

Paul Rosen

paul.rosen@utah.edu
@paulrosenphd
<https://cspaul.com>



Visualization for Data Science

DS-4630 / CS-5630

Introduction to Visualization

Why Study Visualization?



INDUSTRIAL REVOLUTION OF DATA

Joe Hellerstein, UC Berkley, 2008



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. So the complimentary scarce factor is the ability to understand that data and extract value from it.

Google's Chief Economist, Hal Varian, 2009

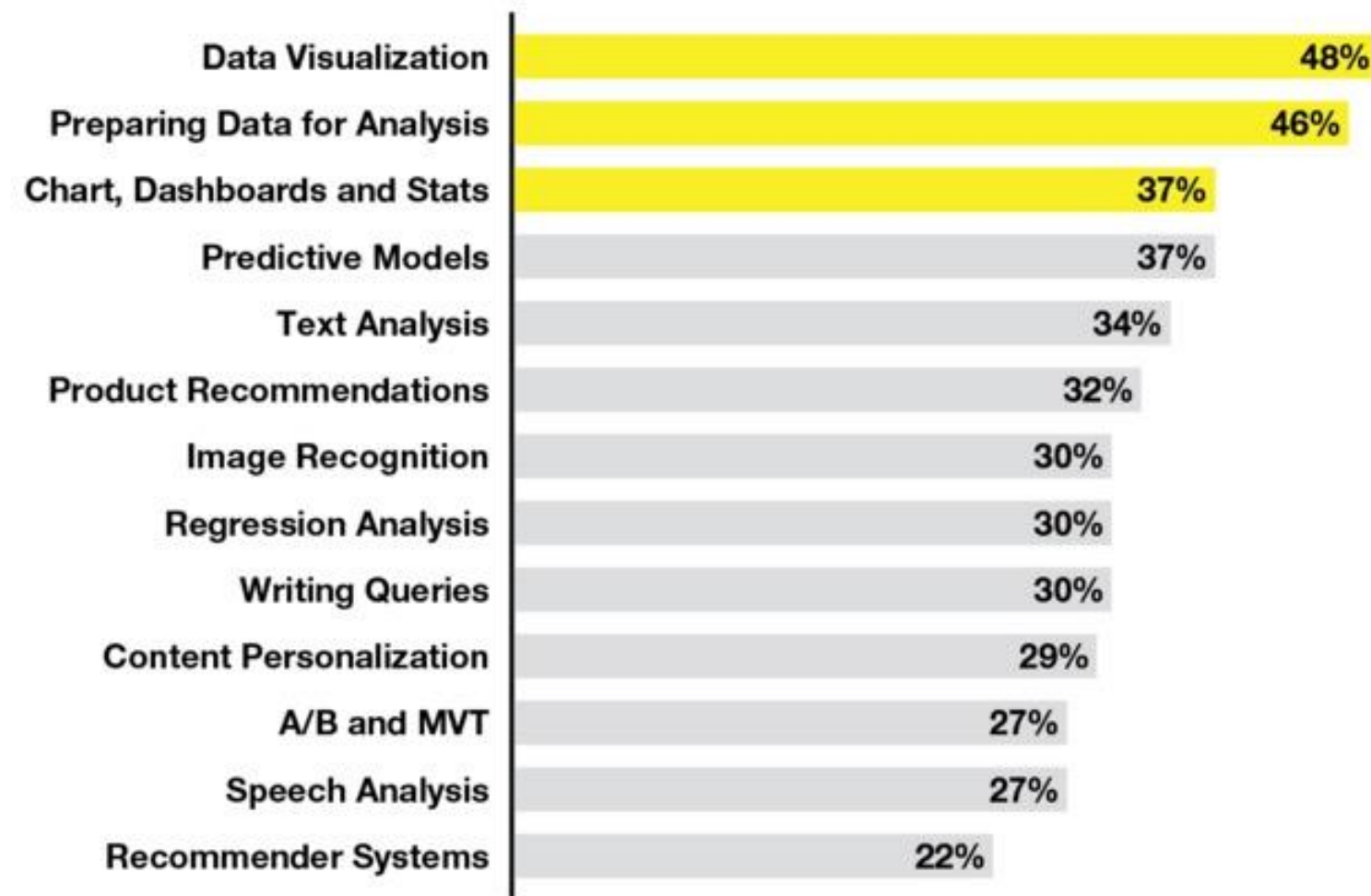
“The purpose of visualization is insight, not pictures.”
— Ben Shneiderman

“The greatest value of a picture is when it forces us to notice what
we never expected to see.”
— John Tukey

“A computer should make both calculations and graphs. Both sorts
of output should be studied; each will contribute to understanding.”
— Francis Anscombe

Expensive, talented resources are **misaligned.**

Which of the following activities does your company's data scientist
(or advanced analytics resource) perform for your marketing analytics team?



Percentage of respondents. Multiple responses allowed.
Base: All respondents, n=503

gartner.com/SmarterWithGartner

© 2018 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. or its affiliates. PR_051_393763

