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

THE UNIVERSITY OF UTAH

What is Data Visualization and Why do We Care About it for Biomedical Applications

visualization design lab

www.sci.utah.edu

visualization design lab

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

Carolina Nobre Sam Quinan Jimmy Moore Kiran Gadhave Alexander Lex

Alex Bigelow Pascal Goffin Nina McCurdy Ethan Kerzner Jennifer Rogers Haihan Lin

Miriah Meyer

Ilkin Safarli

datavisyn.io

datavisyn

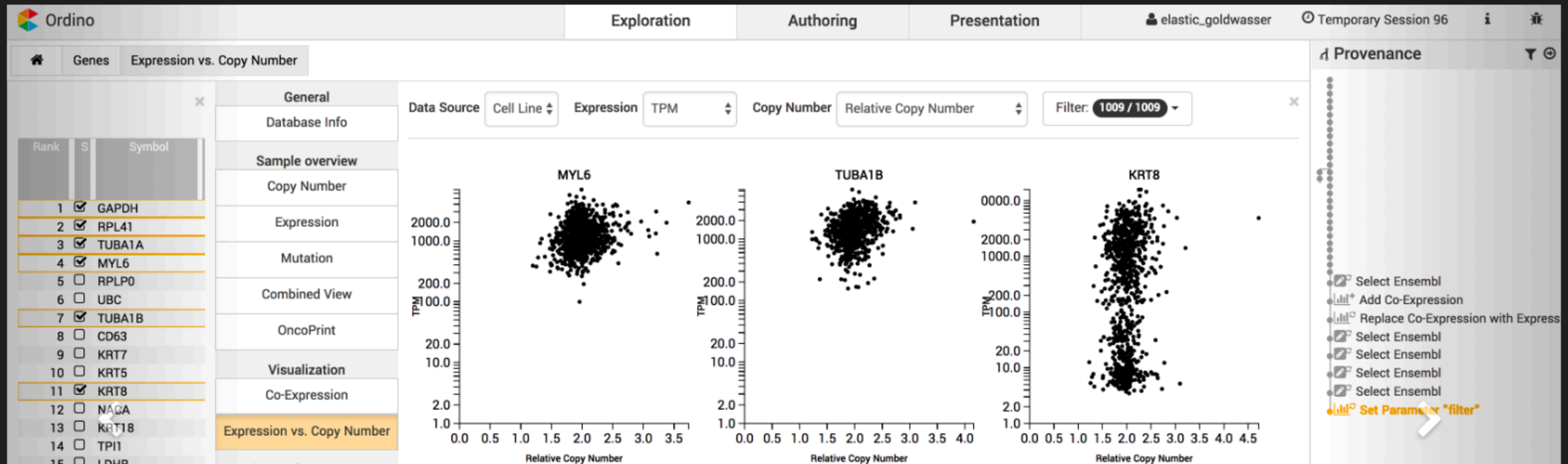
PRODUCTS SERVICES TEAM CONTACT

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

PRODUCTS

TARGET DISCOVERY PLATFORM

Our Target Discovery Platfrom 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.

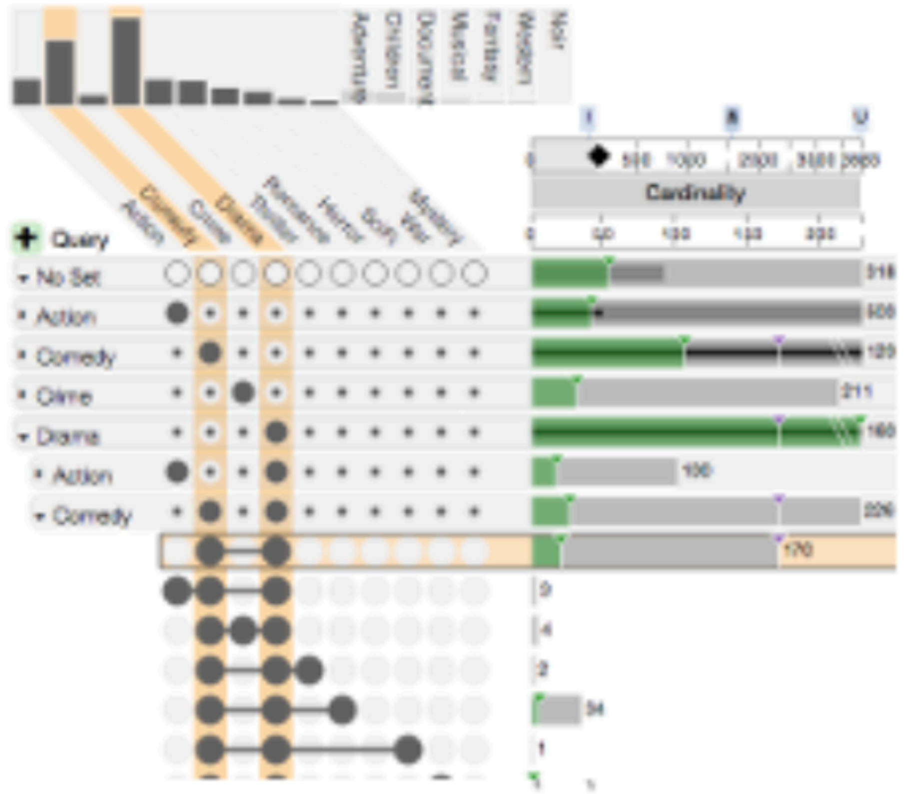


http://dataviscourse.net/professional/

Visualization for Professionals

THE UNIVERSITY OF UTAH

Home   Resources



UpSet visualizing intersecting sets | *Wind map* | *How states have shifted*

The amount and complexity of information produced in business, science, engineering, and everyday human activity is increasing at staggering rates. This course for professionals will expose you to visualization techniques and best practices that will enable you to efficiently communicate complex data.

Visualization for data discovery and communication is an important part of the data science pipeline. Good visualizations not only present a visual interpretation of data, but do so by improving comprehension, communication, and decision making.

In this course you will learn how to read, design, and build visualizations to effectively communicate your data! We tailor the course to the needs of our clients. Among the topics we can cover are:

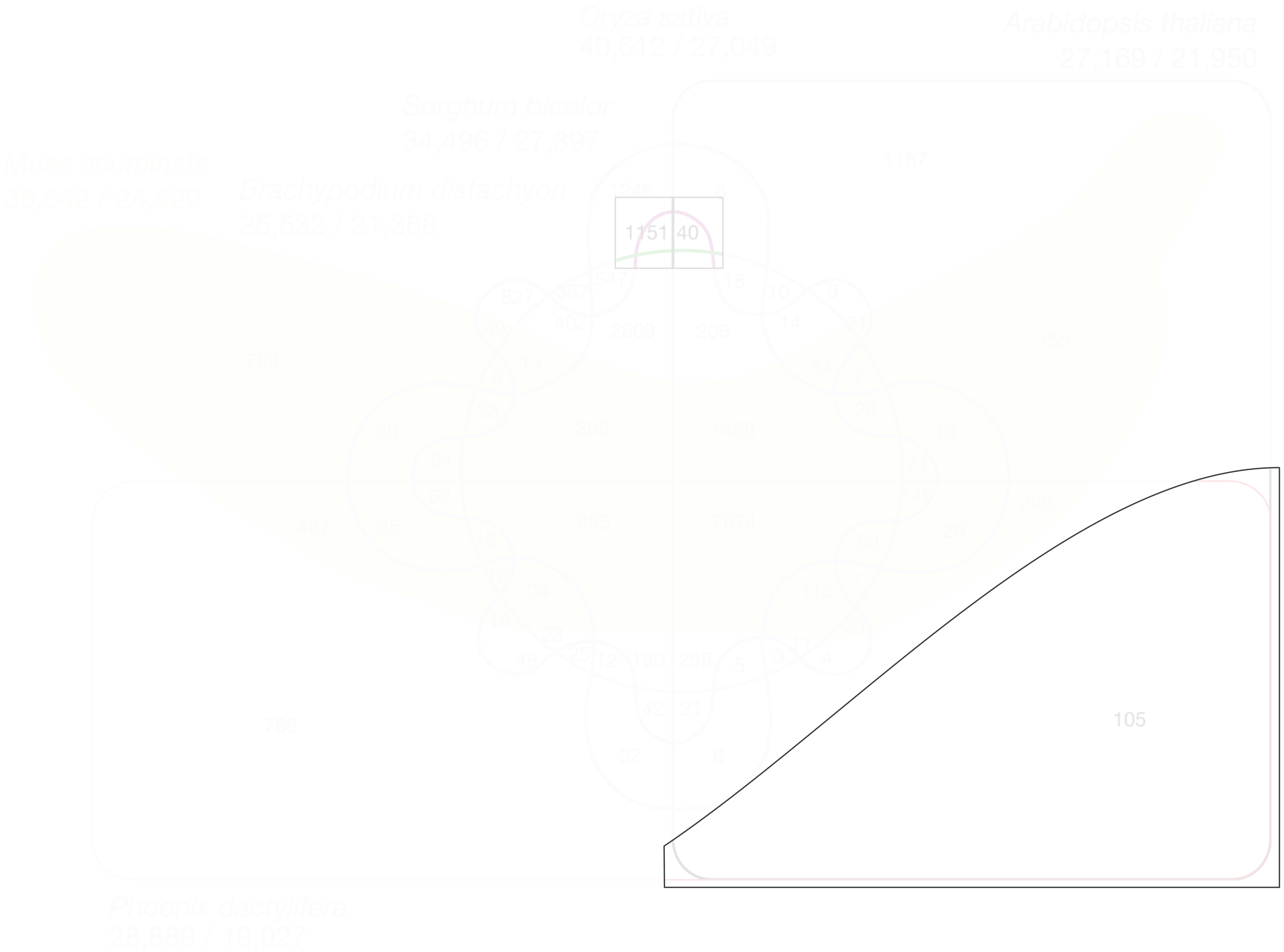
visualization

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***



[D’Hont et al., Nature, 2012]

Good Data Visualization

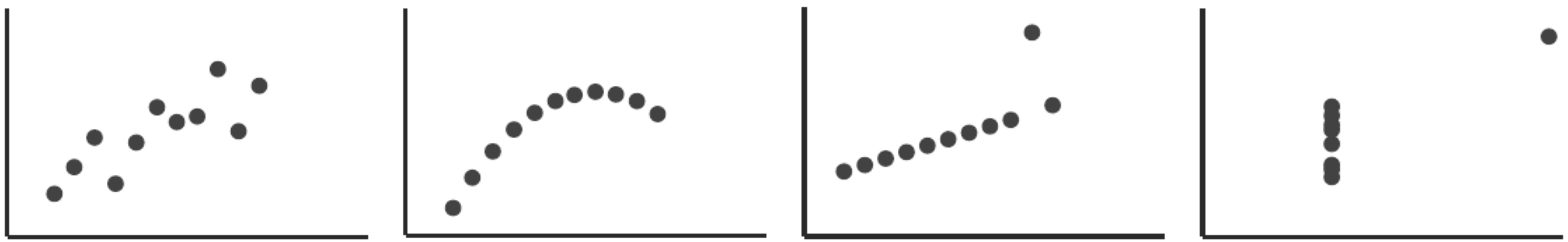
- ... makes data accessible
- ... combines strengths of humans and computers
- ... enables insight
- ... communicates

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.[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	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

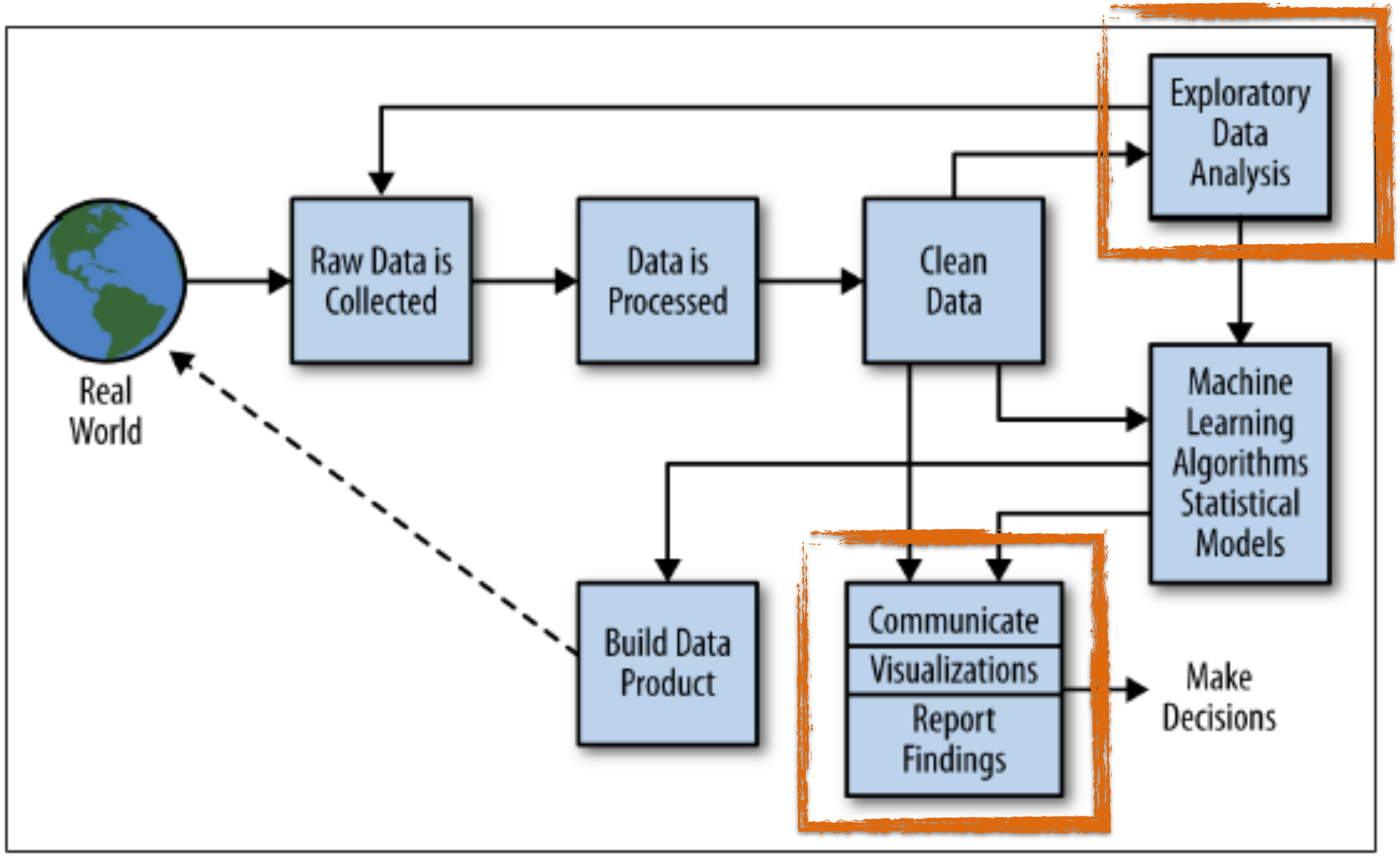
Anscombe’s Quartett



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



Doing Data Science: Straight Talk from the Frontline,

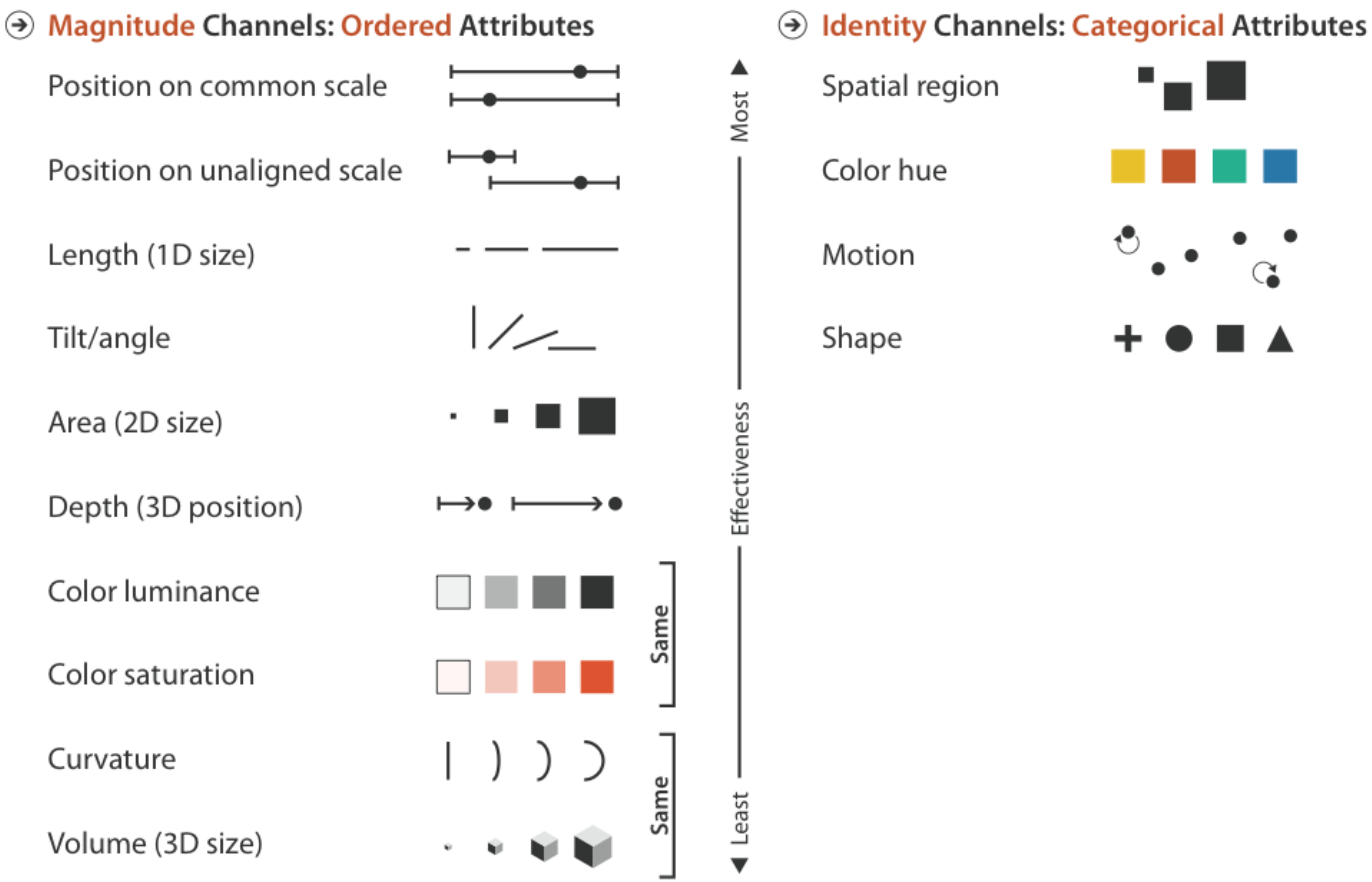
by Cathy O'Neil and Rachel Schutt

Visualization =

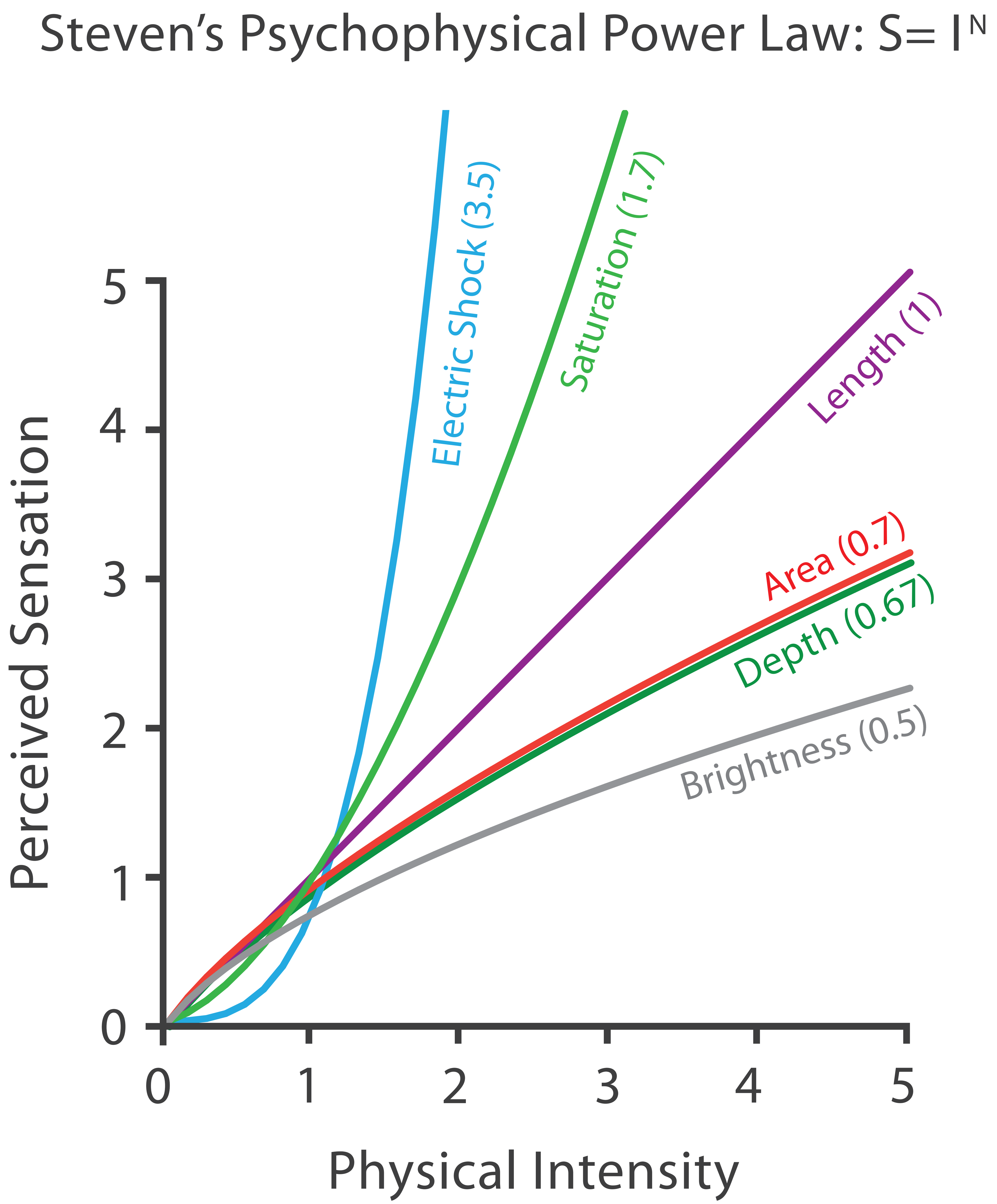
Human Data Interaction

The Vis Vocabulary

How can we effectively encode data using graphical marks?

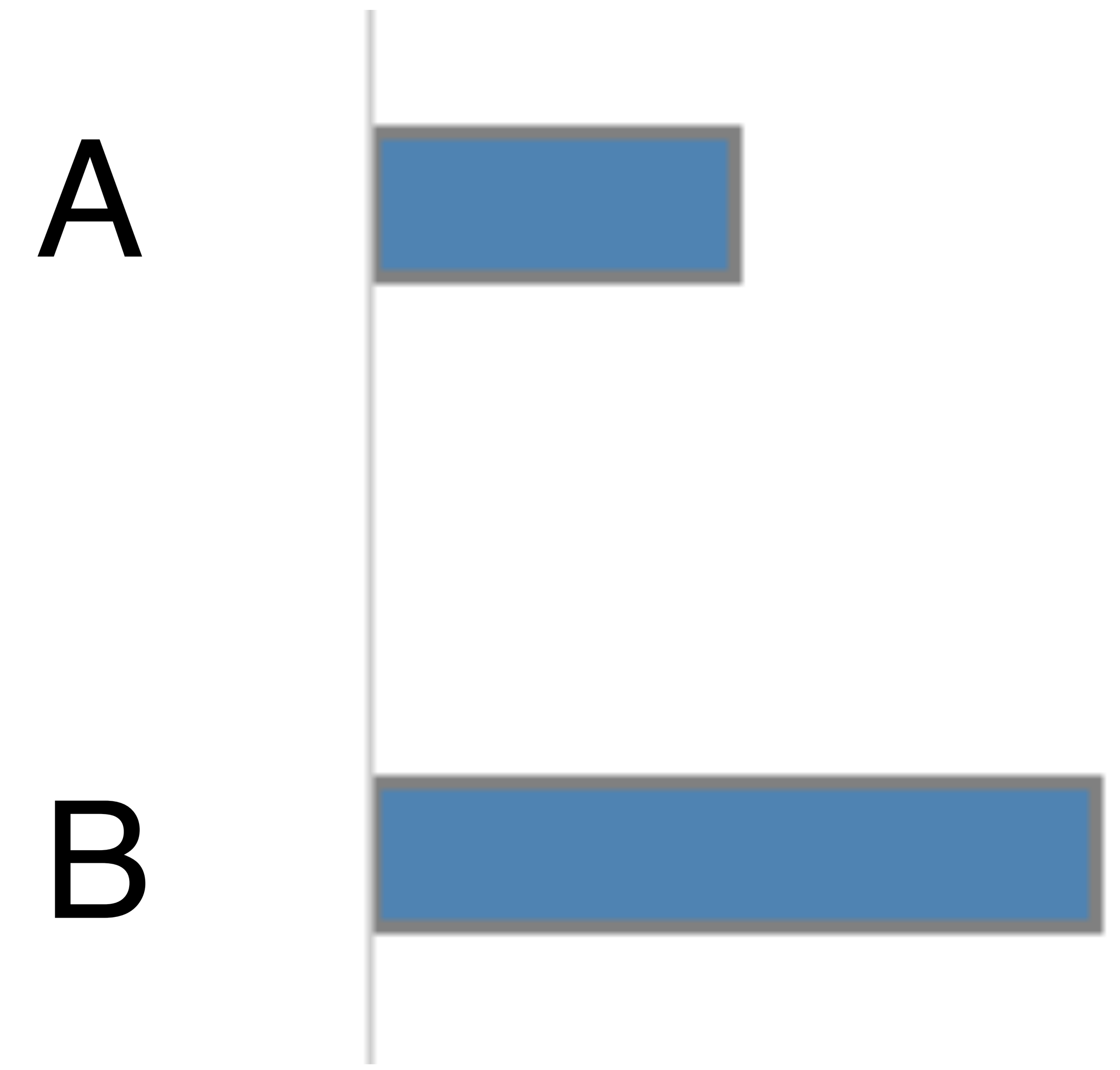


Why are quantitative channels different?



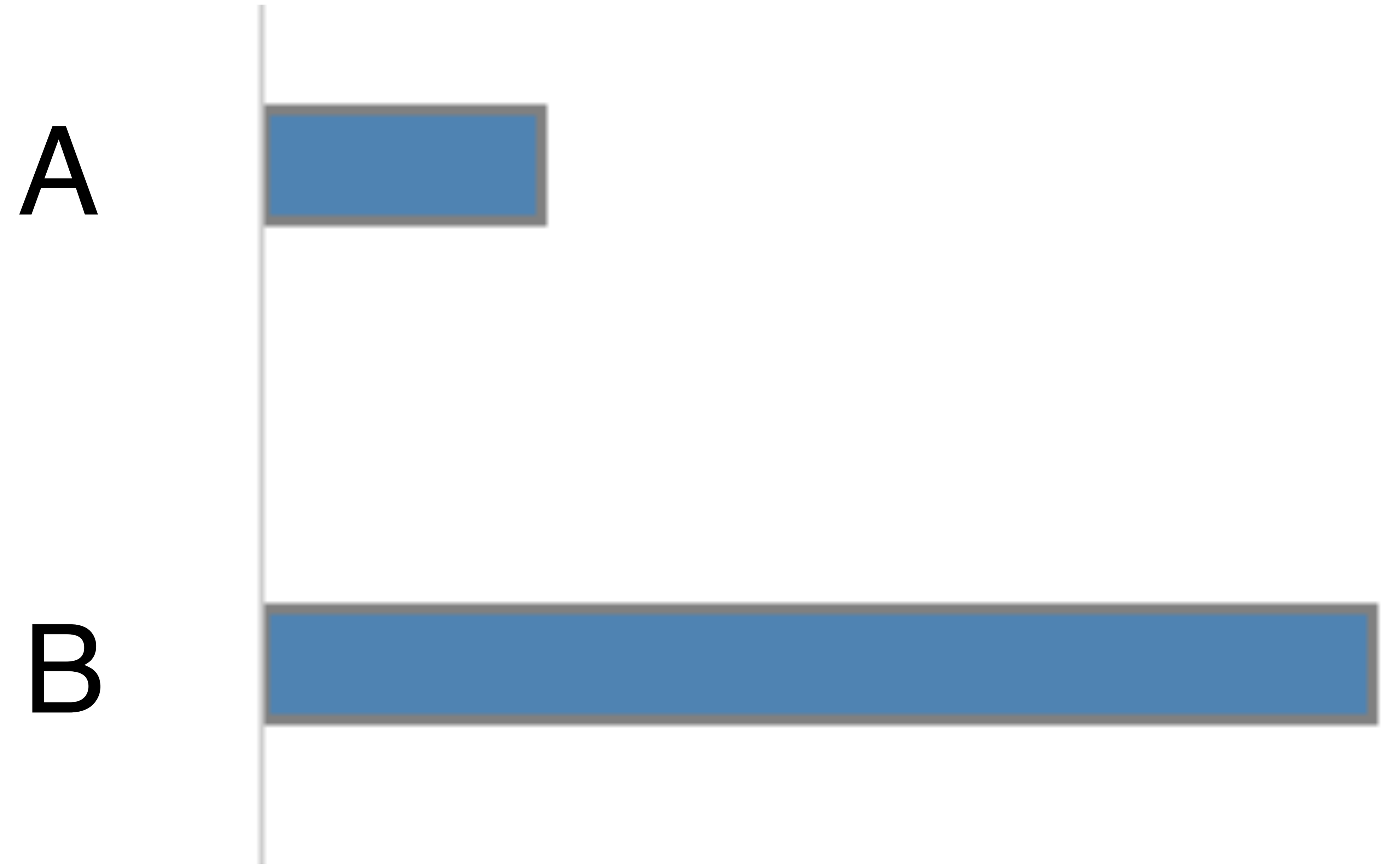
S = sensation
I = intensity

How much longer?



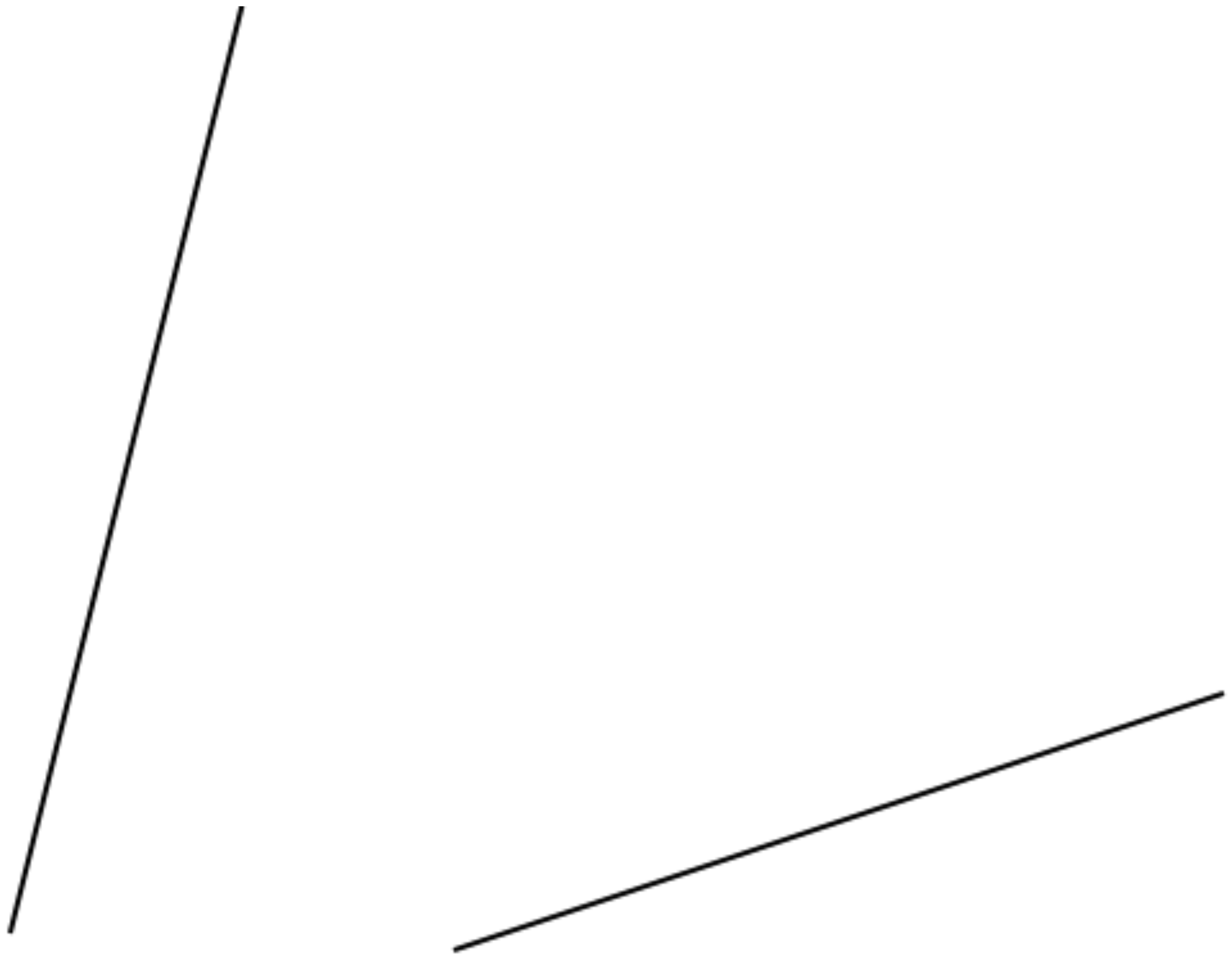
2x

How much longer?



4x

How much steeper?

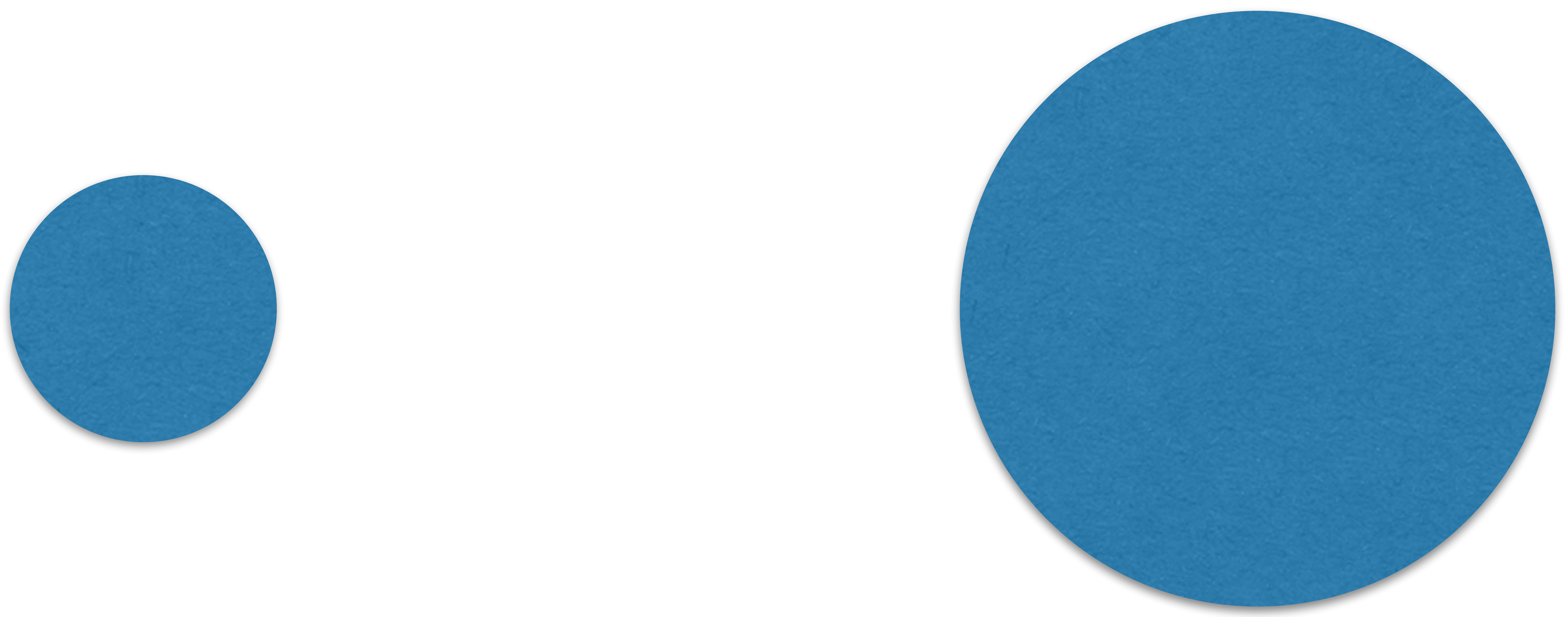


A

B

~4x

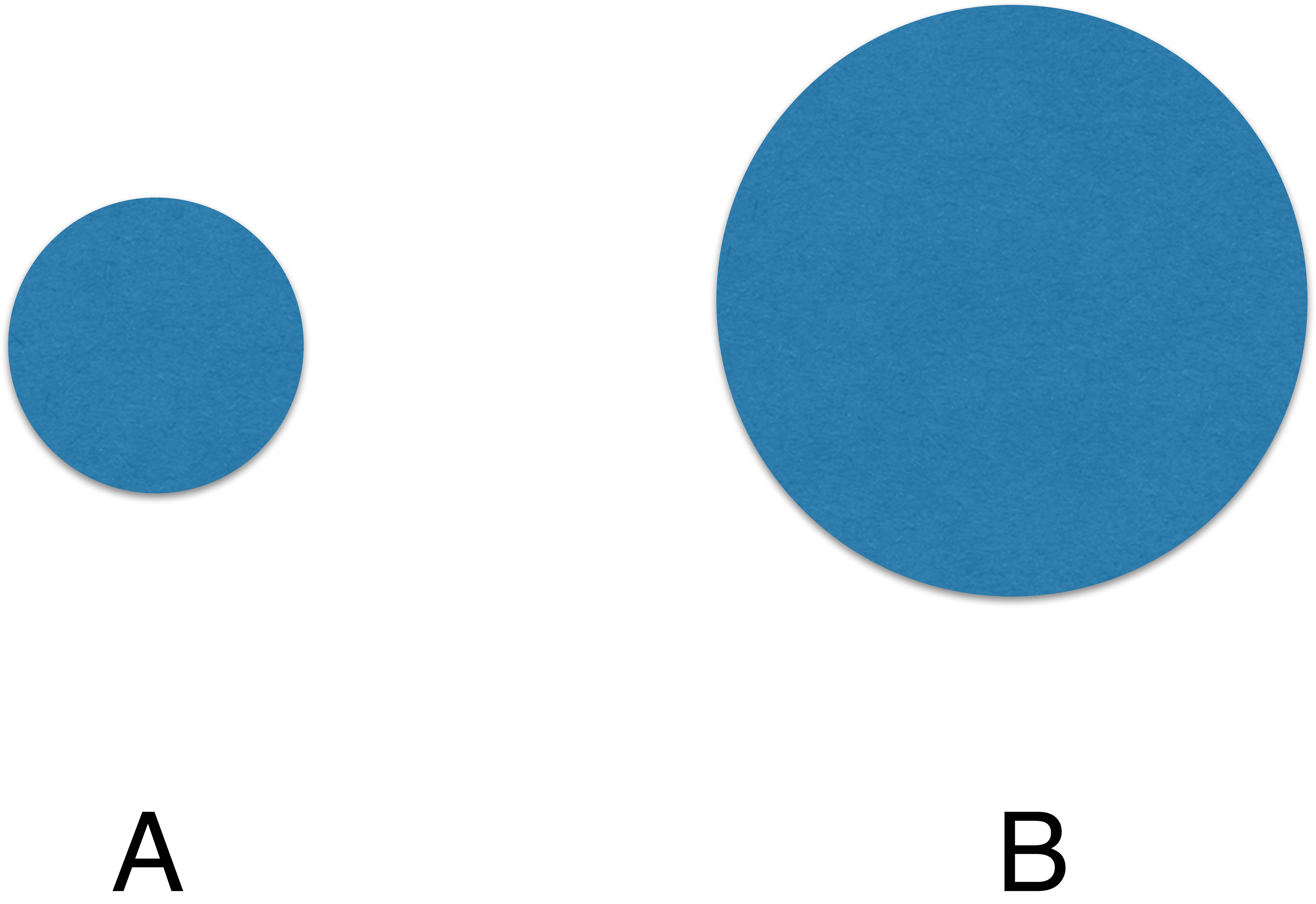
How much larger?



A B

5x

How much larger?

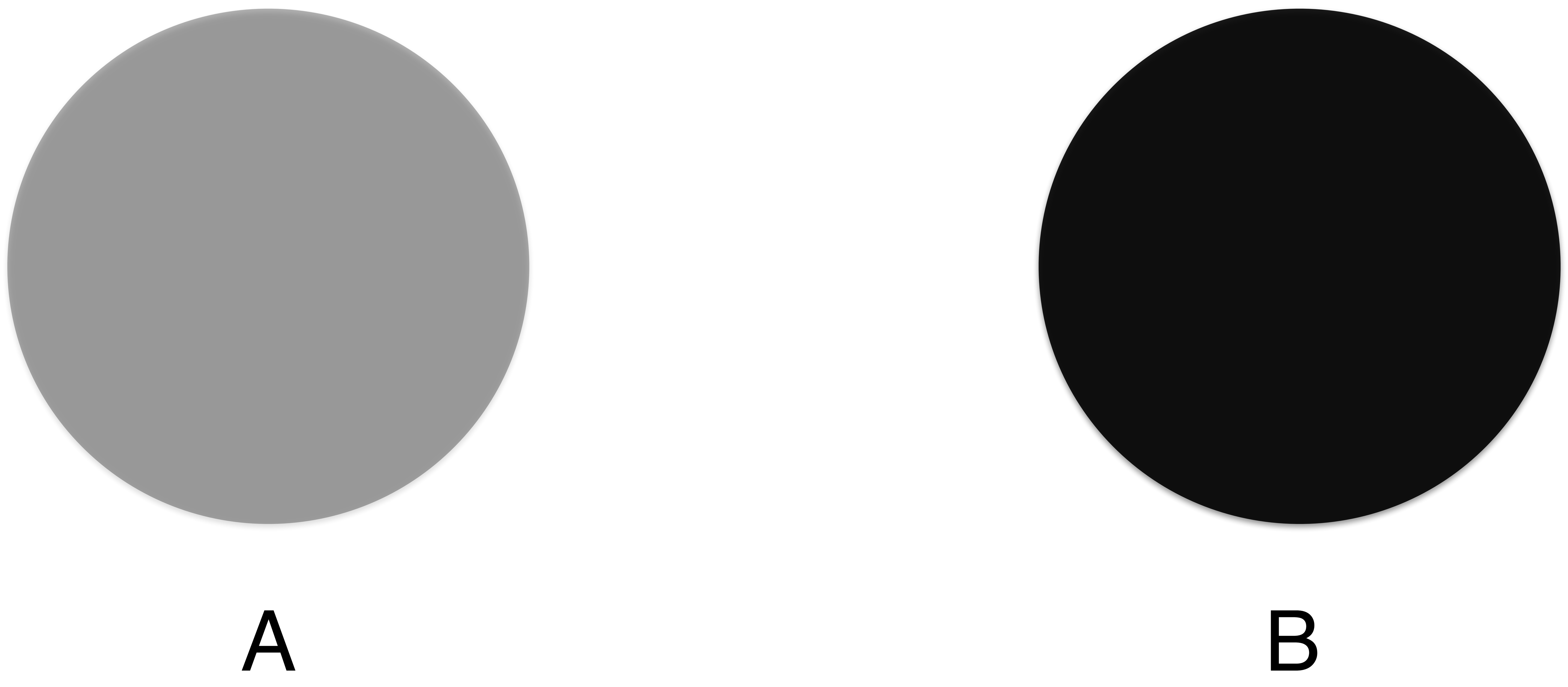


2x diameter

4x area

area is proportional to diameter squared

How much darker?



