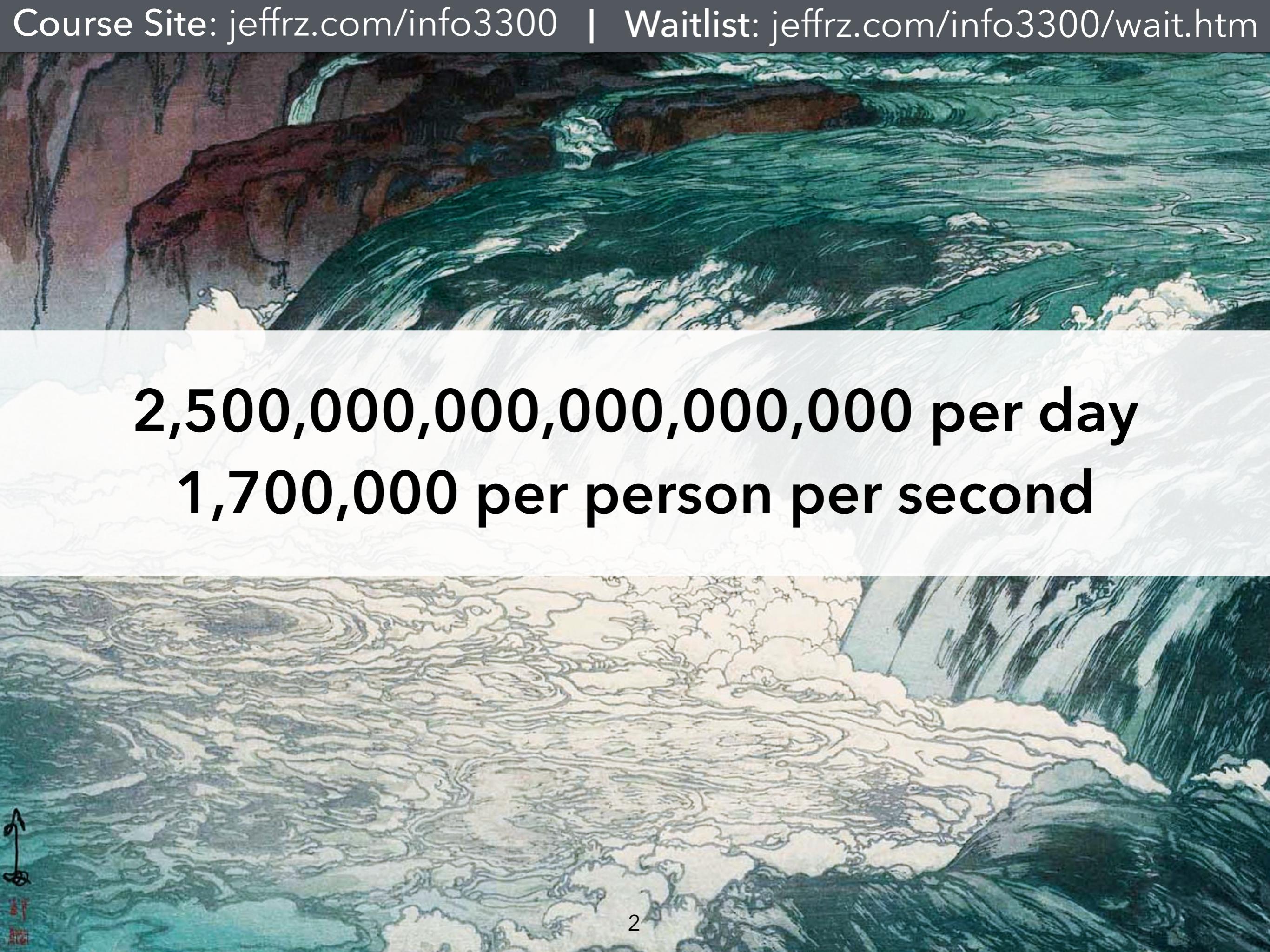


INFO / CS 3300

INFO 5100

Jeff Rzeszotarski
Assistant Prof, InfoSci





**2,500,000,000,000,000 per day
1,700,000 per person per second**

- The daily New York Times now contains more information than the average 17th century person would have encountered *in their lifetime*
- More information has been produced in the last 30 years than in the rest of Earth's known history
- In 2016 office workers get an average of 121 emails per day, while in 1900 1 in 10 people couldn't read

7



4

- Tasks to do; emails to send
- Restaurants to compare
- Ongoing conversations
- Items to put on this slide

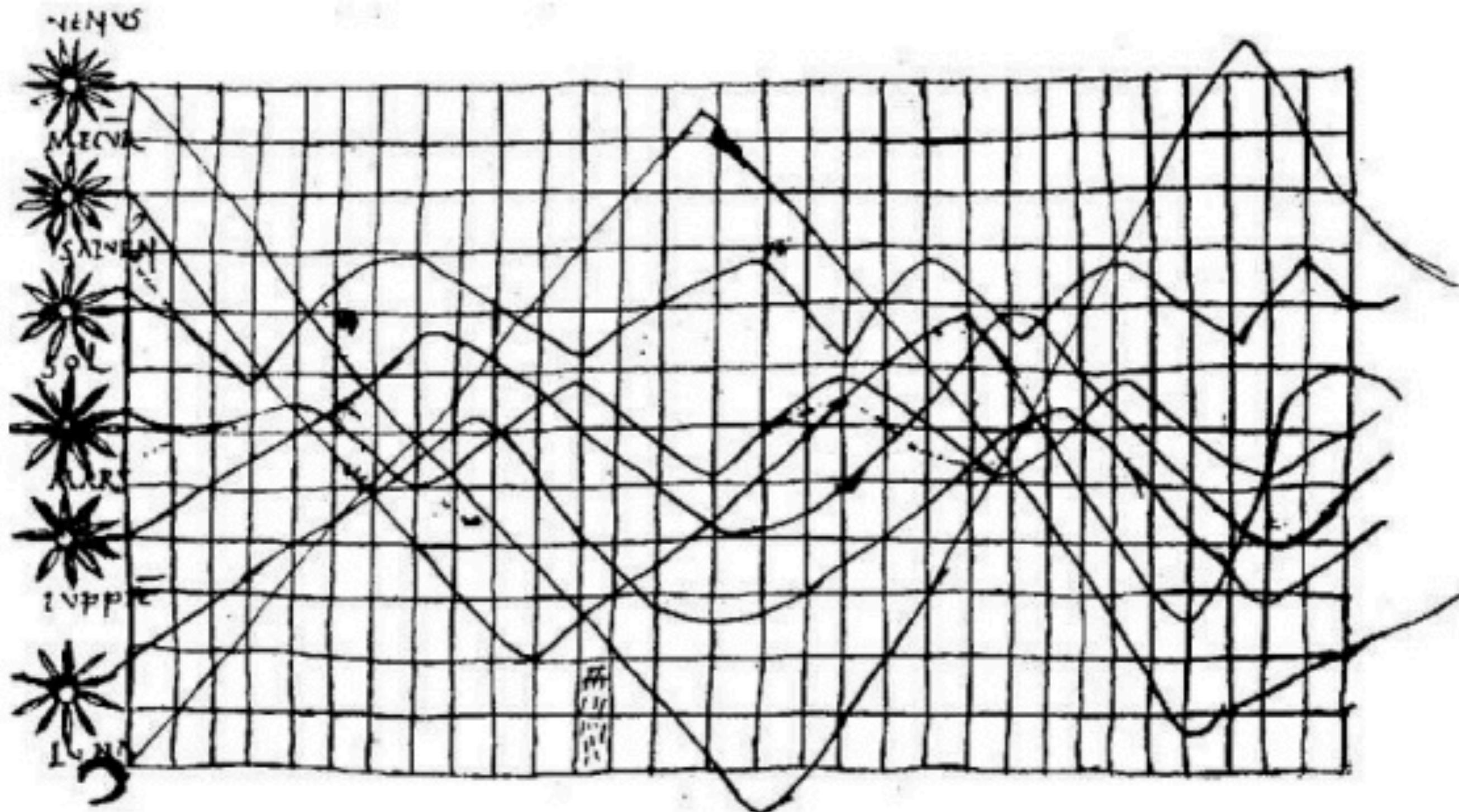


“We are drowning in information
and starving for knowledge.”

- Rutherford D. Rogers
1915

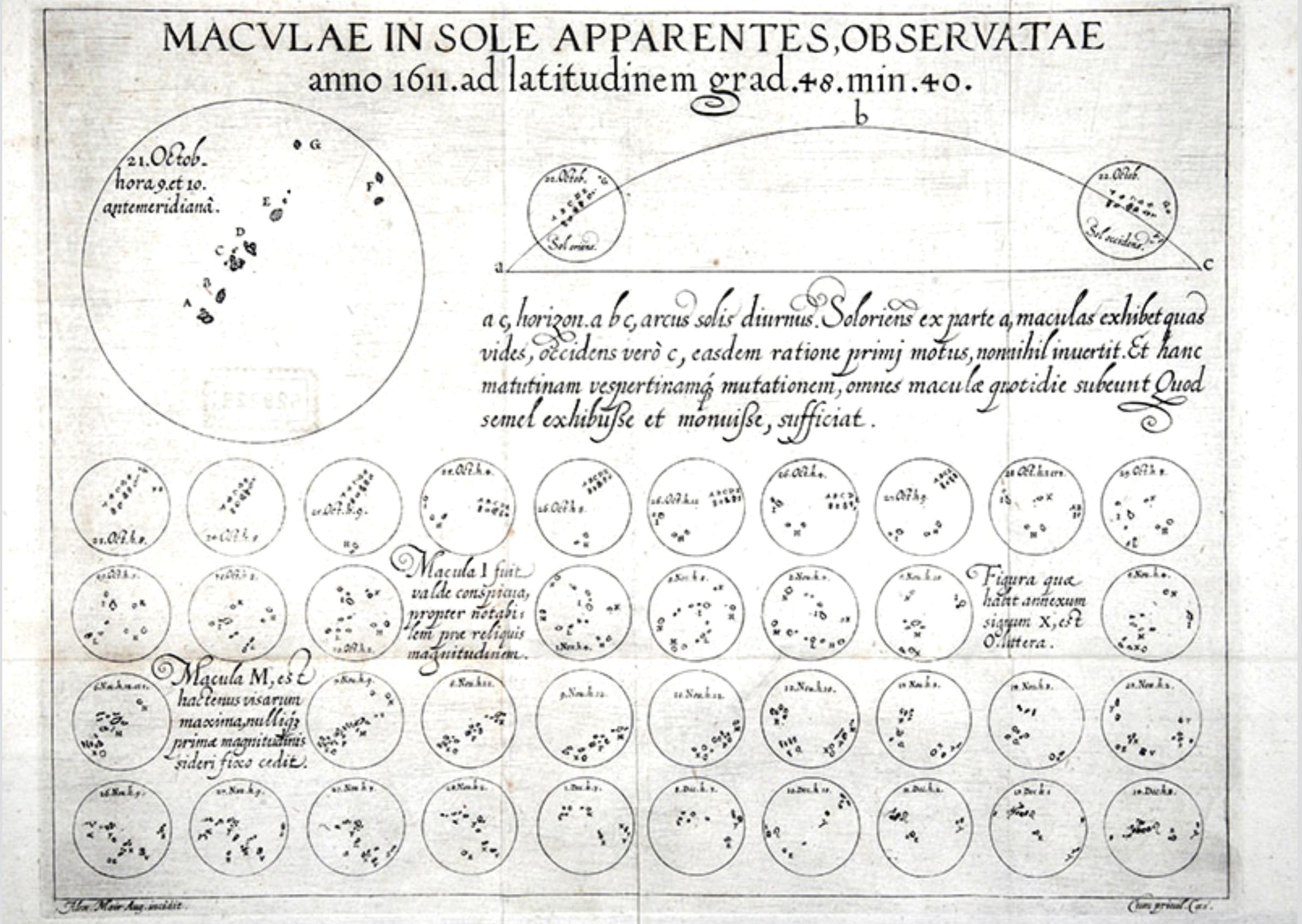




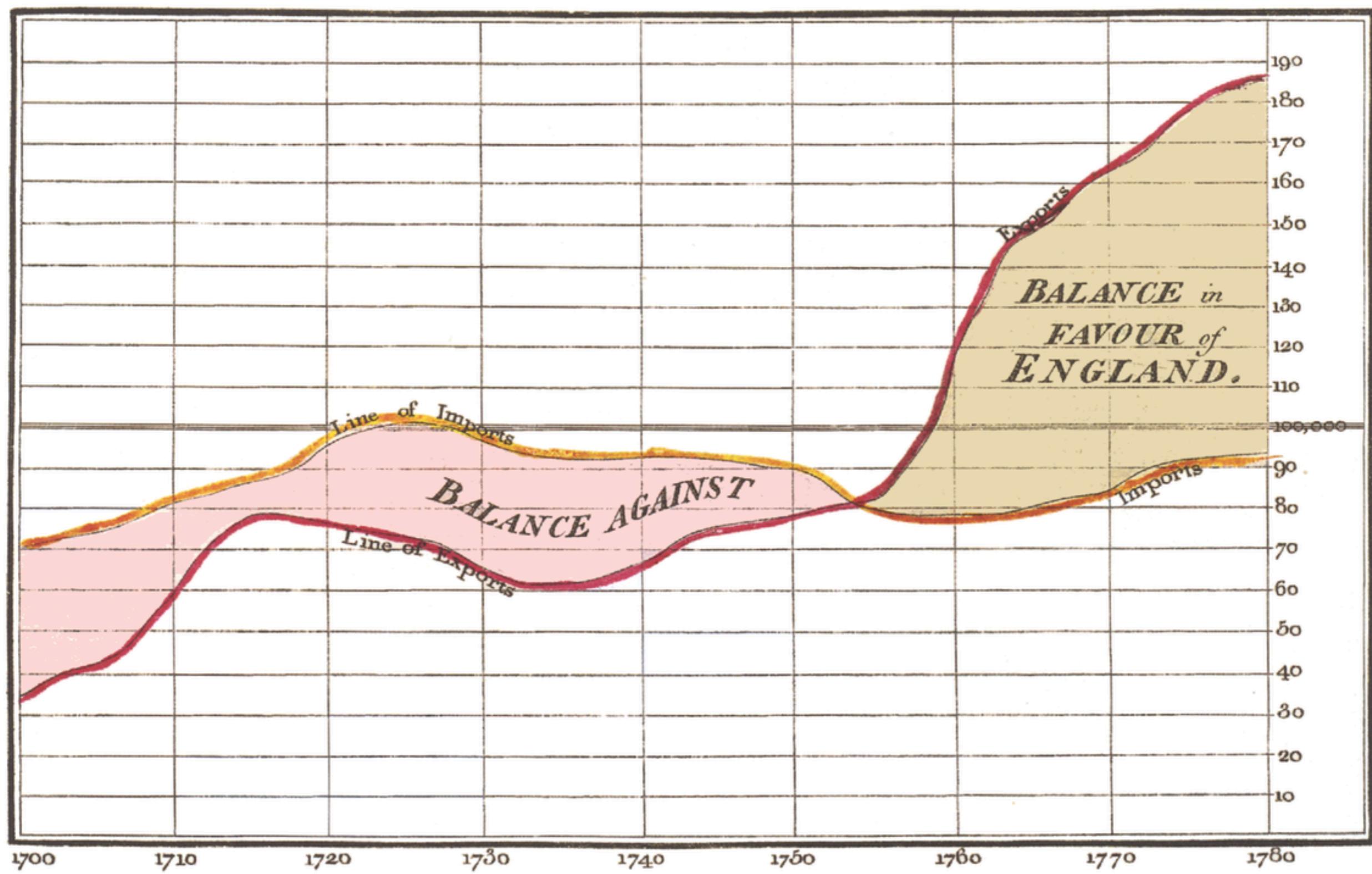


Img credit
Funkhouser, 1936





Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.



The Bottom line is divided into Years, the Right hand line into £10,000 each.

Published as the Act directs, 1st May 1786. by W^m Playfair

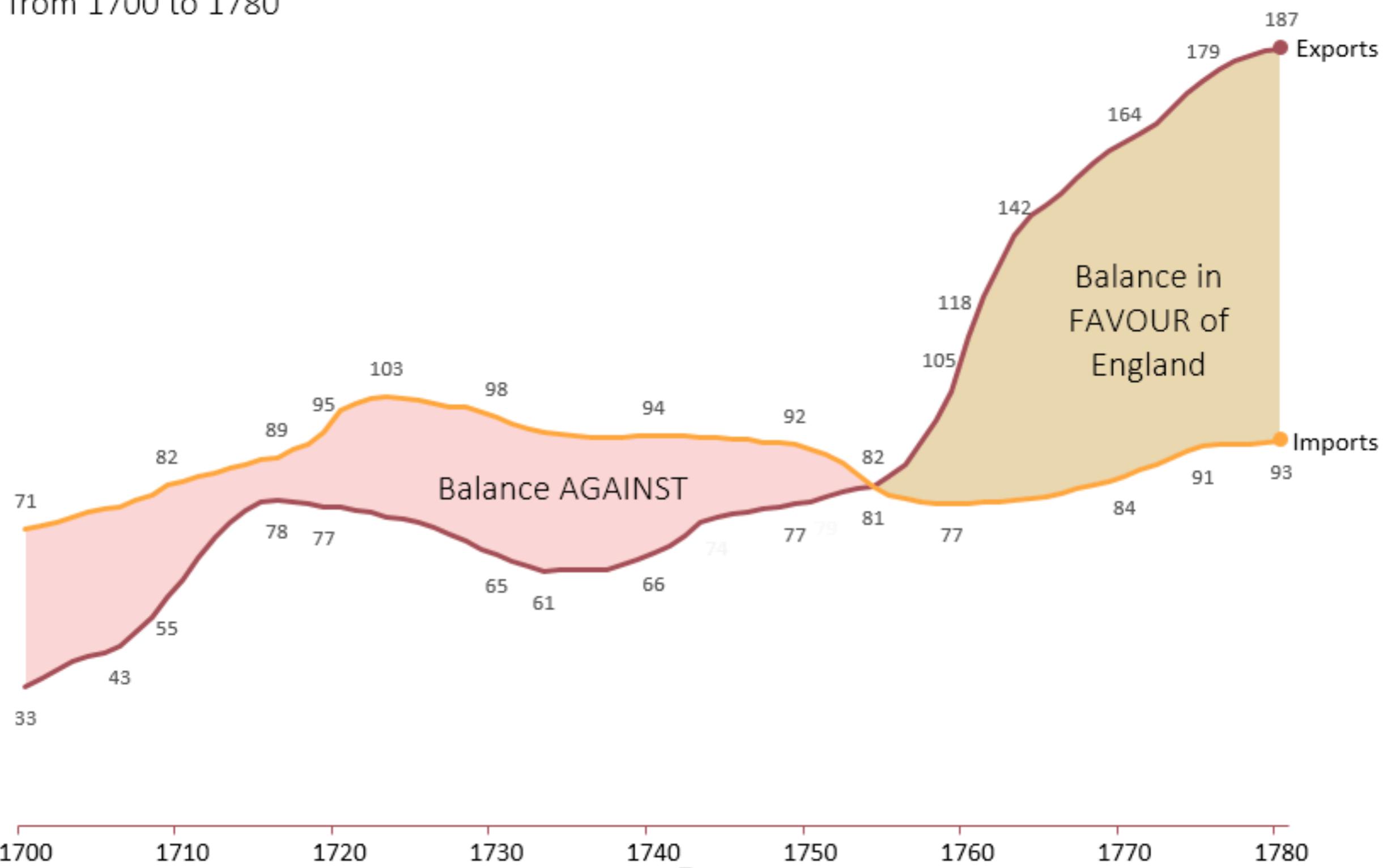
Neale sculpt^r 352, Strand, London.

