



CDS 324 DATA VISUALIZATION

LECTURE 1: INTRODUCTION



**How much data is created
EVERY DAY?**

402.74

million terabytes
of data are created
each day.



402.74
million terabytes
of data are created
each day.

147

zettabytes / year



12

zettabytes / month



0.4

zettabytes / day

Unit	Abbreviation	Approximate size
bit	b	binary digit, a single 0 or 1
byte	B	8 bits
kilobyte	KB	1,024 bytes or 10^3 bytes
megabyte	MB	1,024 KB or 10^6 bytes
gigabyte	GB	1,024 MB or 10^9 bytes
terabyte	TB	1,024 GB or 10^{12} bytes
petabyte	PB	1,024 TB or 10^{15} bytes
exabyte	EB	1,024 PB or 10^{18} bytes
zettabyte	ZB	1,024 EB or 10^{21} bytes
yottabyte	YB	1,024 ZB or 10^{24} bytes

Important Statistics About How Much Data Is Created Every Day

1 How much data is generated every minute?

Source: Domo

 **41,666,667**

messages shared
by WhatsApp users

 **1,388,889**

video / voice calls made
by people worldwide

 **404,444**

hours of video streamed
by Netflix users

 **347,222**

stories posted by Instagram users

 **150,000**

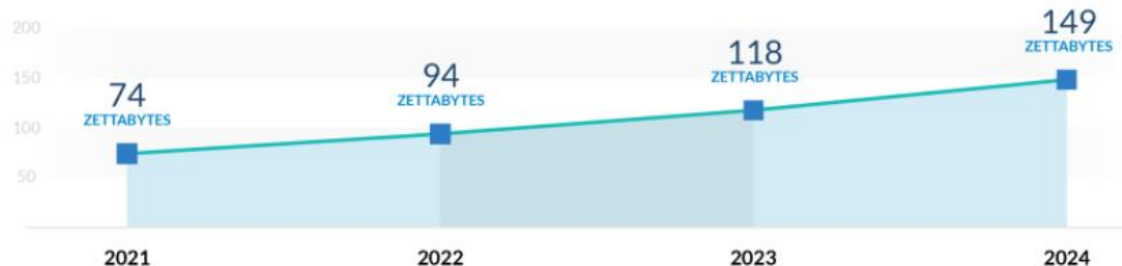
messages shared by Facebook users

 **147,000**

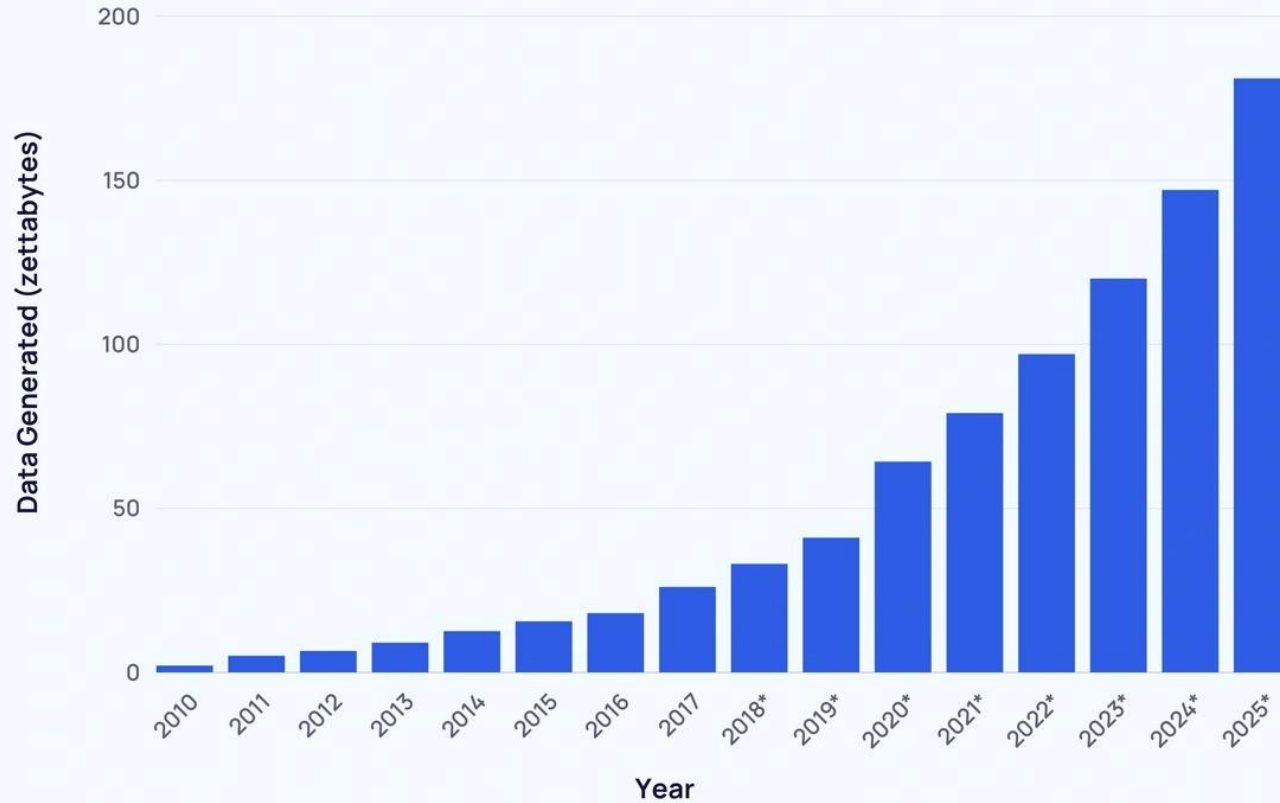
photos shared by Facebook users

2 Estimated Data Consumption from 2021 to 2024

Source: IDC / Statista



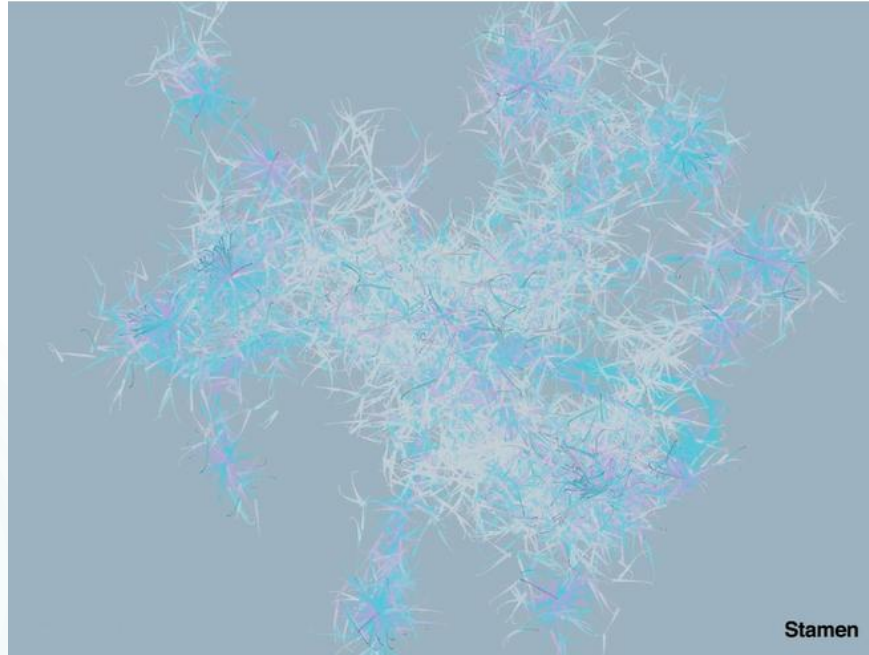
Global Data Generated Annually



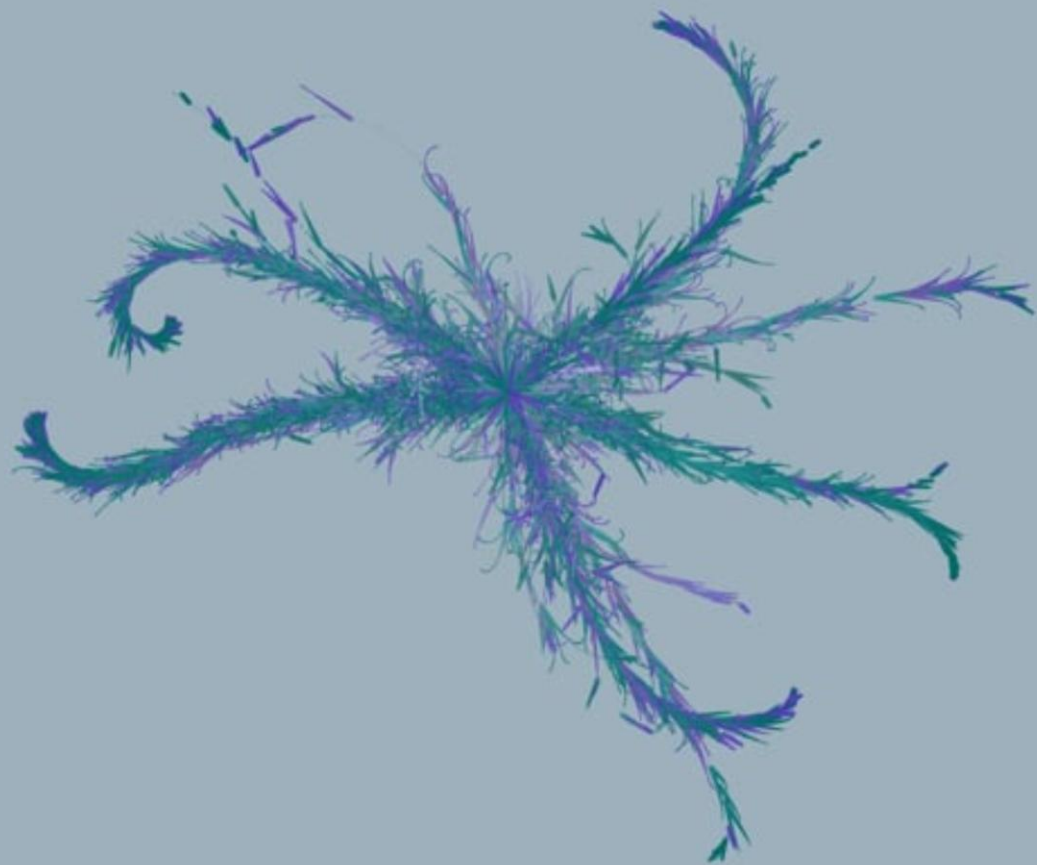
Social Media



Facebook Flower



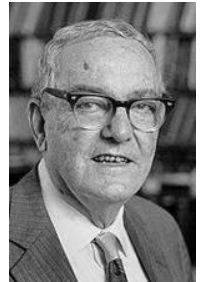
<https://stamen.com/work/facebook-flowers/>



