2015



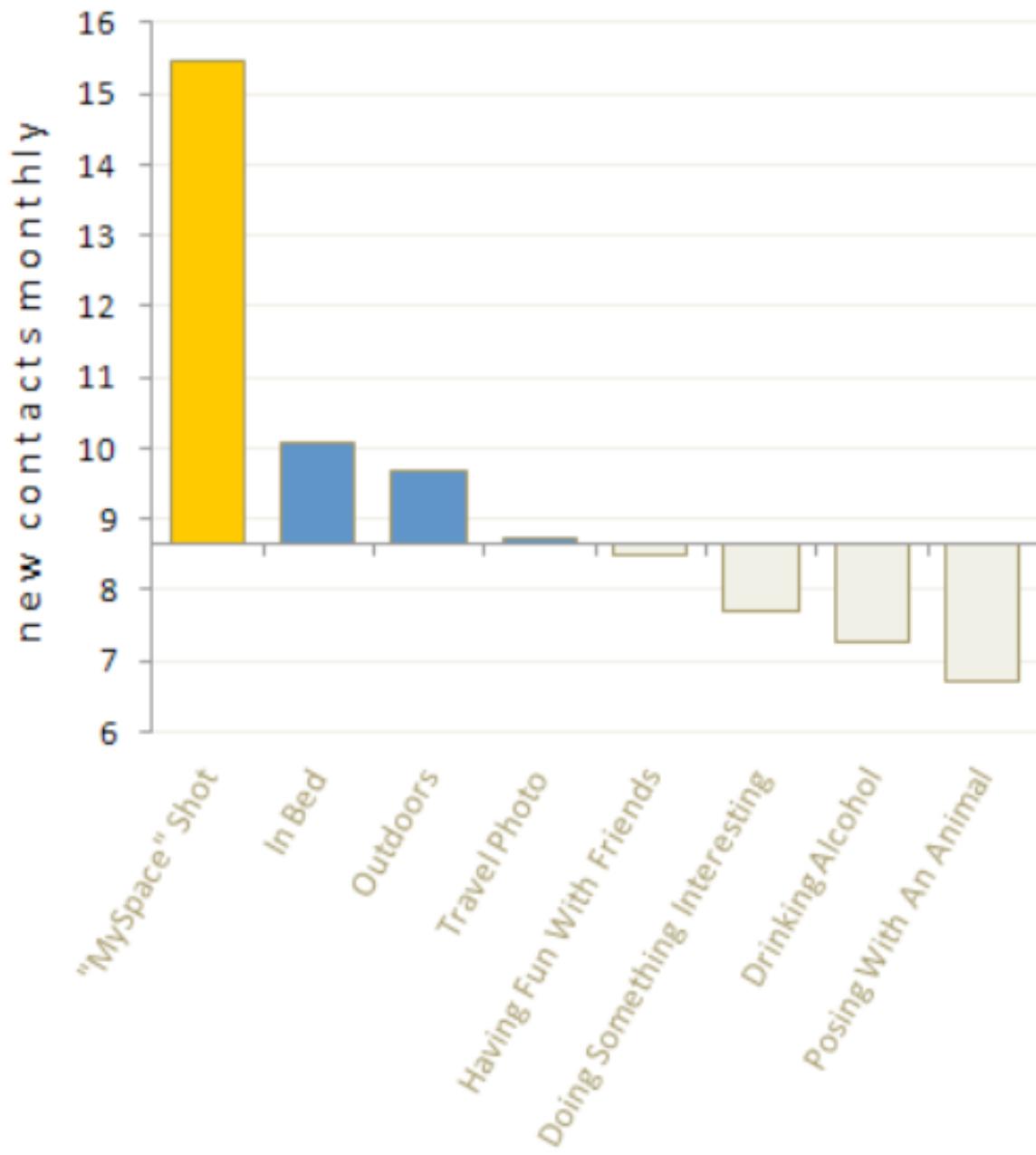




Data structure for treemap/sankey:

Top level category	Sub category	Number
Avon & Somerset	Part-time reserve	
Avon & Somerset	Neighbourhood officers	265.01
Avon & Somerset	Neighbourhood PCSOs	314.36
Bedfordshire	Part-time reserve	
Bedfordshire	Neighbourhood officers	241.75
Bedfordshire	Neighbourhood PCSOs	41.02
Cambridgeshire	Part-time reserve	
Cambridgeshire	Neighbourhood officers	509.18

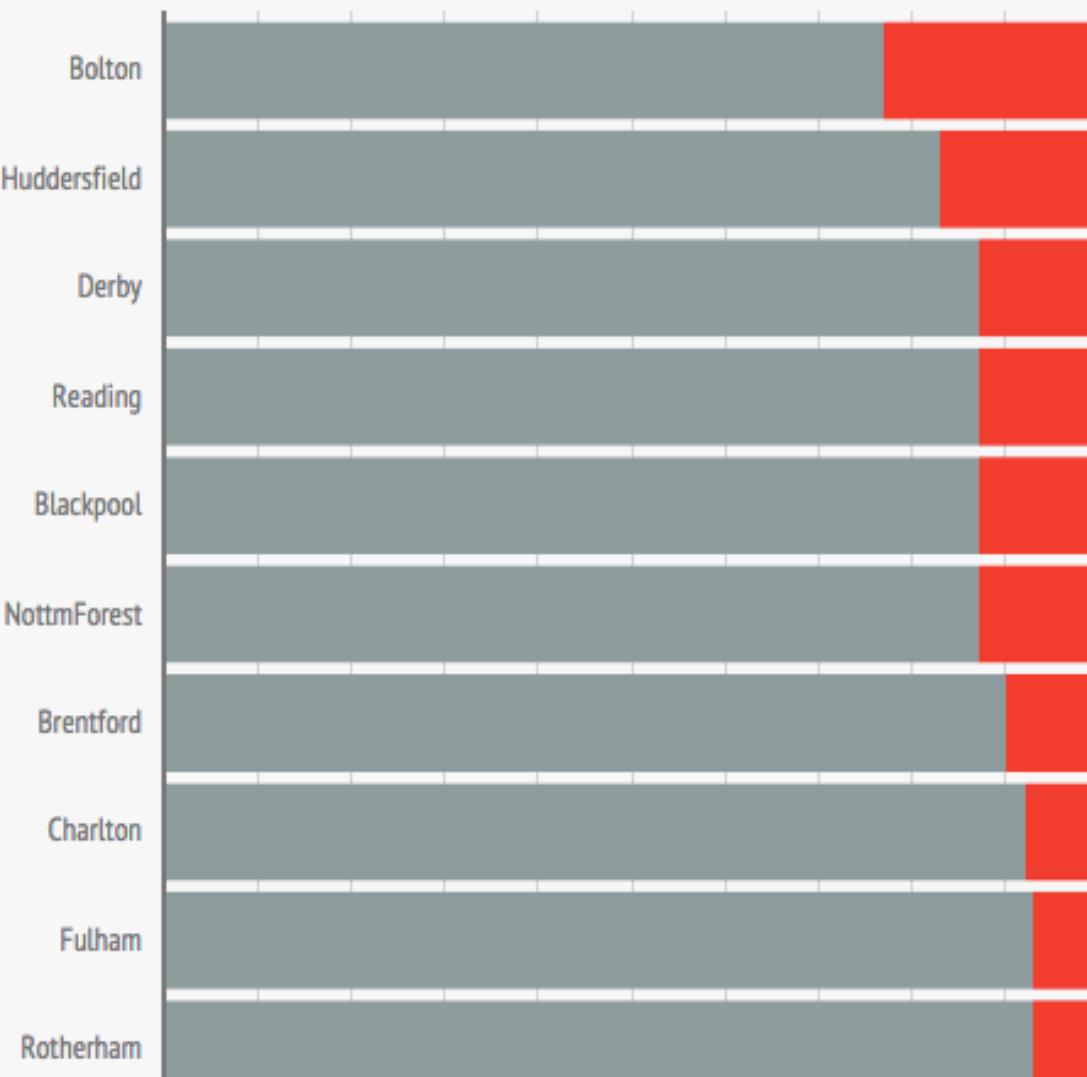
Popular Female Photo Contexts



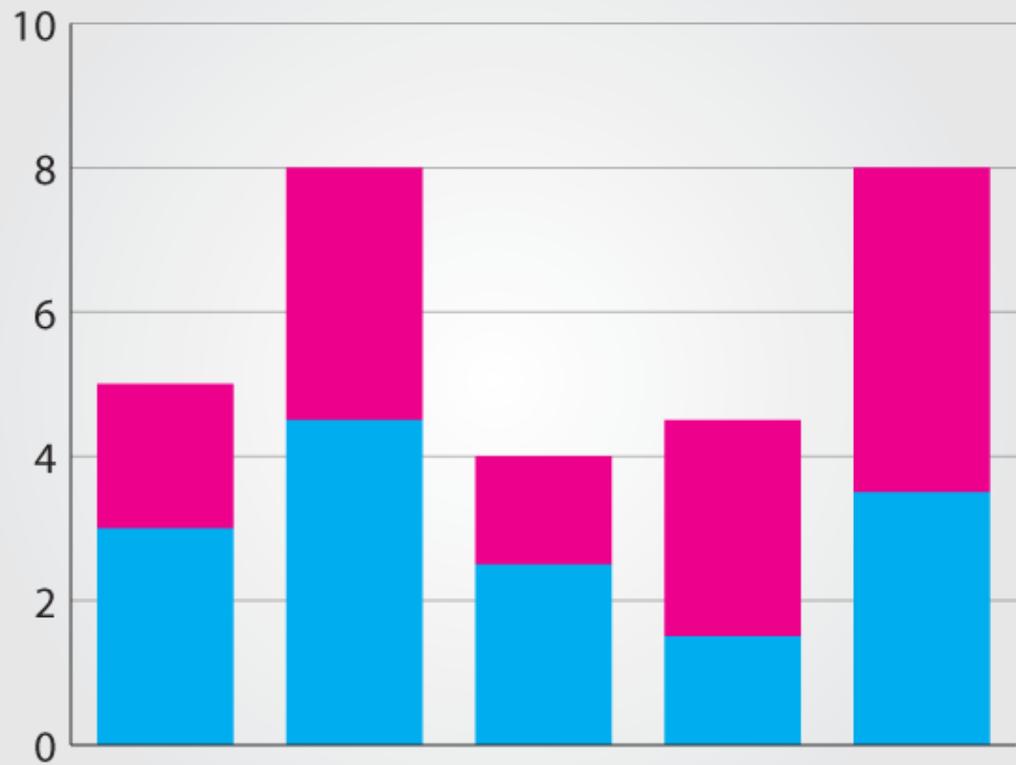
Data structure for bar:

