

Multivariate Networks in *Ecology and Evolution* Design Study

Jen Rogers

5. 17. 2019



Photo by [Edi Libedinsky](#) on [Unsplash](#)

***My name is Jen,
I am a PhD student at the University of Utah
I build visual tools for scientists***

Cohort Control

+ 1 clear all

Demographic Filters

Score Filters

search the web

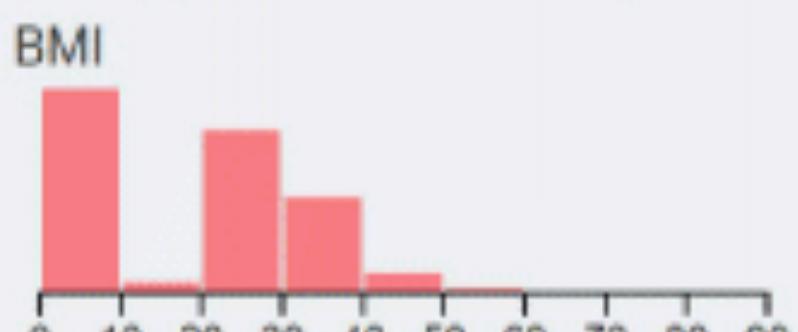
Cohort-1

Cohort-2

Cohort-3 +  X

Cohort-4

Page 1



Cohort-1 Filter Layers

Cohort Size: 165

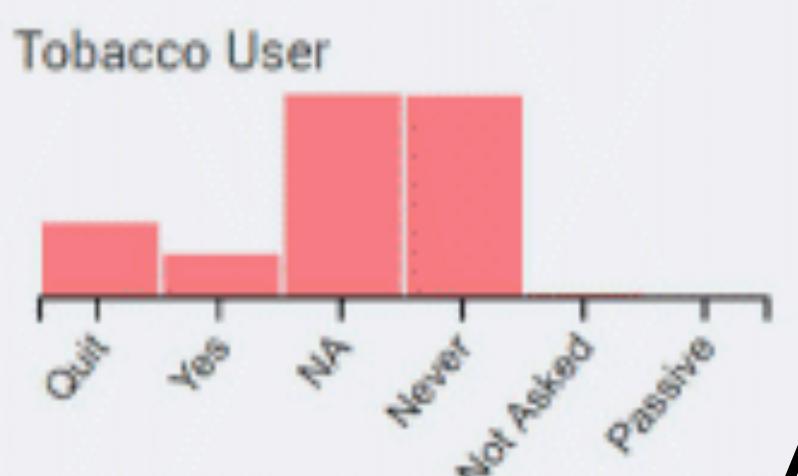
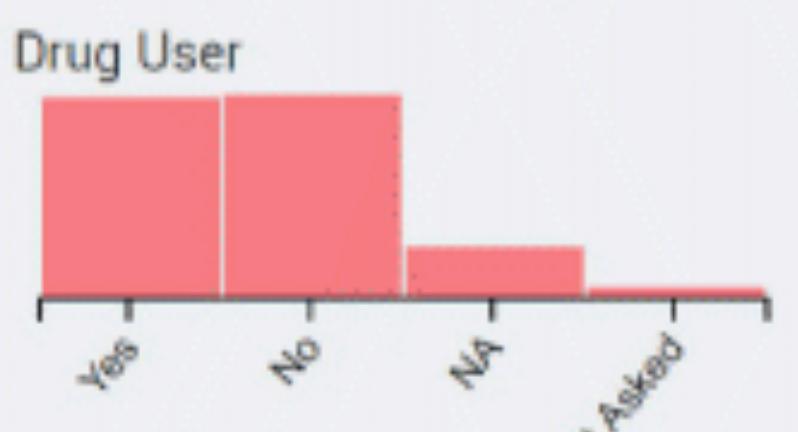
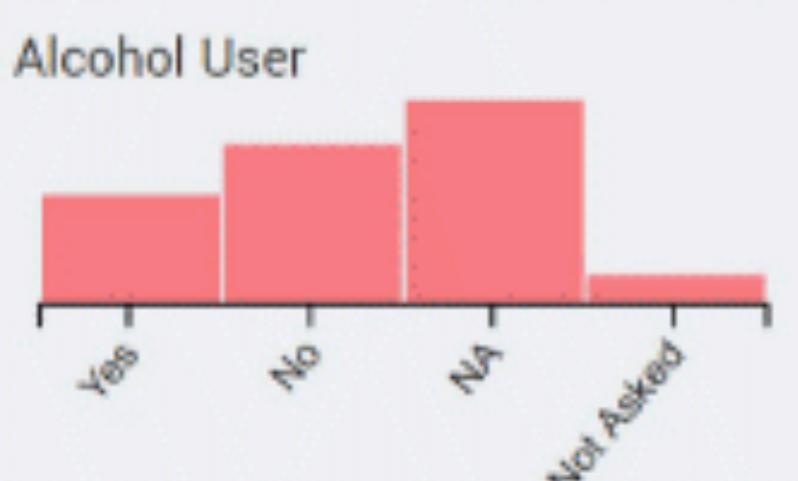
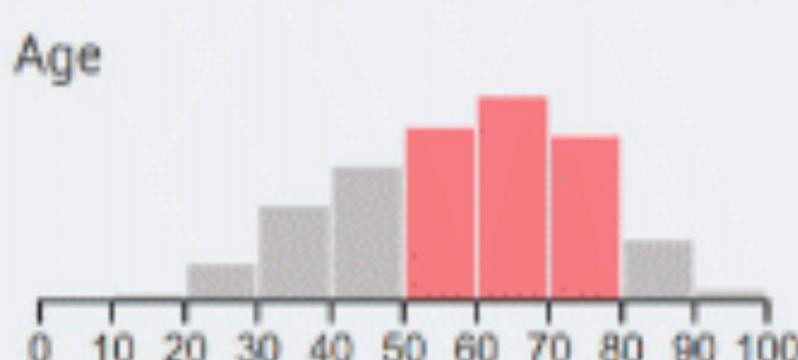
1: All Patients

2. Surgery

3. Excluding injection

1

4. PAGE



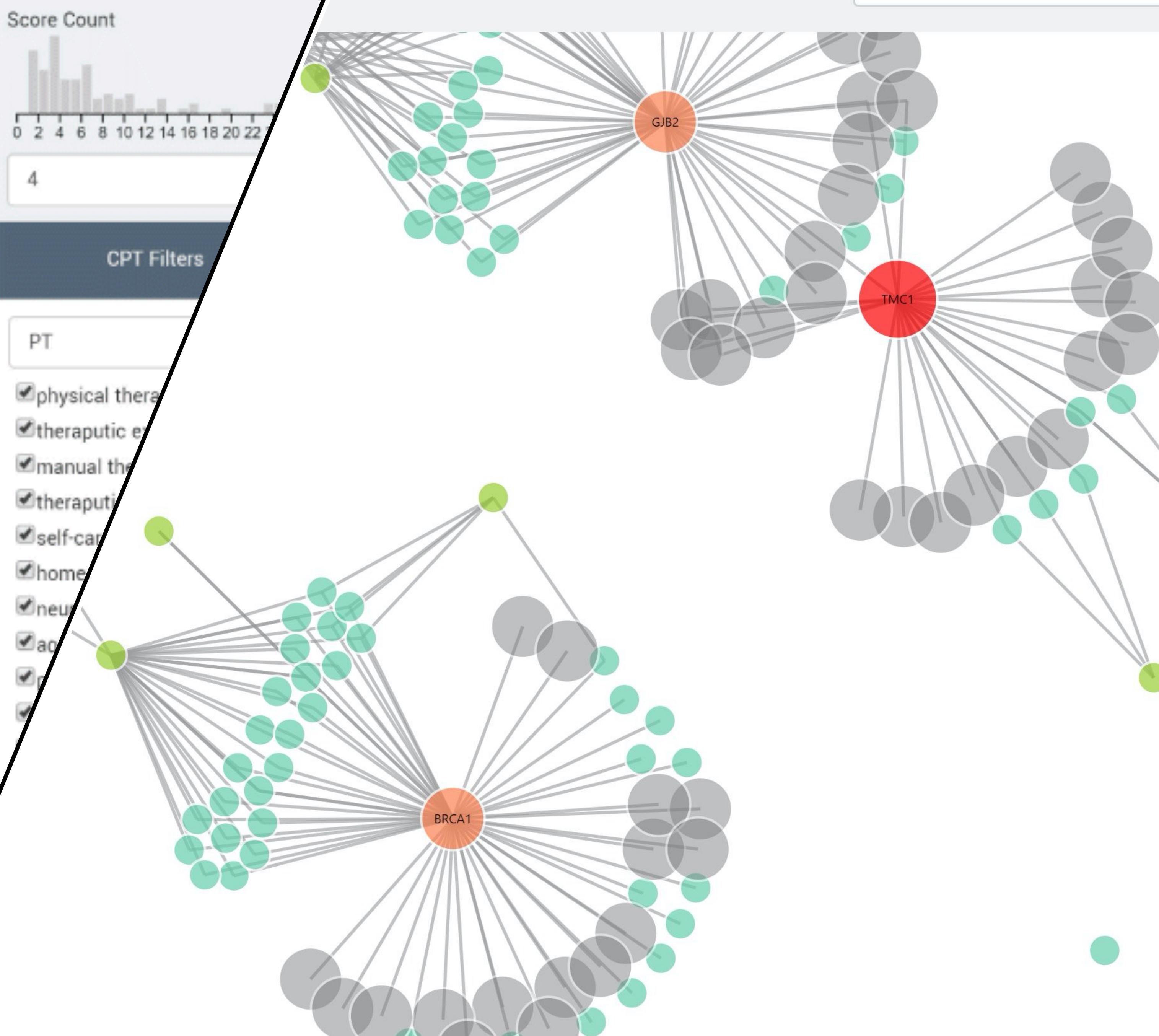
Score Count

1

CPT Filters

PT

- physical therapy
 - therapeutic exercise
 - manual therapy
 - therapeutic ultrasound
 - self-care
 - home exercise
 - neuromuscular



*Used to live here
And stare at this
mountain*



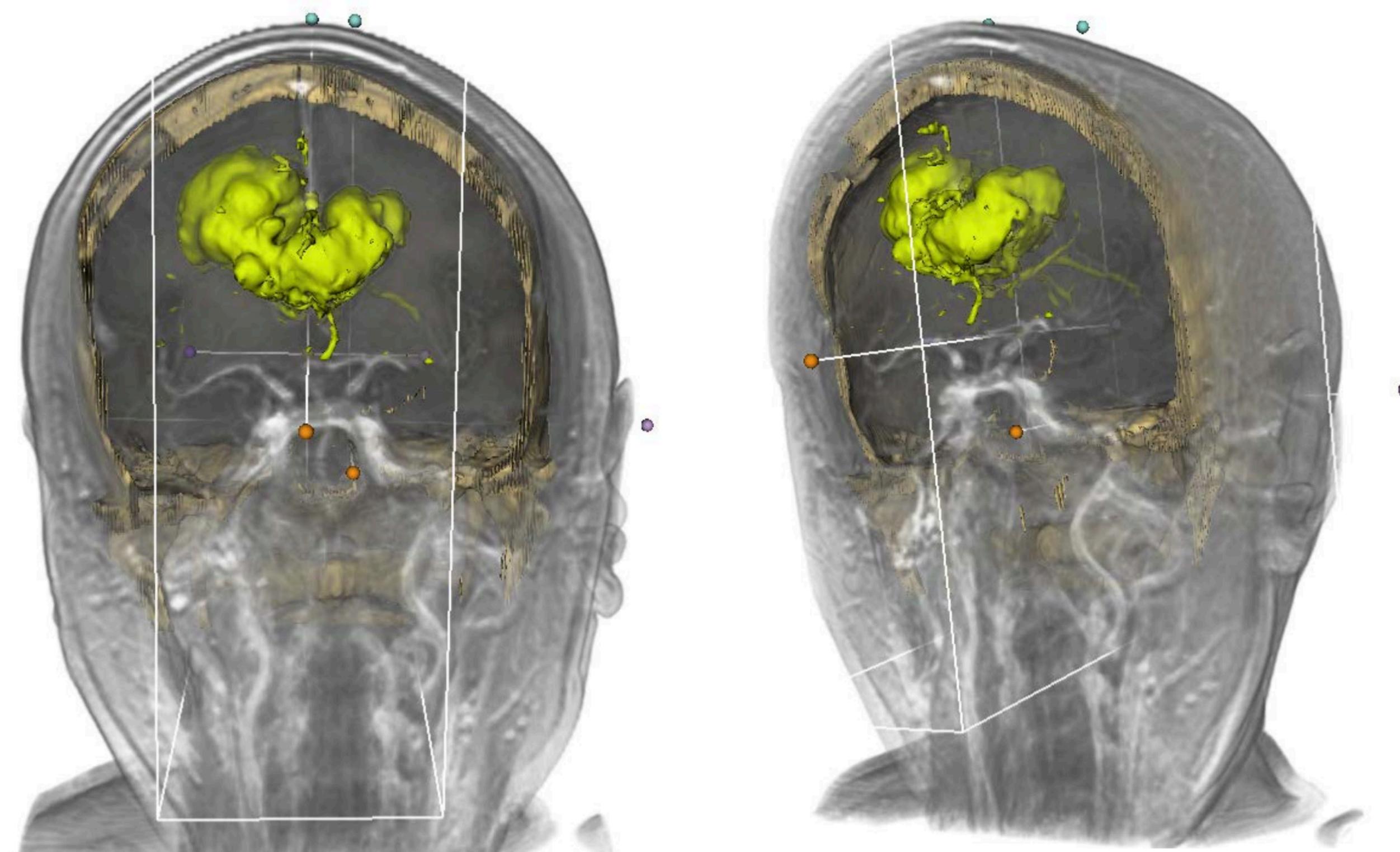


Figure 15 A. Direct MRI Rendering with tumour segmentation combined with CT Skull Segmentation.

Figure 15 B. Direct MRI Rendering with tumour segmentation combined with CT Skull Segmentation.

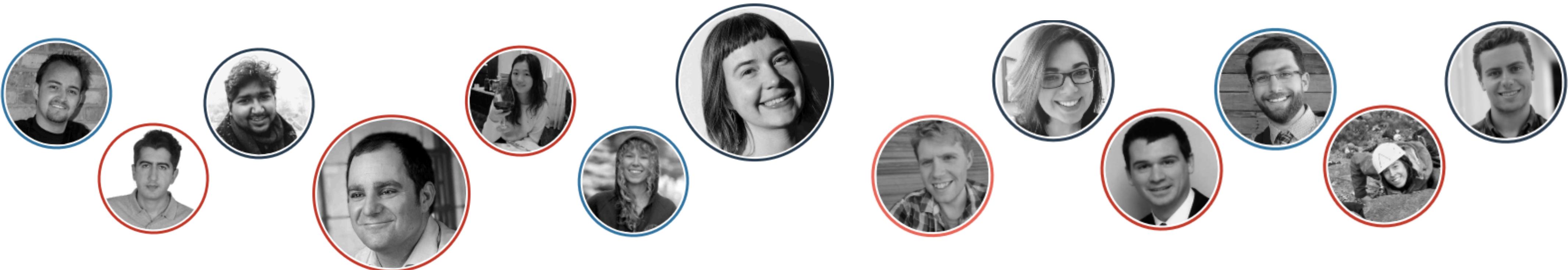




***Probably
find me
somewhere
here***

My academic home

The Visualization Design Lab





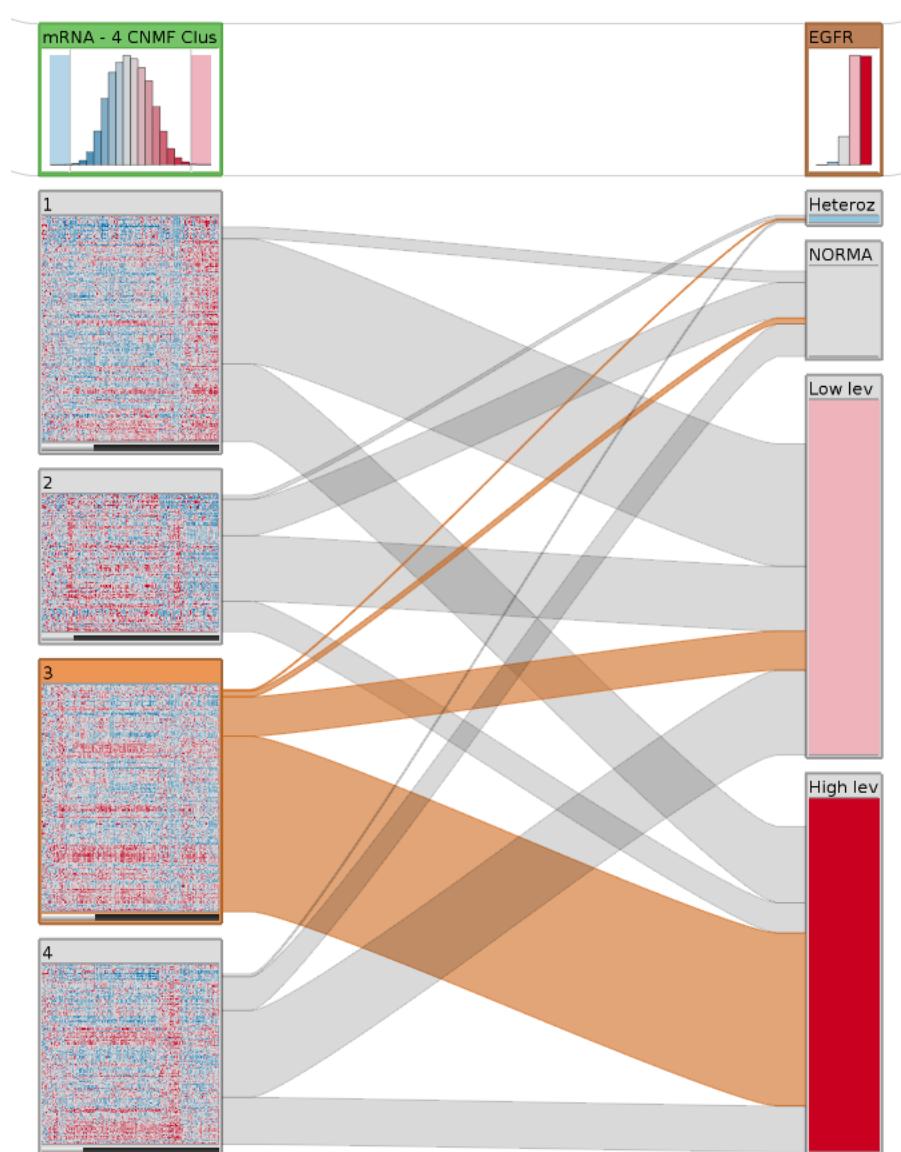
We come from a variety of disciplines..

***Computer Science, Applied Math,
Art, Astrophysics, Graphic Design,
Oceanography,***

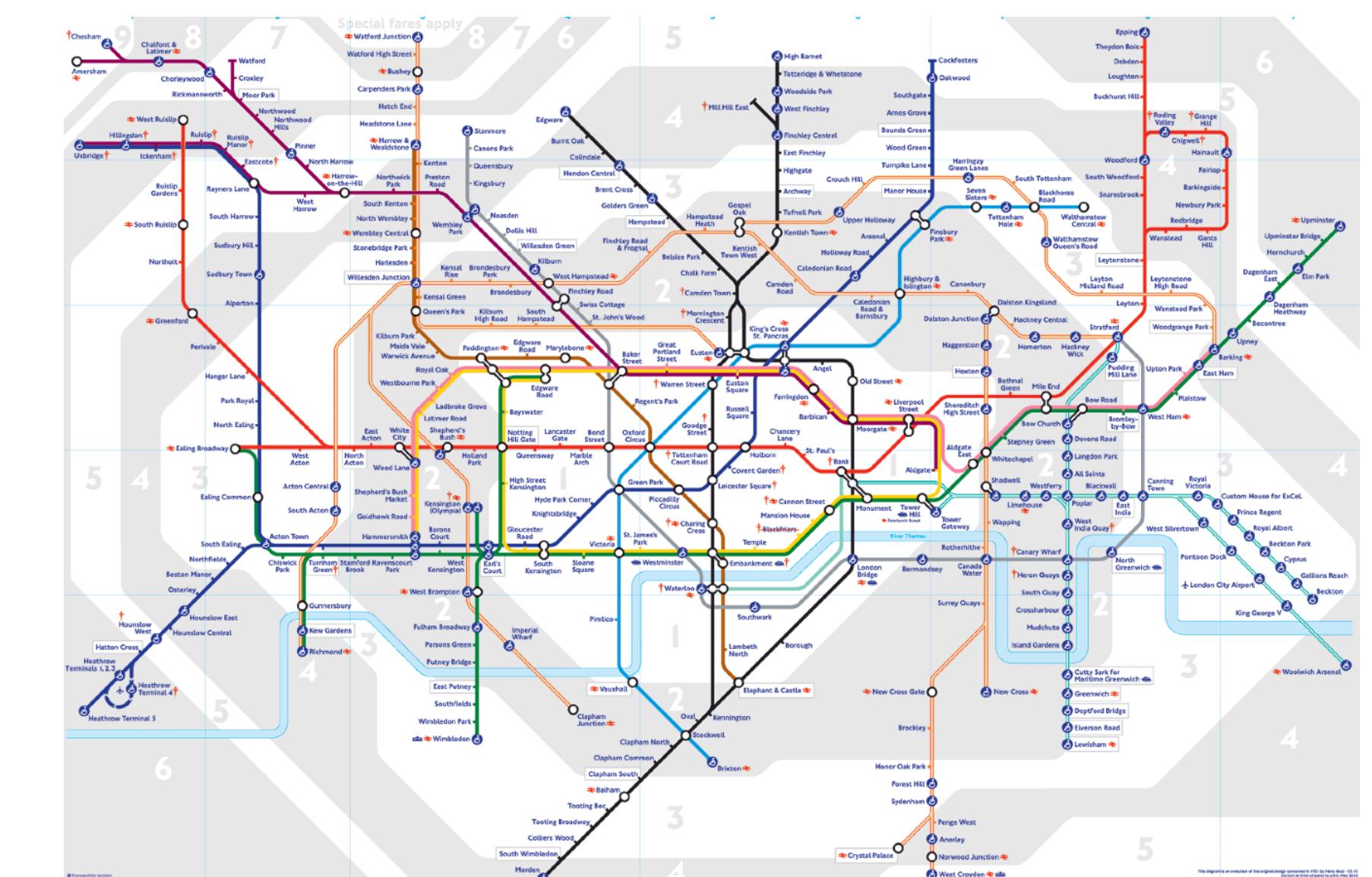
**We are *visualization*
*researchers***

***Visualization is the process
that transforms (abstract)
data into interactive
graphical representations
for the purpose of
exploration, confirmation,
or presentation.***

Why Use visualization?



Confirmation



Exploration

Communication

le more
tter and in less
ne gestalt
view; make
gures are
o understand,
prehensible,
fun, and less

Why Use Figures?

- 1. Communicate more with less***
- 2. Make structure more visible***
- 3. More accessible, easier to comprehend, memorable***

le more
tter and in less
the gestalt
view; make
gures are
to understand,
mprehensible,
fun, and less

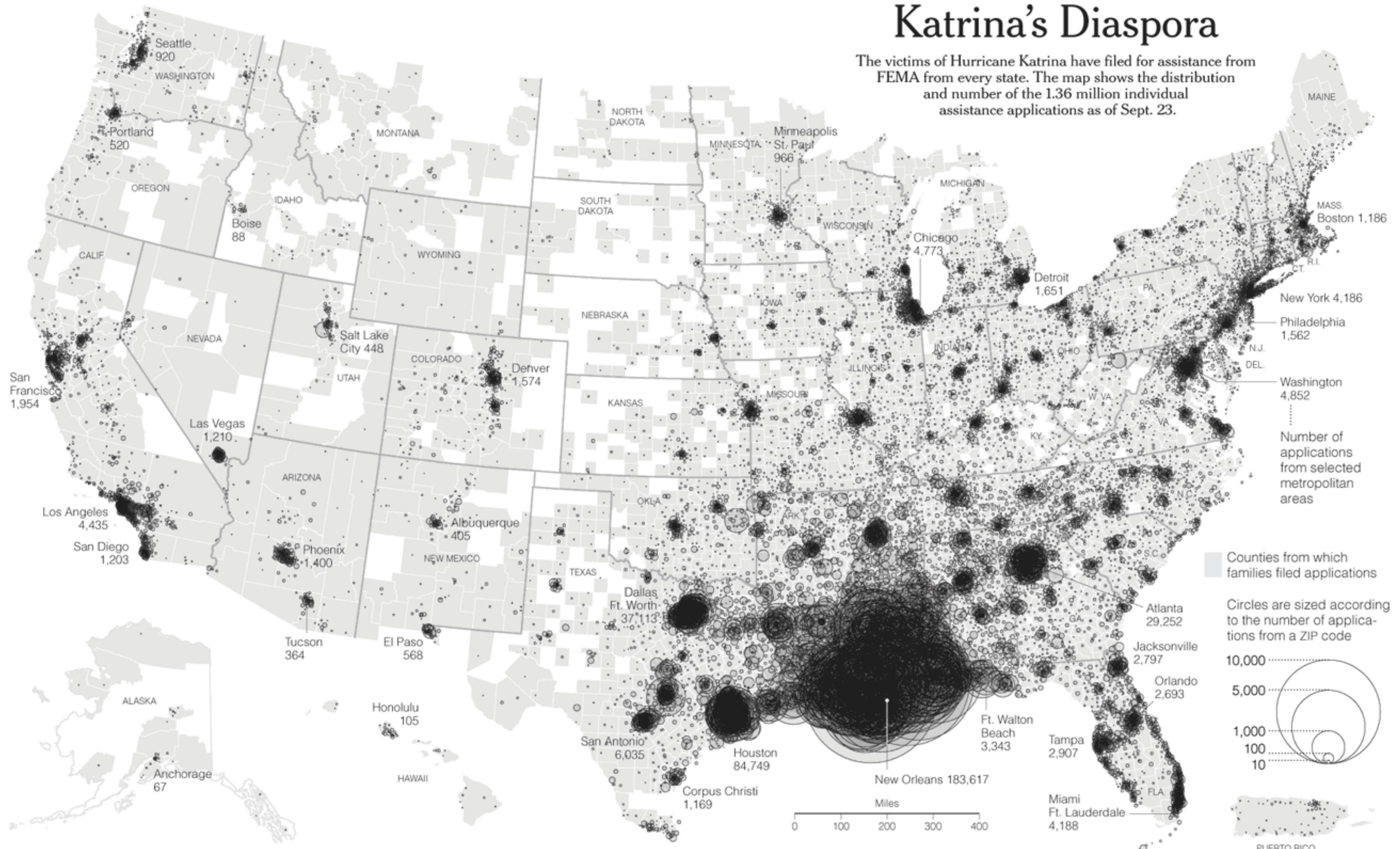
the public schools were shut down, the city's main public hospital was a wreck, and the city's public-housing projects were shuttered.

Campanella then switched to an identically constructed map, only this time based on 2010 census data, and in bits and pieces on the screen there was a simple and arresting picture of what Katrina meant. In the neighborhoods that were once a dense black, many of the little squares had thinned and turned gray. The sharp lines that once separated the teapot from Central City were now blurry: the white areas of the city were pushing north, into the vacuum left by the exodus. The Bywater was graying, as it gentrified still further. "Before Katrina, an American Community Survey estimate of New Orleans Parish population was four hundred and fifty-five thousand, and about sixty-eight per cent black," Campanella said. "Now the latest estimate is three hundred and eighty-four thousand, and it's about

***Textual description of a map
of the effects of hurricane
Katrina on New Orleans.
New Yorker, posted by Alberto Cairo***

Katrina's Diaspora

The victims of Hurricane Katrina have filed for assistance from FEMA from every state. The map shows the distribution and number of the 1.36 million individual assistance applications as of Sept. 23.



They are scattered through all 50 states, the District of Columbia and Puerto Rico — 623 in Utah, 1,114 in Kansas, 101 way out in Alaska. They are clustered by the thousands in large Southern cities like Dallas, Atlanta and Memphis, and huddled in handfuls in unlikely hamlets like Shell Knob, Mo. (pop. 1,393) and Fountain Run, Ky. (pop. 236).

Evacuees fled Hurricane Katrina and the floods that followed in caravans of cars and fleets of buses, on helicopters and chartered planes, by boat and, a few, on foot. A month after the storm, a map

emerges of where they landed, based on ZIP codes from which applications for aid were submitted to the Federal Emergency Management Agency as of Sept. 23.

Of 1,356,704 applications, 86 percent came from Louisiana, Mississippi, Texas and Alabama. But 35,539 families were more than 1,000 miles from the Gulf — among the farthest: one in Nome, Alaska, 3,931 miles from the French Quarter and another in Lihue, Hawaii, 4,279 miles away.

Residents of New Orleans, a city that was two-thirds black, seem to have flocked to the nation's African-American population.

centers. On average, the applicants came from counties where blacks were 28 percent of the population, more than twice the national average.

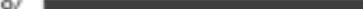
Baton Rouge, La., appears to be temporary home to 10 percent of evacuees from Houston. But after the top 18 hubs, applicants are spread like the winds that whipped through their old neighborhoods: none of the other 900-plus metropolitan areas has even 1 percent of the total.

Some 4,000 ZIP codes — among them Pocahontas, Miss.; Promise City, Iowa; and Hope, Mich. — had just one applicant

Applications by state

Louisiana	523,149	38.6%
Mississippi	383,840	28.3%
Texas	156,895	11.6%
Alabama	109,469	8.1%
Georgia	35,342	2.6%
Florida	31,005	2.3%
Tennessee	15,529	1.1%
Arkansas	11,027	0.8%
California	10,953	0.8%
Illinois	6,430	0.5%
Others	73,065	5.4%

Applications by distance from New Orleans

MILES	APPlicants	PCT.	
0-100	626,232	46.2%	
100-200	338,080	24.9%	
200-400	184,169	13.6%	
400-800	143,497	10.6%	
800-1,600	45,371	3.3%	
1,600-3,200	13,403	1.0%	
3,200+	232	0.0%	

Distances could not be calculated for 0.4 percent of applications.

