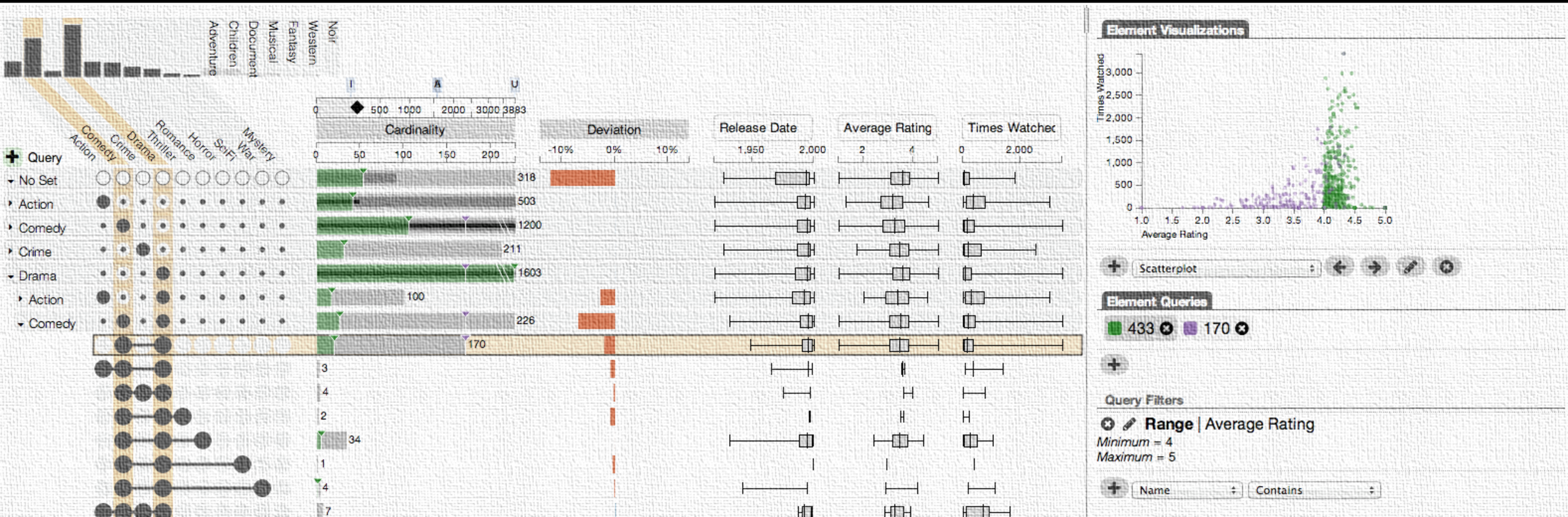
Alexander Lex
@alexander_lex
http://alexander-lex.com

University of Utah
Scientific Computing Institute
December 3, 2014

HARVARD
School of Engineering
and Applied Sciences

Enabling Scientific Discovery through Interactive Visual Data Analysis



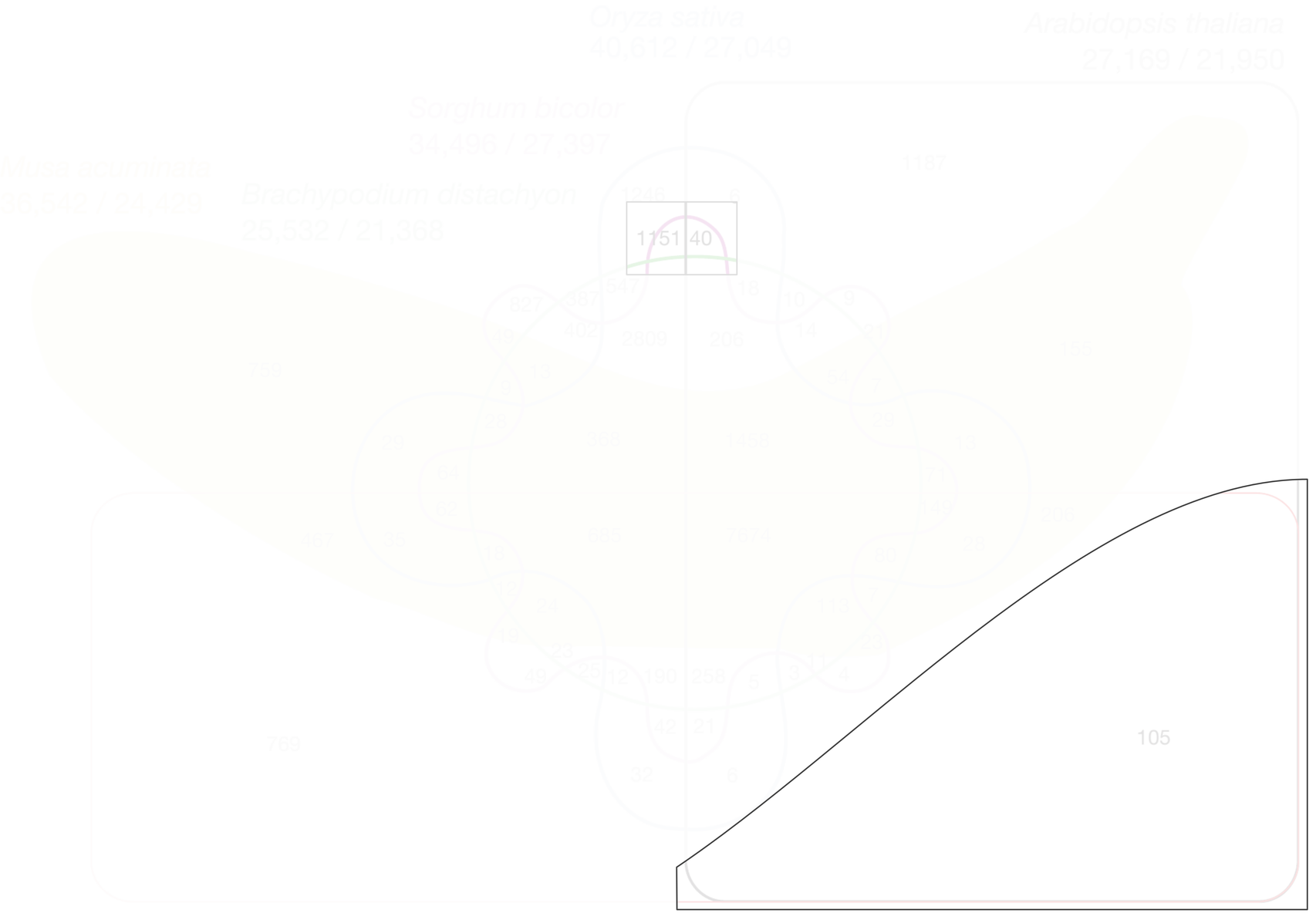
visualization *pictures*

The purpose of computing is insight, not numbers.

- Richard Wesley Hamming

- Card, Mackinlay, Shneiderman

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

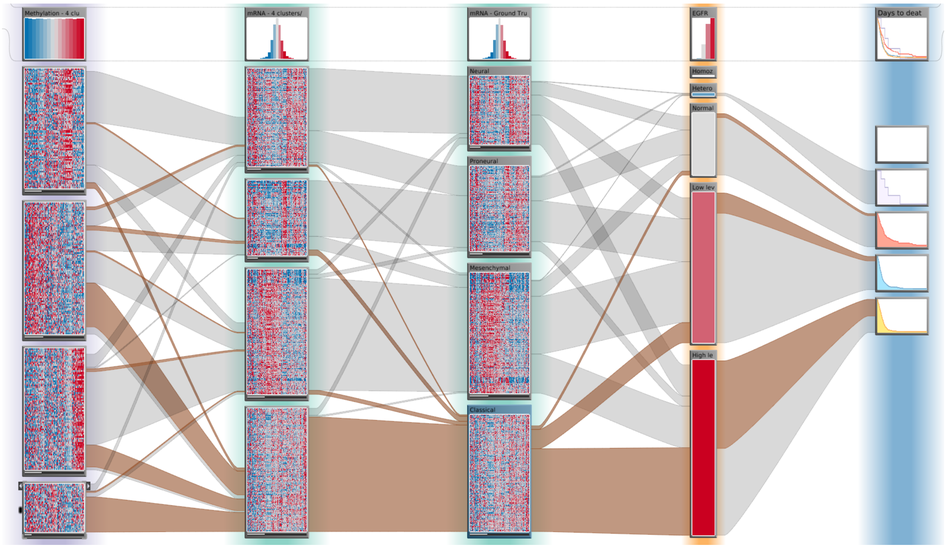
… makes data accessible

… combines strengths of
humans and computers

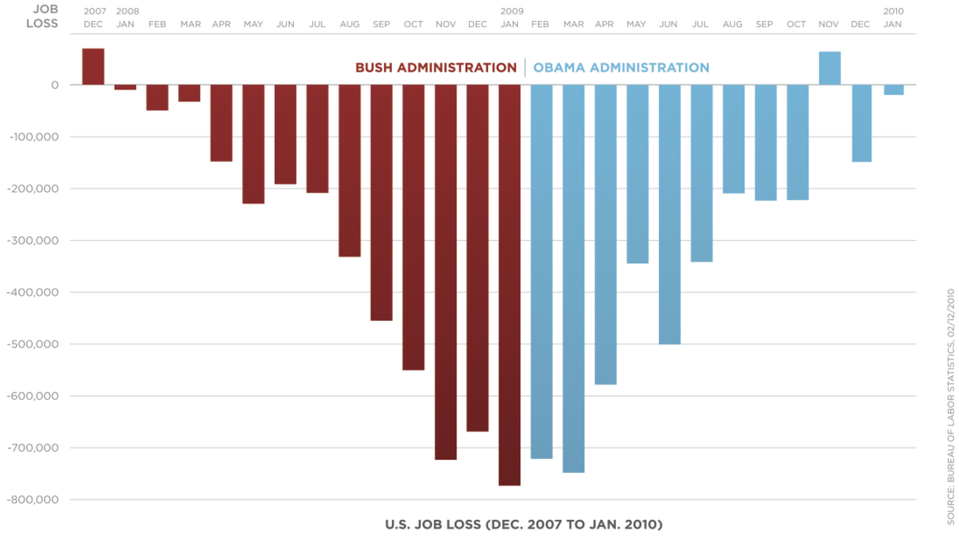
… enables insight

… communicates

Purpose of Visualization



[Obama Administration]



Open Exploration ⟷ **Communication**

Research Approach

Technique Driven

Solving a general visual analysis problem

Broad application

General questions

- ***How to visualize many sets?***
- ***How to make rankings interactive?***

Visualization Method

Problem Driven

Solving a domain problem of collaborator

Complex, targeted systems

Help answer important questions

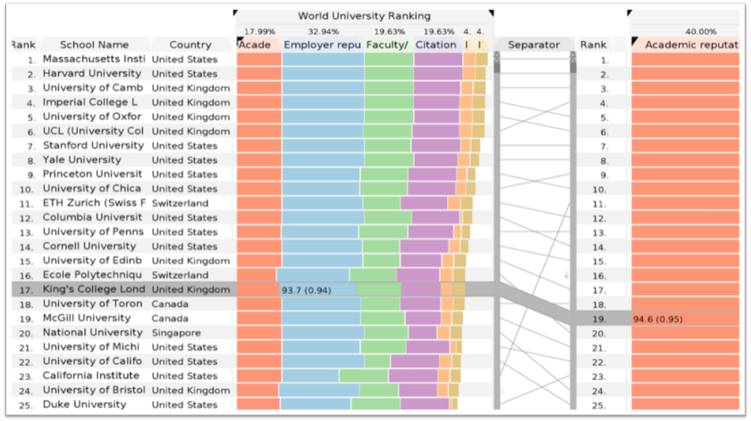
- ***What characterizes a cancer subtype?***
- ***Which drug works for these patients?***

Visualization Application

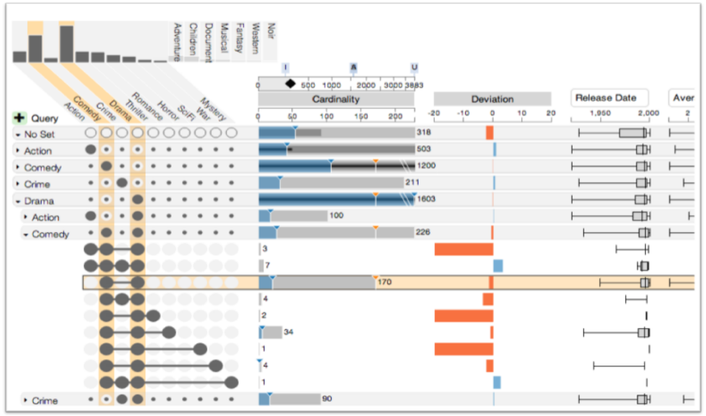
Content

Techniques

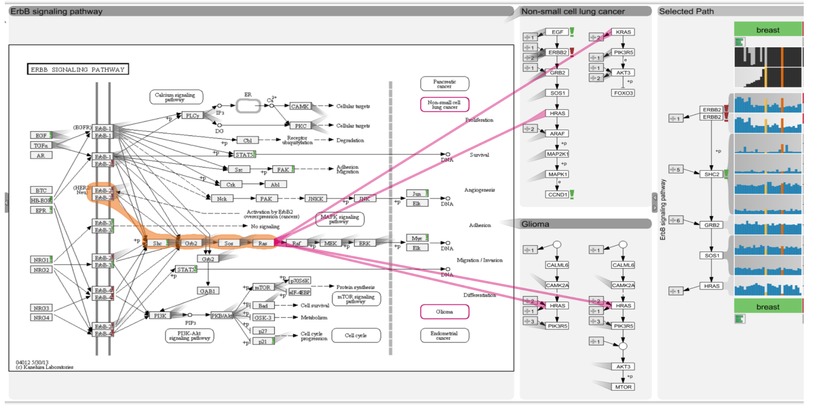
Domain Solution



Multi-Attribute Rankings



Set Visualization



Multivariate Networks

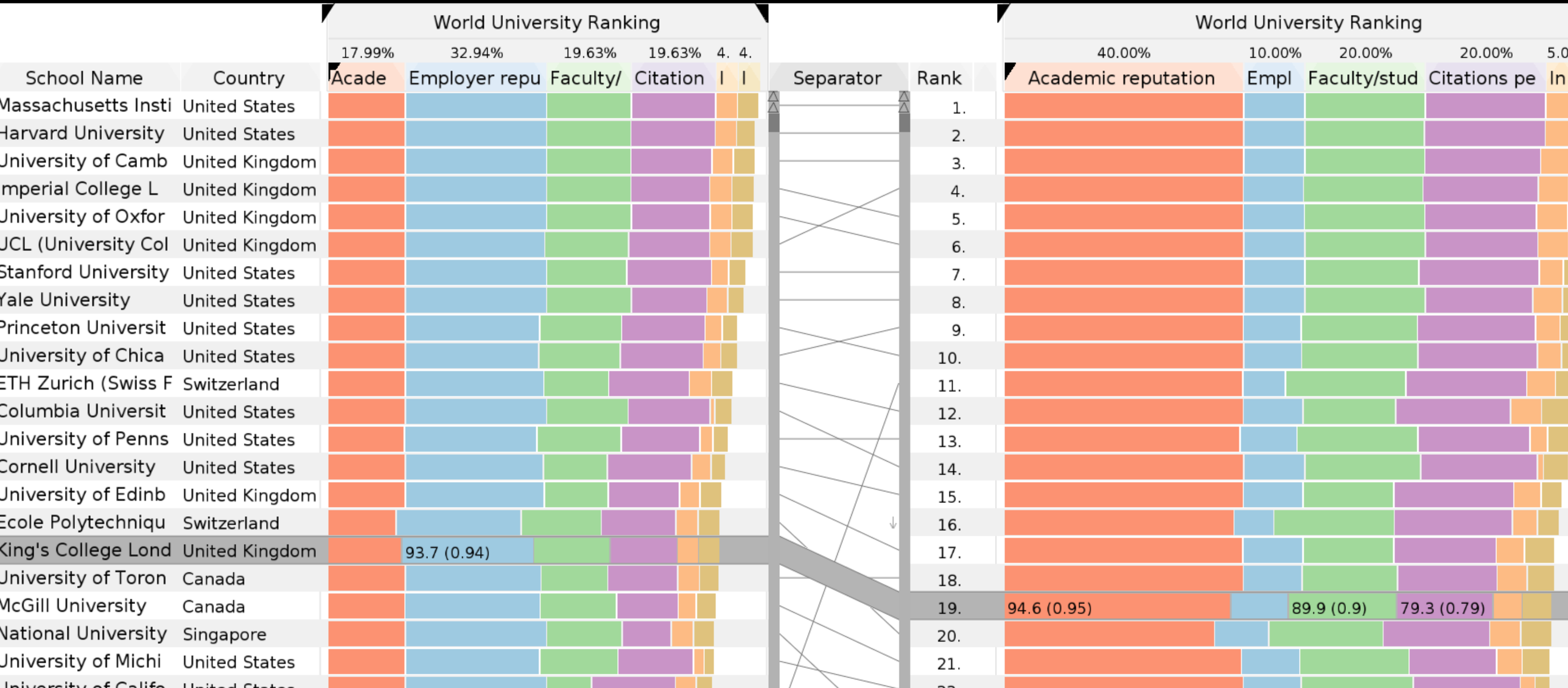
Pathway Analysis

Lay Users

Expert Analysts

[InfoVis ‘13]
Best Paper Award

Rankings




Rankings are popular

Goal

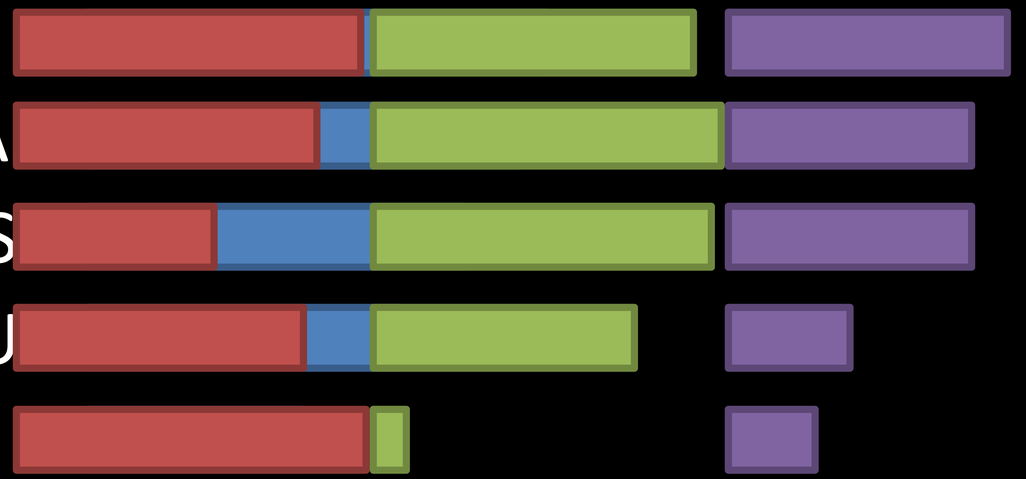
Intuitive
Interactive
Multi-Attribute
Ranking Visualization
To Create
Refine
Explore

Rank	University	Score
1.	MIT, USA	[redacted]
2.	Harvard, U	[redacted]
3.	Princeton,	[redacted]
4.	Cambridge,	[redacted]
5.	Oxford, UK	[redacted]4.0

Support Multiple Attributes

Score = f(A, B, C)

Rank	University	A	B	C
1.	MIT, USA			
2.	Harvard, USA			
3.	Princeton, US			
4.	Cambridge, U			
5.	Oxford, UK			



Combiner functions: f(A,B,C)

(Weighted) sum

$Score = w_a A + w_b B + w_c C$

Maximum

$Score = max(A, B, C)$

Product

Nesting

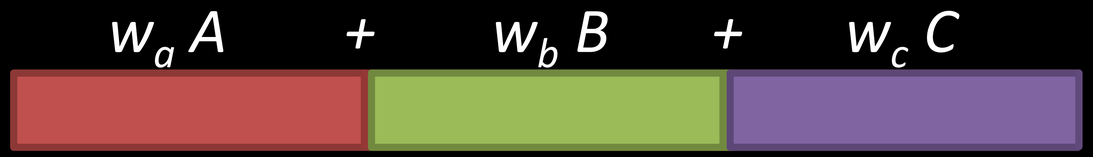
…

→ Serial

→ Parallel

→ Complex Combiners

Serial Combiner *(as Stacked Bar)*



Rank	University	A	B	C
1.	MIT, USA			
2.	Harvard, USA			
3.	Princeton, USA			
4.	Cambridge, UK			
5.	Oxford, UK			

Serial Combiner *(as Stacked Bar)*

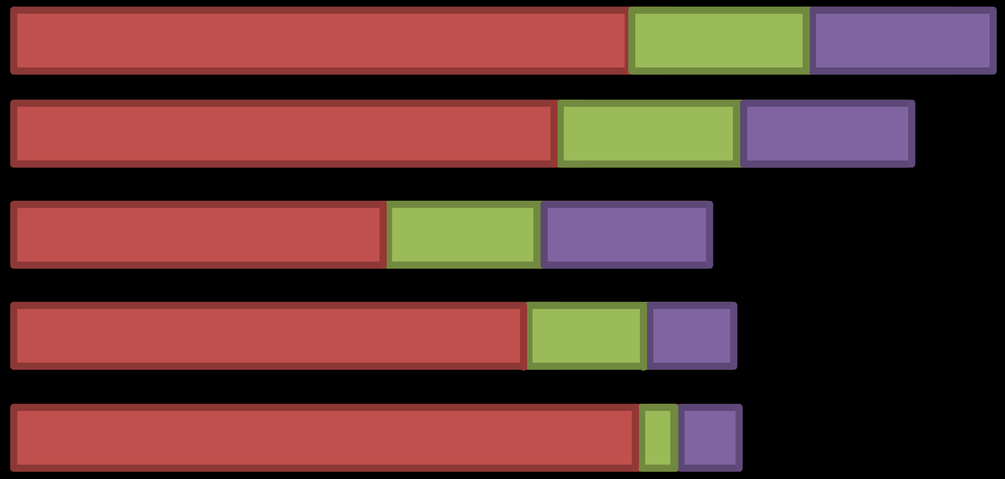
$w_a A + w_b B + w_c C$

Rank	University	A	B	C
1.	MIT, USA			
2.	Harvard, USA			
3.	Princeton, USA			
4.	Cambridge, UK			
5.	Oxford, UK			

Serial Combiner *(as Stacked Bar)*

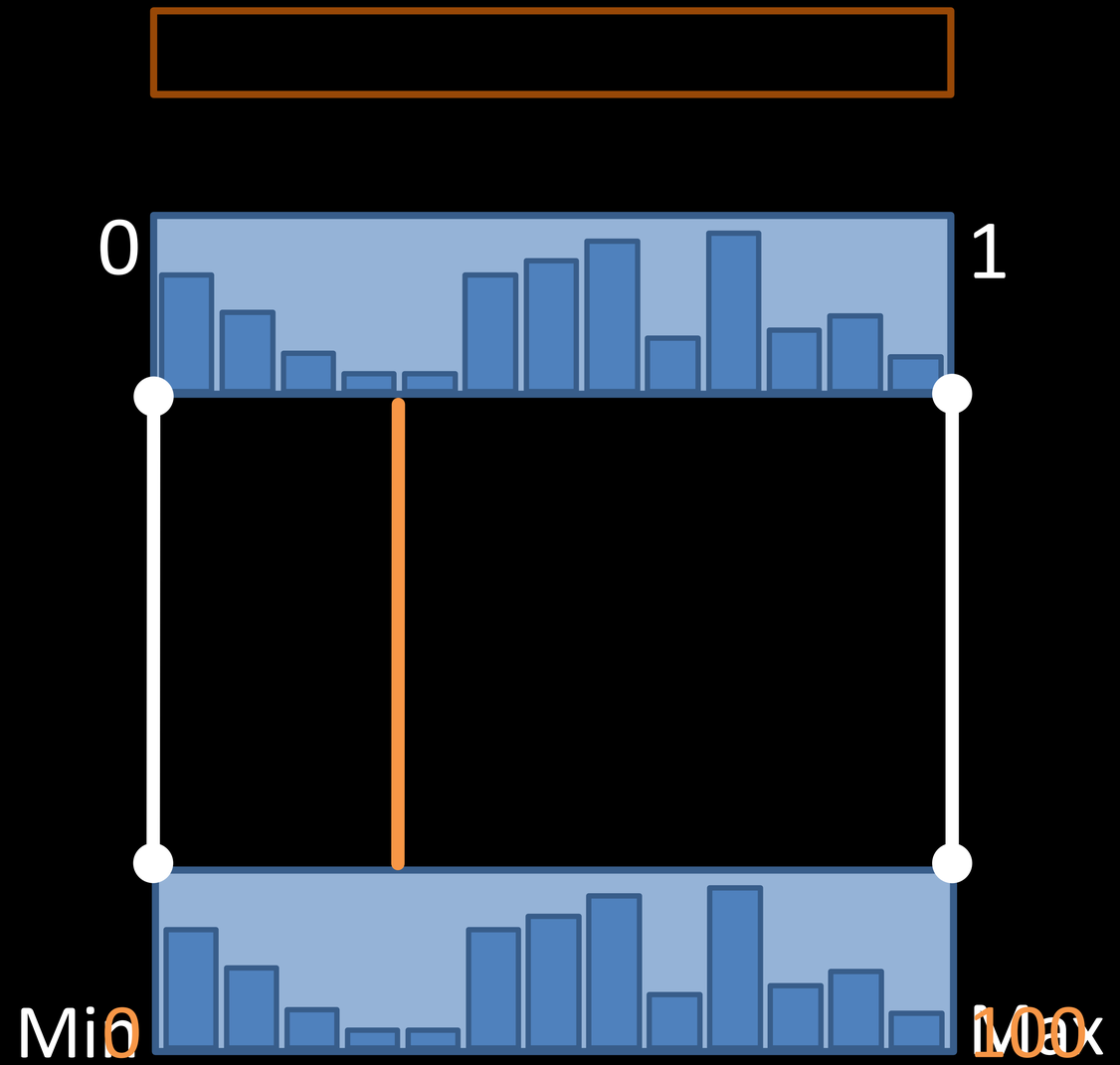
$w_a A + w_b B + w_c C$

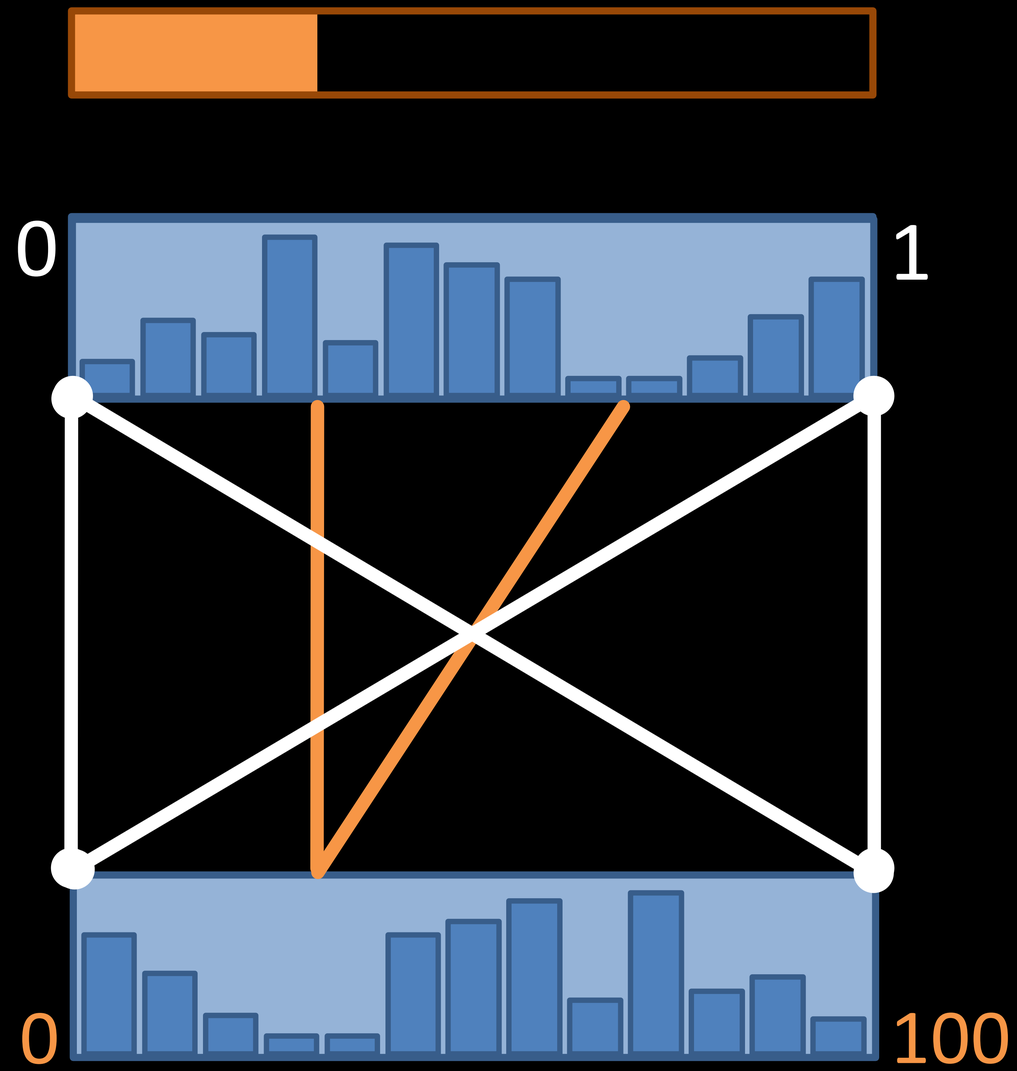
Rank	University	A	B	C
1.	MIT, USA			
2.	Harvard, USA			
3.	Princeton, USA			
4.	Cambridge, UK			
5.	Oxford, UK			

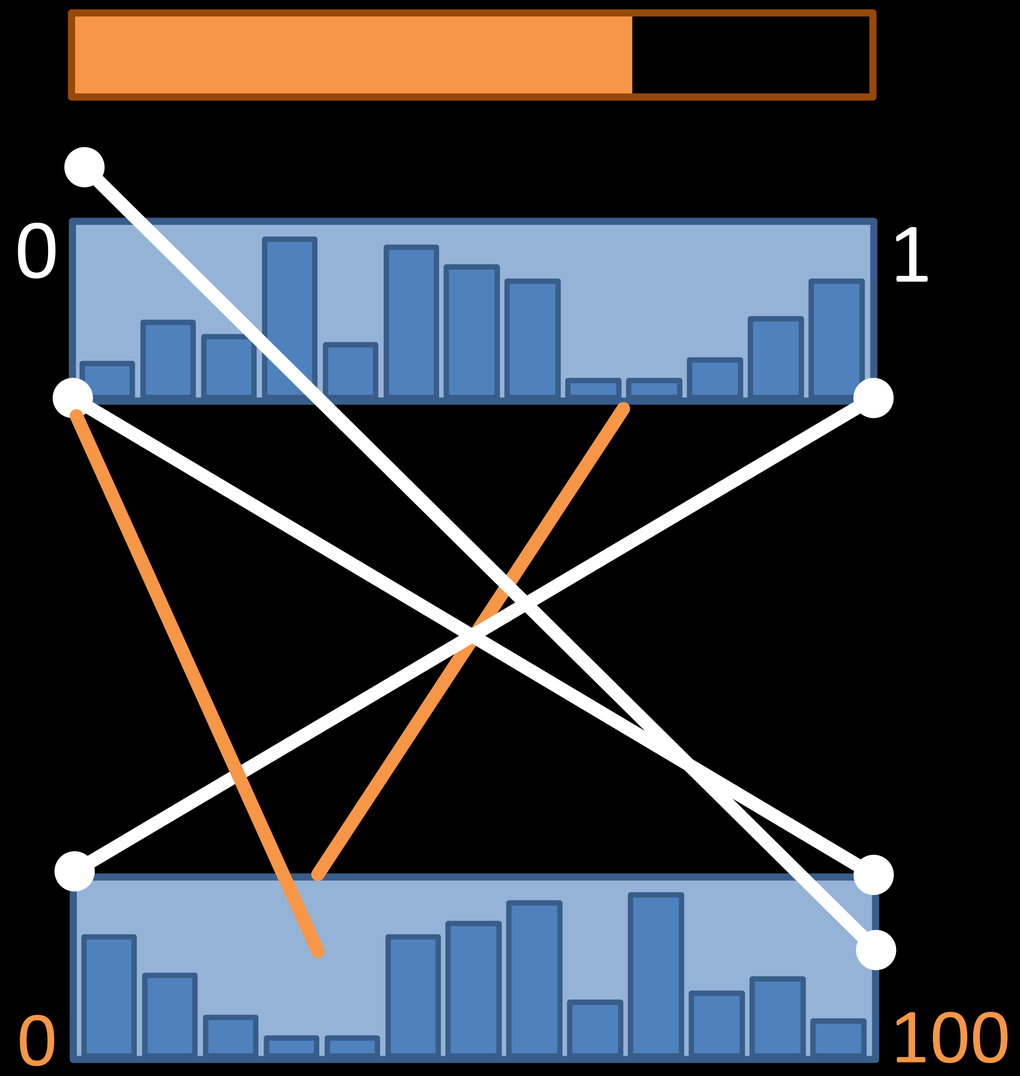



128 visible of 906 (14.13%)
Ran
School Name
Filter:
<None>
Country
Filter:
2 out of 72
Faculty/student ratio
Employer reputation
Citations per faculty
1. American University United States
2. Arizona State University United States
3. Aston University United Kingdom
4. Birkbeck College, University of L United Kingdom
5. Boston College United States
6. Boston University United States
7. Brandeis University United States
8. Brown University United States
9. Brunel University United Kingdom
10. California Institute of Technology United States
11. Cardiff University United Kingdom
12. Case Western Reserve University United States
13. City University London United Kingdom
14. College of William & Mary United States
15. Colorado State University United States
16. Columbia University United States
17. Cornell University United States
18. Cranfield University United Kingdom
19. Dartmouth College United States
20. Drexel University United States
21. Duke University United States
22. Durham University United Kingdom

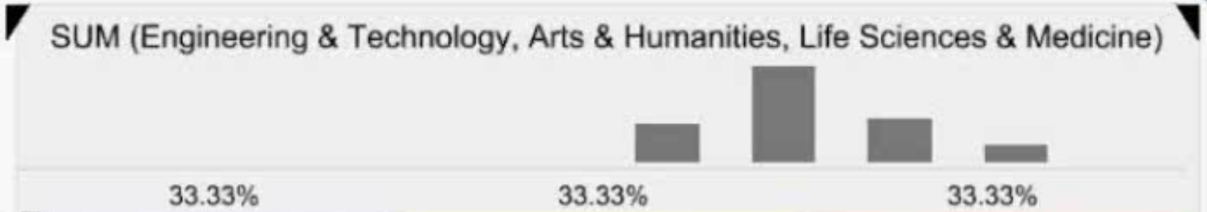
Flexible Mapping of Attributes to Scores


0
1
Min
0
Max
100


0
1
0
100


0
1
0
100

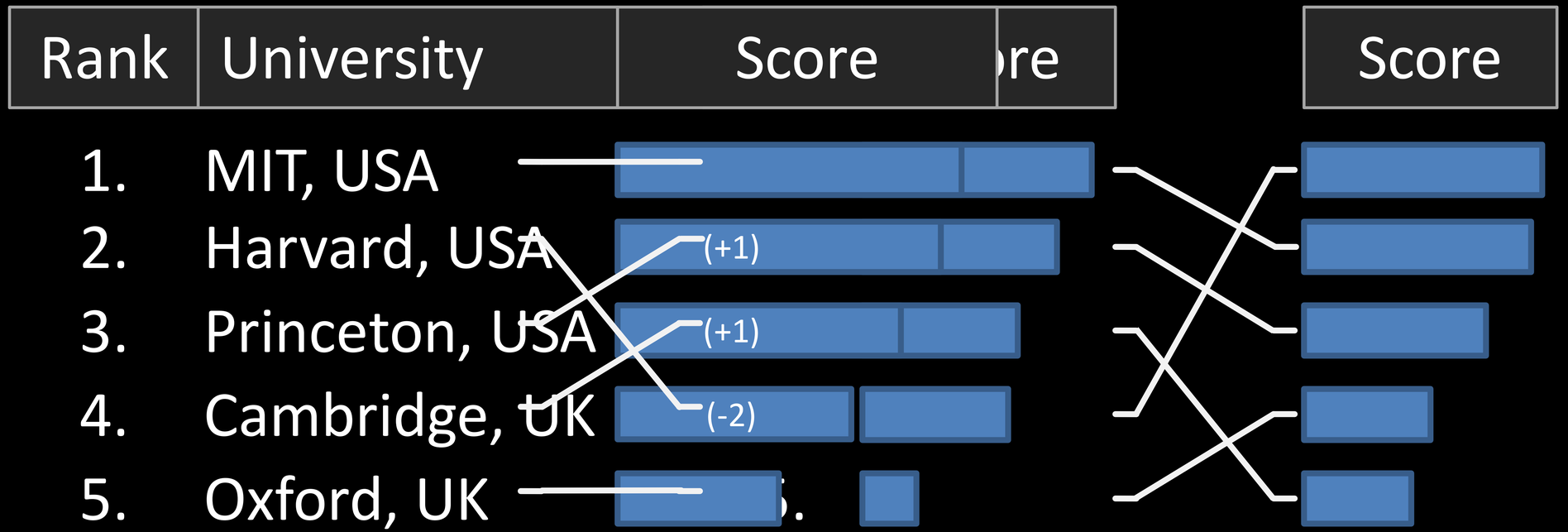
93 visible of 399 (23.31%)



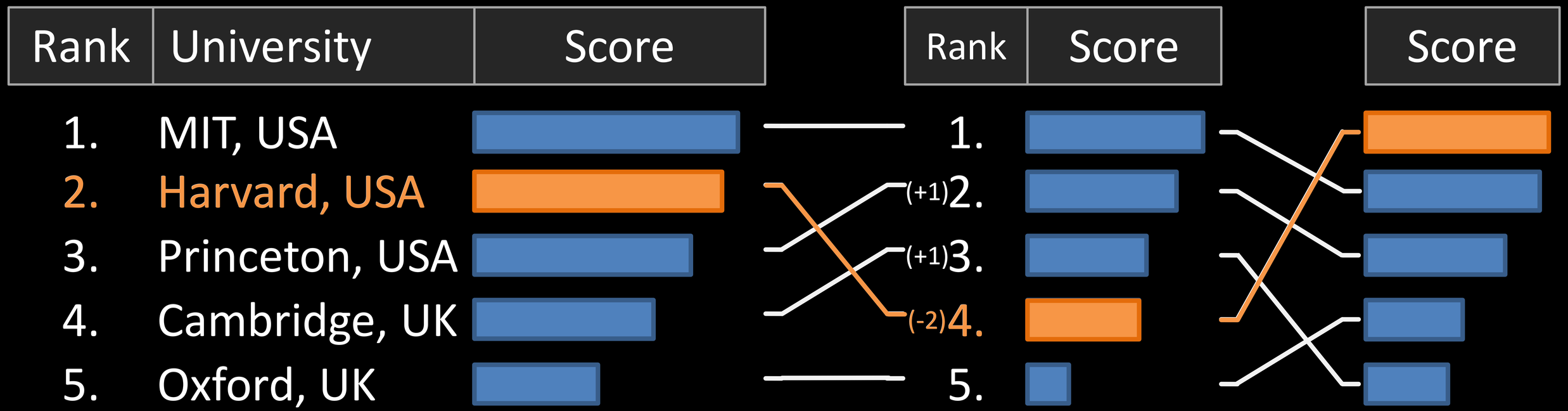
Ran	School Name	Country
	Filter: <None>	Filter: 2 out of 43
1.	University of Oxford	United Kingdom
2.	University of Cambridge	United Kingdom
3.	Harvard University	United States
4.	Stanford University	United States
5.	Massachusetts Institute of Technology (MIT	United States
6.	University of California, Berkeley (UCB)	United States
7.	University of California, Los Angeles (UCL	United States
8.	Yale University	United States
9.	UCL (University College London)	United Kingdom
10.	Columbia University	United States
11.	Princeton University	United States
12.	University of Edinburgh	United Kingdom
13.	University of Michigan	United States
14.	Cornell University	United States
15.	University of Pennsylvania	United States
16.	The University of Manchester	United Kingdom
17.	Imperial College London	United Kingdom
18.	University of Chicago	United States

