Learning to Visualize: Surviving in the World of Data

Nam Wook Kim

Mini-Courses — January @ GSAS
2019

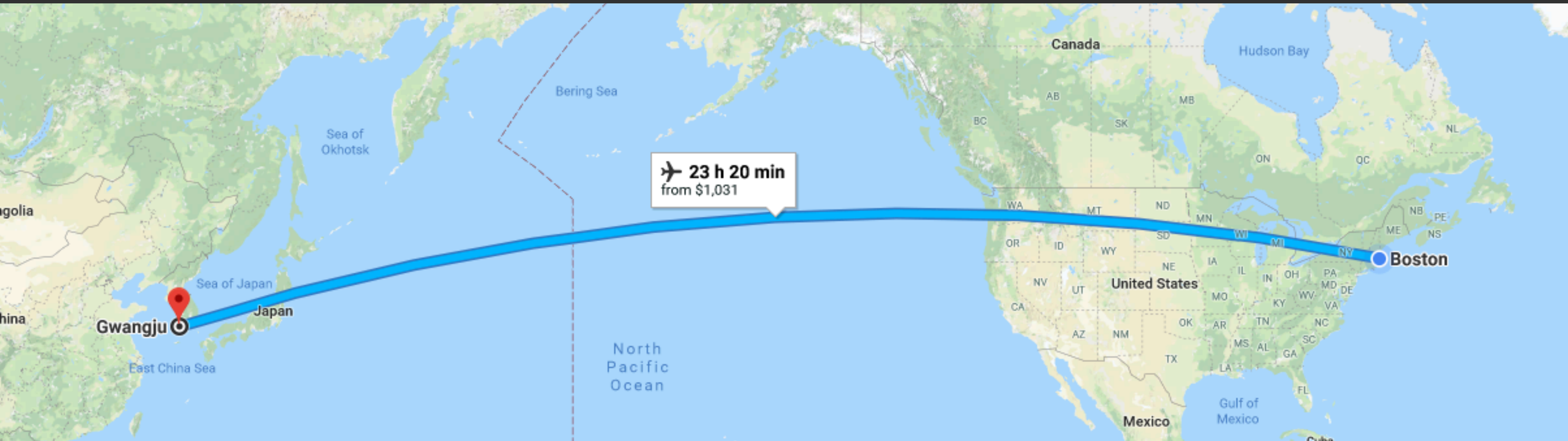
About Me

Nam Wook Kim

5th-Year Ph.D. Student

Computer Science Department

Information Visualization & Human-Computer Interaction



About You

Goal

To learn how to
design effective visualization

Goal

To learn how to
evaluate visualization design

Today

Fundamental

1. Value of visualization
2. Design principles
3. Graphical perception

Tomorrow

Practical

1. Data model and visual encoding
2. Exploratory data analysis
3. Storytelling with data
4. Advanced visualizations

Tomorrow

Practical

Tableau {

1. Data model and visual encoding
2. Exploratory data analysis
3. Storytelling with data
4. Advanced visualizations

The Value of Visualization

Big Data

Small Data

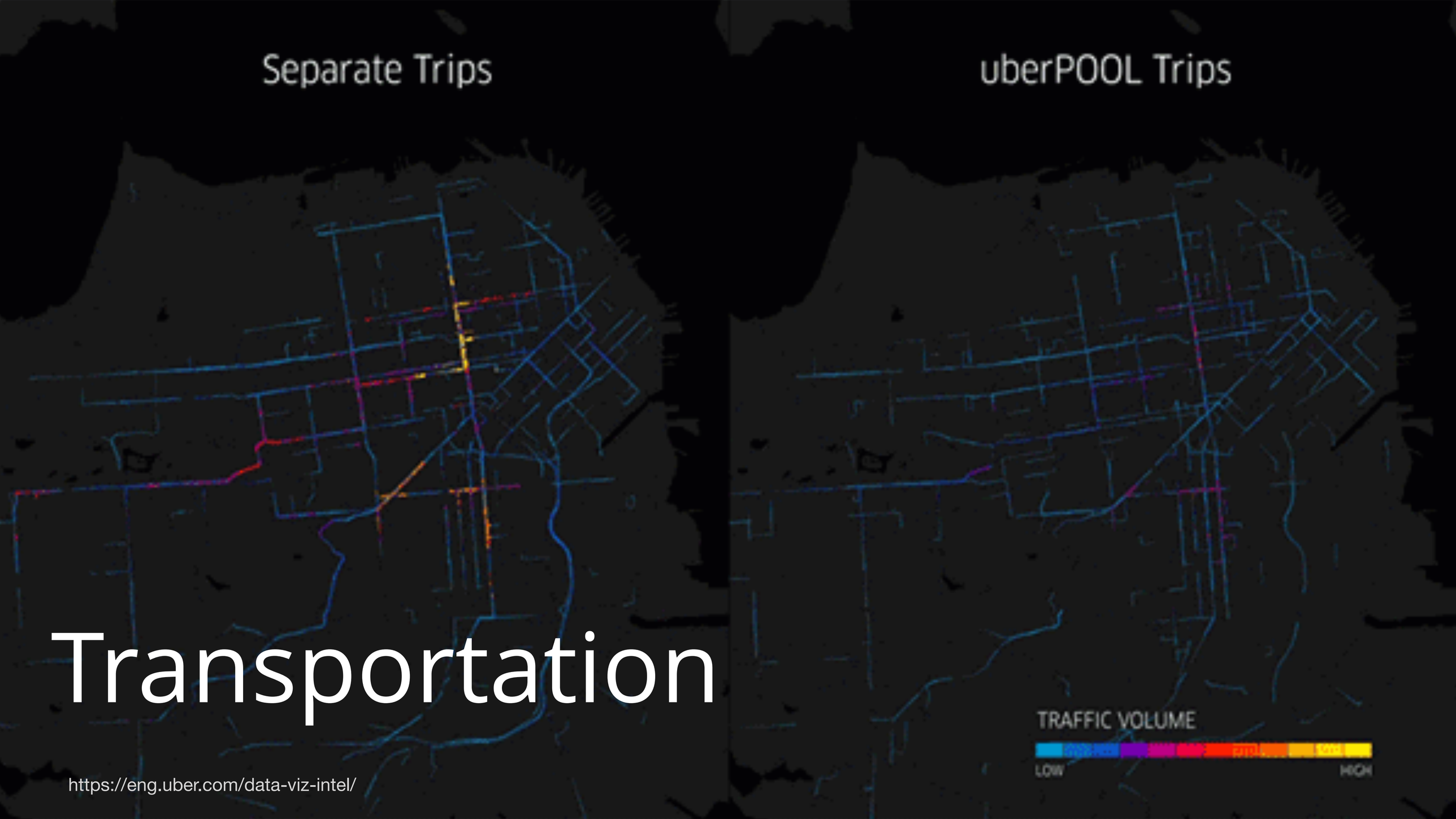
Data Everywhere

Replace with a visualization example.

Event sequence analysis??

Medical visualization (scientific)

Health & Medicine



Transportation

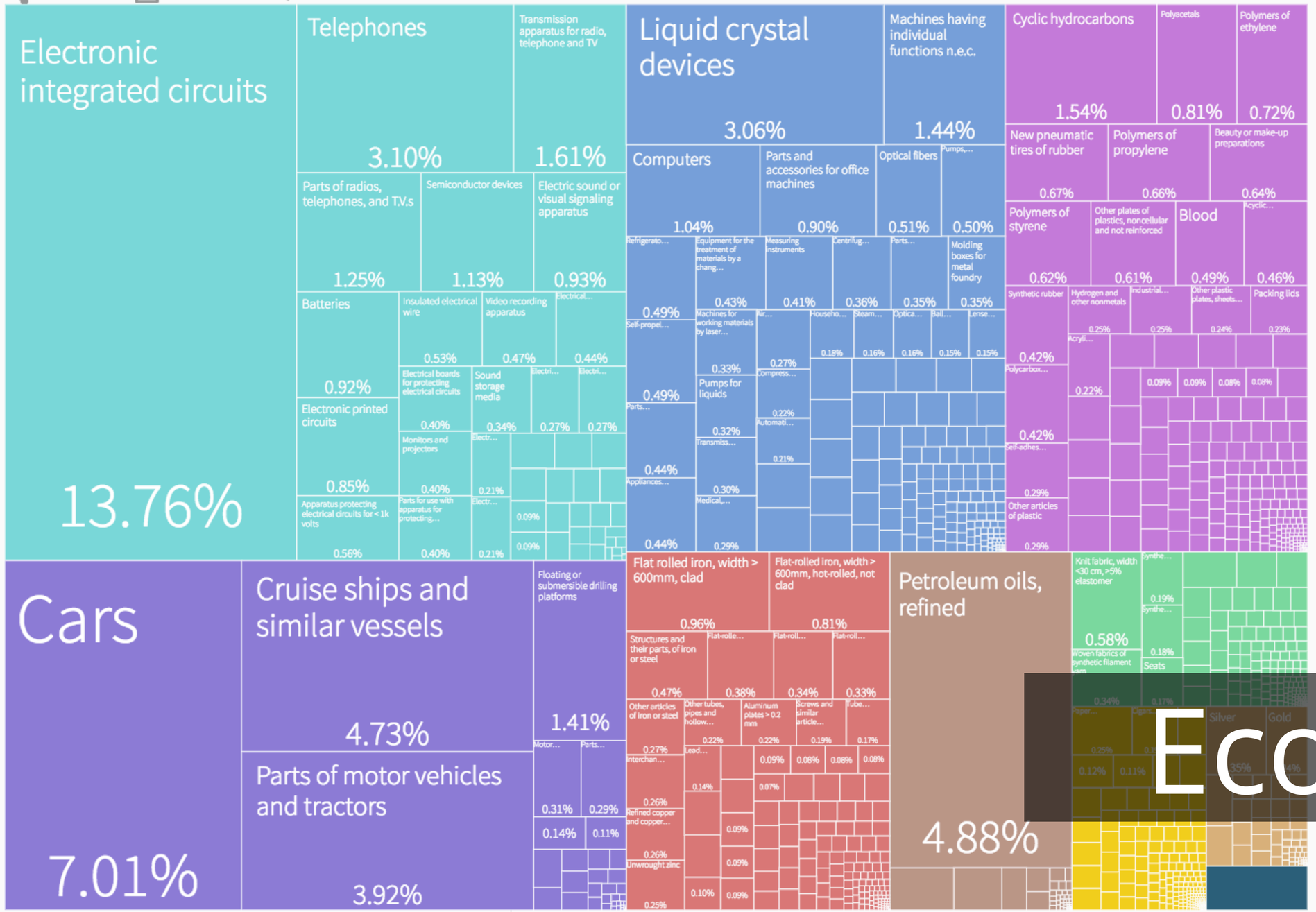
https://eng.uber.com/data-viz-intel/

What did South Korea export in 2016?

http://atlas.cid.harvard.edu/

Share Download Feedback

$518B / $518B



Please select a country, a product, or a combination of the two.

South Korea

Exports | Imports

By Product | By Partner

Select a HS product

Trade Data Visualization

Complexity Visualization

Product Tree Map

The tree map displays the breakdown of exports or imports by country or product, in a given year. **Click on a panel for more information or generate variations using the options below.**