3x

Other Factors Affecting Accuracy

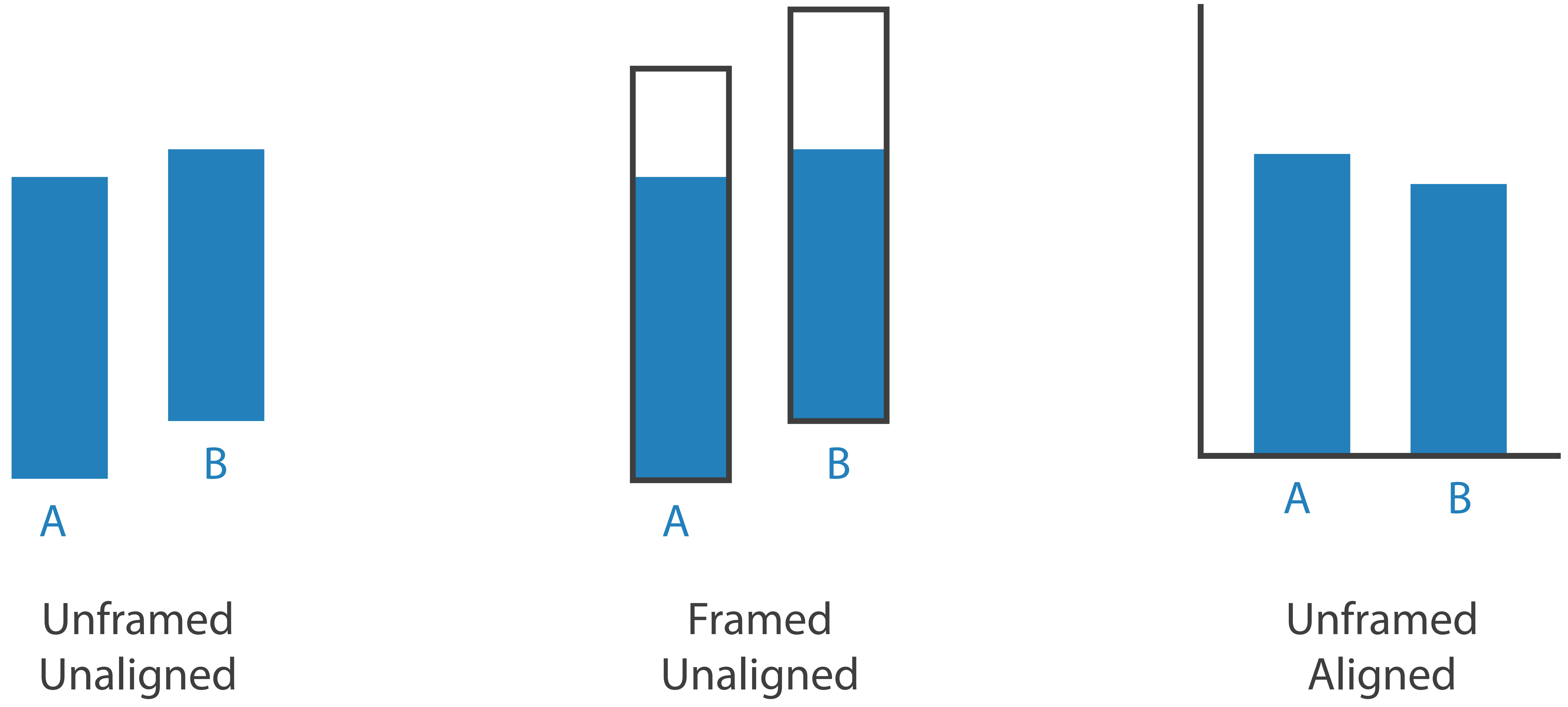
Alignment

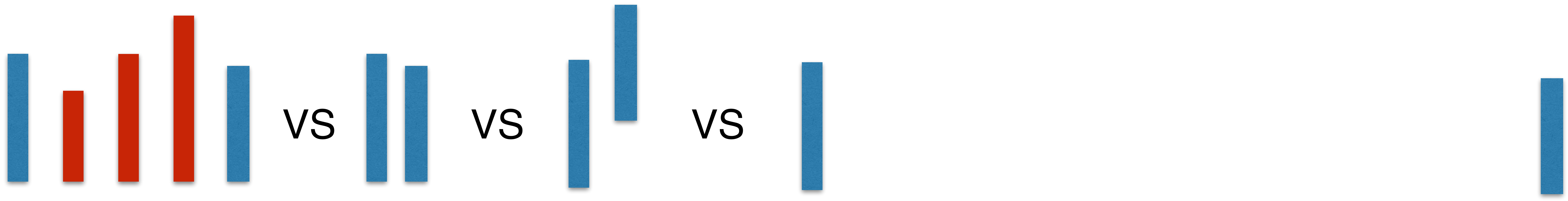
Distractors

Distance

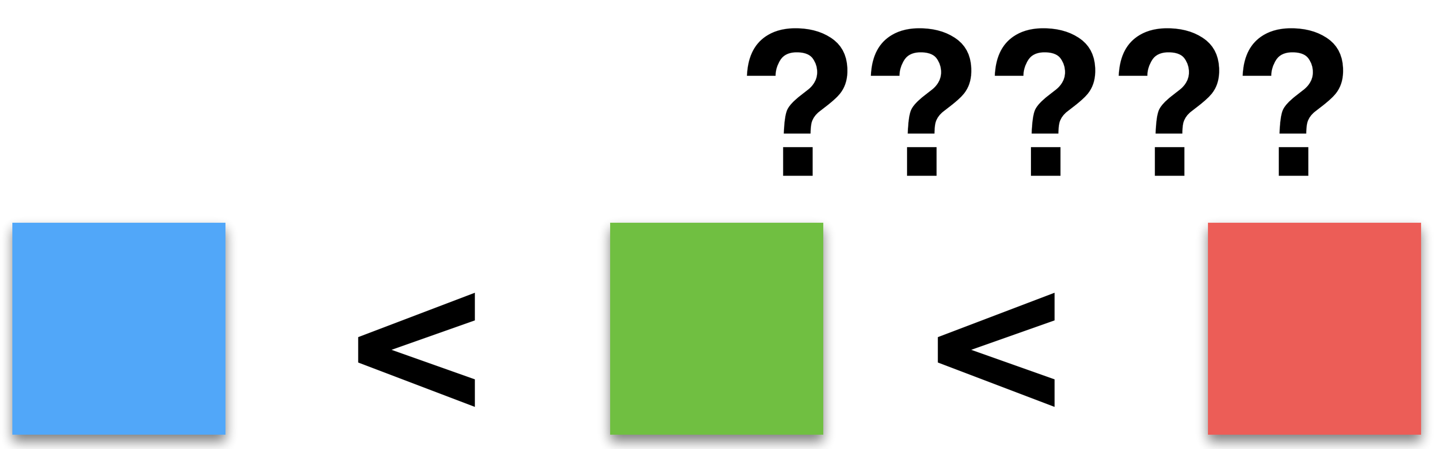
Common scale

…





Color



Good for qualitative data (identity channel)

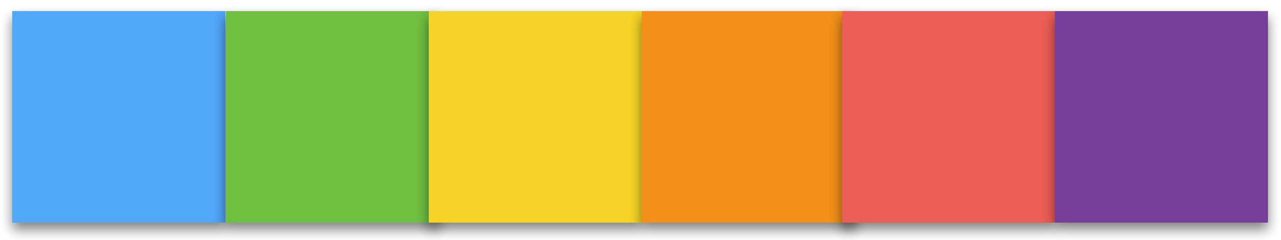
Limited number of classes/length (~7-10!)

Does not work for quantitative data!

Lots of pitfalls! Be careful!

My rule:

- minimize color use for encoding data
- use for highlights



Integrity Principles

1. Show **data variation**, not design variation
2. Clear, detailed, and thorough **labeling** and **appropriate scales**
3. Size of the **graphic effect** should be **directly proportional to the numerical quantities** (“lie factor”)

What’s wrong?

Viele Bezieher mit "ungeklärter Staatsbürgerschaft"

Die größte Gruppe in der Liste der Mindestsicherungsbezieher ist aber jene der "ungeklärten Staatsbürgerschaft". Dass es sich bei den 16.712 Personen um

What’s wrong?

Viele Bezieher mit "ungeklärter Staatsbürgerschaft"

Die größte Gruppe in der Liste der Mindestsicherungsbezieher ist aber jene der "ungeklärten Staatsbürgerschaft". Dass es sich bei den 16.712 Personen um

Viele Bezieher mit "ungeklärter Staatsbürgerschaft"

Die größte Gruppe in der Liste der Mindestsicherungsbezieher ist aber jene der "ungeklärten Staatsbürgerschaft". Dass es sich bei den 16.712 Personen um

What’s wrong?

Grafik der Kronen-zeitung


Zusätzlich geht die Mindestsicherung in Wien auch an 1314 Deutsche, 369 Italiener, 66 Schweden, 59 Schweizer, zehn Kanadier, dazu an einen Liechtensteiner, einen Isländer sowie an einen Bürger von Andorra.

Viele Bezieher mit "ungeklärter Staatsbürgerschaft"

Die größte Gruppe in der Liste der Mindestsicherungsbezieher ist aber jene der "ungeklärten Staatsbürgerschaft". Dass es sich bei den 16.712 Personen um

Resources

http://dataviscourse.net

Slides

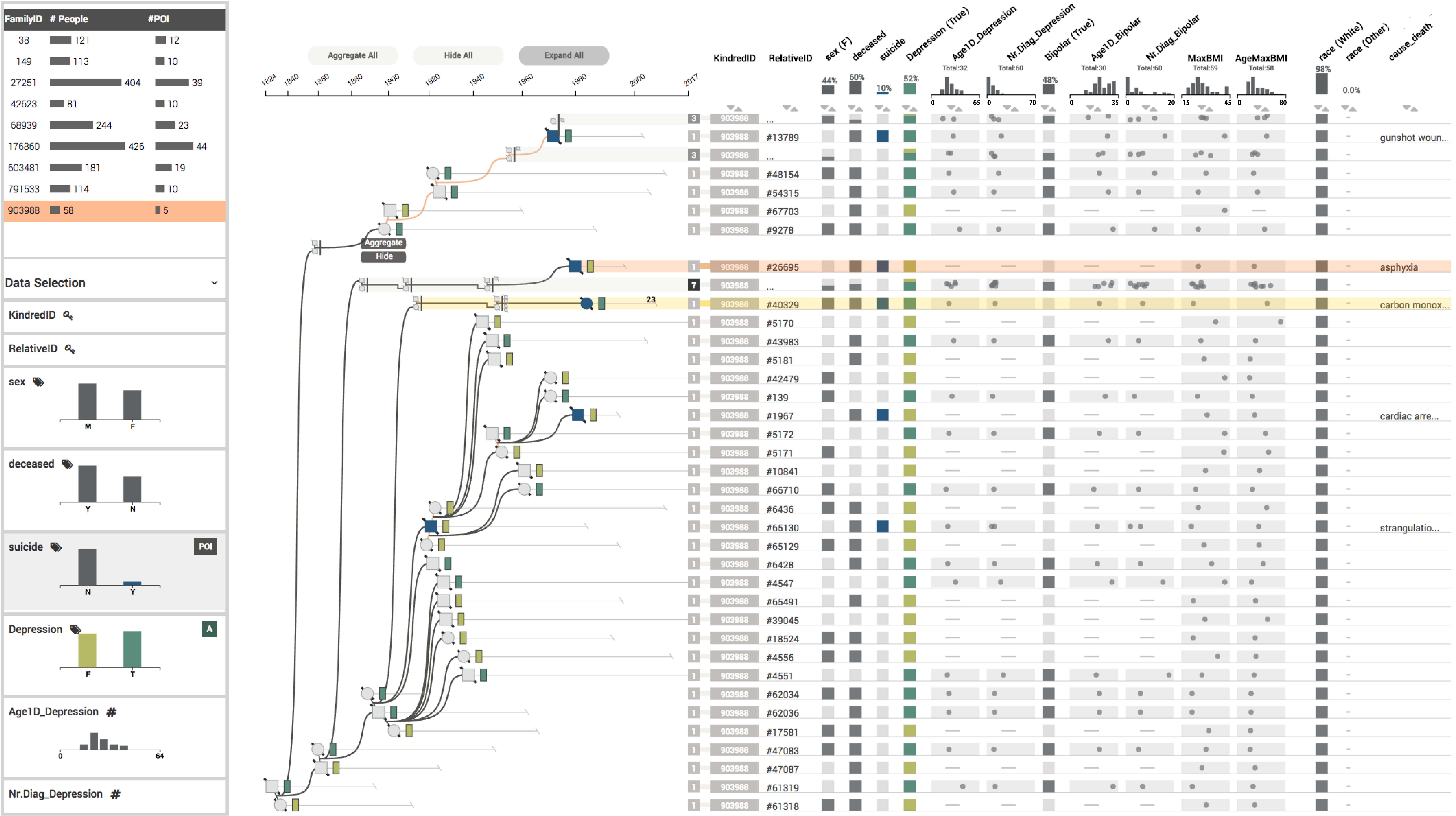
Videos

D3 Tutorials

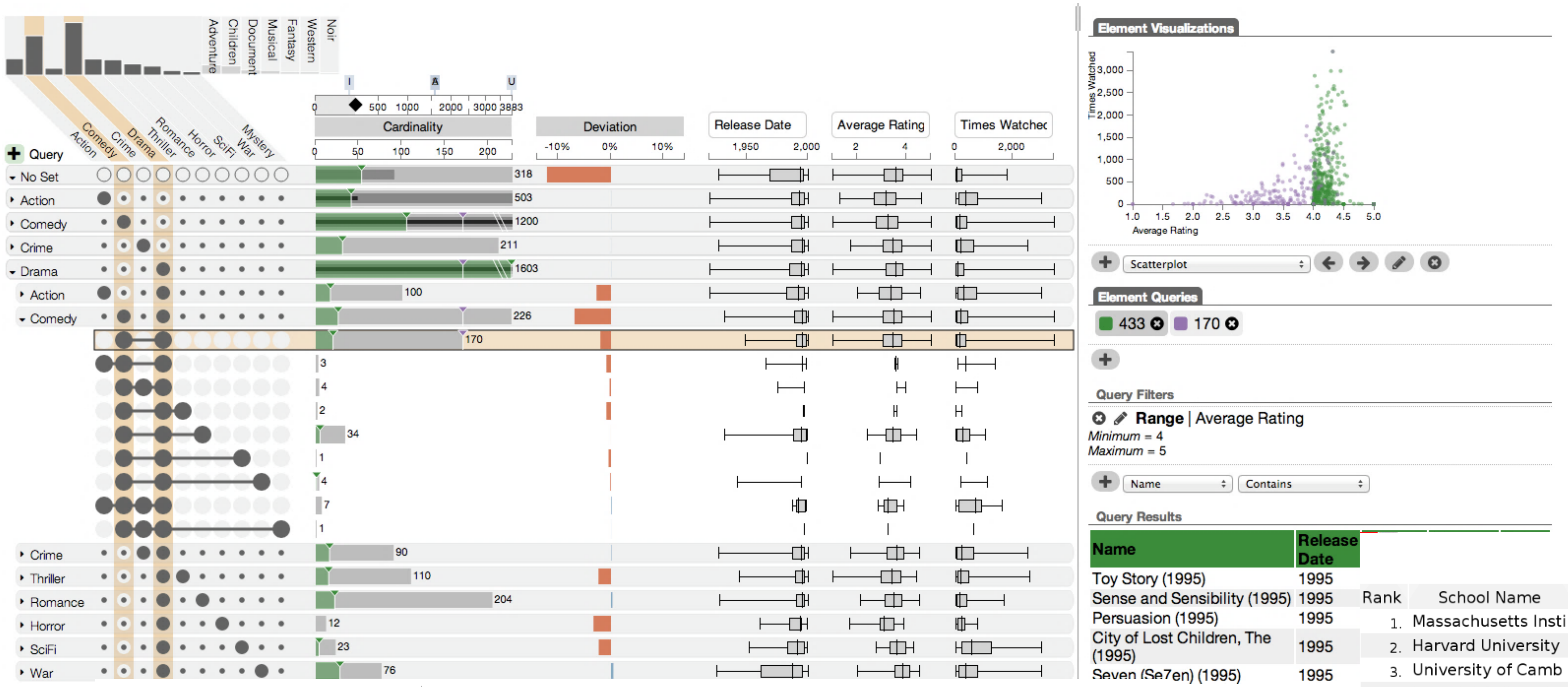
Research Areas

Large, Multivariate (Biological) Networks

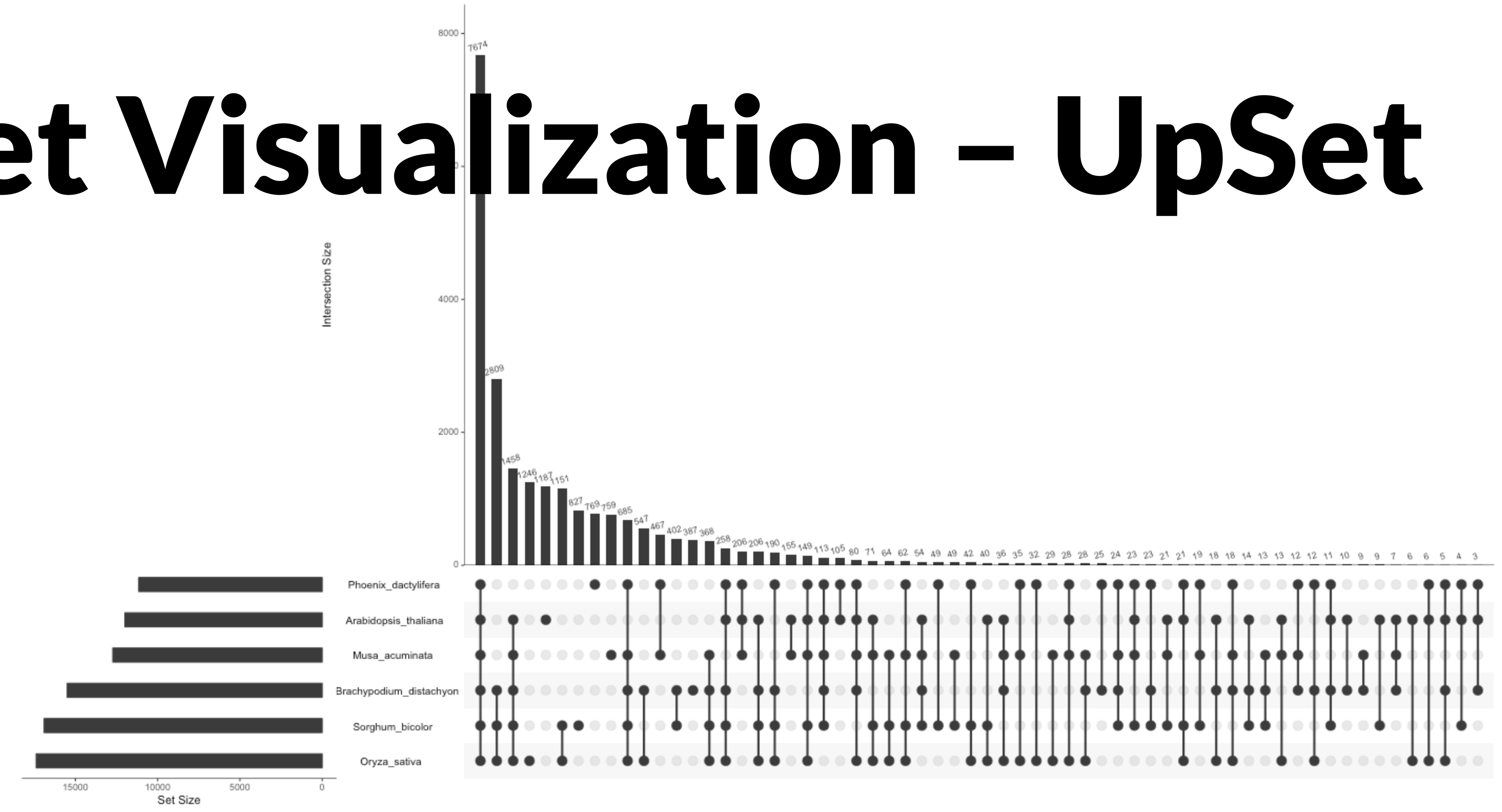
Genealogies & Clinical Data



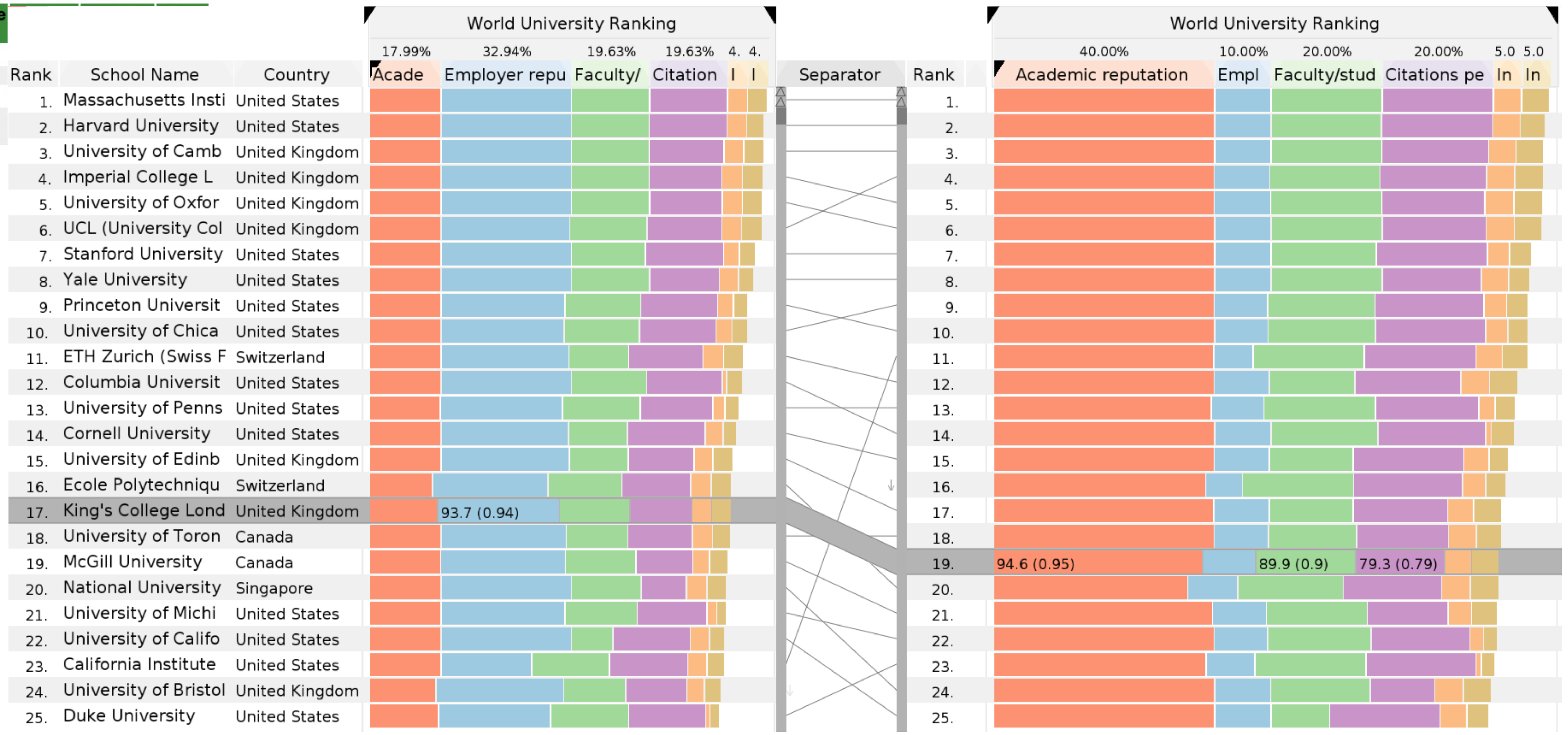
Multidimensional Data



Set Visualization – UpSet



Multivariate Rankings – Lineup

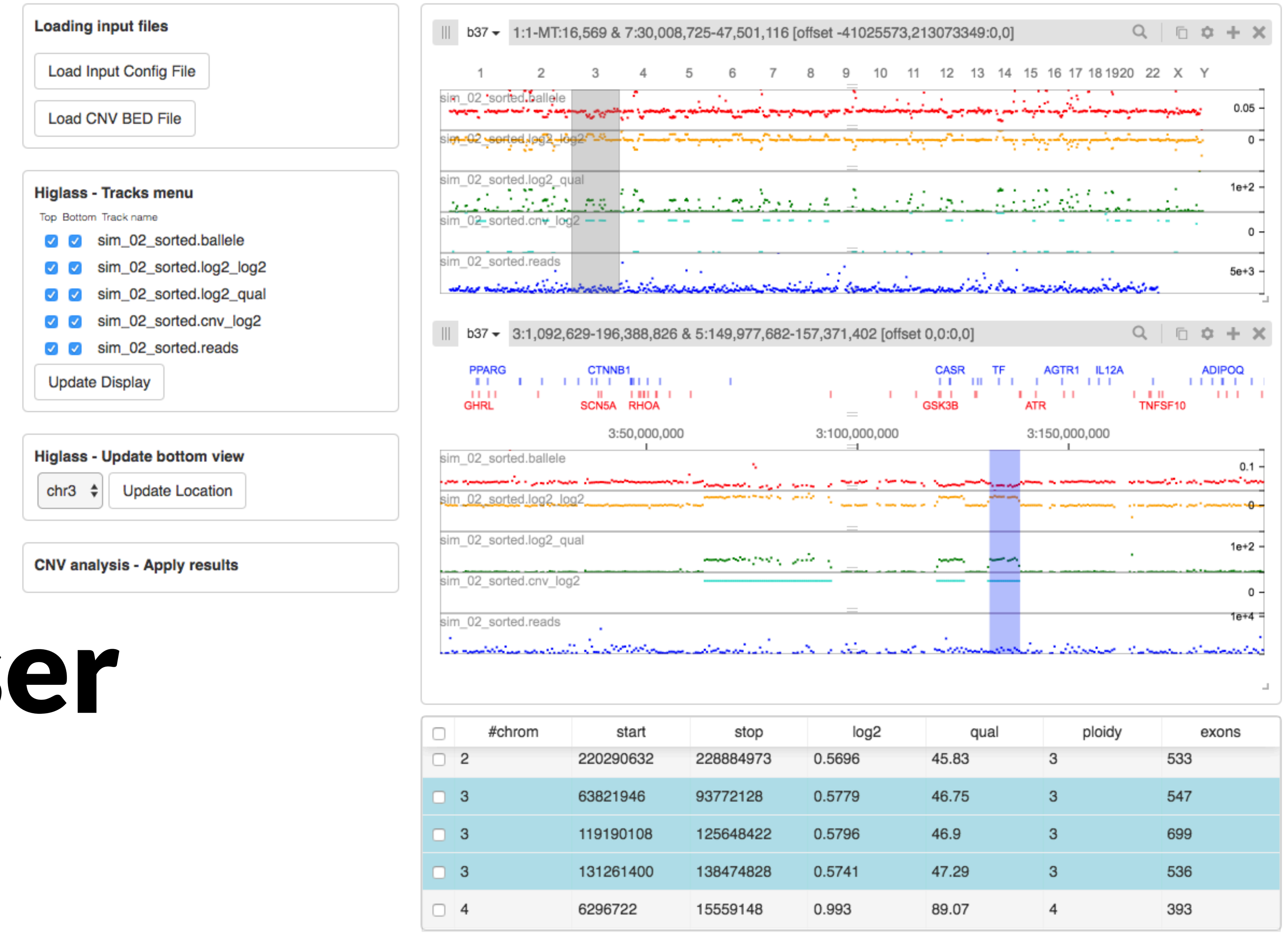


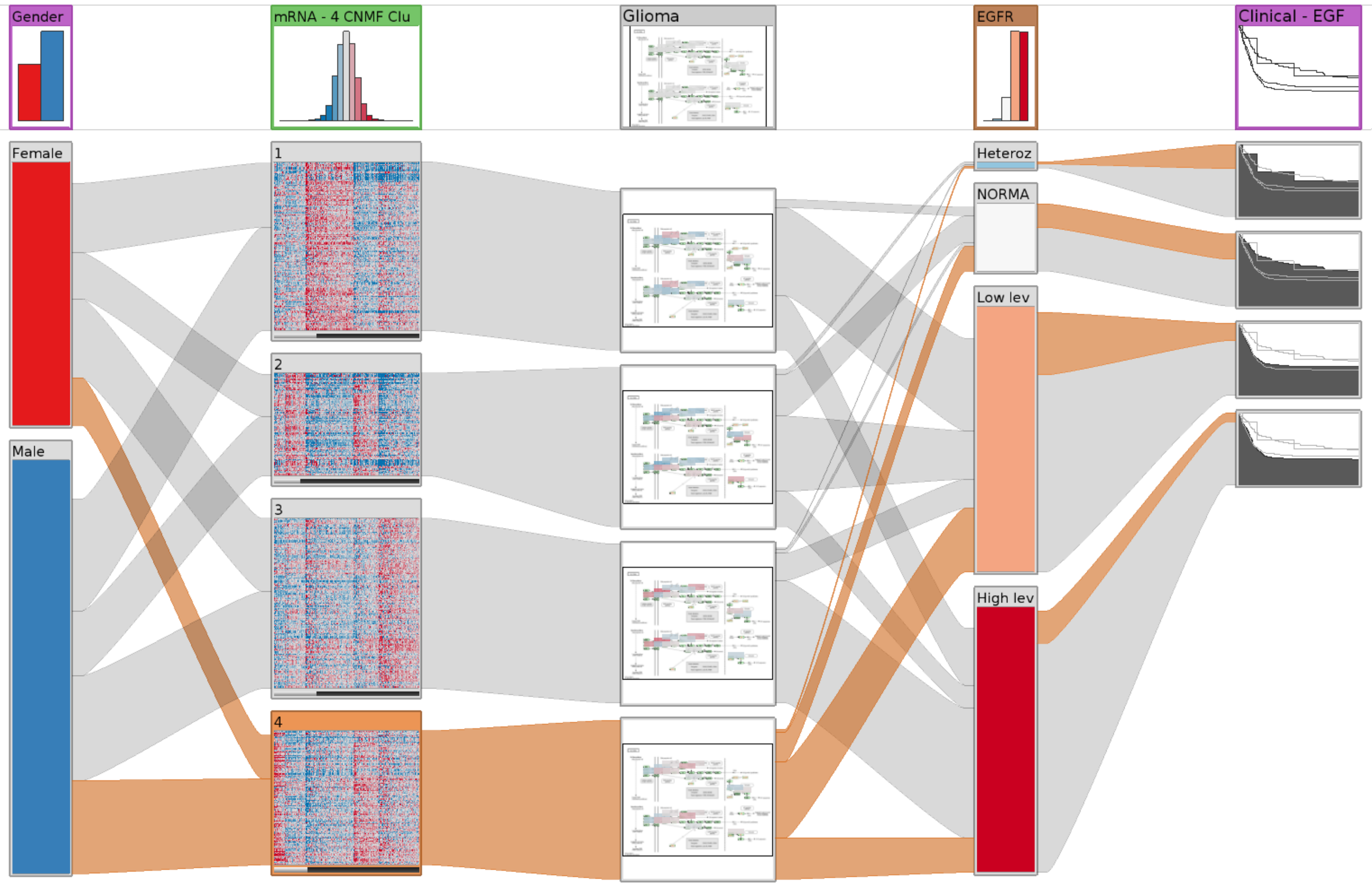
EHRs

COMPOSER
Cohort Control
+ 1
clear all
Cohort-1
Cohort-2
C-2 Branch-1
Cohort-3
Applied Filters
C-2 Branch-1 Filter Layers
Cohort Size: 152
1: All Patients 6069
2: injection 152
Demographic Filters
BMI
CCI
Age
Alcohol User
Drug User
Layer Control
Layer-1
Layer-2
Layer-3
Layer-4
injection
Relative Scale
Aggregate Scores
Separate Quantiles
Layer View
Cohort-3
injection
surgery
PROMIS Bank v1.2 - Physical Function

Genomic Data

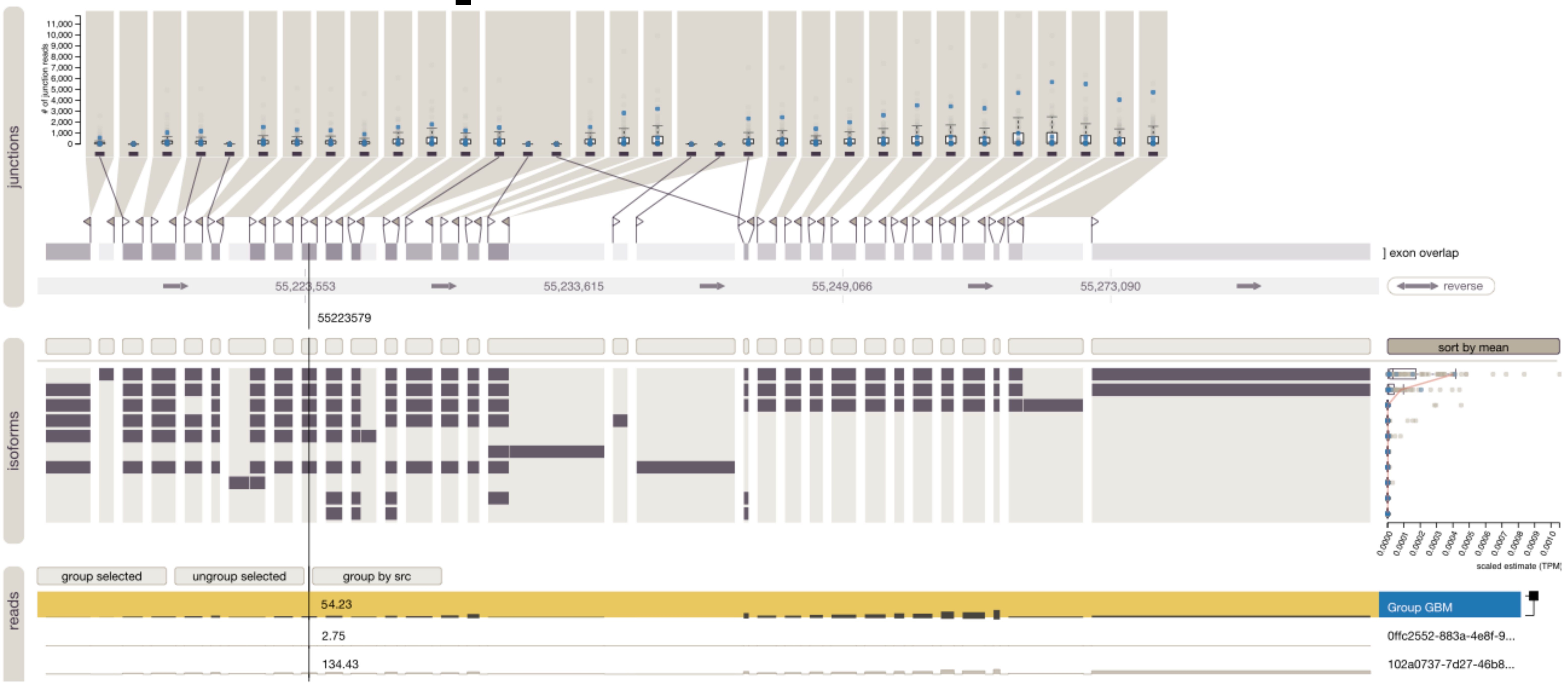
cTracks – Copy Number Browser



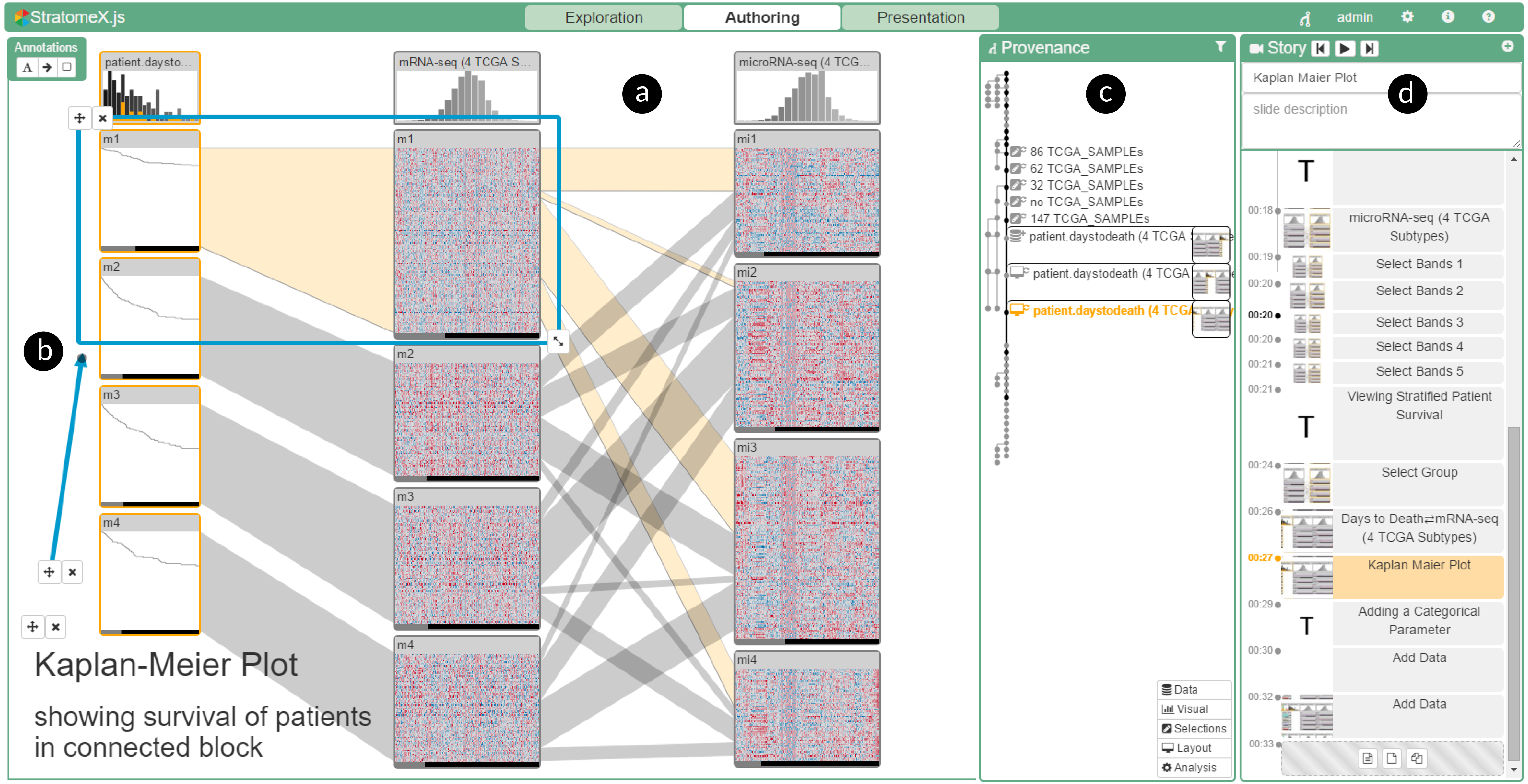


Cancer Subtypes / Omics Clustering and Stratification

Alternative Splicing / mRNA-seq



Reproducibility, Storytelling, Annotation, and Integration in Computational Workflows



