



# Data Visualization

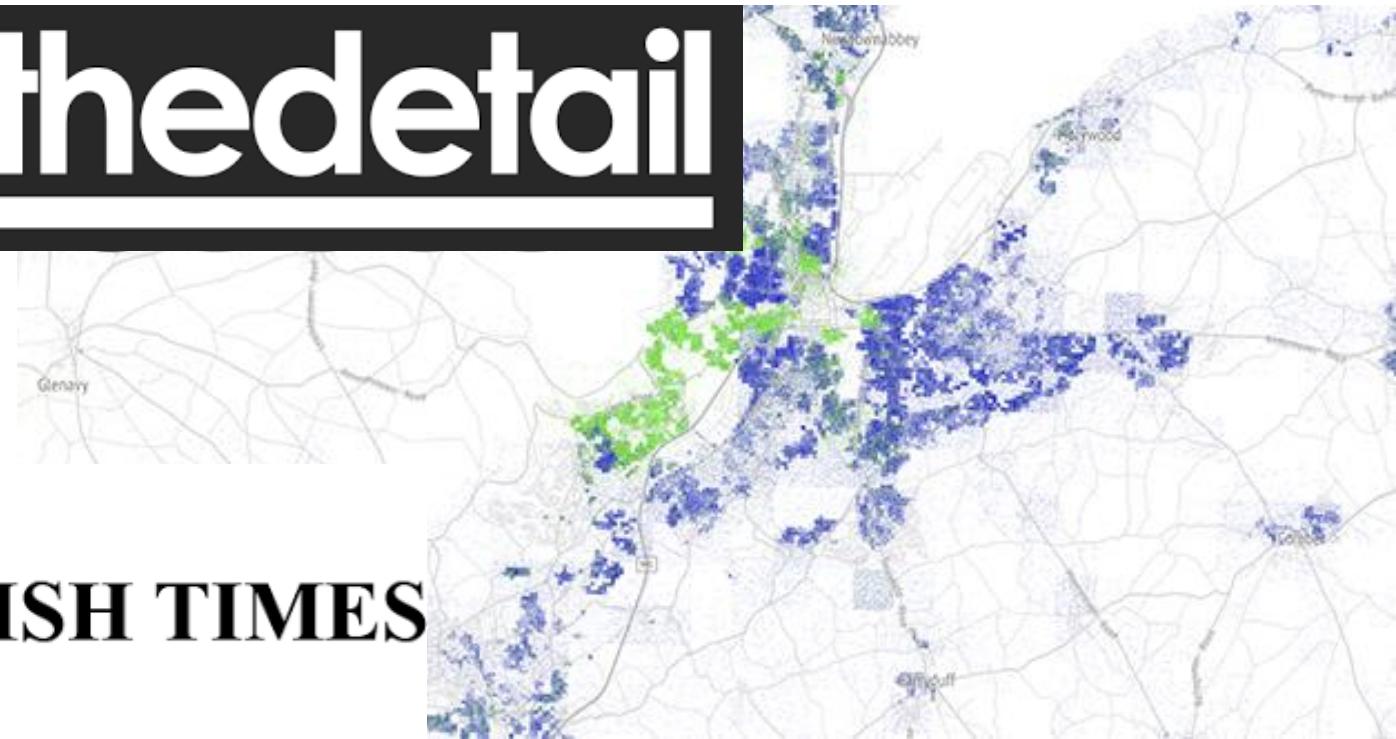
Storytelling with data



# Who am I?

- Matt Doherty
- Principle Engineer at AquaQ Analytics
- @picodoco
- <https://www.picodoc.org/>
- <https://github.com/picoDoc>

**thedetail**



# Intro

- (almost) nothing here about tech
  - Data visualization != software
  - It's about **storytelling**
  - Brings together the language of the eye, and of the mind
- 
- Some rules and guidelines, what to think about
  - Lots of examples (good and bad)
  - Some local (NI) stuff

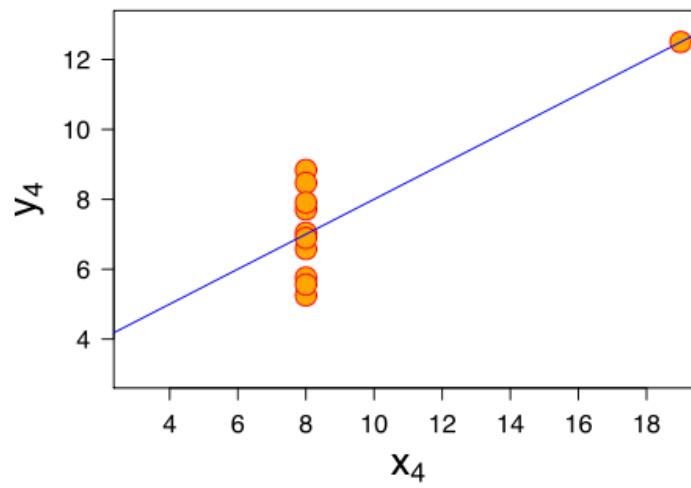
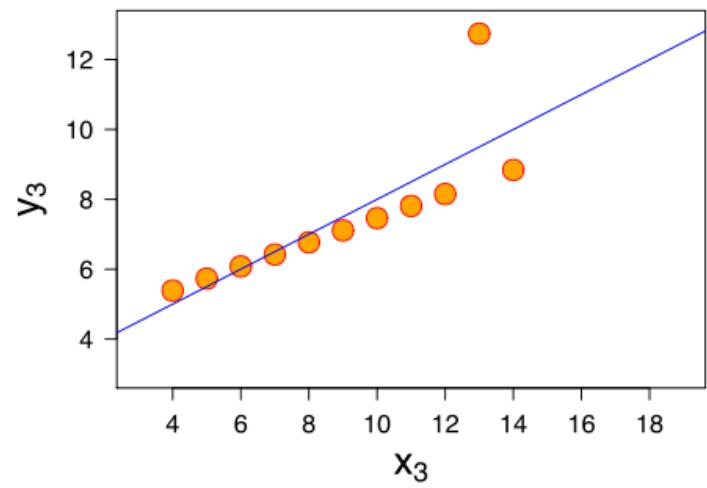
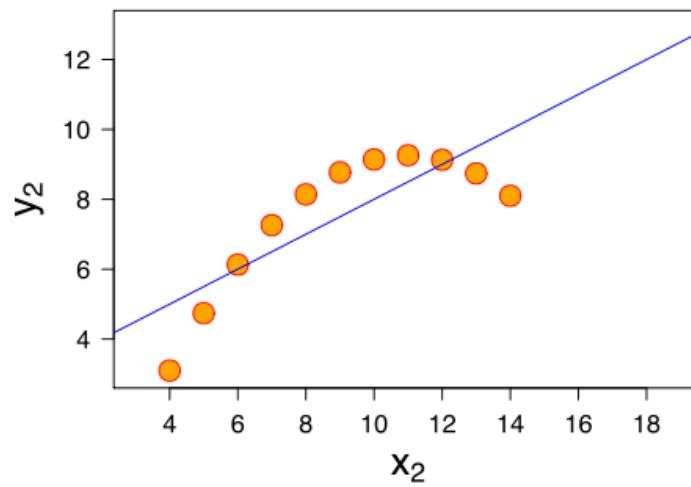
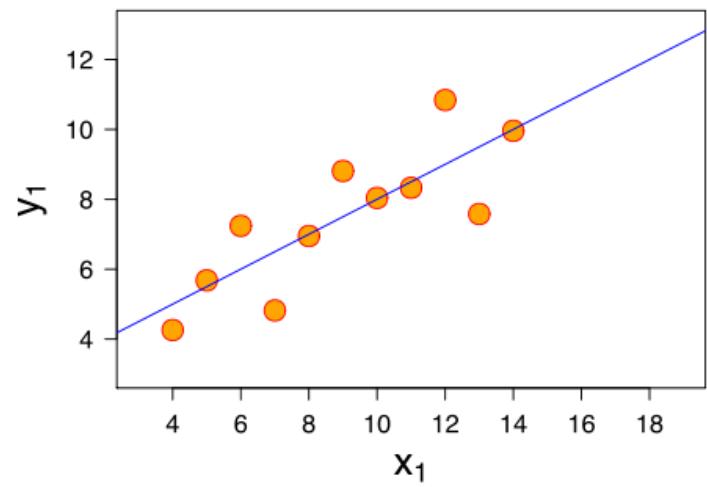
# Resources

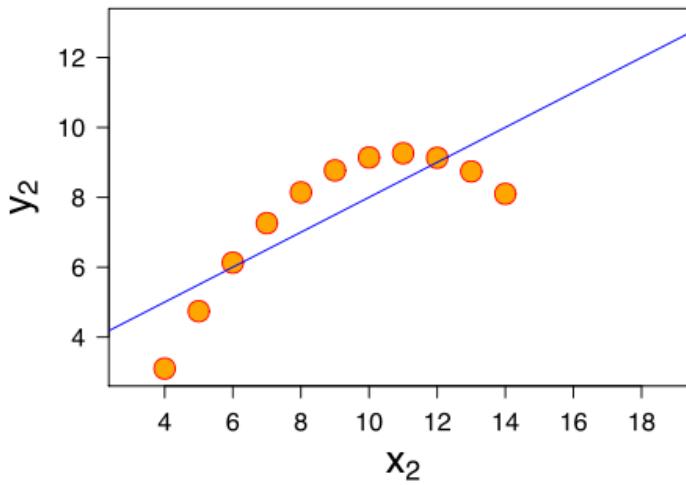
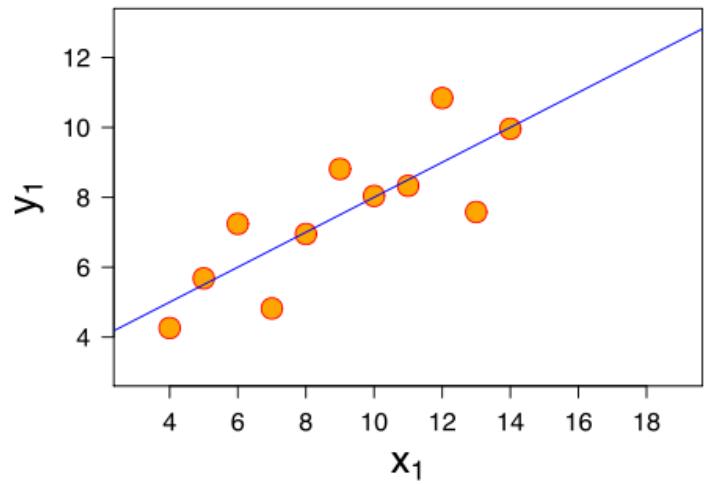
## Books:

- *The Visual Display of Quantitative Information*, E.R. Tufte (2001)

## Online learning:

- *Data Visualization* on Coursera
- *Data Visualization and d3.js* on Udacity



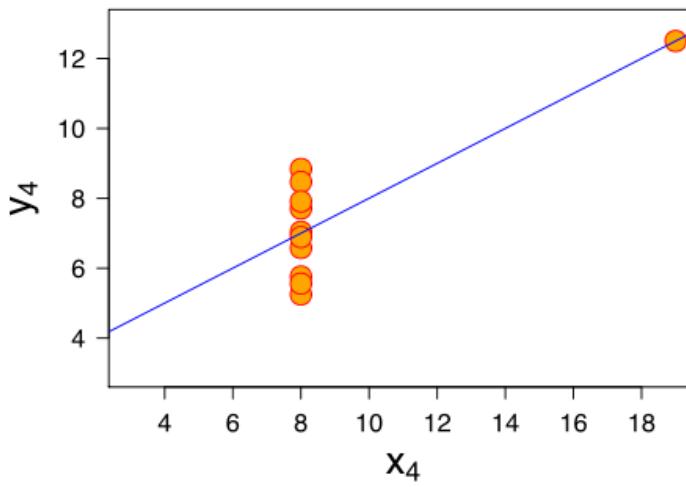
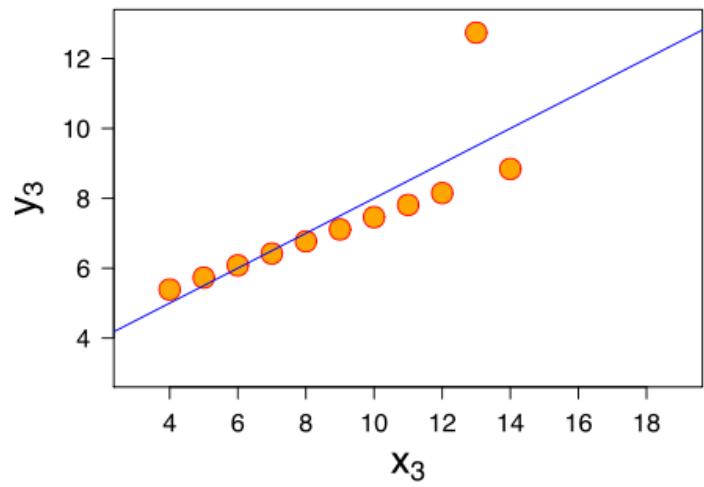


*mean is the same*

*variance is the same*

*correlation is the same*

*fit is the same*



# Visualization is storytelling

- Visualization is storytelling
- Storytelling with shapes and colours instead of words
- *No such thing as an objective visualization*
- You're always choosing what to show

# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)

# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)

**The more detail the story has the better**

**The more ‘things’ the user can see the better**

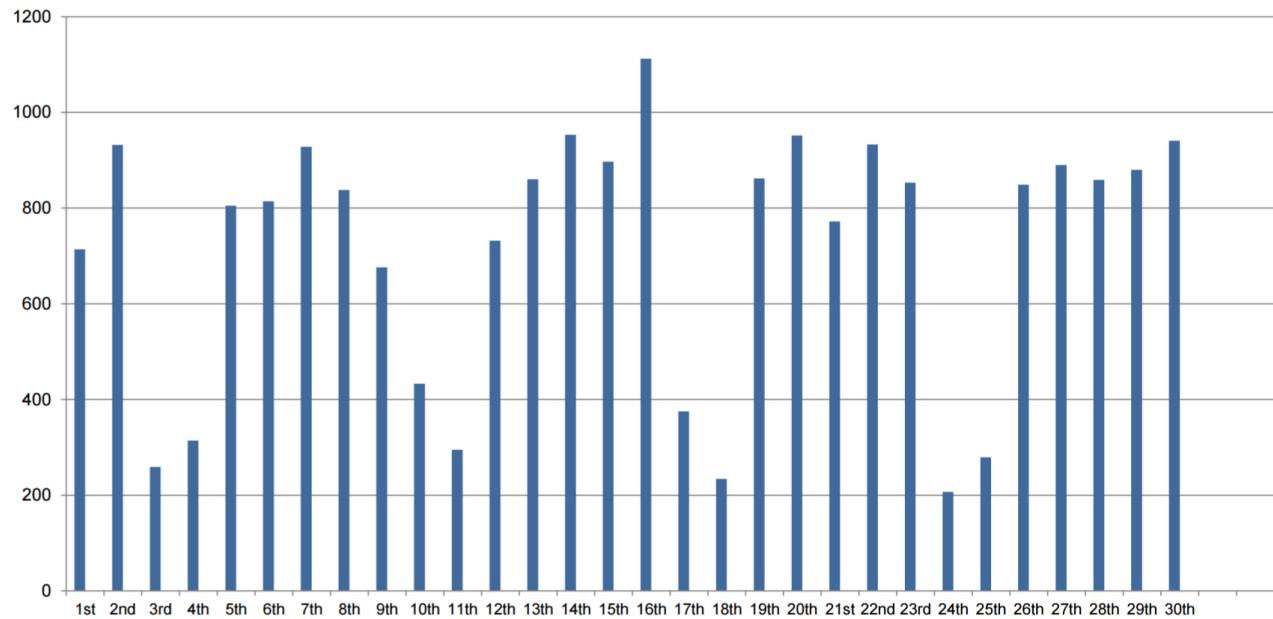
**Tool you’re using shouldn’t matter**

# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)

September 2016

During the month of September, there were 21,448 rentals. The rental fluctuations are shown below:

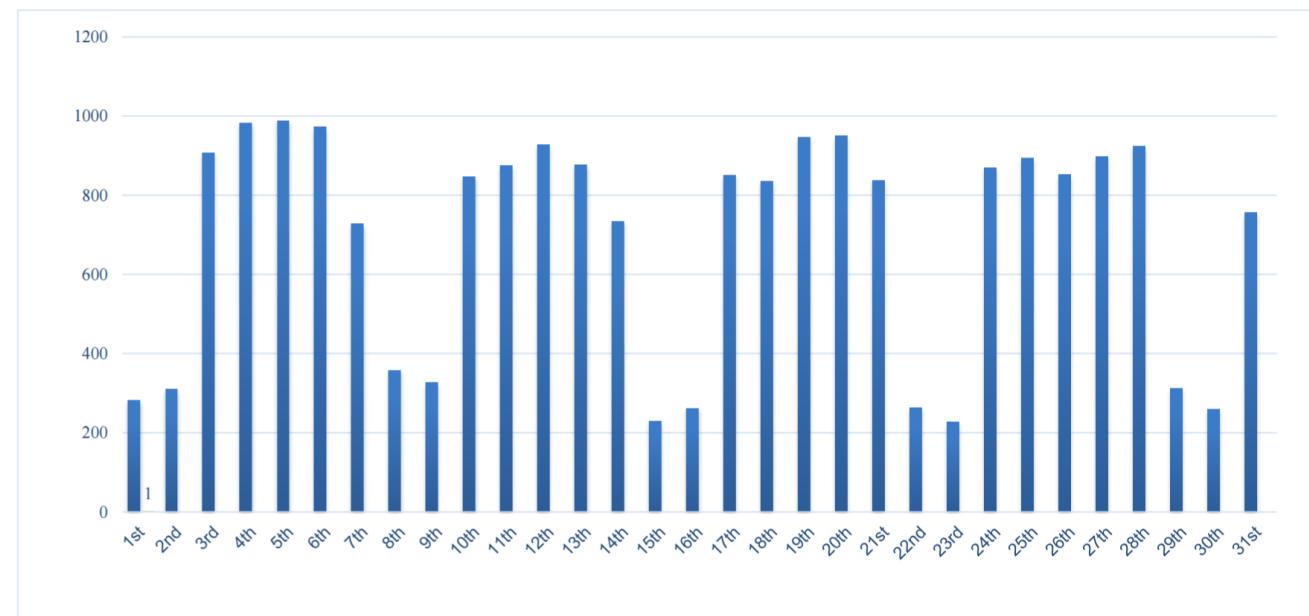


# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)

October 2016

During the month of October, there were **21,270** rentals. The rental fluctuations are shown below:

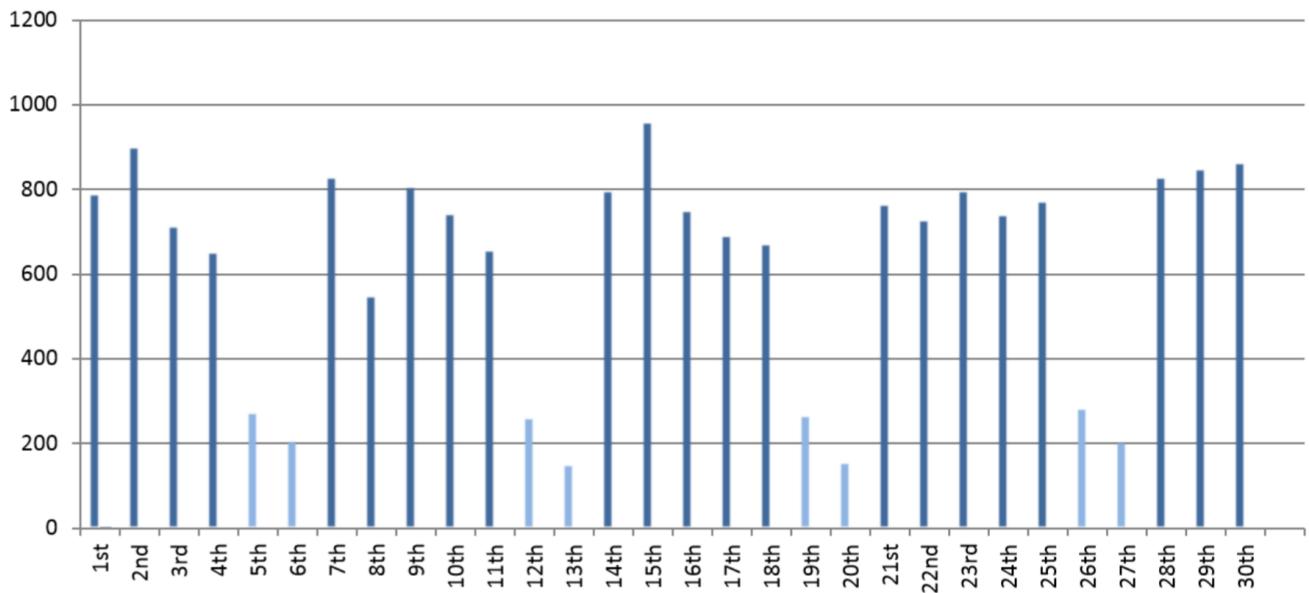


# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)

**November 2016**

During the month of November, there were **18,676** rentals. The rental fluctuations are shown below:

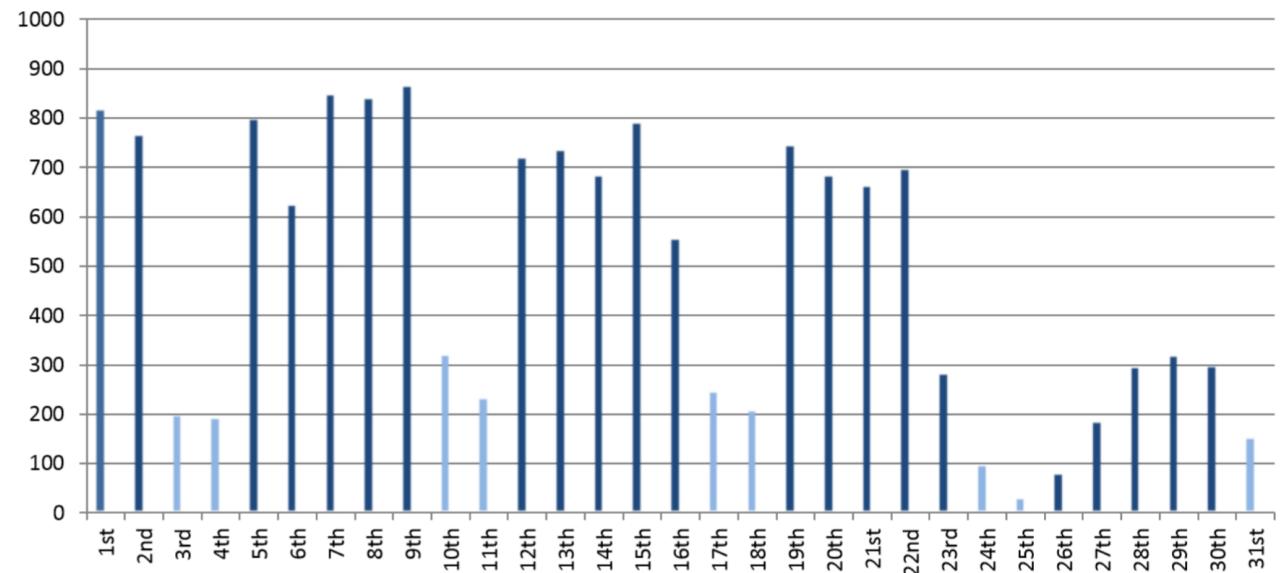


# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)

**December 2016**

During the month of December, there were **15,022** rentals. The rental fluctuations are shown below:



# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)

*“smart city my hole”*

anonymous Belfast open data advocate

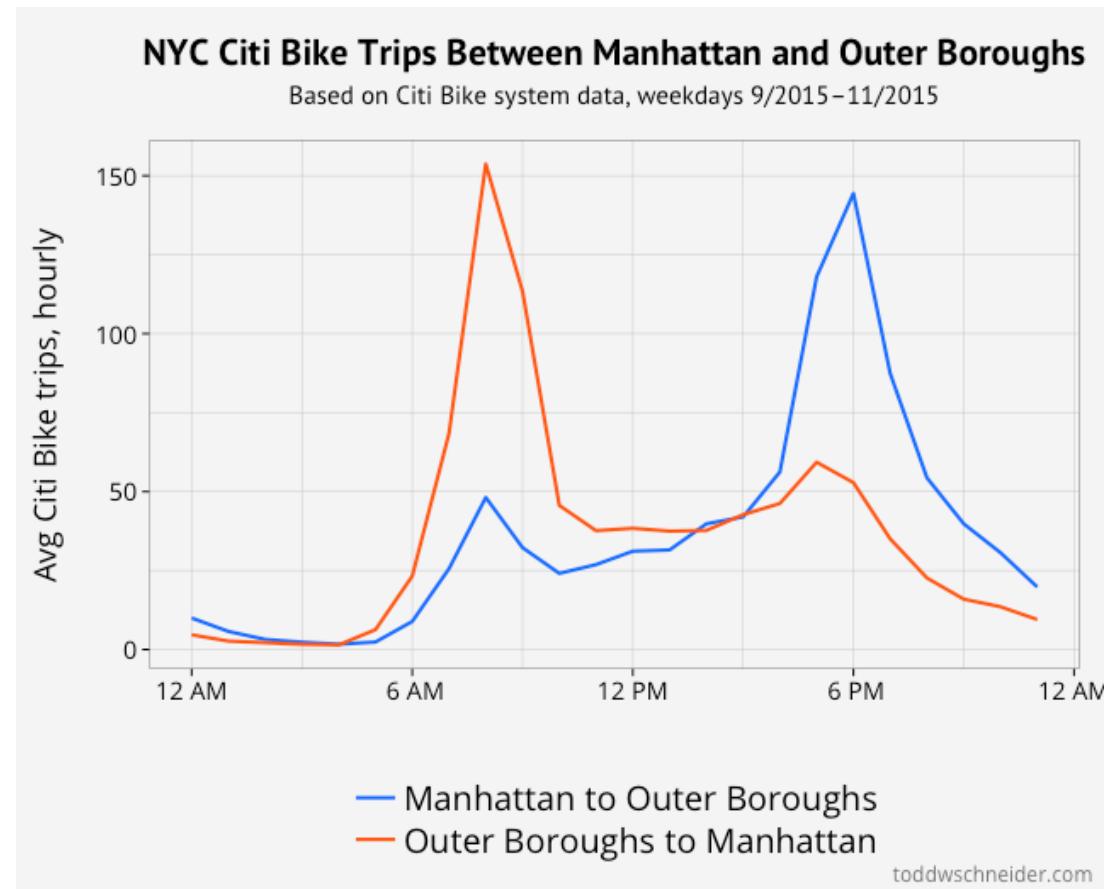
# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)

<http://toddwschneider.com/posts/a-tale-of-twenty-two-million-citi-bikes-analyzing-the-nyc-bike-share-system/>

# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



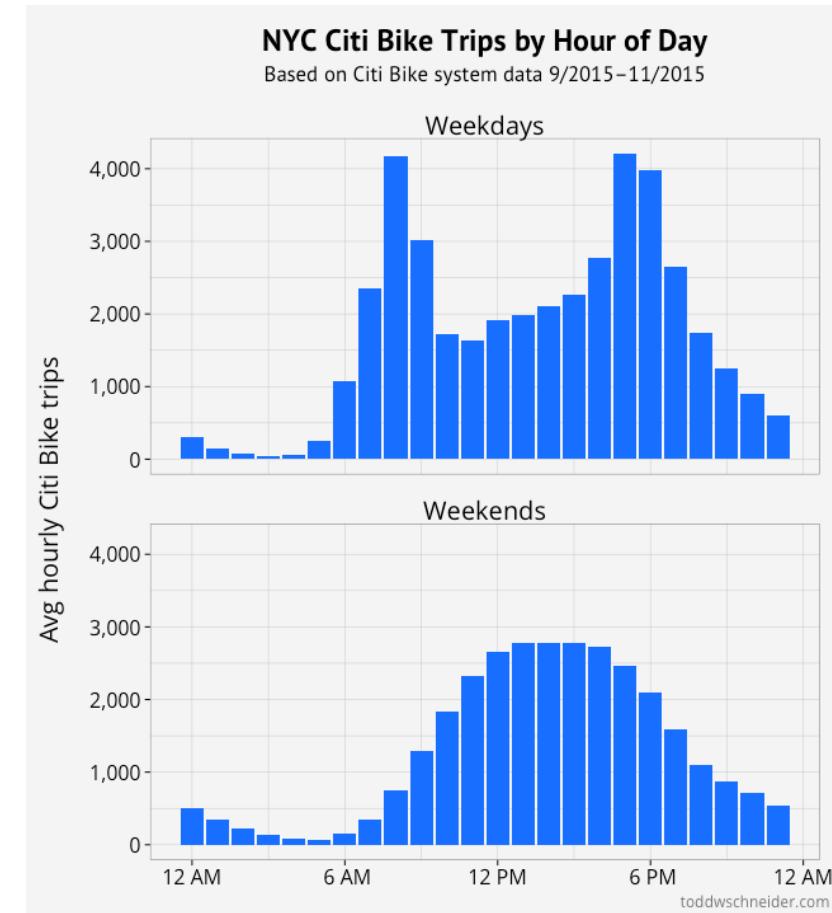
# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



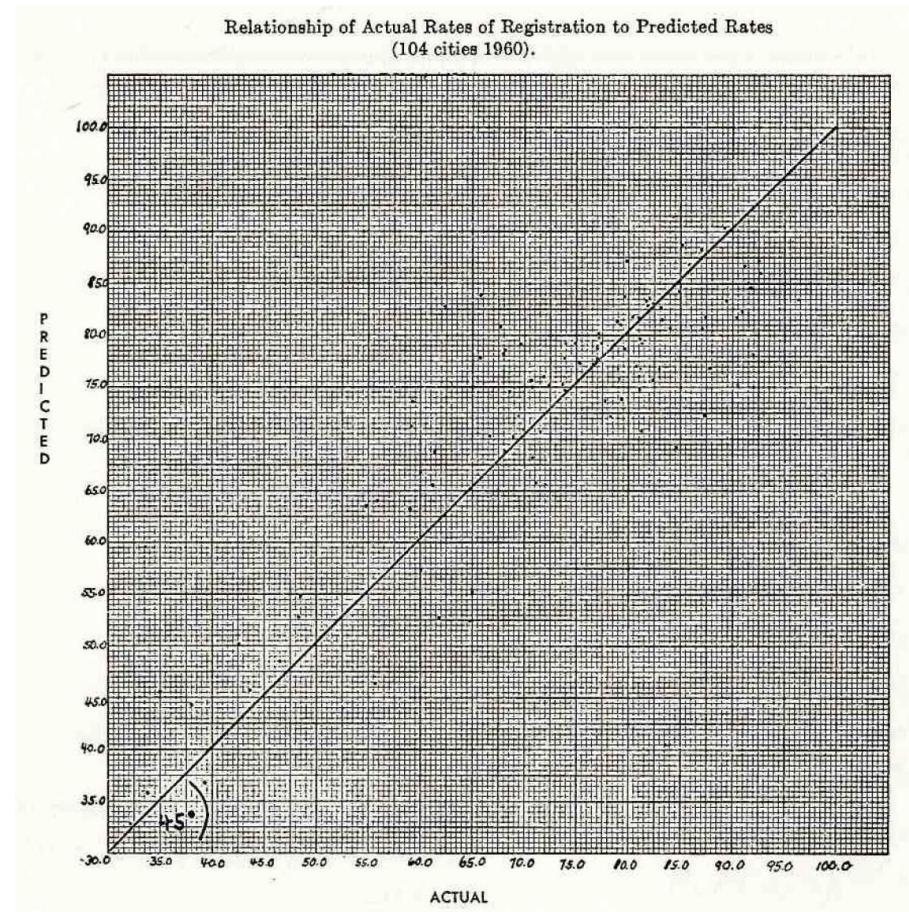
# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



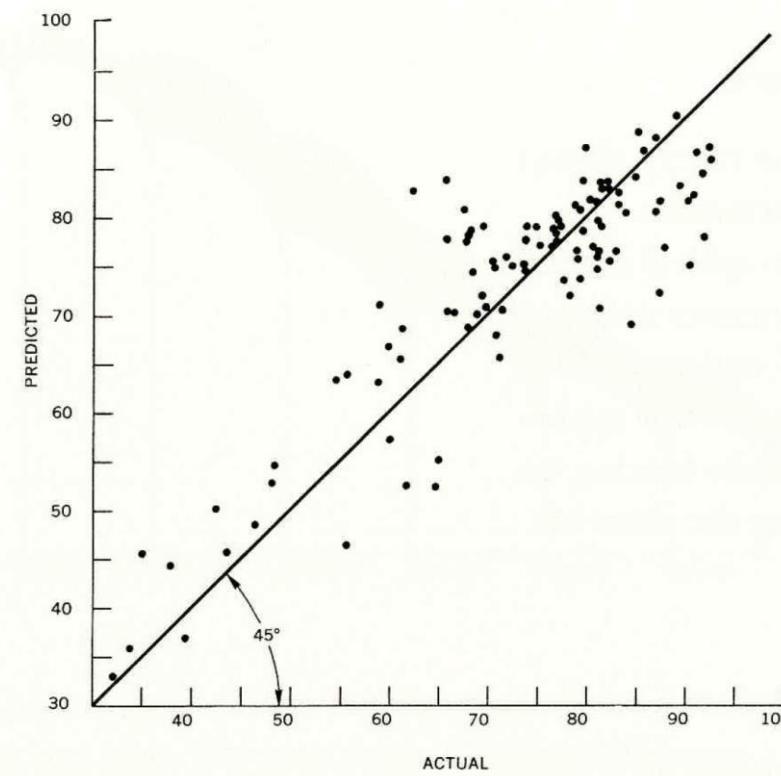
# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



