

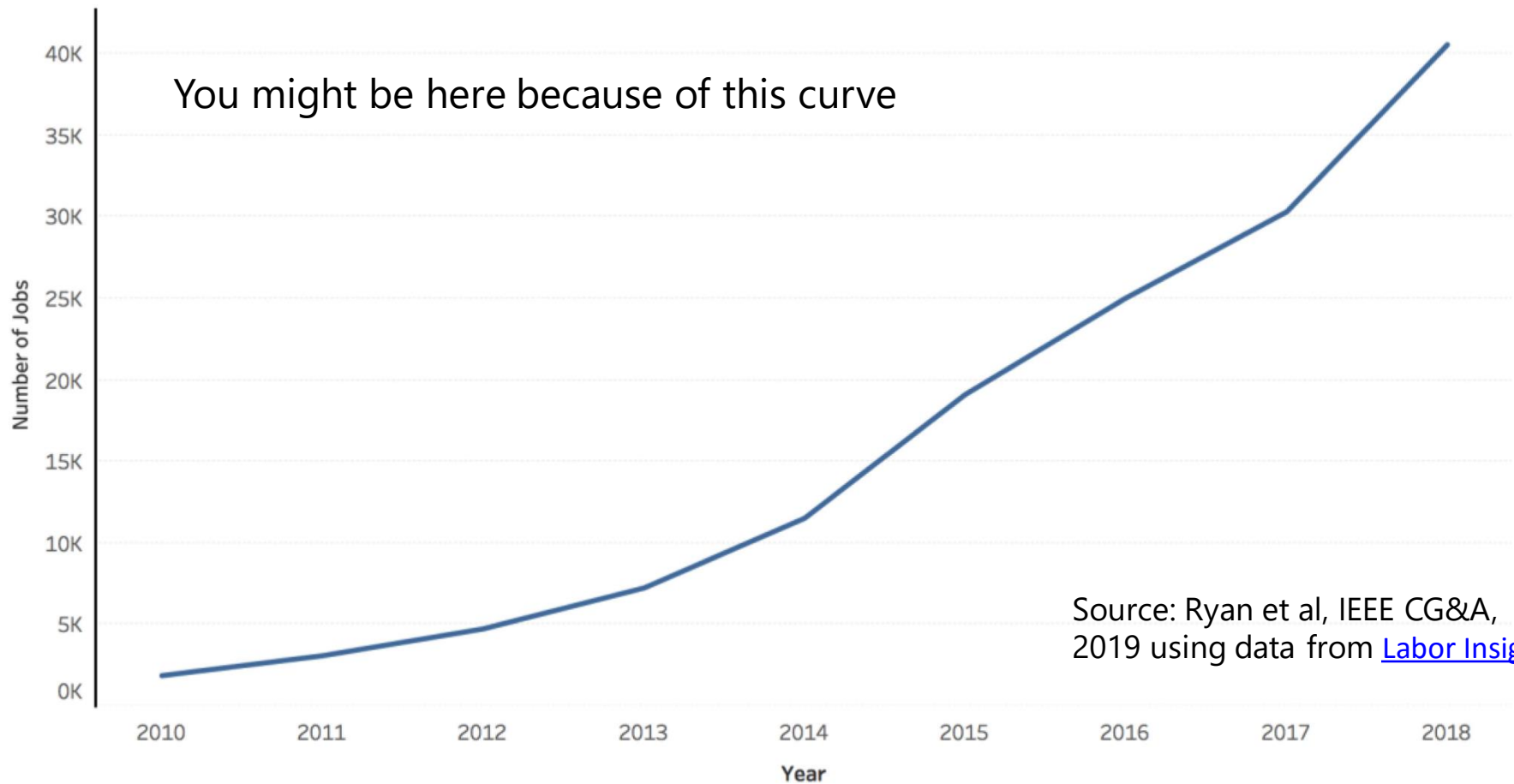
CSE 332
INTRODUCTION TO VISUALIZATION

INTRODUCTION

KLAUS MUELLER

COMPUTER SCIENCE DEPARTMENT
STONY BROOK UNIVERSITY

WHY ARE YOU HERE?

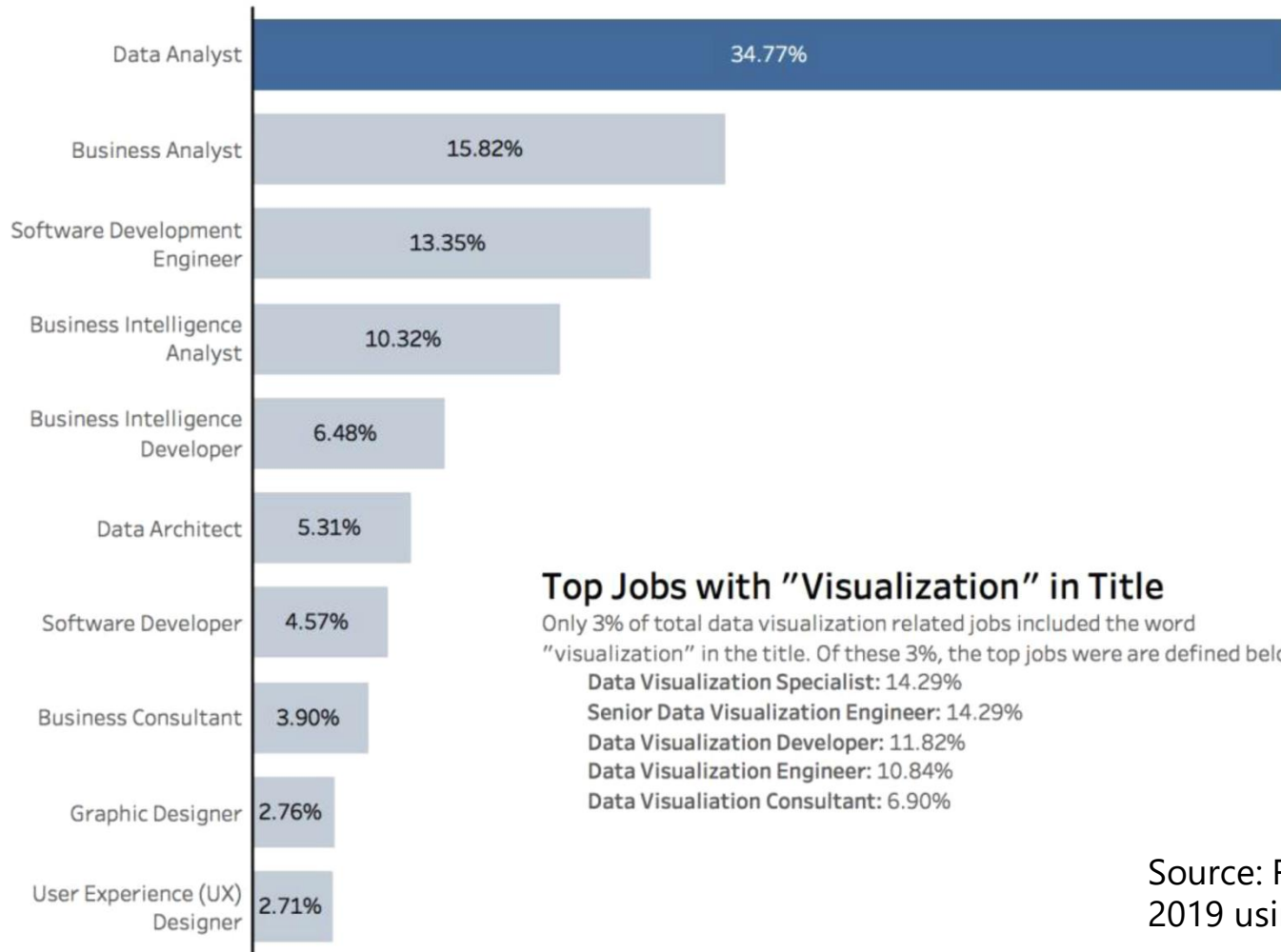


The growth of jobs mentioning “data visualization” as a skill from 2010 through 2017 has steadily increased from only 1,888 jobs in 2010 to 30,327 jobs in 2017 (16× growth)

“VISUALIZATION” SKILL...

Top Job Titles Listing “Data Visualization” as a Skill

... is needed everywhere



Top Jobs with “Visualization” in Title

Only 3% of total data visualization related jobs included the word “visualization” in the title. Of these 3%, the top jobs were defined below:

- Data Visualization Specialist: 14.29%
- Senior Data Visualization Engineer: 14.29%
- Data Visualization Developer: 11.82%
- Data Visualization Engineer: 10.84%
- Data Visualization Consultant: 6.90%

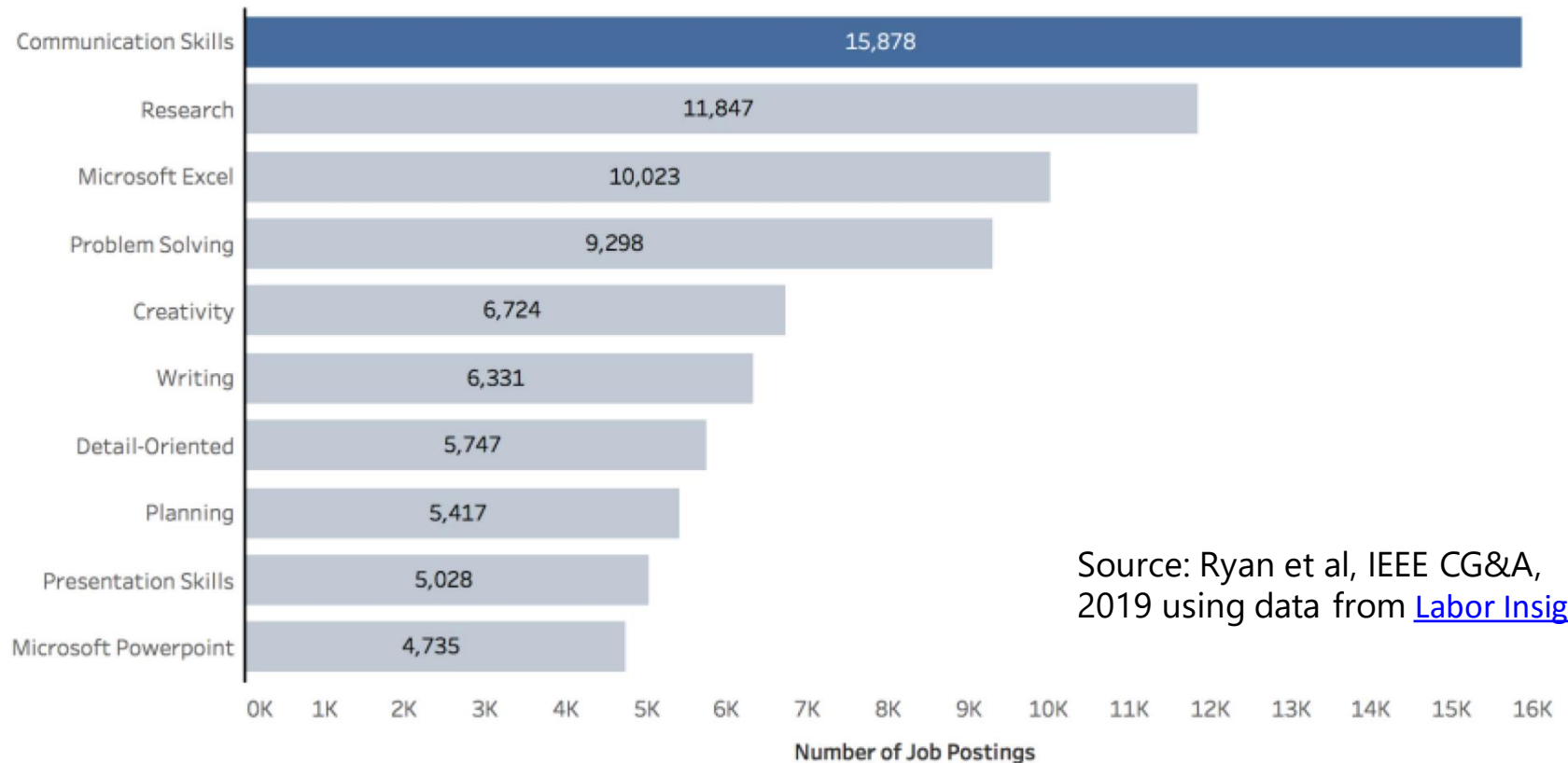
Source: Ryan et al, IEEE CG&A, 2019 using data from [Labor Insight](#)

WHAT OTHER SKILLS?

Data Visualization Top Baseline (Soft) Skills

Of ~31k visualization related jobs posted between March 2017 and February 2018, ~16k listed the broad skill of **communication** as the top “soft” skill. Many of the other top soft skills, including problem solving, detail-oriented, and planned all fall into a larger project management skillset.

Source: Labor Insight (Burning Glass Technologies)



Source: Ryan et al, IEEE CG&A, 2019 using data from [Labor Insight](#)

Baseline, or “soft” skills listed for these 30k “Data Visualization” jobs.

SKILLS, READING BETWEEN THE LINES

Communication, when mentioned in conjunction with data visualization really means:

- communication of information derived from data
- *visual* story telling with data
- half of the data analytics projects fail due to poor communication (according to L. Kart, N. Neudecker, F. Buytendijk, Gartner Report GG0255160, 2013)

Apart from the specialized skills, these general skills (or proficiencies) are also often listed:

- SQL
- Tableau (41%),
- Excel (34%), PowerPoint (16%)
- Python (30%), SAS (22%), R (16%), Plotly (?%)
- JavaScript & JavaScript-based data-driven documents D3.js (13%)

Source: Ryan et al, IEEE CG&A, 2019 using data from [Labor Insight](#)

SKILLS, READING BETWEEN THE LINES

Communication, when mentioned in conjunction with data visualization really means:

- communication of information derived from data
- *visual story telling with data*
- half of the data analytics projects fail due to poor communication (according to L. Kart, N. Neudecker, F. Buytendijk, Gartner Report GG0255160, 2013)

Apart from the specialized skills, these general skills (or proficiencies) are also often listed:

- SQL
- Tableau (41%),
- Excel (34%), PowerPoint (16%)
- Python (30%), SAS (22%), R (16%), Plotly (?%)
- JavaScript & JavaScript-based data-driven documents D3.js (13%)

Source: Ryan et al, IEEE CG&A, 2019 using data from [Labor Insight](#)

VISUALIZATION IS NOT NEW

RICH HISTORY

Let's go back some 160 years to 1854, London, England



NEWSFLASH, 1854

The most terrible outbreak of cholera which ever occurred in this kingdom, is probably that which is taking place in Broad Street, Golden Square, and adjoining streets.

Within two hundred and fifty yards of the spot where Cambridge Street joins Broad Street, there are upwards of five hundred fatal attacks of cholera in ten days.

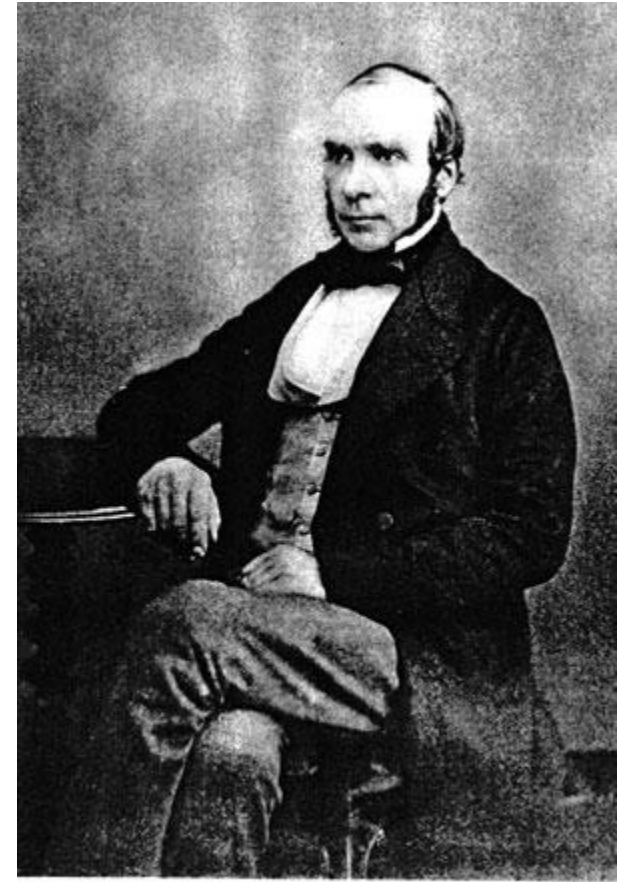
The mortality in this limited area probably equals any that was ever caused in this country, even by the plague; and it is much more sudden, as the greater number of cases terminated in a few hours.

WHAT CAN WE DO?

WHAT IS THE CAUSE?

HOW CAN WE ELIMINATE IT?

TIME FOR “IMAGINATION”



John Snow

PROVED THE HYPOTHESIS

Hypothesis: cholera spreads through water

- and not via some other fantastic causes
- one said it rose out of the burying grounds of plague victims from two centuries earlier
- the bacteria was discovered later, in 1886

A real-life experiment (often the case with observational data)

- established the mode of cholera transmission
- and consequently the method of prevention: keep drinking water, food, and hands clear of infected sewage

Visualization provided

- inspiration
- convincing arguments to justify actions
- led to Dr. John Snow's historic immortality
- a bar near the old Broad Street pump bears his name (safe drinking)

WHAT IS NEEDED FOR VISUALIZATION?

