

# CIS 4930/6930-002

## DATA VISUALIZATION



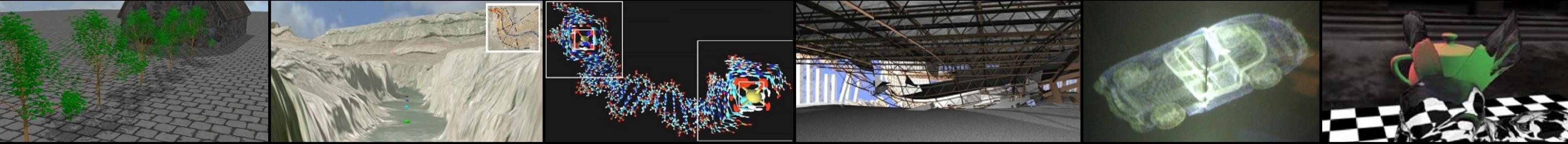
### Introduction to Visualization

Paul Rosen

Assistant Professor

University of South Florida

(slides adapted from Miriah Meyer)



# WHY STUDY VISUALIZATION?





# INDUSTRIAL REVOLUTION OF DATA

Joe Hellerstein, UC Berkeley, 2008



DOMO

# DATA NEVER SLEEPS 5.0

How much data is generated *every minute*?

90% of all data today was created in the last two years—that's 2.5 quintillion bytes of data per day. In our 5th edition of Data Never Sleeps, we bring you the latest stats on just how much data is being created in the digital sphere—and the numbers are staggering.

## HOW MUCH DATA IS THERE?

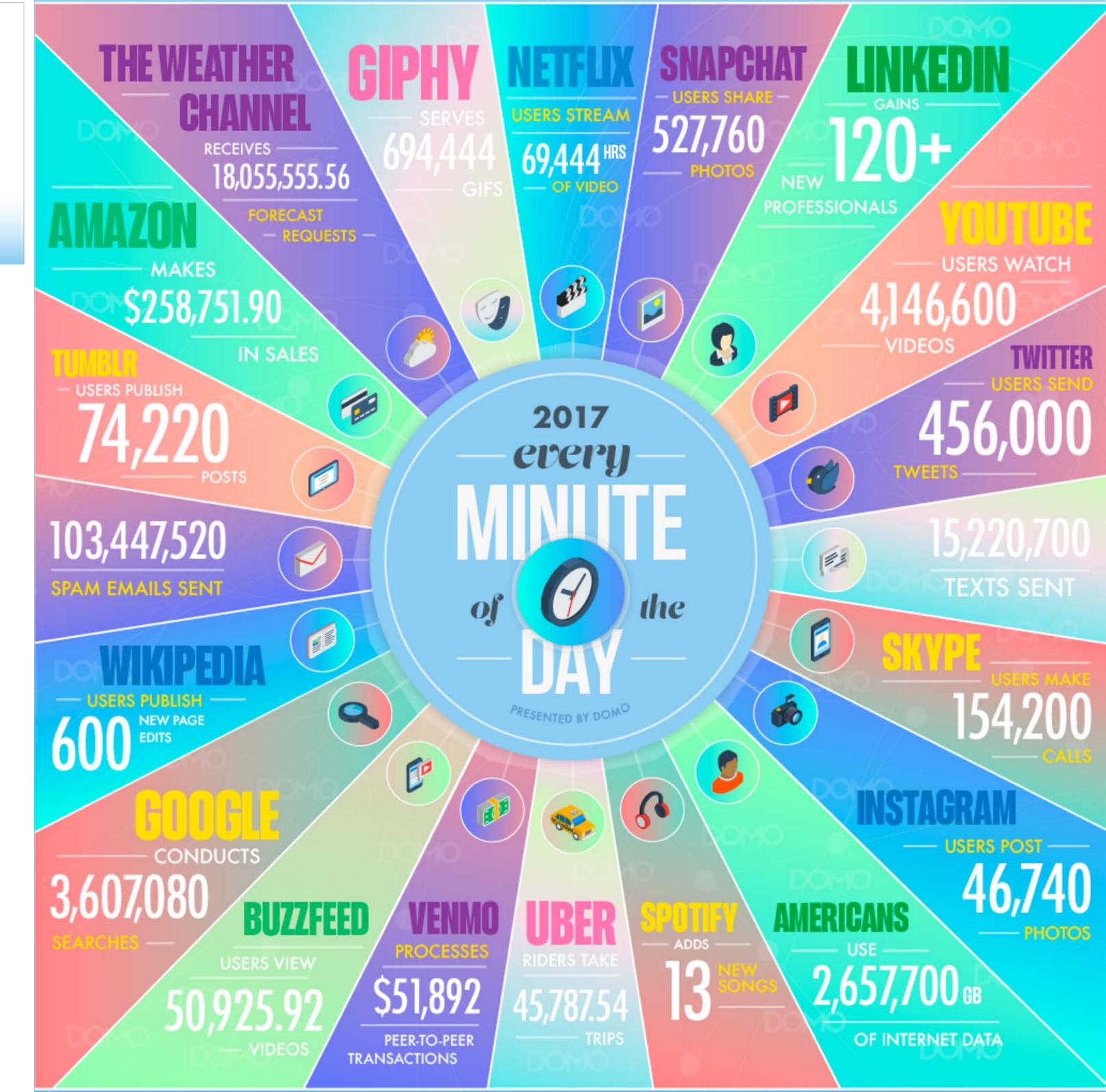
2010: 1.2 zettabytes

2013: 4.4 zettabytes

2020: ~40 zettabytes

2.5 quintillion bytes / day (2.5 EB)

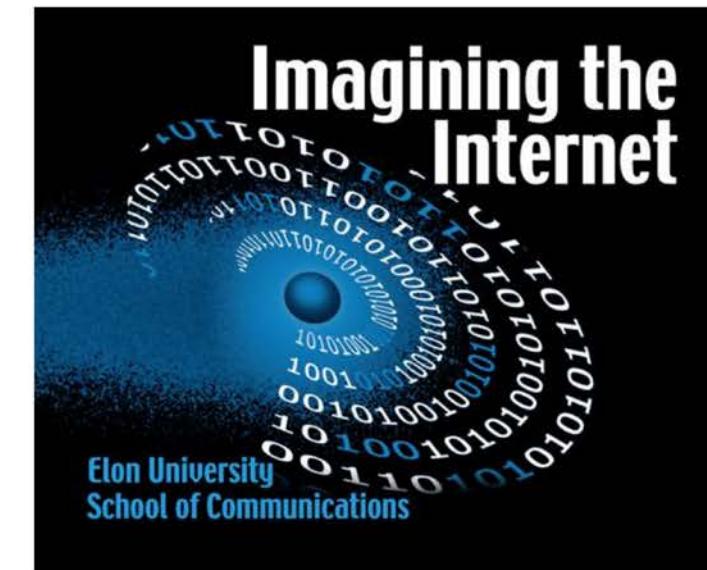
zettabyte ~=  
1,000,000,000,000,000,000 or  $10^{21}$   
(200x all words ever spoken by humans)



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.





## **Big Data: Experts say new forms of information analysis will help people be more nimble and adaptive, but worry over humans' capacity to understand and use these new tools well**

Tech experts believe the vast quantities of data that humans and machines will be creating by the year 2020 could enhance productivity, improve organizational transparency, and expand the frontier of the "knowable future."

But they worry about "humanity's dashboard" beingin government and corporate hands and they are anxious about people's ability to analyze it wisely.

Janna Quitney Anderson, Elon University  
Lee Rainie, Pew Research Center's Internet & American Life Project  
July 20,2012