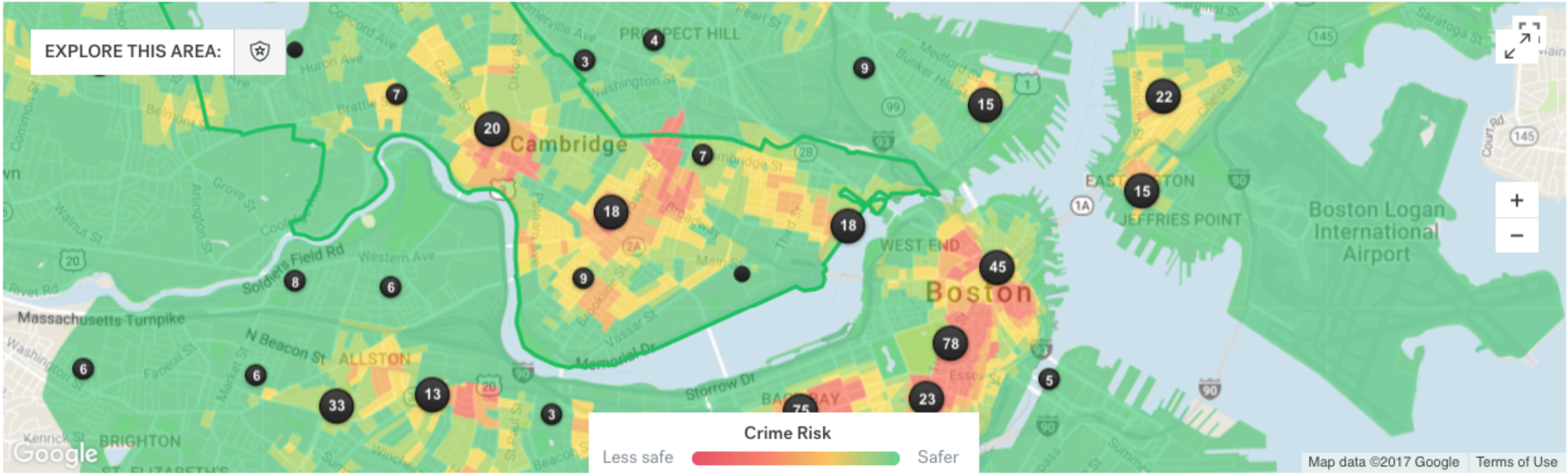
Learn more

Economy

Crime Data in Cambridge

https://www.trulia.com/real_estate/Cambridge-Massachusetts/crime/

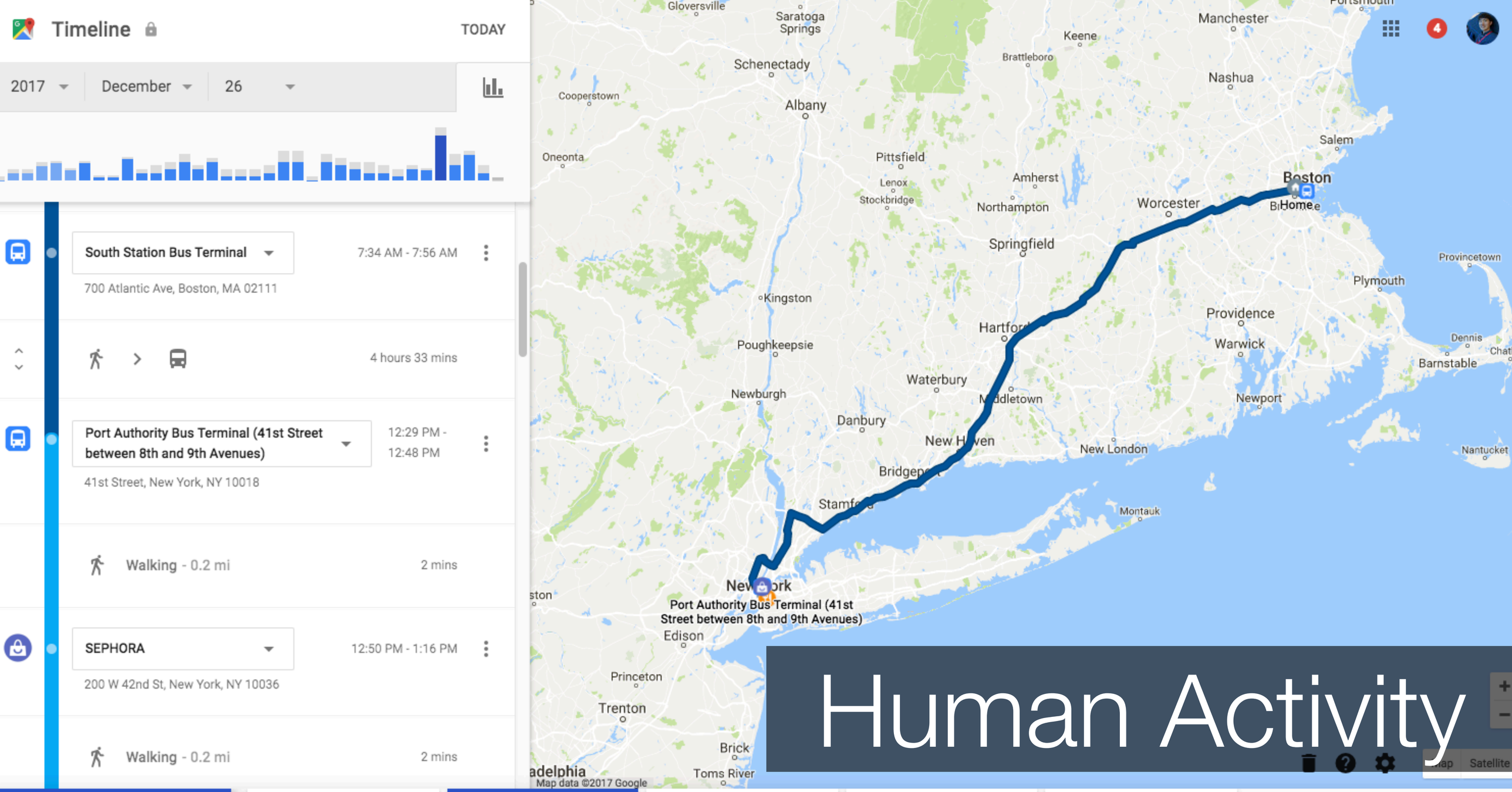
Trulia uses crime reports to provide valuable information on the relative safety of homes in the U.S. Use the map below to learn more about crime activity in and around Cambridge.



Most frequent crimes in the past year

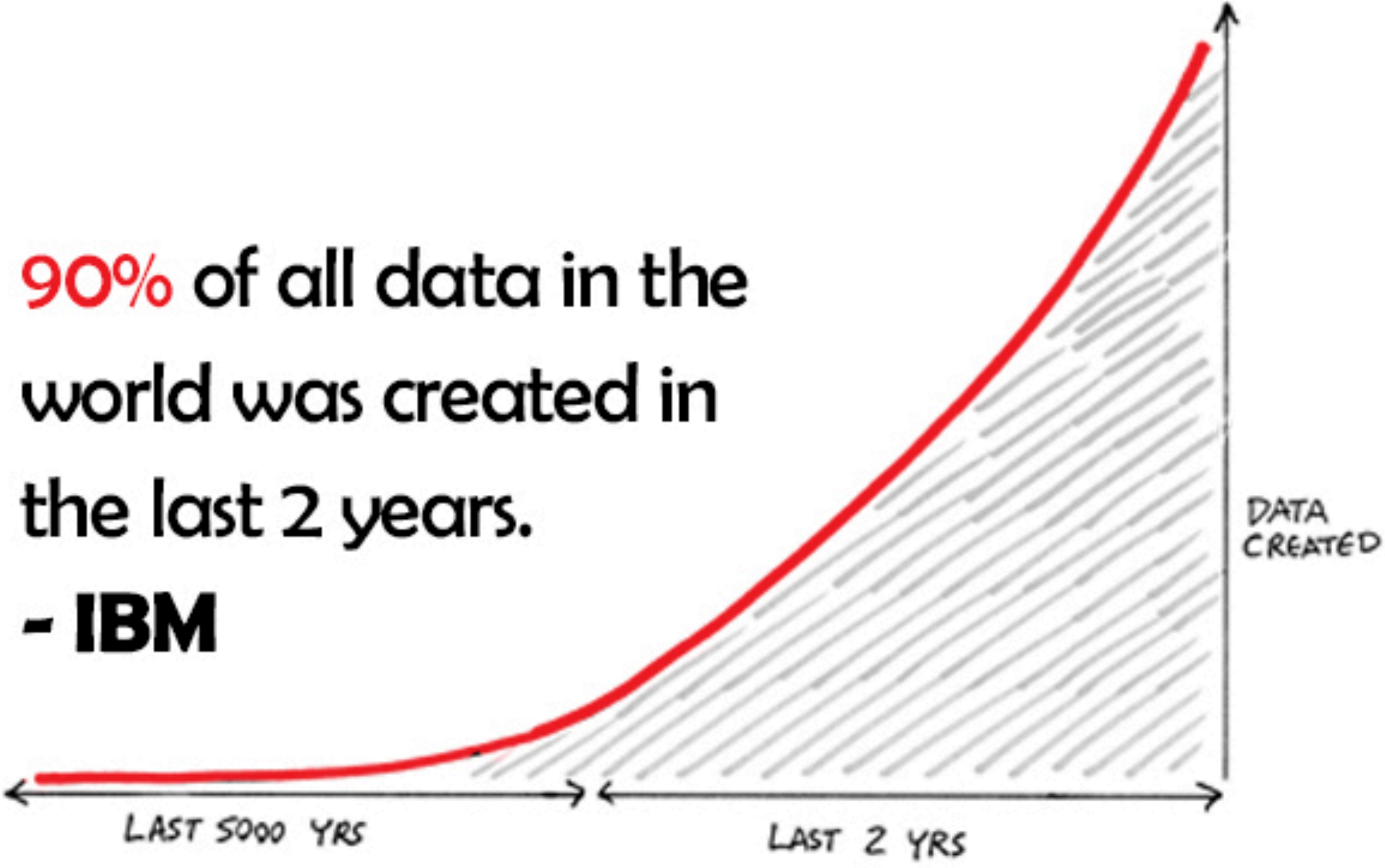
1,754	362	304	235	92
COUNTS OF THEFT	COUNTS OF BURGLARY	COUNTS OF ARREST	COUNTS OF ASSAULT	COUNTS OF VANDALISM

Public Safety


Timeline
TODAY
2017
December
26
South Station Bus Terminal
7:34 AM - 7:56 AM
700 Atlantic Ave, Boston, MA 02111
4 hours 33 mins
Port Authority Bus Terminal (41st Street between 8th and 9th Avenues)
12:29 PM - 12:48 PM
41st Street, New York, NY 10018
Walking - 0.2 mi
2 mins
SEPHORA
12:50 PM - 1:16 PM
200 W 42nd St, New York, NY 10036
Walking - 0.2 mi
2 mins
Boston
Home
Worcester
Springfield
Hartford
Providence
Warwick
Newport
New London
Waterbury
Danbury
Newburgh
Poughkeepsie
Kingston
Albany
Schenectady
Saratoga Springs
Pittsfield
Northampton
Amherst
Manchester
Nashua
Keene
Brattleboro
Salem
Plymouth
Provincetown
Barnstable
Dennis
Nantucket
Montauk
Edison
Princeton
Trenton
Brick
Toms River
Port Authority Bus Terminal (41st Street between 8th and 9th Avenues)
Map data ©2017 Google
Human Activity
Boston and New York
Dec 26 – 30, 2017
25 Monday December 2017
26 Tuesday December 2017
27 Wednesday December 2017
28 Thursday December 2017
29 Friday December 2017

90% of all data in the world was created in the last 2 years.

- IBM



The Industrial Revolution of Data

Joe Hellerstein, UC Berkley, 2008

Data Literacy

“The ability to take data — to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it — that’s going to be a hugely important skill in the next decades …”

Hal Varian, Google’s Chief Economist
The McKinsey Quarterly, January 2009

A Poverty of Attention

“…Information consumes the attention of its recipients. Hence … a need to allocate that attention efficiently among the overabundance of information sources that might consume it.”

Herbert A. Simon
Economist & Psychologist

Visualization can help!

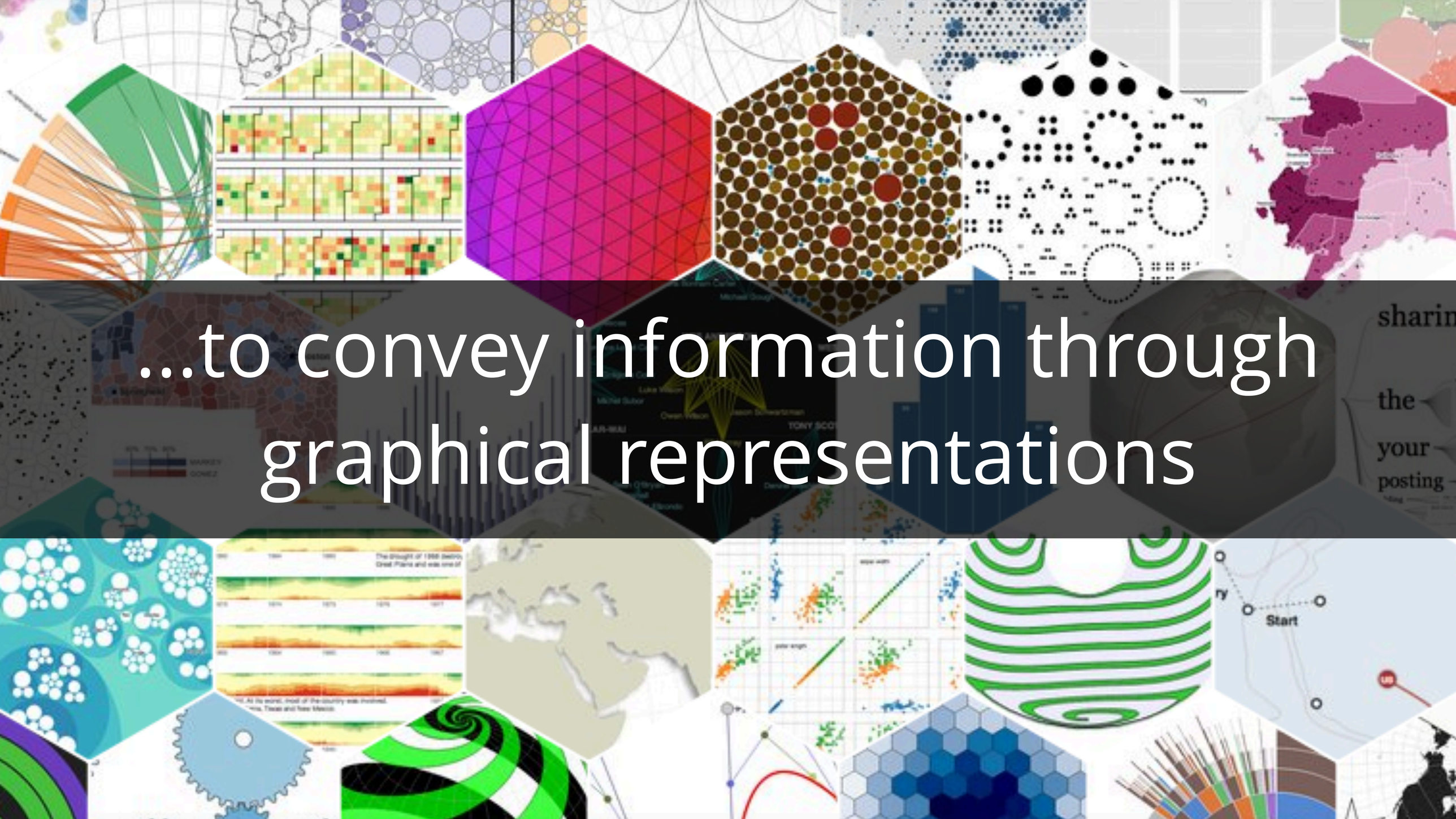
provides a powerful yet accessible way to make sense of large and complex data

What is Visualization?