WHAT IS NEEDED FOR VISUALIZATION – SOME APPROPRIATE ANSWERS

Data (wide variety)

Algorithms

- data mining
- data analytics

Computer

- run those algorithms
- data storage

Humans

- with a purpose/need to understand their data
- endowed with cognitive faculties, creative thought, intuition
- domain expertise

Understanding of humans

- perception, cognition, HCI issues
- we can gain it through experimentation with humans

WHAT IS NEEDED FOR VISUALIZATION – SOME APPROPRIATE ANSWERS

Data (wide variety)

Algorithms

- data mining
- data analytics

Computer

- run those algorithms
- data storage

Humans

- with a purpose/need to understand their data
- endowed with cognitive faculties, creative thought, intuition
- domain expertise

Understanding of humans

- perception, cognition, HCI issues
- we can gain it through experimentation with humans

= Visual Analytics

DR. JOHN SNOW: A VISUAL ANALYTICS PIONEER

Dr. John Snow's London Cholera Map of 1854

- data collection
- data assimilation
- statistical testing
- visualization
- computational analysis (brain)
- domain knowledge



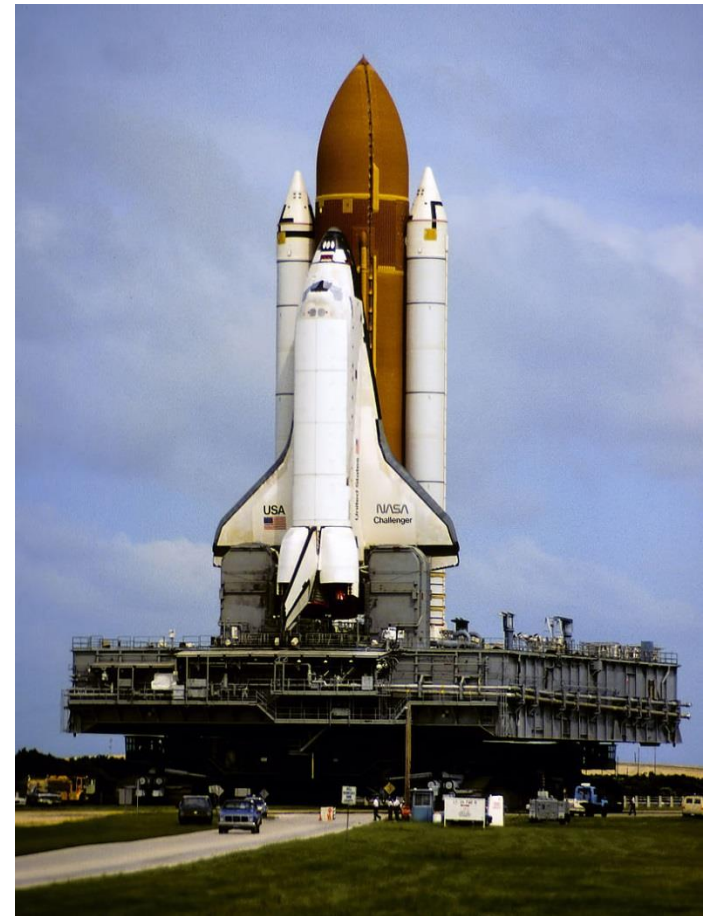
Very early example of
visual analytics

MORE RECENT HISTORY

Let's go back some 40 years to 1986, JFK Space Center, FL



The crew of Space Shuttle mission STS-51-L 11/15/85. Back row, left to right: Ellison S. Onizuka, Sharon Christa McAuliffe, Greg Jarvis, Judy Resnik. Front row, left to right: Michael J. Smith, Dick Scobee, Ron McNair.



73 SECONDS AFTER LIFT-OFF

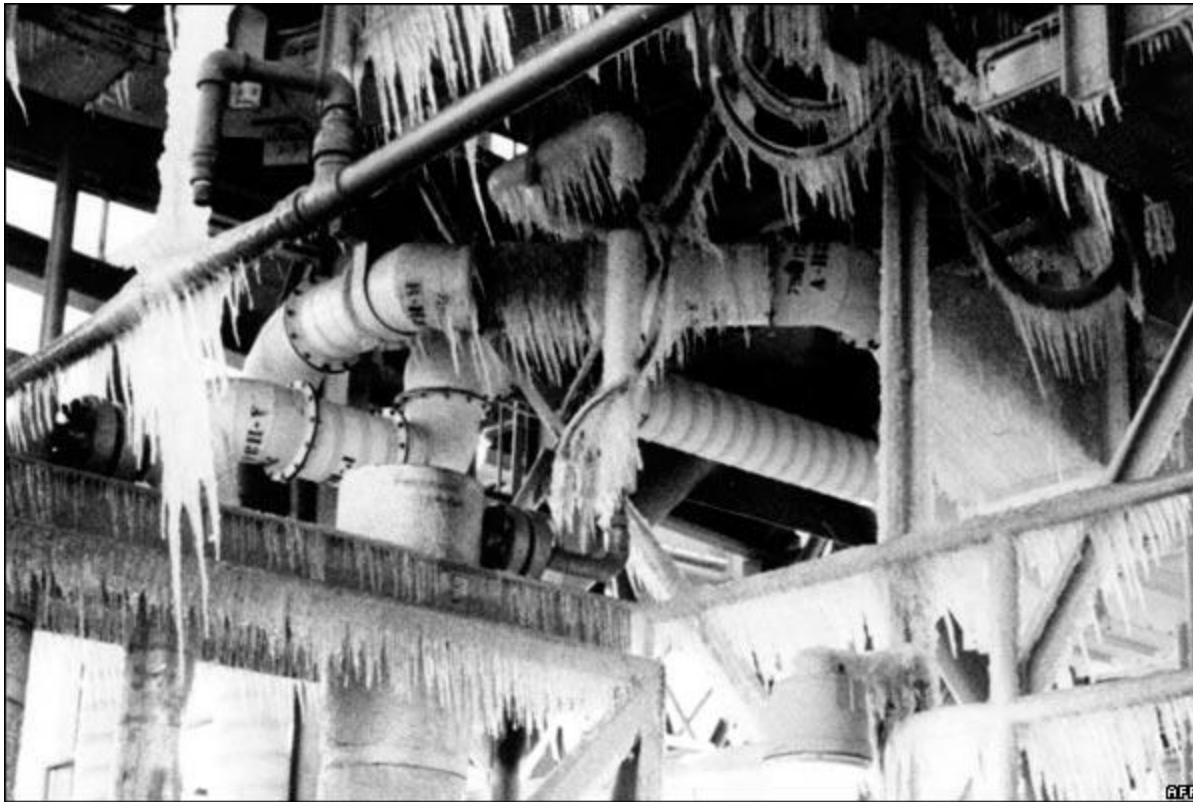


WHAT HAPPENED?

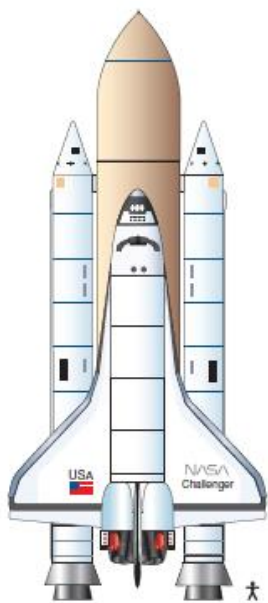
WHAT WAS THE CAUSE?

THE DAY OF THE LAUNCH

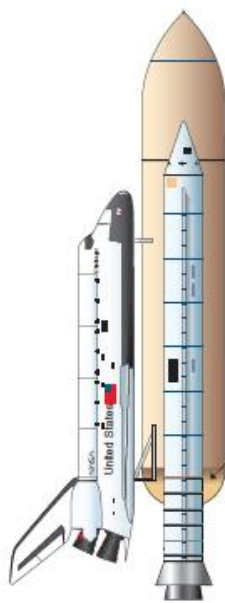
36 degrees F on Launch Pad 39



SPACE SHUTTLE 101

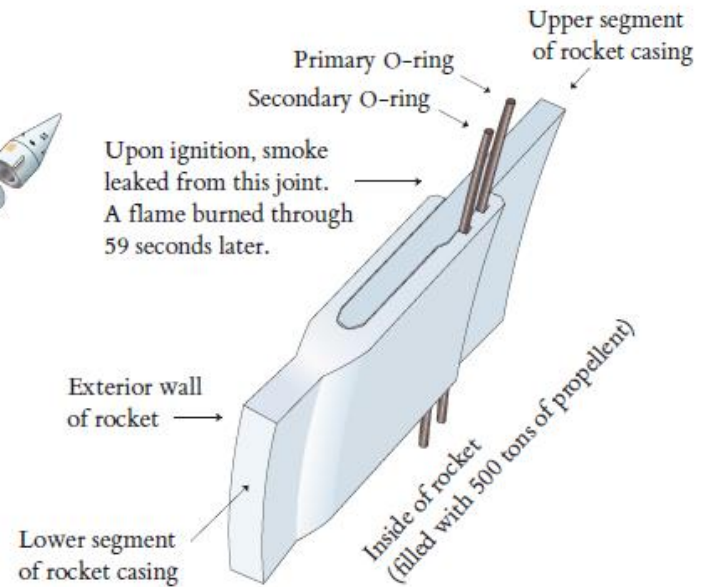
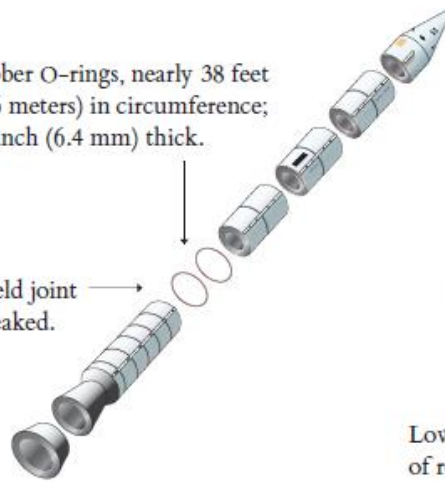


185 feet (56.4 meters)



The field joint that leaked.

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



FAST FORWARD 58 SECONDS AFTER IGNITION



WHAT HAPPENED?

WHAT WAS THE CAUSE?

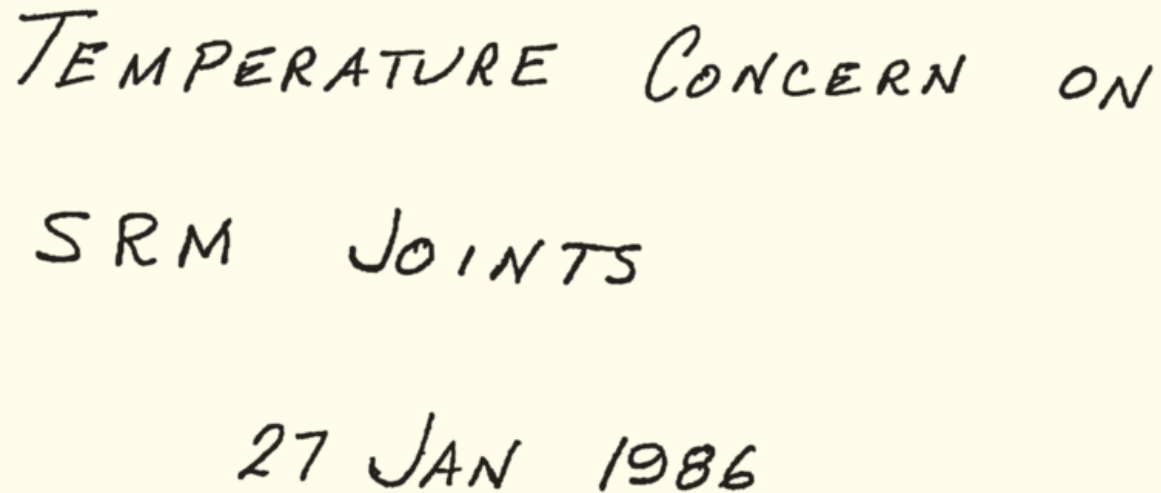
COULD IT HAVE BEEN PREVENTED?

ENGINEERS AT THIOKOL HAD A HUNCH

Two days before launch they presented their concerns

- created 13 charts to make their case

Slide #1:



TEMPERATURE CONCERN ON
SRM JOINTS
27 JAN 1986

- SRM – Solid Rocket Motor

SLIDE #2

Teaches about past damages to O-ring

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.)	
22A	None	None	0.280	None	None	36° -- 66°
22A	NONE	NONE	0.280	NONE	NONE	338° -- 18°
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	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

1161
 Oct 30, 1985
 80
 30
 July

*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.

SLIDES #2 AND 3

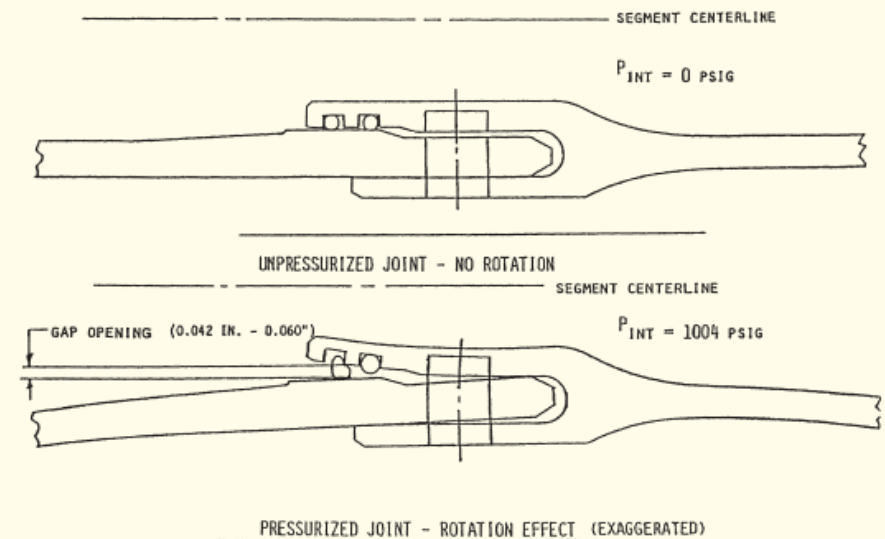
Teaches about O-ring damage mechanics and erosion

PRIMARY CONCERNS -

FIELD JOINT - HIGHEST CONCERN

- o EROSION PENETRATION OF PRIMARY SEAL REQUIRES RELIABLE SECONDARY SEAL FOR PRESSURE INTEGRITY
- o IGNITION TRANSIENT - (0-600 MS)
 - o (0-170 MS) HIGH PROBABILITY OF RELIABLE SECONDARY SEAL
 - o (170-330 MS) REDUCED PROBABILITY OF RELIABLE SECONDARY SEAL
 - o (330-600 MS) HIGH PROBABILITY OF NO SECONDARY SEAL CAPABILITY
- o STEADY STATE - (600 MS - 2 MINUTES)
 - o IF EROSION PENETRATES PRIMARY O-RING SEAL - HIGH PROBABILITY OF NO SECONDARY SEAL CAPABILITY
 - o BENCH TESTING SHOWED O-RING NOT CAPABLE OF MAINTAINING CONTACT WITH METAL PARTS GAP OPENING RATE TO MEOP
 - o BENCH TESTING SHOWED CAPABILITY TO MAINTAIN O-RING CONTACT DURING INITIAL PHASE (0-170 MS) OF TRANSIENT

PRIMARY CONCERNS - CONT



SLIDES #4 AND 5

Lists temperature and blow-by history for two SRMs

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)

<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 27	10 MPH 25 MPH

ASSUME YOU'RE A NASA MANAGER

Given the information provided in the company slides

- would you vote for a launch?
- ignore you know about the consequences



Be keenly aware of the immense PR pressures

- President Reagan's upcoming State of the Union speech
- the first civilian in space
- NASA's funding problems

Launch:

- **No:** OK with a PR disaster & possible budget cuts down the road
- **Yes:** the rocket company is too cautious & concerns are unproven

WHY THE RECOMMENDATION FAILED

Presentation only has exactly two shuttle flights

- one with two blow-by's and high temperature
- one with two blow-by's and low temperature
- ignores all other 22 shuttle flights (SRM)

Statistically weak

Recommendation

- "O-ring temp must be $>53^{\circ}\text{F}$ at launch"
- is only based on a sample size of 1
- context of other flights is missing
- no statistical leverage

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

Test rockets ignited on fixed horizontal platforms in Utah.

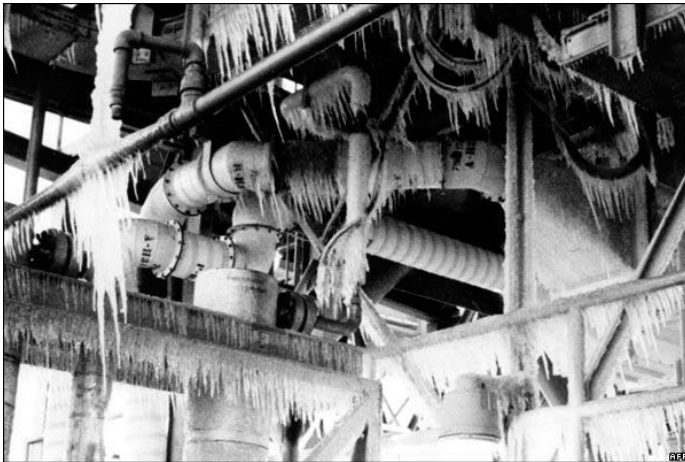
The only 2 shuttle launches (of 24) for which temperatures were shown in the 13 Challenger charts.

Forecasted O-ring temperatures for the Challenger.

DEFICIENCIES

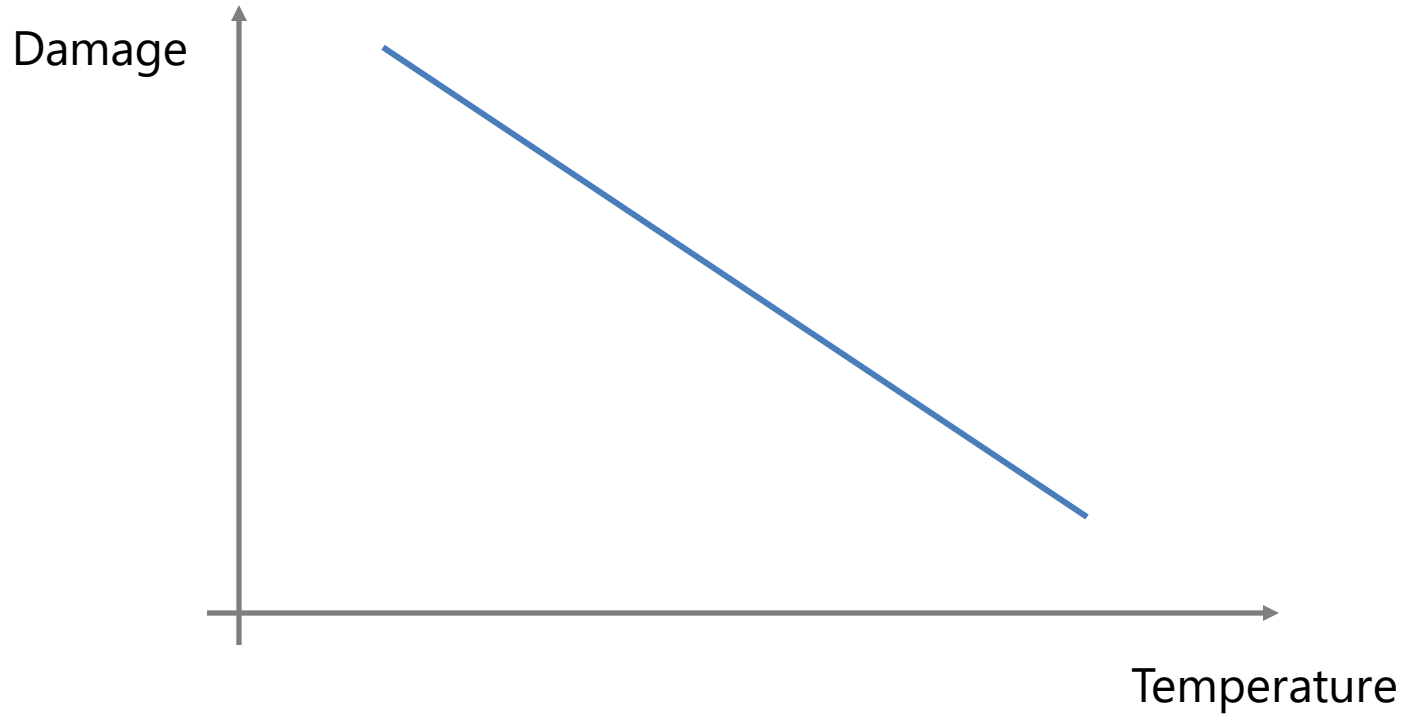
Lots of numbers and facts

But no causal evidence that could predict



What is needed?

