

Not Just Pretty Pictures

Dr Jon Minton

AQMeN Research Fellow

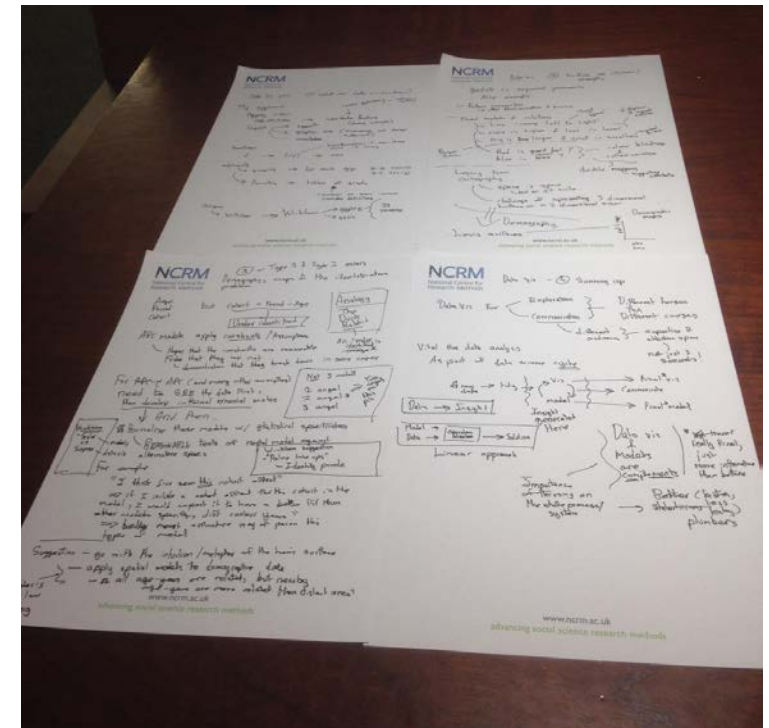
University of Glasgow

Contents

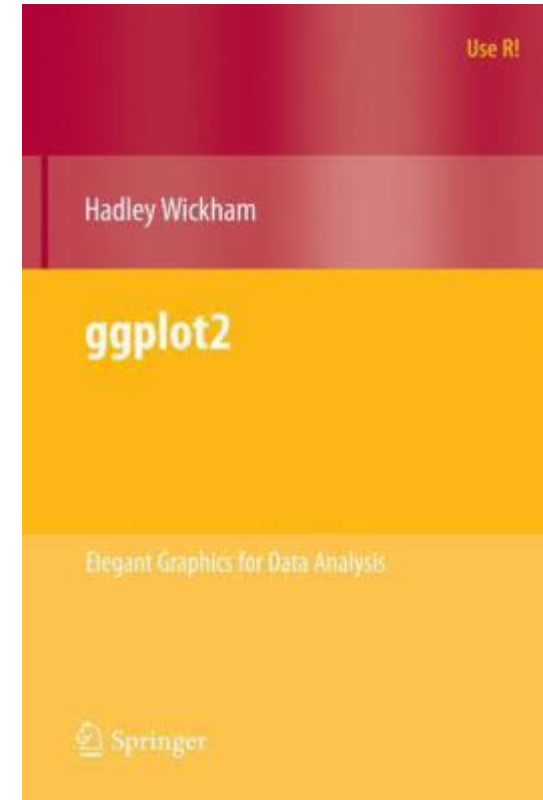
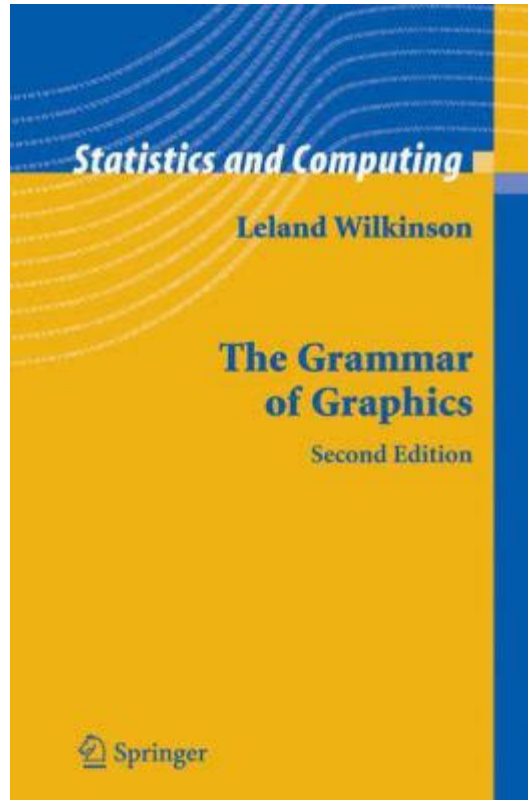
- What are data visualisations?
- How do data visualisations build on human strengths and intuitions?
- Data Visualisation and Statistical Modelling
- Summing up

What are data visualisations?

- Various definitions
- I'm partisan
- **Grammar of Graphics Definition**
- (Excluding my notes, right)



Grammar of Graphics



Grammar of Graphics

Layers

1. Support Layer
 2. Data Layer
 3. Annotation Layer
- Data Layer, defined by consistently applied mapping rules, is a **necessary condition** for an image being a data visualisation rather than something else
 - Not (usually) a **sufficient condition** for a data visualisation to be a good data visualisation

