INM433: Session 01
Introduction to Visual Analytics and Visualisation

INM433 Visual Analytics

Welcome to Visual Analytics!

- Taught by
  - Lectures: Rafael Henkin, Gennady and Natalia Andrienko
  - Computer labs: Mirela, Kevin, Alex and Adam/Olga
  - Aidan Slingsby (module leader)
- 100% coursework
  - Groupwork: literature review + seminar
  - Individual work: analysis project

You?

- Backgrounds?
- People currently in employment
  - Currently “data scientists”
  - Want to introduce more data science to current job
  - Thinking of changing career
- Technical/maths background
  - Low
  - Medium
  - High

Visual Analytics

What is Visual Analytics?

- "The science and practice of **analytical reasoning** by combining **computational analysis** with **interactive visual interfaces**".
- Analytical reasoning
  - Data
    →Information
    →Knowledge
    →Solution/decision

James J. Thomas and Kristin A. Cook (Ed.) (2005)
"Illuminating the Path: The R&D Agenda for Visual Analytics"
National Visualization and Analytics Center

Why Visual Analytics?

- Divide the efforts
  - Too much data to simply explore ourselves
    - need a computer to reduce/filter/generalise/identify
  - Want to be aware of potential problems in data
  - Don’t know what we’re looking for (yet)
  - Want to incorporate “expert knowledge”
  - Want to record and/or understand the process of reasoning

Humans vs Computers

Human

- flexible and inventive, can deal with new situations and problems
- solve problems that are hard to formalise
- can cope with incomplete/inconsistent information
- can recognise things that are hard to compute or formalise

Computer

- large amounts of data
- fast search
- fast data processing
- link to other databases/services
- high quality graphics

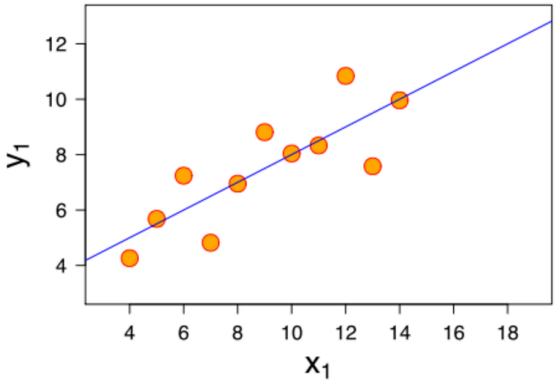
Examples from later in the module

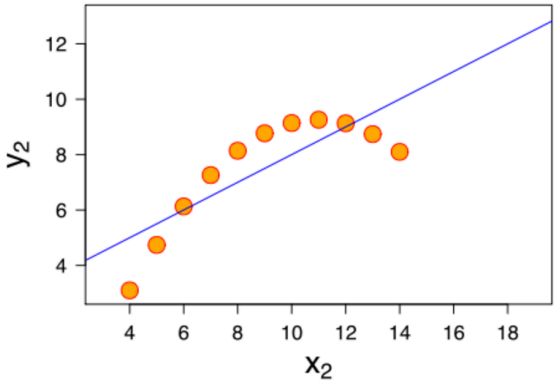
- London Bike Hire scheme
  - Are there distinct bike hiring behaviours? Which? Where? Can we manage the scheme better?
- Tweets
  - Can we detect events from tweets? Where? When?
  - Can we characterise neighbourhoods?
- Journeys
  - What are common journeys made?

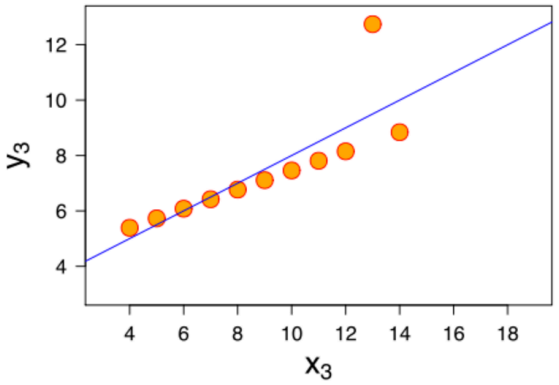
What is “visual”?

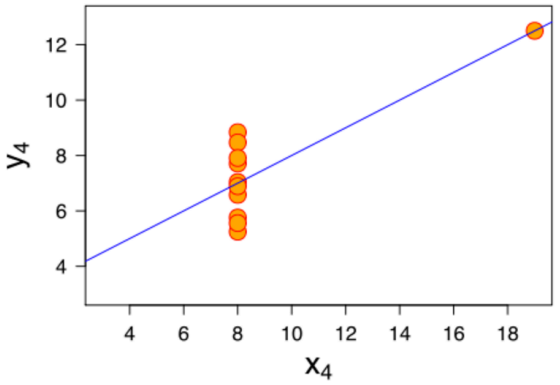
Anscombe's Quartet (1973)

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89









Mean of x: 9
Variance: 11
Mean of y: 7.50
Variance: 4.25
Correlation: 0.816

https://en.wikipedia.org/wiki/Anscombe%27s_quartet

Role of visualisation

- Present data in a way that facilitates comparison
  - Compare hundreds of numbers - humans are good as seeing visual patterns
  - See distributions and uncertainties
  - Compare alternative outputs

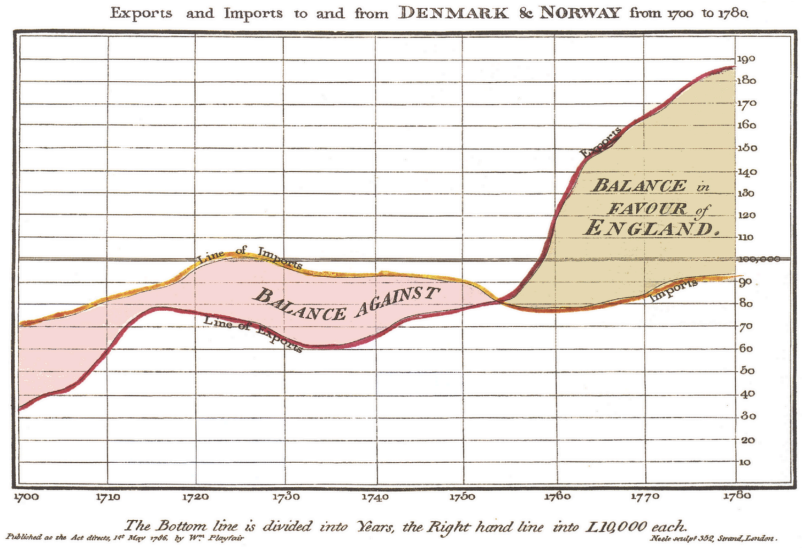
A historical perspective

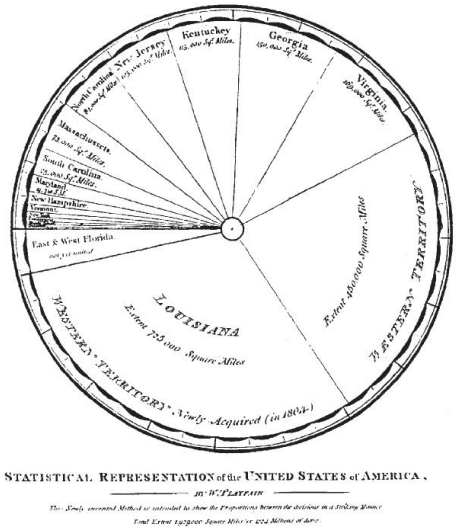
Statistical Graphics

- From the statistics domain
  - Scatter plots
  - Histograms
  - Box plots
  - Bar charts
  - Realisation that statistics may hide important things
- Exploratory data analysis (Tukey, 1977)
  - Interaction from 1970s onwards

William Playfair (1759-1823)

- Scottish engineer and political economist
- Invented modern charts:
  - Line, bar, area, pie
- The Commercial and Political Atlas, 1786





Charles Minard (1781-1870)

- French civil engineer
- Invented flow maps to depict movement between locations using line thickness
- Famous in visualisation research for Napoleon's Russia campaign map
  - 6 data attributes in 2 dimensions

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.

Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zônes colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zônes. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M.M. Thiers, de Ségur, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davoust qui avaient été détachés sur Minsk et Mobilow en ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.

