Visualizing Public Health Data

Anamaria Crisan, MSc
PhD Candidate
University of British Columbia


@amcrisan


acrisan@cs.ubc.ca


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

I will attempt to make two points

- The differences between clinical medicine and public health and the vis implications

- Provide an overview of the state of visualization within a public health domain

Public Health and Clinical Medicine

Clinical Medicine vs. Public Health

Clinical Medicine

- Targets individual patients

Public Health

- Targets populations

Clinical Medicine vs. Public Health

Clinical Medicine

- Targets individual patients
- Diagnosis and treatment focused

Public Health

- Targets populations
- Prevention and control focused

Clinical Medicine vs. Public Health

Clinical Medicine

- Targets individual patients
- Diagnosis and treatment focused
- Interventions are typically pharmaceutical interventions

Public Health

- Targets populations
- Prevention and control focused
- Pharmaceutical, but also more commonly environmental and behavioral interventions

Clinical Medicine vs. Public Health

Clinical Medicine

- Targets individual patients
- Diagnosis and treatment focused
- Interventions are typically pharmaceutical interventions
- Decision makers typically siloed specialists (doctors, nurses, etc.)

Public Health

- Targets populations
- Prevention and control focused
- Pharmaceutical, but also more commonly environmental and behavioral interventions
- Decision makers diverse, not necessarily specialists (community leaders, etc.)

Clinical Medicine vs. Public Health

Clinical Medicine

- Targets individual patients
- Diagnosis and treatment focused
- Interventions are typically pharmaceutical interventions
- Decision makers typically siloed specialists (doctors, nurses, etc.)

Public Health

- Targets populations
- Prevention and control focused
- Pharmaceutical, but also more commonly environmental and behavioral interventions
- Decision makers diverse, not necessarily specialists (community leaders, etc.)

Example:

Treating lung cancer patient

Anti-smoking campaign

Both use data, even the same data, in different ways

Visualization consumers in clinical medicine

- Currently data vis tends to emphasize clinical medicine applications and targets clinicians, researchers, and patients

Clinicians

Researchers

Patients

Visualization consumers in public health

- Public Health has much more multidisciplinary decision making teams
  - More data & diverse data types = more informed decision making
  - BUT – different stakeholder abilities to interpret data & different needs