Gartner



VISUALIZATION GOALS

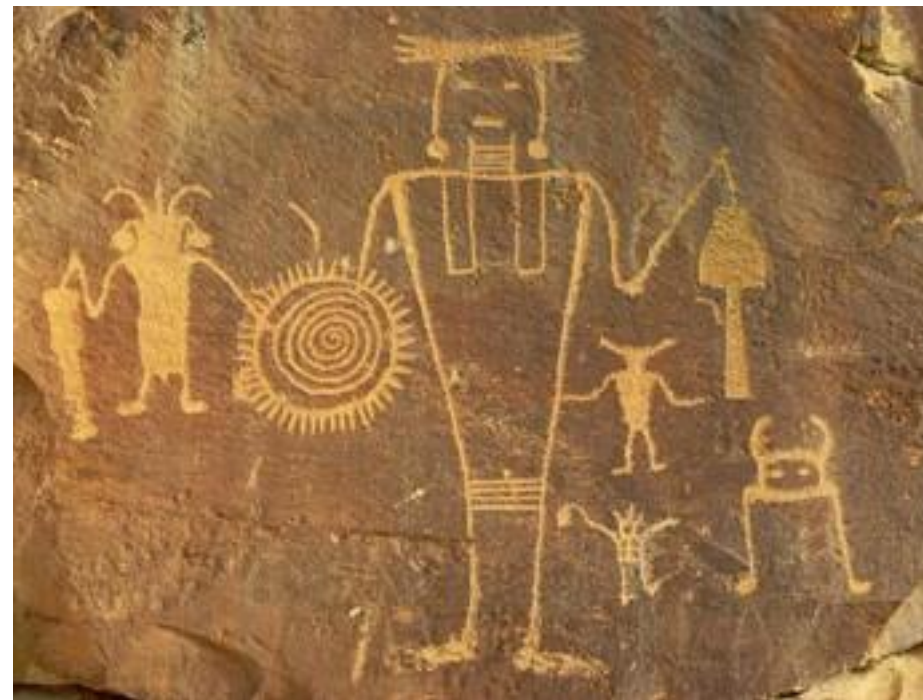
record information

analyze data to support reasoning

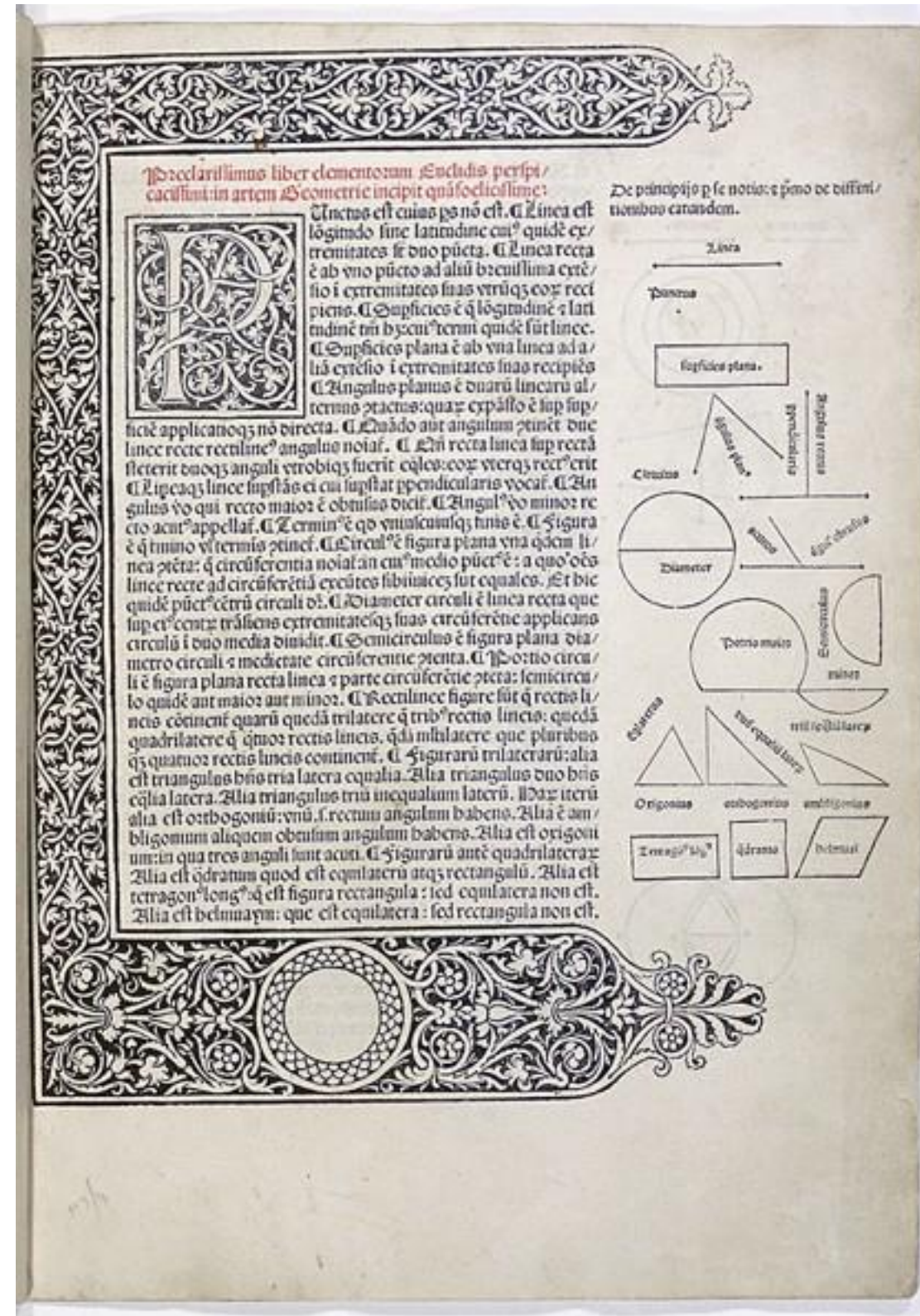
confirm hypotheses

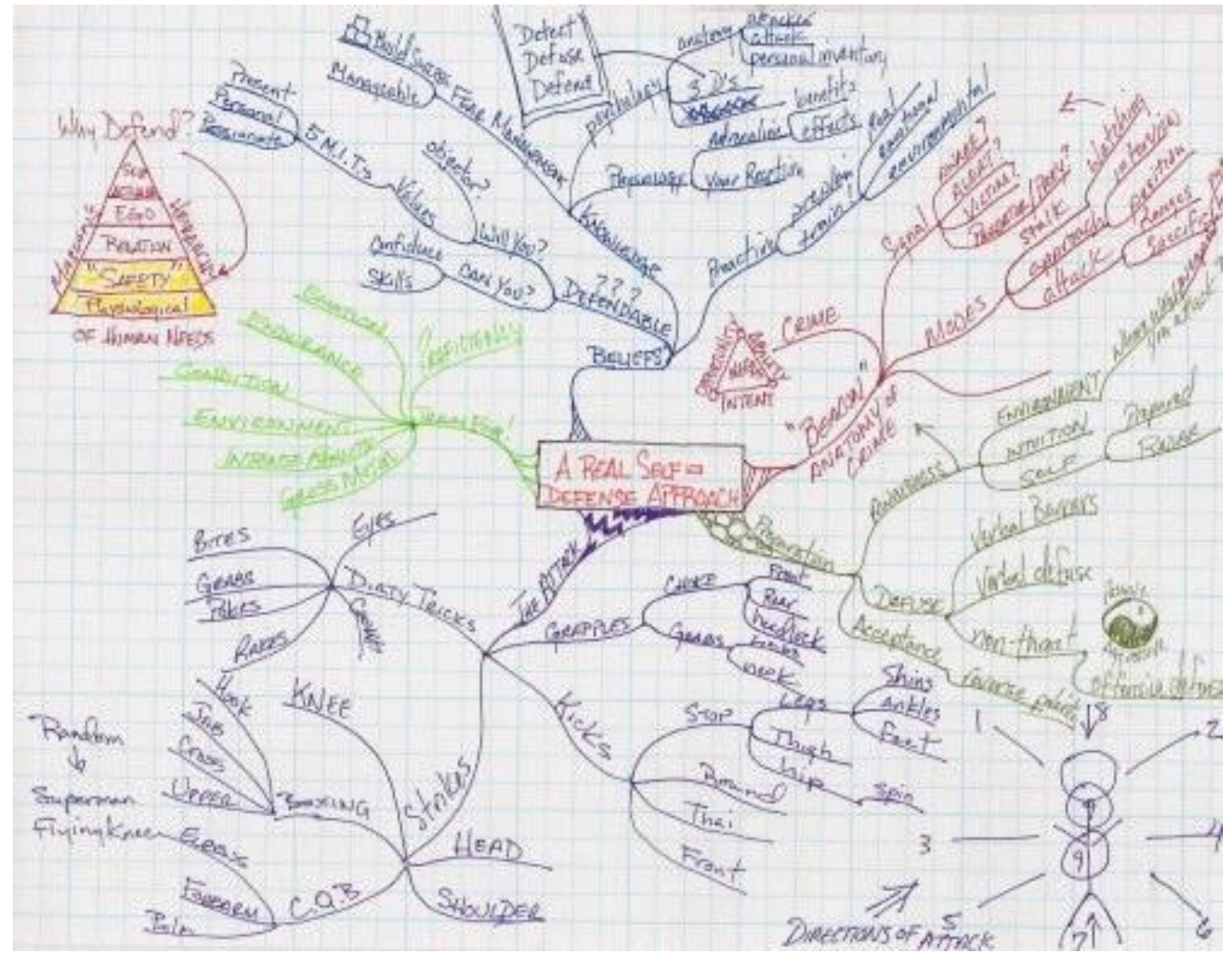
communicate ideas to others

Record Information

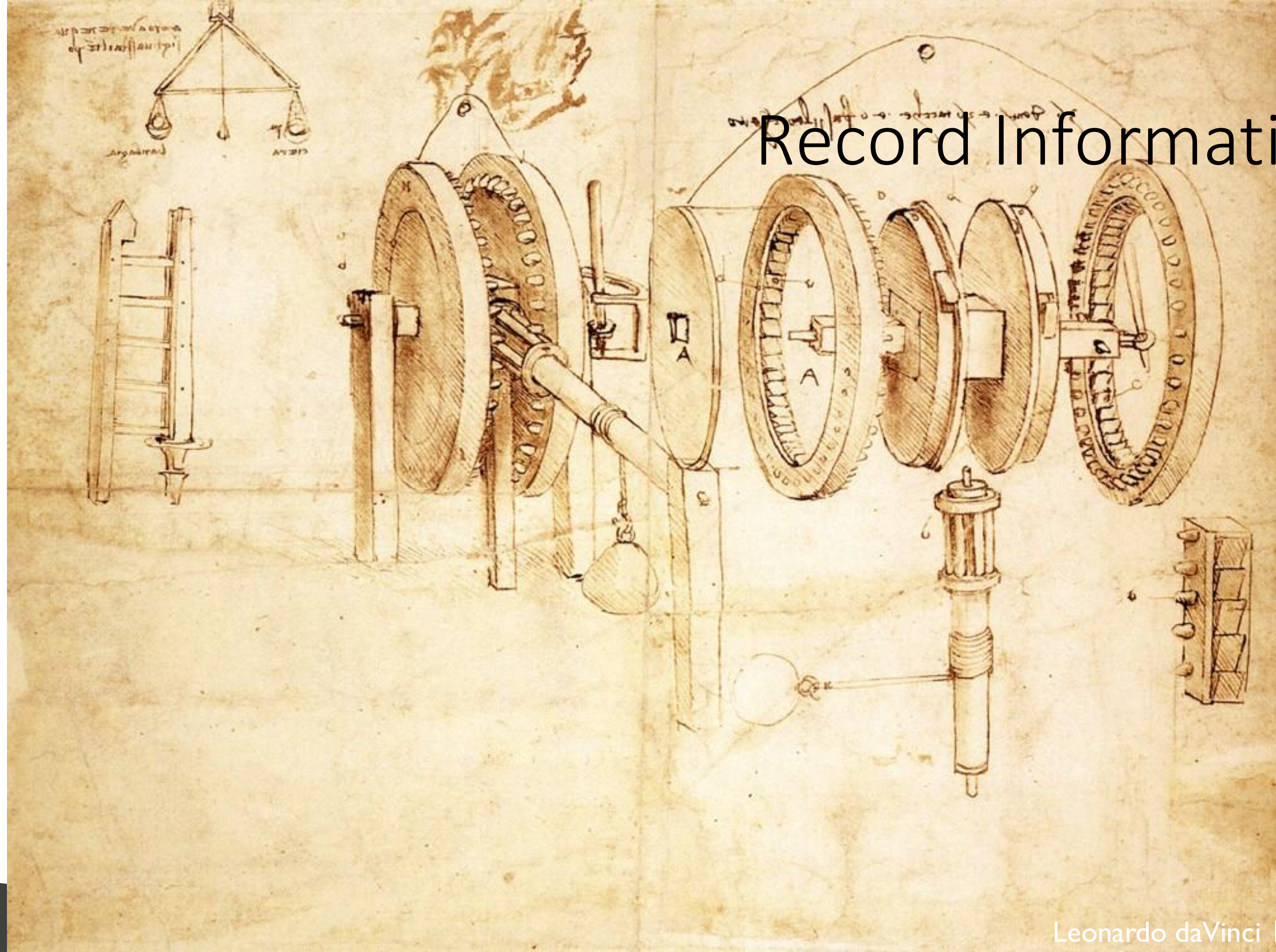


Record Information

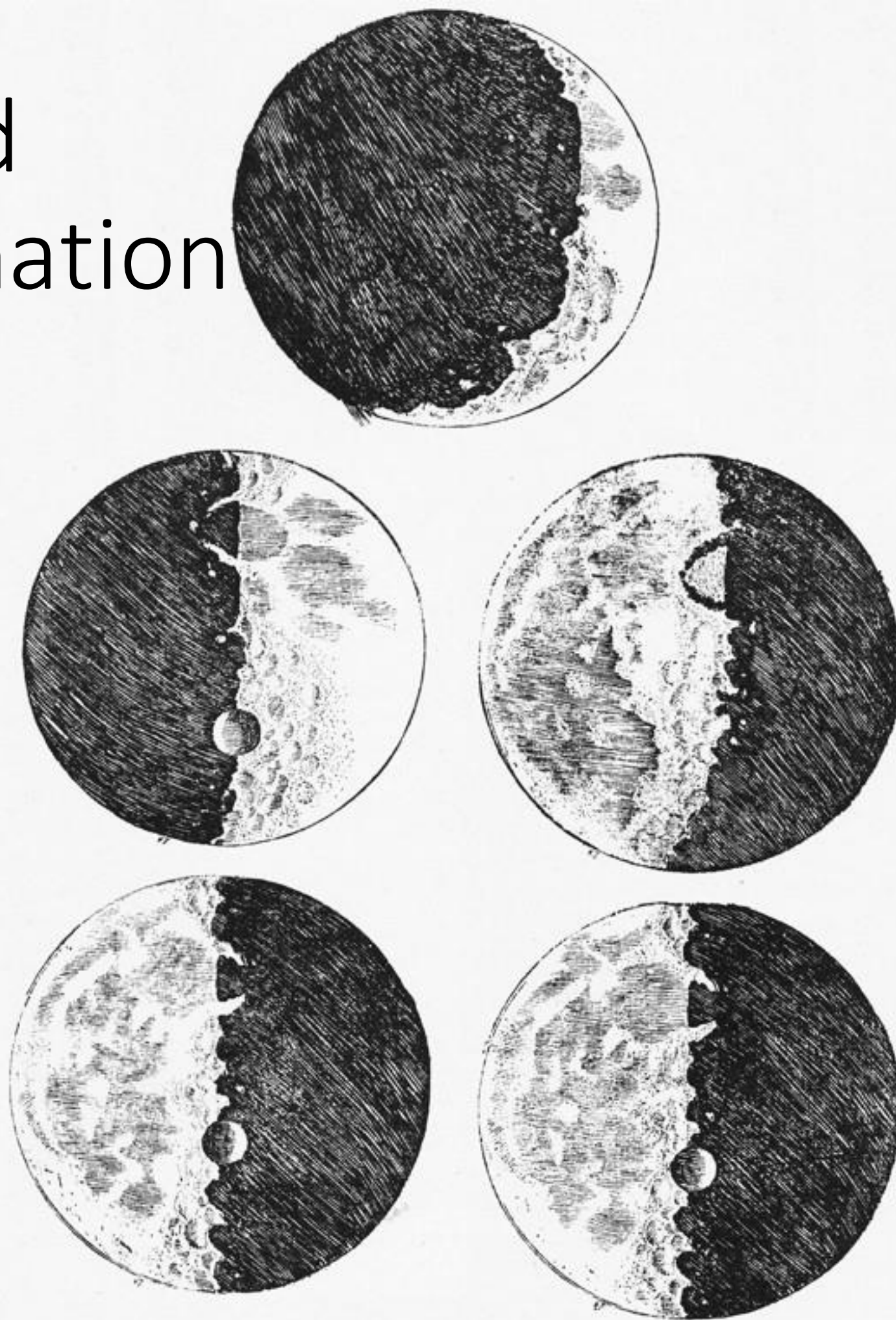


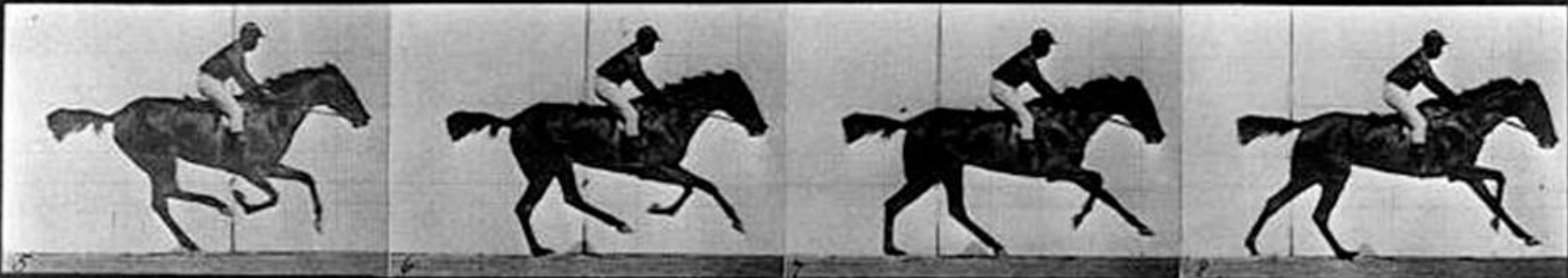


Record Information

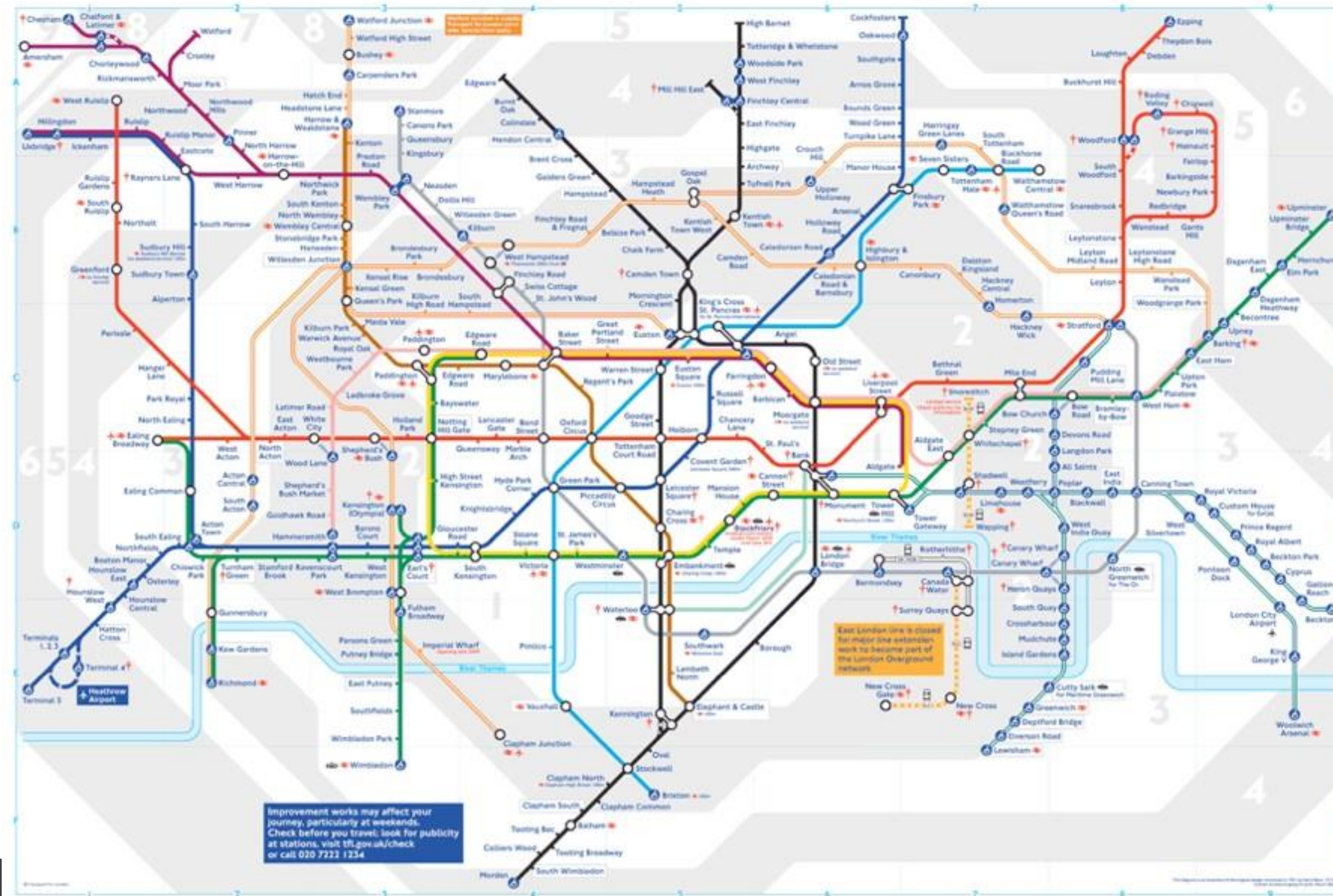


Record Information





ANALYSIS TO SUPPORT REASONING