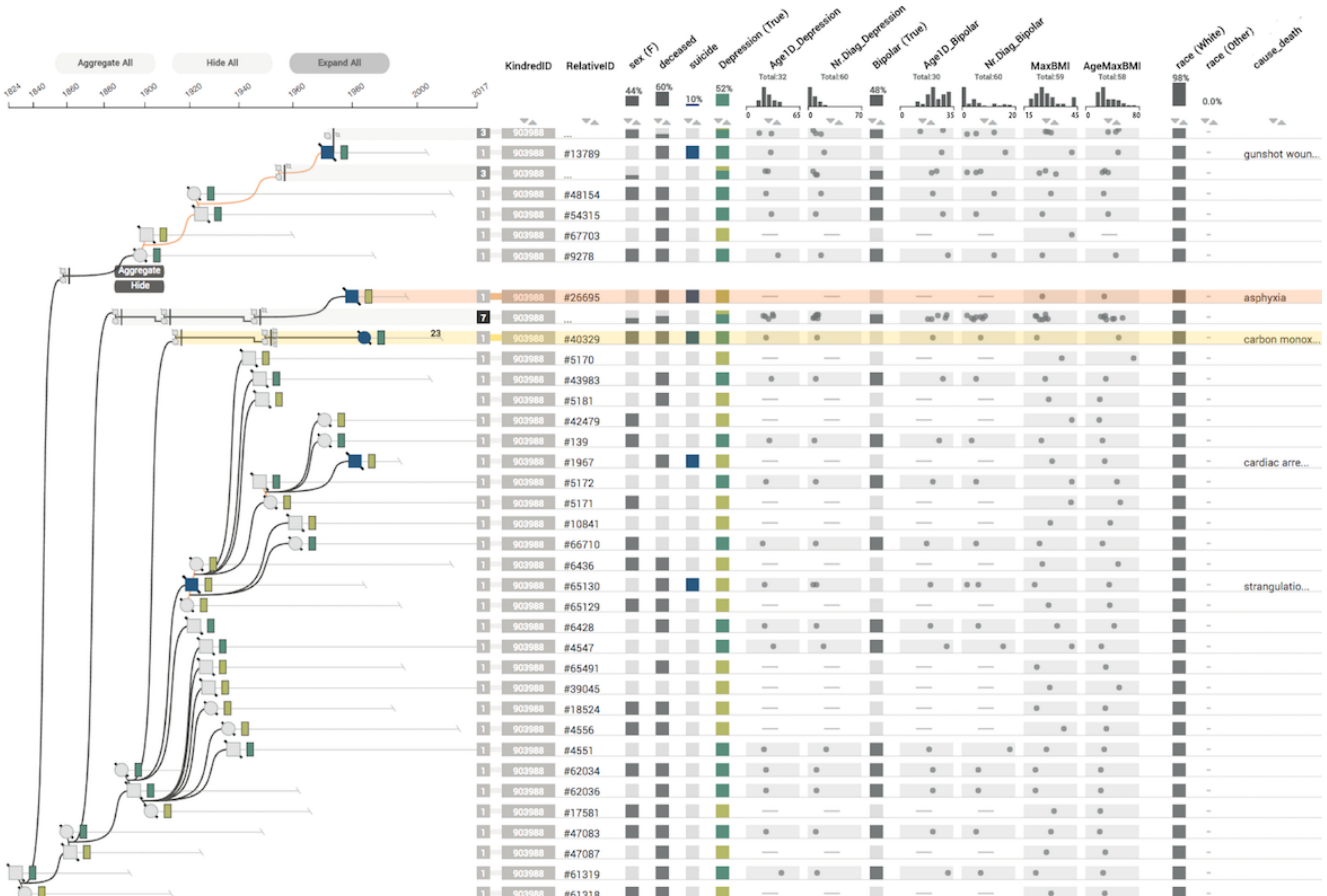
Lineage
Visualizing Clinical Data in Genealogy Graphs

Carolina Nobre

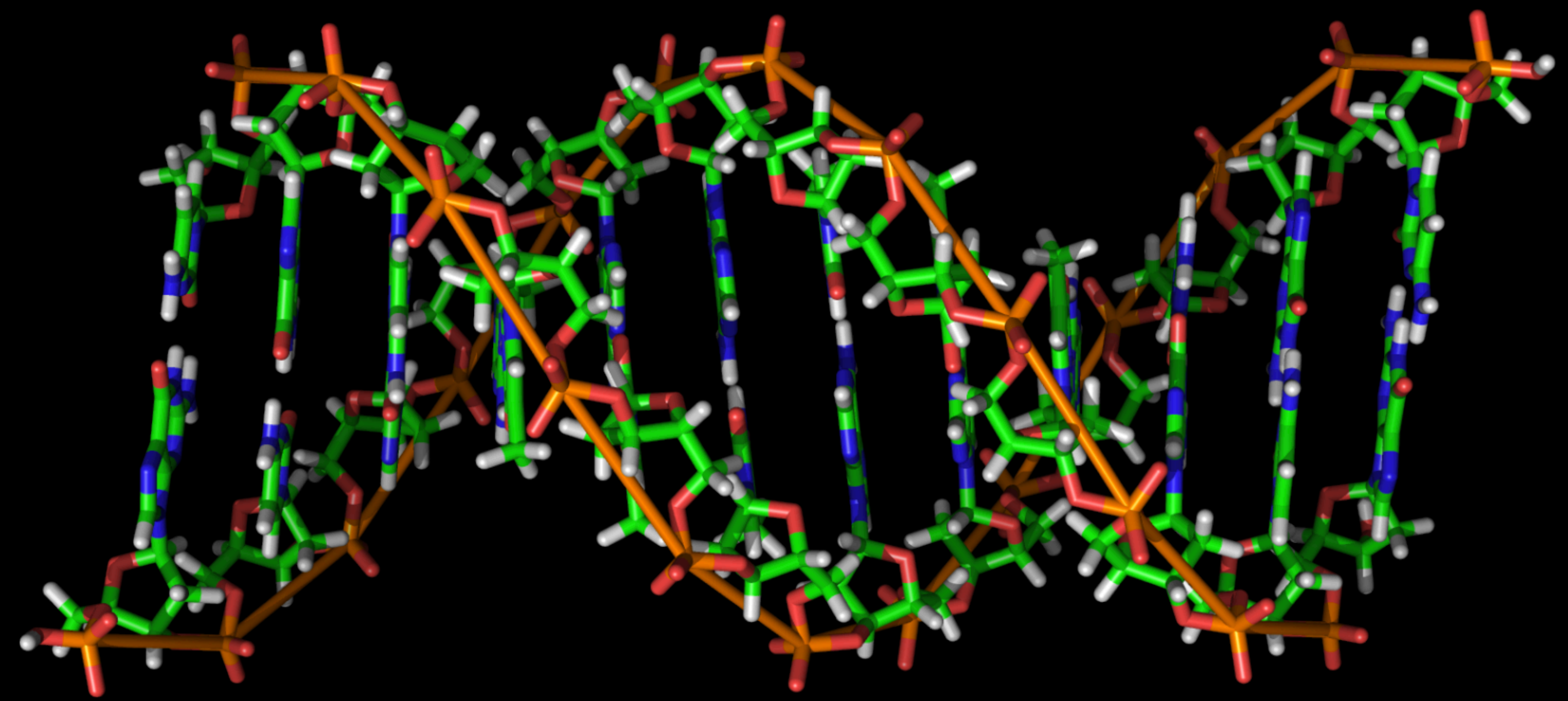
Nils Gehlenborg

Hilary Coon

Alexander Lex



Motivation



Understand Complex Psychiatric Conditions

Discover Genetic Risk Factors

Dataset: 118k people, 19k suicide cases, ~2k with genomic data, 550 families

Based on Utah Population Database

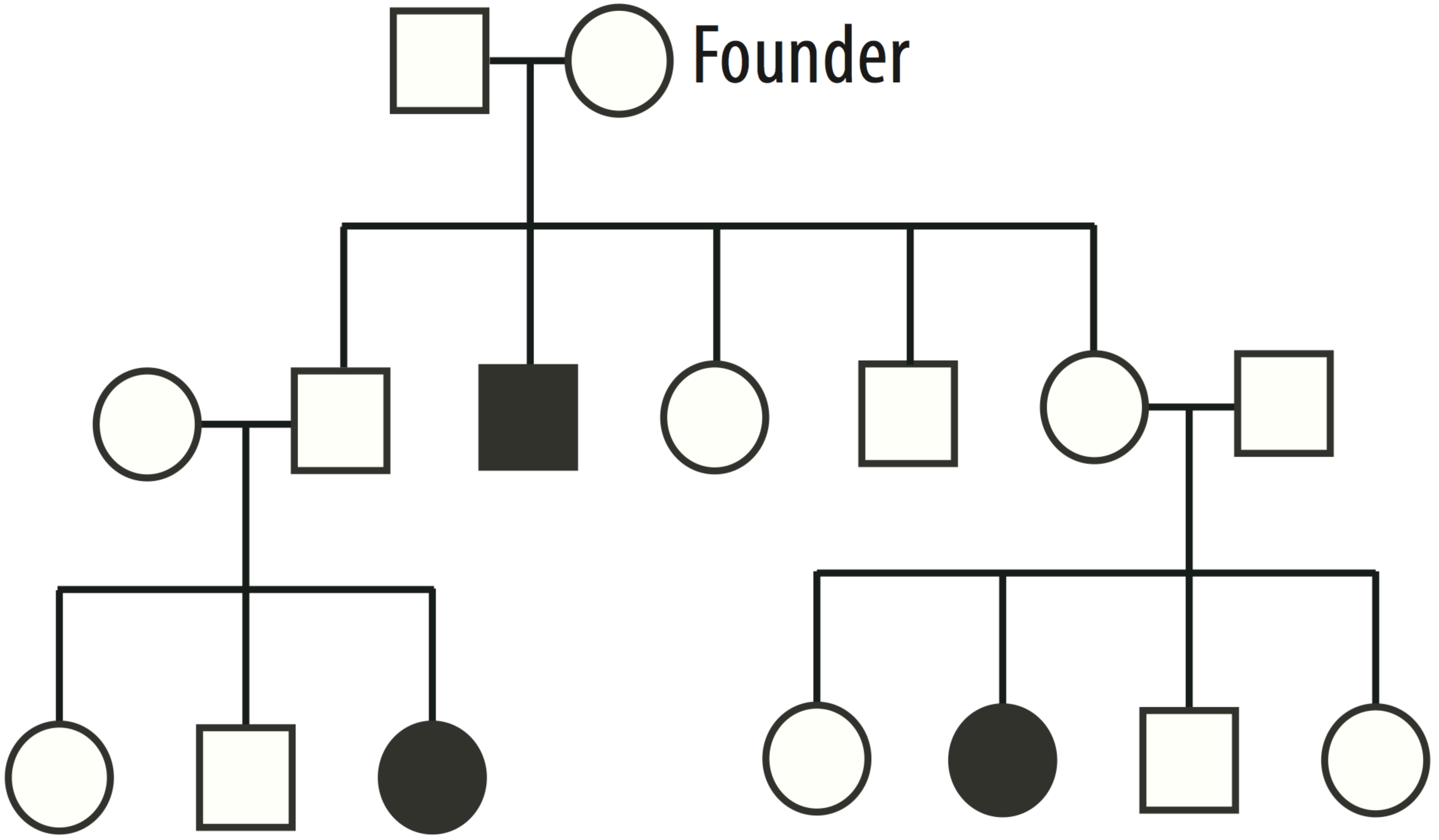
Specific Goals

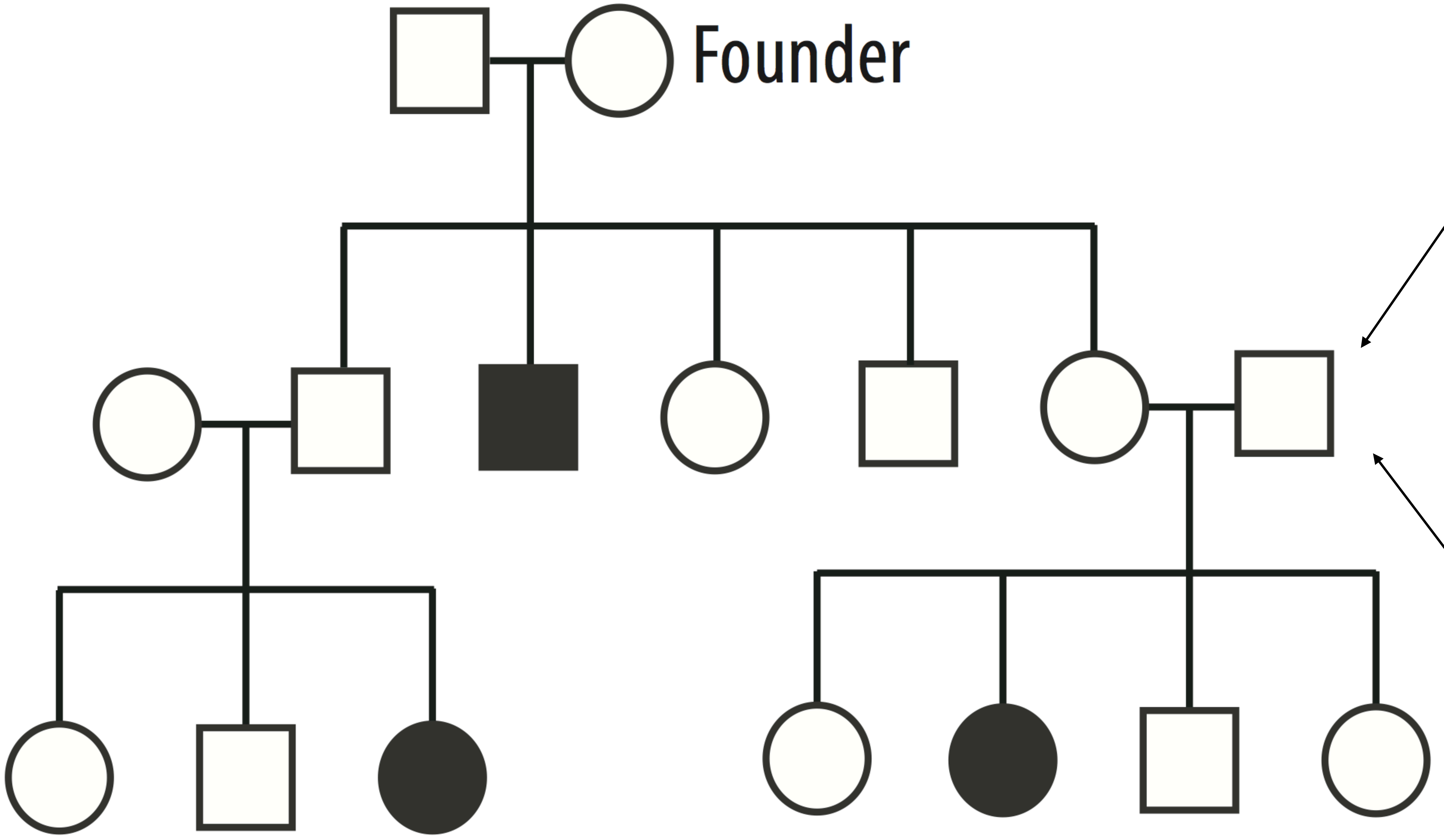
Find familial cases that also have an "interesting" phenotype

- e.g., predominantly female, associated with rare psychiatric disease, etc.

Prioritize those cases for analysis of shared genomic sequences

Proofreading the Data!

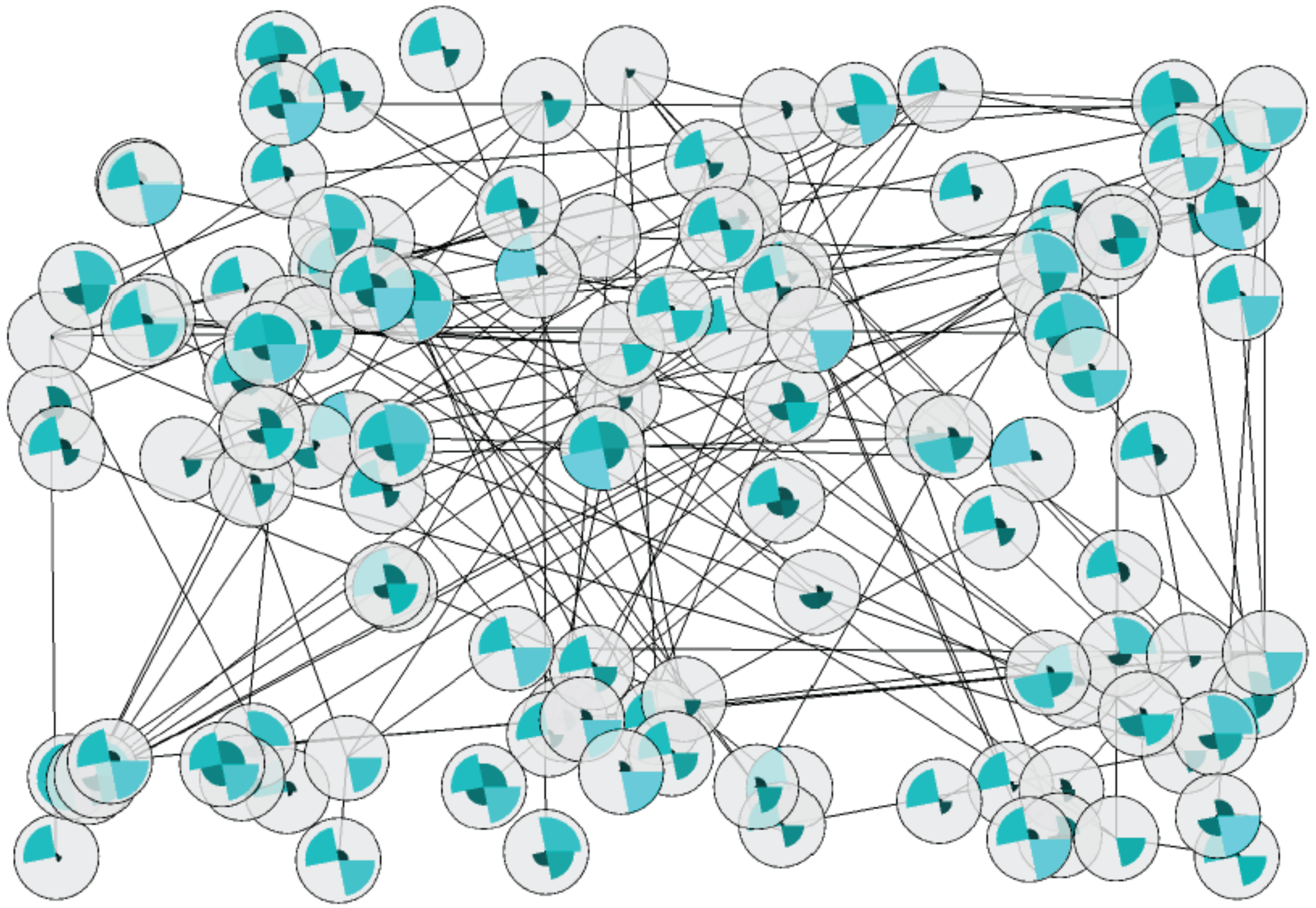

Founder



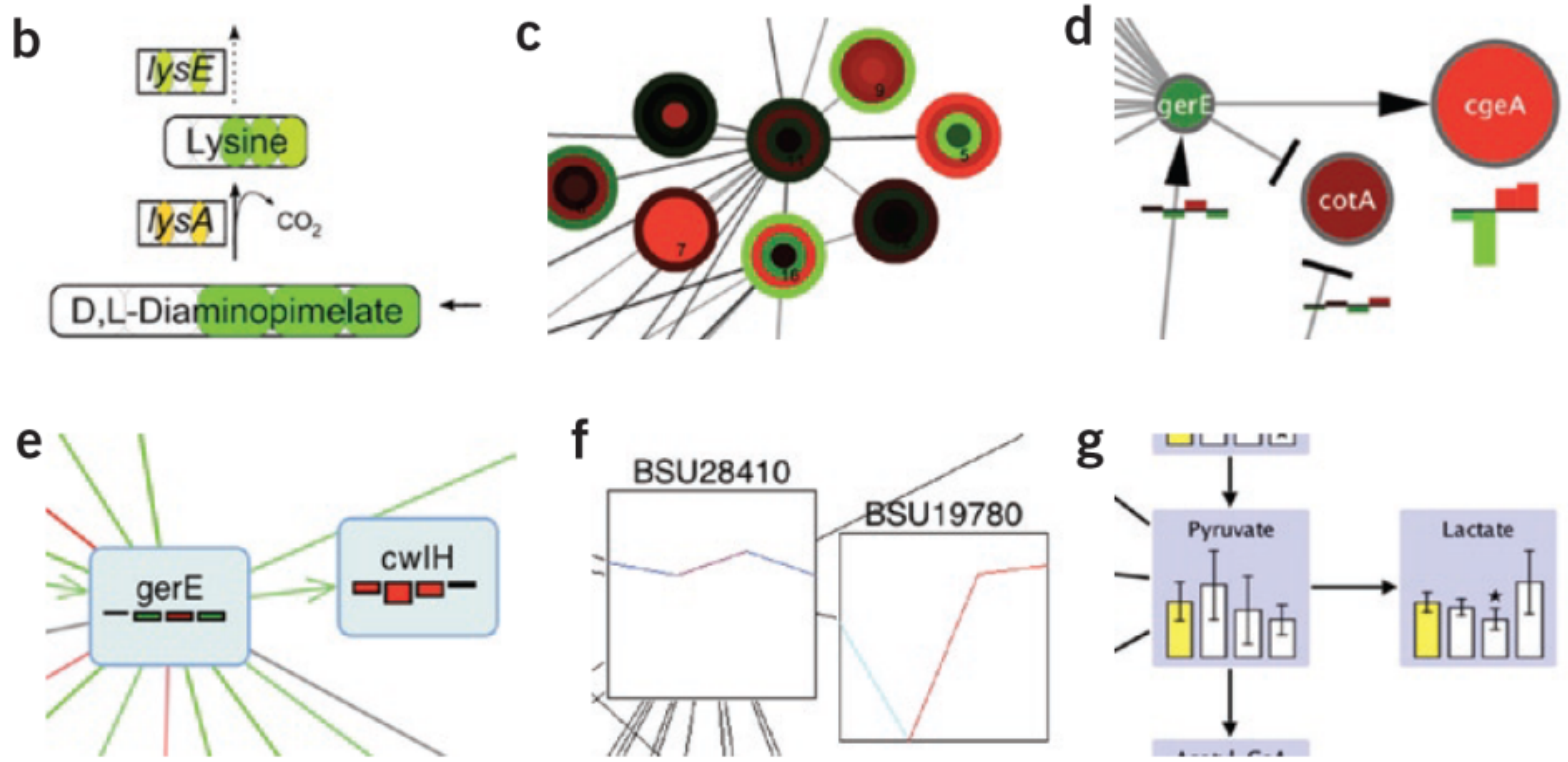
Age
Sex
Race
Bipolar
diagnosis bipolar
Depression
diagnosis depression
Asthma
diagnosis asthma
Obesity
Schizophrenia
Cause of Death
Weapon Used

Multivariate Attributes and Graphs

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

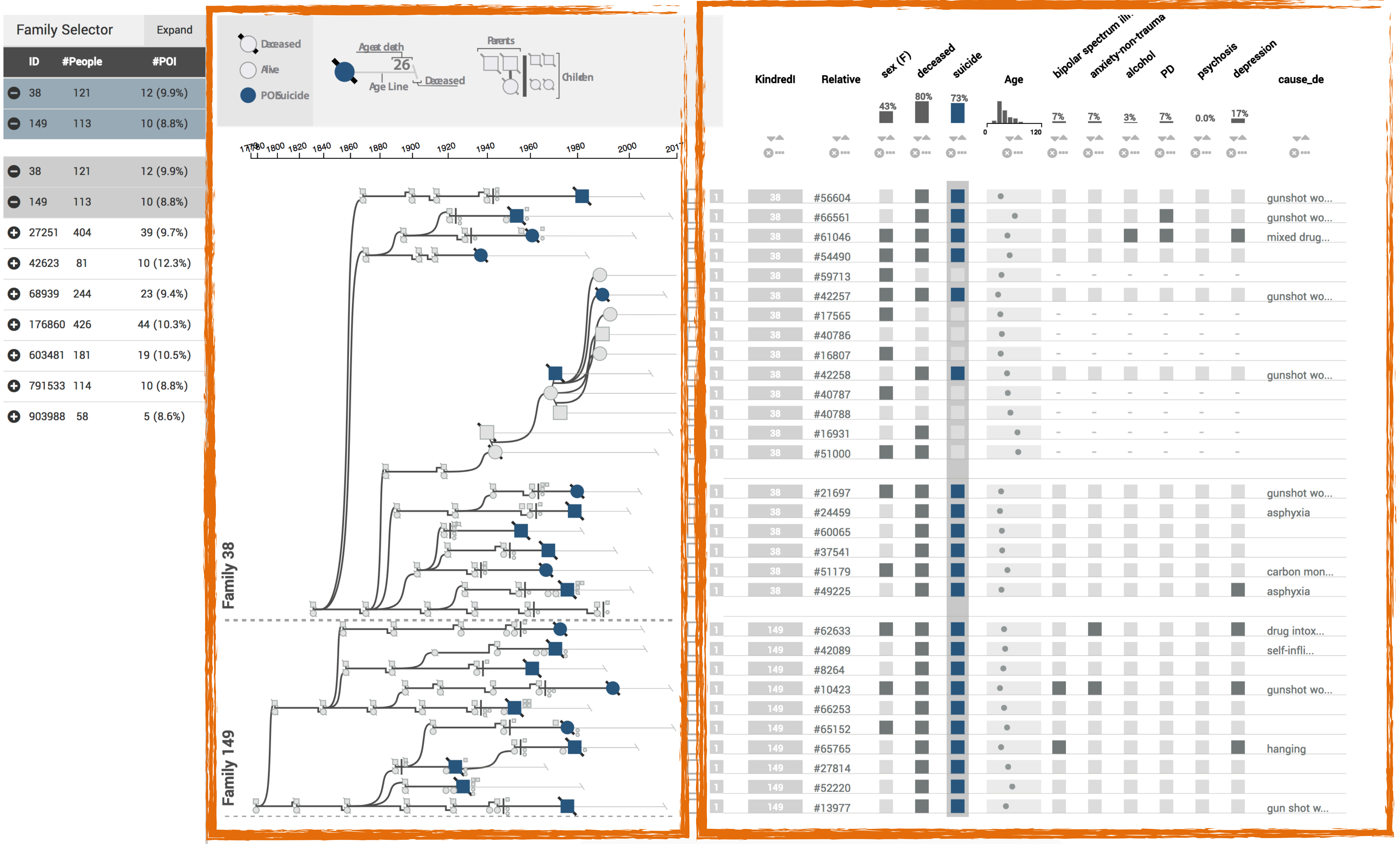


[McDonnel2009]



[Gehlenborg2010]

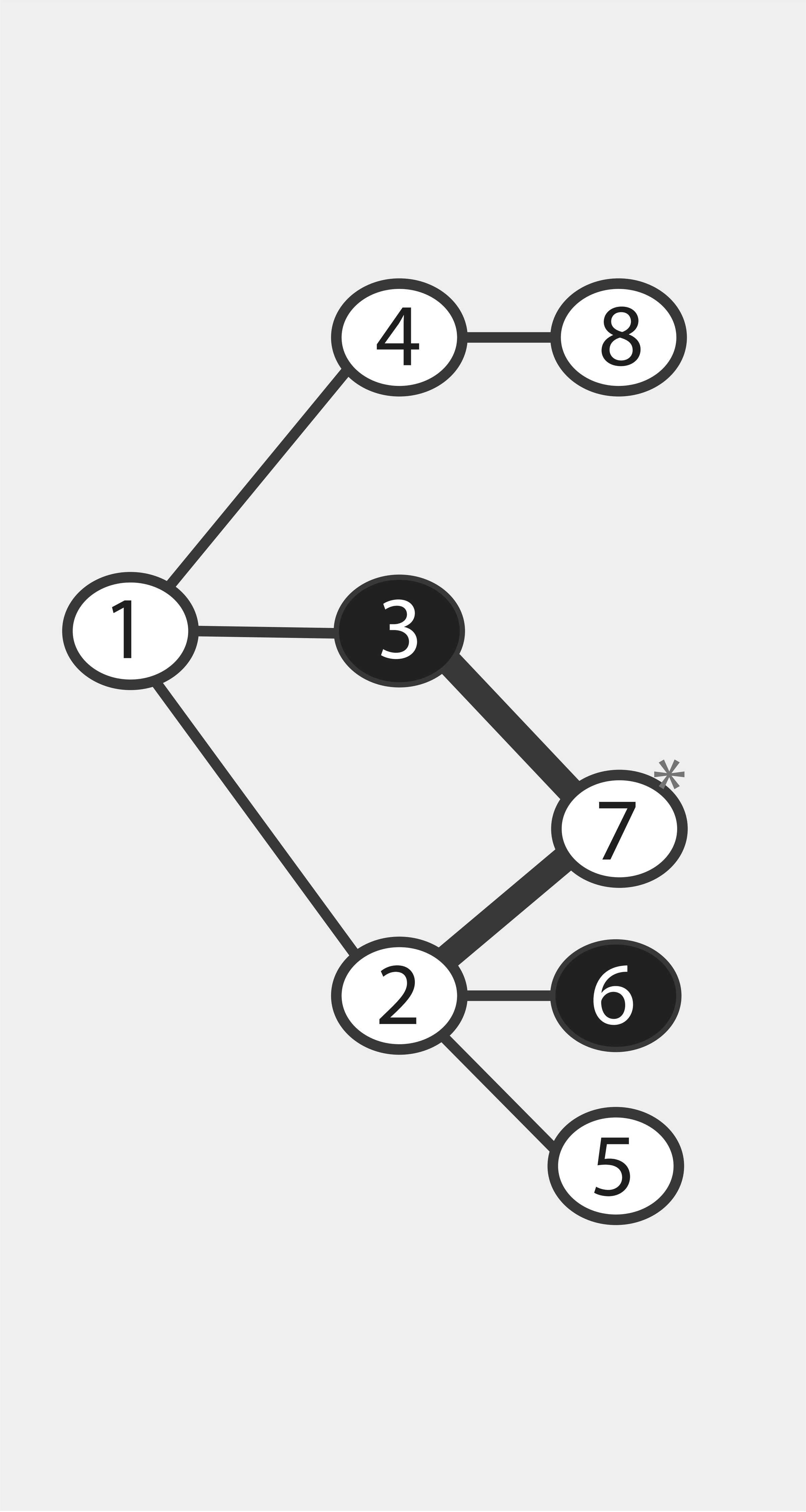
Genealogy with ~400 members rendered with Progeny