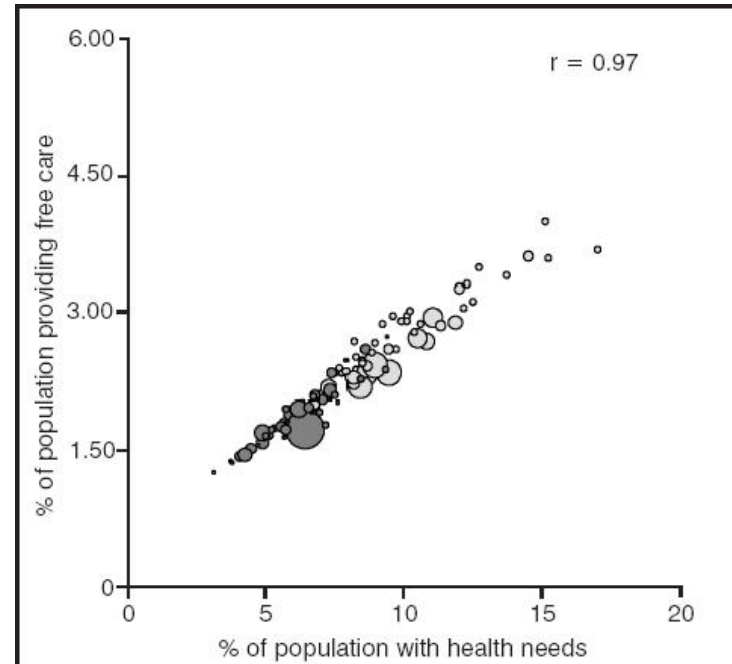
Mapping Rules

- Consistently applied
- Defined a priori

- From variables in data
- To graphical elements on an image

Simple Mapping Rules Example

Graphic



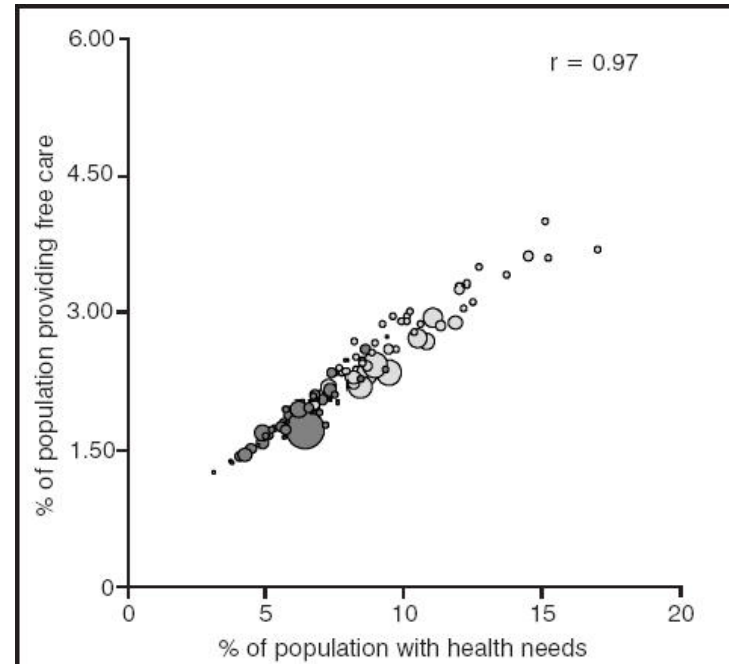
Mapping Rules

Data Variable		Graphical Aesthetic
% of population providing free care		Position along horizontal axis
% of population with health needs		Position along vertical axis
Areal unit population		Size of bubble
Geographical location (North or south)		Colour of bubble

Source: Shaw M & Dorling D (2004) "Who cares in England and Wales? The Positive Care Law: cross-sectional study" *British Journal of General Practice*, 54 (509): 899-903

Simple Mapping Rules Example

Graphic



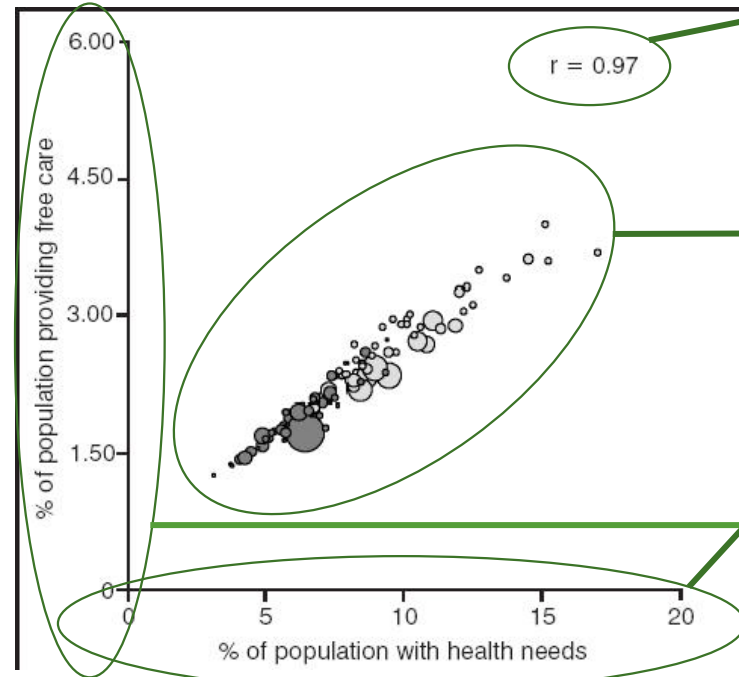
Mapping Rules

Data Variable		Graphical Aesthetic
% of population providing free care		Position along horizontal axis
% of population with health needs		Position along vertical axis
Areal unit population		Size of bubble
Geographical location (North or south)		Colour of bubble

Source: Shaw M & Dorling D (2004) "Who cares in England and Wales? The Positive Care Law: cross-sectional study" *British Journal of General Practice*, 54 (509): 899-903

Simple Mapping Rules Example

Graphic



Layers

Annotation

Data

Support

Source: Shaw M & Dorling D (2004) "Who cares in England and Wales? The Positive Care Law: cross-sectional study" *British Journal of General Practice*, 54 (509): 899-903

Benefits of Data Visualisations

Sequential Processing

"Can you do Addition?" the White Queen asked. "What's one and one and one and one and one and one and one and one and one and one?"

"I don't know," said Alice. "**I lost count.**"

"She can't do Addition," the Red Queen interrupted.

Sources:

<http://www.shmoop.com/alice-in-wonderland-looking-glass/education-quotes-4.html>

<http://www.captcha.net/>

<https://upload.wikimedia.org/wikipedia/en/b/b9/MagrittePipe.jpg>

Gestalt Processing

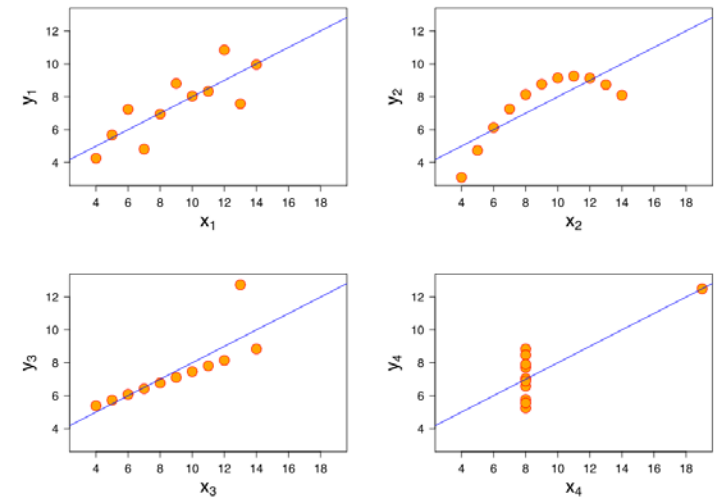


Benefits of Data Visualisations

Scale and Surprise

- Models scale
- Data visualisations
Surprise

Anscombe Quartet



Sources: <http://r4ds.had.co.nz/intro.html>
https://en.wikipedia.org/wiki/Anscombe%27s_quartet

Intuitions that data visualisations build on

Mental models and metaphors

- Time 'runs from left to right' (position)
- More is 'higher' and less is 'lower'. (position)
- More is 'bigger' and less is 'smaller' (size)
- Red is 'bad' & blue is 'good'?