- Gap: few vis applications for public health

Medical Health Officers

Clinicians

Nurses

Researchers

Community Leaders

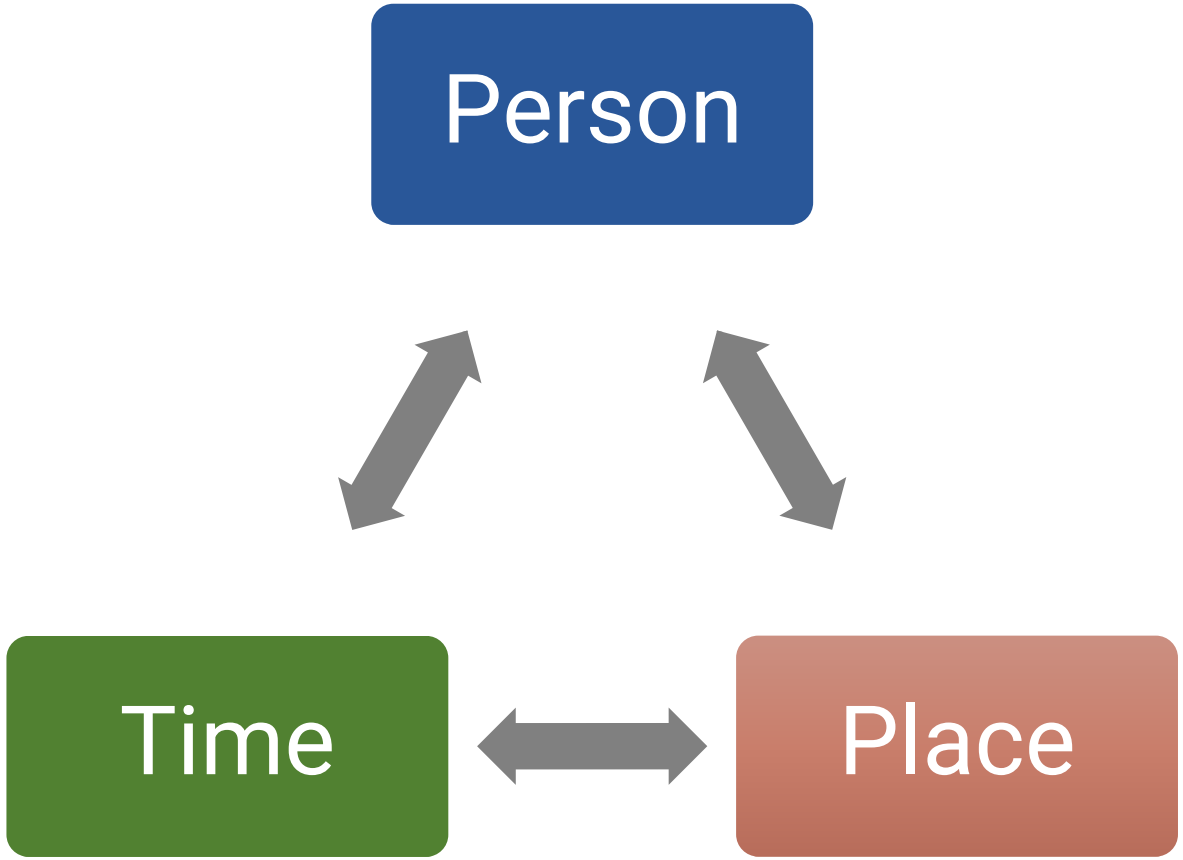
Patients

Politicians

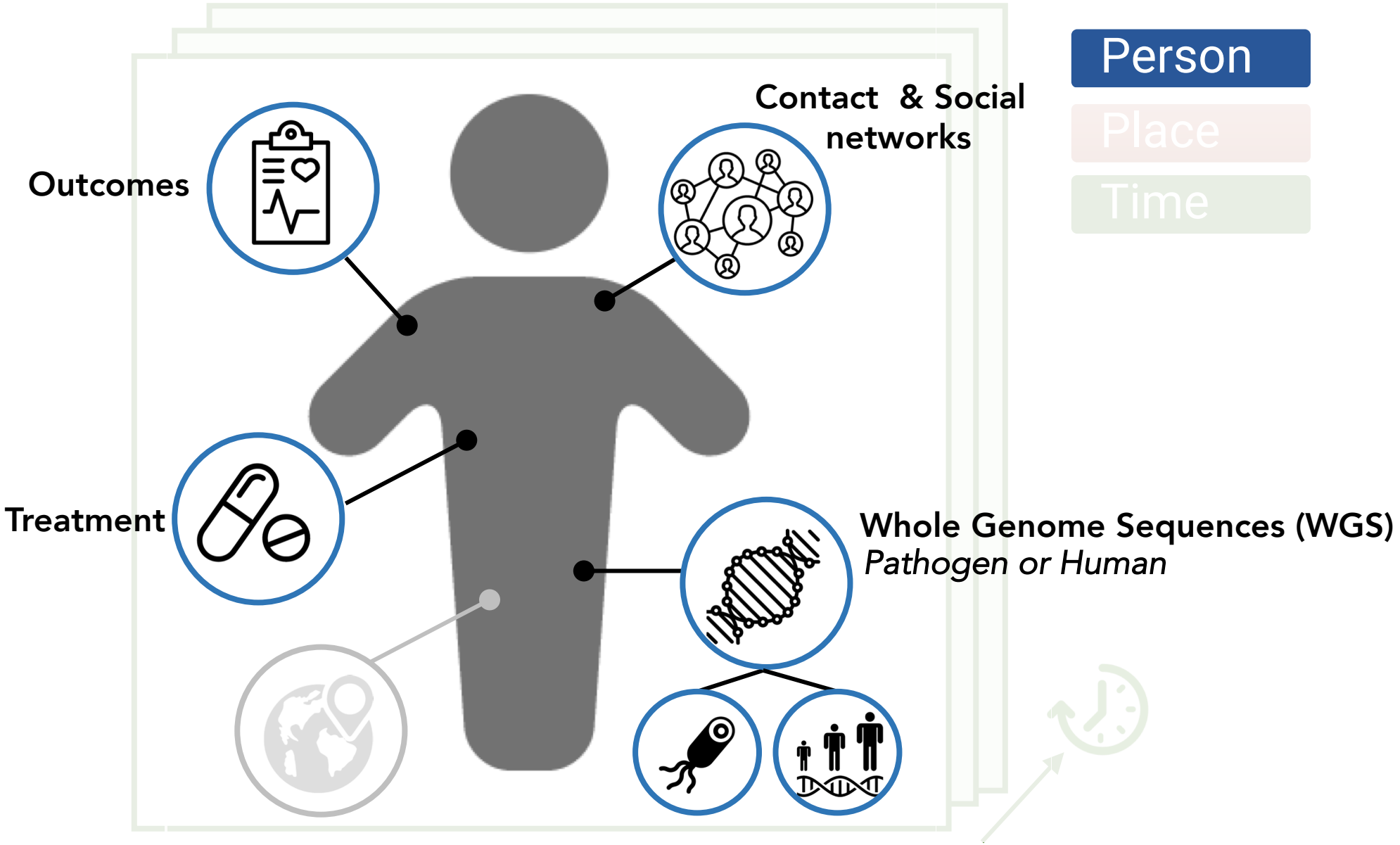
What are Public Health Data?

What are Public Health Data?

