

Lecture 15: Visualization for Communication

DS4200 Fall 2022

Prof. Ab Mosca (they/them)

Looking Forward

- In-class presentations 12/02 and 12/07
- Then you're done with DS4200!

Today

- Visualization for communication
- Intro to Visualization Research (example of my work in visualization for communication)

What is visualization for communication?

Visualization

Visualization (in CS)

Visualization (in CS)

Data Analysis Pipeline

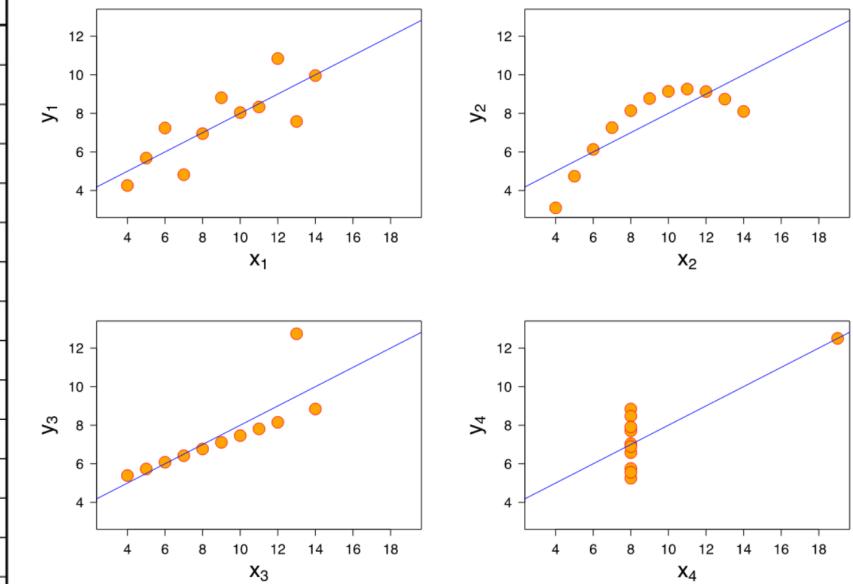
Visualization (in CS)

Data Analysis Pipeline -- Validation

Visualization (in CS)

Data Analysis Pipeline -- Validation

Anscombe's Data									
Observation	x1	y1	x2	y2	x3	y3	x4	y4	
1	10	8.04	10	9.14	10	7.46	8	6.58	
2	8	6.95	8	8.14	8	6.77	8	5.76	
3	13	7.58	13	8.74	13	12.74	8	7.71	
4	9	8.81	9	8.77	9	7.11	8	8.84	
5	11	8.33	11	9.26	11	7.81	8	8.47	
6	14	9.96	14	8.1	14	8.84	8	7.04	
7	6	7.24	6	6.13	6	6.08	8	5.25	
8	4	4.26	4	3.1	4	5.39	19	12.5	
9	12	10.84	12	9.13	12	8.15	8	5.56	
10	7	4.82	7	7.26	7	6.42	8	7.91	
11	5	5.68	5	4.74	5	5.73	8	6.89	
Summary Statistics									
N	11	11	11	11	11	11	11	11	
mean	9.00	7.50	9.00	7.500909	9.00	7.50	9.00	7.50	
SD	3.16	1.94	3.16	1.94	3.16	1.94	3.16	1.94	
r	0.82		0.82		0.82		0.82		



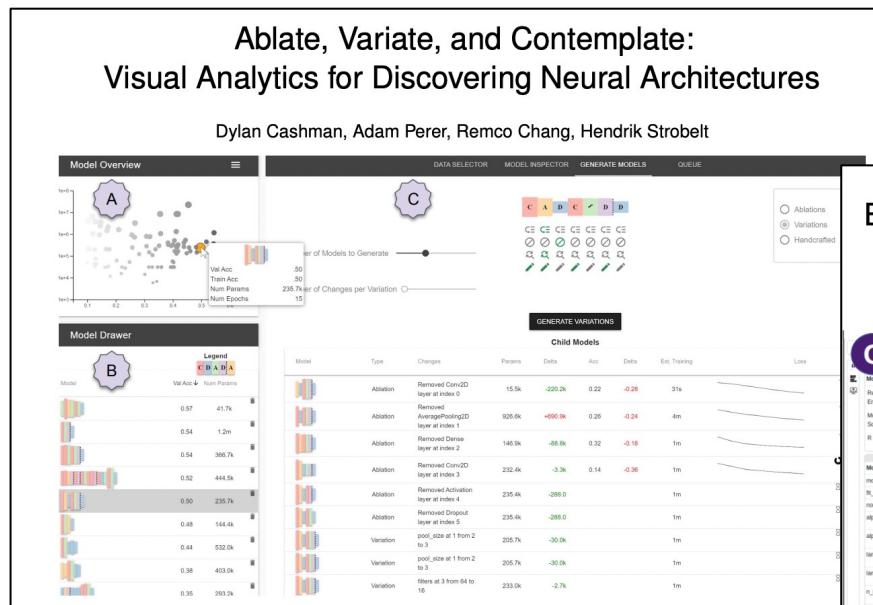
<https://heap.io/blog/data-stories/anscombes-quartet-and-why-summary-statistics-dont-tell-the-whole-story>

Visualization (in CS)

Data Analysis Pipeline -- Modeling

Visualization (in CS)

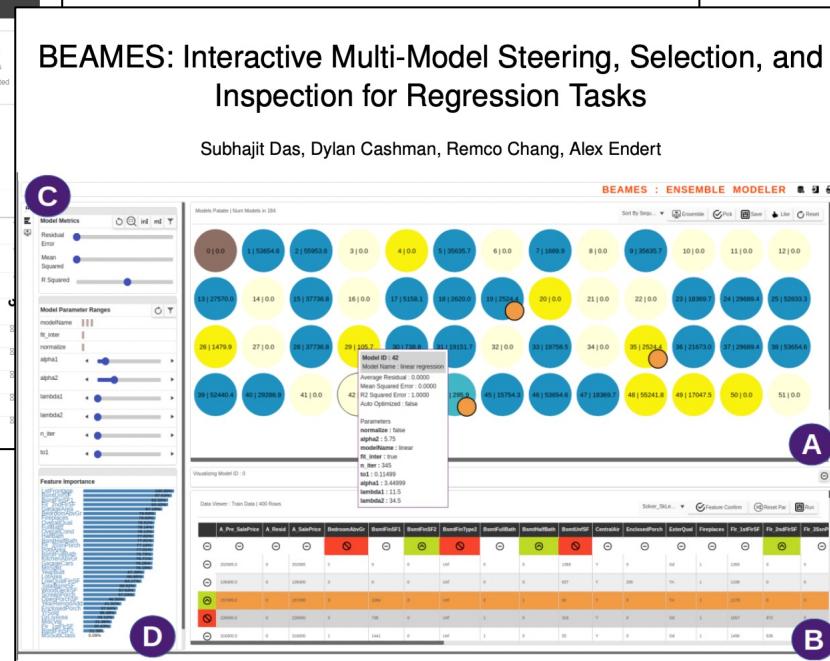
Data Analysis Pipeline -- Modeling



<https://dylancashman.github.io/public/docs/remap.pdf>

BEAMES: Interactive Multi-Model Steering, Selection, and Inspection for Regression Tasks

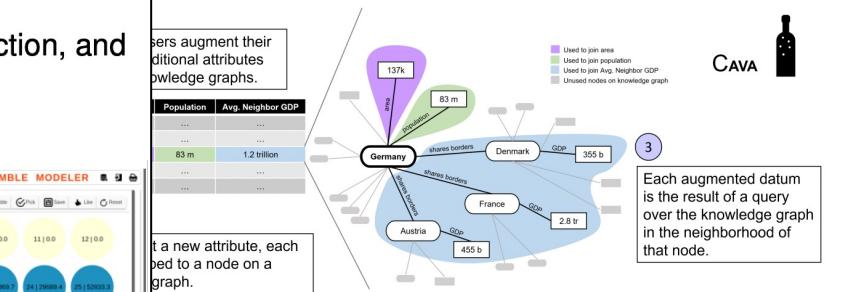
Subhajit Das, Dylan Cashman, Remco Chang, Alex Endert



https://dylancashman.github.io/public/docs/beames_ensemble.pdf

CAVA: A Visual Analytics System for Exploratory Columnar Data Augmentation Using Knowledge Graphs

Dylan Cashman, Shenyu Xu, Subhajit Das, Florian Heimerl, Cong Liu, Shah Rukh Humayoun, Michael Gleicher, Alex Endert, Remco Chang



<https://dylancashman.github.io/public/docs/cava.pdf>

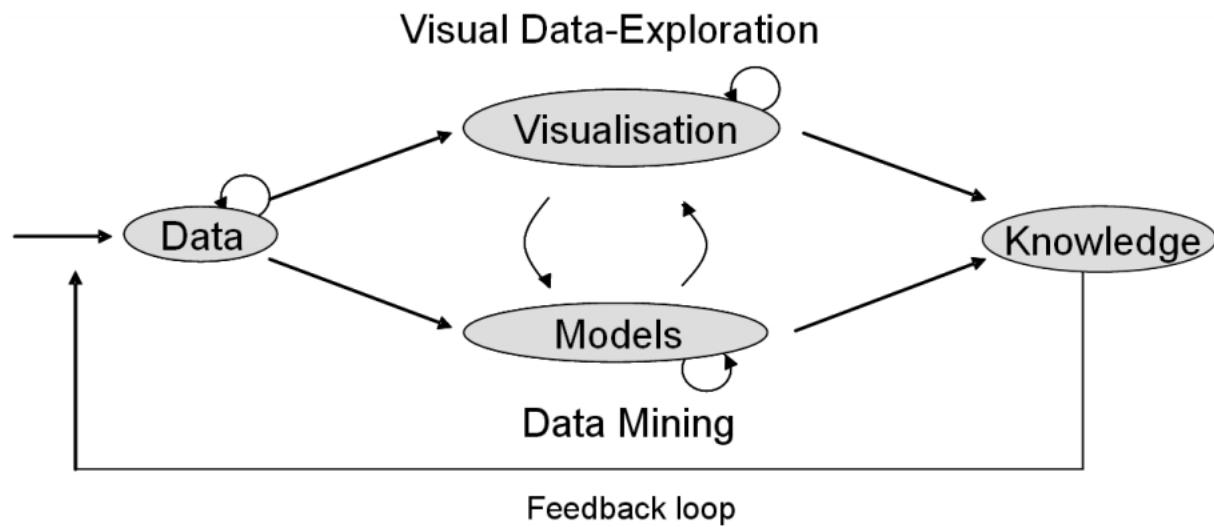
https://dylancashman.github.io/public/docs/ds_vds_2018_gaggle.pdf

Visualization (in CS)

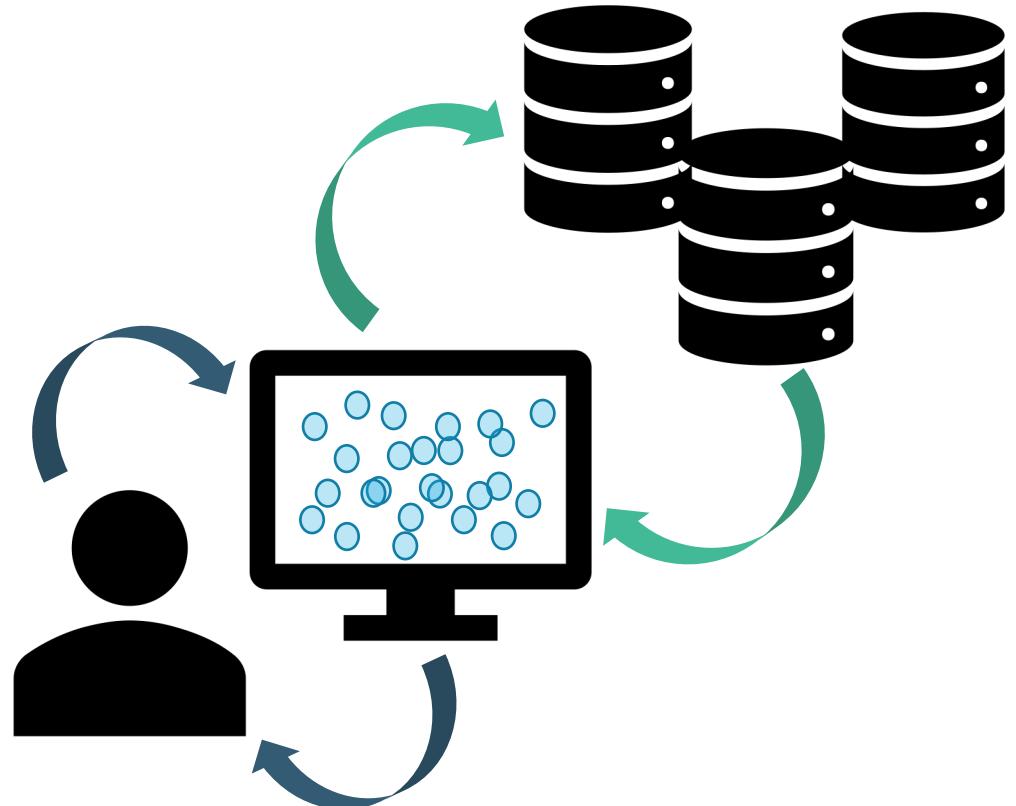
Data Analysis Pipeline -- Exploratory Data Analysis

Visualization (in CS)

Data Analysis Pipeline -- Exploratory Data Analysis



Keim et al.. Visual Analytics: Definition, Process and Challenges (2008).



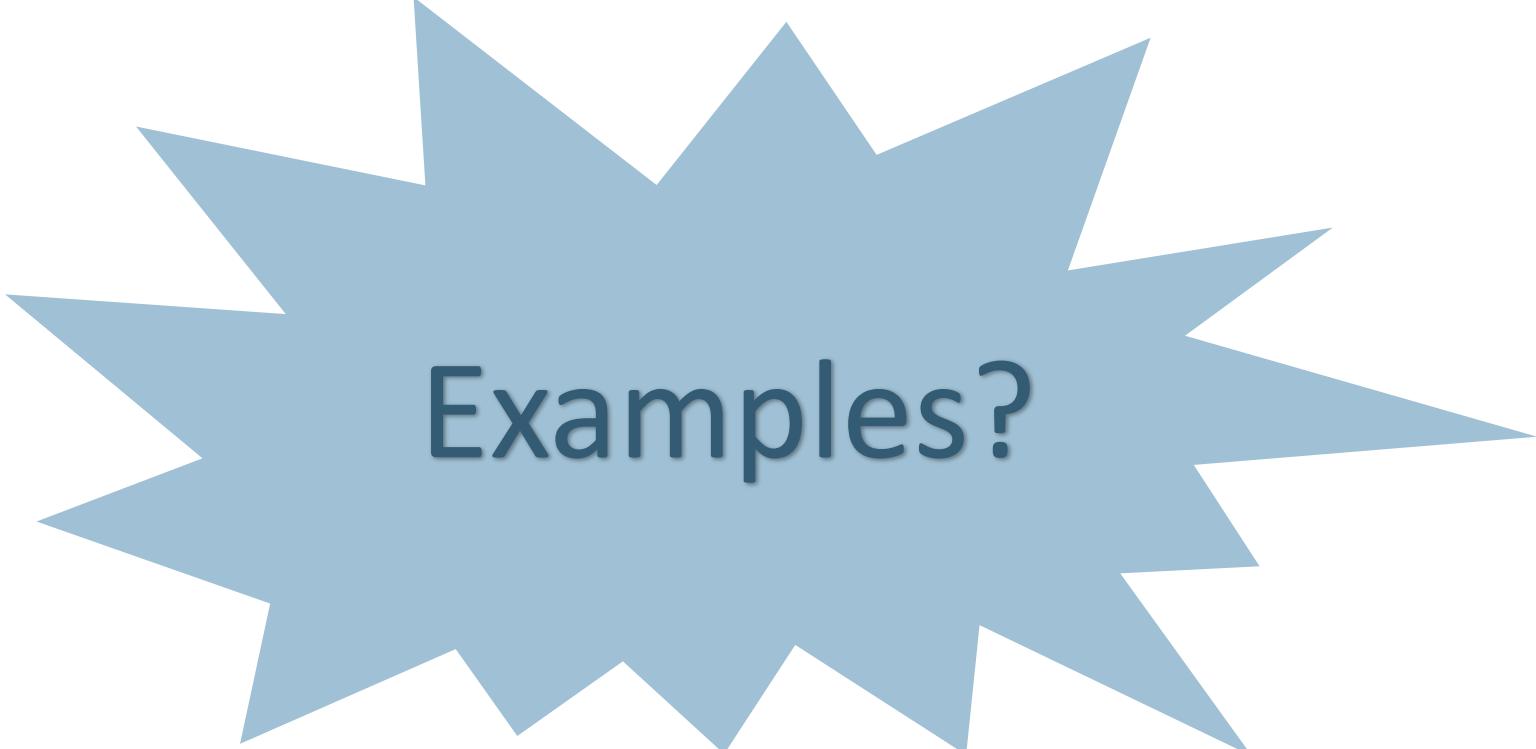
Visualization (in CS)

Medium of Communication

As an alternative to written communication or numerical communication

Visualization (in CS)

Medium of Communication



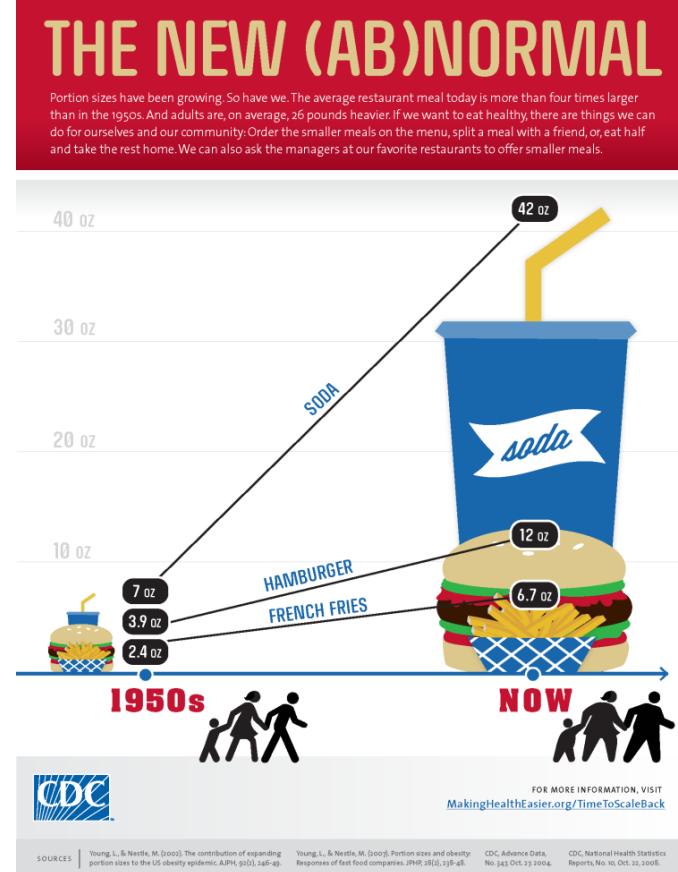
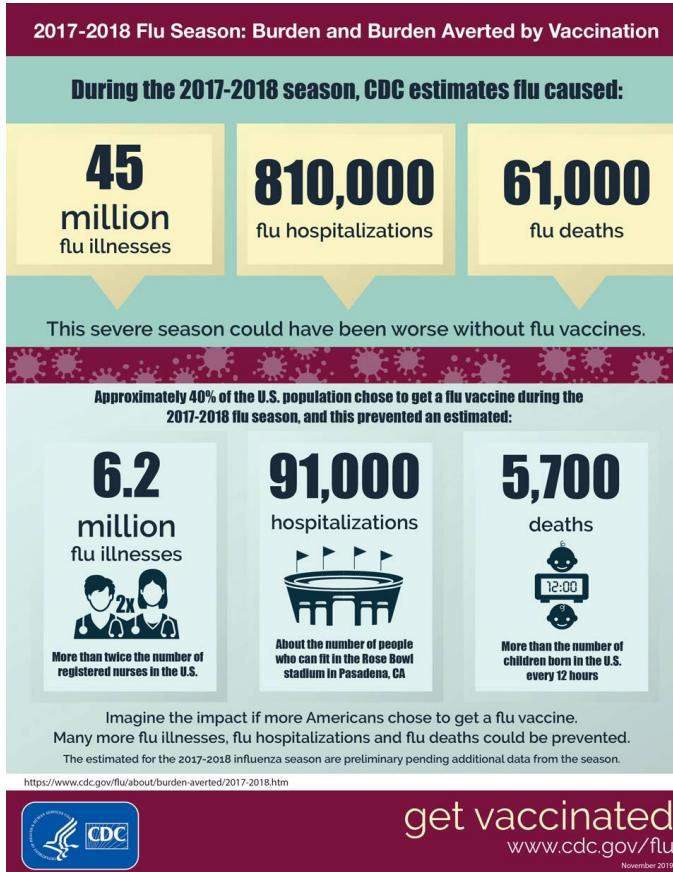
Examples?

Visualization (in CS)

Medium of Communication -- Public Health

Visualization (in CS)

Medium of Communication -- Public Health



Visualization (in CS)

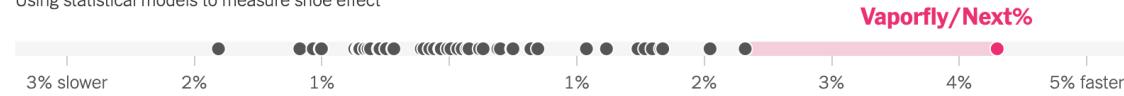
Medium of Communication -- News reporting

Visualization (in CS)

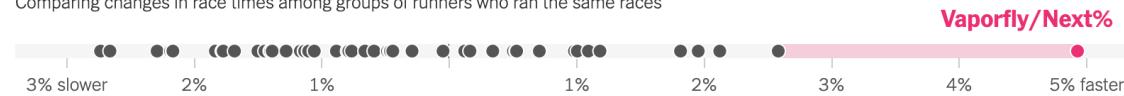
Medium of Communication -- News reporting

How Vaporflys and Next% shoes compare with other popular running shoes:

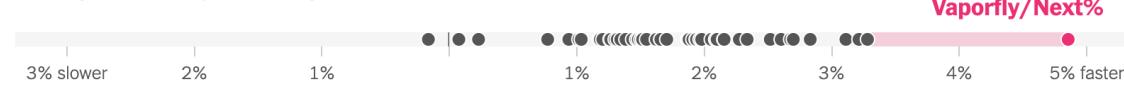
Using statistical models to measure shoe effect



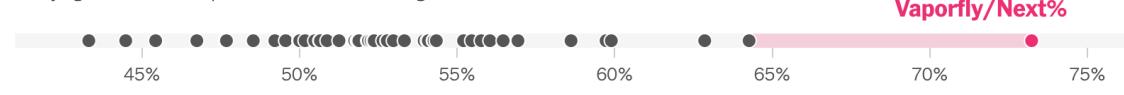
Comparing changes in race times among groups of runners who ran the same races



Following runners as they switch racing shoes

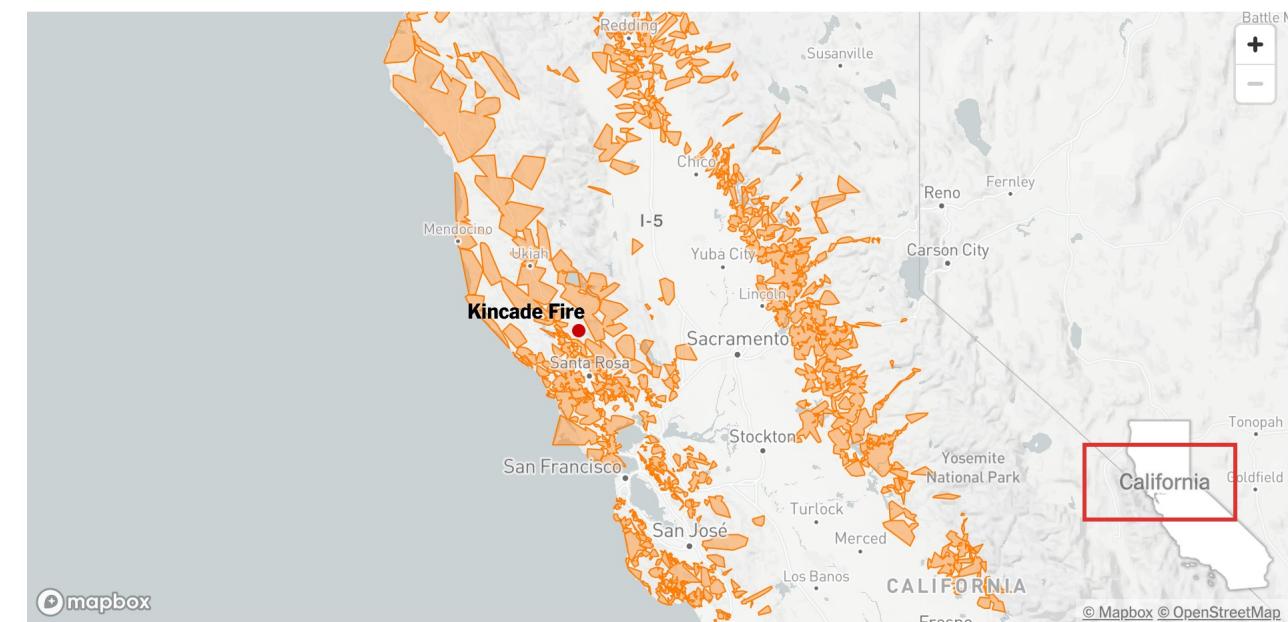


Studying the chance of a personal best while wearing certain shoes



<https://www.nytimes.com/interactive/2019/12/13/upshot/nike-vaporfly-next-percent-shoe-estimates.html>

Power Outage Areas



<https://www.nytimes.com/interactive/2019/10/25/us/california-fire-map.html>

Visualization (in CS)

Medium of Communication -- Medical communication

Visualization (in CS)

Medium of Communication

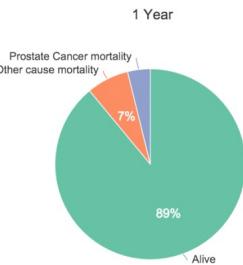
-- Medical communication

How big of a threat is my prostate cancer?