WHAT IS NEEDED?



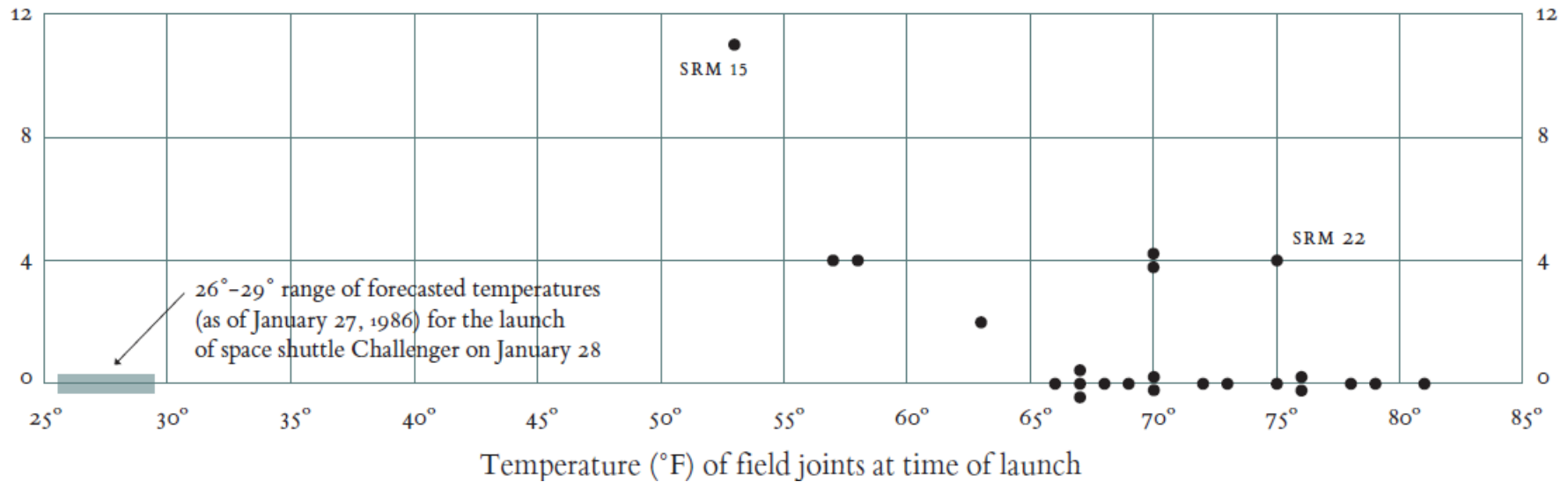
Need a measure for damage

DAMAGE INDEX

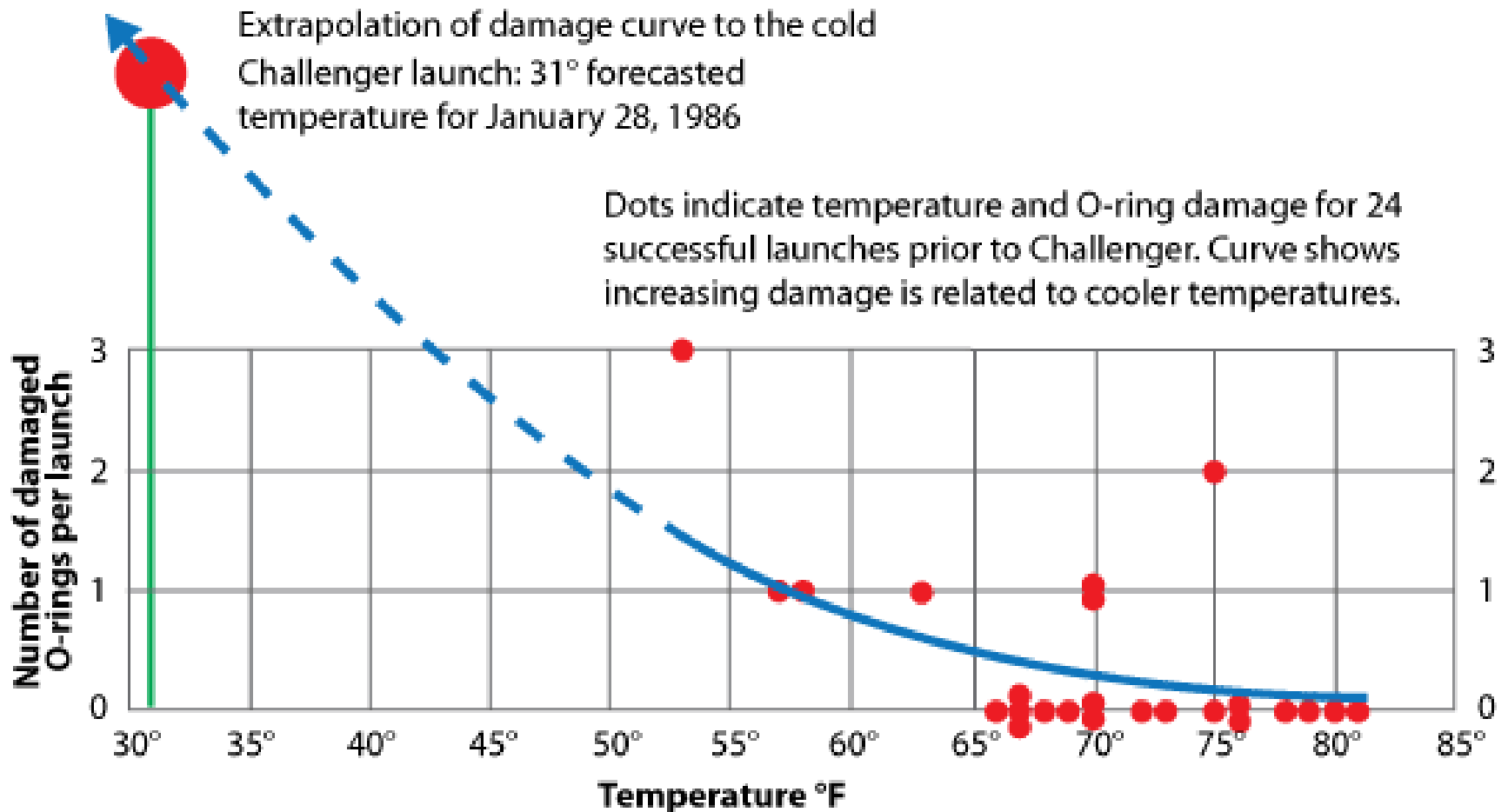
Flight	Date	Temperature °F	Erosion incidents	Blow-by incidents	Damage index	Comments
51-C	01.24.85	53°	3	2	11	Most erosion any flight; blow-by; back-up rings heated.
41-B	02.03.84	57°	1		4	Deep, extensive erosion.
61-C	01.12.86	58°	1		4	O-ring erosion on launch two weeks before Challenger.
41-C	04.06.84	63°	1		2	O-rings showed signs of heating, but no damage.
1	04.12.81	66°			0	Coollest (66°) launch without O-ring problems.
6	04.04.83	67°			0	
51-A	11.08.84	67°			0	
51-D	04.12.85	67°			0	
5	11.11.82	68°			0	
3	03.22.82	69°			0	
2	11.12.81	70°	1		4	Extent of erosion not fully known.
9	11.28.83	70°			0	
41-D	08.30.84	70°	1		4	
51-G	06.17.85	70°			0	
7	06.18.83	72°			0	
8	08.30.83	73°			0	
51-B	04.29.85	75°			0	
61-A	10.30.85	75°		2	4	No erosion. Soot found behind two primary O-rings.
51-I	08.27.85	76°			0	
61-B	11.26.85	76°			0	
41-G	10.05.84	78°			0	
51-J	10.03.85	79°			0	
	06.27.82	80°			?	O-ring condition unknown; rocket casing lost at sea.
51-F	07.29.85	81°			0	

VISUALIZE IT – JUST THE FACTS

O-ring damage index, each launch

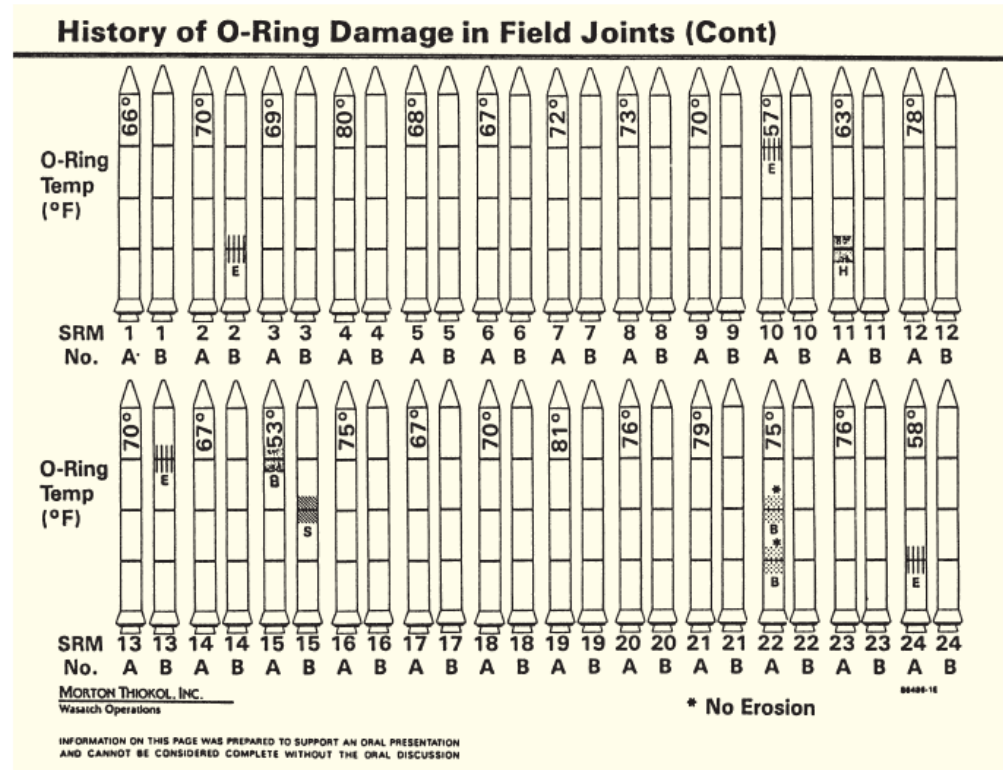


VISUALIZE IT – TELL THE STORY



SHOWN AT CONGRESSIONAL HEARINGS

Used these charts



All information is there

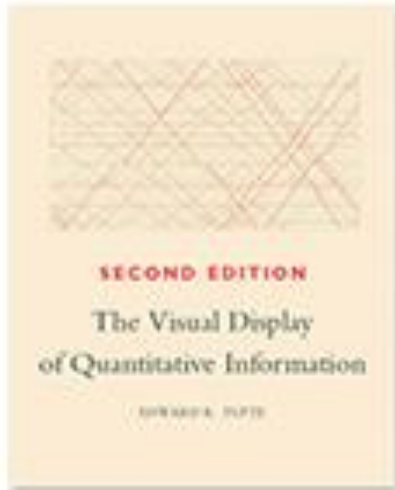
- but very hard to identify and assimilate
- why?

SOURCE: EDWARD TUFTE



Four seminal books

- standard literature for every visualization enthusiast
- written 1983, 1990, 1997, 2006



- taught information design at Princeton University
- now a professor at Yale University

COURSE TOPICS



CSE 332 INTRODUCTION TO VISUALIZATION



NON-SPATIAL DATA

DATA MINING

INSIGHT

SPATIAL DATA

DISPLAY TECHNOLOGY

VISUALIZATION
INTERACTION
ANALYSIS



PERCEPTION & COGNITION

LARGE & BIG DATA

DOMAIN KNOWLEDGE

KNOWLEDGE

HIGH PERFORMANCE COMPUTING

SPATIAL DATA

shock wave

virtual frog

spiral flow

nerve cell

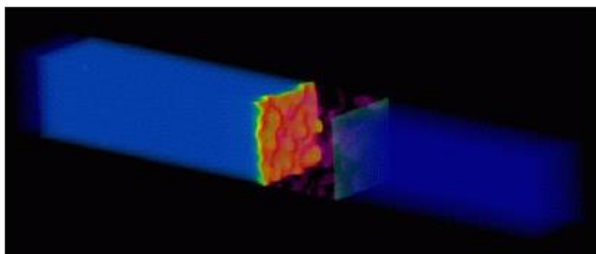
transparent MRI head

wind flow

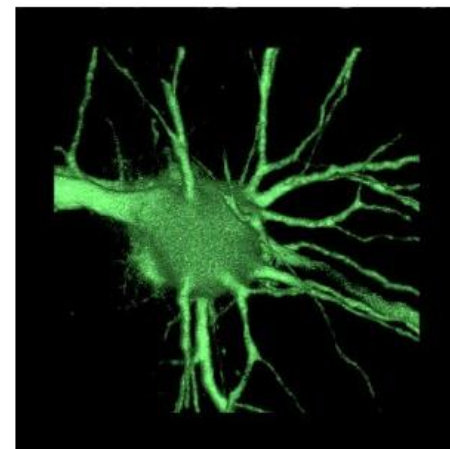
semi-transparent
tomato

MRI head

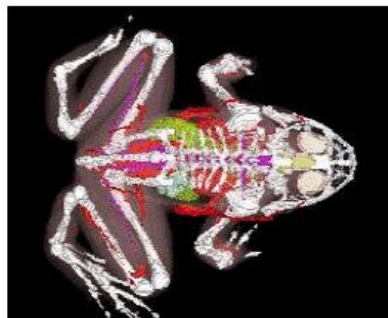
SPATIAL DATA



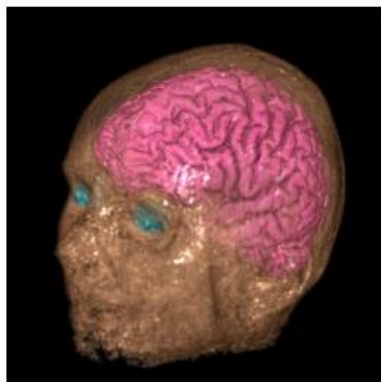
shock wave



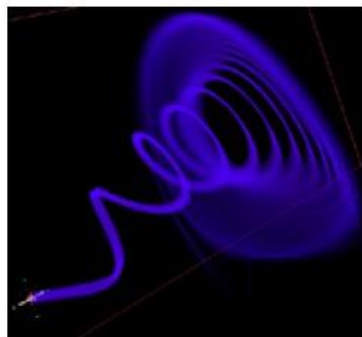
nerve cell



virtual frog



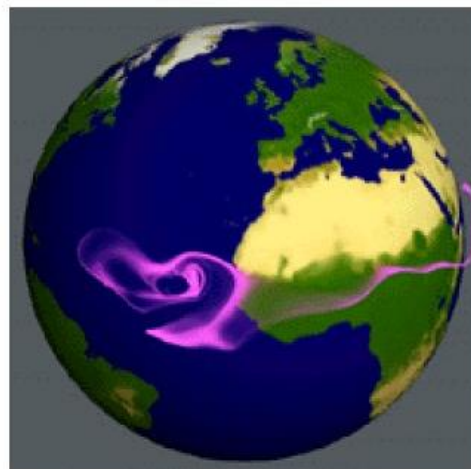
transparent MRI head



spiral flow



semi-transparent tomato



wind flow

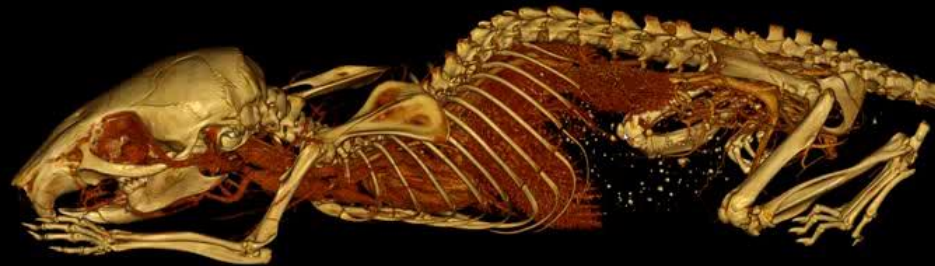


MRI head

SPATIAL DATA

Example: Datasets obtained by 3D volumetric scans (CT, MRI)

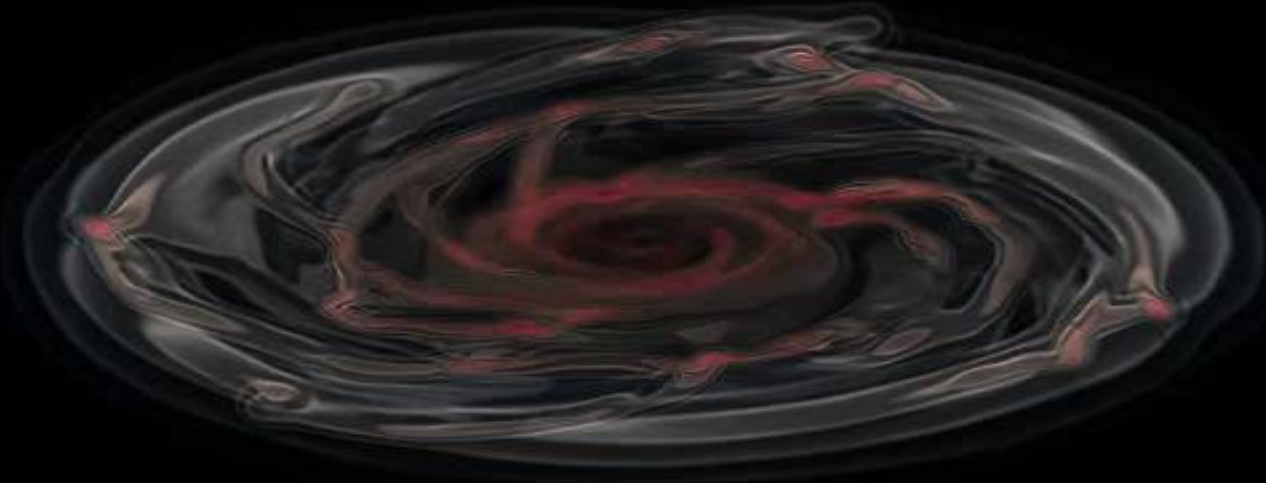
- what are some questions you might have?