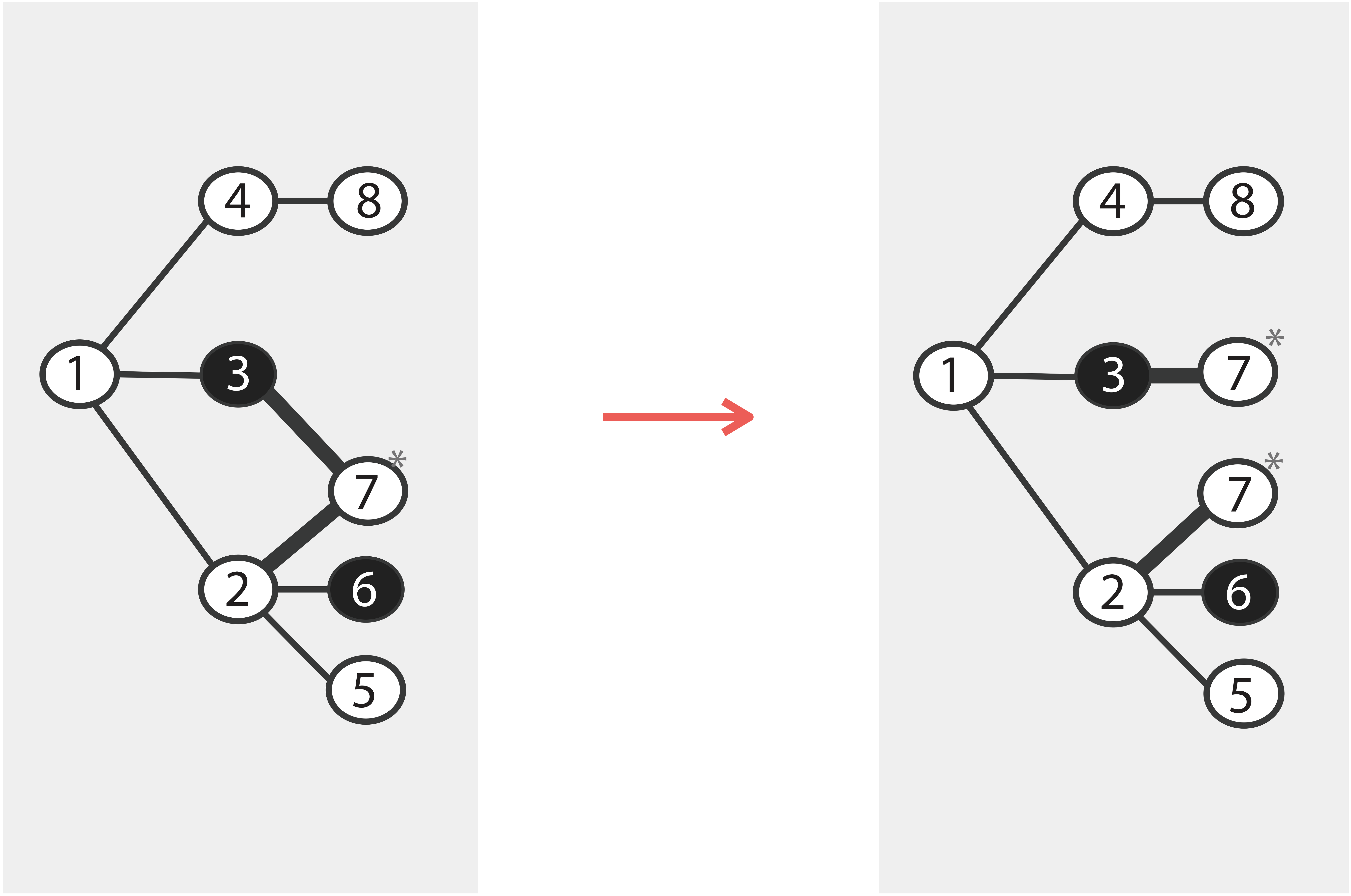
1. De-cycle and linearize graph

2. Plot attributes in table

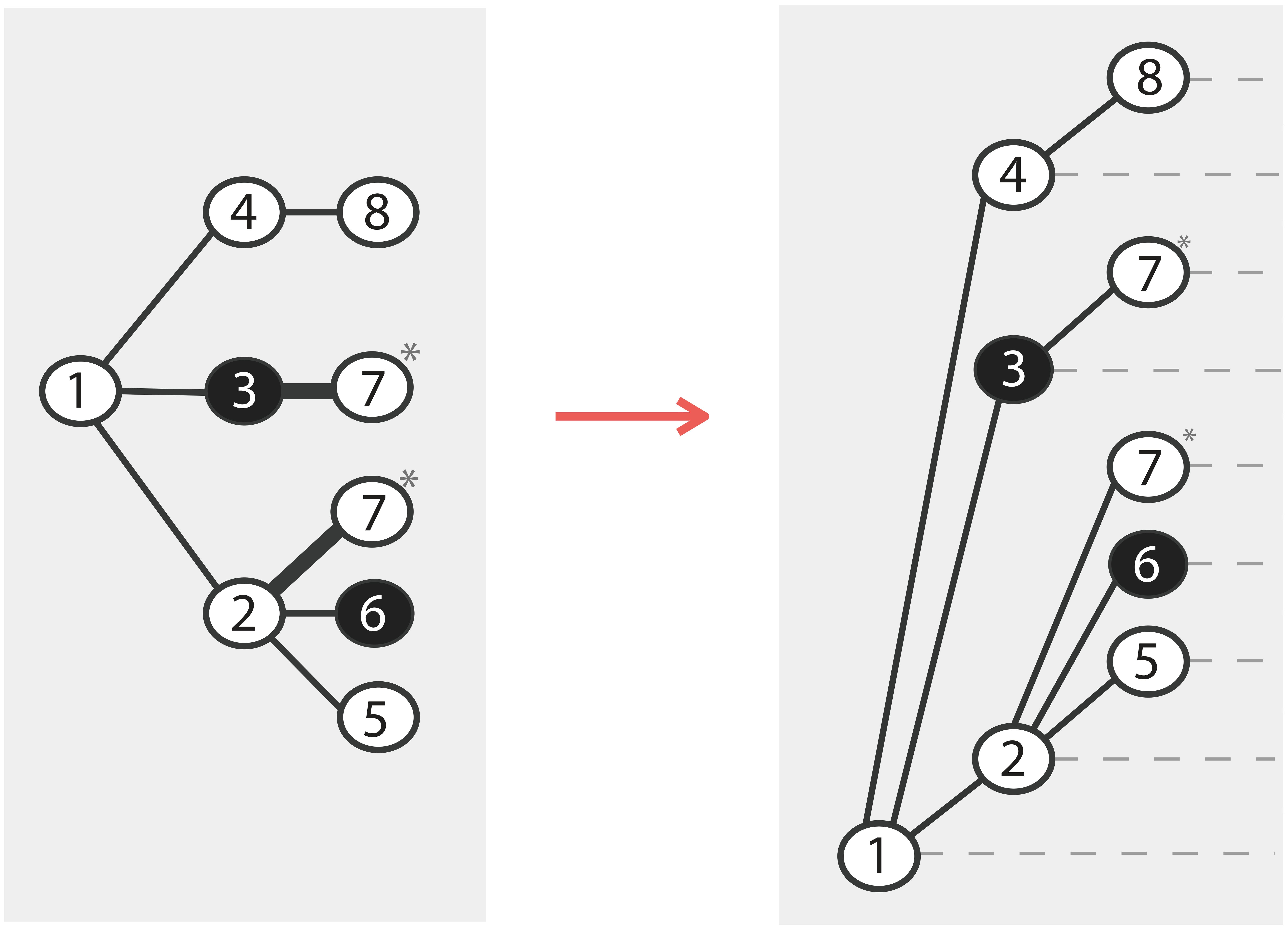
De-Cycling



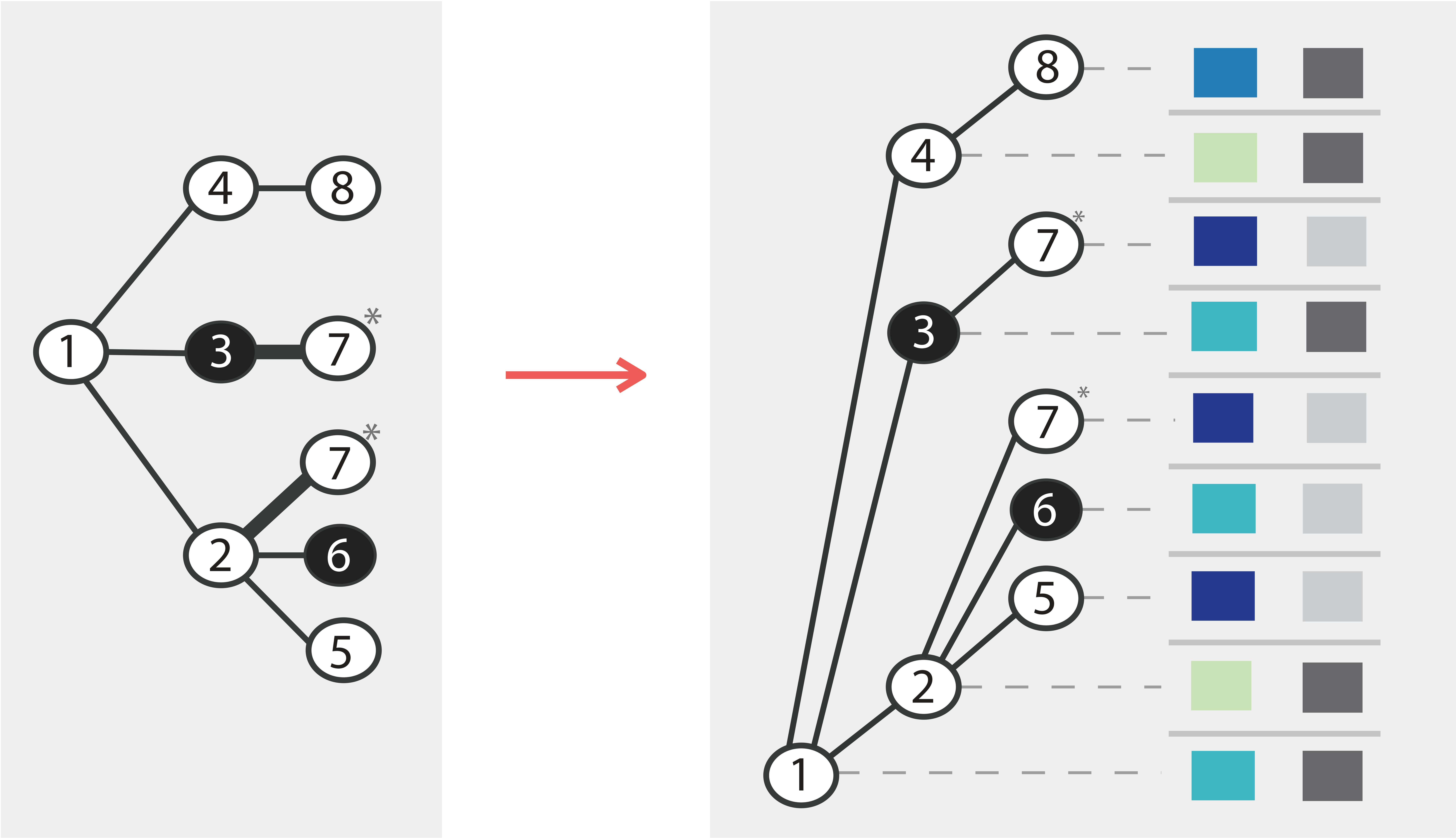
De-Cycling



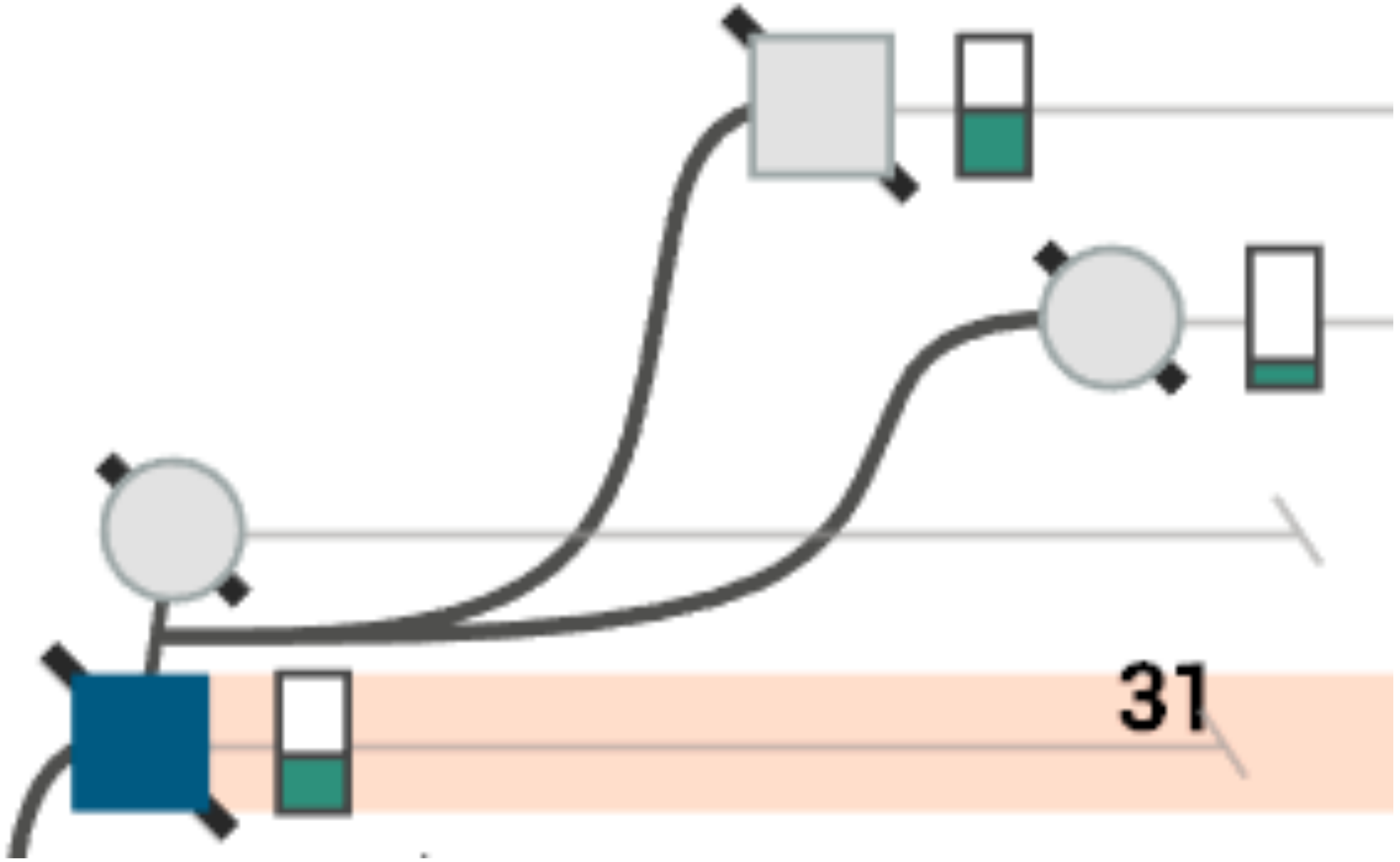
Linearization

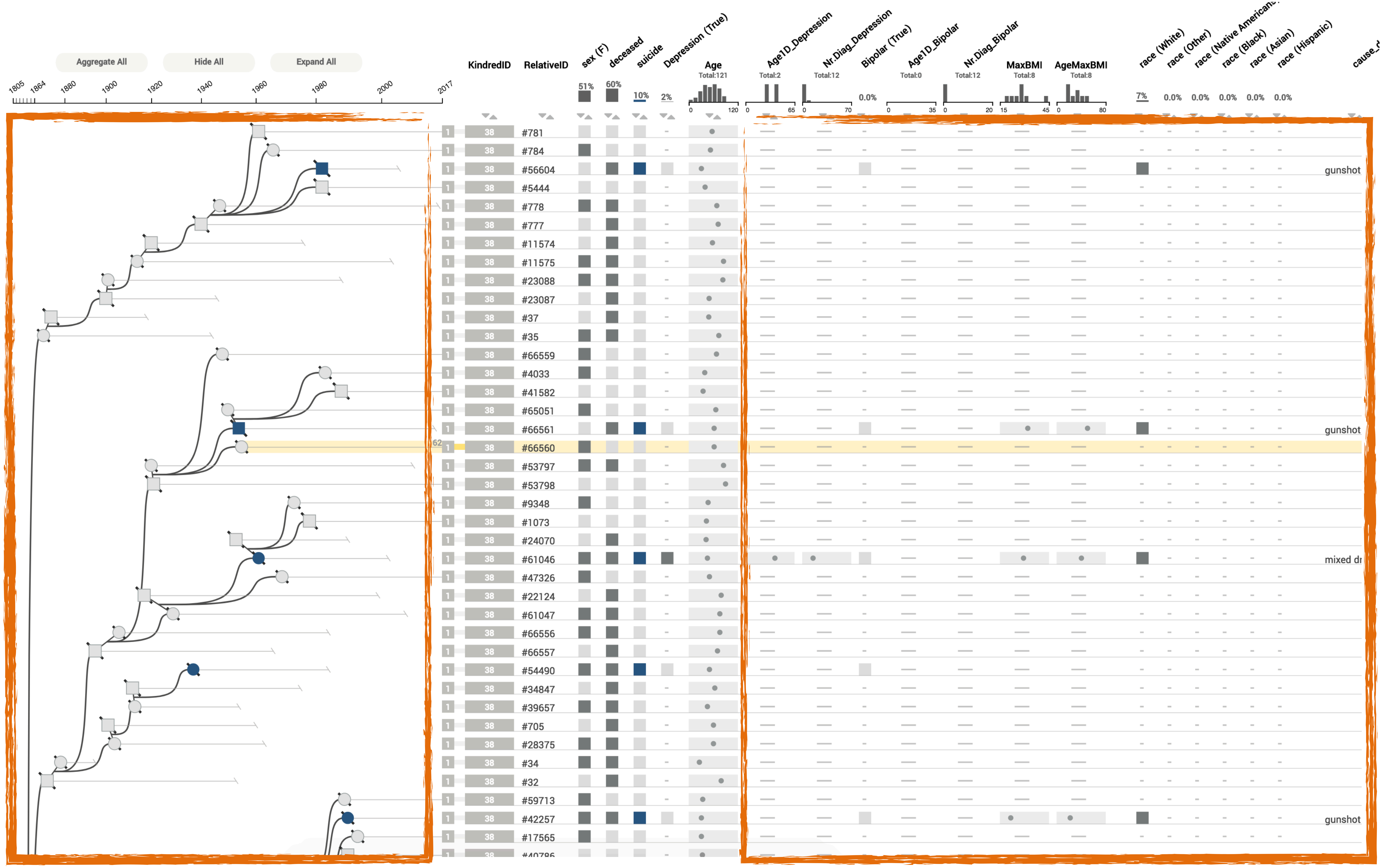


Linearization



Basic Encoding

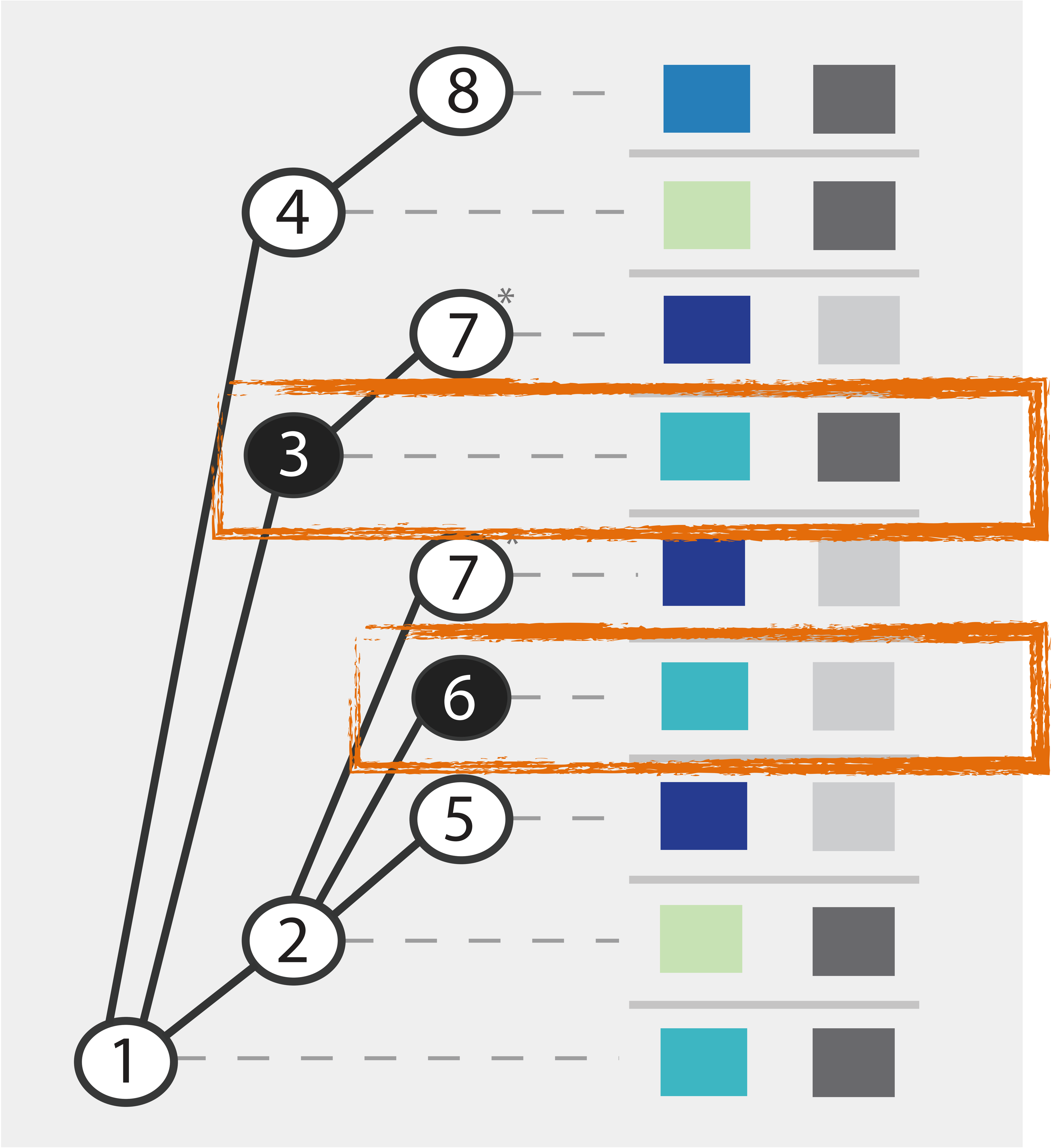




Can’t show many people

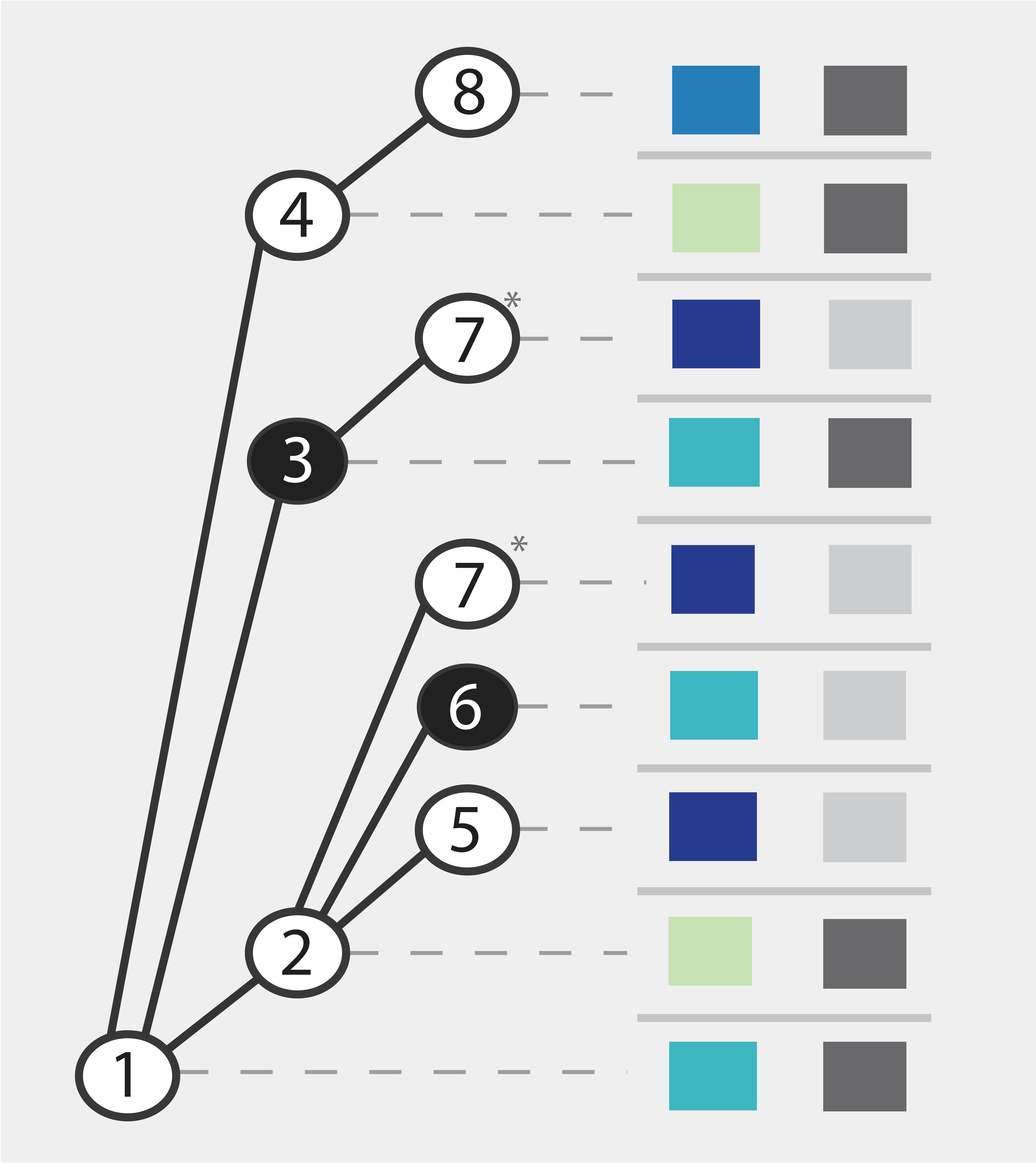
Lots of missing data

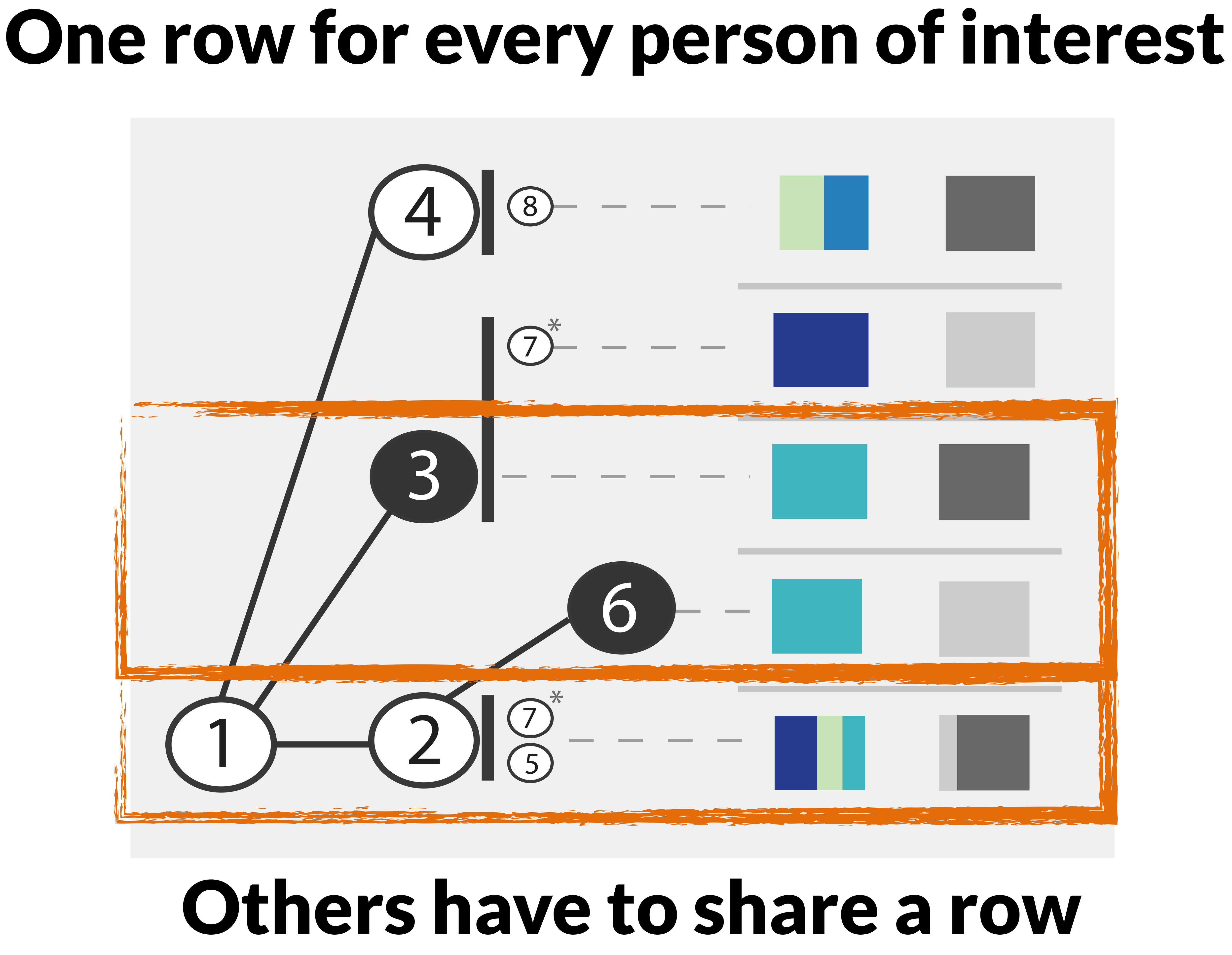
Aggregation

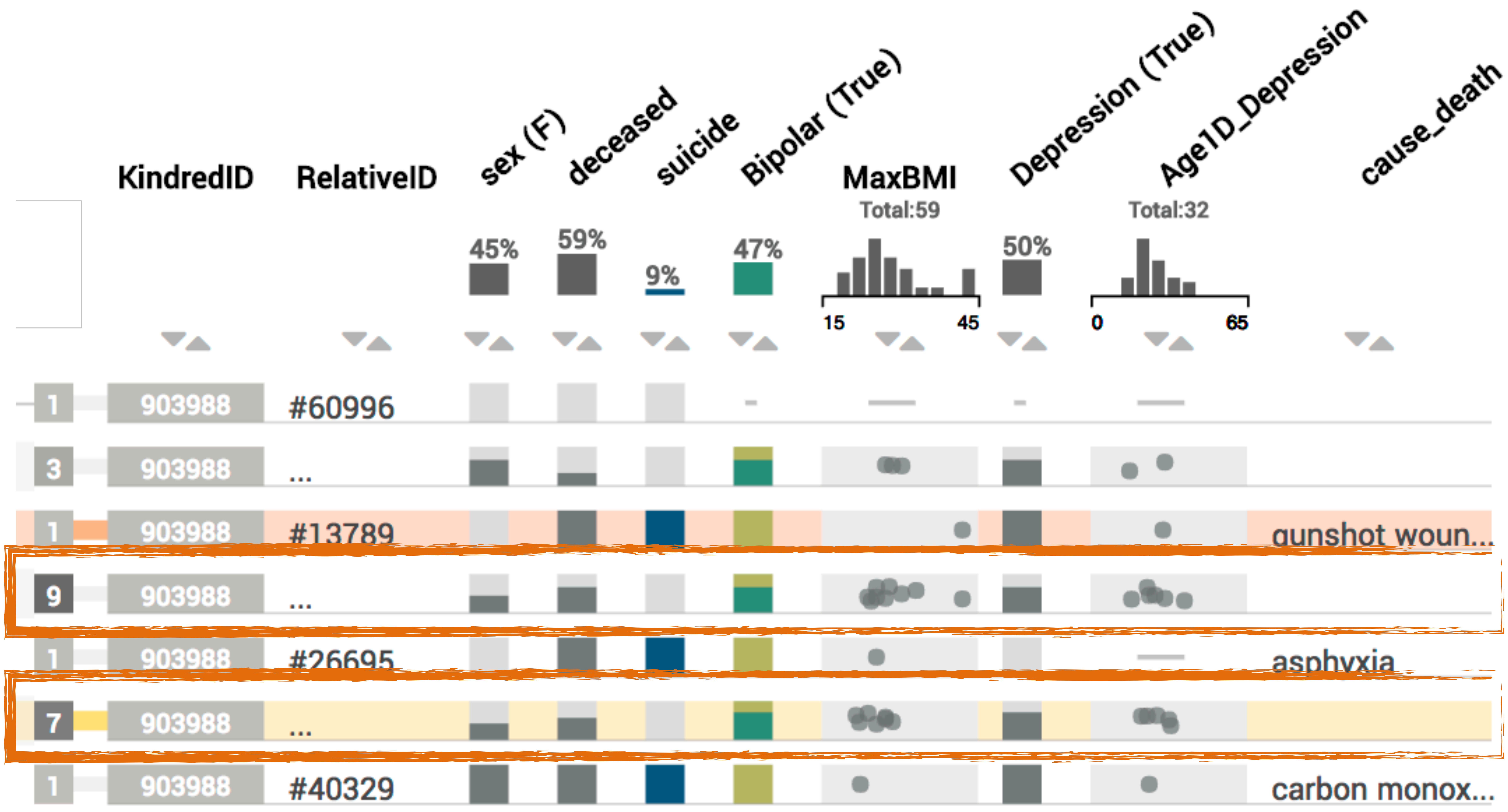


People of Interest

Aggregation

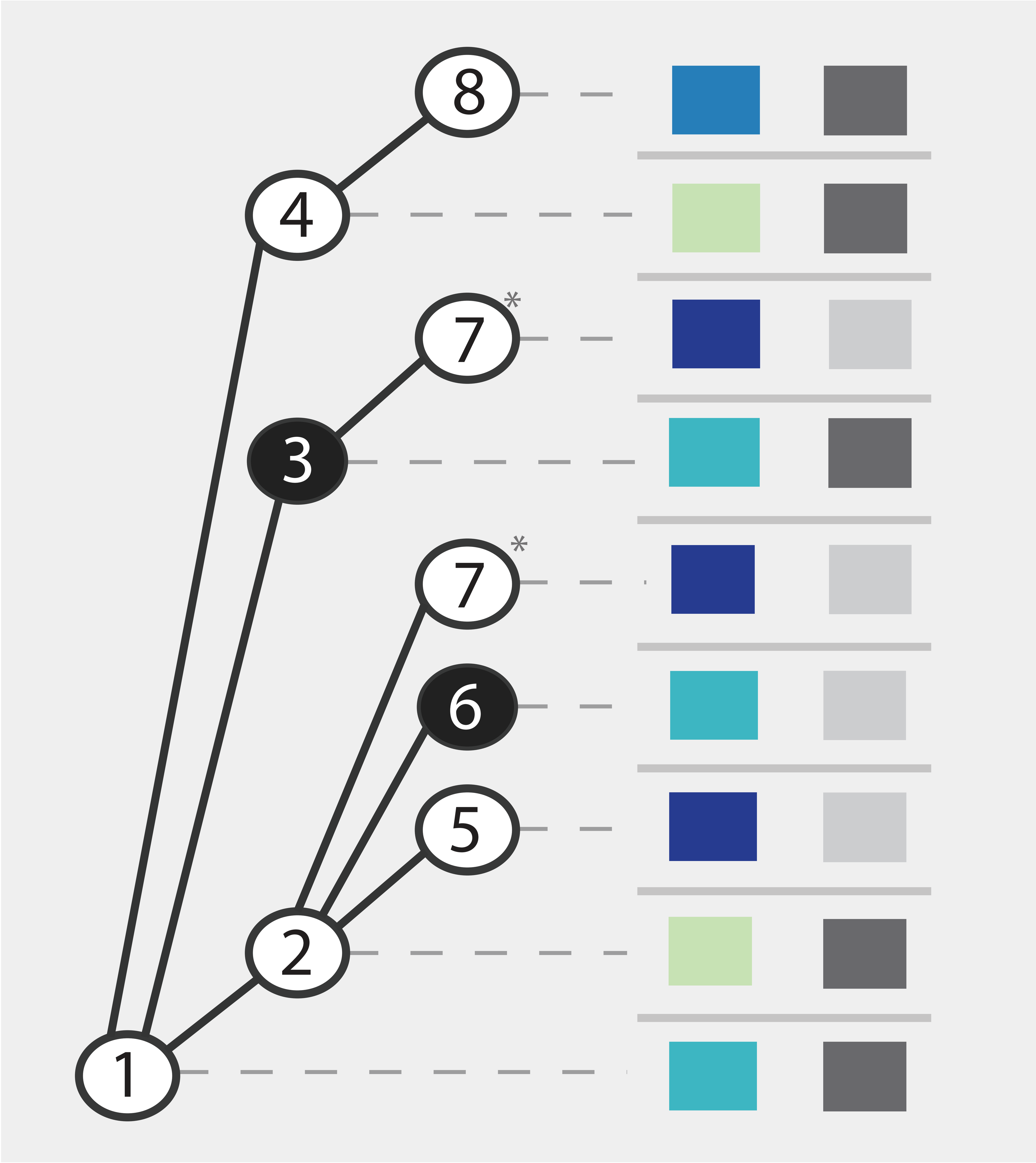


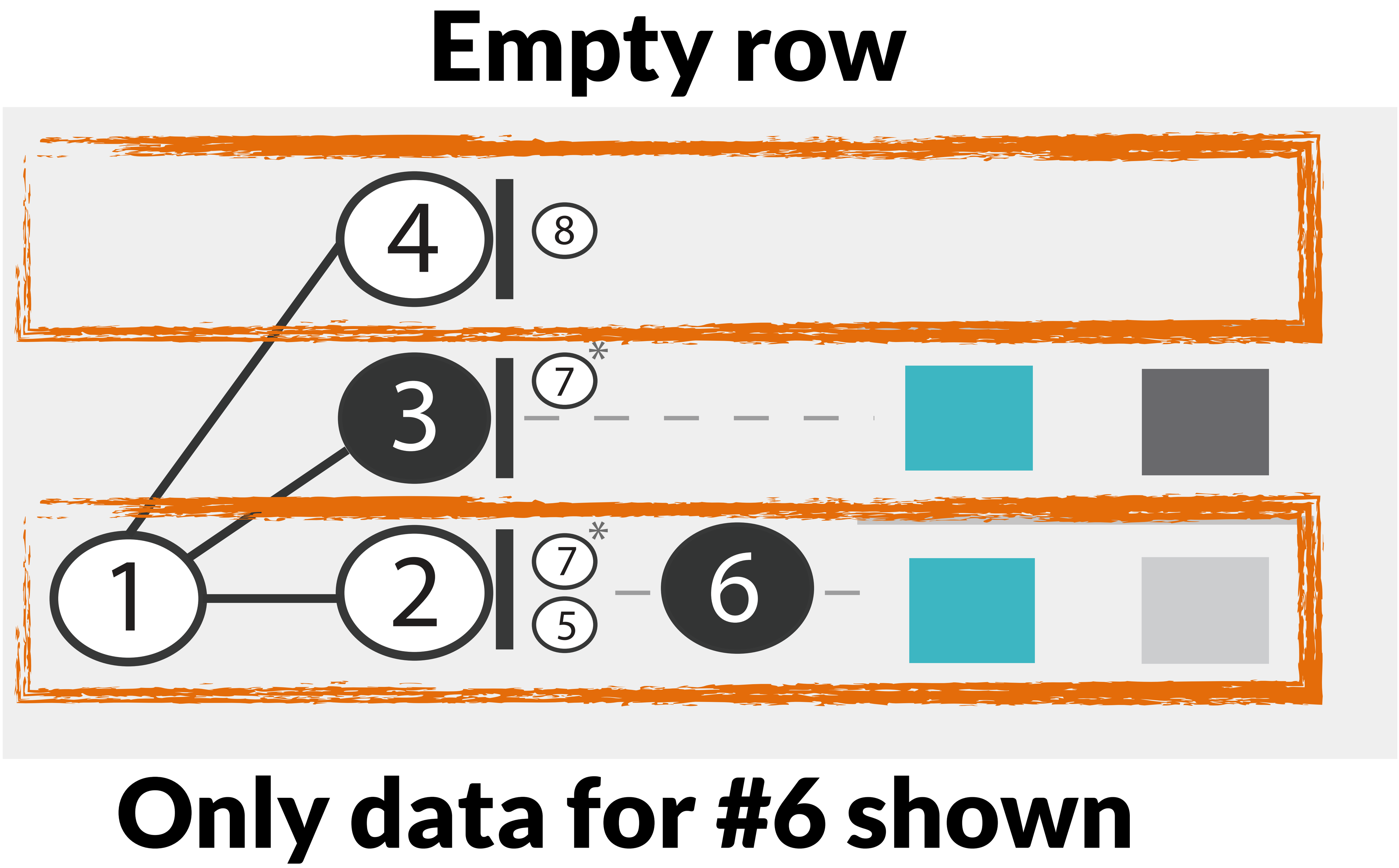




Aggregated Rows

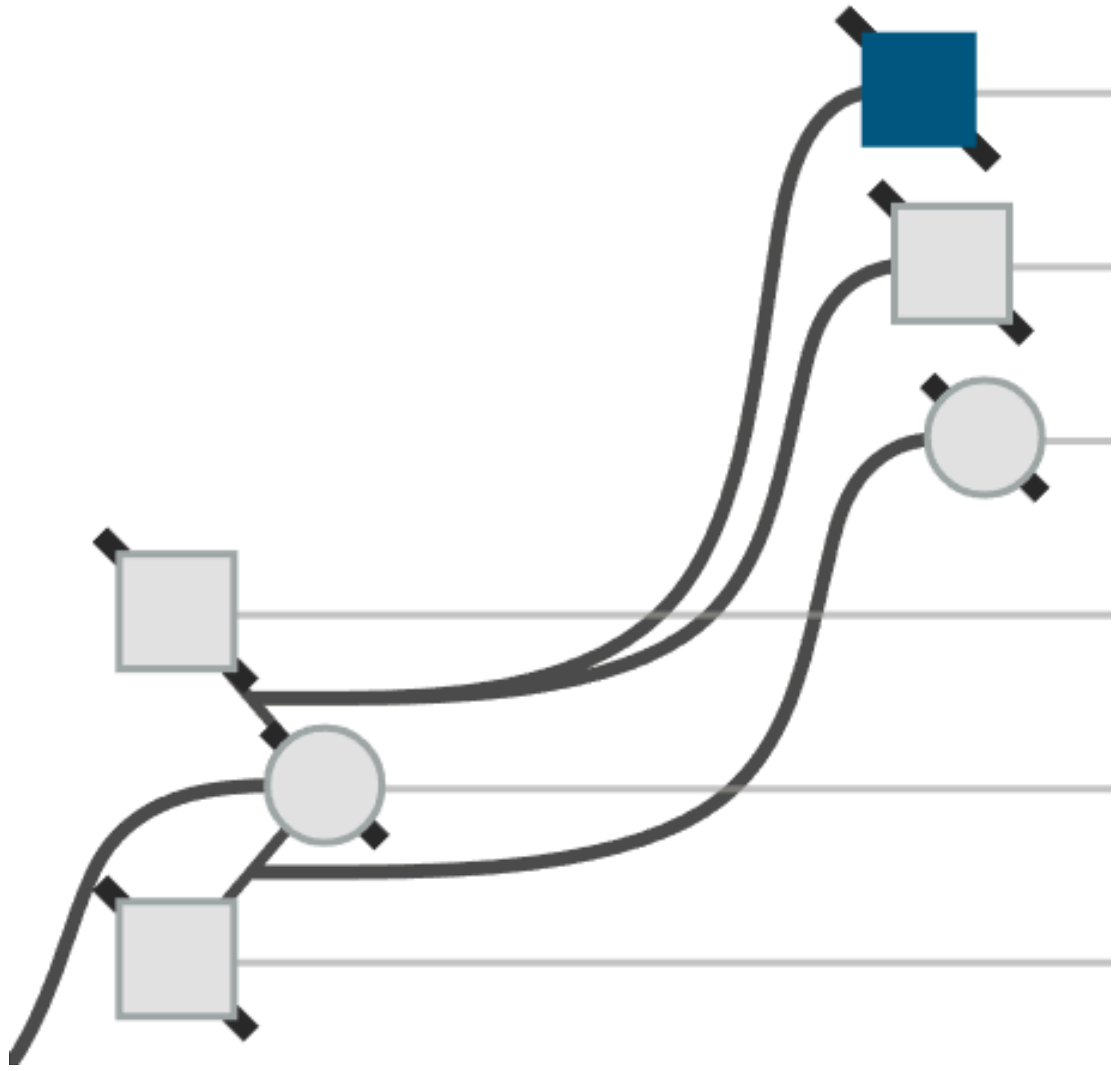
More Aggressive: Hiding



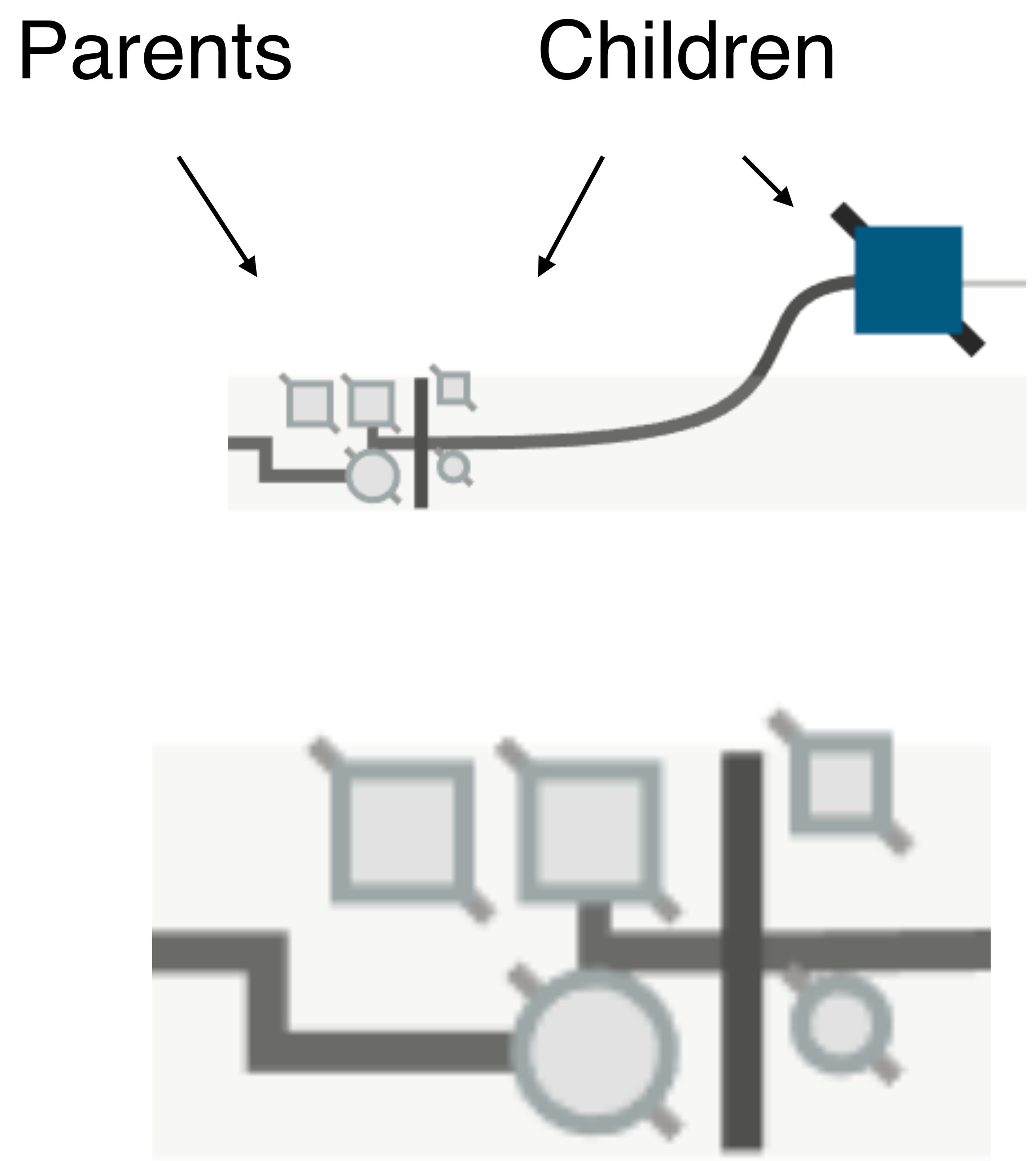


Implicit Encoding of Family

No Aggregation



Aggregation



Search and Filter Strategies

Find families enriched for a trait

Scan relevant cases for relationships

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 Weissbrod[5], Dan Geiger[4], Mary Wahl[1,2,6], Michael Gershovits[2], 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, ...

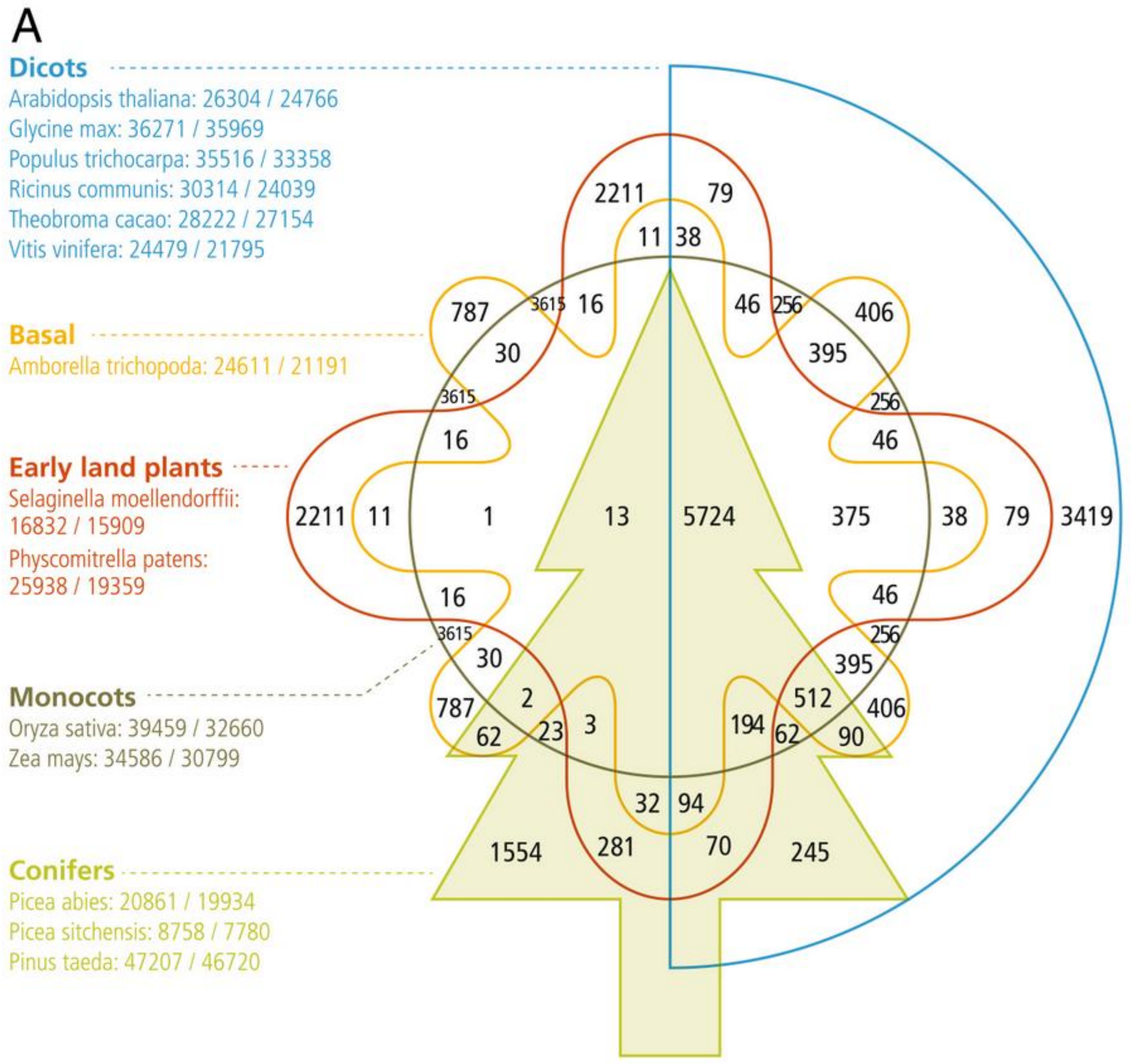
http://lineage.caleydoapp.org

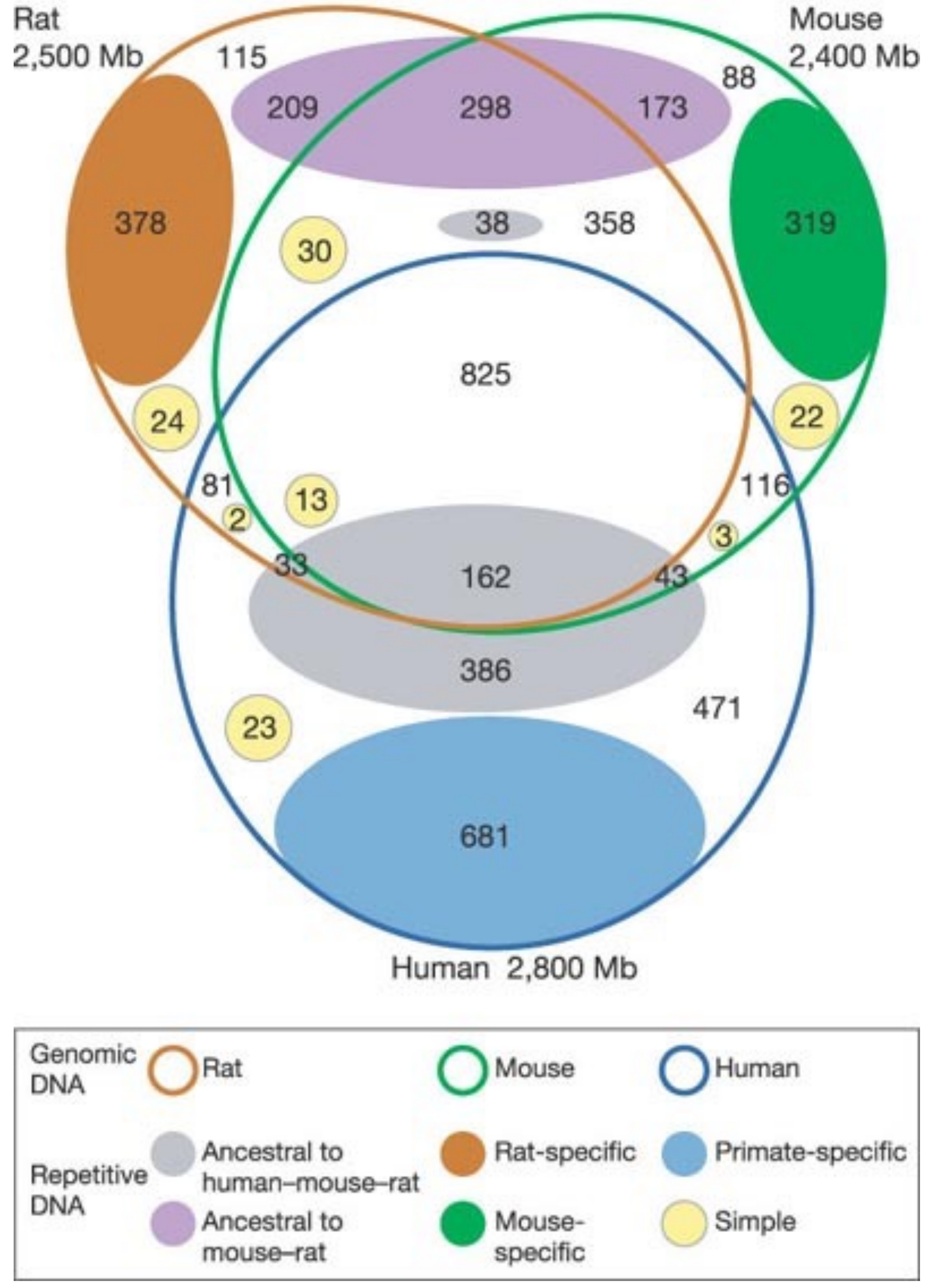
Paper: http://vdl.sci.utah.edu/

[InfoVis'14]

