

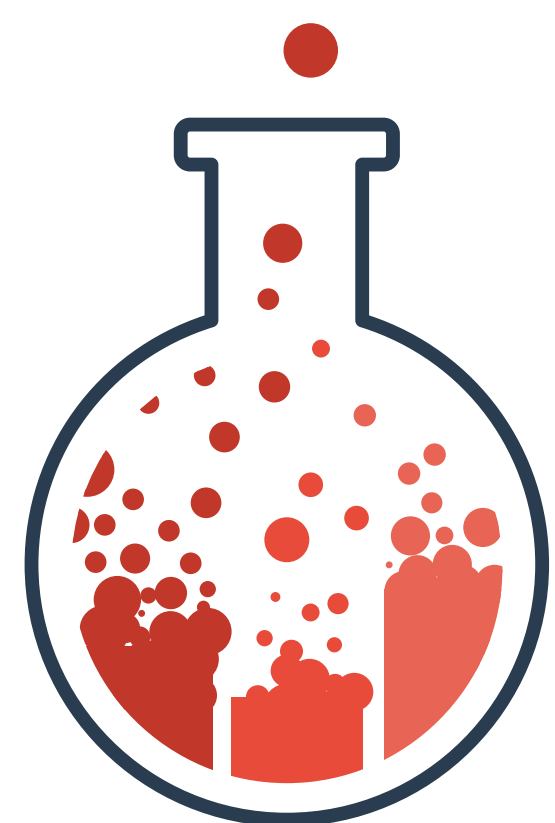
Alexander Lex

@alexander_lex

<http://alexander-lex.net>



Lineage: Visualizing Multivariate Clinical Data in Genealogy Graphs



visualization
design lab





visualization
design lab



Miriah Meyer



Alexander Lex



www.sci.utah.edu

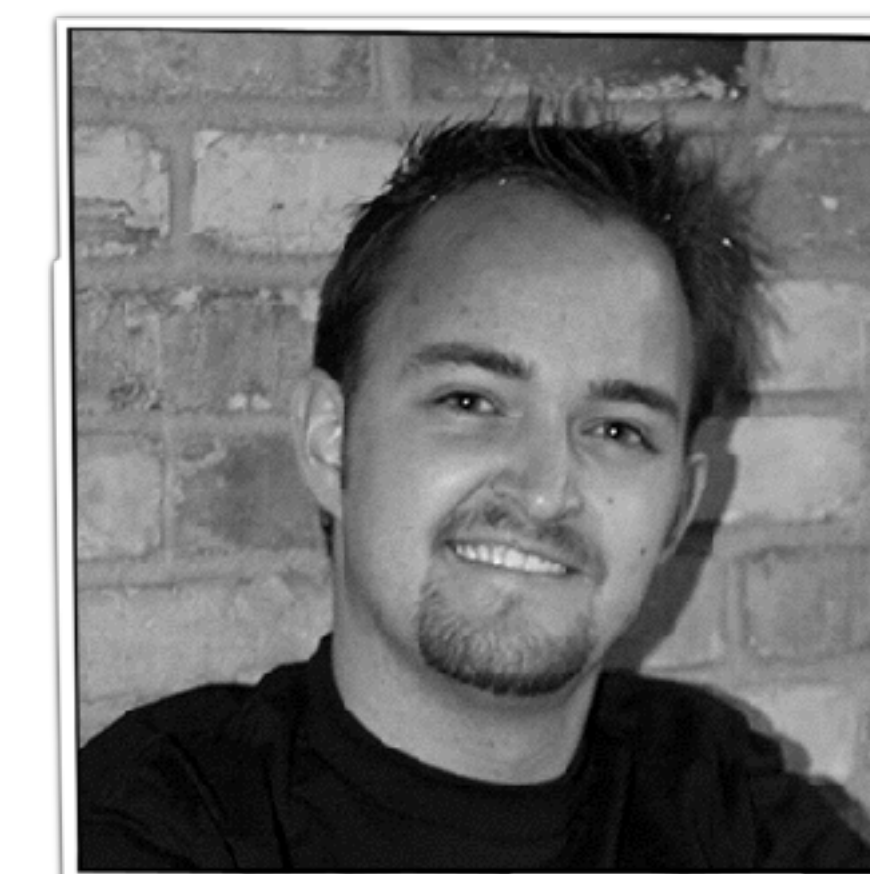
<http://vdl.sci.utah.edu/>



Pascal Goffin



Ethan Kerzner



Alex Bigelow



Nina McCurdy



Jennifer Rogers



Sam Quinan



Jimmy Moore



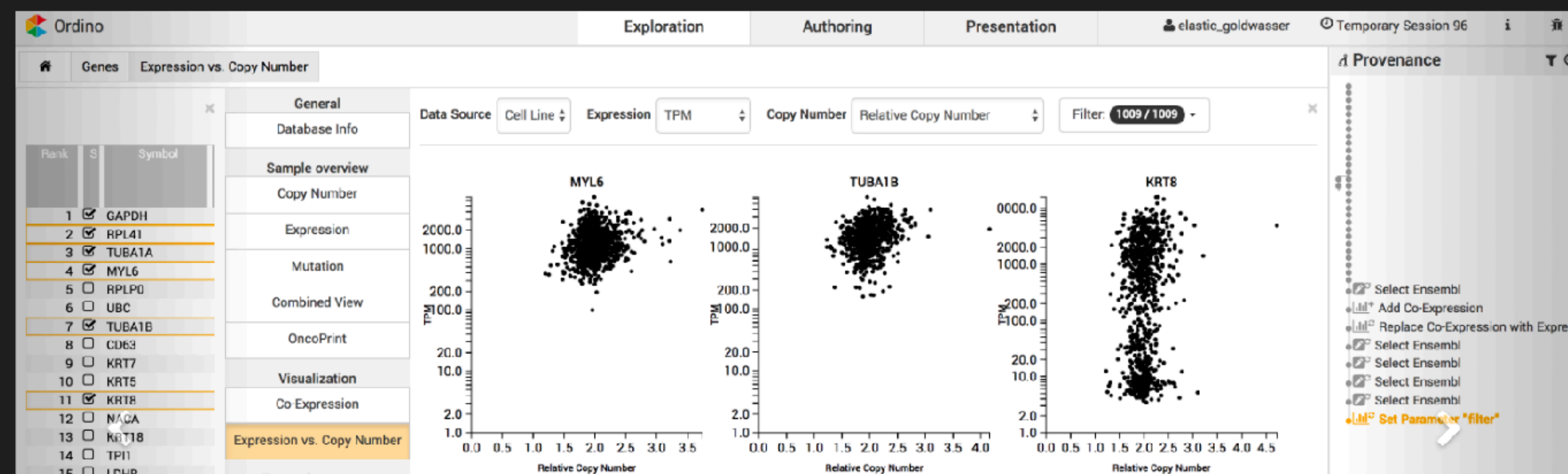
Carolina Nobre

We develop data visualization solutions for applications in pharmaceutical and biomedical R&D.

PRODUCTS

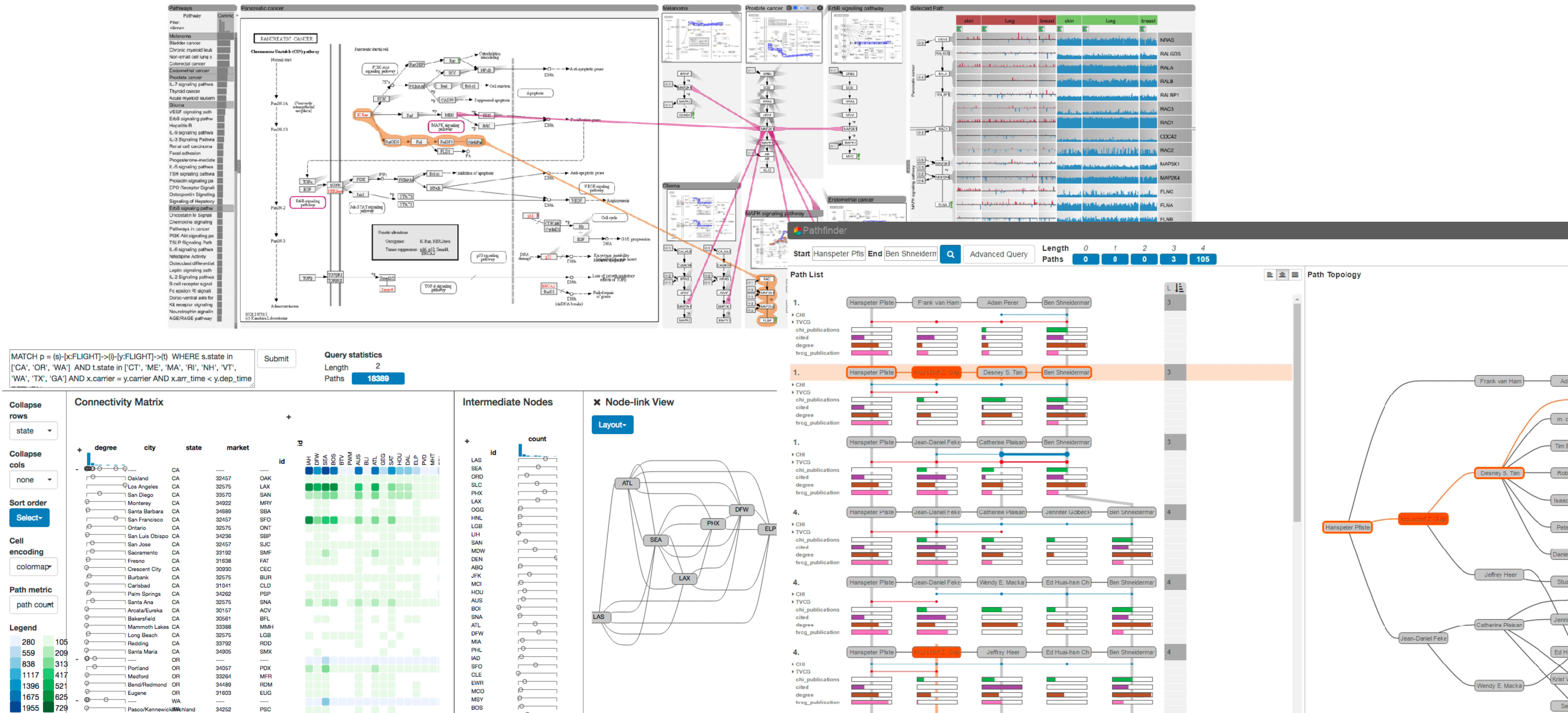
TARGET DISCOVERY PLATFORM

Our Target Discovery Platform is a web-based visual data analysis solution designed to score, rank, filter and visualize datasets that provides all the data and visualizations needed to identify analysis targets.

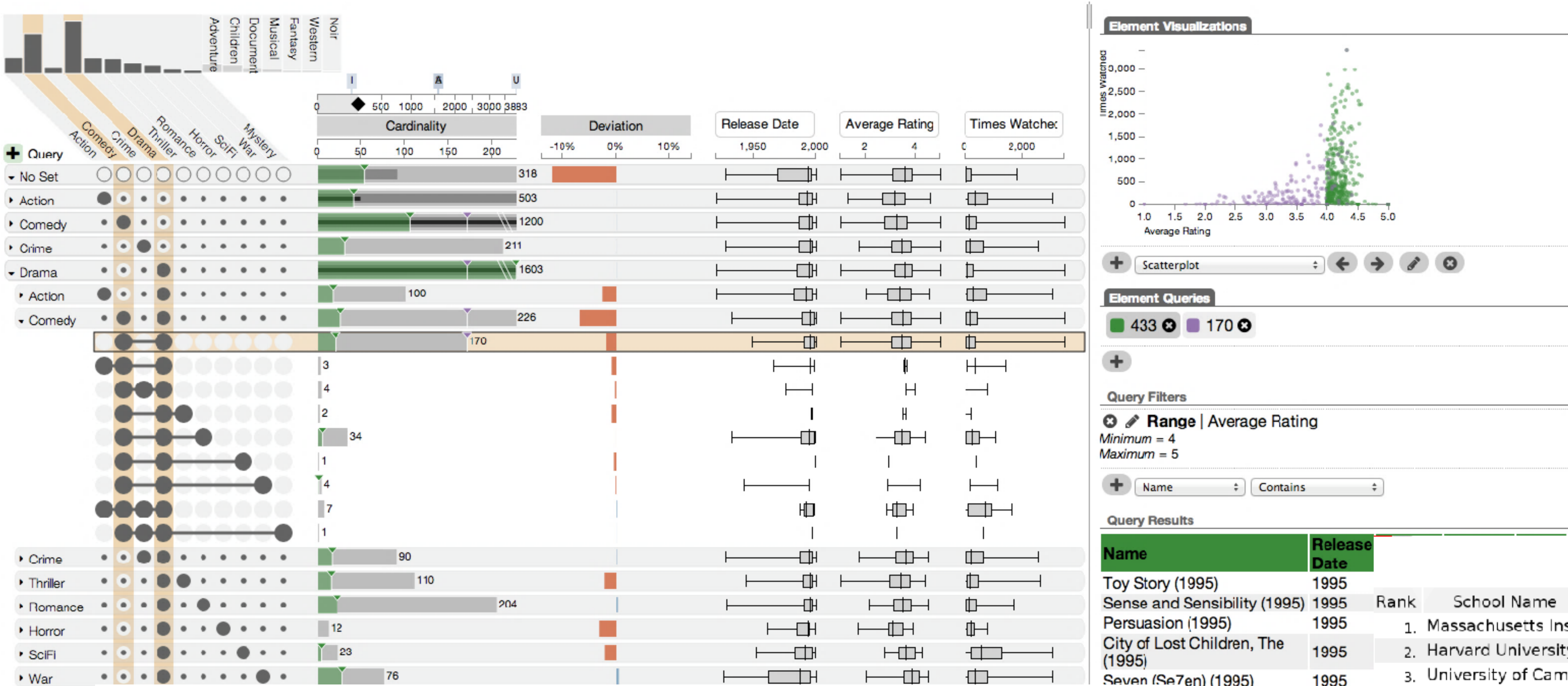


Research Areas

Large, Multivariate (Biological) Networks

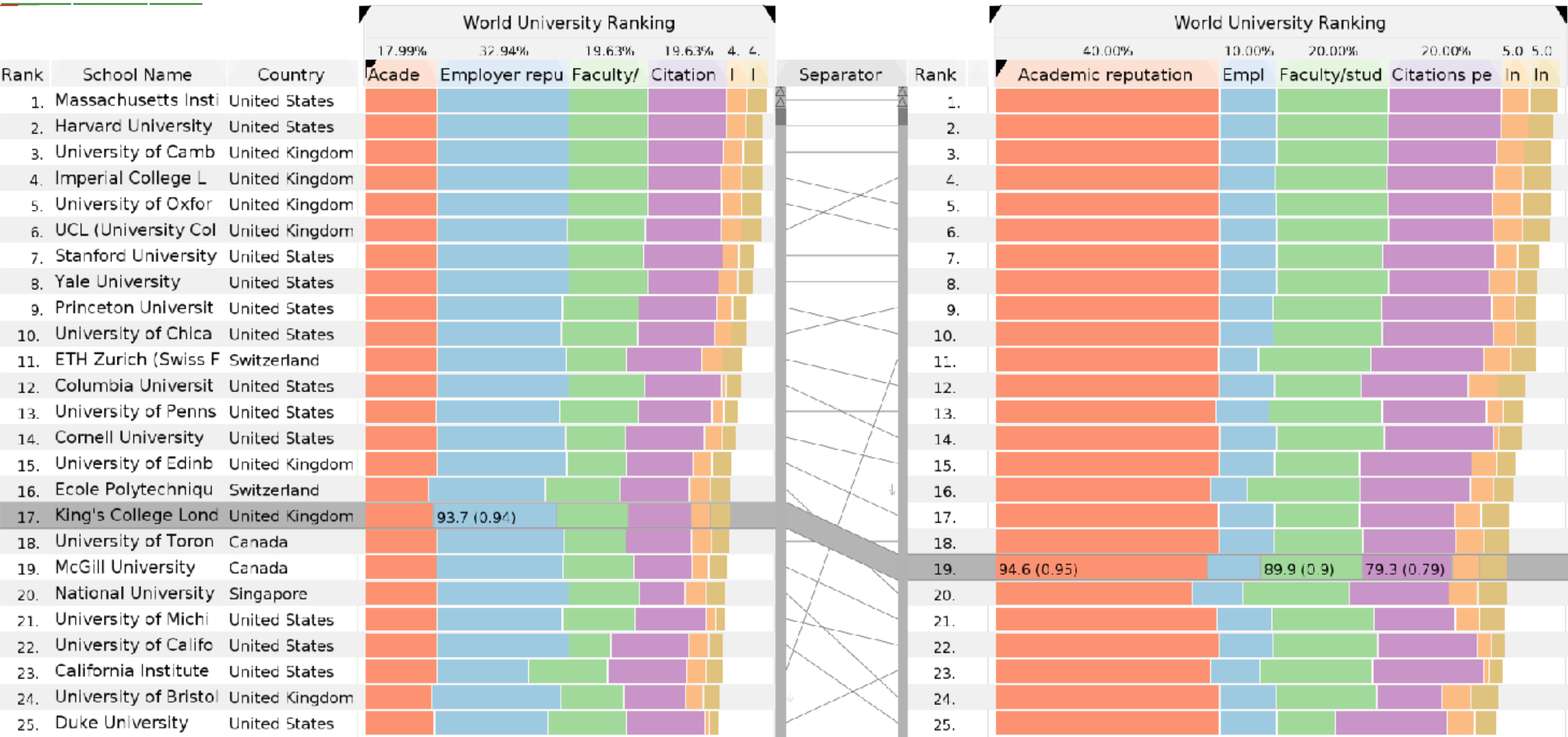
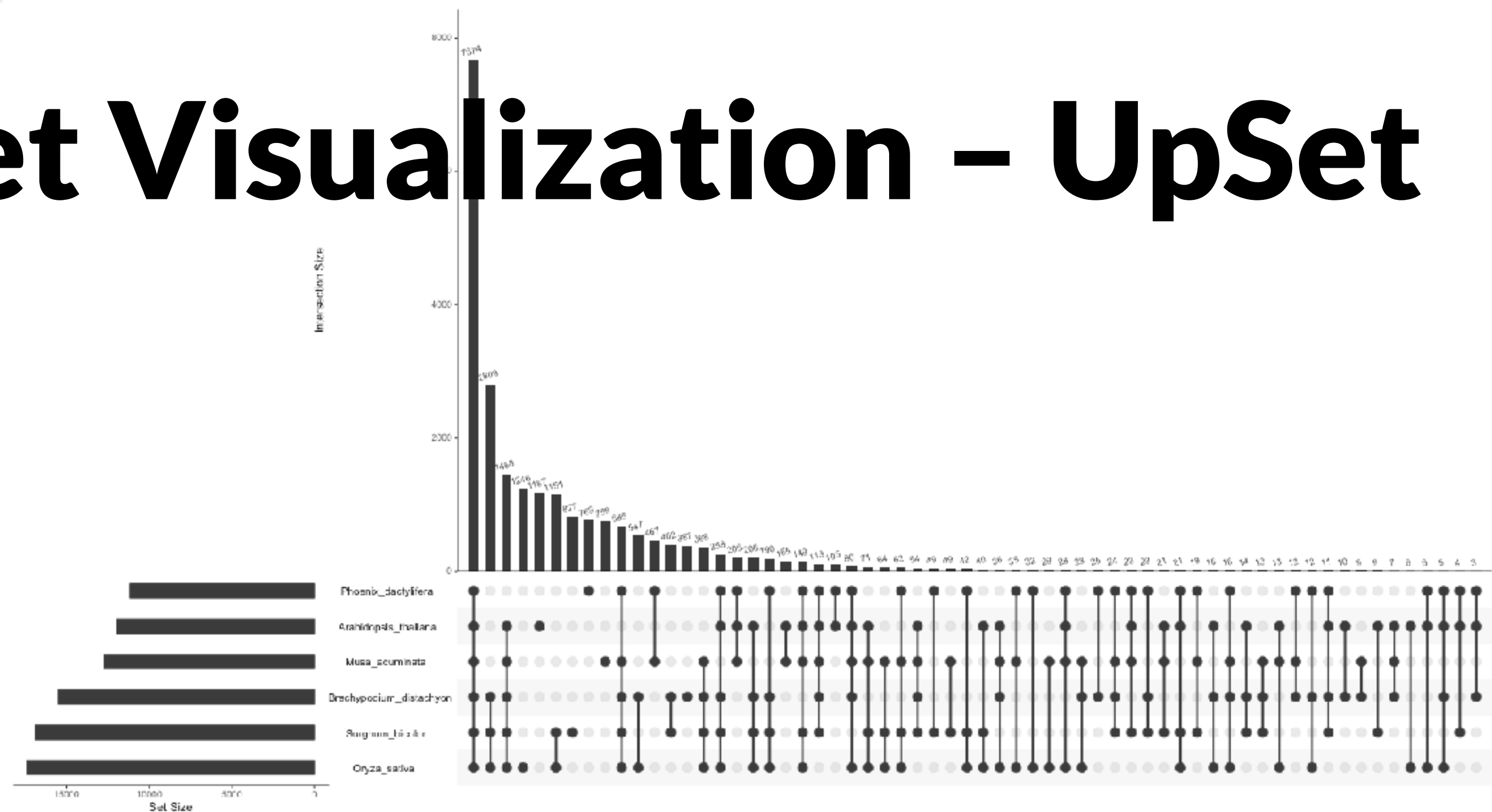