In most cases, prostate cancer **progresses slowly**, and is not lethal.

In all cases, **you have time to think carefully** about what to do about your prostate cancer, before you decide.

For example, based on the data you provided **your risk of dying from prostate cancer after 1 year is 4%**, illustrated in the following chart:



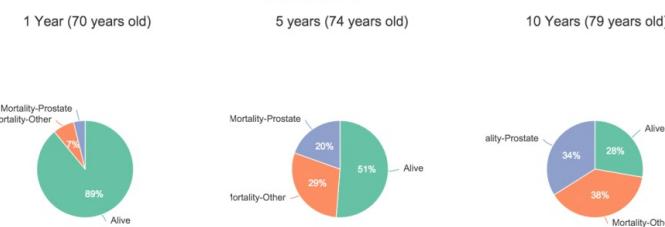
- Your chances of being **alive** (**in GREEN**)
- Your chances of dying from your **prostate cancer** (**in PURPLE**)
- Your chances of dying from **other causes** (**in ORANGE**)

Before thinking about the benefits of specific treatments, it is helpful to first think about **how big of a threat** your prostate cancer is to your future survival.

How big of a threat is my prostate cancer?

Before thinking about the benefits of specific treatments, it's helpful to first think about how big of threat your prostate cancer is to your future survival. The pie chart below shows the following:

- Your chances of being **alive** (**in GREEN**)
- Your chances of dying from your **prostate cancer** (**in PURPLE**)
- Your chances of dying from **other causes** (**in ORANGE**)

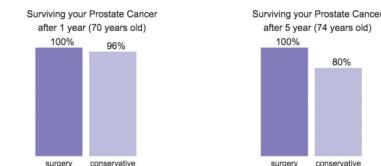


How effective are different treatments for my prostate cancer?

The expected benefits from **surgery** and **conservative management** are listed below.

These results show your estimated chances of either surviving or dying from **your prostate cancer** at 1, 5, and 10 years, depending on whether you choose either **surgery (DARK PURPLE BAR)** or **conservative treatment (LIGHT PURPLE BAR)**.

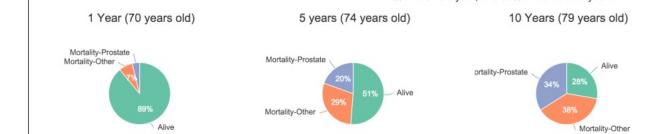
You can view these risks in terms of either survival or mortality.



What do I do next?

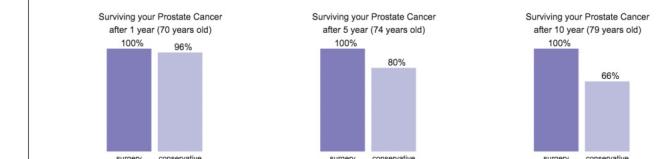
What threats am I facing?

- Your risk of dying from prostate cancer: 4% in the next year, and 34% in the next 10 years.
- Your risk of dying from other causes: 7% in the next year, and 38% in the next 10 years.



What are the possible treatments for me?

- Surgery
- Conservative management
- Radiation
- Other treatments



Visualization (in CS)

Medium of Communication

- Public Health
- News Reporting
- Medical communication
- etc.

Use visualization
as an alternative
to written
communication or
numerical
communication

How does visualization for communication differ from other types of visualization?

