

COS30045

DATA VISUALISATION

TOPIC 01: INTRODUCTION TO DATA VISUALISATION PART 1

Commonwealth of Australia
Copyright Act 1968

Notice for paragraph 135ZXA (a) of the *Copyright Act 1968*

Warning

This material has been reproduced and communicated to you by or on behalf of Swinburne University of Technology under Part VB of the *Copyright Act 1968* (the *Act*).

The material in this communication may be subject to copyright under the *Act*. Any further reproduction or communication of this material by you may be the subject of copyright protection under the *Act*.

Do not remove this notice.

IMAGES USED IN THIS PRESENTATION: FROM MICROSOFT CLIPART (UNLESS OTHERWISE SPECIFIED)

Acknowledgment: This unit is partially based on the Visualisation unit developed by Alex Lex and Hanspeter Hfister (with permission).

Source of images used in this lecture available as links in the PDF version.



INTRODUCTION TO DATA VISUALISATION

What:

- ▶ definition and goals of data visualisation, brief history

Why:

- ▶ big data, limits of statistics and human cognition

How:

- ▶ understand visual perception, become familiar with the design space, represent data accurately



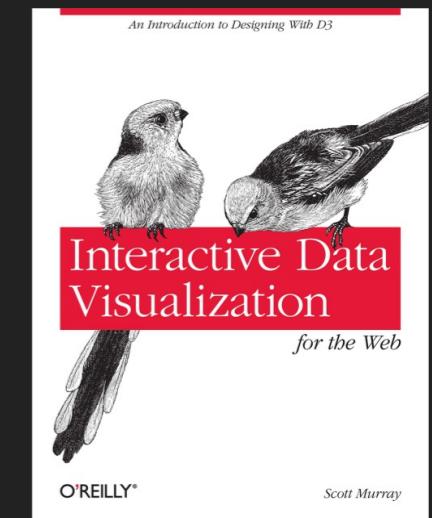
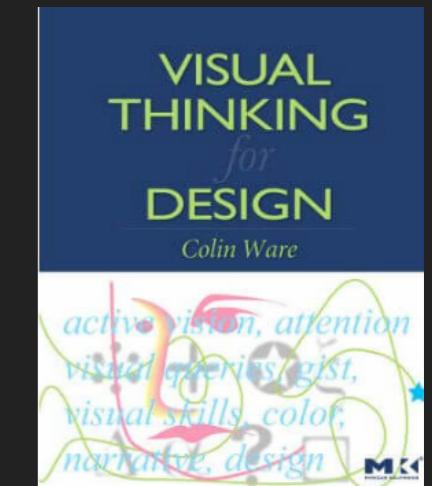
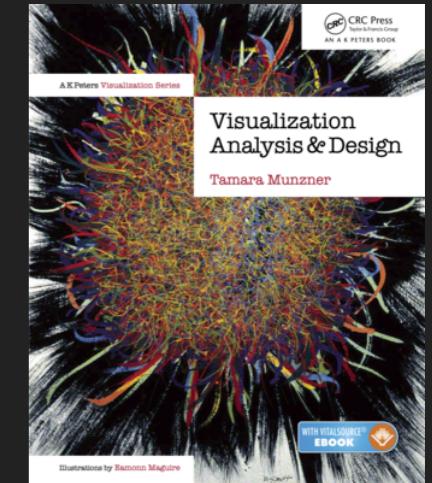
INTRODUCTION TO DATA VISUALISATION

Lecture Readings

- ▶ VAD Ch 1
- ▶ VTD Ch 1

Tutorial

- ▶ IDV Ch 3



WHAT IS VISUALISATION?

- ▶ The formation of a mental image of something (Oxford Dictionary)
- ▶ visualisation as something humans do (i.e, an **internal** representation of something)



WHAT IS VISUALISATION?

- ▶ The representation of an object, situation, or set of information as a chart or other image (Oxford Dictionary)
- ▶ visualisation as something outside the human (i.e, an **external** representation of something)



WHAT IS DATA VISUALISATION?

- ▶ (Data) visualisation is the process of **transforming** (abstract) data into **interactive graphical representations** for the purpose of **exploration, confirmation, or presentation.**
- ▶ To convey information through graphical representations of data



WHY VISUALISE DATA?

- ▶ Humans have limited memory and attentional resources
- ▶ External cognition
 - ▶ using elements outside of our head to help us think



Visualization is really about external cognition, that is, how resources outside the mind can be used to boost the cognitive capabilities of the mind.

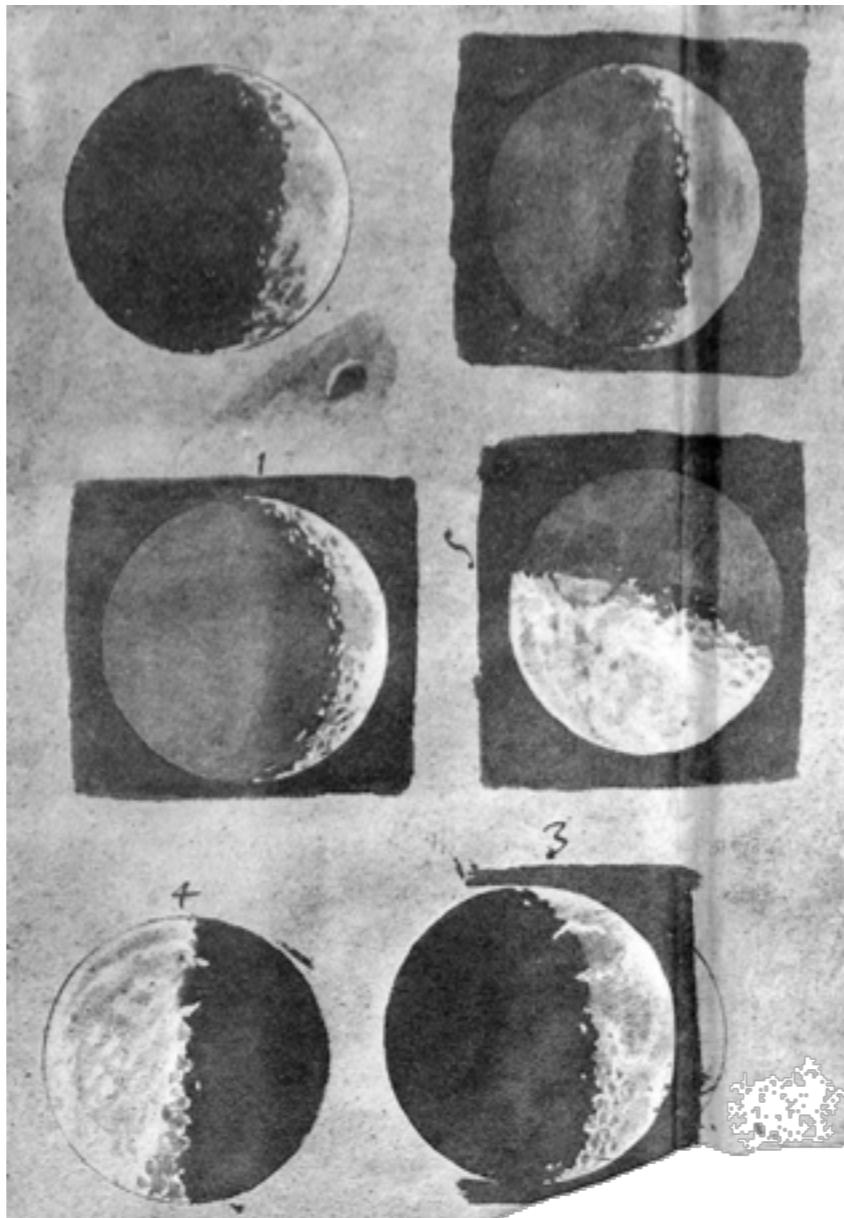


Stuart Card
Pioneering Human-
Computer Interaction
(HCI) researcher

PRESENT INFORMATION



Leonardo da Vinci, circa 1500's



Galileo Galilei, 1616



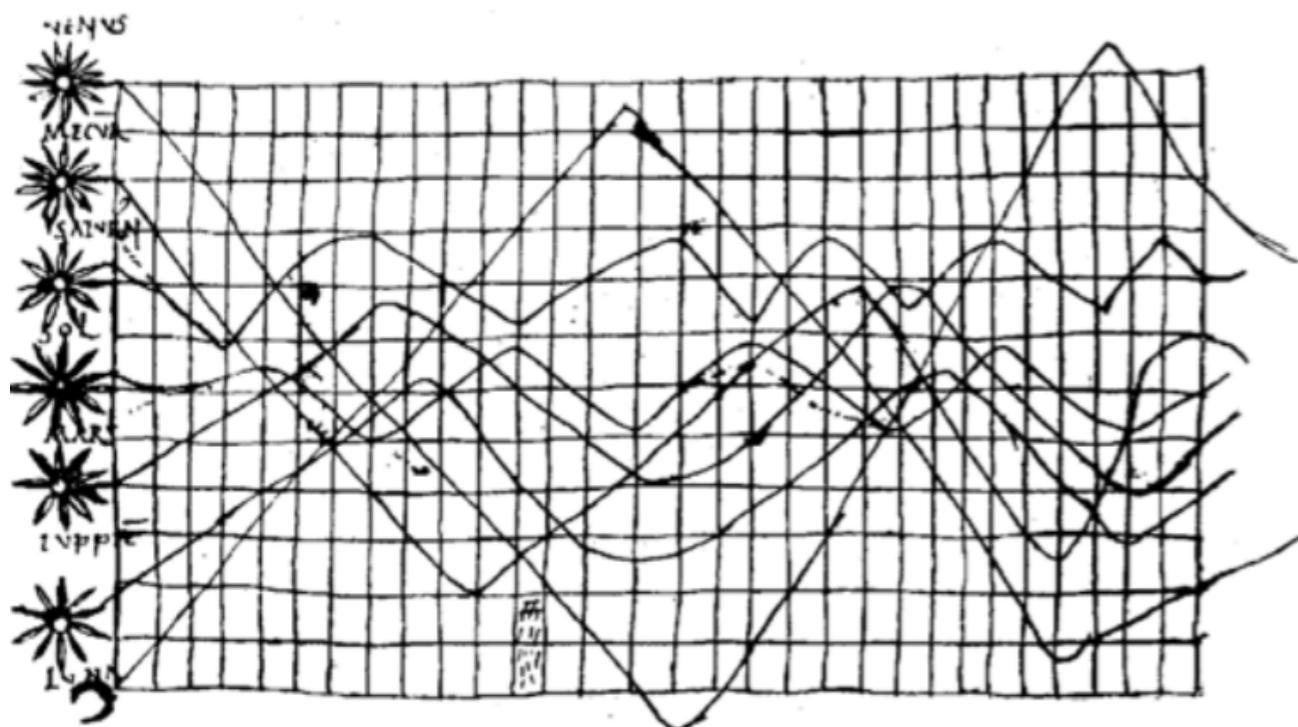
William Curtis (1746-1799)