“Transformation of the **symbolic** into the **geometric**”
—McCormick et al. 1987

“... finding the **artificial memory** that best supports our natural means of perception.”
—Bertin 1967

“visual representations of data to **amplify cognition**.”
—Card, Mackinlay, & Shneiderman 1999


...to convey information through graphical representations
sharin
the
your
posting
Start

Anscombe's Quartet

A		*B*		*C*		*D*	
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	6.8

Summary Statistics

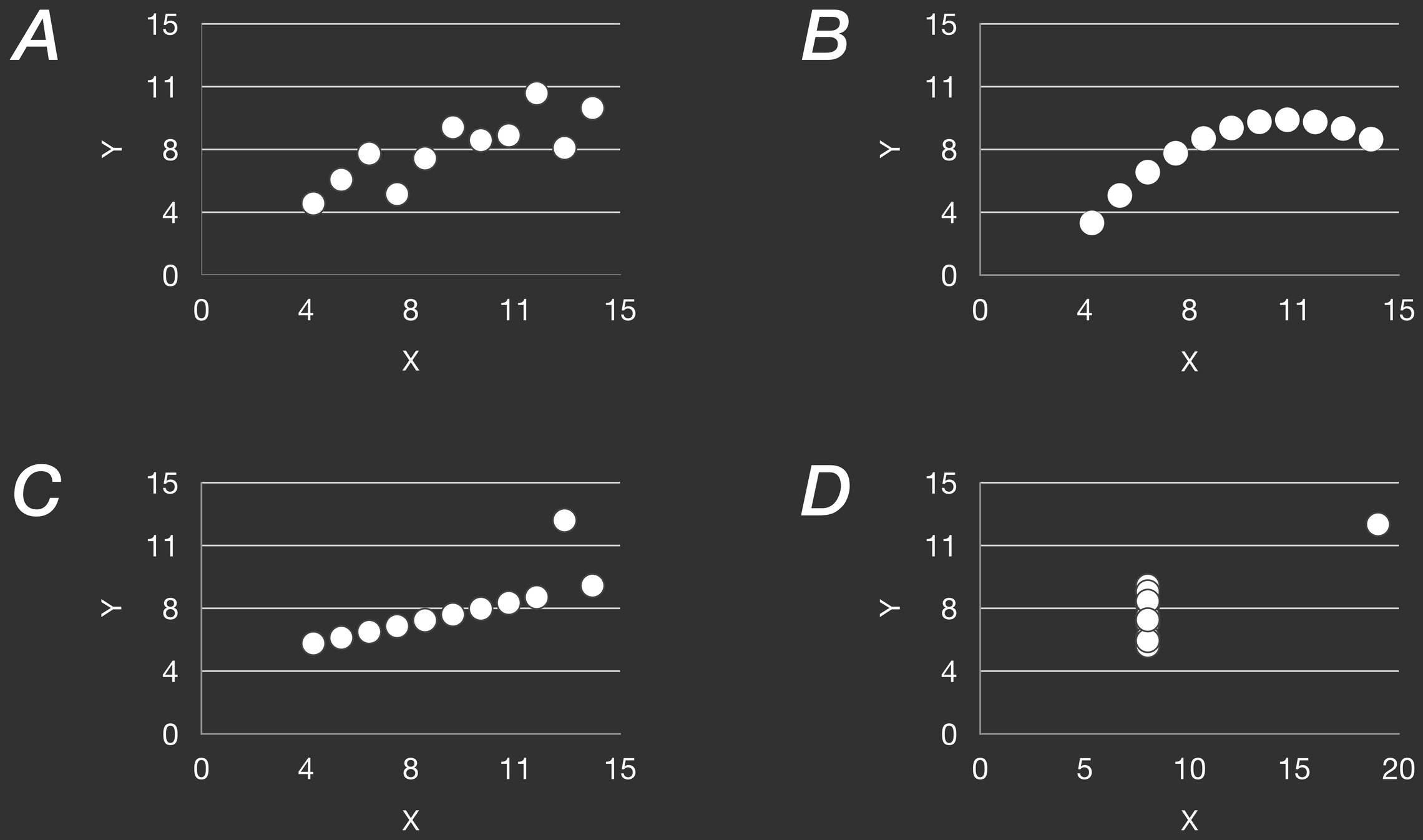
$u_X = 9.0$ $\sigma_X = 3.317$

$u_Y = 7.5$ $\sigma_Y = 2.03$

Linear Regression

$Y = 3 + 0.5\,X$

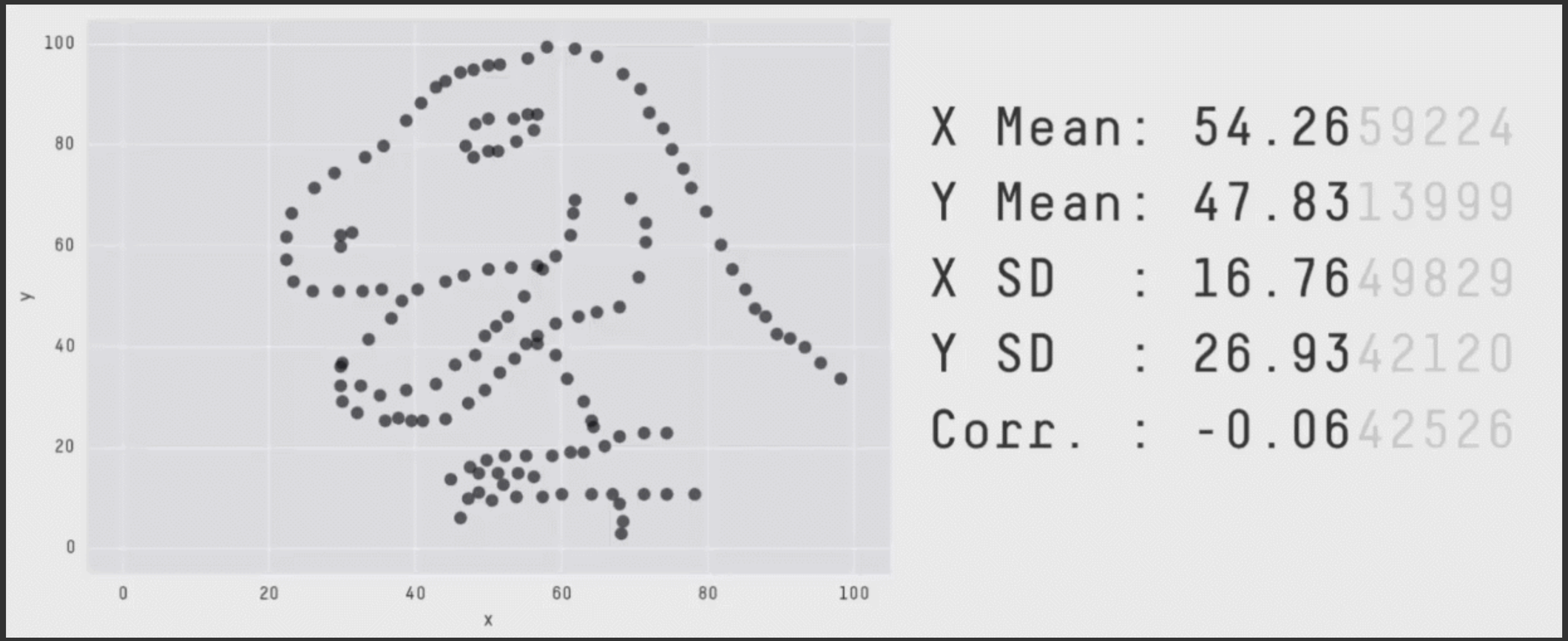
$R^2 = 0.67$


A
15
11
8
4
0
Y
0
4
8
11
15
X
B
15
11
8
4
0
Y
0
4
8
11
15
X
C
15
11
8
4
0
Y
0
4
8
11
15
X
D
15
11
8
4
0
Y
0
5
10
15
20
X

“

…make both calculations and graphs. Both sorts of output should be studied; each will contribute to understanding.

F. J. Anscombe, 1973



All distinct datasets with same statistical properties

Matejka & Fitzmaurice 2017

Why Create Visualizations?

Why Create Visualizations?

- Answer questions (or discover them)
- Make decisions
- See data in context
- Expand memory
- Support graphical calculation
- Find patterns
- Present argument or tell a story
- Inspire

Three functions of visualization

1. Record
2. Analyze
3. Communicate

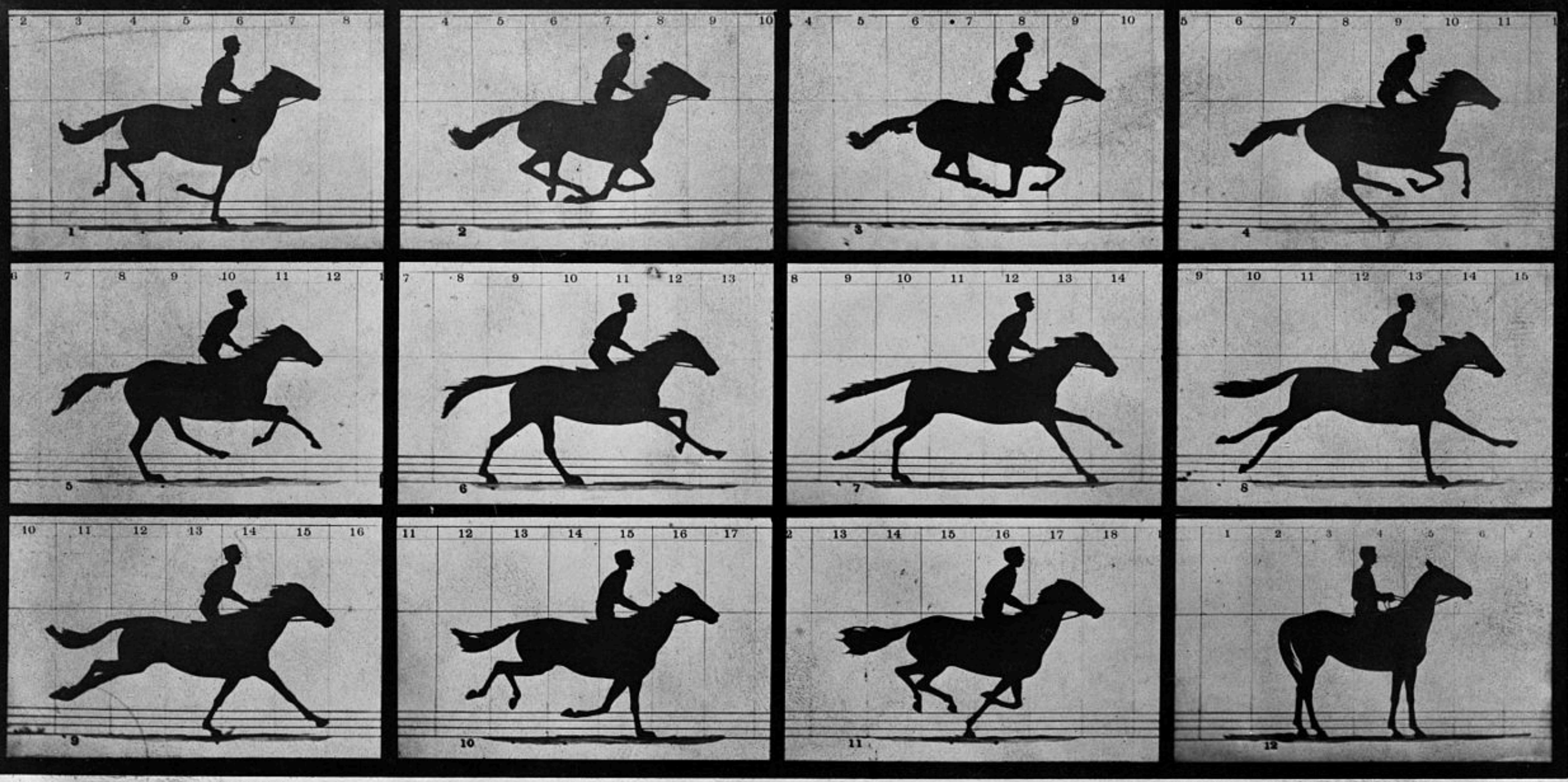
Record Information

6200 BC

Leonardo da Vinci 1485

Galileo Galilei's Sketches of the Moon

(November-December 1609)



Copyright, 1878, by MUYBRIDGE.