Comparison

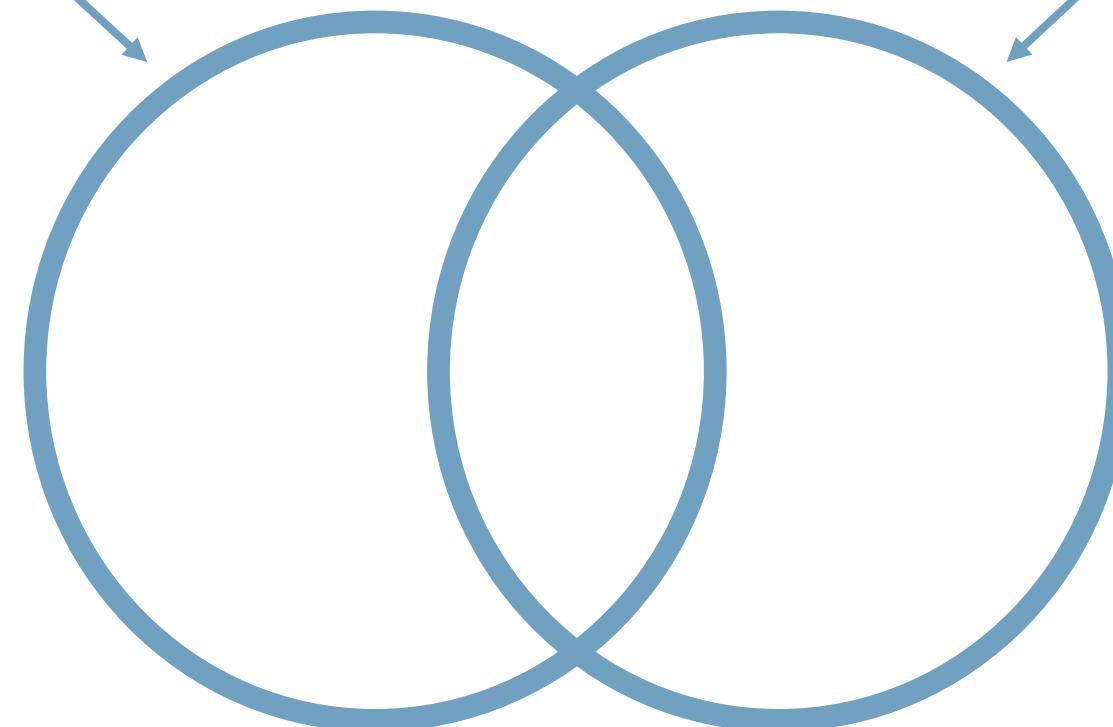
Data Analysis Pipeline

Medium of Communication

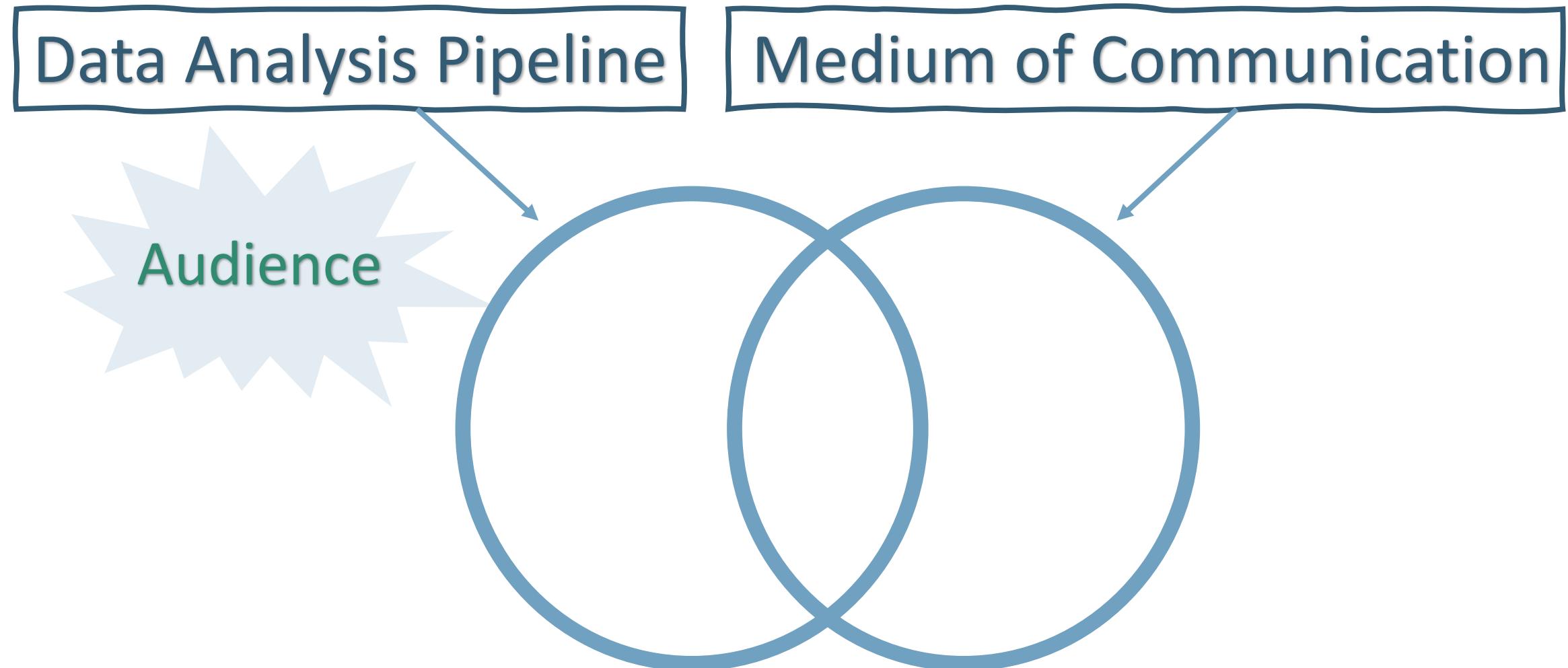
Comparison

Data Analysis Pipeline

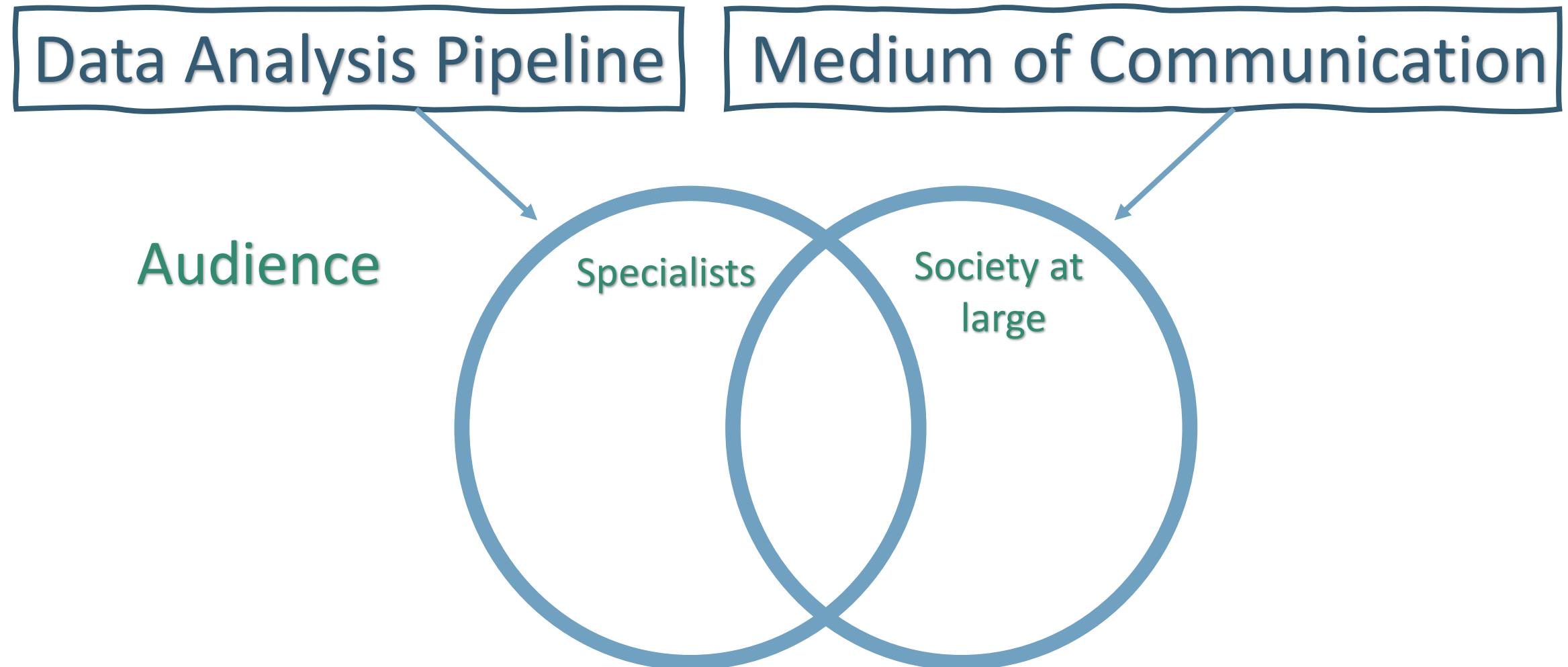
Medium of Communication



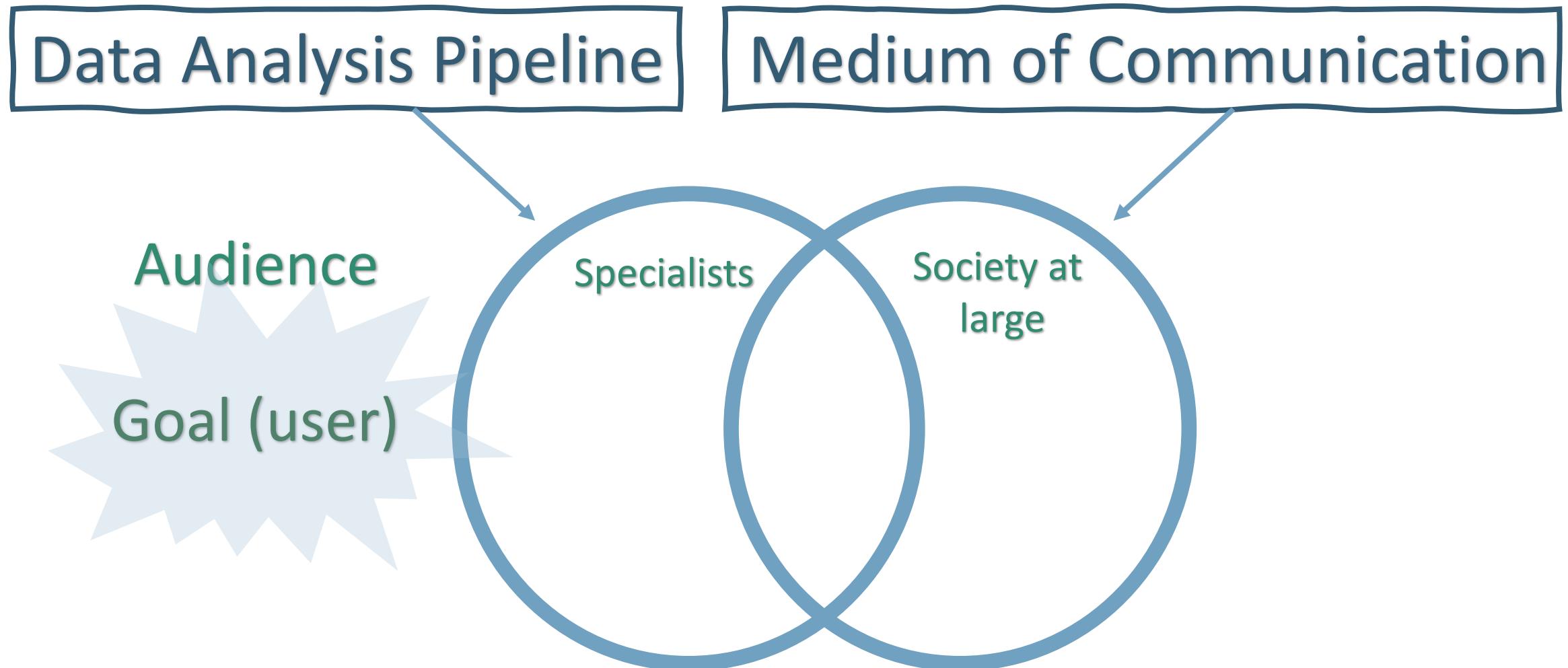
Comparison



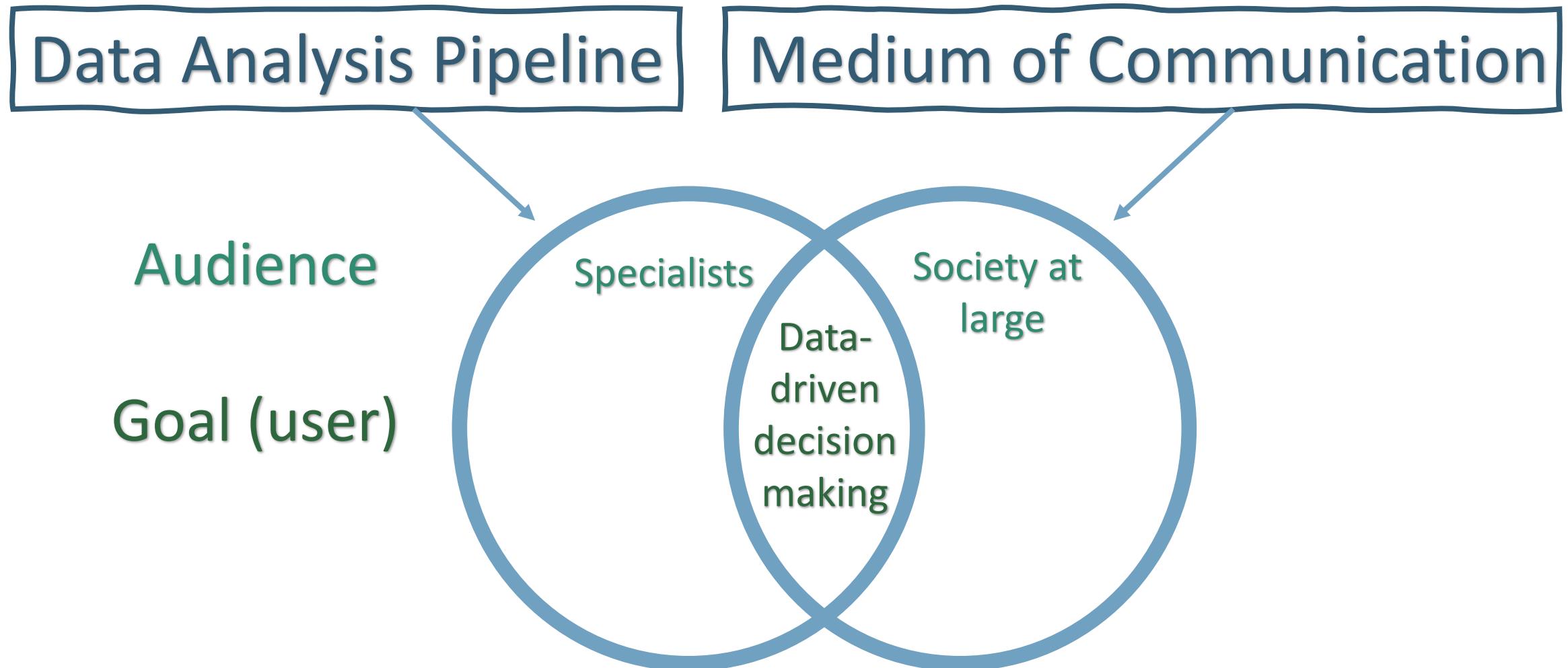
Comparison



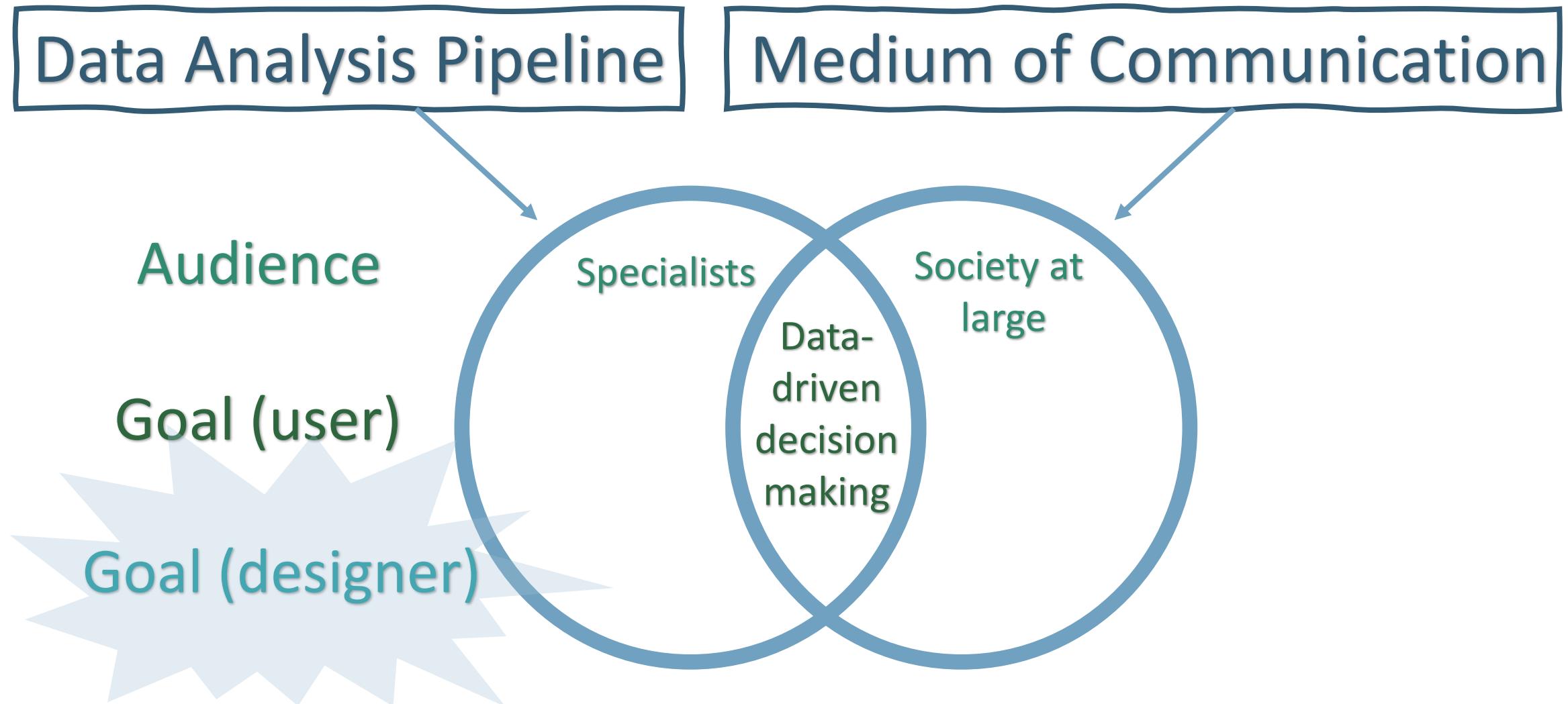
Comparison



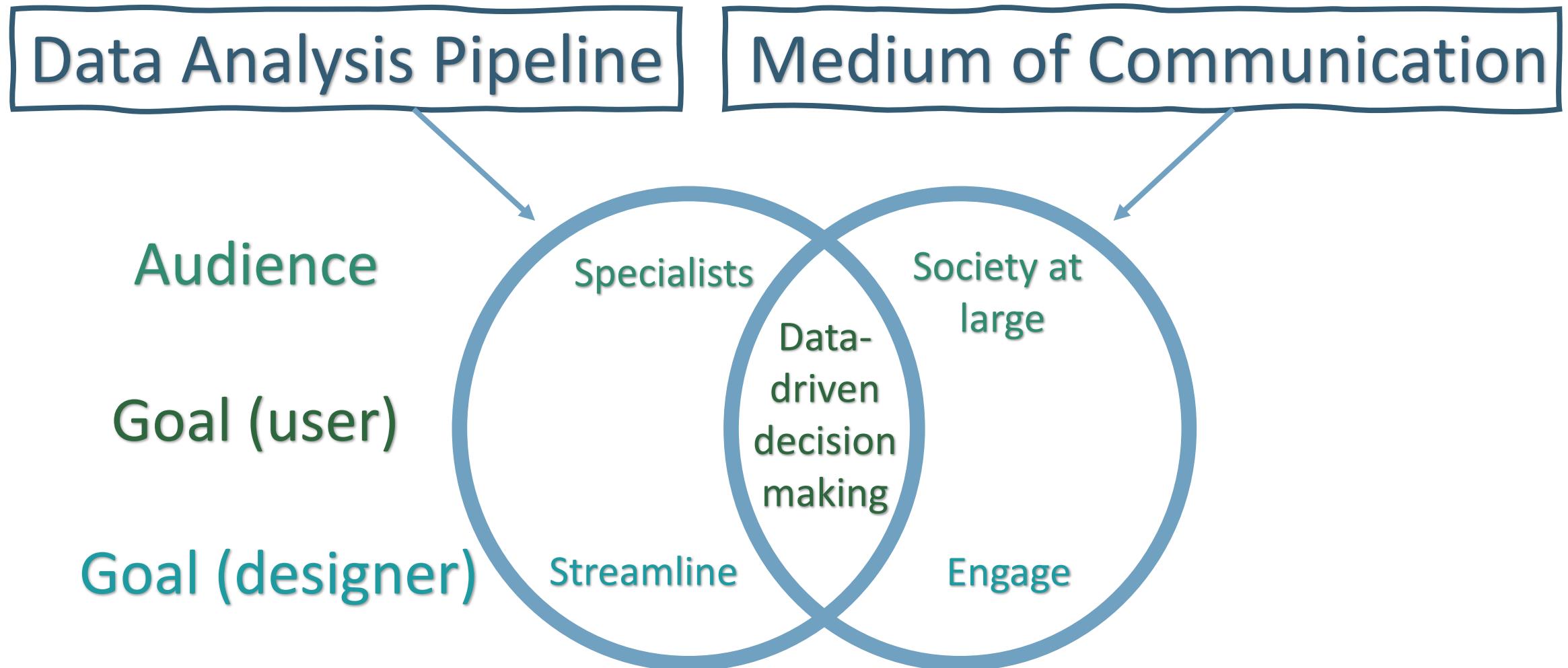
Comparison



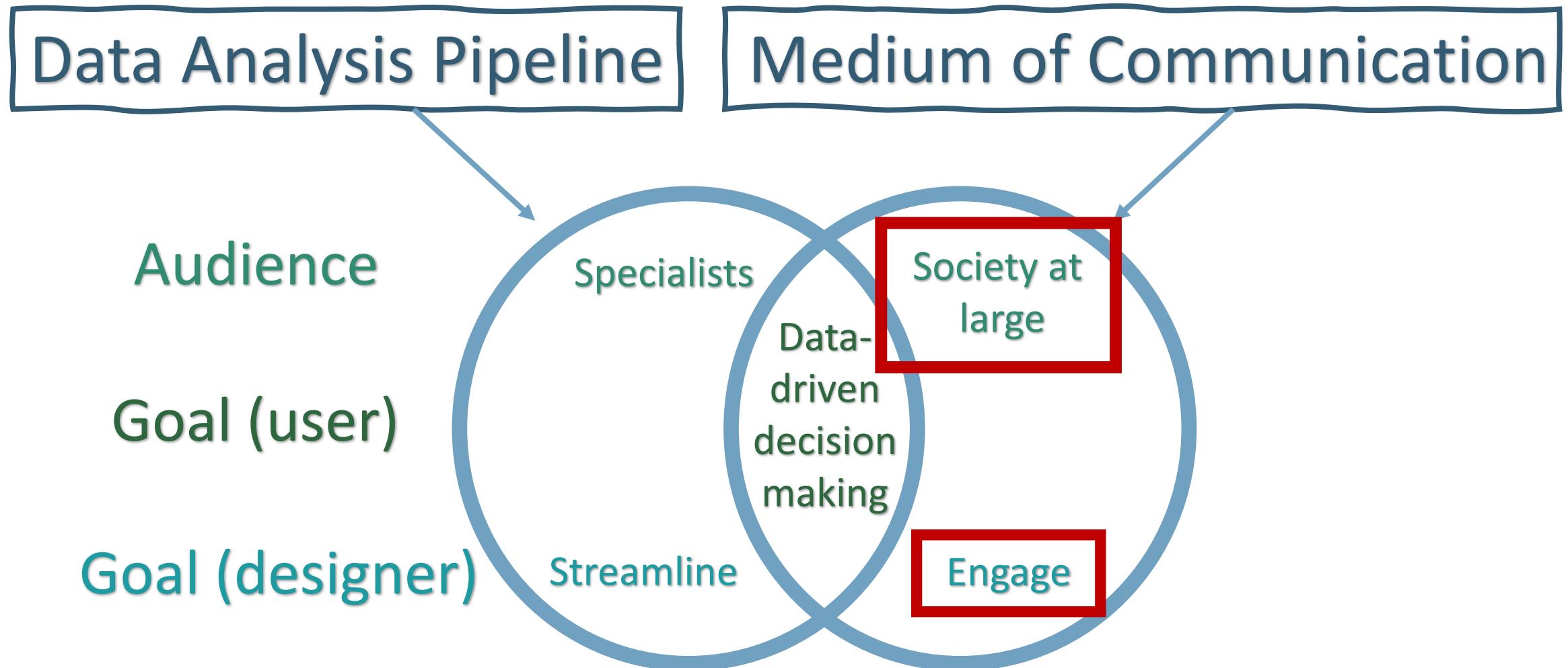
Comparison



Comparison



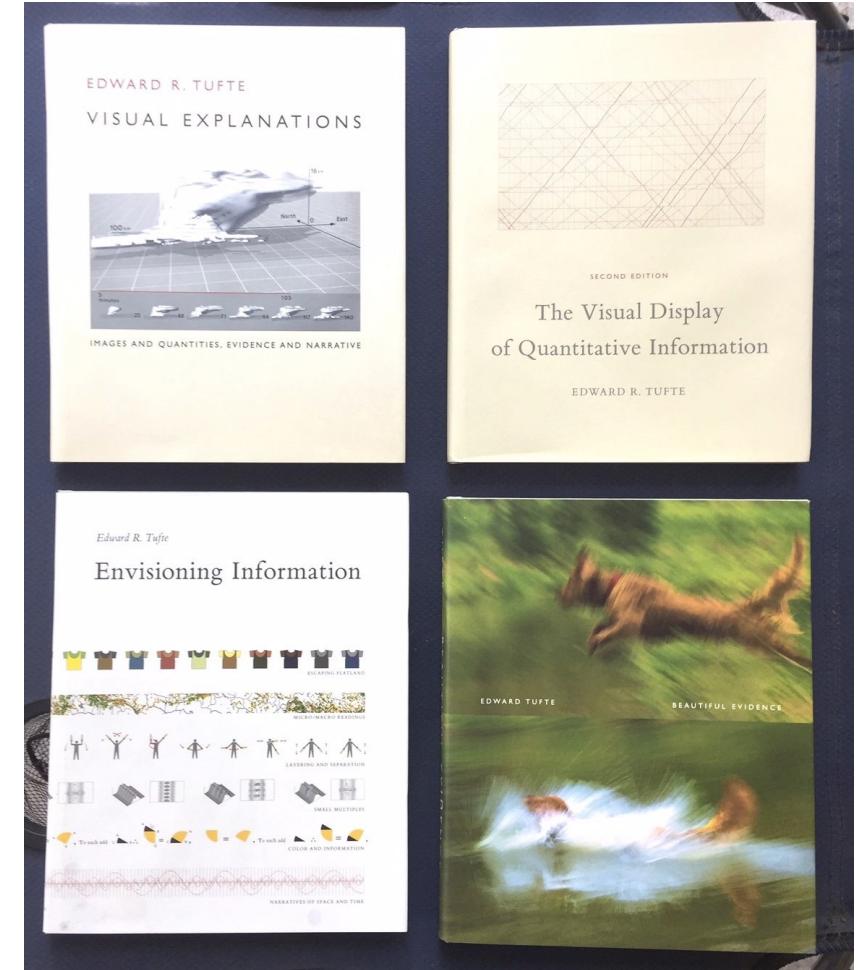
Comparison



Visualization for
communication design

Historically – Tufte’s “Graphical Integrity”

- Essentially, Tufte’s Design Rules of Thumb
- Widely known, but not always adhered to



Historically – Tufte’s “Graphical Integrity”

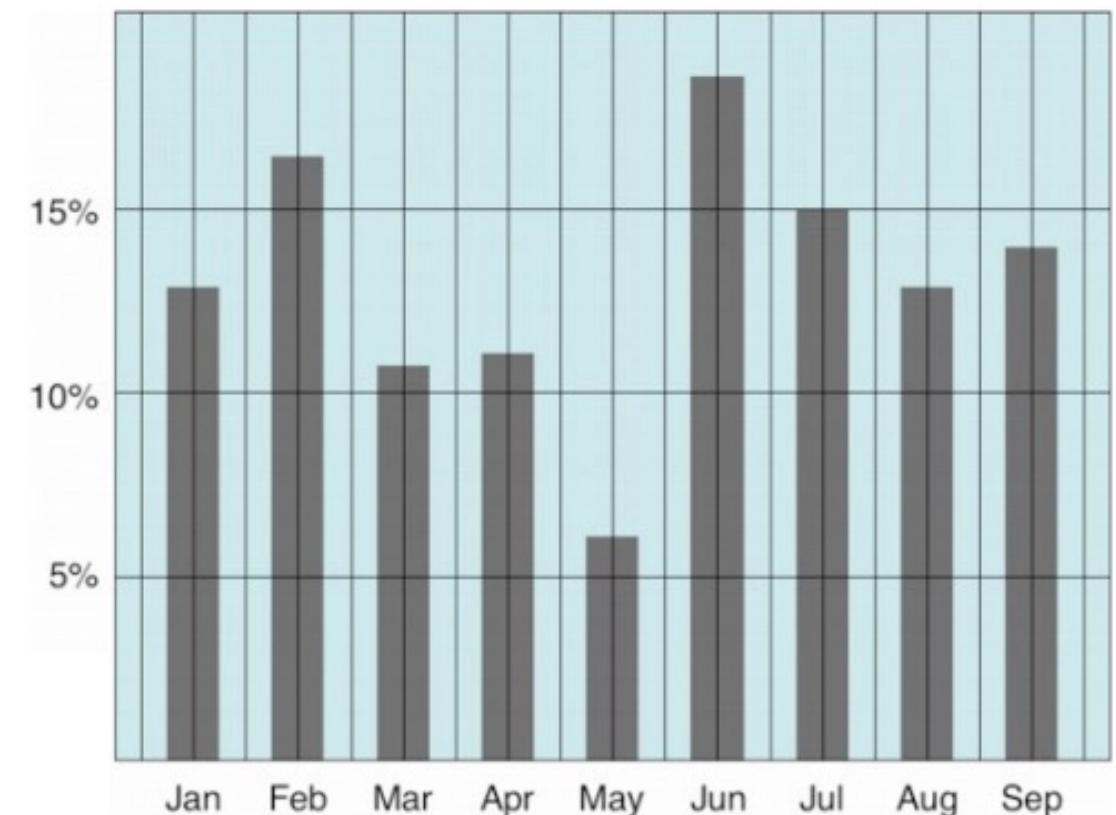
To achieve graphical “excellence”:

1. Above all else show the data.
2. Maximize data to ink ratio.
3. Erase non-data ink.
4. Erase redundant data ink.
5. Revise and edit.

Historically – Tuft's “Graphical Integrity”

To achieve graphical “excellence”:

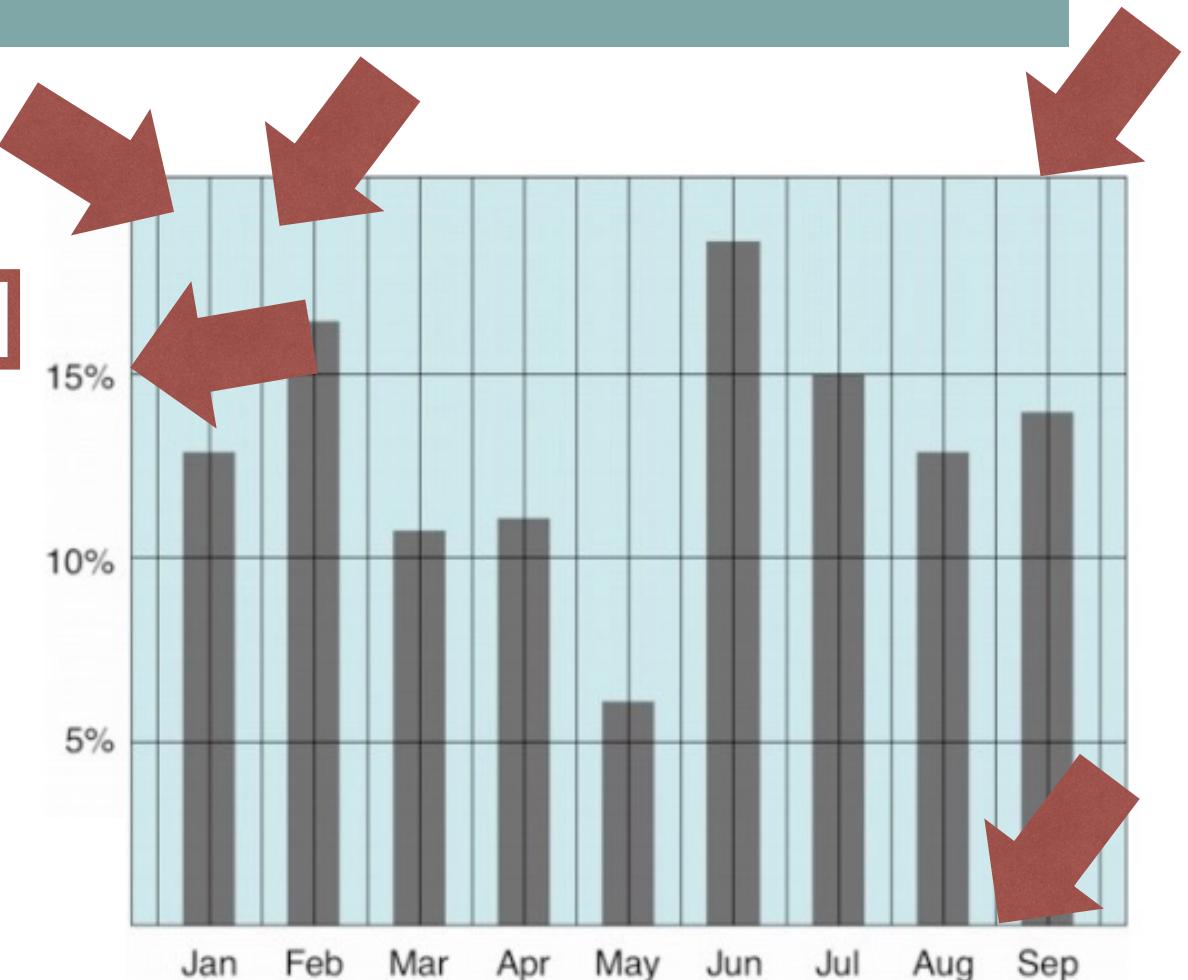
1. Above all else show the data.
2. Maximize data to ink ratio.
3. Erase non-data ink.
4. Erase redundant data ink.
5. Revise and edit.



Historically – Tuft's “Graphical Integrity”

To achieve graphical “excellence”:

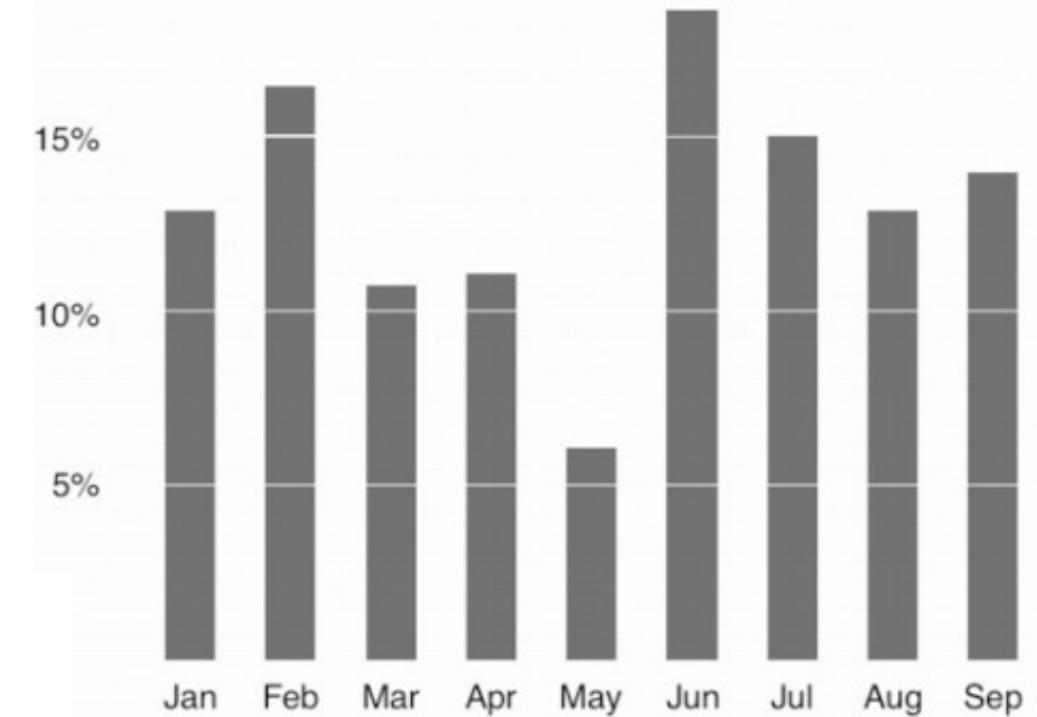
1. Above all else show the data.
2. Maximize data to ink ratio.
3. Erase non-data ink.
4. Erase redundant data ink.
5. Revise and edit.



Historically – Tuft's “Graphical Integrity”

To achieve graphical “excellence”:

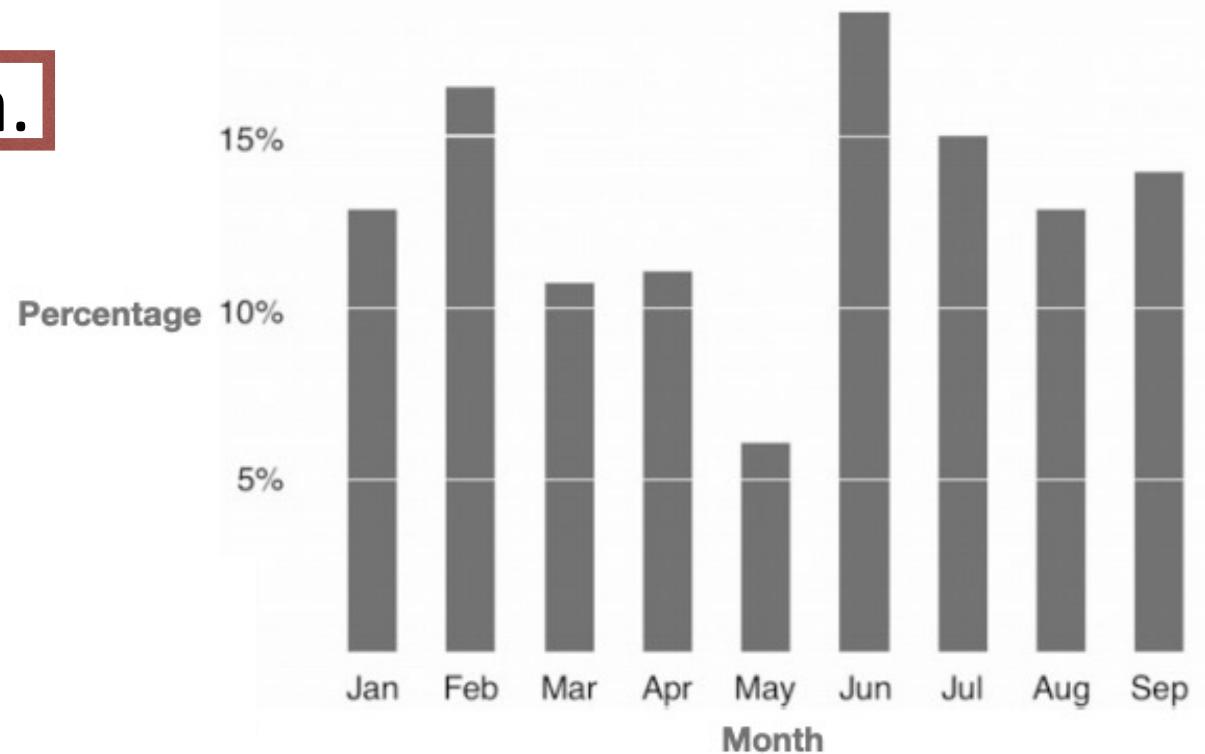
1. Above all else show the data.
2. Maximize data to ink ratio.
3. Erase non-data ink.
4. Erase redundant data ink.
5. Revise and edit.



Historically – Tufte’s “Graphical Integrity”

To achieve graphical “excellence”:

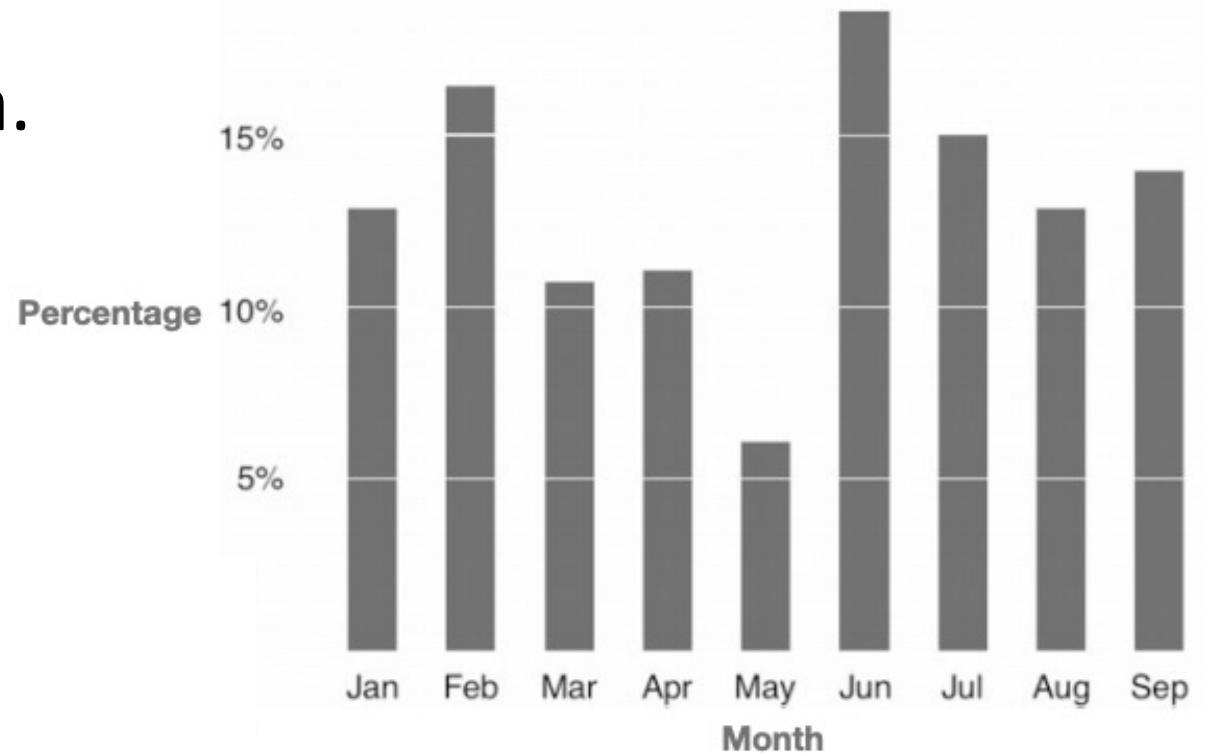
1. Above all else show the data.
2. Maximize data to ink ratio.
3. Erase non-data ink.
4. Erase redundant data ink.
5. Revise and edit.



Historically – Tufte’s “Graphical Integrity”

To achieve graphical “excellence”:

1. Above all else show the data.
2. Maximize data to ink ratio.
3. Erase non-data ink.
4. Erase redundant data ink.
5. Revise and edit.



Historically – Tufte’s “Graphical Integrity”

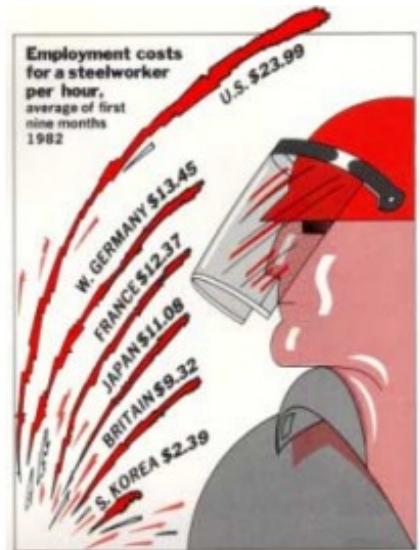
To achieve graphical “excellence”:

1. Above all else show the data.
2. Maximize data to ink ratio.
3. Erase non-data ink.
4. Erase redundant data ink.
5. Revise and edit.



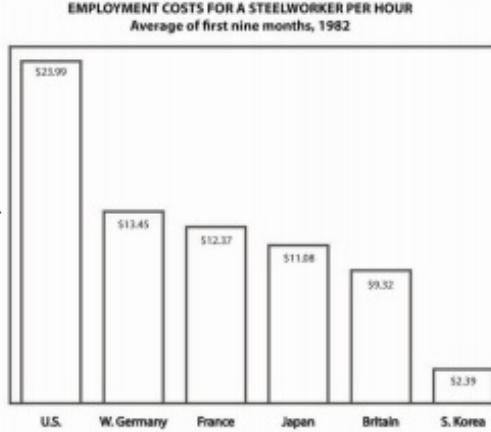
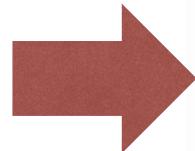
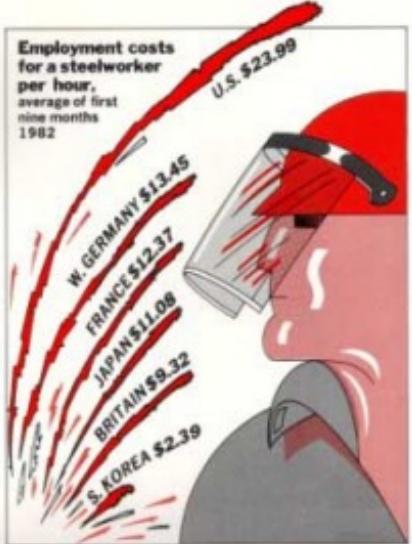
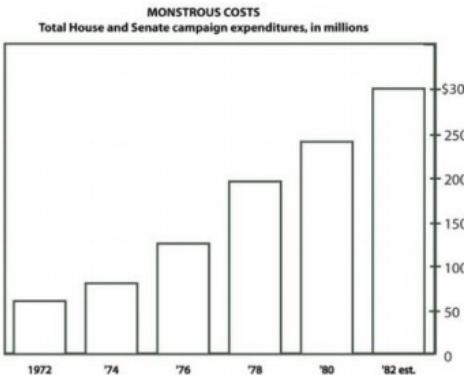
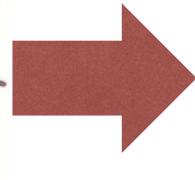
No “Chart Junk” i.e.
no elements that do
not communicate data.

