Paul Rosen
paul.rosen@utah.edu
@paulrosenphd
https://cspaul.com

Visualization for Data Science
DS-4630 / CS-5630 / CS-6630

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

Data Science Top 3 Job since 2017 according to Glassdoor.com

50 Best Jobs in America

Awards

Best Places to Work

Highest Rated CEOs

Best Places to Interview

Lists

Best Jobs

Best Cities for Jobs

Highest Paying Jobs

Oddball Interview Questions

Trends

This report ranks jobs according to each job's Glassdoor Job Score, determined by combining three factors: number of job openings, salary, and overall job satisfaction rating.

Employers: Want to recruit better in 2017? Find out how.

United States | 2017

11k Shares

1 Data Scientist

4.8 / 5
Job Score

4.4 / 5
Job Satisfaction

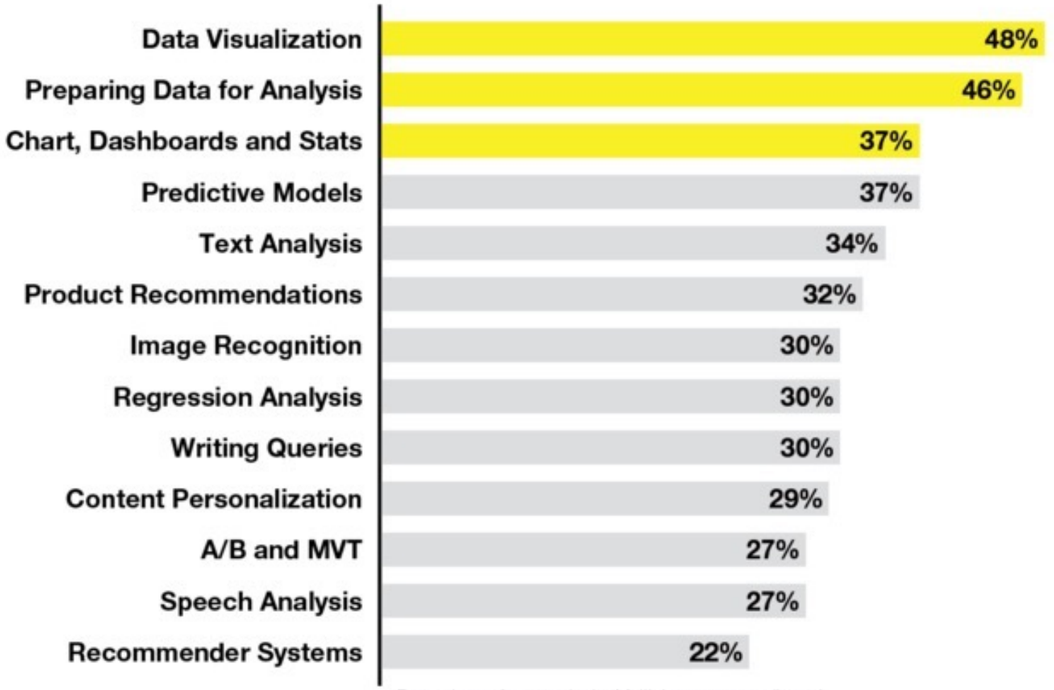
$110,000
Median Base Salary

4,184
Job Openings

View Jobs

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

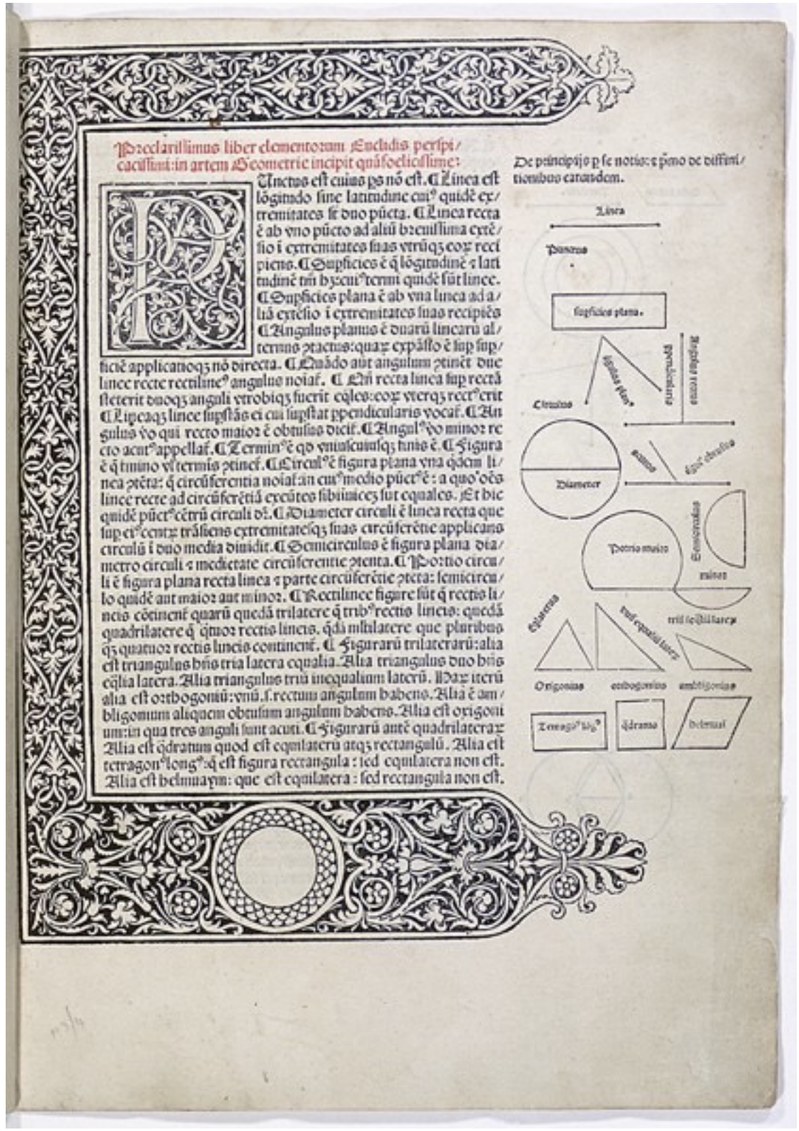
analyze data to support reasoning

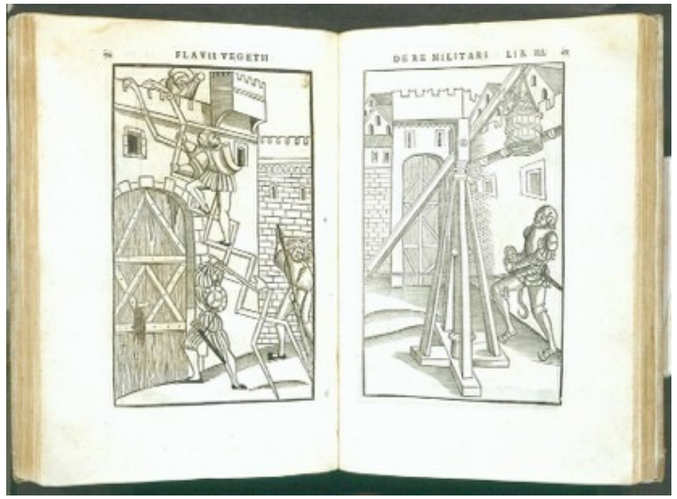
confirm hypotheses

communicate ideas to others

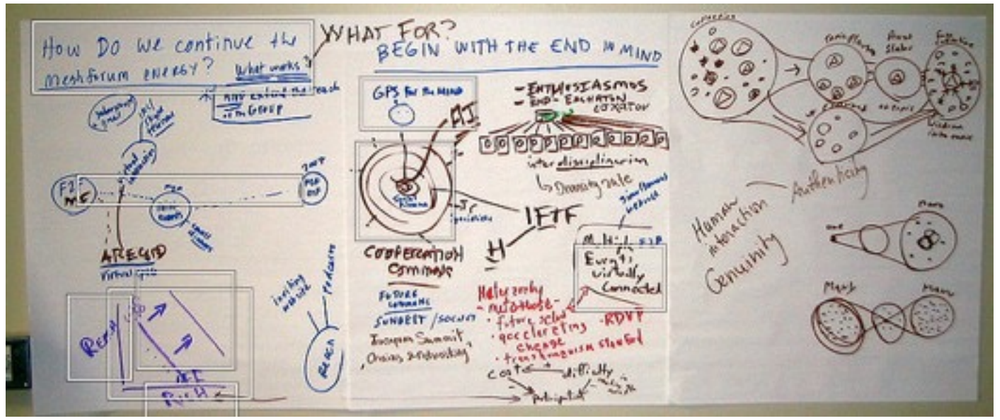
Record Information