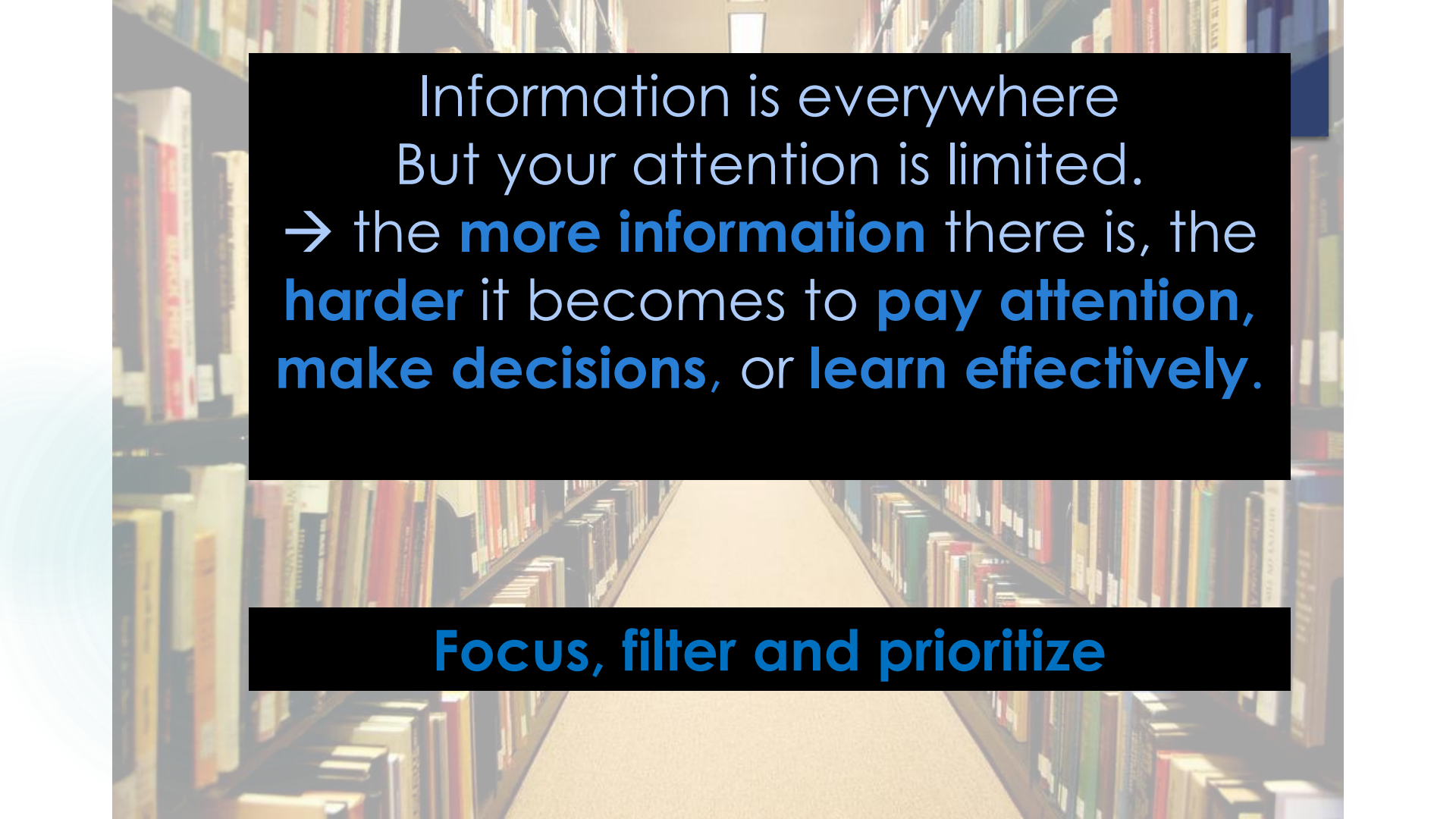
2012.08.06 02:00

Stamen

“In an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes.



~Herbert Simon



Information is everywhere
But your attention is limited.
→ the **more information** there is, the
harder it becomes to **pay attention,**
make decisions, or **learn effectively.**

Focus, filter and prioritize



▶ **HOW** might we use **VISUALIZATION** to **EMPOWER** understanding of data and analysis processes?



What is **VISUALIZATION**?

B. McCormick, T. DeFanti, and M. Brown, 1987

Visualization is a method of computing. It **transforms the symbolic into the geometric**, enabling researchers to observe their simulations and computations. Visualization offers a method for **seeing the unseen**. It enriches the process of scientific discovery and fosters profound and unexpected insights. In many fields it is already revolutionizing the way scientists do science.

Stuart Card, 2007

The purpose of information visualization is to **amplify cognitive performance, not just to create interesting pictures.** Information visualizations should do for the mind what automobiles do for the feet.

Stuart Card, Information visualization, in A. Sears and J.A. Jacko (eds.)
The Human-Computer Interaction Handbook, 2007

Modern definition, 2018-2022

Data visualization is the practice of **translating information into a visual context**, such as a map or graph, to make data easier for the **human brain to understand and pull insights from**. The main goal of data visualization is to make it easier to **identify patterns, trends and outliers in large data sets**.

[Origin unknown]

What should be achieved?

- ✓ show the data
- ✓ induce viewer to think about substance rather than methodology, graphical design or other aspects
- ✓ encourage eye to compare different pieces of data
- ✓ avoid distorting what the data represents
- ✓ present many numbers in a small space
- ✓ make large data sets coherent
- ✓ reveal data at several levels of detail
- ✓ serve a reasonably clear purpose



Class Activity

Set A		Set B		Set C		Set D	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.11	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

[Anscombe 1973]

Task 1 [15 minutes]

Calculate summary statistics:

- Mean
- Standard Deviation
- Regression equation
- R^2

Task 2 [10 minutes]

Plot the data points for each sets and compare.
Share your observation with the class.

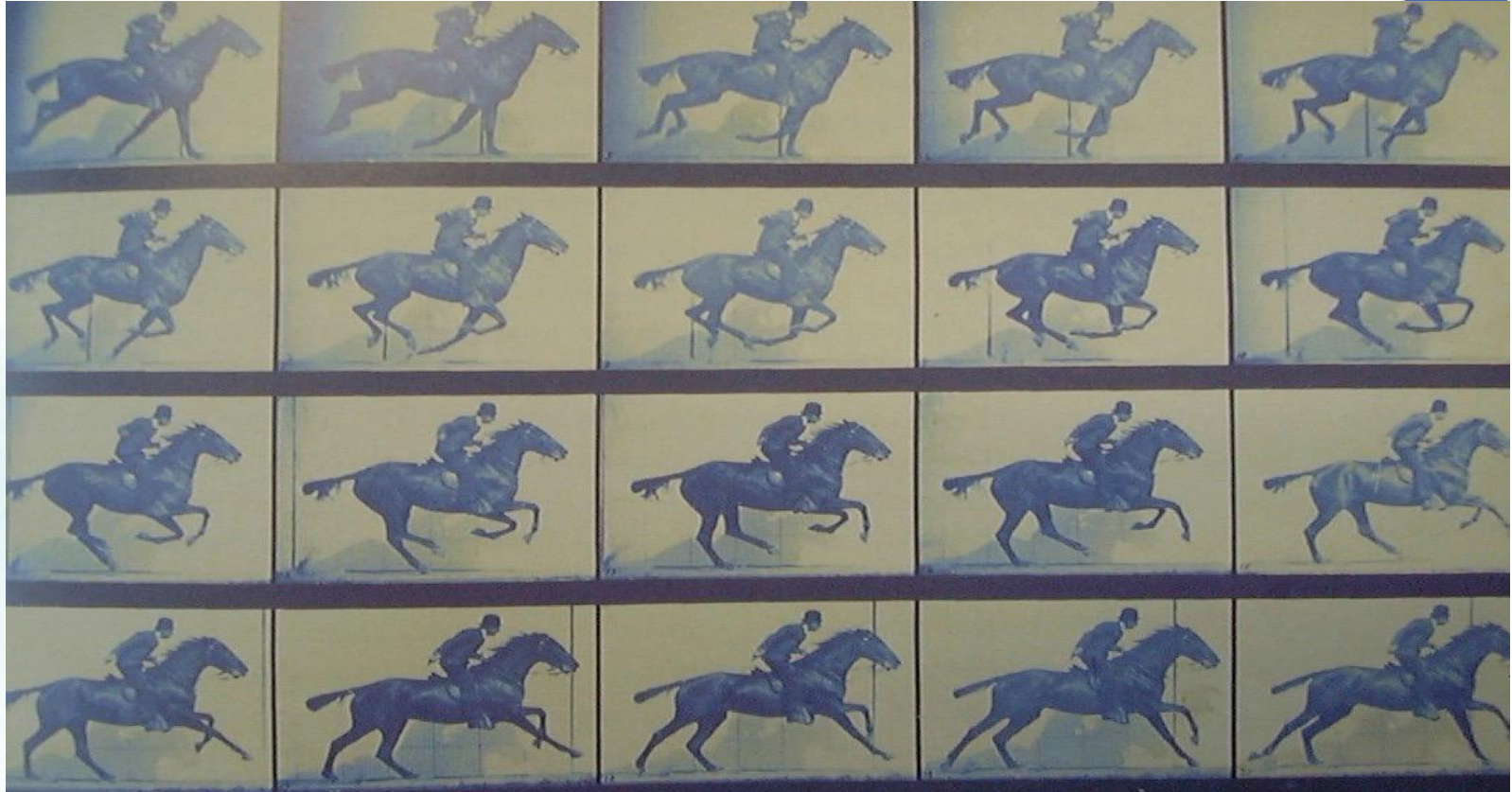
Why Create Visualizations?