Originally created by William Playfair



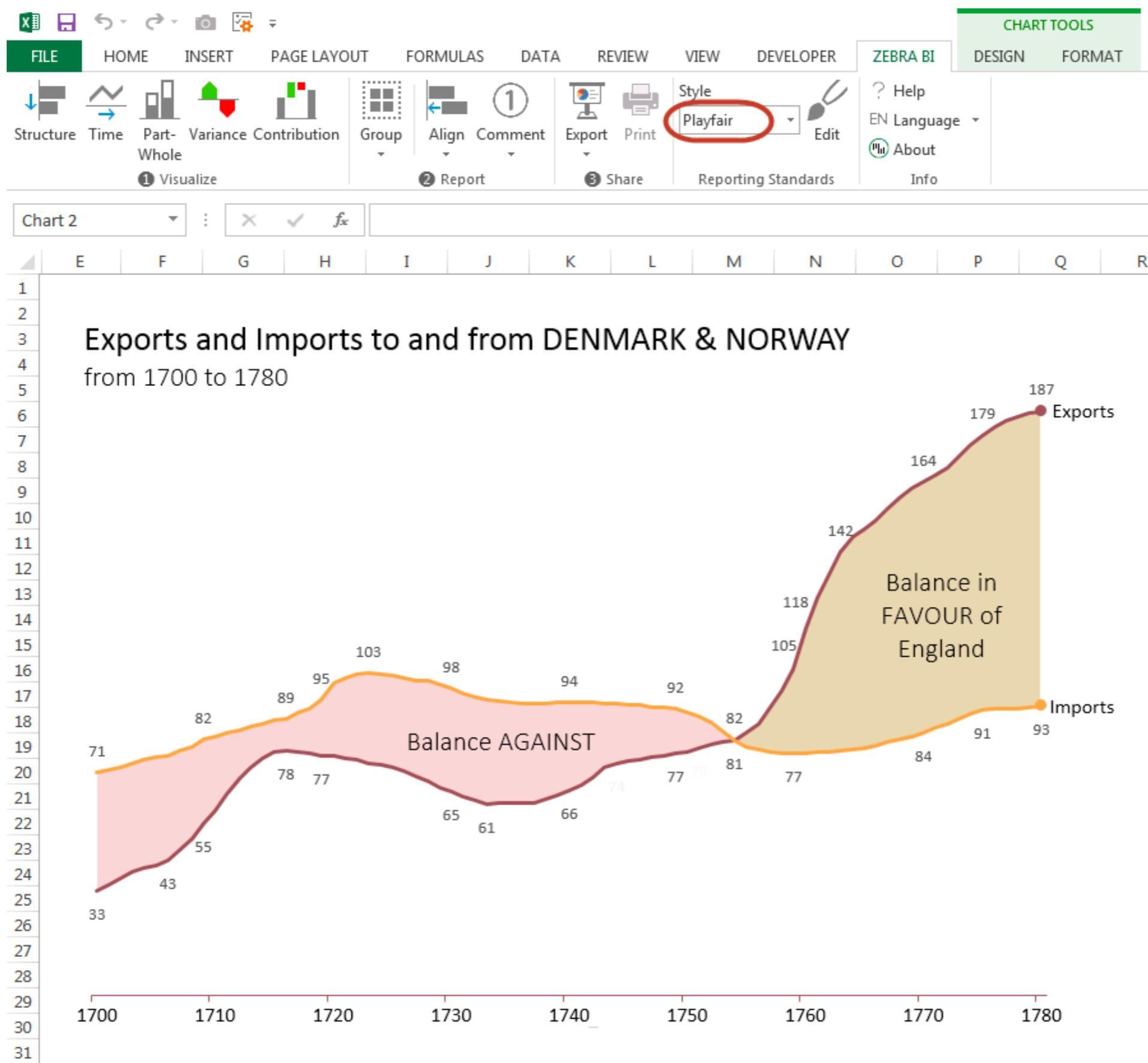
Exports and Imports to and from DENMARK & NORWAY

from 1700 to 1780



Created by Zebra BI





Created by Zebra BI



What makes these effective or ineffective data visualizations?

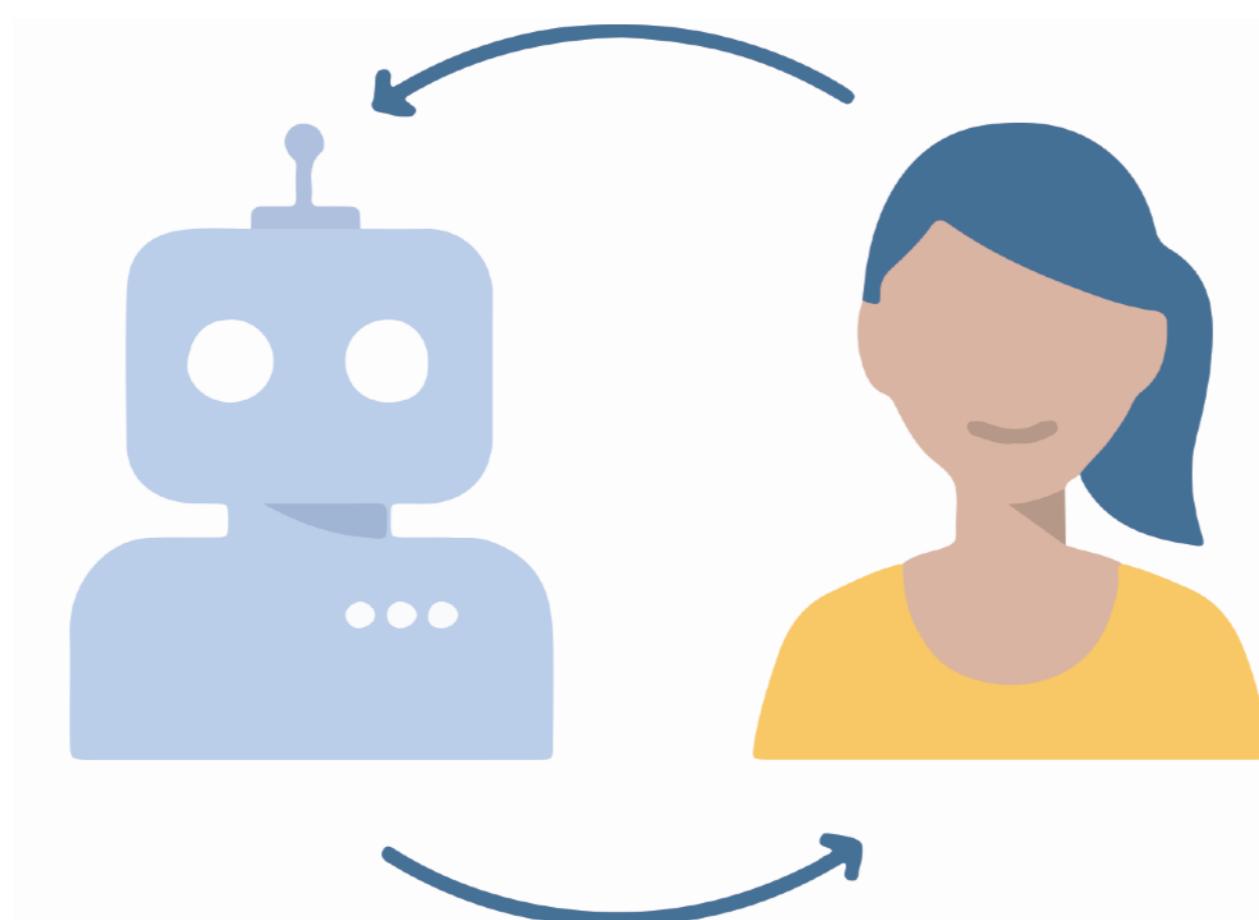
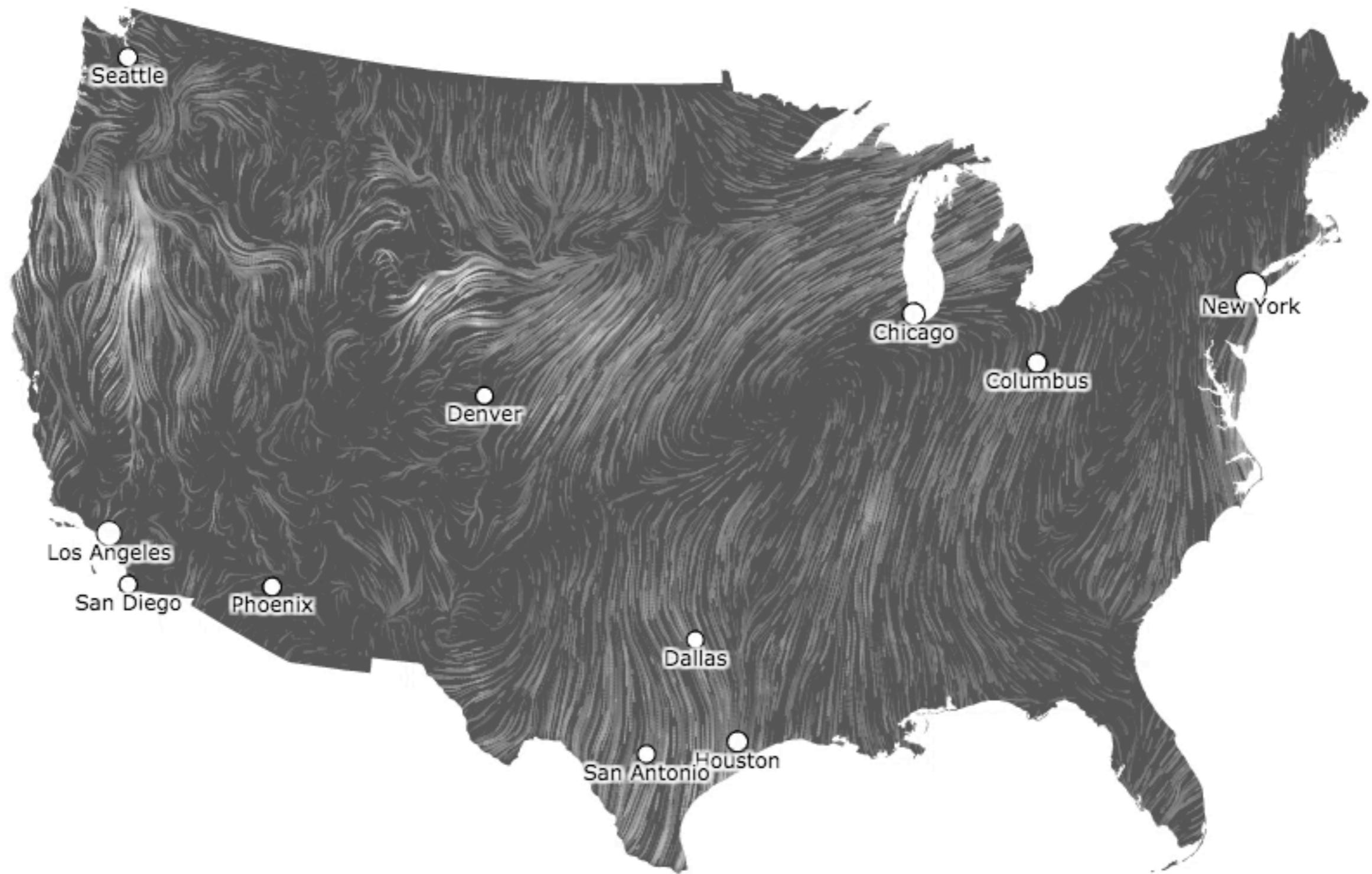


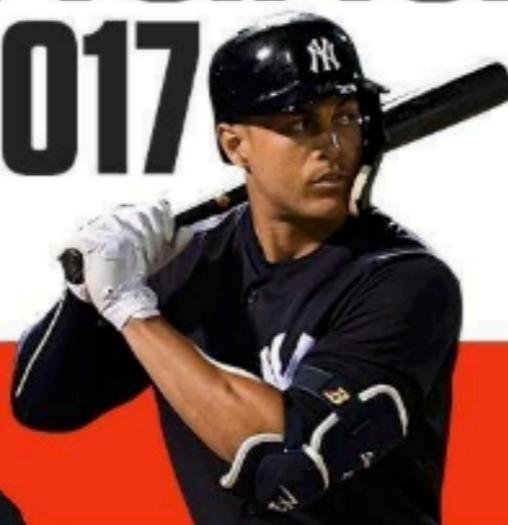
Image credit: CrowdFlower



2018 SALARY FOR TOP HR HITTERS IN 2017

GIANCARLO STANTON

HOME RUNS: 59



\$25M

AARON JUDGE

HOME RUNS: 52



\$622,300

J.D. MARTINEZ

HOME RUNS: 45



\$23.75M

KHRIS DAVIS

HOME RUNS: 43

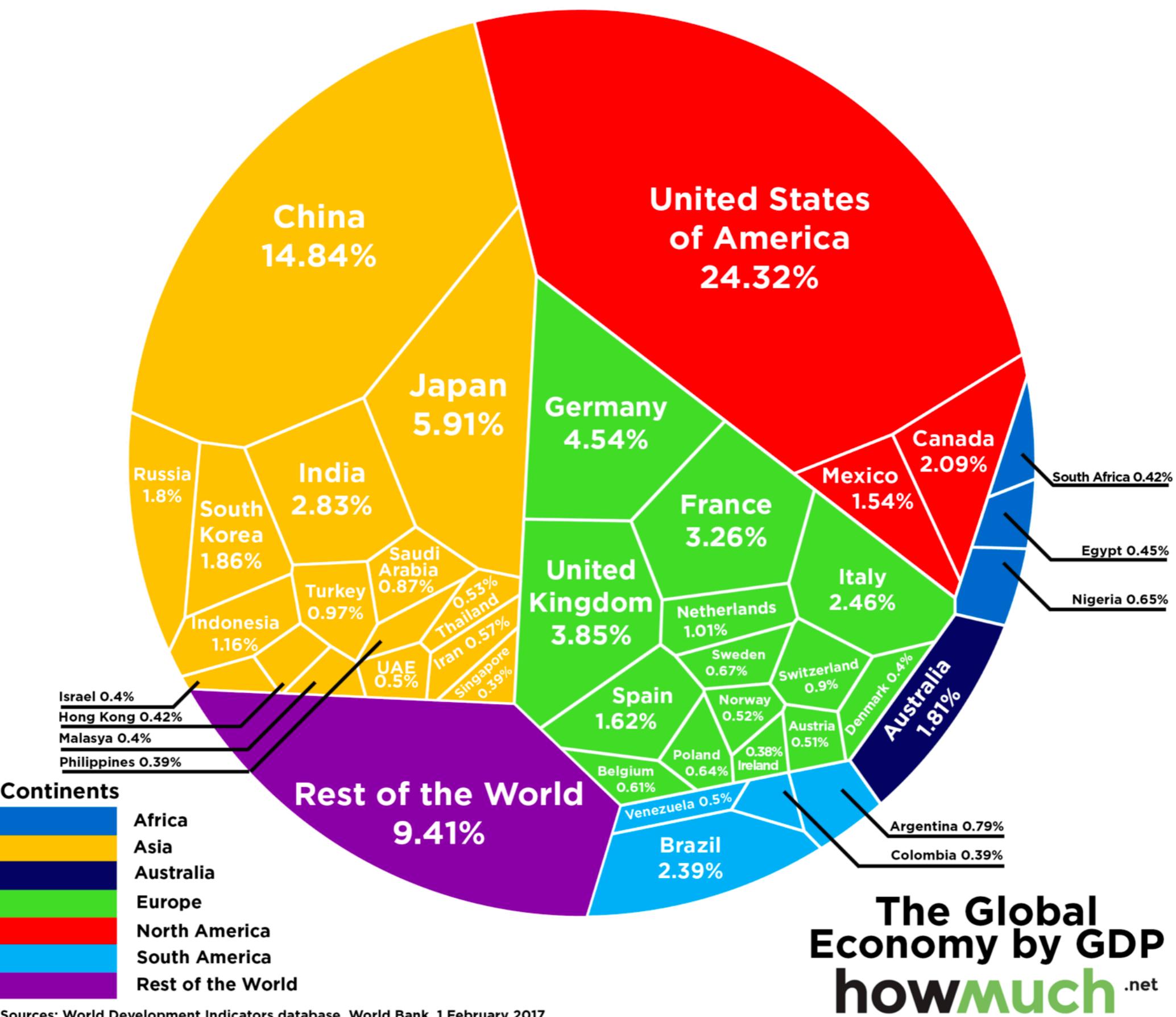


\$10.5M

A word cloud centered around the Star Wars universe, featuring words like Federation, Jedi, Naboo, and Chancellor.

The word cloud includes the following words:

- doorway, Captain, moves, move, runs, Mom, continue
- appears, swamp, Supreme, Force, main, right, think, take, large, sure, continues
- Majesty, beeps, know, Federation, hanger, begin, smiles
- Trade, starts, Coruscant, like, Federation, hanger, enters, Skywalker
- droids, You, a, going, two, onto, sees, Jar, stand, Republic, Qui-Gon
- puts, mesa, followed, battle, can, last, young, toward, help, walk, city
- group, followed, battle, side, can, Jar, stand, Republic, Qui-Gon
- treaty, GROUP, SITH, control, side, can, Jar, stand, Republic, Qui-Gon
- whistles, One, stop, get, ship, turns, start, cockpit, ramp
- spacecraft, sits, side, can, Jar, stand, Republic, Qui-Gon
- Wesa, lands, Mesa, goes, door, life, sir, see, Jar, people, place
- fight, left, stands, boy, one, little, come, shield, cruiser
- Sith, rush, shop, tries, distance, bridge, screen, walks, room, just, far, now, behind, hallway, stay
- begins, good, fire, Yes, eyes, look, Naboo, way, small, takes, several
- lets, Ani, great, energy, next, makes
- shop, tries, distance, bridge, screen, walks, room, just, far, now, behind, hallway, stay
- Senator, Lord, inside, want, gets, flies, gives, planet, run, Jedi, must, need, Now, center
- make, jumps, sword, race, Highness, Chancellor, power, engines, Sebulba
- Anakin's, Pod, hologram, seen, Anakin, box, much, MAIN, end, droid, parts, things, view





Photographed by Jeff Rzeszotarski

Ibghy and Lemmens, The Prophets, 2013-5



There is no guarantee that a sophisticated data visualization will be an effective visualization.

