Alexander Lex

@alexander_lex http://alexander-lex.net



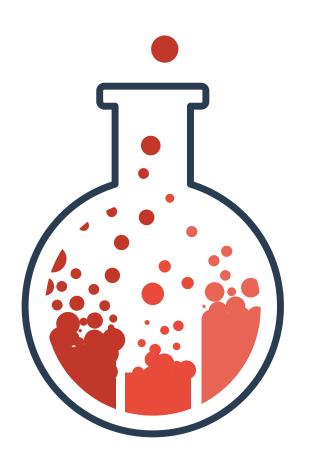


Visualization design lab









Visualization design lab



Carolina Nobre

Alex Bigelow





Jimmy Moore Kiran Gadhave Sam Quinan Alexander Lex Nina McCurdy Jennifer Rogers Haihan Lin Ethan Kerzner

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



Miriah Meyer



Ilkin Safarli

C ŋ

datavisyn

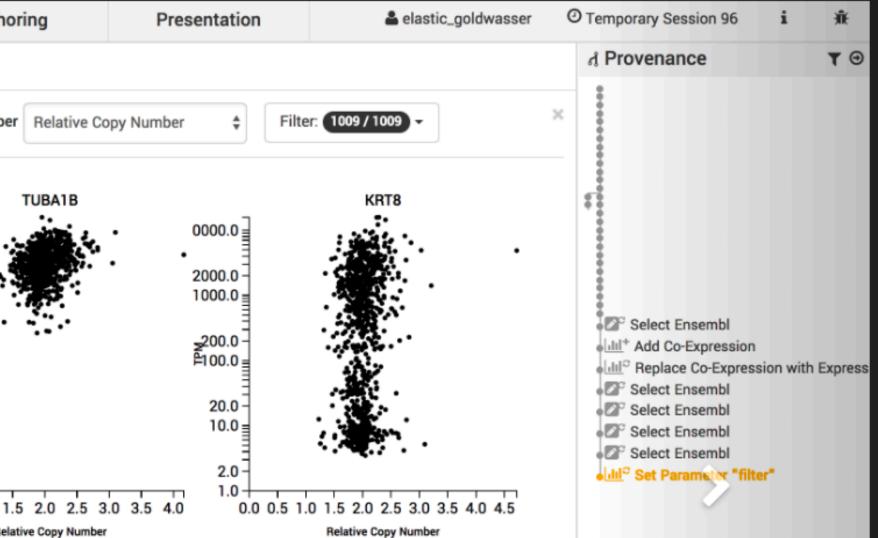
Our Target Discovery Platfrom is a web-based visual data analysis solution designed to score, rank, filter and visualize datasets that provides all the data and visualizations needed to identify analysis targets.

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

PRODUCTS

TARGET DISCOVERY PLATFORM

💲 Orc	lino					E	xplora	tion		Autho
*	Genes	Expression vs	s. Copy Number							
		×	General	Data Source	Call Line	Everees				Conv Number
			Database Info	Data Source	Cell Line	Express		PM	\$	Copy Number
Rank	S	Symbol	Sample overview							
			Copy Number	Е		MYL6			Ŧ	
1	GA GA	PDH L41	Expression	2000.0				. 2000		
3		BA1A	Mutation	1000.0				1000	.0	
4	MY	L6 LP0	Mutation	200.0			•	200	E	
	O UB		Combined View	200.0 -		:		¥100		
7		BA1B	OncoPrint					¥.	1	
8				20.0 -					.0-	
10			Visualization	10.0				10	.0	
11	KR		Co-Expression						=	
12	O NA	GA	OU Expression	2.0 -				2	-0.	
13		18	Expression vs. Copy Number	1.0+					+0.	
14	О ТР	1	,	0.0		.5 2.0 2.5		3.5	0.0	
		10			Relat	ive Copy Numbe	er 🛛			Rela





http://dataviscourse.net/professional/

Resources Home



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:





THE UNIVERSITY OF UTAH

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









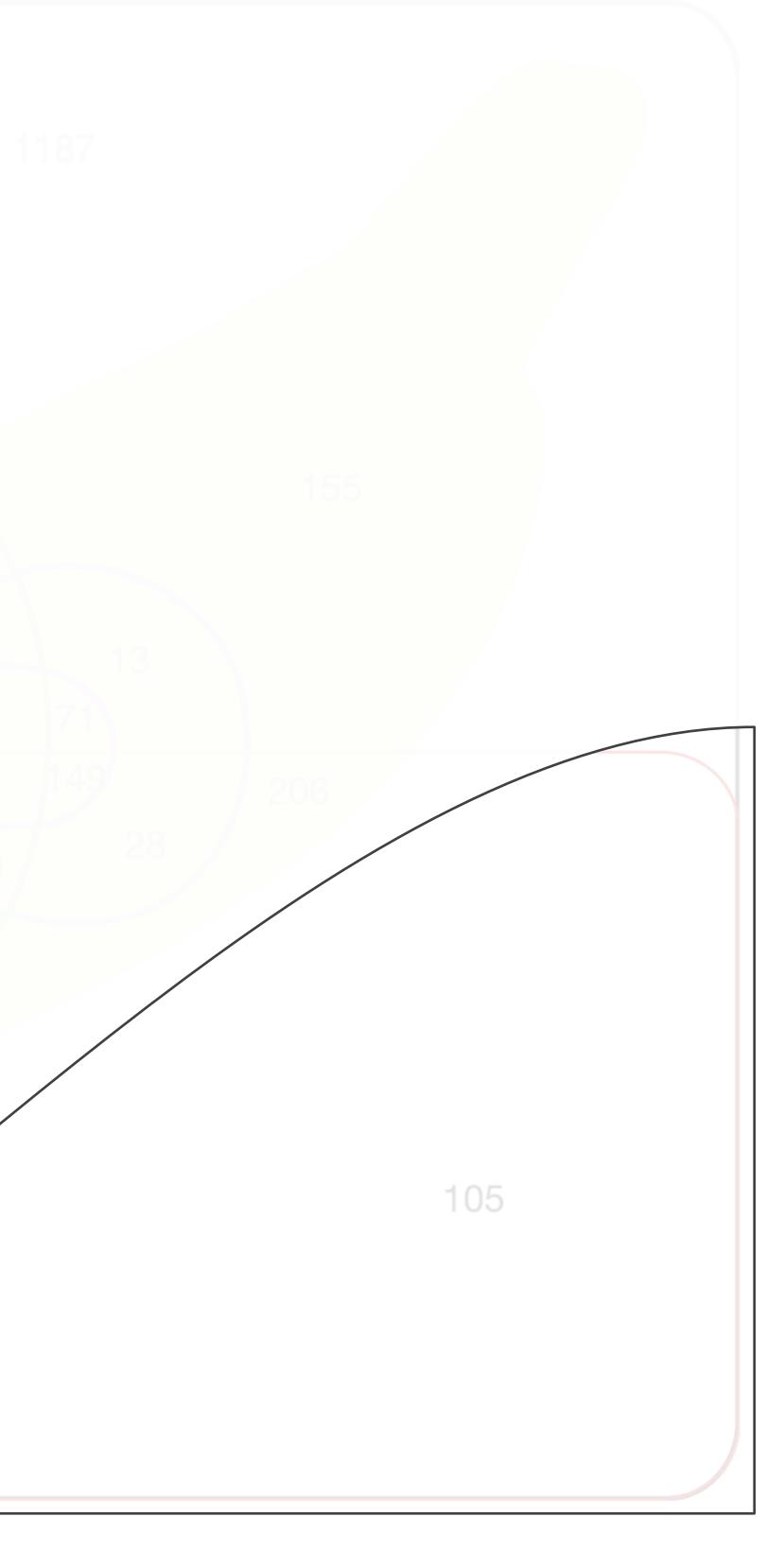
- Card, Mackinlay, Shneiderman - Richard Wesley Hamming

Banana Date Cress Rice Brome

M. acuminata P. dactylifera Arabidopsis thaliana Oryza sativa Sorghum Sorghum bicolor Brachypodium distachyon



Brachypodium distachyon / 1246







VESUALEZAELOM

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

Can We Trust Statistics?

	I	
Х		У
	10	8.(
	8	6.9
	13	7.5
	9	8.8
	11	8.3
	14	9.9
	6	7.2
	4	4.2
	12	10.8
		4.8

II

III

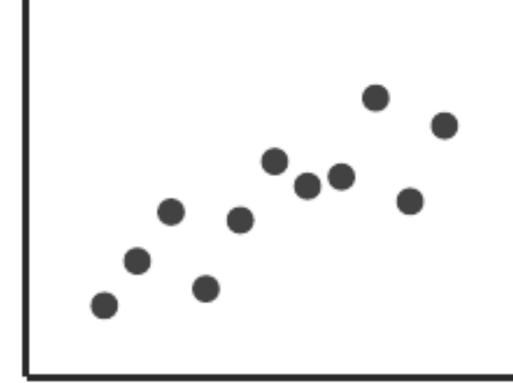
	У	X	У	X	У
10	8.04	10	9.14	10	7.46
8	6.95	8	8.14	8	6.77
13	7.58	13	8.74	13	12.74
9	8.81	9	8.77	9	7.11
11	8.33	11	9.26	11	7.81
14	9.96	14	8.1	14	8.84
6	7.24	6	6.13	6	6.08
4	4.26	4	3.1	4	5.39
12	10.84	12	9.13	12	8.15
7	4.82	7	7.26	7	6 1 2
5	5. M	ean x: 9	y: 7.5	0	

Variance x: 11 y: 4.122 **Correlation x - y: 0.816** Linear regression: y = 3.00 + 0.500x

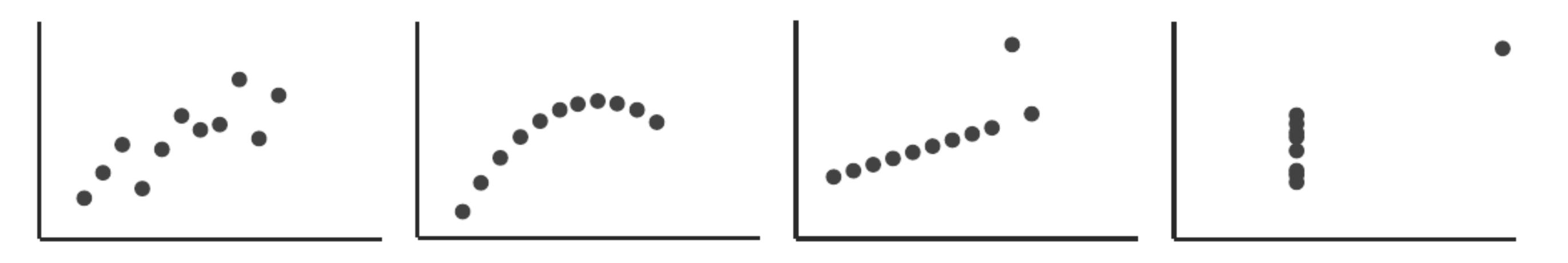
IV

X		У
8	8 6	5.58
8	3 5	.76
8	37	.71
8	8	8.84
8	8	8.47
8	37	.04
8	3 5	5.25
19) 1	2.5
8	3 5	5.56
2	ş 7	.91
	\$ 6	5.89



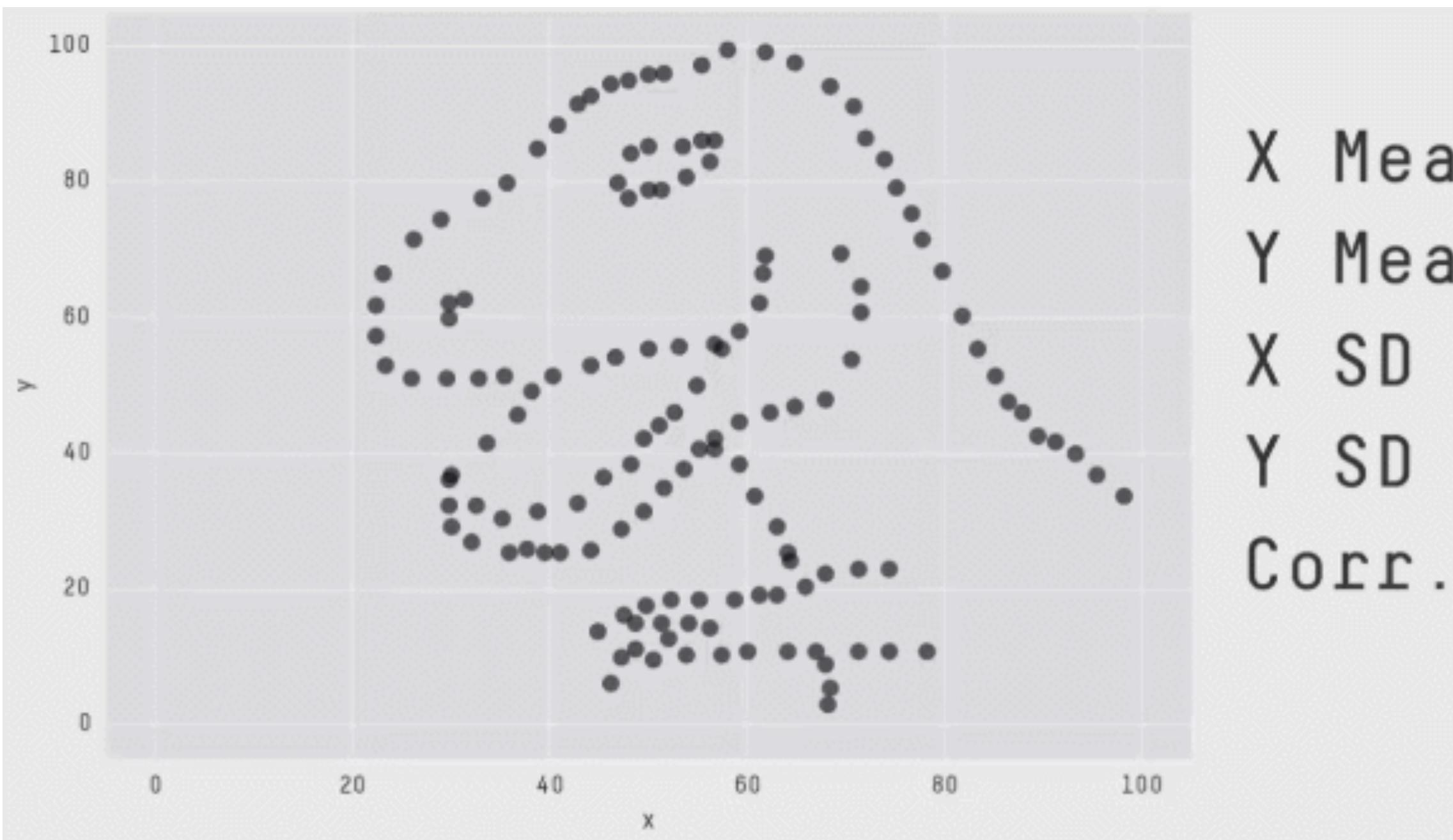






Mean x: 9 y: 7.50 Variance x: 11 y: 4.122 **Correlation x – y: 0.816** Linear regression: y = 3.00 + 0.500x

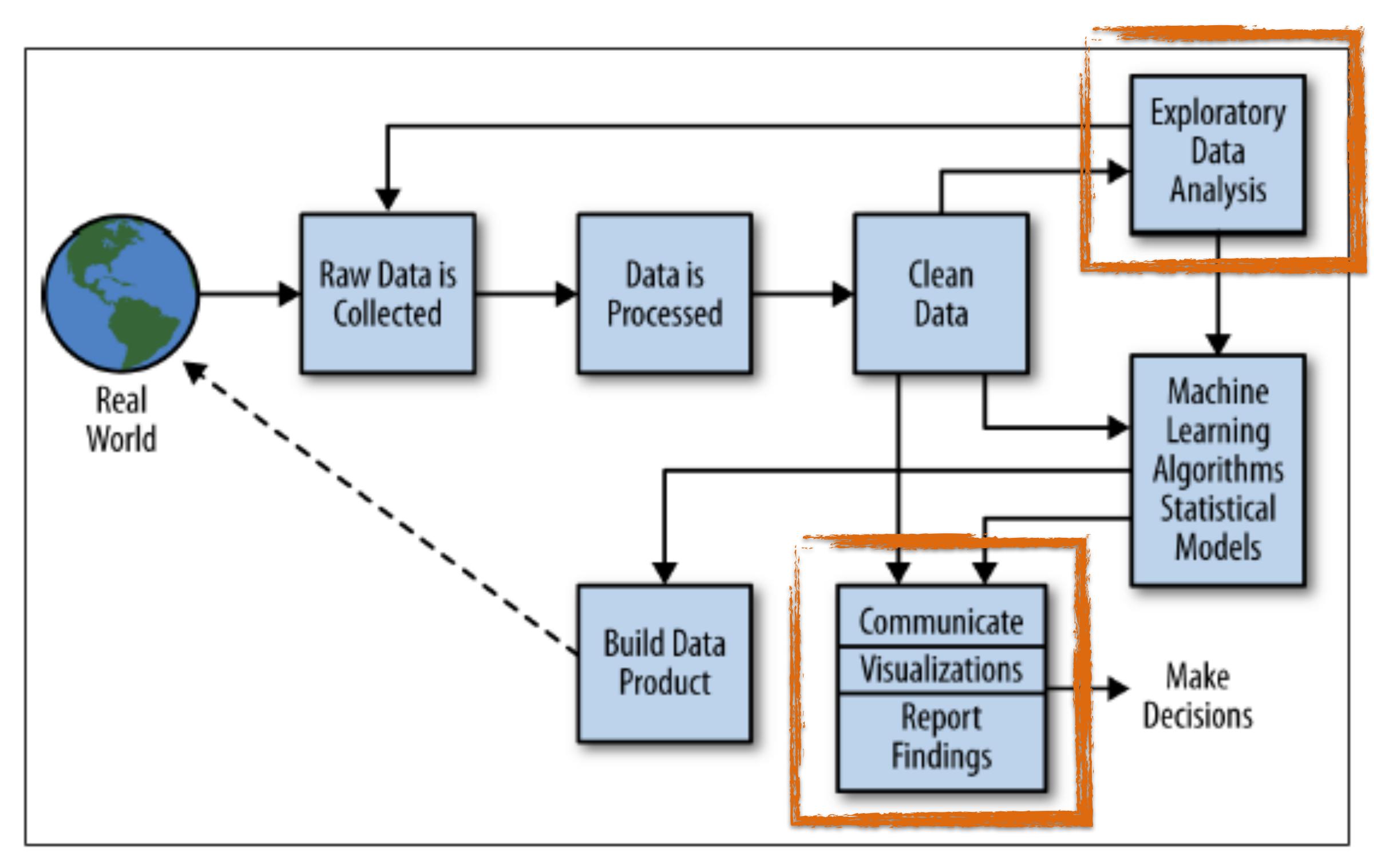




Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing, CHI 2017, Justin Matejka, George Fitzmaurice

X Mean: 54.2659224 Y Mean: 47.8313999 X SD : 16.7649829 Y SD : 26.9342120 Corr. : -0.0642526

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?



Magnitude Channels: Ordered Attributes Position on common scale Position on unaligned scale Length (1D size) Tilt/angle Area (2D size) Depth (3D position) →• ⊢ Color luminance Color saturation Curvature Volume (3D size)



Most

Effective

Least

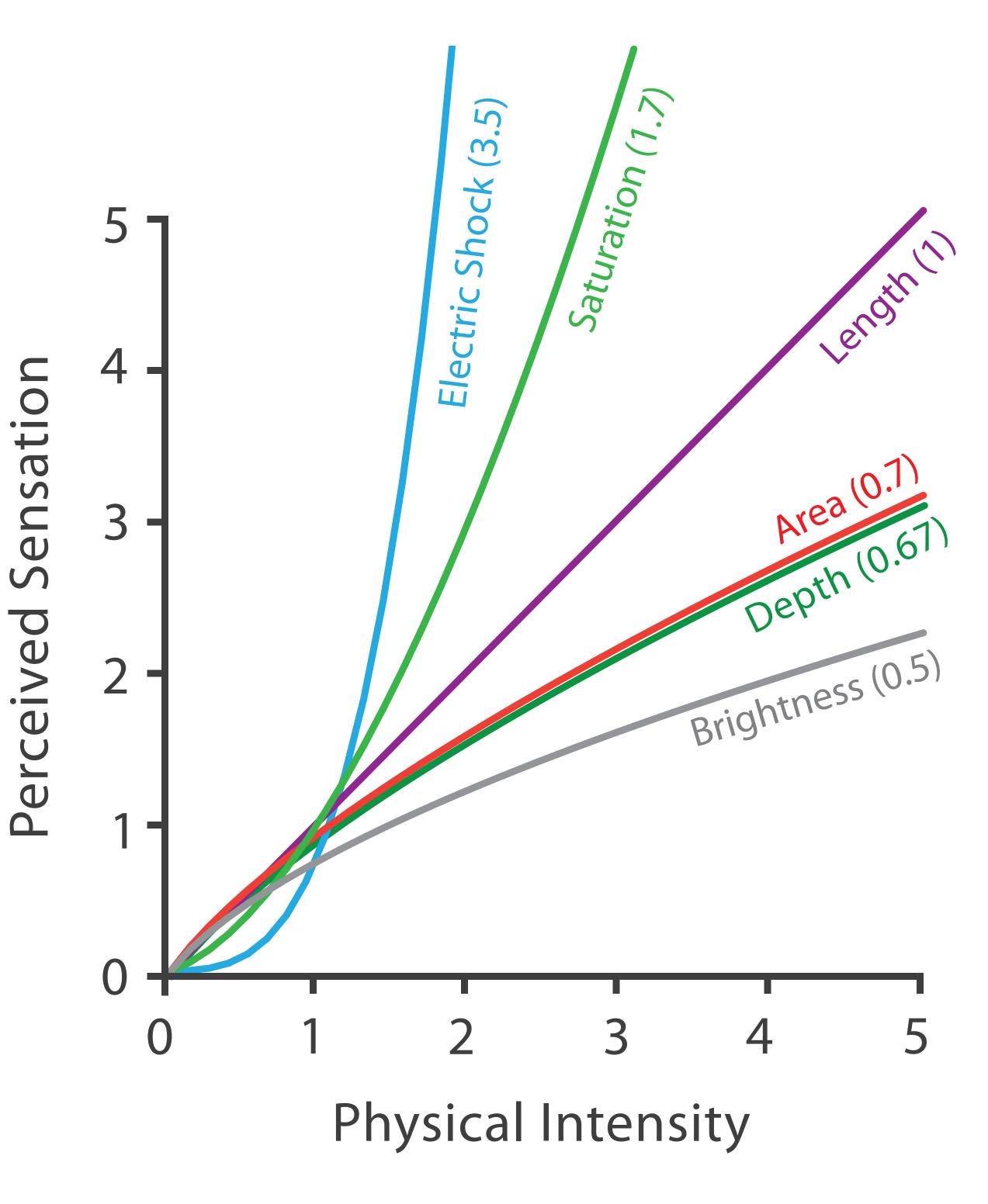
Identity Channels: Categorical Attributes

Spatial region	
Color hue	
Motion	.
Shape	+•



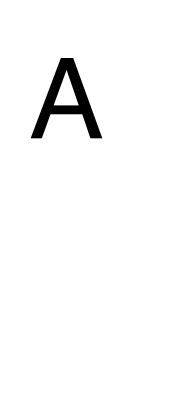
Why are quantitative channels different?

Steven's Psychophysical Power Law: S= I^N

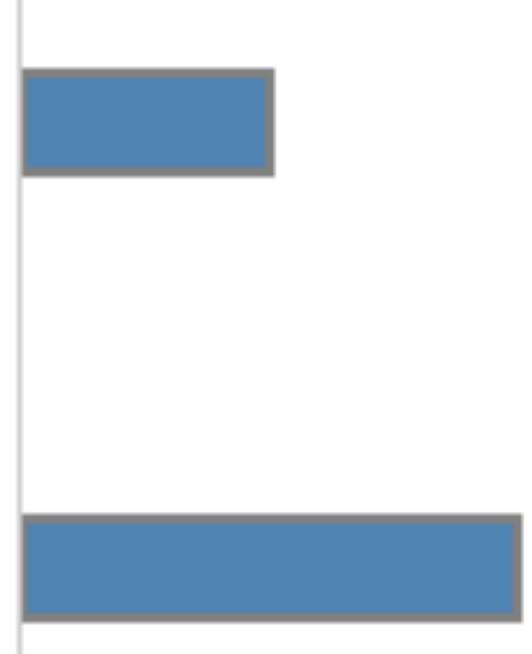


S = sensation I = intensity

How much longer?



B





How much longer?



B



How much steeper?



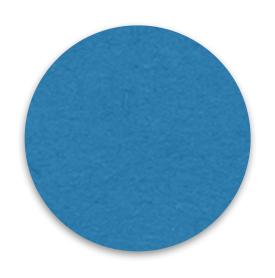
B

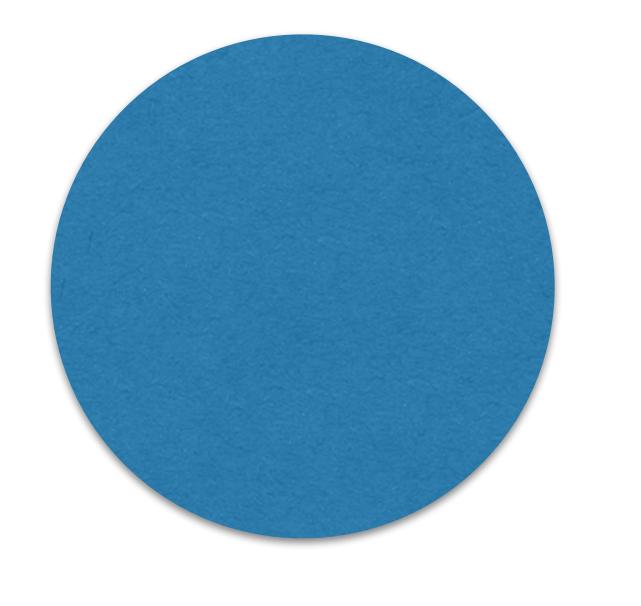




How much larger?







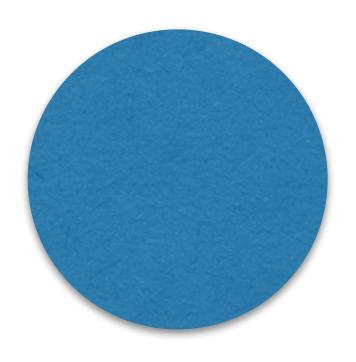


B

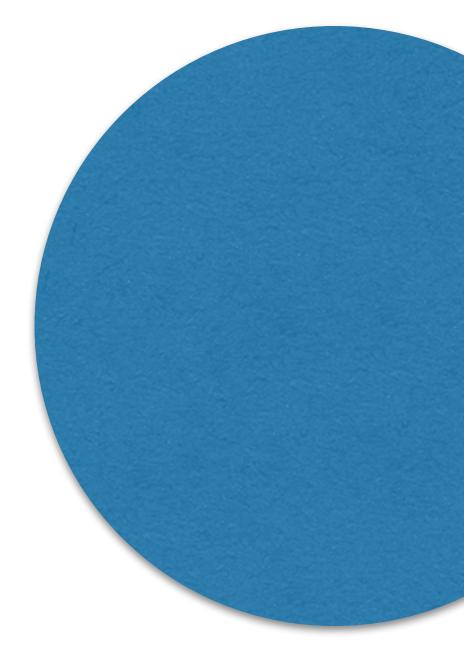
5X

How much larger?









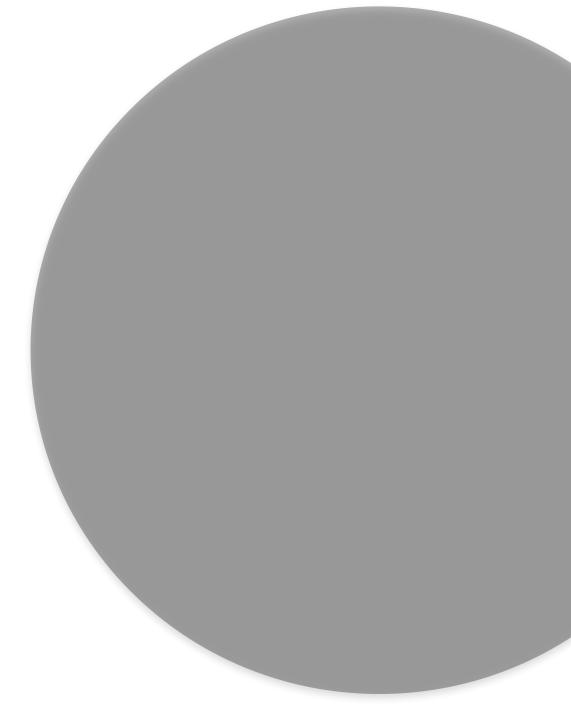
B

2x diameter 4x area area is proportional to diameter

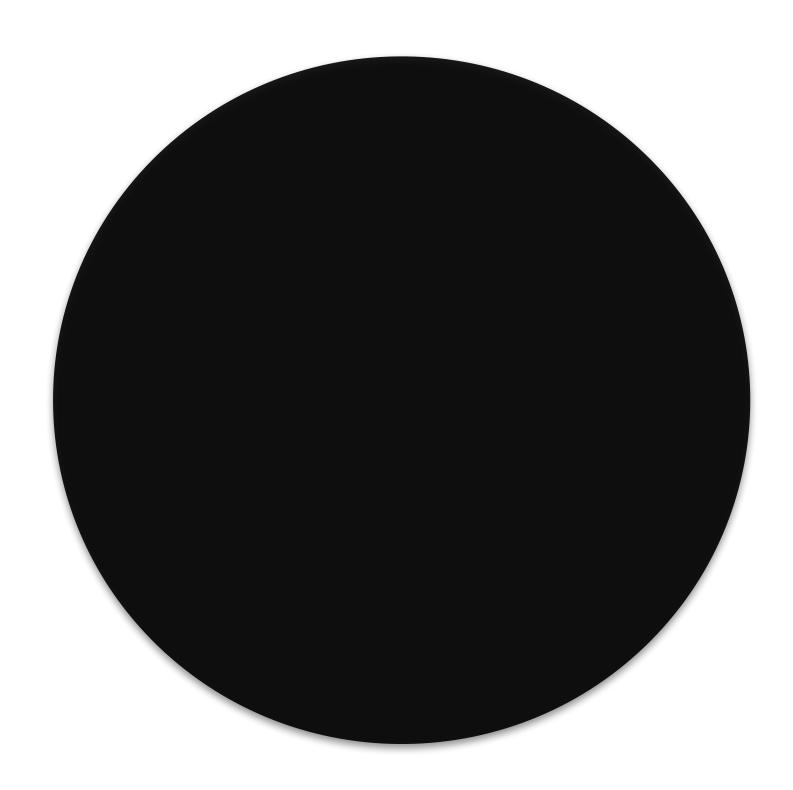
squared



How much darker?





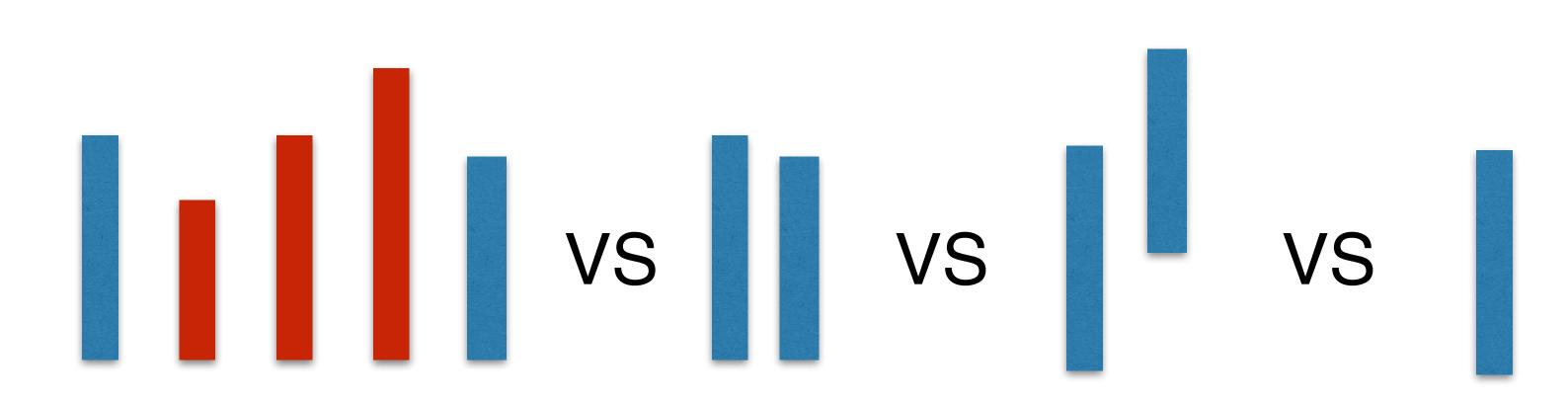


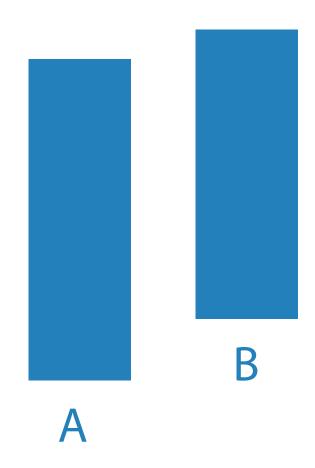


3x

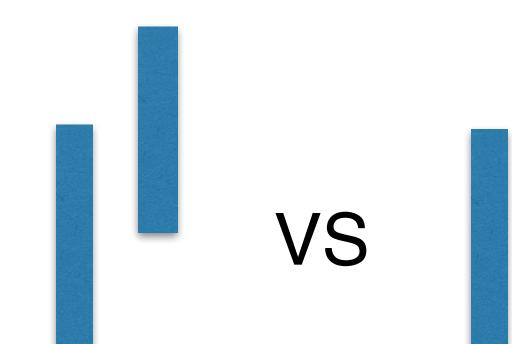
Other Factors Affecting Accuracy

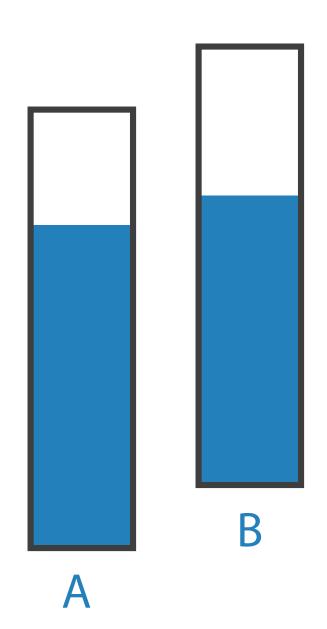
Alignment Distractors Distance Common scale



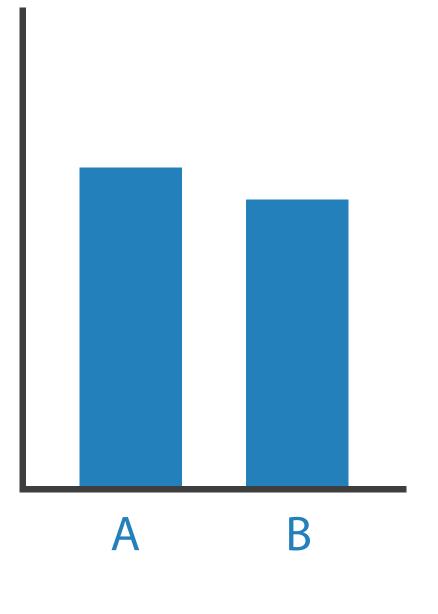


Unframed Unaligned





Framed Unaligned

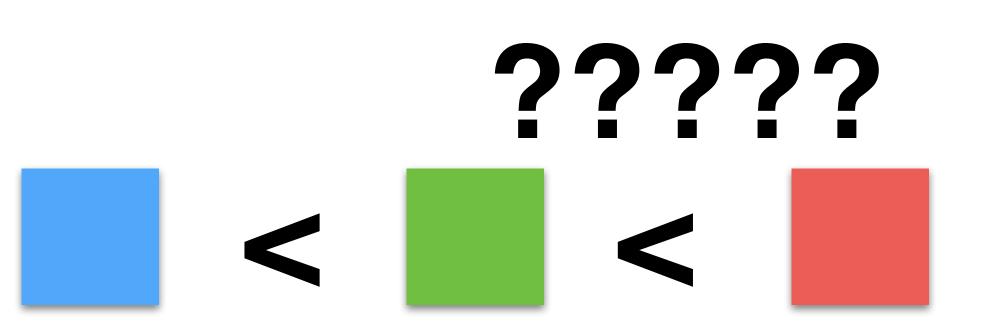


Unframed Aligned



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

- scales

1. Show data variation, not design variation 2. Clear, detailed, and thorough labeling and appropriate

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

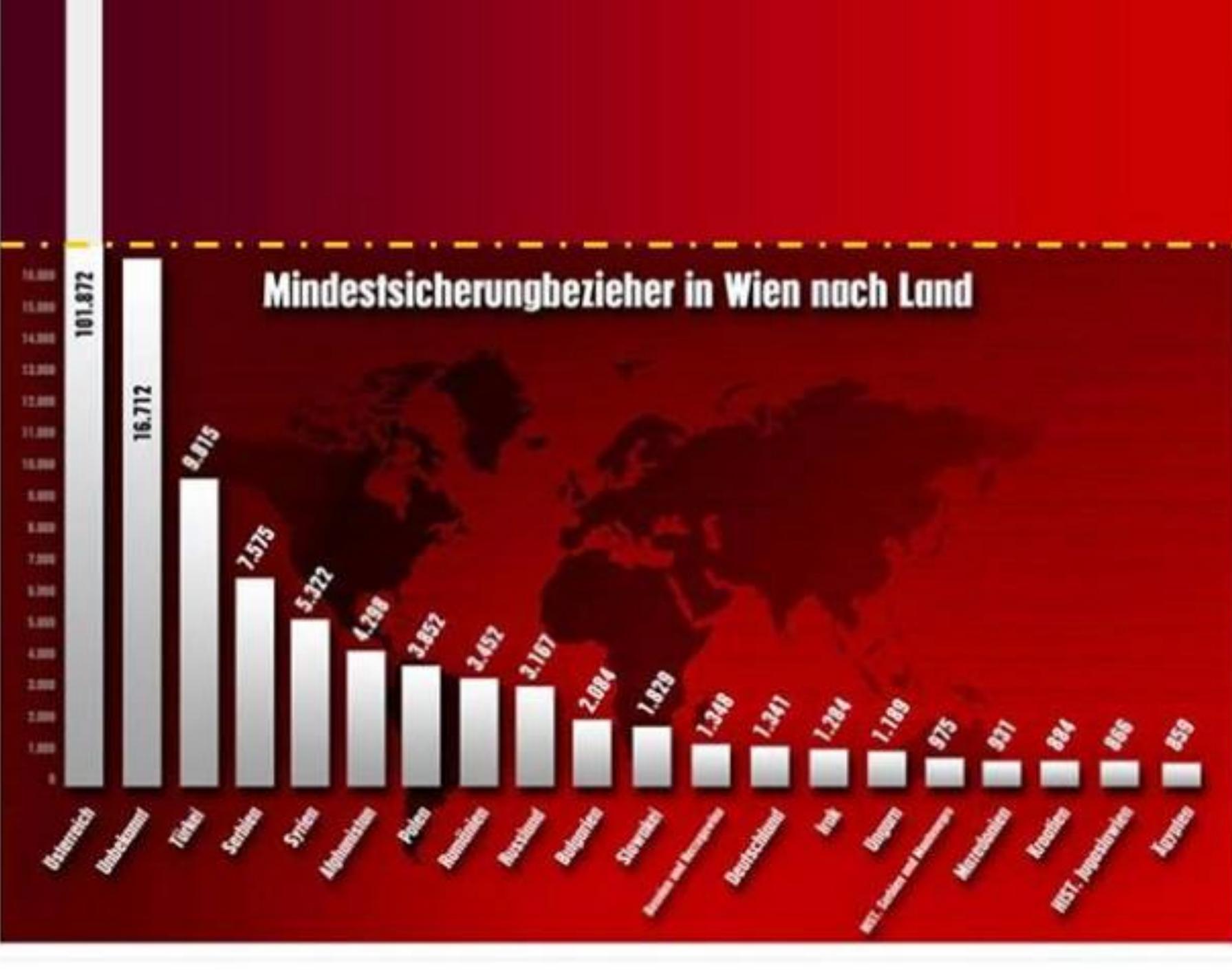
Asy Opf frei "Krc Räd geh Zwe Mit Wol Am

Stre

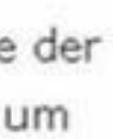
Mes

Abe

2



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?

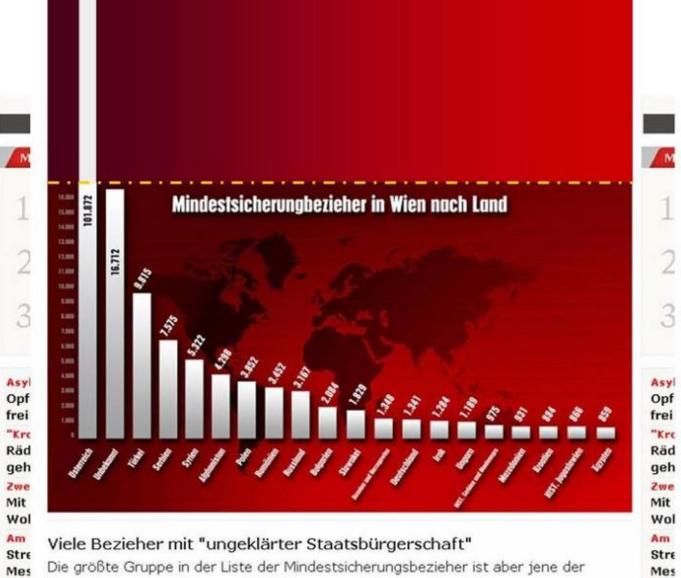
Grafik der Kronenzeitung

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

Grafik in echt



Abe

Abe "ungeklärten Staatsbürgerschaft". Dass es sich bei den 16.712 Personen um

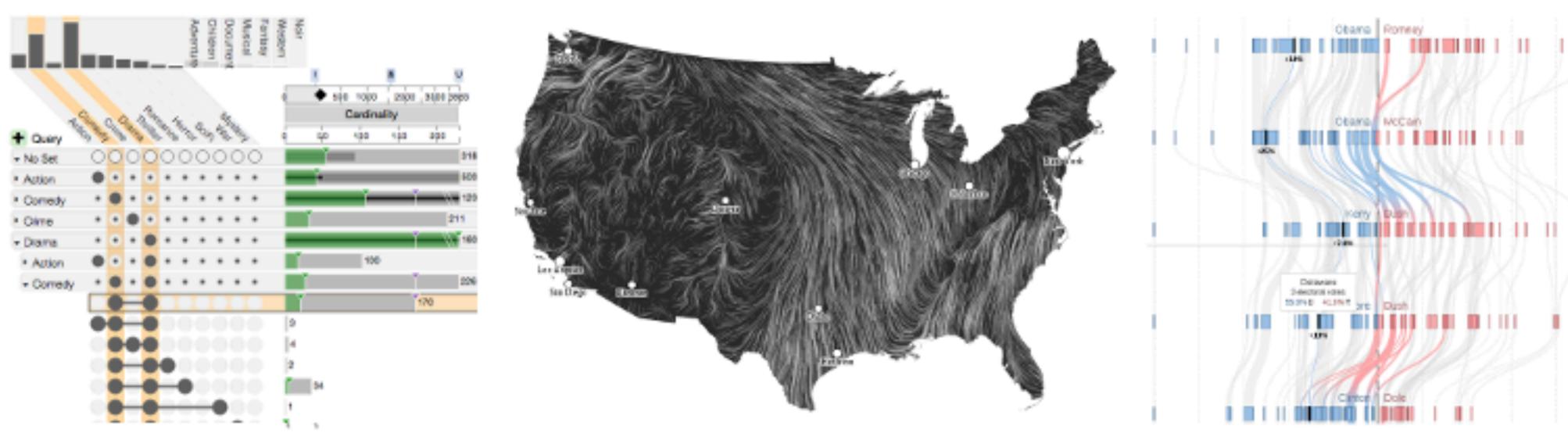
Resources

http://dataviscourse.net

Slides Videos D3 Tutorials



Uisualization CS-5630 / CS-6630



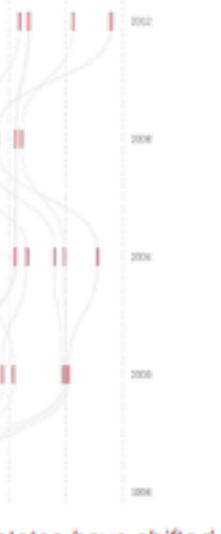
The amount and complexity of information produced in science, engineering, business, and everyday human activity is increasing at staggering rates. The geal of this source is to expanse you to vieual representation methods and techniques that

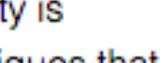


Resources

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

Fame

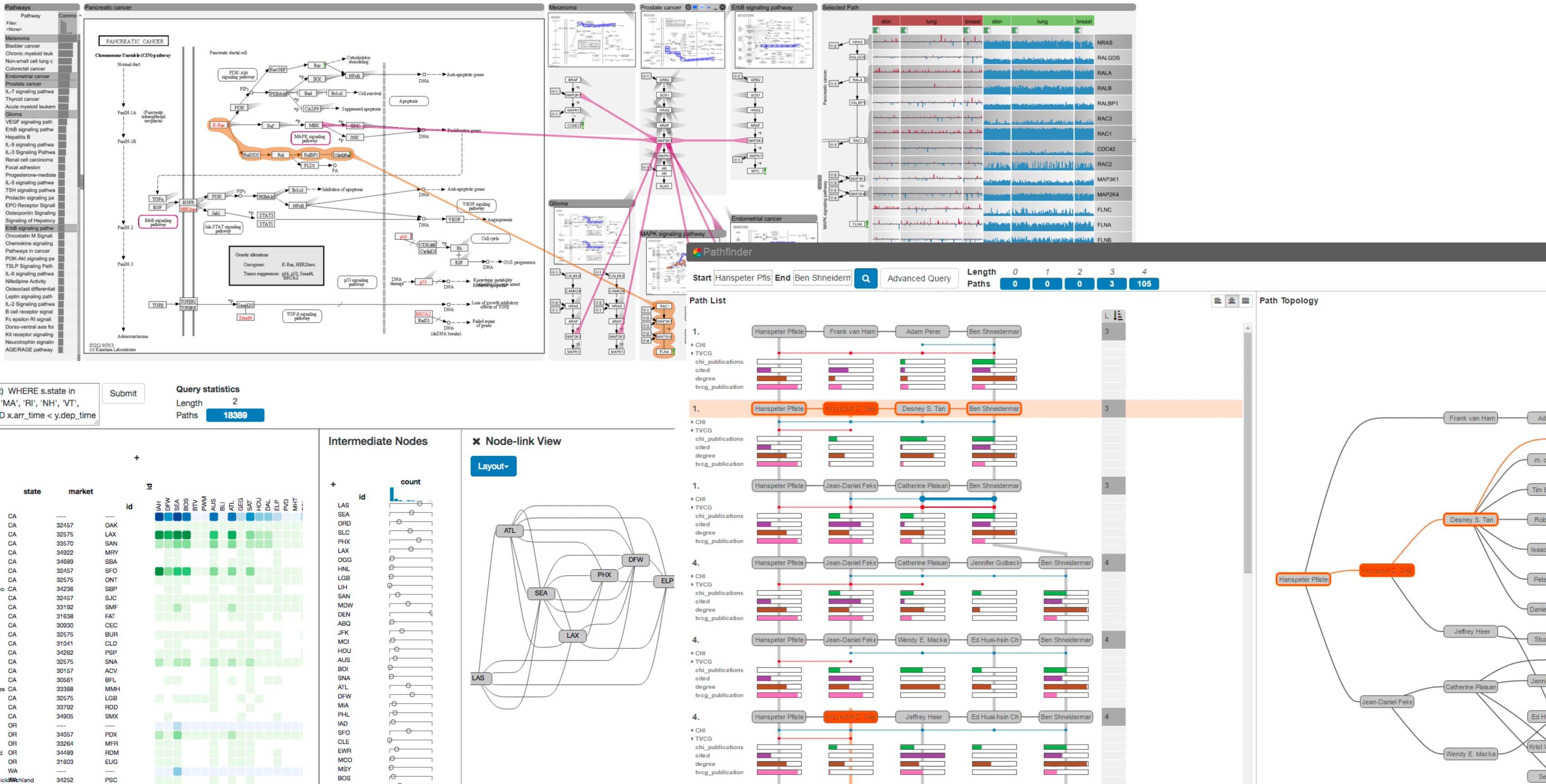


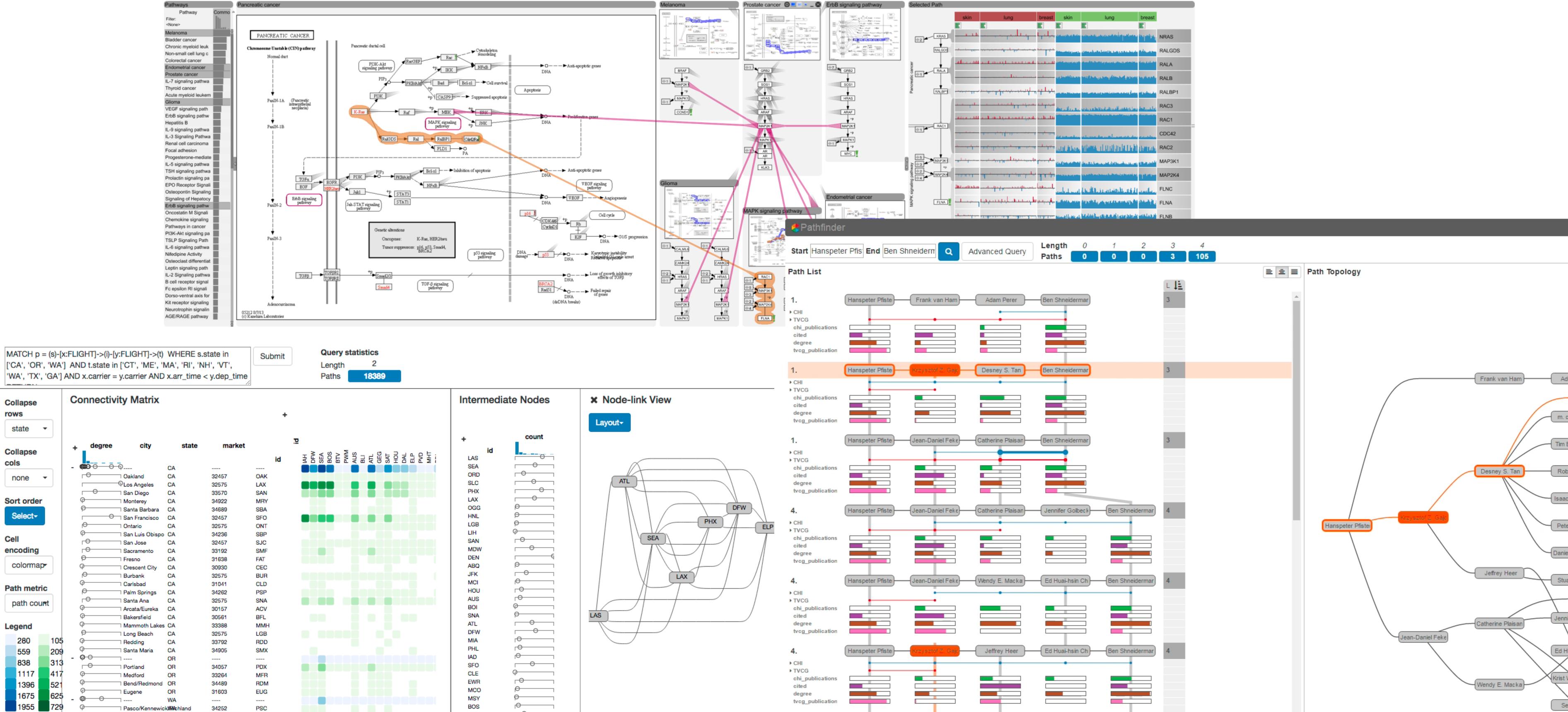






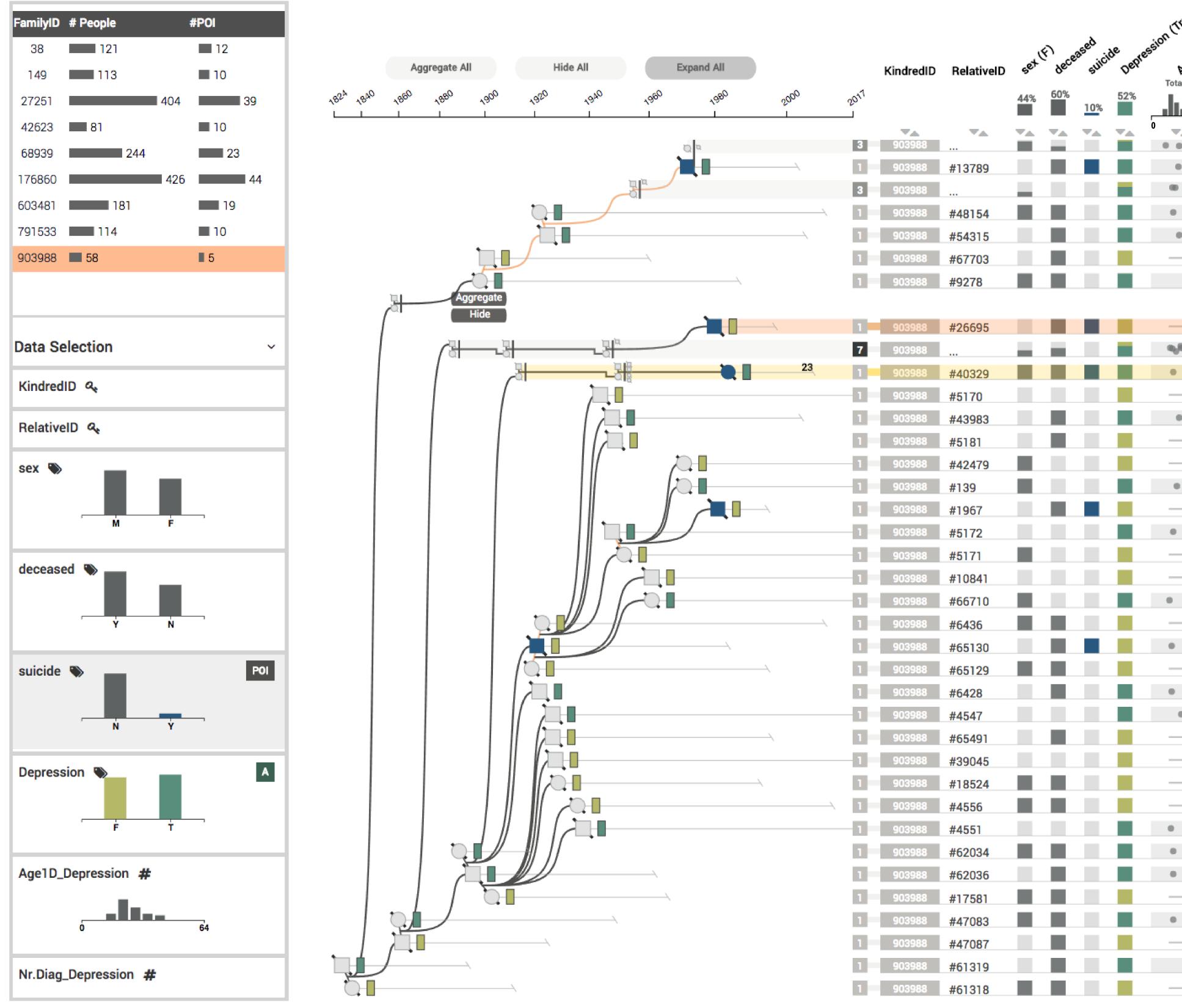
Large, Multivariate (Biological) Networks





