

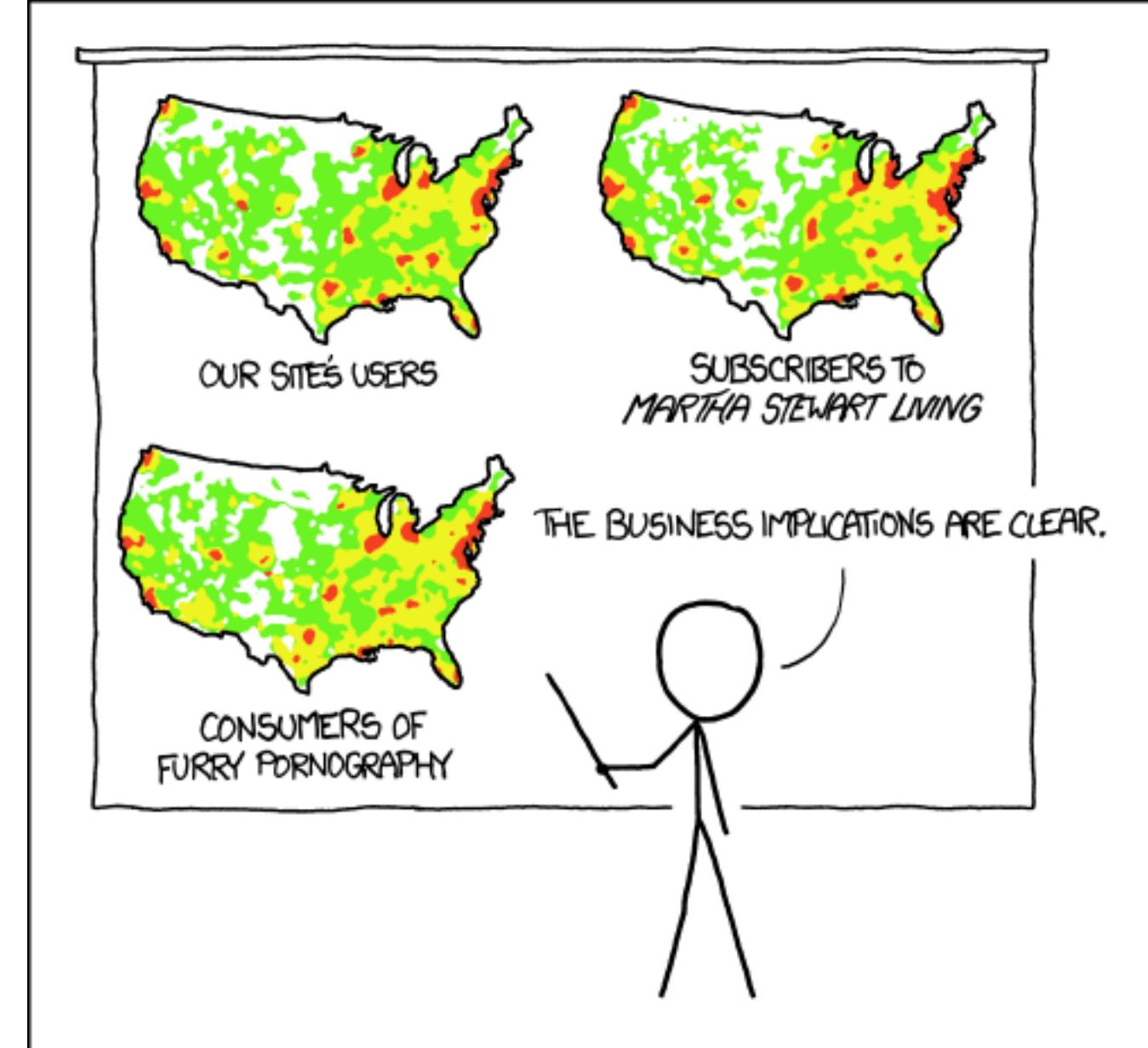
# Introduction to *Urban* Data Science

## Geo-Visualisation

(EPA1316)

Lecture 6

Trivik Verma



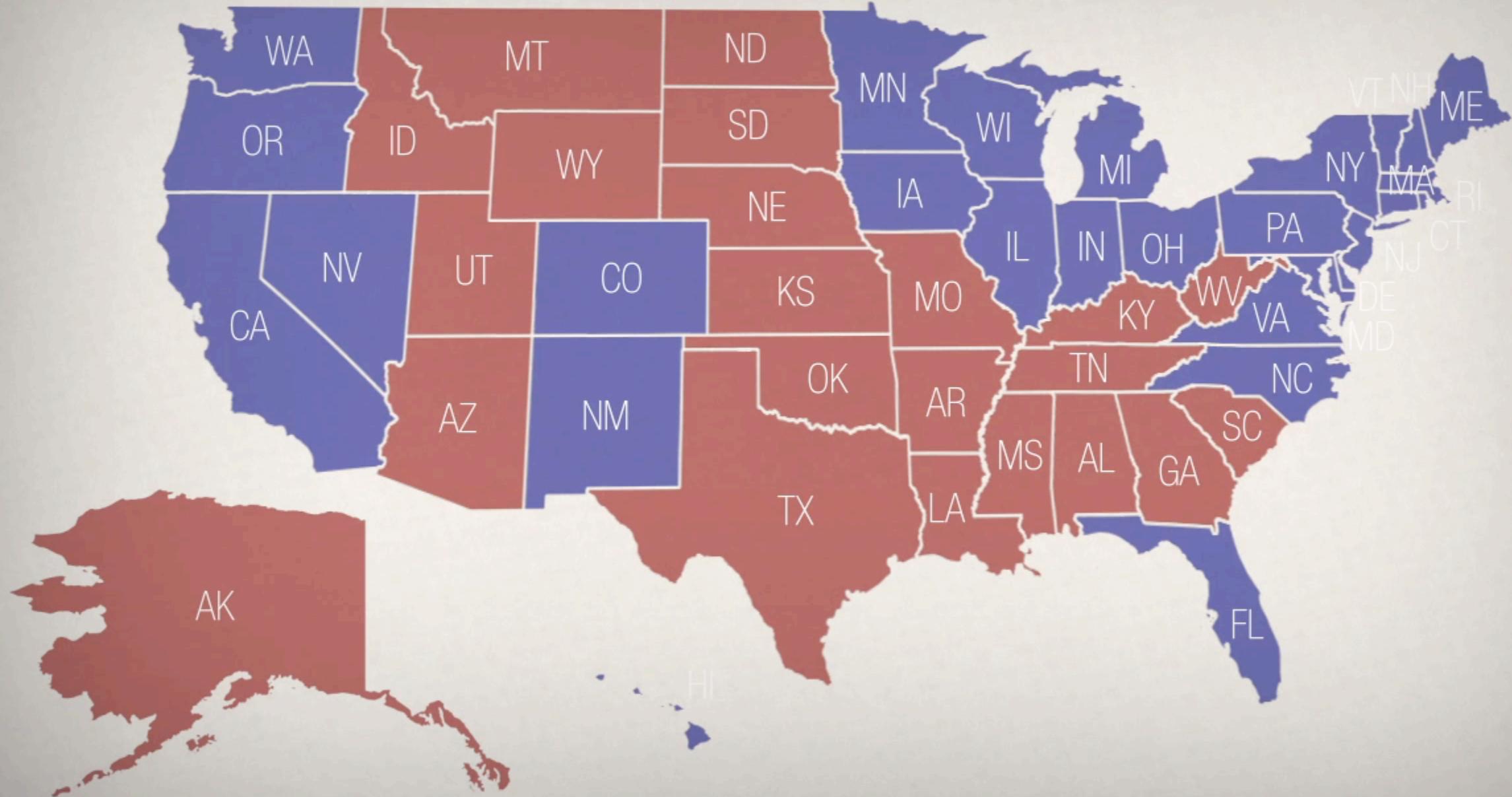
# Last Time

- History of Visualisations
- Exploratory Data Analysis
- Types of Visualisations
- Effective Visualisation

# Today

- Geo-Visualisation
- Dangers of Geo-Vis
- Mapping Data
  - MAUP
  - Choropleths

# 2008 Election



# *Geo-Visualisation*

# Tufte (1983)

*“The most extensive data maps place millions of bits of information on a single page before our eyes. No other method for the display of statistical information is so powerful”*

# MacEachren (1994)

*“Geographic visualization can be defined as the use of concrete visual representations – whether on paper or through computer displays or other media – to make spatial contexts and problems visible, so as to engage the most powerful human information processing abilities, those associated with vision.”*

# Geo Visualisation

- End goal is not to replace the human *in the loop*, but to augment her.
- Augmentation here comes through engaging the **pattern recognition** capabilities that our brain inherently has.
- Combines:
  - Traditional maps
  - Statistical maps
  - Statistical devices of other kind (charts, scatter plots, etc.)
  - **Different roles** in the analysis process...