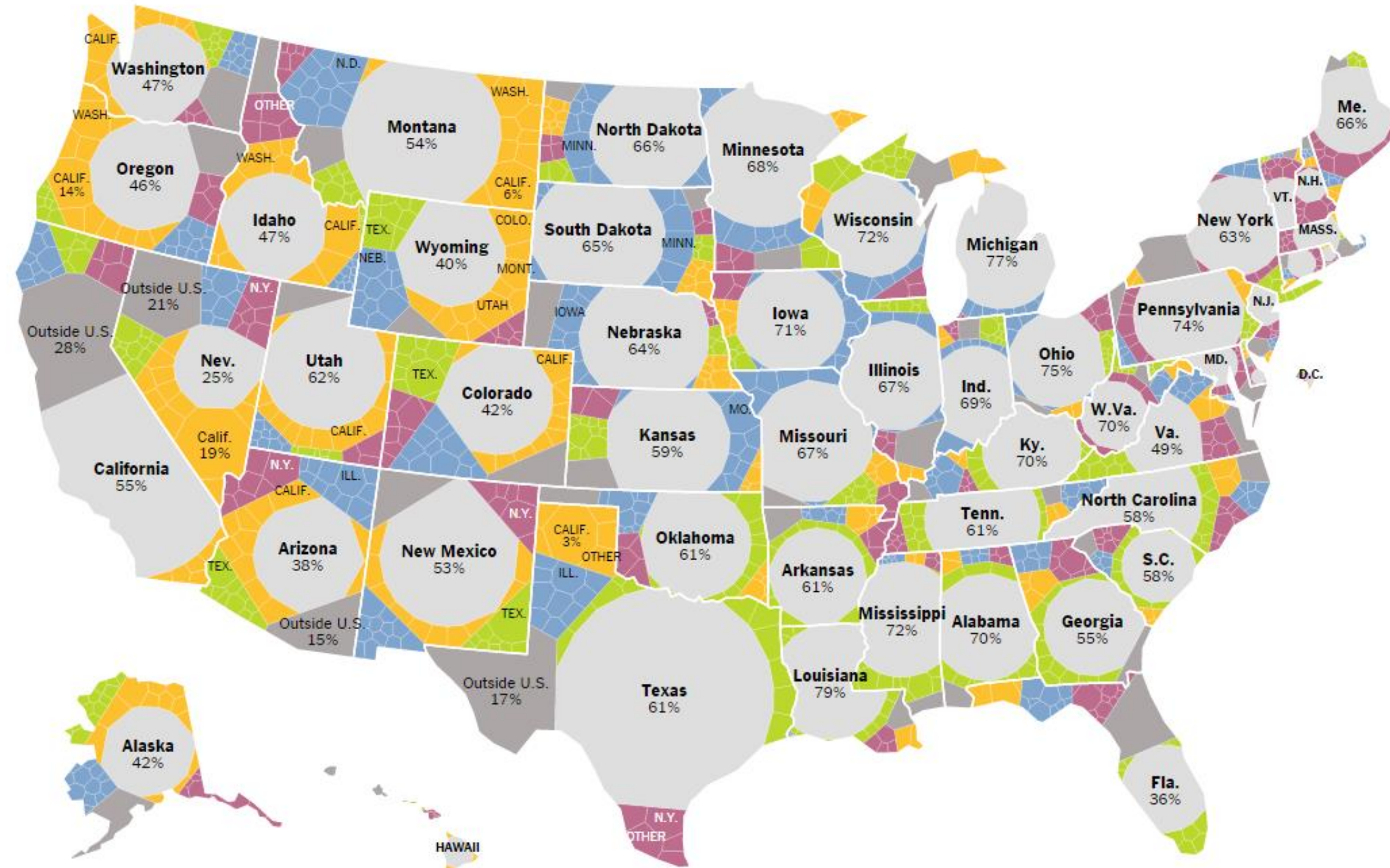
Mapping Migration in the United States

AUG. 15, 2014

Where people who lived in each state in 2012 were born

Each shape represents where the people living in a state were born. Within a state, larger shapes mean a group makes up a larger share of the population.

■ Northeast ■ South ■ Midwest ■ West ■ Outside the U.S.*



THE CHALLENGER DISASTER, 1986

Confirm Hypotheses





HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

112
Oct 30, 1985
y
y

AFT

		Cross Sectional View			Top View			<u>MOTOR</u>	<u>O-RING</u>