1	lung	breast	skin	lung	breast	
ьb		┟┰┵┠┍┑	an a		n printe	NRAS
		•• T+•		and the latentic party of the basis	a la sur	RALGDS
	and more to react task.	1 n 41			1-1, 2 (L-1-)	RALA
			a de stille as	with the provident of provi		RALB
		4Lau			- Juliu Ar	RALBP1
- I -				ومالو فانتب أأدبر فلياب الطام ويردوه	-	RAC3
	and marks of constitutions.	a a ba a ba				RAC1
				والمتعادية المعادية المعادية	en alalan	CDC42
		-1-1 ⁰ -1	attl. Doo	nterna, dun dan da bahi	1. Totals	RAC2
• •		r 144		des la freces certe del becelor de certe	the sector	MAP3K1
-	ali a construiste des	44.40			111	MAP2K4
n han		ا	استا باس	Locille, scorearcher et d		FLNC
	un per un de la come de	a an	ullets, and	the first of the second states of the	n Alu	FLNA
	a talan a tala	nde tre-	بلد بعد العبدية	and and the state of the state of the	- Independent	FLNB

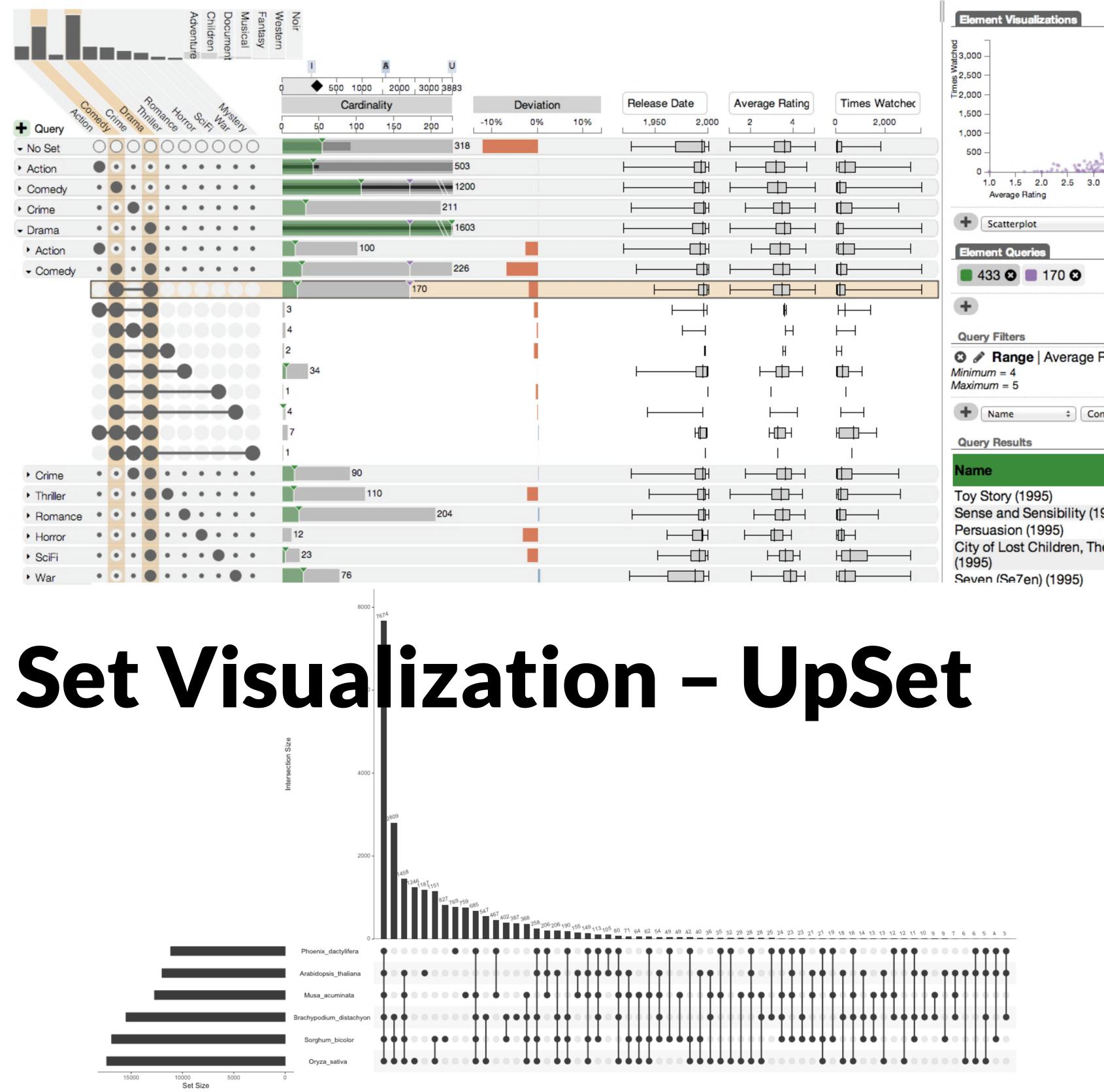
Genealogies & Clinical Data



(True)	epression Nr.Dia	g. Depression Bipola	AgeID B	polar Nr.Diag	aipolar		چے	(white)	other) cause death
otal:32	Total:60	814	Total:30	Total:60	MaxBMI Total:59	AgeMaxBMI Total:58	10 98%	120	Car
65		48%	0 35	0 _ 20	15 45	0 80		0.0%	
	10					• •		.	▼▲
•	•		•	•	0	•		-	gunshot woun
0	%	1.00	00	0.00	•• •	6		-	
•		100						-	
•			•					-	
								-	
	•							-	

						-	asphyxia
n,6		±0°₿	9499 ·	%	46.0	-	
•		•			•	-	carbon monox
						-	
	0	•				-	
	_		_	•		-	
						-	
•	0			•		-	
	_	_	_			-	cardiac arre
•	•					-	
						-	
						-	
0						-	
			_			-	
•	60					-	strangulatio
	_		_			-	
•			•		0	-	
						-	
						-	
						-	
						 -	
						-	
0		•				-	
•						-	
0						-	
				•		-	
0	•					-	
					0	-	
		•	•	•	•	-	
						-	

Multidimensional Data



erage Rating	Times Watchec
2 4	0 2,000
[]]	
—- () —-I	
—[]]I	₽
$\vdash \Box \vdash \dashv$	
H	₩
Н	$\vdash \vdash$
H	Н
\vdash	l∰i
	I
\vdash	⊢
НШН	HH
	I
——————————————————————————————————————	
\vdash	
— <u></u>	ф-I
ΗЩΗ	
┝━━━━╋╋	H

Eleme	nt Visualizations
patos 1000 – 1000 – 1000 – 1000 –	•
23,000 -	
g2,500 -	
₽2,000 -	
1,500 -	
1,000 -	
500 -	
0-	
1.	
1	Average Rating
+	Scatterplot 💠 🗲 🔶 🙆 🔞
Eleme	nt Queries
43	3 3 9 🔳 170 3
Querv	Filters
	Range Average Rating
+	Name
Query	Results
	Deleger

V
Ra

	Date				
	1995				17.99%
995)	1995	Rank	School Name	Country	Acade
	1995	1.	Massachusetts Insti	United States	
ie	1995	2.	Harvard University	United States	
	1995	3.	University of Camb	United Kingdom	
		4.	Imperial College L	United Kingdom	
		5.	University of Oxfor	United Kingdom	
		6.	UCL (University Col	United Kingdom	
		7.	Stanford University	United States	
		8.	Yale University	United States	
		9.	Princeton Universit	United States	
		10.	University of Chica	United States	
		11.	ETH Zurich (Swiss F	Switzerland	
		12.	Columbia Universit	United States	
		13.	University of Penns	United States	
		14.	Cornell University	United States	
		15.	University of Edinb	United Kingdom	
		16.	Ecole Polytechniqu	Switzerland	
		17.	King's College Lond	United Kingdom	

18. University of Toron Canada

19. McGill University Canada

20. National University Singapore

21. University of Michi United States

22. University of Califo United States

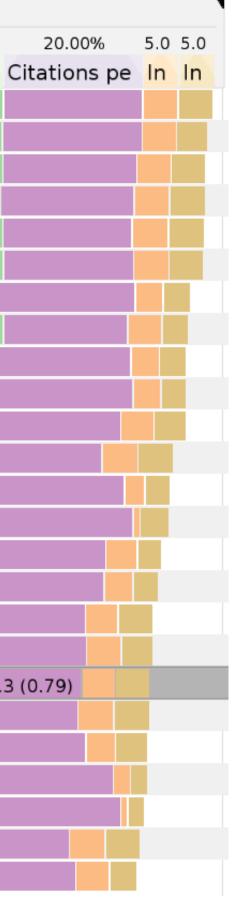
23. California Institute United States

25. Duke University United States

24. University of Bristol United Kingdom

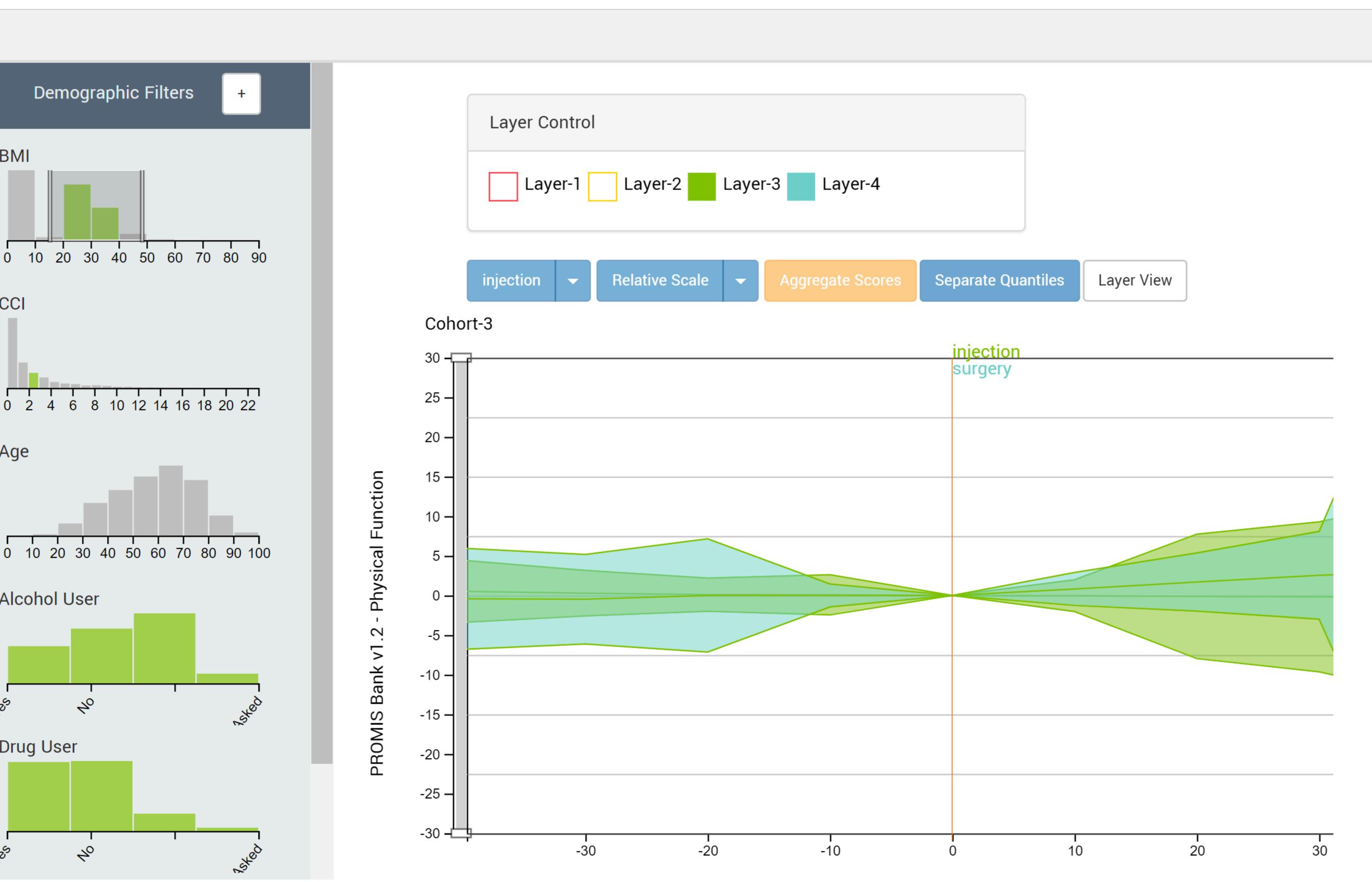
Multivariate ankings – Lineup

	World Unive	rsity Rank	king						World	d Unive	rsity Ranl	king	
%	32.94%	19.63%	19.63%	4.4					40.00%	10.00%	20.00%	6	
•	Employer repu	Faculty/	Citation	1 1		Separator	Rank	7	Academic reputation	Empl	Faculty/s	tud	C
						<u> </u>	1.						
							2.						
							З.						
							4.						
							5.						
							6.						
							7.						
							8.						
							9.						
							10.						
							11.						
						$ \rightarrow $	12.						
							13.						
							14.						
						$\sim 7^{-1}$	15.						
							16.						
	93.7 (0.94)						17.						
							18.						
							19.	94	4.6 (0.95)	8	9.9 (0.9)	79.3	3
					4	\langle / \rangle	20.						
						A	21.						
						\checkmark	22.						
						$' \searrow$	23.						
							24.						
							25.						
							-						

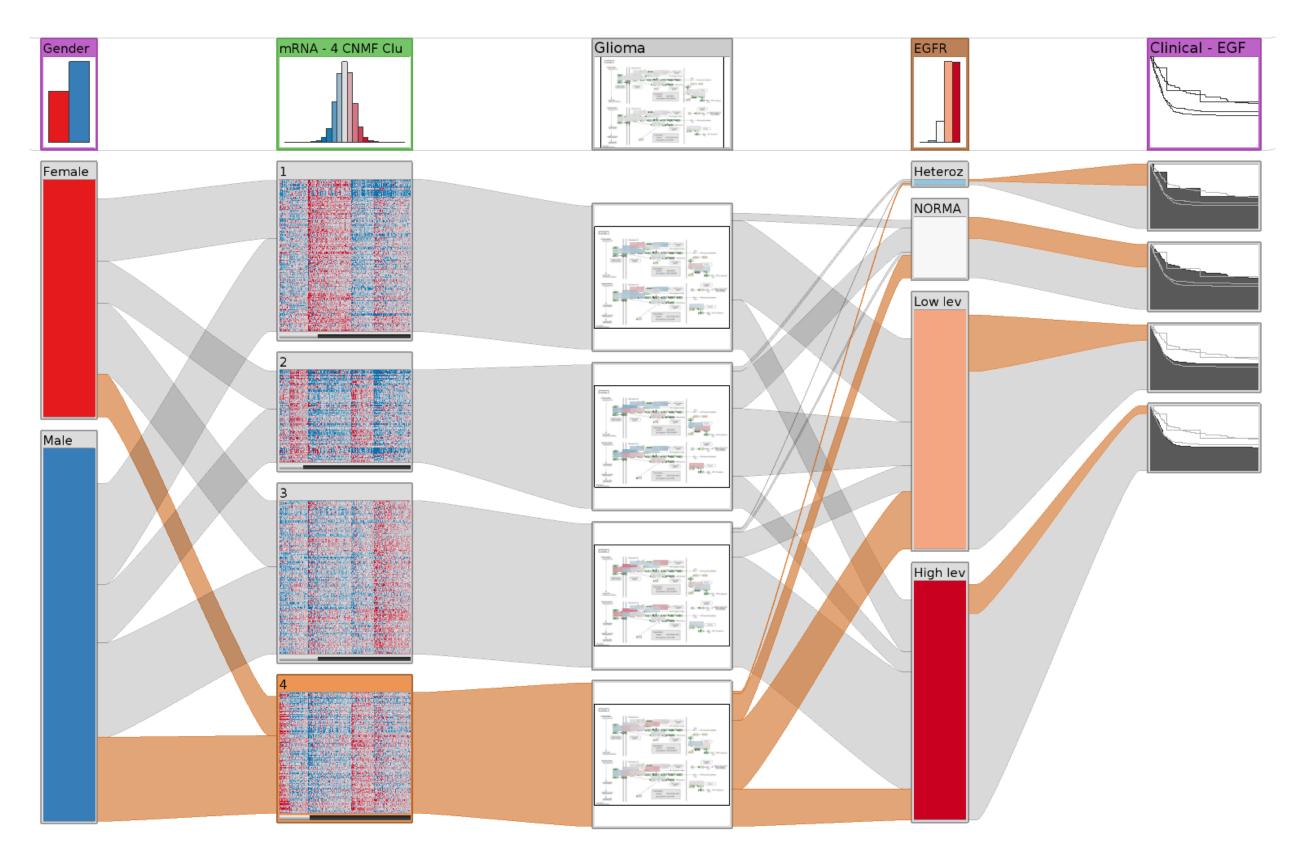








Genomic Data

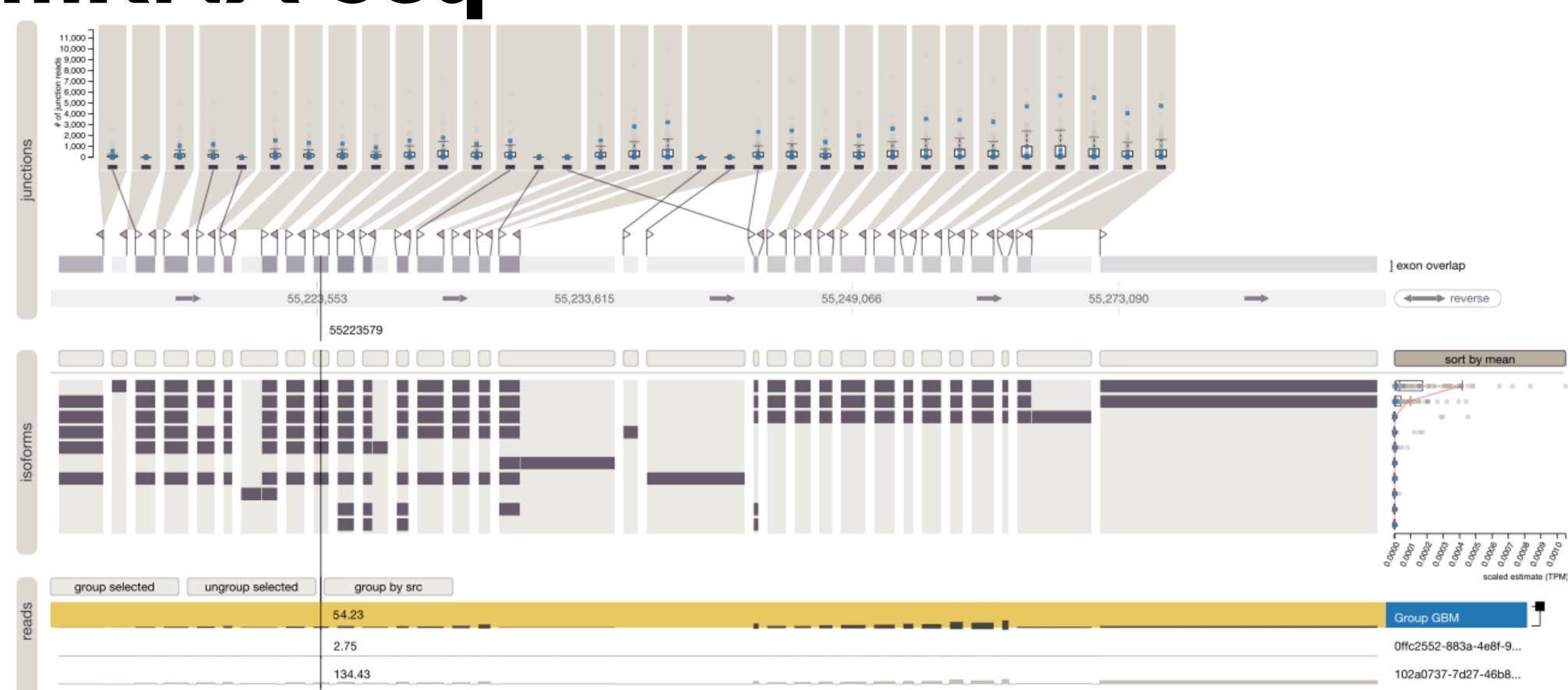


Cancer Subtypes / Omics Clustering and Stratification



cTracks - Copy Number Browse

Alternative Splicing / mRNA-seq

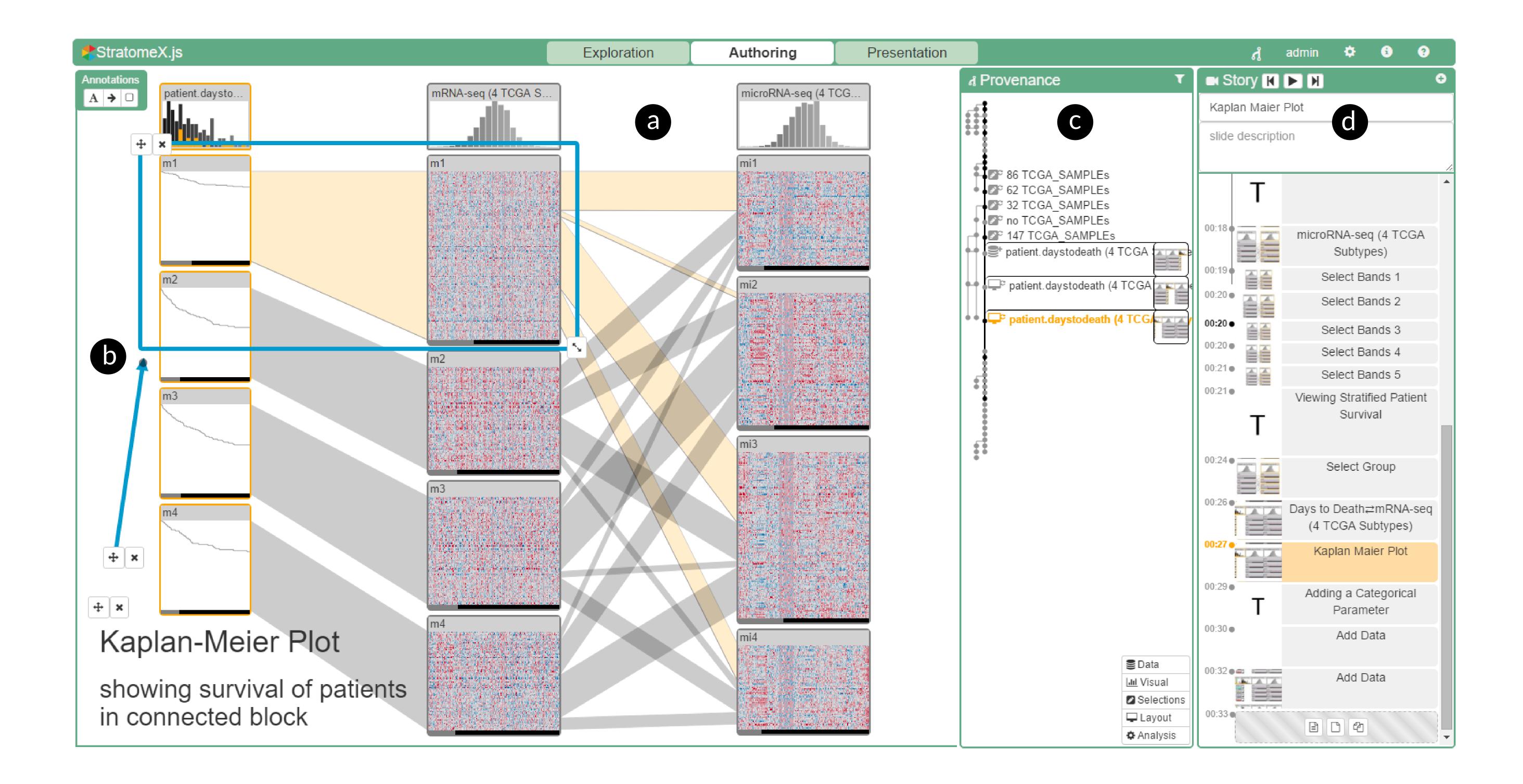


	sim_02_sorte	d.reads				<u></u>		-				. 1 300.00			5°
CNV analysis - Apply results	sim_02_sorte						nin y								
chr3 Update Location	sim_02_sorte sim_02_sorte		g2											 	~. ~
Update Display	PPARG IIII GHRL	11	II SCN5A				3:10	ا == 00,000,000	GS	CASR	TF AT	AGTR1 IL I I I R 150,000,000	TN		AD
 sim_02_sorted.cnv_log2 sim_02_sorted.reads 	b37 - 3	8:1,092,6	829-196,	388,826	& 5:149,9	977,682	-157,37	'1,402 [of	ffset 0,	,0:0,0]			Q		\$
 sim_02_sorted.log2_log2 sim_02_sorted.log2_qual 	sim_02_sorte		whenever	- Shaga	e . Navelski		وأحدادت	Contral		and the	ligen	an she was a she was	Constation		
Higlass - Tracks menu Top Bottom Track name Sim_02_sorted.ballele	sim_02_sorte		7.9	<u>}</u>			4	:	42	<u></u>					
Load CNV BED File	sim_02_sorte	er an generet	- Carrier								-	a da anti- Se francé	y langer an Iyonitanan		• •
Load Input Config File	1	2	3	4	5 6	7	8	9 10				16 17 18	1920 2	2 X	Υ
Louding input mes	b37 - 1	:1-MT:1	6,569 &	7:30,008	,725-47,	501,116	[offset	-4102557	73,213	073349:	0,0]		Q		\$
.oading input files															

	Group GBM
	0ffc2552-883a-4e8f
	102a0737-7d27-46



Reproducibility, Storytelling, Annotation, and Integration in Computational Workflows



Lineage Visualizing Clinical Data in Genealogy Graphs



Carolina Nobre



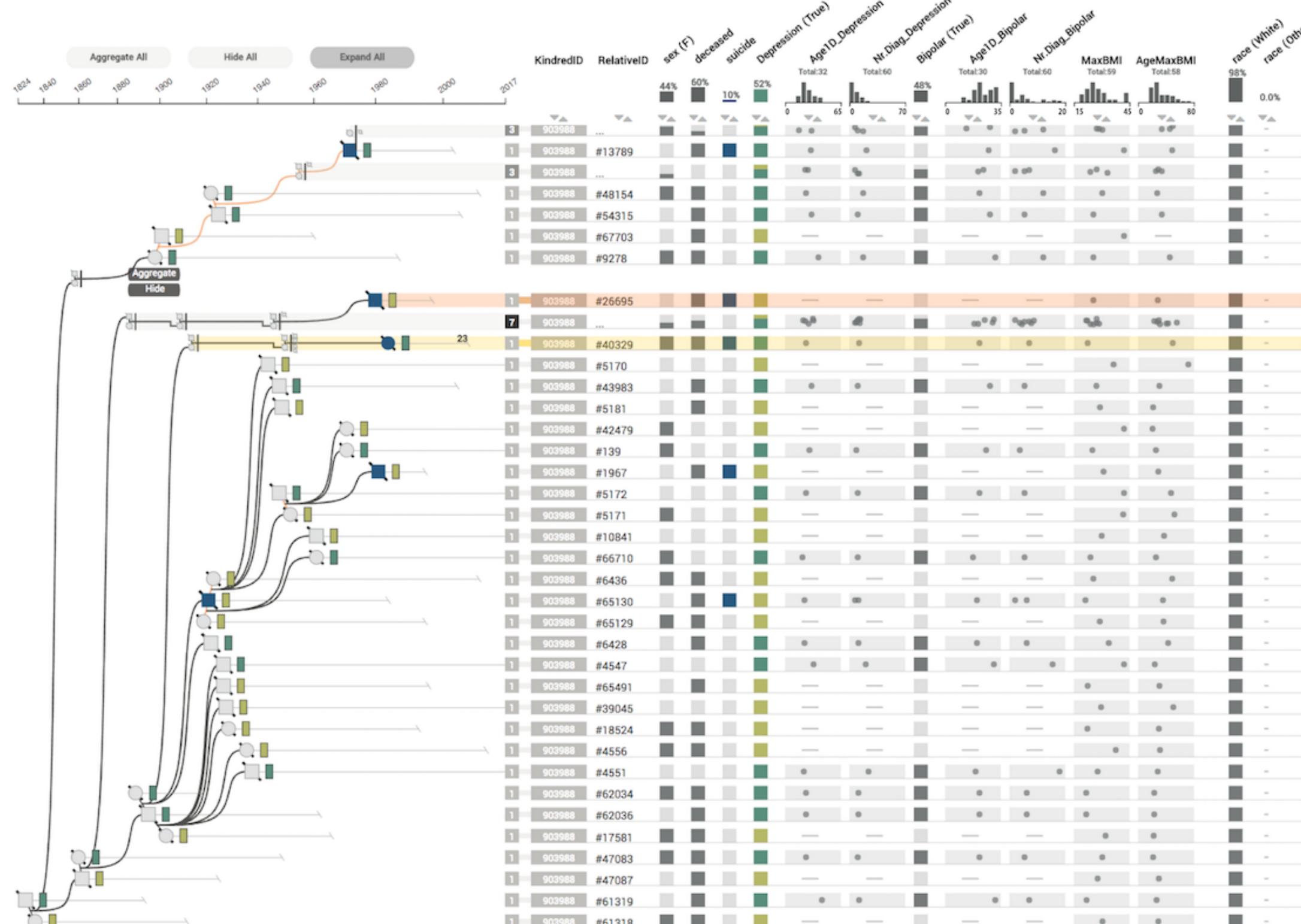
Hilary Coon



Nils Gehlenborg



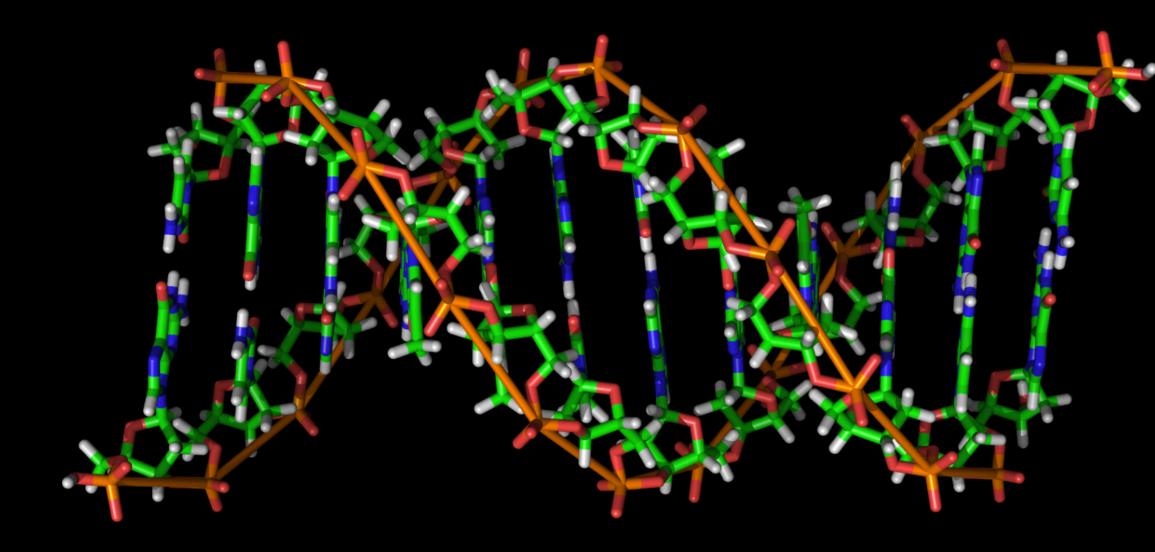
Alexander Lex



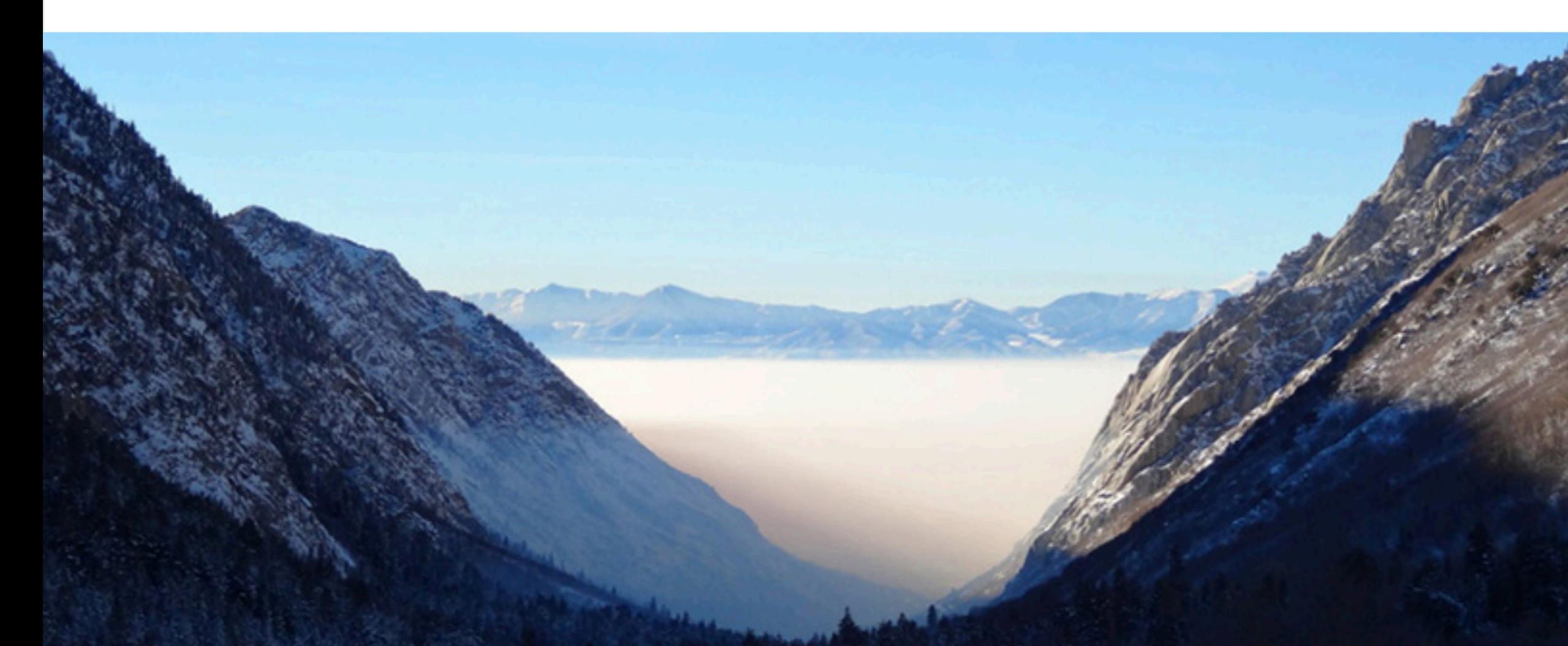
	#26695			—	—	—	—	0	0	-	a
				- 4	æ		1494P	- 54	46.00	-	
	#40329					0	0		0	-	Ca
	#5170							0	0	-	
	#43983			0	0	0	0		0	-	
	#5181			_	_	_	_	0	0	-	
1	#42479				_		_	0	0	-	
1	#139			0	0	0	0	0	0	-	
I.	#1967						_		0	-	Ca
I.	#5172				0					-	
I.	#5171			_	_	_	_		0	-	
I.	#10841			_			_	0	0	-	
I.	#66710			0	0		0	0	0	-	
I.	#6436			_		_	_	0	0	-	
I.	#65130			0	60	0	0 0	0	0	-	st
I.	#65129						_	0	0	-	
I.	#6428			0	0	0	0	0	0	-	
I.	#4547			0	0				0	-	
I.	#65491			_				0	0	-	
I.	#39045						_	0	0	-	
I.	#18524							0	0	-	
I.	#4556						_	0	0	-	
I.	#4551			0	0	0	0	0	0	-	
I.	#62034			0	0	0	0	0	0	-	
I.	#62036			0	0	0	0	0	0	-	
I.	#17581						_	0	0	-	
	#47083			0	0	0	0	0	0	-	
	#47087							0		-	
	#61319				0			0		-	
l	#61319			_		_		0	0	-	







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**





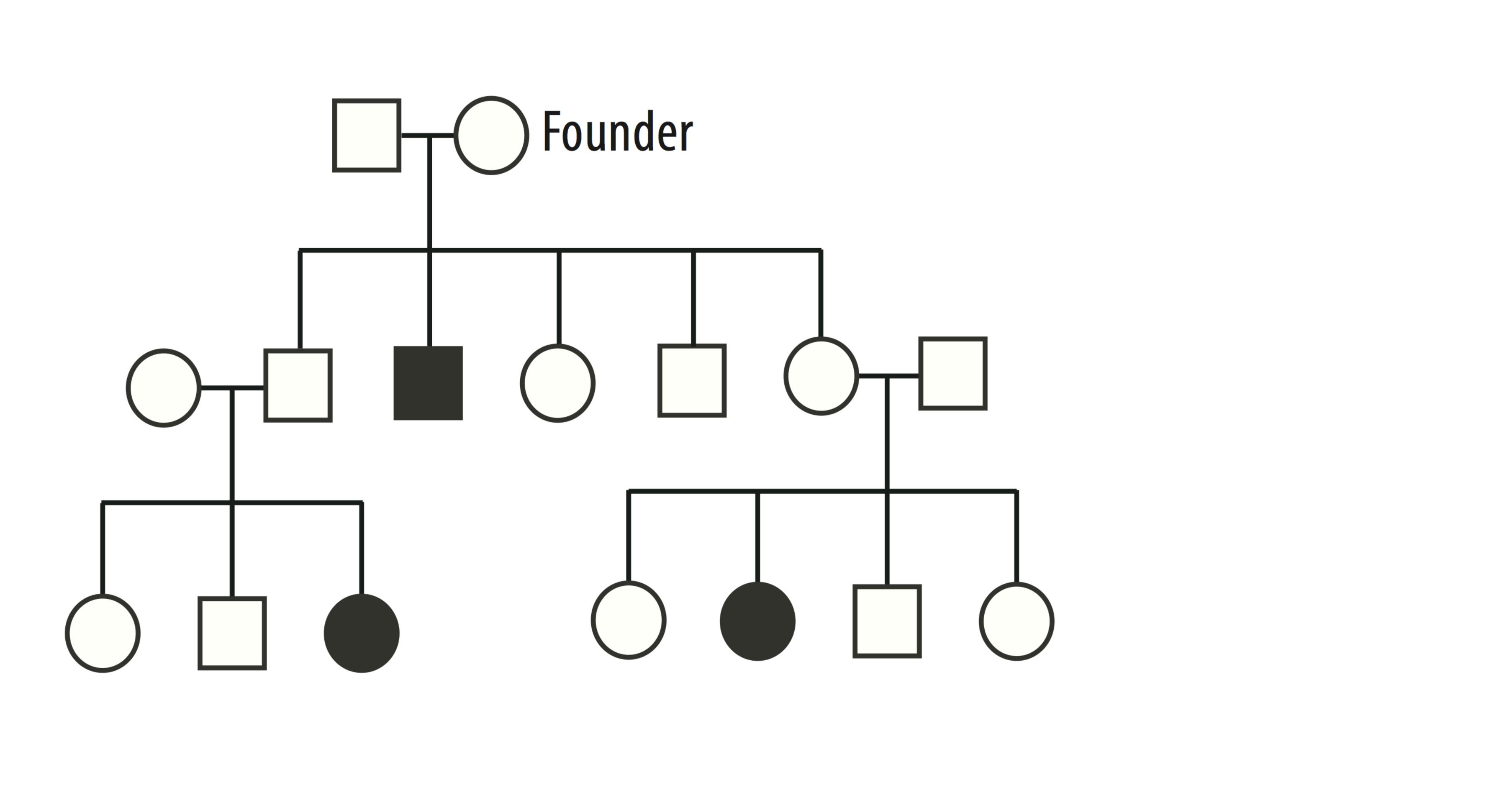


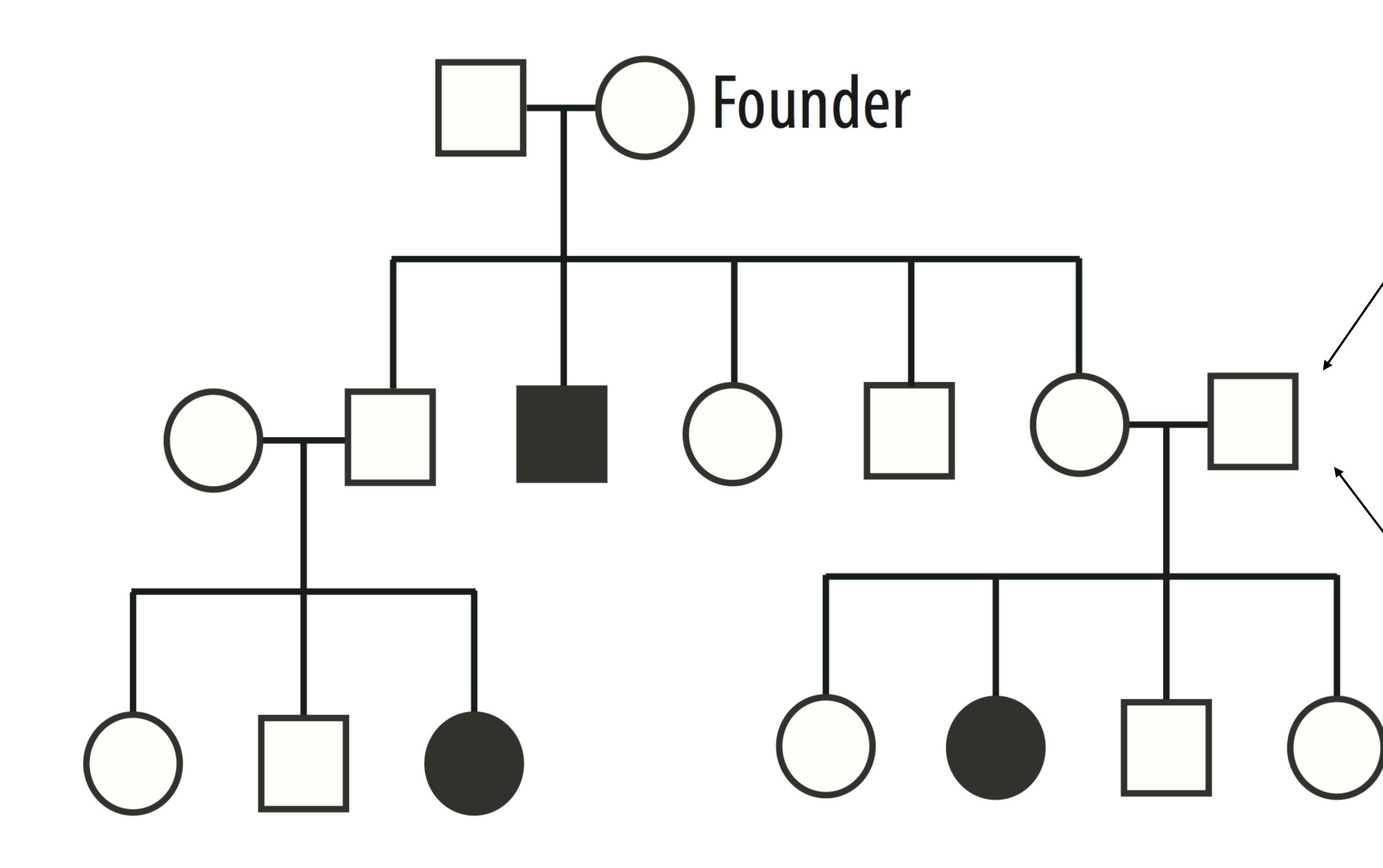
Find familial cases that also have an "interesting" phenotype e.g., predominantly female, associated with rare psychiatric disease, etc.

Prioritize those cases for analysis of shared genomic sequences

Proofreading the Data!