Category	2012
One category	481.0
Another category	551.1
A third category	527.4

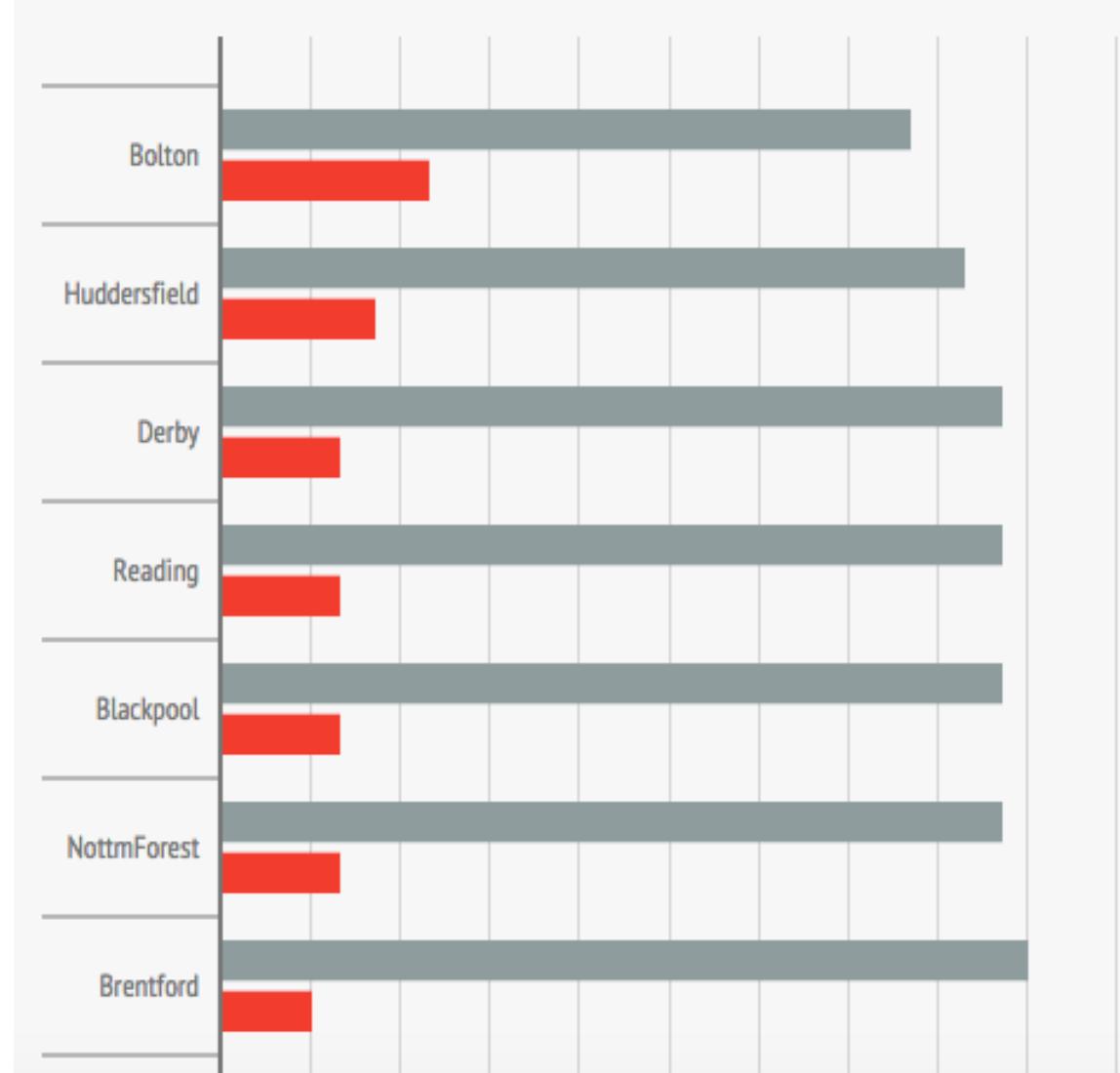
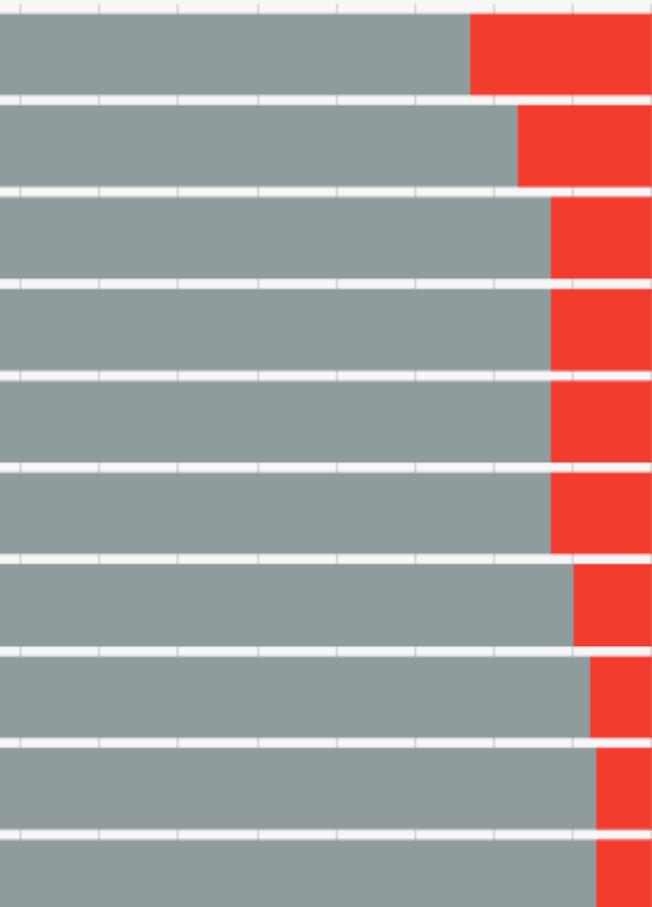
This is higher than any other team in the championship



This is higher than any other team in the championship

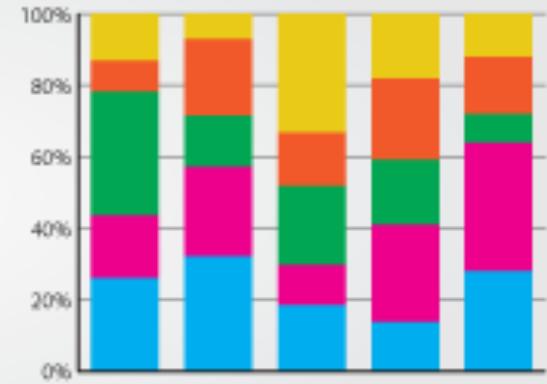
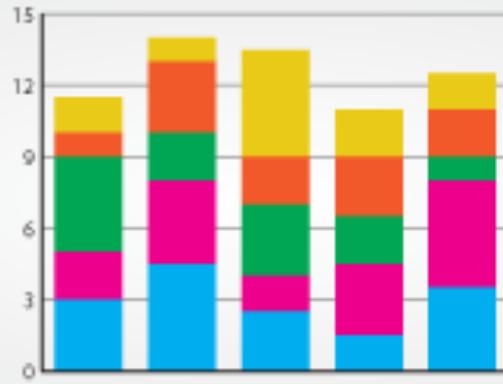
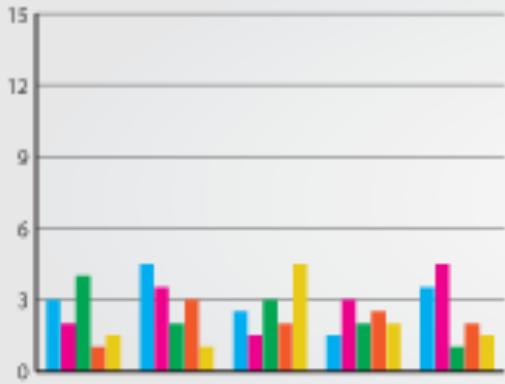


n any other team in the championship



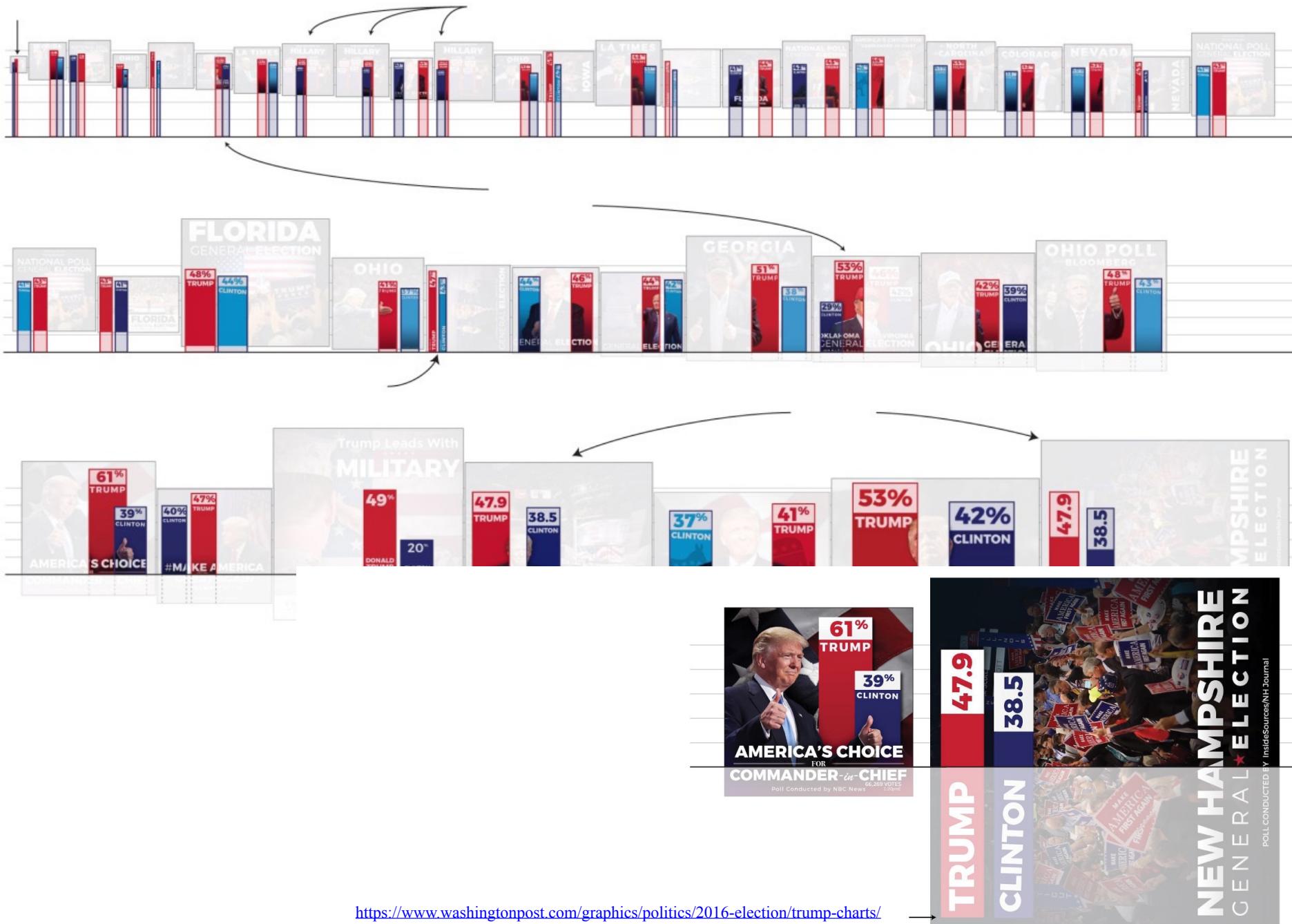
Data structure for multiple/stacked bar:

Category	2012	2017
One category	481.0	0
Another category	551.1	727
A third category	527.4	707



Too many categories per group add visual noise, making it hard to see the patterns in the data.