# DATA SCIENCE TOP JOB IN 2017 & 2018 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

**\$110,000**  
Median Base Salary

**4.4 / 5**  
Job Satisfaction

**4,184**  
Job Openings

**View Jobs**



# WHY DO WE CREATE VISUALIZATIONS?

answer questions

generate hypotheses

make decisions

see data in context

expand memory

support computational analysis

find patterns

tell a story

inspire



## VISUALIZATION GOALS

**record** information

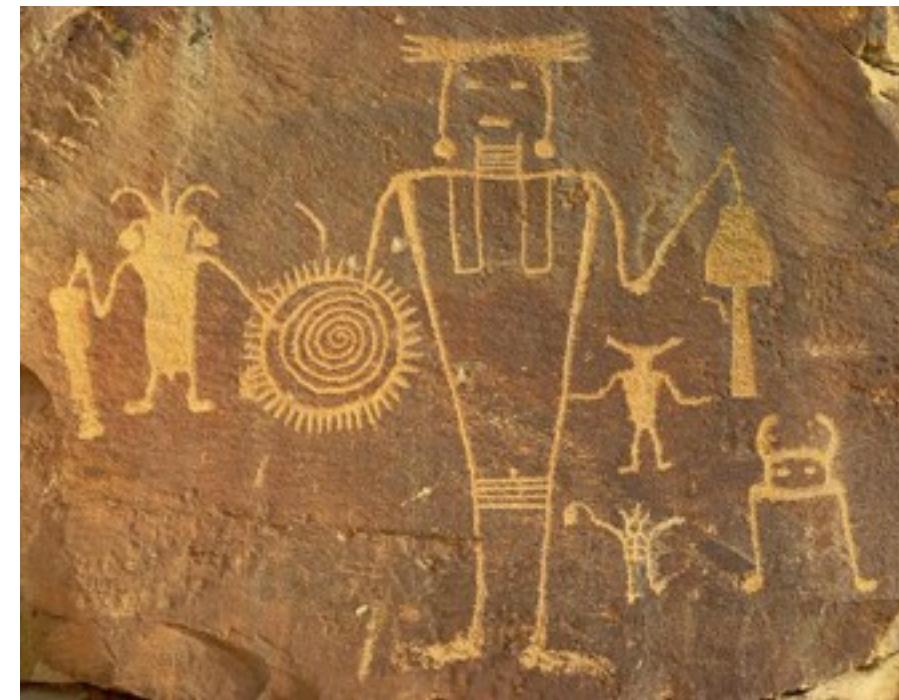
**analyze** data to support reasoning

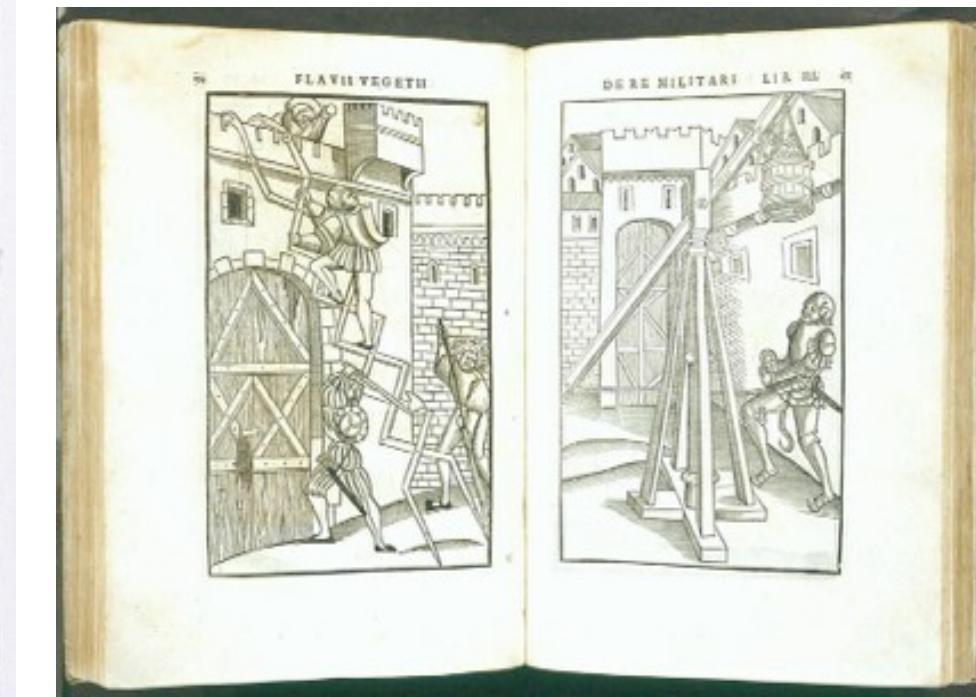
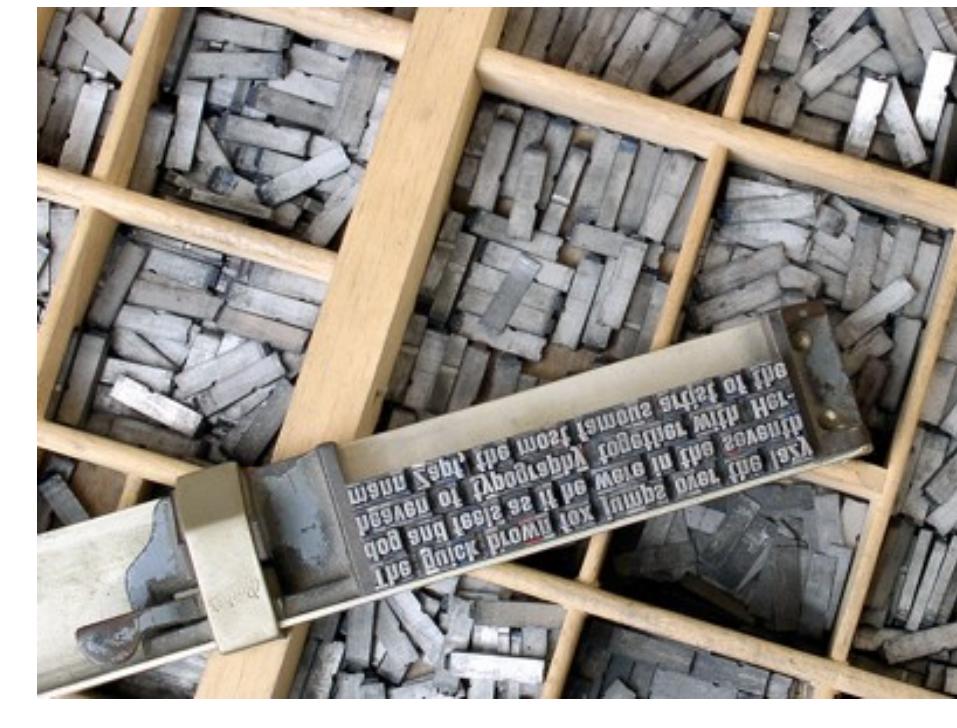
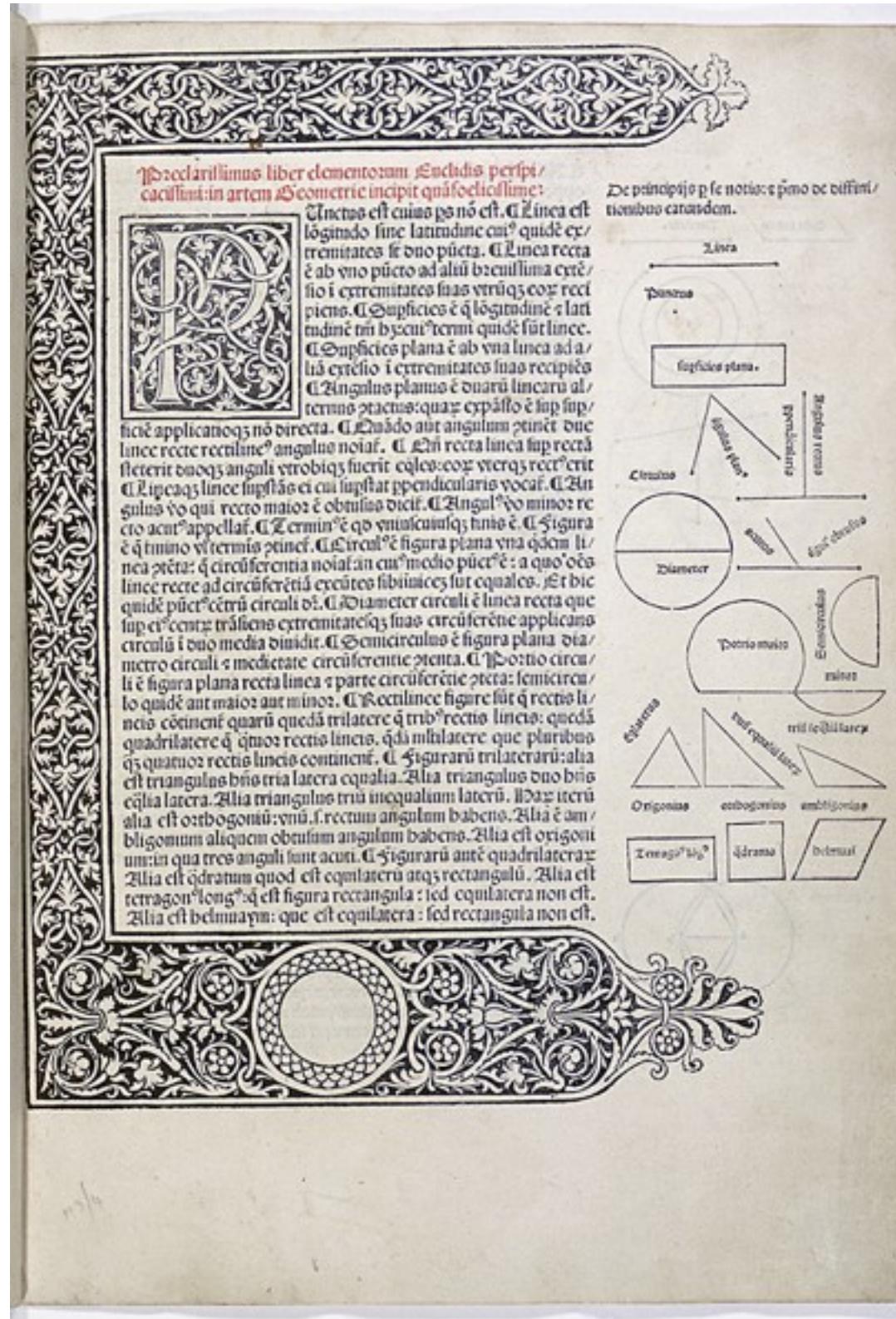
**confirm** hypotheses