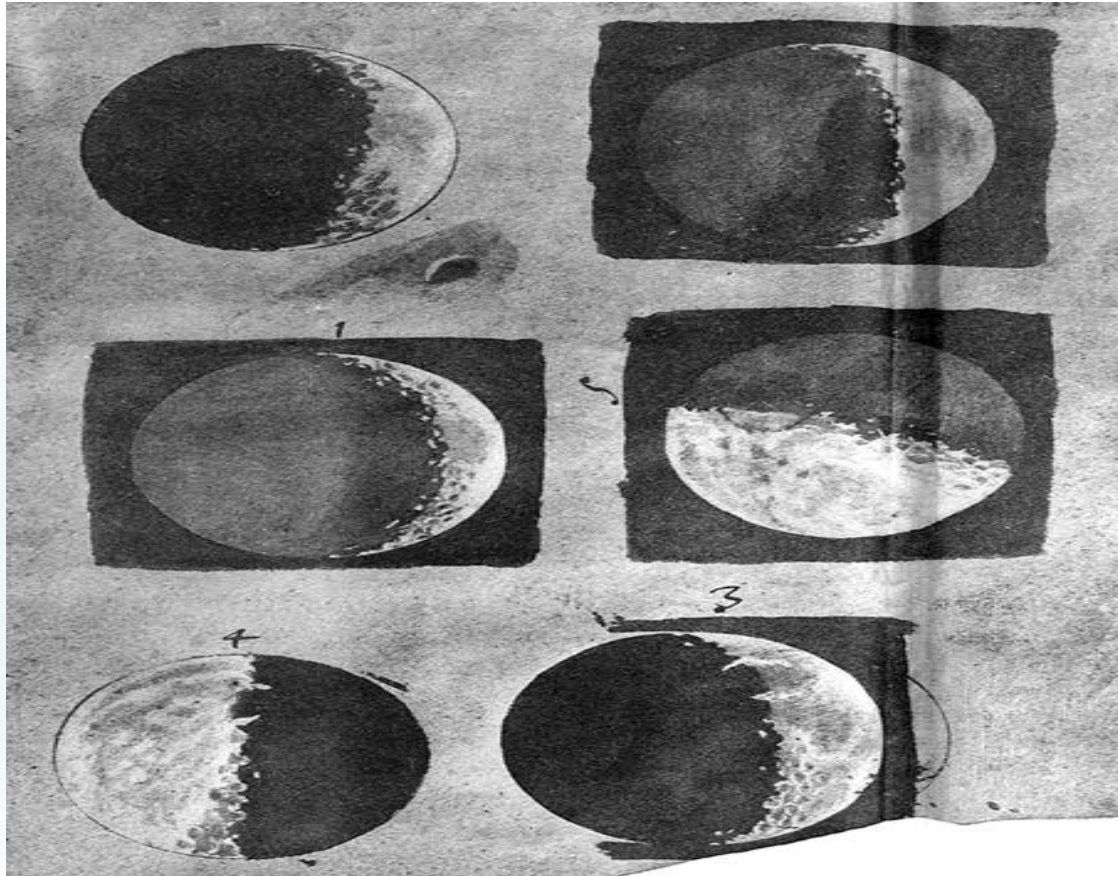
Record information

- Blueprints, photographs, seismographs, record historical data, ...

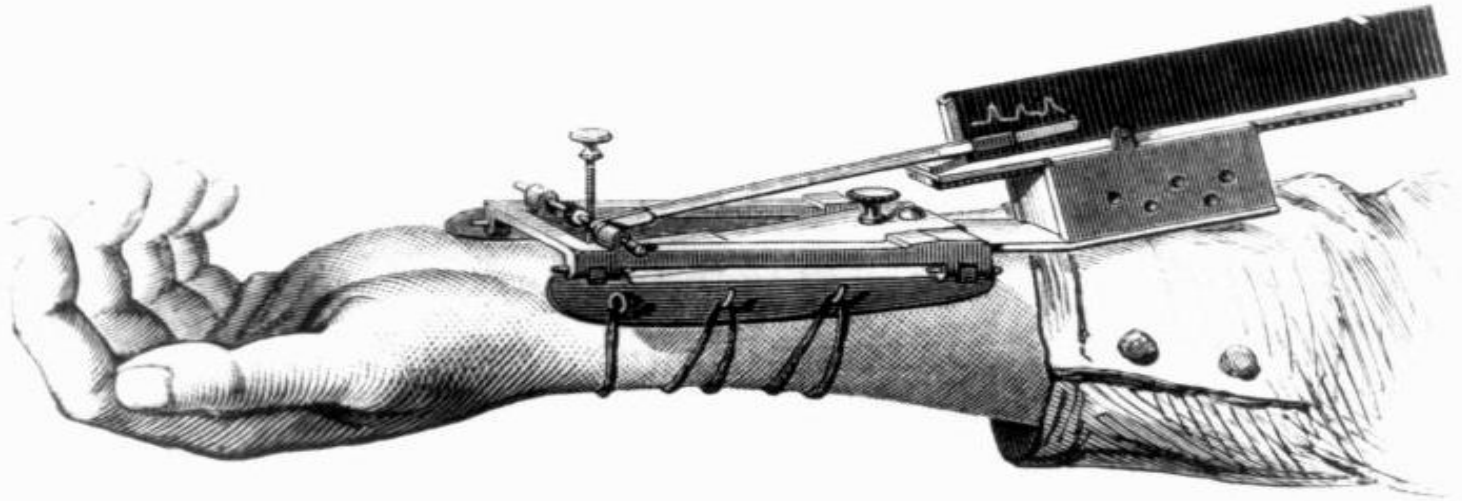
Record Information : Answer Question



Record Information : Drawing: Phases of the moon



Record Information : Recording Instruments



1.
Marey's **sphygmograph** in use.
1860. *La méthode graphique dans
les sciences expérimentales et
principalement en physiologie et en
médecine.*

Why Create Visualizations?

Record information

- Blueprints, photographs, seismographs, record historical data, ...

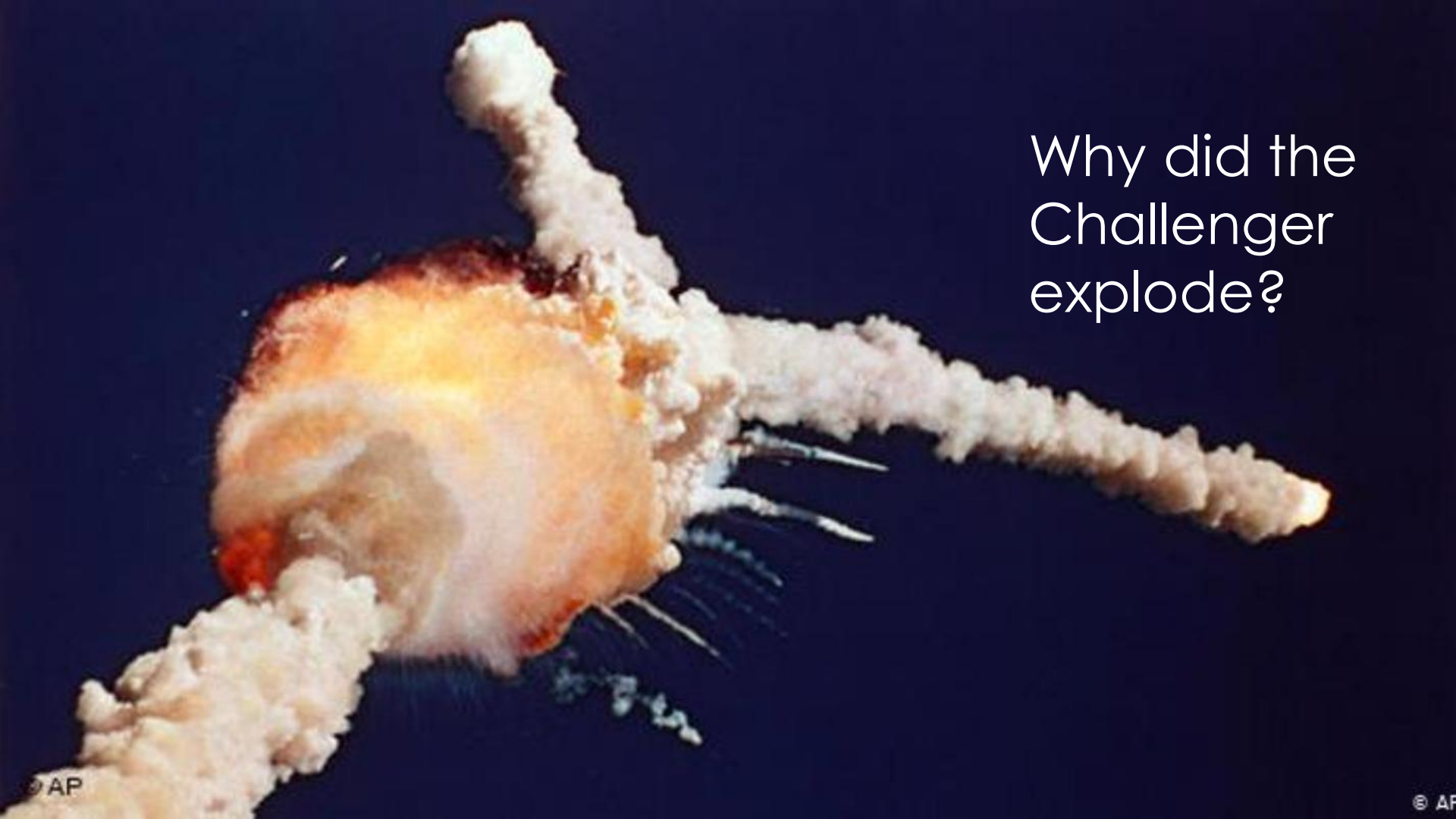
Analyze data to support reasoning (exploratory visualization)

- Develop and assess hypotheses
- Find patterns / Discover errors in data
- Expand memory

Launching of the
Challenger @ Jan
1986



Why did the Challenger explode?



Support Reasoning : Make a Decision: Challenger

HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

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	--
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.

