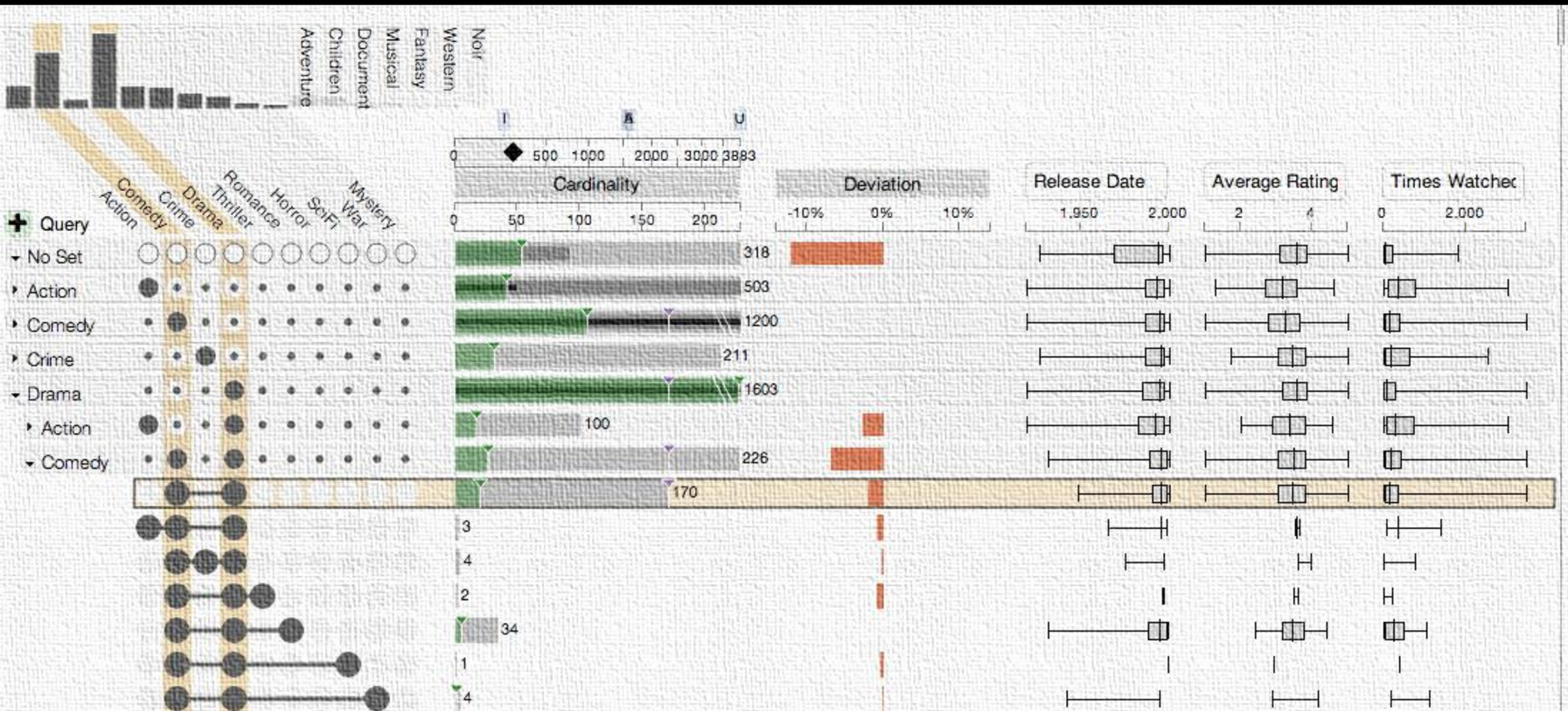
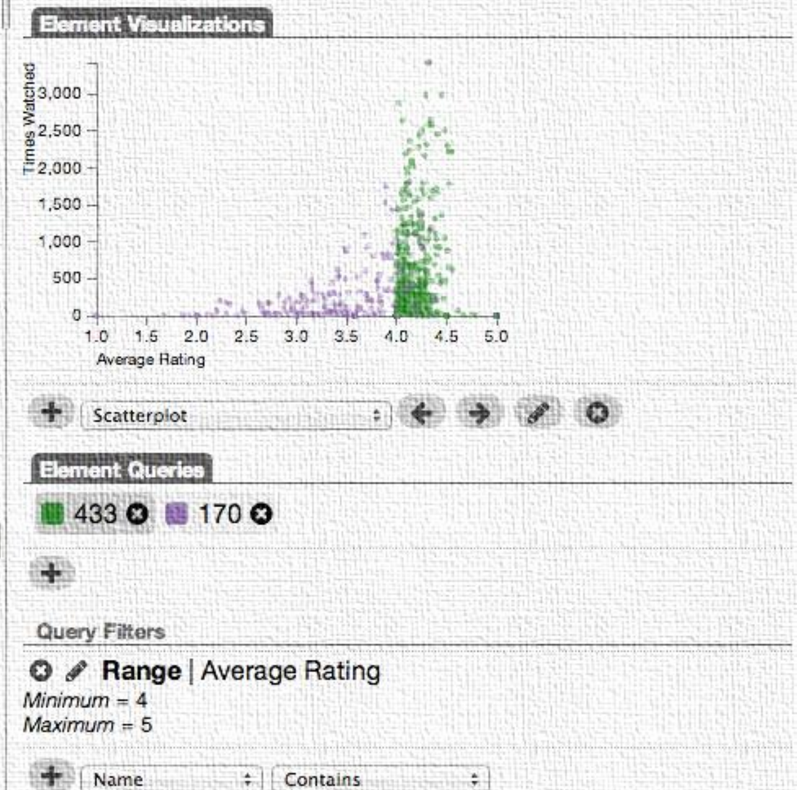
Alexander Lex
@alexander_lex
http://alexander-lex.com