PRESENT INFORMATION

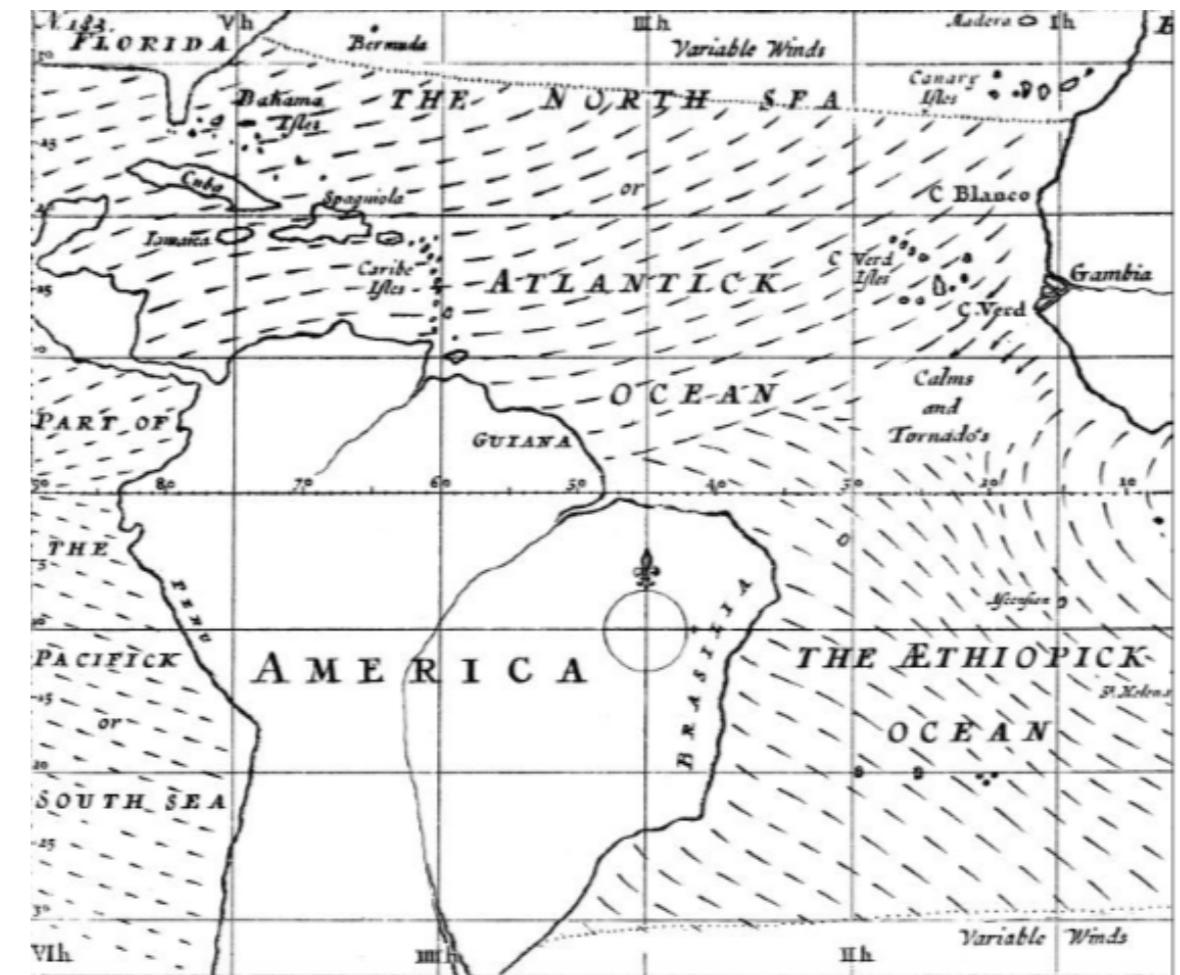


Konya town map, Turkey, c. 6200 BC

ANALYSE INFORMATION: DISCOVER AND PRESENT



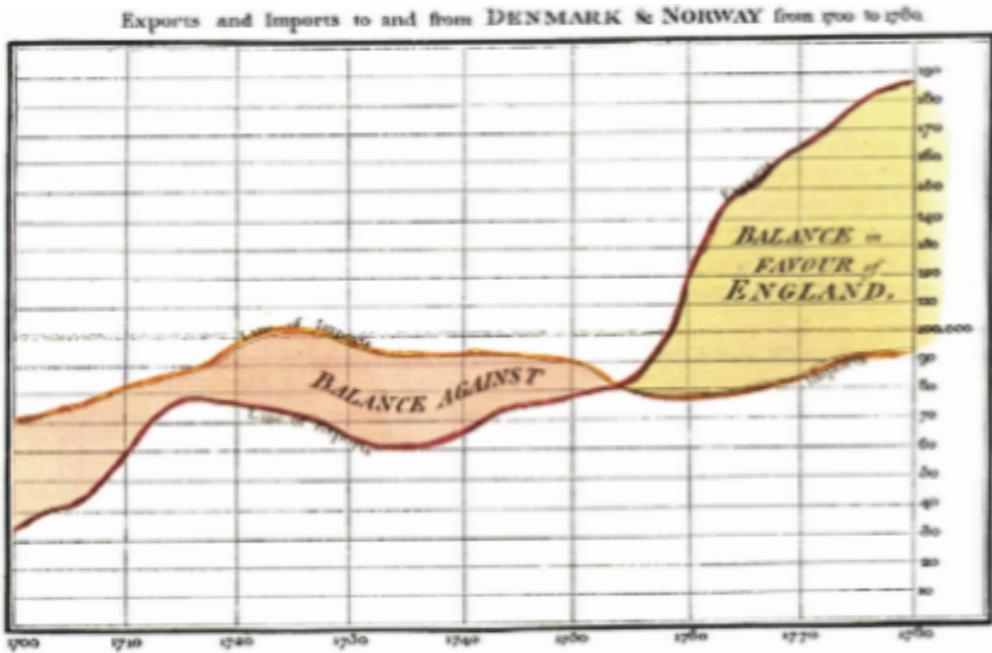
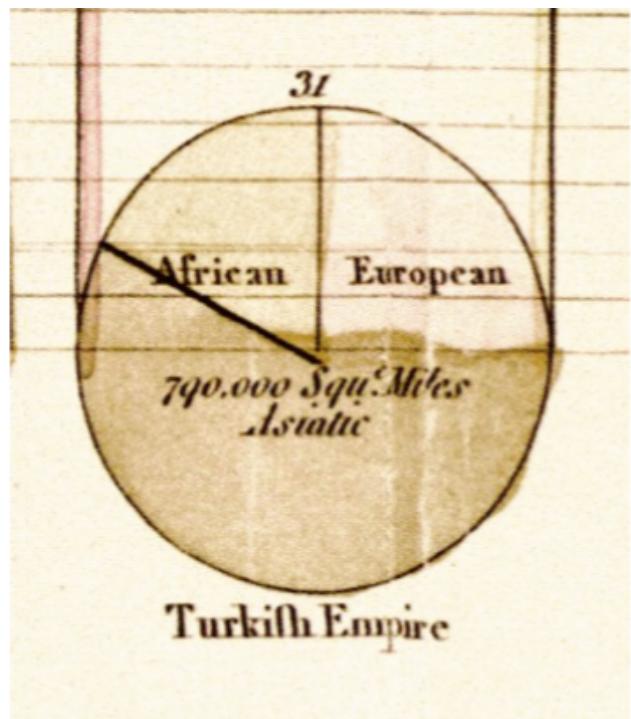
Planetary Movement Diagram, c. 950