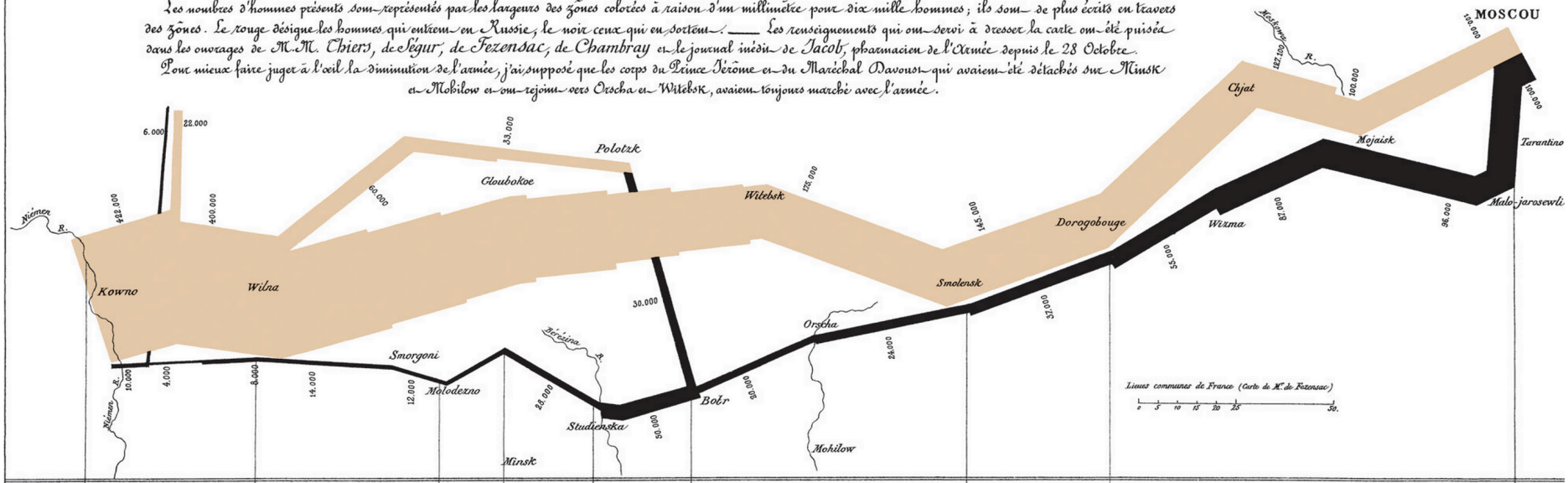
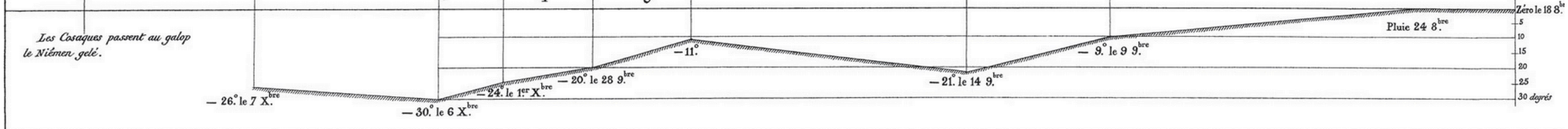


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.



Autog. par Regnier, 8. Pas. Ste Marie St Gain à Paris.

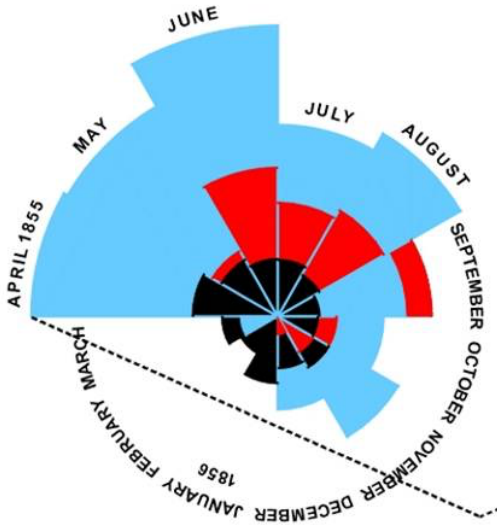
Imp. Lith. Regnier et Dourdet.

Florence Nightingale (1820-1910)

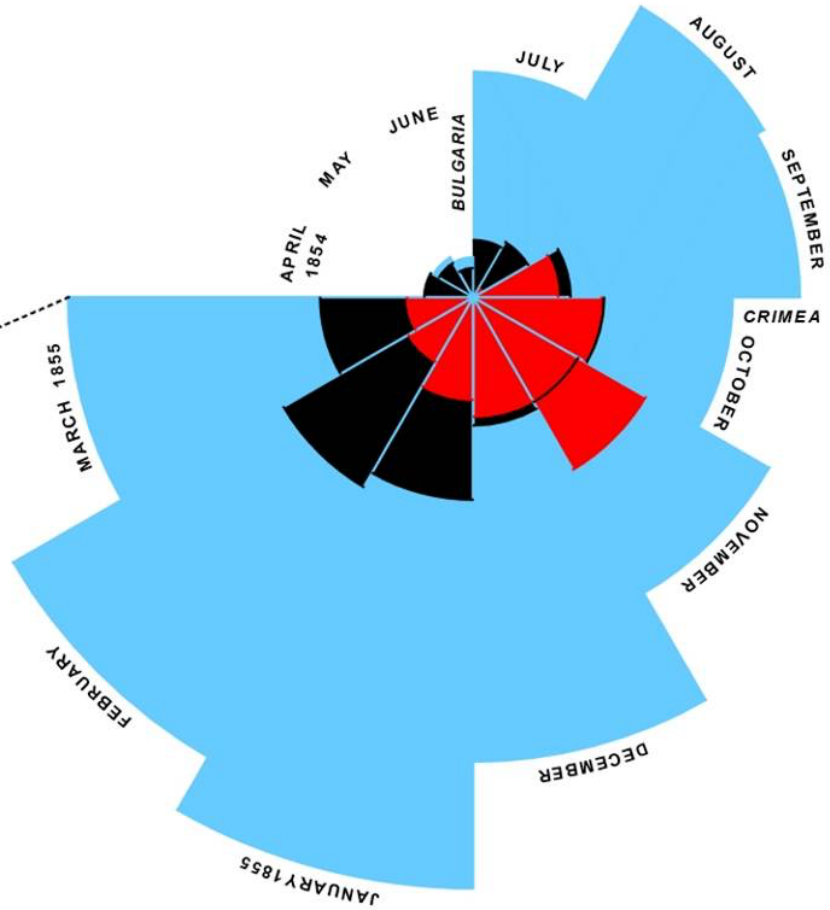
- Italian-born English nurse
- Use of statistics to advocate for sanitary reforms
  - Experience in Crimean War and India
- Applied and invented charts
- Famous in visualisation for polar area charts (also called ”Coxcomb” or “wedges”)

DIAGRAM of the CAUSES of MORTALITY
IN THE ARMY IN THE EAST.

2.
APRIL 1855 TO MARCH 1856.



1.
APRIL 1854 TO MARCH 1855.



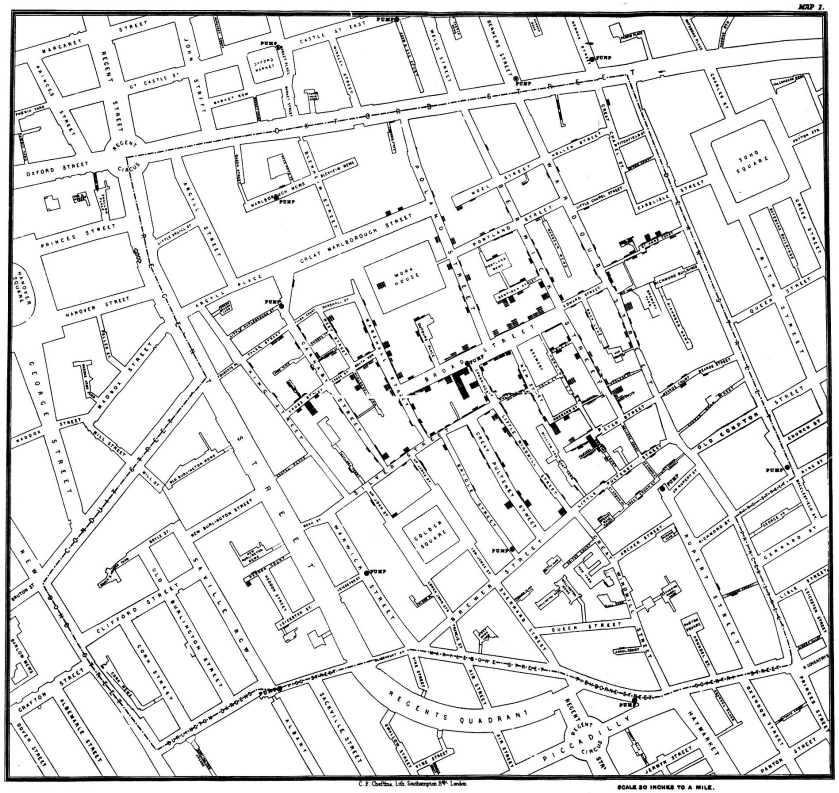
The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic Diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes
The black line across the red triangle in Nov^r 1854 marks the boundary of the deaths from all other causes during the month
In October 1854, & April 1855, the black area coincides with the red, in January & February 1856, the blue coincides with the black
The entire areas may be compared by following the blue, the red & the black lines enclosing them

John Snow (1813-1858)

- English physician
- Role in trying to dsiprove “miasma theory” of disease spread
- Used maps to communicate findings
  - Locating a pump that was the source of cholera

https://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak


50 0 50 100 150 200 Yards
× Pump • Deaths from cholera
OXFORD STREET
REGENT STREET
GT. MARLBOROUGH ST.
BROAD STREET
WARDOUR STREET
DEAN STREET
KING STREET
CONDUIT STREET
SAVILLE ROW
GOLDEN SQUARE
BREWER STREET
NEW BOND STREET
REGENTS QUADRANT
PICCADILLY


MAP 2.
OXFORD STREET
REGENT STREET
GREAT MARLBOROUGH STREET
BROAD STREET
GOLDEN SQUARE
REGENTS QUADRANT
PICCADILLY
SCALE 20 INCHES TO A MILE.