Use of colour

Cultural variation



Accessibility



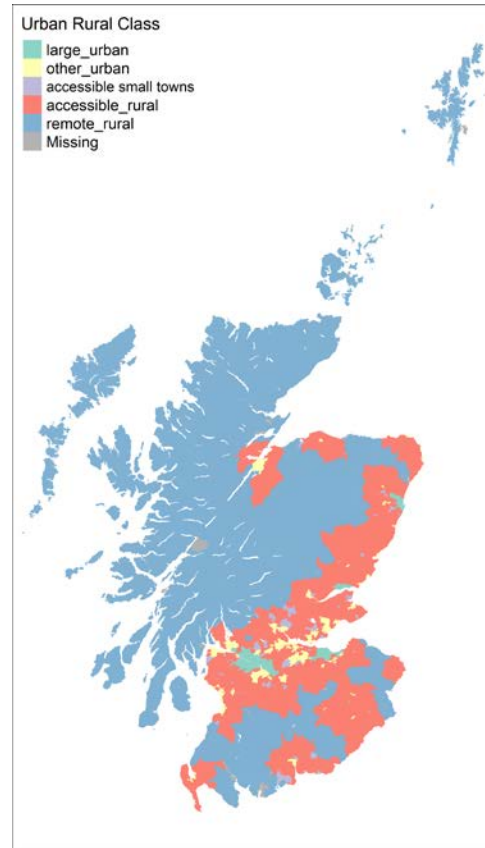
Data Variable		Graphical Aesthetic
Variable 1		
Variable 2		Colour
Variable 3		Something else
Variable 4		

Source:

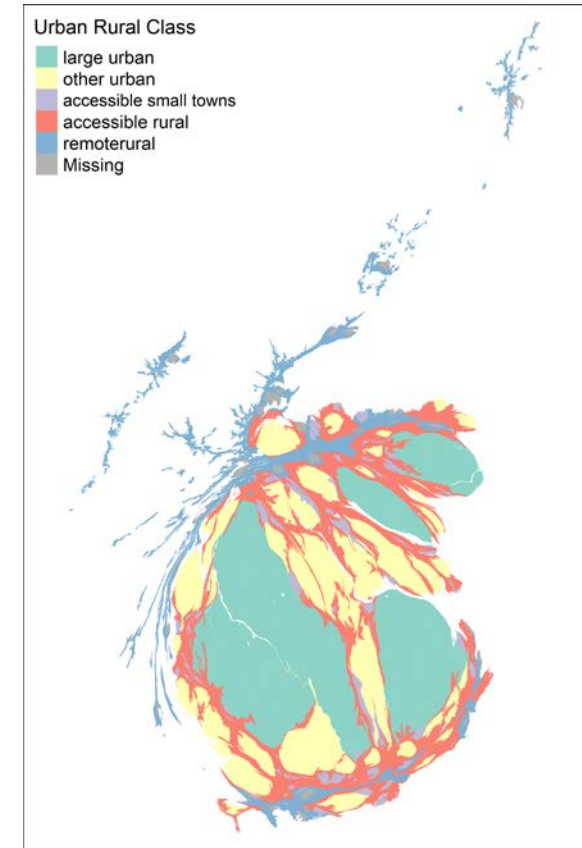
<http://www.ebay.co.uk/itm/like/261727816290?lpid=122&chn=ps&googleloc=9046939&poi=&campaignid=620865095&device=c&adgroupid=27378760866&rlsarget=pla-181484341026&adtype=pla&crdt=0>