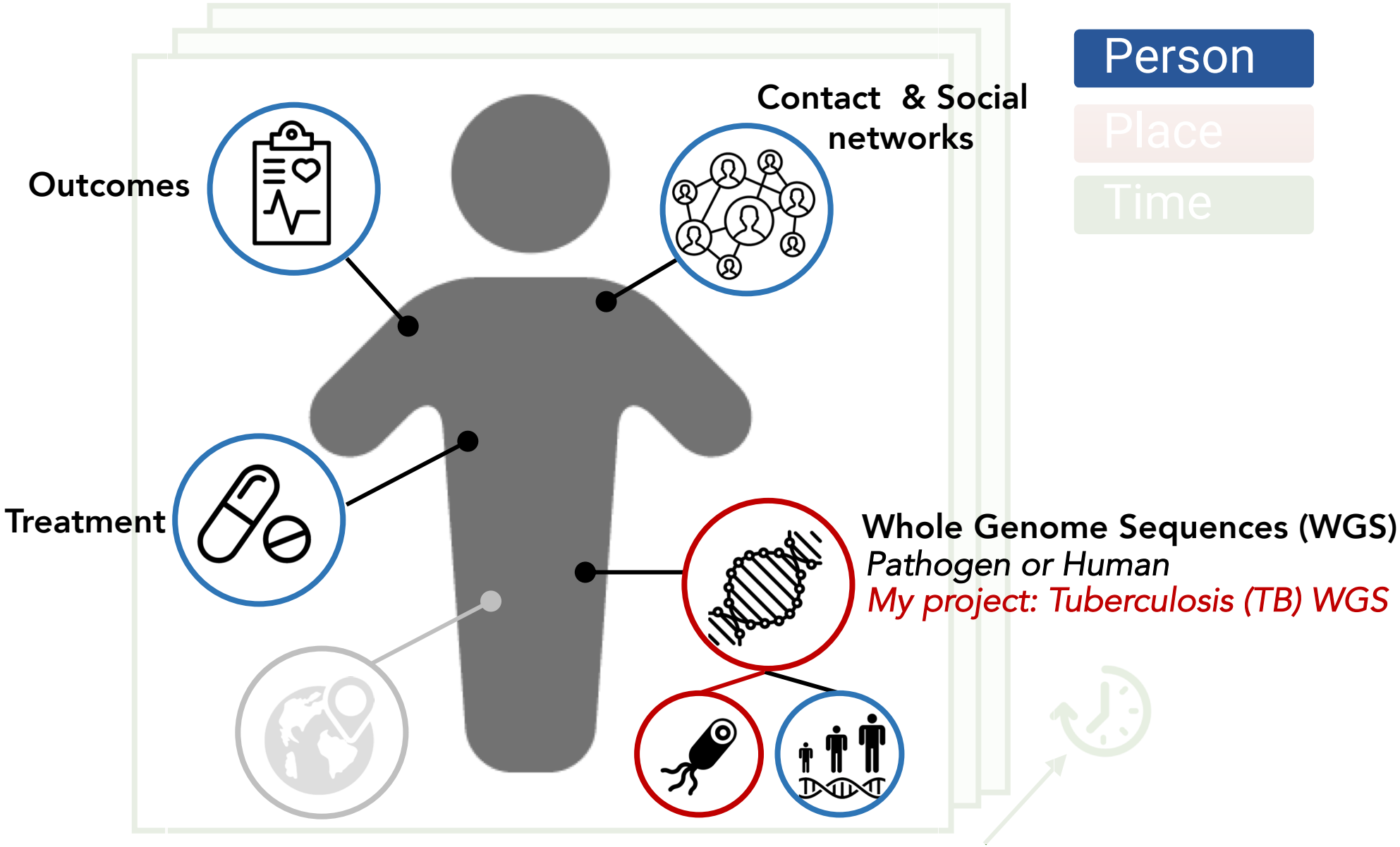
The Epidemiological Trinity



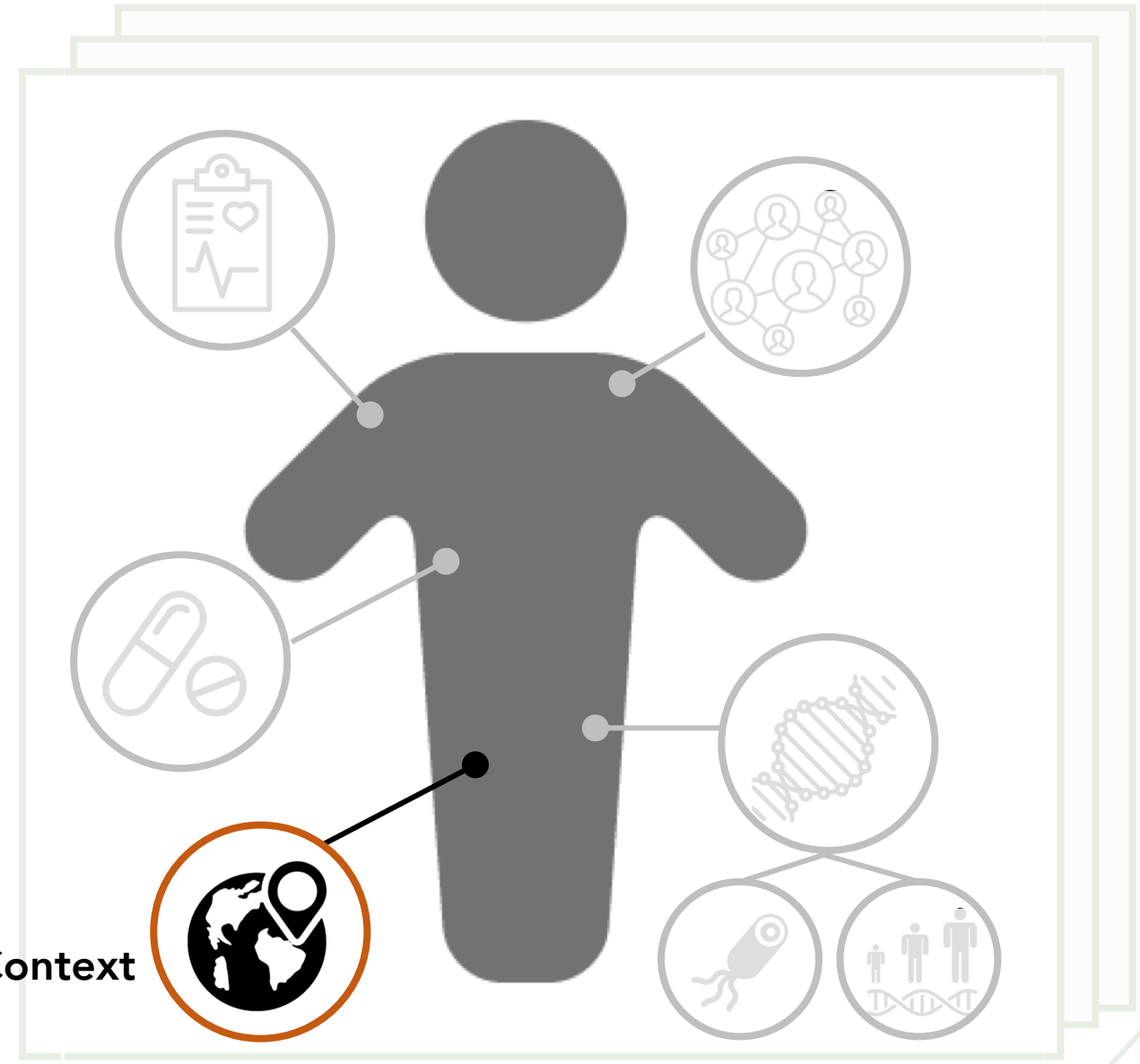
What are Public Health Data?