# A map for everyone

Maps can fulfill several needs

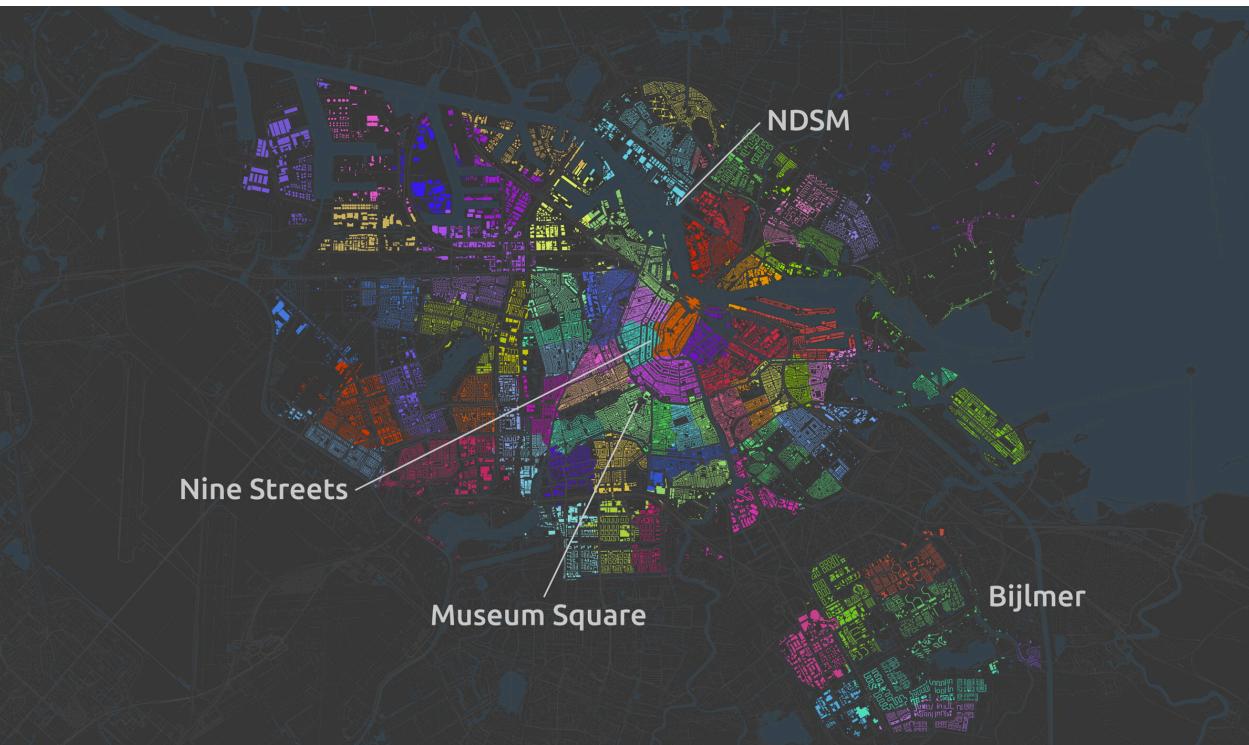
Depending on which one we want to stress, the best map will look very different

MacEachren & Kraak (1997) identify three main dimensions

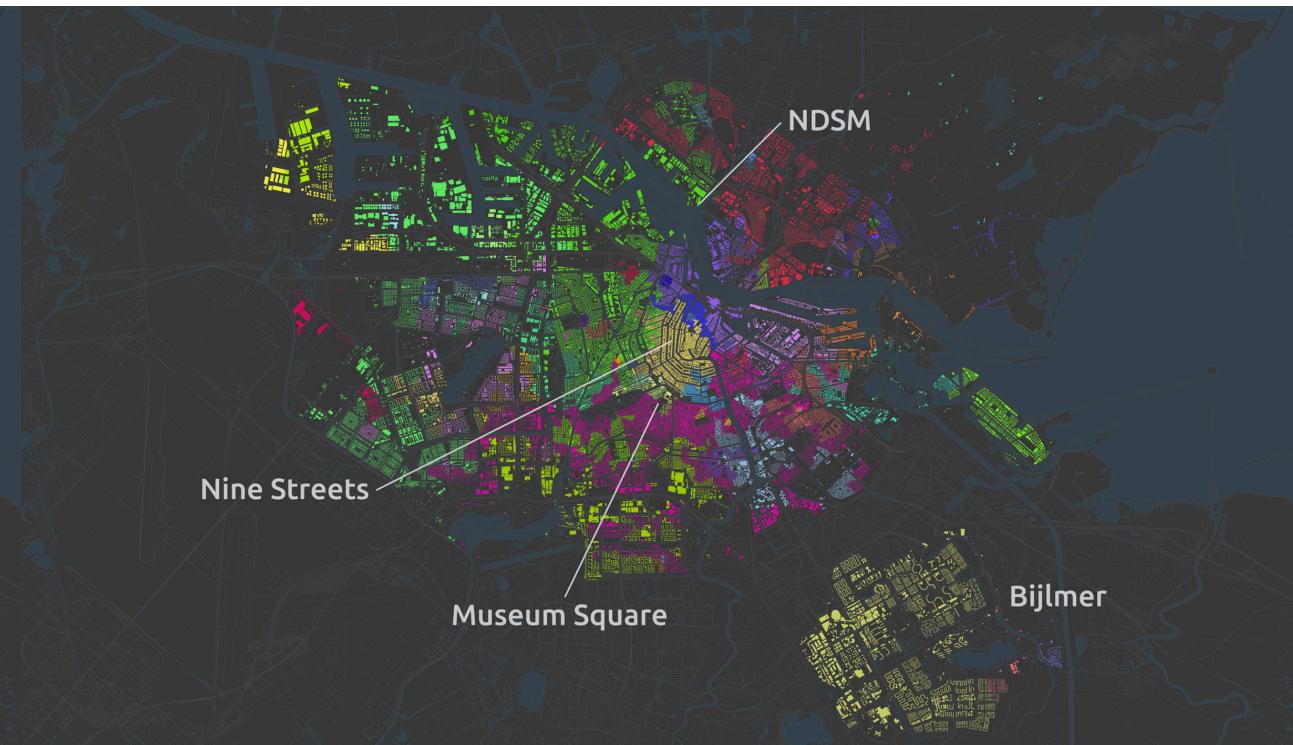
- Knowledge of what is being plotted
- Target audience
- Degree of interactivity