Compare Rankings

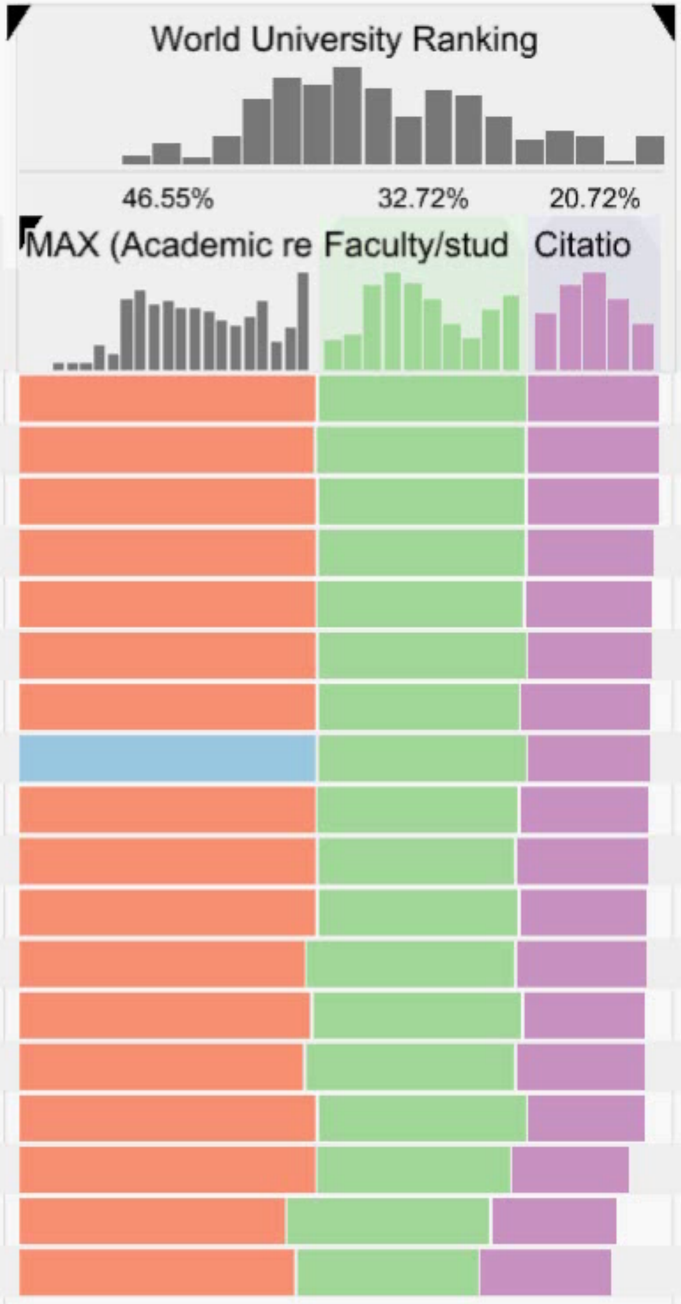
Bump Charts


Rank
University
Score
ore
Score
1. MIT, USA
2. Harvard, USA
3. Princeton, USA
4. Cambridge, UK
5. Oxford, UK
(+1)
(+1)
(-2)

Bump Charts


Rank
University
Score
1. MIT, USA
2. Harvard, USA
3. Princeton, USA
4. Cambridge, UK
5. Oxford, UK
Rank
Score
1.
(+1)2.
(+1)3.
(-2)4.
5.
Score

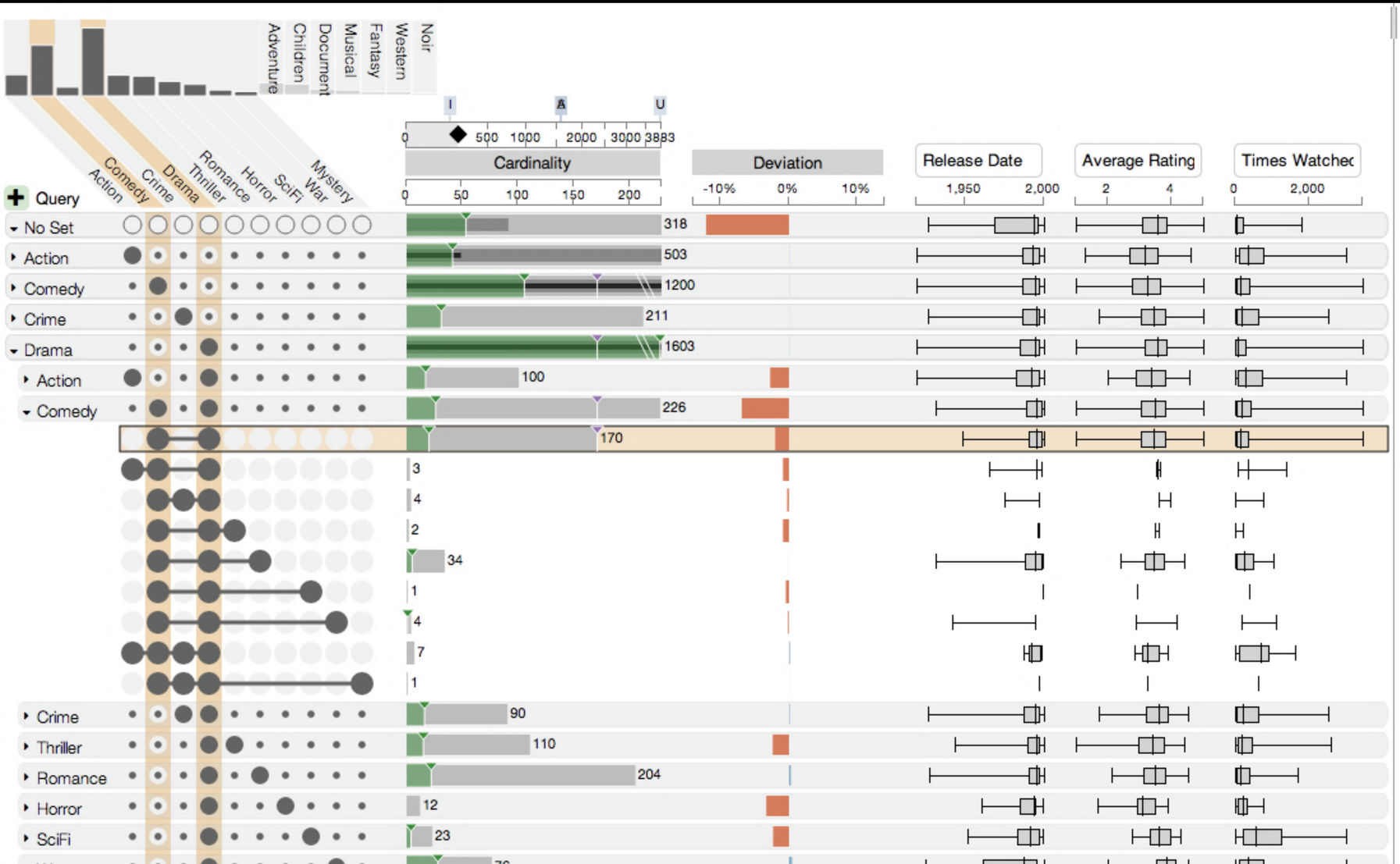
453 visible of 453 (100.00%)

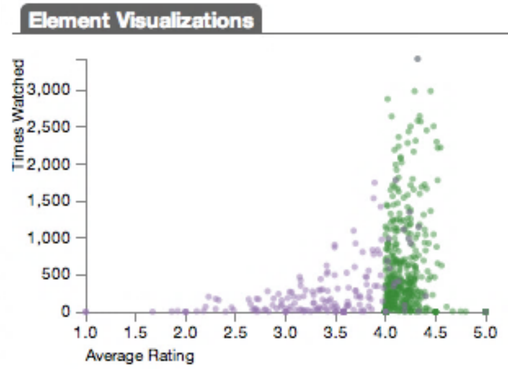


Rank	School Name
	Filter: <None>
1.	Massachusetts Institute of Te
2.	California Institute of Technol
3.	Harvard University
4.	University of Cambridge
5.	UCL (University College Lond
6.	University of Oxford
7.	Princeton University
8.	Imperial College London
9.	University of Chicago
10.	Stanford University
11.	Columbia University
12.	Duke University
13.	University of Pennsylvania
14.	Johns Hopkins University
15.	Yale University
16.	University of Michigan
17.	Ecole normale supérieure, Pa
18.	Northwestern University

http://lineup.caleydo.org
http://caleydo.github.io/lineup.js/

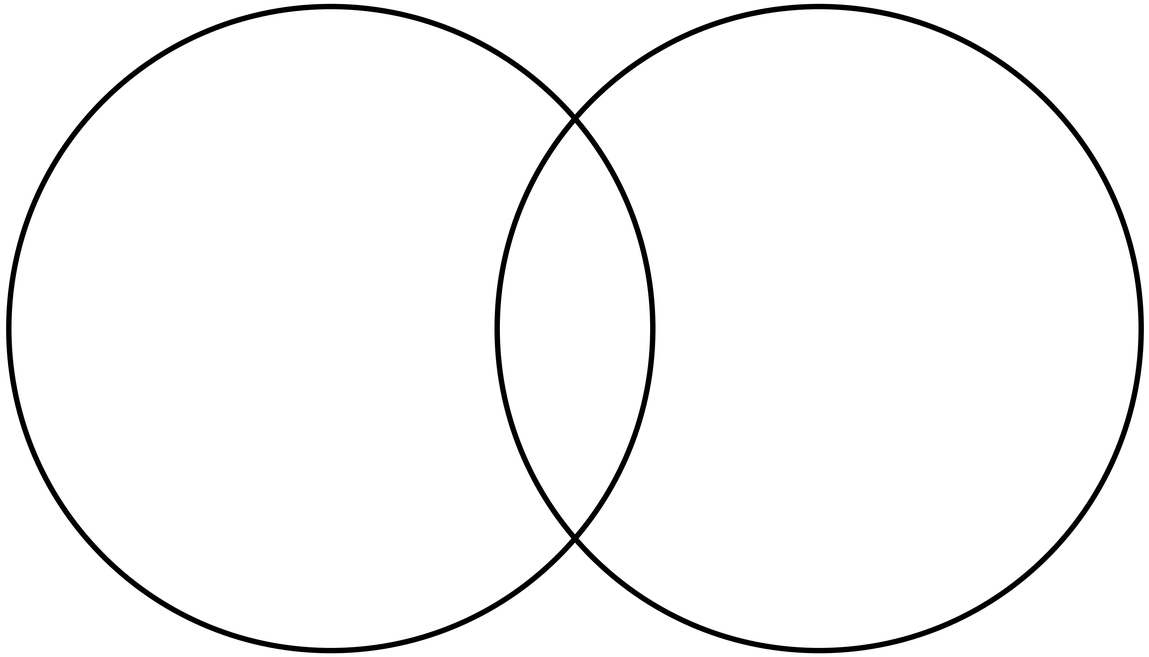
[InfoVis'14] UpSet
Visualizing Intersecting Sets



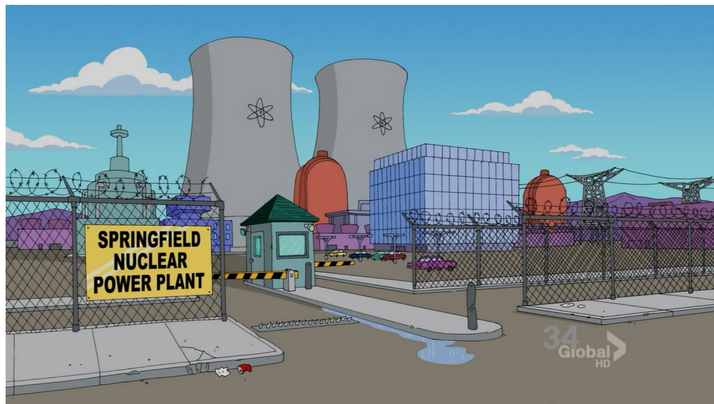


Name	Release Date	Average Rating	Times Watched	Set Count
Toy Story (1995)	1995	4.15	2077	2
Sense and Sensibility (1995)	1995	4.03	835	2
Persuasion (1995)	1995	4.06	179	1
City of Lost Children, The (1995)	1995	4.06	403	2

School



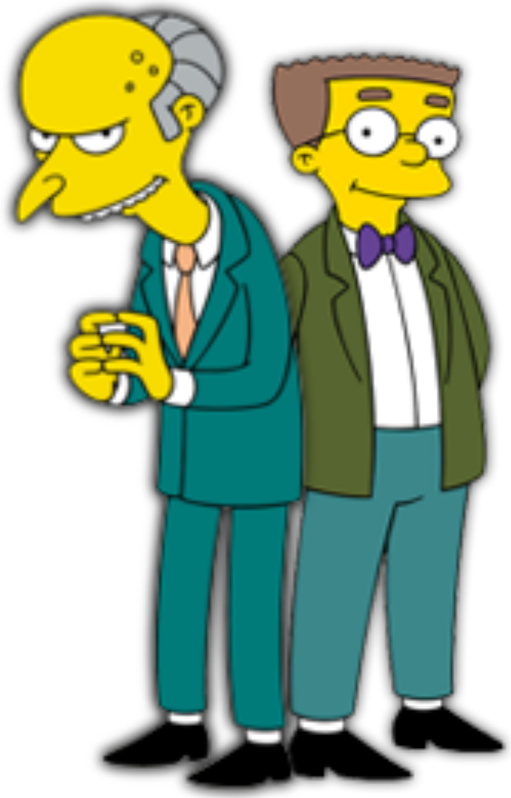
School **Power Plant**

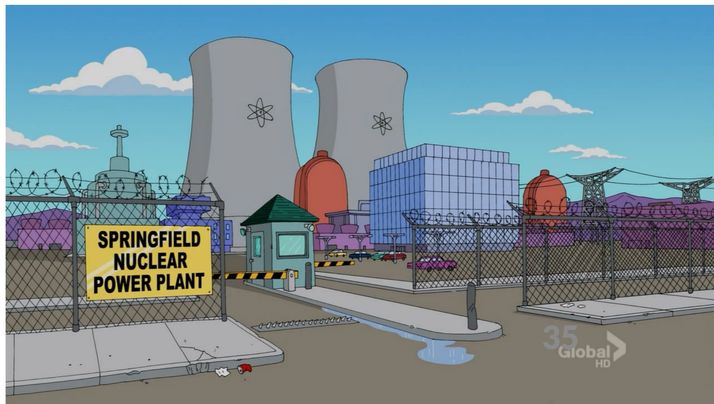


Evil

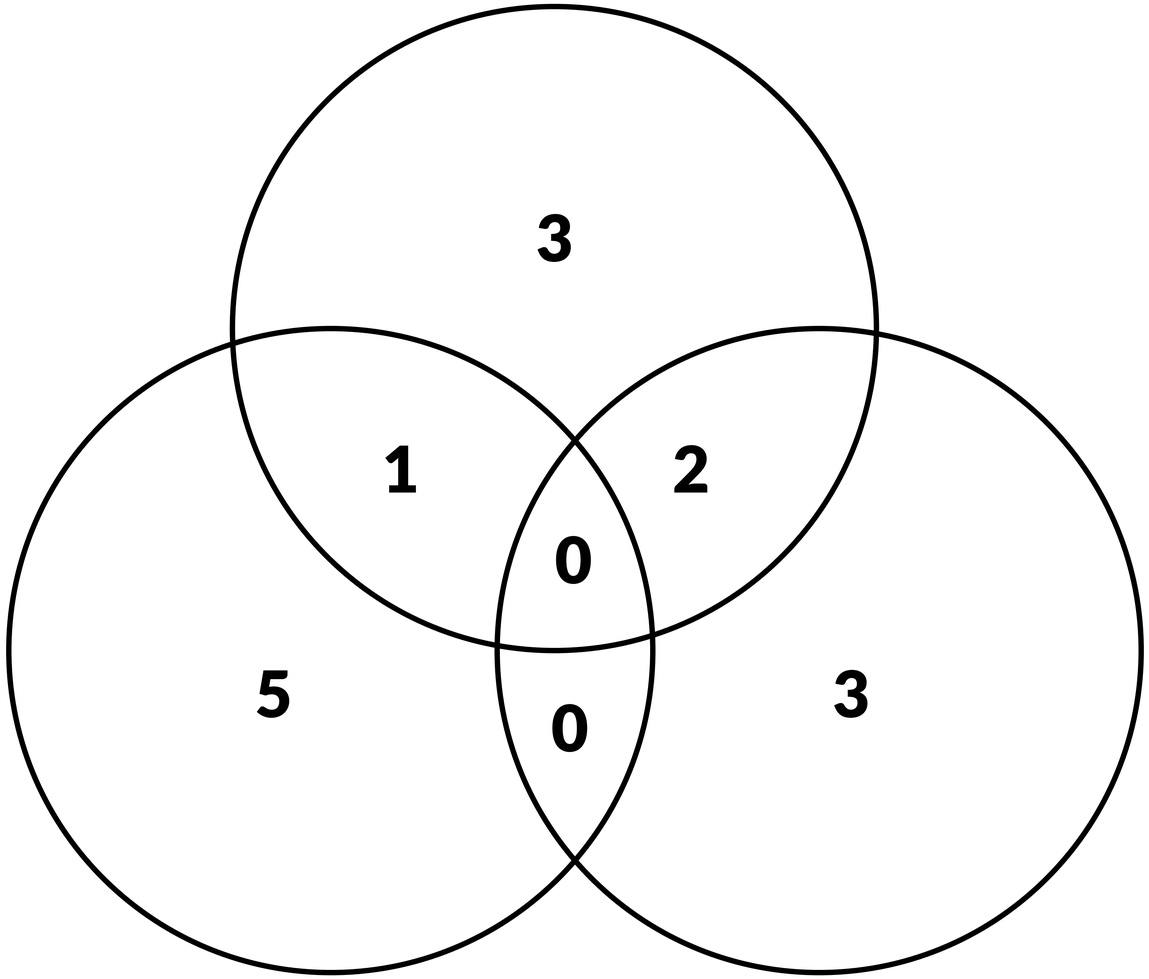
School

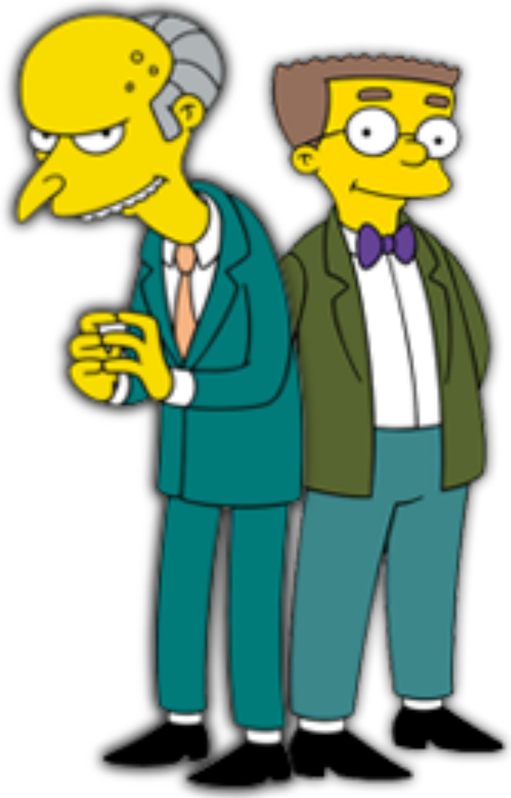
Power Plant



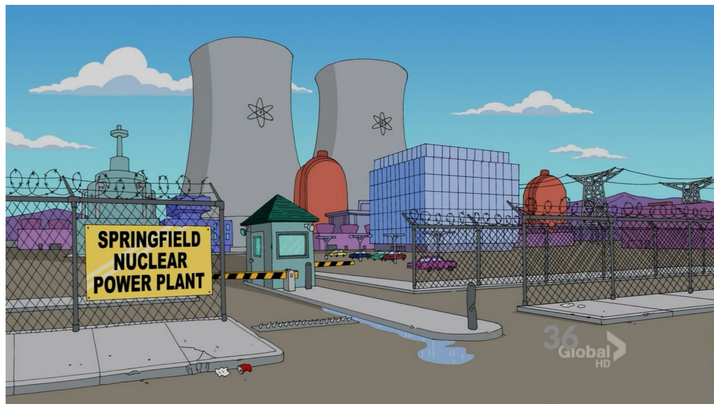



MATT
GROENING


Evil
3
1
2
0
5
0
3
School
Power Plant

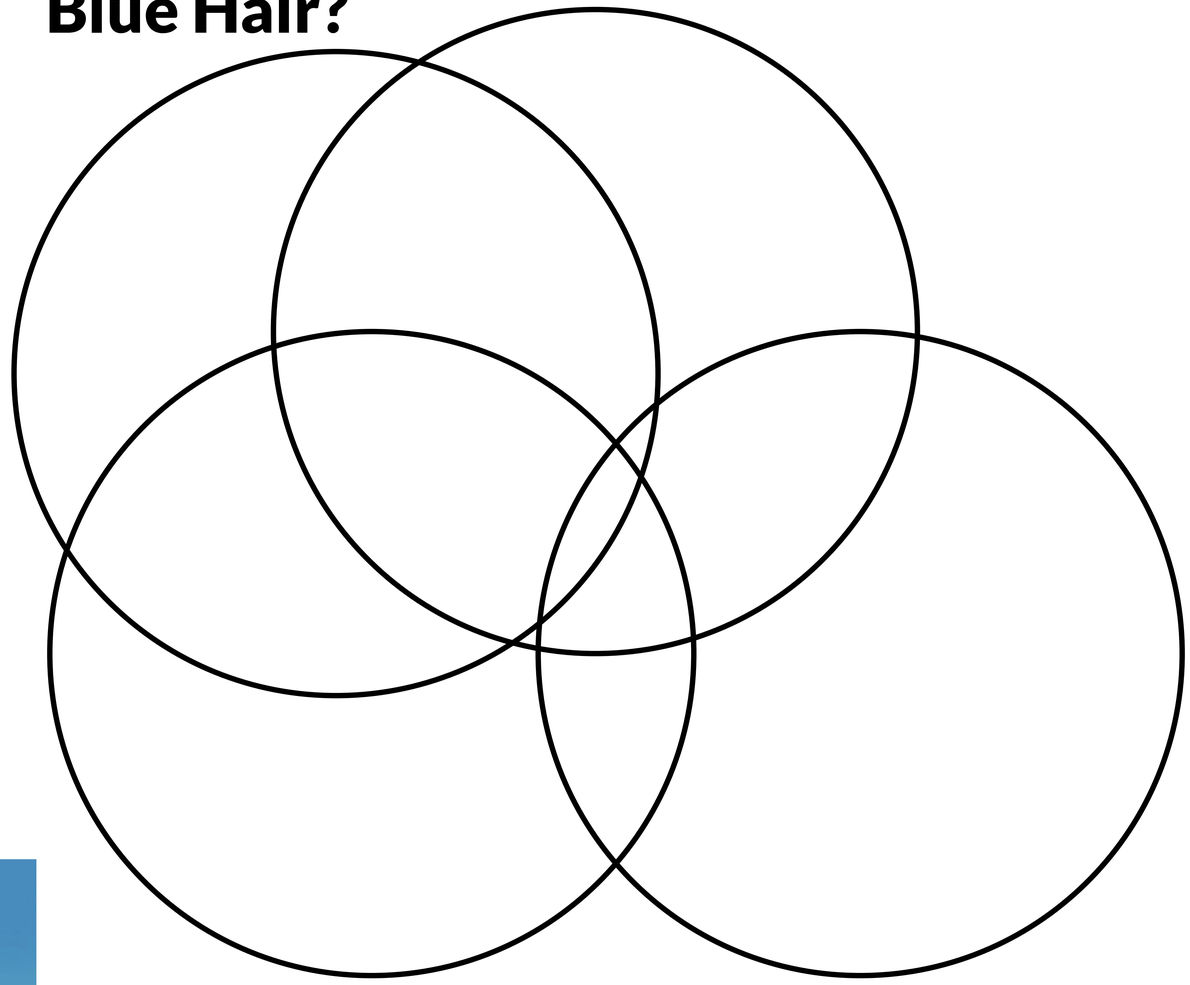


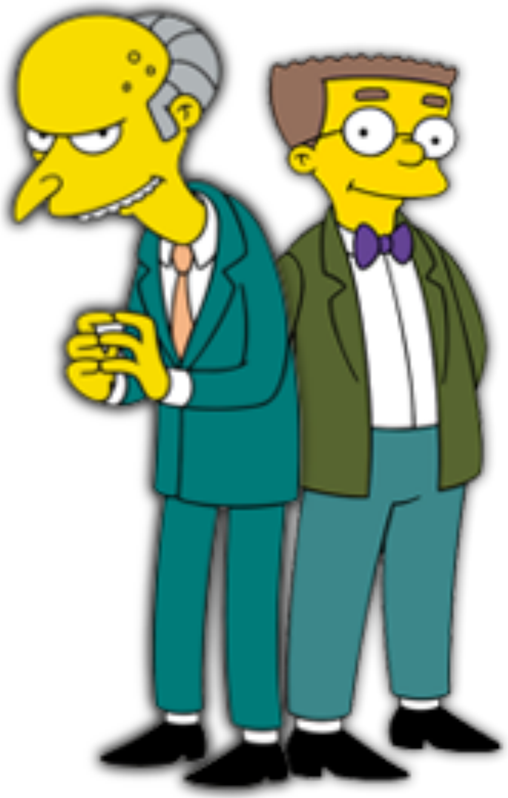

SPRINGFIELD
ELEMENTARY
SCHOOL


SPRINGFIELD
NUCLEAR
POWER PLANT

Evil

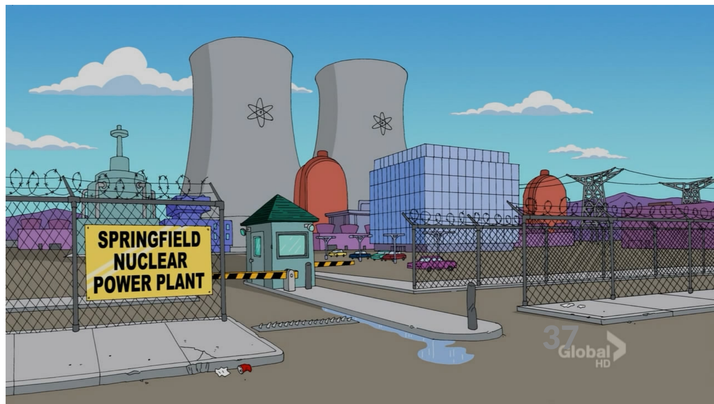
Blue Hair?





School

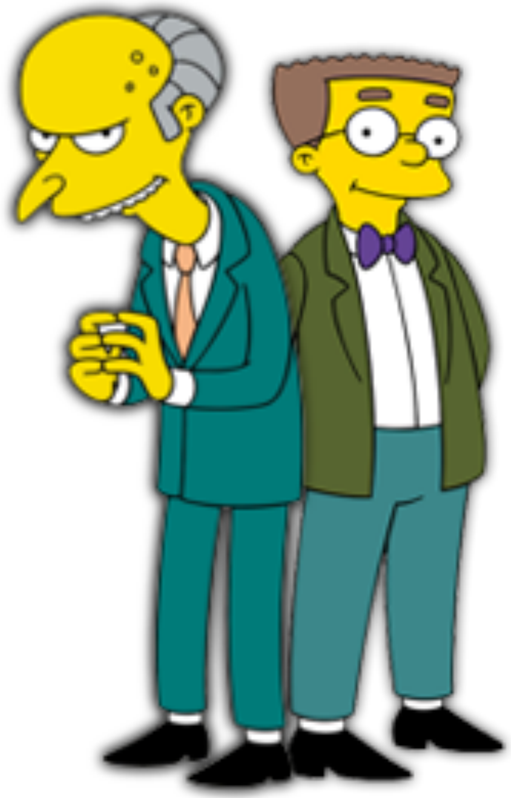
Power Plant



Blue Hair?

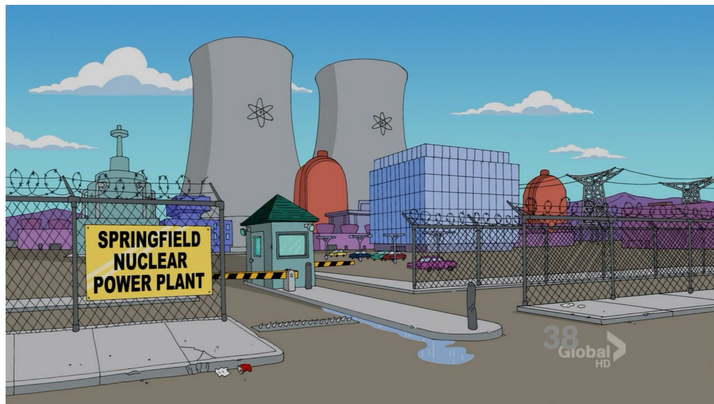
Evil

Duff Beer?

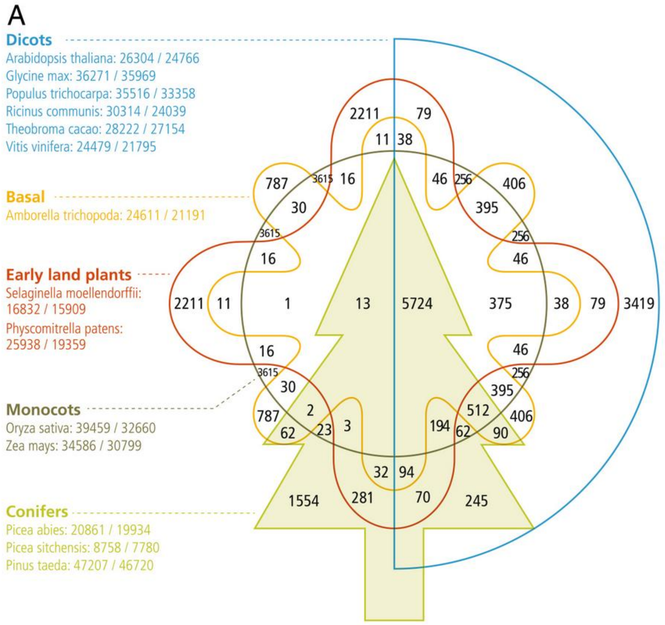


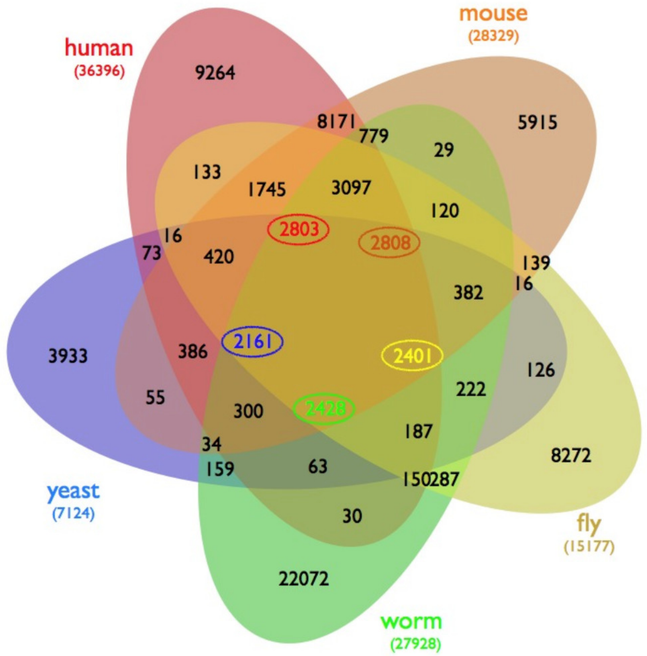
School

Power Plant

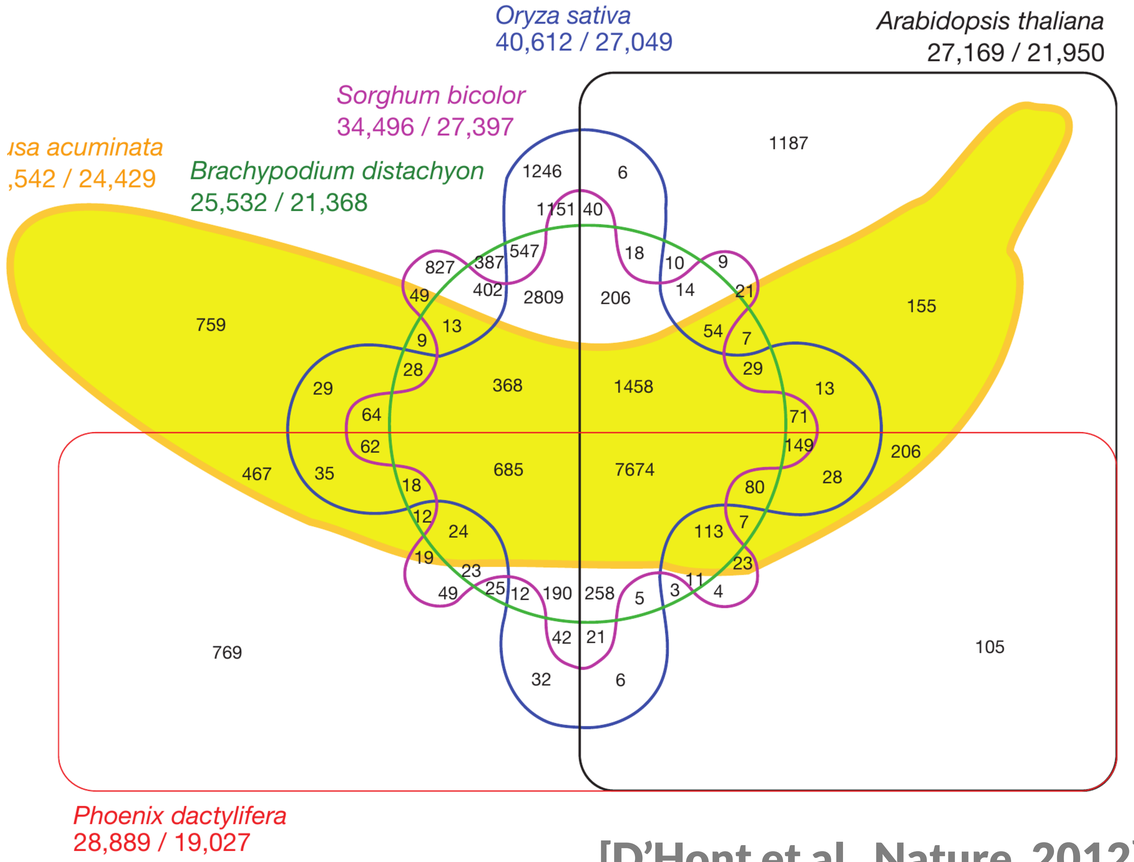


[Neale et al., BMC Genome Biology, 2014]

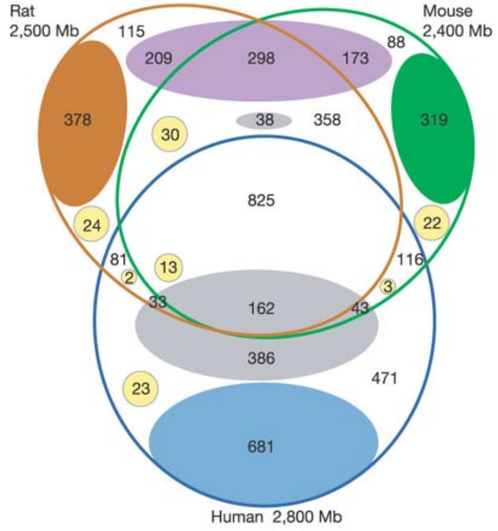


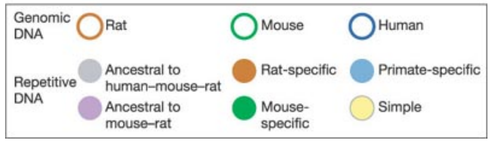


[Wiles et al., BMC Systems Biology]

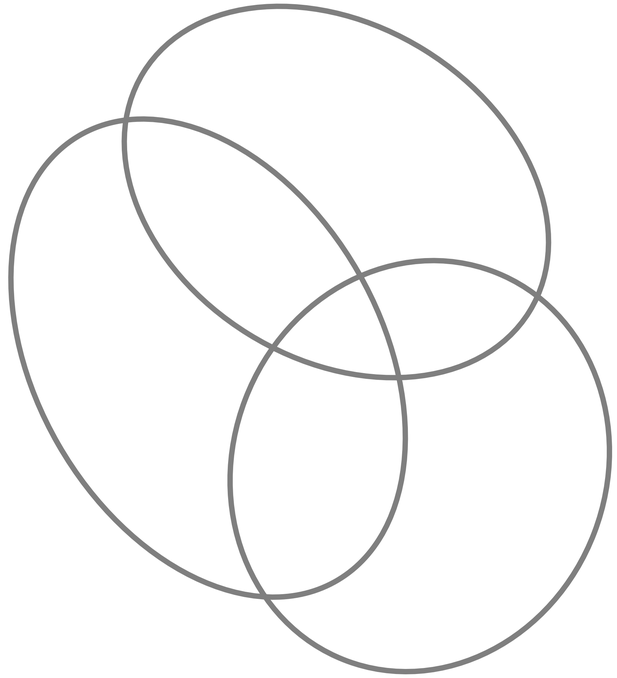


[D'Hont et al., Nature, 2012]



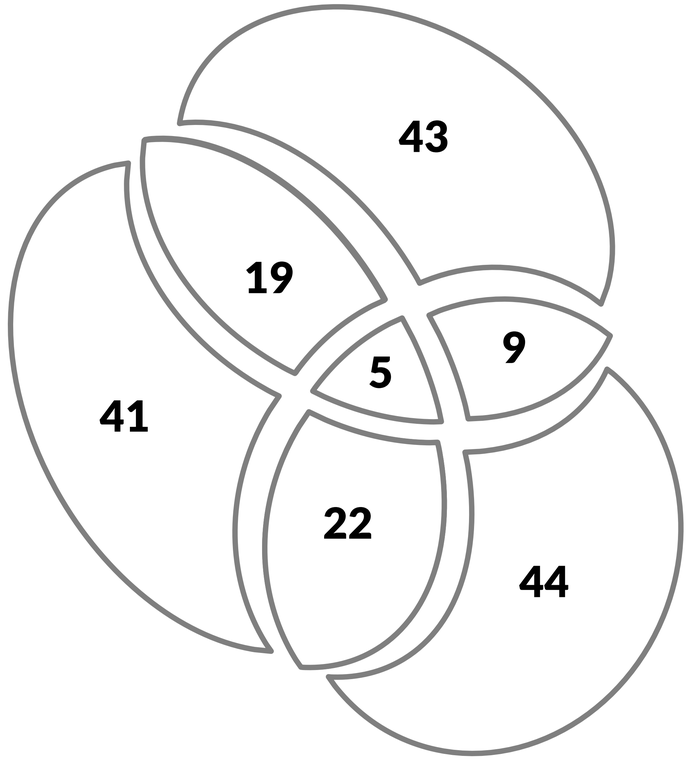


[Gibbs et al., Nature, 2004]



[created with EulerAPE]

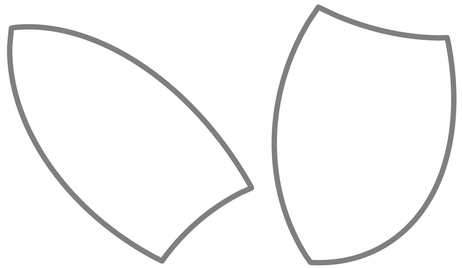
22 ? 19



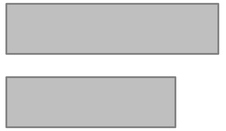
[created with EulerAPE]

Set Vis Goals

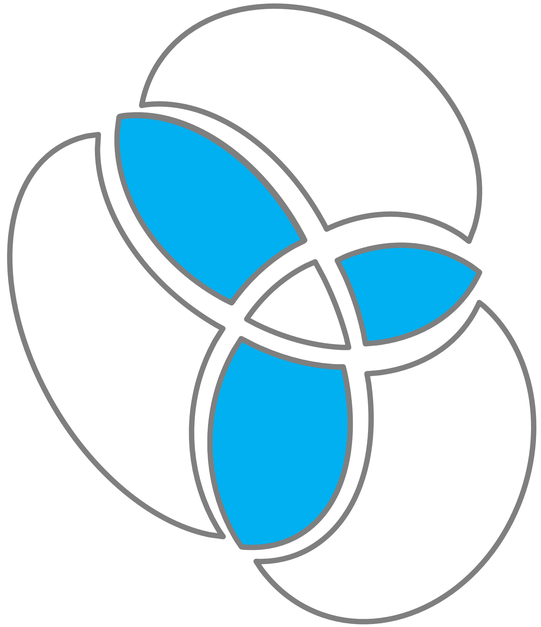
1. Efficient visual encoding



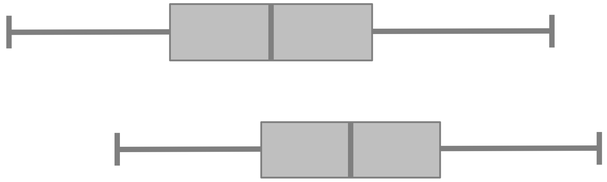
vs.