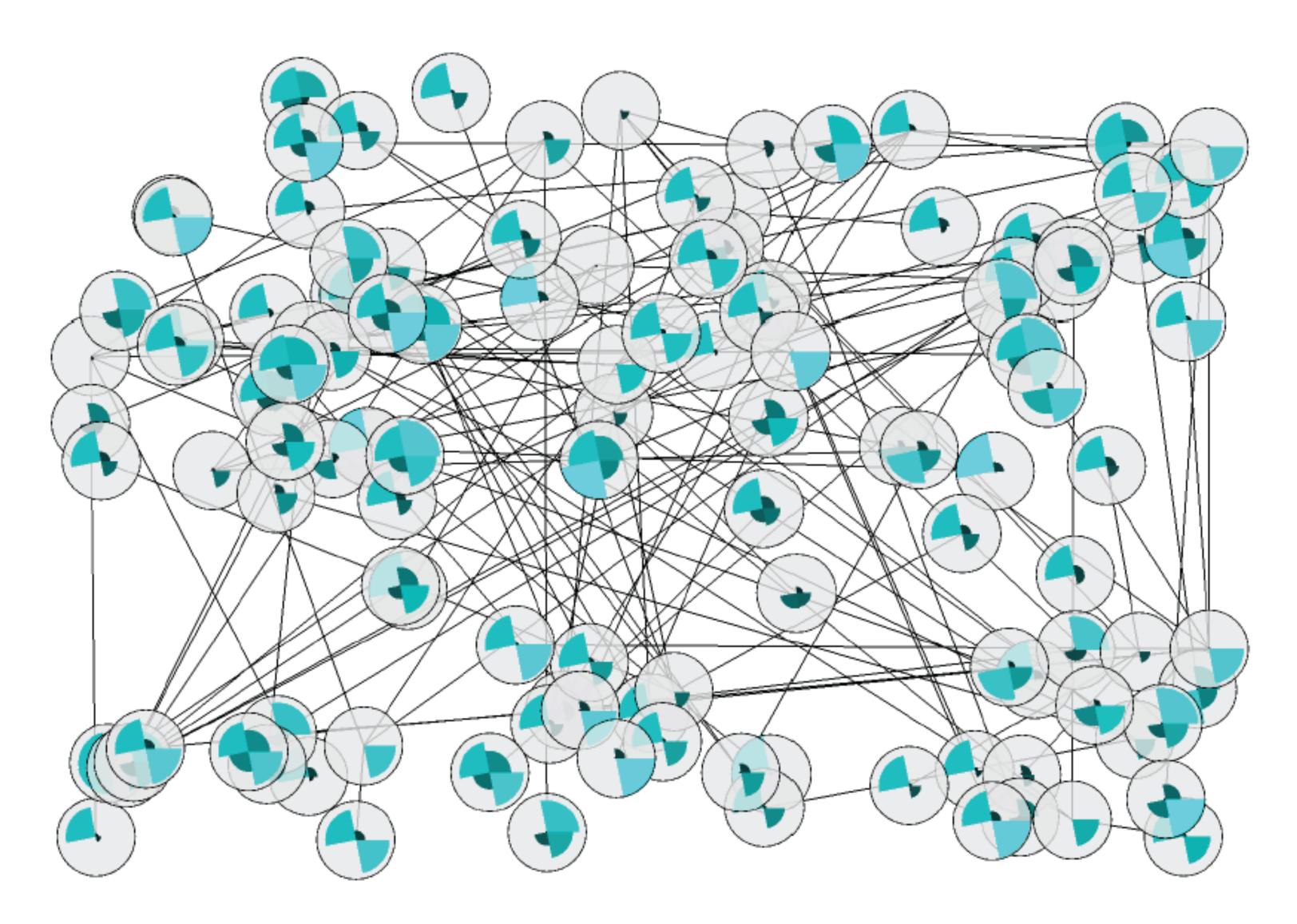




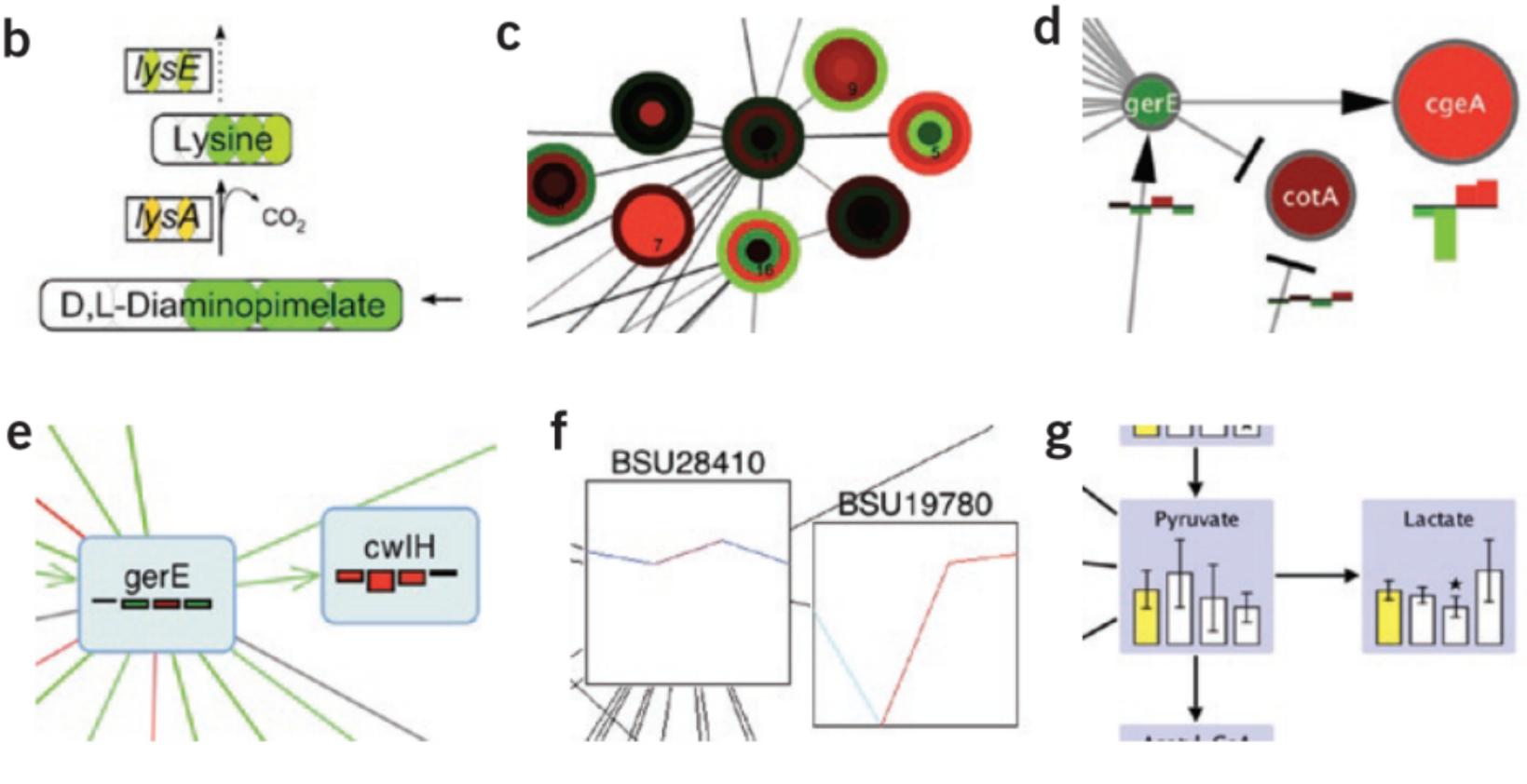


Age Sex Race Bipolar # diagnosis bipolar Depression **# diagnosis depression** Asthma # diagnosis asthma Obesity Schizophrenia **Cause of Death** Weapon Used

Multivariate Attributes and Graphs How can we deal with graphs that contain **rich attribute data**?



b lysA CO2 D,L-Diaminopimelate

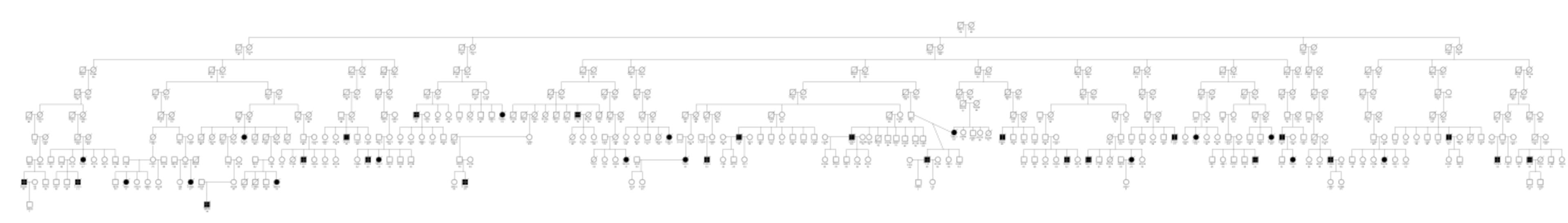


[McDonnel2009]

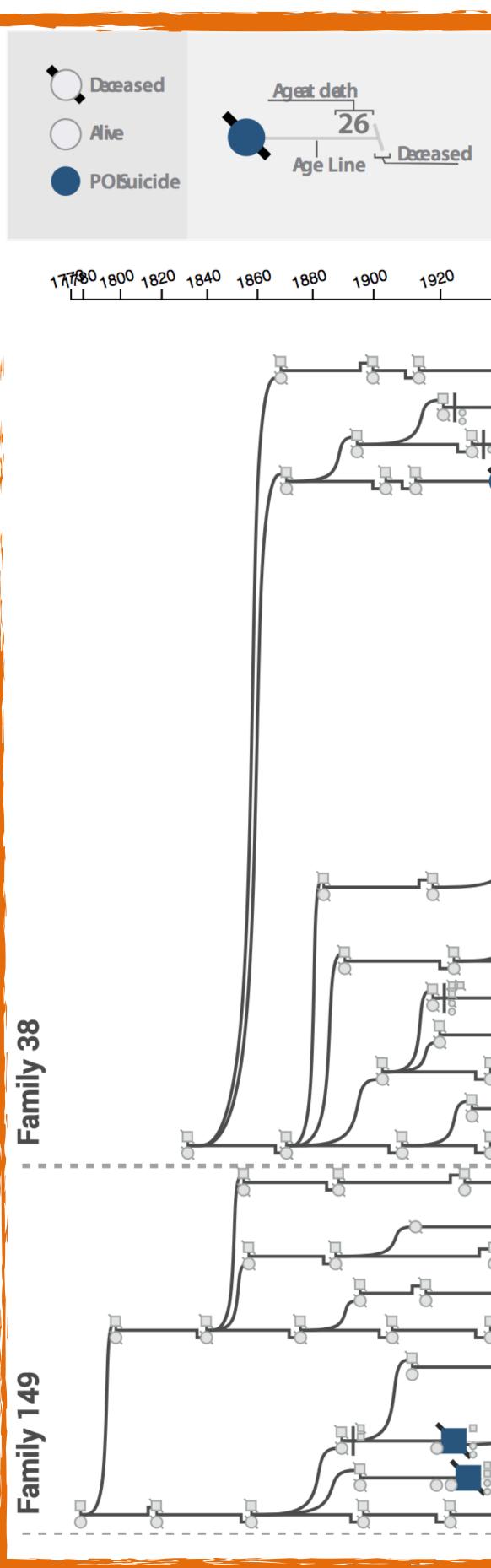
[Gehlenborg2010]



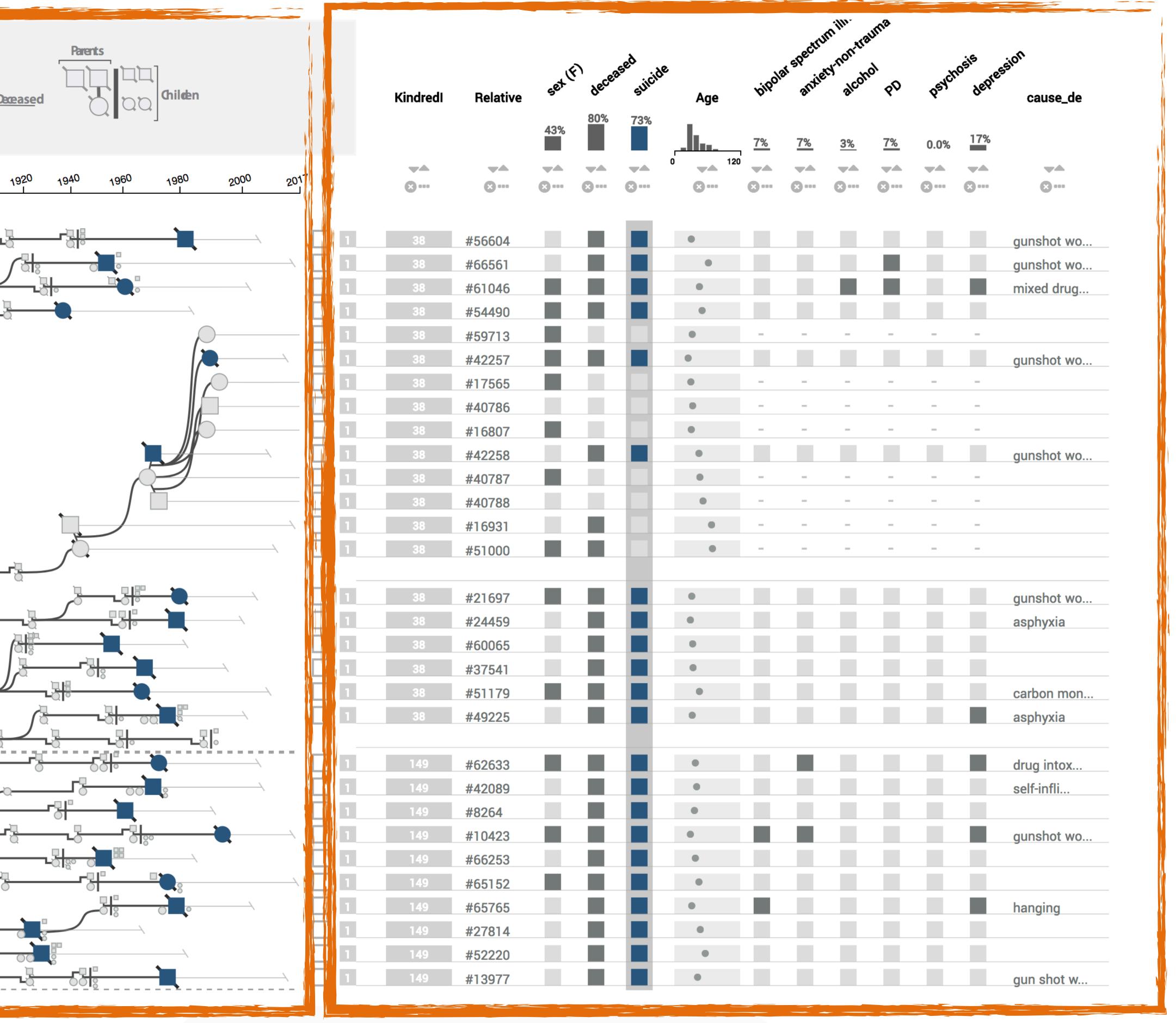
Genealogy with ~400 members rendered with Progeny



F	amily S	elector	Expand
	ID #	People	#POI
0	38	121	12 (9.9%)
0	149	113	10 (8.8%)
0	38	121	12 (9.9%)
0	149	113	10 (8.8%)
0	27251	404	39 (9.7%)
0	42623	81	10 (12.3%)
0	68939	244	23 (9.4%)
Đ	176860	426	44 (10.3%)
0	603481	181	19 (10.5%)
0	791533	114	10 (8.8%)
0	903988	58	5 (8.6%)

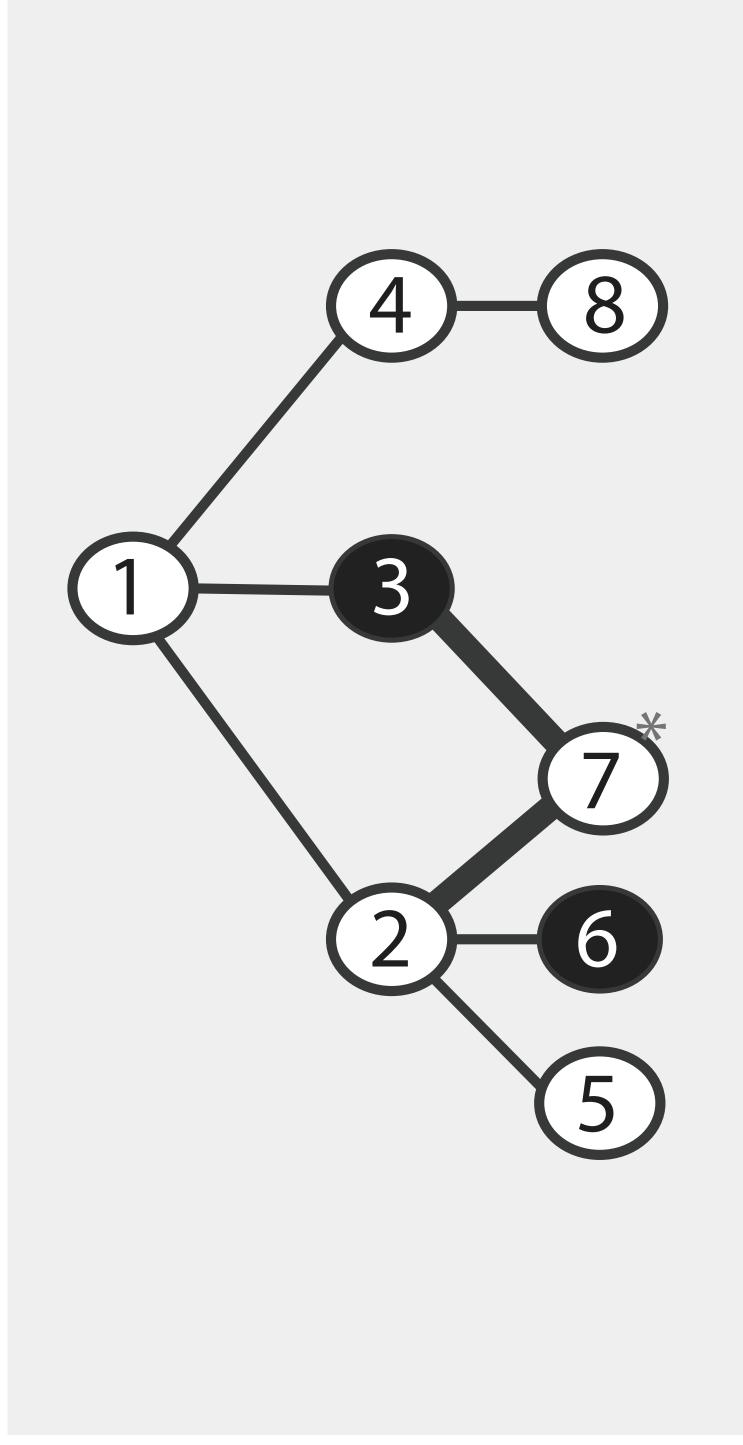


1. De-cycle and linearize graph

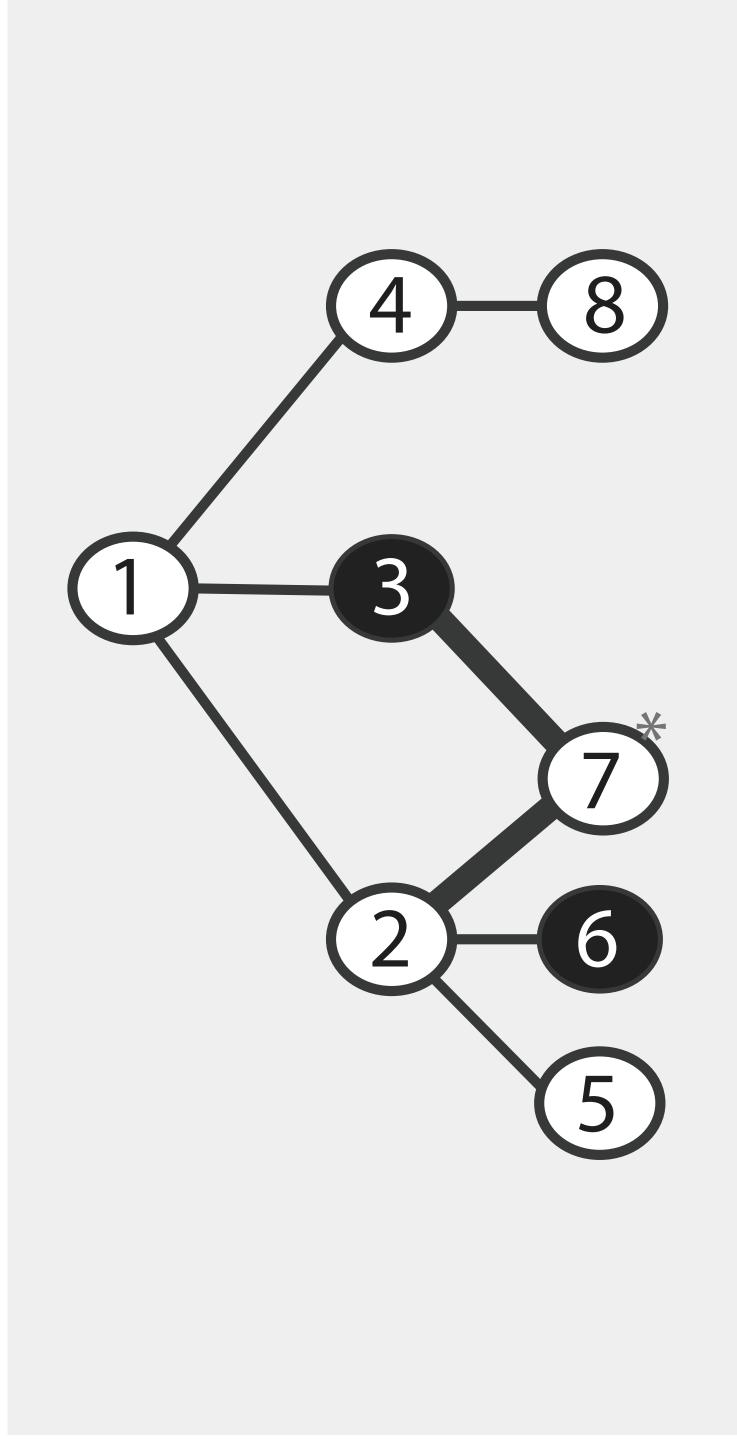


2. Plot attributes in table

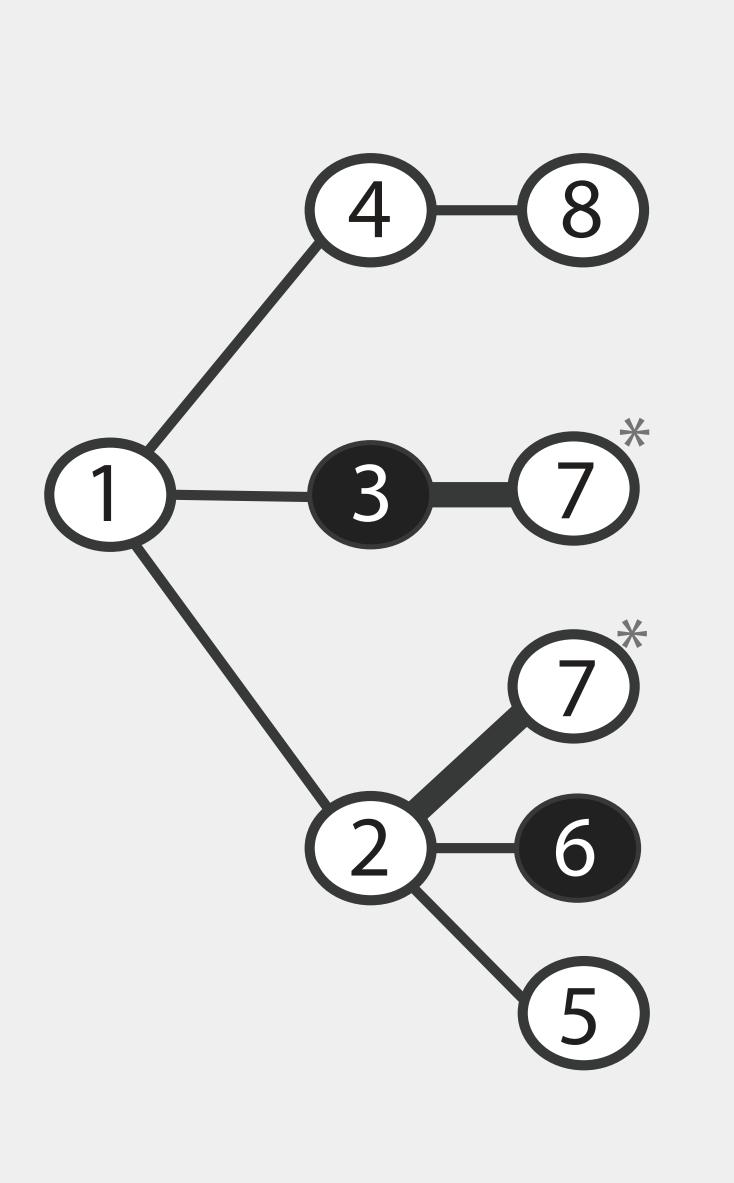
De-Cycling



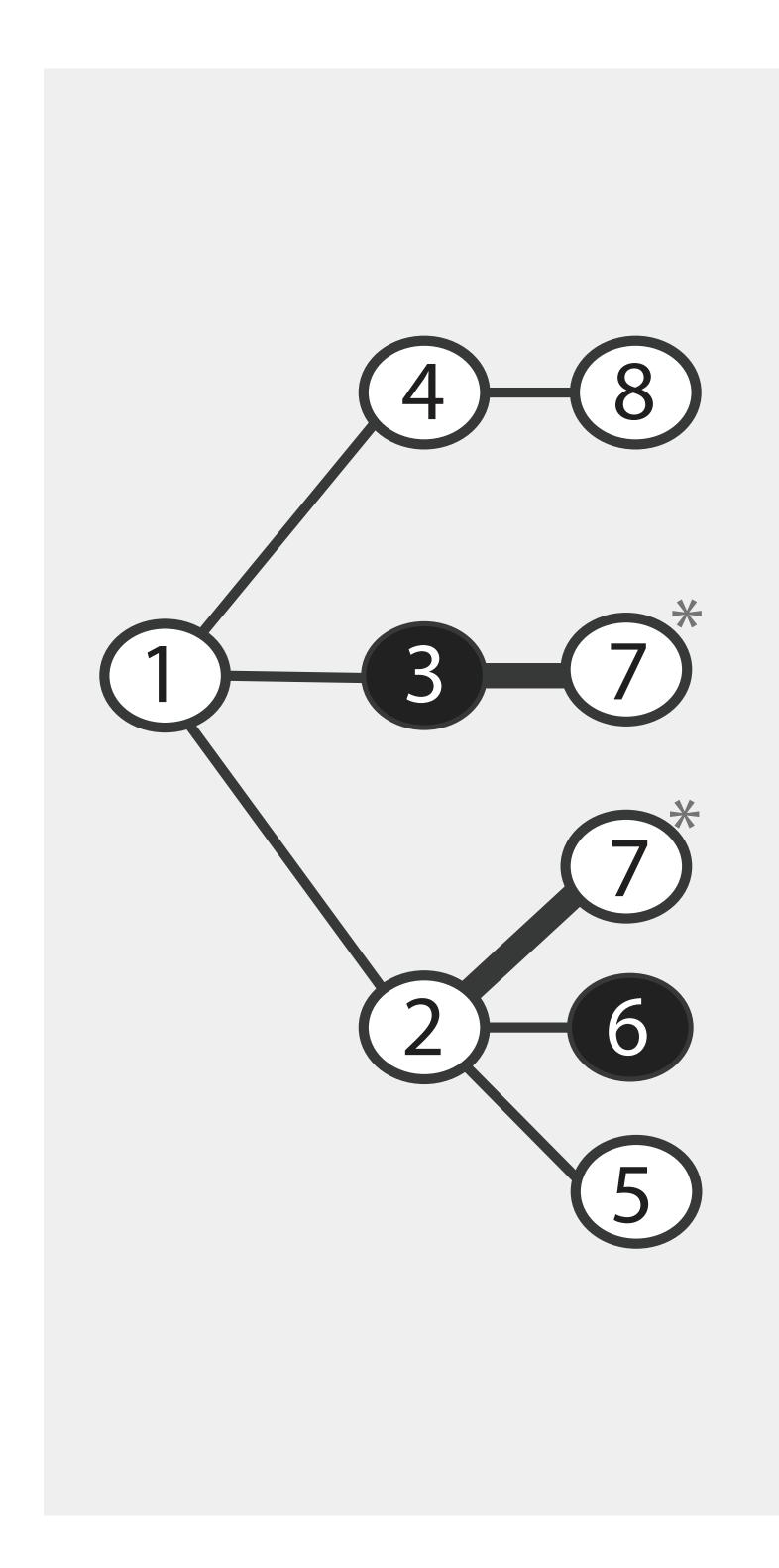
De-Cycling





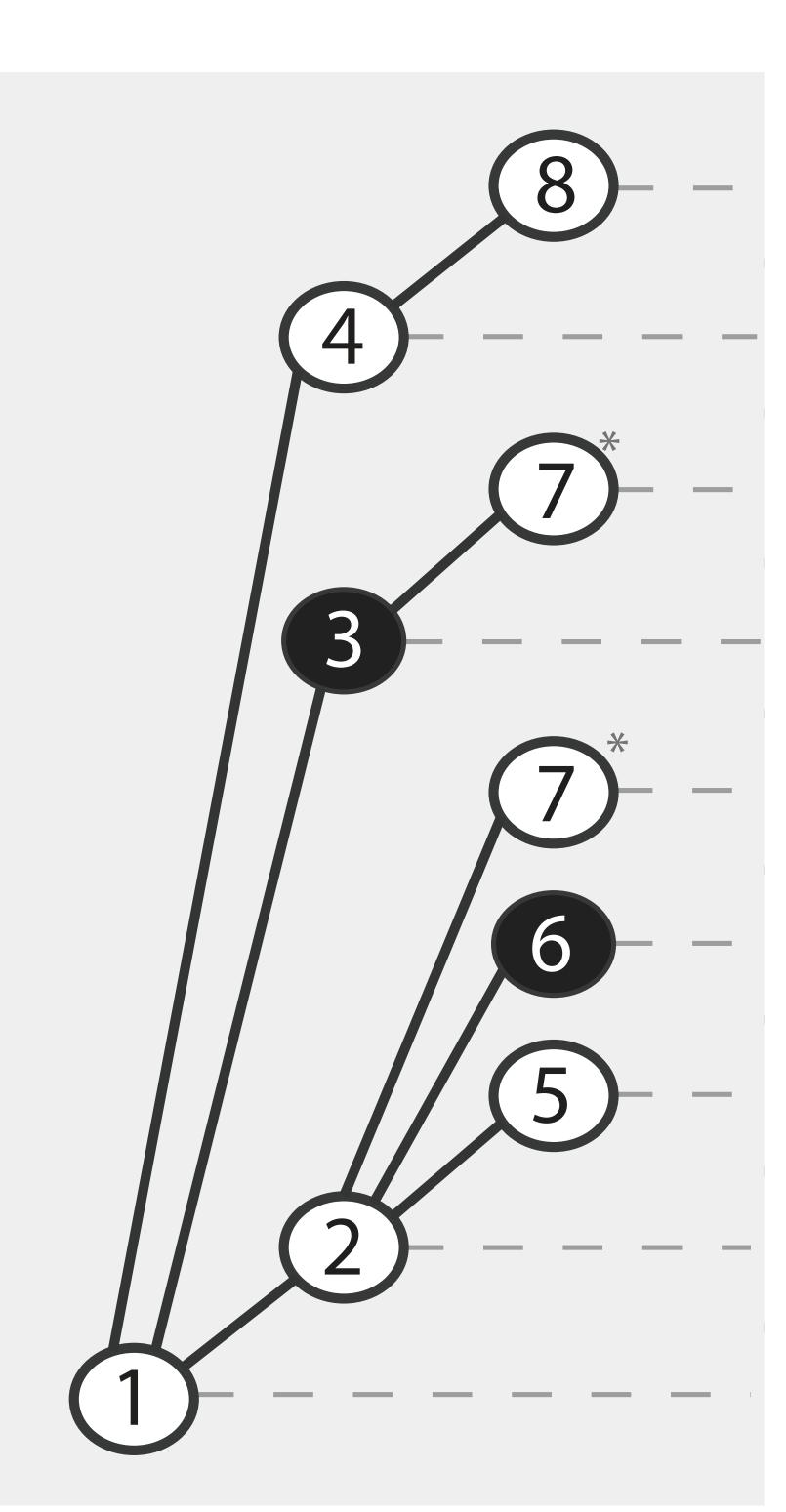


Linearization

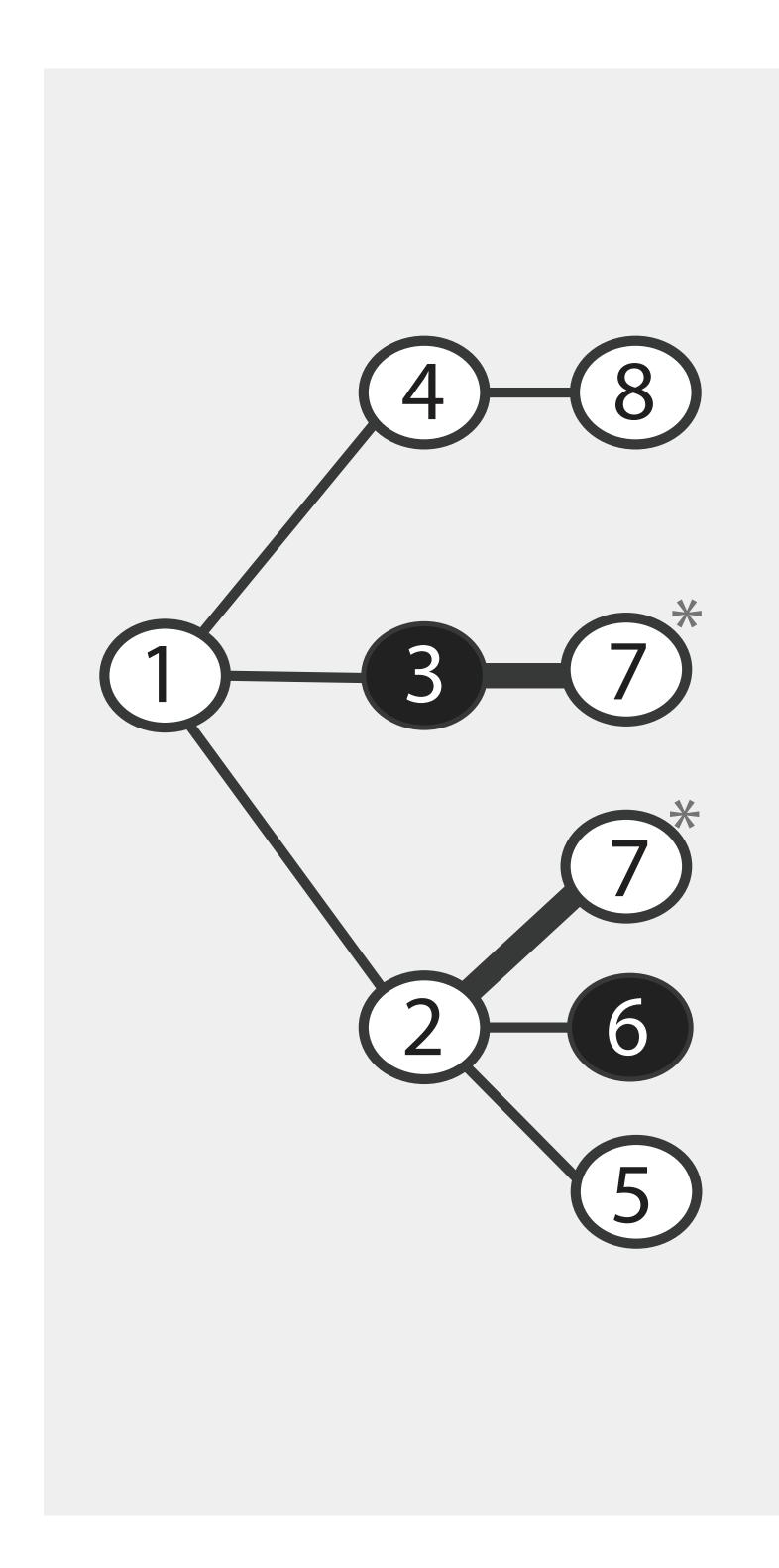




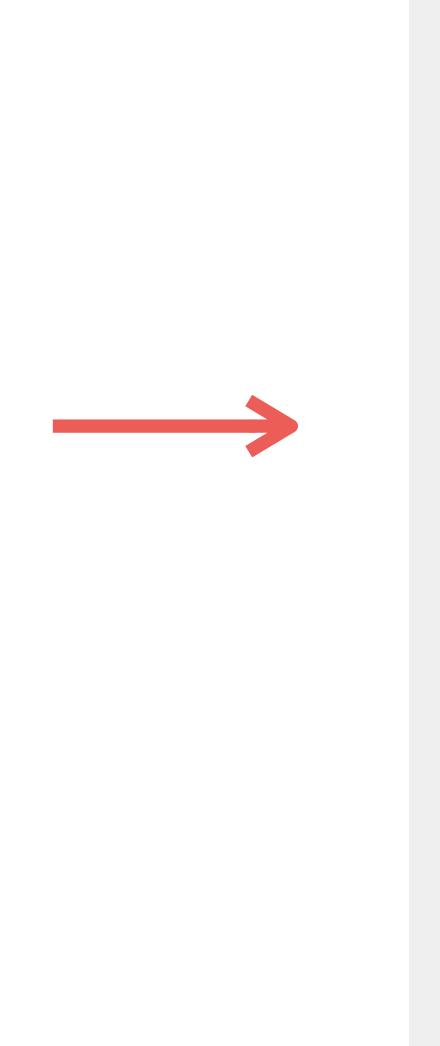


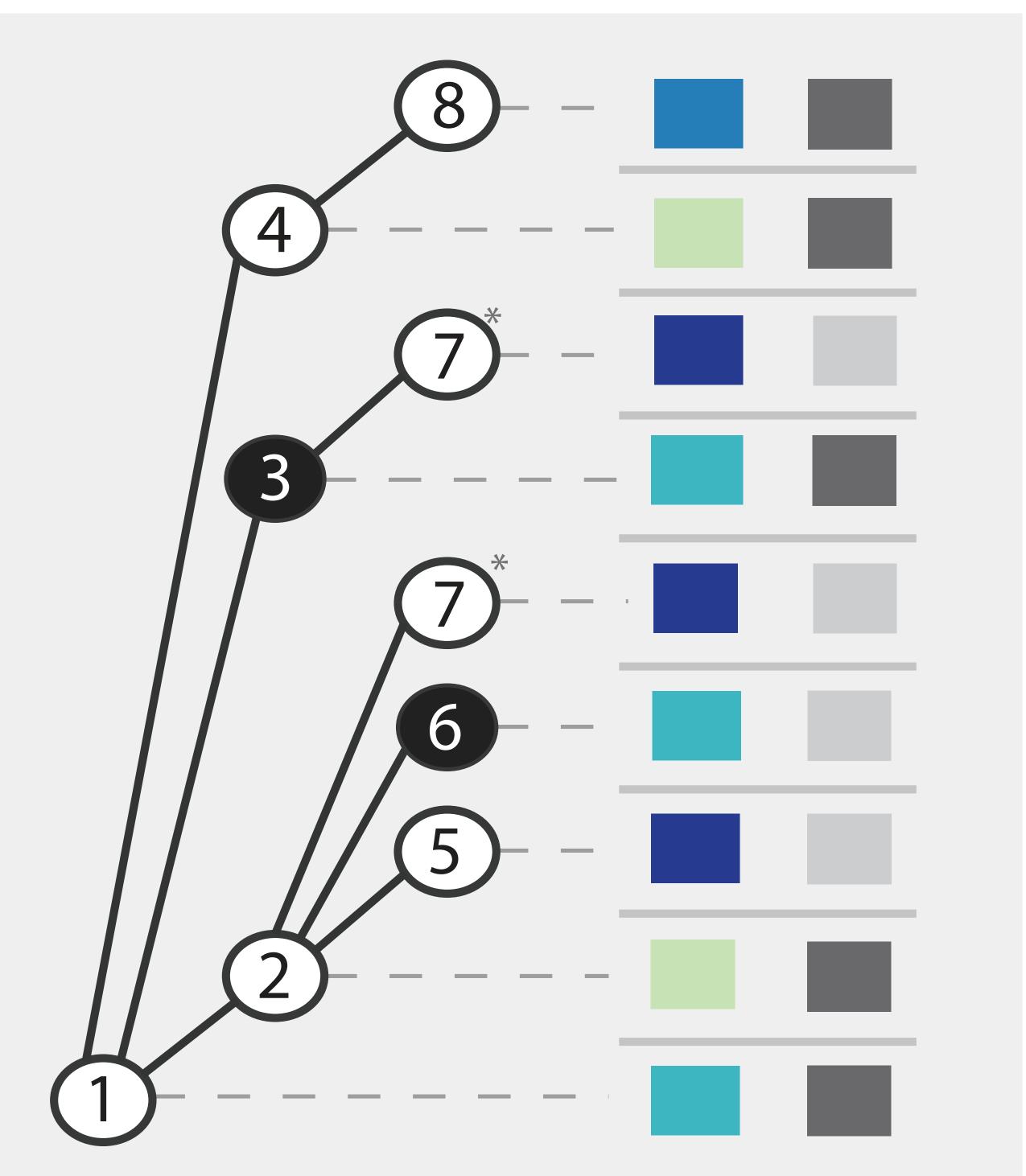


Linearization



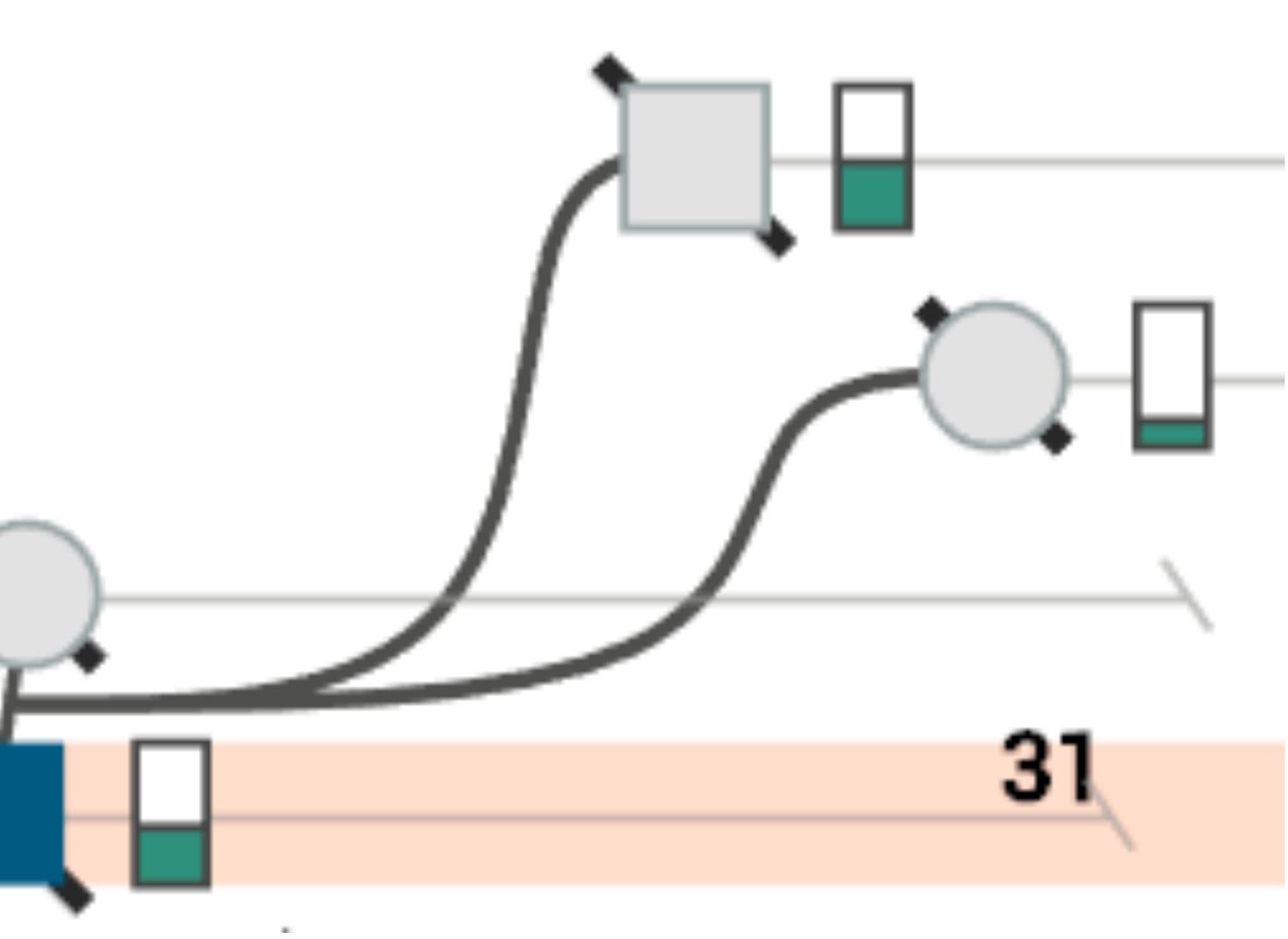




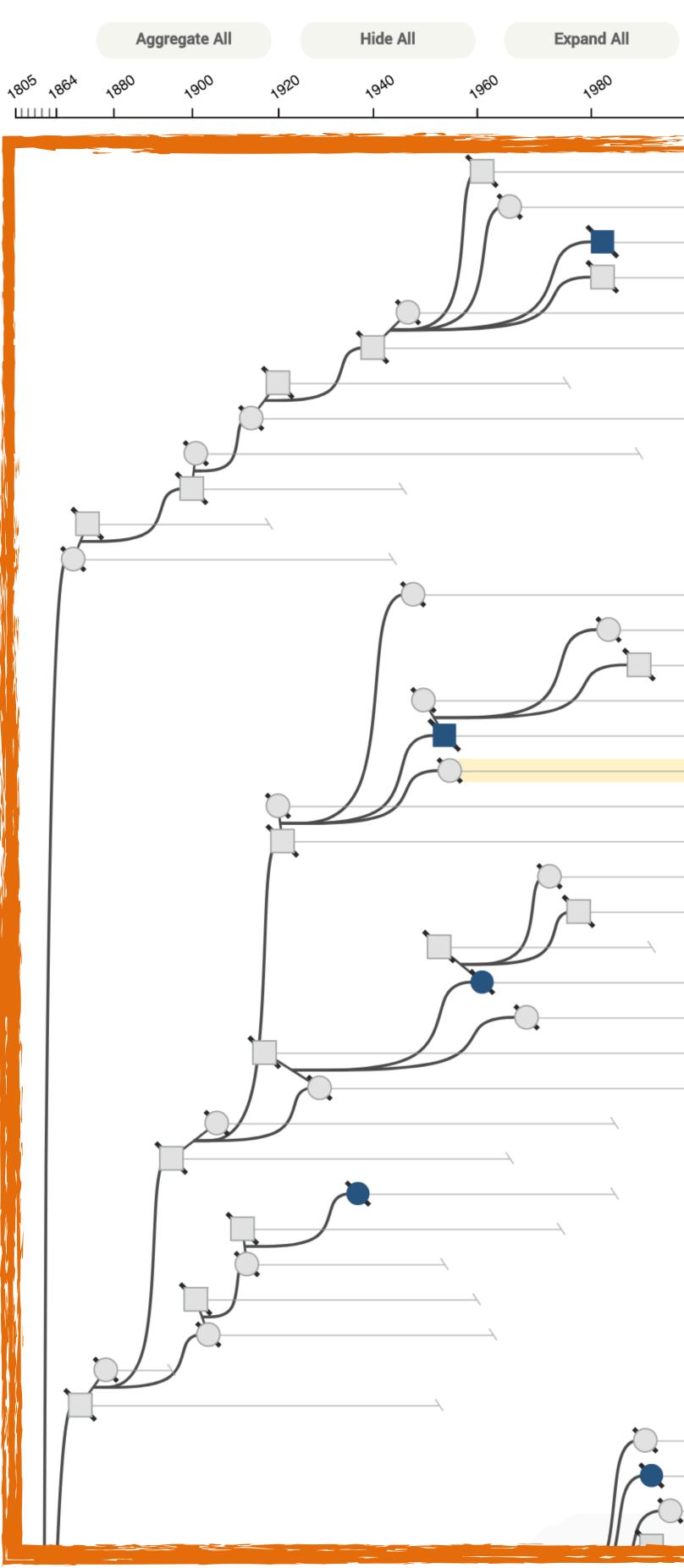


Basic Encoding





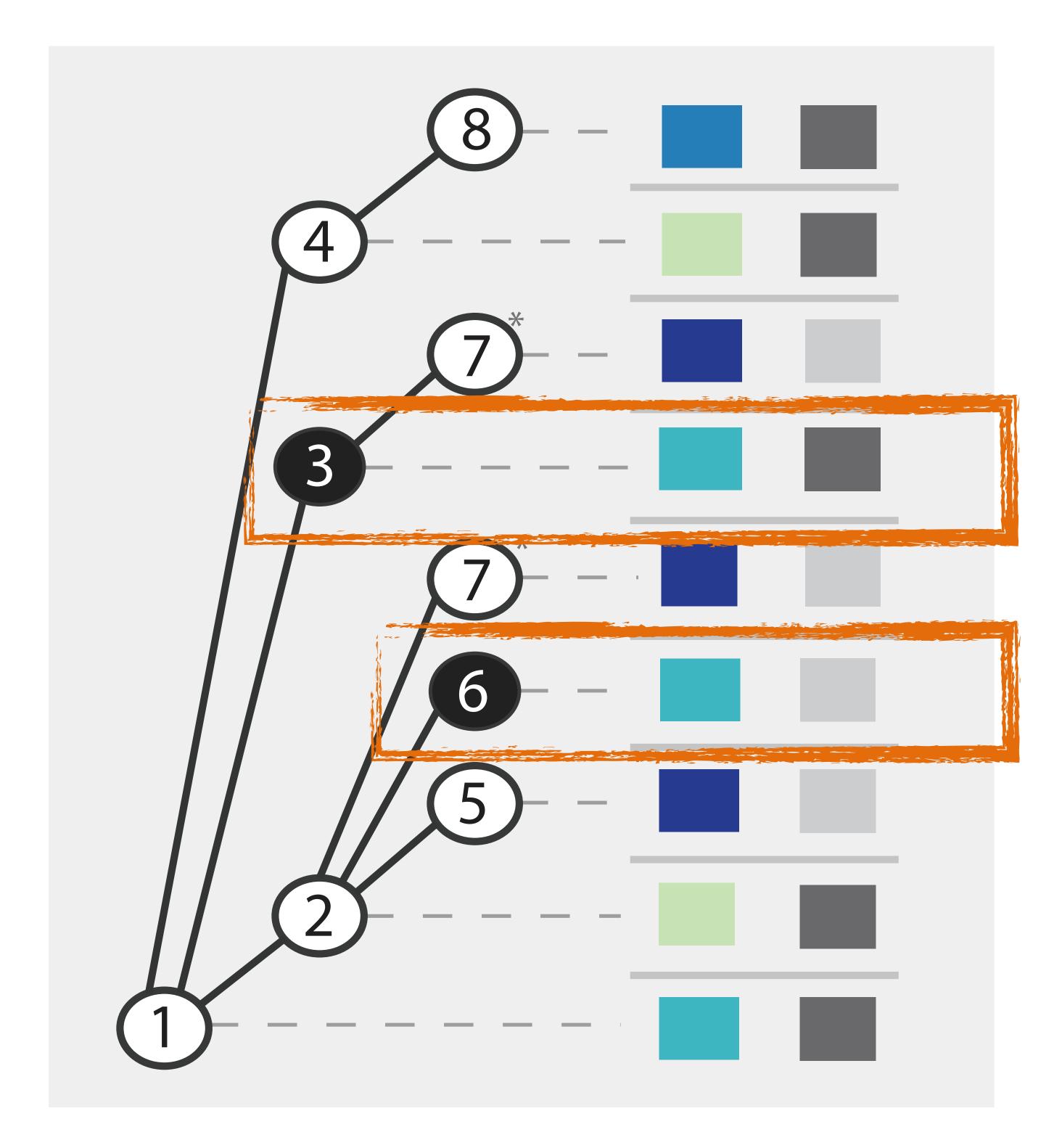
Can't show many people



					0	sed	16	ession (True) Age		Depression	9. Depressi Bipole	on In (True) Age 10	Bipolar Nr.Dia	Eipolar			white	other	Native	Elack)	Asian Hispan	ic
		KindredID	RelativeID	set	E) dece	asecuic	loc Depre	Age Total:121	Agen Total:2	D.Depite Nr.Dia Total:12	Bipole	Total:0	Nr.Dia Total:12	MaxBMI Total:8	AgeMaxBMI Total:8	1ace	1ace	(other)	tace	tace	Asian Hispan	cal
2000	2017			51%	60%	10%	2%				0.0%	rotul.o				<u>7%</u>	0.0%	0.0%	0.0%	0.0%	0.0%	
			▼▲	••	•	**	•	0 12	0 0	65 0	70	0 38	5 0 2	20 15 45	0 80				T .	V .	V	
	1	38	#781	а.	2	2	-	•			-					-	-	-	-	-	_	
		38	#784 #56604	71	÷.	Ξ.		•									_	_	_	_	_	gunchot
`		38	#5444		11	Ξ.		•									_	_	_	_	-	gunshot
		38	#778	Π.		1	_	•			_					_	_	_	_	_	_	
	1	38	#777	Ē.			-	٠			-					-	-	-	_	-	-	
	1	38	#11574				-	٠			-					-	-	-	-	-	-	
	1	38	#11575				-	•	L		-					-	-	-	-	-	-	
	1	38	#23088				-	٠	■		-					-	-	-	-	-	-	
	1	38	#23087				-	•			-					-	-	-	-	-	-	
	1 -	38	#37	_			-	•	I. —		-					-	-	-	-	-	-	
	1	38	#35				-	•			-					-	-	-	-	-	-	
		38	#66559			-	-	•			-					-	-	-	-	-	-	
		38	#4033	2.		-	-	•			-					-	-	-	-	-	-	
		38	#41582	-			-	•	_		-					-	-	-	-	-	-	
		38	#65051	21	-	2	-	•			-					-	-	-	-	-	-	
	62	38	#66561					•						•	•		-	-	-	-		gunshot
\ \		38	#66560	Ξ.				•								-	_	_	_	_	_	
``		30	#53797 #53798	71	11		_	•			_						_	_	_	_	_	
		38	#9348	Ξ.	10	1	_	•	N		_					_	_	_	_	_	_	
	1	38		Ξ.	10		-	•			_					_	_	_	_	_	_	
	1		#24070				-	•	M		_					_	_	-	_	_	-	
	1 1 -	38	#61046					٠	•	٠				•	•		-	_	-	_	-	mixed d
	- 1 -	38	#47326				-	•			-					-	-	-	-	-		
\rightarrow	1 -	38	#22124				-	•			-					-	-	-	-	-	-	
	1	38	#61047				-	•			-					-	-	-	-	-	-	
	1	38	#66556				-	٠			-					-	-	-	-	-	-	
	1	38	#66557				-	•	L		-					-	-	-	-	-	-	
	1 -	38	#54490					•	L							-	-	-	-	-	-	
	1	38	#34847				-	•	<u> </u>		-					-	-	-	-	-	-	
		38	#39657			-	-	•			-					-	-	-	-	-	-	
	1	38	#705				-	•			-					-	-	-	-	-	-	
A	1	38	#28375	21			-	•			-					-	-	-	-	-	-	
		38	#34	21			-	•			-					-	-	-	-	-	-	
		38	#32				-	•			_						-	-	-	-	_	
		38	#59713	Ξ.			-				-						_	_	-	_		au un a h- a
		38	#42257 #17565				-	•								-	_	-	_	_	_	gunshot
		38		Ξ.	10																	

