

# Data visualisation for humanities researchers

CHASE Going Digital programme, June 2015

Mia Ridge, Open University  
@mia\_out <http://miaridge.com>

# While we're getting started...

Check you can log on, use the mouse and go online to access further reading:

<http://bit.ly/OUInfoVis>

Check you can use the browsers Firefox or Chrome

Unzip ('extract') the file containing the slides and exercise handouts and copy the folder to your desktop

# Timetable

10.00-10.30 Registration and coffee

10.30-12.30 Workshop part 1

12.30-13.30 Lunch

13.30-15.30 Workshop part 2

15.30-15.45 Coffee

15.45-16.15 Conclusion, Q&A

# Overview

Introductions; what is data visualisation?

History and types of visualisations

Critiquing visualisations

Visualisations for scholarly analysis

Dealing with heritage and humanities data

Planning and designing visualisations



Data visualisation is the graphical display of quantitative or qualitative information to create insights by highlighting patterns, trends, variations and anomalies.

# From this...

Discipline	Description	Description so	Image_url	Image_source	Link	Year	Birth_place	Place_of_birth	Combined_birf	Country	Places_lived	Spouse	Death_year	Death_place	Place_of_burial
Polymath	Hildegard of B	Wikipedia	https://upload	Wikipedia	http://en.wikipedia.org/wiki/	1098	Bernersheim	Germany	Bernersheim	Germany			1179	Bingen am Rhein	
Writer	Christine de P	Based on Wiki	https://upload	Wikipedia		1363	Venice	Italy	Venice, Italy	France	Paris		1434	Poissy	
Writer	Bethusa	Wikipedia	https://upload	Wikipedia	http://herstor	1600				England			1675		
Artist	Anna Maria va	Wikipedia	https://upload	Wikipedia	http://en.wikipedia.org/wiki/	1607	Cologne		Cologne,	Germany	Utrecht, Amsterdam		1678	Wuwert	
Writer	Her lengthy n	Wikipedia	https://upload	Wikipedia	https://en.wiki	1607	Le Havre	France	Le Havre, Fra	France	Le Havre		1701	Paris	
Intellectual	Katherine Jon	Wikipedia and	COINB		http://en.wiki	1615	Ireland	St Martin-in-tr	Ireland, St Martin-in-the-Fields	London			1691	London	
Writer	Hannah Woolf	Wikipedia	https://upload	Wikipedia	https://en.wiki	1622				England	Newport, Esse		1675		
Scientist	Margaret, Cav	Wikipedia	https://upload	Wikipedia	http://en.wiki	1623				England	William Caven		1673		
Intellectual	Anne-Thérèse	Wikipedia	https://upload	Wikipedia	https://en.wiki	1647	Paris		Paris,	France	Henri de Lamé		1733	Paris	
Astronomer	Maria Margare	Wikipedia	https://upload	Wikipedia	http://en.wiki	1670	Leipzig	Germany	Leipzig, Germ	Germany	Danzig; King	Godfried Kirn	1720	Berlin	
Private	Mary Read wa	Wikipedia and	https://upload	Wikipedia	http://en.wiki	1685	Leipzig	England	Leipzig, Eng	England	London		1721	Jamaica	
Writer	Mary Delany	Wikipedia	https://upload	Wikipedia	http://www.br	1700	Coustant	United Kingd	Coustant, Unit	England	Roscrow Cast	Alexander Pen	1788		St James's Ch
Artist	Madeline Basse	pointe was a	https://women	National Muse	http://women	1701	Paris	France	Paris, France	France			1780	Paris	
Private	Anne Bonny	Wikipedia	https://upload	Wikipedia	http://en.wiki	1702	County Cork	Ireland	County Cork, I	Ireland	Bahamas; Nas	James Bonny	1782		
Mathematician	Gabrielle Emil	Wikipedia	https://upload	Wikipedia	http://en.wiki	1706	Paris	France	Paris, France	France	Marquis Flor		1749	Lunéville	
Scientist	Laura Maria C	Wikipedia	https://upload	Wikipedia	http://en.wiki	1711	Bologna	Italy	Bologna, Italy	Italy	Bologna		1778	Bologna	
Intellectual	Margaret Bent	Wikipedia	https://upload	Wikipedia	http://en.wiki	1715	Welbeck Abbe	United Kingd	Welbeck Abbe	United Kingdom		William Bentin	1785	Bulstrode Park, Buckingham	
Physician	Dorothea Chr	Wikipedia	https://upload	Wikipedia	http://en.wiki	1717	Quedlinburg	Germany	Quedlinburg,	Germany	1762	Quedlinburg			
Mathematician	Elizabeth Carl	Wikipedia	https://upload	Wikipedia	http://en.wiki	1717	Deal, Kent	England	Deal, Kent,	England	1806				
Writer	Maria Gastard	Wikipedia	https://upload	Wikipedia	http://en.wiki	1718	Milan	Italy	Milan	Italy	1799				
Intellectual	Elizabeth Mon	Wikipedia	https://upload	Wikipedia	http://en.wiki	1718	Yorkshire	England	Yorkshire, Eng	England	Henry Scott, 3		1800	London	
Composer	Francois Evelyn	Wikipedia	https://upload	Wikipedia	http://en.wiki	1719	Kent	England	Kent, England	England	Edward Bosca		1805	London	
Astronomer	Maria Teresa F	Wikipedia	https://upload	Wikipedia	http://en.wiki	1720	Milan	Italy	Milan	Italy	1795				
Private	Nicole-Reine L	Wikipedia	https://upload	Wikipedia	http://en.wiki	1723	Luxembourg P	France	Luxembourg P	France		Jean-André Le	1788		
ing Writer	Louise Florenc	Wikipedia	https://upload	Wikipedia	http://en.wiki	1726	Valenciennes	France	Valenciennes,	France	Geneva		1783	Paris	
Historian	Catharine Mac	Wikipedia	https://upload	Wikipedia	http://en.wiki	1731	Wye	United Kingd	Wye, United K	England	Dr. George Ma		1791	Binfield	
Writer	Suzanne Curd	Wikipedia	https://upload	Wikipedia	http://en.wiki	1737	Craissler	Switzerland	Craissler, Swit	Switzerland	Wye	Jacques Neck	1794	Lausanne	
uff Painter	Maria Anna Ar	Wikipedia	https://upload	Wikipedia	http://en.wiki	1741	Gräubünden	Switzerland	Gräubünden,	Austria	Chur	Antonio Zuch	1807	Rome	Sant'Andrea d
Post	Anna Seward	Wikipedia	https://upload	Wikipedia	http://en.wiki	1742	Eyam	England	Eyam, England	England	Lichfield		1809	Lichfield	Lichfield Cath
Political	Princess Yekat	Wikipedia	https://upload	Wikipedia	http://en.wiki	1744	Saint Petersburg	Russia	Saint Petersburg	Russia	Moscow	Prince Mikhail	1810	Moscow	
er-Painter	Anne Vallayer	Wikipedia	https://upload	Wikipedia	http://en.wiki	1744	France	France	France	France	Paris		1818	Paris	
er-Writer	Hannah More	Wikipedia	https://upload	Wikipedia	http://en.wiki	1745	Fishponds, Bristol	England	Fishponds, Bri	England	Bristol		1833	Clifton, Bristol	Church of All
y Writer	Josefa Amar y	Wikipedia	https://upload	Wikipedia	http://en.wiki	1749	Zaragoza	Spain	Zaragoza, Spa	Spain	Joaquín Fuert		1833		
ill Artist	Adelaide Labil	Wikipedia	https://upload	Wikipedia	http://en.wiki	1749	Paris	France	Paris, France	France	Paris	Louis-Nicolas	1803	Paris	
ey Writer							King's Lynn	Norfolk	King's Lynn, N	England			1840	London	
Artist	Louise Elisabe	Wikipedia	https://upload	Wikipedia	http://en.wiki	1755	Paris	France	Paris, France	France	Epemont; Sain	Jean-Baptiste	1842	Paris	Cimetière de L
Private	Marie-Anne Pi	Wikipedia	https://upload	Wikipedia	http://en.wiki	1758	Montbrison	Loire	Montbrison, L	France	Paris	Antoine Lavois	1836	Paris	
Writer						1759				England			1843		Abney Park Cr
Writer	Mary Wollston	Wikipedia	https://upload	Wikipedia	http://en.wiki	1759	Spitalfields	London	Spitalfields, L	United Kingdom	Spitalfields	William Godwi	1797	London	St Pancras Ok
Artist	Marguerite G	Wikipedia	https://en.wik	Wikipedia	http://en.wiki	1761	Grasse	France	Grasse, France	France			1837	Paris	
Mathematician							Paris	France	Paris, France	France			1831	Paris	
Intellectual							Jedburgh	United Kingd	Jedburgh, Unit	Scotland			1872	Naples	
Painter			https://upload	wikimedia.org	wikipedia/commons/thumb/3/38/Hortense		Paris	France	Paris, France	France			1845	Paris	
							London	England	London, Engla	England		John Franklin	1875	London	
nie Painter,										France			1867		
ret Photographer	Augusta Ada K	Wikipedia	https://upload	Wikipedia	http://en.wiki	1815	Kolkata	India	Kolkata, India	United Kingdom	Kolkata		1879	Kalutara	
Scientist	María Mitchell	Wikipedia	https://upload	Wikipedia	http://en.wiki	1818	Nantucket	Massachusetts	Nantucket, Ma	United States	Nantucket	William King-3	1852	Marylebone	Hucknall
ell Astronomer						1820	Dorchester Co	England	Dorchester Co	United States	Maryland	Nelson Davies	1913	Auburn	Fort Hill Ceme
Writer							Bristol	England	Bristol, Englan	United States	Bristol		1910	Hastings	
Physician							Bristol	England	Bristol, Englan	United States	of America		1910	York Cliffs, Maine	
Physician							Berlin	Germany	Berlin, Germ	Poland			1902		
Physician							Oswego	New York	Oswego, New	United States of America			1919	Oswego	
Physician							Skillbreen	Ireland	Skillbreen, Ir	Republic of Ire	County Cork		1907	London	
Physician							London	England	London, Engla	United States of America			1906	New York City	
Physician							Liverpool	England	Liverpool, Eng	England	Liverpool	Walter Sheppe	1934	Christchurch	
Physician							Aldesburgh	United Kingd	Aldesburgh, Un	United Kingdom	Suffolk	Henry Fawcett	1929	London	
Physician	Annie Besant	Wikipedia	https://upload	Wikipedia	http://en.wiki	1847	Clapham	London	Clapham, Lon	England	Sibsey, Varan	Frank Besant	1933	Acyar	
Physician						1848	Dublin	Republic of Ire	Dublin, Repub	United Kingdom		William Huggi	1915		
Mathematician			https://upload	Wikipedia			Moscow	Soviet Union	Moscow, Sovi	Russian Empire	Moscow		1891	Stockholm	
Writer							Hobart	Tasmania	Hobart, Tasme	England	Hobart	Thomas Hume	1920	London	

# ...to this

Discipline	Description	Description so	Image_url	Image_source	Link	Year	Birth_year	Birth_place	Place_of_birth	Combined_birl	Country	Places_lived	Spouse	Death_year	Death_place	Place_of_burial
Polymath	Hildegard of B	Wikipedia	<a href="https://upload.wikimedia.org/wikipedia/commons/4/4c/Hildegard_of_Bingen.jpg">https://upload.wikimedia.org/wikipedia/commons/4/4c/Hildegard_of_Bingen.jpg</a>	Wikipedia	<a href="http://en.wikipedia.org/wiki/Hildegard_of_Bingen">http://en.wikipedia.org/wiki/Hildegard_of_Bingen</a>	1098	1098	Bernersheim	Germany	Bernersheim	Germany			1179	Bingen am Rhein	4
Writer	Christine de Pisan	Based on Wiki	<a href="https://upload.wikimedia.org/wikipedia/commons/5/5d/Christine_de_Pisan.jpg">https://upload.wikimedia.org/wikipedia/commons/5/5d/Christine_de_Pisan.jpg</a>	Wikipedia	<a href="http://en.wikipedia.org/wiki/Christine_de_Pisan">http://en.wikipedia.org/wiki/Christine_de_Pisan</a>	1363	1363	Venice	Italy	Venice, Italy	France	Paris		1434	Paris	4
Writer	Bethune	Wikipedia	<a href="https://upload.wikimedia.org/wikipedia/commons/1/1a/Bethune.jpg">https://upload.wikimedia.org/wikipedia/commons/1/1a/Bethune.jpg</a>	Wikipedia	<a href="http://en.wikipedia.org/wiki/Bethune">http://en.wikipedia.org/wiki/Bethune</a>	1600	1600	Cologne	England	Cologne, Germany	England	Utrecht, Amsterdam		1675	Wuivert	5
Artist	Anna Maria van Schrieck	Wikipedia	<a href="https://upload.wikimedia.org/wikipedia/commons/1/1a/Anna_Maria_van_Schrieck.jpg">https://upload.wikimedia.org/wikipedia/commons/1/1a/Anna_Maria_van_Schrieck.jpg</a>	Wikipedia	<a href="http://en.wikipedia.org/wiki/Anna_Maria_van_Schrieck">http://en.wikipedia.org/wiki/Anna_Maria_van_Schrieck</a>	1607	1607	Cologne	Germany	Cologne, Germany	Germany	Utrecht, Amsterdam		1678	Wuivert	5



Writer						1820	Dorchester Co	Maryland	Dorchester Co	United States	Maryland	Nelson Davies		1913	Auburn	Fort Hill Ceme	3
Physician							Bristol	England	Bristol, England	United States	Bristol			1910	Hastings		4
Physician							Bristol	England	Bristol, England	United States	Bristol			1910	York Cliffs, Maine		4
Physician							Berlin	Germany	Berlin, Germany	Poland	Berlin, Germany			1902			5
Surgeon							Oswego	New York	Oswego, New York	United States	Oswego, New York			1919	Oswego		5
Astronomer							Skibbereen	Ireland	Skibbereen, Ireland	Republic of Ireland	County Cork			1907	London		5
Physician							London	England	London, England	United States	London, England			1906	New York City		3
Physician							Liverpool	England	Liverpool, England	United States	Liverpool, England	Walter Sheppe		1934	Christchurch		5
Physician							Aideburgh	United Kingdom	Aideburgh, United Kingdom	United Kingdom	Suffolk	Henry Fawcett		1929	London		5
Physician							Clapham	London	Clapham, London	England	Sibsey, Varant	Frank Besant		1933	Adyar		5
Physician							Dublin	Republic of Ireland	Dublin, Republic of Ireland	United Kingdom	Sibsey, Varant	William Hugg		1915			5
Physician							Moscow	Soviet Union	Moscow, Soviet Union	Russian Empire	Moscow			1891	Stockholm		5
Physician							Hobart	Tasmania	Hobart, Tasmania	England	Hobart	Thomas Hume		1920	London		5

# About me



# Introductions

In a sentence or two, what's your interest in data visualisation?