2. Creating complex slices of a dataset



3. Visualize attributes



[Movie Lens Dataset]

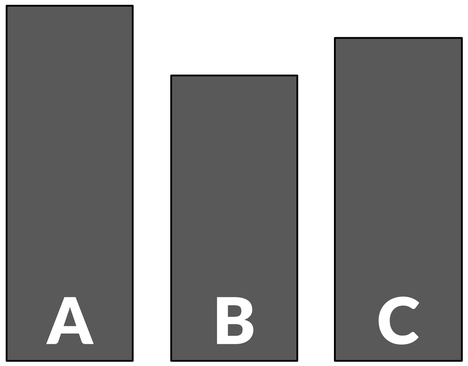
Attribute Details

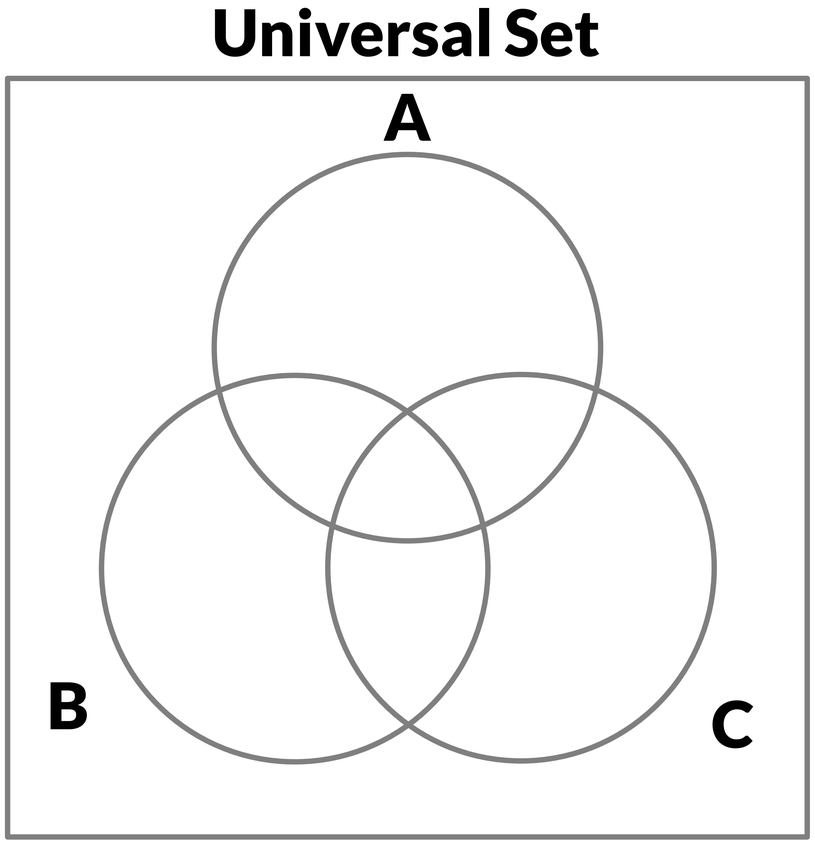
Visualizing Intersections

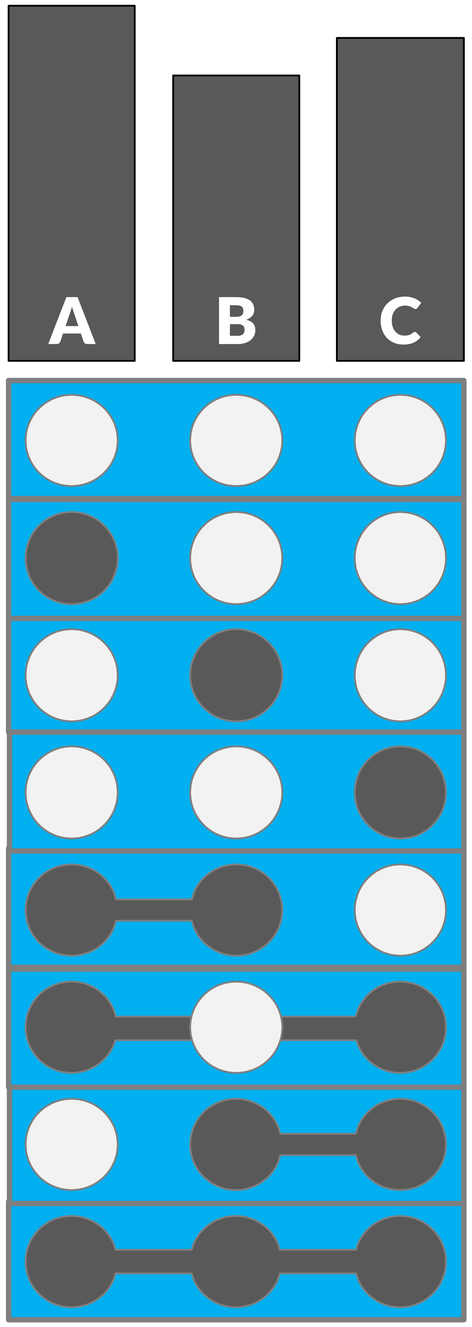
Visualizing Properties

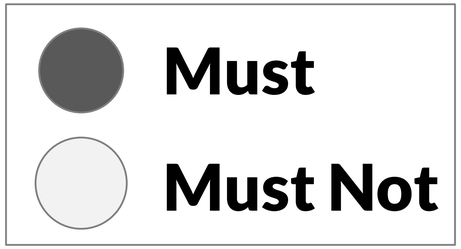
Element List & Queries

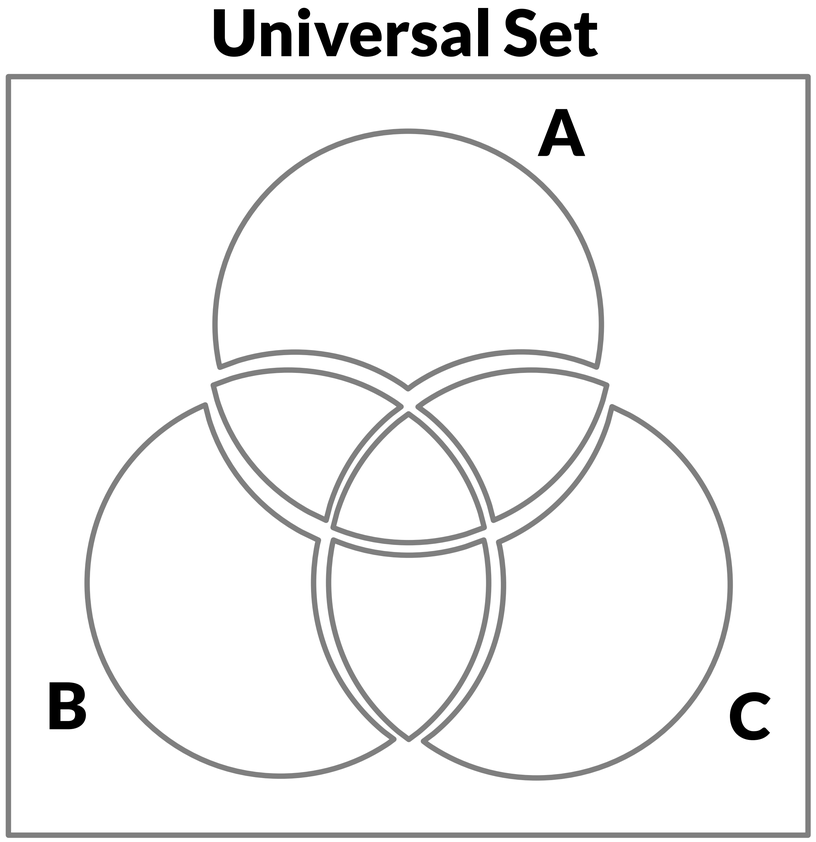
Visualizing Intersections

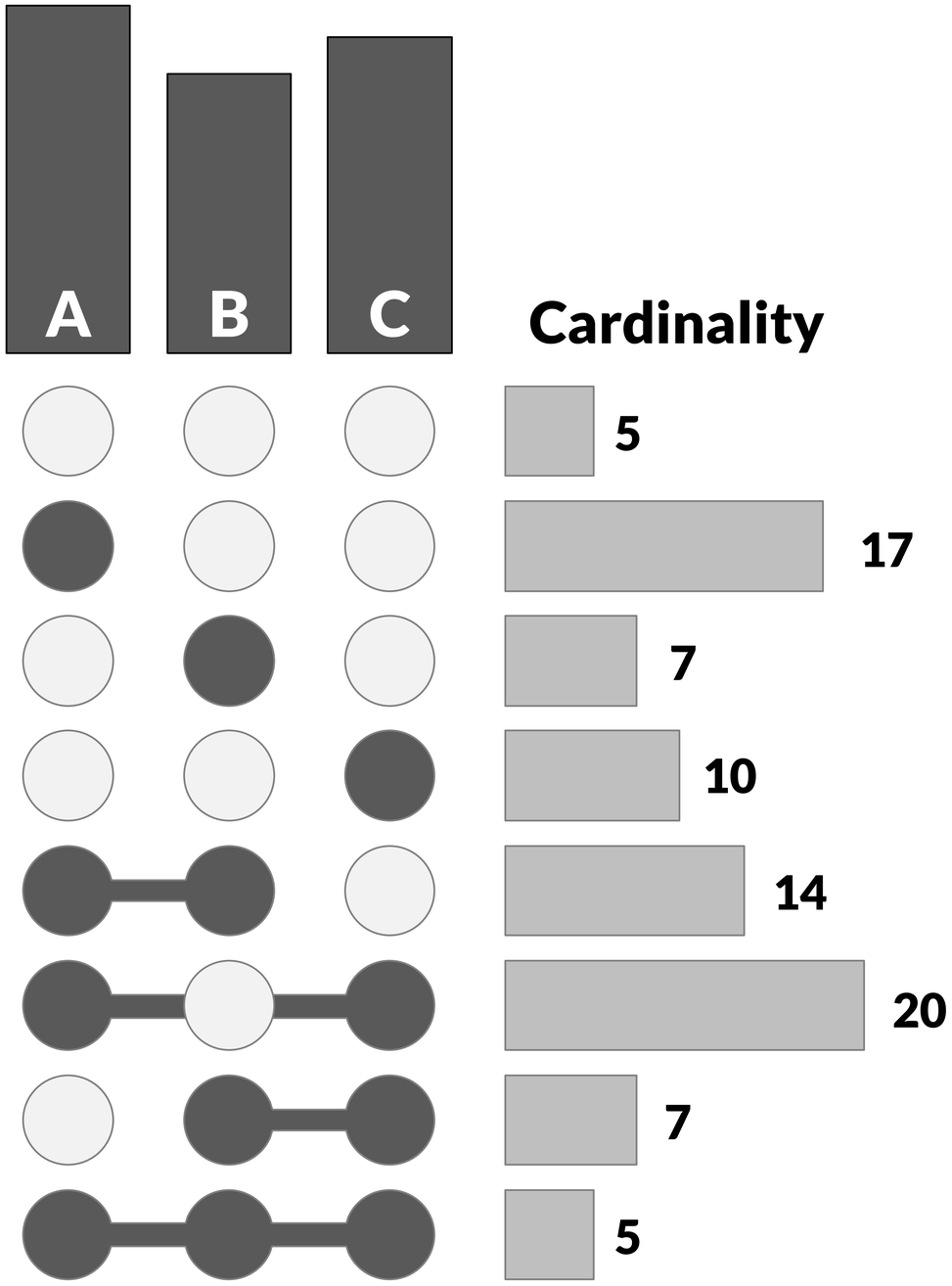

A
B
C

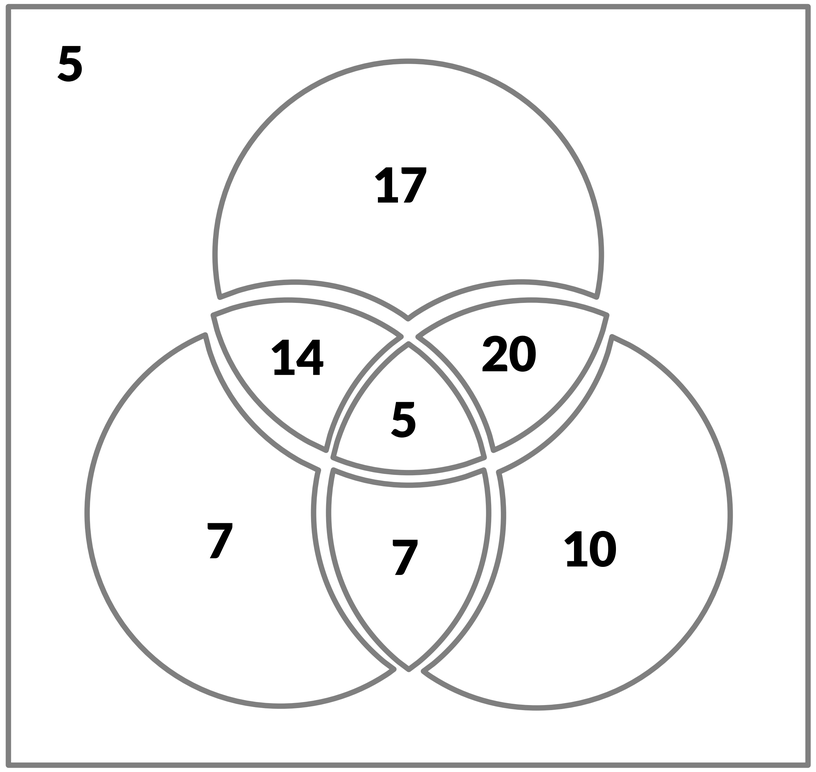

Universal Set
A
B
C


A
B
C

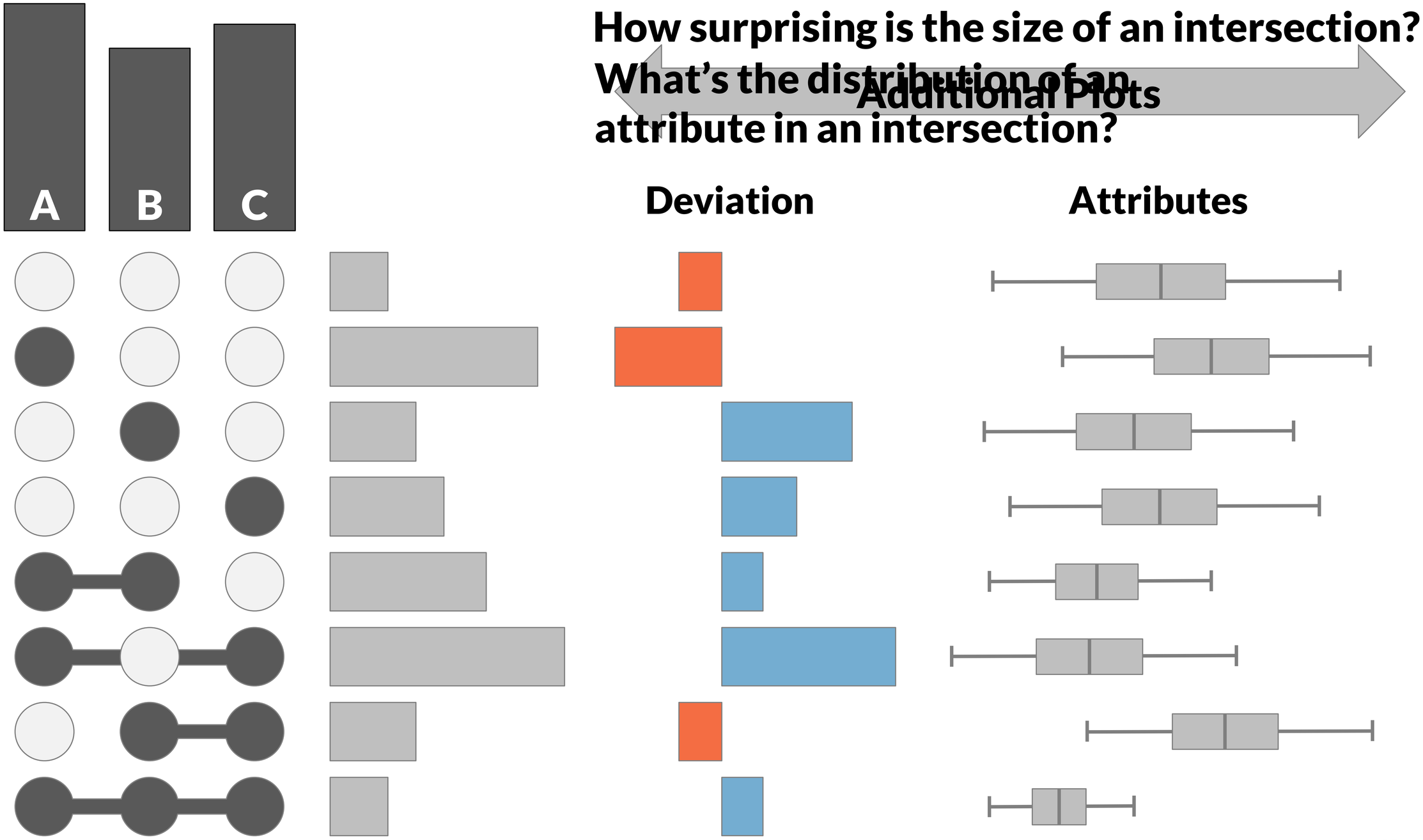

Must
Must Not

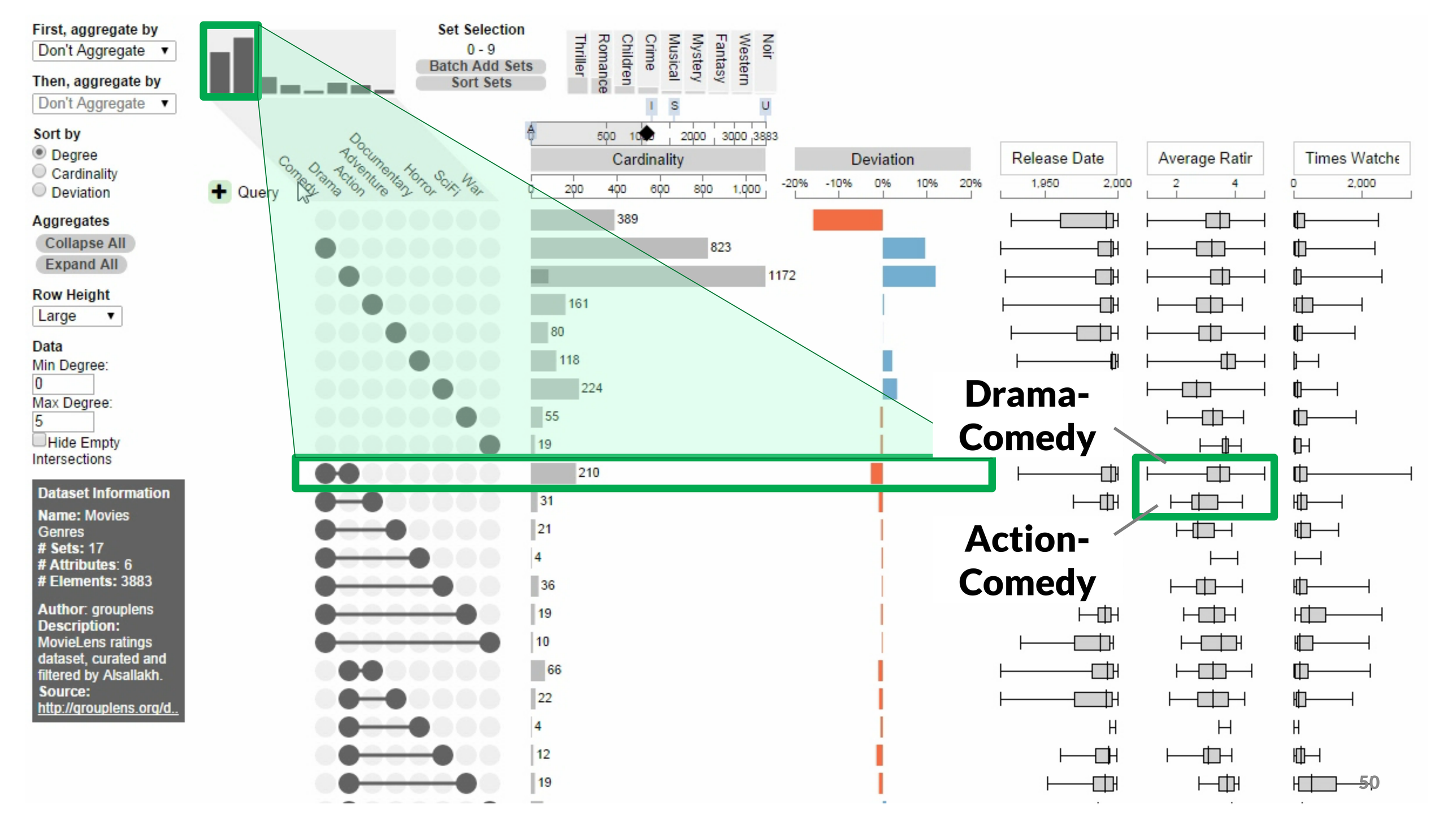

Universal Set
A
B
C


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

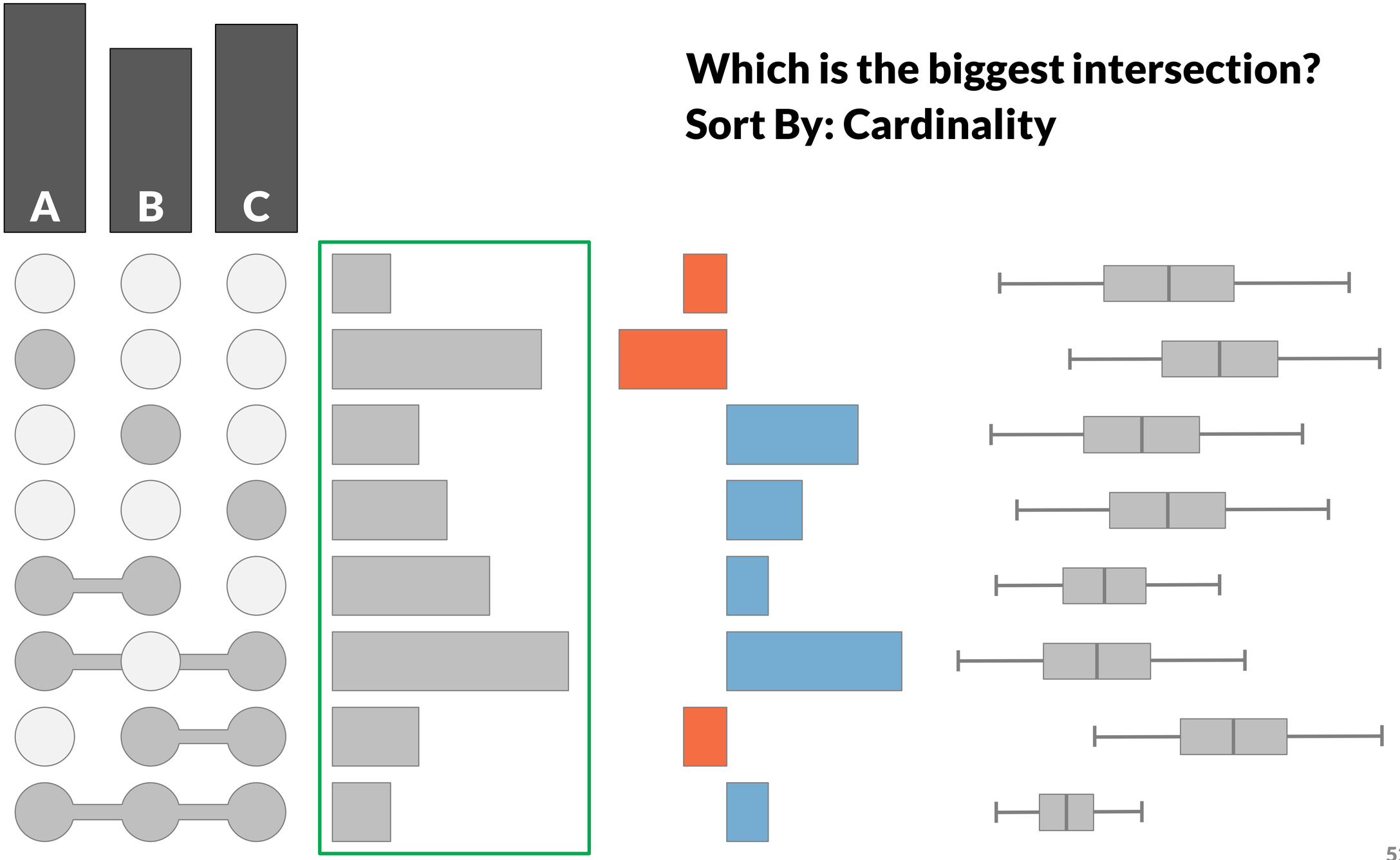

5
17
14
20
5
7
7
10

Plotting Attributes

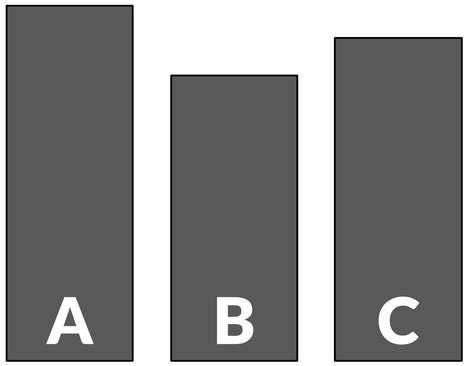

How surprising is the size of an intersection?
What's the distribution of an attribute in an intersection?
Additional Plots
A
B
C
Deviation
Attributes


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
Comedy
Drama
Action
Adventure
Documentary
Horror
SciFi
War
Query
Cardinality
Deviation
Release Date
Average Ratir
Times Watche
Drama-Comedy
Action-Comedy

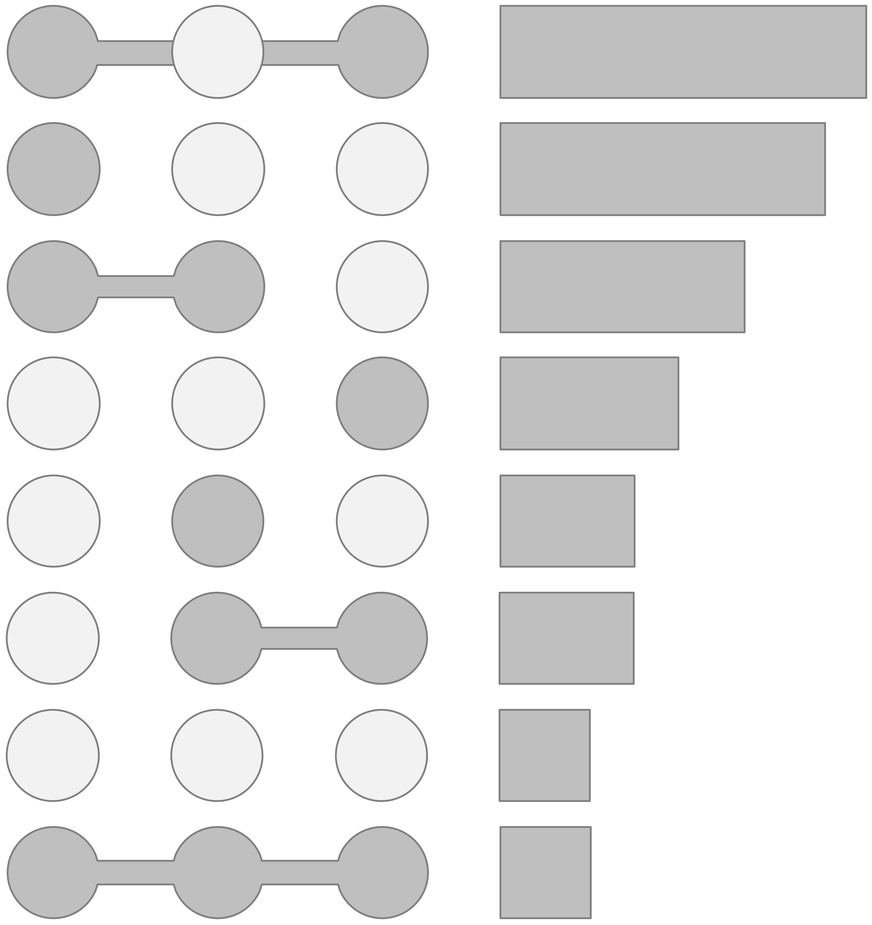
Sorting

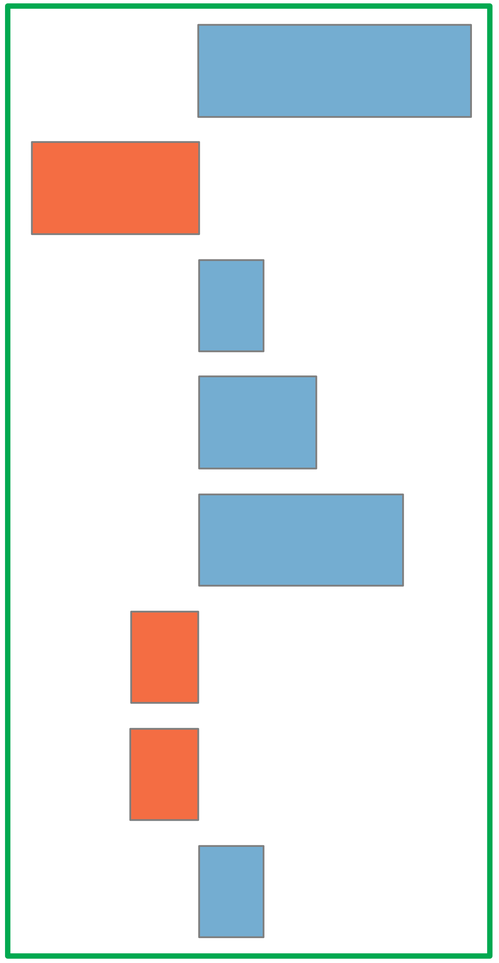


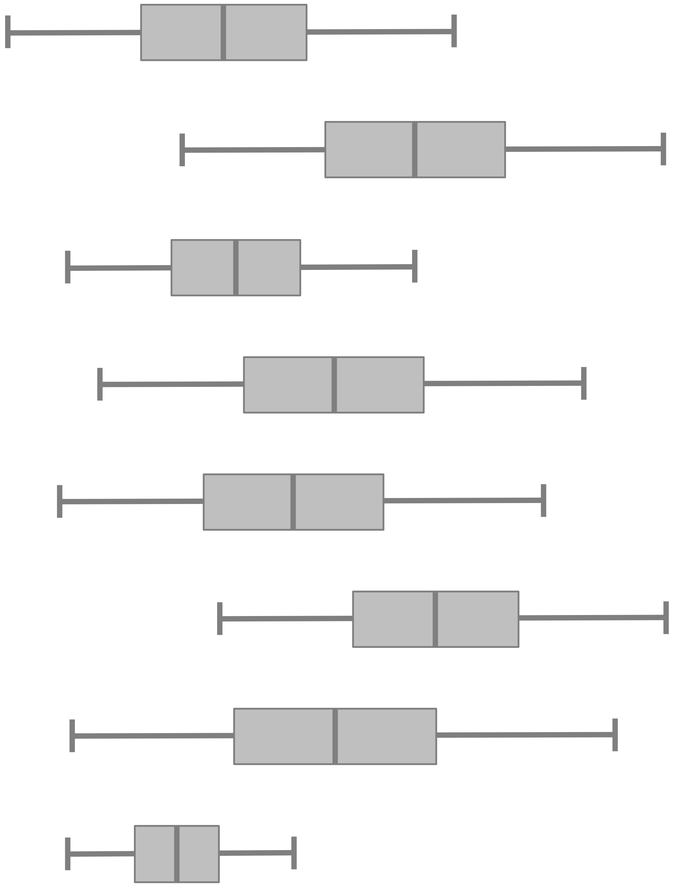
Which is the biggest intersection?
Sort By: Cardinality

