

Working Group on Data Visualisation: Good and Bad Practices in Data Vis

September 26th 2017

Oli Chalkley o.chalkley@bristol.ac.uk

Bobby Stuijzand bg.stuijzand@bristol.ac.uk

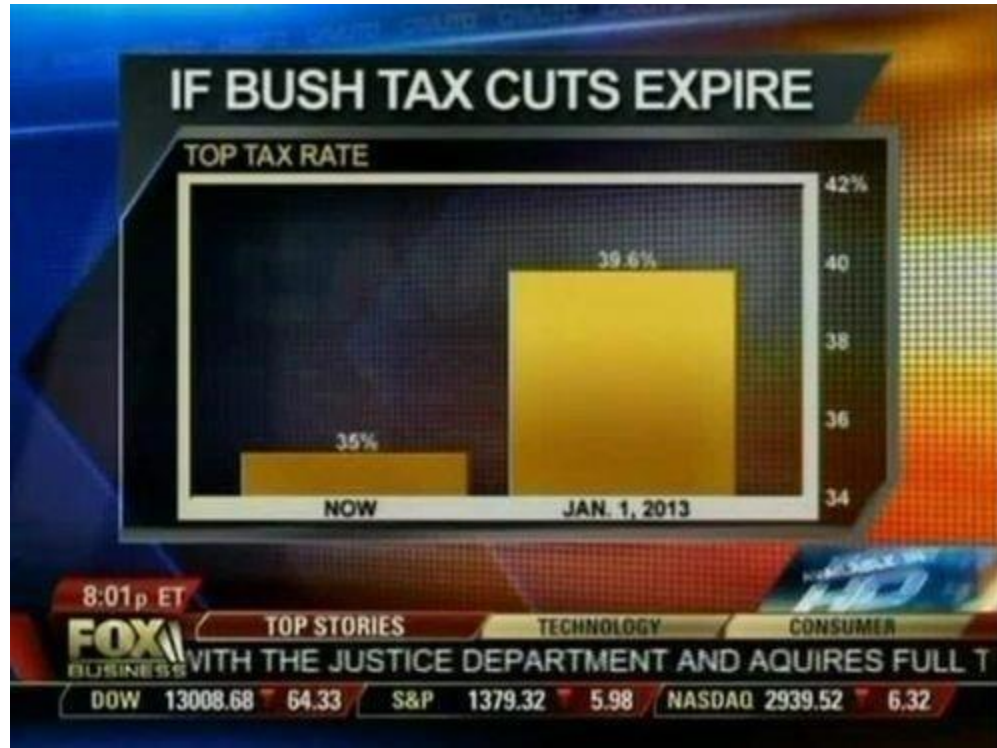
Harriet Mills harriet.mills@bristol.ac.uk

Today's session

- Basic visualisation mistakes
- Good examples
- Small groups discussion
- Working group format

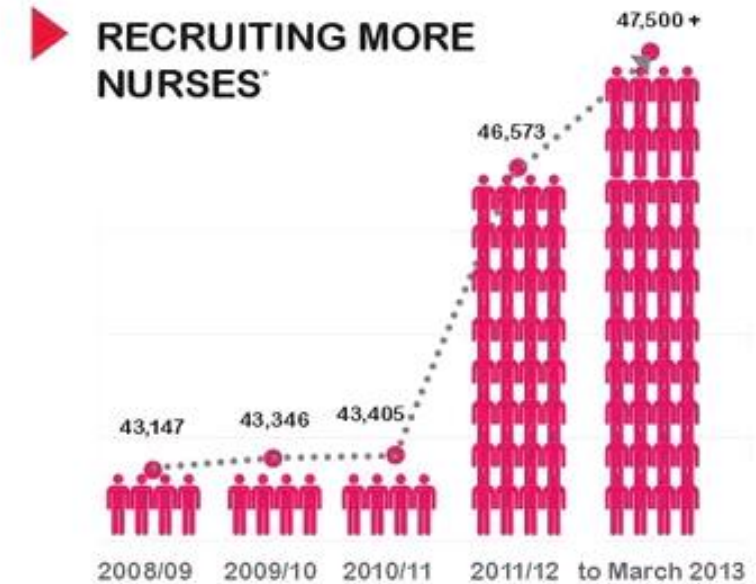
Basic mistakes

Truncated y-axis / unusual scales



<https://blog.heapanalytics.com/how-to-lie-with-data-visualization/>

The NSW Health system is...



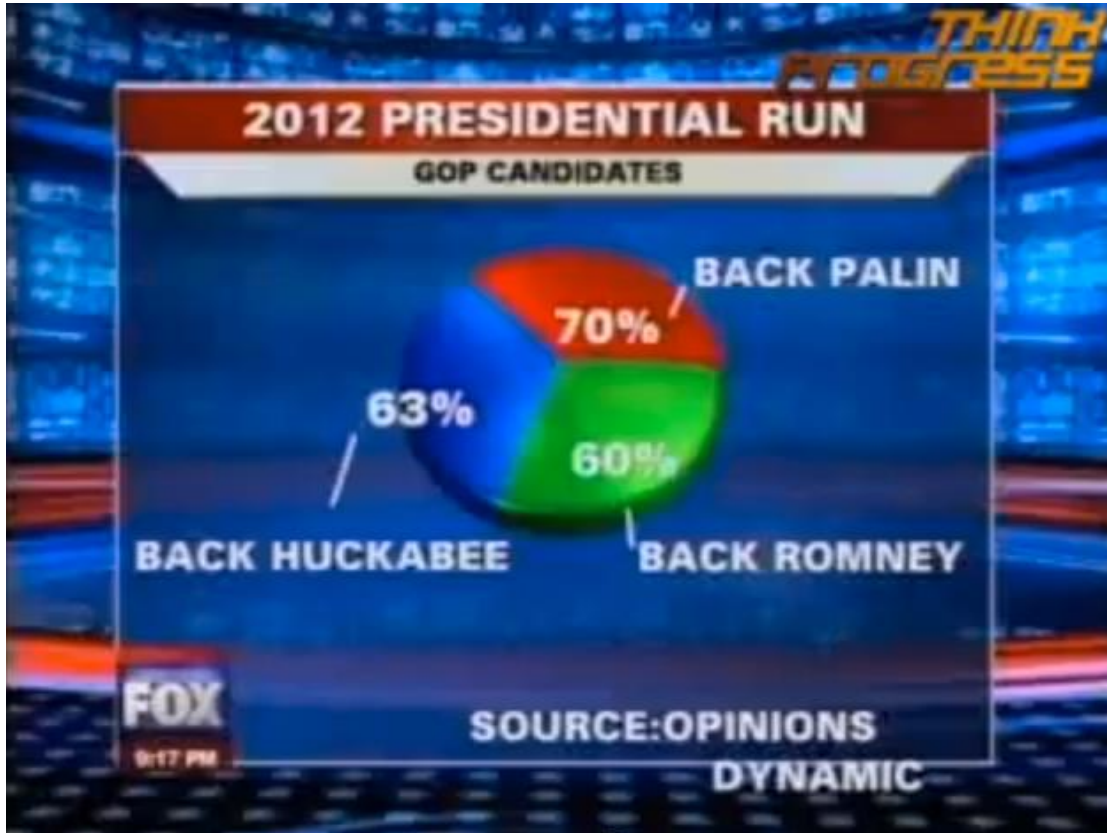
* Nursing headcount figures at June includes non casual staff and 3rd schedule