HARVARD
School of Engineering
and Applied Sciences

Visual Data Analysis for Biology & Pharmacology

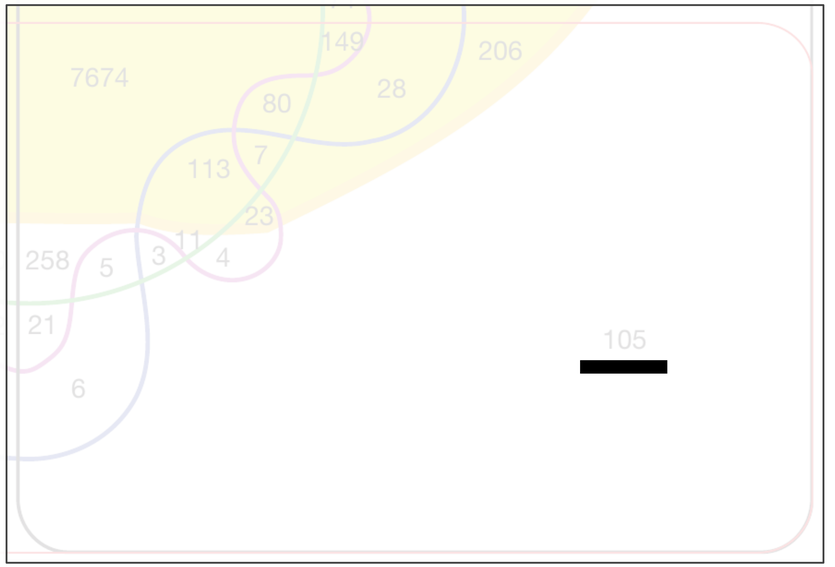




visualization *pictures*

The purpose of computing is insight, not numbers.