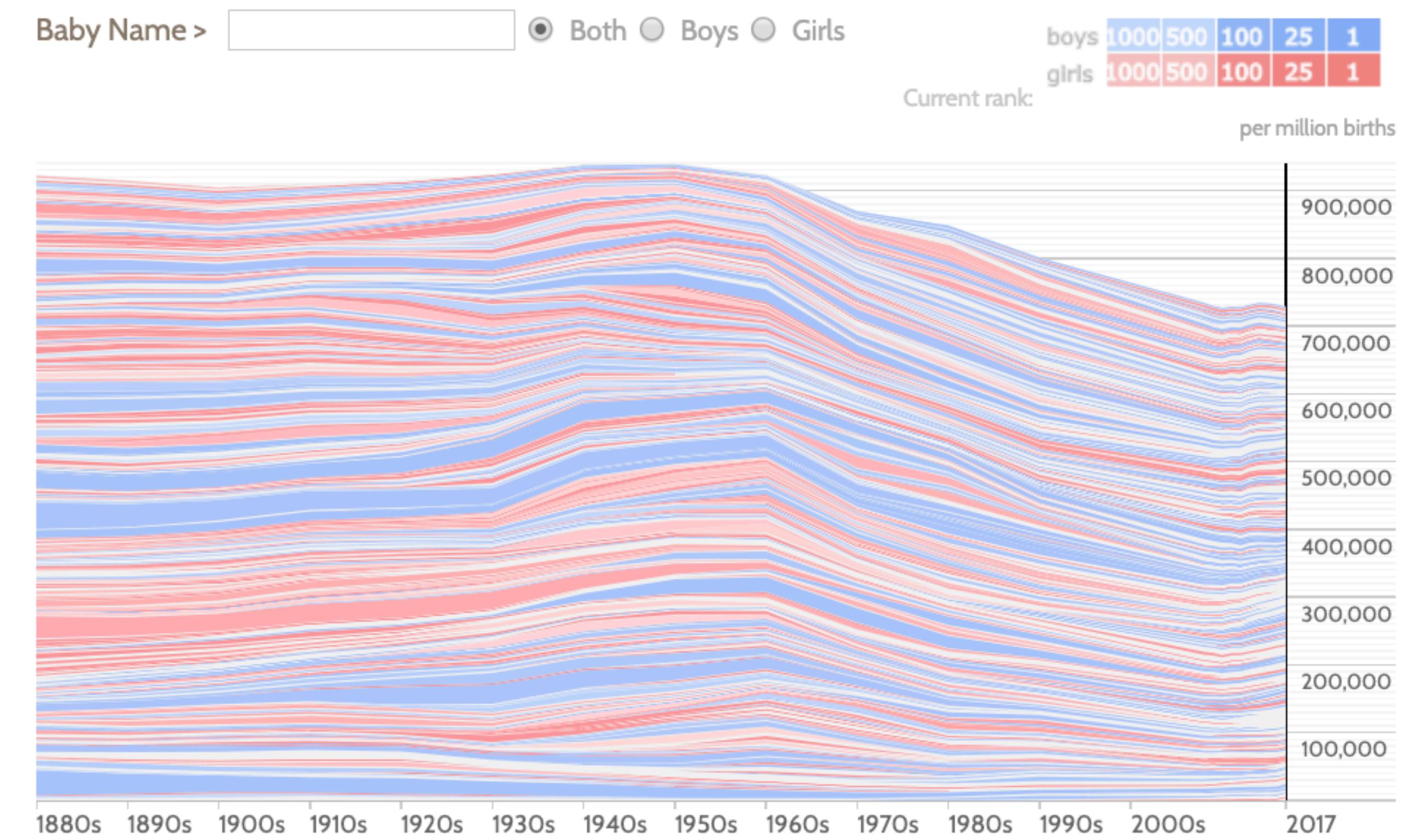
Sources: FEMA; Census Bureau; Queens College Sociology Department
Matthew Ericson, Archie Tse and Jodi Wilgoren/The New York Times

**A good data visualization
Makes data accessible,
Enables insight**

Exploratory

**Readily accessible
exploratory
visualization:
Name Voyager**

**Martin
Wattenburg 2005**



<http://www.babynamewizard.com/voyager#prefix=&sw=both&exact=false>

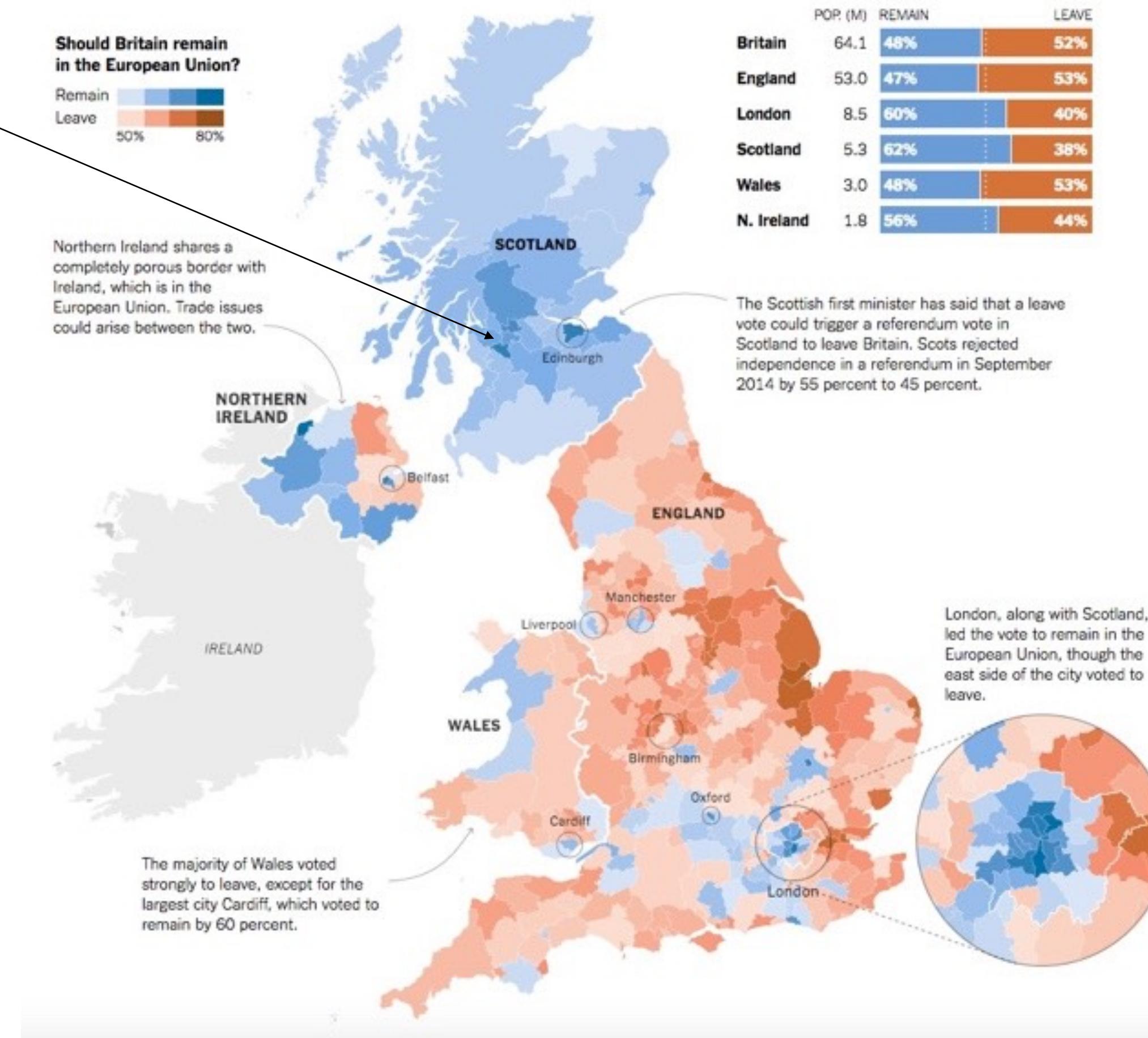
A good data visualization
Makes data accessible,
Enables insight,
Communicates

How Britain Voted in the E.U. Referendum

By GREGOR AISCH, ADAM PEARCE and KARL RUSSELL UPDATED June 24, 2016

Britons voted on Thursday to leave the European Union. The Leave side led with 17.4 million votes, or 52 percent, versus the Remain side's 16.1 million, or 48 percent, with a turnout of around 72 percent. [RELATED ARTICLE](#)

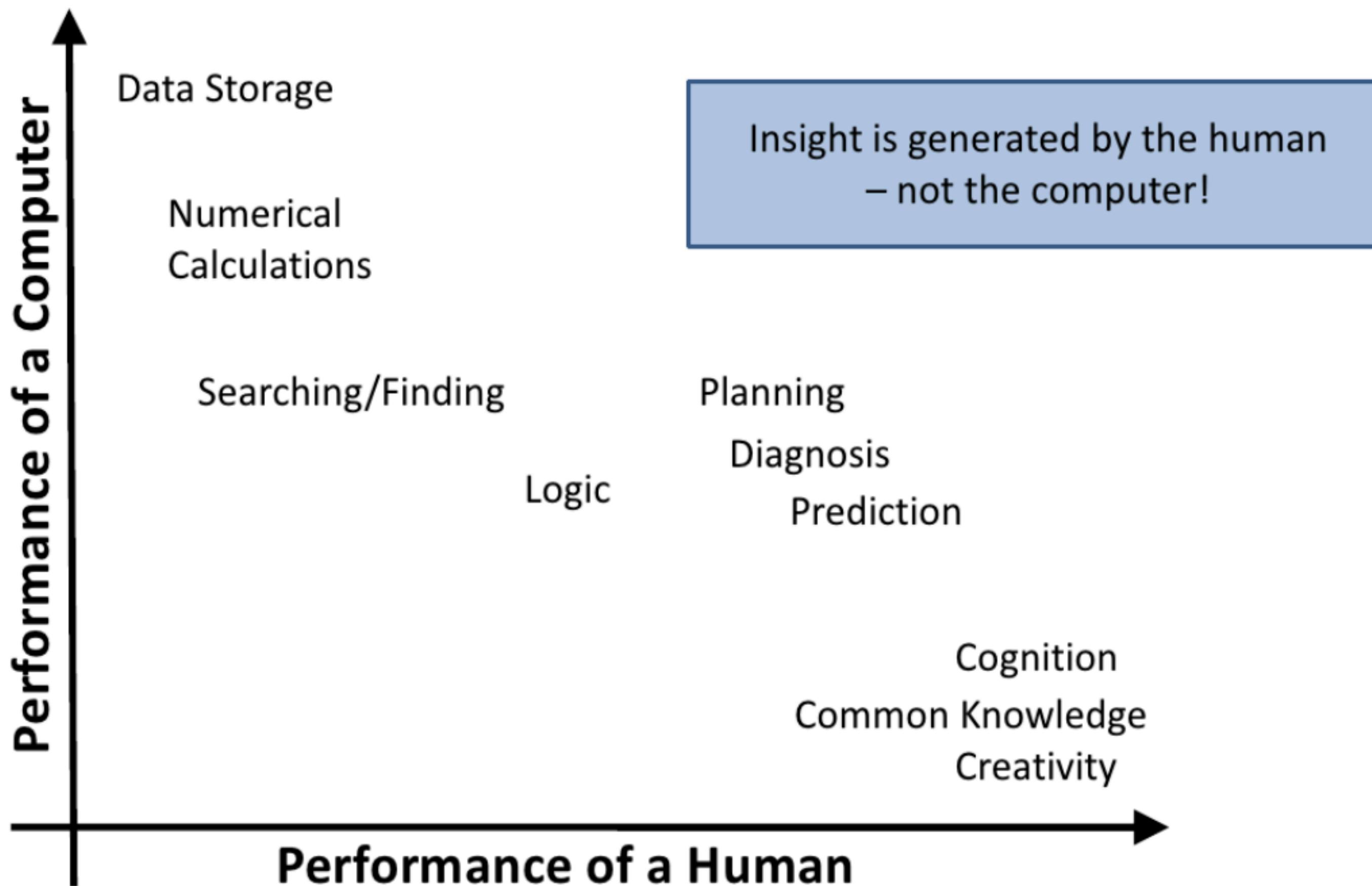
**People of
Glasgow were
not impressed**





***Human vs Computer
Who does it better?***

Human vs Computer



**A good data visualization
makes data accessible,
Enables insight,
Communicates
combines strengths of humans and
computers**

Can't we just use statistics?

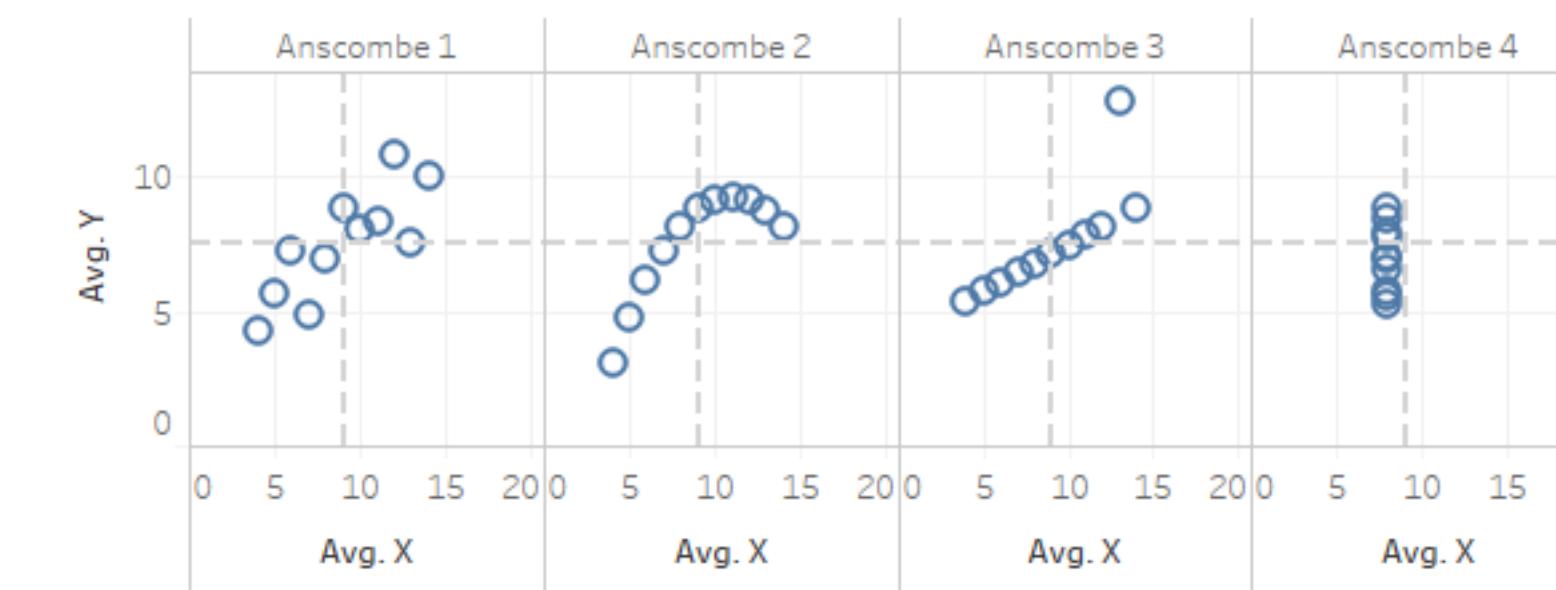
“numerical calculations are exact, but graphs are rough”?

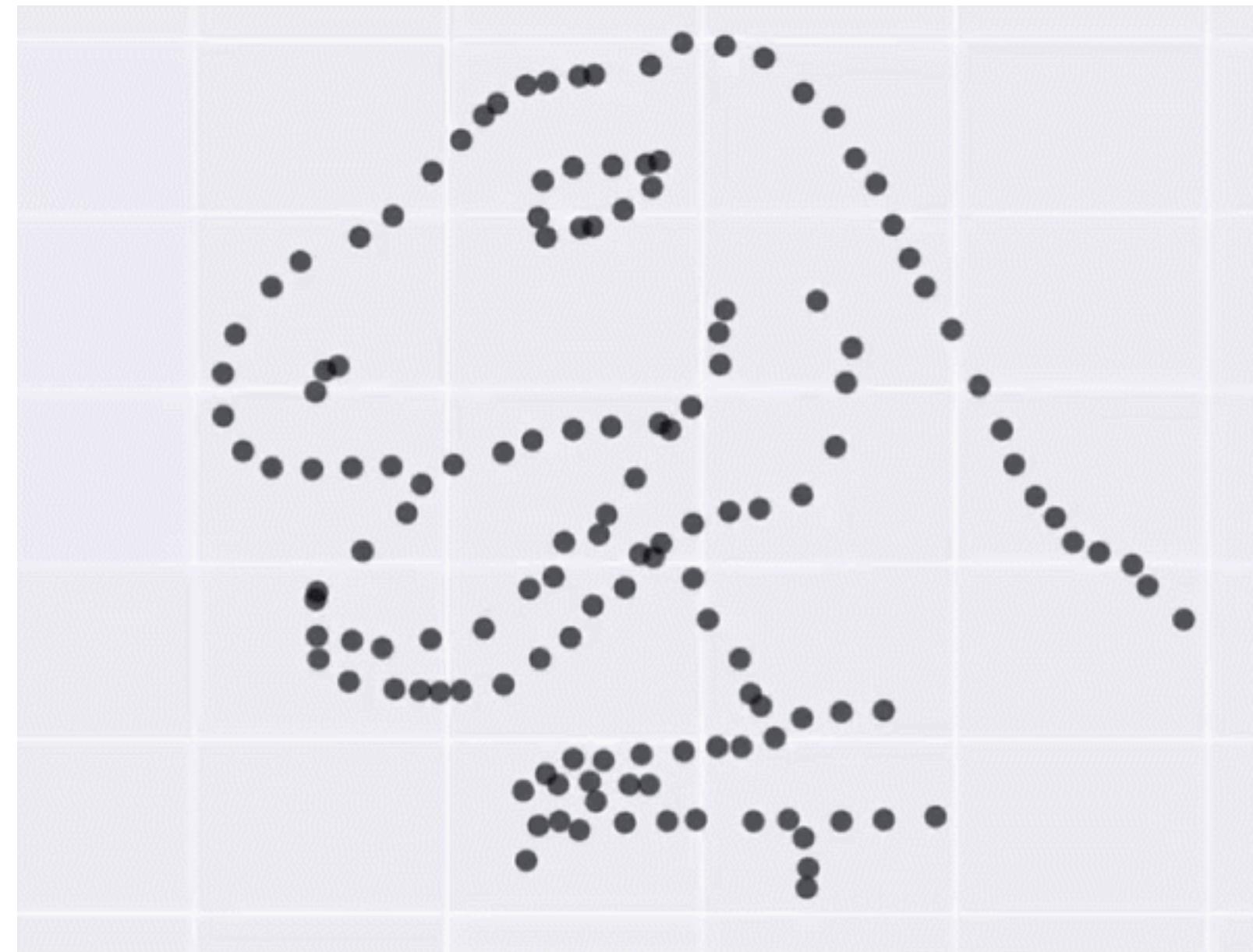
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Anscombe's Quartet.

- The average x value is **9** for each dataset
- The average y value is **7.50** for each dataset
- The variance for x is **11** and the variance for y is **4.12**
- The correlation between x and y is **0.816** for each dataset
- A linear regression (line of best fit) for each dataset follows the equation **$y = 0.5x + 3$**

Anscombe quartet





X Mean: 54.26
Y Mean: 47.83
X SD : 16.76
Y SD : 26.93
Corr. : -0.06

Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing, CHI 2017, Justin Matejka, George Fitzmaurice

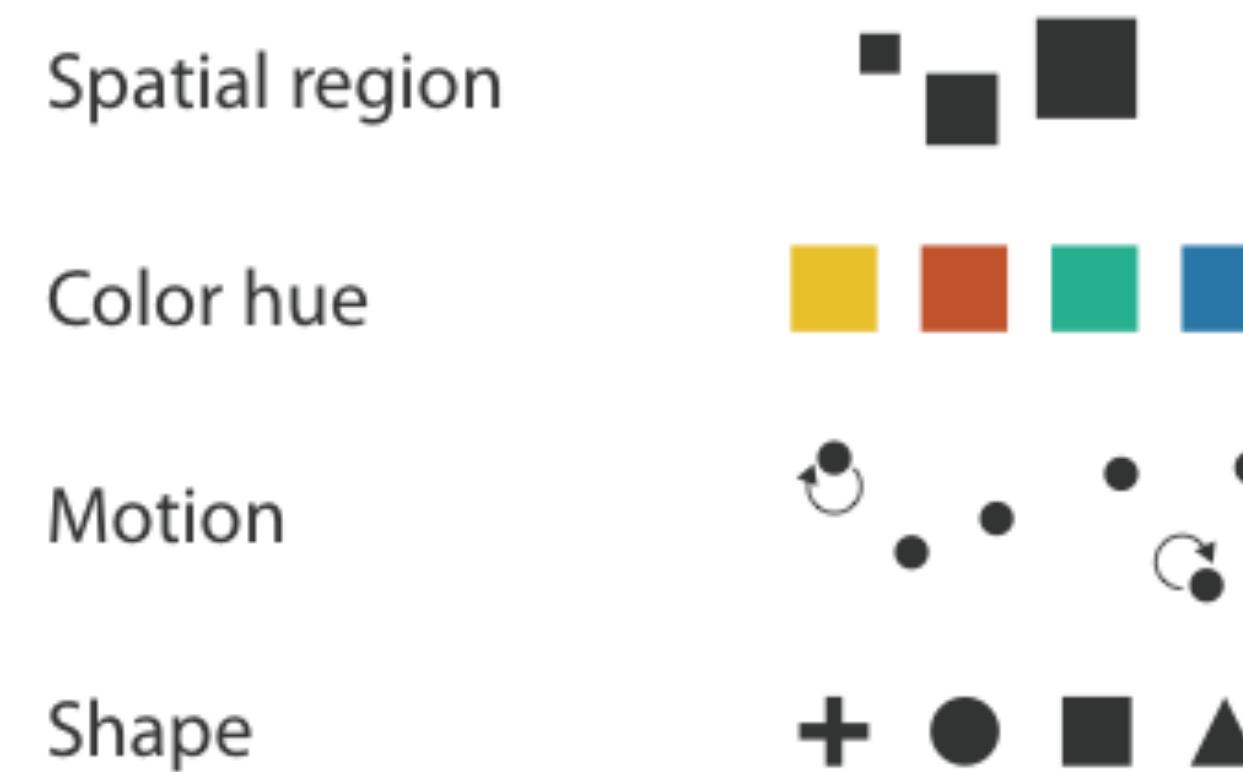
[Link to video](#)

<https://www.autodeskresearch.com/publications/samestats>

④ **Magnitude Channels: Ordered Attributes**

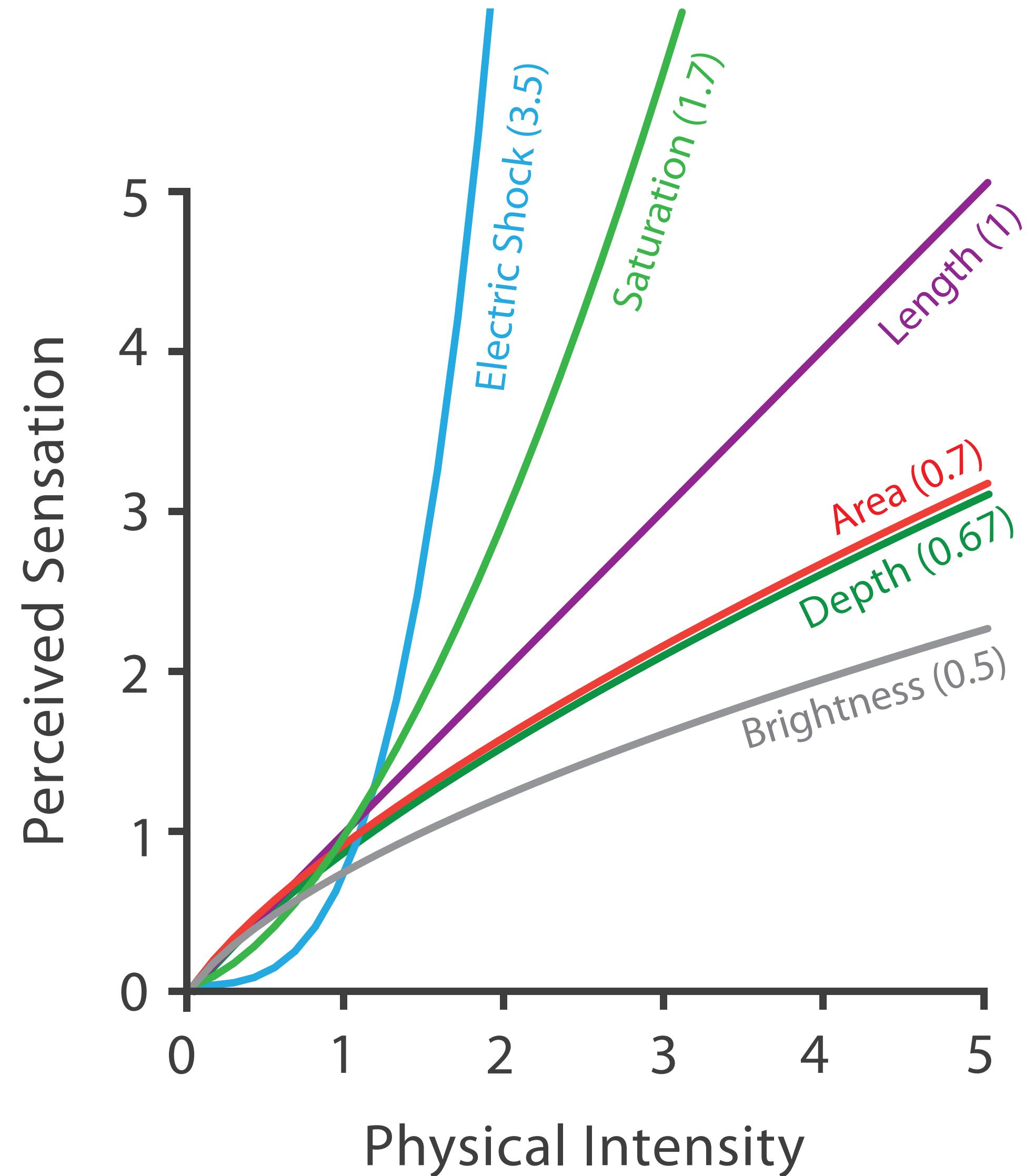


④ **Identity Channels: Categorical Attributes**



Data Encoding

Steven's Psychophysical Power Law: $S= I^n$

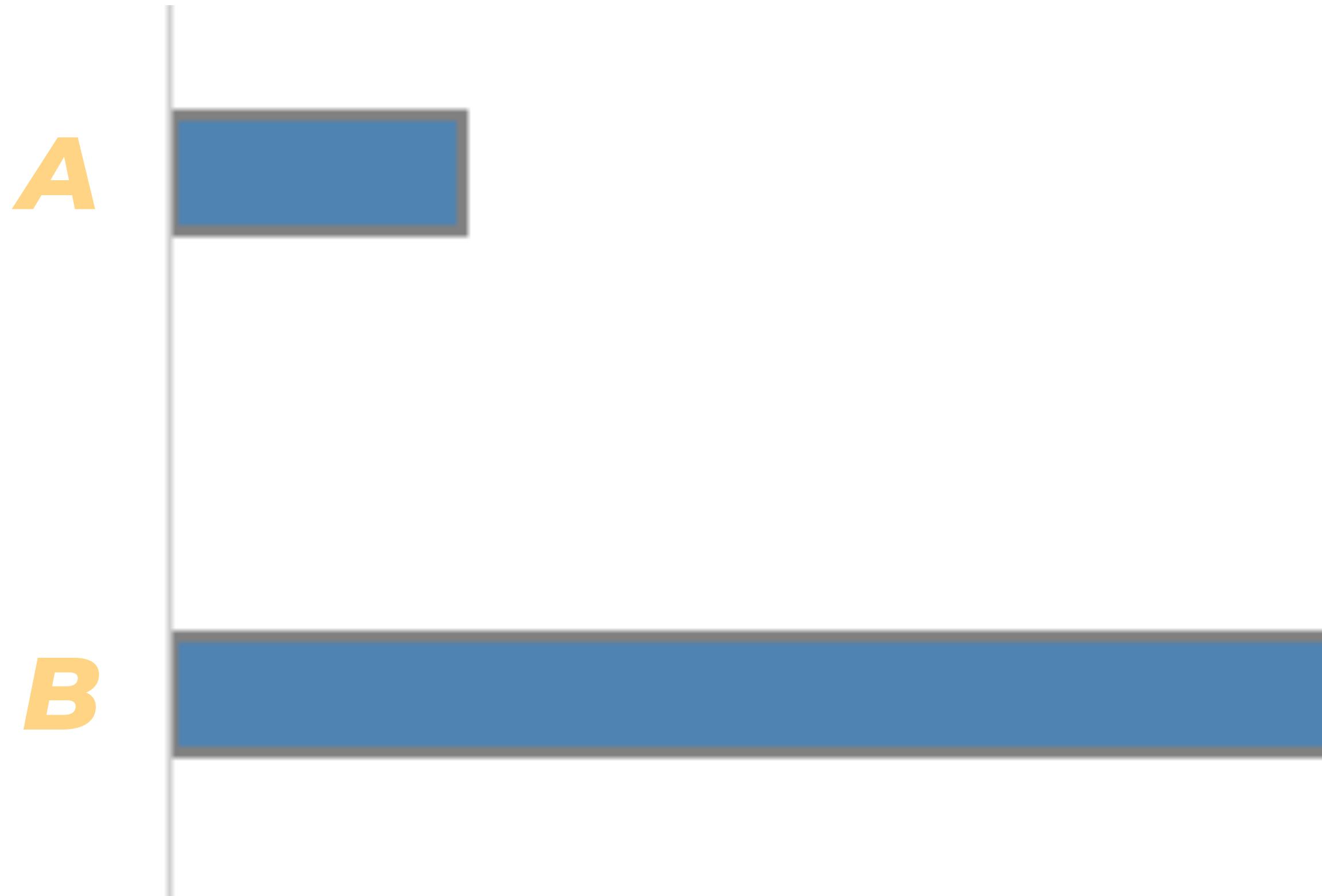


Not all quantitative channels are created equal

Relationship between intensity of a stimulus and perceived magnitude increase in sensation

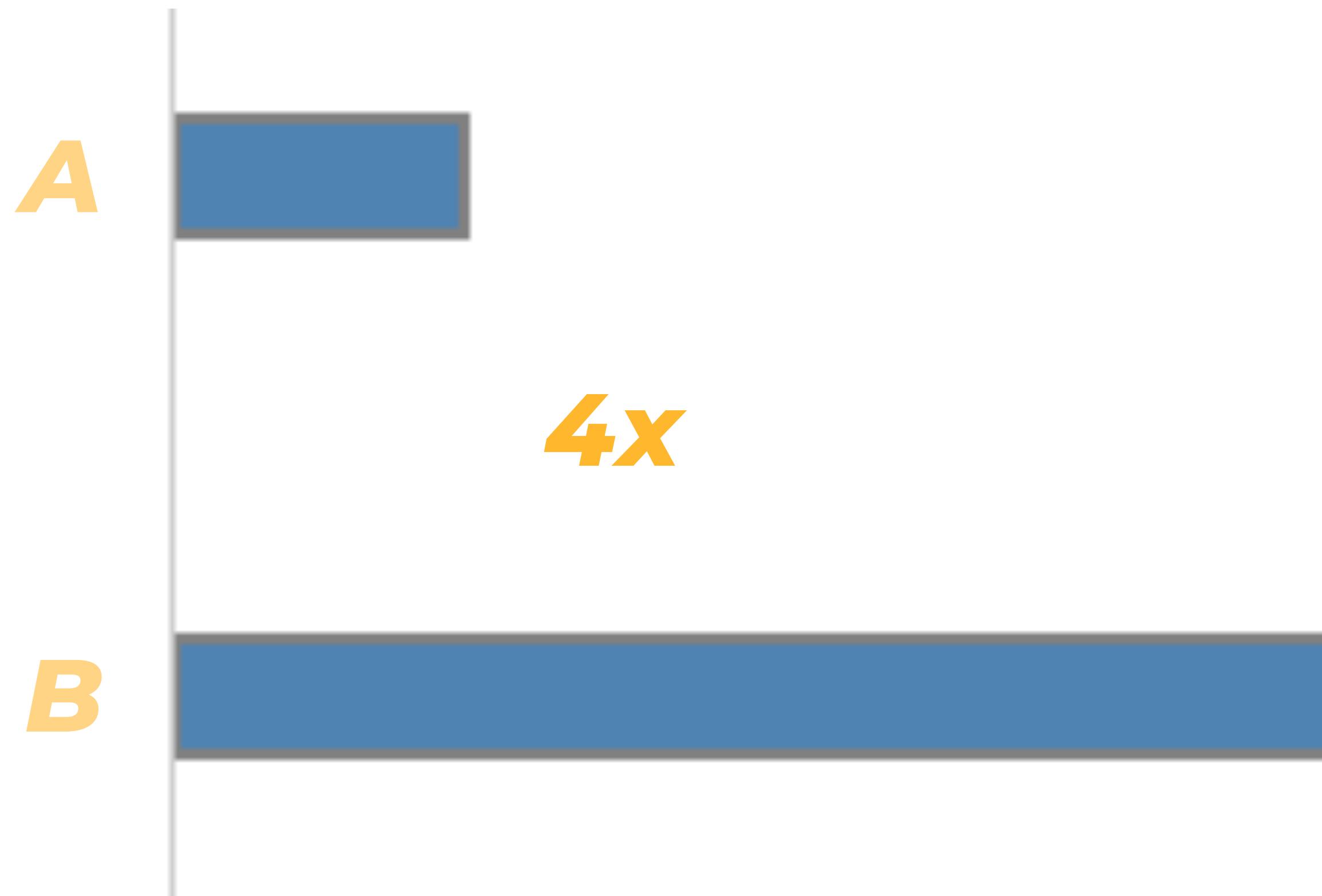
Quantitative channels are different

***How much
longer is B?***



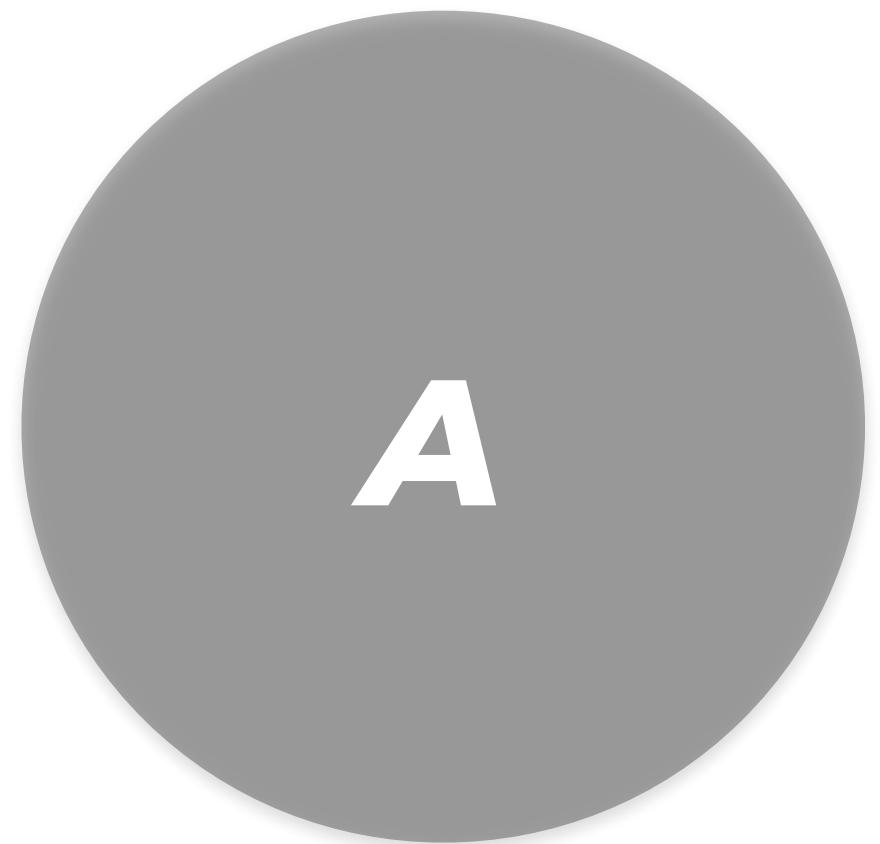
Quantitative channels are different

**How much
longer is B?**

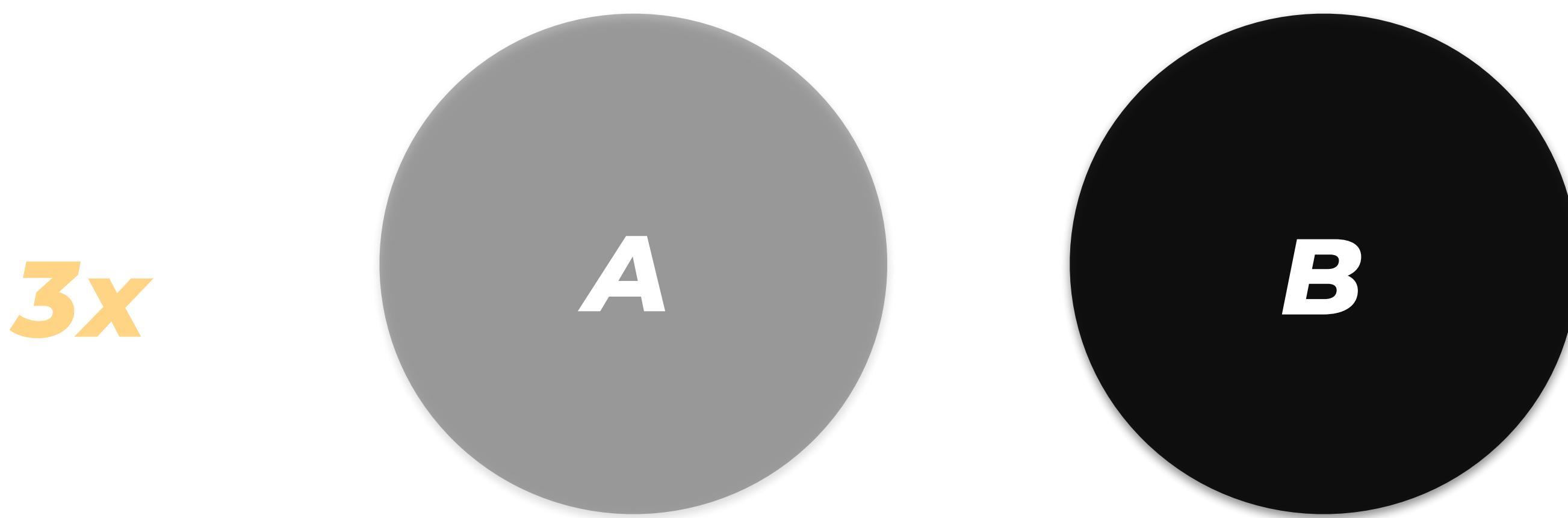


Quantitative channels are different

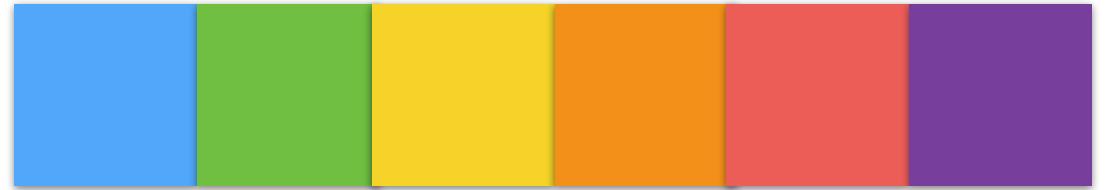
***How much
darker is B?***



Quantitative channels are different

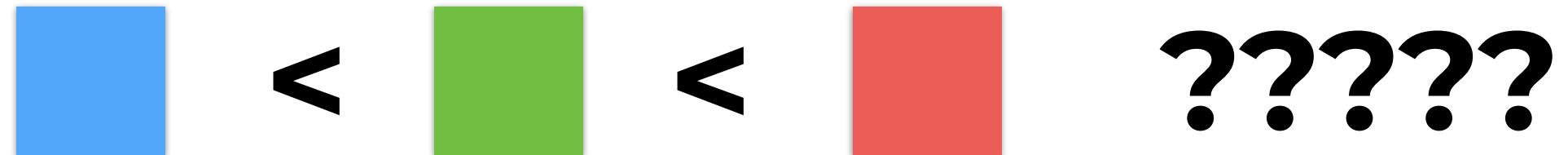


Color



Good for qualitative data (identity channel)