SPATIAL DATA

Example: Datasets obtained by 3D Simulations

- what are some questions you might have?



SPATIAL DATA

Example: Data obtained by observation-supported simulations

- what are some questions you might have?

NON-SPATIAL DATA

The salient features of a car:

- miles per gallon (MPG)
- top speed
- acceleration
- number of cylinders
- horsepower
- weight
- year
- country origin
- brand
- number of seats
- number of doors
- reliability (# of breakdowns)
- and so on...

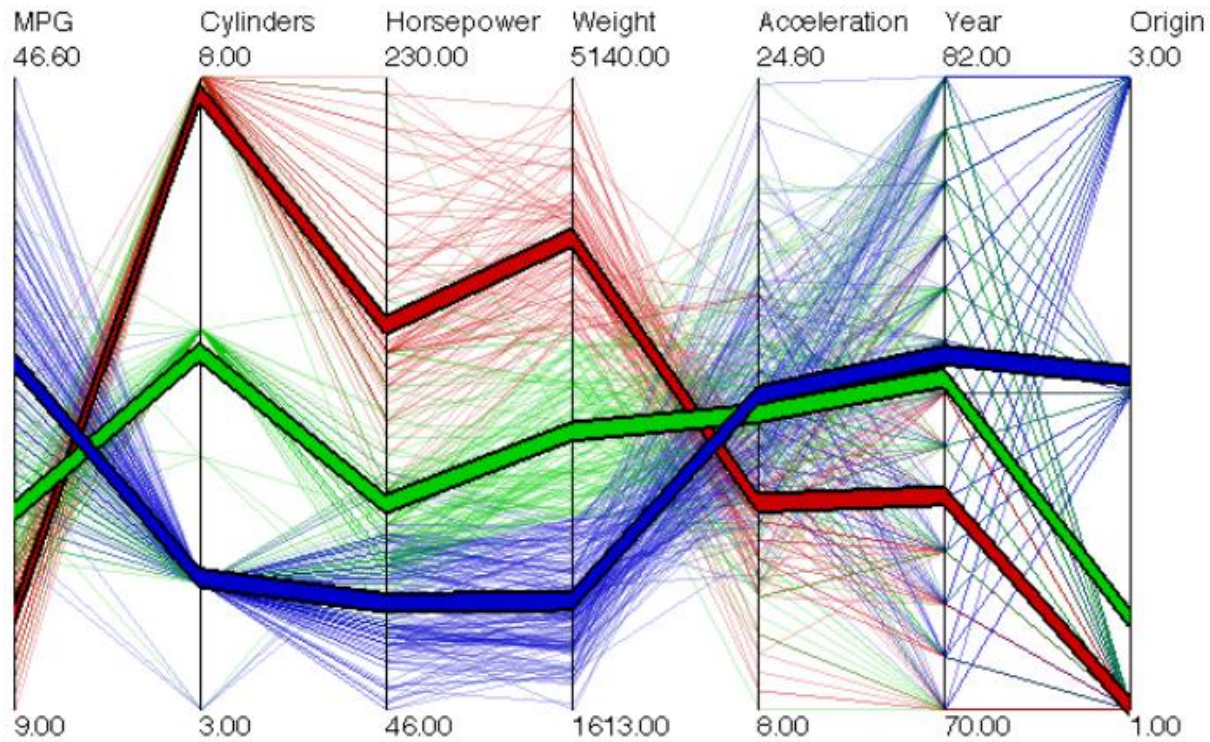


CAN YOU VISUALIZE THEM LIKE THIS?

A1	Urban population															
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Urban population	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974
2	Afghanistan	769308	811389	855131	900646	948060	997499	1053104	1110728	1170961	1234664	1302370	1391081	1483942	1579748	1676656
3	Albania	494443	511637	529182	547024	565117	583422	601897	620508	639234	658062	676985	698179	719561	741149	762972
4	Algeria	3293999	3513320	3737362	3969886	4216744	4483048	4644898	4822860	5015071	5218184	5429743	5618190	5813978	6017932	6231383
5	American Samoa	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	Andorra	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	Angola	521205	552777	585121	618345	652638	688181	729595	772643	817418	863993	912486	982944	1056617	1133936	1215437
8	Antigua and Barbuda	21699	21737	21878	22086	22309	22513	22717	22893	23053	23218	23394	24046	24718	25342	25826
9	Argentina	15224096	15588864	15957125	16328045	16700303	17073371	17432905	17793789	18160868	18540720	18938137	19335571	19750609	20180707	20621674
10	Armenia	957974	1008899	1061551	1115546	1170414	1225785	1281346	1337060	1393199	1450241	1508526	1565054	1622558	1680709	1739019
11	Aruba	24996	25514	26019	26498	26941	27337	27683	27984	28247	28491	28726	28959	29188	29409	29610
12	Australia	8375329	8585577	8840666	9055650	9279777	9508980	9770529	9937118	10157212	10416192	10668471	11050785	11271606	11461308	11771589
13	Austria	4560057	4589541	4621666	4653194	4685421	4715750	4754585	4778506	4798552	4817322	4849178	4871380	4904030	4932109	4939292
14	Azerbaijan	1857673	1929429	2004258	2080816	2157307	2232355	2306310	2378380	2448728	2517815	2586000	2660687	2734631	2807879	2880491
15	Bahamas	65457	69655	74179	78961	83902	88918	93931	98974	103944	108721	113219	117339	121142	124761	128393
16	Bahrain	128480	133815	139791	146052	152097	157596	162844	167630	172373	177677	183997	191379	199768	209201	219678
17	Bangladesh	2761049	2947191	3141372	3344120	3556037	3777716	4047121	4329144	4624445	4933701	5257558	5710277	6184871	6682073	7202503
18	Barbados	84884	85284	85761	86285	86797	87259	87707	88117	88526	88986	89532	90518	91596	92713	93796
19	Belarus	2656152	2774166	2896449	3022217	3150553	3280410	3415984	3554673	3695363	3836802	3977600	4131179	4285735	4439788	4591705
20	Belgium	8435075	8489549	8548773	8620194	8709437	8796088	8865259	8924327	8968568	9003536	9040444	9086816	9134227	9175144	9217085
21	Belize	49165	50608	52156	53734	55226	56561	57756	58820	59746	60532	61186	61883	62445	62984	63665
22	Benin	211033	229172	248065	267765	288321	309788	332782	366019	396065	427482	460341	500355	542251	586179	632320
23	Bermuda	44400	45500	46600	47700	48800	50100	51000	52000	53000	54000	55000	56000	57000	58000	59000
24	Bhutan	8064	8778	9526	10311	11137	12010	13089	14230	15445	16750	18158	19926	21827	23858	26008
25	Bolivia	1233398	1271250	1310294	1350615	1392328	1435536	1480255	1526529	1574517	1624419	1676370	1730434	1786553	1844596	1904355
26	Bosnia and Herzegovi	604204	637337	671124	705395	739884	774380	812856	851325	890011	929301	969514	1008688	1048890	1089898	1131315
27	Botswana	16240	17379	18583	19855	21203	22631	28191	34090	40352	46995	54038	61638	69689	78254	87422
28	Brazil	32662018	34463344	36353068	38320171	40346703	42418482	44548227	46722996	48945984	51223962	53563179	56042505	58587770	61207586	63913385
29	Brunei	35501	38753	42173	45802	49699	53916	58461	63355	68595	74157	80024	83802	87671	91616	95629
30	Bulgaria	2918659	3085061	3251675	3418610	3588246	3756058	3889518	4022040	4159890	4301340	4440270	4554810	4667059	4782931	4907107
31	Burkina Faso	221872	230199	238713	247472	256558	266039	275958	286311	297074	308196	319642	332556	345877	359655	373966
32	Burundi	58810	61055	63344	65696	68137	70683	73370	76186	79034	81779	84324	90879	97308	103757	110494
33	Cambodia	559631	578678	598248	618631	640243	663272	747219	835638	927177	1019449	1110079	962037	806676	645287	479631
34	Cameroon	751711	801009	852578	906523	962928	1021891	1088521	1158289	1231375	1307967	1388275	1522958	1664410	1813278	1970385
35	Canada	12375125	12764121	13145207	13536503	13941055	14345262	14727261	15108962	15470875	15800439	16142268	16381341	16640381	16920220	17221765
36	Cape Verde	32791	34353	35972	37672	39487	41435	43592	45884	48200	50383	52314	54103	55620	56940	58184
37	Cayman Islands	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
38	Central African Rep.	302157	317715	333986	351001	368787	387357	408129	429825	452326	475441	499036	526414	554452	583376	613530
39	Chad	198777	213406	228652	244499	260903	277834	305390	333898	363523	394530	427153	467662	510348	554973	601045
40	Channel Islands	42565	42665	42792	42941	43102	43269	43437	43604	43765	43916	44051	44028	43987	43907	43762

How are MPG, weight, HP, and reliability related? Are there tradeoffs?
Which car is best for me?

HIGH-DIMENSIONAL DATA VISUALIZATION



BIG DATA

12+ TBs
of tweet data
every day

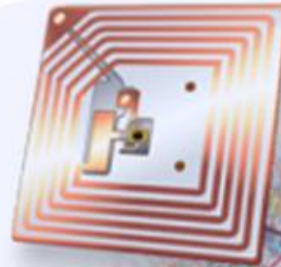


25+ TBs of
log data every day

? TBs of
data every day



30 billion RFID
tags today
(1.3B in 2005)



76 million smart
meters in 2009...
200M by 2014



4.6 billion
camera
phones
world wide



100s of millions
of GPS
enabled
devices
sold
annually



2+ billion
people on the Web
by end
2011

THE SCIENTIFIC METHOD

IN THE AGE OF DATA SCIENCE

Formulate
Question

Generate
Hypothesis

Form Experiment
(find data sources)

Form Testable
Prediction

Collect Data
(scrape, mine)

Analyze Data

Test Prediction
(visualize)

Publish Results



MODERN DATA SCIENTIST

21st century, requires a mixture of multidisciplinary skills ranging from computer science, communication and business. A modern data scientist is, is equally hybrid. The modern data scientist really is:

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative



PROGRAMMING & DATABASE

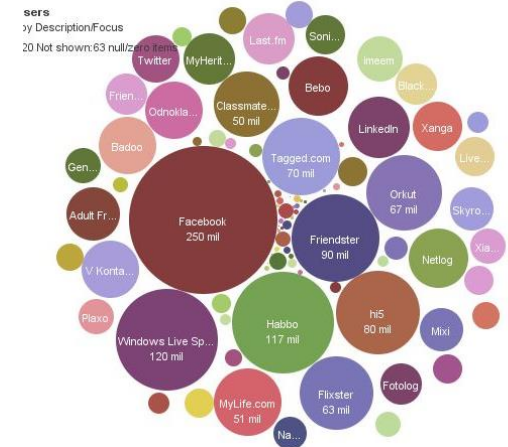
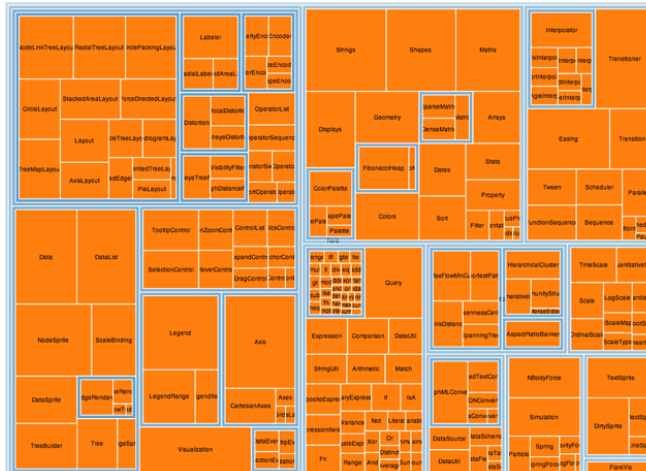
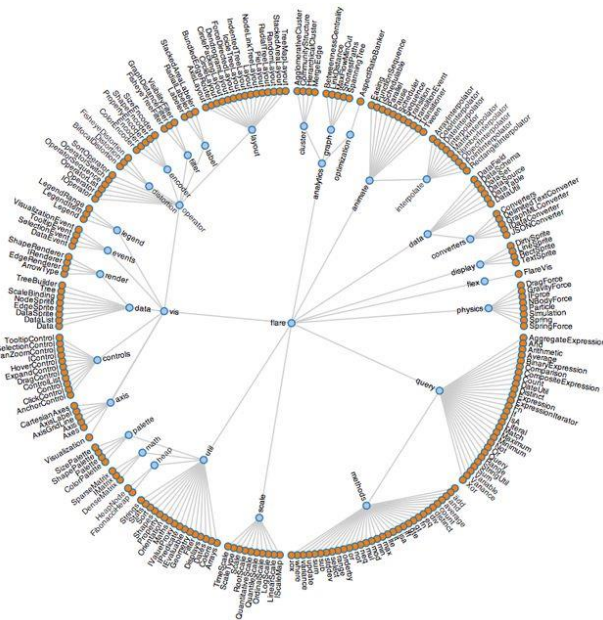
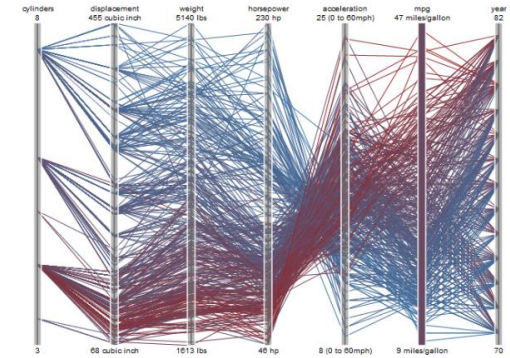
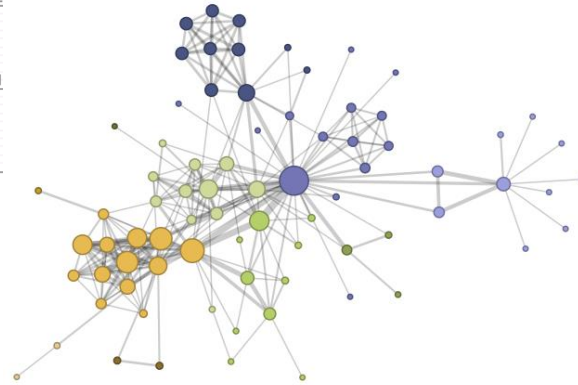
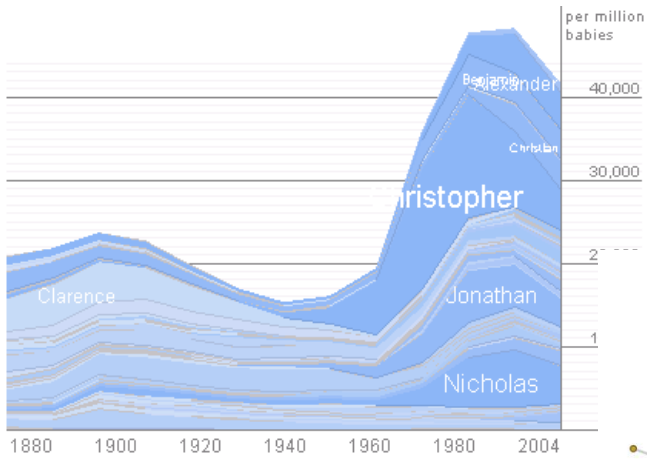
- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing packages, e.g., R
- ☆ Databases: SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query

COMMUNICATION & VISUALIZATION

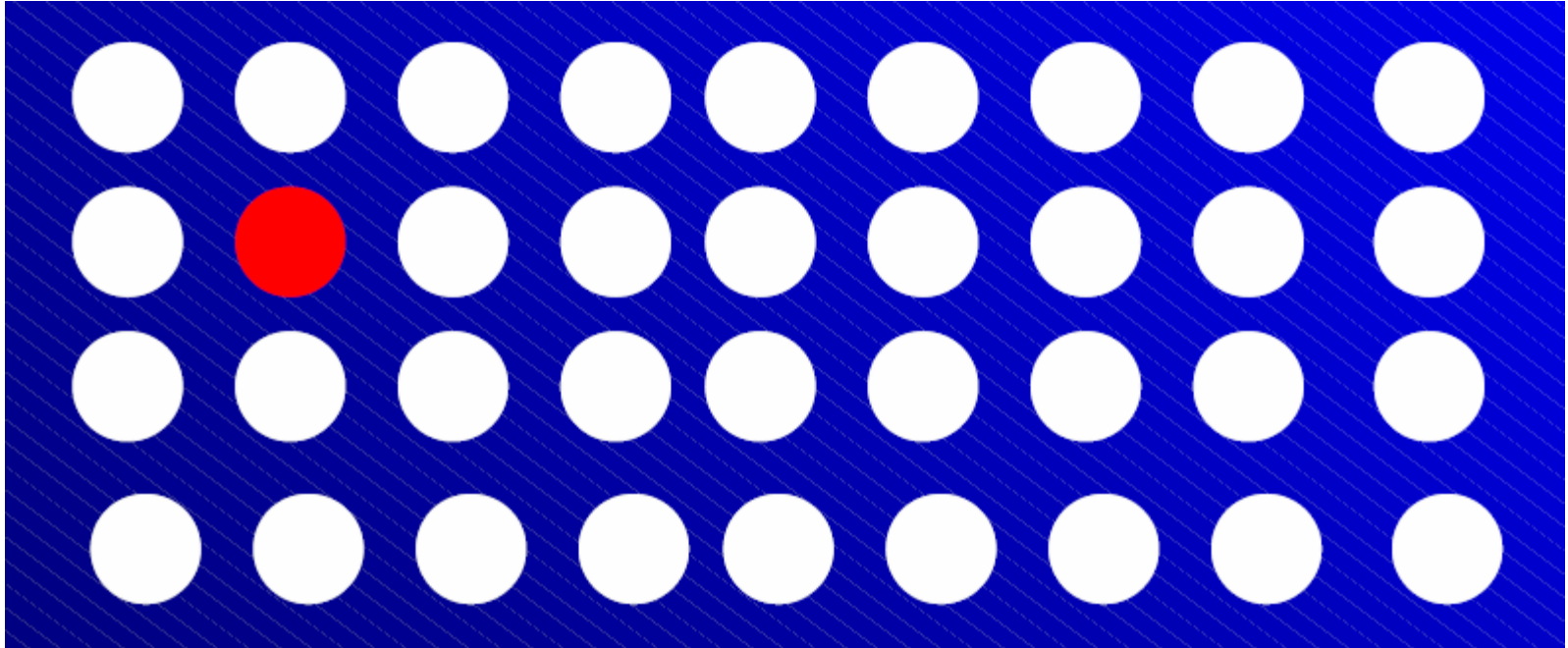
- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