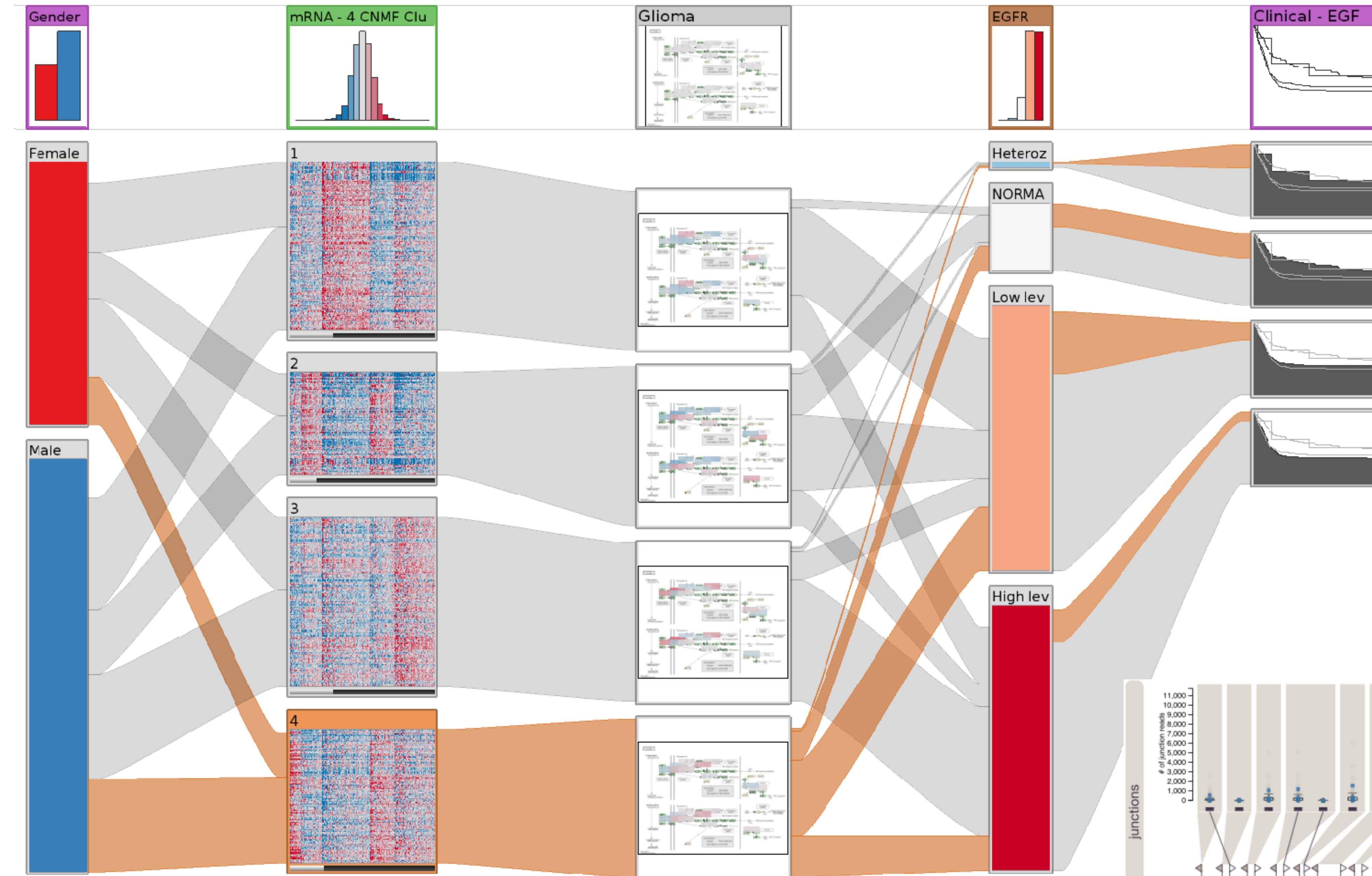
Multidimensional Data



Multivariate Rankings – Lineup

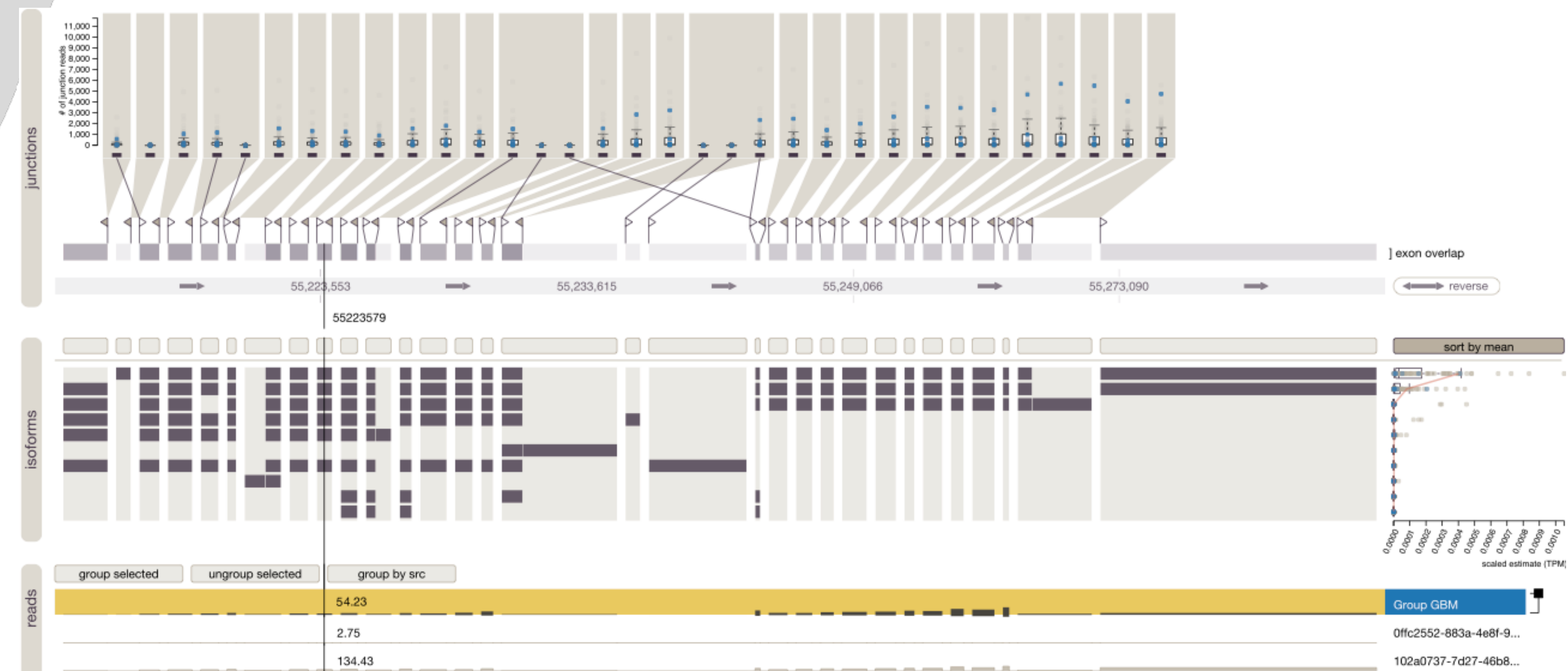
Set Visualization – UpSet



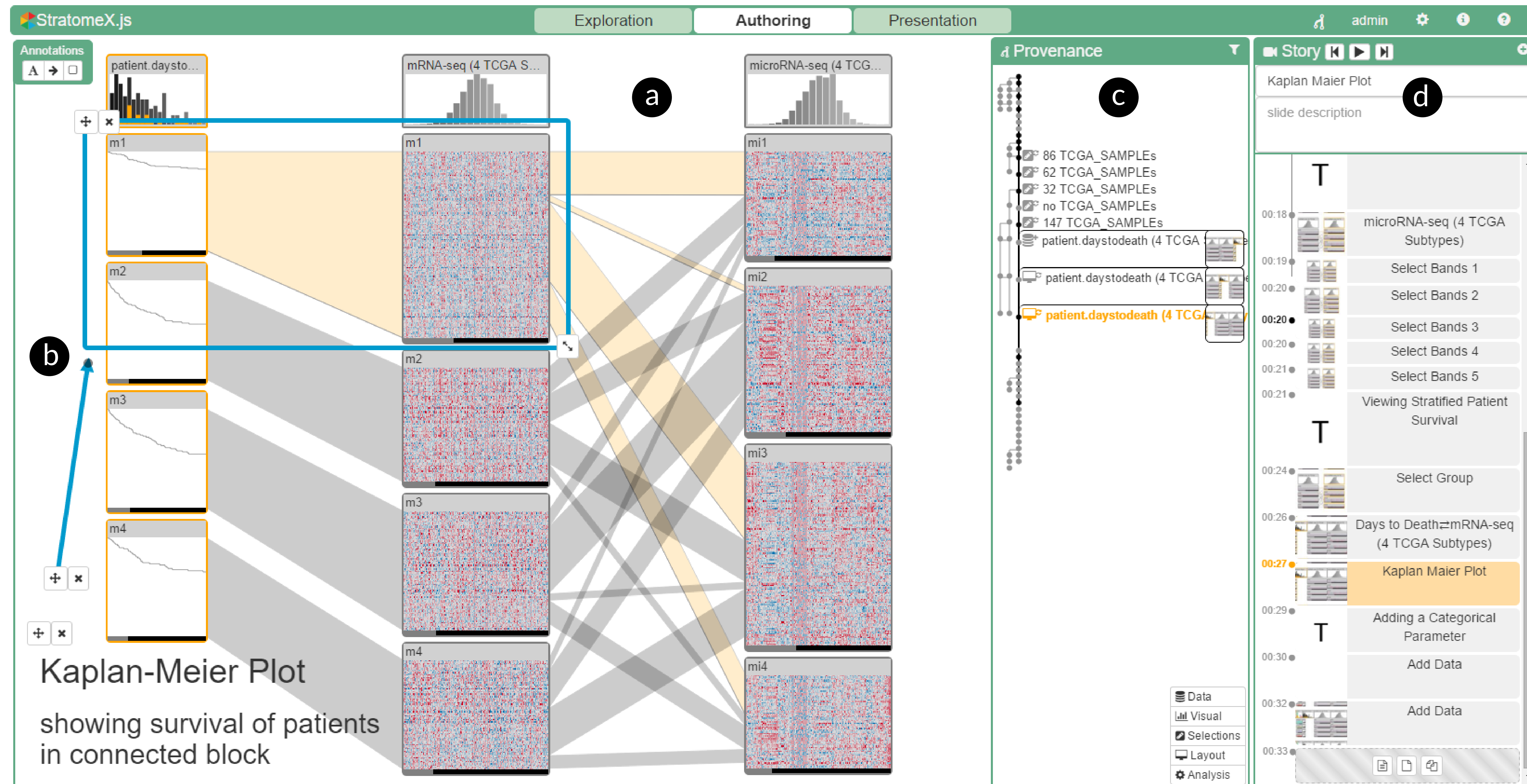


Alternative Splicing / mRNA-seq

Cancer Subtypes / Omics Clustering and Stratification



Reproducibility, Storytelling, Annotation, and Integration in Computational Workflows



Genealogies & Clinical Data

FamilyID # People #POI

3812112

14911310

2725140439

426238110

6893924423

17686042644

60348118119

79153311410

903988585

Data Selection

KindredID

RelativeID

sex

M

F

deceased

Y

N

suicide

N

Y

POI

Depression

F

T

A

Age1D_Depression

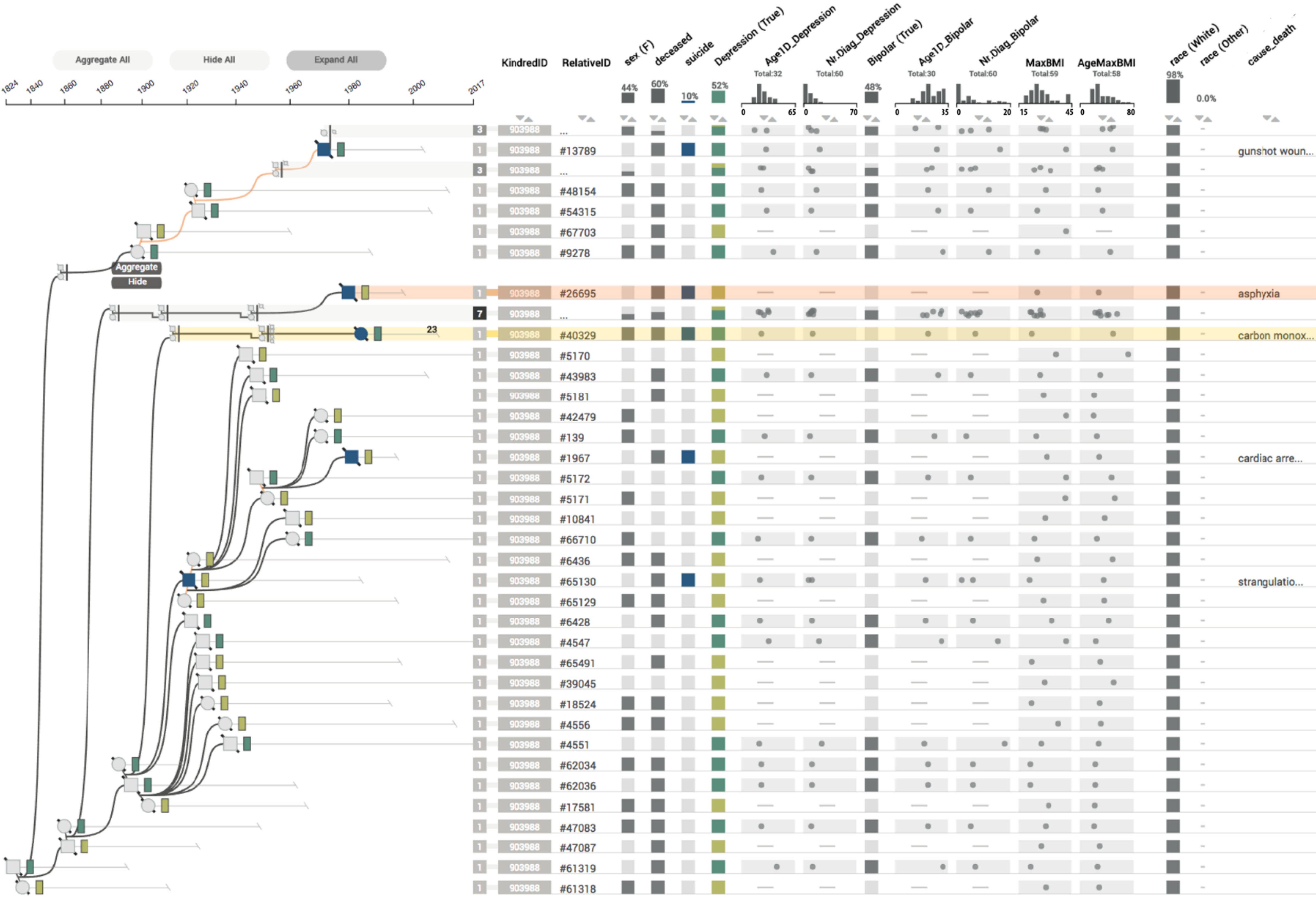
#

0

64

Nr.Diag_Depression

#



visualization

The purpose of computing is insight,
not numbers.

pictures

- Card, Mackinlay, Shneiderman
- Richard Wesley Hamming

Banana ***M. acuminata***

Date ***P. dactylifera***

Cress ***Arabidopsis thaliana***

Rice ***Oryza sativa***

Sorghum ***Sorghum bicolor***

Brome ***Brachypodium distachyon***

Good Data Visualization

- ... makes data **accessible**
- ... combines strengths of **humans and computers**
- ... enables **insight**
- ... **communicates**

Can We Trust Statistics?