- What kinds of data do you work with?
- What's the goal of any visualisations you're interested in creating?
- Do you have any potential audiences in mind?

# What is data visualisation?

**sense-making** (also called data analysis) and **communication'** Stephen Few

...showing quantitative and qualitative information so that a viewer can see **patterns, trends, or anomalies, constancy or variation'** Michael Friendly

...interactive, visual representations of abstract data to **amplify cognition'** Card et al

# and purpose

Exploratory or explanatory: find new insights, or tell a story?

Product or process?

Pragmatic, emotive?

Static or interactive? Print or digital?

'Distant reading' - focus on the shape rather than detail of a collection

Franco Moretti

Data visualisation can help you...

Explore your data

Explain your results



# Exploring data

Discipline

(Missing this field)

Activist Anatomist

Archaeologist Artist

Astronomer Aviator

Biologist Botanist Chemist

Collector Composer Computer

Engineer Doctor Farmer

Historian Hostess Intellectual

Inventor Mathematician

Mountaineering Natural historian

Painter Philanthropist

Photographer Physician

Physicist Pirate Poet

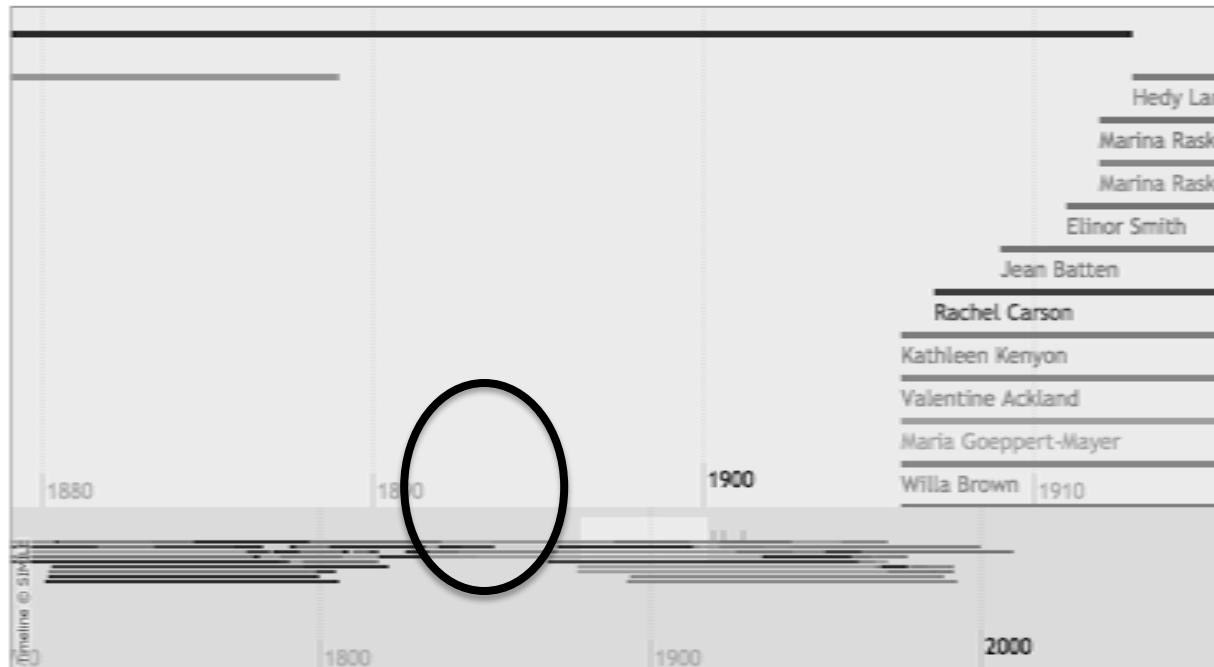
Politician Polymath Scientist

Surgeon Writer

139 Items

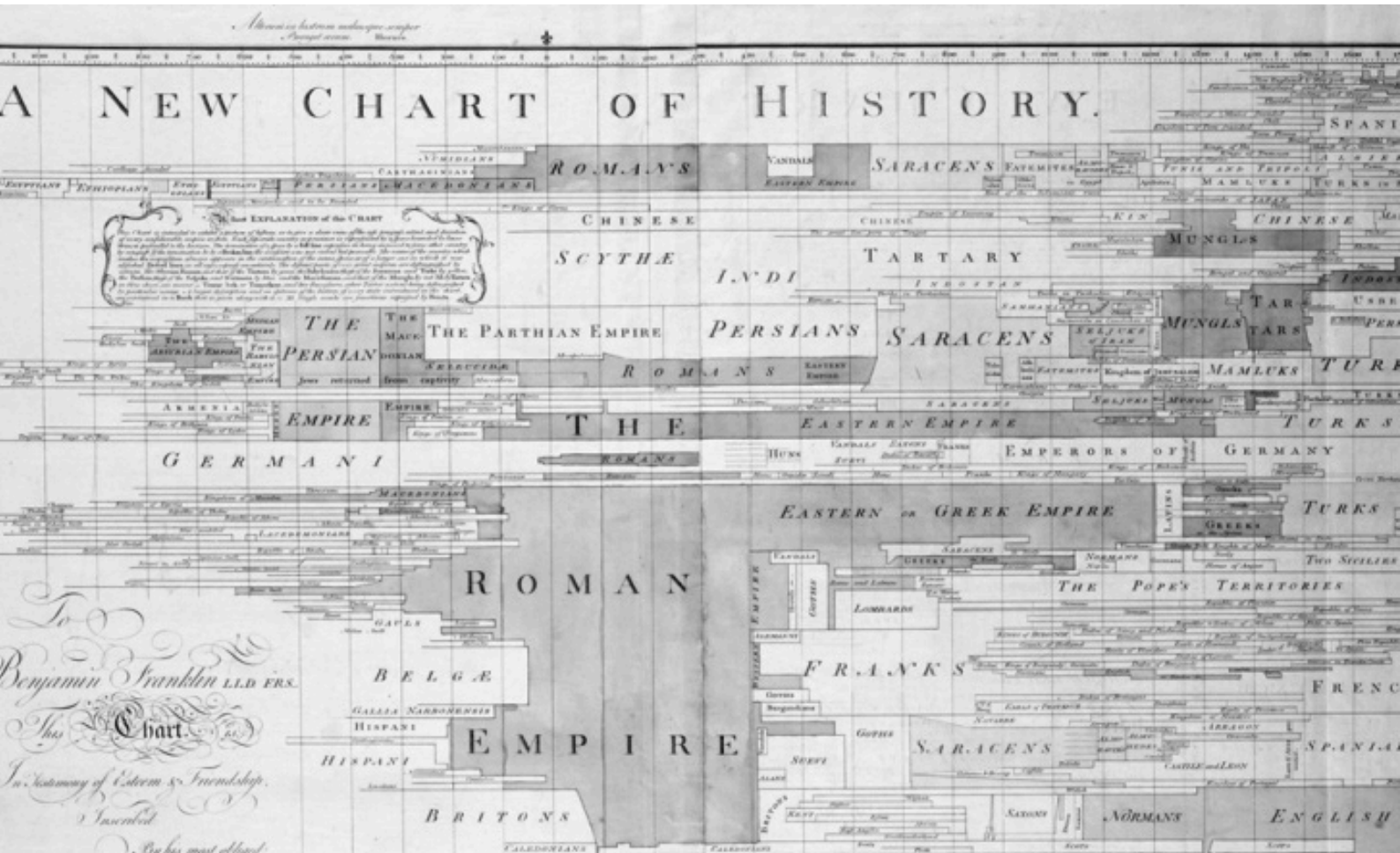
[LIST](#) • [MAP](#) • [GALLERY](#) • [TIMELINE](#)

29 results out of 139 cannot be plotted.