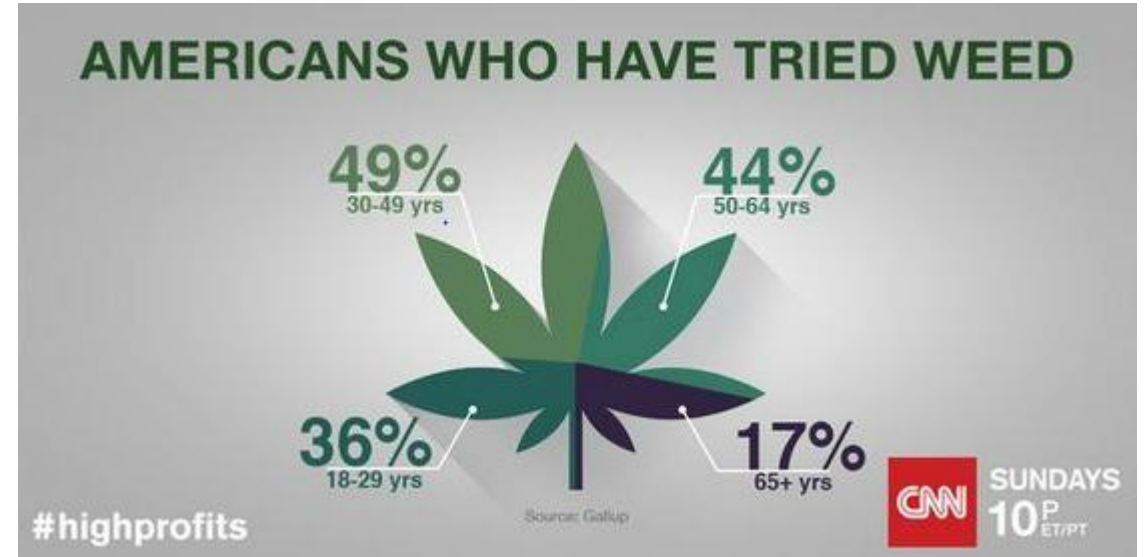
NSW Ministry of Health March 2013

<https://www.theguardian.com/news/datablog/gallery/2013/aug/01/16-useless-infographics>

Ignoring conventions



<https://blog.heapanalytics.com/how-to-lie-with-data-visualization/>

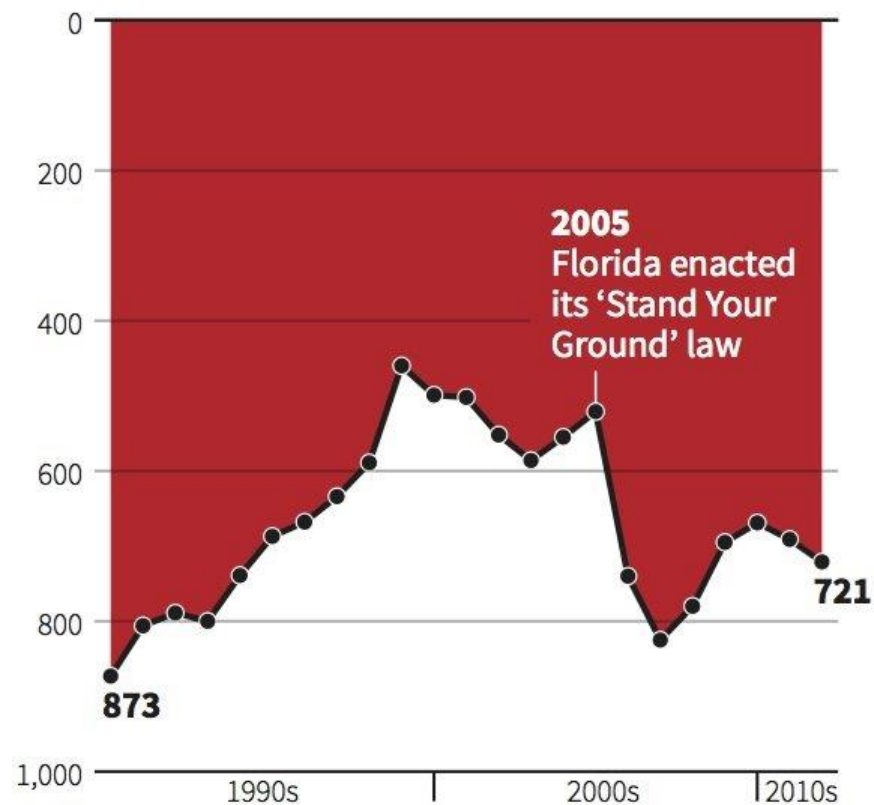


<http://skepchick.org/2015/06/bad-chart-thursday-cnn-pot-pie/>

Ignoring conventions

Gun deaths in Florida

Number of murders committed using firearms



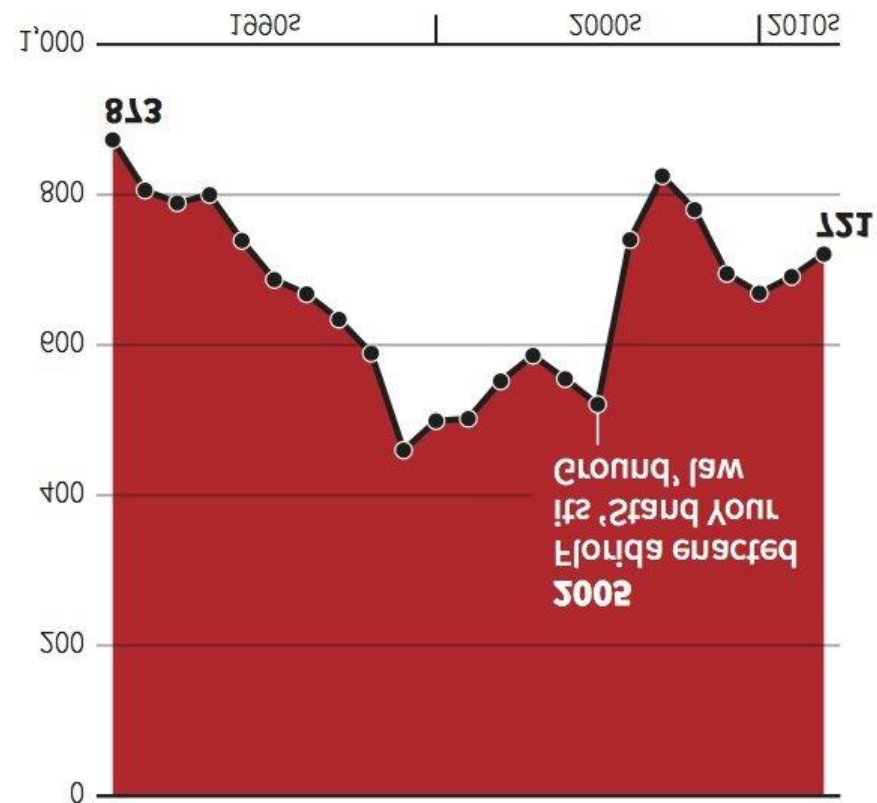
Source: Florida Department of Law Enforcement