Lots of missing data



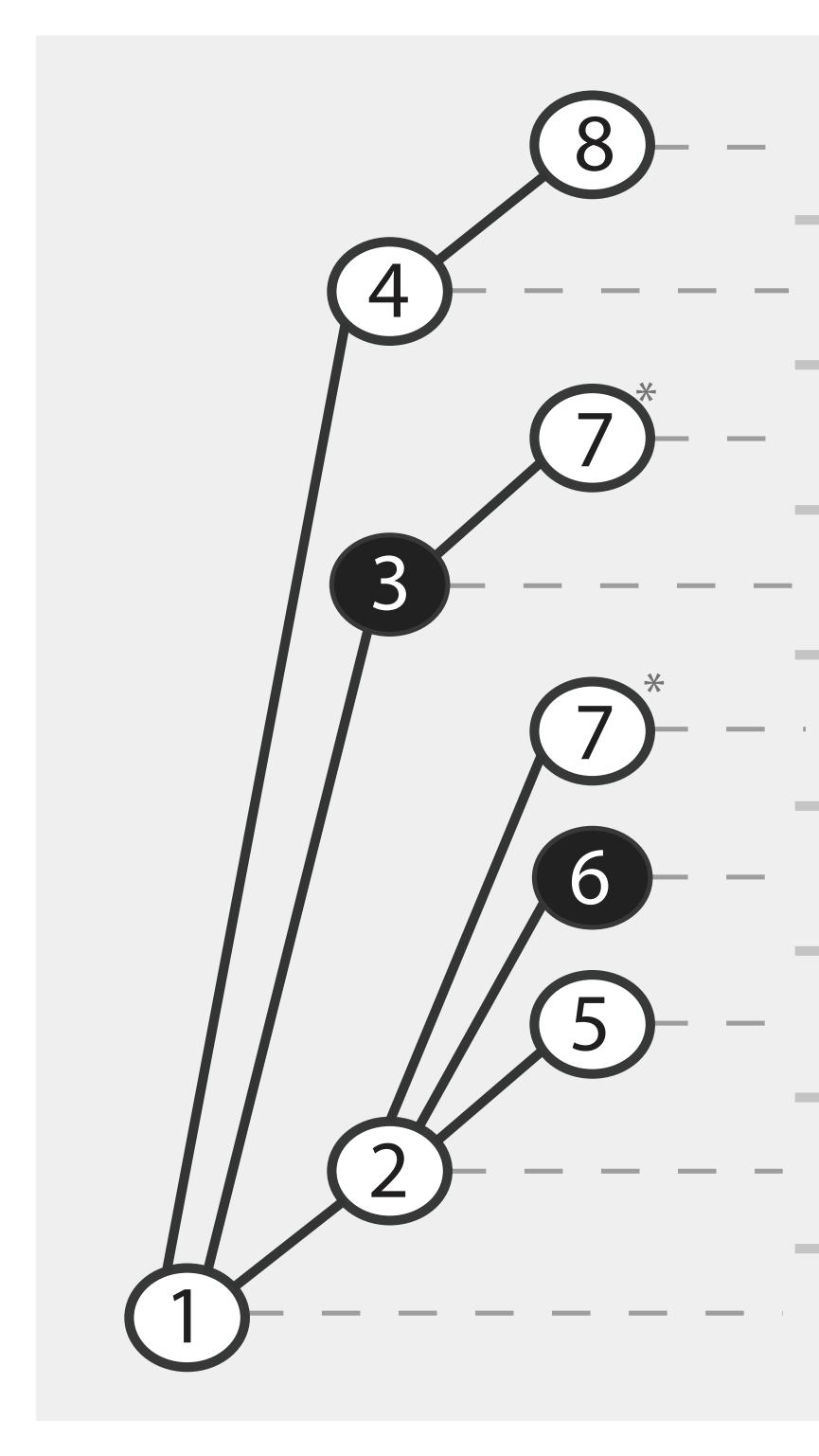




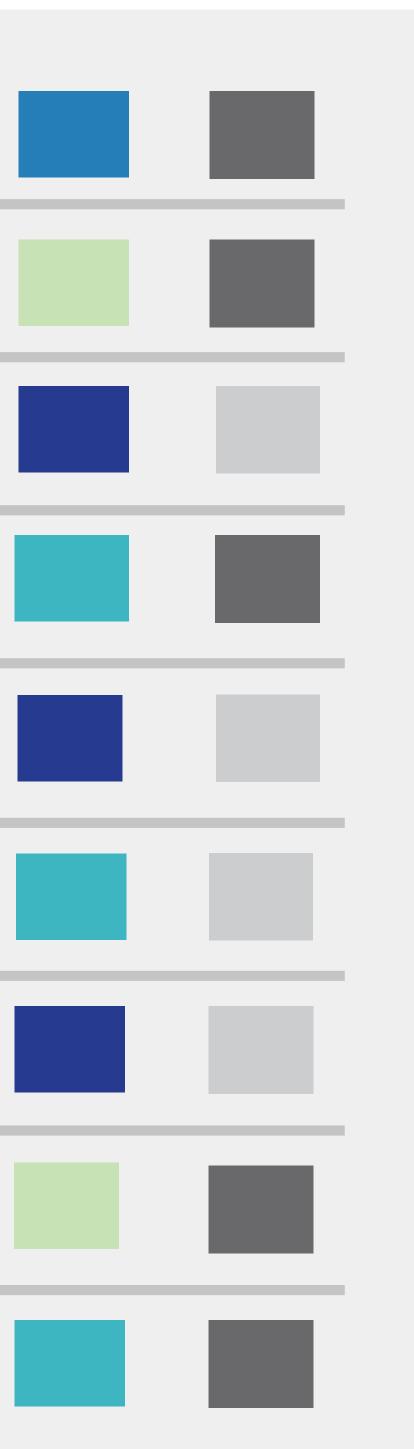
People of Interest



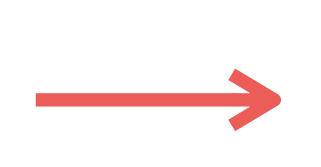


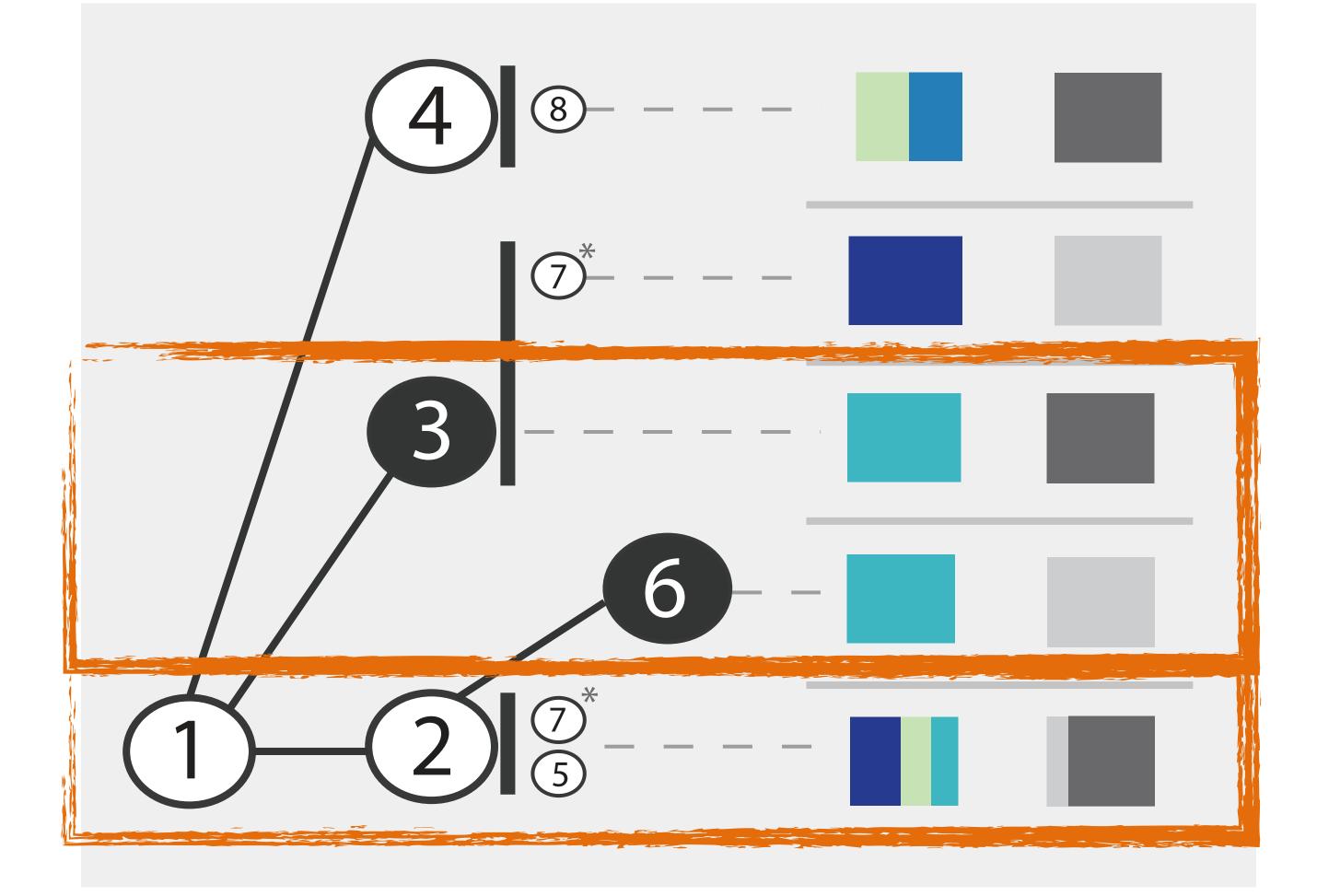






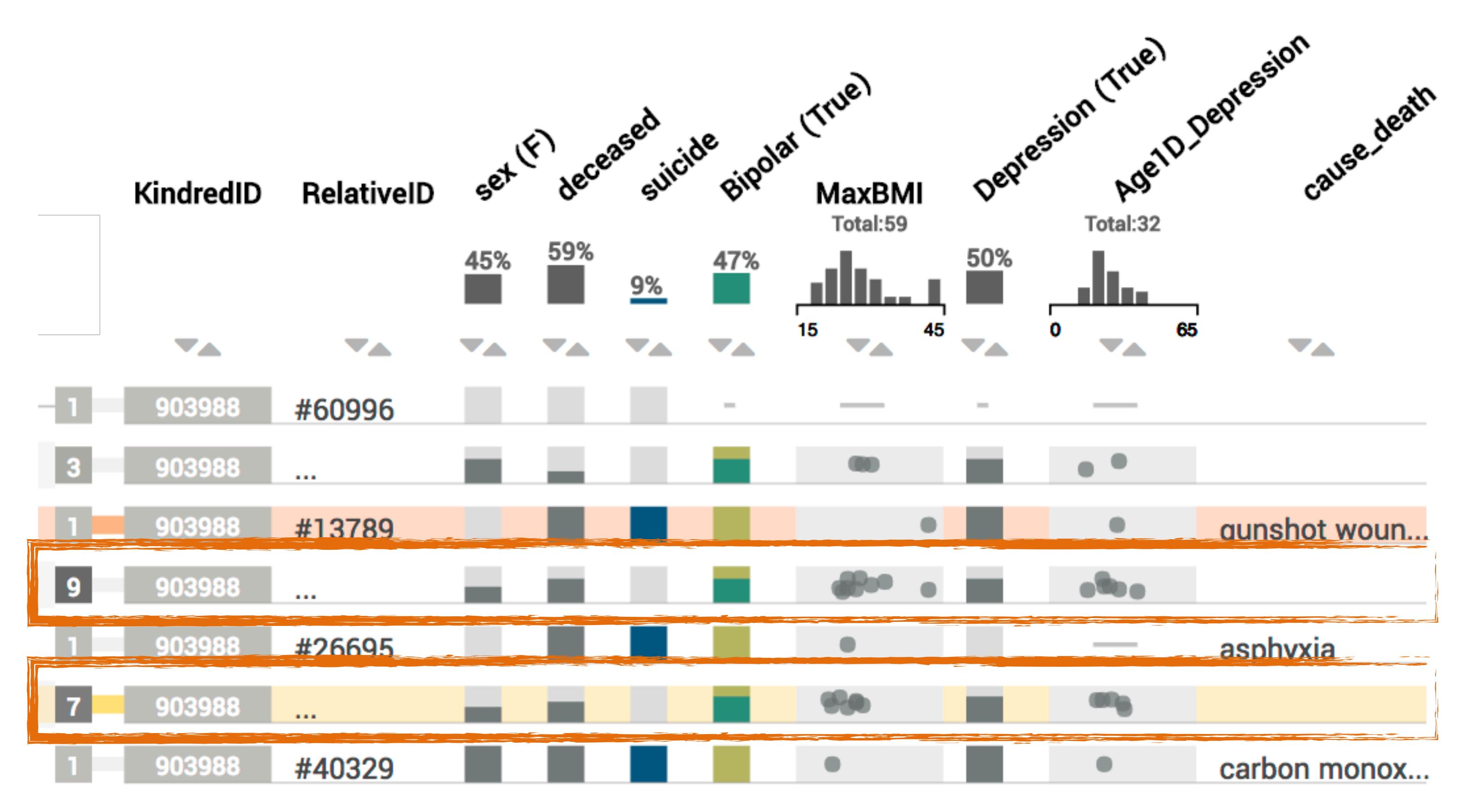
One row for every person of interest



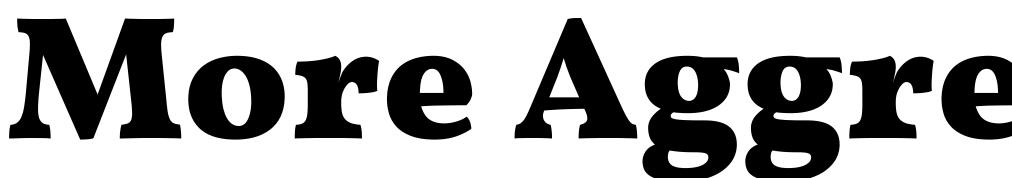


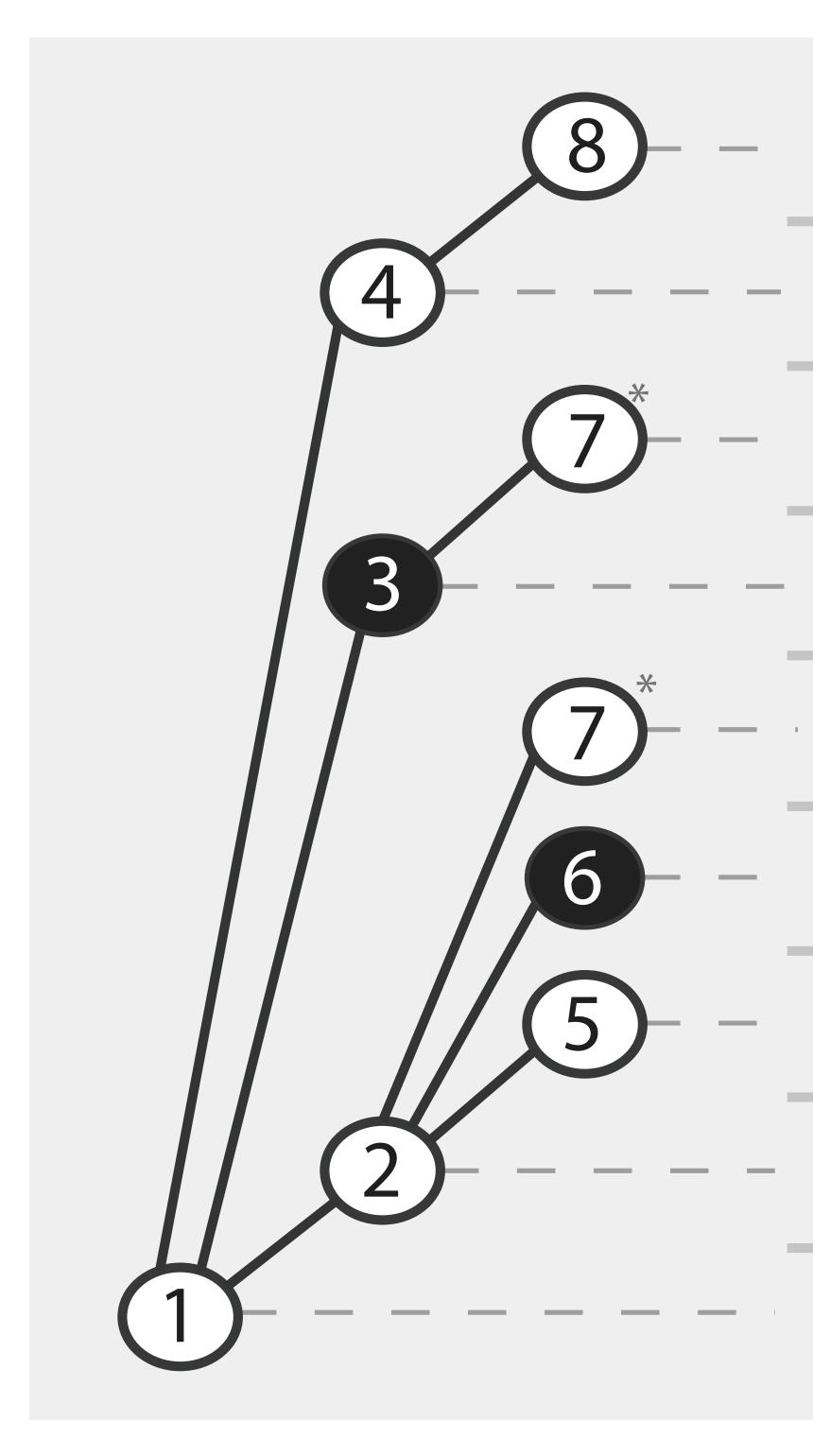


Others have to share a row

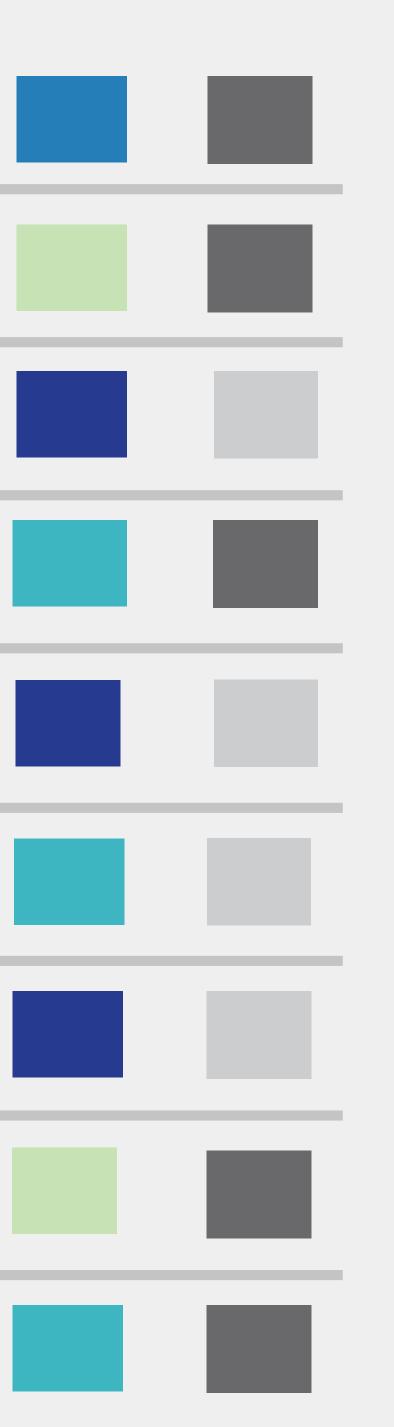


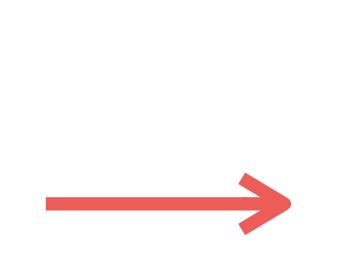
Aggregated Rows

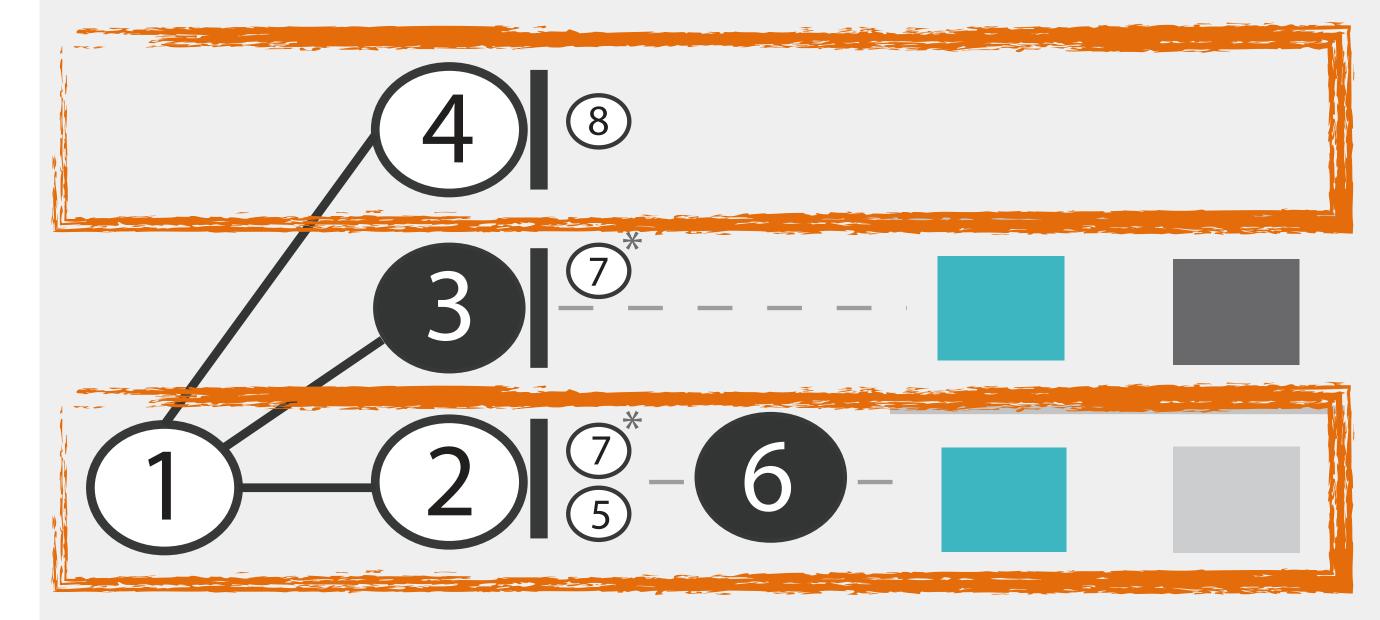




More Aggressive: Hiding





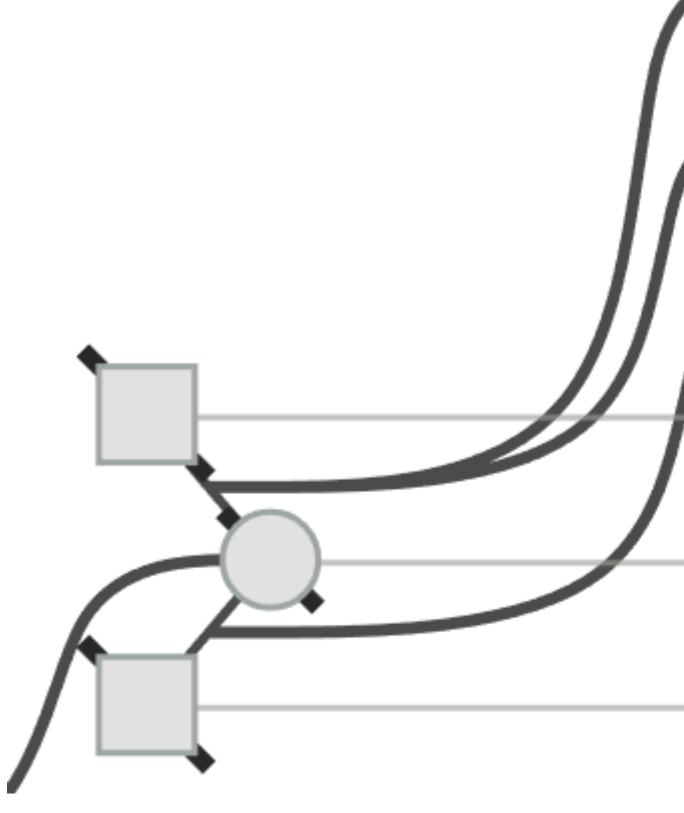


Empty row

Only data for #6 shown

Implicit Encoding of Family

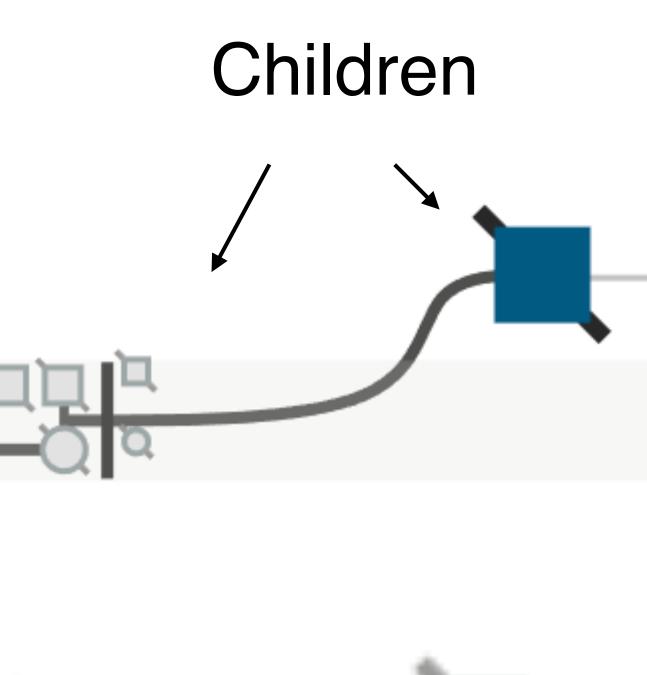
No Aggregation

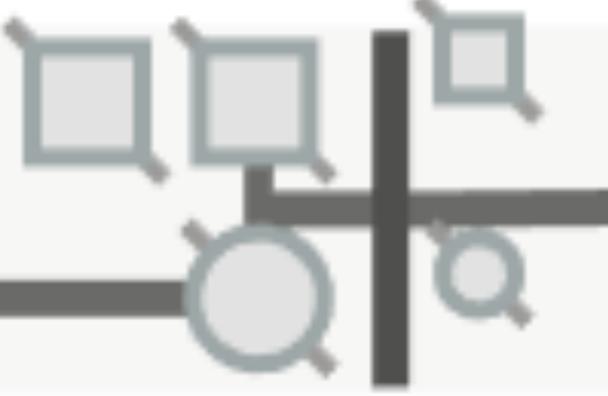


Parents



Aggregation









Find families enriched for a trait

Scan relevant cases for relationships

Show variants and other genetic data Figure mode **Extend to other datasets UPDB Users: Cancer, other psychiatric,** cardiovascular, etc. Genealogical datasets becoming more common Quantitative analysis of population-scale family trees

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,...

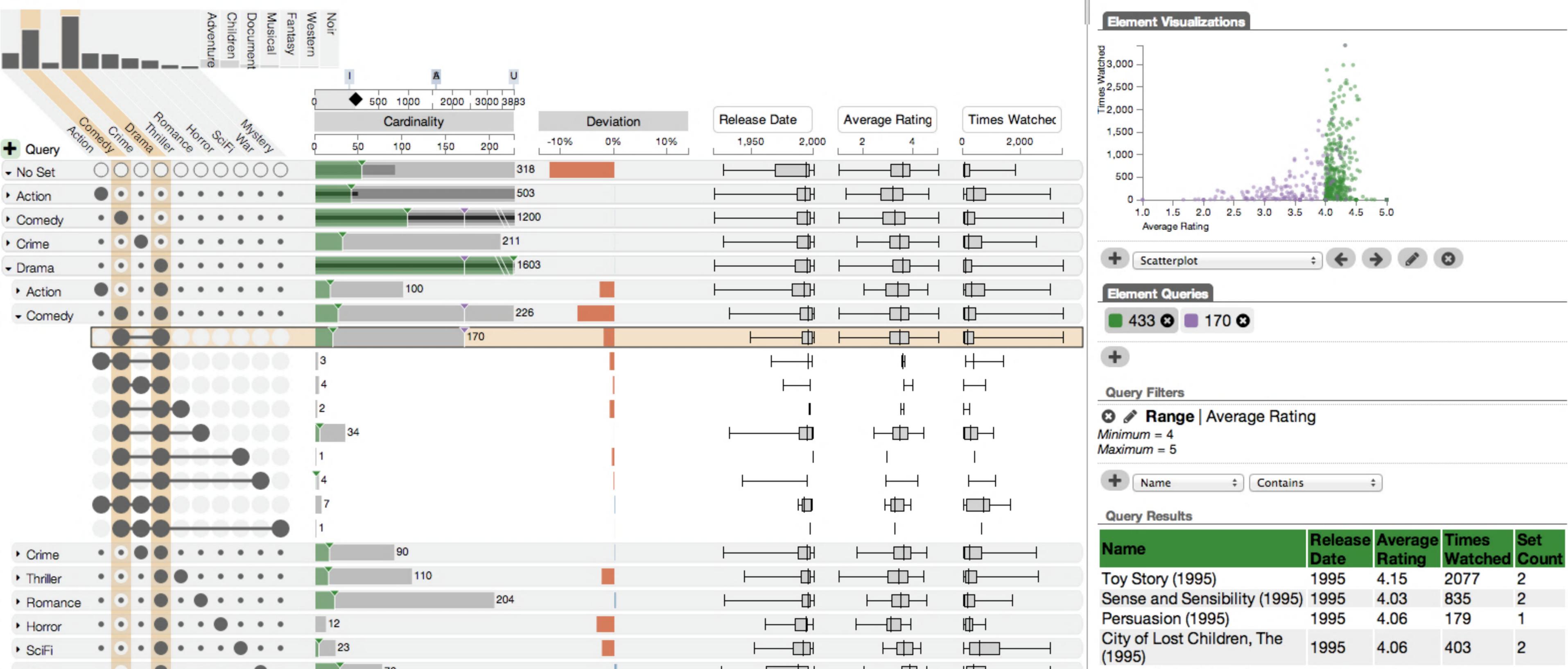
with millions of relatives







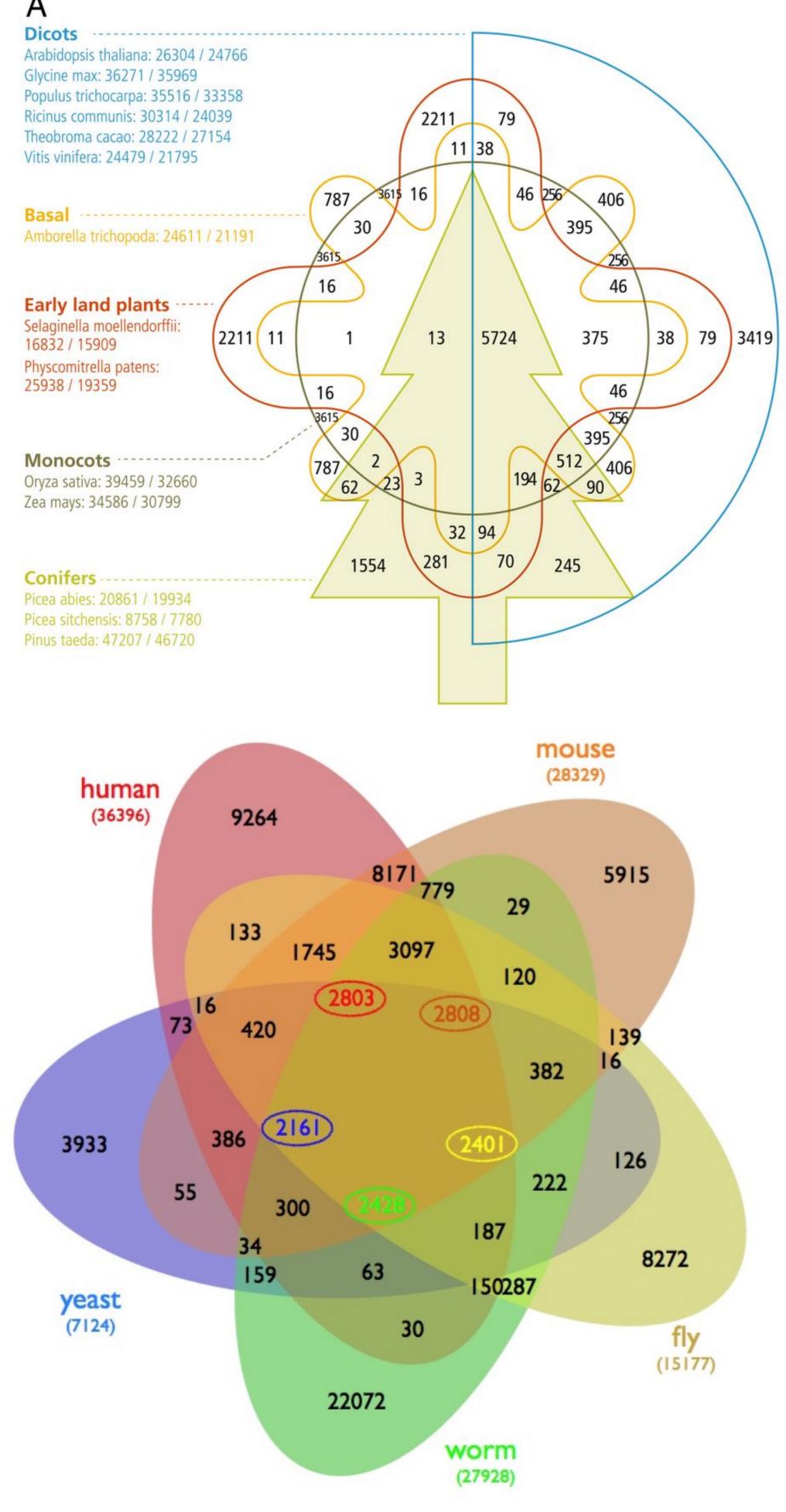
[InfoVis'14]



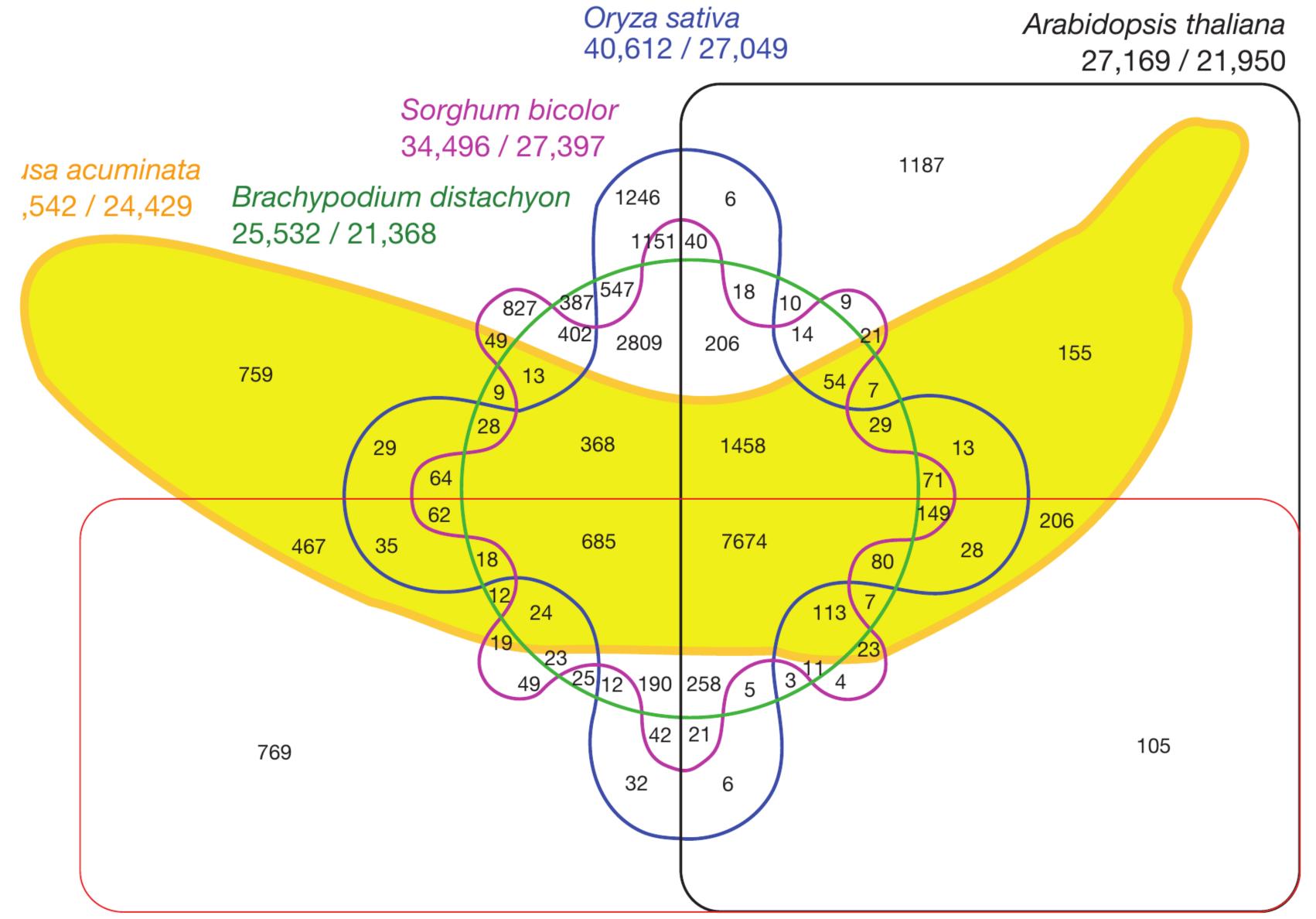
Visualizing Intersecting Sets



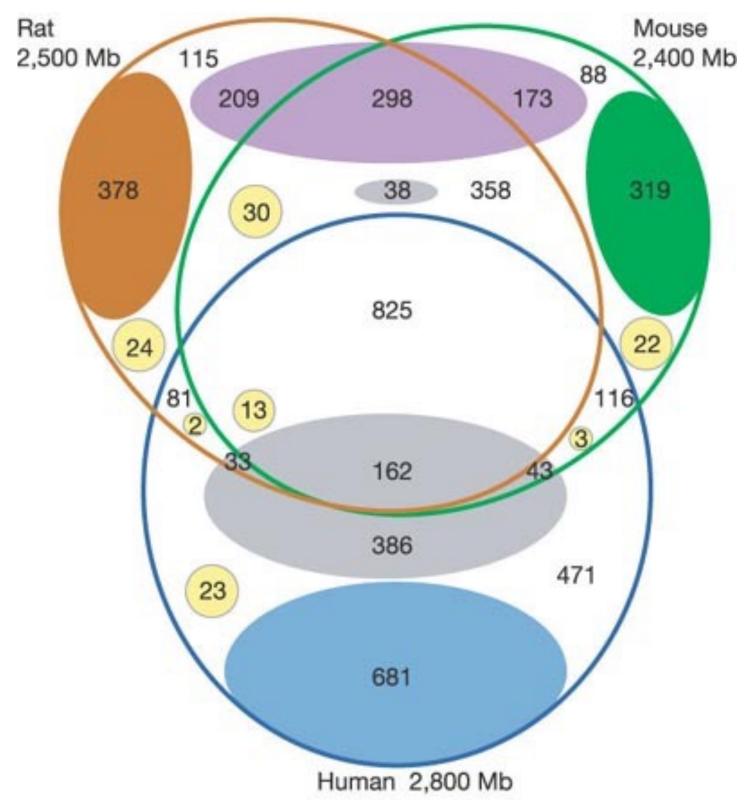
[Neale et al., BMC Genome Biology, 2014]

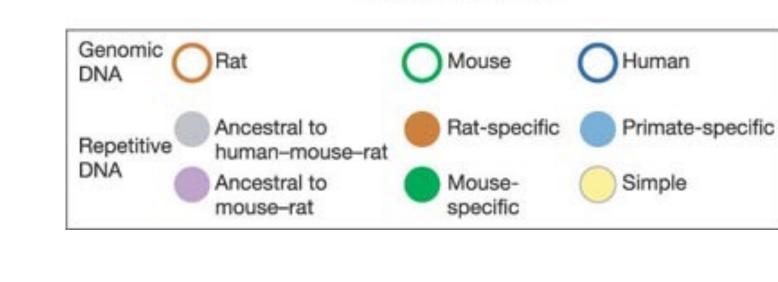


[Wiles et al., BMC Systems Biology]



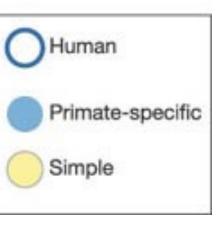
Phoenix dactylifera 28,889 / 19,027



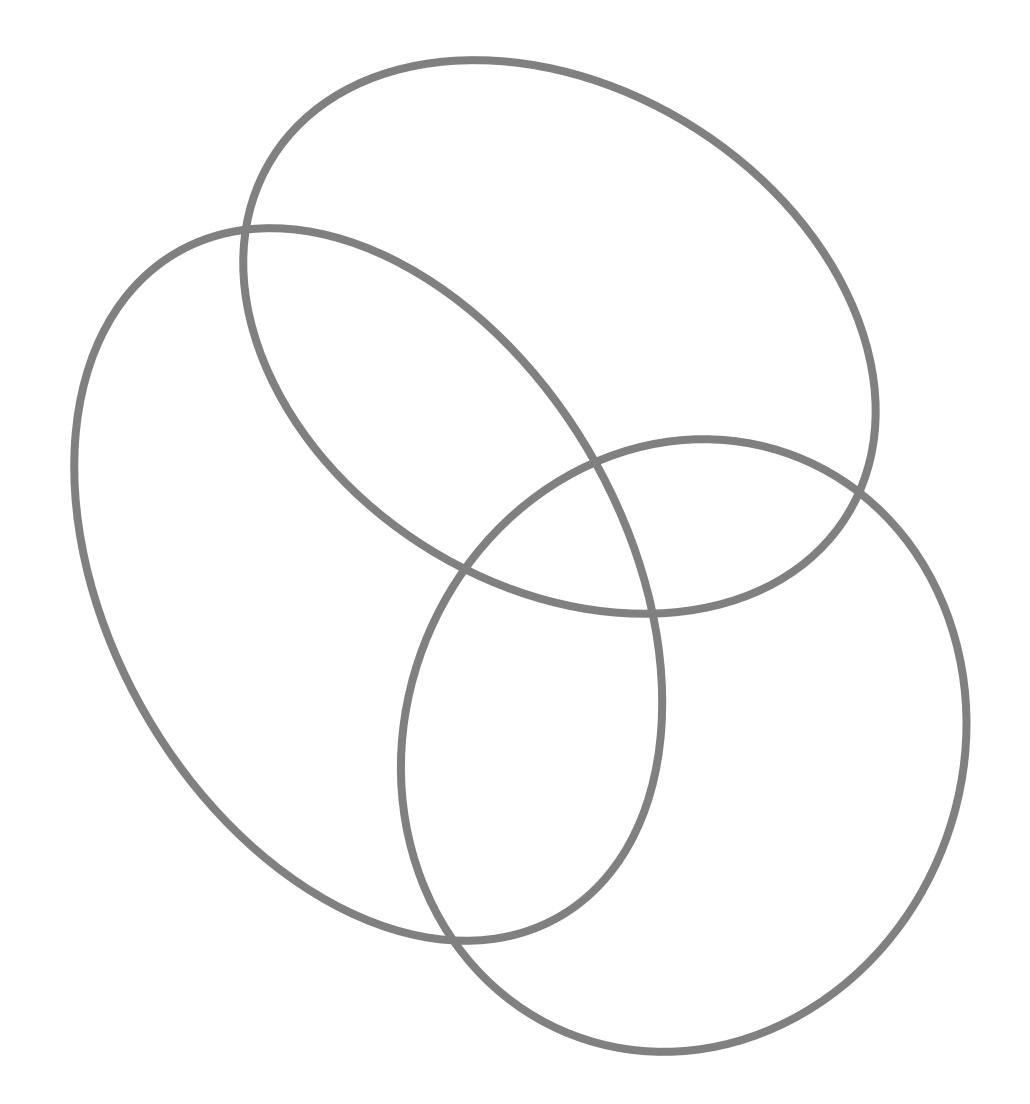


[Gibbs et al., Nature, 2004]

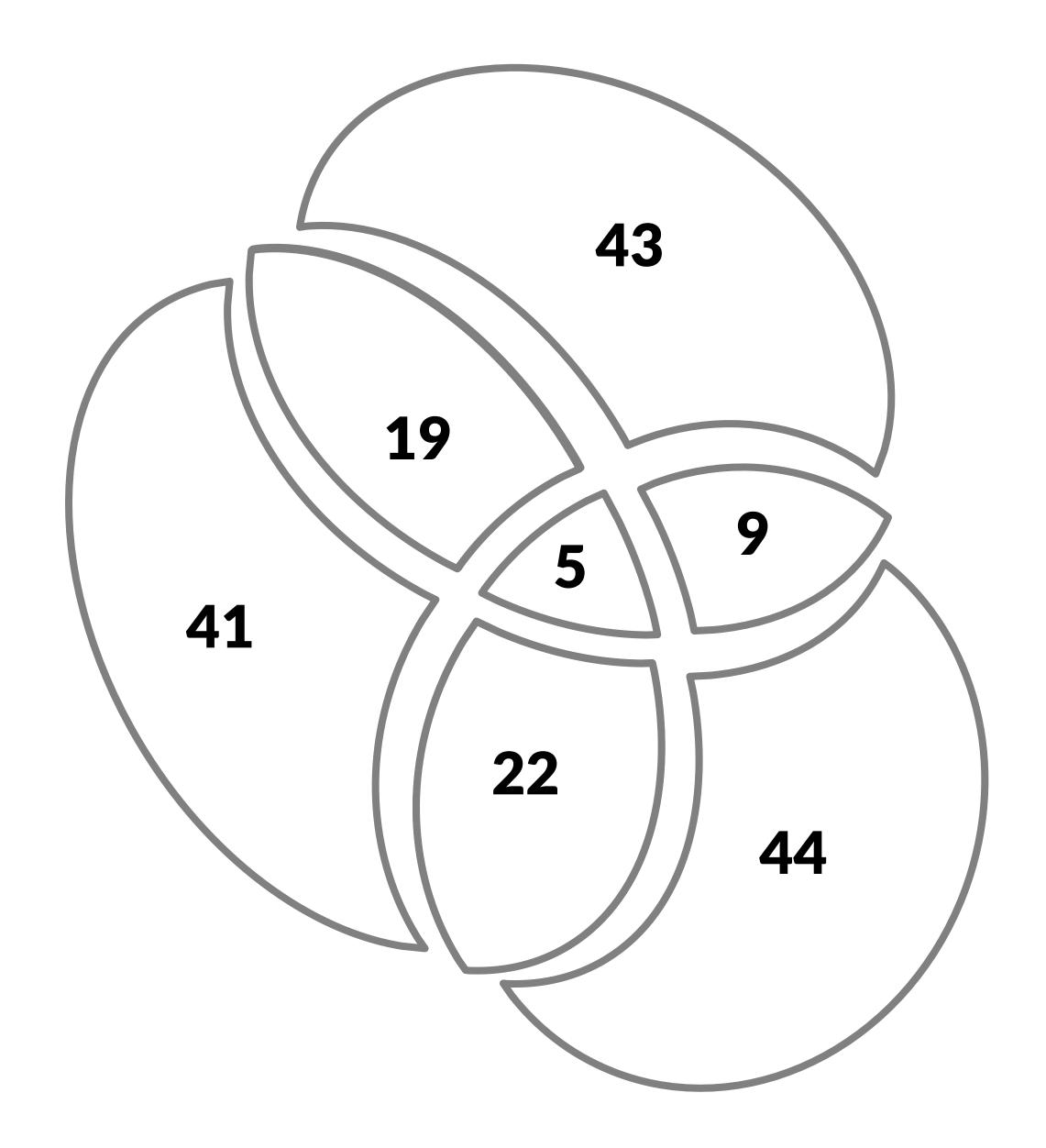
[D'Hont et al., Nature, 2012]



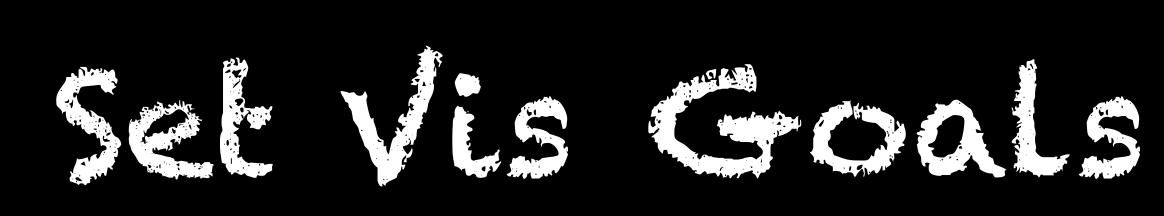




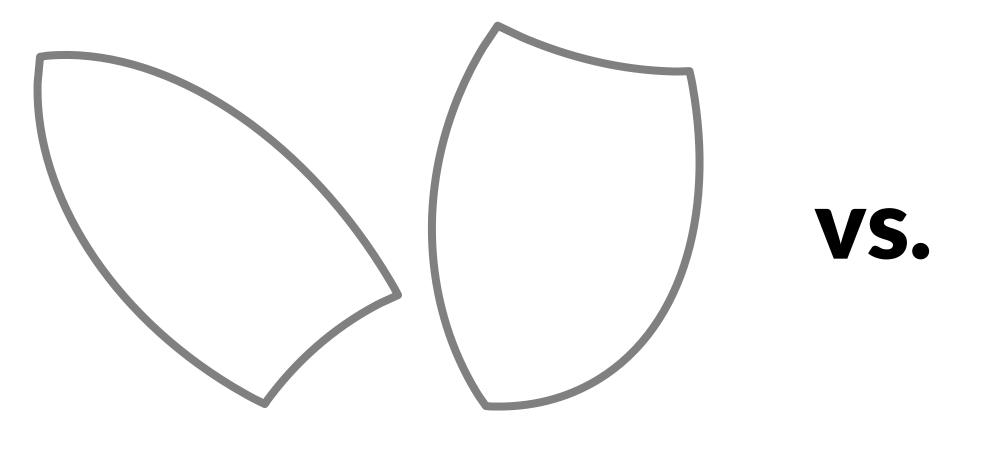
[created with EulerAPE]