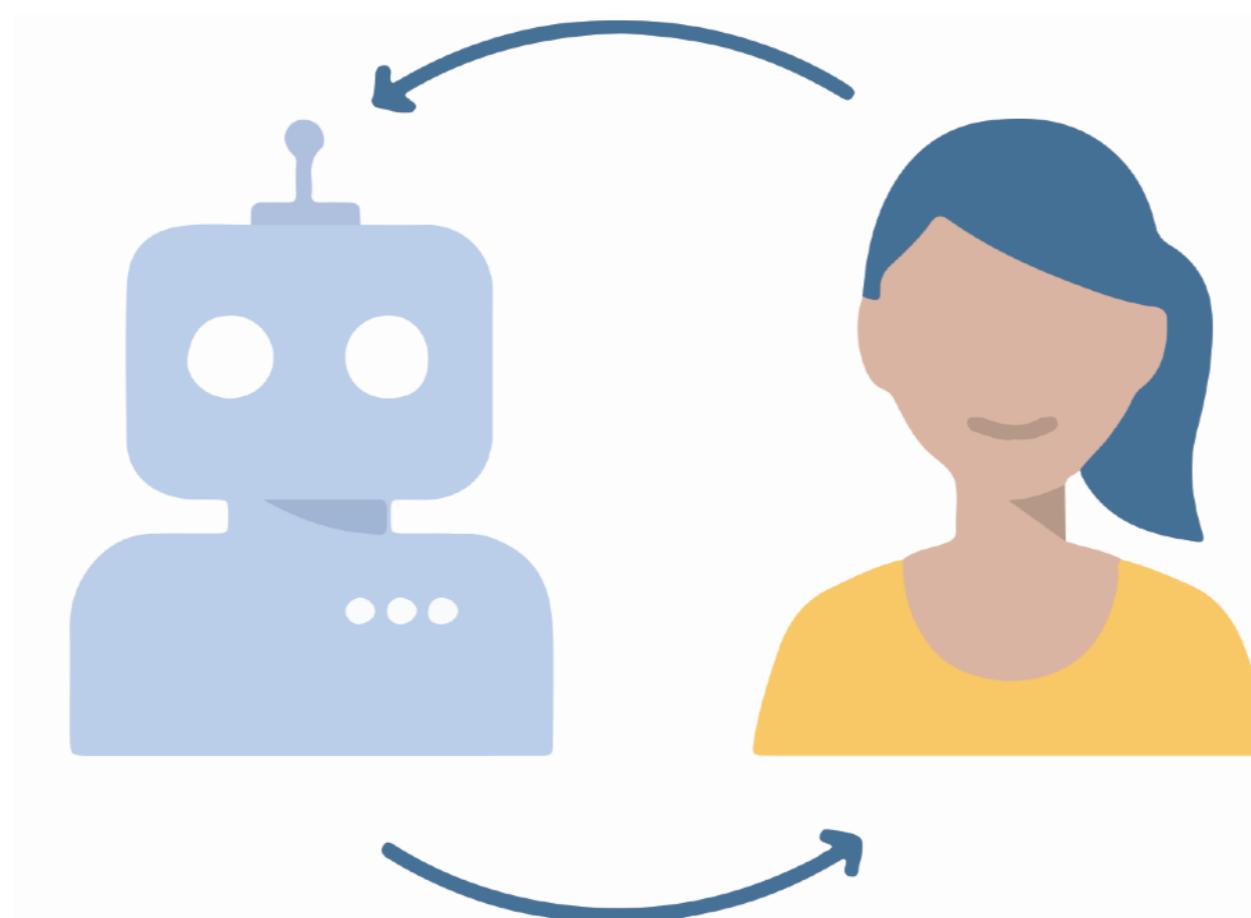


Image credit: CrowdFlower

*Showing
data*

*Visualizing
data*

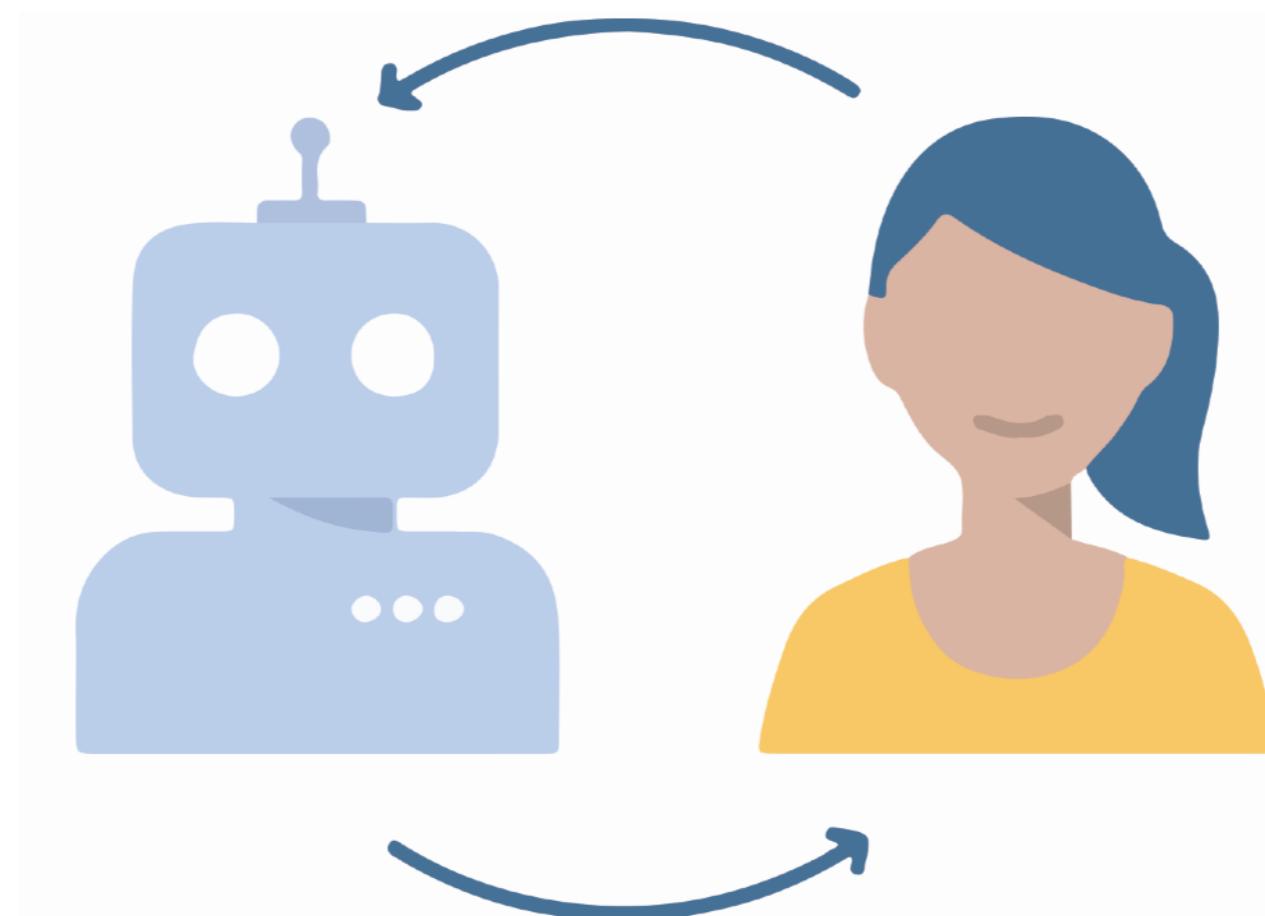


Image credit: CrowdFlower

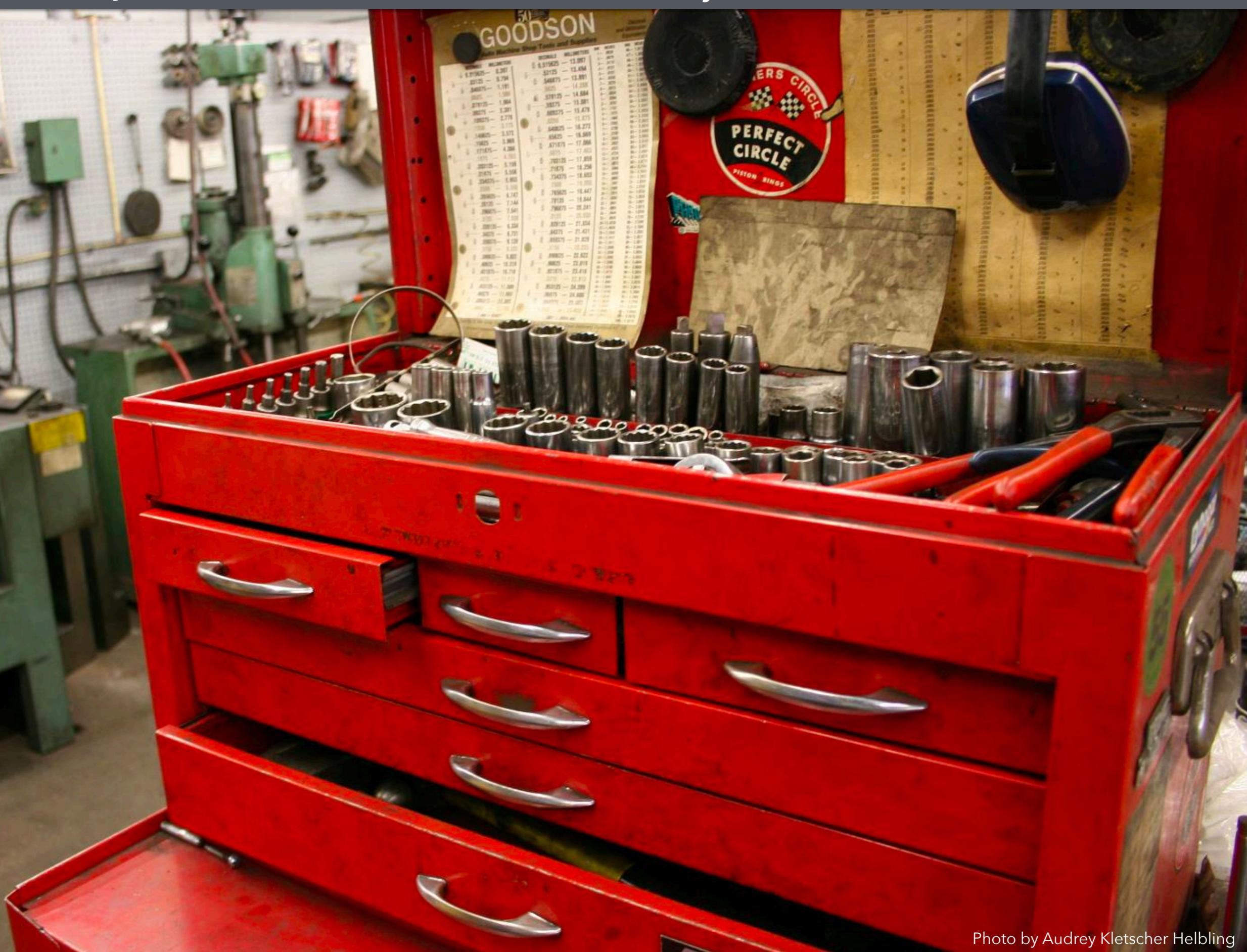
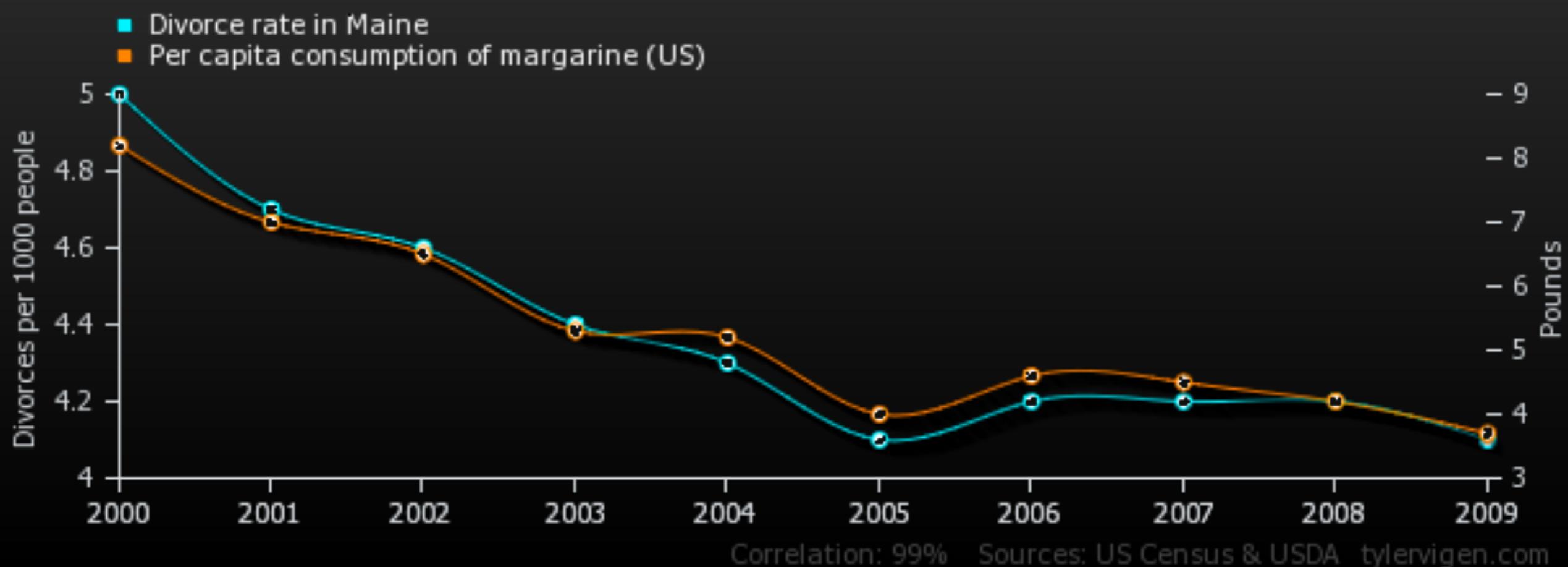
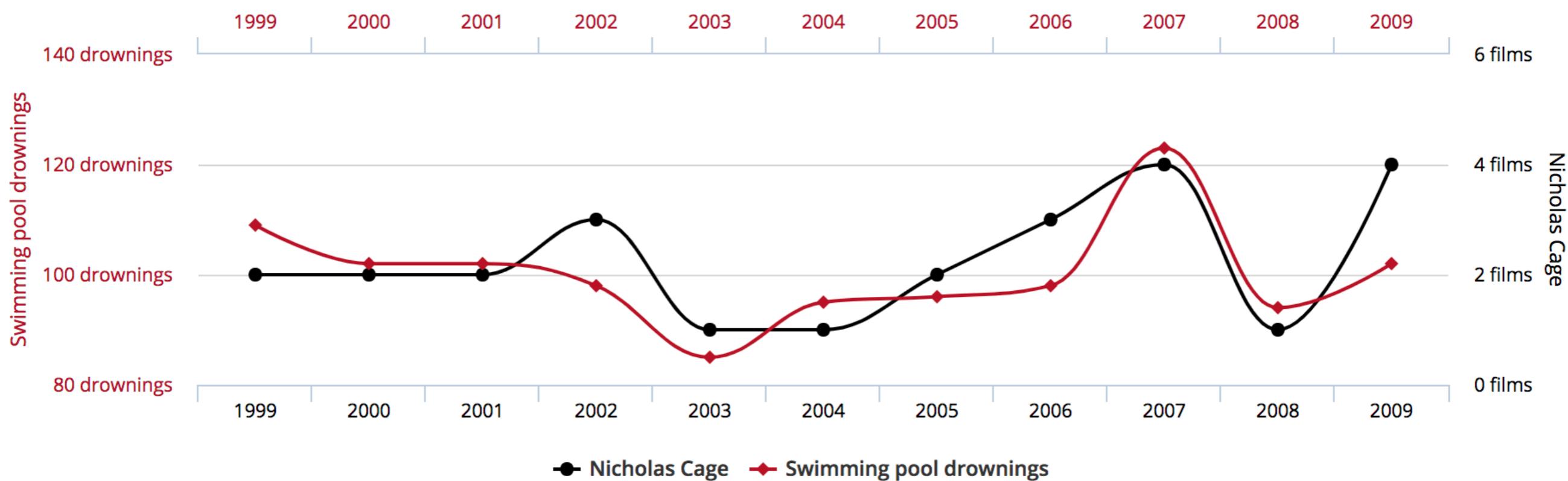


Photo by Audrey Kletscher Helbling



Number of people who drowned by falling into a pool correlates with Films Nicolas Cage appeared in

Correlation: 66.6% ($r=0.666004$)



Data sources: Centers for Disease Control & Prevention and Internet Movie Database

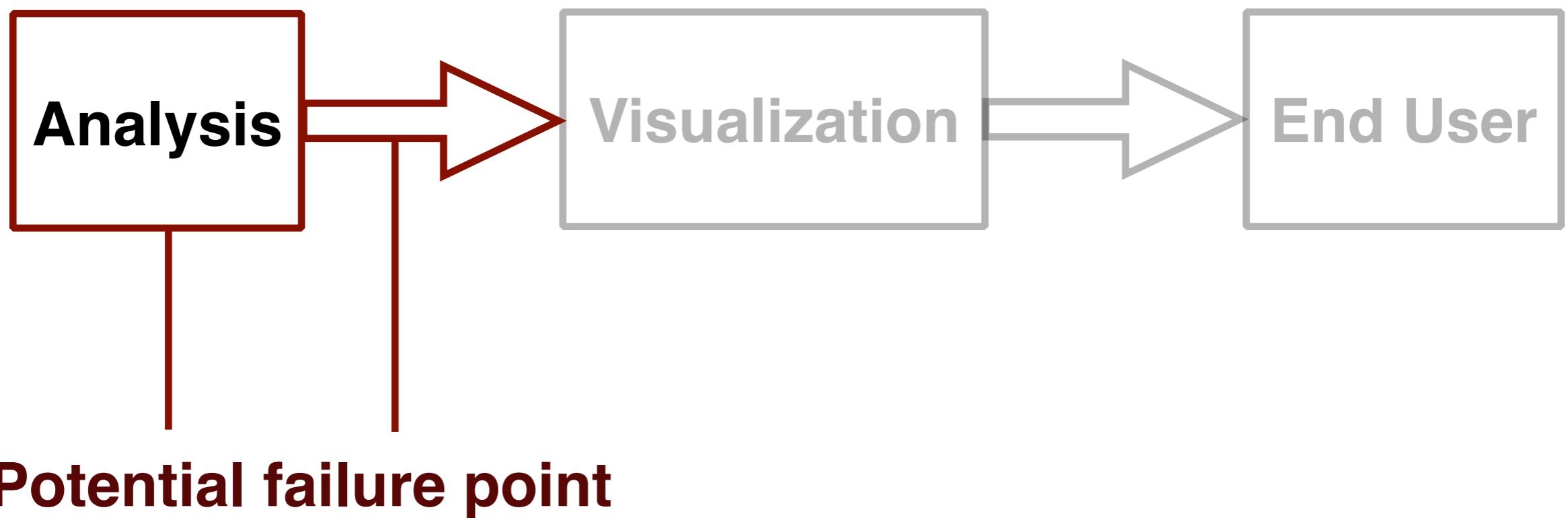
tylervigen.com



Datavis Process Model

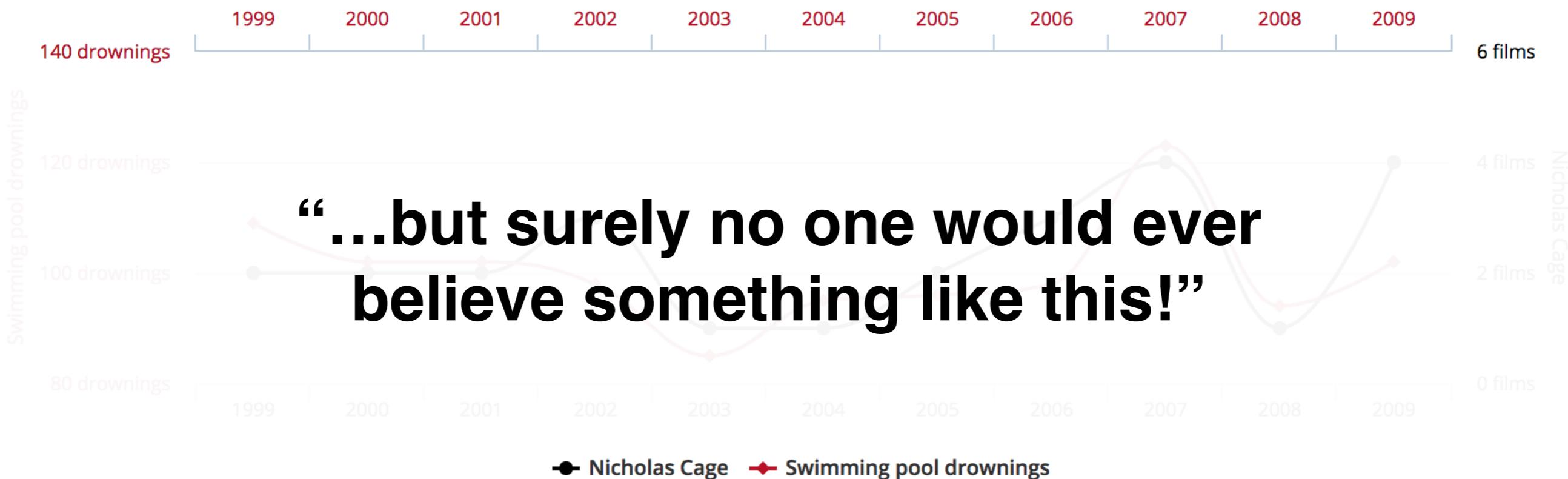


Datavis Process Model



Number of people who drowned by falling into a pool correlates with Films Nicolas Cage appeared in