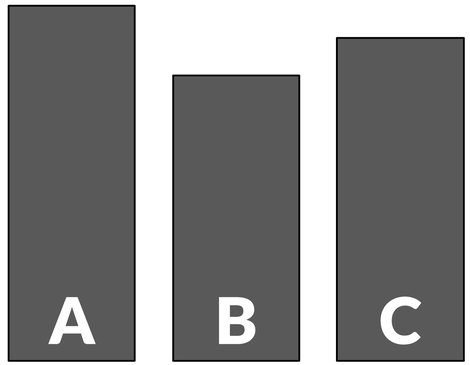


Which is the most ‘surprising’ intersection?
Sort By: Deviation

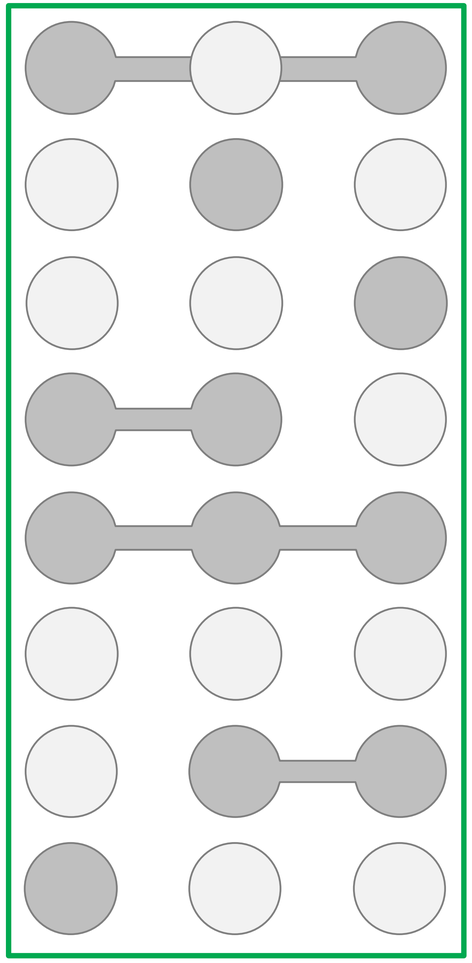


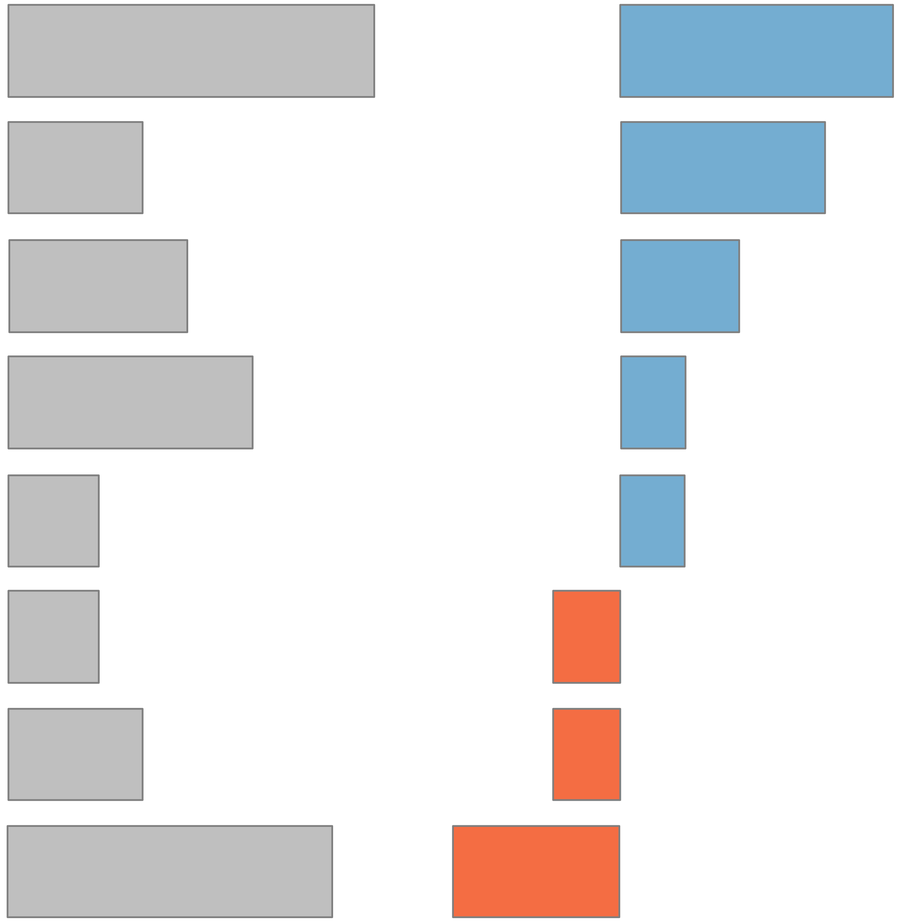


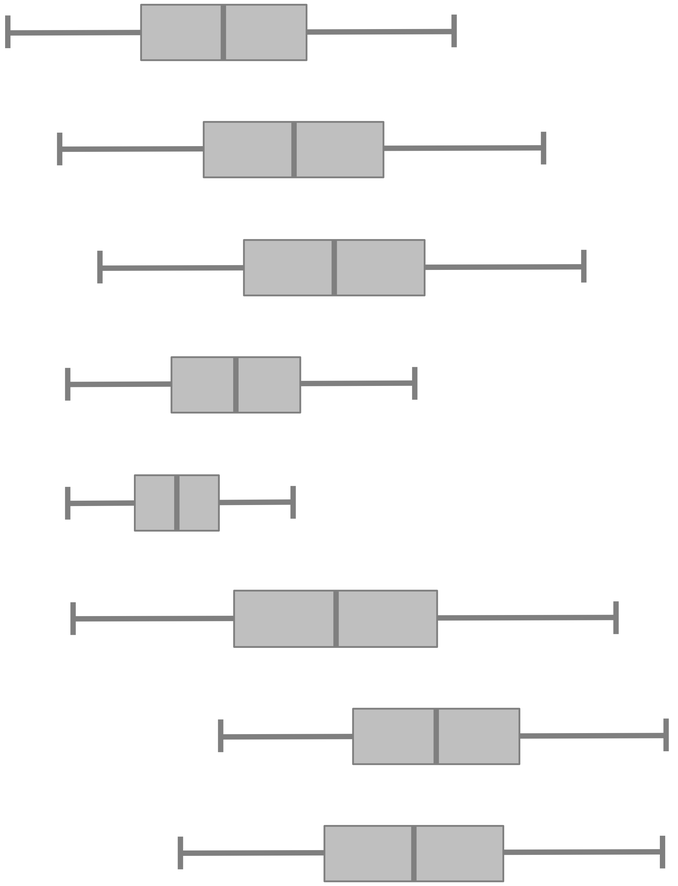


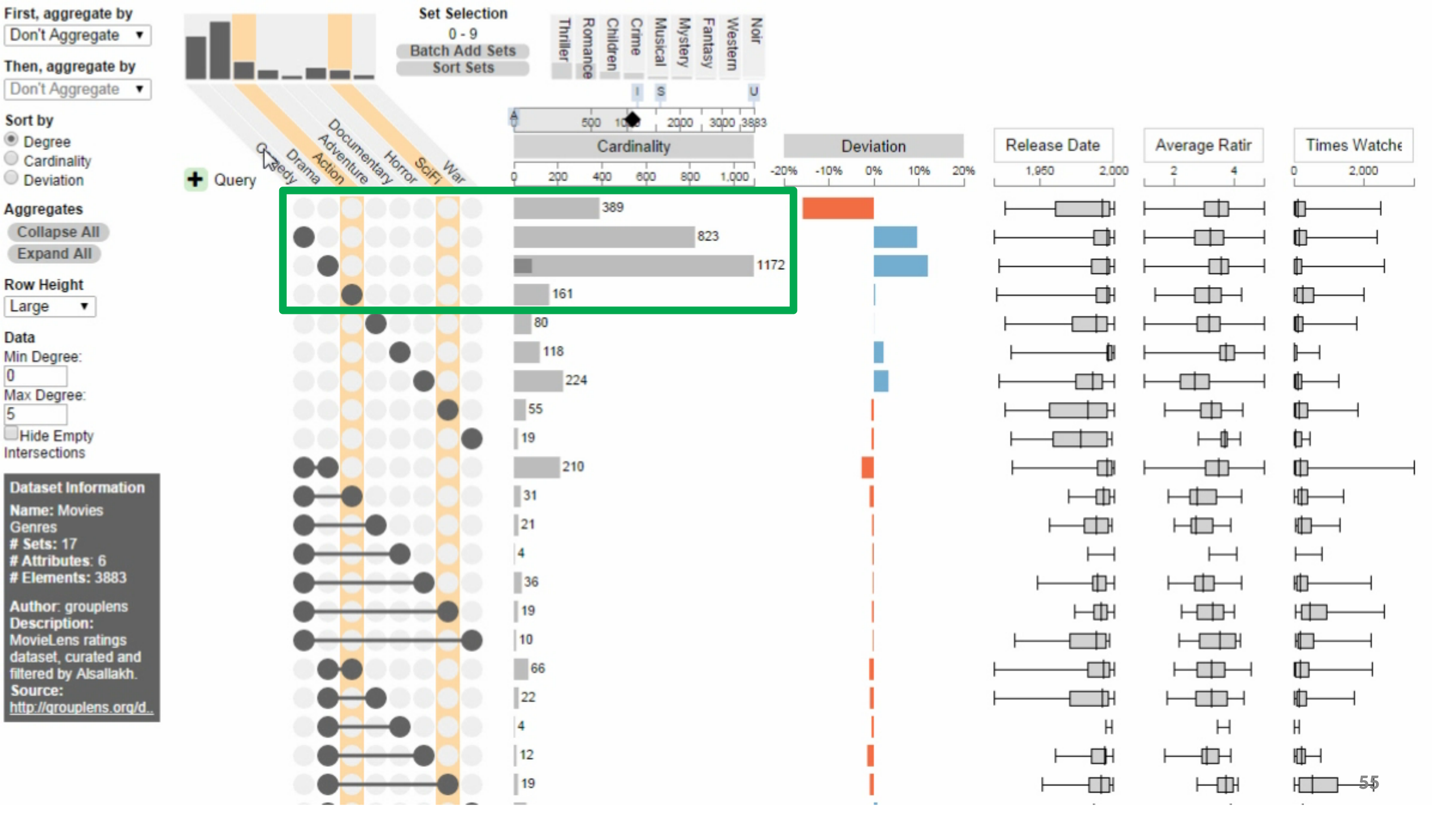


What are the properties of the intersections involving ‘A’?
Sort By: Set

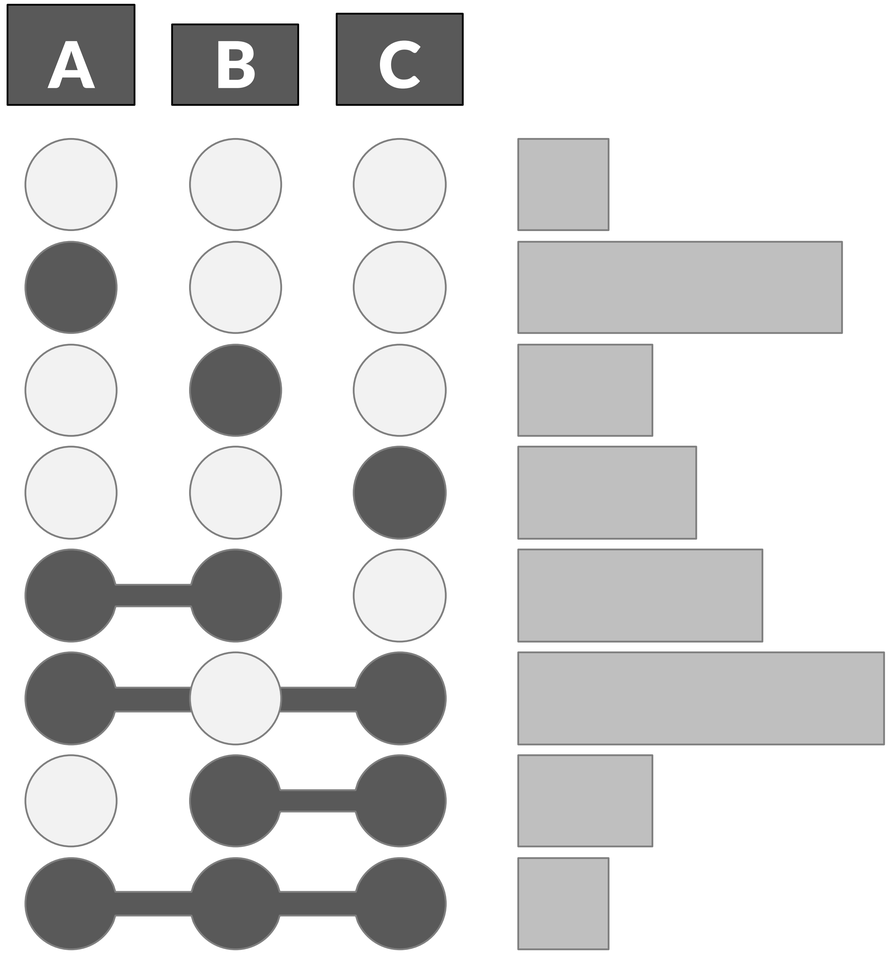




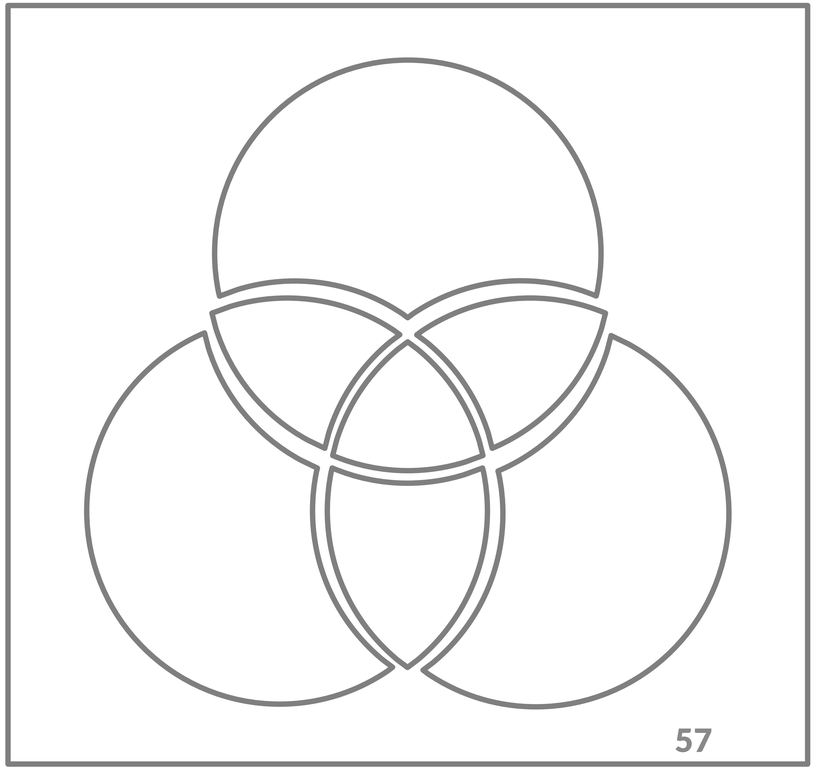


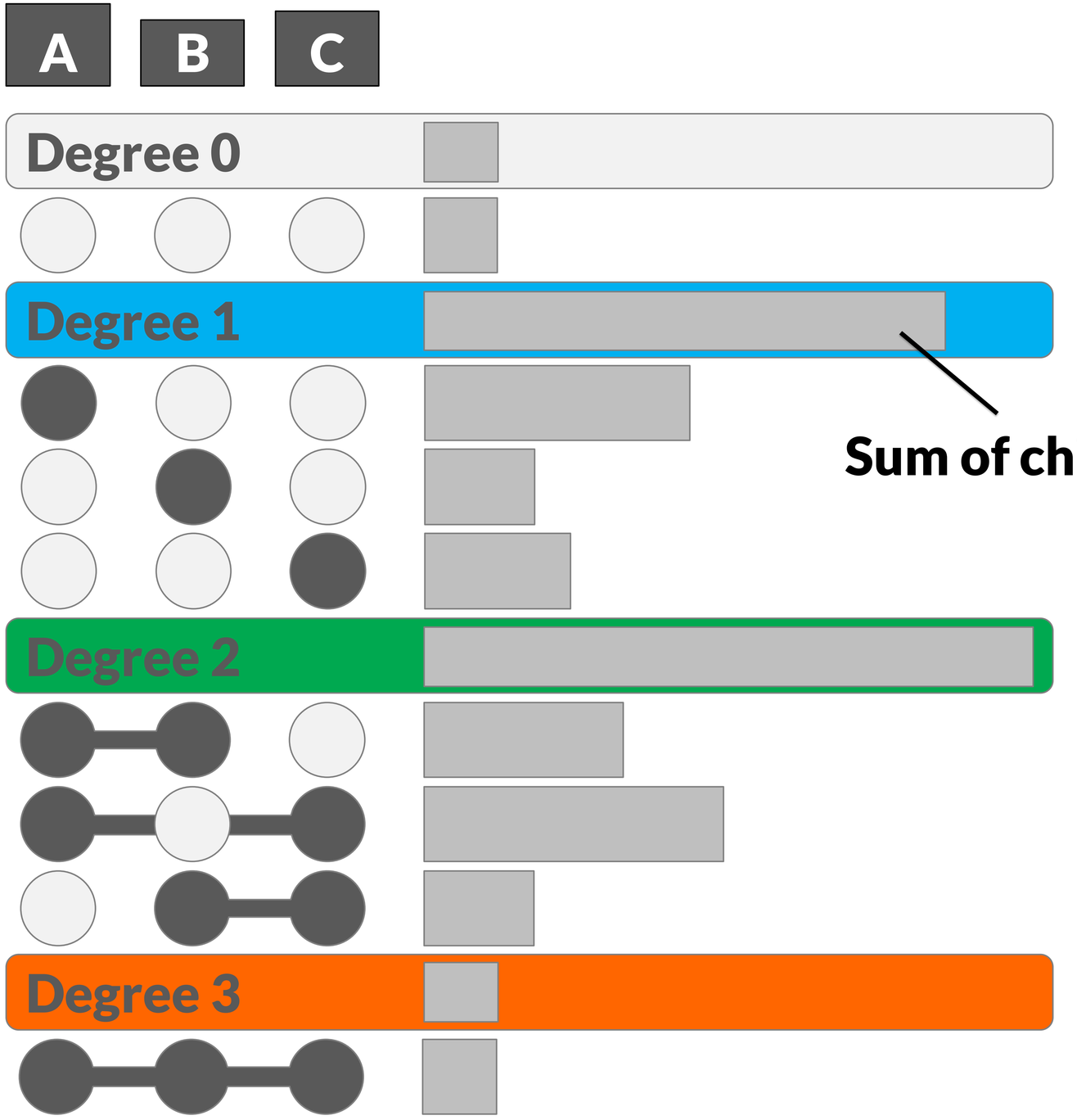

First, aggregate by
Don't Aggregate
Then, aggregate by
Don't Aggregate
Sort by
Degree
Cardinality
Deviation
Aggregates
Collapse All
Expand All
Row Height
Large
Data
Min Degree:
0
Max Degree:
5
Hide Empty Intersections
Dataset Information
Name: Movies Genres
Sets: 17
Attributes: 6
Elements: 3883
Author: 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
Comedy
Drama
Action
Adventure
Documentary
Horror
SciFi
War
Query
Cardinality
Deviation
Release Date
Average Ratir
Times Watche
389
823
1172
161
80
118
224
55
19
210
31
21
4
36
19
10
66
22
4
12
19

Aggregation



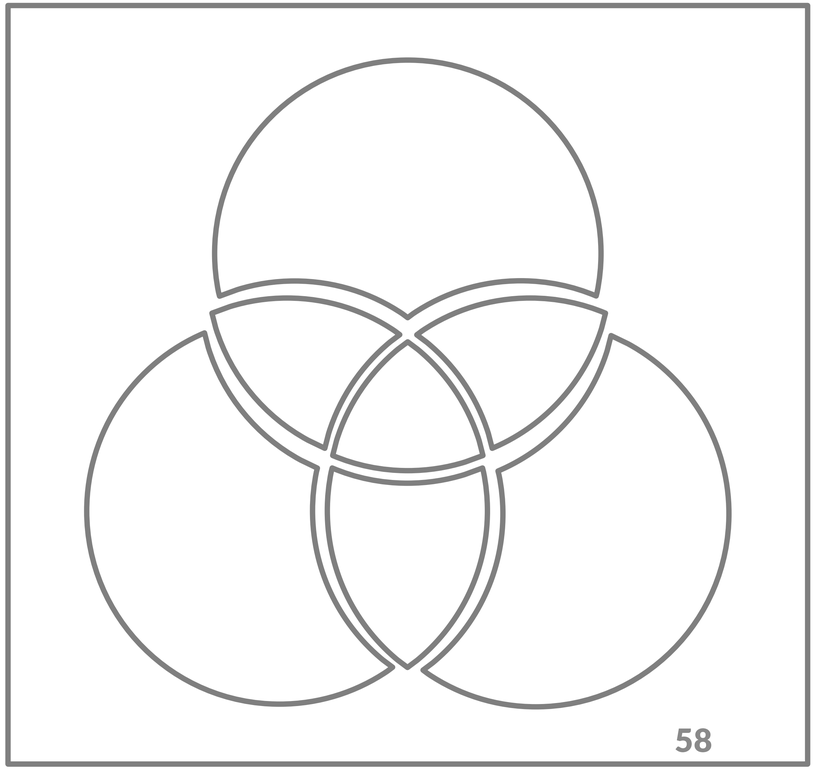
Are many items shared between two sets?
Aggregate By: Degree



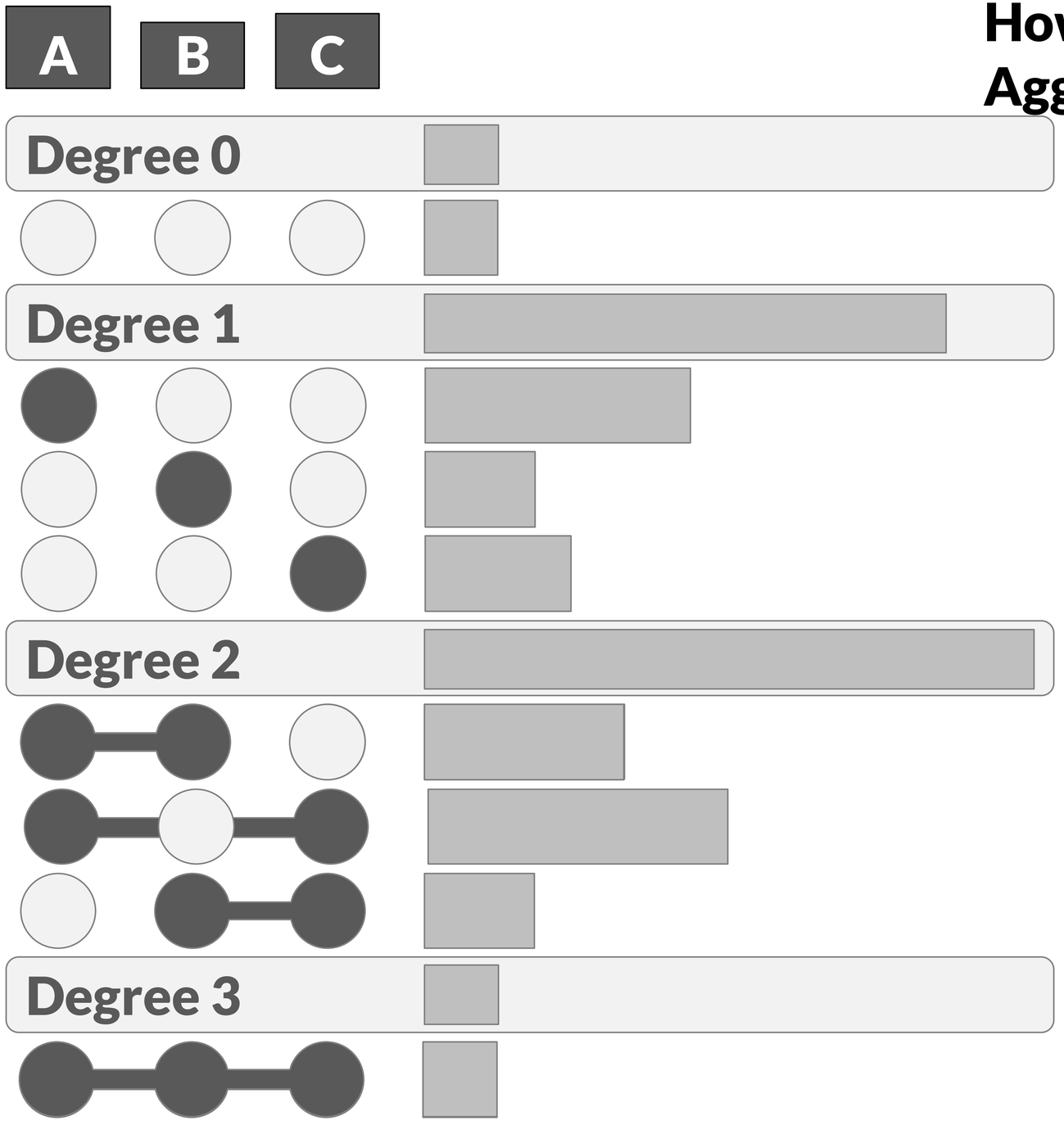


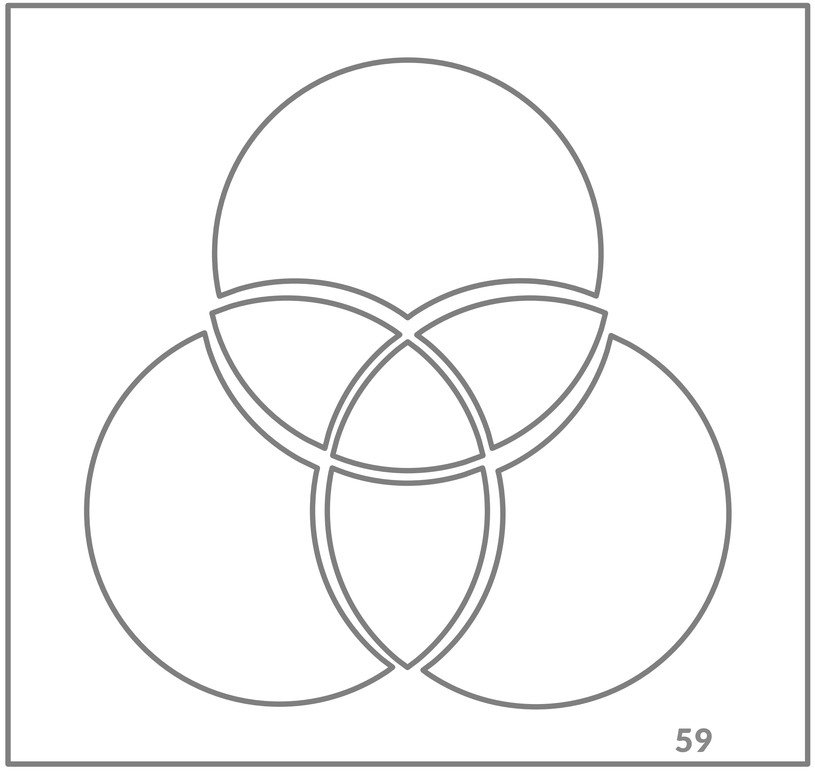
Are many items shared between two sets?
Aggregate By: Degree

Sum of children

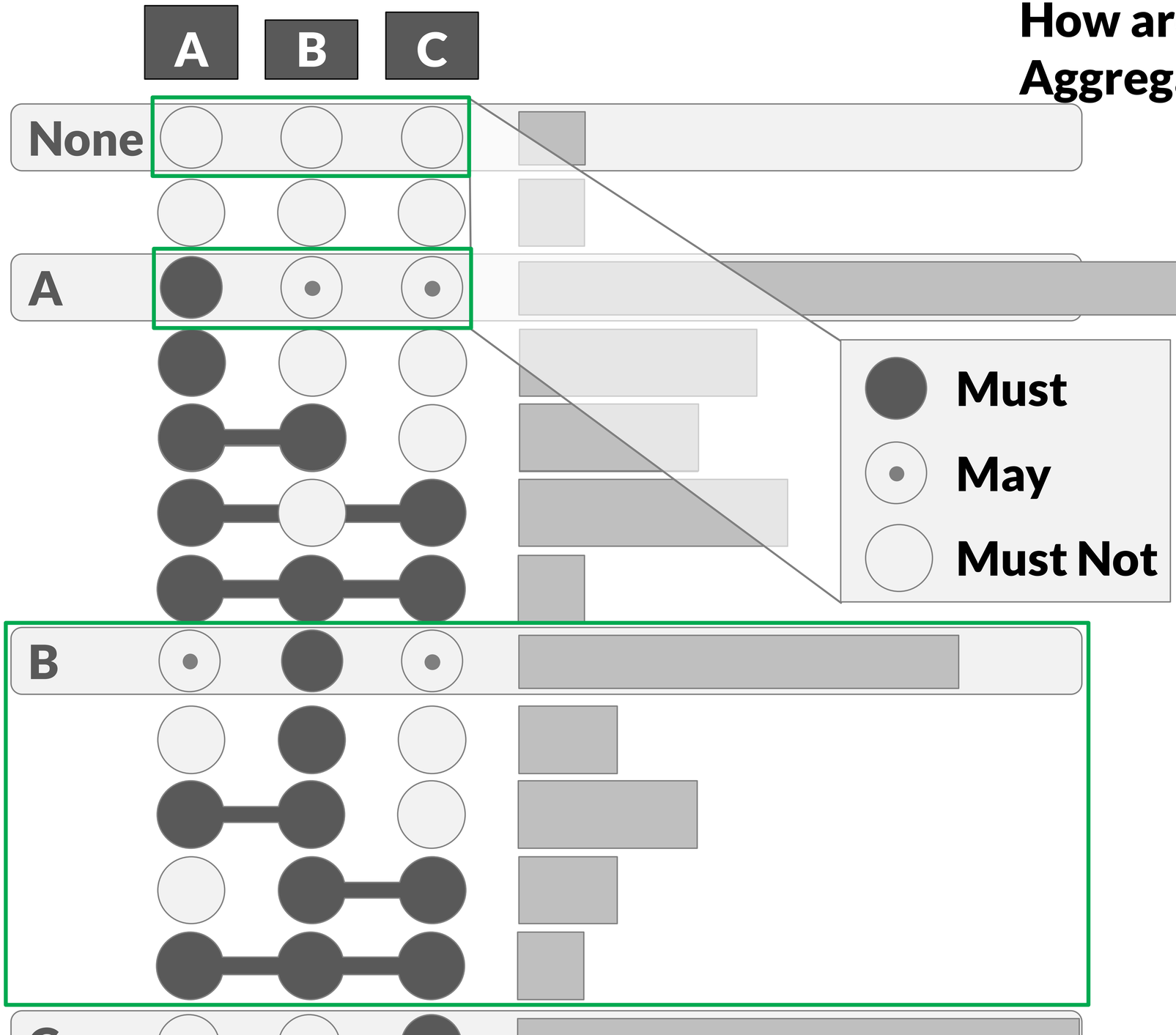


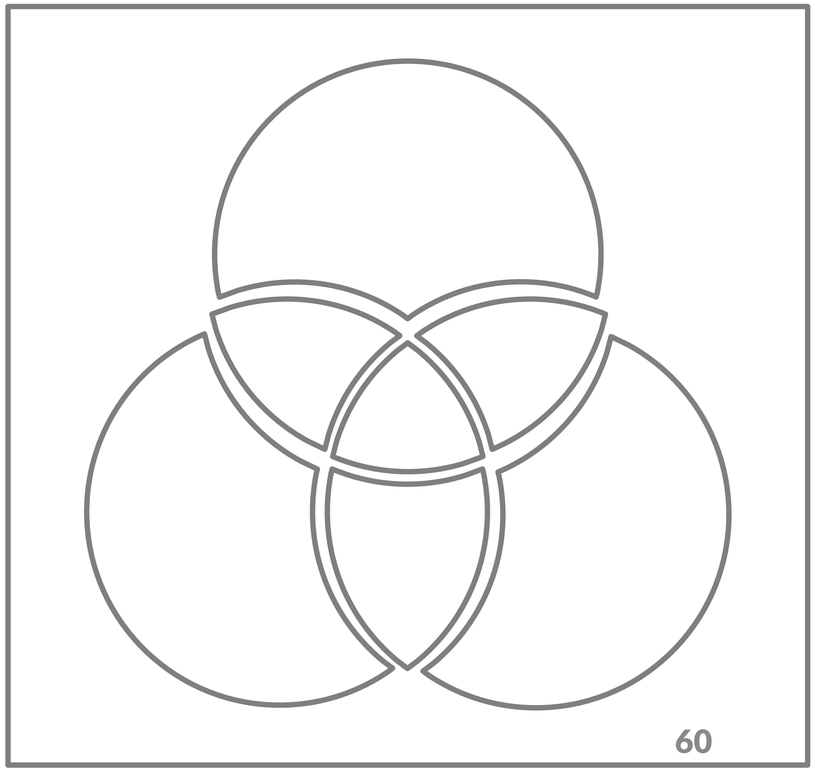
How are the elements of ‘B’ distributed?
Aggregate By: Set



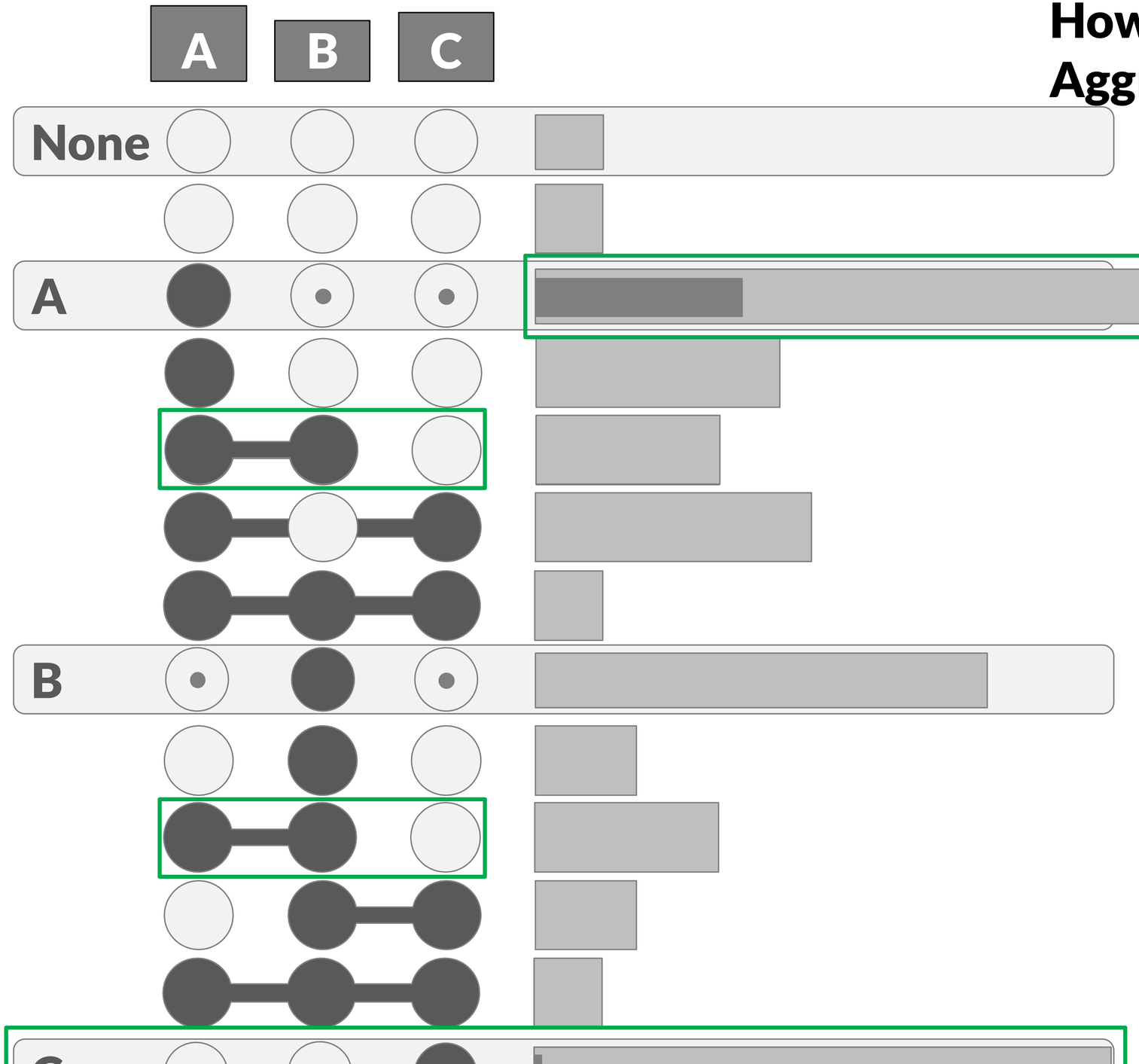


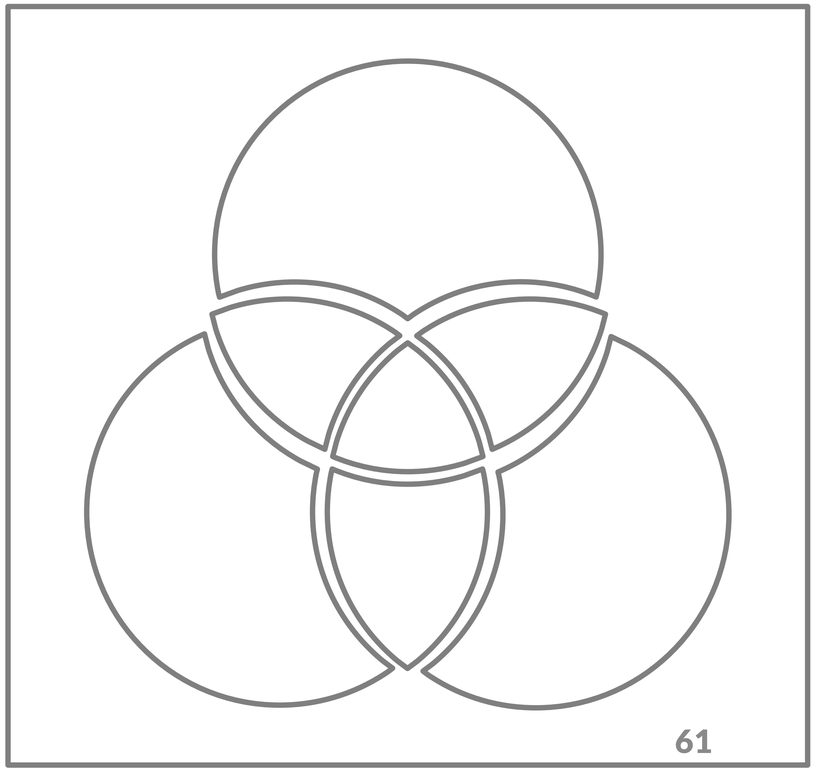
How are the elements of ‘B’ distributed?
Aggregate By: Set

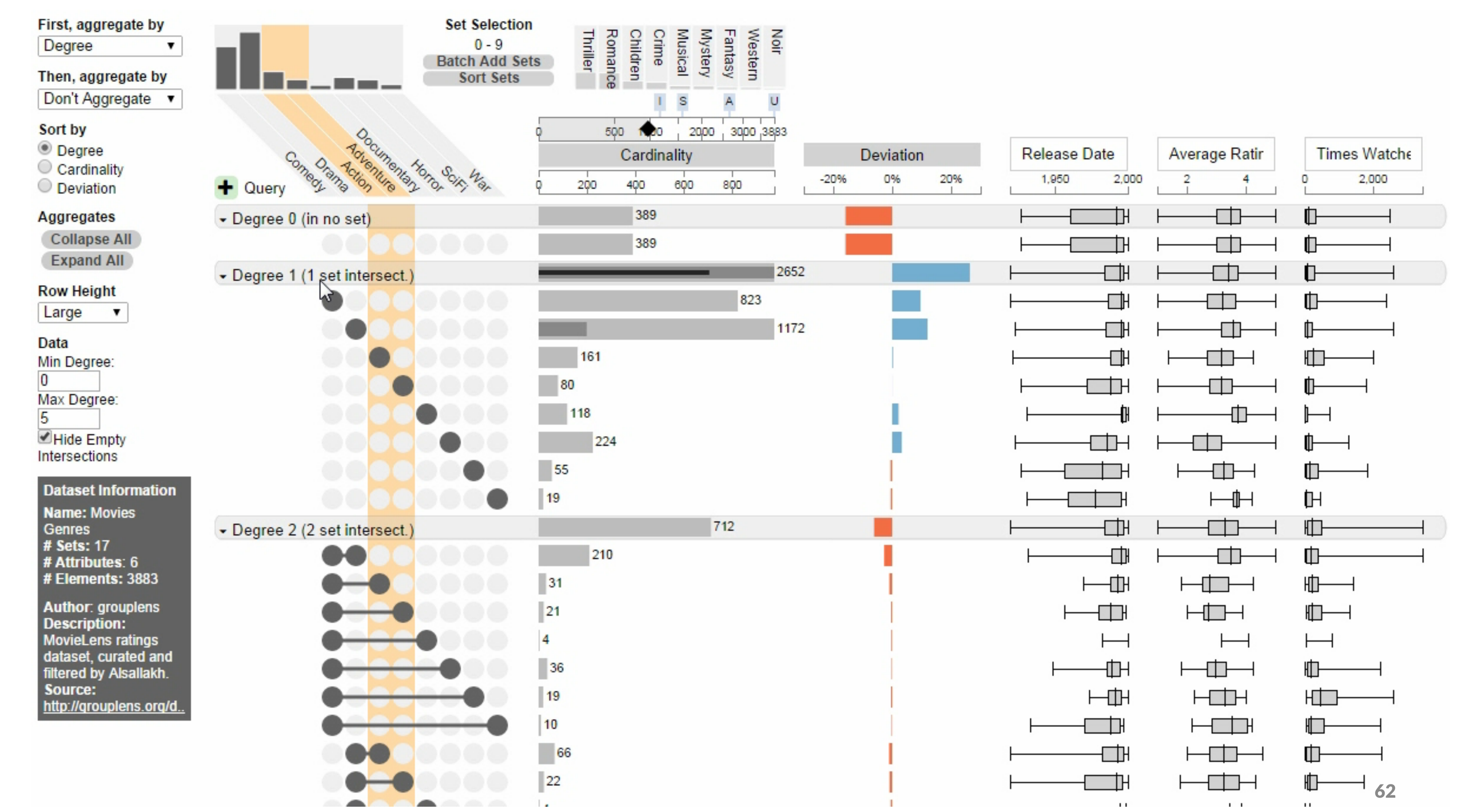




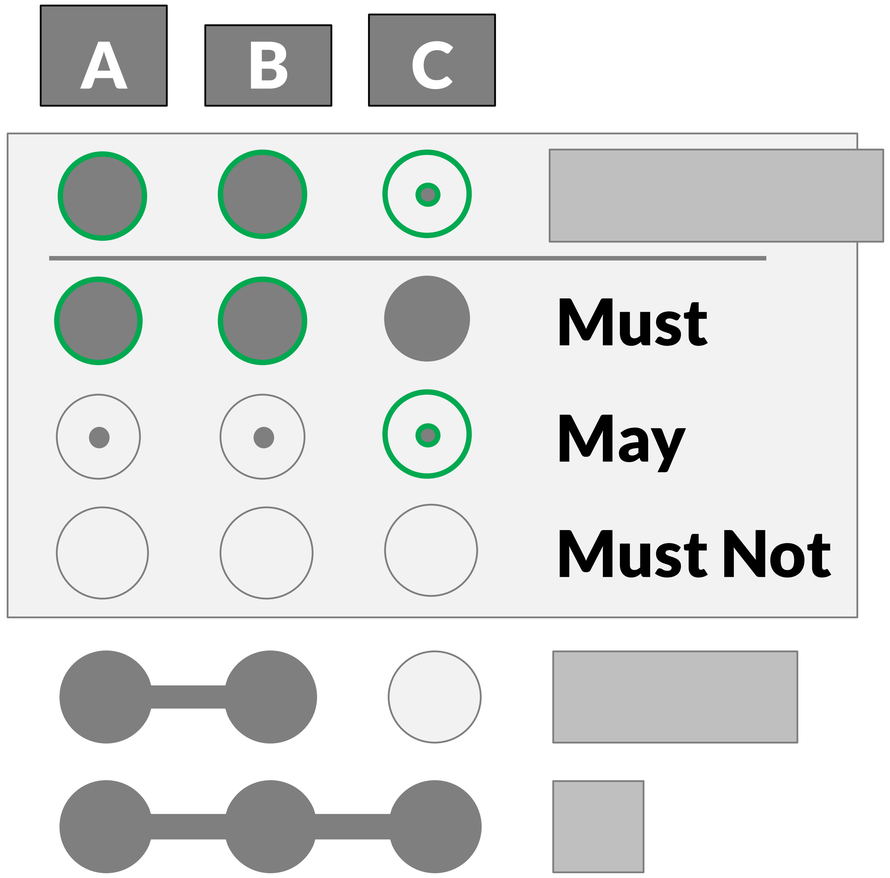
How are the elements of ‘B’ distributed?
Aggregate By: Set

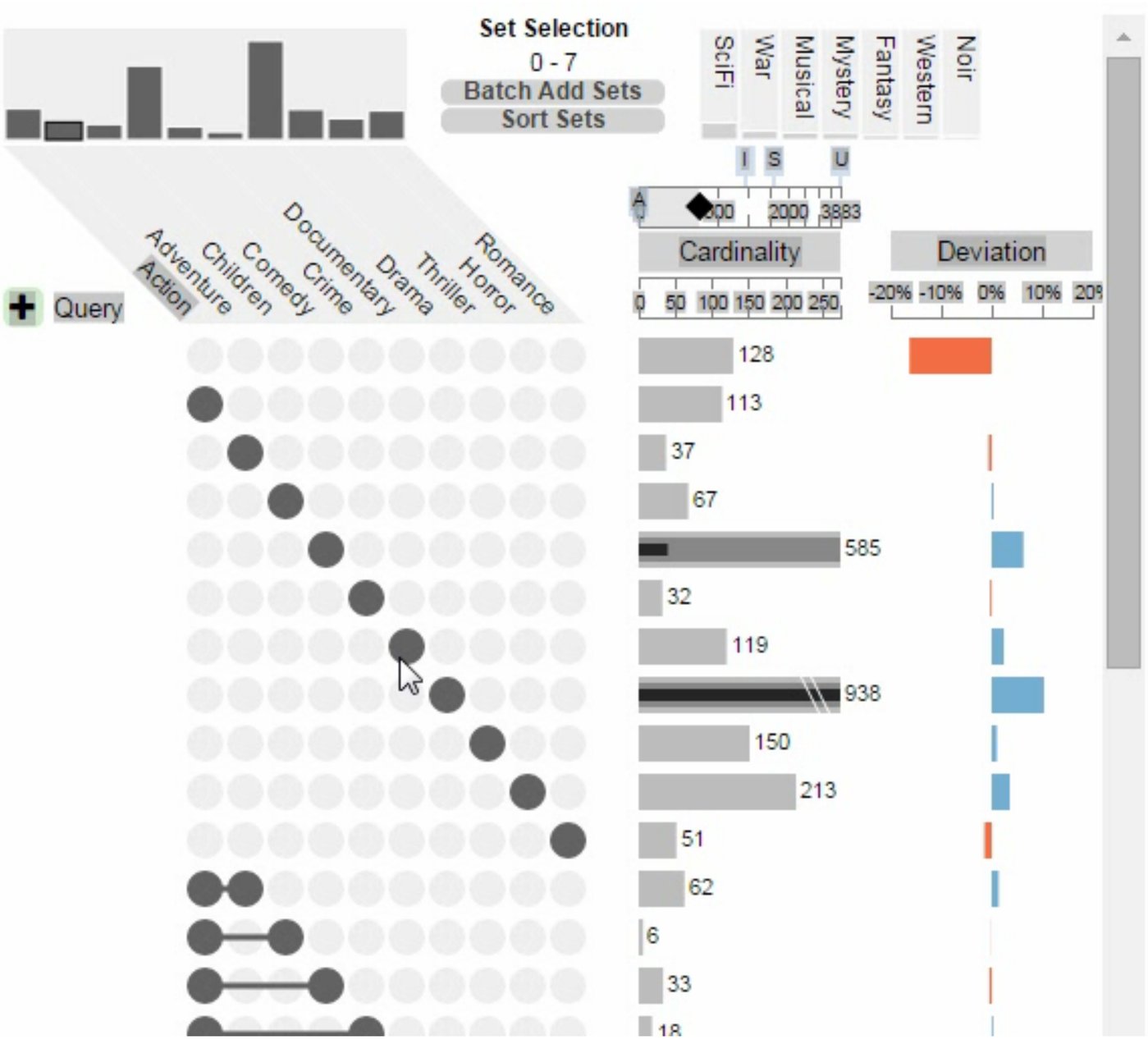





First, aggregate by
Degree
Then, aggregate by
Don't Aggregate
Sort by
Degree
Cardinality
Deviation
Aggregates
Collapse All
Expand All
Row Height
Large
Data
Min Degree:
0
Max Degree:
5
Hide Empty Intersections
Dataset Information
Name: Movies Genres
Sets: 17
Attributes: 6
Elements: 3883
Author: grouplens
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
Comedy
Drama
Action
Adventure
Documentary
Horror
SciFi
War
Query
Cardinality
Deviation
Release Date
Average Ratir
Times Watche
Degree 0 (in no set)
389
389
Degree 1 (1 set intersect.)
2652
823
1172
161
80
118
224
55
19
Degree 2 (2 set intersect.)
712
210
31
21
4
36
19
10
66
22

Queries


A
B
C
Must
May
Must Not



Element Visualizations

No visualizations configured. Click + button to add a new visua

Scatterplot

Element Queries

No queries. Click + button to add a new query.