Some Examples

Standard Urban Rural Map

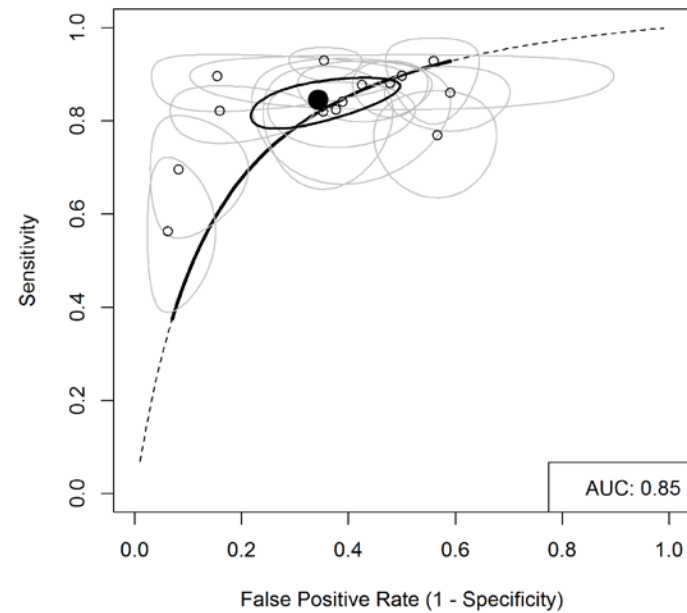


Cartogram by population

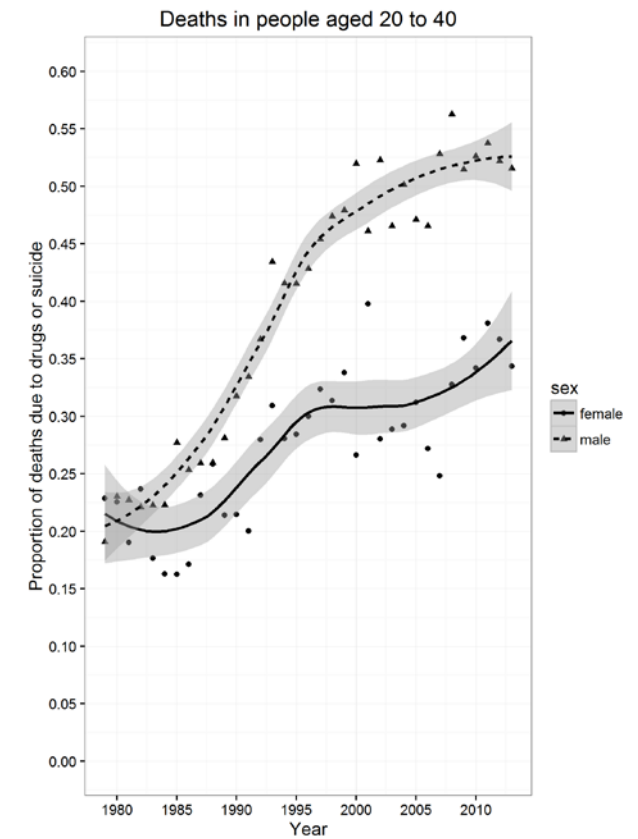


Some Examples

Diagnostic Meta-Analysis

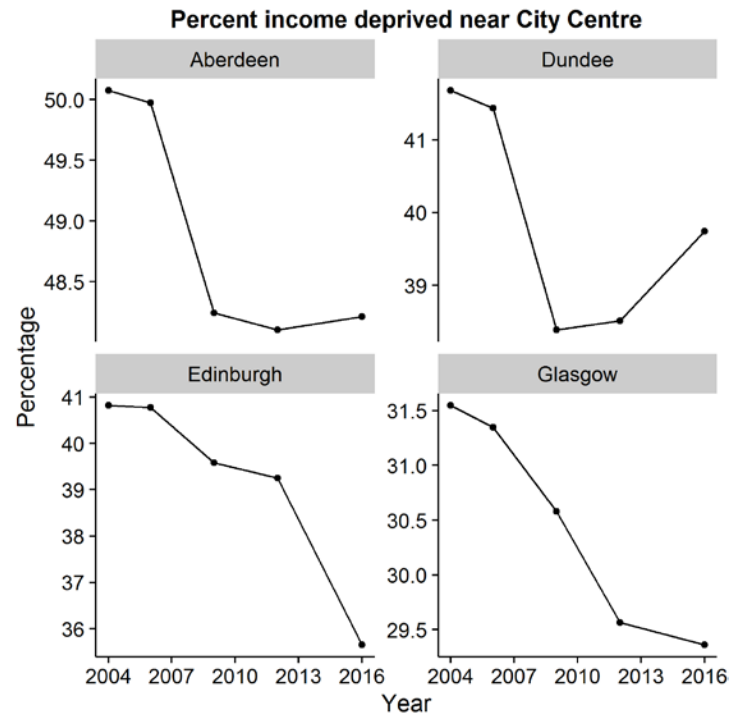


Suicide and Drug Deaths

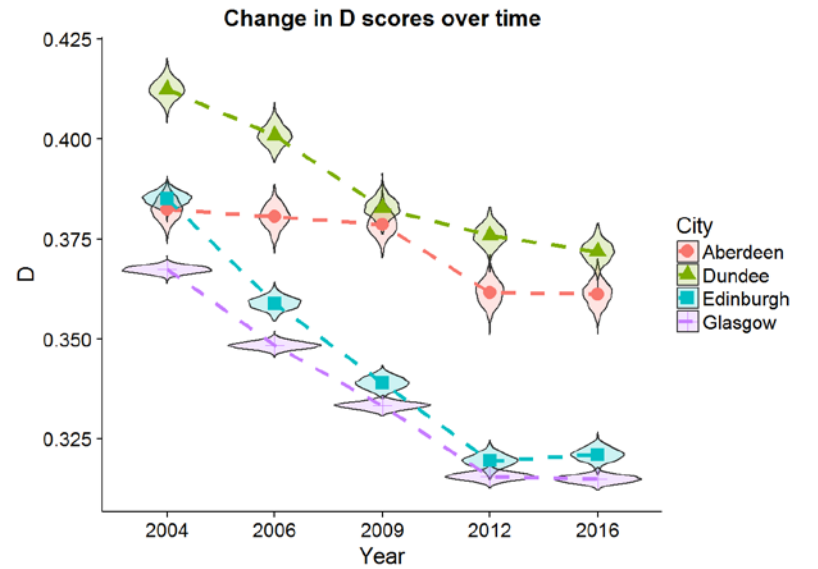


Some Examples

Decentralisation of poverty

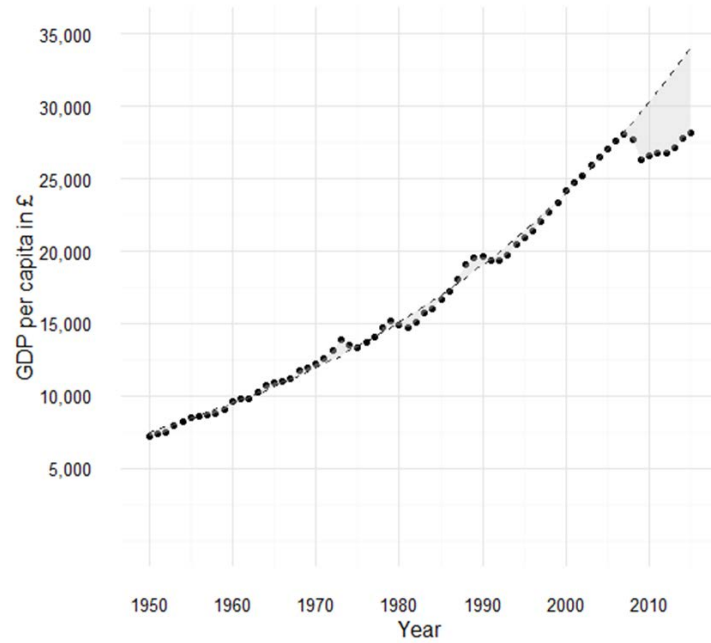


Index of Dissimilarity

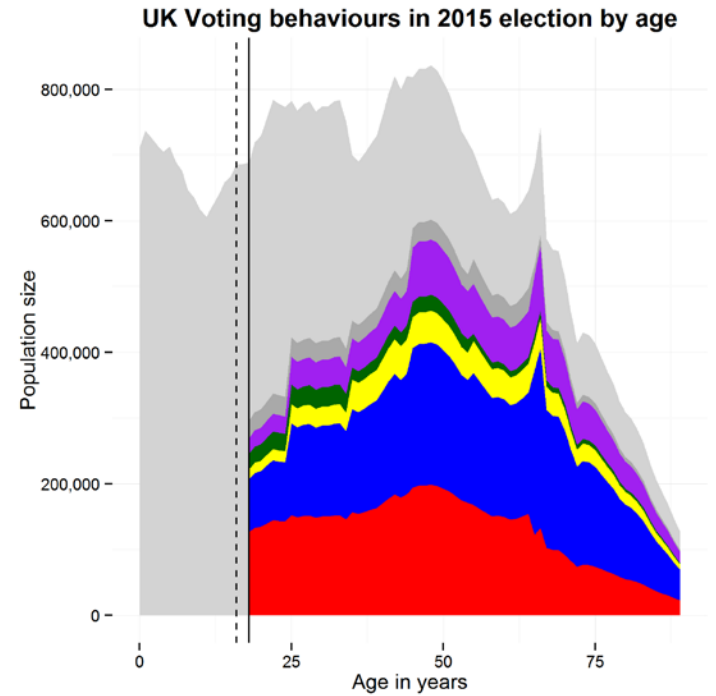


Some Examples

Per Capita GDP trends in UK



(Non)voting by age in 2015



My special interest

Lexis surface

- Demography as cartography
- Latitude, Longitude, Height
- Age, Year, something else
- Age: relative time
- Year: absolute time