Does not work for quantitative data



*****Alex's Rule of Thumb:***

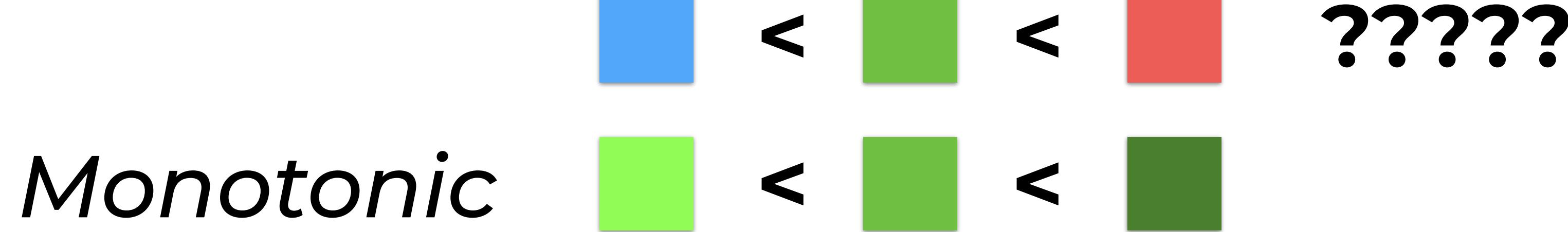
***minimize color use for encoding data
use for highlights***

Color



Good for qualitative data (identity channel)

Does not work for quantitative data



Monotonic

****Alex's Rule of Thumb:**
*minimize color use for encoding data
use for highlights*

***What brings
me to Idaho?***

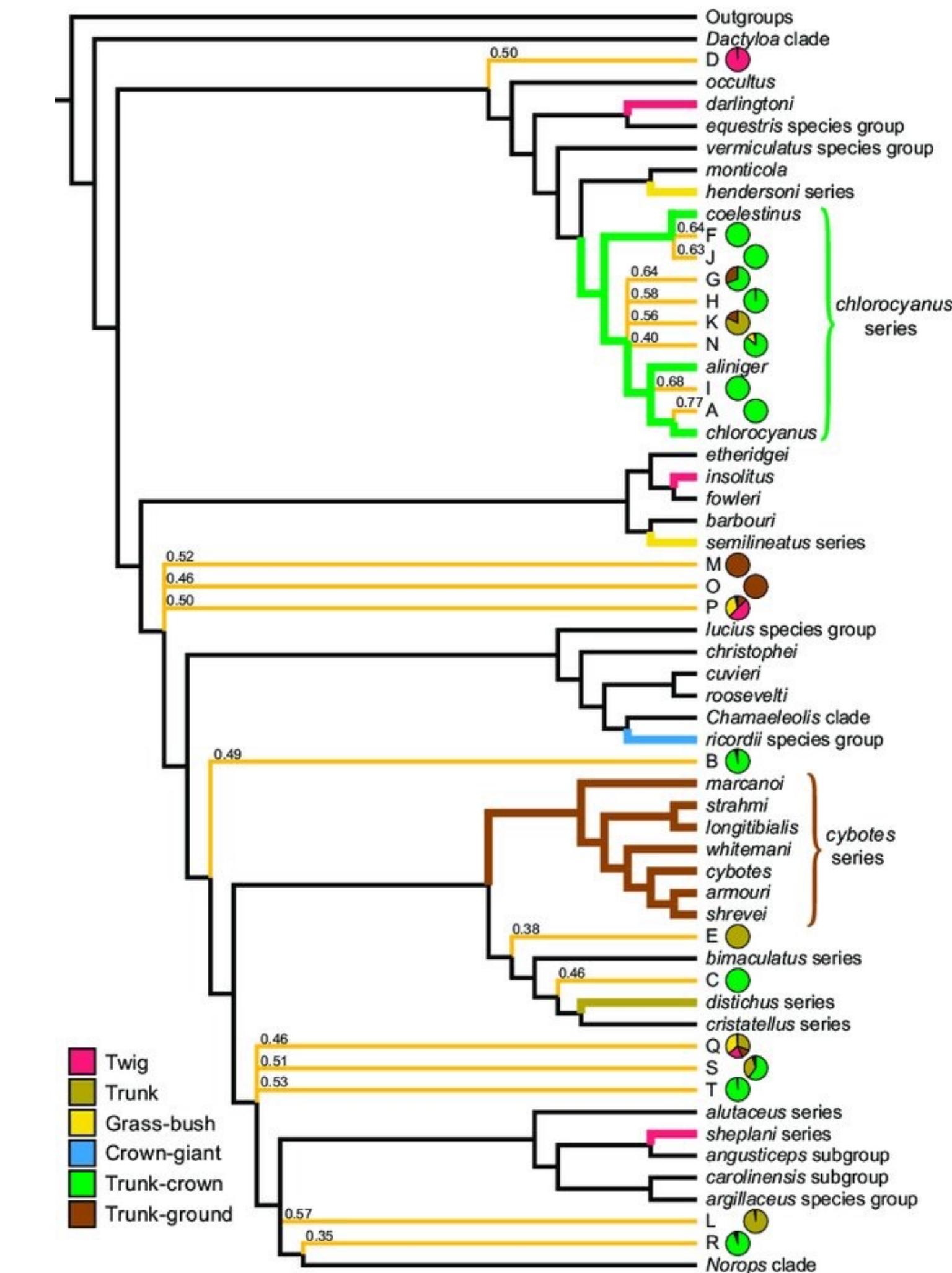


© University of Idaho



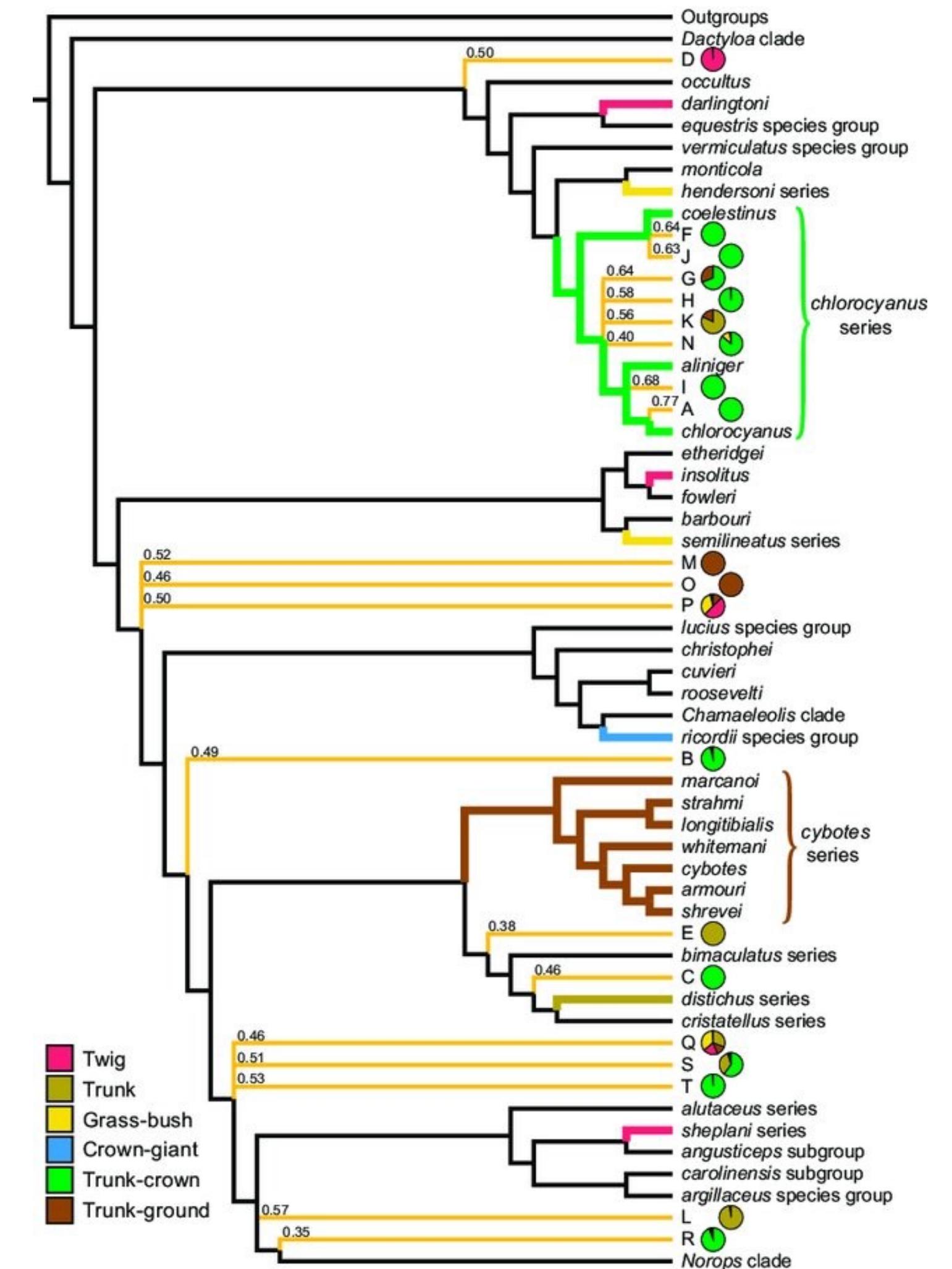
***NSF-funded
Multinet project
software for analysis
of multivariate
networks & trees***

Multivariate networks & trees in the context of phylogenies



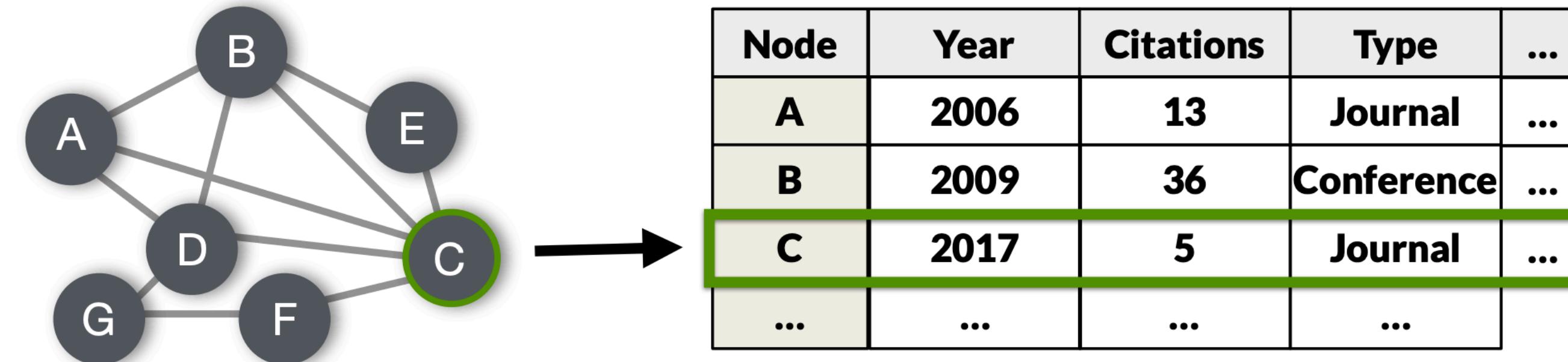
Sherratt, E., del Rosario Castañeda, M., Garwood, R.J., Mahler, D.L., Sanger, T.J., Herrel, A., de Queiroz, K. and Losos, J.B., 2015. Amber fossils demonstrate deep-time stability of Caribbean lizard communities. *Proceedings of the National Academy of Sciences*, 112(32), pp.9961-9966.

Multivariate networks & trees in the context of phylogenies *Ecology and Evolution*



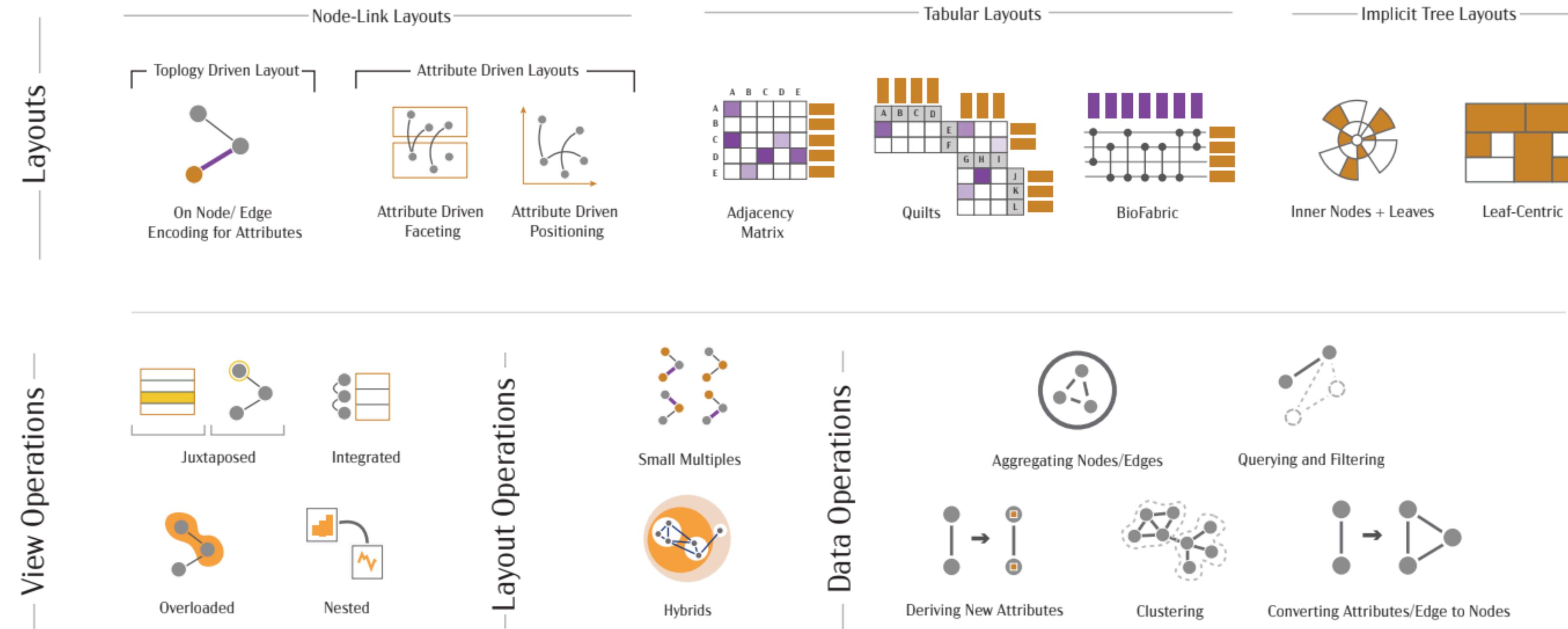
Sherratt, E., del Rosario Castañeda, M., Garwood, R.J., Mahler, D.L., Sanger, T.J., Herrel, A., de Queiroz, K. and Losos, J.B., 2015. Amber fossils demonstrate deep-time stability of Caribbean lizard communities. *Proceedings of the National Academy of Sciences*, 112(32), pp.9961-9966.

What is a multivariate network?



NSF GRANT TO BUILD robust
software for analysis of multivariate
networks and trees

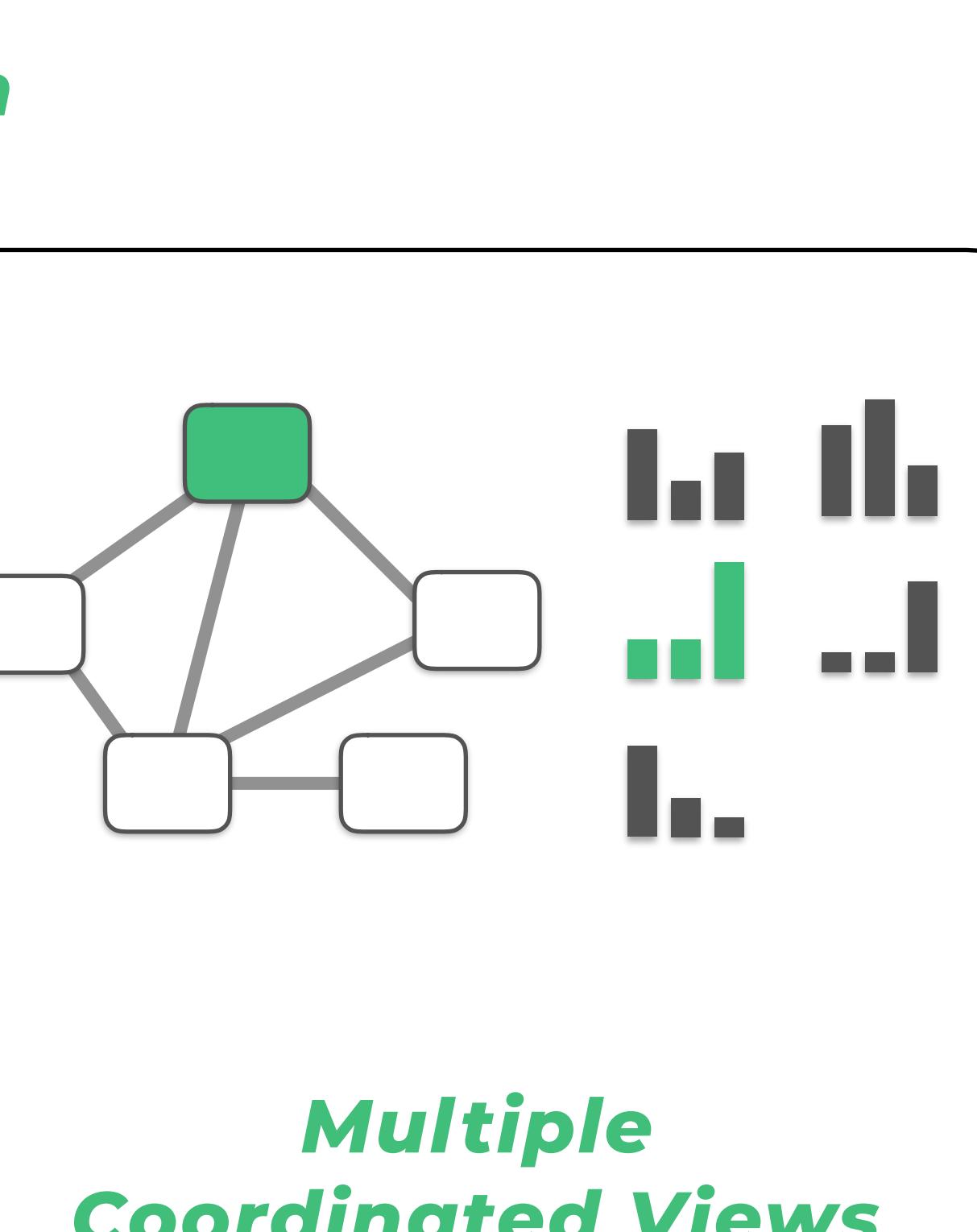
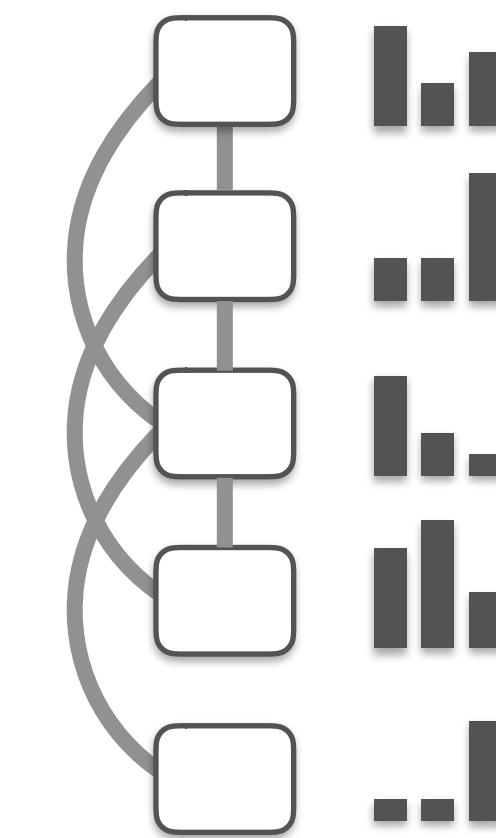
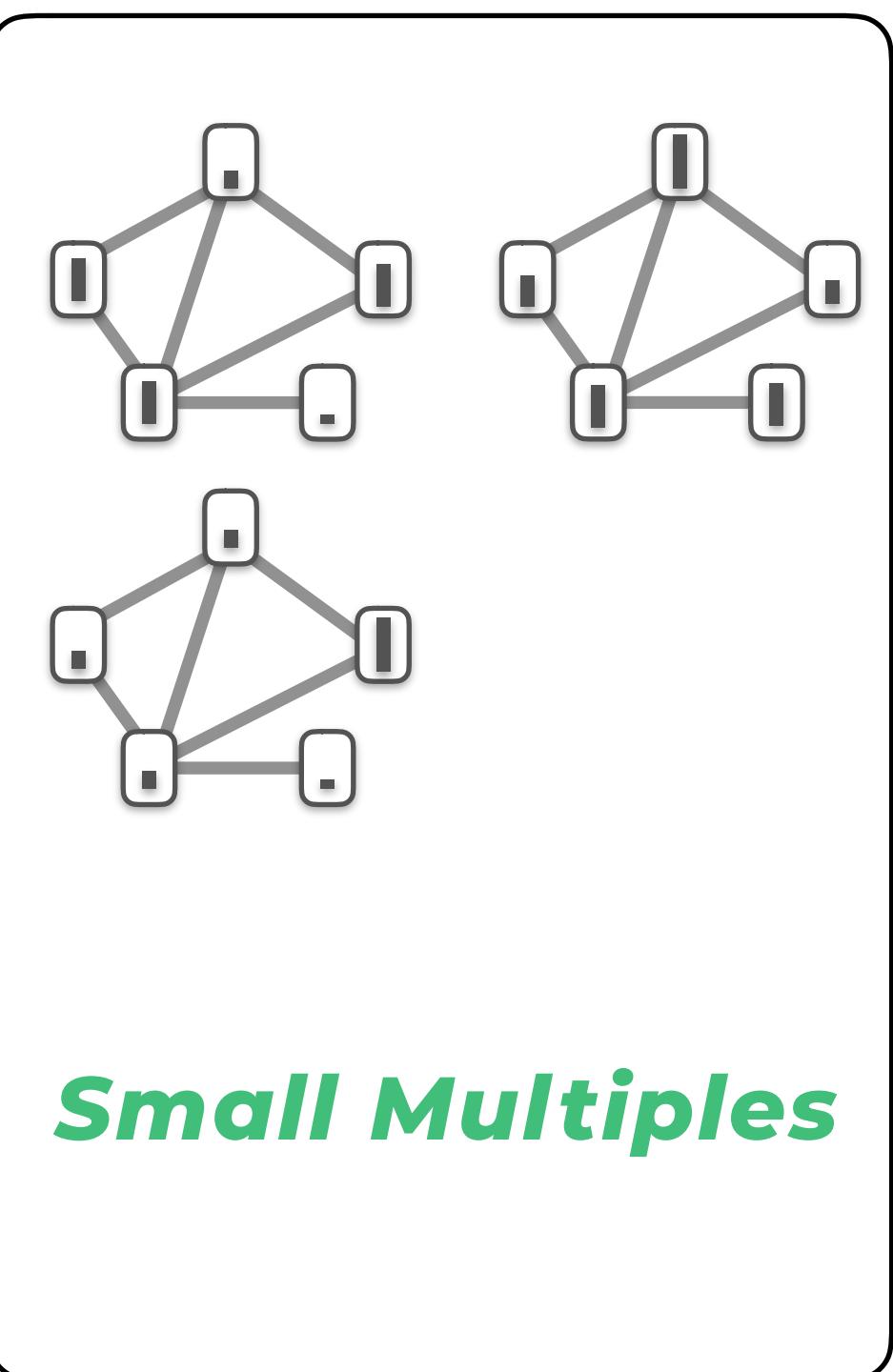
Multivariate network visualization techniques



