

Learning to Visualize: Surviving in the World of Data

Nam Wook Kim

Mini-Courses — January @ GSAS
2018

Who am I?

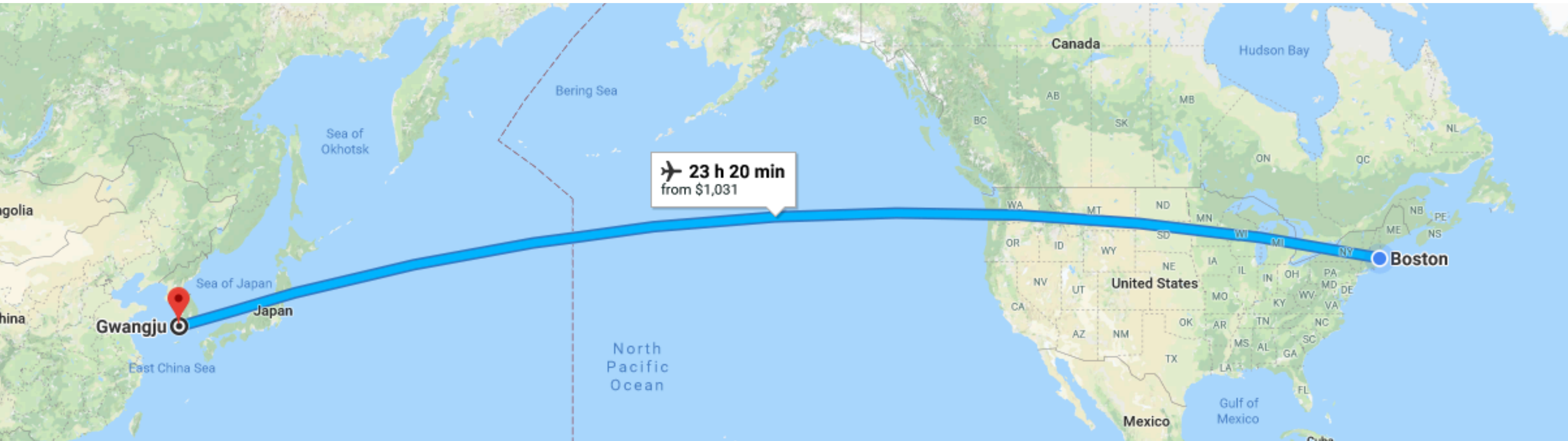
Nam Wook Kim

Ph.D. Student

Computer Science Department

Information Visualization & Human-Computer Interaction

or **Human-Data Interaction**



About You

Goal

To learn how to
visualize data **effectively**

Goal

To learn how to

evaluate visualization design

Goal

Not to learn about Tableau

IACS ComputeFest Workshop:

Introduction to Tableau

Wednesday, January 11, 2017

12:00 PM - 2:30 PM

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

The Value of Visualization

Big Data, **Small** Data, Data **Everywhere**



Health & Medicine

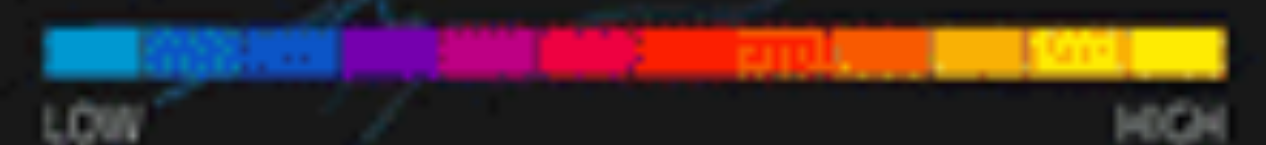
Separate Trips

uberPOOL Trips

Transportation

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

TRAFFIC VOLUME



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

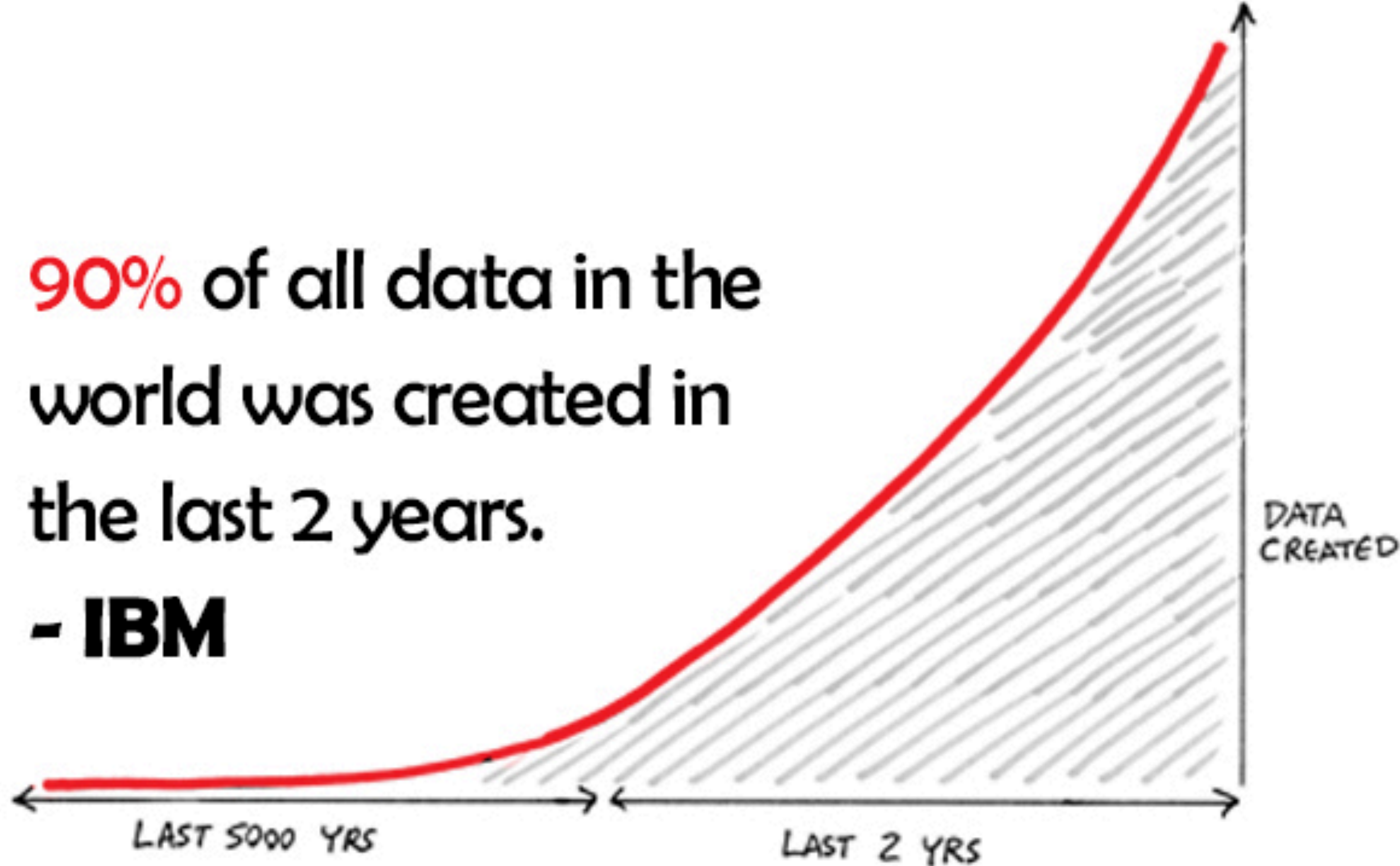
🚶 **Walking - 0.2 mi** 2 mins



Human Activity

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, ... Because now we really do have essentially **free** and **ubiquitous** data.”

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



A Poverty of Attention

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

Herbert A. Simon
Economist & Psychologist

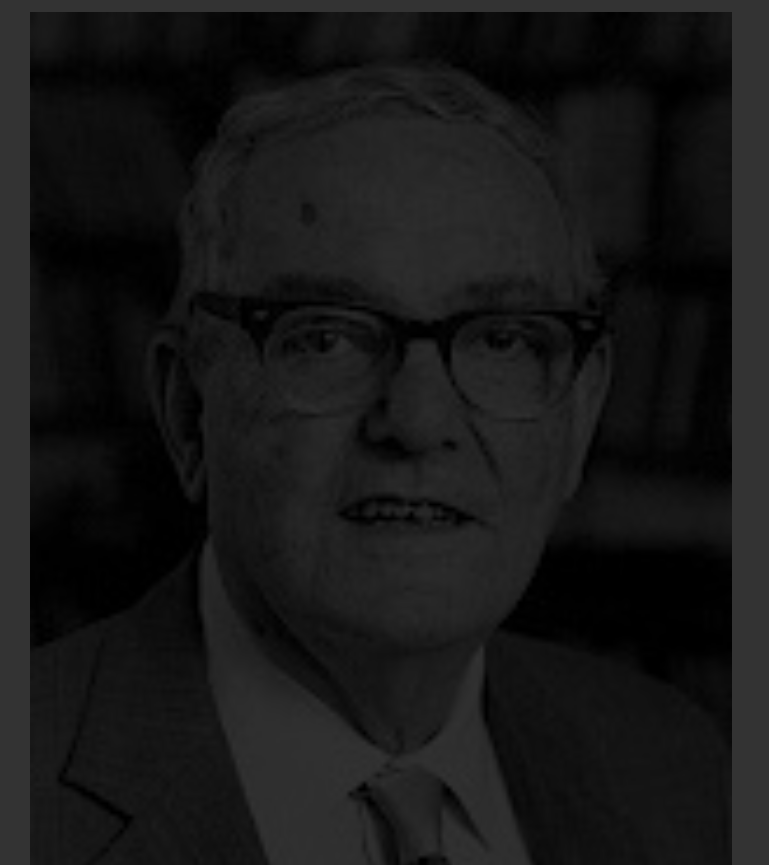


A Poverty of Attention

Visualization can help!

“...Information consumes the attention of its recipients. provides a powerful yet accessible way to make sense of large and complex data that attention efficiently among the overabundance of information sources that might consume it.”