What are Public Health Data?



What are Public Health Data?



Person

Place

Time

Location
Geographic Context

What are Public Health Data?

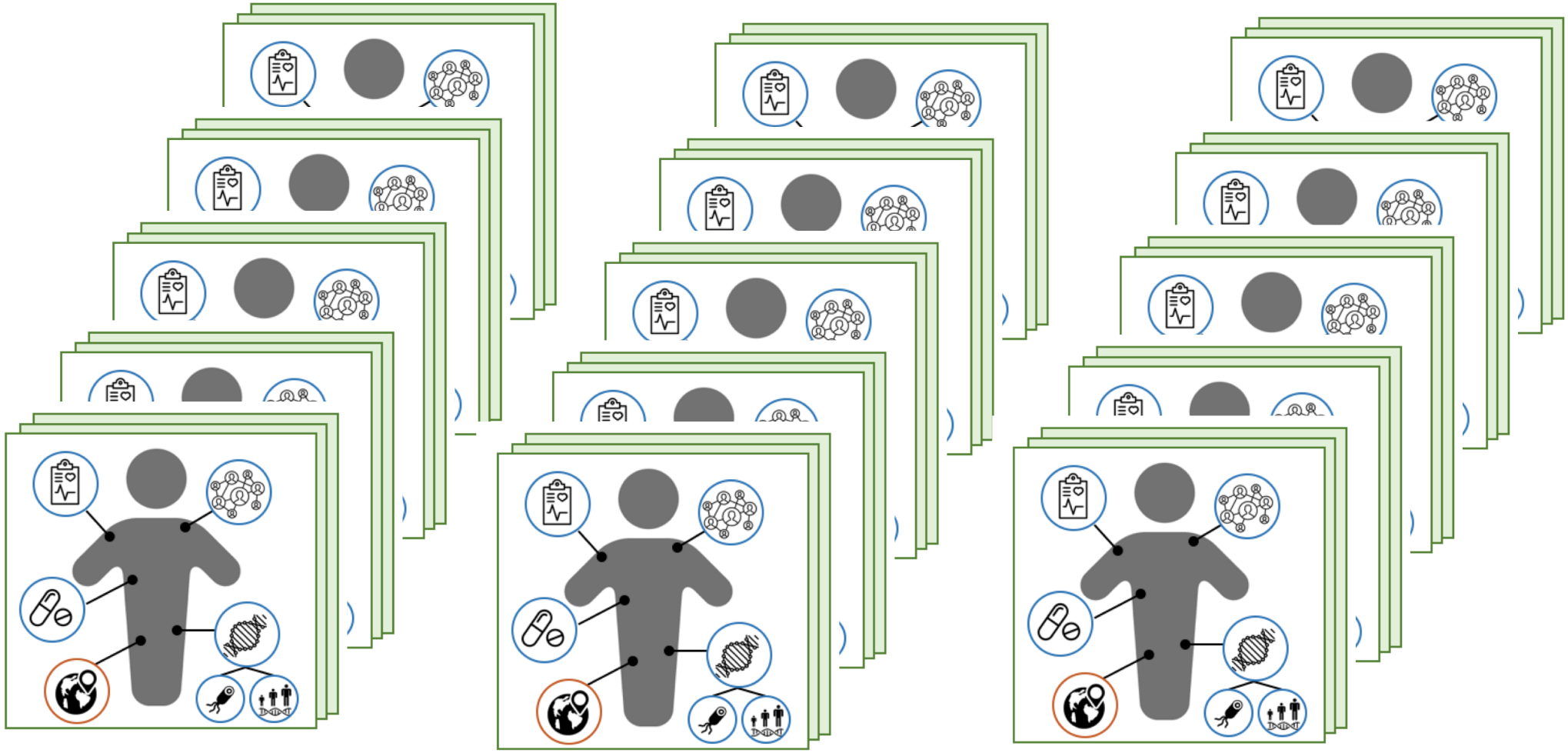
Person

Place

Time

What are Public Health Data?

Via EHRs data are passively collected about entire populations over time



The State of Data Vis in Public Health

The state of visualization in public health

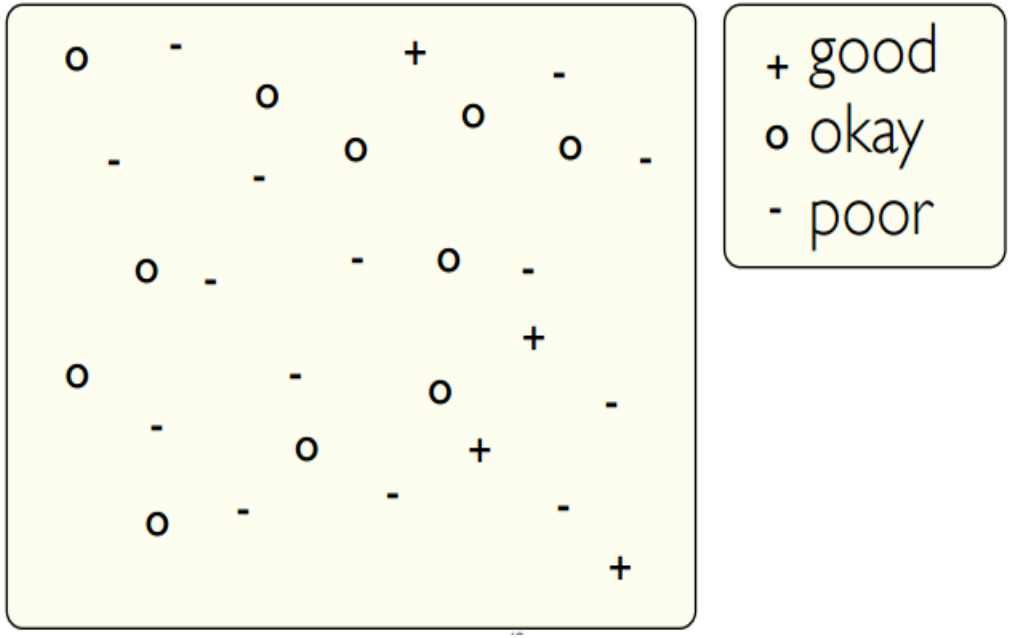
- Barriers for creating data visualizations are lowering
  - Many domain specialists (scientists, public servants) routinely create data visualizations
- Guidance on what makes a good data visualization is absent
  - Domain specialists don't read the vis literature
- Lack of guidance = inefficient unsupervised exploration of vis design space
  - "Hit or Miss" ad hoc design solutions

