

Data visualization strategies and tools for microbial genomic epidemiology

Anamaria Crisan

Research Scientist

Tableau Research

Tableau, Seattle, WA



@amcrisan



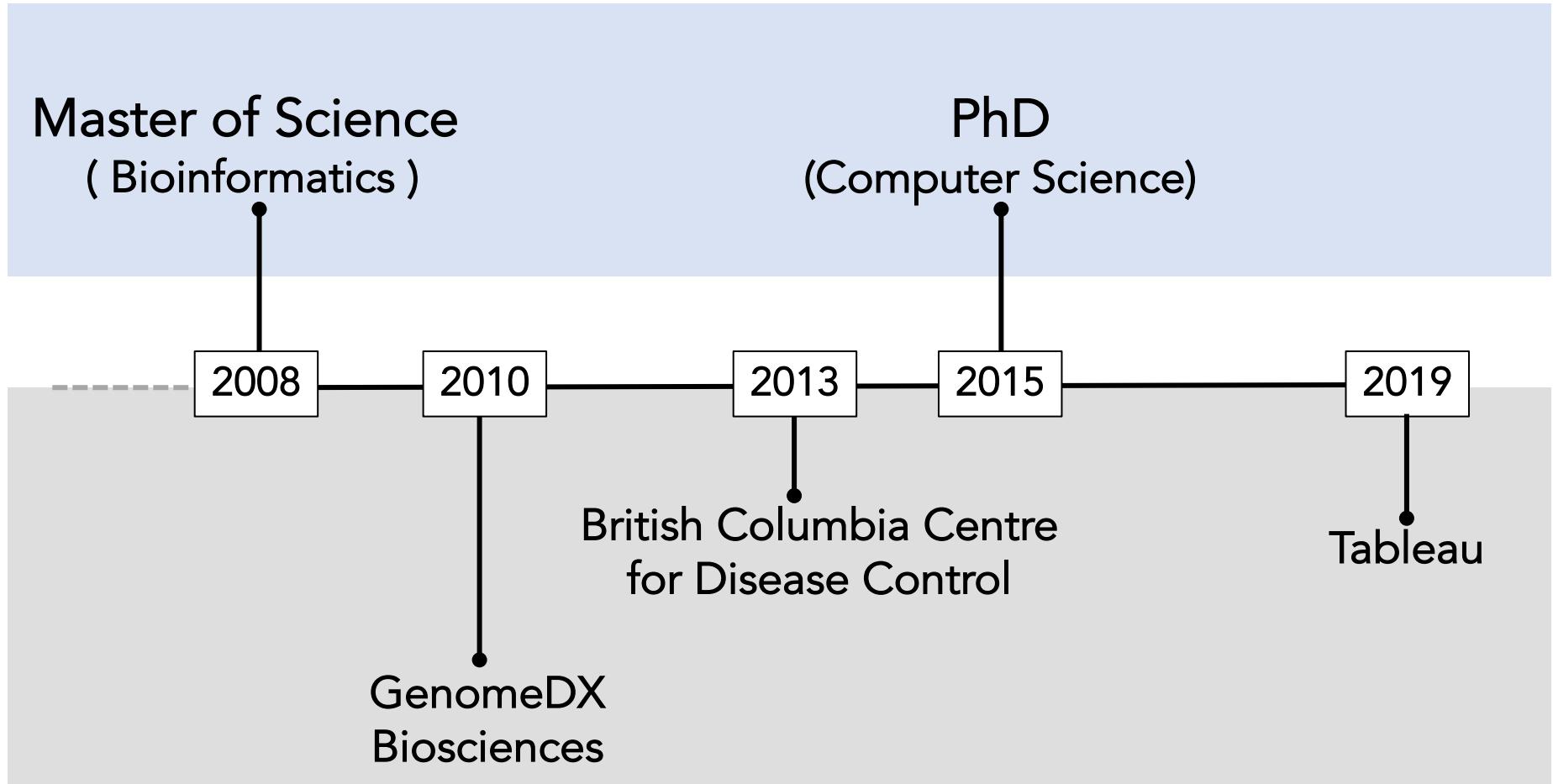
acrisan@tableau.com



<https://amcrisan.github.io>



Research Scientist
Tableau Research



What we'll talk about

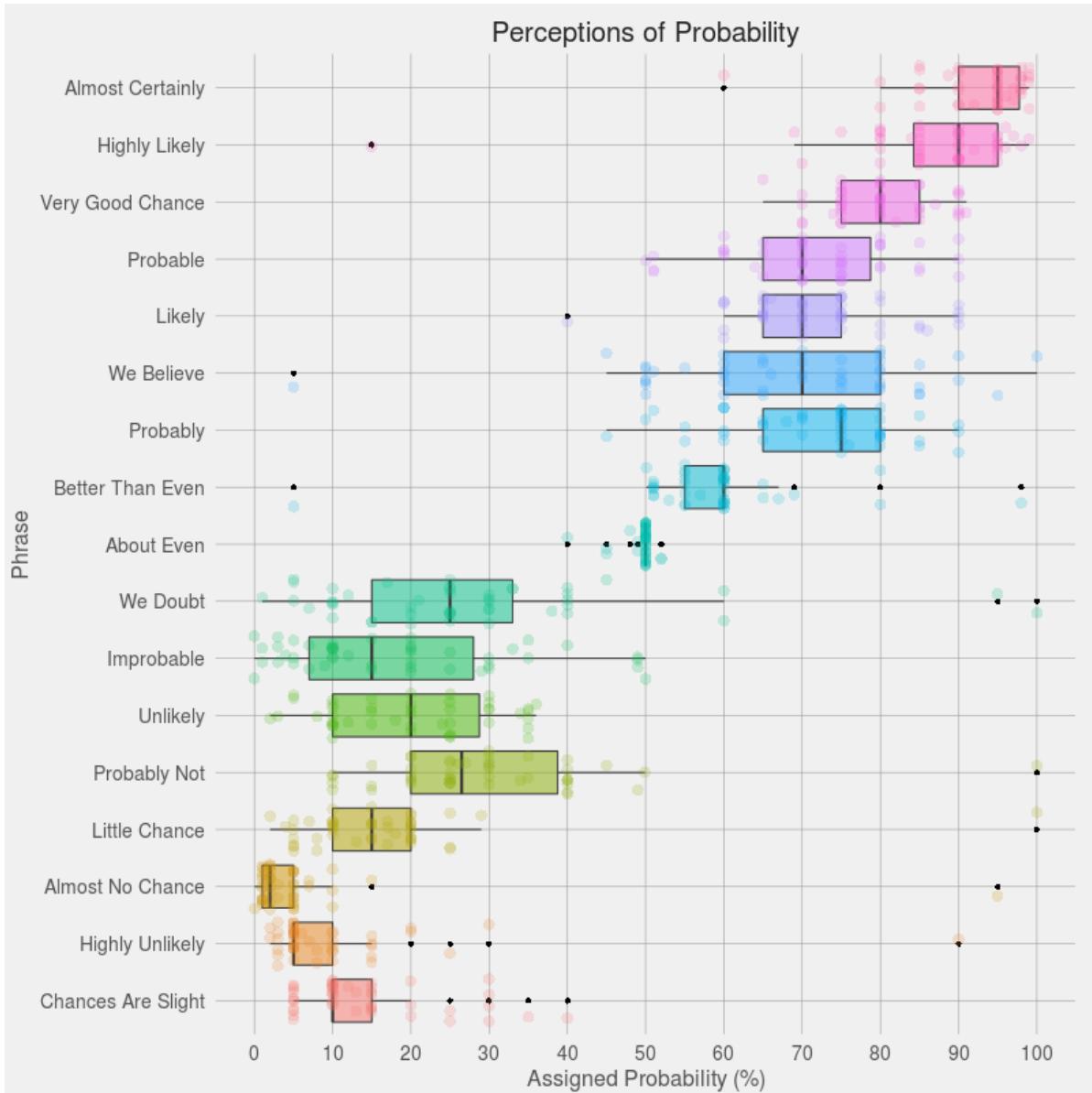
Why should we visualize data?

How should we visualize data?

What datavis tools are available?

**Why should we
visualize data?**

Translating Numbers to Words



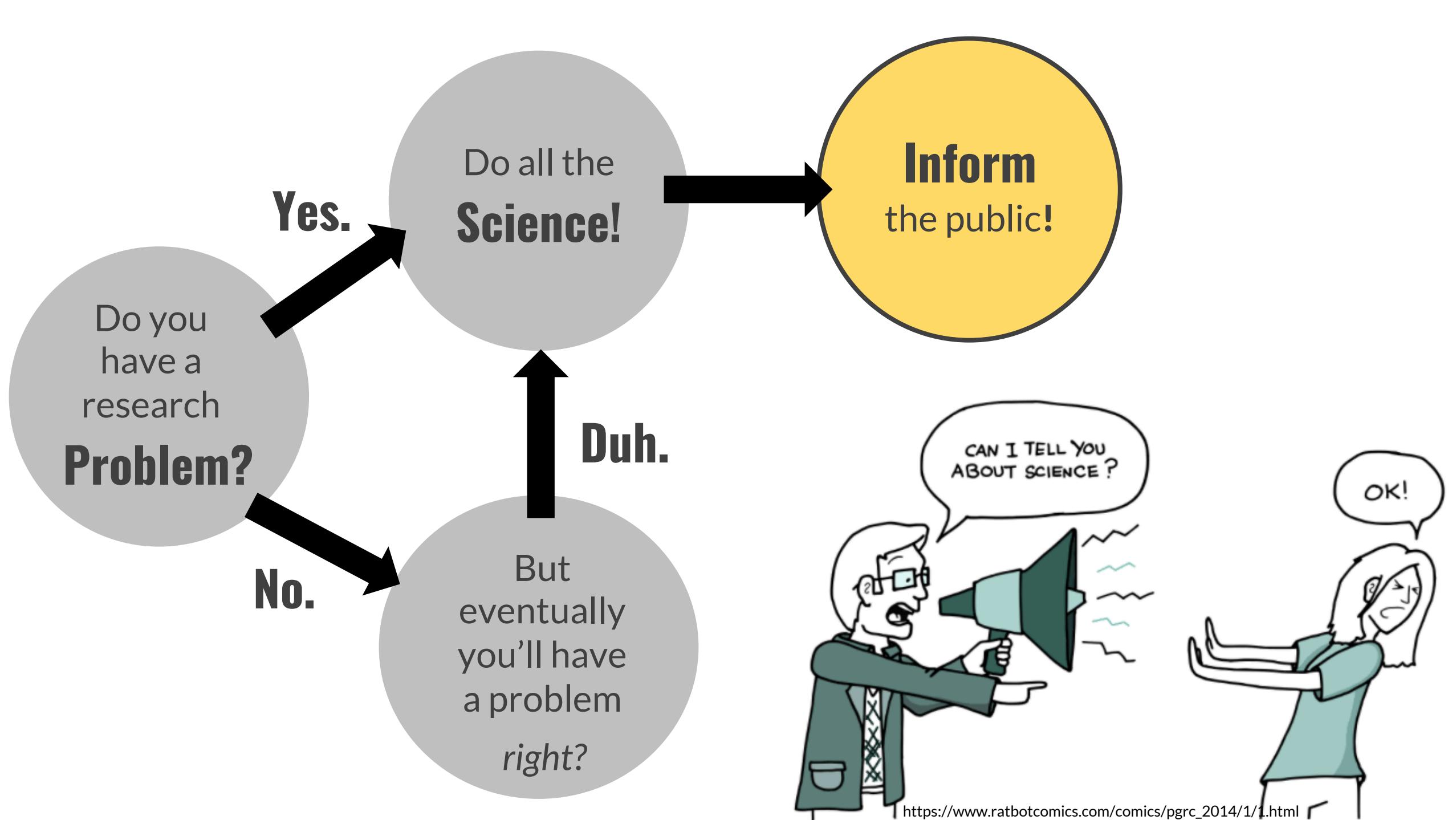
It is not always easy to reason consistently with numbers

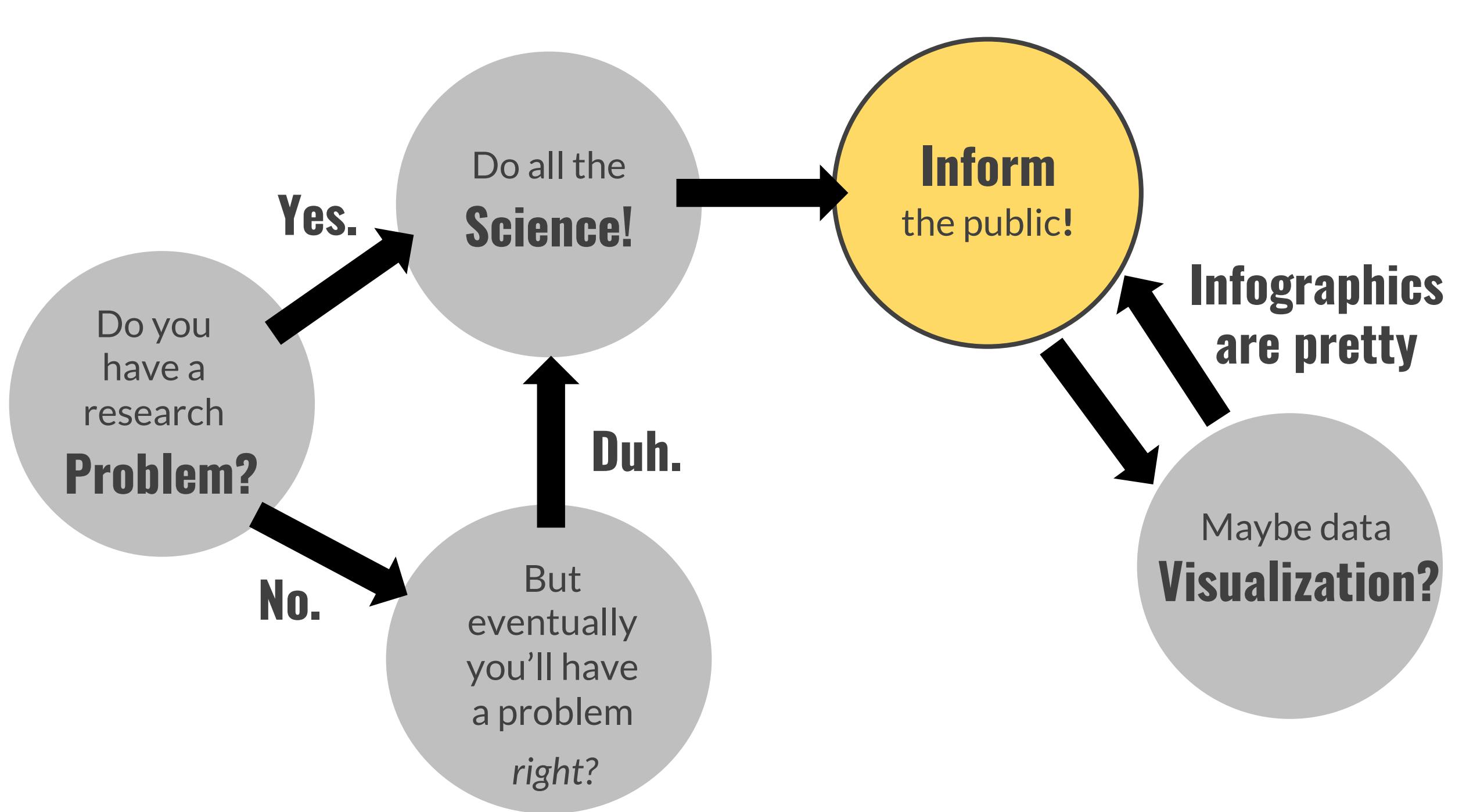
Data Visualization is a Powerful Medium

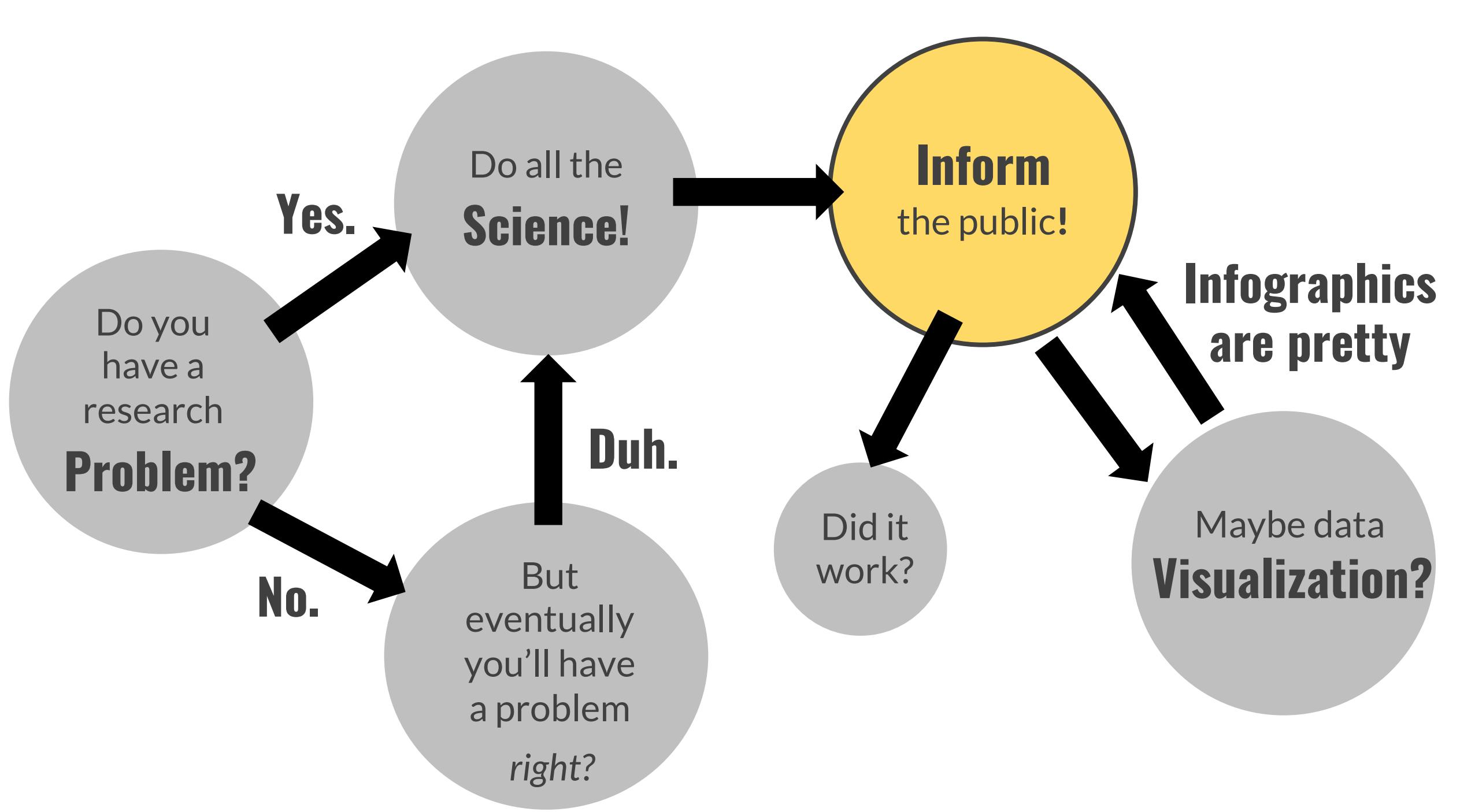


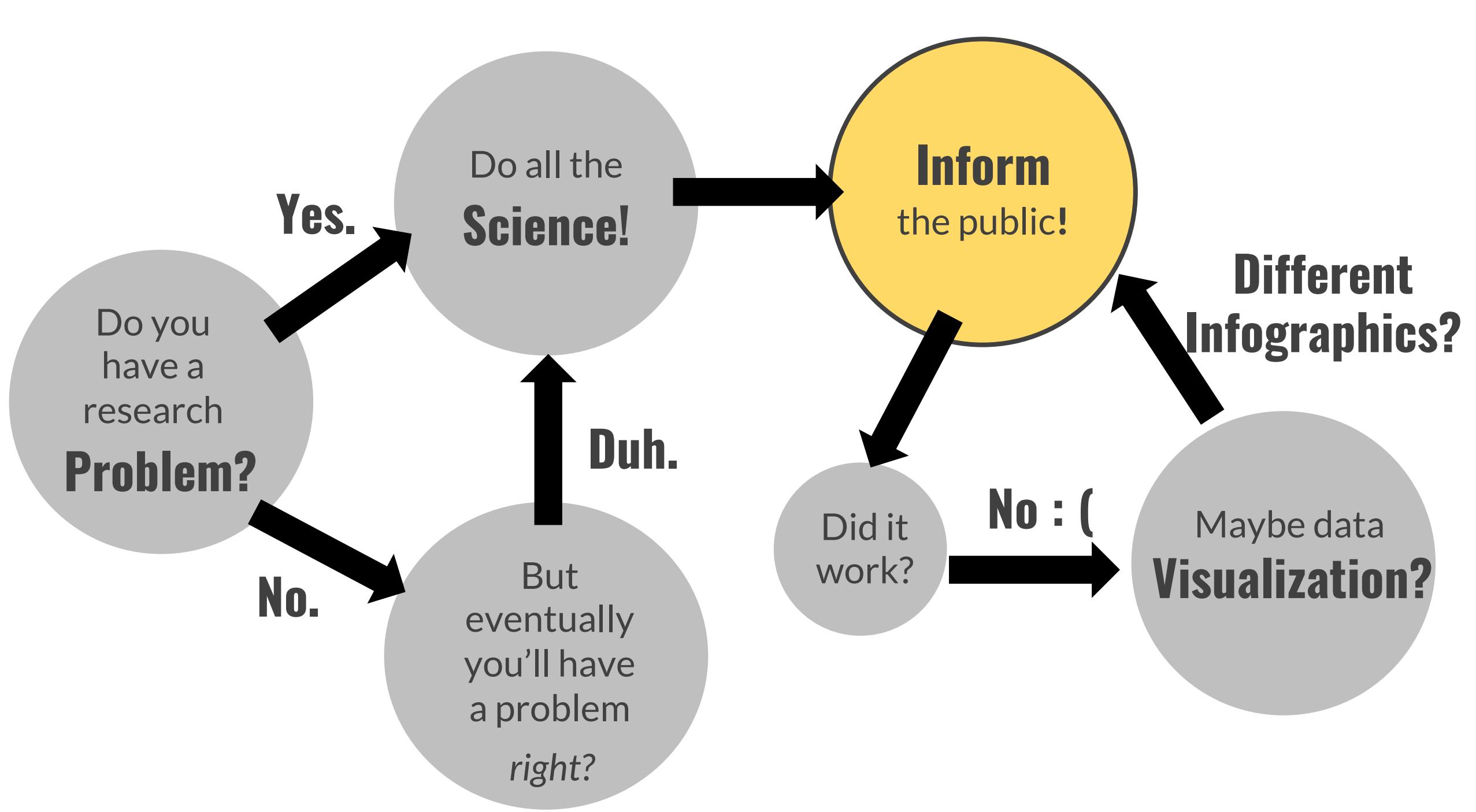
Role of data visualization
in the current paradigm of
scientific research

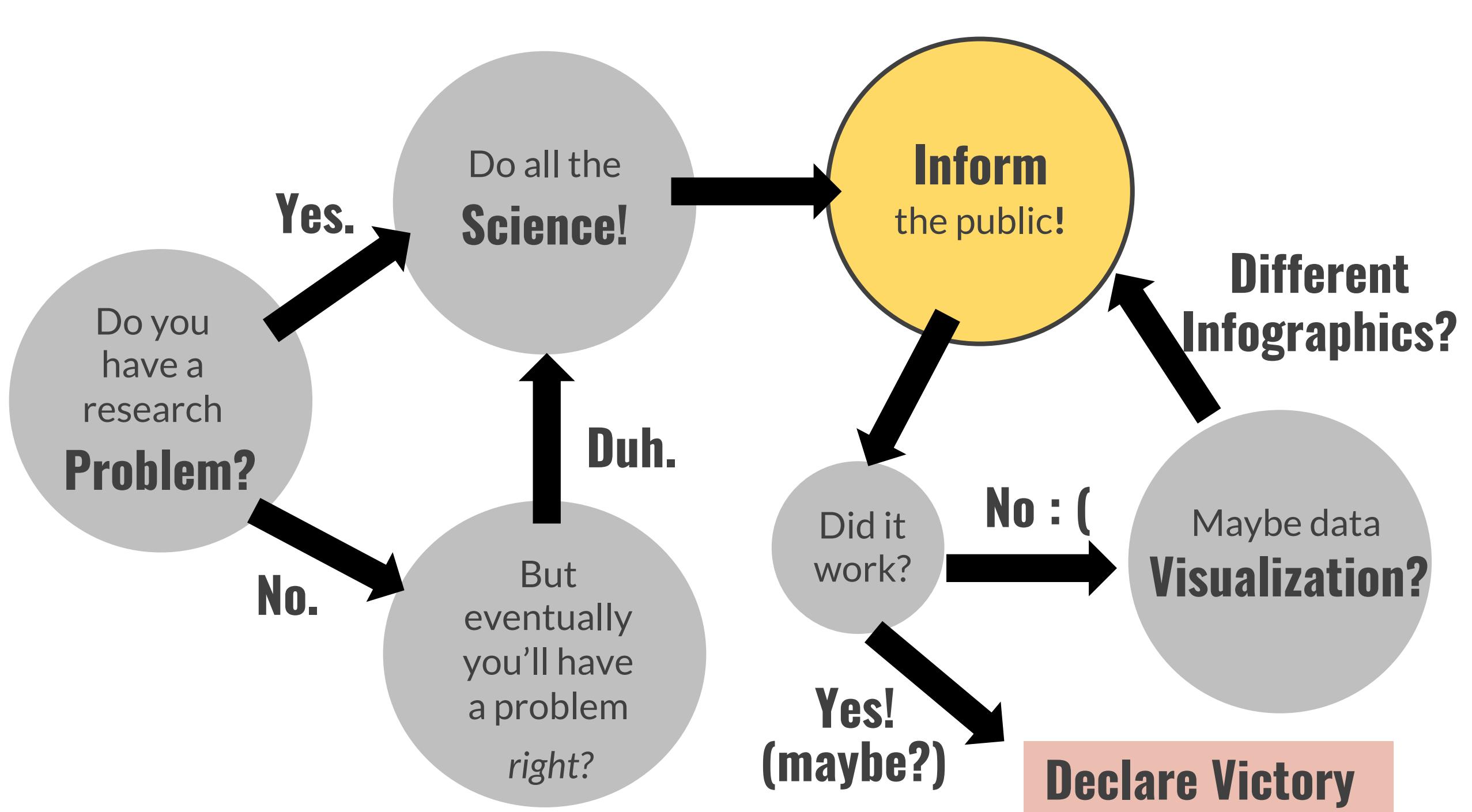
= Communication







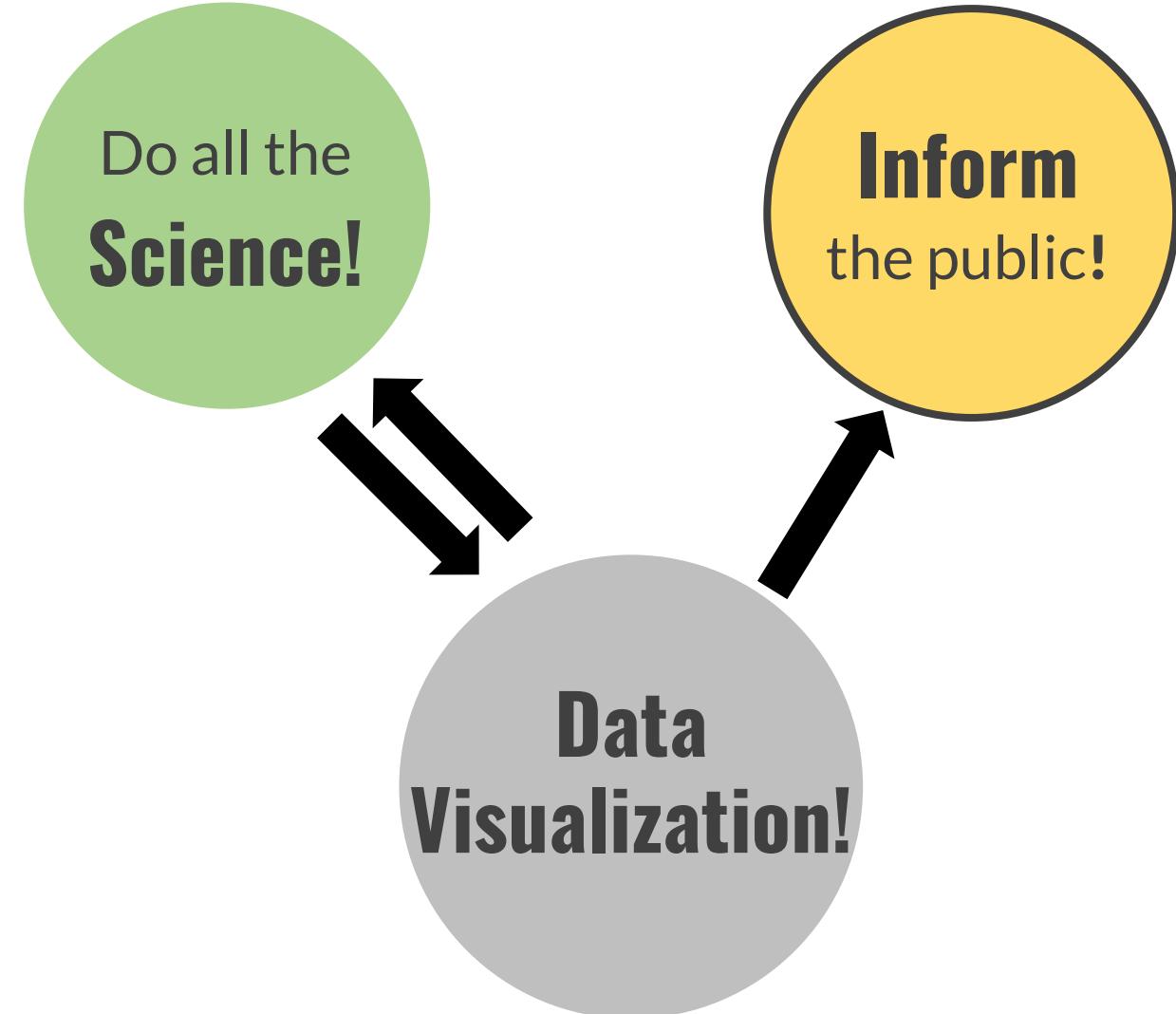




Limitation #1 : Missed Opportunity in Exploration

Missed Opportunity for Exploration

- Exploration is looking at your data, trying different analysis methods, assessing if there are outliers or missing data etc.