[created with EulerAPE]

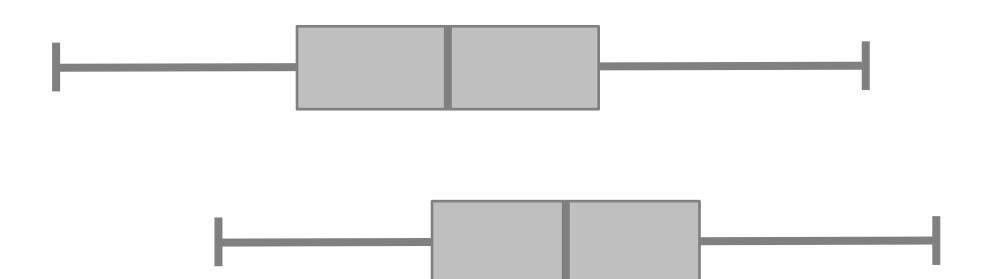


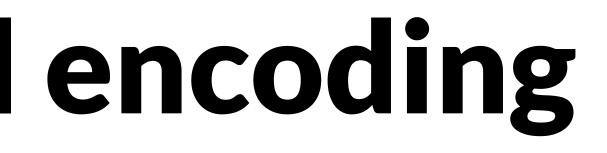
1. Efficient visual encoding



2. Creating complex slices of a dataset

3. Visualize attributes

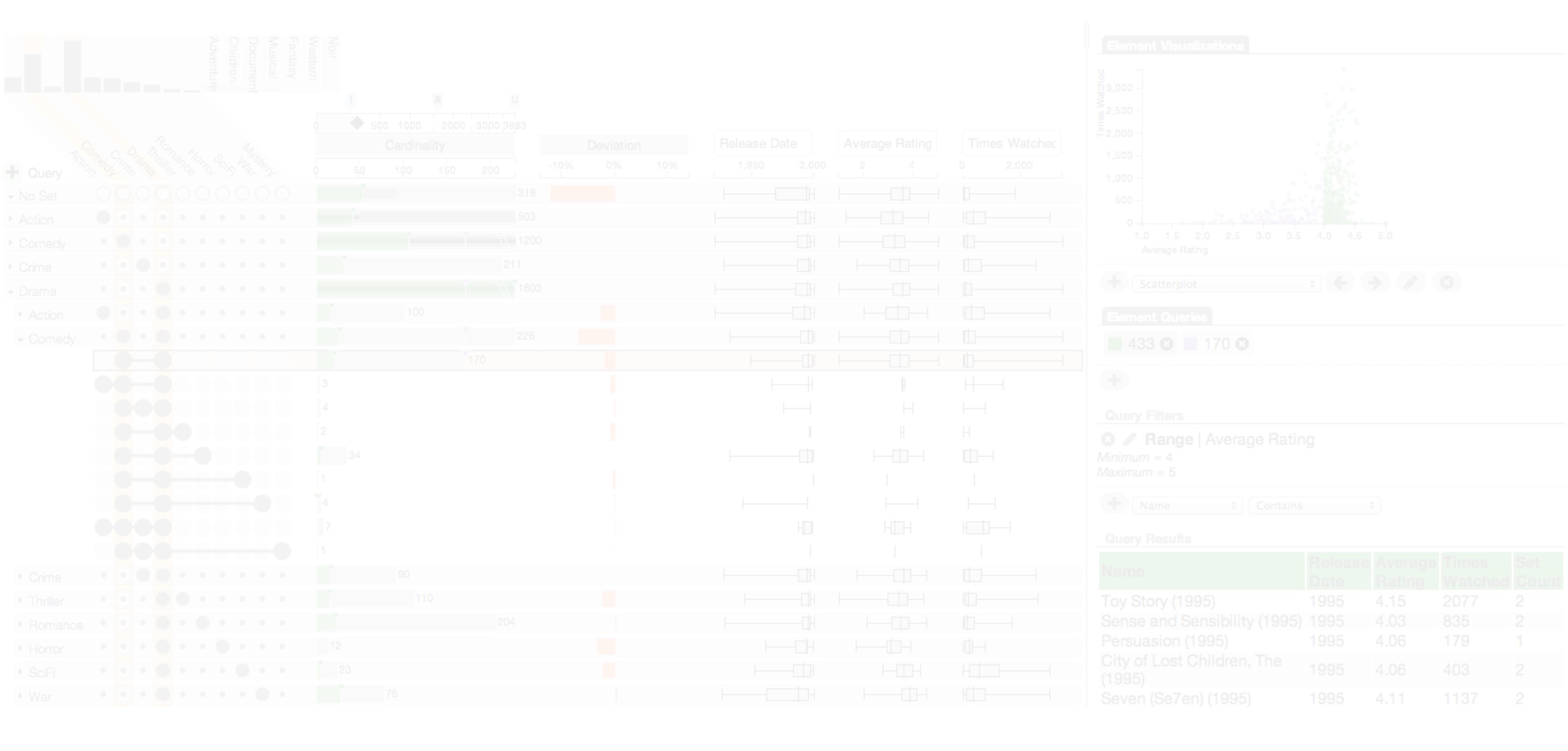








[Movie Lens Dataset]