Correlation: 66.6% ($r=0.666004$)



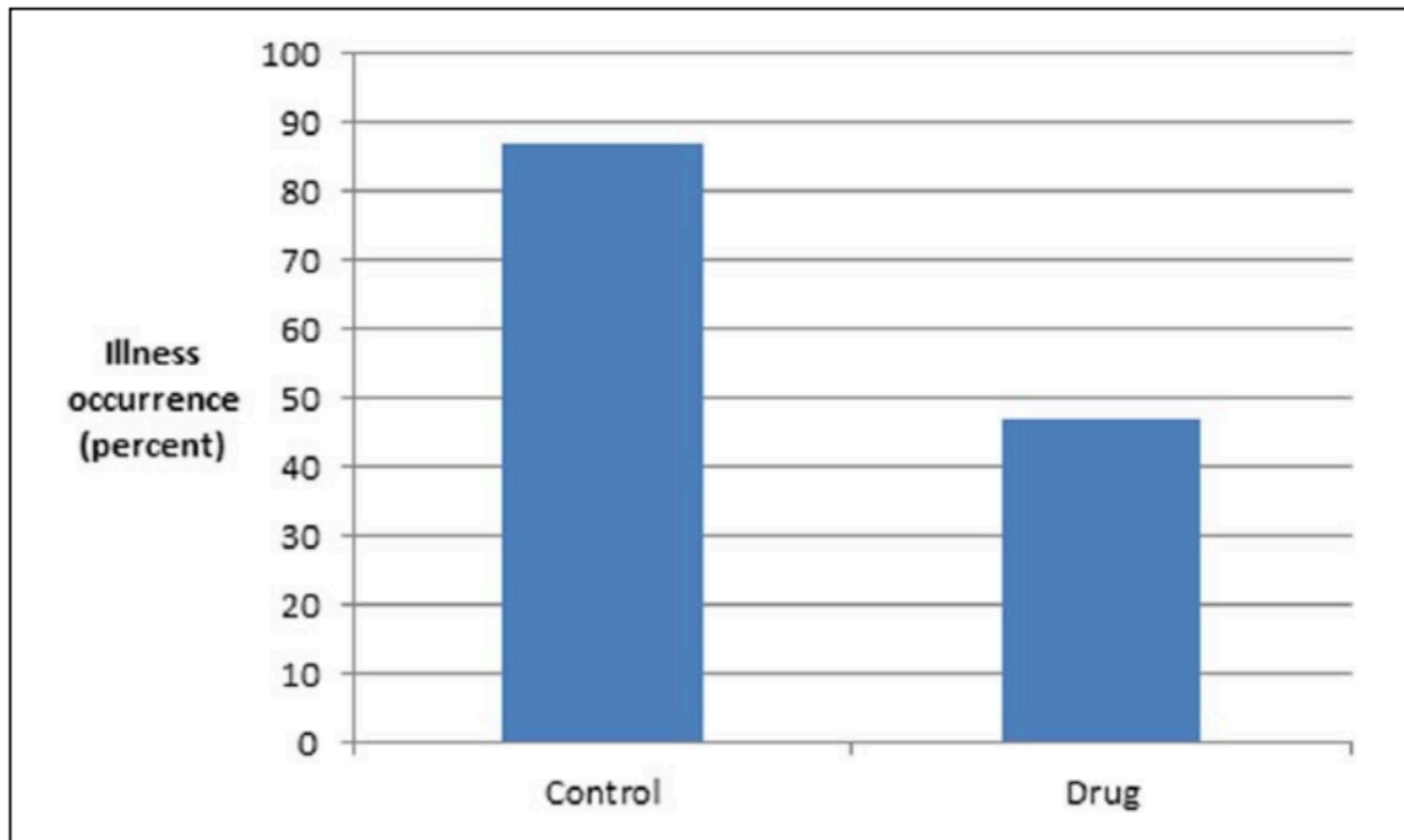
Data sources: Centers for Disease Control & Prevention and Internet Movie Database

tylervigen.com



A large pharmaceutical company has recently developed a new drug to boost peoples' immune function. It reports that trials it conducted demonstrated a drop of forty percent (from eighty seven to forty seven percent) in occurrence of the common cold. It intends to market the new drug as soon as next winter, following FDA approval.





A large pharmaceutical company has recently developed a new drug to boost peoples' immune function. It reports that trials it conducted demonstrated a drop of forty percent (from eighty seven to forty seven percent) in occurrence of the common cold. It intends to market the new drug as soon as next winter, following FDA approval.

Tal, A. and Wansink, B., 2016. Blinded with science: Trivial graphs and formulas increase ad persuasiveness and belief in product efficacy. *Public Understanding of Science*

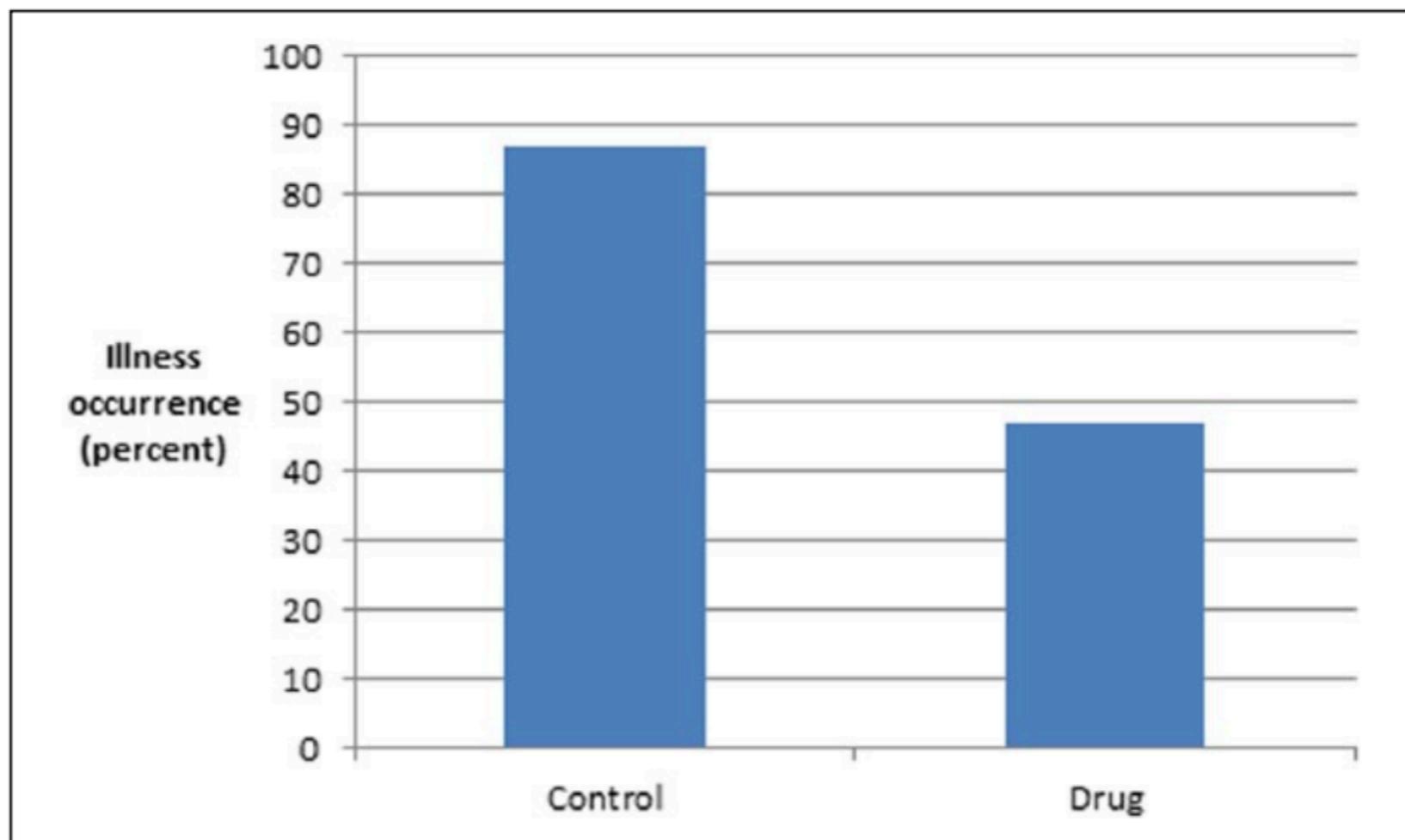
NOTE: Though this paper has not been retracted, Brian Wansink has had several other papers retracted due to statistical and ethical issues.

This example is one of many like it, so I'm less concerned about the general applicability of these findings



Rate trustworthiness from 1 to 9

w/graph 6.83 / 9 vs without 6.12 / 9



$$t(59) = -2.1, p = .04$$

Displaying a graph made people trust the findings more

Tal, A. and Wansink, B., 2016. Blinded with science: Trivial graphs and formulas increase ad persuasiveness and belief in product efficacy. Public Understanding of Science

NOTE: Though this paper has not been retracted, Brian Wansink has had several other papers retracted due to statistical and ethical issues.

This example is one of many like it, so I'm less concerned about the general applicability of these findings



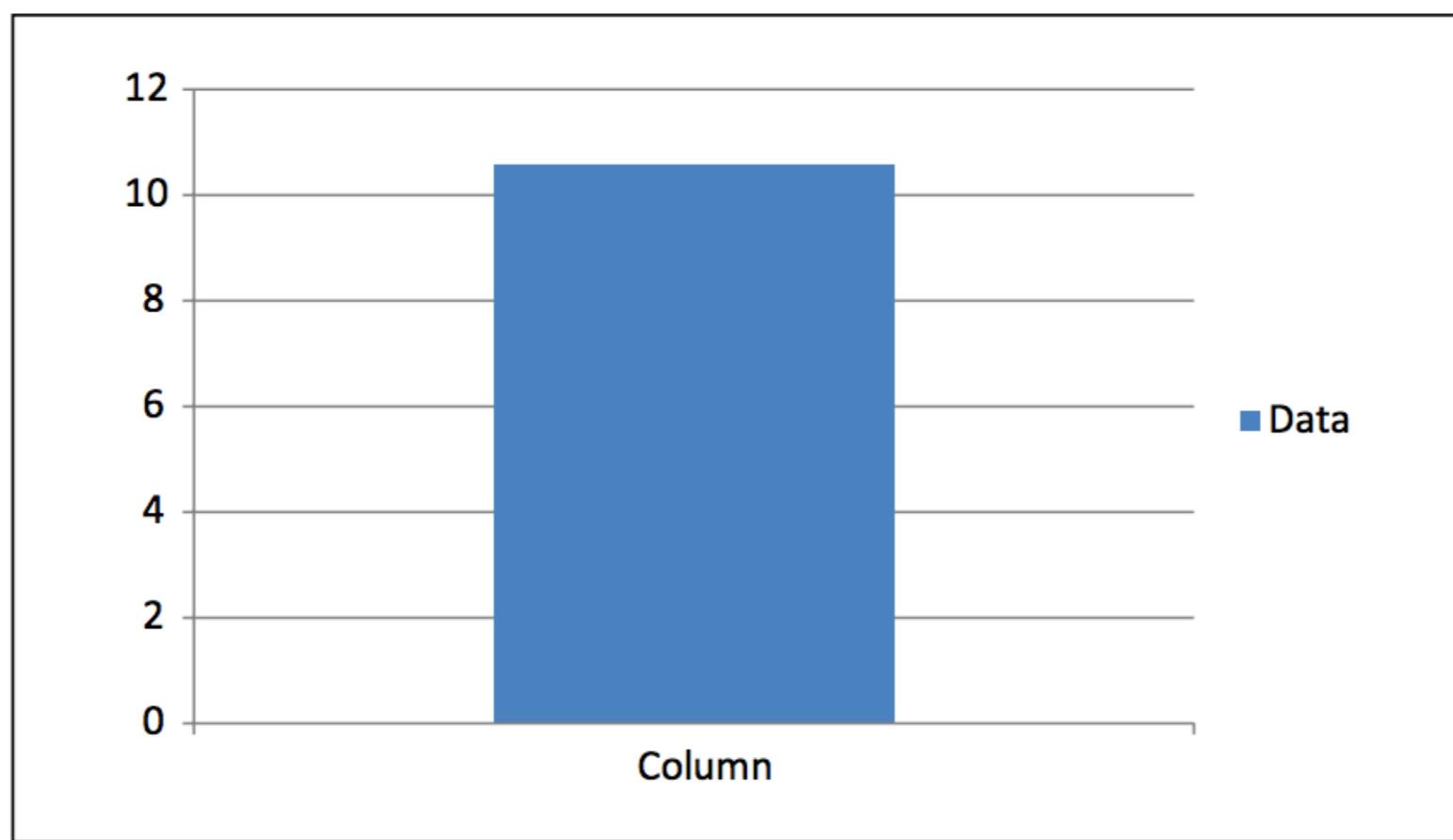
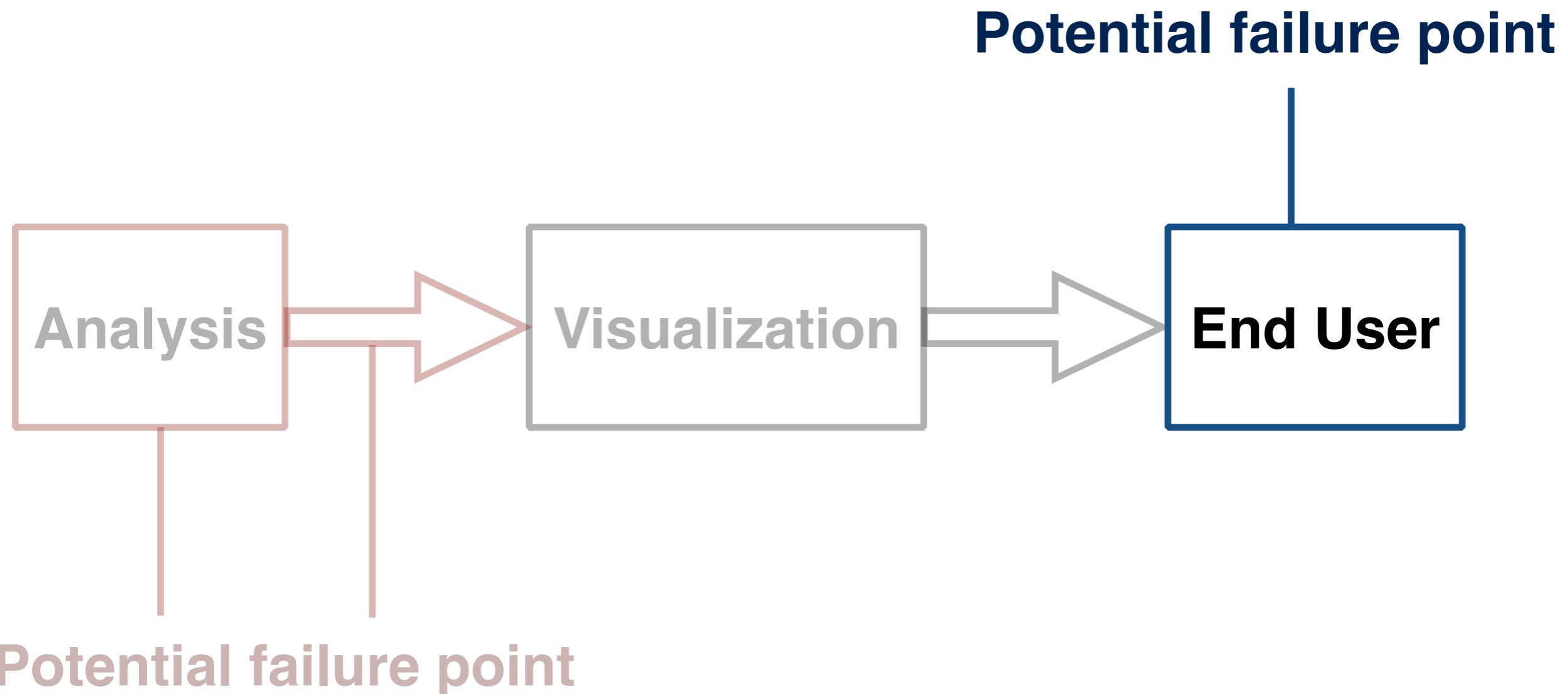


Figure 2. A highly convincing graph.

Further research can investigate the links between the “science-signaling” of graphs and their effects on persuasion. In addition, research may extend to study what elements other than graphs and formulas increase persuasion through “science signaling.” Finally, research can also examine whether particular populations would be more or less vulnerable to these effects, and investigate methods of mitigating such effects.

