Visualization of Public Health Data

Anamaria Crisan

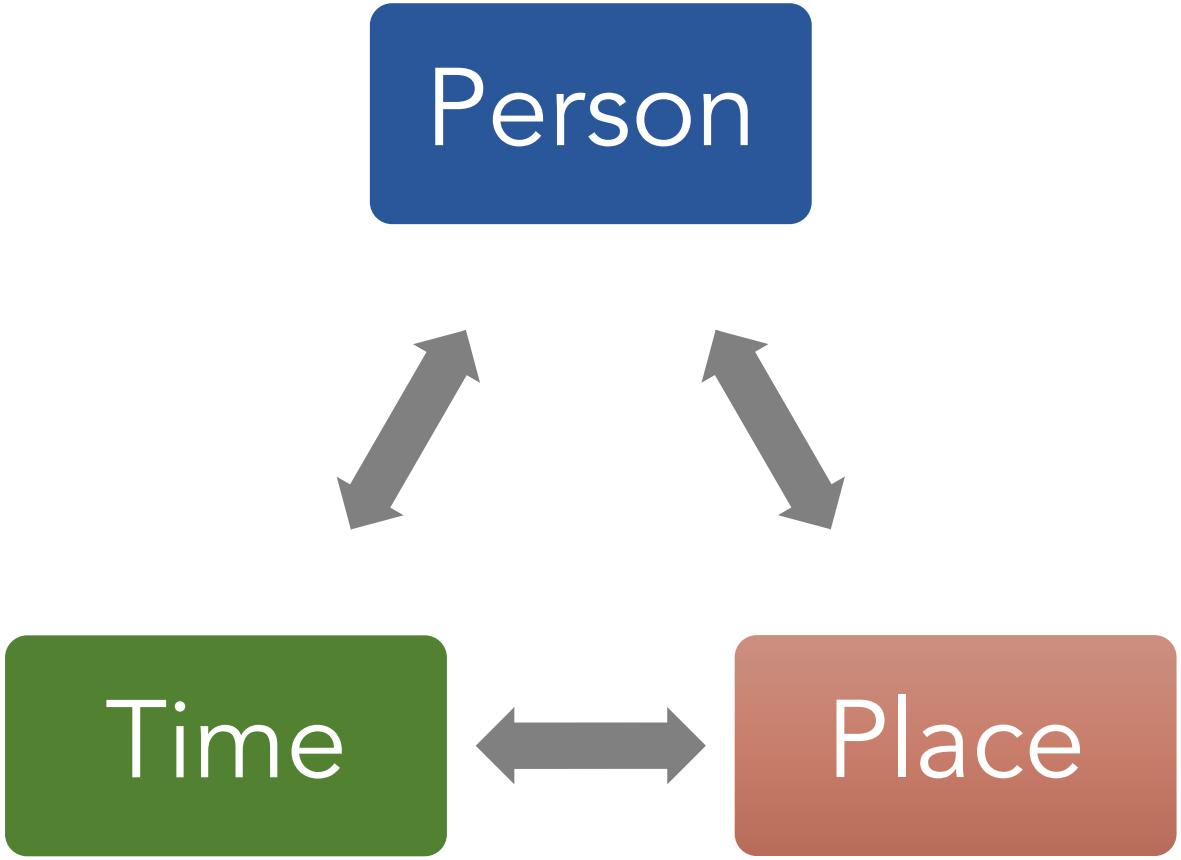
PhD Student at UBC in Computer Science

Supervisors:

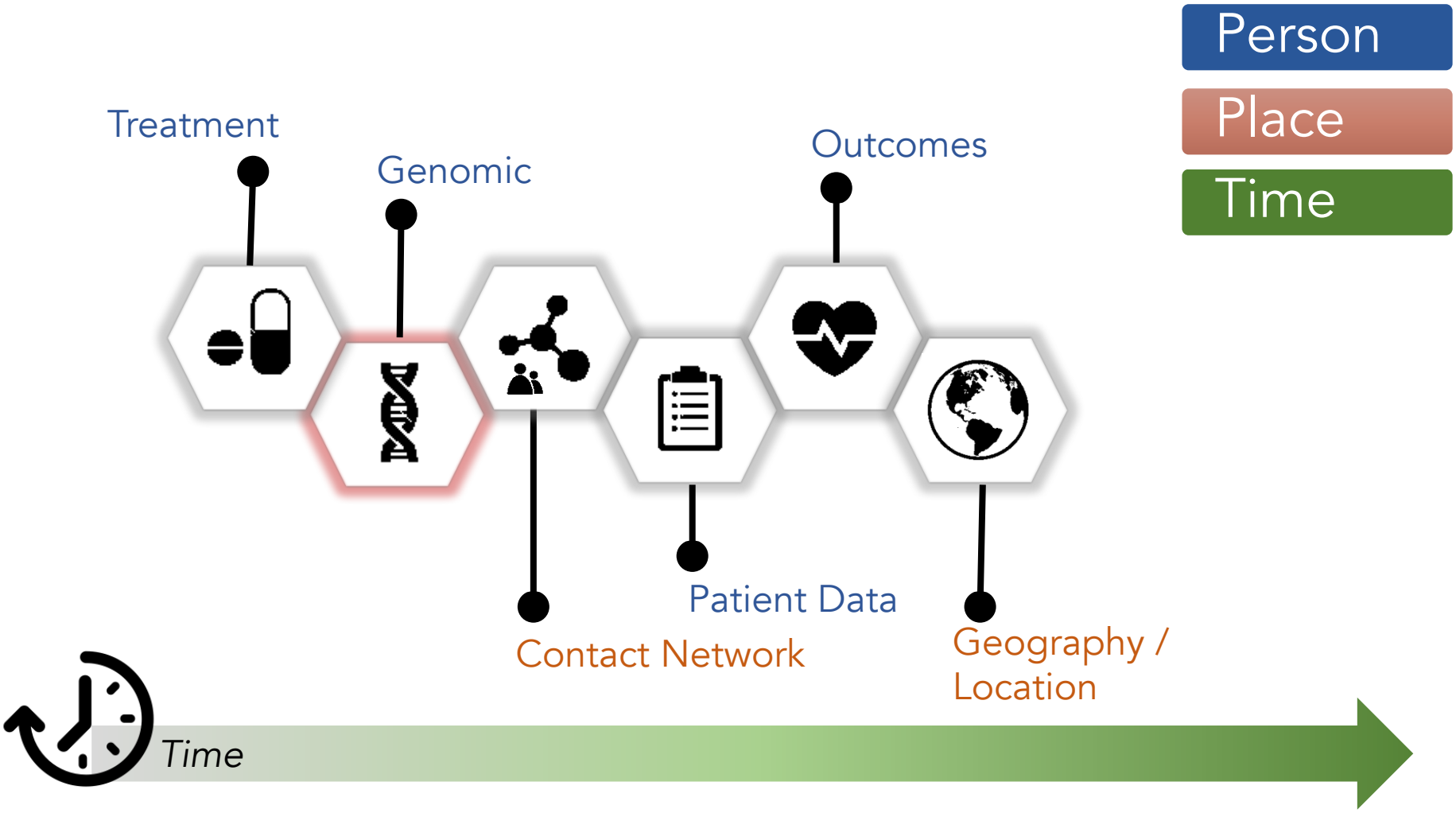
Jennifer Gardy , *Population and Public Health*

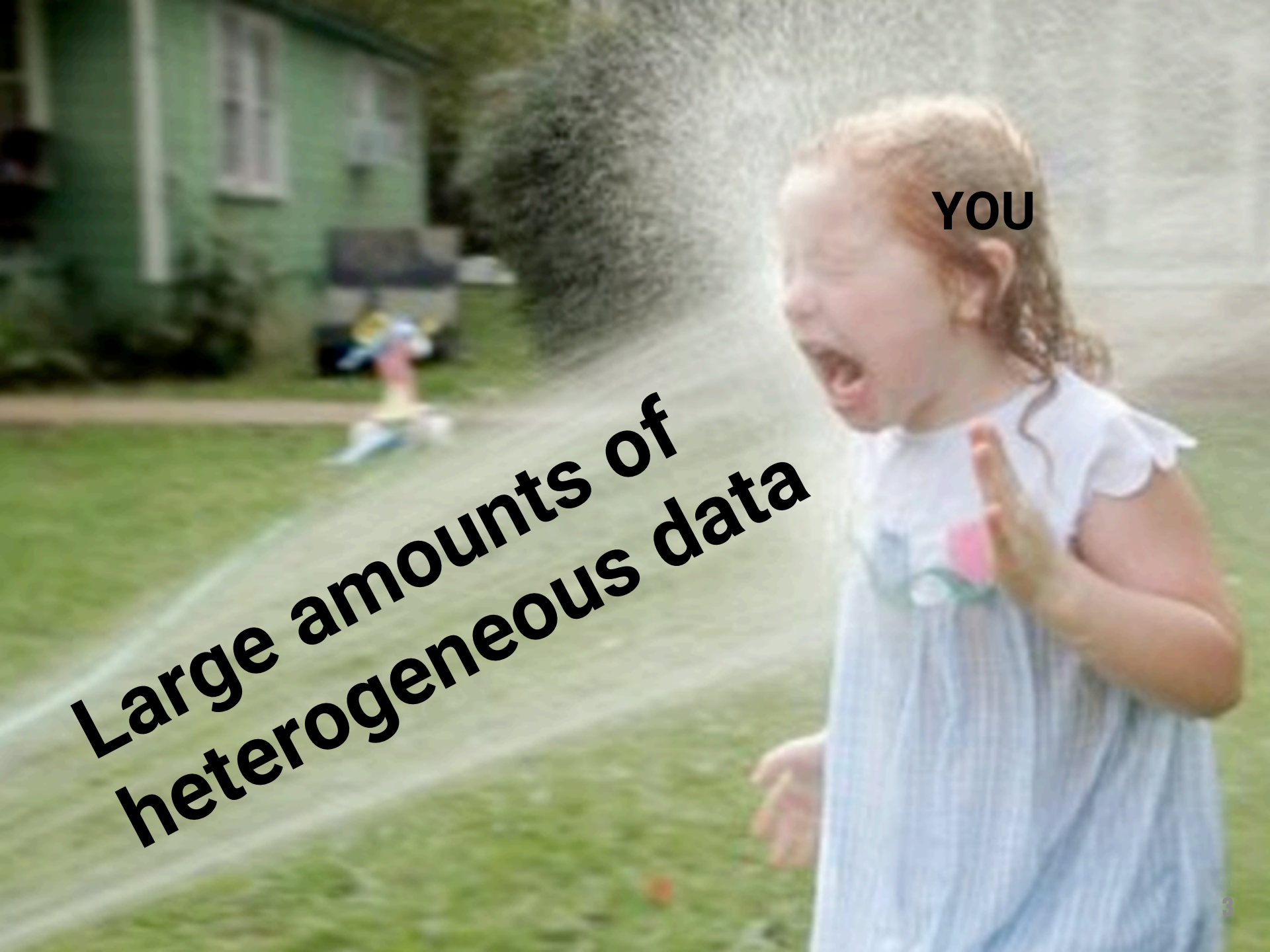
Tamara Munzner, *Computer Science*

WHAT ARE PUBLIC HEALTH DATA?
(FOR INFECTIOUS DISEASE MANAGEMENT)