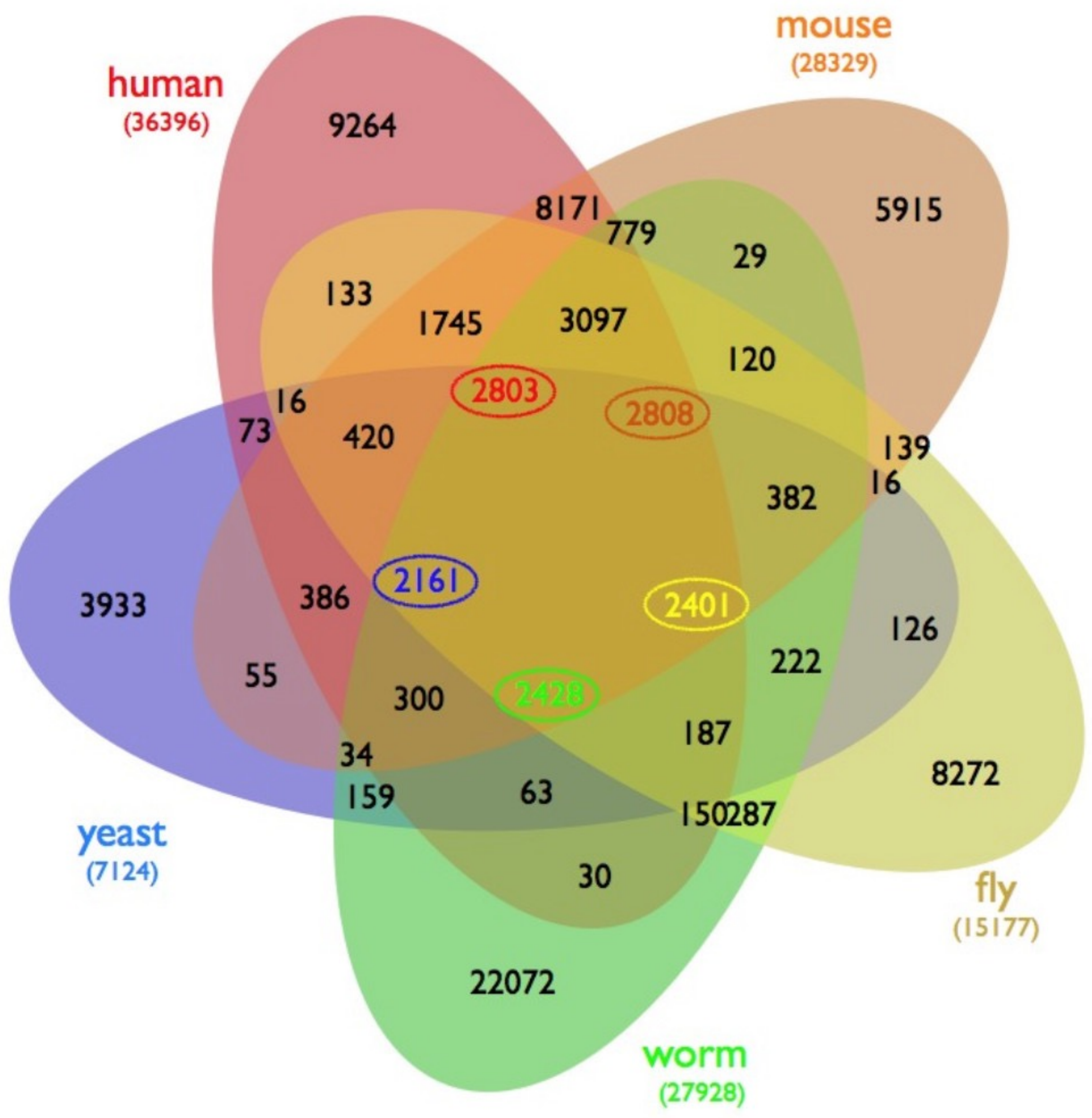
UpSet
Visualizing Intersecting Sets

[Neale et al., BMC Genome Biology, 2014]

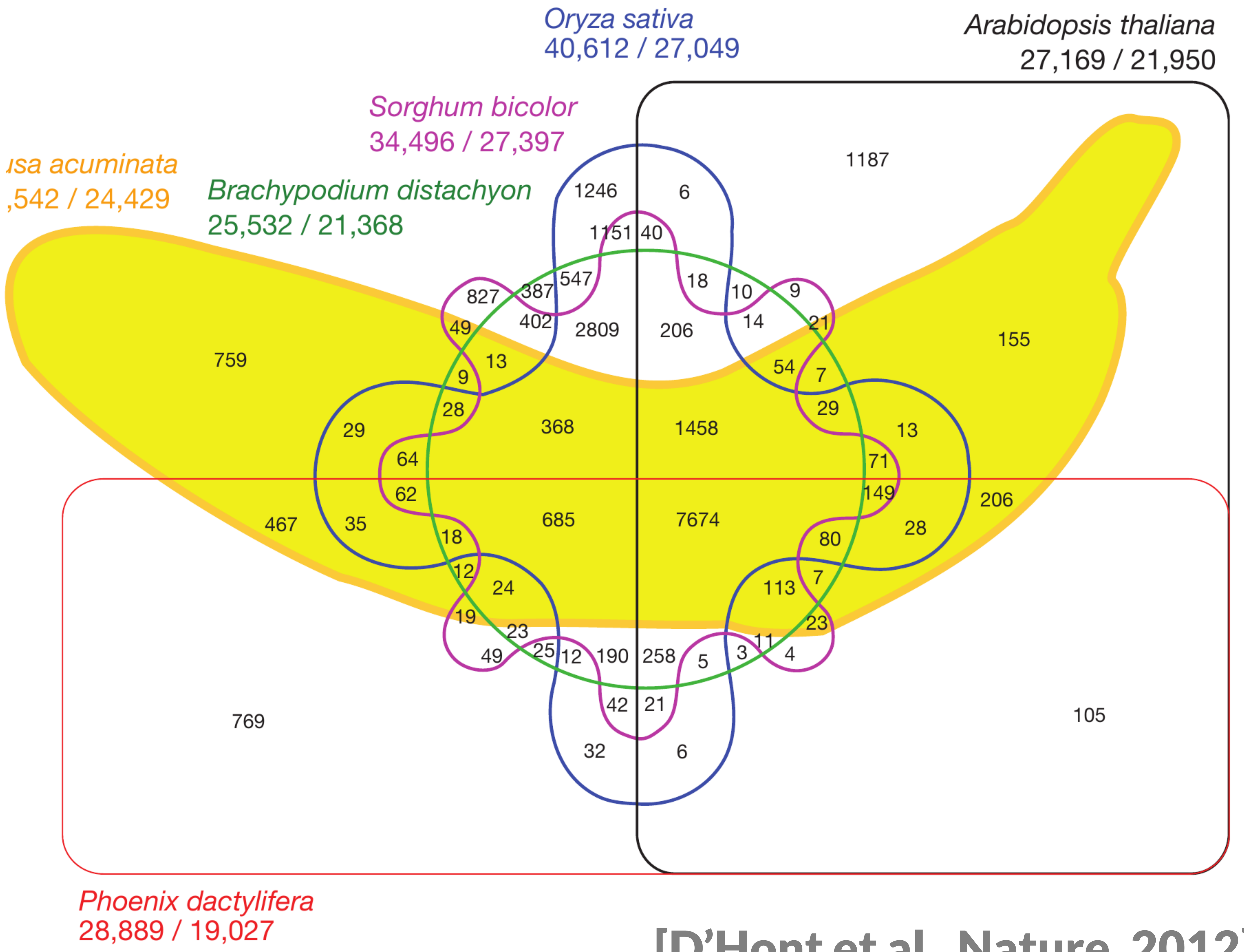




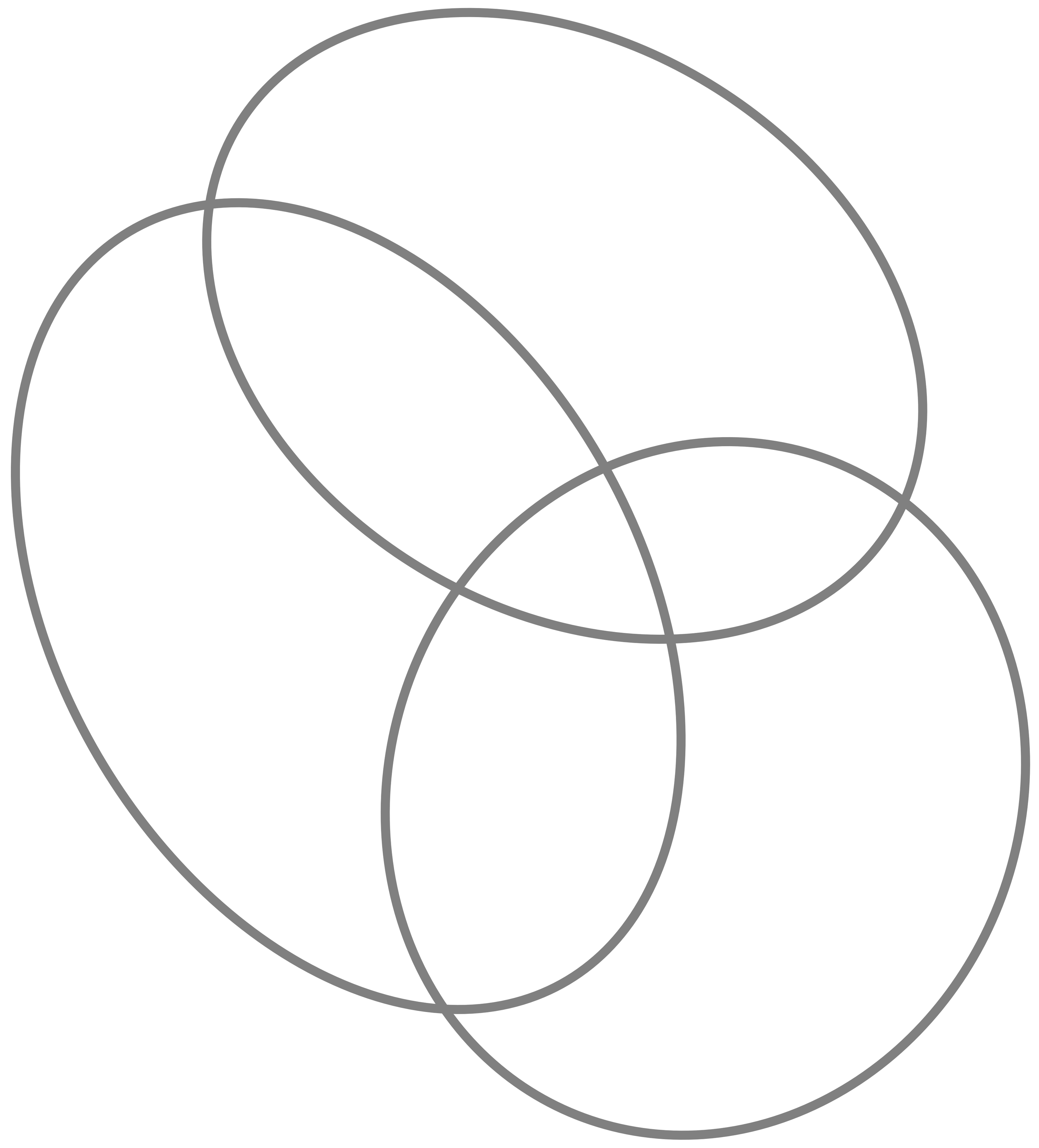
[Gibbs et al., Nature, 2004]



[Wiles et al., BMC Systems Biology]



[D'Hont et al., Nature, 2012]



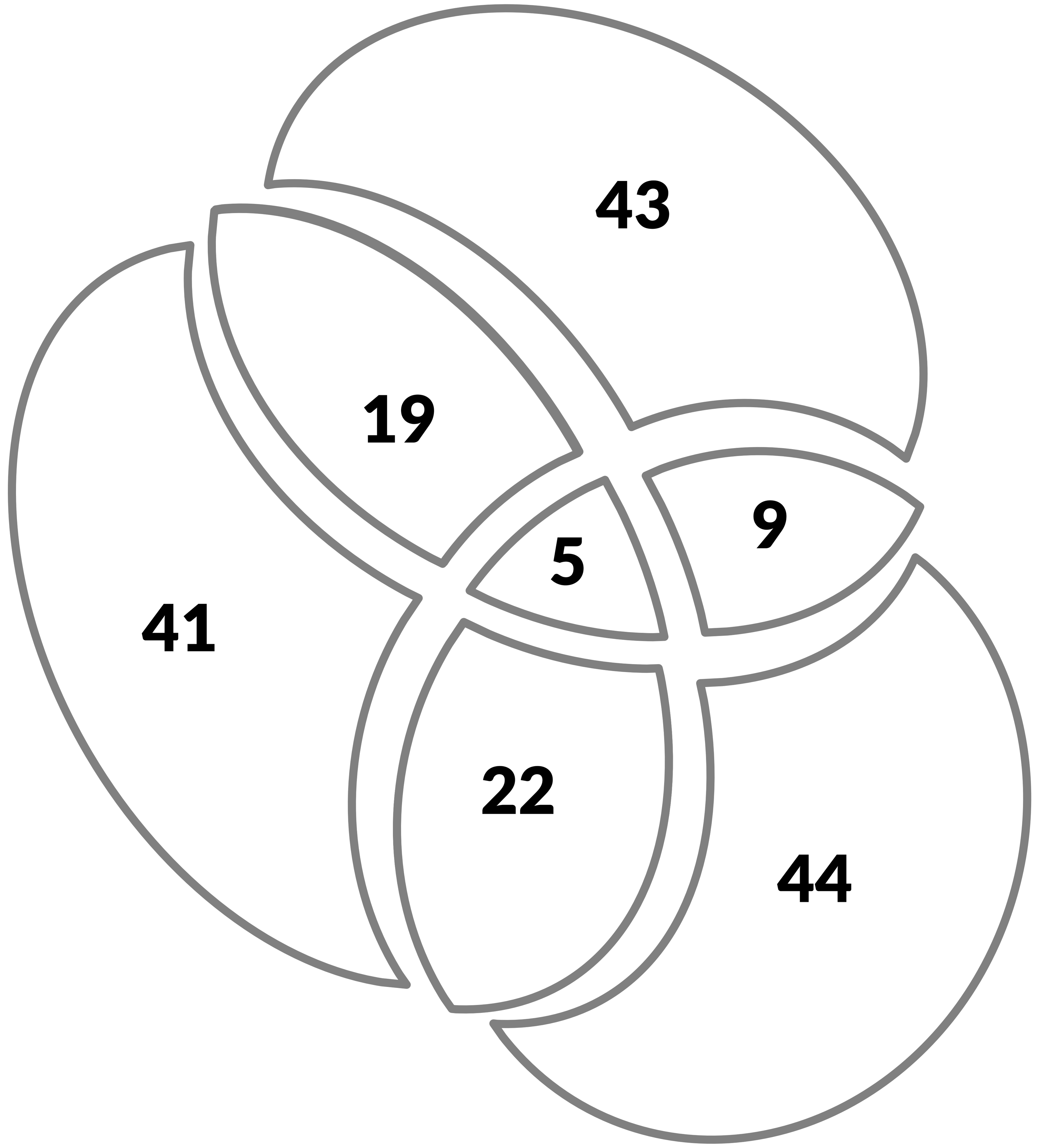
[created with EulerAPE]

22

?

>
<

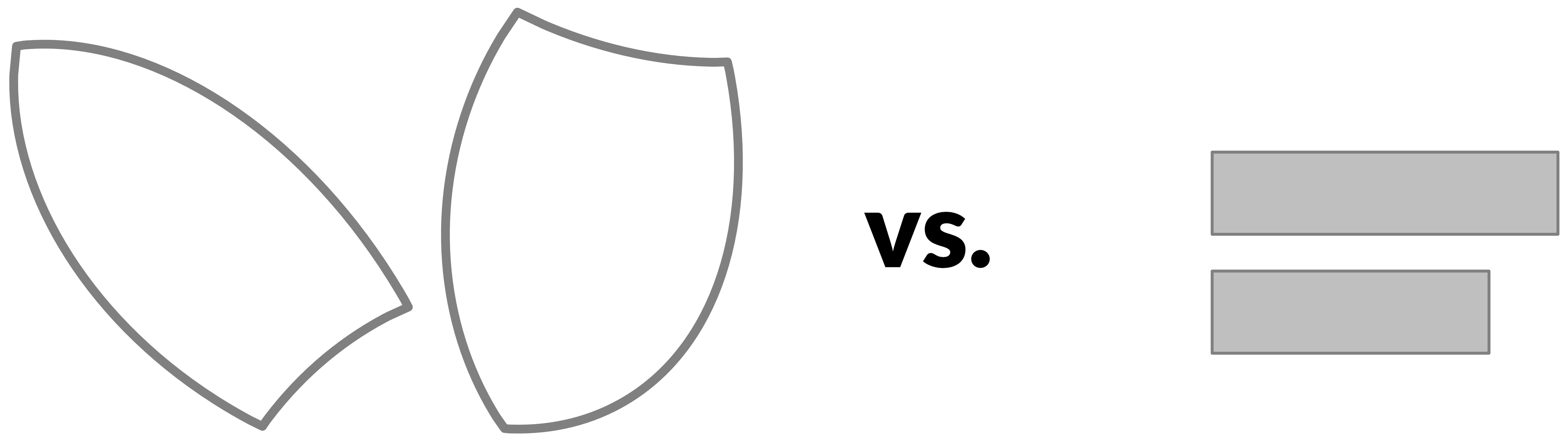
19



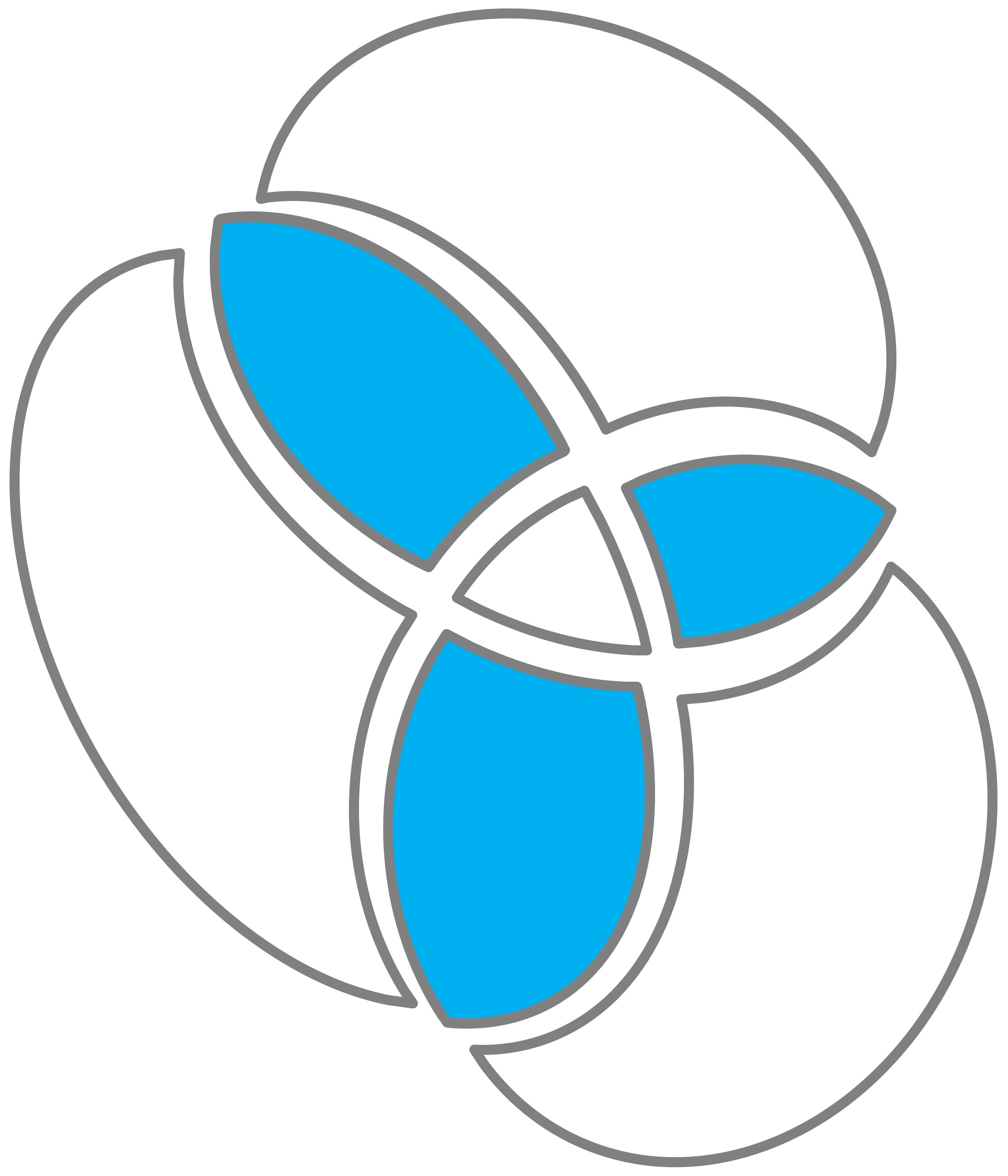
[created with EulerAPE]

Set Vis Goals

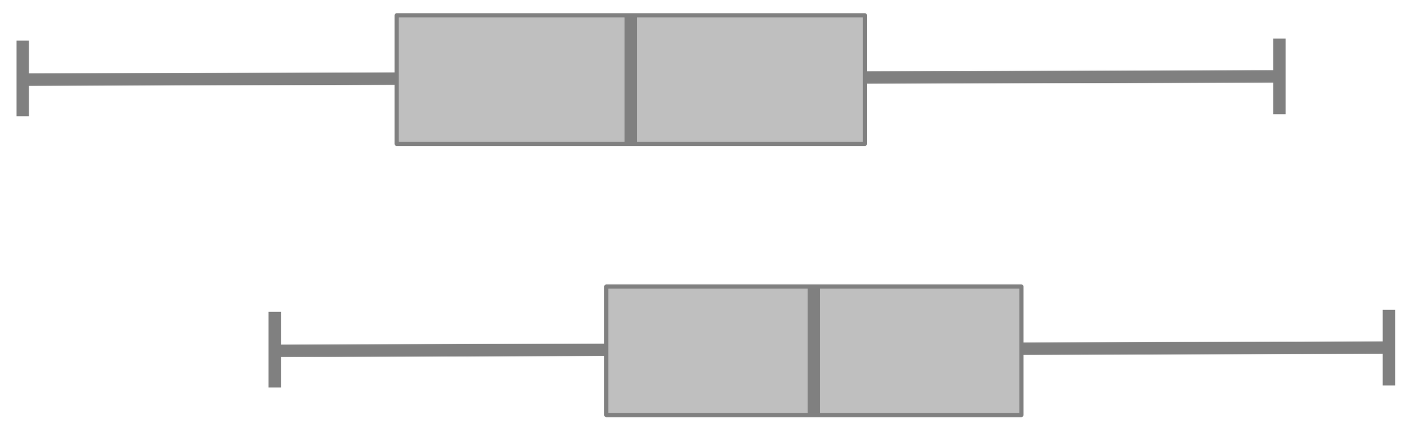
1. Efficient visual encoding



2. Creating complex slices of a dataset



3. Visualize attributes



[Movie Lens Dataset]