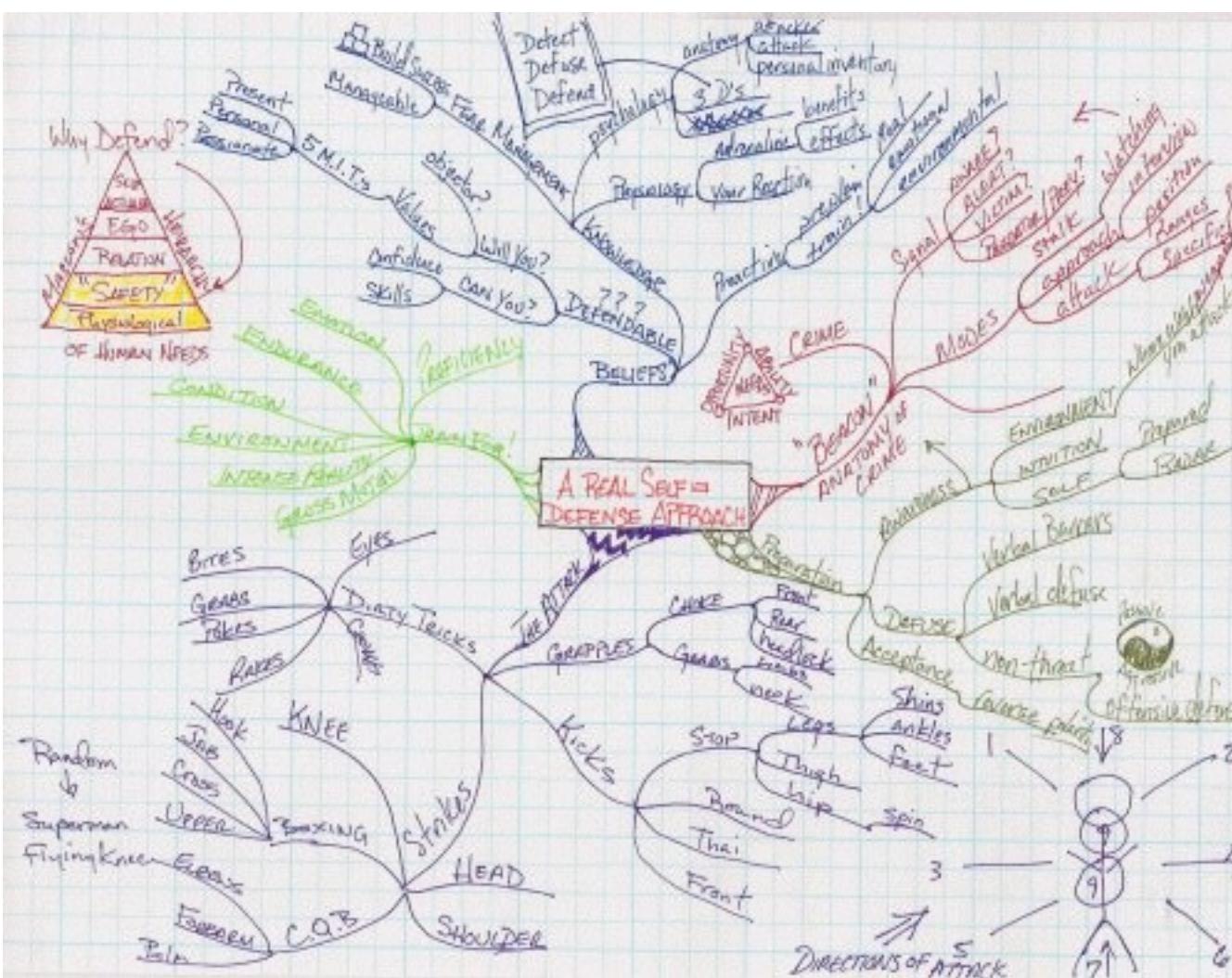
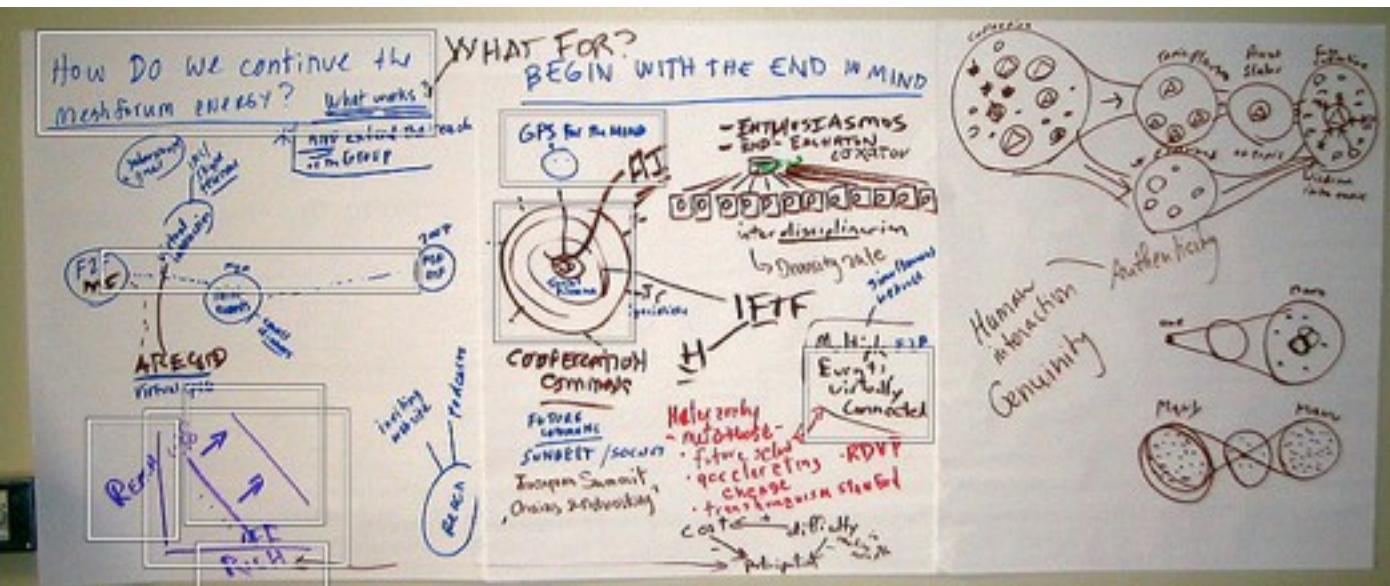
**communicate** ideas to others