Challenge of representing 3rd variable

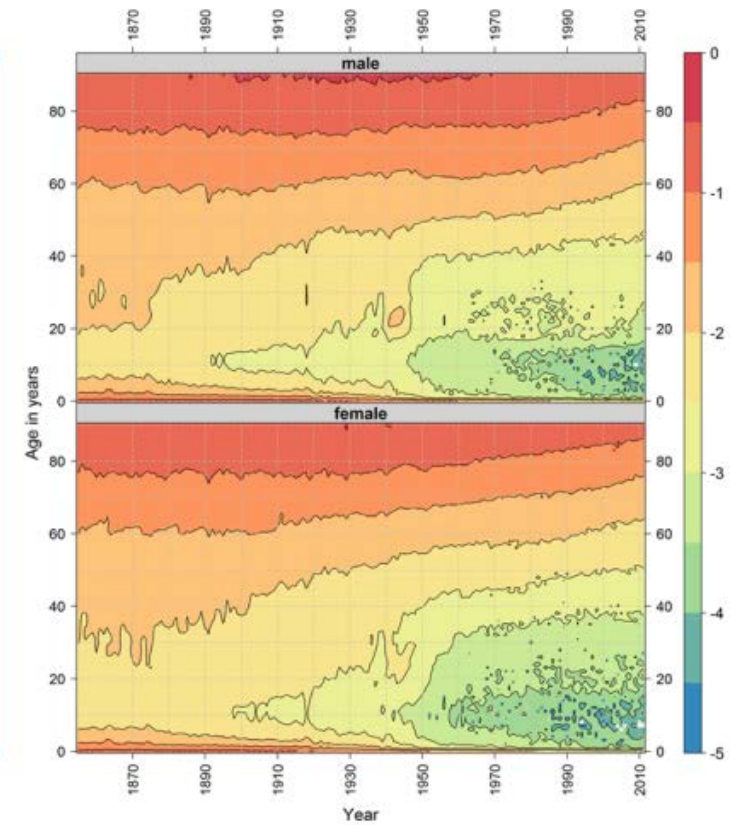
- Two dimensional 'canvas'
- Three dimensional relationship
- Double-mapping of 3rd attribute
 - Colour/shade
 - Contour lines
- Contour lines link places on surface w/ same 3rd value

Maps of space and time

Topographic map of Scotland



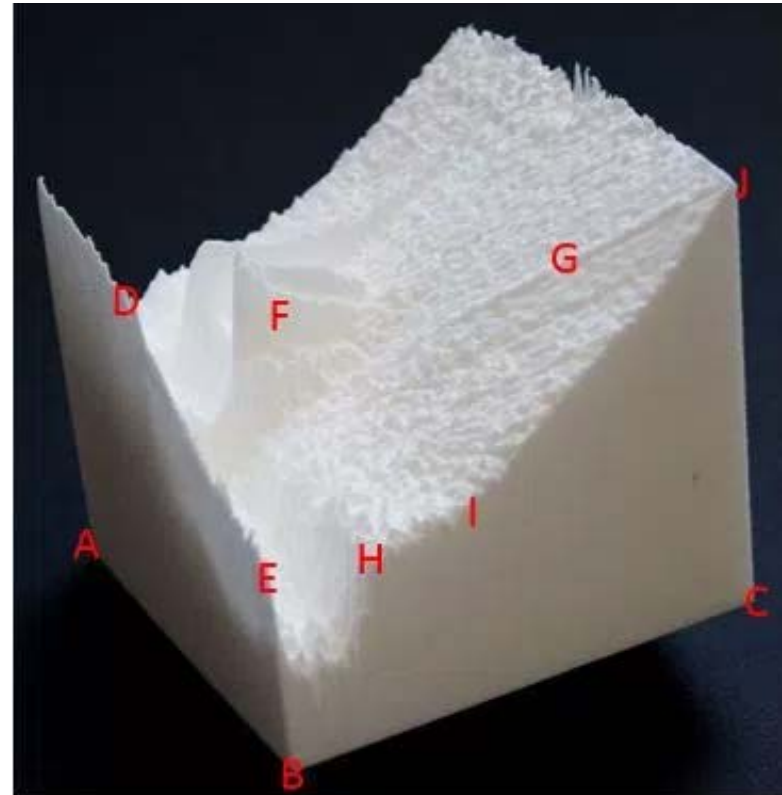
Log mortality surface map



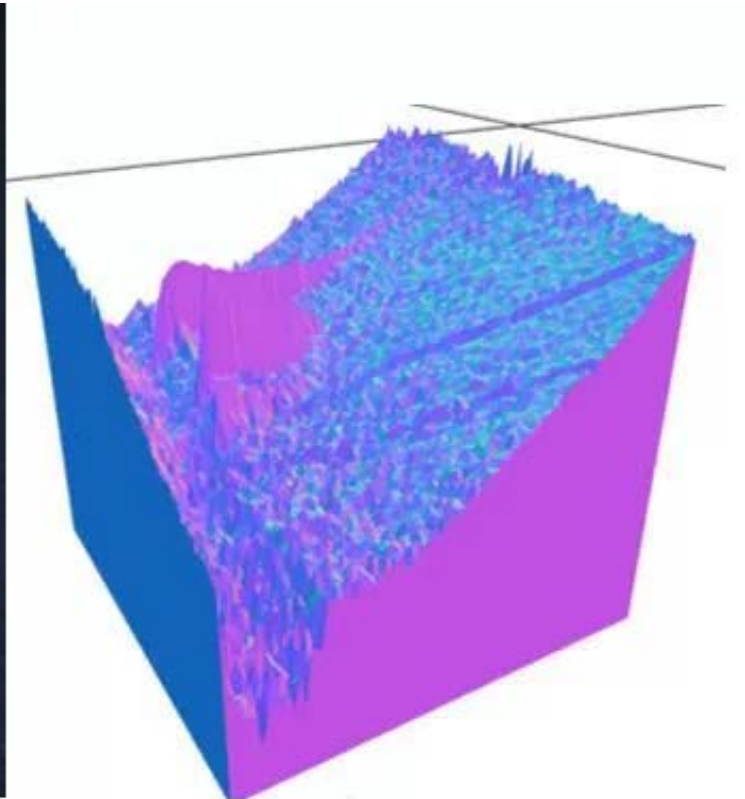
Source: <https://ije-blog.com/2016/06/27/lexis-cubes-1-from-maps-of-space-to-maps-of-time/>

Escaping the Canvas

3D Printed 'Data Cube'



Github rendering



Source: <https://ije-blog.com/2016/06/27/lexis-cubes-2-case-study-log-mortality-for-males-in-finland-1878-to-2012/>
[https://github.com/JonMinton/Statistical_Sculpture/blob/master/stl/individual/lmorts/fin_male_\(1878-2012\).stl](https://github.com/JonMinton/Statistical_Sculpture/blob/master/stl/individual/lmorts/fin_male_(1878-2012).stl)