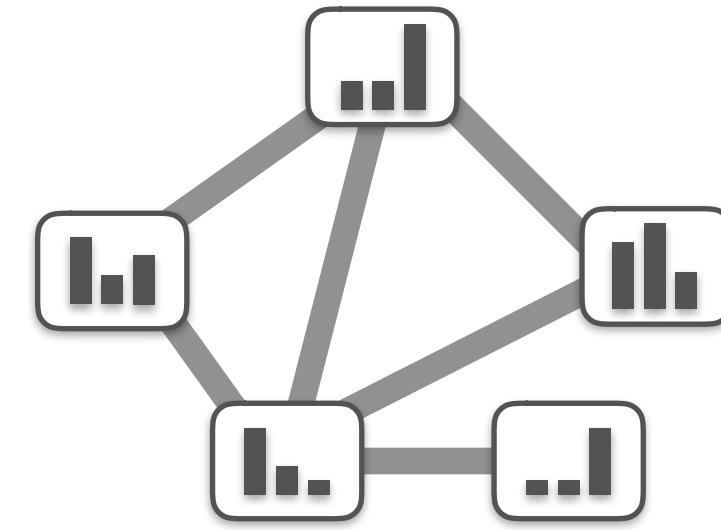
Matrix



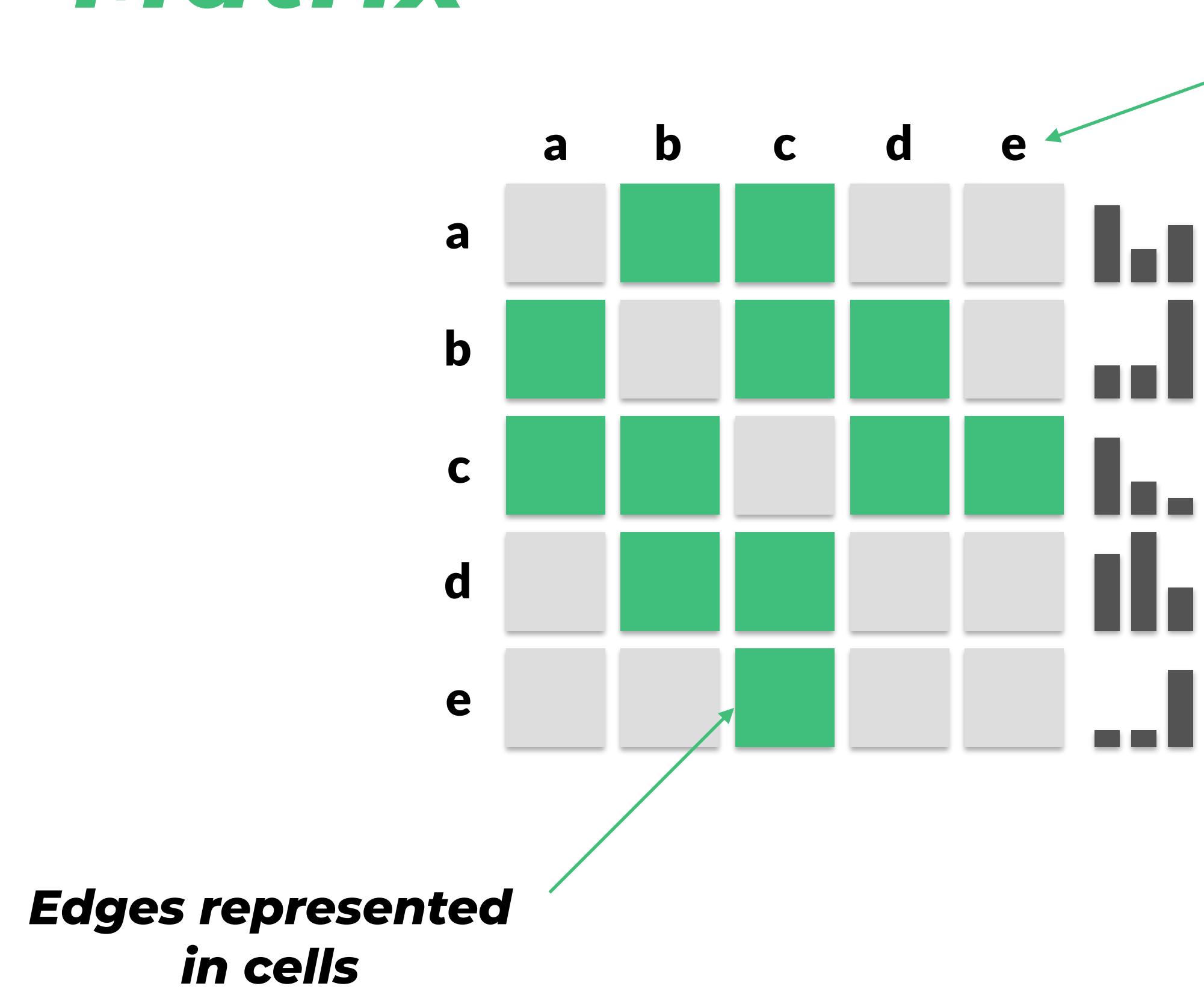
Layout Adaption



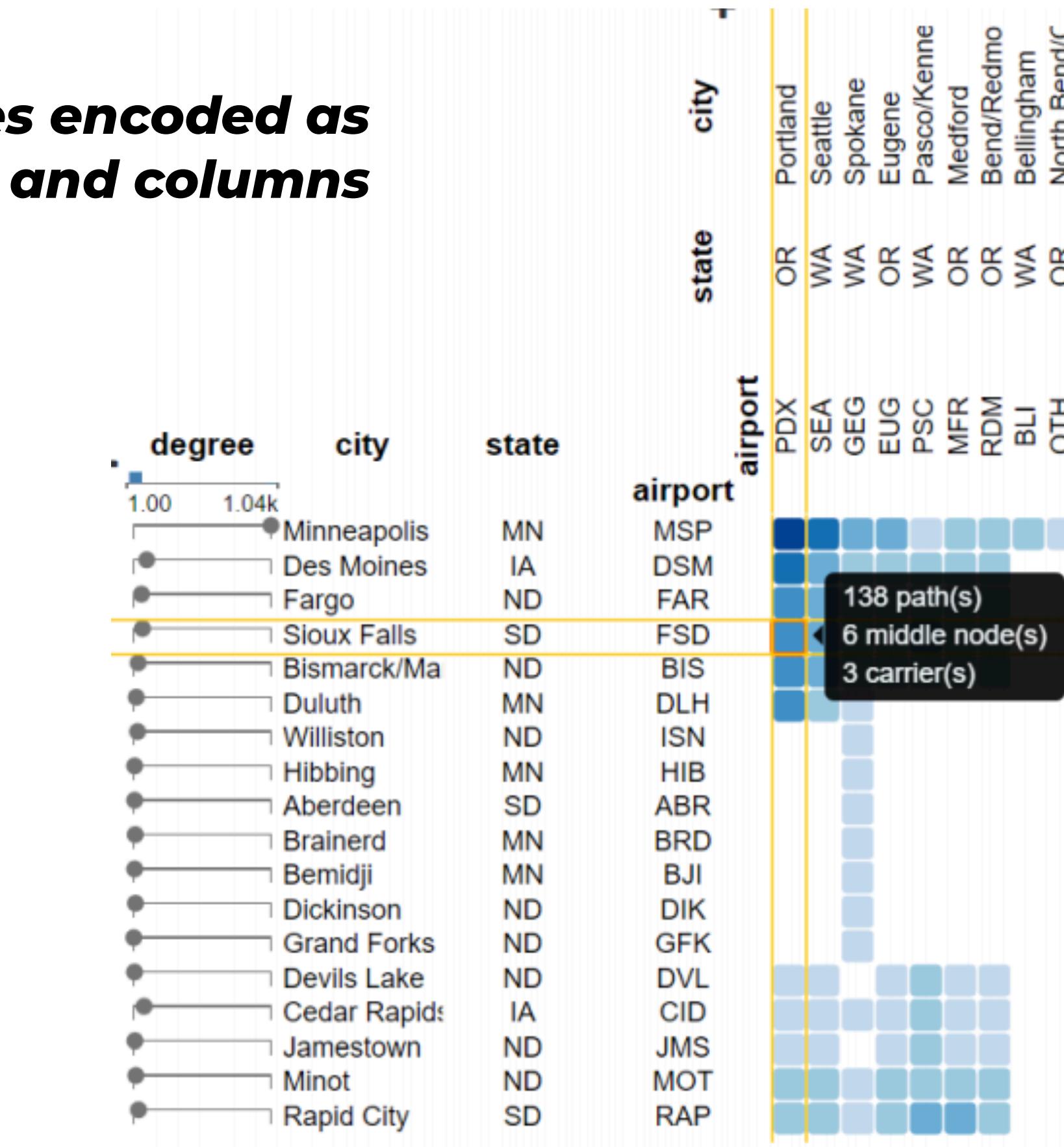
On node encoding



Matrix



Nodes encoded as rows and columns



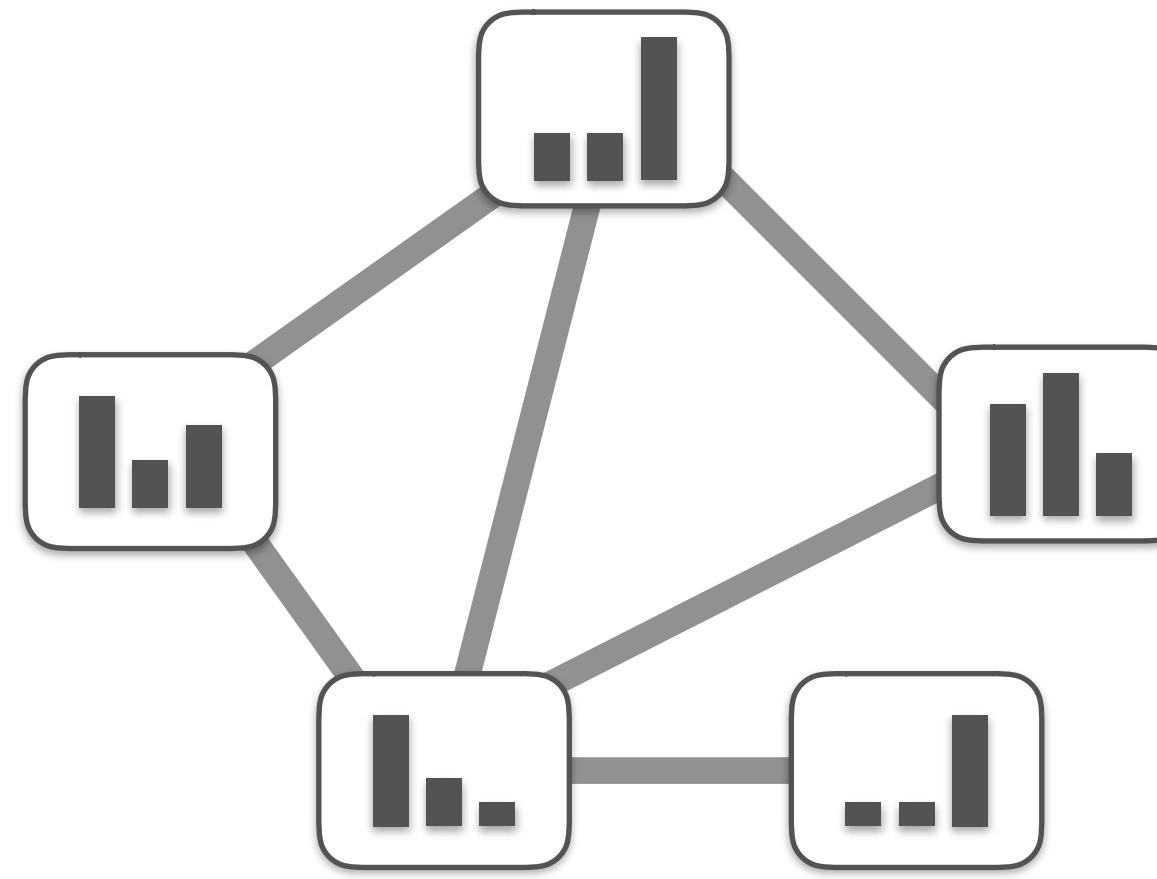
[Kerzner et al., Graffinity, 2017]

Pro: neighborhoods and clusters

Con: paths

On-Node/ On-Edge Encoding

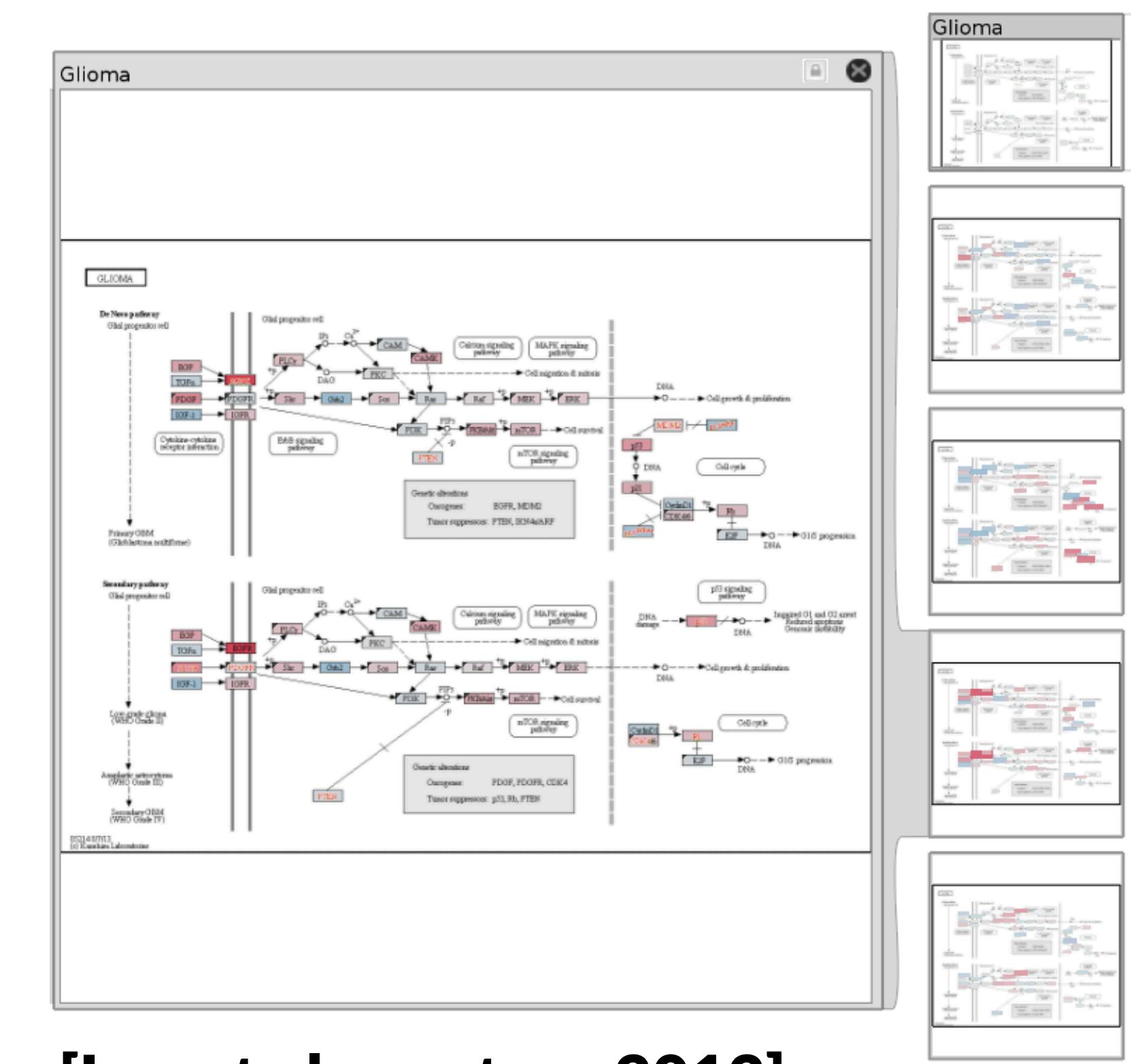
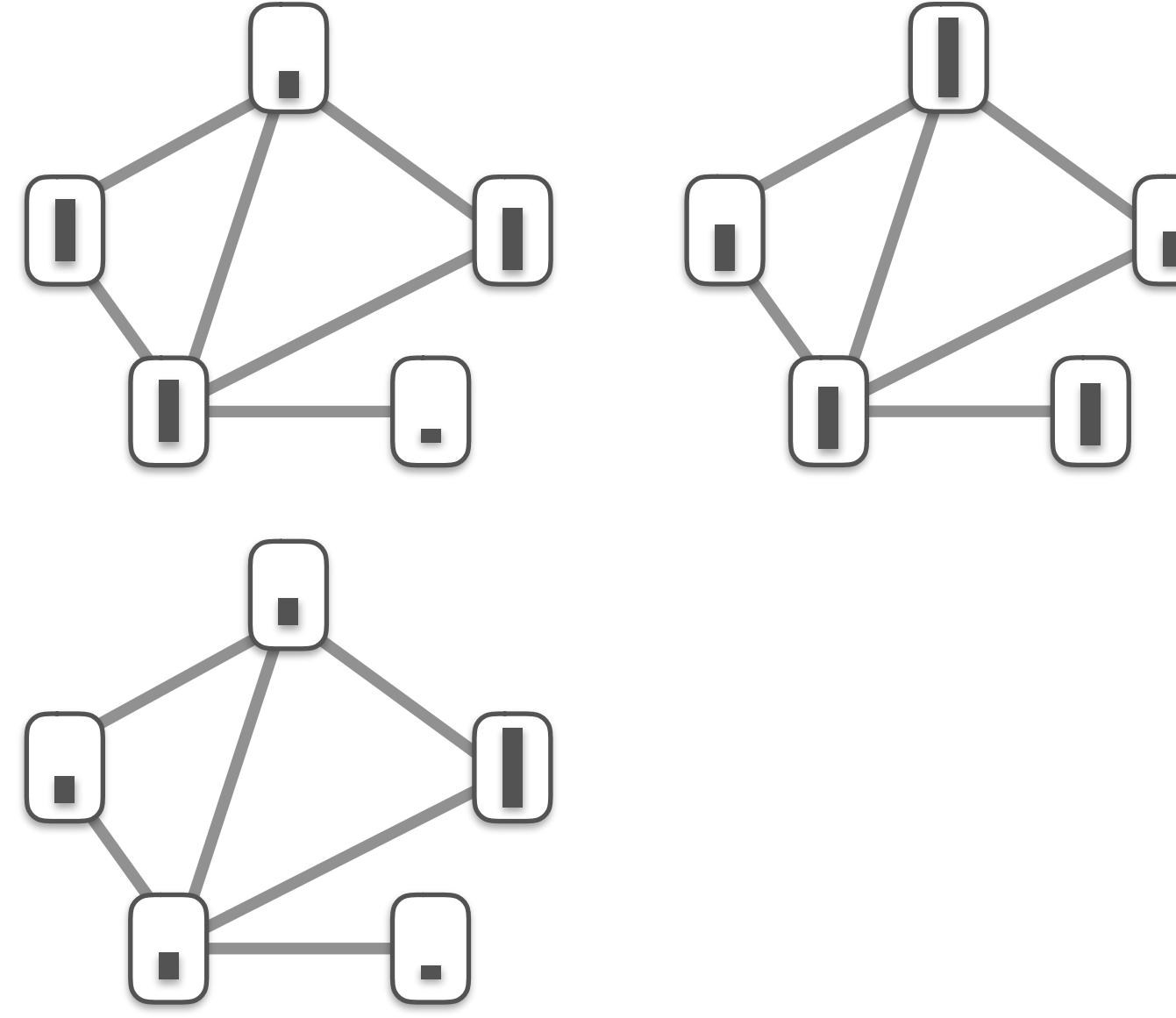
Attributes in the context of topology.



From the Interactive Tree of Life <https://itol.embl.de/>

Pro: sparse complex networks, layered networks, and trees.
Con: scalability trade-offs.

Small Multiples



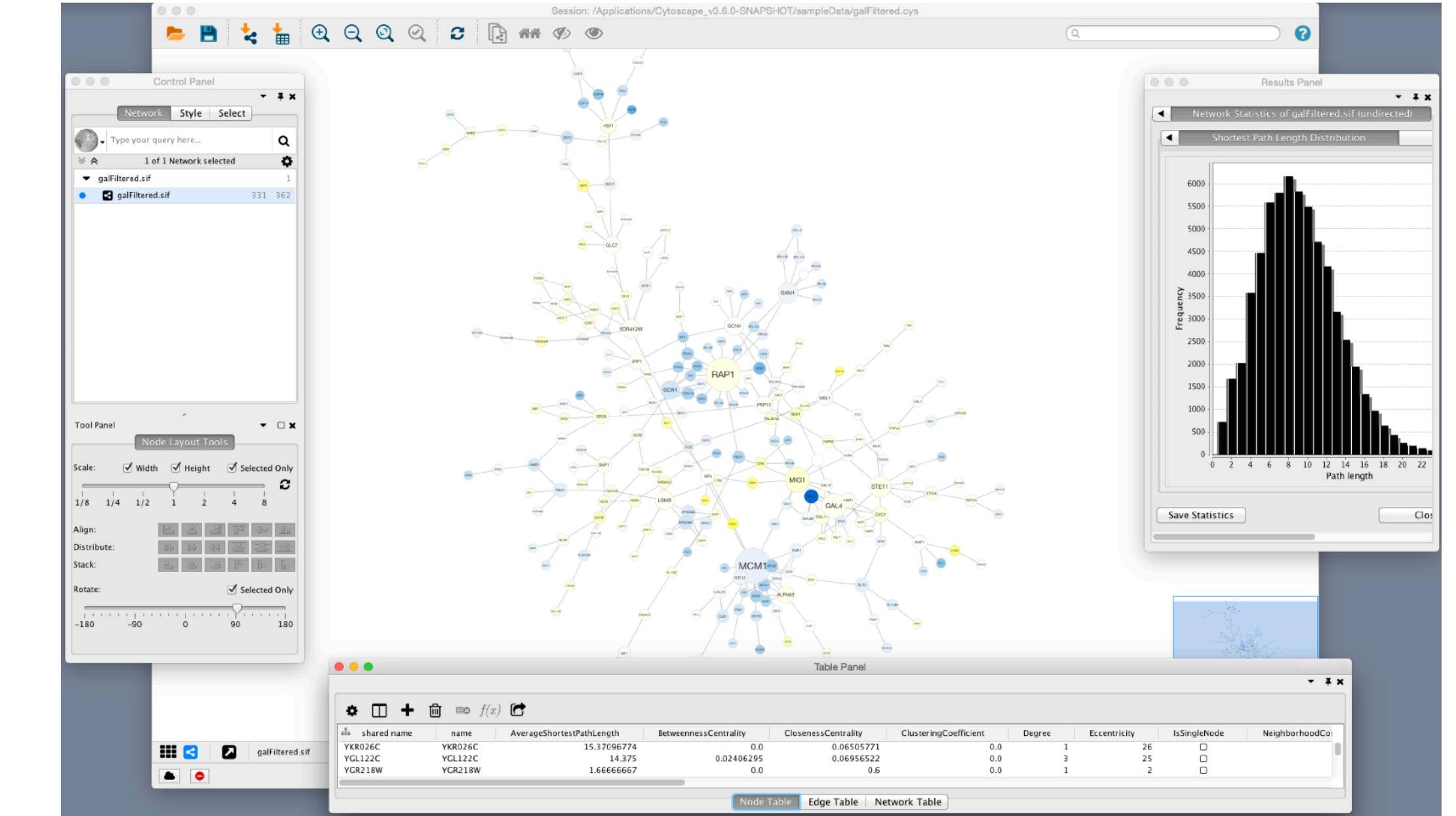
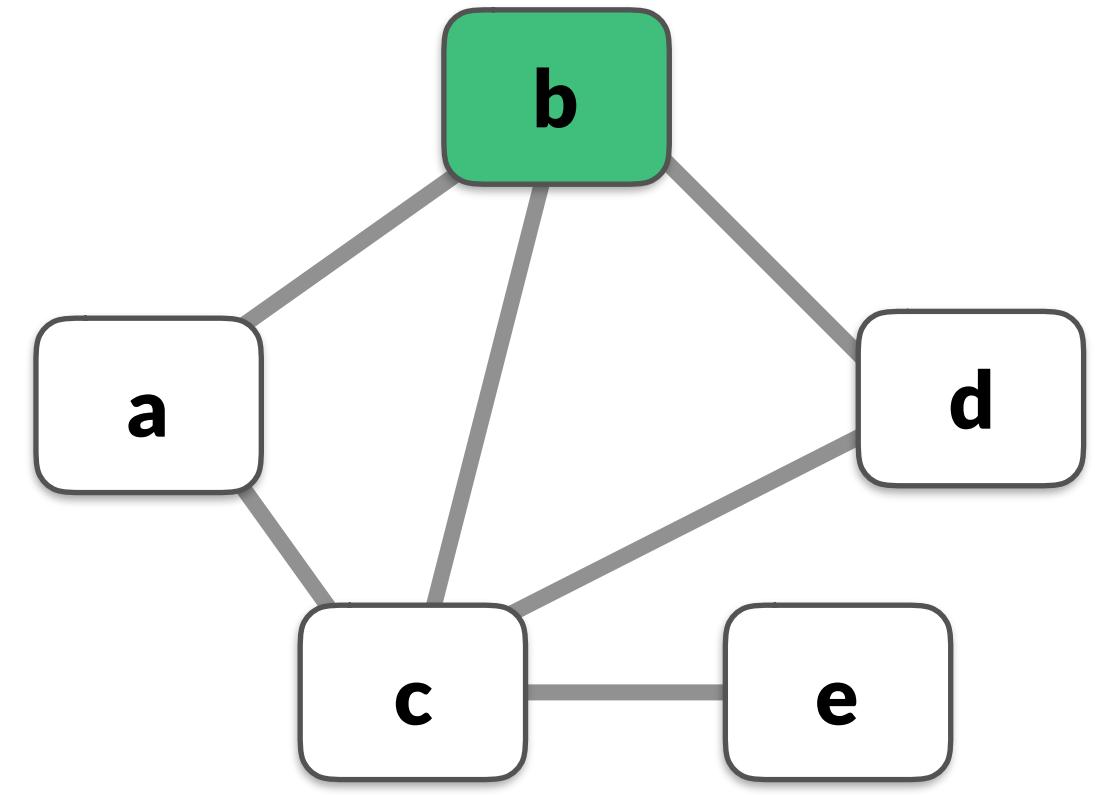
[Lex et al., contour 2012]

Often encode single attribute per view with on-node/edge encoding

Pro: comparing several attributes of a single network

Con: Scalability

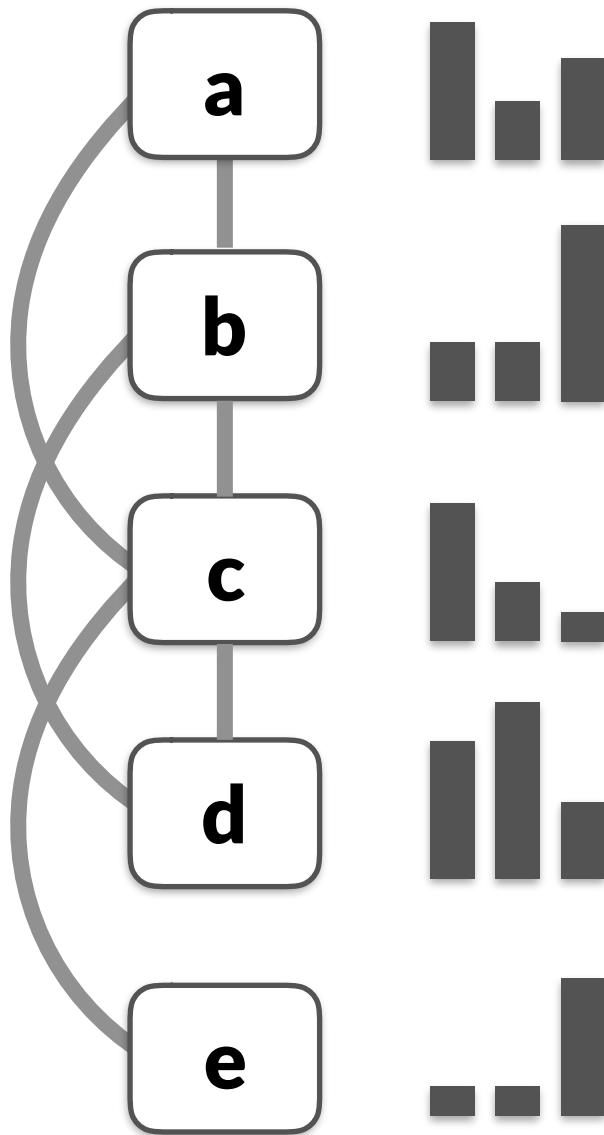
Multiple Coordinated Views



[Shannon et al., Cytoscape, 2003]

Separate views for the attributes + topology.
Juxtaposed MCV.
Links are apparent through linking and brushing

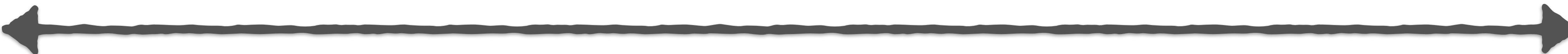
Layout Adaption



***Adapt node position in a
node-link diagram so it is
well suited for attribute
visualization***

*Layout driven
by attributes*

*Layout driven
by Topology*



Layout Adaptation: Attribute Driven Layout

[Bezerianos et al, 2010]

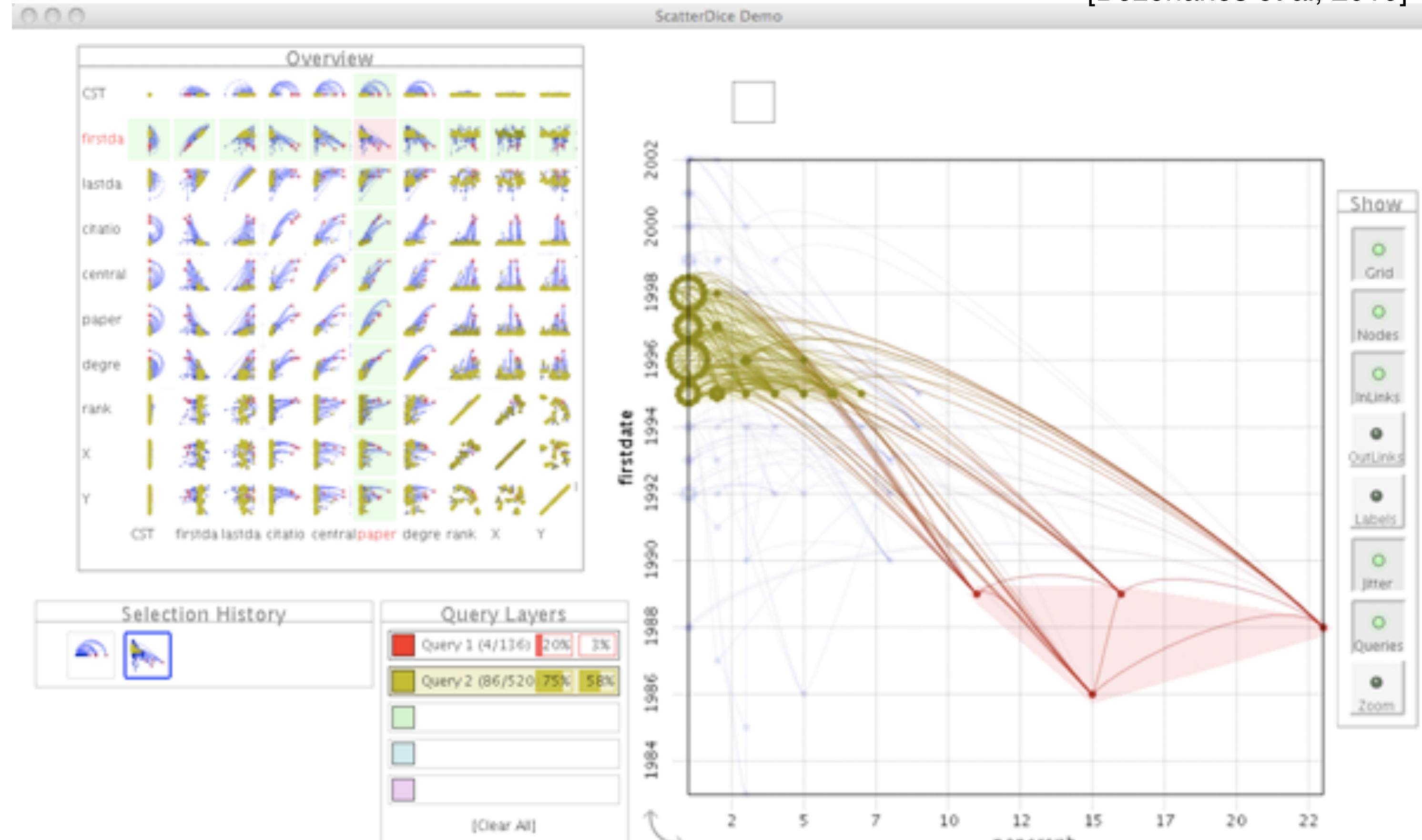
Node position defined by attribute values.

Focus on relationship of limited number of attributes (or traits)

Cons :