Herbert A. Simon
Economist & Psychologist



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 of data

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

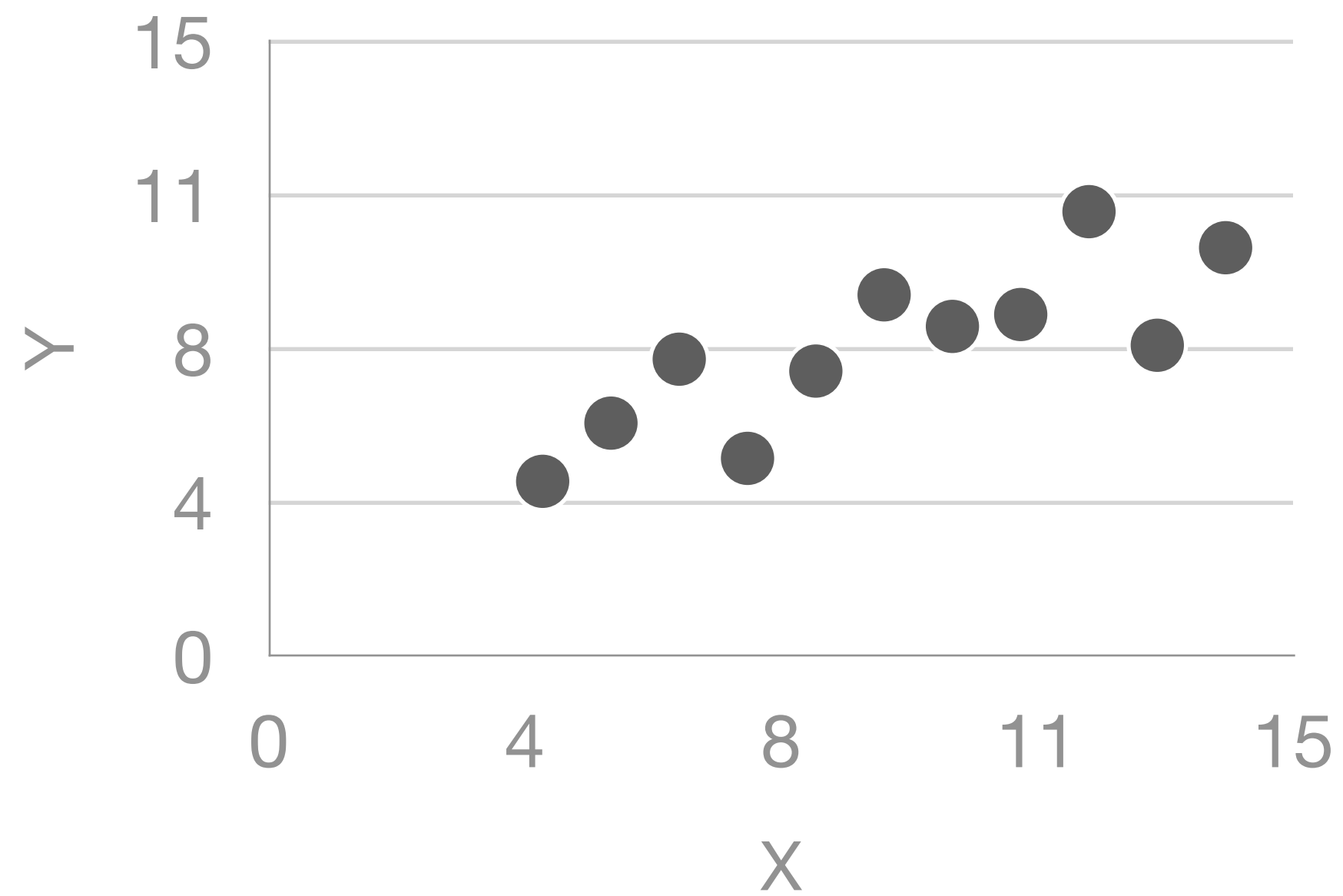
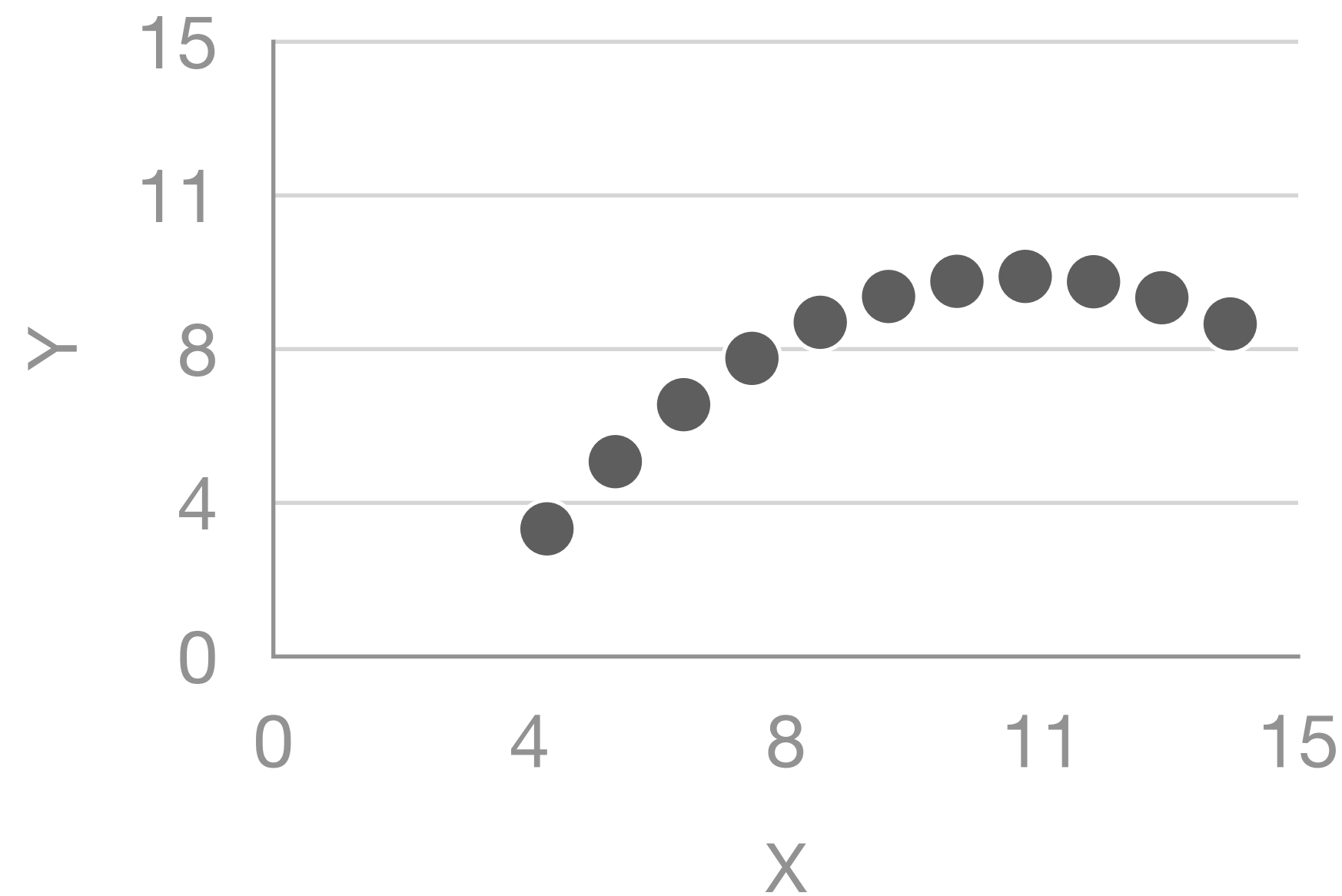
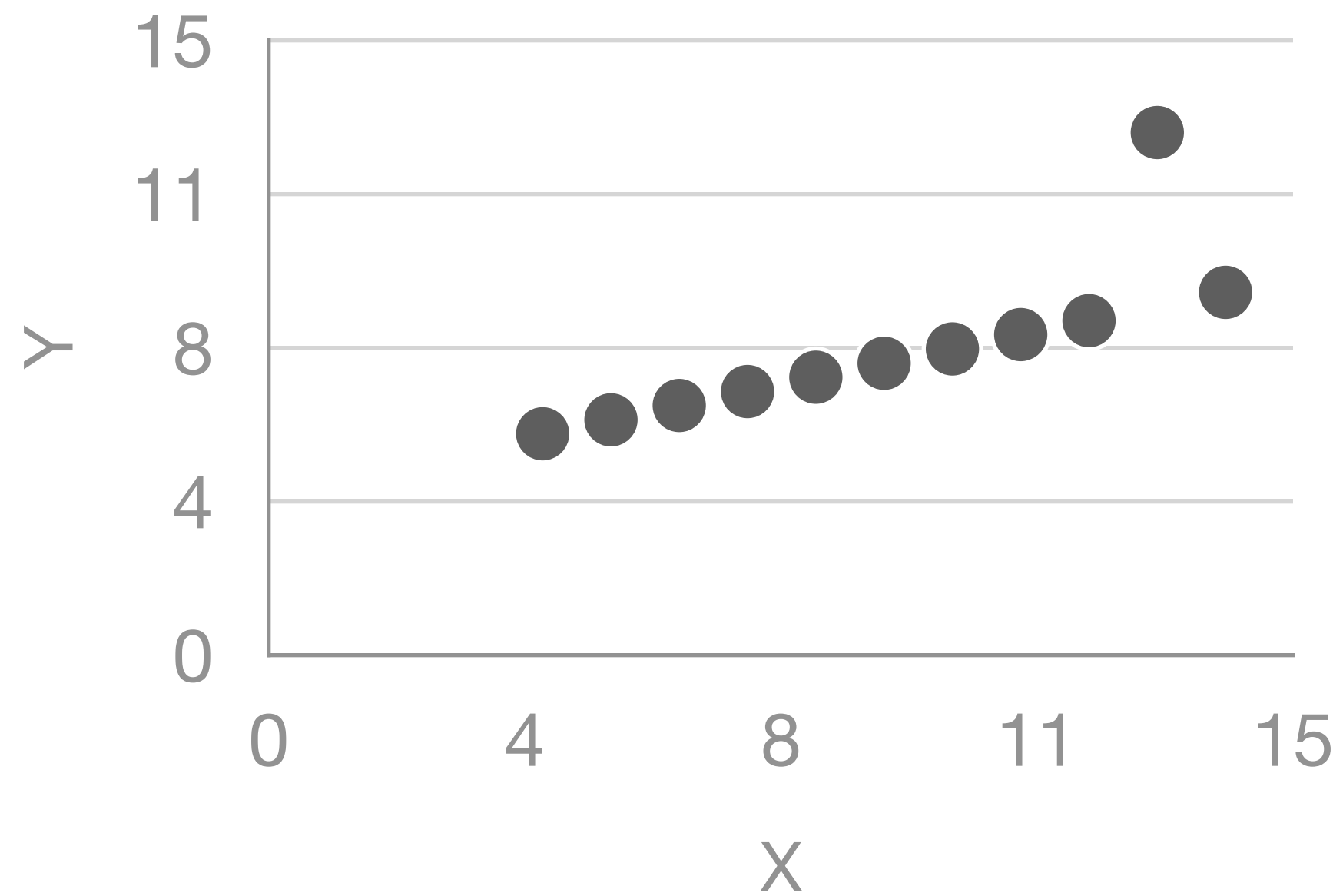
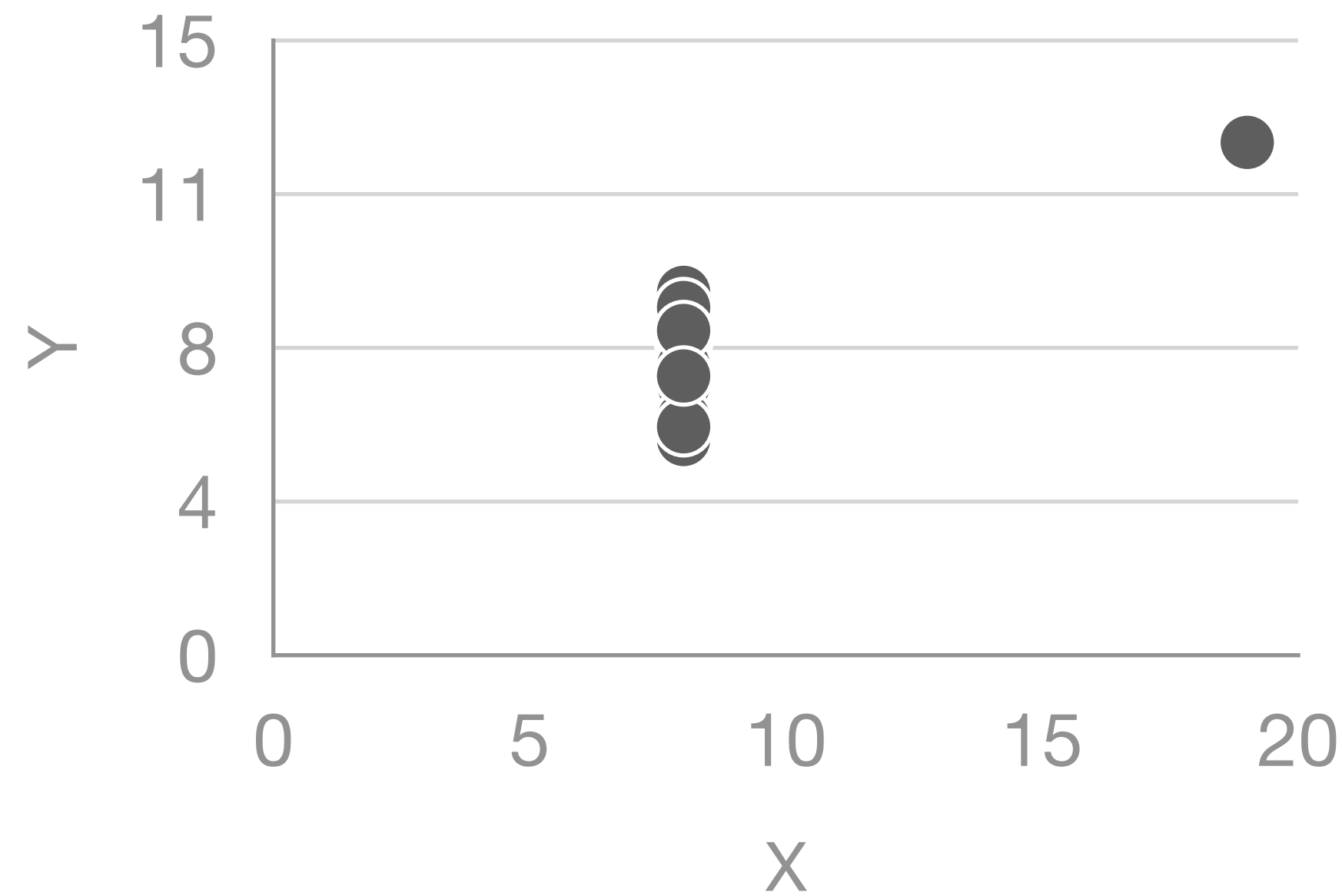
$$\mu_X = 9.0 \quad \sigma_X = 3.317$$

$$\mu_Y = 7.5 \quad \sigma_Y = 2.03$$

Linear Regression

$$Y = 3 + 0.5 X$$

$$R^2 = 0.67$$

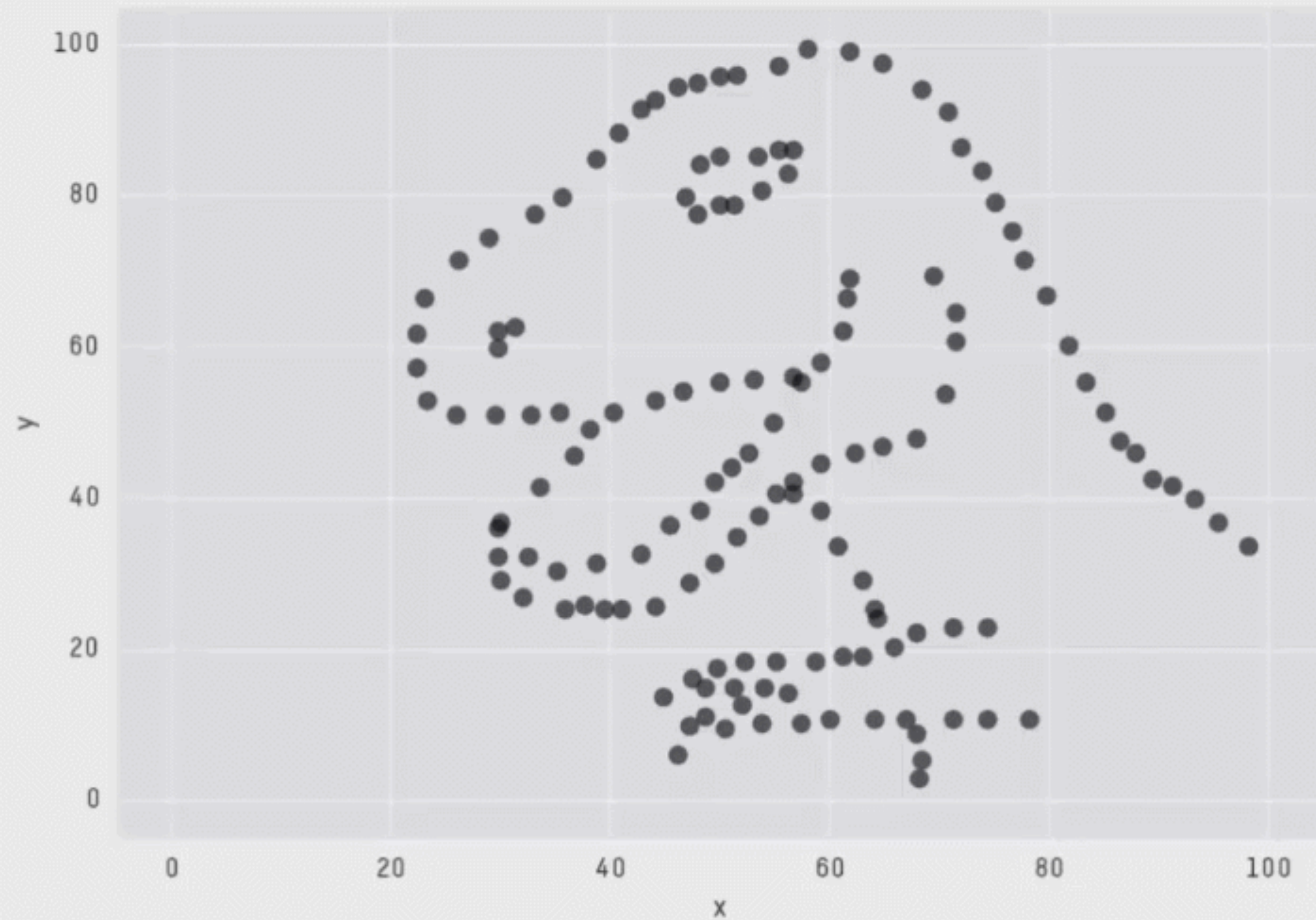
A*B**C**D*

“

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

—F. J. Anscombe, 1973





```
X Mean: 54.2659224  
Y Mean: 47.8313999  
X SD  : 16.7649829  
Y SD  : 26.9342120  
Corr. : -0.0642526
```

*clearly different and visually distinct datasets with
same statistical properties*

[Matejka & Fitzmaurice 2017]

Why Create Visualizations?