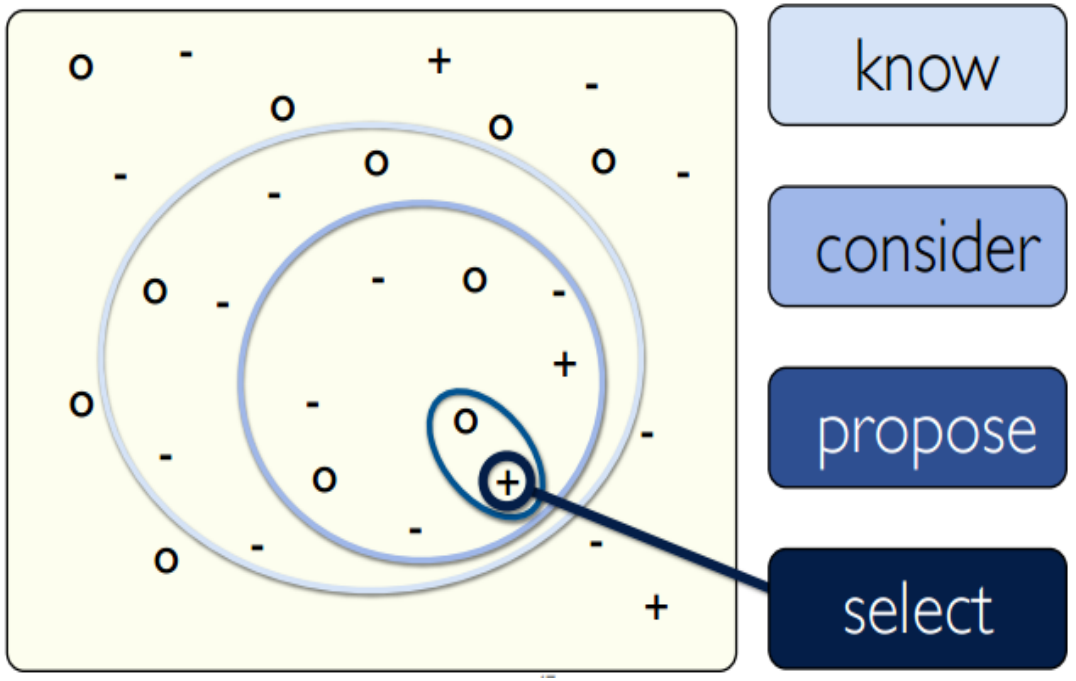
The state of visualization in public health

- Barriers for creating data visualizations are *lowering*
  - Many domain specialists (scientists, public servants) routinely create data visualizations
- Guidance on what makes a good data visualization is *absent*
  - Domain specialists don't read the vis literature
- Lack of guidance = inefficient unsupervised exploration of vis design space
  - "Hit or Miss" ad hoc design solutions
- **Our proposed solution: systematically create an explorable vis design space**

Design Spaces : A quick primer

Design spaces are made of visualization design choices or varying utility (+ 0 -)