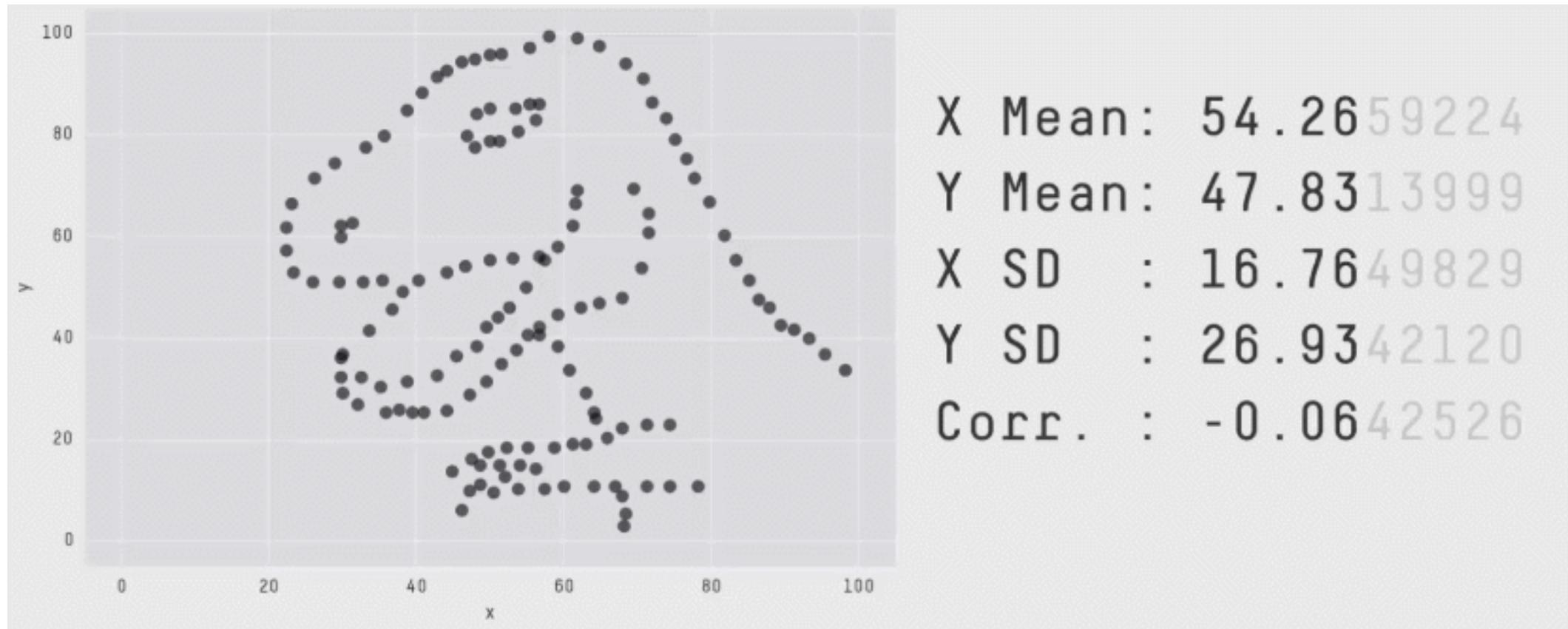
Limitation #1 : Missed Opportunity in Exploration

Same stats, different graphs

X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526

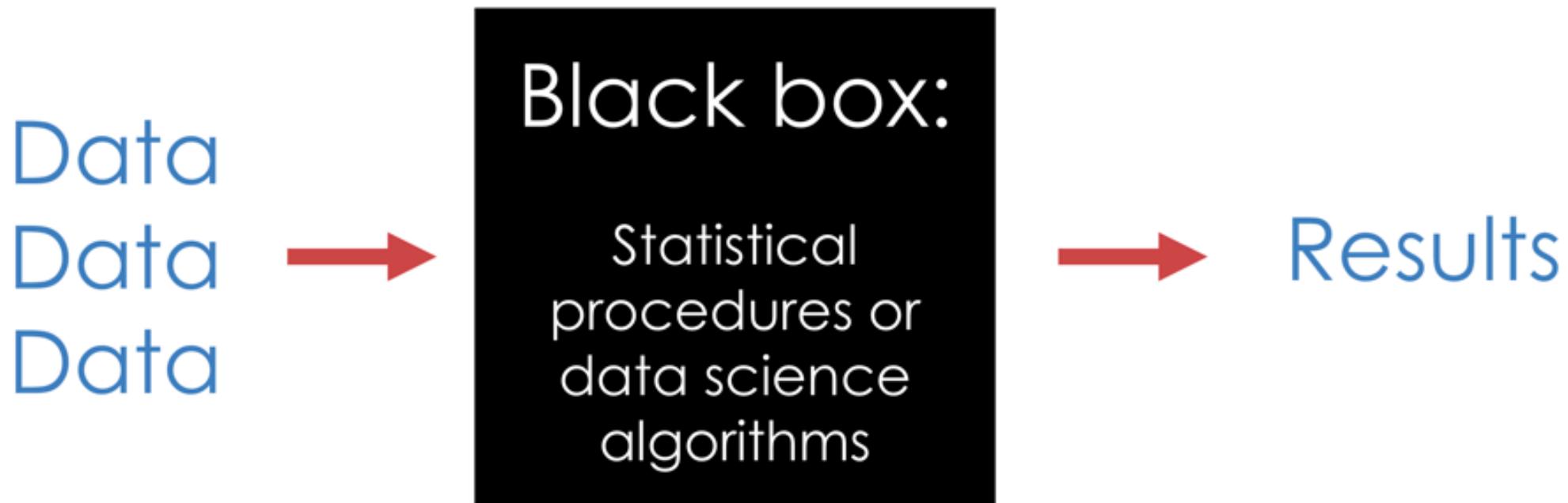
Limitation #1 : Missed Opportunity in Exploration

Same stats, different graphs (*Datasaurus*)



Limitation #1 : Missed Opportunity in Exploration

Opening up the machine learning black box



Limitation #1 : Missed Opportunity in Exploration

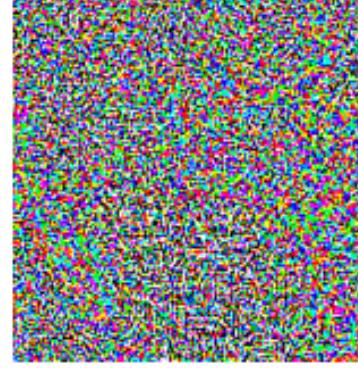
Chihuahua or muffin?



Mop or sheep dog?

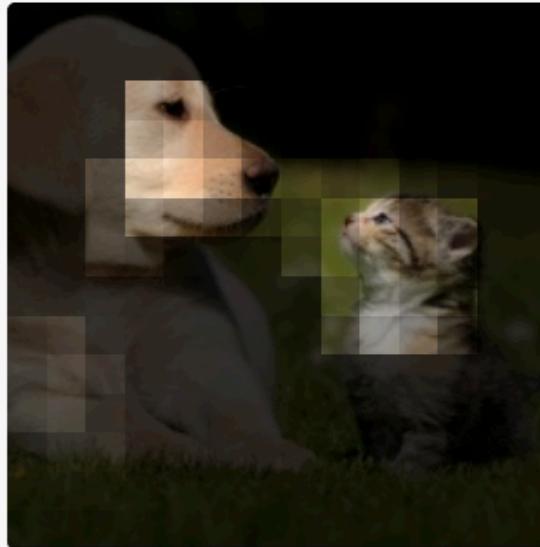


Limitation #1 : Missed Opportunity in Exploration

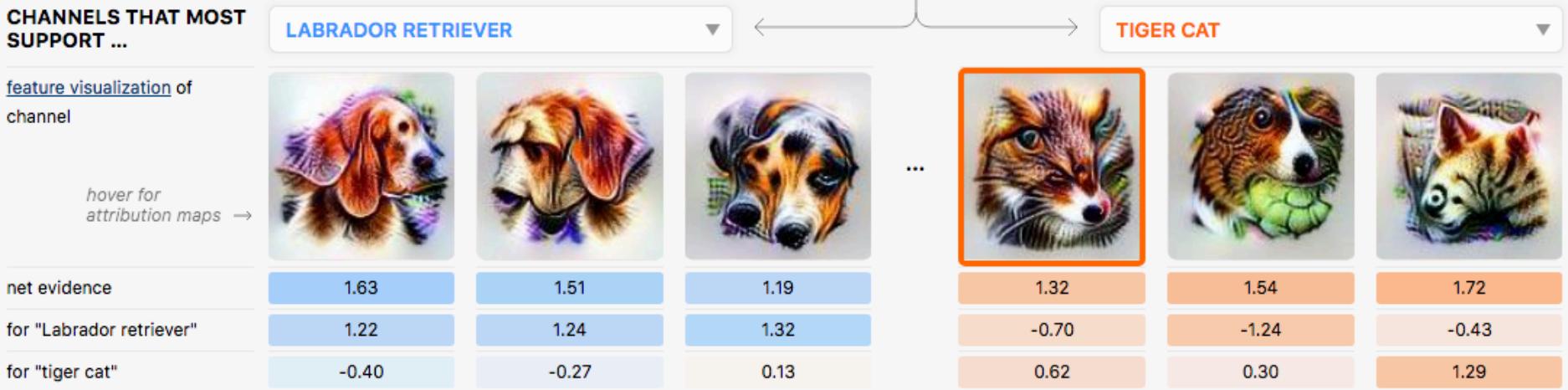
$$\begin{array}{ccc} \text{} & + .007 \times & \text{} \\ \boldsymbol{x} & & \text{sign}(\nabla_{\boldsymbol{x}} J(\theta, \boldsymbol{x}, y)) \\ \text{"panda"} & & \text{"nematode"} \\ 57.7\% \text{ confidence} & & 8.2\% \text{ confidence} \\ & = & \\ \text{} & & \epsilon \text{sign}(\nabla_{\boldsymbol{x}} J(\theta, \boldsymbol{x}, y)) \\ & & \text{"gibbon"} \\ & & 99.3 \% \text{ confidence} \end{array}$$

Example : Trying to understand the black box

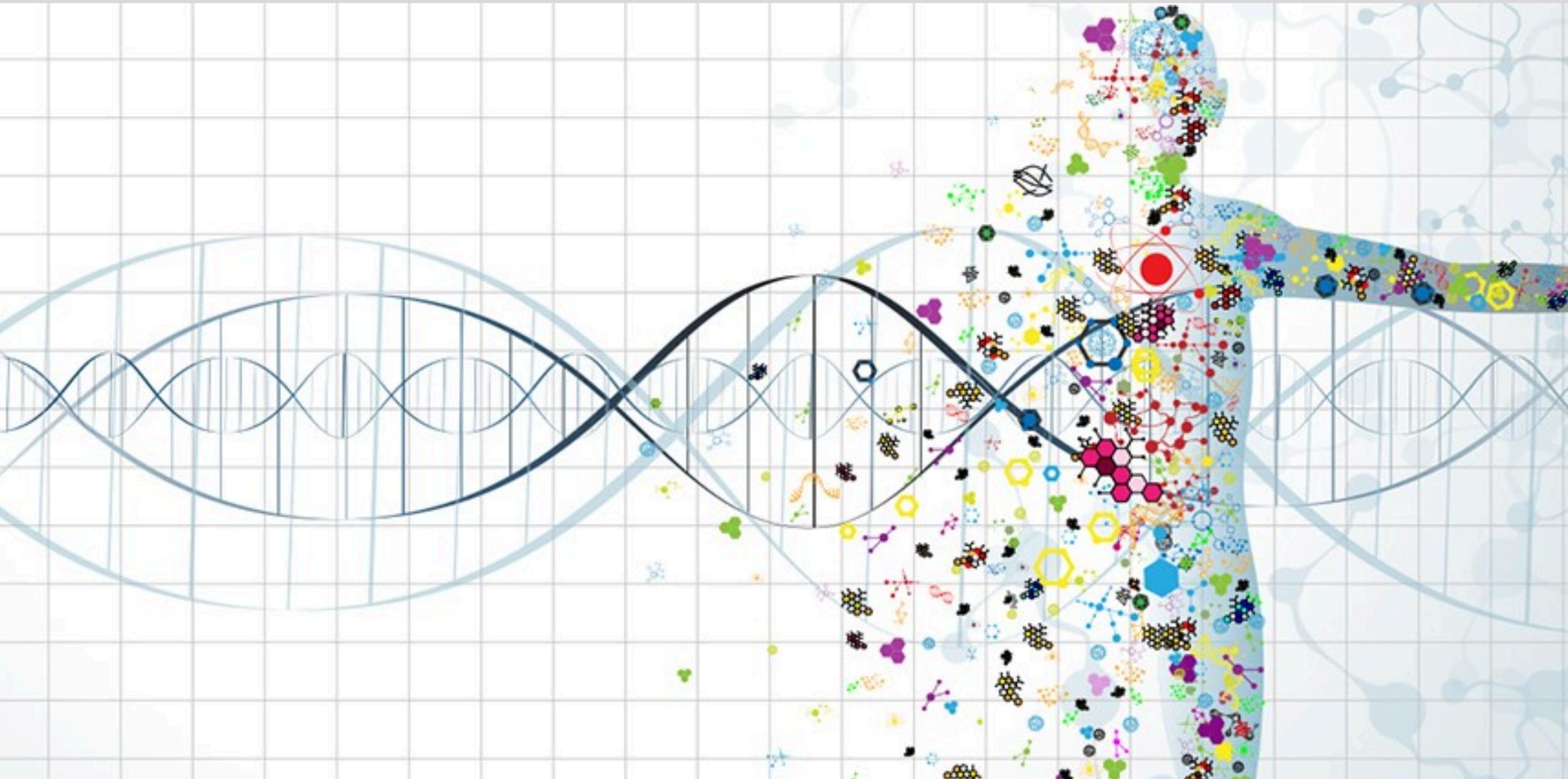
For instance, by combining feature visualization (*what is a neuron looking for?*) with attribution (*how does it affect the output?*), we can explore how the network decides between labels like **Labrador retriever** and **tiger cat**.



Several floppy ear detectors seem to be important when distinguishing dogs, whereas pointy ears are used to classify "tiger cat".



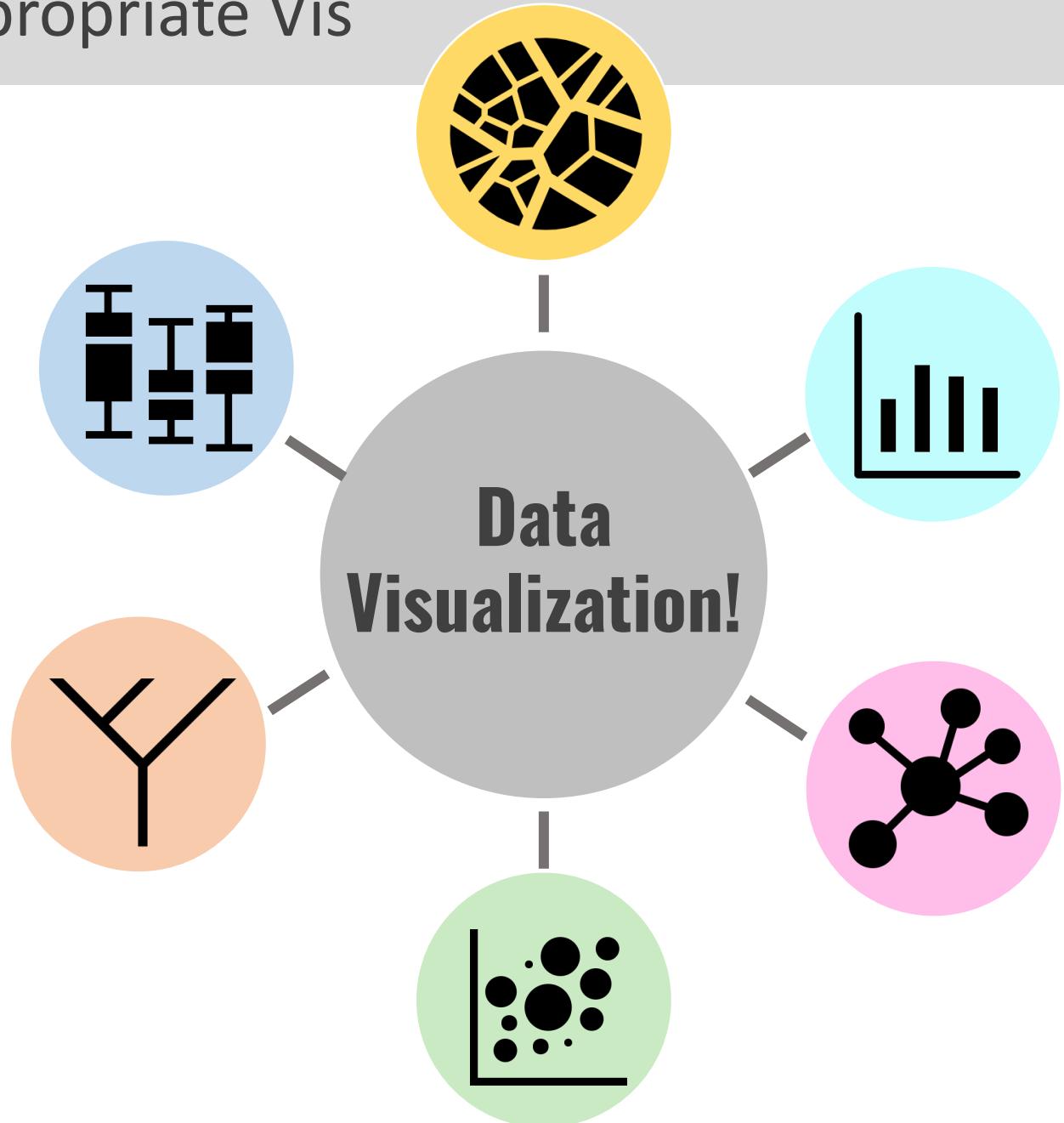
Health data are complex to analyze and visualization



Limitations #2 : Identifying the Appropriate Vis

Selecting the appropriate data visualization is challenging

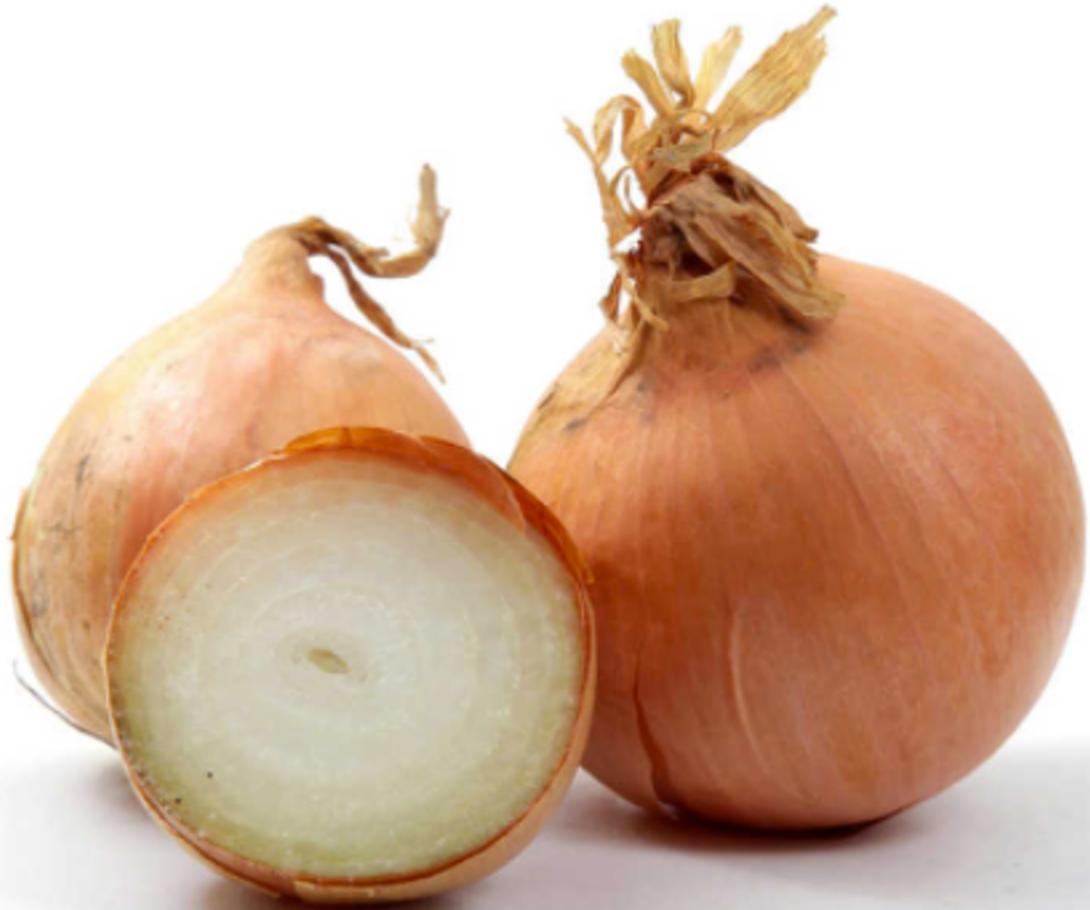
- True for exploration & communication applications



Visualization Design ALSO matters

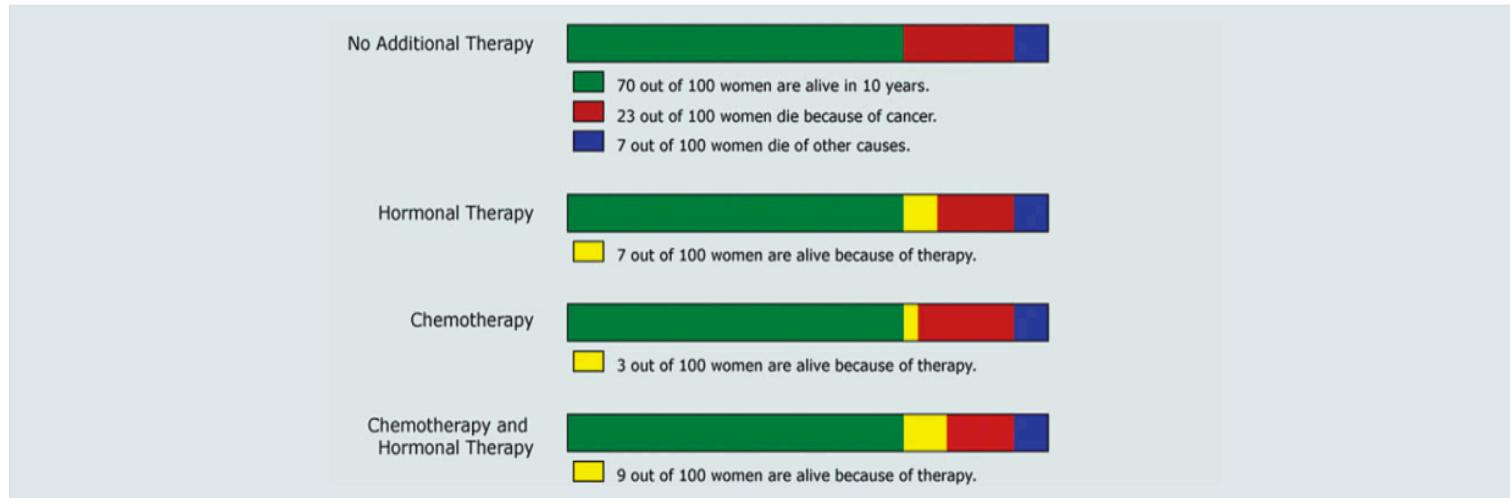
Design is a funny word.
Some people think
design means how it
looks. But of course, if
you dig deeper, it's really
how it works.