VISUALIZATION CAN BE BEAUTIFUL

VISUALIZATION CAN BE BEAUTIFUL

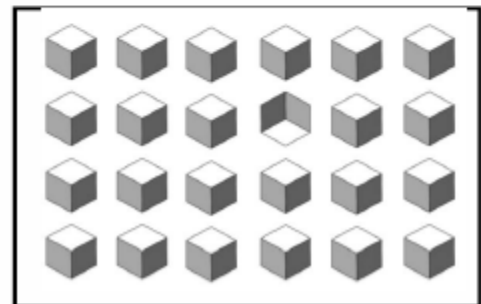
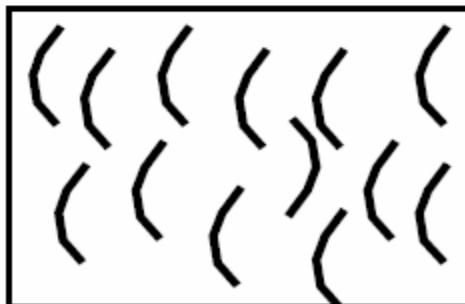
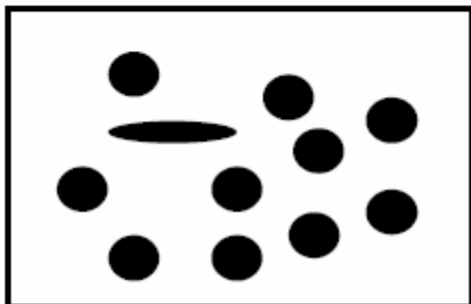
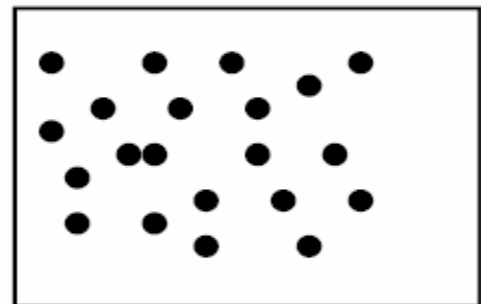
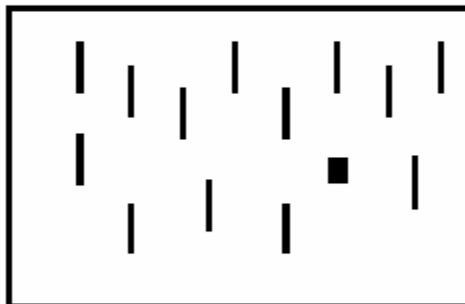
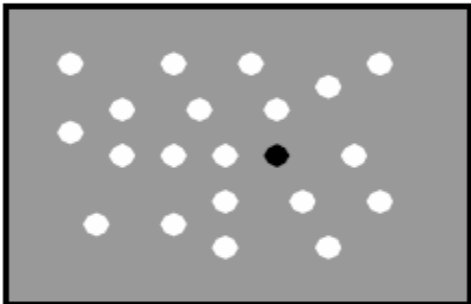
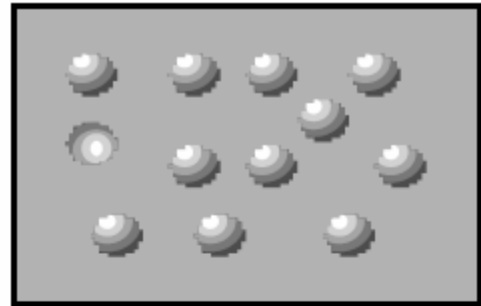
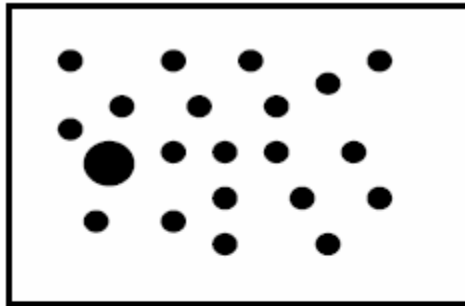
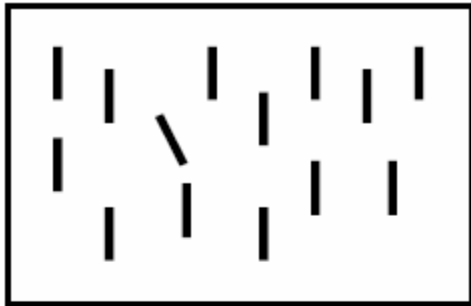


VISUALIZATION IS FAST

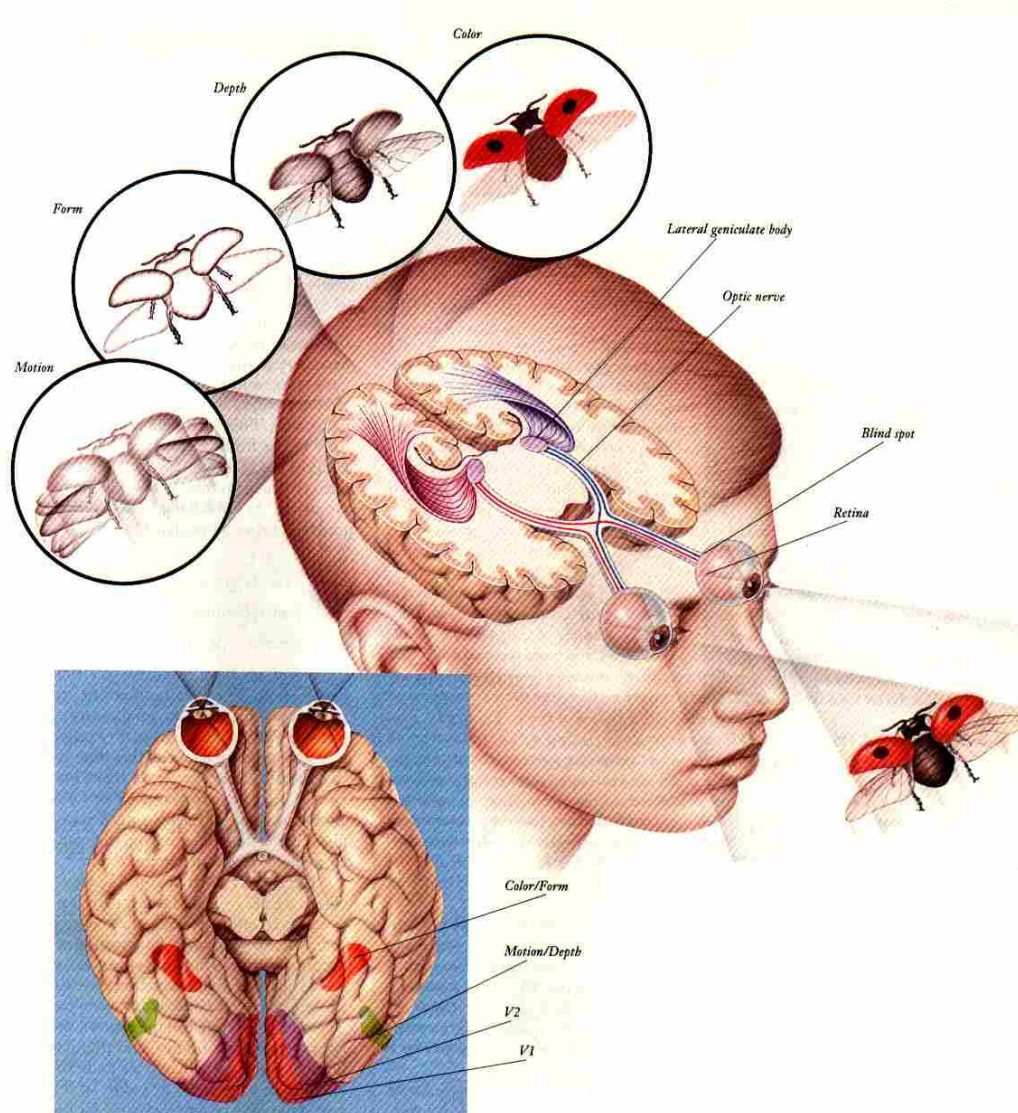


< 200 ms to recognize the red dot

VISUALIZATION IS FAST



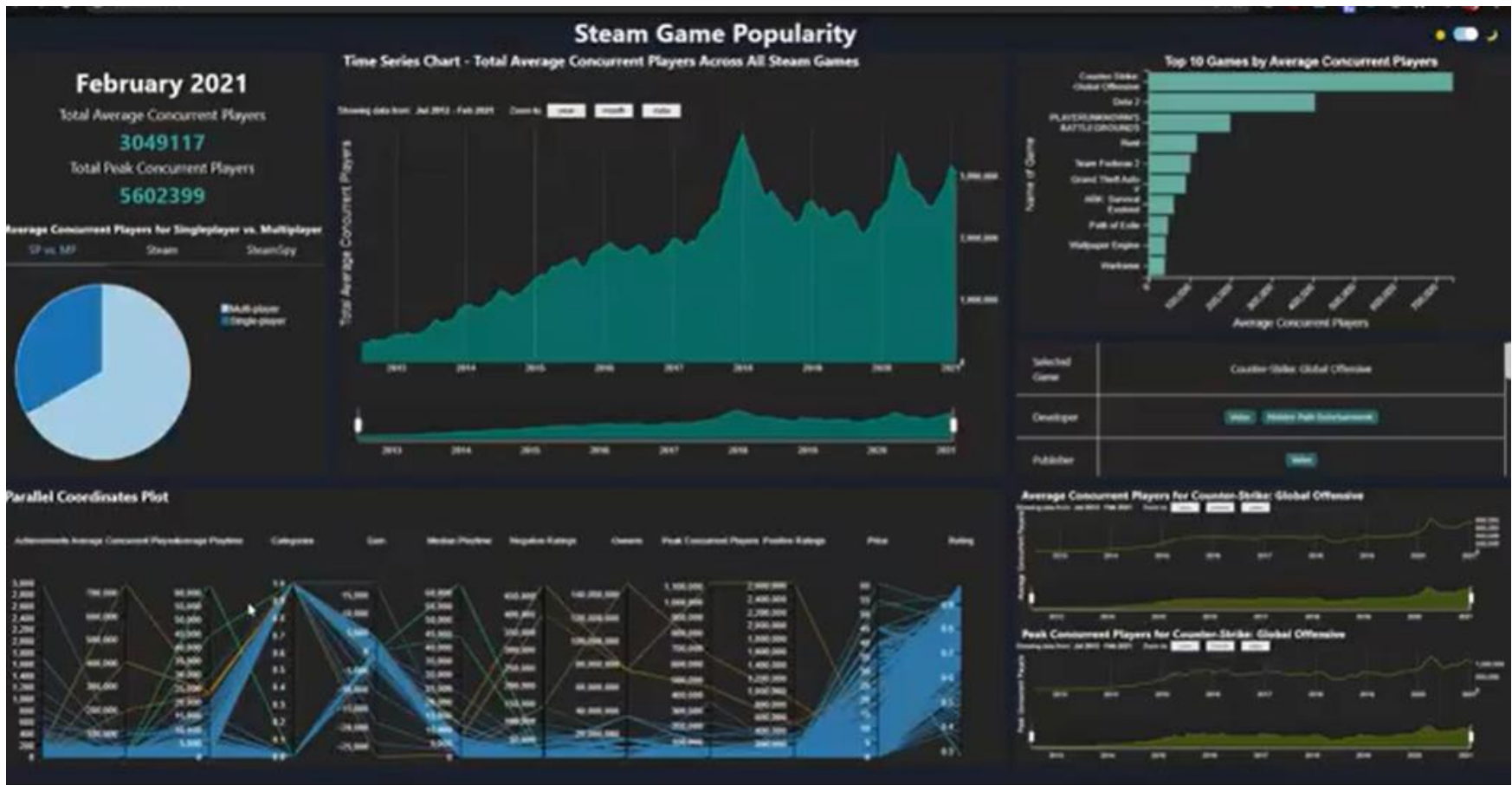
VISION IS MASSIVELY PARALLEL



more than
50% of the
brain

VISUALIZATION CAN BE INTERACTIVE

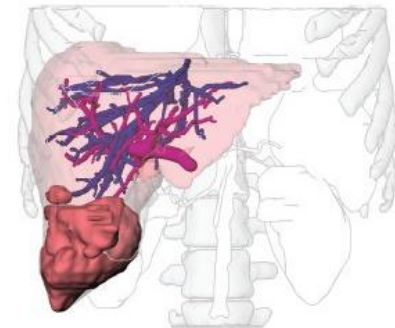
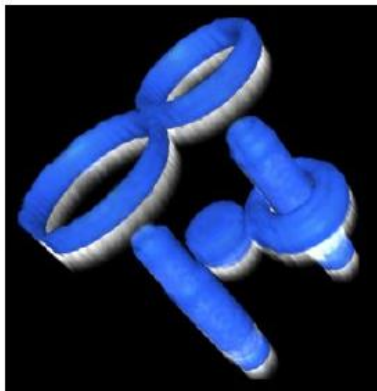
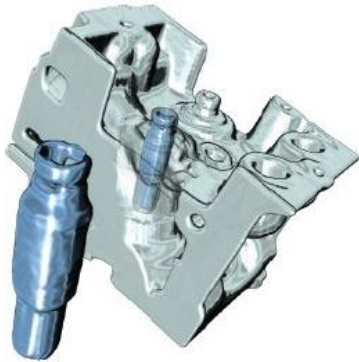
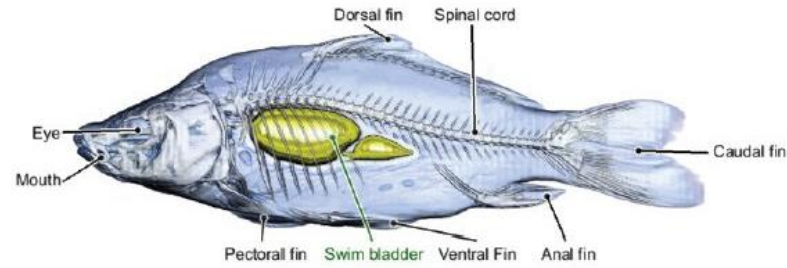
[youtube](#)