BLOW BY HISTORY

SRM-15 WORST BLOW-BY

- o 2 CASE JOINTS (80°), (110°) ARE
- o MUCH WORSE VISUALLY THAN SRM-22

SRM 12 BLOW-BY

- o 2 CASE JOINTS (30-40°)

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

- o NOZZLE BLOW-BY

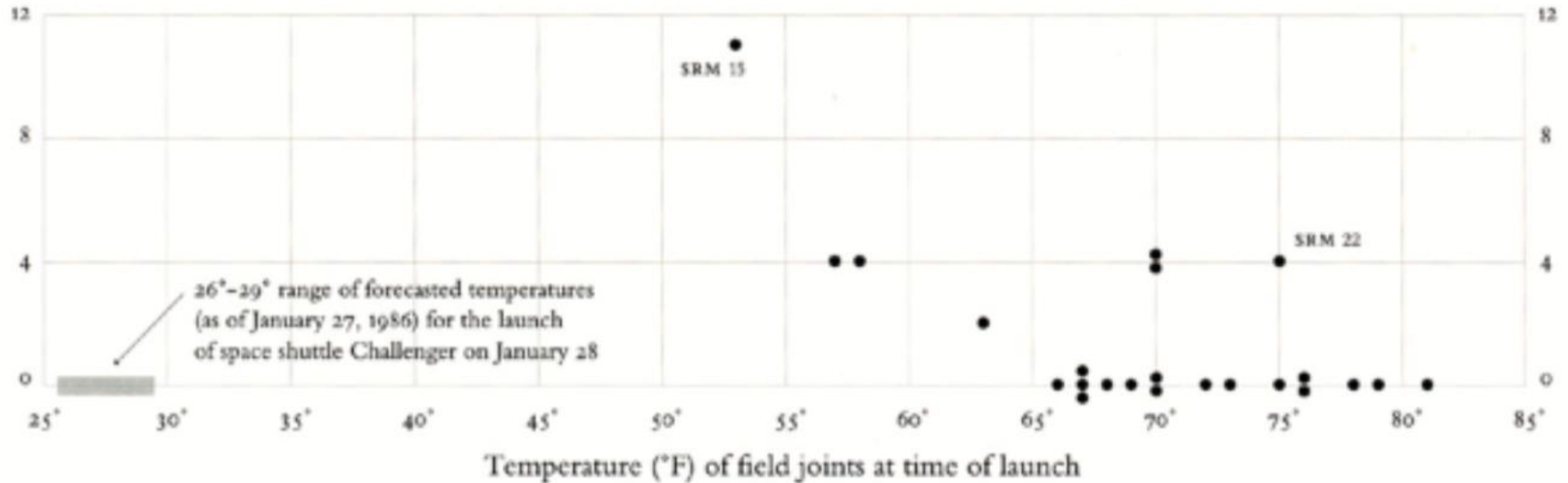
HISTORY OF O-RING TEMPERATURES (DEGREES - F)

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	10 MPH
			27	25 MPH

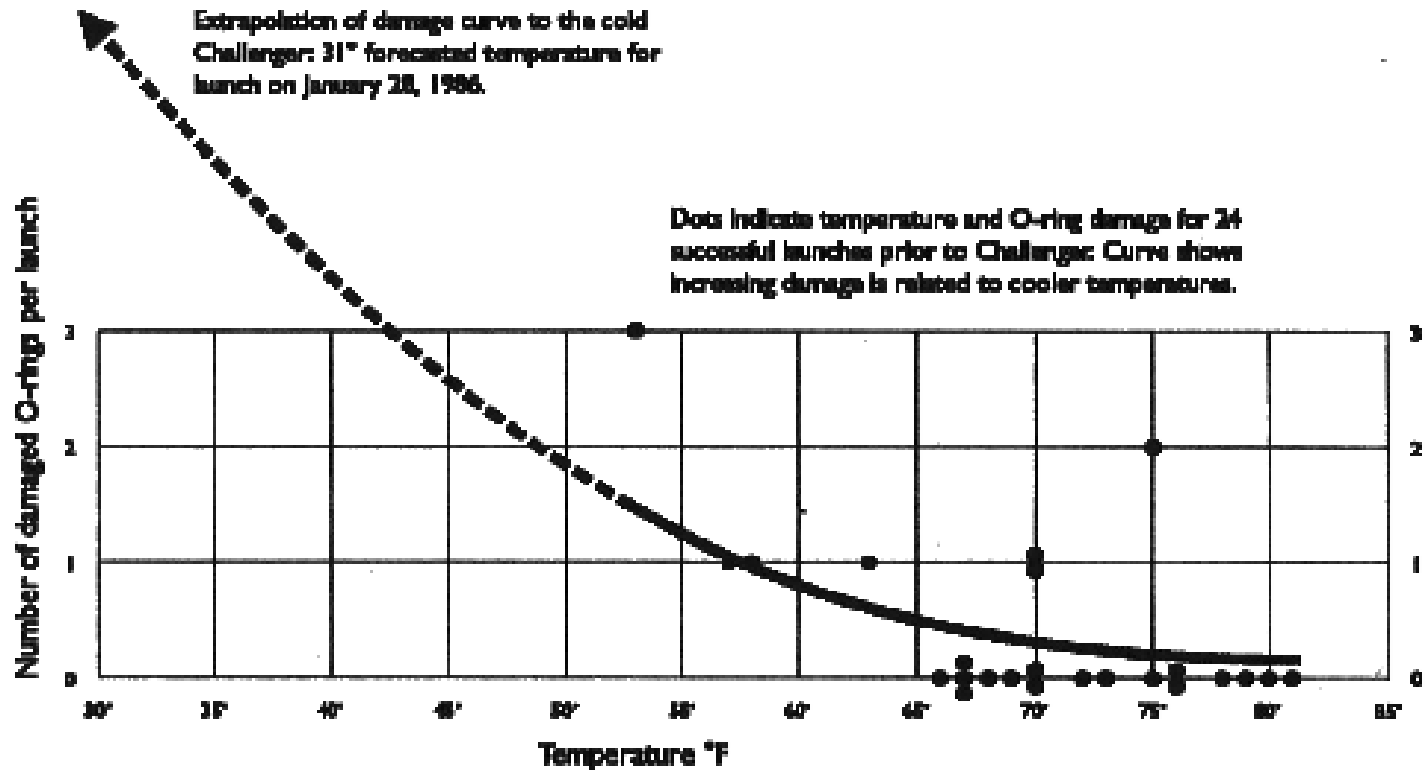


But wait! What is an appropriate “damage index”?
Which temperatures, O-ring or outside air?

O-ring damage index, each launch



Visualizations drawn by Tufte show how low temperatures damage O-rings [Tufte 97]
<https://www.asktog.com/books/challengerExerpt.html>



Tufte's close analysis demonstrates that the engineers had the information they needed — that O-ring failure rates rose as temperature declined—but didn't display it clearly. Seven astronauts' lives could have been saved with a simple graph of previous O-ring damage level against temperature.

Support Reasoning

- Data in Context: Cholera Outbreak In 1854
- In 1854, John Snow plotted the position of each cholera case on a map.



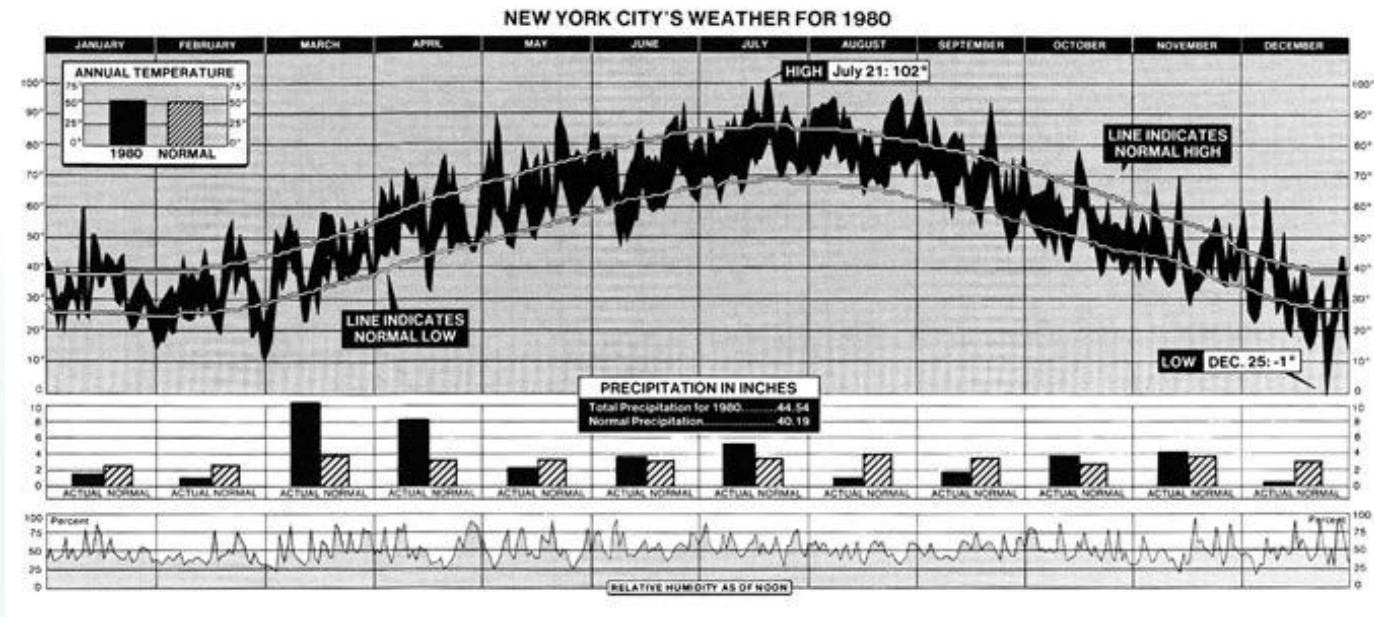
Support Reasoning

- Used map to hypothesize that pump on Broad St. was the cause.