**Baselines should be 0 for
bar charts...**



Difference of Weight vs. Day 1 Weight

Starting weight was 168 pounds.

POUNDS

10 —

8 —

6 —

4 —

2 —

0 —

-2 —

-4 —

WEIGHT GAIN

WEIGHT LOSS

Day 1

20

40

60

80

100

120

This shifts focus to the weight difference and bar length represents something again.

**...But baselines don't need
to be 0 for line charts.**

Weight Over Time

POUNDS

178

176
Tell the reader to compare position and slope. No more bars
to muddy up the place.

174

172

170

168

166

164

|
Day 1

|
20

|
40

|
60

|
80

|
100

But. But. The value axis doesn't start at zero. What gives? The line chart doesn't need a zero baseline, because bar length is out of the picture. There's no more conflict between visual encoding and context.



Paul Bradshaw retweeted



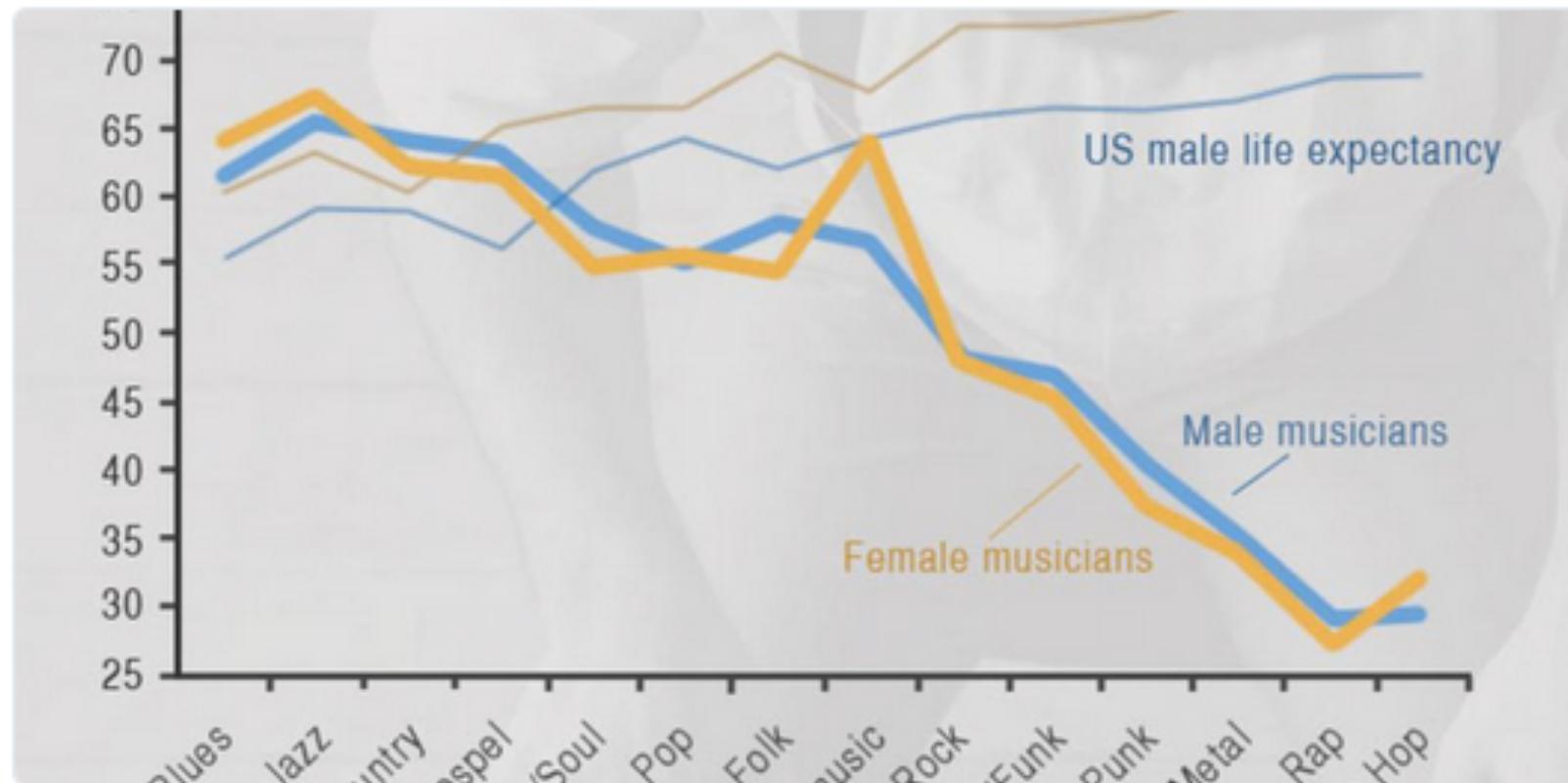
Katie Park @katiepark · Apr 2

repeat after me:

line charts are for time series

line charts are for time series

line charts are for time series



687



682

...

[View more photos and videos](#)

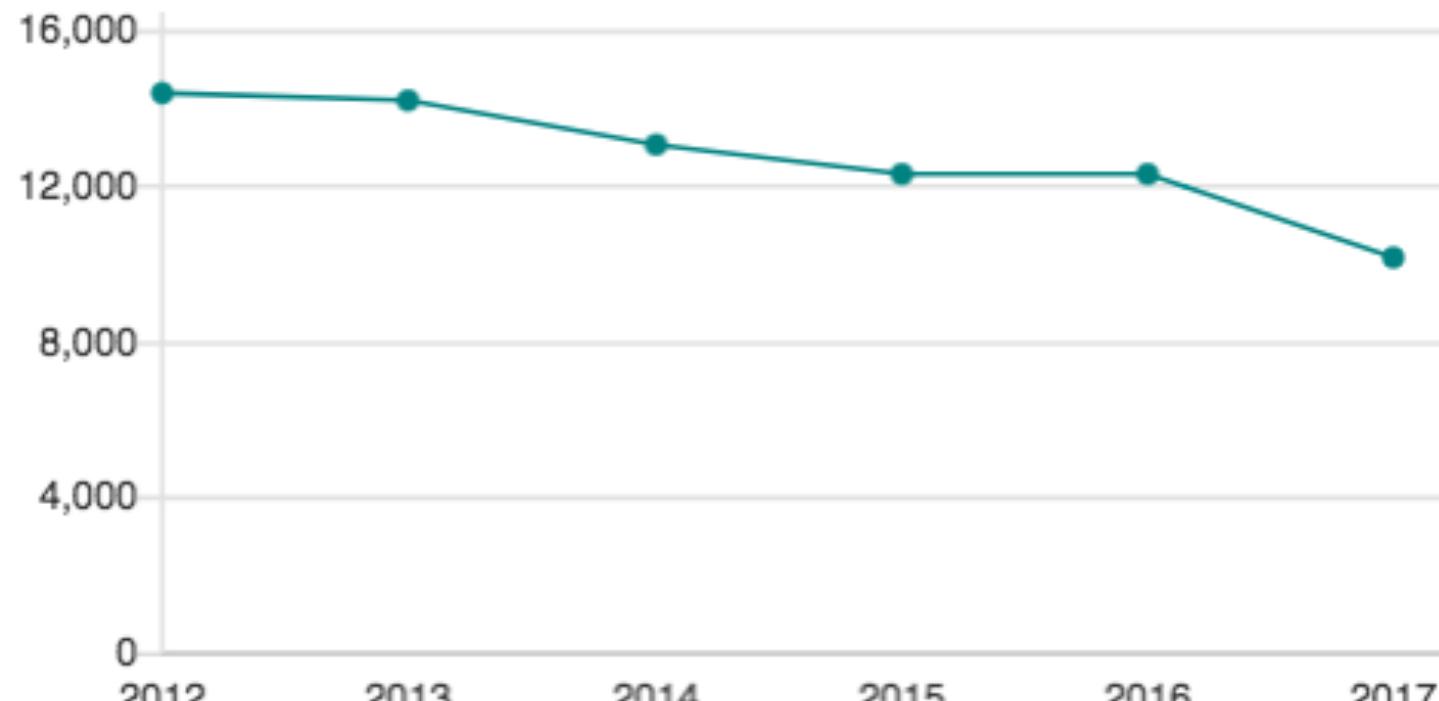
Neighbourhood police: One in seven officers axed were beat bobbies

By Alex Homer
BBC News

Police Community Support Officers (PCSOs)

Forces in England and Wales

— Numbers of PCSOs (FTE)