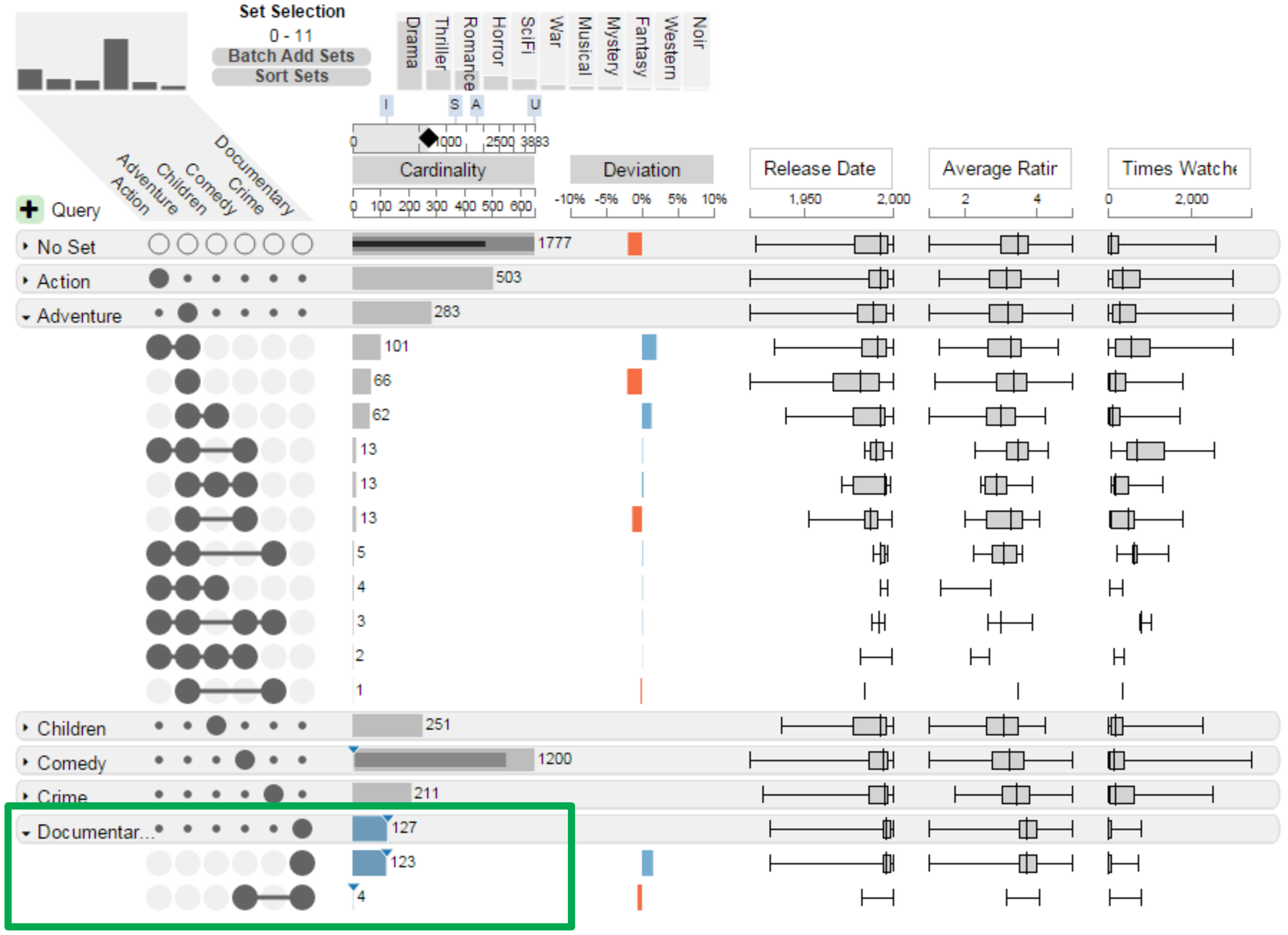
Query Filters

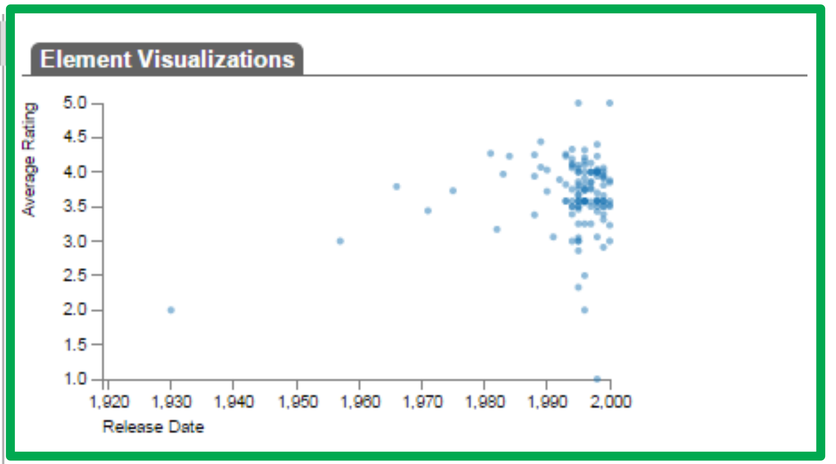
No active query.

Query Results

No active query.

Elements & Attributes

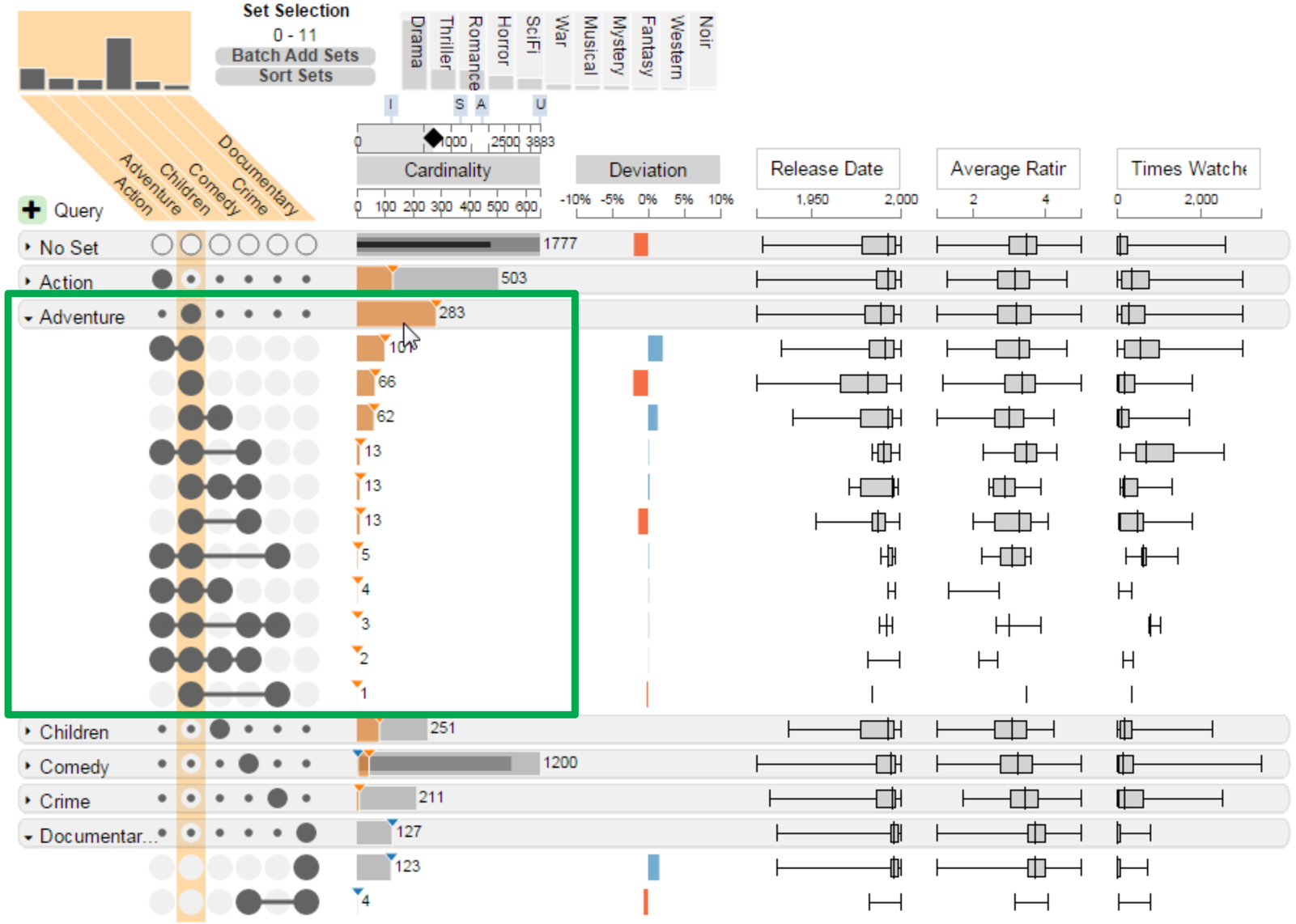




Query Results

Name	Release Date
Across the Sea of Time (1995)	1995
Nico Icon (1995)	1995
Heidi Fleiss: Hollywood Madam (1995)	1995
Catwalk (1995)	1995
Anne Frank Remembered (1995)	1995
Jupiter's Wife (1994)	1994
Sonic Outlaws (1995)	1995
From the Journals of Jean Seberg (1995)	1995
Man of the Year (1995)	1995

How do documentaries compare to adventure movies?



Element Queries

127 283

Query Filters

Subset | Sets

Name Contains

Query Results

Name	Release Date
Jumanji (1995)	1995
Tom and Huck (1995)	1995
GoldenEye (1995)	1995
Cutthroat Island (1995)	1995
City of Lost Children, The (1995)	1995
Wings of Courage (1995)	1995
Mortal Kombat (1995)	1995
Kids of the Round Table (1995)	1995
Indian in the Cupboard, The (1995)	1995
White Squall (1996)	1996
Muppet Treasure Island (1996)	1996

How do documentaries compare to adventure movies?

Applications

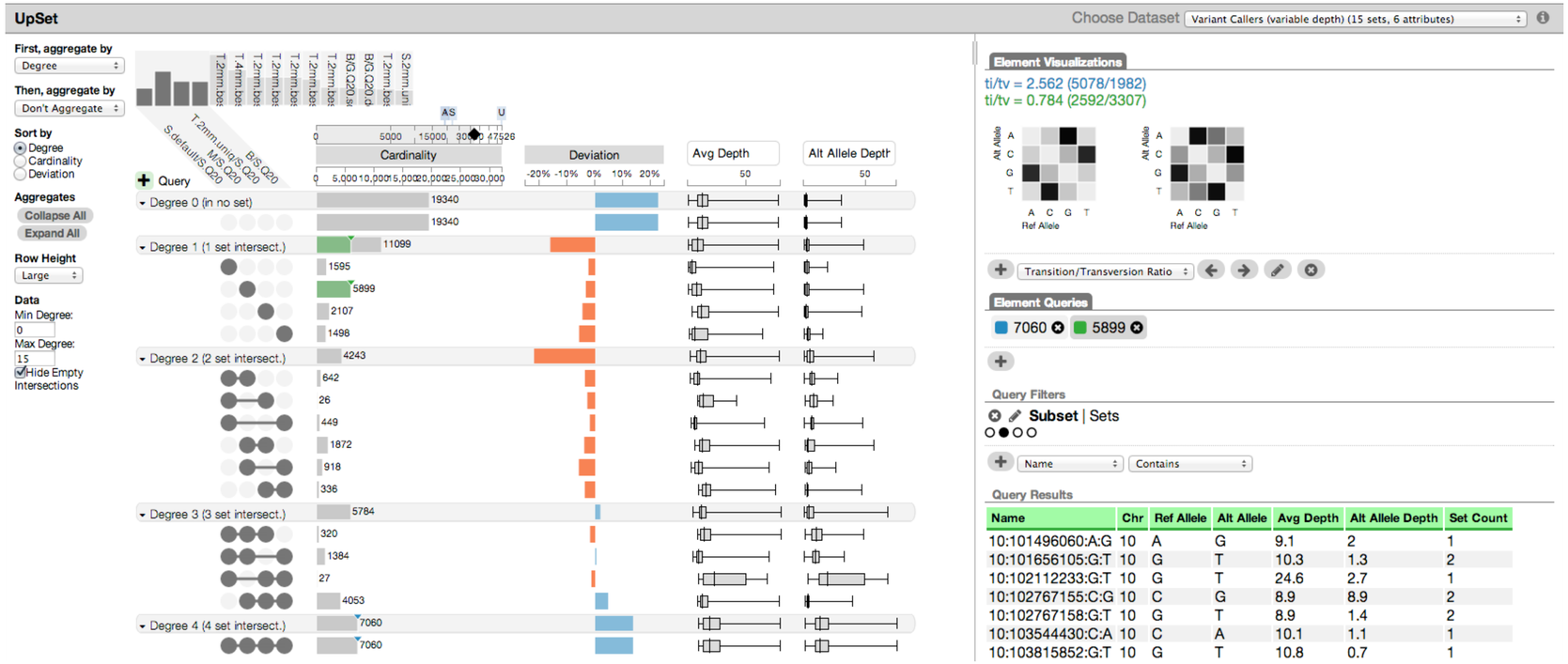
Applications

Genetics

Economics

Pharmacology

Social Networks



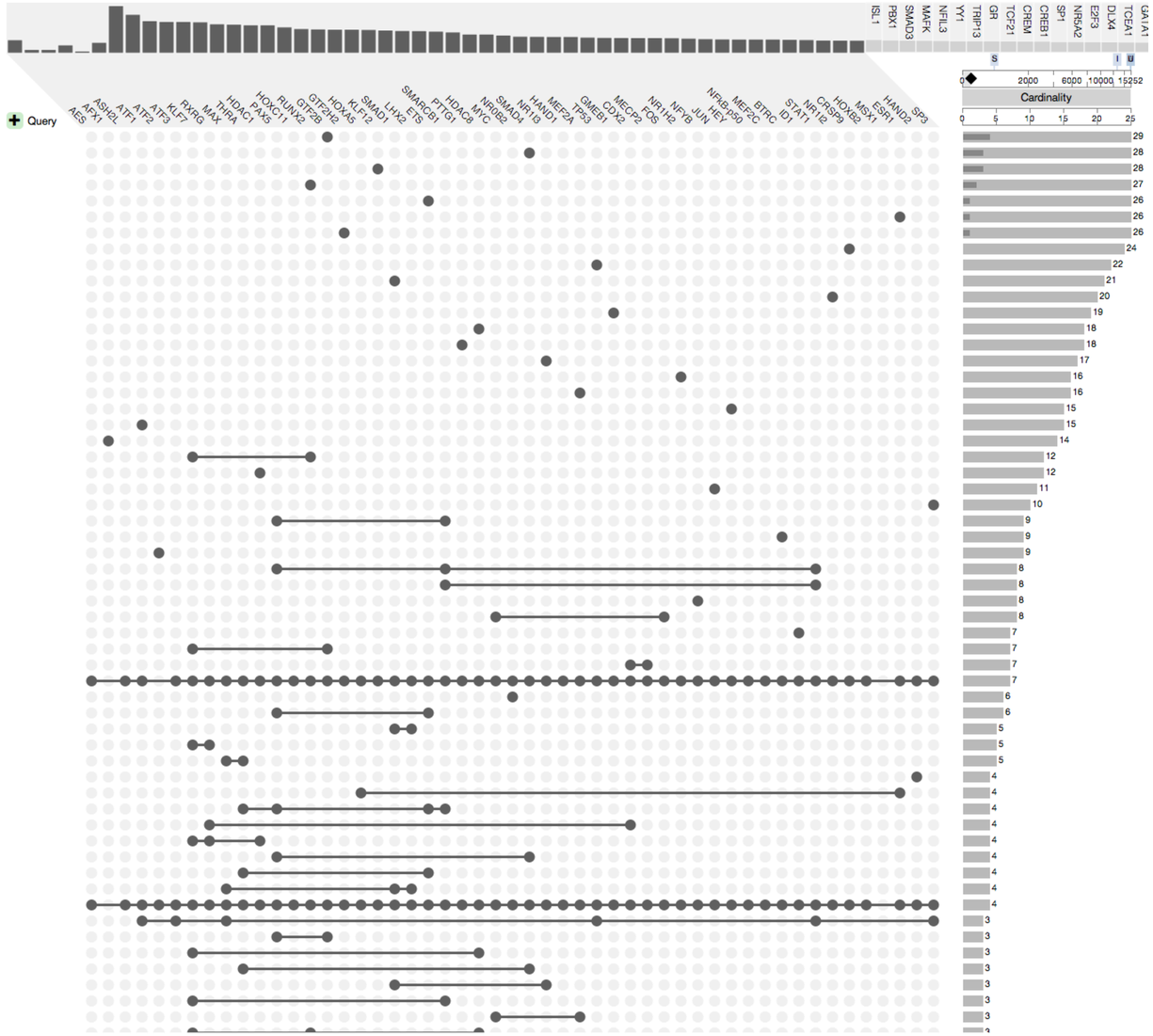
Scalability

Comfortable: ~15 sets

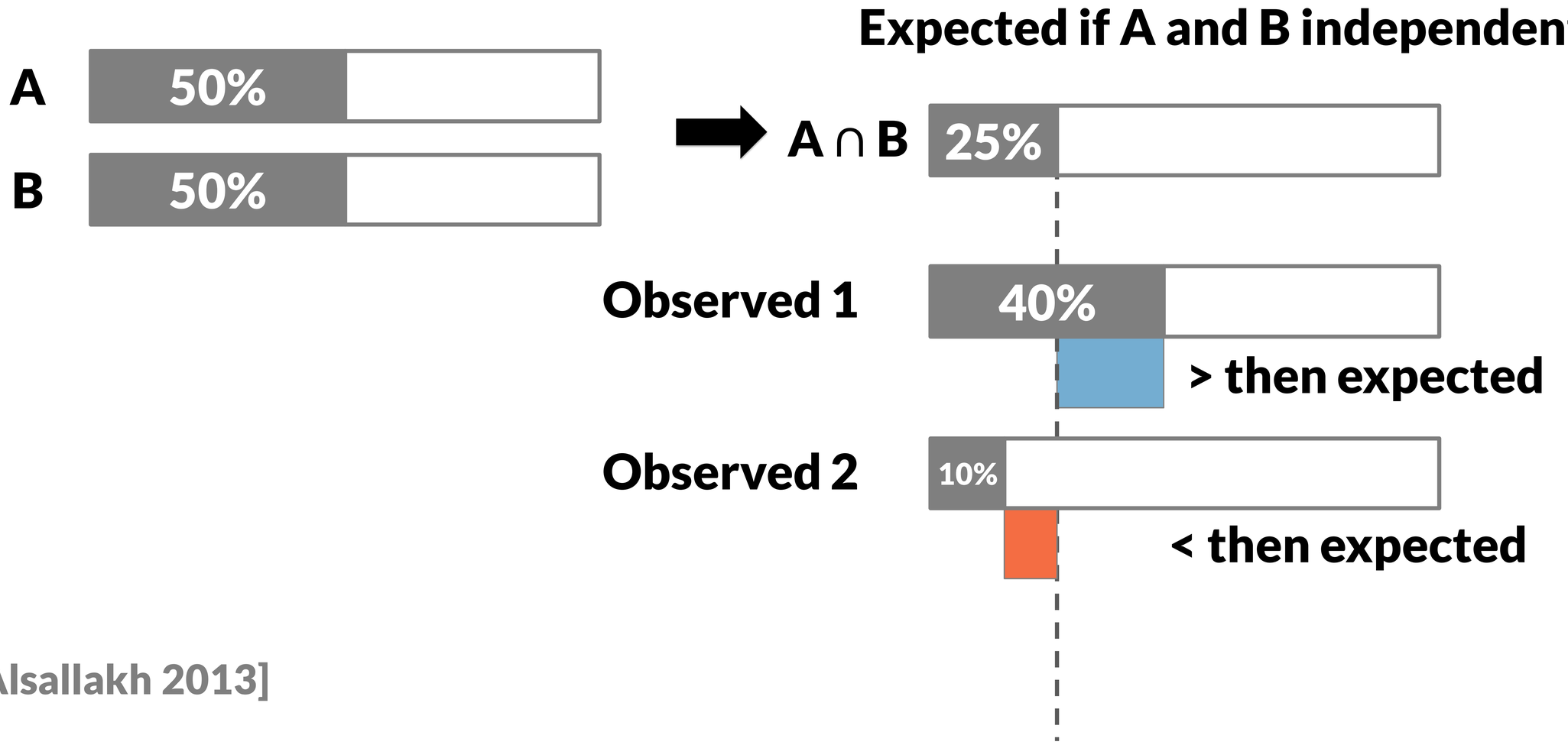
Possible: ~40 sets

Scales with the number of non-empty intersections

- **Most datasets are sparse**



Deviation Measure

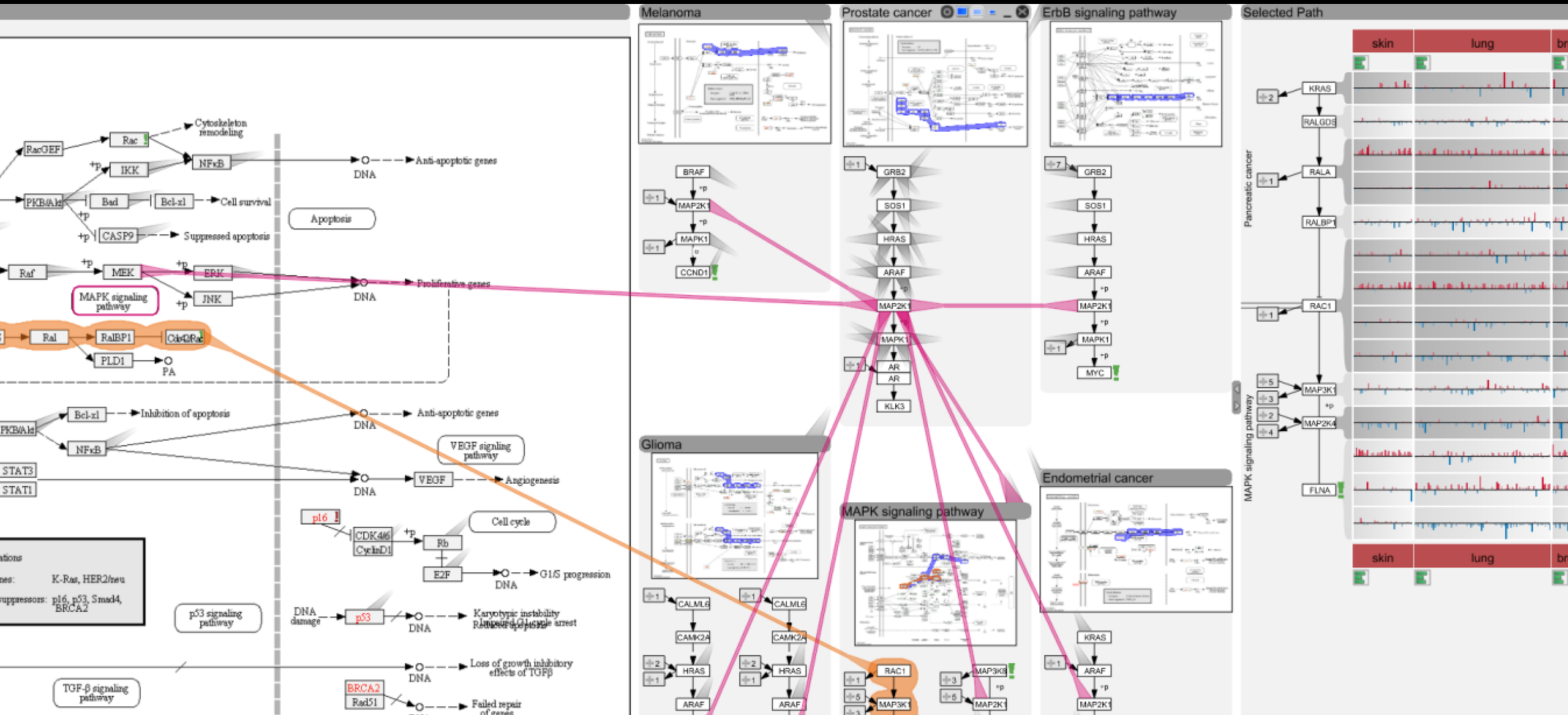


[Alsallakh 2013]

http://vcg.github.io/upset

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

Pathways



Publications

Alexander Lex, Christian Partl, Denis Kalkofen, Marc Streit, Anne Mai Wasserman, Samuel Gratzl, Dieter Schmalstieg and Hanspeter Pfister

Entourage: Visualizing Relationships between Biological Pathways using Contextual Subsets

IEEE TVCG (InfoVis '13), 2013

Christian Partl, Alexander Lex, Marc Streit, Denis Kalkofen, Karl Kashofer and Dieter Schmalstieg

enRoute: Dynamic Path Extraction from Biological Pathway Maps for Exploring Heterogeneous Experimental Datasets

BMC Bioinformatics, 2013

Christian Partl, Alexander Lex, Marc Streit, Denis Kalkofen, Karl Kashofer, Dieter Schmalstieg

enRoute: Dynamic Path Extraction from Biological Pathway Maps for In-Depth Experimental Data Analysis

BioVis, 2012

Best Paper Award

A Pathway

KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions, and relations

KEGG2 PATHWAY BRITE MODULE KO GENOME GENES LIGAND DISEASE DRUG DBGET

Select prefix: map | Organism | Enter keywords | Go | Help

[New maps | Update history]

Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps representing our knowledge on the molecular interaction and reaction networks for:

1. **Metabolism**
   Global/overview Carbohydrate Energy Lipid Nucleotide Amino acid Other amino Glycan
   Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Chemical structure
2. **Genetic Information Processing**
3. **Environmental Information Processing**
4. **Cellular Processes**
5. **Organismal Systems**
6. **Human Diseases**

and also on the structure relationships (KEGG drug structure maps) in:

7. **Drug Development**

Pathway Mapping

KEGG PATHWAY mapping is the process to map molecular datasets, especially large-scale datasets in genomics, transcriptomics, proteomics, and metabolomics, to the KEGG pathway maps for biological interpretaion of higher-level systemic functions.

- Search Pathway - basic pathway mapping tool
- Search&Color Pathway - advanced pathway mapping tool
- Color Pathway - selected pathway map coloring tool

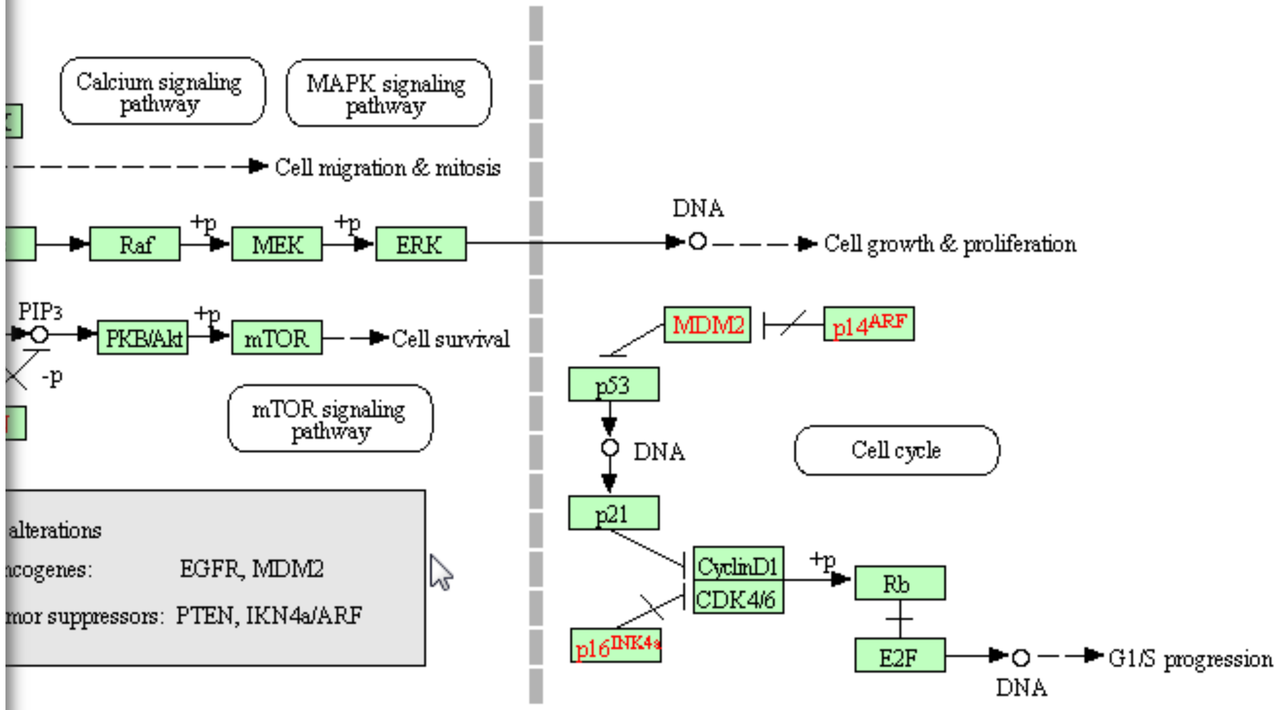
1. Metabolism

1.0 Global and overview maps

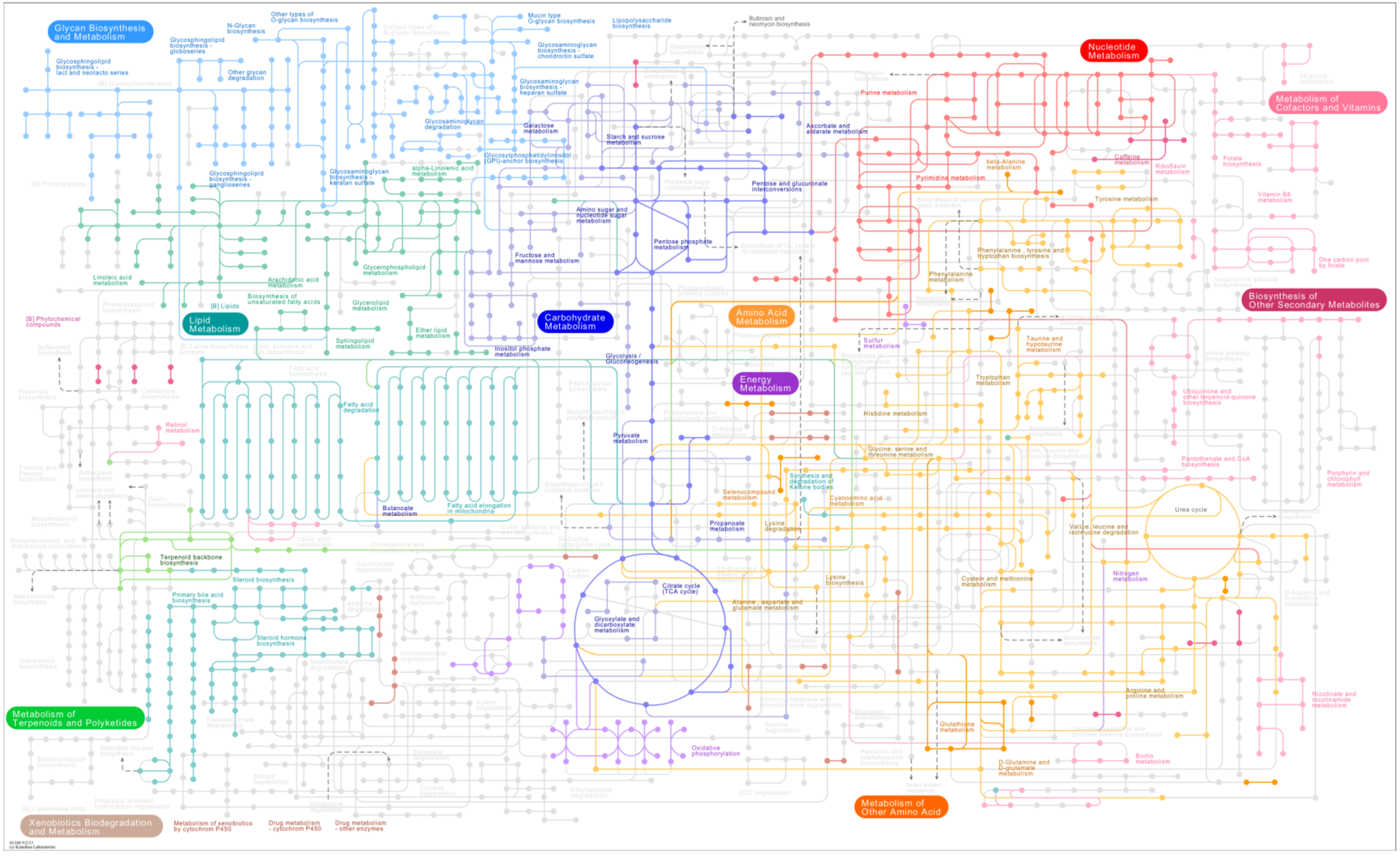
Metabolic pathways	[KEGG Atlas]	KEGG modules
Biosynthesis of secondary metabolites	[KEGG Atlas]	KEGG reaction modules
Microbial metabolism in diverse environments	[KEGG Atlas]	
Carbon metabolism	[KEGG Atlas]	
2-Oxocarboxylic acid metabolism	[KEGG Atlas]	
Fatty acid metabolism	[KEGG Atlas]	
Biosynthesis of amino acids	[KEGG Atlas]	
Degradation of aromatic compounds	[KEGG Atlas]	

1.1 Carbohydrate metabolism

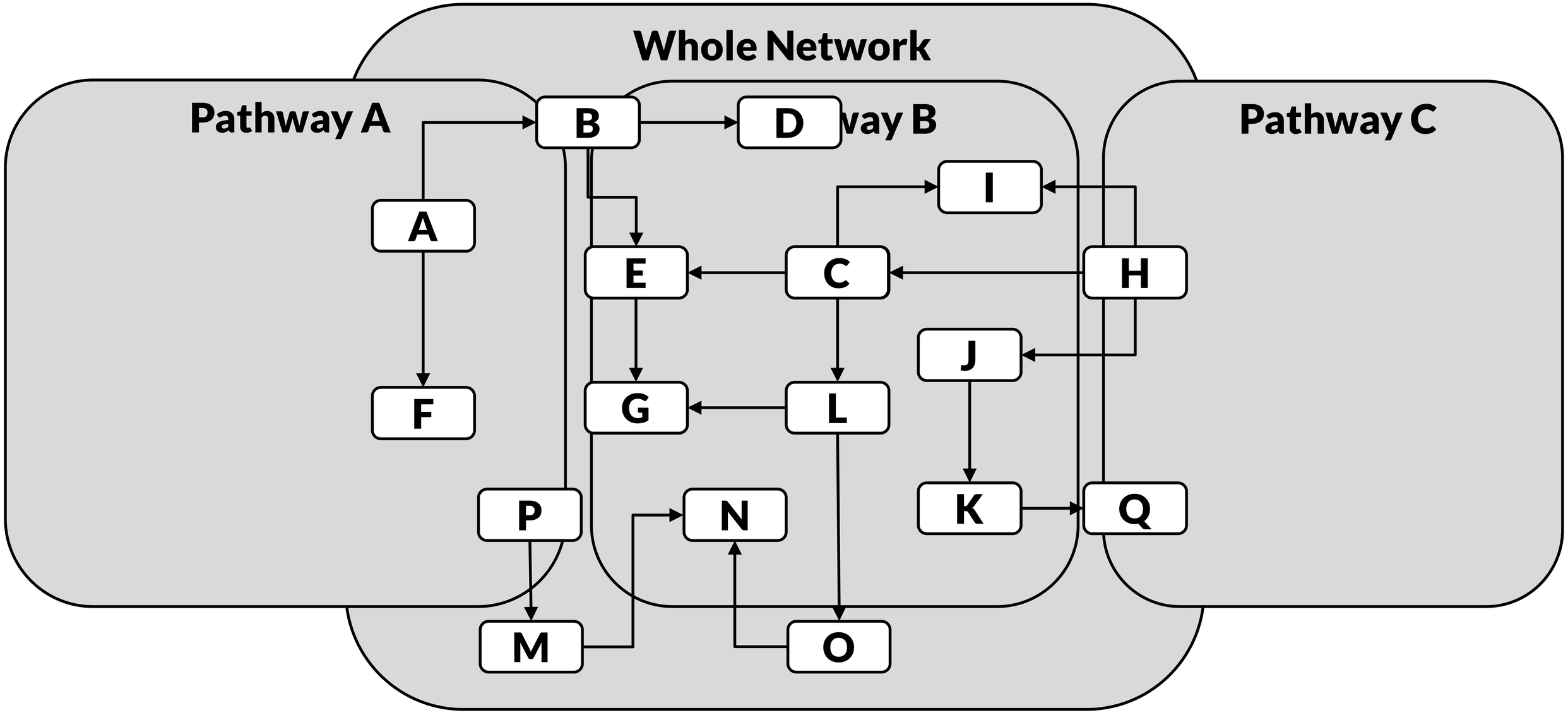
Glycolysis / Gluconeogenesis Enzymes



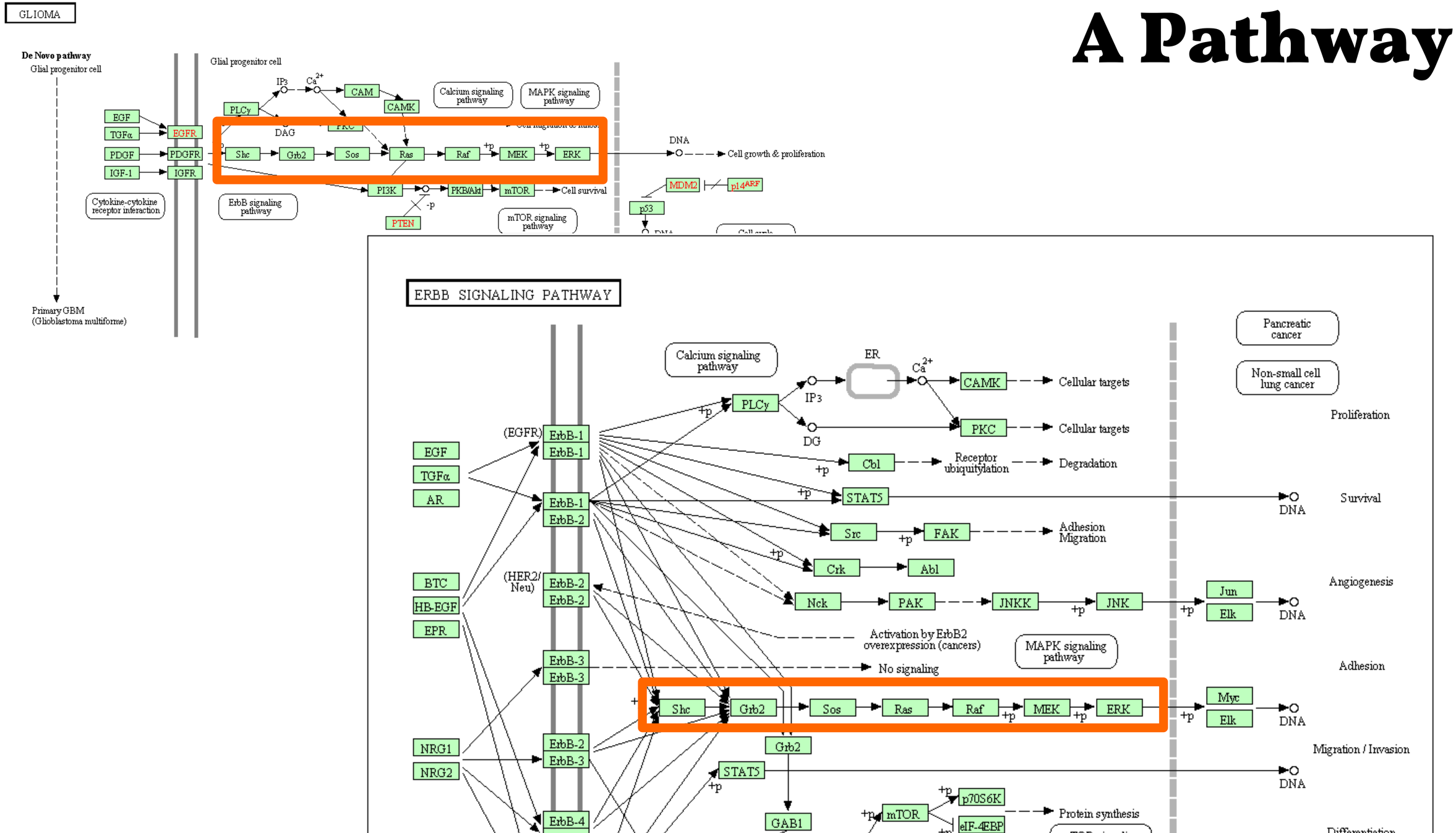
The bigger picture



Background


Whole Network
Pathway A
Pathway B
Pathway C
A
B
D
E
C
H
I
F
G
L
J
P
N
K
Q
M
O

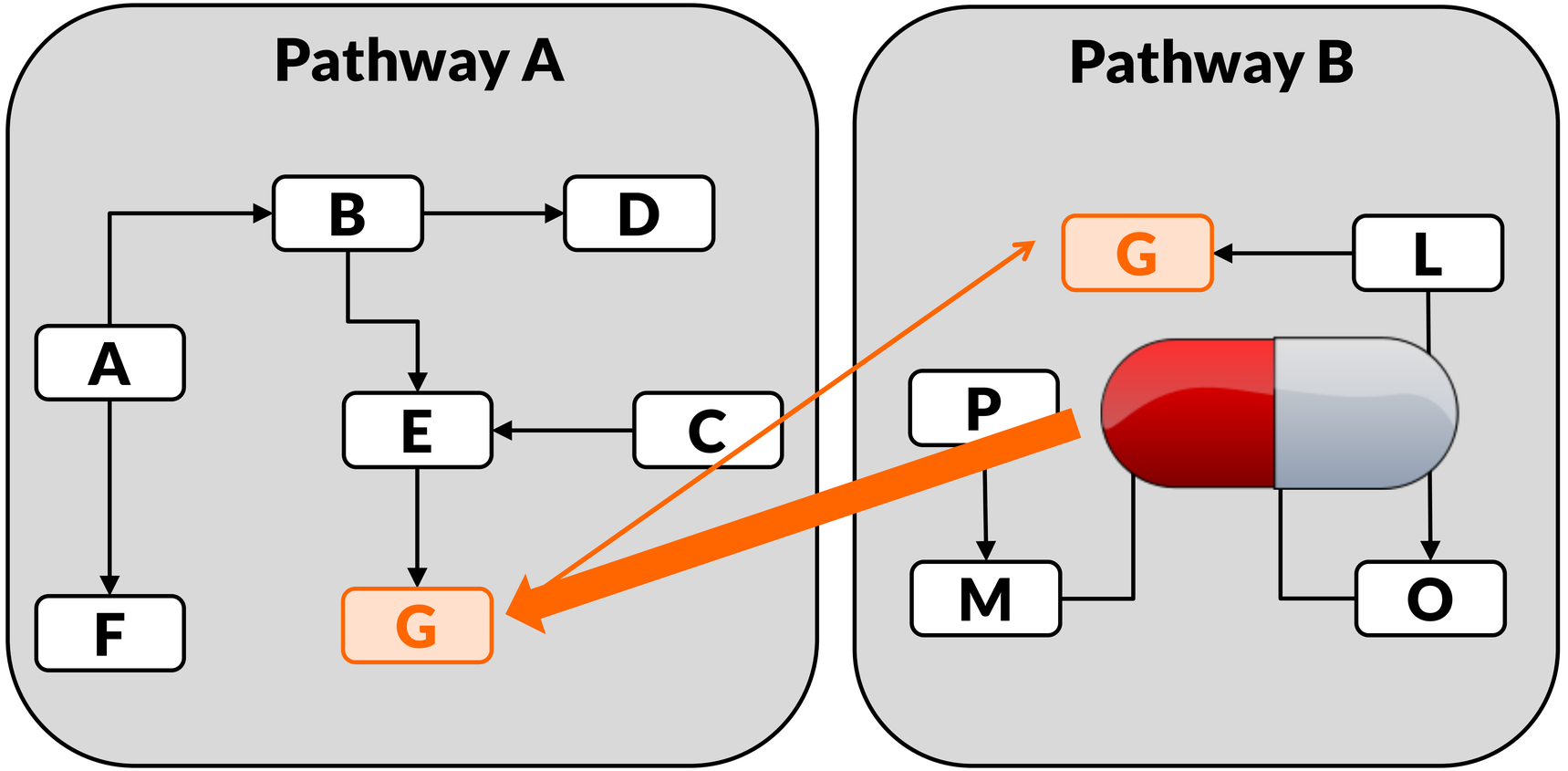
A Pathway


GLIOMA
De Novo pathway
Glial progenitor cell
Primary GBM
(Glioblastoma multiforme)
Glial progenitor cell
EGF
TGFα
PDGF
IGF-1
EGFR
PDGFR
IGFR
Cytokine-cytokine receptor interaction
IP3
Ca2+
PLCγ
CAM
CAMK
Calcium signaling pathway
MAPK signaling pathway
DAG
Shc
Grb2
Sos
Ras
Raf
MEK
ERK
DNA
Cell growth & proliferation
PI3K
PKB/Akt
mTOR
Cell survival
PTEN
ErbB signaling pathway
mTOR signaling pathway
MDM2
p14ARF
p53
ERBB SIGNALING PATHWAY
Pancreatic cancer
Non-small cell lung cancer
Proliferation
Survival
Angiogenesis
Adhesion
Migration / Invasion
Calcium signaling pathway
ER
Ca2+
CAMK
Cellular targets
PKC
Cellular targets
PLCγ
IP3
DG
Cbl
Receptor ubiquitylation
Degradation
STAT5
DNA
Src
FAK
Adhesion Migration
Crk
Abl
Nck
PAK
JNKK
JNK
Jun
Elk
DNA
Activation by ErbB2 overexpression (cancers)
No signaling
MAPK signaling pathway
Shc
Grb2
Sos
Ras
Raf
MEK
ERK
Myc
Elk
DNA
Grb2
STAT5
DNA
GAB1
mTOR
p70S6K
eIF-4EBP
Protein synthesis
(EGFR)
ErbB-1
ErbB-1
ErbB-1
ErbB-2
(HER2/ Neu)
ErbB-2
ErbB-2
ErbB-3
ErbB-3
ErbB-2
ErbB-3
ErbB-4
EGF
TGFα
AR
BTC
HB-EGF
EPR
NRG1
NRG2
+p