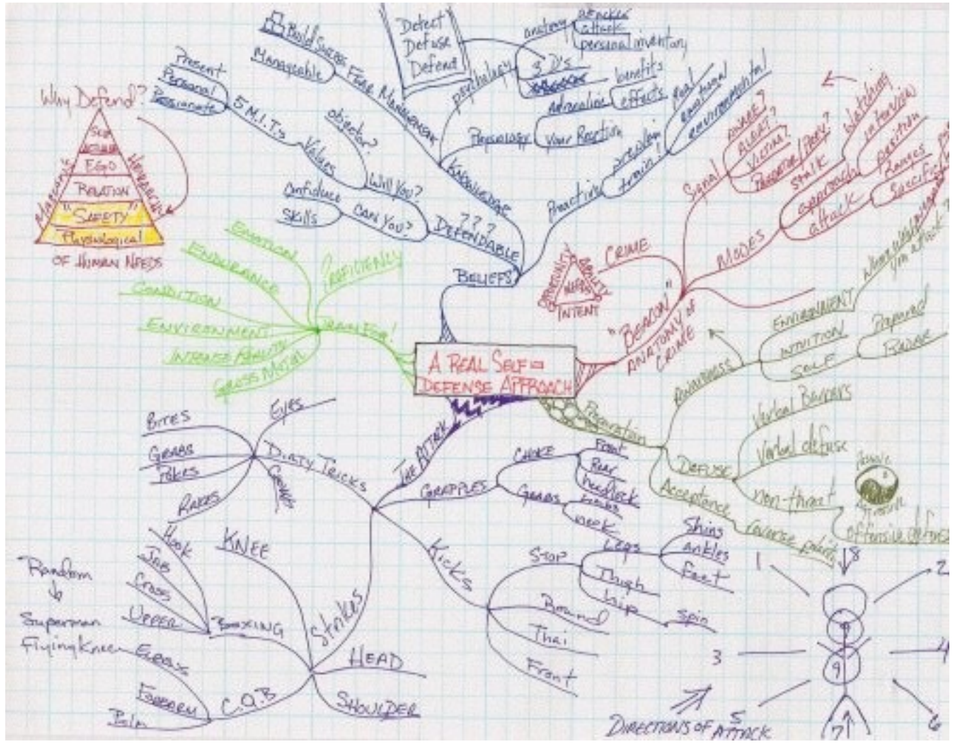
Record Information



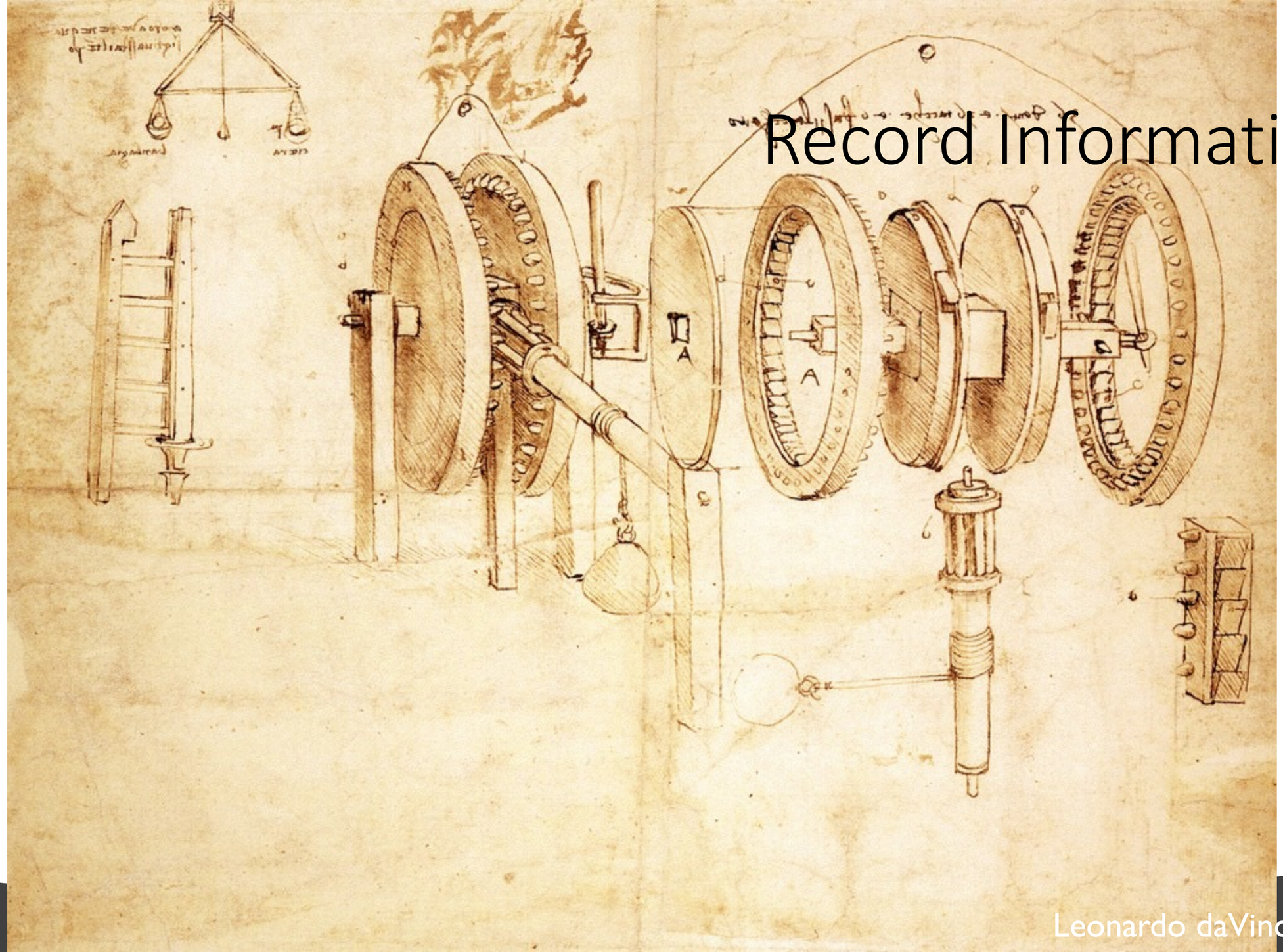


Record Information



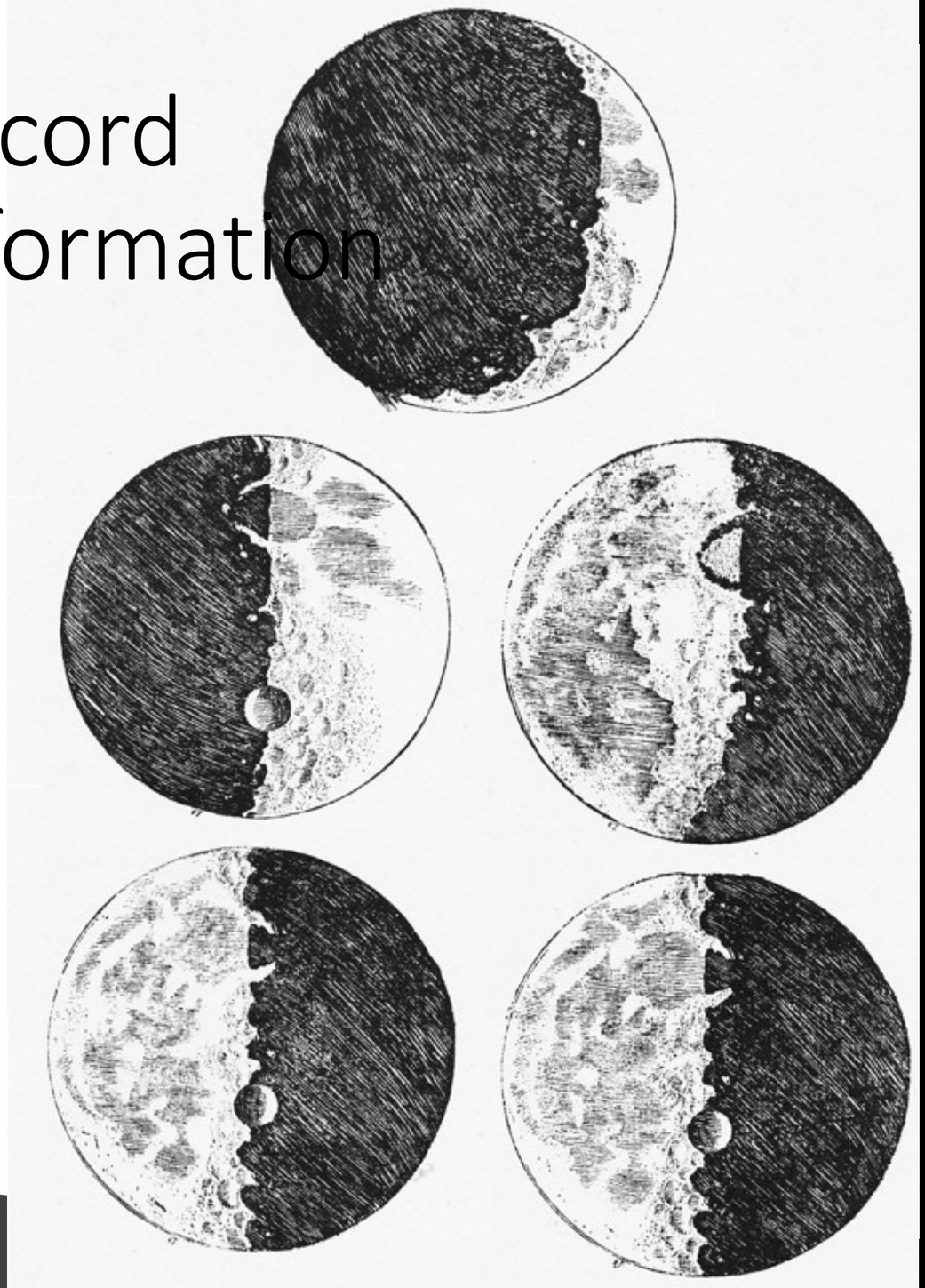


Record Information

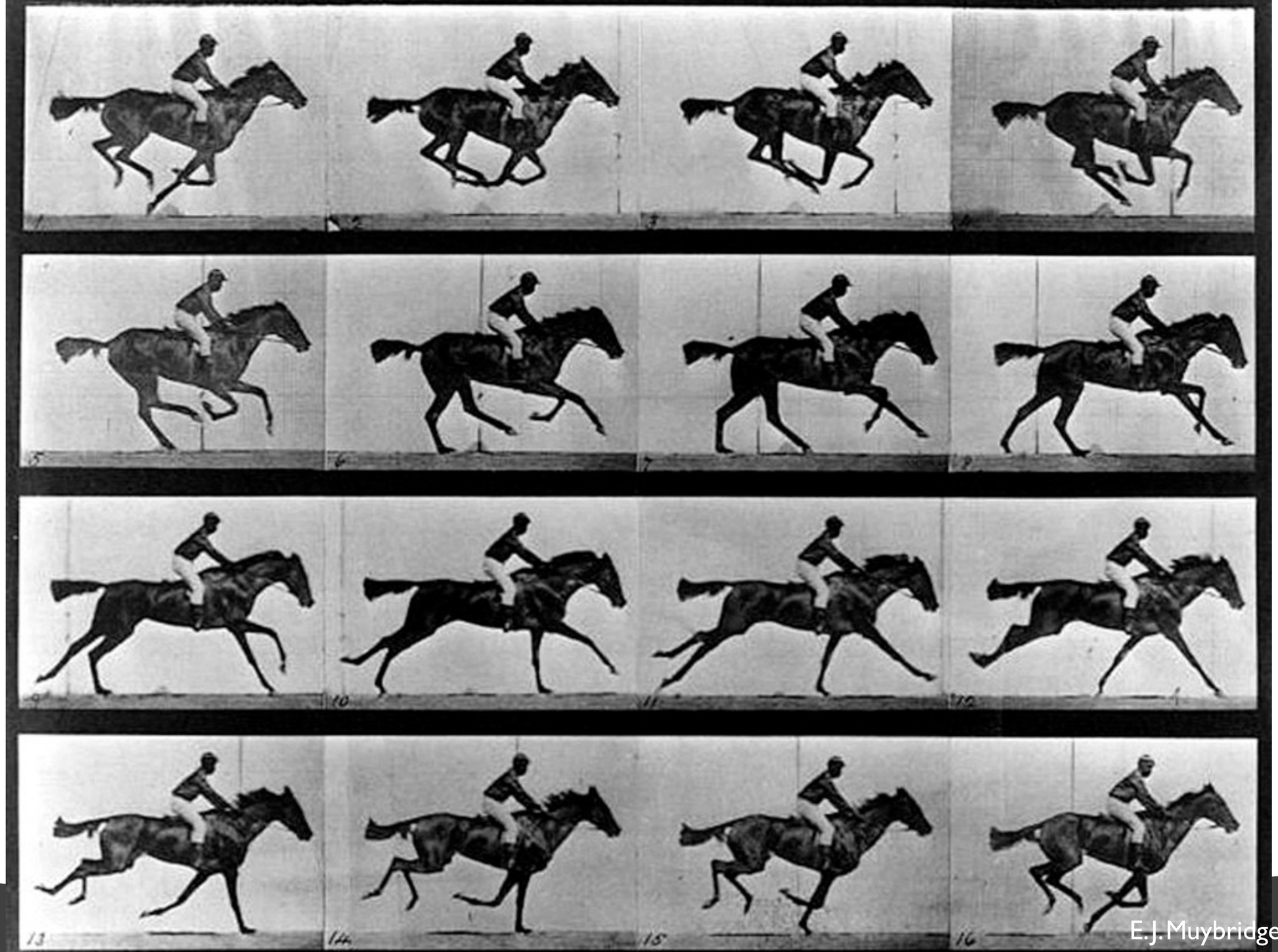


Leonardo daVinci 1485

Record Information



Galileo 1610



E.J. Muybridge 1878

ANALYSIS TO SUPPORT REASONING

ANALYSIS TO SUPPORT REASONING

Mapping Migration in the United States

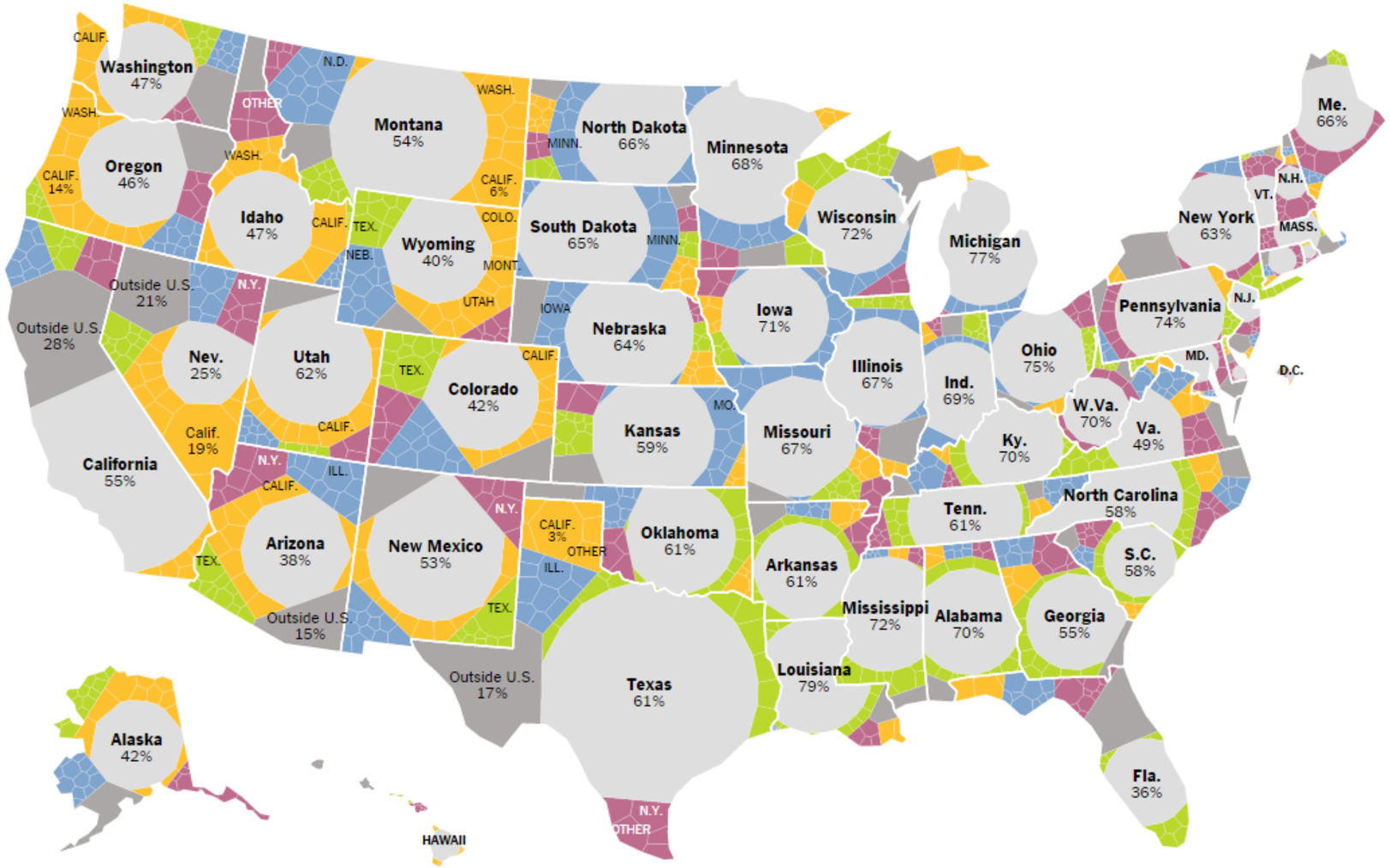
AUG. 15, 2014

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.

SELECT A YEAR 1900 1950 2012