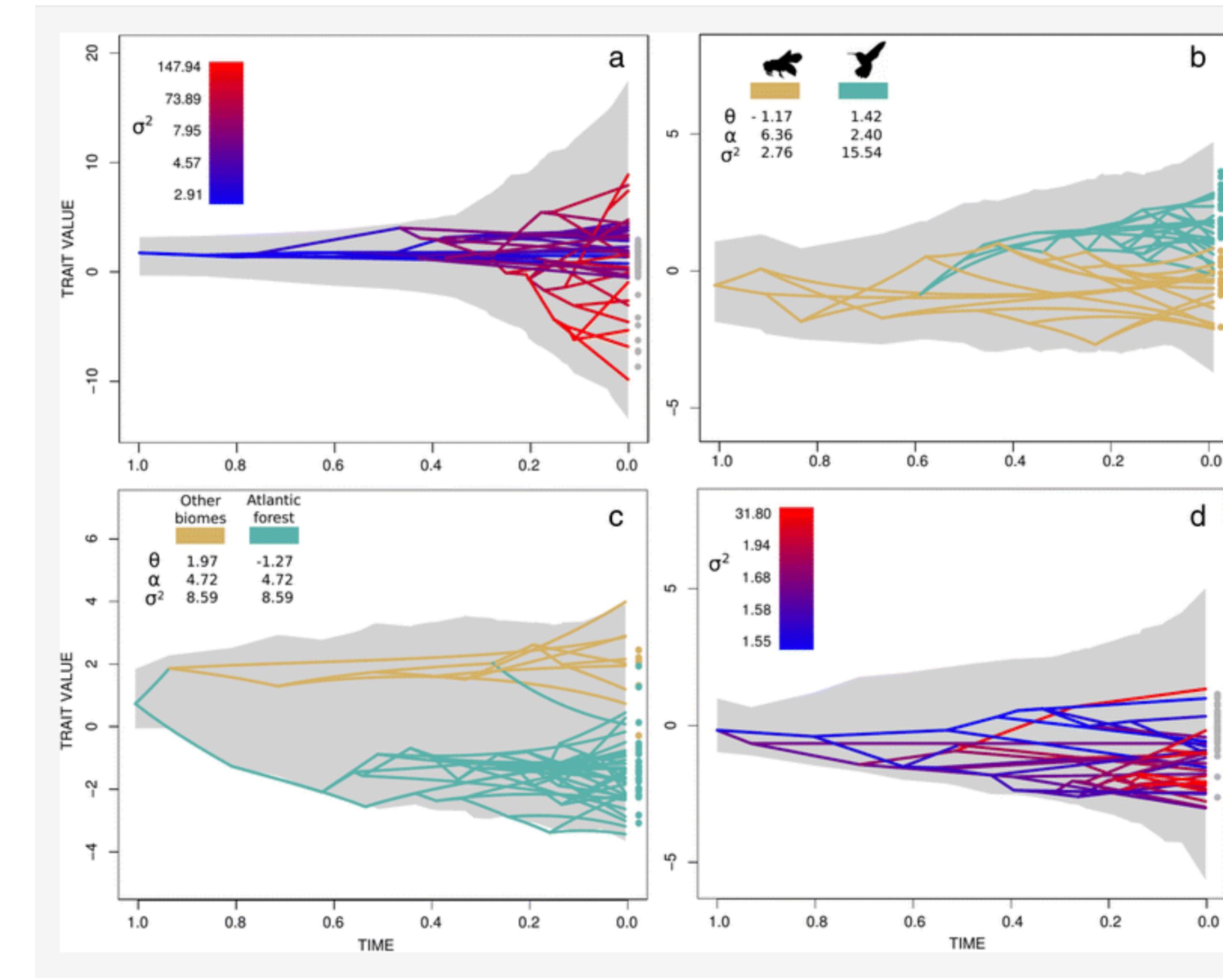
Topology hard to read

Layout driven by Topology

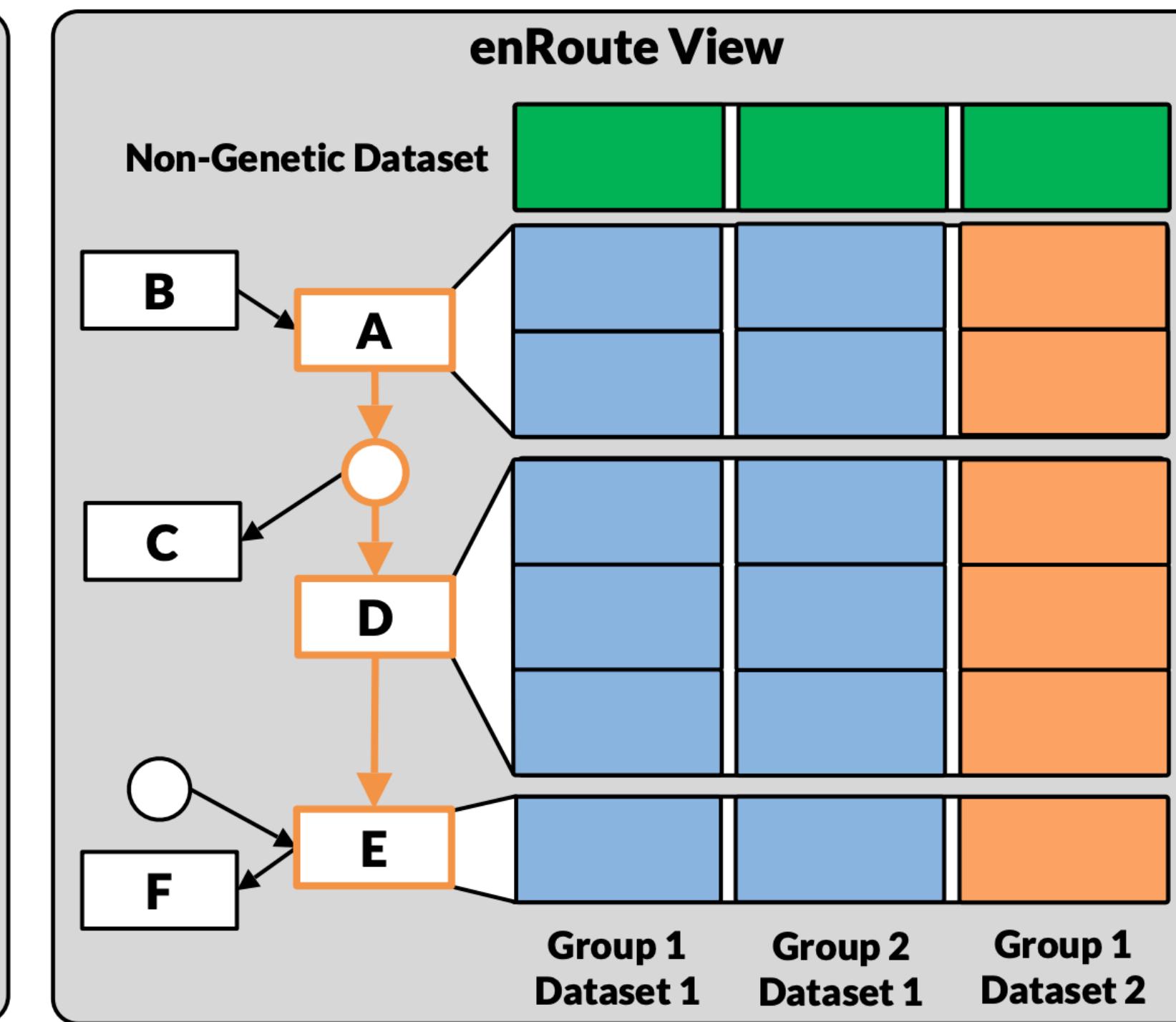
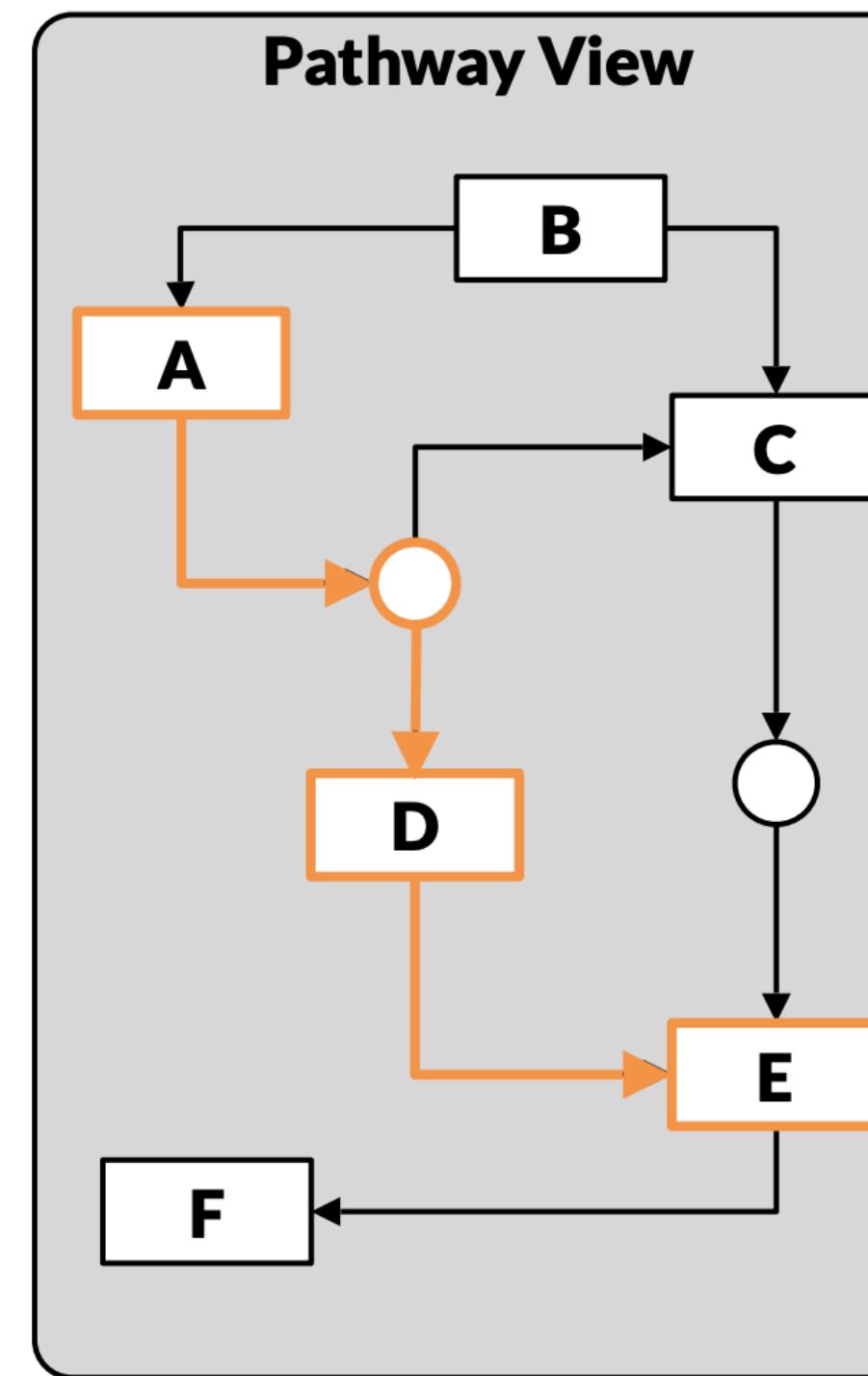


Layout driven by attributes

Layout Adaption: *Fixed Layout*



Layout Adaption: Linearization



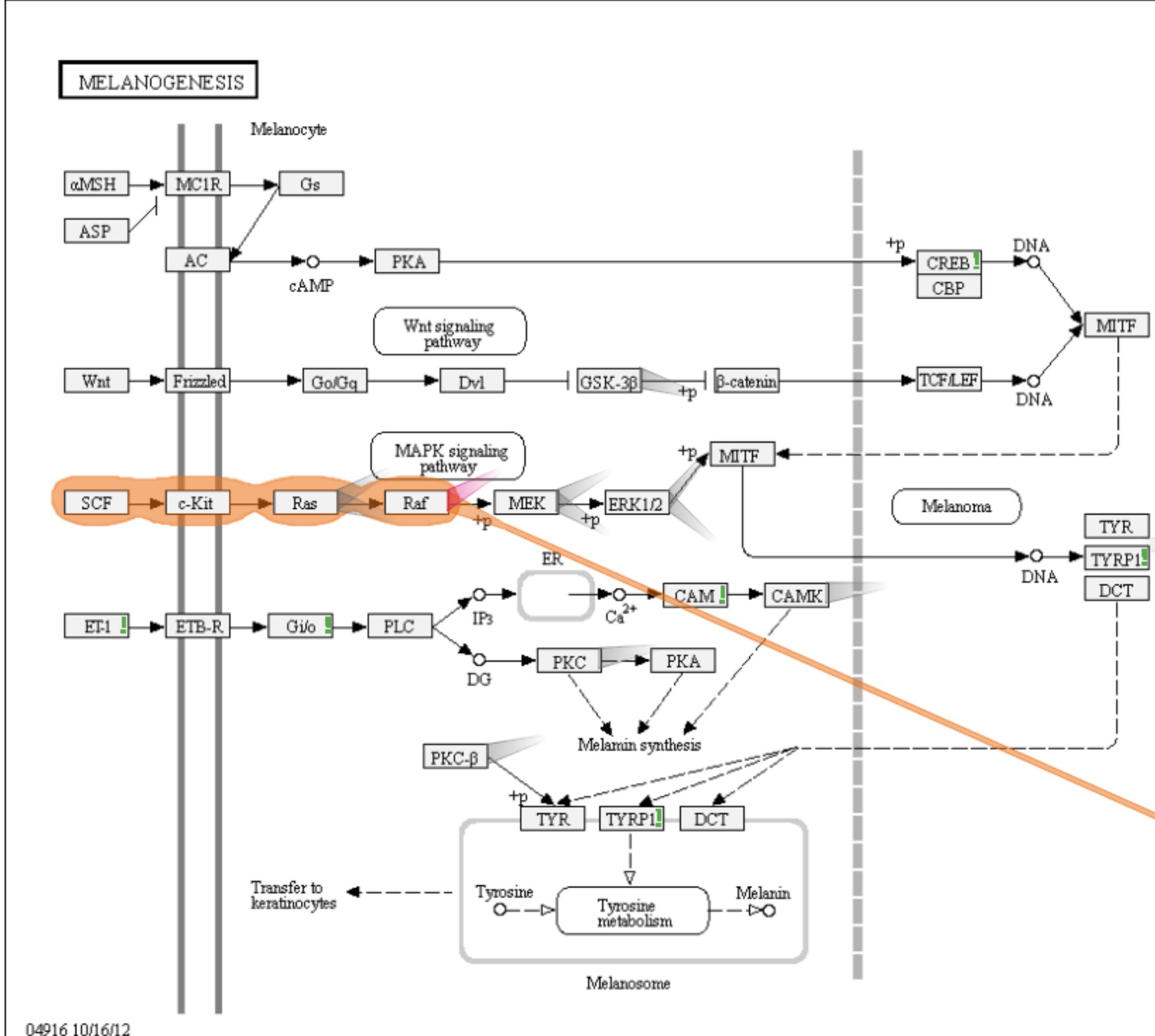
**Layout driven
by Topology**

**Layout driven
by attributes**



Linearization with enRoute

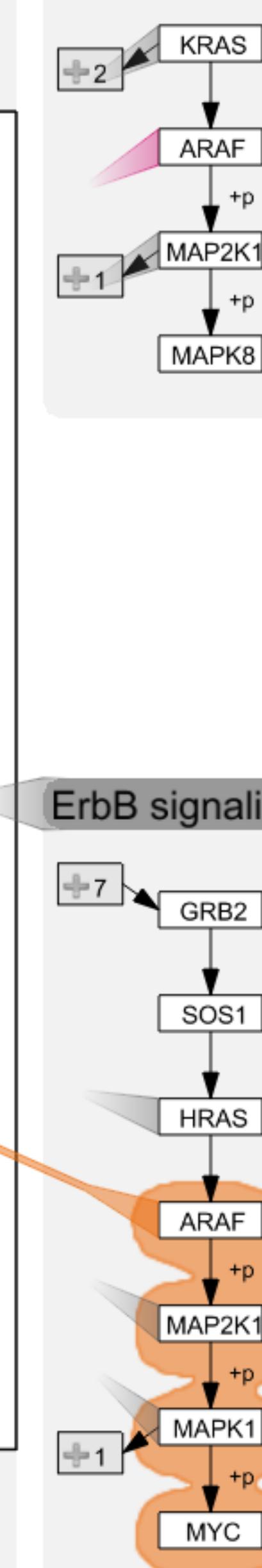
Melanogenesis



Pancreatic cancer

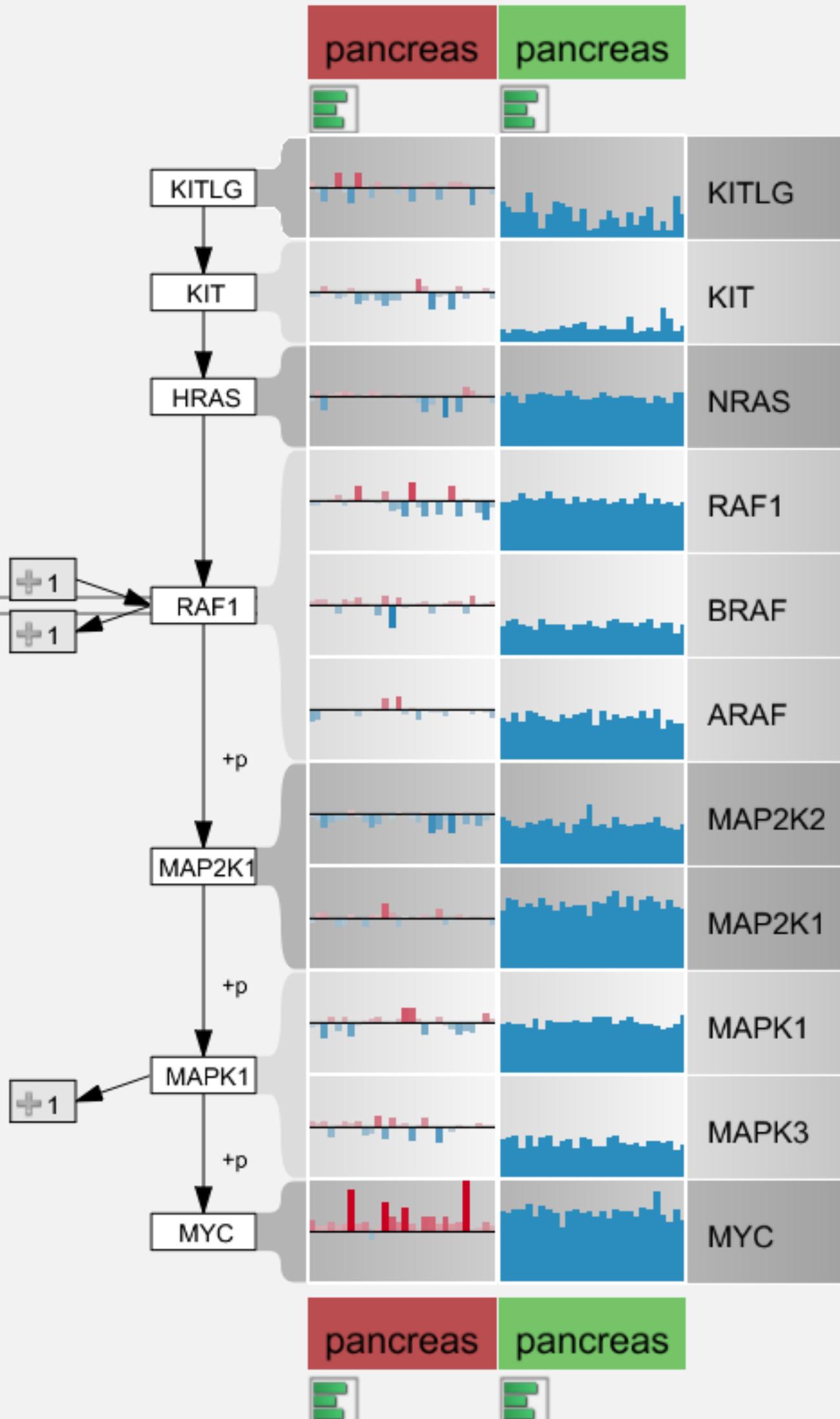
Pancreatic cancer

Selected Path



Melanogenesis

ErbB signaling pathway



*Large part of our work is
applied visualization research
in the form of design studies.*

What is a design study?

What is a design study?

Method of conducting applied visualization research

What is a design study?

**Method of conducting applied visualization research
To design systems for real-world problems**

What is a design study?

**Method of conducting applied visualization research
To design systems for real-world problems
based on the understanding and interpretation of a domain problem.**

There is more than one way to skin a cat design study

We have multiple frameworks to choose from in the literature

N. McCurdy, J. Dykes, and M. Meyer. Action design research and visualization design. In Proc. Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization, pp. 10–18. ACM, 2016.

S. McKenna, D. Mazur, J. Agutter, and M. Meyer. Design activity framework for visualization design. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2191–2200, 2014.

Slingsby, A. and Dykes, J., 2012, October. Experiences in involving analysts in visualisation design. In *Proceedings of the 2012 BELIV Workshop: Beyond Time and Errors-Novel Evaluation Methods for Visualization* (p. 1). ACM.

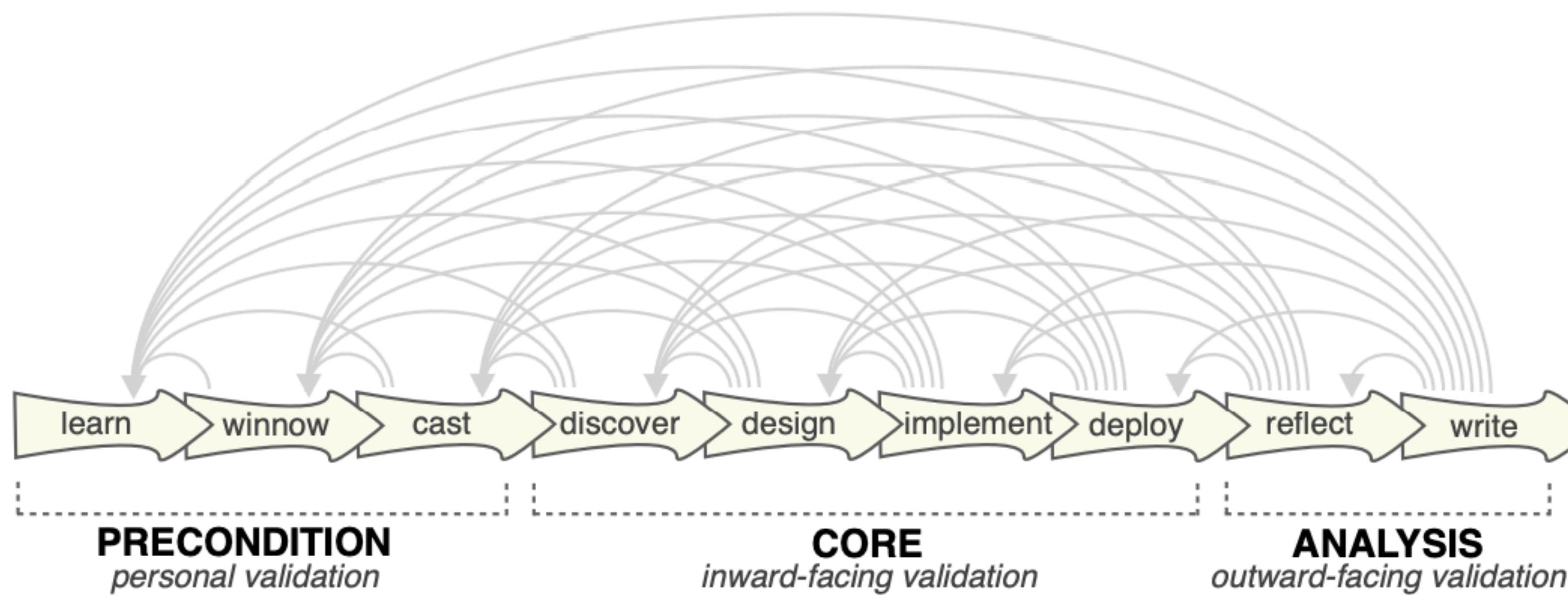
M. K. Sein, O. Henfridsson, S. Purao, M. Rossi, and R. Lindgren. Action design research. *MIS Q.*, 35(1):37–56, Mar. 2011.

M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012.

Meyer, M., Sedlmair, M., Quinan, P.S. and Munzner, T., 2015. The nested blocks and guidelines model. *Information Visualization*, 14(3), pp.234–249.

Munzner, T., 2009. A nested model for visualization design and validation. *IEEE transactions on visualization and computer graphics*, 15(6), pp.921–928.

9 Stage framework

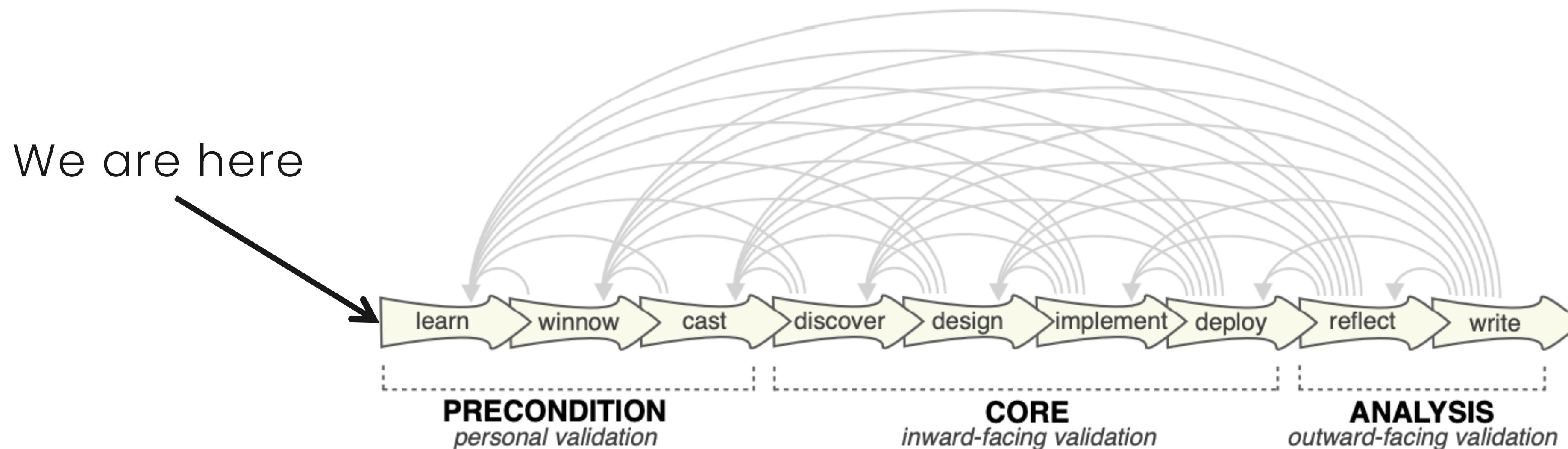


Design Study Methodology: Reflections from the Trenches and the Stacks

Michael Sedlmair, Miriah Meyer, and Tamara Munzner

IEEE Transactions on Visualization and Computer Graphics (InfoVis '12),
18(12): 2431-2440, 2012.

Life stages of a design study



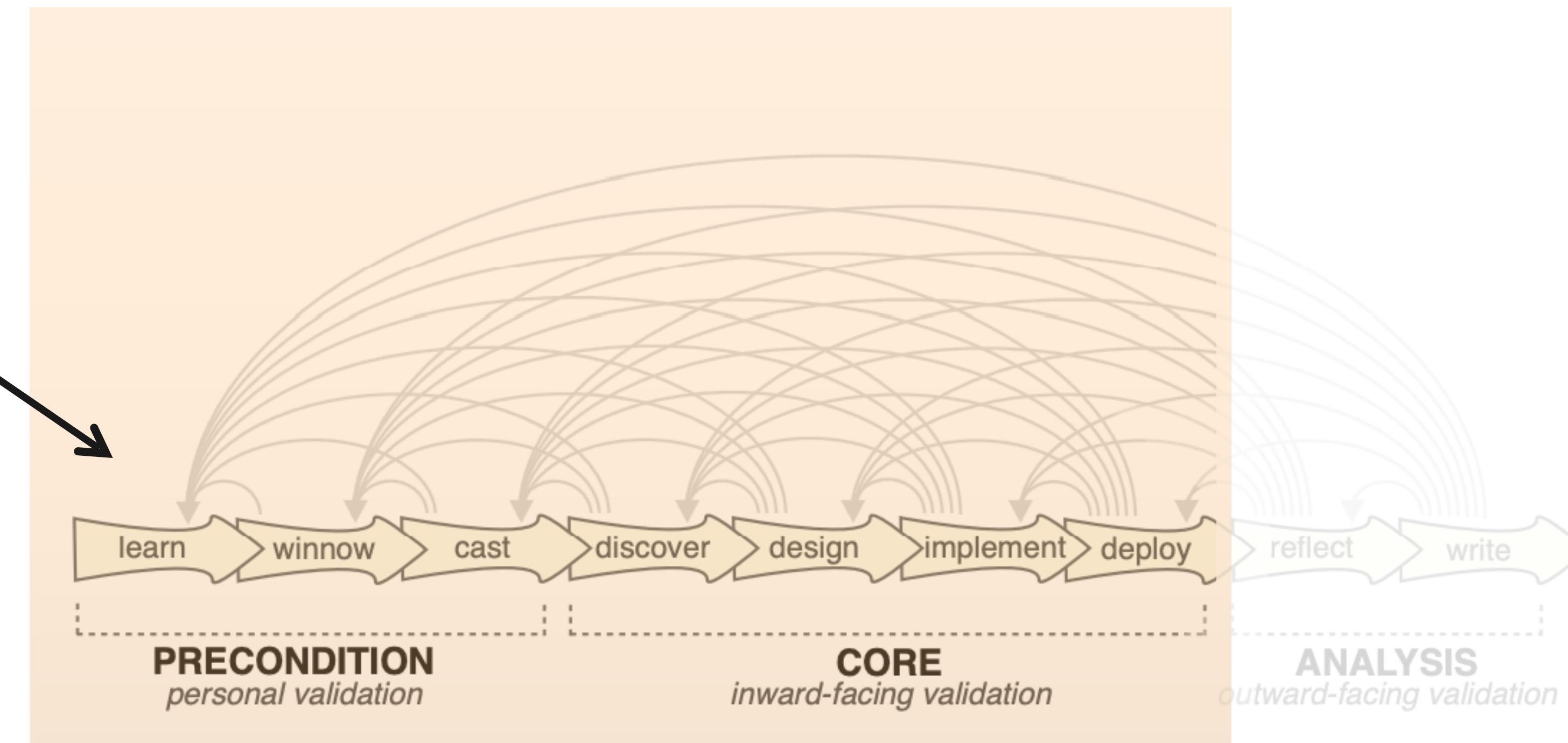
Design Study Methodology: Reflections from the Trenches and the Stacks

Michael Sedlmair, Miriah Meyer, and Tamara Munzner

IEEE Transactions on Visualization and Computer Graphics (InfoVis '12),
18(12): 2431-2440, 2012.

Life stages of a design study

We are going to focus on these



Design Study Methodology: Reflections from the Trenches and the Stacks

Michael Sedlmair, Miriah Meyer, and Tamara Munzner

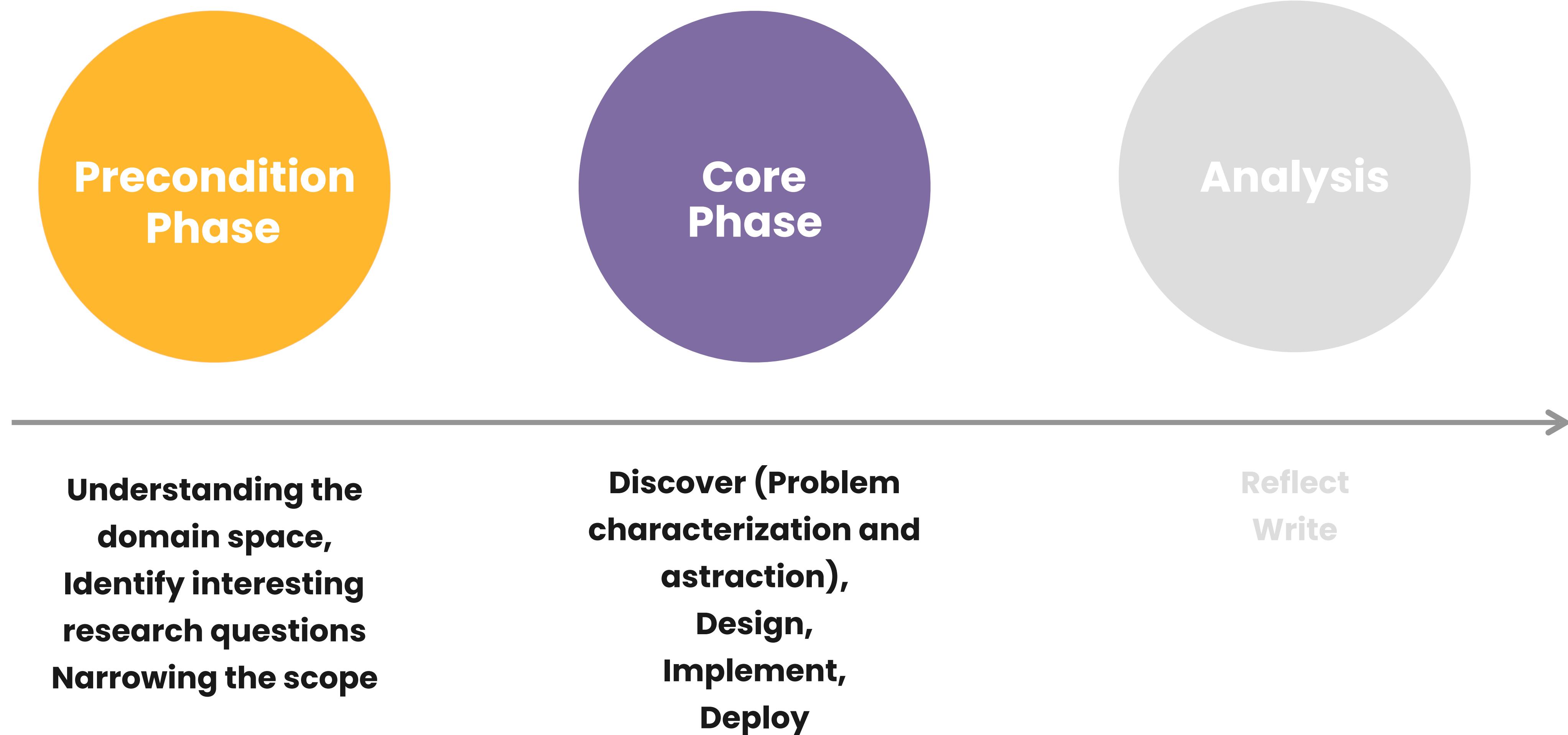
IEEE Transactions on Visualization and Computer Graphics (InfoVis '12),
18(12): 2431-2440, 2012.

Life stages of a design study

This phase centers around reflection of lessons learned and writing a research paper



Design Study Methodology: Reflections from the Trenches and the Stacks
Michael Sedlmair, Miriah Meyer, and Tamara Munzner
IEEE Transactions on Visualization and Computer Graphics (InfoVis '12),
18(12): 2431-2440, 2012.



Understanding the Domain space

“it is essential to learn about the target domain and
the

practices,

needs,

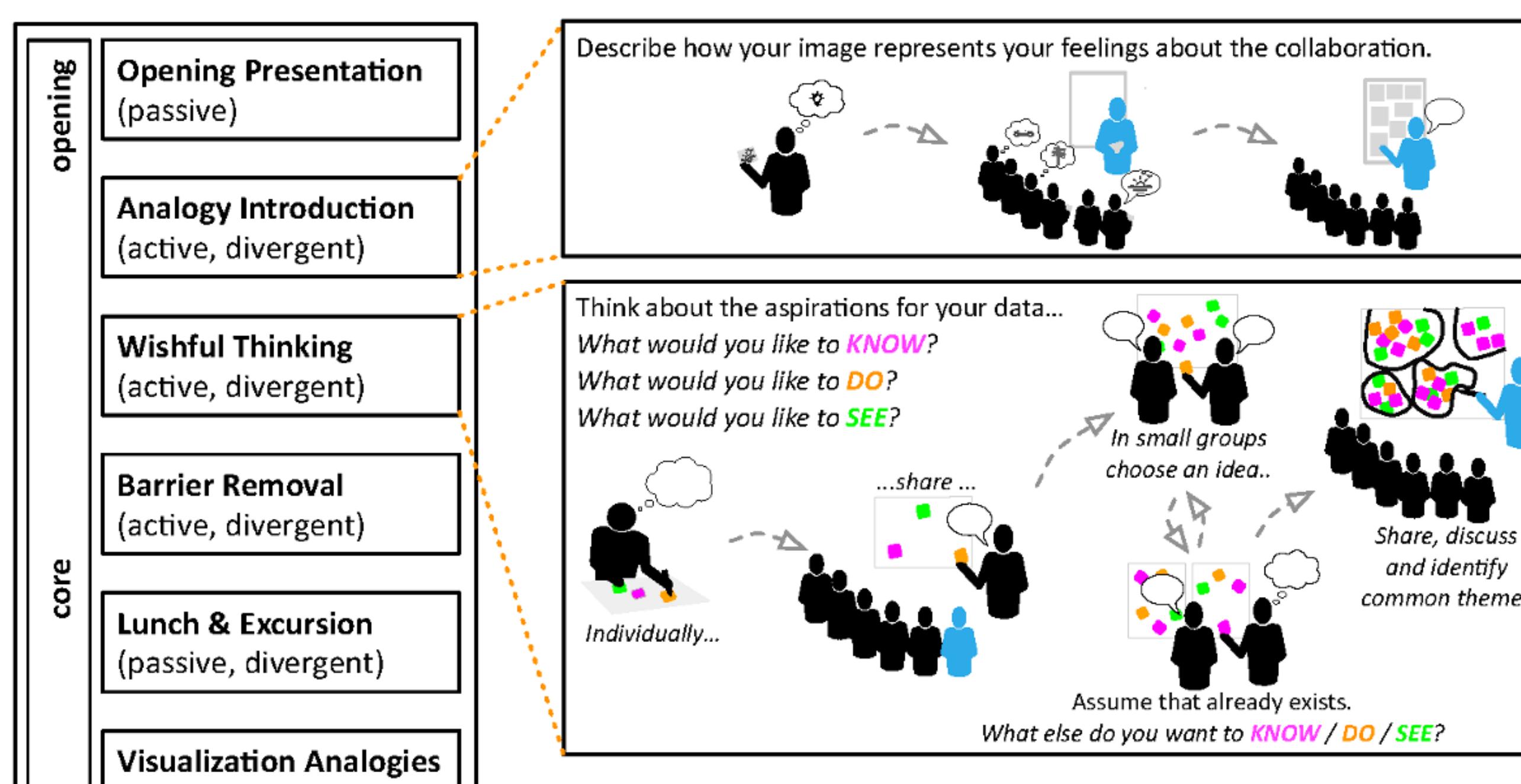
problems,

and ***requirements***

of the domain experts in order to discover
if and how visualization can enable insight and
discovery”

-Sedlmair et al (2012)

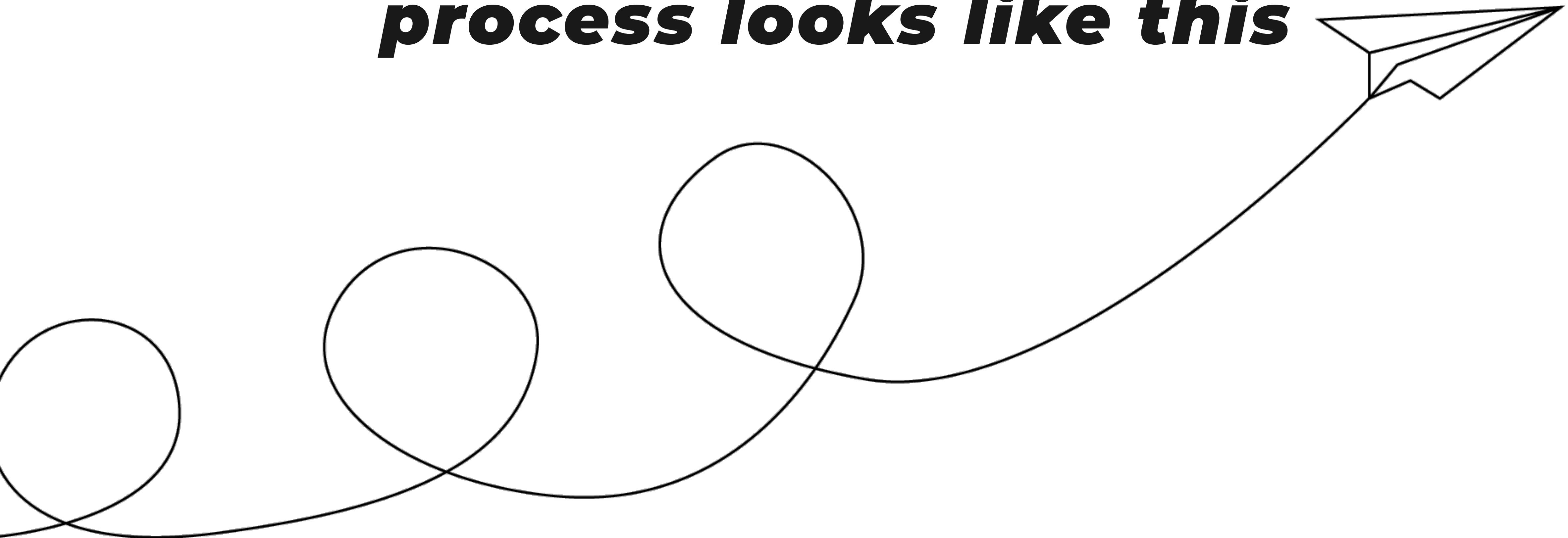
Creativity workshops method of accelerating the early foundational work of a design study.