Two Problems - Two Solutions

1. Large Partitioned Network
2. Many Node Attributes

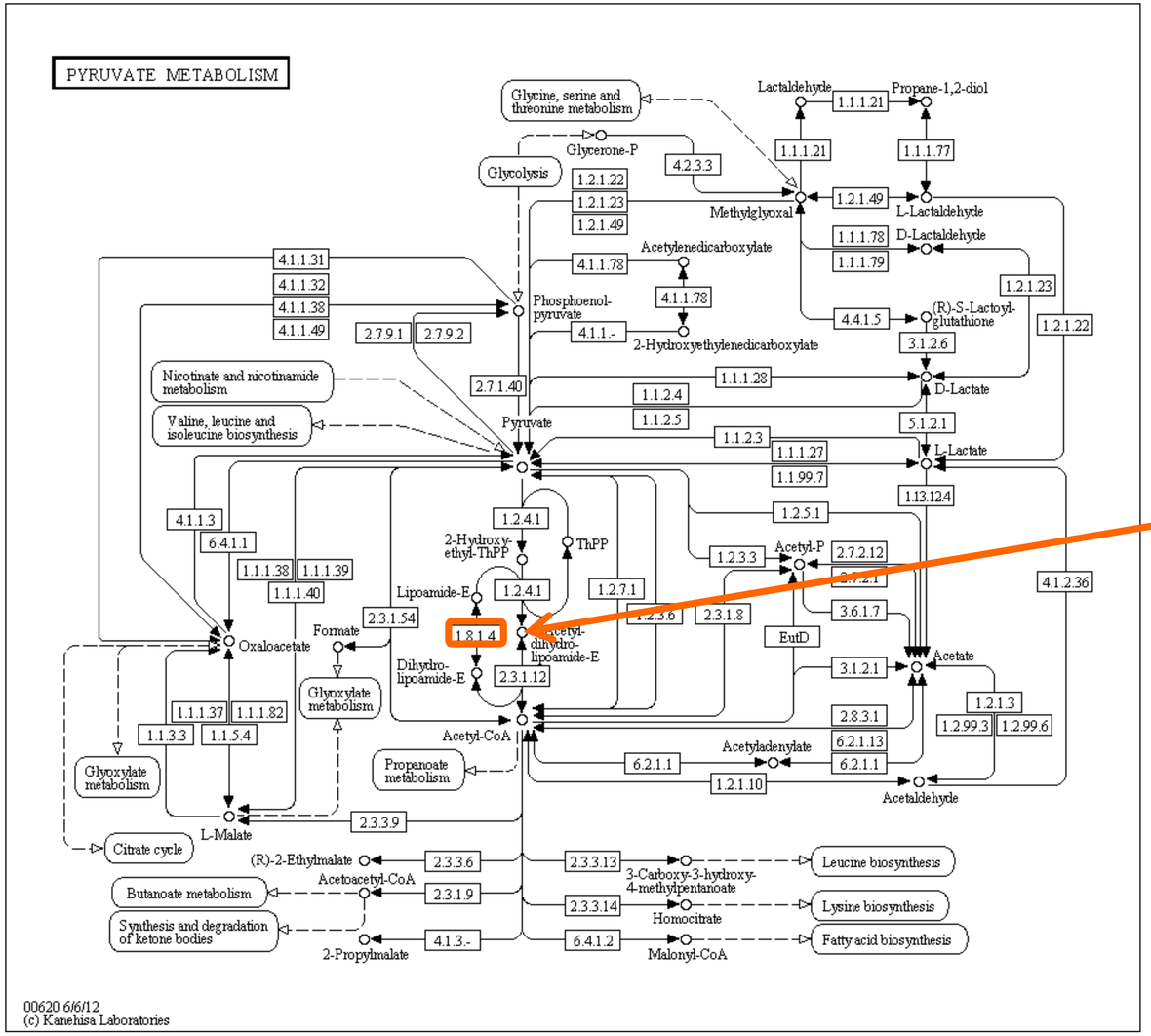
Large Partitioned Network

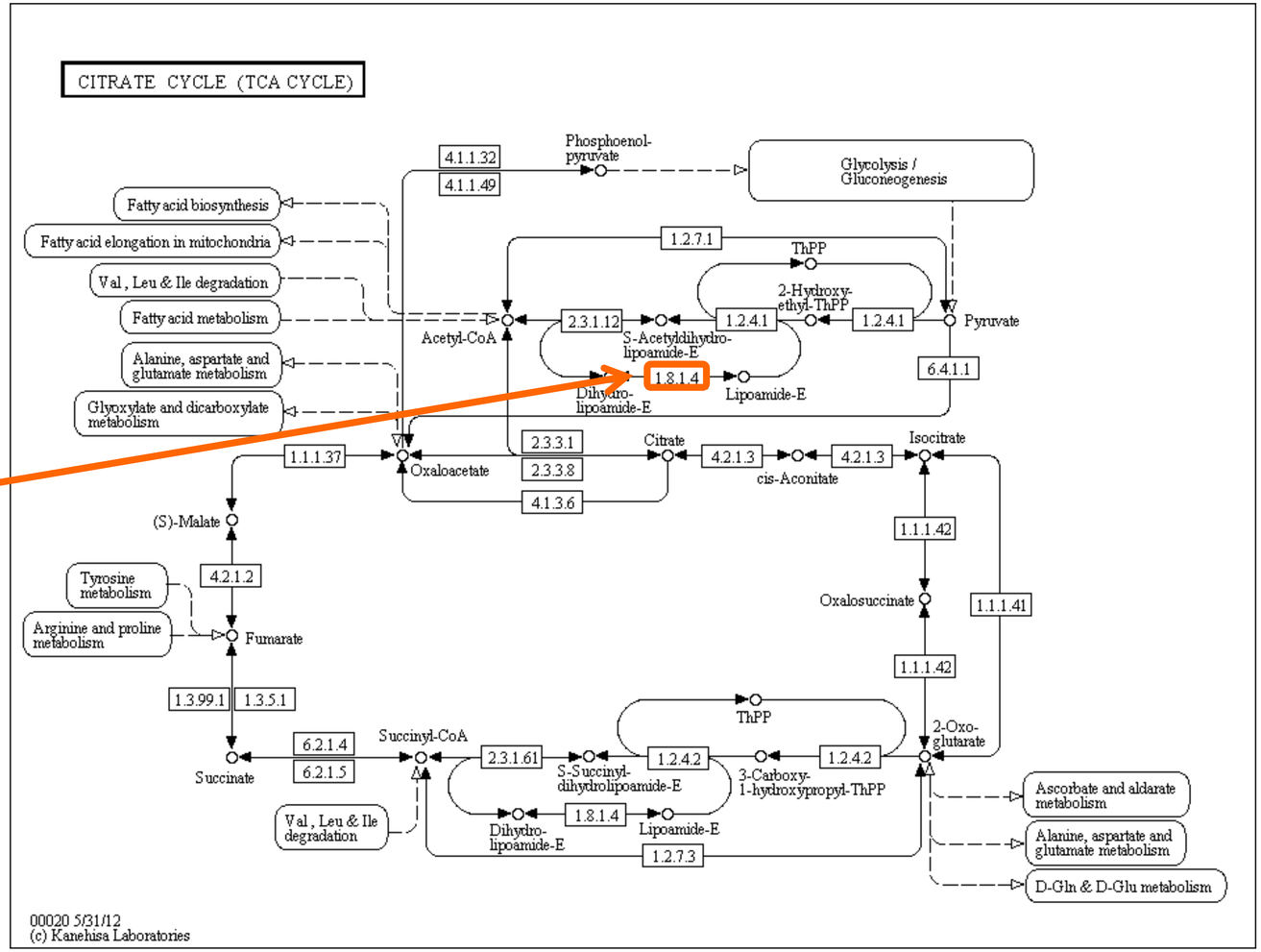


Drug side-effects

Drug repositioning

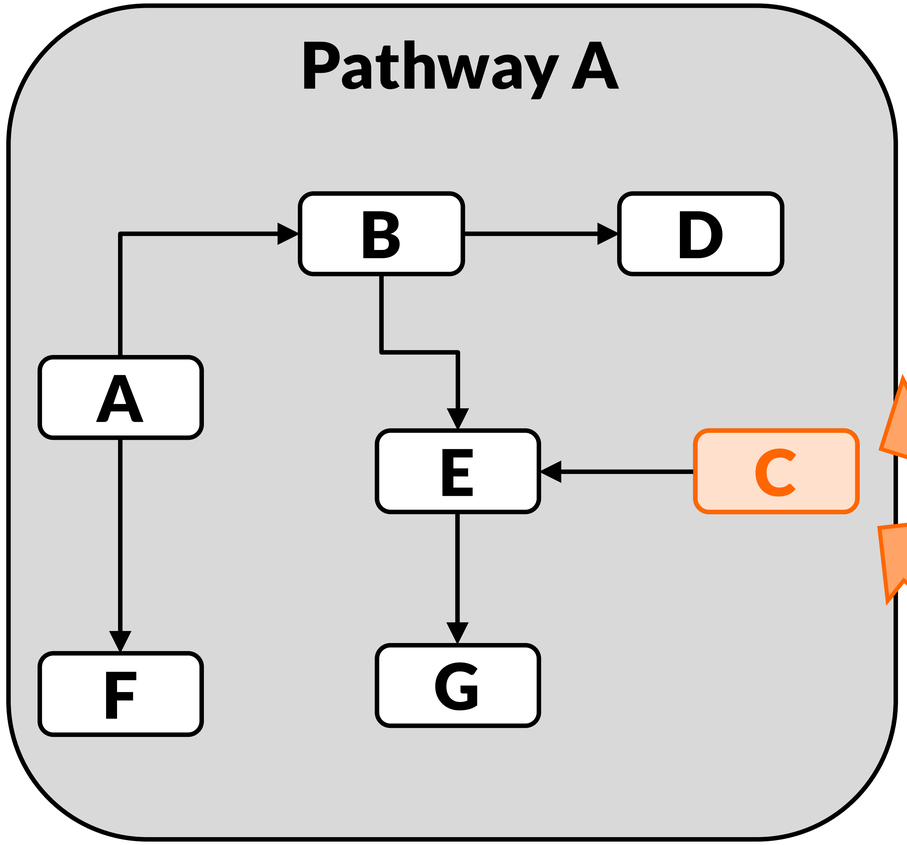
Large Partitioned Network





How to visualize pathway relationships?

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	...
...	...	...	...	

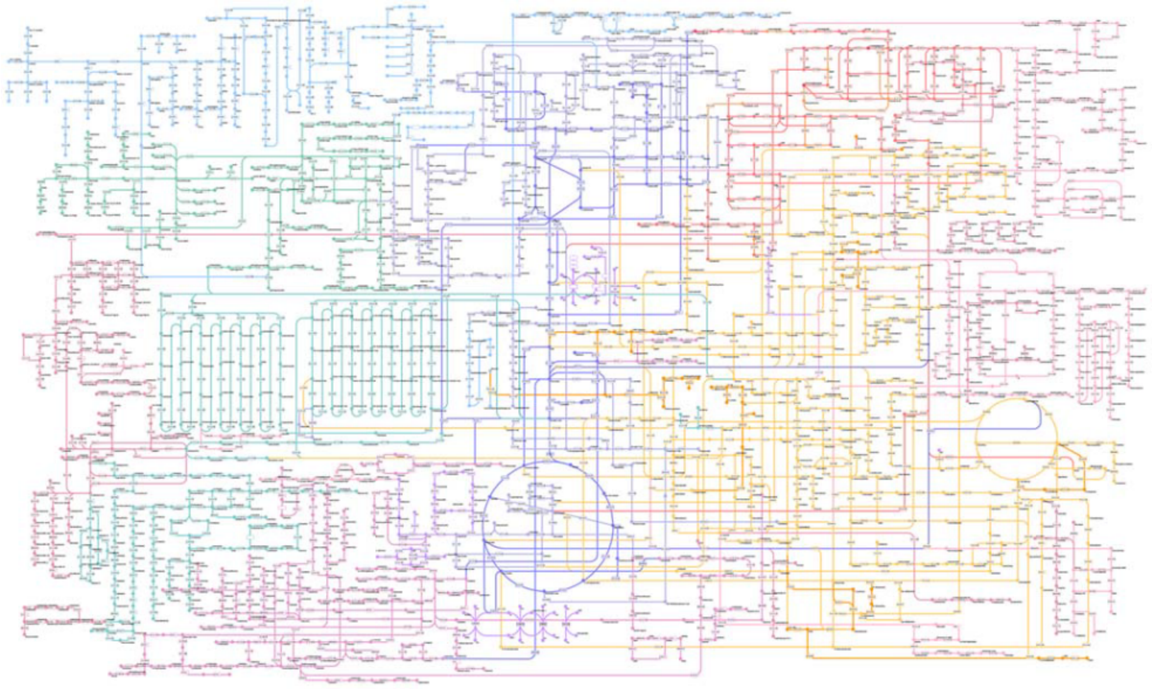
Node	Sample 1	Sample 2	Sample 3	...
A	low	low	very high	...
B	normal	low	high	...
C	high	very low	normal	...
...	...	...	...	

How to visualize experimental data on pathways?

Visualizing
Large Partitioned Networks

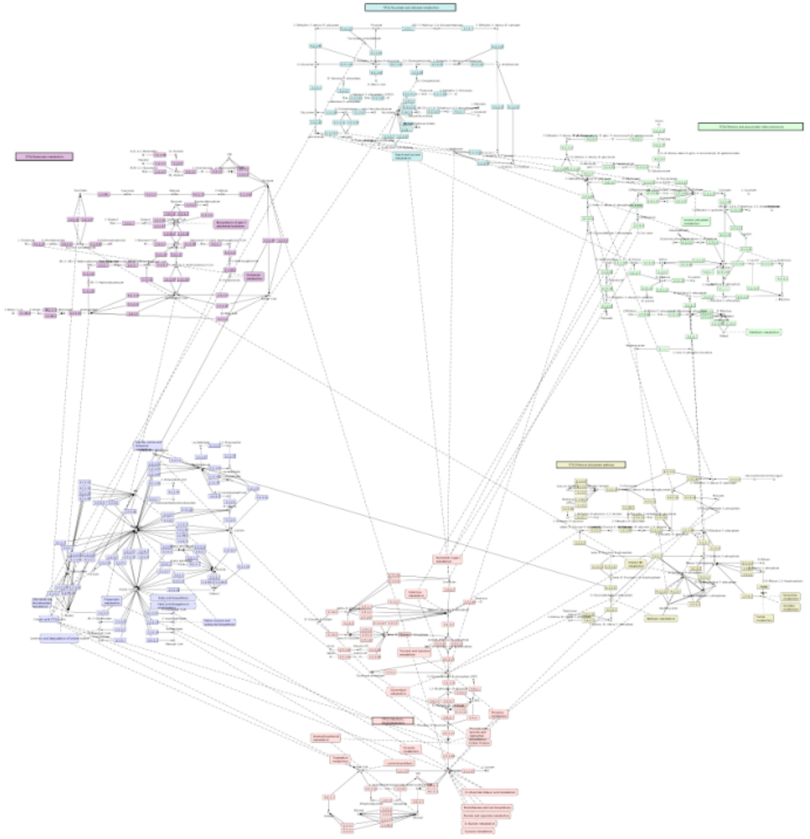
Approaches

Whole Network



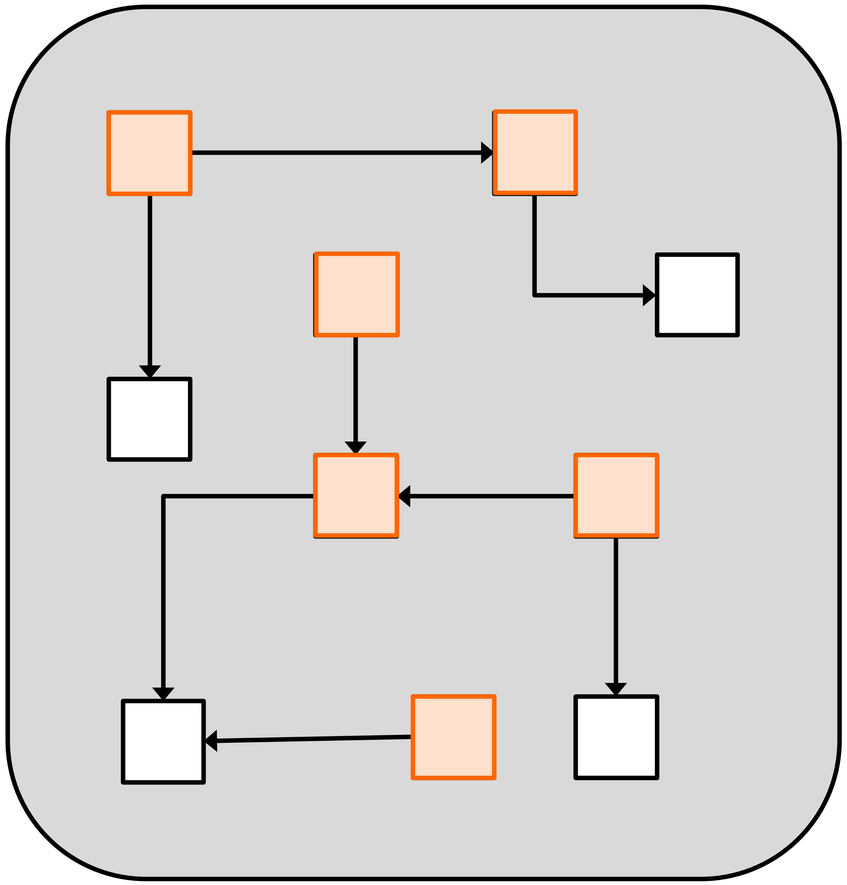
[Kono2009]

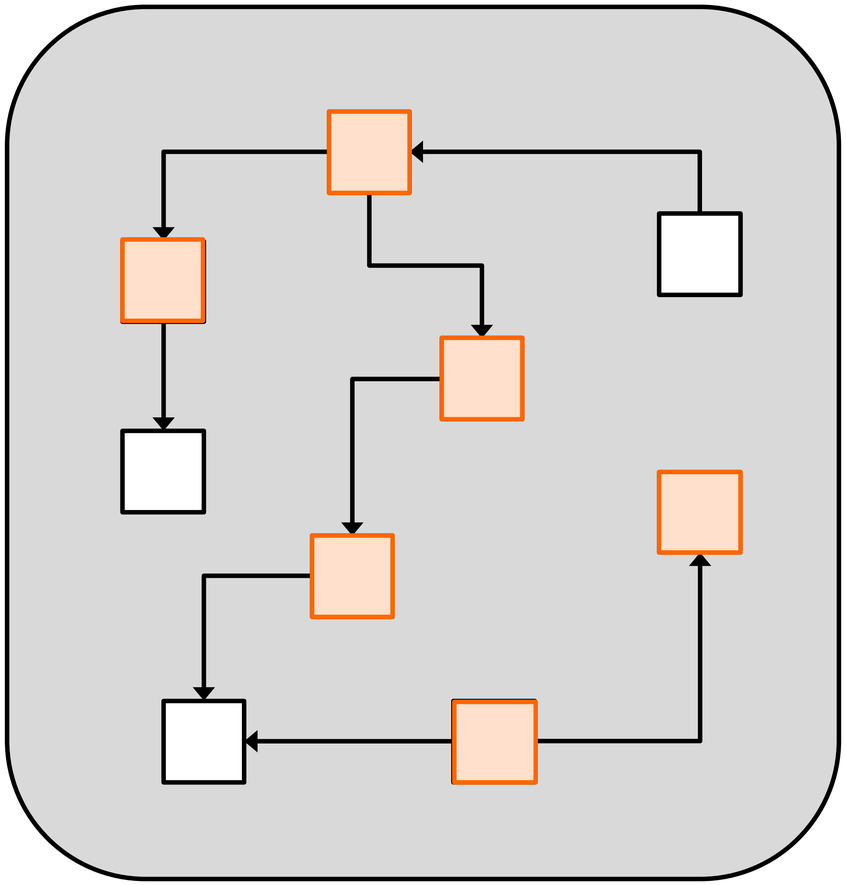
Connected Pathways



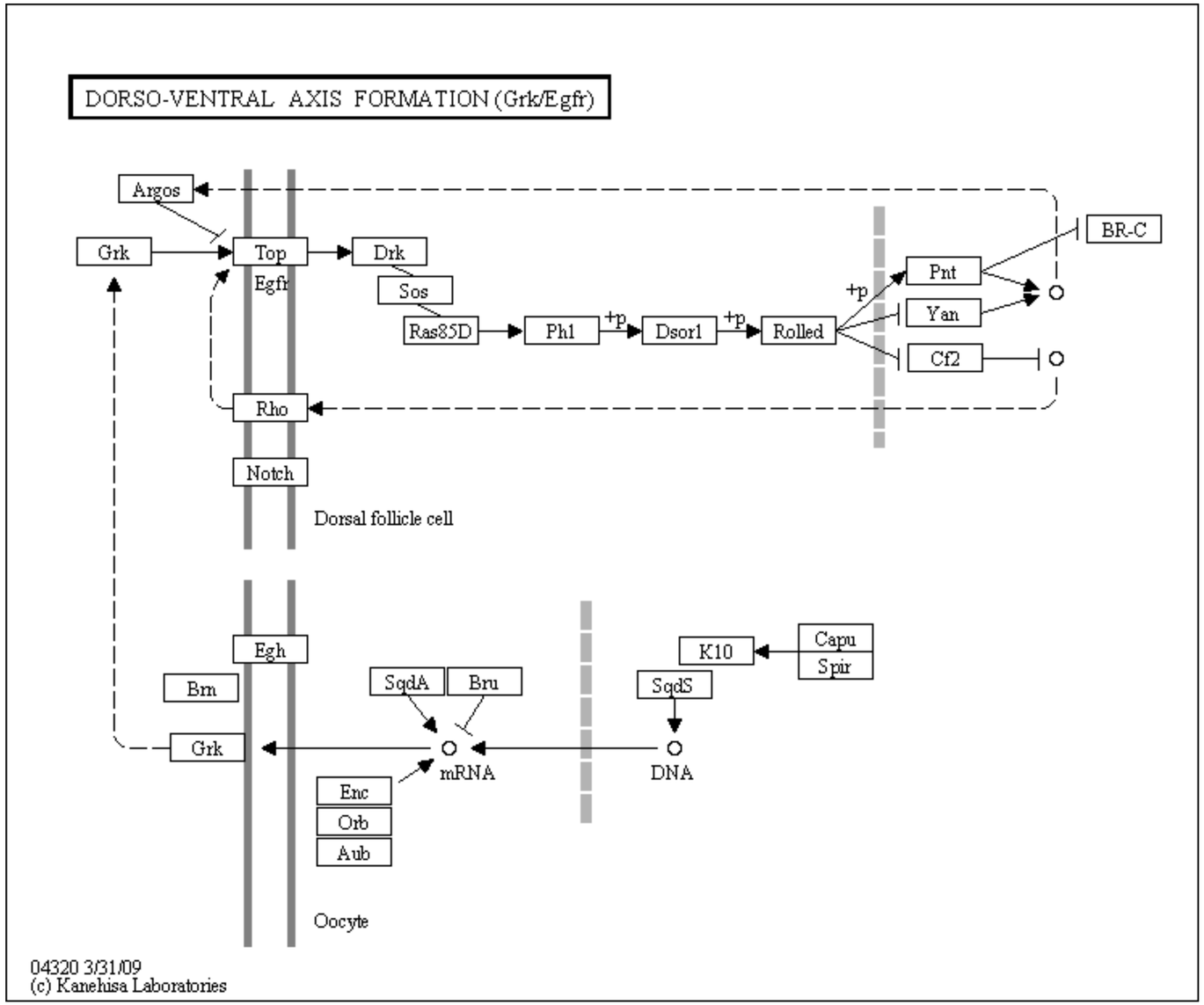
[Klukas2006]