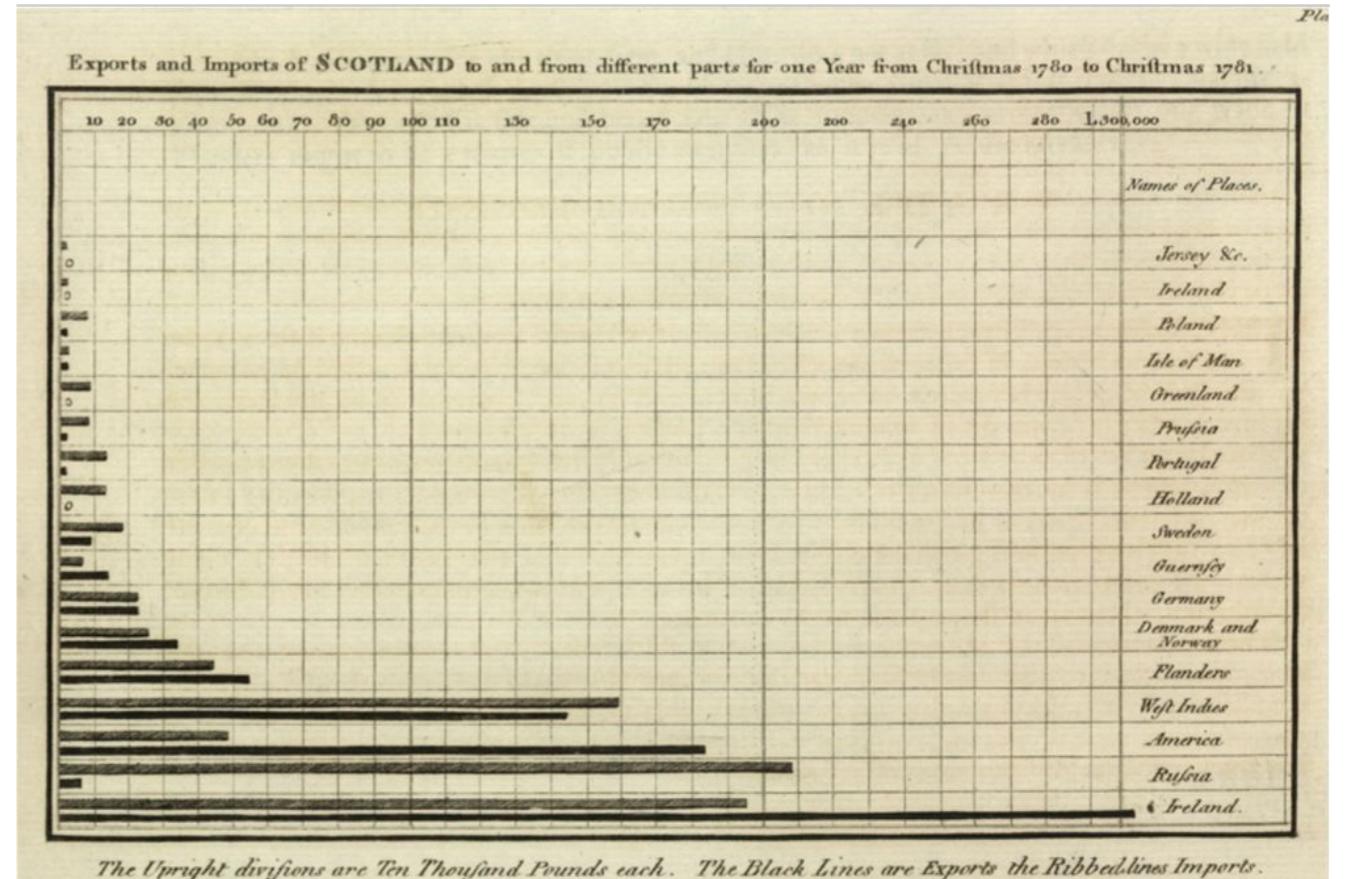
Halley's Wind Map, 1686

ANALYSE INFORMATION: DISCOVER AND PRESENT

First Pie Charts



First Bar Charts

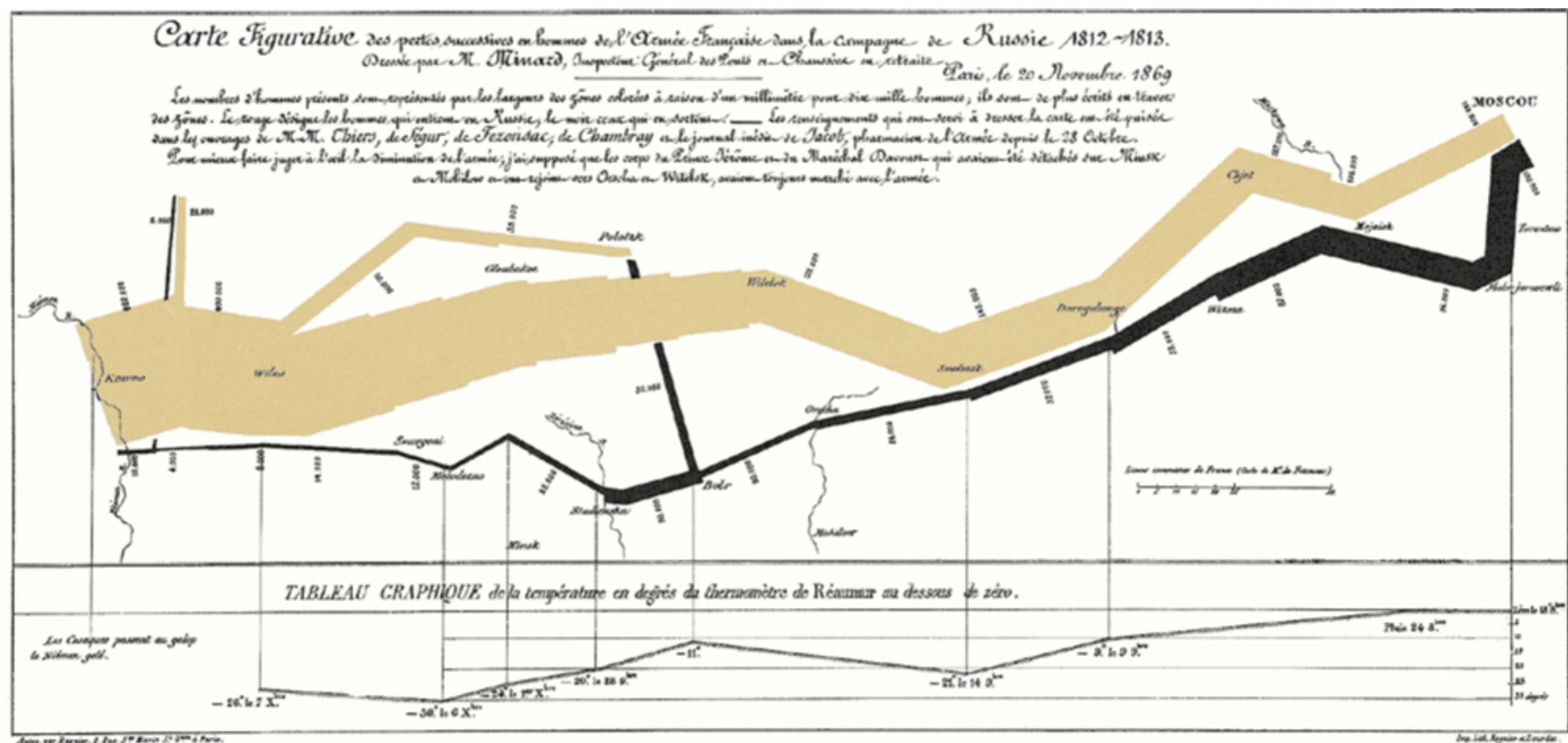


William Playfair
Scottish Engineer and
political economist, 1780's

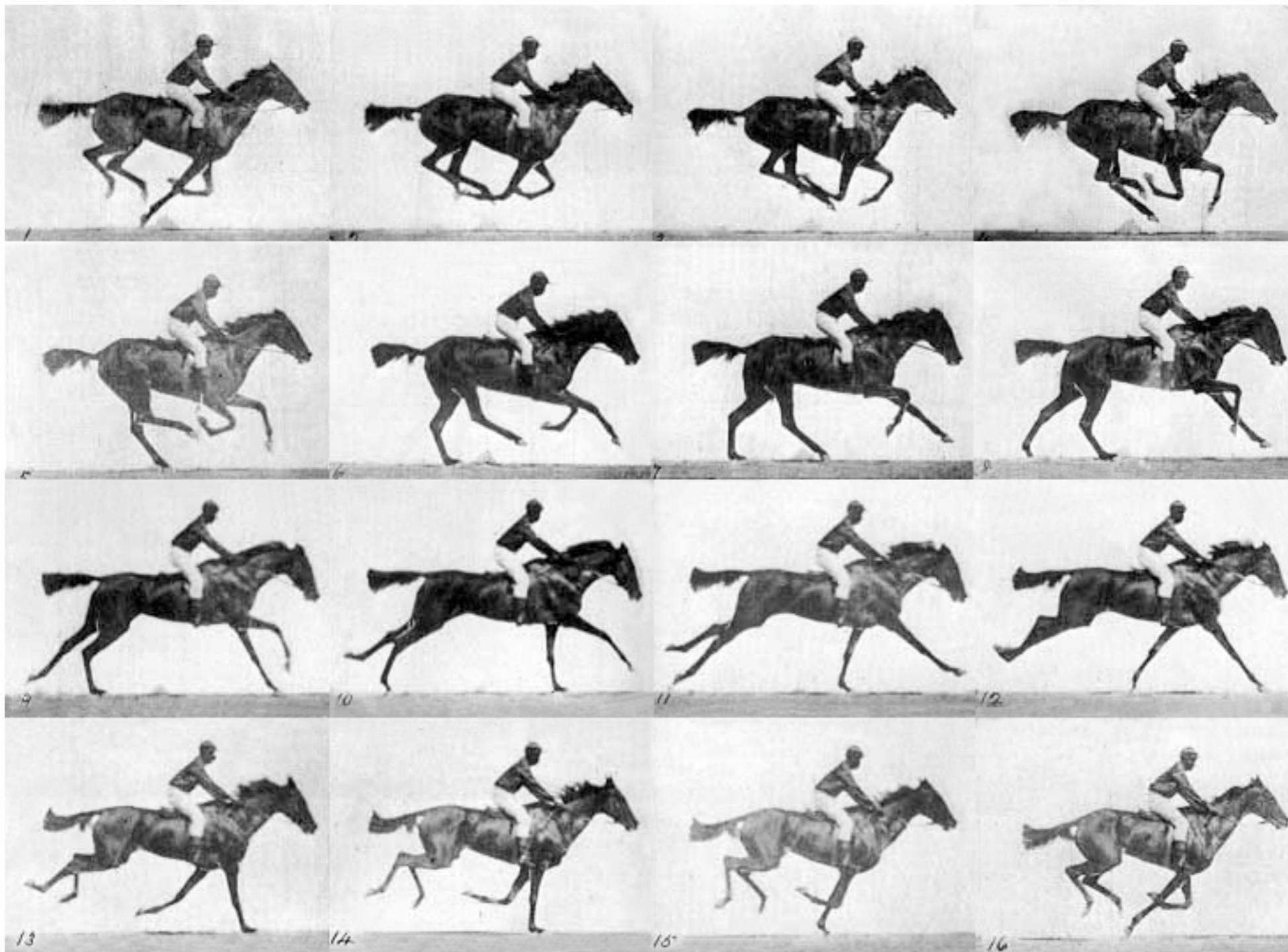


PRESENT INFORMATION

Losses in the French army in the Russian campaign (1812-1813) by C. R. Minard (1869)



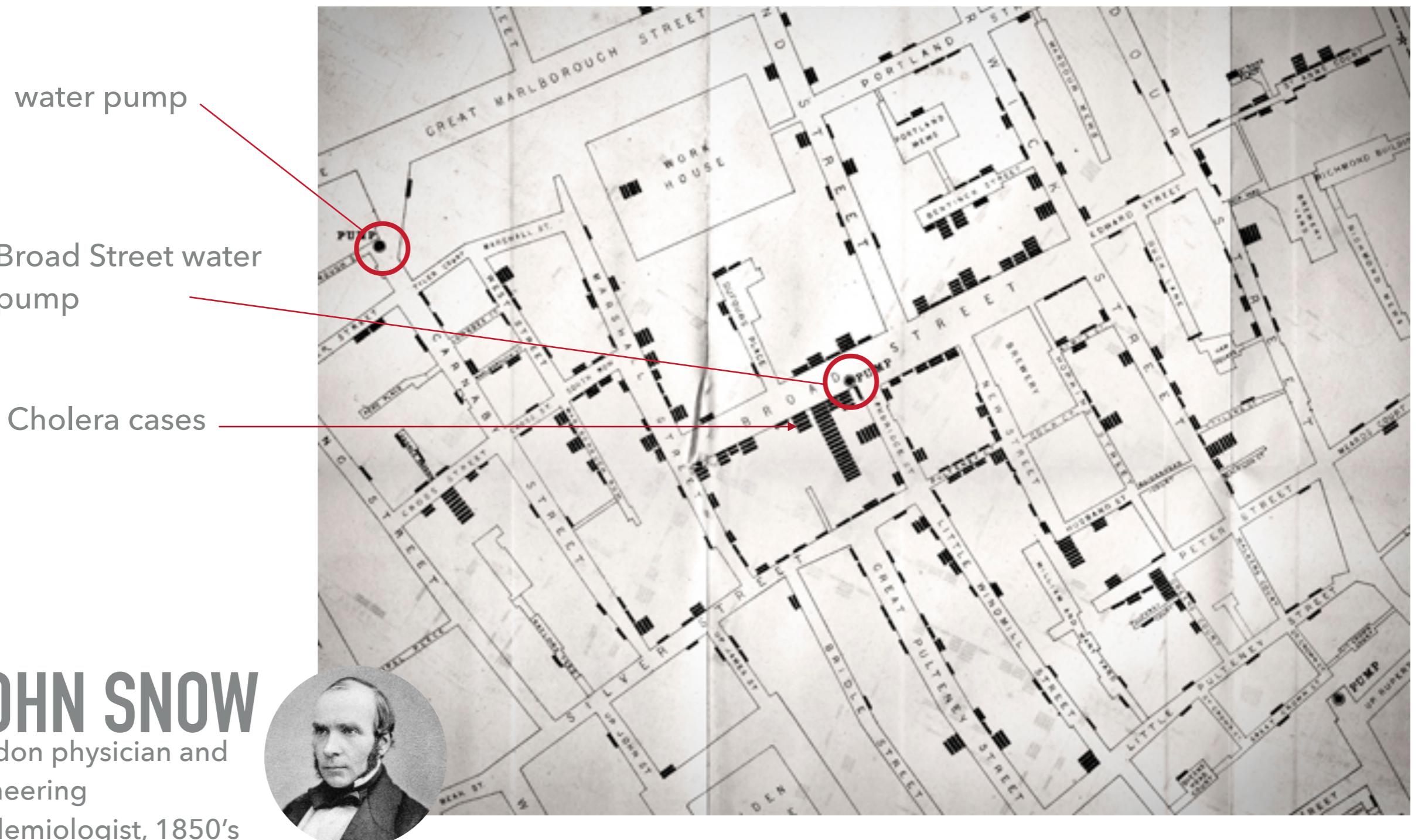
ANALYSE INFORMATION: DISCOVER AND PRESENT



E. J. Muybridge, 1878

Watch the video: https://www.youtube.com/watch?v=VJ86D_DtyWg

ANALYSE INFORMATION: DISCOVER AND PRESENT



JOHN SNOW
London physician and
pioneering
epidemiologist, 1850's



COMMUNICATION

2015 - There were nearly 13,000 gun Homicides in USA, but not everyone was affected equally . . .

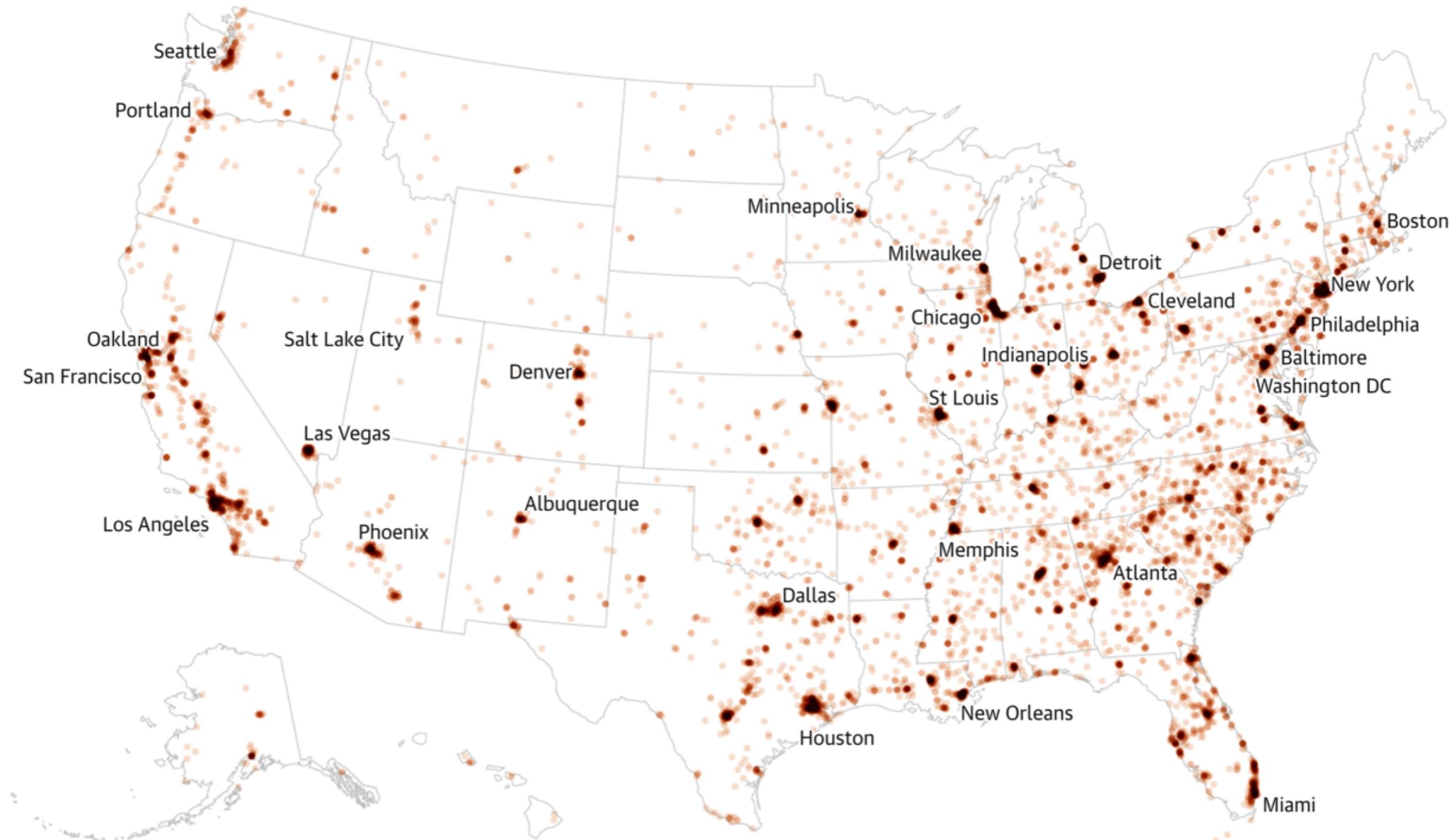


North St Louis is home to blocks of abandoned and dilapidated homes. Houses like this one, where the rear wall has collapsed, are known locally as 'dollhouses'. Photograph: Sid Hastings for the Guardian