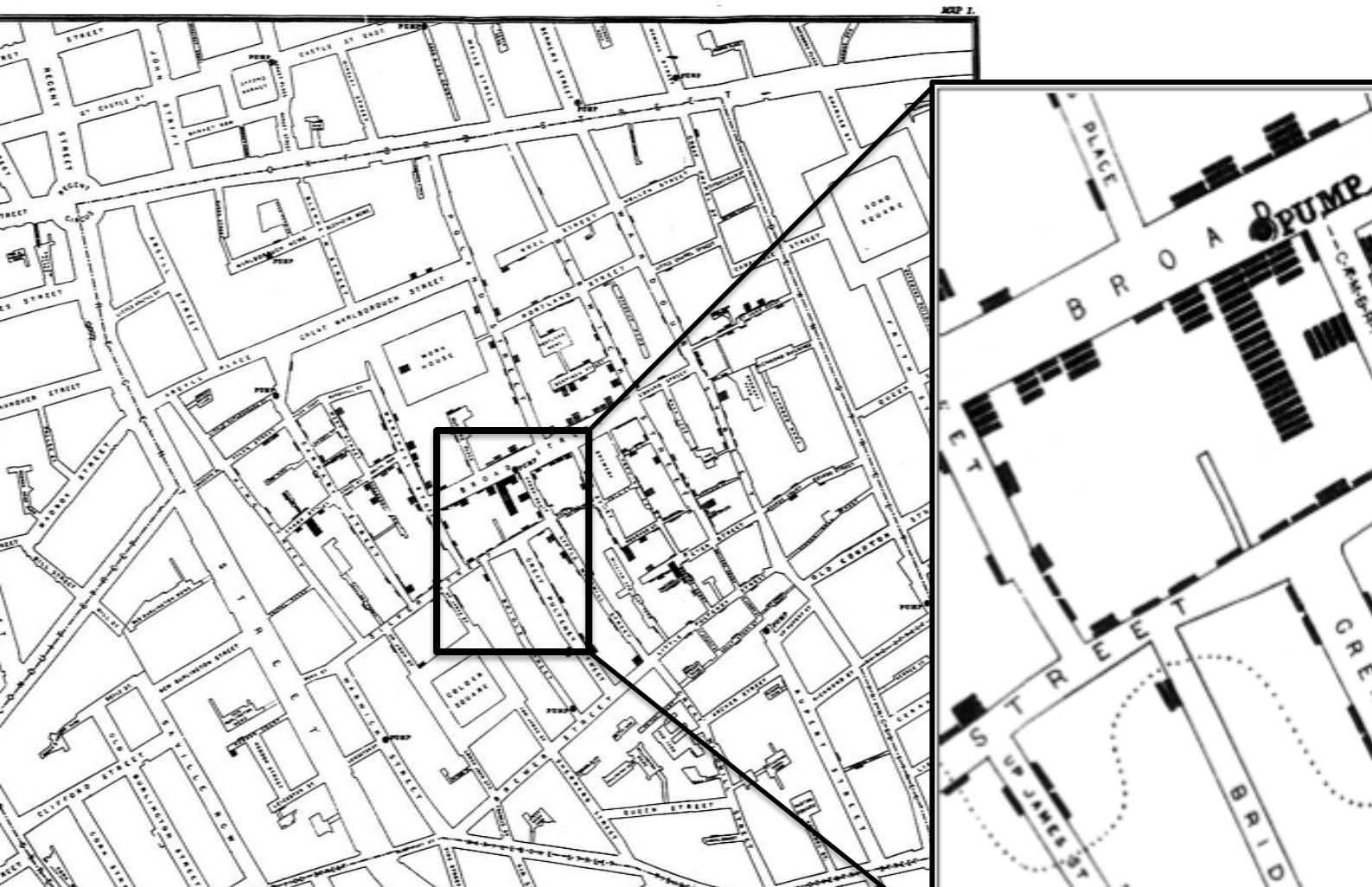


# **HISTORY AND TYPES OF VISUALISATIONS**

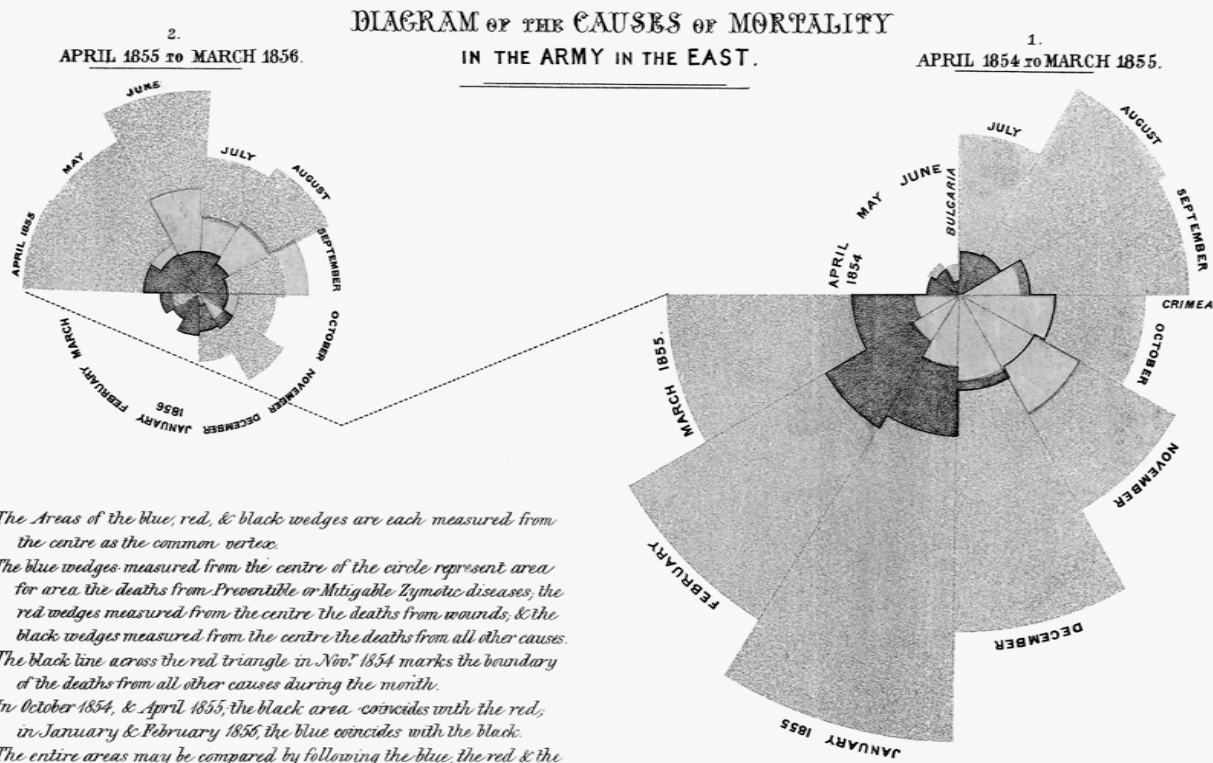
# Joseph Priestley, 1/69



# John Snow's cholera map, 1854



# Florence Nightingale's petal charts, 1855

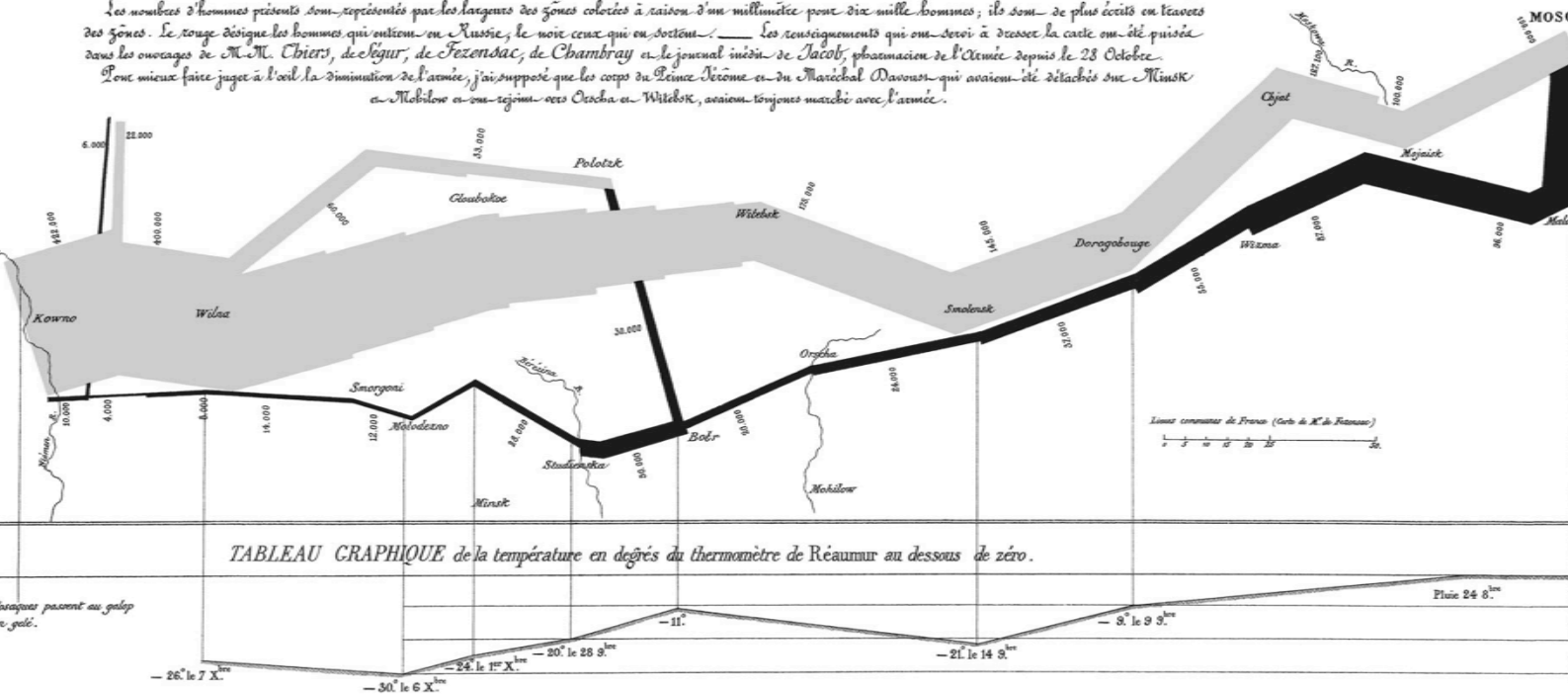


# Charles Minard's figurative map, 1869

## 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 larges des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en lettres dans les zones. Le rouge désigne les hommes qui ont été en Russie, le noir ceux qui en sont restés. Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Chéguir, de Fozensac, 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 Mohilew et qui se rejoignent vers Orscha et Witebsk, avaient toujours marché avec l'armée.



# ...updated



# UNDERGROUND

## RAILWAYS

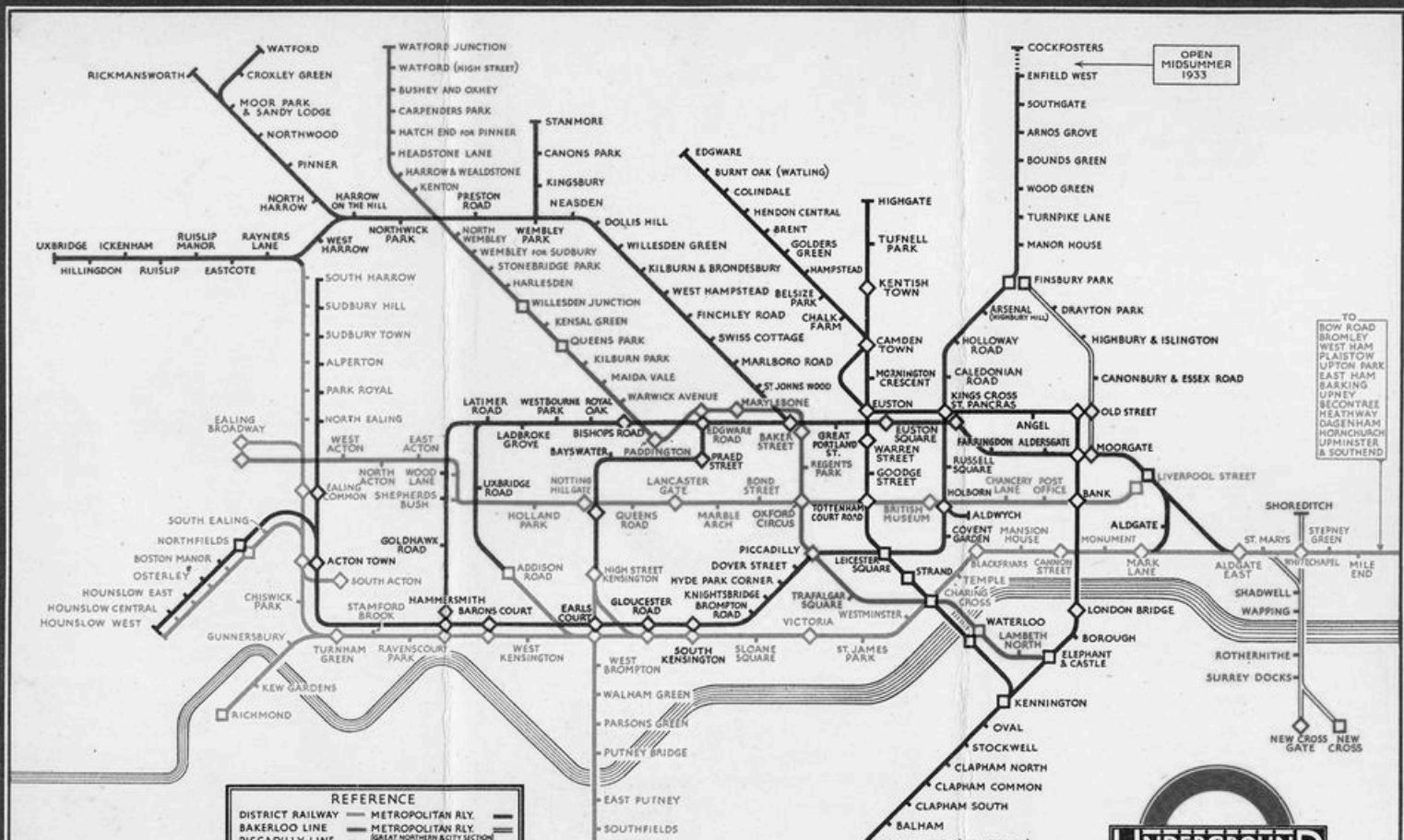
**Abstract**

BAKERLOO RY.	Shown True	
CENTRAL LONDON RY.	-	
CITY & SOUTH LONDON RY.	-	
DISTRICT RAILWAY.	-	
GREAT NORTHERN & CITY RY.	-	
RAMPSTEAD RY.	-	
METROPOLITAN RY.	-	
PICCADILLY RY.	-	
LONDON UNITED TRAMS.	-	
OTHER TRAMWAYS,	-	
SUBWAY CONNECTIONS,	-	
OTHER RAILWAYS.	-	





# Harry Beck, 1931

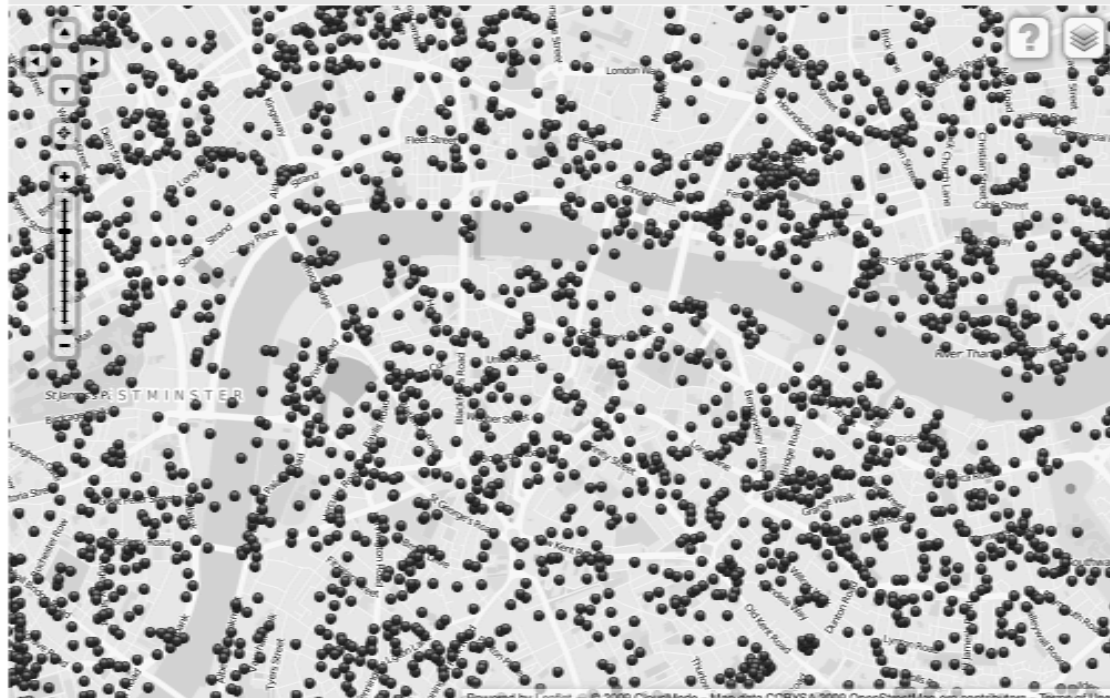


# Web 2.0 and the mashup, 2006

**BOMB SIGHT**  
Mapping the WW2 bomb census

[Map](#) [Explore London](#) [About](#) [FAQ](#) [Data](#) [Android App](#)

## EXPLORE THE LONDON BLITZ during 7th October 1940 to 6th June 1941

[Information](#)

Powered by Leaflet - © 2008 OpenStreetMap contributors - Map data © OpenStreetMap contributors, Imagery © Mapbox  
The National Archives give no warranty to the accuracy, completeness or fitness for purpose of the information provided. Images should be used only for purposes of research, private study or education. Applications for any other use should be made to The National Archives Image Library, Kew, Richmond, Surrey TW9 4DU. Tel: 02083925255 Fax: 02083925266