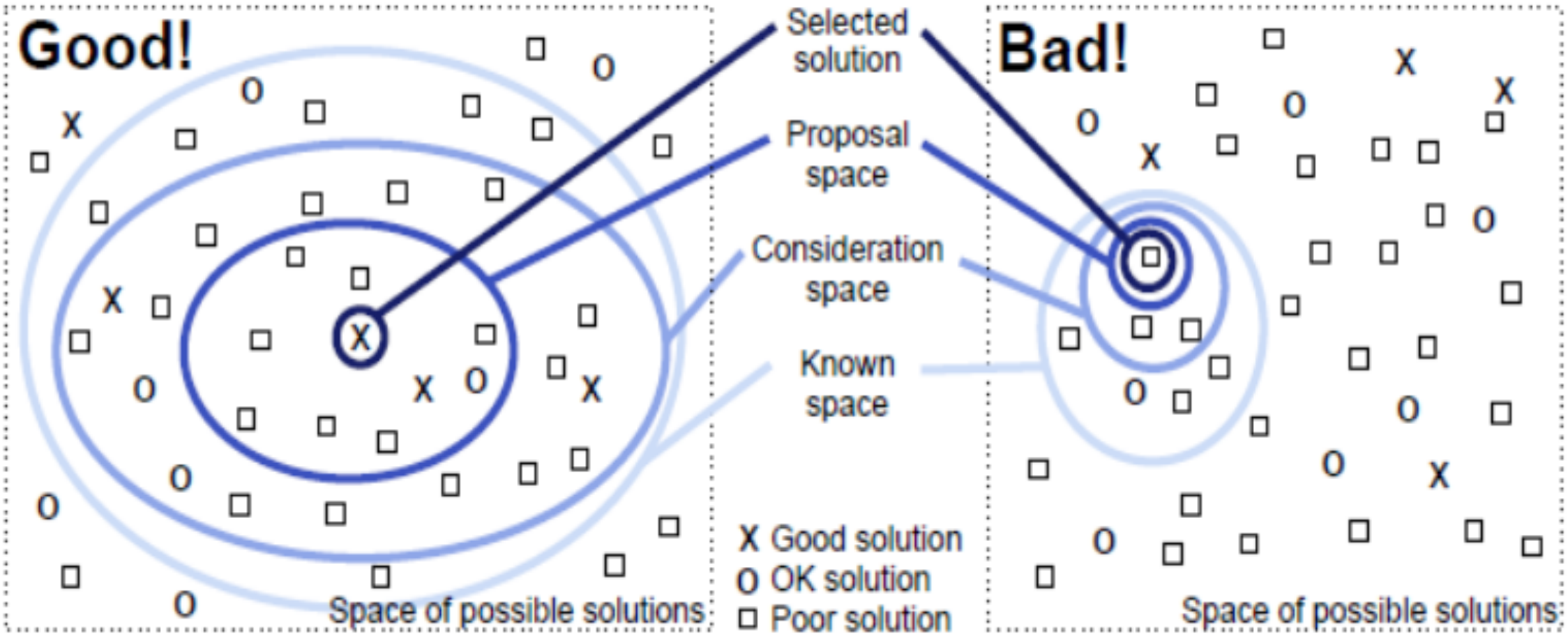




Source: Sedlmair (2012) "Design Study Methodology"

Design Spaces : A quick primer

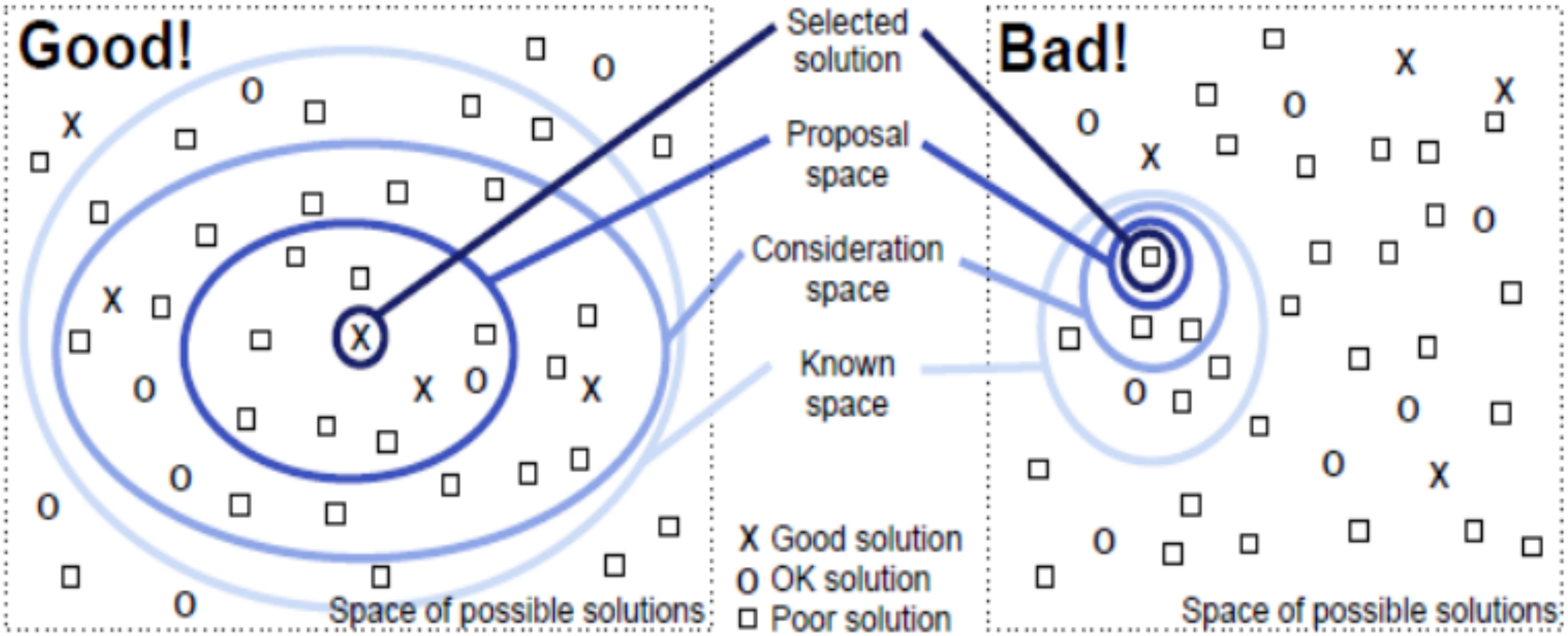
GOAL – nudge domain specialists toward better design choice solutions



Source: Munzner (2014) "Visualization Analysis and Design"

Design Spaces : A quick primer

BUT – how do we **systematically** describe design space to promote good exploration?