PRESENT INFORMATION

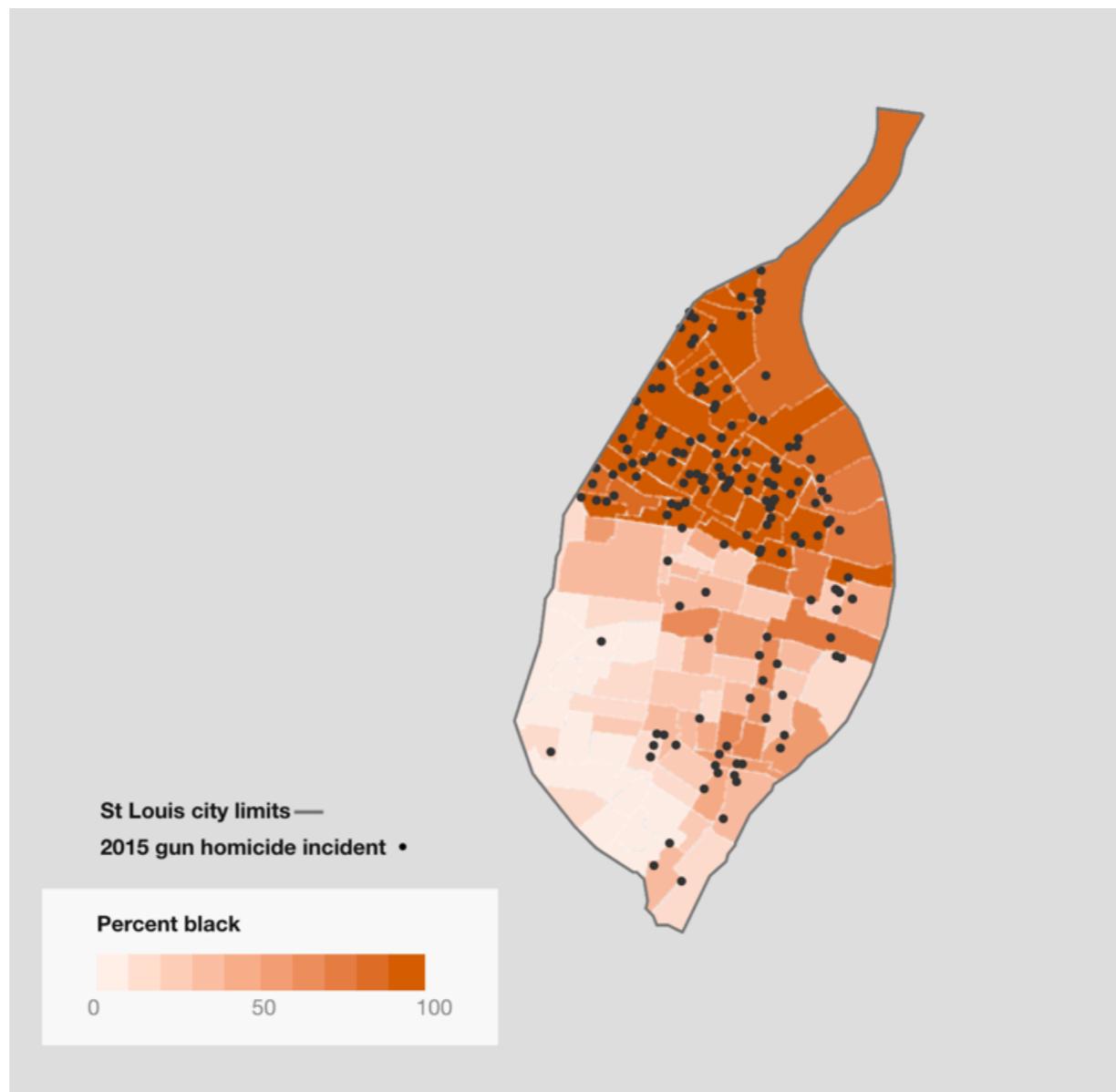
Gun Murders in USA

The Guardian



PRESENT INFORMATION

2015 St Louis (USA) Gun Murders



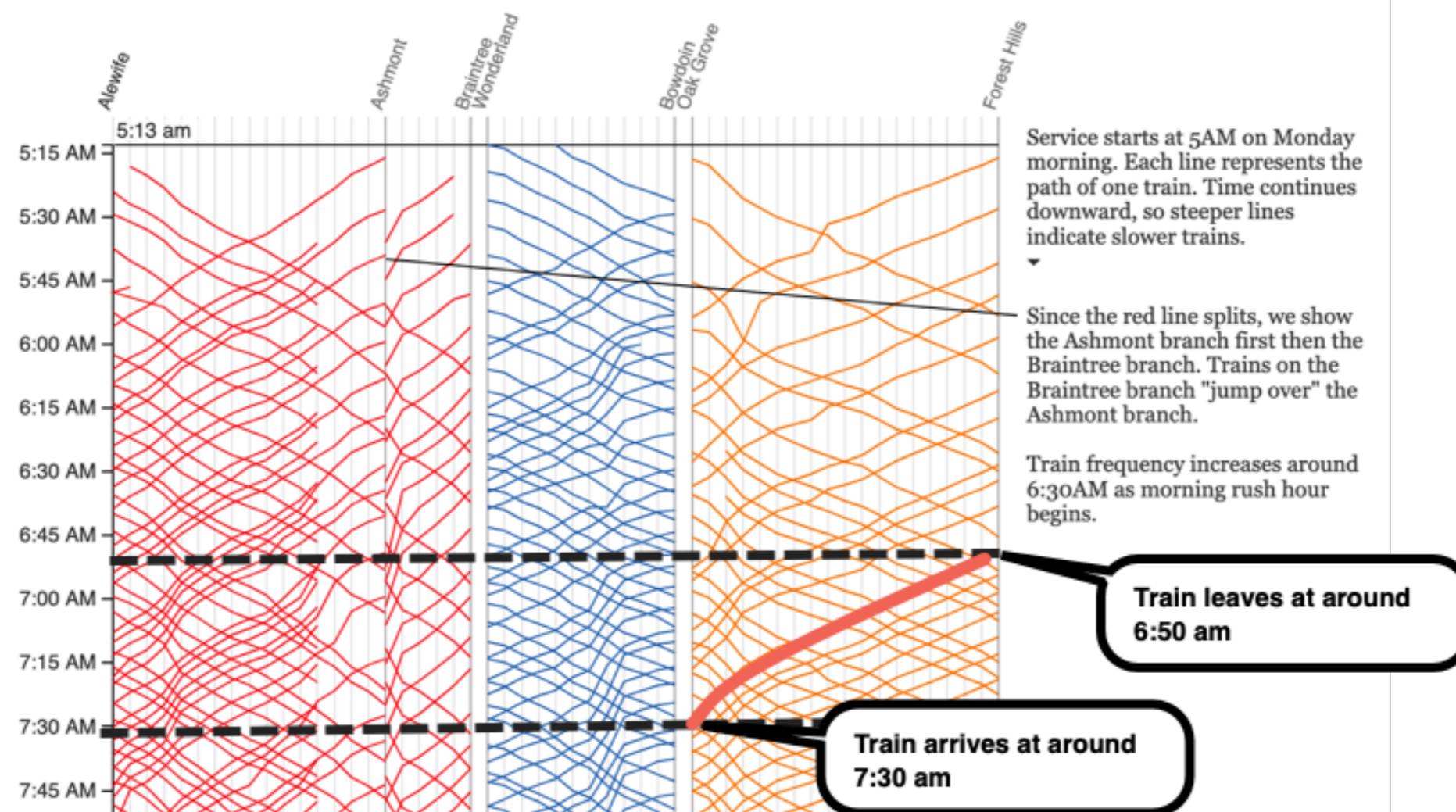
COS30045

DATA VISUALISATION

TOPIC 01.1: INTRODUCTION TO DATA VISUALISATION PART 2

EXPLORE AND DISCOVER

- ▶ Singapore's MRT Circle Line train disruptions...
- ▶ Marey chart, plots time against distance



EXPLORE AND DISCOVER

- ▶ Singapore's MRT Circle Line train disruptions what was causing 'random' breakdowns?
- ▶ plot of location and time of breakdowns

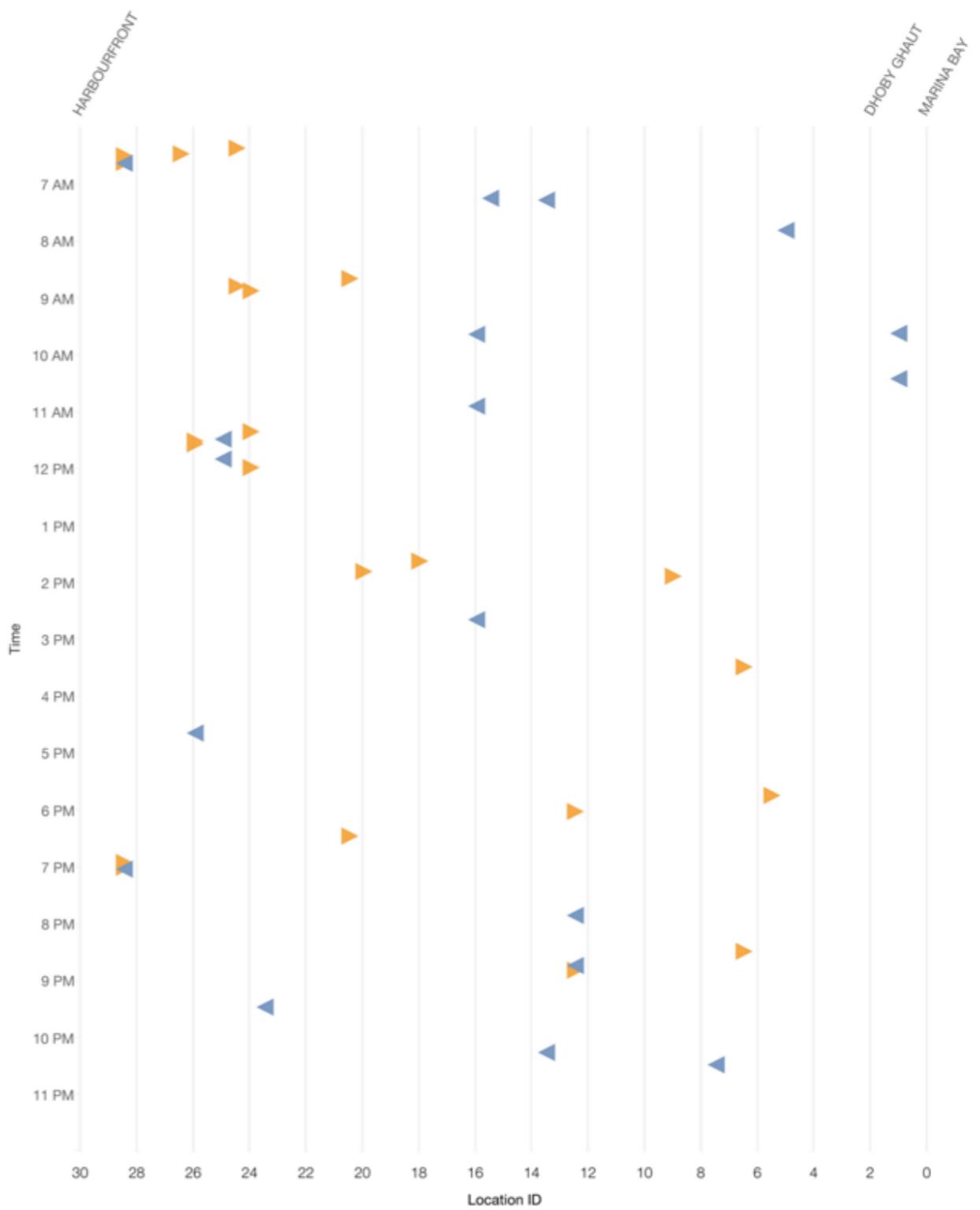


Figure 7: Direction is represented by arrows and colour.

EXPLORE AND DISCOVER

- ▶ Singapore's MRT Circle Line train disruptions...
- ▶ visual inspection to find patterns