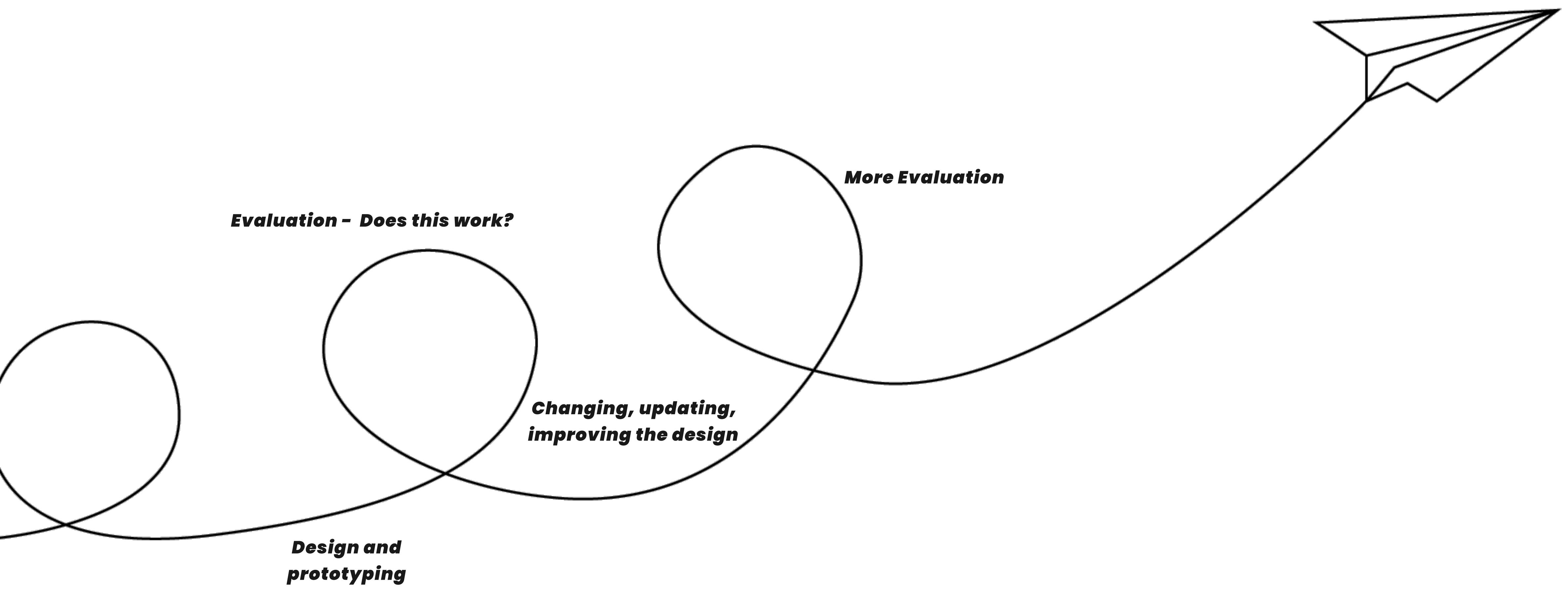


Kerzner, E., Goodwin, S., Dykes, J., Jones, S. and Meyer, M., 2019. A framework for creative visualization-opportunities workshops. *IEEE transactions on visualization and computer graphics*, 25(1), pp.748-758.

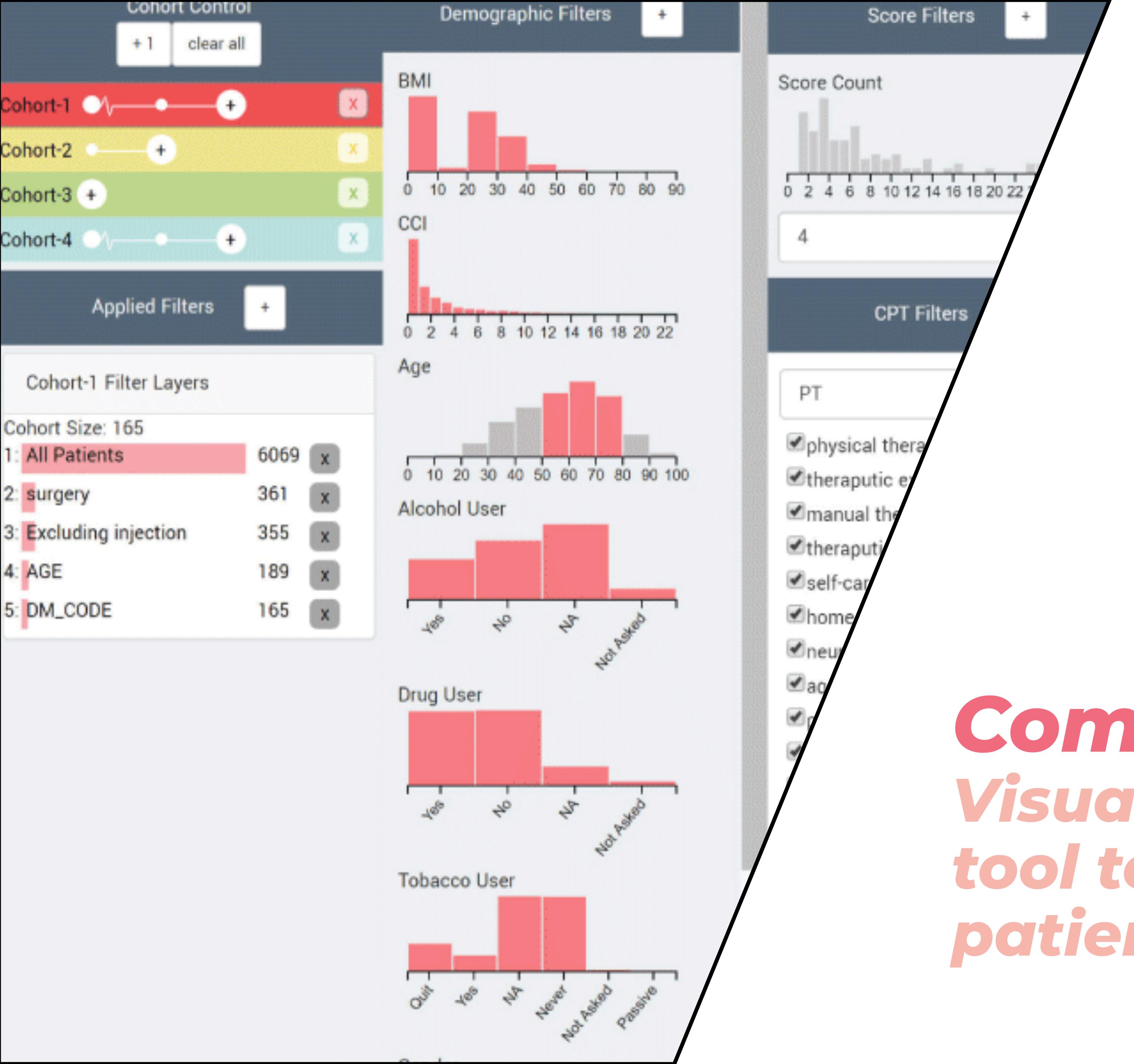
***Designing and developing is
not a linear process.***

***A lot of the time, the design
process looks like this***





*Looking at these in the context of
my first project.*



Composer

Visual cohort analysis

tool to compare patient outcomes



***collaboration with Orthopedic
surgeons and researchers at
the University of Utah's
Orthopedic Center.***

Understanding the Domain space

Many of our collaborators' patients are seeking treatment for herniated discs



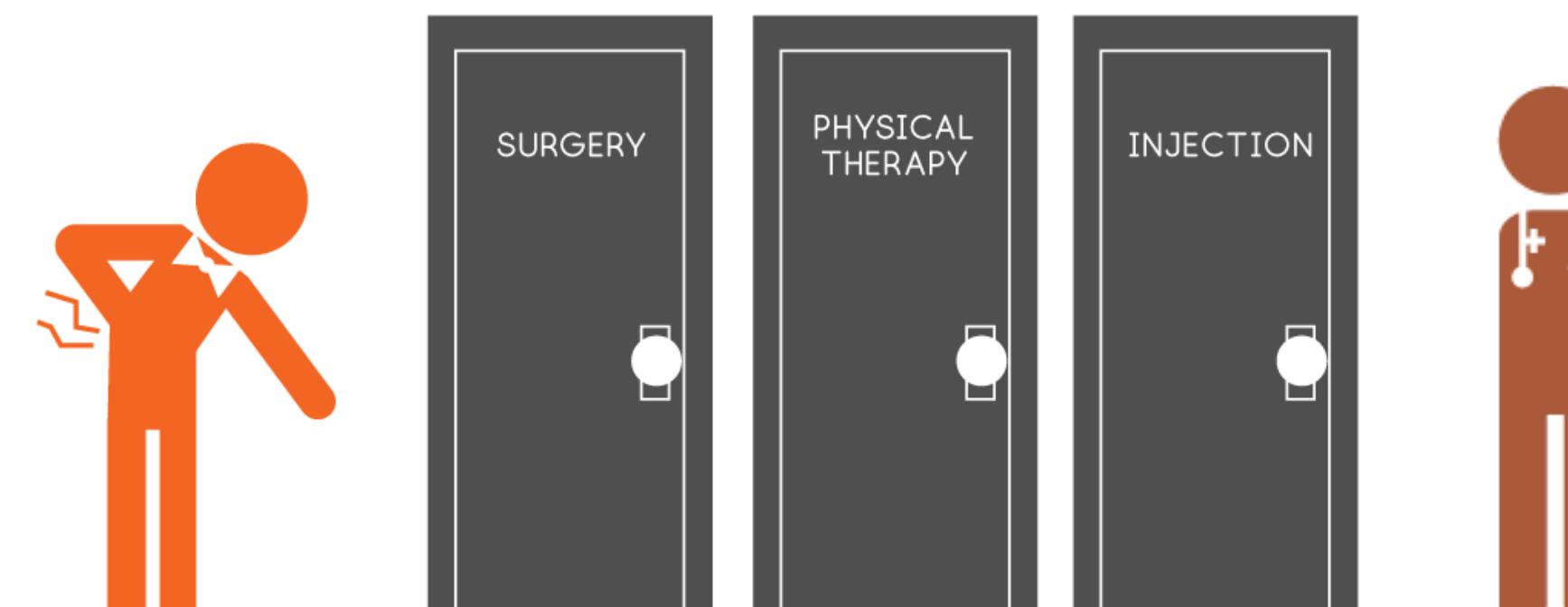
Understanding the Domain space

***Herniated discs can cause back pain
and inhibition of physical function***



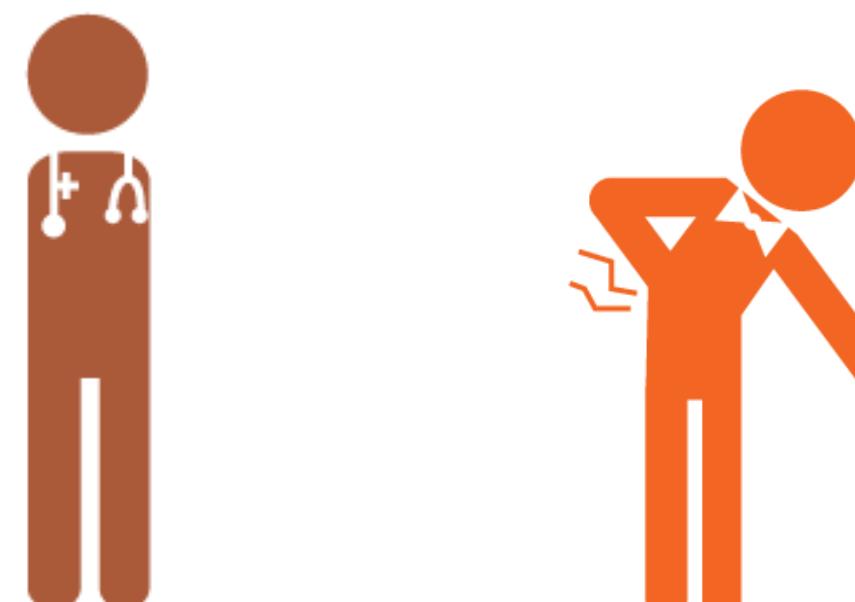
Understanding the Domain space

Main treatment options:
Surgery, injection, and physical therapy



Understanding the Domain space

Surgeons need to determine the best treatment options for their patients





Understanding Current practices..

***Rely on clinical guidelines and past
experience..***



Identifying potential Limitations..

Most clinical guidelines, based on data from a general population, may not provide an accurate reflection of potential outcomes for patients with pre-existing conditions.



And opportunities to address these limitations..

Cohort analysis has been used for identifying factors that could affect patient recovery and Treatment.



**And
opportunities to address
these *limitations*..**

Cohort analysis has been used for
identifying factors that could affect patient
recovery and Treatment.

**Subset of the general
population that share
common characteristics**



And opportunities to address these limitations..

Lacked tools to compare different treatment outcomes of cohorts that they could define from demographic information and medical histories



***Area that we
can help with***

***And
opportunities to address
these limitations..***

***Lacked tools to compare different
treatment outcomes of cohorts that they
could define from demographic
information and medical histories***

Understanding User needs..



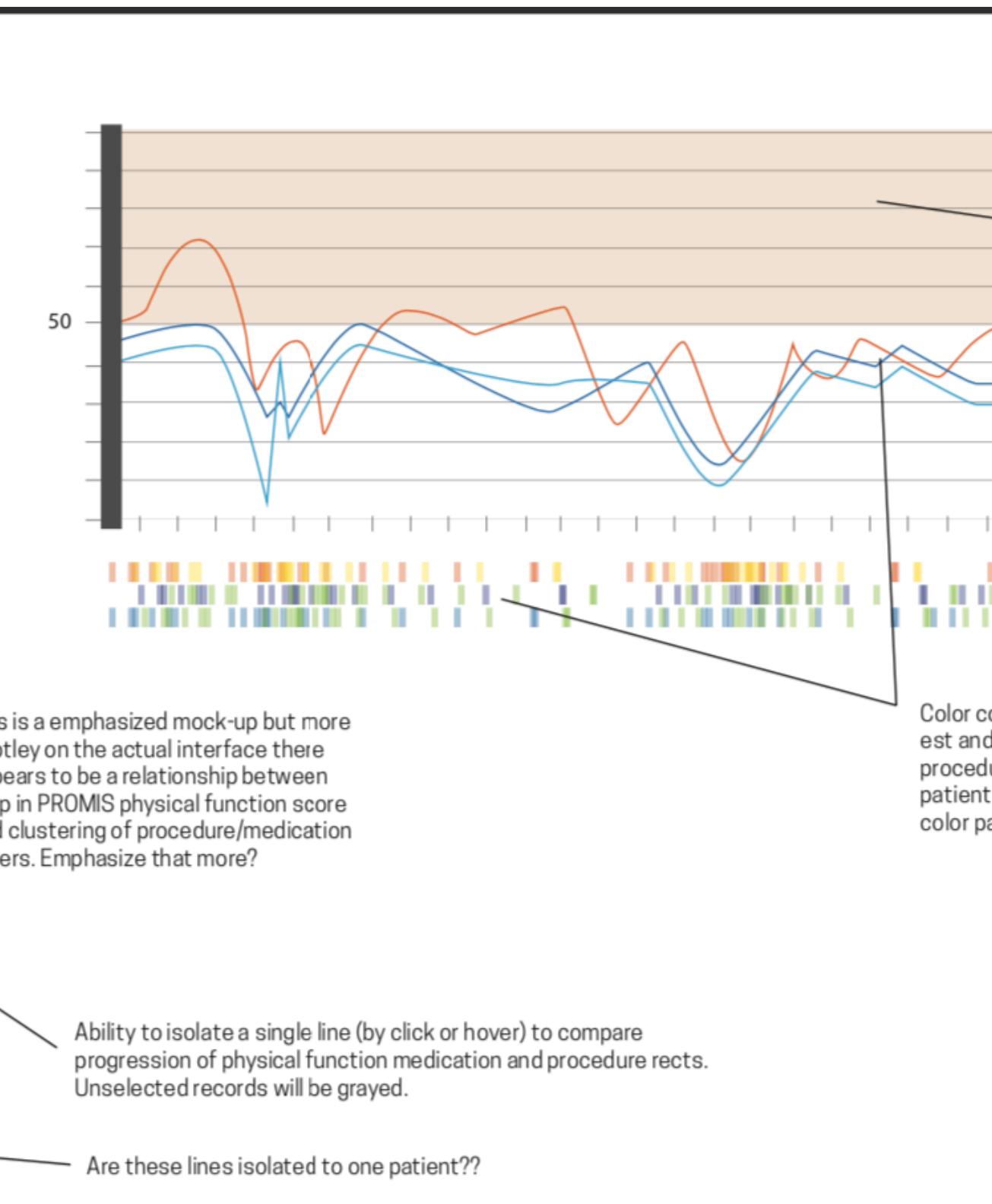
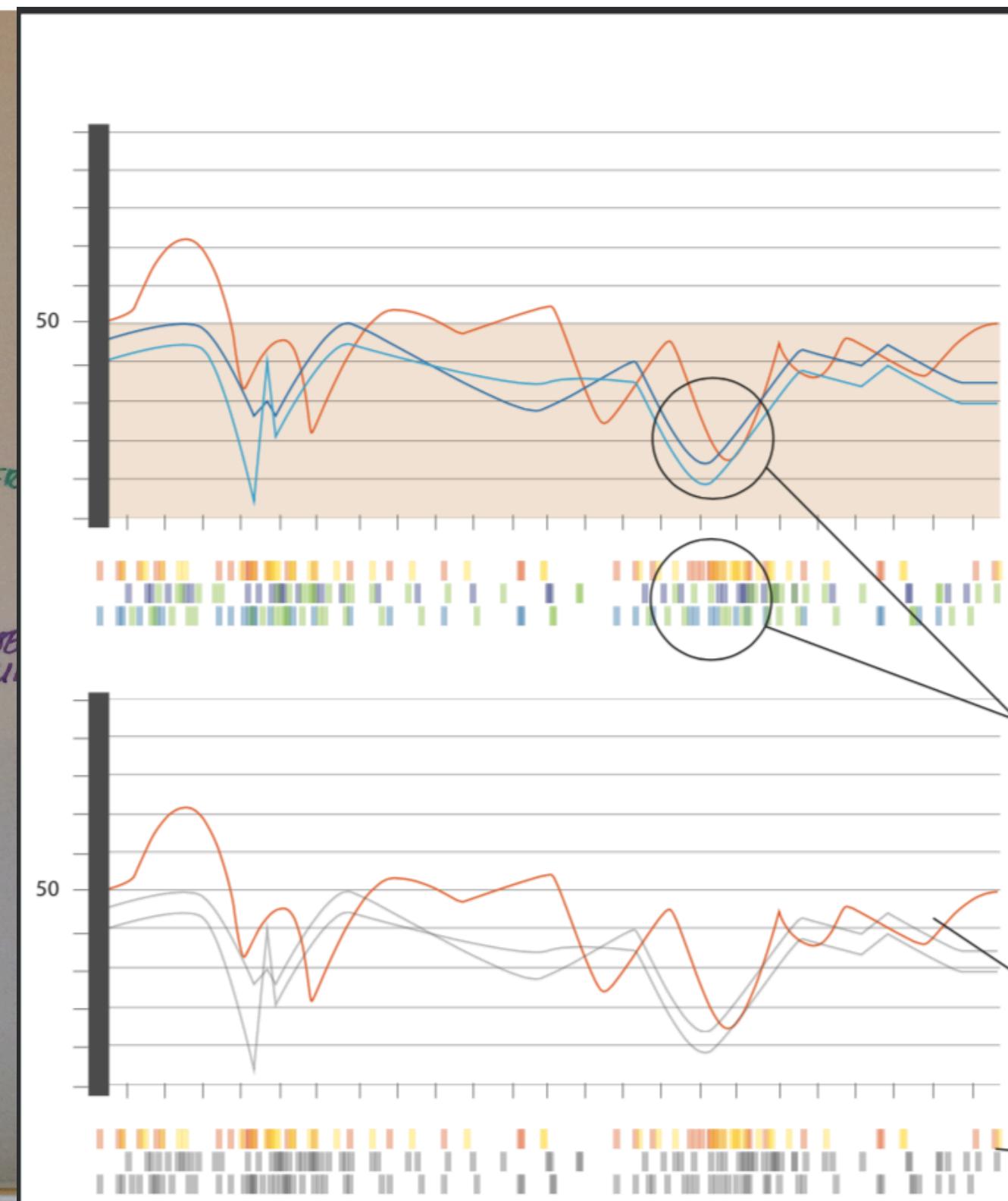
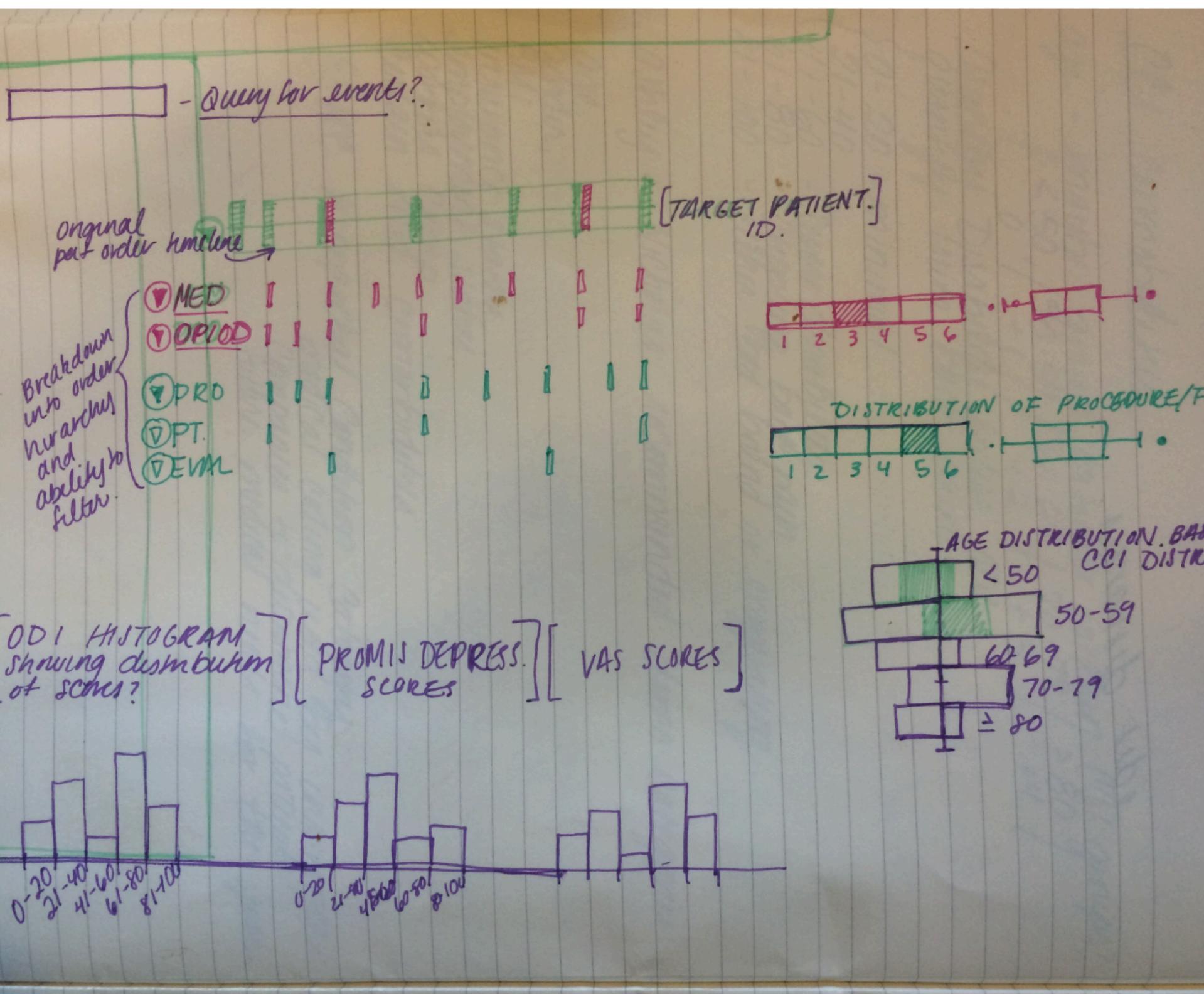
Ability to evaluate patient progress and predict how a given patient would progress after treatment, especially for patients with pre-existing conditions

Defined requirements to meet the users needs..

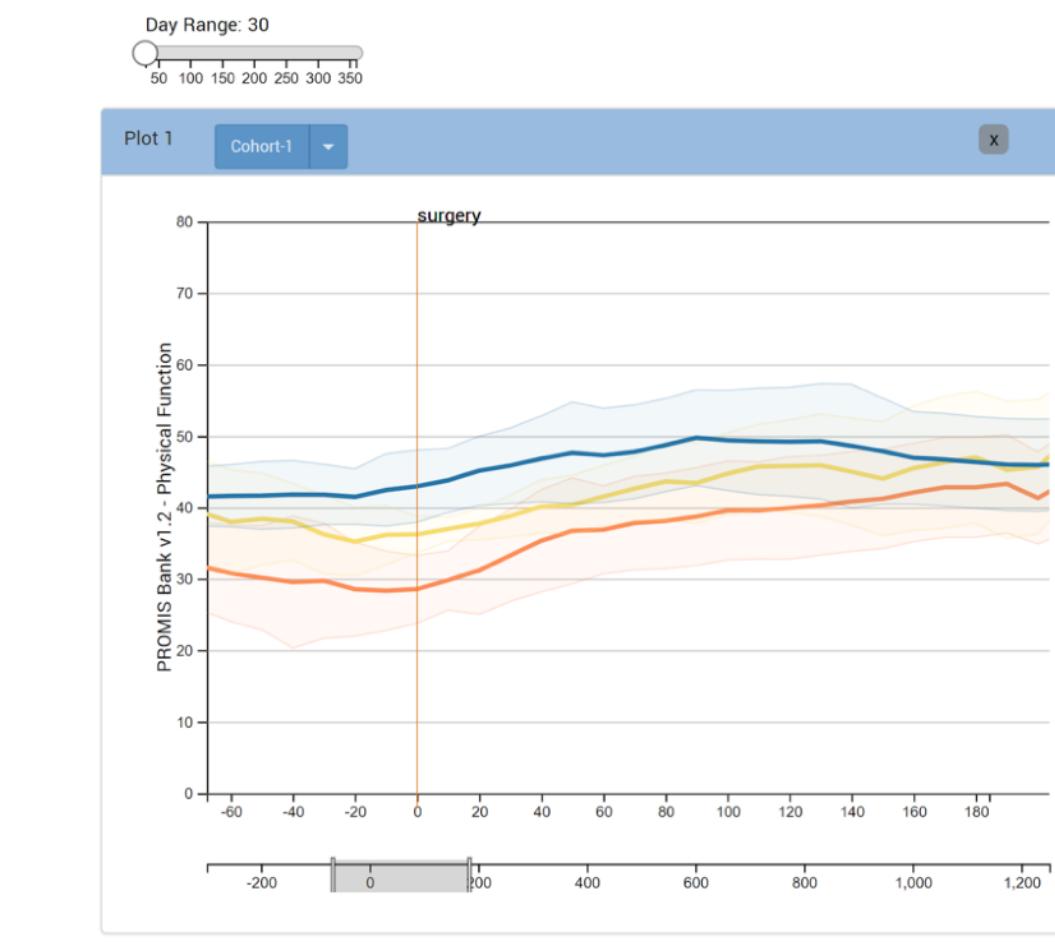
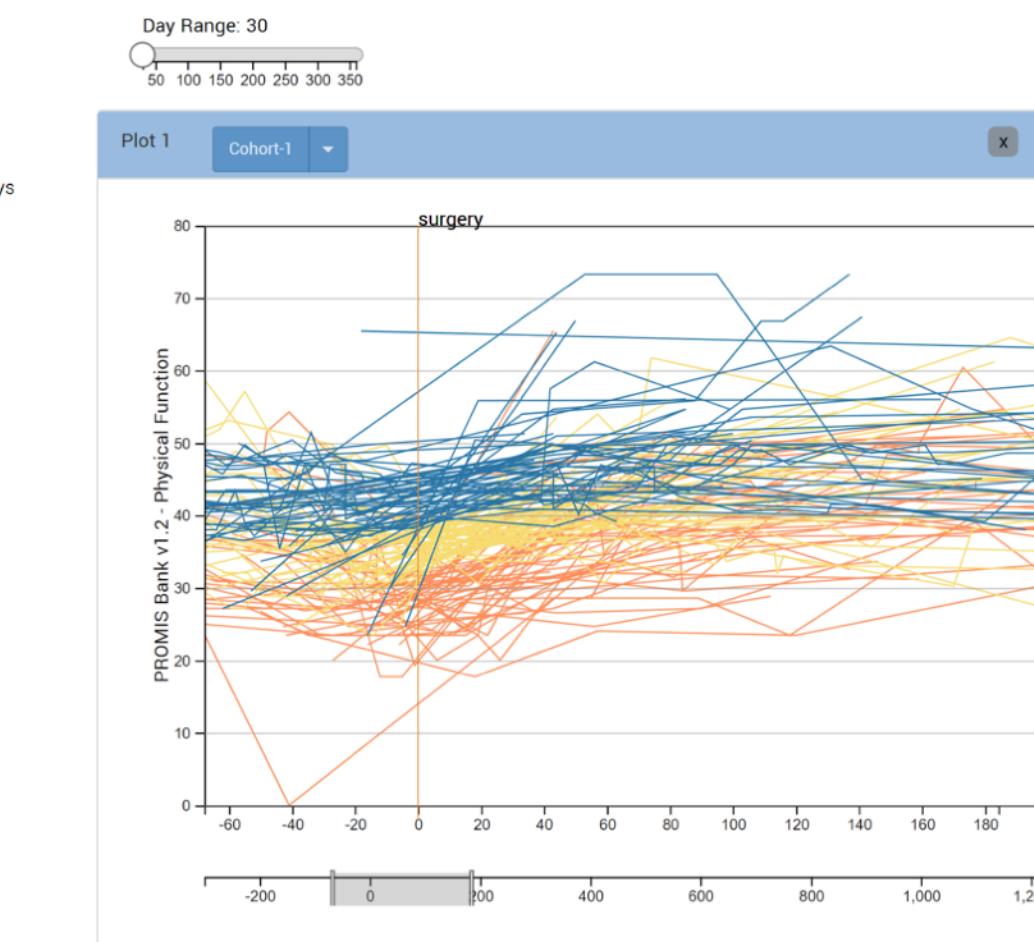
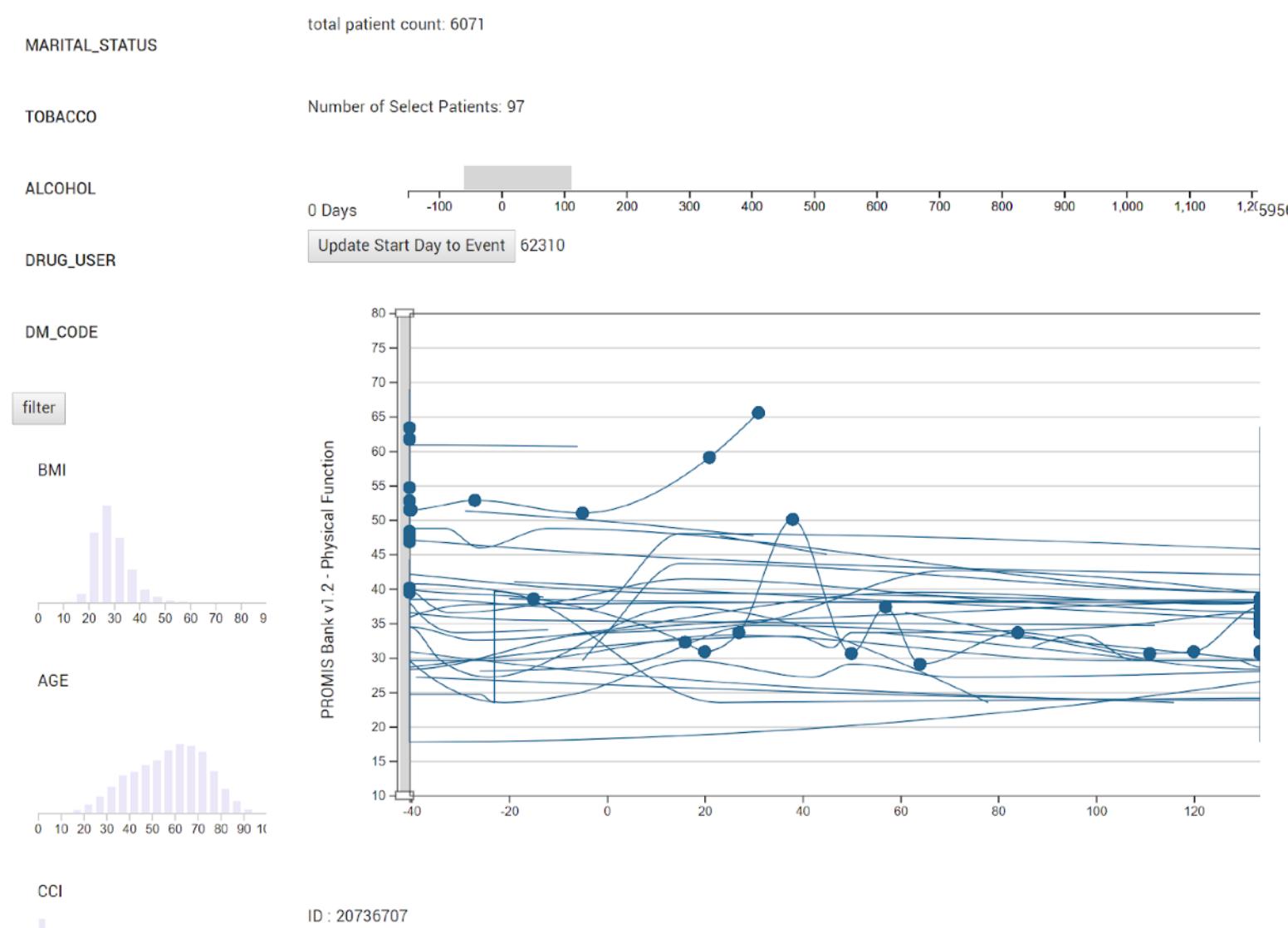
- 1. Define meaningful cohorts of patients**
- 2. Compare outcomes of different cohorts**
- 3. Compare outcomes of different treatments**