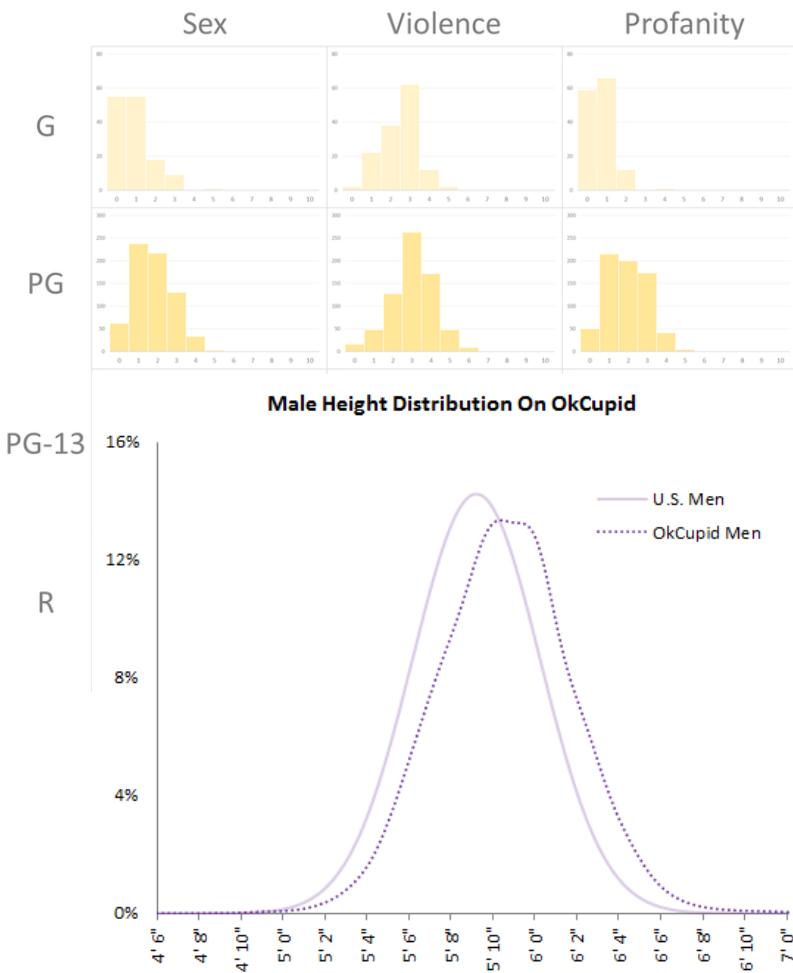
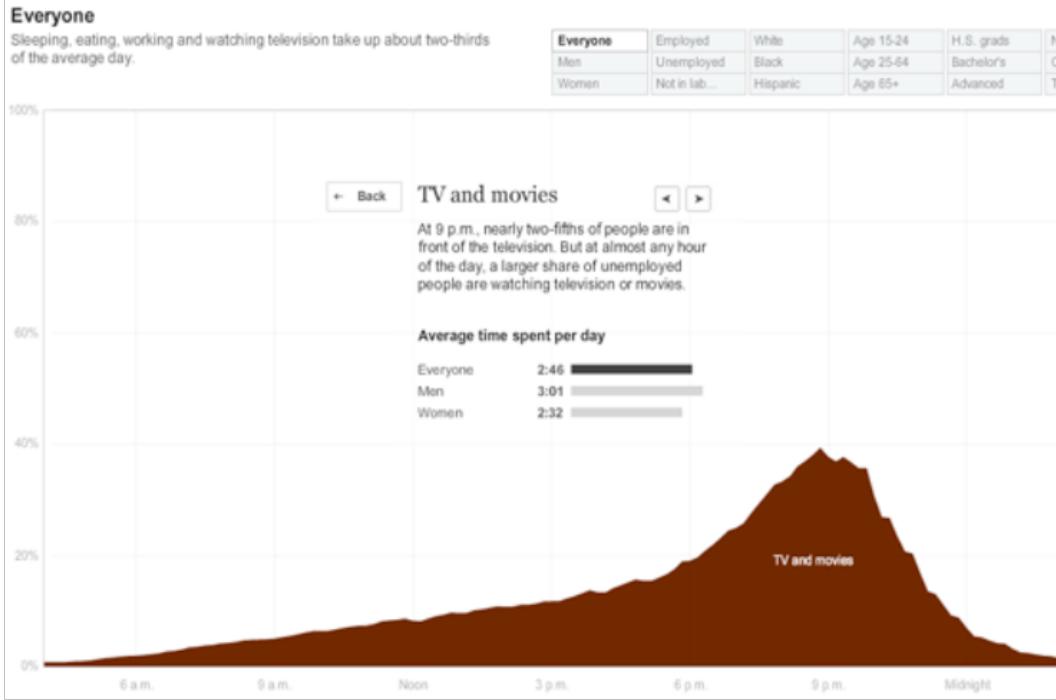
Source: Home Office

BBC

Data structure for line chart:

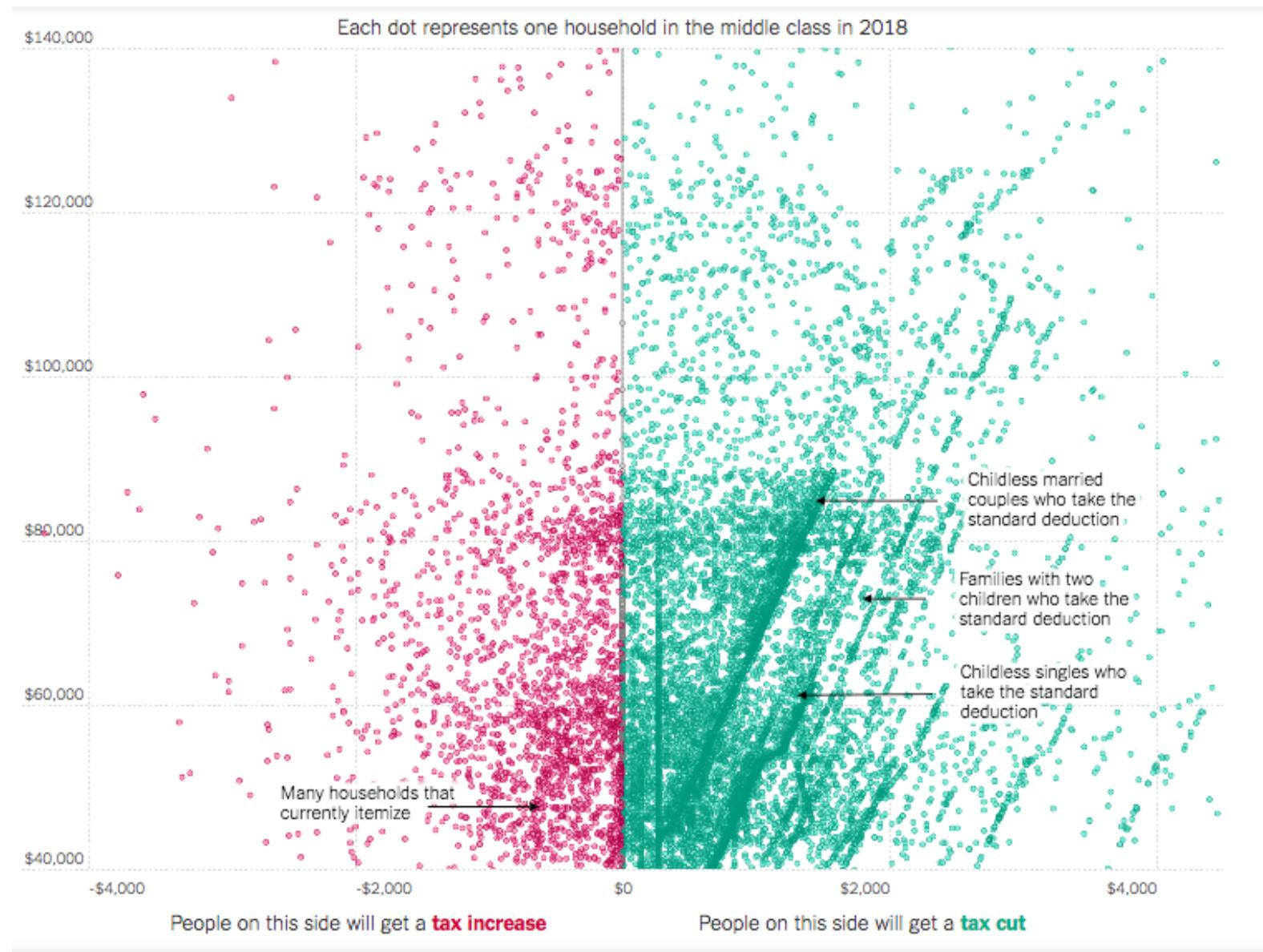
Year	Total
2012	192195.25
2013	185722.69
2014	184078.235
2015	184016.779
2016	181418.911
2017	179708.9055

Distribution stories are told with histograms...



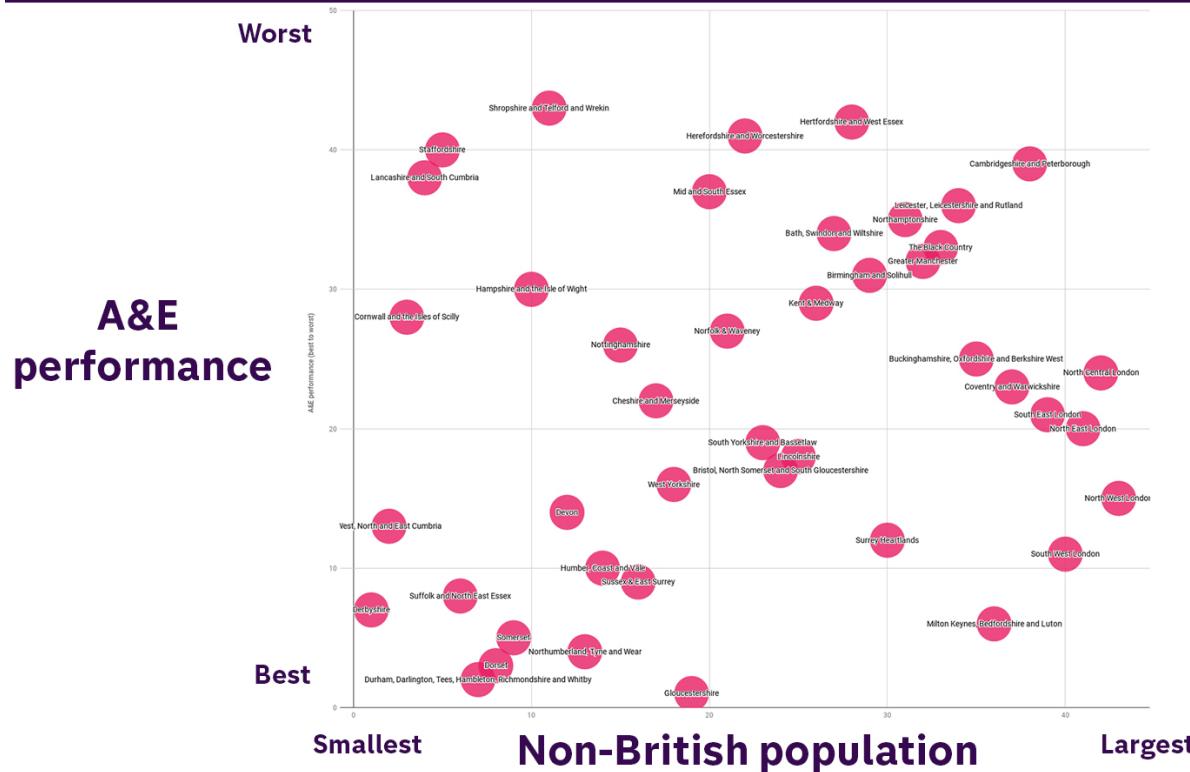
[Sex, Violence & Profanity in the Movies](#); [The Big Lies People Tell In Online Dating](#)

...and scatterplots (which can also show relationships)



...for factchecking/debunking

Relative size of non-British population versus A&E performance by area, England



A&E performance based on figures for November 2017 on percentage of patients discharged, treated or transferred within 4 hours of arrival, broken down by Sustainability and Transformation Plan (STP) area. A score of 1 means the highest proportion of patients treated within 4 hours. Non-British population based on estimates from the Office for National Statistics, 2016, broken down by administrative county and ranked for these purposes. A score of 1 means the lowest proportion of non-British people compared to the wider local population of any STP area in England.