Visualizing Intersections

Visualizing Properties

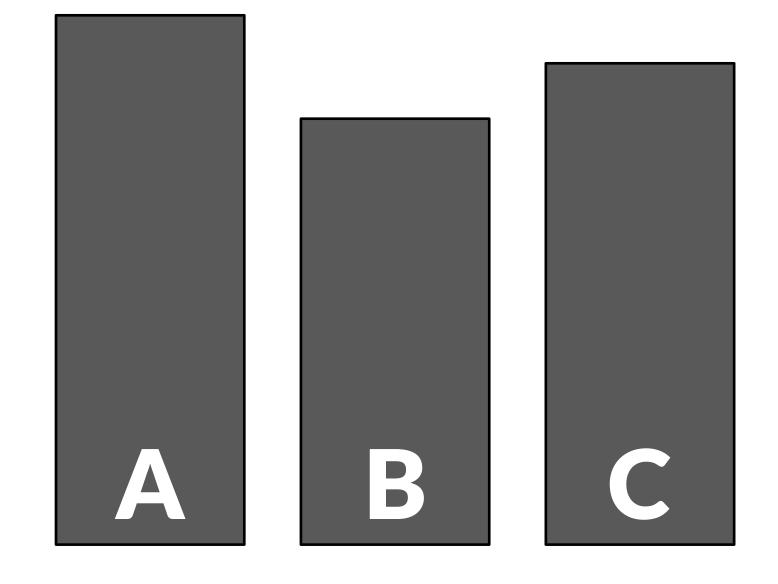
Attribute Details

Element List & Queries

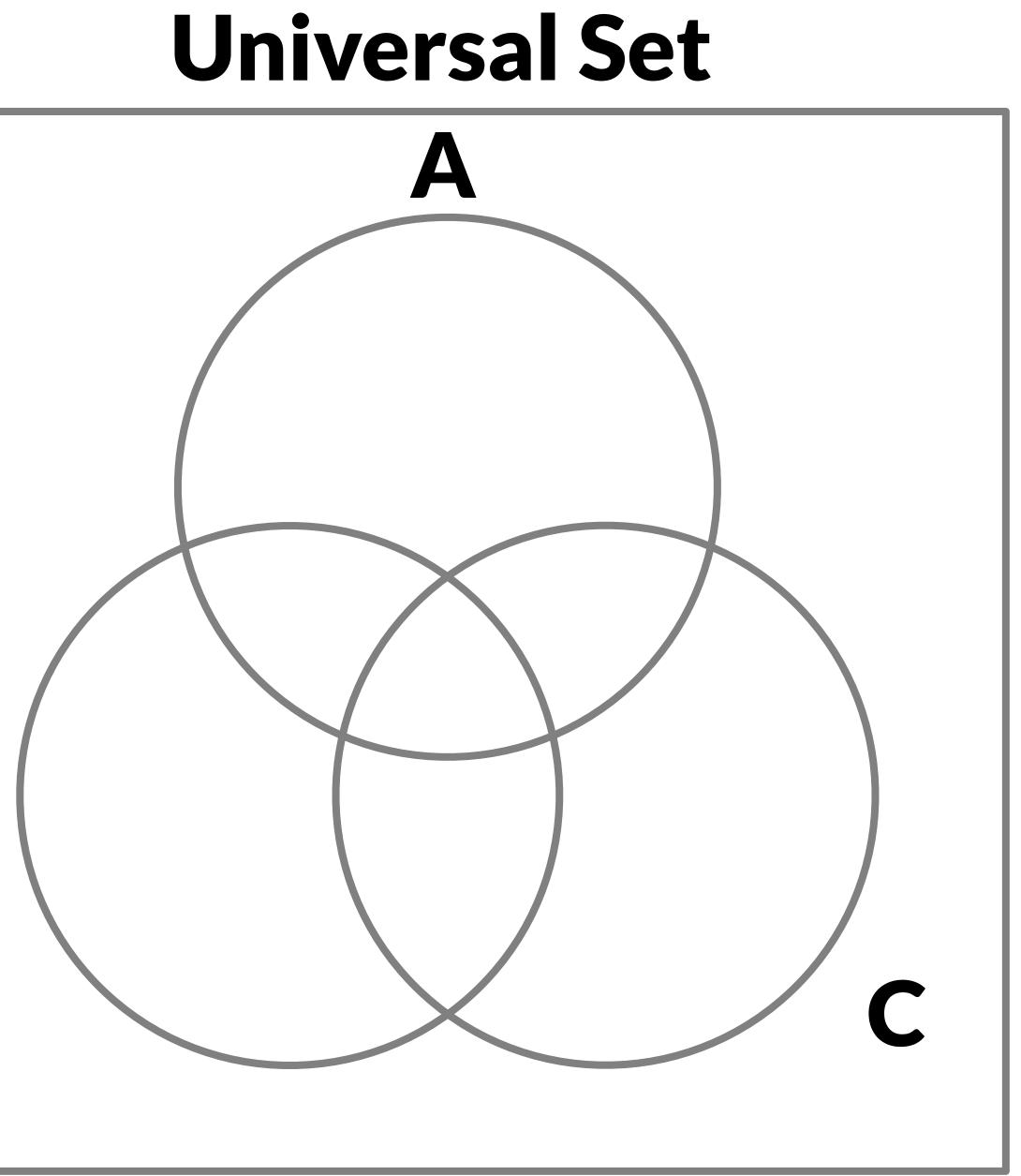


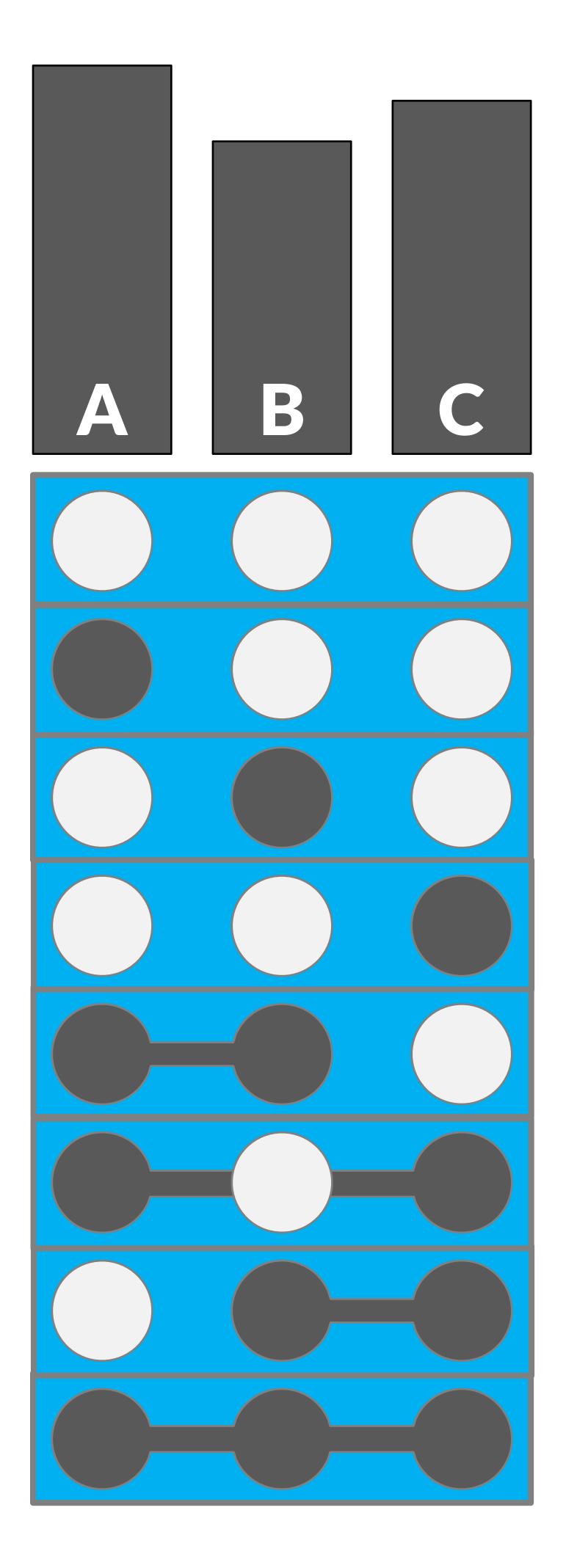
VESEALEZEME

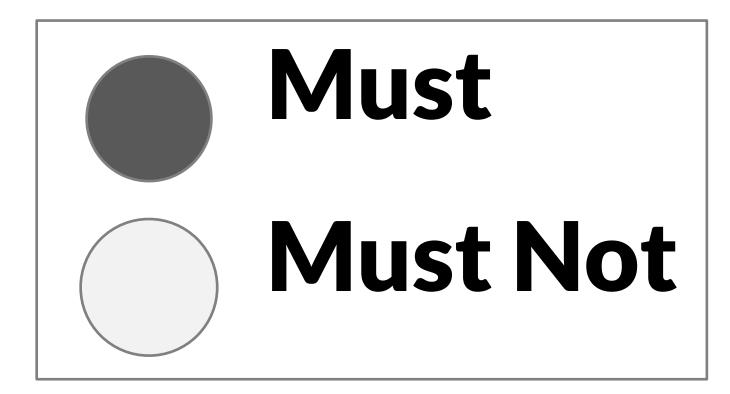






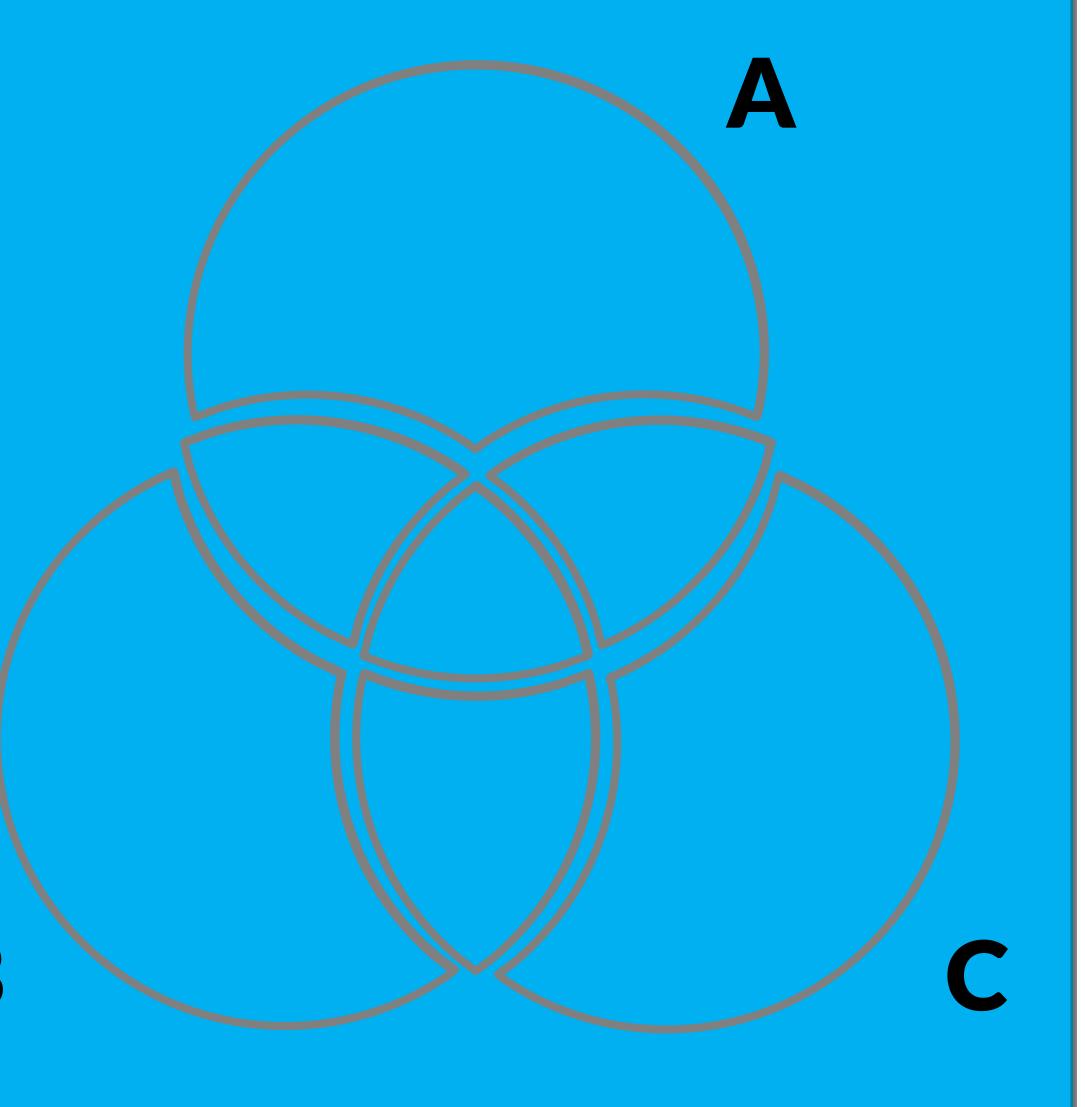


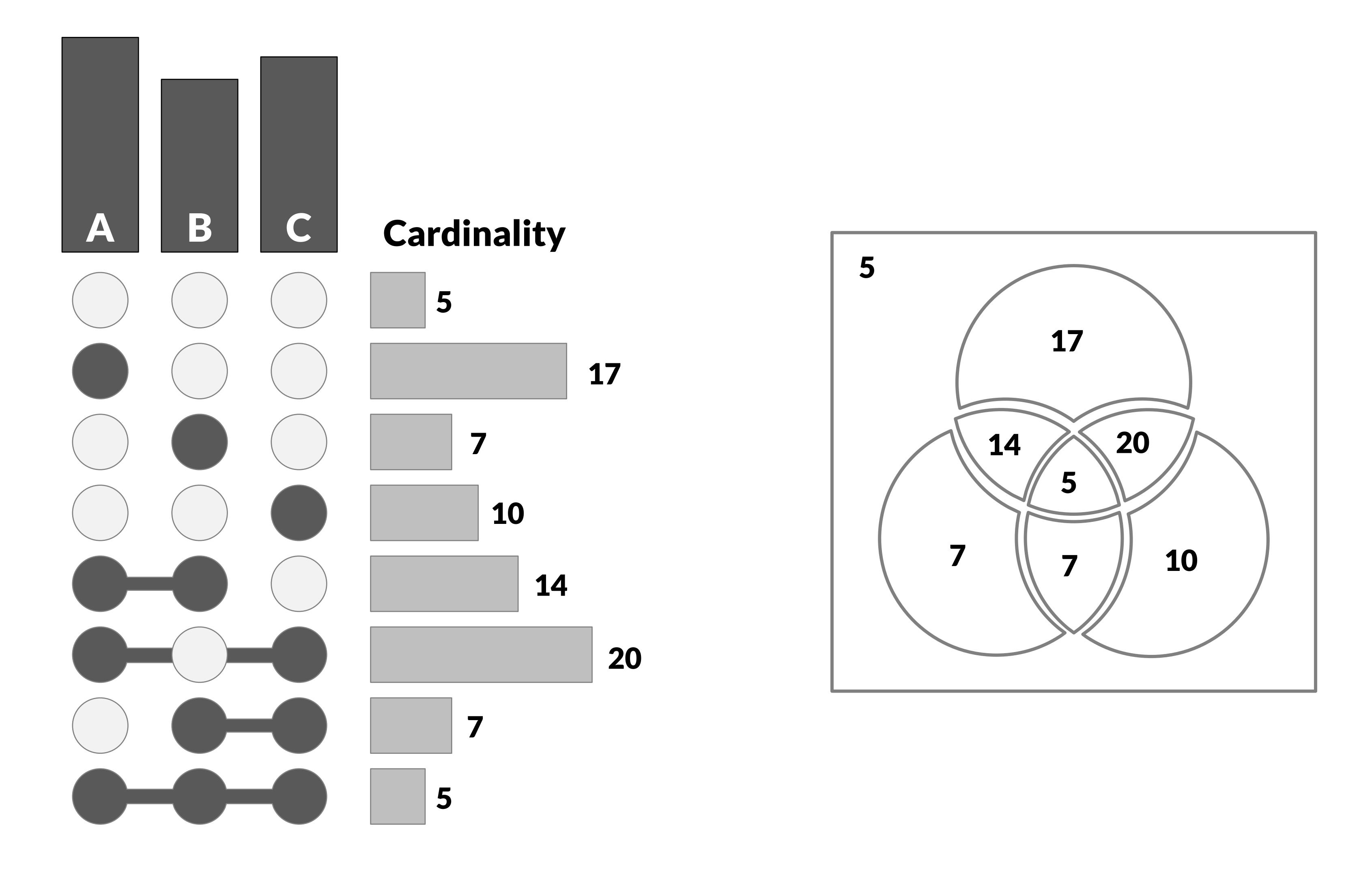






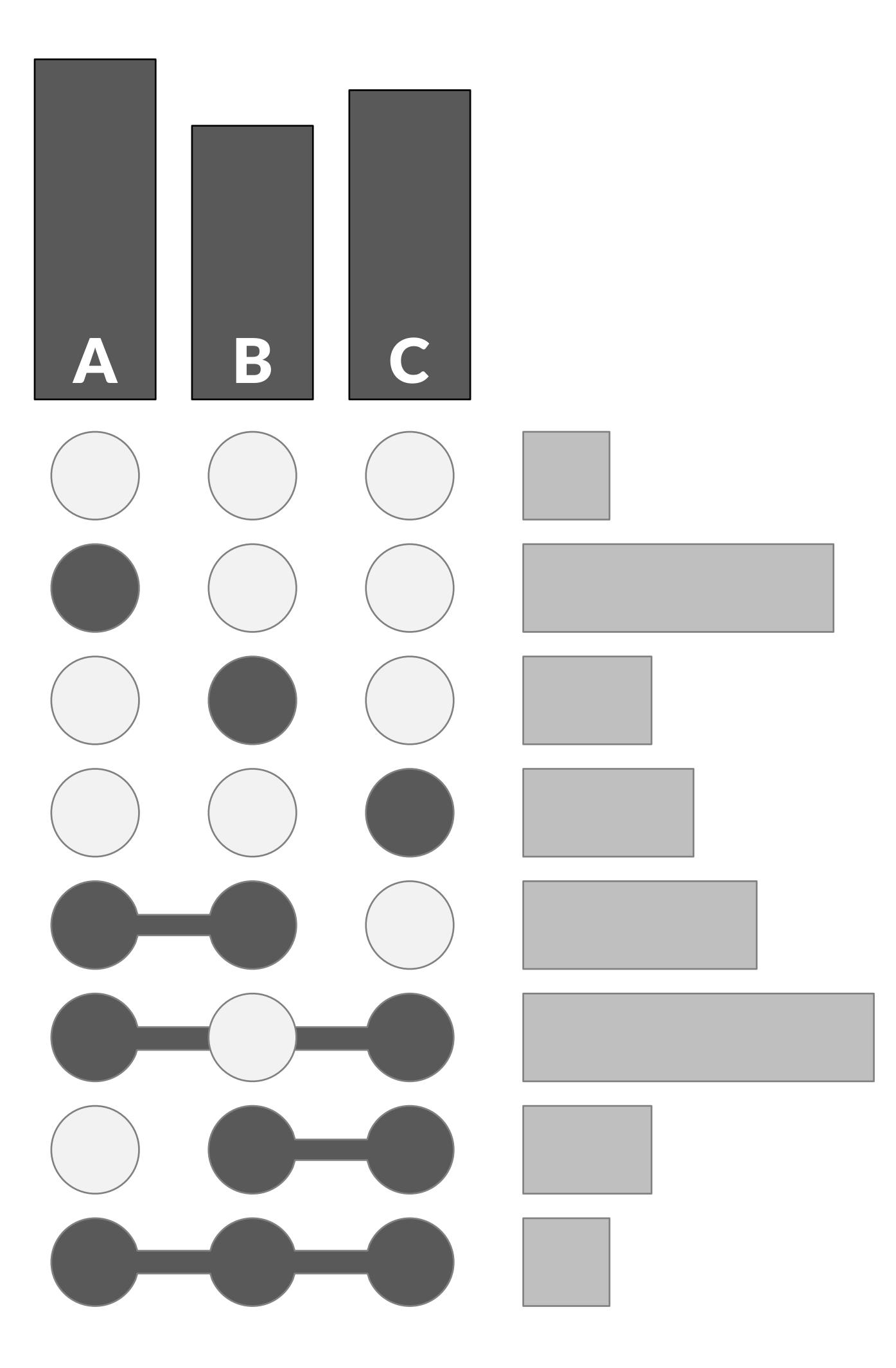






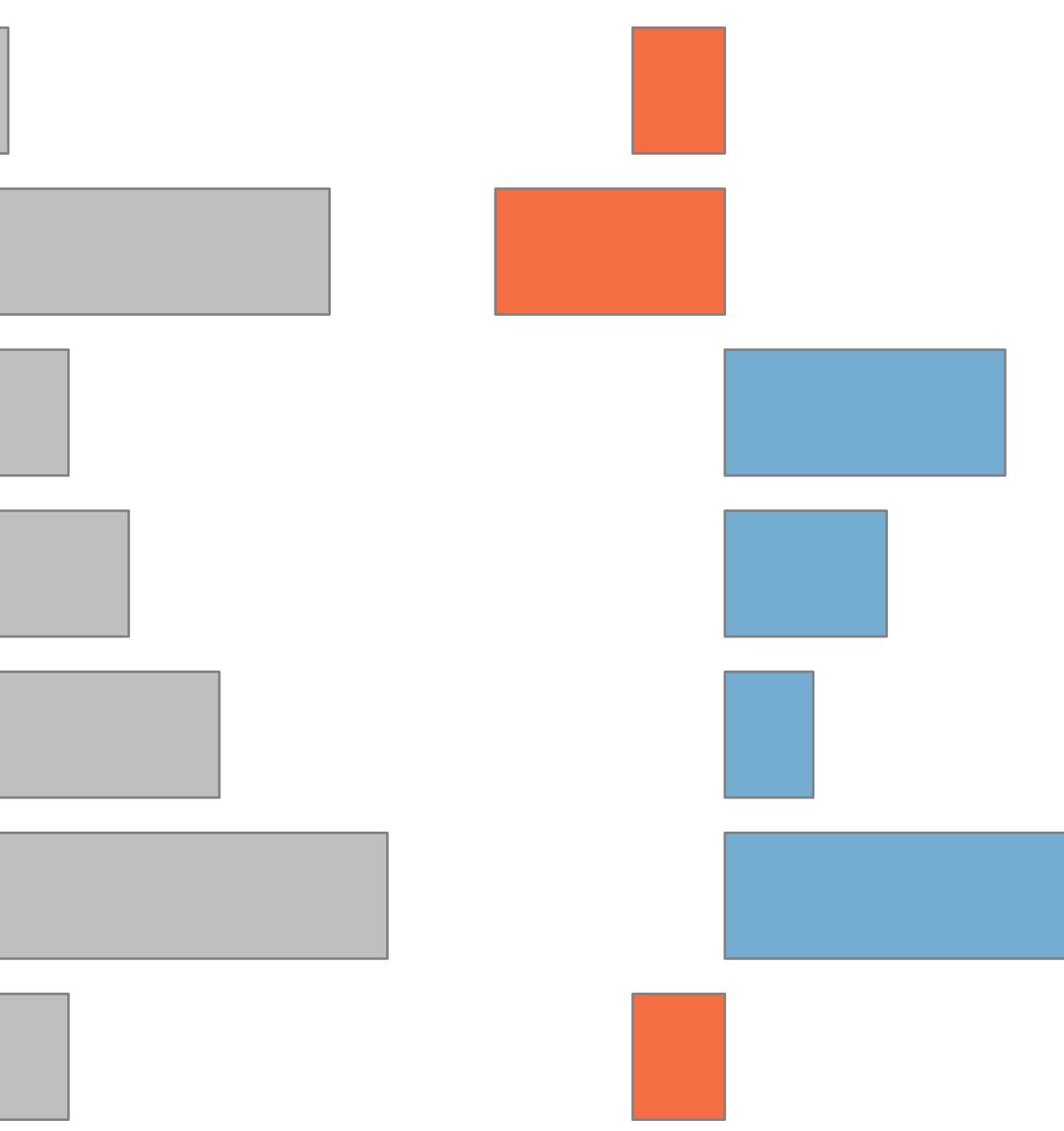




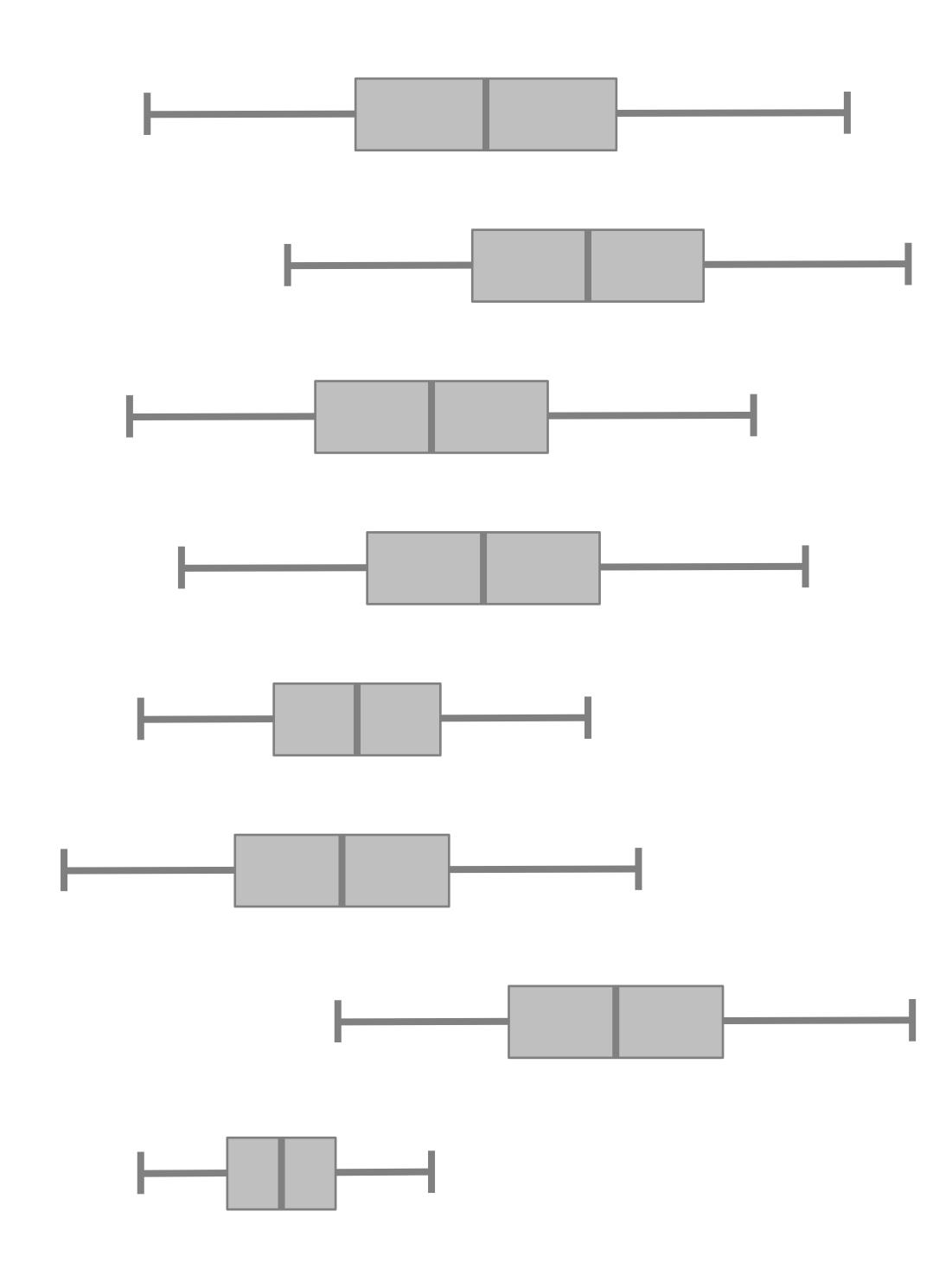


Magutheisigistibuteoizeform intersection? attribute in an Anterse fipports

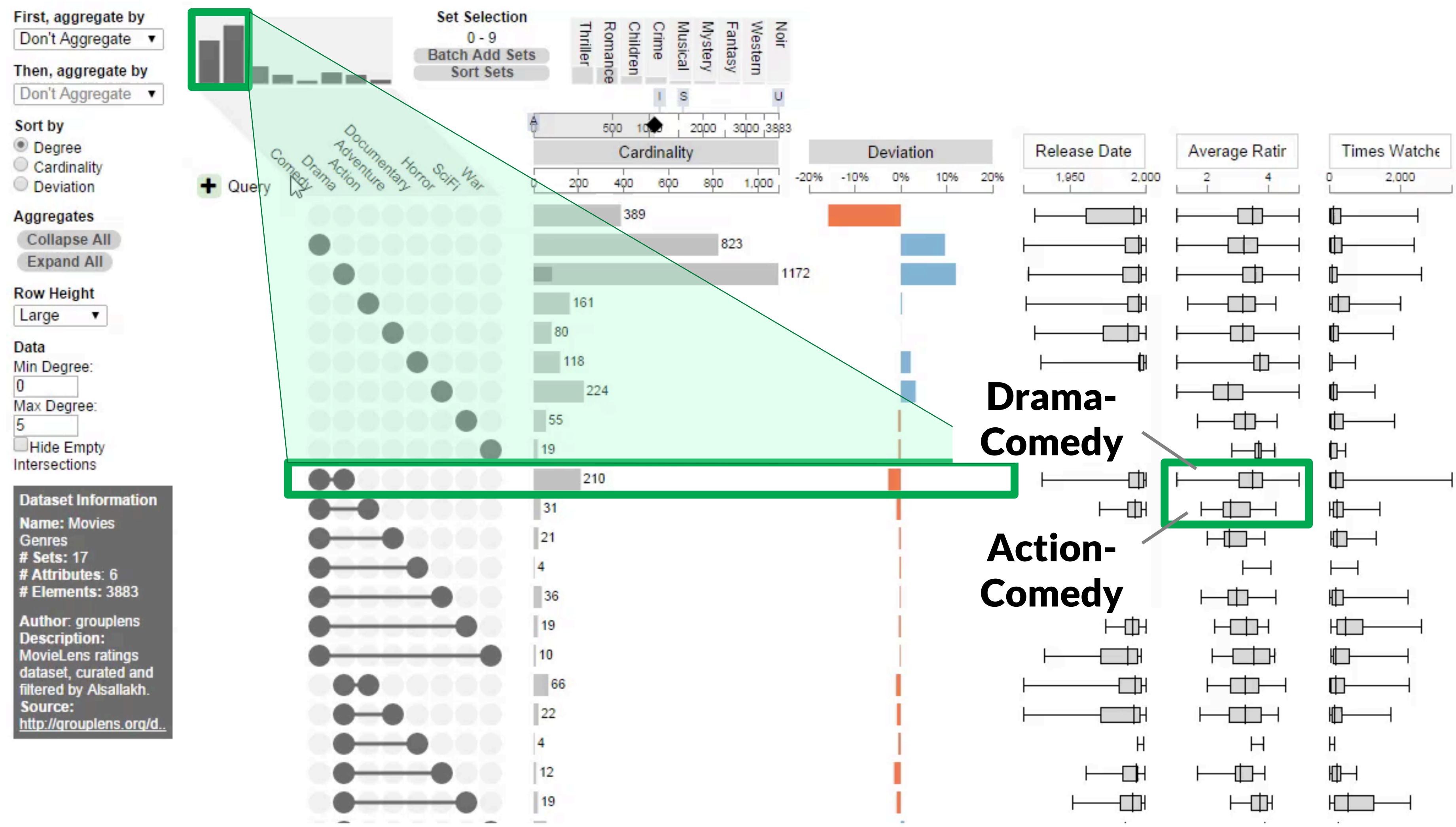
Deviation



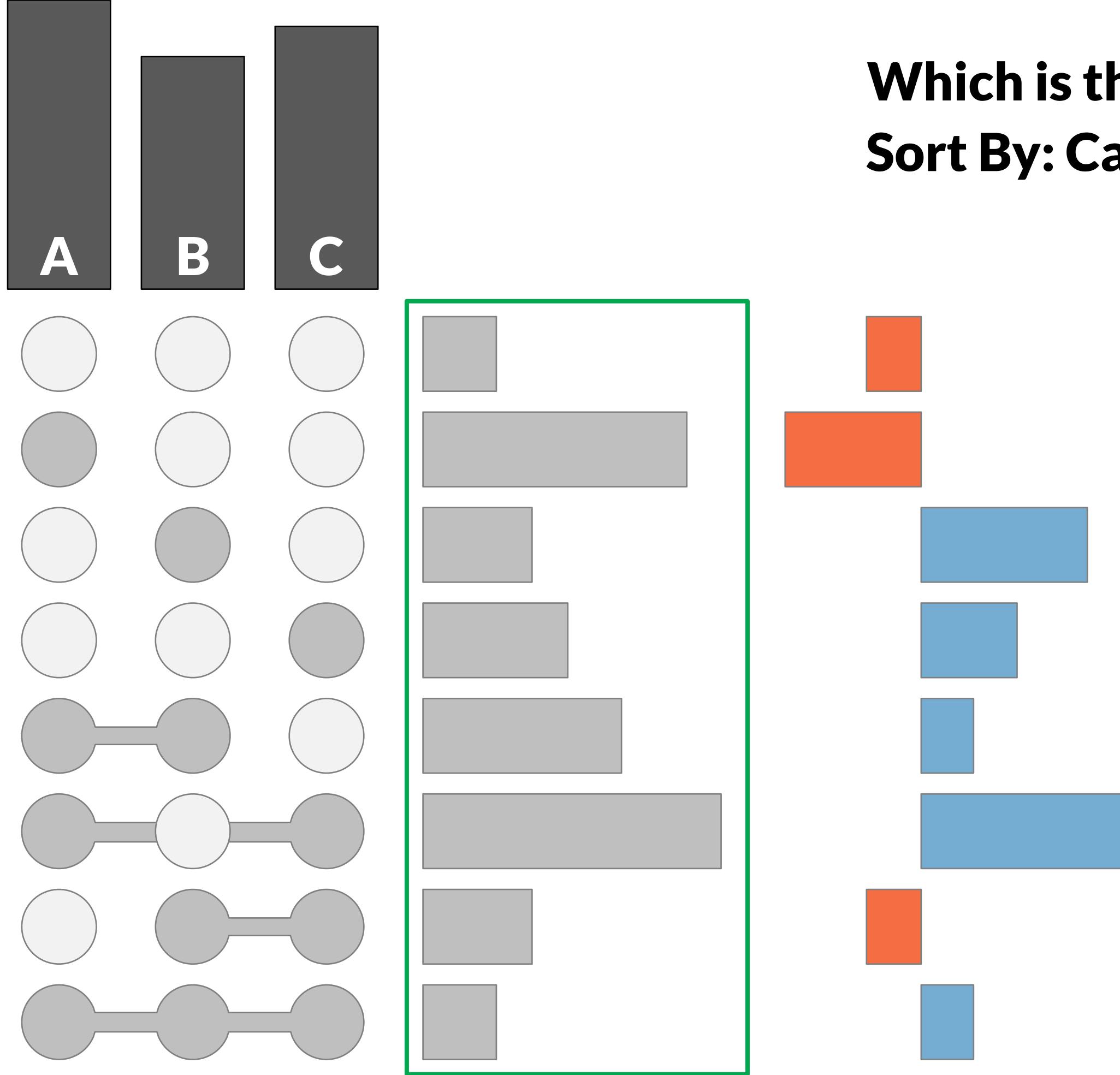
Attributes



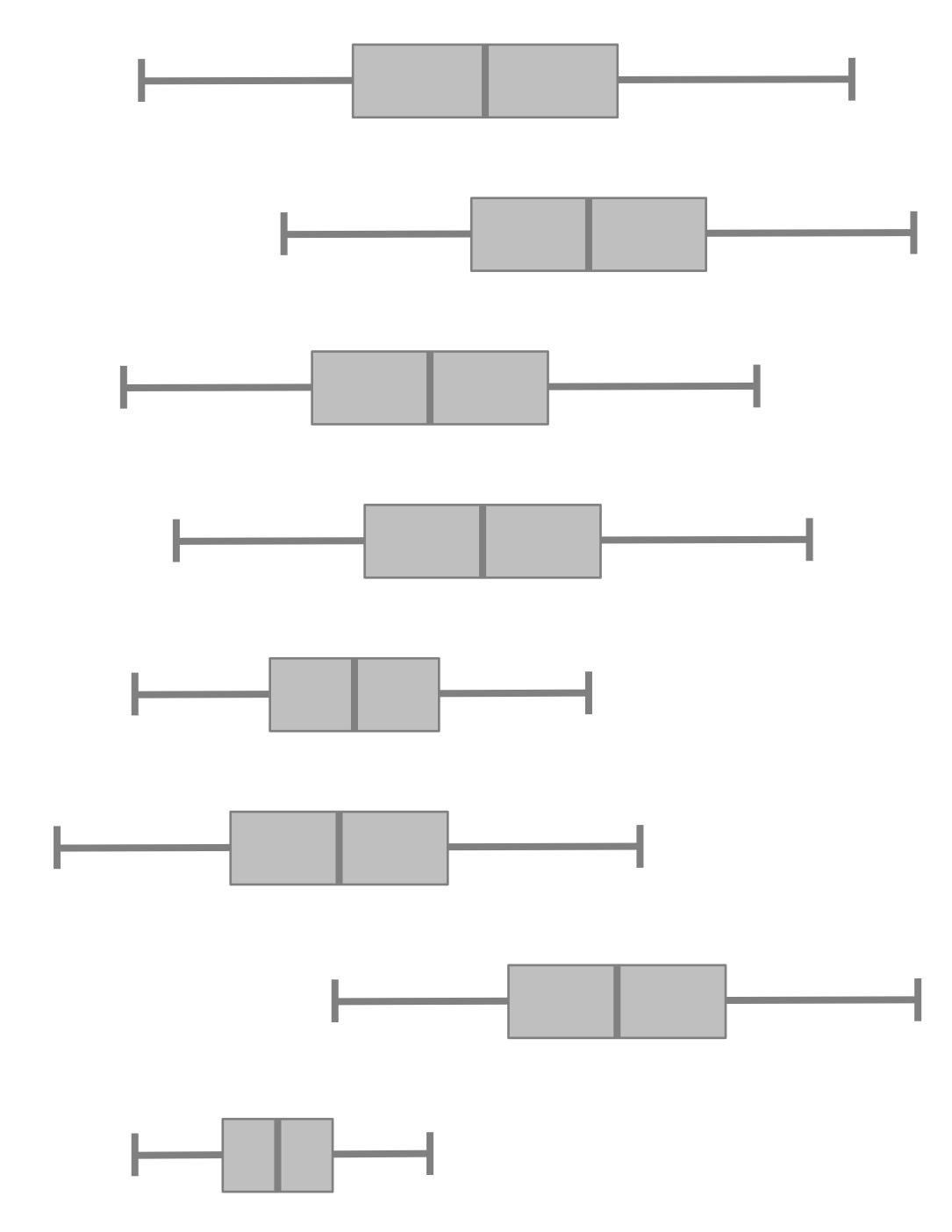


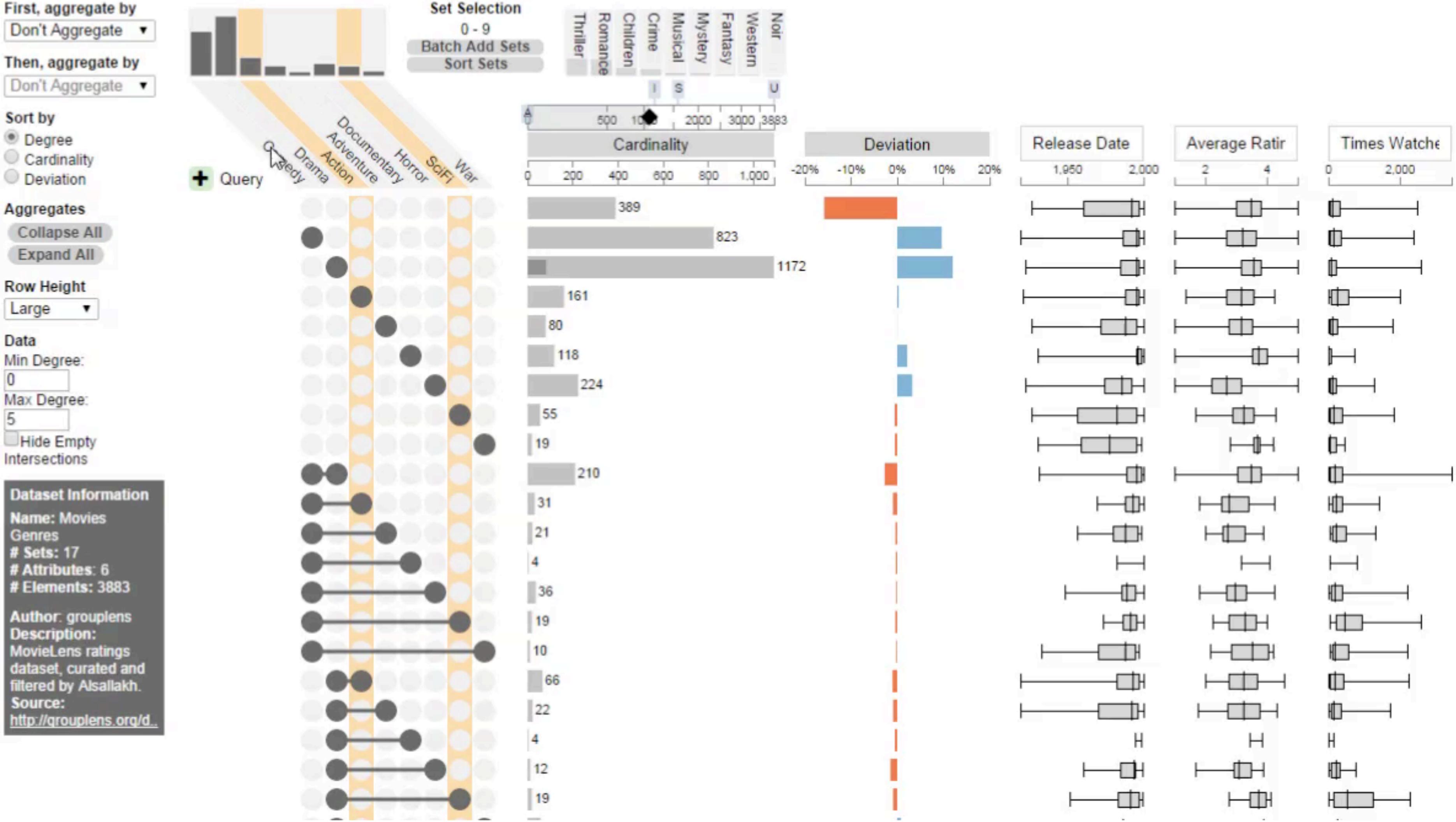




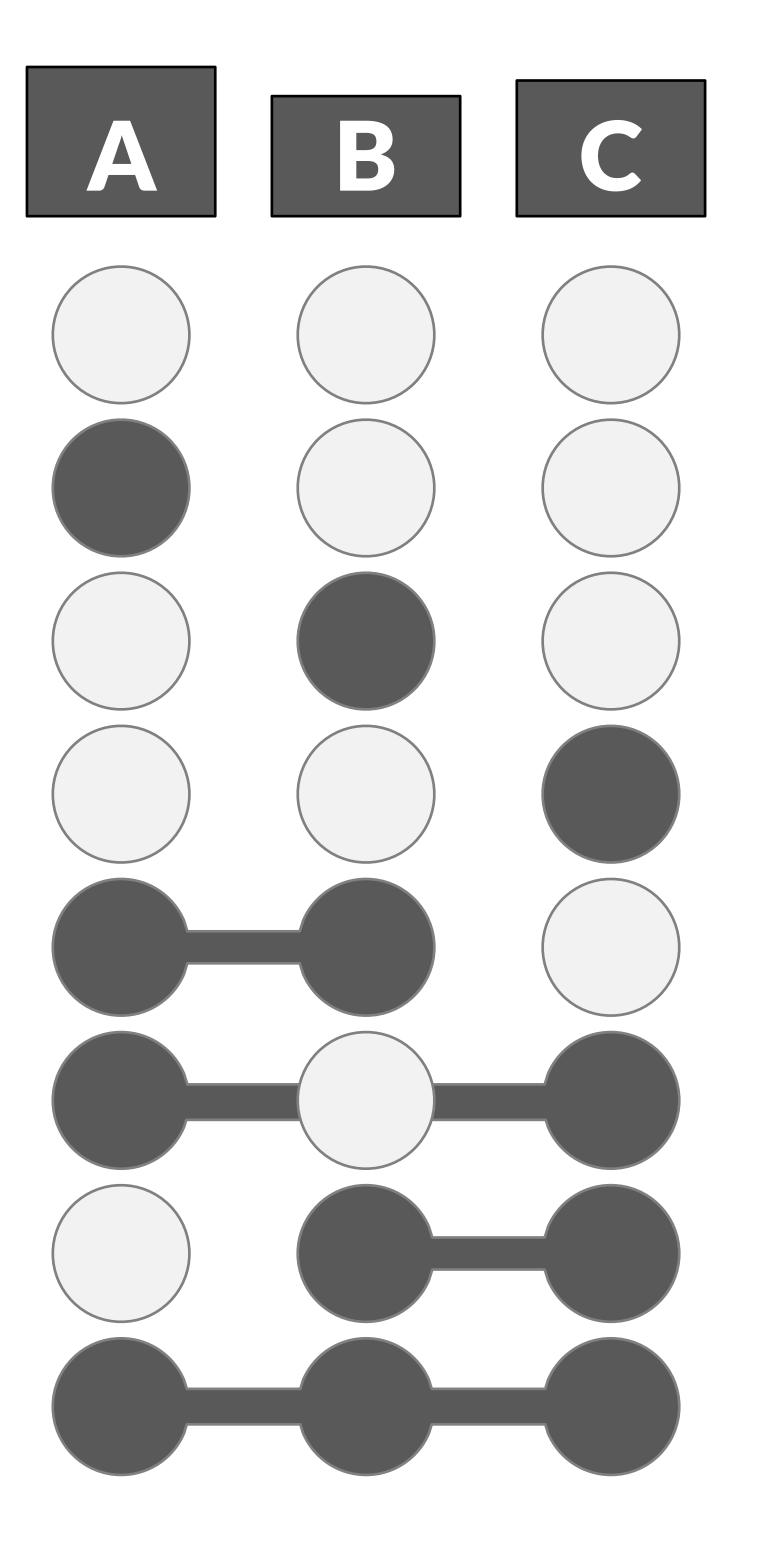


Which is the biggest intersection? Sort By: Cardinality

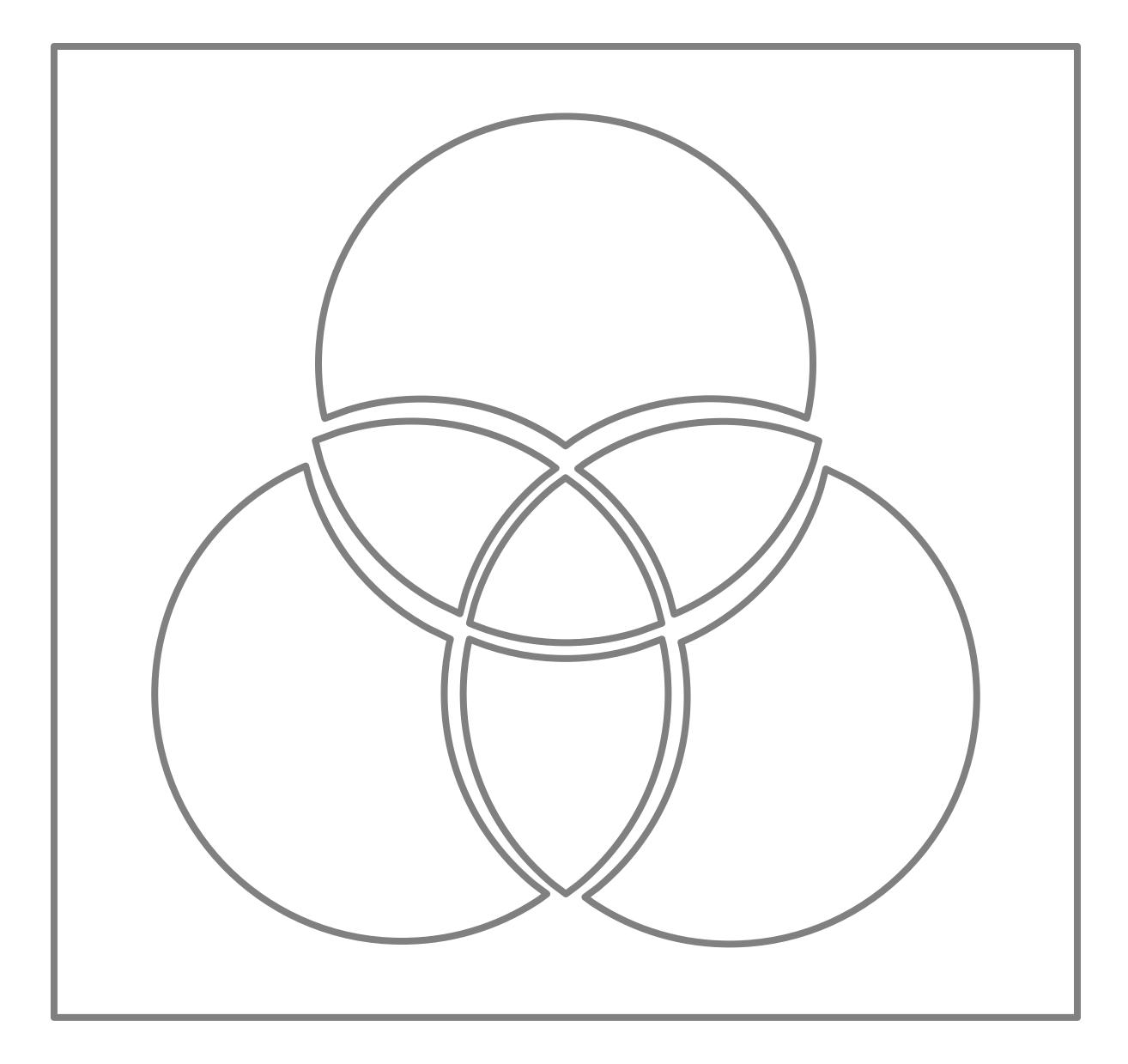




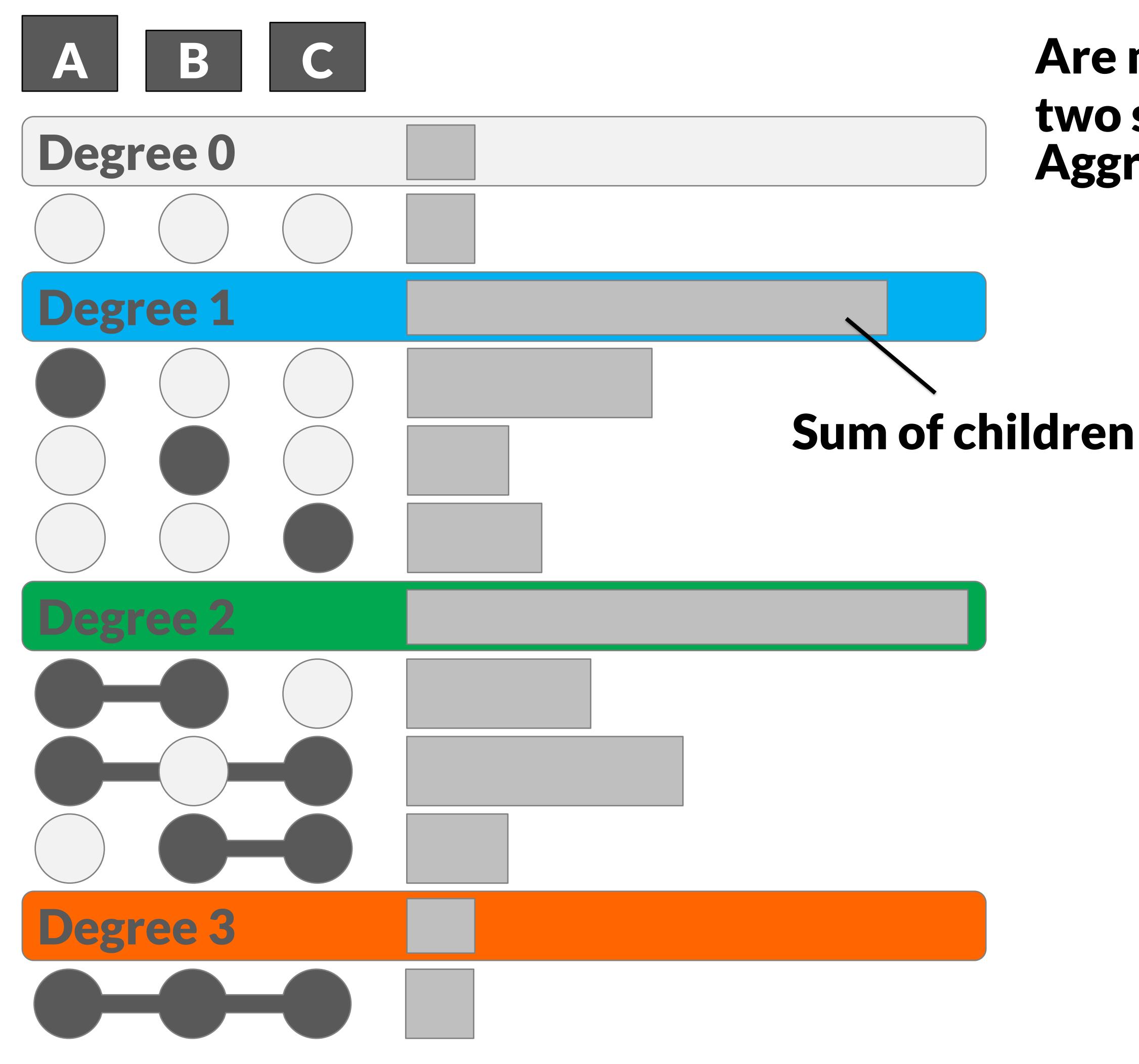
Aggregation



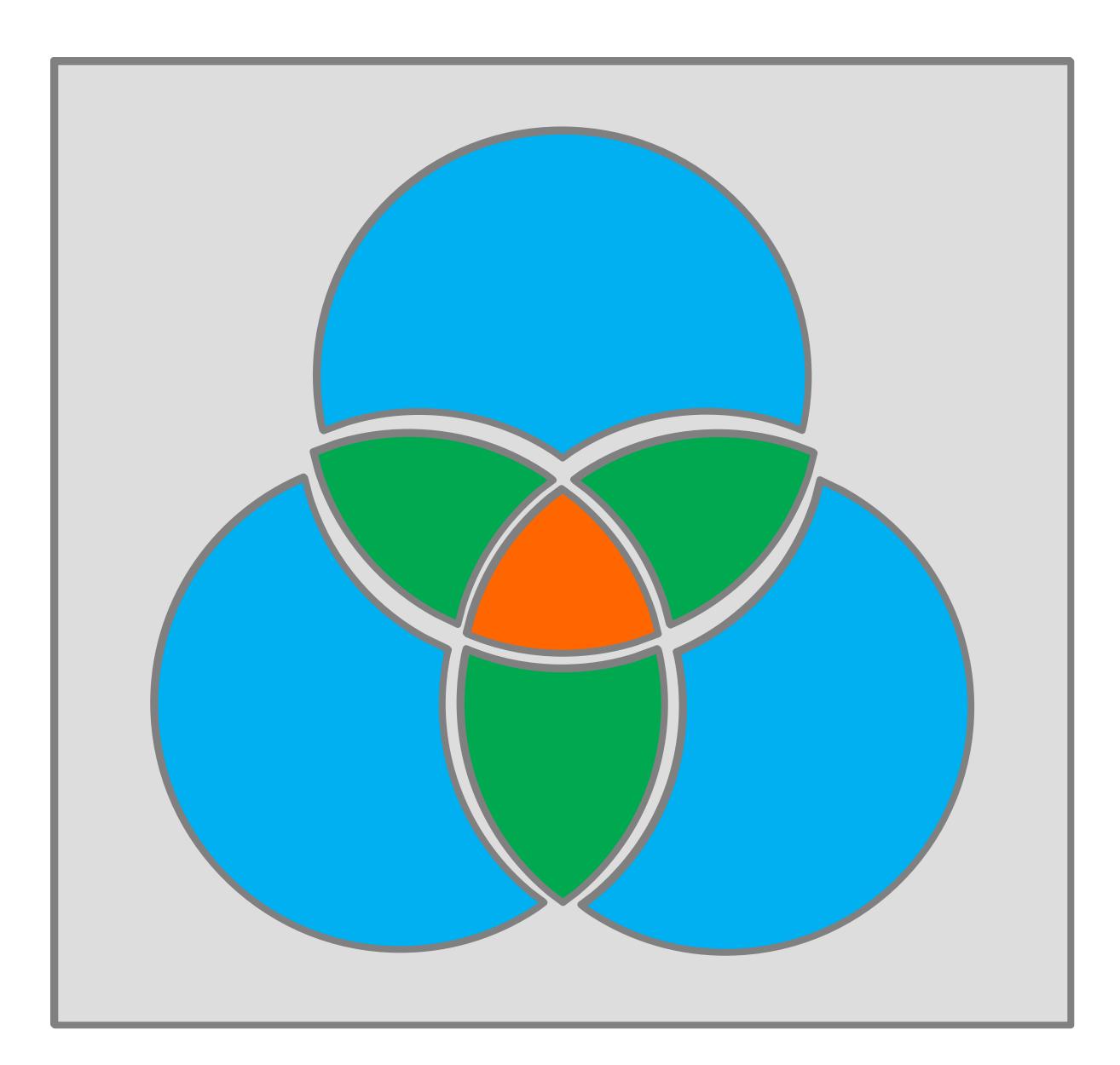
Are many items shared between two sets? Aggregate By: Degree



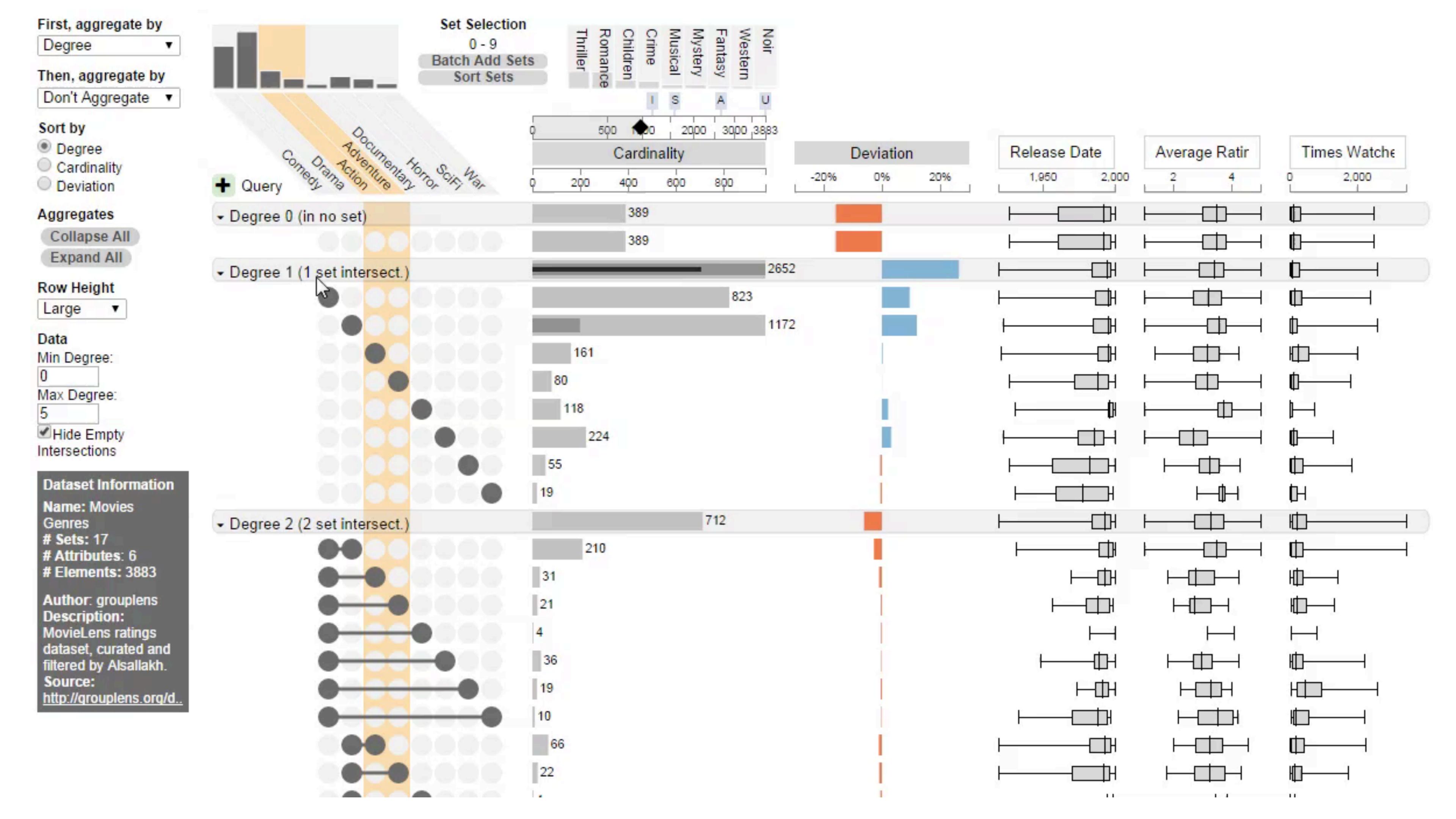




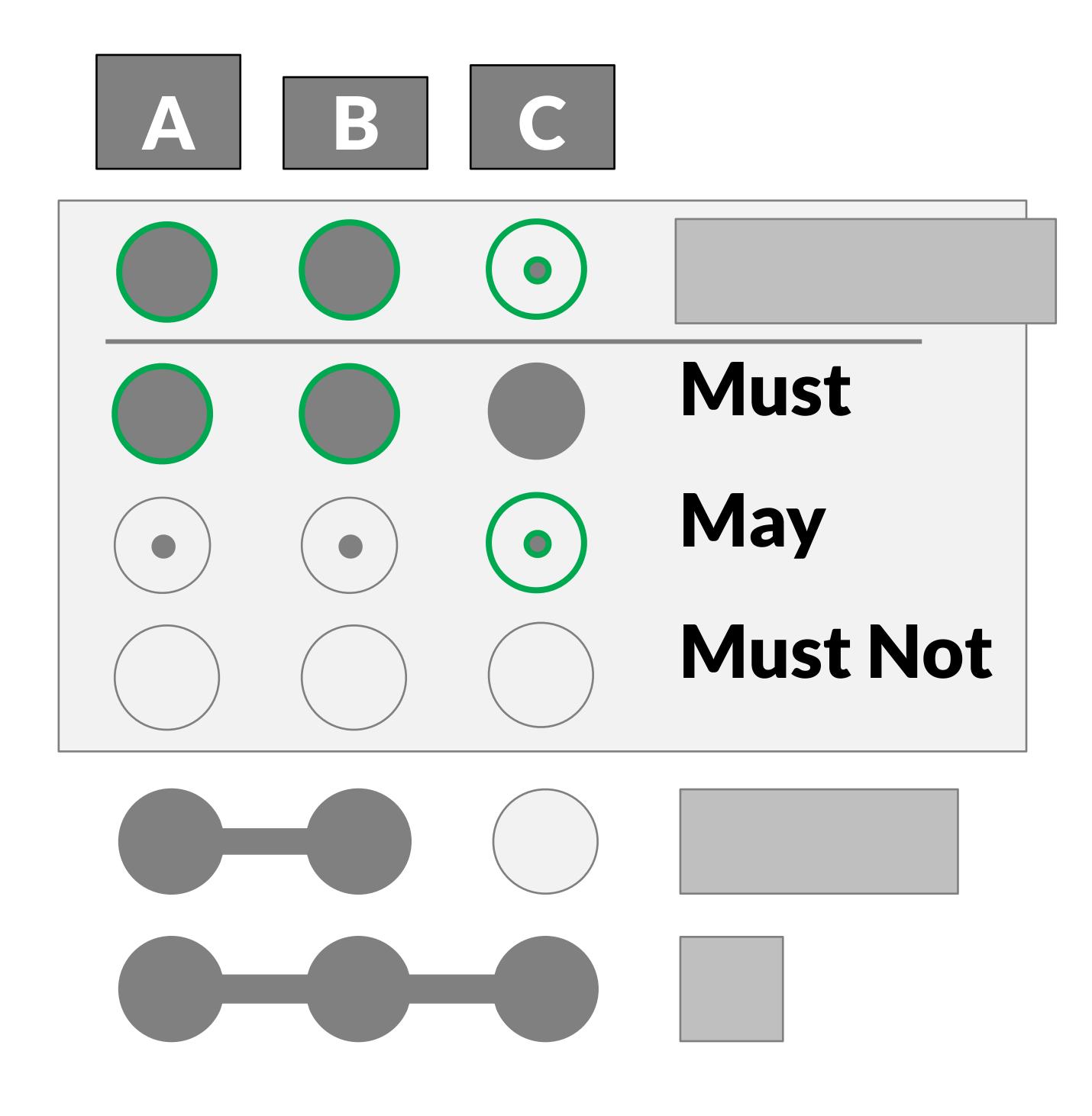
Are many items shared between two sets? Aggregate By: Degree

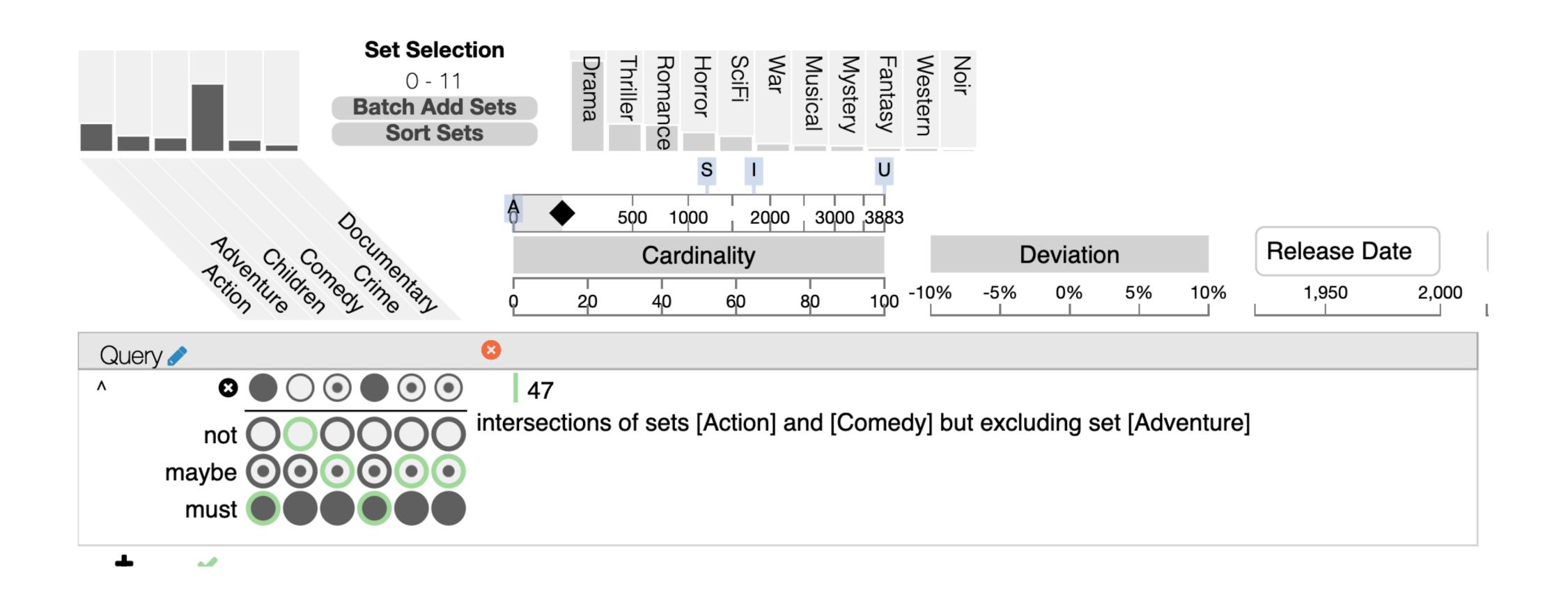


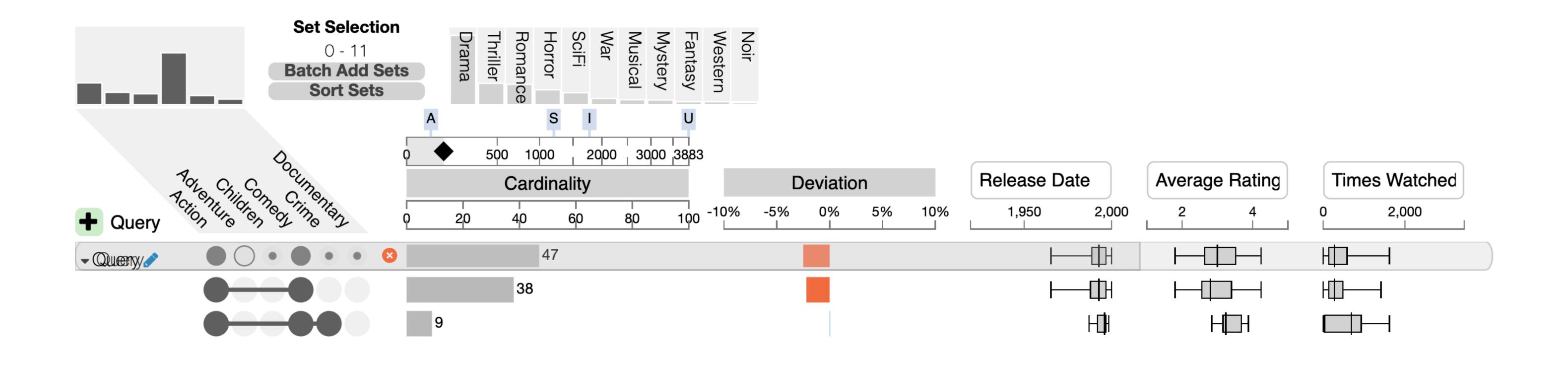




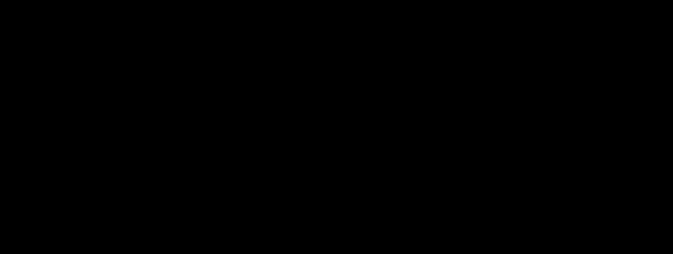


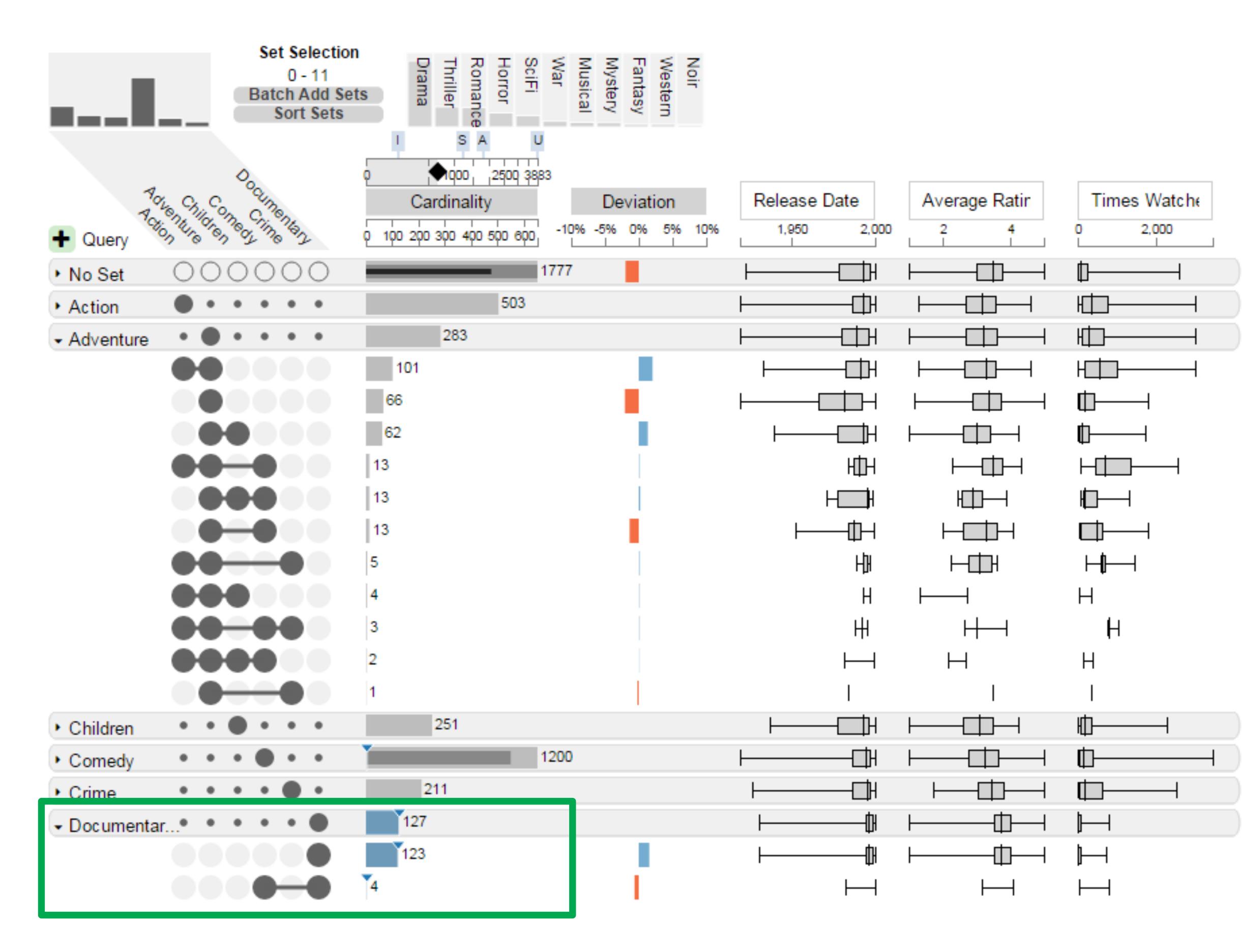






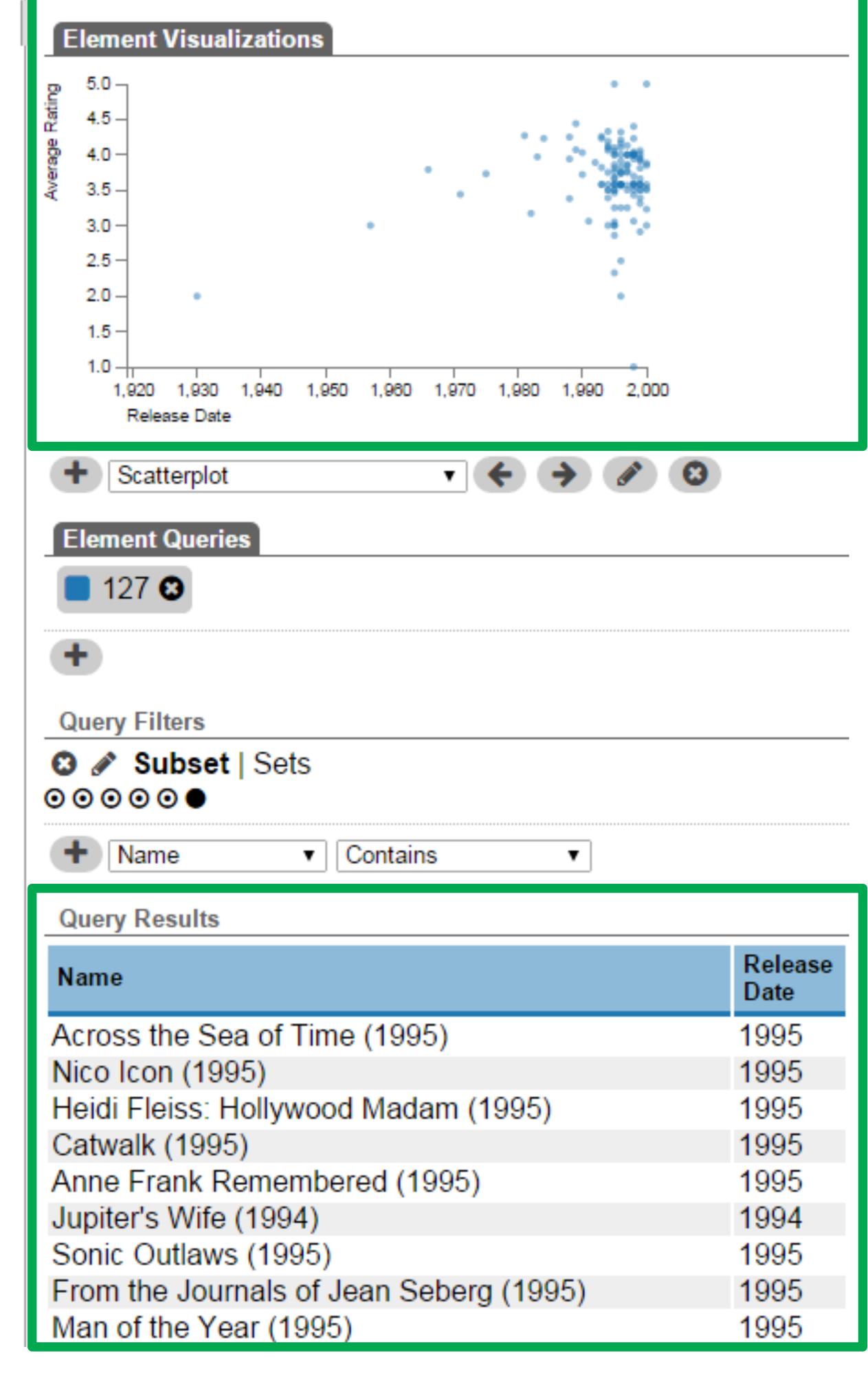


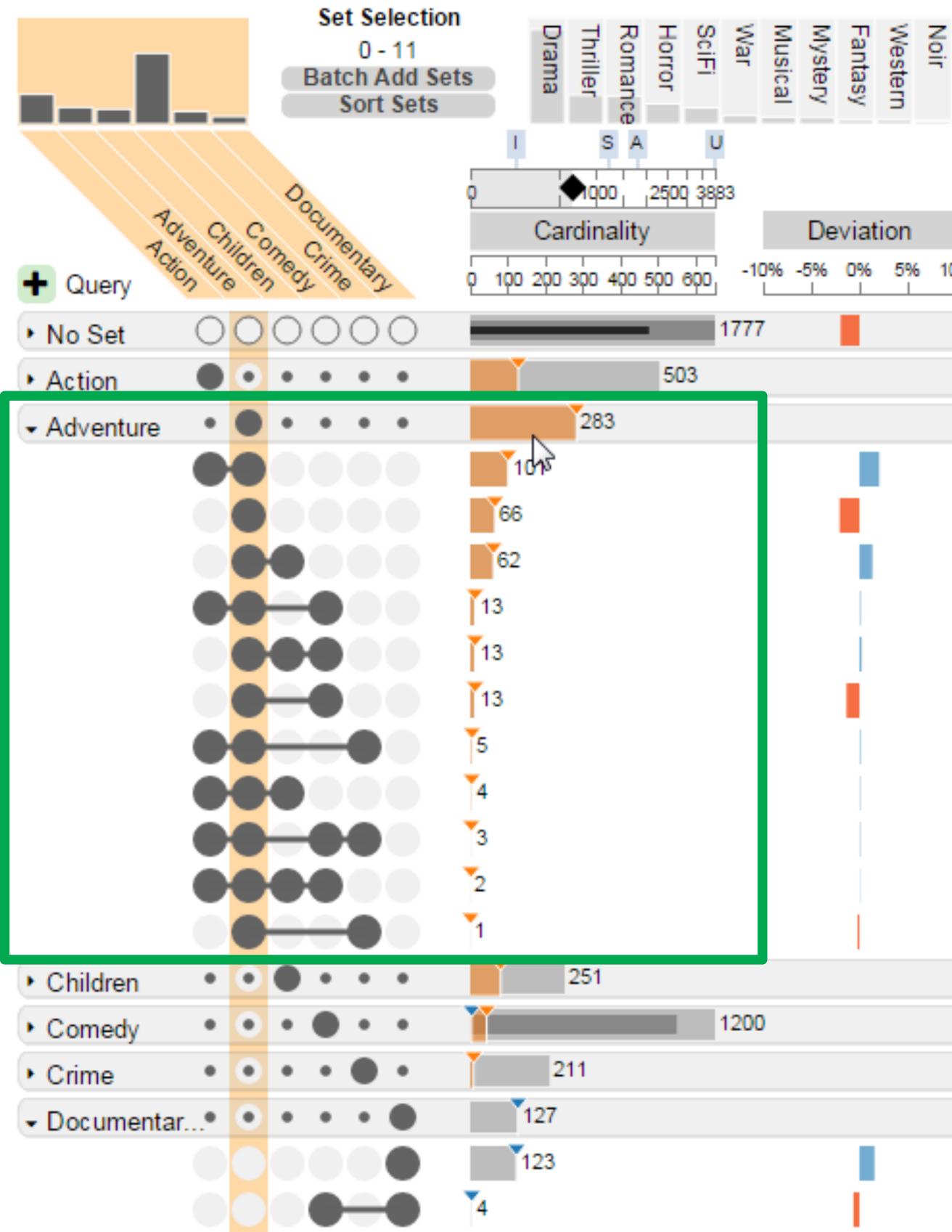




3

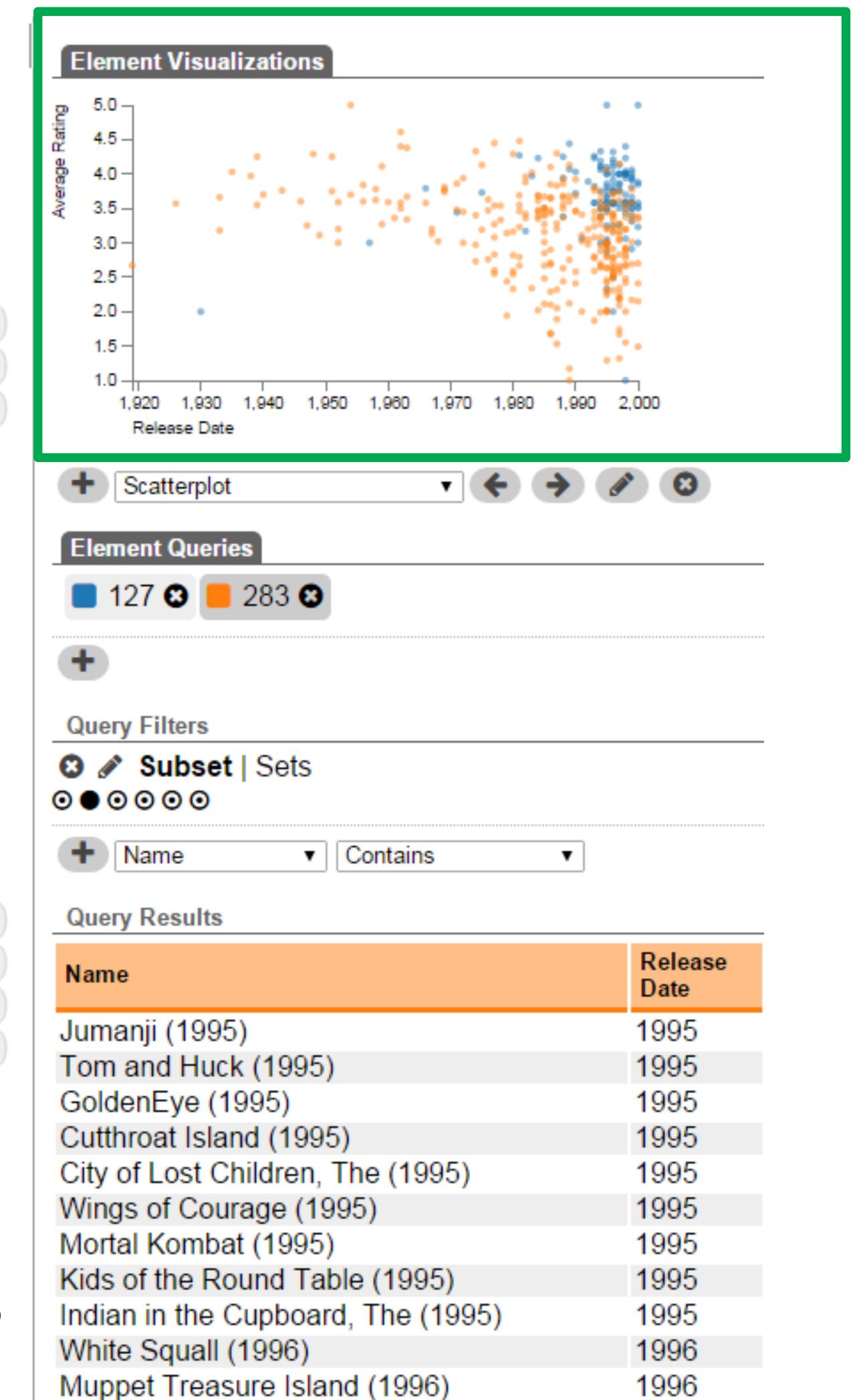
How do documentaries compare to adventure movies?





Deviation	Release Date	Average Ratir	Times Watche
-5% 0% 5% 10%	1,950 2,000	2 4	0 2,000
			₽
	⊢Ū		H
			H
			₫——
			┣───┤
	НШН		
	H	H	₩□
	⊢		
	Н	H	⊢┨──┤
	Н	\vdash	Н
	Н	-	H
	\vdash	Н	Н
	I		Ι
			₩□
	⊢H		$\square \longrightarrow$
	₽	├─────────┤	
	┣	⊢	
	\vdash	\vdash	\vdash

How do documentaries compare to adventure movies?





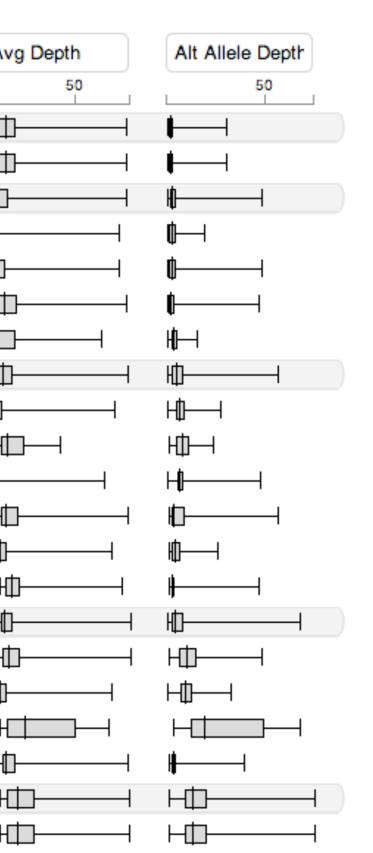
Applications

Genetics Economics Pharmacology **Social Networks**



UpSet

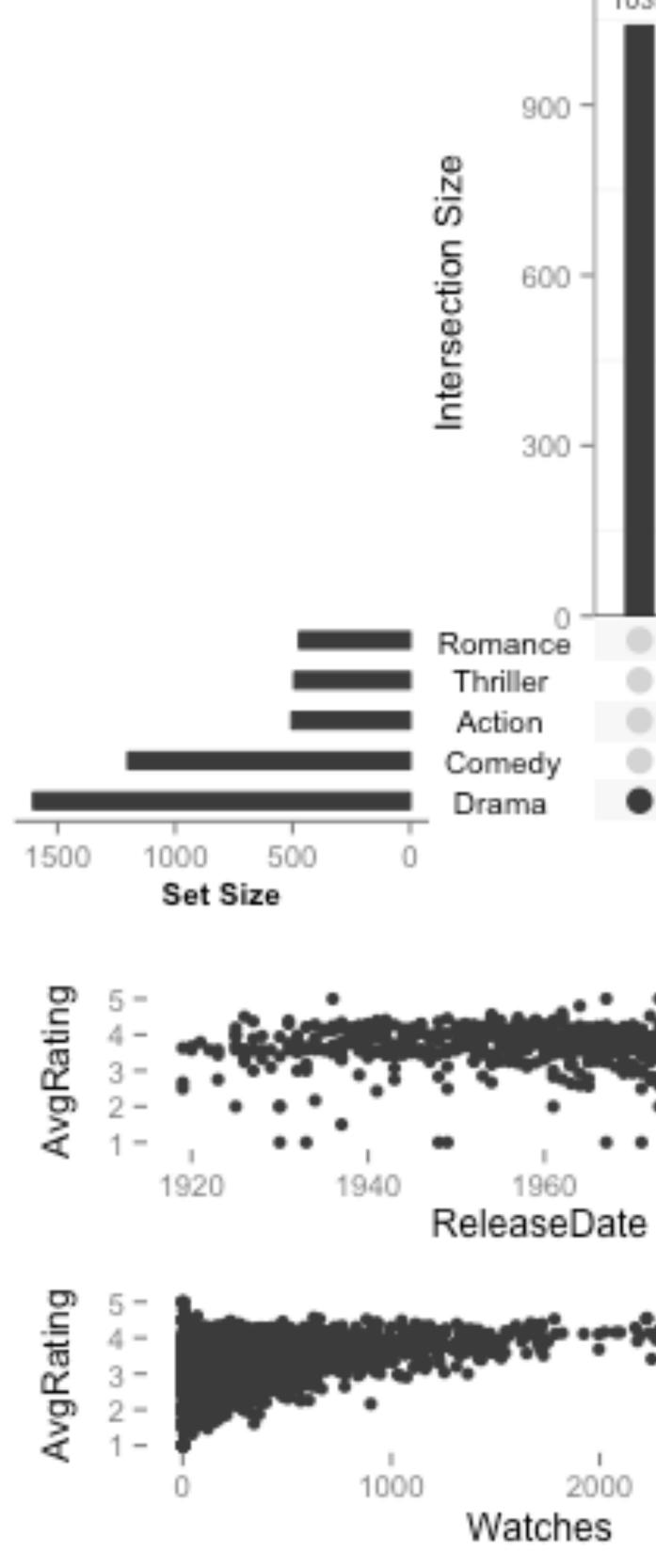
First, aggregate by Degree ‡ Then, aggregate by Don't Aggregate ‡ Sort by	T.2mm.bes T.2mm.bes T.4mm.bes T.2mm.bes S.defaults Query Query Decree 0 (in no set)	S.2mm.uni T.2mm.bes B/G.Q20.d B/G.Q20.s T.2mm.bes 0 15000 3000 47526		
 Degree Cardinality 	9411 Ma 19 80	Cardinality	Deviation	Avg
Deviation	+ Query 200 000 000 000	o 5,00010,0005,00@0,00@5,00@0,000	-20% -10% 0% 10% 20%	
Aggregates	 Degree 0 (in no set) 	19340		НФ-
Collapse All	0000	19340		нф-
Expand All	- Degree 1 (1 set intersect.)	11099		Ю—
Row Height		1595		Ф
		5899		Ю—
Data Min Degree:		2107		НФ-
0		1498		HD-
Max Degree: 15	- Degree 2 (2 set intersect.)	4243		HD-
Hide Empty Intersections		642		Ю—
Intersections		26		۱D-
		449		₩
		1872		HD-
		918		Ю—
		336		Ю—
	- Degree 3 (3 set intersect.)	5784		HD
		320		Ю-
		1384		Ю—
		27		ΗŢ
		4053		Ю—
	- Degree 4 (4 set intersect.)	7060		ΗŢ
		7060		НŢ

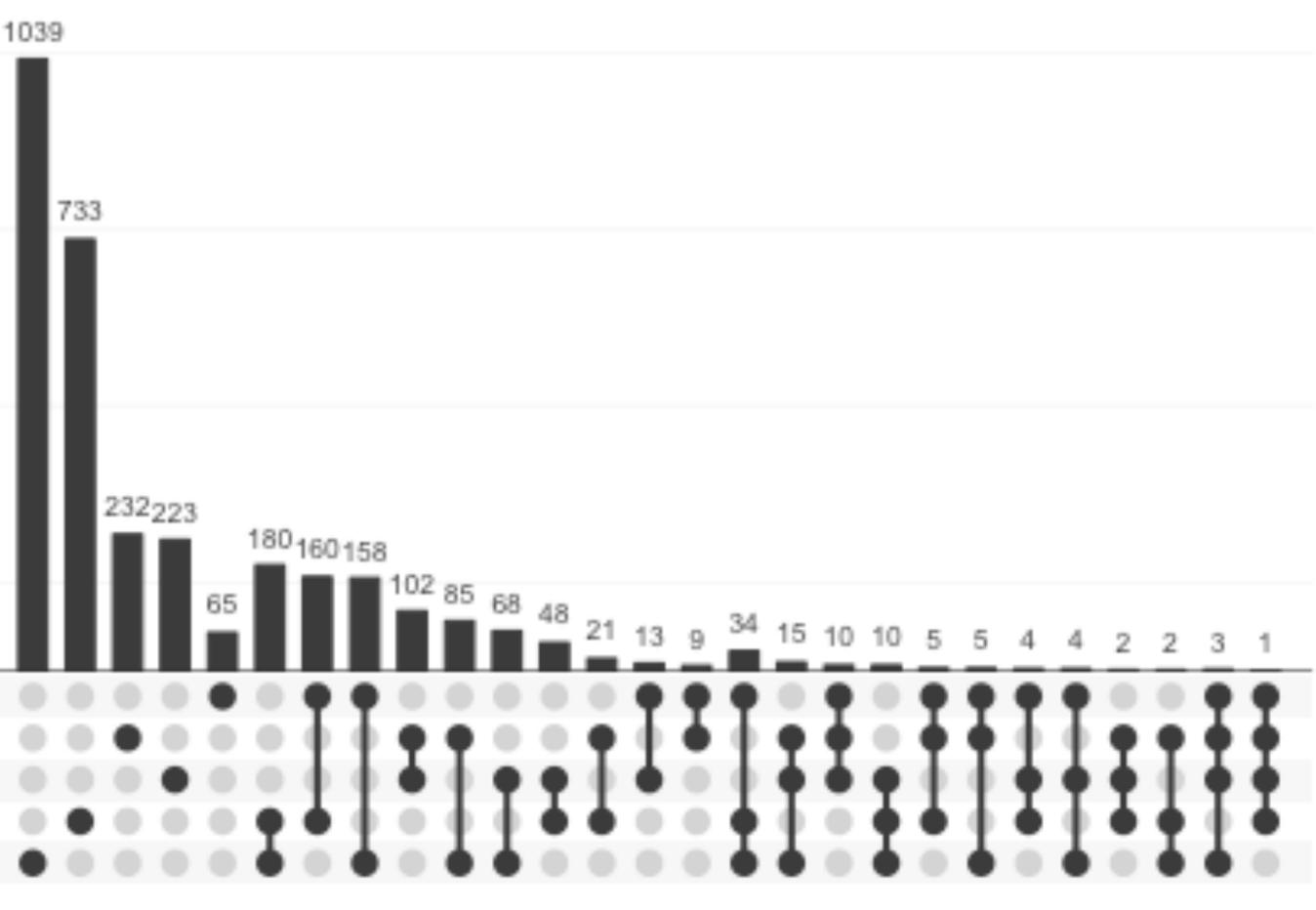


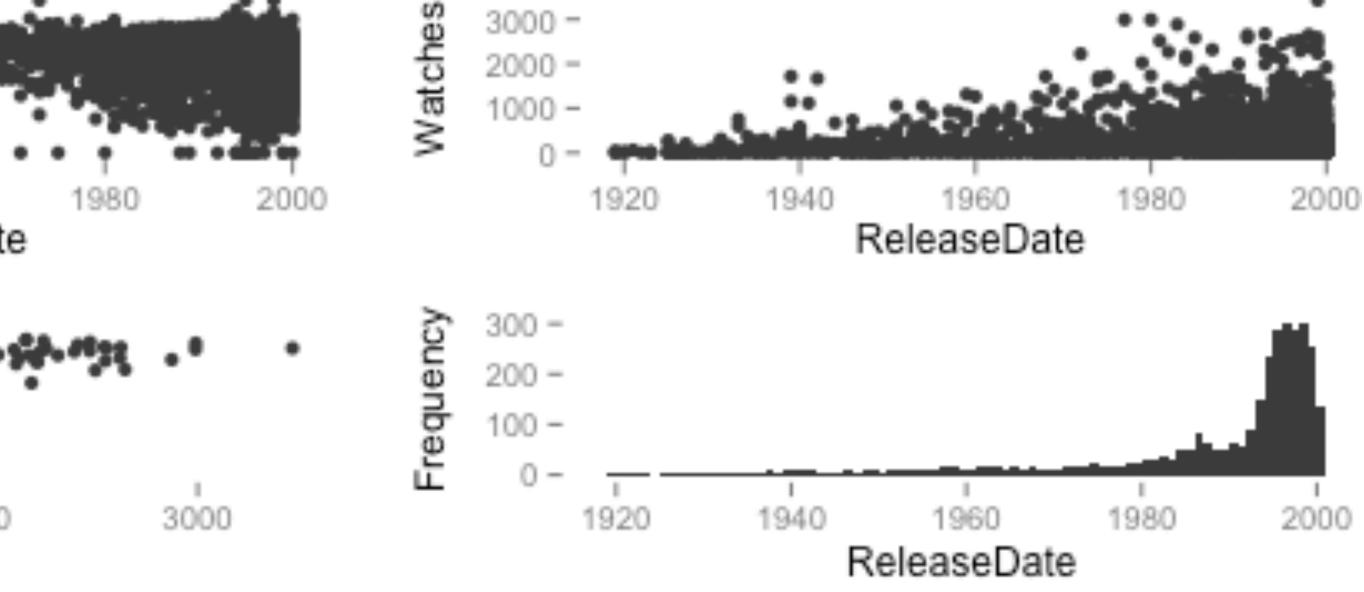
Choos	e Da	taset Vari	ant Callers (variable depth)	(15 sets, 6 attribute	es) ‡	0
Element Visualization ti/tv = 2.562 (5078/1 ti/tv = 0.784 (2592/3	982)						
Periodic A C G T Ref Allele	Alt Allele	A C G T A C Ref Allele	G T				
 Transition/Transv Element Queries 7060 S 589 	ersion 9 8	Ratio 💠	€ →	Ø 8			
Query Filters							
O ● O O O ● O O	S						
+ Name	*) Co	ontains	\$				
Name	Chr	Ref Allele	Alt Allele	Avg Depth	Alt Allele Depth	Set Count	
10:101496060:A:G	10	А	G	9.1	2	1	
10:101656105:G:T		G	Т	10.3	1.3	2	
10:102112233:G:T		G	Т	24.6	2.7	1	
10:102767155:C:G		С	G	8.9	8.9	2	
10:102767158:G:T		G	Т	8.9	1.4	2	
10:103544430:C:A	10	С	A	10.1	1.1	1	