WHAT ARE PUBLIC HEALTH DATA?
(FOR INFECTIOUS DISEASE MANAGEMENT)




YOU
Large amounts of heterogeneous data

SUPPORT FOR DATA DRIVEN DECISIONS

- Public health has multidisciplinary decision making teams
  - More data & diverse data types = more informed decision making
  - BUT - not all stakeholders can interpret / understand data
- Support needed for decision making with heterogeneous data

Medical Health Officers

Clinicians

Nurses

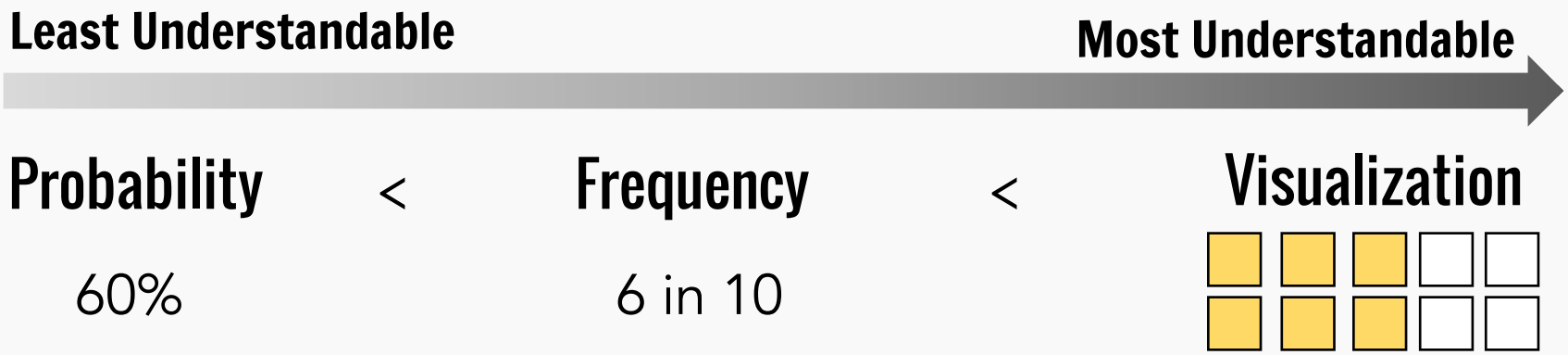
Researchers

Community Leaders

PROPOSAL

Visualization of public health data **can improve knowledge sharing and decision making** in infectious disease prevention and control

WHY VISUALIZATION?

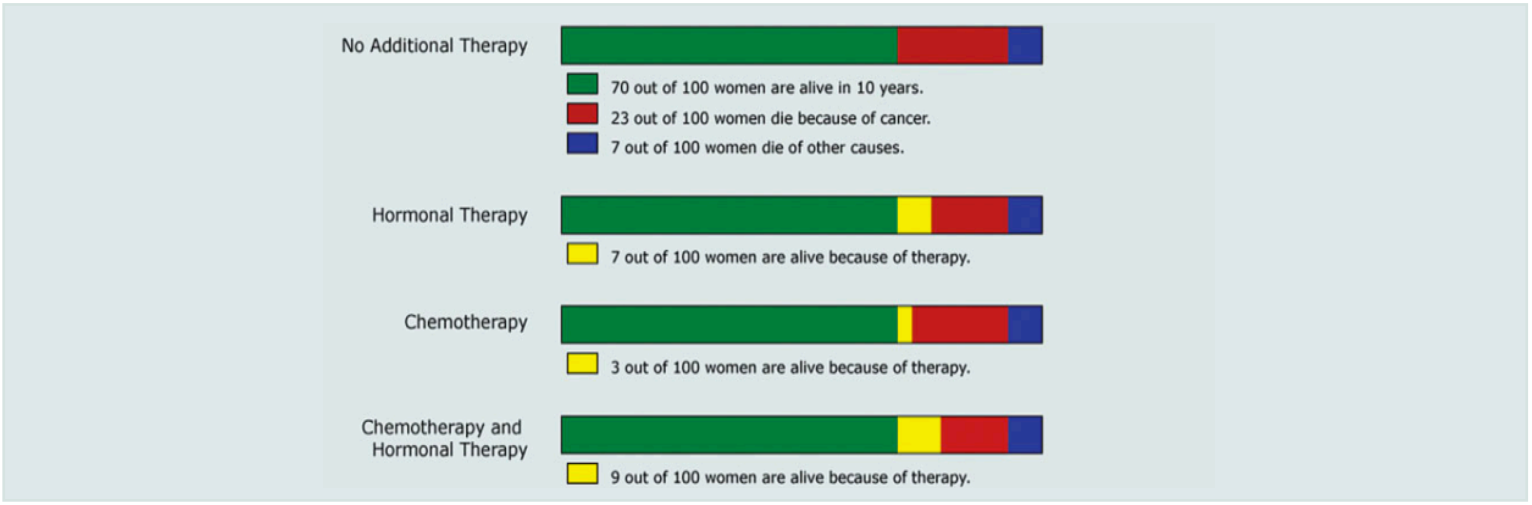


- Numeracy : the ability to reason with numbers
  - Individuals with **low numeracy** have a difficulty interpreting numbers and probabilities
  - Also true amongst educated professionals
- **Visualization can make data more accessible to diverse stakeholders on decision making teams**

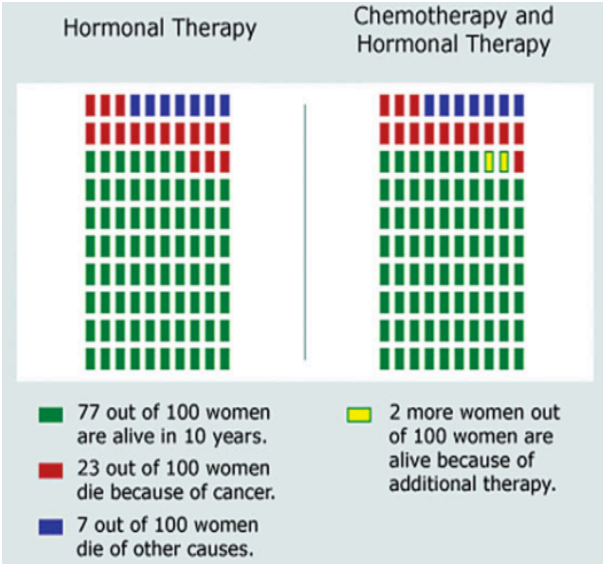
Whiting (2015) “How well do health professionals interpret diagnostic information? A systematic review”

BUT! VISUAL DESIGN ALSO MATTERS

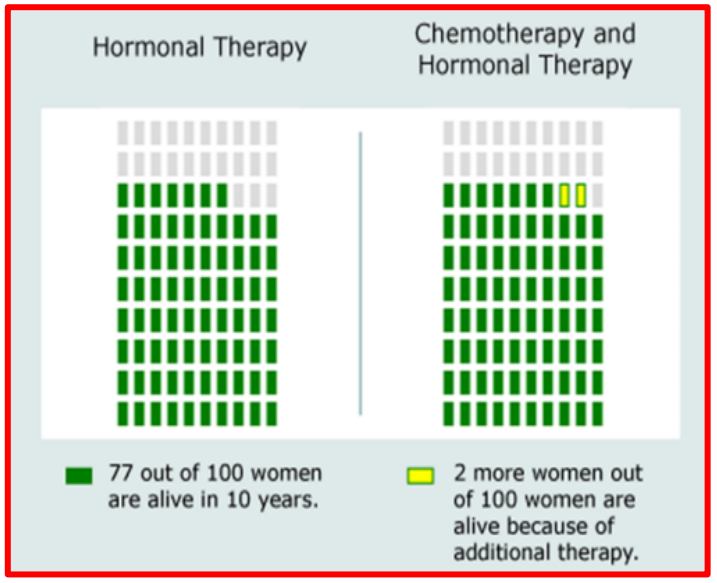
Baseline Visualization



Alternative 1



Alternative 2



Zikmund-Fisher (2013). A demonstration of "less can be more" in risk graphics.

EXAMPLE OF GUIDANCE : WWW. VIZHEALTH.ORG

VISUALIZING HEALTH

A Scientifically Vetted Style Guide for Communicating Health Data

VISUALIZING HEALTH | About | The Wizard | Browse the Gallery | Using Visualizing Health | Disclaimer | Download Repor

My goal

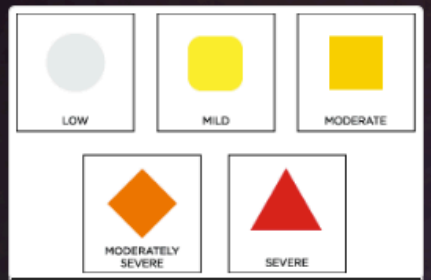
- Classifying risks
- Raise or lower concern
- Awareness of risk
- Differences in likelihood
- Risk tradeoffs

Details or gist?