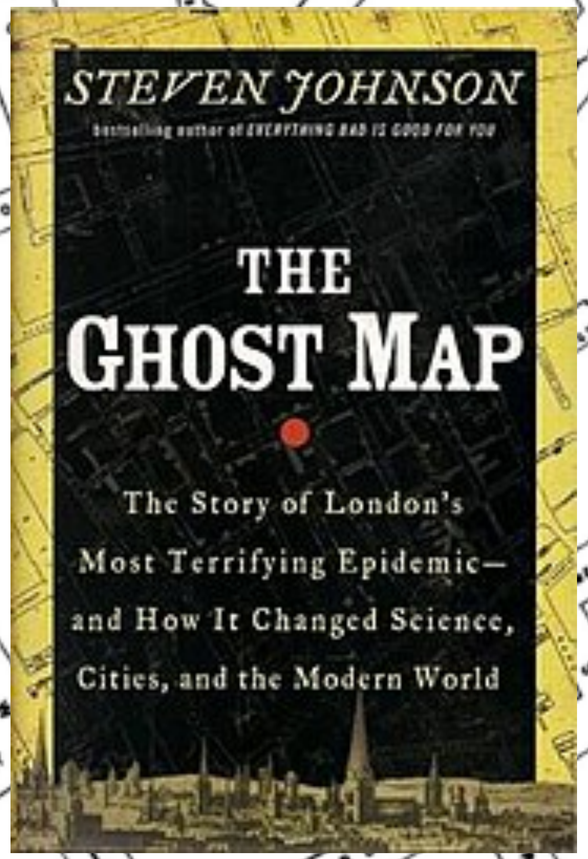
MORSE'S Gallery, 417 Montgomery St., San Francisco.

The Horse in Motion.

Illustrated by

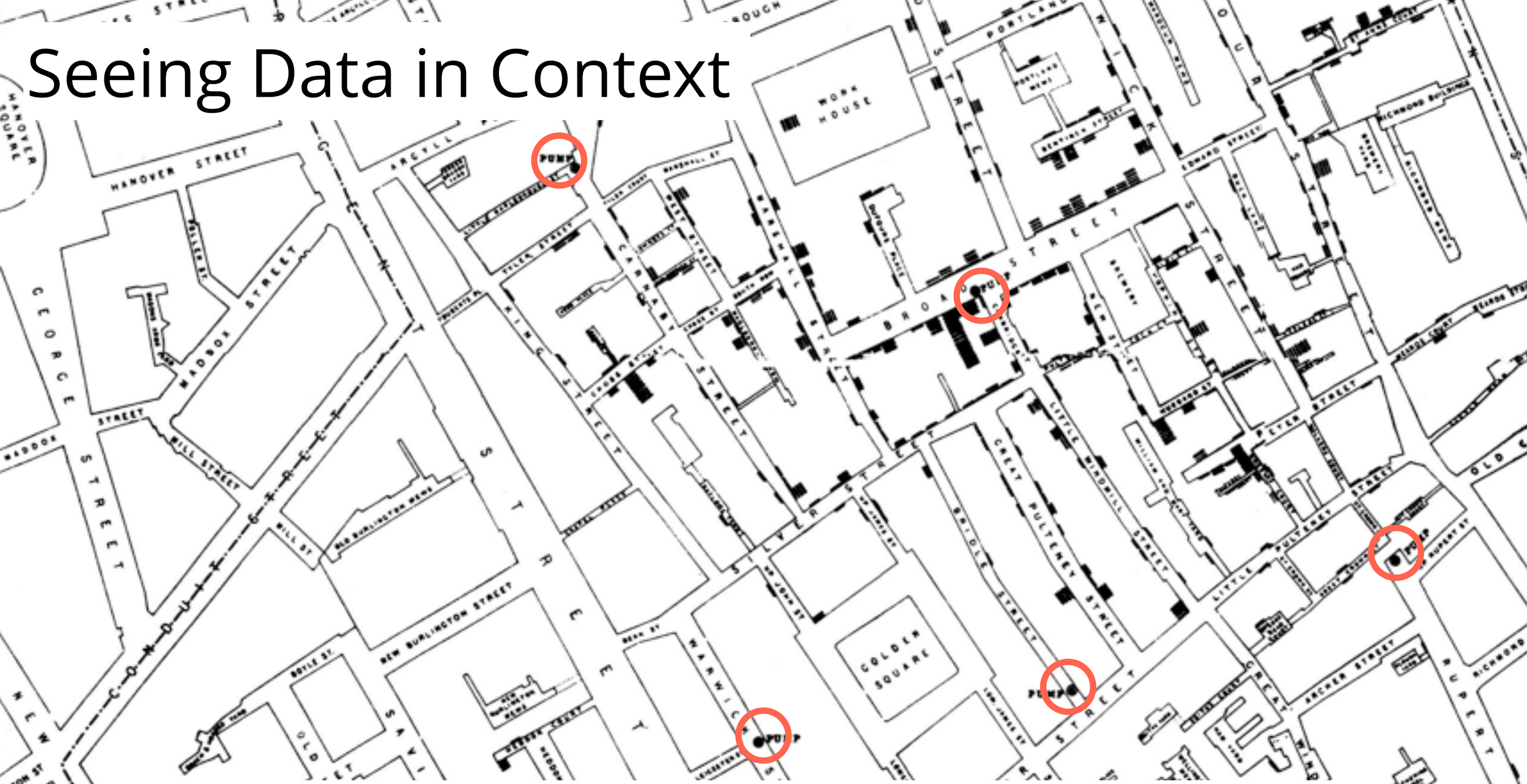
E. J. Muybridge, 1878

Support Reasoning

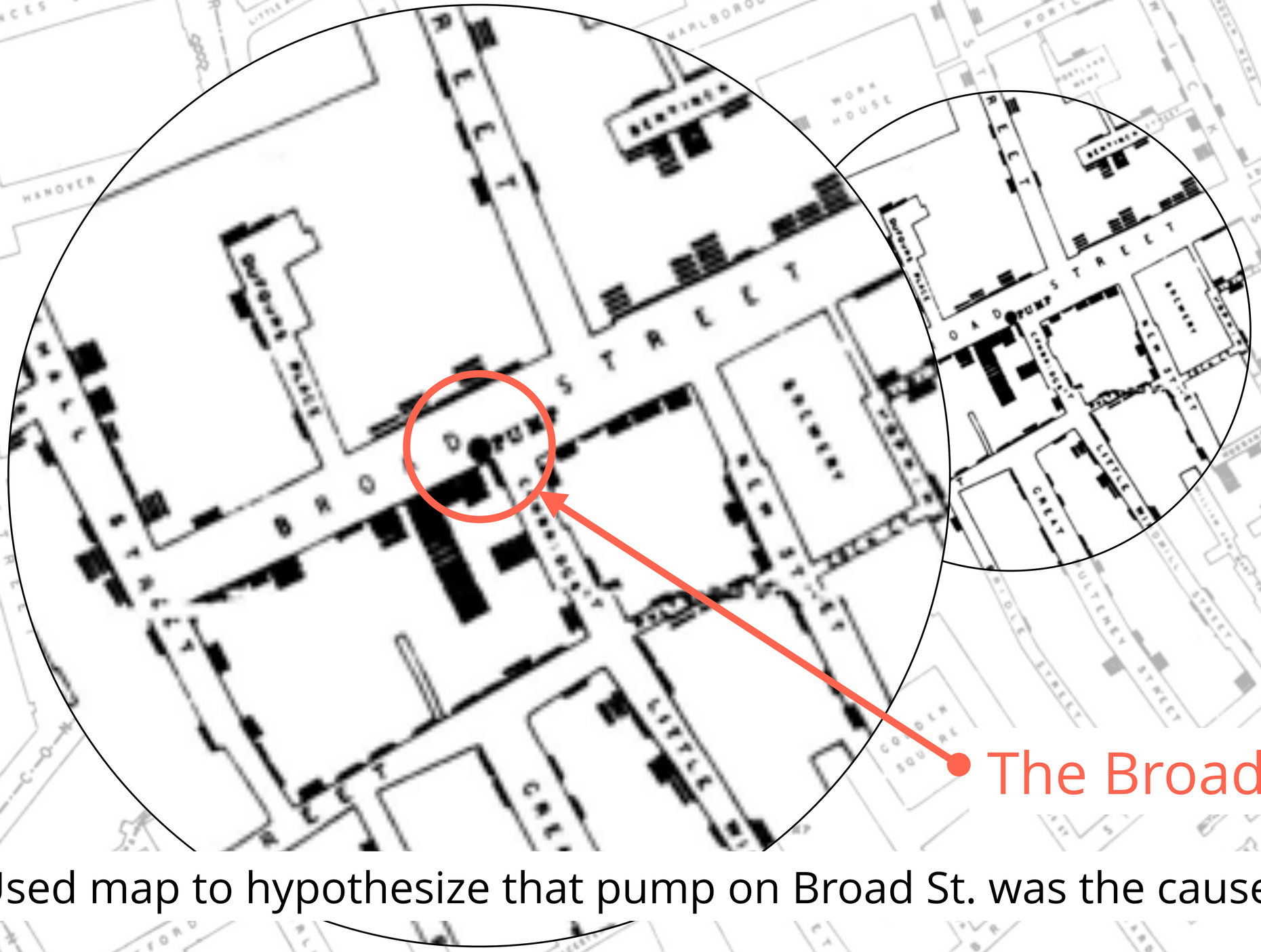


John Snow, the Cholera Epidemic 1854

Seeing Data in Context



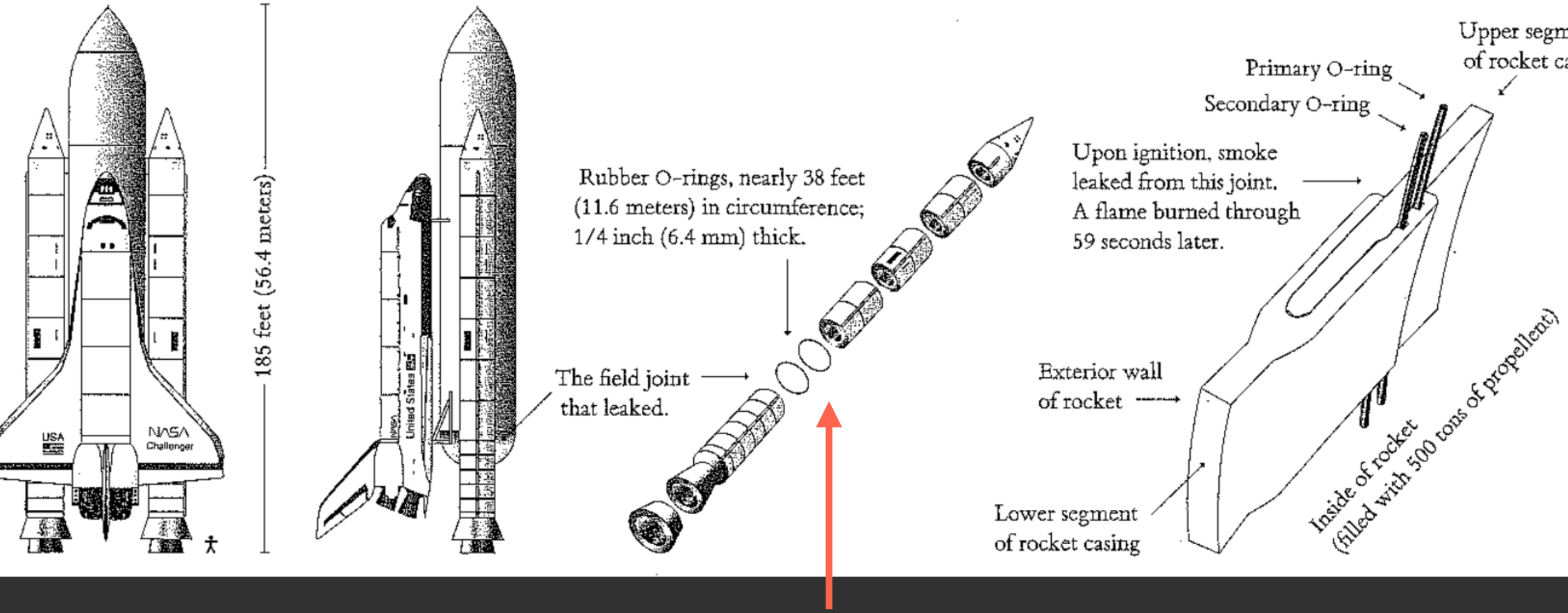
Plotted the position of each cholera case on a map. [from Tufte 83]



Used map to hypothesize that pump on Broad St. was the cause. [from Tufte 83]

Space Shuttle Challenger Disaster (1986)

approx. 73 seconds after



Rubber O-rings
had problems with cold temperatures.