Northeast South Midwest West Outside the U.S.*



Source: New York Times http://www.nytimes.com/2014/08/16/upshot/mapping-migration-in-the-united-states-since-1900.html?_r=0

Confirm Hypotheses

THE CHALLENGER DISASTER, 1986

source: Wikipedia

USA
NASA
Challenger

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**	None	None	0.280	None	None	36°--66°
61A LH ~~CENTER~~ FIELD**	NONE	NONE	0.280	NONE	NONE	338°-18°
51C LH Forward Field**	0.010	154.0	0.280	4.25	5.25	163
51C RH Center Field (prim)***	0.038	130.0	0.280	12.50	58.75	354
51C RH Center Field (sec)***	None	45.0	0.280	None	29.50	354
41D RH Forward Field	0.028	110.0	0.280	3.00	None	275
41C LH Aft Field*	None	None	0.280	None	None	--
41B LH Forward Field	0.040	217.0	0.280	3.00	14.50	351
STS-2 RH Aft Field	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.

MOTOR	O-RING
DM-1	47
DM-2	52
QM-3	48
QM-4	51
SRM-15	53
SRM-22	75
SRM-25	29 27

BLOW BY HISTORY

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

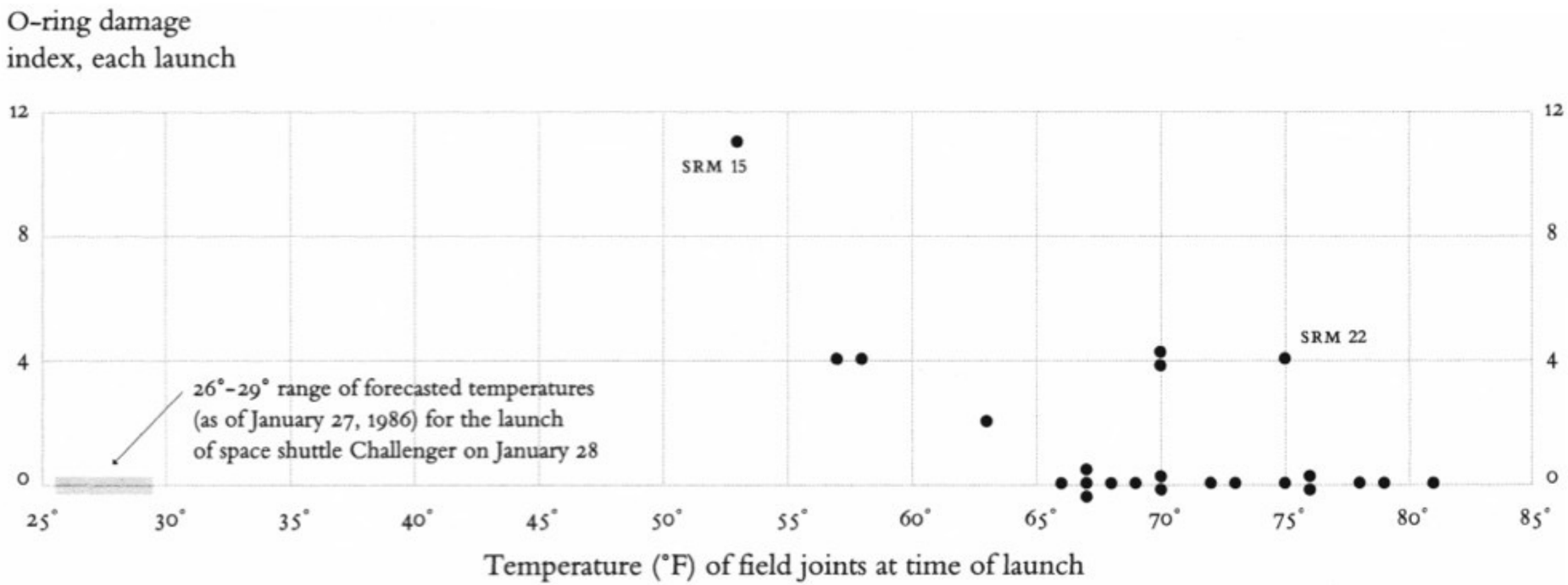
SRM 22 BLOW-BY
- 2 CASE JOINTS (30-40°)

SRM-13A, 15, 16A, 18, 23A 24A
- NOZZLE BLOW-BY

HISTORY OF O-RING TEMPERATURES (DEGREES-F)