Geographical Information Science (GIS)

- Geographical data storage, visualisation and analysis
  - Facility managements
  - Automated cartography
  - Layers and geographical linking
  - Data representation
  - Spatial analysis

Information Visualisation

- From computer graphics and human-computer interaction
- Emphasis in novelty in visual encoding and interaction
- More recently, greater focus on vision science

Visual Analytics

- Incorporation of computational analysis methods
  - particularly for large data
- Emphasis on reasoning and sensemaking in an application domain
  - Using interactive visualisation tools

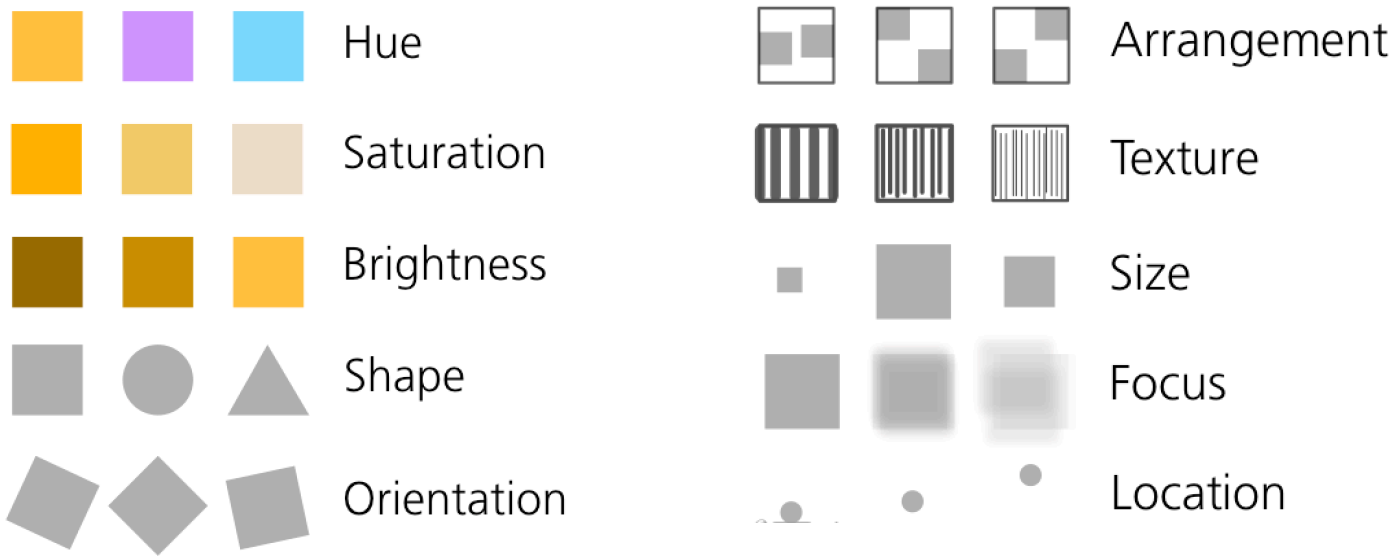
Why visual?

Why visualize things?

- Offload cognitively heavy tasks to sensorial system
- Use vision for tasks like comparing values and detect outliers
- Use high-level cognitive reasoning to understand and interpret
- See:
  - Patterson, R. E., et al. (2014). A human cognition framework for information visualization. *Computers and Graphics (Pergamon)*, *42*(1), 42–58. https://doi.org/10.1016/j.cag.2014.03.002

Visual variables

- How we place and modify *marks* on a 2D (or 3D) space
  - Bertin, 1983
- Basic visual properties



Visual variables

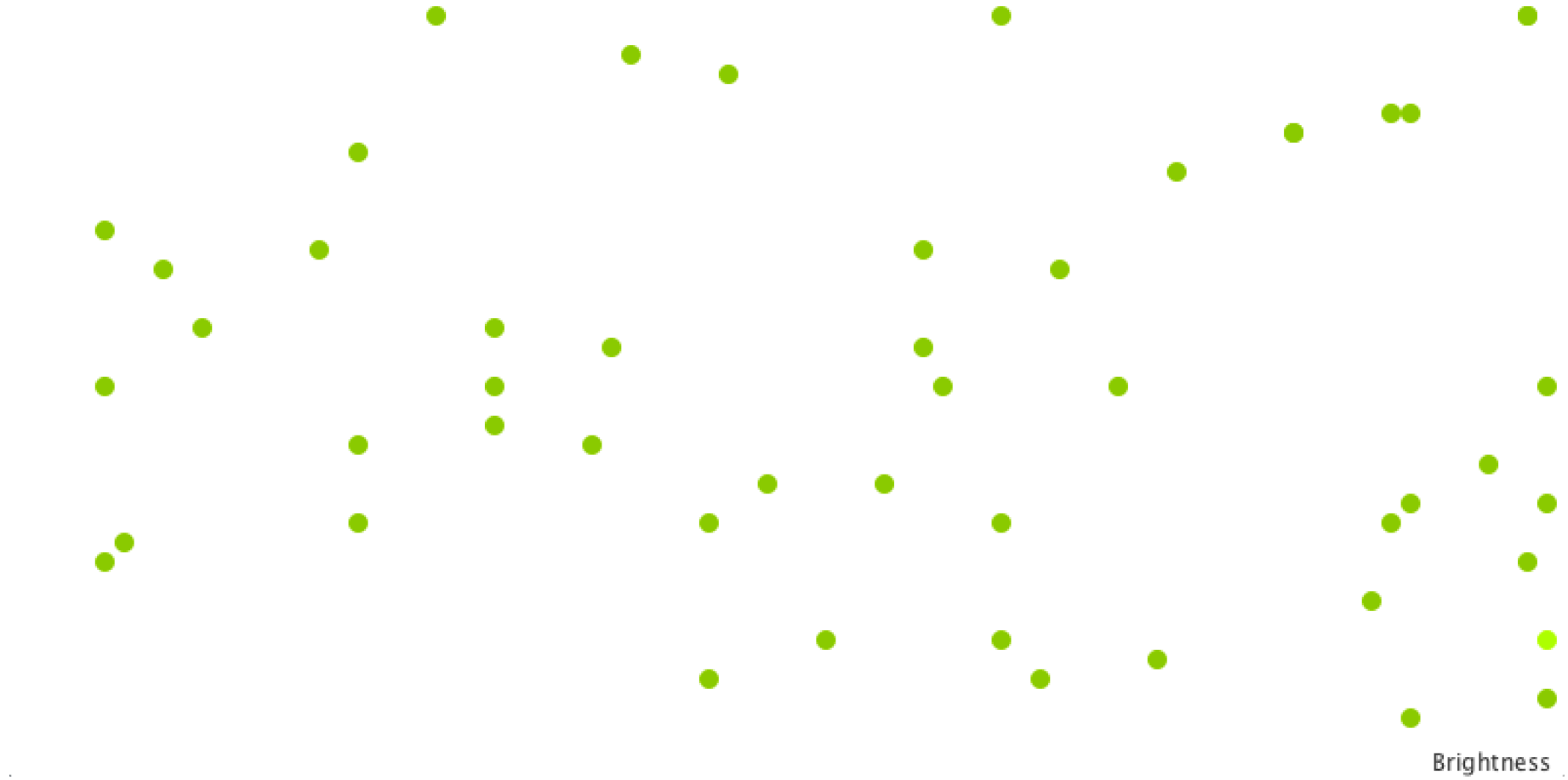
- Why do they “work”?
  - Preattentiveness
- Our vision can process information without conscious effort
- Non-preattentive features require us to *think*
  - Example: combining certain pre-attentive variables result in non-preattentive charts