# rules for good visualization

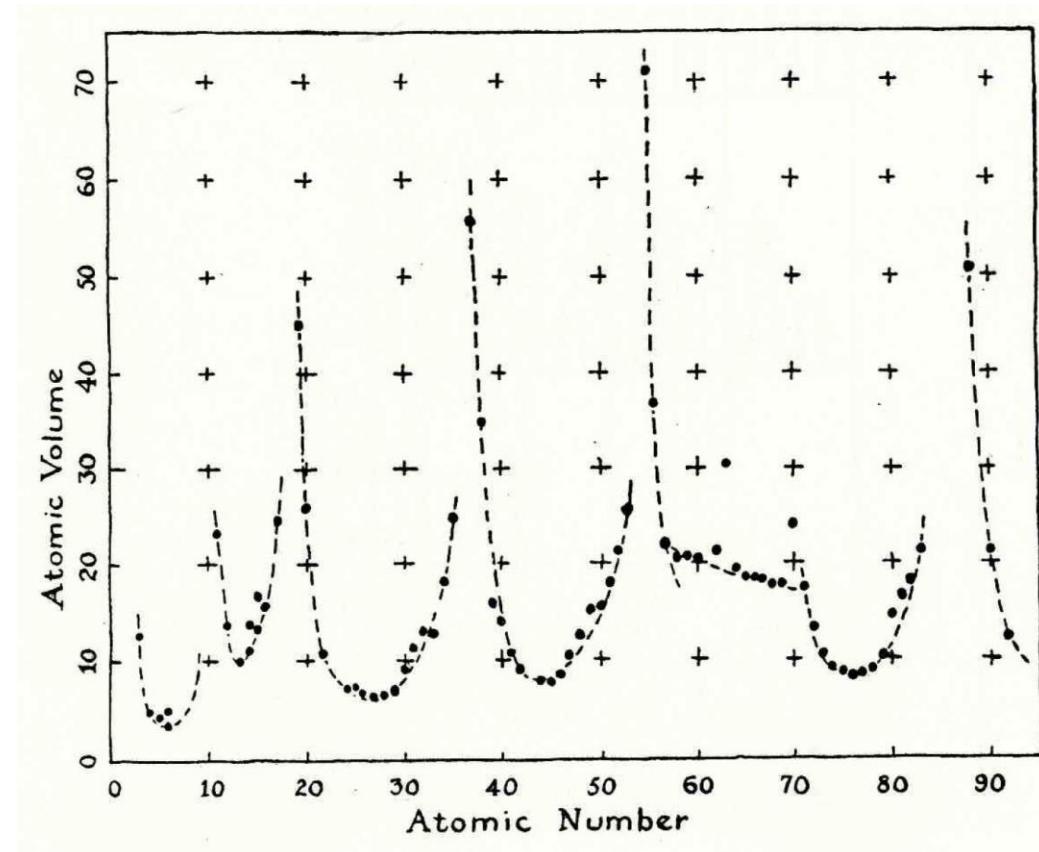
1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



Relationship of Actual Rates of Registration to Predicted Rates (104 cities 1960).

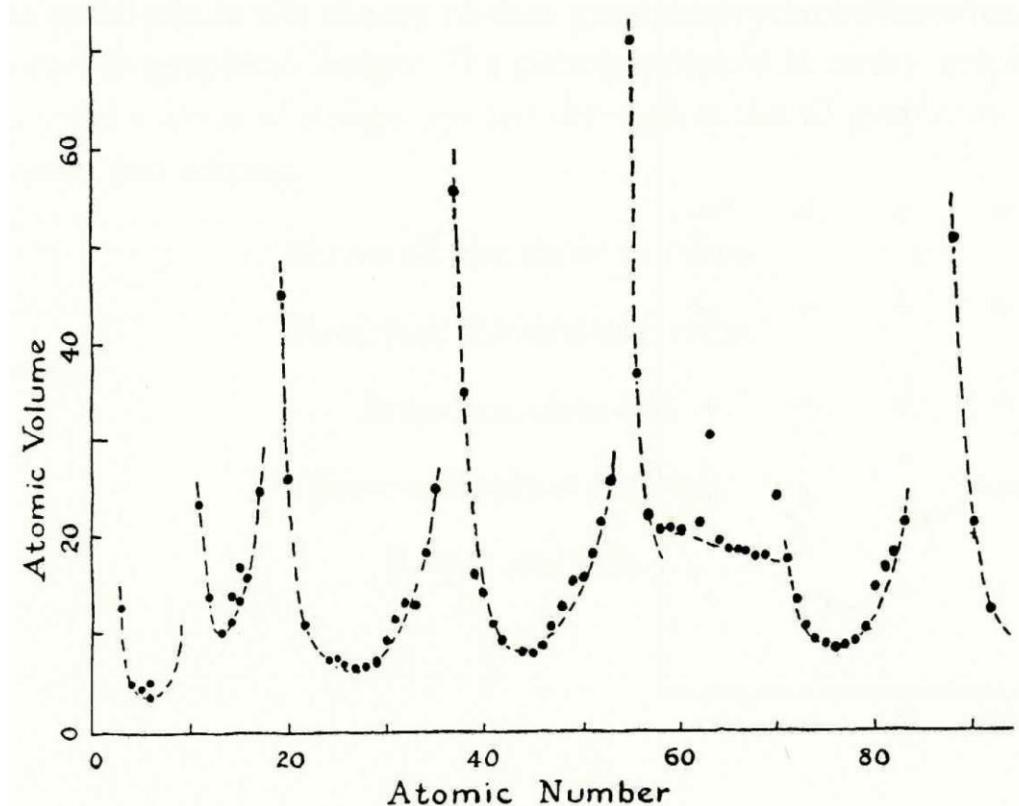
# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



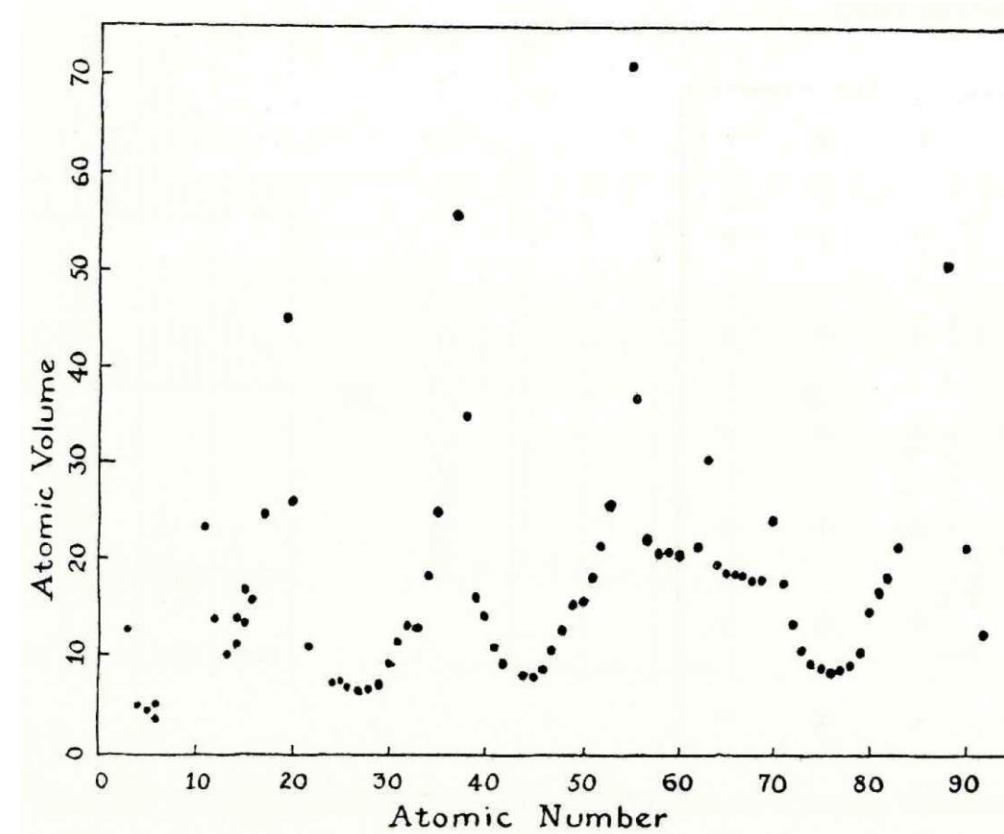
# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



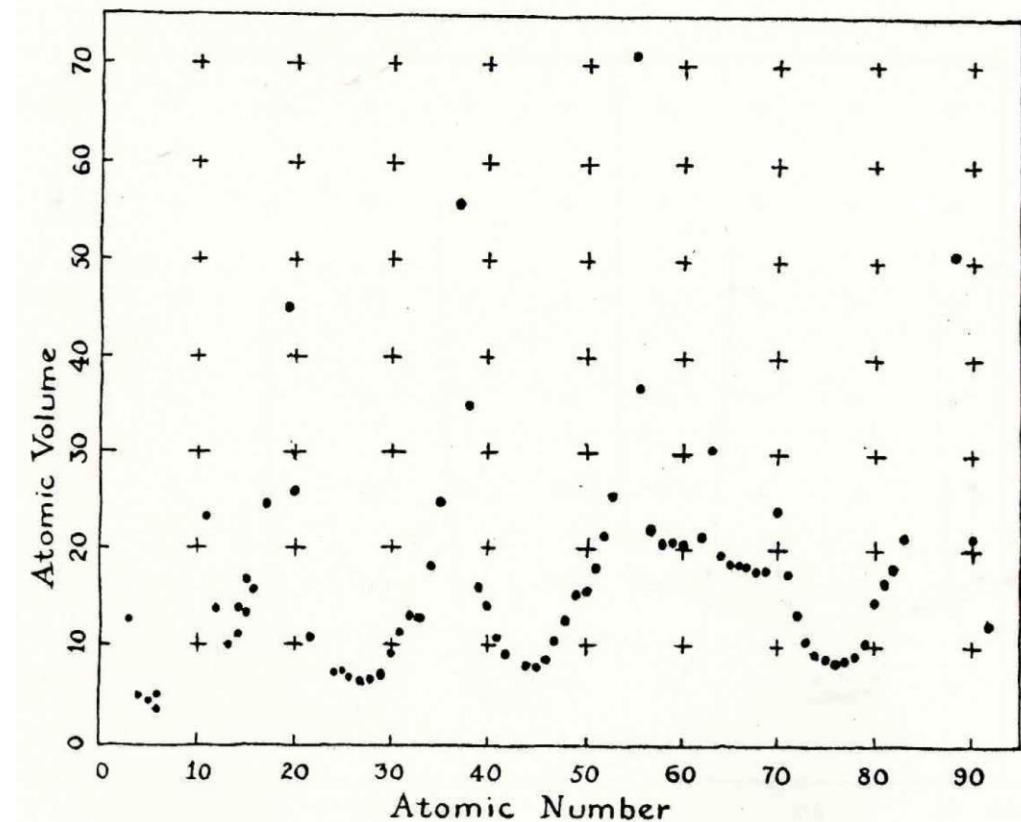
# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



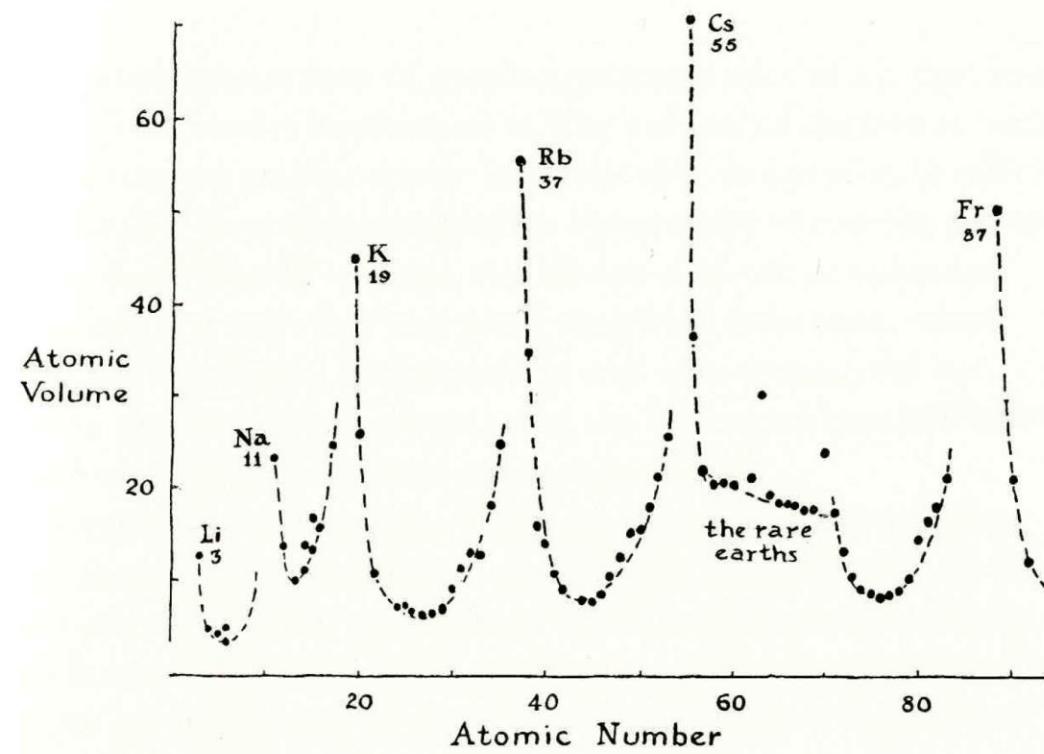
# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



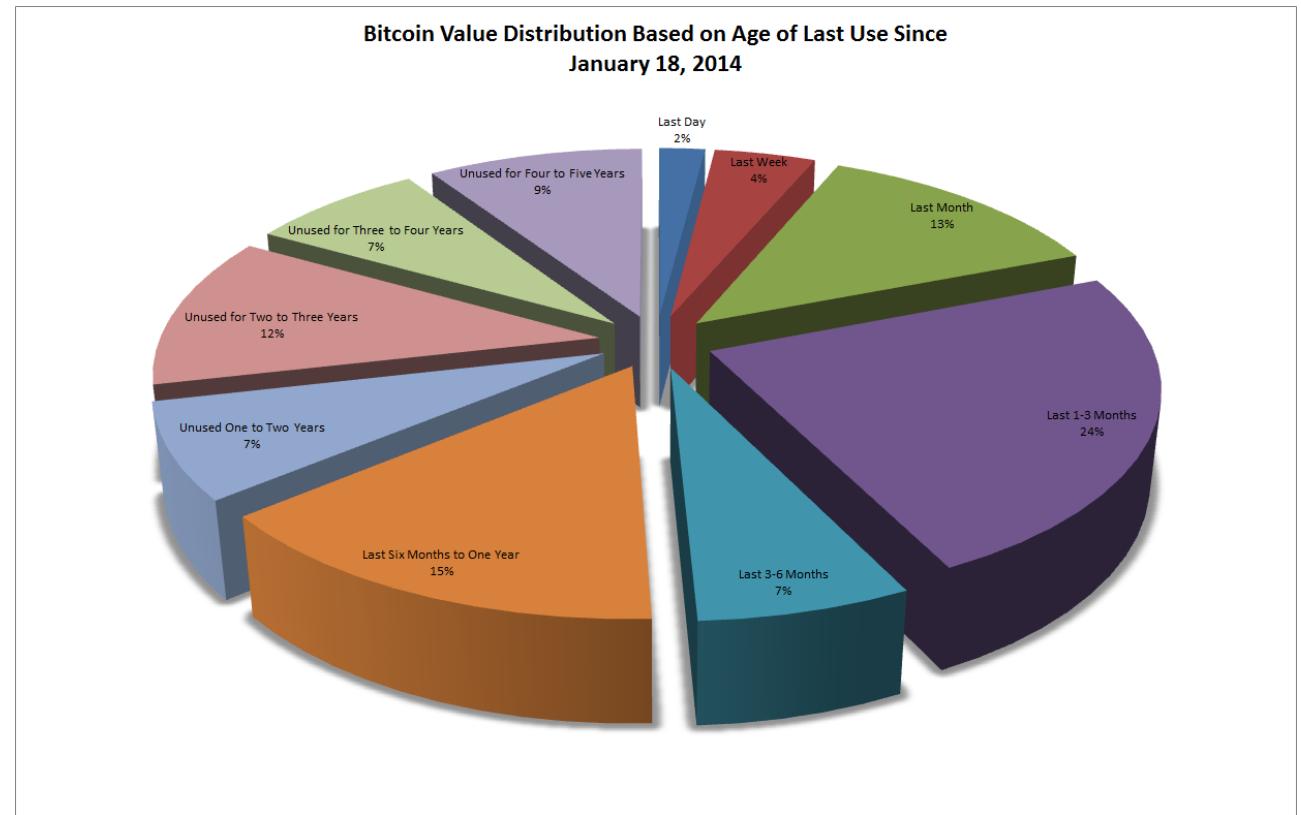
# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