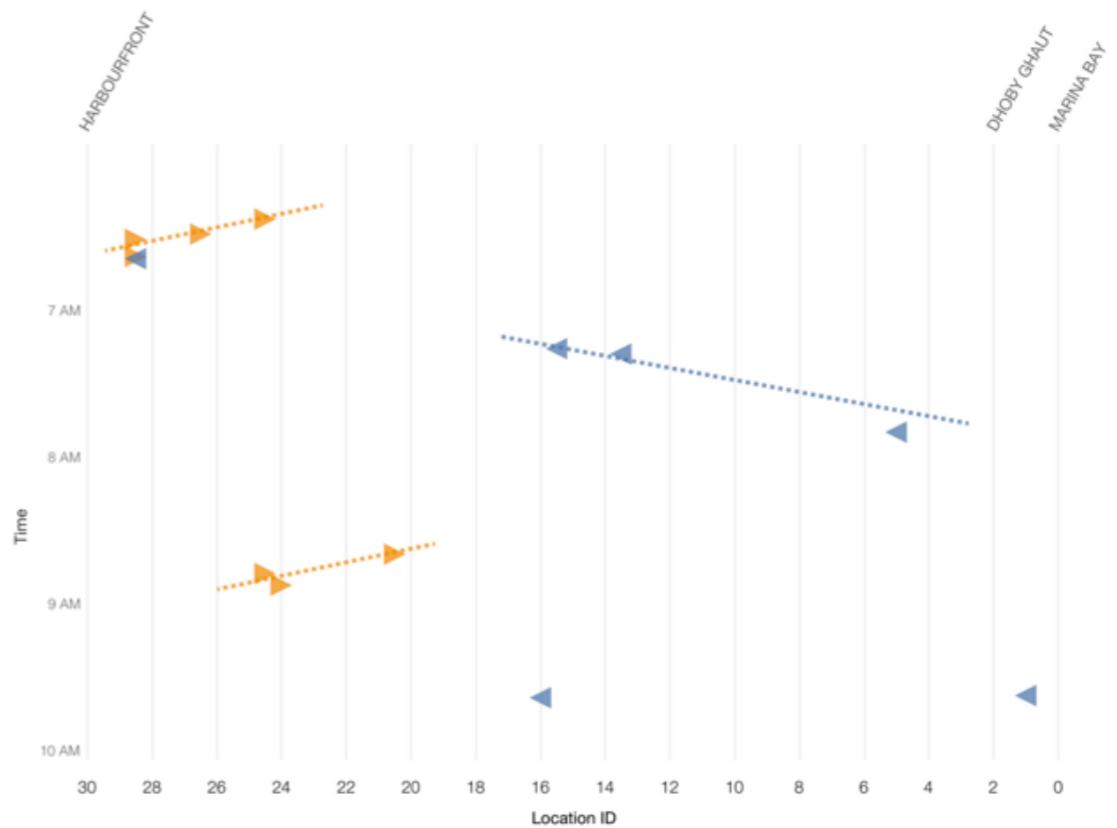


Figure 8: Incidents between 6am and 10am

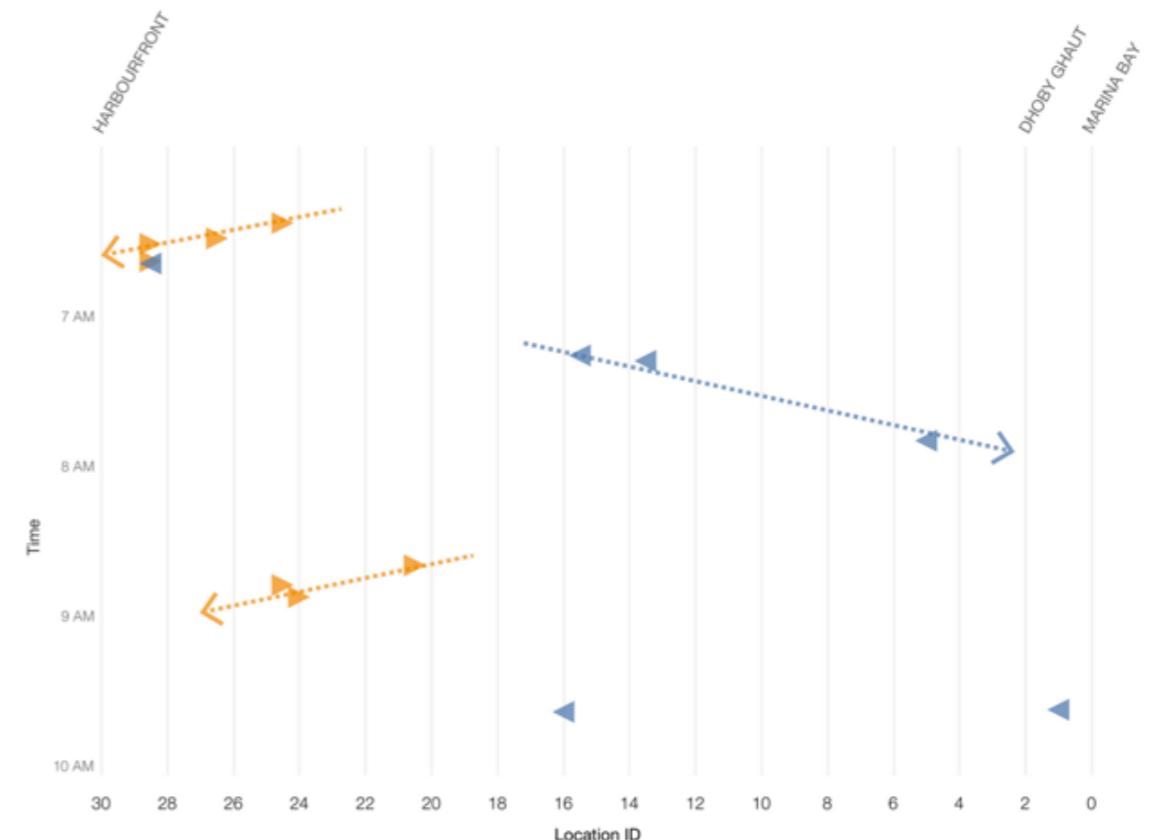


Figure 9: Could it be a train moving in the opposite direction?

EXPLORE AND DISCOVER

- ▶ Singapore's MRT Circle Line train disruptions...
- ▶ one rogue train?

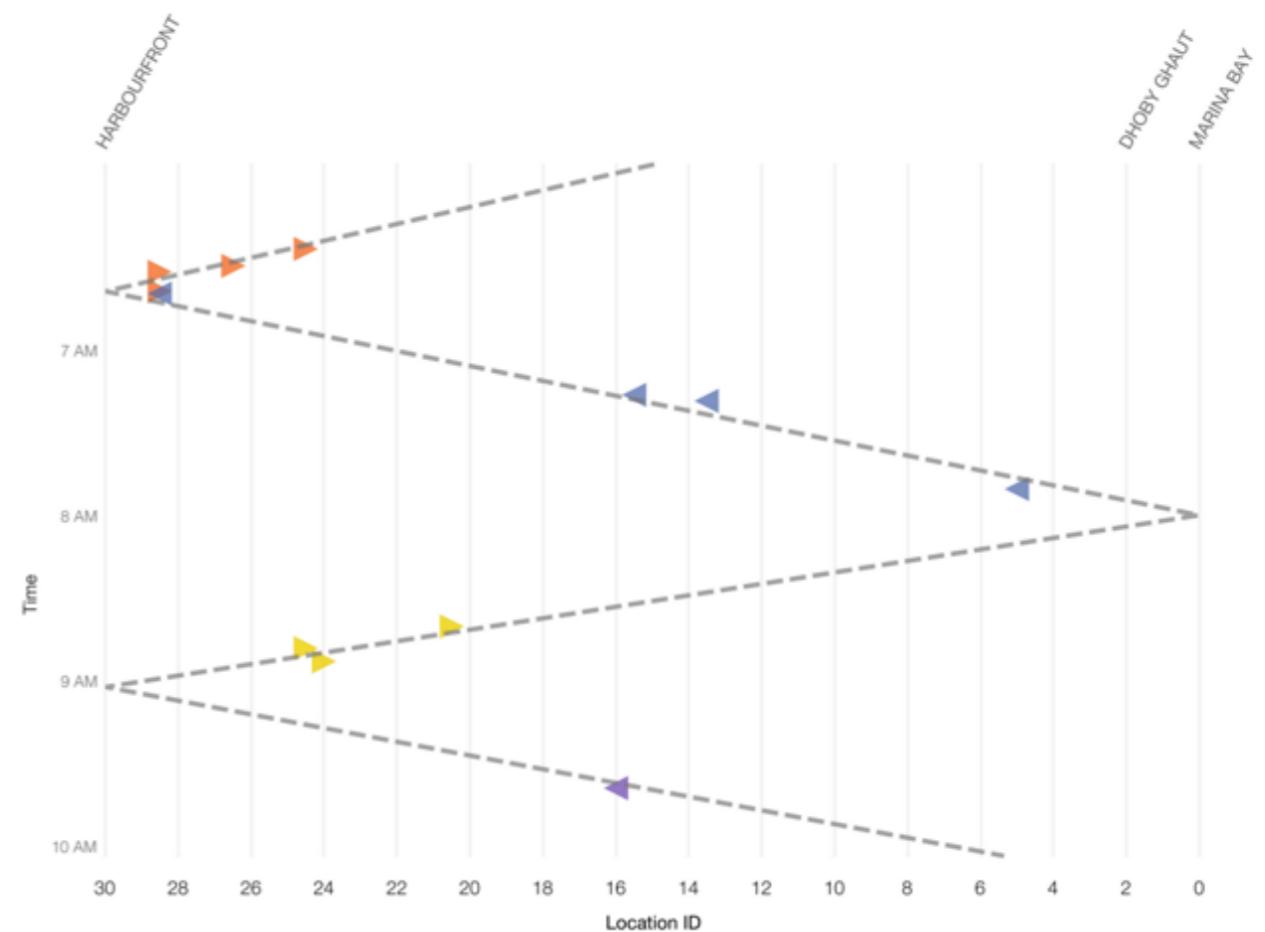
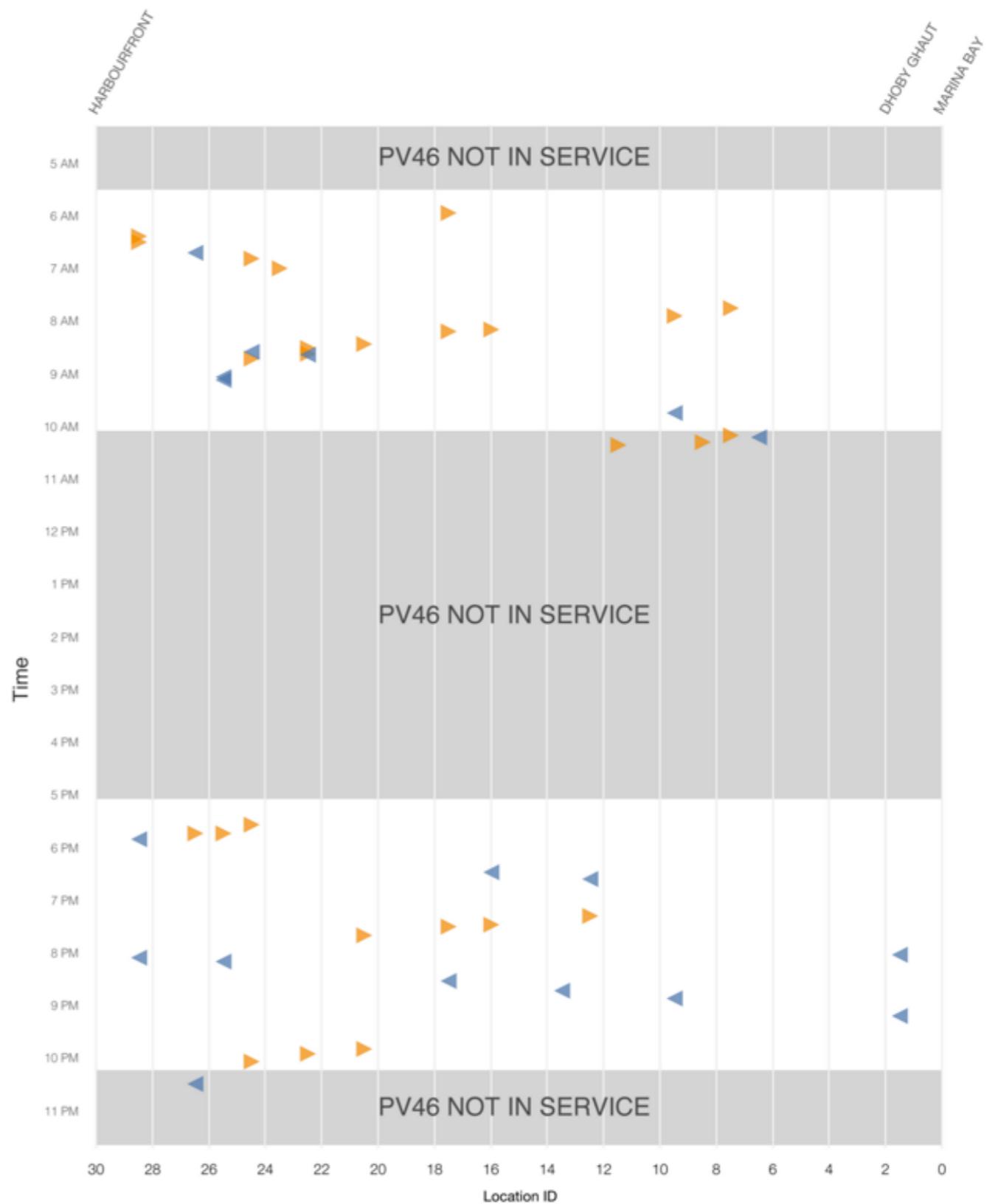


Figure 11: Time of clustered incidents strongly implies that the interference could be linked a single train