I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.69					8	6.89

Mean x: 9 y: 7.50
Variance x: 11 y: 4.122
Correlation x – y: 0.816
Linear regression: $y = 3.00 + 0.500x$

Anscombe's Quartett

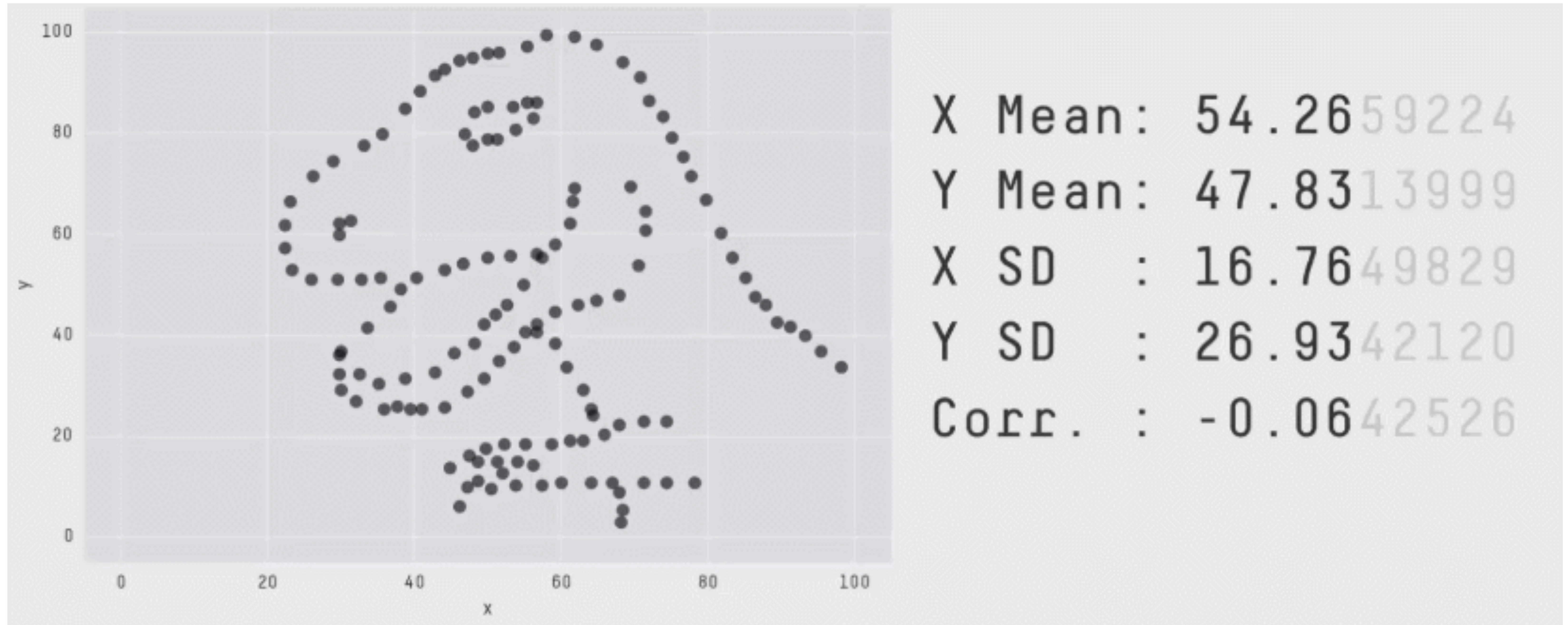


Mean x: 9 y: 7.50

Variance x: 11 y: 4.122

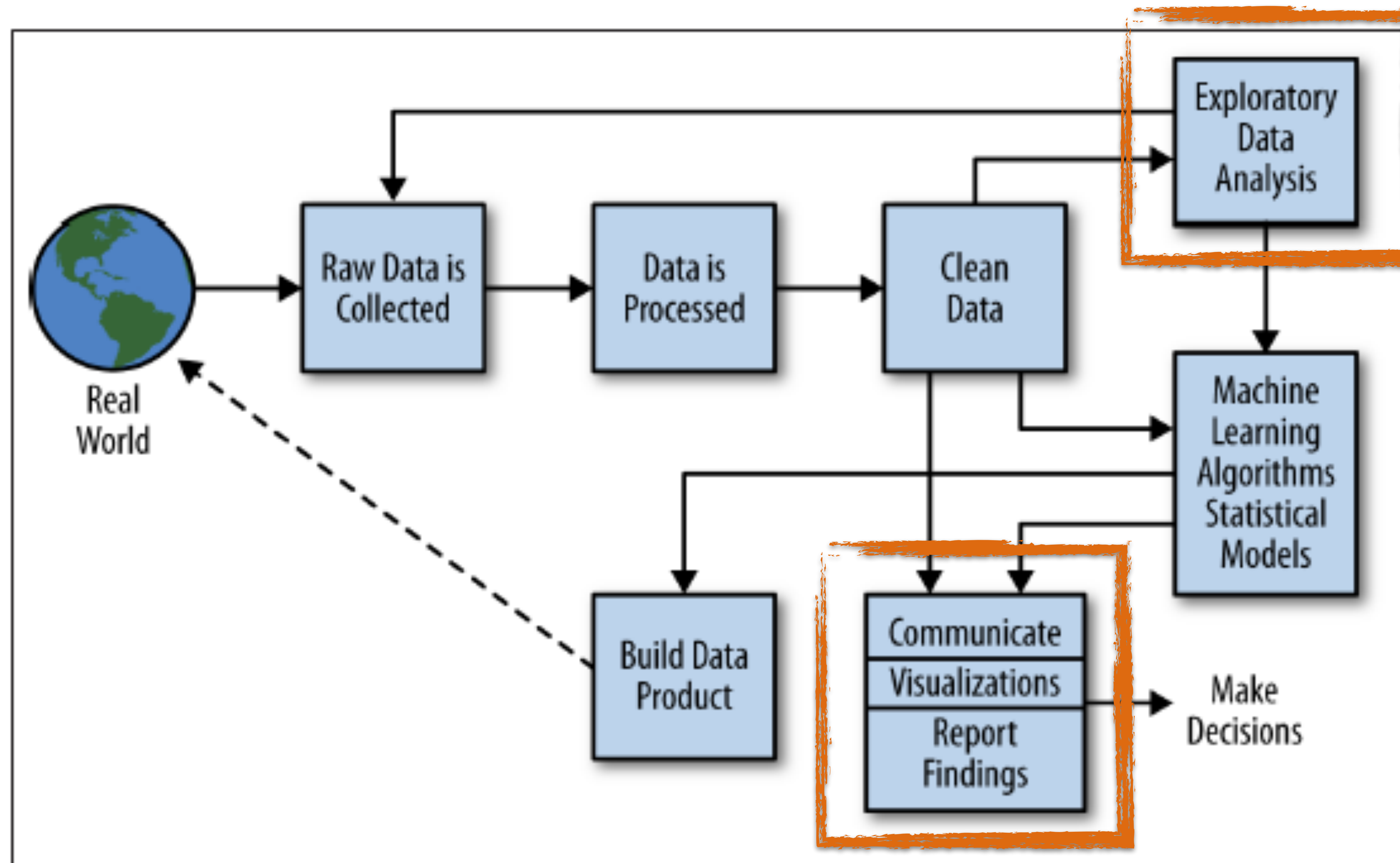
Correlation x – y: 0.816

Linear regression: $y = 3.00 + 0.500x$



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

Visualization in the Data Science Process



Visualization =

Human Data Interaction

Visualizing Clinical Data in Genealogy Graphs



Carolina Nobre



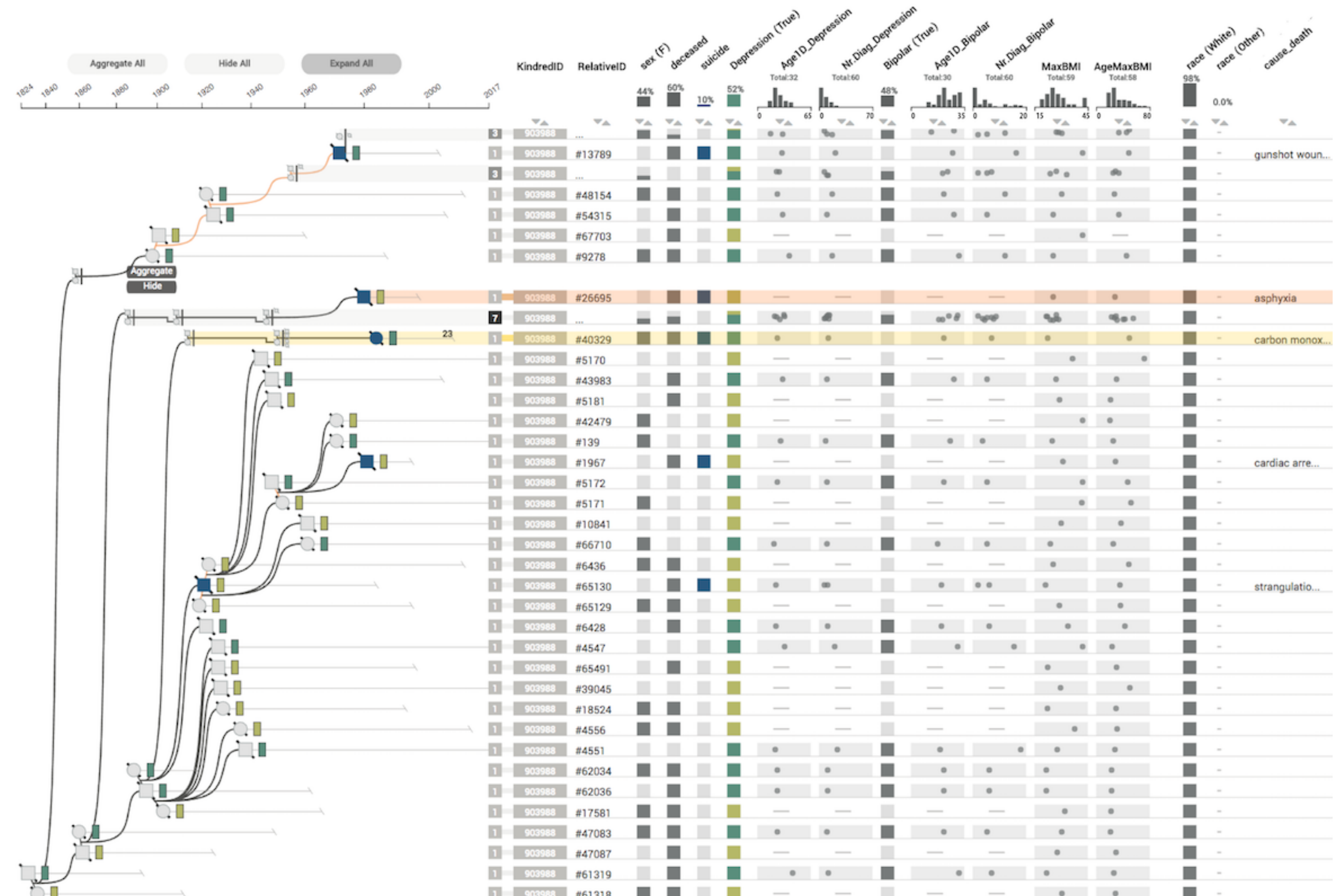
Nils Gehlenborg



Hilary Coon



Alexander Lex

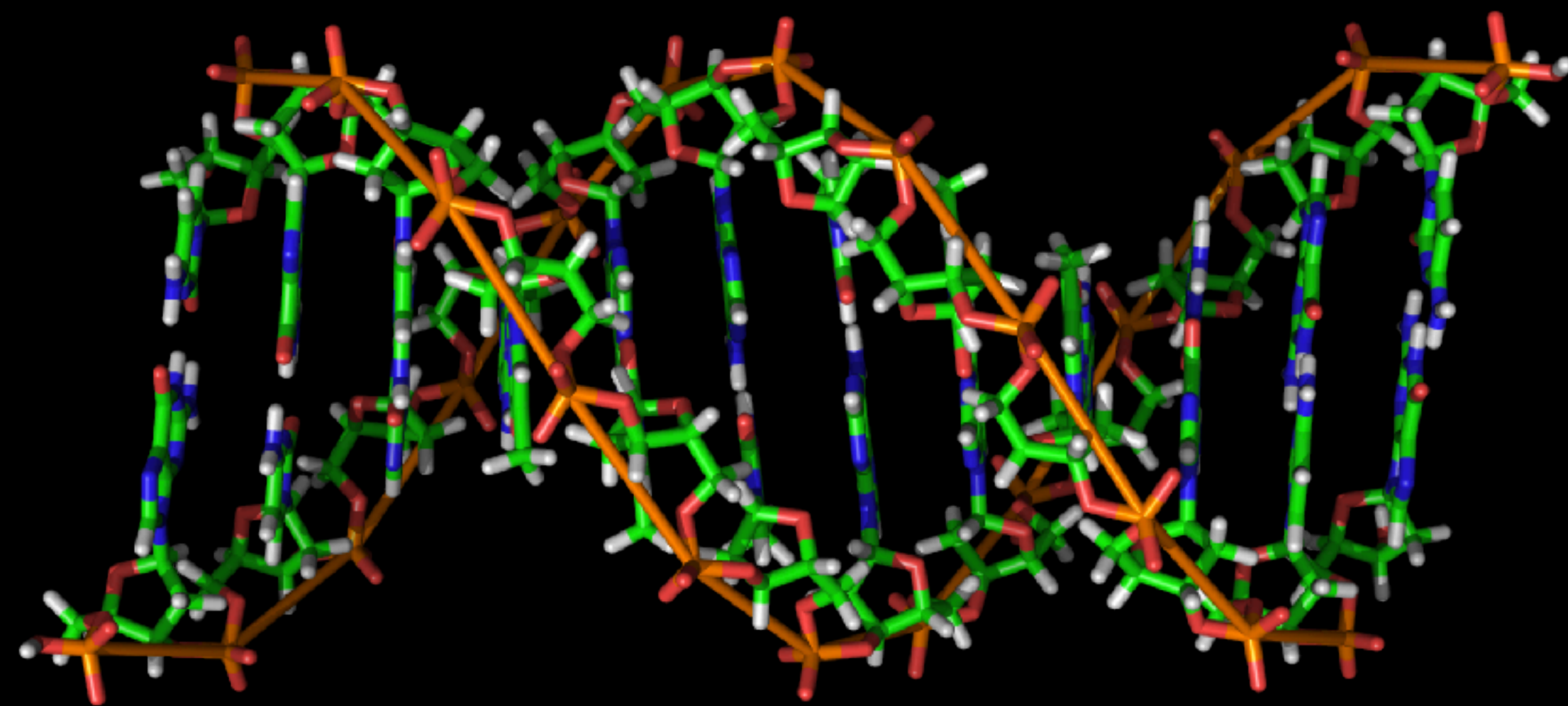


Motivation

**Understand Complex Psychiatric
Conditions**