<https://blog.heapanalytics.com/how-to-lie-with-data-visualization/>

C. Chan 16/02/2014

REUTERS

Source: Florida Department of Law Enforcement



Number of murders committed using firearms

Gun deaths in Florida

Mind the defaults

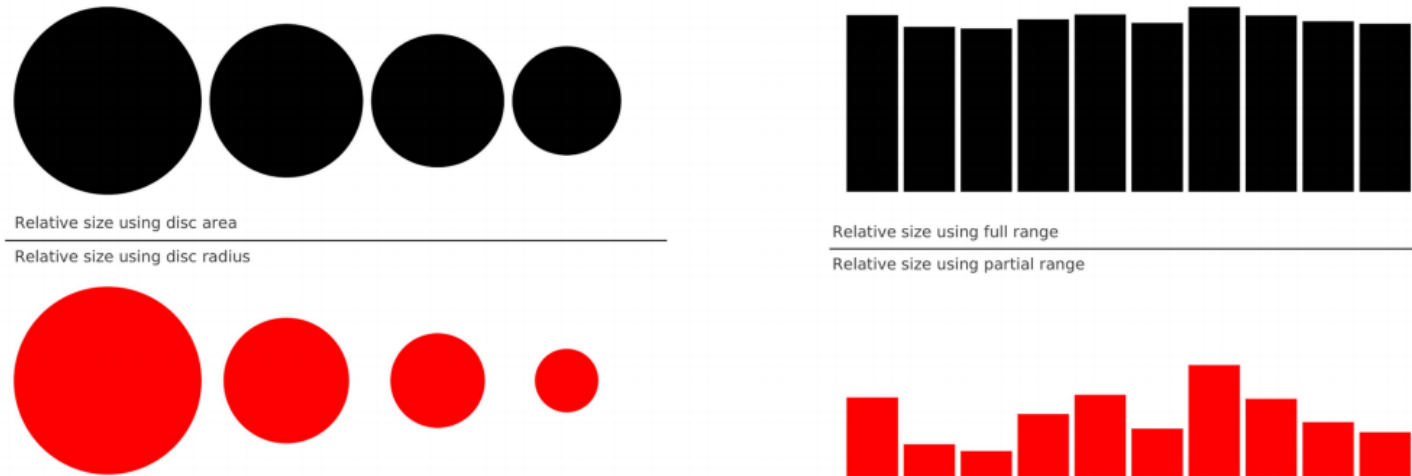
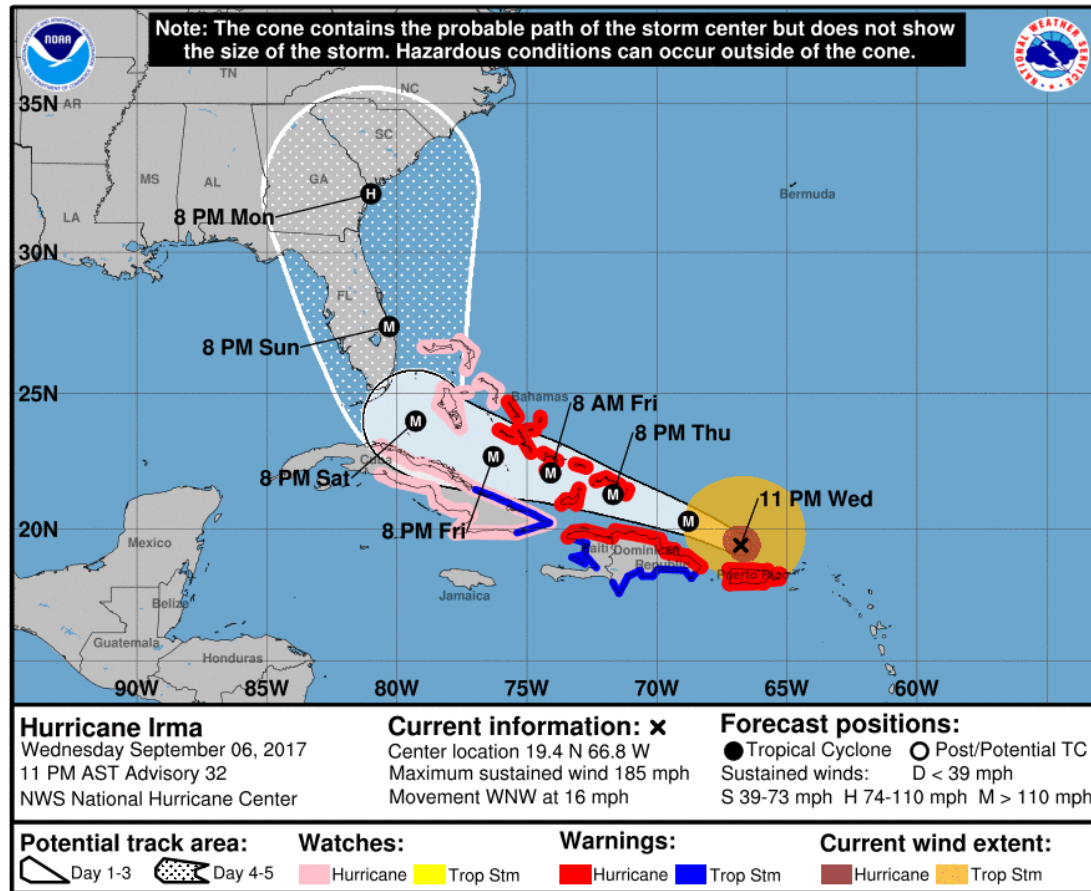


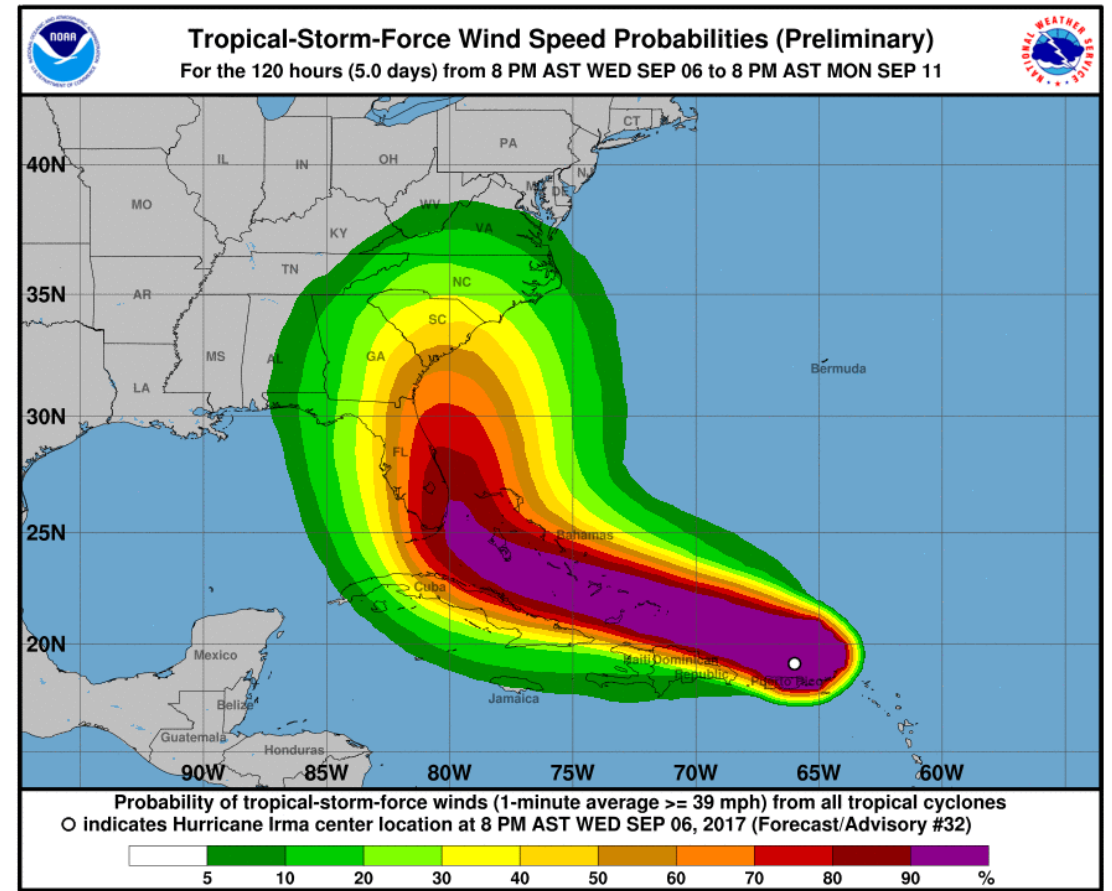
Figure 6. Do not mislead the reader. On the left part of the figure, we represented a series of four values: 30, 20, 15, 10. On the upper left part, we used the disc area to represent the value, while in the bottom part we used the disc radius. Results are visually very different. In the latter case (red circles), the last value (10) appears very small compared to the first one (30), while the ratio between the two values is only 3:1. This situation is actually very frequent in the literature because the command (or interface) used to produce circles or scatter plots (with varying point sizes) offers to use the radius as default to specify the disc size. It thus appears logical to use the value for the radius, but this is misleading. On the right part of the figure, we display a series of ten values using the full range for values on the top part (y axis goes from 0 to 100) or a partial range in the bottom part (y axis goes from 80 to 100), and we explicitly did not label the y-axis to enhance the confusion. The visual perception of the two series is totally different. In the top part (black series), we tend to interpret the values as very similar, while in the bottom part, we tend to believe there are significant differences. Even if we had used labels to indicate the actual range, the effect would persist because the bars are the most salient information on these figures.

doi:10.1371/journal.pcbi.1003833.g006

Ambiguity



Accurate but confusing



Better?