- Richard Wesley Hamming, 1962



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

Good Data Visualization

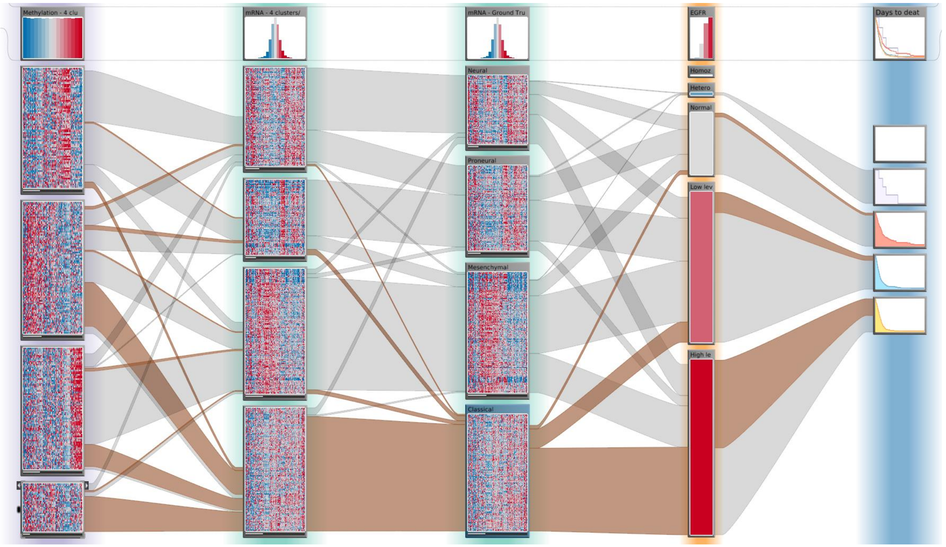
… makes data accessible

… combines strengths of humans and computers

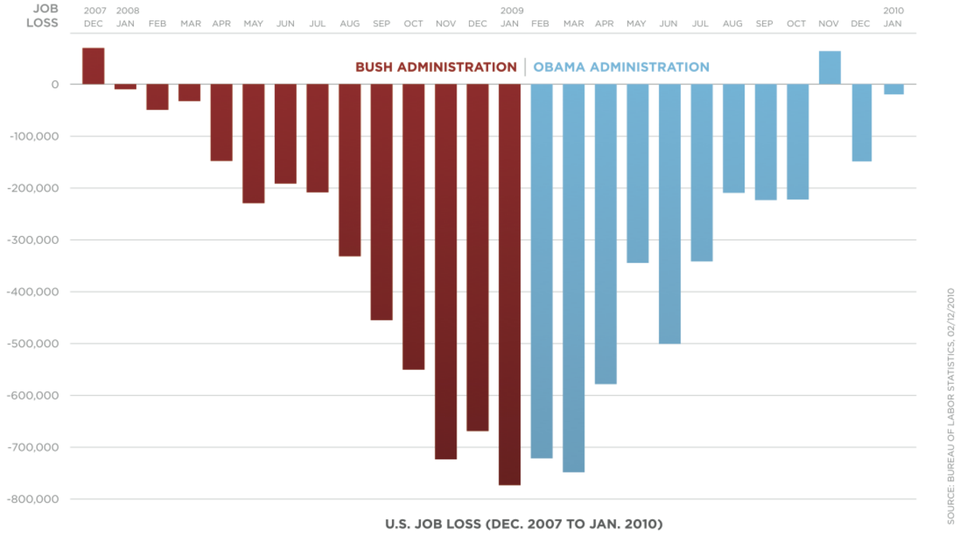
… enables insight

… communicates

Purpose of Visualization

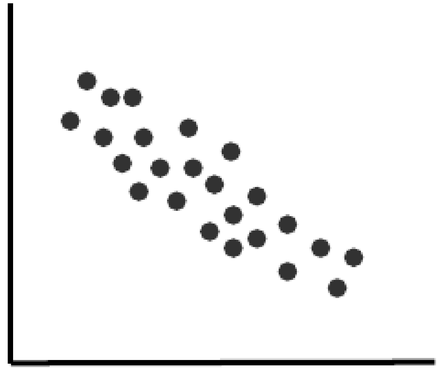


[Obama Administration]