Why Create Visualizations?

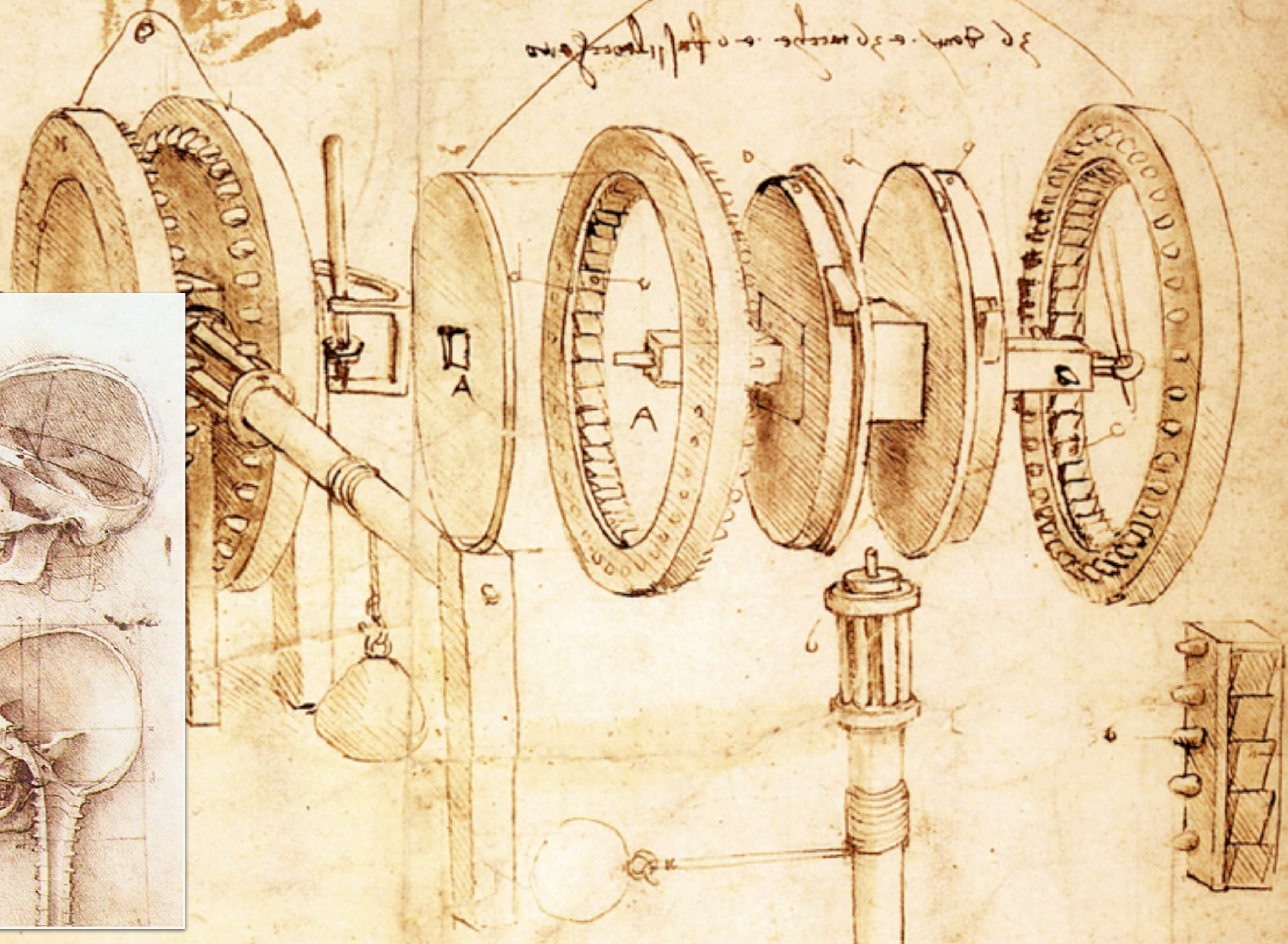
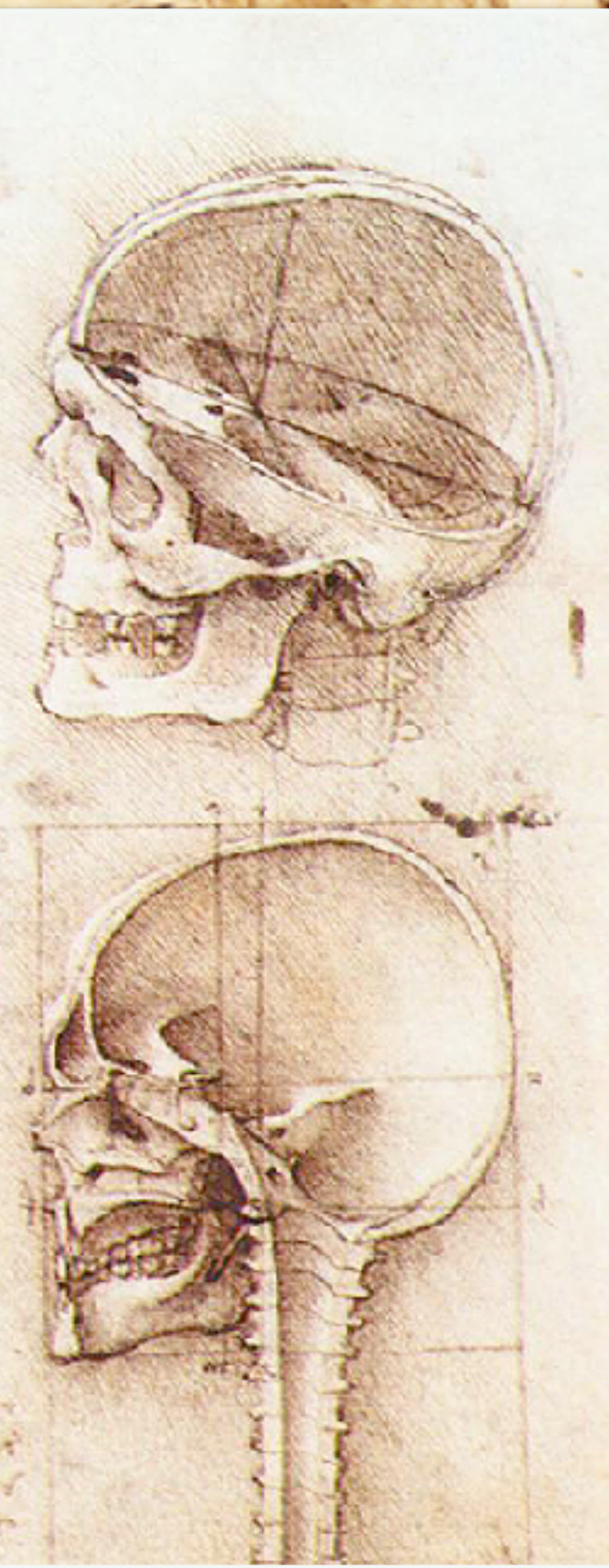
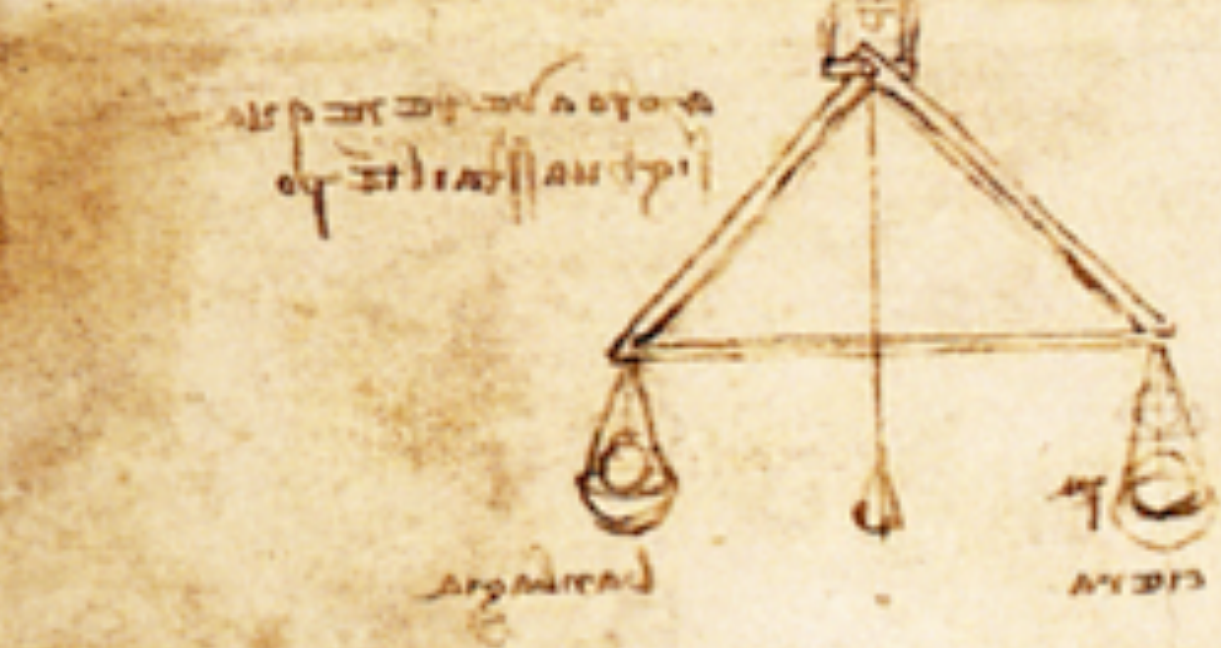
- Answer questions (or discover them)
- Make decisions
- See data in context
- Expand memory
- Support graphical calculation
- Find patterns
- Present argument or tell a story
- Inspire