Bad big data visualisations

<https://maxcooper.net/chromos-ep>

Bad big data visualisations

The latest data from the U.S. Census's American Community Survey paints a fascinating picture of the United States at the county level. We've looked at the educational achievement and the median income of the entire nation, to see where people are going to school, where they're earning money, and if there is any correlation.

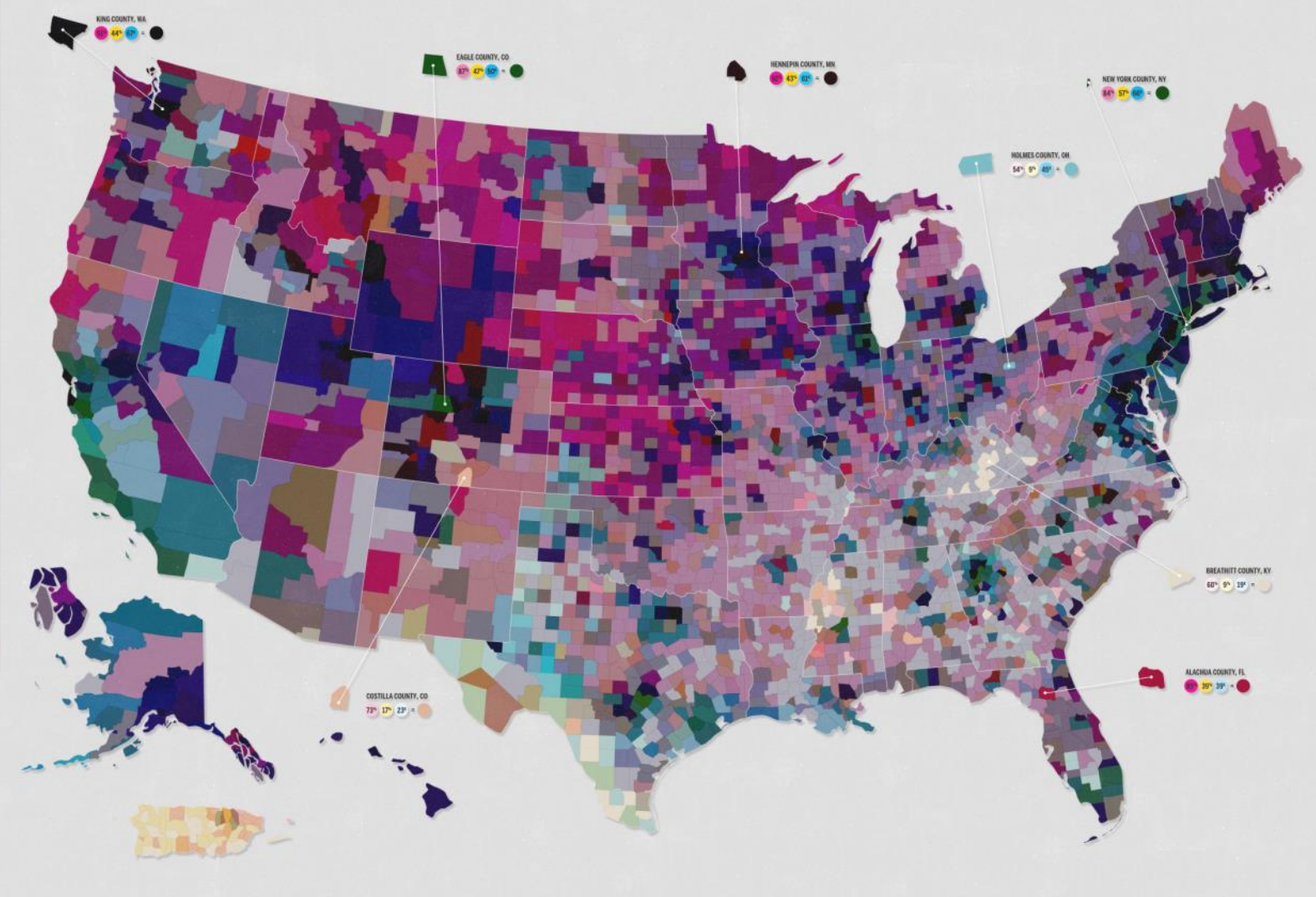
READING, WRITING, AND EARNING MONEY

The latest data from the U.S. Census's American Community Survey paints a fascinating picture of the United States at the county level. We've looked at the educational achievement and the median income of the entire nation, to see where people are going to school, where they're earning money, and if there is any correlation.

- A HIGH SCHOOL GRADUATES 60% 75% 82%
- B COLLEGE GRADUATES 15% 22% 30% 40%
- C MEDIAN HOUSEHOLD INCOME 20% 40% 50% 60%

The map to the right is a product of overlaying the three sets of data. The variation in hue and value has been produced from the data shown above. In general, darker counties represent a more educated, better paid population while lighter areas represent communities with fewer graduates and lower incomes.

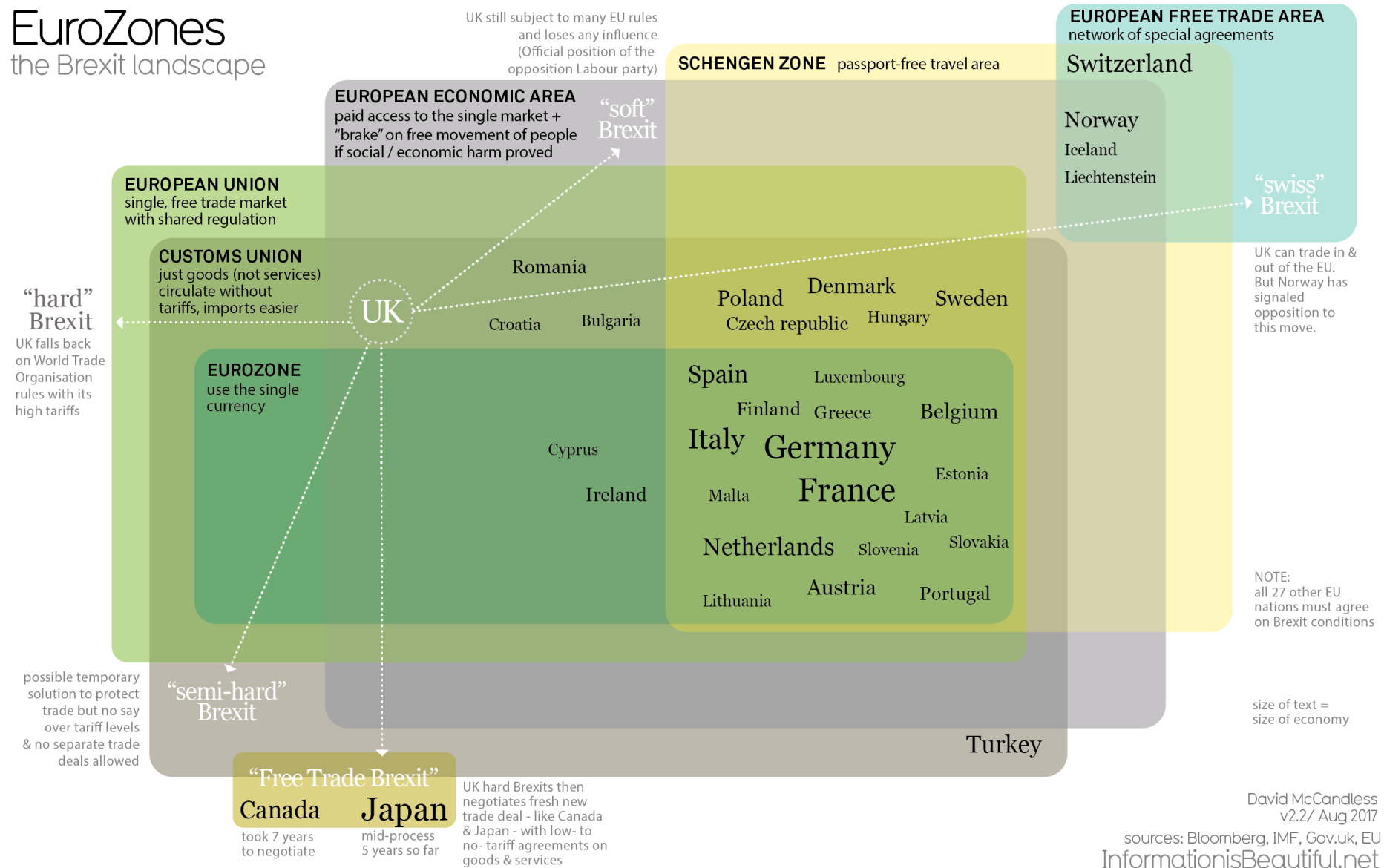
A collaboration between GOOD and Gregory Hubacek
SOURCE: US Census