- Verbatim recall
- Gist understanding

Data I have

- Benefit estimate
- Risk over time
- Case counts

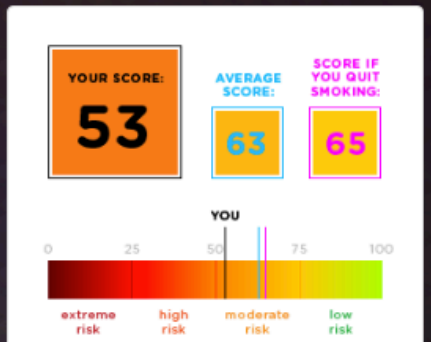


(84) Icons to show severity of side effects

...MORE LIKE THIS

Classifying risks

ALL TAGS



(43) Visualizing health scores

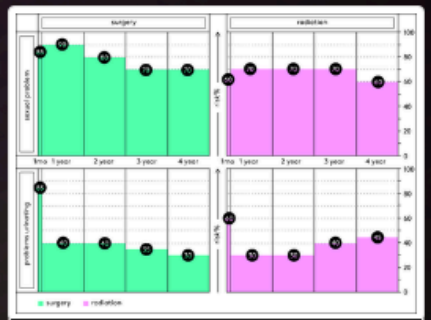
...MORE LIKE THIS

Classifying risks

Raise or lower concern

Awareness of risk

ALL TAGS



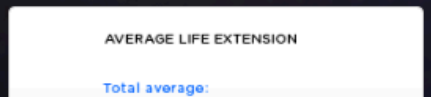
(66) Showing how side effects change over time

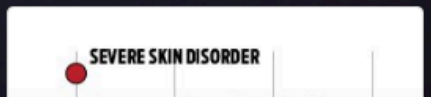
...MORE LIKE THIS

Risk tradeoffs ★

Differences in likelihood ★

ALL TAGS





APPLICATION TO PUBLIC HEALTH

- Lots of interest in Visualization in Public Health
- But - mainly developing *ad hoc* solutions
  - Visualization designers usually bioinformaticians (high numeracy, lack stakeholder context)
  - Stakeholders relying on Excel for visualizations
- **Need to make a case for better visualizations**
- **Need to treat data visualization as a research process**

VISUALIZATION DESIGN & ANALYSIS

Steps for visual design

1. Partner with a group of stakeholders that have a **problem**

VISUALIZATION DESIGN & ANALYSIS

Steps for visual design

1. Partner with a group of stakeholders that have a **problem**
2. Ask **what data** stakeholders use (is it available)?

VISUALIZATION DESIGN & ANALYSIS

Steps for visual design

1. Partner with a group of stakeholders that have a **problem**
2. Ask **what data** stakeholders use (is it available)?
3. Ask **what stakeholders do** with the data [**tasks**]

VISUALIZATION DESIGN & ANALYSIS

Steps for visual design

1. Partner with a group of stakeholders that have a **problem**
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**

VISUALIZATION DESIGN & ANALYSIS

Steps for visual design

1. Partner with a group of stakeholders that have a **problem**
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**
5. Test multiple alternatives (including new ones you develop) with stakeholders

VISUALIZATION DESIGN & ANALYSIS

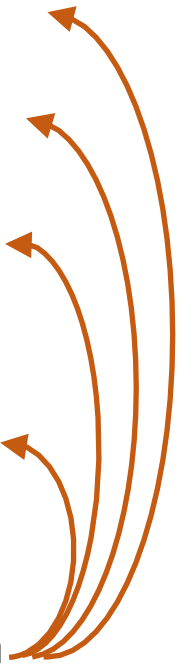
Steps for visual design

1. Partner with a group of stakeholders that have a **problem**
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**
5. Test multiple alternatives (including new ones you develop) with stakeholders
6. Gather qualitative & quantitative evaluation data

VISUALIZATION DESIGN & ANALYSIS

Steps for visual design AN ITERAVTIVE PROCESS

1. Partner with a group of stakeholders that have a **problem**
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**
5. Test multiple alternatives (including new ones you develop) with stakeholders
6. Gather qualitative & quantitative evaluation data



EXAMPLE: TB GENOMIC CLINICAL REPORT

Current Report

Mycobacterium Whole Genome Sequencing Report from MGIT Positive Samples

Not for diagnostic use 11/12/2015

Sample Details

Sequencing Location	Oxford	**Date received in Lab**	
Local Lims Specimen ID	12.0610882	**Run date**	01/01/20151008
Guuid	b7aa98e0-3612-4c0b-a47b-471e0e78c72d		

Organism Identification

Predicted/closest match	
TBCOMP/microti	100%
TBCOMP	100%
TBCOMP/TB	96.77%
TBCOMP/tuberculosis-canettii	35.71%
MACCOMP	21.21%

Sample/Sequencing Quality

Total reads (~millions)	**Mapped %**	**No reads mapped (~millions)**	**Coverage %**
4.73	99.47	4.7	91.99

Resistance Summary

INH	**RIF**	**EMB**	**PZA**	**QUI**	**SM**	**AG**
U	S	S	S	S	S	S

Resistotype

Drug	**Mutation**	**Nucleotides**	**Support (ACGT)**	**Source – (R/Total)**	**Prediction**
INH	katG_A727T	GCC->ACC	(160/0/1/0) (0/164/0/0) (0/167/0/0)	Unclassified	UNK

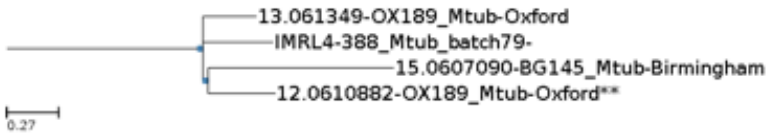
Relatedness

NB: This data may be added or updated at a later date

Nearest neighbour(s)

Sample -Plate Name	**Date received in Lab**	**Centre**	**No. of SNPs apart**
12.0610882-OX189_Mtub		Oxford	0
IMRL4-388_Mtub_batch79	1900-01-01		13
15.0607090-BG145_Mtub	2015-05-26	Birmingham	16
13.061349-OX189_Mtub		Oxford	8

The alignment width is 285. Multiply this number by the tree metrics.



Comments

Authorised

Signature:	**Print name:**
Position:	**Date:**

DESIGN PROCESS OVERVIEW

Question: Can we improve upon the existing report design

Note: Not a data vis project, but uses data vis methods and result will feed into other data vis projects

Phase 1: Expert consultations

Phase 2: Task Questionnaire

Design Sprint

Phase 3: Design choice Questionnaire

Phase 4: Evaluation of final report design

DESIGN PROCESS OVERVIEW

Question: Can we improve upon the existing report design

Note: Not a data vis project, but uses data vis methods and result will feed into other data vis projects

Phase 1: Expert consultations

Phase 2: Task Questionnaire

Design Sprint

Phase 3: Design choice Questionnaire

Phase 4: Evaluation of final report design

PHASE 1
EXPERT CONSULTATIONS

Participants: 7 = physicians (clinical & laboratory), public health researchers

Key Findings

- Different needs between physicians and researchers
- Physicians had greater time pressure
- Trust in lab and procedures
- Some data on report not necessary, other data confusing
- Constraints on delivery report due EHR

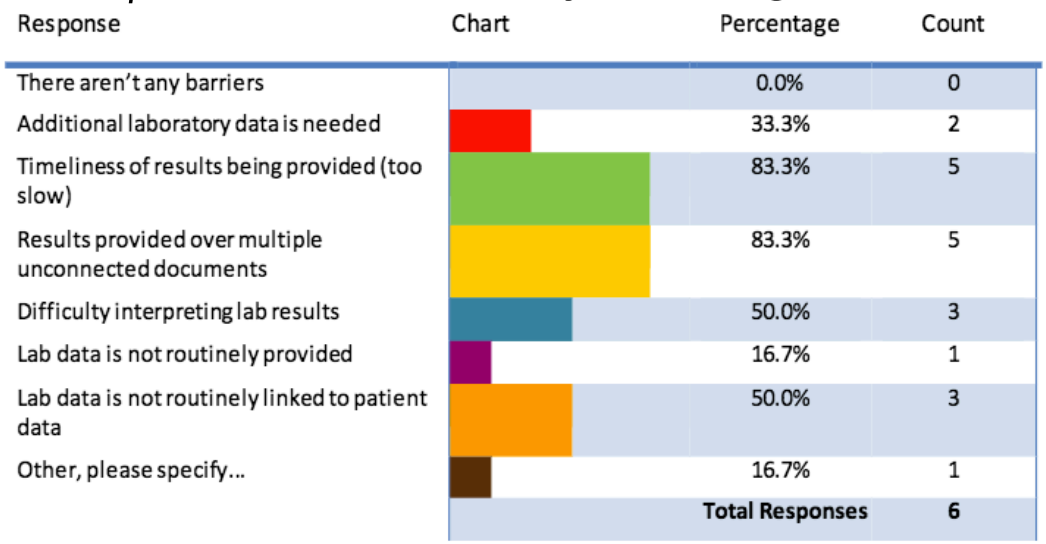
PHASE 2
TASK QUESTIONNAIRE

Participants: 17 = physicians (clinical & laboratory), nurses, public health researchers, surveillance experts

Key findings

- Quantitative support for earlier qualitative findings
- Better granularity of data used, and confidence performing, different tasks