Three functions of visualization

1. Record
2. Analyze
3. Communicate

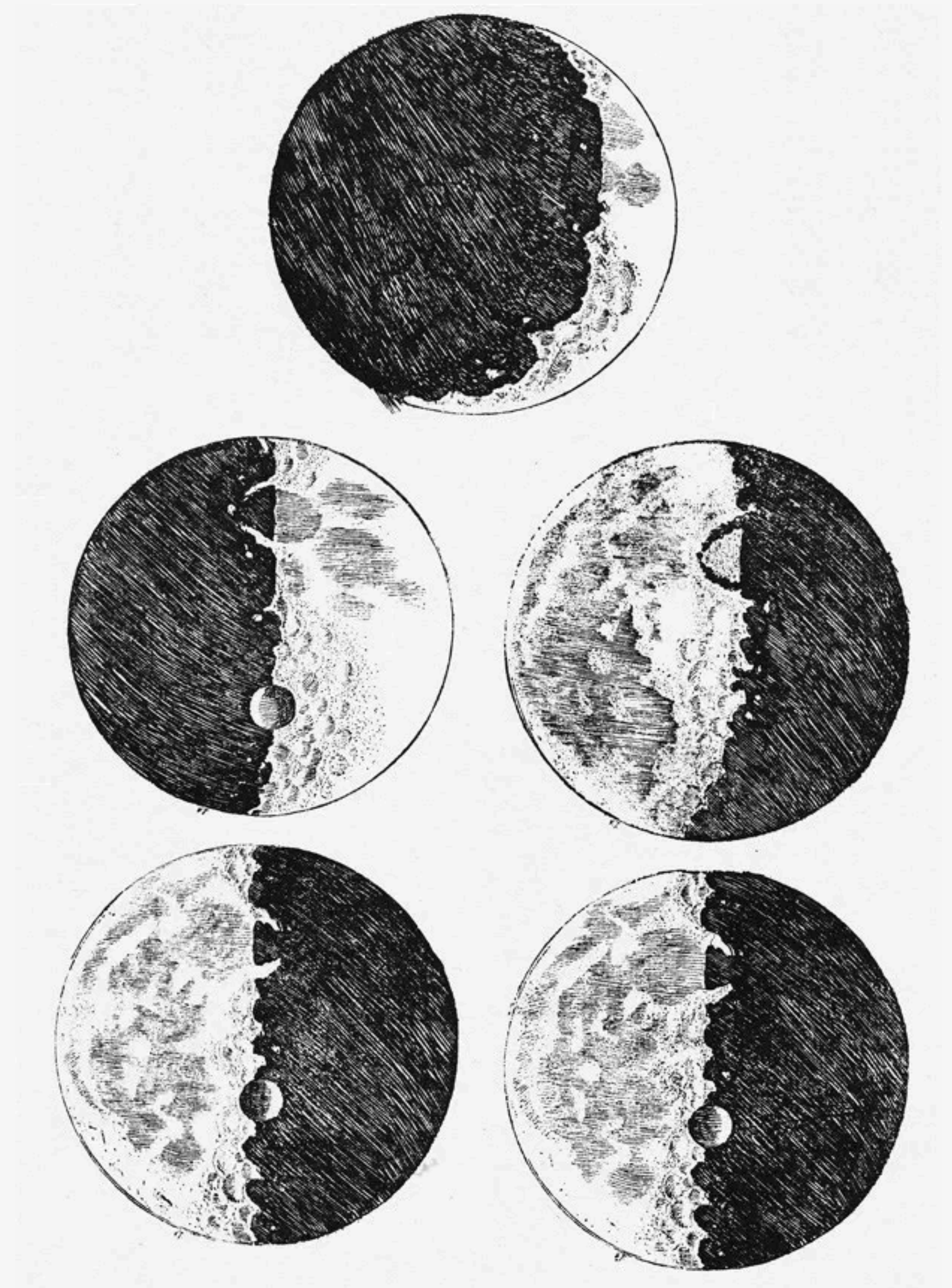
Record Information

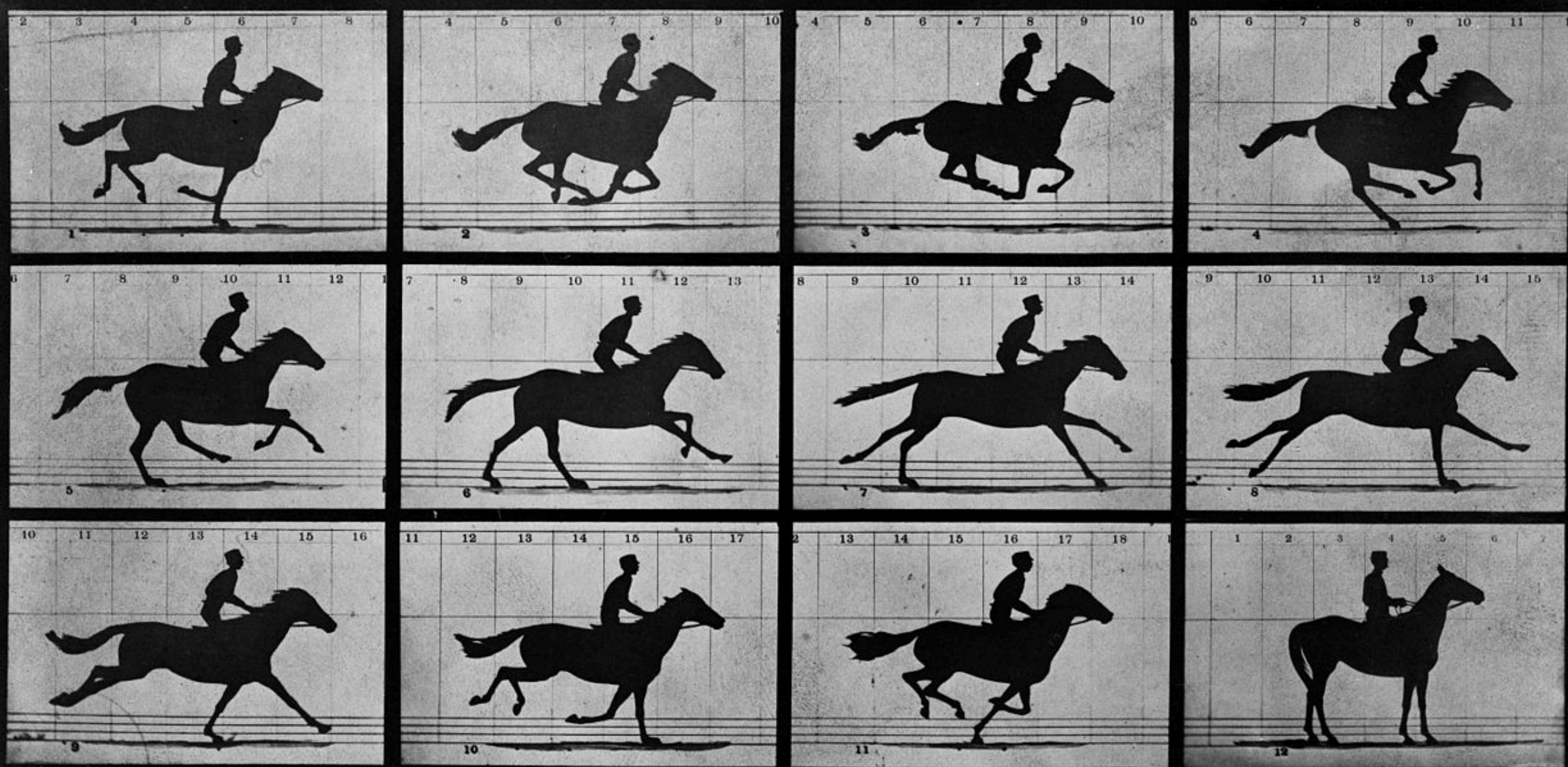
Leonardo da Vinci 1485



Galileo Galilei's Sketches of the Moon

(November-December 1609)





Copyright, 1878, by MUYBRIDGE.

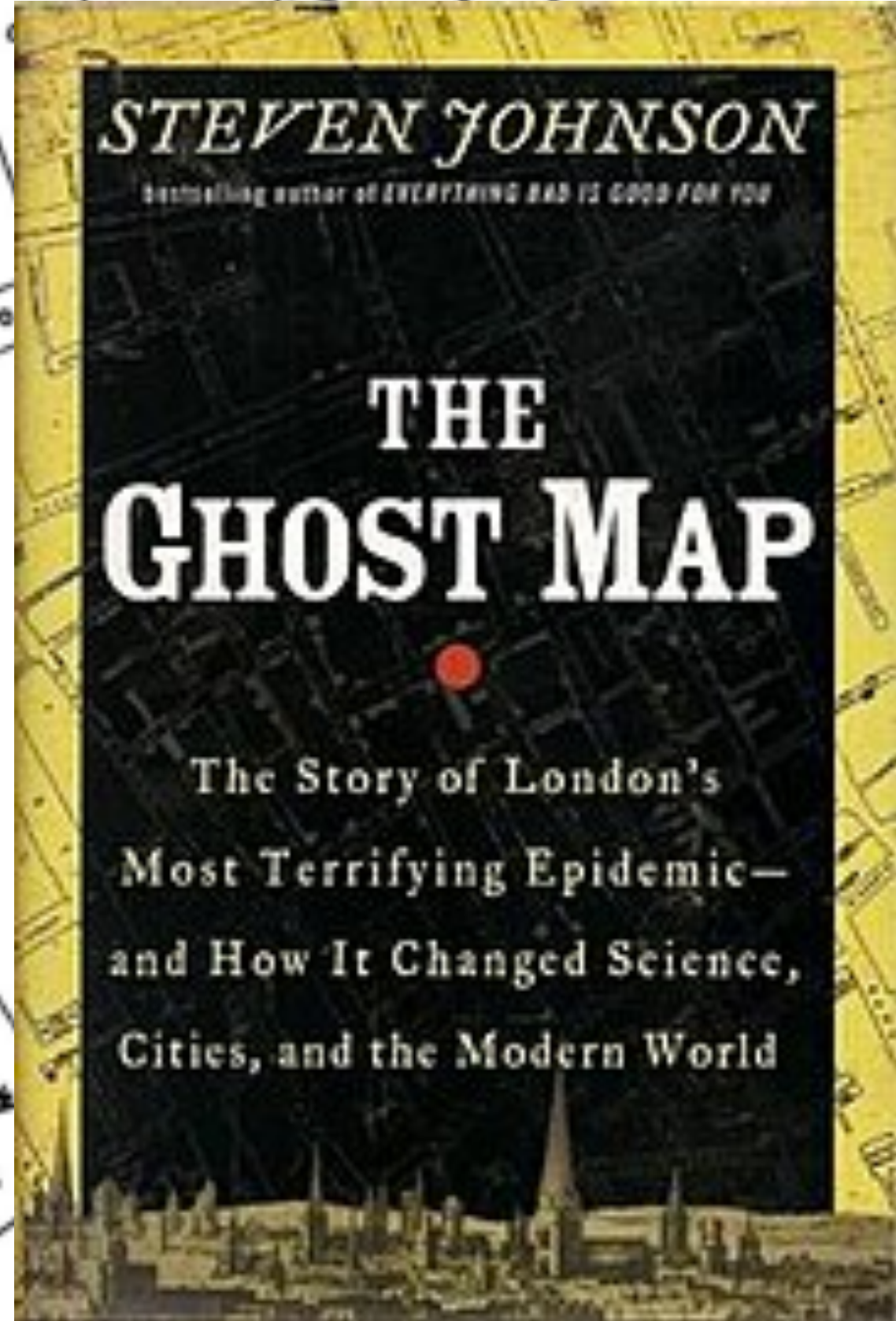
MORSE'S Gallery, 417 Montgomery St., San Francisco.

THE HORSE IN MOTION.