Modelling of Age-Period- Cohort Effects

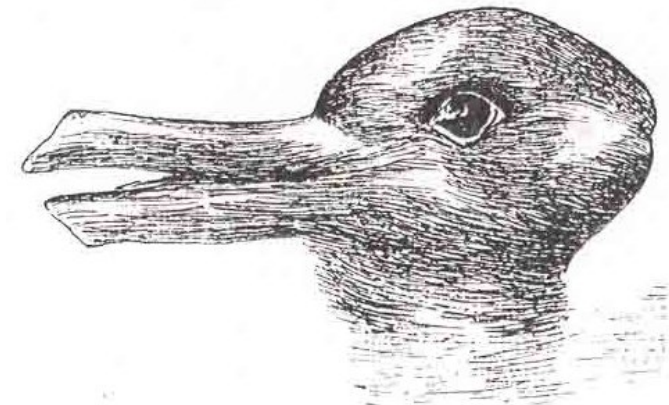
The Identification Problem

- Age matters
- Period matters
- Cohort matters

- But Cohort = Period – Age
 - An accounting identity

- Under-Identified

Visual Analogy



- Two ways of resolving the data

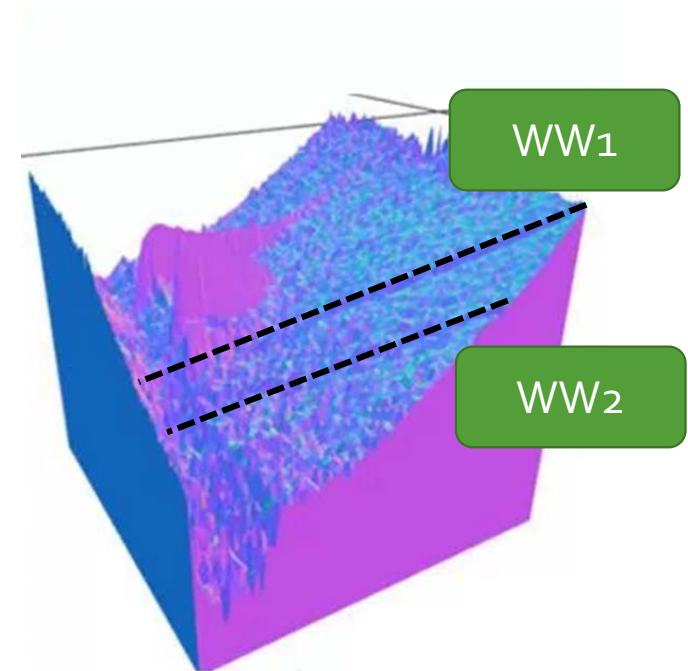
Source: https://commons.wikimedia.org/wiki/File:Duck-Rabbit_illusion.jpg

APC Debates

Argument Wars (Good for citations)

- Yang et al:
 - Intrinsic Estimators
 - Hierarchical APCs
 - Assumptions reasonable
- Bell et al
 - Assumptions still don't defeat logic
 - Simulation modelling shows assumptions not always reasonable

But (some) cohort effects are real



Data Vis Driven Approaches

Modelling as formalisation

- See the data
- Build intuitions
 - Important Age-relationships
 - Important Period relationships
 - Important Cohorts
- Also
 - Trends in age-relationships
 - Sex differences
 - Age-localised period effects
- Formalise intuitions
- Test intuitions

Lexis surfaces as a modelling metaphor

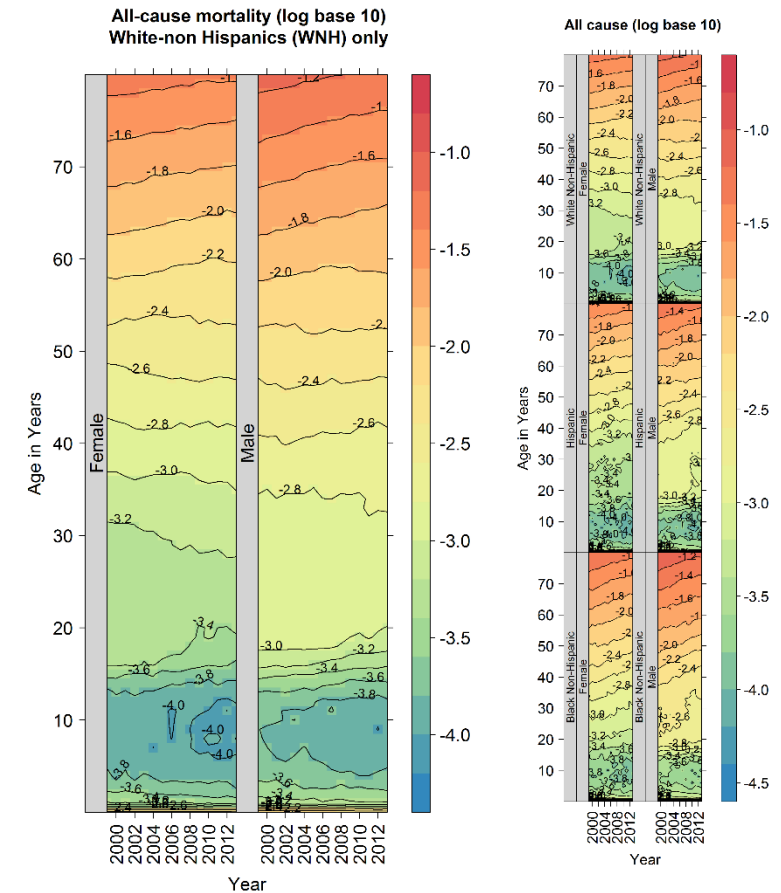
- Tobler's First Law:
 - Everything related
 - Near things more related
- Applies to space
- Can it apply to age-time?
- BYM models
 - Spatial Stats and Image Processing
- CARBayes package