Open Question – Chart Junk



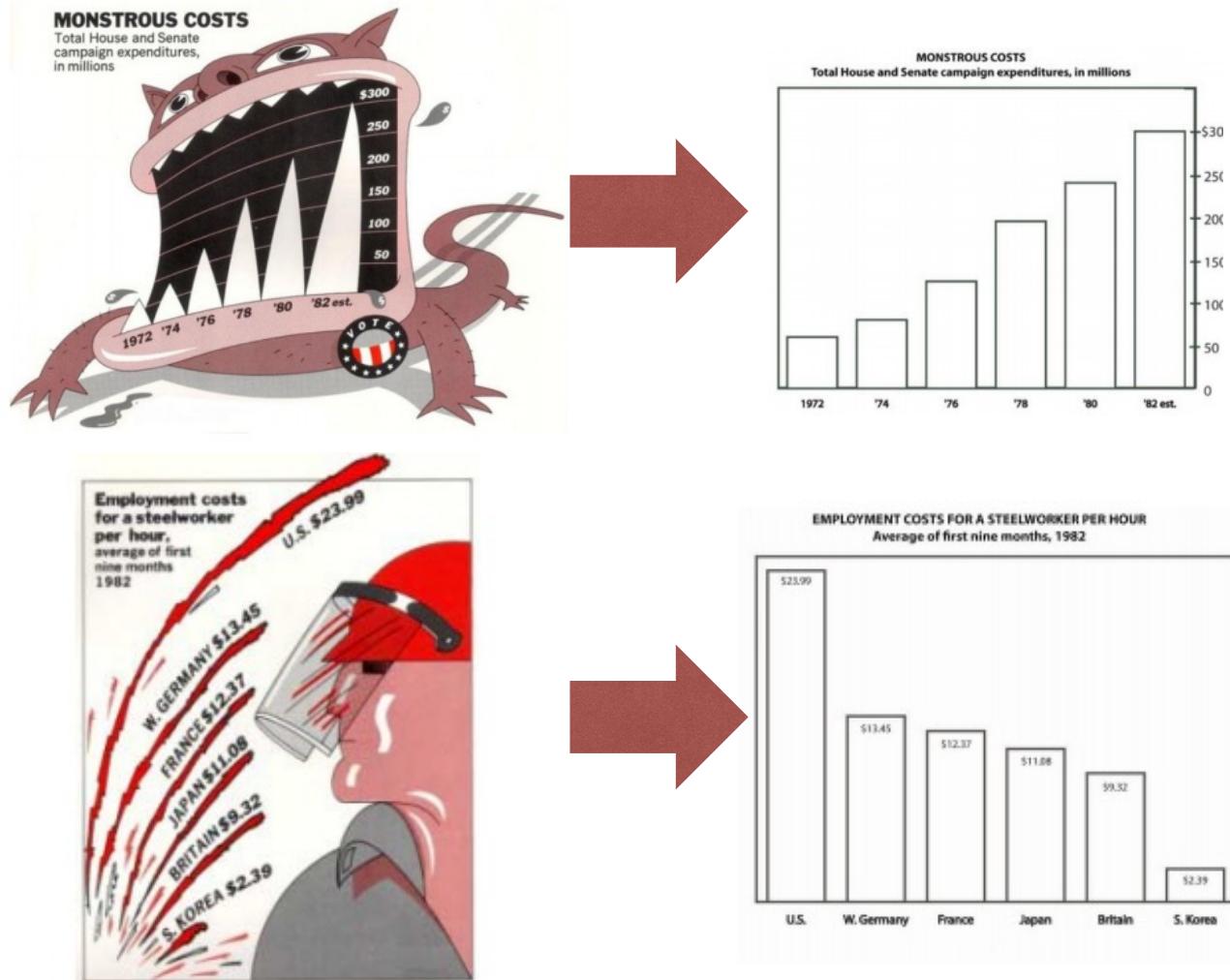
Useful Junk? The effects of visual embellishment on comprehension and memorability of charts. Bateman et al.

Open Question – Chart Junk



Useful Junk? The effects of visual embellishment on comprehension and memorability of charts. Bateman et al.

Open Question – Chart Junk



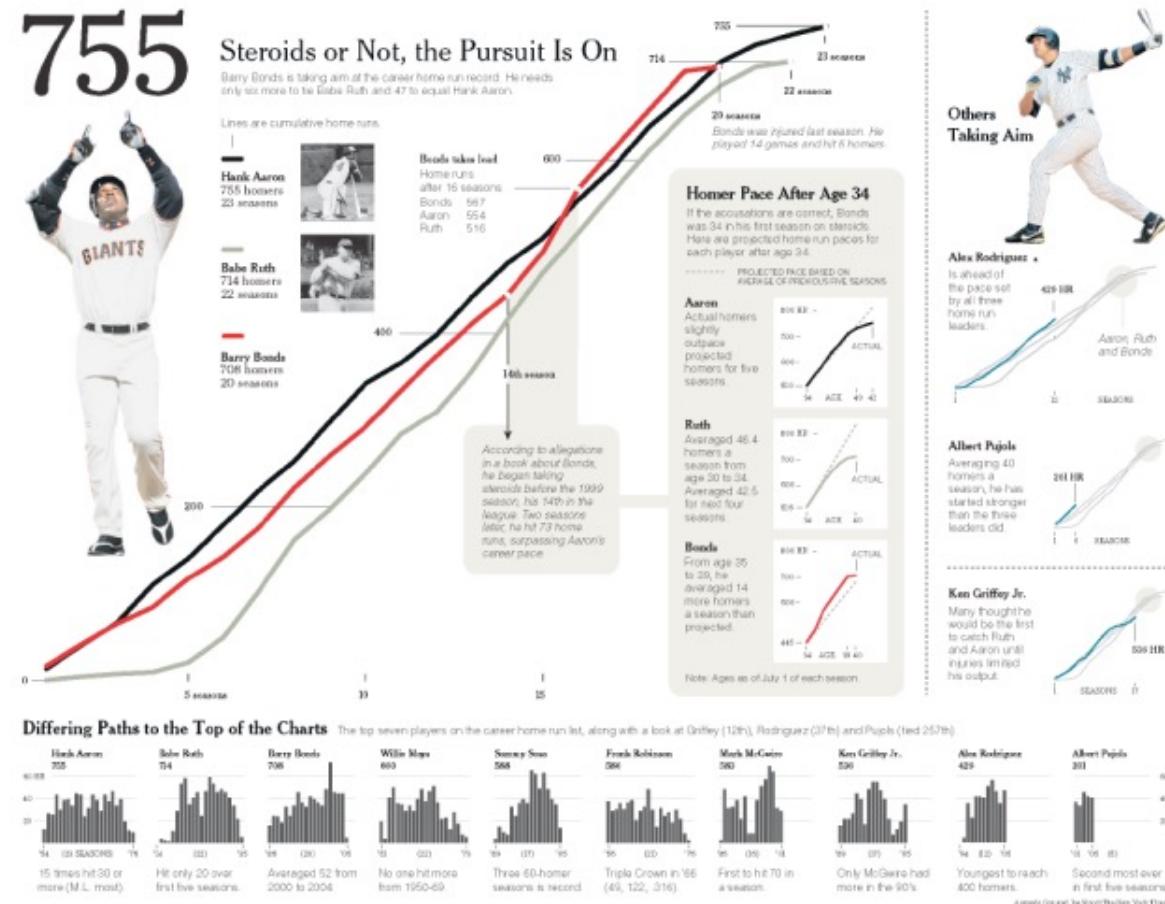
Debate Amongst Contemporary Researchers

- + Chart junk has been shown to persuade, help with memorability, engage users
- Chart junk has been shown to cause bias, reduce data-ink ratio, clutter, degrade user trust

Useful Junk? The effects of visual embellishment on comprehension and memorability of charts. Bateman et al.

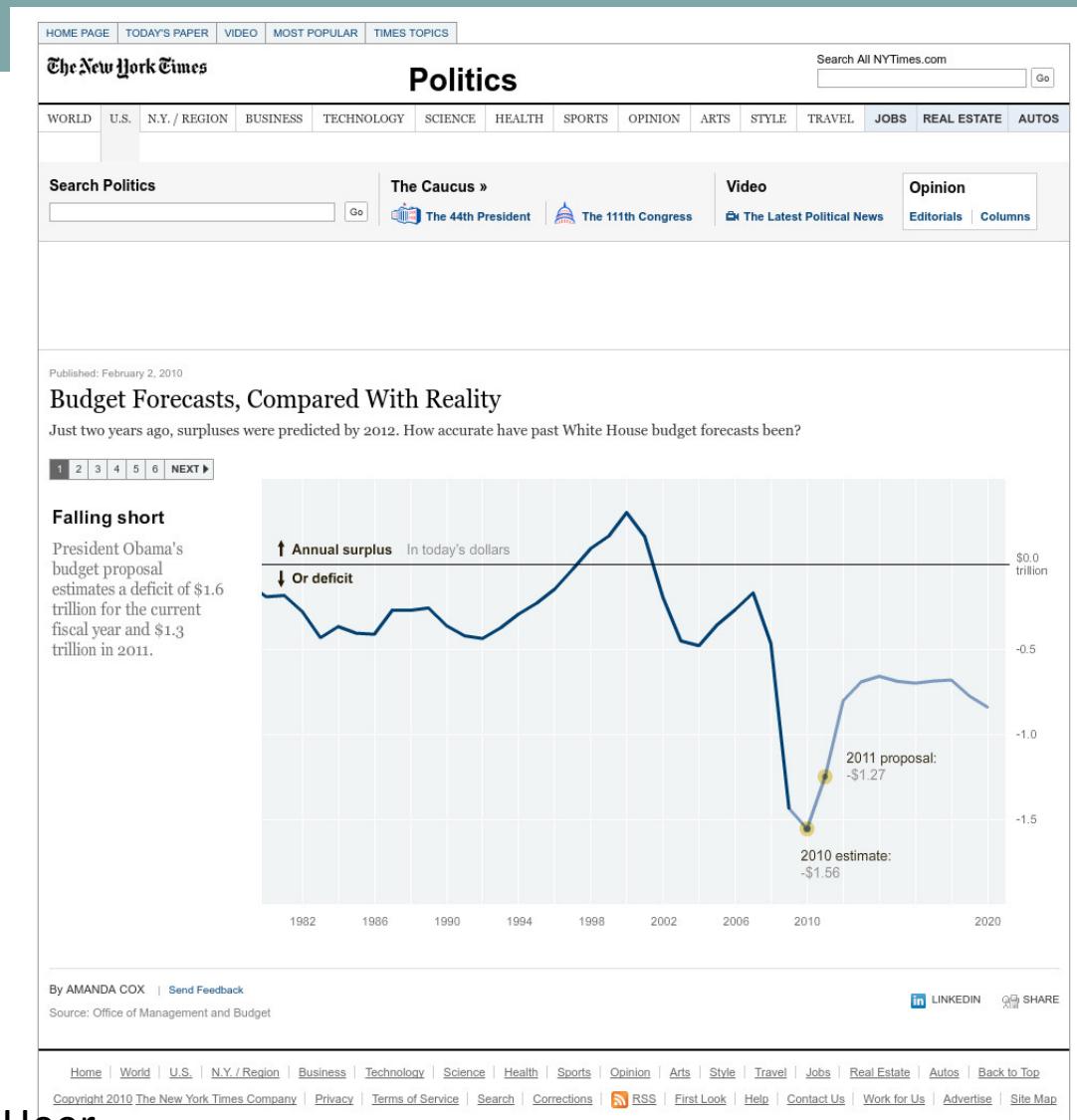
Storytelling – Narrative Visualization

Visualizations that combine narratives with interactive graphics.



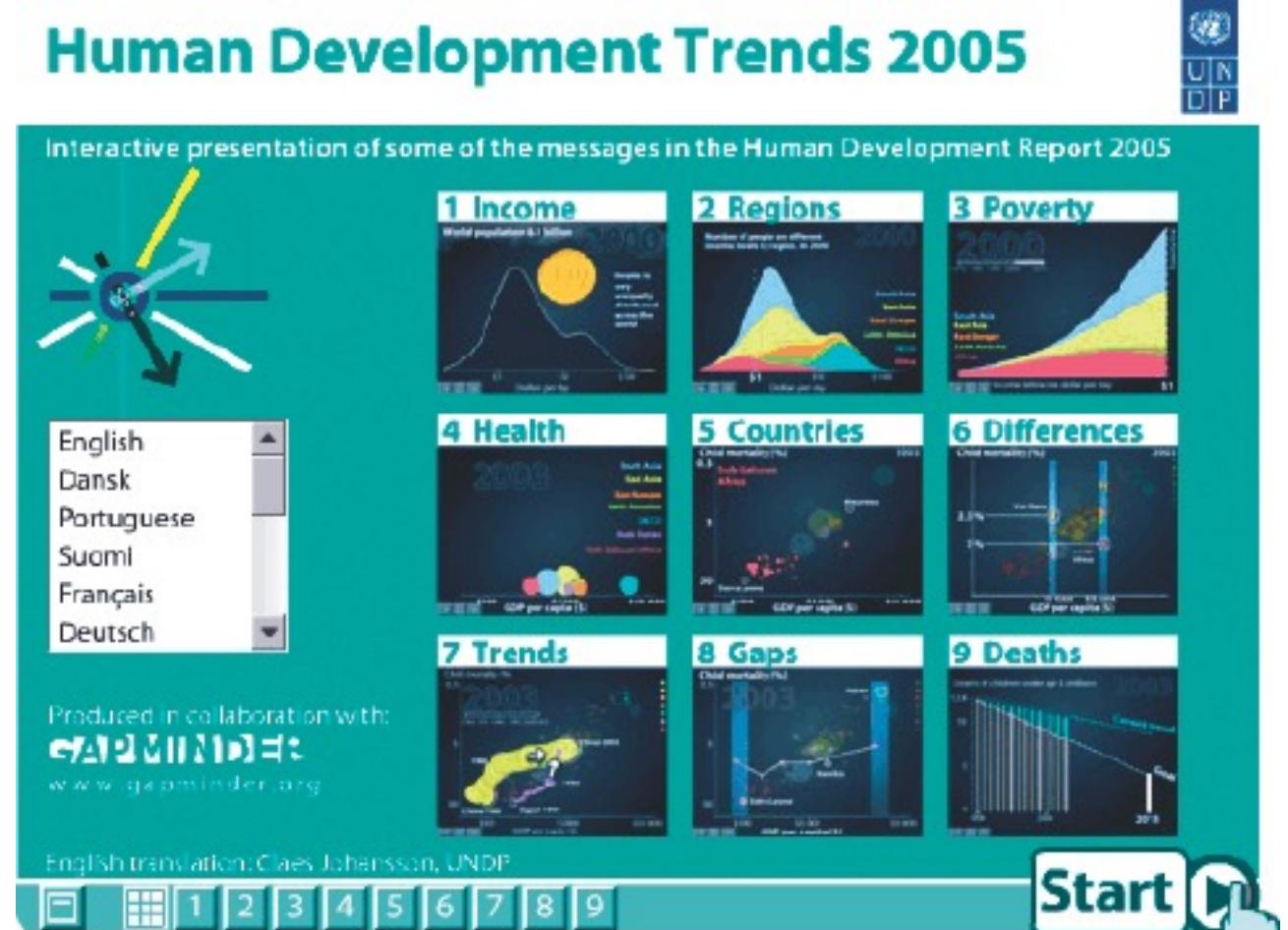
Storytelling – Narrative Visualization

Visualizations that combine narratives with interactive graphics.

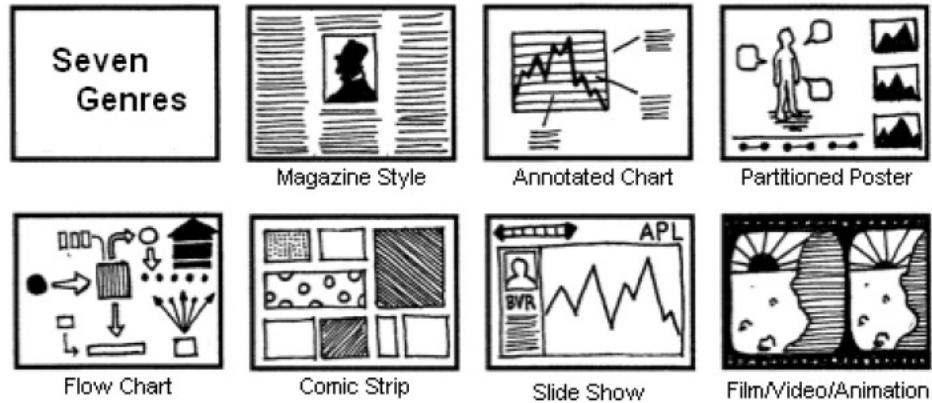


Storytelling – Narrative Visualization

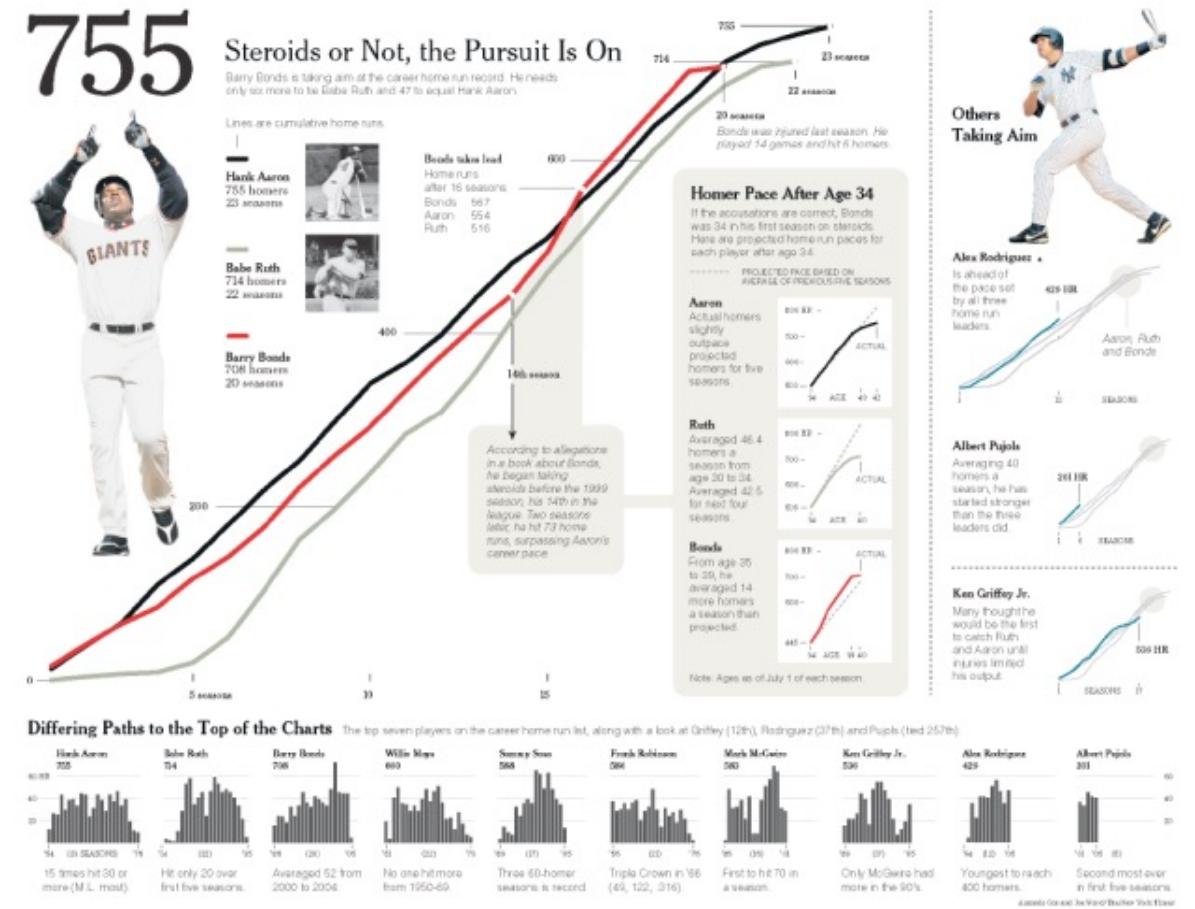
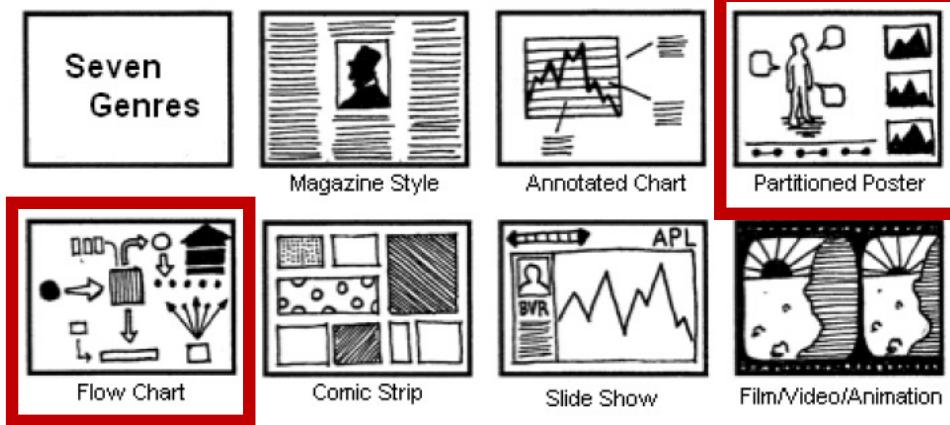
Visualizations that combine narratives with interactive graphics.



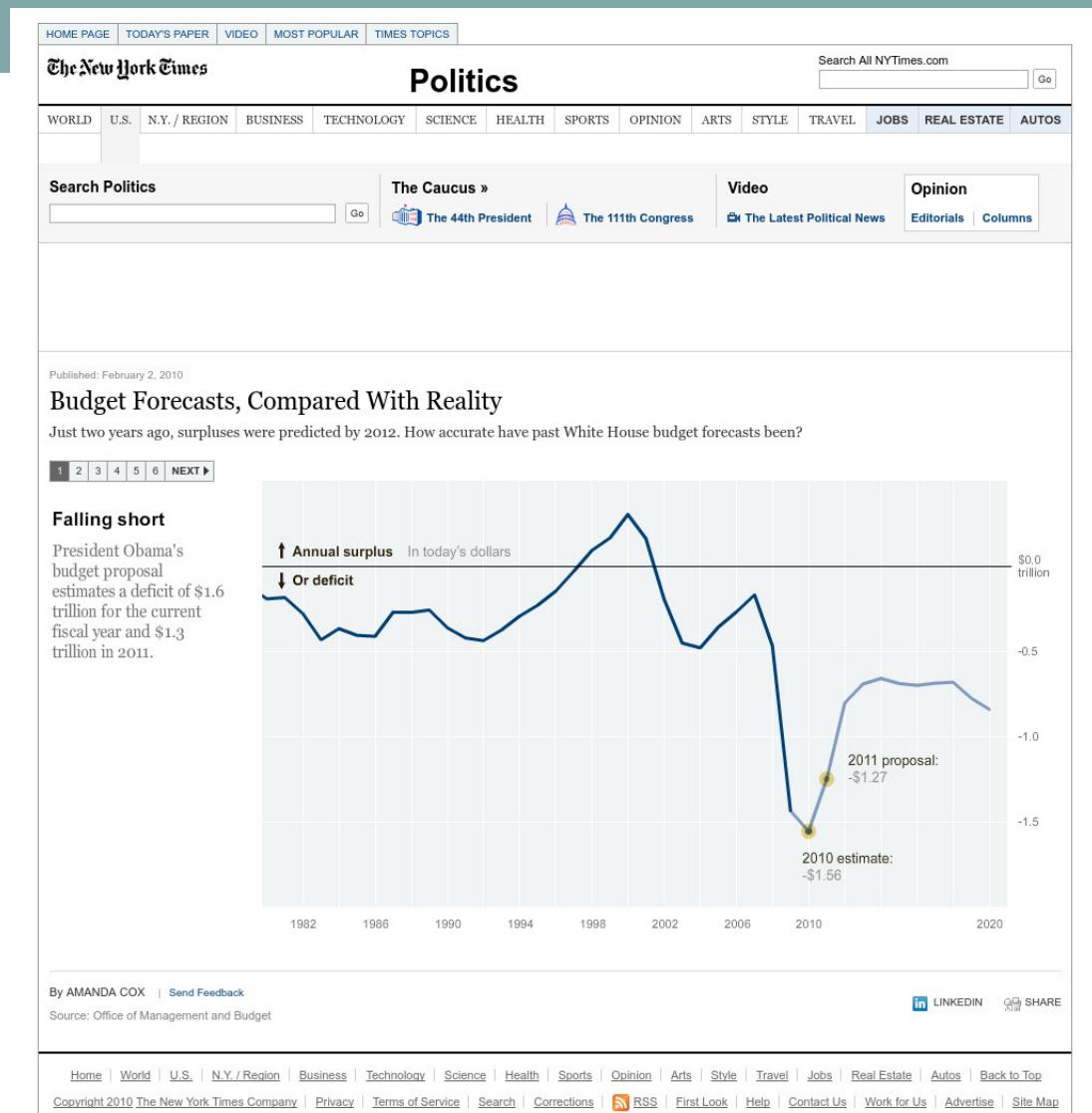
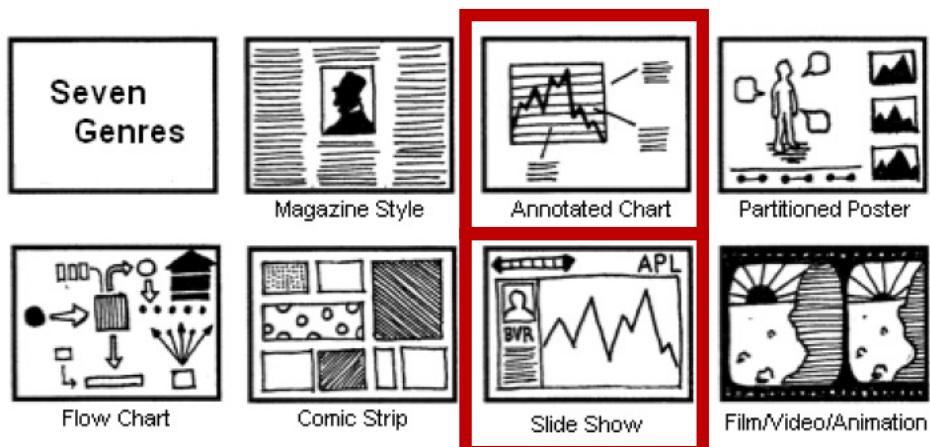
Storytelling – Narrative Visualization