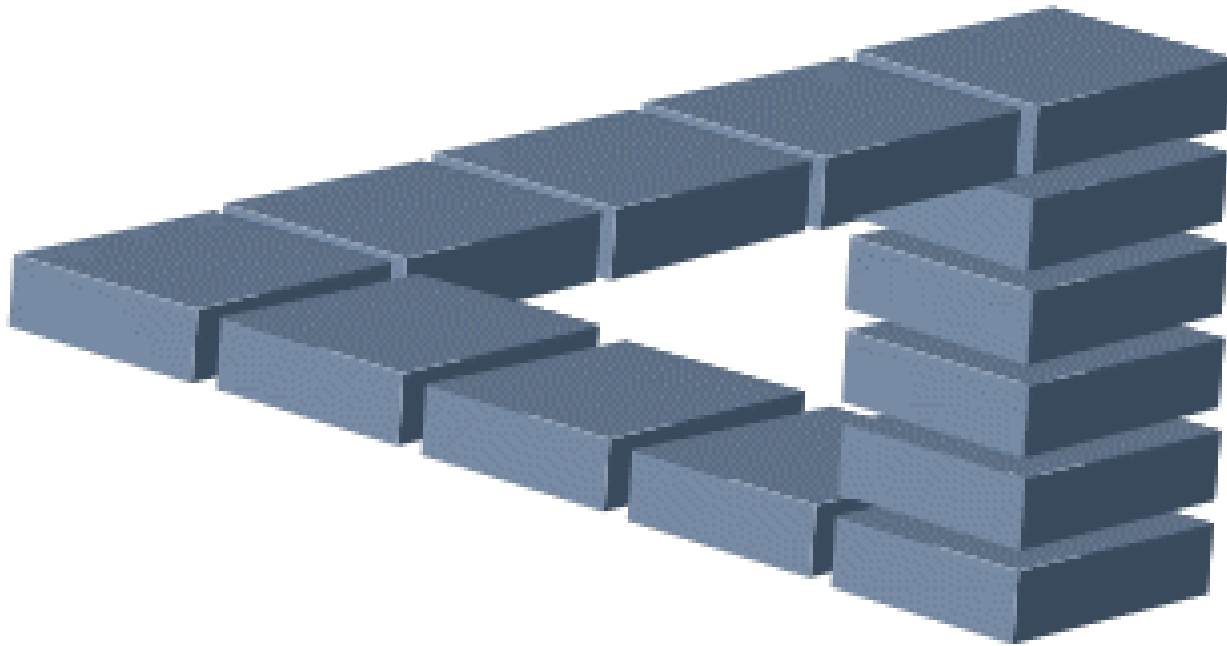
VISUALIZATION HAS A LONG HISTORY



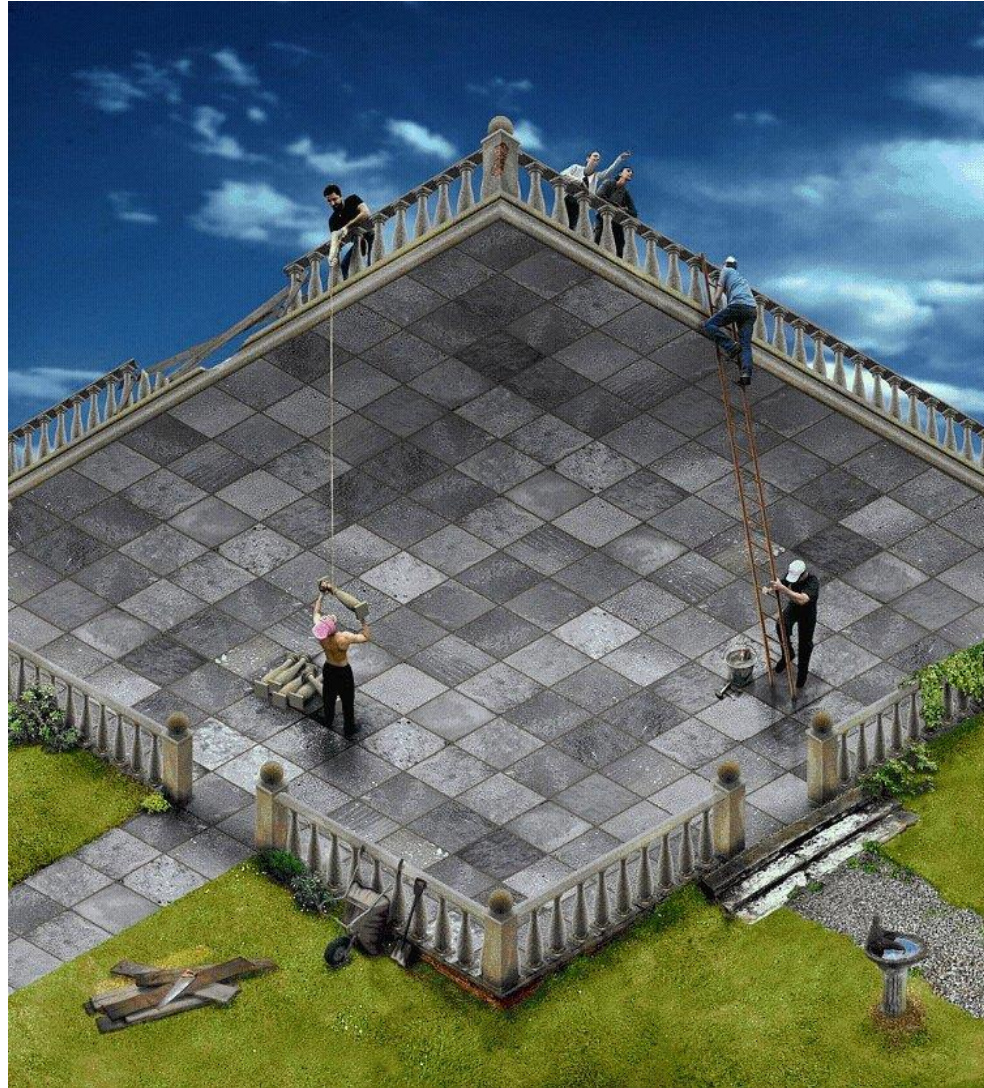
VISUALIZATION CAN BE INSPIRED BY ART



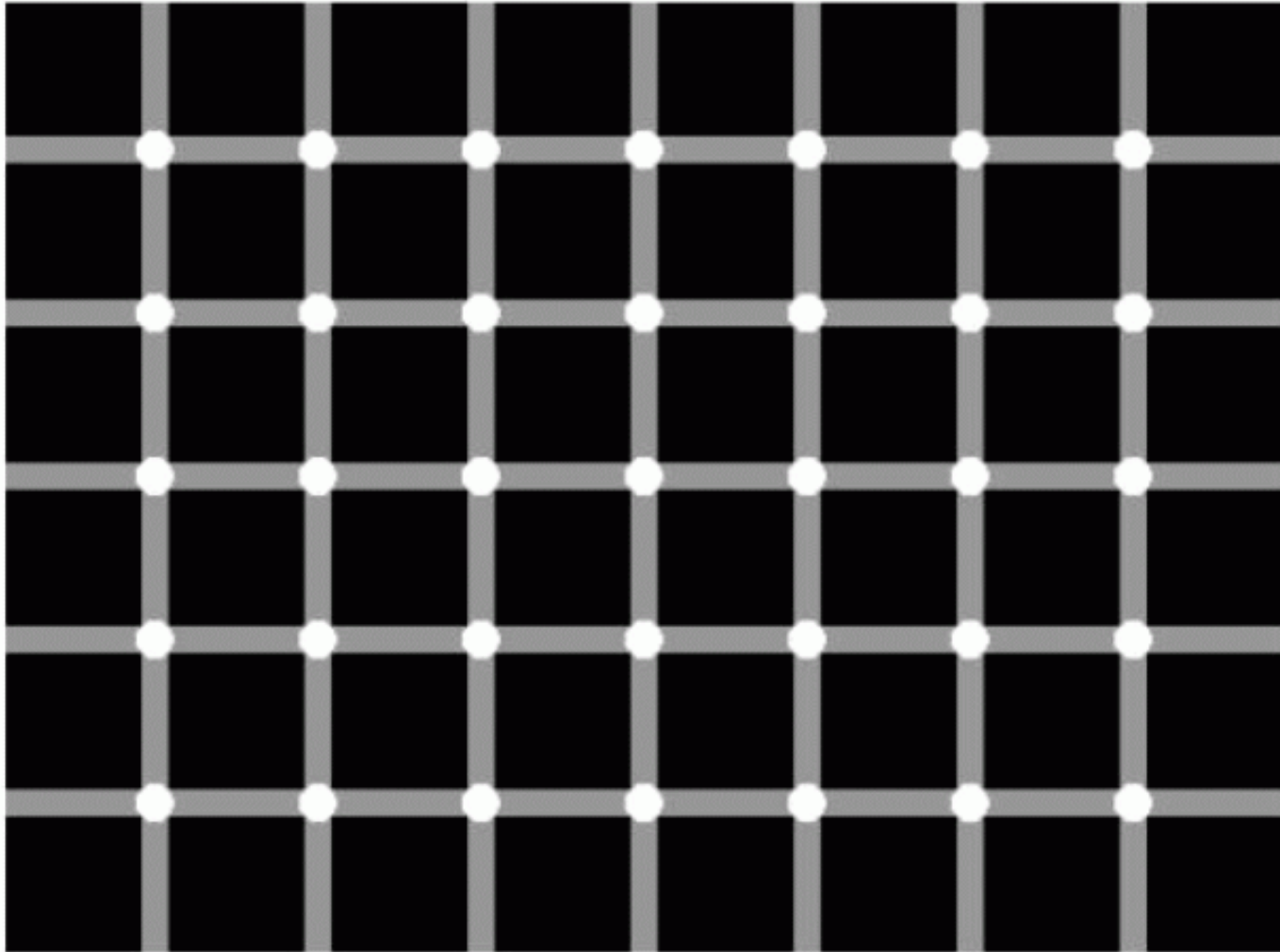
VISUALIZATION CAN BE DECEPTIVE



VISUALIZATION CAN BE DECEPTIVE

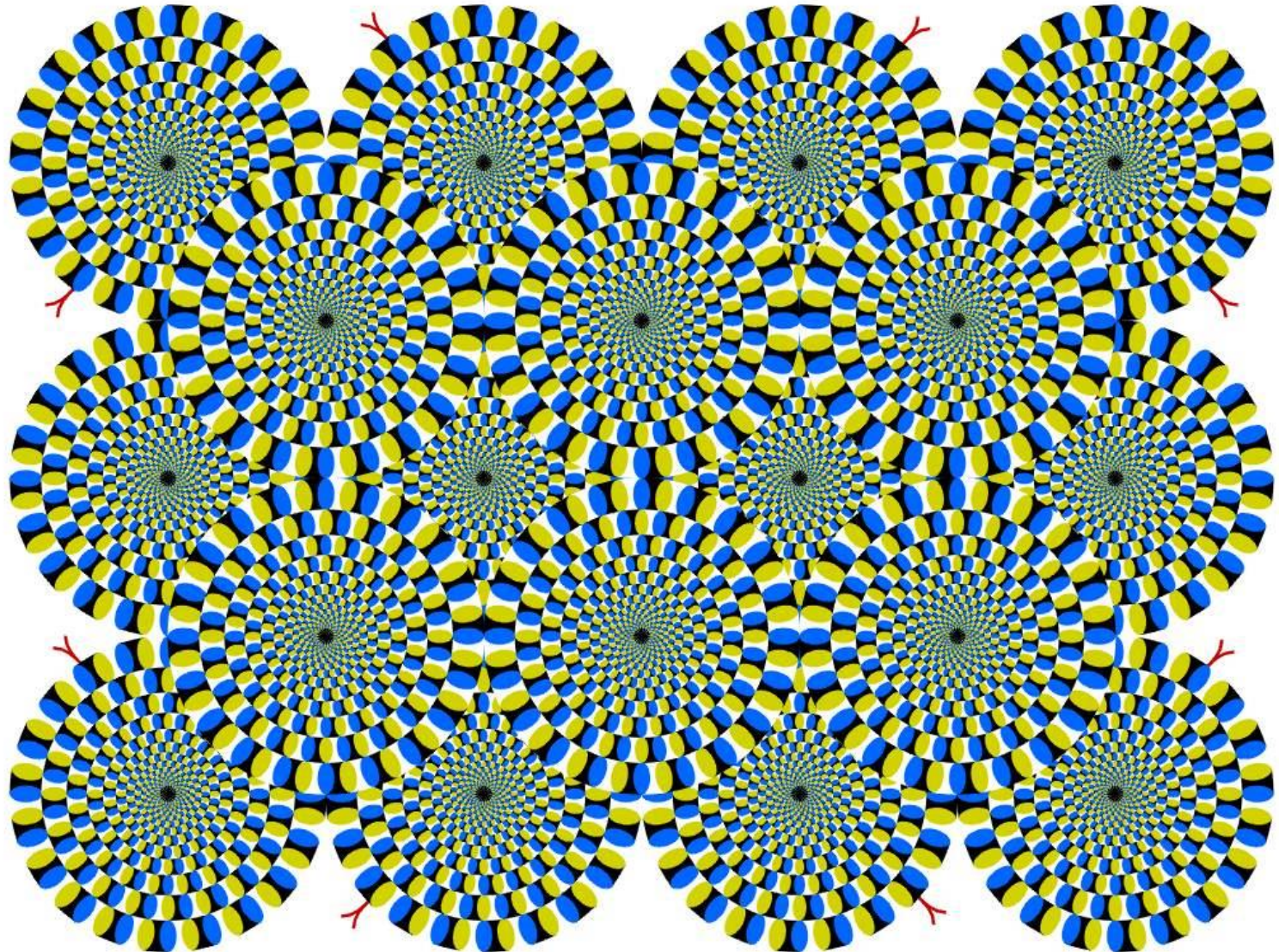


VISUALIZATION CAN BE DECEPTIVE

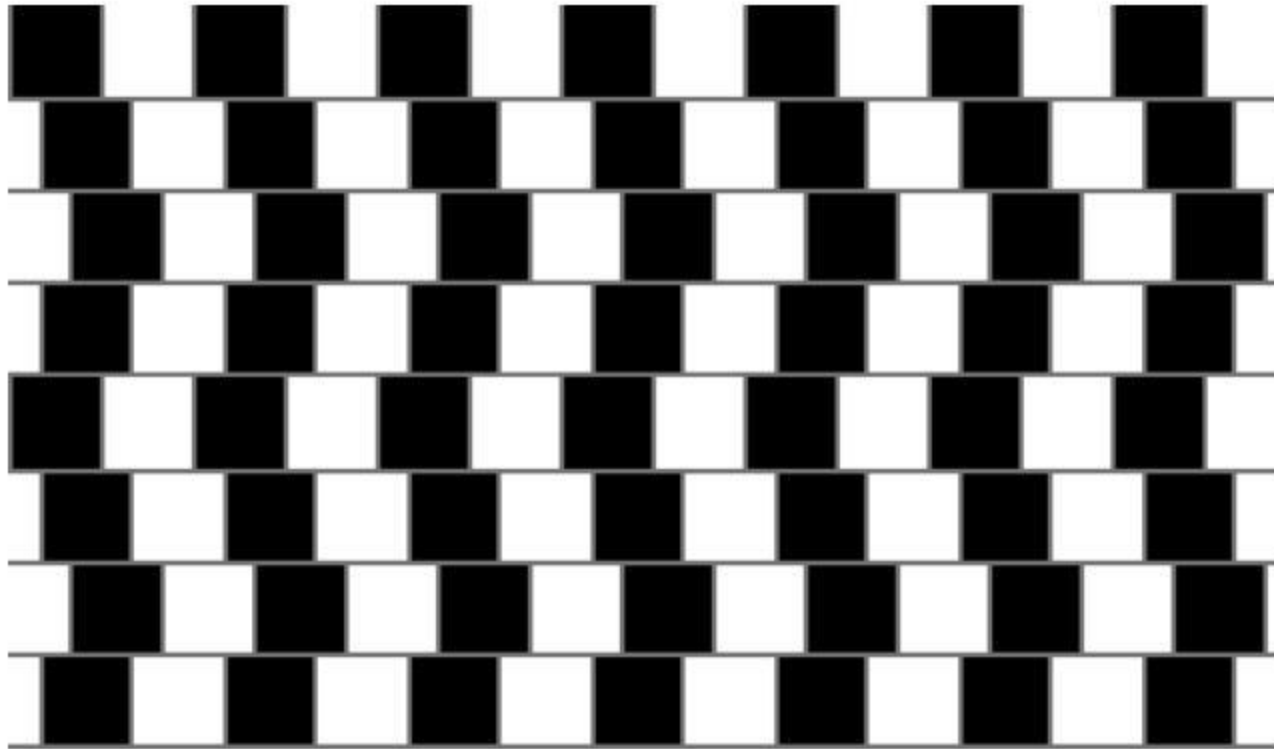


Count the number of black dots

VISUALIZATION CAN BE DECEPTIVE

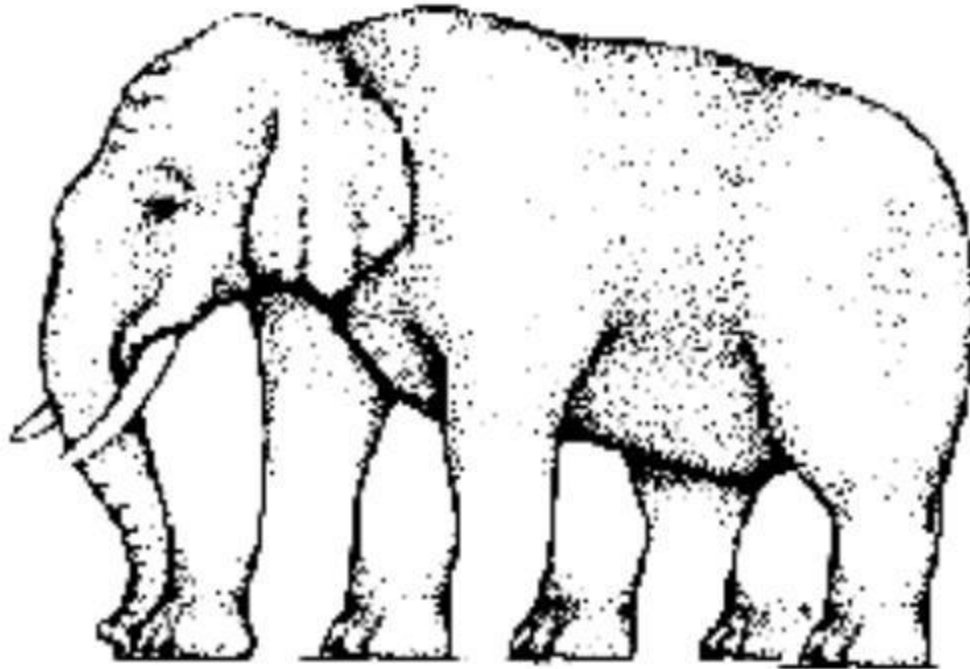


VISUALIZATION CAN BE DECEPTIVE



Are the horizontal lines parallel or do they slope?

VISUALIZATION CAN BE DECEPTIVE



How many legs does this elephant have?

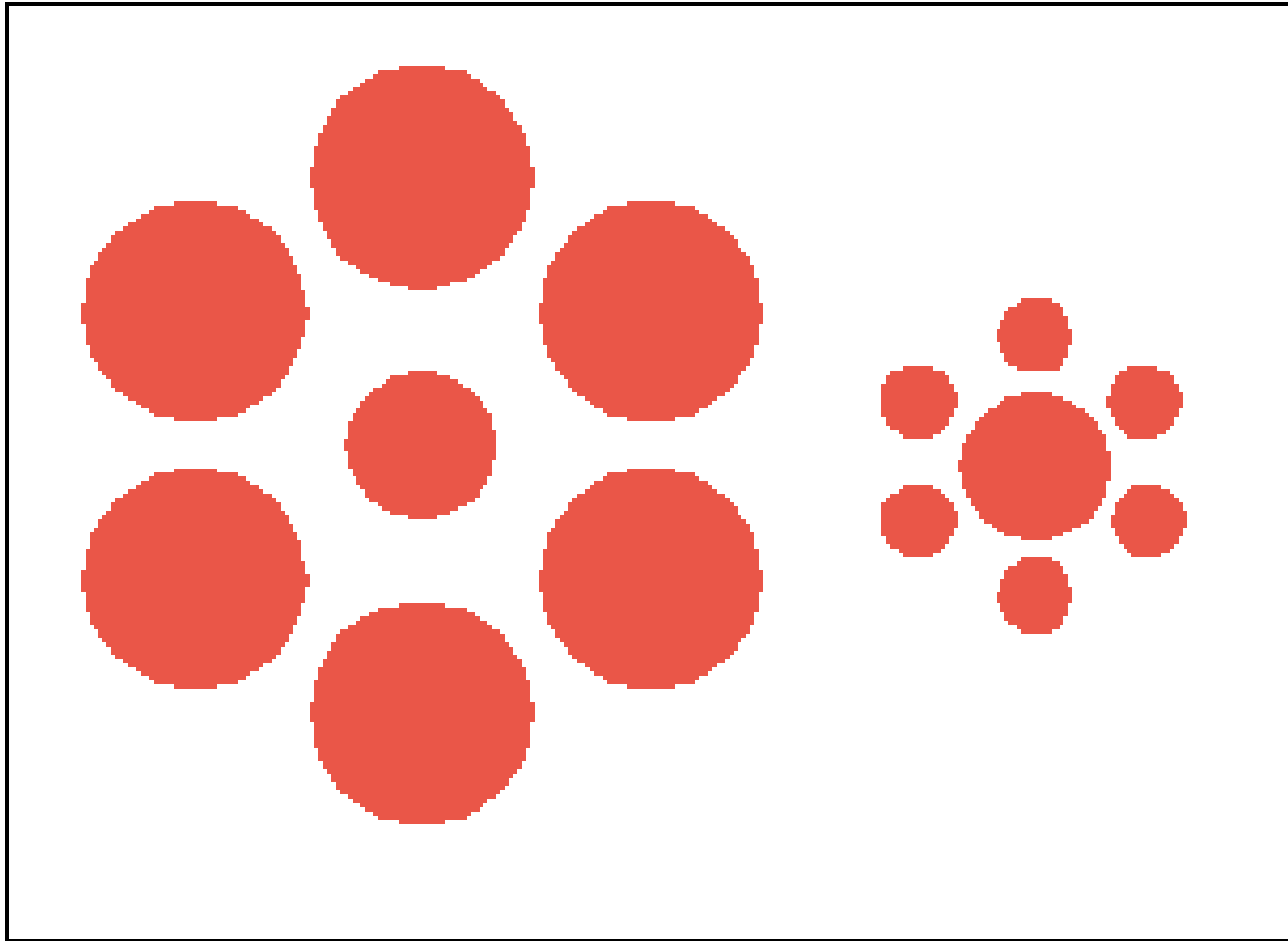
VISUALIZATION CAN BE DECEPTIVE



Julian Beever



VISUALIZATION CAN BE DECEPTIVE

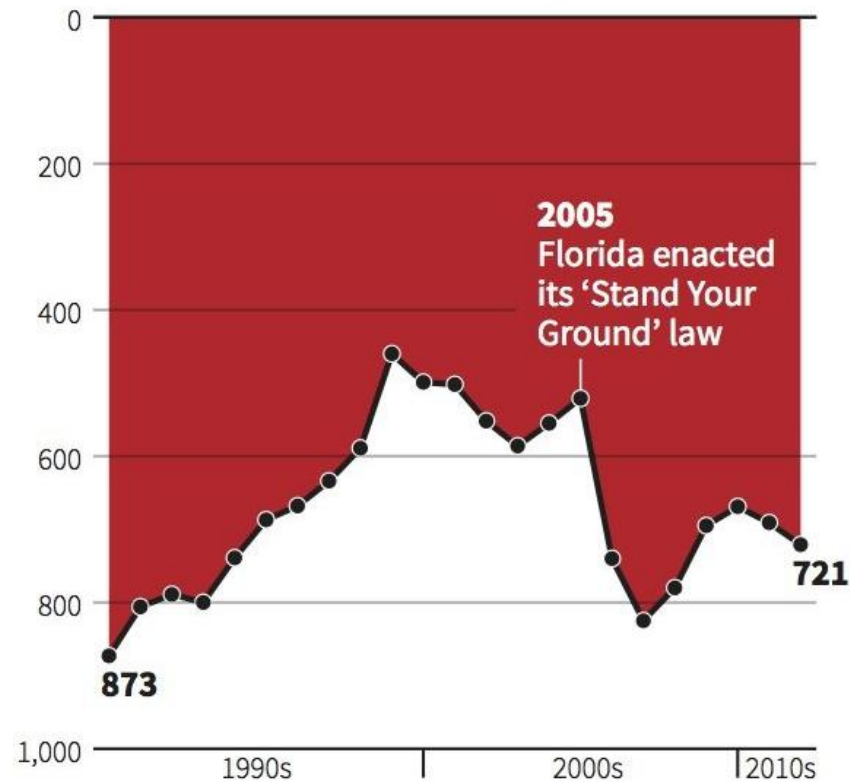


Which circle in the middle is bigger?