Un/known: *fast* and *slow* maps

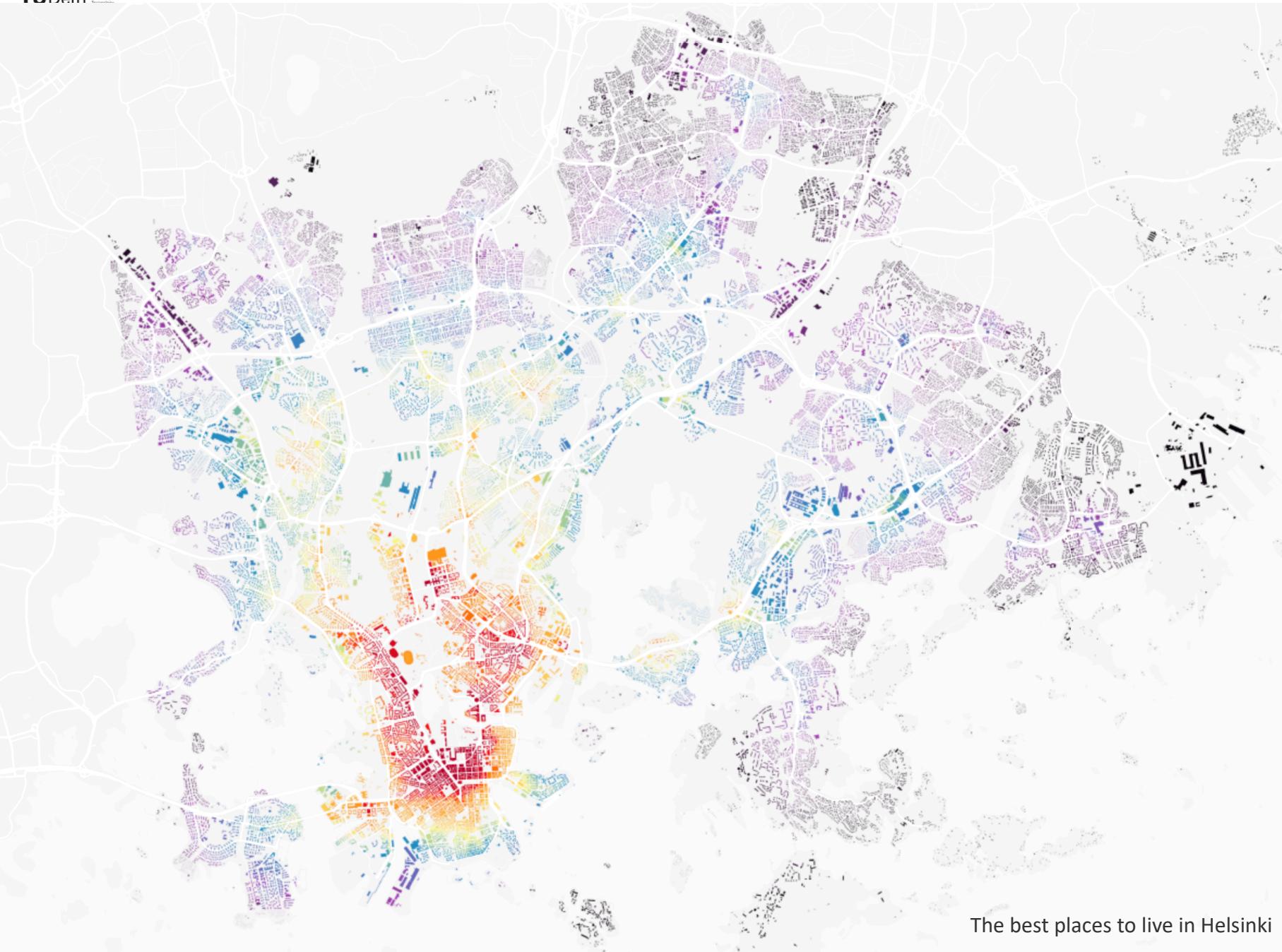
Digit postcodes : fast



Twitter : slow



Audience: *easy* and *hard* maps



# easy map

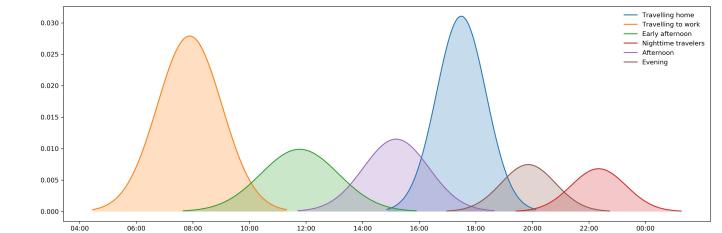
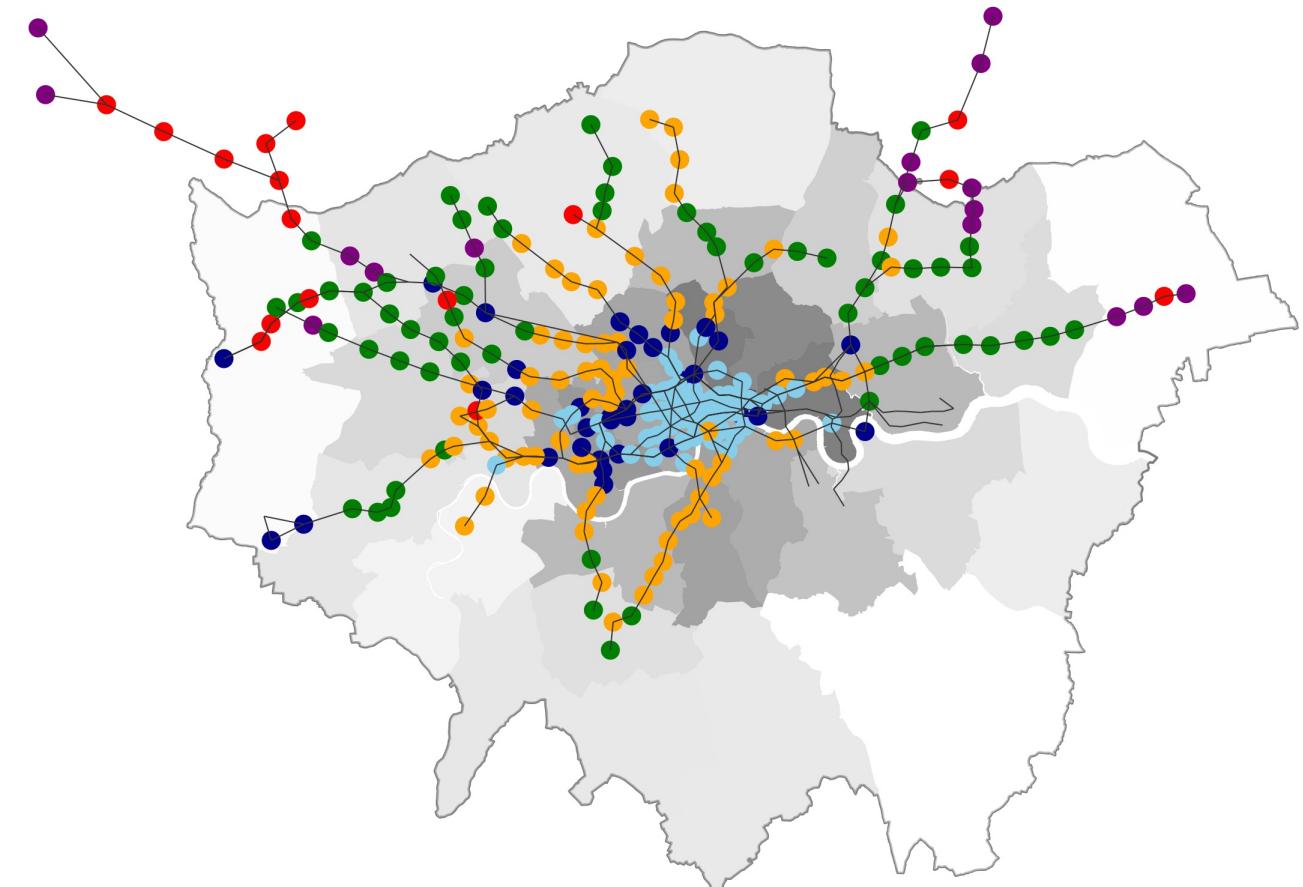
Ambitions Travel Index



[source]

## Greater London, TfL stations

*hard* map

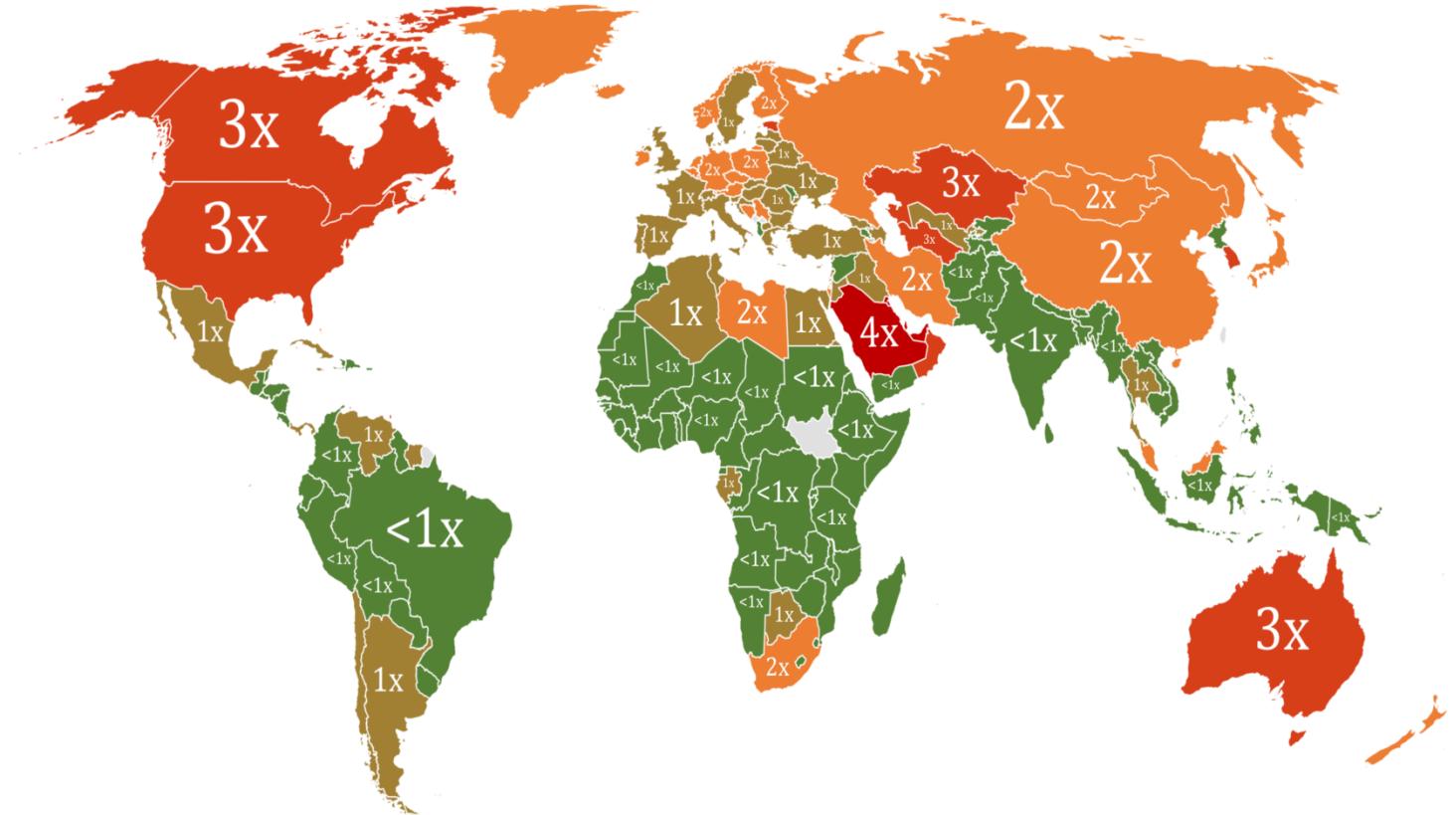


Interaction: *one or many* maps in one

# More Than Their Fair Share:

Which countries produce a greater proportion of global CO<sub>2</sub> emissions than their proportion of the global population?