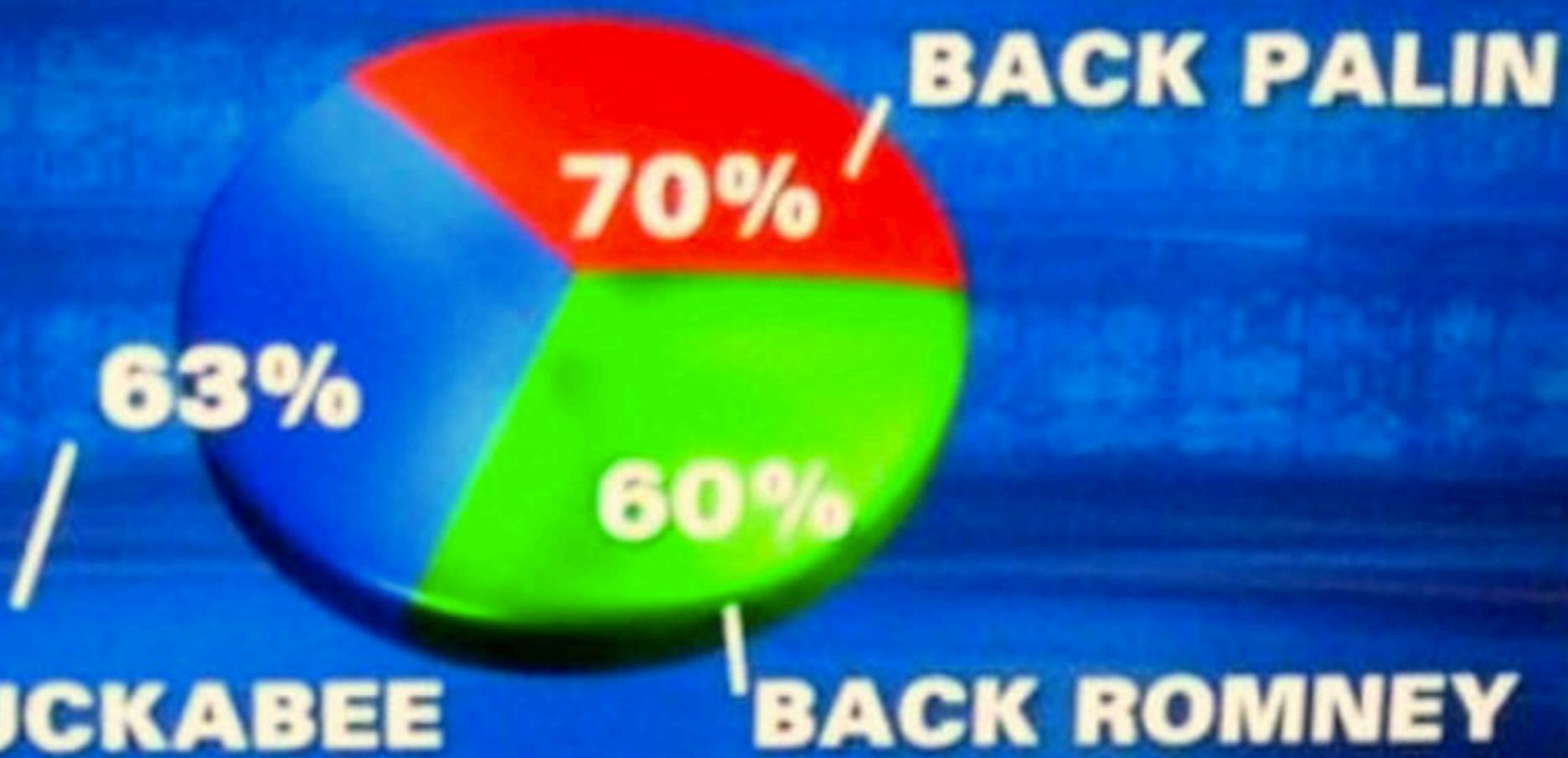
On the practical front, the research demonstrates how easily companies can create a scientific appearance. Accordingly, it advises caution when encountering communications hinting at sci-

Dataviz Process Model



2012 PRESIDENTIAL RUN

GOP CANDIDATES



SOURCE: OPINIONS

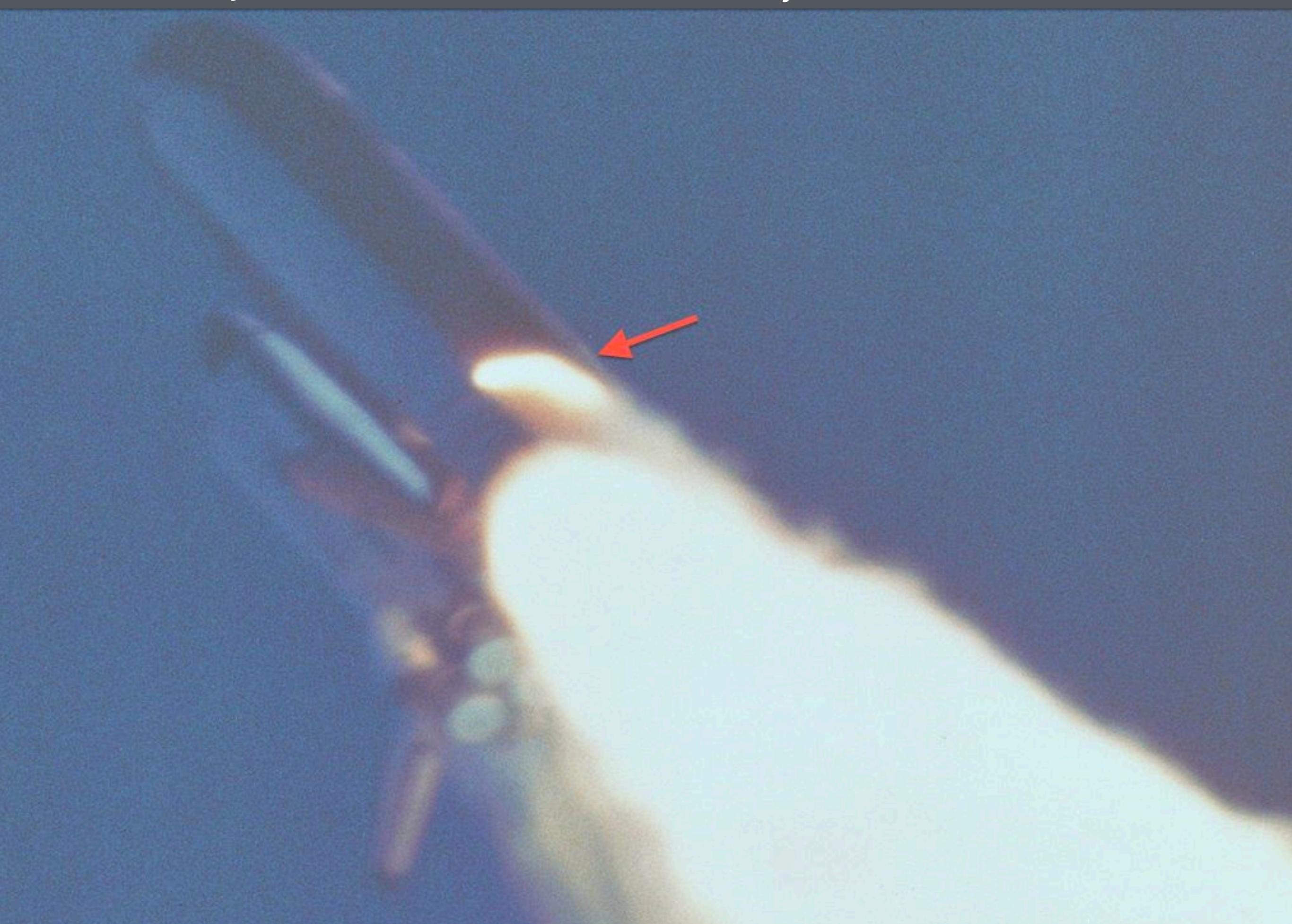
DYNAMIC

FOX

47°









"regardless of the indirect cultural causes of the accident, there was a clear proximate cause: an inability to assess the link between cool temperature and O-ring damage on earlier flights."

	ART	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

- 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

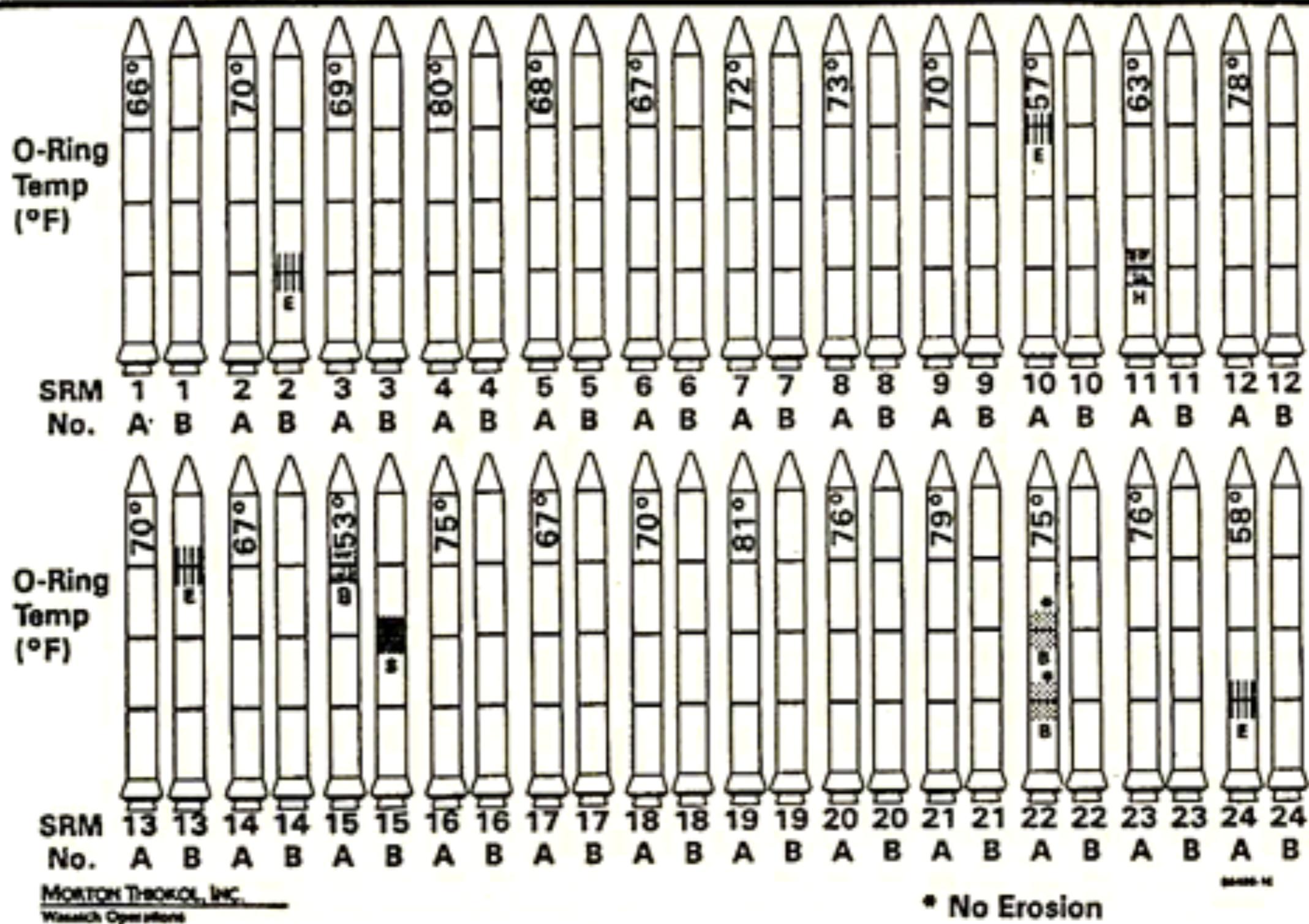
HISTORY OF O-RING TEMPERATURES (DEGREES - F)

MOTOR	MGT	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

Figures from Tufte, 1997



History of O-Ring Damage in Field Joints (Cont)

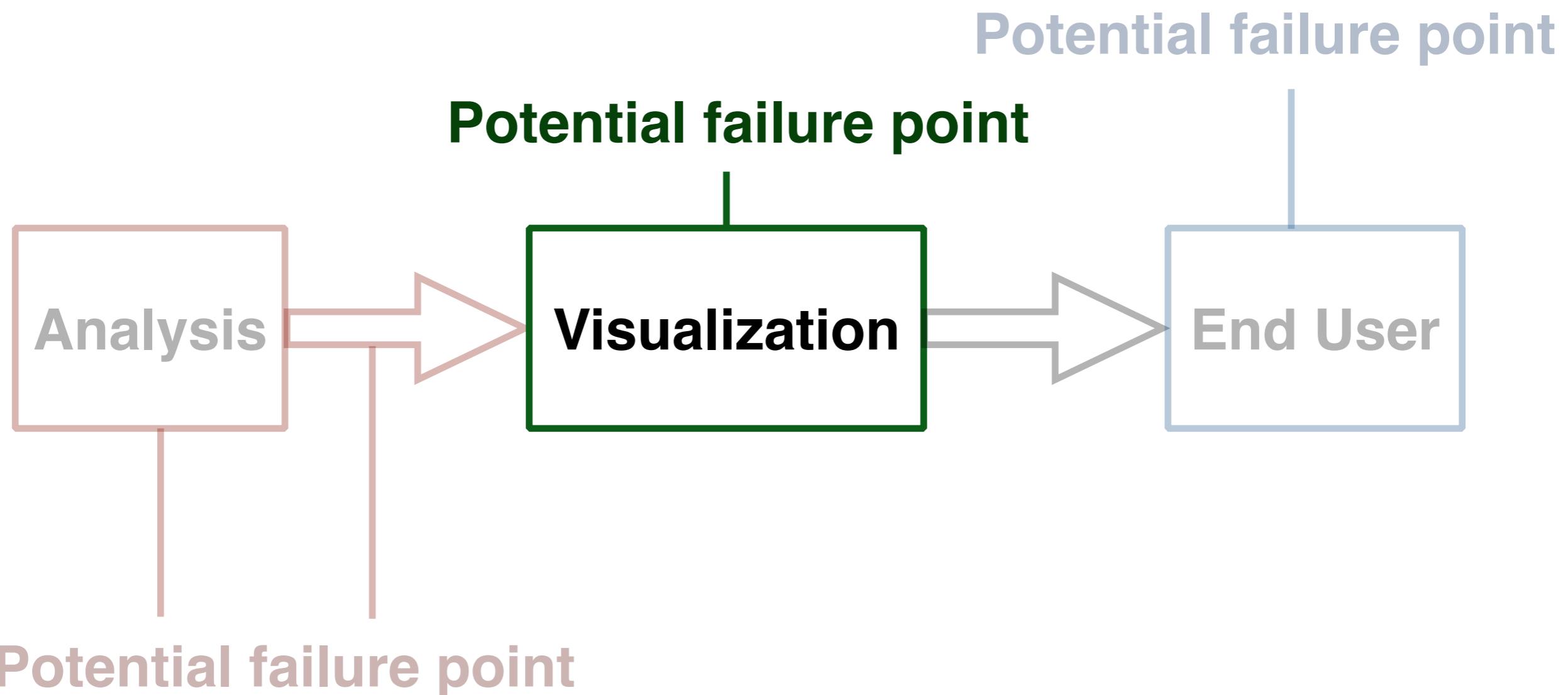


INFORMATION ON THIS PAGE WAS PREPARED TO SUPPORT AN ORAL PRESENTATION
AND CANNOT BE CONSIDERED COMPLETE WITHOUT THE ORAL DISCUSSION

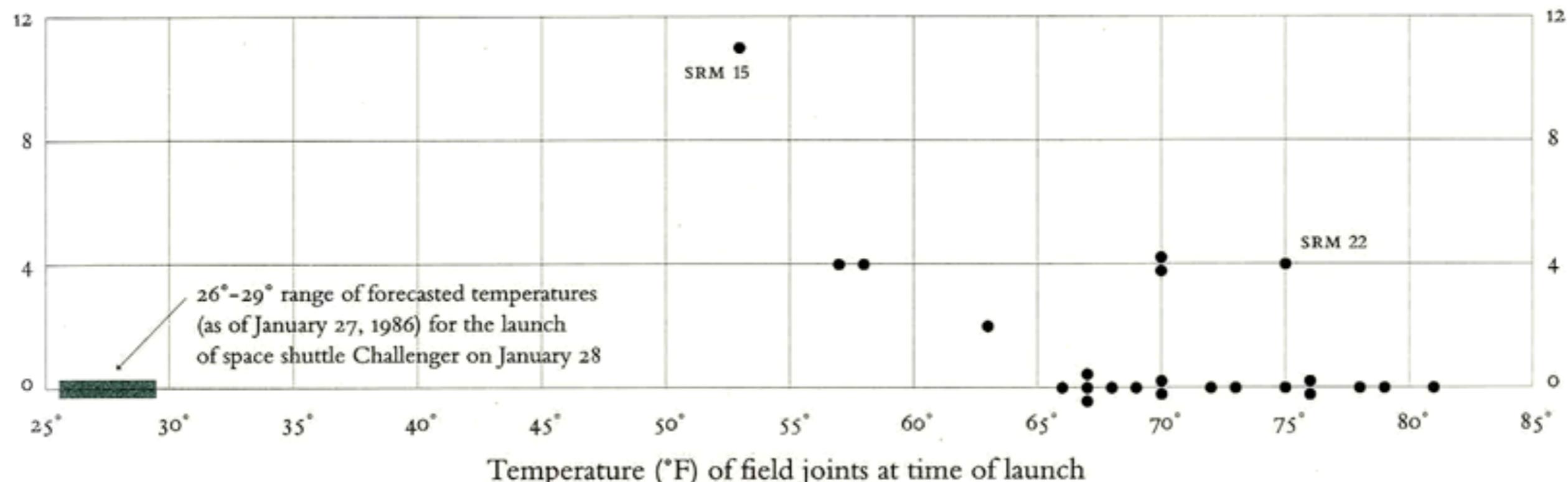
Figure from Tufte, 1997



Dataviz Process Model



O-ring damage
index, each launch



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	Coolest (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	

Figure from Tufte, 1997



Maybe we just need better visualizations!