One of original reports sent to NASA officials before launch

HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

	SRM No.	Cross Sectional View: Erosion Depth (in.)	Cross Sectional View: Perimeter Affected (deg)	Cross Sectional View: Nominal Dia. (in.)	Top View: Length Of Max Erosion (in.)	Top View: Total Heat Affected Length (in.)	Clocking Location (deg)
61A LH Center Field**	22A	None	None	0.280	None	None	36° --66°
61A LH ~~CENTER~~ FIELD**	22A	NONE	NONE	0.280	NONE	NONE	338°-18°
51C LH Forward Field**	15A	0.010	154.0	0.280	4.25	5.25	163
51C RH Center Field (prim)***	15B	0.038	130.0	0.280	12.50	58.75	354
51C RH Center Field (sec)***	15B	None	45.0	0.280	None	29.50	354
41D RH Forward Field	13B	0.028	110.0	0.280	3.00	None	275
41C LH Aft Field*	11A	None	None	0.280	None	None	--
41B LH Forward Field	10A	0.040	217.0	0.280	3.00	14.50	351
STS-2 RH Aft Field	2B	0.053	116.0	0.280	--	--	90

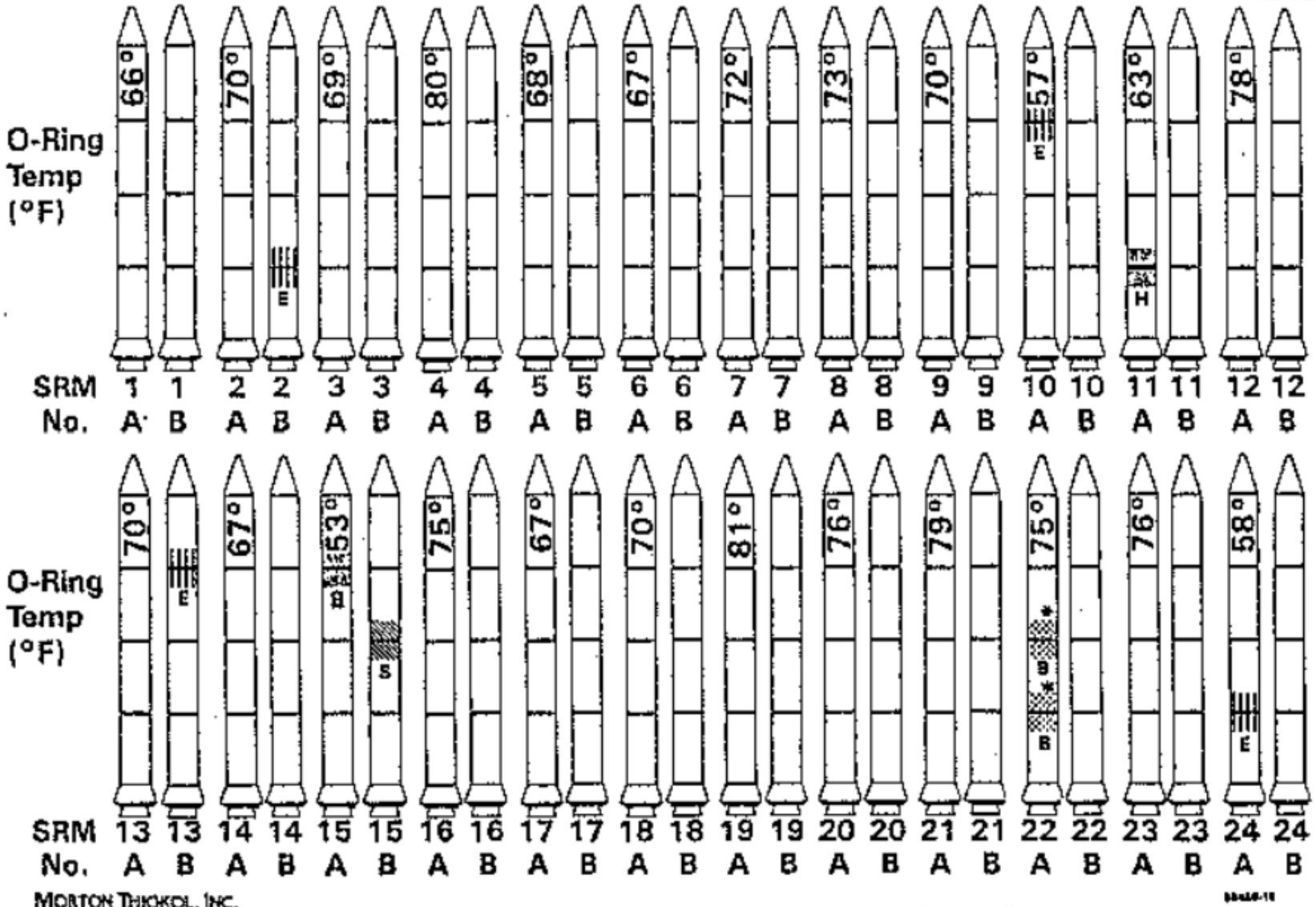
*Hot gas path detected in putty. Indication of heat on O-ring, but no damage.
**Soot behind primary O-ring.
***Soot behind primary O-ring, heat affected secondary O-ring.

Clocking location of leak check port - 0 deg.

OTHER SRM-15 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY AND NO SOOT NEAR OR BEYOND THE PRIMARY O-RING.

SRM-22 FORWARD FIELD JOINT HAD PUTTY PATH TO PRIMARY O-RING, BUT NO O-RING EROSION AND NO SOOT BLOWBY. OTHER SRM-22 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY.

History of O-Ring Damage in Field Joints (Cont)



Code

- S = Heating of Secondary O-Ring
- B = Primary O-Ring Blowby
- E = Primary O-Ring Erosion
- H = Heating of Primary O-Ring
- = No Damage

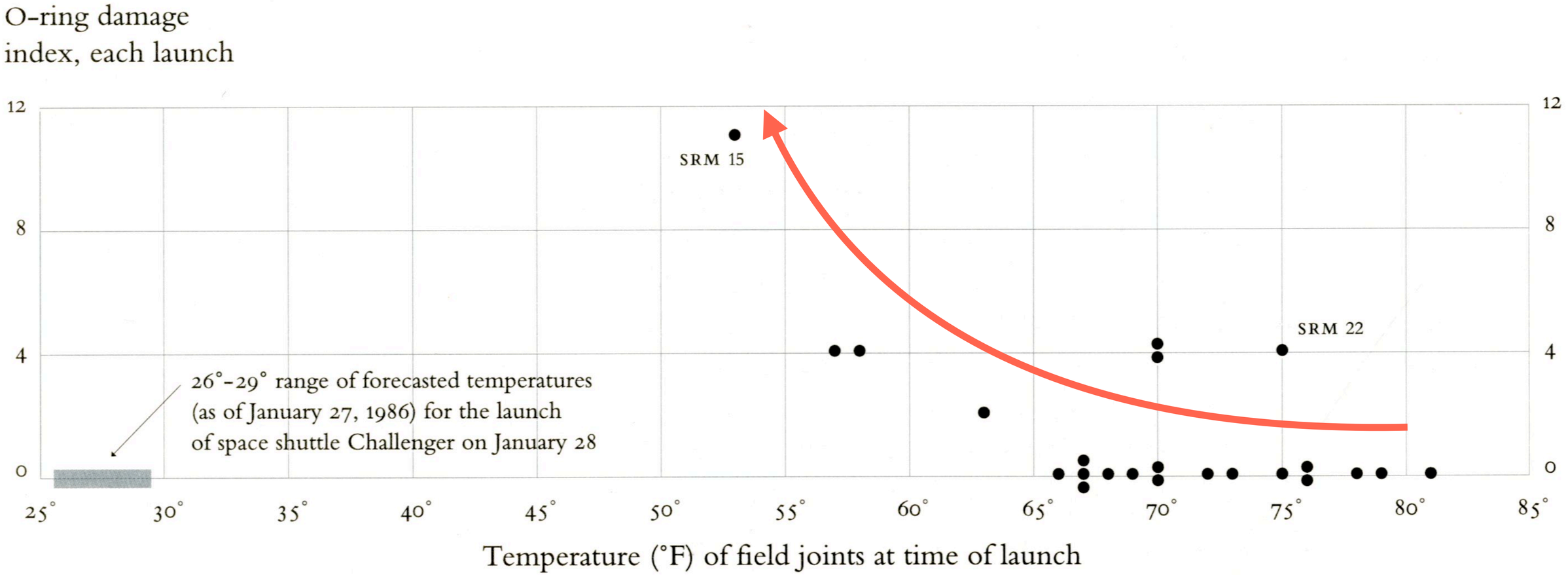
STATIC TEST MOTORS

- HORIZONTAL ASSEMBLY
- SOME PUTTY REPAIRED

MORTON THIOKOL, INC.
Wasatch Operations

INFORMATION ON THIS PAGE WAS PREPARED TO SUPPORT AN ORAL PRESENTATION AND CANNOT BE CONSIDERED COMPLETE WITHOUT THE ORAL DISCUSSION

Use a right visualization to make a right decision



[Edward Tufte 1997]

Expand Memory: Feynman Diagram (1948)