VISUALIZATION CAN BE DECEPTIVE

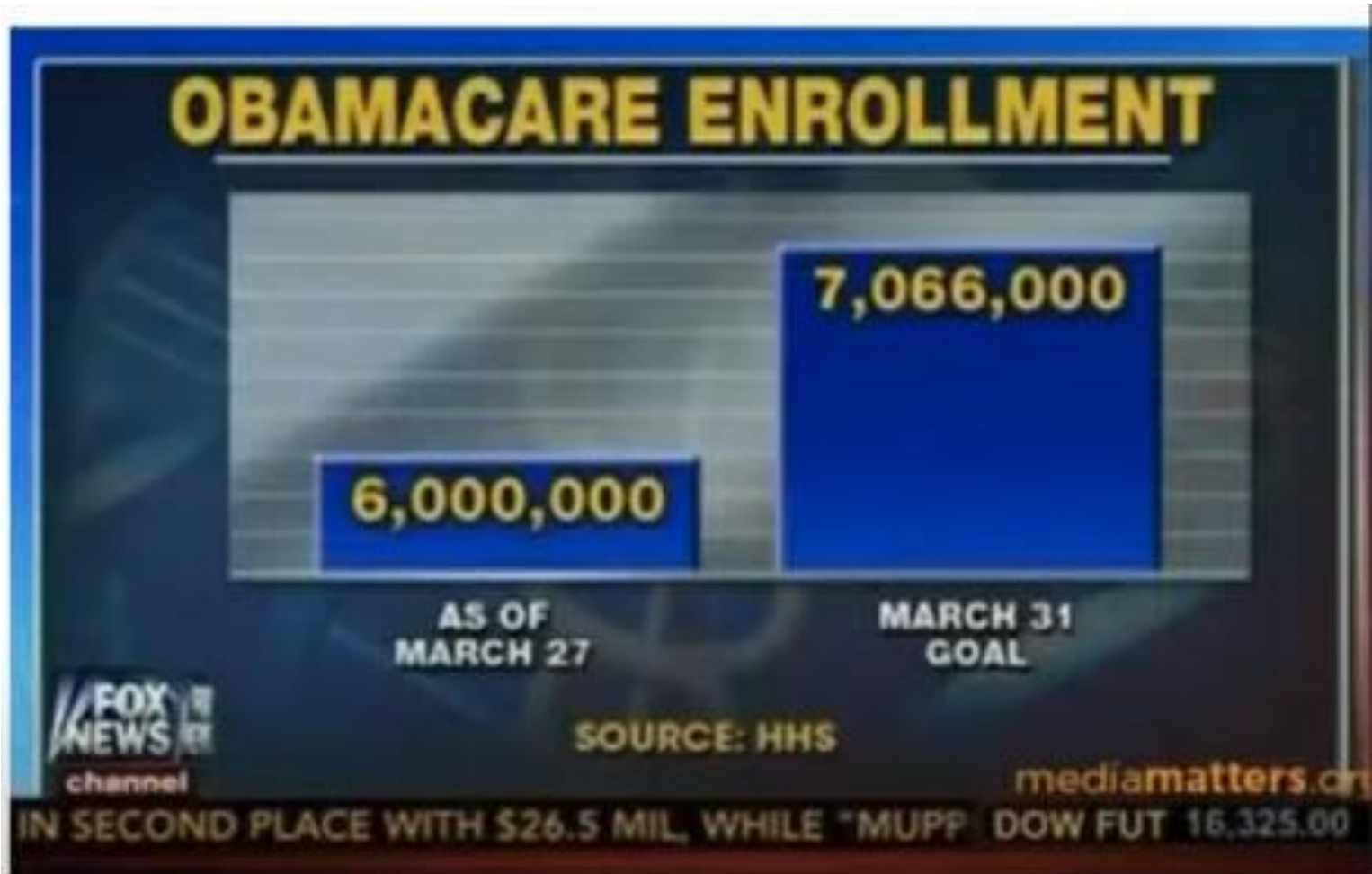
Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

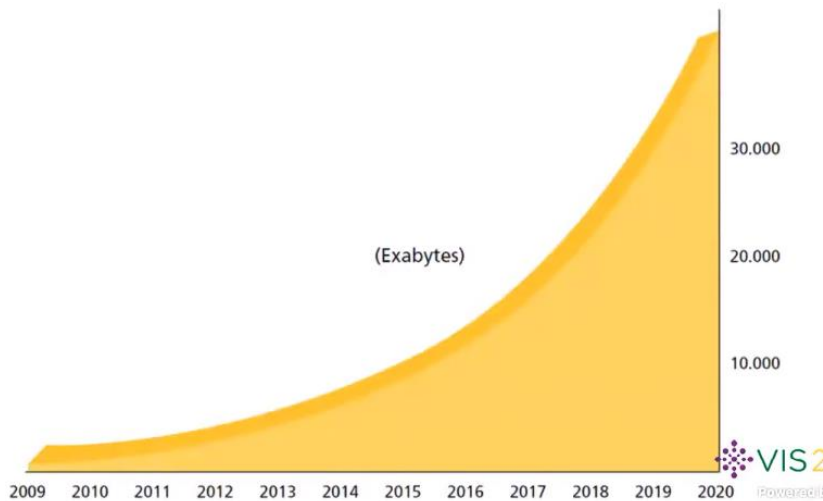
VISUALIZATION CAN BE DECEPTIVE



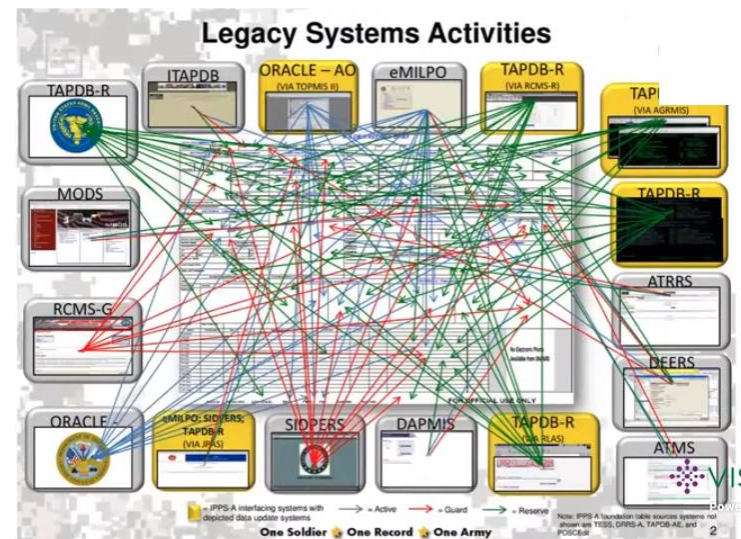
VISUALIZATION CAN BE BS*

From Michael Correll's alt.vis 2021 talk ([link](#))

- Don't relate to the real world
- Don't really help people understand their data
- Don't even have the decency to *lie* to you



“Stock footage chart”

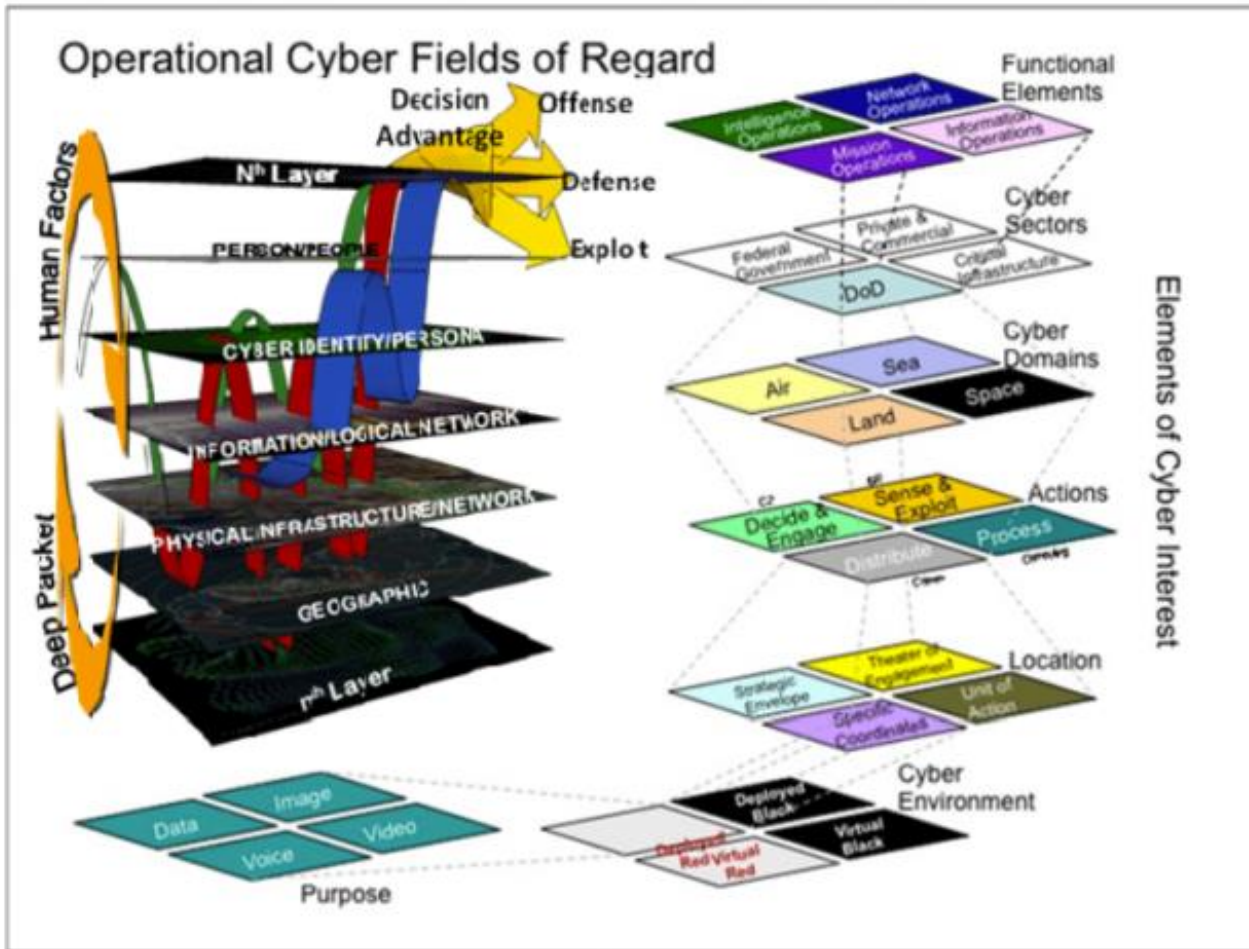


“Novocaine chart”

This stuff is way too complex for you to understand. Aren't you glad there's somebody smart like me taking care of it?

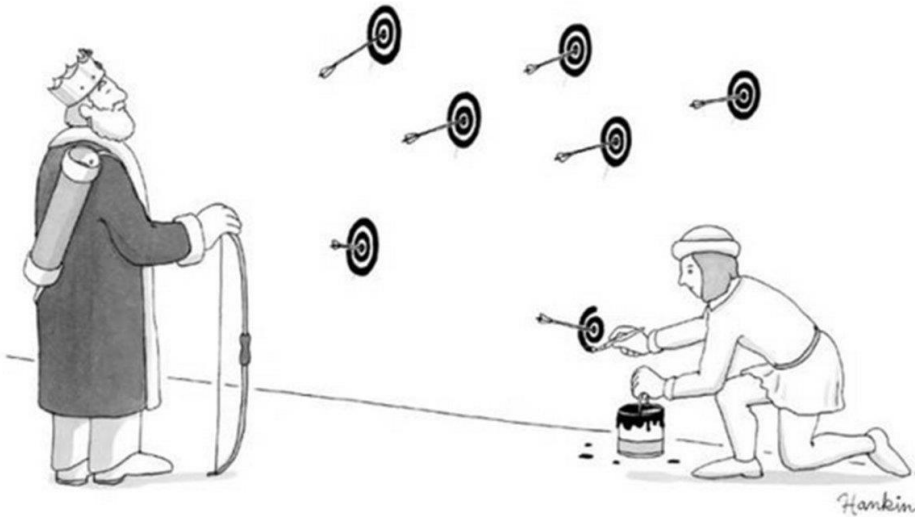
* Bullshit

VISUALIZATION CAN BE BS



Hwang's @DefenseCharts Twitter account, "dedicated to the presentational aesthetics of the defense-industrial complex"

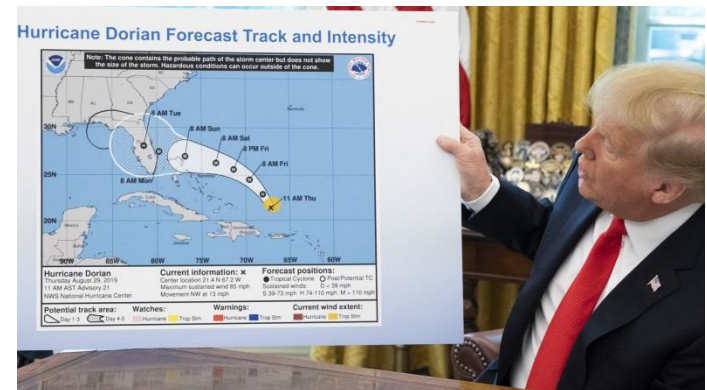
VISUALIZATION CAN BE BS



“Texas sharpshooter chart”