*static map*



Ratio of Global Emissions Contribution to Share of Global Population



# WebCAT

Address or co-ordinates

 X GoAccess level (PTAL) Time mapping (TIM)

TIM: a new measure, looking at how far you can travel in a given journey time.

Map key - Travel Time

< 15 mins	15 - 30 mins
30 - 45 mins	45 - 60 mins
60 - 75 mins	75 - 90 mins
90 - 105 mins	105 - 120 mins
120 - 135 mins	135 - 150 mins

Change travel time bands +

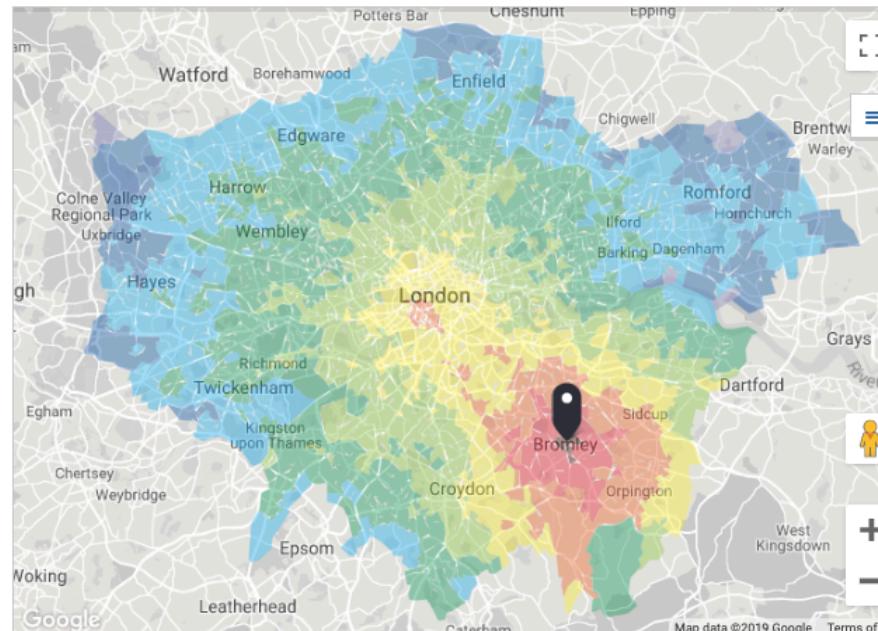
Map layers



Scenario

Base Year ▼

Mode

All public transport modes ▼

You can click anywhere on the map to change the selected location.

## TIM output for Base Year

Scenario: **Base Year Mode: All public transport modes**, Time of day: **AM peak**, Direction: **From location**

**Bromley**

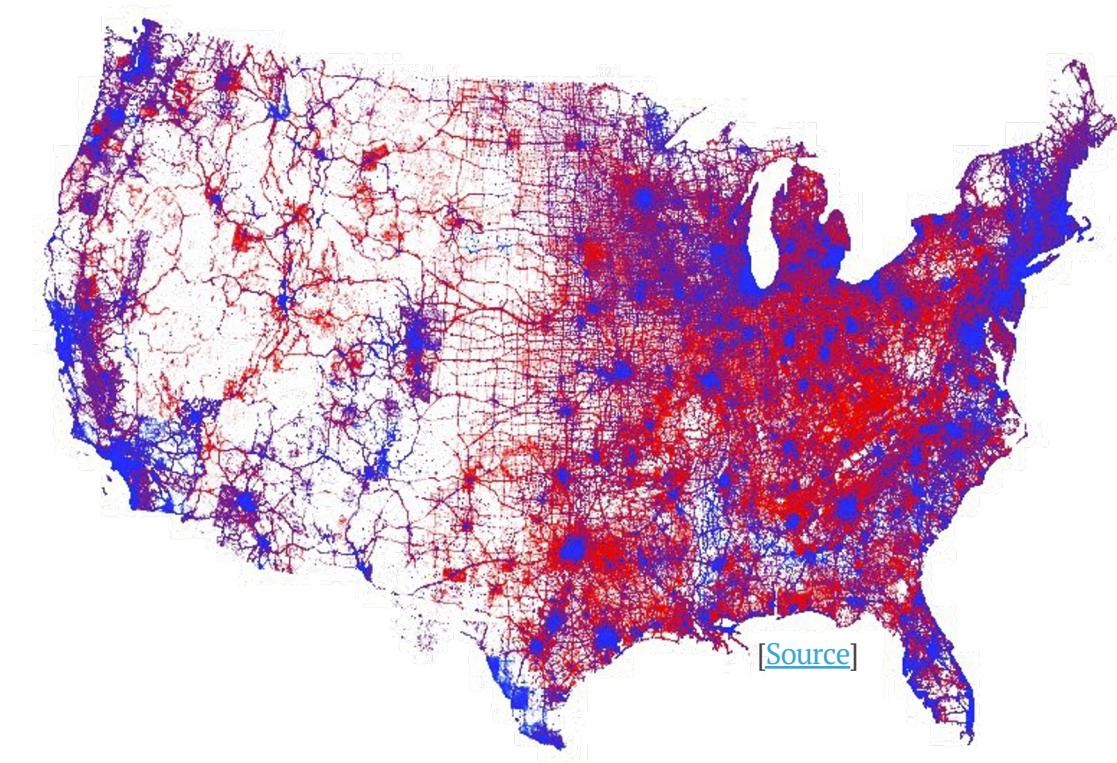
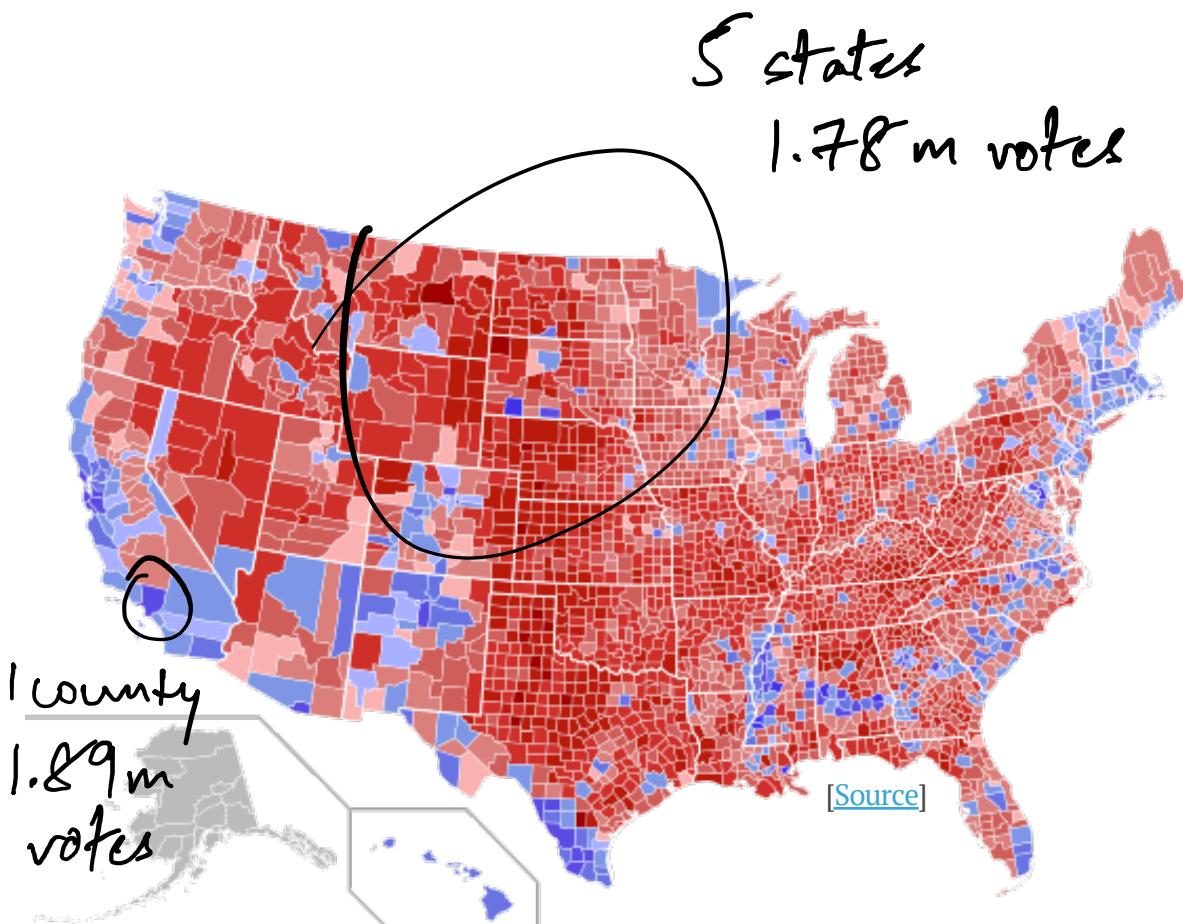
Bromley, UK

Easting: **540120**, Northing: **I69366**

# interactive map

# *Dangers* of GeoVisualisation

## 2016 Presidential Election in the USA



# *How to lie with maps\**

The human brain is so good at picking up patterns...  
... that it finds them even where they don't exist!