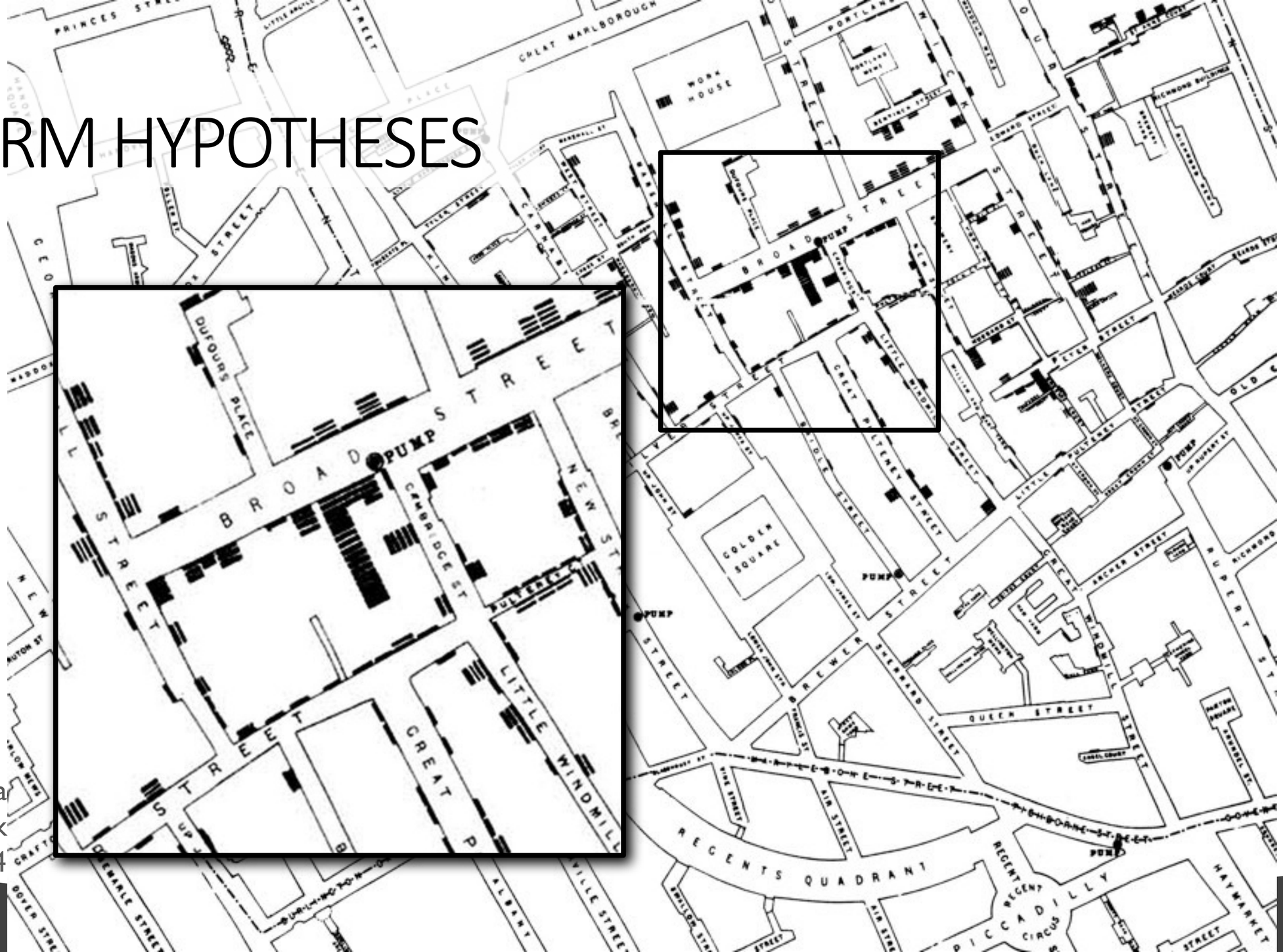
MOTOR	MBT	AMB	O-RING	WIND
DM-1	68	36	47	10 MPH
DM-2	76	45	52	10 MPH
QM-3	72.5	40	48	10 MPH
QM-4	76	48	51	10 MPH
SRM-15	52	64	53	10 MPH
SRM-22	77	78	75	10 MPH
SRM-25	55	26	29 27	10 MPH 25 MPH