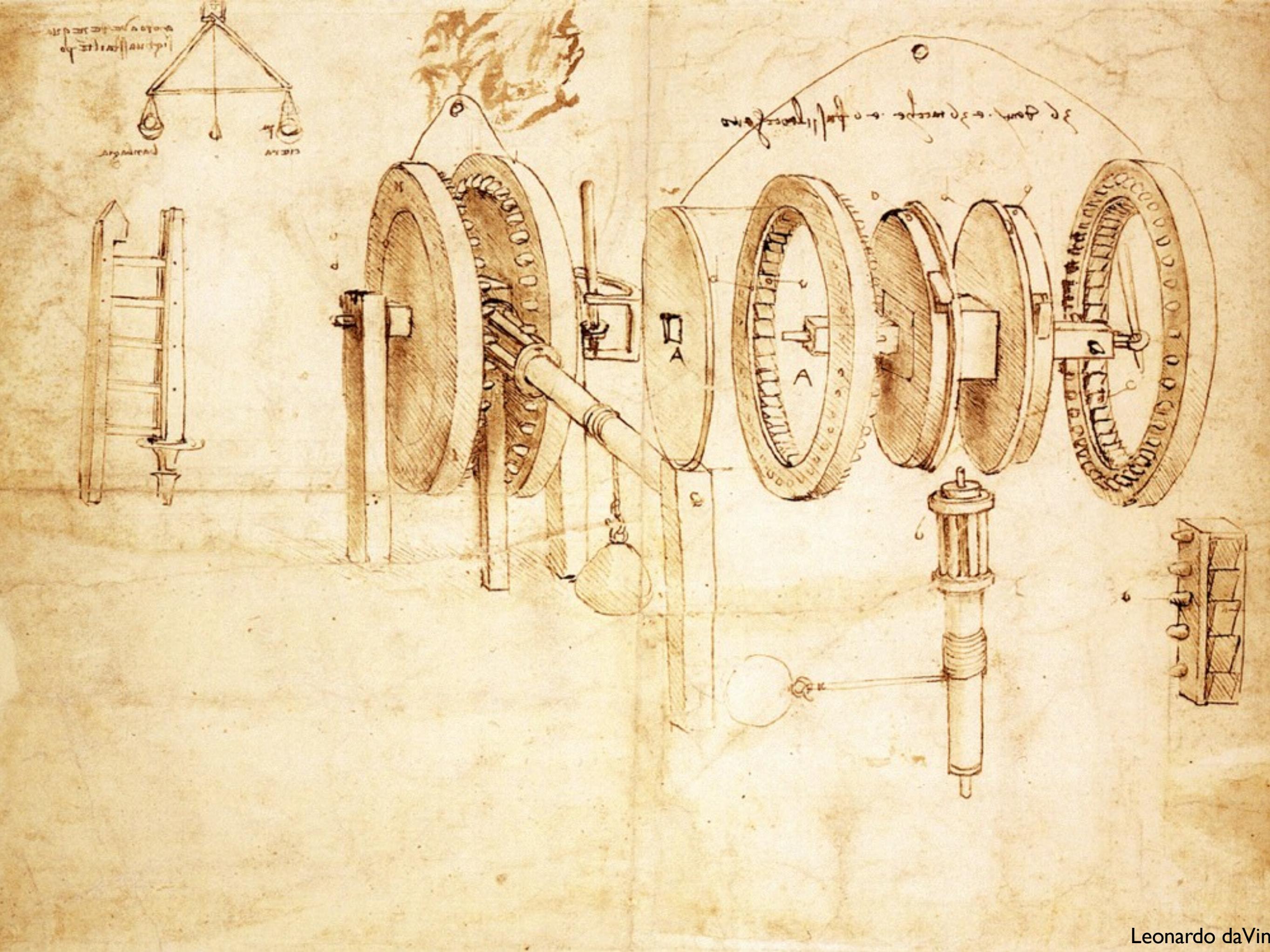
# Record Information





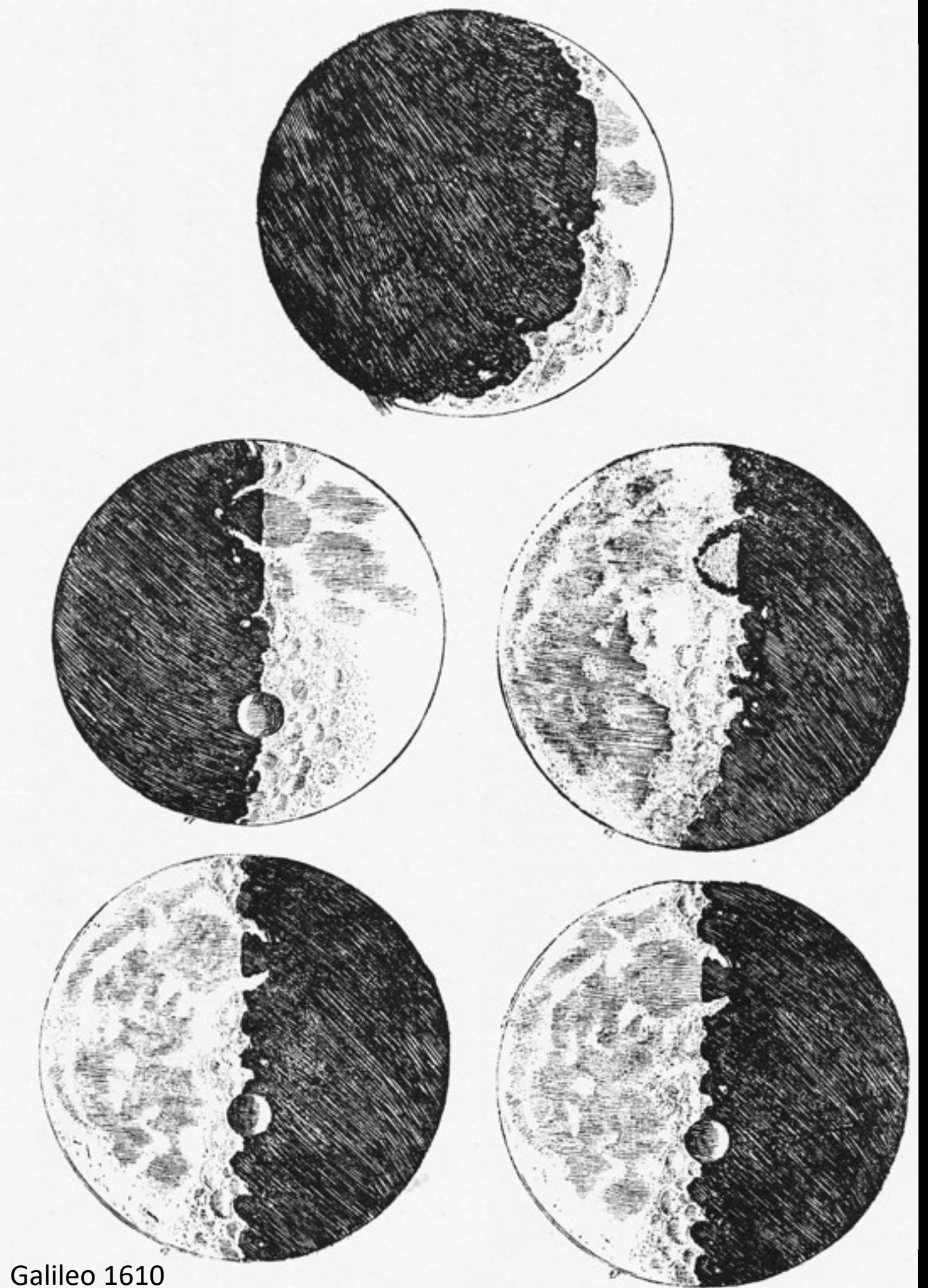


Visual Thinking Collection, Dave Grey

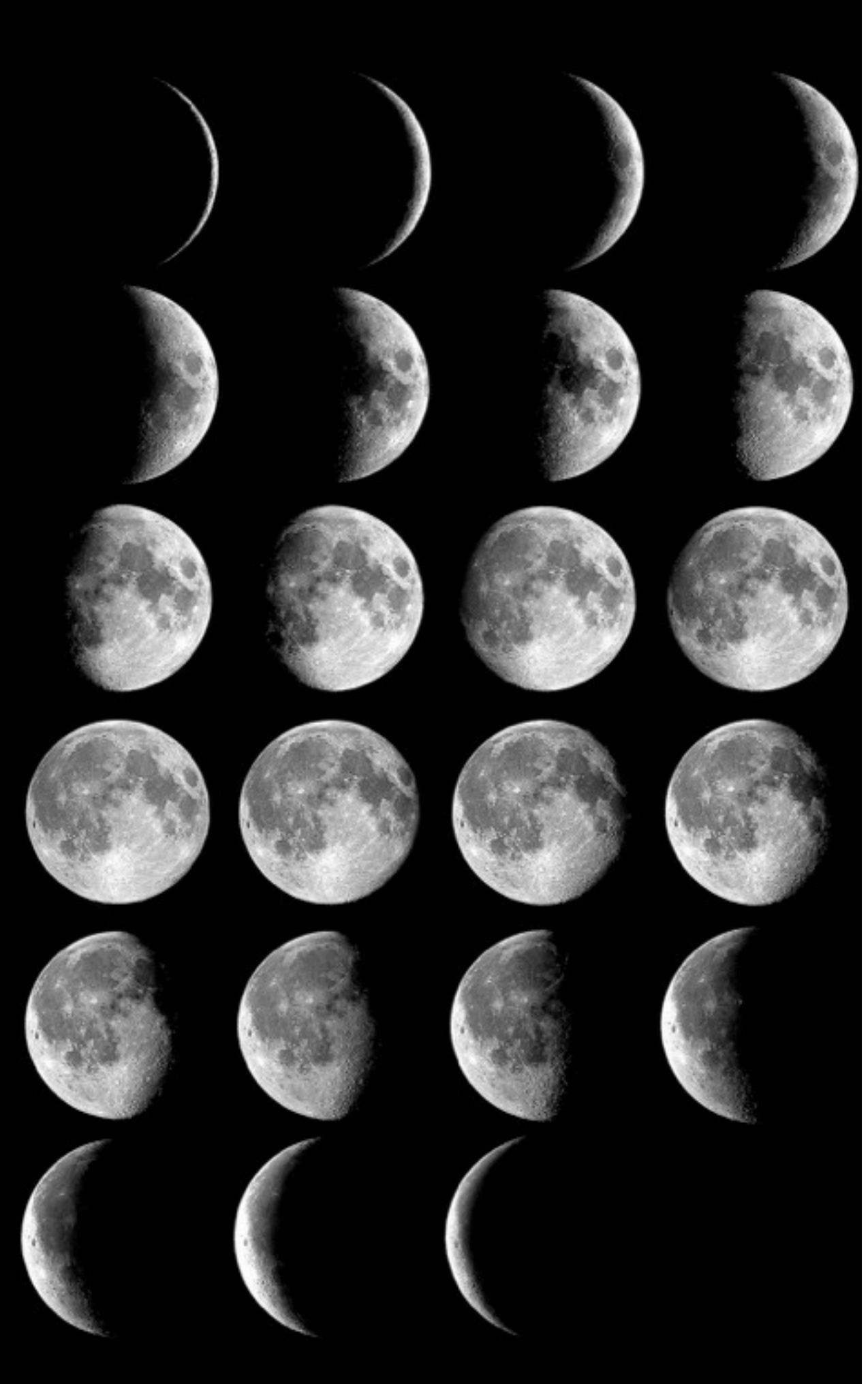


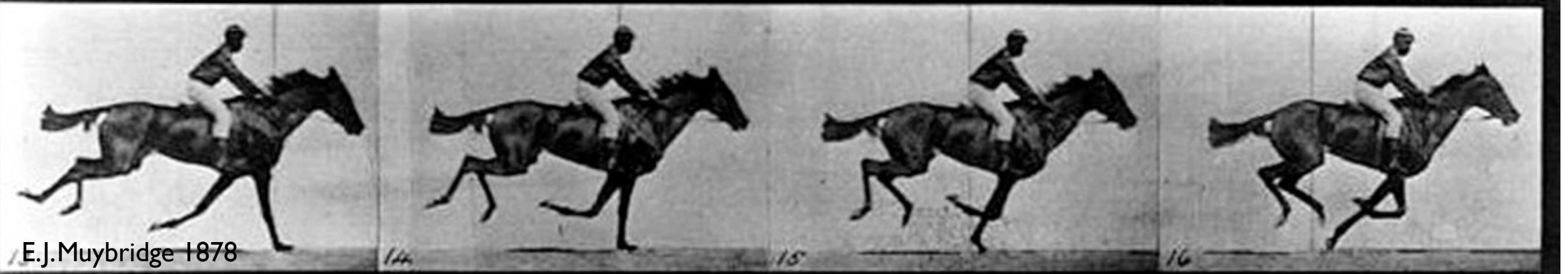
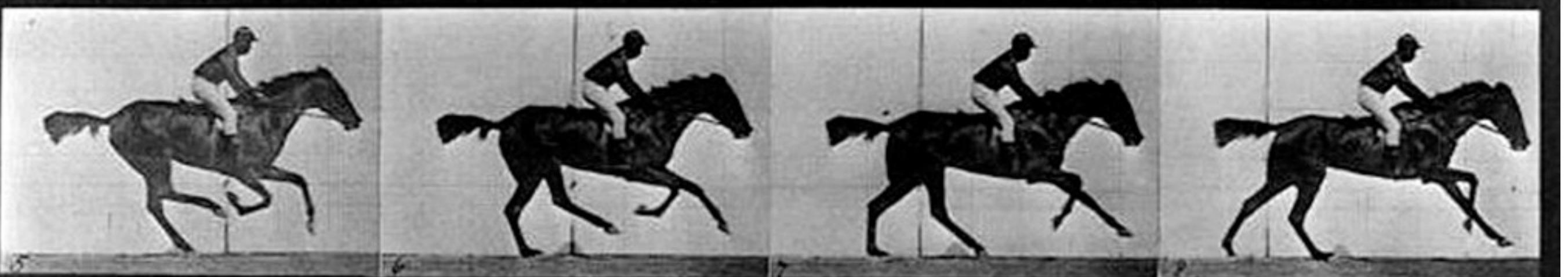
Leonardo daVinci 1485





Galileo 1610





E.J. Muybridge 1878

14

15

16



# ANALYSIS TO SUPPORT REASONING



# Mapping Migration in the United States

AUG. 15, 2014

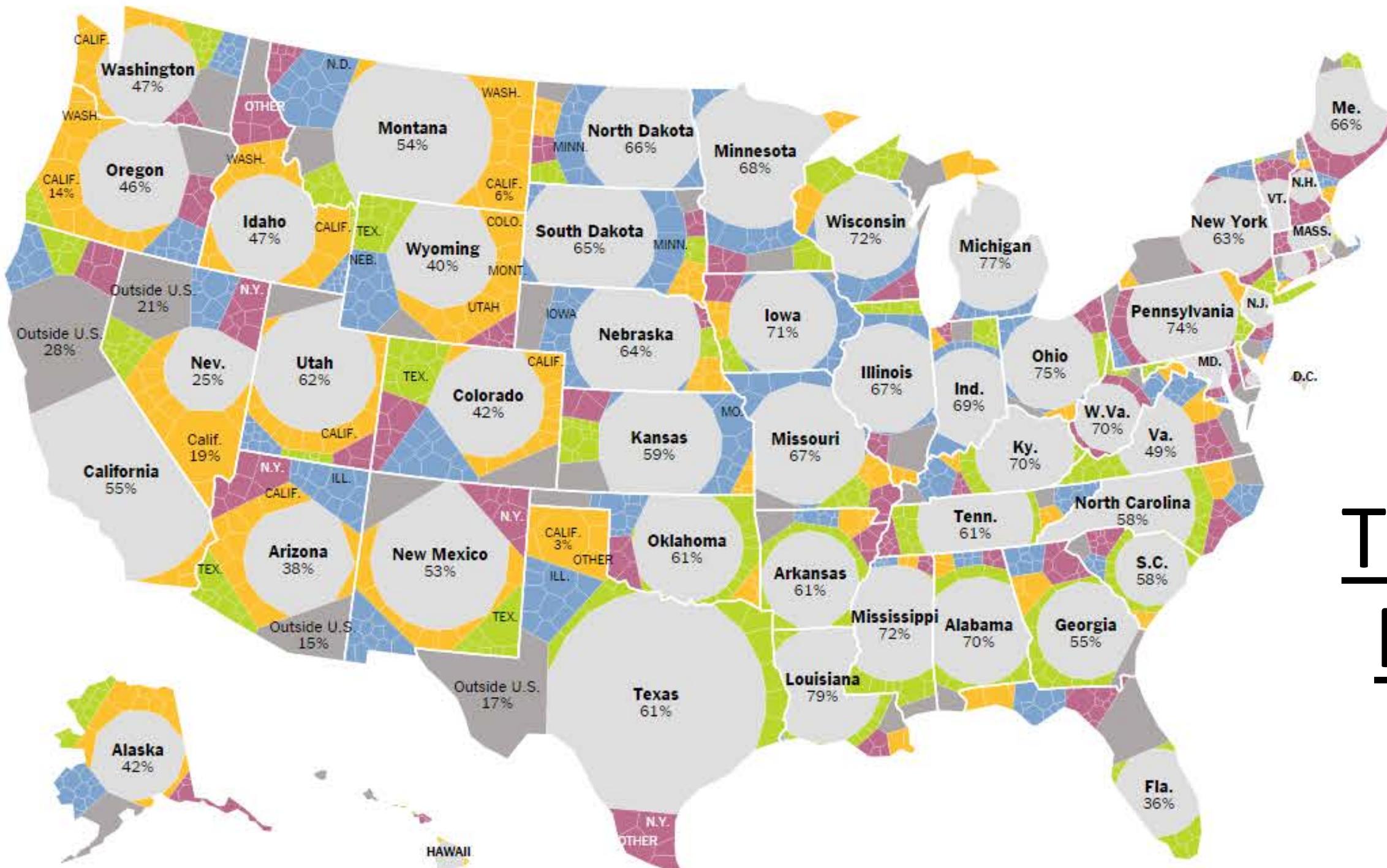
## Where people who lived in each state in 2012 were born

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.\*



# THE CHALLENGER DISASTER, 1986

Confirm  
Hypotheses



source:Wikipedia





HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

	SRM No.	Cross Sectional View			Top View			Clocking Location (deg)	MOTOR	O-RING
		Erosion Depth (in.)	Perimeter Affected (deg)	Nominal Dia. (in.)	Length Of Max Erosion (in.)	Total Heat Affected Length (in.)				
Oct 20, 1987 AFT	22A	None	None	0.280	None	None	36° - 66°	DM-4	47	
	22A	NONE	NONE	0.280	NONE	NONE	338° - 18°	DM-2	52	
15A	0.010	154.0	0.280	4.25	5.25	163				
15B	0.038	130.0	0.280	12.50	58.75	354				
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	QM-3	48	
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	QM-4	51	
STS-2 RH Aft Field	2B	0.053	116.0	0.280	--	--	90			
								SRM-15	53	
								SRM-22	75	
								SRM-25	29	
									27	