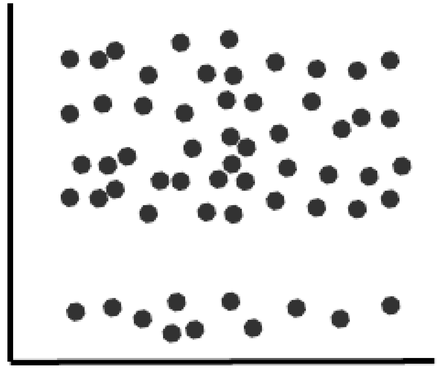


Open Exploration ⟷ **Communication**

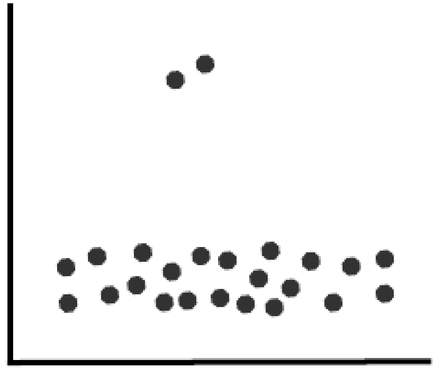
Visualization for Pattern Discovery



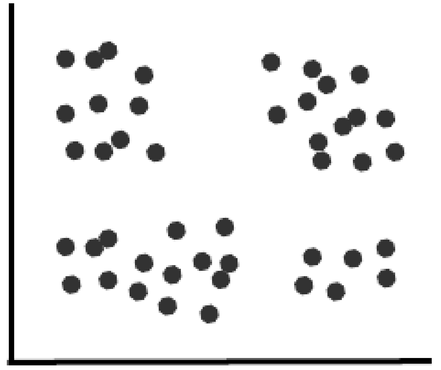
trends



gaps



outliers



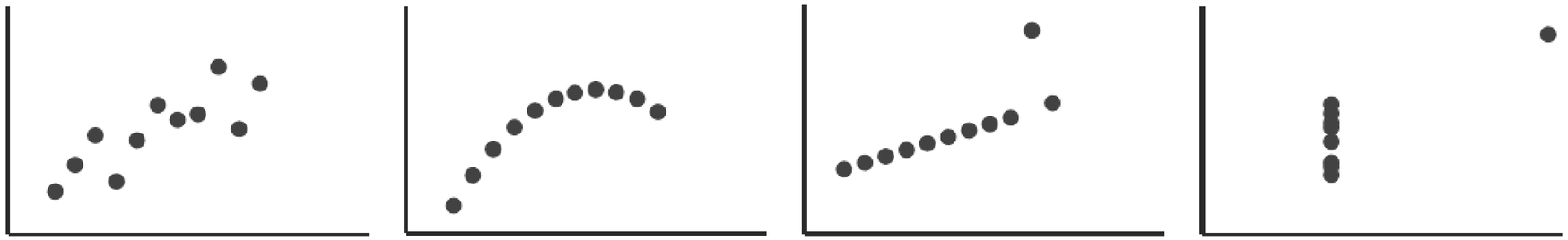
clusters

Can We Trust Statistics?