Name	Chr	Ref Allele	Alt Allele	Avg Depth	Alt Allele Depth	Set C
10:101496060:A:G	10	Α	G	9.1	2	1
10:101656105:G:T	10	G	Т	10.3	1.3	2
10:102112233:G:T	10	G	Т	24.6	2.7	1
10:102767155:C:G	10	С	G	8.9	8.9	2
10:102767158:G:T	10	G	Т	8.9	1.4	2
10:103544430:C:A	10	С	Α	10.1	1.1	1
10:103815852:G:T	10	G	Т	10.8	0.7	1

R-Version: UpSetR Developed at HMS Some design adaptions







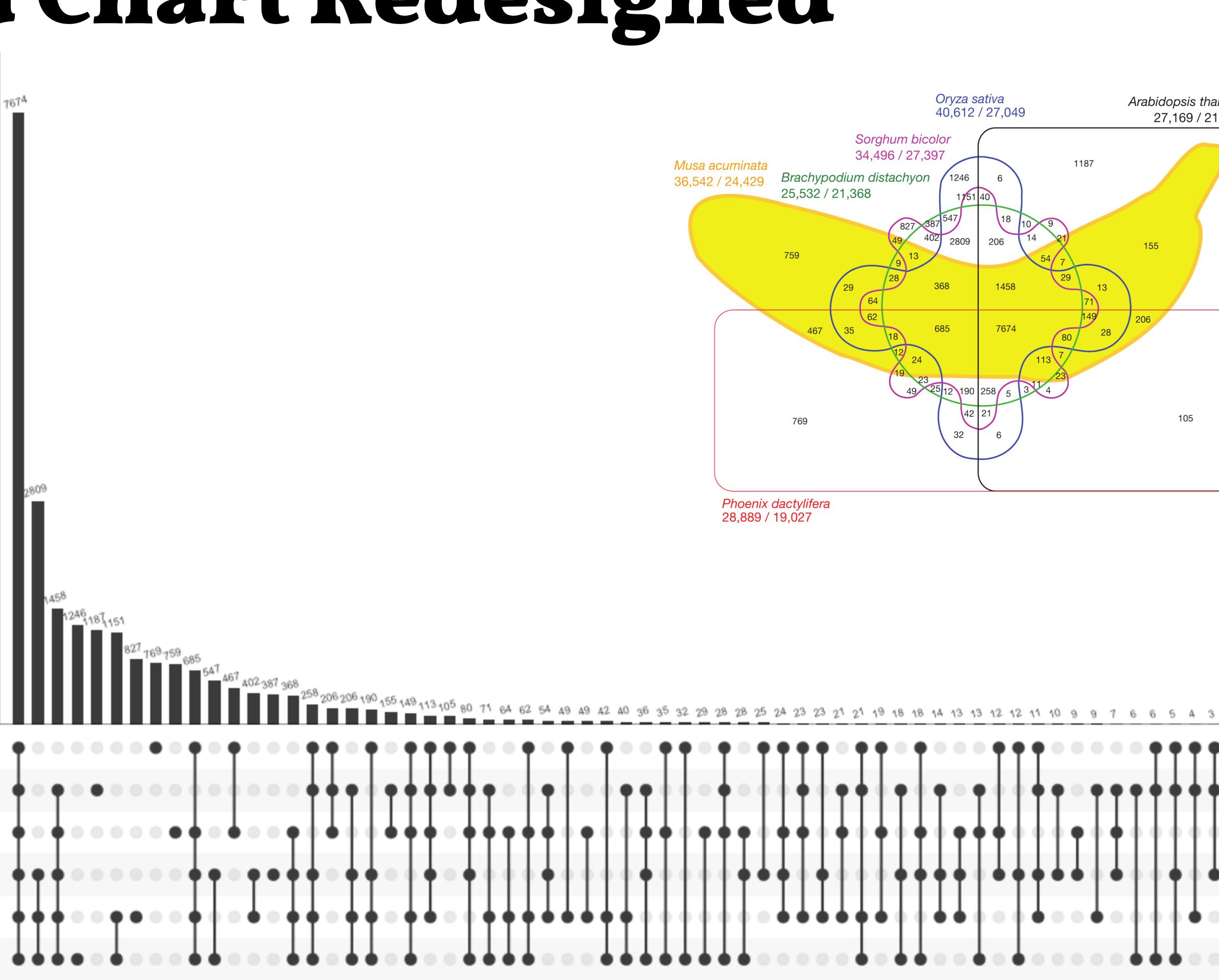
8000

6000

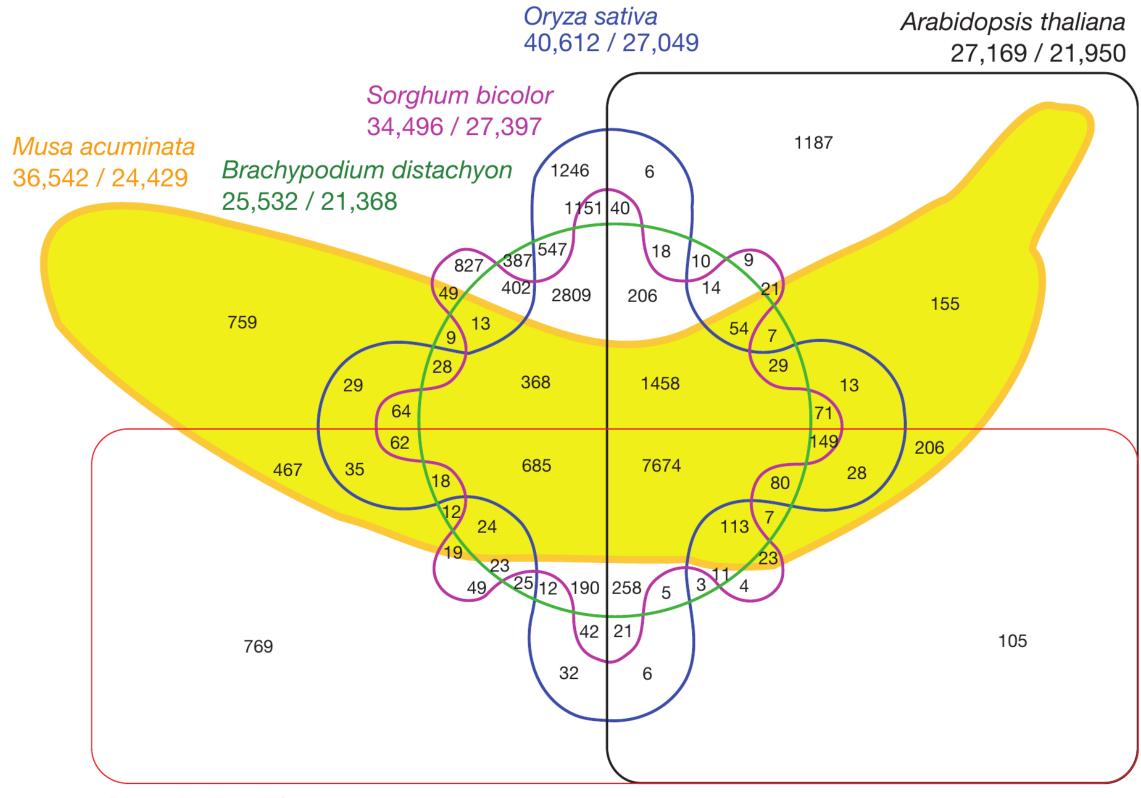
4000

2000 -

				Phoenix_dactylifera
				Arabidopsis_thaliana
				Musa_acuminata
			Br	achypodium_distachyo
				Sorghum_bicolor
				Oryza_sativa
15000	10000 Set Size	5000	ò	



The Banana Chart Redesigned



UpSetR UpSet Plot

Welcome to the UpSetR Shiny App!

UpSetR generates static UpSet plots. The UpSet technique visualizes set intersections in a matrix layout and introduces aggregates based on groupings and queries. The matrix layout enables the effective representation of associated data, such as the number of elements in the aggregates and intersections, as well as additional summary statistics derived from subset or element attributes.

To begin, input your data using one of the three input styles.

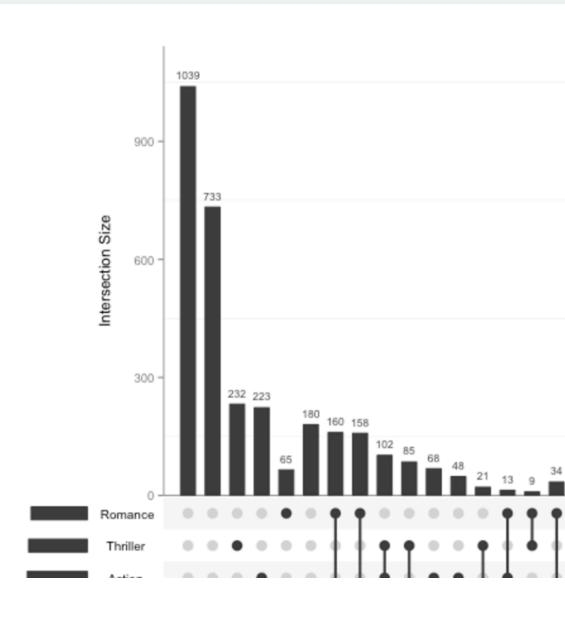
- 1. "File" takes a correctly formatted.csv file.
- 2008) and jvenn (Bardou et al., 2014)
- 3. "Expression" takes the input used by the venneuler R package (Wilkinson, 2015)

To view and explore your data click on the "Plot!" button.

For further details about the original technique see the UpSet website. You can also check out the UpSetR R package and its source code.

If you use UpSetR in a paper, please cite:

Alexander Lex, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, Hanspeter Pfister, UpSet: Visualization of Intersecting Sets, IEEE Transactions on Visualization and Computer Graphics (InfoVis '14), vol. 20, no. 12, pp. 1983–1992, 2014. doi:10.1109/TVCG.2014.2346248



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

2. "List" takes up to 6 different lists that contain unique elements, similar to that used in the web applications BioVenn (Hulsen et al.,

Option 1: File Option 2: List Option 3: Expression

<u>Instructions</u>

The input style of lists is useful when wanting to compare sets by supplying, say a list of gene IDs or SNPs. To use this format enter a list of elements seperated by a comma to each input box. These elements can be entered as numbers, letters, IDs, words, etc. The only limitation to entering the lists is having spaces in the element names. As an alternative an underscore ('_') character can be used to to substitute for the spaces. To give each set a name, enter the names into the bars where the word 'List' followed by a number is grayed out.

To see how the list format works copy and paste each list of letters into their respective input boxes.

List 1: A, B, C, D, E, F, G, H List 2: A, B, D, F, I, J, K, L List 3: A, H, J, M, N, O, P, Q List 4: B, L, O, P, R, S, T, U

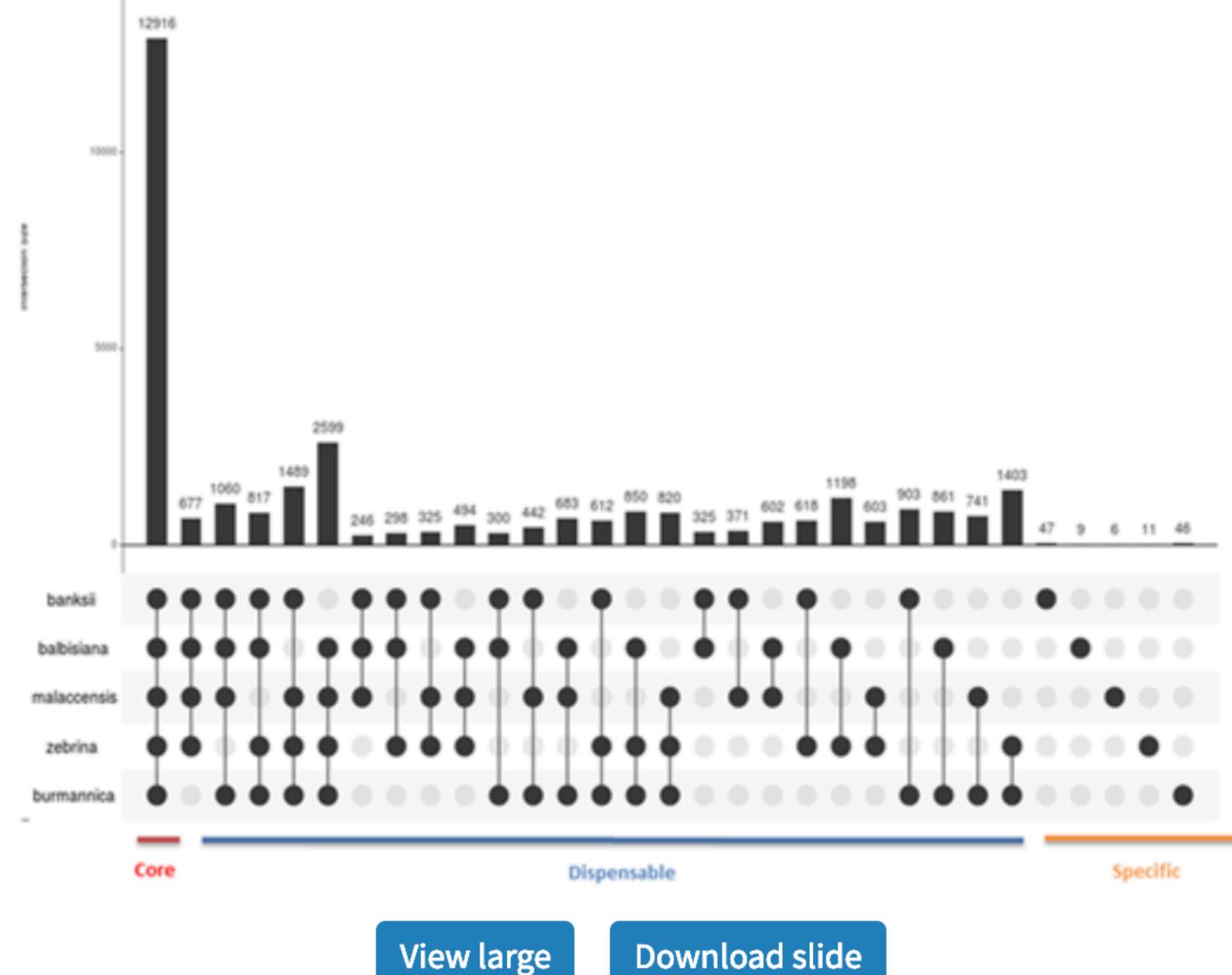
List 1	1,
	//
List 2	11
List 3	1,
	//
Plot!	

☆ 🖽 🗞 🗀 📣 🗊 🚺 🕨 🐾 😫

List 4	1,
List 5	1,
	1,
List 6	1,
	1,



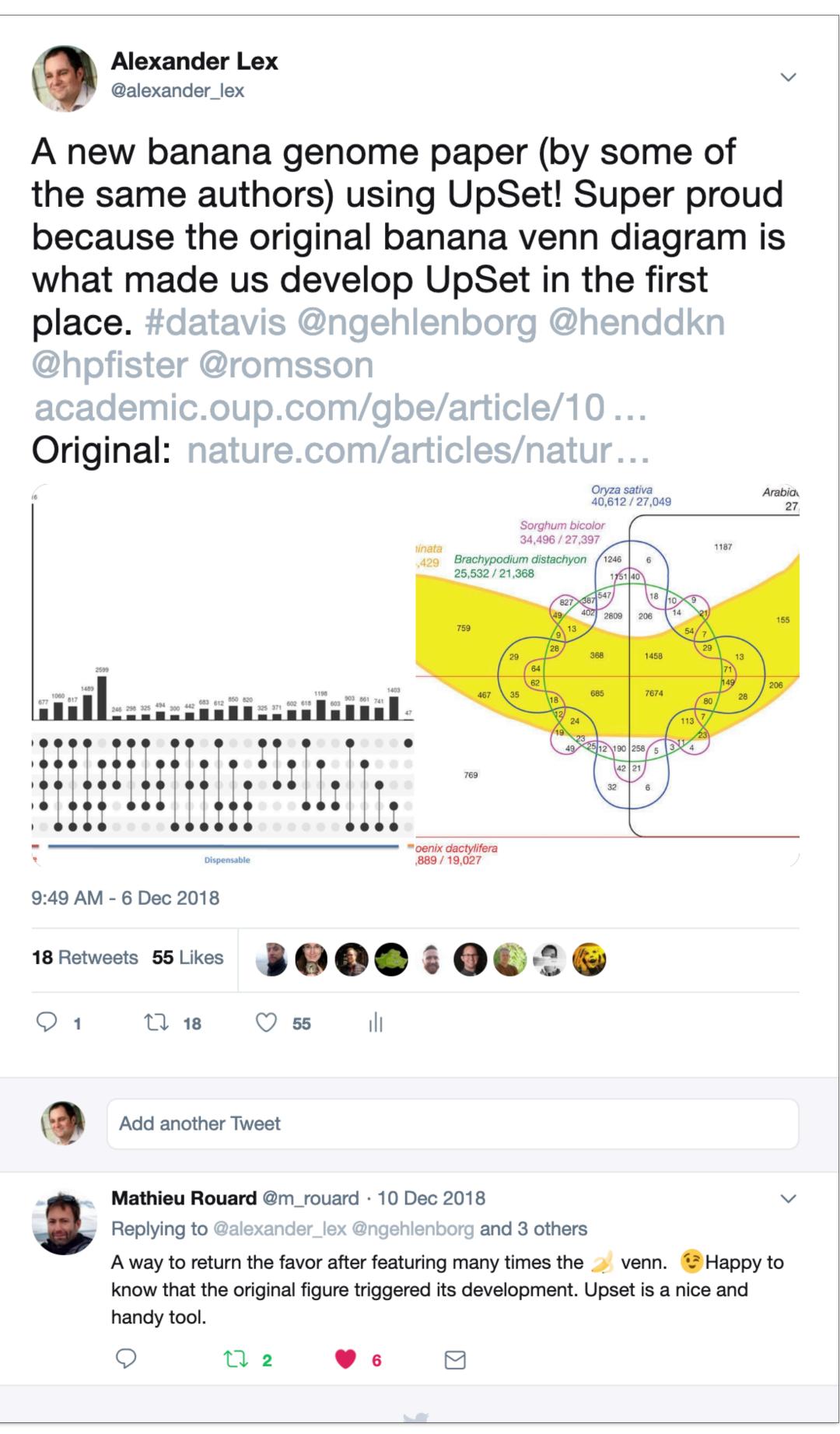
FIG. 1.



—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).

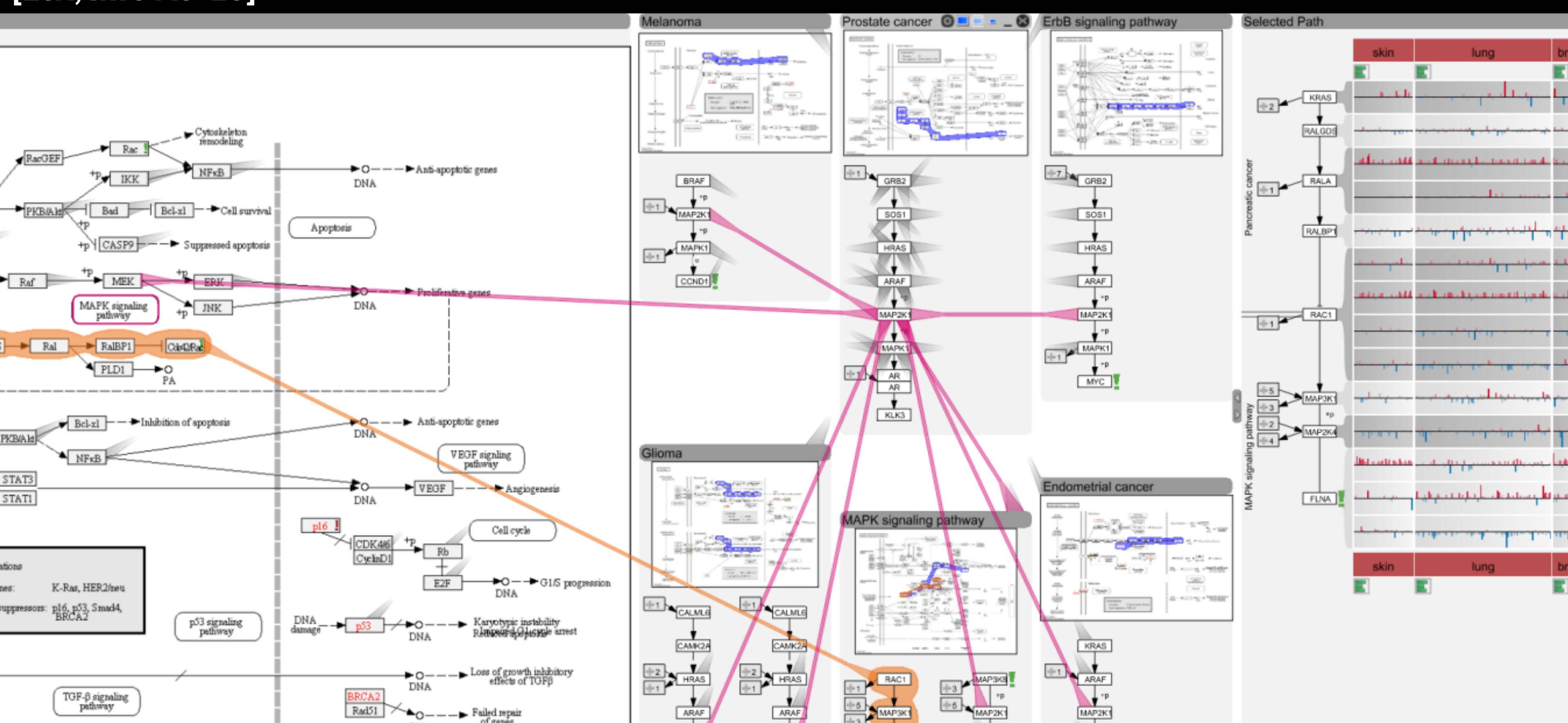
Download slide

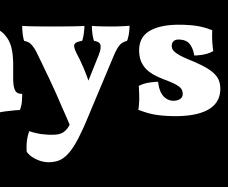






[Partl, BioVis '12] **Best Paper Award** [Partl, BMC Bioinf. '13] [Lex, InfoVis '13]





K MACHENNAC A CAL Data and Palmasaus

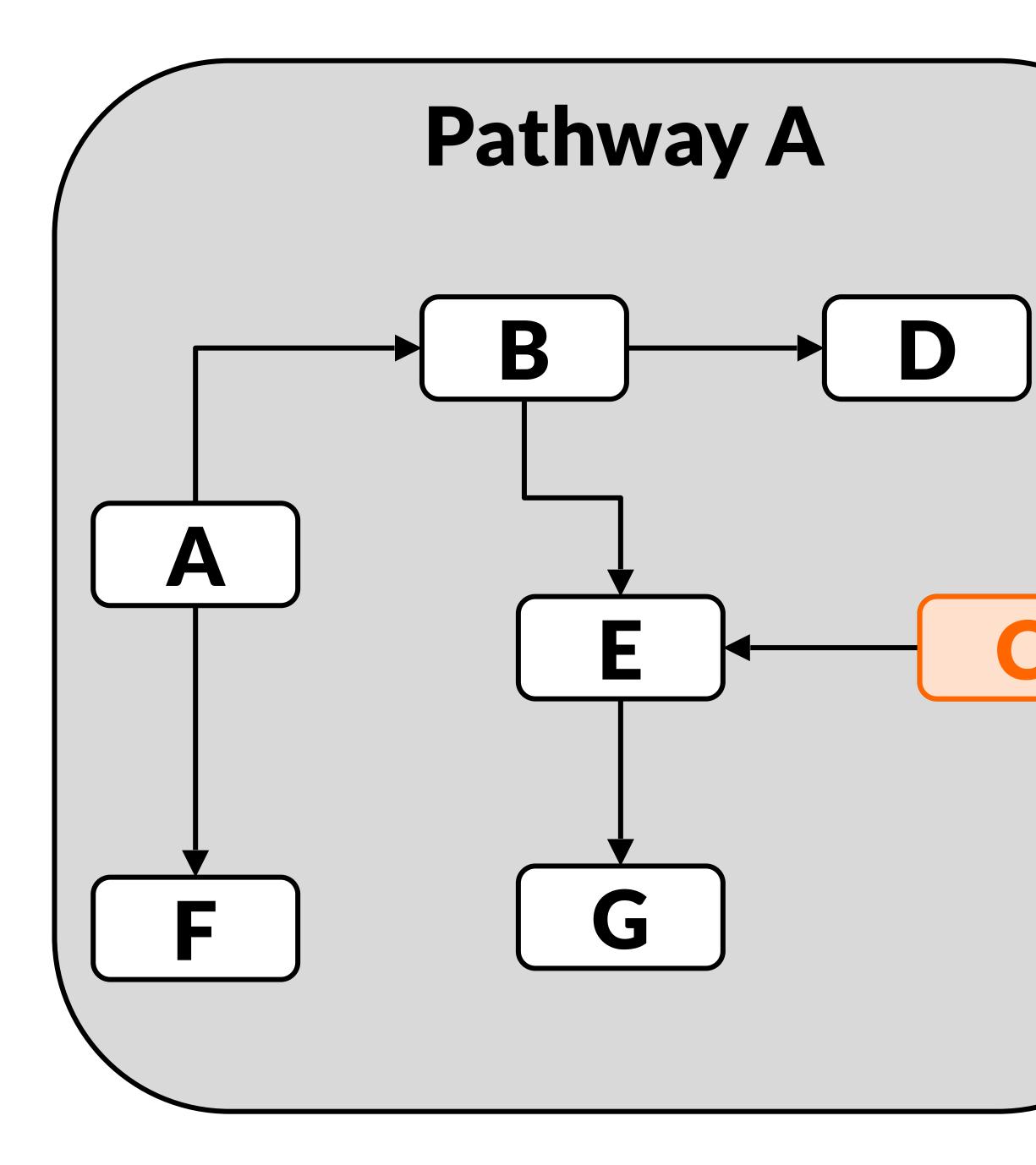




Cannot account for variation found in real-world data Branches can be (in)activated due to mutation, changed gene expression, modulation due to drug treatment,

etc.

Many Node Attributes

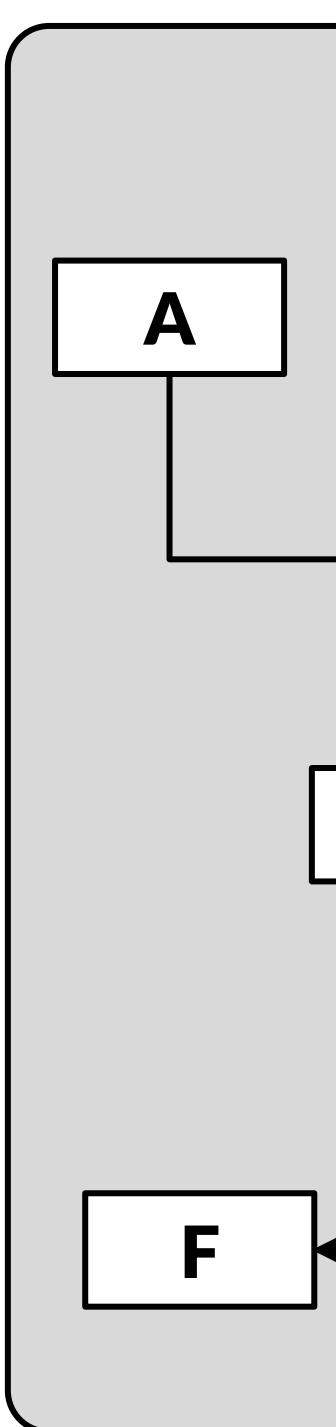


How to visualize experimental data on pathways?

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

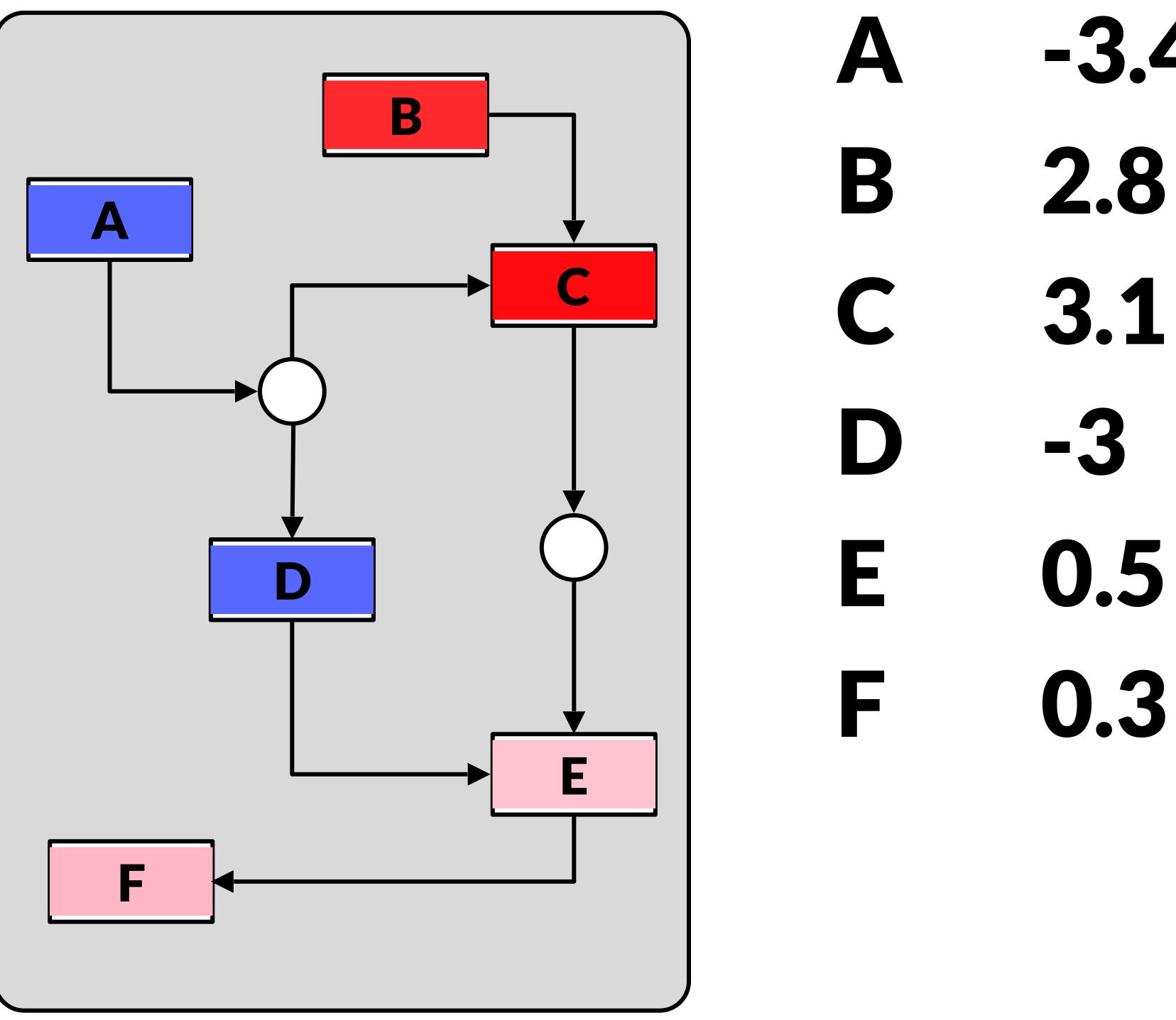
Node	Sample 1	Sample 2	Sample 3	•••
A	low	low	very high	•••
B	normal	low	high	•••
С	high	very low	normal	•••
•••	•••	•••	•••	

Good Old Color Coding



A B B Ε D

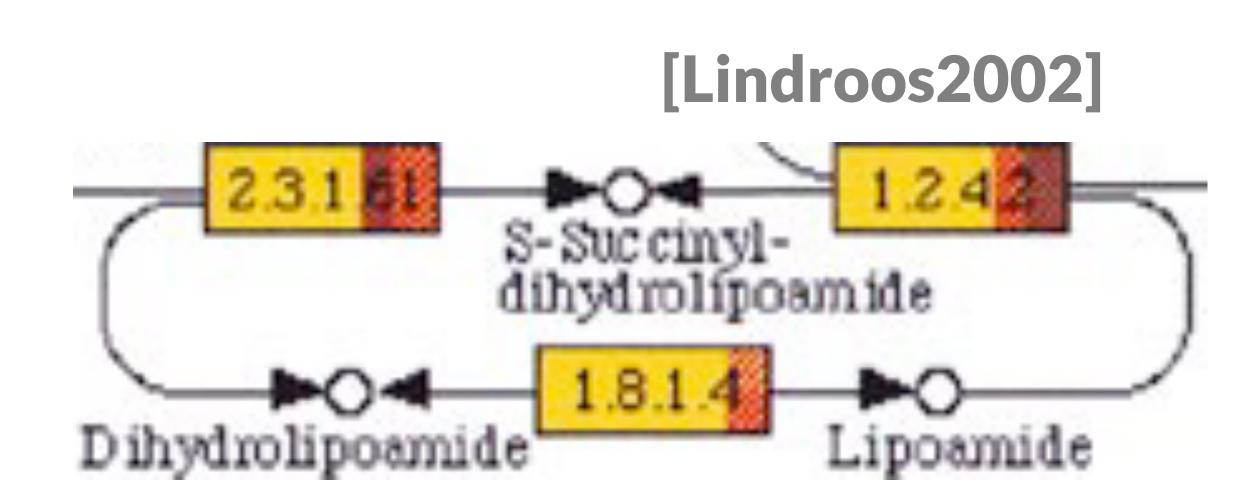
-3.4 2.8 3.1 -3 0.5 0.3



Good Old Color Coding

- A-3.44.25.14.2B2.81.81.31.1C3.1-2.22.42.2

 - -2.8 1.6 1.0
 - 0.3 1.1 1.3
 - 0.3 1.8 0.3

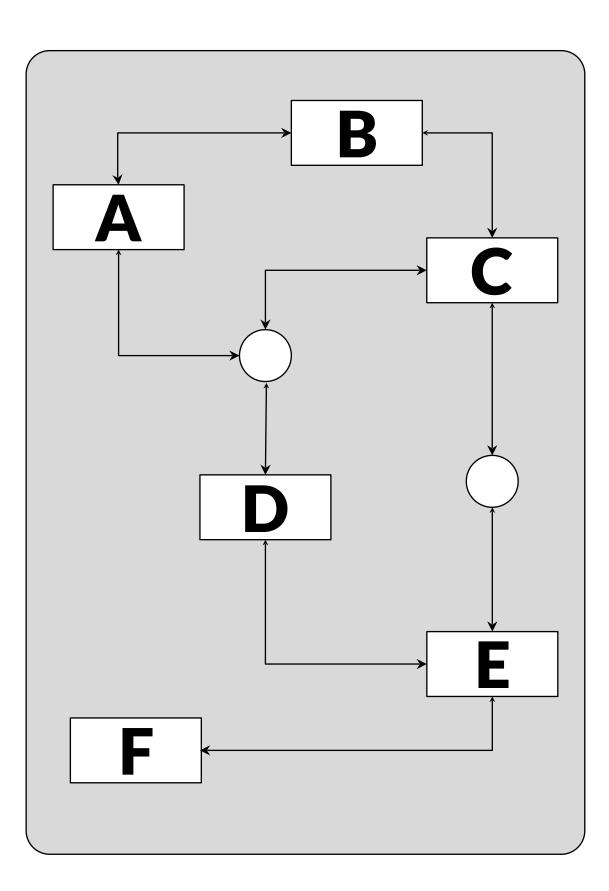




Challenge: Supporting Multiple Tasks

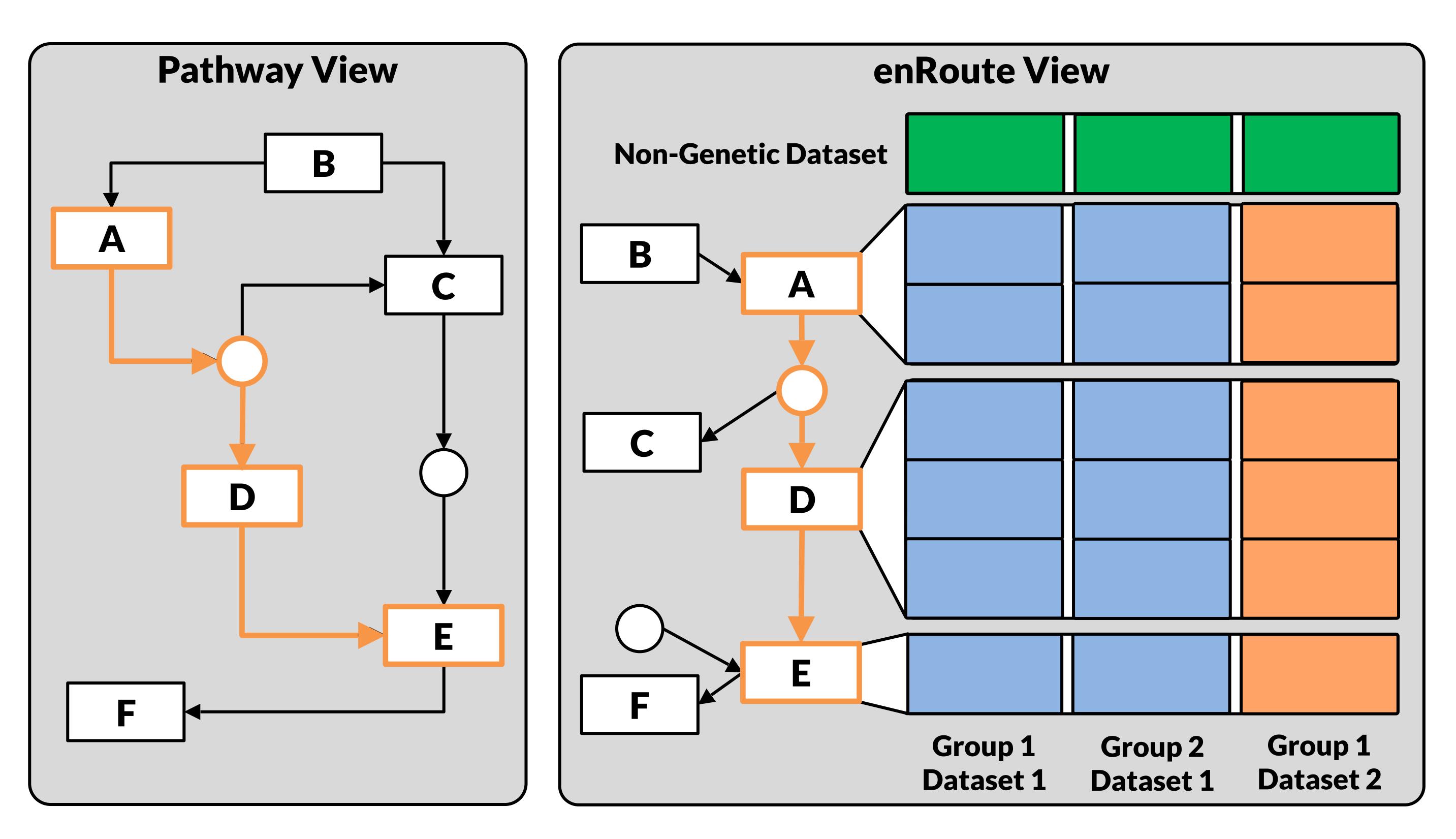
Two central tasks: Explore topology of pathway Explore the attributes of the nodes (experimental data) Need to support both!



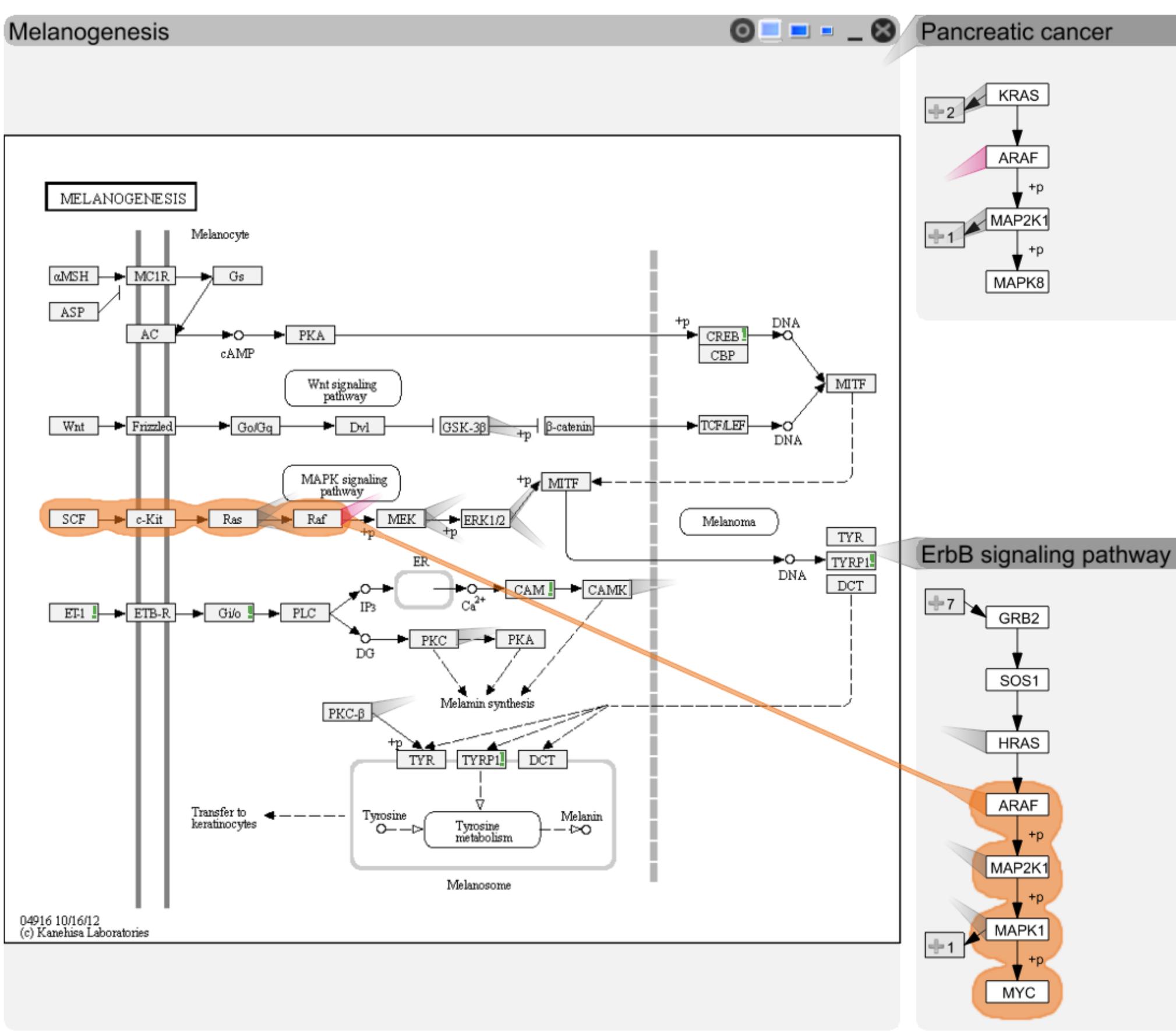


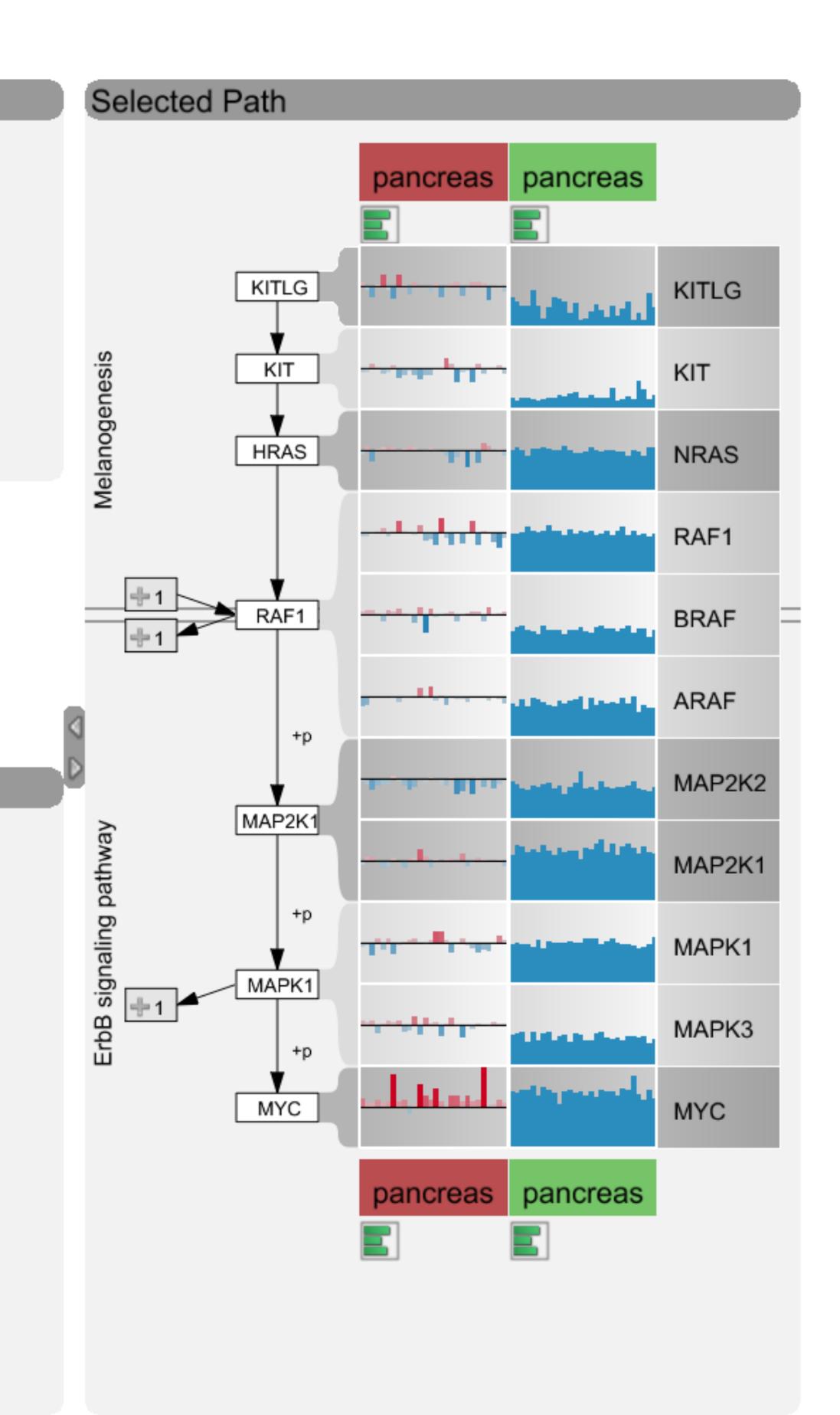
	Sample 1	Sample 2	Sample 3
e 1	1	1.1	0.4
e 2	2	0.5	1.2
e 3	1.4	0.2	0.5
e 4	0.3	0.5	0.7

Concept



enRoute



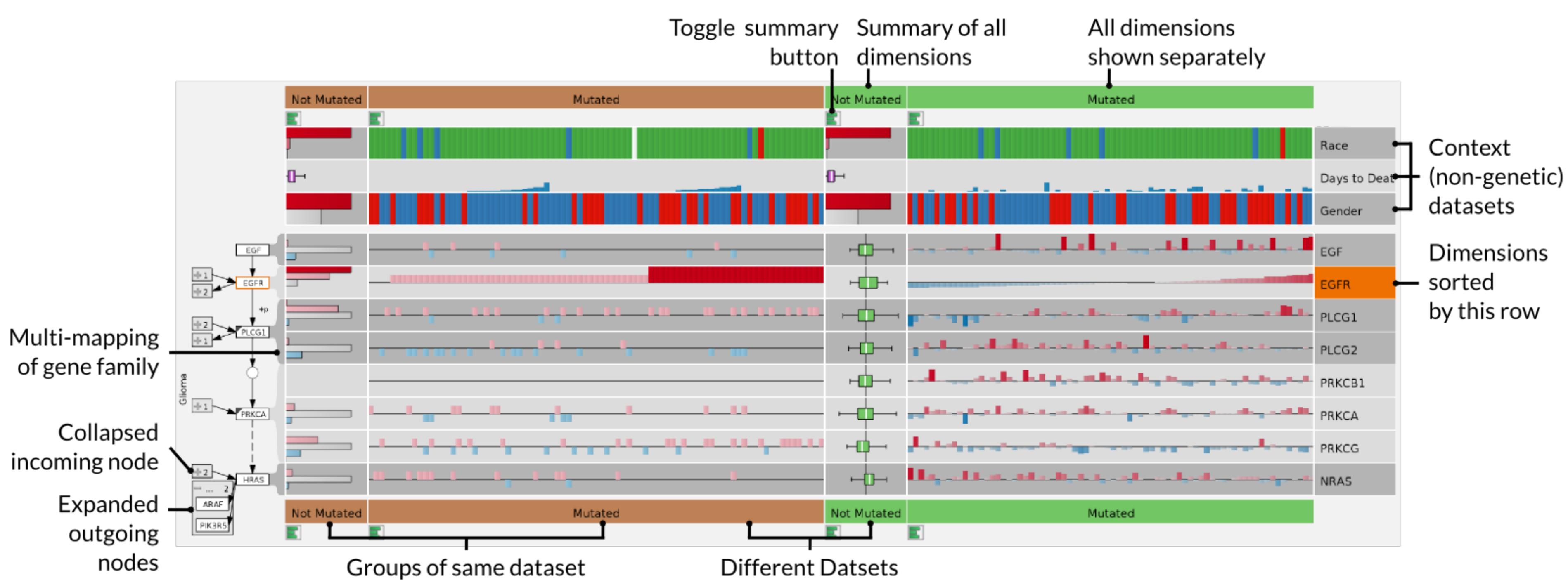


Entourage 😂	
Pathways	
Pathwaya	
Filter:	
<none></none>	
1 C donor	8
2-Oxocarboxylic acid	
ABC transporters	
ABC-family proteins	
ACE Inhibitor Pathwa	
Acetylcholine Synthes	
Acute myeloid leukem	
Adherens junction	
Adipocyte TarBase	
Adipocytokine signali	
Adipogenesis	
Advanced glycosylatio	
Aflatoxin B1 metaboli	
African trypanosomias	
AGE/RAGE pathway	
AhR pathway	
Alanine and aspartate	
Alanine, aspartate an	
Alcoholism Aldosterene regulated	
Aldosterone-regulated Allograft rejection	le la constante de la constante
Allograft rejection	3
Alpha 6 Beta 4 signal	P
alpha-Linolenic acid	
Alzheimer's disease	
Alzheimers Disease	
amino acid conjugatio	
amino acid conjugatio	
Amino sugar and nucl	
Aminoacyl-IRNA 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	8
	-