I'VE SEEN THIS WEEK

3,274

Estimate that looks definite when it's big

THE PERCENTAGE OF THOSE  
WHICH ADDED NOTHING BUT  
SIMPLE GRAPHS AND COLOR

100%

Footnote adding weight

THAT SAME FIGURE AS A PIE  
CHART WHICH MAKES THIS  
LOOK MORE COMPLICATED

YES ☒ NO ☐



Maps make it look more serious and complex

HOW MUCH MORE INTERESTING  
THOSE "INFOGRAPHICS" COULD  
HAVE BEEN IF THE CREATORS  
HAD TRIED TO ILLUMINATE  
SOME POINT OTHER THAN THE  
SIZE OF ONE-DIMENSIONAL  
SETS OF NUMBERS RELATIVE  
TO EACH OTHER AND TELLING  
US NOTHING THAT WE  
COULDN'T HAVE DISCOVERED  
BY SEEING A COUPLE OF  
NUMBERS PRINTED IN  
NORMAL TEXT RATHER  
THAN SOME OVERBLOWN,  
SELF-IMPORTANT  
"VISUALIZATION" THAT  
GIVES THE TALENTED  
CREATORS OF GOOD,  
EYE-OPENING  
INFOGRAPHICS  
A BAD NAME

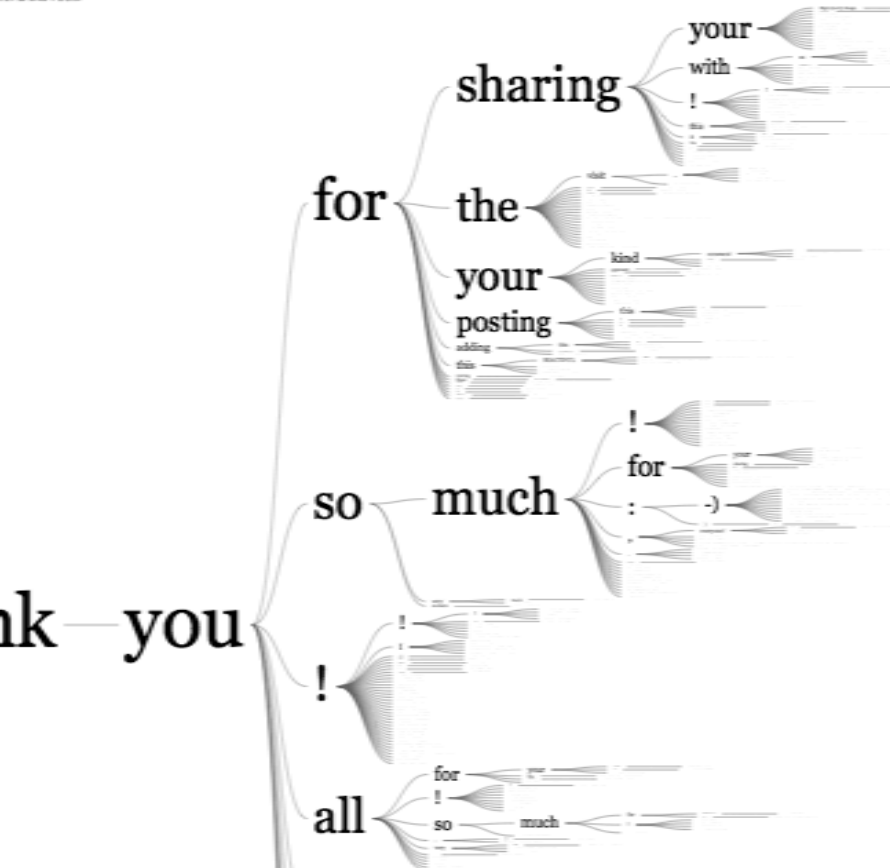
# Exploring words

## word tree

Thank you

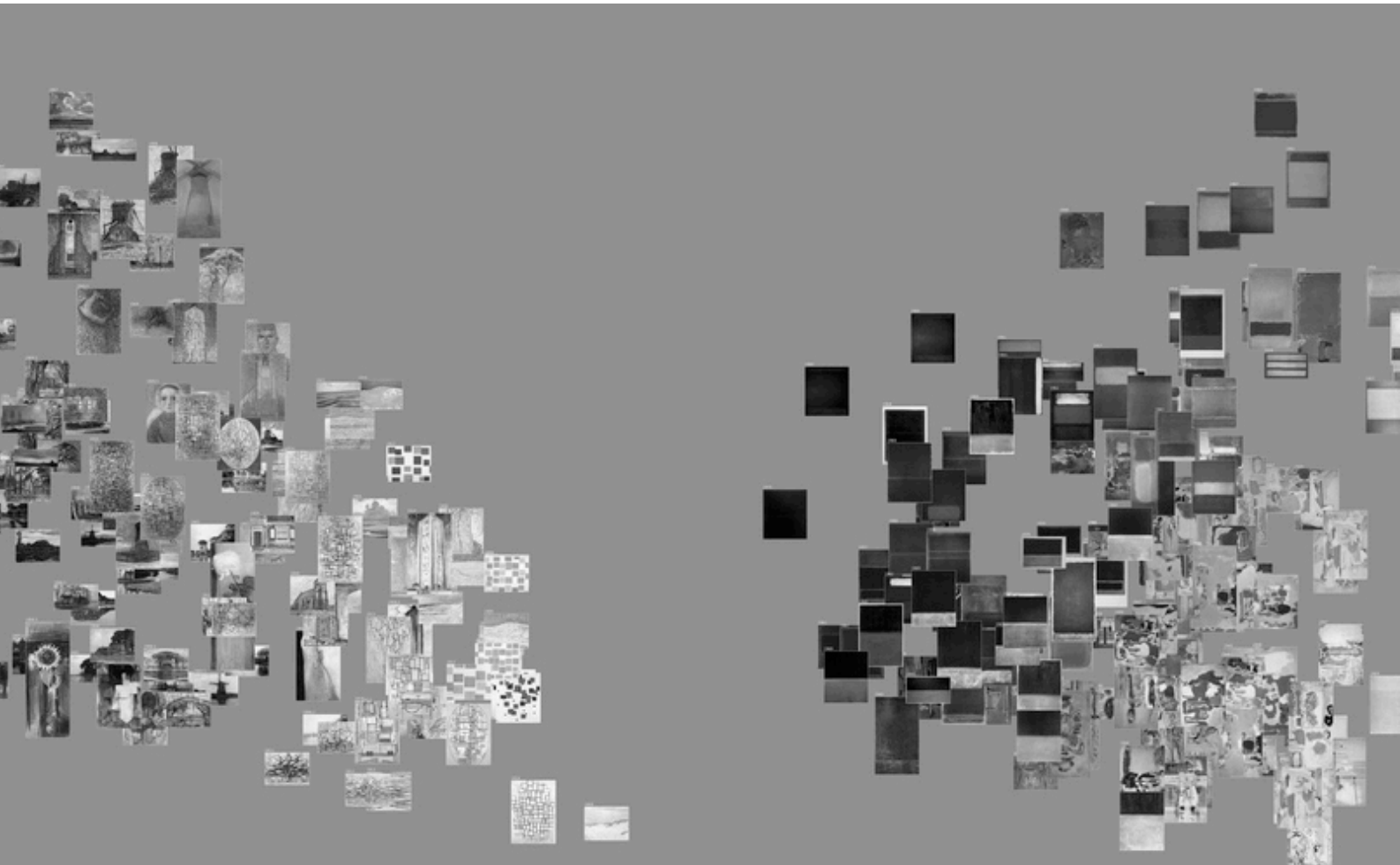
☐ reverse tree    ☐ one phrase per line

board the roost.



congrats! i have  
COOL shot! I love  
earrings! I NEED  
this!!!Awww those  
cute!so cool!!!C  
That's awesome  
congratulations!  
published three  
time feels like  
LOL. Great earring  
covet them.HI Hel  
good to see you p  
would love a pair  
really so sweet!  
presentation and  
Congratulations  
one Helen. I like  
Seen on your photo  
? )Oh i Love the  
! Really need a p  
now to give me a  
Lovely shot aswe  
Conrats =DColl sh  
cute, I thought a  
they were key rin  
DOF and color and  
on yet another p  
well deserved !w  
that's great!congr  
,no-one deserves  
moreAwwwwwwwwww  
cute! I wish I ha  
best...Df

# Visualising images and video, 2012



# Data types

quantitative / qualitative

geographic, time series

media

entities (people, places, events, concepts, things)

# **CRITIQUING VISUALISATIONS**

# 'sentiment'

## Olympics in review – the big Twitter sentiment timeline

Average Sentiment



negative

positive

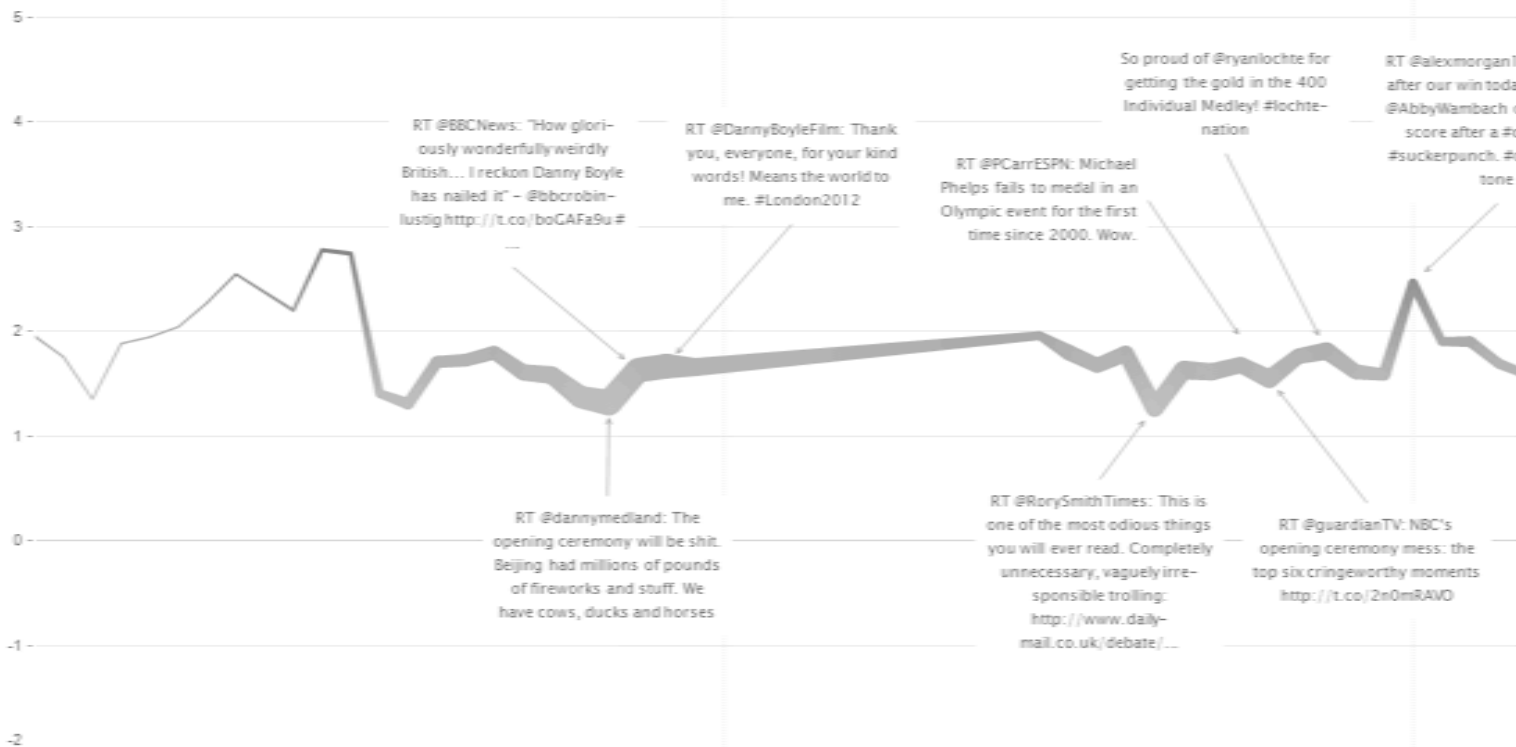
Tweets / hour



1000

50'000

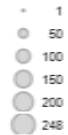
100'000





# Visualisations and 'truthiness'

Number of Records

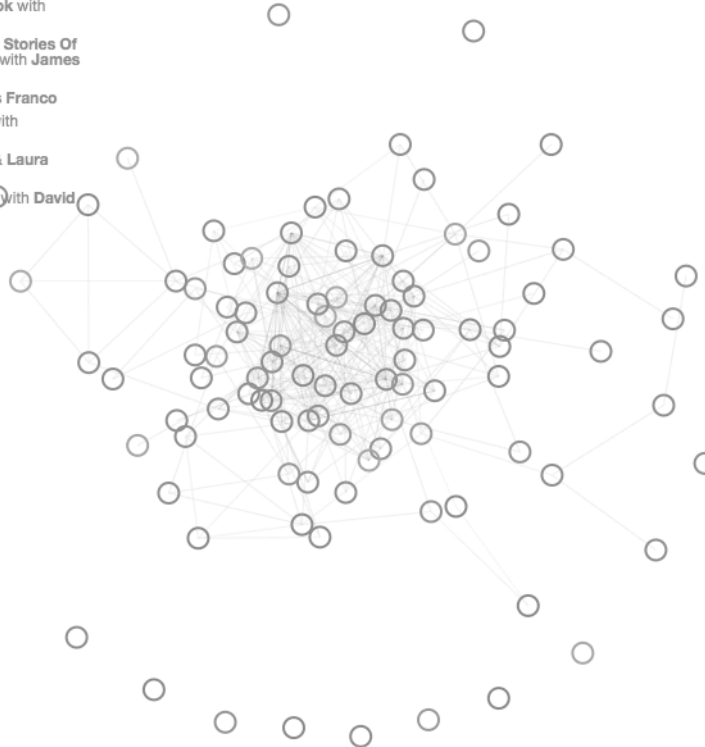


# Network visualisations

Hover over each node to see the contributor's neighbors.