	SRM No.	Erosion Depth (in.)	Perimeter Affected (deg)	Nominal Dia. (in.)	Length Of Max Erosion (in.)	Total Heat Affected Length (in.)	Clocking Location (deg)		
61A LH Center Field**	22A	None	None	0.280	None	None	36° - 56°	DM-1	47
61A LH CENTER FIELD**	22A	NONE	NONE	0.280	NONE	NONE	338° - 18°	DM-2	52
51C LH Forward Field**	15A	0.010	154.0	0.280	4.25	5.25	163	QM-3	48
51C RH Center Field (prim)***	15B	0.038	130.0	0.280	12.50	58.75	354	QM-4	51
51C RH Center Field (sec)***	15B	None	45.0	0.280	None	29.50	354	SRM-15	53
41D RH Forward Field	13B	0.028	110.0	0.280	3.00	None	275	SRM-22	75
41C LH Aft Field*	11A	None	None	0.280	None	None	--	SRM-25	29
41B LH Forward Field	10A	0.040	217.0	0.280	3.00	14.50	351		27
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.

BLOW BY HISTORY

- SRM-15 WORST BLOW-BY
 - o 2 CASE JOINTS (80°), (110°) ARC
 - o MUCH WORSE VISUALLY THAN SRM-22

SRM 22 BLOW-BY

- o 2 CASE JOINTS (30-40°)

SRM-13A, 15, 16A, 18, 23A 24A

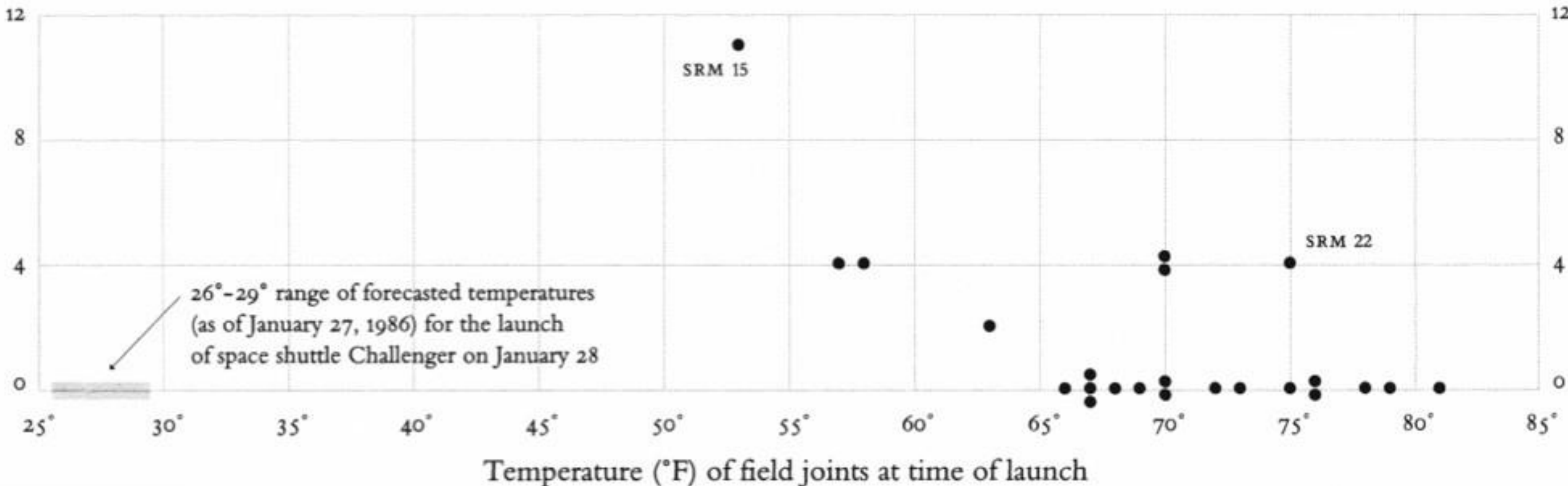
- o NOZZLE BLOW-BY

HISTORY OF O-RING TEMPERATURES (DEGREES - F)