***Patternicity*** (Shermer, 2008) The tendency to find meaningful patterns in meaningless noise

***Apophenia*** (Konrad, 1958) The experience of seeing patterns or connections in random or meaningless data

# How to *be truthful* with maps

“With great power comes great responsibility”

Statistics to the rescue!!!

- Complement and enhance visuals
- Help disentangling **true** from **spurious** patterns
- **Reciprocity:** GeoVis can also enhance statistics and make them more useful

# How to *be truthful* with maps

- What is the story?
- Who is the audience?
- Did the visual communicate the story?
- What story was not told?
- How are the design elements used?

# All plots are composed of

- *Data* that you want to visualize
- *Layers* made up of geometric elements
- *Scales* which map values to aesthetics
- *Systems* of coordinates
- *Facets* and their specification
- *Themes* controlling finer points

# Break



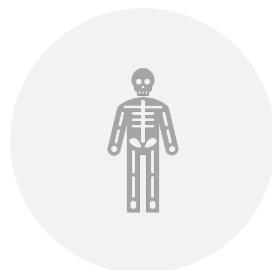
WATER



WALK



COFFEE OR TEA

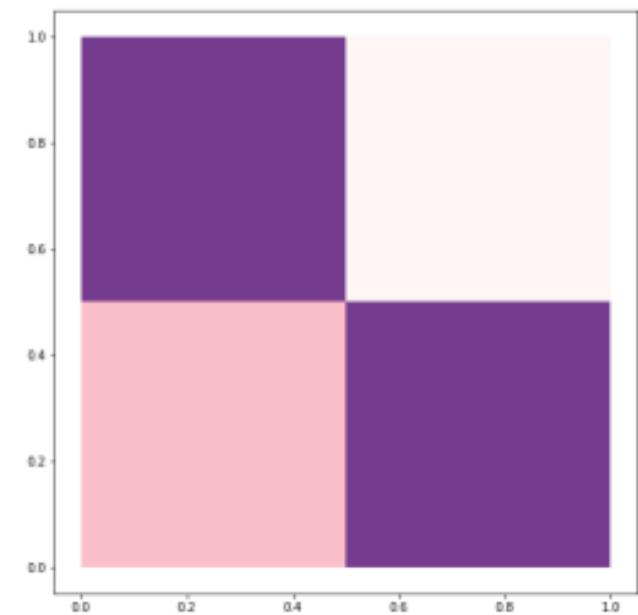
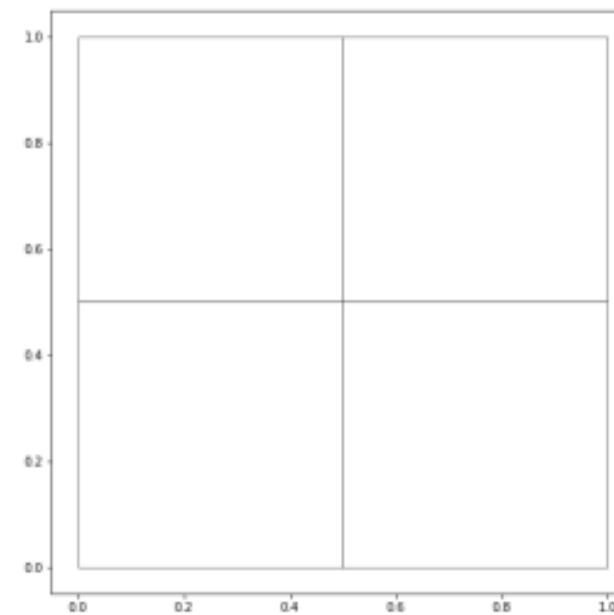
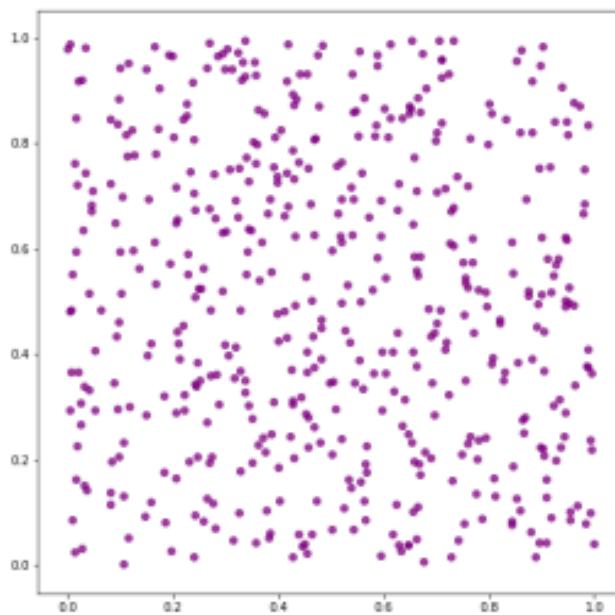


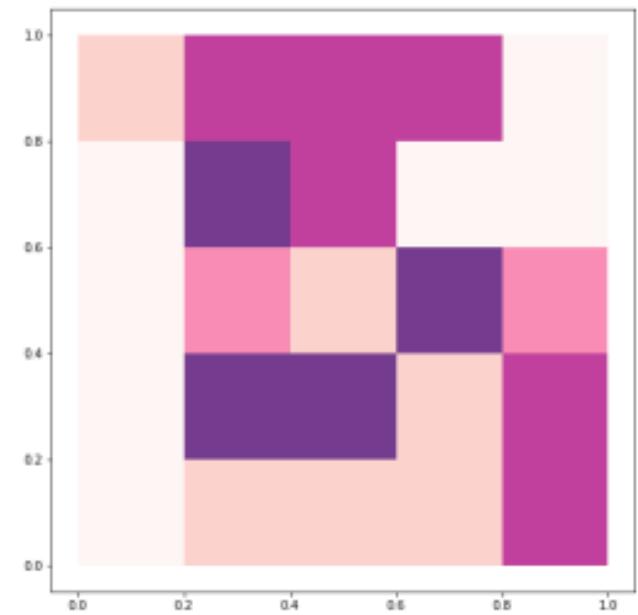
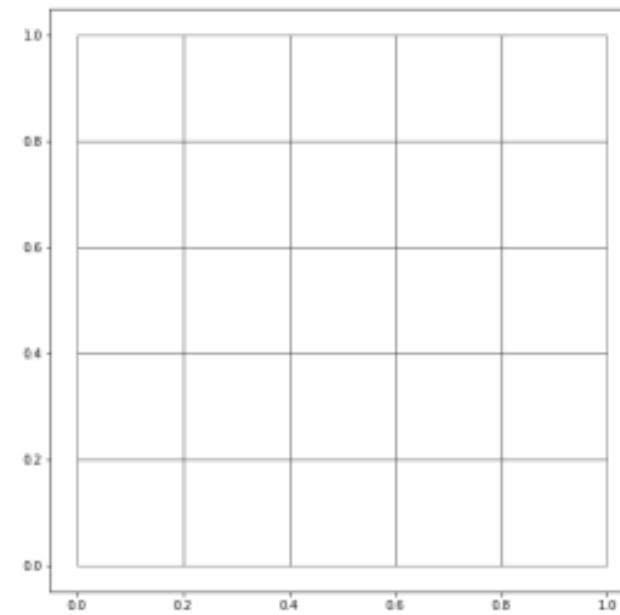
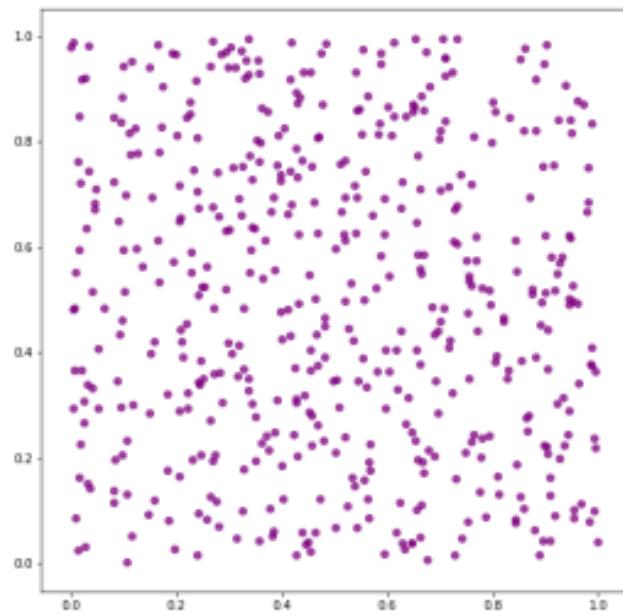
MAKE FRIENDS