- Steve Jobs

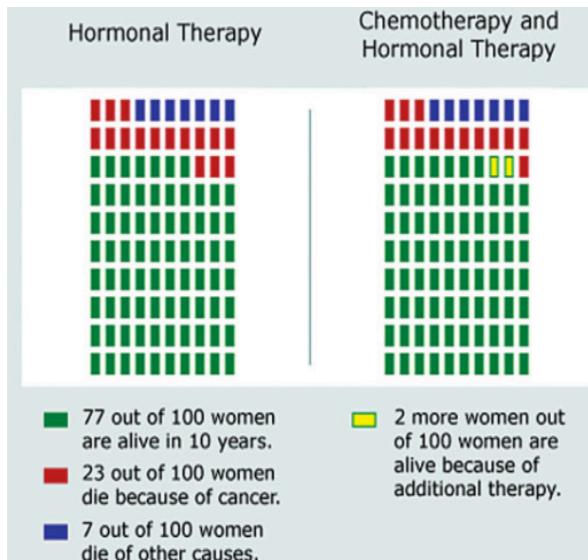


Example: Communicating Survival Benefit of Cancer Therapy

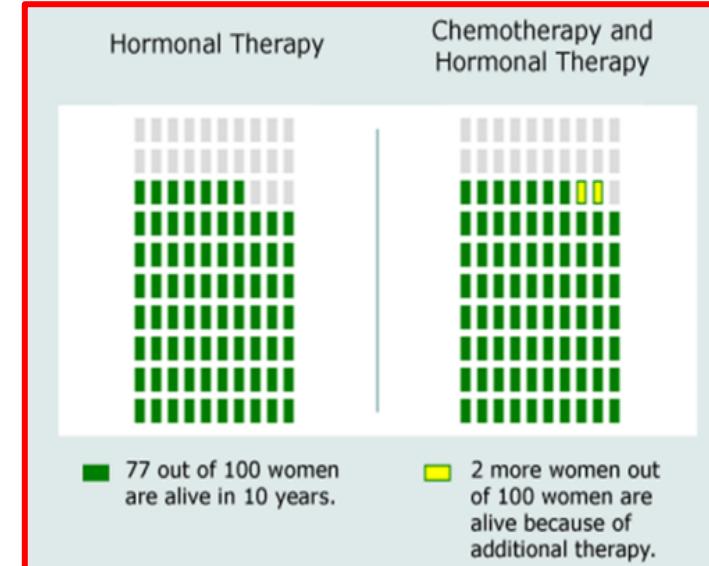
Baseline Visualization



Alternative 1

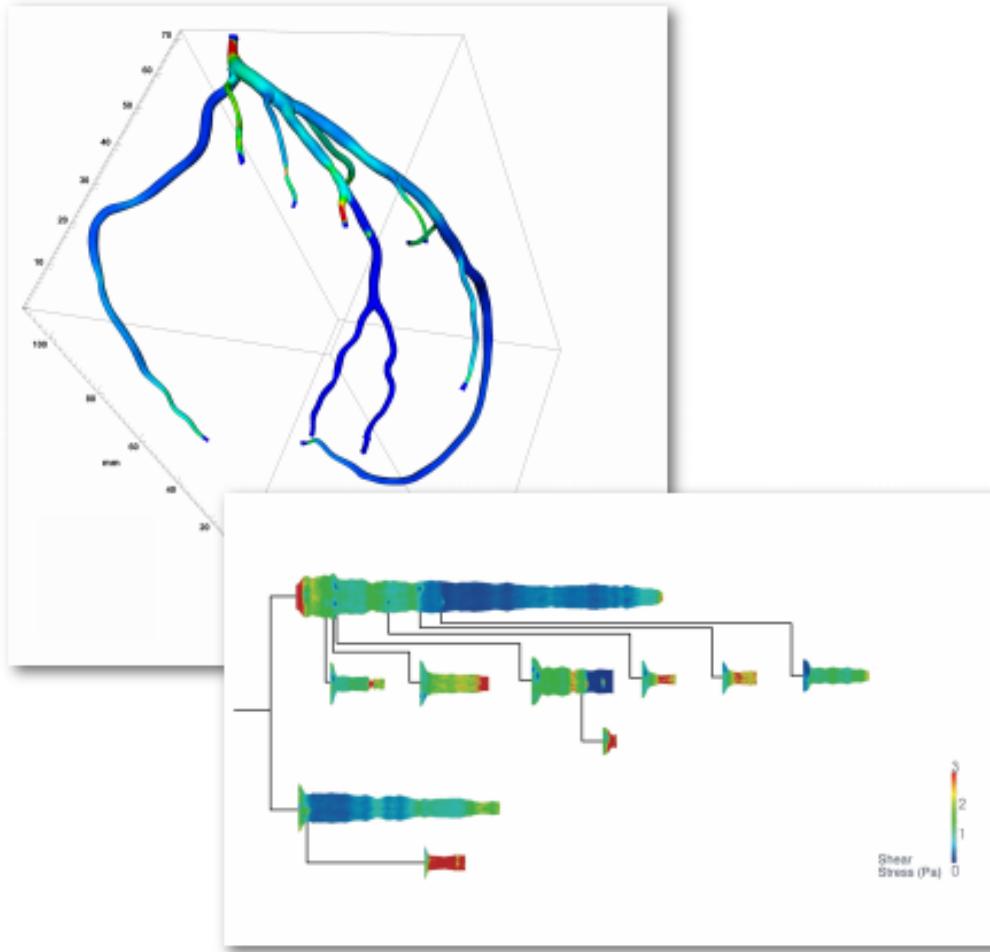


Alternative 2

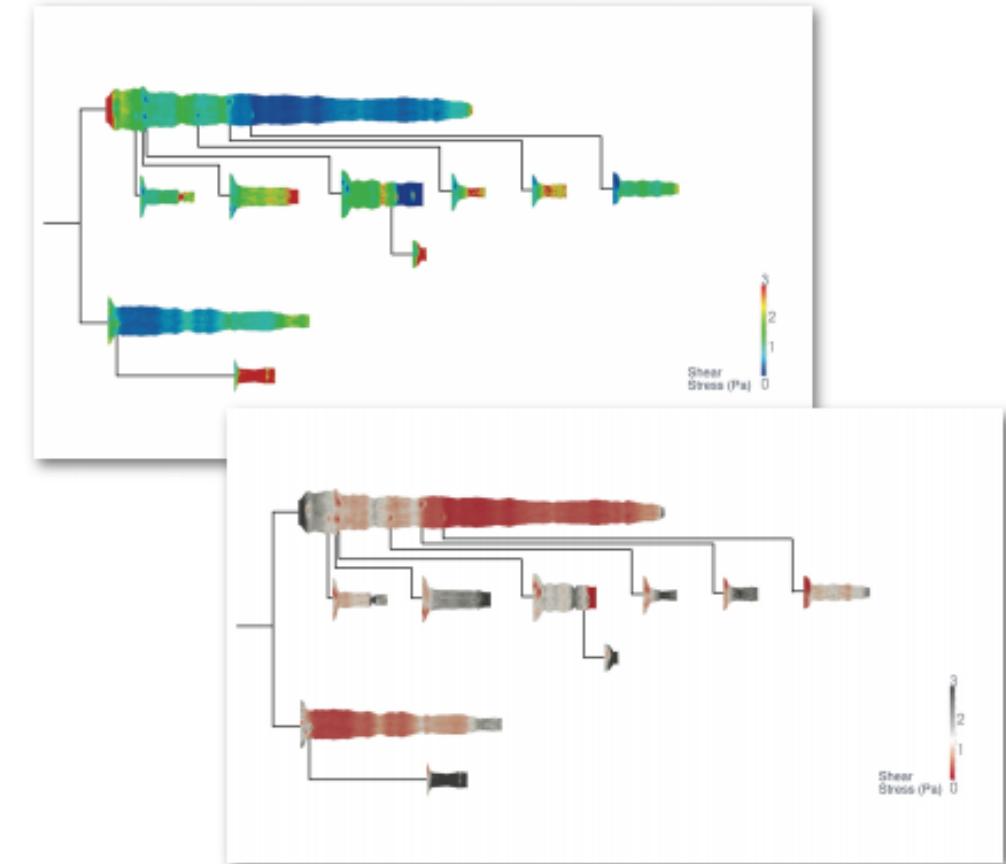


Example: Visualizing Arteries of the Heart for Surgery Planning

3D vs. 2D



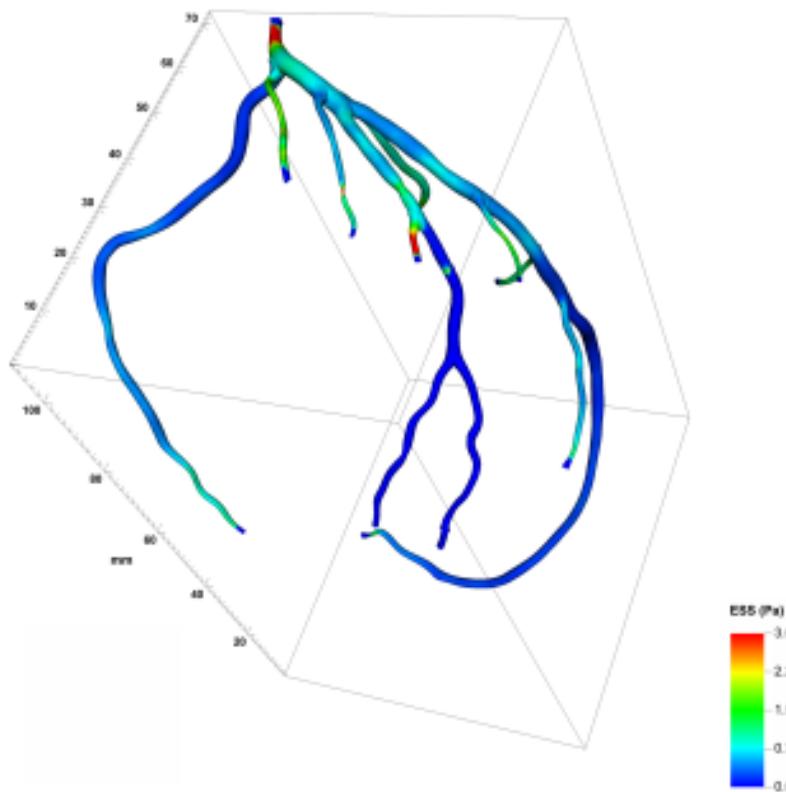
rainbow vs. diverging



Example: Visualizing Arteries of the Heart for Surgery Planning

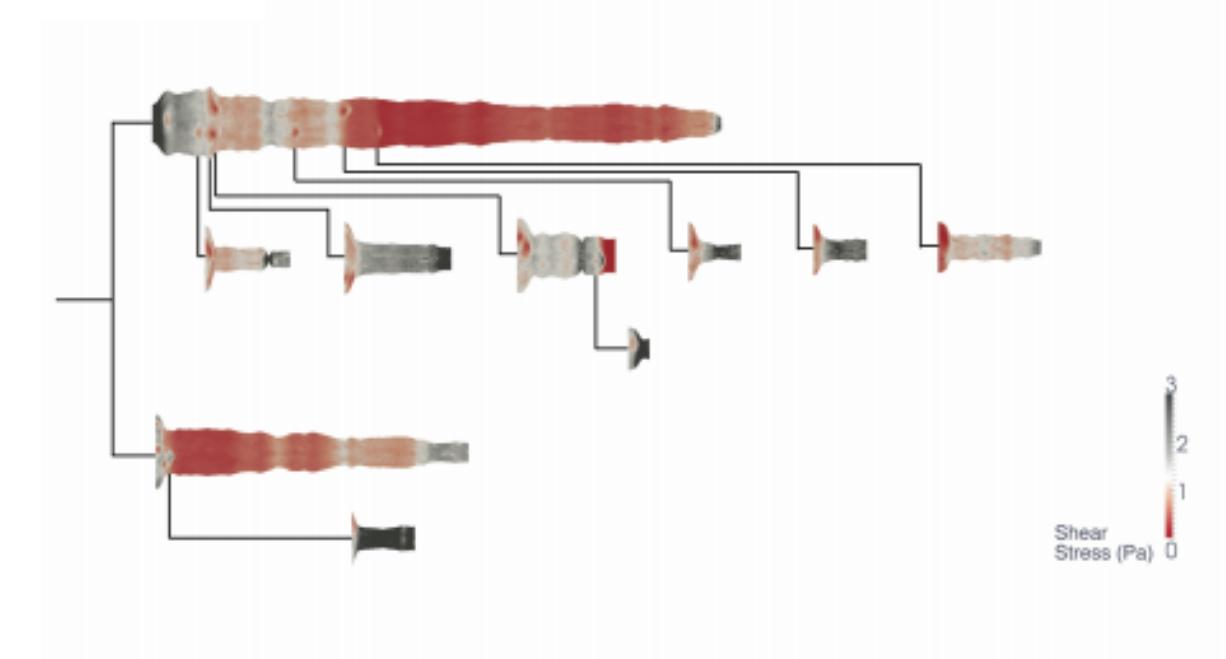
EXISTING STANDARD

Accuracy : 39%



REVISED VISUALIZATION

Accuracy: 91%



There are two aspects of visualizations to think about:

Is it the appropriate visualization?

How should we visualize data?

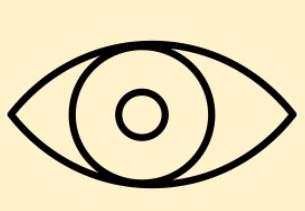
How do you make a visualization?

What datavis tools are available?

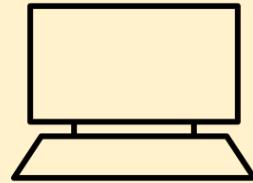


**How should we
visualize data ?**

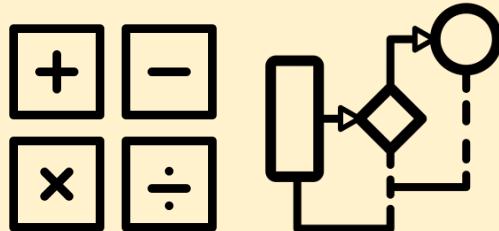
Cross Cutting Disciplines in Information Visualization



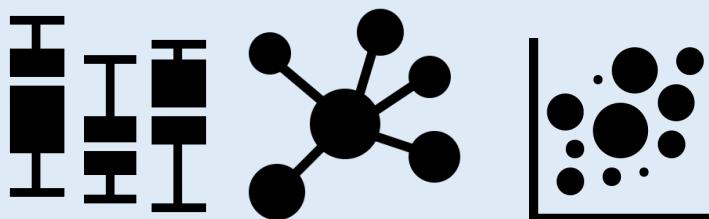
Human Perception & Cognition



Computer Graphics

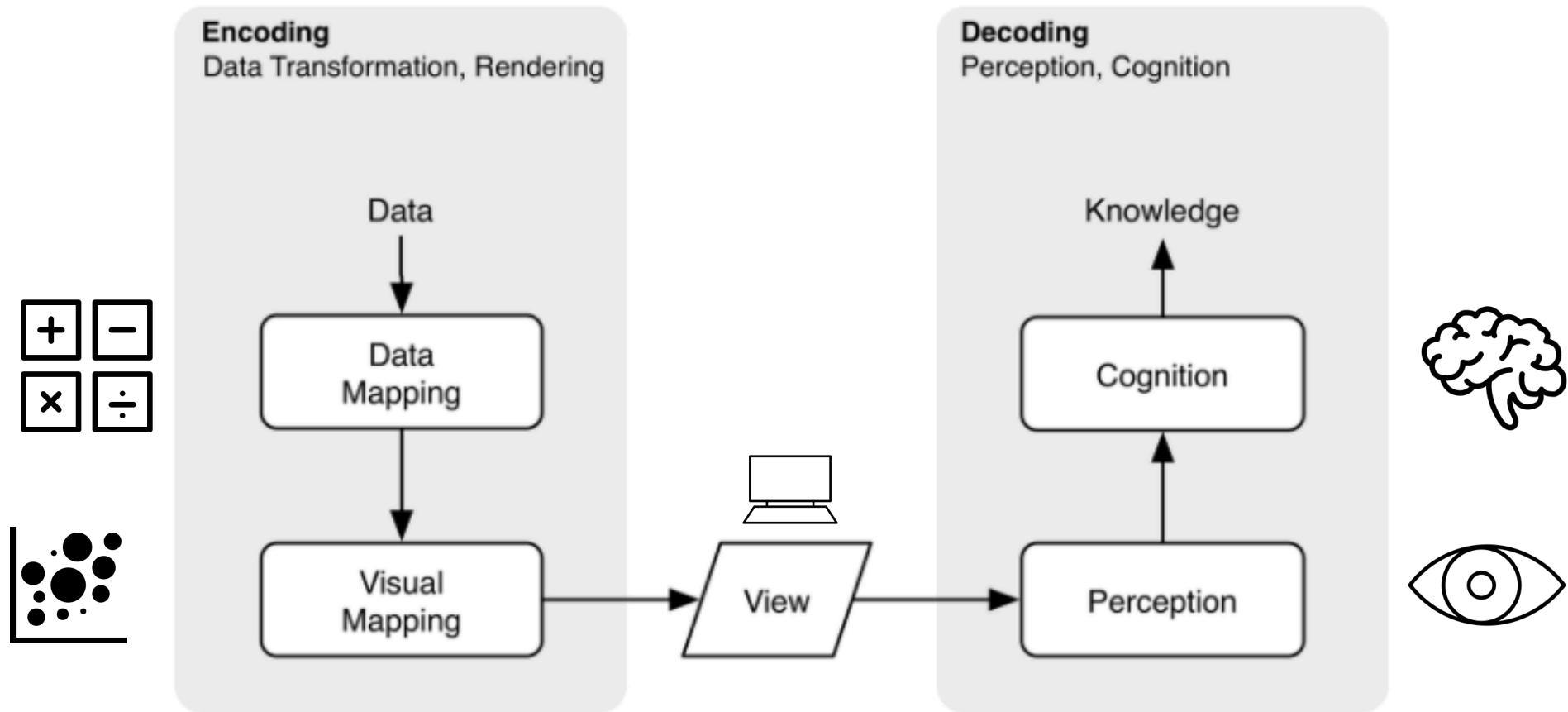


Data Analysis

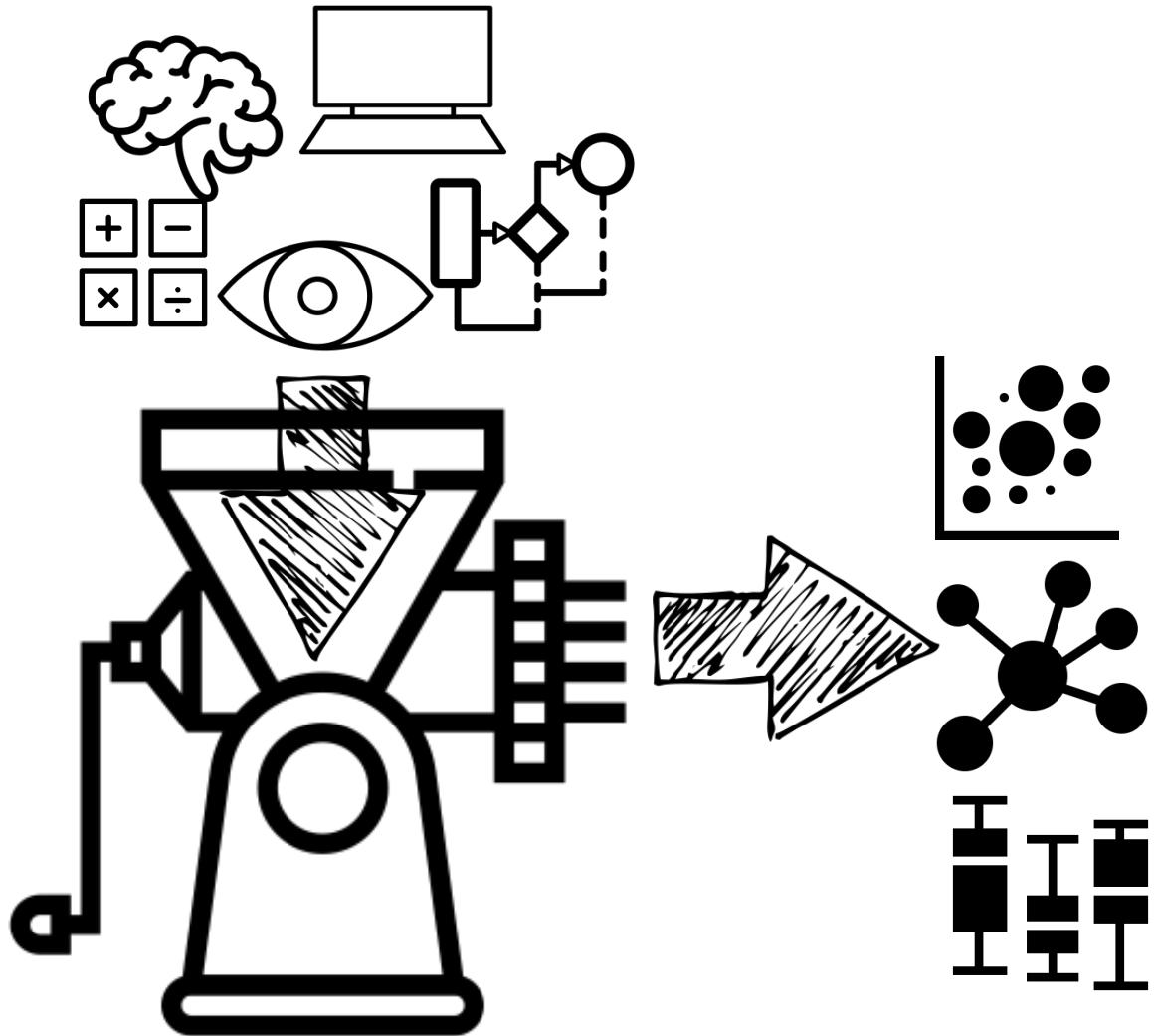


Visualization Design & Analysis

Encoding and Decoding Information



Putting it all Together for Visualization Design & Analysis



- Non-trivial to condense knowledge across all these areas
- Still an ongoing area of research
- **I will try convey a simpler intuition about design & analysis**

Guiding Principles for Visualizing your Data



Image Source: Valentin Antonucci via Pexels

Breaking Down a Visualization in Three Questions

Why? (Motivation)

Why do you need to visualize data?

How will you, or others, use the visualization?

Breaking Down a Visualization in Three Questions

Why? (Motivation)

Why do you need to visualize data?

How will you, or others, use the visualization?

What? (Data & Tasks)

What kind of data is being visualized?

What tasks are performed with the data?

Breaking Down a Visualization in Three Questions

Why? (Motivation)

Why do you need to visualize data?

How will you, or others, use the visualization?

What? (Data & Tasks)

What kind of data is being visualized?

What tasks are performed with the data?

How? (Visual & Interactive Design)

How do you make the visualization?

Is it the right visualization?



People tend to jump
to this level and
ignore why and what

Design & Evaluation with Three Questions

Design

Why?

What?

How?

Evaluation

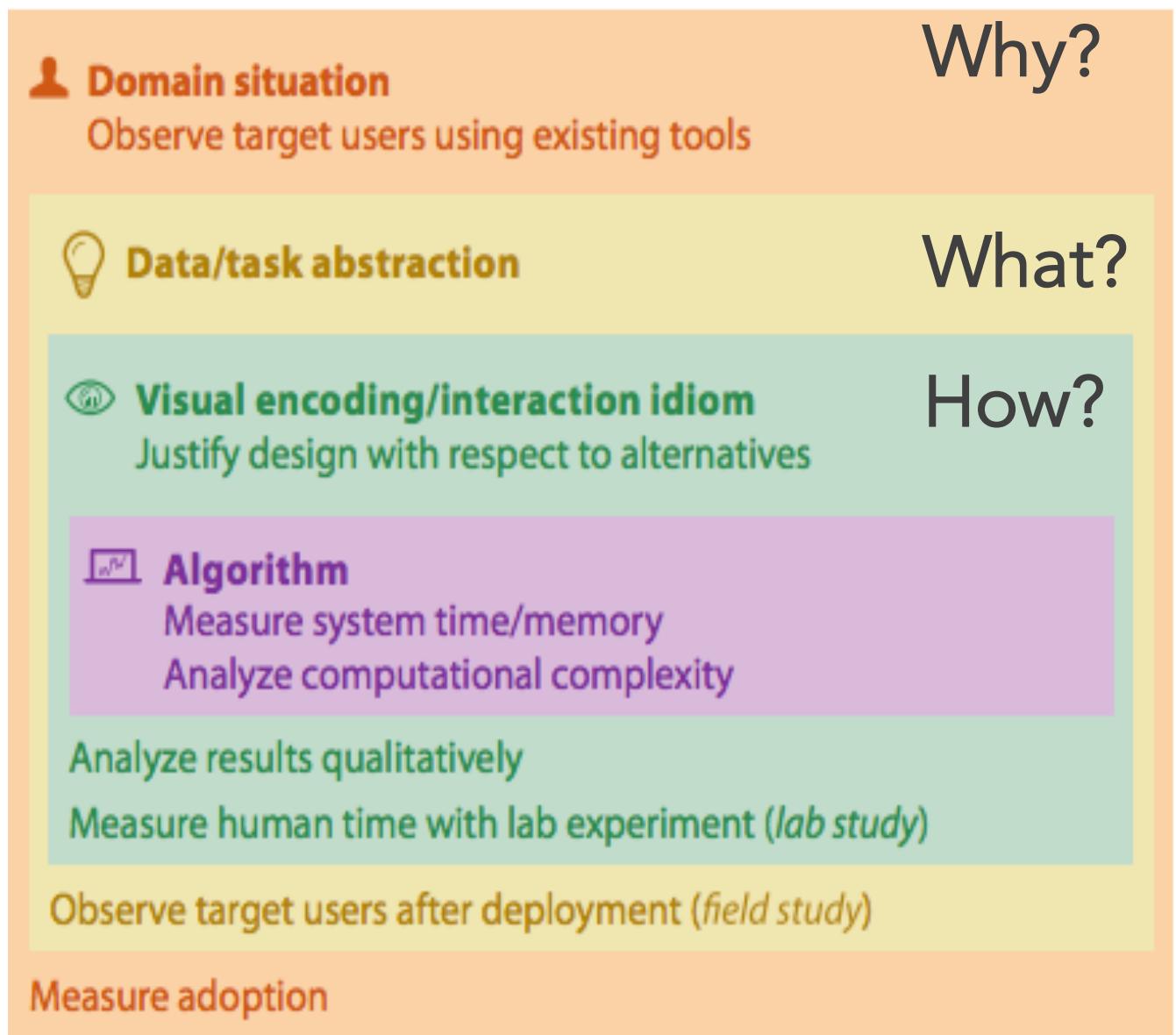
Does the visualization address the the intended need?

Are you using the right data, or *deriving* the right data?
Does the visualization support the tasks using that data?

Are the visual & interactive choices appropriate for the data and tasks?

If interactive / computer based, is the visualization easy to use and reliable (i.e doesn't crash all the time)

Ideas from the research literature : the nested-model



Design

Evaluation

Steps to Systematic Thinking in Data Visualization

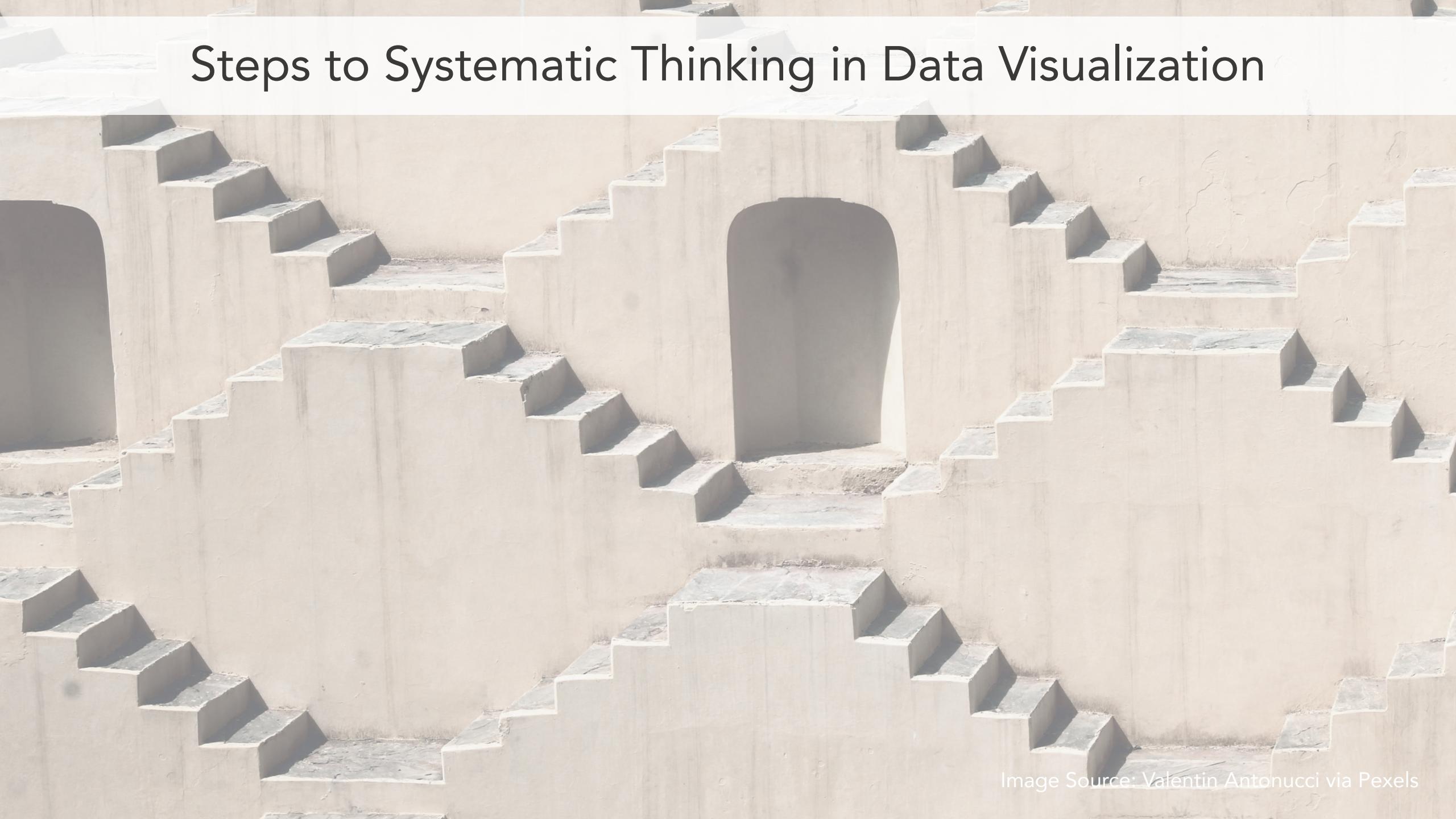
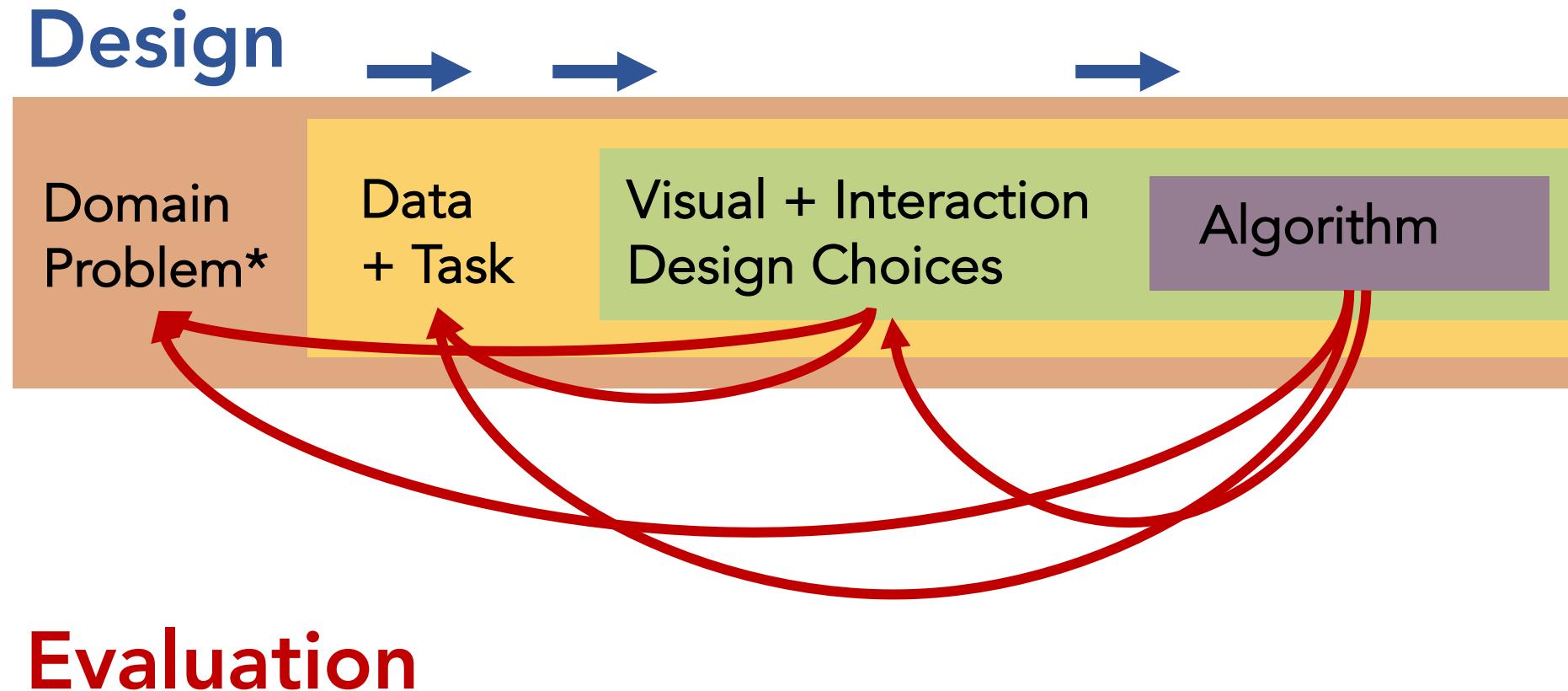