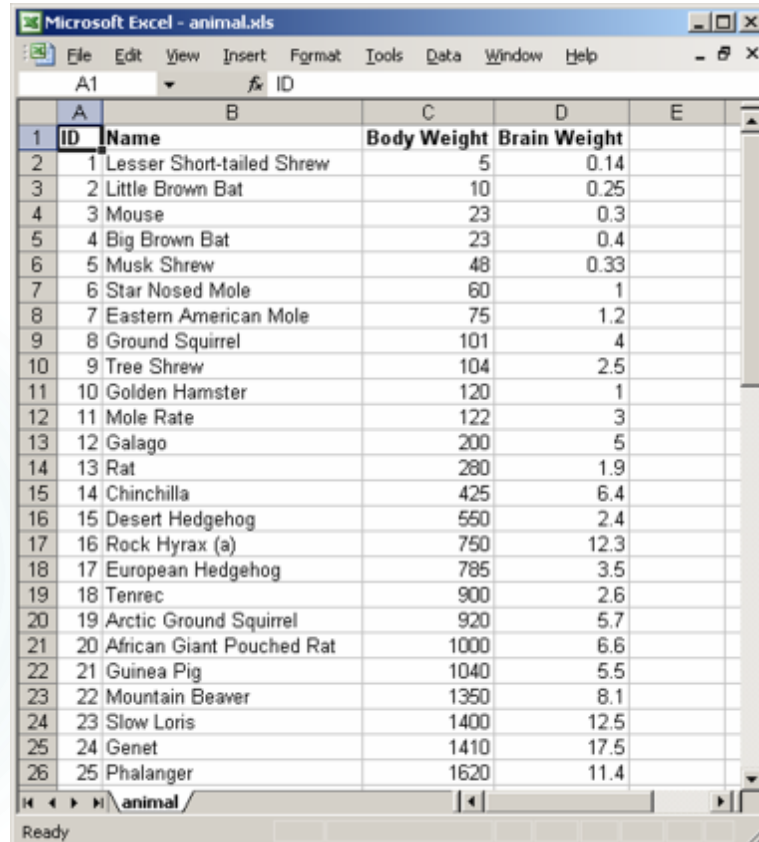
Find Patterns

NYC Weather



[New York Times 1981]

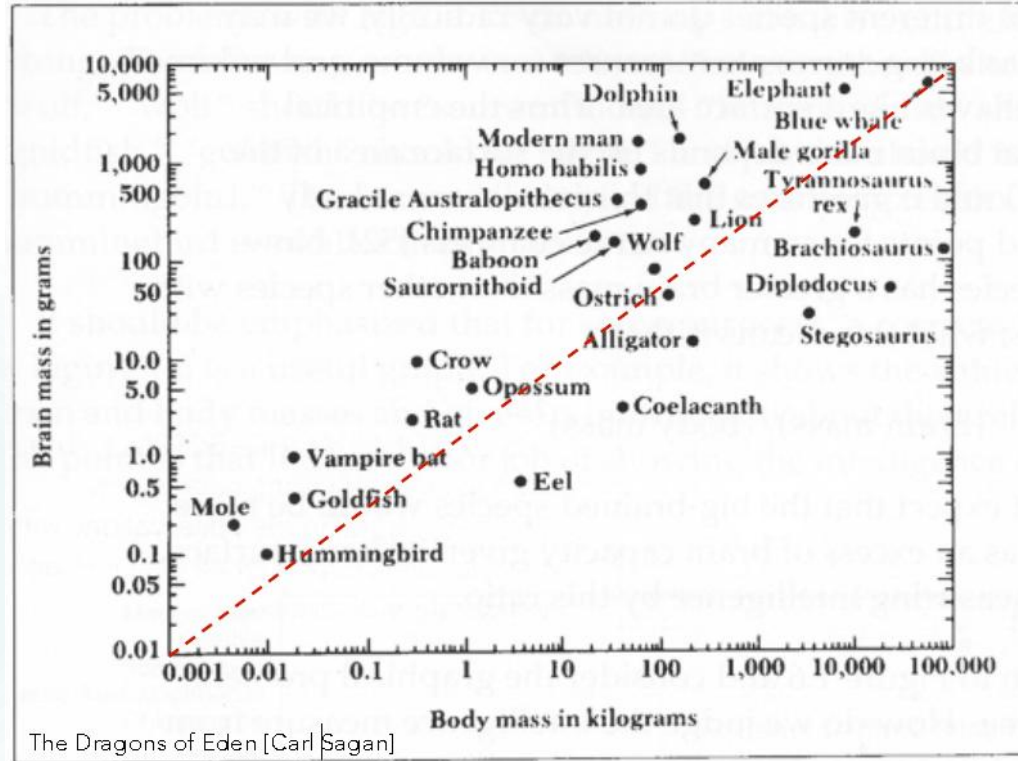
The Most Powerful Brain?



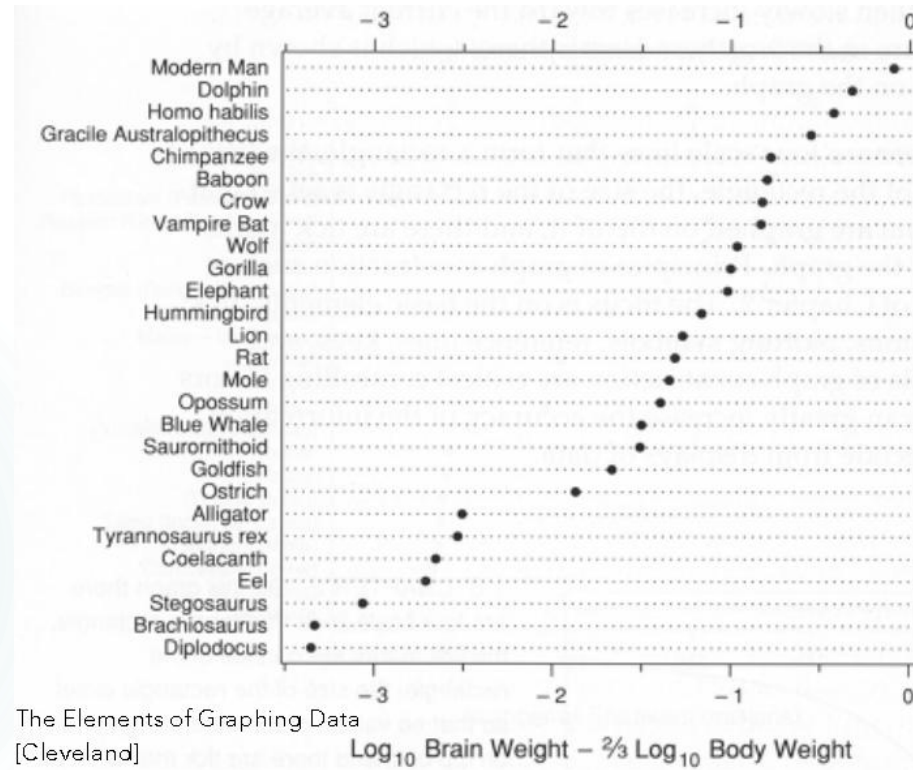
The image shows a screenshot of a Microsoft Excel spreadsheet titled "animal.xls". The spreadsheet contains a table with 26 rows of animal data. The columns are labeled "ID", "Name", "Body Weight", and "Brain Weight". The data is as follows:

ID	Name	Body Weight	Brain Weight
1	Lesser Short-tailed Shrew	5	0.14
2	Little Brown Bat	10	0.25
3	Mouse	23	0.3
4	Big Brown Bat	23	0.4
5	Musk Shrew	48	0.33
6	Star Nosed Mole	60	1
7	Eastern American Mole	75	1.2
8	Ground Squirrel	101	4
9	Tree Shrew	104	2.5
10	Golden Hamster	120	1
11	Mole Rate	122	3
12	Galago	200	5
13	Rat	280	1.9
14	Chinchilla	425	6.4
15	Desert Hedgehog	550	2.4
16	Rock Hyrax (a)	750	12.3
17	European Hedgehog	785	3.5
18	Tenrec	900	2.6
19	Arctic Ground Squirrel	920	5.7
20	African Giant Pouched Rat	1000	6.6
21	Guinea Pig	1040	5.5
22	Mountain Beaver	1350	8.1
23	Slow Loris	1400	12.5
24	Genet	1410	17.5
25	Phalanger	1620	11.4

The Most Powerful Brain?



The Most Powerful Brain?




Expand Memory

Class Exercise

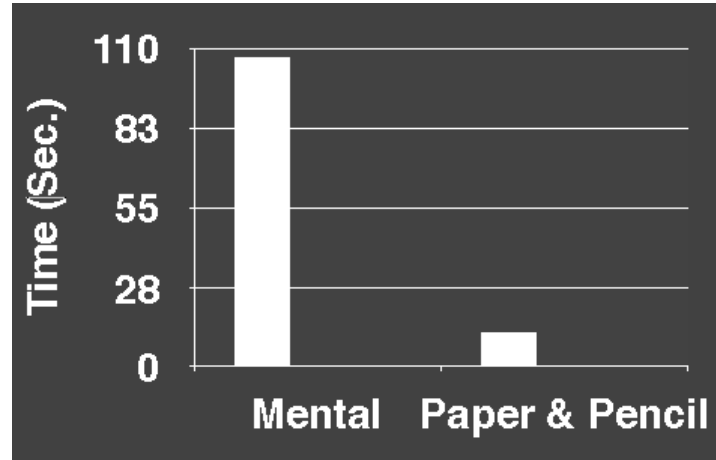
34
x 72

00:00

mins: secs: type:
 

Expand Memory

Class Exercise

$$\begin{array}{r} 34 \\ \times 72 \\ \hline 68 \\ 2380 \\ \hline 2448 \end{array}$$


Why Create Visualizations?

Record information

- Blueprints, photographs, seismographs, record historical data, ...