Discover Genetic Risk Factors

**Dataset: 118k people, 19k suicide cases,
550 families based on Utah Population
Database**

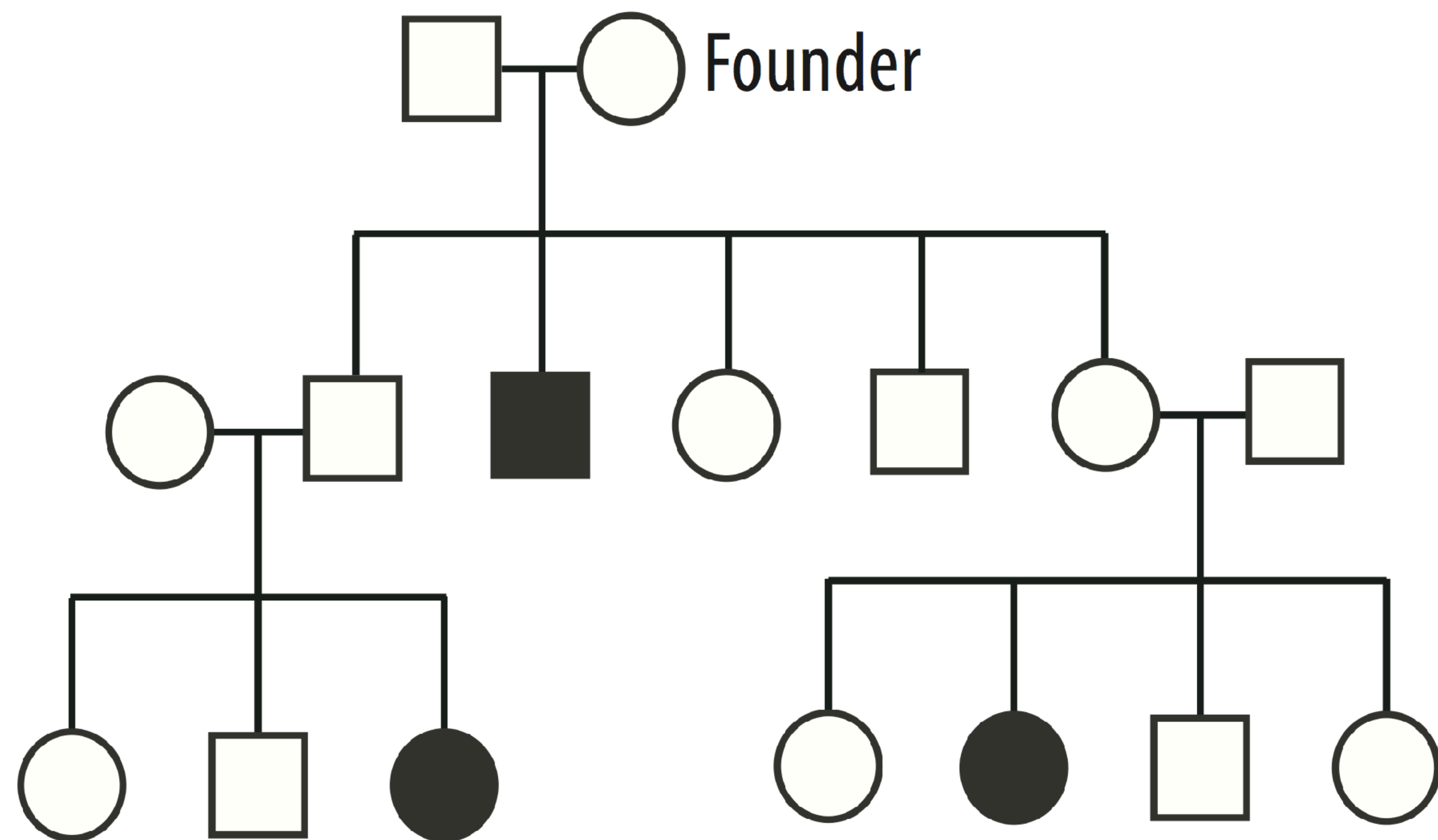


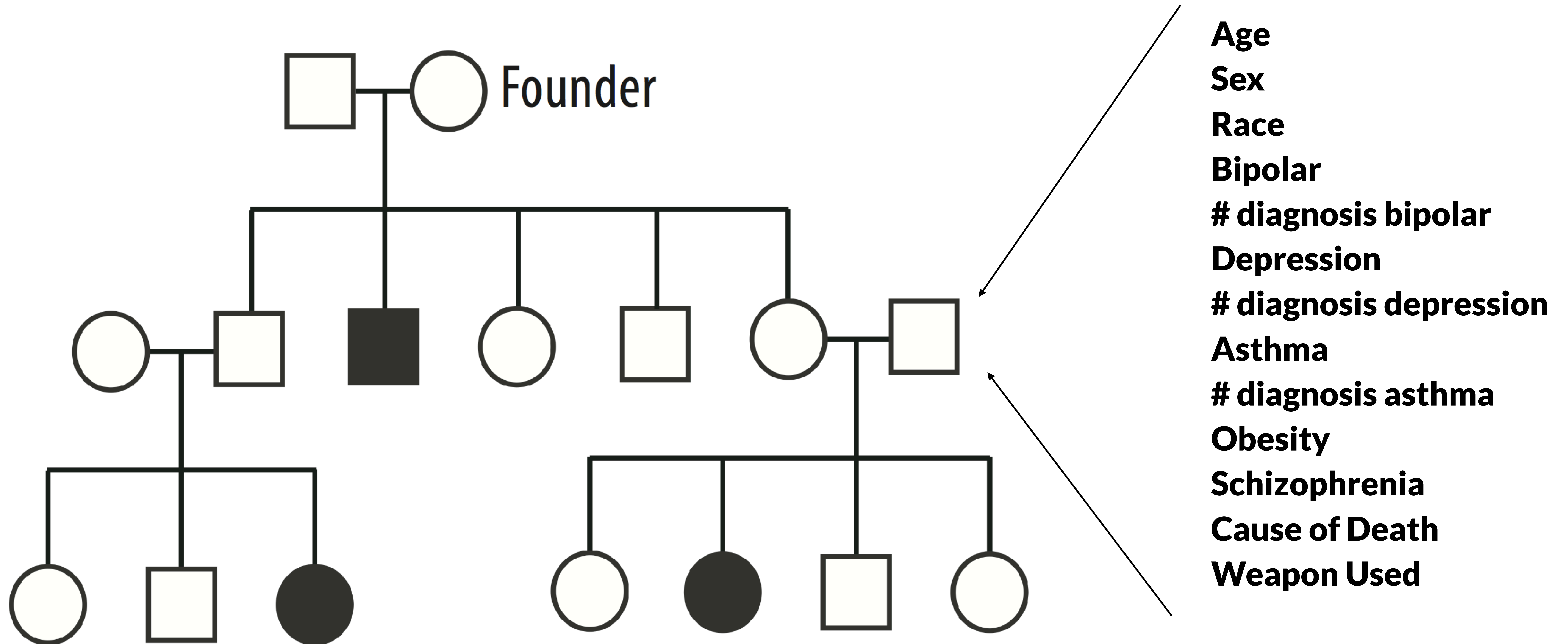
Specific Goals

Find familial cases that also have an “interesting” phenotype
e.g., predominantly female, associated with rare psychiatric disease, etc.

Prioritize those cases for analysis of Shared Genomic Sequences

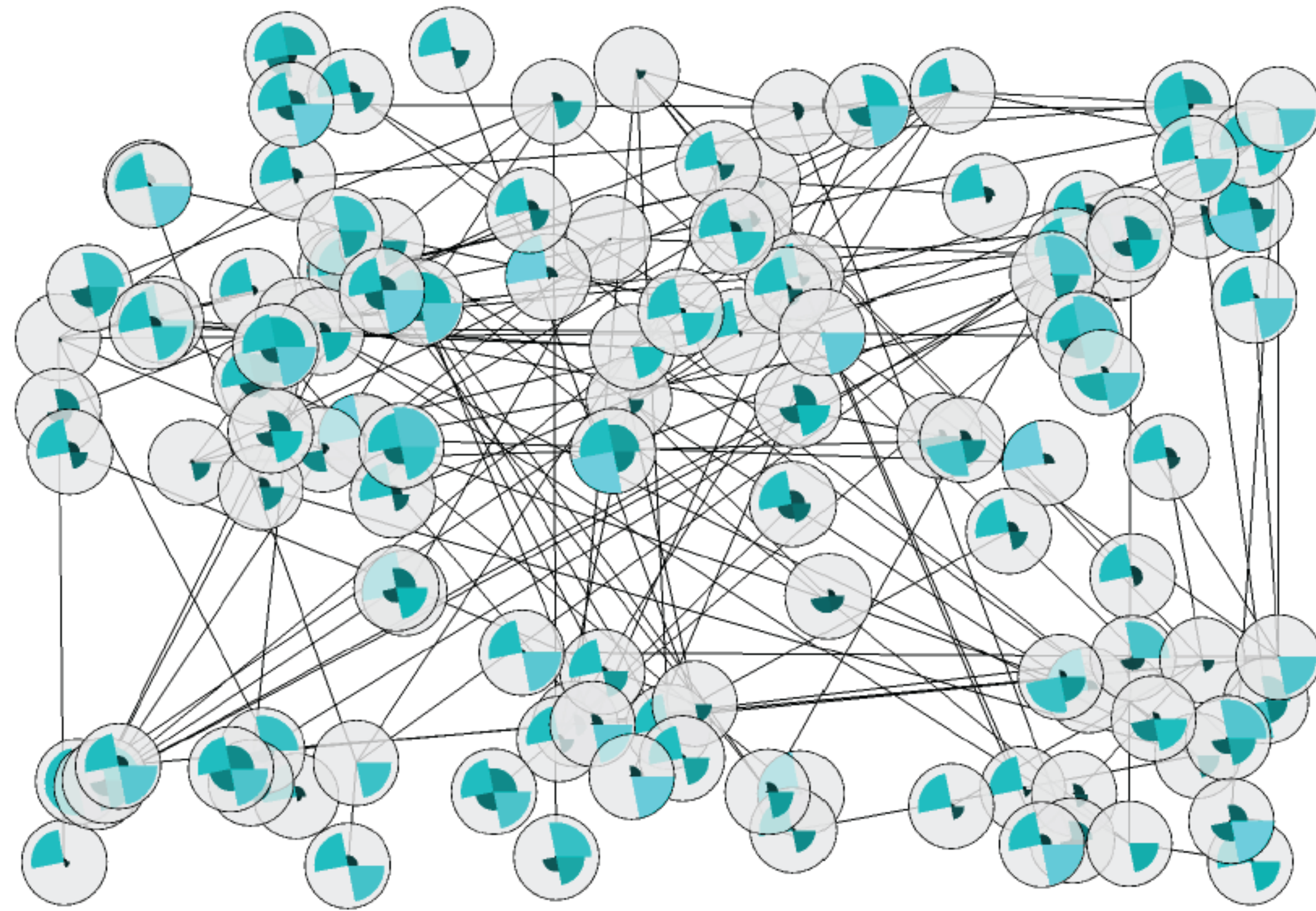
Proofreading the Data!



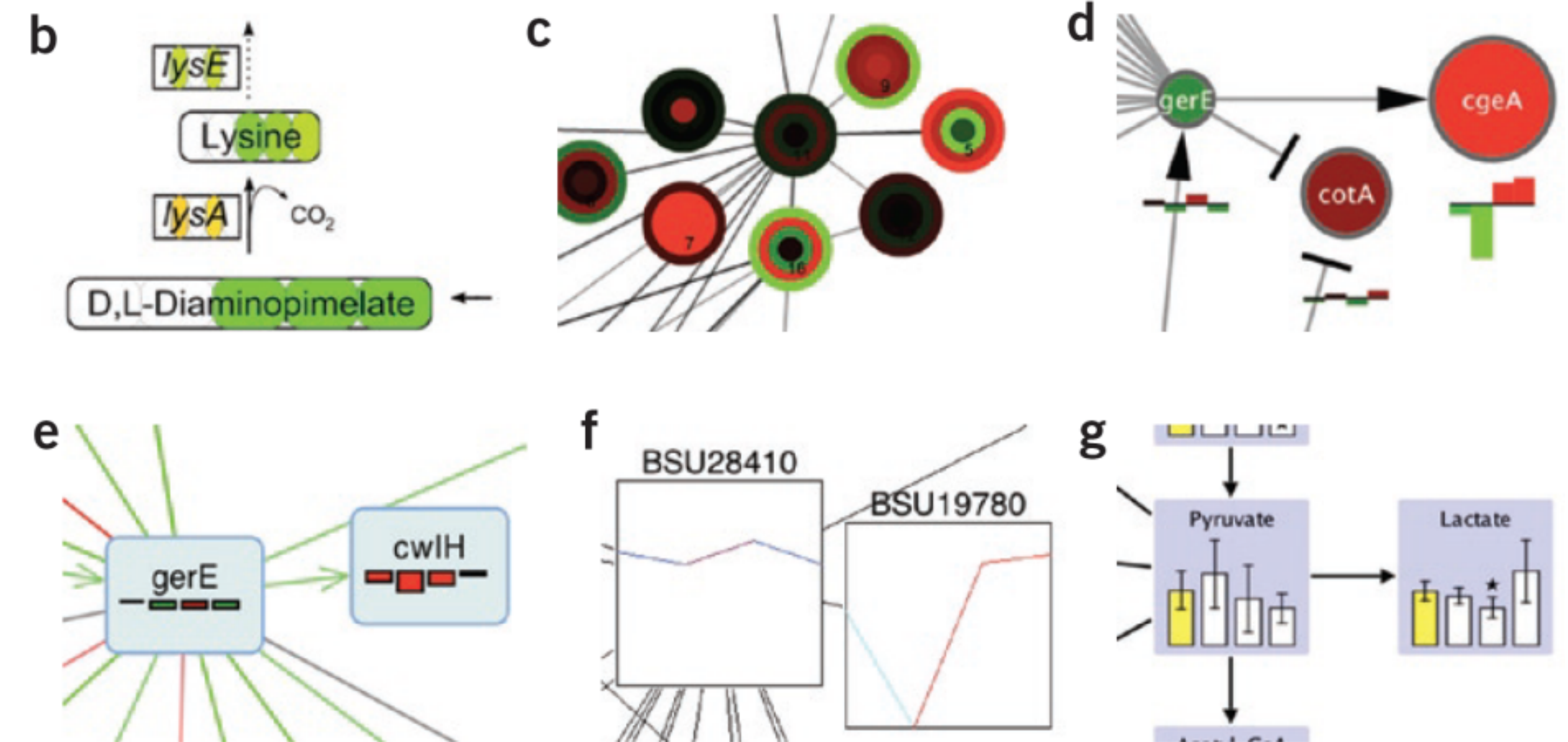


Multivariate Attributes and Graphs

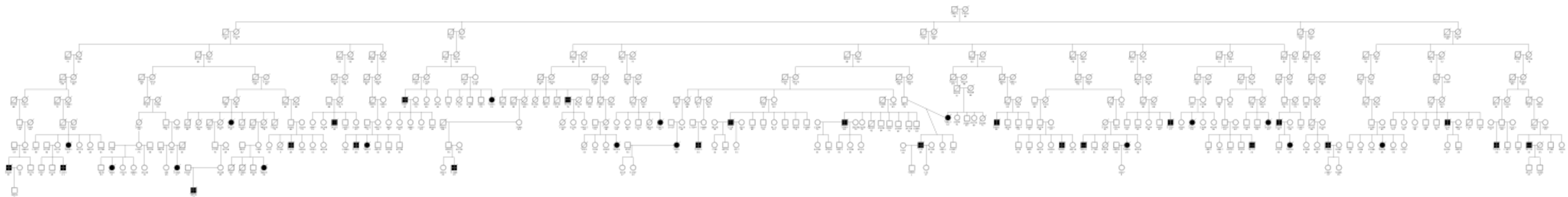
How can we deal with graphs that contain **rich attribute data**?



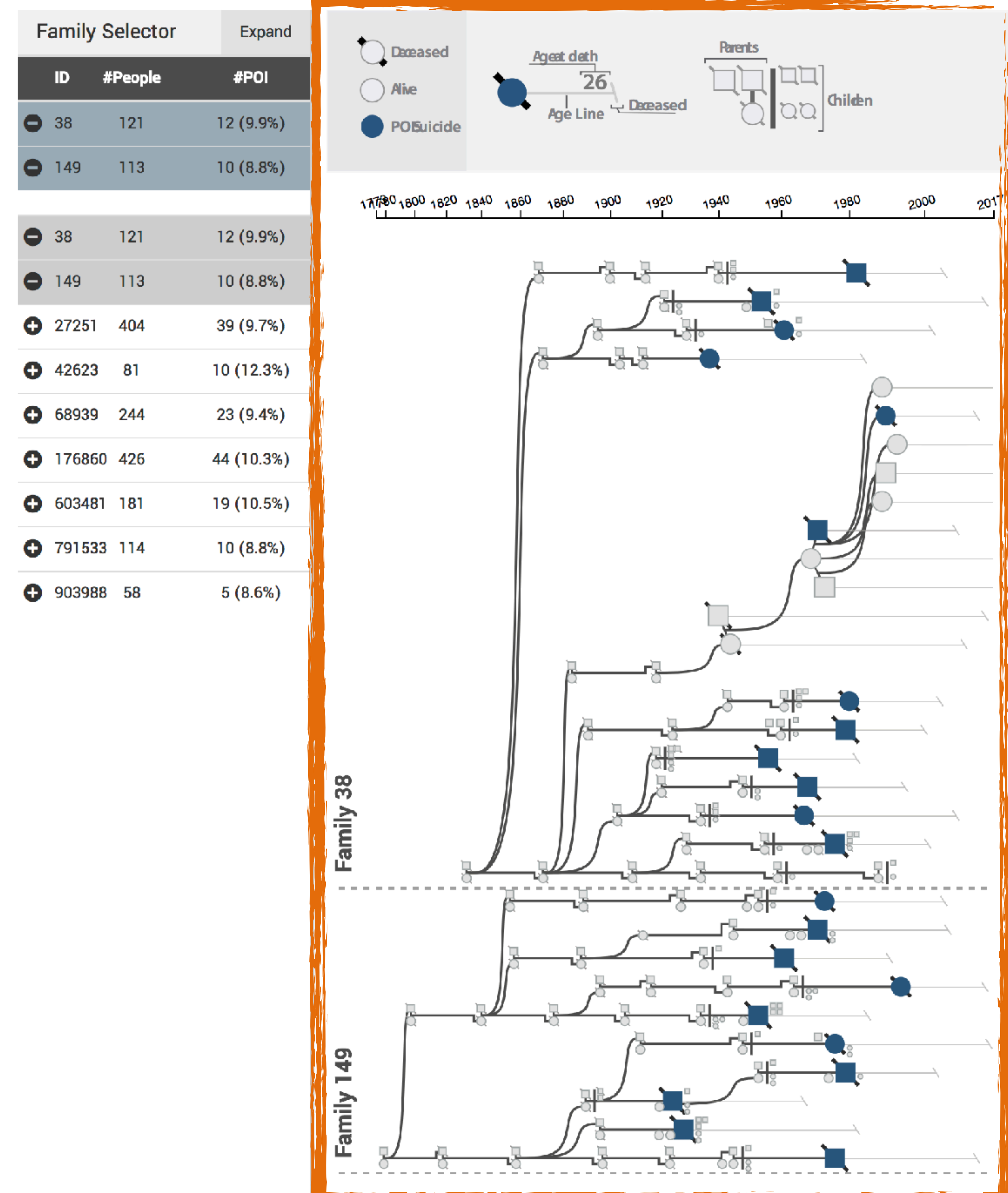
[McDonnell2009]



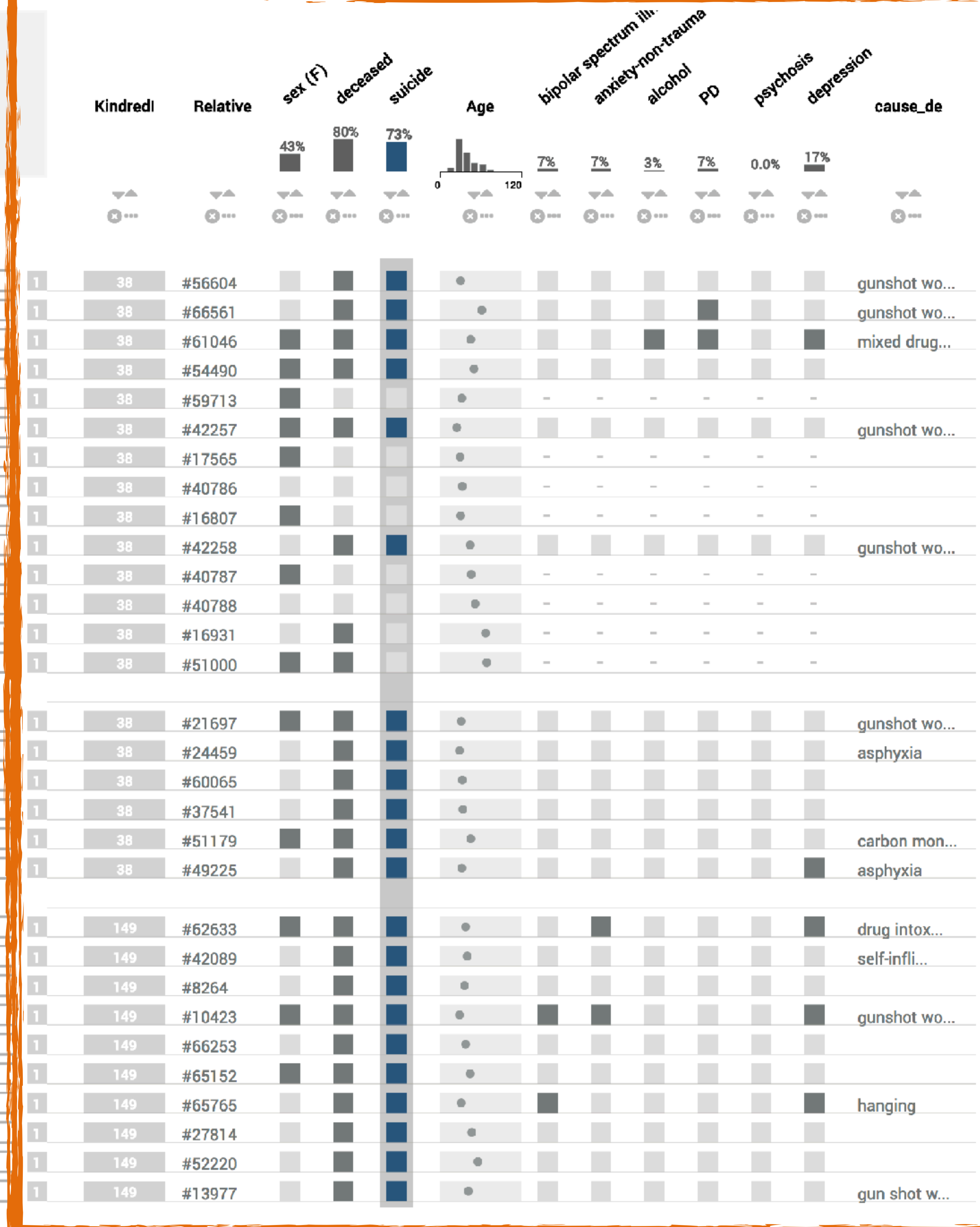
[Gehlenborg2010]