Image Source: Valentin Antonucci via Pexels

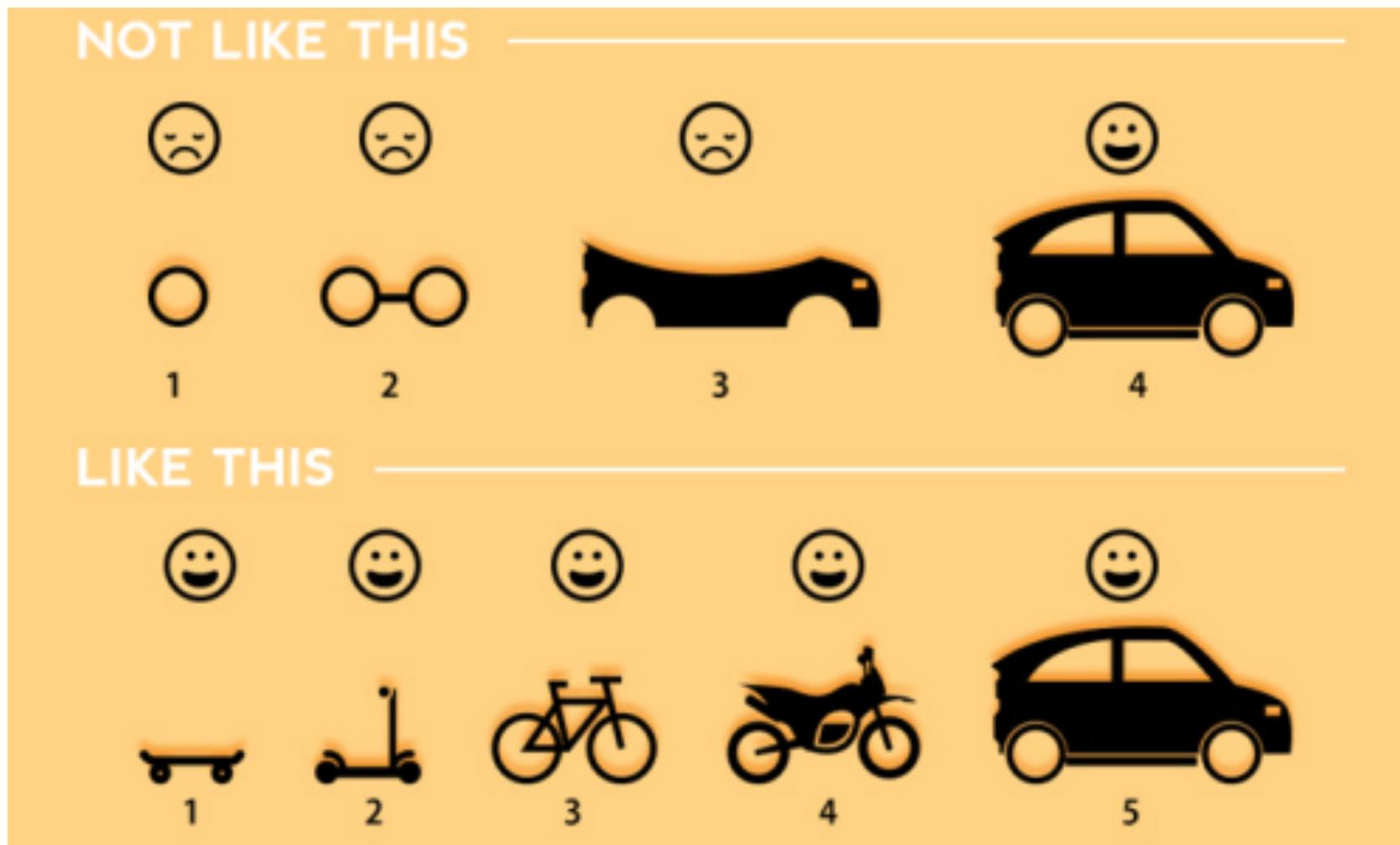
Thinking Systematically about Data Visualization

Infovis (Information Visualization) research advocates an **iterative** process



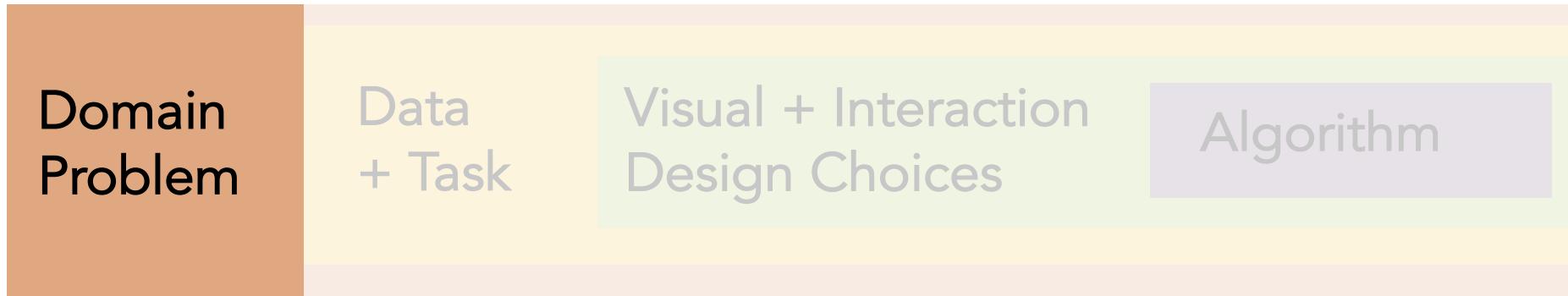
*Domain Problem = Motivation

An Iterative Process



An iterative approach to development allows us to get feedback before committing to ineffective design choices

Thinking Systematically about Data Visualization



1. Identify a relevant **problem** that effects you or a group of stakeholders

Public Health Stakeholders

- Multidisciplinary decision making teams

- More data & diverse data types = more informed decision making
- BUT – different stakeholder abilities to interpret data & different needs

Medical
Health
Officers



Clinicians



Nurses



Researchers



Community
Leaders



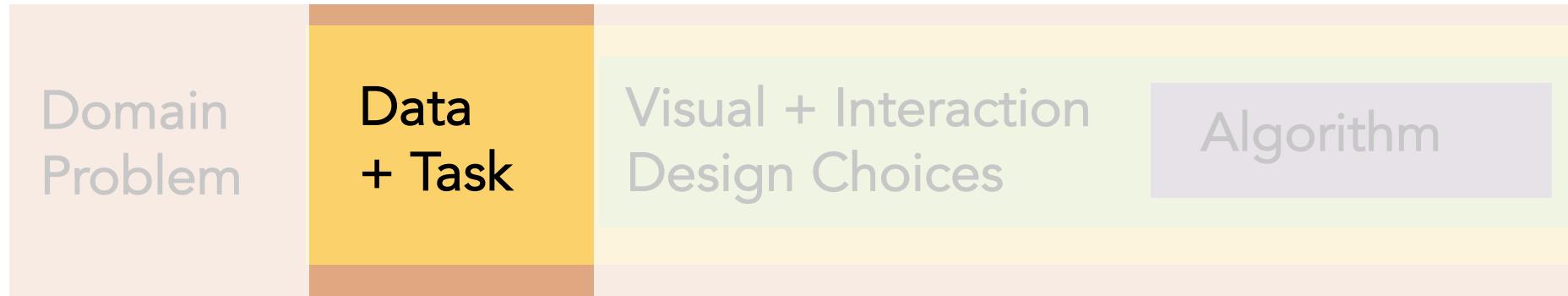
Patients



Policy Makers



Thinking Systematically about Data Visualization

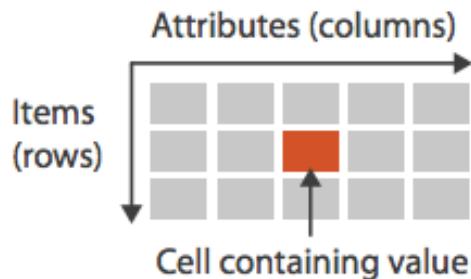


2. Ask **what data** stakeholders use (is it available)?
3. Ask **what stakeholders do** with the data [**tasks**]

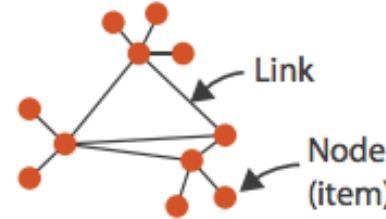
Data - Many Different Types of Data!

→ Dataset Types

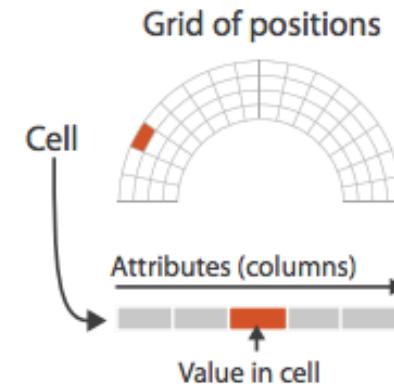
→ Tables



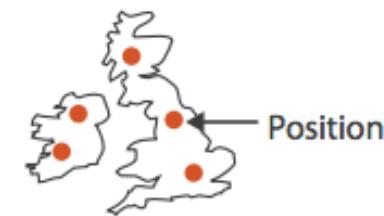
→ Networks



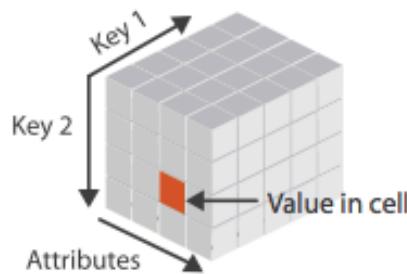
→ Fields (Continuous)



→ Geometry (Spatial)



→ Multidimensional Table



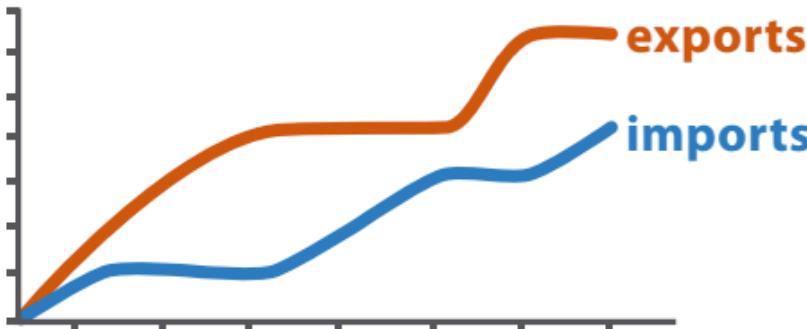
→ Trees



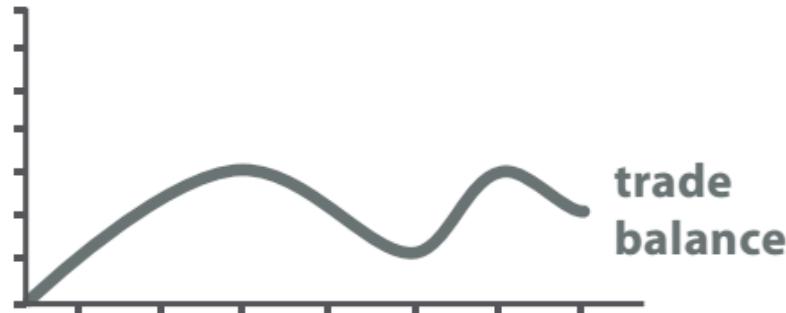
Data - Don't Just Visualize the Raw Data!

Example

Original (Raw) Data



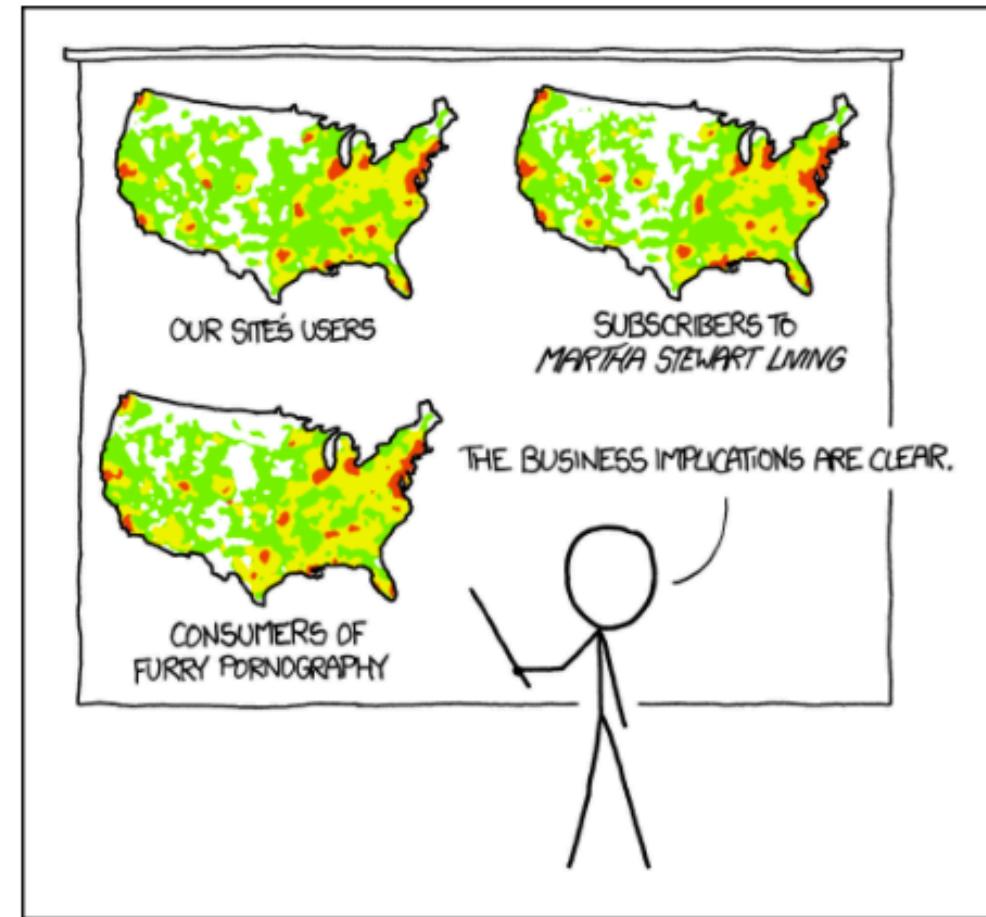
Derived Data



$$\text{trade balance} = \text{exports} - \text{imports}$$

T. Munzner (2014) – Visualization Design and Analysis

Example when this advice is ignored



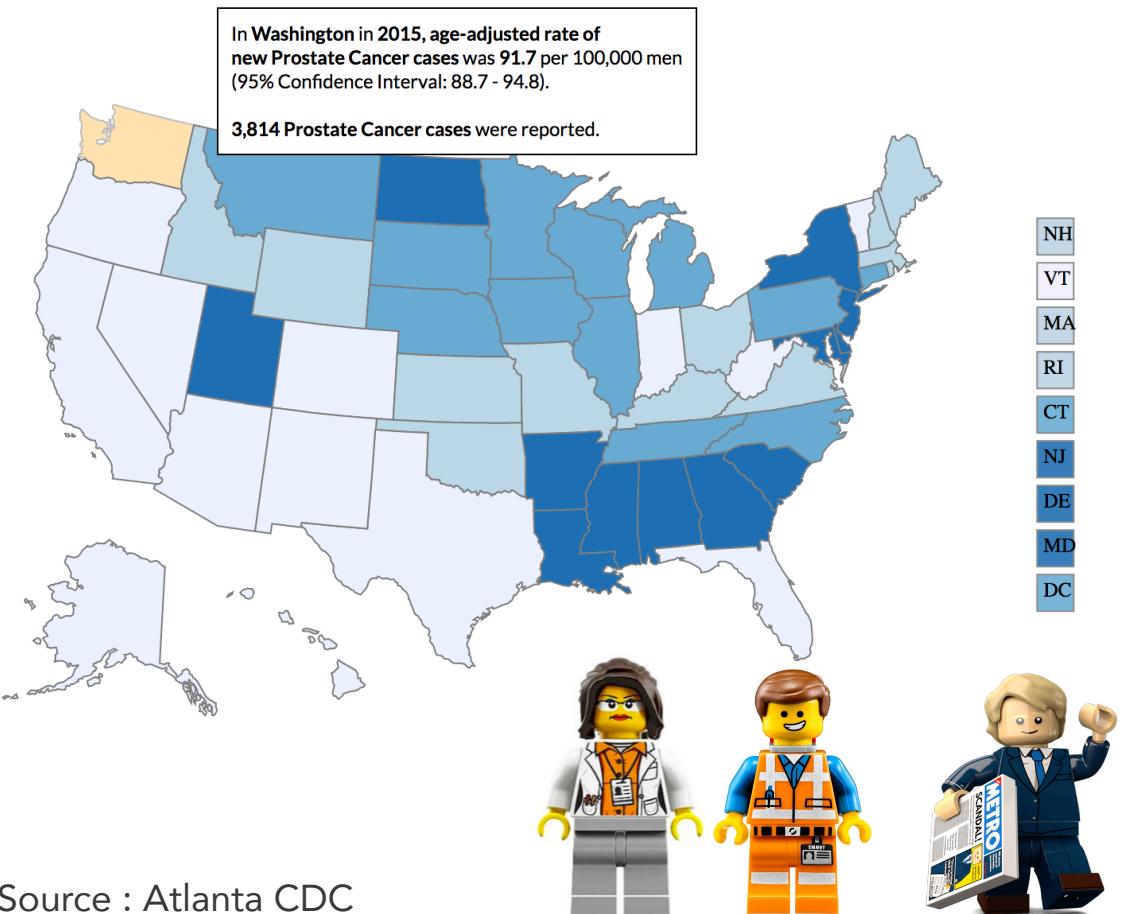
PET PEEVE #208:
GEOGRAPHIC PROFILE MAPS WHICH ARE
BASICALLY JUST POPULATION MAPS

XKCD

Tasks - How People Use the Data

Geographic Overview of Prostate Cancer

- Useful for epidemiologists and policy makers
- Supports surveillance tasks



Individual Prostate Cancer Risk

- Good for patients and doctors
- Supports treatment decision making tasks

Risk of prostate cancer if biopsy were to be performed

Based on the provided risk factors a prostate biopsy performed would have a:

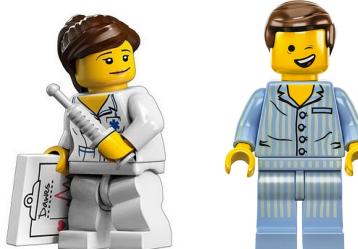
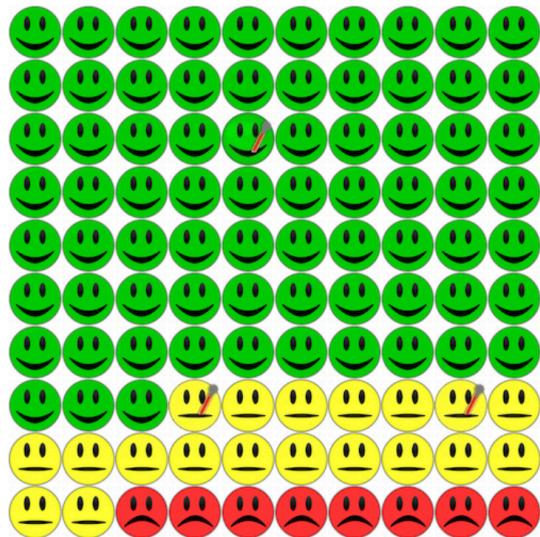
8% chance of high-grade prostate cancer,

19% chance of low-grade cancer,

73% chance that the biopsy is negative for cancer.

About 2 to 4% of men undergoing biopsy will have an infection that may require hospitalization.

Please consult your physician concerning these results.

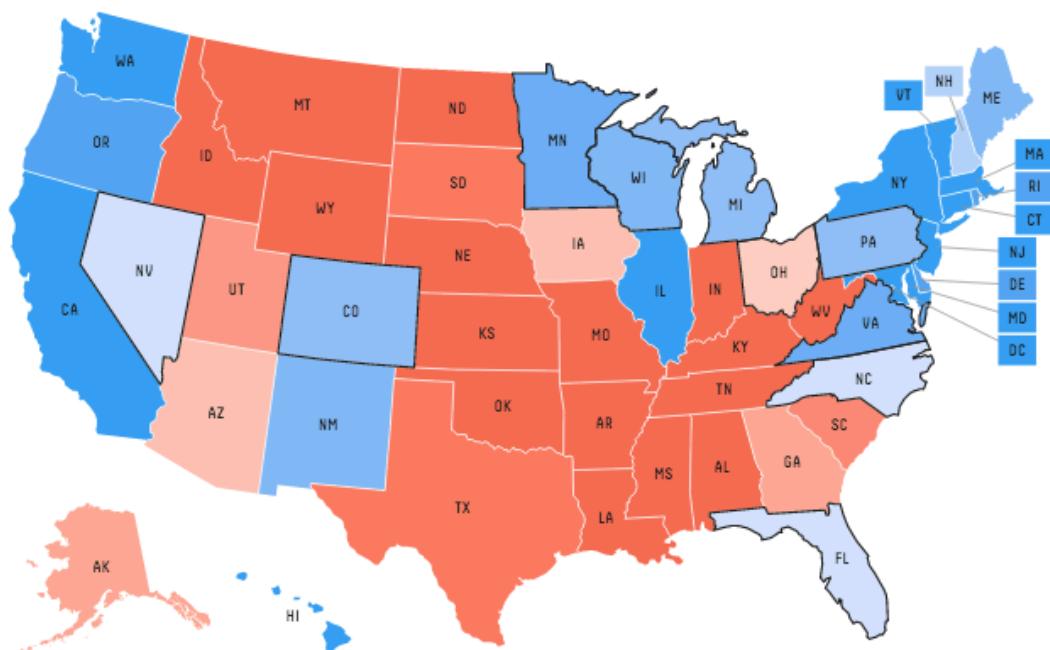


Source : <http://riskcalc.org/PCPTRC/>
(UT San Antonio)

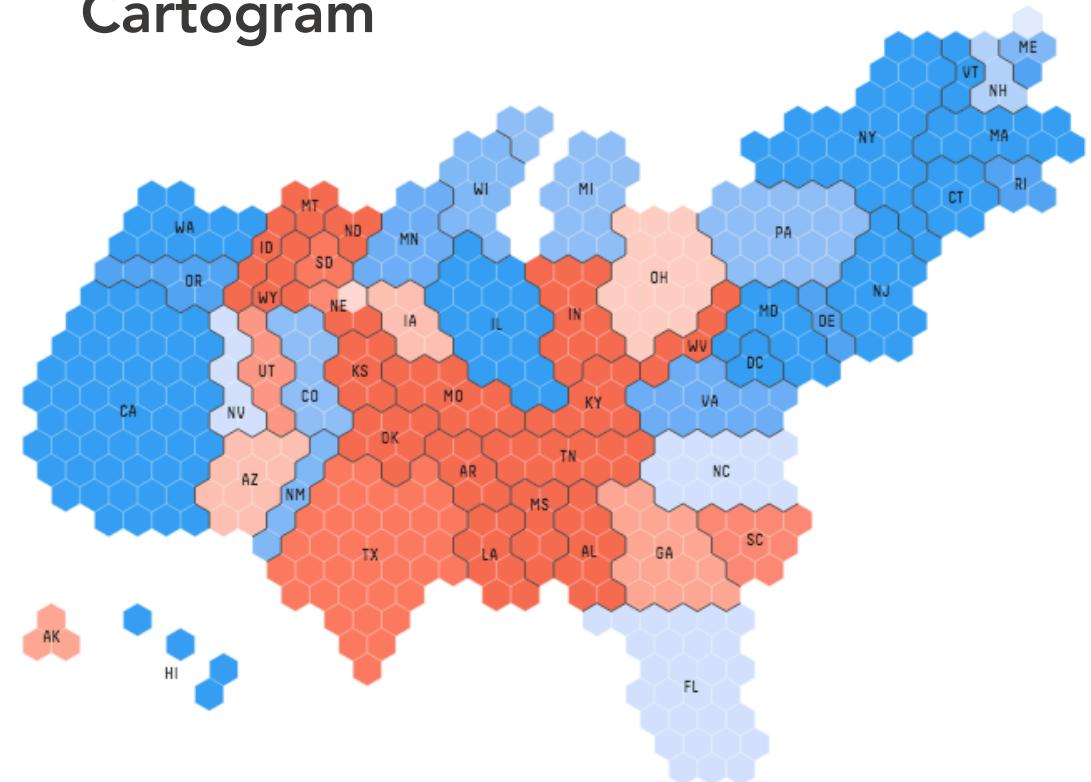
Tasks - How People Use the Data

- Tasks can also change how the same data should be visualized
 - Example: representing US electoral collage results

Standard Map



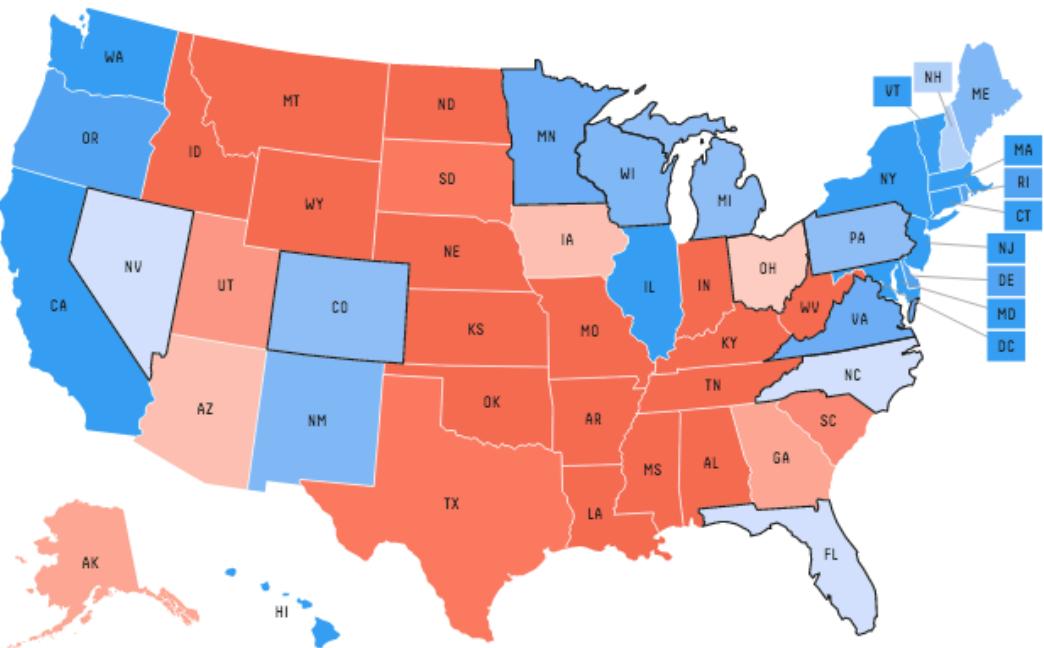
Cartogram



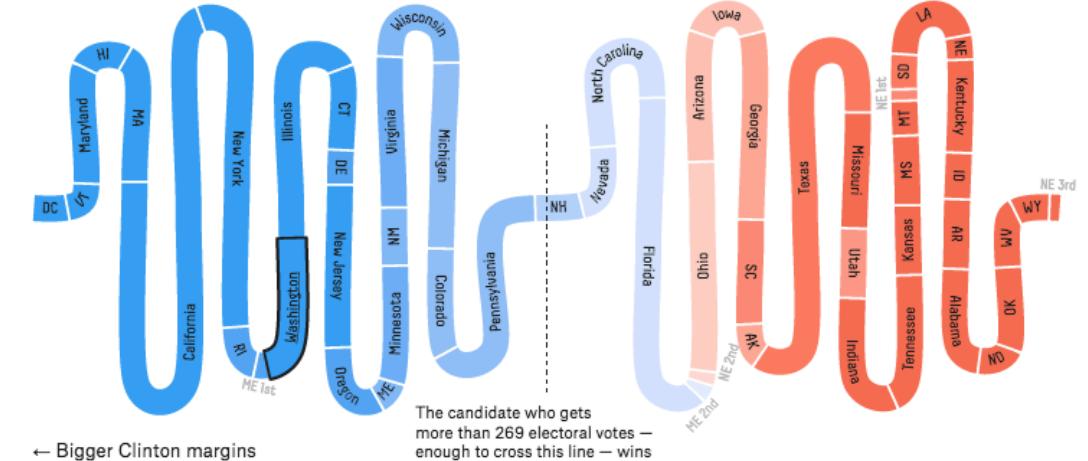
Tasks - How People Use the Data

- Tasks can also change how the same data should be visualized
- Example: representing US electoral college results