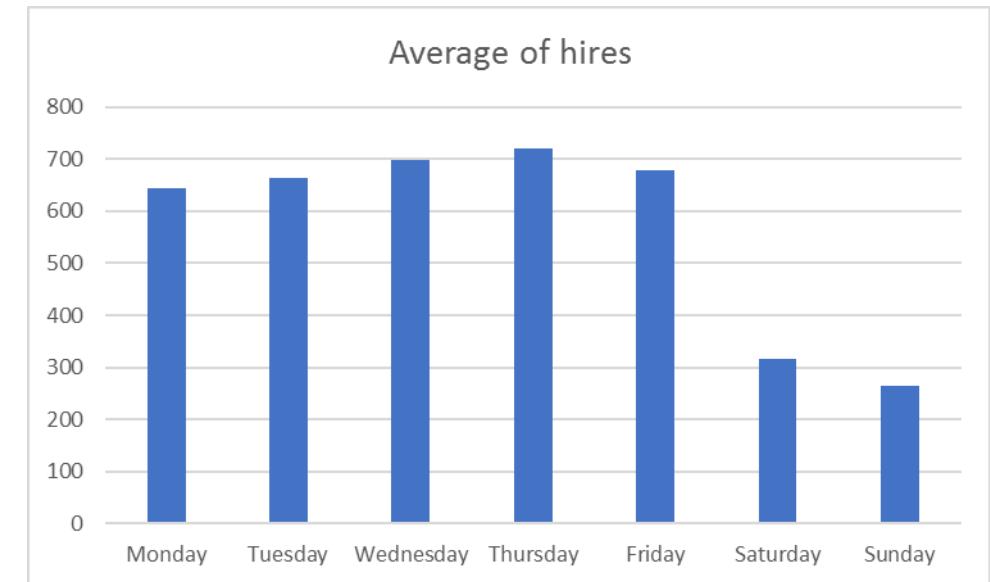
# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



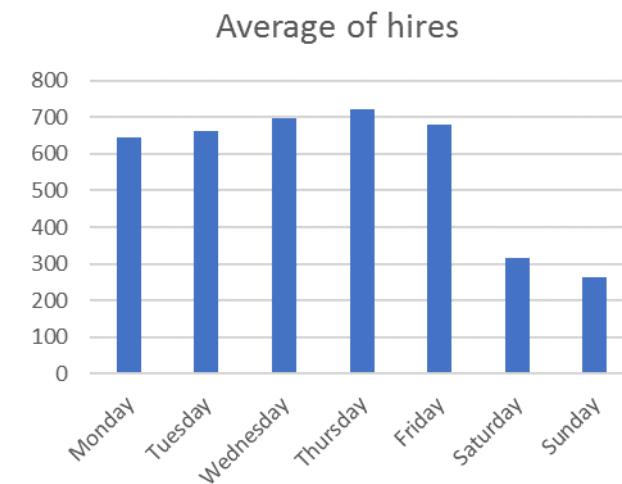
# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



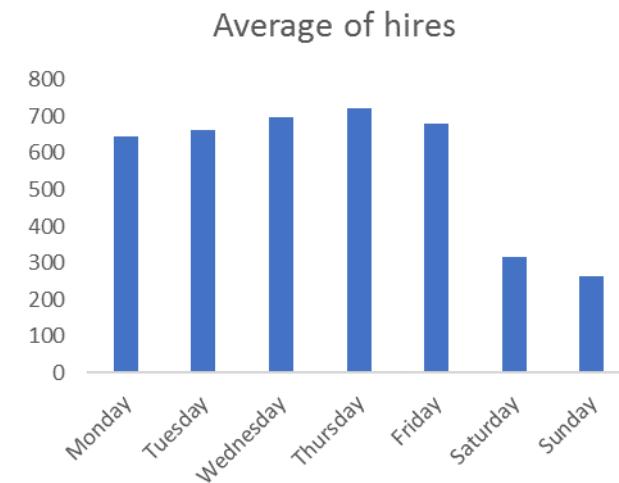
# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



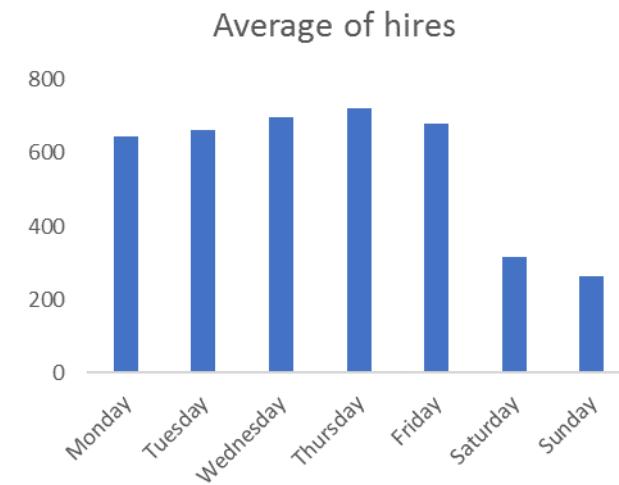
# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



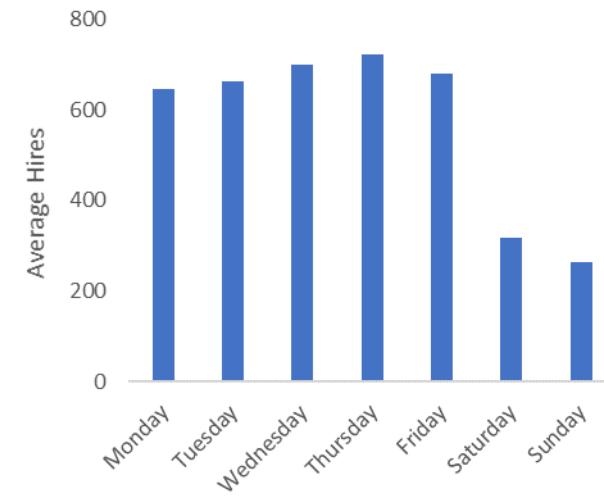
# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



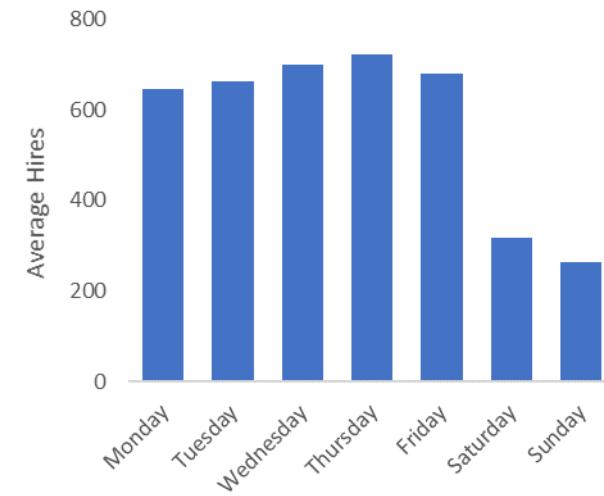
# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



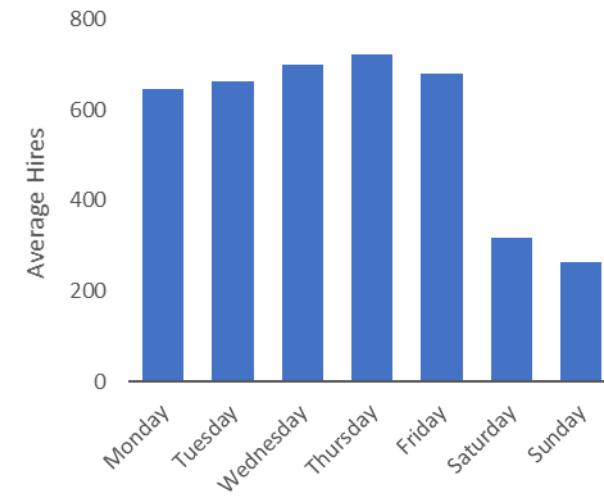
# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



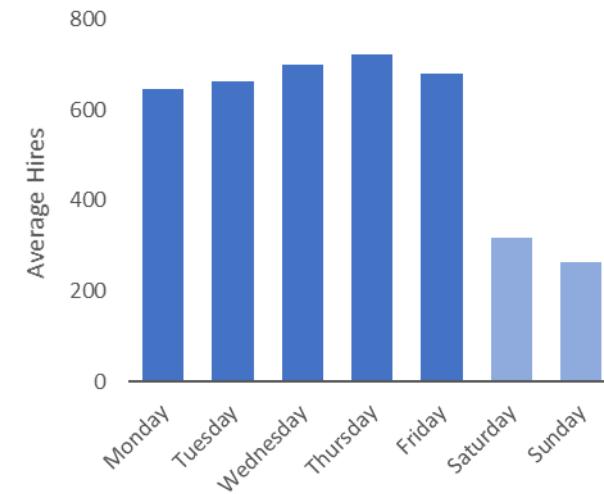
# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



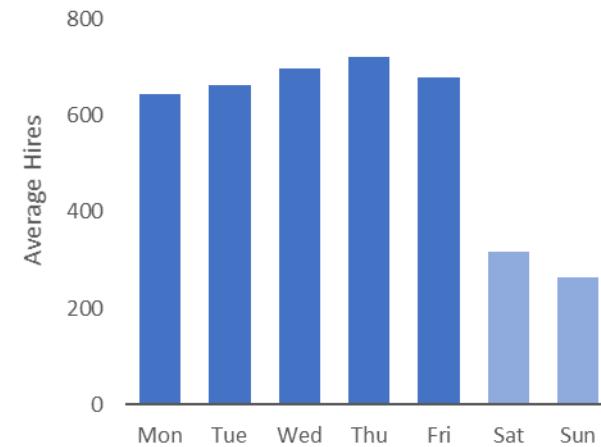
# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



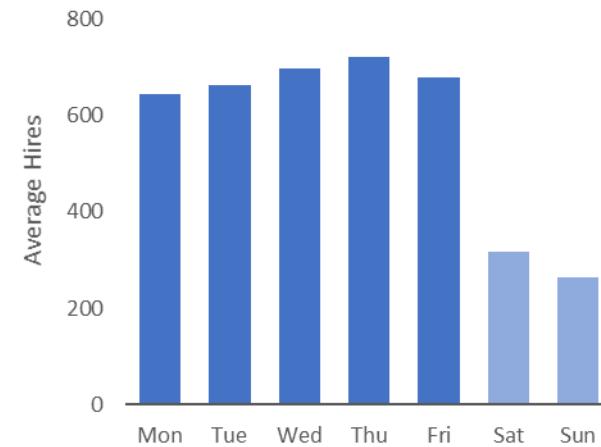
# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



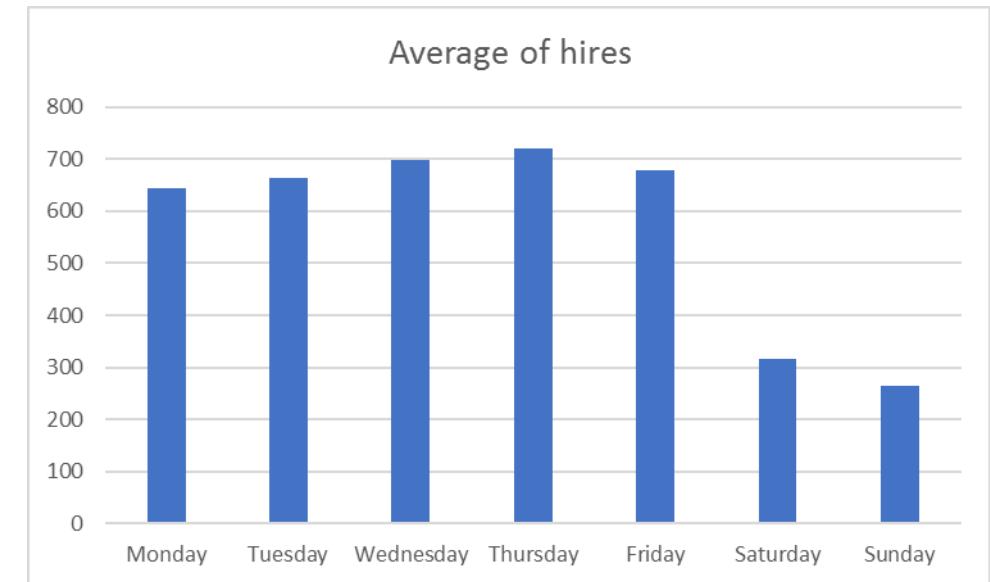
# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)

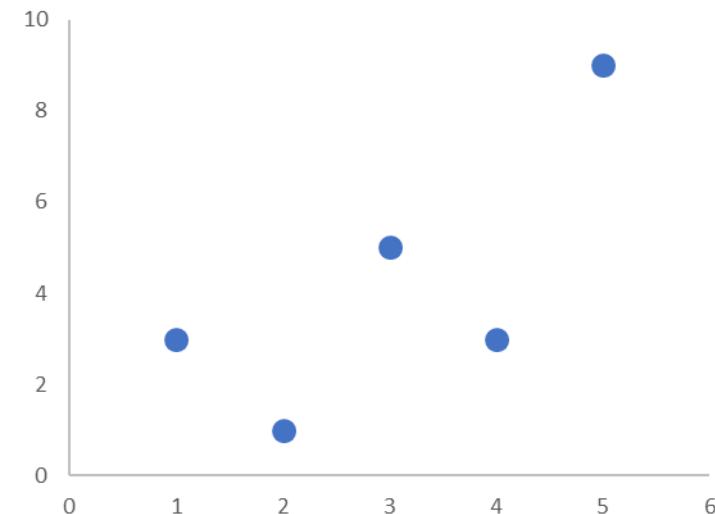
For small datasets tables are pretty good

Average hires	
Mon	644.7
Tue	663.3
Wed	698.0
Thu	720.2
Fri	678.5
Sat	317.1
Sun	263.8

# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)

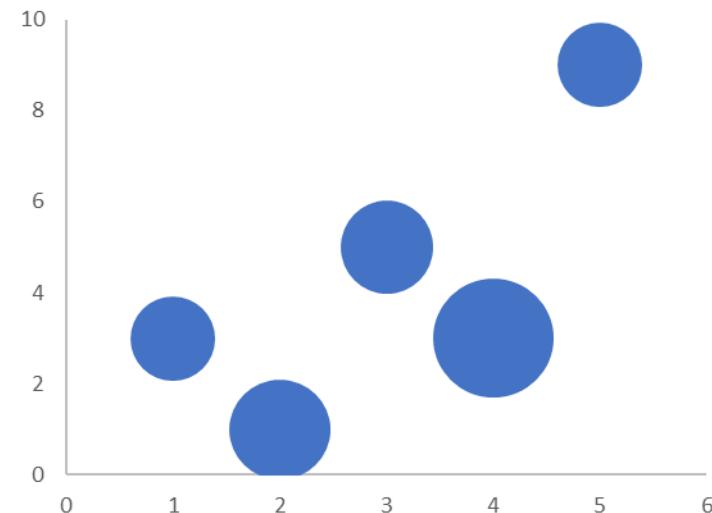
2 dimensions



# rules for good visualization

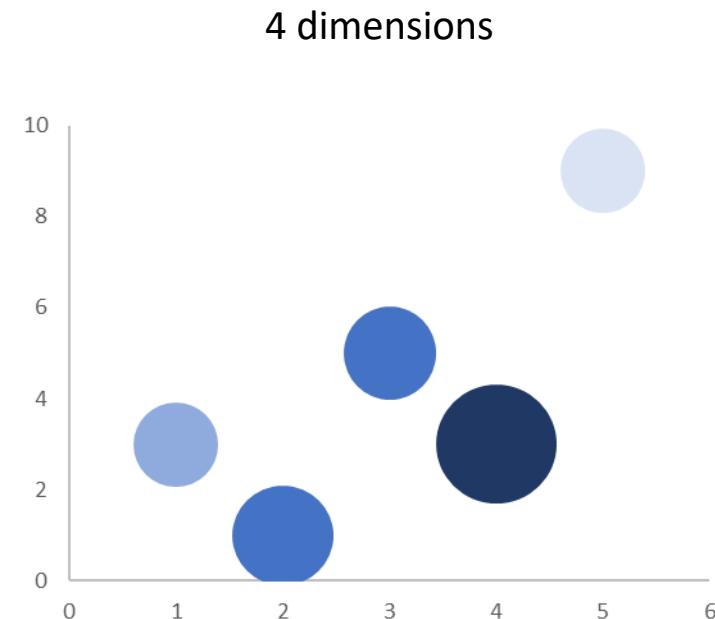
1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)

3 dimensions



# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)



# rules for good visualization

1. Tell a story/know what you're trying to show
2. The quicker someone can understand what they're seeing the better
3. Have as little *non-data ink* as possible
4. Increase the data density (without interfering with 1 & 2)

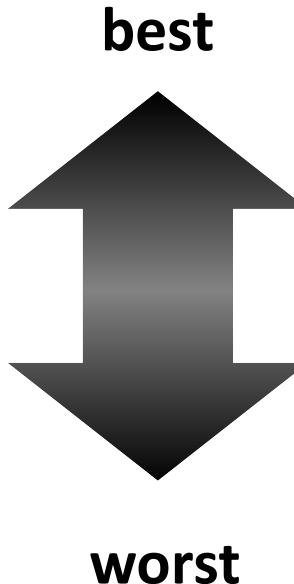
1 & 2 metrics to judge your visualization by

3 & 4 ways to improve it

# which chart?

Humans can see patterns in data more easily when it's present in certain ways.