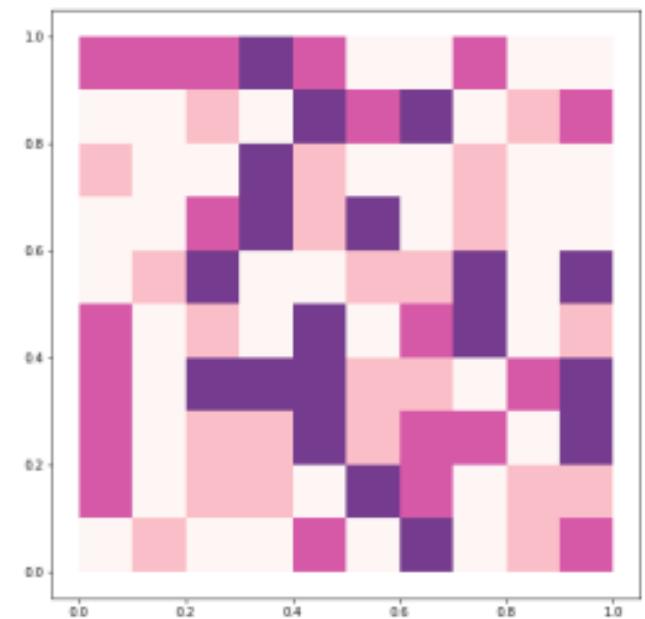
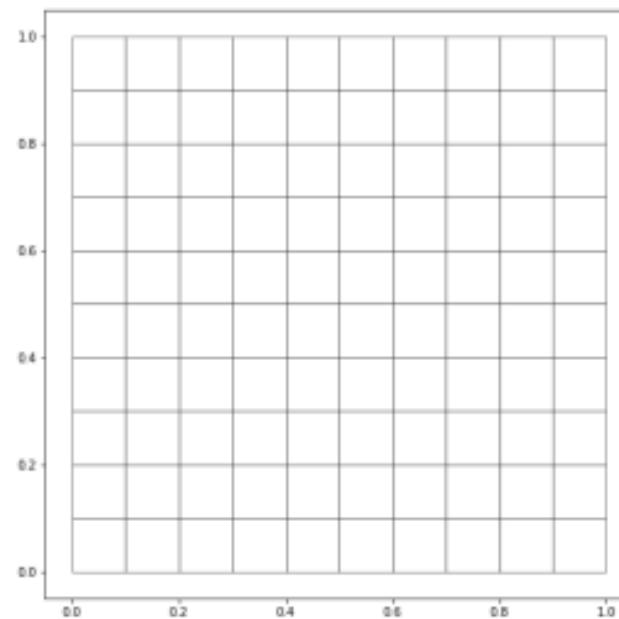
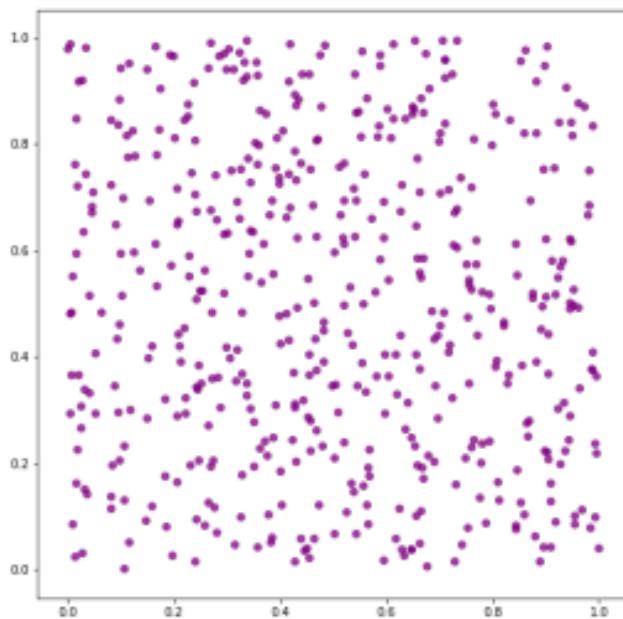
# Making good data maps

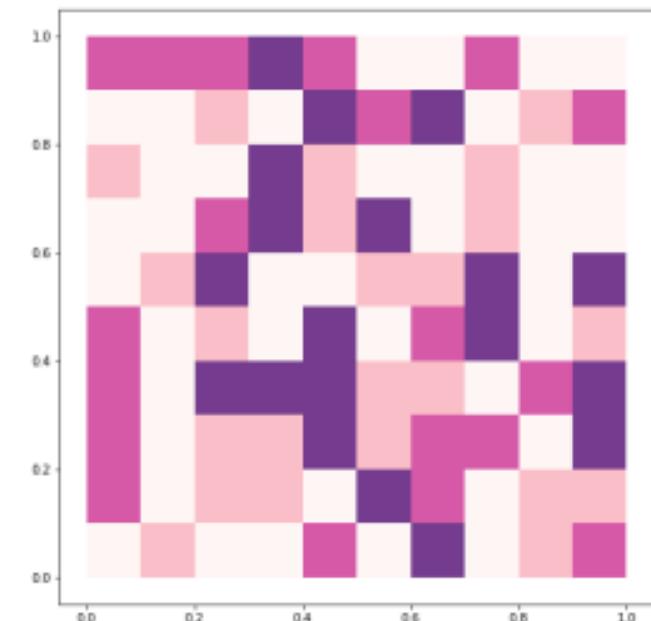
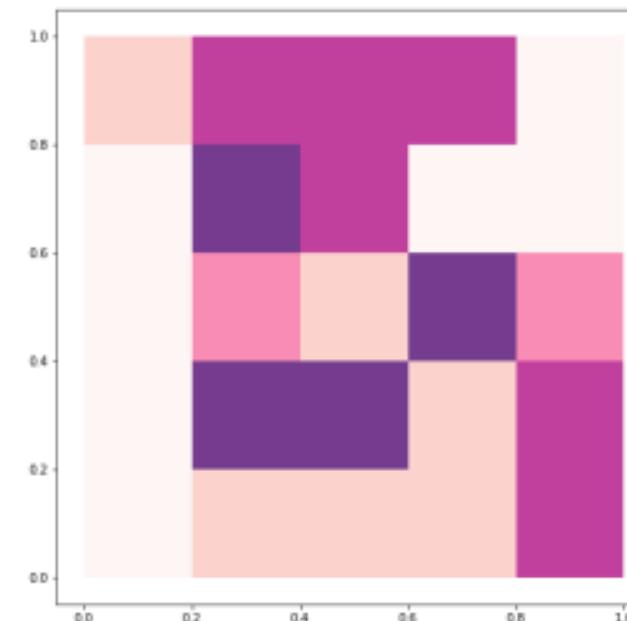
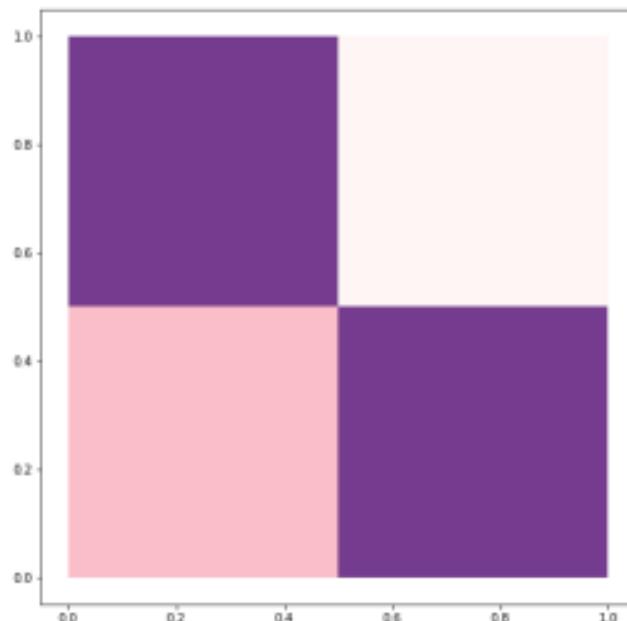
- MAUP
- Choropleths

# Modifiable Areal Unit Problem (Openshaw, 1984)









# MAUP

**Scale and delineation mismatch** between:

- Underlying process (e.g. individuals, firms, shops)
- Unit of measurement (e.g. neighborhoods, regions, etc.)

In some cases, it can **seriously mislead** analysis on aggregated data (e.g. [FLINT, MI](#))

Always keep MAUP in mind when exploring aggregated data!!!