Storytelling – Narrative Visualization



Storytelling – Narrative Visualization



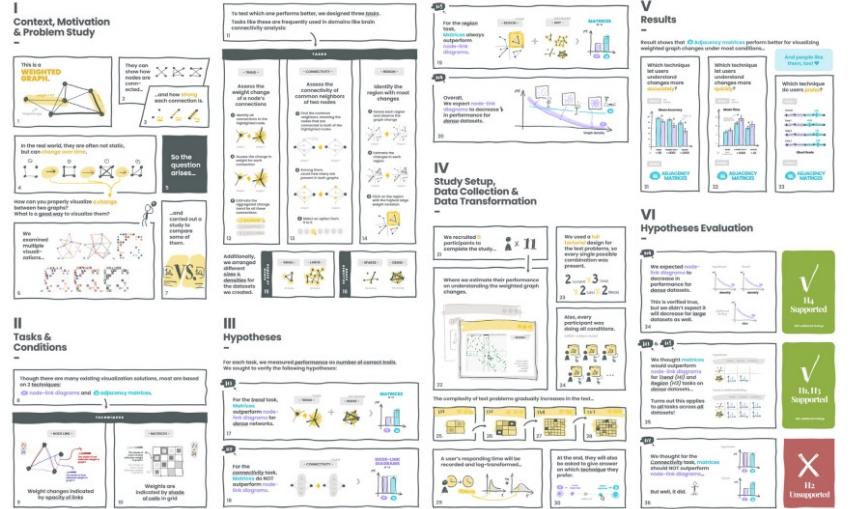
Storytelling – Narrative Visualization



Author Driven

Reader Driven

Storytelling – Narrative Visualization



Data Comics

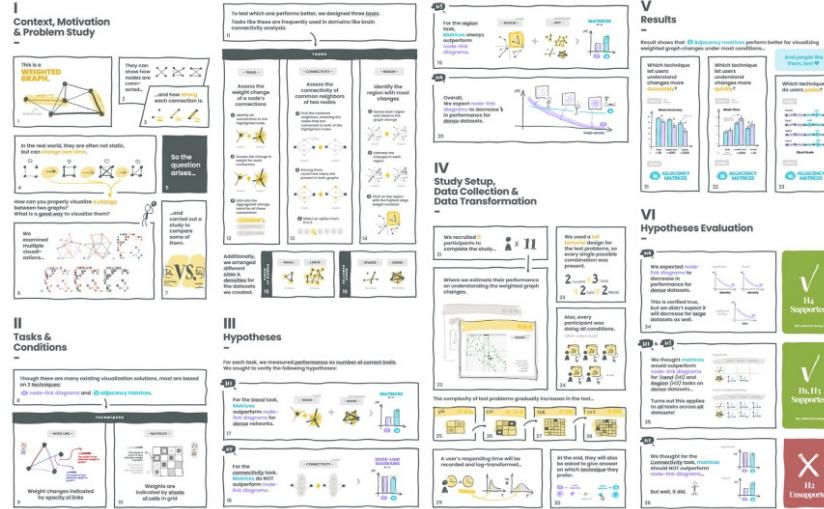
- Strict linear ordering of scenes
- Heavy messaging
- No interactivity



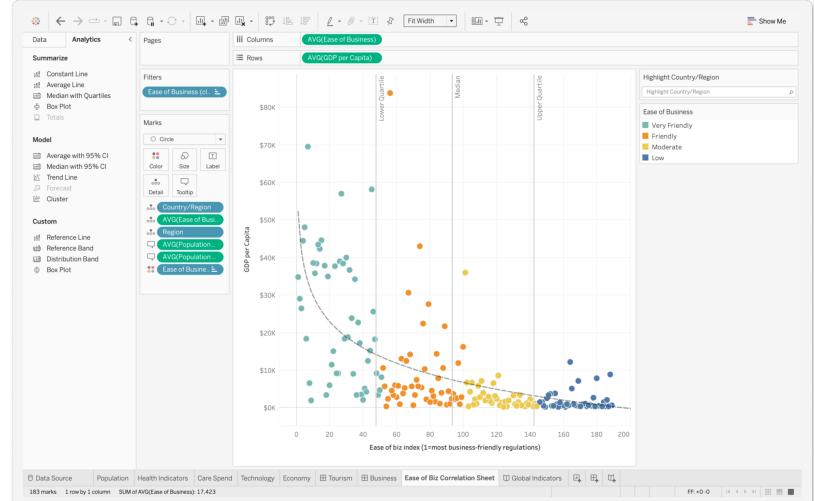
Author Driven

Reader Driven

Storytelling – Narrative Visualization



Data Comics



Tableau

- Strict linear ordering of scenes
- Heavy messaging
- No interactivity



Author Driven

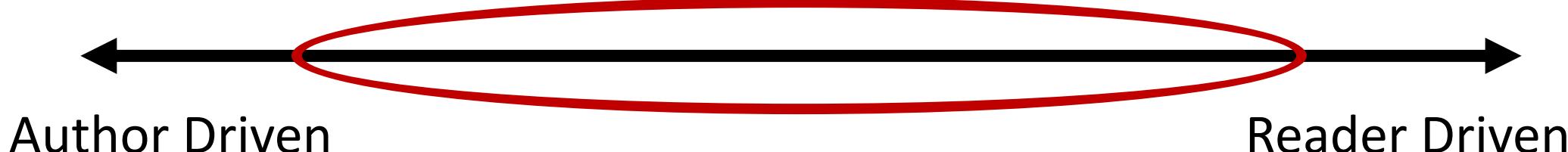
- No prescribed ordering
 - No messaging
 - Free interactivity

Reader Driven

Storytelling – Narrative Visualization

Martini Glass Structure Interactive Slideshow Drill-Down Story

- Strict linear ordering of scenes
 - Heavy messaging
 - No interactivity
- No prescribed ordering
 - No messaging
 - Free interactivity



Storytelling – Narrative Visualization

Martini Glass Structure
Interactive Slideshow
Drill-Down Story

In groups, brainstorm what each of these structures might look like. Sketch a (simple) example to illustrate your thoughts.

- Strict linear ordering of scenes
 - Heavy messaging
 - No interactivity
- No prescribed ordering
 - No messaging
 - Free interactivity



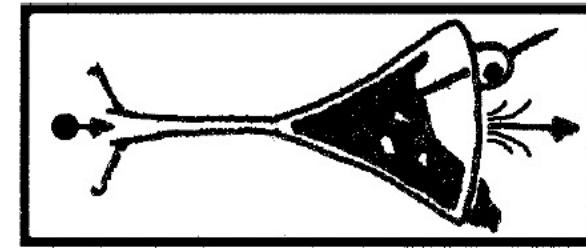
Author Driven

Reader Driven

Storytelling – Narrative Visualization

Martini Glass Structure

- Starts with an author-driven approach, then opens to a reader-driven stage



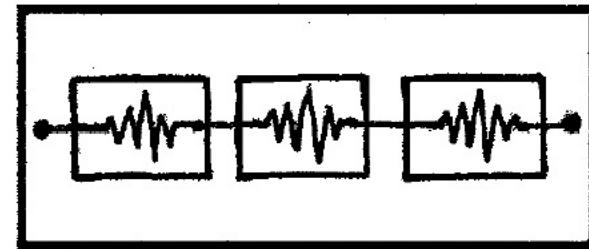
Author Driven

Reader Driven

Storytelling – Narrative Visualization

Interactive Slideshow

- Slideshow that incorporates interaction within each slide
- More even mix of author-driven and reader-driven



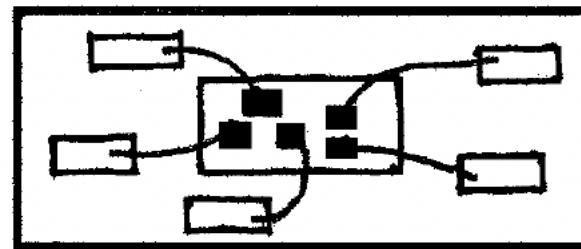
Author Driven

Reader Driven

Storytelling – Narrative Visualization

Drill-Down Story

- Presents a general theme, then allows the user to explore specific instances of that theme
- More emphasis on reader-driven

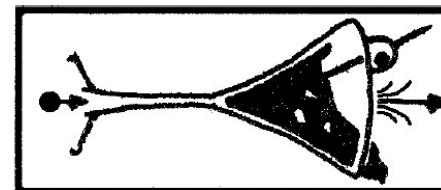


Author Driven

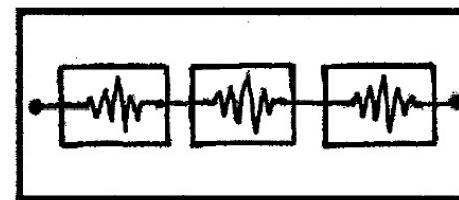
Reader Driven

Storytelling – Narrative Visualization

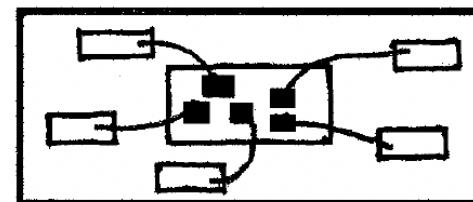
Martini Glass Structure



Interactive Slideshow



Drill-Down Story



In groups, find an example of each story structure. Add your examples to the [Narrative-Vis Jamboard](#).

(Try looking on New York Times, Washington Post, FiveThirtyEight, Boston Globe, etc.)

Author Driven

Reader Driven



A more artistic approach

Giorgia Lupi – Dear Data

GIORGIA

week seven

a week of complaints

Stefanie

“DEAR DATA
WEEK 07: MUSICAL COMPLAINTS

HOW TO READ IT: Each “note” is a single complaint I said. (i.e. every single time I expressed dissatisfaction or annoyance about a situation or particular thing.)
Each “score” represents a typology of things I complained about, starting complaints in chronological order.

SCORES:

- ME AS A PERSON (e.g. "I am so... angry / obsessive...")
- THE PLACE (e.g. "I should've done...")
- WORK (e.g. "this project isn't going well!")
- TECHNOLOGICAL (e.g. "the summer is not working!")
- SERVICE/IOD (e.g. "and the water is so slow!")
- SOMEBODY (e.g. "he's really a jerk...")
- COLD (e.g. "I am freezing! The A.C. is crazy!")
- HOW I FEEL (e.g. "so tired!!", "so bored!")
- BOYFRIEND (e.g. "you're snoring! you haven't...")
- OTHER (e.g. "I spent 1 hour waiting for...")

POSITIONS OF NOTES:

- 1 → ACTUAL need to complain
- 2 → OUTAGE
- 3 → NORMAL
- 4 → MISSED COMPLAINTS:
Thought of complaining
But didn't do!

ATTRIBUTES

- 0 to boyfriend
- 0 to friendly family
- 0 to strangers
- 99 → in english
(earlier others were in ITA)

DELIVERED BY HAND (SPECIAL NYC DELIVERY!)

FROM:
GIORGIA LUPI
NYC BROOKLYN
NY - USA

SEND TO:
STEFANIE POSAVEC
LONDON

TO:
GIORGIA LUPI
BROOKLYN, NY
USA

MAIN STATS

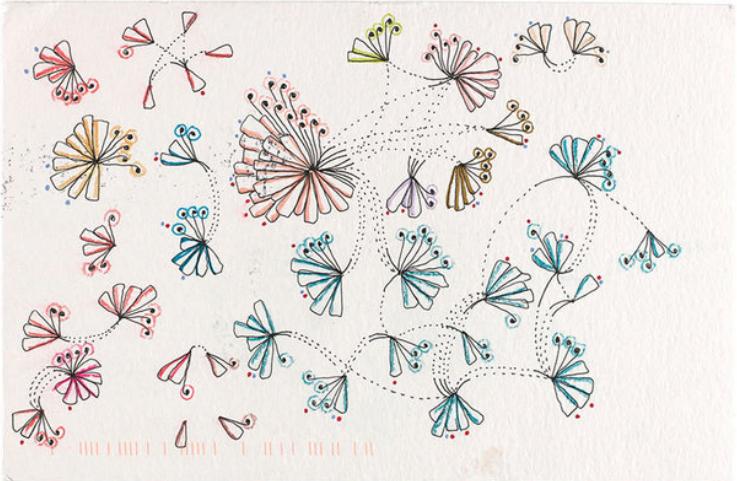
PRIVATE COMPLAINTS: 67	OUTWARD COMPLAINTS: 100
COMPLAINED TO ME: 43	HUNGER: 100
# OF PEOPLE WHO COMPLAINED TO ME: 5	TECHNOLOGY/MEDIA: 100
COMPLAINED TO ME: 10	INNATE: 100
# OF PEOPLE I COMPLAINED TO: 10	STRANGERS: 100
COMPLAINED TO ME: 10	TRANSPORT: 100
FAMILY: 100	WORK: 100
ACQUAINTANCES: 100	MY APPEARANCE: 100
STRANGERS: 100	MY FRIENDS: 100
INNATE: 100	MY WORK: 100
TRANSPORT: 100	MY BODY: 100
WORK: 100	MY HUSBAND: 100
MY APPEARANCE: 100	LEAKED/SHOVED: 100
MY FRIENDS: 100	MY DRAWING + GOT ALL OVER MY HANDS: 100
MY WORK: 100	AND A WEEK OF COMPLAINTS ABOUT I FORKED UP THIS DRAWING! (COMPLAINT #5)

Note the hand-drawn stamps: these postcards were delivered in person in New York!

Giorgia Lupi – Dear Data

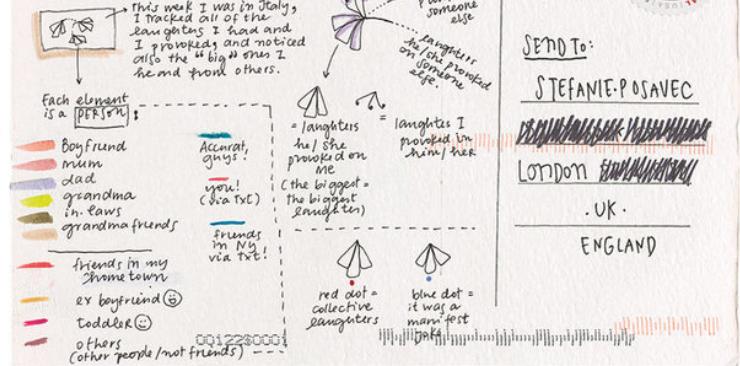
GIORGIA

week forty-two



66 Dear Data
WEEK 42: Laughters!

HOW TO READ IT:



When data-gathering gets in the way: Giorgia laughs and then has to report it. So, is she fully enjoying her cheerful moments as they pop up?

a week of laughters

Stefanie



DEAR DATA - WEEK 42

FOR A CARD ABOUT LAUGHTER I AM SAD ABOUT HOW THIS CARD TURNED OUT. I TRIED TO CAPTURE MY LAUGHS WHICH WAS REALLY HARD + GOT LOST IN THE WAY OF ENJOYING LIFE, HENCE THE DATA VOIDS :)

HOW TO READ IT:



ALL LAUGHS ARE DRAWN IN CHRONOLOGICAL ORDER FROM START TO END. MARKERS INDICATE END OF ONE DAY + BEGINNING OF OTHER. IF MARKER IS IN PEN, IT MEANS I HAD A "DATA VOID" DUE TO THE FOLLOWING:

DRINKING W FRIENDS, MY BDAY!! , MY 8-DAY DINNER

E WAS LAUGHING WITH I WAS LAUGHING ABOUT

000

WITH MYSELF, MY HUSBAND, MY FRIEND, STUDIO MATES, GROUP OF ANIMALS, PARENTS, SIBLINGS, YOU!!

LARGEST HEARTET! A DOT MEANS YOU LAUGHED WITH THEM LAUGHING LAUGHING WITH PEOPLE (GENERALLY IN FUN)

BEING IN A GOOD MOOD, GENERAL LAUGHS (WE GET SECONDS)

FRIENDS/FAMILY MEMBERS REHABILITATING

MESSES UP MY DATA :)

LAUGHING LAUGHING WITH PEOPLE (GENERALLY IN FUN)

BEING TICKLED (KICKIN TO MESSES UP MY DATA :)

ANIMALS, PARENTS, SIBLINGS, YOU!!

FROM: S POSAVEC

TO: GIORGIA LUPI

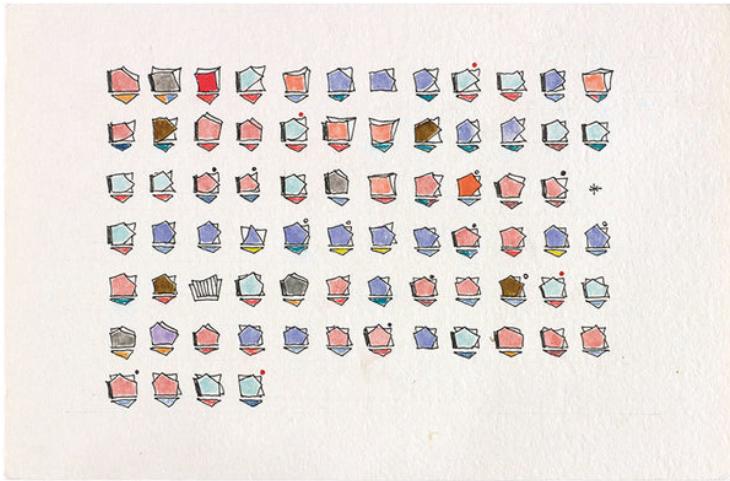


Stefanie was banned by her husband from gathering data at her birthday dinner in a posh restaurant (for obvious reasons). Hence the "data void".

Giorgia Lupi – Dear Data

GIORGIA

week fifty-two



66 Dear Data
WEEK 52 - ... Goodbye! LAST WEEK OF Dear Data

HOW TO READ IT:

This week I tracked all the "Goodbyes / sayByes / goodnight" I said. Each filament is a goodbye I said, in chronological order.

SHAPE = "HOW"

- = in real life
- = over the phone
- = Skype / hangout
- = in public ("public speech")
- = farewell to my old apartment!
- * = missed goodnight to my boyfriend cause I fell asleep too early!

COLOR = TO

color of the triangle = did I add something?

- = kiss
- = hand shake
- = other physical contact
- = if dat is missing
- = a good luck!
- = have fun, enjoy / devuerti!
- = talk soon / see you later / or drop!
- = thanks! / thanks from... / grace!
- = acquaintance
- = client
- = grandma
- = stranger (sales man / waitress...)

from:
GIORGIA LUPI
ONLINE
BROOKLYN
NY - USA

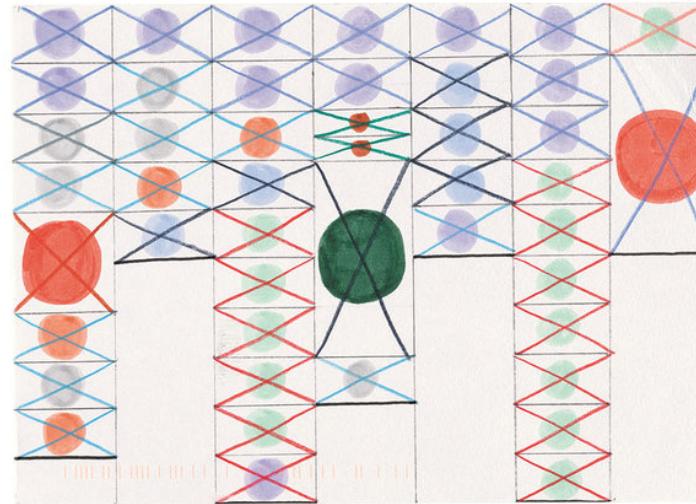


SEND TO:
STEFANIE POSAVEC
LONDON
UK
ENGLAND

The most shameful revelation. Of course, the last week of Dear Data was about "goodbyes", to include a special goodbye to Dear Data in their postcards. Guess what Giorgia forgot to add? :)

a week of goodbyes

Stefanie



DEAR DATA - WEEK 52

A WEEK OF GOODBYES "SAYING GOODBYE, BOTH FOR A WHILE AND FOREVER."

HOW TO READ IT: EACH IS ONE GOODBYE.

GOODBYES ARE ORGANISED IN CHRONOLOGICAL ORDER FROM L-R AND TOP-BOTTOM.

AM PM
M T W T F S S

HOW / WHERE I WAS SAYING THE GOODBYE:
AT HOME
AT WORK (INCLUDE MY ANTIQUE PUB CRAWL)
ON SOCIAL MEDIA

AT STUDIO (NEW+GO)
WHO / WHAT I WAS SAYING GOODBYE TO:
COLLEAGUE
STUDIOMATE
FRIEND
YOU!

A FINAL GOODBYE!
AND A FINAL HANGOVER HAHA

FROM:
S. POSAVEC
LONDON
UK

Royal Mail
Mount Pleasant
Mail Centre
15-09-2015
44009732



TO:
GIORGIA LUPI
BROOKLYN, NY
USA

BY AIR MAIL
par avion

Royal Mail®

"My amazing pub crawl": Stefanie hit ten pubs in an afternoon and was feeling pretty proud. "My old studio": the pub crawl was in honour of the disbanding of the studio she shared with friends.

Giorgia Lupi

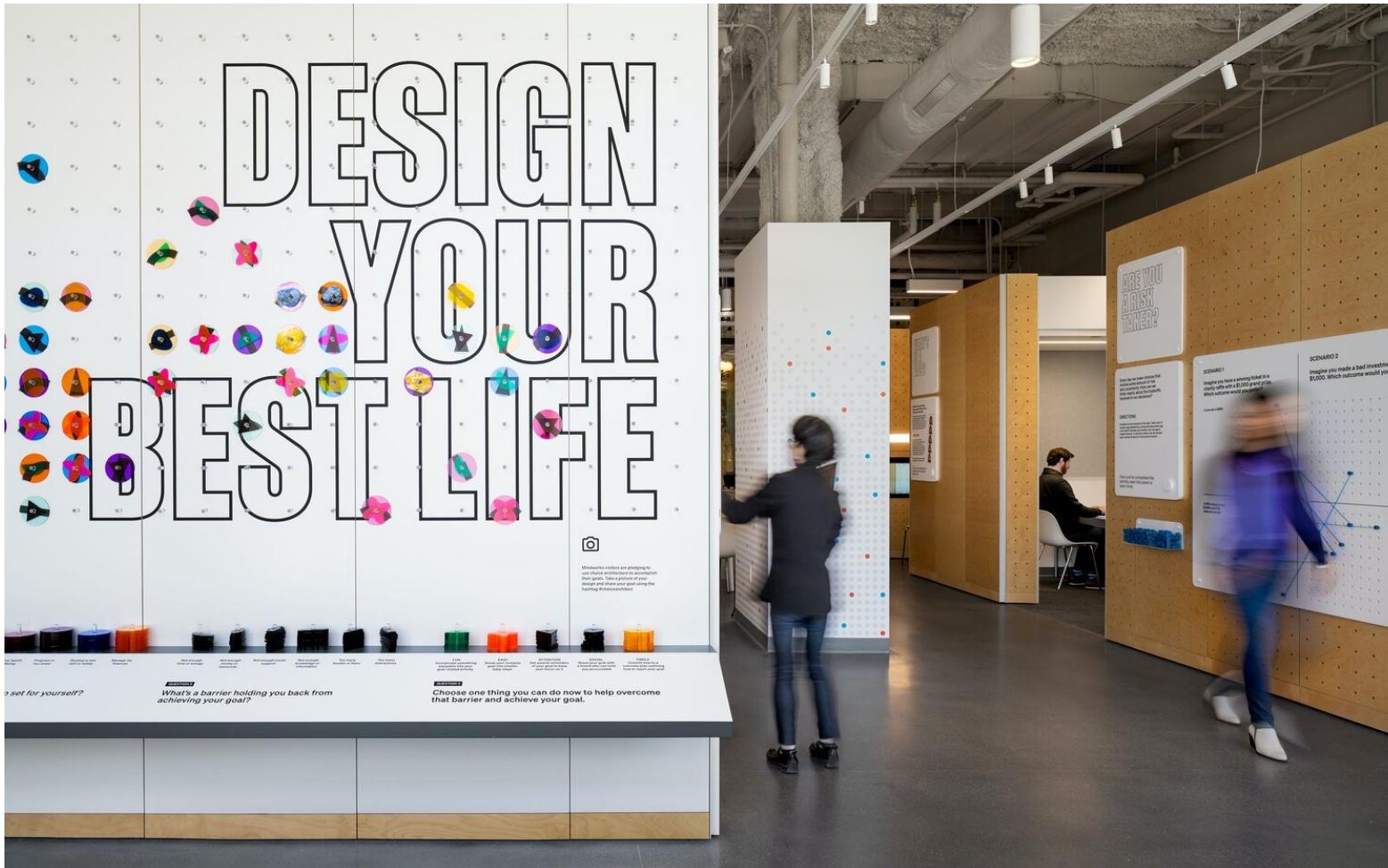


What sets Giorgia apart is her [humanistic approach](#) to the world of data. Data is considered to be impersonal, boring, and clinical, but her work proves the opposite.

She uses data as a lens to better understand our human nature and every aspect of our society. By [distilling our personal experiences](#) (our activities, thoughts, behaviors, relationships) into what we so coldly call data, and by actively building her datasets and expressing them as a designer and artist, she seeks to grasp glimpses of humanity and discover overlooked details.

When Giorgia is presented with data, [she seeks to humanize it](#), to make it speak our language and represent our human nature, because, in her opinion, this is the ultimate goal of any design work, especially with data. She often combines the original data with layers of softer and more qualitative information that renders and presents its more nuanced and more human aspects of us.