Analyze data to support reasoning (exploratory visualization)

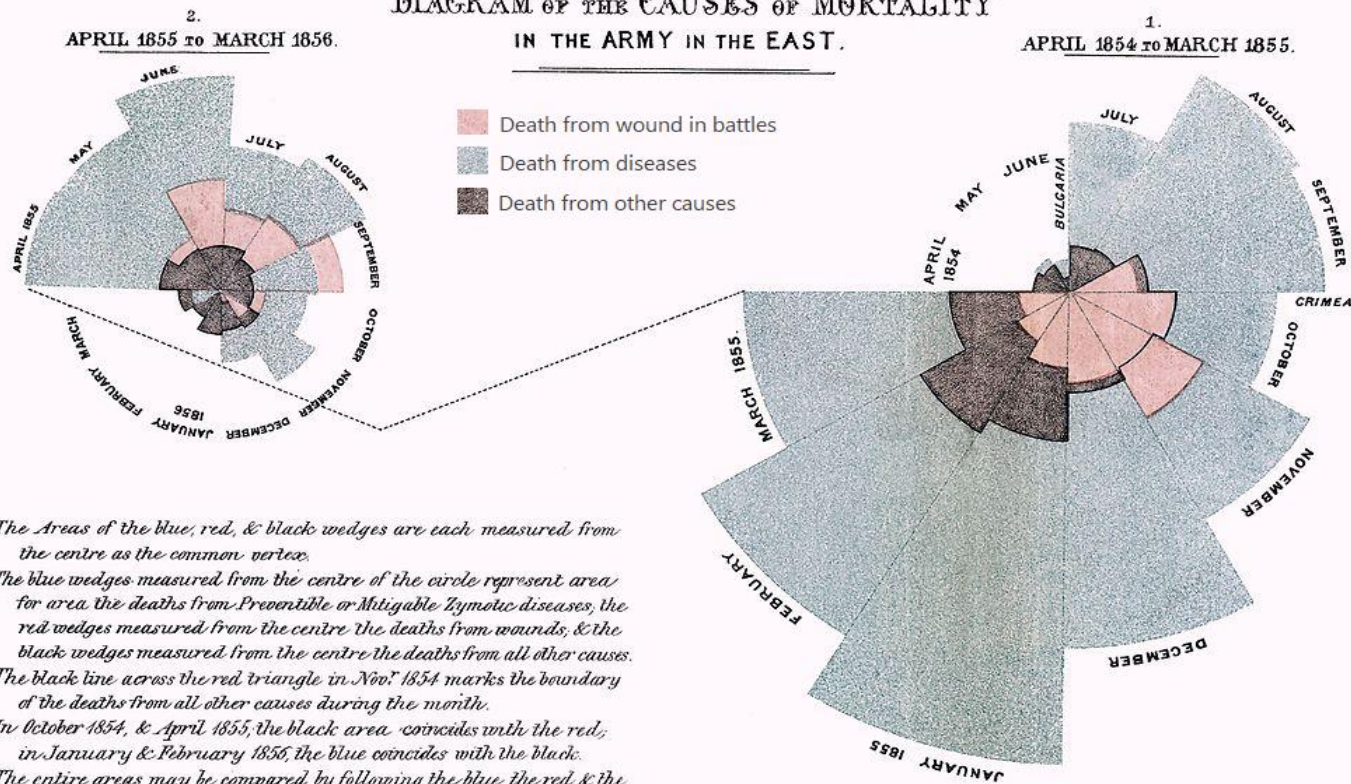
- Develop and assess hypotheses
- Find patterns / Discover errors in data
- Expand memory

Communicate information to others (explanatory visualization)

- Share and persuade
- Collaborate and revise
- Emphasize important aspects of data

Share and persuade

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



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the 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.

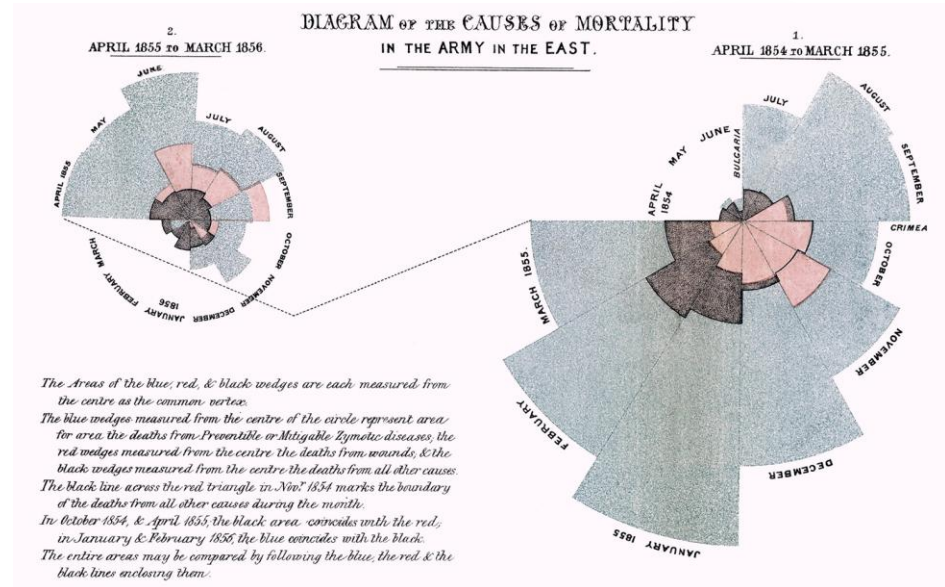
In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.

The entire areas may be compared by following the blue, the red & the black lines enclosing them.

Share and persuade

Insights:

- most of the fatalities during the war were from sickness caused by deficient sanitary measures
- improvements in hygiene dramatically reduced the death rate



Why Create Visualizations?

Record information

- Blueprints, photographs, seismographs, record historical data, ...

Analyze data to support reasoning (exploratory visualization)

- Develop and assess hypotheses
- Find patterns / Discover errors in data
- Expand memory

Communicate information to others (explanatory visualization)

- Share and persuade
- Collaborate and revise
- Emphasize important aspects of data

Visualization Research

Challenge

More and more unseen data

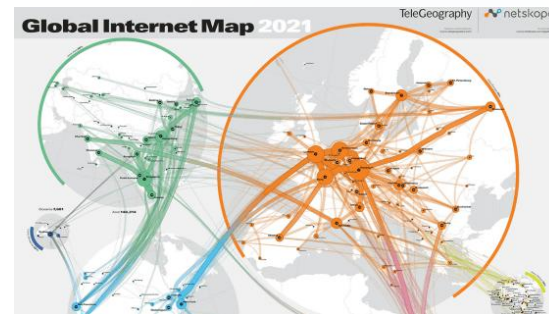
- Faster creation and collection
- Faster dissemination



Photo sharing/annotation



Wikipedia



Map of the Internet

Top Visualization Research Labs:

- UW Interactive Research Lab: <https://idl.cs.washington.edu/>

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.

Hal Varian
Google's Chief Economist
The McKinsey Quarterly
Jan 2009



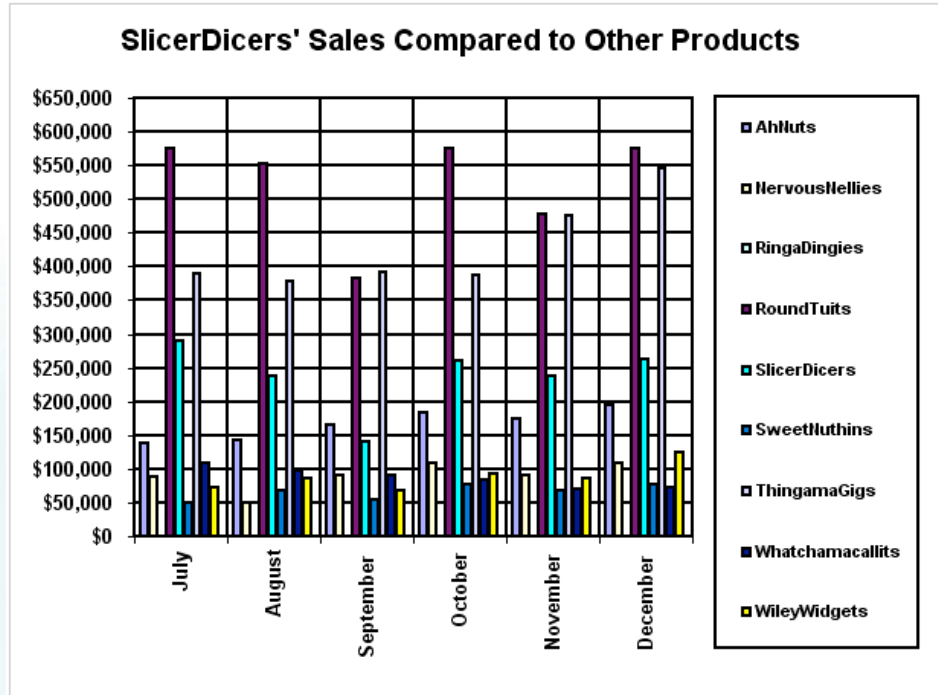


Course Topics

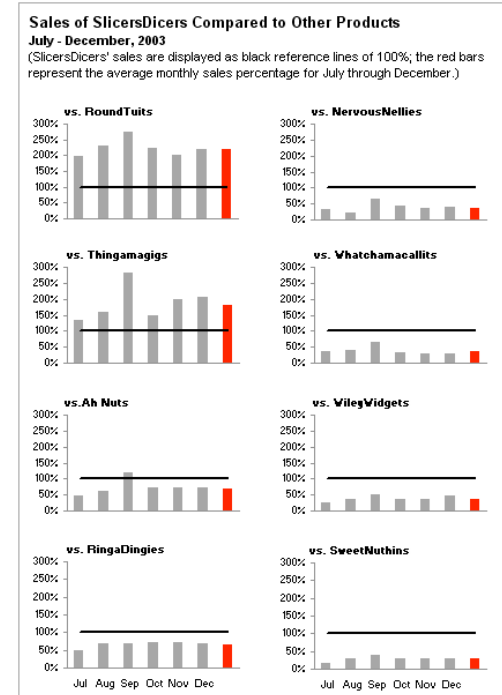
Data and Image Models

		LES VARIABLES DE L'IMAGE						12	14		
		POINTS		LIGNES		ZONES					
XY 2 DIMENSIONS DU PLAN		x	x	x	/	~	/	14 15 9 14 1 1 10 21 2 2 14 15 1	2 18 2 1 21 15 2 2 9	OO	≠
	Z									OO	≠
TAILLE		■	■	■	/	~	/	■	■	OO	≠
VALEUR		■	■	■	/	~	/	■	■	○	≠
		LES VARIABLES DE SÉPARATION DES IMAGES						13			
GRAIN		■	■	■	/	~	/	■	■	○	≠
COULEUR		■	■	■	/	~	/	■	■	≡	≠
ORIENTATION		■	■	■	/	~	/	■	■	≡	≠
FORME		■	▲	●	/	~	/	■	■	≡	≠