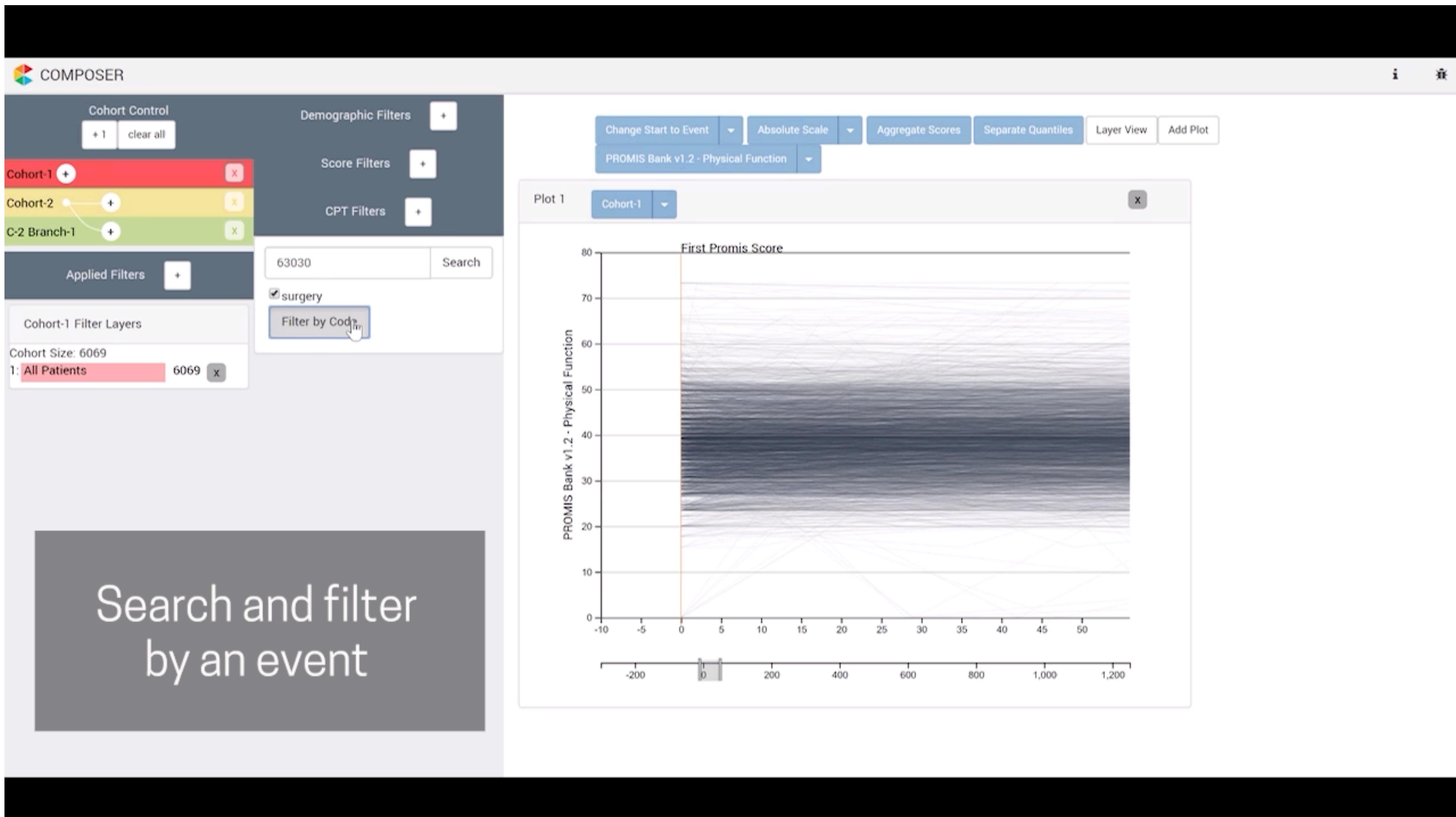
**Need: ability to evaluate patient progress and
predict how a given patient would progress after treatment**

Design began with sketching and prototyping to develop functionality that would support these requirements:

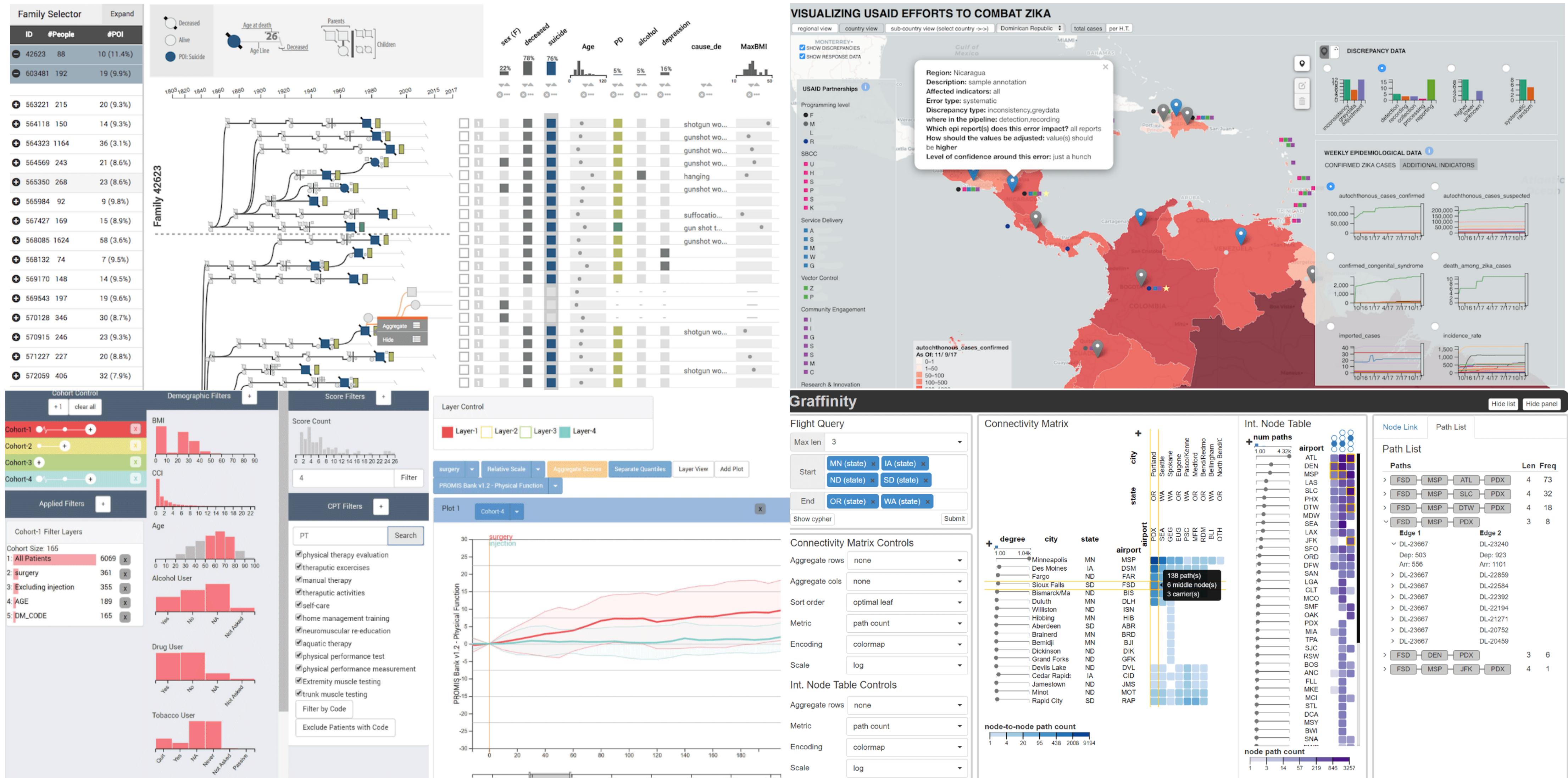


met with our collaborators on a bi-weekly basis to go over design iterations





Design studies from the lab..



These are some of the design studies from the lab. Carolin's tool in the context of Zika reporting. Jimmys work to deelope an air monitor that captures annotations. Ethans work to develop a tool for connectivity.

air monitor that supports user annotations
Ethans work with Neuroscientists to develop a tool to summarize graph connectivity

Graffinity

Flight Query

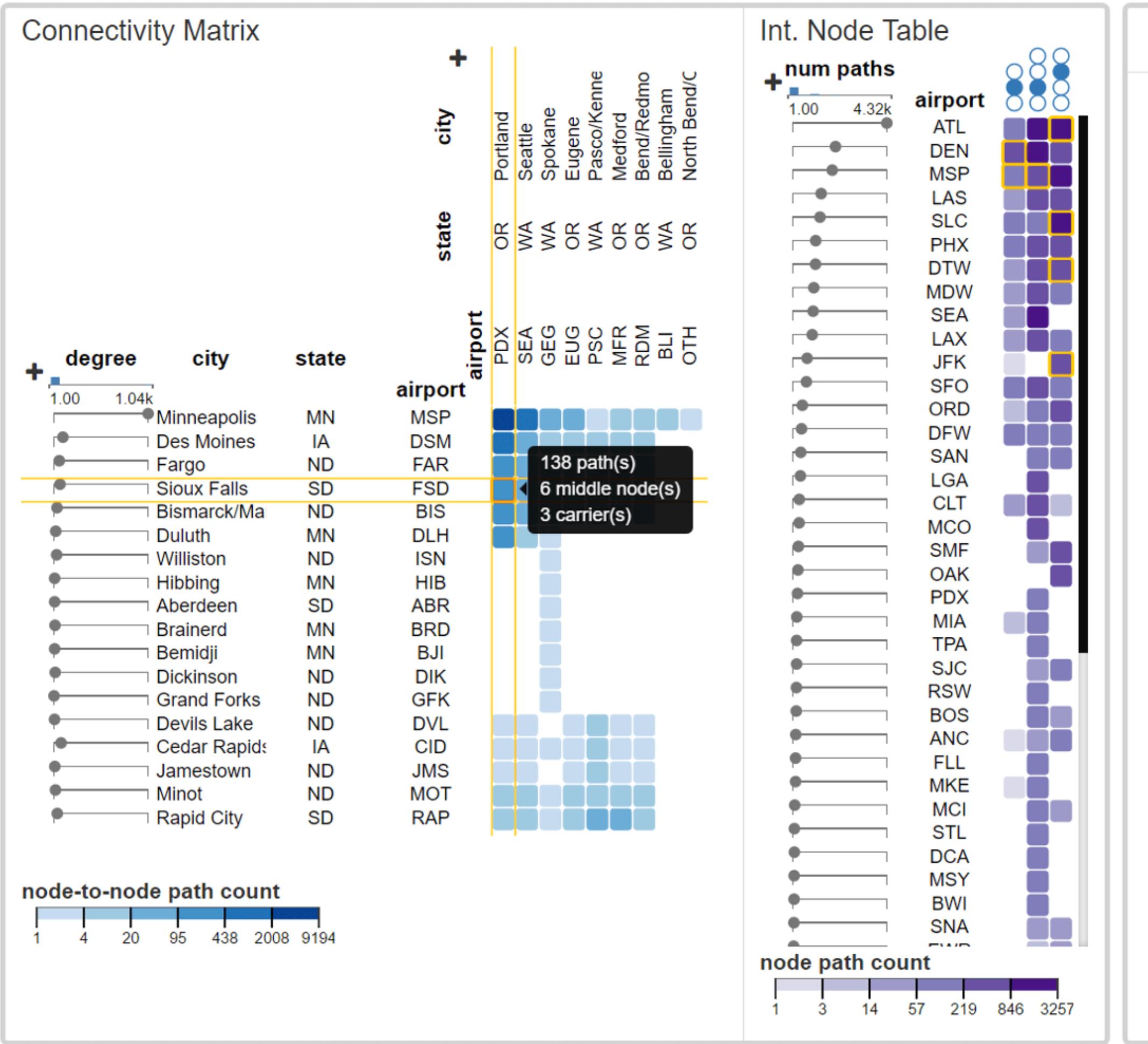
Max len	3
Start	MN (state) x IA (state) x
	ND (state) x SD (state) x
End	OR (state) x WA (state) x
Show cypher	Submit

Connectivity Matrix Controls

Aggregate rows	none
Aggregate cols	none
Sort order	optimal leaf
Metric	path count
Encoding	colormap
Scale	log

Int. Node Table Controls

Aggregate rows	none
Metric	path count
Encoding	colormap
Scale	log



Path List

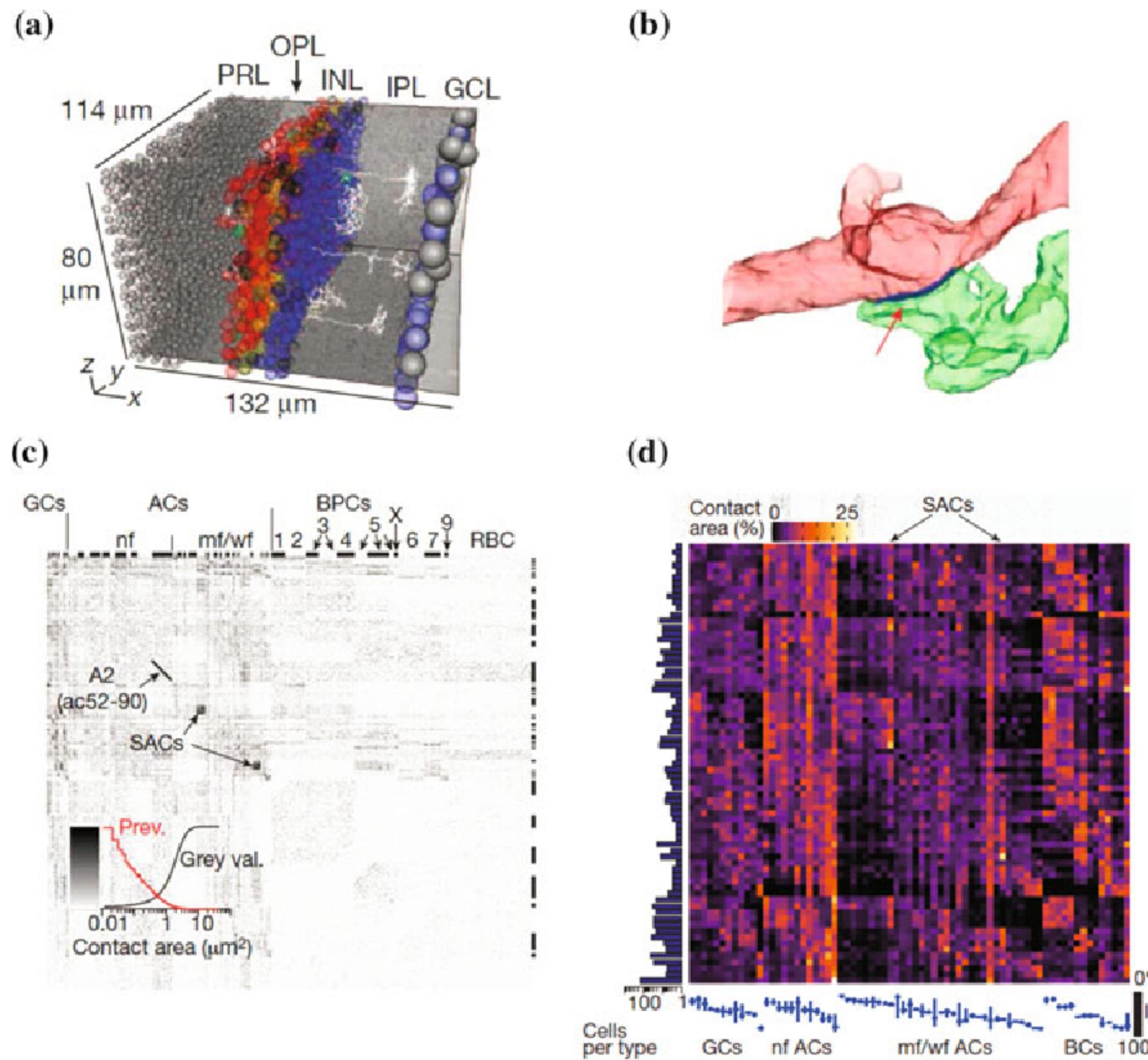
Node Link	Path List	
Paths		
> FSD -> MSP -> ATL -> PDX		
> FSD -> MSP -> SLC -> PDX		
> FSD -> MSP -> DTW -> PDX		
> FSD -> MSP -> PDX		
Edge 1	DL-23667	
	Dep: 503	
	Arr: 556	
Edge 2	DL-23240	
	Dep: 923	
	Arr: 1101	
Edge 1	DL-23667	
	Dep: 503	
	Arr: 556	
Edge 2	DL-22859	
	Dep: 923	
	Arr: 1101	
Edge 1	DL-23667	
	Dep: 503	
	Arr: 556	
Edge 2	DL-22584	
	Dep: 923	
	Arr: 1101	
Edge 1	DL-23667	
	Dep: 503	
	Arr: 556	
Edge 2	DL-22392	
	Dep: 923	
	Arr: 1101	
Edge 1	DL-23667	
	Dep: 503	
	Arr: 556	
Edge 2	DL-22194	
	Dep: 923	
	Arr: 1101	
Edge 1	DL-23667	
	Dep: 503	
	Arr: 556	
Edge 2	DL-21271	
	Dep: 923	
	Arr: 1101	
Edge 1	DL-23667	
	Dep: 503	
	Arr: 556	
Edge 2	DL-20752	
	Dep: 923	
	Arr: 1101	
Edge 1	DL-23667	
	Dep: 503	
	Arr: 556	
Edge 2	DL-20459	
	Dep: 923	
	Arr: 1101	
Edge 1	FSD -> DEN -> PDX	3 6
Edge 2	FSD -> MSP -> JFK -> PDX	4 1

Graffinity



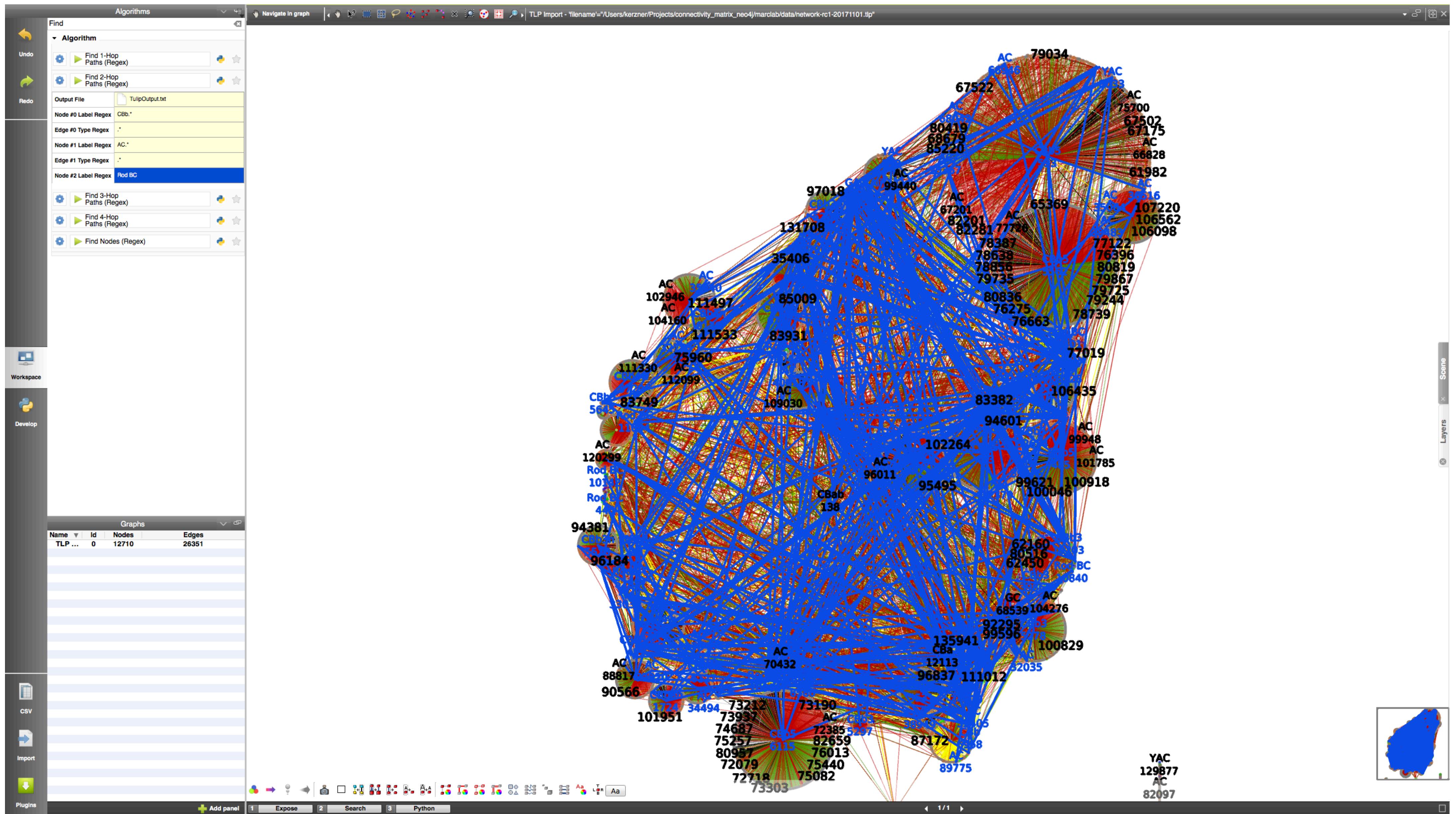
tool to summarize graph connectivity

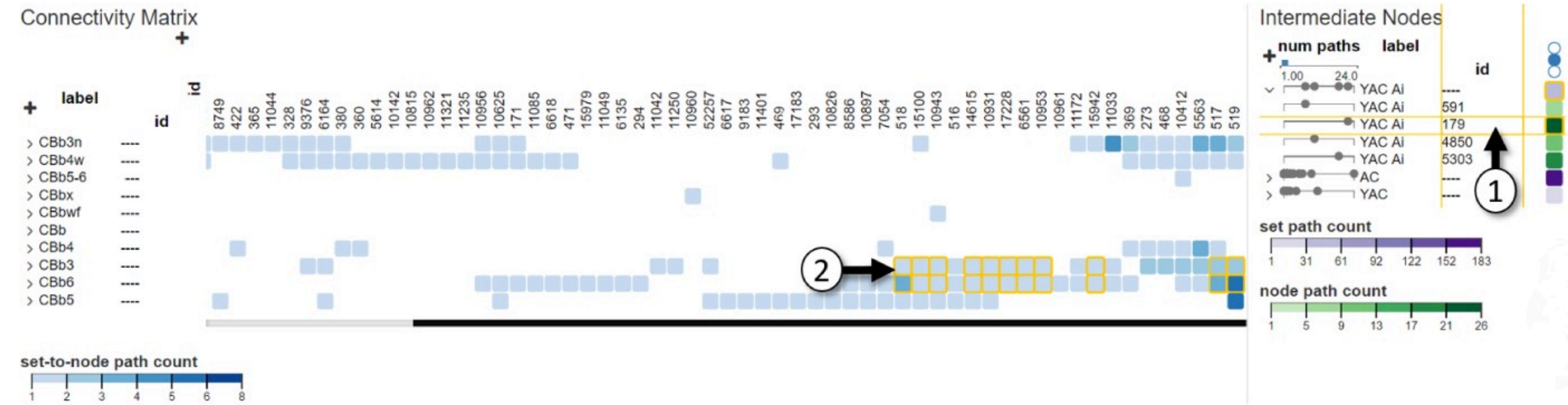
Kerzner, E., Lex, A., Sigulinsky, C.L., Urness, T., Jones, B.W., Marc, R.E. and Meyer, M., 2017, June. Graffinity: Visualizing connectivity in large graphs. In *Computer Graphics Forum* (Vol. 36, No. 3, pp. 251-260).



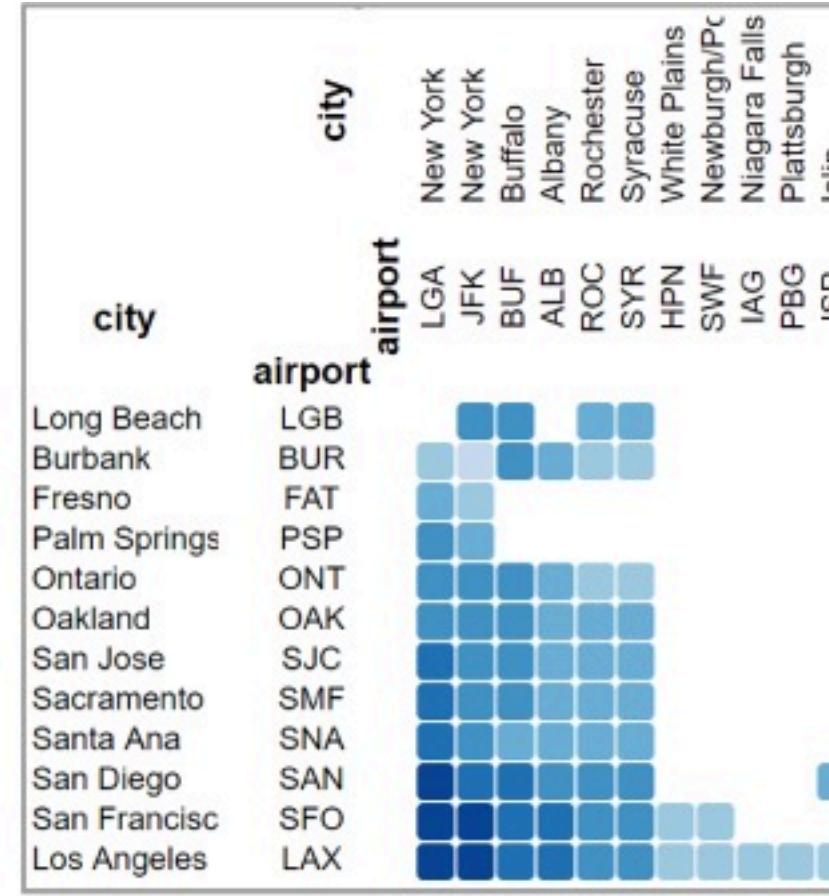
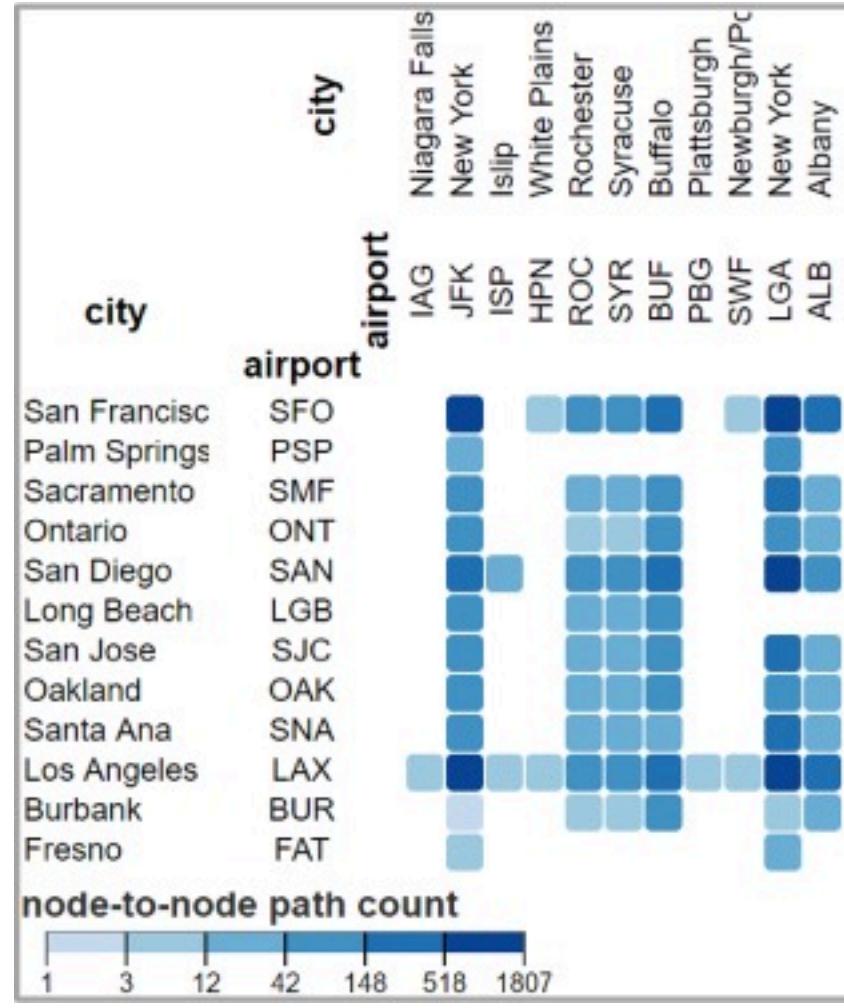
direct and indirect connections in the retina

Briggman, Kevin L.. "Chapter 2 Retinal Connectomics." (2017).