Figures have been cross referenced based on closest match between administrative county, local authority and STP area.

X and Y can be coordinates

Tour the data

Explore on my own

+

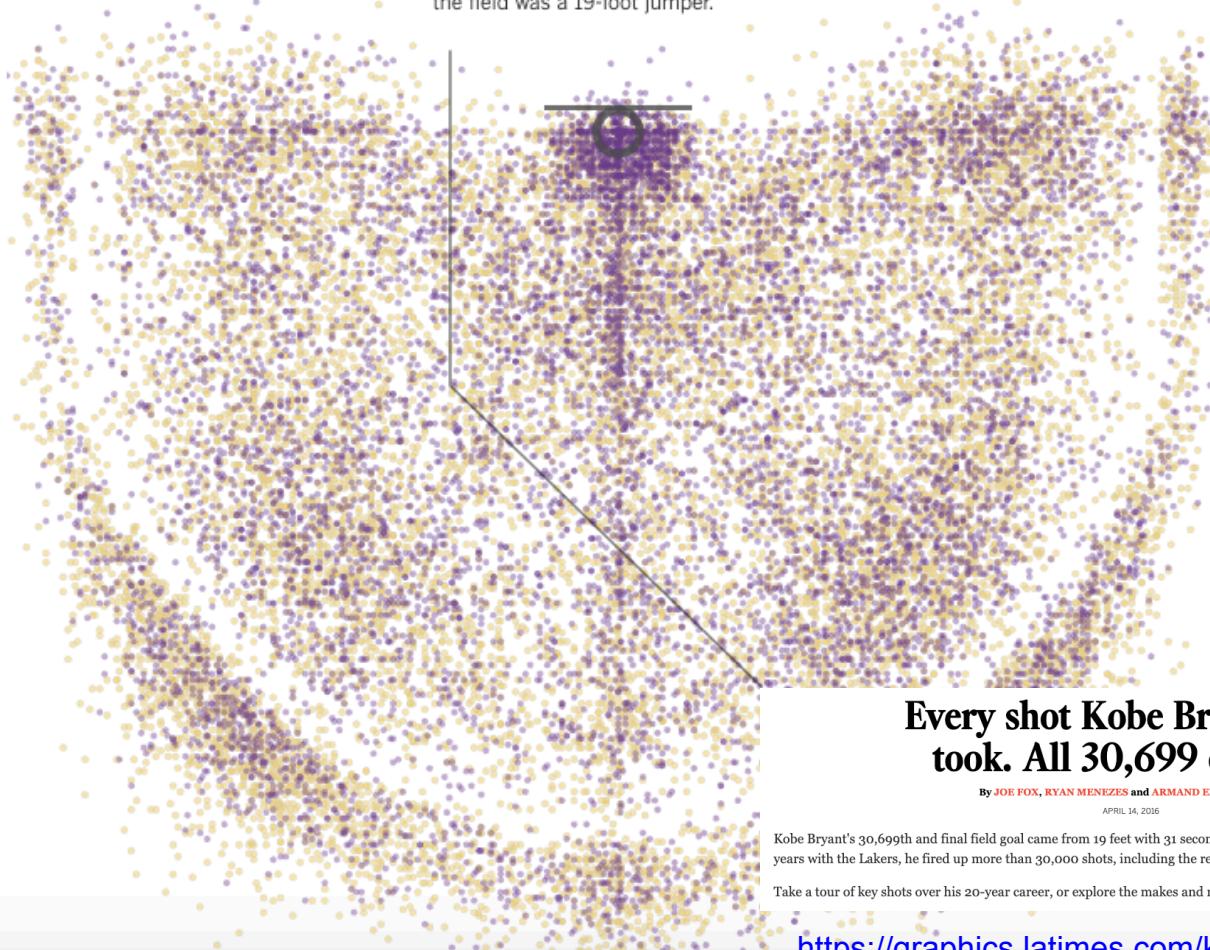
Bryant attempted
30,699 shots
throughout his
career.

-

- Made
- Missed

Kobe's last shot:

April 13, 2016: Capping a 60-point performance, Kobe's final shot from the field was a 19-foot jumper.



Every shot Kobe Bryant ever took. All 30,699 of them

By JOE FOX, RYAN MENEZES and ARMAND EMAMDJOMEH

APRIL 14, 2016

Kobe Bryant's 30,699th and final field goal came from 19 feet with 31 seconds left against the Utah Jazz. During his 20 years with the Lakers, he fired up more than 30,000 shots, including the regular season and playoffs.

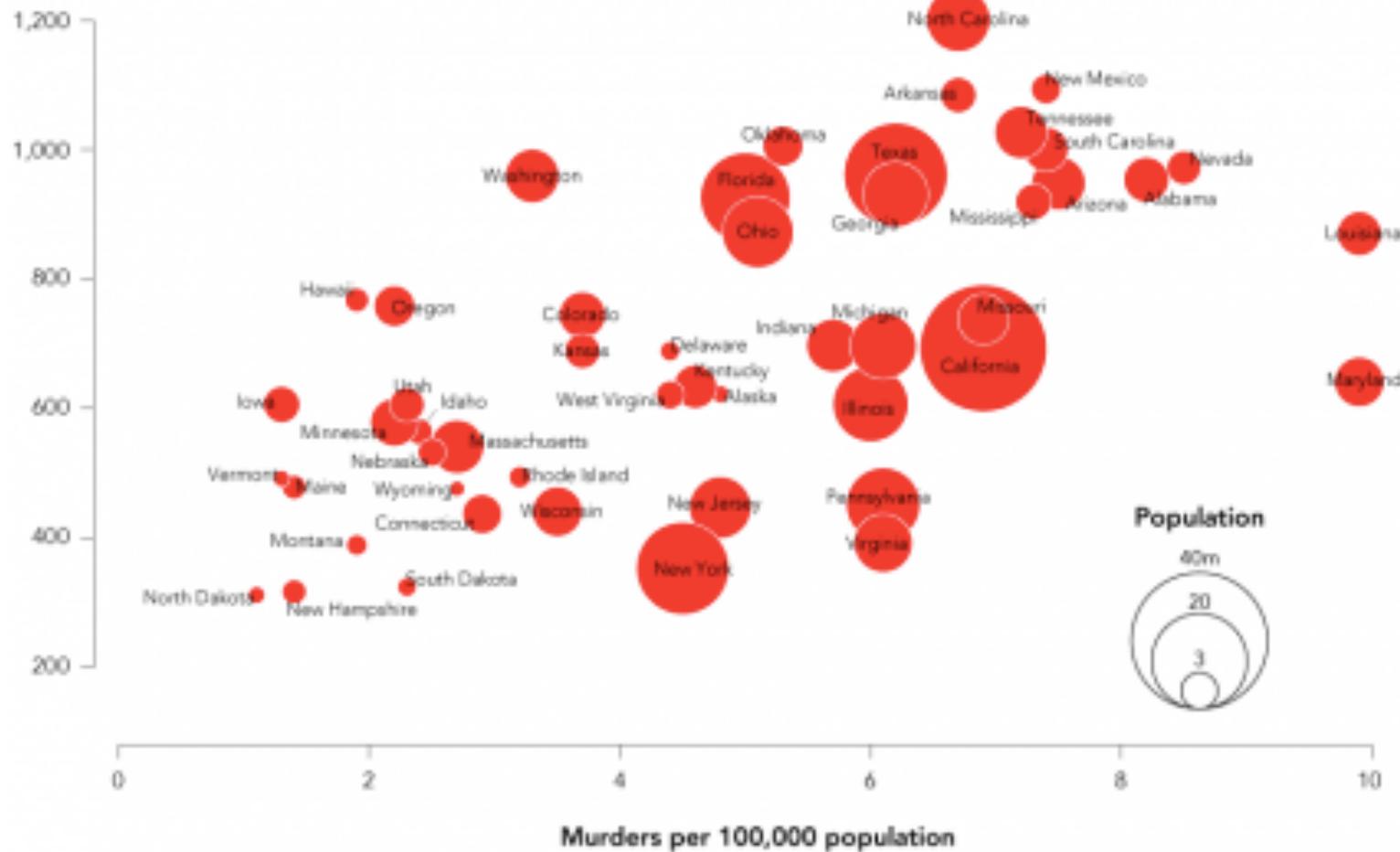
Take a tour of key shots over his 20-year career, or explore the makes and misses over his long career on your own.