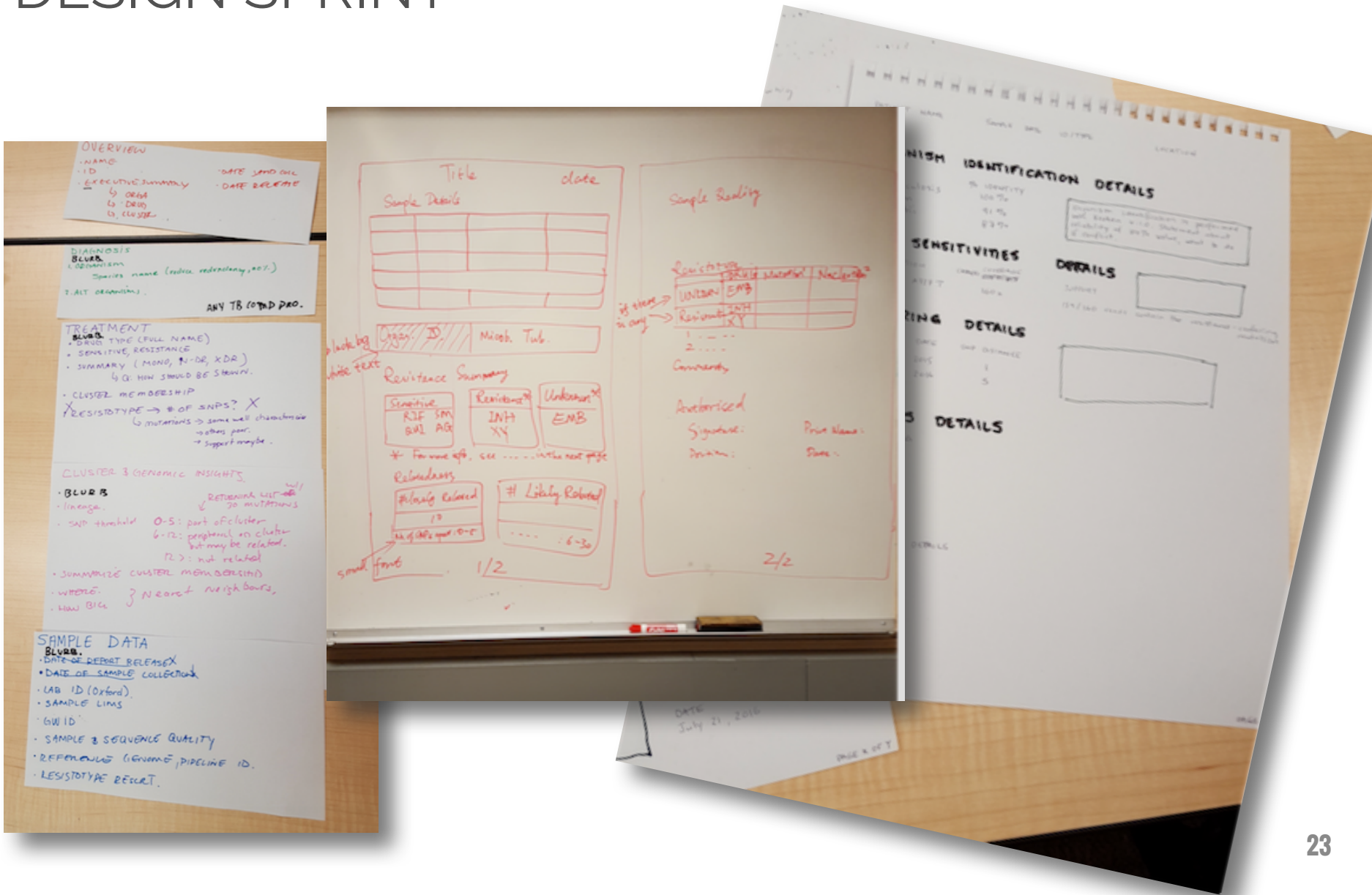
Q: What could improve the efficiency of using molecular data?



Response	Chart	Percentage	Count
There aren't any barriers		0.0%	0
Additional laboratory data is needed		33.3%	2
Timeliness of results being provided (too slow)		83.3%	5
Results provided over multiple unconnected documents		83.3%	5
Difficulty interpreting lab results		50.0%	3
Lab data is not routinely provided		16.7%	1
Lab data is not routinely linked to patient data		50.0%	3
Other, please specify...		16.7%	1
		Total Responses	**6**

PHASE 2
DESIGN SPRINT

PHASE 2
DESIGN SPRINT

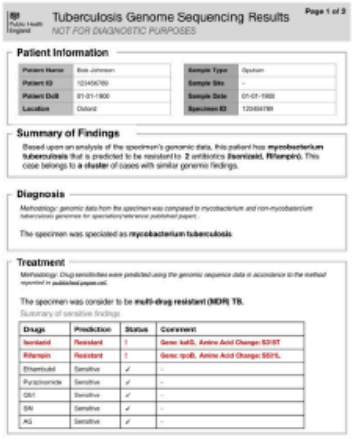


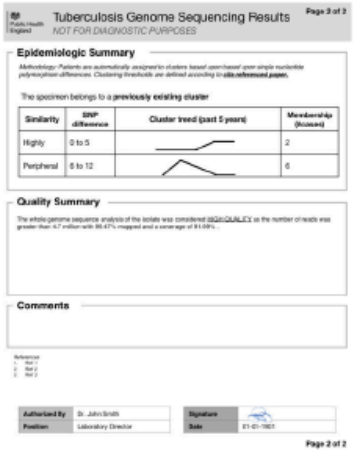
PHASE 3

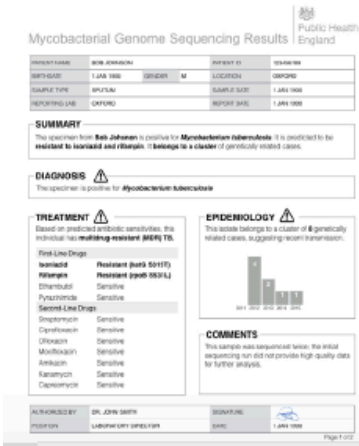
DESIGN CHOICE QUESTIONNAIRE

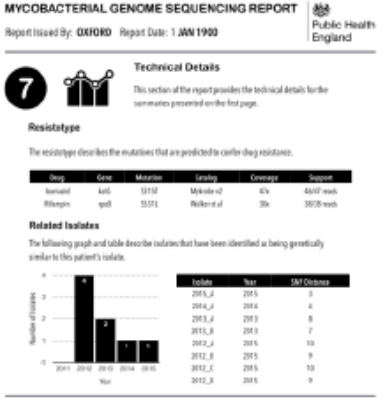
Participants: 42

Goal: Compare control (existing report) with options developed in the design sprint



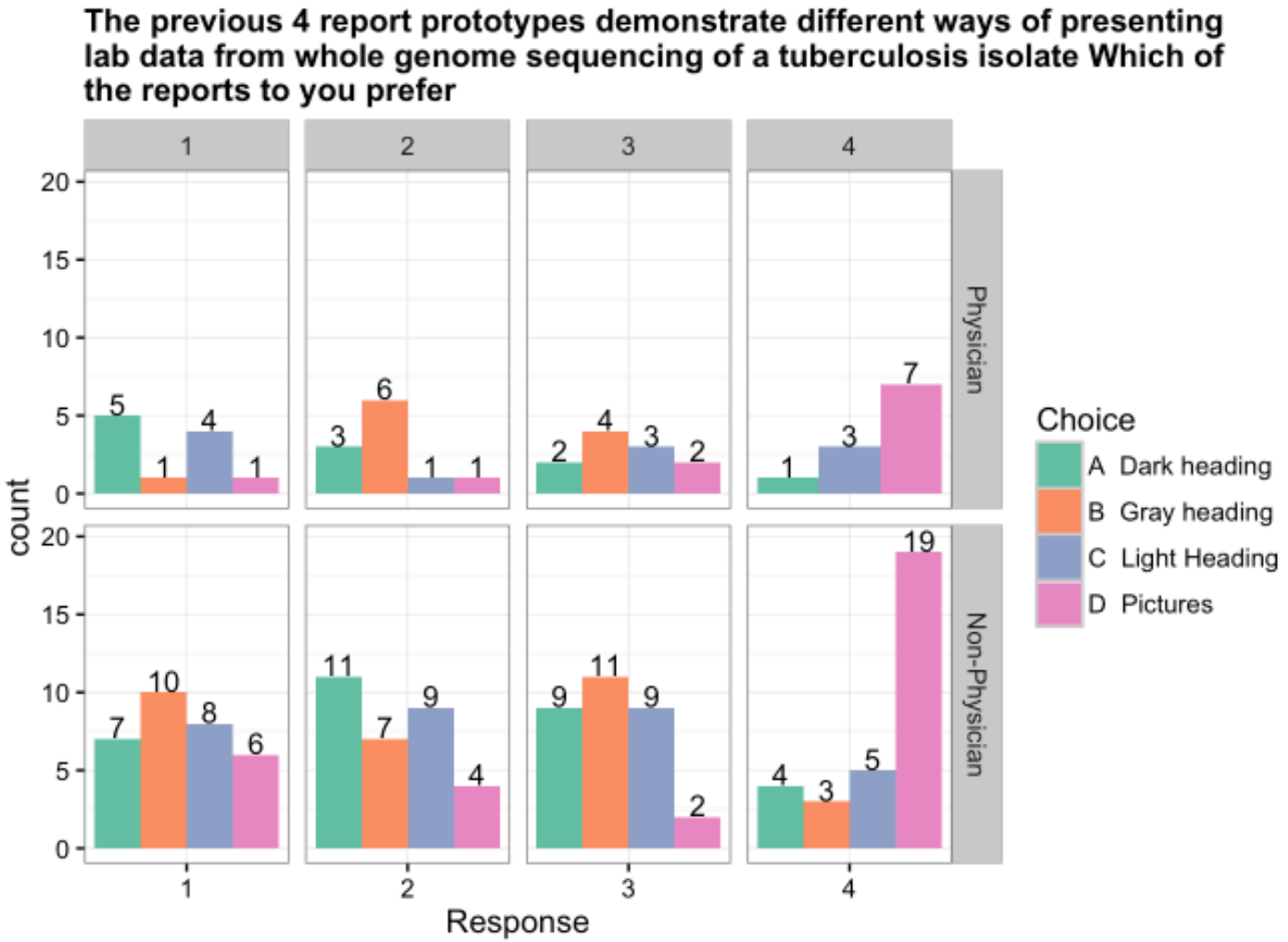






PHASE 3
DESIGN CHOICE QUESTIONNAIRE

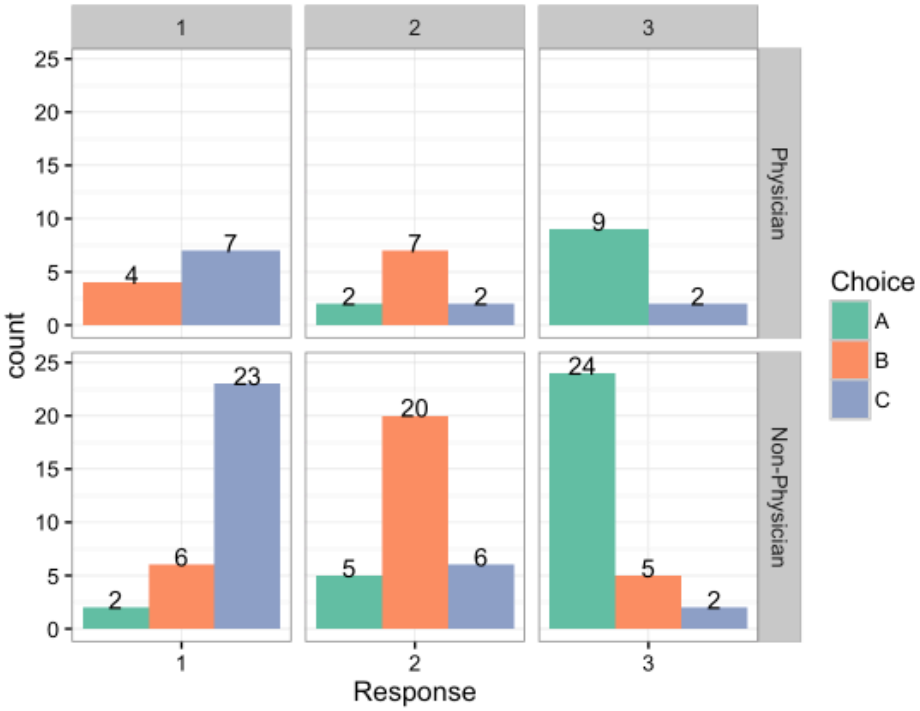
Key finding #1: Comparing whole reports not very useful