E. J. Muybridge, 1878

Illustrated by

Support Reasoning

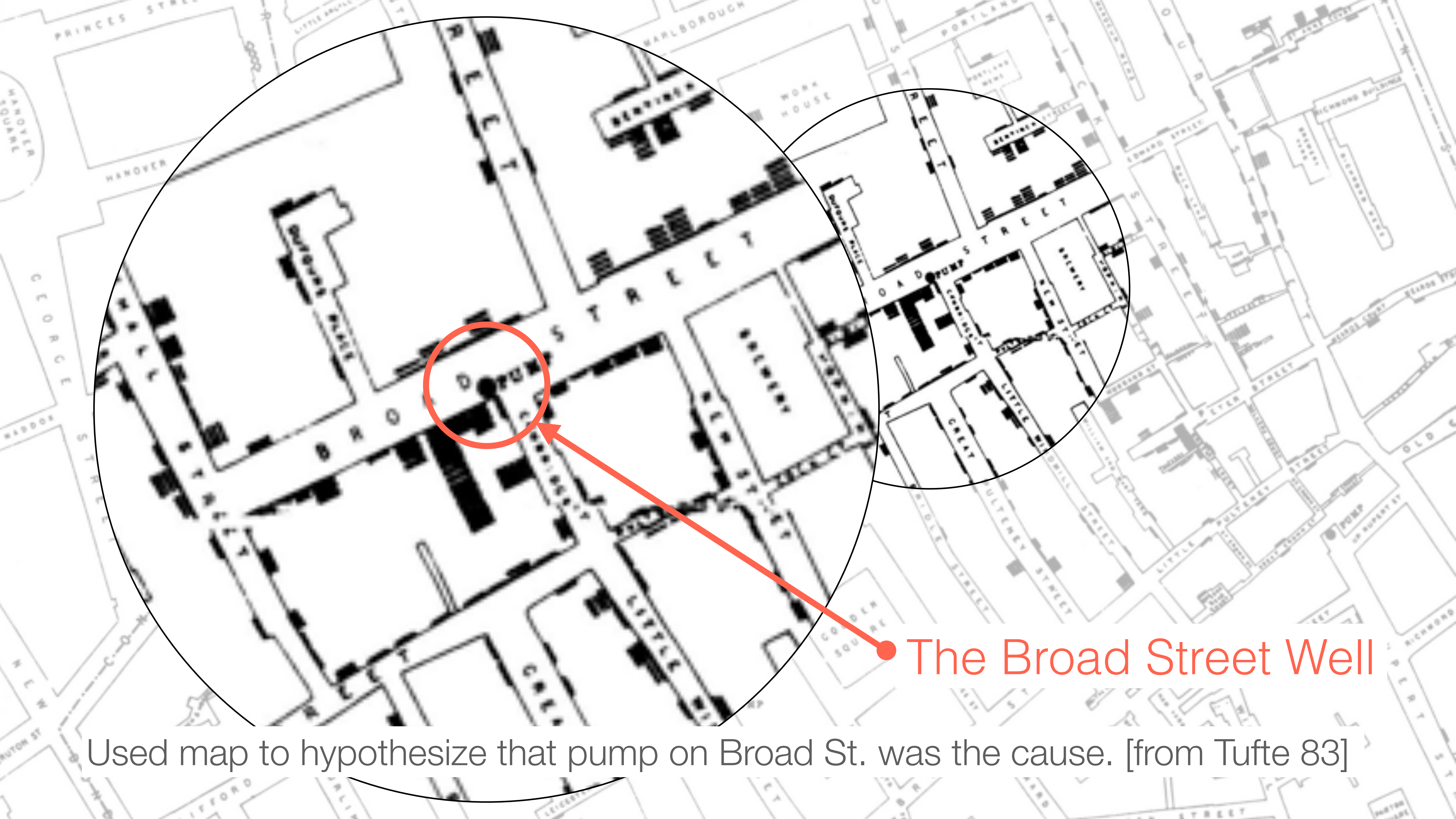


John Snow, the Cholera Epidemic 1854

Seeing Data in Context



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



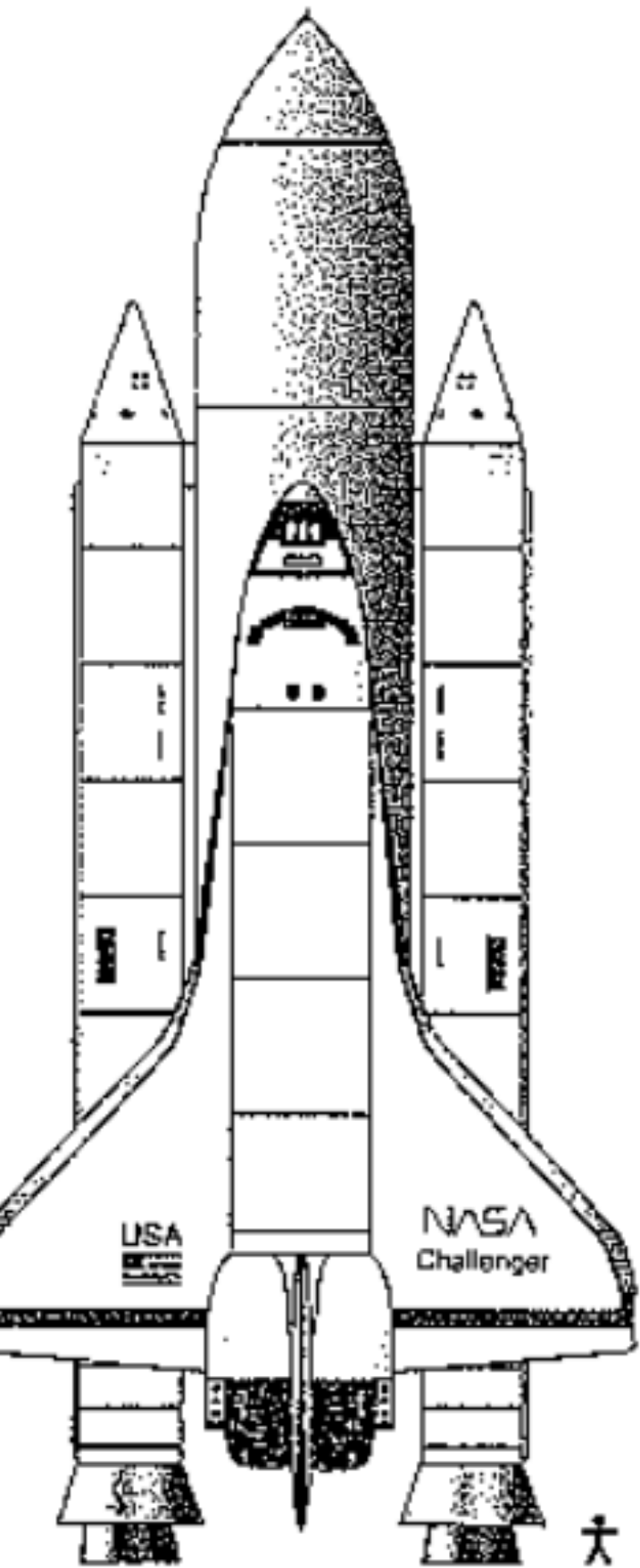
The Broad Street Well

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

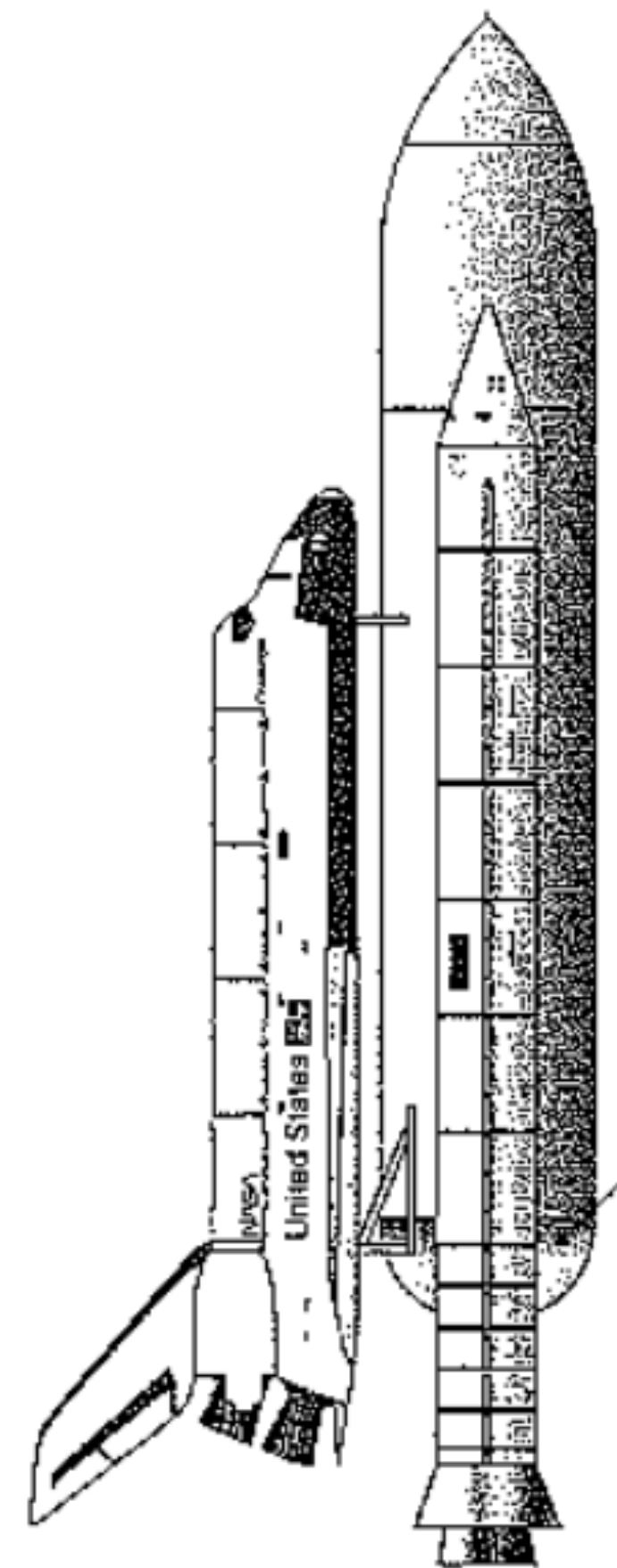
Space Shuttle Challenger Disaster (1986)



approx. 73 seconds after

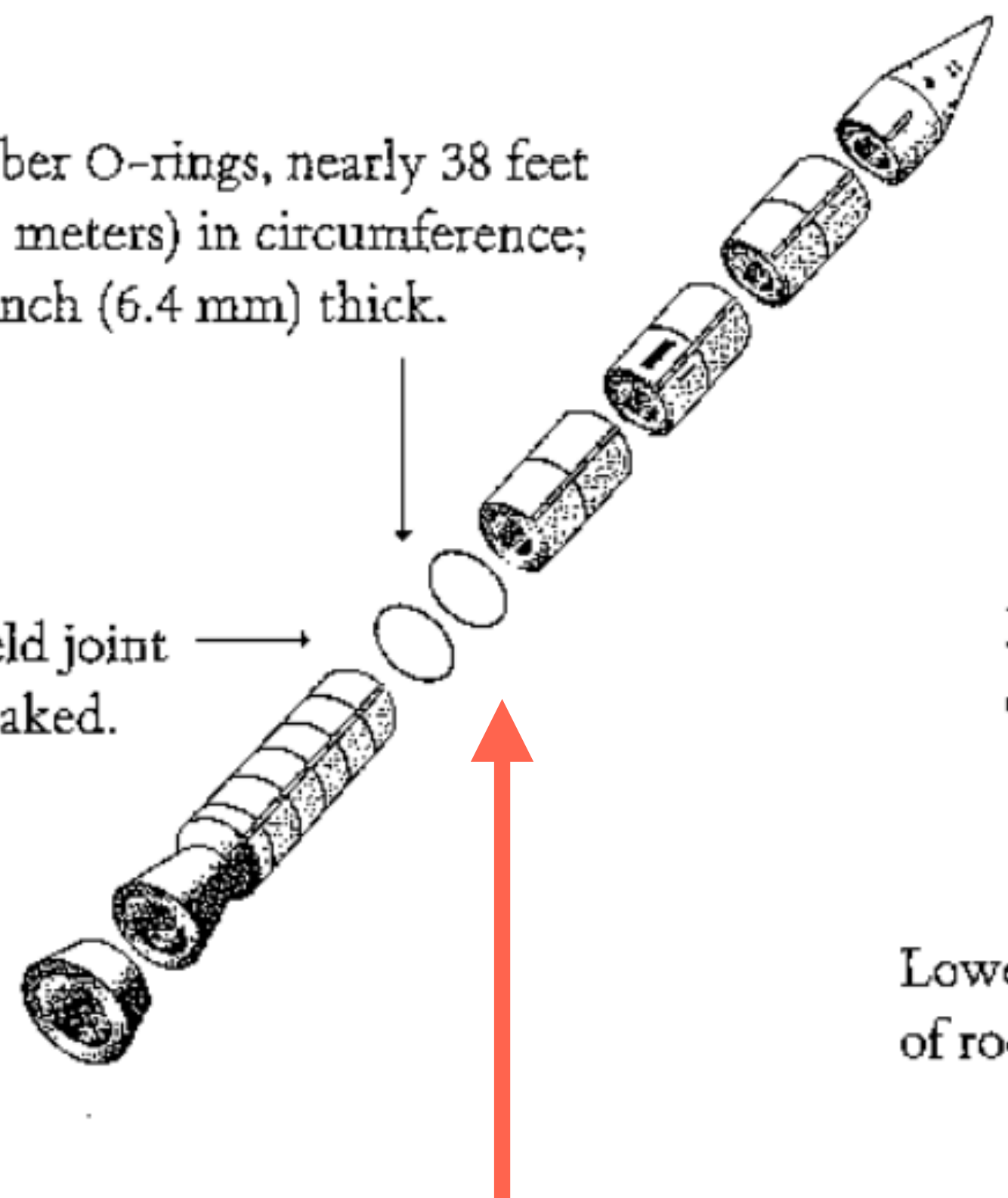


185 feet (56.4 meters)



Rubber O-rings, nearly 38 feet (11.6 meters) in circumference; 1/4 inch (6.4 mm) thick.

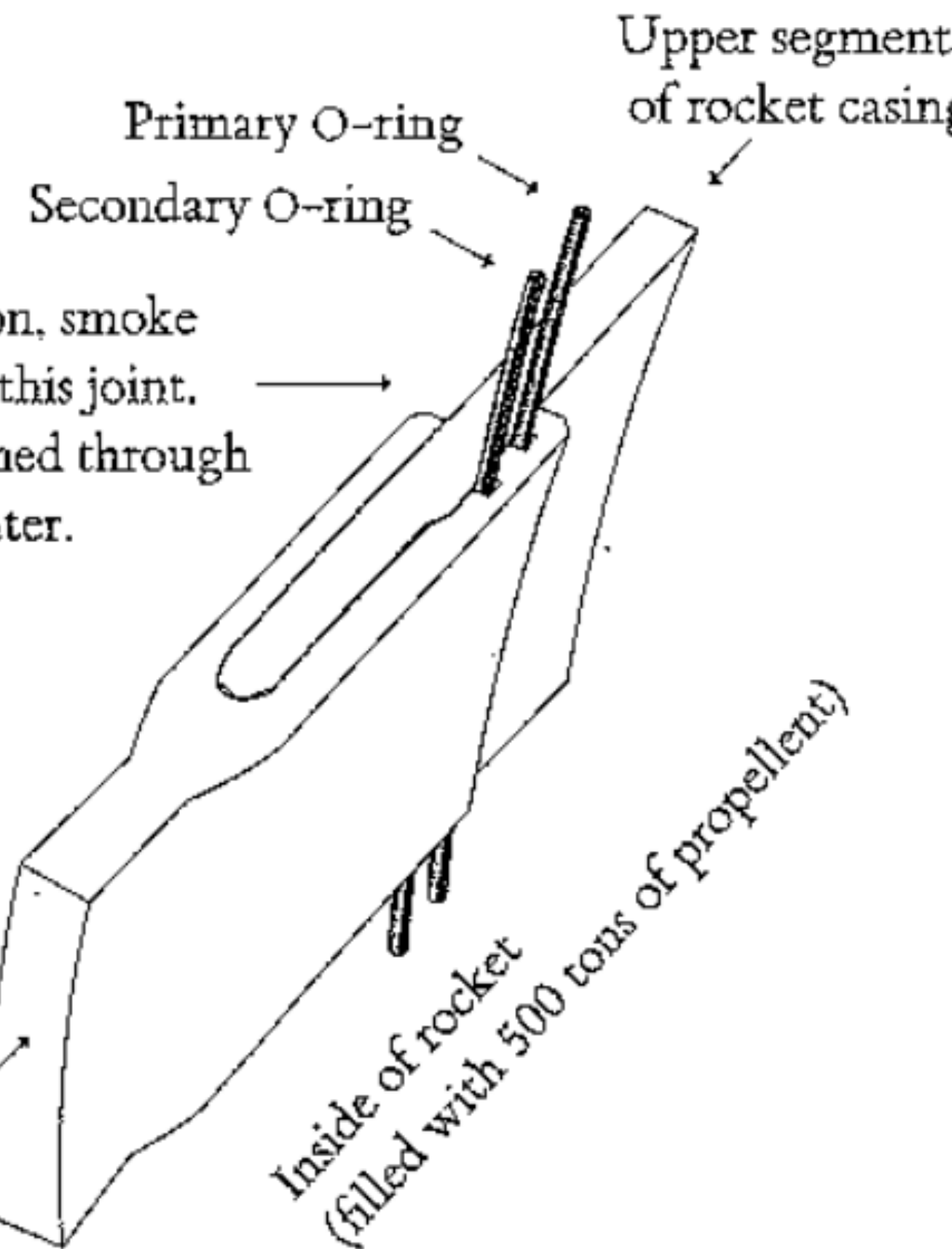
The field joint that leaked.