Standard Map

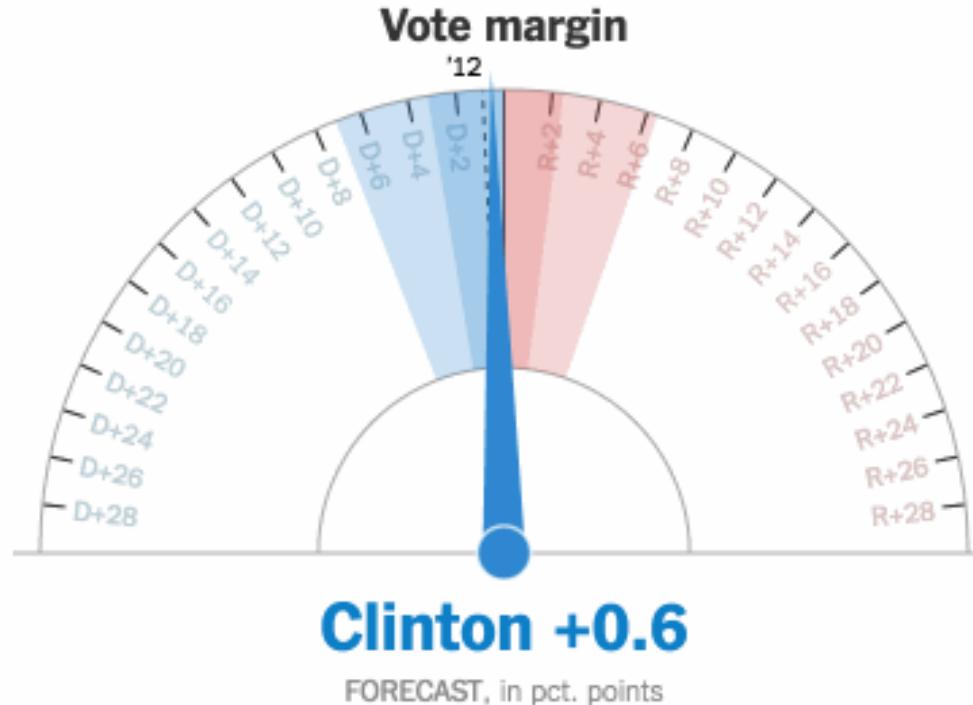


Snakey Diagram



Tasks - How People Use the Data

- Tasks can also change how the same data should be visualized
- Example: representing US electoral college results

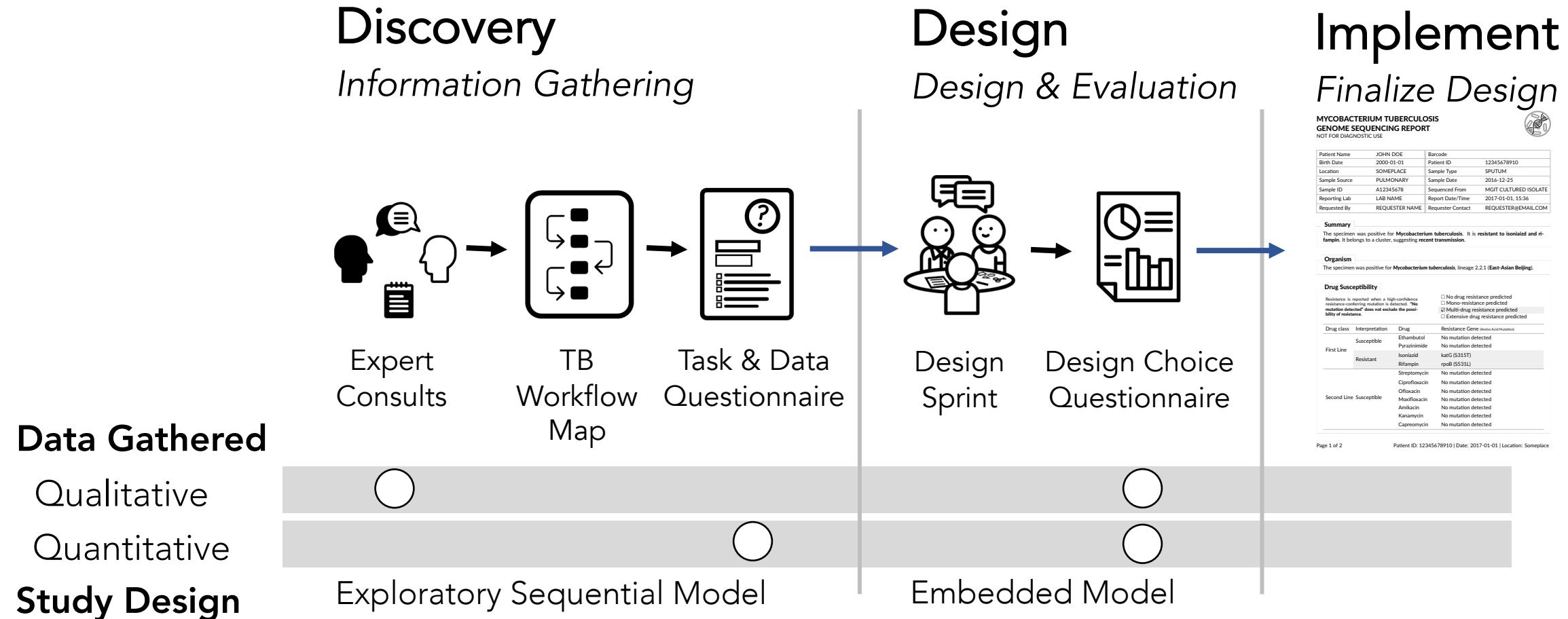


Examples from
my own research

How can we identify
tasks and data?

My research : making a clinical report for tuberculosis

- Mixed methods approach to gathering data and tasks



My research : making a clinical report for tuberculosis

Data

	WGS equivalent	DIAGNOSIS TASKS				TREATMENT TASKS			SURVEILLANCE TASKS					TOTAL SCORE
		Diagnose Latent TB	Diagnose Active TB	Reactive vs New Infection	Characterize Transmission Risk	Choose Meds	Choose Tx Duration	Assess Response to Tx	Guide Contact Tracing	Report to Public Health	Define a Cluster	Connect Case to Existing Cluster	Guide Public Health Response	
Patient Identifier	Same	3	3	3	3	3	3	3	2	1	1	1	1	26
Sample Collection Date	Same	3	3	2	3	3	3	3	1	1	1	1	1	24
Patient Prior TB Results	Same	3	2	3	3	3	3	3	1	1	1	0	1	23
Speciation	Speciation	1	3	2	3	3	3	3	2	1	1	1	1	23
Sample Type (sputum, fine needle aspirate etc.)	Same	2	3	2	3	3	3	3	1	1	1	0	1	22
Culture results	NA	1	3	2	3	3	3	3	2	1	1	0	1	22
Sample Collection Site (lymph node, lung etc..)	Same	2	3	2	3	3	3	3	1	1	0	0	1	21
Acid Fast Bacilli Smear	Speciation	2	3	2	3	2	3	3	1	1	1	0	1	21
Resistotype	Predicted DST	0	2	3	1	3	3	2	2	1	1	1	1	19
Phenotypic DST	Predicted DST	0	2	3	2	3	3	2	1	1	1	0	1	18
Chest x-ray	NA	3	3	2	3	0	2	3	1	0	0	0	0	17
Report Release Date	Same	2	2	1	2	2	2	2	1	0	1	0	1	15
Requester IDs	Same	2	2	2	2	2	2	2	1	0	0	0	0	15
Interpretation or comments from reviewer	Same	2	2	1	2	2	2	3	1	0	0	0	0	15
Predicted DST	Predicted DST	0	2	2	1	3	3	2	1	0	1	0	0	15
MIRU-VNTR	SNPs	0	2	3	1	1	1	1	1	1	1	1	1	13
Cluster Assignment	Same	0	2	2	1	1	1	0	1	1	1	1	1	11
SNP/variant distance	SNPs	0	1	2	1	1	1	0	1	1	1	1	1	10
Phylogenetic Tree	Same	0	2	1	1	1	1	0	1	0	1	1	1	9
Reviewer ID	Same	1	1	1	1	1	1	1	1	0	0	0	0	8
TST results	Speciation*	3	1	1	1	0	0	0	1	0	0	0	0	7
IGRA results	Speciation*	3	1	1	1	0	0	0	1	0	0	0	0	7
Lab QC	WGS Specific	0	1	2	1	1	1	0	1	0	0	0	0	7
Spoligotype	SNPs	0	1	1	1	0	0	0	0	0	0	0	0	3
RFLP	SNPs	0	1	1	1	0	0	0	0	0	0	0	0	3

Consensus among participants

cat.	% agree
3	(>75%)
2	(50% - 25%)
1	(25% - 50%)
0	(<25%)

My research : making a clinical report for tuberculosis

MYCOBACTERIUM TUBERCULOSIS GENOME SEQUENCING REPORT NOT FOR DIAGNOSTIC USE



Patient Name	JOHN DOE	Barcode
Birth Date	2000-01-01	Patient ID
Location	SOMEPLACE	Sample Type
Sample Source	PULMONARY	Sample Date
Sample ID	A12345678	Sequenced From
Reporting Lab	LAB NAME	Report Date/Time
Requested By	REQUESTER NAME	Requester Contact

Summary

The specimen was positive for *Mycobacterium tuberculosis*. It is resistant to isoniazid and rifampin. It belongs to a cluster, suggesting recent transmission.

Organism

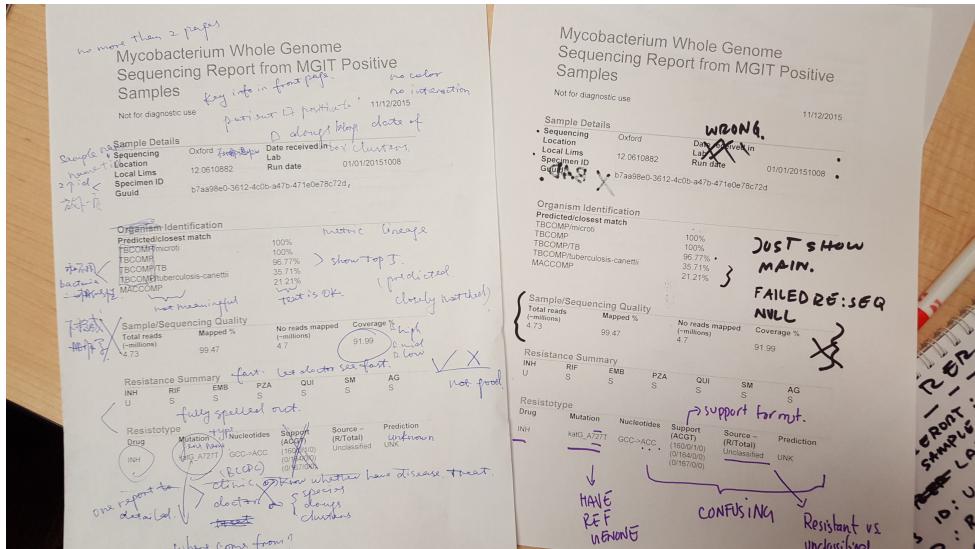
The specimen was positive for *Mycobacterium tuberculosis*, lineage 2.2.1 (East-Asian Beijing).

Drug Susceptibility

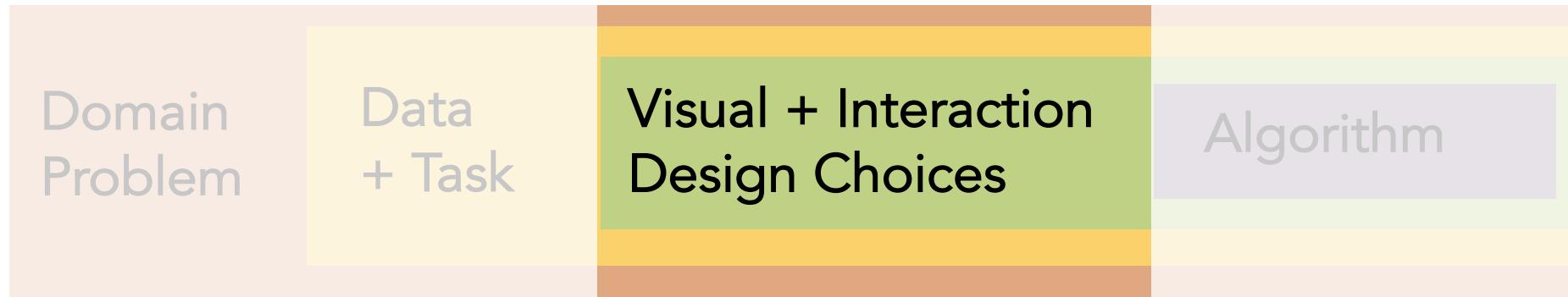
Resistance is reported when a high-confidence resistance-confering mutation is detected. "No mutation detected" does not exclude the possibility of resistance.

- No drug resistance predicted
- Mono-resistance predicted
- Multi-drug resistance predicted
- Extensive drug resistance predicted

Drug class	Interpretation	Drug	Resistance Gene (Amino Acid Mutation)
First Line	Susceptible	Ethambutol	No mutation detected
		Pyrazinamide	No mutation detected
Second Line	Resistant	Isoniazid	katG (S315T)
		Rifampin	rpoB (S531L)
	Susceptible	Streptomycin	No mutation detected
		Ciprofloxacin	No mutation detected
		Ofloxacin	No mutation detected
		Moxifloxacin	No mutation detected
		Amikacin	No mutation detected
		Kanamycin	No mutation detected
		Capreomycin	No mutation detected



Thinking Systematically about Data Visualization



4. Explore if other visualizations have addressed this **problem** and set of **tasks & data**
5. Implement **your own solution** (**remember this include interaction!**)

Marks & Channels : Basic Building Blocks

Mark:

Basic Graphical Element
(basic building block)



④ Points



④ Lines

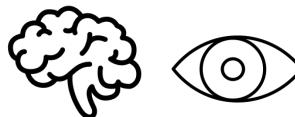


④ Areas

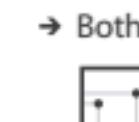
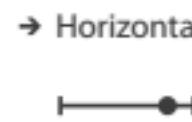


Channel:

Controls the
appearance of marks



④ Position



④ Color



④ Shape



④ Tilt



④ Size

→ Length



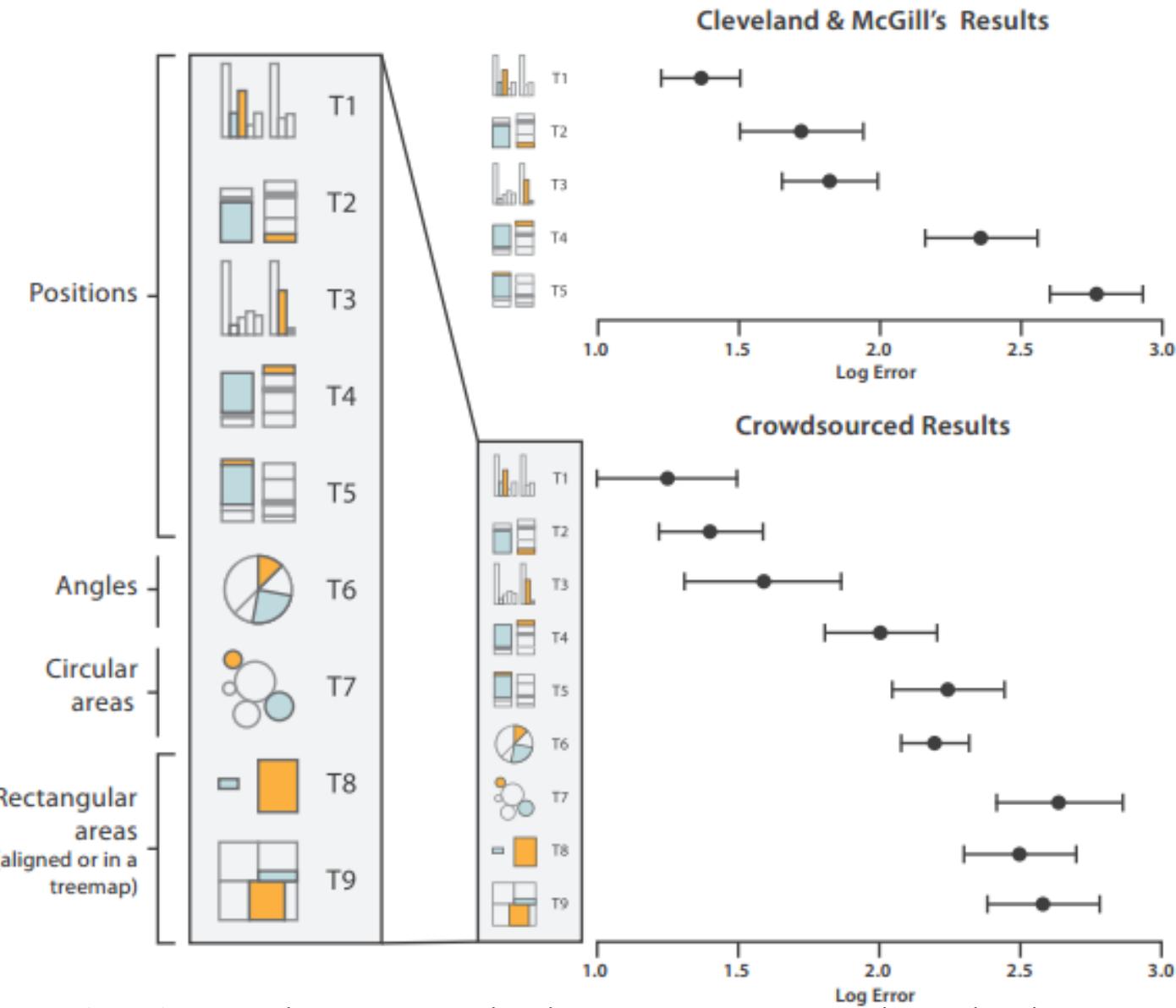
→ Area



→ Volume



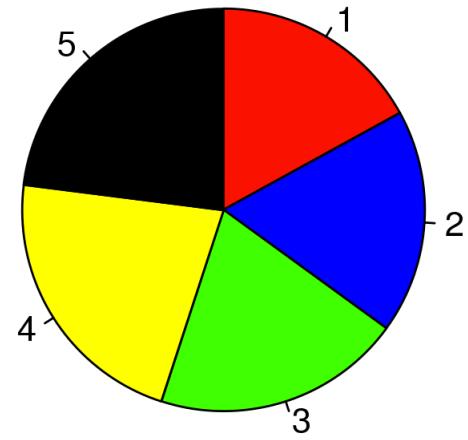
Marks Vary in their Effectiveness



Example

Pie Chart

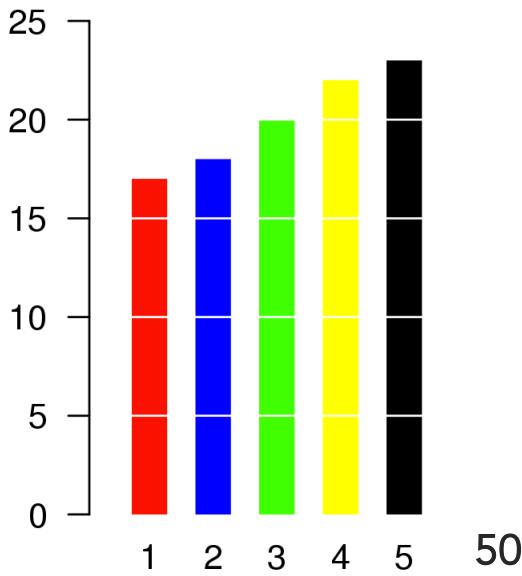
Angle & Area



Bar Chart

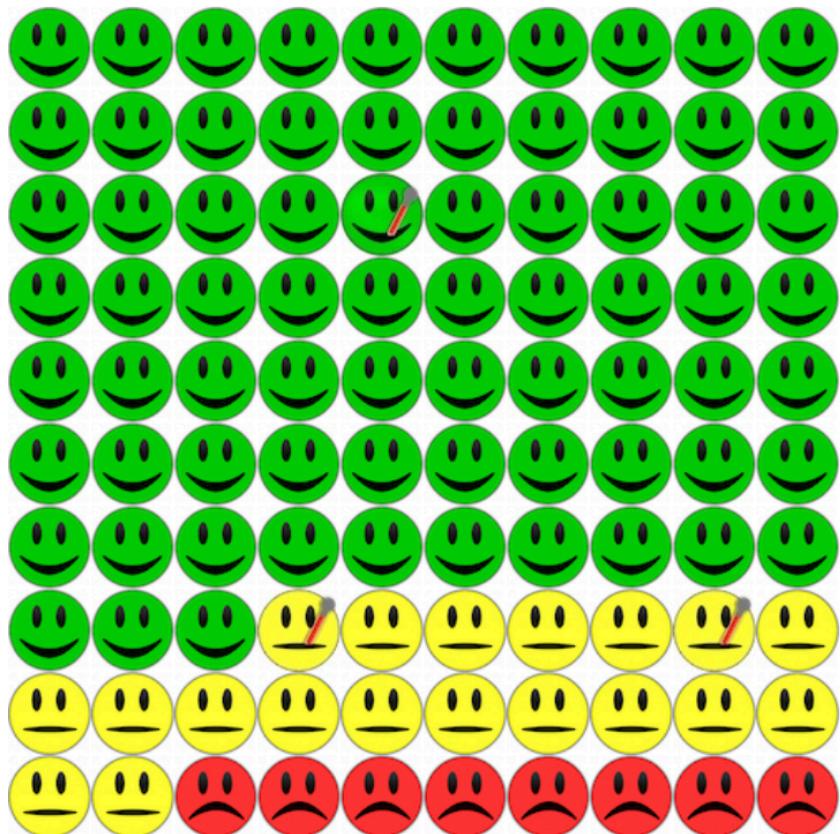
Position

Common Scale



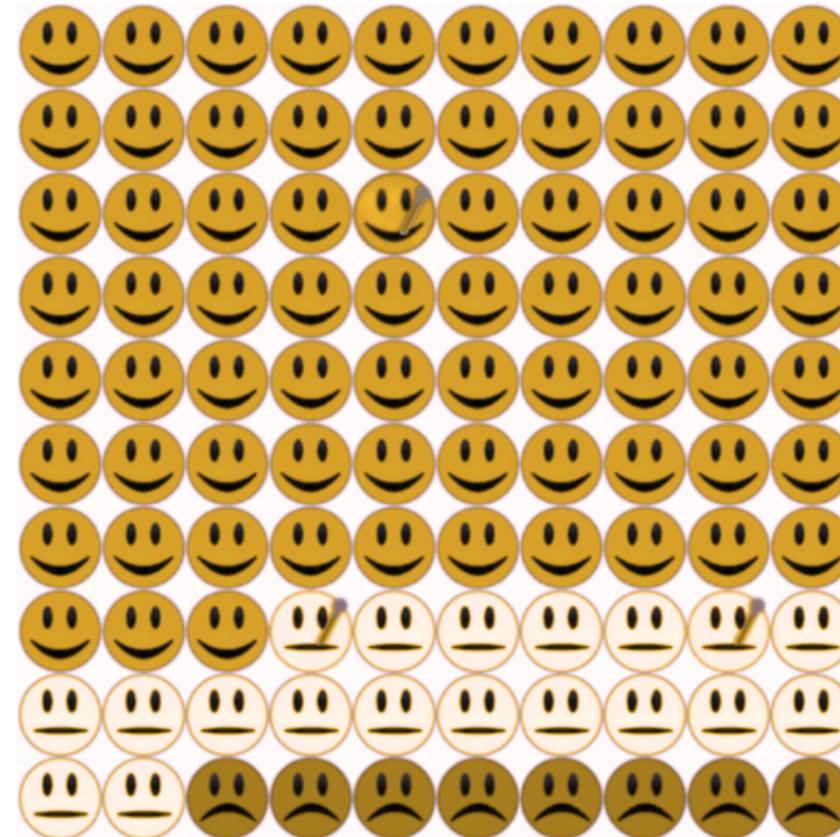
Perception and Cognition Matter Too!

Original Visualization



Visualization as seen by color blind person

(color blindness (deutanopia) impacts men more often)



Perception and Cognition Here too!

Colour scales also impact interpretation!

Perceptual research from Liu et al (2018)

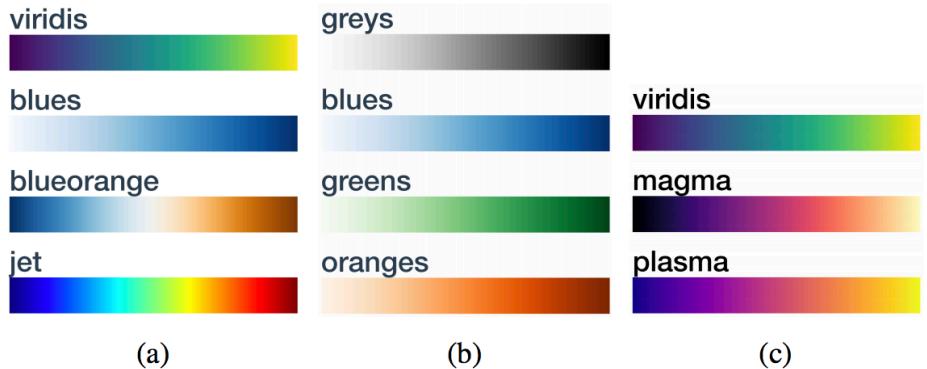


Figure 1: **Colormaps under study.** We evaluate four single-hue, three perceptually-uniform multi-hue, a diverging, and a rainbow colormap(s). We divide them into (a) assorted, (b) single-hue and (c) multi-hue groups, with two colormaps repeated across groups for replication.

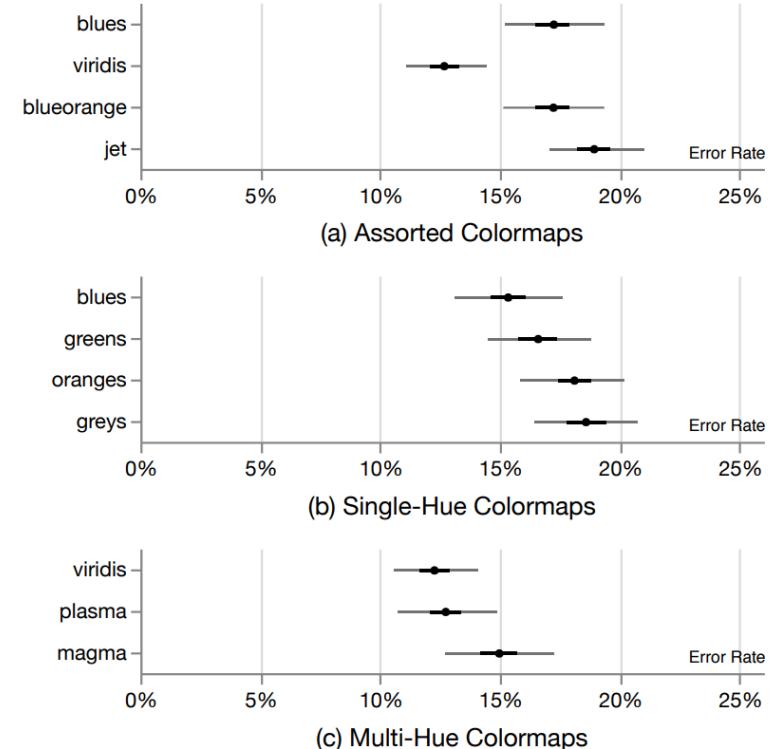
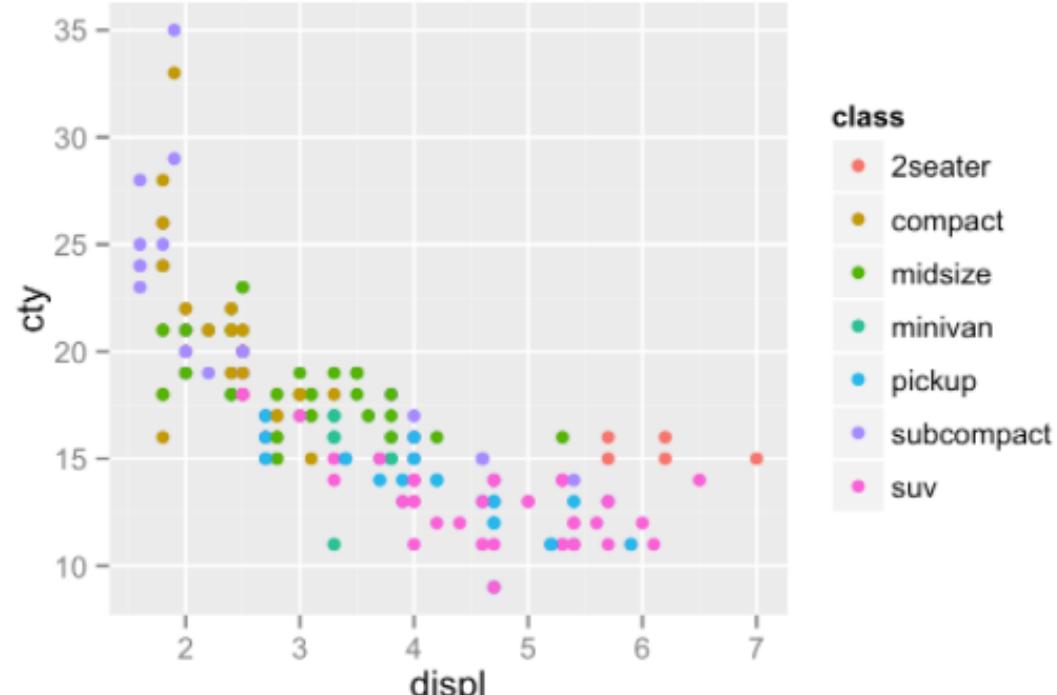


Figure 4: **Error rate by colormap for each study.** Plots depict bootstrapped means, with 50% (thick) and 95% (thin) CIs. (a) Assorted colormaps. *Viridis* excels in accuracy while *jet* is the most error-prone. (b) Single-hue colormaps. Though slightly faster, *blues* and *greens* have overlapping confidence intervals with the slower colormaps, *oranges* and *greys*. (c) Multi-hue colormaps. Multi-hue colormaps have comparable accuracy within group. The per-colormap average error rate of *magma* is higher as it contains degenerate cases.