PHASE 3
DESIGN CHOICE QUESTIONNAIRE

Key finding #2: Generally strong preference patterns, consistent between clinicians and non-clinicians

Depending on the resistance mutations observed an isolate might be identified as having multidrug resistant TB MDR TB There are many ways this could be noted on the report



A

Drug Susceptibility

Drug	Prediction
Isoniazid	Resistant
Rifampin	Resistant
Ethambutol	Sensitive
Pyrazinimide	Sensitive

B

Drug Susceptibility

Based on predicted antibiotic sensitivities, this individual has **multidrug-resistant (MDR) TB.**

Drug	Prediction
Isoniazid	Resistant
Rifampin	Resistant
Ethambutol	Sensitive
Pyrazinimide	Sensitive

C

Drug Susceptibility

Mono-resistant ☐
Multidrug-resistant (MDR) ☑
Extremely Drug Resistant (XDR) ☐

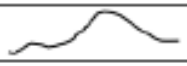
Drug	Prediction
Isoniazid	Resistant
Rifampin	Resistant
Ethambutol	Sensitive
Pyrazinimide	Sensitive

PHASE 3
DESIGN CHOICE QUESTIONNAIRE

Key finding #2: Generally strong preference patterns, consistent between clinicians and non-clinicians

A

Relatedness

Similarity	SNP difference	Cluster trend (past 5 years)	#cases
Highly	0 to 5		10
Peripheral	6 to 12		25

B

Relatedness

	Likely Related (less than 5 SNP Difference)	**Possibly Related** (6-30 SNP Differences)
Number of isolates	2	6

C

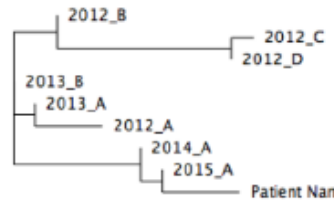
Relatedness

	Likely Related (< 5 SNPs)	**Possibly Related** (6-30 SNPs)
Number of isolates	2	6

D

Relatedness

	Likely Related (less than 5 SNP Difference)	**Possibly Related** (6-30 SNP Differences)
Number of isolates	2	6



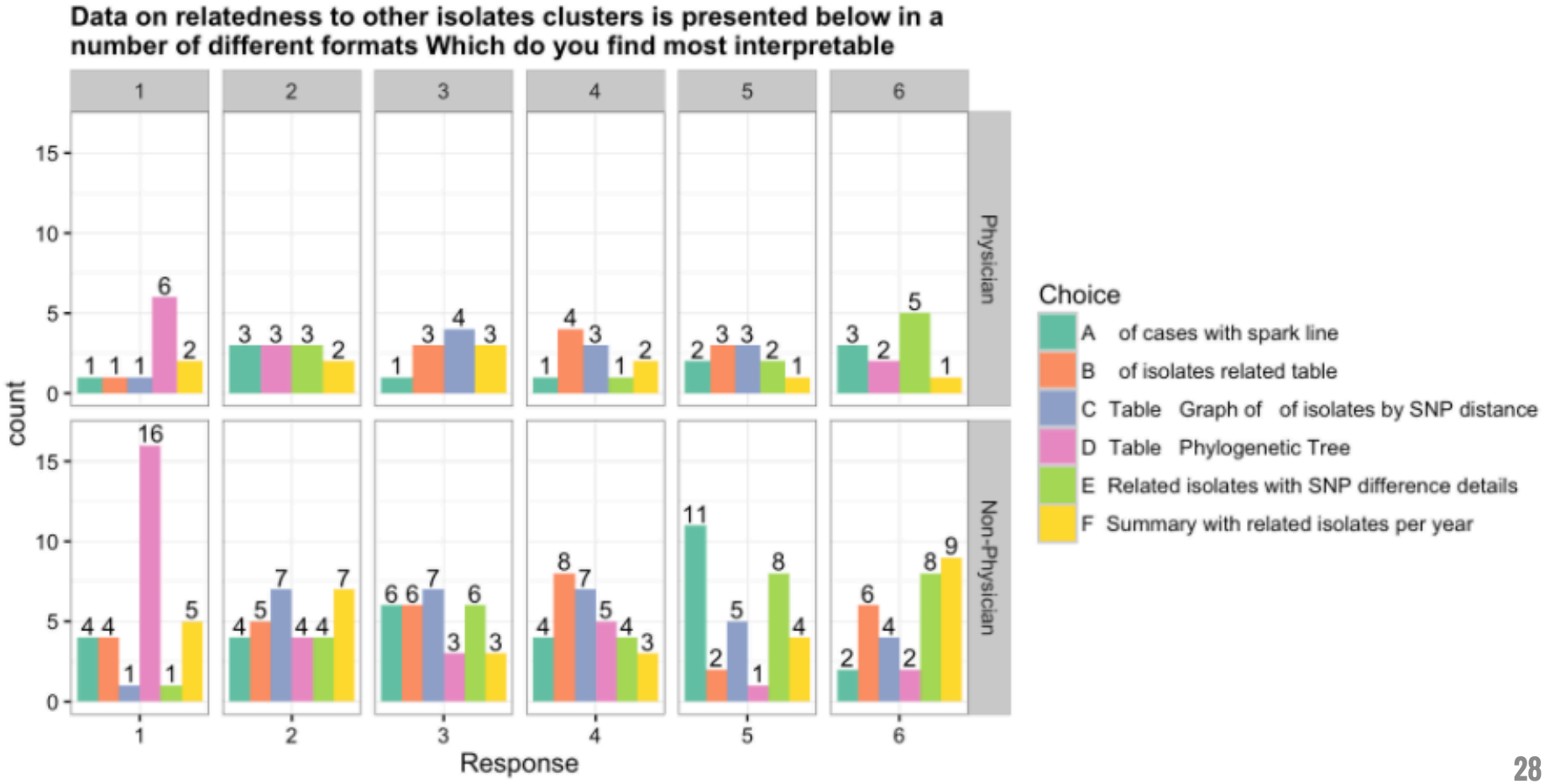
E

Relatedness

Isolate Name	SNP difference
2015_A	3
2014_A	4
2013_A	8
2013_B	7
2012_A	10
2012_B	9
2012_C	10
2012_D	9

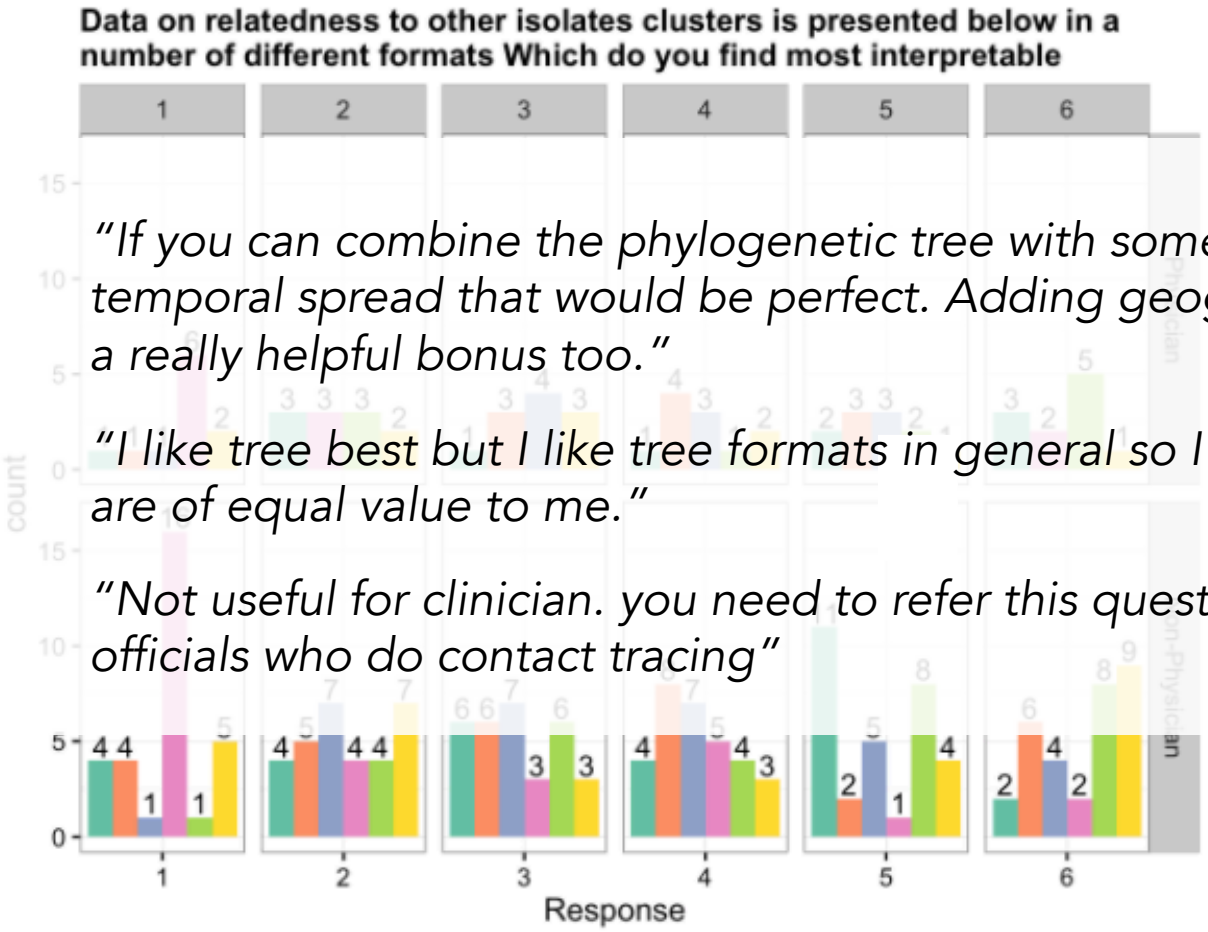
PHASE 3
DESIGN CHOICE QUESTIONNAIRE

Key finding #2: Generally strong preference patterns, consistent between clinicians and non-clinicians



PHASE 3
DESIGN CHOICE QUESTIONNAIRE

Key finding #2: Generally strong preference patterns, consistent between clinicians and non-clinicians



"If you can combine the phylogenetic tree with some kind of graph showing temporal spread that would be perfect. Adding geographical data would be a really helpful bonus too."