Genealogy with ~400 members rendered with Progeny

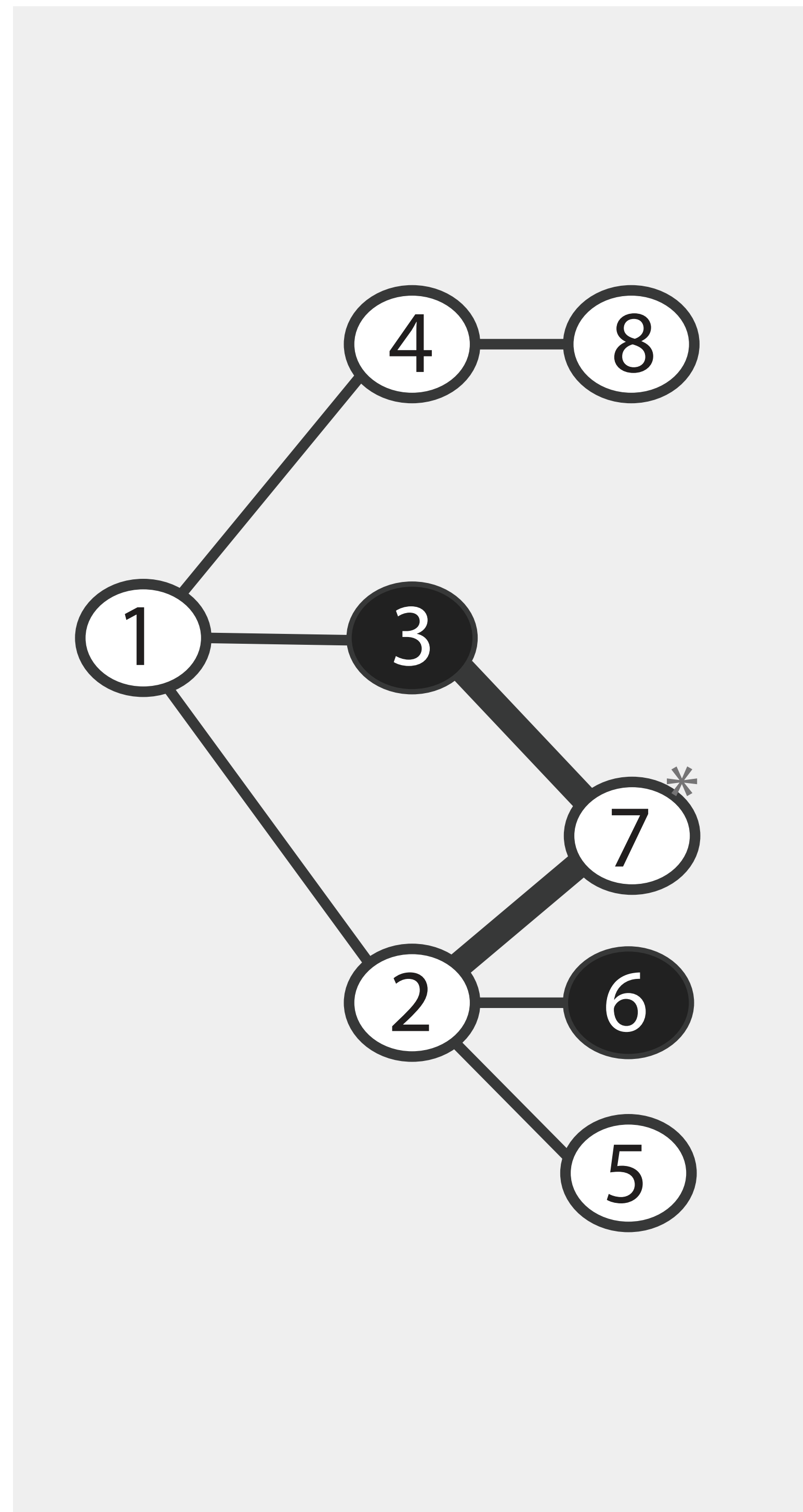


1. De-cycle and linearize graph

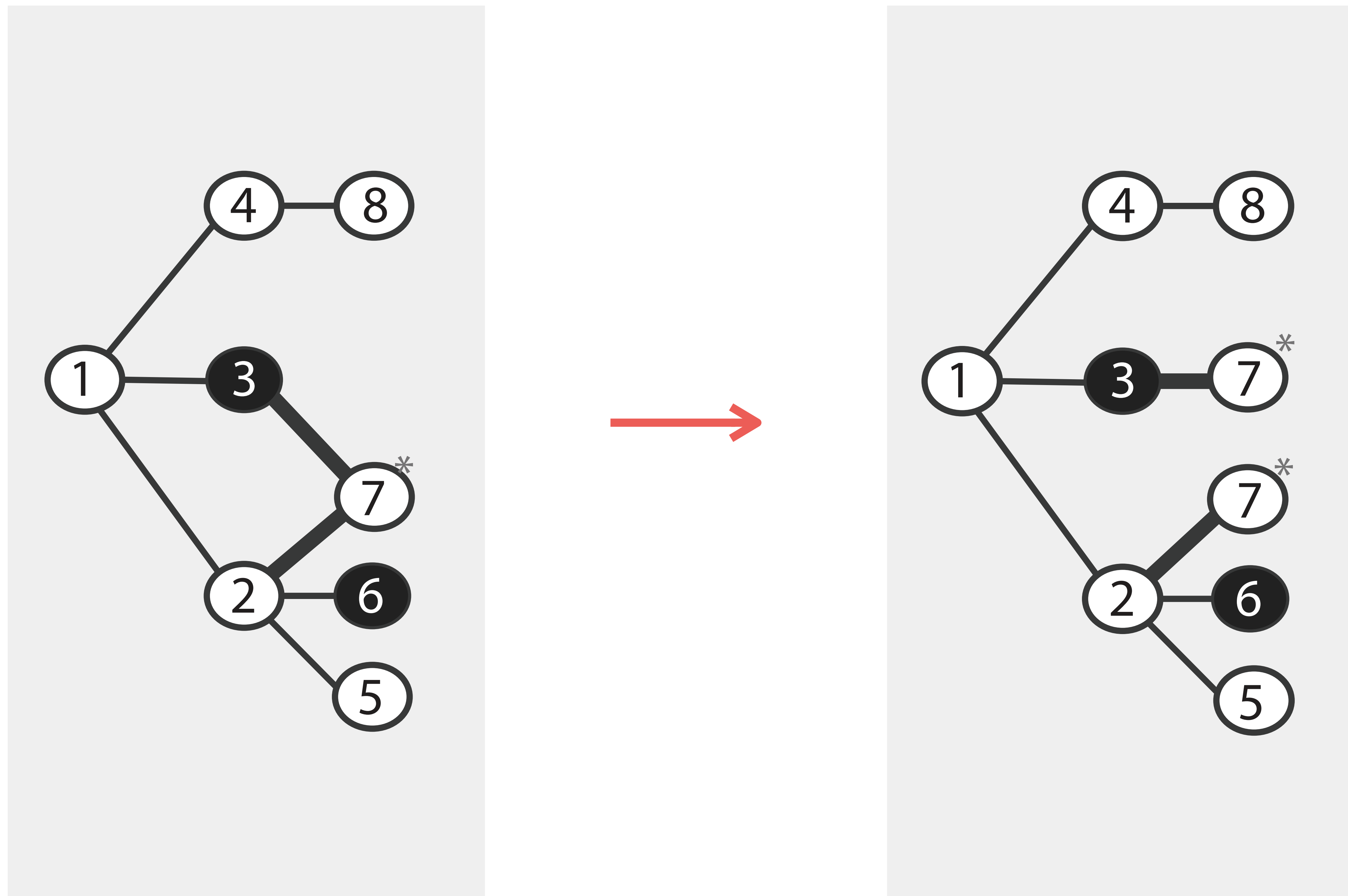


2. Plot attributes in table

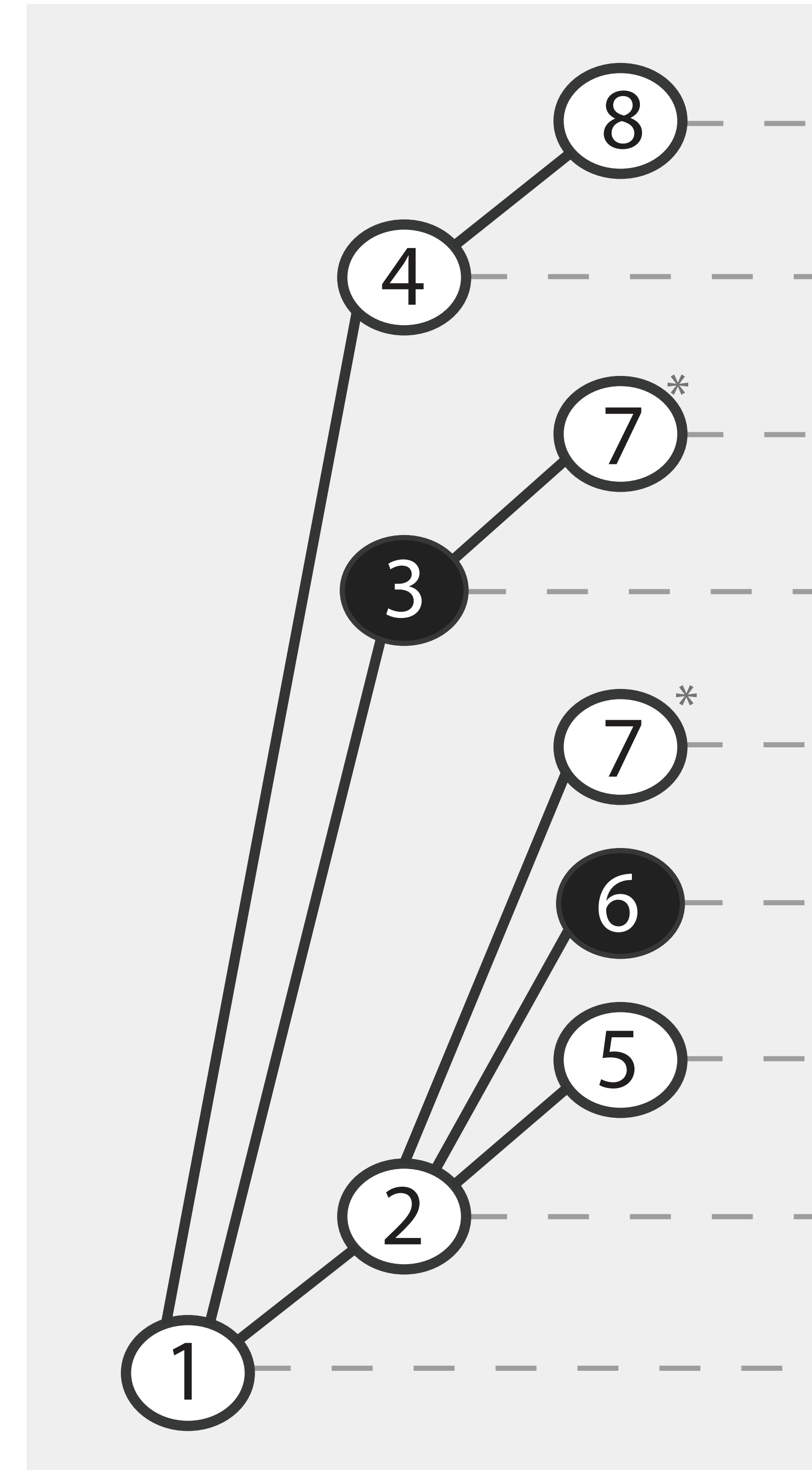
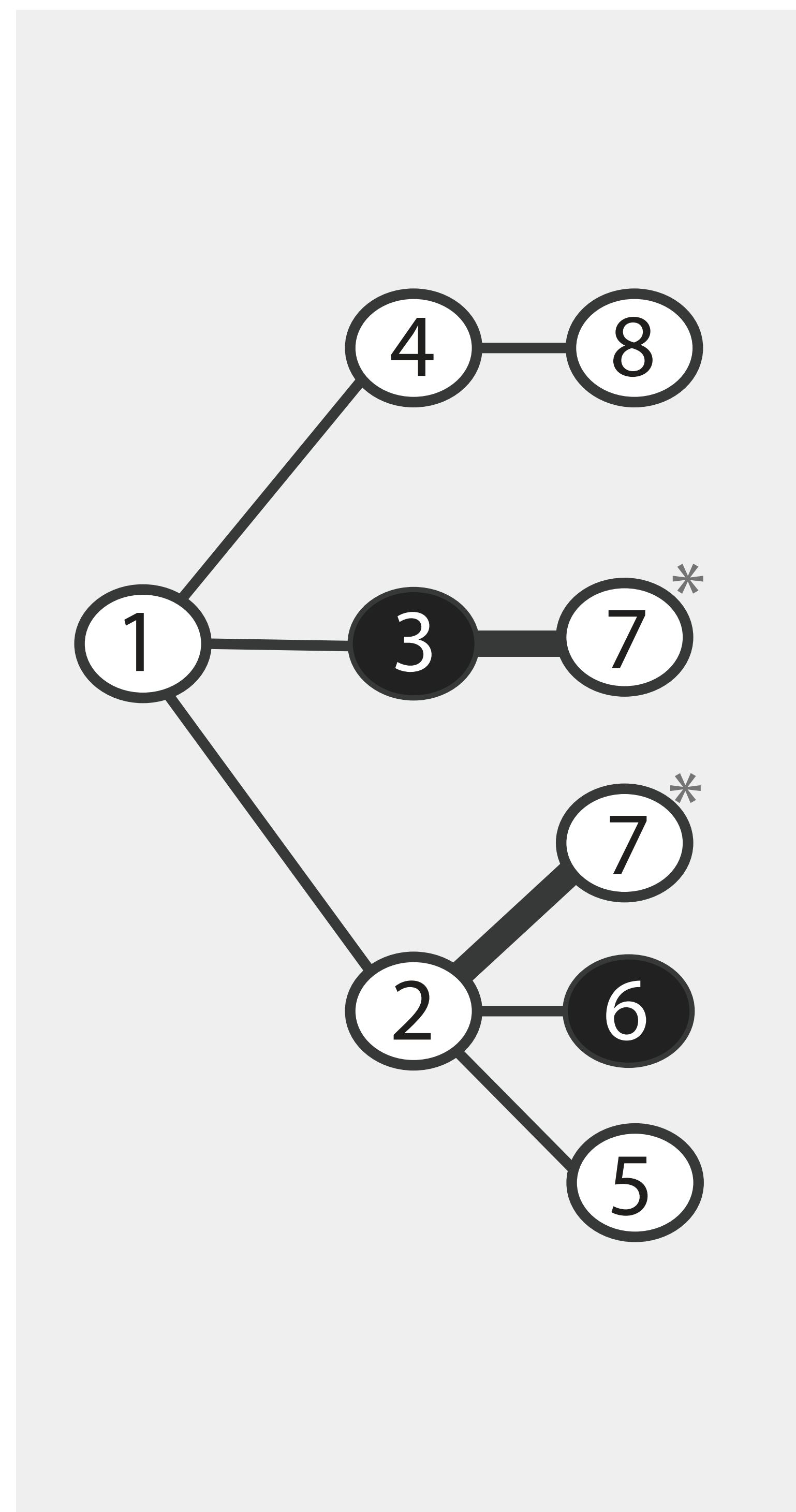
De-Cycling