Source: Munzner (2014) "Visualization Analysis and Design"

Constructing a design space

- Our observation: there's a lot of figures in research papers, let's study them!
- Challenge: methods for systematic assessment of data visualizations don't exist
  - Systematic matters! Shows the good, the bad, and the common
  - Existing studies (setvis, treevis, vishealth) are not systematic reviews of specialist's domain
- We combined methods from epidemiology with infovis to construct a design space

Our approach allows us to answer three different questions

- Scope: Infectious Disease Genomic Epidemiology literature
- Objective: Identify and enumerate the kinds of visualizations generated for different topics of infectious disease genomic epidemiology

Our approach allows us to answer three different questions

- Scope: Infectious Disease Genomic Epidemiology literature
- Objective: Identify and enumerate the kinds of visualizations generated for different topics of infectious disease genomic epidemiology

Literature Analysis

Qualitative Data Visualization Analysis

Quantitative Data Visualization Analysis

Our approach allows us to answer three different questions

- Scope: Infectious Disease Genomic Epidemiology literature
- Objective: Identify and enumerate the kinds of visualizations generated for different topics of infectious disease genomic epidemiology

Literature Analysis	WHY are researchers visualizing data?
Qualitative Data Visualization Analysis	HOW are researchers visualizing data, WHAT are they visualizing?
Quantitative Data Visualization Analysis	HOW MANY examples are there of specific visualizations?

GEViT Gallery

Use the different filters below to navigate the GEViT Gallery. To get more ifnromation about each filter click the ℹ icon

Paper Lookup (PMID):

Visualization Context

Pathogen:

Topic:

Data (from figure captions):

Visualization Graphical Properties

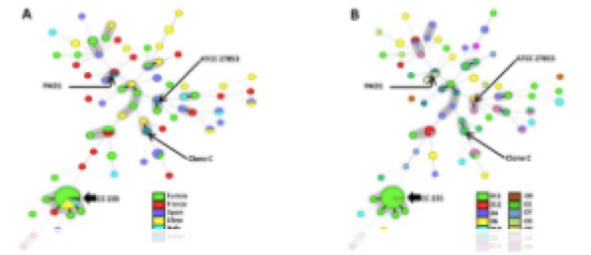
Chart Type

Special Chart Type

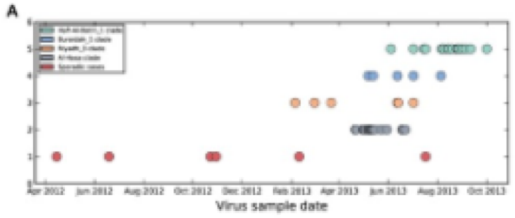
Catalogue | Figure

Only show images with the following tags (select to activate): Missed Opportunity | ✔ Good Practice

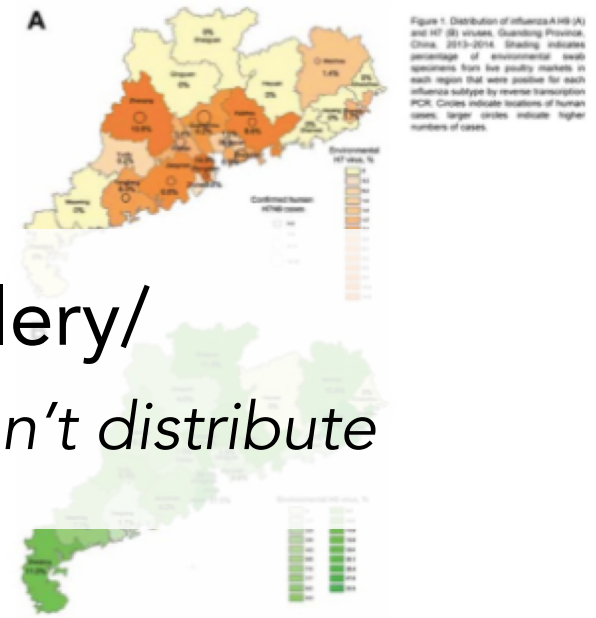
Good Practice



Good Practice



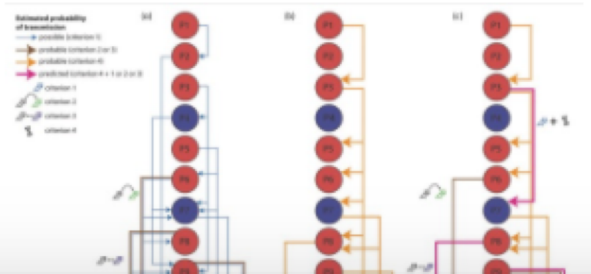
Good Practice



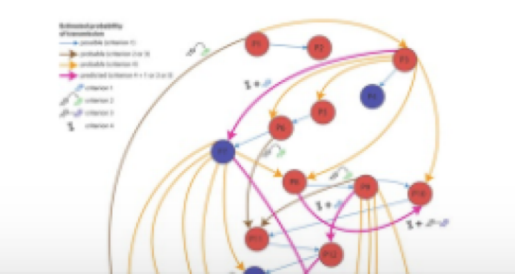
https://amcrisan.shinyapps.io/gevit_gallery/

Unpublished & still some work to be done so please don't distribute

Good Practice



Good Practice



NA

Implications of our findings

- **Surprise finding: a lot of data in data visualizations were not visualized!**
- Pedagogical implications :
  - Can we give people more complex vis applications when their vis skills are kind of low?
  - How can we improve vis literacy?
  - I think a design space is a useful discussion tool

- Software develop implications:
  - Discussion of a design space in bioinformatics development
  - GEViT is resource to provide alternative designs
  - Alternative designs also see gaps in the where vis research is needed

- Human-in-the-loop implications:
  - Need to think beyond image recognition problems
  - Might be premature to apply AI methods (no good training data)