Length

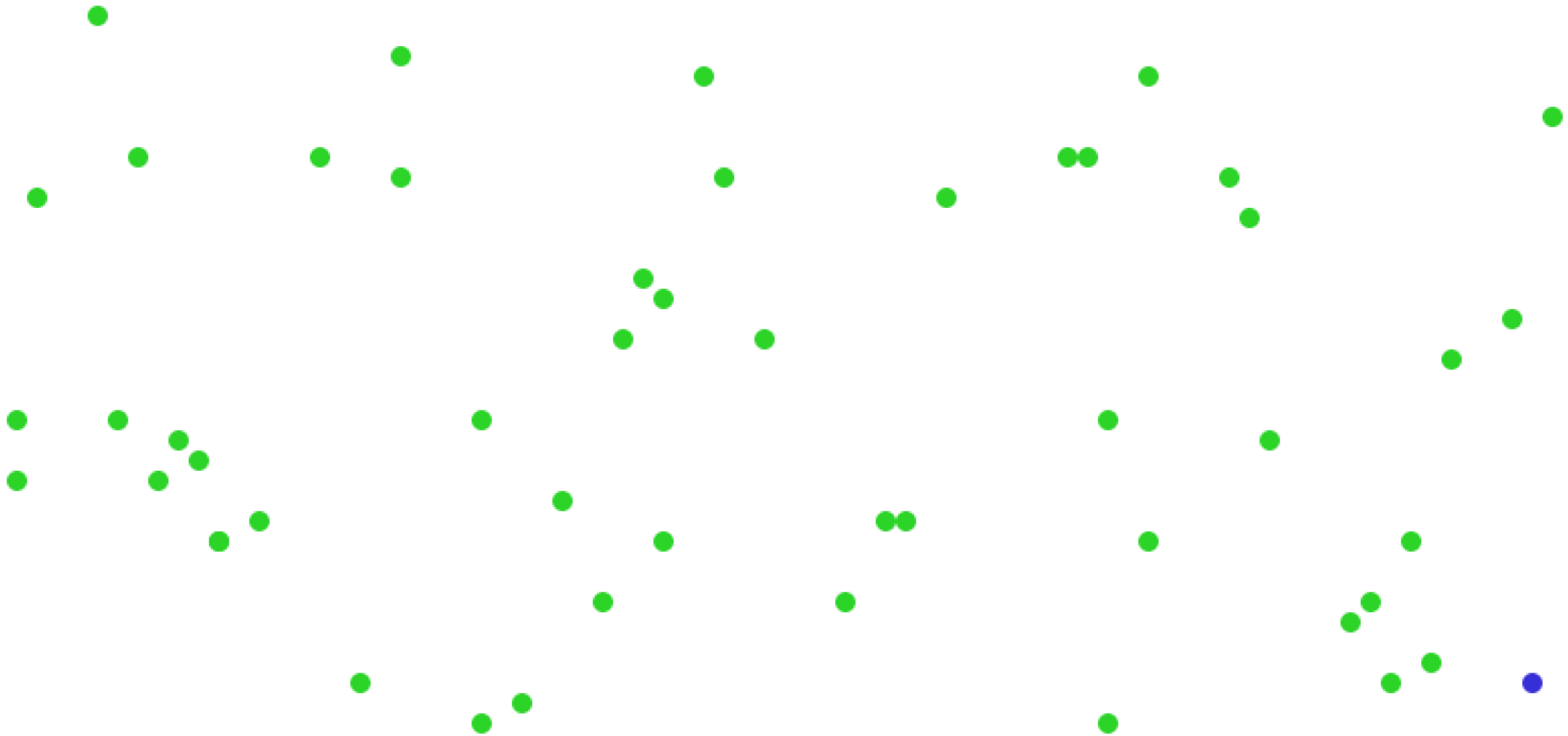
Size

Orientation

Brightness



Hue

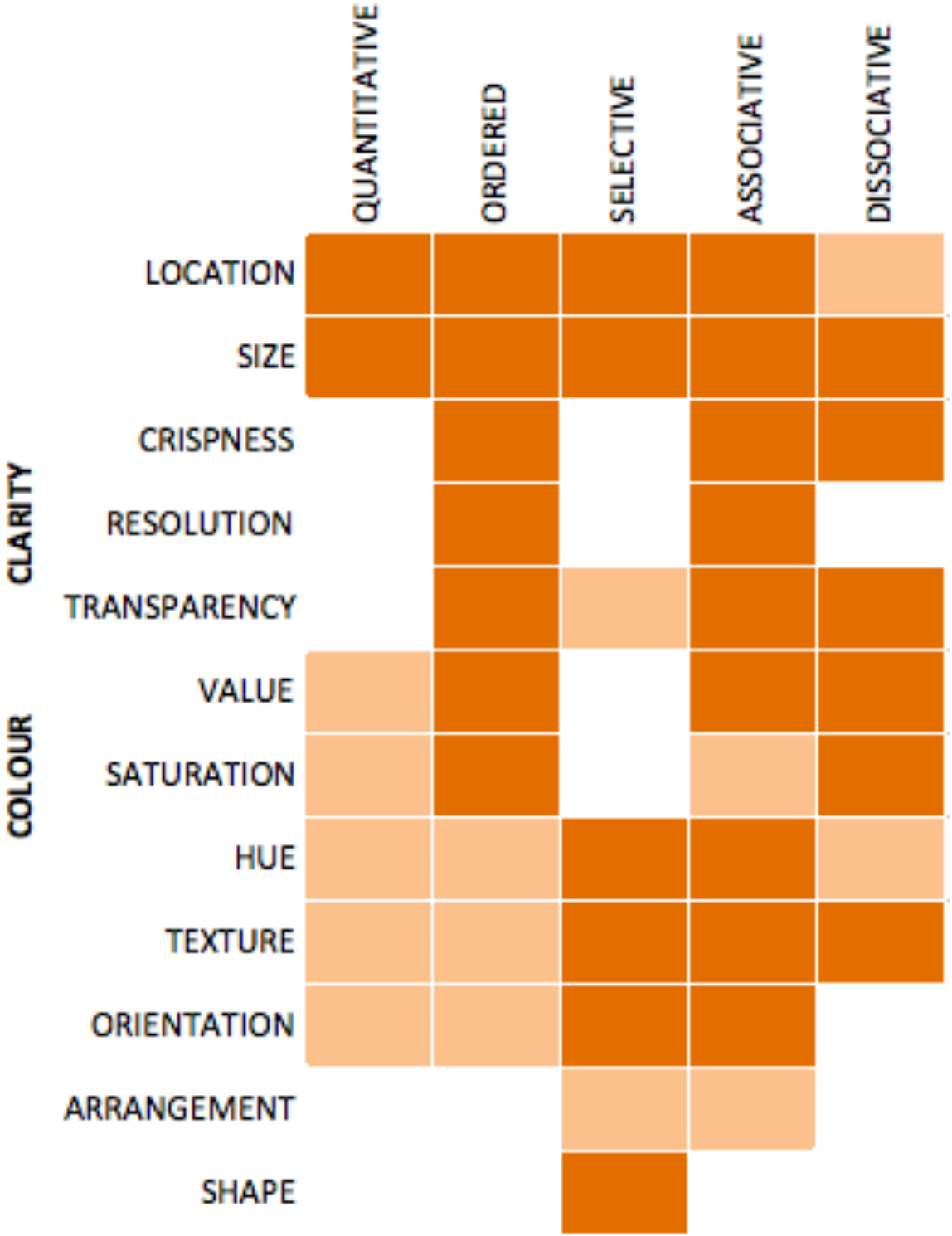


Hue

Visual variables not created equal

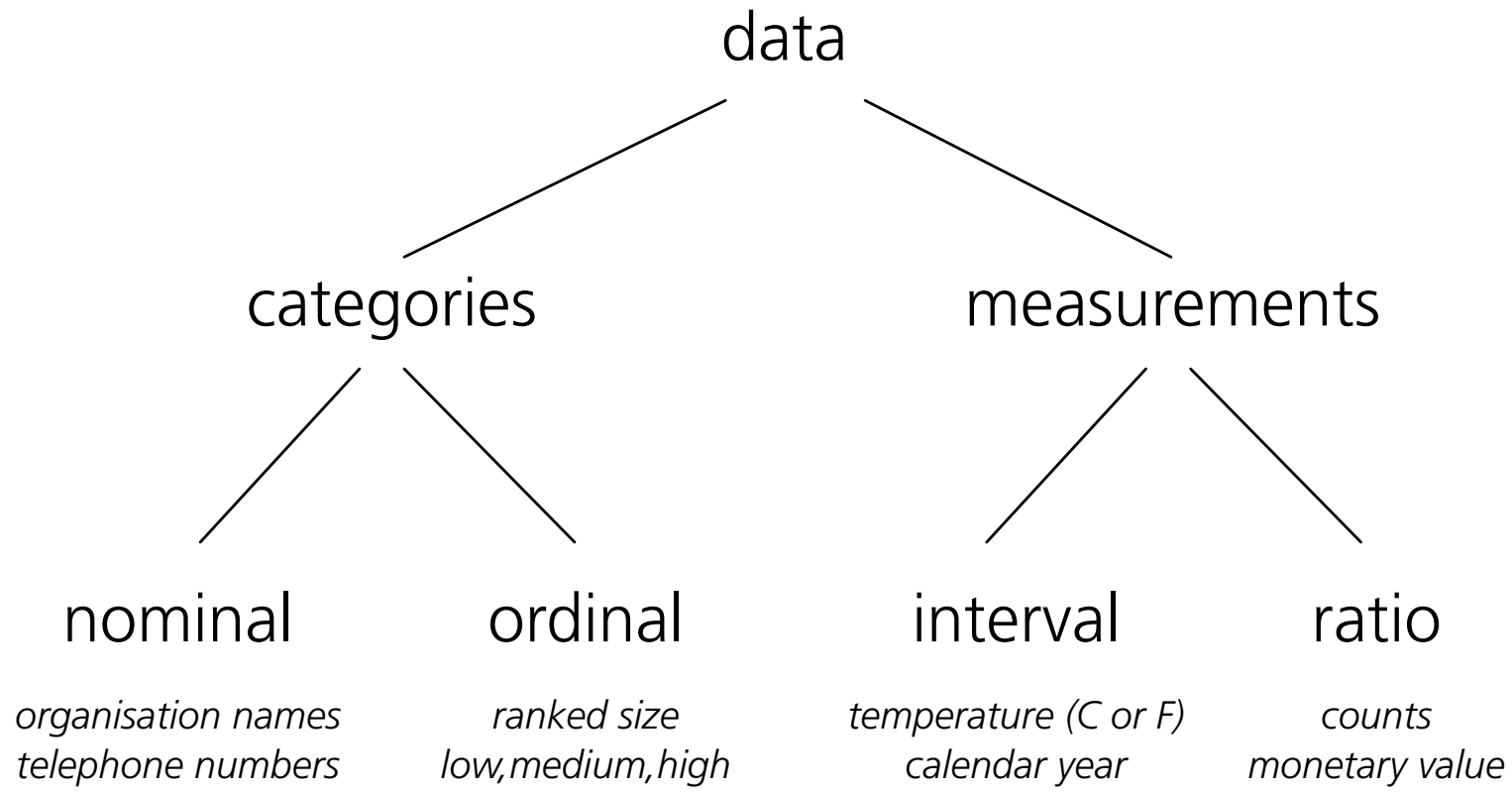
- Quantitative
  - Allowing **quantity** to be estimated
- Ordered
  - Allowing **order** to be determined
- Selective
  - Allowing **particular things** to be identified
- Associative
  - Allowing **groups of things** to be identified