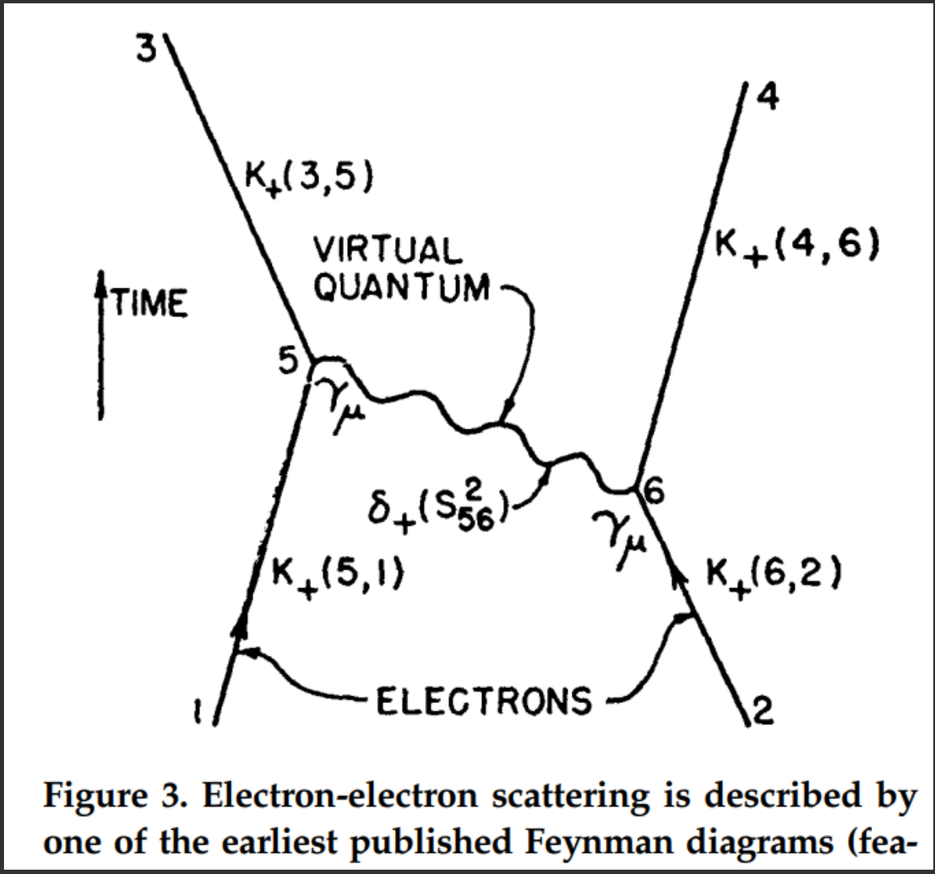
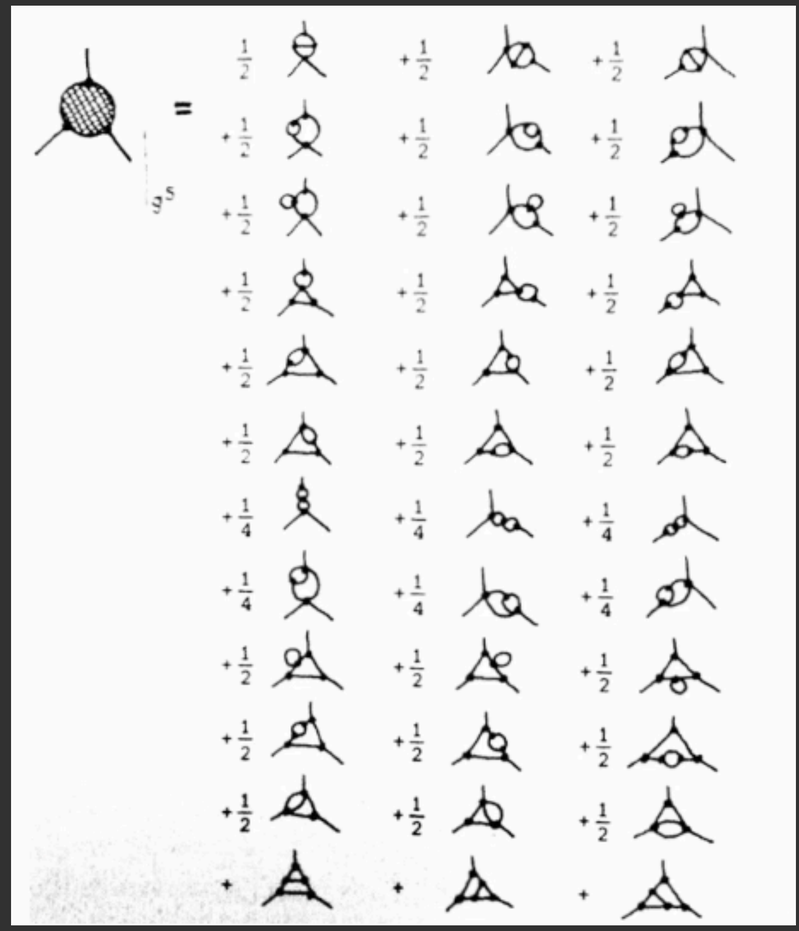
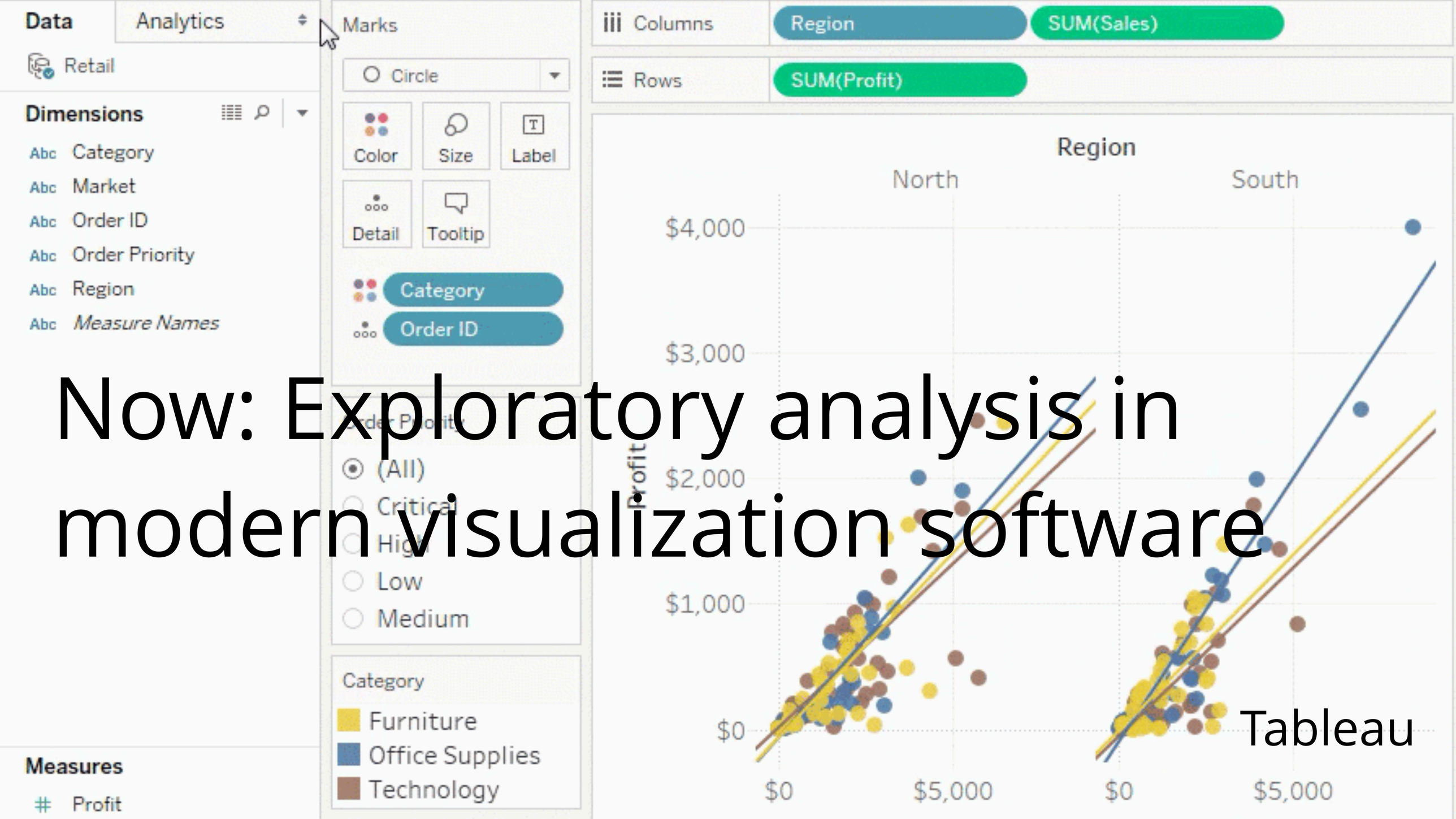


Figure 3. Electron-electron scattering is described by one of the earliest published Feynman diagrams (fea-



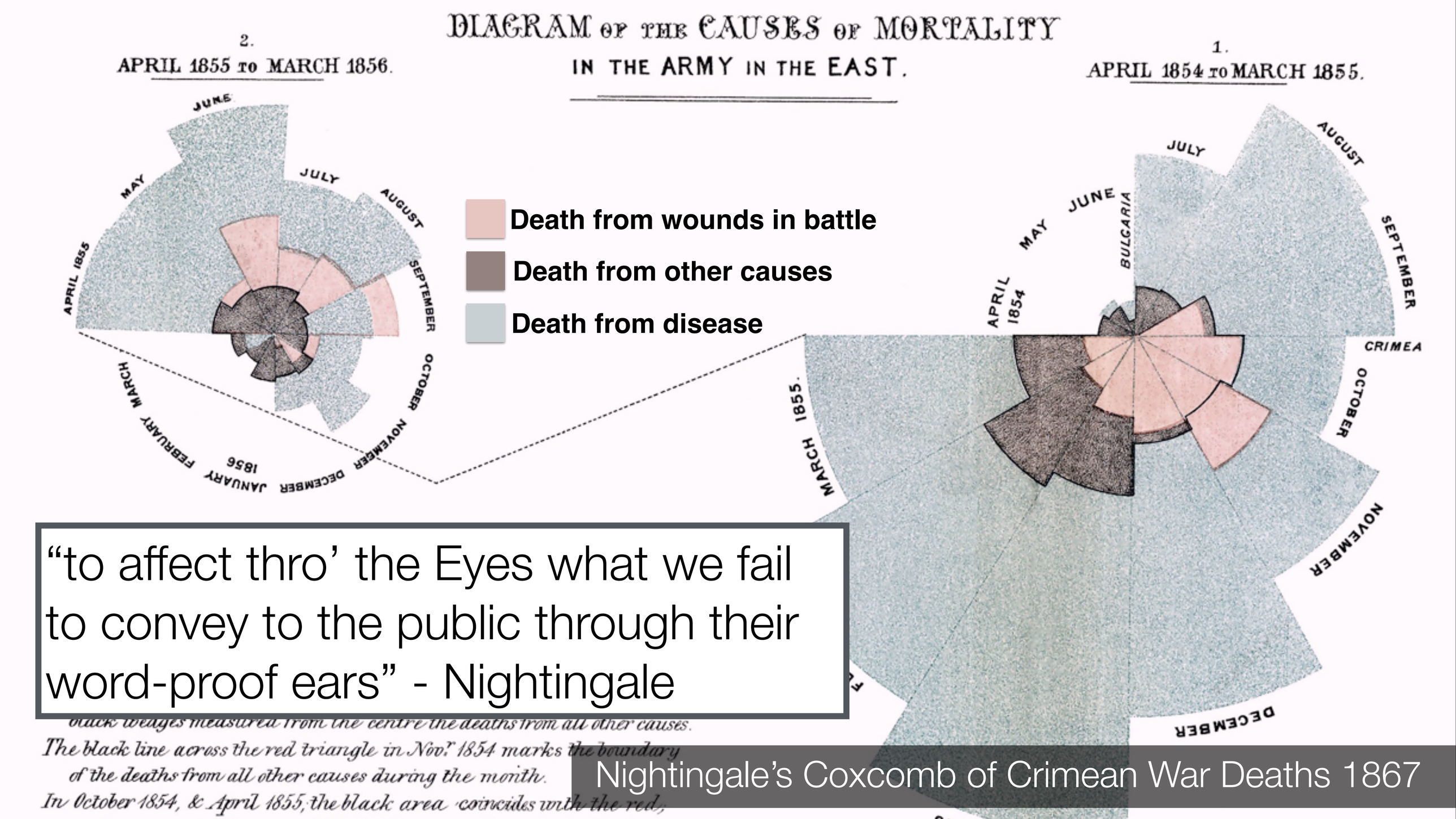
"since the middle of the 20th century, theoretical physicists have increasingly turned to this tool to help them undertake critical calculations" — David Kaiser