O-ring damage
index, each launch

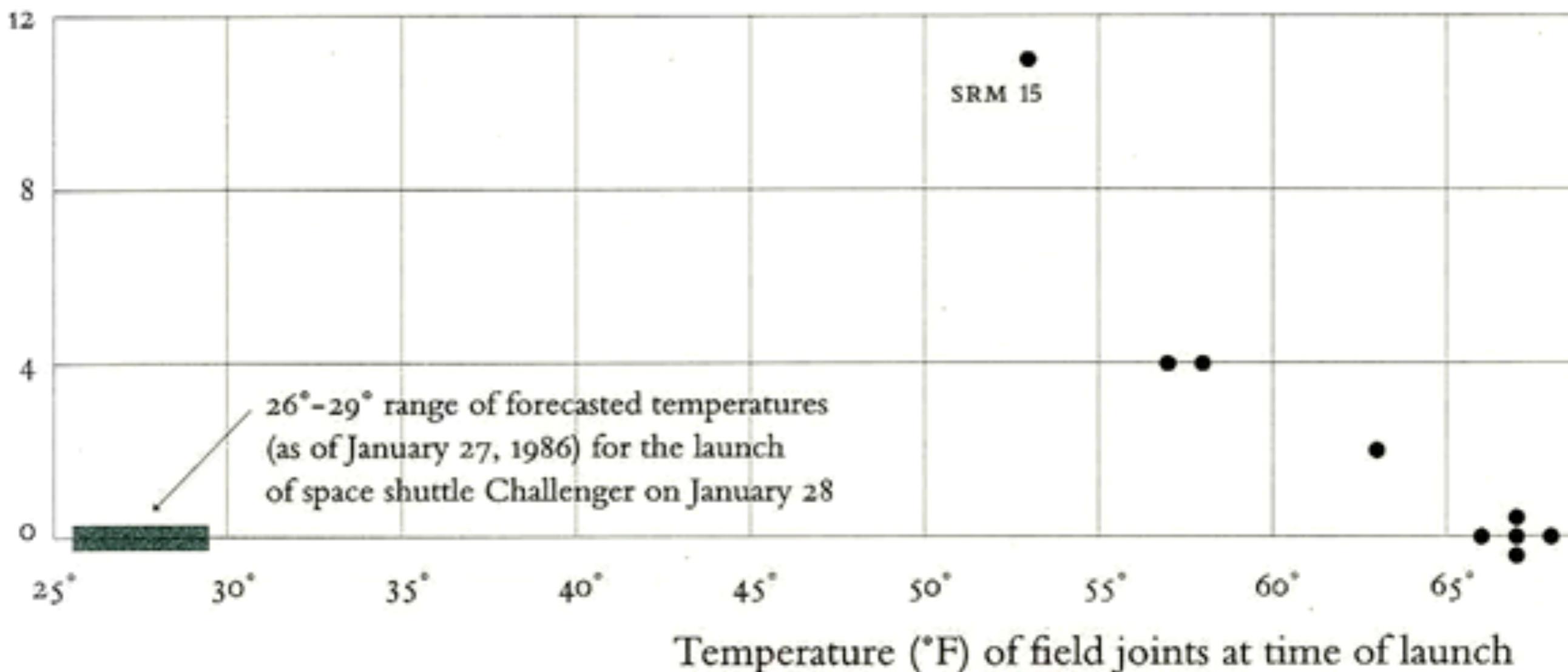
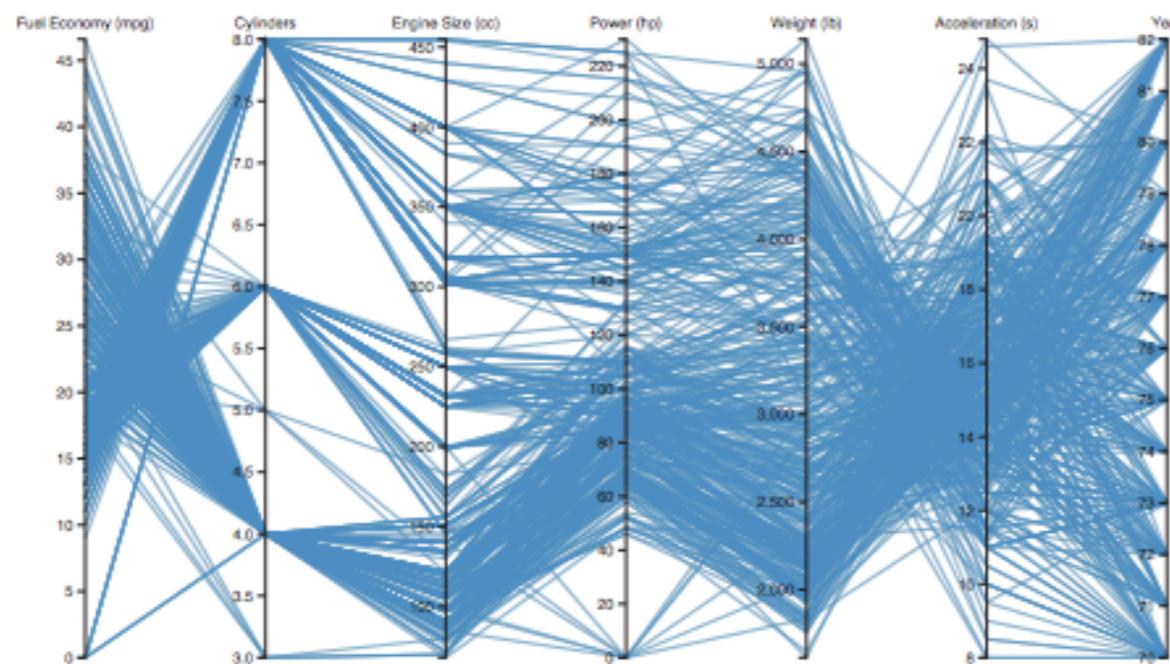


Figure from Tufte, 1997

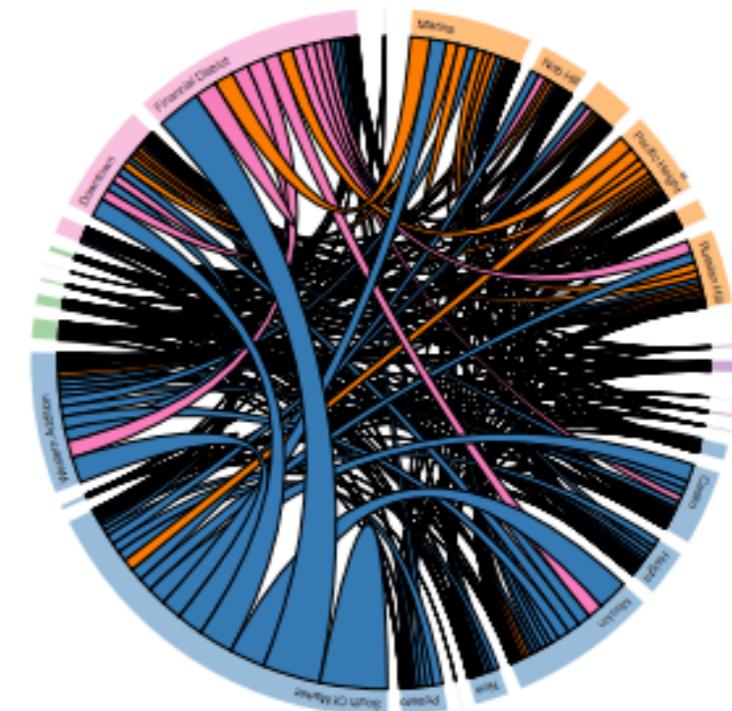


"Please verbalize your thoughts and behavior while trying to make sense of the visualization"

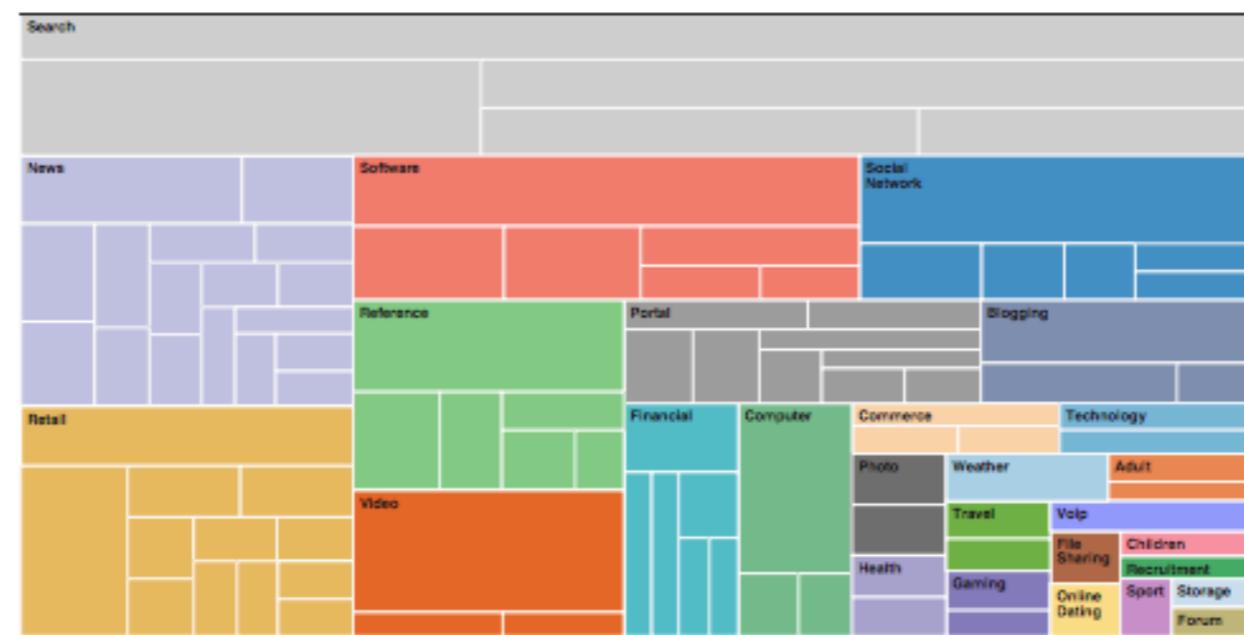
Cars From The 1970s And 1980s

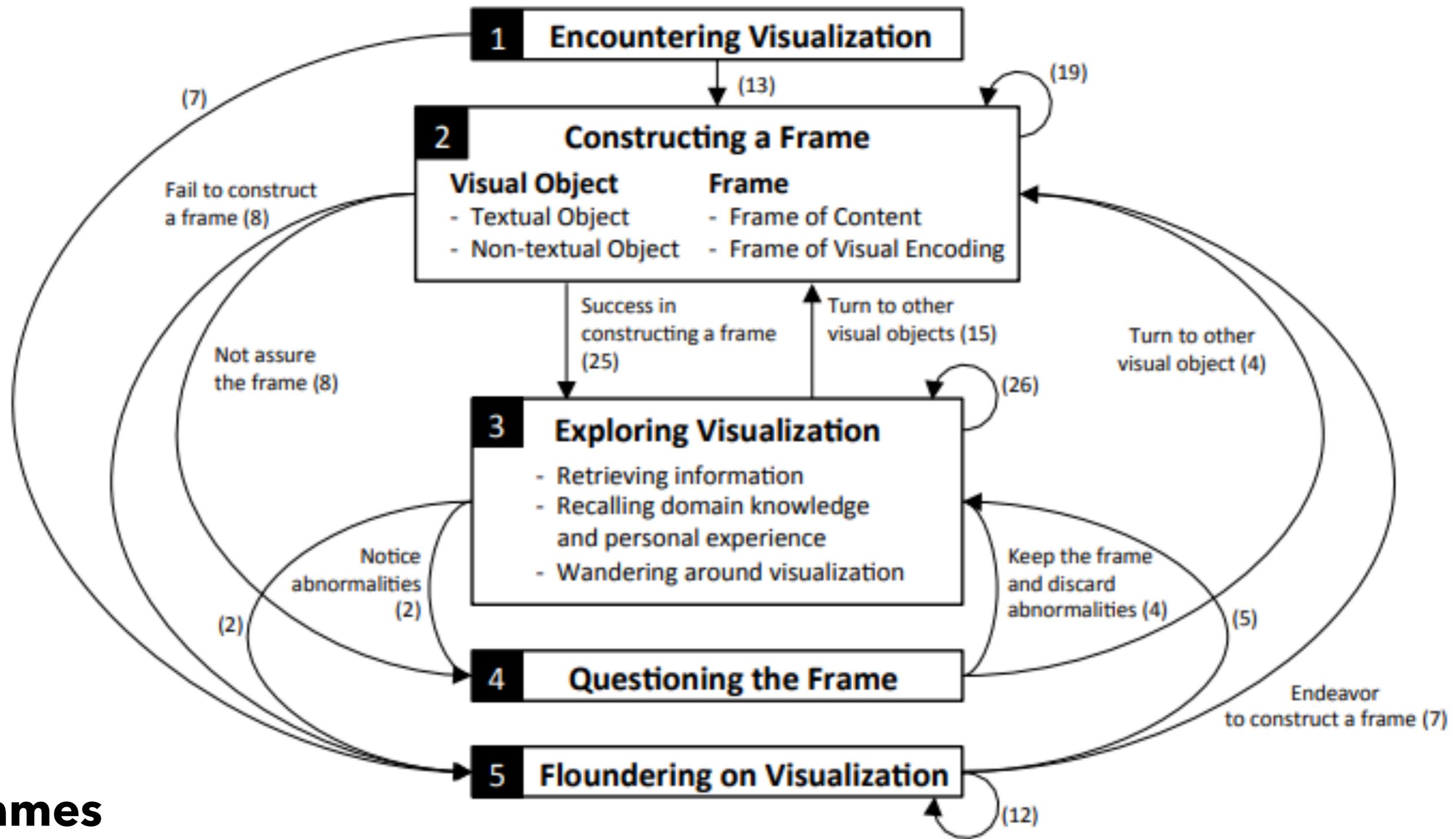


Private Driver Service in San Francisco



Unique Users Of The Internet

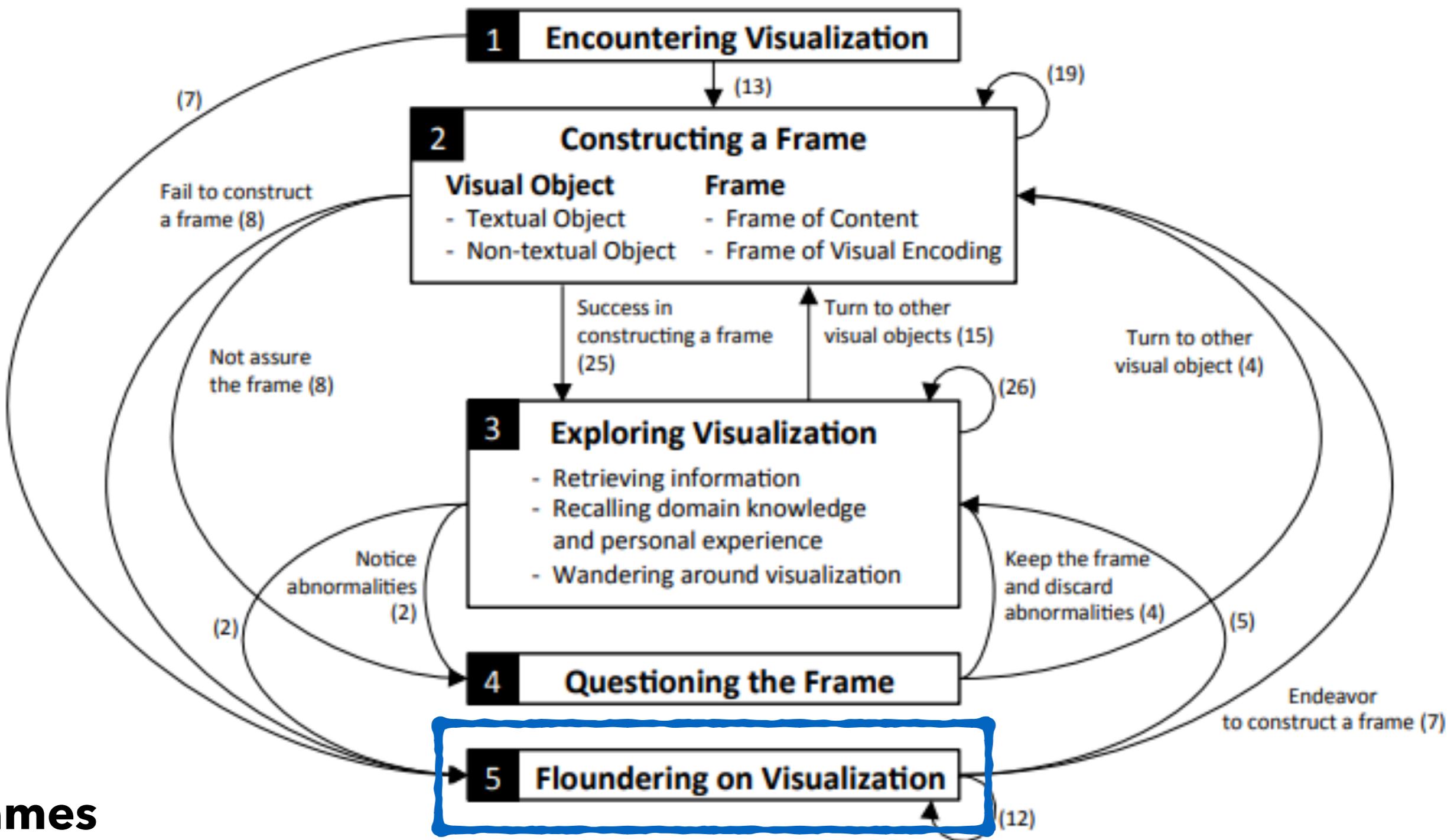




Frames

“Hmm, this visualization is about different websites”

“Ah, I think that the size of each box means a number of users for each website”