Visualization Design

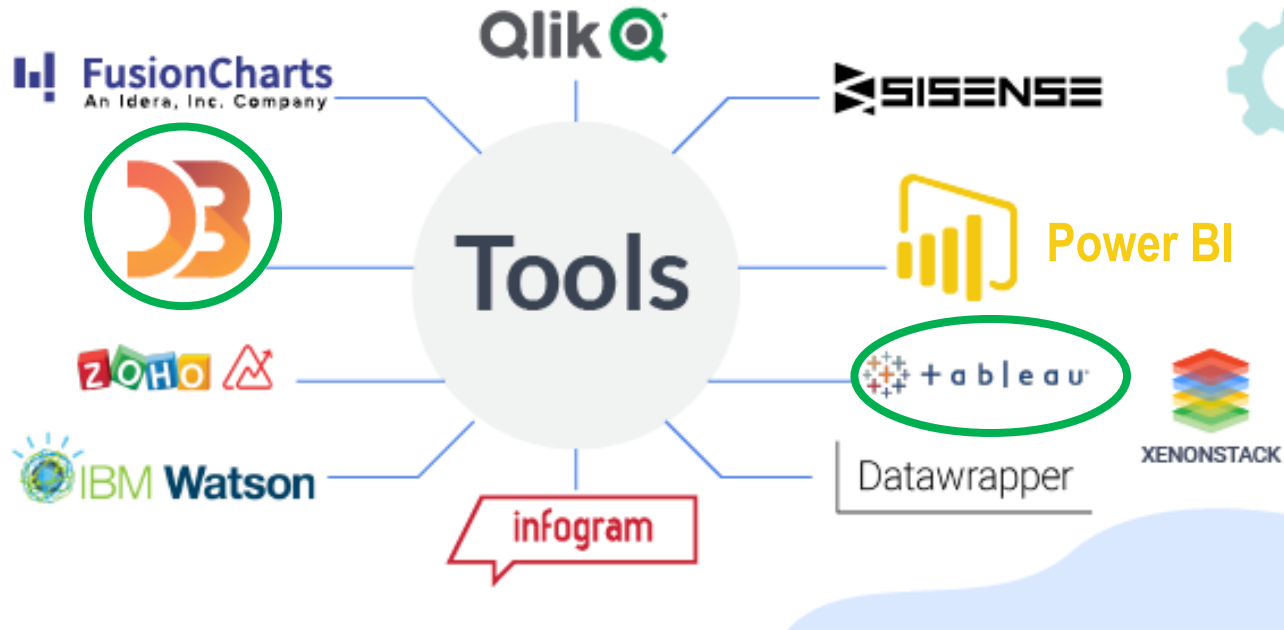


Problematic design



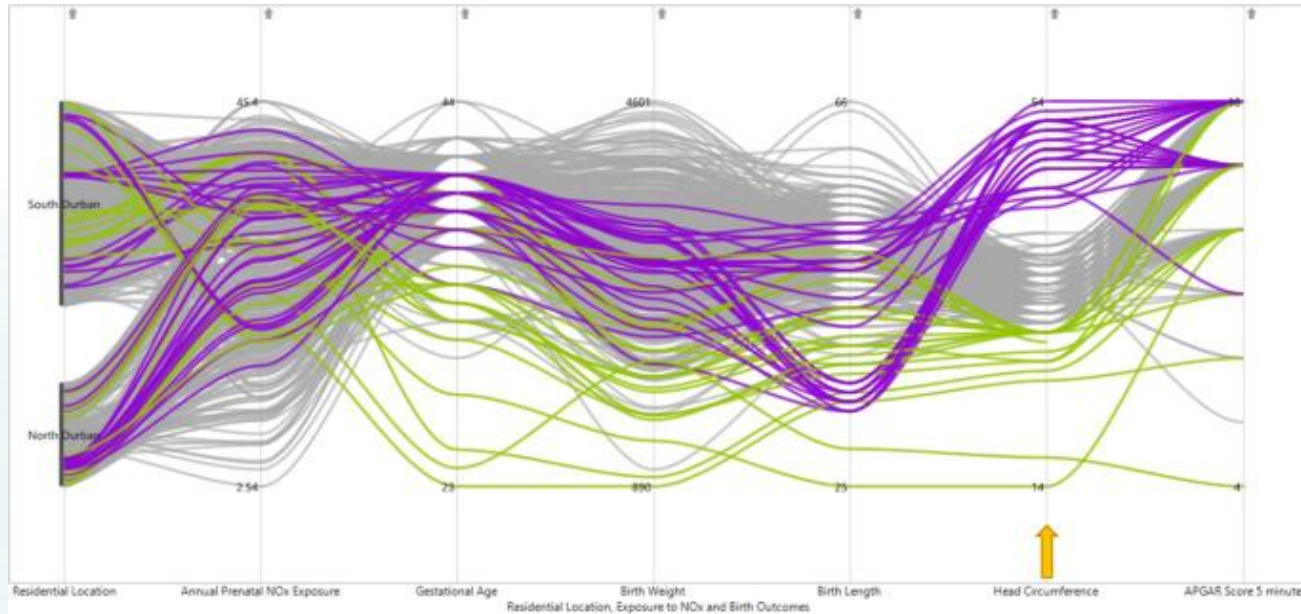
Redesign

Visualization Tools



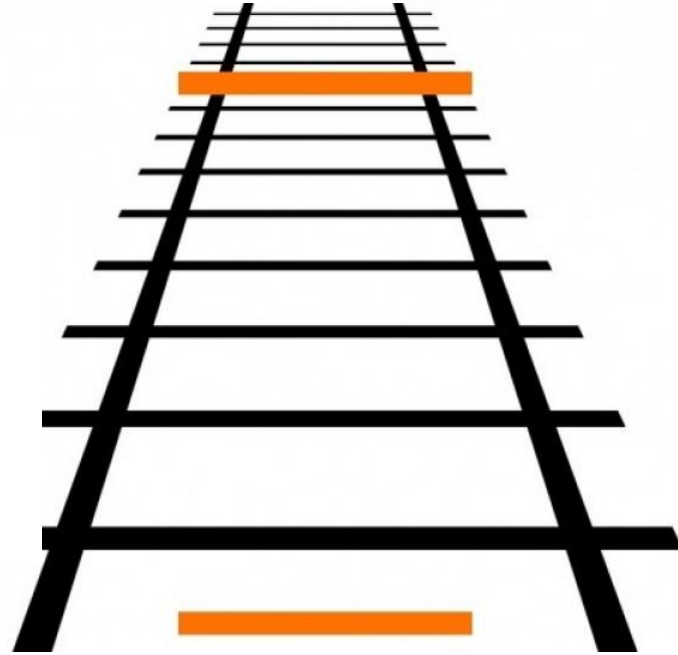
<https://www.xenonstack.com/blog/top-data-visualization-tools>

Multidimensional Data Visualization



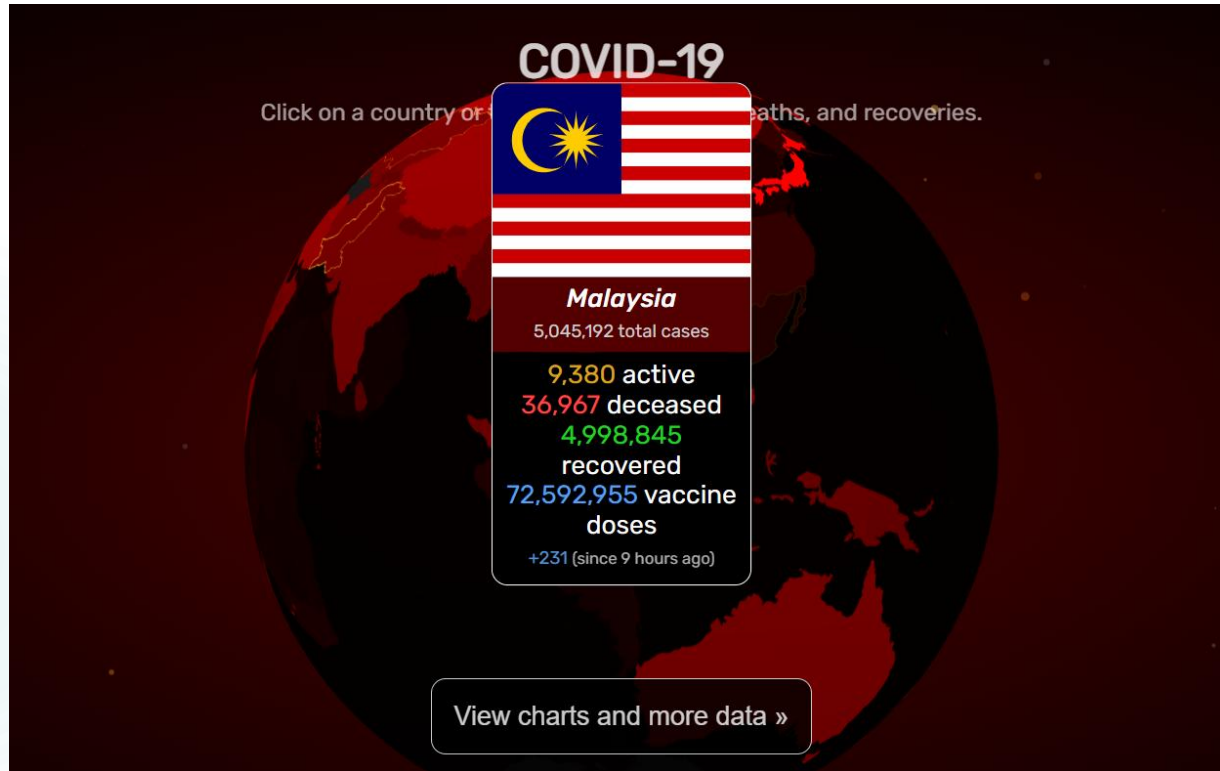
Exploratory Data Analysis of Adverse Birth outcomes and exposure to oxides of nitrogen Using interactive parallel coordinates plot technique. *Scientific reports*, 2020

Graphical Perception



[Don't Believe Your Eyes: How Visual Illusions Work](#)