<https://graphics.latimes.com/kobe-every-shot-ever/>

Data structure for scatterplot:

Police force	Events 1	Events 2
Avon & Somerset	3339.42	3039.42
Bedfordshire	1457.06	1157.06
Cambridgeshire	1677.45	1377.45
Cheshire	2311.46	2011.46
Cleveland	1828.86	1528.86
Cumbria	1425.3	1125.3
Derbyshire	2118.51	1818.51
Devon & Cornwall	3525.09	3225.09
Dorset	1677.7	1377.7
Durham	1663.1	1363.1
Dyfed-Powys	1430.61	1130.61
Essex	3708.47	3408.47
Gloucestershire	1506.92	1206.92
Greater Manchester	7798.39	7498.39

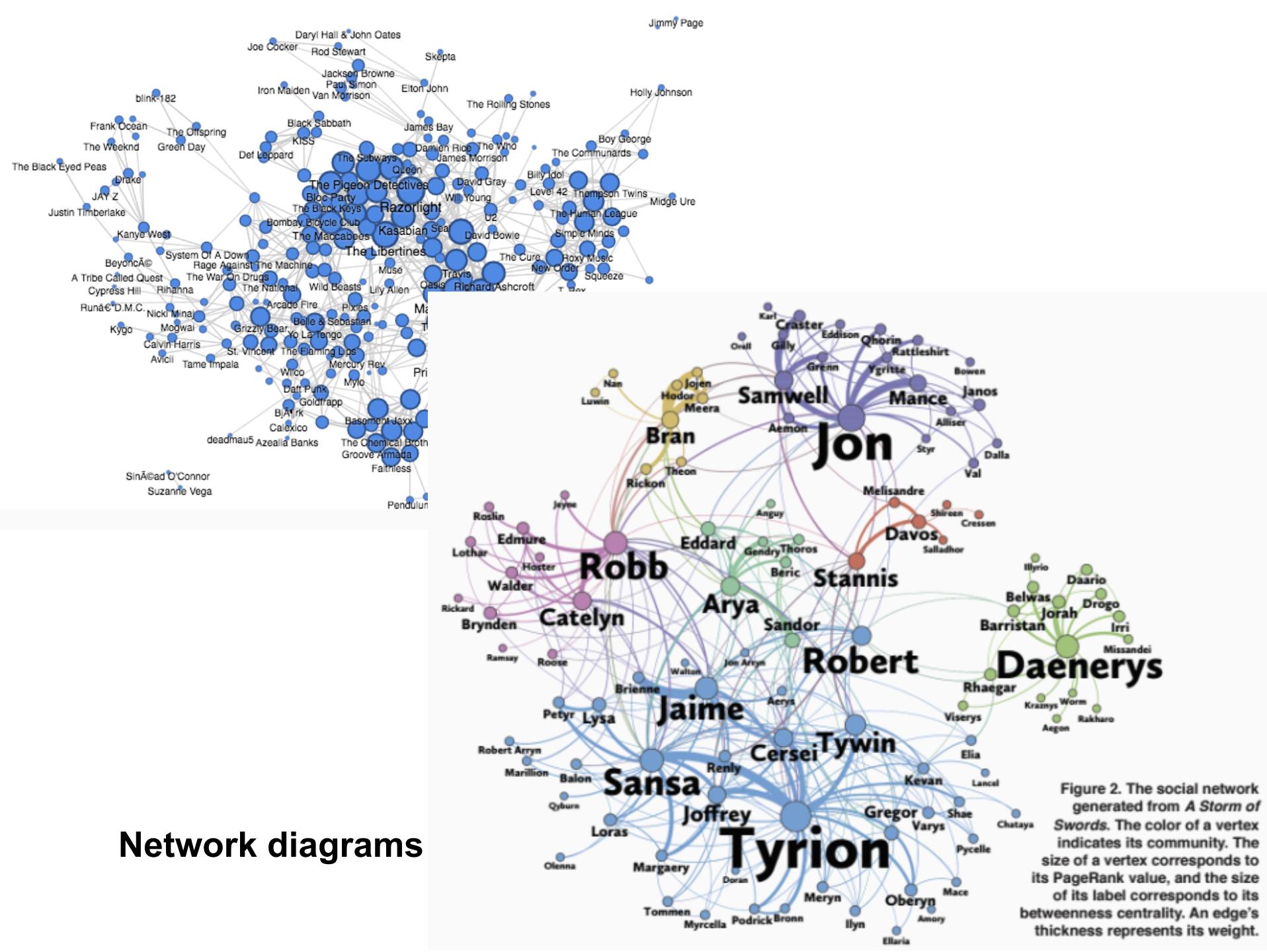
Burglaries per
100,000 population



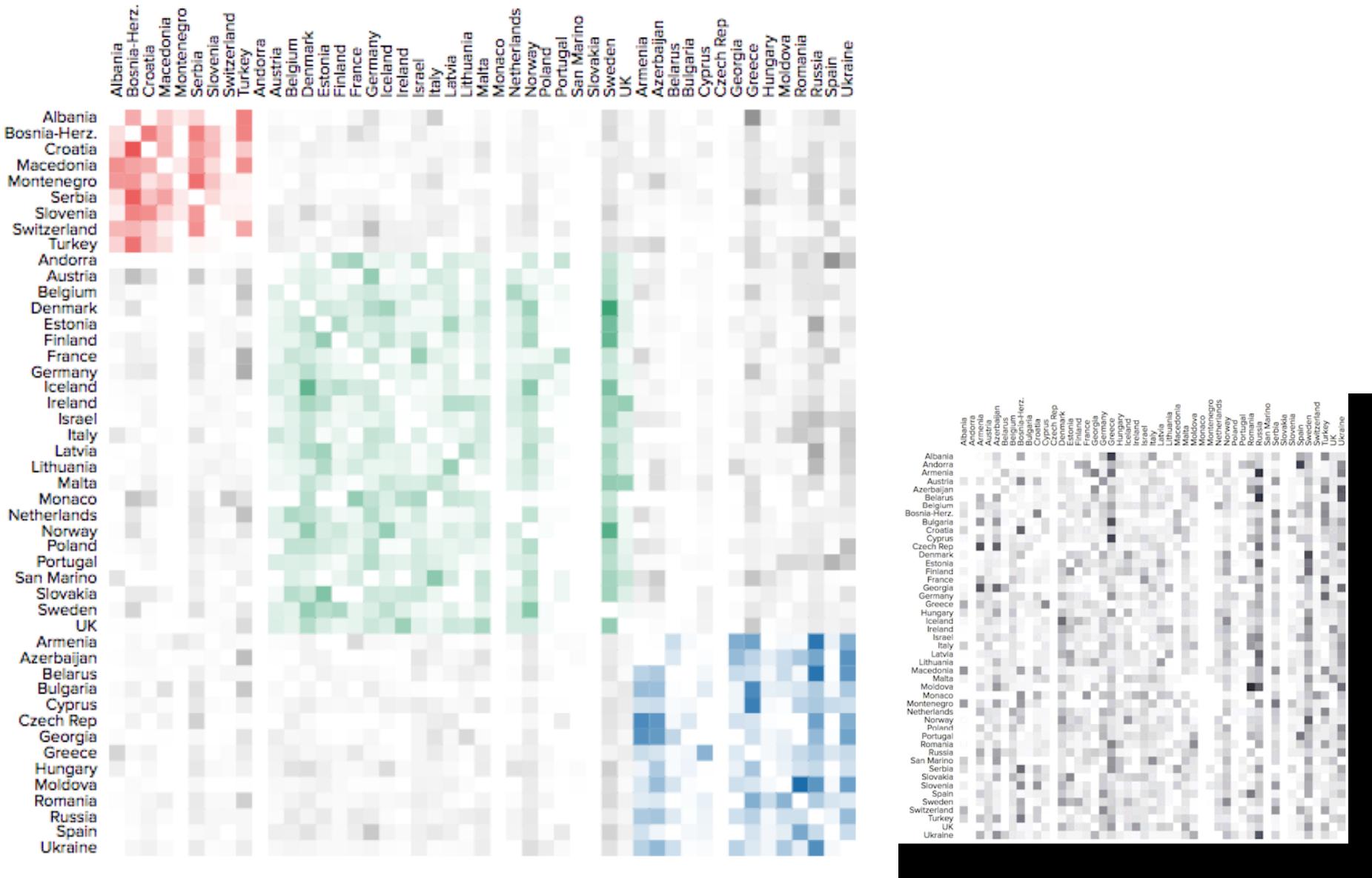
Bubble chart: comparison across 3 variables

Data structure for bubble plot:

Police force	Events 1	Events 2	Population
Avon & Somerset	3339.42	3039.42	1680700
Bedfordshire	1457.06	1157.06	664500
Cambridgeshire	1677.45	1377.45	849000
Cheshire	2311.46	2011.46	1048100
Cleveland	1828.86	1528.86	564300
Cumbria	1425.3	1125.3	497900
Derbyshire	2118.51	1818.51	1042000
Devon & Cornwall	3525.09	3225.09	1733900
Dorset	1677.7	1377.7	771900
Durham	1663.1	1363.1	627800
Dyfed-Powys	1430.61	1130.61	515900
Essex	3708.47	3408.47	1802200
Gloucestershire	1506.92	1206.92	623100
Greater Manchester	7798.39	7498.39	2782100



X-Y heatmap (relationships)



Maps: why?

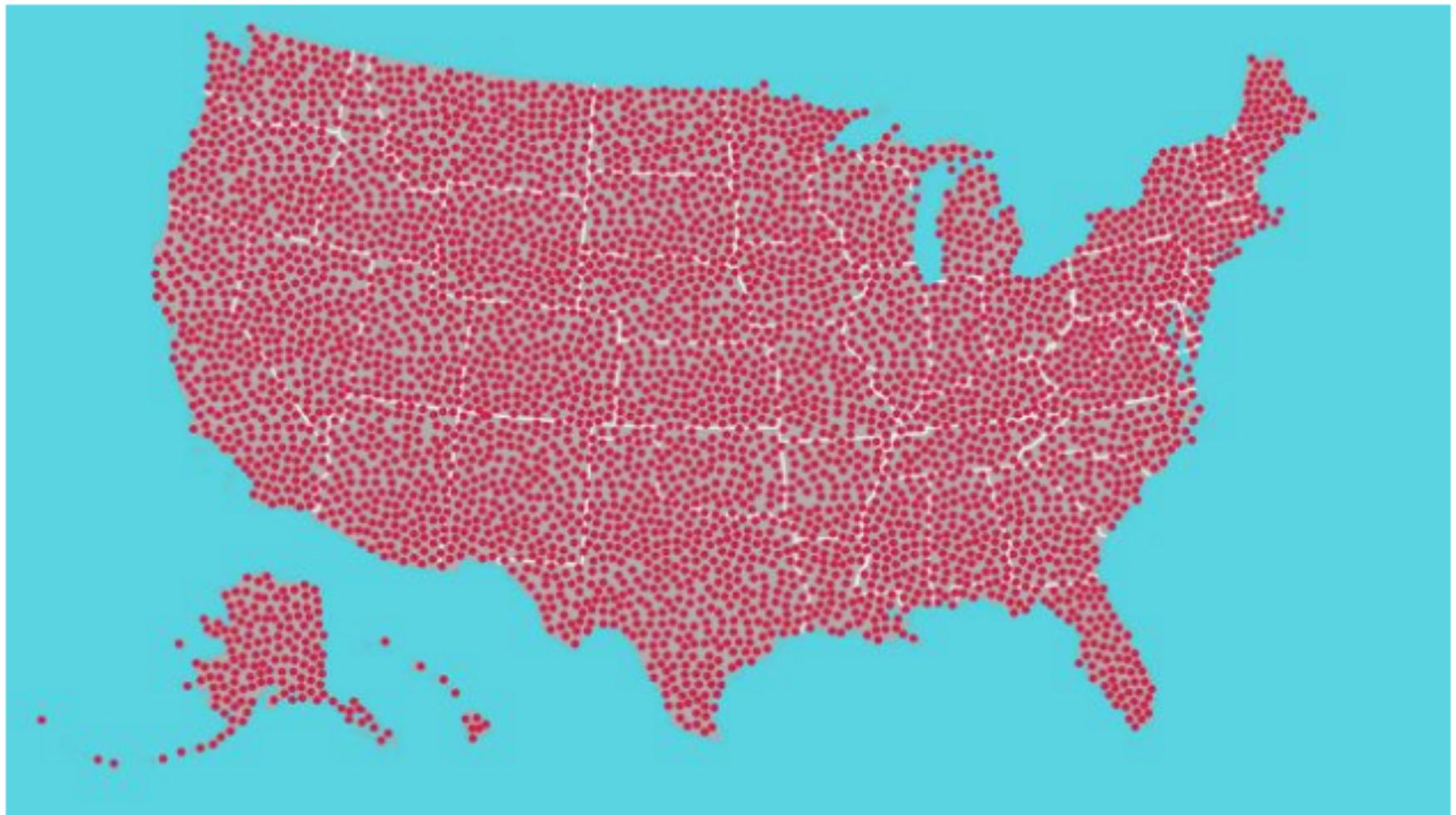
Because it's geographical?