Connectivity matrix



**generalizable to use
with other datasets
such as airport flight
connections**

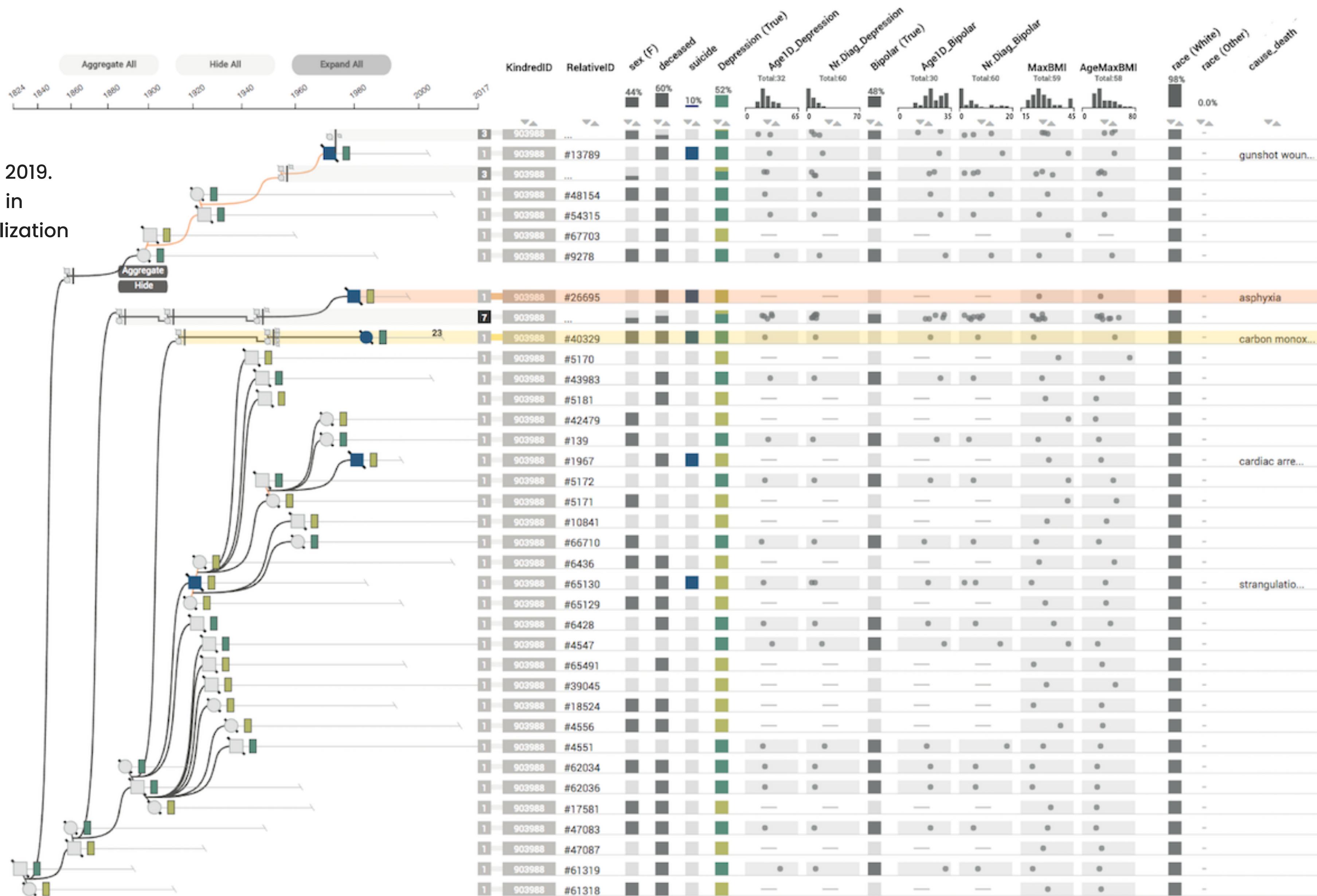
https://www.youtube.com/watch?time_continue=57&v=tTabH-8eh_c

Lineage

Nobre, C., Gehlenborg, N., Coon, H. and Lex, A., 2019.
Lineage: Visualizing multivariate clinical data in
genealogy graphs. IEEE transactions on visualization
and computer graphics, 25(3), pp.1543–1558.



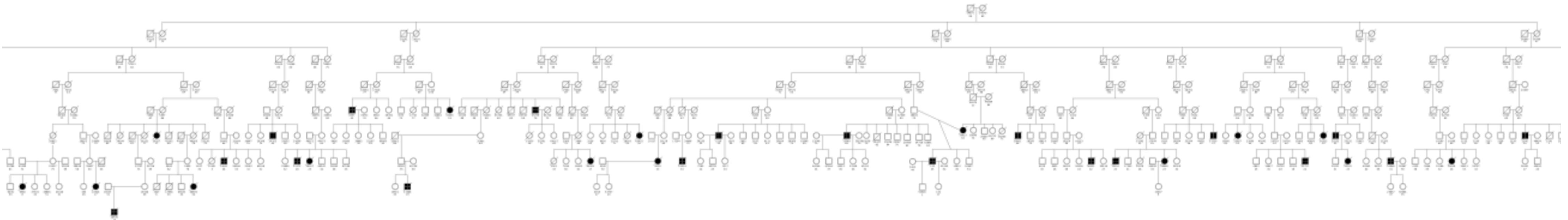
**disease in the
context of
genealogies.**



***Investigating genetic and
environmental factors that contribute
to complex psychiatric deceases.***

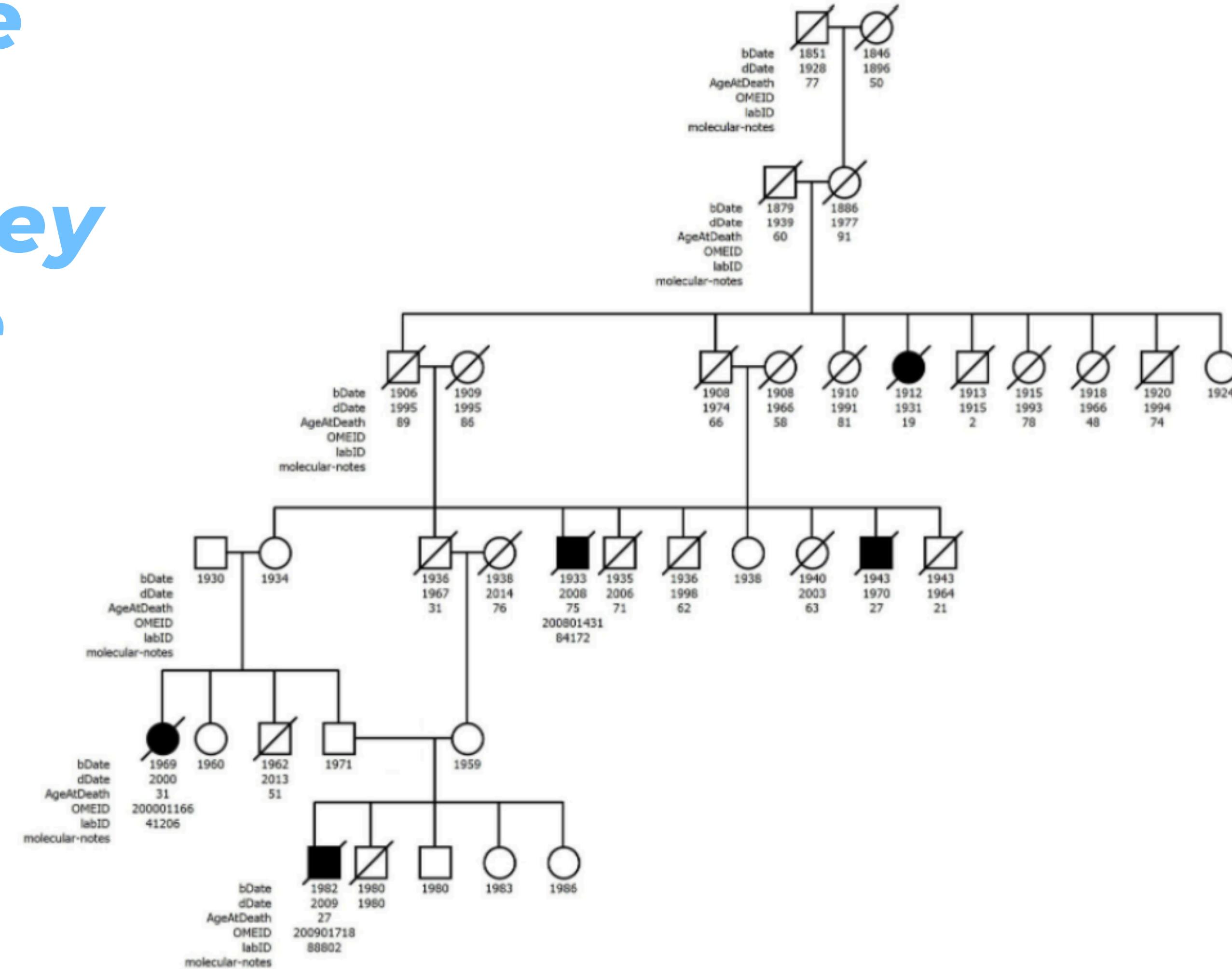
Dataset: sample of suicide cases

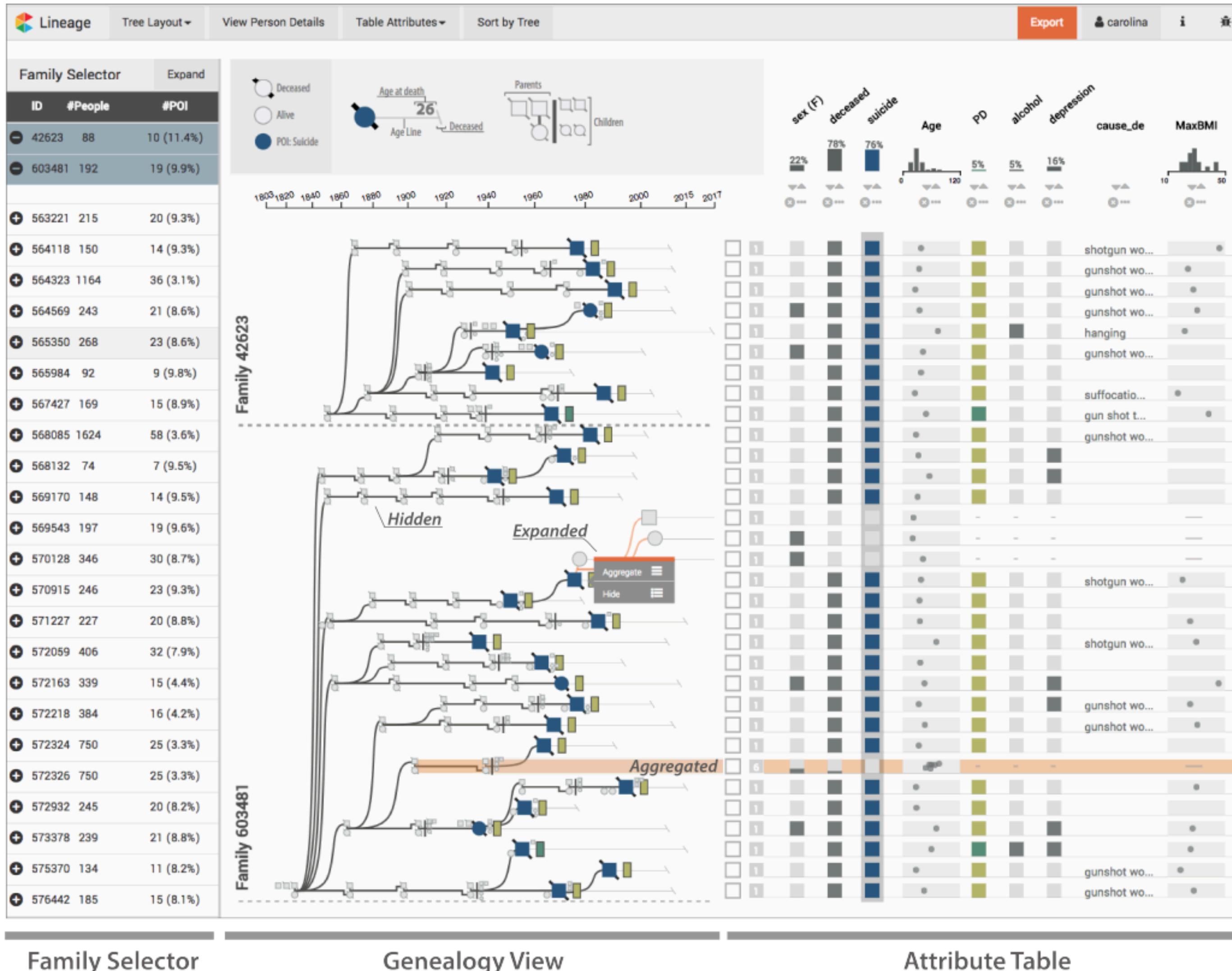
***Collaborators use
genealogical graphs
in their workflow,***



400 families in Progeny

**limited in the
number of
attributes they
can visualize**

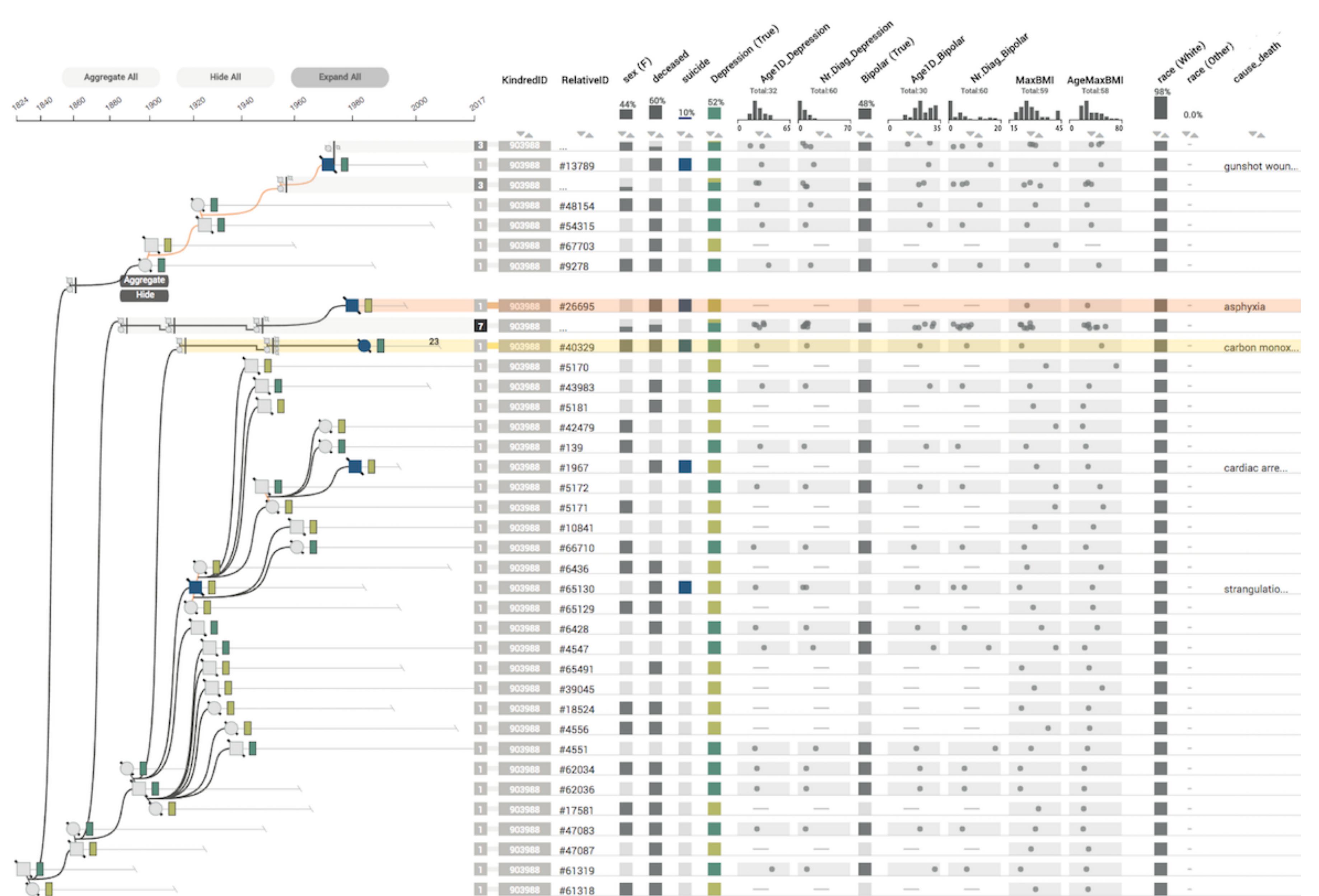




**scalable in
respect to the
number of
attributes for
every person in
the tree**



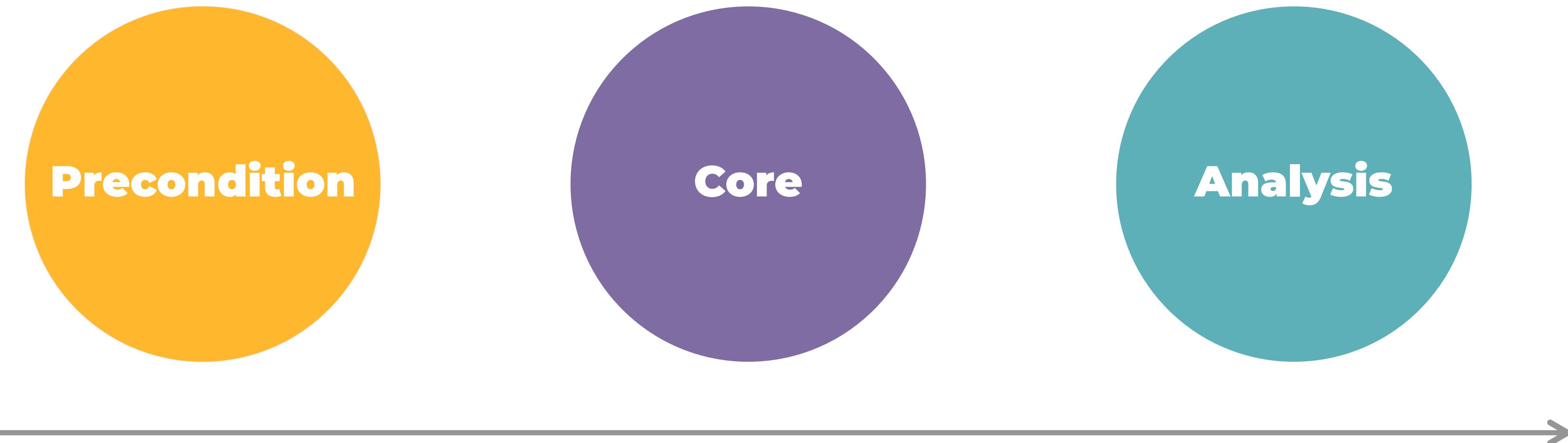
***Worked closely with
researchers,
conducted creativity
workshop***



<https://youtu.be/dG8z6T97hgE>

I'll be here for little over 2 months..

Revisiting the design study phases

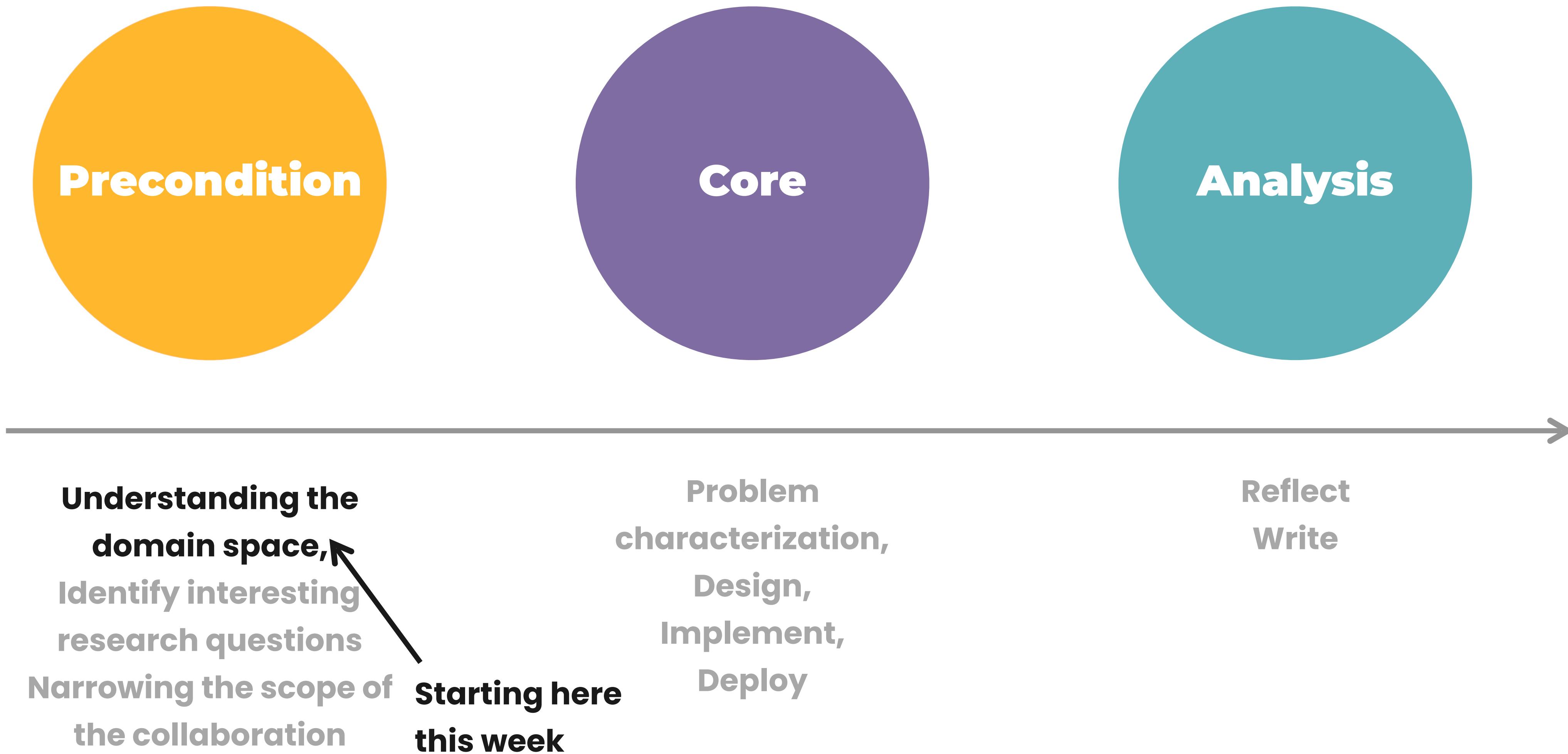


Understanding the domain space,
Identify interesting research questions
Narrowing the scope of the collaboration

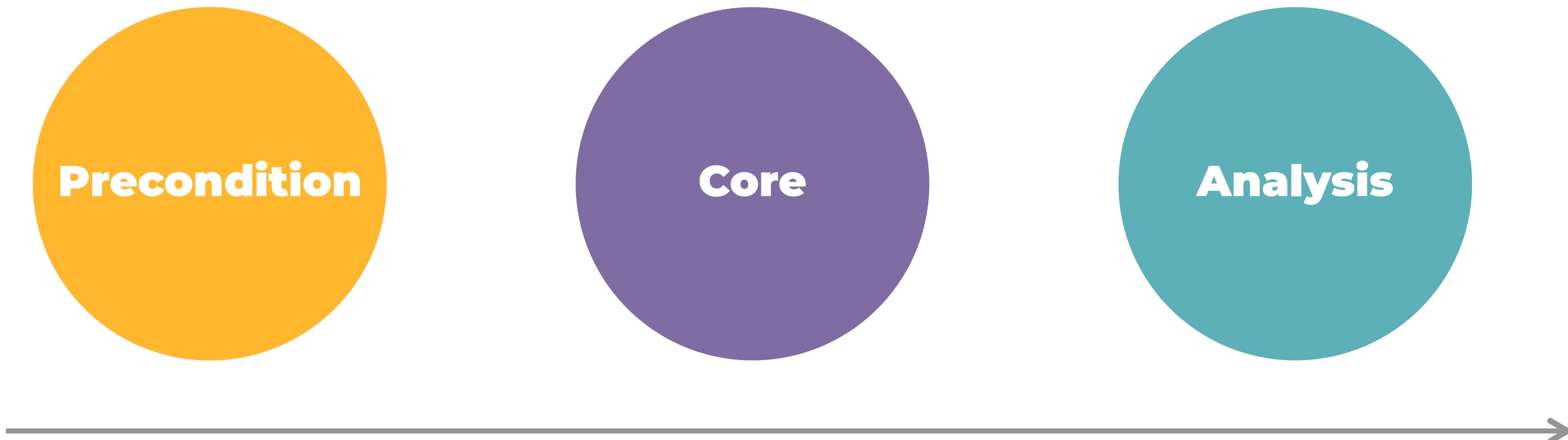
Problem characterization,
Design,
Implement,
Deploy

Reflect
Write

Revisiting the design study phases



Revisiting the design study phases



Understanding the domain space,
Identify interesting research questions
Narrowing the scope of the collaboration

Problem characterization,
Design, Implement, ←
Deploy

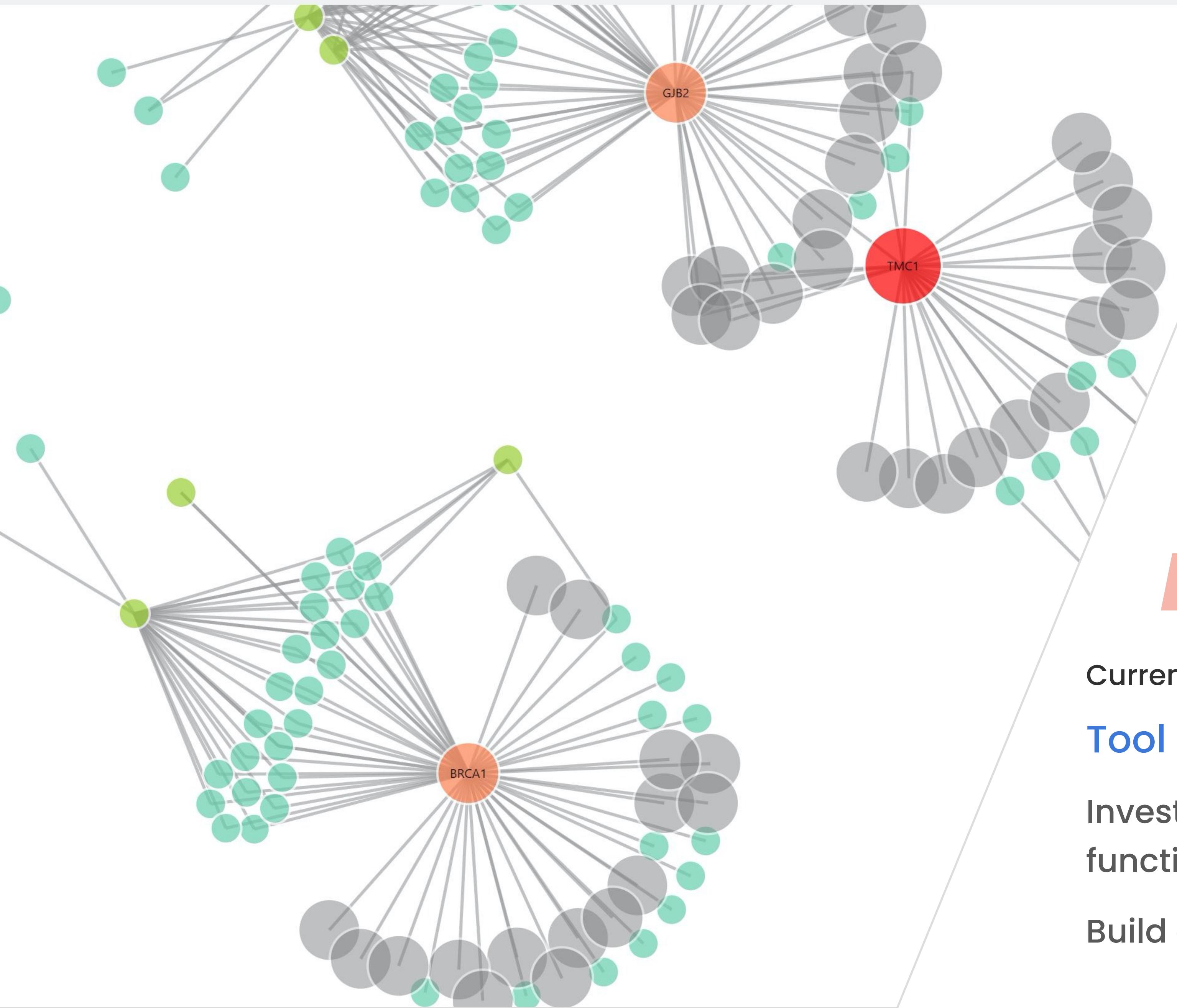
Reflect
Write
We will probably get somewhere about here

***Want to meet with all of you to get
an understanding of what you do***

Thank you!

Whole Network ▾

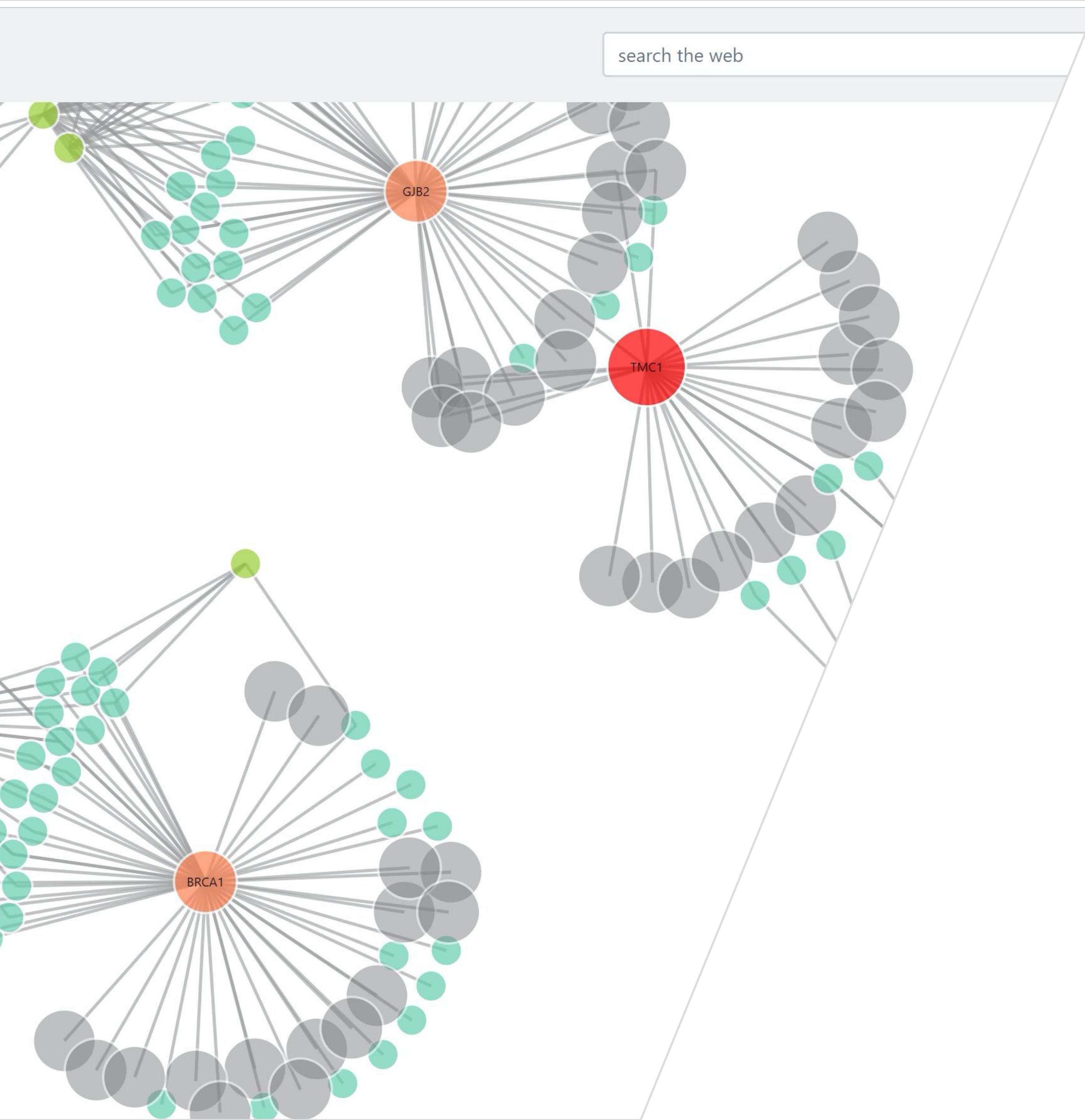
search the web



ARUP Pathways Project

Currently collaborating with genetic analysts from ARUP.
Tool for genetic analysts,
Investigate mutated genes of lesser known functions
Build evidence for associations between entities.

search the web



Understanding Current workflow..

Initial interview with analyst focused on exam sequencing.

Studied her work flow

Her goal for each analysis, find variants that could help explain a patients phenotype.

Understanding Current Workflow..

To do this:

Begins with a sample of identified variant calls

References variety of resources and databases

Builds mental model of evidence and interprets findings for report.

Understanding Workflow limitations..

Not all genes and variants are well documented.

Extract clues from a combination of resources to understand importance

(Gene, known variants, associated pathways, gene expression and products, conservation, protein interaction partners of products,

Understanding User Needs..

Assistive tool to build a network of evidence, making connections between entities explicit

Understanding User requirements..

Look for **connections between entities**

Extract evidence from relevant literature and external databases
support associations between entities in the graph