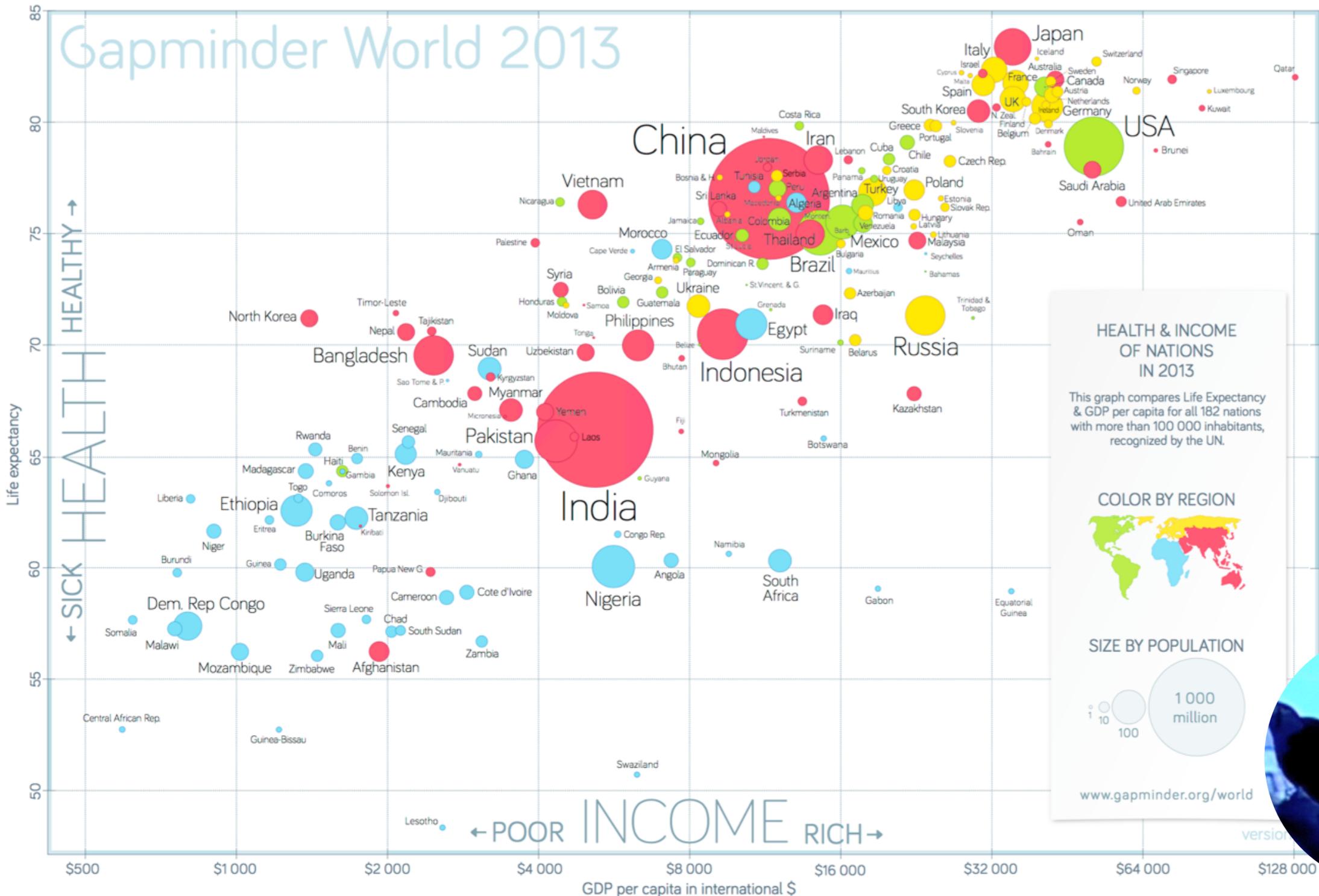
EXPLORE AND DISCOVER

- ▶ Singapore's MRT Circle Line train disruptions...
- ▶ Testing the rouge train



Watch the TED talk: <https://www.youtube.com/watch?v=hVimVzgtD6w>
up to 6:18

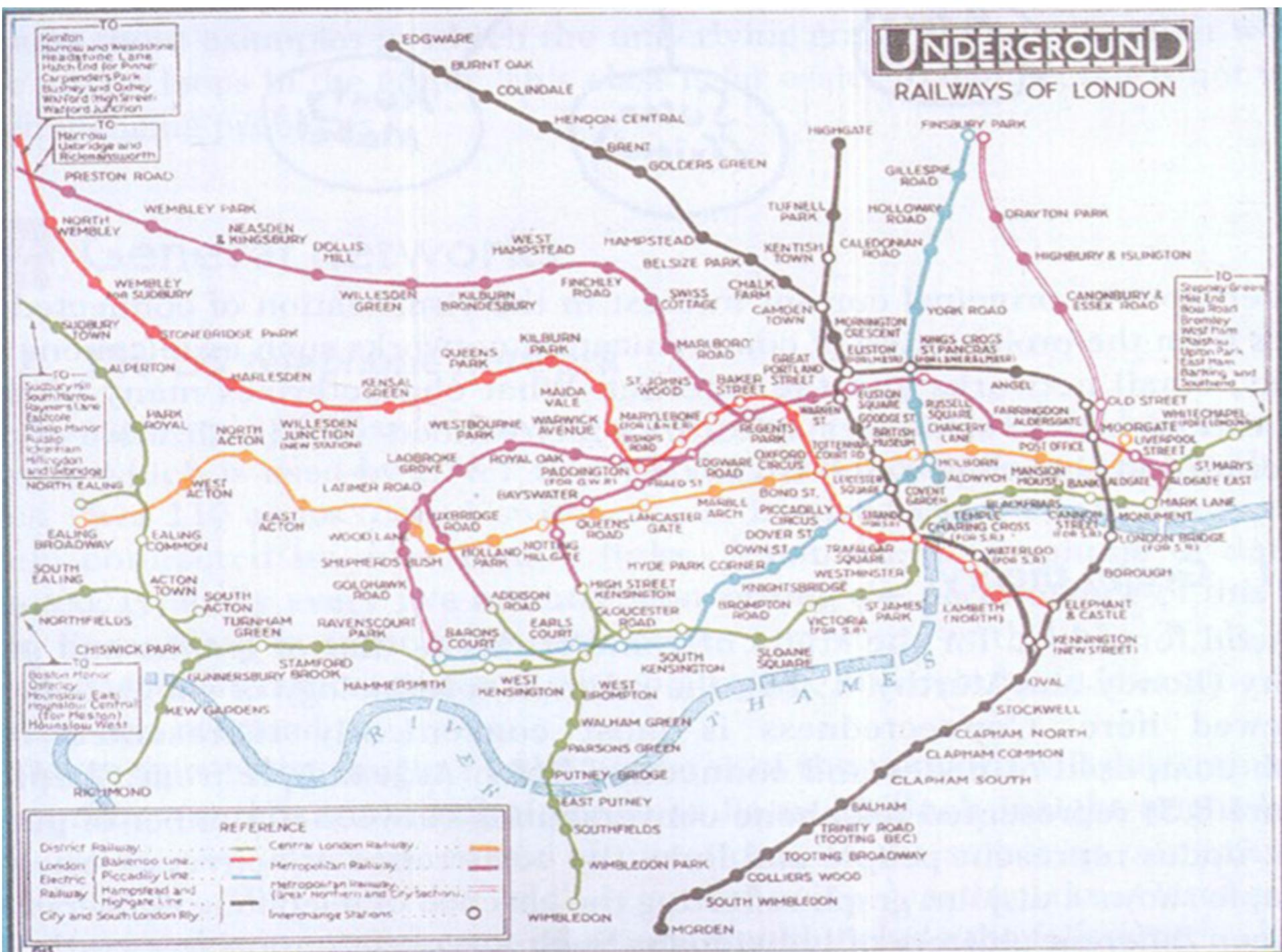
PRESENT INFORMATION



HANS ROSLING

Swedish physician and
statistician

PRESENT INFORMATION



London Subway Map (1927)

PRESENT INFORMATION



Henry Beck's London Subway Map (1933)



Photographic images
used in this presentation:

mostly from Microsoft Clipart circa 2009
(unless otherwise specified)
Note references are linked to images in PDF
version of this presentation.

COS30045

DATA VISUALISATION

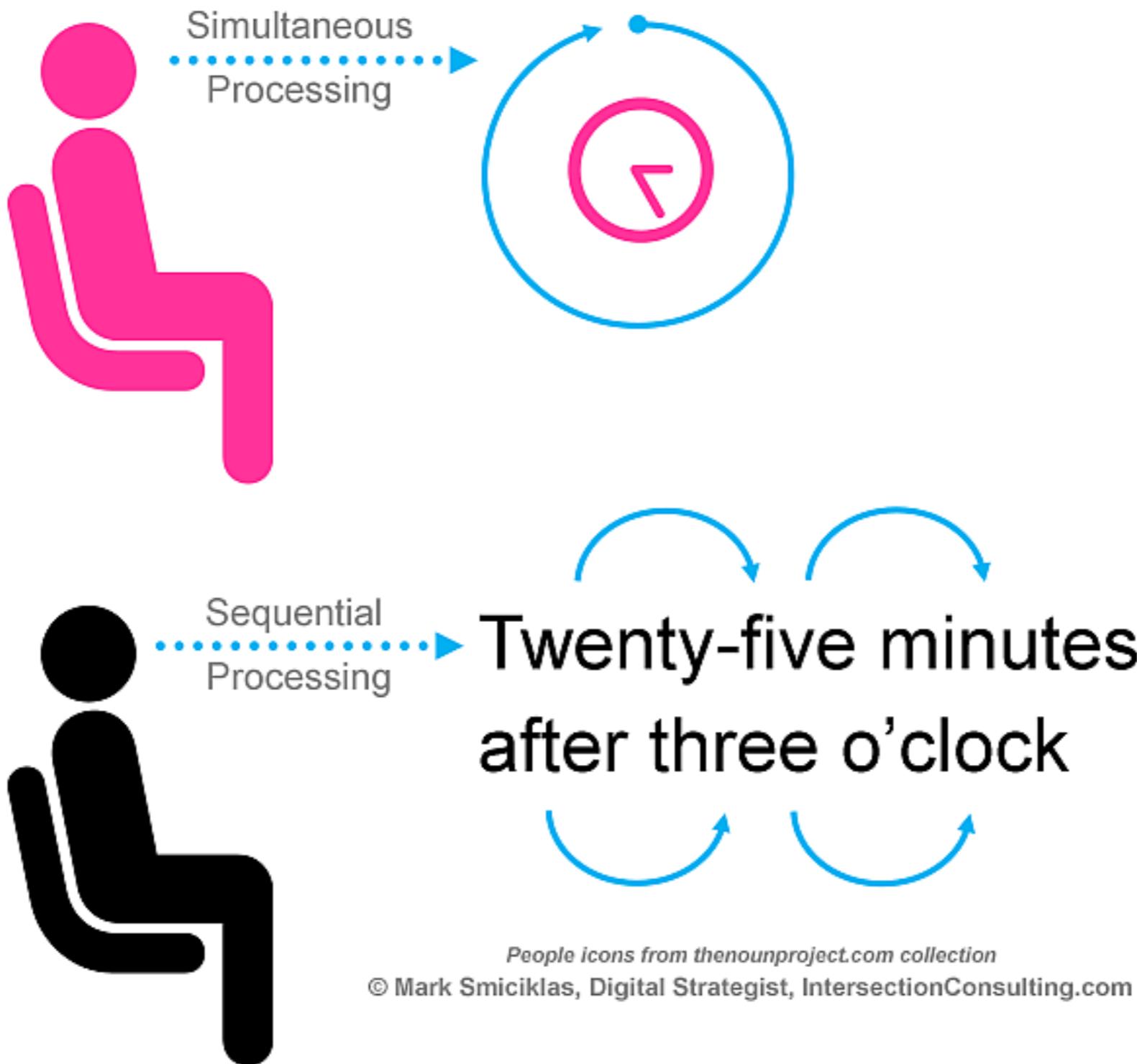
TOPIC 01.2: INTRODUCTION TO DATA VISUALISATION – WHY DATA VIS?

WHY USE VISION?

- ▶ Very high-bandwidth channel
- ▶ Parallel processing of information (pre-attentive/'sub-conscious')
- ▶ Sound processed sequentially and merged over time
- ▶ Currently no good ways of producing and displaying taste, touch and smell