Upon ignition, smoke leaked from this joint. A flame burned through 59 seconds later.

Exterior wall of rocket

Lower segment of rocket casing



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

1169

OCT 30, 1985

85

113

AFT

SRM No.	Cross Sectional View			Top View		Clocking Location (deg)	
	Erosion Depth (in.)	Perimeter Affected (deg)	Nominal Dia. (in.)	Length Of Max Erosion (in.)	Total Heat Affected Length (in.)		
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	--
418 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.

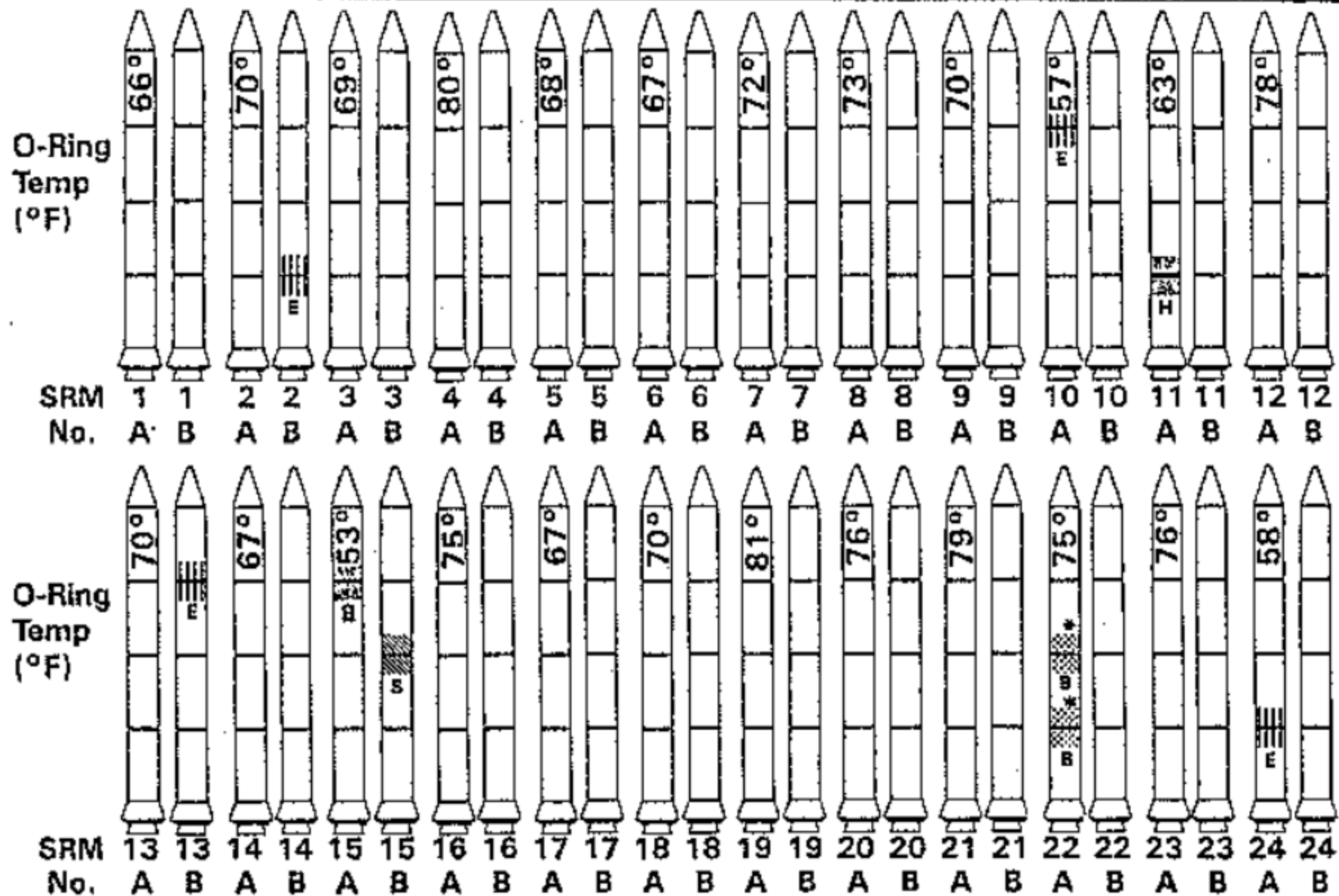
***Soot behind primary O-ring, heat affected secondary O-ring.

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

-  = Heating of Secondary O-Ring
-  = Primary O-Ring Blowby
-  = Primary O-Ring Erosion
-  = Heating of Primary O-Ring
-  = No Damage

STATIC TEST MOTORS

- HORIZONTAL ASSEMBLY
- SOME PUTTY REPAIRED

MORTON THIKOL, INC.
Wasatch Operations

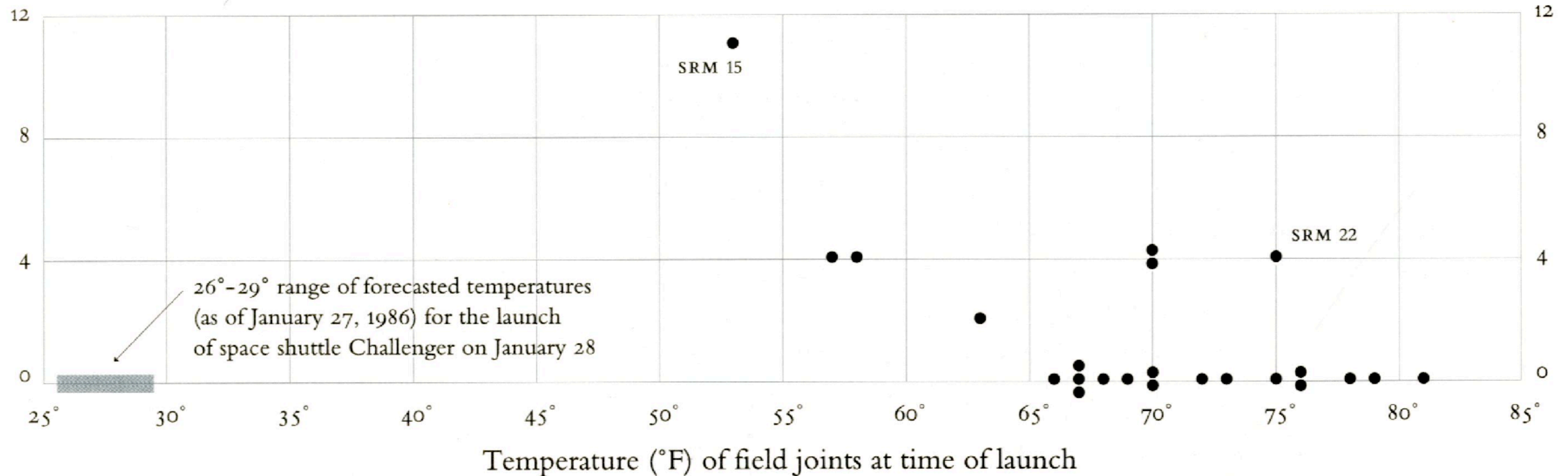
* No Erosion

10-10-10

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

O-ring damage index, each launch



[Edward Tufte 1997]

Expand Memory: Feynman Diagram (1948)

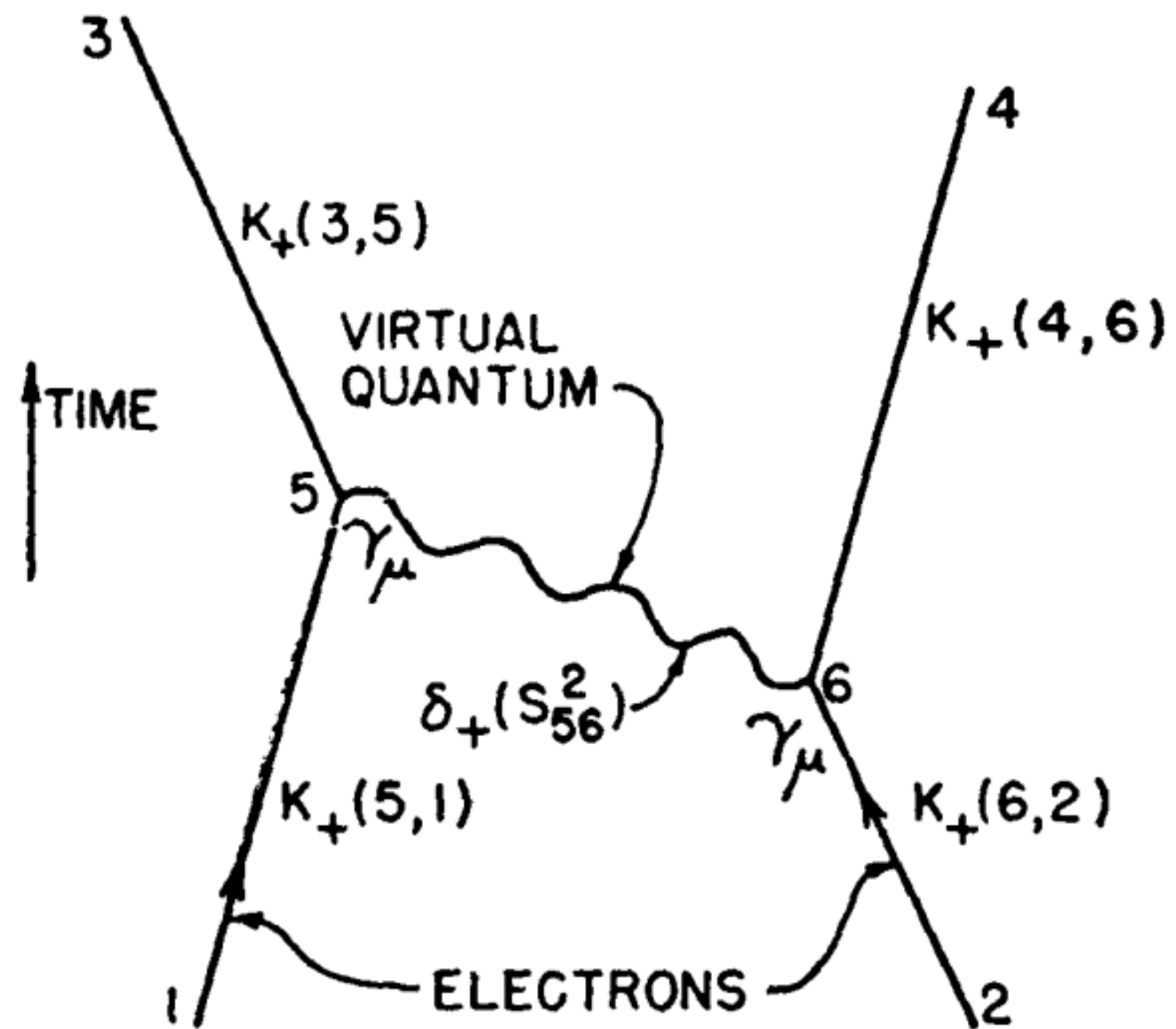
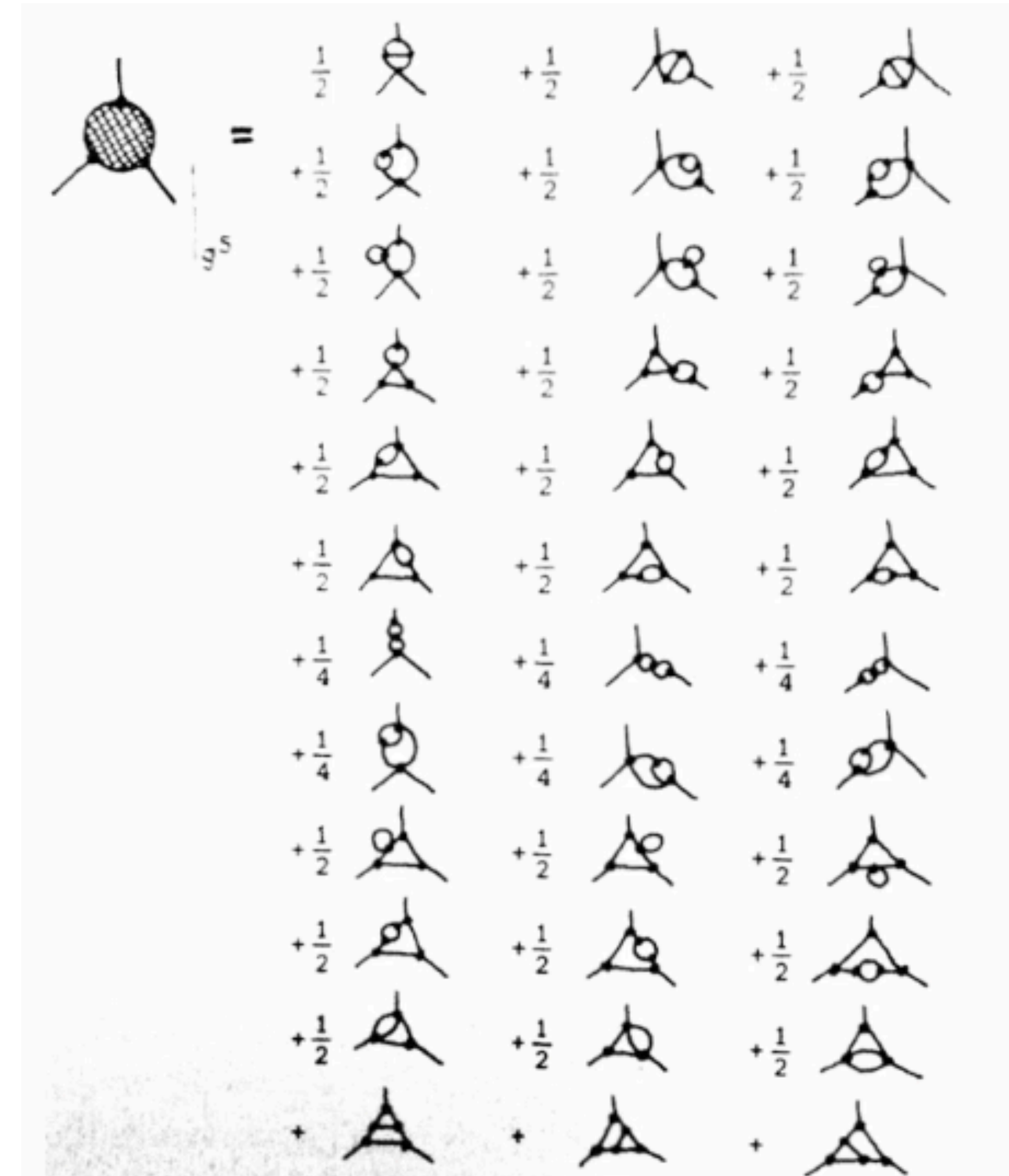