Good examples

Information is Beautiful - David McCandless

EuroZones the Brexit landscape



David McCandless
v2.2/ Aug 2017

sources: Bloomberg, IMF, Gov.uk, EU
InformationisBeautiful.net

Information is Beautiful - David McCandless

Ted talk

https://www.ted.com/talks/david_mccandless_the_beauty_of_data_visualization

Information is Beautiful Awards

<http://www.informationisbeautiful.net/2017/the-kantar-information-is-beautiful-awards-2017-is-now-open/#more-4727>

Spatial image - Facebook

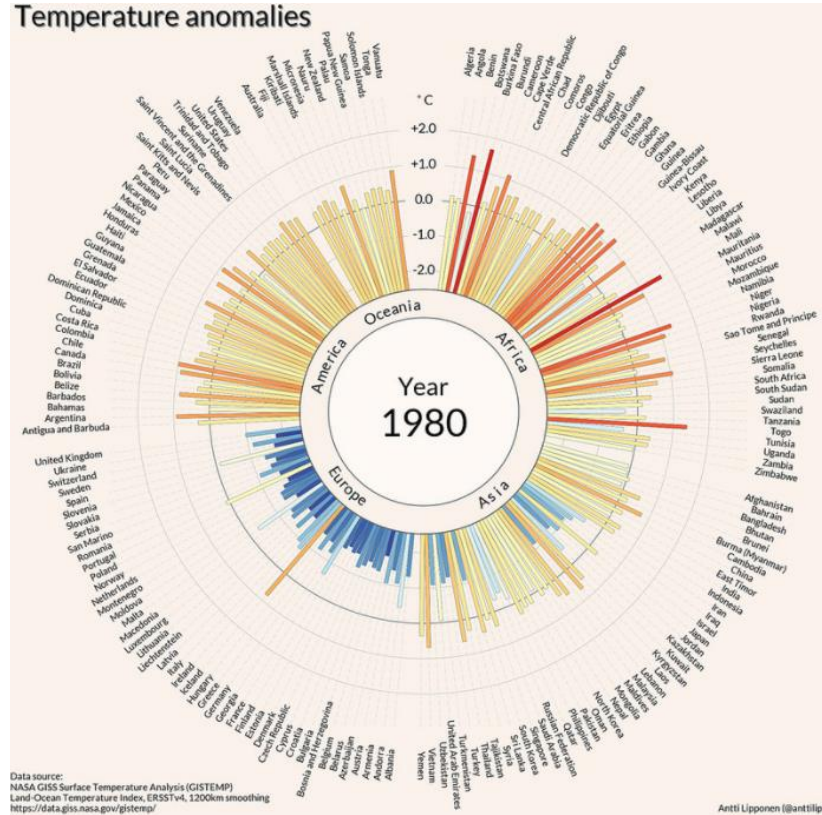


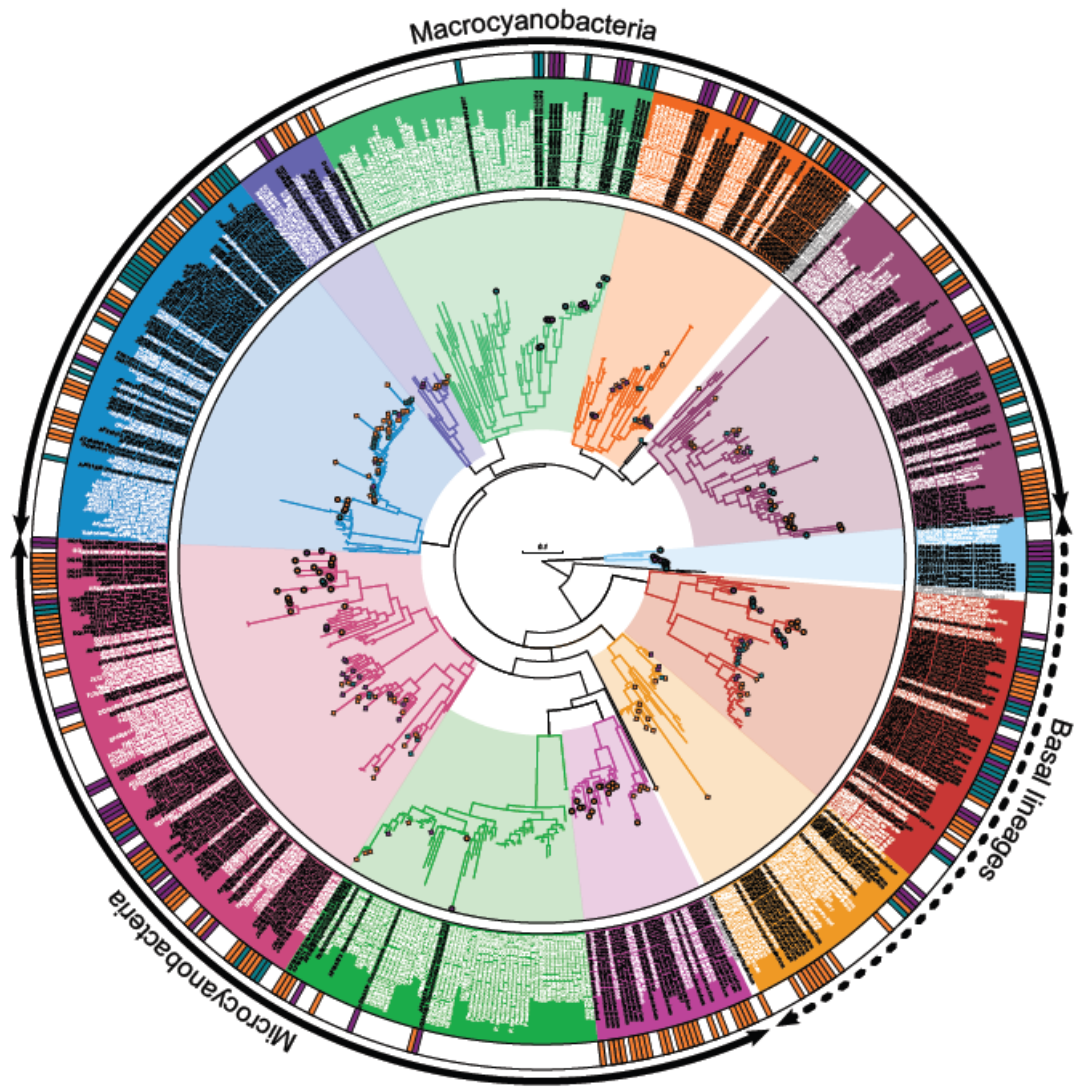
<https://dabrownstein.files.wordpress.com/2014/02/paul-butlers-map-of-friendships.jpg>

The power of animation

Temperature anomalies:

<https://www.flickr.com/photos/150411108@N06/35471910724/>





Key to sub-clades:

Gloeocapsa*
Pseudanabaena†
Acaryochloris, Aphanocapsa, Thermosynechococcus*
Leptolyngbya (Group I)†
Synechococcus, Prochlorococcus, Cyanobium†
Leptolyngbya (Group II), Phormidesmis, Chamaesiphon†
Trichodesmium, Phormidium, Arthrospira, Microcoleus†
Wilmozia†
Microcystis, Cyanothece, Pleurocapsa, Snowella etc*
Gloeocapsa, Chroococcidiopsis
Calothrix, Anabaena, Nodularia, Nostoc etc†