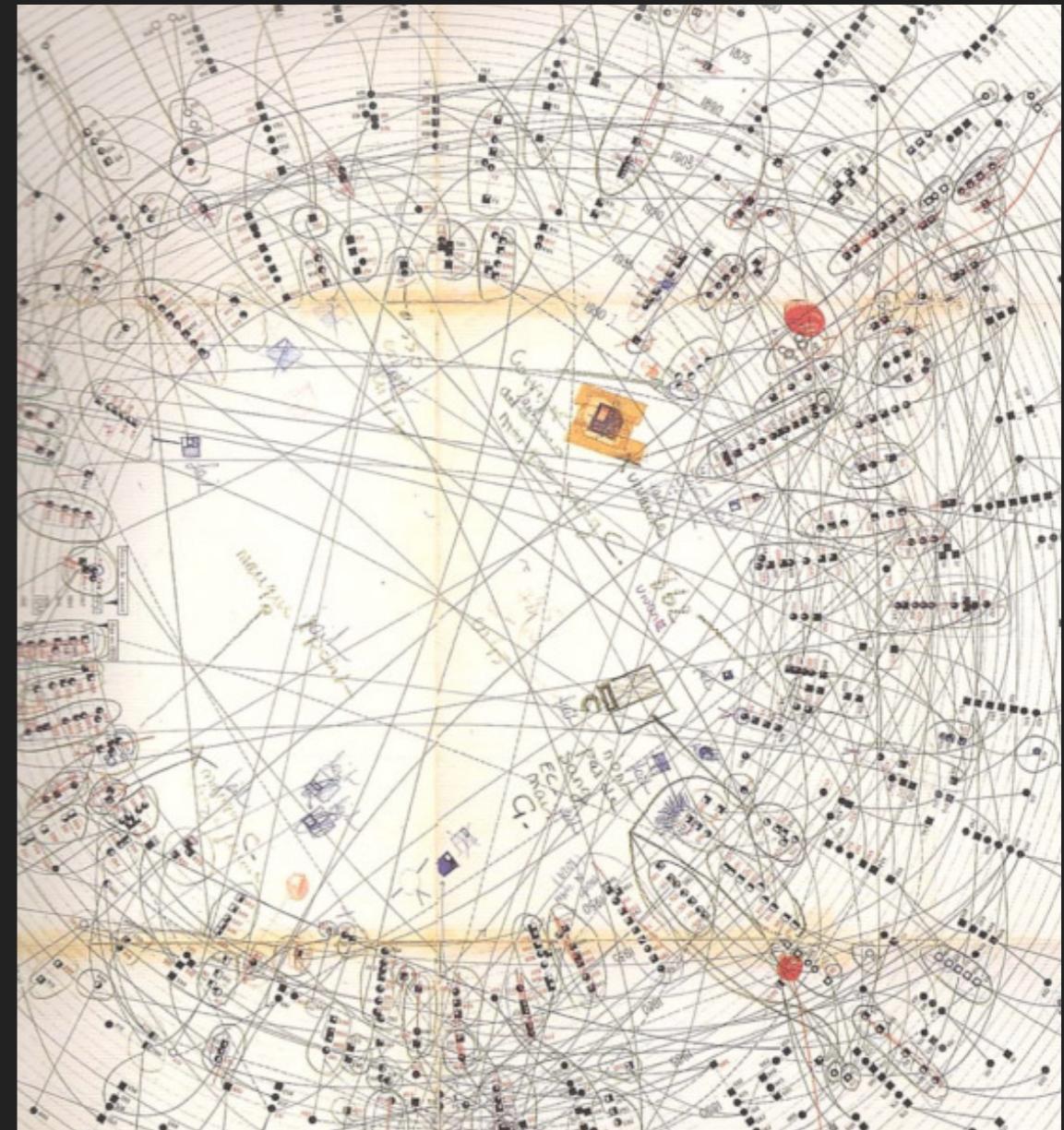


WHY USE GRAPHICS?



WHY USE COMPUTERS?

- ▶ Hand drawn (or Illustrator)
 - ▶ Time consuming to create when data sets very large (infeasible)
 - ▶ Difficult to update data set (inflexible)
 - ▶ Static
- ▶ Computer program
 - ▶ Able to deal with large data sets
 - ▶ Easy to update and reuse (efficiency)
 - ▶ Interactive



Hand drawn genealogical diagram from mid 1950's

WHY USE COMPUTERS? BIG DATA!!

Google Reader (1000+)

tinyURL mss-01 Google Reader keybr.com - Be prod... CourseCalendar - HF... Archive in Yojimbo Bookmark in Yojimbo Google Scholar last.fm hpfister@gmail.com | Office | Seminars | My Account | Help | Sign out

Home All Items (1000+) Starred Items Trends Your shared items Friends' shared items (70) Add subscription Discover Show updated - all Creativity (39) CreativeIQ How to Change the World kottke.org Obscure Design Design (42) Because We Can, Blogs beyond bulletins Cool Hunting (4) dezen (2) NOTCOT.ORG (24) Presentation Zen

A look at what's new Data Mining: Text Mining, Visualization and Social Media (2) Wired - The Secret Life of a Blog Post TechMeme - Beyond the List Calvin and Hobbes (Unofficial) (2) daily dose of imagery (2)

Subtraction (2) The Host with the Most Bought and Paid For Calvin and Hobbes for January 26, 2008 Calvin and Hobbes for January 25, 2008 mozart's house corner streetcar

See more from Visualization (3) See more from WebDesign (3) See more from Mac (4) See more from Mac (4)

See more from tips (2)

Top Recommendations View all JMLR Lightroom-Blog.Com decor8 Recently started How to Shave Ten Hours Off Your Work Week from Where I Sit Sensor Size and Depth of Field from O'Reilly Digital Media Blog

See more from Visualization (3) See more from WebDesign (3) See more from Mac (4) See more from Mac (4)

See more from tips (2)

See more from Blogger? If so, it's even easier! Add to Blogger

What are you doing? 140

hpfister 14 following 0 followers 0 updates

Latest: Check out this SlideShare Presentation : Doooold http://tinyurl.com/3rttme about 1 month ago

update

guykawasaki Ancestor For All An http://adix.com/awh9 Also see about 1 hour ago from Adix

guykawasaki RT @szetela: screen Contextual Advertising sometime /1812f Kawasaki Steel! about 1 hour ago

guykawasaki RT @roehmholdt: F mantra belief: http://bit.ly/fWZzE about 1 hour ago

timoreilly RT @bobgourley: Blog Security and Cyber Conflict http://hour ago from bobgourley

guykawasaki Will Check-up Char professionals http://adix.com/tr http://gaming.alltop.com about 1 hour ago

presentationzen My new Starbucks http://tinyurl.com/3tq4a about 1 hour ago

Home

Google youtube cat videos

Web Videos Shopping Images News More Search tools

About 593,000,000 results (0.44 seconds)

TOP 10 BEST CAT VIDEOS OF ALL TIME! - YouTube www.youtube.com/watch?v=... YouTube Sep 6, 2012 - Uploaded by WatchTheDaily We've scoured the internet and found the cutest and funniest cat videos of all time. Any we missed? Let us ...

Image:
Hanspeter Pfister

Wikipedia.org/

WIKIPEDIA

English Free Encyclopedia 1 162 000+ articles

Deutsch Die freie Enzyklopädie 667 000+ Artikel

Polski Wolna encyklopedia 463 000+ haseł

Neder De vrije encyclopedie 480 000+ artikelen

Português A encyclopédia livre

facebook Barack Obama for President in 2008 Global

Information Group Info Name: Barack Obama for President in 2008 Type: Student Groups - Political Groups

Students for Barack Obama

View Discussion Board Join this Group Report Group Share +

HP & AL

Digg All News, Videos, & Images

Look beneath the surface of Digg

Barack Obama wins South Carolina Democratic primary

Driver Who Killed Teen Sues for Damaged Vehicle

The naked truth about women in Russian politics

The Terrible Secret Behind the World's Greatest Card Trick

How to Reverse a Car Out of a Moving Plane

New Jersey Souche Bag!

Brussels airport ETER (F1Q)

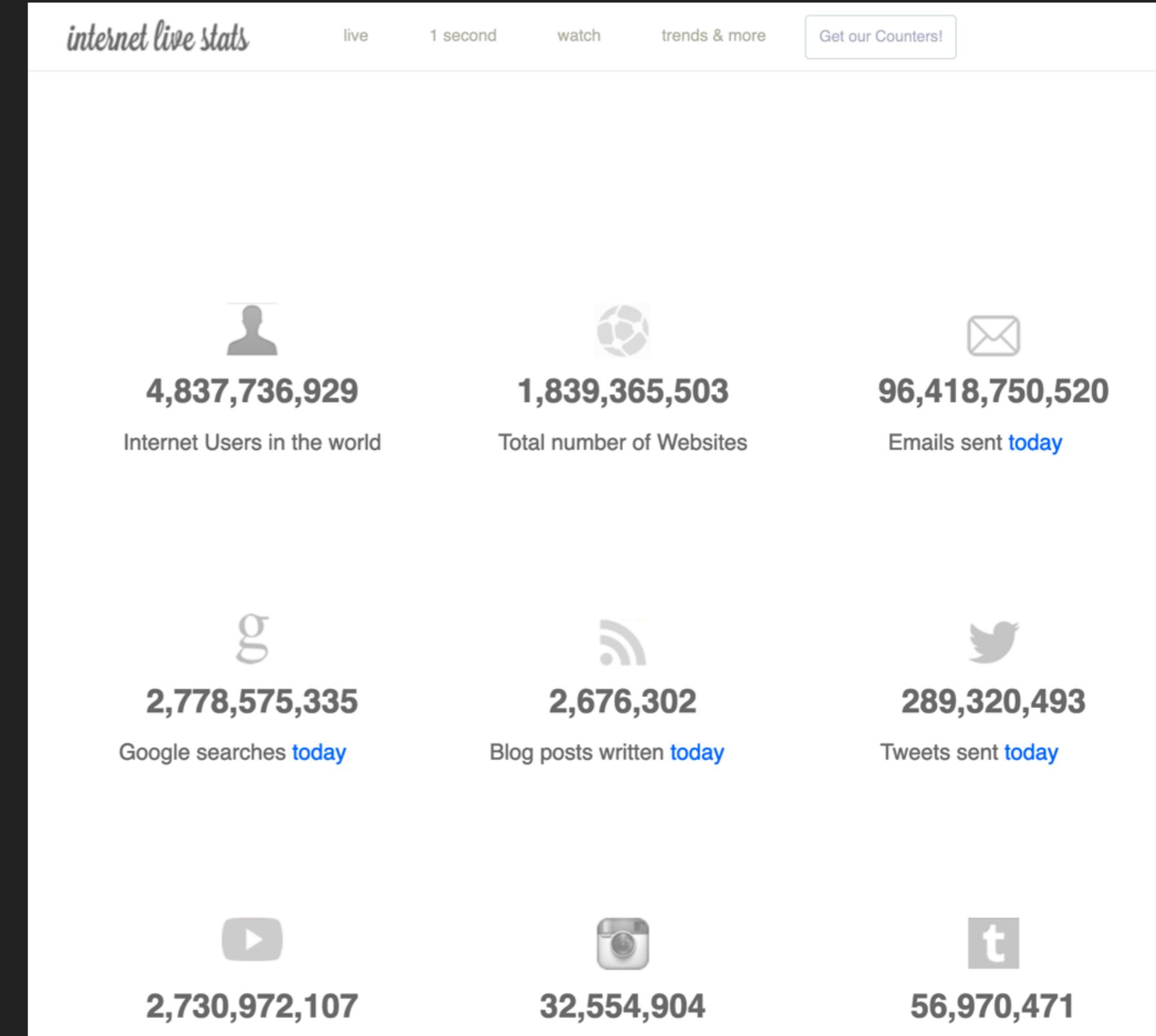
But how? The only clues were the pictures on the car...

Sam Macheane Written Short as Arms Strike Negotiation

UNIDENTIFIED: Oppose Telecom Immunity and Contact Your Senator

Science: Build First Man-Made

BIG DATA



<https://www.internetlivestats.com/>

WHY USE COMPUTERS?

- ▶ Interaction
- ▶ Exploring large data sets:
 - ▶ Overview First
 - ▶ Zoom and Filter
 - ▶ Details on demand

Shneiderman's mantra (1996)



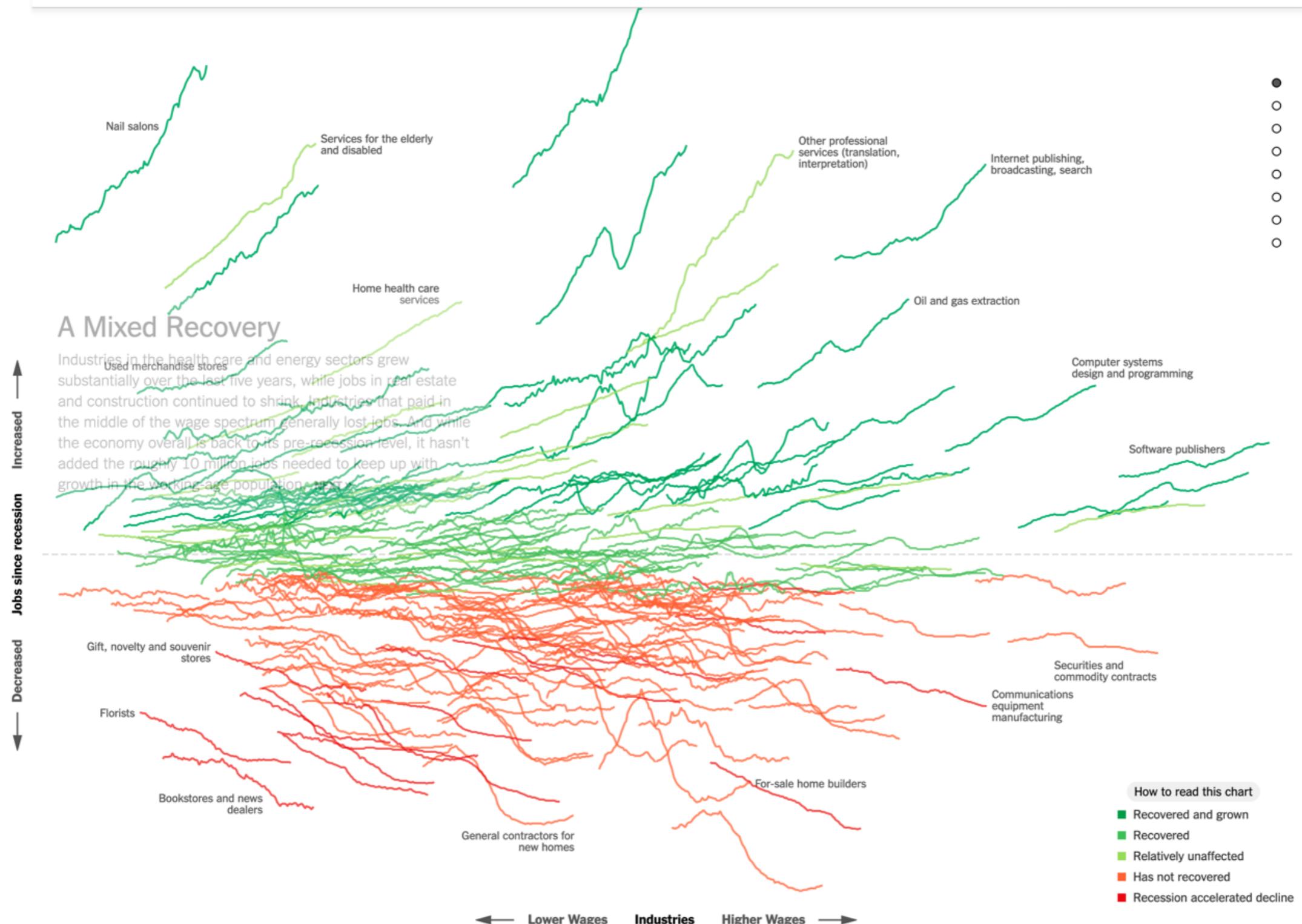
IVAN SUTHERLAND

Pioneering computer
graphics researcher

Sketchpad 1960's

WHY USE COMPUTERS?