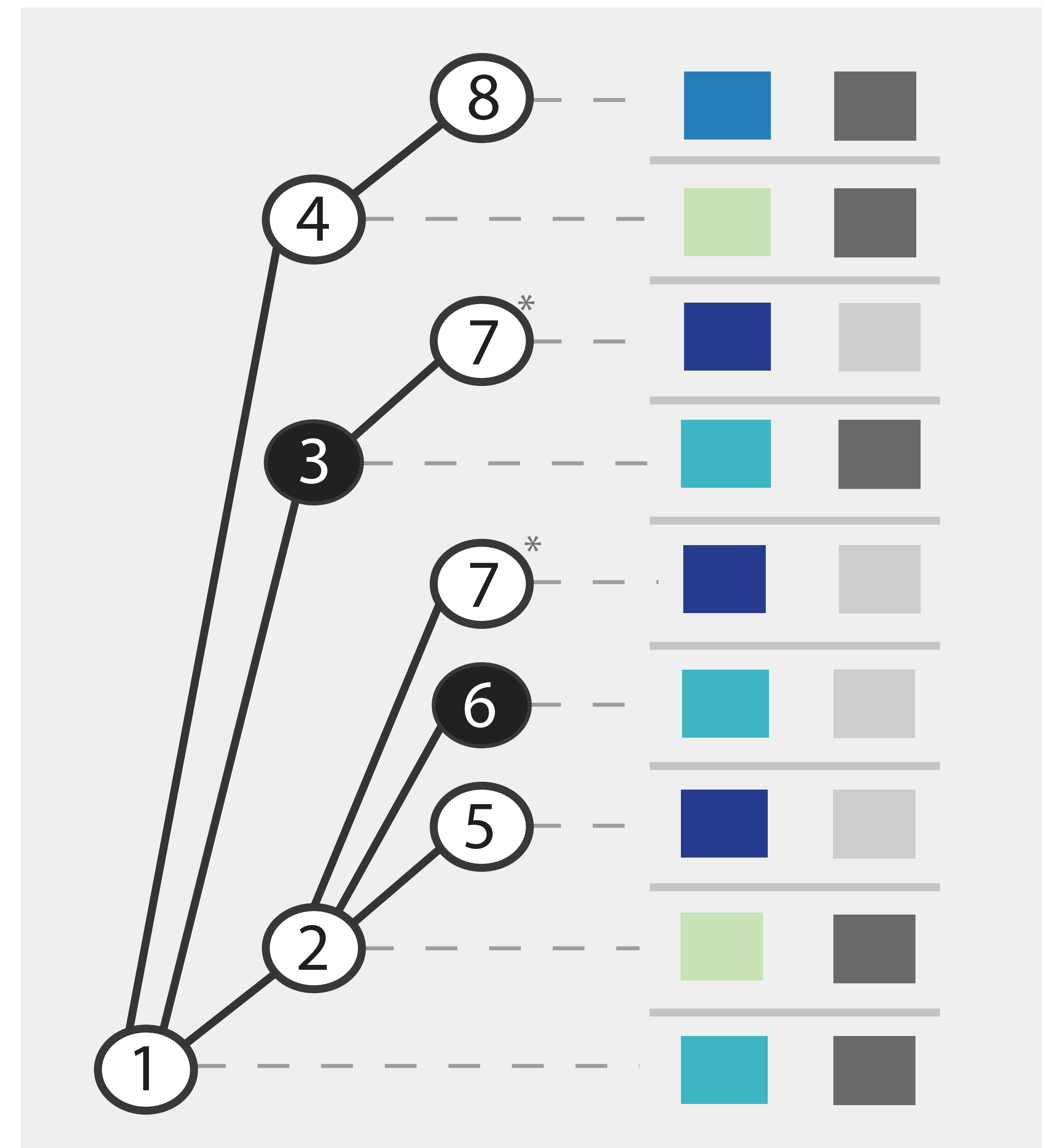
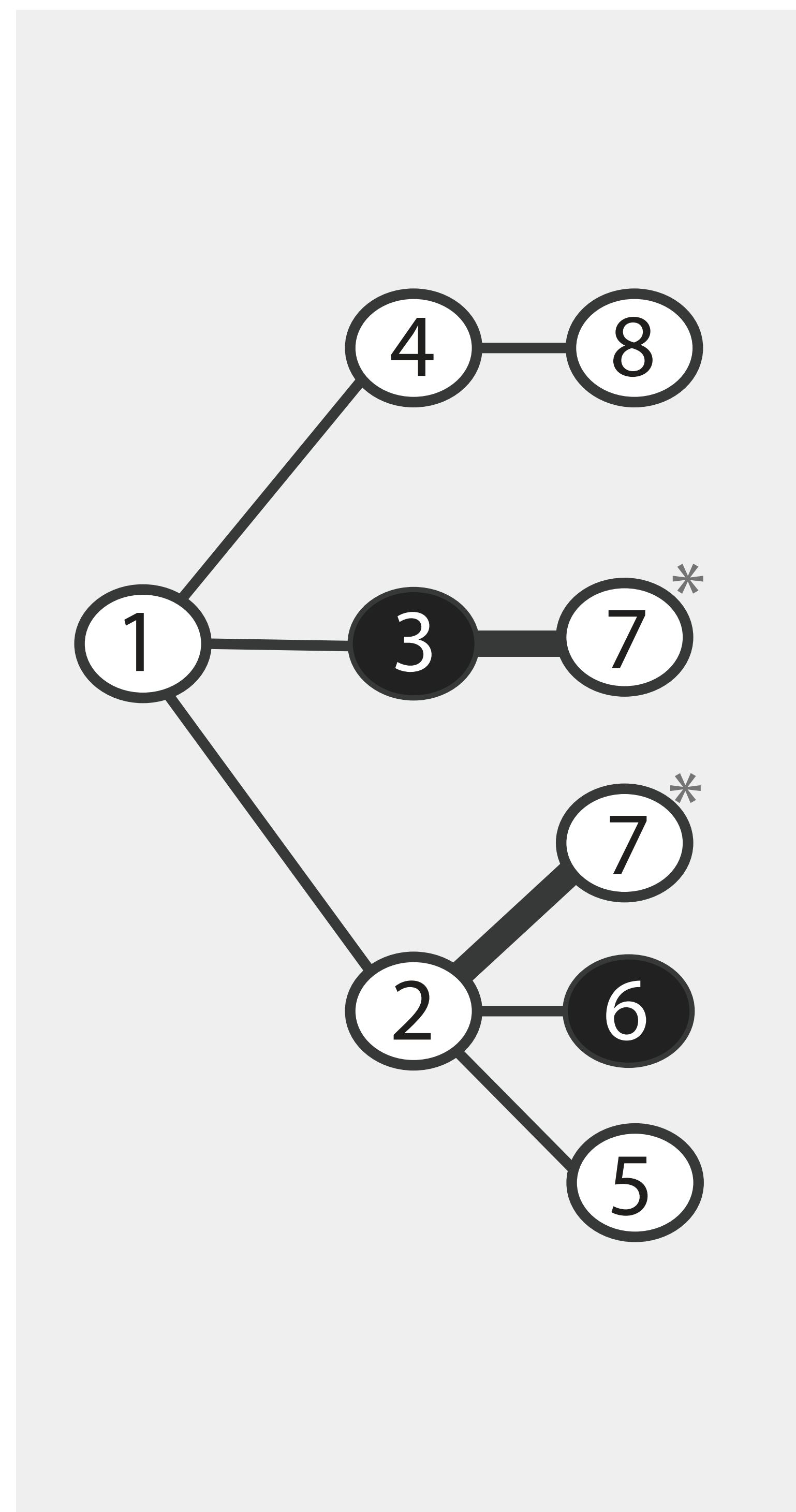
De-Cycling