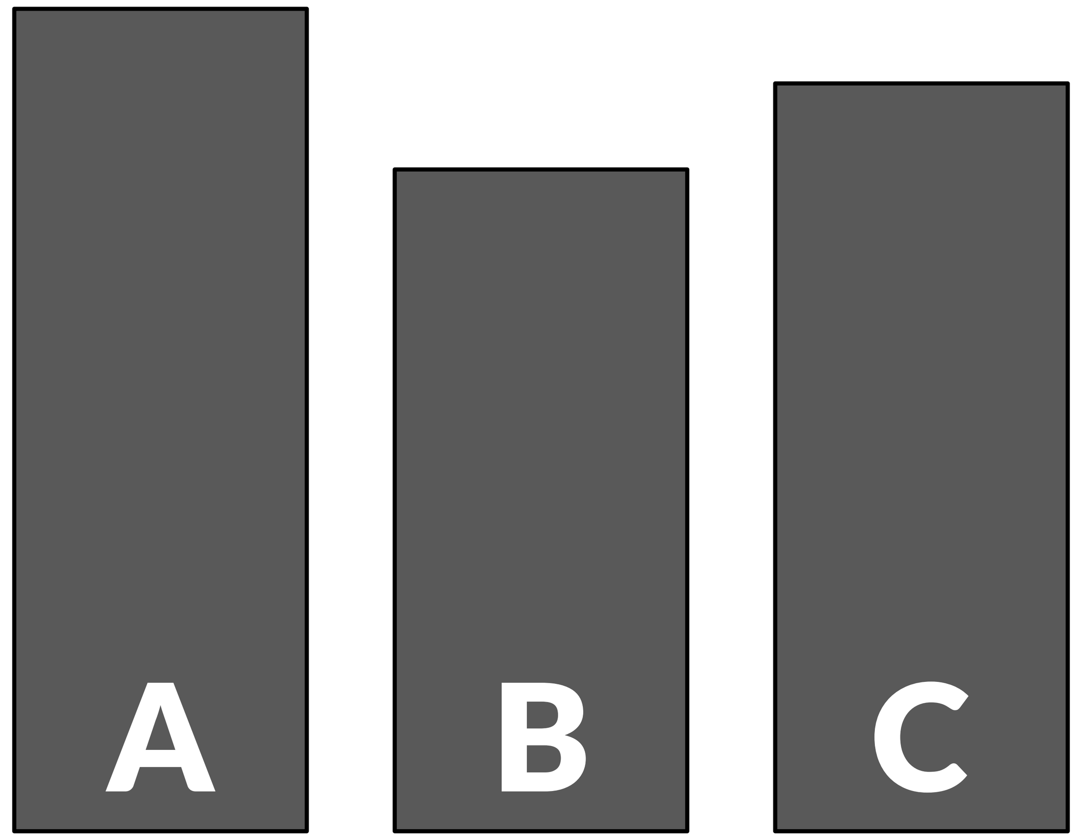
Attribute Details

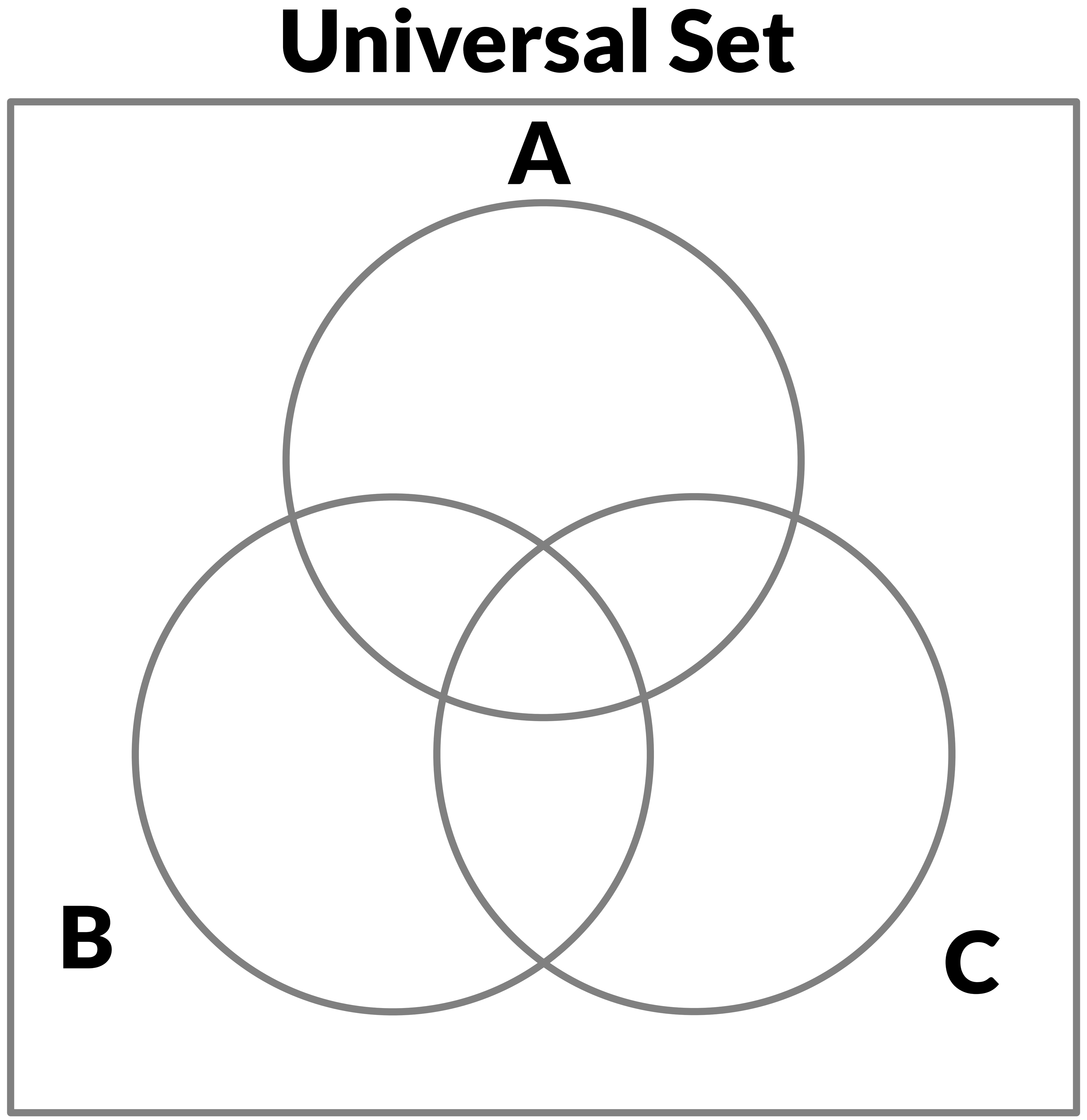
Visualizing Intersections

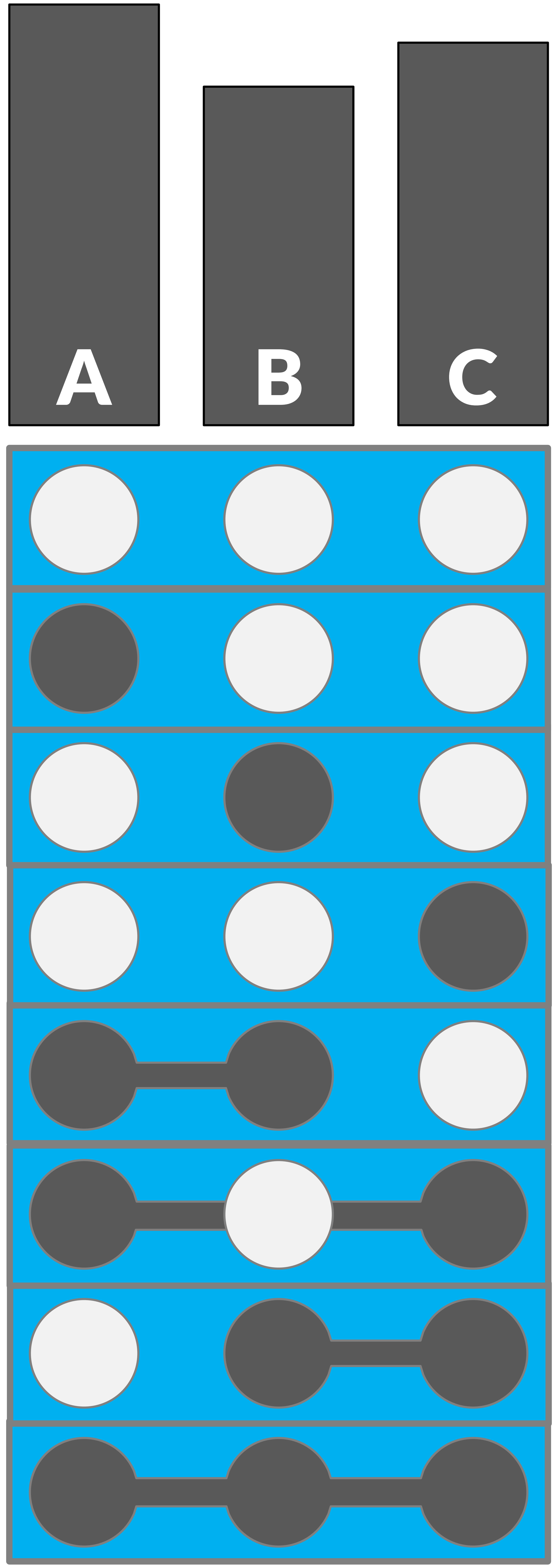
Visualizing Properties

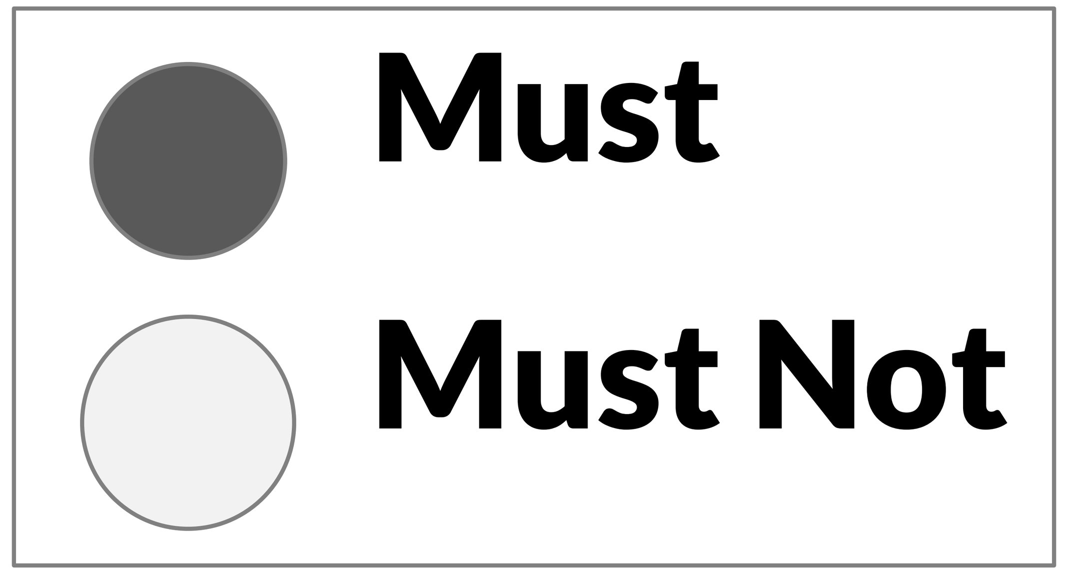
Element List & Queries

Visualizing Intersections

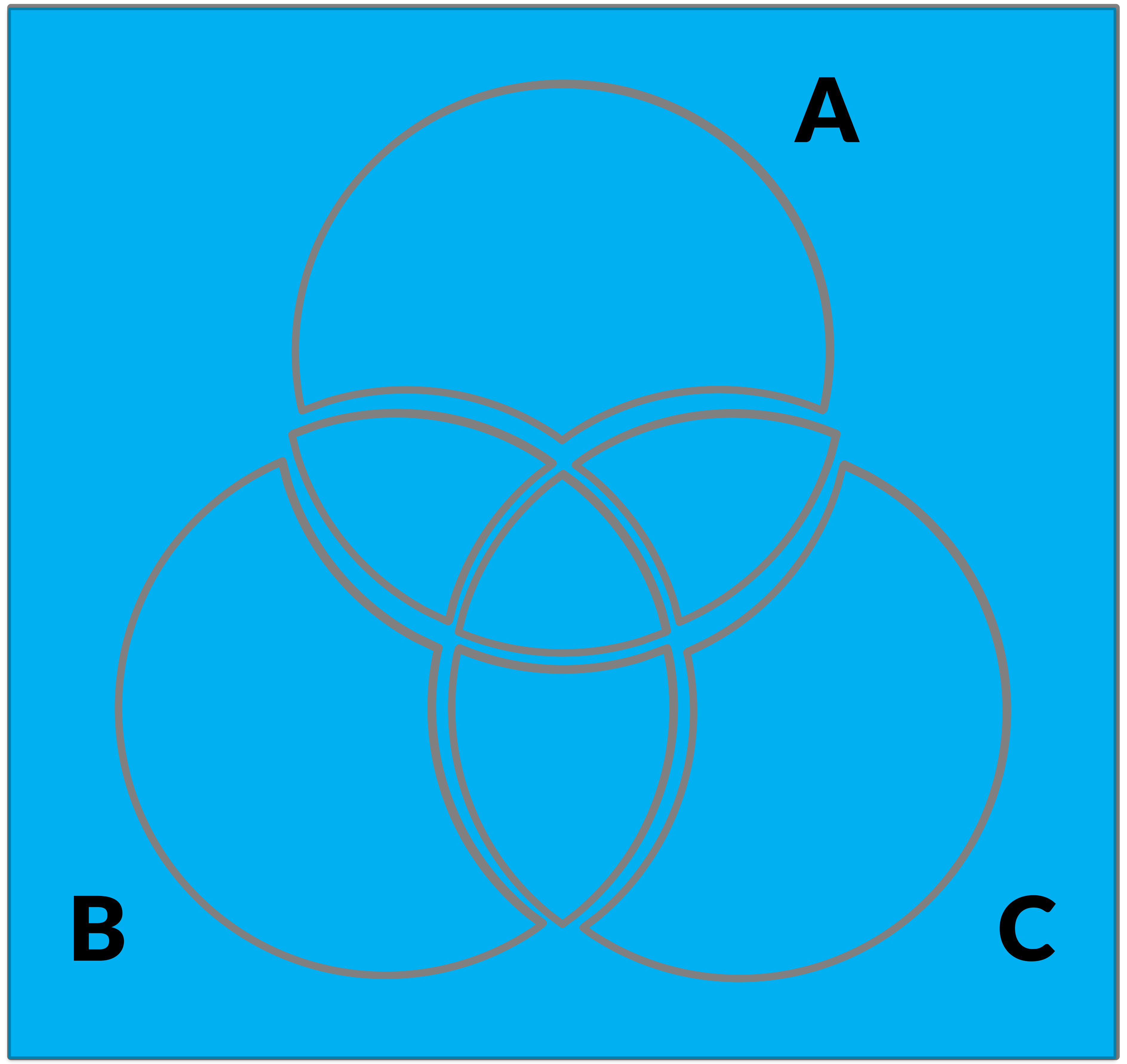

A
B
C

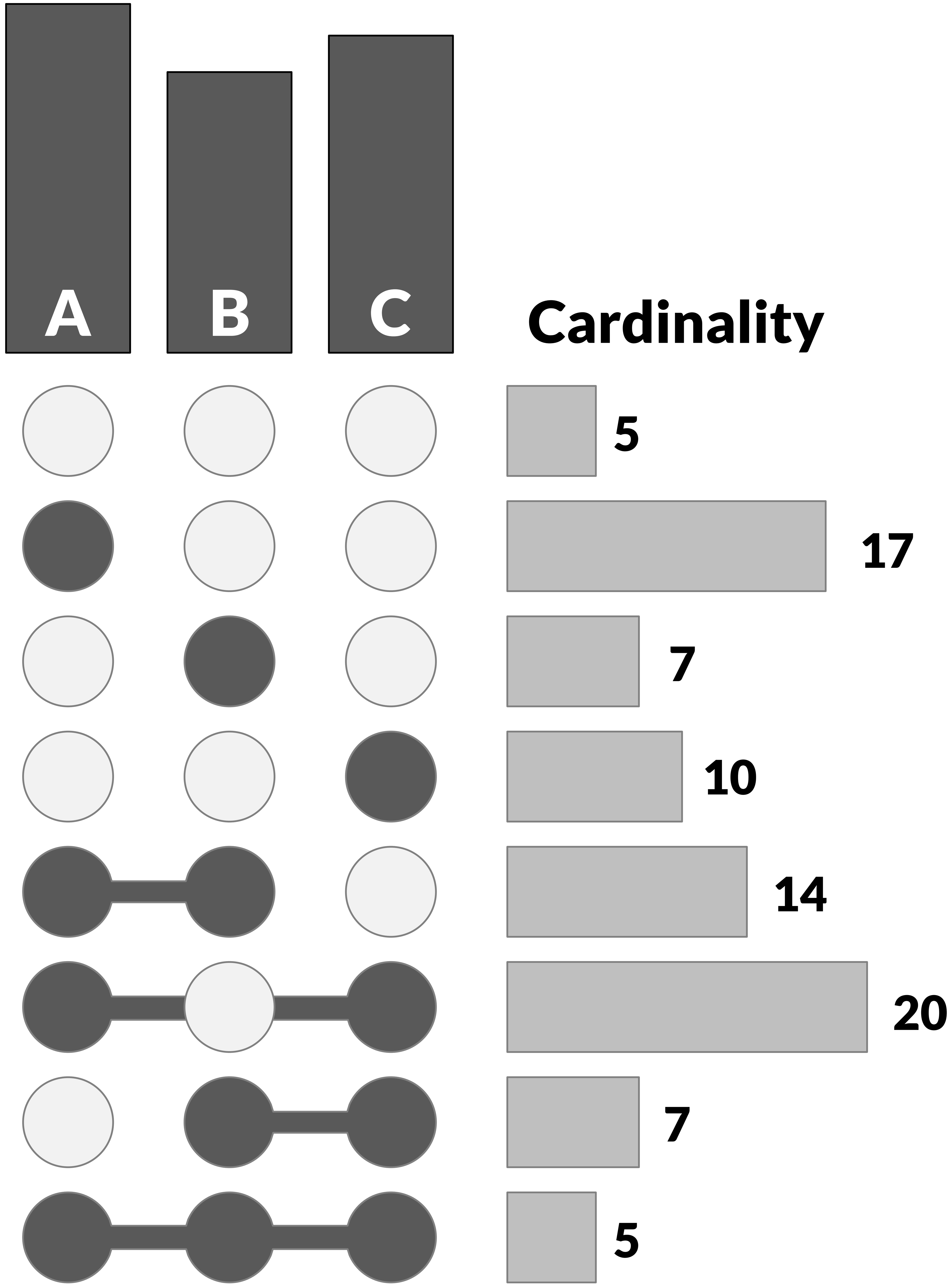

Universal Set
A
B
C

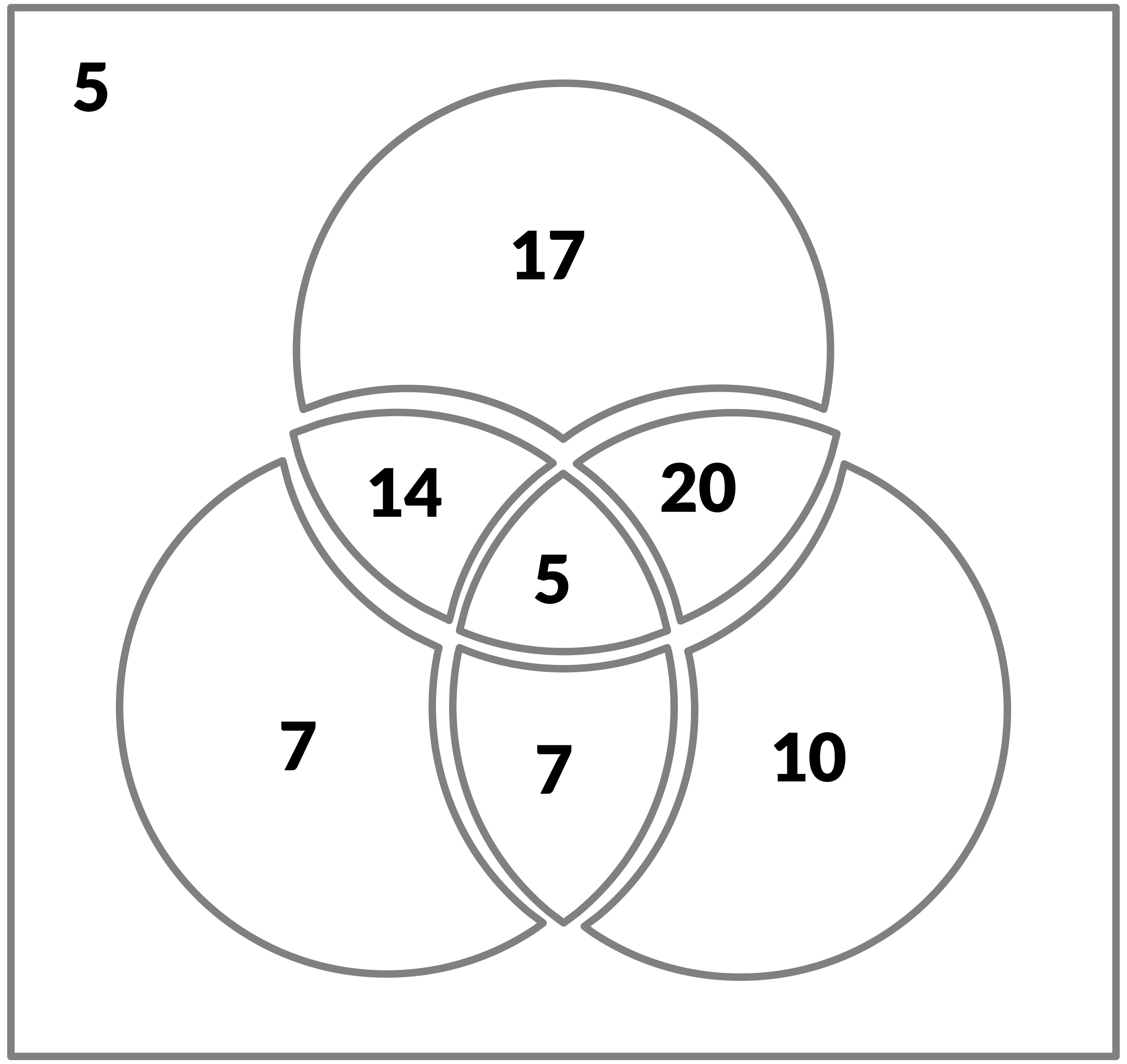




Universal Set




A
B
C
Cardinality
5
17
7
10
14
20
7
5

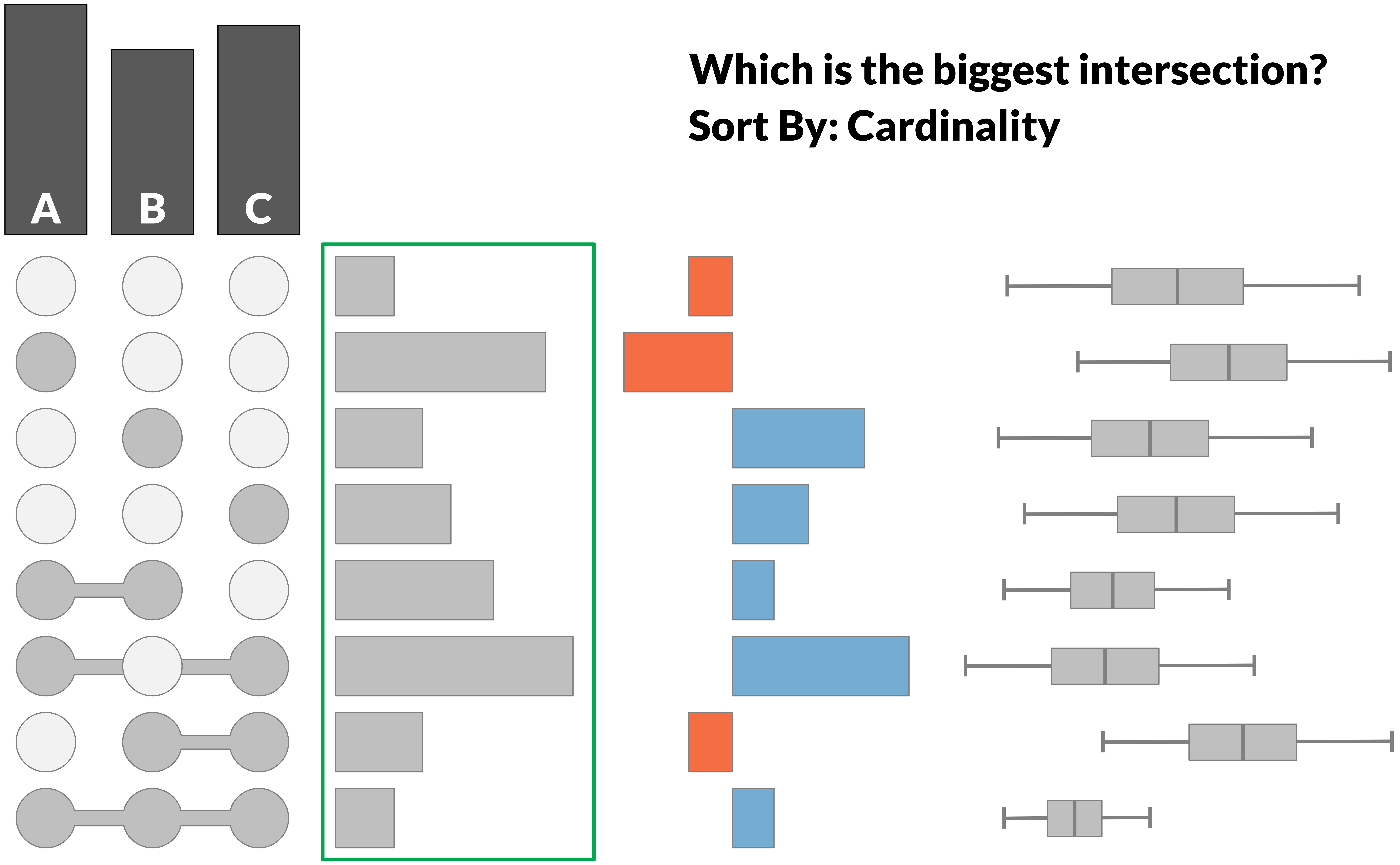

5
17
14
20
5
7
7
10

Plotting Attributes


How surprising is the size of an intersection?
What's the distribution of an attribute in an intersection?
Additional Plots
A
B
C
Deviation
Attributes

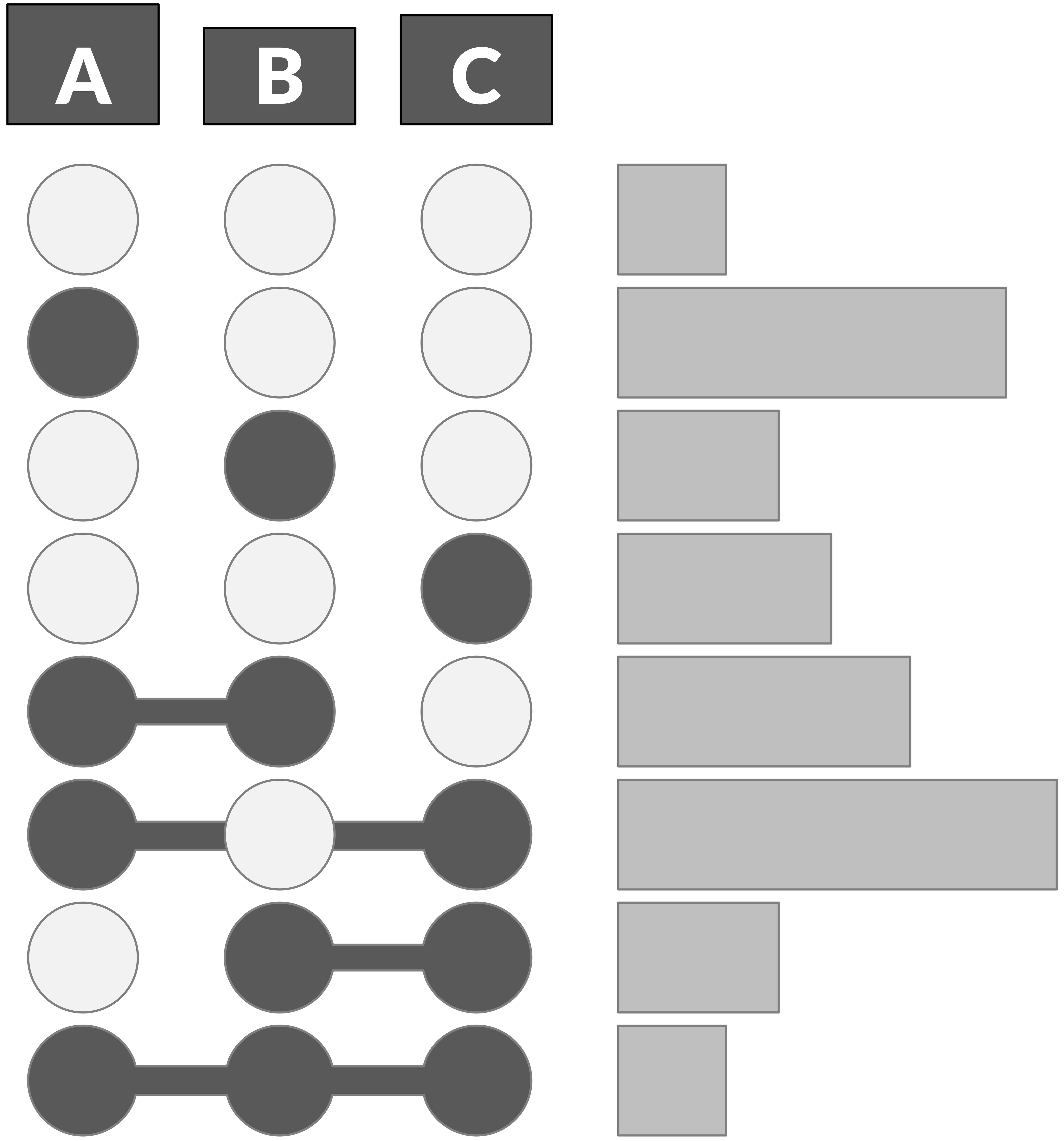
First, aggregate by
Don't Aggregate
Then, aggregate by
Don't Aggregate
Sort by
Degree
Cardinality
Deviation
Aggregates
Collapse All
Expand All
Row Height
Large
Data
Min Degree:
0
Max Degree:
5
Hide Empty Intersections
Dataset Information
Name: Movies Genres
Sets: 17
Attributes: 6
Elements: 3883
Author: grouplens
Description: MovieLens ratings dataset, curated and filtered by Alsallakh.
Source: http://grouplens.org/d..
Set Selection
0 - 9
Batch Add Sets
Sort Sets
Thriller
Romance
Children
Crime
Musical
Mystery
Fantasy
Western
Noir
Query
Comedy
Drama
Action
Adventure
Documentary
Horror
SciFi
War
Cardinality
Deviation
Release Date
Average Ratir
Times Watche
Drama-Comedy
Action-Comedy

Sorting

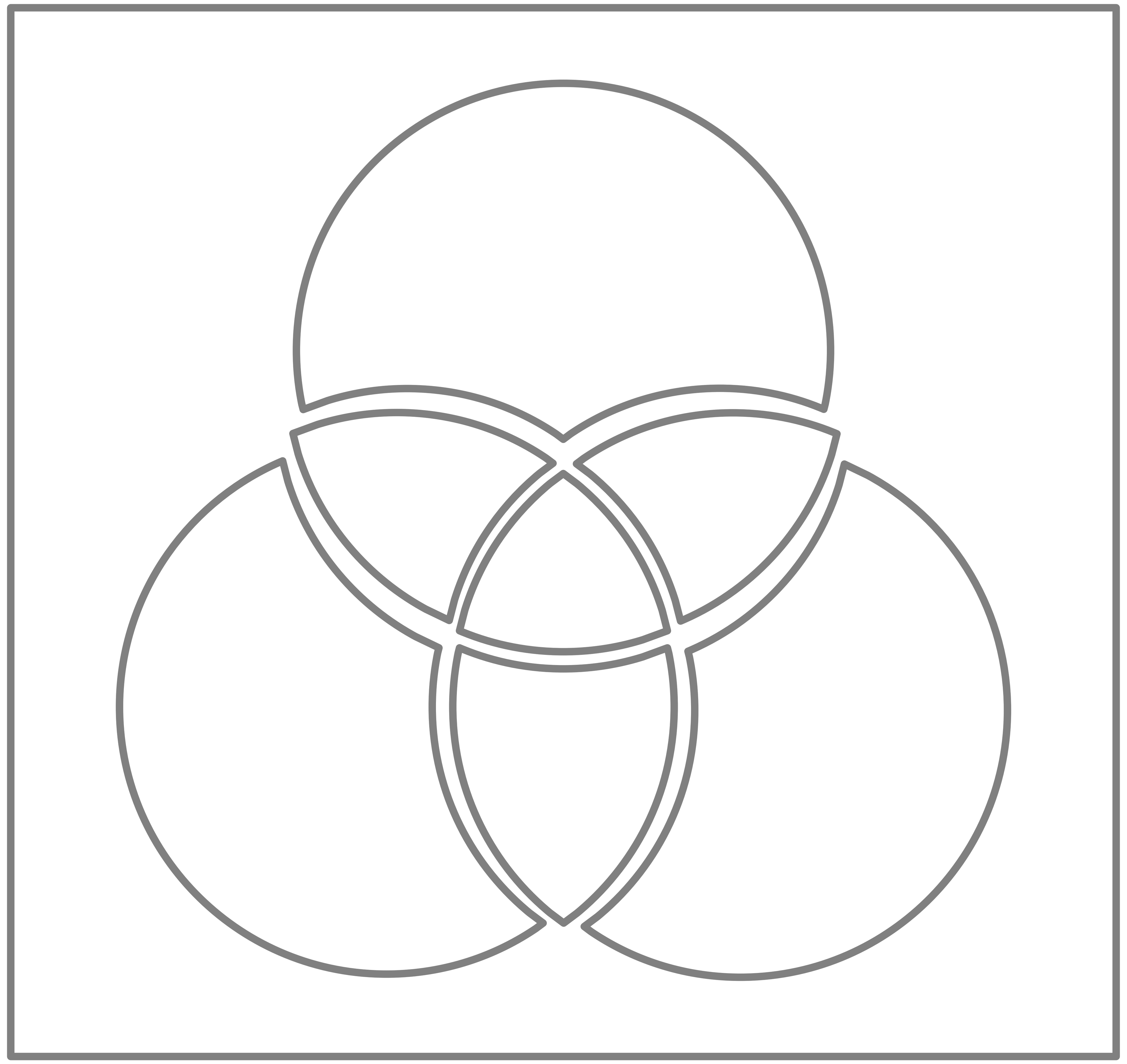

Which is the biggest intersection?
Sort By: Cardinality
A
B
C

First, aggregate by
Don't Aggregate
Then, aggregate by
Don't Aggregate
Sort by
Degree
Cardinality
Deviation
Aggregates
Collapse All
Expand All
Row Height
Large
Data
Min Degree:
0
Max Degree:
5
Hide Empty Intersections
Dataset Information
Name: Movies Genres
Sets: 17
Attributes: 6
Elements: 3883
Author: grouplens
Description: MovieLens ratings dataset, curated and filtered by Alsallakh.
Source: http://grouplens.org/d...
Query
Set Selection
0 - 9
Batch Add Sets
Sort Sets
Thriller
Romance
Children
Crime
Musical
Mystery
Fantasy
Western
Noir
Comedy
Drama
Action
Adventure
Documentary
Horror
SciFi
War
Cardinality
Deviation
Release Date
Average Ratin
Times Watche
389
823
1172
161
80
118
224
55
19
210
31
21
4
36
19
10
66
22
4
12
19

Aggregation

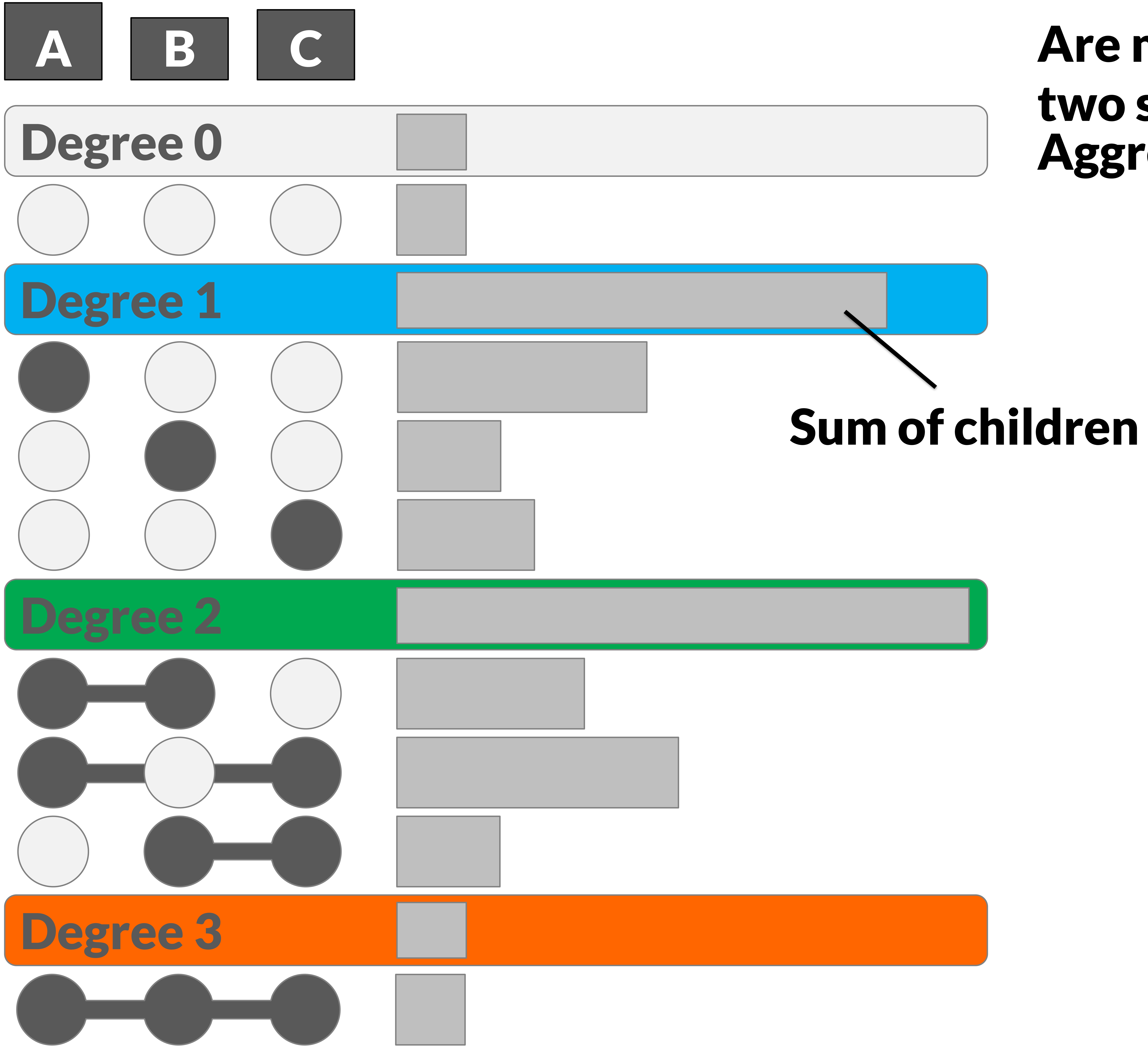


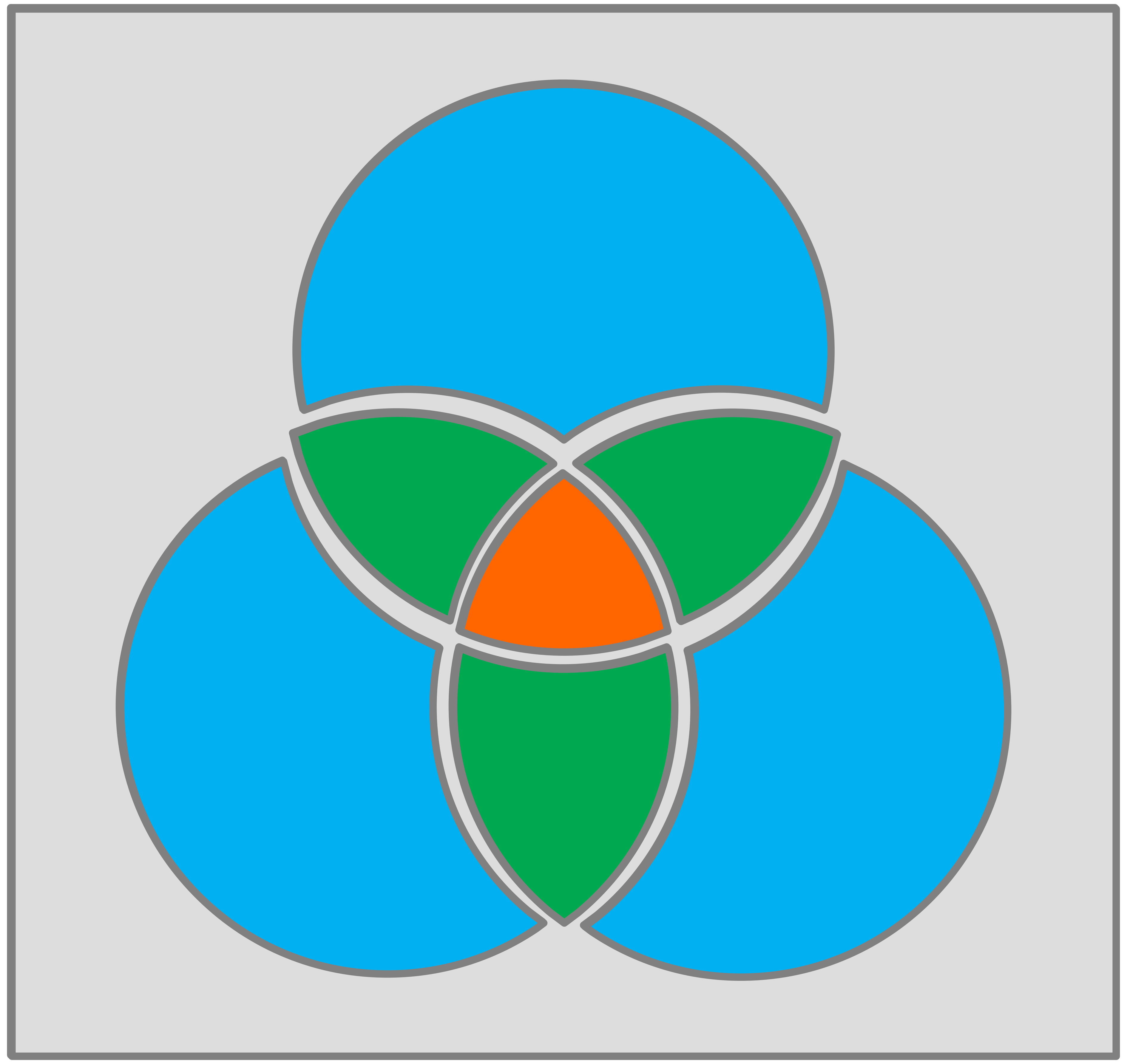
Are many items shared between two sets?
Aggregate By: Degree



Are many items shared between two sets?

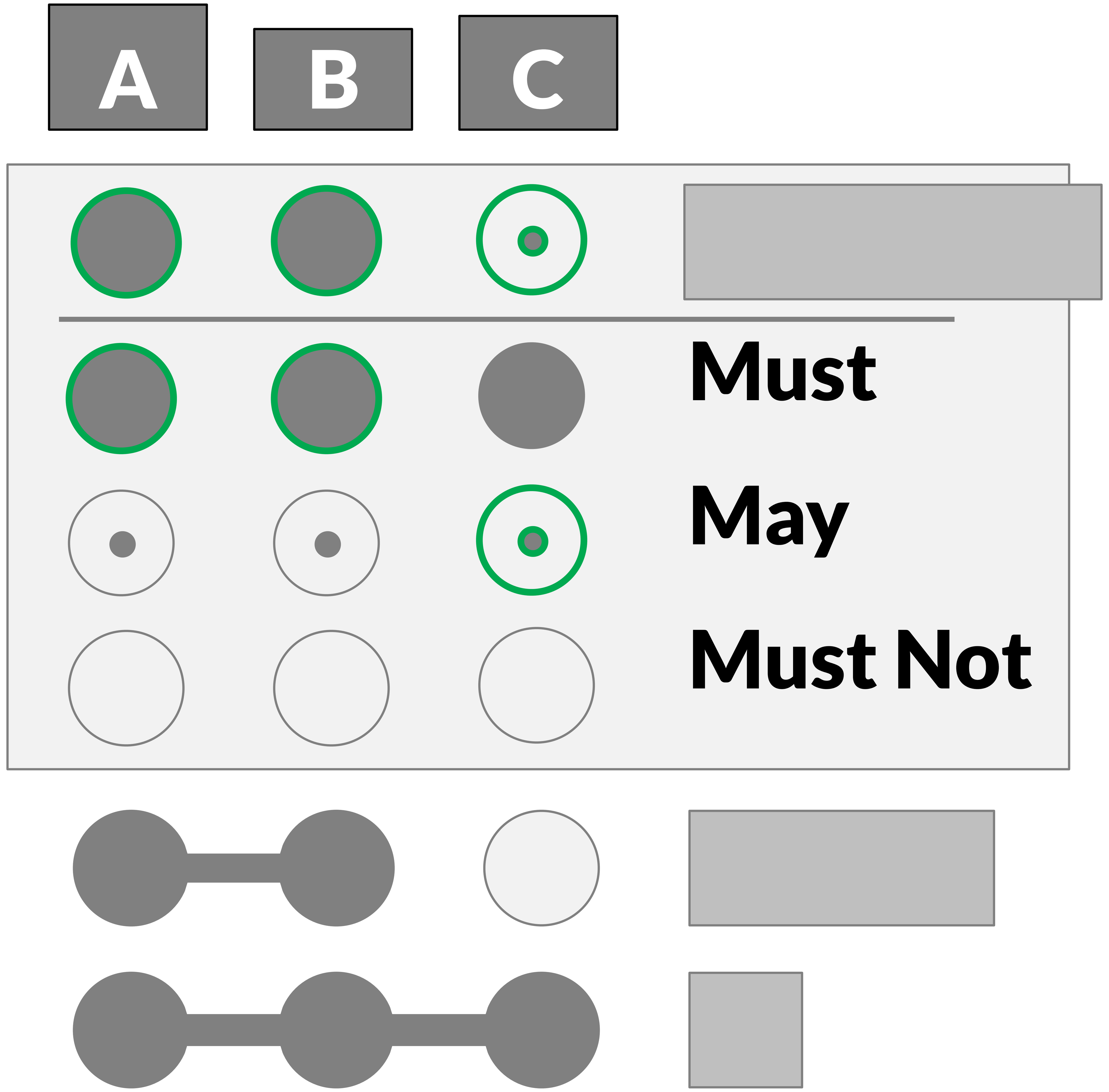
Aggregate By: Degree

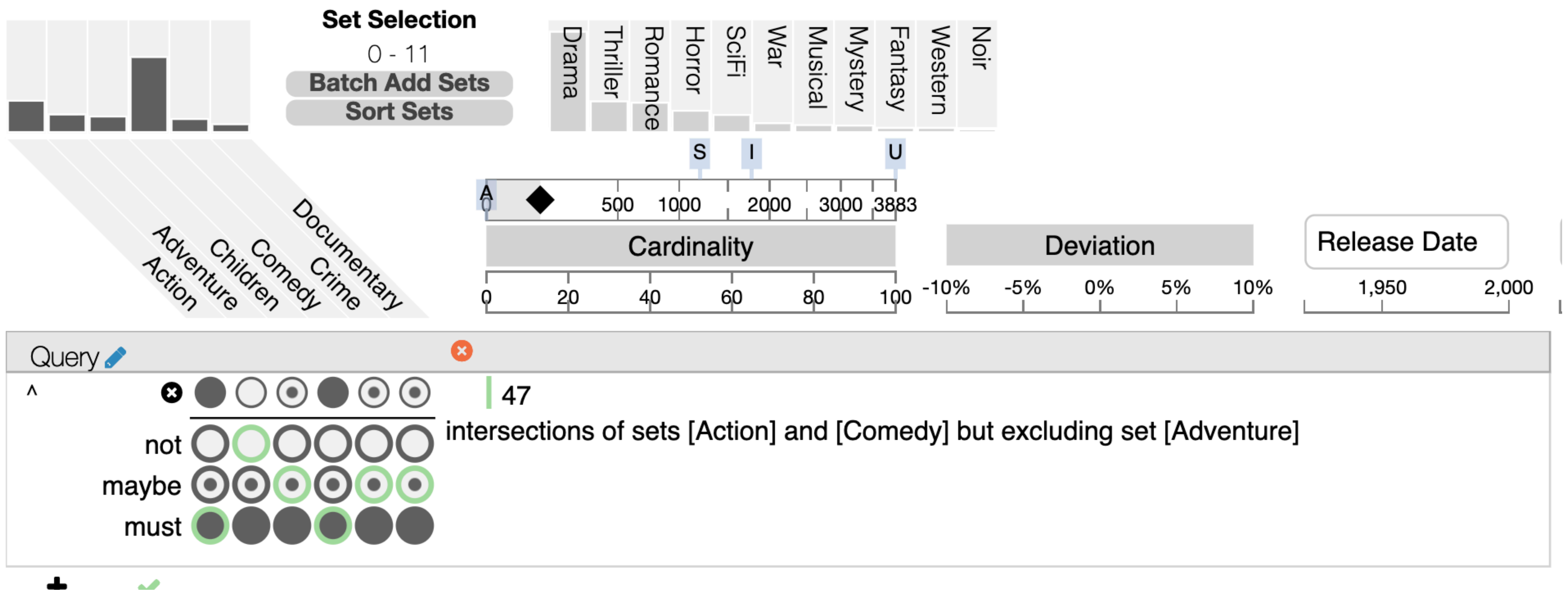




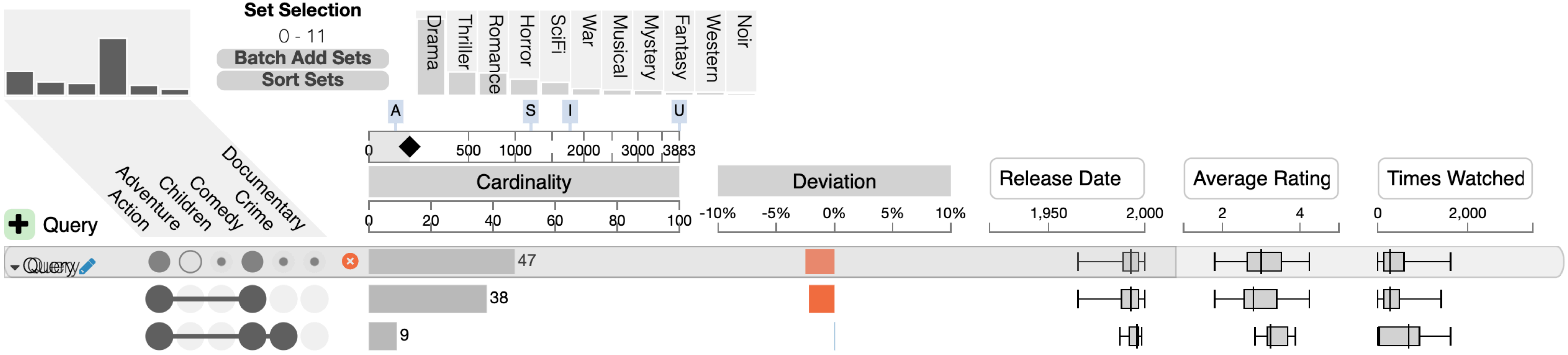
First, aggregate by
Degree
Then, aggregate by
Don't Aggregate
Sort by
Degree
Cardinality
Deviation
Aggregates
Collapse All
Expand All
Row Height
Large
Data
Min Degree:
0
Max Degree:
5
Hide Empty Intersections
Dataset Information
Name: Movies Genres
Sets: 17
Attributes: 6
Elements: 3883
Author: grouplens
Description: MovieLens ratings dataset, curated and filtered by Alsallakh.
Source: http://grouplens.org/d..
Set Selection
0 - 9
Batch Add Sets
Sort Sets
Thriller
Romance
Children
Crime
Musical
Mystery
Fantasy
Western
Noir
Query
Comedy
Drama
Action
Adventure
Documentary
Horror
SciFi
War
Cardinality
Deviation
Release Date
Average Ratir
Times Watche
Degree 0 (in no set)
Degree 1 (1 set intersect.)
Degree 2 (2 set intersect.)
389
389
2652
823
1172
161
80
118
224
55
19
712
210
31
21
4
36
19
10
66
22

Queries


A
B
C
Must
May
Must Not



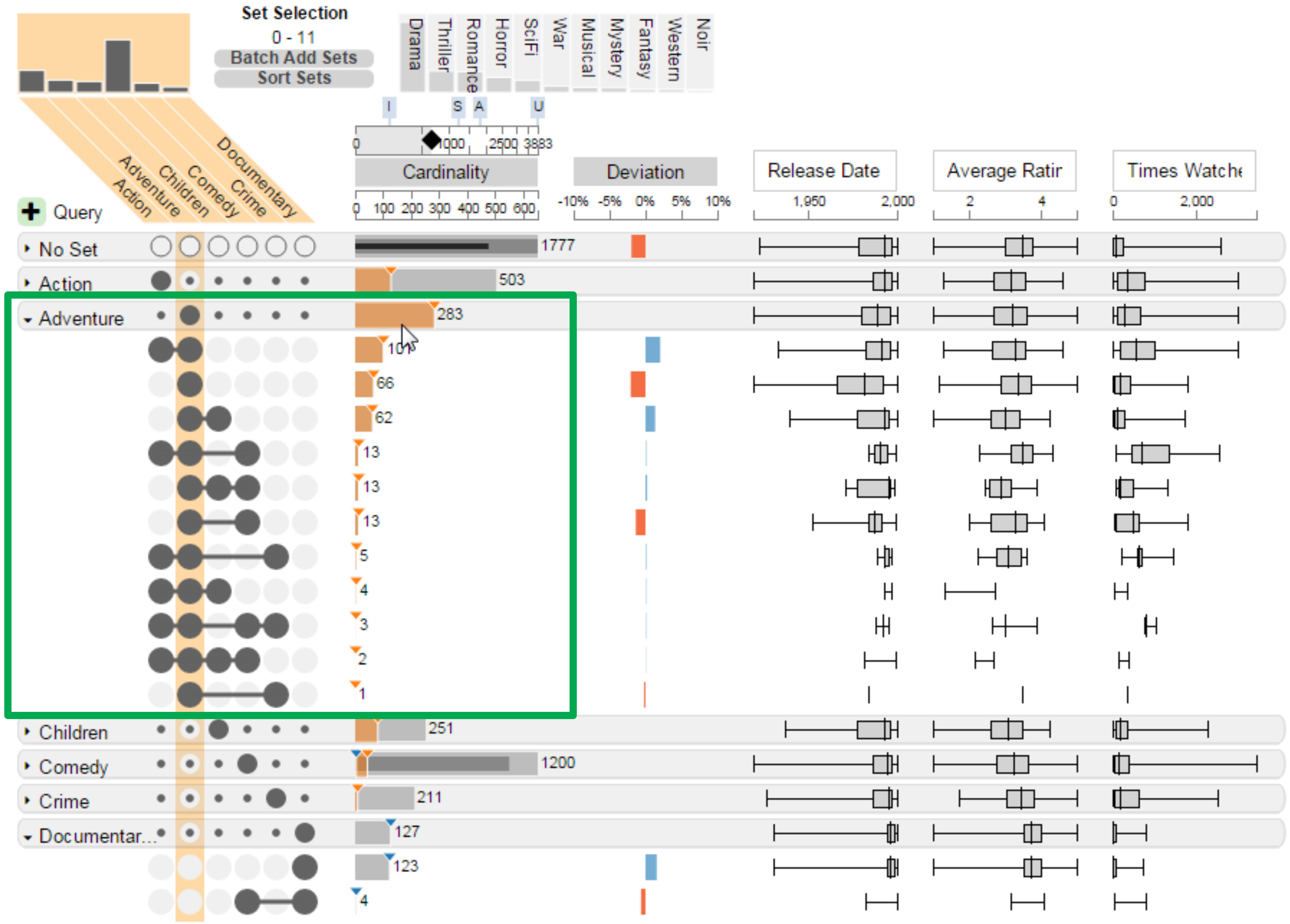
Set Selection
0 - 11
Batch Add Sets
Sort Sets
Drama
Thriller
Romance
Horror
SciFi
War
Musical
Mystery
Fantasy
Western
Noir
Documentary
Crime
Comedy
Children
Adventure
Action
Cardinality
Deviation
Release Date
Query
47
intersections of sets [Action] and [Comedy] but excluding set [Adventure]
not
maybe
must

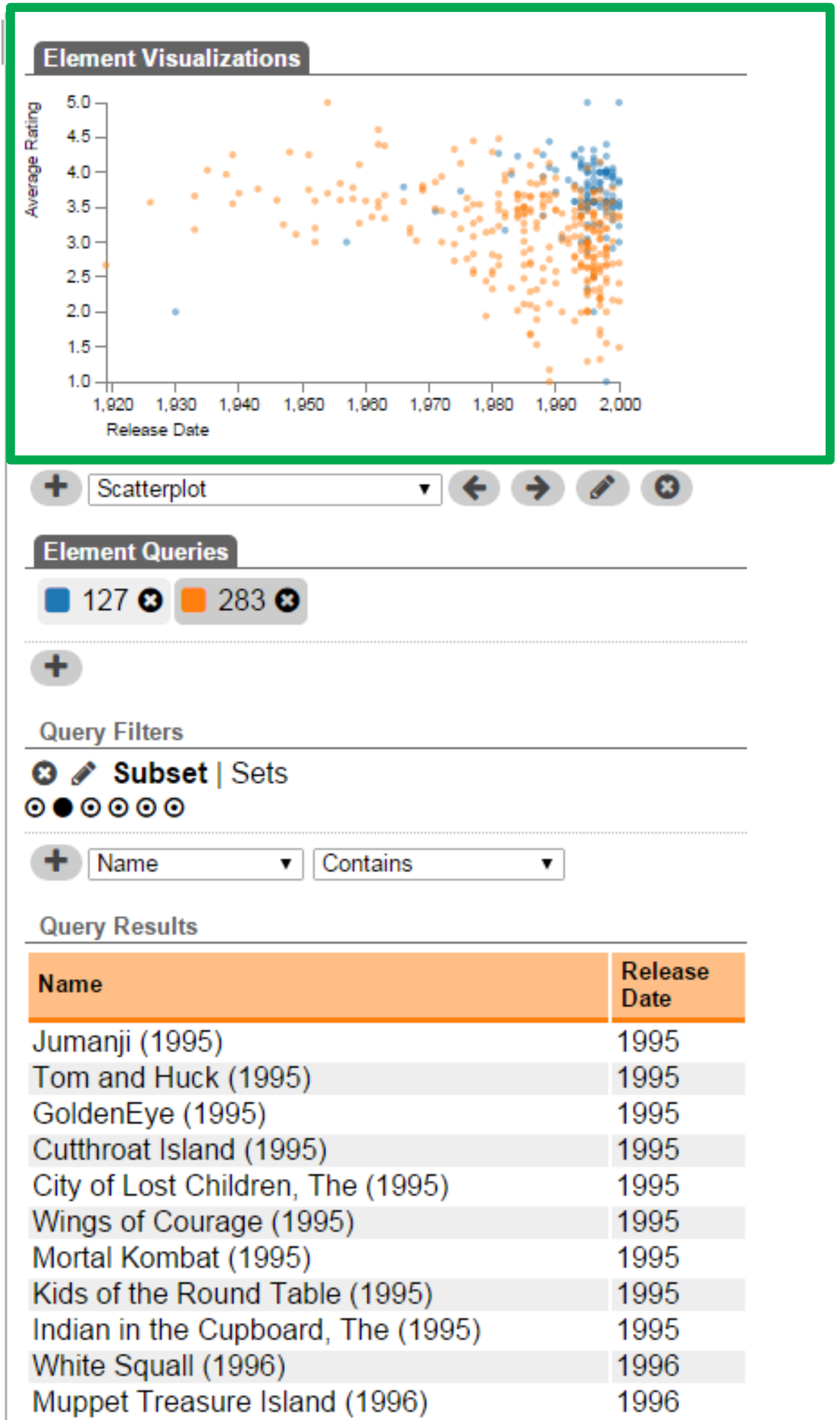


Set Selection
0 - 11
Batch Add Sets
Sort Sets
Drama
Thriller
Romance
Horror
SciFi
War
Musical
Mystery
Fantasy
Western
Noir
Documentary
Crime
Comedy
Children
Adventure
Action
Query
Cardinality
Deviation
Release Date
Average Rating
Times Watched
47
38
9

Elements & Attributes

How do documentaries compare to adventure movies?