THE UPSHOT | How the Recession Reshaped the Economy, in 255 Charts



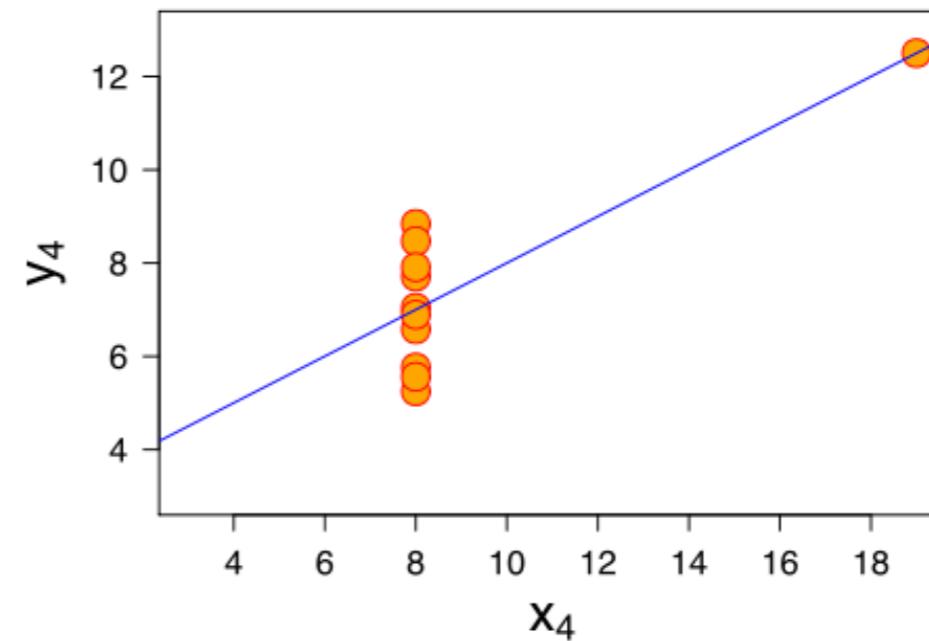
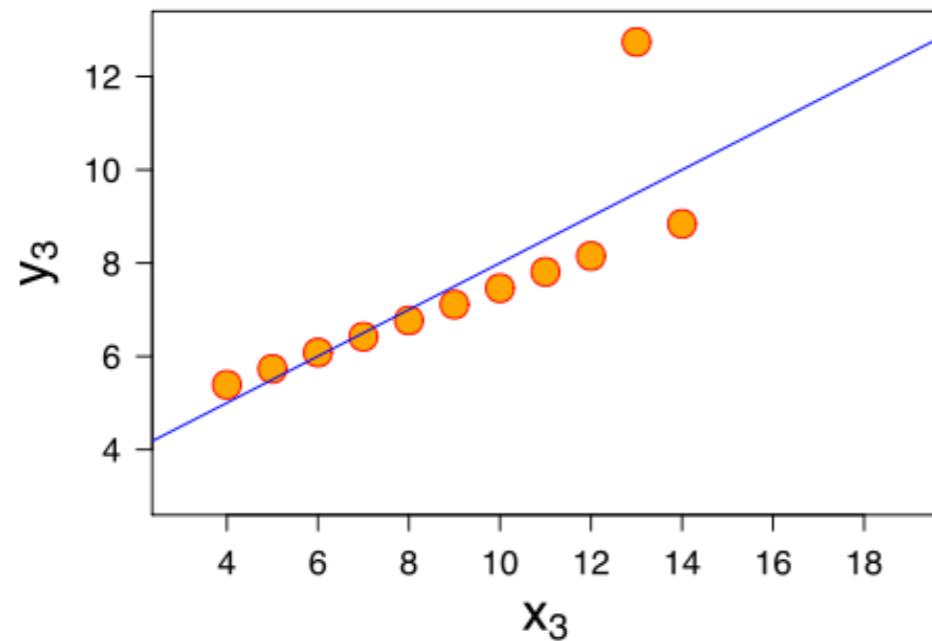
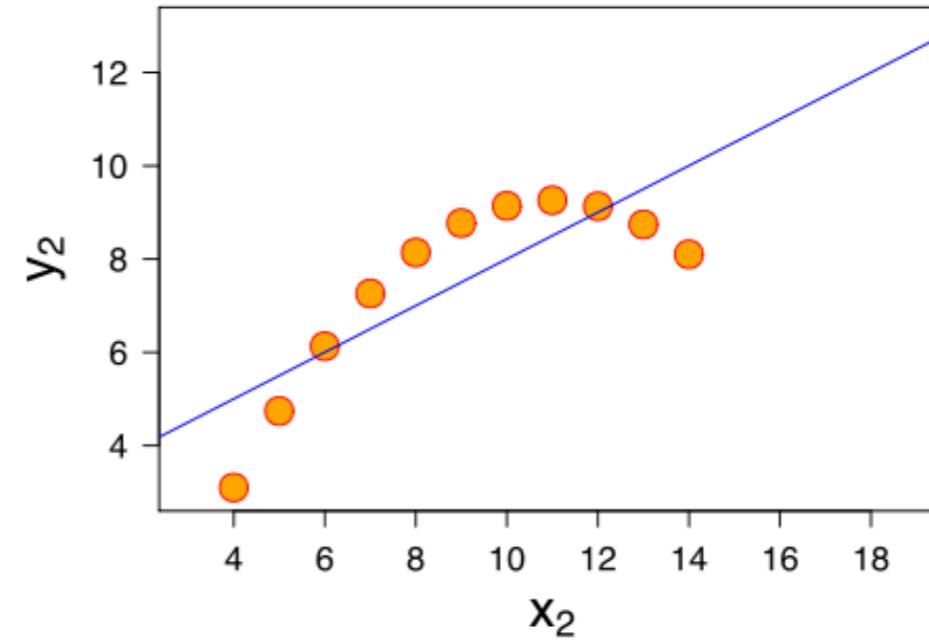
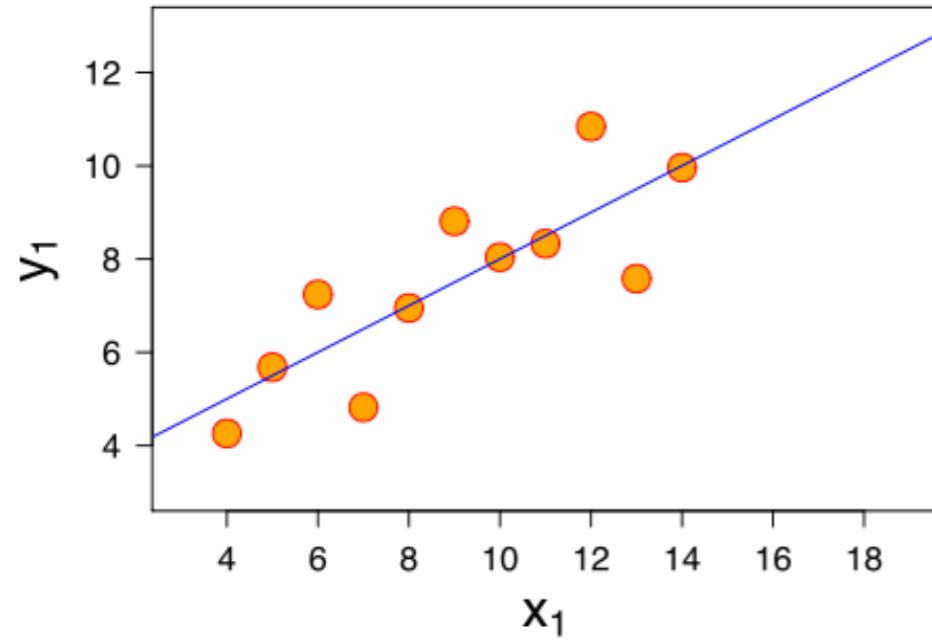
WHY NOT USE STATISTICS?

ANSCOME'S QUARTET

I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

- ▶ Mean, variance, correlation and linear regression line are all **identical**

VISUALISATION OF ANSCOME'S QUARTET



WHY USE VISUALISATIONS?

Abortion Rates By Age in the US

	1980	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
	300	297	294	288	286	275	280	275	275	274	266	259	259	255	251	246	245	245	242	238
Less than 15 years old	607	624	605	578	553	523	515	502	511	492	488	479	493	498	504	497	512	519	529	537
15 to 19 years old	451	462	457	449	444	418	403	379	370	364	353	347	350	346	341	337	339	341	339	337
20 to 24 years old	310	328	328	327	327	318	328	330	333	334	326	317	314	307	301	297	296	298	296	293
25 to 29 years old	213	219	219	216	216	213	224	224	228	230	227	224	228	226	224	221	220	219	215	211
30 to 34 years old	213	203	201	197	194	189	196	192	192	189	183	179	178	176	174	171	169	171	169	167
35 to 39 years old	317	280	277	265	254	244	249	241	239	234	226	219	215	208	203	200	195	195	190	186
40 years old and over	461	409	381	374	361	350	354	339	338	329	320	309	301	291	290	283	276	276	278	268

Which group has the lowest rate?

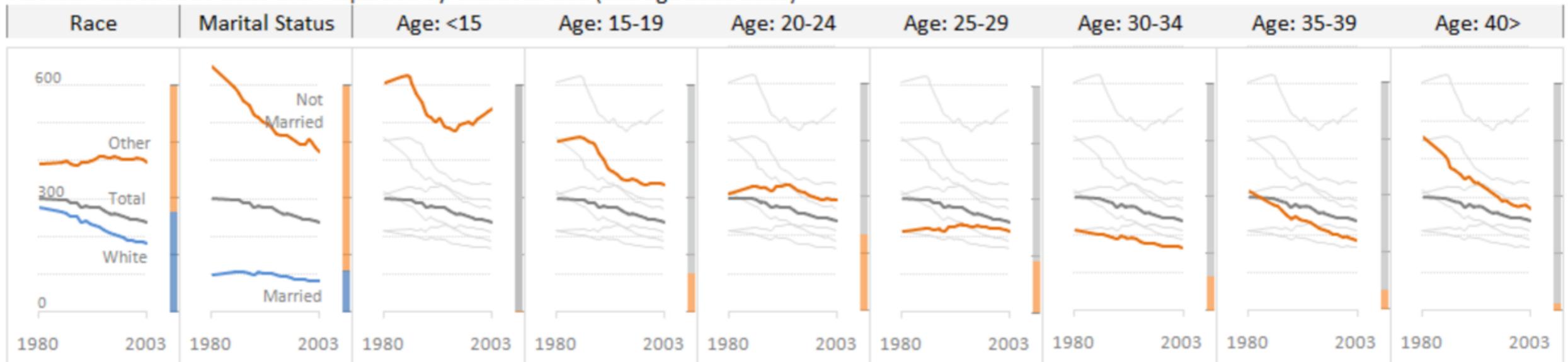
Are rates increasing or decreasing for any groups?

How fast/slow is the rate of change?

Source: L.B. Finer and S.K. Henshaw,
"Estimates of U.S. abortion incidence
in 2001-2003",
New York: Guttmacher Institute, 2006,
<http://www.guttmacher.org/pubs/2006/>
and unpublished data from Guttmacher
Institute.

WHY USE VISUALISATIONS?

Abortion Ratios 1980-2003 and Proportion by Characteristic (average 2001-2003)



Jorge Camoes / Excelcharts

Which group has the lowest rate?

Are rates increasing or decreasing for any groups?

Small multiples

How fast/slow is the rate of change?

- ▶ The right visualisation makes finding the patterns easy!



Photographic images
used in this presentation:

mostly from Microsoft Clipart circa 2009
(unless otherwise specified)
Note references are linked to images in PDF
version of this presentation.