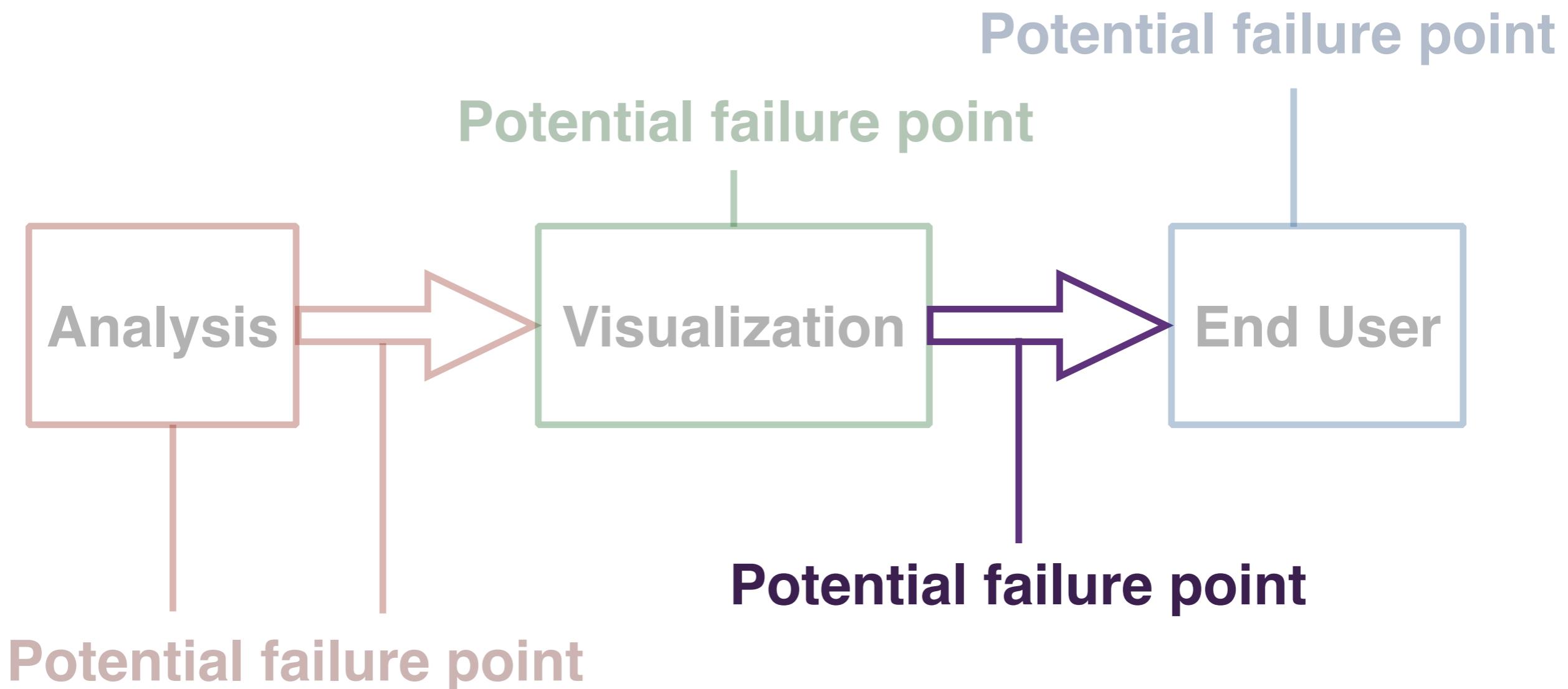


Frames

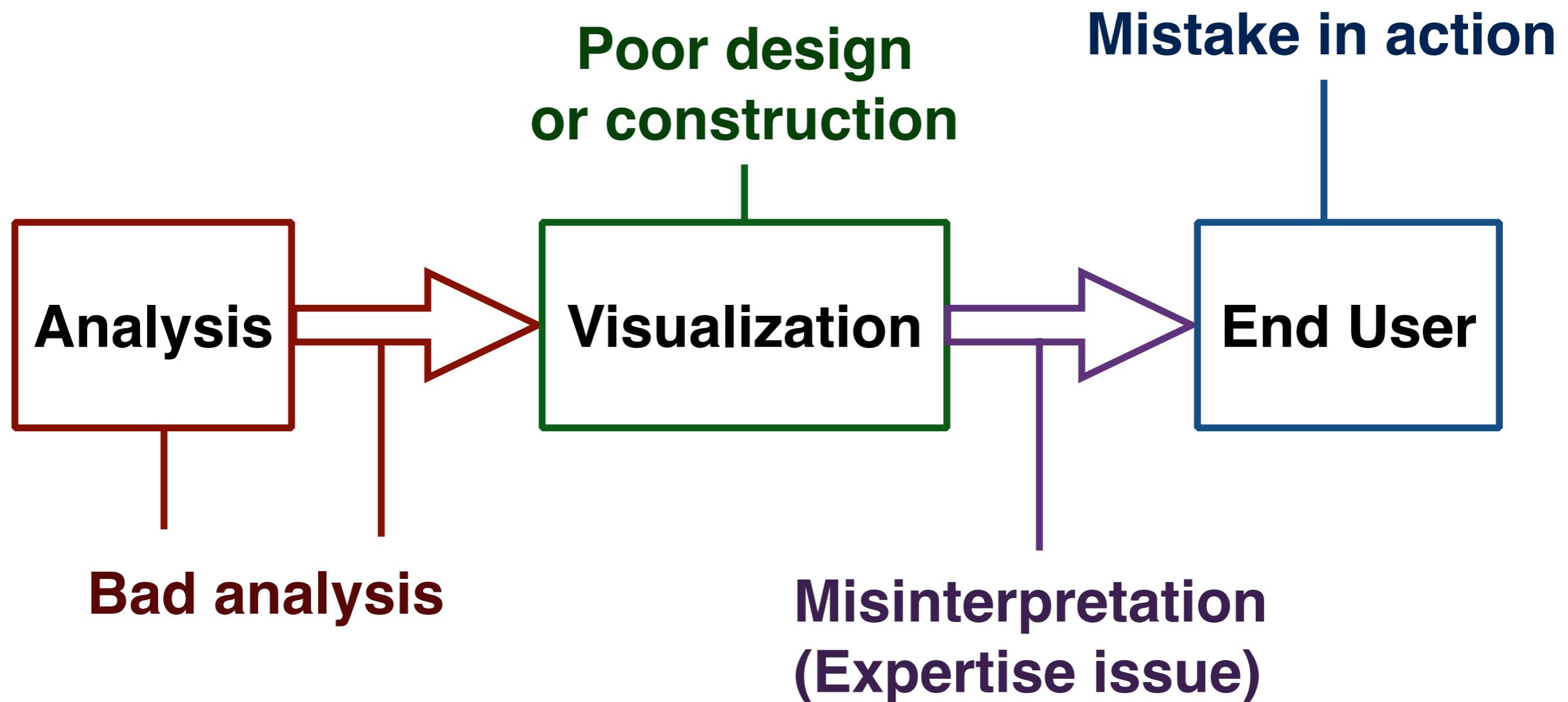
"Hmm, this visualization is about different websites"

"Ah, I think that the size of each box means a number of users for each website"

Dataviz Process Model

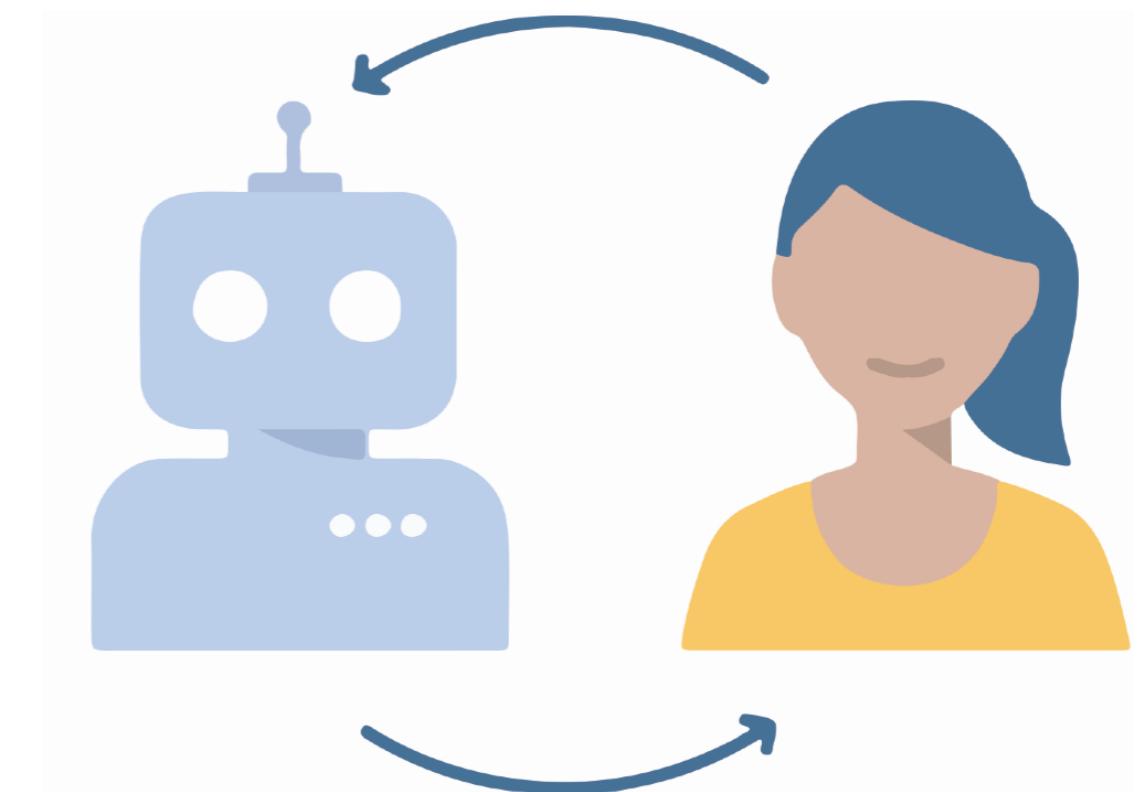


There are many potential failure points!



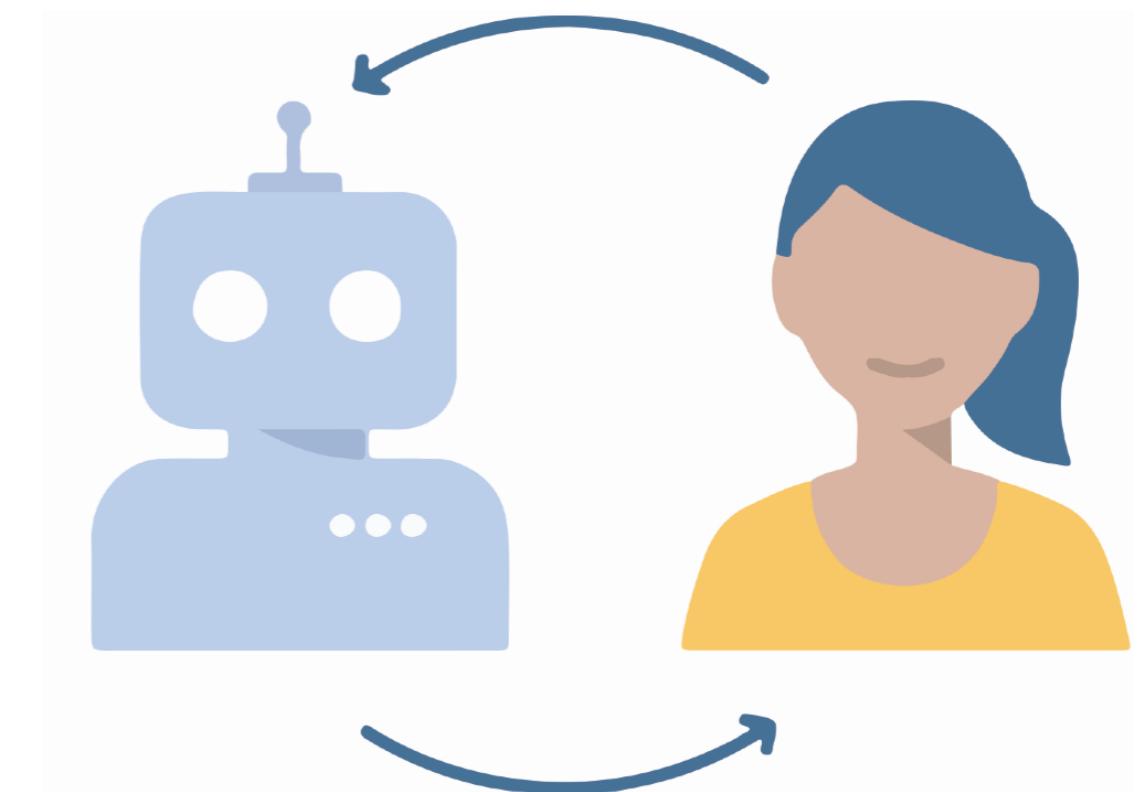
How do we design an effective visualization?

- Data we have



How do we design an effective visualization?

- Data we have
- Message to convey



How do we design an effective visualization?

- Data we have
- Message to convey
- Audience of vis

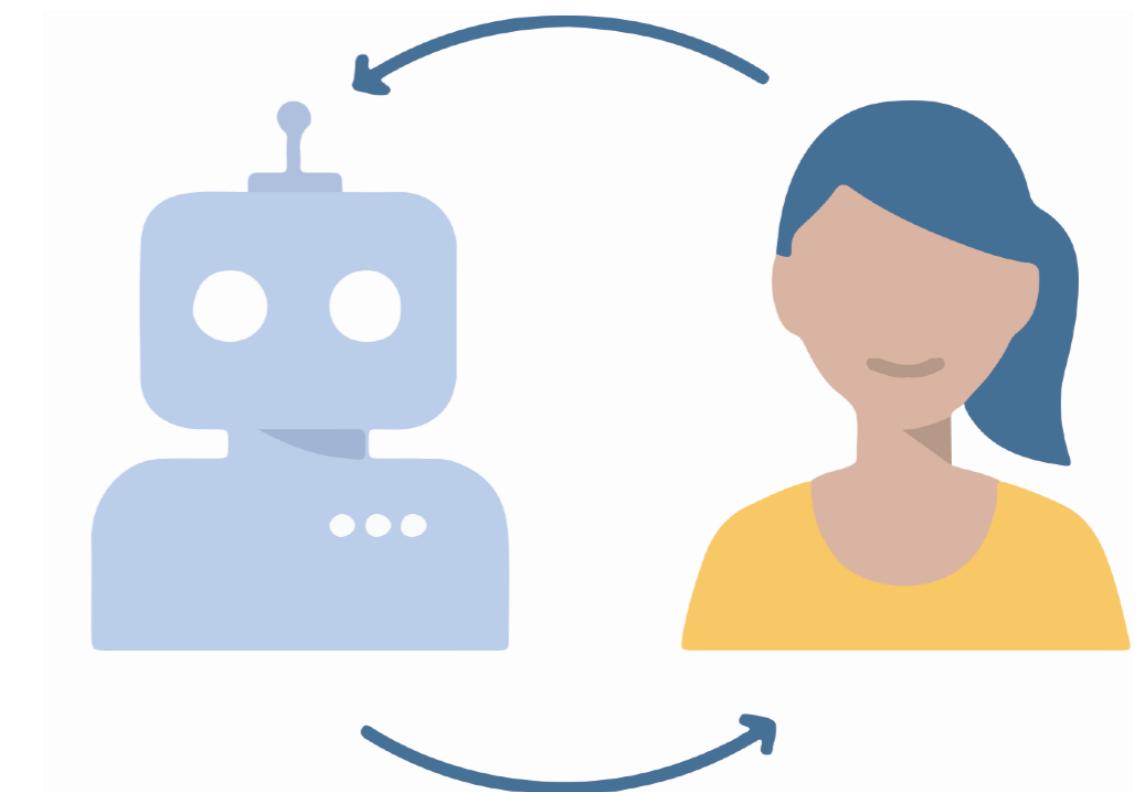


Image credit: CrowdFlower

How do we design an effective visualization?

- Data we have
- Message to convey
- Audience of vis
- System capabilities

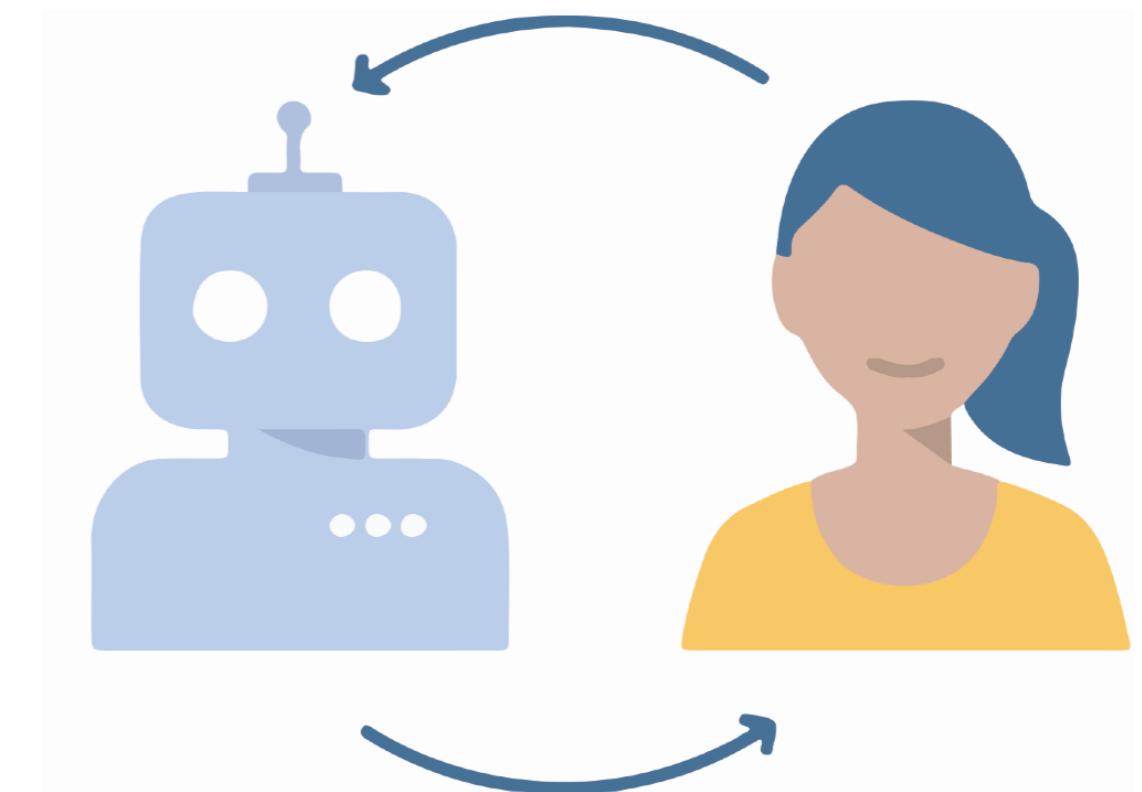


Image credit: CrowdFlower

INFO3300 / CS3300 / INFO5100

Develop skills in client-side visualization development (using common libraries)

Create a variety of static and interactive web visualizations for different datatypes

Learn trade-offs and best practices for matching data to visual elements

Get practical experience gathering and visualizing real-world data



Pre-requisites

CS 2110

Make sure you know object-oriented
programming and data structures!

INFO/CS 2300

Significant front- and back-end experience with JS
and web-dev. Projects rather than small examples.



Weekly Classes

2 / 3 - Live coding & demonstration

1 / 3 - Lectures on visualization design

Always check the course schedule page!

Attendance policy:

Short quizzes via CMS each week

Laptops on the left and right sides of the room



Class Activities

Every other week, Friday remote-only

Brainstorming and sketching are critical design learning skills...

...but they don't work in a classroom this big!

We will do design activities in small groups via
Zoom breakout rooms

Make sure to have sketching materials ready



Course Work

10 homework assignments:

Weekly, assigned via CMS on Mondays

Due Wed @ 11:59PM via CMS

No late assignments accepted, except...