\*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

- 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

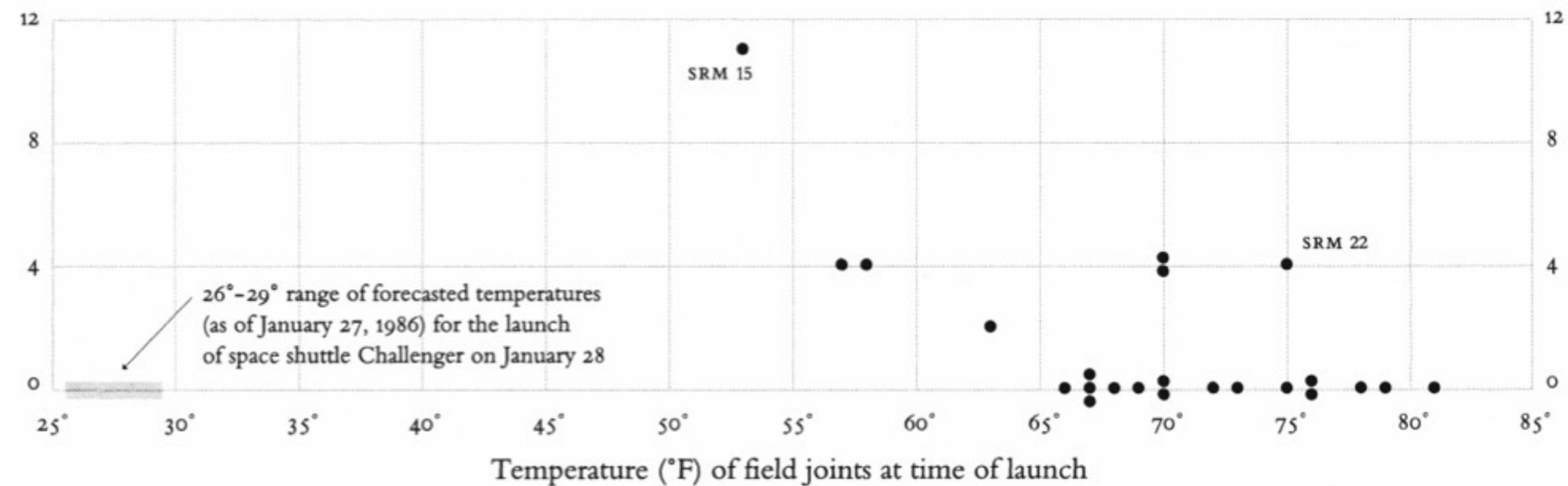
- NOZZLE BLOW-BY

### HISTORY OF O-RING TEMPERATURES (DEGREES - F)

MOTOR	MBT	AMB	O-RING	WIND
DM-4	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