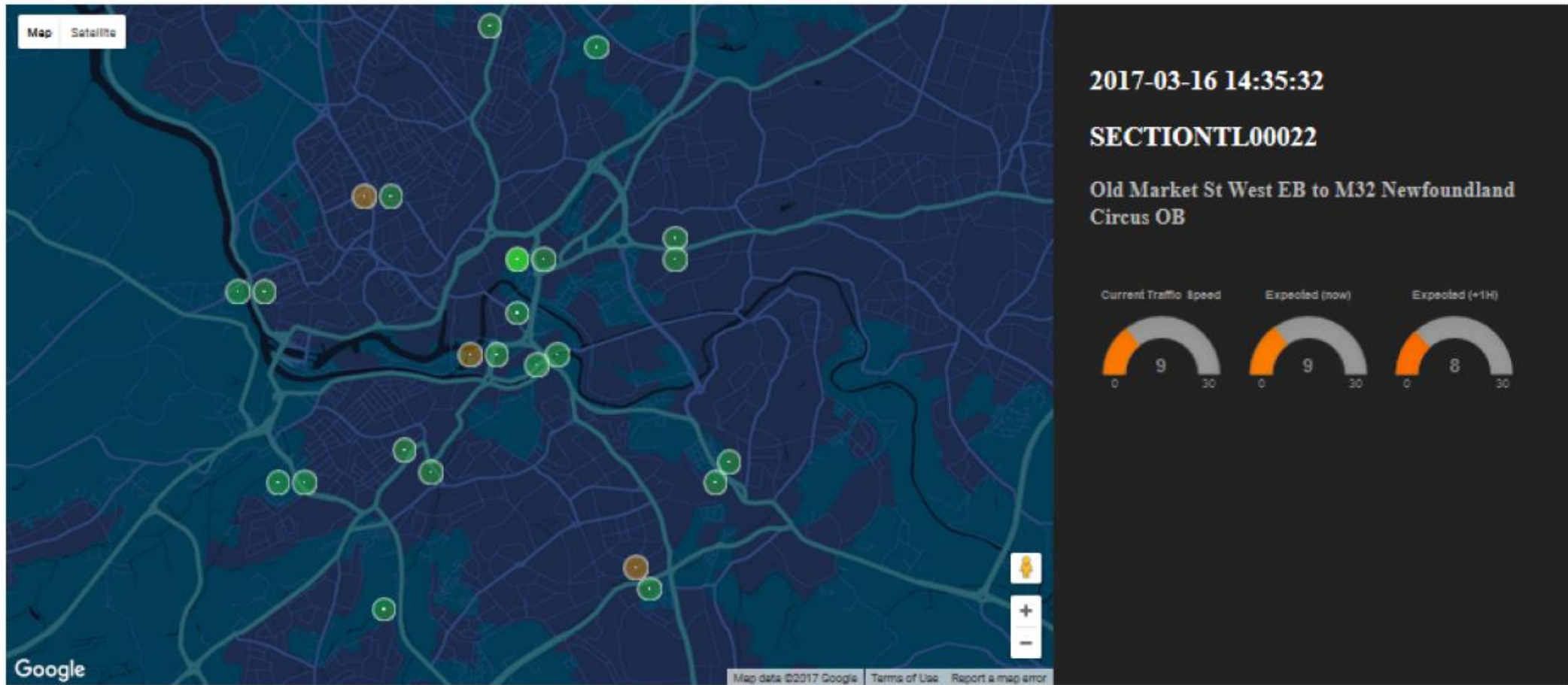
* = unicellular and colonial lineages
 † = filamentous lineages

Key to locations:

High Arctic, Northern high latitude
Alpine, high altitude
Antarctica, Southern high latitude

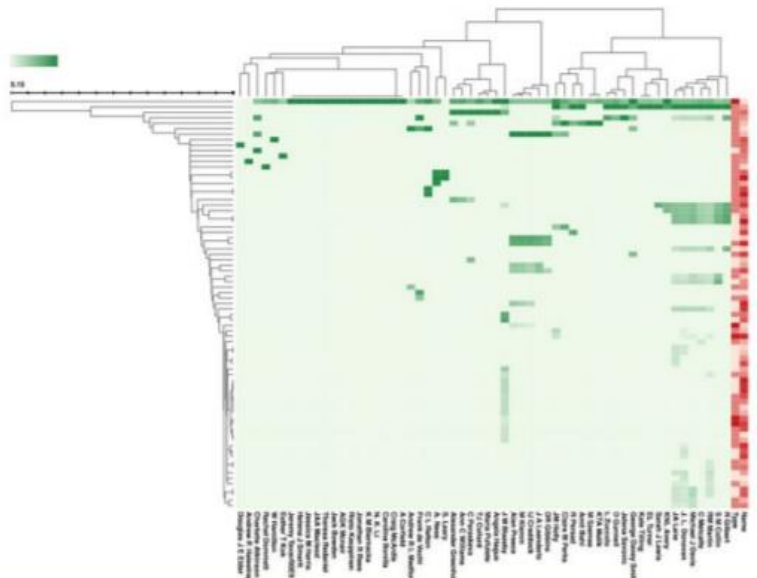
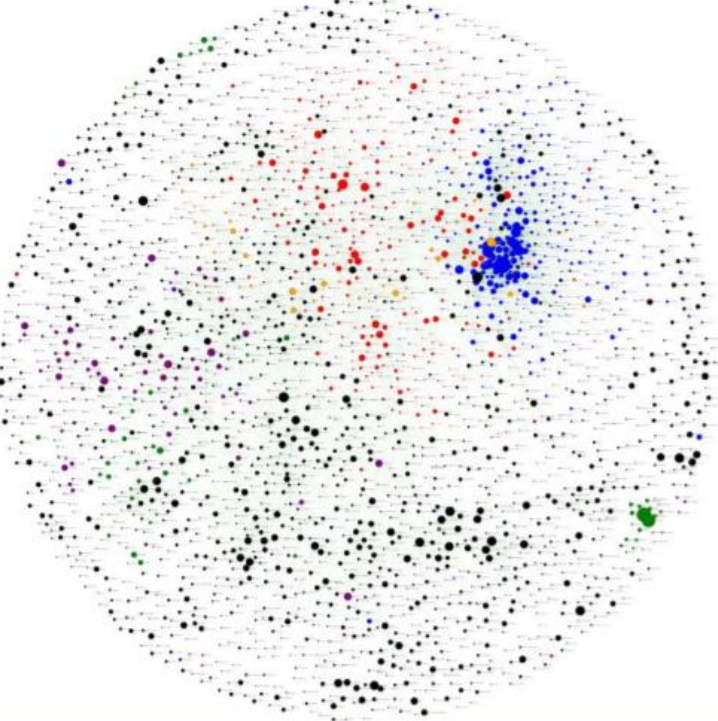
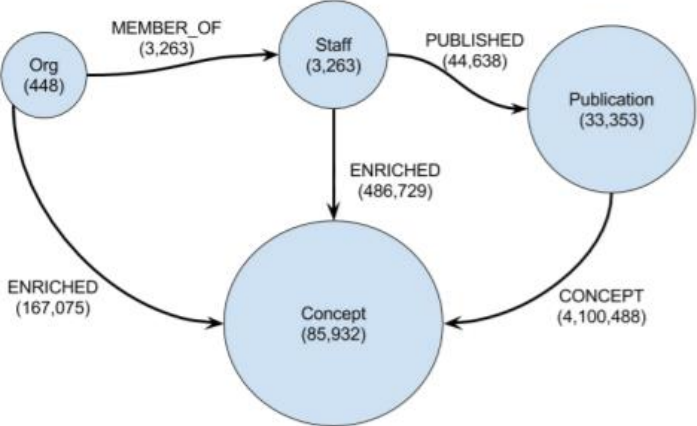
Key to major groups:

Monophyletic
Paraphyletic



Tom Bland & Kathryn Leeming

Networks and graphs - Ben Elsworth - MRC Integrative Epidemiology Unit, SSCM



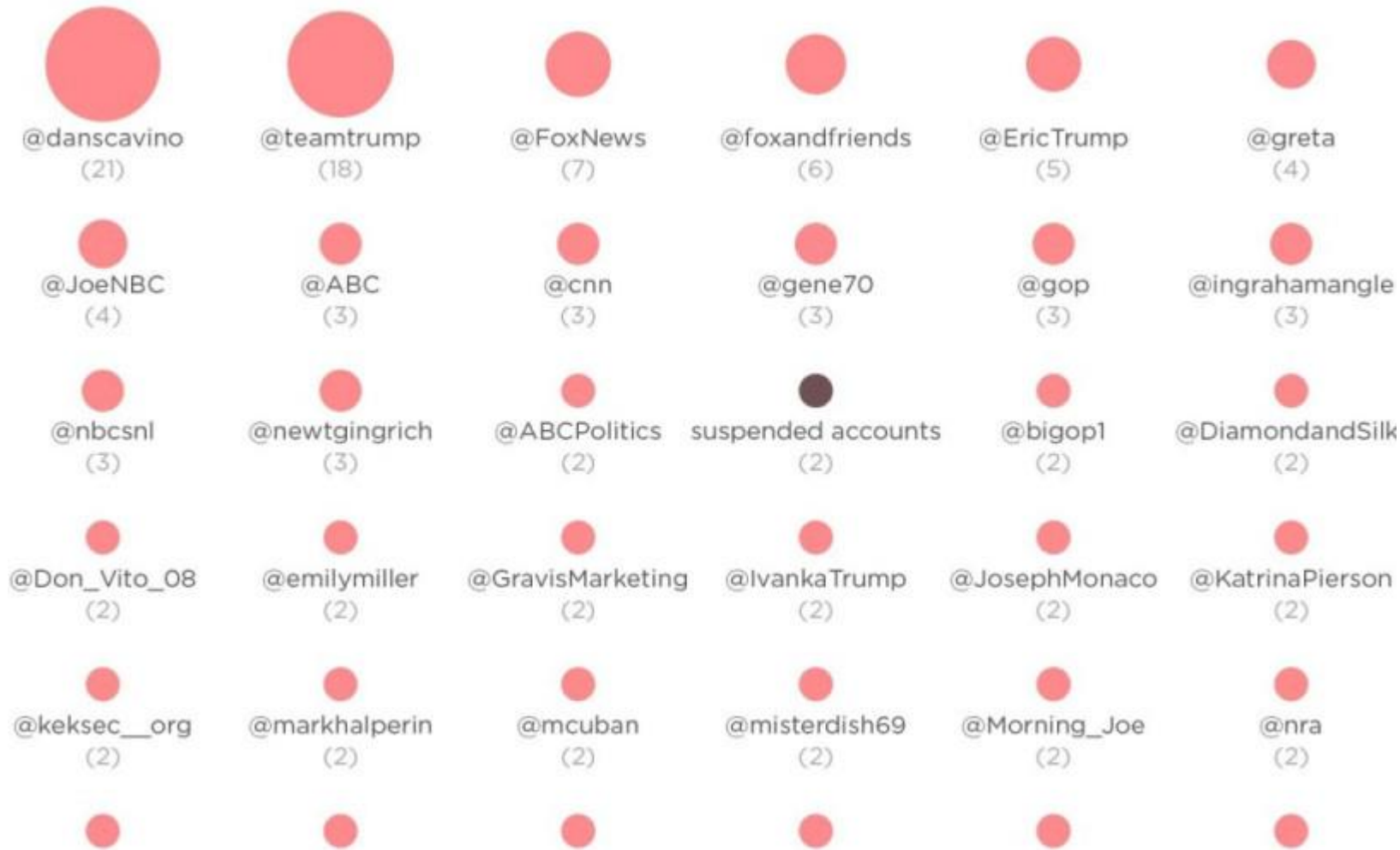
Small groups discussion

Makeover Monday

“Each week we post a link to a chart, and its data, and then you rework the chart.”

<http://www.makeovermonday.co.uk/>

Makeover Monday – Trump’s tweets



Week 3 – January 16th 2017

https://www.buzzfeed.com/harliewarzel/here-are-the-294-accounts-donald-trump-retweeted-during-the?utm_term=.ijAgeyPd7l#.ch0GJX2MOV

Makeover Monday – Trump's tweets

Trump and his thumb

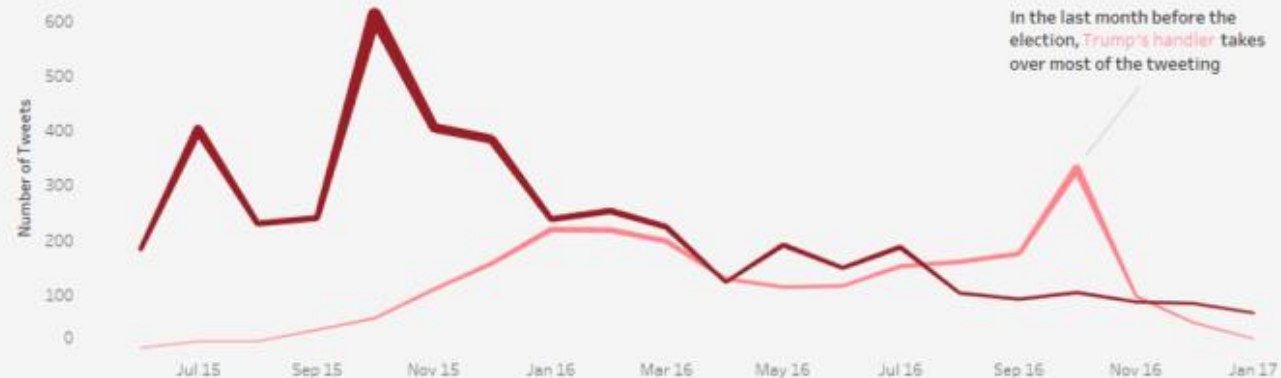
The Tweets from @realDonaldTrump since announcing his candidacy on 16 June 2015

Number of Tweets	Tweets from Android & iPhone	Tweets related to Barack Obama	Tweets related to Hillary Clinton
8,125	6,728	234	709

Tweets from mobile: **Android & iPhone**



As the election campaign progresses, **Trump's handler** appears to take on more of the tweeting while **Trump's own tweets** are gradually becoming less frequent



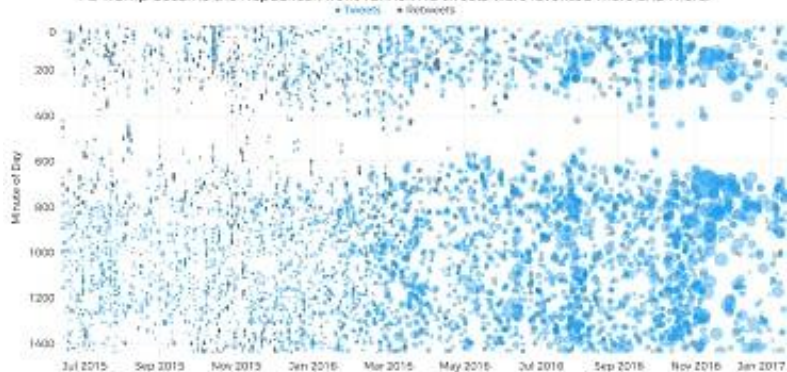
Interactive version:

[https://public.tableau.com/profile/eva.murray#!/vizhome/MakeoverMondayWeek3-
Trumpandhisthumb/Trumpandhisthumb](https://public.tableau.com/profile/eva.murray#!/vizhome/MakeoverMondayWeek3-Trumpandhisthumb/Trumpandhisthumb)

The Tweeting Habits of President-Elect Trump

Tweets from @realDonaldTrump since announcing his candidacy on 14 June 2015

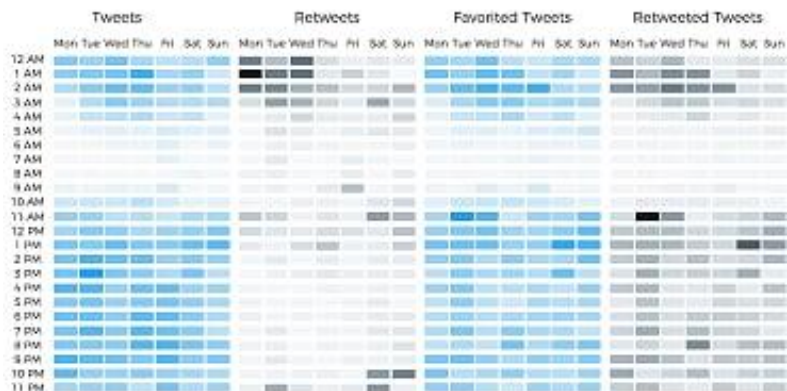
As Trump became the Republican front-runner, his tweets were favorited more and more.