Additional Slides

An overview of our results so far

- Literature Analysis: Understanding the structure of genomic epidemiology papers promotes systematicity via intelligent sampling
  - Total sample ~18K papers on genomic epidemiology
  - Defined strata by pathogen (document structure) and a priori concepts (domain knowledge)
  - Literature analysis stratified sampling yielded ~850 figures for analysis from 221 papers
- Qualitative Analysis: Developed GEViT, a Genomic Epidemiology Visualization Typology
  - Developed a typology to systematically described charts using three descriptive axes: chart types, chart combinations, and chart enhancements
- Quantitative Analysis: It's nearly all phylogenetic trees, across all pathogens and concepts, but there's also a lot of tables and plain text
- Surprising general conclusion: most data is these data visualizations are not visualized

An overview of our approach

“Identify and enumerate the kinds of visualizations generated per topic of i.d. genomic epidemiology”

An overview of our approach

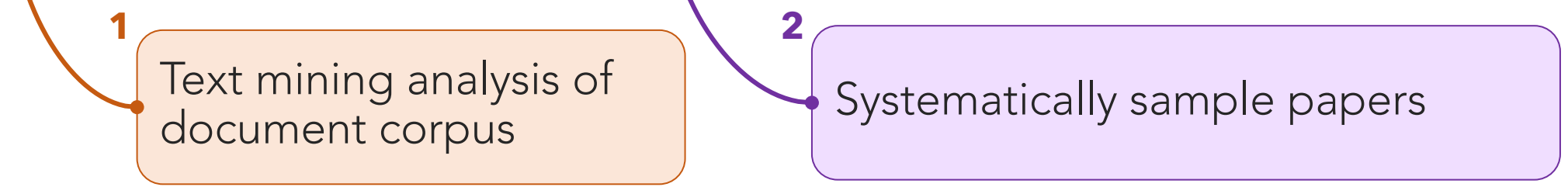
"Identify and enumerate the kinds of visualizations generated per topic of i.d. genomic epidemiology"

1

Text mining analysis of document corpus

An overview of our approach

"Identify and enumerate the kinds of visualizations generated per topic of i.d. genomic epidemiology"



An overview of our approach

3

Create a codeset to classify research figures consistently

"Identify and enumerate the kinds of visualizations generated per topic of i.d. genomic epidemiology"

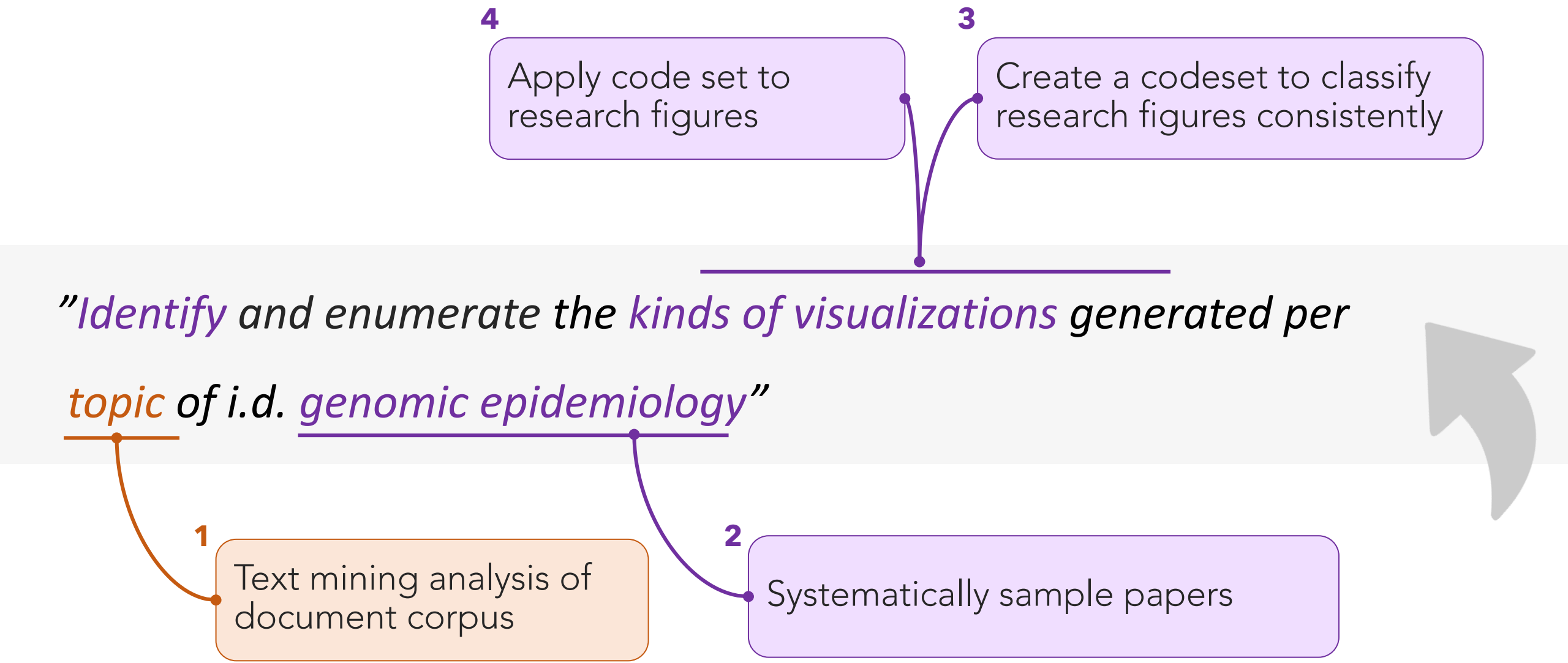
1

Text mining analysis of document corpus

2

Systematically sample papers

An overview of our approach



An overview of our approach

5 Descriptive statistics (literally count)

4 Apply code set to research figures

3 Create a codeset to classify research figures consistently

"Identify and enumerate the kinds of visualizations generated per topic of i.d. genomic epidemiology"

1 Text mining analysis of document corpus

2 Systematically sample papers