Marks & Channels : ggplot2 example

```
ggplot (data = mpg, aes( x= displ, y = cty, colour = class)) +  
  geom_point()
```

Mark: Point



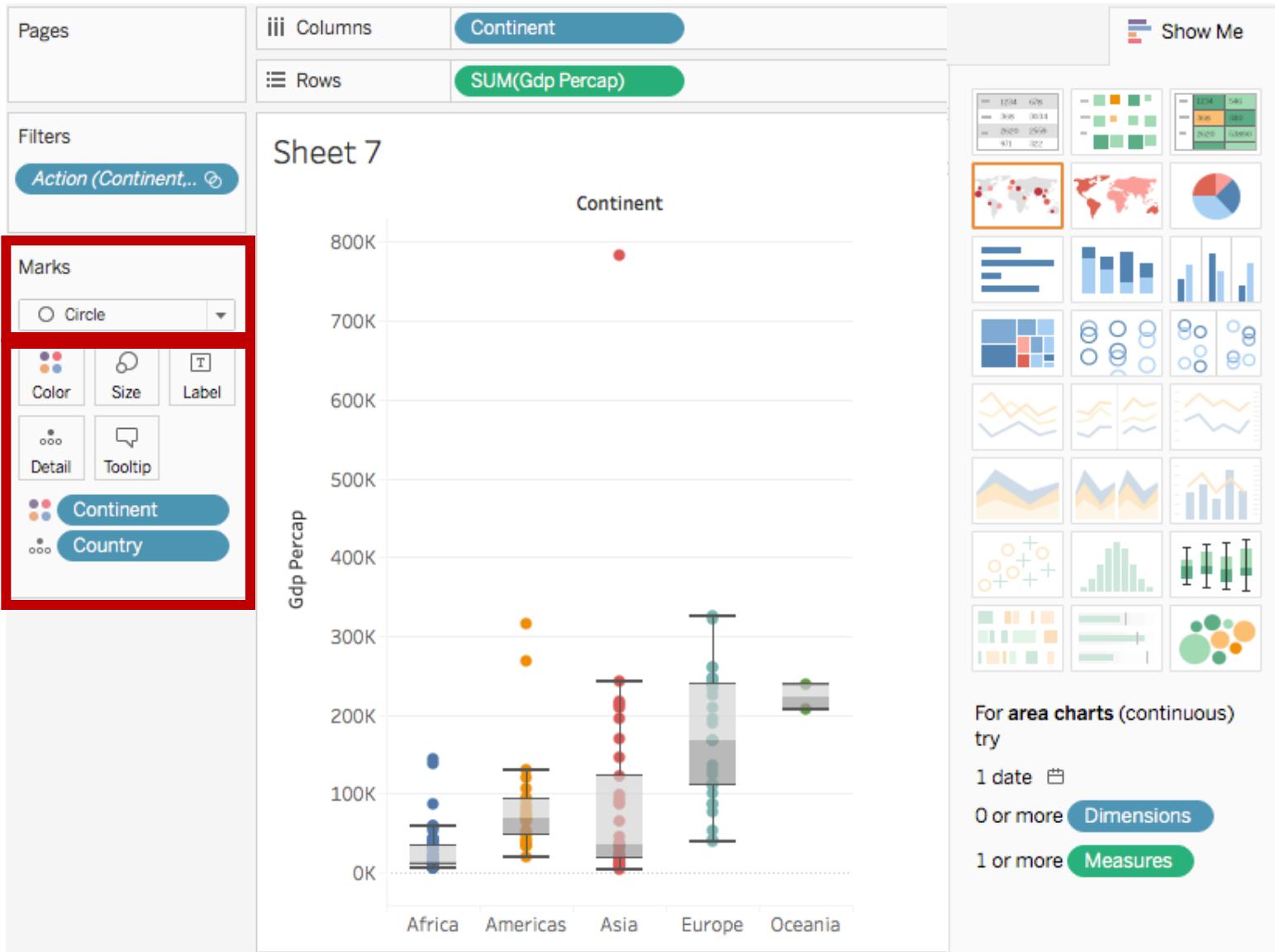
Note: Generally in ggplot2 aesthetics refer to channels and geoms refer to marks, but there are complex geoms that aren't simple marks but chart types (i.e. geom_density) and there are aesthetics that have little to do with the visual channels directly (i.e. group)

Marks & Channels : Tableau example

Marks



Channels



For area charts (continuous)
try

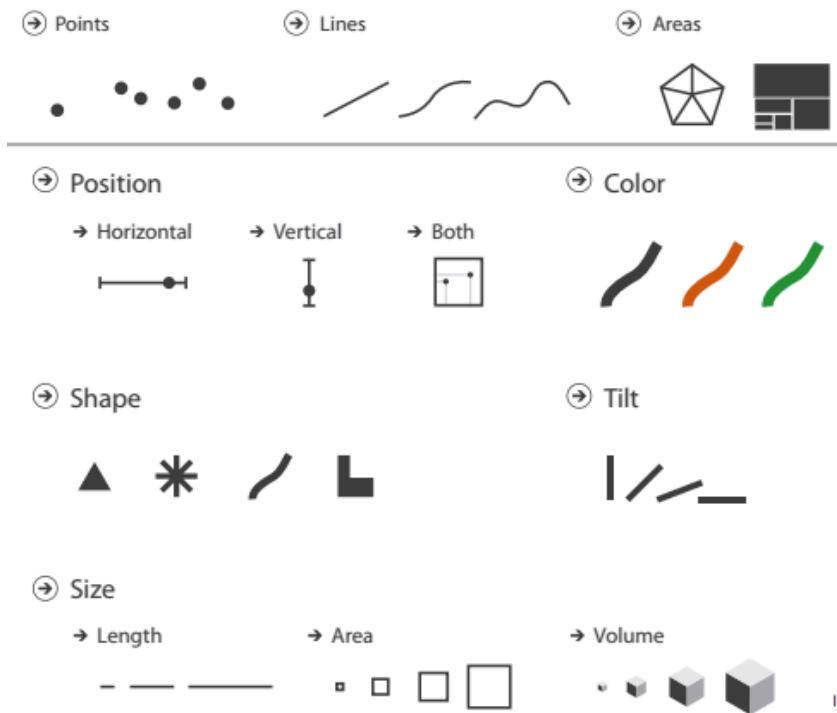
1 date

0 or more Dimensions

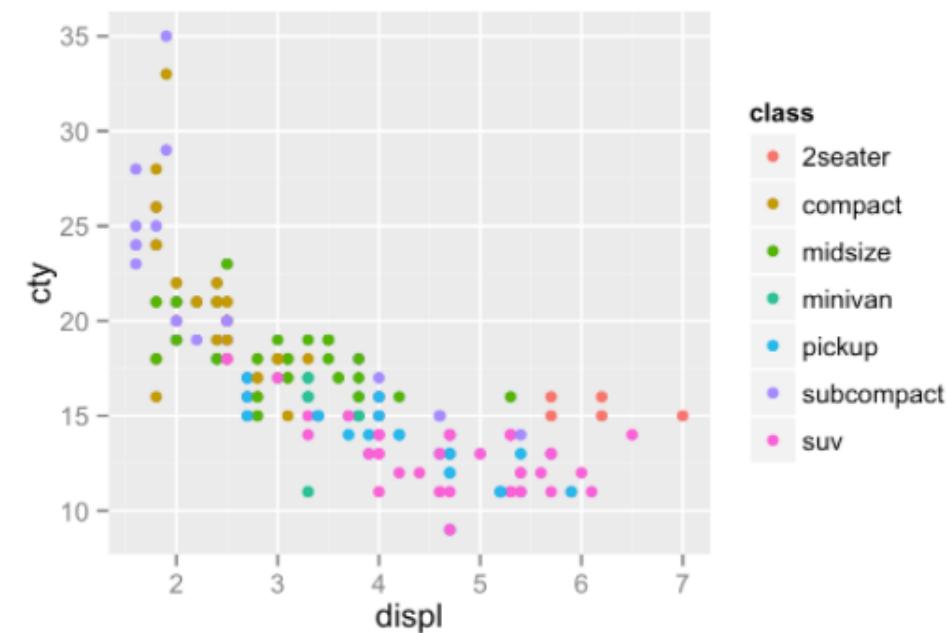
1 or more Measures

Linking Data to Mark and Channels to Make Visualizations

Data → Marks & Channels → Visualization



19



Linking Data to Mark and Channels to Make Visualizations

Data to viz

<https://www.data-to-viz.com/>

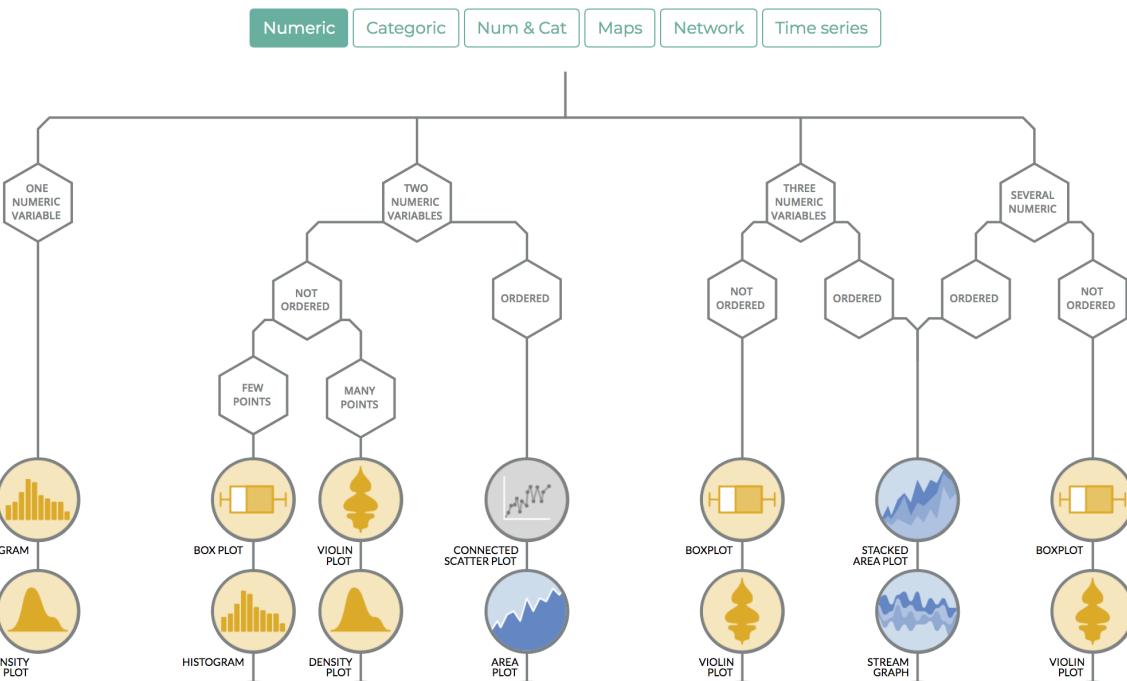


Chart Chooser

<https://bit.ly/2P9zLEW>

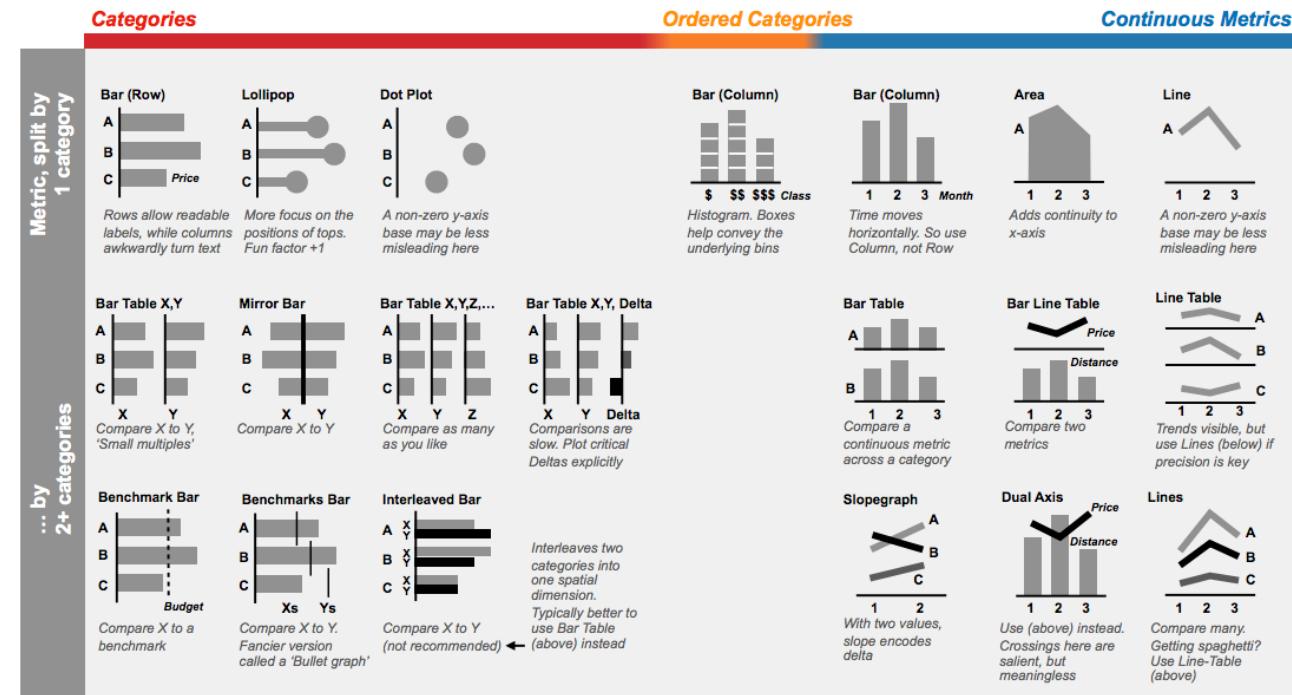
Which Visualization?

A Quick Reference

You have the following data table:
Categories in the first two columns,
 a single **ordered category**,
 and a few **continuous metrics**.

Categories	Airline	Ordered Cats	Continuous Metrics	FlightTime	Price
City	XeroTrip	\$	Month	300	120
	YoloFly	\$\$		500	185
	ZeusAir	\$\$\$	Distance	650	240
...

Here's how to plot them.



Tradeoffs between different medical visualizations

Detailed experiments on efficacy of different vis designs

<http://www.vizhealth.org/gallery/>

VISUALIZING
HEALTH

About The Wizard Browse the Gallery Using Visualizing Health Disclaimer

My goal

- Classifying risks
- Raise or lower concern
- Awareness of risk
- Differences in likelihood
- Risk tradeoffs

Details or gist?

- Verbatim recall
- Gist understanding

Data I have

- Benefit estimate
- Risk over time
- Case counts
- Population risks
- Risk disparity
- Risk estimate

(78) The benefits of risk reduction

...MORE LIKE THIS

Differences in likelihood Raise or lower concern

ALL TAGS

(82) Icons to show severity of side effects

...MORE LIKE THIS

Classifying risks

ALL TAGS

(85) Icons to show

RISK OF DEVELOPING TYPE 2 DIABETES WITHIN THE NEXT 8 YEARS

Current risk Absolute risk reduction (with 25lb. weight loss)

25%

10%

(59) The benefits of risk reduction

...MORE LIKE THIS

Differences in likelihood ★ Raise or lower concern

ALL TAGS

COLON CANCER RISK %

RISK PERCENTAGE

1.8% ← 0.7% RISK REDUCTION

2.5%

WITHOUT SCREENING TESTS WITH REGULAR SCREENING TESTS RISK DIFFERENCE

patients who use a cancer drug

life 6 average of months patients live longer

2.5 years that 20% of patients live longer

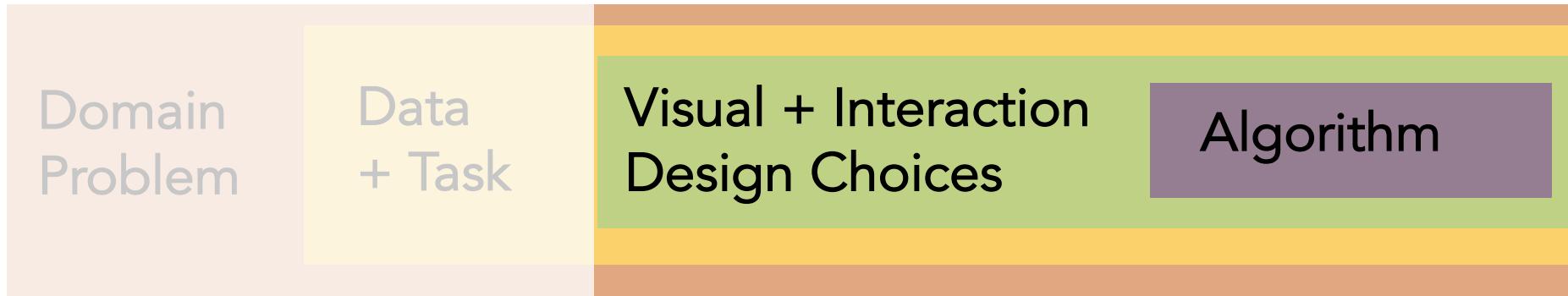
80% patients who don't benefit

10%

The screenshot displays a dark-themed website for 'Visualizing Health'. On the left, a sidebar lists user goals (e.g., 'Classifying risks', 'Raise or lower concern') and data types (e.g., 'Benefit estimate', 'Risk over time'). The main area contains five cards, each illustrating a different visualization concept:

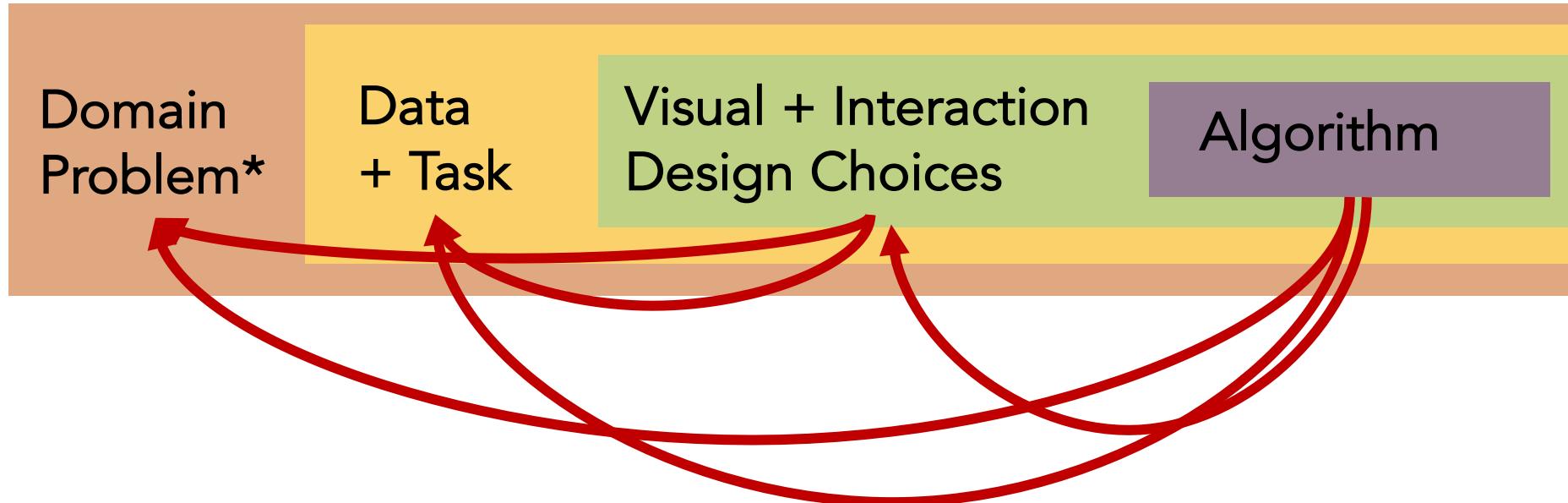
- (78) The benefits of risk reduction:** Shows a grid of colored hexagons representing risk levels (purple for low, pink for moderate, yellow for high) and a bar chart showing a 10% reduction in risk from 25% to 15%.
- (82) Icons to show severity of side effects:** Displays five icons of faces with increasing levels of distress (from neutral to severely distressed).
- (85) Icons to show:** A card with the heading 'RISK OF DEVELOPING TYPE 2 DIABETES WITHIN THE NEXT 8 YEARS' and a grid of colored squares representing risk levels.
- (59) The benefits of risk reduction:** Shows a bar chart comparing colon cancer risk without screening (2.5%) and with screening (1.8%), highlighting a 0.7% risk reduction.
- patients who use a cancer drug:** Illustrates the impact of a cancer drug on patient survival, showing an average increase of 6 months and 2.5 years for 20% of patients.

Thinking Systematically about Data Visualization



4. Explore if other visualizations have addressed this **problem** and set of **tasks**
5. Implement **your own solution** (part or all of that **solution** could be a new algorithm)

Thinking Systematically about Data Visualization

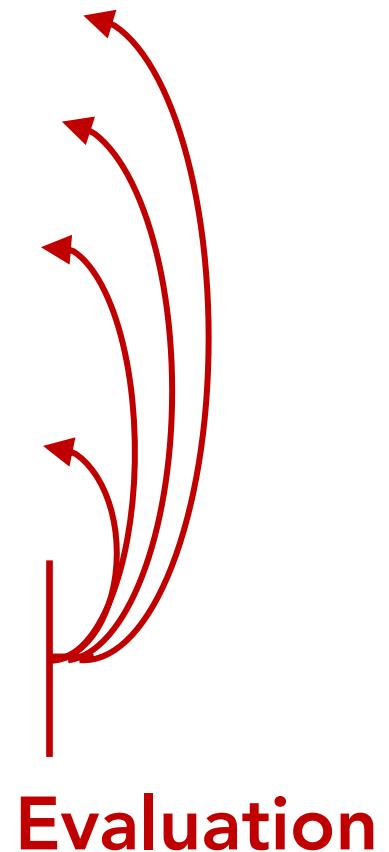


6. Test **multiple alternatives** (including new ones you develop) with stakeholders
7. Gather **qualitative & quantitative** evaluation data

Thinking Systematically about Data Visualization

Design

1. Identify a relevant **problem** that effects you or a group of stakeholders
2. Ask **what data** stakeholders use (is it available)?
3. Ask **what stakeholders do** with the data [**tasks**]
4. Explore if other visualizations have addressed this **problem** and set of **tasks & data**
5. Implement **your own solution** (vis and/or algorithm)
6. Test **multiple alternatives** (including new ones you develop) with stakeholders
7. Gather **qualitative & quantitative** evaluation data



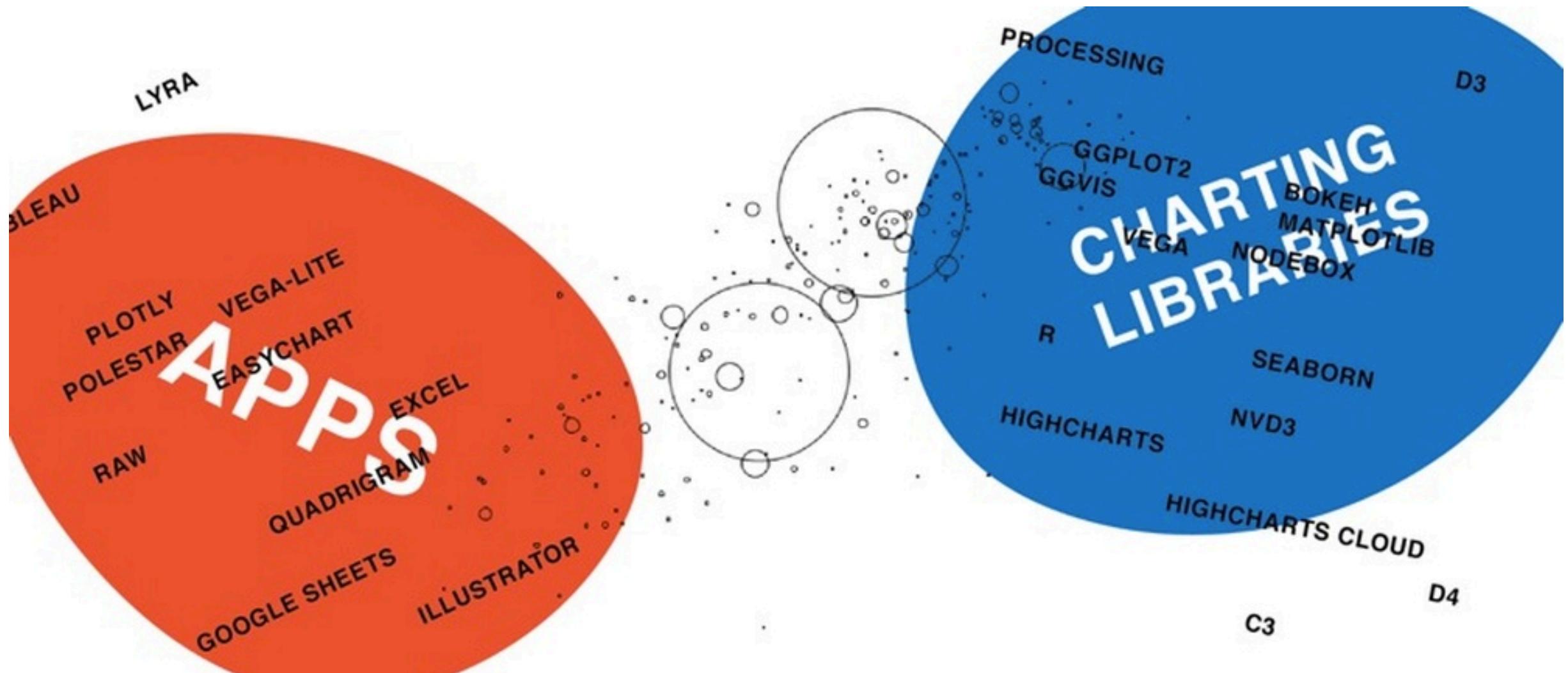
**What datavis
tools are available?**

Data Visualization Tools to Get You Started



Tools & Libraries for data visualization

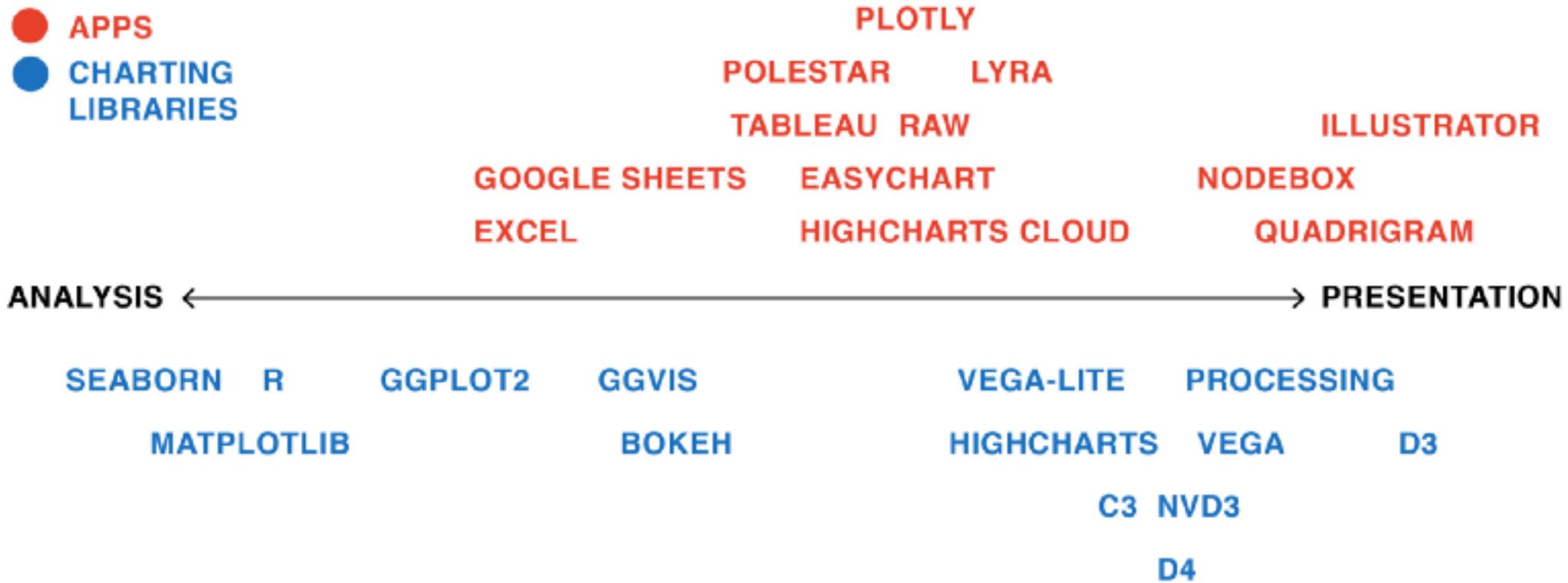
Lisa Charlotte Rost has an excellent blog post about this: <http://bit.ly/2gRGx1J>
I am presenting her figures here