No.

Identify the purpose:
To explore? To show
pattern?

We Put 700 Red Dots On A Map

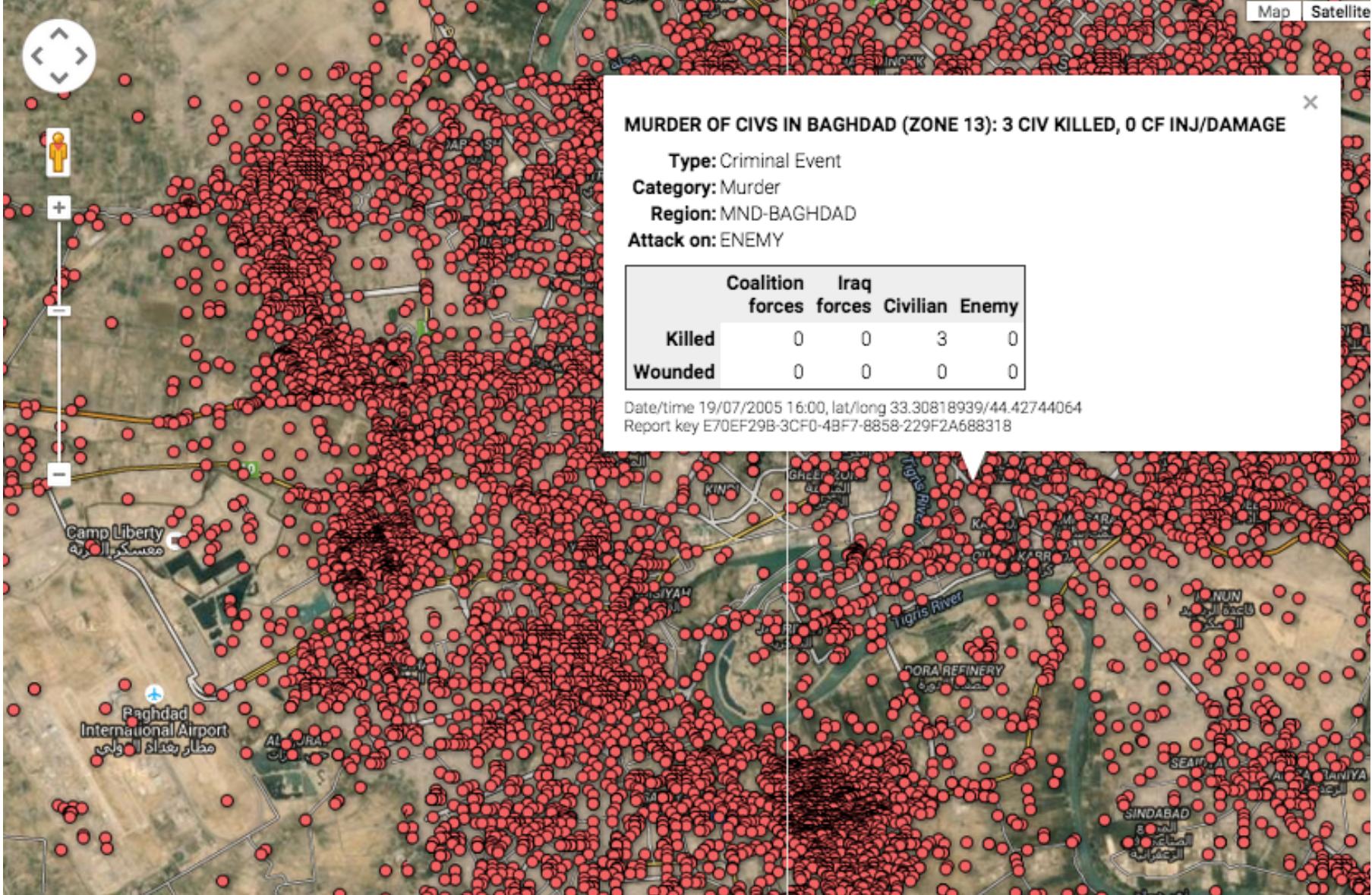
Posted Oct. 22, 2014



But use with care: don't use just because data is geographical

Map tools...

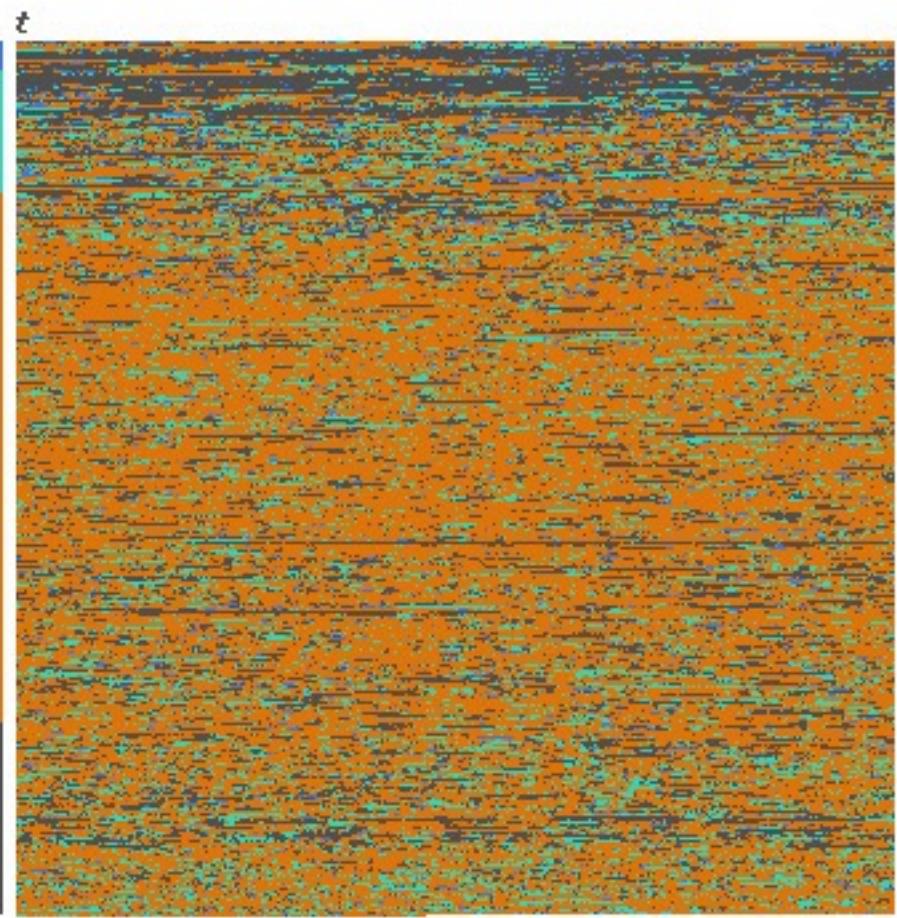
Datawrapper
Flourish
BatchGeo (£)
Carto (closed)
Storymap JS
QGIS



Wikileaks Iraq war logs: every death mapped

The Wikileaks Iraq war logs provide us with a unique picture of every death in Iraq. These are those events mapped using Google Fusion tables

- Download the data from the Datablog



Kamel Makhloifi

+ Follow

function

WIP

Blue = *Friendly*, Green = *Host* Nation, Orange = Civilians, Grey = Enemies.

First one is function of sum, second one is function of time, or how you can dilute the media impact of a massacre by killing a few people each day for 6 years. Just remember that host nation + civilian + enemies = mostly Iraqis.

Used the cleaned dump from The Guardian.

Maarten Lambrechts

@maartenzam

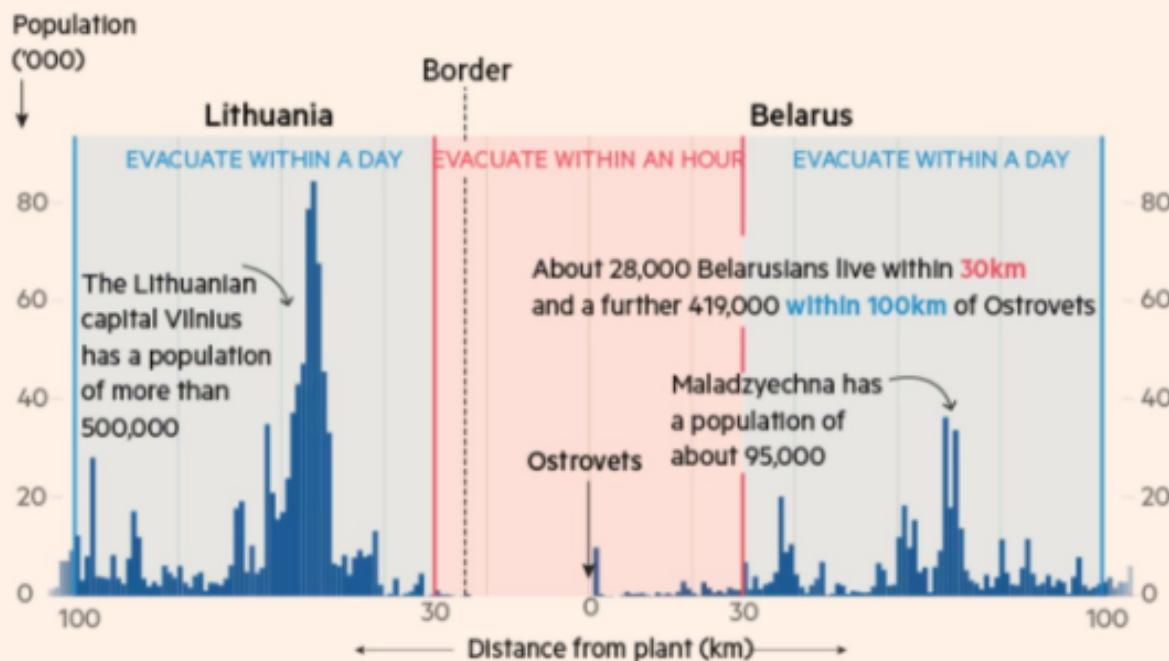
Follow



Sometimes the best map is not a map,
illustrated by @BillyEhrenberg
[ft.com/content/a98322 ...](https://ft.com/content/a98322)