# Choropleths

# Choropleths

*Thematic map in which values of a variable are encoded using a colour gradient of some sort*

- Counterpart of the histogram
- **Values are classified** into specific colours: value → bin
- **Information loss** as a trade off for **simplicity**

# Classification Choices

- No. of bins
- How to bin?
- Colours

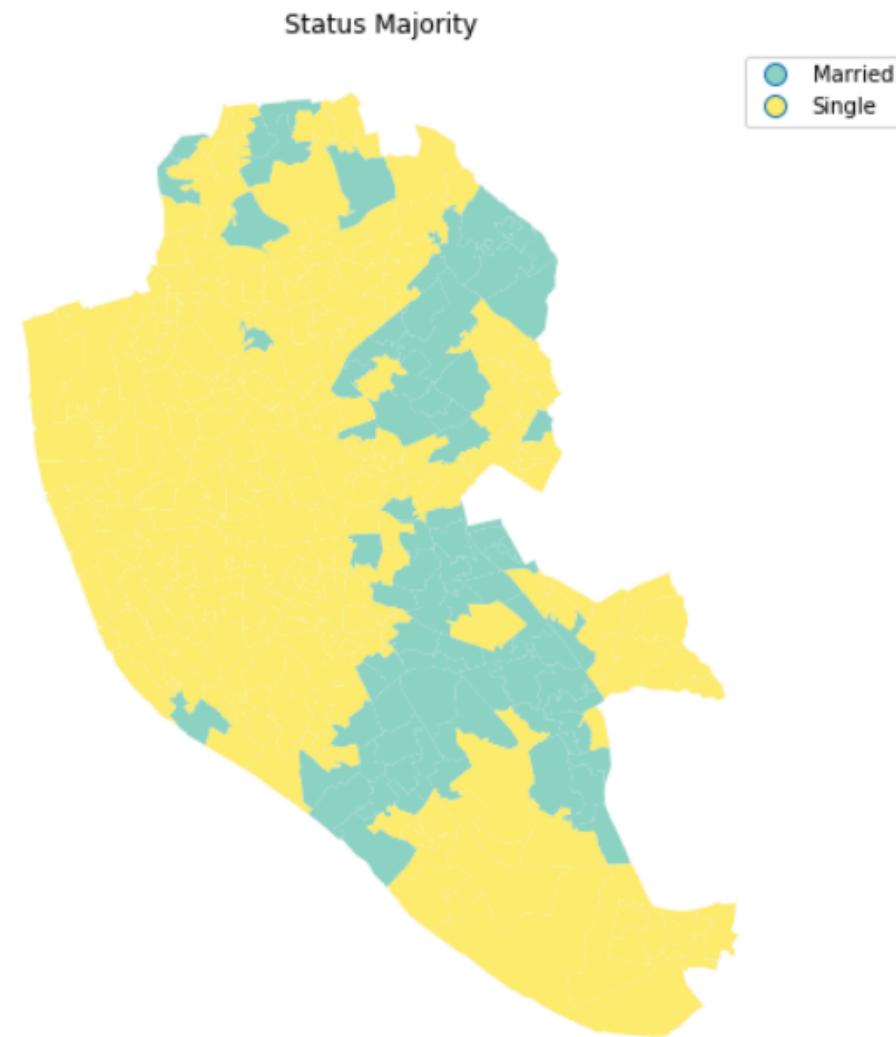
# How many bins

- Trade-off: detail vs cognitive load
- Exact number depends on purpose of the map
- Usually not more than 12

# How to bin?

# Unique values

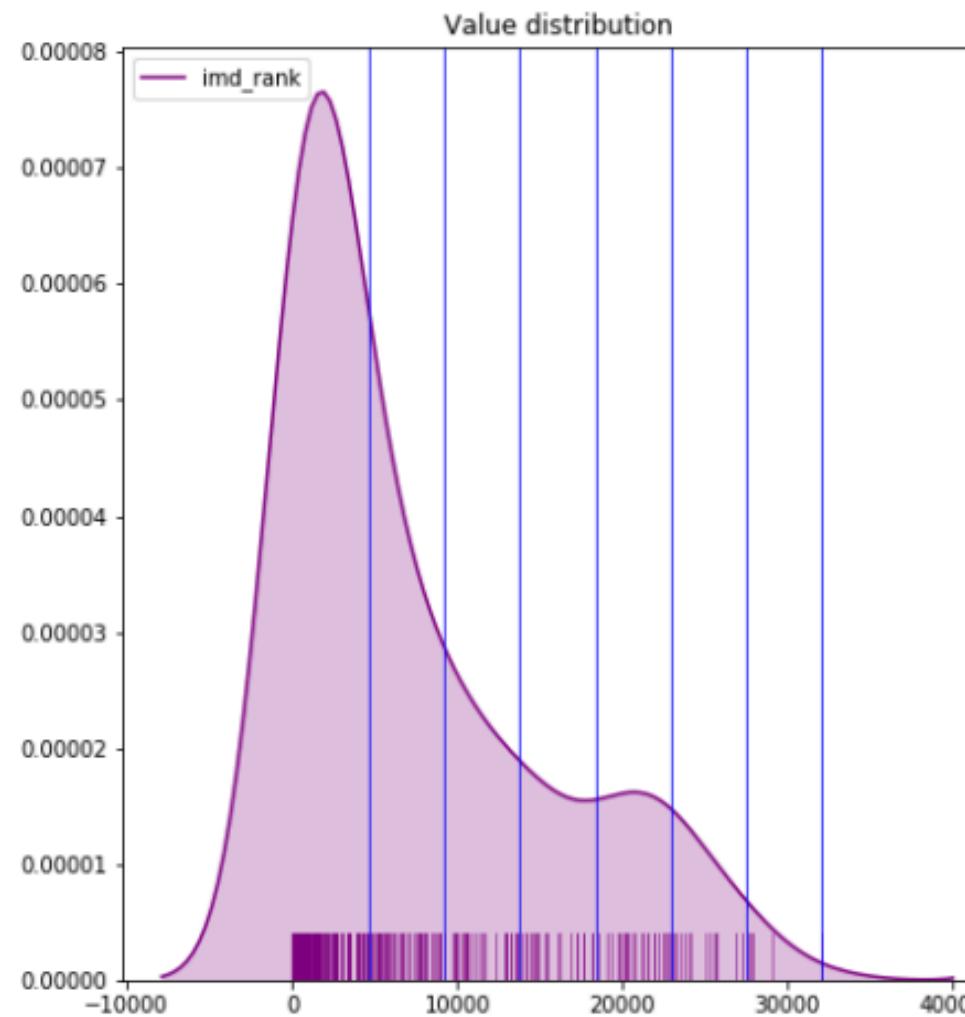
- Categorical data
- No gradient (reflect it with the colour scheme!!!)
- Examples: Religion, country of origin...



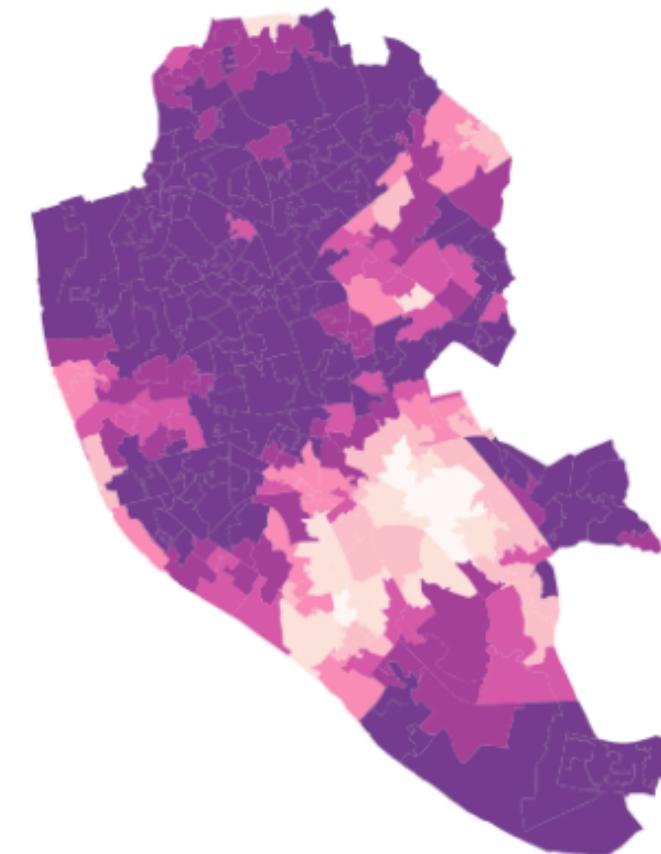
# Equal interval

- Take the **value** span of the data to represent and split it equally
- **Splitting** happens based on the **numerical value**
- Gives more weight to outliers if the distribution is skewed