Scott Spencer's shelf shares:

- » The Stories of John Cheever with Michael Chabon
- » The Stories of John Cheever with Sloane Crosley
- » The Golden Notebook with Jennifer Egan
- » The Complete Short Stories Of Ernest Hemingway with James Franco
- » Pale Fire with James Franco
- » A Mother's Kisses with Jonathan Lethem
- » Pale Fire with Kate & Laura Mulleavy
- » Revolutionary Road with David Sedaris



# Exercise 1: network visualisations

Instructions on the hand-out.

# N-grams

## Google books Ngram Viewer

Graph these case-sensitive comma-separated phrases:

between  and  from the corpus  with smoothing of .

 Tweet

[Search lots of books](#)



Search in Google Books:

[1650 - 1767](#)

[1768 - 1774](#)

[1775 - 1780](#)

[1781 - 1842](#)

[1843 - 1850](#)

[electricity](#)

[English](#)

[1650 - 1774](#)

[1775 - 1783](#)

[1784](#)

[1785 - 1790](#)

[1791 - 1850](#)

[phlogiston](#)

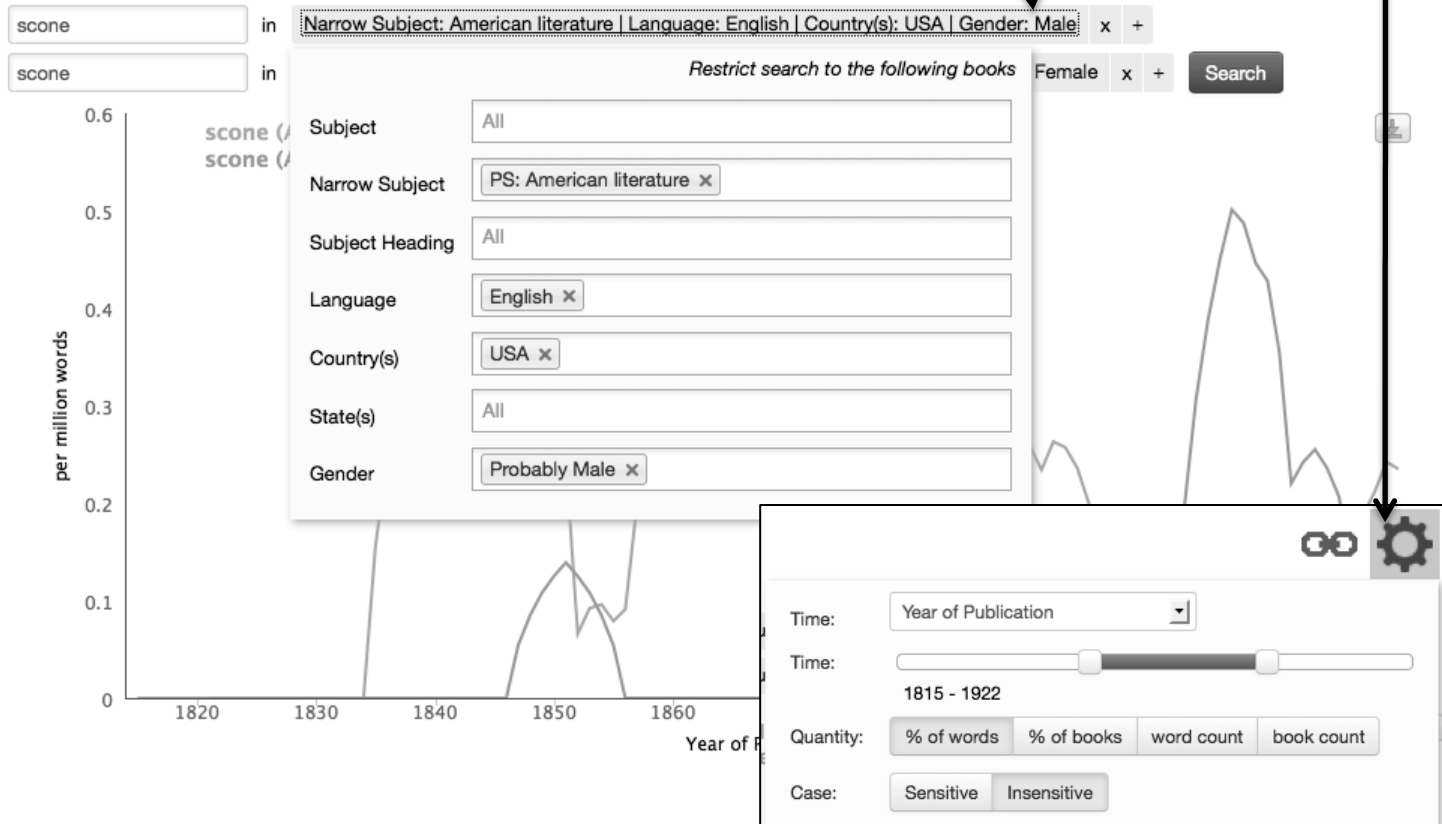
[English](#)

# Exercise 2: comparing N-gram tool

Bookworm tip: [click here to change options](#)

## bookworm Open Library

Search for trends in hundreds of thousands of books on [OpenLibrary.org](#)



# Topic modelling

## Topics

future man people men world great life day time present years country things make work good past human true

australia year wool industry future trade production years world prices australian present war price cent time market increase great

It is the lie that is the arc good because of the local mid game mini industry

air future motor power car speed miles made engine day flying years light cars high great ship machine oil

land wheat good farmers sheep sugar years stock time country cattle mr queensland soil farm acres fruit farmer agricultural

city railway water future land town north miles port south building river years area present street sydney line great

war british empire germany britain great german peace states united nations france europe india world russia government military china

mine company mines mining gold and ore industry work with them mr futuro thom boen thero which wore ft

mr government state australia future party made minister members public present sir general council federal meeting labour work commonwealth

school women future home children young work years day woman good house boys mrs time man year made miss

# Other forms of text analysis

JB/002/004/001

1818 April 19 4

Annuity Notes. Advertisement

4

Aware of the manner in which to a degree beyond my power of estimation the scheme would <sup>could not but be</sup> prejudicial to the interests of the Bank of England, the success <sup>never had the ultimate</sup> adoption of the

scheme[?] had from the first been regarded by bus[?]

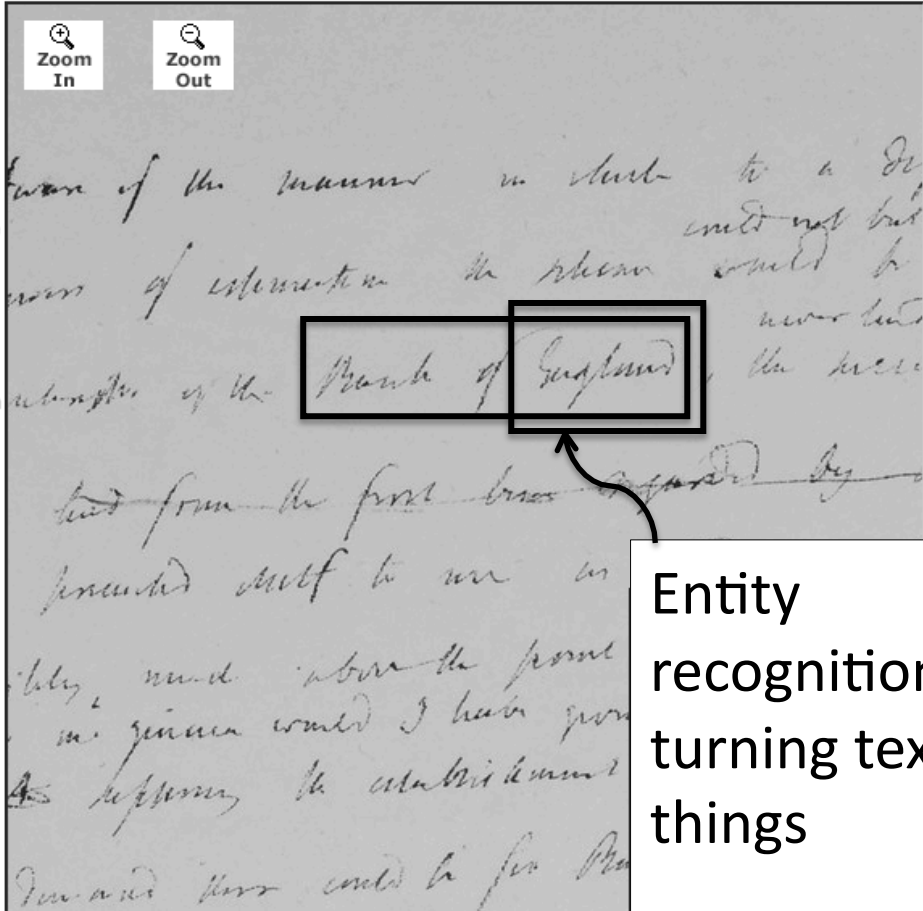
scheme presented itself to me as standing in the scale of probability much above the point of hopelessness. Not so much as one guinea would I have given to receive a hundred in one of success. But in comparison of the advantage <sup>in that event</sup>

which presented itself to my eyes, the offering I cast <sup>cast as above</sup> into the Treasury was but a mite!

As Supposing the establishment of this currency, what use or demand there could be for Bank paper I neither saw nor saw <sup>found</sup> any reason for being anxious to enquire.

But the particular interest of that corporation is in a way now understood which is not now a secret to any body when seen truth to all eyes understood with the interest of subservient to the most insatiable

of all lusts the lust of ambition <sup>power</sup> and in that way inseparably linked and interwoven with that not only with the sinister interest of any administration but but thence with



Entity  
recognition:  
turning text  
things

# Exercise 3: trying entity recognition

Instructions on the hand-out.



# Entity recognition examples

## Entity Recognition:

Speaking at a UN conference in Sendai, Japan, on Monday, said 90% of buildings in Port Vila had  
aged or destroyed by the category five storm, which saw winds of up to 250km/h (150mph).  
à setback for the government and for the people of Vanuatu, he said.  
erall the development that has taken place, all this development has been wiped out."  
communications in the Port Vila province of the island have now been 'almost fully restored', according  
telecommunications provider Digicel, allowing information to flow more freely to and from disaster  
s.  
eral countries have also pledged additional aid and funding for the stricken island nation.

Misc Person MONEY Location

# **VISUALISATIONS FOR SCHOLARLY ANALYSIS**

# Scholarly data visualisations

Visualisations as 'distant reading' where distance is 'a specific form of knowledge: fewer elements, hence a sharper sense of their overall interconnection' Moretti, 2005

Inspiring curiosity and research questions

But - which questions do they privilege and what do they leave out?

Exercise 4: explore scholarly visualisation

pair up and discuss together before reporting  
back.

Instructions on the hand-out.

## C

# Letters"



# GAPVis

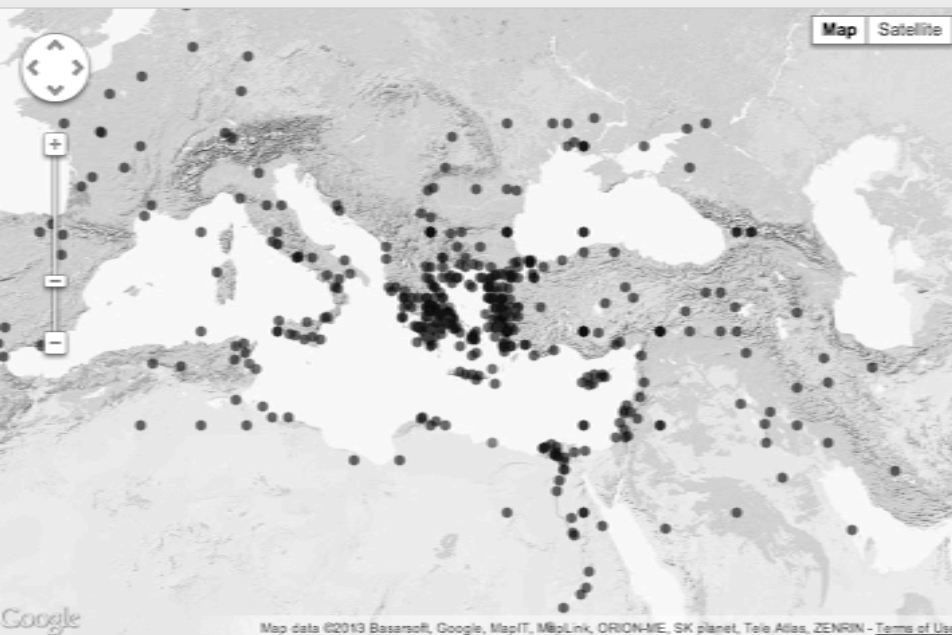
## Herodotus

By Herodotus (translated by W. Beloe)

Published 1830 · View on Google Books

The Google Books version of "Herodotus", by Herodotus (translated by W. Beloe), was published in 1830. It references 486 identified ancient places. The place most frequently referenced is Hellas, followed by Sardis/Hyde?, Athenae, and Dirphys M.