source: E. Tufte



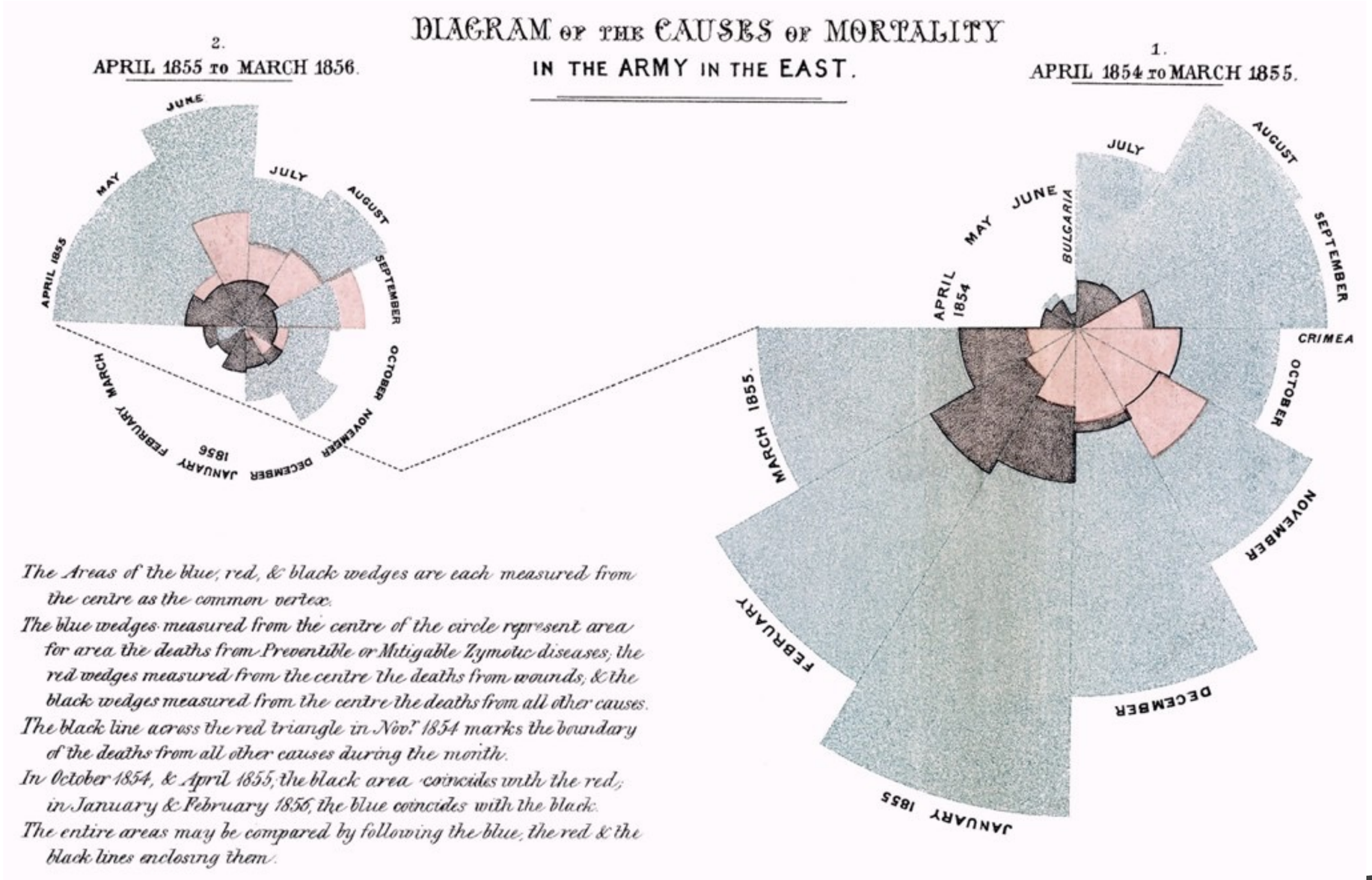
source: E. Tufte

CONFIRM HYPOTHESES

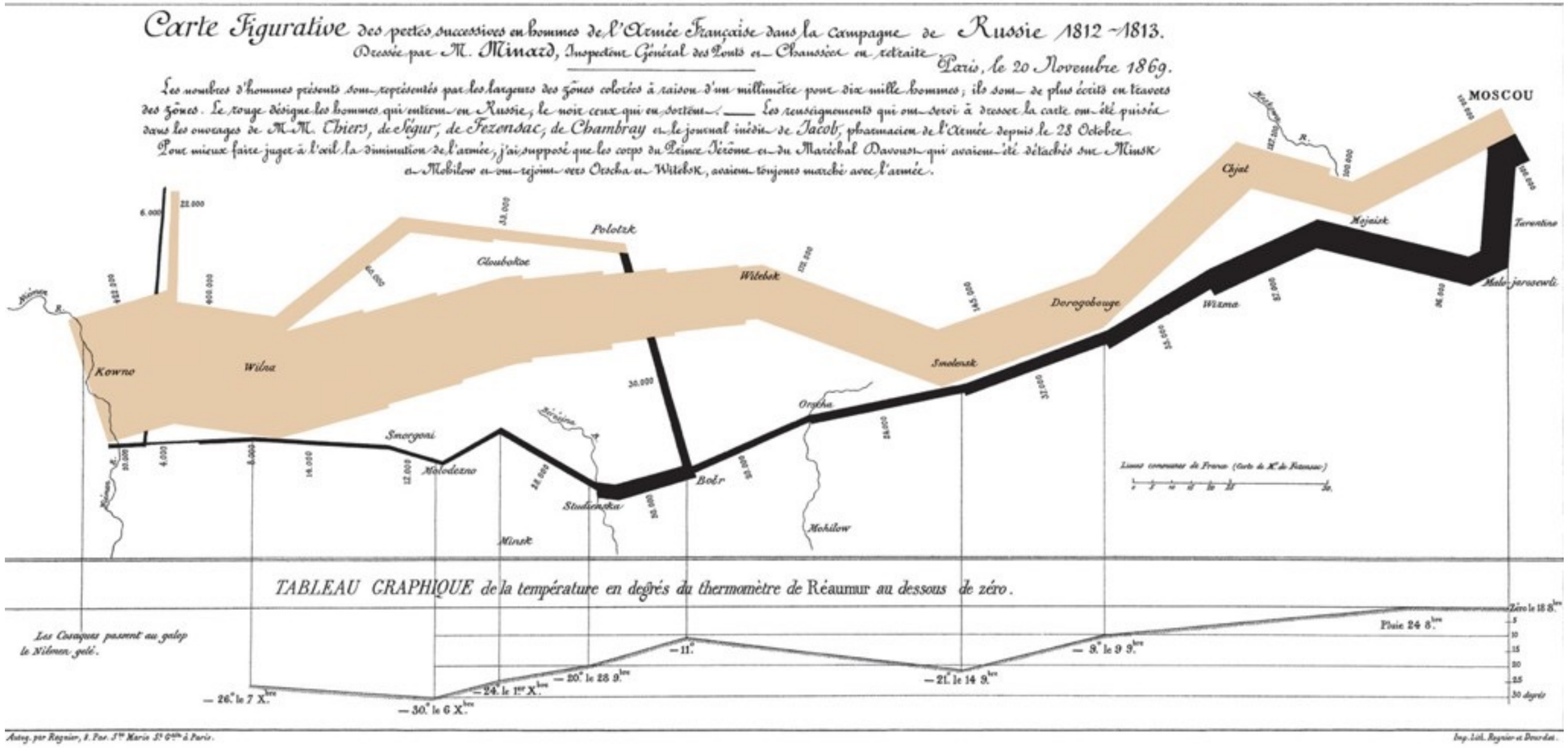


London Cholera Outbreak
John Snow 1854

COMMUNICATE IDEAS



COMMUNICATE IDEAS



Joseph Minard 1861

why does visualization work?

why does visualization work?

cognition is limited

Visualization uses perception to point out interesting things

MTHIVLWYADCEQGHKILKMTWYN
ARDCAIREQGHLVKMFPSTWYARN
GFPSVCEILQGKMFPSNVRCEQDI
PSGHLMFHKMVPSTWYACEQTWRN

MTHI**V**LWYADCEQGHKILKMTWYN

ARDCAIREQGHL**V**KMFPSTWYARN

GFPS**V**CEILQGKMFPSNVRCEQDI

PSGHLMFHKM**V**PSTWYACEQTWRN

MTHI**V**LWYADCEQGHKILKMTWYN

ARDCAIREQGHL**V**KMFPSTWYARN

GFPS**V**CEILQGKMFPSN**V**RCEQDI

PSGHLMFHKM**V**PSTWYACEQTWRN

(but can also deceive you)

why does visualization work?

memory is limited

$$\begin{array}{r} 34 \\ \text{calculation exercise . . .} \quad \underline{\times\ 28} \end{array}$$