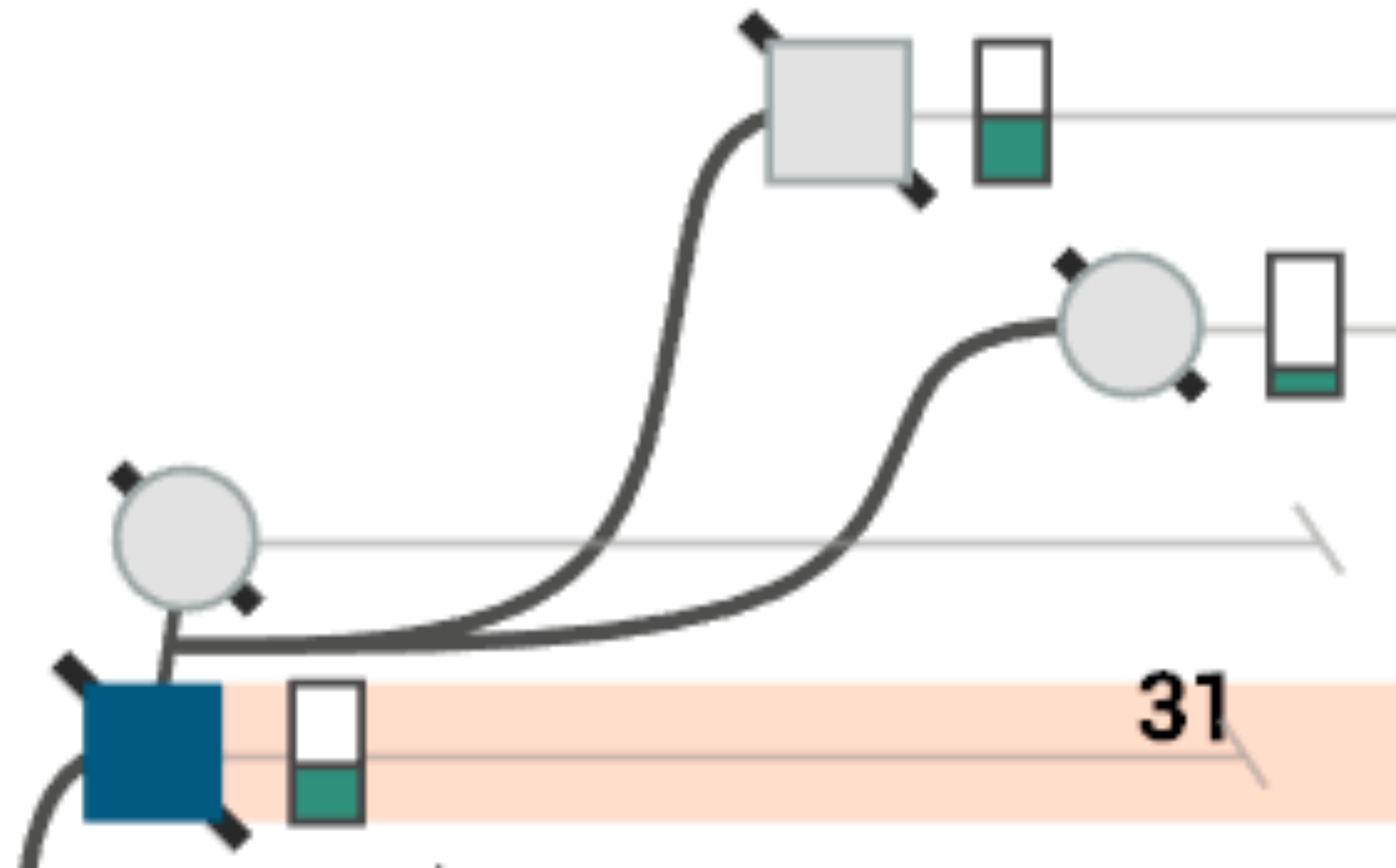
Linearization

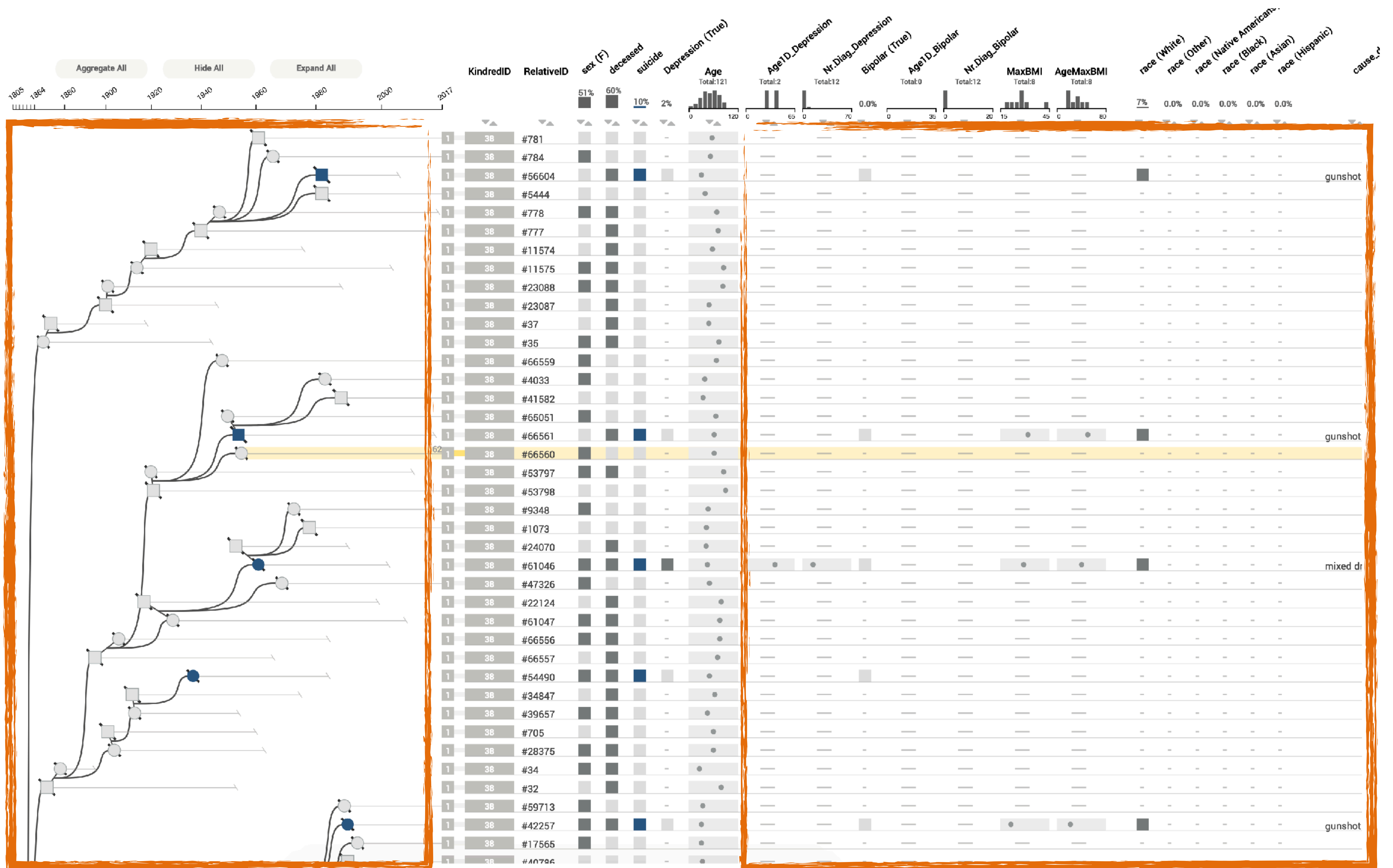


Linearization



Basic Encoding





Can't show many people

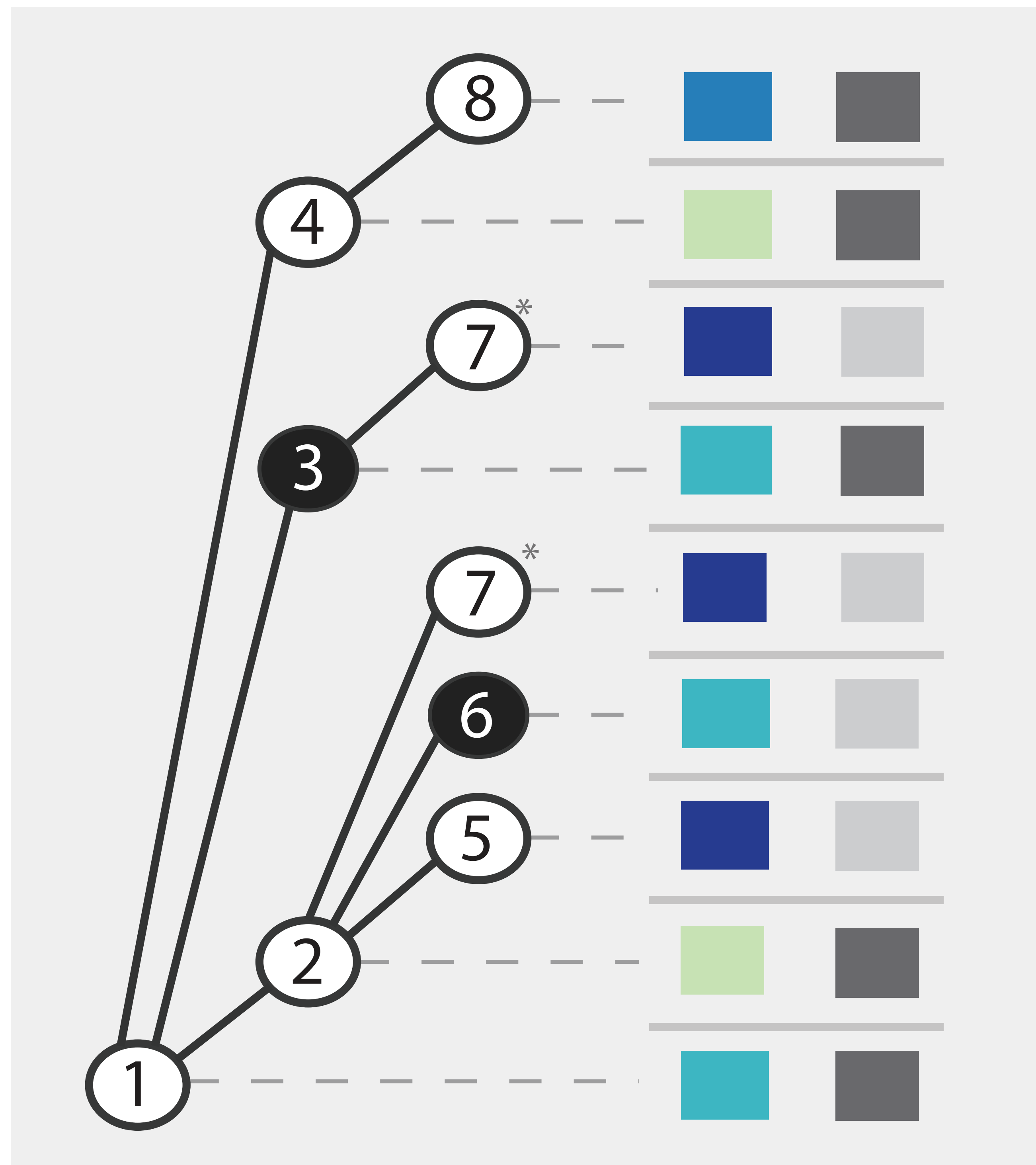
Lots of missing data

Aggregation

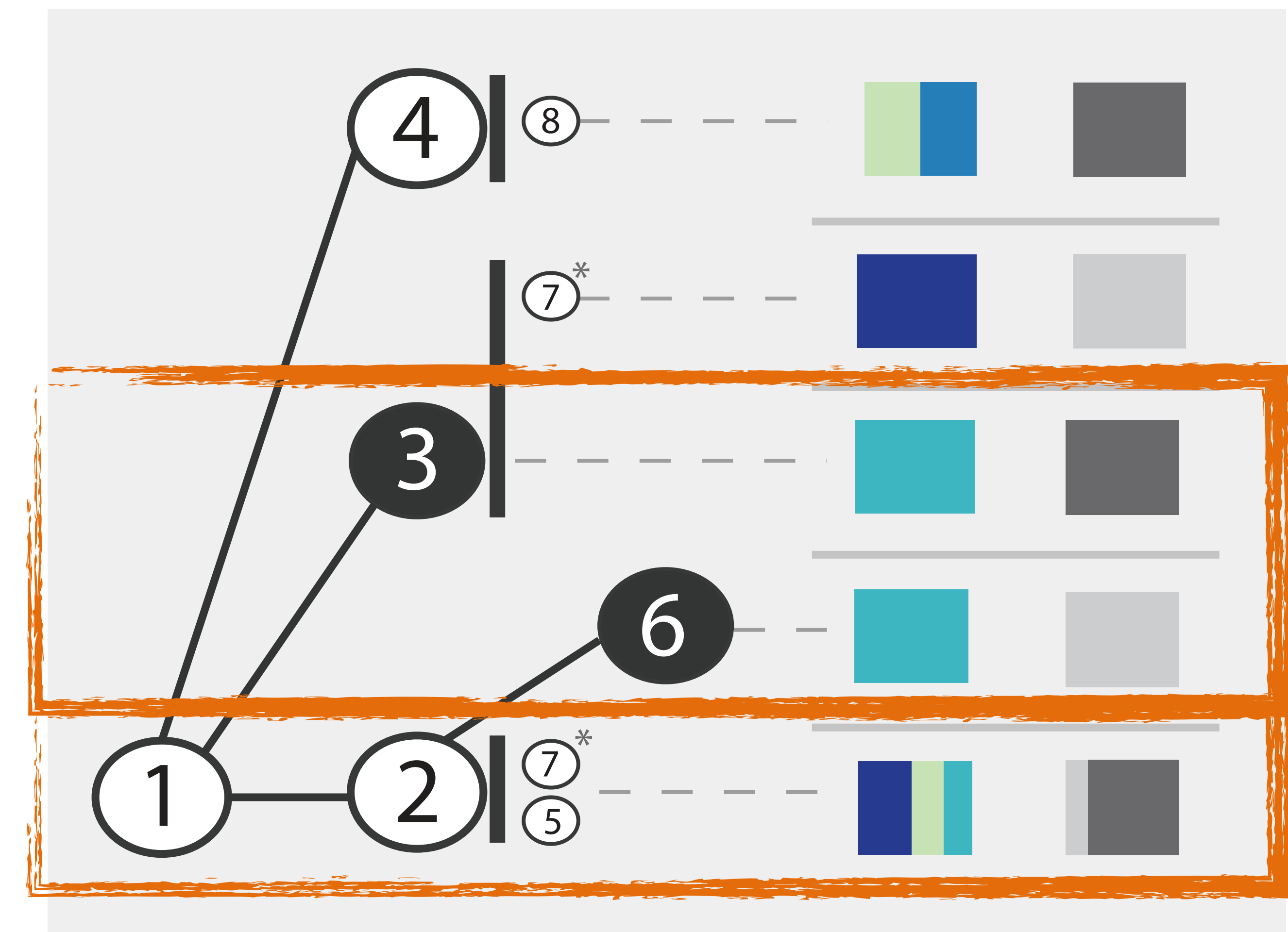


People of Interest

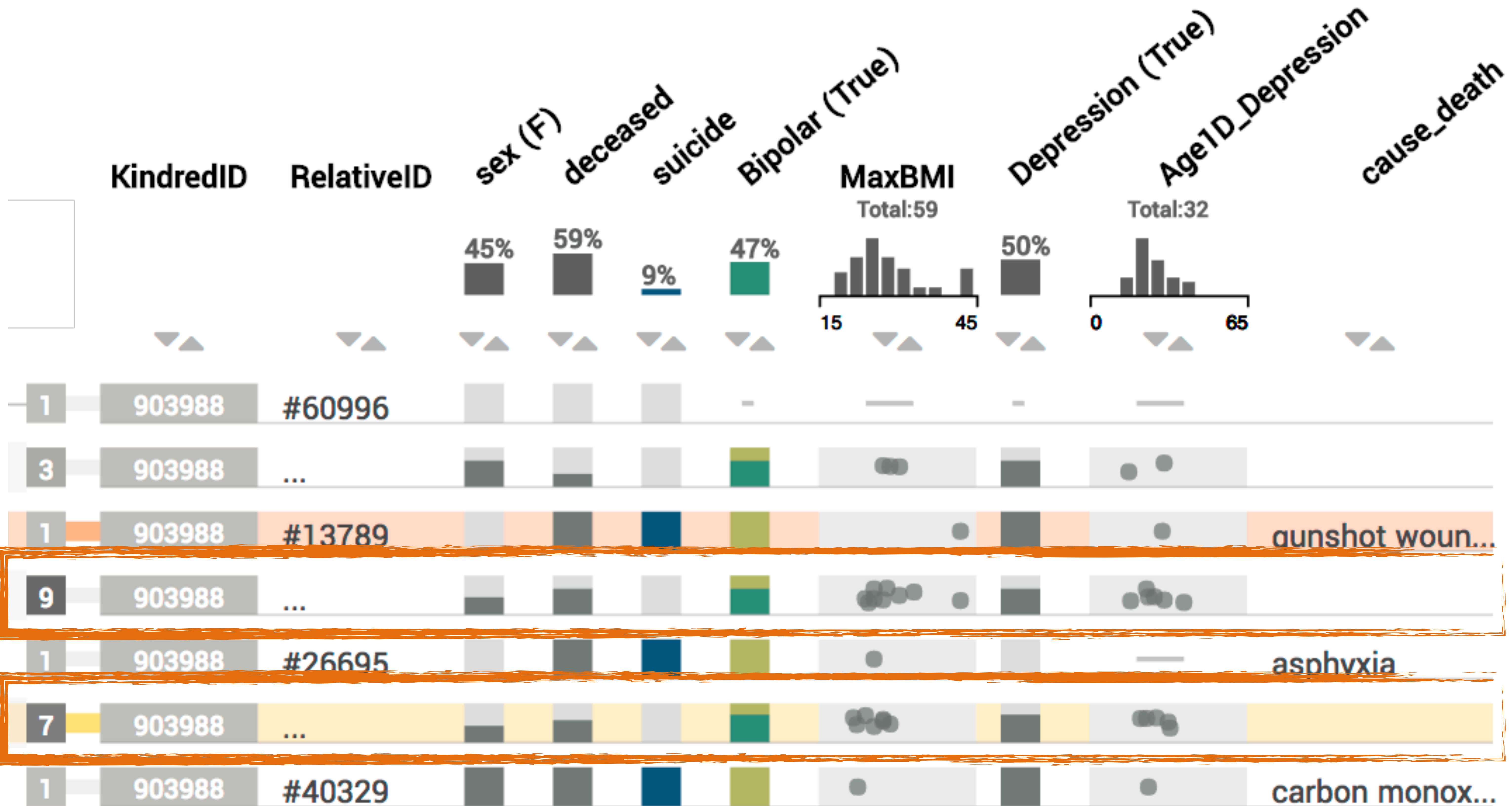
Aggregation



One row for every person of interest

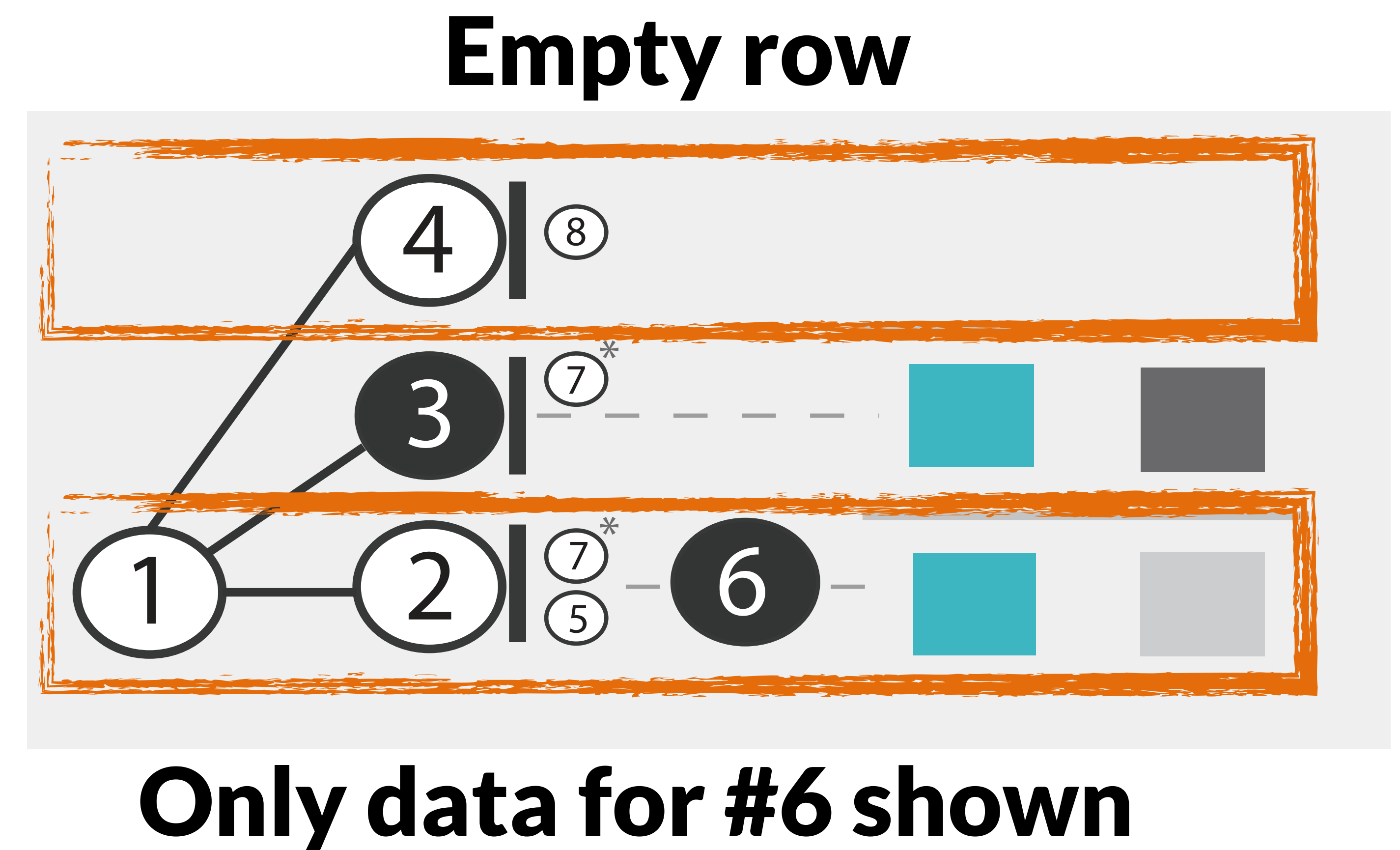
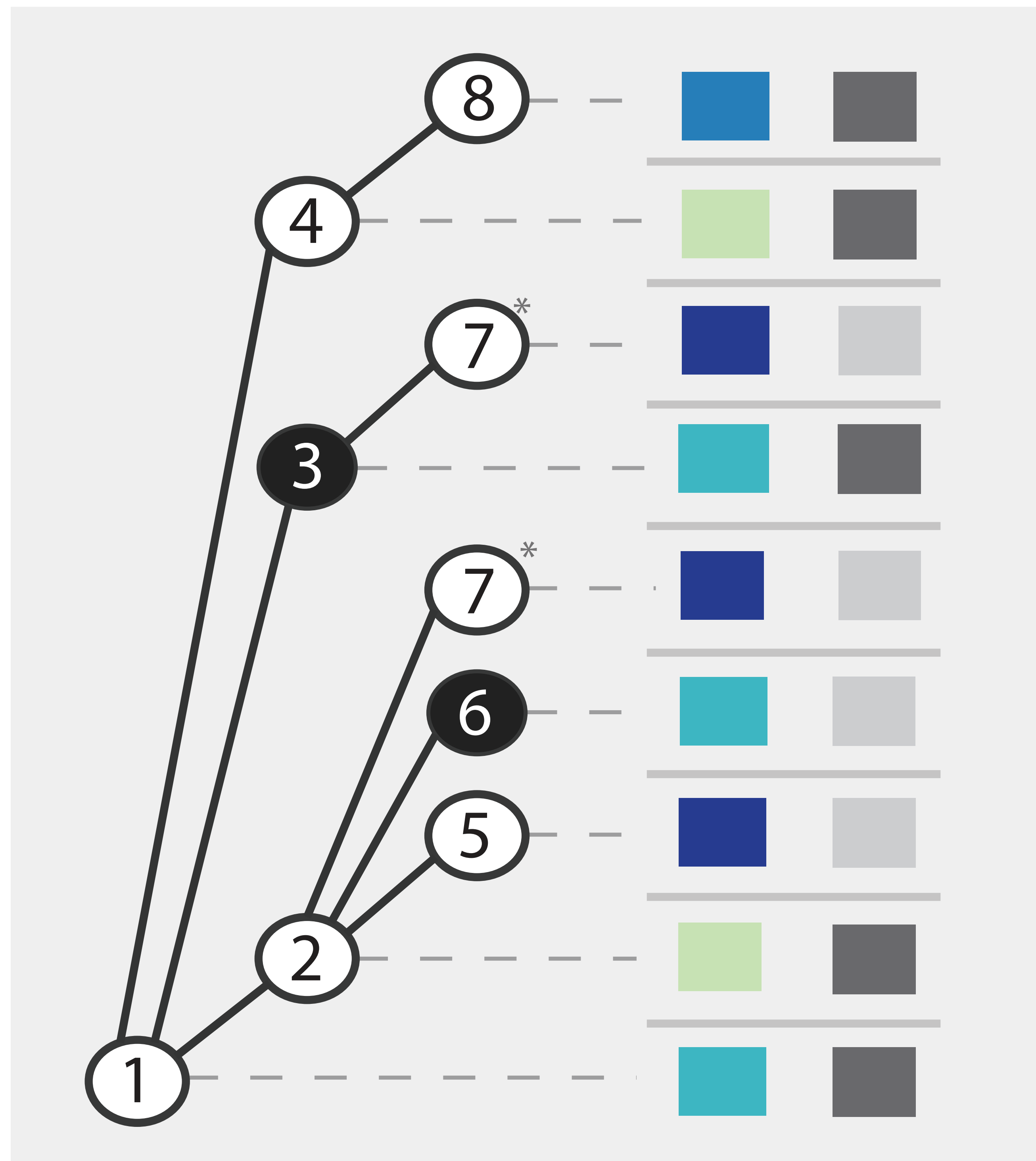