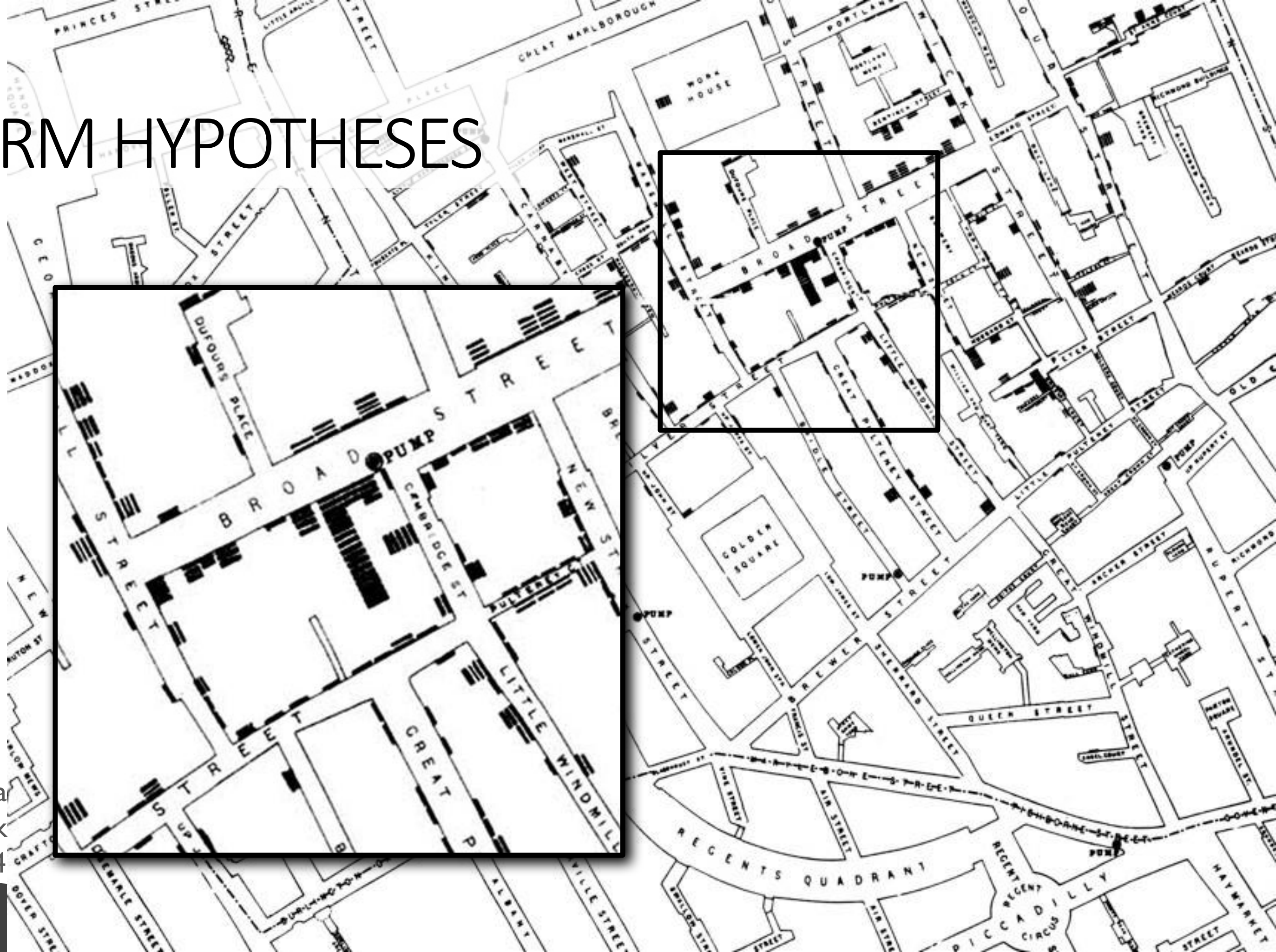
<u>MOTOR</u>	<u>MBT</u>	<u>AMB</u>	<u>O-RING</u>	<u>WIND</u>
DM-1	68	36	47	10 MPH
DM-2	76	45	52	10 MPH
QM-3	72.5	40	48	10 MPH
QM-4	76	48	51	10 MPH
SRM-15	52	64	53	10 MPH
SRM-22	77	78	75	10 MPH
SRM-25	55	26	29	10 MPH
			27	25 MPH