Figure 3. Electron-electron scattering is described by one of the earliest published Feynman diagrams (fea-



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

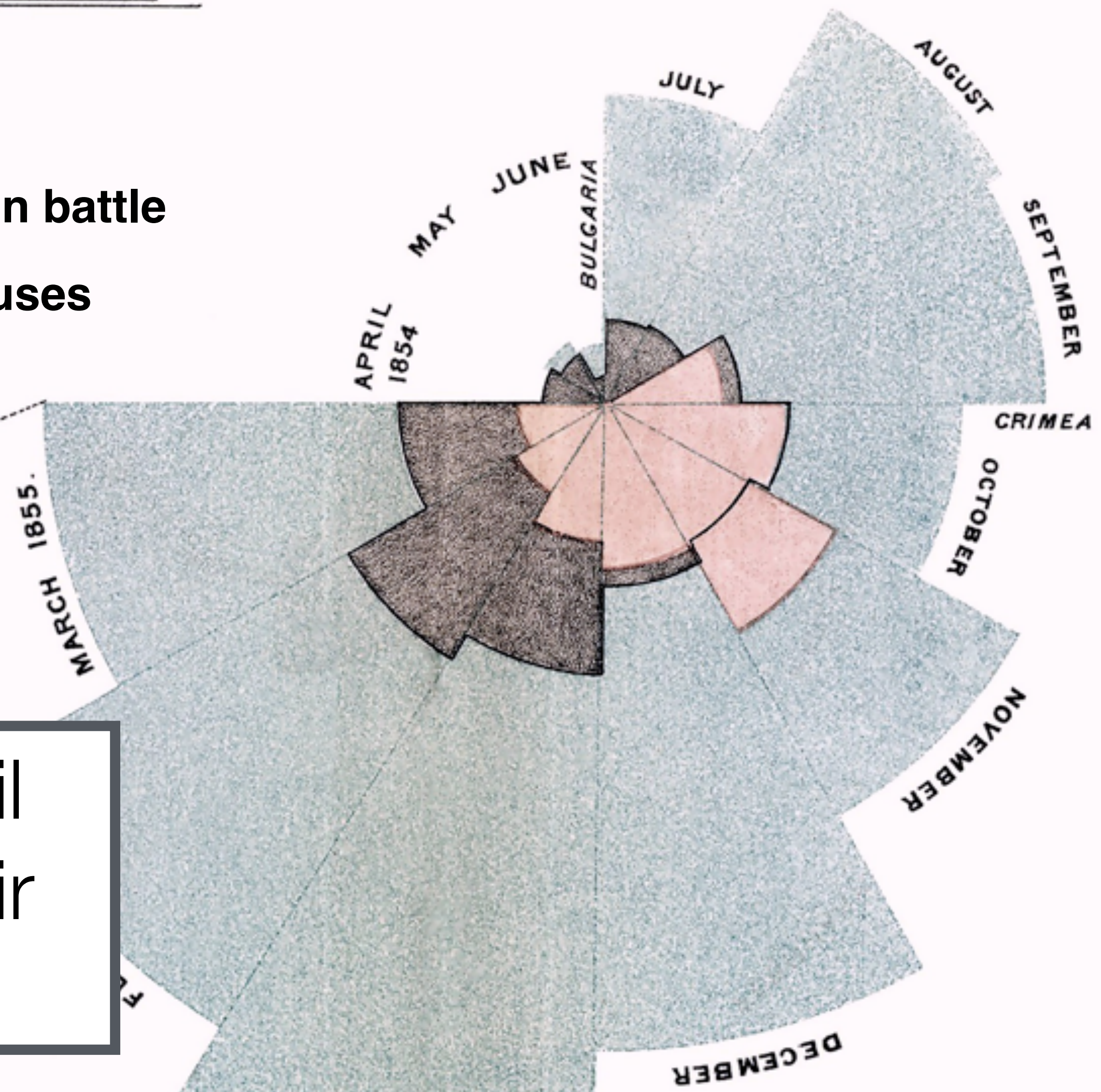
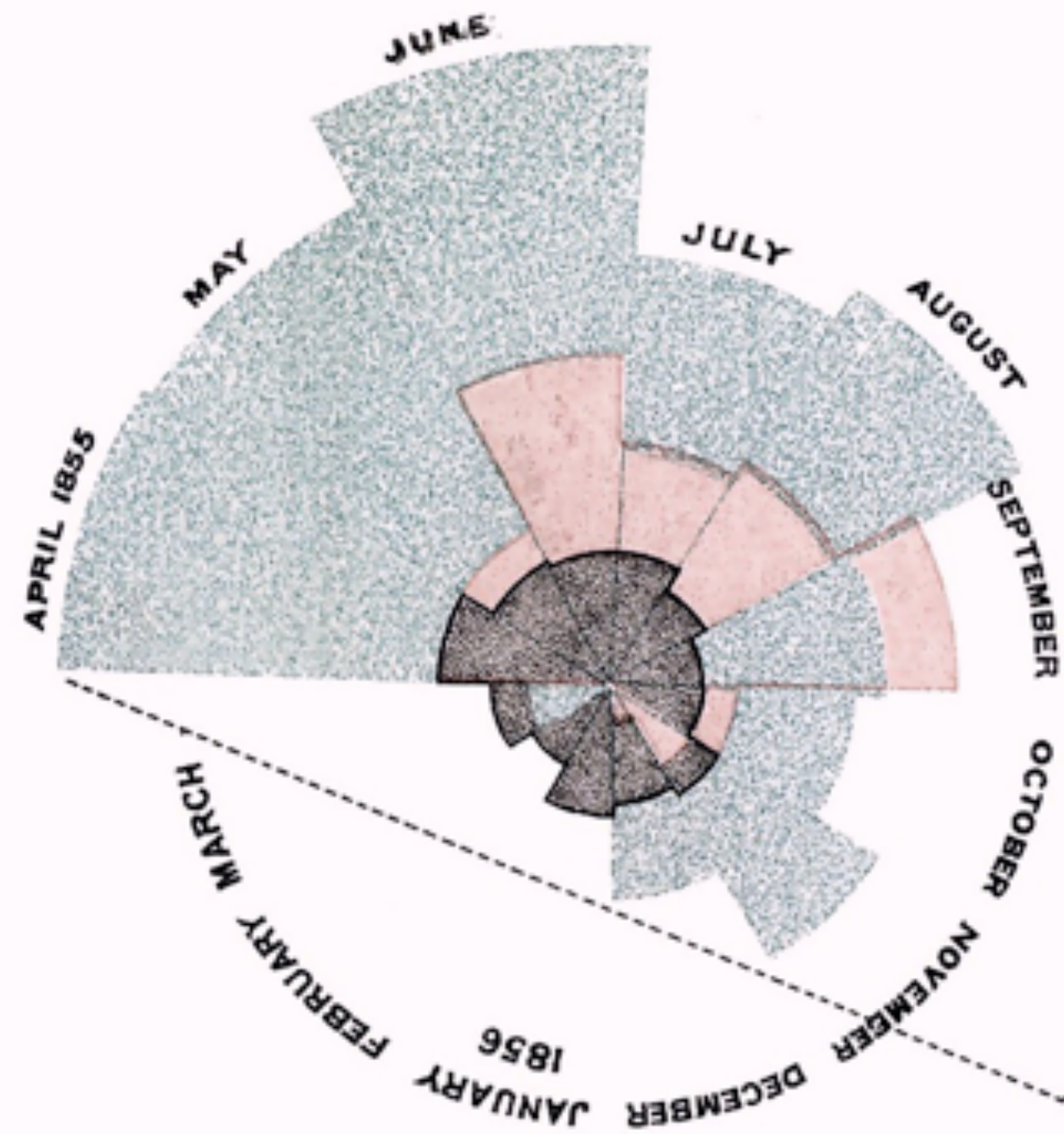
Convey Information to Others

DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

2.
APRIL 1855 TO MARCH 1856.

1.
APRIL 1854 TO MARCH 1855.

- Death from wounds in battle
- Death from other causes
- Death from disease



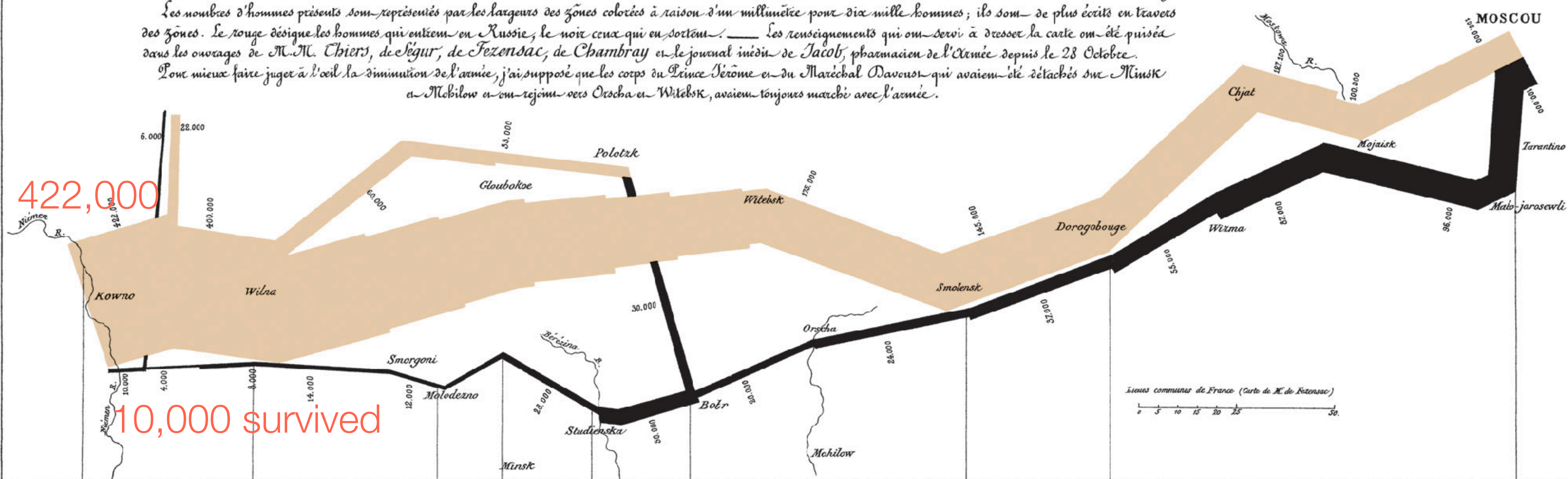
“to affect thro’ the Eyes what we fail to convey to the public through their word-proof ears” - Nightingale

*black wedges measured from the centre the deaths from all other causes.
The black line across the red triangle in Nov. 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,*

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

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 zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. 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. Chiers, de L'égur, 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 Davout qui avaient été détachés sur Minsk et Mohilow et renvoyés vers Orscha et Witebsk, avaient toujours marché avec l'armée.

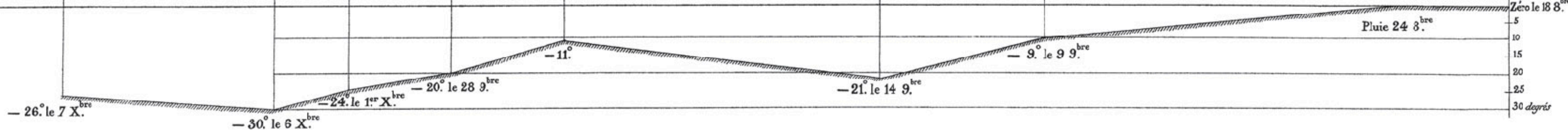


422,000

10,000 survived

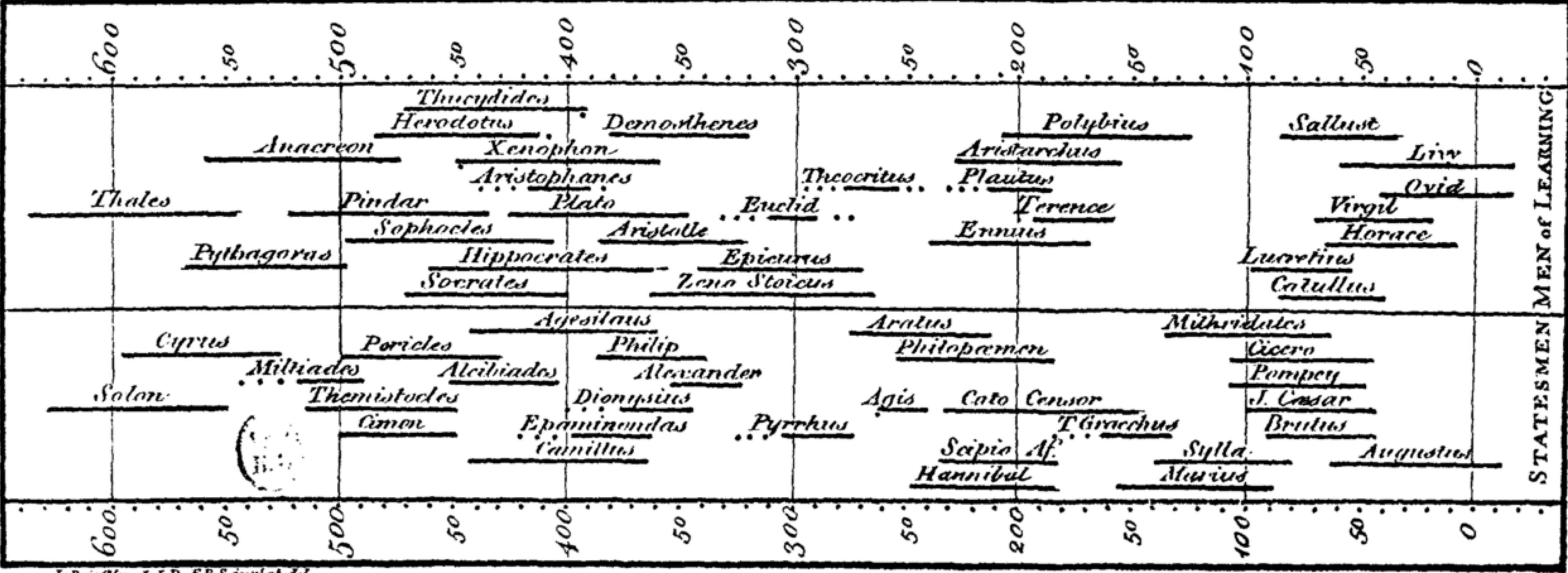
TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro. Temperature drops during the retreat

Les Cosaques passent au galop le Niemen gelé.