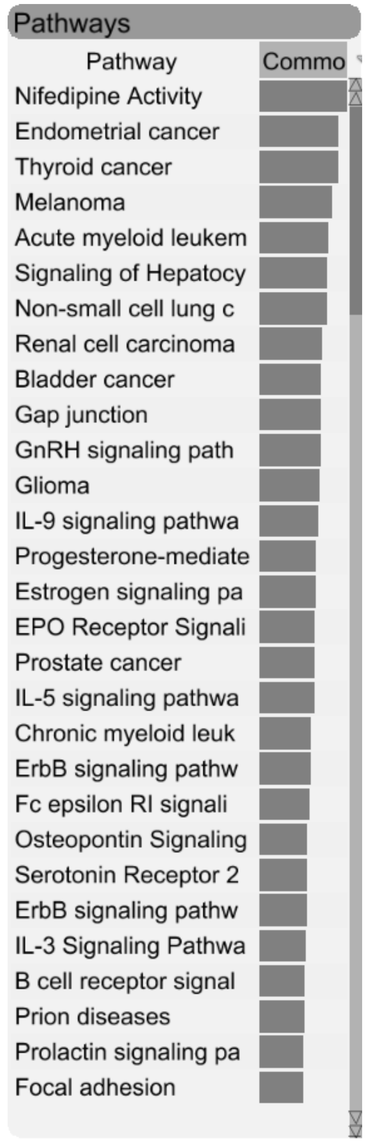
Step 1: Finding Related Pathways



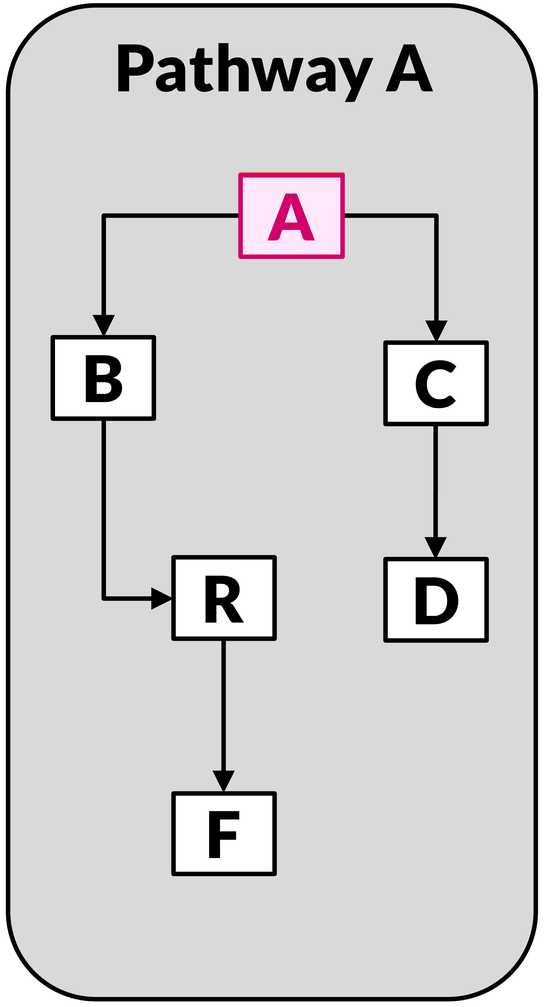


Finding Related Pathways

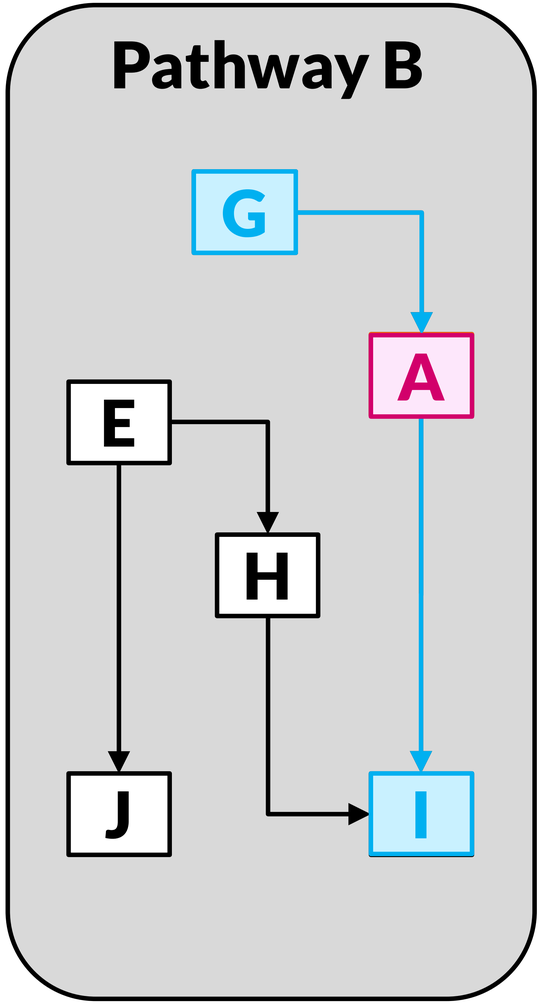




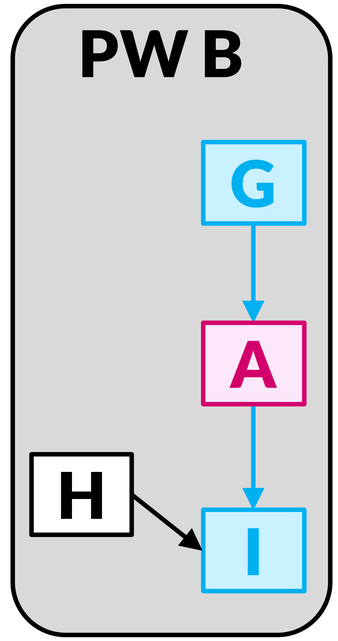
Step 2: Managing Display Space



Focus Pathway

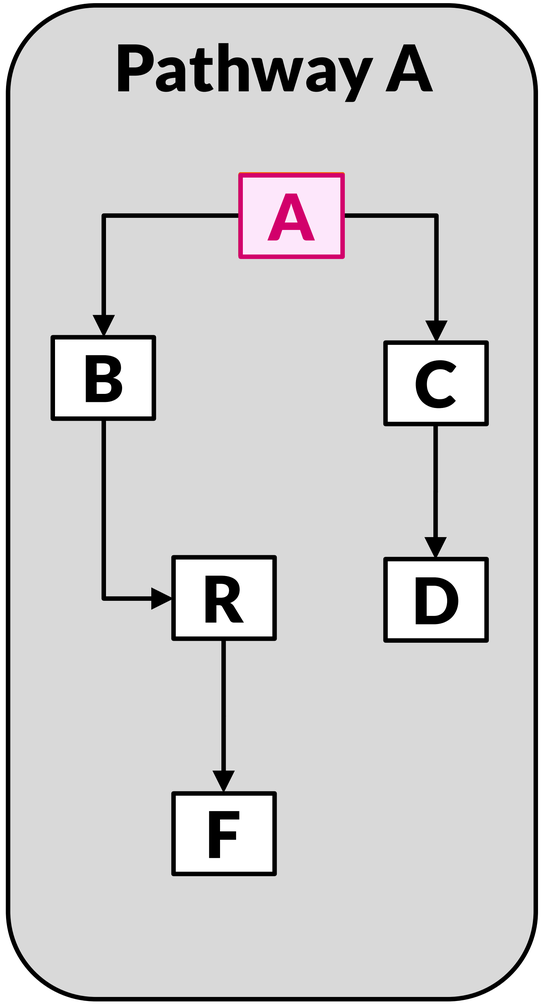


Context Pathway

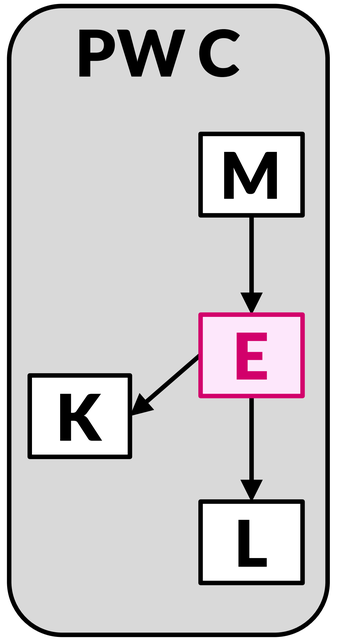


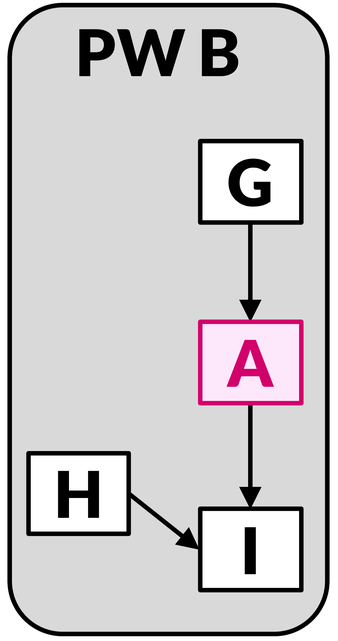
Contextual Subset

Step 2: Managing Display Space



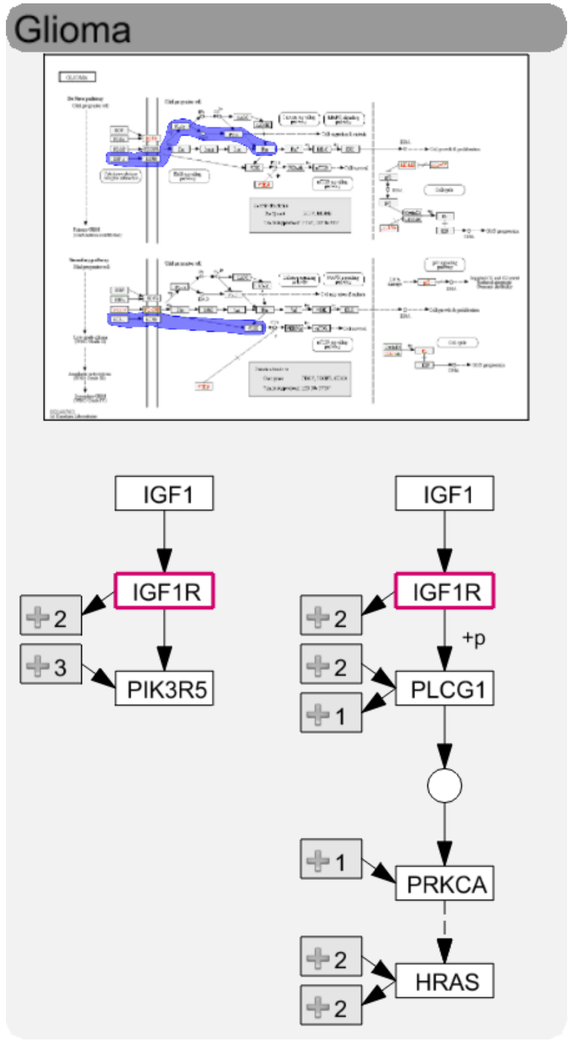
Focus Pathway



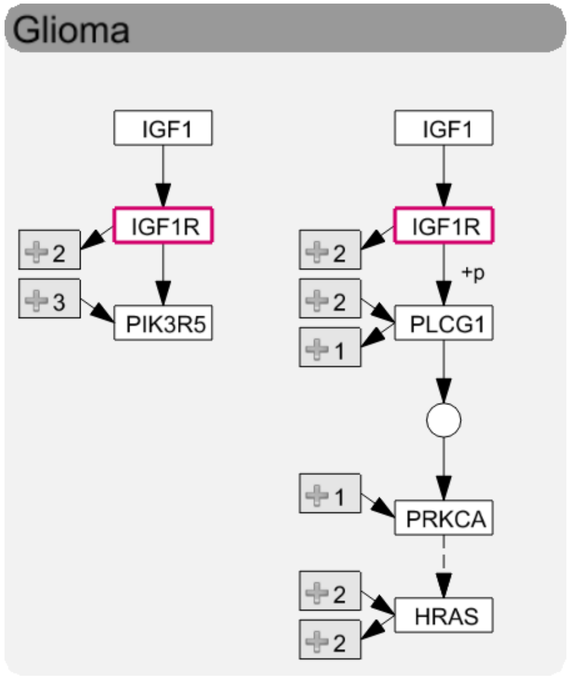


Context Pathways

Levels of Detail



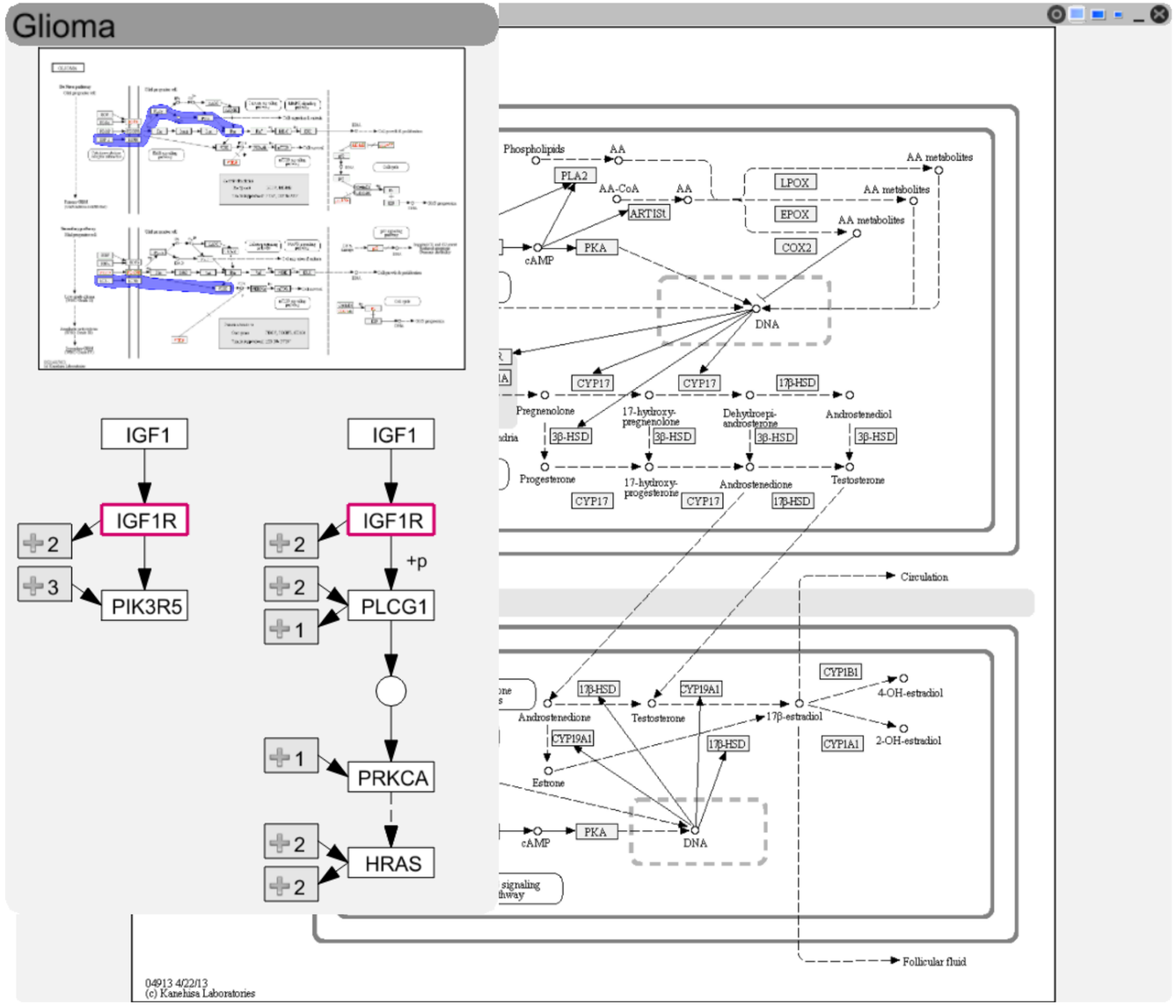
High

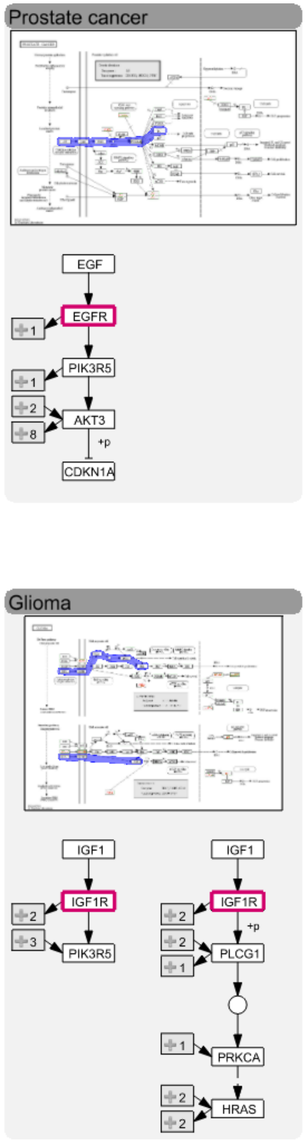


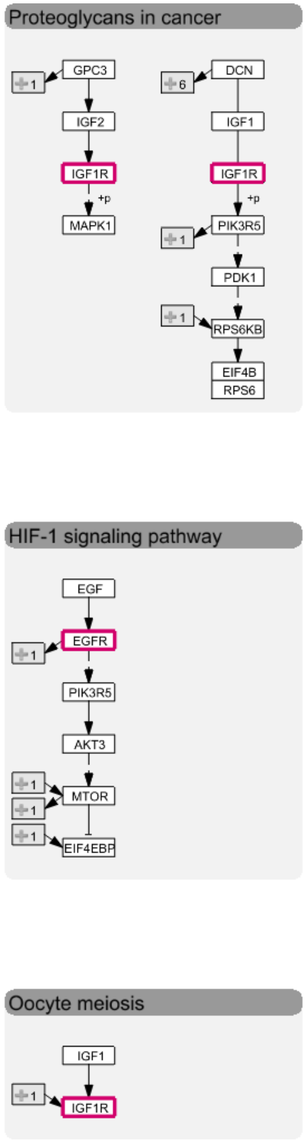
Medium

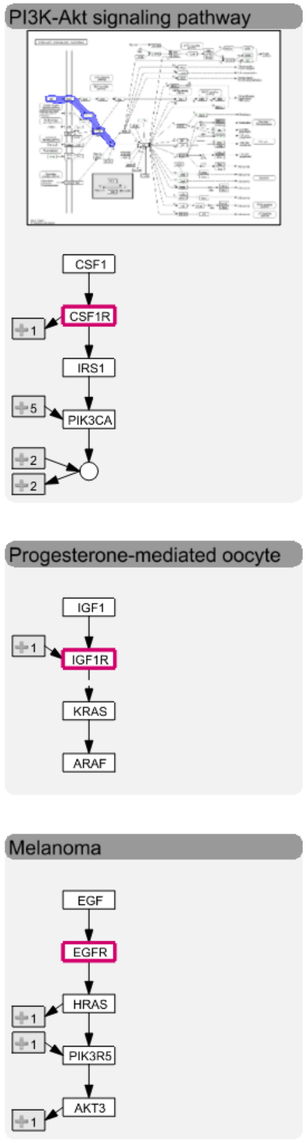
Low

Layout

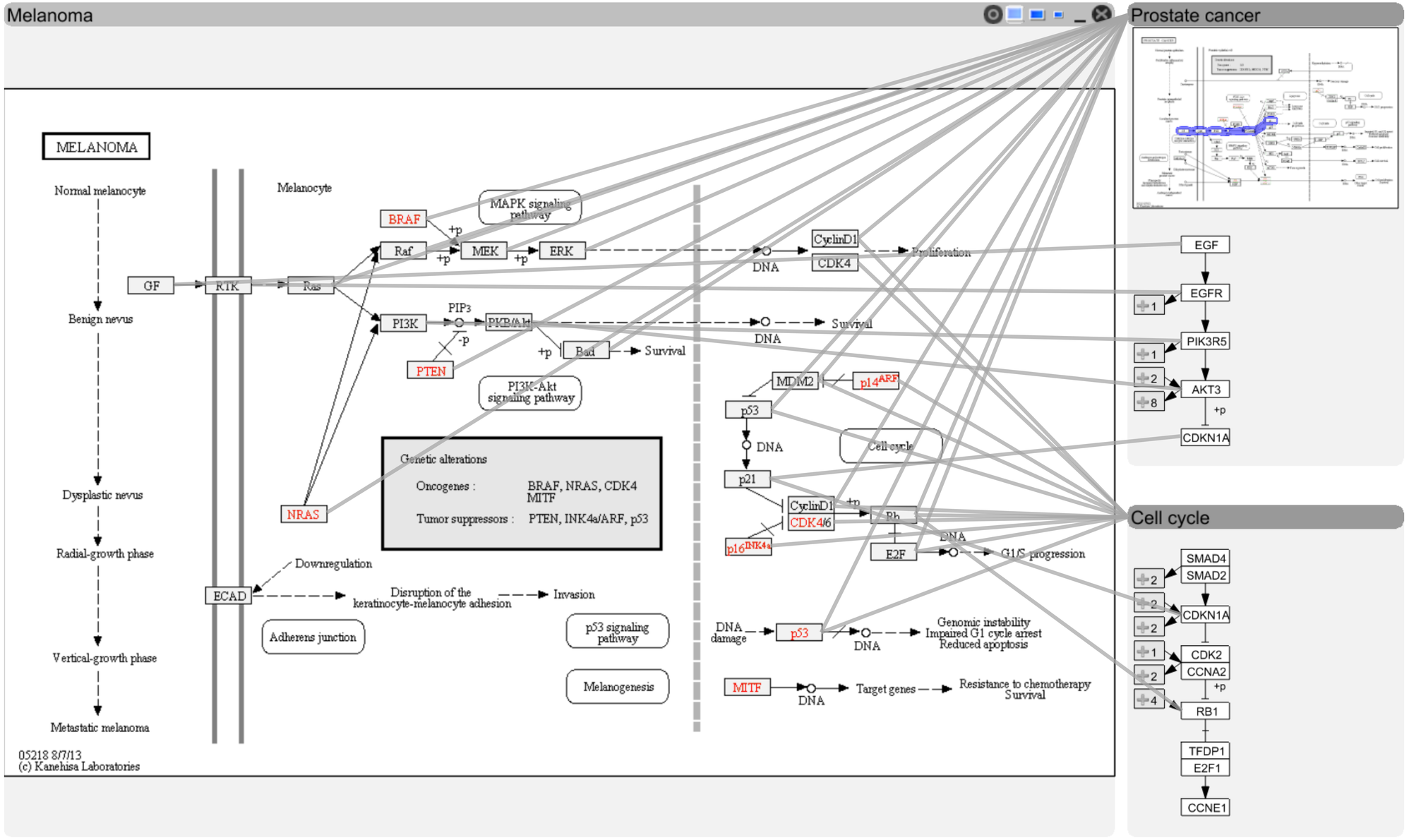

Glioma
IGF1
IGF1R
2
3
PIK3R5
IGF1
IGF1R
+p
2
2
1
PLCG1
1
PRKCA
2
2
HRAS
Phospholipids
AA
PLA2
AA-CoA
AA
ARTISt
LPOX
EPOX
COX2
AA metabolites
AA metabolites
AA metabolites
cAMP
PKA
DNA
CYP17
CYP17
17β-HSD
Pregnenolone
17-hydroxy-pregnenolone
Dehydroepi-androsterone
Androstenediol
3β-HSD
3β-HSD
3β-HSD
3β-HSD
Progesterone
17-hydroxy-progesterone
Androstenedione
Testosterone
CYP17
CYP17
17β-HSD
Circulation
17β-HSD
CYP19A1
CYP1B1
4-OH-estradiol
Androstenedione
Testosterone
17β-estradiol
CYP19A1
17β-HSD
CYP1A1
2-OH-estradiol
Estrone
cAMP
PKA
DNA
signaling pathway
Follicular fluid
04913 4/22/13
(c) Kanehisa Laboratories


Prostate cancer
EGF
1
EGFR
1
PIK3R5
2
8
AKT3
+p
CDKN1A
Glioma
IGF1
2
3
IGF1R
PIK3R5
IGF1
2
2
1
IGF1R
+p
PLCG1
1
PRKCA
2
2
HRAS

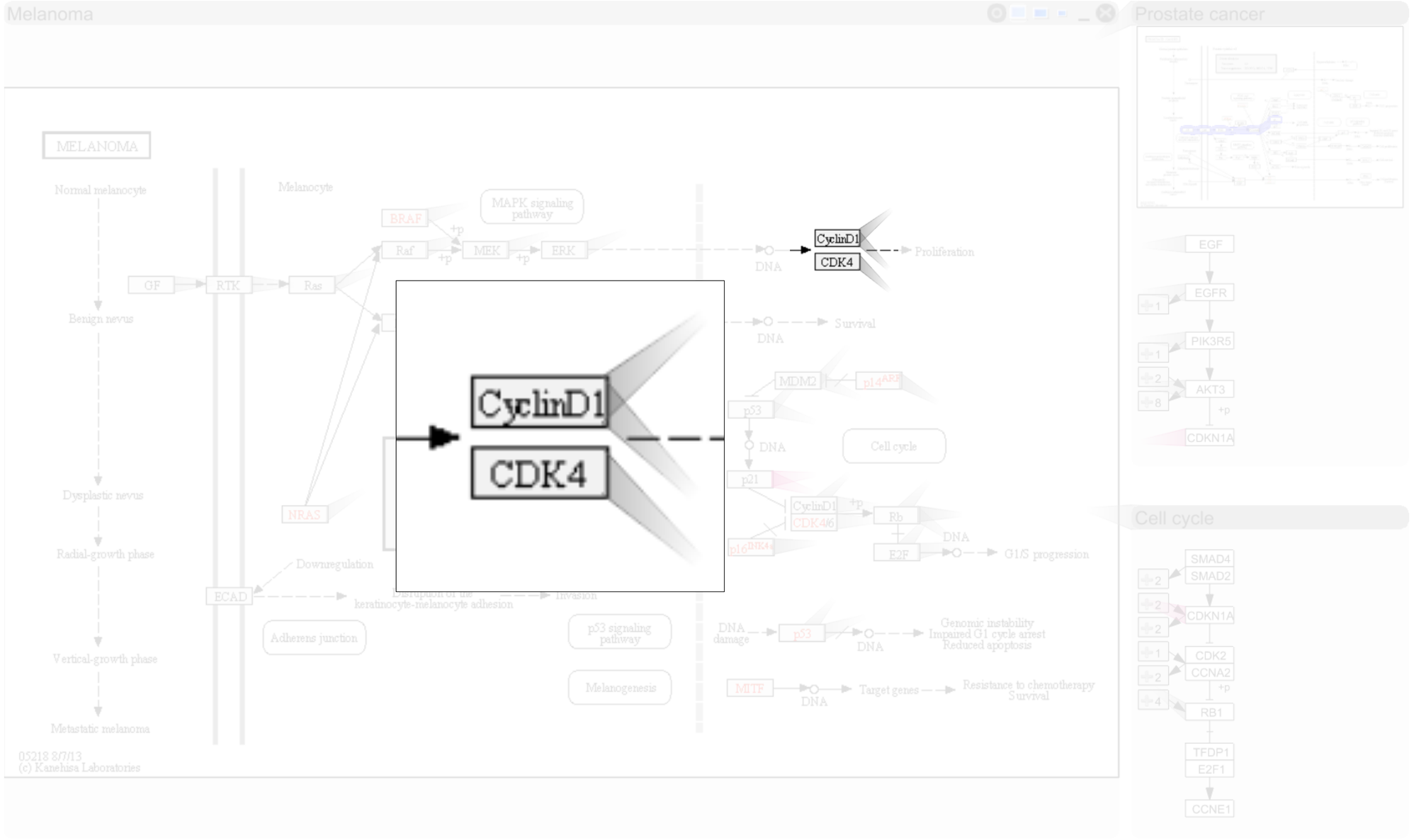

Proteoglycans in cancer
1
GPC3
IGF2
IGF1R
+p
MAPK1
6
DCN
IGF1
IGF1R
+p
1
PIK3R5
PDK1
1
RPS6KB
EIF4B
RPS6
HIF-1 signaling pathway
EGF
1
EGFR
PIK3R5
AKT3
1
1
1
MTOR
EIF4EBP
Oocyte meiosis
IGF1
1
IGF1R


PI3K-Akt signaling pathway
CSF1
1
CSF1R
IRS1
5
PIK3CA
2
2
Progesterone-mediated oocyte
IGF1
1
IGF1R
KRAS
ARAF
Melanoma
EGF
EGFR
1
HRAS
1
PIK3R5
1
AKT3

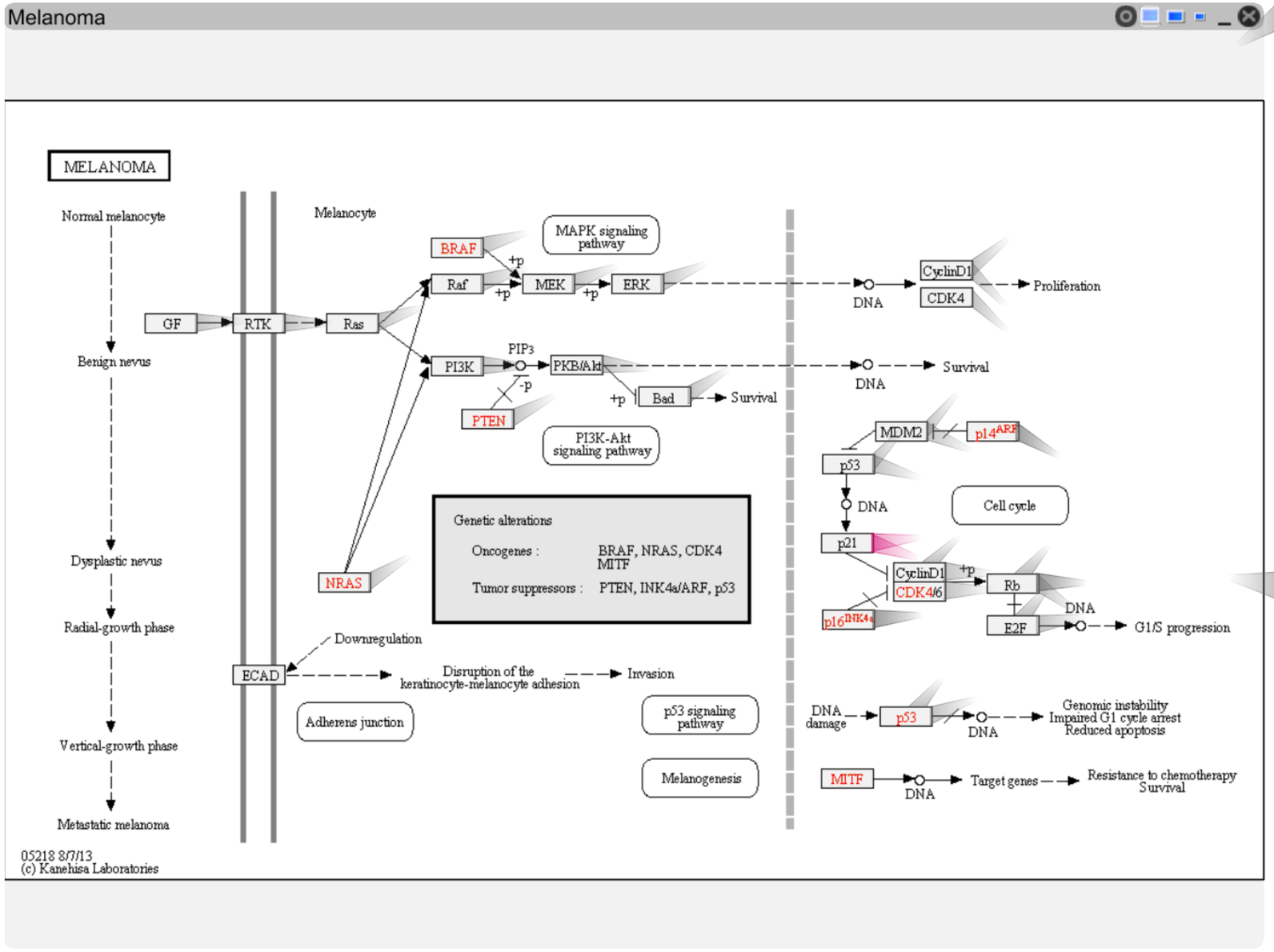
Step 3: Visualizing Relationships

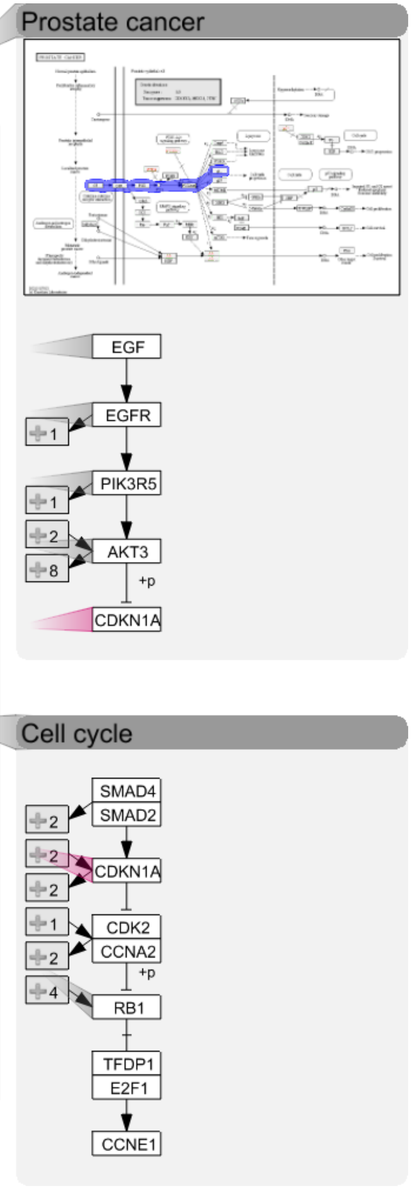


Step 3: Visualizing Relationships

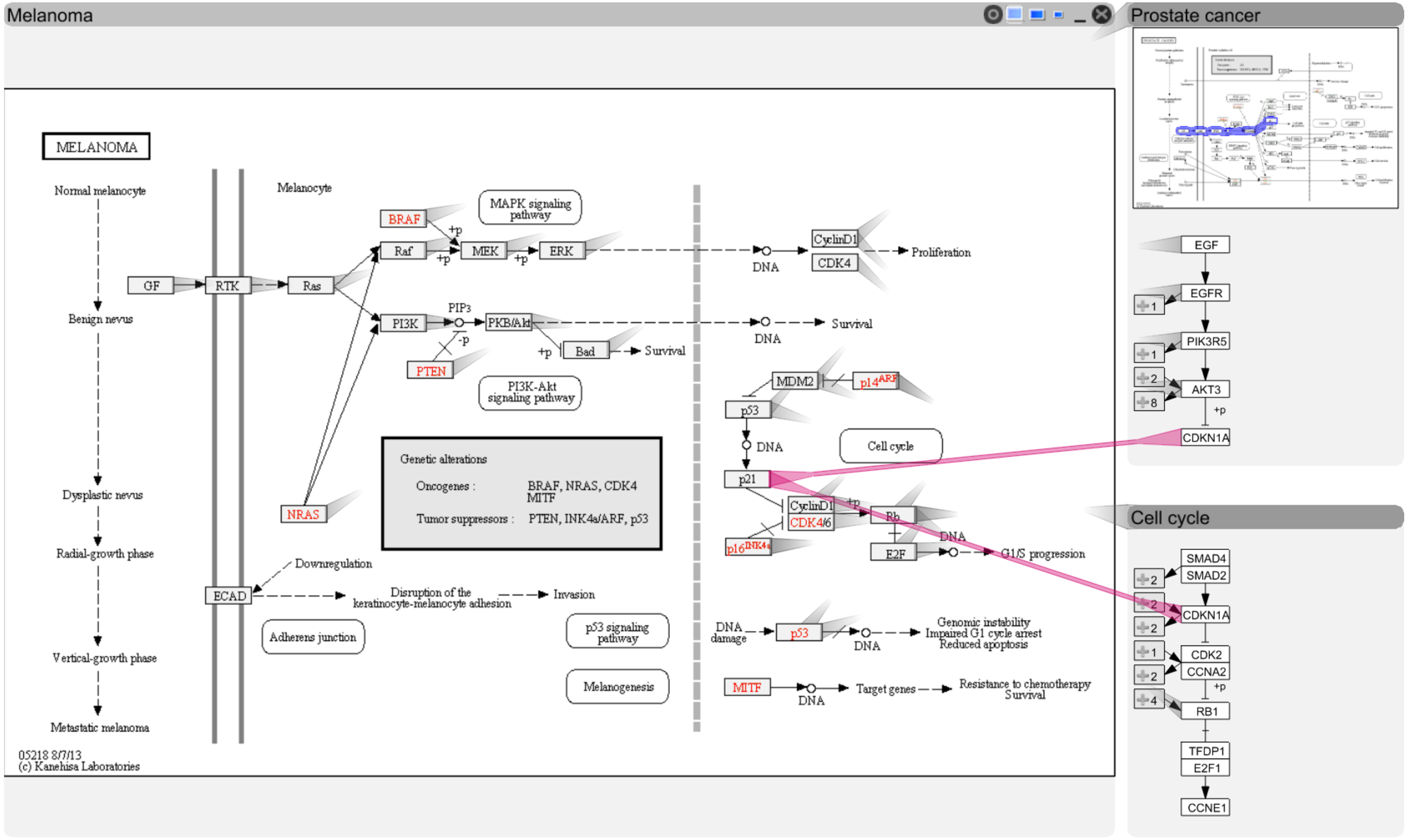


Step 3: Visualizing Relationships

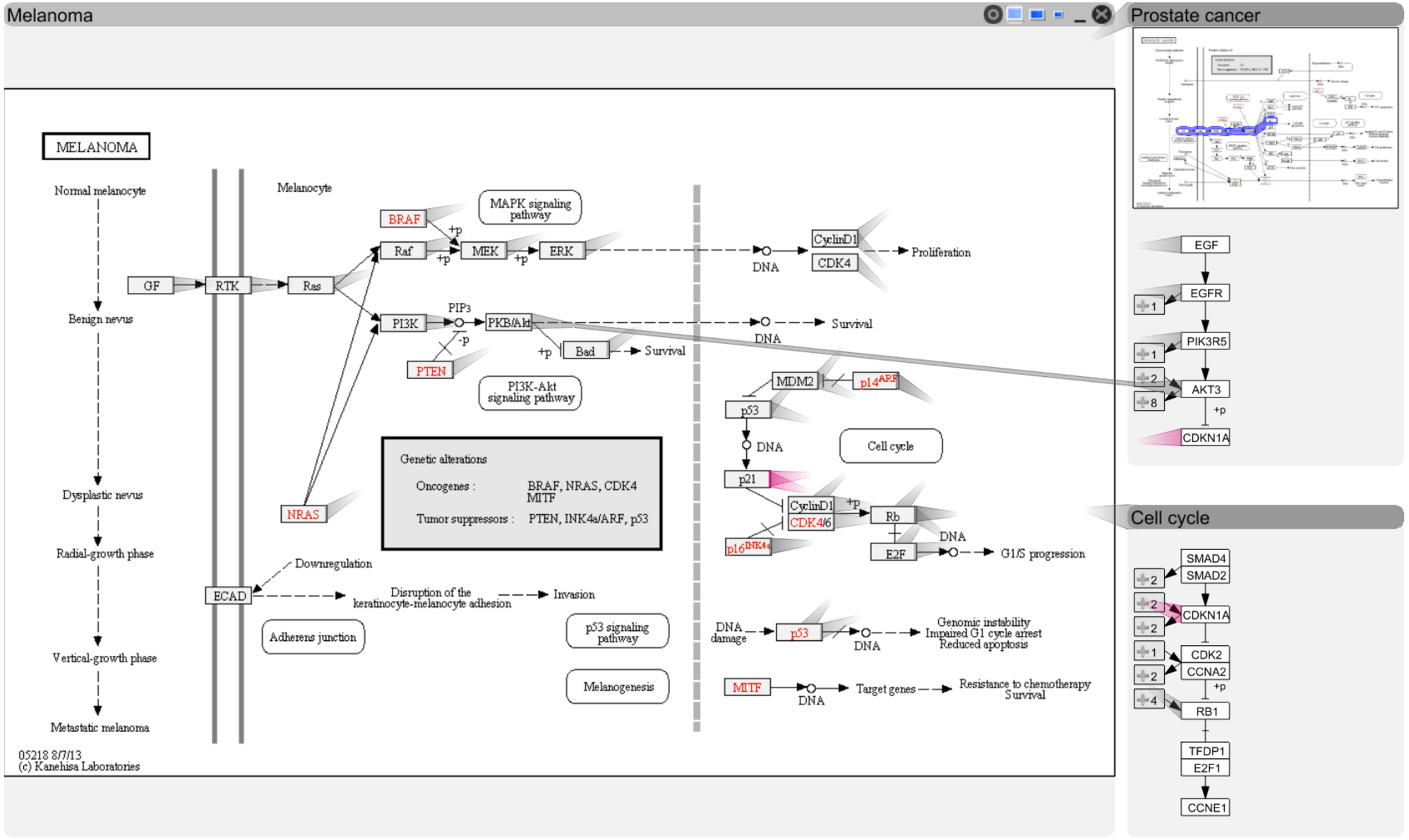




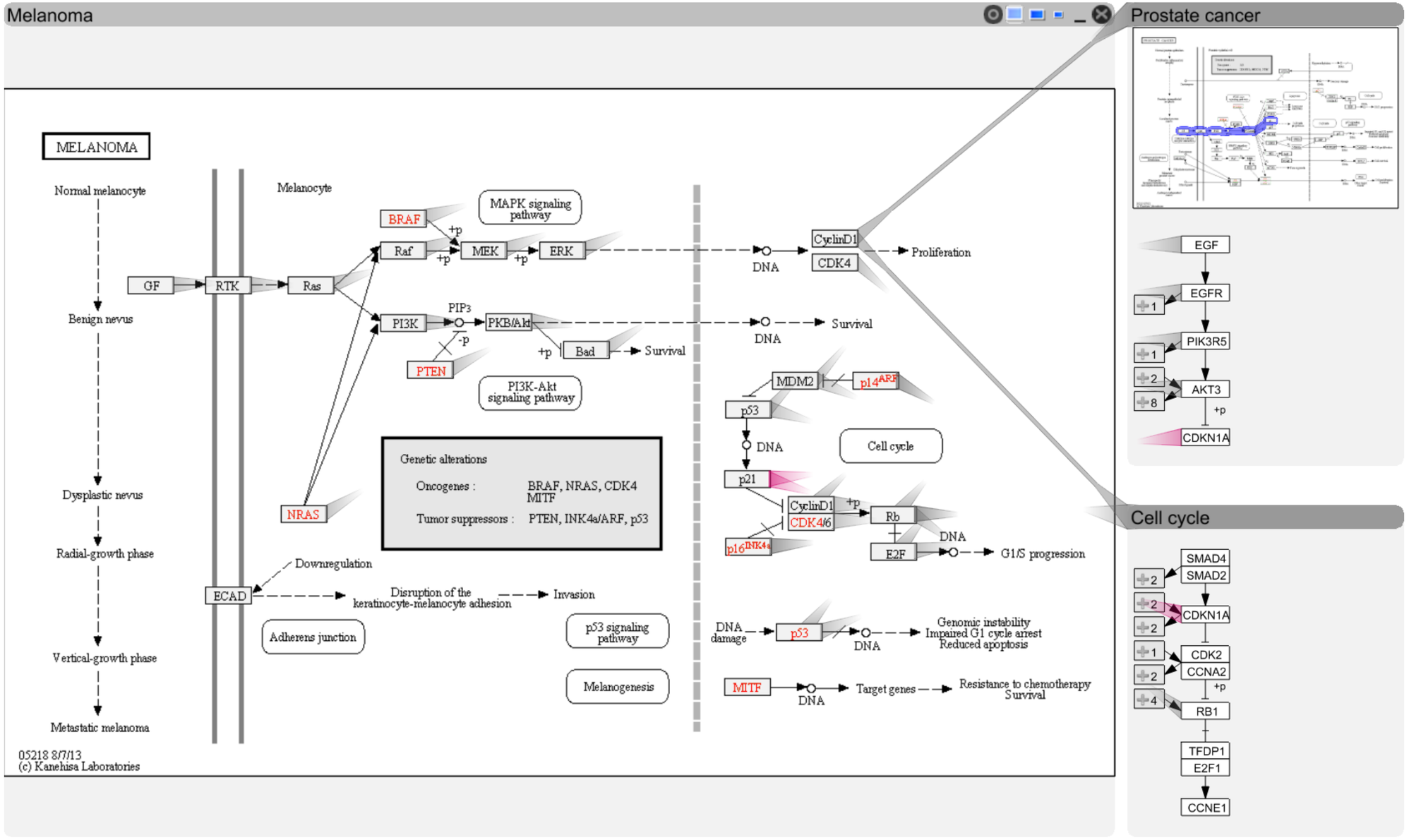
Step 3: Visualizing Relationships



Step 3: Visualizing Relationships



Step 3: Visualizing Relationships



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

Visualizing
Many Node Attributes

Experimental Data and Pathways

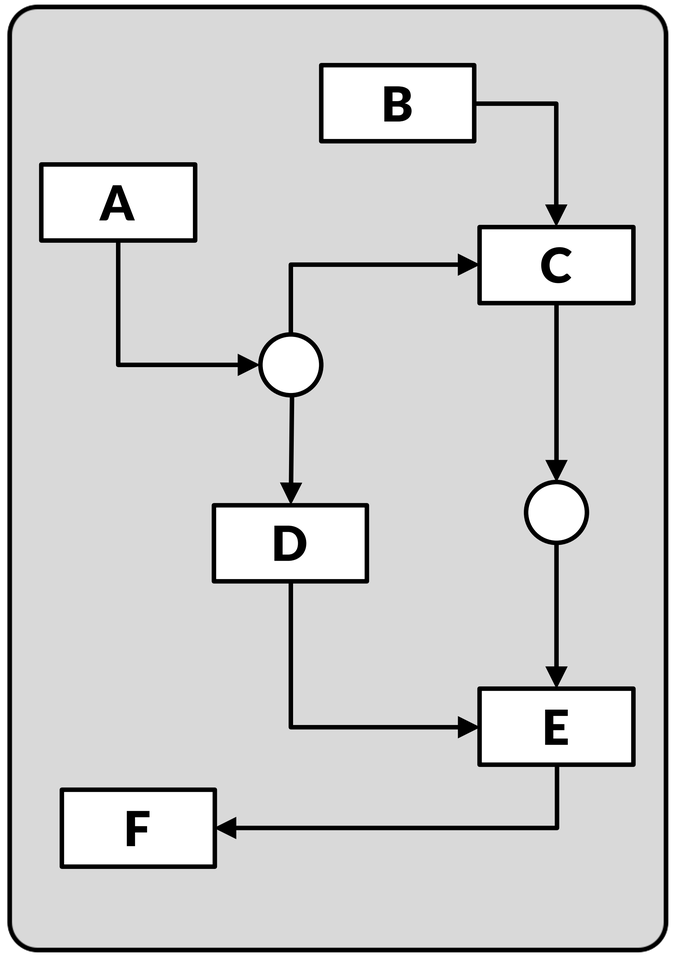
[Partl, BioVis ‘12]

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.

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: Data Scale & Heterogeneity

Large number of experiments

- Large datasets have more than 500 experiments

Multiple groups/conditions

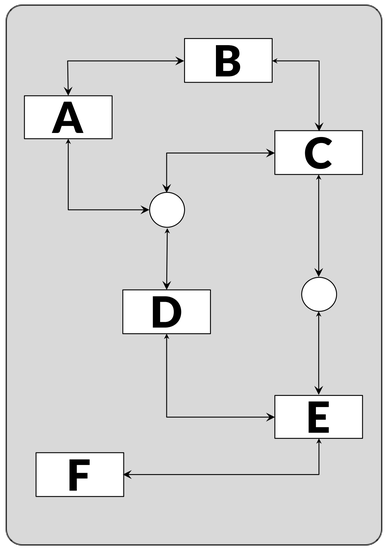
Different types of data

Challenge: Supporting Multiple Tasks

Two central tasks:

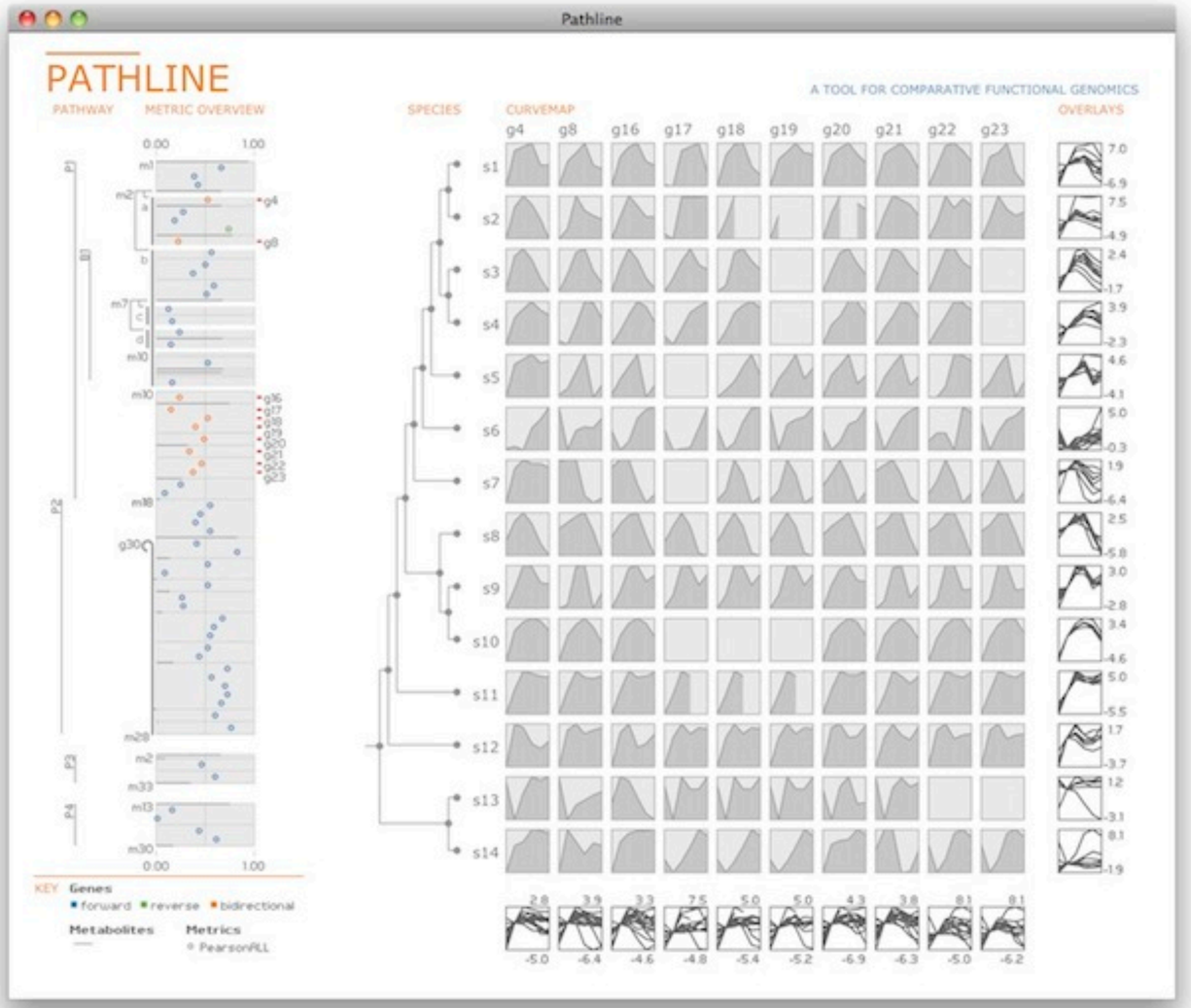
- Explore topology of pathway
- Explore the attributes of the nodes (experimental data)

Need to support both!