[Go to Reading View >](#)



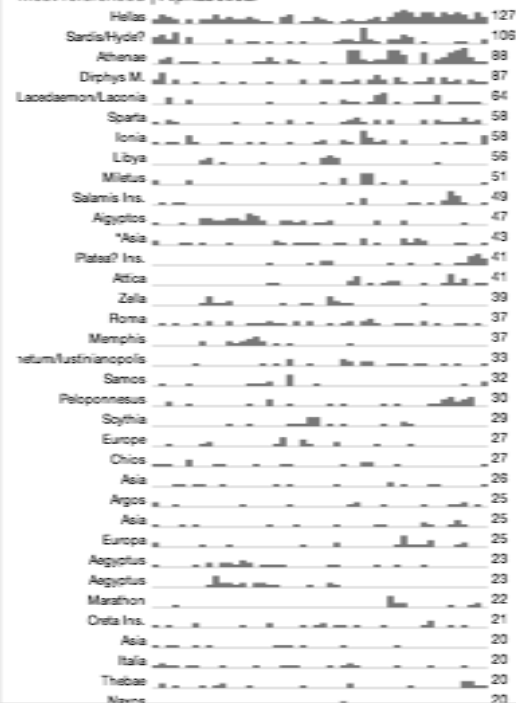
★ Book Summary

📖 Reading View

📍 Place Data

## Top Places

Most referenced | Alphabetical



# Digital Harlem

## Digital Harlem

Everyday Life 1915-1930

[HOME](#)[ABOUT](#)[THE MAP](#)[SOURCES](#)[HOW TO...](#)[FEATURED](#)[EVENTS](#) [PEOPLE](#) [PLACES](#)

Type of Event

Select ...

Keyword(s)

From Date...

To Date...

Charge / Conviction

Select ...

Birthplace of Participant(s)

Select ...

Occupation of Participant(s)

Select ...

Race

Gender

Select ...

Select ...

Surname of Participant(s)

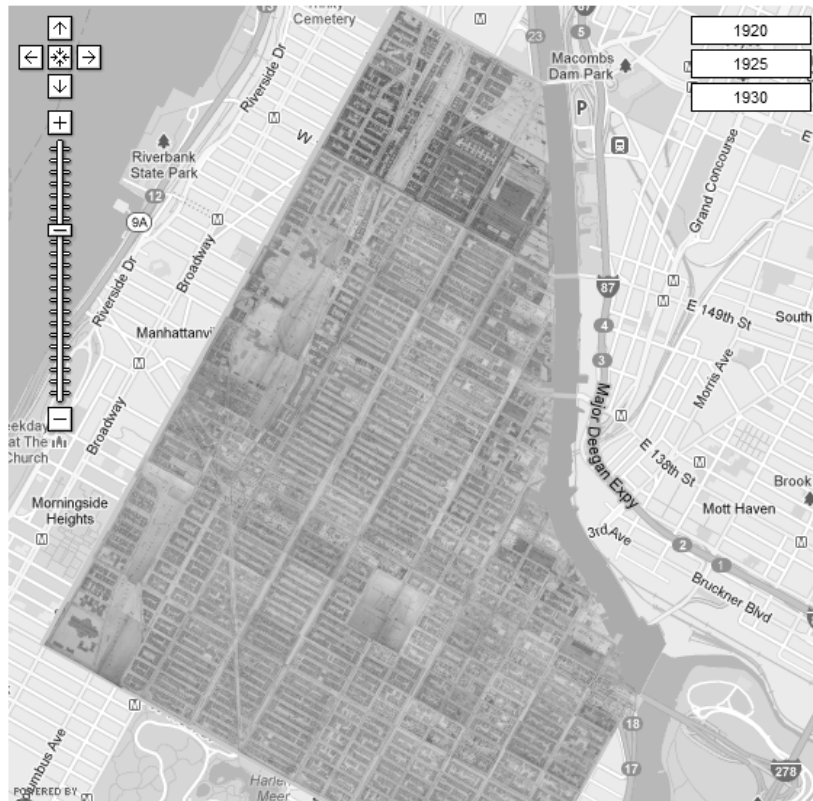
Street / Intersection Name

Select ...

Location Type

Select ...

Submit

[MAPS](#)[LAYERS](#)[LEGEND](#)

January 1925

Numbers-Arrests

Fuller Long

Churches



# Orbis

ORBIS

The Stanford Geospatial Network Model of the Roman World



CREATED BY  
Walter Scheidel  
Elijah Meeks

Home Introducing ORBIS Understanding ORBIS Building ORBIS Using ORBIS **Mapping ORBIS** Applying ORBIS What's New ORBIS in the Media Credits

Mapping ORBIS

Interactive distance cartogram

Map gallery



Query and Results

calculate route reset options results grid

Start:   
Destination:   
Month of travel:

Priority  
☒ fastest ☐ cheapest ☐ shortest

Network mode  
☒ Road ☒ Coastal sea  
☒ River ☐ Coastal sea (daylight)  
☒ Open sea

Aquatic options  
River: ☐ Military ☒ Civilian  
Sea: ☒ Faster ☐ Slower

Road options  
☒ Speed options  
☒ Foot/army/pack animal (mod. load)/mule cart/camel caravan  
☐ Rapid military march  
☐ Ox cart  
☐ Porter/fully loaded mule  
☐ Horseback rider (routine travel)  
☐ Private travel (routine, vehicular)  
☐ Private travel (accelerated, vehicular/horseback)

# Digital Public Library of America

[About](#)[Get Involved](#)[For Developers](#)[Help](#)[Follow](#)[Contact](#)[Donate](#)[Log](#)

DIGITAL PUBLIC LIBRARY  
OF AMERICA

[Home](#)[Exhibitions](#)[Map](#)[Timeline](#)[Bookshelf](#)[Apps](#)

## Exhibitions

[View all »](#)

Explore  
by Place

[Map »](#)

## Wealth of Knowledge

5,537,853 items from libraries, archives, and museums

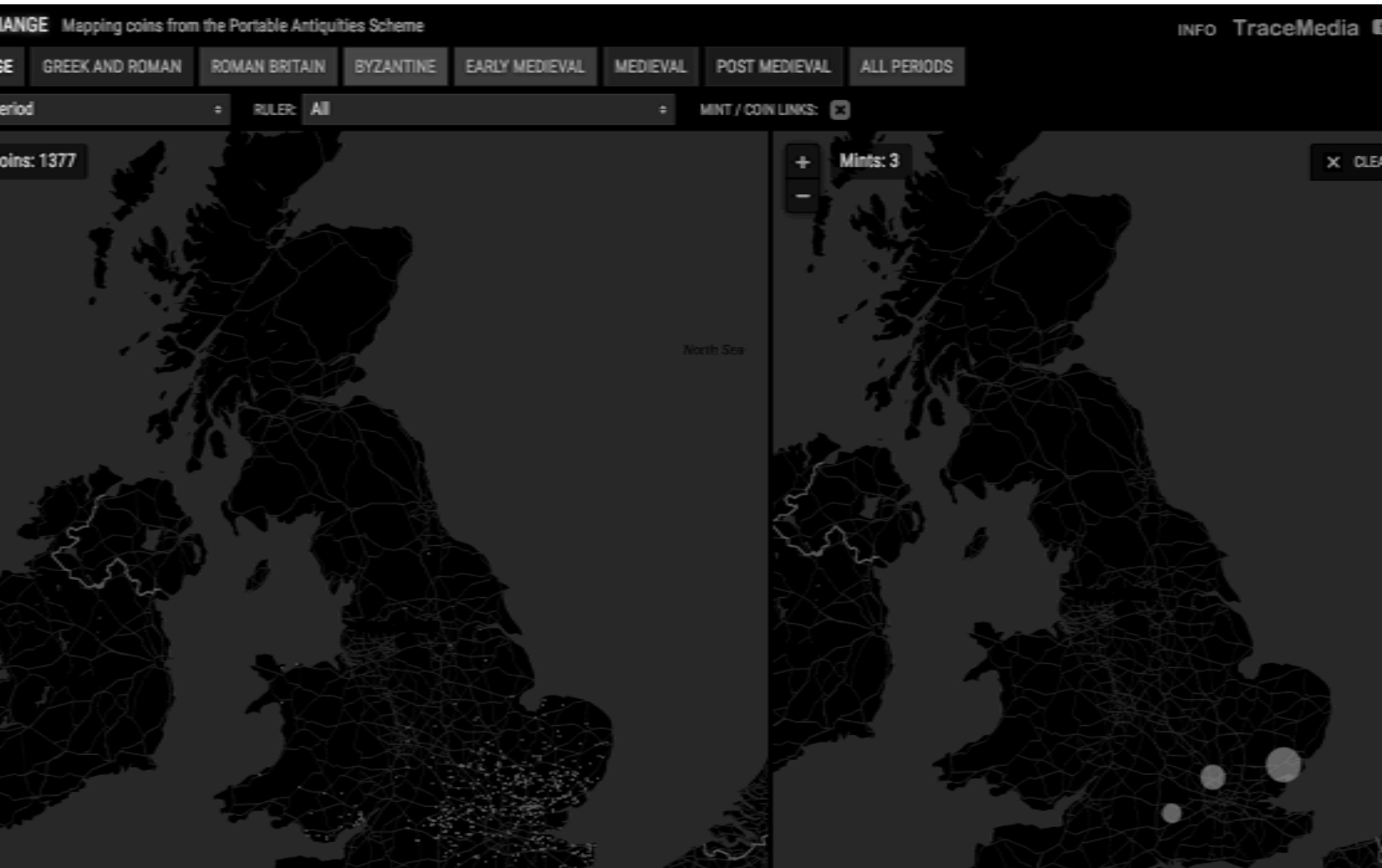
Search the Library



Explore  
by Date



# Lost Change



# State of the Union

Read Obama's State of the Union in context

Made By Ben Schmidt

Using Bookworm

## The State of the Union in Context

This site lets you use a Bookworm database to read Obama's State of the Union in the context of all the other State of the Union messages given by American presidents. For any word in the message, just **click on the text**: the word will turn red. If you want to search a two-word phrase, just highlight both words. (You can't search for phrases longer than two words).

If you then click on a bar, you'll get an (ugly, unformatted, alas) list of all the times that President used the phrase. Certain forms of punctuation (hyphenated words, etc.) won't work quite right yet. Keep in mind that the State of the Union has changed a lot in its history. Of particular importance to understanding these charts is knowing that it was written, not spoken, before Woodrow Wilson. You're looking at the  address: change that dropdown to use any other year as a baseline.


Also, check out the two pieces Mitch Fraas and I wrote for the Atlantic on the language of past State of the Unions and the locations mentioned in past speeches. I owe the idea for this layout to Yoni Appelbaum and John Gould at the Atlantic. The text of the speeches is from the American Presidency project at UCSB, except for the 2015 text which I took from the White House's post at Medium.

ker, Mr. Vice President, Members of Congress, my fellow Americans:

teen years into this new century. Fifteen years that dawned with terror touching  
s; that unfolded with a new generation fighting two long and costly wars; that  
ious recession spread across our nation and the world. It has been, and still is,  
e for many.

nt, we turn the page.

foreign oil



President	Frequency (approximate)
James Carter	100
Gerald Ford	95
Barack Obama	45
George W. Bush	35

James Carter  
Gerald Ford  
Barack Obama  
George W. Bush

# Comments or questions?

**ISSUES WITH HUMANITIES DATA**

# Considerations for data

Commercial tools often assume complete, born-digital datasets – no missing fields or changes in data entry over time

Records often contain uncertainty and fuzziness (e.g. date ranges, multiple values, uncertain or unavailable information)

Includes metadata, data, digital surrogates

# Messiness in GLAM data

'Begun in Kiryu, Japan, finished in France'

'Bali? Java? Mexico?'

Variations on USA:

- U.S.
- U.S.A
- U.S.A.
- USA
- United States of America
- USA ?
- United States (case)

Inconsistency in uncertainty

- U.S.A. or England
- U.S.A./England ?
- England & U.S.A.



# When were objects collected?

Many Eyes

Log in

IBM



An experiment brought to you by IBM Research and the IBM Cognos software group

Visualizations

Search

## Visualizations : Cooper Hewitt accessions by department over time

Uploaded by: miah

Created at: Jun 15 2012

Description: Experimental visualisation of collecting by different departments over time.

### Explore

Visualizations  
Data sets  
Comments  
Topic centers

### Participate

Create a visualization  
Upload a data set  
Create a topic center  
Register

### Learn more

Quick start  
Visualization types  
Data format and style  
About Many Eyes  
FAQ  
Blog

### Contact us

Contact  
Report a bug

### Legal

Terms of use  
Privacy  
Provide Feedback

#### Legend

Click to select,  
Ctrl-Click: multiple  
Shift-Click: range

Wallcoverings\_objects  
Textiles\_objects  
ProductDesignDecArts  
DrawingsPrintsGraphicl  
Total\_objects



Click or ctrl-click to highlight points on graph.