I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68					8	6.89

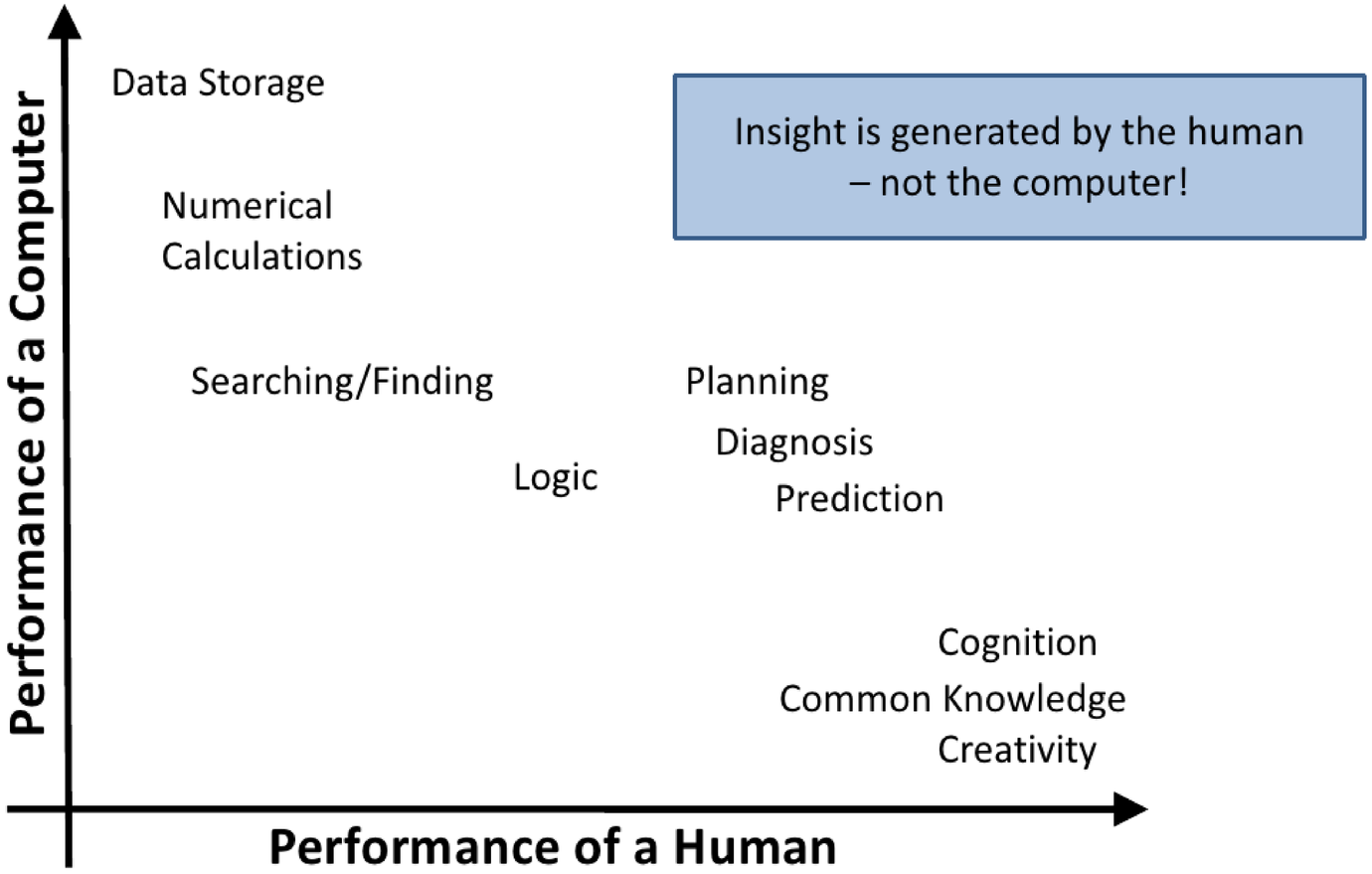
Mean x: 9 y: 7.50
Variance x: 11 y: 4.122
Correlation x – y: 0.816
Linear regression: y = 3.00 + 0.500x