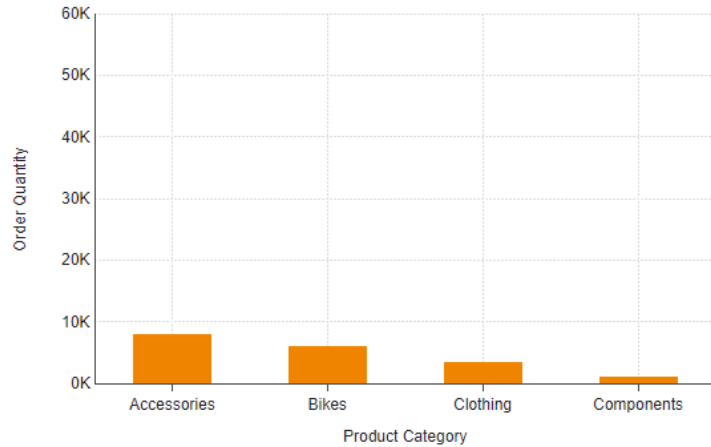
Interaction



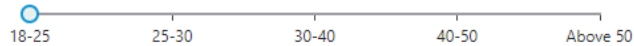
<https://www.covidvisualizer.com/>

Animation

Orders by Products

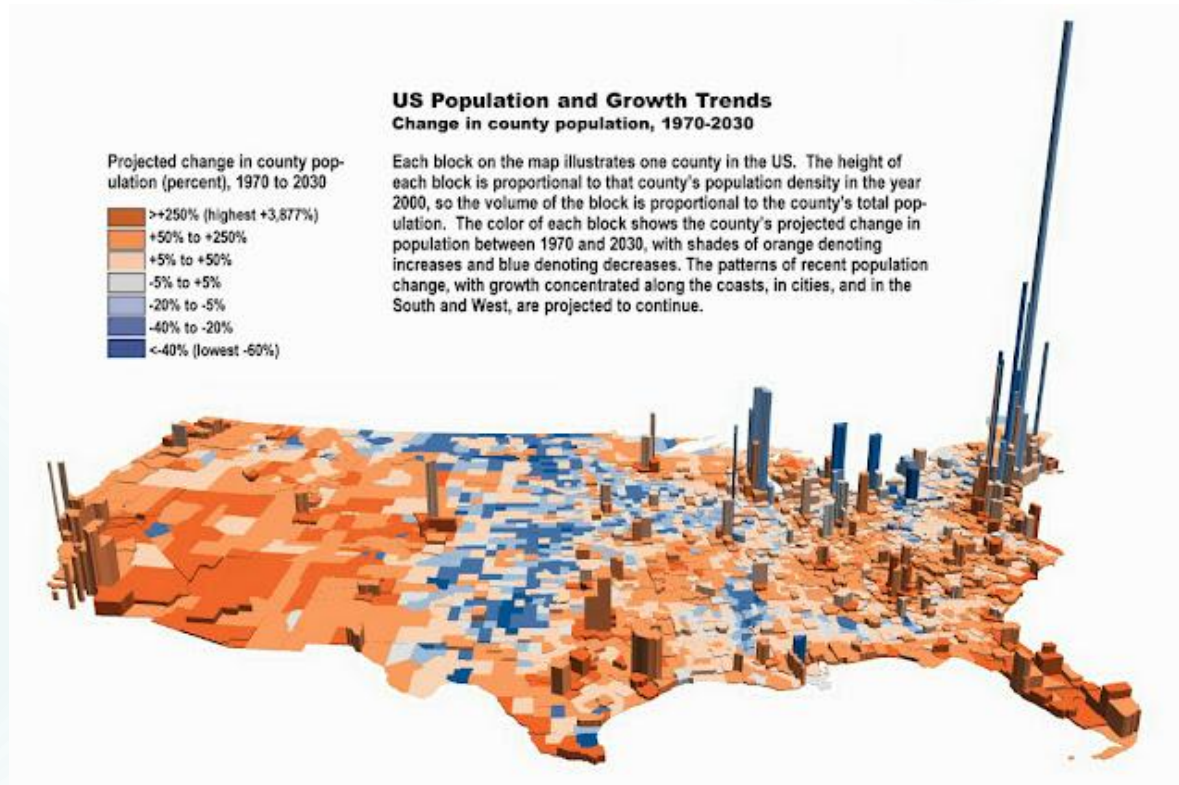


Select Age Group



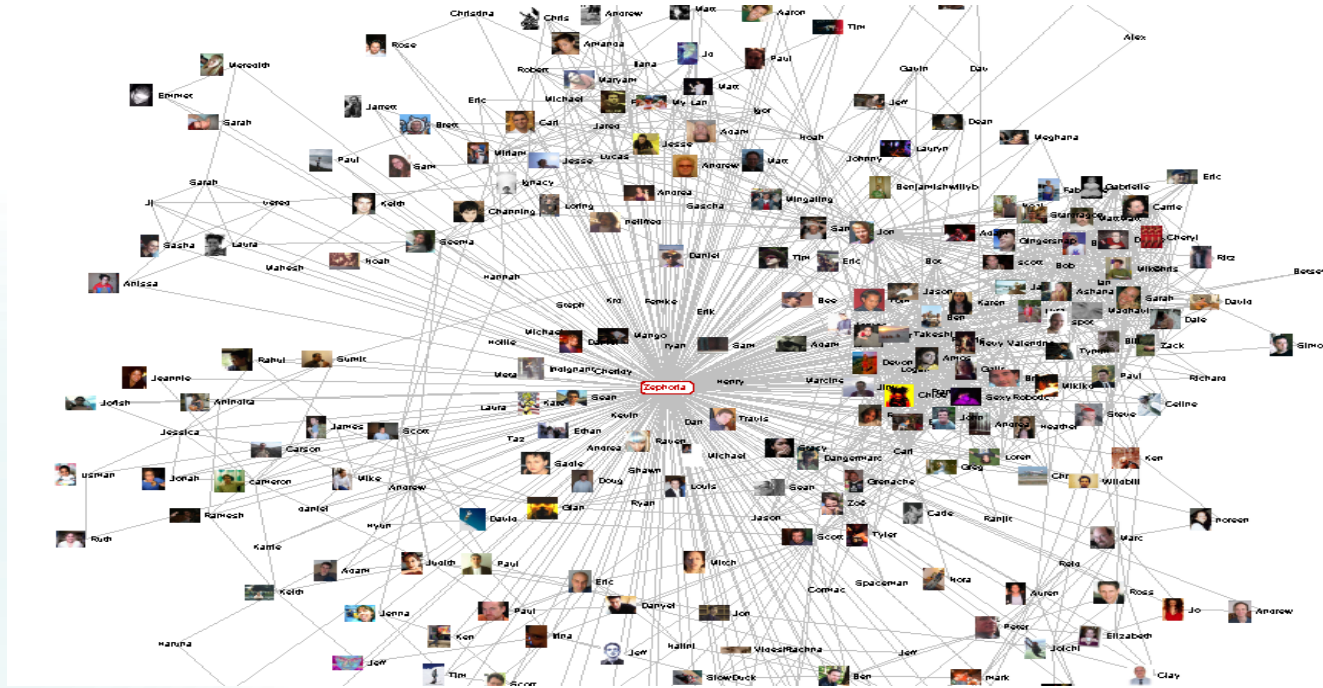
<https://www.dundas.com/resources/blogs/benefits-of-bi/enhance-your-data-storytelling-with-animated-charts>

Geospatial Data Visualization



<http://www.flintexpats.com/2010/05/interview-with-frank-popper-about.html>

Graph Visualization



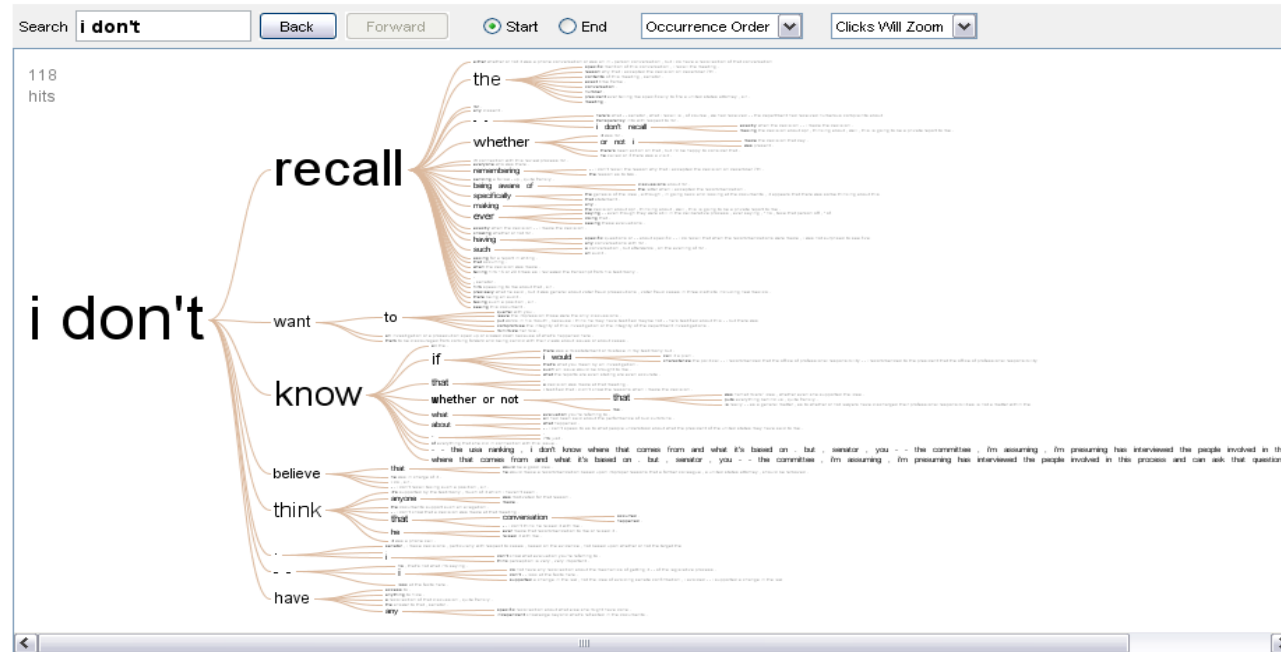
Record of Human Activity on Facebook [Zheporia Digital Marketing]

Text Visualization

Visualizations : Word tree / Alberto Gonzales

Creator: Martin Wattenberg

Tags:



References

- M Tufte, E. (2001). The Visual Display of Quantitative Information (2nd Edition). Graphics Press. <https://www.edwardtufte.com/>
- Data Visualization Course (2022), University of Washington <https://courses.cs.washington.edu/courses/cse442>