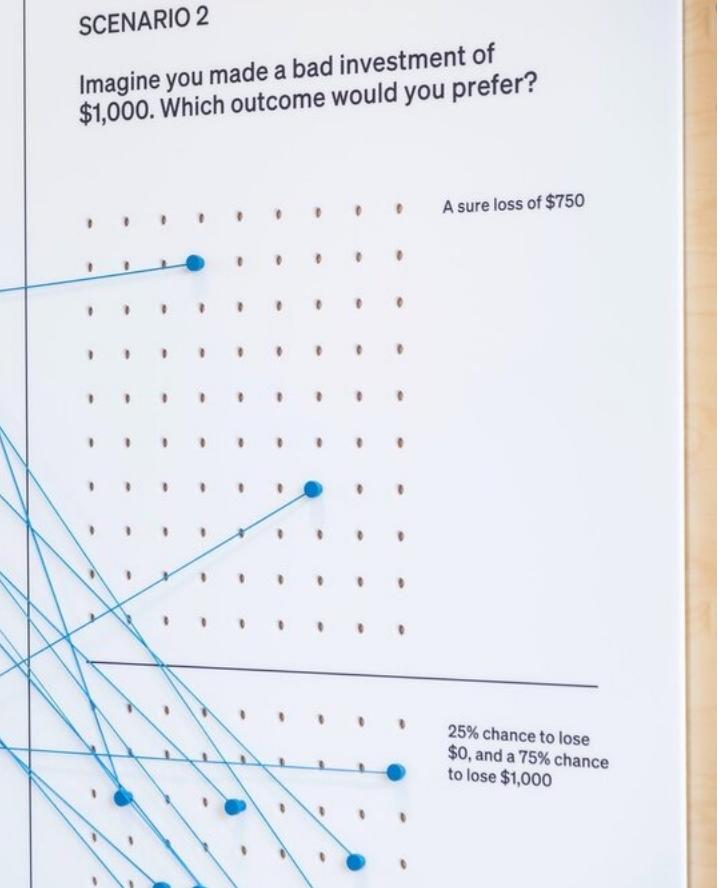
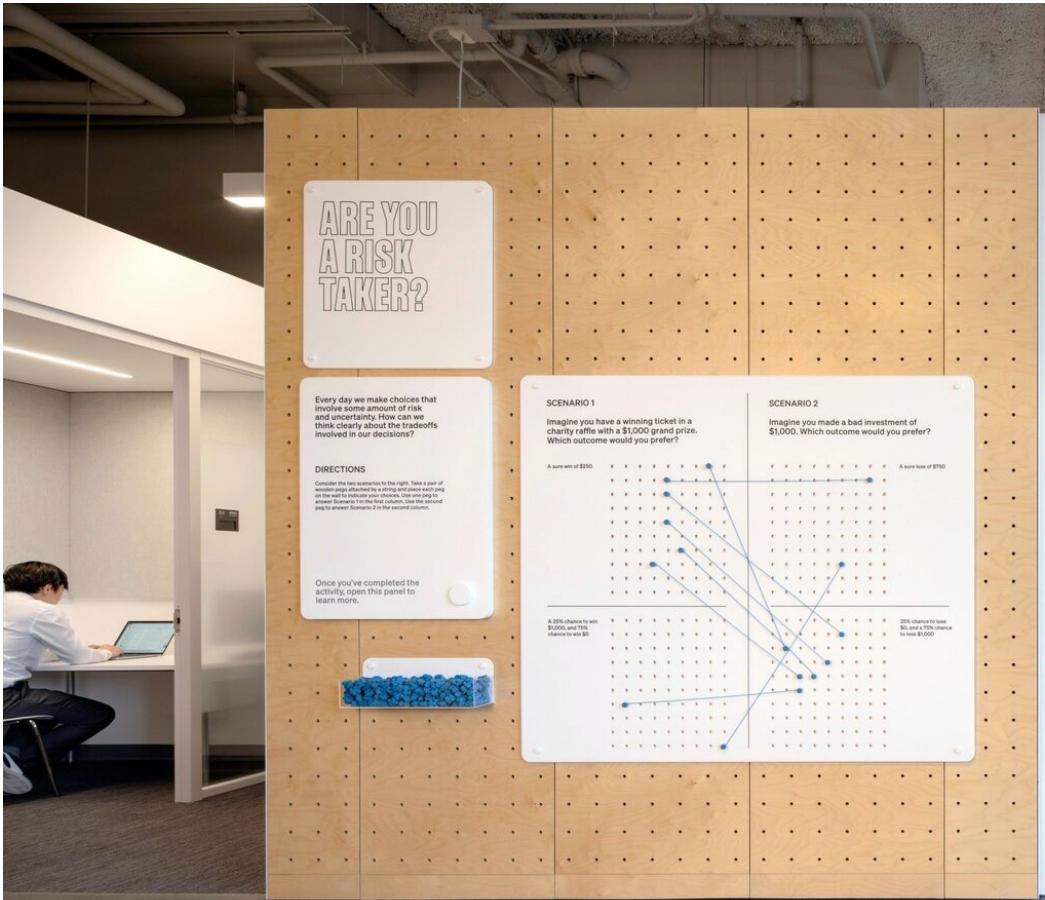
Giorgia Lupi – Physical Visualizations



Giorgia Lupi – Physical Visualizations



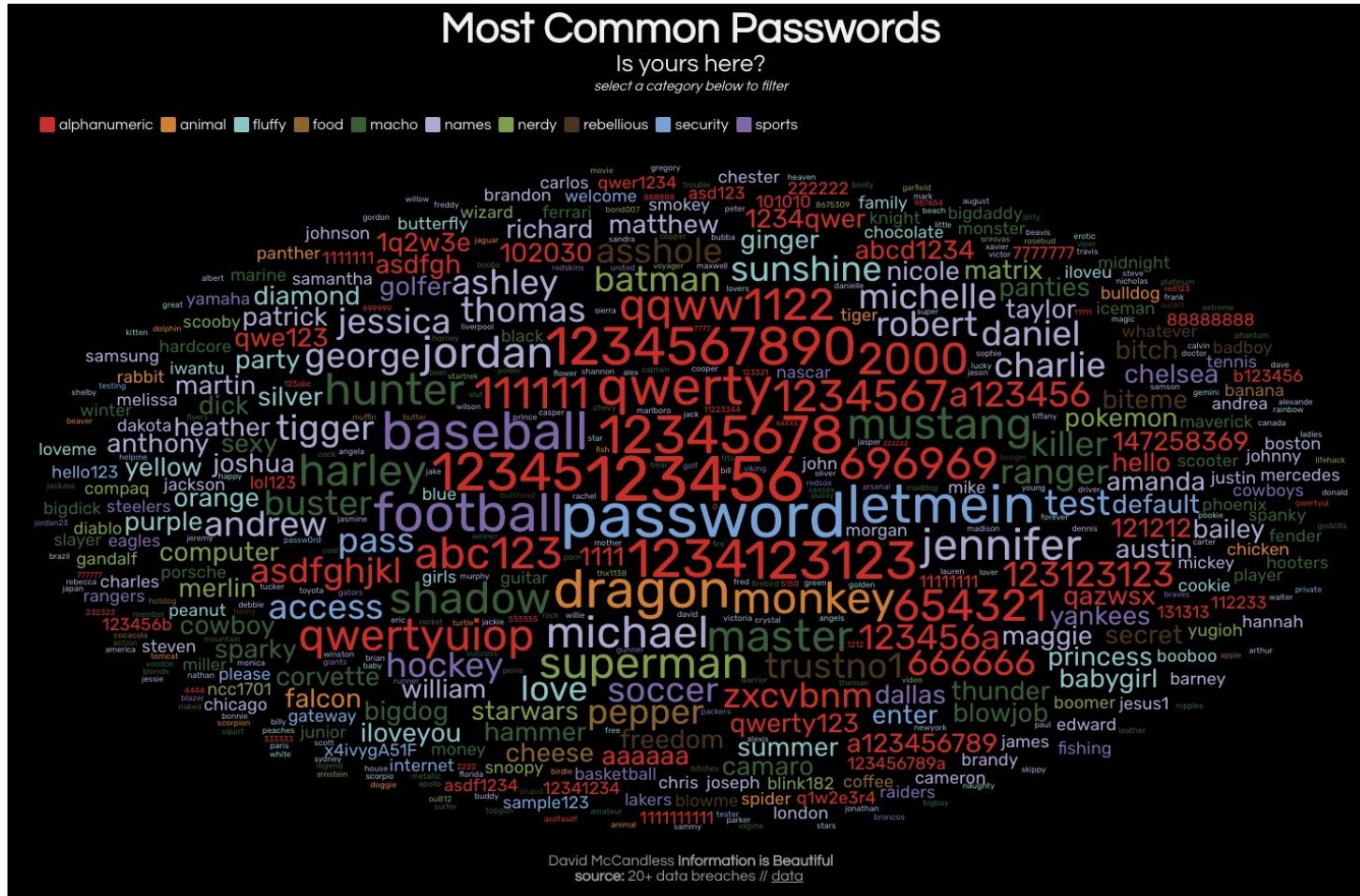
Giorgia Lupi – Physical Visualizations



information is beautiful

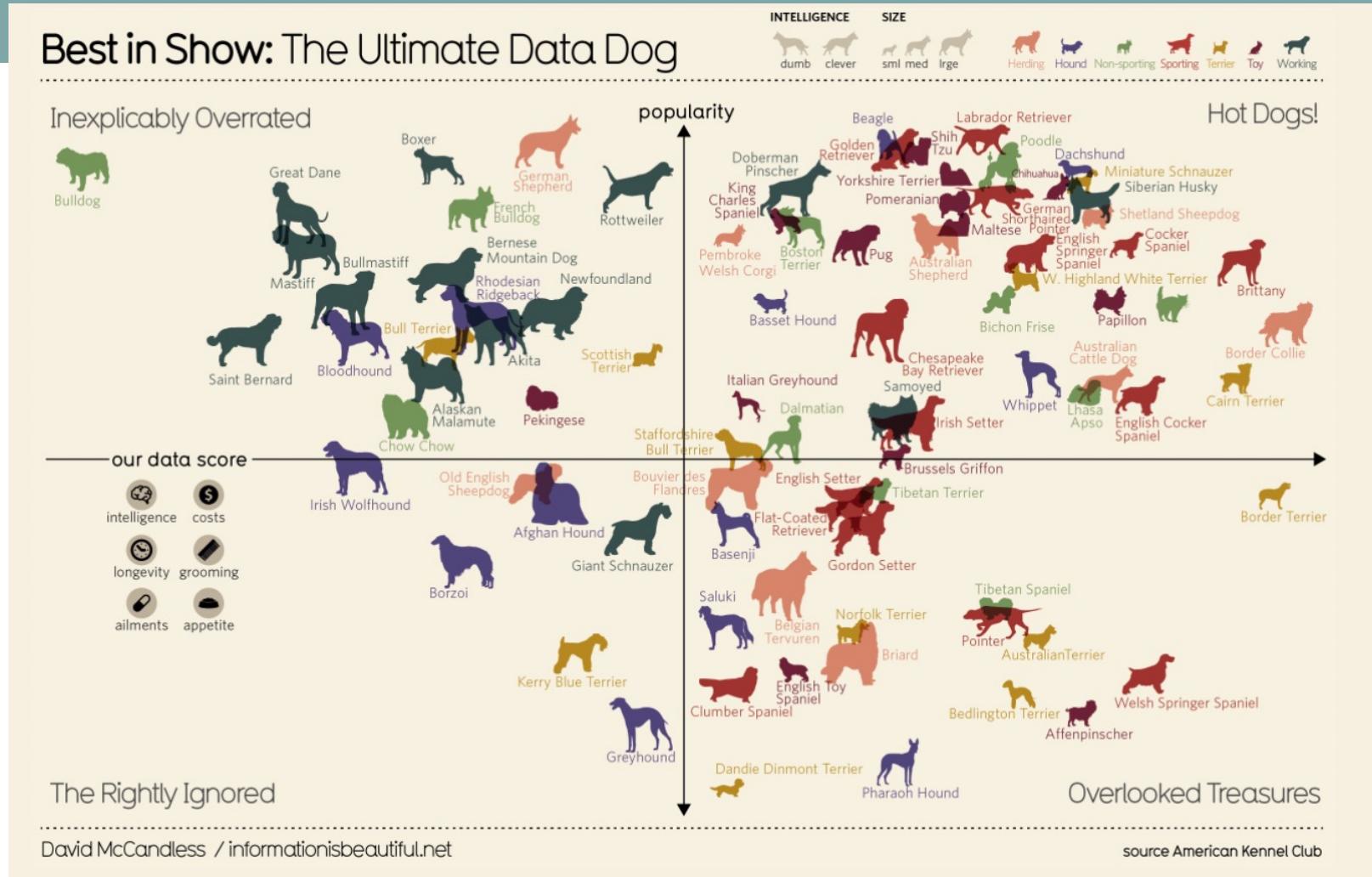


information is beautiful

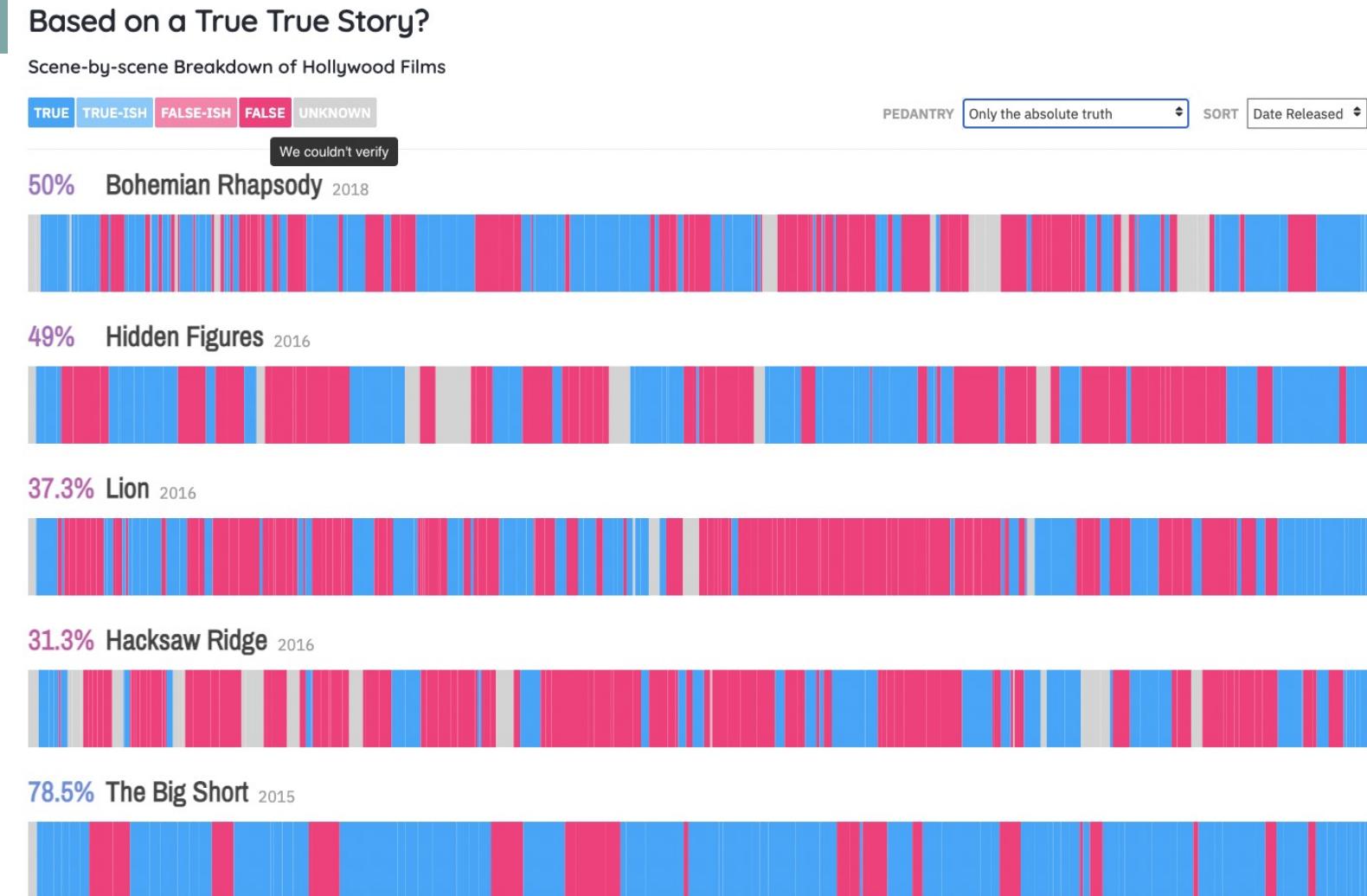


<https://informationisbeautiful.net/visualizations/top-500-passwords-visualized/>

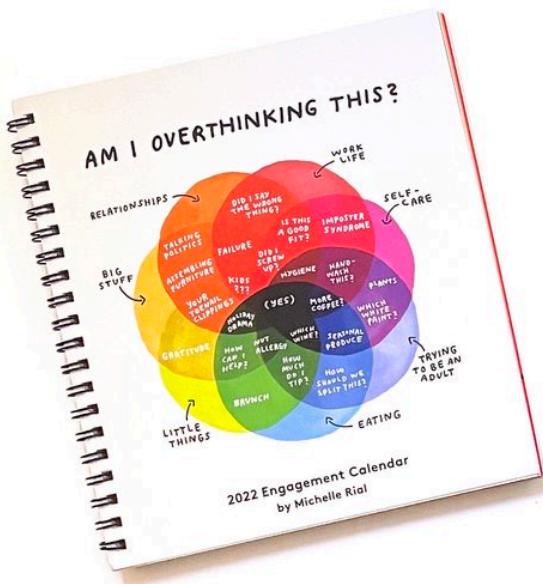
information is beautiful



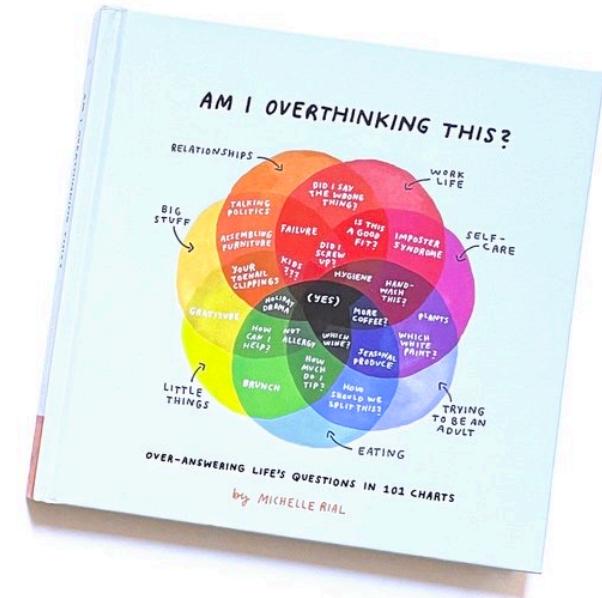
information is beautiful



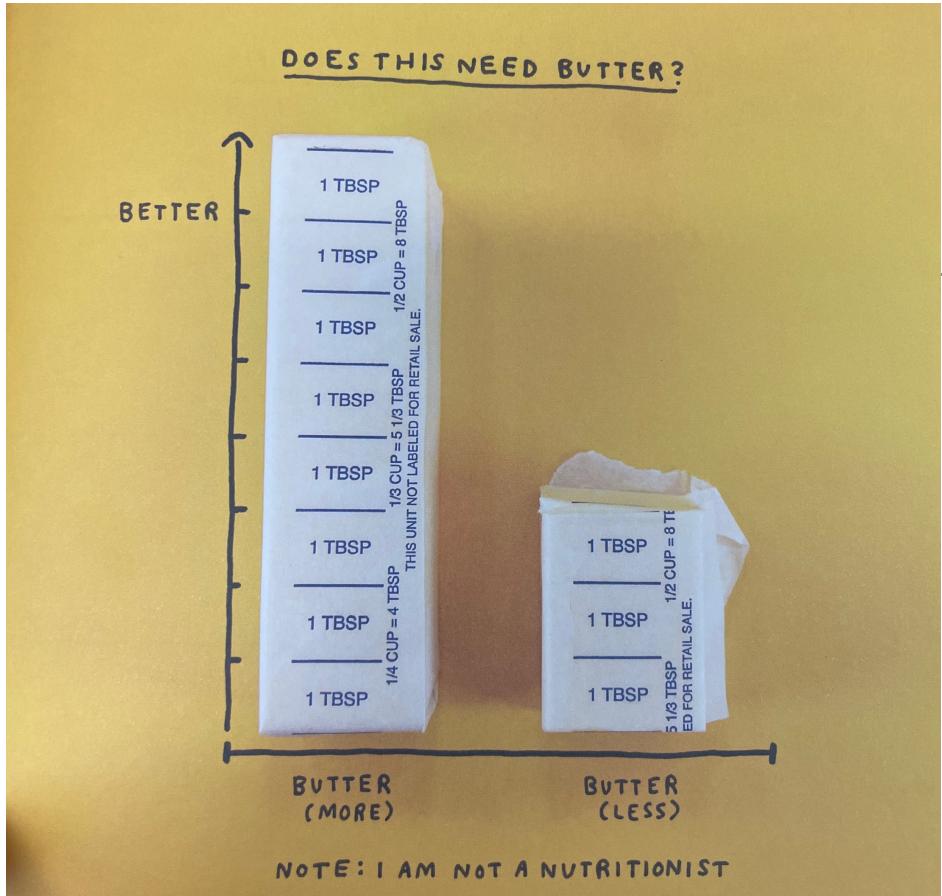
Michelle Rial



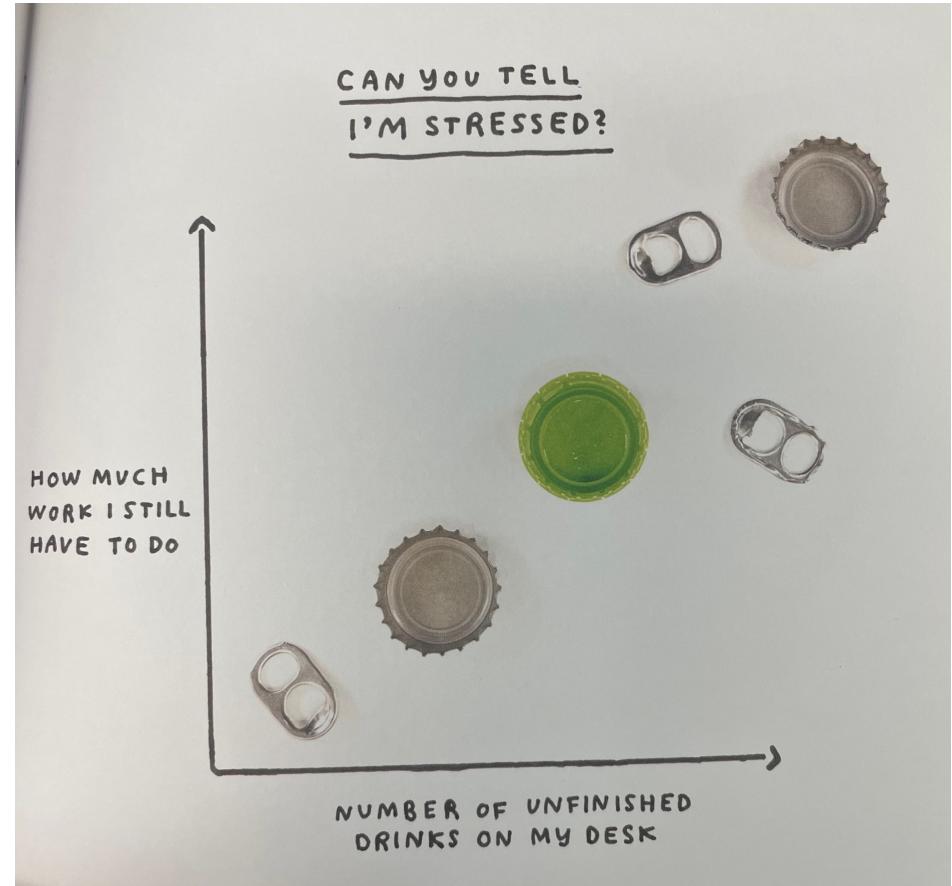
G FAY ABOUT ME? WHAT SHOULD WE LISTEN TO? HOW I HAVE ANOTHER CUP OF COFFEE? HOW AM I FULL ALREADY? HOW SHOULD I WHAT DOES ALLY IRRESPONSIBLE? SHOULD I EAT THE REST OR SPLIT IT FEED A MILLENNIAL? HOW DO I START THIS EMAIL? WHY DO I FEEL TERRIBLE? STILL OR SPARKLING EXCLAMATION POINT? ARE PEOPLE JUDGING ME BY DO SMALL TALE? IS THERE SOMETHING MORE INT POP THIS ZIP? **AM I GETTING FIRED?** IS IT NORM AY? WHAT'S SO GREAT ABOUT CAMPING? DOES THIS COULD I LEND OUT MY BOOKS? SHOULD I GO OUT OR NOT KEEP THIS ALIVE? DO I HAVE TOO MANY PLANTS? SHOULD I USE? WHY DON'T MY DIYZ LOOK LIKE HGTV? AM I IMPLICATED? **OVERTHINKING** THE EVER IT? SHOULD I GET RID OF THIS? WHICH BIN DOES TH AM I USING TOO MUCH PLASTIC? AM I GOING TO RE WILL OUR RELATIONSHIP SURVIVE THIS STEP? AN WHEN CAN I CLIP MY TOENAILS IN FRONT OF YOU AM I A BAD FRIEND? MAYBE WE SHOULD TALK A TE? HOW MUCH DO I TIP FOR **THIS?** WILL IE IS MY HAIR GROWFUL? ARE YOU SURE? HOW DO I F OR? WHAT A E REALLY IS SOMETHING WRONG W DID I SO JOURNAL WILL PARENTHOOD CHANGE EVERYT AM I DOING UGH? WHERE ARE MY HAIRIES? I FAIL? DID I SCREW UP? by MICHELLE RIAL NHE ANY THOUGHTS? SHOULD I GIVE MYSELF A BREAK? SHILL IT BE OK? ARE YOU SURE? WHAT IF THERE TO WHICH WINE SHOULD I BRING? DO I HAVE TO HAND OUT? SHOULD I GIVE THEM A THICK? AND



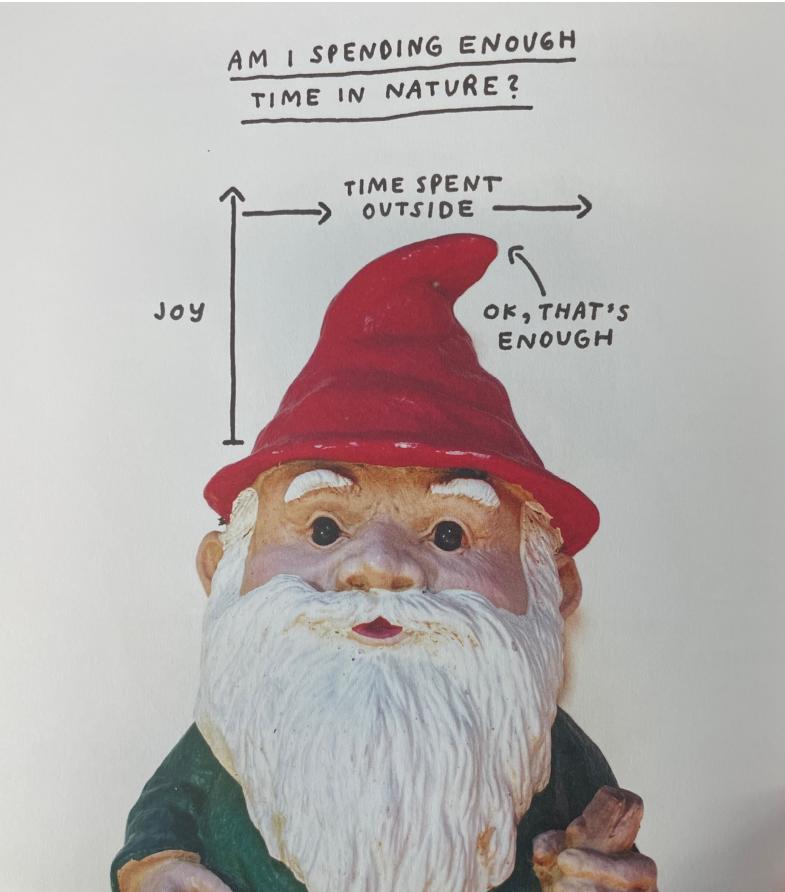
Michelle Rial – Am I Overthinking This?