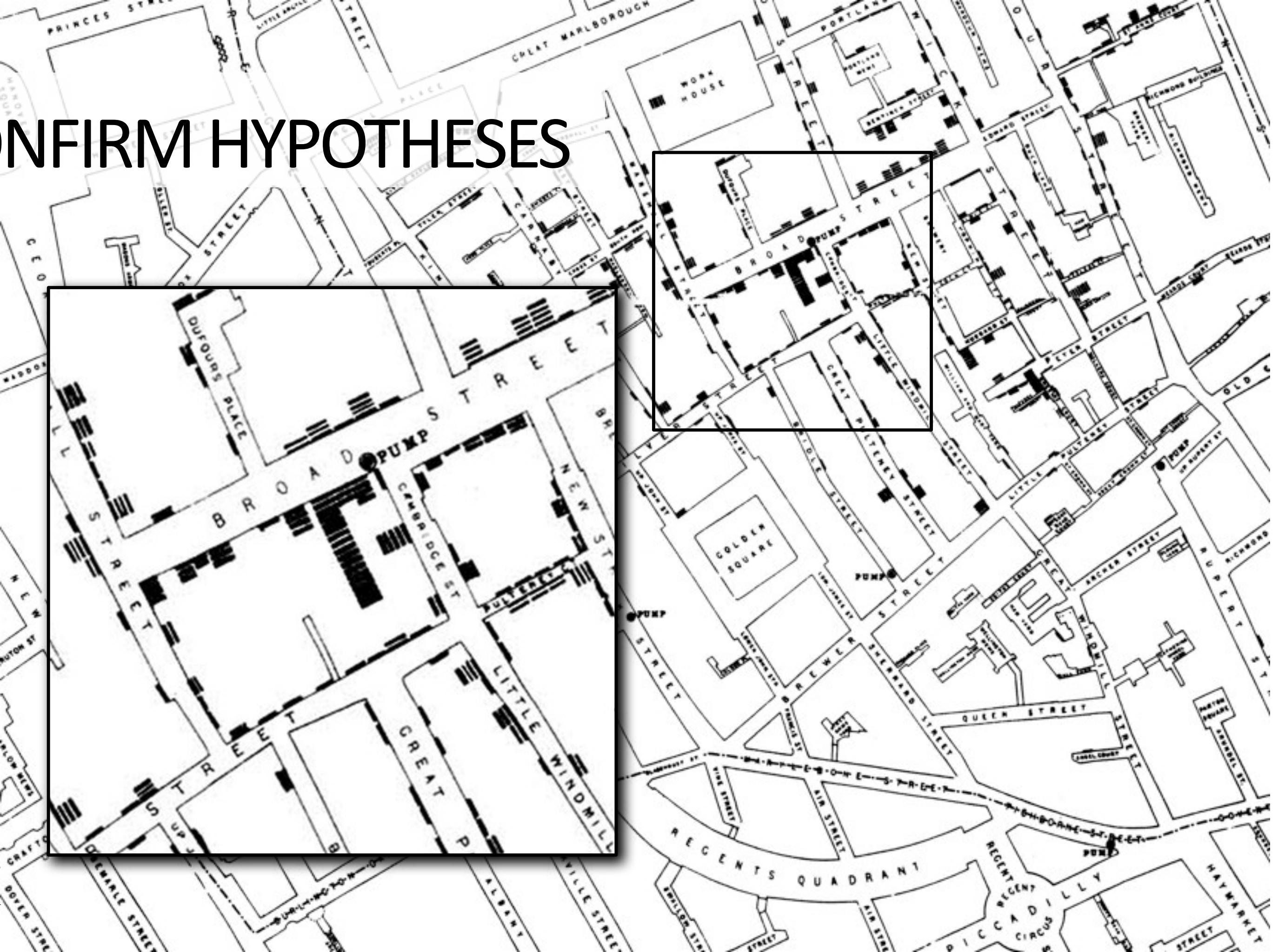
O-ring damage  
index, each launch



source: E. Tufte



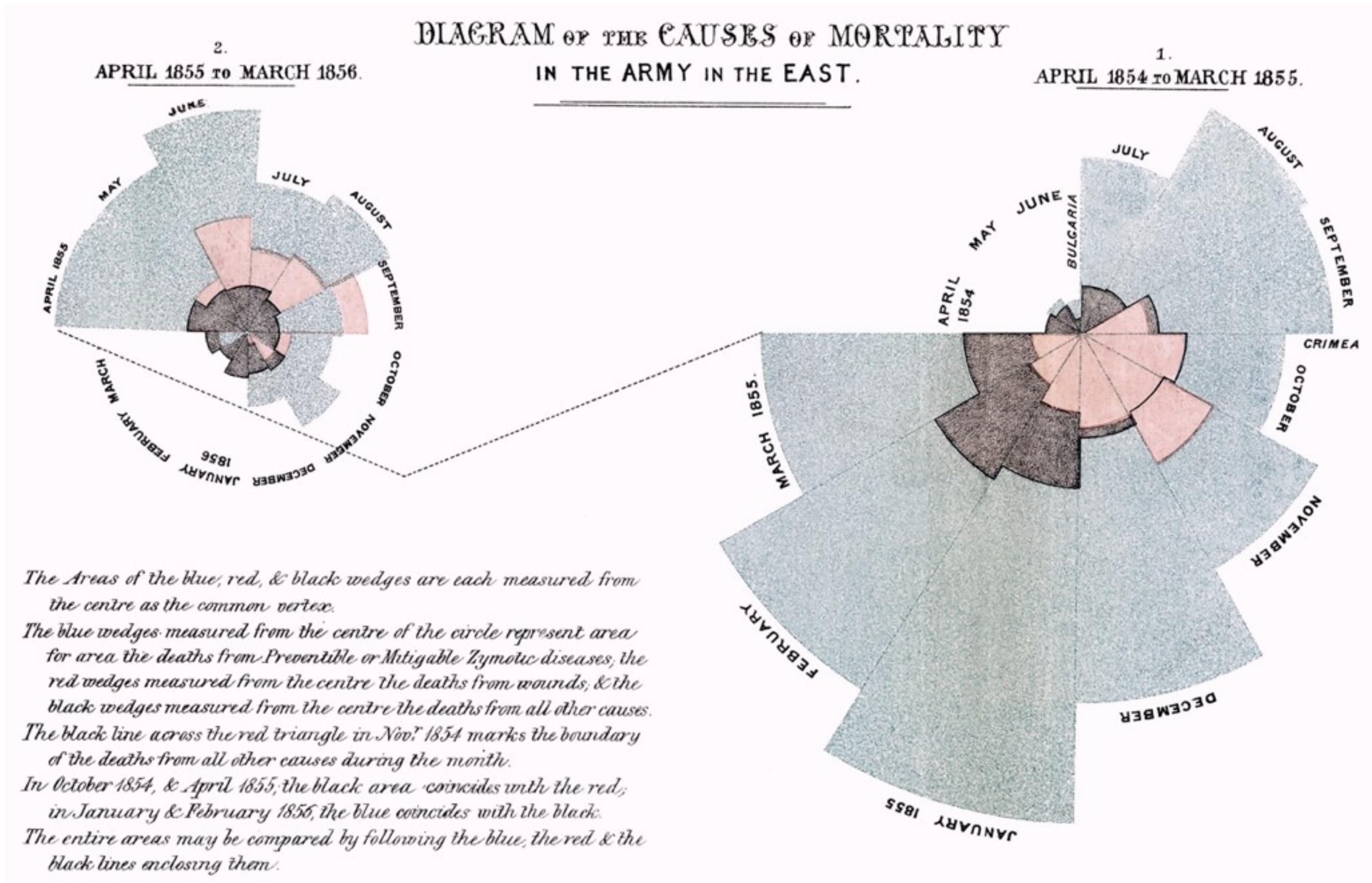
# CONFIRM HYPOTHESES



London Cholera  
Outbreak  
John Snow 1854



# COMMUNICATE IDEAS



Florence Nightingale 1856



*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 perdus 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 dessiner la carte ont été pris dans les ouvrages de M. Chiers, de Séguir, de Fezensac, de Chambray et le journal intime 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 Napoléon et du Maréchal Davout, qui avaient été détachés sur Minsk et Malibor et qui rejoignirent Oryha au Wilejka, avaient toujours marché avec l'armée.

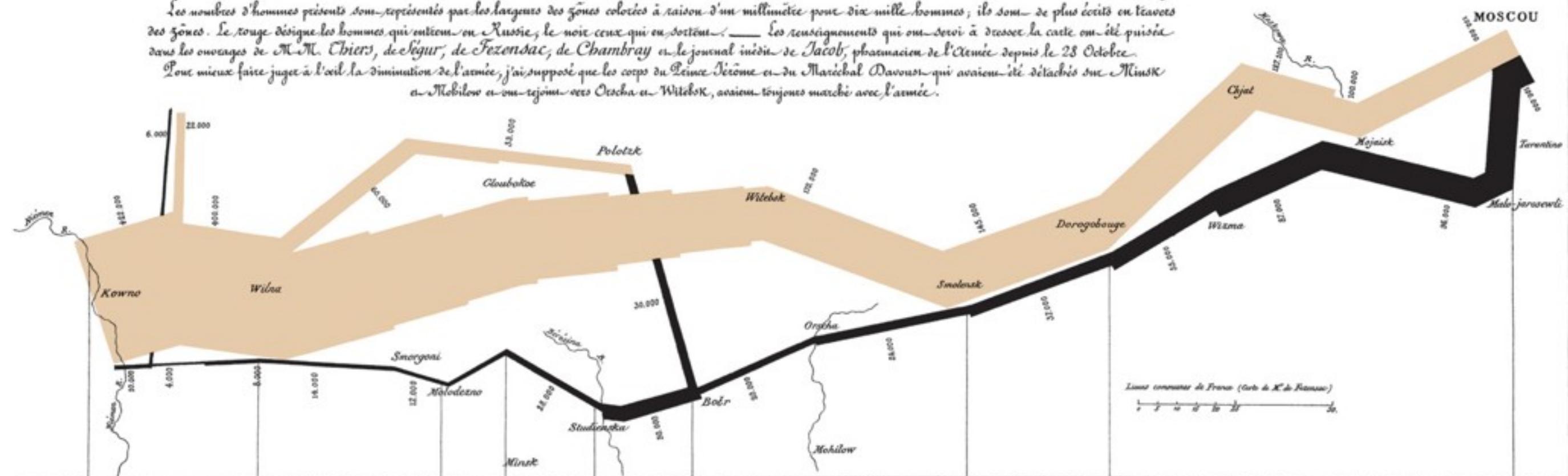


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Cosaques passent au galop  
le Nilovka gelé.

— 26° le 7 X.  
— 30° le 6 X.  
— 24° le 1<sup>er</sup> X.  
— 20° le 28 9<sup>me</sup>.  
— 11°.

— 21° le 14 9<sup>me</sup>.

— 9° le 9 9<sup>me</sup>.

Zéro le 15 8<sup>me</sup>.  
Pluie 24 8<sup>me</sup>.  
10  
15  
20  
25  
30 degrés.

Impr. par Regnier, 1. Rue S<sup>e</sup> Marie S<sup>e</sup> G<sup>e</sup> de Paris.

Imp. L. Regnier et Bourdet.

Joseph Minard 1861

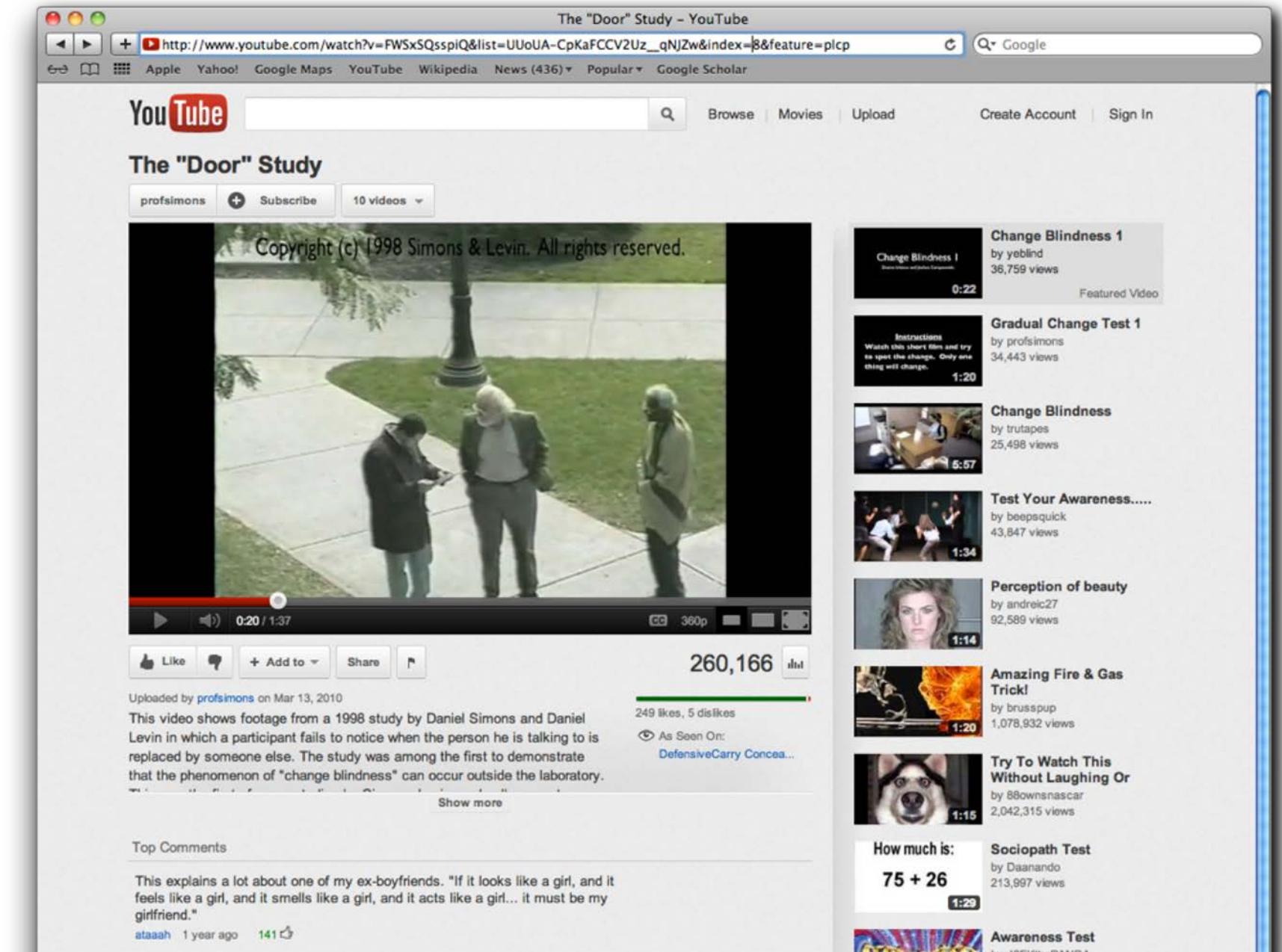


# WHY DOES VISUALIZATION WORK?



# WHY DOES VISUALIZATION WORK?

## cognition is limited



## WHY DOES VISUALIZATION WORK?

cognition is limited  
memory is limited



calculation exercise . . .