Bertin, 1983



MacEachren, 1995

Depends on your data



Use theory to inform design choices

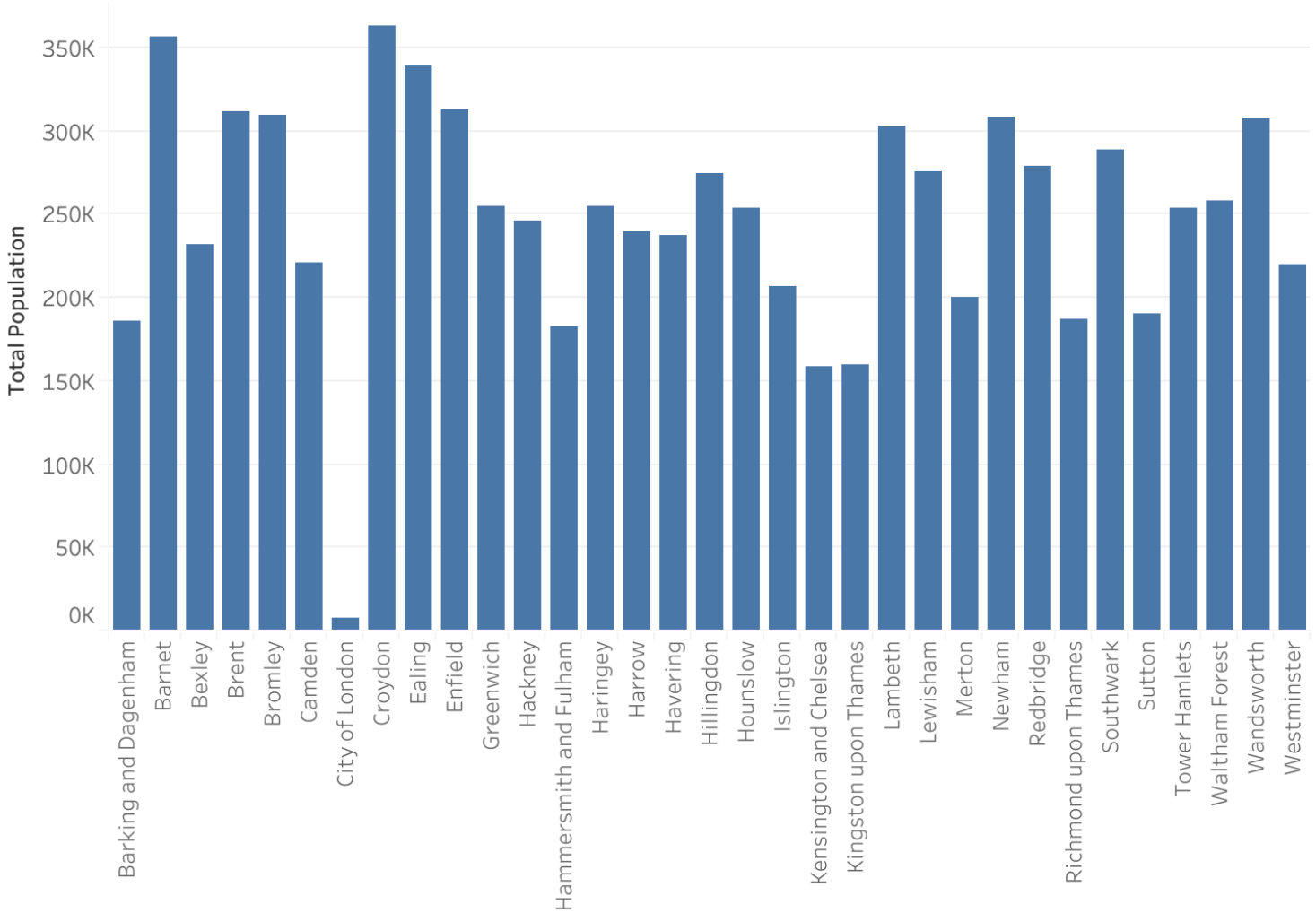
- For example:
  - Position or (aligned) length is the strongest
    - implies quantity/order, so take care with categories
  - Hue and shape good for unordered categories (<8ish)
  - Lightness has a much lower resolution than position
    - can be good for ordered categories.
  - Maps use position to show geography
    - but if geography isn't the main point, are you "wasting" this important visual variable?

Types of charts

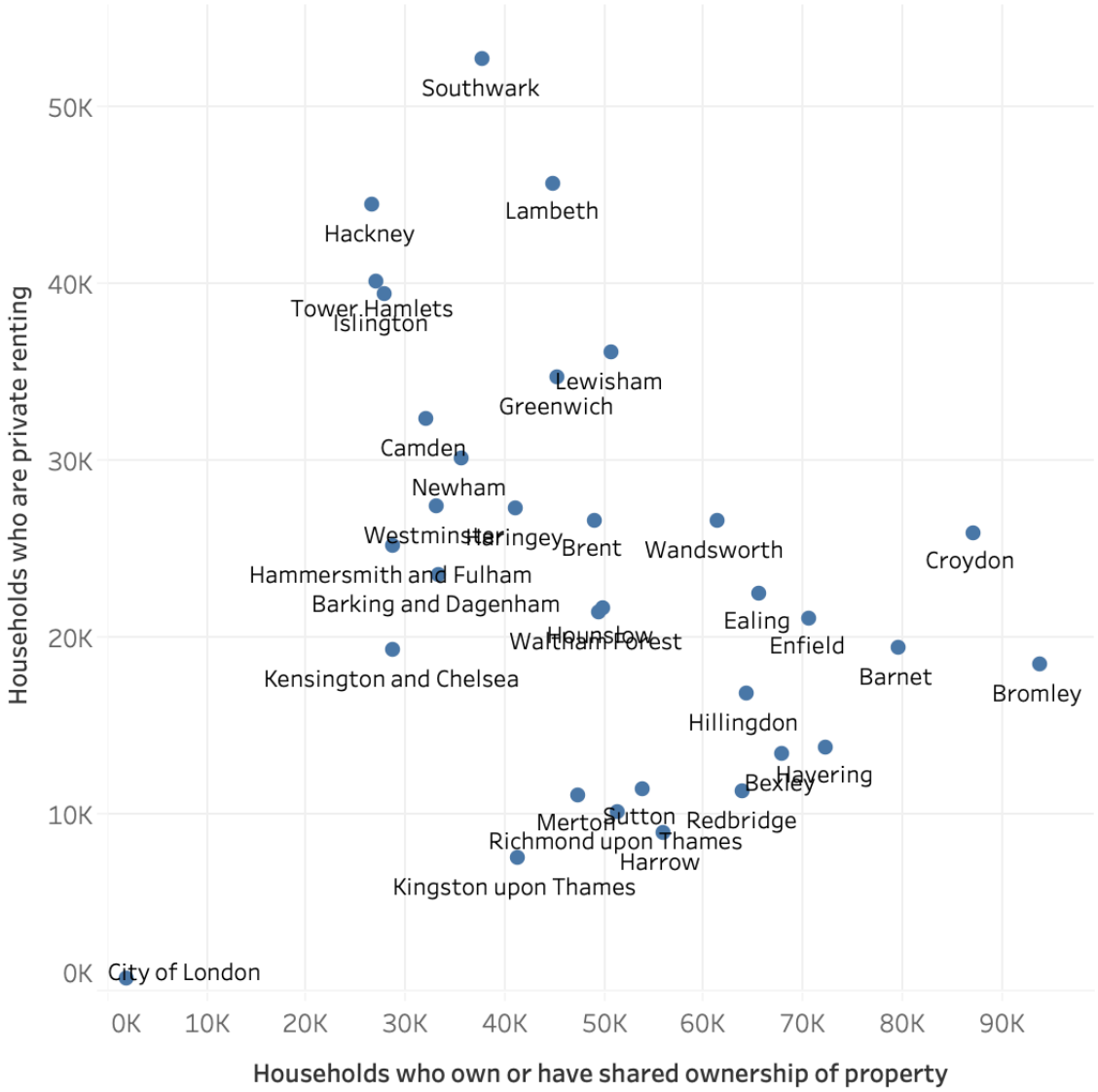
Some common types of charts

- Data → visual variables
- Bar chart:
  - Reference: X position of bar
  - Numeric attribute: Bar length
- Other important visual attributes
  - Bar width, axes, gridlines, etc.