1. **position along a common scale**
2. **position along a non-aligned scale**
3. **length**
4. **angle and slope**
5. **area**
6. **volume, density, and color saturation**
7. **colour hue**



*Graphical perception: theory, experimentation and application to the development of graphical methods (Cleveland & McGill)*

# Design is important too

Good design can't be summarised very easily. Try to think about:

- Fonts
- Weight of lines
- Size of chart
- Colours
- Patterns
- Consistency

# Design is important too

Good design can't be summarised very easily. Try to think about:

- Fonts
- Weight of lines
- Size of chart
- Colours
- Patterns
- Consistency

use sans-serif fonts

not serif fonts

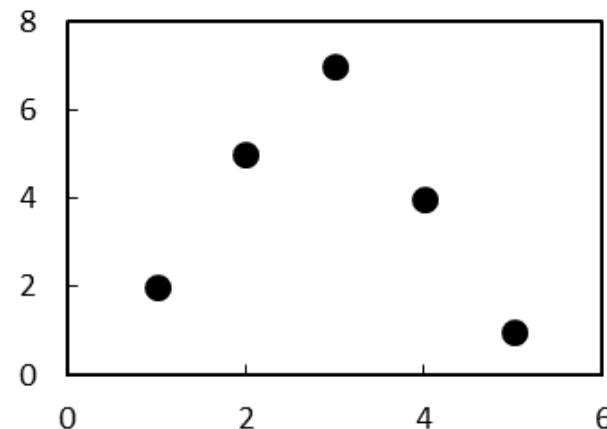
use size  
to convey  
importance

Use as little text as  
possible!

# Design is important too

Good design can't be summarised very easily. Try to think about:

- Fonts
- **Weight of lines**
- Size of chart
- Colours
- Patterns
- Consistency

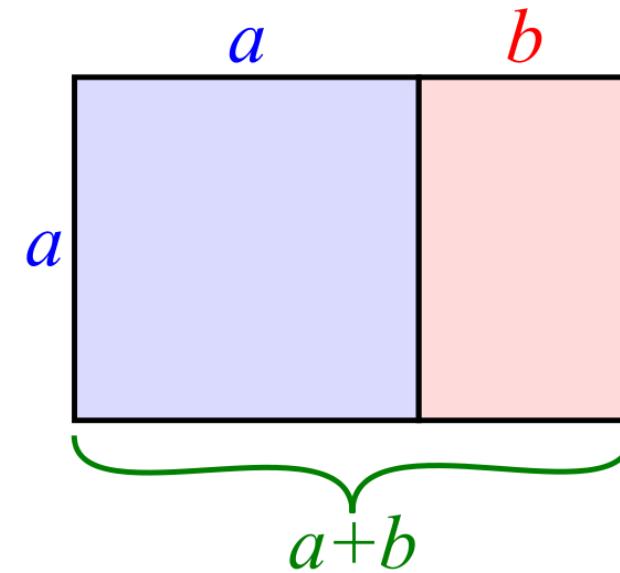


Draw the eye to the right place first

# Design is important too

Good design can't be summarised very easily. Try to think about:

- Fonts
- Weight of lines
- **Size of chart**
- Colours
- Patterns
- Consistency



golden ratio

# Design is important too

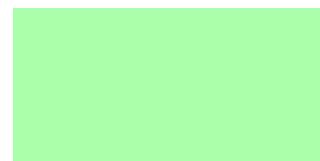
Good design can't be summarised very easily. Try to think about:

- Fonts
  - Weight of lines
  - Size of chart
  - Colours
  - Patterns
  - Consistency
- wider than tall
- If taller than wide,  
give it rows

# Design is important too

Good design can't be summarised very easily. Try to think about:

- Fonts
- Weight of lines
- Size of chart
- Colours
- Patterns
- Consistency



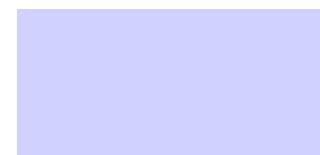
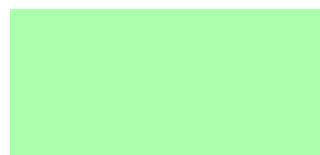
don't use

better

# Design is important too

Good design can't be summarised very easily. Try to think about:

- Fonts
- Weight of lines
- Size of chart
- Colours
- Patterns
- Consistency



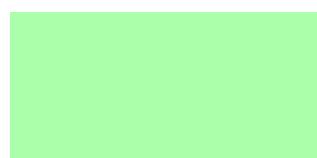
categories  
(only if number is small)

numbers

# Design is important too

Good design can't be summarised very easily. Try to think about:

- Fonts
- Weight of lines
- Size of chart
- Colours
- Patterns
- Consistency



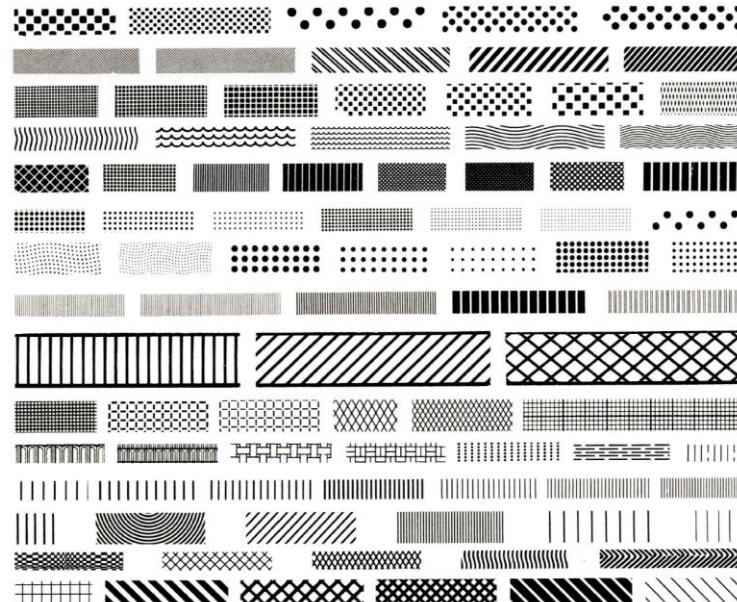
colour is a good way to indicate similarity

# Design is important too

Good design can't be summarised very easily. Try to think about:

- Fonts
- Weight of lines
- Size of chart
- Colours
- Patterns
- Consistency

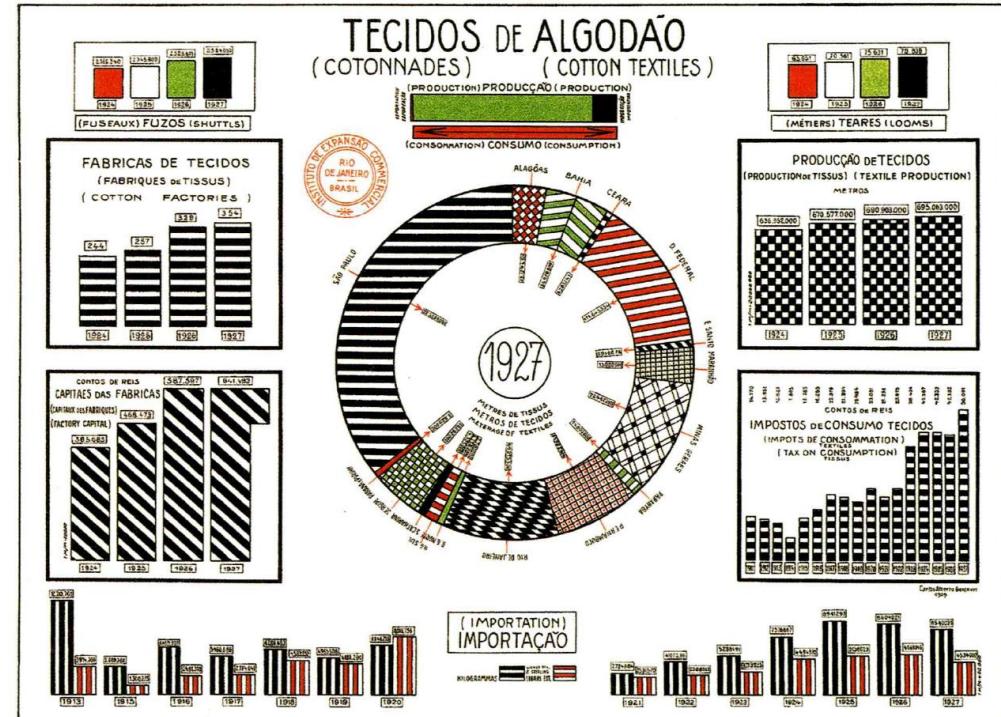
don't use patterns



# Design is important too

Good design can't be summarised very easily. Try to think about:

- Fonts
- Weight of lines
- Size of chart
- Colours
- Patterns
- Consistency



# Design is important too

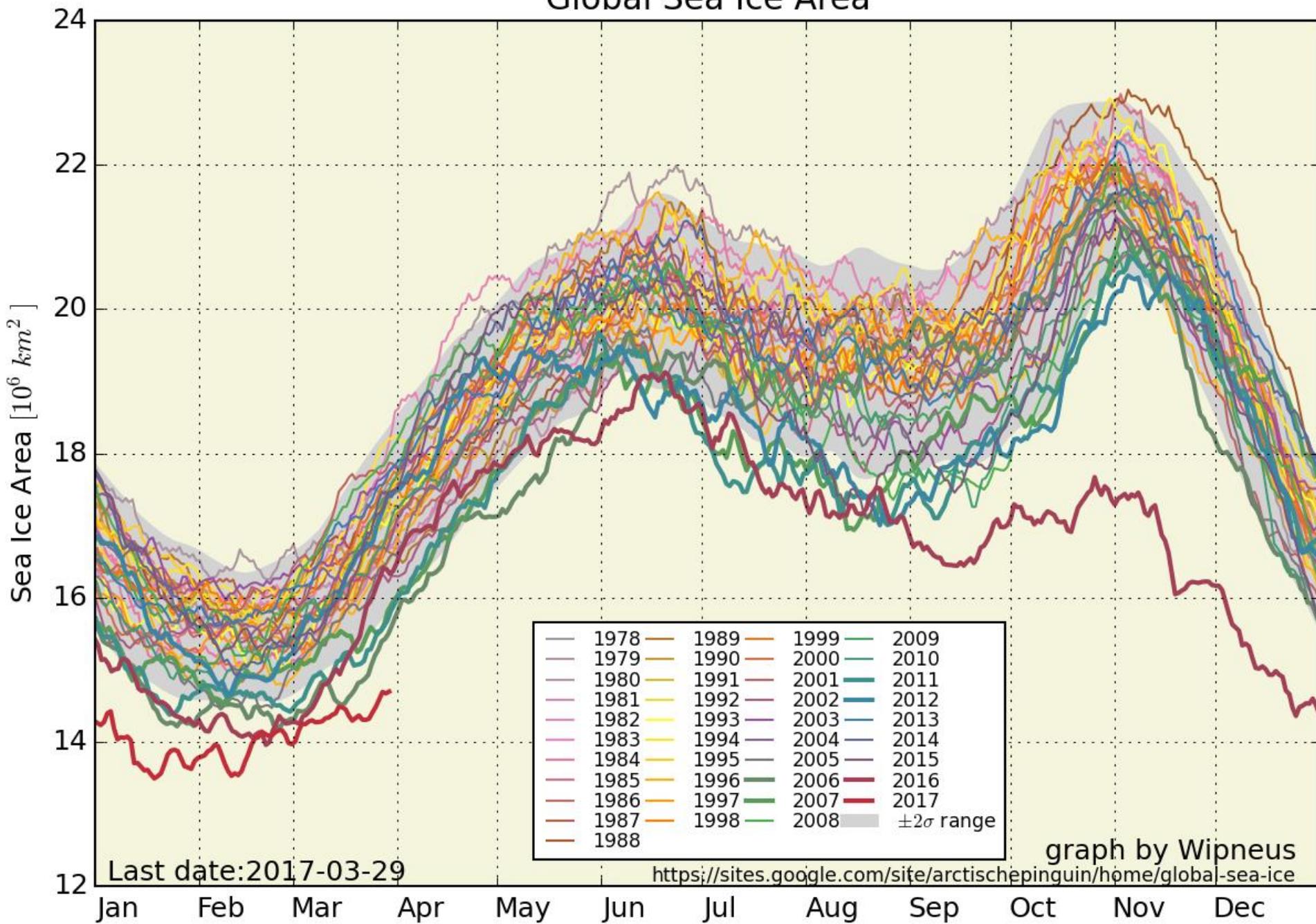
Good design can't be summarised very easily. Try to think about:

- Fonts
- Weight of lines
- Size of chart
- Colours
- Patterns
- Consistency

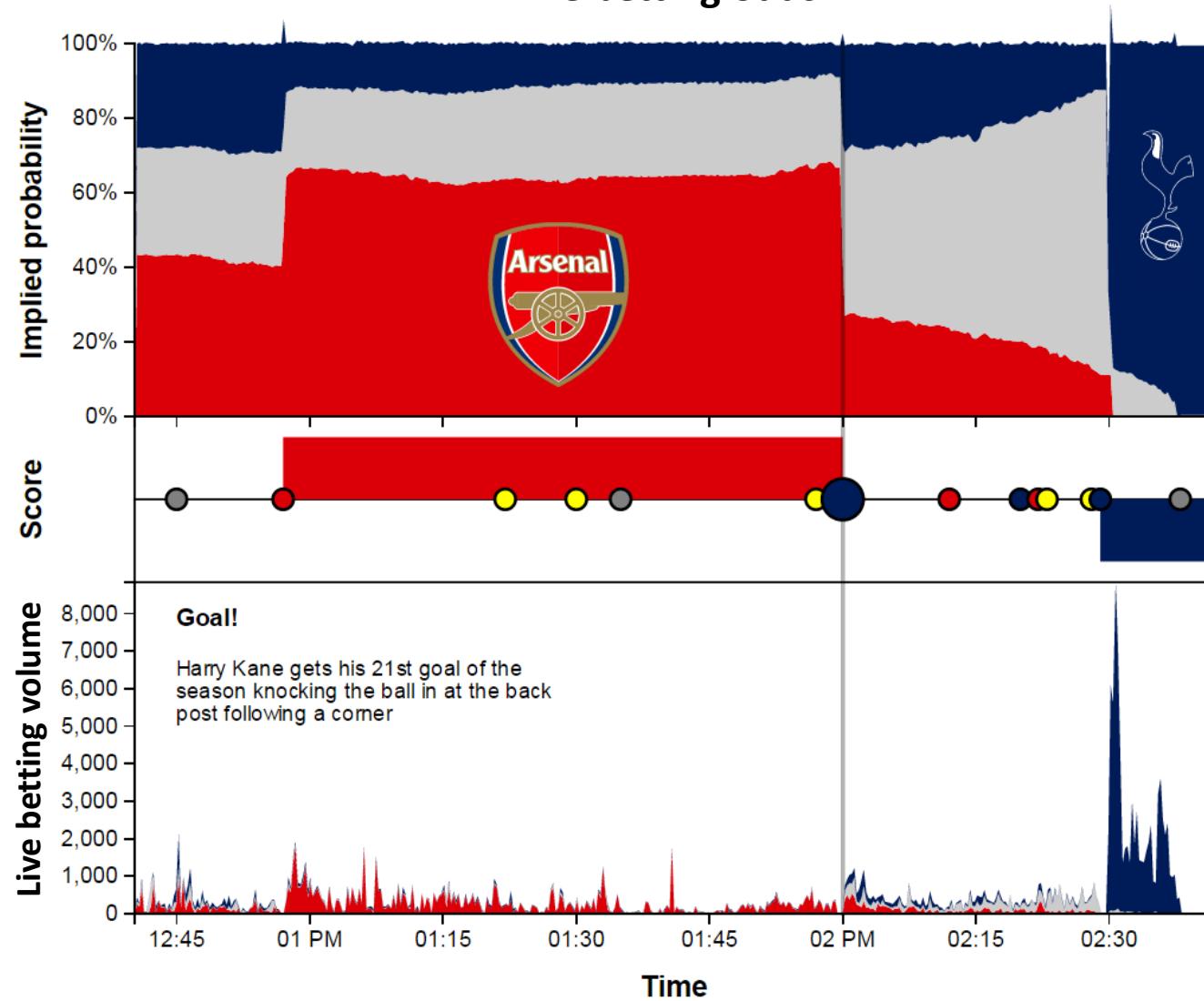
# Tufte's rules

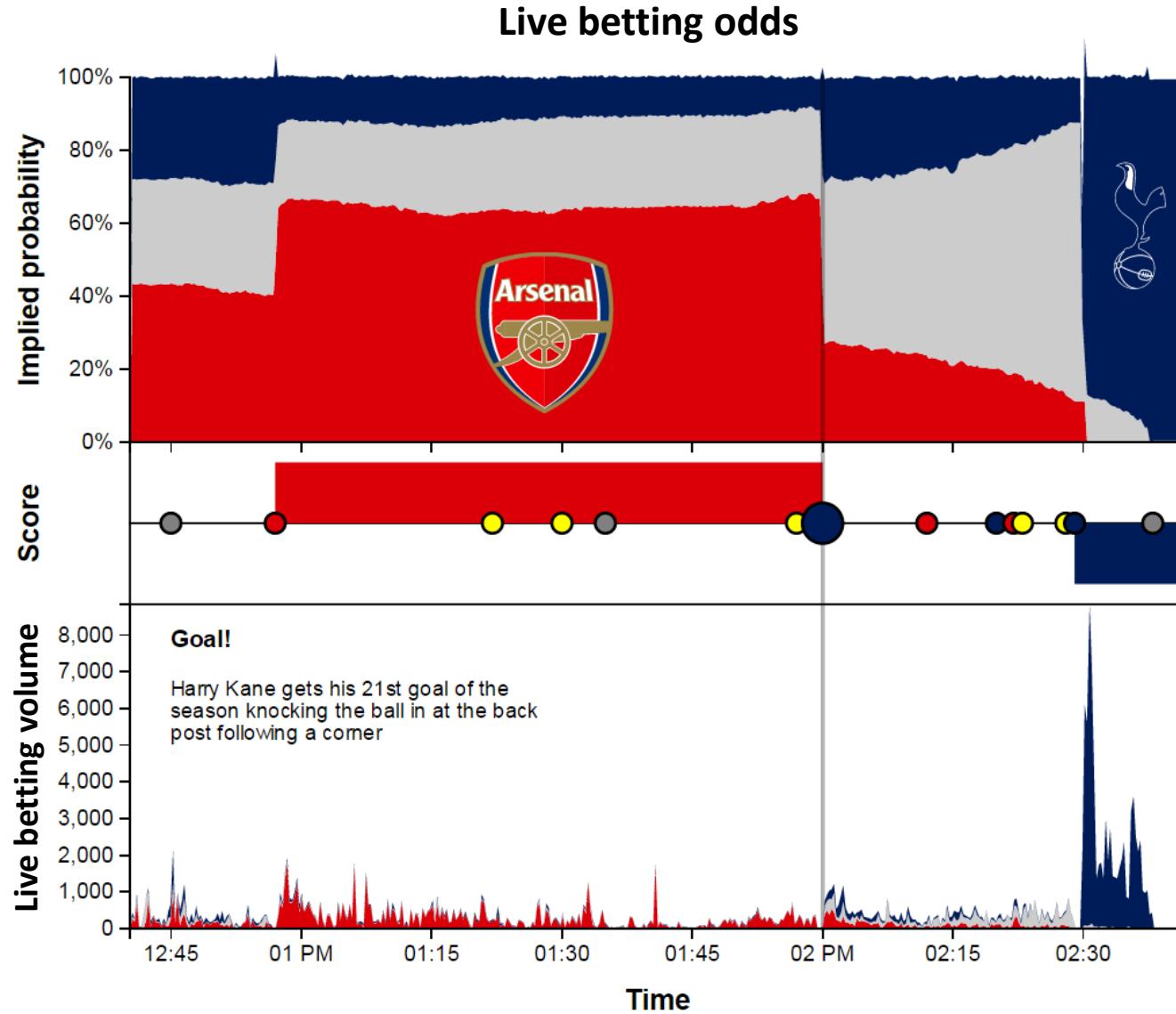
- Make the viewer think about the meaning of the data, not technical stuff like design, the chart etc.
- The first thing they think about should be to do with the data, not the design
- Avoid distorting the data
- Reveal the data at several different levels of detail
- Try to get as much out of the data as possible

from NSIDC sea ice concentration data  
Global Sea Ice Area



## Live betting odds

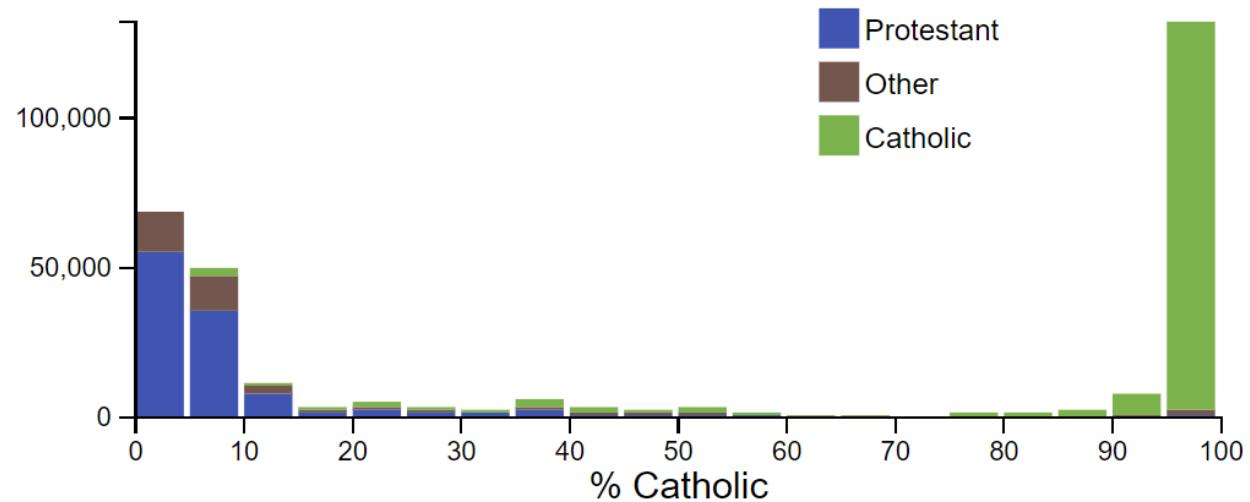




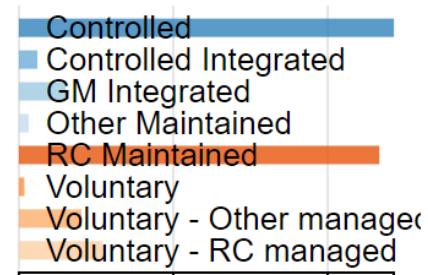
- *stack emphasises odds move from one outcome to another*
- *sum is always 100%*
- *Independent variable on x*
- *X axis aligned*
- *2<sup>nd</sup> graph lowers data density...*
- *Minimal non-data ink*
- *Colours on two charts match*

# Education in Northern Ireland

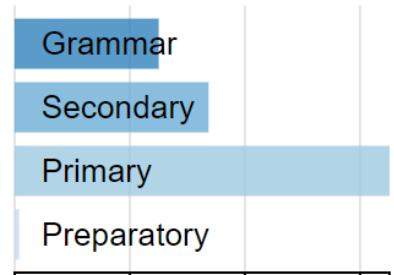
Pupil numbers:



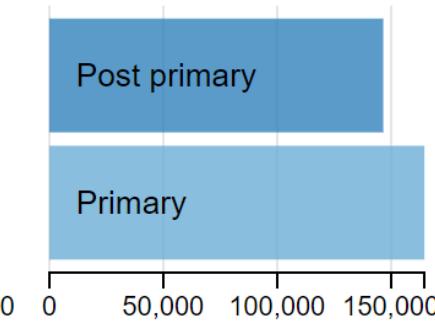
Management type:



School type:

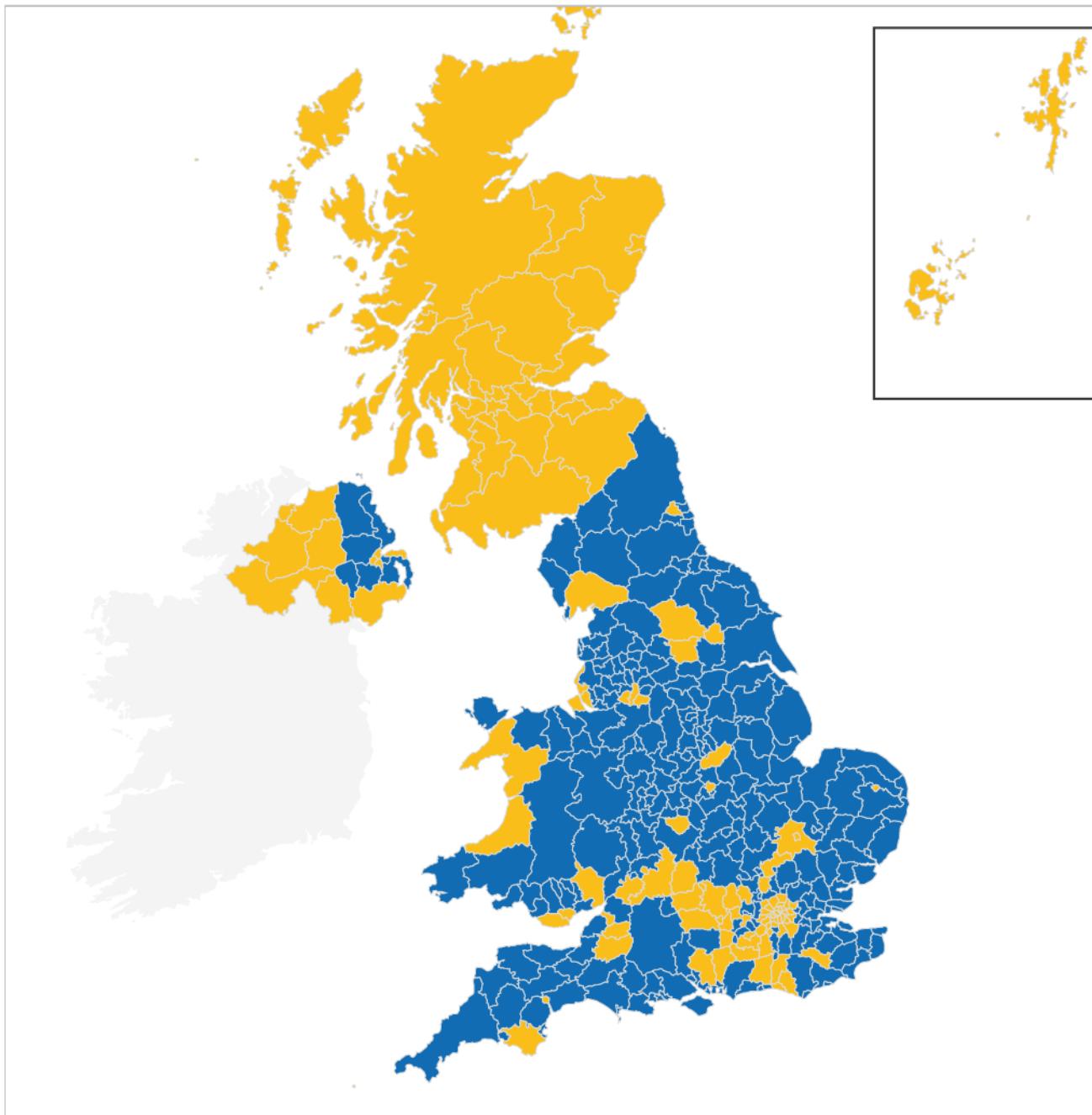


Age:

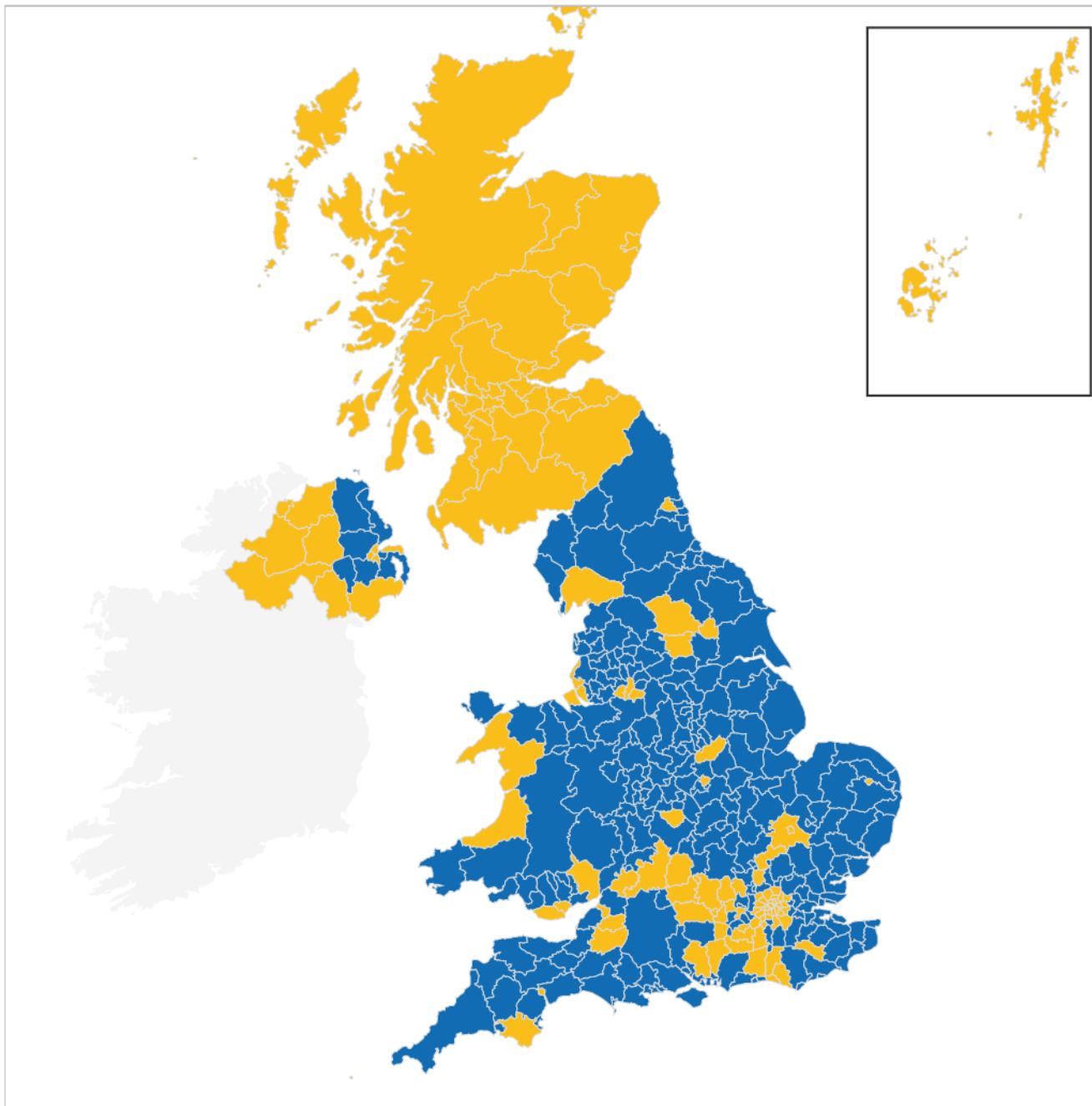


<https://dc-js.github.io/dc.js/docs/stock.html>

Key: Majority leave Majority remain Tie Undeclared



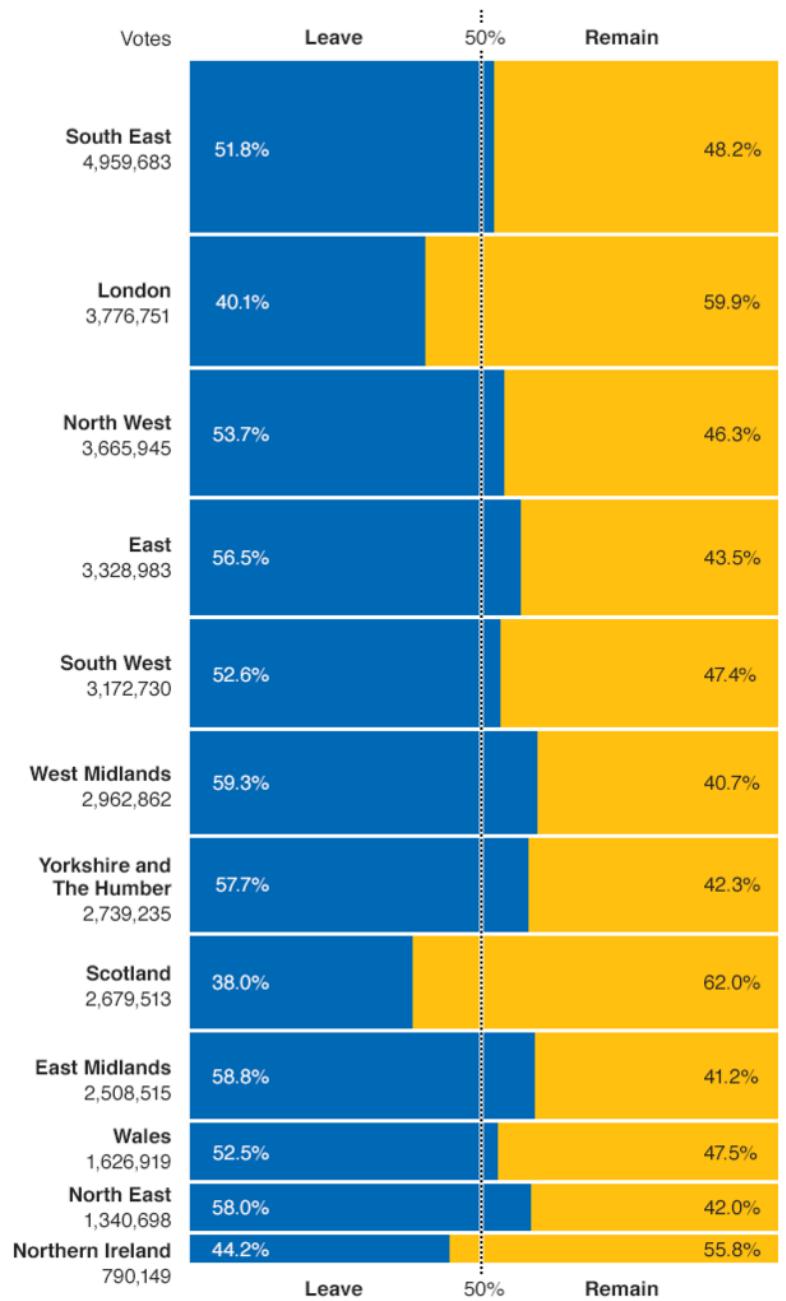
Key: Majority leave Majority remain Tie Undeclared



*plotting area, but area doesn't matter  
no sense of how big majorities were*

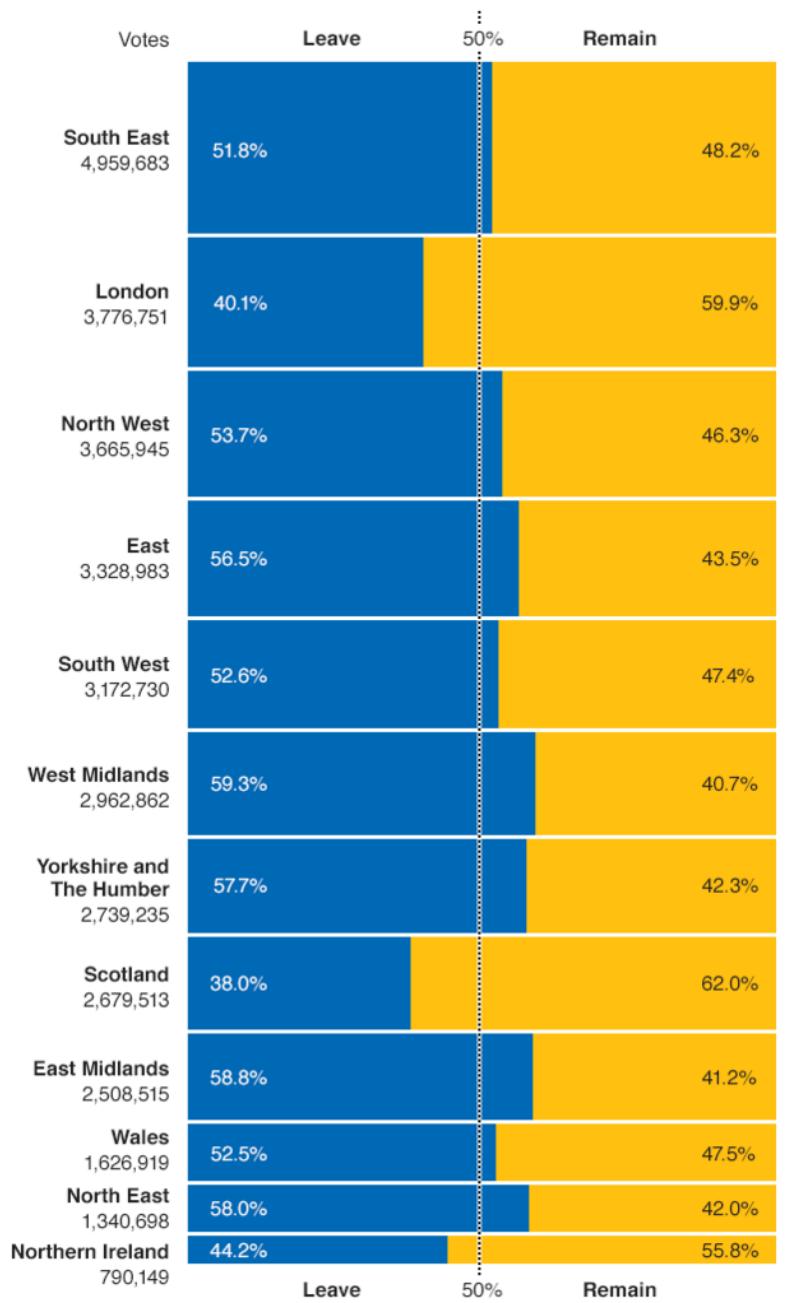
## How Leave won the referendum

Depth of bars is proportional to votes cast, largest areas shown first



## How Leave won the referendum

Depth of bars is proportional to votes cast, largest areas shown first



*Can see clearly the most populous areas*

*Can clearly see the size of majorities*

Majority >15%

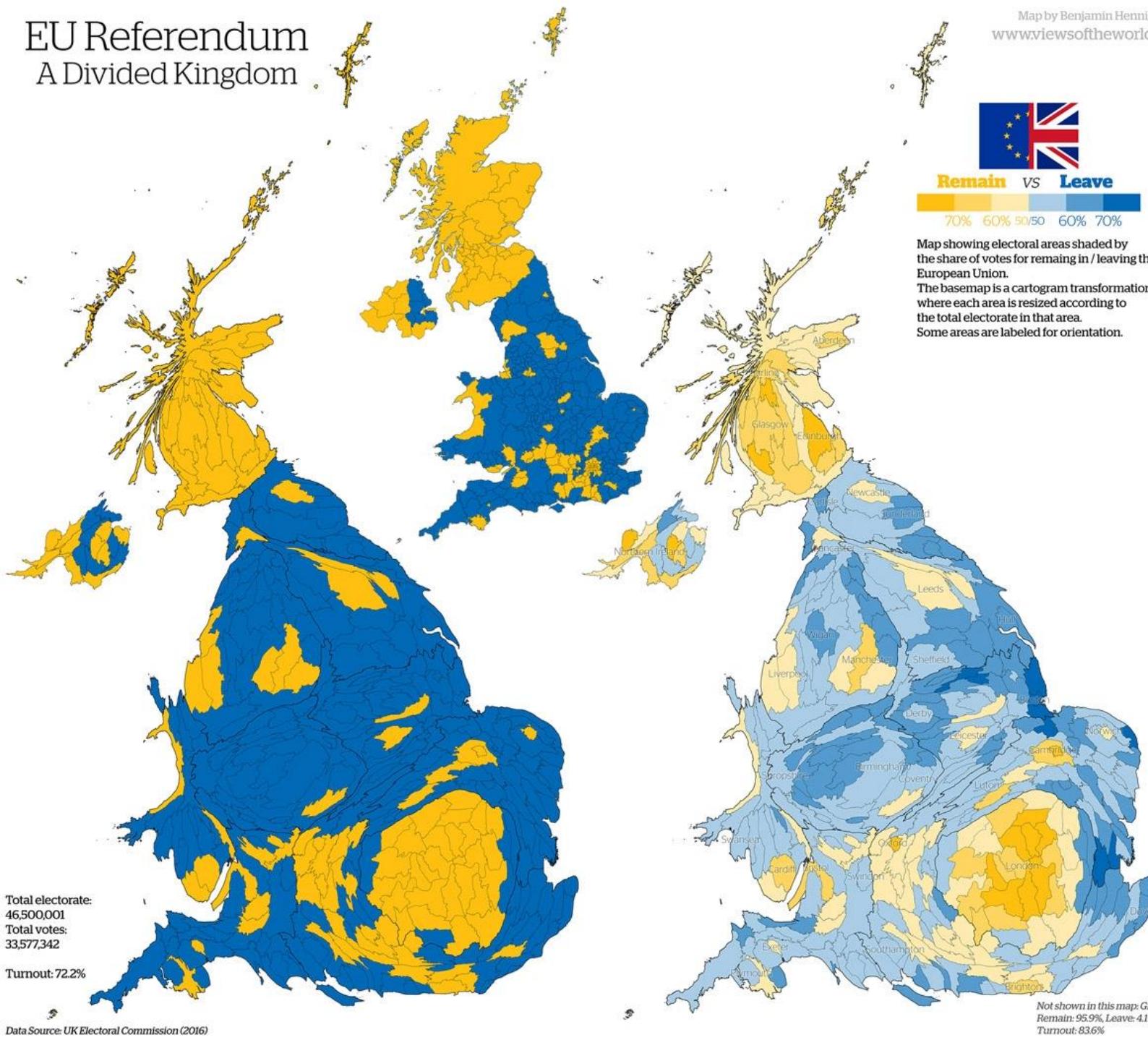
Remain  
Leave



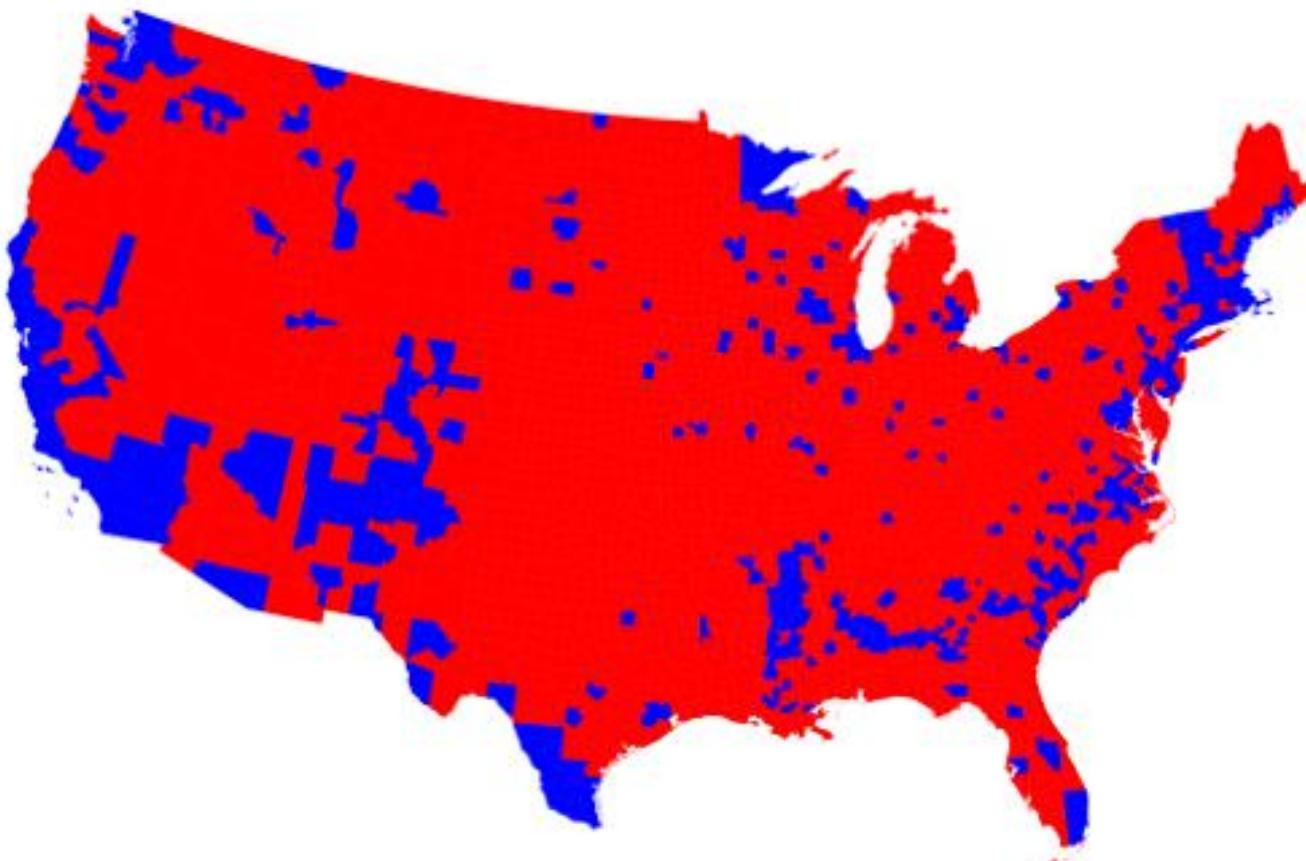
# EU Referendum

## A Divided Kingdom

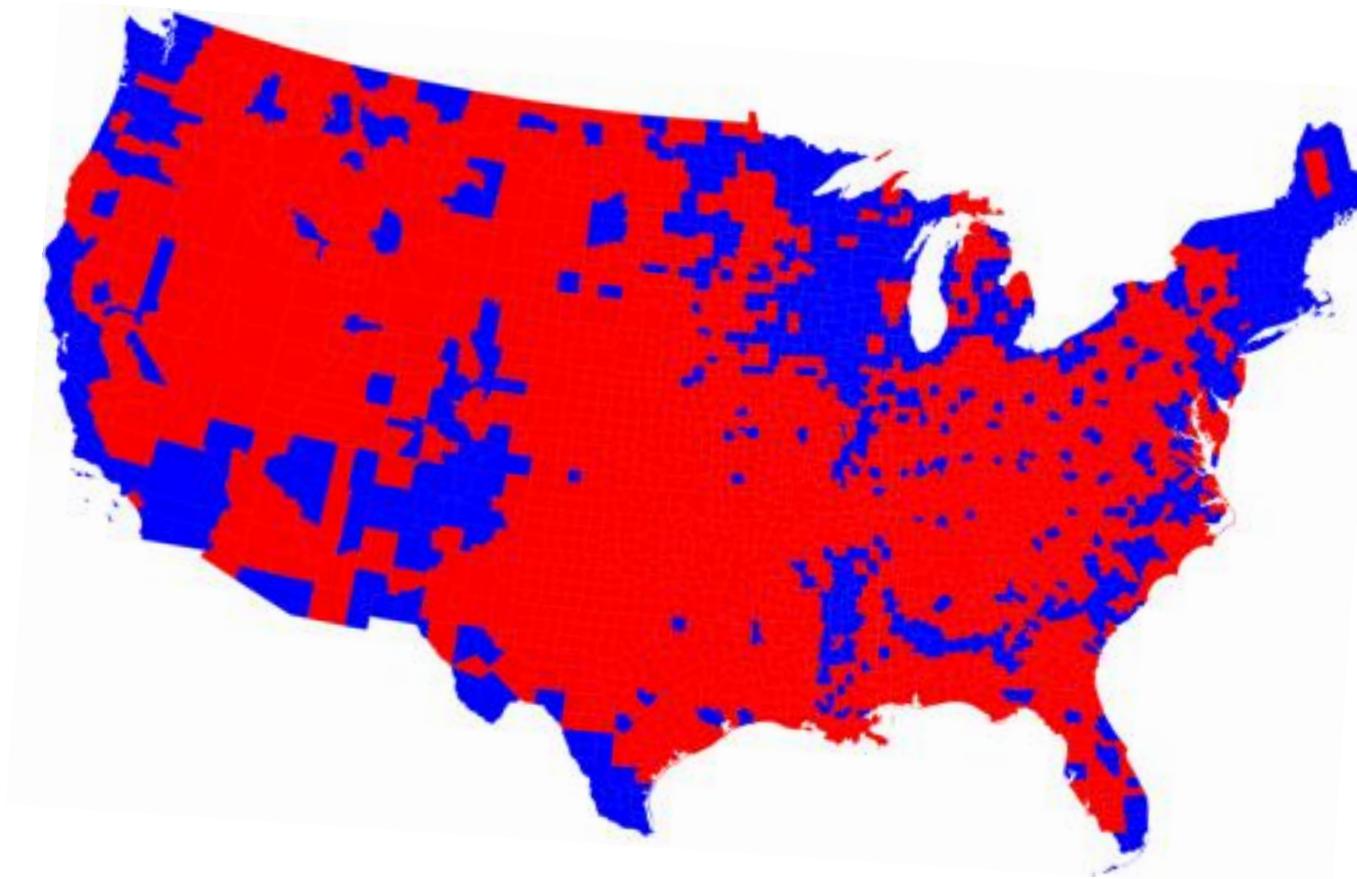
Map by Benjamin Hennig  
[www.viewsoftheworld.net](http://www.viewsoftheworld.net)