# equal\_interval



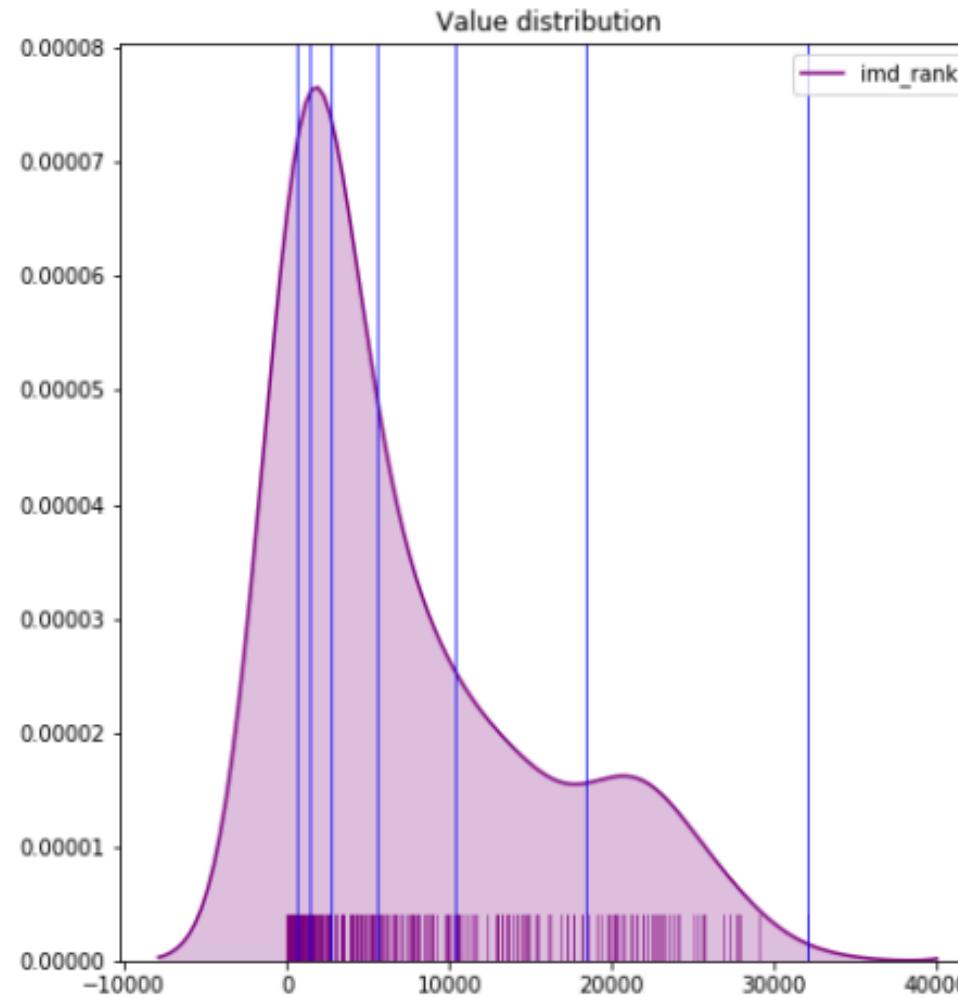
Geographical distribution



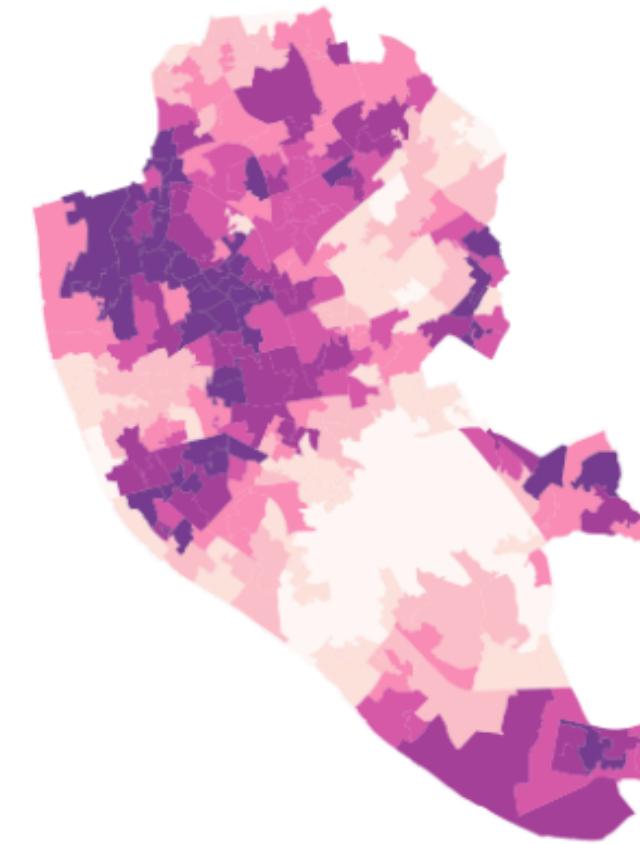
# Quantiles

- Regardless of numerical values, split the distribution keeping the same amount of values in each bin
- **Splitting** based on the **rank** of the value
- If distribution is skewed, it can put very different values in the same bin

# quantiles



Geographical distribution



# Other..

- Fisher-Jenks
- Natural breaks
- Outlier Maps: box maps, std. maps

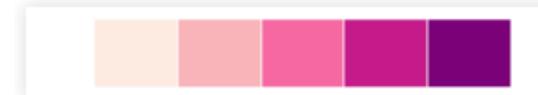
# Colour Schemes

Align with your purpose

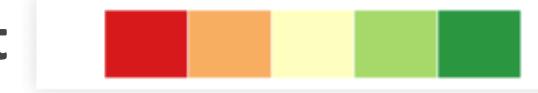
- **Categories**, non-ordered



- Graduated, **sequential**



- Graduated, **divergent**



TIP: check [ColorBrewer](#) for guidance

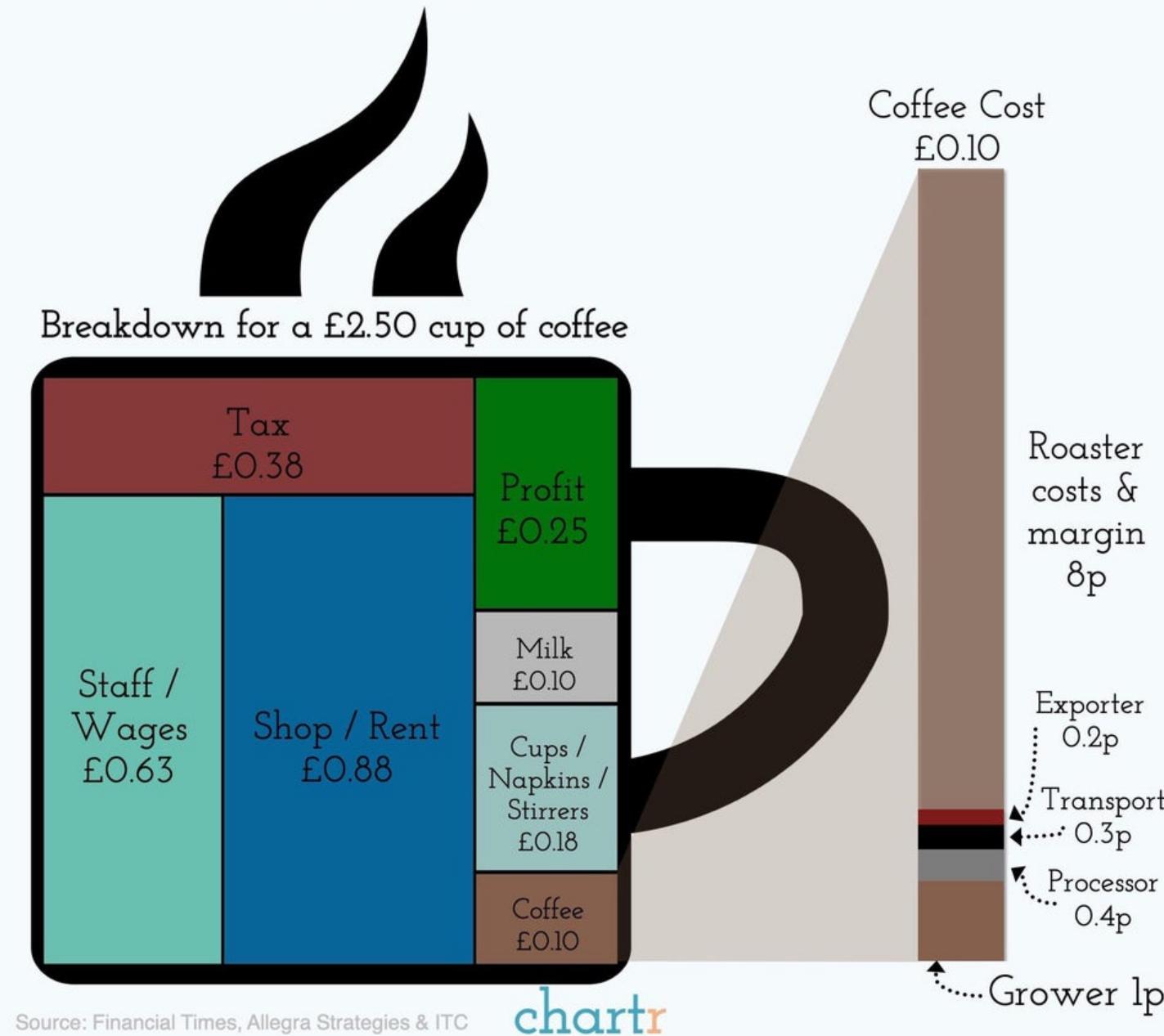
# Tips

- Think of the purpose of the map
- Explore by trying different classification alternatives
- Combine (Geo)-visualisation with other statistical devices

# The Economics behind Coffee

## Recap

- Visualization of statistical data is a recent phenomenon.
- Needs of the audience are key.
- Data can tell more than one story at a time.
- Its power comes from engaging and augmenting the human in the loop, rather than replacing her.
- Its power can be misused, so think twice.



# For next class..



**Finish** Lab 04 to practice programming



**Submit** Homework 04 for peer review on Brightspace



**Check** Assignment 2 – due in **Week 5** on Friday at **2330**



**See** “To do before class” for every lecture (~ 1 hour of self study)



**Read** paper for **Discussion** session before every Friday



**Post** questions on the **Discussion** forum on Brightspace (especially on **Visualisation** for this week)