Dangers of not looking enough first

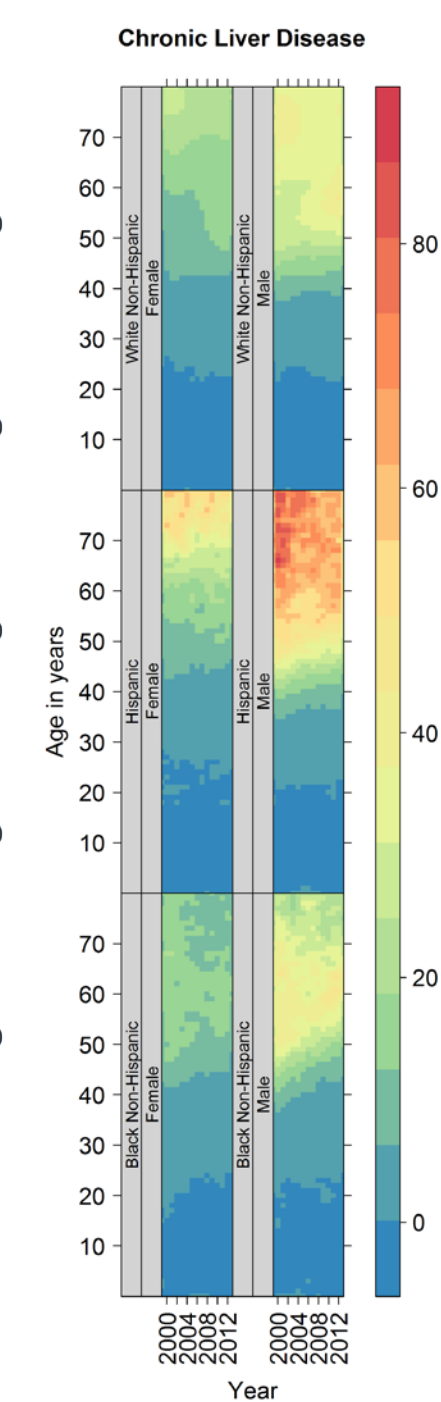
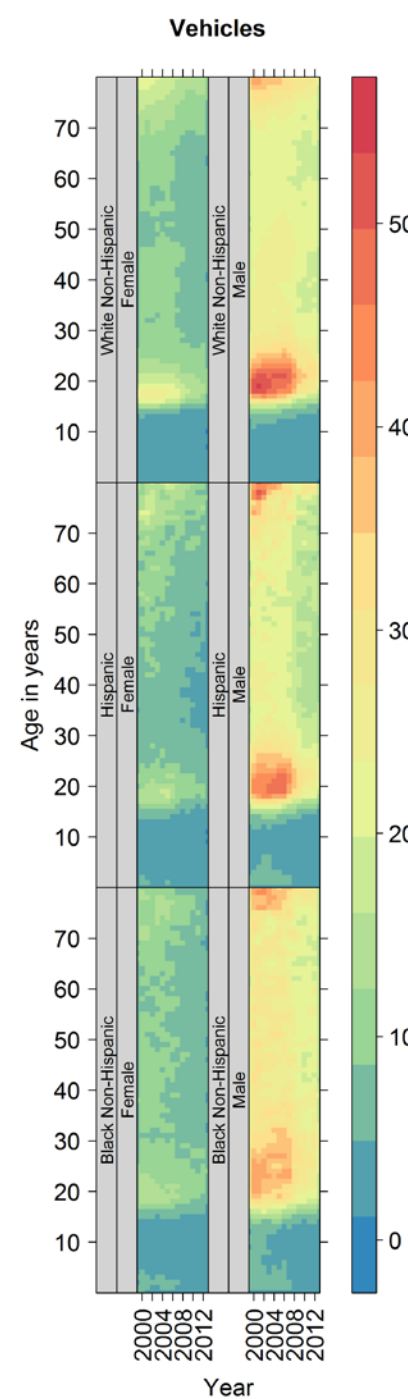
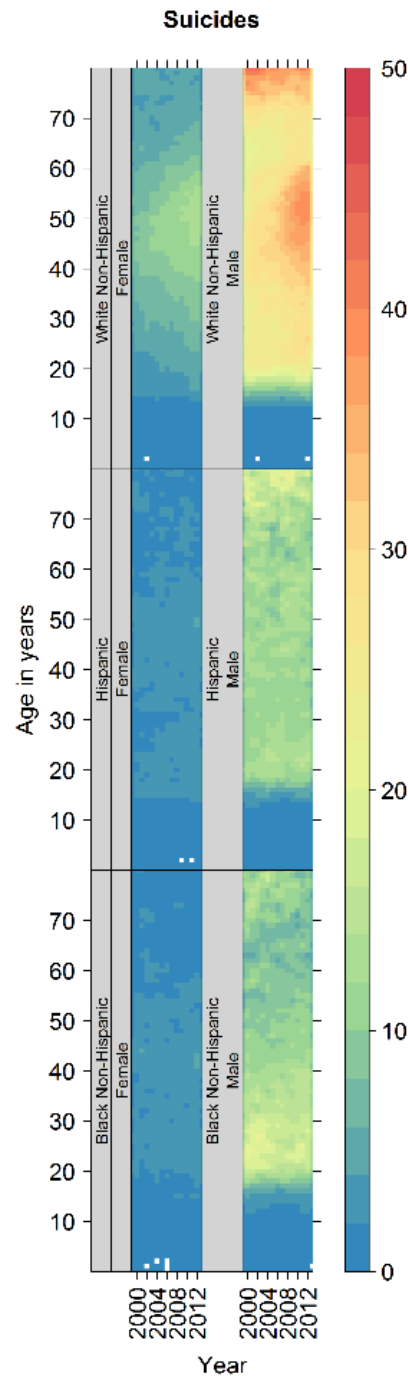
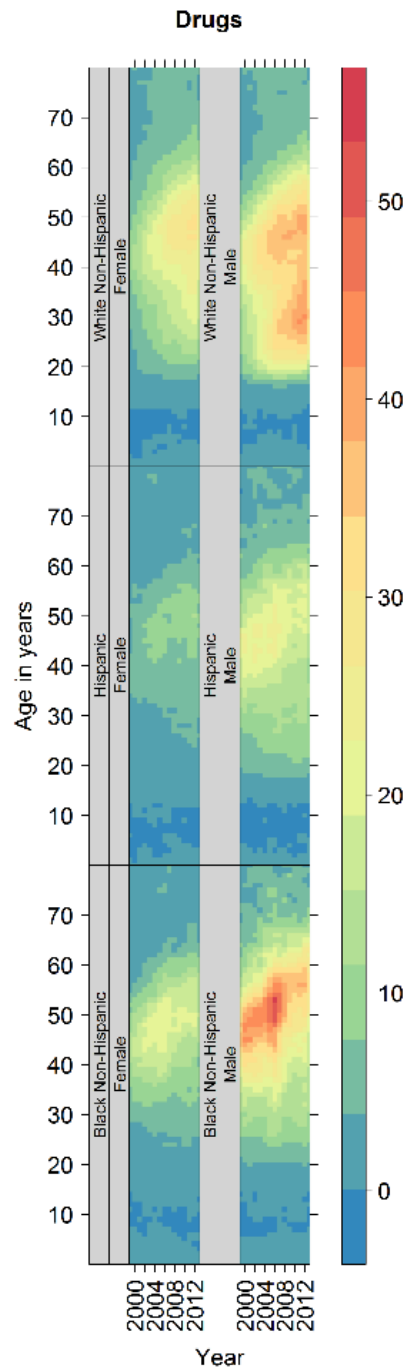
Aggregation Bias

- Case & Deaton
- Collapsing two genders and 10 years of age into one number
- Changing age-composition over time
- IJE Commentary