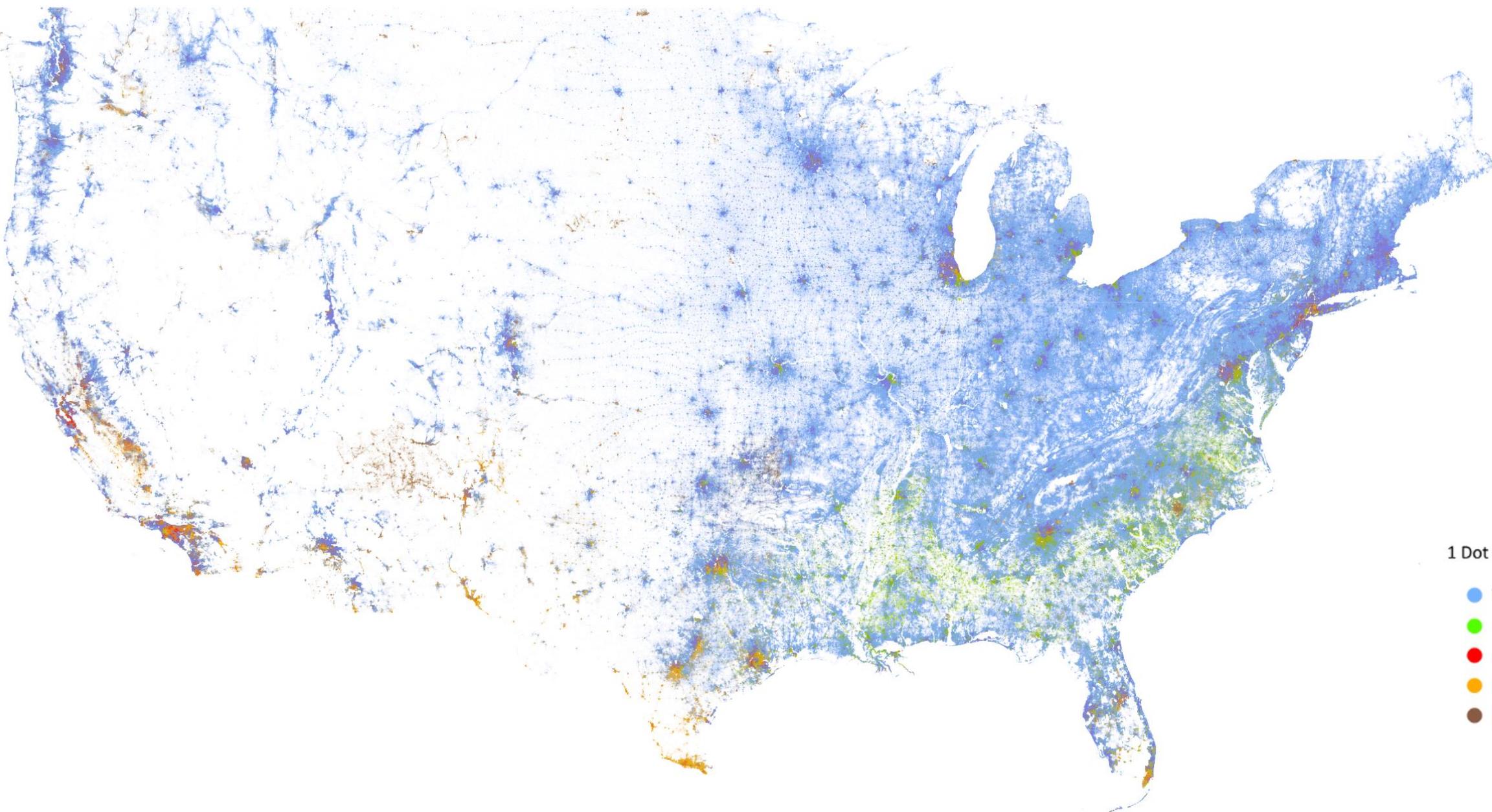


**2016 US election**



**2008 US election**



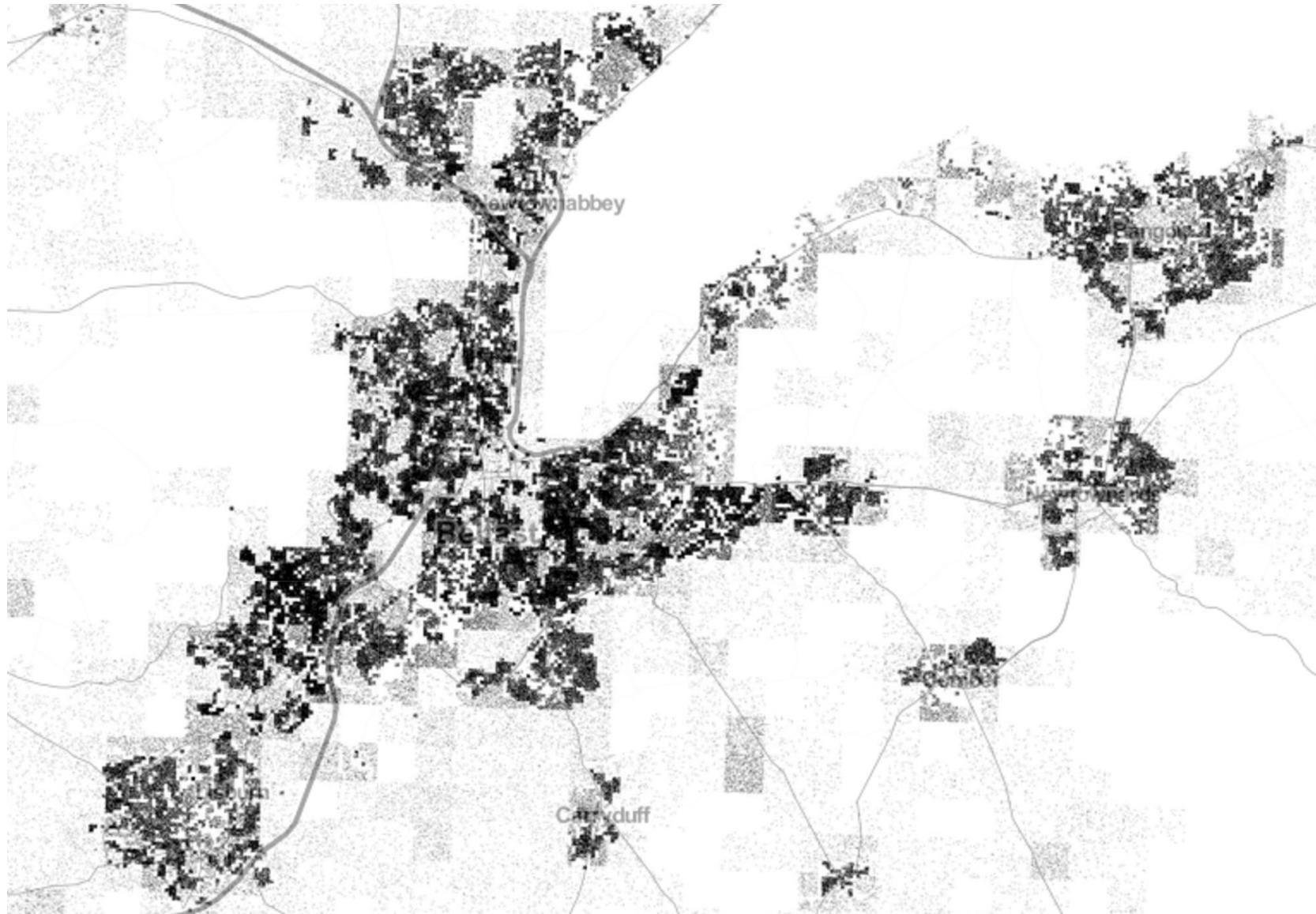


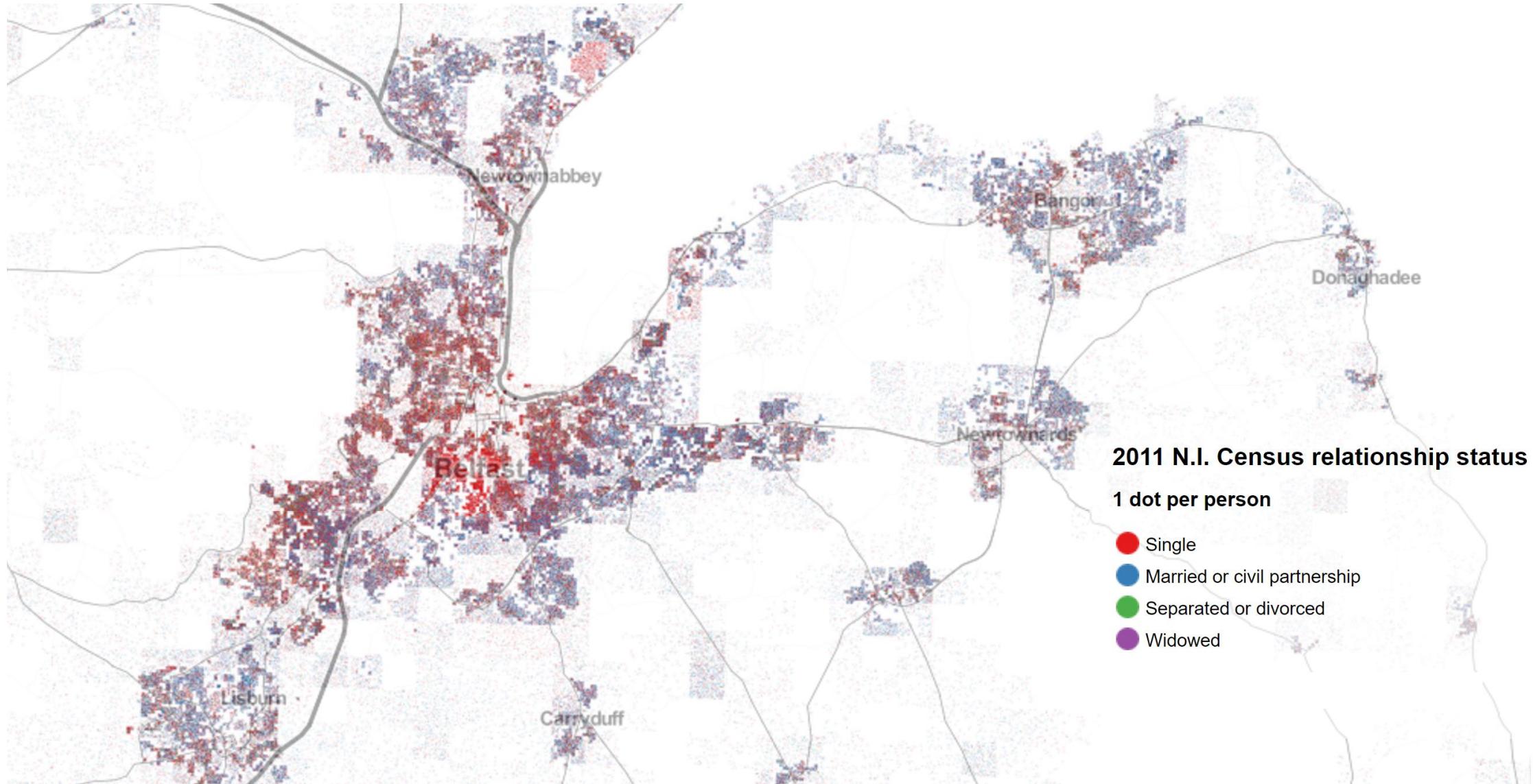
1 Dot = 1 Person

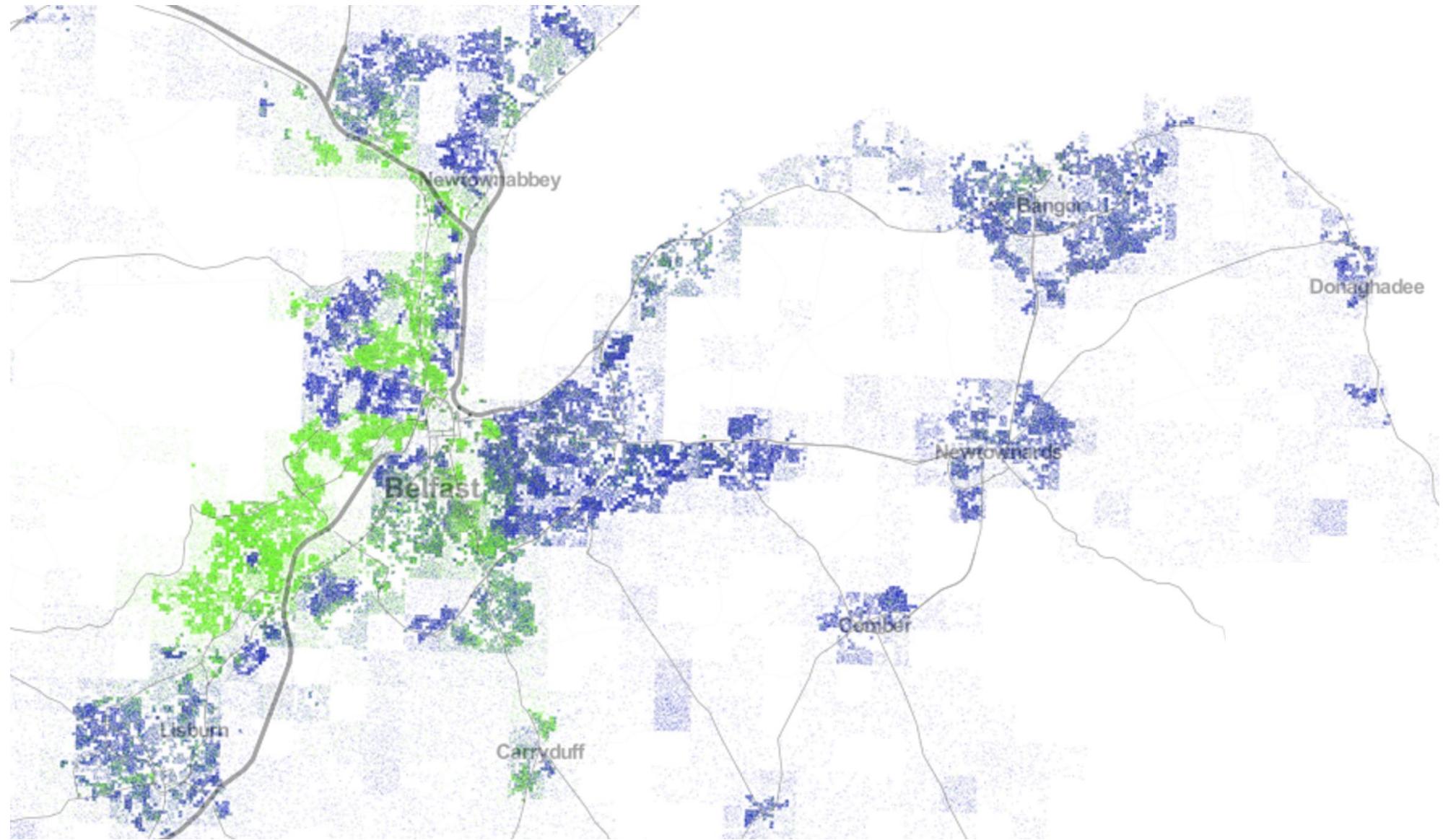
- White
- Black
- Asian
- Hispanic
- Other Race / Native American / Multi-racial











<https://www.picodoc.org/wp-content/uploads/councilni/>

<https://www.bloomberg.com/politics/graphics/2015-redistricting/>