O-ring damage
index, each launch

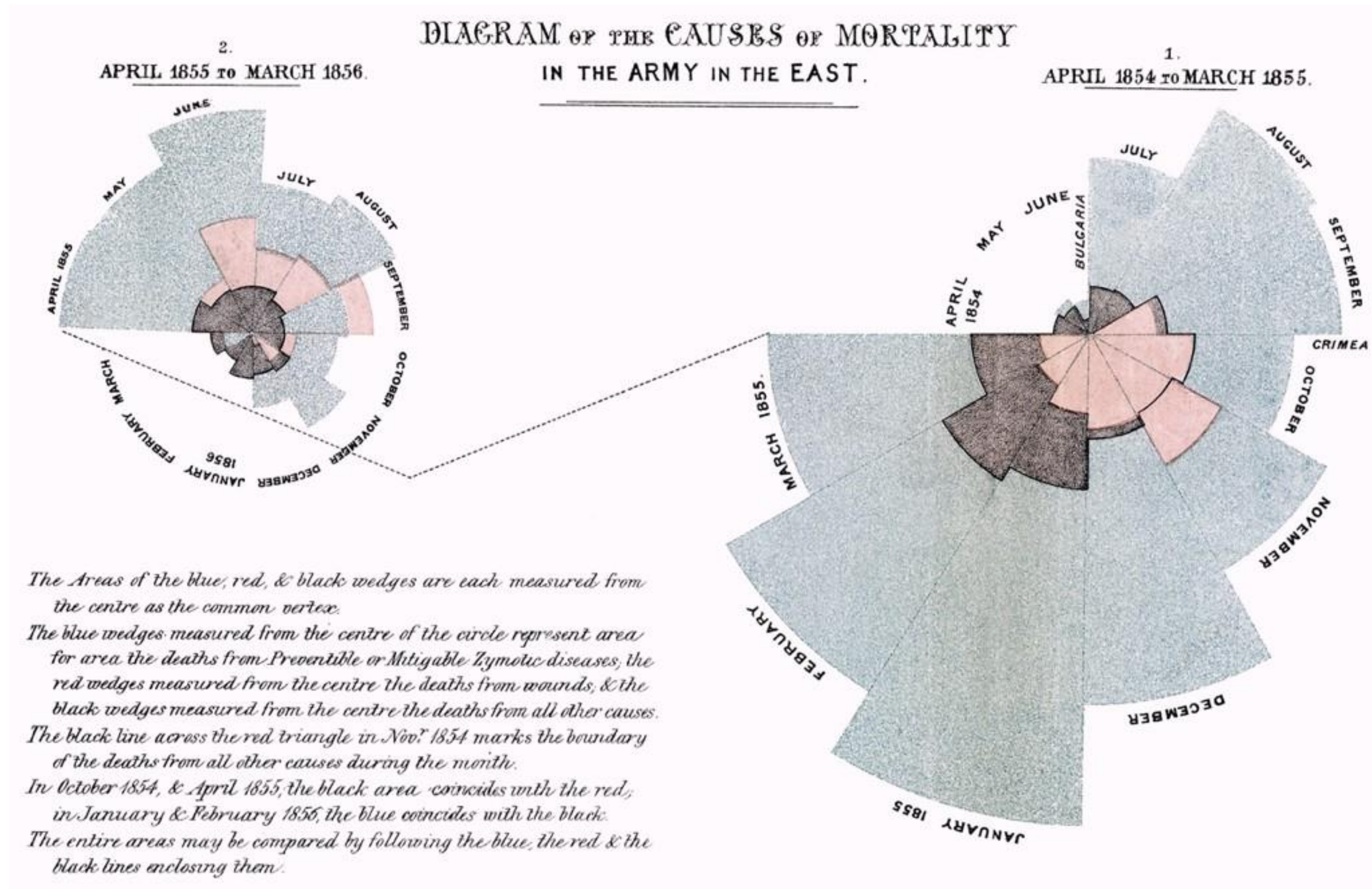


CONFIRM HYPOTHESES



London Cholera
Outbreak
John Snow 1854

COMMUNICATE IDEAS

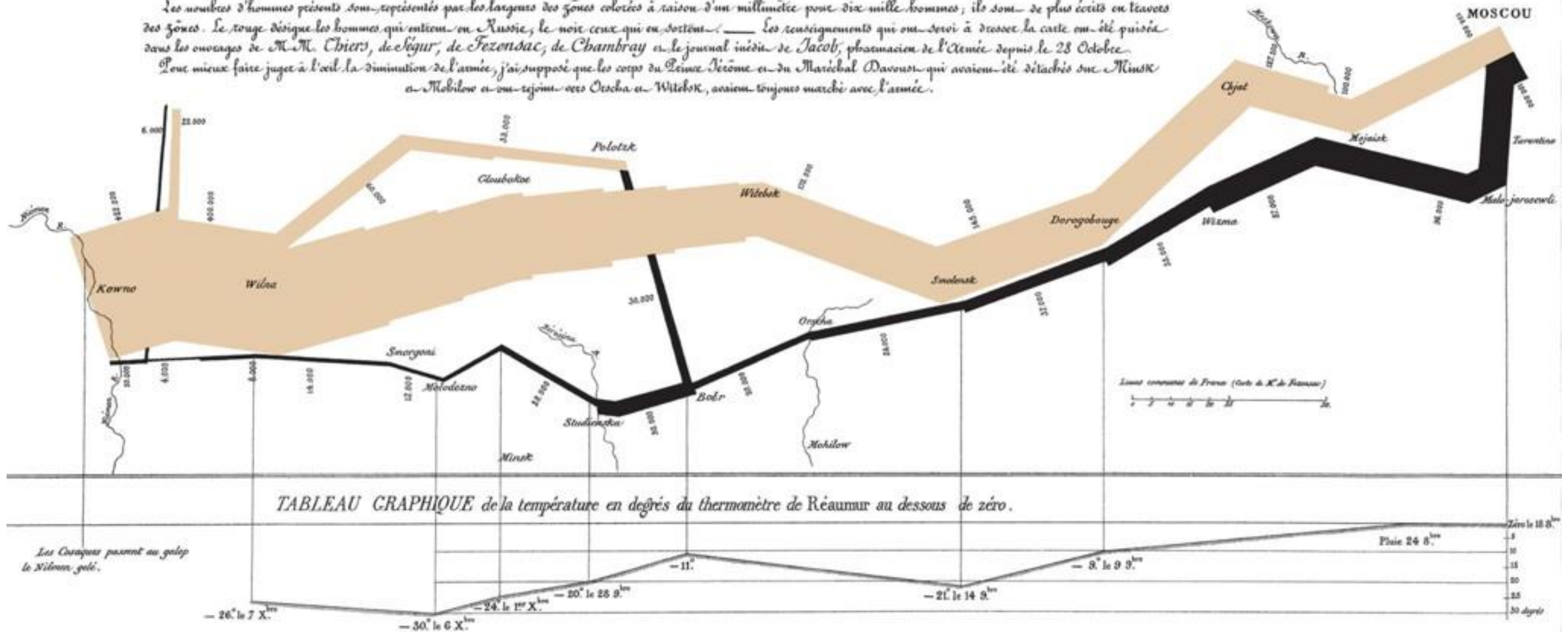


Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dessiné 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 lettres 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 Léger, 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 qui rejoindront Orescha et Witebsk, avaient toujours marché avec l'armée.



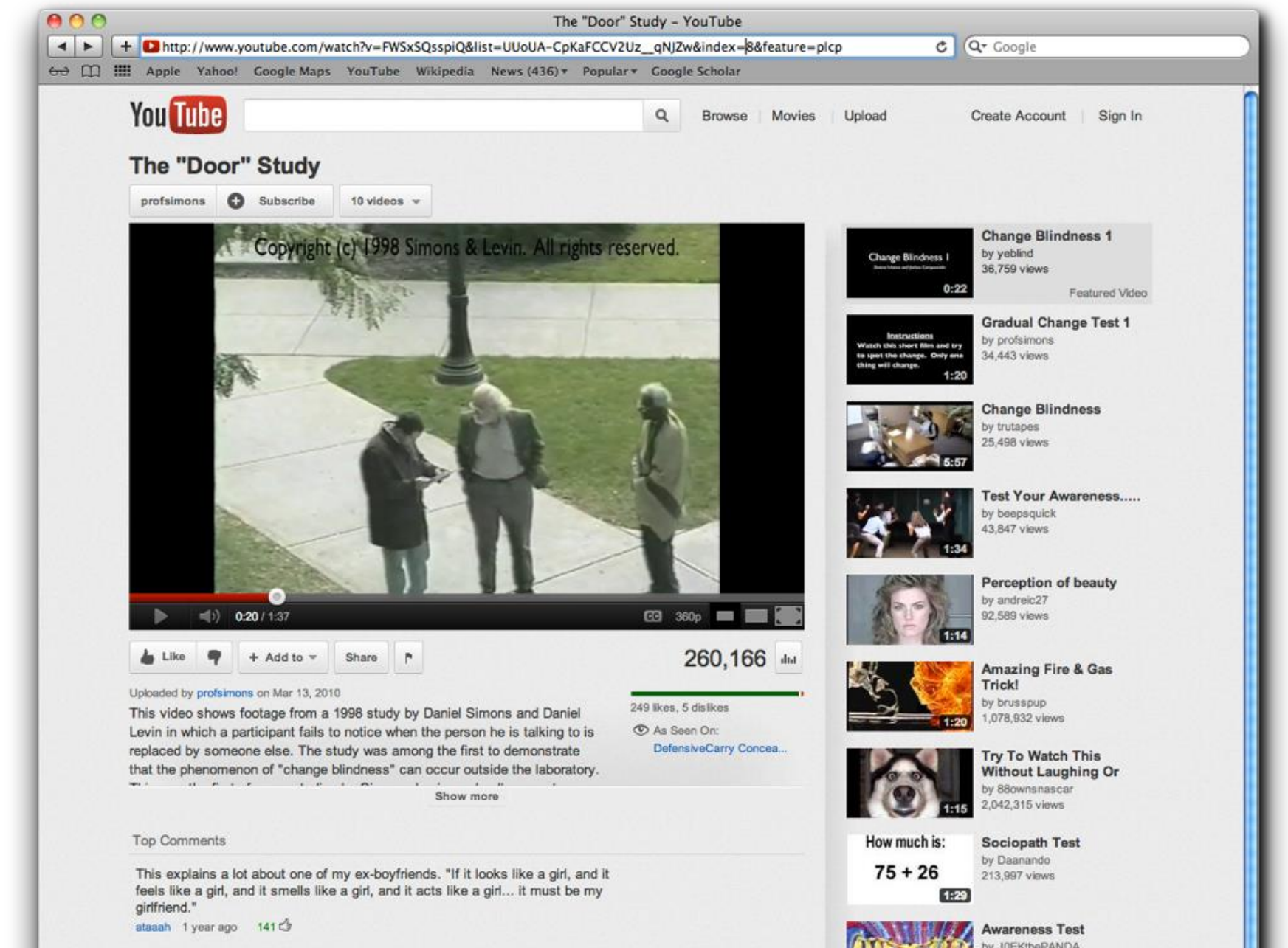
Autog. par Reynier, à Paris, 3^e Mars 1869.