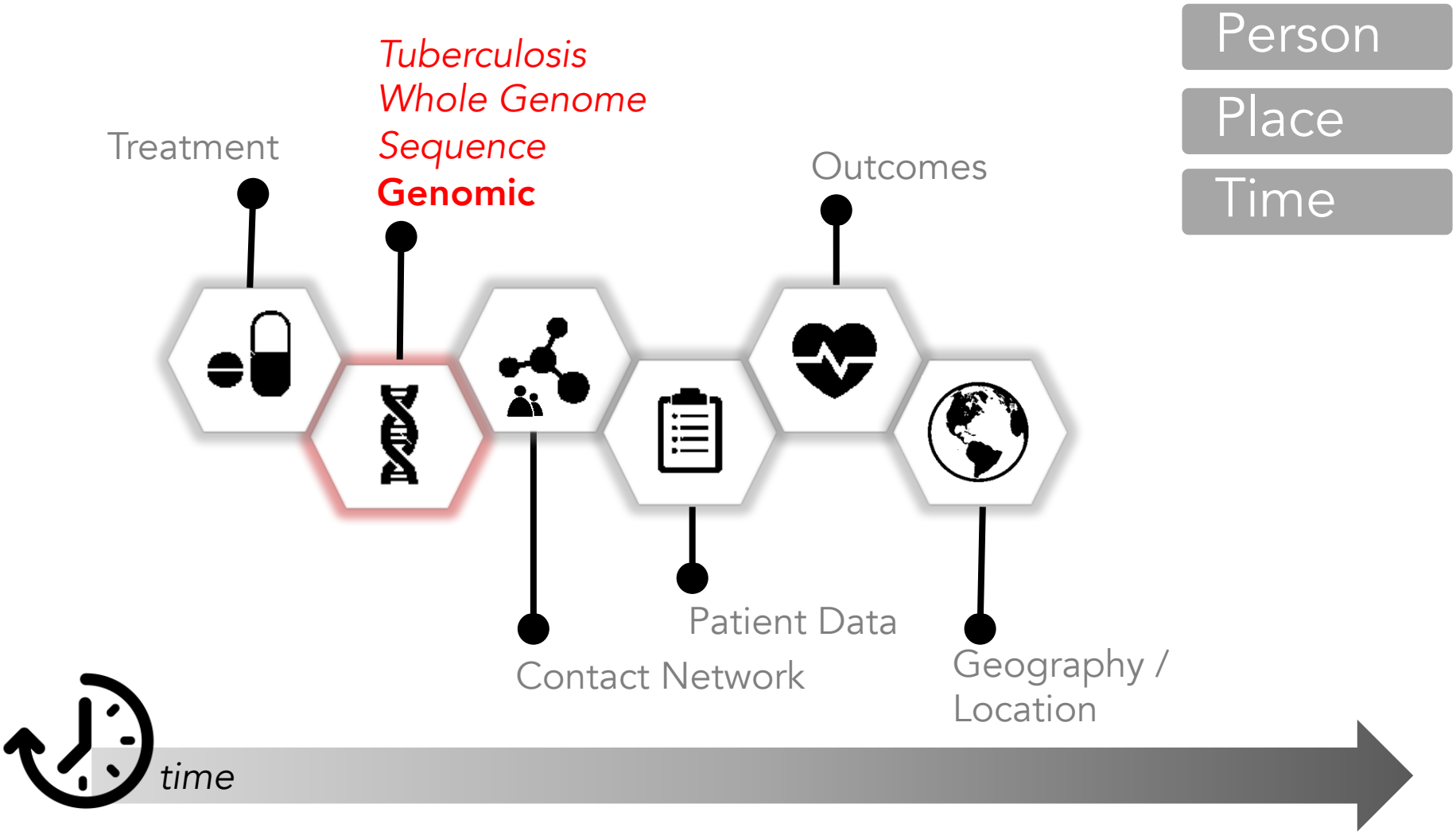
"I like tree best but I like tree formats in general so I am biased. C;A and F are of equal value to me."

"Not useful for clinician. you need to refer this question to public health officials who do contact tracing"

Problem & task data will be used to construct more complex visualizations in future*

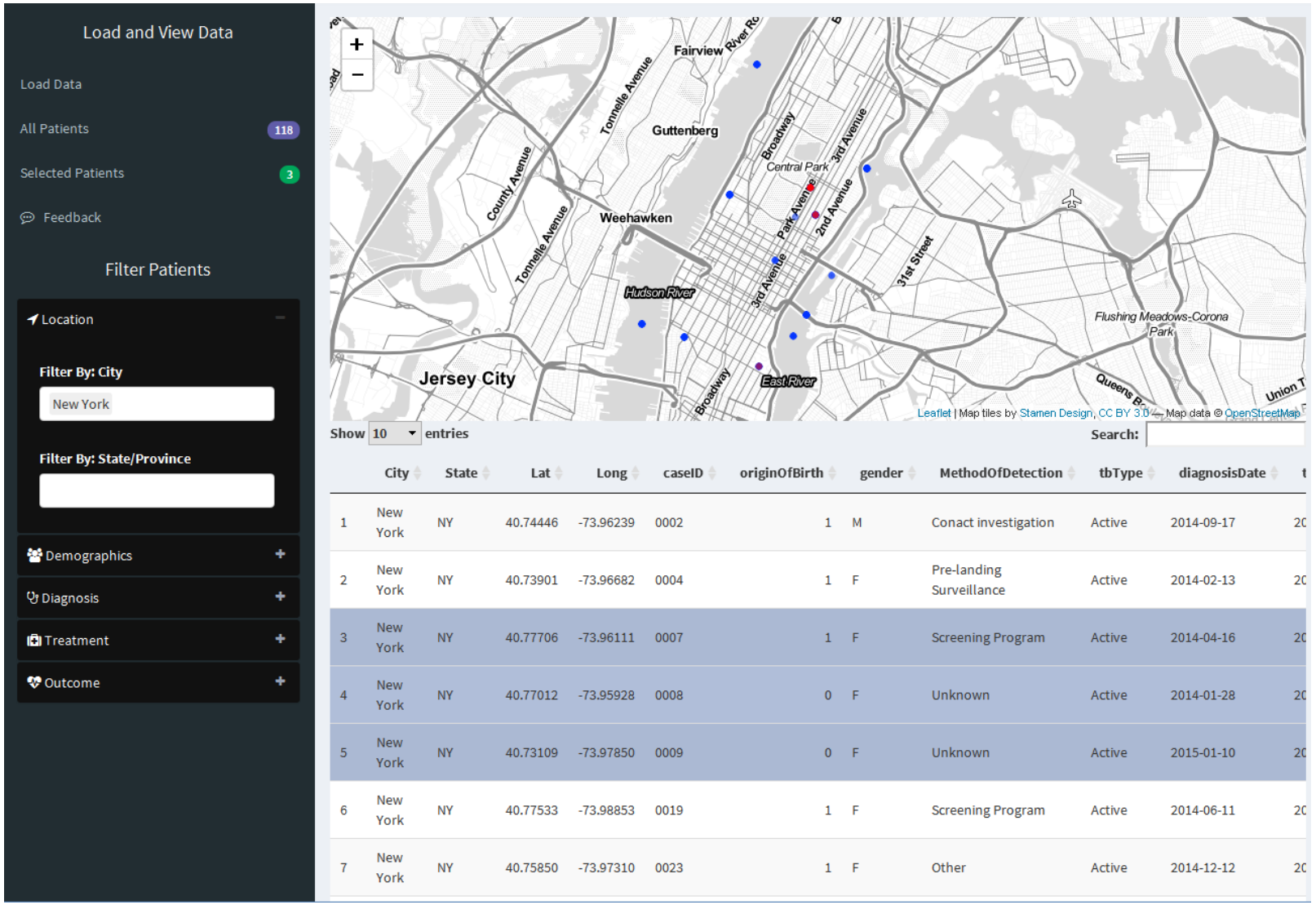
*like my PhD work

WHERE IS MY WORK HEADED?



EpiCOGS

https://amcrisan.shinyapps.io/EpiCOGSDEMO/



	City	State	Lat	Long	caseID	originOfBirth	gender	MethodOfDetection	tbType	diagnosisDate	t
1	New York	NY	40.74446	-73.96239	0002	1	M	Conact investigation	Active	2014-09-17	20
2	New York	NY	40.73901	-73.96682	0004	1	F	Pre-landing Surveillance	Active	2014-02-13	20
3	New York	NY	40.77706	-73.96111	0007	1	F	Screening Program	Active	2014-04-16	20
4	New York	NY	40.77012	-73.95928	0008	0	F	Unknown	Active	2014-01-28	20
5	New York	NY	40.73109	-73.97850	0009	0	F	Unknown	Active	2015-01-10	20
6	New York	NY	40.77533	-73.98853	0019	1	F	Screening Program	Active	2014-06-11	20
7	New York	NY	40.75850	-73.97310	0023	1	F	Other	Active	2014-12-12	20

DECOMPOSING VIS TO TWO LEVELS

PROBLEM & TASK BASED DESIGN

Working with stakeholders to solve relevant problems & provide workable solutions

ABSTRACTIONS & VISUAL ENCODINGS

Common terminology to describe & compare visualizations

IN CONCLUSION

- Data visualization can support decision making in diverse stakeholder groups
- Visual design, not just presence of visualization, matters
- Visualization is a research process in design
- Consider and evaluate alternative choices
- Stay tuned for future developments!