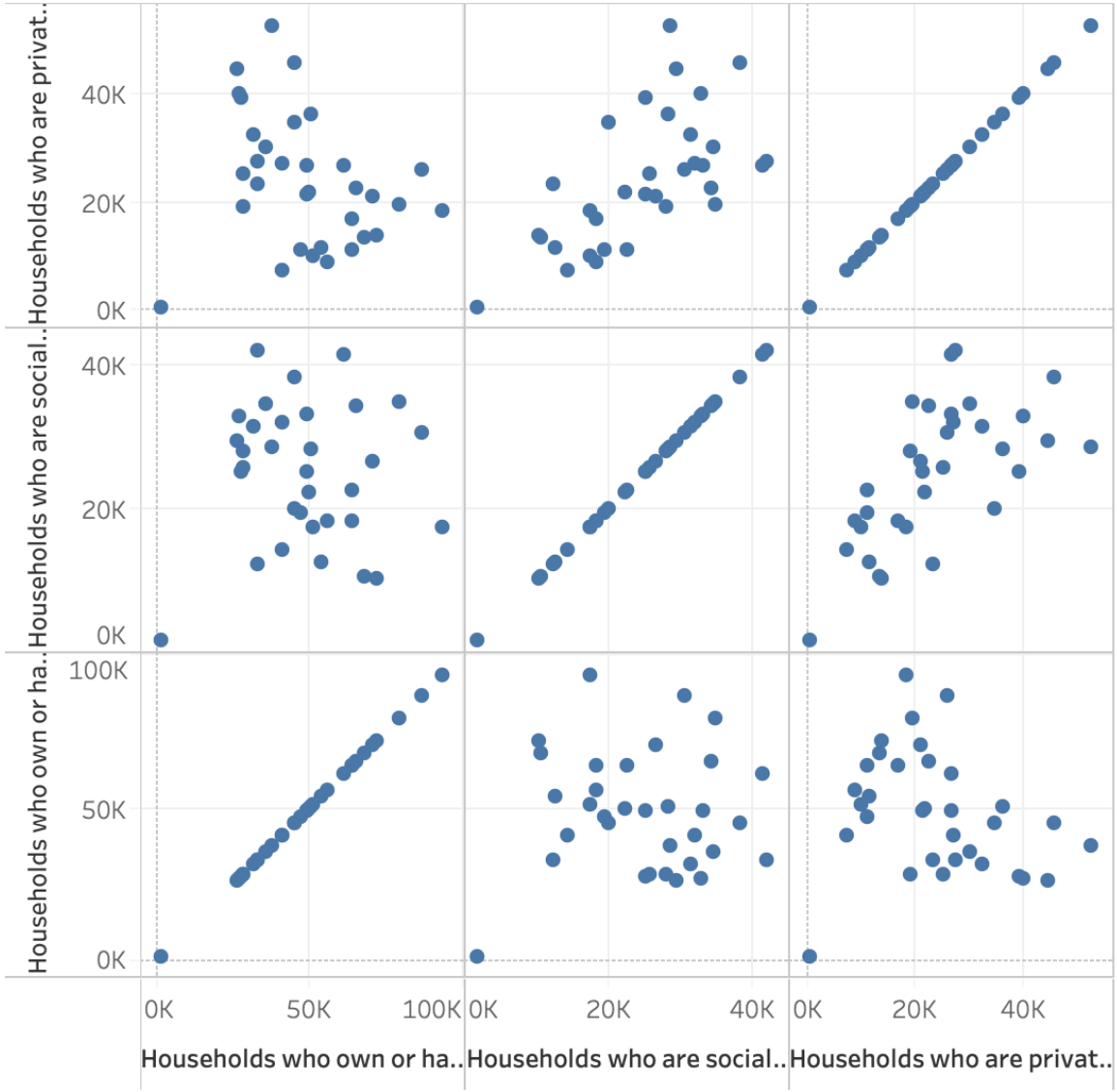
Bar chart


Total Population
350K
300K
250K
200K
150K
100K
50K
0K
Barking and Dagenham
Barnet
Bexley
Brent
Bromley
Camden
City of London
Croydon
Ealing
Enfield
Greenwich
Hackney
Hammersmith and Fulham
Haringey
Harrow
Havering
Hillingdon
Hounslow
Islington
Kensington and Chelsea
Kingston upon Thames
Lambeth
Lewisham
Merton
Newham
Redbridge
Richmond upon Thames
Southwark
Sutton
Tower Hamlets
Waltham Forest
Wandsworth
Westminster