952



calculation exercise . . .

1264



34

calculation exercise . . . x28



# How VISUALIZATION WORKS



## VISUALIZATION

uses perception to point  
out interesting things



MTHIVLWYADCEQGHKILKMTWYN  
ARDCAIREQGHLVKMFPSTWYARN  
GFPSVCEILQGKMFPSNDRCEQDI  
PSGHLMFHKMVPSTWYACEQTWRN



MTHIVLWYADCEQGHKILKMTWYN  
ARDCAIREQGHLVKMFPSTWYARN  
GFPSSVCEILQGKMFPSNDRCEQDI  
PSGHLMFHKMVPSTWYACEQTWRN



## 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

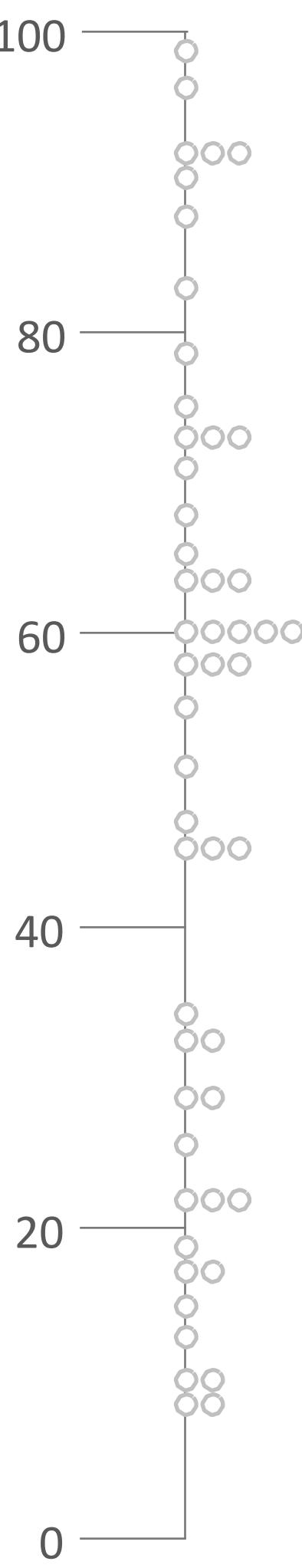
91 29 57

96 12 47

**GIVEN THESE NUMBERS . . .**

. . . what number appears most often?





GIVEN THESE NUMBERS . . .  
. . . what number appears most often?



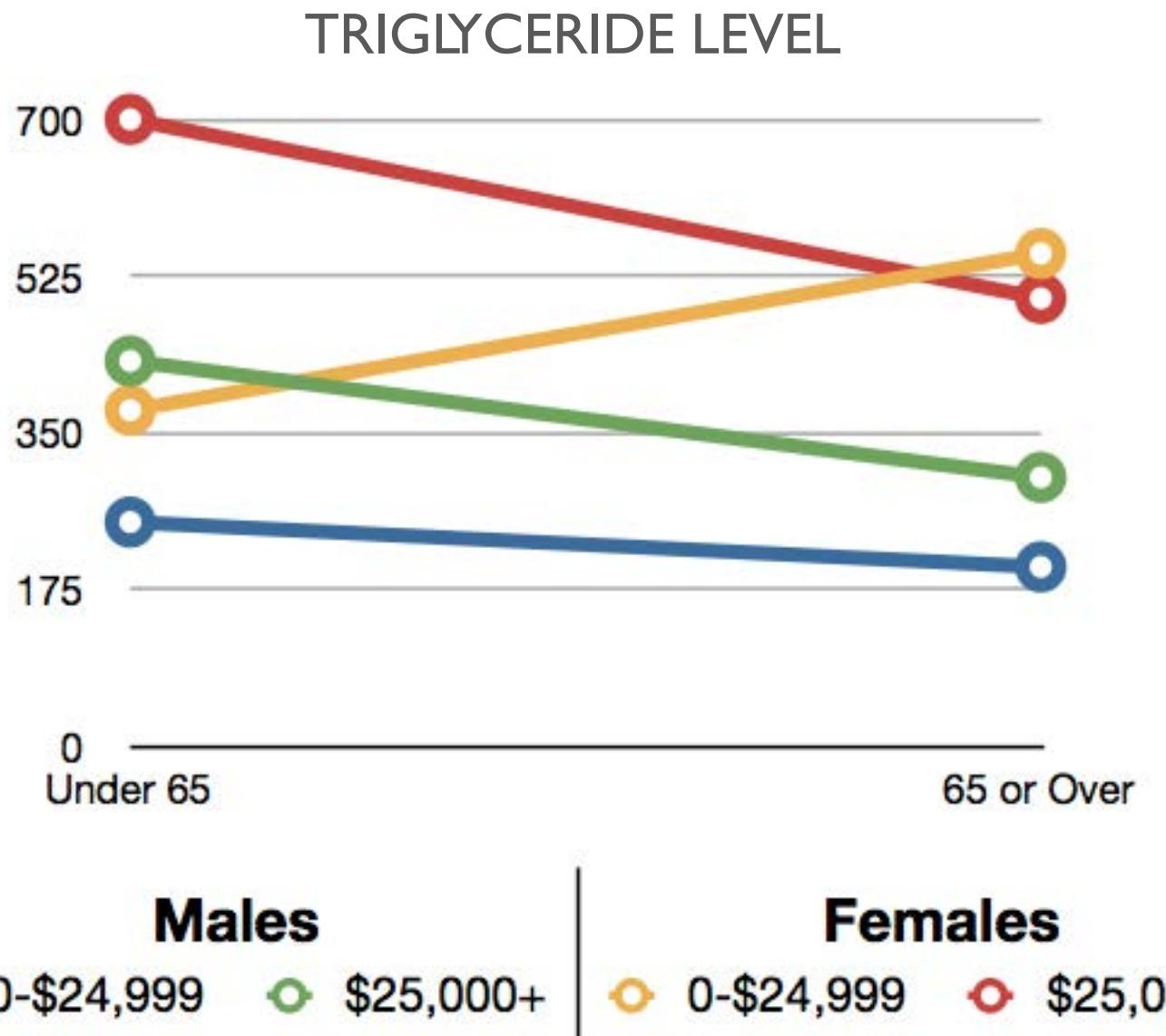
## QUERY EXERCISE . . .

TRIGLYCERIDE LEVEL

Income Group	Males		Females	
	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** using visual analysis tools, as well as in **developing your own** interactive visualization tools.



# FOUNDATIONS

design

perception

data and tasks

visual encoding

interaction

multiple views

filtering and aggregation



# TECHNIQUES

tabular data

graphs and trees

text

sets

maps



## ADVANCED TOPICS

High dimensional data

Vis + Stats

Vis + Machine Learning

Vis + Topology



## GROUND RULES

be respectful of everyone's time

come to class prepared

critique ideas, not people

discuss ideas together, do the  
assignments on your own

be engaged!



# Recommended Watching

Hans Rosling shows the best stats you've ever seen | Video on TED.com

http://www.ted.com/talks/hans\_rosling\_shows\_the\_best\_stats\_you\_ve\_ever\_seen.html

RSS Google

Apple Yahoo! Google Maps YouTube Wikipedia News (457) Popular Google Scholar

**TED** Ideas worth spreading

**TALKS**

**Hans Rosling shows the best stats you've ever seen**

TED2006, Filmed Feb 2006; Posted Jun 2006



3,471,109 Views [Like](#) 33k

**INTERACTIVE TRANSCRIPT**

**ABOUT THE SPEAKER**

**ABOUT THIS TALK**

You've never seen data presented like this. With the drama and urgency of a sportscaster, statistics guru Hans Rosling debunks myths about the so-called "developing world."

THE ROLEX ARTS INITIATIVE PAIRS ESTABLISHED MENTORS WITH EMERGING PROTÉGÉS FOR A YEAR OF CREATIVE COLLABORATION

WHAT TO WATCH NEXT

 Hans Rosling's new insights on poverty  
18:57 Posted: Jun 2007 Views 1,616,080 | Comments 193

Subtitles Available in: 45 languages [Off]

00:17 | 19:53 Share Rate



## RECOMMENDED READING

Visualization Analysis & Design: Chapter 1 (pp. 1-19)