Almost 1m Lithuanians live in the **100km** evacuation zone around
Ostrovets nuclear plant

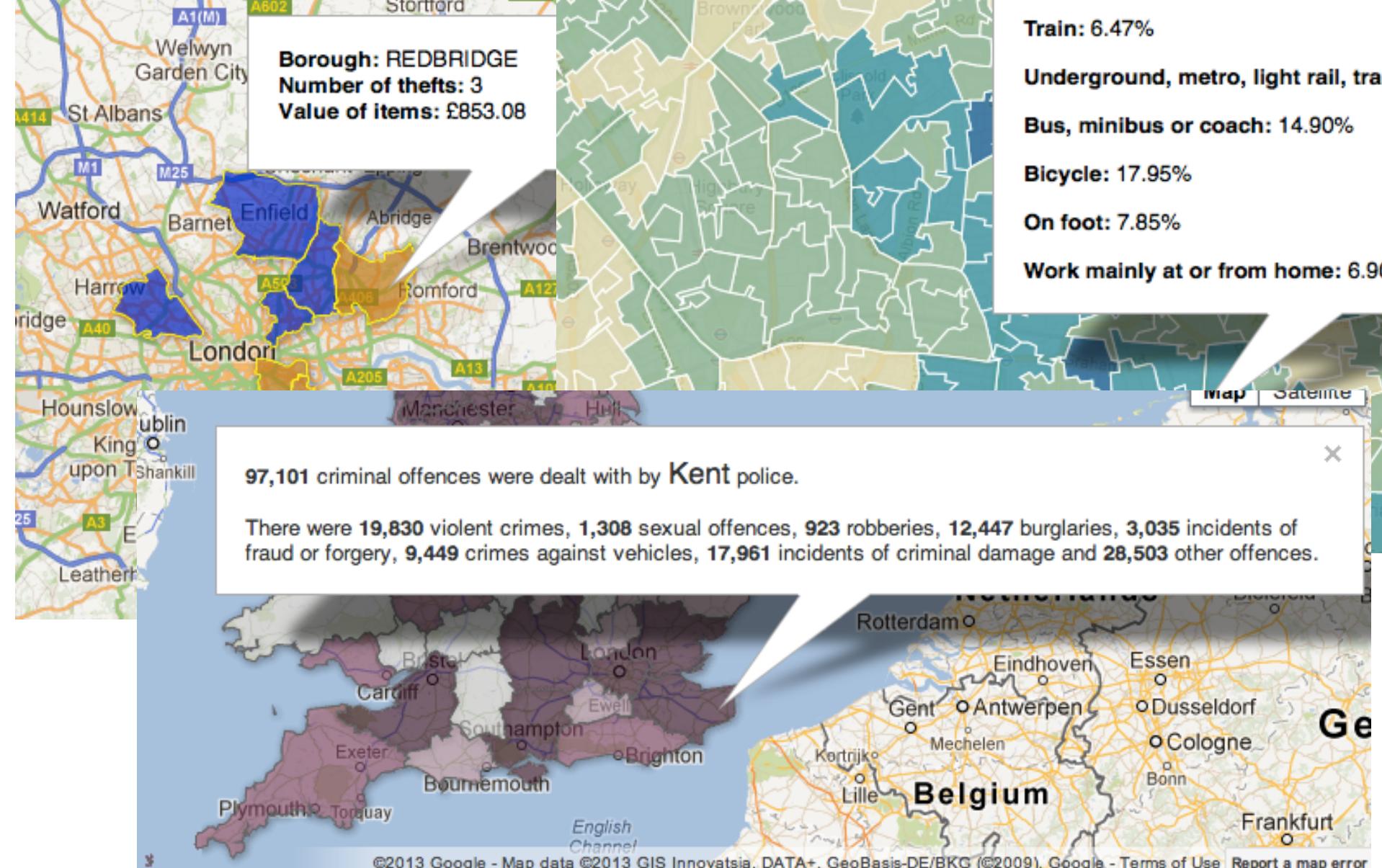
Areas **within 30km** of Ostrovets must have plans to evacuate within 1hr of a serious nuclear incident.
For areas **between 30km and 100km** from the plant, the timescale is one day



FT graphic Billy Ehrenberg-Shannon

Source: JRC Open Data repository

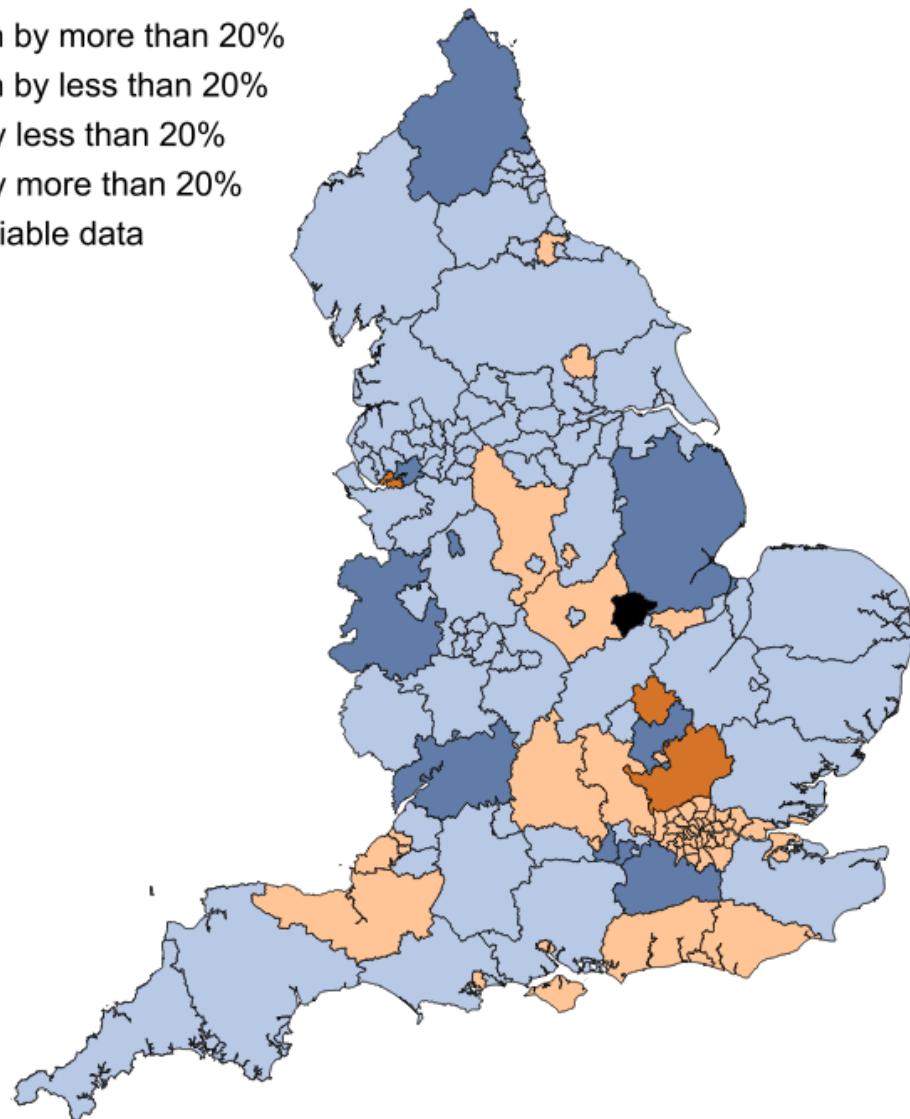
© FT



Outside of London, English bus networks are mostly getting smaller

Percentage change in miles travelled by buses in 2016/17 compared with 2013/14

- Down by more than 20%
- Down by less than 20%
- Up by less than 20%
- Up by more than 20%
- Unreliable data



More tools...

Rawgraphs.io

Kumu.io

JuxtaposeJS

TimelineJS

Thinglink

GIF makers (e.g. Giphy)

Quiz makers: Playbuzz

Deviation

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

Example FT uses
Trade surplus/deficit, climate change

Correlation

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

Example FT uses
Inflation & unemployment, income & life expectancy

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

Distribution

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

Example FT uses
Income distribution, population
(age/gender) distribution

Change over Time

Give emphasis to changing trends. These can be short (intra-day) movements or extended series (over months or years). Choosing the correct time period is key to providing suitable context for the reader.

Example FT uses
Share price movements, economic time series

Magnitude

Show size comparisons. These can be relative (just being able to see larger/bigger) or absolute (need to see the raw numbers). It's good to have a 'counted' number (for example, barrels, customers etc.) rather than a calculated rate or per cent.

Example FT uses
Commodity production, market capitalisation

Spatial

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

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

Flow

Show changes in flows from one location to another. Good for tracing the eventual outcome of a complex process.

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

Sankey

Show changes in flows from one location to another. Good for tracing the eventual outcome of a complex process.

Example FT uses

The standard approach for showing part-to-whole relationships but should always be tends rather than totals to avoid base geography.

Waterfall

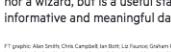
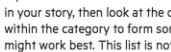
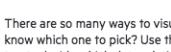
Designed to show the sequencing of a flow process. Good for budgets. Can include +/- components.

Chord

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

Network

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



FT graphic: Alan Smith; Chris Campbell; Ben Bell; Li Huamei; Graham French; Billy Oberberg; Paul McCullagh; Martin Sibley
Inspired by the Graphic Contrarium by Jon Schmidbauer and Svenneth Blawie

Visual vocabulary

Designing with data

There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

ft.com/vocabulary

FT

Summing up: tips

What story are you telling?
Choose the chart for the story
Maps aren't a given. Do they say/do anything?
Colour - repetition + contrast (& don't discriminate)
Tools - consider gifs, sliders timelines and other options