Imp. Lit. Reynier et Desclaux.

why does visualization work?

why does visualization
work?

cognition is limited



Visualization uses perception to point
out interesting things

MTHIVLWYADCEQGHKILKMTWYN
ARDCAIREQGHLVKMFPSTWYARN
GFPSVCEILQGKMFP SNVRCEQDI
PSGHLMFHKMVPSTWYACEQTWRN

MTHI**V**LWYADCEQGHKILKMTWYN
ARDCAIREQGH**L**VKMFPSTWYARN
GFPS**V**CEILQGKMFP SNVRCEQDI
PSGHLMFHKM**V**PSTWYACEQTWRN

MTHI**V**LWYADCEQGHKILKMTWYN
ARDCAIREQGH**L**KMFPSSTWYARN
GFPS**V**CEILQGKMFPSN**V**RCEQDI
PSGHLMFHKM**V**PSTWYACEQTWRN

(but can also deceive you)

why does visualization work?

memory is limited

calculation exercise . . . $\begin{array}{r} 34 \\ \times 28 \\ \hline \end{array}$

79
calculation exercise . . . x 16

calculation exercise . . . $\begin{array}{r} 34 \\ \times 28 \\ \hline \end{array}$

Visualization
uses pictures to enhance working
memory

15 19 60

33 11 75

57 34 79

18 51 92

73 22 13

71 60 22

17 10 68

73 18 55

65 46 29

60 73 22

46 92 97

10 58 46

57 17 83

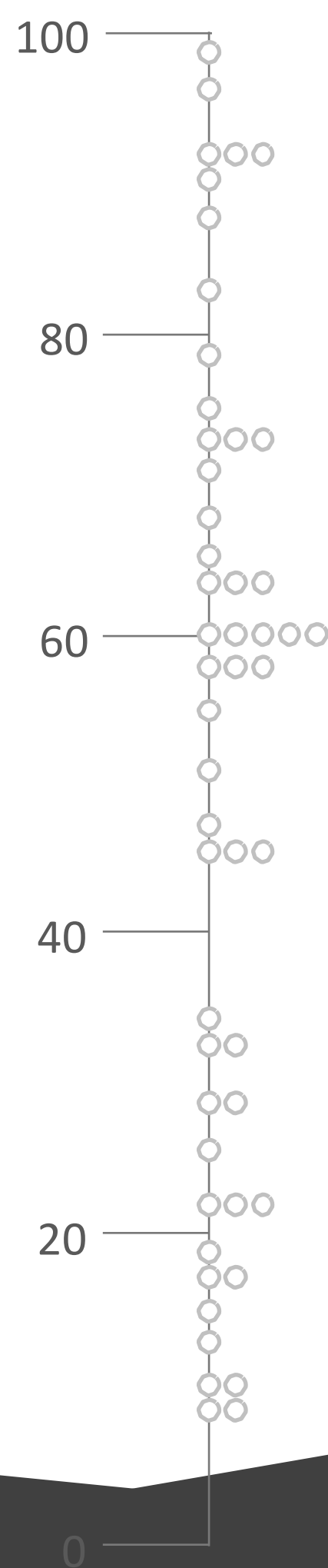
26 99 33

88 92 60

91 29 57

96 12 47

given these numbers . . .
. . . what number
appears most often?

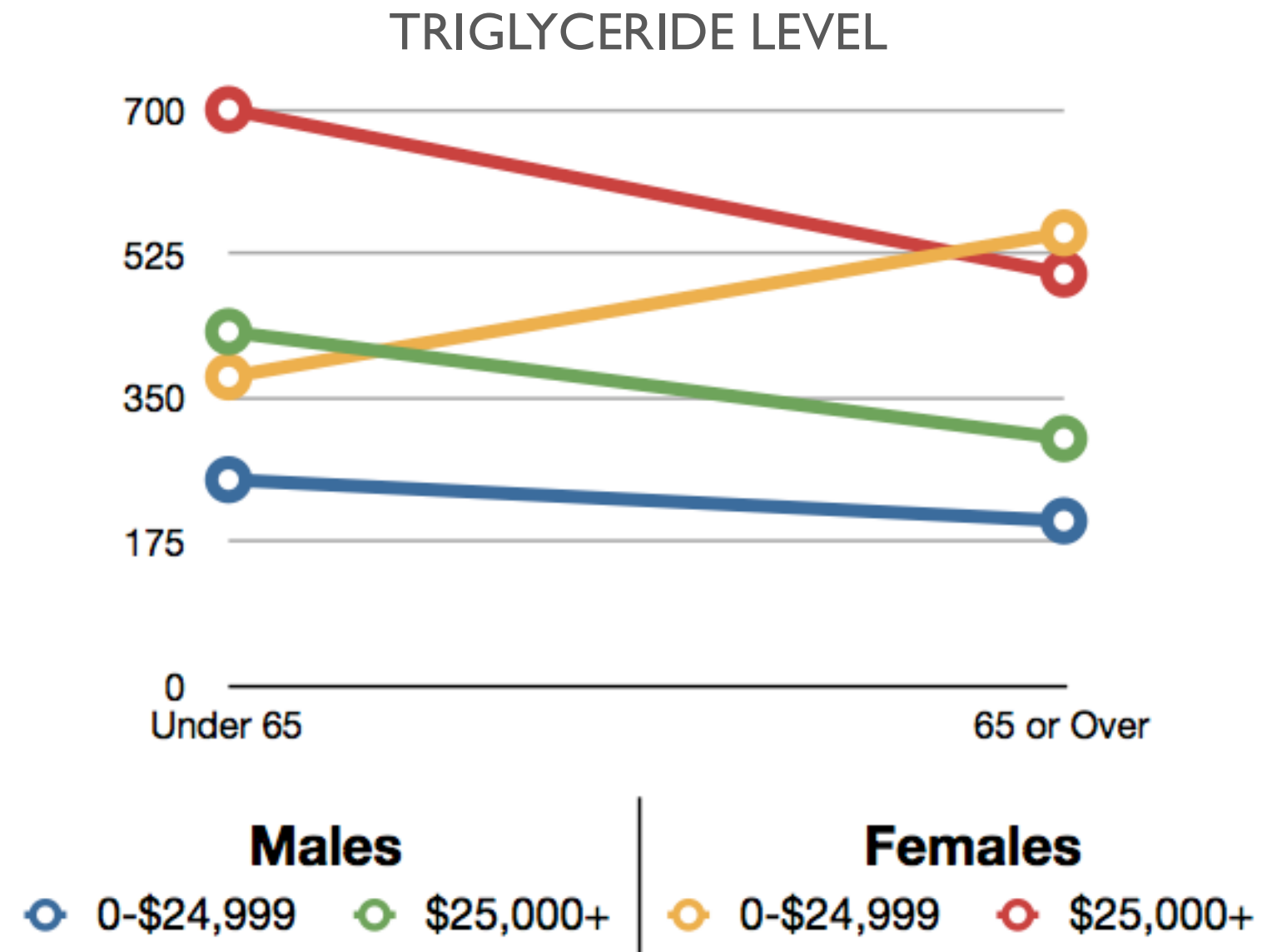


given these numbers . . .
. . . what number
appears most often?

query exercise . . .

TRIGLYCERIDE LEVEL				
	Males		Females	
Income Group	Under 65	65 or Over	Under 65	65 or Over
0-\$24,999	250	200	375	550
\$25,000+	430	300	700	500

QUESTION: Which gender and income level shows a different effect of age on triglyceride levels?



QUESTION: Which gender and income level shows a different effect of age on triglyceride levels?

The **goal of this course** is to introduce students to the principles, methods, and techniques for effective visual analysis of data

We will discuss visualization **techniques for a broad range of data types.**

You will **gain experience in developing your own** interactive visualization tools.

This course requires independent knowledge in (or learning of) HTML, CSS, and JavaScript. We will cover these topics but move quickly.