Tools & Libraries for data visualization

Lisa Charlotte Rost has an excellent blog post about this: <http://bit.ly/2gRGx1J>

Analysis vs Presentation

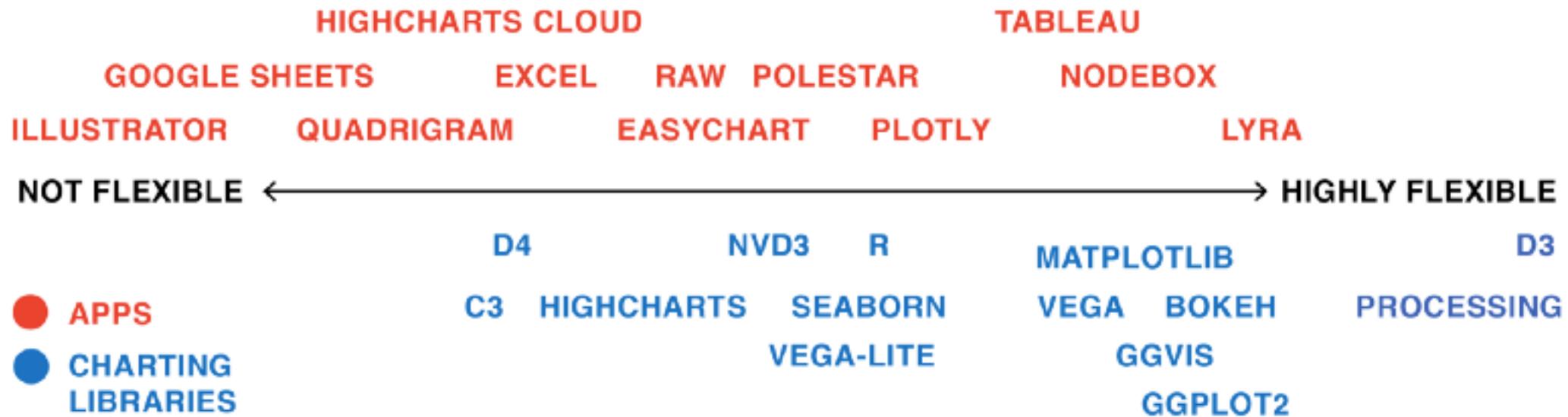


Tools & Libraries for data visualization

Lisa Charlotte Rost has an excellent blog post about this: <http://bit.ly/2gRGx1J>

Extent of Flexibility

How easy/hard it is to make data visualizations (including custom/novel visualizations)



Tools & Libraries for data visualization

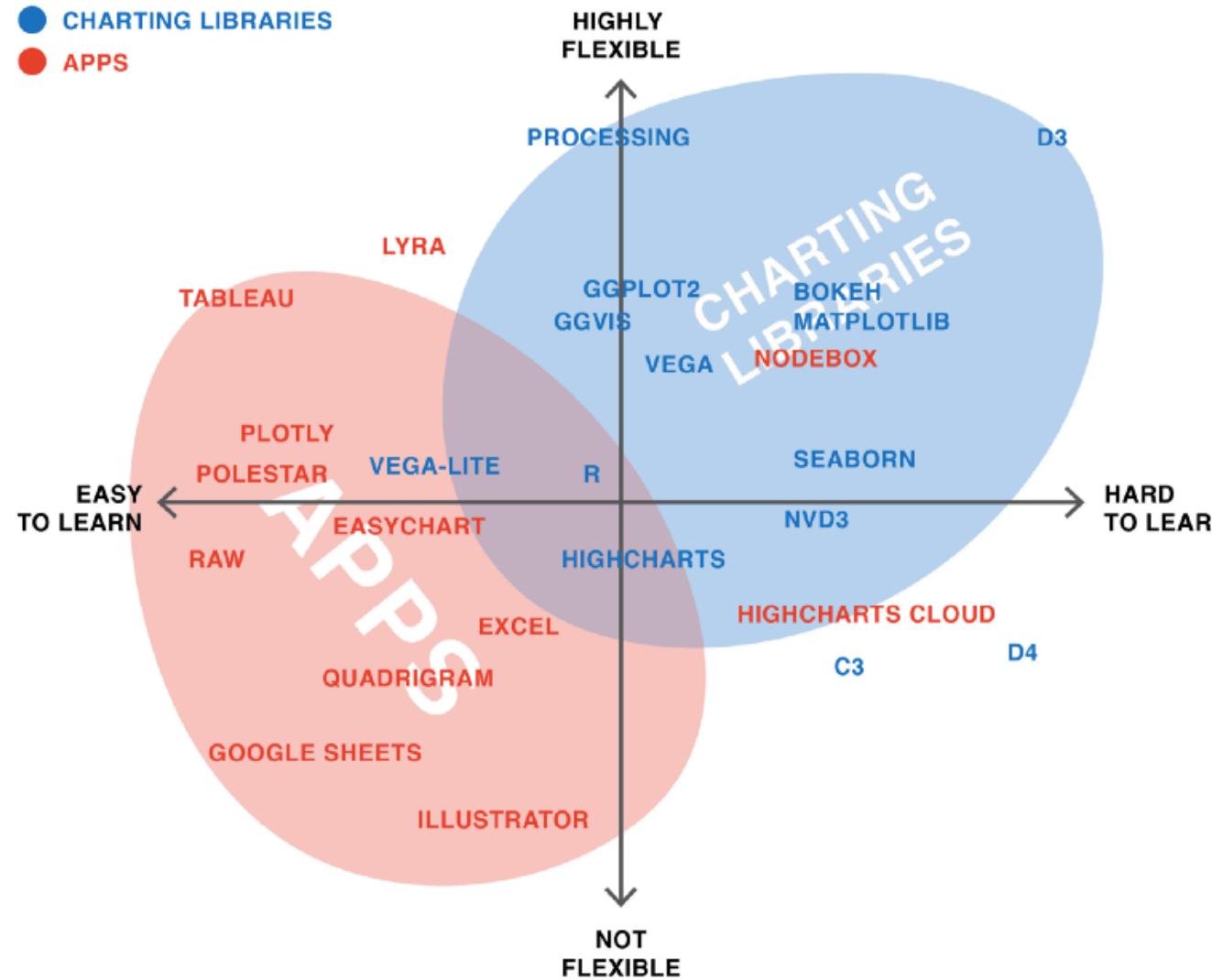
Lisa Charlotte Rost has an excellent blog post about this: <http://bit.ly/2gRGx1J>

Static vs Interactive

	STATIC	WEB - INTERACTIVE
APPS	ILLUSTRATOR, NODEBOX, EXCEL, POLESTAR, RAW	HIGHCHARTS CLOUD, QUADRIGRAM, EASYCHRT, DATAWRAPPER, TABLEAU, PLOTLY, GOOGLE SHEETS
CHARTING LIBRARIES	GGPLOT2, MATPLOTLIB, R, SEABORN, BOKEH, PROCESSING	D3, D4, C3, NVD3, GGVIS, HIGHCHARTS, SHINY, VEGA, VEGA-LITE

Tools & Libraries for data visualization

Lisa Charlotte Rost has an excellent blog post about this: <http://bit.ly/2gRGx1J>

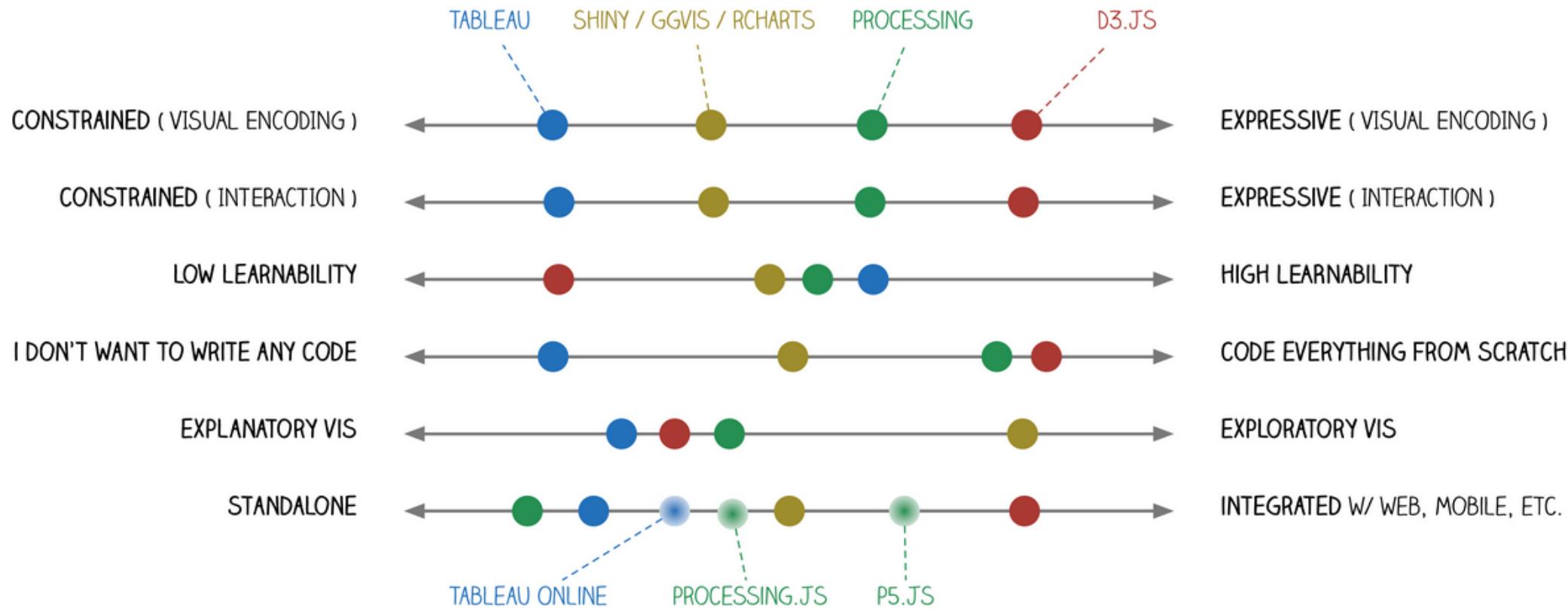


*"There are no perfect tools,
just good tools for people
with certain goals"*

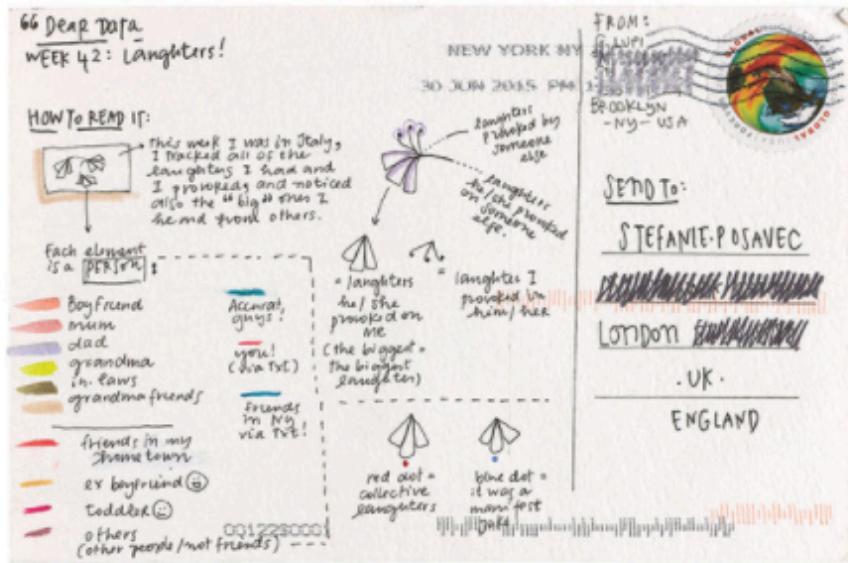
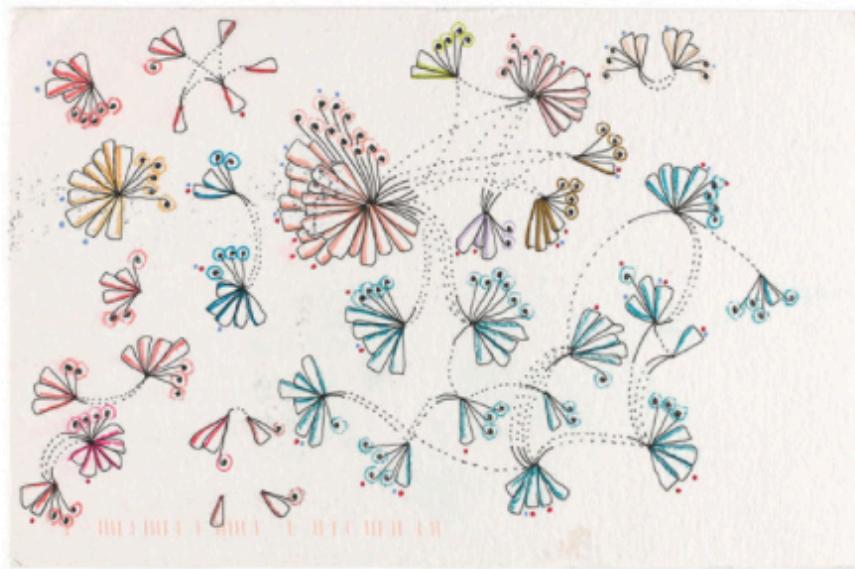
See a detailed table here:
<http://bit.ly/2DeWPwV>

Tools & Libraries for data visualization

Another take with commonly used tools : <https://bit.ly/2SgrOzS>



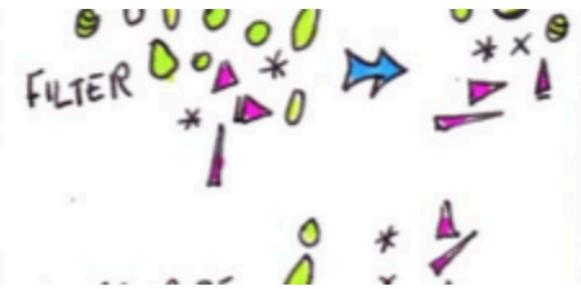
Don't forget that pen and paper is an option too!



Dear Data Project (Lipi & Posavec)

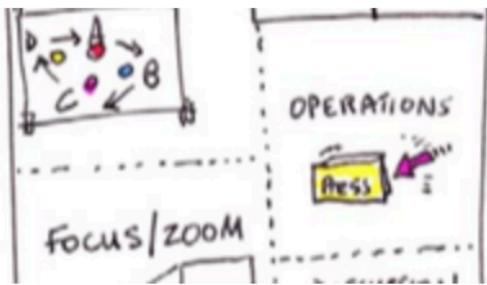
Don't forget that pen and paper is an option too!

Five Design Sheet Methodology



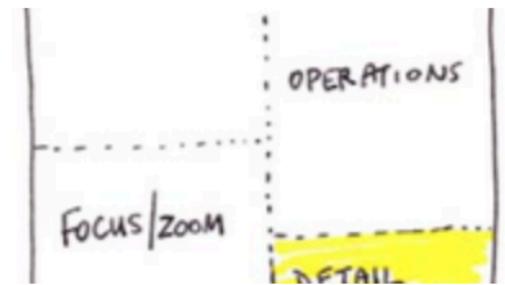
Sheet 1 : Brain storm

The idea of brain-storming is to enlarge the design space of possibilities. There should be a focus on quantity – to generate all the possible designs.



Sheet 2/3/4: initial designs

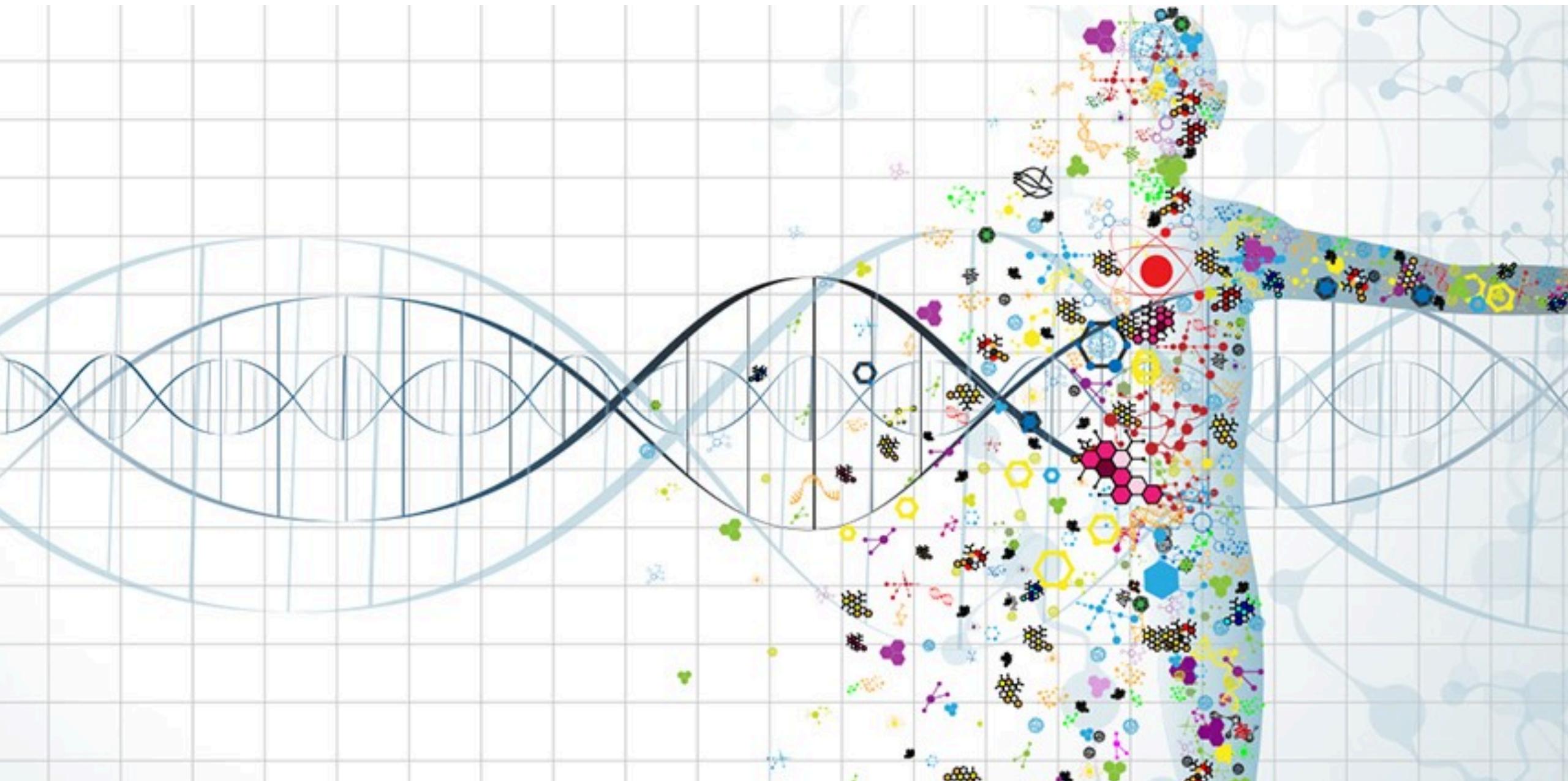
The individual design sheets are used to record three ideas from the brain-storming exercise in greater detail. These sheets should represent three completely different designs.



Sheet 5: Realization design

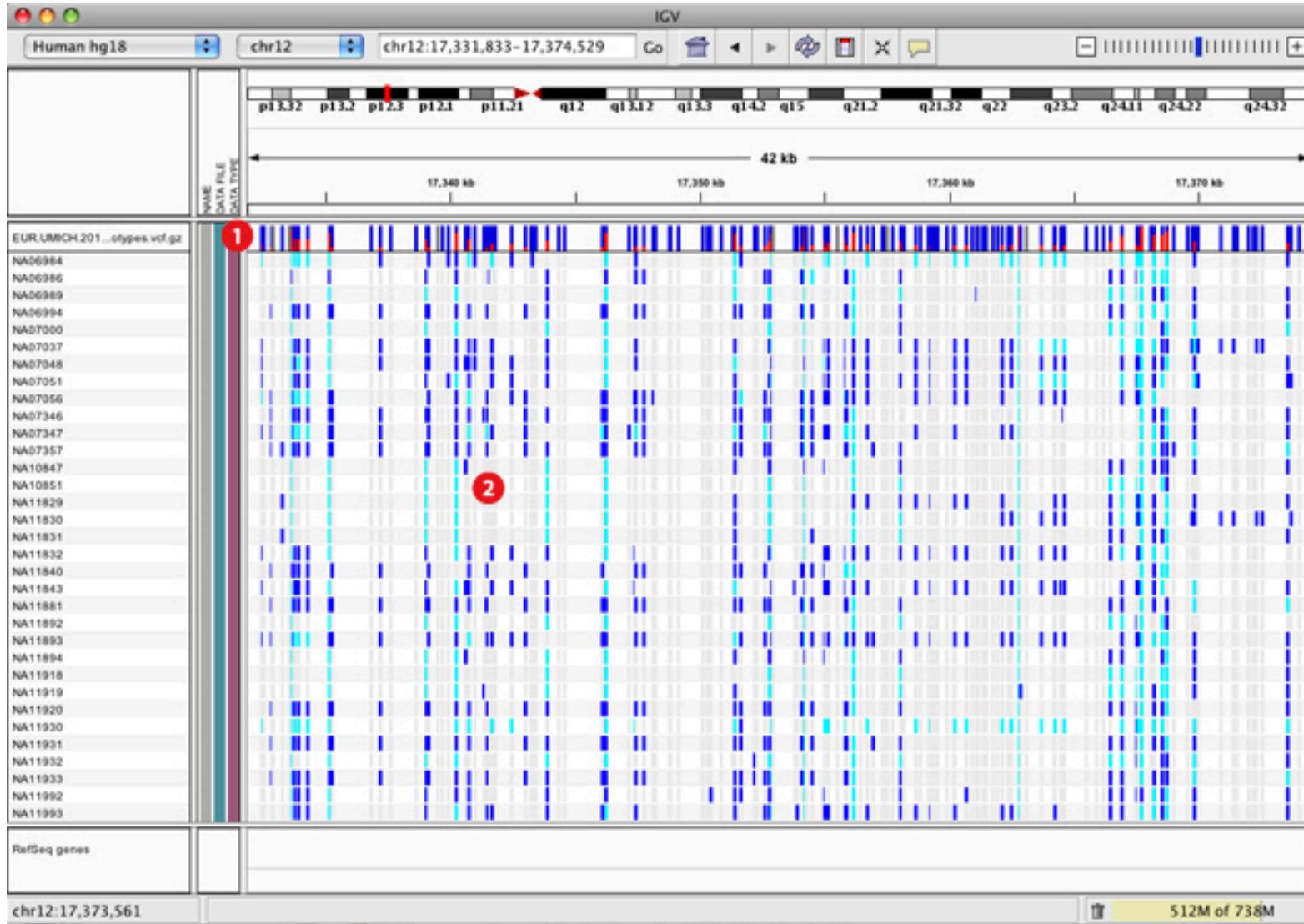
This is the realization design. This is what you think the visualization tool may look like, what specific visualization technique it principally uses and how users operate it.