3

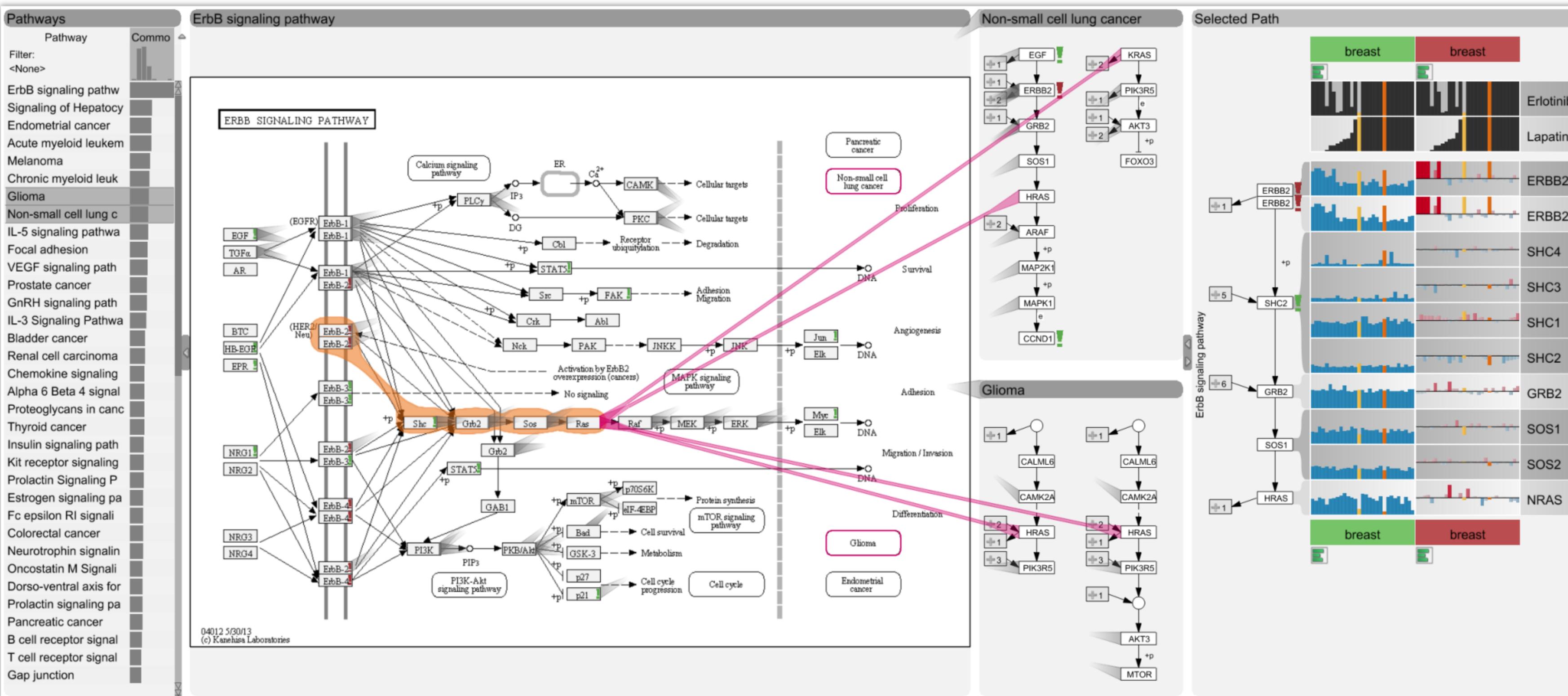
10 Au

Selected Path



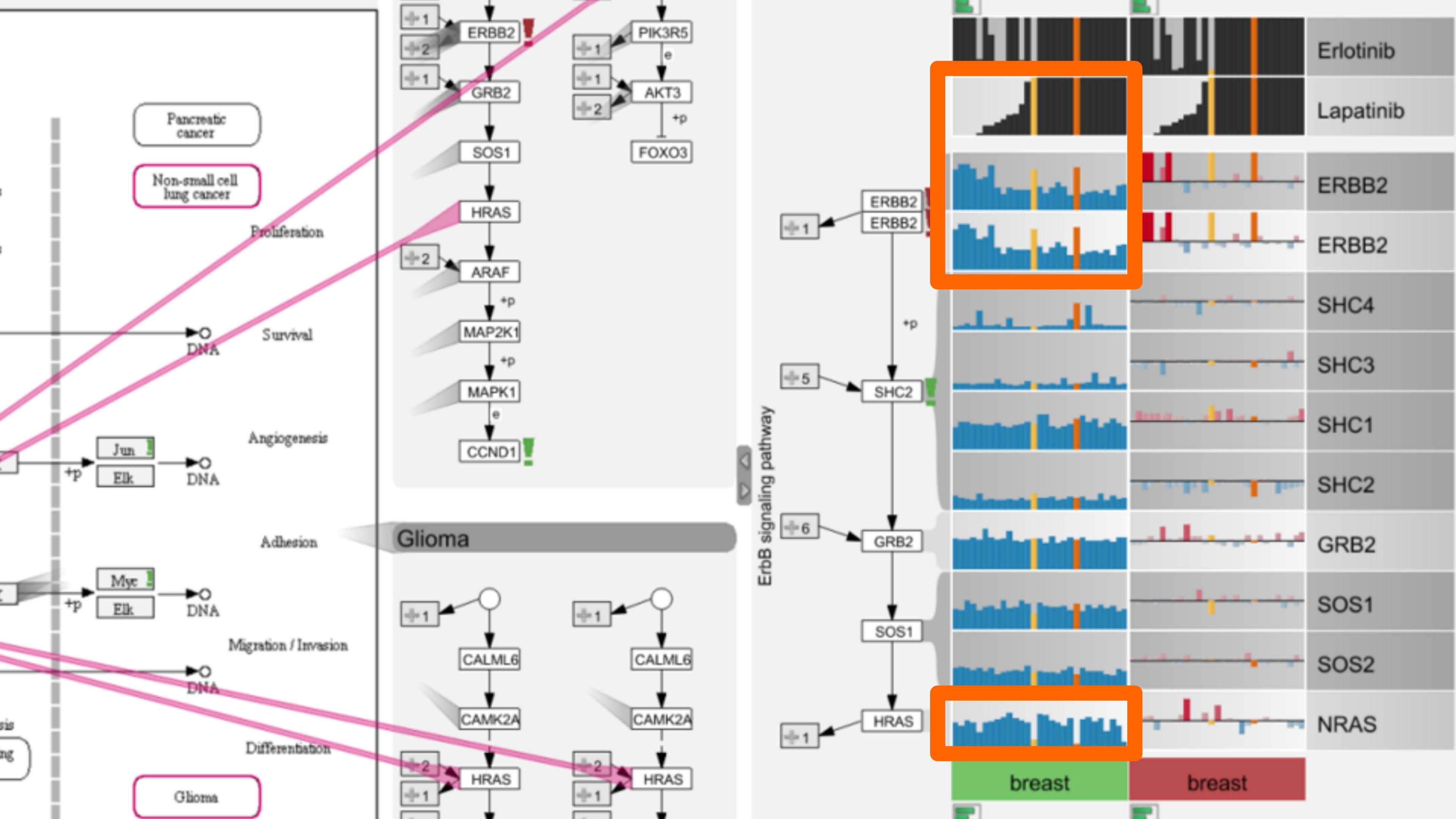
Different Datsets

Case Study: CCLE Data



		1	1	l	
ib)			ļ	
ni	b				
1	1		1	Ĺ	
2				l	
2					
	1	1	1	l	
				ļ	
				l	
				i	
				ł	
				l	
				ł	











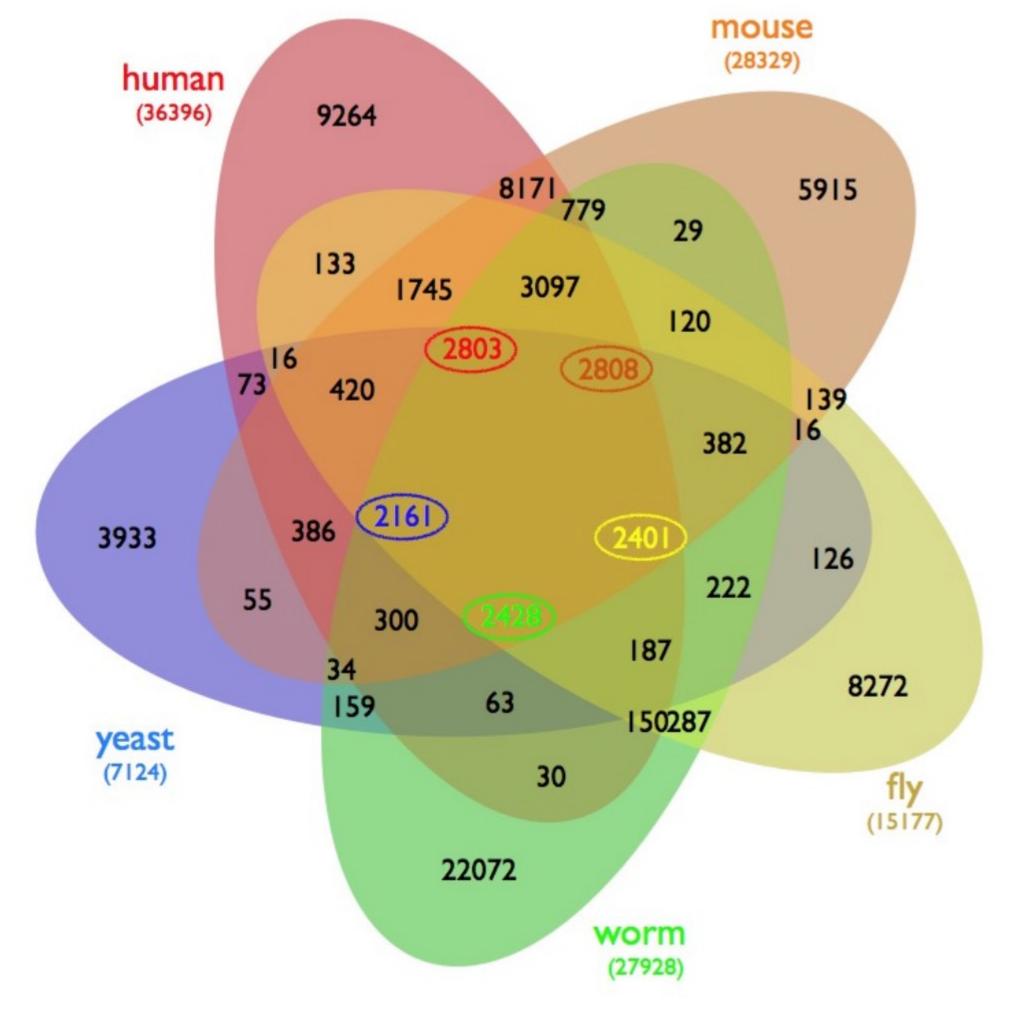
Visualization Design Strategies

3. Use queries

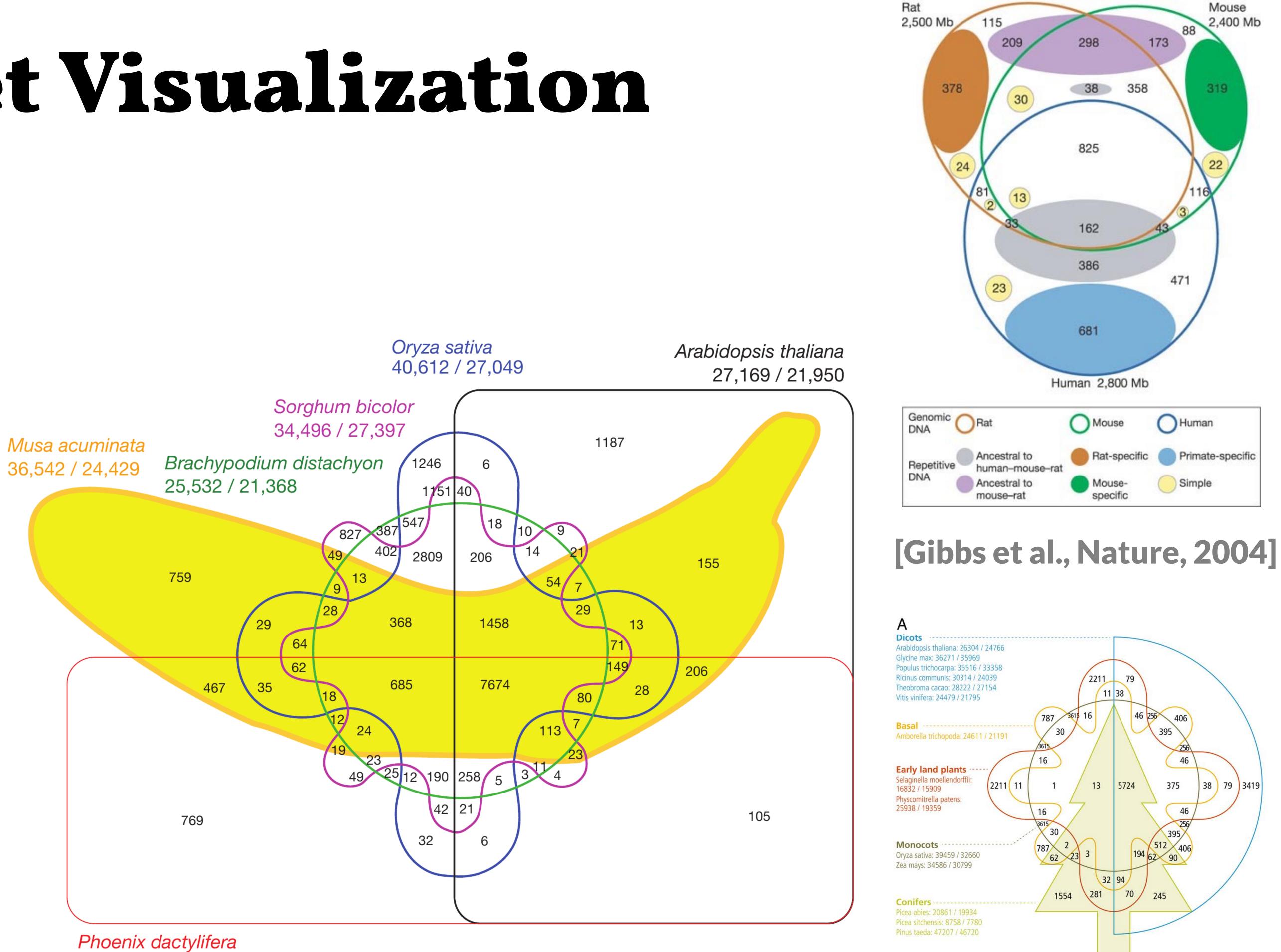
1. Encoding channel primacy 2. Show relationships explicitly 4. User color sparingly 5. Enable annotation / provenance

1. Encoding Channel Primacy Most important data is assigned most powerful encoding channel (position)

Example: Set Visualization



[Wiles et al., BMC Systems Biology]



28,889 / 19,027

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

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



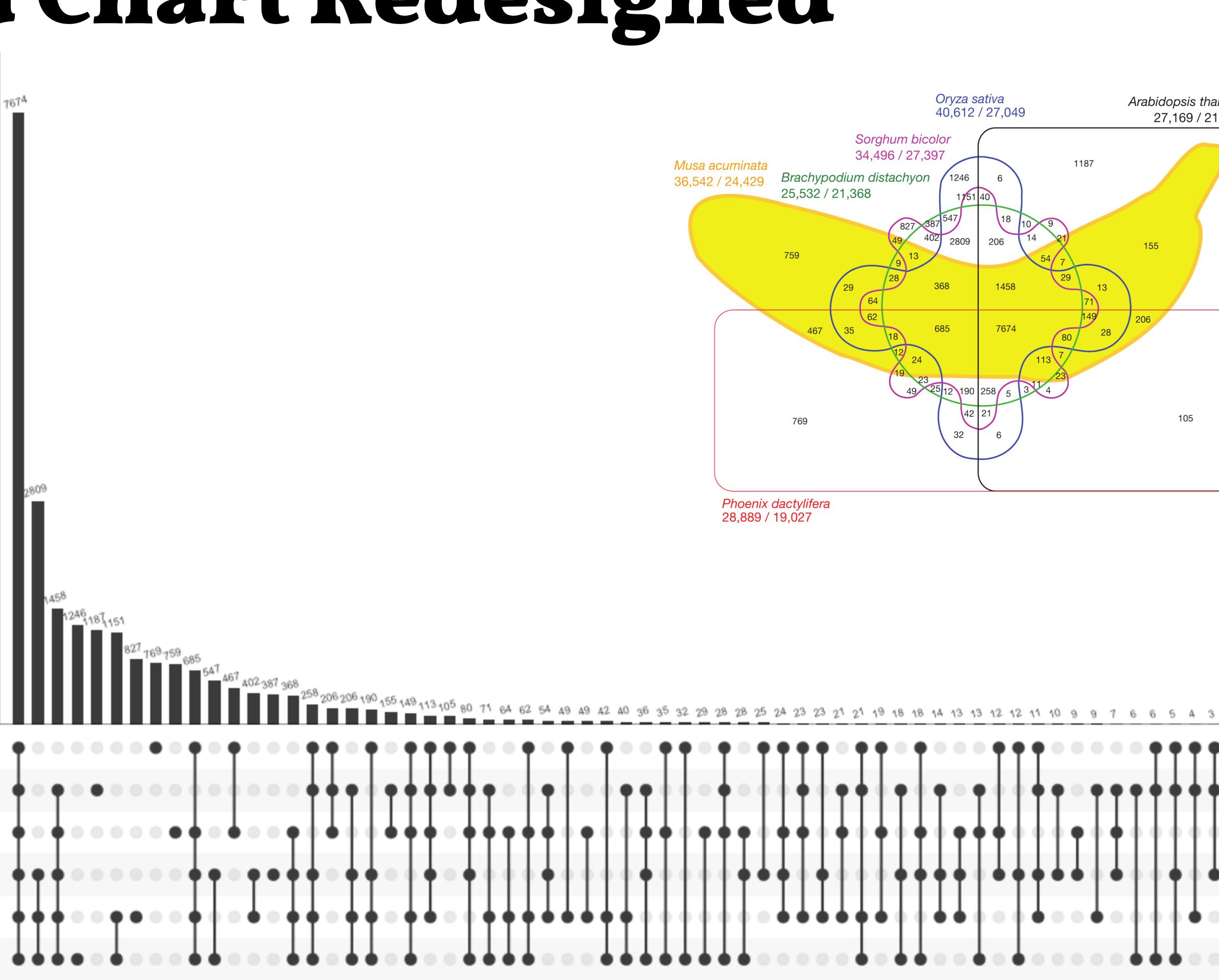
8000

6000

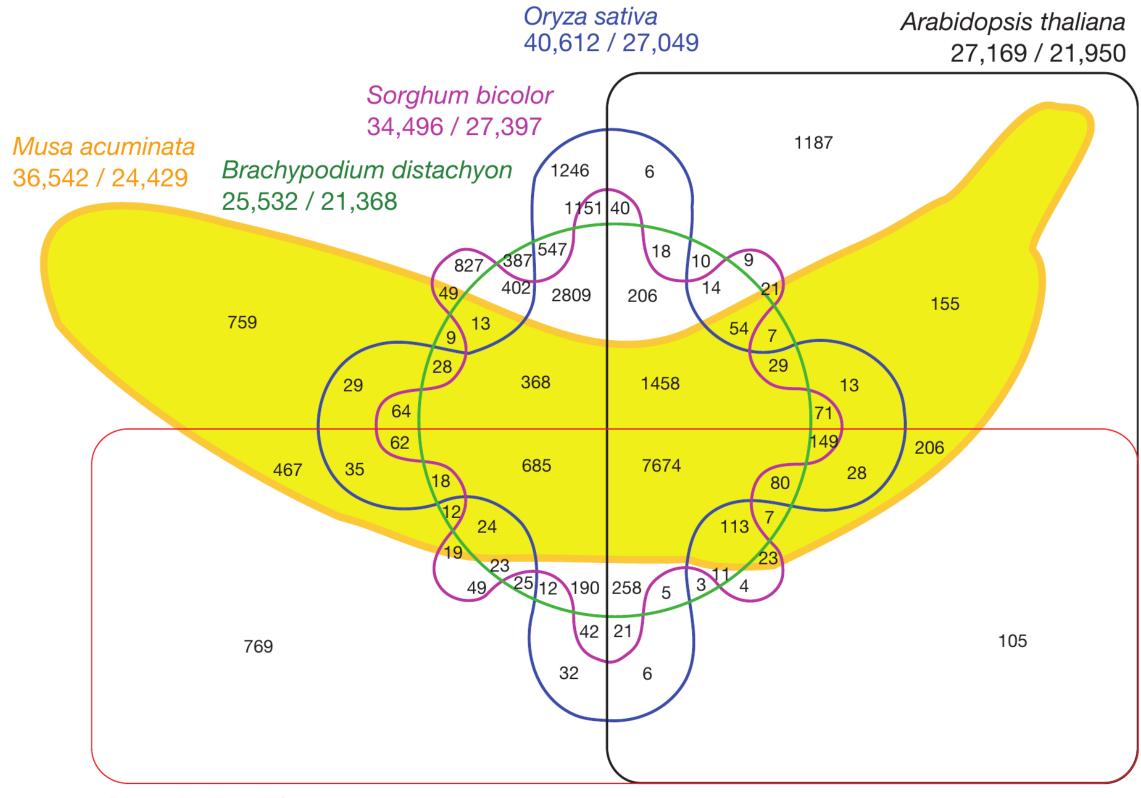
4000

2000 -

				Phoenix_dactylifera
				Arabidopsis_thaliana
				Musa_acuminata
			Br	achypodium_distachyo
				Sorghum_bicolor
				Oryza_sativa
15000	10000 Set Size	5000	ò	



The Banana Chart Redesigned



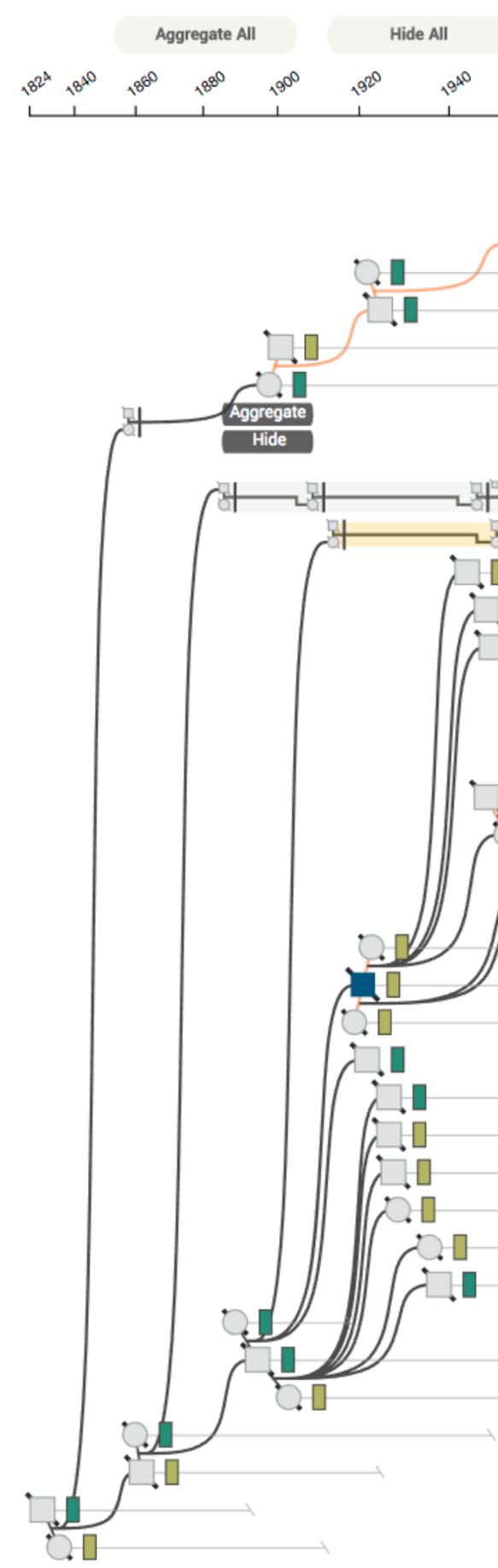
different views

2. Show relationships explicitly Don't use highlighting to connect

Use smart layouts (position) or connectivel

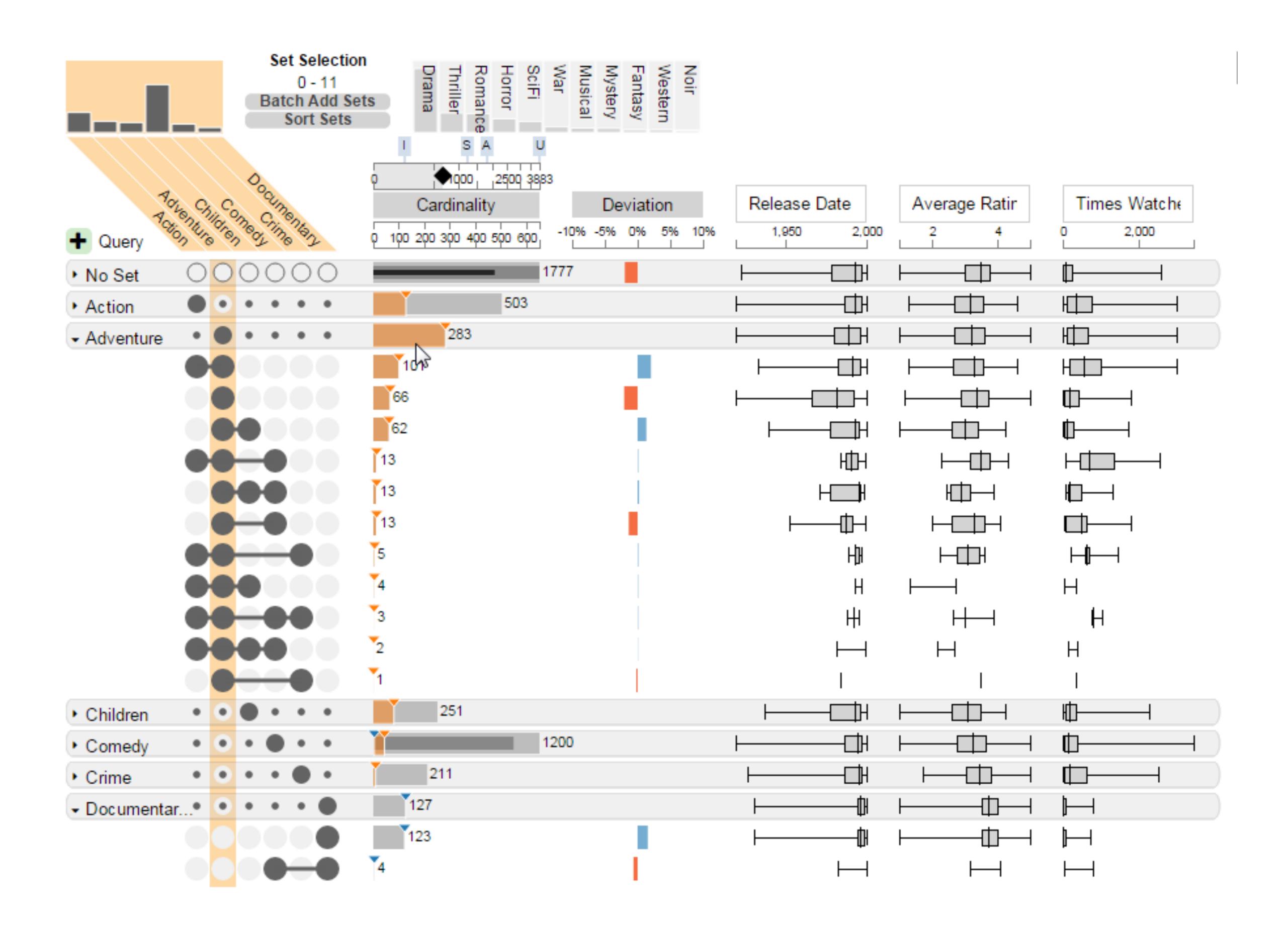




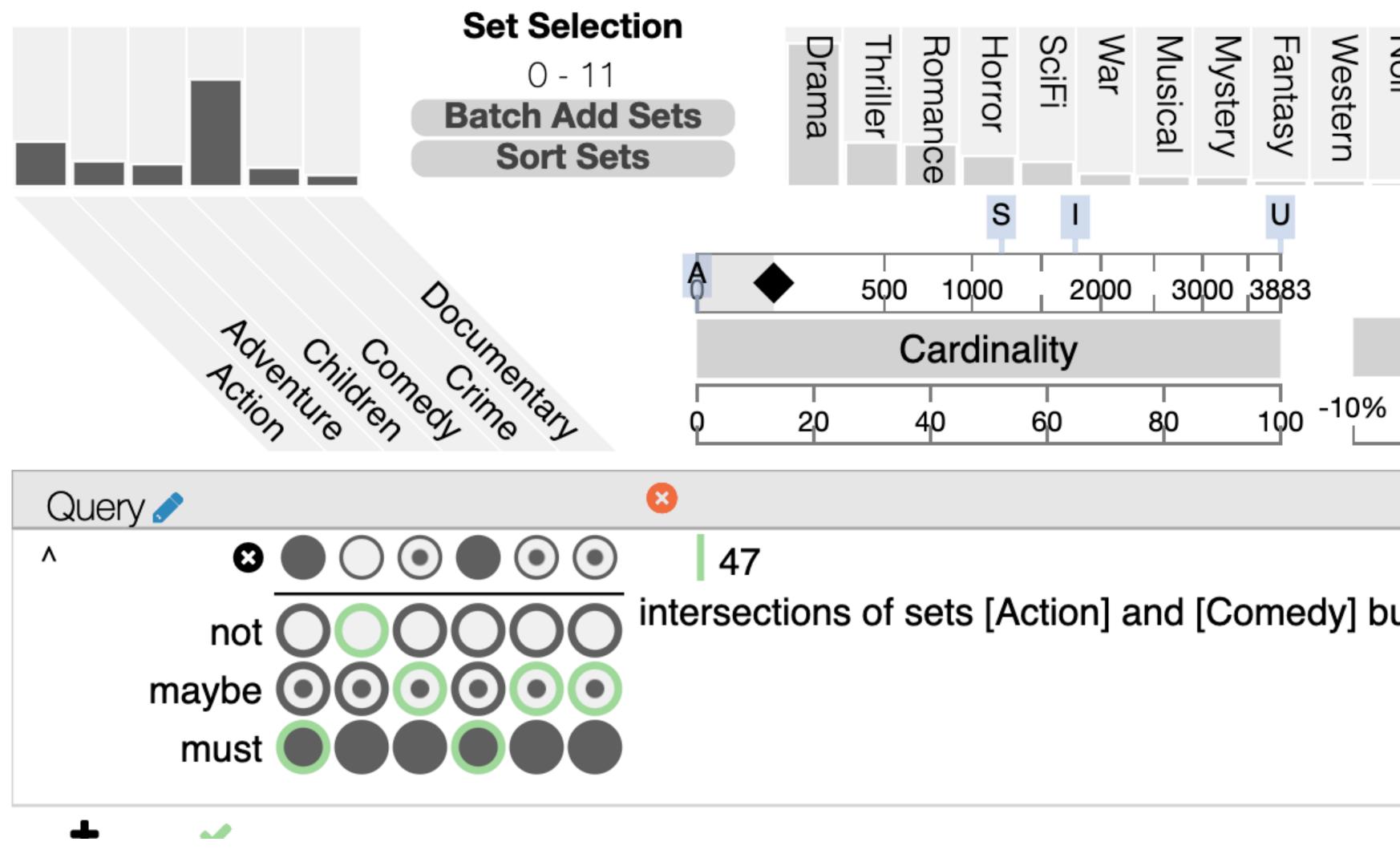


	Expand All				KindredID	RelativelD	set	f) dece	ased	ide Deptes	asion (True) Age1D / Total:32	Depression Nr.Dian Total:60	J.Depressi Bipole	on Agen Bi Agen Bi Total:30	polar Nr.Diag.P Total:60	bipolat MaxBMI Total:59	AgeMaxBMI Total:58	1ace 1	white)	other)	Native f	America Black) Face	Asian) (His	panic)
196	0867 0	206	\$	2017			44%	60%	10%	52%		<u> </u>	48%			ى مىللى		50 %	0.0%	0.0%	0.0%	0.0%	0.0%	
				3	903988						0 65	i 0 7	oʻ	0 <u>36</u>	0 20	15 45	о <u>во</u>		*	*	*	T	*	
			4		903988	#13789	Ē.			÷.	0	•		•			•		_	_	-	_	_	gun
2 ª				3	903988			1	ī.			%		00	0.00		e b		_	_	_	-	-	guin
				1	903988	#48154					•								_	_	-	-	_	
			_	1	903988	#54315													-	-	-	-	-	
				1	903988	#67703					_			_	_				-	-	-	-	-	
				1	903988	#9278					•								-	-	-	-	-	
				1	903988	#26695					—			—		0	0		-	-	-	-	-	aspl
2				7	903988						0,8			• • •	9494P	%	46.0 0		-	-	-	-	-	
			23	1	903988	#40329					•						•		-	-	-	-	-	carb
				1	903988	#5170											•		-	-	-	-	-	
			\rightarrow	1	903988	#43983					•			•	•		•		-	-	-	-	-	
				1	903988	#5181													-	-	-	-	-	
				1	903988	#42479										•	•		-	-	-	-	-	
				1	903988	#139						0							-	-	-	-	-	
		\rightarrow		1	903988	#1967													-	-	-	-	-	card
				1	903988	#5172								•					-	-	-	-	-	
				1	903988	#5171													-	-	-	-	-	
				1	903988	#10841											0		-	-	-	-	-	
				1	903988	#66710													-	-	-	-	-	
//				1	903988	#6436													-	-	-	-	-	
/	\			1	903988	#65130						00					•		-	-	-	-	-	stra
		\rightarrow		1	903988	#65129													-	-	-	-	-	
				1	903988	#6428					•								-	-	-	-	-	
				1	903988	#4547						•		•			0		-	-	-	-	-	
				1	903988	#65491										•	•		-	-	-	-	-	
				1	903988	#39045										•			-	-	-	-	-	
		~		1	903988	#18524																-		
			\	1	903988	#4556																-		
				1	903988	#4551				_							0					-		
					903988																	-		
	ι				903988														-	-	-	-	-	
	\rightarrow				903988							_		_								-		
\rightarrow																	•						-	
					903988																			
					903988												•							
				1	903988	#61318										0			-	-	-	-	-	





3. Use queries especially for big data

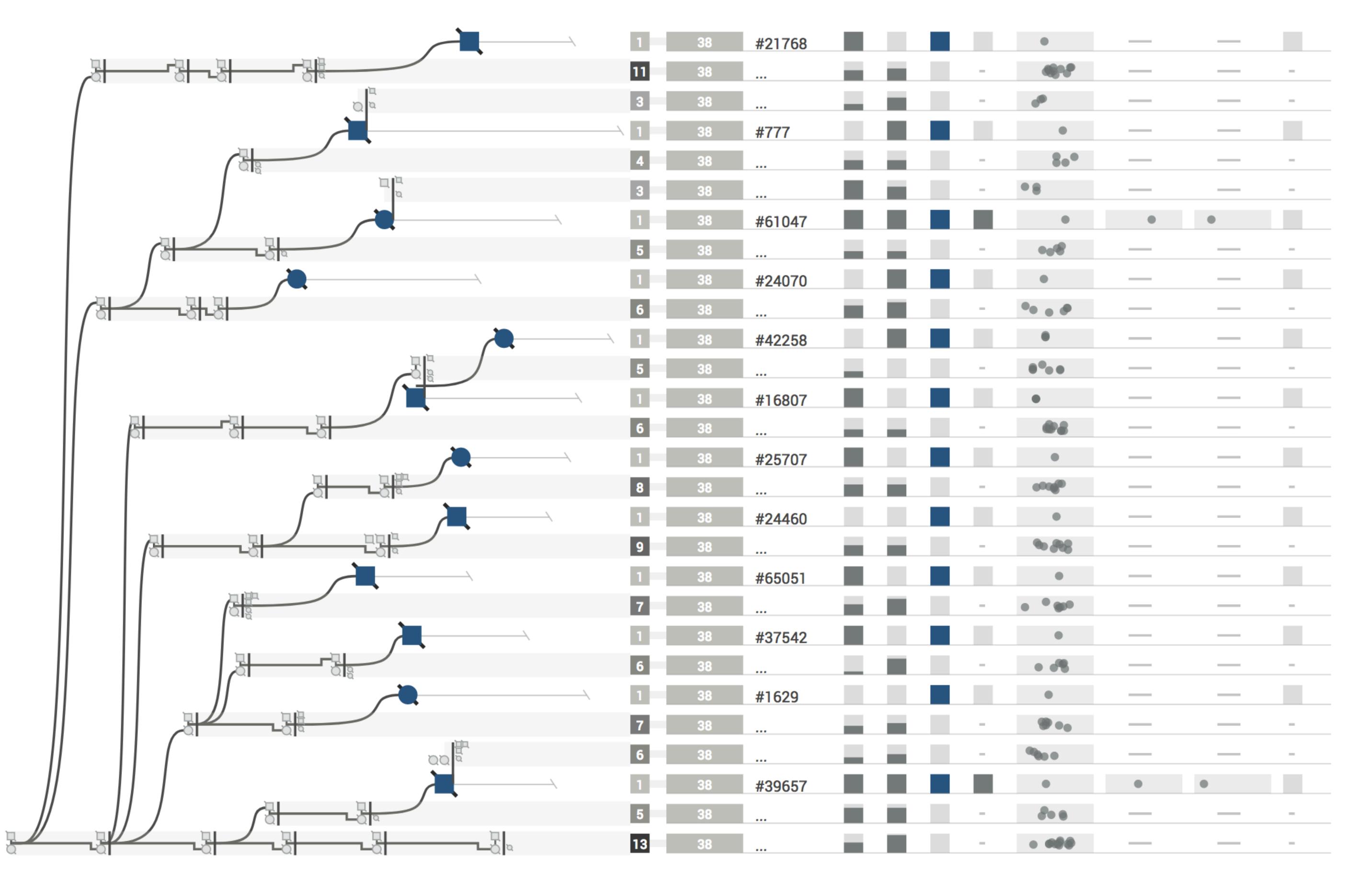


	D	eviatio	n		Release Da	ate		
%	-5%	0%	5%	10%	1,950	2,000		
				'				
but excluding set [Adventure]								
		U	-	-				

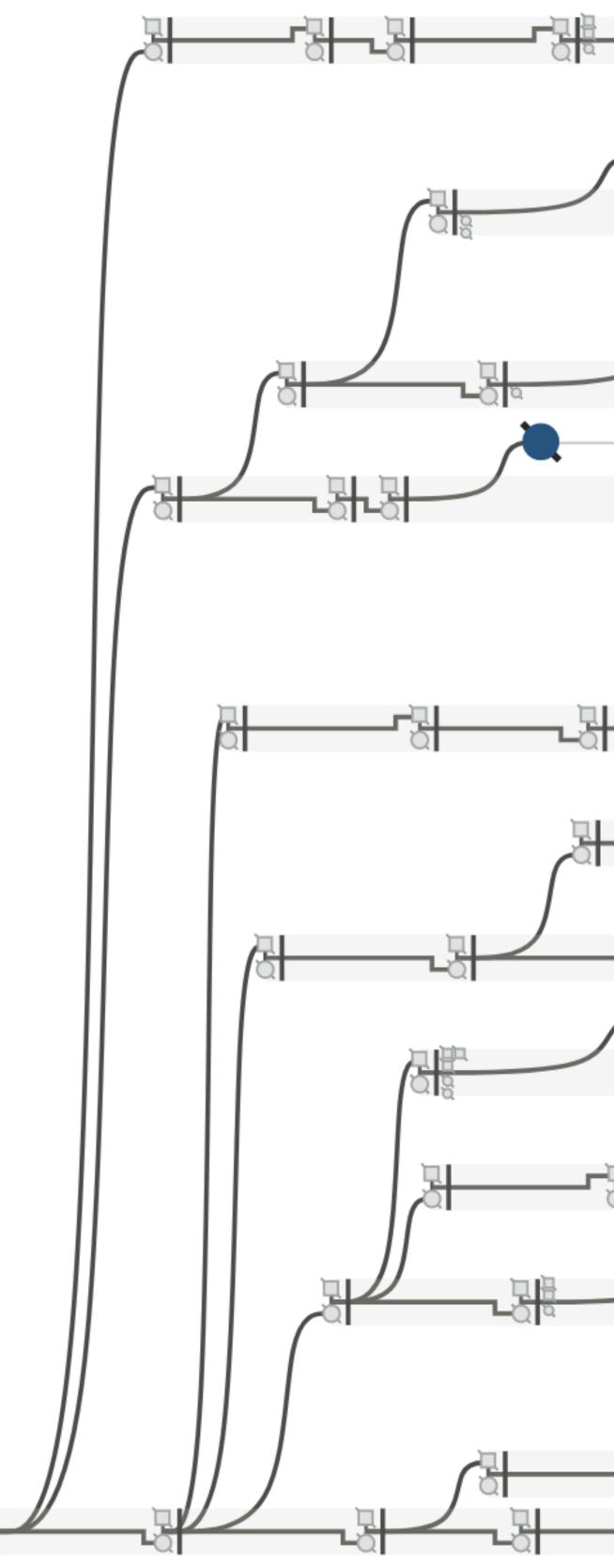
Noir

4. User color sparingly Limit use to encode data Primarily use it to highlight items of





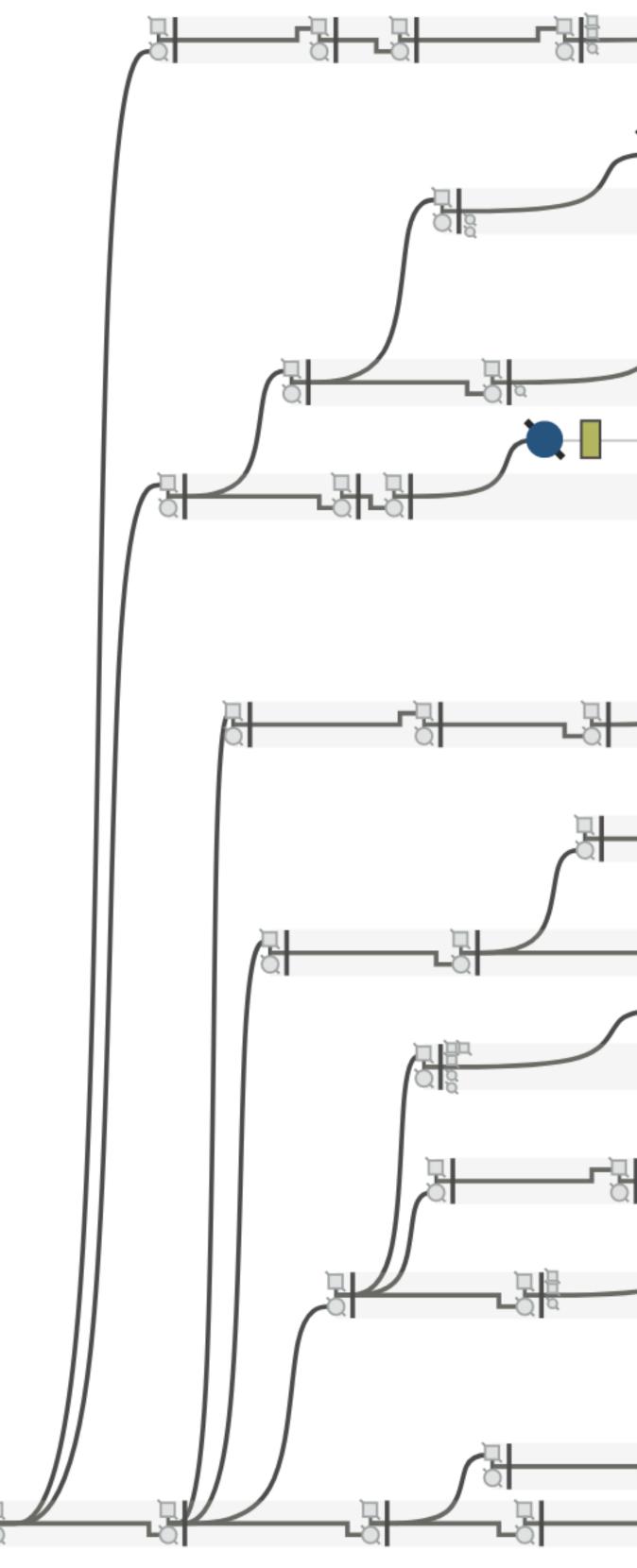
Only one color for primary attribute (suicide, blue)



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

1	38	#21768			٠			10.1
1				-				-
3				-				-
	_	#777			•			
4	38			-	8.*			-
۲ <mark>۲</mark> ۵ ۵	38			-	• 8			-
	38	#61047			•	•	•	
5	38		- He - He	-	•••			-
1	38	#24070			•			
6	38			-	••••			-
	38	#42258			۲	—		
5	38			-	••••			-
	38	#16807			•			
6	38			-	•			-
	38	#25707			•			
8	38			-	0.000			-
	38	#24460			•			
9	38			-	%			-
	38	#65051			•			
7	38			-	• •			-
	38	#37542			•			
6	38			-	• • • •			-
		#1629			•			
7				-	9 •••			-
00 6	38			-	*** **			-
	38	#39657			•	•	•	
				-	8			-
<u></u>	38			-	• ••••			-



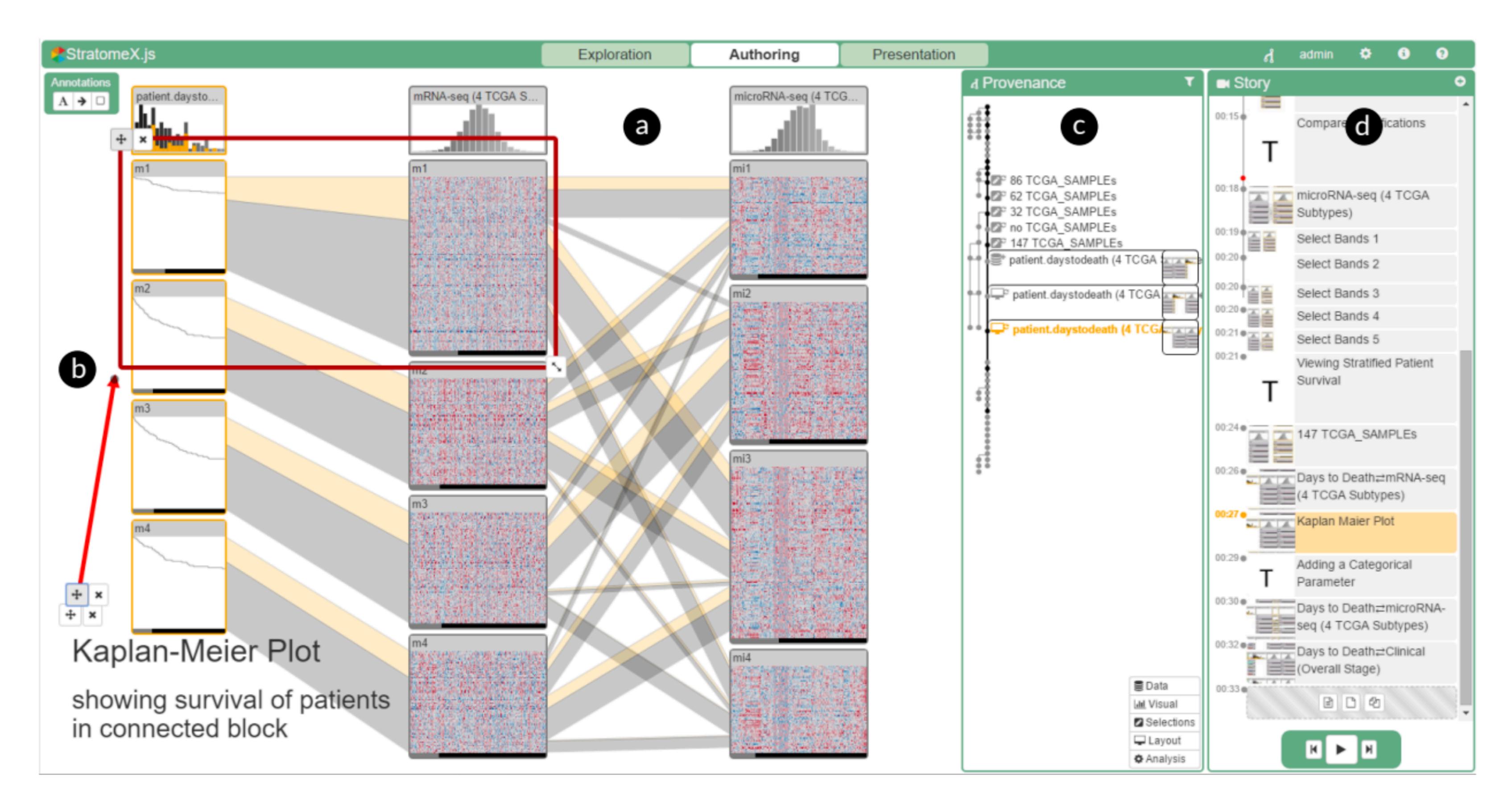


Adding a color for an additional attribute (deceased yes/no, green)

#21768 . . 38 11 _ _ _ 38 ... 3 #777 ... #61047 5 #24070 6 #42258 5 38 #16807 ... #25707 38 ... #24460 #65051 #37542 #1629 38 6 #39657 13 38 ... ÷.

	•			
-	800			-
	9			-
	•			
	•••			-
	••			-
	•	•	•	
	••••			-
	•			
-	•••			-
	۲			
-	8000			-
	8			
-	•			-
	•			
-	000 ⁰⁰			-
	•			
-	00 0 <u>6</u> 8			-
	•			
-	• • •			-
	•			
	• %			-
	•			
	•••			-
	90.0			-
	•	•	•	
	Ø ⁰ e			-
	• •			-

5. Enable annotation / provenance What did you see / think when looking at this visualization? How did I get here? Can I go back?



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

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















THE UNIVERSITY OF UTAH

visualization design lab