Now: Exploratory analysis in modern visualization software

Tableau

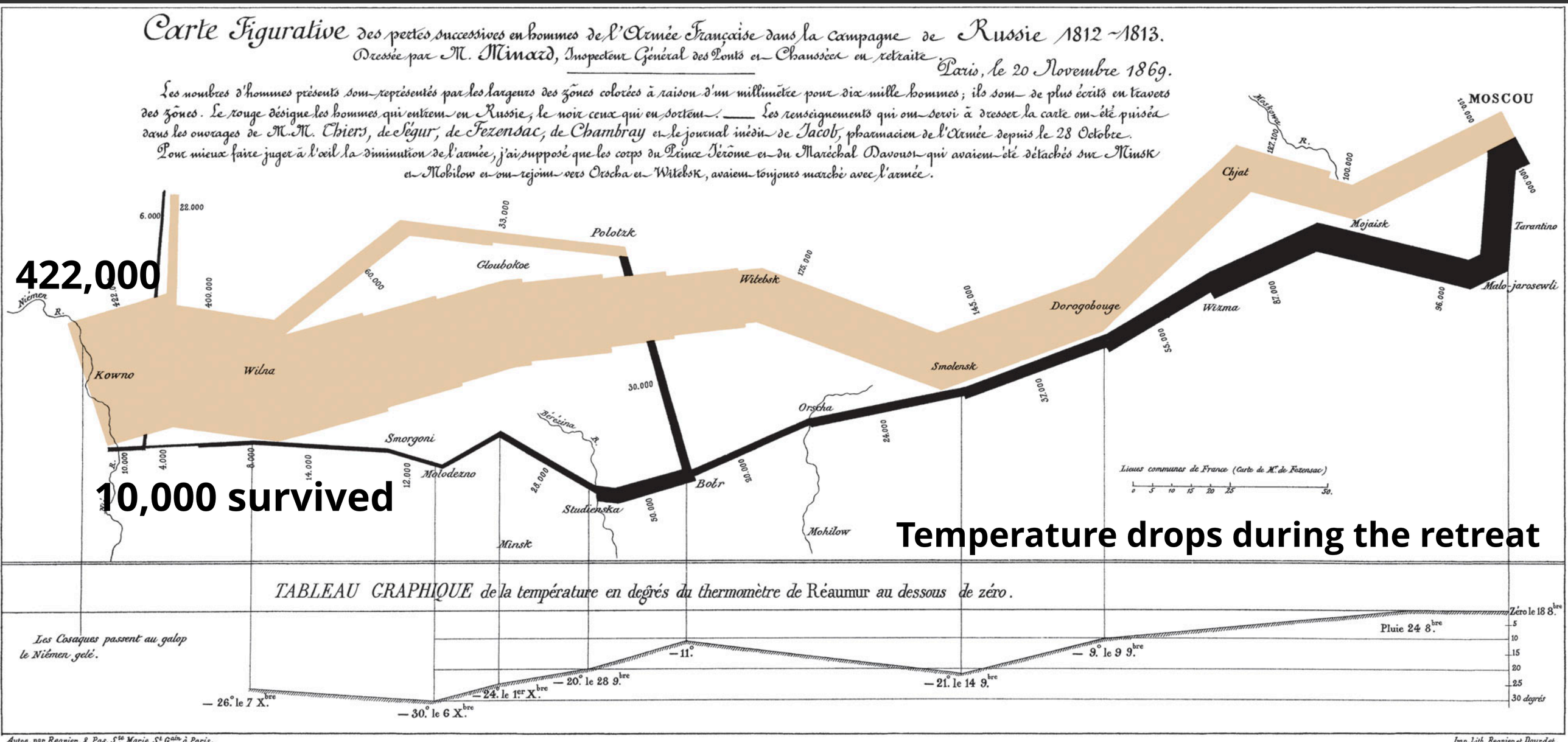
Convey Information to Others



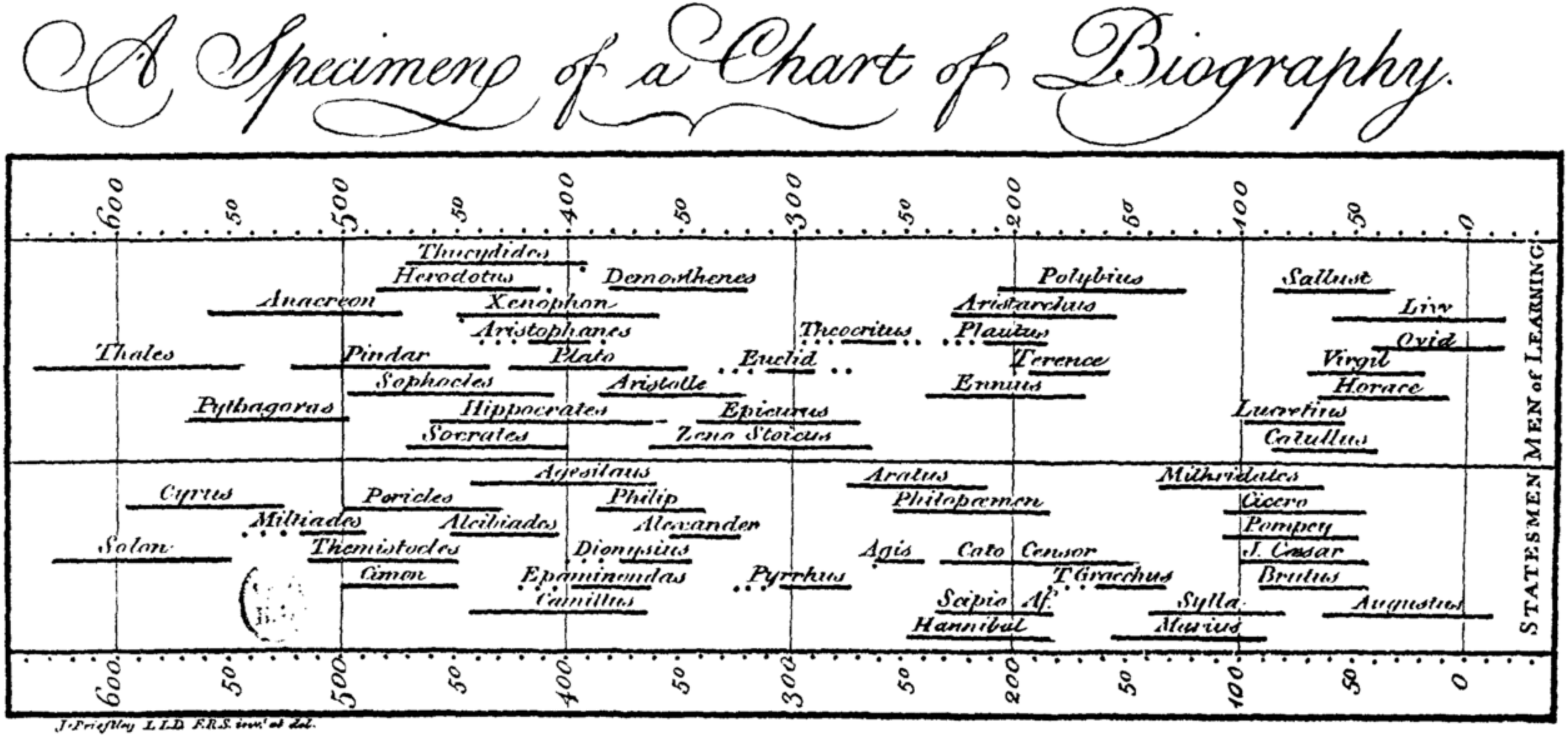
"to affect thro' the Eyes what we fail to convey to the public through their word-proof ears" - Nightingale

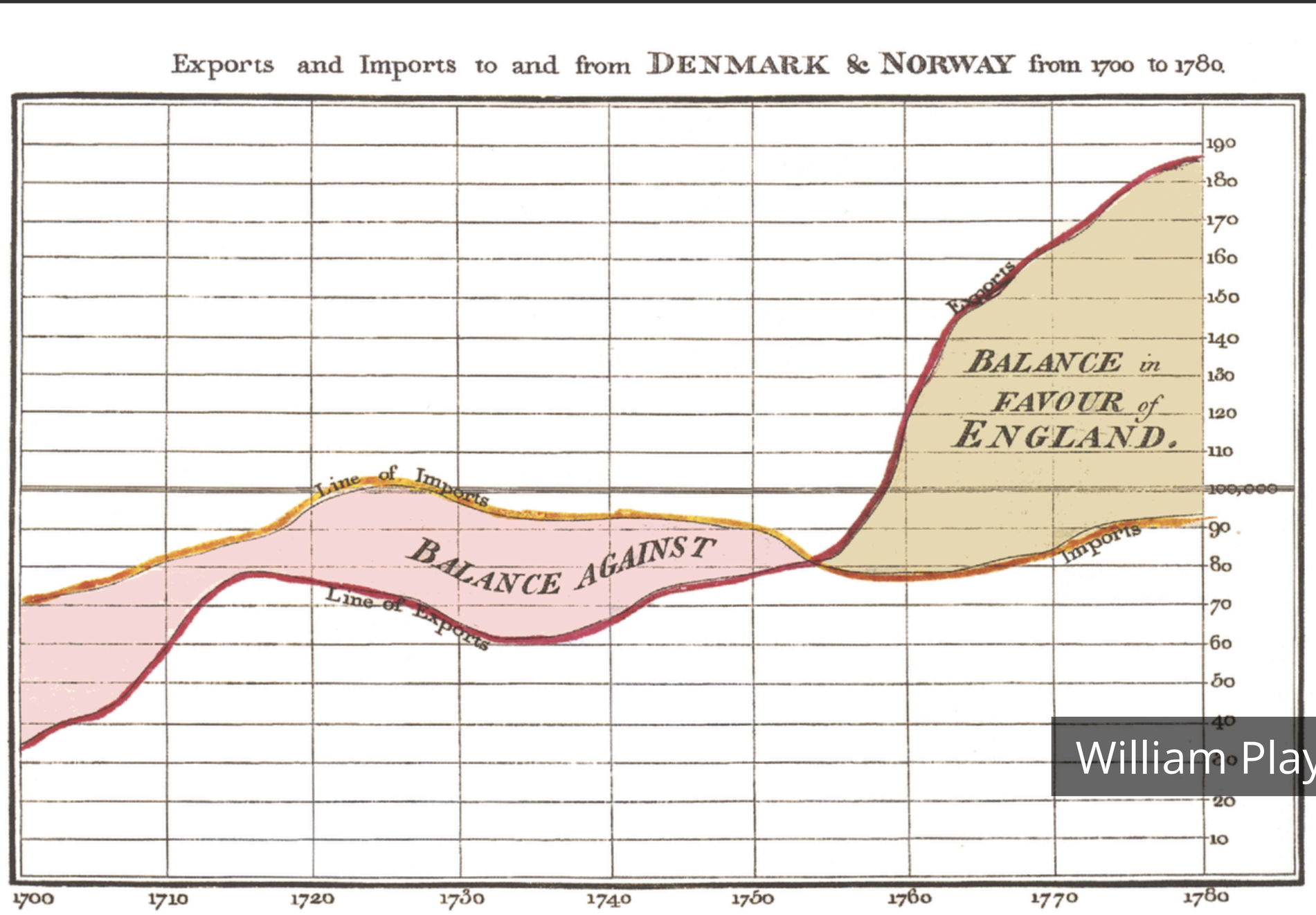
Nightingale's Coxcomb of Crimean War Deaths 1867

Napoleon's March to Moscow [Charles Joseph Minard 1812]



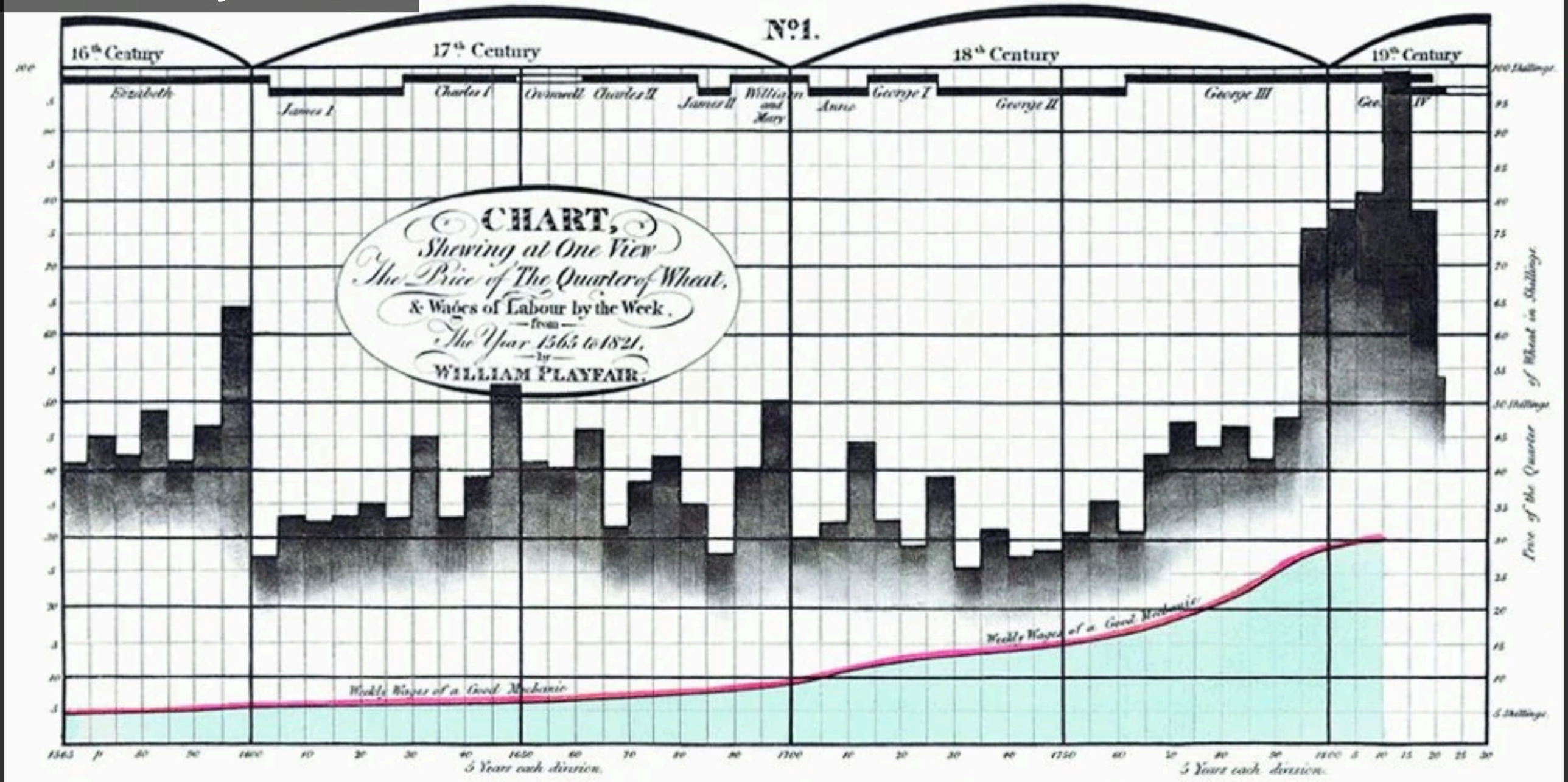
[Joseph Priestley 1765]