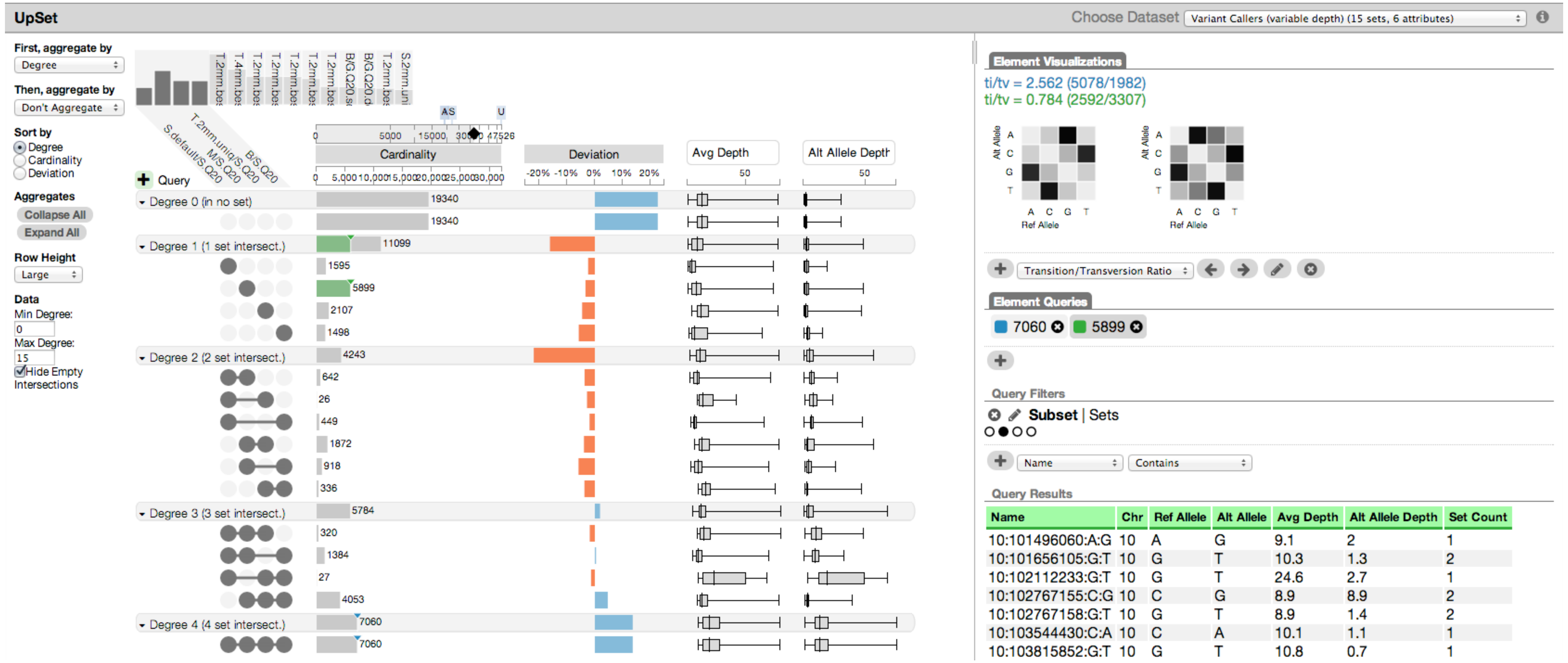


How do documentaries compare to adventure movies?

Applications

Applications

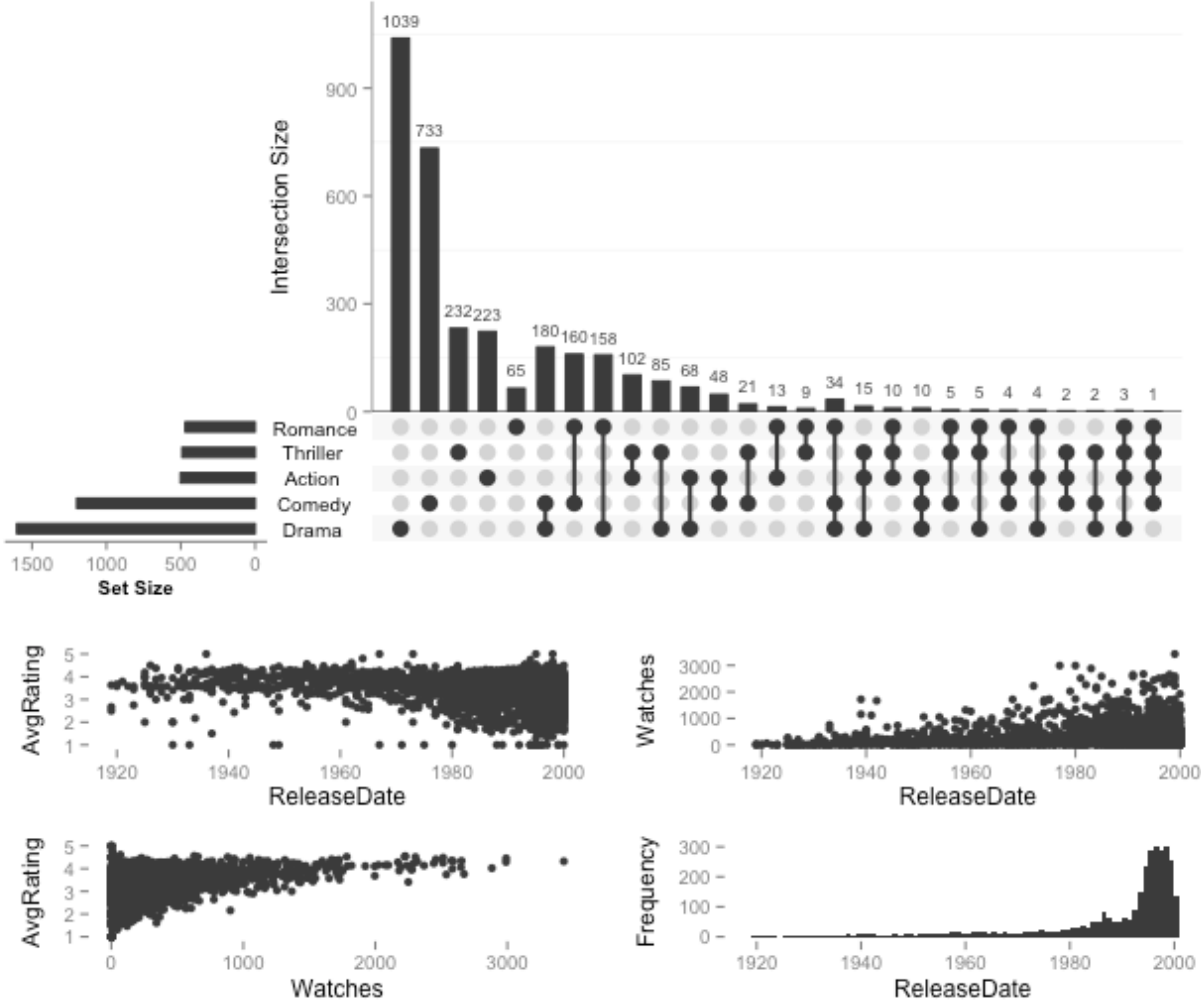
- Genetics
- Economics
- Pharmacology
- Social Networks



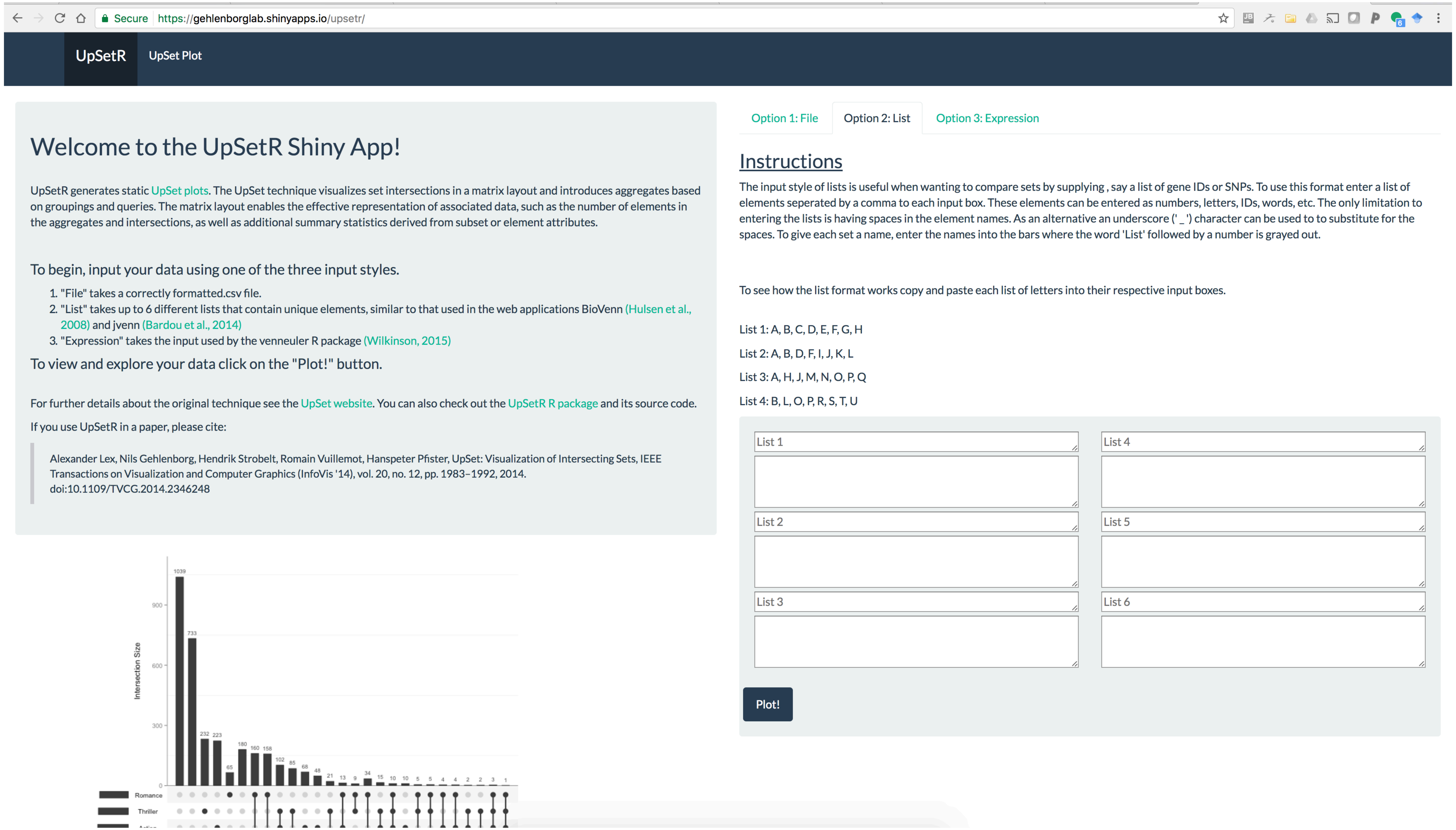
R-Version: UpSetR

Developed at HMS

Some design adaptions



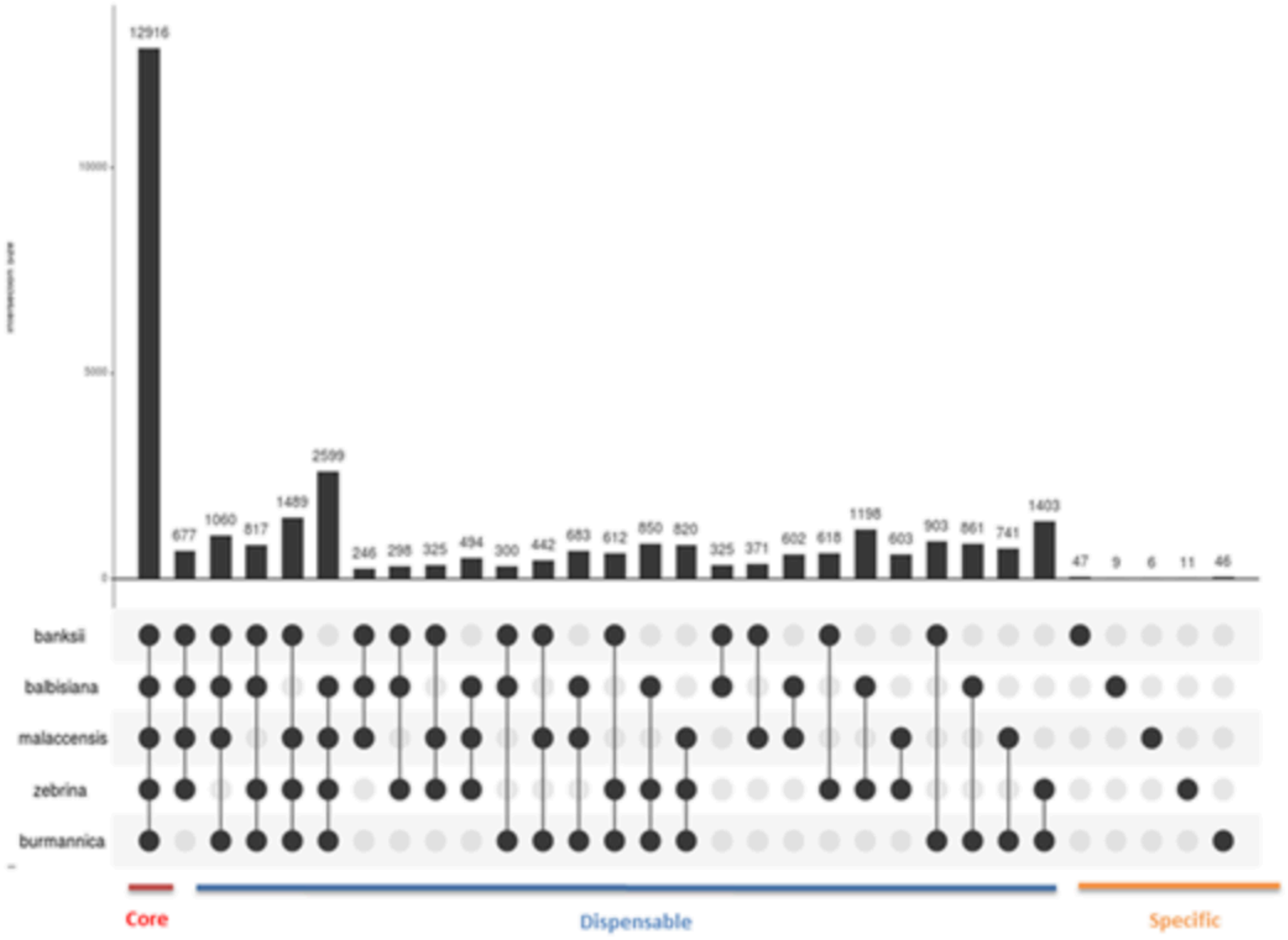
The Banana Chart Redesigned



https://gehlenborglab.shinyapps.io/upsetr/

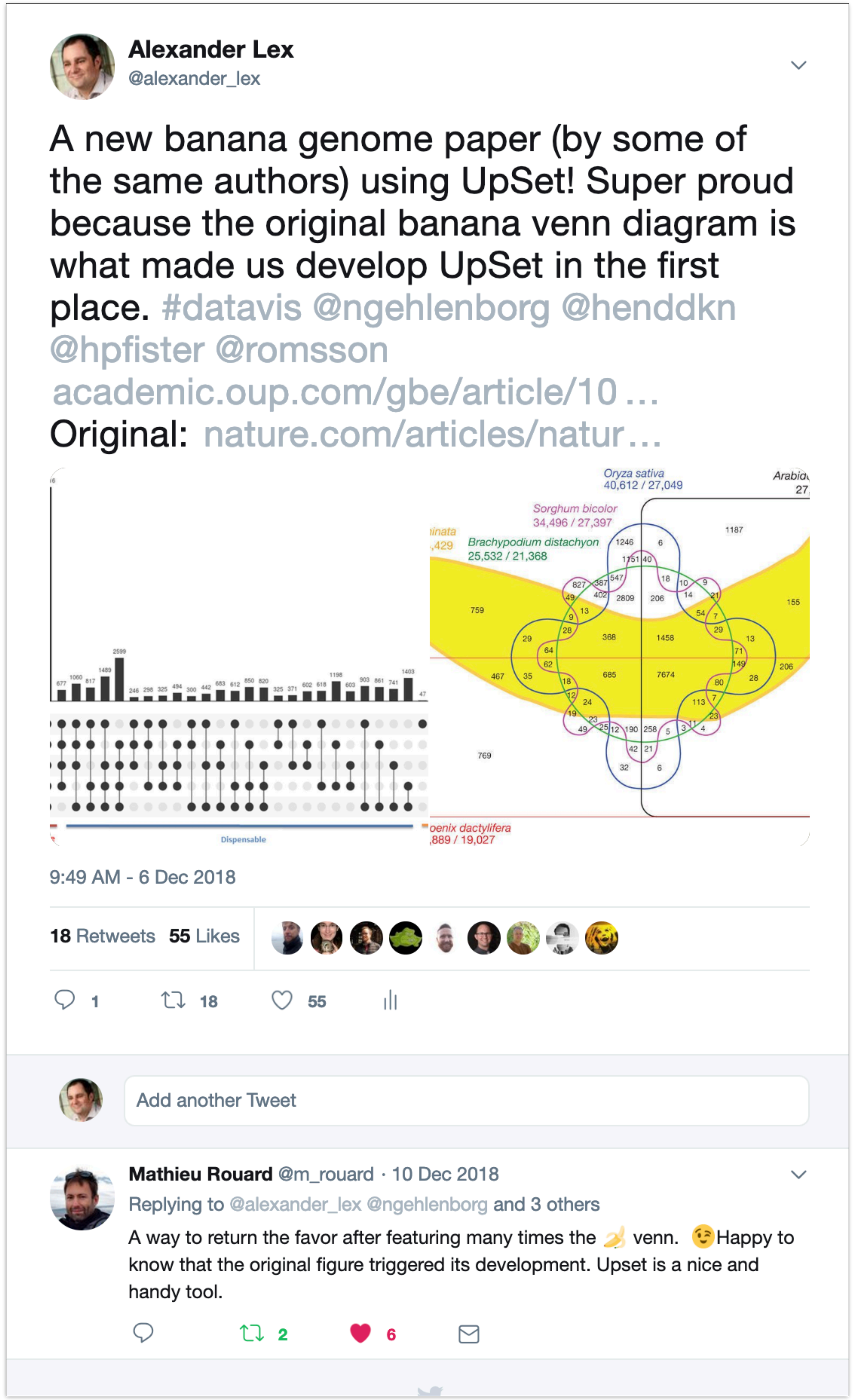
A new Banana Paper

FIG. 1.



View large Download slide

—Intersection diagram showing the distribution of shared gene families (at least two sequences per OG) among *M. a. banksii* "Banksii," *M. a. zebrina* "Maia Oa," *M. a. burmannica* "Calcutta 4," *M. a. malaccensis* "DH Pahang," and *M. balbisiana* "PKW" genomes. The figure was created with UpsetR (Lex et al. 2014).



caleydo.org/tools/upset

[Partl, BioVis '12]
Best Paper Award
[Partl, BMC Bioinf. '13]
[Lex, InfoVis '13]

Pathways

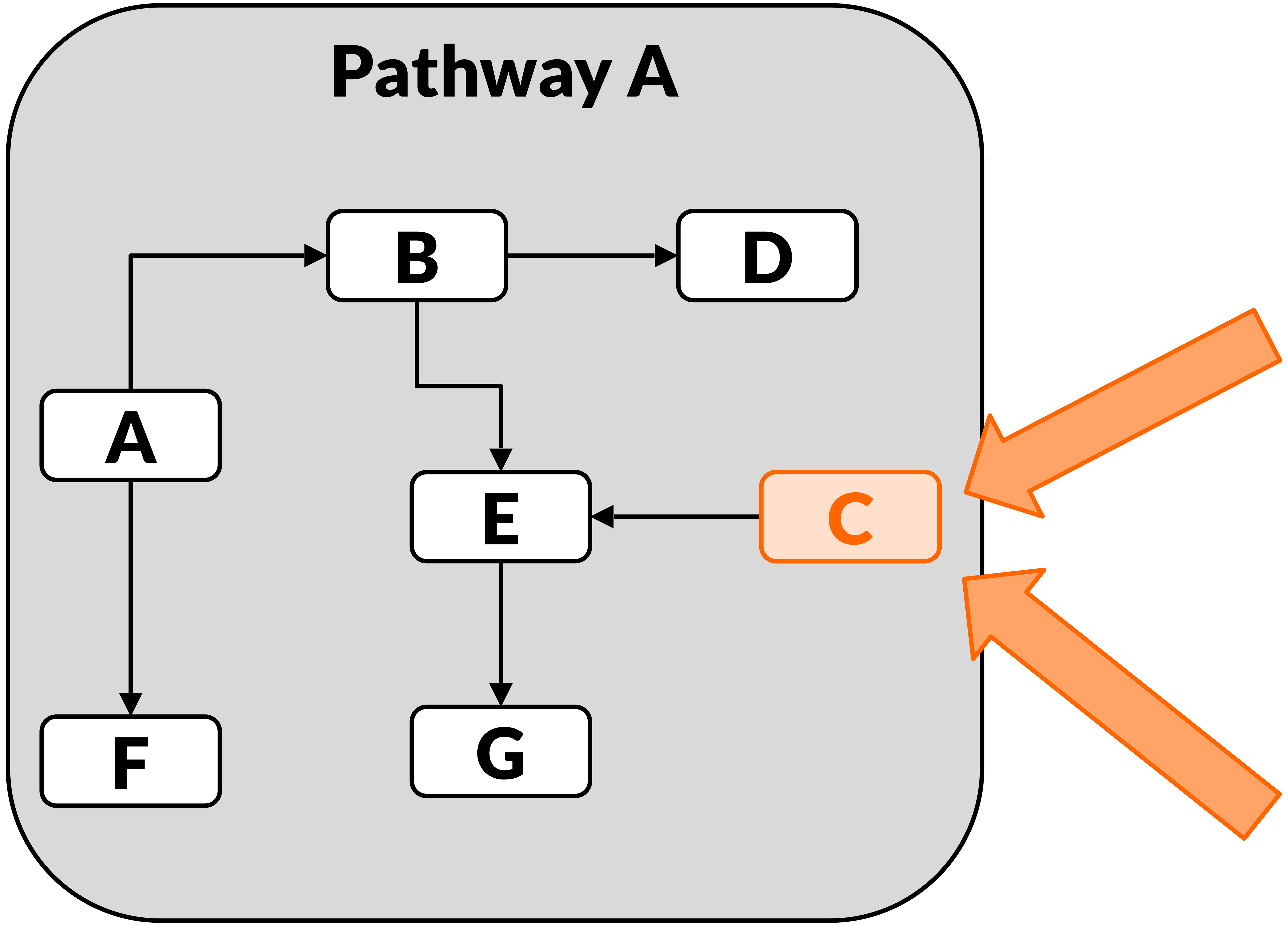
Experi-mental Data and Pathways

Cannot account for variation found in real-world data

Branches can be (in)activated due to

- mutation,
- changed gene expression,
- modulation due to drug treatment,
- etc.

Many Node Attributes

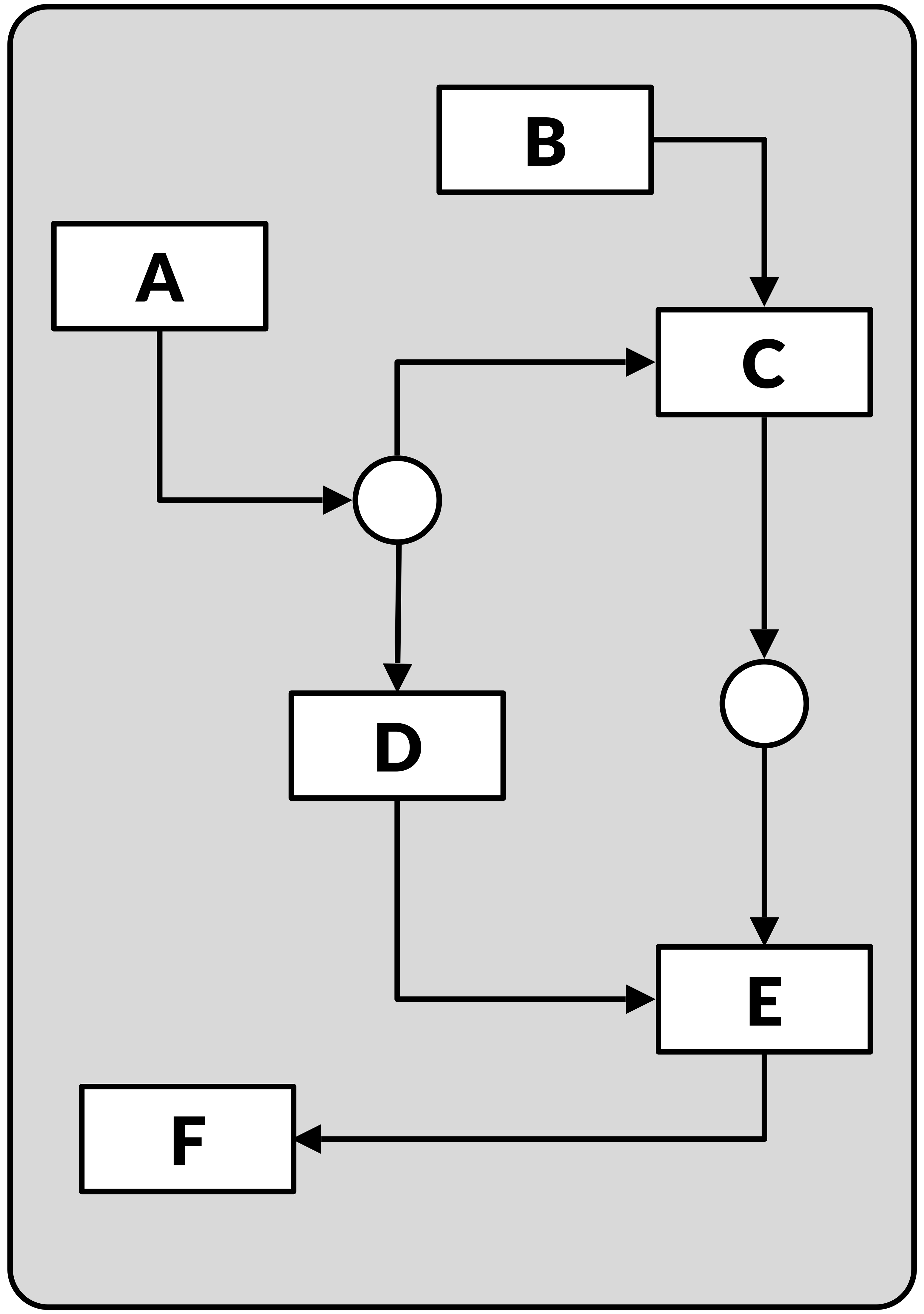


Node	Sample 1	Sample 2	Sample 3	...
A	0.55	0.95	0.83	...
B	0.12	0.42	0.16	...
C	0.33	0.65	0.38	...
...	...	...	...	

Node	Sample 1	Sample 2	Sample 3	...
A	low	low	very high	...
B	normal	low	high	...
C	high	very low	normal	...
...	...	...	...	

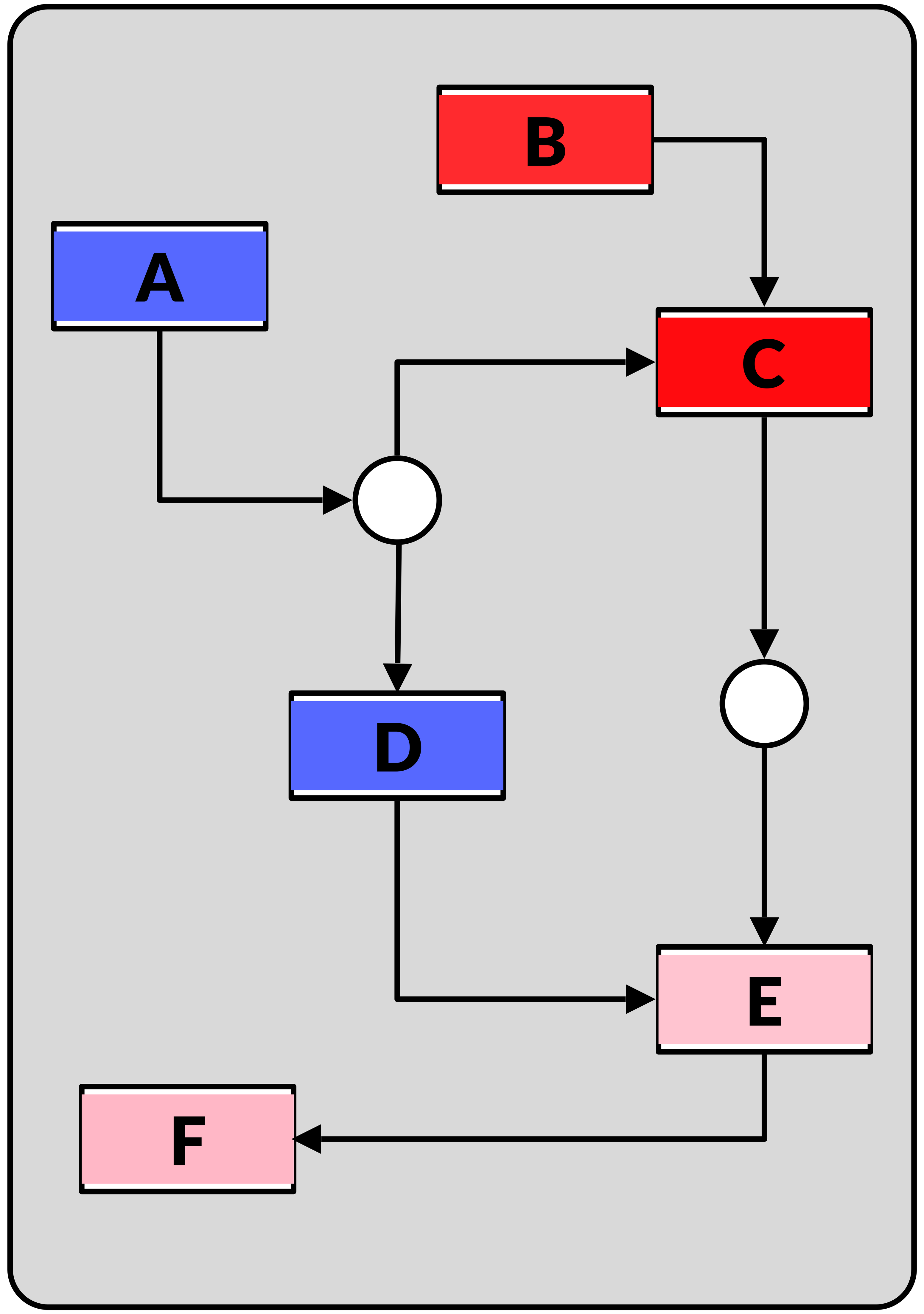
How to visualize experimental data on pathways?

Good Old Color Coding



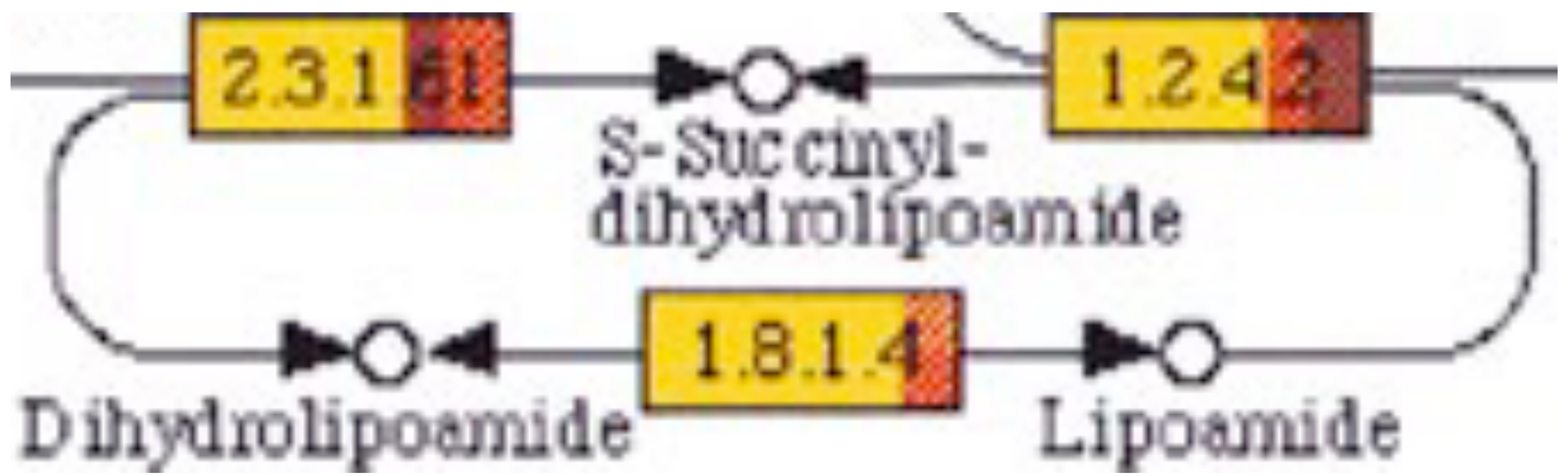
A	-3.4
B	2.8
C	3.1
D	-3
E	0.5
F	0.3

Good Old Color Coding



A	-3.4	4.2	5.1	4.2
B	2.8	1.8	1.3	1.1
C	3.1	-2.2	2.4	2.2
D	-3	-2.8	1.6	1.0
E	0.5	0.3	-1.1	1.3
F	0.3	0.3	1.8	-0.3

[Lindroos2002]

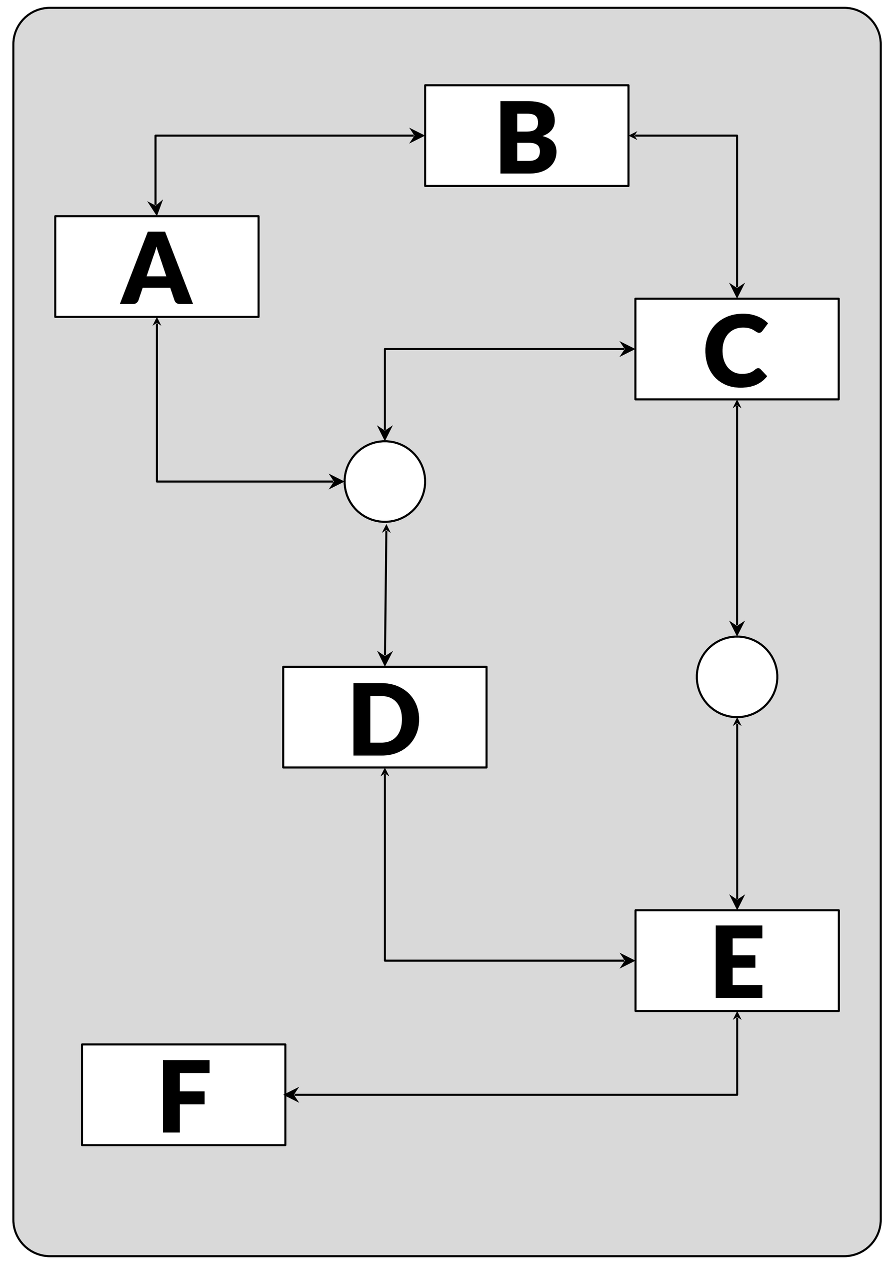


Challenge: Supporting Multiple Tasks

Two central tasks:

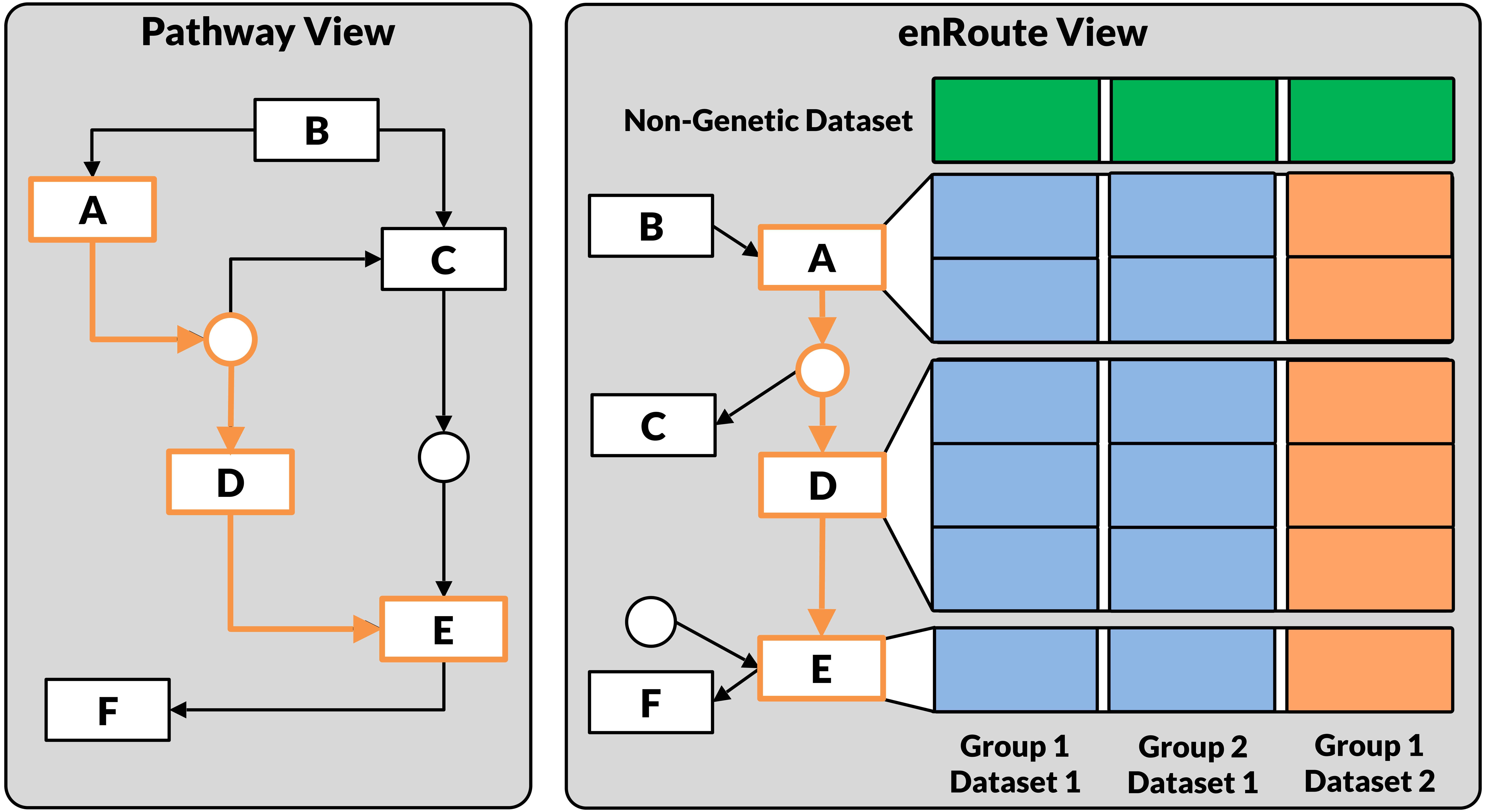
- Explore **topology** of pathway
- Explore the **attributes** of the nodes (experimental data)

Need to support both!