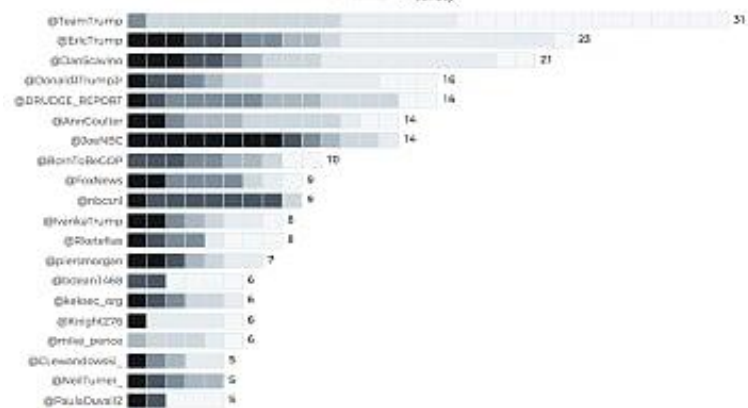
Makeover Monday – Trump's tweets

When is Trump most likely to tweet or retweet?

When do his tweets get favorited or retweeted?



Trump is most likely to retweet those from his inner circle and friendly members of the press



Interactive version:

<http://www.vizwiz.com/2017/01/makeover-monday-tweeting-habits-of.html>

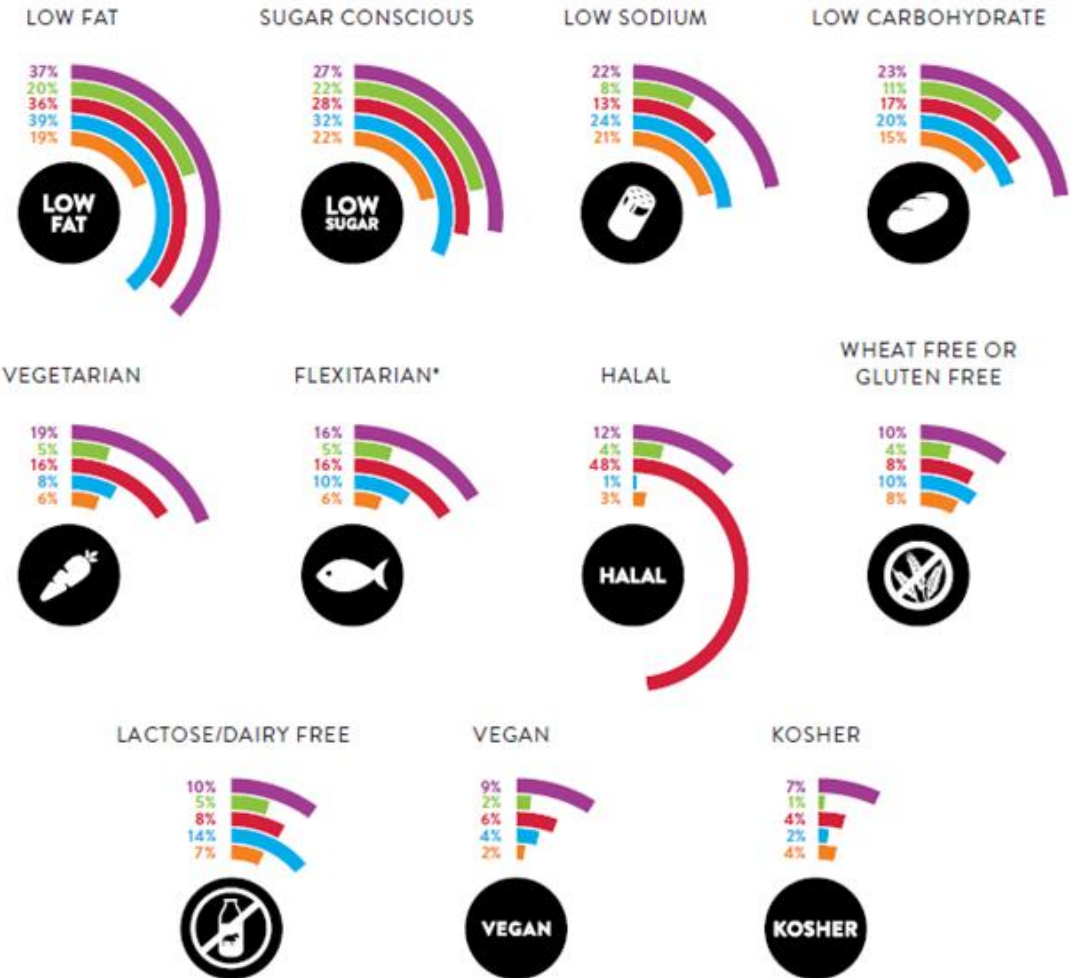
Makeover Monday – this week

- Group work – discuss chart and suggest new version
- <http://www.makeovermonday.co.uk/data/>
- "What's in our food and on our mind – ingredient and dining out trends around the world" ([Nielson report](#))

RESTRICTED DIETARY REQUIREMENTS AROUND THE GLOBE

PERCENTAGE WHO SAY THEY FOLLOW A SPECIAL DIET THAT LIMITS OR RESTRICTS SPECIFIED FOODS OR INGREDIENTS

● ASIA-PACIFIC ● EUROPE ● AFRICA/MIDDLE EAST ● LATIN AMERICA ● NORTH AMERICA



Source: The Nielsen Global Health and Ingredient-Sentiment Survey, Q1 2016

*Primarily vegetarian, but occasionally eat meat, poultry or seafood

Working Group format

Questions

- Monthly or bimonthly?
- Format?
- Mailing list or Yammer group?

Topics people suggested they could lead

Topics:

- Graph databases – networks etc
- 3D printing for data visualisation
- Bristol is Open
- Icons and infographics
- 3D plots
- Design and interaction design
- User experience (Psychological principles)

Specific projects:

- Developing visuals for an eBooks project
- Visualisations of text

Specific Tools:

- d3
- R
- Python (inc pandas and seaborn libraries)
- Adobe Illustrator
- SAP Lumira
- Virtual reality and interactive visualisation

Topics people suggested they would like

Topics:

- Spatial data
- Whole genome sequence and metagenomics data
- Multi-dimensional data
- High-dimensional data
- Improving UX for web applications
- Infographics
- Visualising processes - e.g. a work flow
- Visualising time series data
- Visualisations for model checking and model development
- Interactive/animated visualisations for large data
- Reducing “big” data in a meaningful way in order to visualise, while retaining insight
- Communicating uncertainty
- Bristol Data Dome
- Patient data

Specific Tools:

- Javascript – particularly d3
- R
- Python
- Tableau
- Virtual Reality
- Adobe illustrator

Discussion:

- What do people mean by data visualisation?
- What do people need to help develop their data visualisations?
- Funding strategies for data visualisation projects
- Data visualisation for the social sciences
- Design: how to make stylish, intuitive and informative visuals
- Critical approaches to data visualisation

Oli Chalkley o.chalkley@bristol.ac.uk
Bobby Stuijfzand bg.stuijfzand@bristol.ac.uk
Harriet Mills harriet.mills@bristol.ac.uk

www.bristol.ac.uk/golding