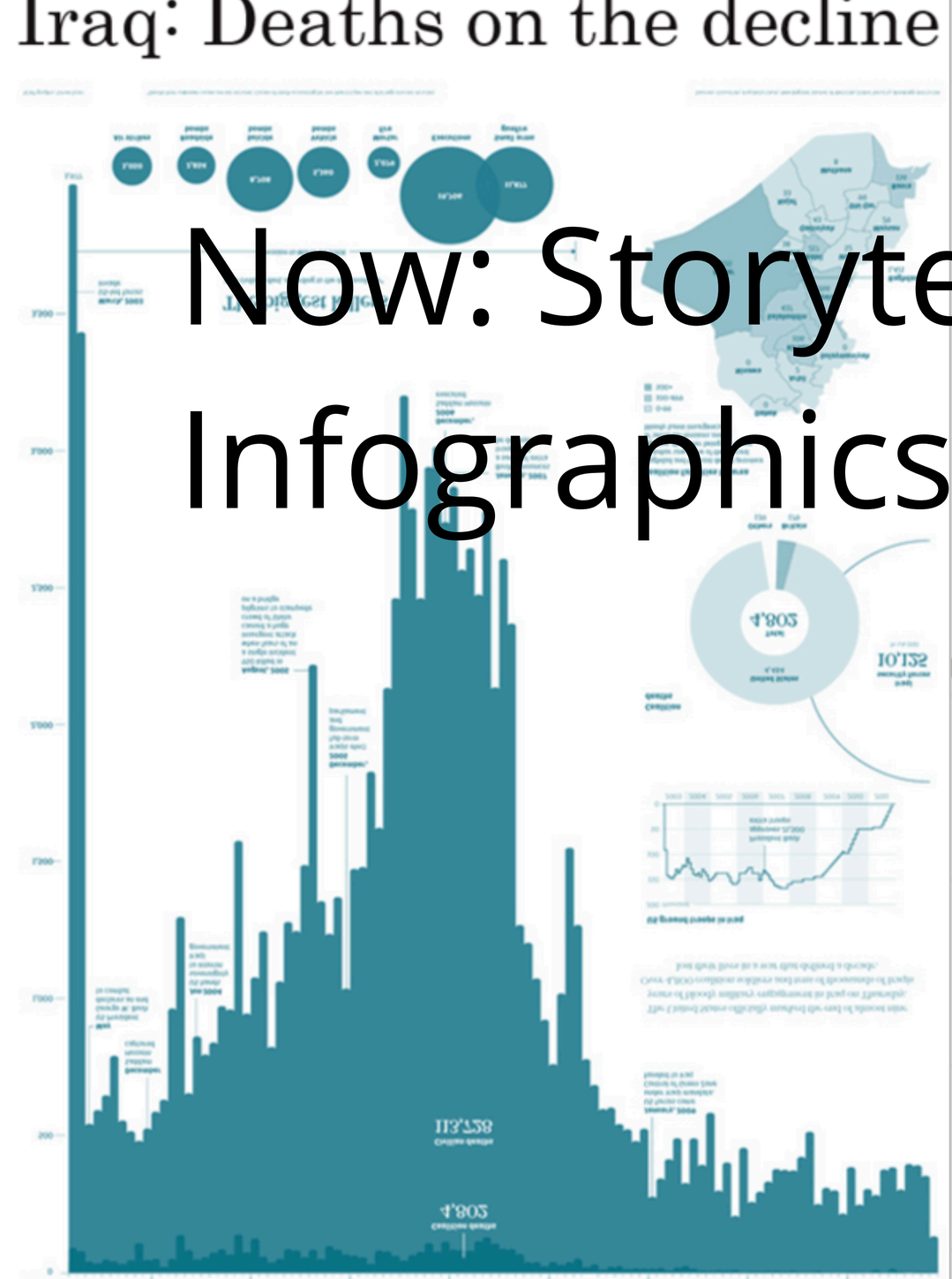




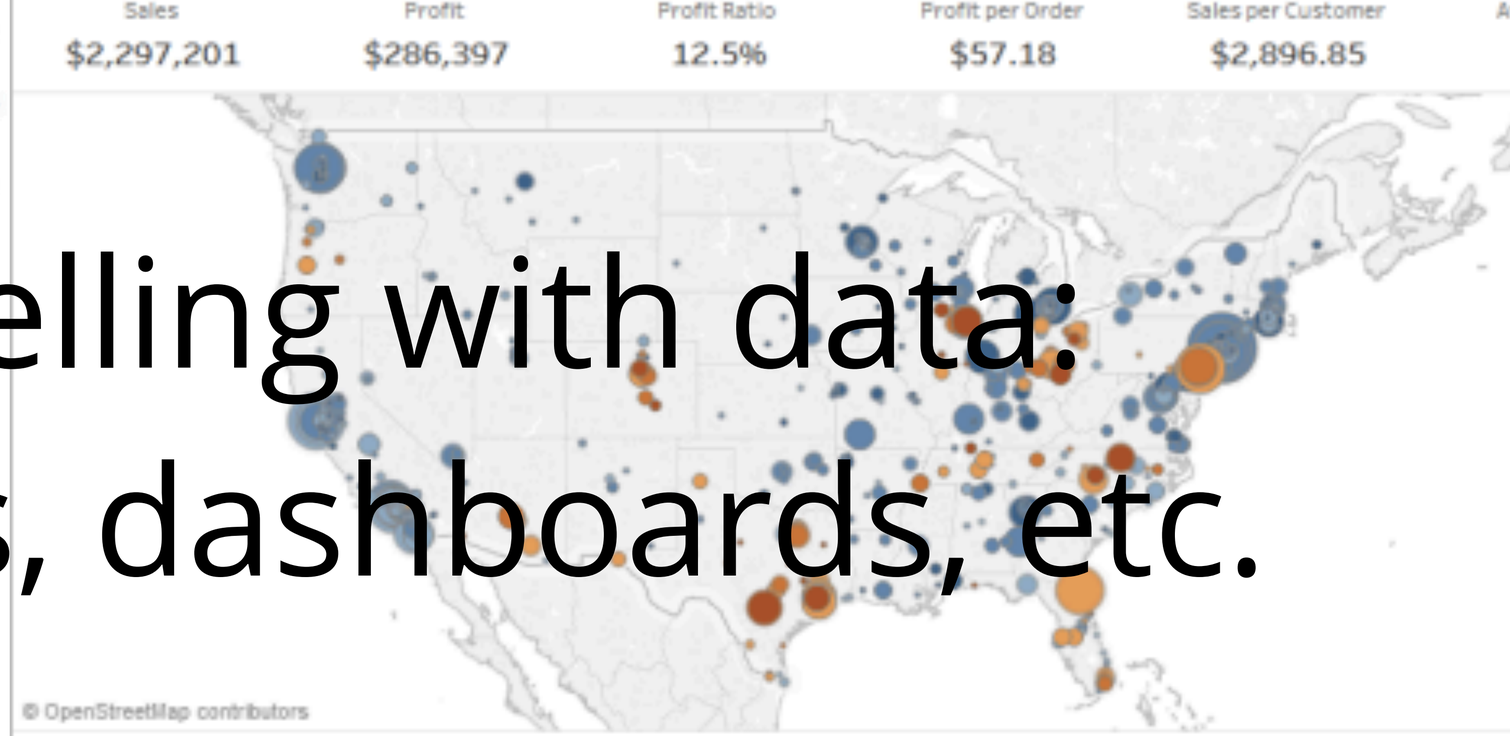
William Playfair 1786

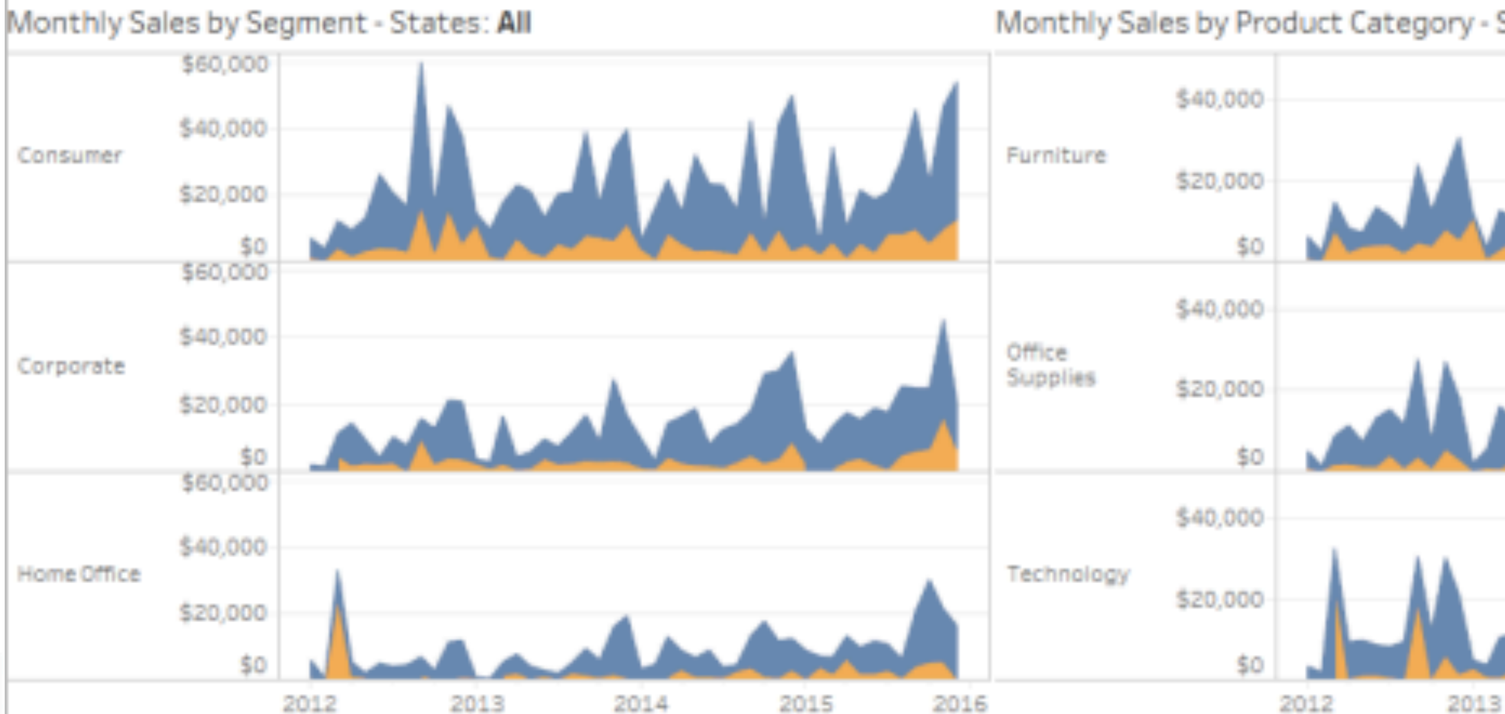
William Playfair 1821





Now: Storytelling with data: Infographics, dashboards, etc.





The Value of Visualization

Record information

- Blueprints, photographs, seismographs, ...

Analyze data to support reasoning

- Develop and assess hypotheses
- Explore patterns and discover the unknown
- Expand memory

Communicate information to others

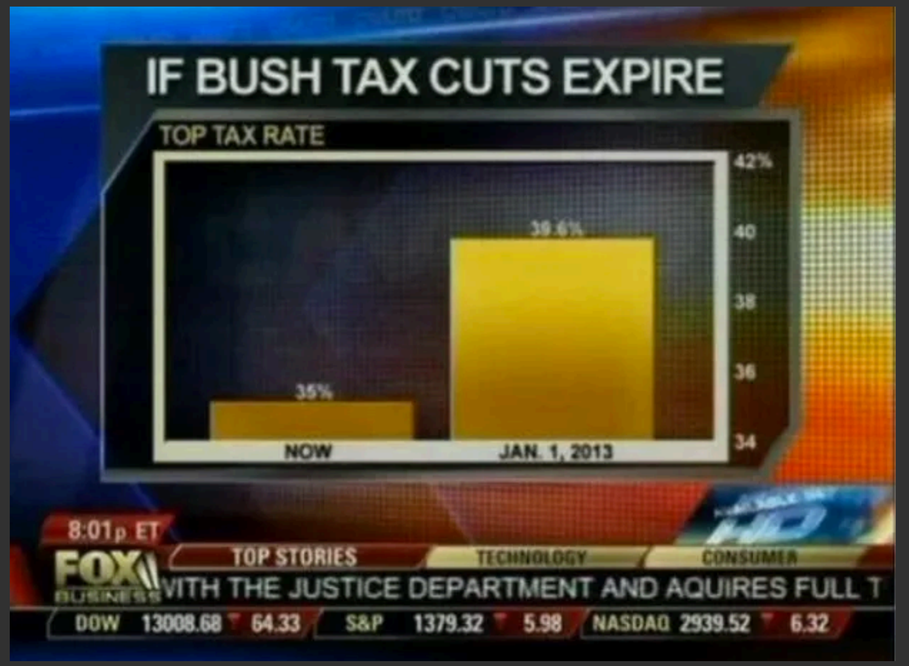
- Explain and persuade
- Share and inspire

Goals of Visualization Research

Understand how people perceive/comprehend visualizations

Develop principles and techniques for effective visualizations

Next



Is this good, bad or weird?

Data Visualization:
The Good, the Bad, the Weird

5 min break