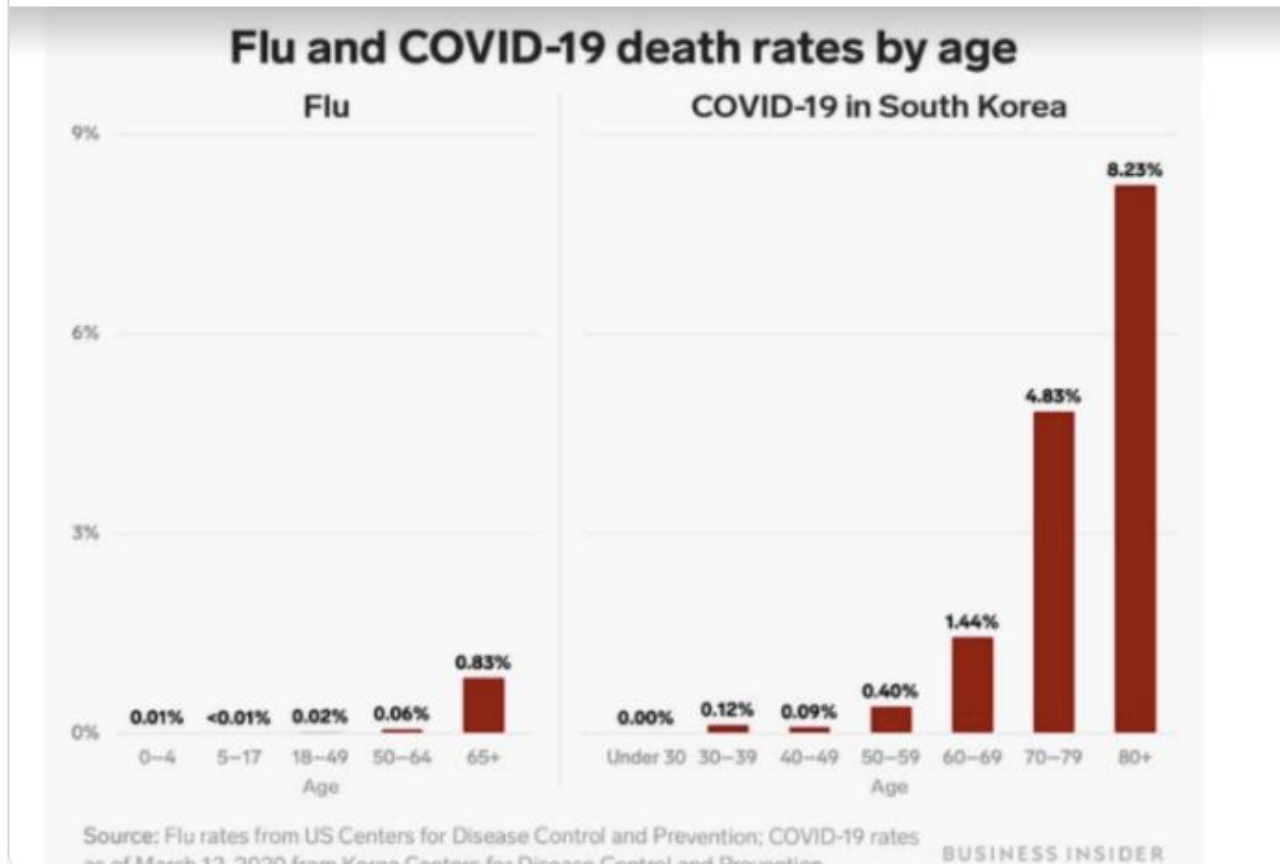
Example: Sharpigate



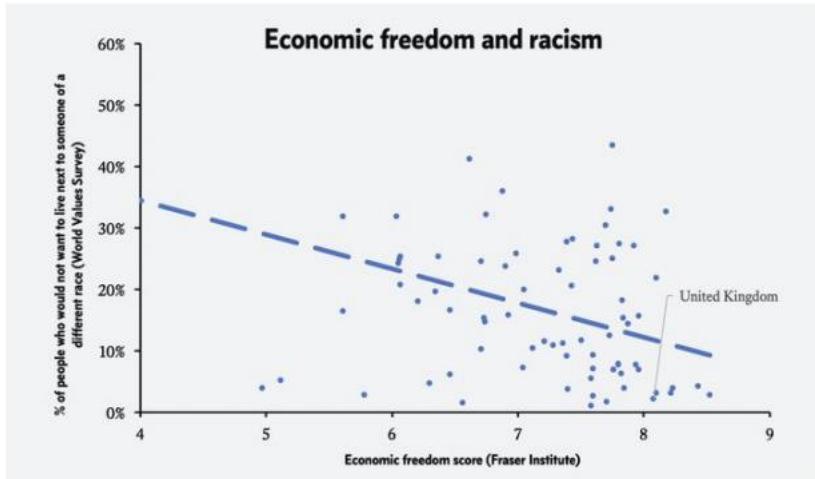
VISUALIZATION CAN BE BS

 **Ann Coulter**  @AnnCoulter

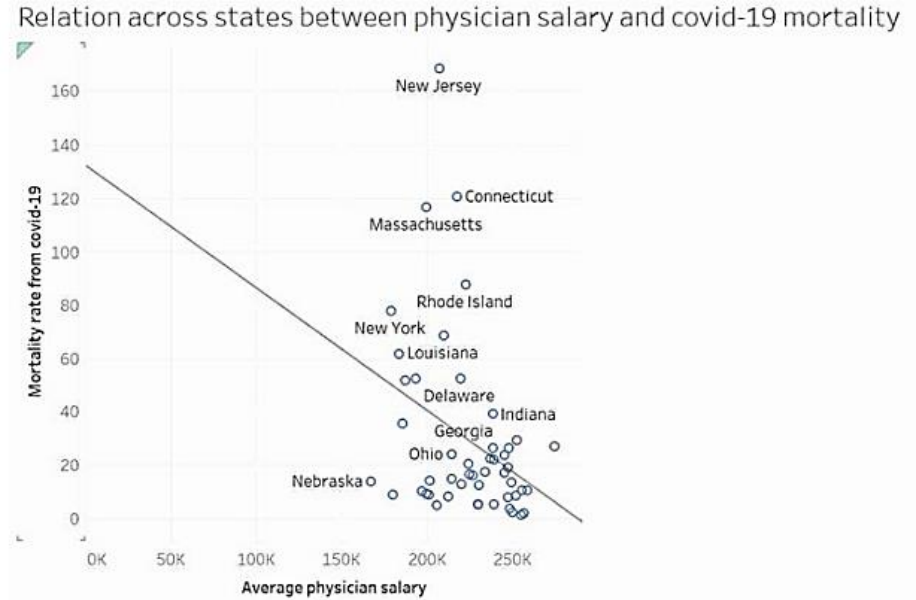
For people under 60, coronavirus is LESS dangerous than the seasonal flu:



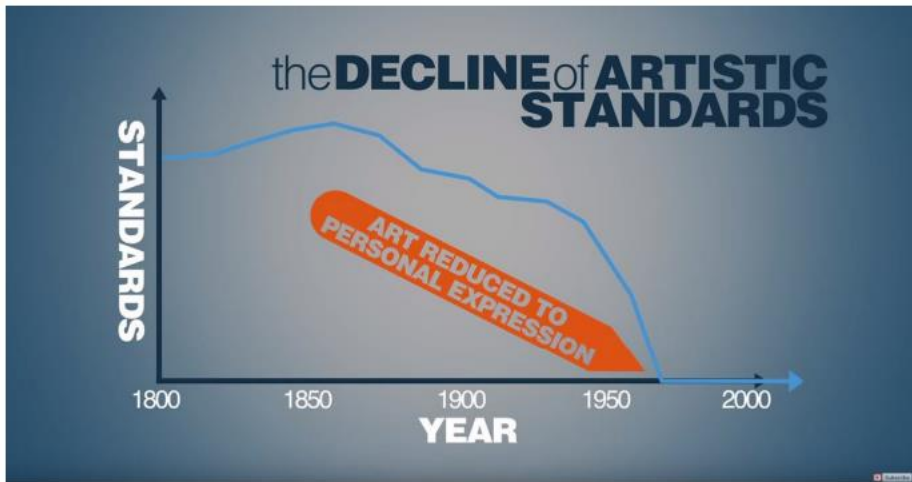
VISUALIZATION CAN BE BS



To show: “Countries with more economic freedom have less racist attitudes”



To show: “States where physicians are highly paid have lower COVID-19 mortality per capita”



Artificial noise added to make the chart look like there is a complex metric being measured precisely over time (when it is really not)

THE POWER OF THE VISUAL SYSTEM

The human visual system is not perfect, but it's extremely powerful

Vision is an integral part of life

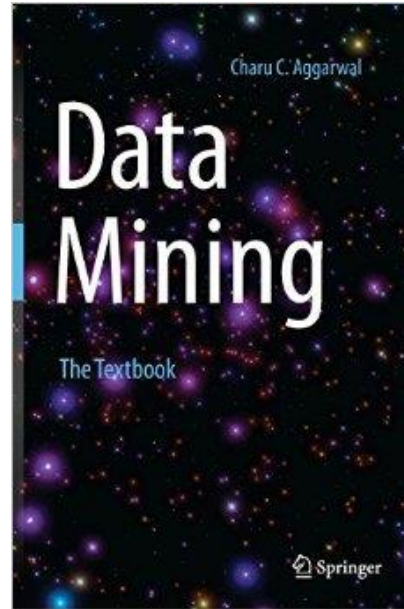
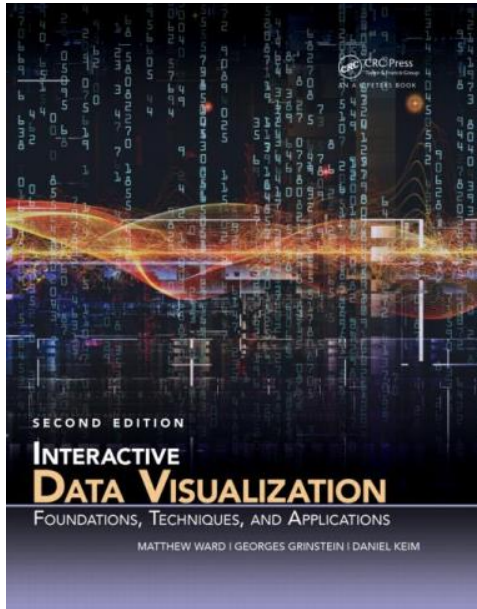
Vision is the gateway to higher-level regions of the brain

Exploit this fast and powerful processor for

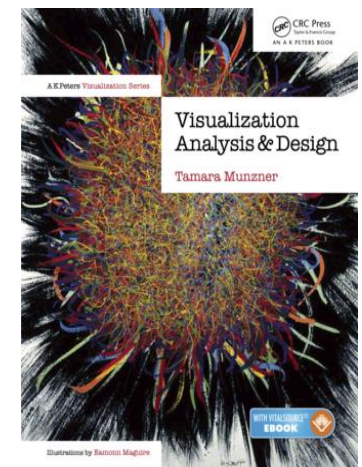
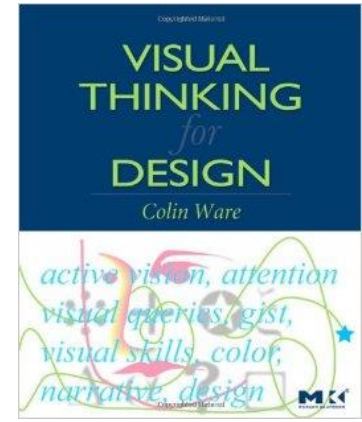
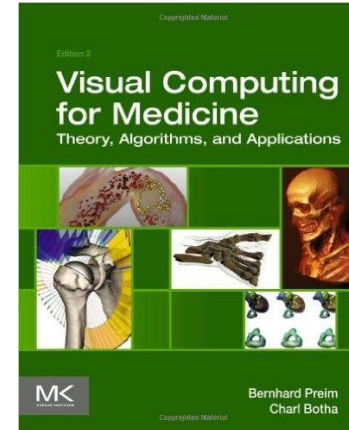
- complex data analyses, creative tasks, communicating ideas

→ The science of visualization and visual analytics

TEXT BOOKS



Required



Optional

TENTATIVE SCHEDULE

Lecture	Topic	Projects
1	Intro, schedule, and logistics	
2	Applications of visual analytics, data, and basic tasks	
3	Basic vis techniques for non-spatial data	Project 1 out
4	Data preparation and reduction	
5	Perception and cognition, visual design and aesthetics	
6	Foundations of statistics	
7	Introduction to D3 and Vega-Lite	Project 2 out
8	Data types, notion of similarity and distance	
9	Data mining techniques: clusters, text, patterns, classifiers	
10	Data mining techniques: clusters, text, patterns, classifiers	
11	High-dimensional data, dimensionality reduction	
12	Computer graphics and volume rendering	Project 3 out
13	Techniques to visualize spatial (3D) data	
14	Scientific and medical visualization	
15	Scientific and medical visualization	
16	Non-photorealistic rendering	
17	Midterm	
18	Principles of interaction	Project 4 out
19	Visual analytics and the visual sense making process	
20	Correlation and causal modeling	
21	Big data: data reduction, summarization	
22	Visualization of graphs and hierarchies	
23	Visualization of text data	Project 5 out
24	Visualization of time-varying and time-series data	
25	Memorable visualizations, visual embellishments	
26	Evaluation and user studies	
27	Narrative visualization and storytelling	
28	Data journalism	

GRADING

Midterm (1st part of the course): 30%

Final (2nd part of the course): 40%

Projects (5): 30%

choose wisely

- propose a dataset DS and argue why you think it's interesting (5%)
- code up a set of basic interactive visualizations for DS (5%)
- implement a set of suitable data analytics (python) for DS (5%)
- interlude: create some spatial visualizations using ImageVis3D (5%)
- create an interactive visual analytics dashboard for DS (10%)

Participation:

- expected to attend each lecture (attendance is not taken)

For late submission policy see [course website](#)

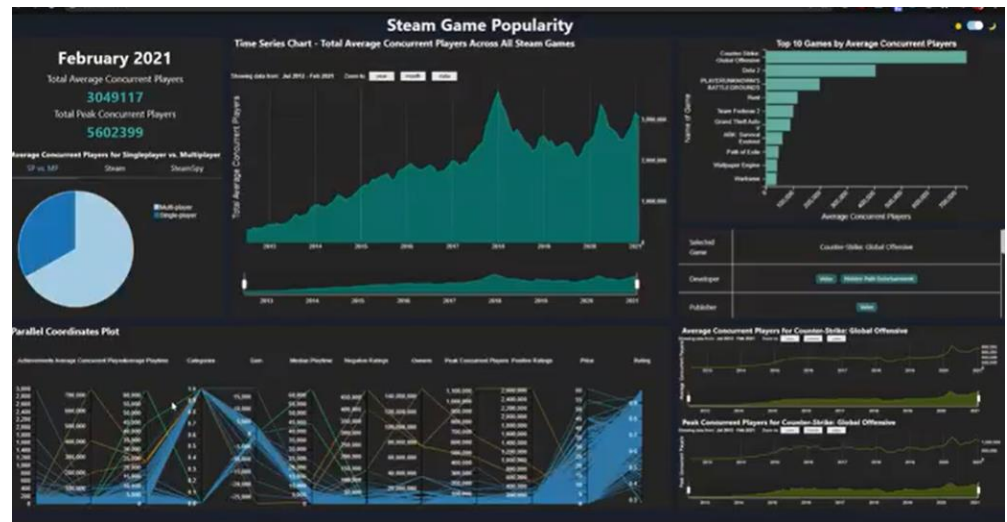
- course website will publish all course materials

WHAT'S A VA DASHBOARD?

See a really good example on [youtube](#)

Programmed with:

- python
- html
- JavaScript
- D3 API



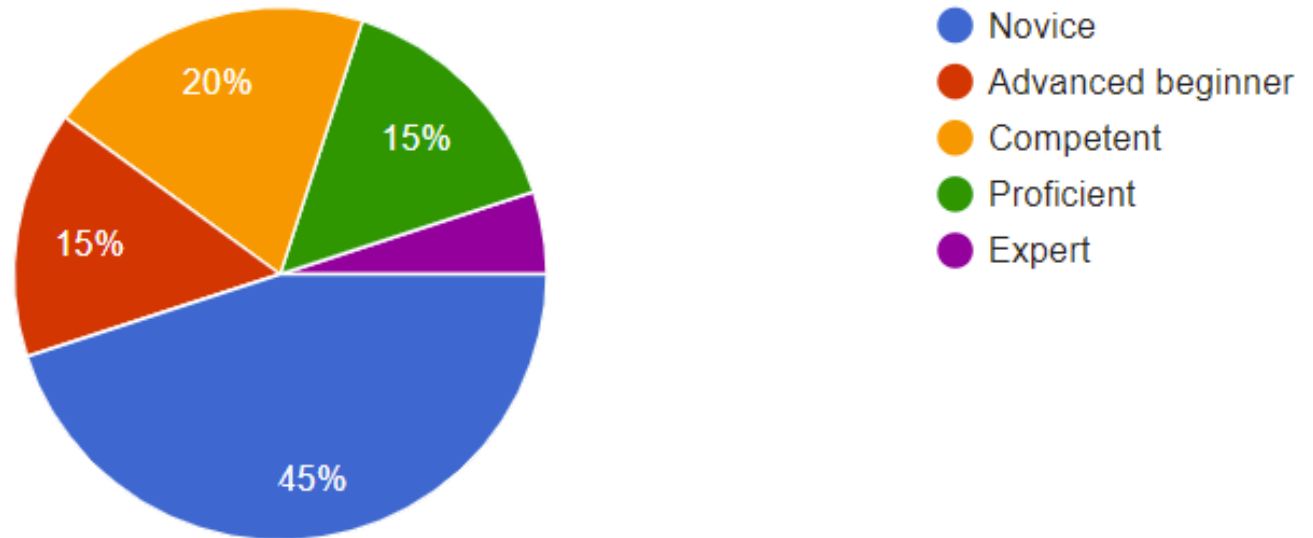
Your path to this:

- a dashboard is a collection of data visualizations linked together
- you will program all the individual dashboard components in lab 2 and lab 3
- then, in lab 5 you will put (some of) them all on one page and connect them in a meaningful way so they together can support users in interactive data analysis explorations

CAPABILITIES OF CSE/ISE 332 FALL 21

HTML programming

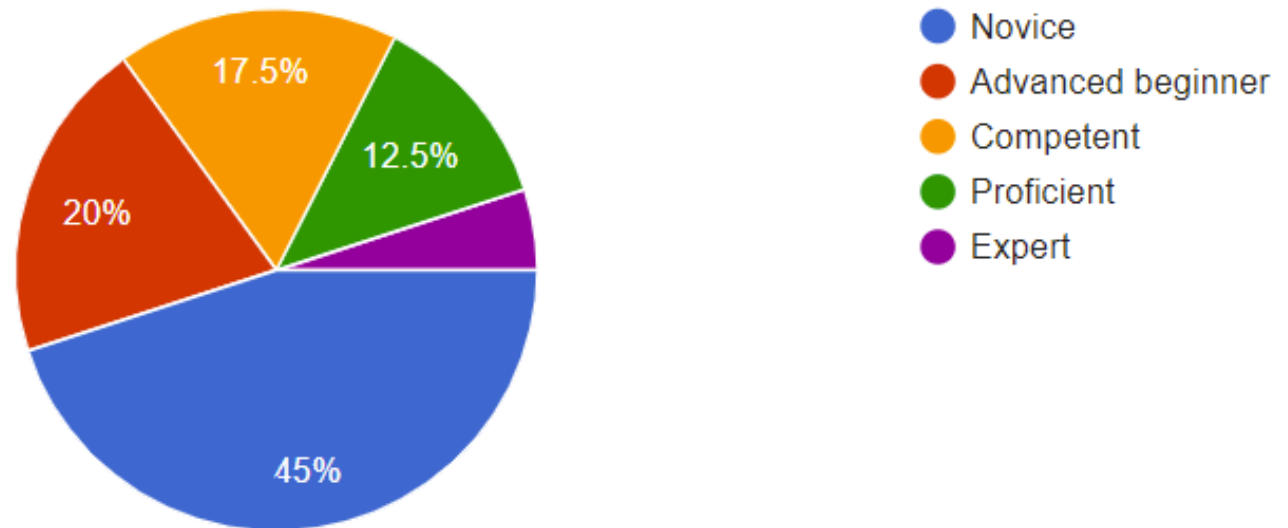
40 responses



CAPABILITIES OF CSE/ISE 332 FALL 21

JavaScript programming (note, JavaScript is not Java)

40 responses



READING ASSIGNMENT

You have 3 weeks to get up to speed with html and js

- 9/13 is the first programming assignment requiring it (3 weeks from now)

Fortunately there is a great and easy resource

- [W3schools html](#)
- [W3schools JavaScript](#)

HTML part, focus on:

- HTML Tutorial (specifically the sections *Home* to *Layout*)
- HTML Graphics
- will take you 2 days max

JavaScript part, focus on:

- JS Tutorial
- JS Objects, JS Functions, JS Async
- JS HTML DOM (Document Object Model)
- JS JSON (JavaScript Object Notation)
- will take you 2 weeks (one hour each day, ~15-20 hours total)

from [here](#)



TOP 10 Popular Programming Languages in 2020	
1	Python
2	JavaScript
3	Java
4	C#
5	C
6	C++
7	GO
8	R
9	Swift
10	PHP

WWW.NORTHEASTERN.EDU/GRADUATE

PROGRAMMING ASSIGNMENT

This will check your newly gained html and js capabilities

- successful completion will be evidence that you're ready for the lab assignments
- please do this yourself, copying it from somewhere will defeat the purpose of the exercise
- submit to Brightspace and gain 1% of extra credit (due 9/14)

Create a html page which supports the following

- allow a user to enter some text into a webpage text widget
- write a simple js program that will
 - count the number of non-white space characters
 - produce a list that shows many instances of each character are in the text (called a histogram, say a: 5, b: 4, c: 3,....)
 - calculate and show the total number of empty spaces and words
 - do some simple spellchecking, like too many instances of the same character in sequence (more than 2 is generally impossible in the English language) and either highlight them in the text or write our an error message

CODE EDITOR

Several free code development environments are available

- [Visual Studio Code](#) (recommended)
- [Atom](#)
- [Sublime Text](#)

Browsers to run and develop your code

- Chrome
- Firefox
- IE and Edge are not overly suitable
- Chrome and Firefox also have panels where you can see and edit your code
- comes in handy when you want to change values of variables