5 homework slip days

(see policy doc for more details)

Lowest 2 grades dropped (including zeroes)

2 projects:

Static & interactive

Instructor assigned teams (w/input)

TA mentorship opportunities

In-class demo/critique days



Exams

3300:

Take-home final exam

(Cumulative over term)

5100:

Large scale final projects



Grading

We are using an **adjusted grading scale** (A+ for over 100%)
See syllabus for more details

	Exam Bias	HW Bias
Class Attendance	8%	8%
Weekly Homework	32%	40%
Project 1 & 2	30%	30%
Project 3 OR Take-home Exam	30%	22%

Lots of extra credit offered!
See syllabus for more details



Waitlist Status

>50 people on waitlist

Priority will be given to graduating seniors
and IS majors/grad students

Have a backup plan if you are a
sophomore, junior, or non-IS senior

(Individuals joining from the waitlist will be granted
extra days on HW and quizzes based on add date)

Support

Q & A resources:

Experimenting with in-class Discord

(more info on Monday)

Post long form course questions to Ed Discussions

Office Hours:

To be posted on ED and Canvas, begin next week

Personal questions:

Use course email info3300staff@gmail.com

Do not email TAs or the instructor about course concerns unless otherwise instructed



Support

Instructor

Jeff Rzeszotarski

Assistant Prof, IS

jeffrz.com, jeffrz@cornell.edu

(you can call me Jeff or Prof. Rz)

Graduate Assistants

Ji Yong Cho

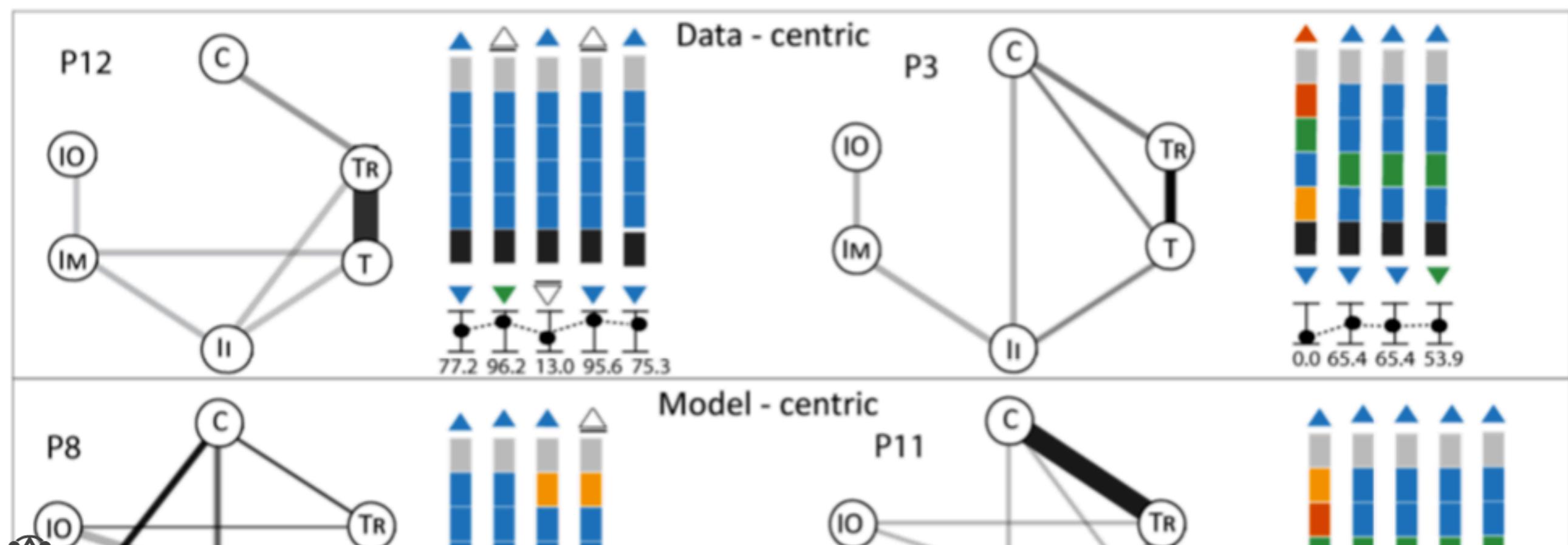
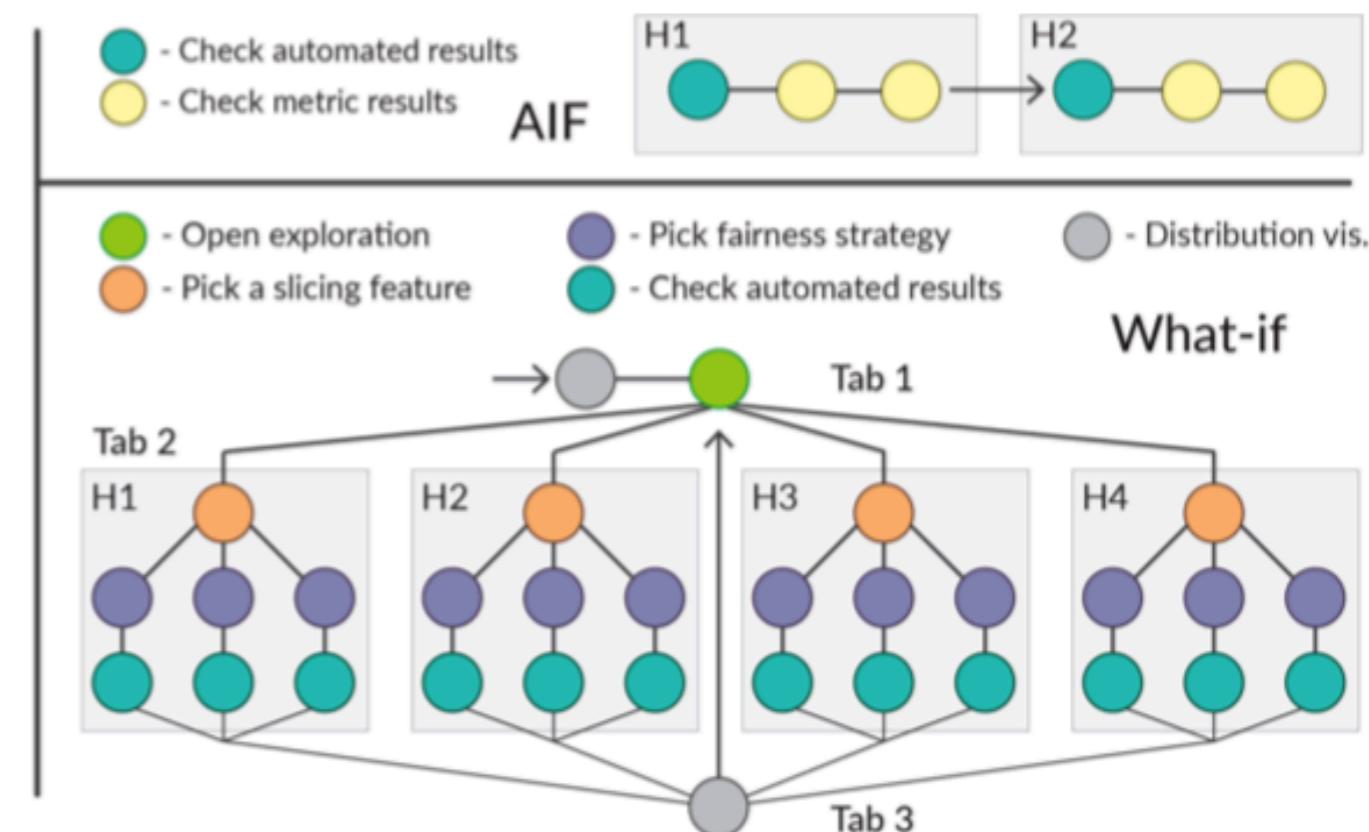
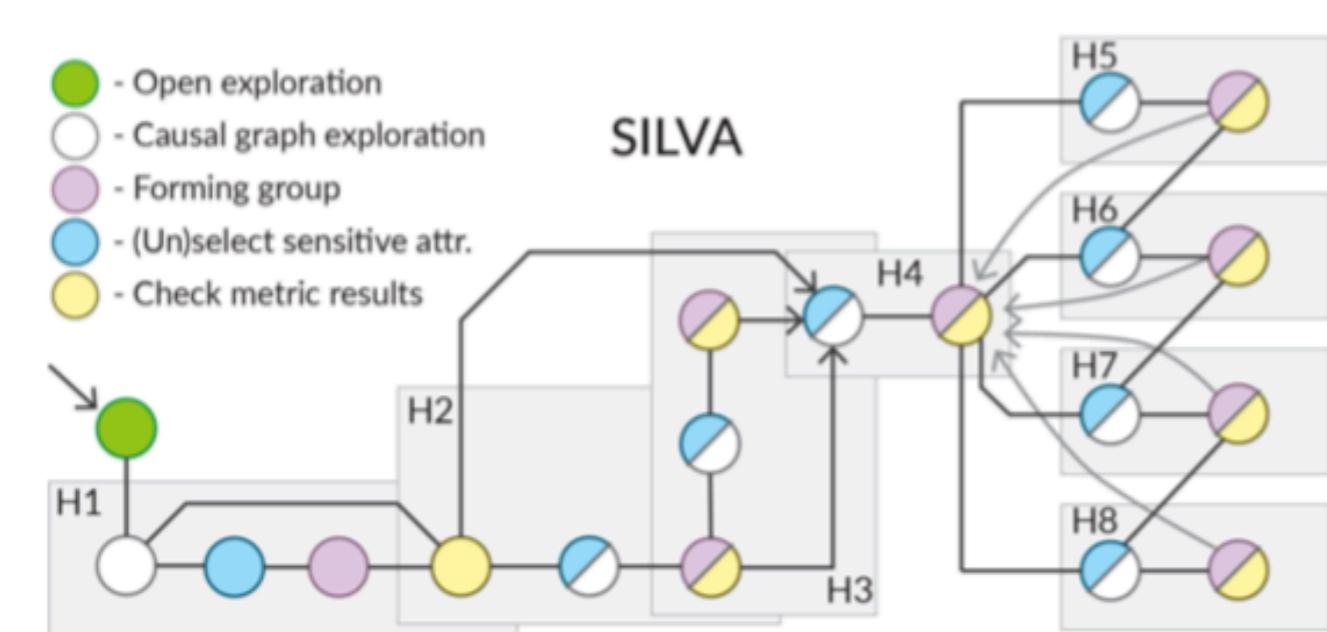
jc3374@cornell.edu

Michael Suguitan

mjs679@cornell.edu

Reach all of us at info3300staff@gmail.com





ML Fairness Explorer

Choose a Dataset:

Adult Income

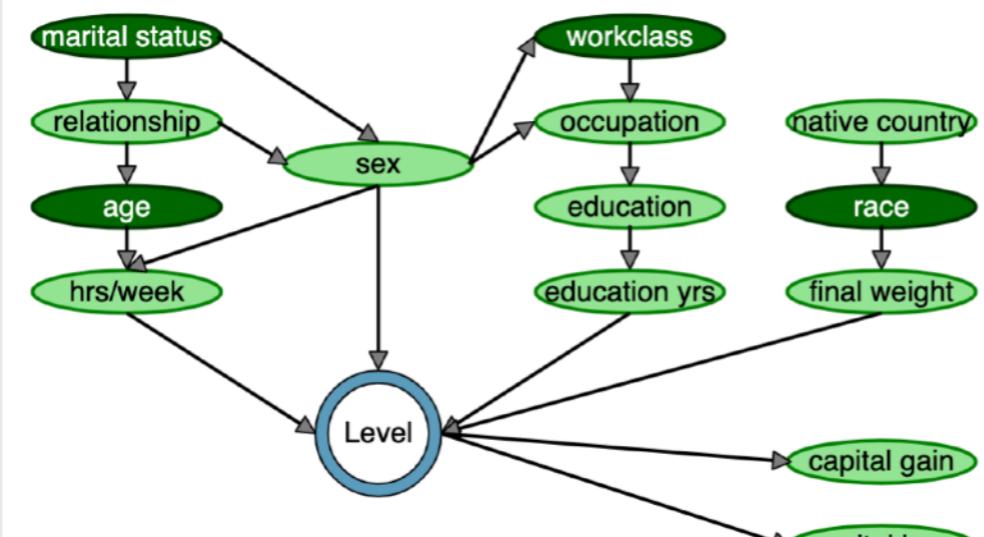
Attributes

- marital-status
- relationship
- age
- hours-per-week
- sex
- workclass
- occupation
- education
- education-num
- native-country
- race
- fnlwgt
- capital-gain
- capital-loss

Sensitivity

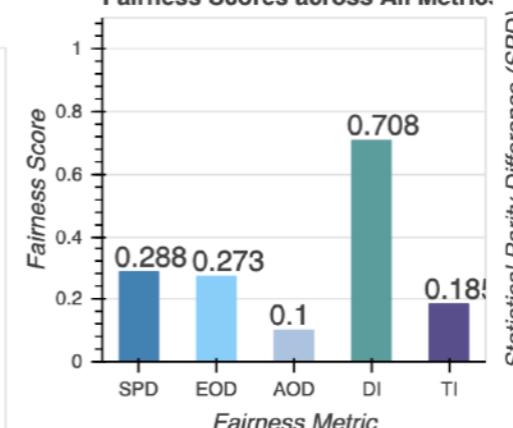


Causal Graph Demonstration

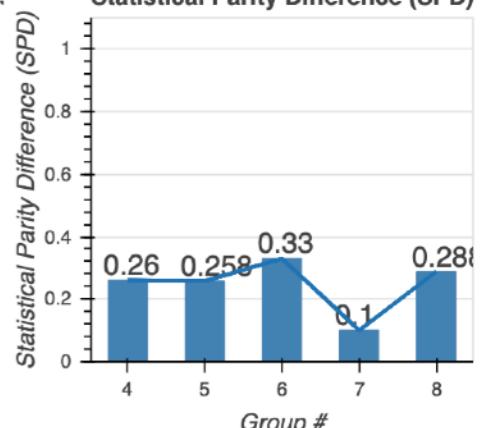


MLP Random Forest Logistic Regression

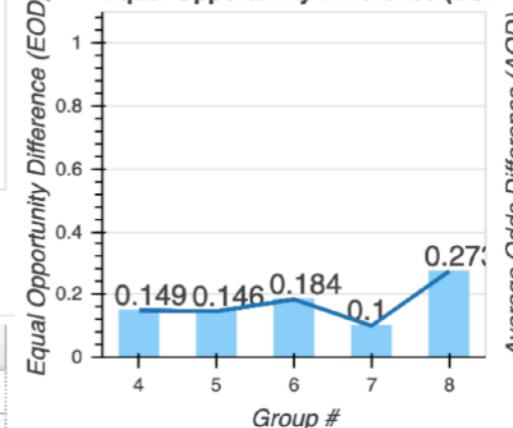
Fairness Scores across All Metric:



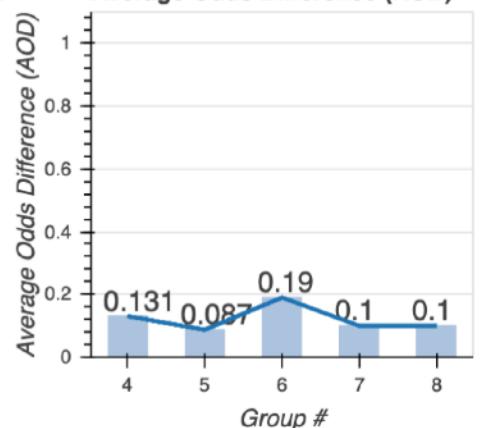
Statistical Parity Difference (SPD)



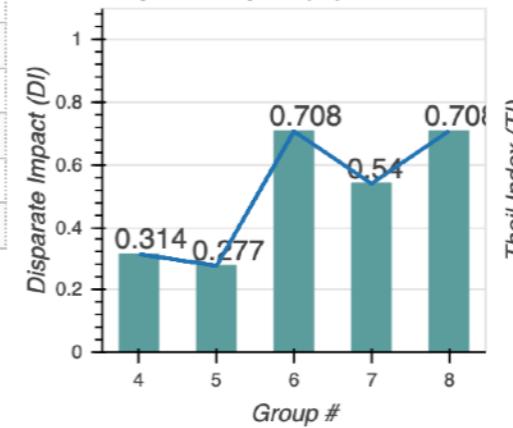
Equal Opportunity Difference (EOD)



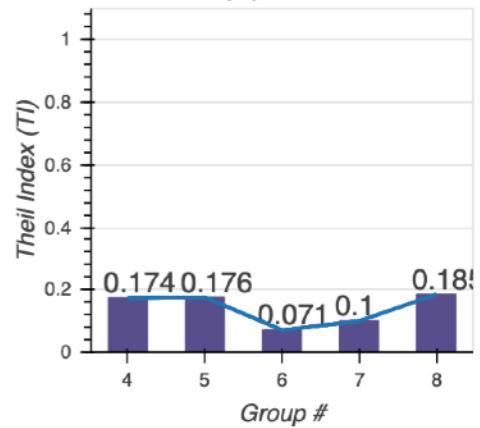
Average Odds Difference (AOD)



Disparate Impact (DI)



Theil Index (TI)



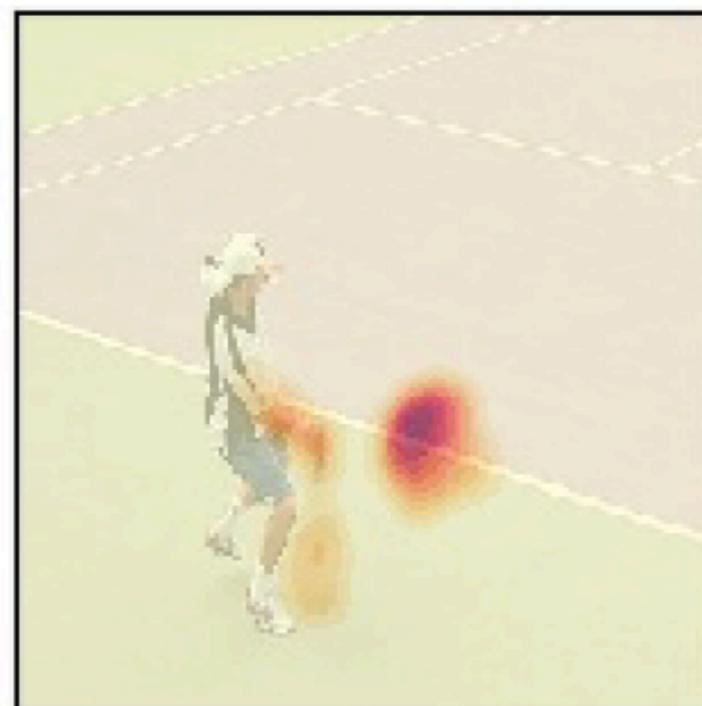
Group Table Dataset Table

Group #	Sensitive Attributes	Non-sensitive Attributes	Avg SPD	Avg EOD	Avg AOD	Avg DI	Avg TI	Data SPD	Data DI
1	race	relationship, sex, hours-per-week	0.201	0.113	0.118	0.474	0.093	0.355	0.206
2	sex, race	marital-status, relationship, age, hours-per-week, w...	0.255	0.188	0.144	0.598	0.083	0.355	0.206
3	marital-status, relationship	race, sex	0.202	0.158	0.115	0.453	0.117	0.355	0.206
4	workclass, occupation, education	marital-status, relationship, race, fnlwgt, age, hours...	0.213	0.16	0.137	0.466	0.119	0.355	0.206
5	marital-status, relationship, hours-per-week, workcl...	race, fnlwgt, age, sex, education-num, native-country	0.223	0.138	0.135	0.455	0.118	0.355	0.206
6	sex	marital-status, relationship, fnlwgt, age, hours-per-...	0.333	0.169	0.2	0.728	0.086	0.355	0.206
7	age, sex	relationship, fnlwgt, hours-per-week, occupation, ed...	0.1	0.1	0.1	0.68	0.1	0.355	0.206
8	marital-status, race, age, workclass	relationship, fnlwgt, capital-gain, capital-loss, hours...	0.169	0.187	0.155	0.636	0.178	0.355	0.206



MobileNet GradCAM

Areas identified by humans



Human+machine feedback

Importance of each part

- ball
- handle
- hand
- frame



About me



Action Items

Visit course website: jeffrz.com/info3300

Make sure you can access Canvas and CMS

We'll talk through Discord and other course materials on Monday

Coding lectures starting next week

If you have questions, please email staff rather than approaching the podium