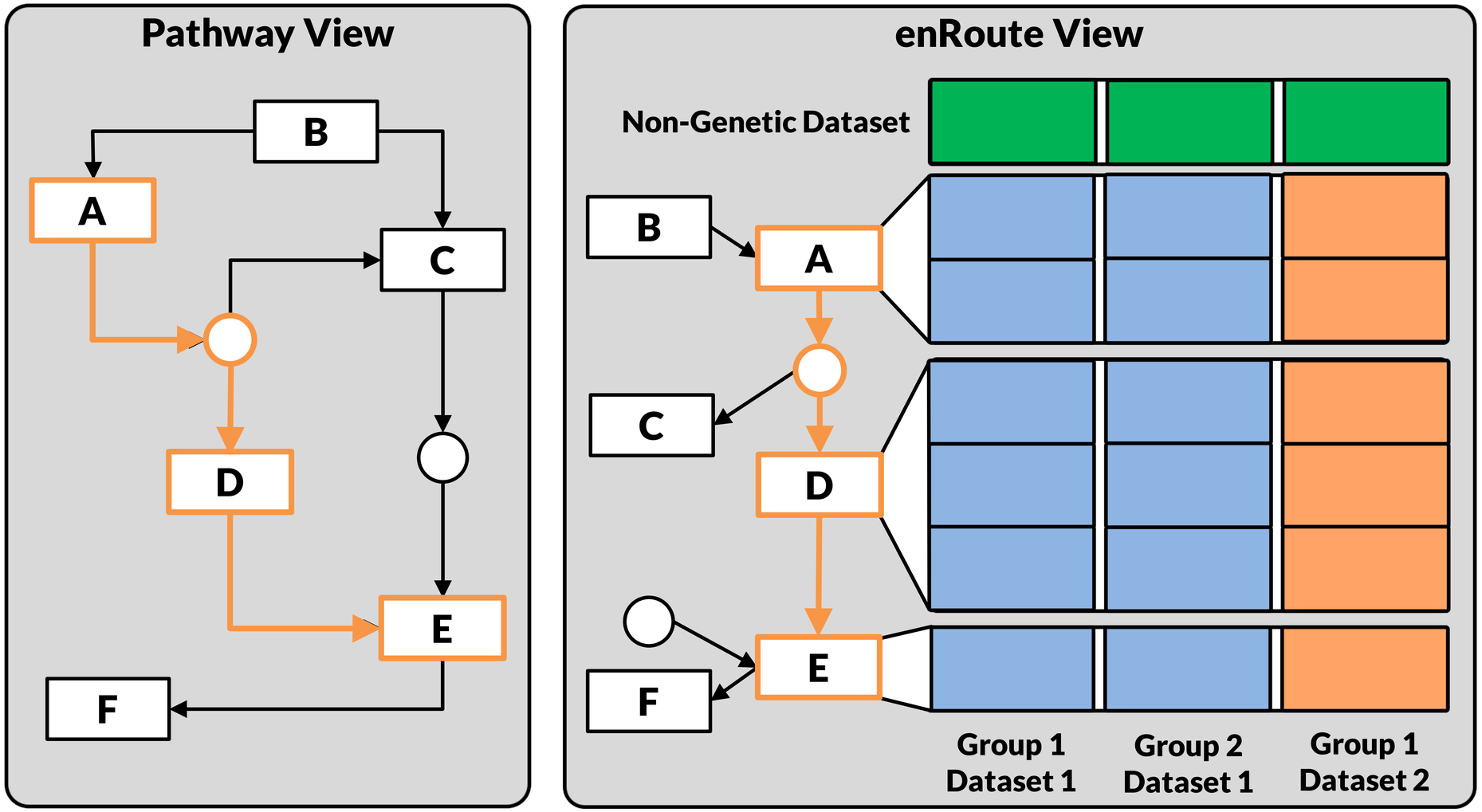
	Sample 1	Sample 2	Sample 3
Gene 1	1	1.1	0.4
Gene 2	2	0.5	1.2
Gene 3	1.4	0.2	0.5
Gene 4	0.3	0.5	0.7

Inspiration

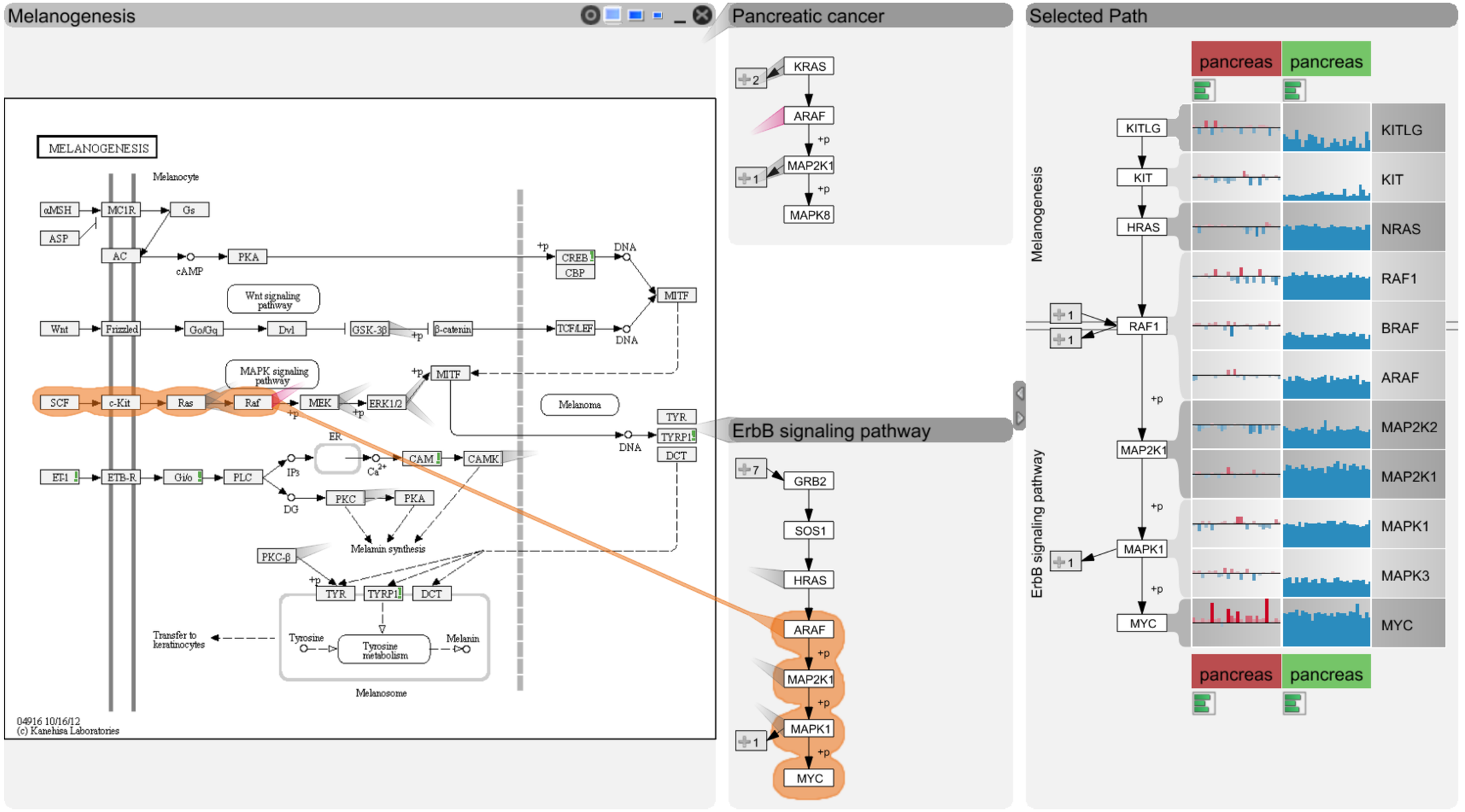


[Meyer 2010]

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
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
ET1
ETB-R
Gi/o
PLC
IP3
Ca2+
CAM
CAMK
DG
PKC
PKA
Melamin synthesis
PKC-β
Transfer to keratinocytes
Tyrosine
Tyrosine metabolism
Melanin
Melanosome
04916 10/16/12
(c) Kanehisa Laboratories
Pancreatic cancer
KRAS
ARAF
MAP2K1
MAPK8
ErbB signaling pathway
GRB2
SOS1
HRAS
ARAF
MAP2K1
MAPK1
MYC
Selected Path
pancreas
pancreas
Melanogenesis
ErbB signaling pathway
KITLG
KIT
HRAS
RAF1
MAP2K1
MAPK1
MYC
NRAS
BRAF
ARAF
MAP2K2
MAP2K1
MAPK1
MAPK3
MYC
pancreas
pancreas

File Data Window View Help

Entourage

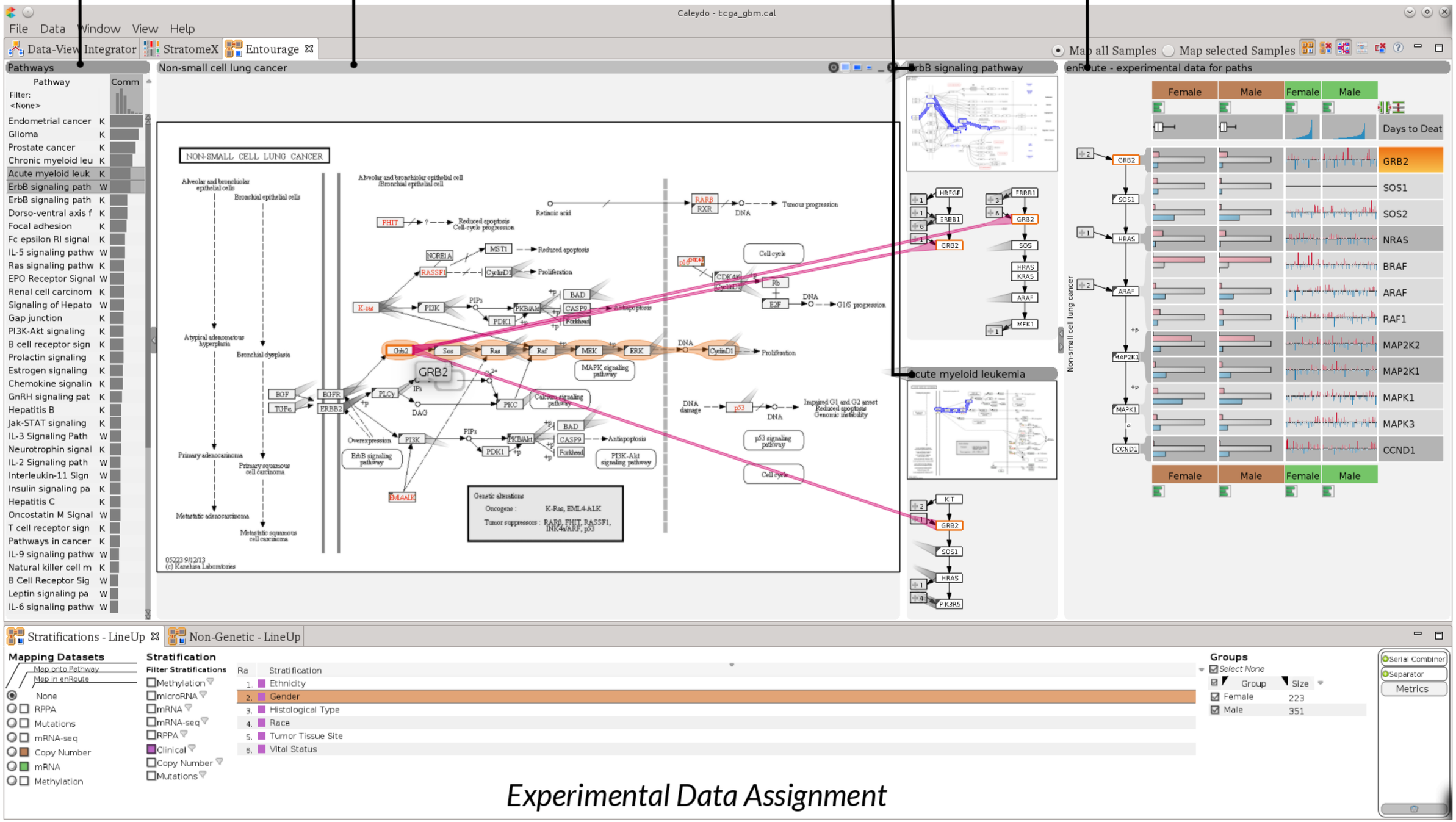
Pathways

Pathway

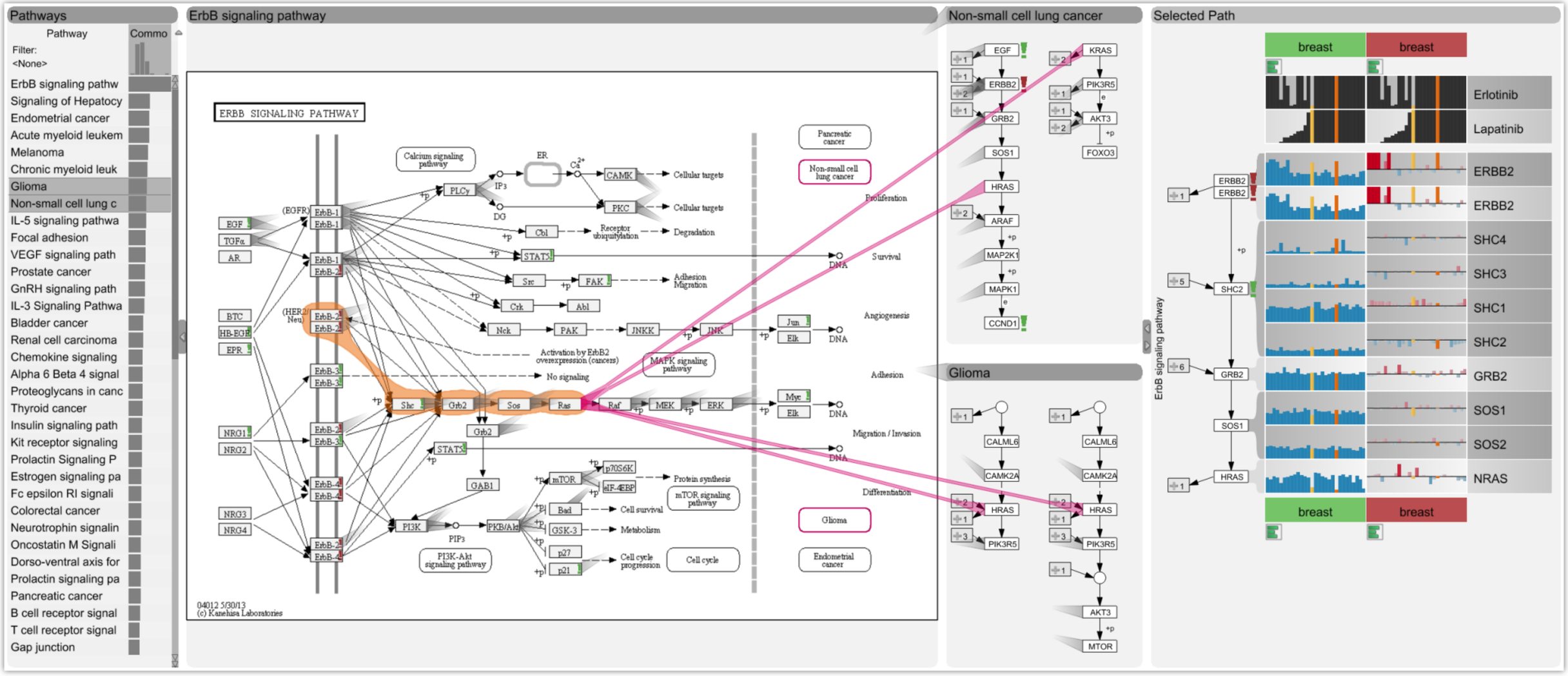
Filter:
<None>

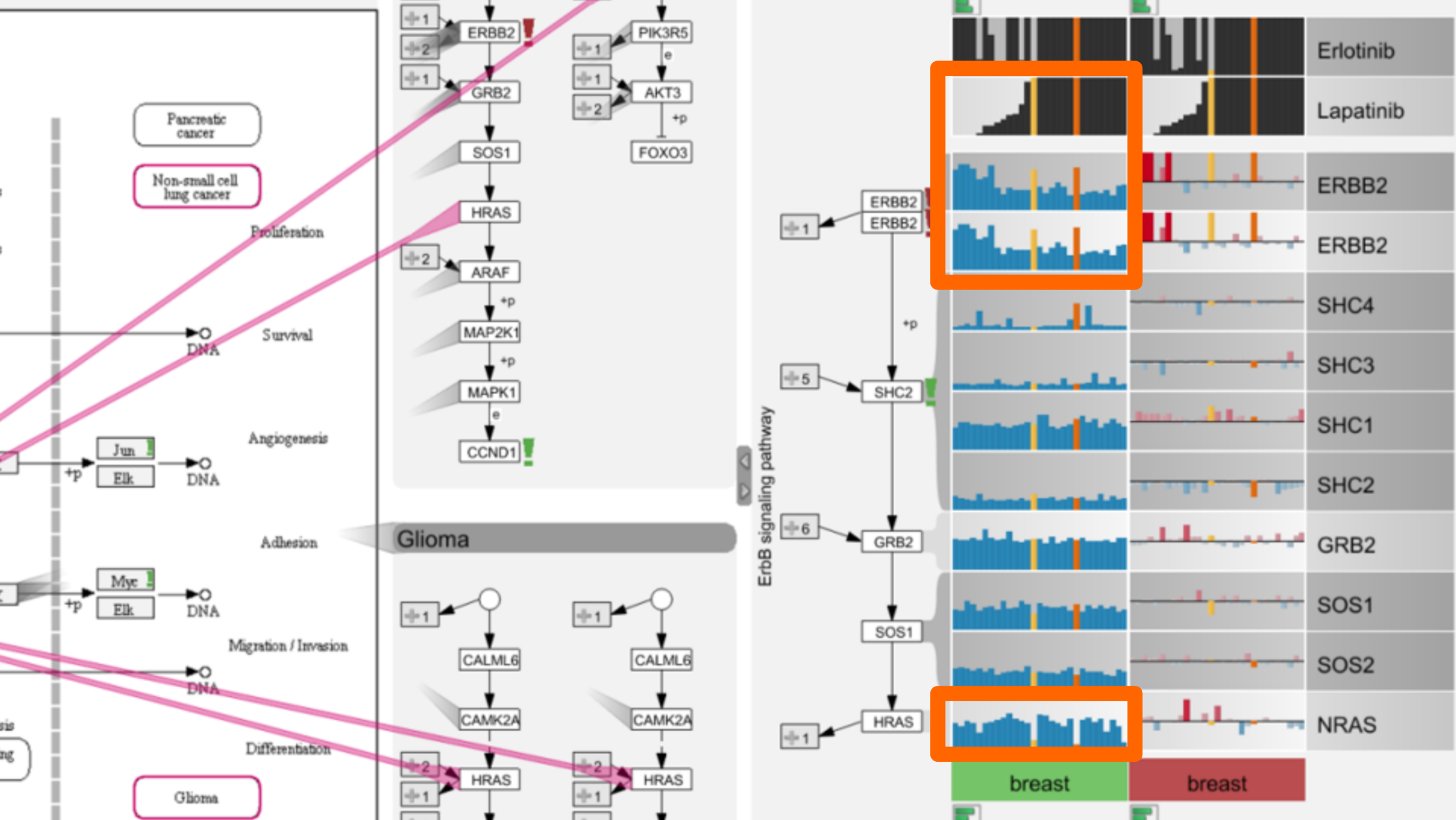
1 C donor
2-Oxocarboxylic acid
ABC transporters
ABC-family proteins
ACE Inhibitor Pathwa
Acetylcholine Synthes
Acute myeloid leukem
Adherens junction
Adipocyte TarBase
Adipocytokine signali
Adipogenesis
Advanced glycosylatio
Aflatoxin B1 metaboli
African trypanosomias
AGE/RAGE pathway
AhR pathway
Alanine 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


Pathway List
Focus Pathway
Context Pathways
enRoute View
Caleydo - tcga_gbm.cal
File Data Window View Help
Data-View Integrator
StratomeX
Entourage
Map all Samples
Map selected Samples
Pathways
Pathway
Comm
Filter:
<None>
Endometrial cancer K
Glioma K
Prostate cancer K
Chronic myeloid leu K
Acute myeloid leuk K
ErbB signaling path W
ErbB signaling path K
Dorso-ventral axis f K
Focal adhesion K
Fc epsilon RI signal K
IL-5 signaling pathw W
Ras signaling pathw K
EPO Receptor Signal W
Renal cell carcinom K
Signaling of Hepato W
Gap junction K
PI3K-Akt signaling K
B cell receptor sign K
Prolactin signaling K
Estrogen signaling K
Chemokine signalin K
GnRH signaling pat K
Hepatitis B K
Jak-STAT signaling K
IL-3 Signaling Path W
Neurotrophin signal K
IL-2 Signaling path W
Interleukin-11 Sign W
Insulin signaling pa K
Hepatitis C K
Oncostatin M Signal W
T cell receptor sign K
Pathways in cancer K
IL-9 signaling pathw W
Natural killer cell m K
B Cell Receptor Sig W
Leptin signaling pa W
IL-6 signaling pathw W
Non-small cell lung cancer
NON-SMALL CELL LUNG CANCER
GRB2
ErbB signaling pathway
Acute myeloid leukemia
enRoute - experimental data for paths
Female
Male
Days to Deat
GRB2
SOS1
SOS2
NRAS
BRAF
ARAF
RAF1
MAP2K2
MAP2K1
MAPK1
MAPK3
CCND1
Non-small cell lung cancer
Stratifications - LineUp
Non-Genetic - LineUp
Mapping Datasets
Map onto Pathway
Map in enRoute
None
RPPA
Mutations
mRNA-seq
Copy Number
mRNA
Methylation
Stratification
Filter Stratifications
Methylation
microRNA
mRNA
mRNA-seq
RPPA
Clinical
Copy Number
Mutations
Ra
Stratification
1. Ethnicity
2. Gender
3. Histological Type
4. Race
5. Tumor Tissue Site
6. Vital Status
Groups
Select None
Group
Size
Female 223
Male 351
Serial Combiner
Separator
Metrics
Experimental Data Assignment

Case Study: CCLE Data

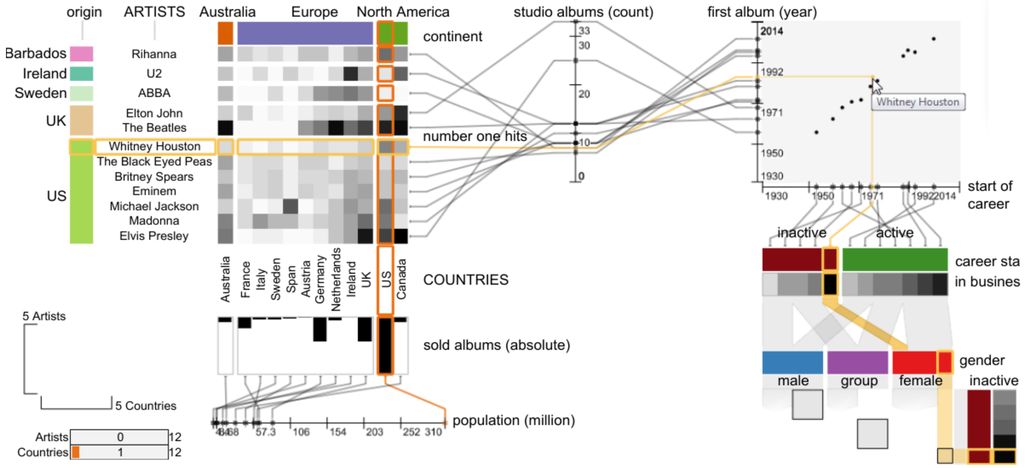



Pancreatic cancer
Non-small cell lung cancer
Proliferation
Survival
Angiogenesis
Adhesion
Migration / Invasion
Differentiation
Glioma
DNA
Jun
Elk
Myc
+p
ERBB2
GRB2
SOS1
HRAS
ARAF
MAP2K1
MAPK1
CCND1
PIK3R5
AKT3
FOXO3
CALML6
CAMK2A
ErbB signaling pathway
SHC2
Erlotinib
Lapatinib
ERBB2
SHC4
SHC3
SHC1
SHC2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

http://entourage.caleydo.org

Other Techniques

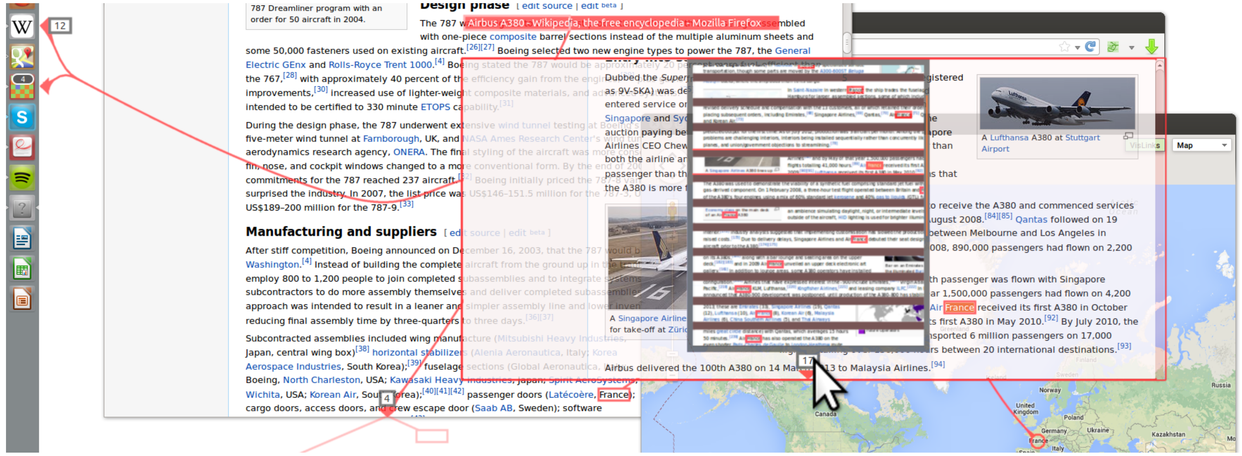
Creating Integrated Visualizations from Subsets



IEEE TVCG (InfoVis '14)

HONORABLE MENTION AWARD

Visualizing Hidden Content

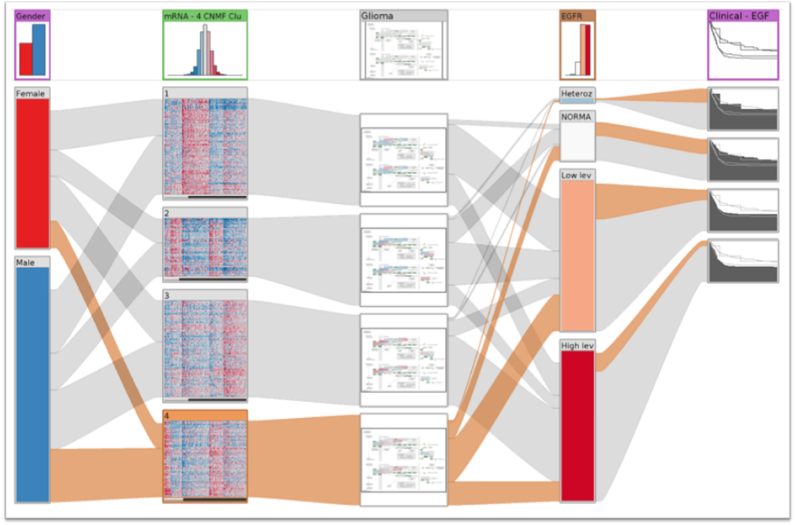


ACM CHI'14

HONORABLE MENTION AWARD

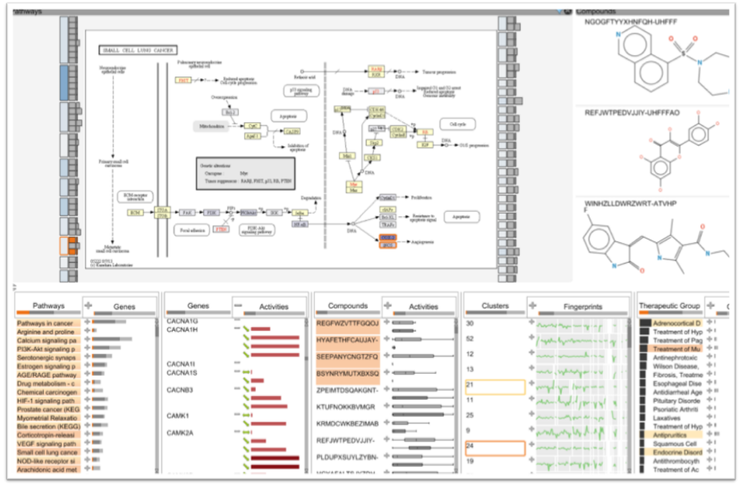
Other Domain Problem Projects

Heterogeneous Datasets
Cancer Subtype Analysis



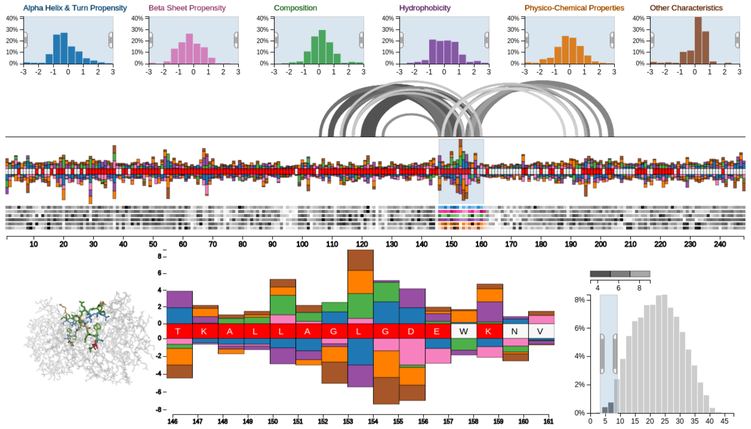
Nature Methods, 2014
CG & A 2014
EuroVis 2012
3rd Best Paper

Faceted Browsing of Graphs
Drug Discovery



IEEE TVCG (VAST '14)

Protein Sequence Analysis



BMC Proceedings 2014

CALEYDO

http://caleydo.org

CALEYDO

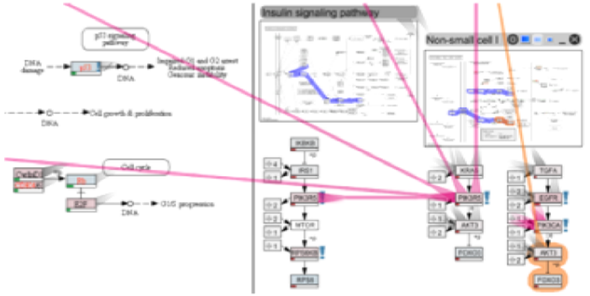
Home | Team | Projects | Downloads | Publications | Help

Home

About

Caleydo is an open source visual analysis framework targeted at biomolecular data. The biggest strength of Caleydo is the visualization of interdependencies between multiple datasets. Caleydo can load tabular data and groupings/clusterings. You can explore relationships between multiple groupings, between different datasets and see how your data maps onto pathways. Caleydo has been successfully used to analyze mRNA, miRNA, methylation, copy number variation, mutation status and clinical data as well as other datset types.

All published parts of the software are available as open source. Developers and researchers that want to use or enhance the current tools are invited to contribute. See our Github page for further details.



Pathway Visualization.

News

Partners

- Graz University of Technology
  Graz, Austria
- Johannes Kepler University Linz
  Linz, Austria
- Harvard University
  Cambridge, MA, USA
- Harvard Medical School
  Boston, MA, USA

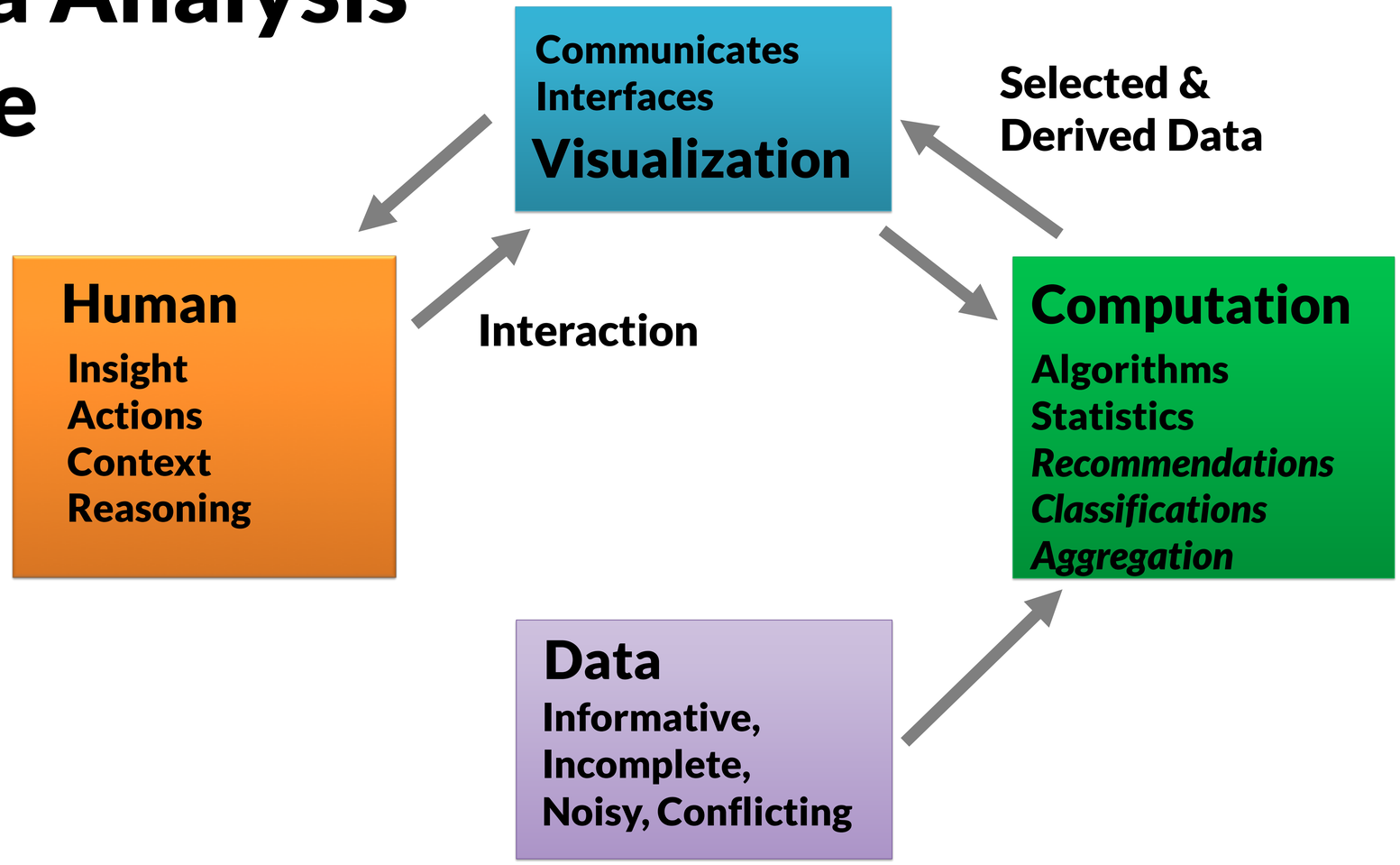
Projects

- StratomeX
  Genomic Stratification of Diseases
- Entourage & enRoute
  Experimental Data in Pathways
- LineUp
  Multi-Attribute Rankings
- ConTour
  Relational Data in Drug Discovery
- Domino
  Exploring Subsets of Tabular Datasets
- Furby
  Bicluster Exploration
- Visual Linking & Hidden Content
  Showing Relationships

RESEARCH CHALLENGES

Interacting with Data

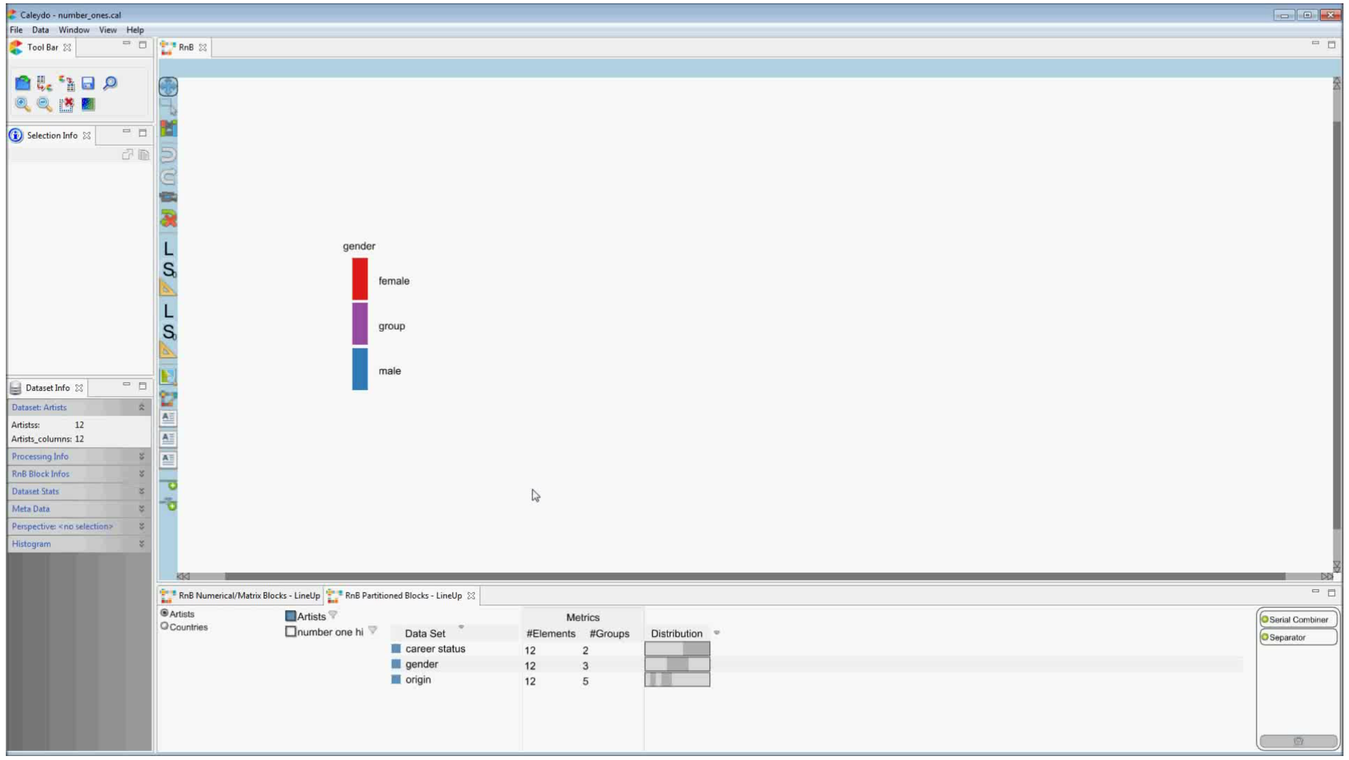
The Future of Data Analysis is (also) Interactive



Can we make creating interactive vis systems easier?

Right now: prerogative of skilled software developers

Goal: democratize the creation of visualizations

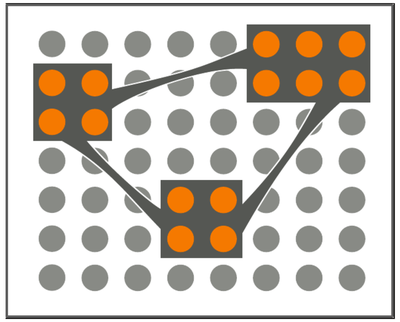


How can we scale to increasingly big datasets?

No local copies of datasets

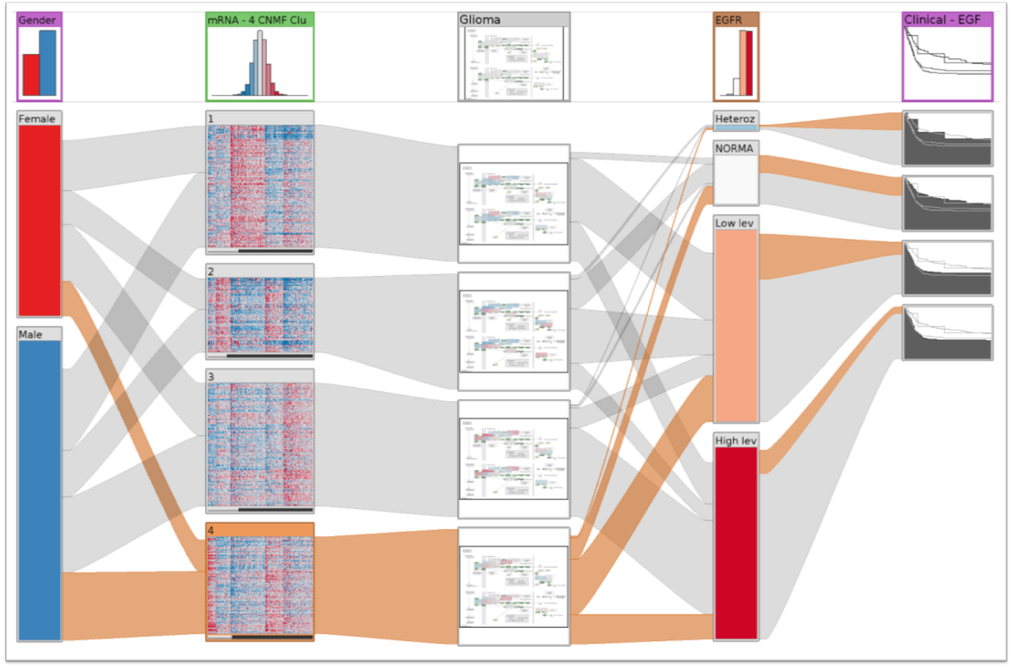
No in-memory processing

....



Approach:

- Automatic slicing into subsets
- Query or explore ranked subsets
- Pull in relationships for subset

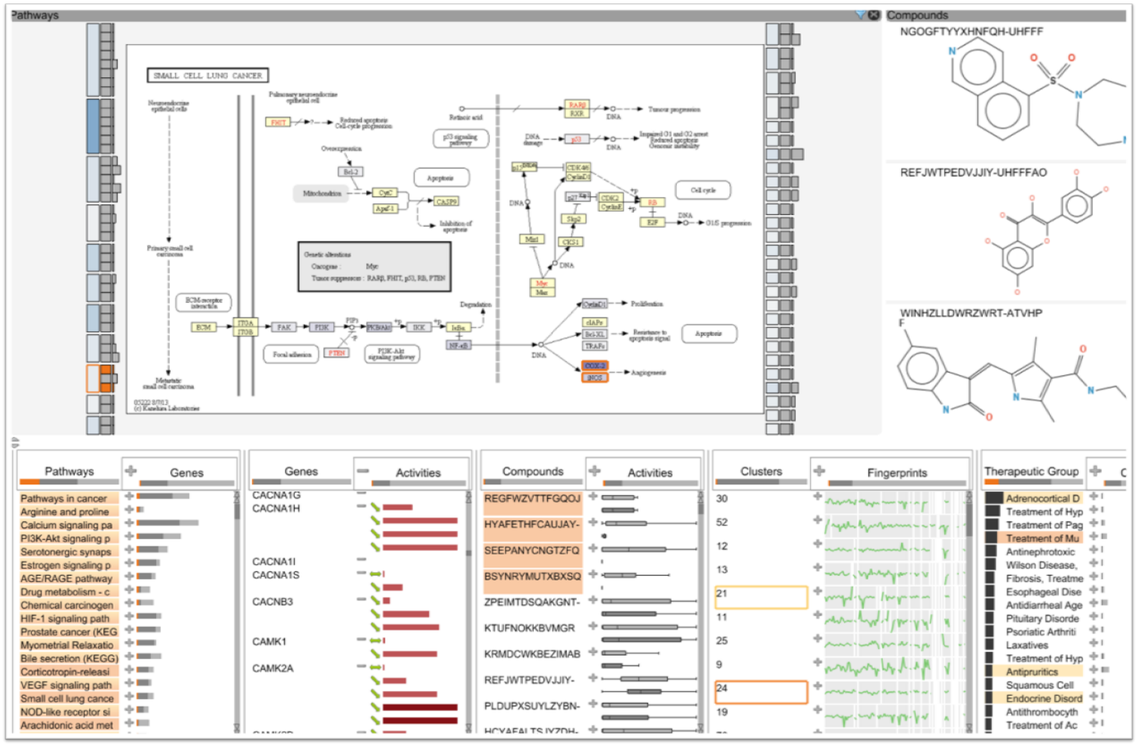


Solve Important Domain Problems

Some problems too specific for general methods

High impact when solving them

Inspiration for research



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

University of Utah
Scientific Computing Institute
December 3, 2014

HARVARD
School of Engineering
and Applied Sciences

Enabling Scientific Discovery through Interactive Visual Data Analysis

?

Credits:

Marc Streit, Nils Gehlenborg, Christian Partl, Samuel Gratzl, Markus Steinberger, Manuela Waldner, Hendrik Strobelt, Romain Vuillemot, Dieter Schmalstieg, Denis Kalkofen, Mark Borowsky, Anne Mai Wasserman, Hanspeter Pfister