Contact Info


@amcrisan


http://cs.ubc.ca/~acrisan


acrisan@cs.ubc.ca

Thanks

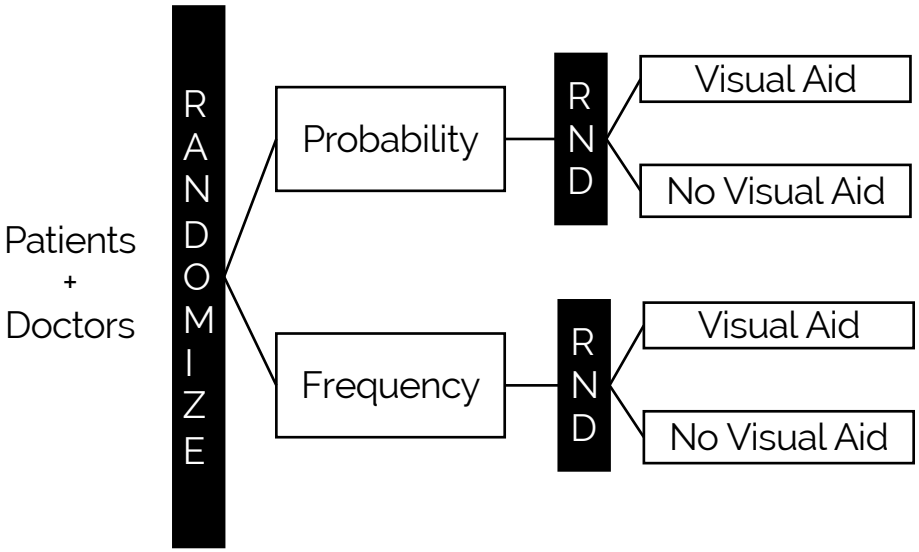
Dr. James Johnston, Dr. Maureen Mayhew, Dr. Victoria Cook, Nash Dahlla, Dr. Jason Wong, Dr. James Brooks, Johnathan Spence, Laura MacDougall, Michael Coss, Ciaran Aiken, and David Roth, Matthew Brehmer, Madison Elliott, Zipeng Liu, Dylan Dong, and Kimberly Dextras-Romagnino

Additional Slides

EXAMPLE : SHARED DECISION MAKING

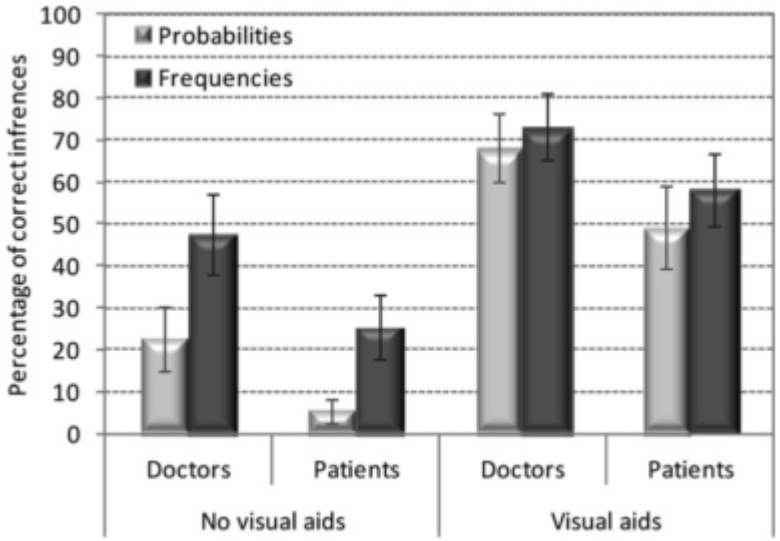
STUDY DESIGN

Quasi-randomized trial with four conditions
Outcome : correctly calculating the risk (essentially a math test)



RESULTS

Visualization improved comprehension of both doctors and patients
Visualization improved concordance between doctors and patients



Garcia-Retamero *et. al* (2013) "Visual representation of statistical information improves diagnostic inferences in doctors and their patients"

DECOMPOSING VIS TO TWO LEVELS

PROBLEM & TASK BASED DESIGN

Working with stakeholders to solve relevant problems & provide workable solutions

ABSTRACTIONS & VISUAL ENCODINGS

Common terminology to describe & compare visualizations

PROBLEM & TASK BASED DESIGN

Why is data being visualized?
Different stakeholders have different needs!

Example

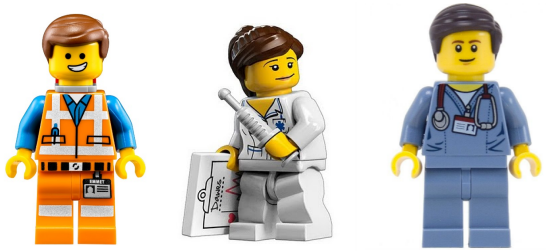
Problem: Understanding Disease Dynamics

Tasks:

"How is a pathogen changing over time?" *"Are there clusters of disease?"*

?

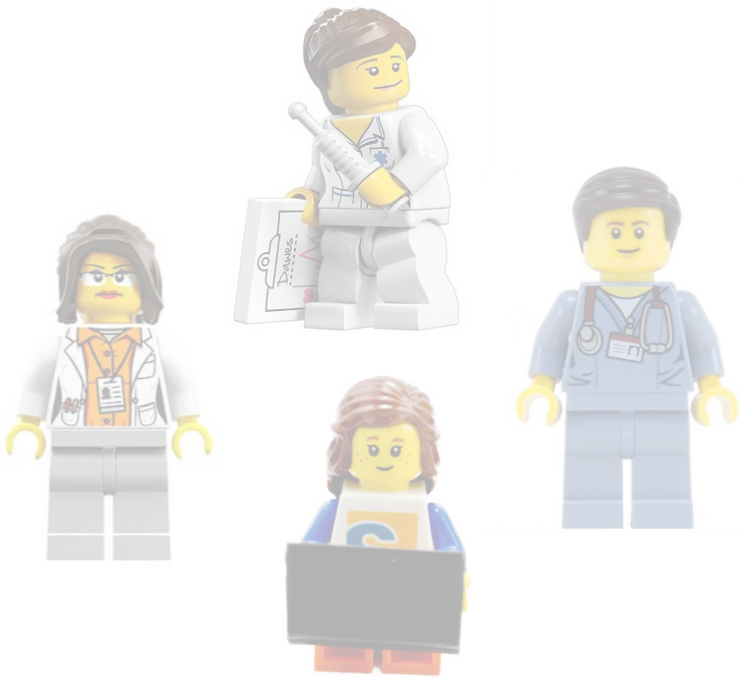
?



DECOMPOSING VIS TO TWO LEVELS

PROBLEM & TASK BASED DESIGN

Working with stakeholders to solve relevant problems & provide workable solutions

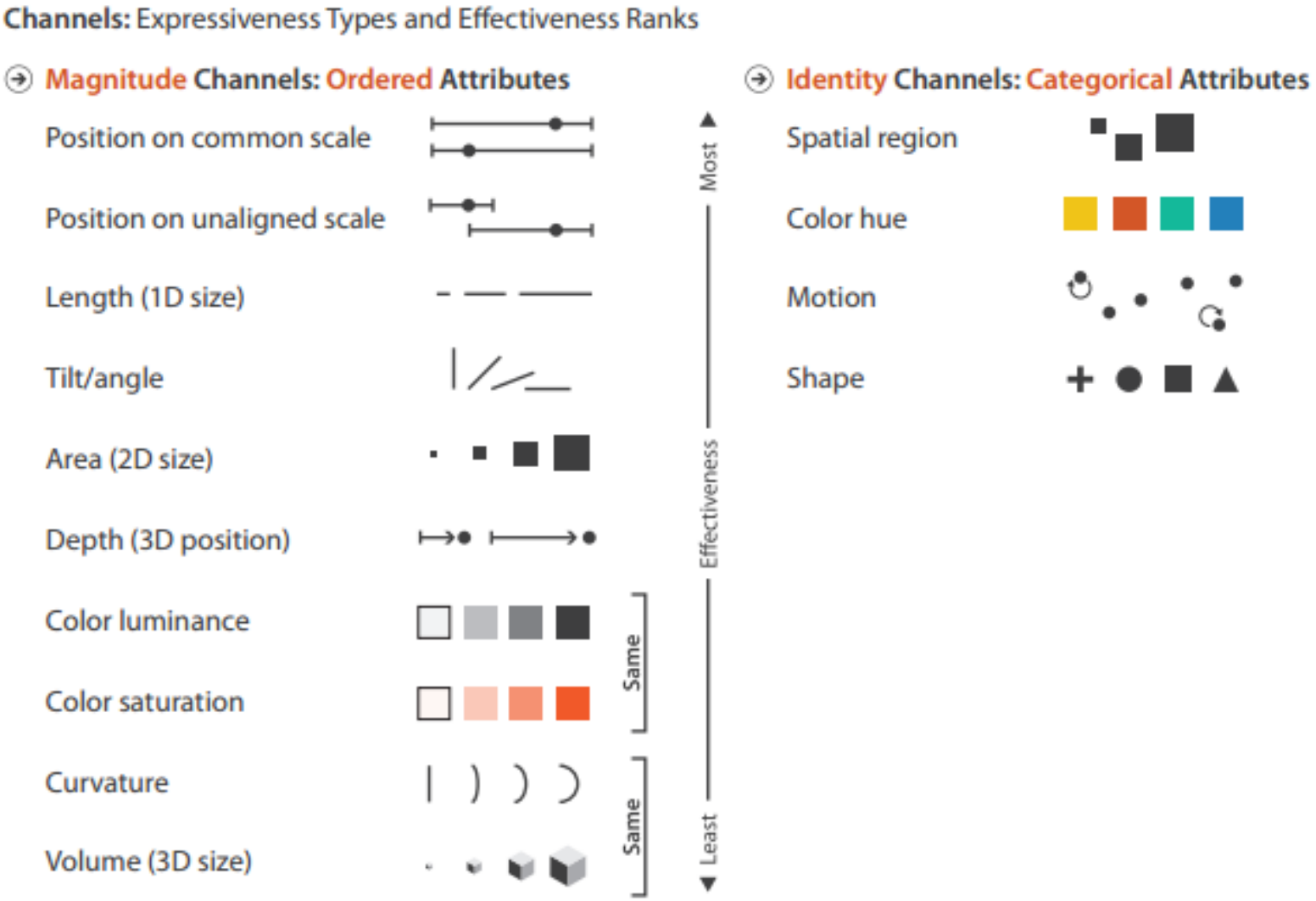


ABSTRACTIONS & VISUAL ENCODINGS

Common terminology to describe & compare visualizations

ABSTRACTIONS & VISUAL ENCODINGS

Decomposition Visualizations into geometric shapes & properties



Munzner (2014) “Visualization Analysis and Design”