Relative: Set Start = 100

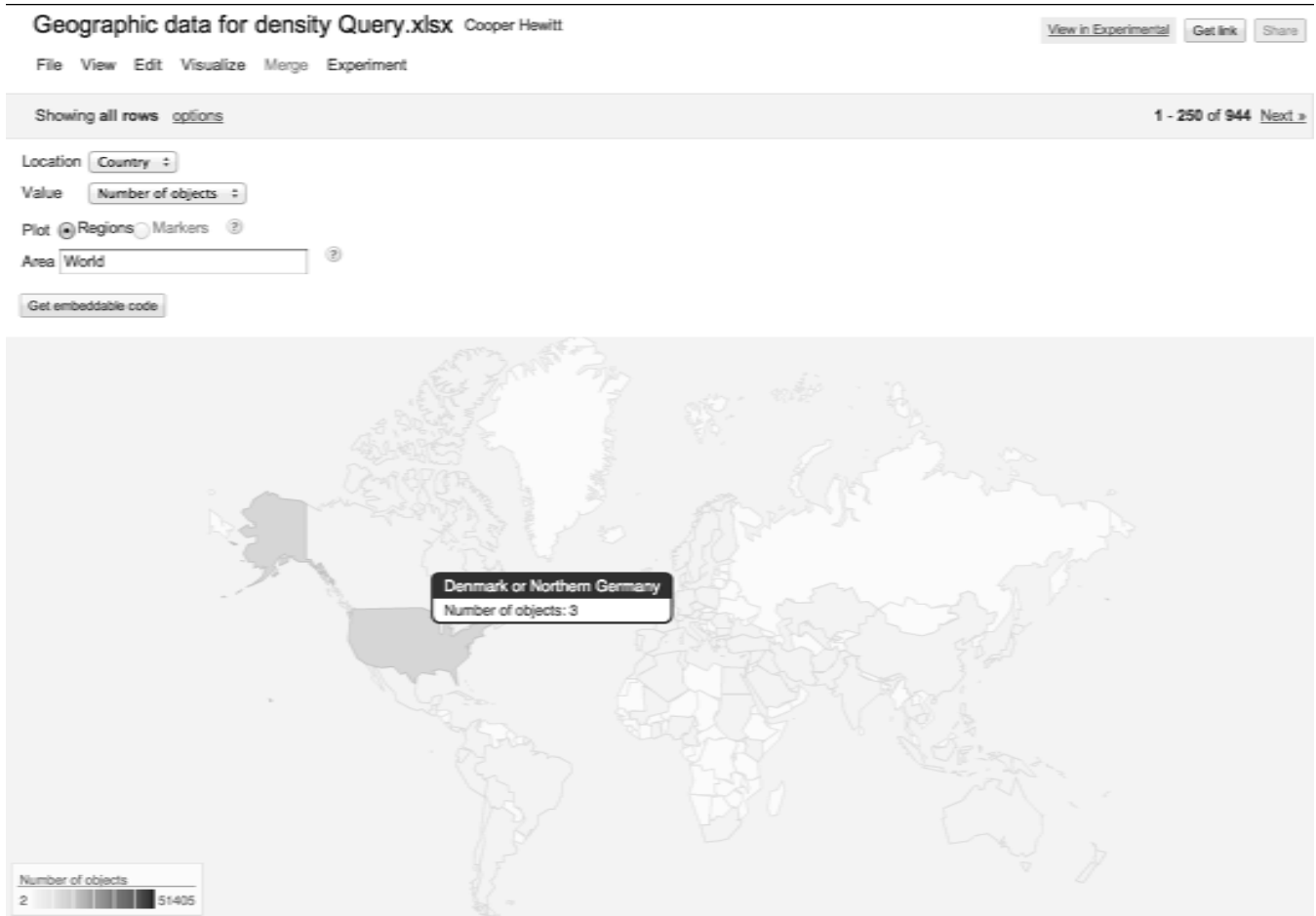
Aggregate items with same label: Average

Sort: labels data order

Visualizations Data sets

2008 2009 2010 2011

# Computers don't cope



# Preparing data for visualisations

data often needs manual cleaning to:

- remove rows where vital information is missing

- tidy inconsistencies in term lists or spelling

- convert words to numbers (e.g. dates)

- remove hard returns and non-ASCII characters (or change data format)

- split multiple values in one field into other columns (e.g. author name, date in single field)

- expand coded values (e.g. countries, language)

# Data Preparation

Generally needs to be in tables, one row per item, one column per value

Might need to calculate values in advance

Data should be made as consistent as possible with tools like

- Excel
- OpenRefine <http://openrefine.org>

# Open Refine

0.0.1:3333/project?project=2326347516179



DuckDuckGo



## Cluster & Edit column "Place of publication"

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. Find out more ...

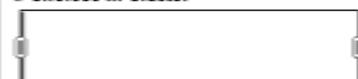
Method **key collision**

Keying Function **fingerprint**

111 clusters found

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
6	3773	<ul style="list-style-type: none"><li>London (1980 rows)</li><li>London (1785 rows)</li><li>London? (4 rows)</li><li>London? (2 rows)</li><li>London (1 rows)</li><li>london (1 rows)</li></ul>	<input checked="" type="checkbox"/>	London
4	5	<ul style="list-style-type: none"><li>Edinburgh printed and London (2 rows)</li><li>Edinburgh and London printed (1 rows)</li><li>London and Edinburgh printed? (1 rows)</li><li>London printed and Edinburgh (1 rows)</li></ul>	<input checked="" type="checkbox"/>	Edinburgh and London
4	12	<ul style="list-style-type: none"><li>Edinburgh London (7 rows)</li><li>London Edinburgh (3 rows)</li><li>Edinburgh ; London (1 rows)</li><li>London ; Edinburgh (1 rows)</li></ul>	<input checked="" type="checkbox"/>	Edinburgh London
4	209	<ul style="list-style-type: none"><li>Edinburgh (127 rows)</li><li>Edinburgh (80 rows)</li><li>Edinburgh (1 rows)</li><li>Edinburgh? (1 rows)</li></ul>	<input type="checkbox"/>	Edinburgh

# Choices in Cluster



2 — 6

# Rows in Cluster



0 — 3800

Average Length of Choices



4 — 34

Length Variance of Choices



0 — 1

Select All

Deselect All

Merge Selected & Re-Cluster

Merge Selected & Close

Close

# ...but be careful

## Cluster & Edit column "Place of publication"

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "New York City" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. Find out more about this feature.

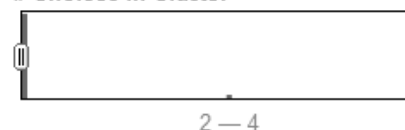
Method **key collision**

Keying Function **cologne-phonetic**

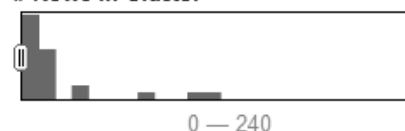
25 clusters found

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
32		<ul style="list-style-type: none"><li>Boston (24 rows)</li><li>Pesten (4 rows)</li><li>Pesthini (2 rows)</li><li>Pestini (2 rows)</li></ul>	<input type="checkbox"/>	<input type="text" value="Boston"/>
12		<ul style="list-style-type: none"><li>Москва (4 rows)</li><li>Санктпетербургъ (4 rows)</li><li>въ Санктпетербургъ (4 rows)</li></ul>	<input type="checkbox"/>	<input type="text" value="Москва"/>
109		<ul style="list-style-type: none"><li>Paris (108 rows)</li><li>Portsea (1 rows)</li></ul>	<input type="checkbox"/>	<input type="text" value="Paris"/>
69		<ul style="list-style-type: none"><li>New York (65 rows)</li><li>Newark (4 rows)</li></ul>	<input type="checkbox"/>	<input type="text" value="New York"/>
3		<ul style="list-style-type: none"><li>Chatham (2 rows)</li><li>Catania (1 rows)</li></ul>	<input type="checkbox"/>	<input type="text" value="Chatham"/>
8		<ul style="list-style-type: none"><li>Basil (4 rows)</li><li>Paislev (4 rows)</li></ul>	<input type="checkbox"/>	<input type="text" value="Basil"/>

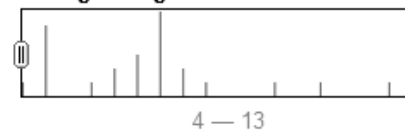
# Choices in Cluster



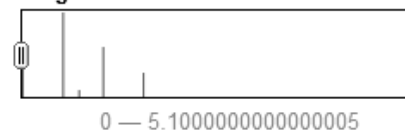
# Rows in Cluster



Average Length of Choices



Length Variance of Choices



# PLANNING VISUALISATIONS

Data

Purpose

Audience

Structure

```
graph TD; Data[Data] --> Structure((Structure)); Purpose[Purpose] --> Structure; Audience[Audience] --> Structure;
```

The diagram illustrates a conceptual model where three distinct inputs—Data, Purpose, and Audience—converge to define a central Structure. Each input is contained within a light gray box with rounded corners, while the central Structure is represented by a light gray circle. Arrows from each box point towards the central circle, indicating a flow of information or influence.



Purpose, data, audiences (revision)

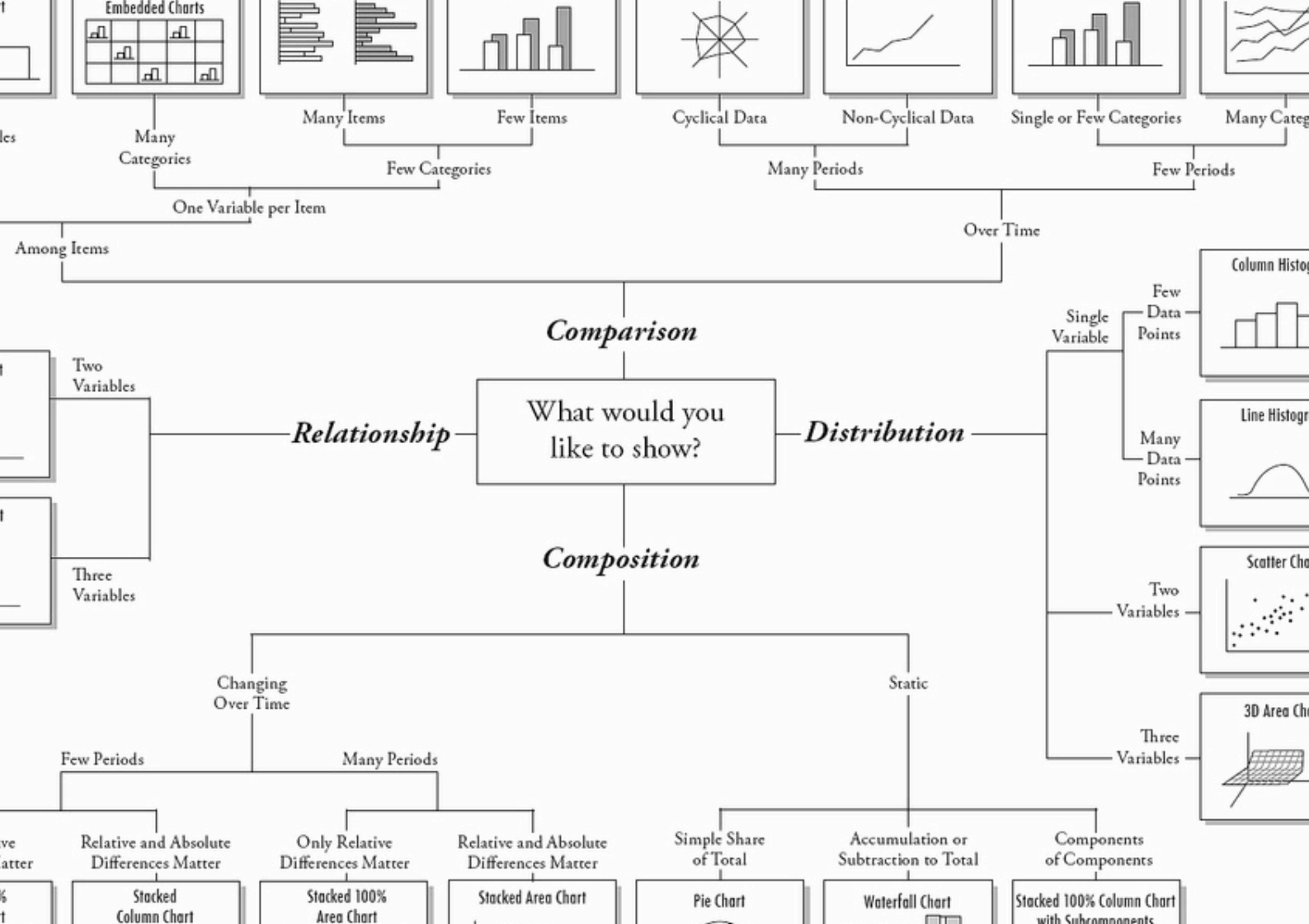
Intersections of format and purpose

Data types: quantitative, qualitative,  
geographic, time series, media, entities  
(people, places, events, concepts, things)

Static, interactive; print, digital; product,  
process

Exploratory, explanatory: find new insights  
or tell a story? Pragmatic, emotive?