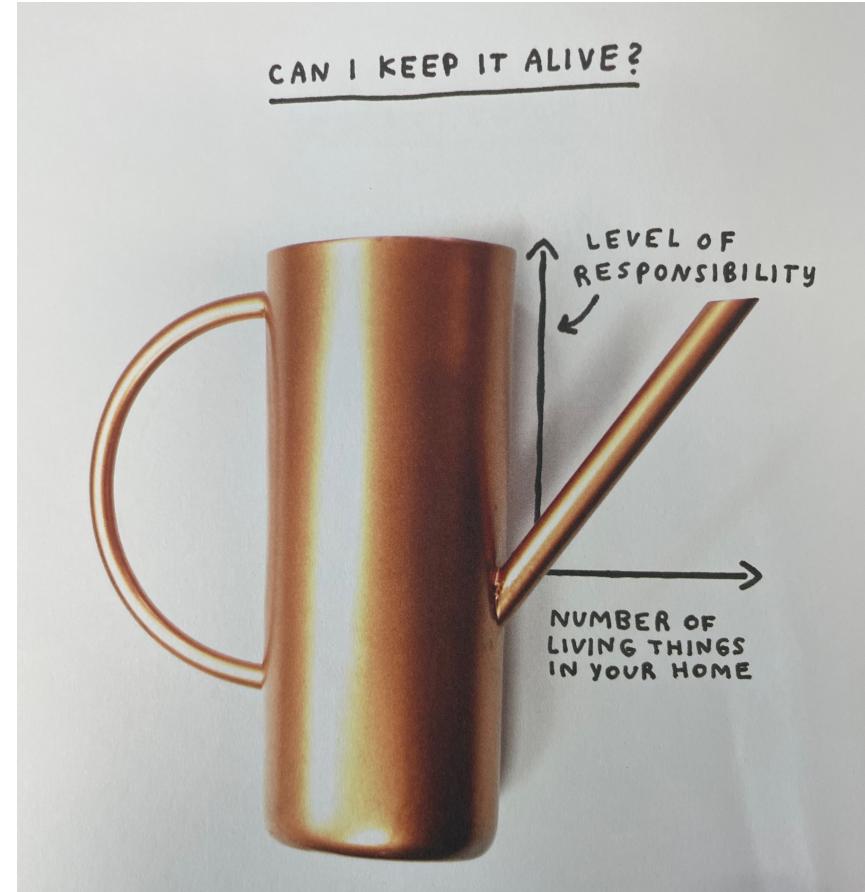
73



Michelle Rial – Am I Overthinking This?



74



Take 15 mins for TRACE evals

Intro to Visualization Research: Example of
my work in visualization for communication

Medical Risk Communication



Bayesian Reasoning

The probability of breast cancer is 1% for women at age forty who participate in routine screening. If a woman has breast cancer, the probability is 80% that she will get a positive mammography. If a woman does not have breast cancer, the probability is 9.6% that she will also get a positive mammography.

A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?



Bayesian Reasoning

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$



Bayesian Reasoning

The probability of breast cancer is 1% for women at age forty who participate in routine screening. If a woman has breast cancer, the probability is 80% that she will get a positive mammography. If a woman does not have breast cancer, the probability is 9.6% that she will also get a positive mammography.



A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?



Bayesian Reasoning

The probability of breast cancer is 1% for women at age forty who participate in routine screening. If a woman has breast cancer, the probability is 80% that she will get a positive mammography. If a woman does not have breast cancer, the probability is 9.6% that she will also get a positive mammography.

A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

Bayesian Reasoning

Probability -> Frequency

The probability of breast cancer is 1% for women at age forty who participate in routine screening. If a woman has breast cancer, the probability is 80% that she will get a positive mammography. If a woman does not have breast cancer, the probability is 9.6% that she will also get a positive mammography.

A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

Bayesian Reasoning

Probability -> Frequency

The probability of breast cancer is 1% for women at age forty who participate in routine screening. If a woman has breast cancer, the probability is 80% that she will get a positive mammography. If a woman does not have breast cancer, the probability is 9.6% that she will also get a positive mammography.

A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

Bayesian Reasoning

Probability -> Frequency

10 out of every 1,000 women at age forty who participate in routine screening have breast cancer. 8 of every 10 women with breast cancer will get a positive mammography. 95 out of every 990 women without breast cancer will also get a positive mammography.

A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

Bayesian Reasoning

10 out of every 1,000 women at age forty who participate in routine screening have breast cancer. 8 of every 10 women with breast cancer will get a positive mammography. 95 out of every 990 women without breast cancer will also get a positive mammography.

A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

Bayesian Reasoning

Use a probe question

10 out of every 1,000 women at age forty who participate in routine screening have breast cancer. 8 of every 10 women with breast cancer will get a positive mammography. 95 out of every 990 women without breast cancer will also get a positive mammography.

A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

Bayesian Reasoning

Use a probe question

10 out of every 1,000 women at age forty who participate in routine screening have breast cancer. 8 of every 10 women with breast cancer will get a positive mammography. 95 out of every 990 women without breast cancer will also get a positive mammography.

A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

Bayesian Reasoning

Use a probe question

10 out of every 1,000 women at age forty who participate in routine screening have breast cancer. 8 of every 10 women with breast cancer will get a positive mammography. 95 out of every 990 women without breast cancer will also get a positive mammography.

Imagine 1000 people are tested for the disease.

(a) How many people will test positive?

(b) Of those who test positive, how many will actually have the disease?

Bayesian Reasoning

10 out of every 1,000 women at age forty who participate in routine screening have breast cancer. 8 of every 10 women with breast cancer will get a positive mammography. 95 out of every 990 women without breast cancer will also get a positive mammography.

Imagine 1000 people are tested for the disease.

(a) How many people will test positive?

(b) Of those who test positive, how many will actually have the disease?

Bayesian Reasoning

Use visualization

10 out of every 1,000 women at age forty who participate in routine screening have breast cancer. 8 of every 10 women with breast cancer will get a positive mammography. 95 out of every 990 women without breast cancer will also get a positive mammography.

Imagine 1000 people are tested for the disease.

(a) How many people will test positive?

(b) Of those who test positive, how many will actually have the disease?

Bayesian Reasoning

Use visualization

10 out of every 1,000 women at age forty who participate in routine screening have breast cancer. 8 of every 10 women with breast cancer will get a positive mammography. 95 out of every 990 women without breast cancer will also get a positive mammography.

Imagine 1000 people are tested for the disease.

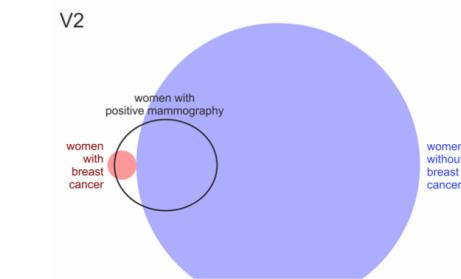
(a) How many people will test positive?

(b) Of those who test positive, how many will actually have the disease?

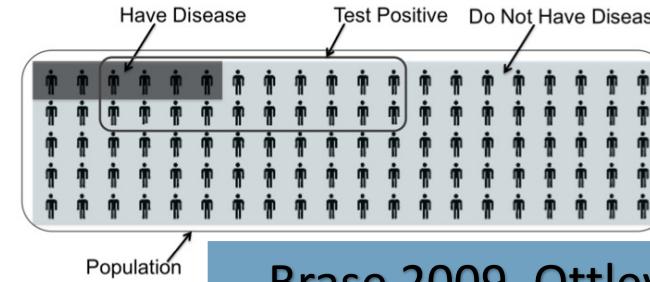
Bayesian Reasoning

Use visualization

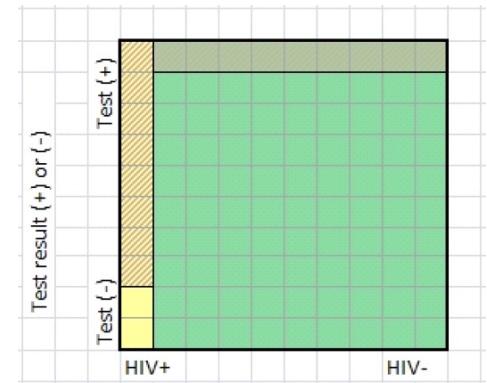
- Euler diagrams
- Frequency grids or icon arrays
- Decision trees
- “Beam cut” diagrams
- Probability curves
- Contingency tables
- Doubletrees
- Flow charts
- Pipe diagrams
- Sankey diagrams
- Unit squares



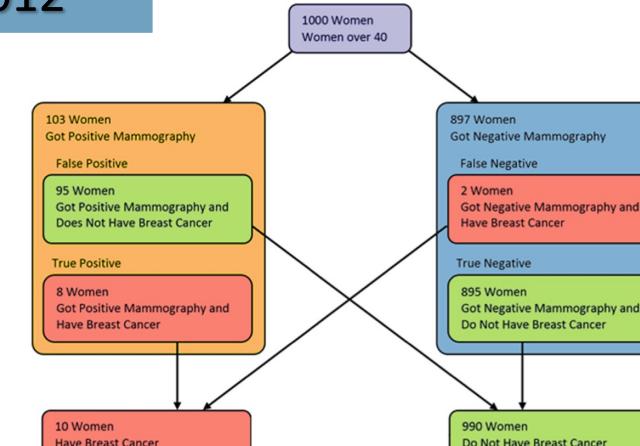
Micallef et al. 2012



Bräse 2009, Ottley et al. 2016



Tsai et al. 2011



Khan et al. 2015



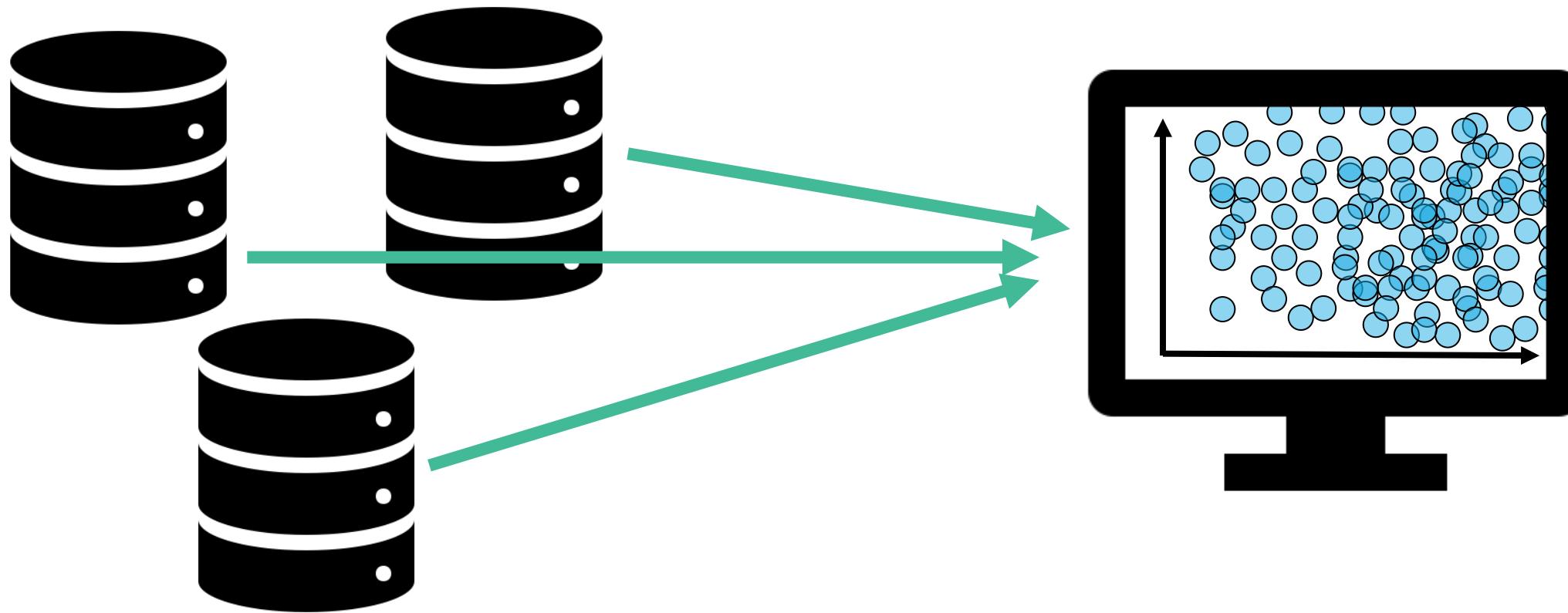
Visual Data Analysis Group



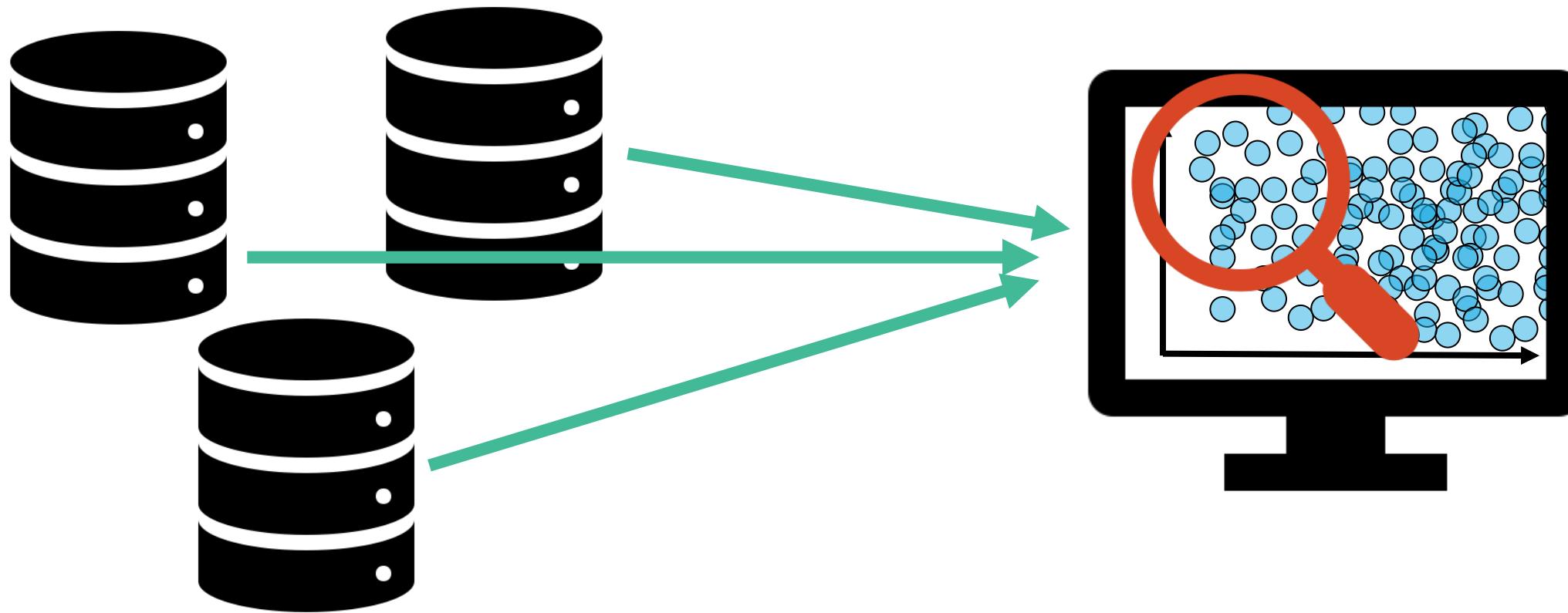
Can interaction help?