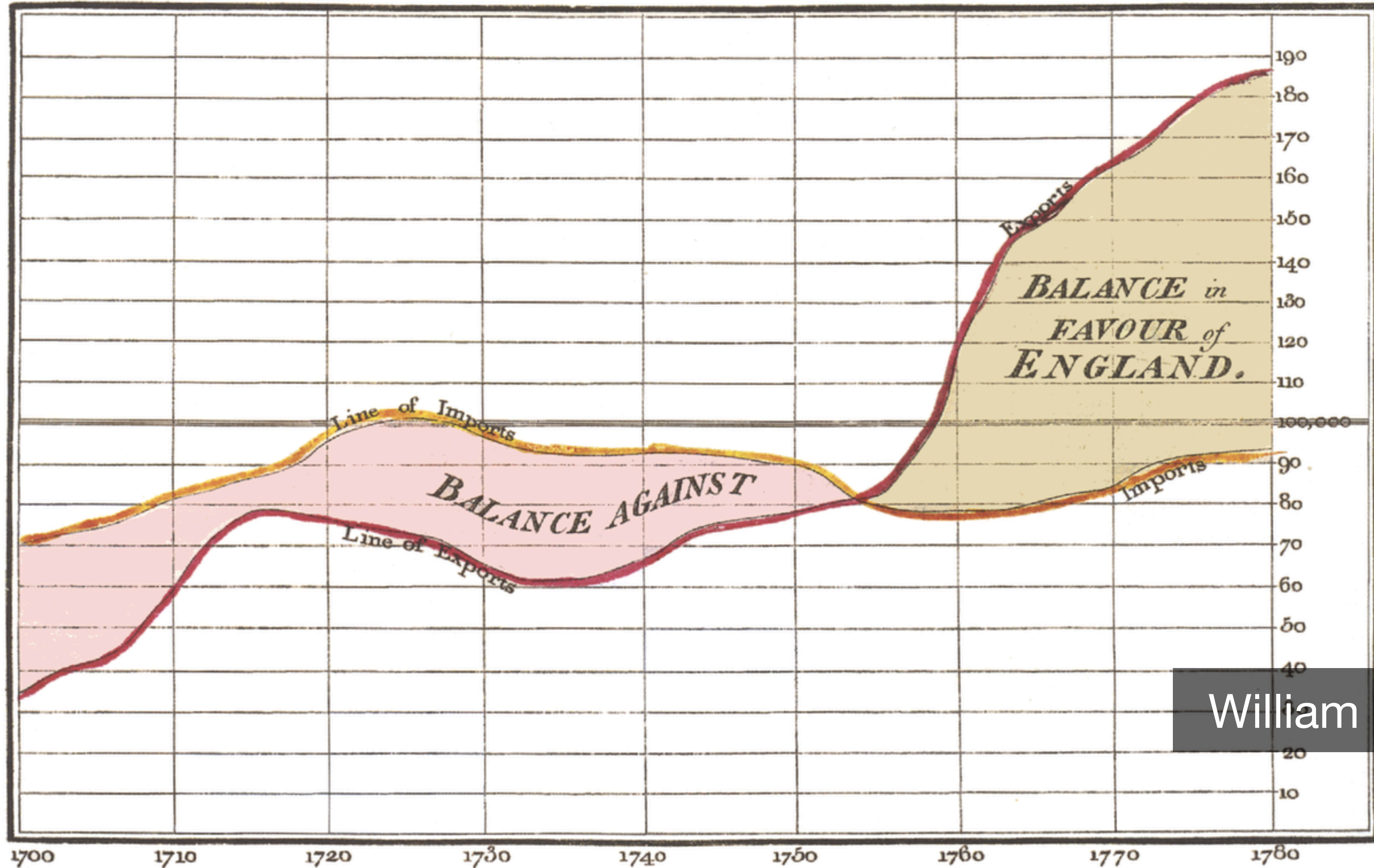
[Joseph Priestley 1765]

A Specimen of a Chart of Biography.



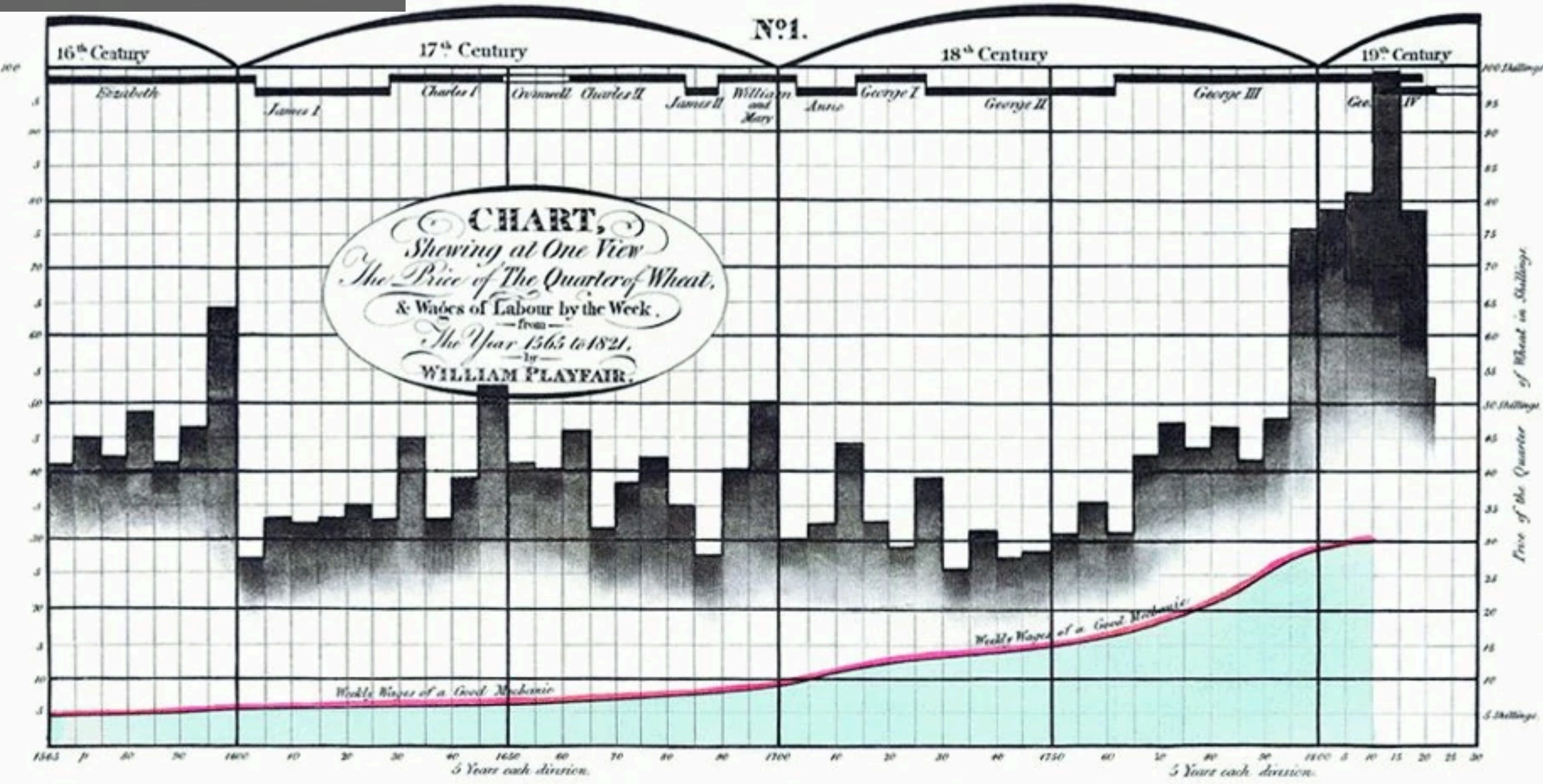
J. Priestley L.L.D. F.R.S. inv. et del.

Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.



William Playfair 1786

The Bottom line is divided into Years, the Right hand line into £10,000 each.
 Published as the Act directs, 1st May 1786, by W^m Playfair
 Neale sculpt 352, Strand, London.



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 visualizations convey information

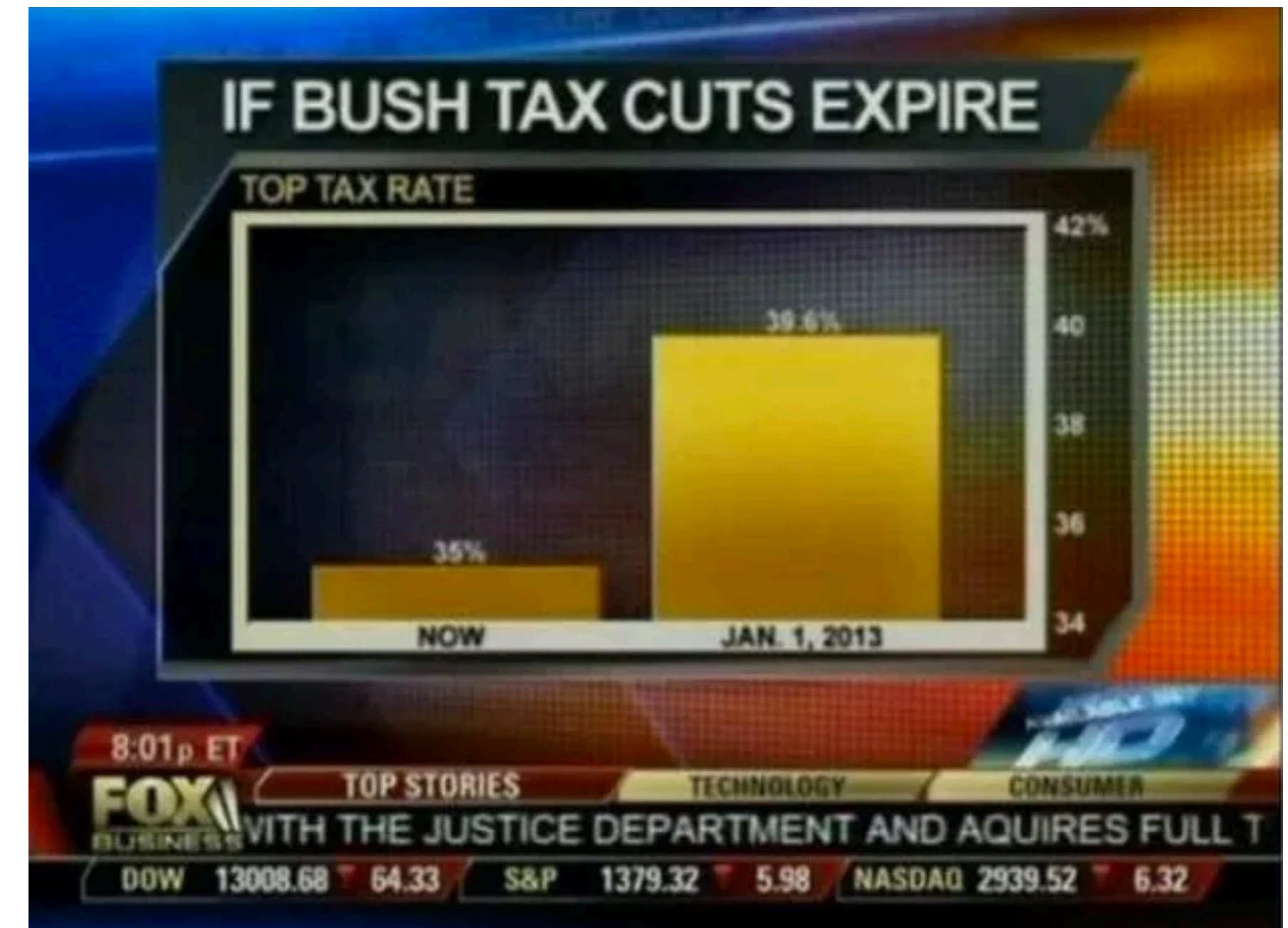
What do people perceive/comprehend?

Develop principles and techniques for effective visualizations

Leverage perception & augment cognition

Create interactive visualization tools & systems

Next



Is this good, bad or weird?

Data Visualization:
The Good, the Bad, the Weird

5 min break