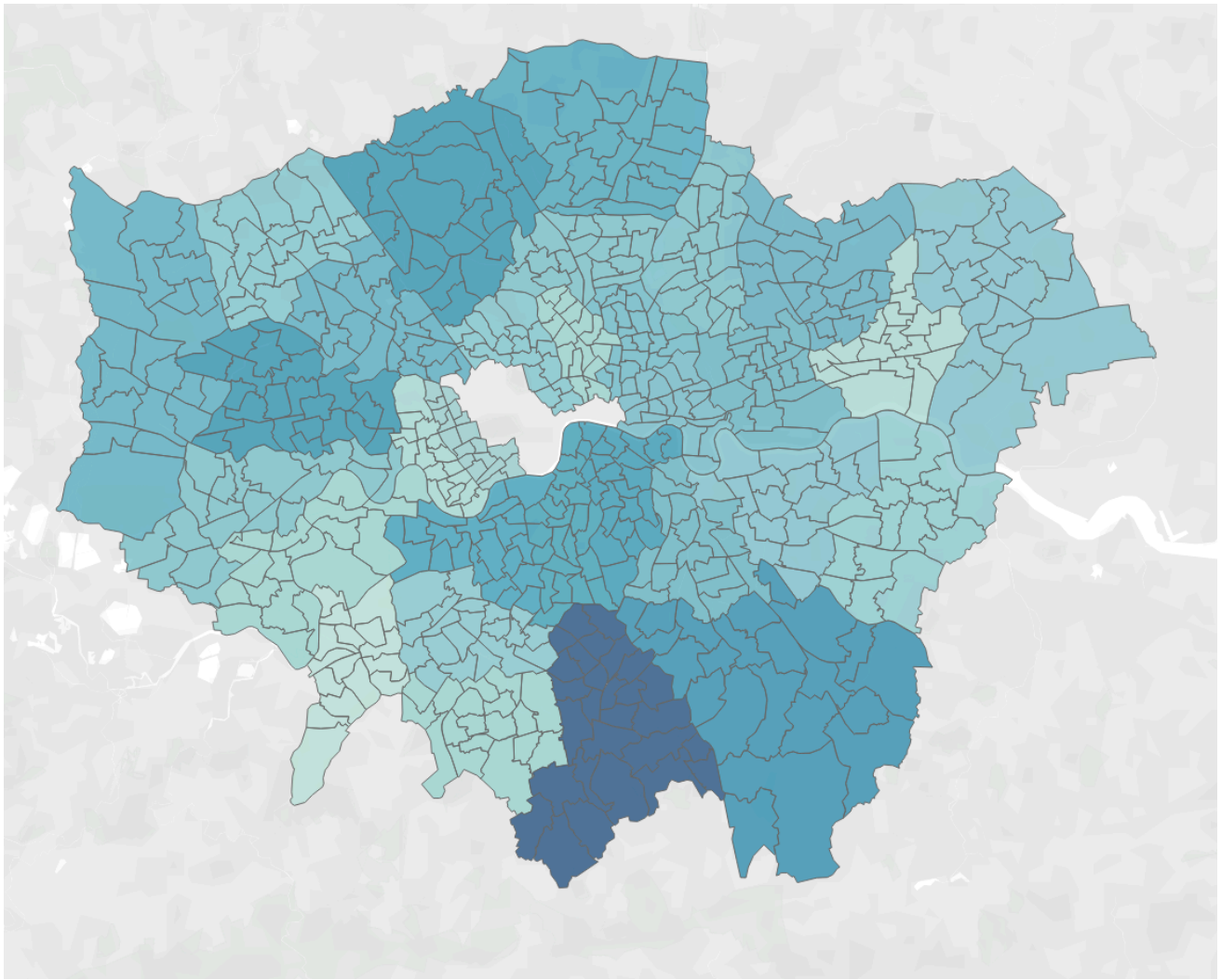
Scatterplot



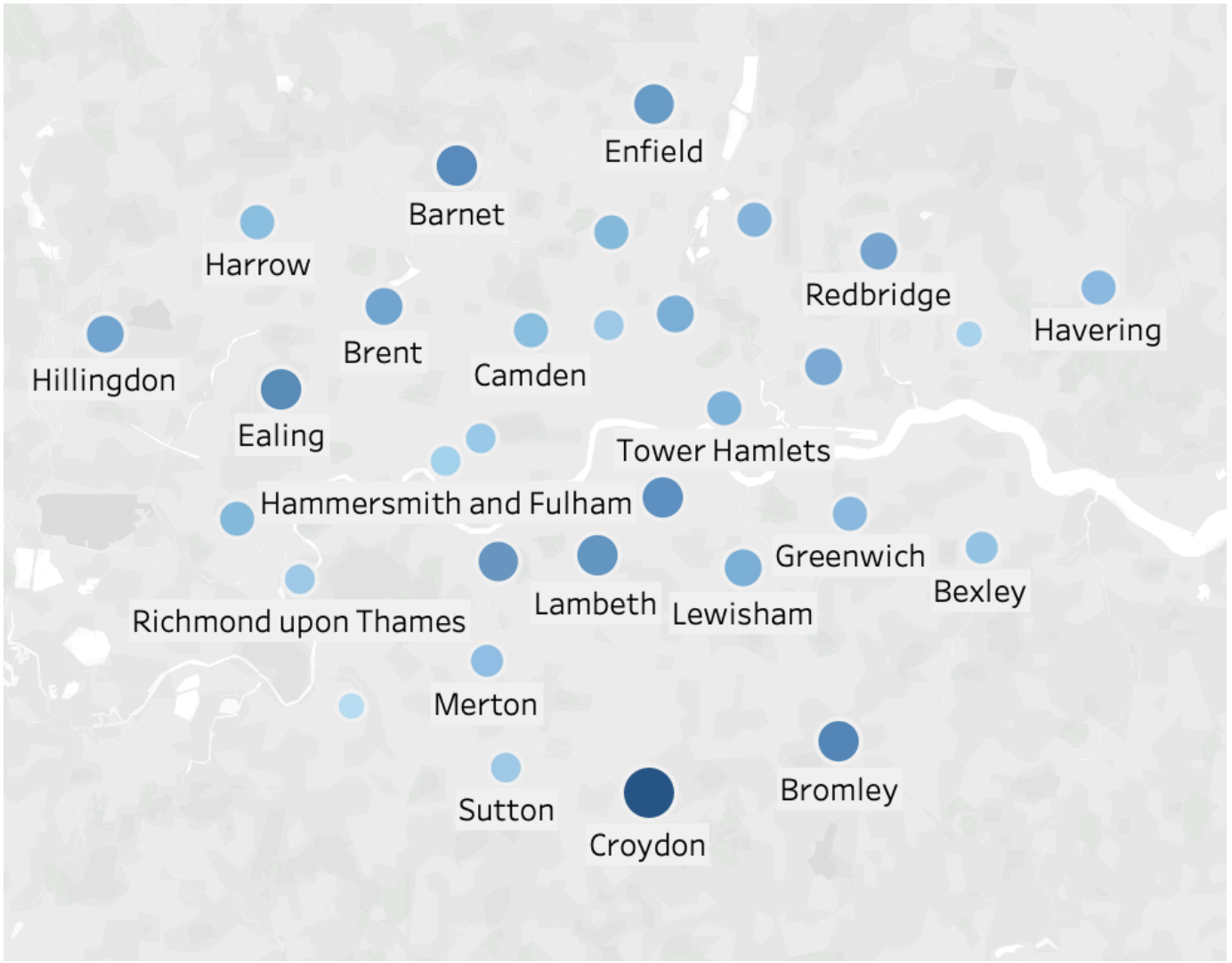
Scatterplot matrix



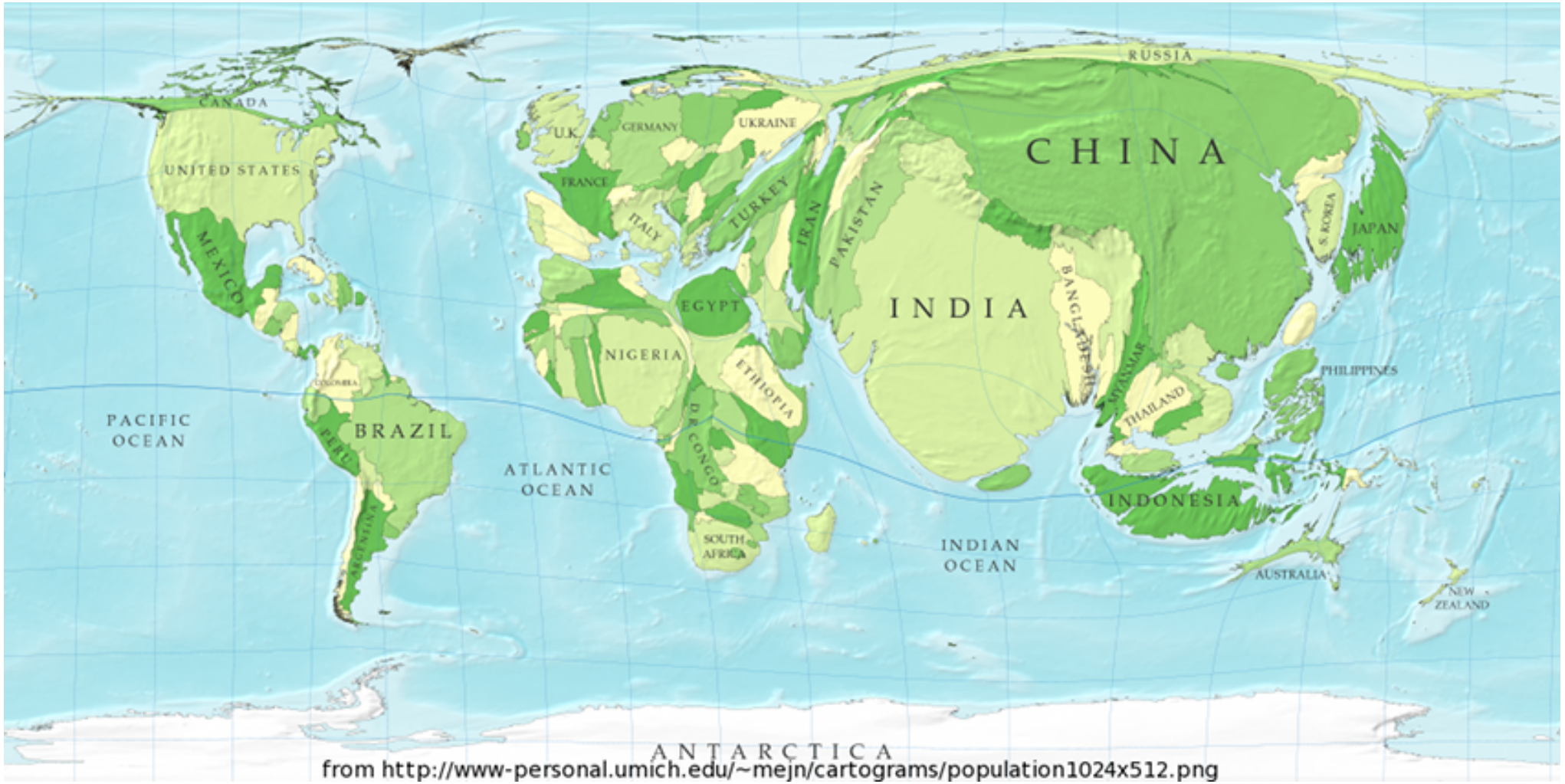
Choropleth map



Proportional symbol map



Area cartograms

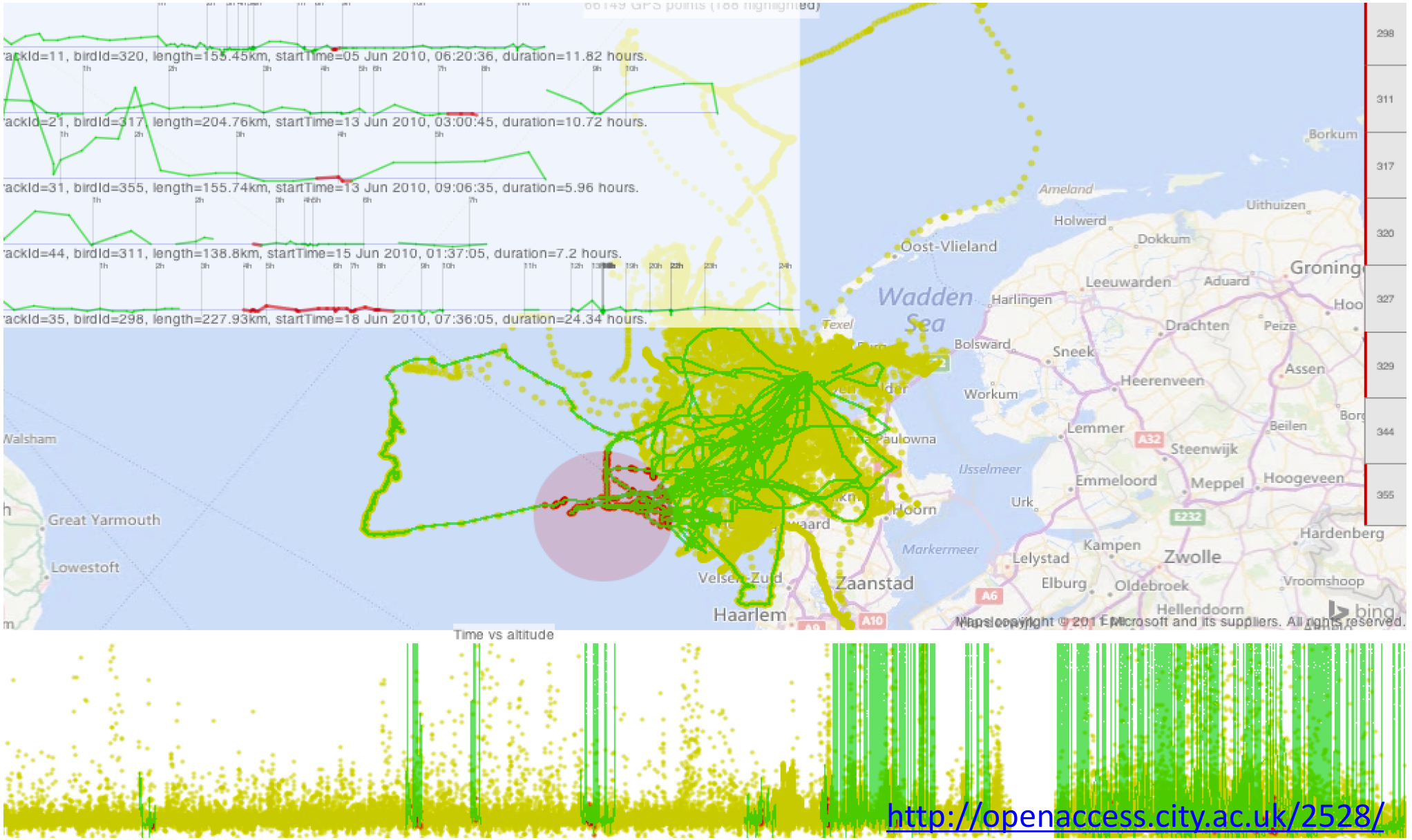


from http://www-personal.umich.edu/~mejn/cartograms/population1024x512.png

A tour through the visualisation zoo

- Jeffrey Heer, Michael Bostock, and Vadim Ogievetsky
  - http://queue.acm.org/detail.cfm?id=1805128