79

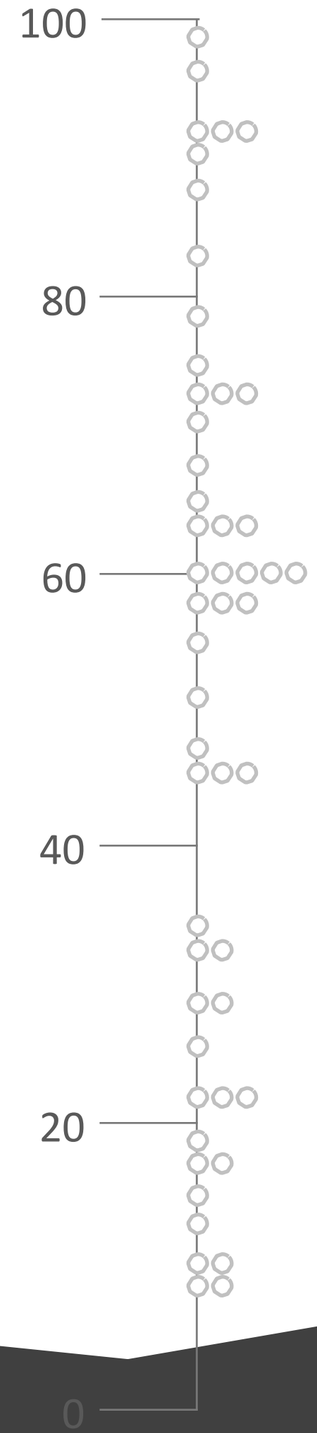
calculation exercise . . . x 16

$$\begin{array}{r} 34 \\ \text{calculation exercise . . .} \quad \underline{\times\ 28} \end{array}$$

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?



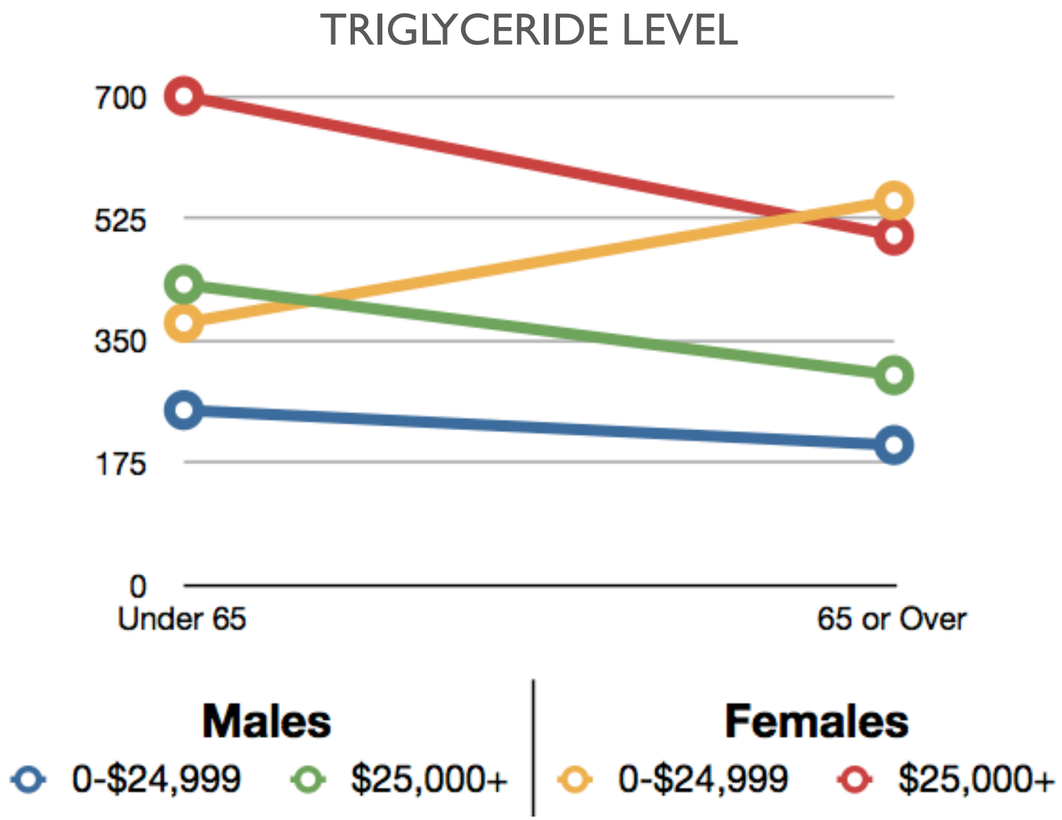
given these numbers . . .
. . . what number
appears most often?

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?



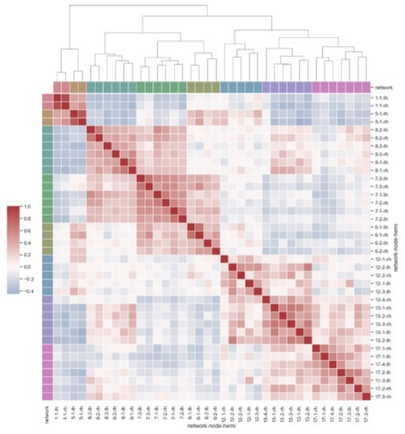
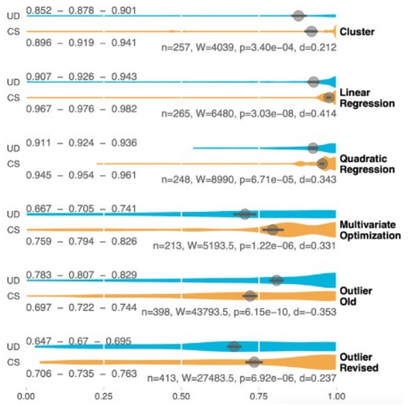
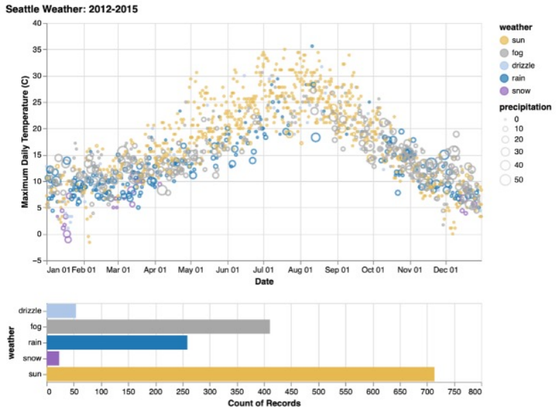
QUESTION: Which gender and income level shows a different effect of age on triglyceride levels?

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 and ability in HTML, CSS, and JavaScript. We will cover these topics but move quickly.

Alternative: COMP 5960 – Applied Data Visualization

- New Course!
- Designed for **Non-CS Students**
- Less focus on visualization research, more on practical data visualization
- How to use Python and Tableau for custom visualizations
- Requires less programming expertise
- **https://www.dataviscourse.net/2023-applied/**