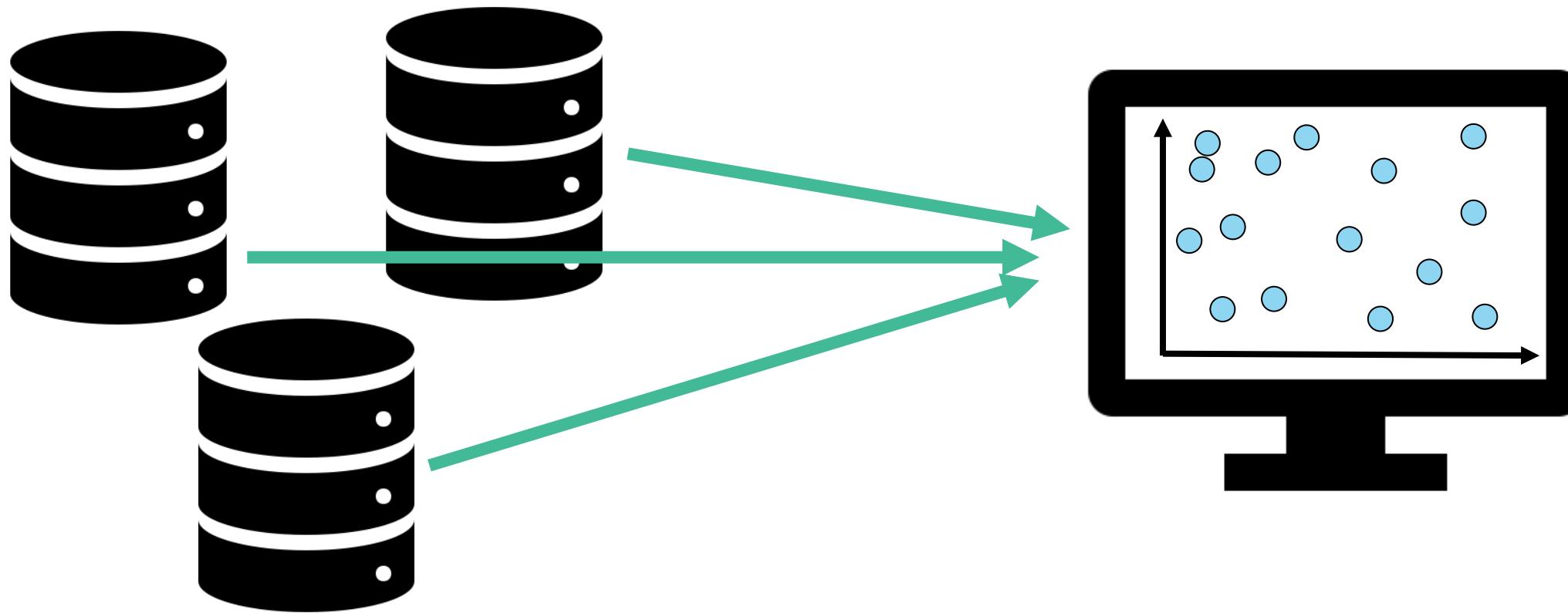
Large Data Exploration



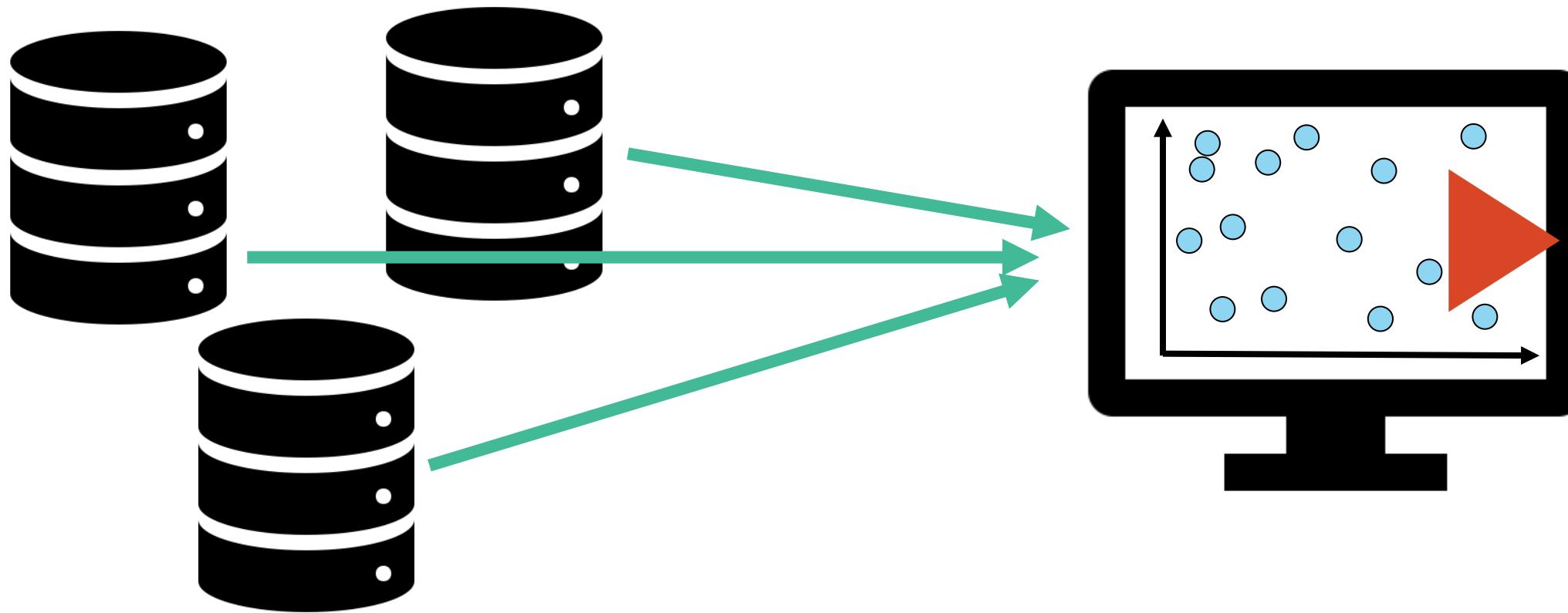
Large Data Exploration



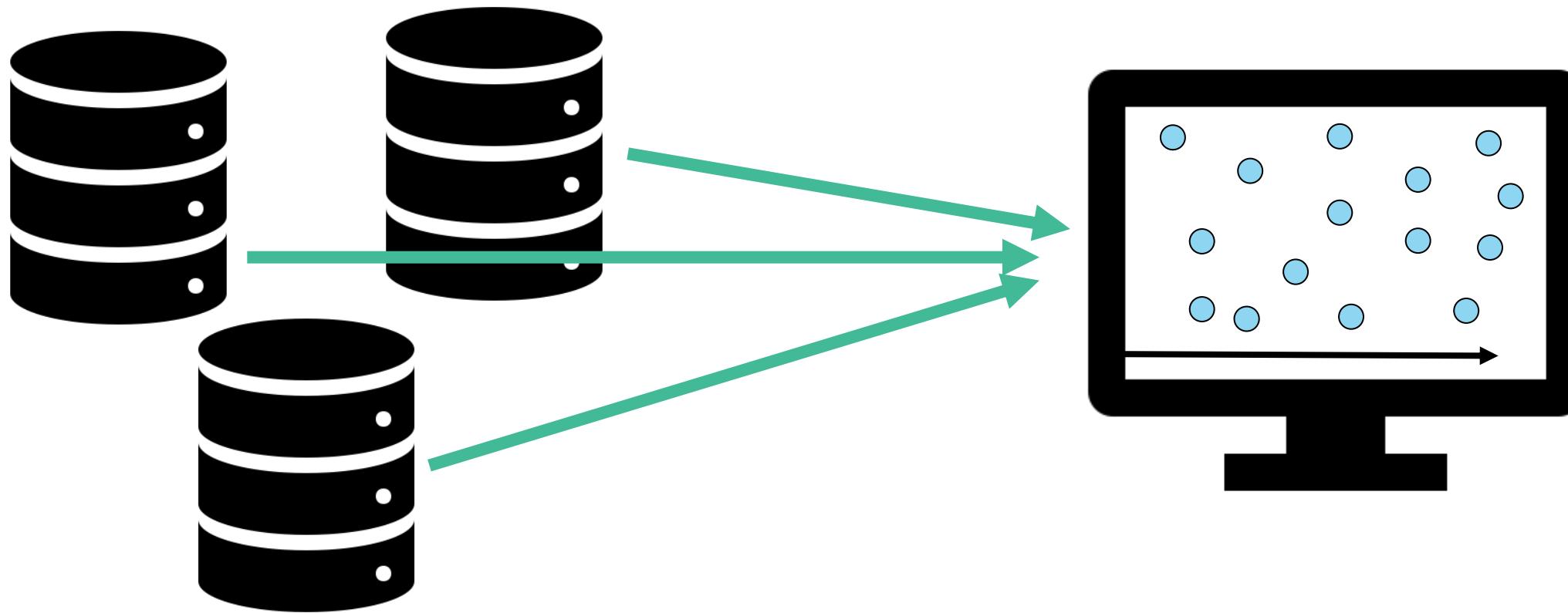
Large Data Exploration



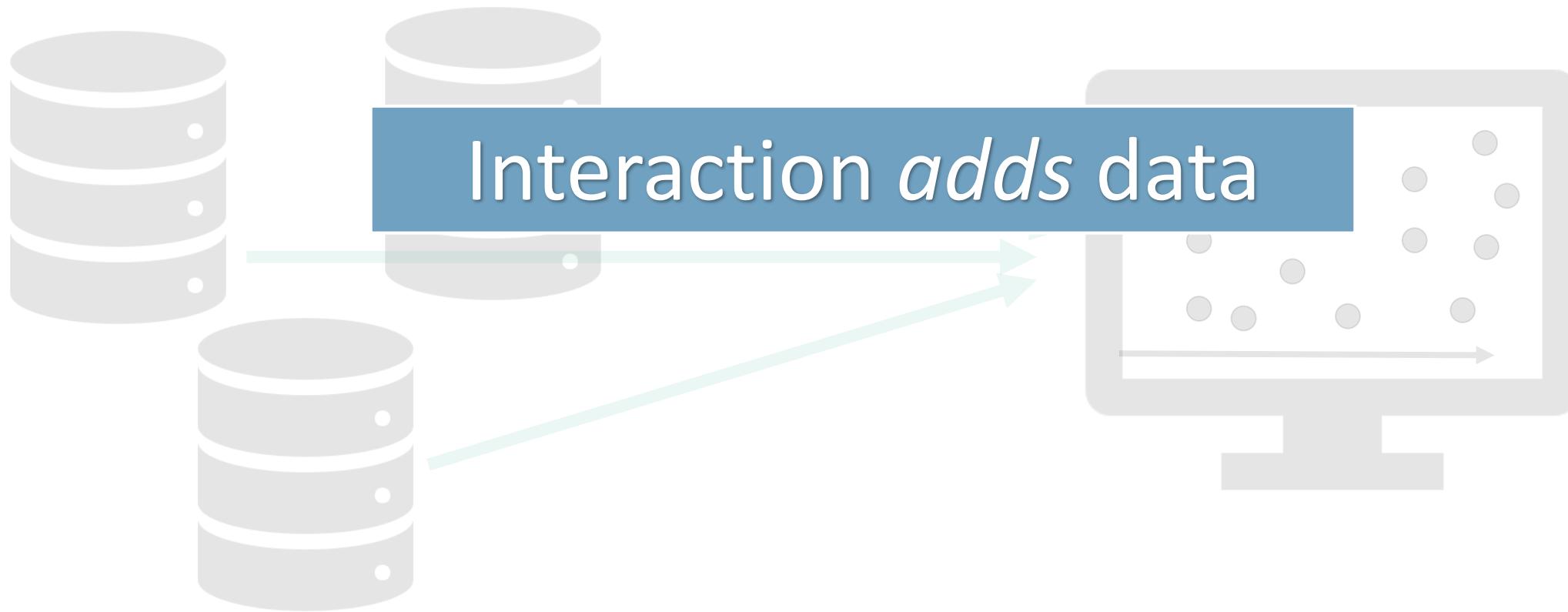
Large Data Exploration



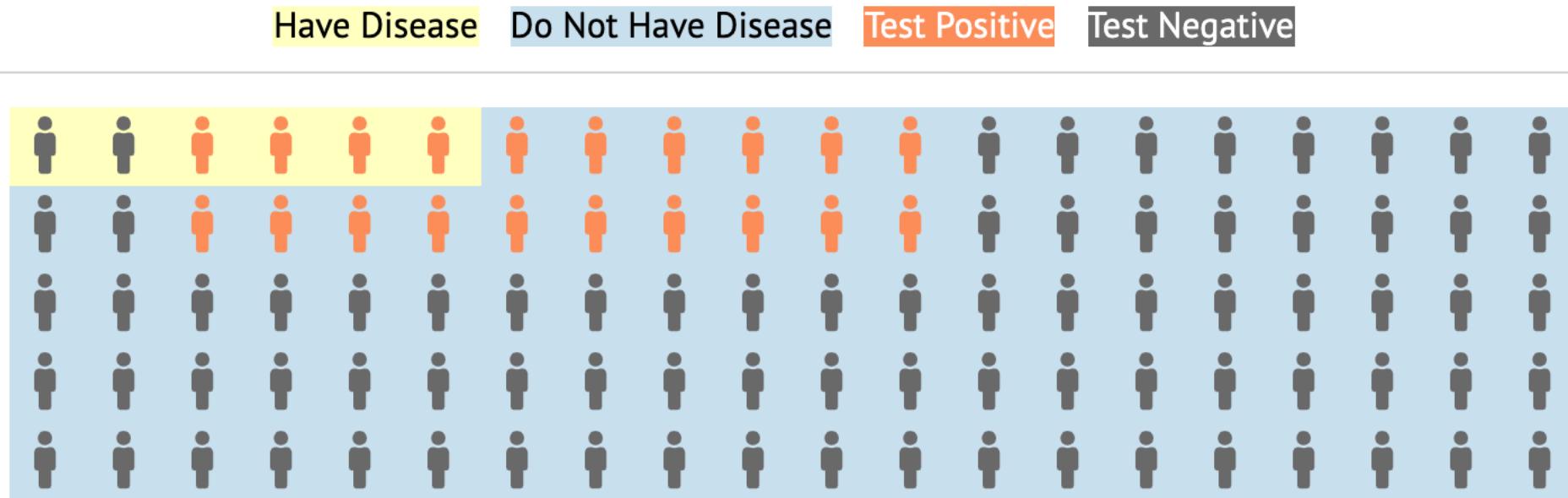
Large Data Exploration



Large Data Exploration



Static Visualization



Static Visualization

True Positive

True Negative

False Positive

False Negative

Have Disease	Do Not Have Disease	Test Positive	Test Negative
● ● ○ ○ ○ ○	○ ○ ○ ○ ○ ○	● ● ○ ○ ○ ○	● ● ○ ○ ○ ○
● ● ○ ○ ○ ○	○ ○ ○ ○ ○ ○	● ○ ○ ○ ○ ○	● ○ ○ ○ ○ ○
● ● ○ ○ ○ ○	○ ○ ○ ○ ○ ○	● ○ ○ ○ ○ ○	● ○ ○ ○ ○ ○
● ● ○ ○ ○ ○	○ ○ ○ ○ ○ ○	● ○ ○ ○ ○ ○	● ○ ○ ○ ○ ○



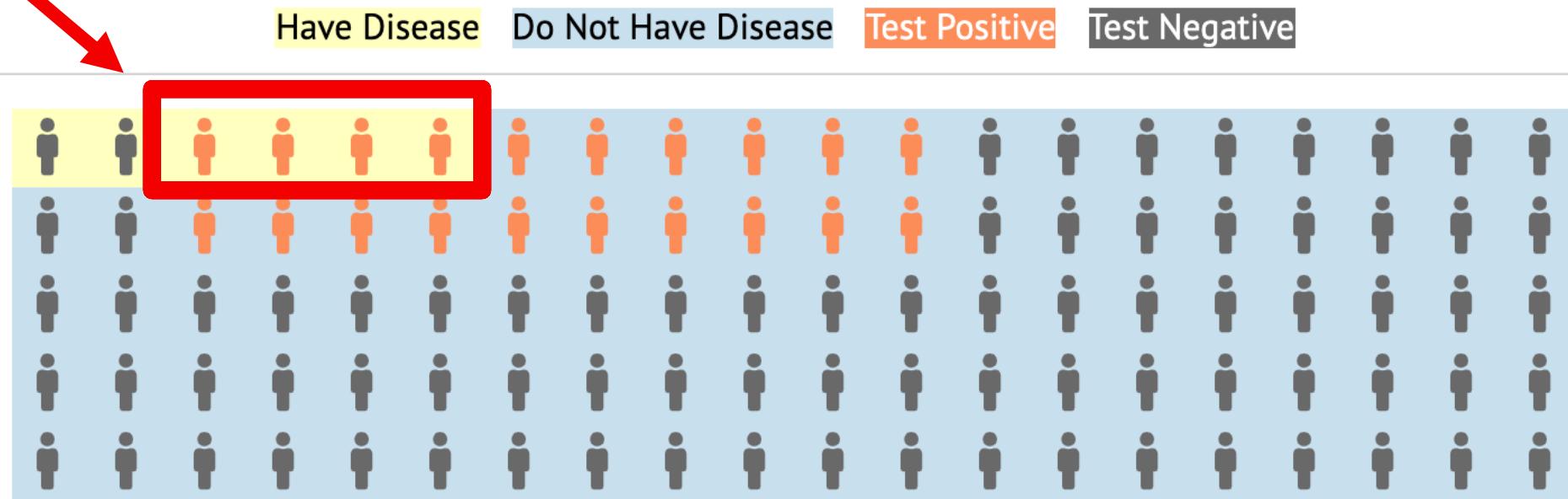
Static Visualization

True Positive

True Negative

False Positive

False Negative



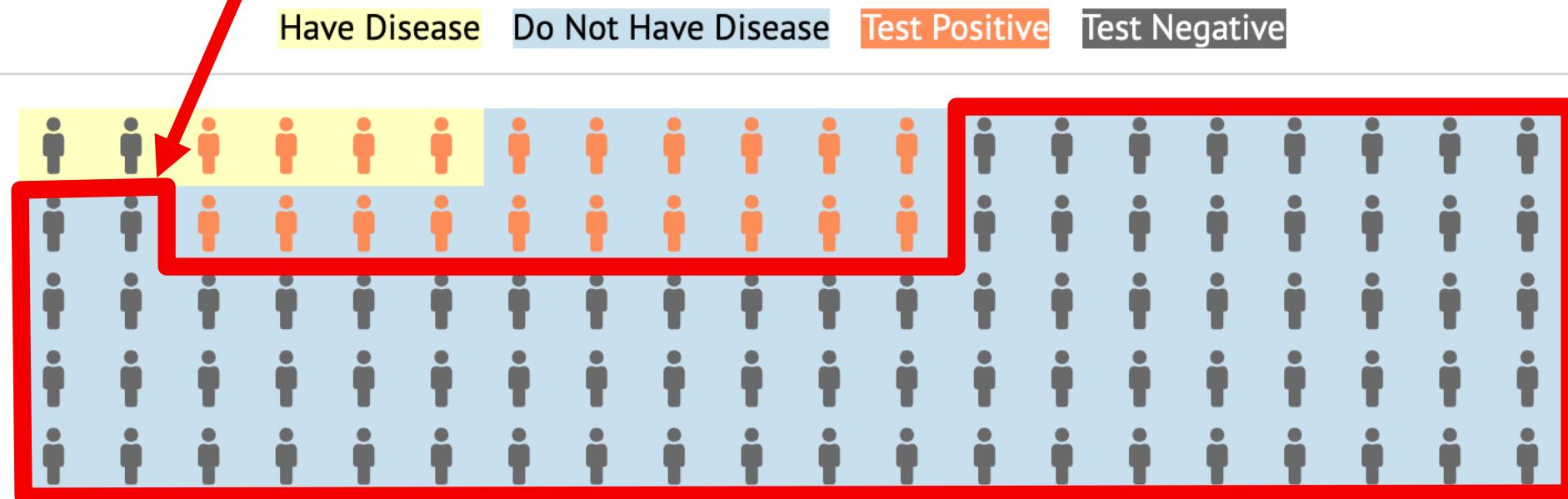
Static Visualization

True Positive

True Negative

False Positive

False Negative



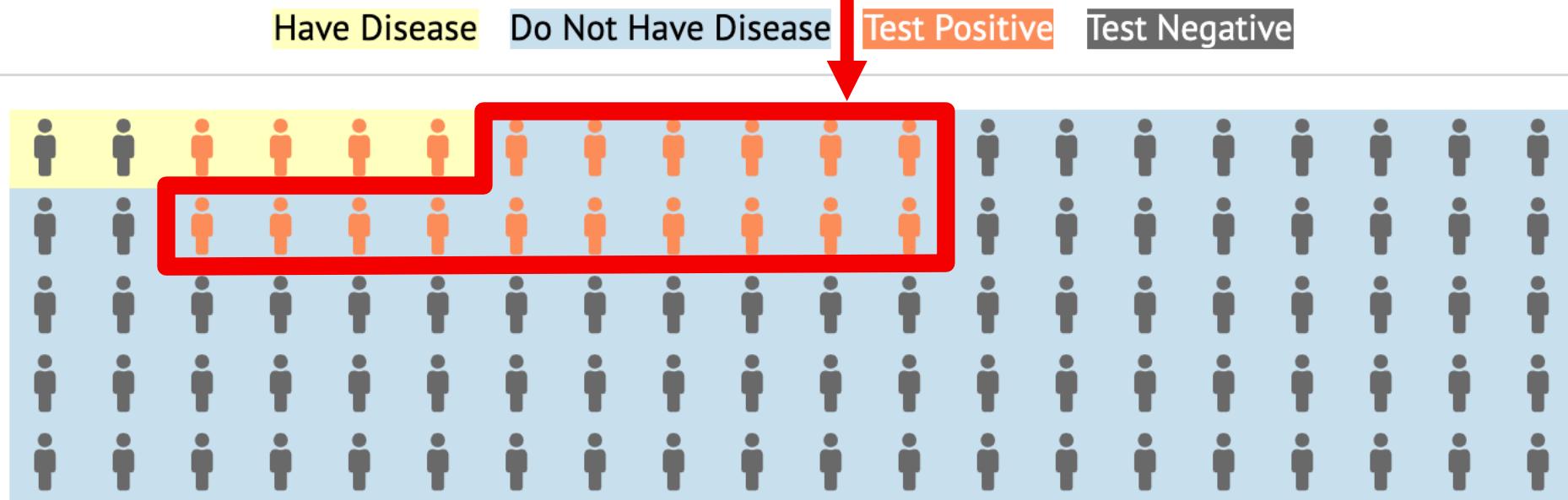
Static Visualization

True Positive

True Negative

False Positive

False Negative



Static Visualization

True Positive

True Negative

False Positive

False Negative



Driving questions

1. Can we help people better perform Bayesian reasoning in the context of medical risk communication?
2. What is the value add of interaction to a *static Bayesian reasoning visualization*?





2 Human-subject Experiments



Visual Data Analysis Group

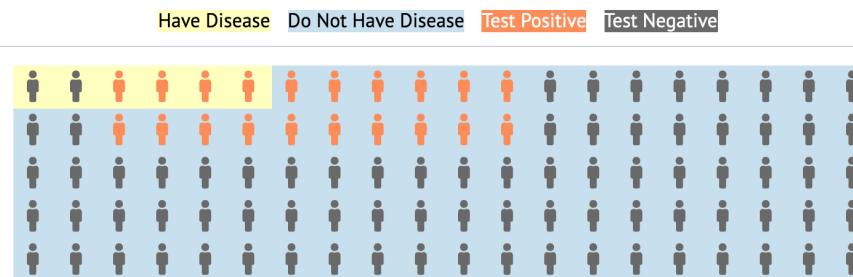
[v]alt

3 Base Visualization Designs

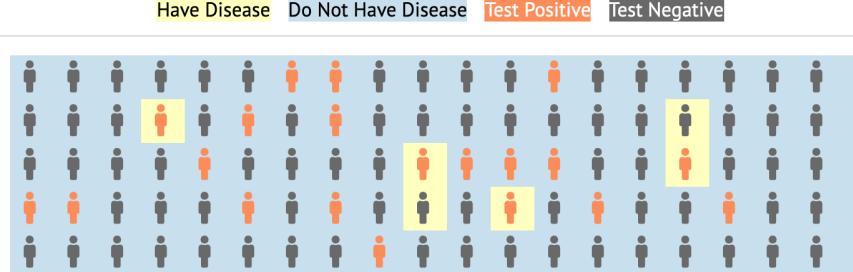
Grouped



Aligned



Randomized



Interactive vs Static

Experiment 1

cbAll

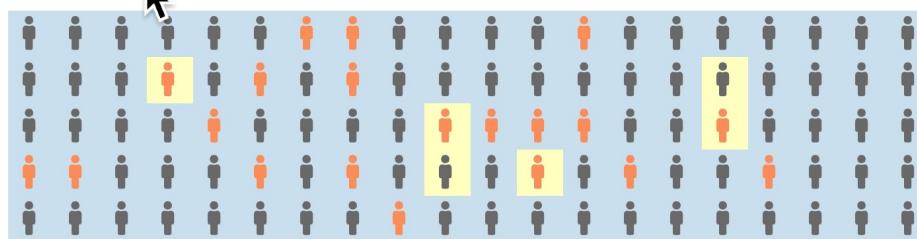


Interactive vs Static

Experiment 2

Use checkboxes in the legend below to add and subtract populations from the visualization.

Have Disease Do Not Have Disease Test Positive Test Negative



cbAll

Use checkboxes in the legend below to add and subtract populations from the visualization.

Have Disease Do Not Have Disease Test Positive Test Negative



cbNone

Drag the representative members of the population shown here onto the visualization below to see similar members of the population. Drag representative members of the population off the visualization to remove similar members.

Has Disease Does Not Have Disease Tests Positive Tests Negative

Visualization shows:



I'm ready to see the question.



Interactive vs Static

Experiment 2

hover

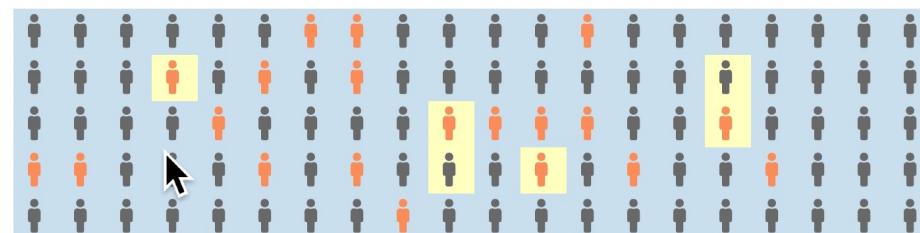
Hover over underlined text to highlight corresponding population in the visualization.

There is a total of 100 people in the population. Out of the 100 people in the population, 6 people actually have the disease. Out of these 6 people, 4 will receive a positive test result and 2 will receive a negative test result. On the other hand, 94 people do not have the disease (i.e., they are perfectly healthy). Out of these 94 people, 16 will receive a positive test result and 78 will receive a negative test result.



tooltips

Hover over an icon for more details.



Experimental Task

Textual description: There is a newly discovered disease, Disease X, which is transmitted by a bacterial infection found in the population. There is a test to detect whether or not a person has the disease, but it is not perfect. Here is some information about the current research on Disease X and efforts to test for the infection that causes it.

There is a total of 100 people in the population. Out of the 100 people in the population, 6 people actually have the disease. Out of these 6 people, 4 will receive a positive test result and 2 will receive a negative test result. On the other hand, 94 people do not have the disease (i.e., they are perfectly healthy). Out of these 94 people, 16 will receive a positive test result and 78 will receive a negative test result.

Questions:

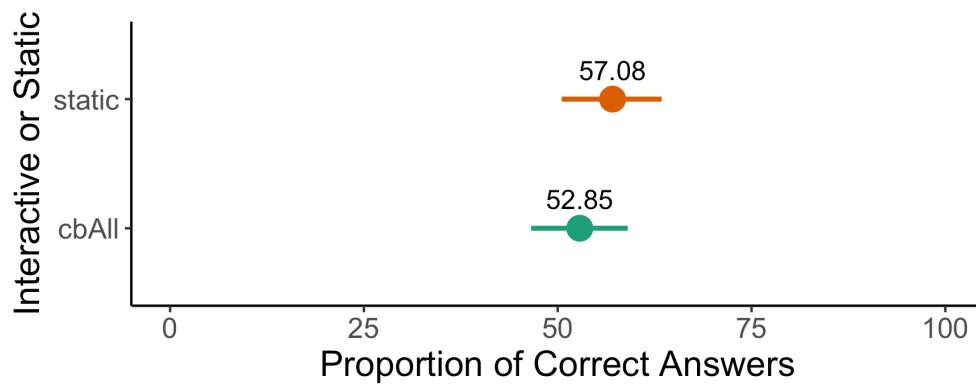
- (a) How many people will test positive? _ _ _
- (b) Of those who test positive, how many will actually have the disease? _ _ _



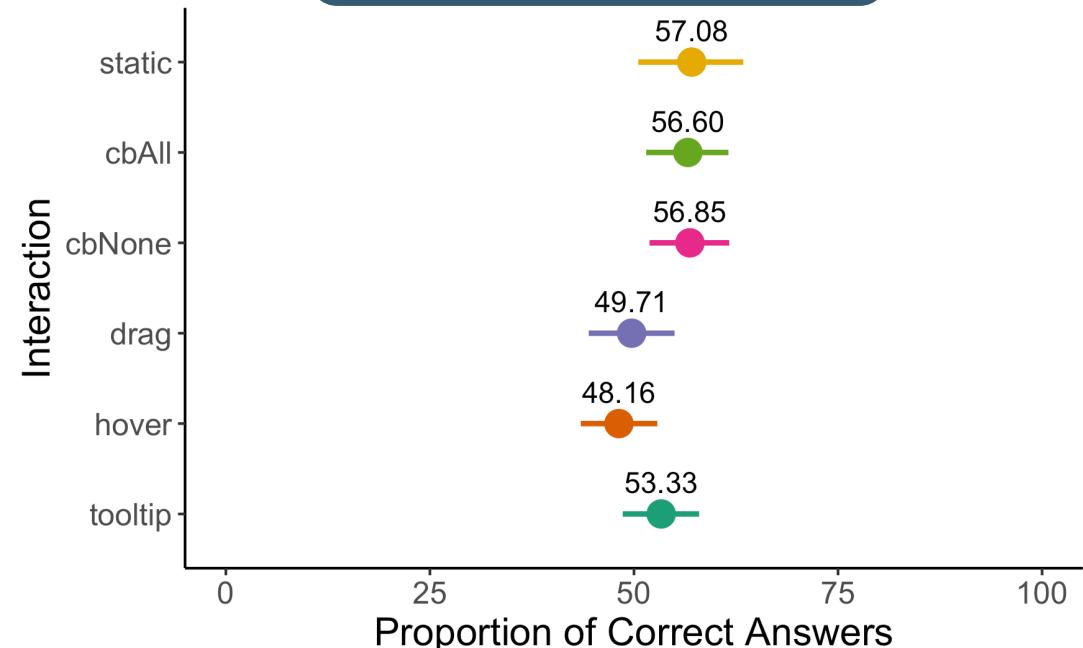
Findings

1. We found no situations in which interaction improves performance on the Bayesian Reasoning task.

Experiment 1



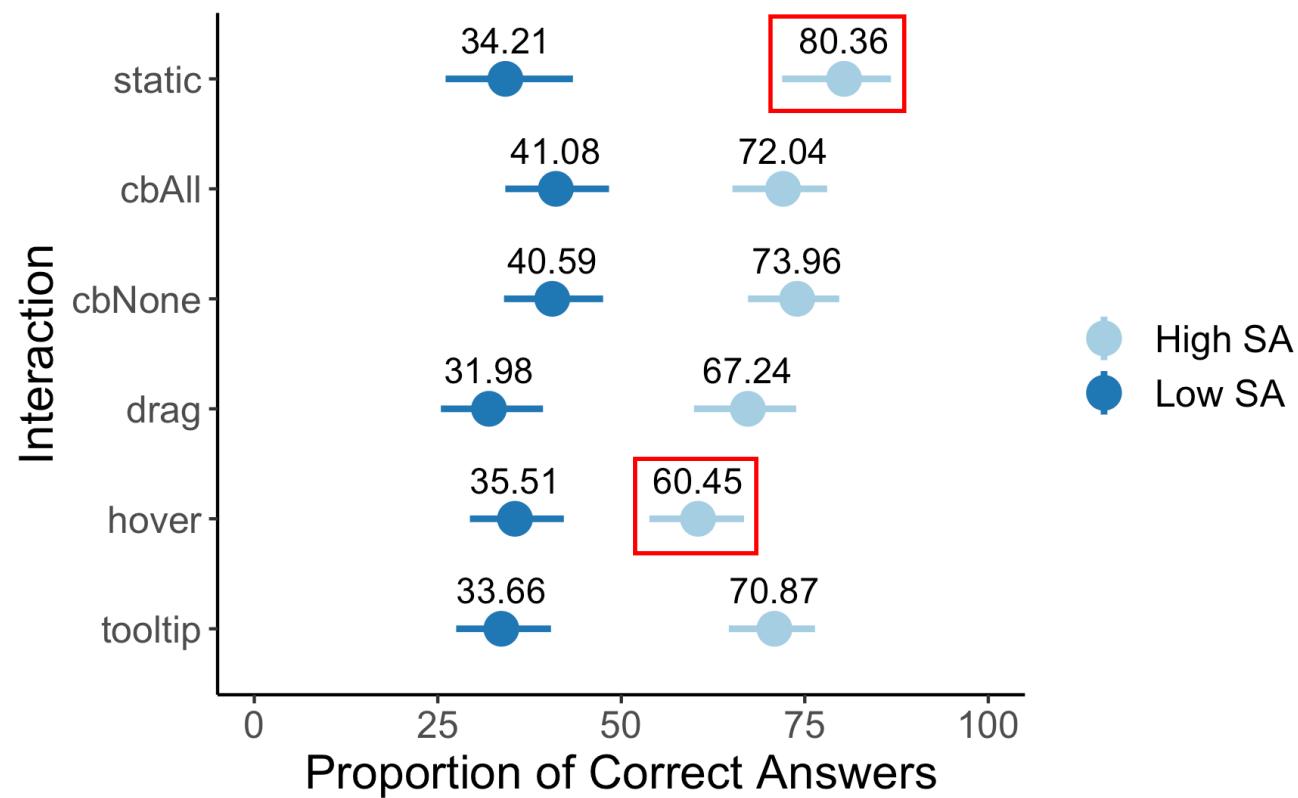
Experiment 2



Findings

2. We observe that **in certain cases interaction results in significantly decreased accuracy** on the Bayesian reasoning task.

Experiment 2



Driving questions

1. Can we help people better perform Bayesian reasoning in the context of medical risk communication?
2. What is the value add of interaction to a *static Bayesian reasoning visualization*?



Driving questions

1. Can we help people better perform Bayesian reasoning in the context of medical risk communication?

Our results suggest interaction may not be the solution.
2. What is the value add of interaction to a *static Bayesian reasoning visualization*?



Driving questions

1. Can we help people better perform Bayesian reasoning in the context of medical risk communication?

Our results suggest interaction may not be the solution.

2. What is the value add of interaction to a *static Bayesian reasoning visualization*?

Our results suggest interaction does not add value in terms of increased accuracy on a Bayesian reasoning task. But more research needs to be done! Interaction is a cost / benefit game and we don't understand all of the nuances yet :)

Does Interaction Improve Bayesian Reasoning with Visualization? (CHI 2021)

Alvitta Ottley,

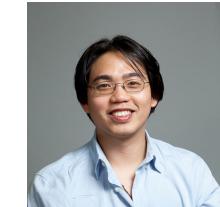
Washington University in St. Louis



W
STL Visual Data Analysis Group

Remco Chang,

Tufts University



[v]alt

Funding sources: Walmart Foundation (OAC-1940175, OAC-1939945, IIS-1452977, DGE-1855886), DARPA D3M (FA8750-17-2-0107), and NSF Grant No. 1755734.



Summary

Today we:

- Reviewed Visualization for Communication
- Took a look at a Visualization Research Project