# Choosing a structure



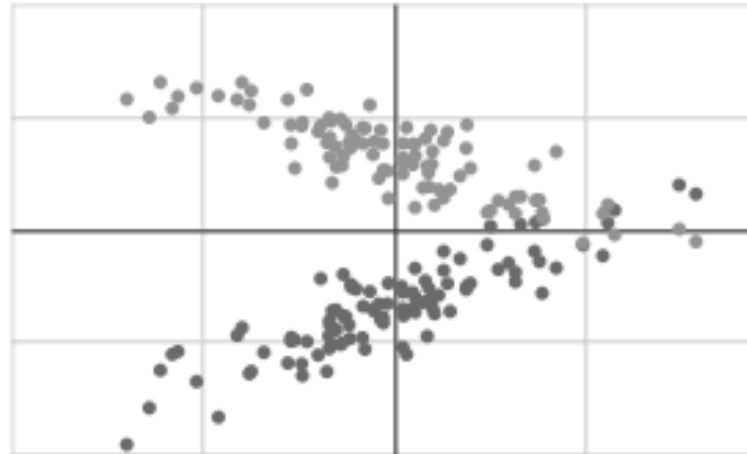
# See relationships among data points

Scatterplot

Matrix

Network diagram

Scatter Chart



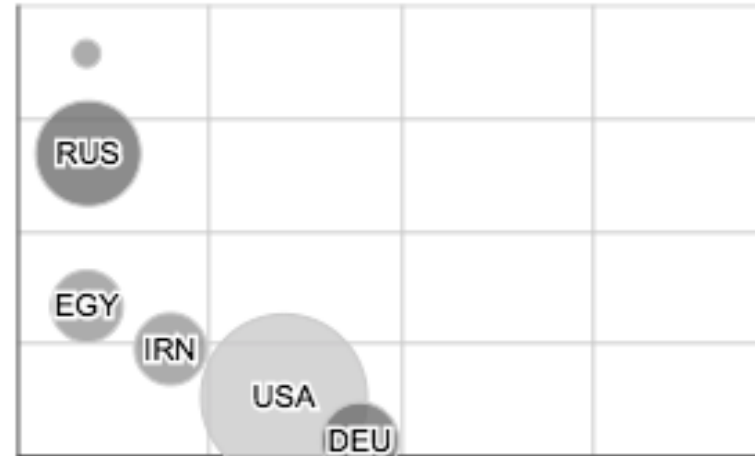
# Compare a set of values

Bar chart

Bubble chart

Histogram

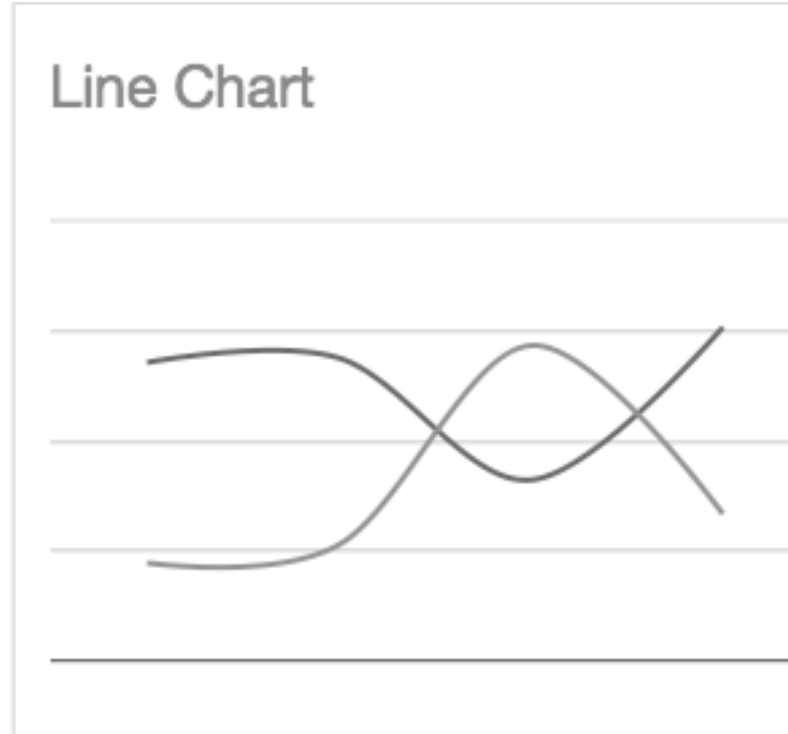
Bubble Chart



# Track change over time

Line graph

Stack graph



# See the parts of a whole

Pie chart

Treemap



# Fusion Tables

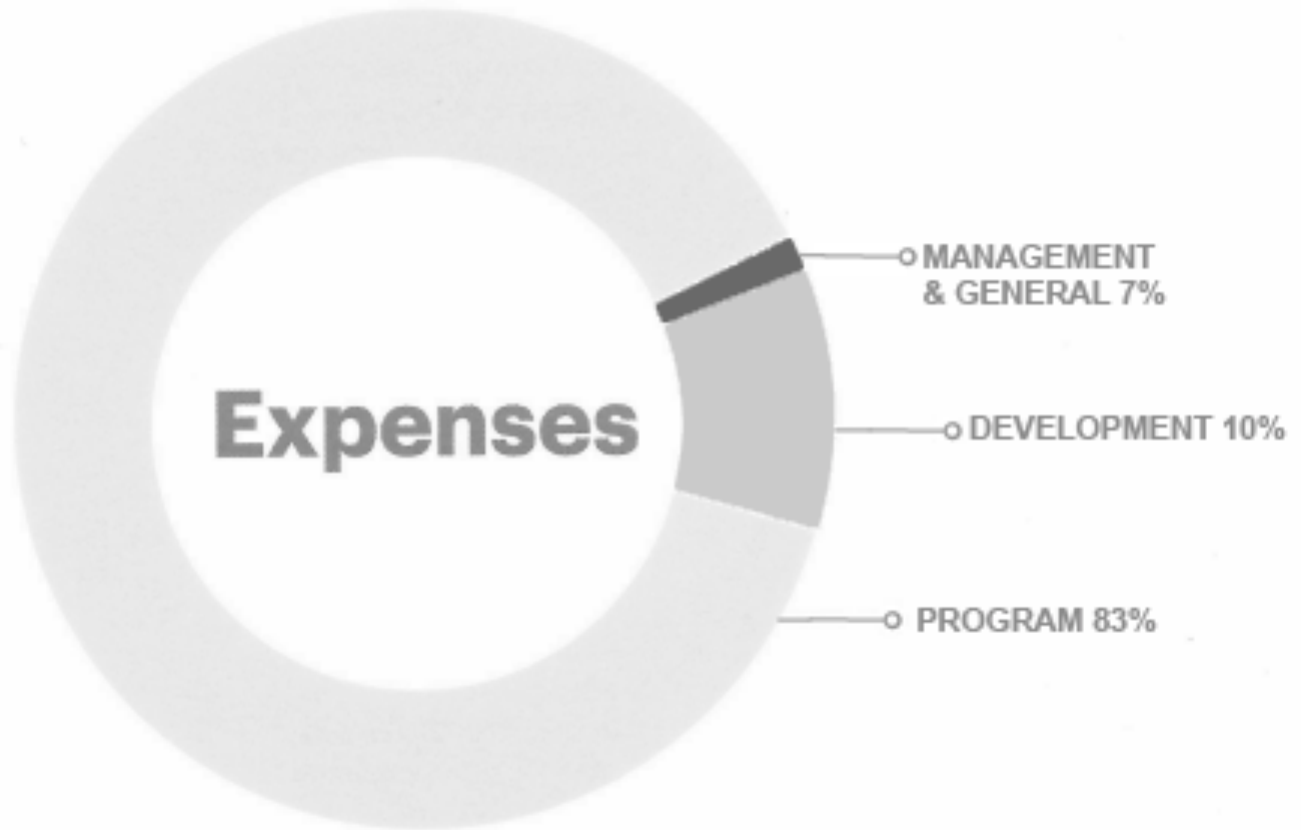
Instructions on the hand-out

If you would rather try an exercise in Excel,  
see instructions for  
creating simple graphs with Excel's Pivot  
Tables and Tate's artist data

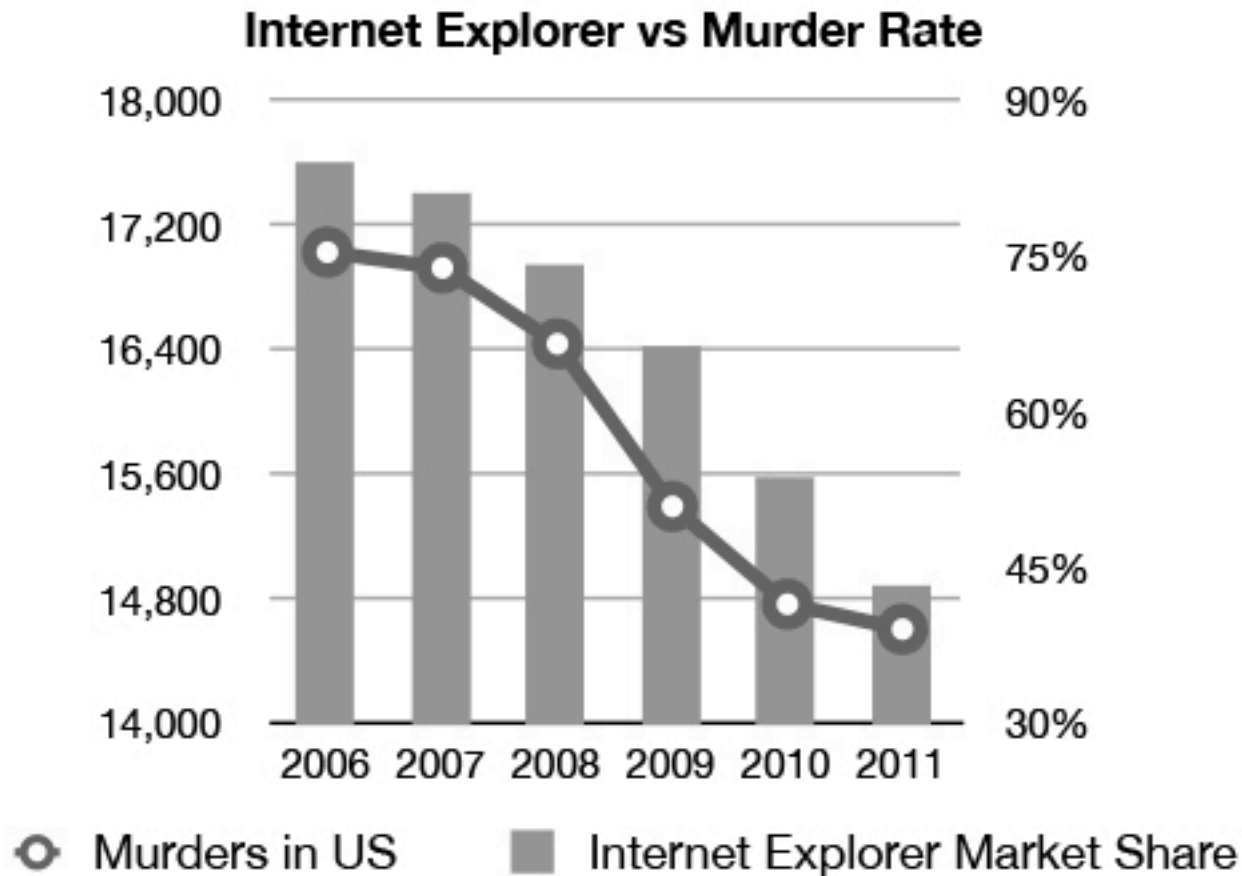


**DESIGNING VISUALISATIONS**

# Worst practice in data visualisations



# Worst practice in data visualisations



# Best practice for design

How effectively does the visualisation support cognitive tasks?

The most important and frequent visual queries/pattern finding should be supported with the most visually distinct objects

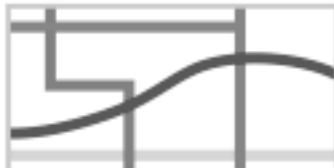
# Visually distinct objects













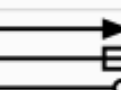

Colour (hue, lightness)

Elementary shape (orientation, size, elongation)

Motion

Spatial grouping

**Points****Lines****Areas****Show****Shape***possible, but too weird to show**cartogram**qualitative differences***Size***cartogram**quantitative differences***Color Hue***qualitative differences***Color Value***quantitative differences***Color Intensity***qualitative differences***Texture***qualitative & quantitative differences*

	position, placement	yes	infinite	Good	Good	Good	Good
2, 3; A, B, C	text labels	optional (alphabetical or numbered)	infinite	Good	Good	Good	Good
	length	yes	many	Good	Good		
	size, area	yes	many	Good	Good		
	angle	yes	medium/few	Good	Good		
	pattern density	yes	few	Good	Good		
	weight, boldness	yes	few		Good		
	saturation, brightness	yes	few		Good		
	color	no	few (< 20)			Good	
	shape, icon	no	medium			Good	
	pattern texture	no	medium			Good	
	enclosure, connection	no	infinite			Good	Good
	line pattern	no	few				Good
	line endings	no	few				Good
	line weight	yes	few		Good		

# Dealing with complex data

Find a visualisation type that can harbour the data in a meaningful way or reduce the data in a meaningful way.

- e.g. go from individual values to distribution of values
- e.g. introduce interaction: overview, zoom and filter, details on demand

Ben Shneiderman



# Do you really need a visualisation:

Use tables when:

- doc will be used to look up individual values
- to compare individual values
- precise values are required
- the quantitative info to be communicated involves more than one unit of measure

Use graphs when:

- the message is contained in the shape of the values
- the document will be used to reveal relationships among values

# Publishing visualisations

How can you contextualise, explain any limitations of your visualisations? e.g.

- provenance and qualities of original dataset;
- what you needed to do to it to get it into software (how transformed, how cleaned);
- what's left out of the visualisation, and why?

Tools that don't require programming

Excel

Viewshare

Google Fusion Tables, Google Drive

IBM Many Eyes

Tableau Public

# Exercise 6: geocoding data and creating a map

Instructions on the hand-out

# Review: planning a visualisation

With a dataset in mind, consider...

Exploratory or explanatory? Static or dynamic? Small- or large-scale?

Choose a type of visualisation (map, timeline, chart, etc)

- Is your dataset in a suitable format for your visualisation type? How can you clean it?
- Is more cleaning or transformation needed? You may need to iterate with different versions of your data

# If all else fails...

Sketch out your visualisation on paper to test it

Iteration is key, and...

Stubbornness is a virtue!

# Exercise /: taking things further

Instructions on the hand-out

# Review: visualisation tools

Any data cleaning tips?

What did you learn about the data?

What did the tool do well? Poorly?

Were the tool and the data a good match for each other?

What other data could you link to?



# Thank you!

Mia Ridge, Open University

<http://miaridge.com>

@mia\_out