Others have to share a row



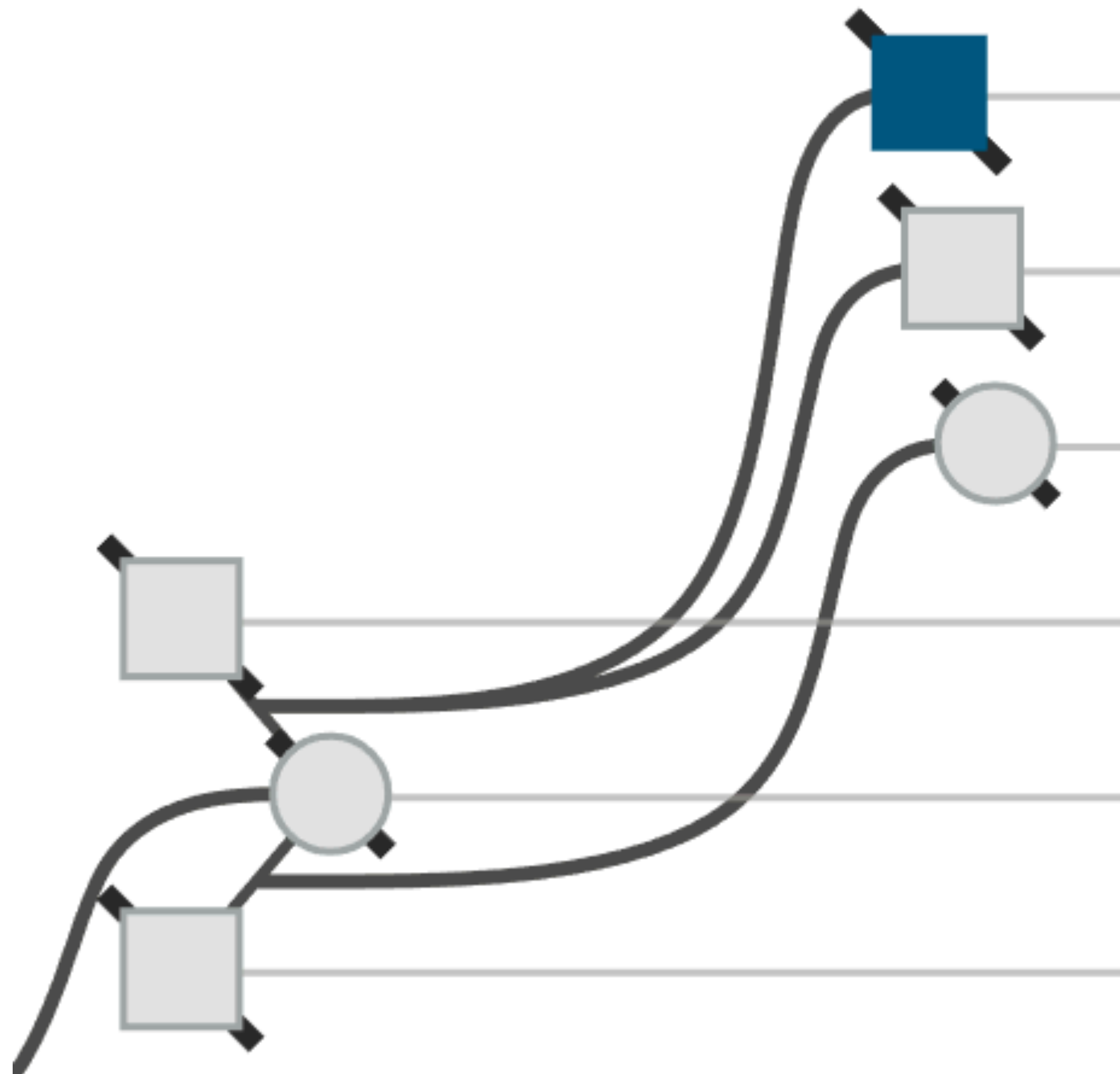
Aggregated Rows

More Aggressive: Hiding

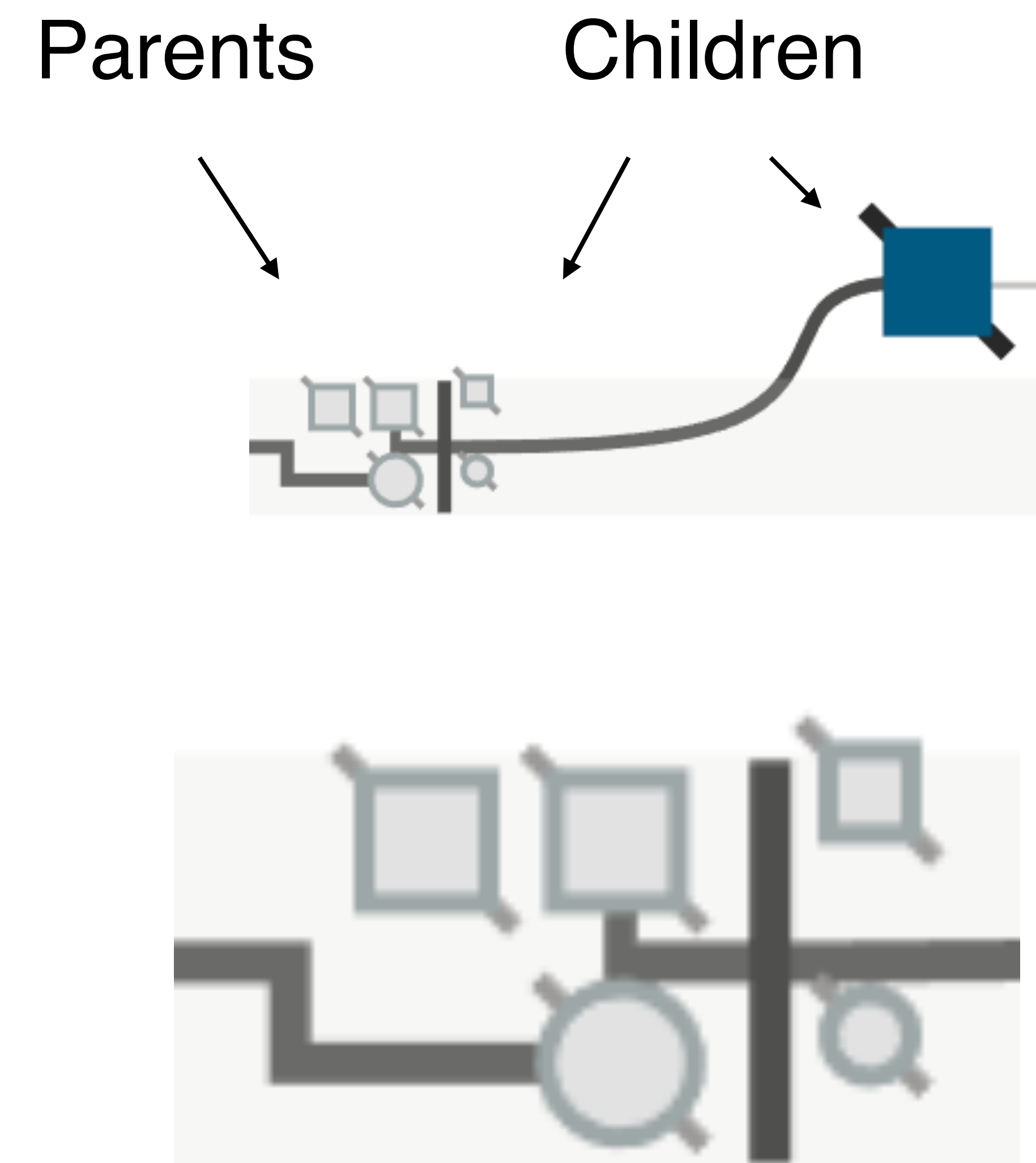


Implicit Encoding of Family

No Aggregation



Aggregation



Search and Filter Strategies

Find families enriched for a trait

Scan relevant cases for relationships

<http://lineage.caleydoapp.org>

Paper: <http://vdl.sci.utah.edu/>

Live Demo

Family Selector			Expand
ID	#People	#POI	PD
38	121	12 (9.9%)	2 (16.7%)
27251404		39 (9.7%)	6 (15.4%)

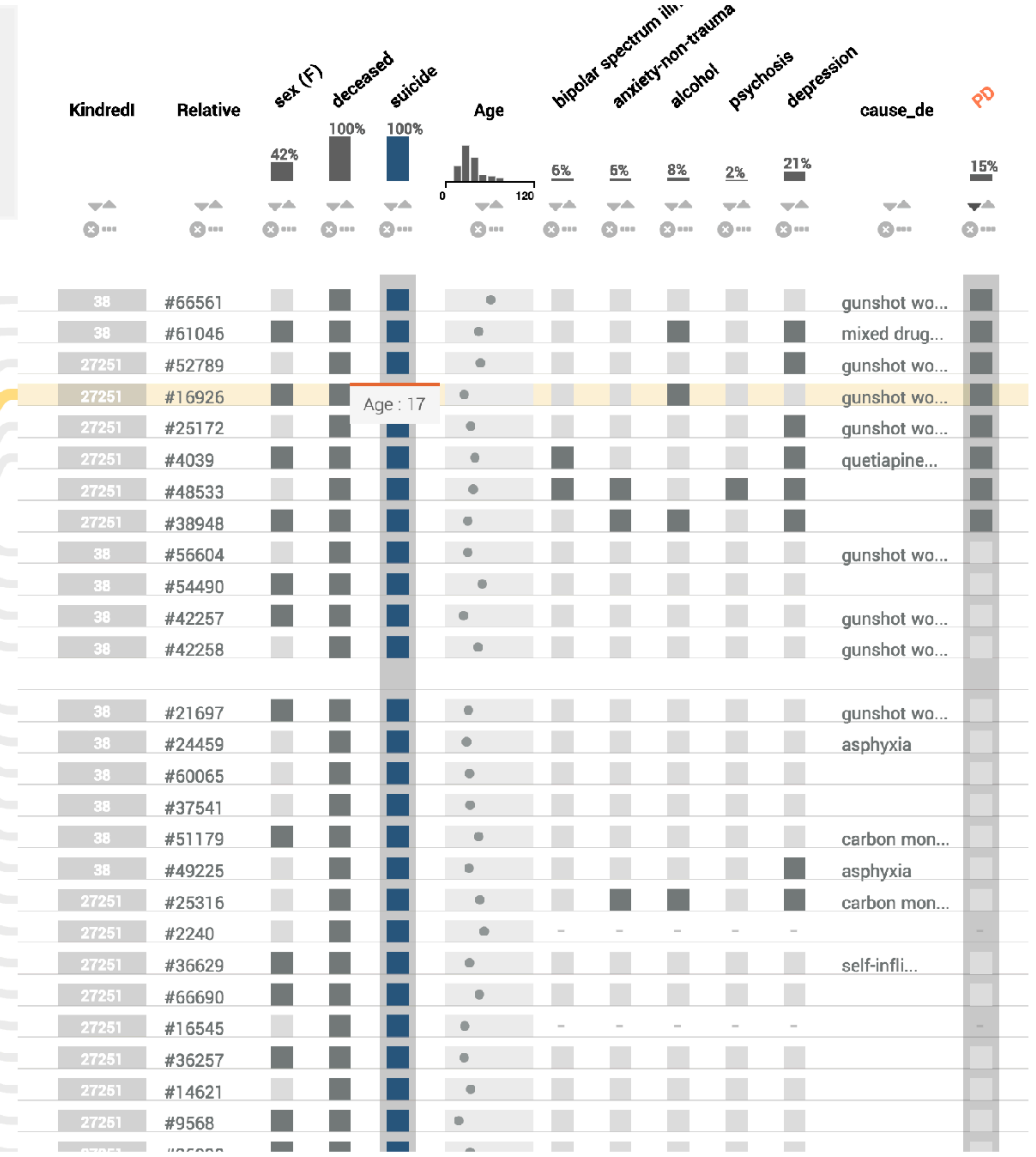
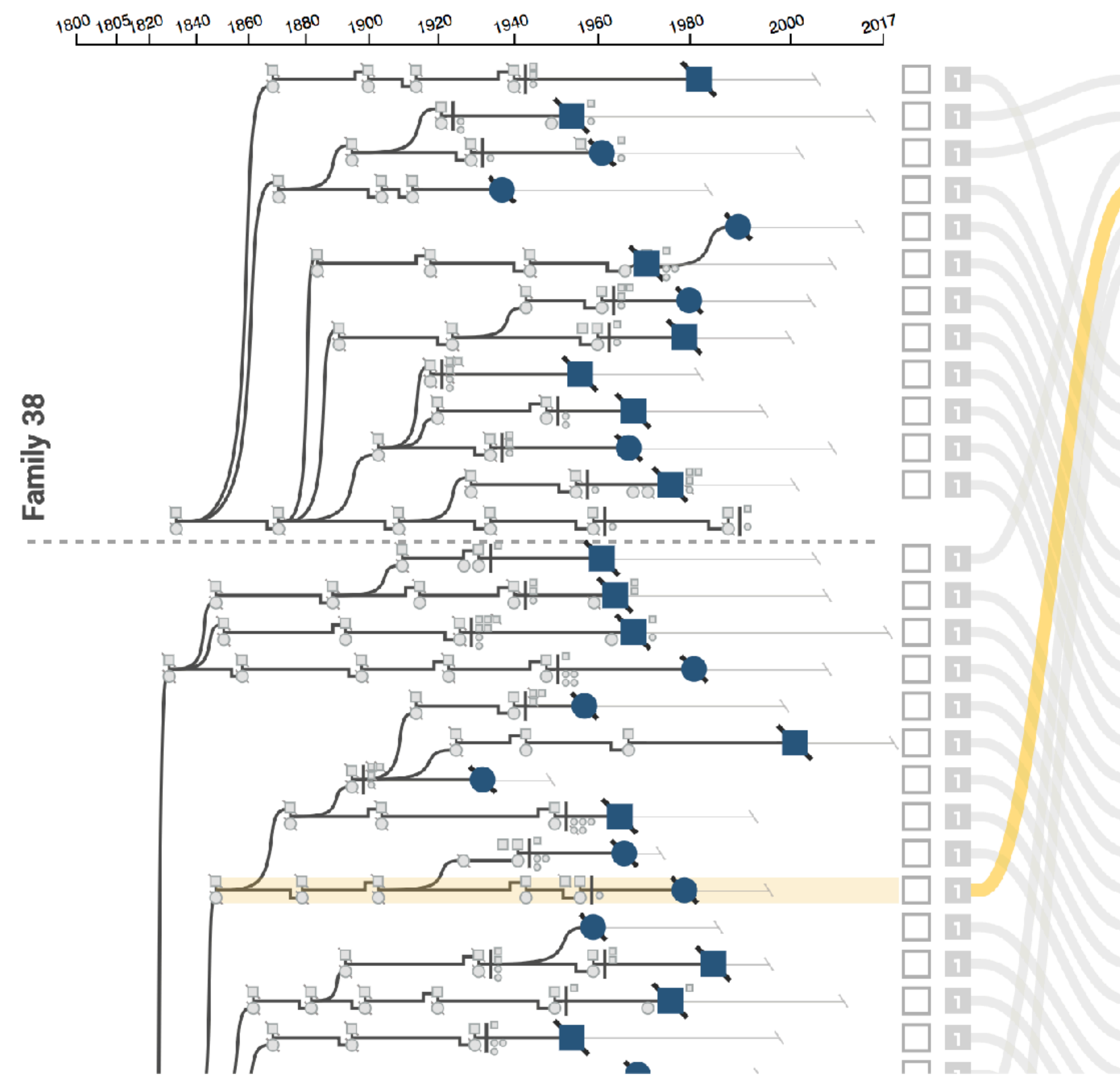
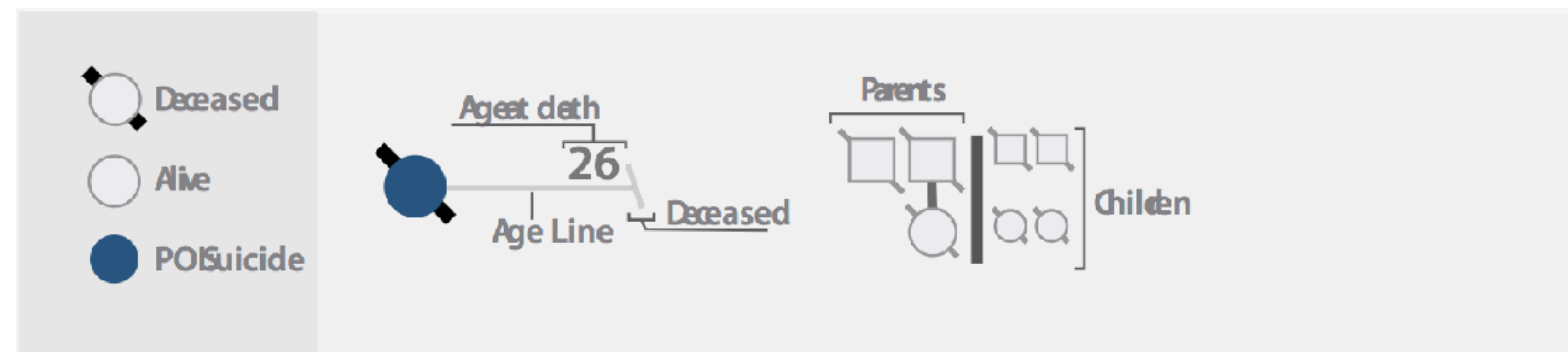
38	121	12 (9.9%)	2 (16.7%)
149	113	10 (8.8%)	0 (0%)
27251404		39 (9.7%)	6 (15.4%)

+	42623 81	10 (12.3%)	1 (10%)
+	68939244	23 (9.4%)	4 (17.4%)
+	17686026	44 (10.3%)	4 (9.1%)

60348181	19	1
	(10.5%)	(5.3%)

79153314	10	0 (0%)
	(8.8%)	

90398858	5	0 (0%)
	(8.6%)	



Next Steps

Show variants and other genetic data

Figure mode

Extend to other datasets

UPDB Users: Cancer, other psychiatric, cardiovascular, etc.

Genealogical datasets becoming more common

RESEARCH ARTICLE

Quantitative analysis of population-scale family trees with millions of relatives

Joanna Kaplanis^{1,2,*}, Assaf Gordon^{1,2,*}, Tal Shor^{3,4}, Omer Weissbrod⁵, Dan Geiger⁴, Mary Wahl^{1,2,6}, Michael Gershovits², Barak ...

+ See all authors and affiliations

Science 13 Apr 2018:
Vol. 360, Issue 6385, pp. 171-175
DOI: 10.1126/science.aam9309

Smaller pedigrees, like trios?

Phylogenies, ...

Could we use something like
Lineage for general
Multivariate Networks?

Juniper: A Tree+Table Approach to Multivariate Graph Visualization



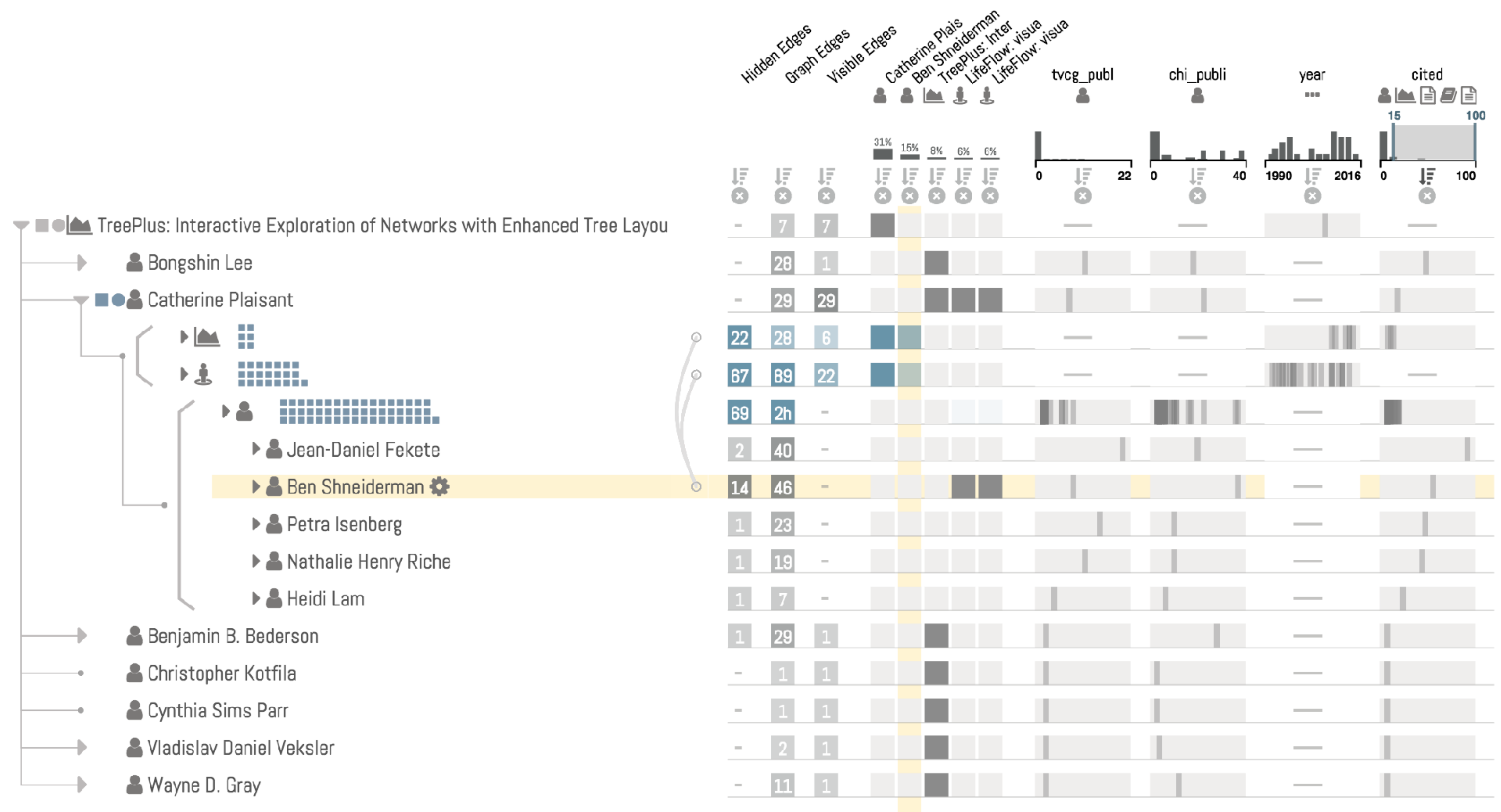
Carolina Nobre



Marc Streit



Alexander Lex



We're hiring PostDocs and accept PhD Students!



visualization
design lab



Miriah Meyer



Alexander Lex



Alexander Lex
@alexander_lex
<http://alexander-lex.net>



Thanks to: Carolina Nobre, Hilary Coon, Marc Streit, Nils Gehlenborg

Funded by the Utah Genome Project, NIH, NSF and DoD



**visualization
design lab**