Example: GPS tracking of birds



Software

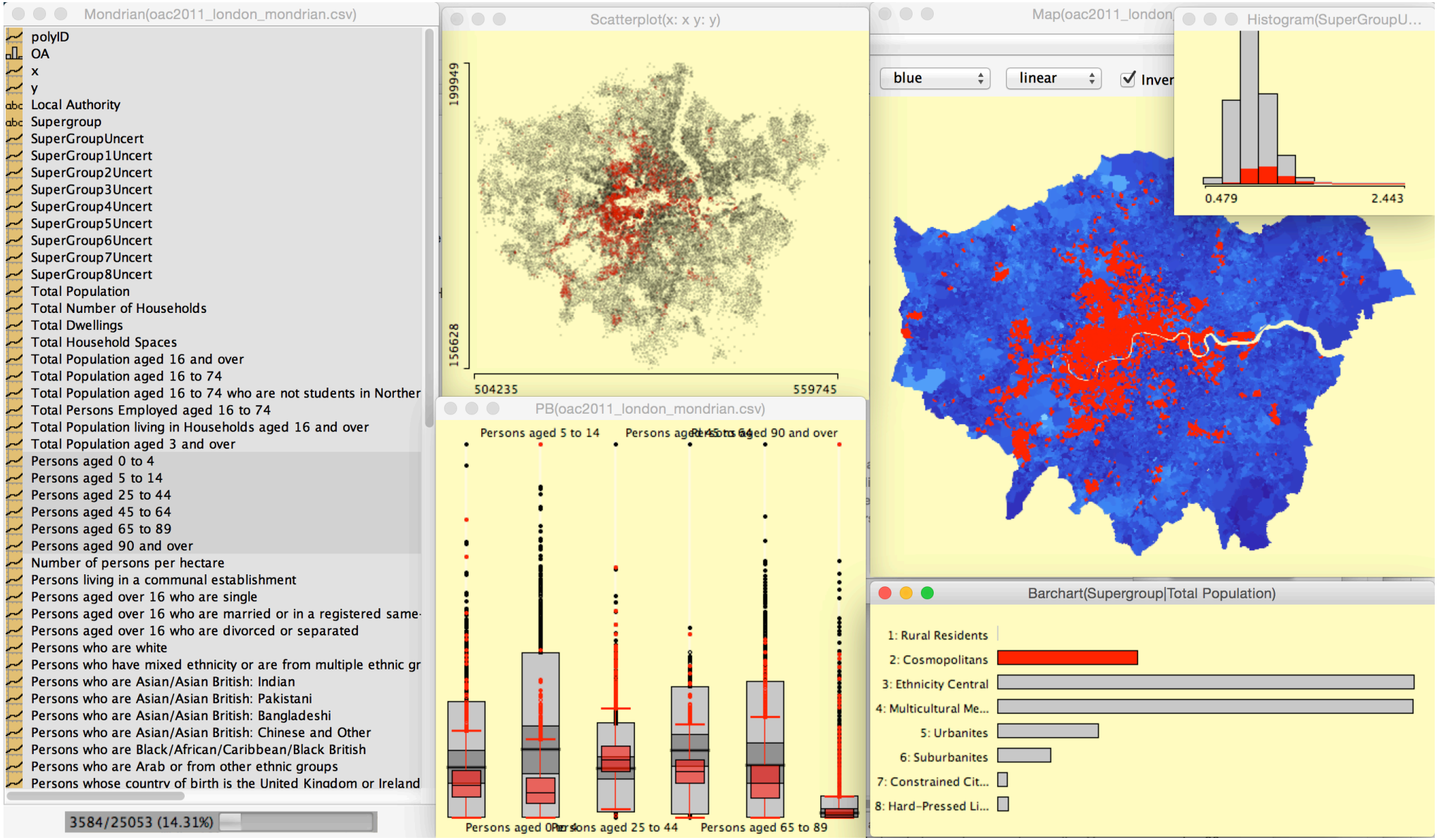
Software

- Bad news
  - No silver bullet “visual analytics” tool
- Good news
  - Orange: workflows + Python
  - Mondrian: interactive exploration
  - Tableau: visualisations/dashboards
  - Notebooks (Python, R, etc.)
  - Others: GIS tools (QGIS, ArcGIS)

Mondrian

Mondrian

- Interactive visual exploration software
  - Written by Martin Theus
  - http://www.theusrus.de/Mondrian/
  - Very fast, simple and effective
  - Reads CSV/TSV files


Mondrian(oac2011_london_mondrian.csv)
polyID
OA
x
y
Local Authority
Supergroup
SuperGroupUncert
SuperGroup1Uncert
SuperGroup2Uncert
SuperGroup3Uncert
SuperGroup4Uncert
SuperGroup5Uncert
SuperGroup6Uncert
SuperGroup7Uncert
SuperGroup8Uncert
Total Population
Total Number of Households
Total Dwellings
Total Household Spaces
Total Population aged 16 and over
Total Population aged 16 to 74
Total Population aged 16 to 74 who are not students in Norther
Total Persons Employed aged 16 to 74
Total Population living in Households aged 16 and over
Total Population aged 3 and over
Persons aged 0 to 4
Persons aged 5 to 14
Persons aged 25 to 44
Persons aged 45 to 64
Persons aged 65 to 89
Persons aged 90 and over
Number of persons per hectare
Persons living in a communal establishment
Persons aged over 16 who are single
Persons aged over 16 who are married or in a registered same
Persons aged over 16 who are divorced or separated
Persons who are white
Persons who have mixed ethnicity or are from multiple ethnic gr
Persons who are Asian/Asian British: Indian
Persons who are Asian/Asian British: Pakistani
Persons who are Asian/Asian British: Bangladeshi
Persons who are Asian/Asian British: Chinese and Other
Persons who are Black/African/Caribbean/Black British
Persons who are Arab or from other ethnic groups
Persons whose country of birth is the United Kingdom or Ireland
3584/25053 (14.31%)
Scatterplot(x: x y: y)
199949
156628
504235
559745
PB(oac2011_london_mondrian.csv)
Persons aged 5 to 14
Persons aged 90 and over
Persons aged 25 to 44
Persons aged 65 to 89
Map(oac2011_london
blue
linear
Inver
Histogram(SuperGroupU...
0.479
2.443
Barchart(Supergroup|Total Population)
1: Rural Residents
2: Cosmopolitans
3: Ethnicity Central
4: Multicultural Me...
5: Urbanites
6: Suburbanites
7: Constrained Cit...
8: Hard-Pressed Li...

Mondrian

- Simple individual charts… but coordinated!
  - Barchart (& weighted)
  - Histogram (& weighted)
  - Scatterplot
  - Mosaic plot
  - Parallel coordinates
  - Scatter plot matrix (SPLOM)
  - Map

Mondrian: Data

- Standard single CSV or TSV file
  - Comma/tab-separated values
    - Rows = records; columns = attributes
  - Plus an additional `.map` file for choropleth maps
    - Polygons need to be specified in a particular format. Very fiddly. See example datasets

Mondrian: operation

- List of attributes
  - Icon identifies type –click to change
- Select one or more (SHIFT/CTRL for multiple)
  - Can weight by something (e.g. population)
- Use the plot menu to plot simple charts
  - Coordinated: selecting items will select corresponding items
- Calculate: min/max; can also connect to R
- Selection types: see options menu

Tableau

Tableau

- Interactive data visualisation/creation tool focused on business intelligence
- Perhaps less suited for exploratory visual analysis

Tableau: data

- Tabular data
  - Many types: CSV, database connection, Excel, etc
- Concept of
  - Dimensions: categorical variables for pivoting
  - Measures: quantitative (numerical) variables for mapping to colour/size/etc
- May need to reshape data
  - http://kb.tableau.com/articles/knowledgebase/preparing-excel-files-analysis

Tableau: operation

- Column and row shelves:
  - Broadly, x-position and y-position
- Marks: the shapes that represent data
  - Map to “visual channels”: size/colour/label
- Tableau will aggregate data according to what’s in the shelves and marks
  - Drag the attribute for disaggregation to marks “detail”.

Wrap up

Conclusions

- Visual Analytics is:
  - **analytical reasoning** by combining **computational analysis** with **interactive visual interfaces**
- Visualisation has an important role:
  - Well-designed visualisation present data effectively and facilitate comparison
  - Well-design interactions act as an interface to data, particularly in multiple-linked views

Intended learning outcomes

- Know what Visual Analytics is
- Know the role of interactive visualisation in Visual Analytics
  - Visual variables and when to use
  - Types of visualisation display and when to use
  - Types of interaction
  - Coordinated linked views
- (Practical) how to use Mondrian and Tableau

Reading

- Exploratory Analysis of Spatial and Temporal Data A Systematic Approach.
  - *Chapter 4.3 Visualization in a Nutshell*
  - *Download whole book* http://0-dx.doi.org.wam.city.ac.uk/10.1007/3-540-31190-4
- **Munzner, T.** Visualization Analysis & Design, CRC Press
  - Chapters 2 and 5.