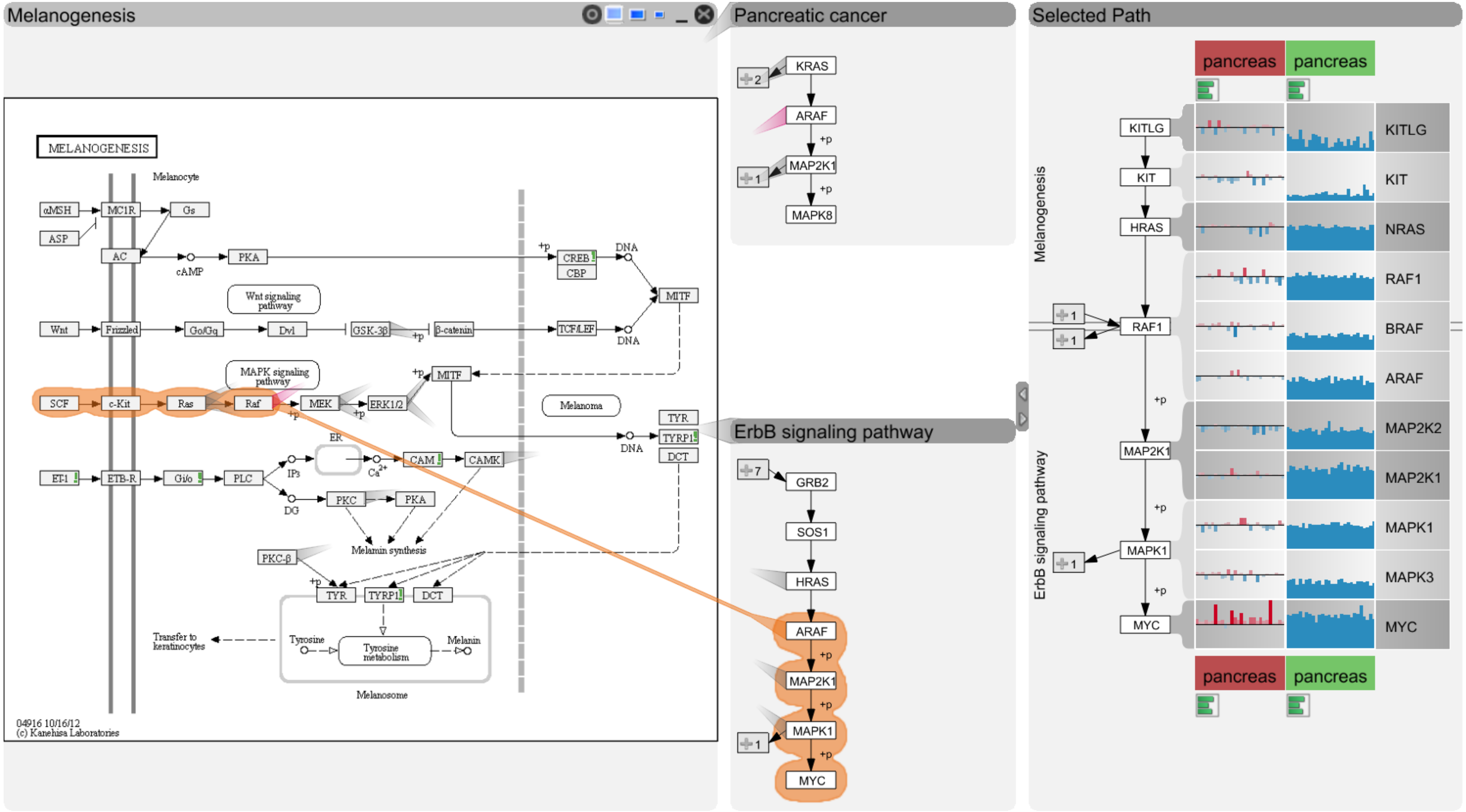
	Sample 1	Sample 2	Sample 3
Gene 1	1	1.1	0.4
Gene 2	2	0.5	1.2
Gene 3	1.4	0.2	0.5
Gene 4	0.3	0.5	0.7

Concept



Pathway View
B
A
C
D
E
F
enRoute View
Non-Genetic Dataset
B
A
C
D
E
F
Group 1
Dataset 1
Group 2
Dataset 1
Group 1
Dataset 2

enRoute



Melanogenesis
Pancreatic cancer
Selected Path
MELANOGENESIS
Melanocyte
αMSH
MC1R
Gs
ASP
AC
cAMP
PKA
CREB
CBP
DNA
MITF
Wnt signaling pathway
Wnt
Frizzled
Go/Gq
Dvl
GSK-3β
β-catenin
TCF/LEF
MAPK signaling pathway
SCF
c-Kit
Ras
Raf
MEK
ERK1/2
Melanoma
TYR
TYRP1
DCT
ER
ET-1
ETB-R
Gi/o
PLC
IP3
CAM
CAMK
DG
PKC
PKA
Melanin synthesis
PKC-β
Transfer to keratinocytes
Tyrosine
Tyrosine metabolism
Melanin
Melanosome
04916 10/16/12
(c) Kanehisa Laboratories
KRAS
ARAF
MAP2K1
MAPK8
ErbB signaling pathway
GRB2
SOS1
HRAS
MAPK1
MYC
pancreas
pancreas
KITLG
KIT
HRAS
RAF1
NRAS
BRAF
MAP2K2
MAPK3
Melanogenesis
ErbB signaling pathway

File Data Window View Help

Entourage

Pathways

Pathway

Filter:
<None>

1 C donor
2-Oxocarboxylic acid
ABC transporters
ABC-family proteins
ACE Inhibitor Pathwa
Acetylcholine Synthes
Acute myeloid leukem
Adherens junction
Adipocyte TarBase
Adipocytokine signali
Adipogenesis
Advanced glycosylatio
Aflatoxin B1 metaboli
African trypanosomias
AGE/RAGE pathway
AhR pathway
Alanine and aspartate
Alanine, aspartate an
Alcoholism
Aldosterone-regulated
Allograft rejection
Allograft rejection
Alpha 6 Beta 4 signal
alpha-Linolenic acid
Alzheimer's disease
Alzheimers Disease
amino acid conjugatio
amino acid conjugatio
Amino sugar and nucl
Aminoacyl-tRNA bios
Amoebiasis
Amphetamine addicti
AMPK signaling
Amyotrophic lateral sc
Androgen receptor si
Angiogenesis
Angiogenesis
angiogenesis overvie
Antigen processing an
APC/C-mediated degra
Apoptosis
Apoptosis
Apoptosis Meta Path
Apoptosis Modulation
Apoptosis Modulation
Apoptosis, anoikis an

Selected Path

Toggle summary button
Summary of all dimensions
All dimensions shown separately
Not Mutated
Mutated
Not Mutated
Mutated
Race
Days to Deat
Gender
Context (non-genetic) datasets
EGF
EGFR
PLCG1
PLCG2
PRKCB1
PRKCA
PRKCG
NRAS
Dimensions sorted by this row
Glioma
EGF
EGFR
+p
PLCG1
PRKCA
HRAS
ARAF
PIK3R5
Multi-mapping of gene family
Collapsed incoming node
Expanded outgoing nodes
Not Mutated
Mutated
Not Mutated
Mutated
Groups of same dataset
Different Datsets

Case Study: CCLE Data

Pancreatic cancer
Non-small cell lung cancer
Proliferation
Survival
Angiogenesis
Adhesion
Migration / Invasion
Differentiation
Glioma
DNA
Jun
Elk
Myc
ERBB2
GRB2
SOS1
HRAS
ARAF
MAP2K1
MAPK1
CCND1
PIK3R5
AKT3
FOXO3
CALML6
CAMK2A
ErbB signaling pathway
SHC2
Erlotinib
Lapatinib
SHC4
SHC3
SHC1
SOS2
NRAS
breast

http://
entourage.caleydo.org

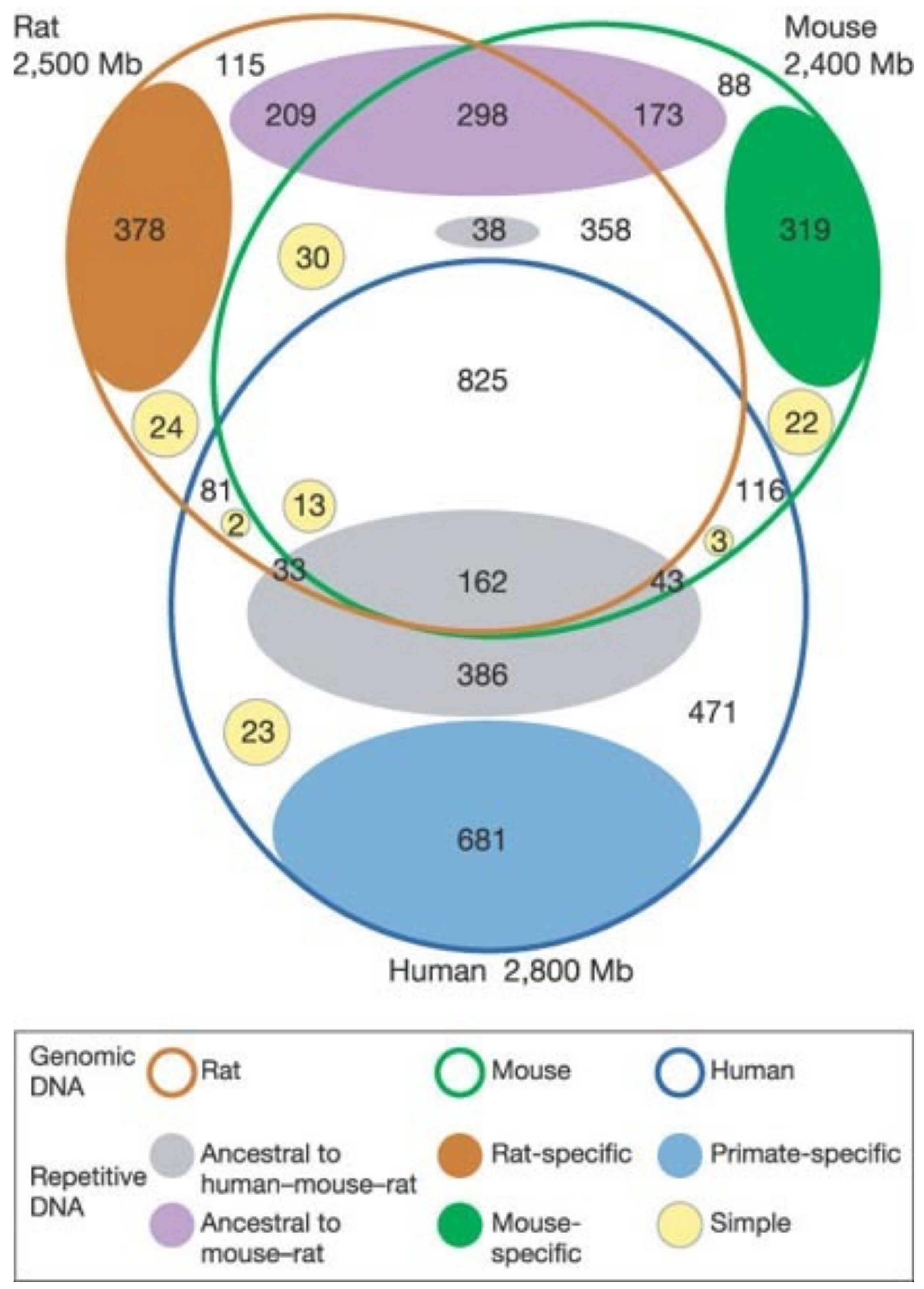
Visualization Design Strategies

1. Encoding channel primacy
2. Show relationships explicitly
3. Use queries
4. User color sparingly
5. Enable annotation / provenance

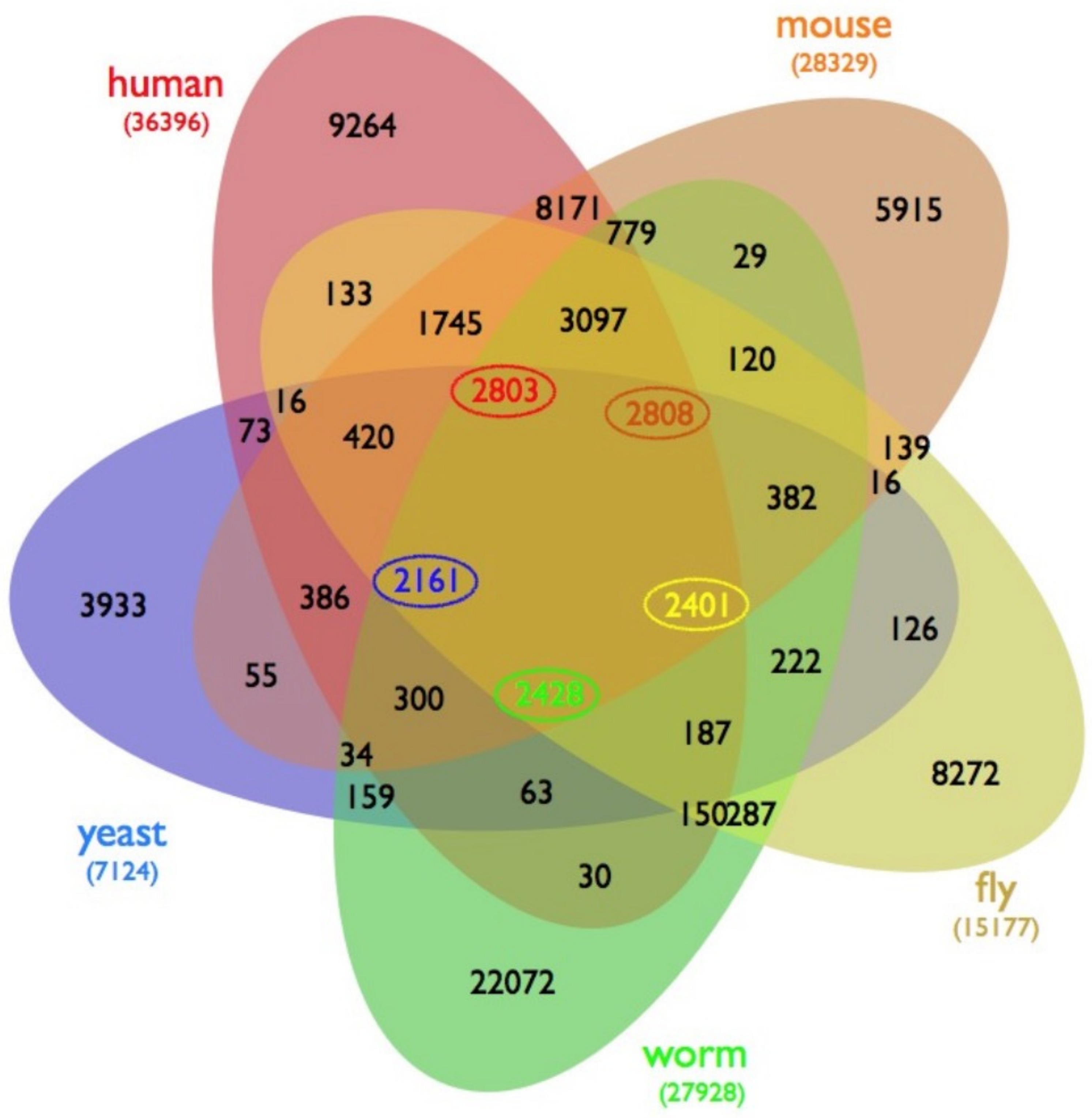
1. Encoding Channel Primacy

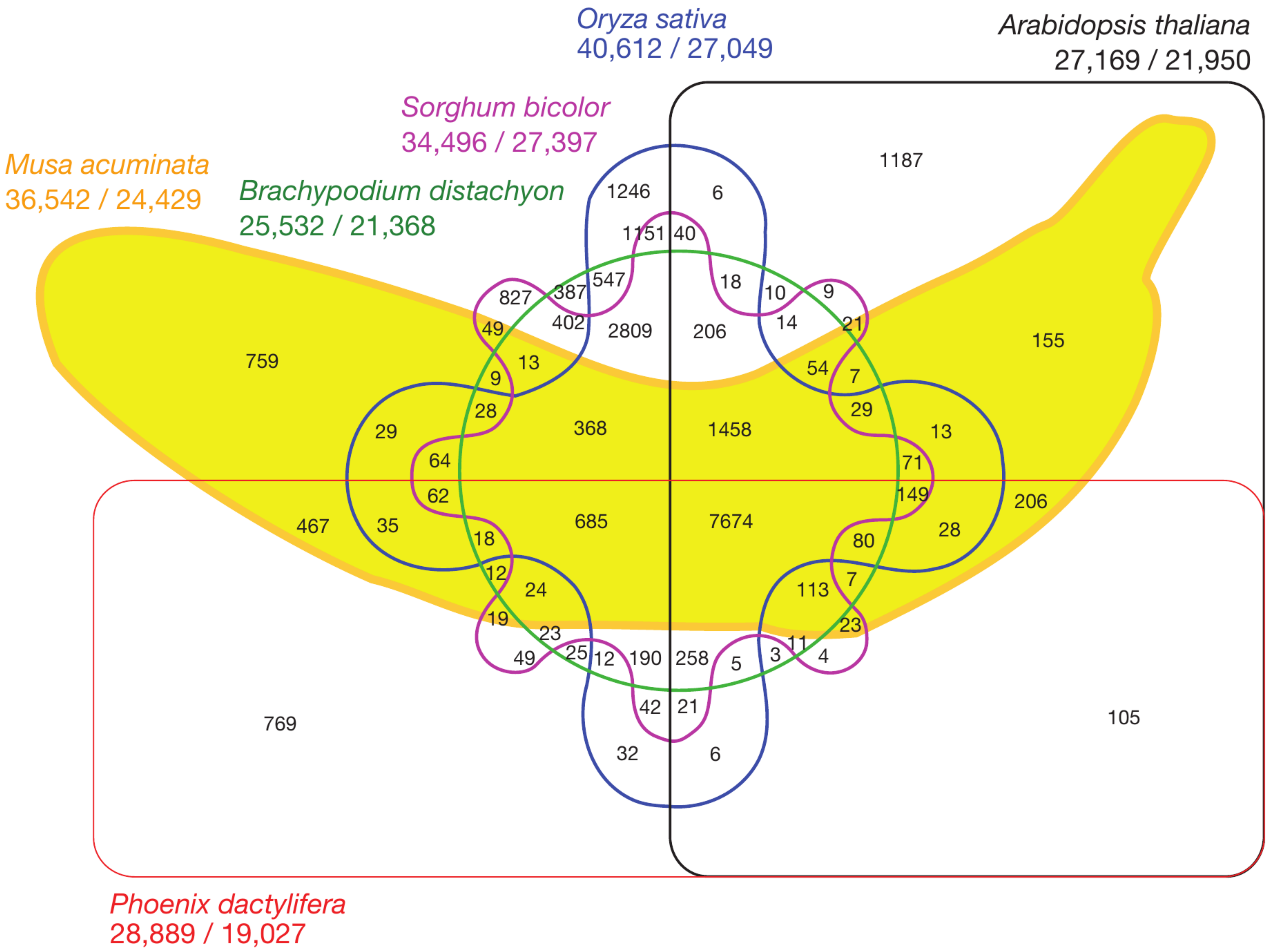
Most important data is assigned most powerful encoding channel (position)

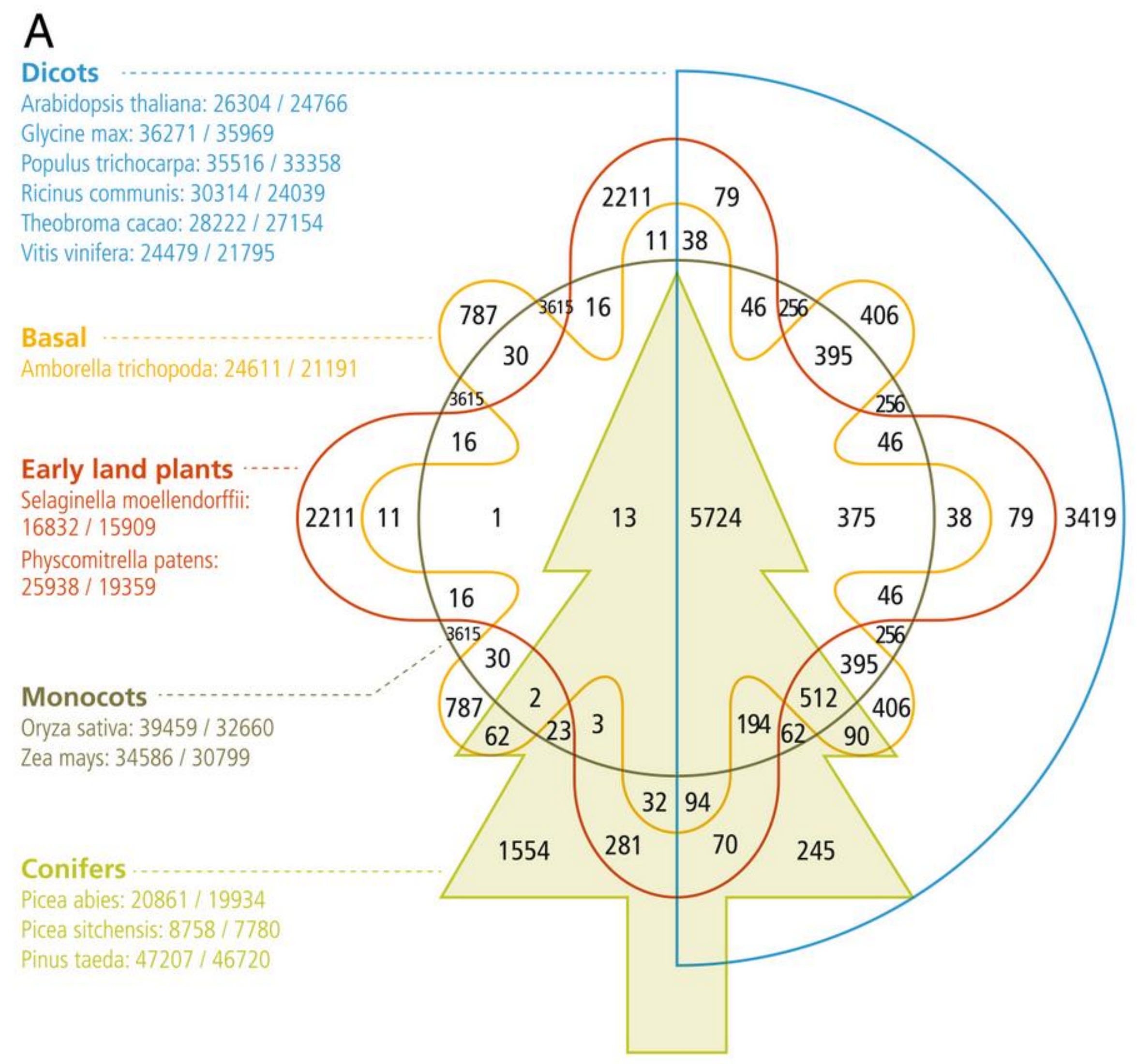
Example: Set Visualization



[Gibbs et al., Nature, 2004]







[Wiles et al., BMC Systems Biology]

[D'Hont et al., Nature, 2012]

[Neale et al., BMC Genome Biology, 2014]

The Banana Chart Redesigned

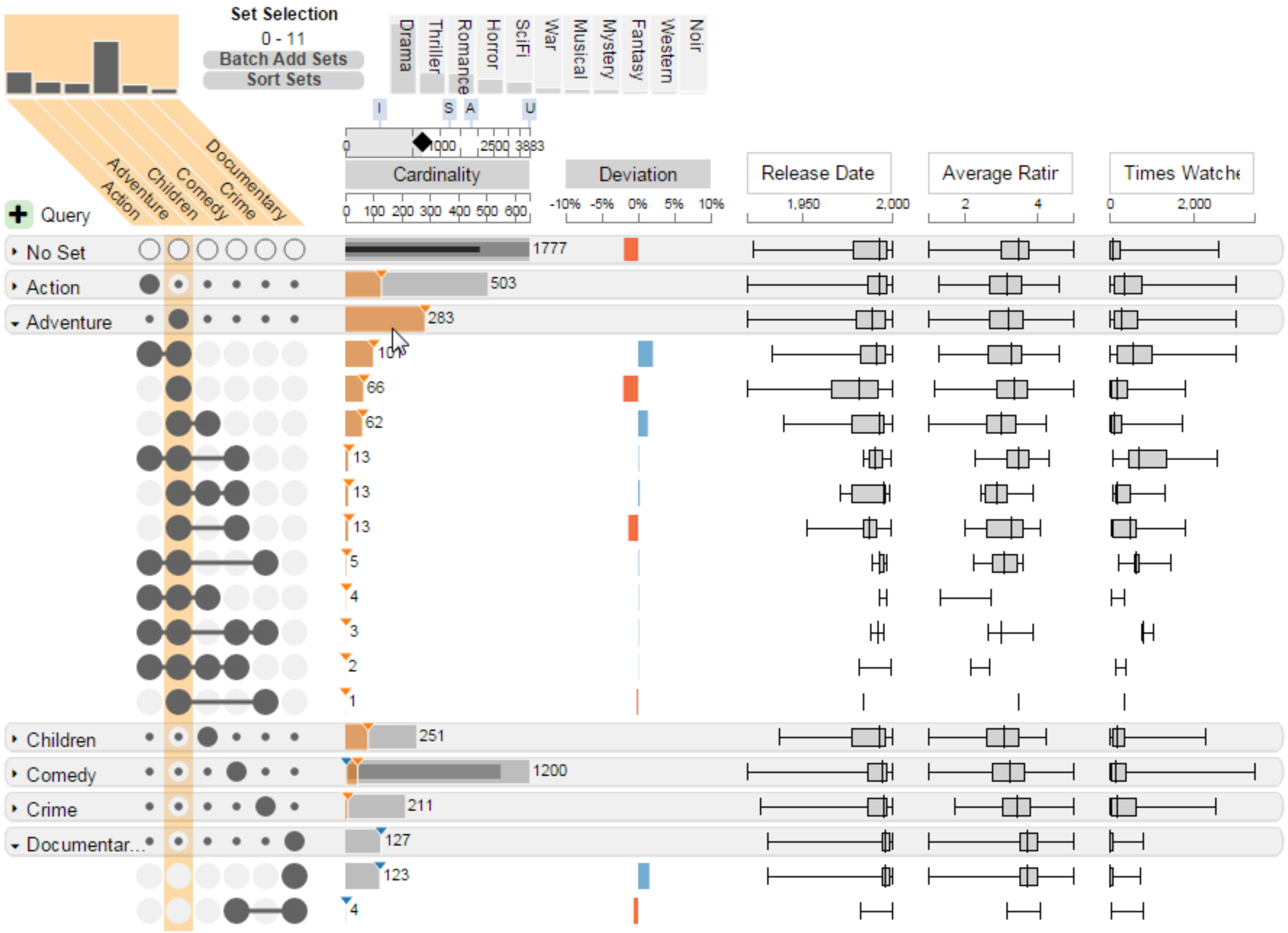
2. Show relationships explicitly

Dont use highlighting to connect different views

Use smart layouts (position)

or connectivity

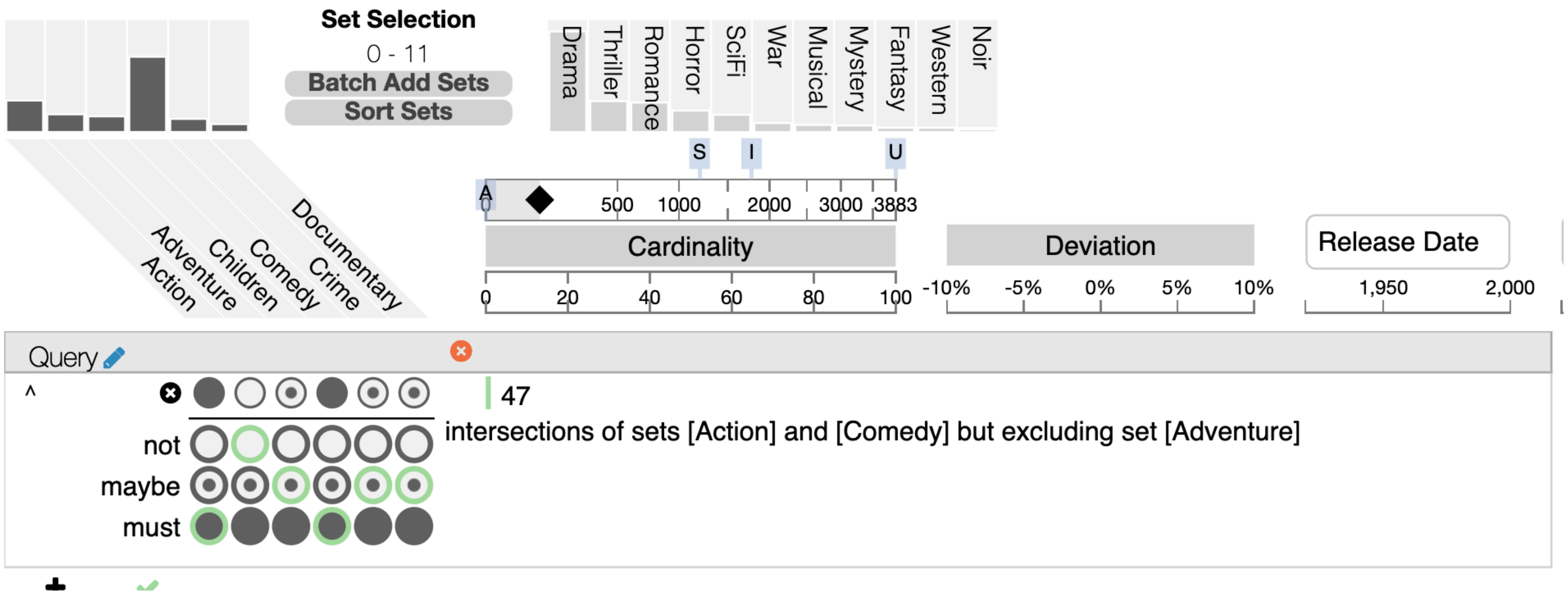
FamilyID
People
#POI
38 121 12
149 113 10
27251 404 39
42623 81 10
68939 244 23
176860 426 44
603481 181 19
791533 114 10
903988 58 5
Data Selection
KindredID
RelativeID
sex
M
F
deceased
Y
N
suicide
POI
N
Y
Depression
A
F
T
Age1D_Depression
0
64
Nr.Diag_Depression
Aggregate All
Hide All
Expand All
1824
1840
1860
1880
1900
1920
1940
1960
1980
2000
2017
Aggregate
Hide
23
KindredID
RelativeID
sex (F)
deceased
suicide
Depression (True)
Age1D_Depression
Nr.Diag_Depression
Bipolar (True)
Age1D_Bipolar
Nr.Diag_Bipolar
MaxBMI
AgeMaxBMI
race (White)
race (Other)
race (Black)
race (Asian)
race (Hispanic)
cause_death
44%
60%
10%
52%
48%
98%
0.0%
Total:32
Total:60
Total:30
Total:59
Total:58
#13789
#48154
#54315
#67703
#9278
#26695
#40329
#5170
#43983
#5181
#42479
#139
#1967
#5172
#5171
#10841
#66710
#6436
#65130
#65129
#6428
#4547
#65491
#39045
#18524
#4556
#4551
#62034
#62036
#17581
#47083
#47087
#61319
#61318
asphyxia
carbon monox...
cardiac arre...
strangulatio...



Set Selection
0 - 11
Batch Add Sets
Sort Sets
Drama
Thriller
Romance
Horror
SciFi
War
Musical
Mystery
Fantasy
Western
Noir
I
S
A
U
0
1000
2500
3883
Documentary
Crime
Comedy
Children
Adventure
Action
Query
Cardinality
Deviation
Release Date
Average Ratir
Times Watche
0 100 200 300 400 500 600
-10% -5% 0% 5% 10%
1,950
2,000
2
4
0
2,000
No Set
1777
Action
503
Adventure
283
101
66
62
13
13
13
5
4
3
2
1
Children
251
Comedy
1200
Crime
211
Documentar...
127
123
4

3. Use queries

especially for big data



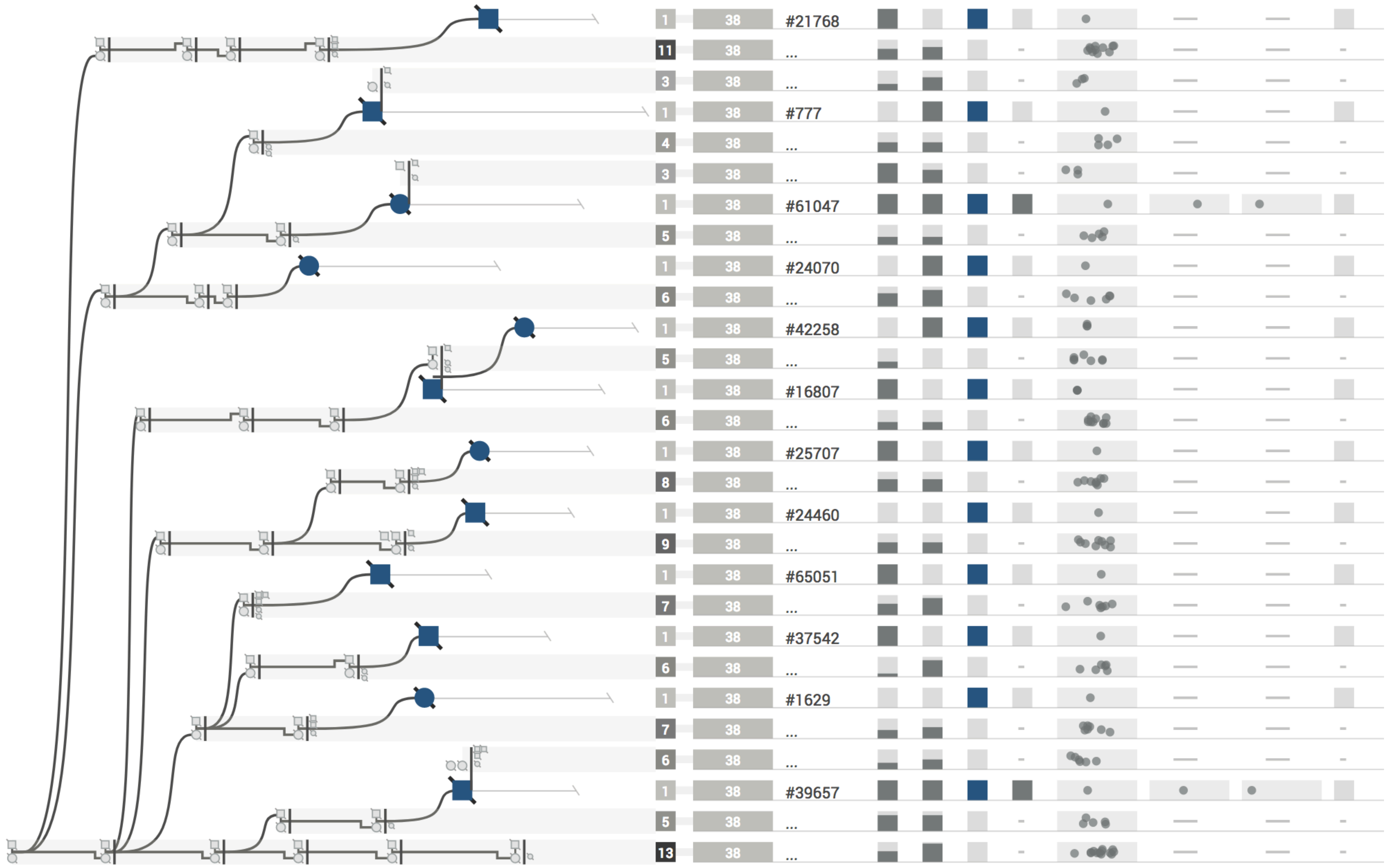
Set Selection
0 - 11
Batch Add Sets
Sort Sets
Drama
Thriller
Romance
Horror
SciFi
War
Musical
Mystery
Fantasy
Western
Noir
Action
Adventure
Children
Comedy
Crime
Documentary
S
I
U
A
0
500
1000
2000
3000
3883
Cardinality
0
20
40
60
80
100
Deviation
-10%
-5%
0%
5%
10%
Release Date
1,950
2,000
Query
47
intersections of sets [Action] and [Comedy] but excluding set [Adventure]
not
maybe
must

4. User color sparingly

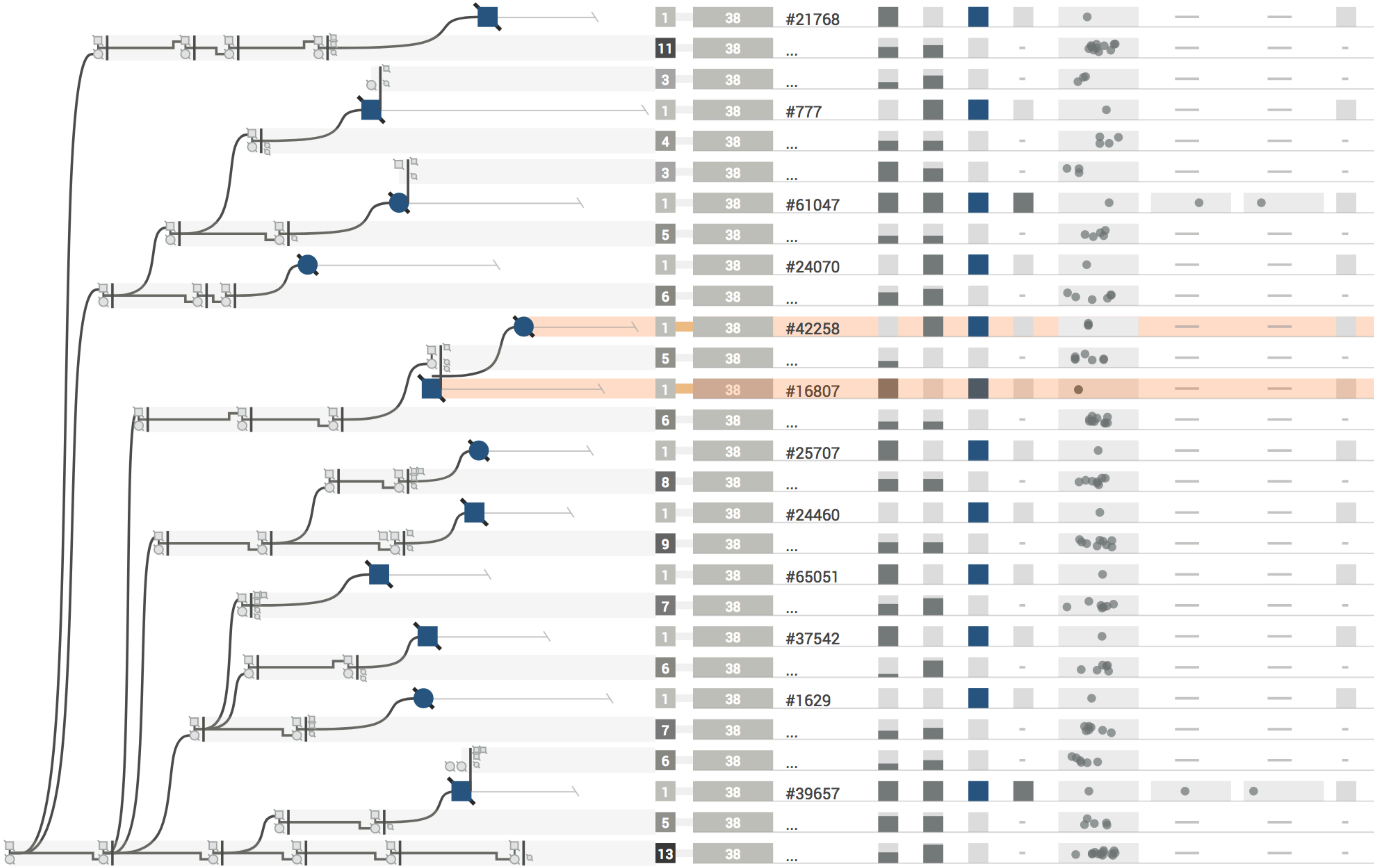
Limit use to encode data

Primarily use it to highlight items of interest

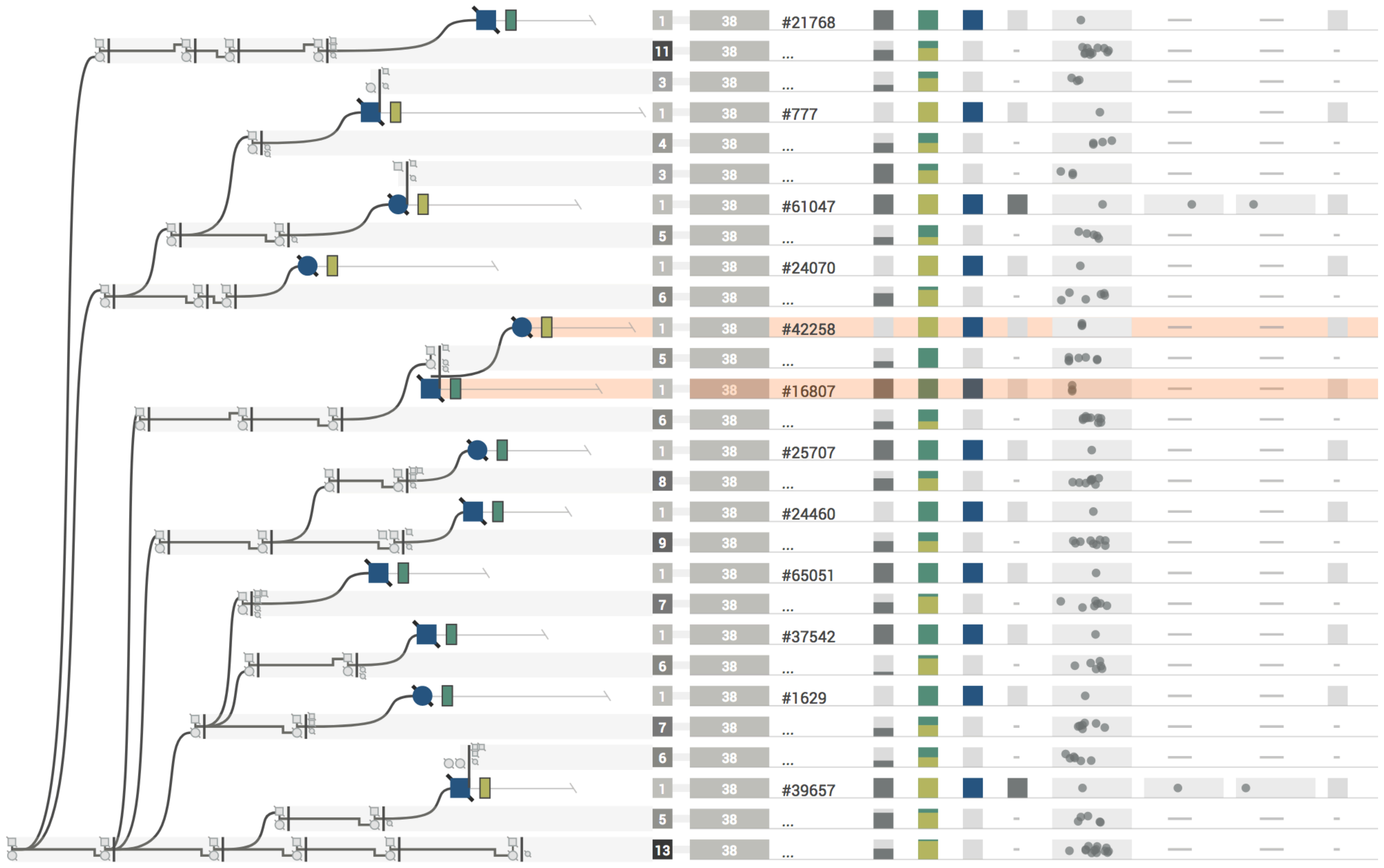
Pop out effect!



Only one color for primary attribute (suicide, blue)



Another color for highlights, to emphasize parent-child relationship (orange)



Adding a color for an additional attribute (deceased yes/no, green)

5. Enable annotation / provenance

What did you see / think when looking at this visualization?

How did I get here? Can I go back?

StratomeX.js
Exploration
Authoring
Presentation
admin
Annotations
patient.daysto...
mRNA-seq (4 TCGA S...
microRNA-seq (4 TCG...
a
b
c
d
m1
m2
m3
m4
mi1
mi2
mi3
mi4
Kaplan-Meier Plot
showing survival of patients
in connected block
Provenance
86 TCGA_SAMPLEs
62 TCGA_SAMPLEs
32 TCGA_SAMPLEs
no TCGA_SAMPLEs
147 TCGA_SAMPLEs
patient.daystodeath (4 TCGA
patient.daystodeath (4 TCGA
patient.daystodeath (4 TCGA
Data
Visual
Selections
Layout
Analysis
Story
microRNA-seq (4 TCGA Subtypes)
Select Bands 1
Select Bands 2
Select Bands 3
Select Bands 4
Select Bands 5
Viewing Stratified Patient Survival
147 TCGA_SAMPLEs
Days to Death⇄mRNA-seq (4 TCGA Subtypes)
Kaplan Maier Plot
Adding a Categorical Parameter
Days to Death⇄microRNA-seq (4 TCGA Subtypes)
Days to Death⇄Clinical (Overall Stage)

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

THE UNIVERSITY OF UTAH

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

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

HEALTH UNIVERSITY OF UTAH

NSF

NIH

DEPARTMENT OF DEFENSE UNITED STATES OF AMERICA

visualization design lab

SCI www.sci.utah.edu