DESCRIBING VISUALIZATIONS

Using geometric marks and their properties (channels)

Vertical Position

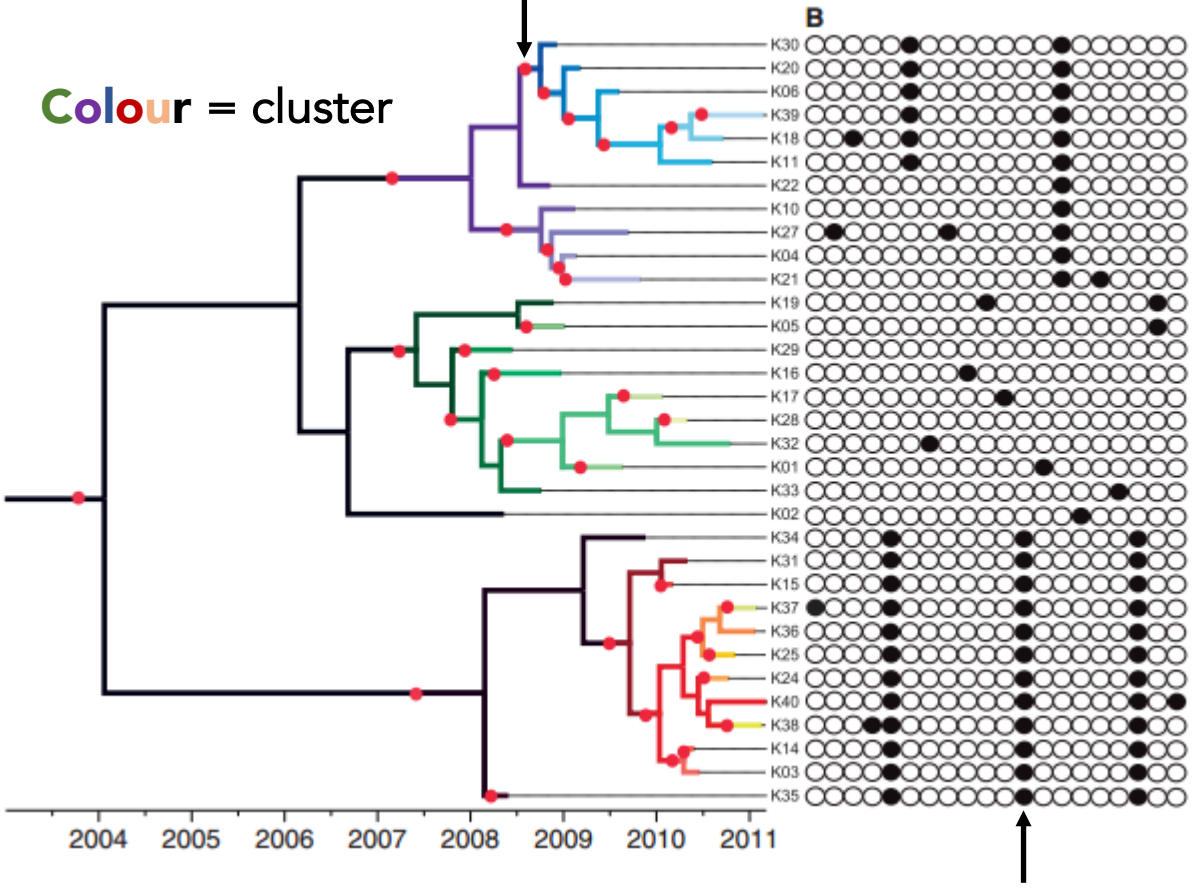
Vertical Position
Horizontal Position

Vertical Position
Horizontal Position
Colour

Vertical Position
Horizontal Position
Colour
Size

DESCRIBING DISEASE DYNAMICS

EXAMPLE 1: PHYLOGENETIC TREE + DOT PLOT



Vertical Pos.
Case Similarity

Colour + Dot
← SNP Absent
← SNP Present

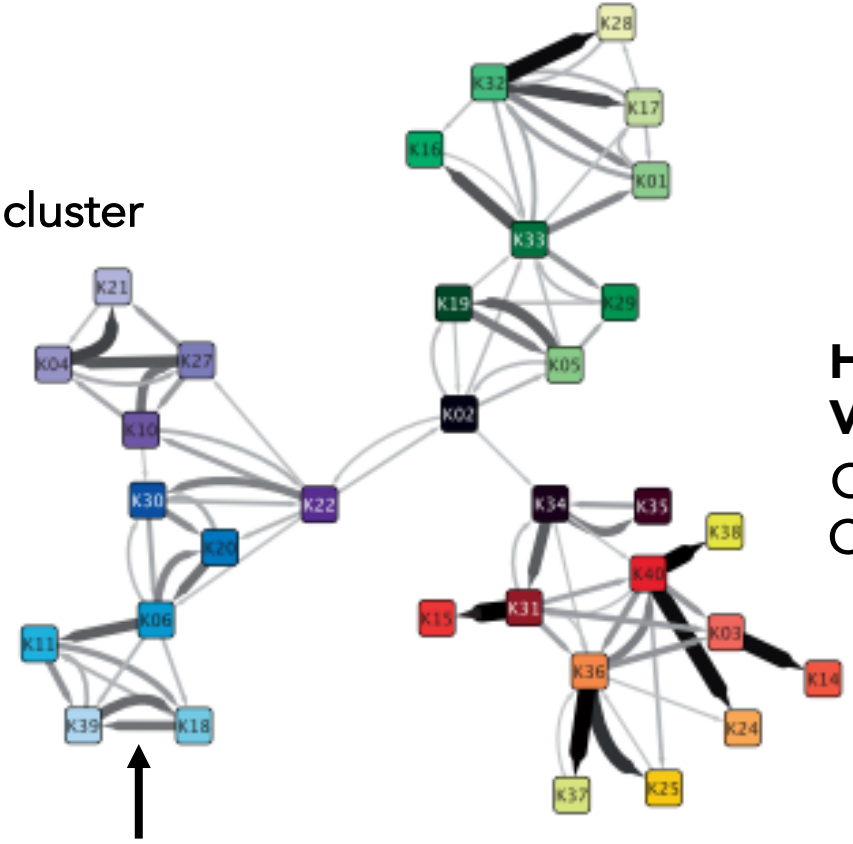
← **Horizontal Position** →
Transmission Timing

← **Horizontal Pos.** →
Genetic Similarity

DESCRIBING DISEASE DYNAMICS

EXAMPLE 2: NETWORK DIAGRAM

Colour = cluster



Horizontal + Vertical Position

Chain of transmissions
Case Similarity

Thickness of line

Probability of transmissions

DESCRIBING DISEASE DYNAMIC

COMPARING VISUALIZATIONS

What information does the visualization show?
How does the visualization show that information?

WHAT	HOW	TREE	NETWORK
Transmission Timing	Horizontal pos.	■	
	Colored Dot	■	
Transmission Confidence	Thickness		■
	Color		■
Case Similarity	Horizontal + Vertical pos.	■	■
	Black/White Dot	■	
	Colored Shape	■ (line)	■ (square)
SNP presence	Black/White Dot	■	

LINKING VIS ABSTRACTIONS TO BIOINFORMATIC ONTOLOGIES

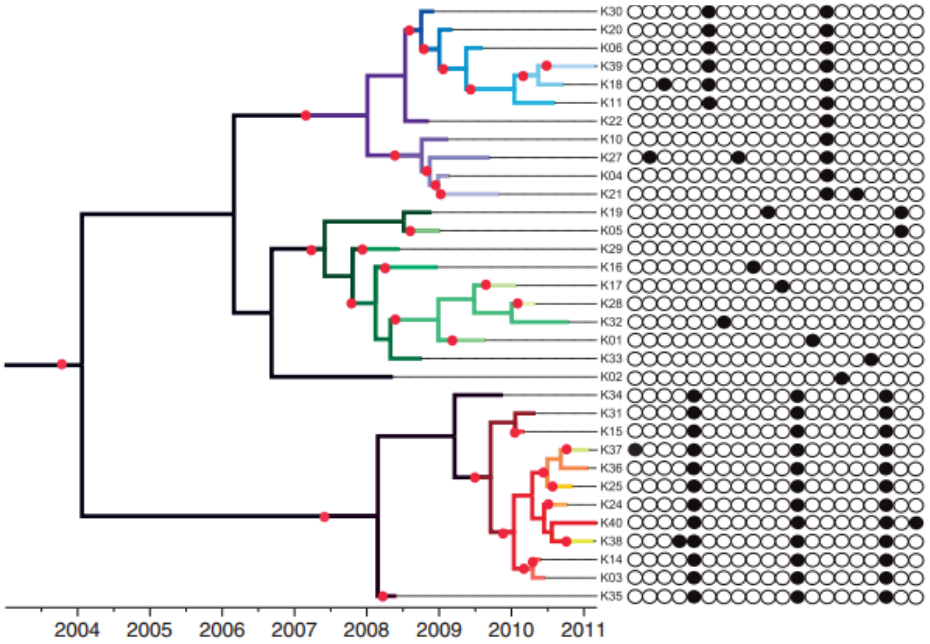
- Can connect ontologies to visualizations through abstractions
- *Suggest* visualizations based on available data
- Need to know *what kinds* of visualizations are suitable for different research questions
  - Currently working on this ☝️
  - Similar idea to www.vizhealth.org

PROBLEM & TASK BASED DESIGN

Why is data being visualized?

"How is a pathogen changing over time?"

Transmission timing & genetic similarity



"Are there clusters of disease?"

Transmission between & within clusters of related cases