Datavis tools for (Microbial) Genomics



IGV Browser for all your genomic needs

<https://software.broadinstitute.org/software/igv/>



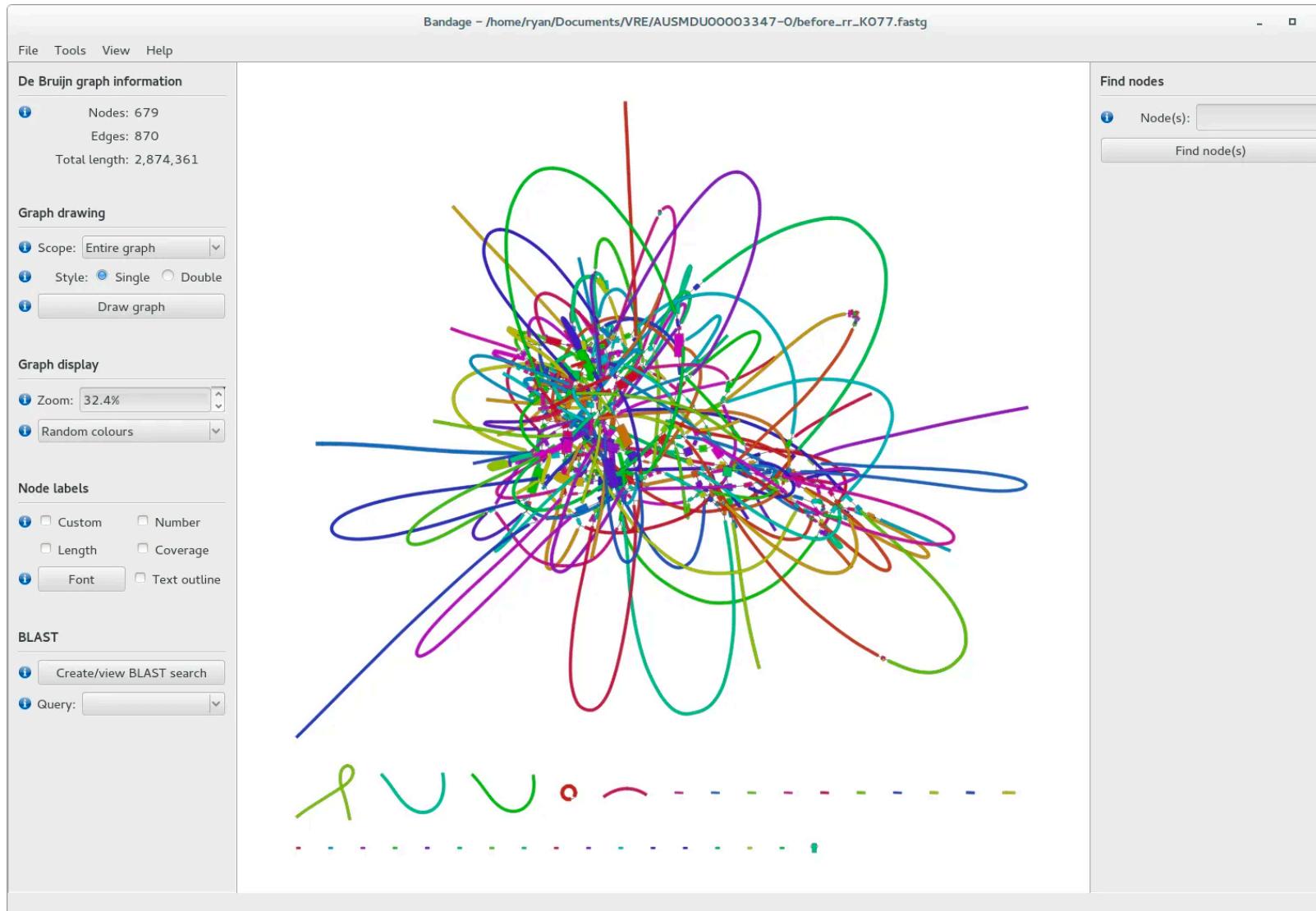
The classic UCSC genome browser

<https://genome.ucsc.edu>



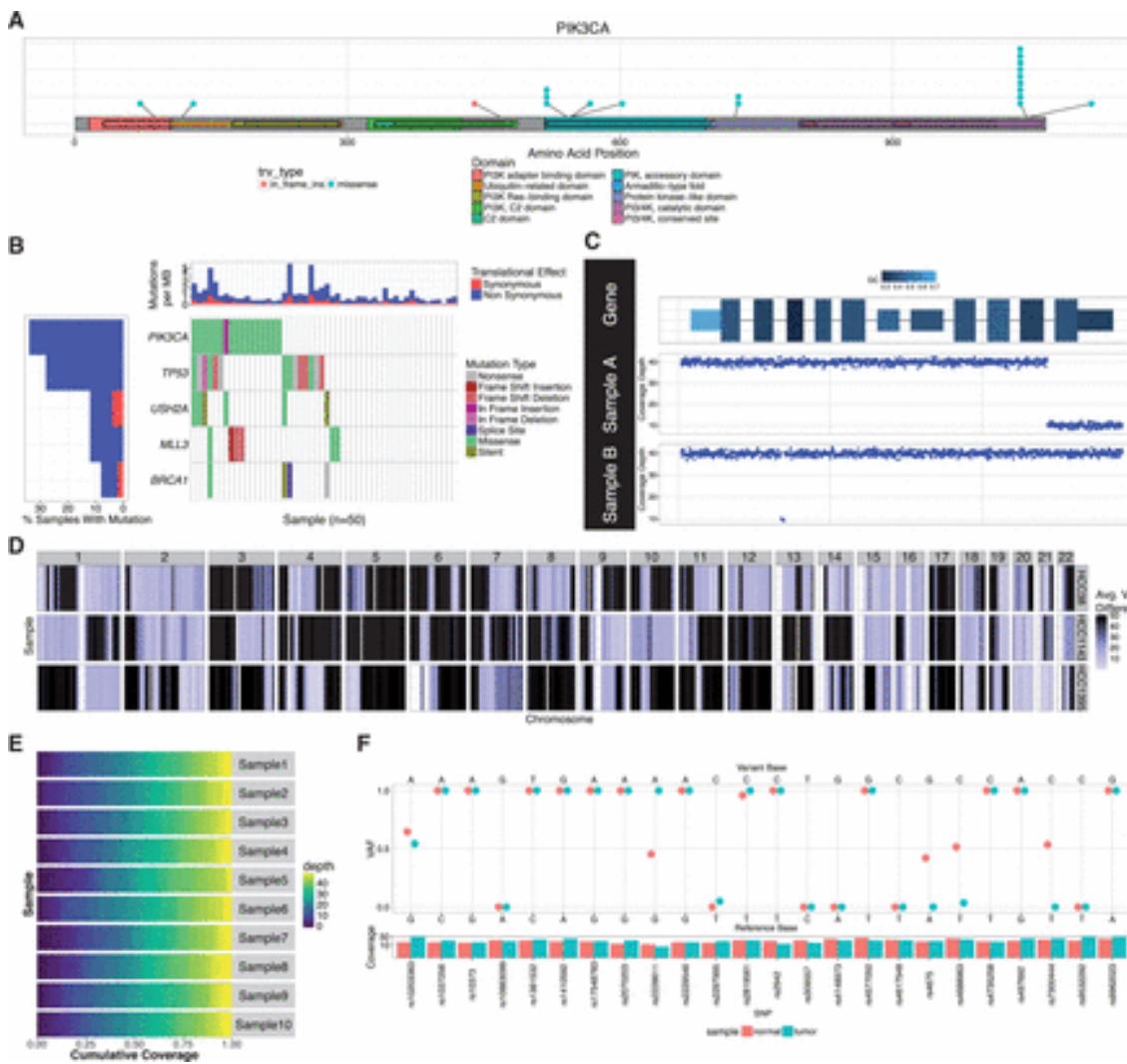
Bandage Plots

<https://rrwick.github.io/Bandage/>



GenVisR: Human Genomes in R

<https://academic.oup.com/bioinformatics/article/32/19/3012/2196360>

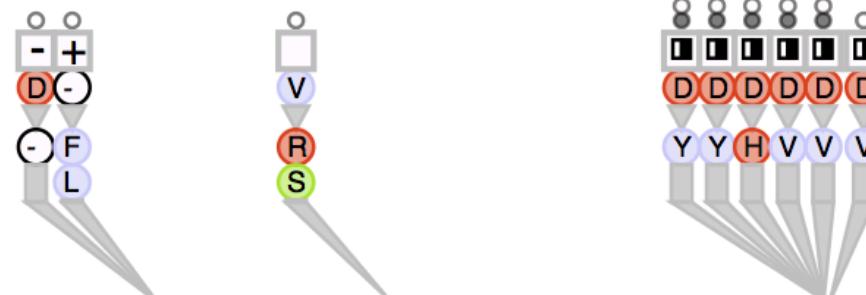


Variant Viewer: Human Genomes

<http://www.cs.ubc.ca/labs/imager/tr/2013/VariantView/>

Variants

Mutation Type
Reference A.A.s
Variant A.A.s



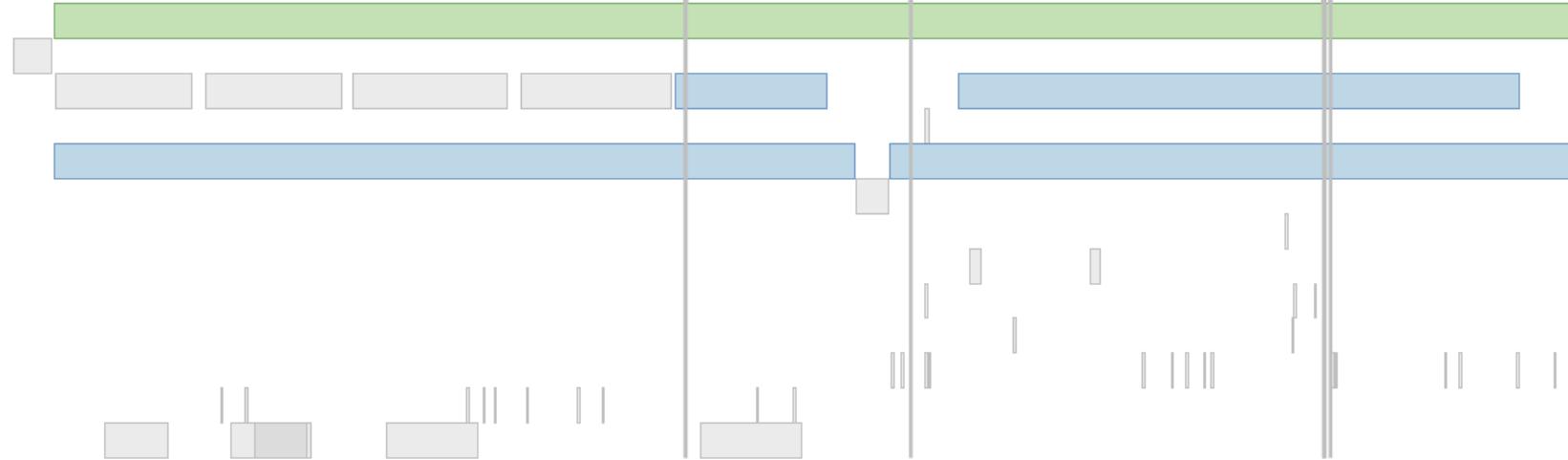
Transcript

trans-anon



Protein

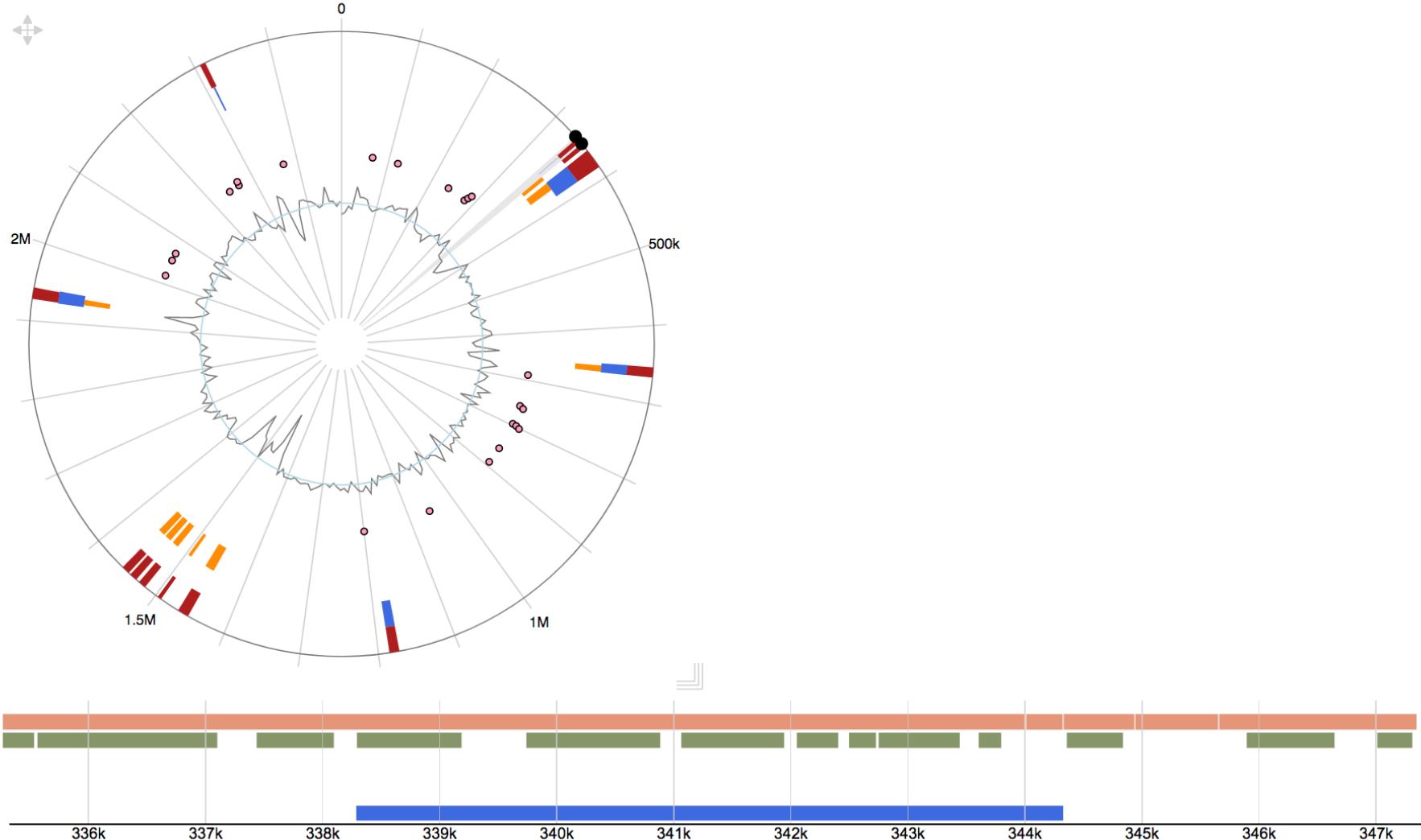
A.A. Chain
Signals
Domains
Regions
Topo. Domains
Transmem.
Active Sites
NP Binding
Metal Bind.
Bindings
Mod. Residue
Carbohyd.
Disulf.



Island Viewer: Microbial Genomics

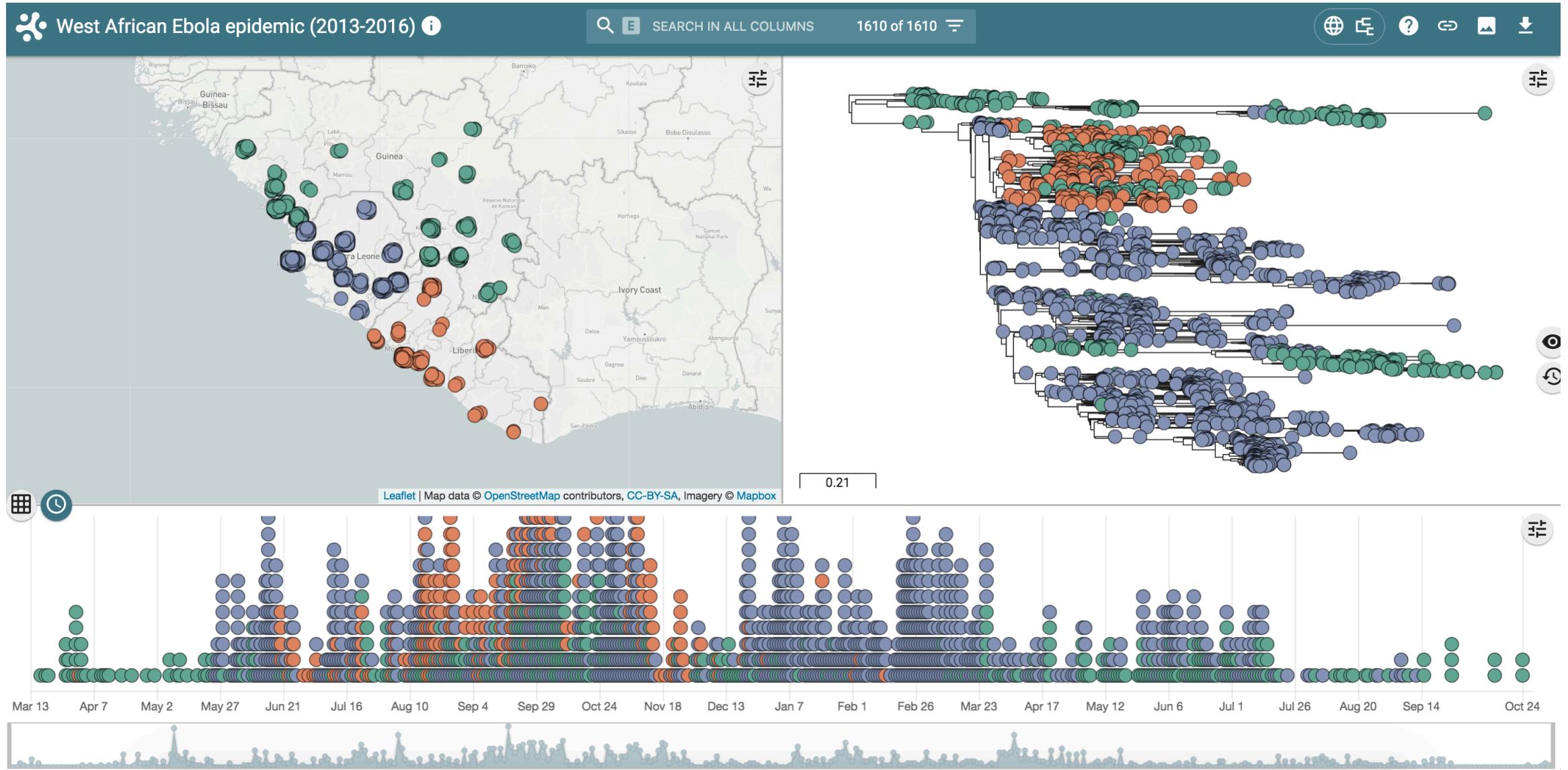
https://www.pathogenomics.sfu.ca/islandviewer/accession/NZ_CP012358.1/

PSEUDOMONADACEAE BACTERIUM B4199, COMPLETE GENOME.



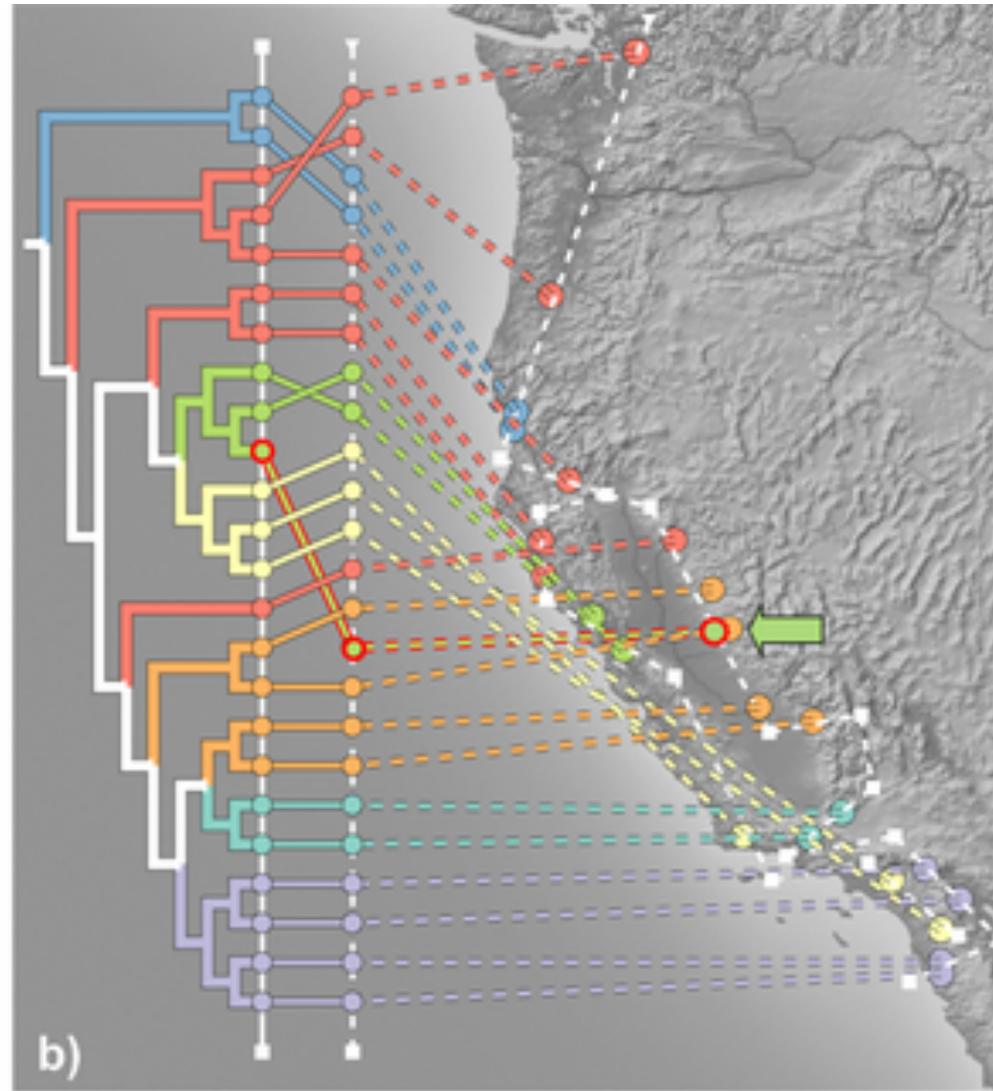
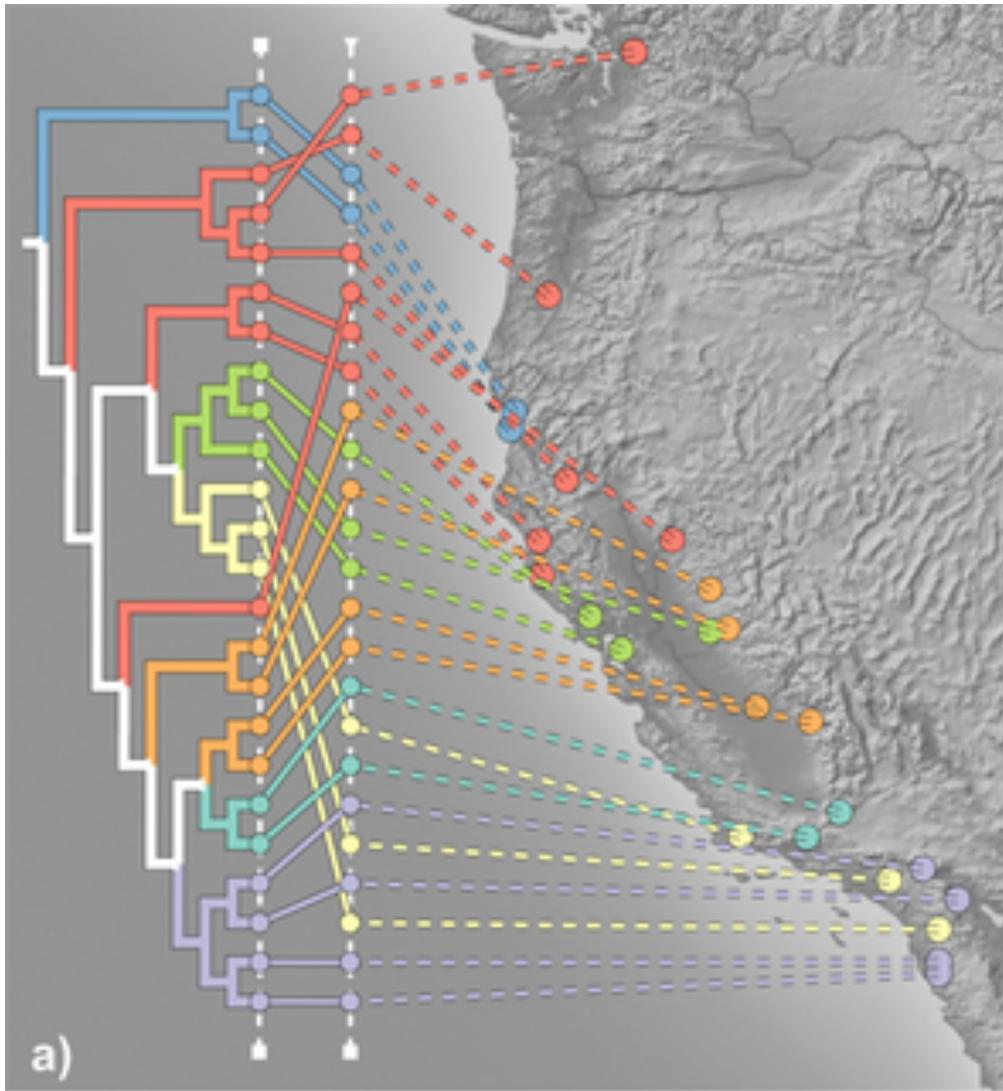
Microreact: Microbial Genomics

<https://microreact.org/>



GenGIS: Microbial Genomics (Made in Canada!)

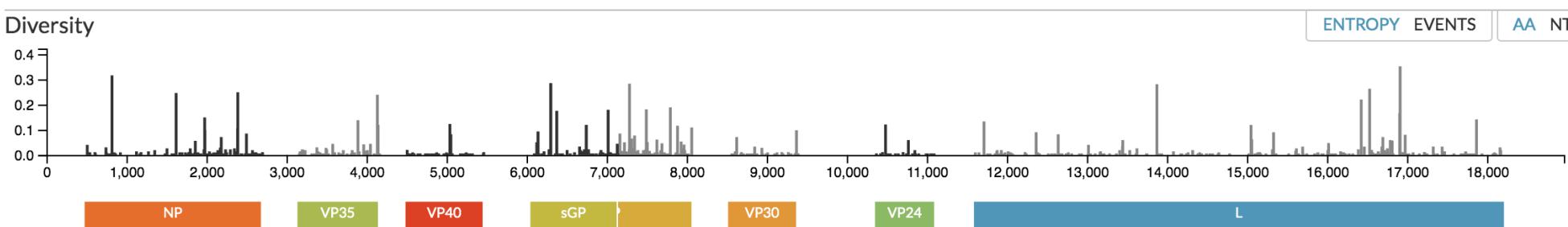
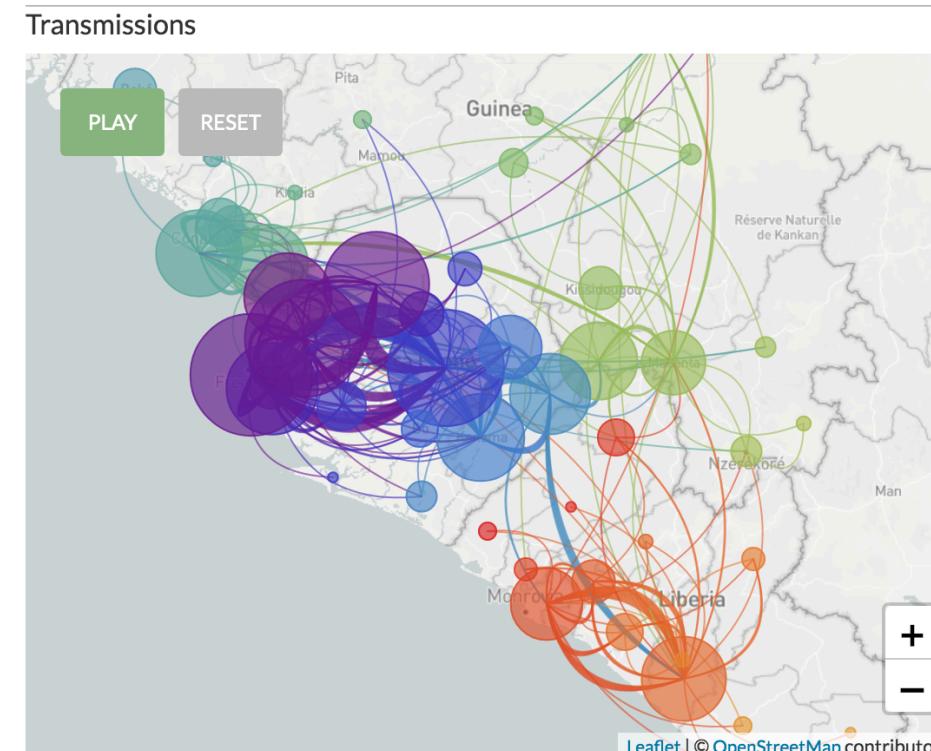
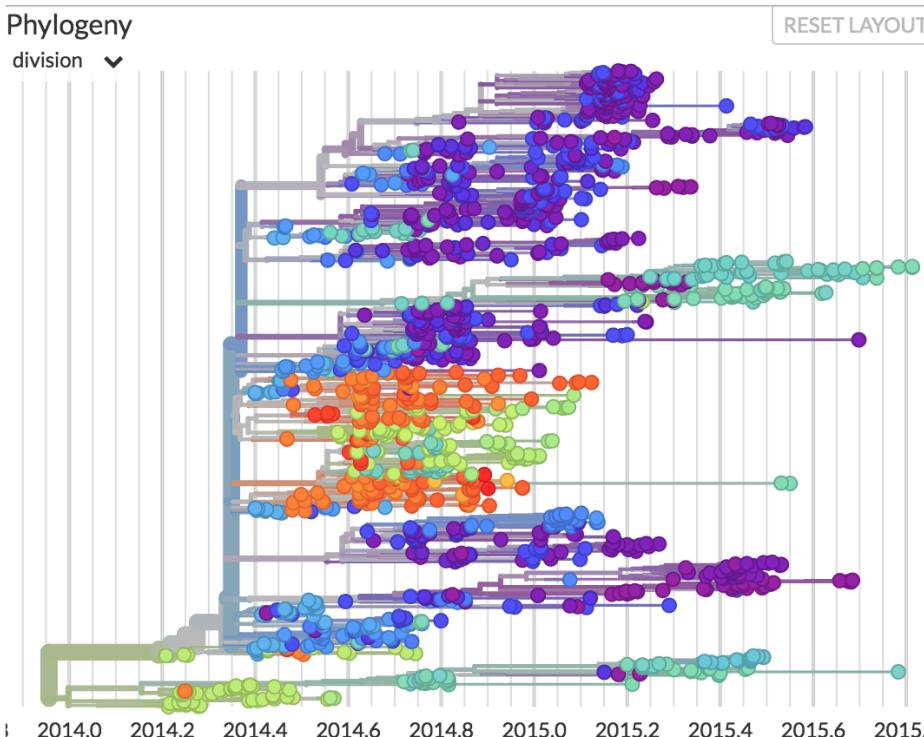
http://kiwi.cs.dal.ca/GenGIS/Main_Page



Nextstrain: Microbial Genomics

<https://nextstrain.org>

Showing 1493 of 1493 genomes sampled between Mar 2014 and Oct 2015 and comprising 3 countries, 50 divisions and 19 authors.



Wrapping up



**DATA VISUALIZATION
IS NOT
JUST AN ART PROJECT**



Key take-aways from this talk

- **Visualizations of data are useful**
 - Helpful in instance of low numeracy
 - Can used in communication and exploration
- **But.. visualization design also matters**
 - Many different alternatives, important to test
- **It's possible to think systematically about visualizations**
 - Many disciplines cross cut information visualization research
 - At the minimum think "Why", "What", "How"
- **Encode data well so that others can decode it later**
- **Data visualization is a *research process* with open and interesting problems**

Additional Resources

- Books to consider:
 - Interpretable Machine Learning: <https://christophm.github.io/interpretable-ml-book/>
 - Making Data Visual: A Practical Guide to Using Visualization for Insight by Danyel Fisher and Miriah Meyer
 - Visualization Design and Analysis by Tamara Munzner (more technical)
- Online resources:
 - Distill Publication : <https://distill.pub/>
 - UBC Infovis Resource Page : <http://www.cs.ubc.ca/group/infovis/resources.shtml>
 - UW Interactive Data Lab : <https://medium.com/@uwdata>
 - Data stories podcast : <http://datastori.es/>
- Inspiration :
 - Information is Beautiful : <https://informationisbeautiful.net/>
 - Visualization WTF (examples of what not to do) : <http://viz.wtf/>

What to get involved? Consider the biovis challenge!



BioVis at IEEE VIS (BioVis@VIS)

October 2020, Salt Lake City, USA - In conjunction with [IEEE VIS 2020](#)



BioVis Challenges

[VIEW DETAILS](#)



Program

[VIEW DETAILS](#)

Data visualization strategies and tools for microbial genomic epidemiology

Anamaria Crisan

Research Scientist

Tableau Research

Tableau, Seattle, WA



@amcrisan



acrisan@tableau.com



<https://amcrisan.github.io>