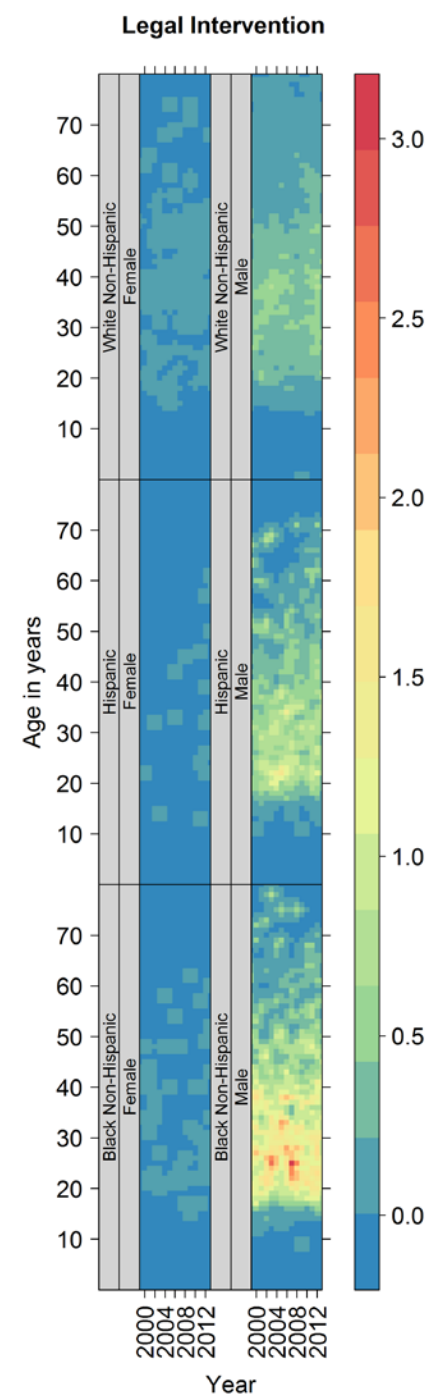
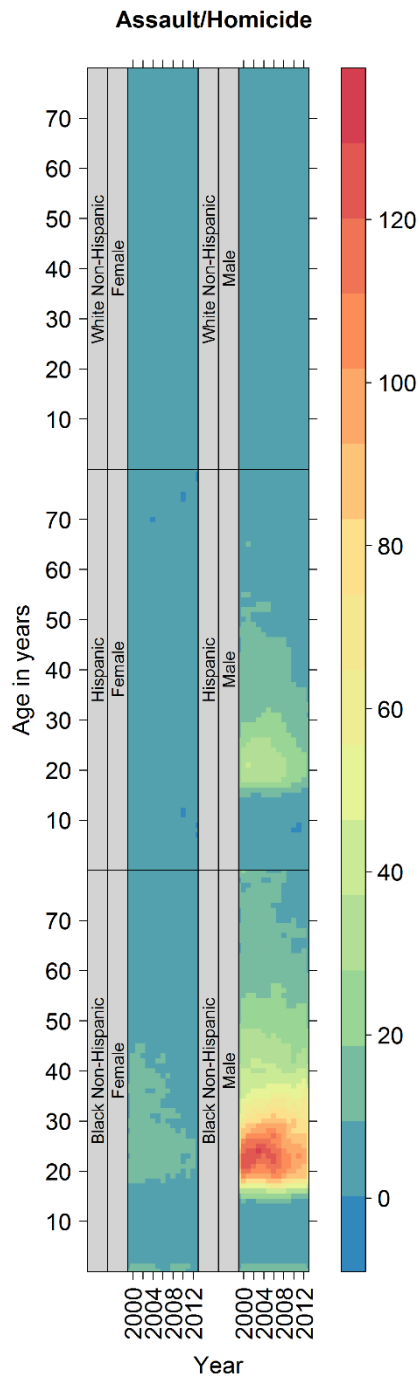
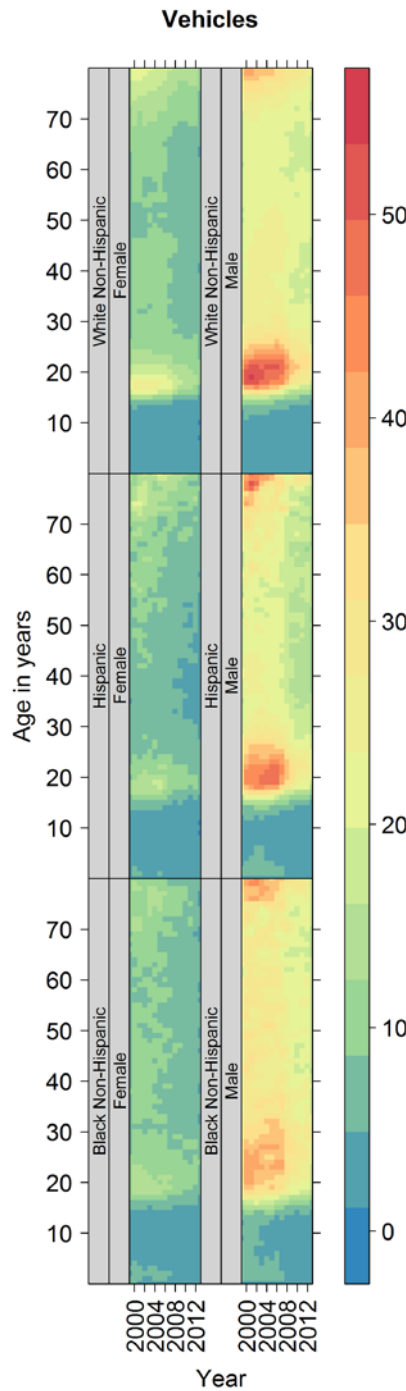
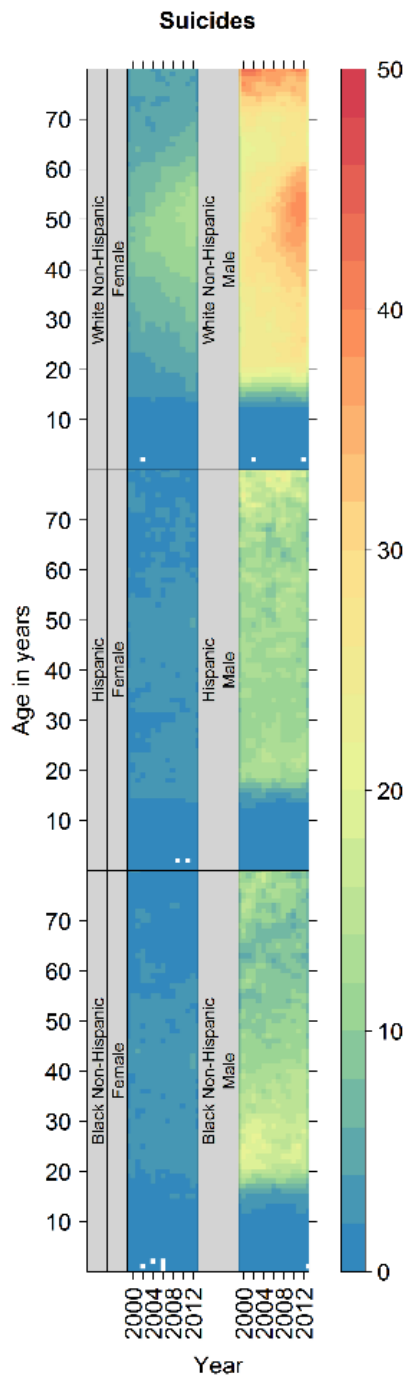
Disaggregated Data



Missed Opportunities



Missed Opportunities

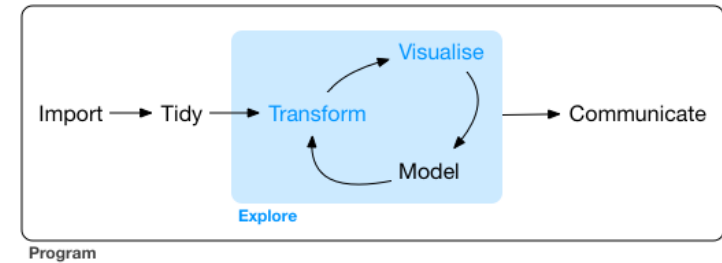


Final Thoughts

Types of Data Visualisation

- Communicatory
 - General Audience
 - Specialist Audience
- Exploratory
- “Different horses for different courses”
- The ‘three second rule’

Data Vis in Data Science



- Data to knowledge/insight
- Data Vis vital ‘in the loop’ as surprise is vital for learning from data
- Models and vis are complements:
 - Develop intuitions: Data Vis
 - Test intuitions: Modelling
 - Type 1/Type 2 error

Source: <http://r4ds.had.co.nz/explore-intro.html>