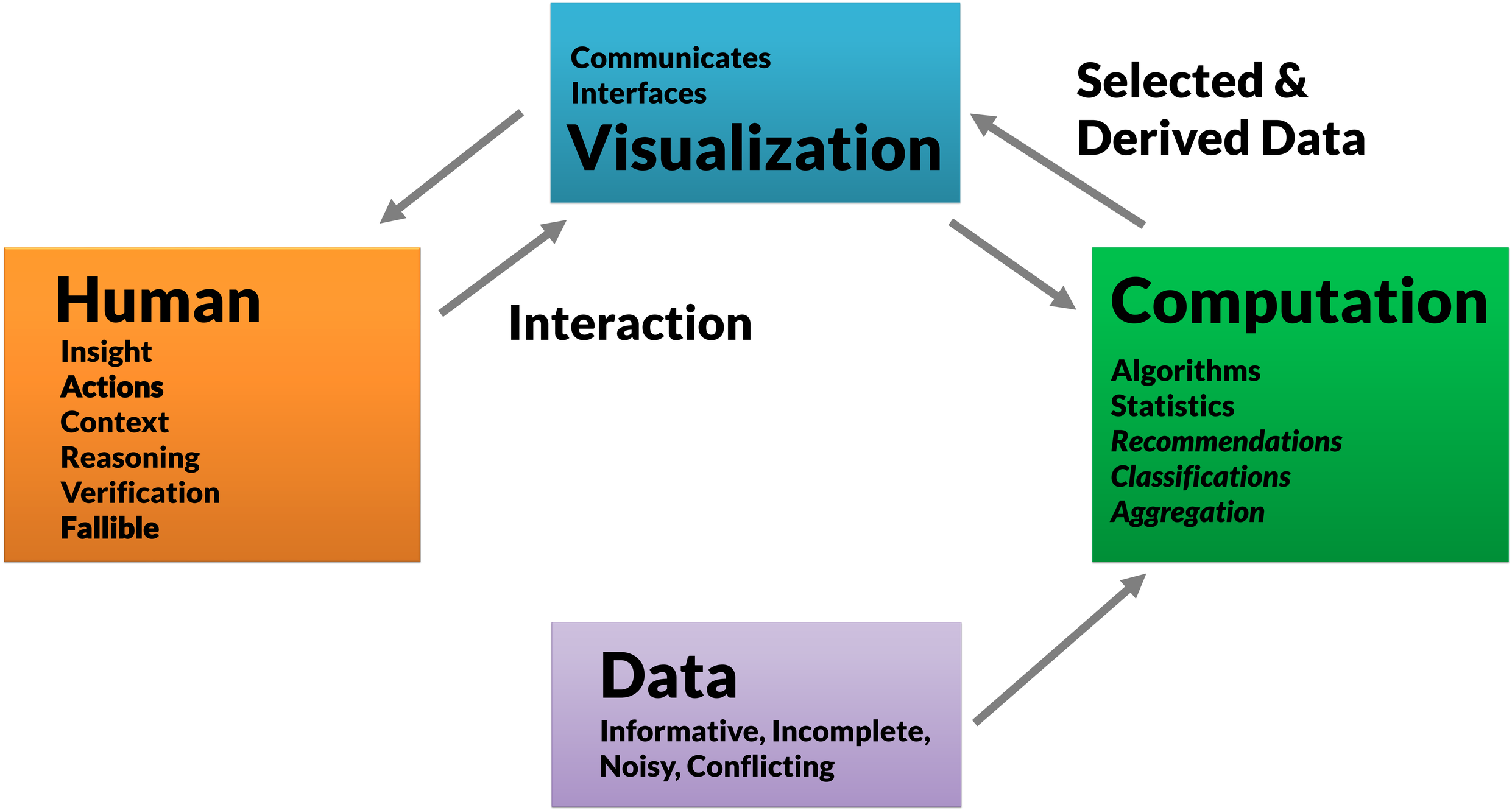
Anscombe’s Quartett



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

The Ability Matrix




Communicates
Interfaces
Visualization
Selected &
Derived Data
Human
Insight
Actions
Context
Reasoning
Verification
Fallible
Interaction
Computation
Algorithms
Statistics
Recommendations
Classifications
Aggregation
Data
Informative, Incomplete,
Noisy, Conflicting

Who am I?

alexander-lex.com
@alexander_lex

PostDoc & Lecturer @ Harvard,
Visual Computing Group,
PI: Hanspeter Pfister

Associated PostDoc @ Novartis
Institute of BioMedical Research,
Scientific Data Analysis,
PI: Mark Borowsky

How it looks like at home

Credits

Marc Streit — *Johannes Kepler University Linz, AT*

Nils Gehlenborg — *Harvard Medical School, Boston, USA*

Christian Partl — *Graz University of Technology, AT*

Samuel Gratzl — *Johannes Kepler University Linz, AT*

Hendrik Strobelt — *Harvard University, Cambridge, USA*

Dieter Schmalstieg — *Graz University of Technology, AT*

Mark Borowsky — *Novartis (NIBR), Cambridge, USA*

Hanspeter Pfister — *Harvard University, Cambridge, USA*

TU Graz

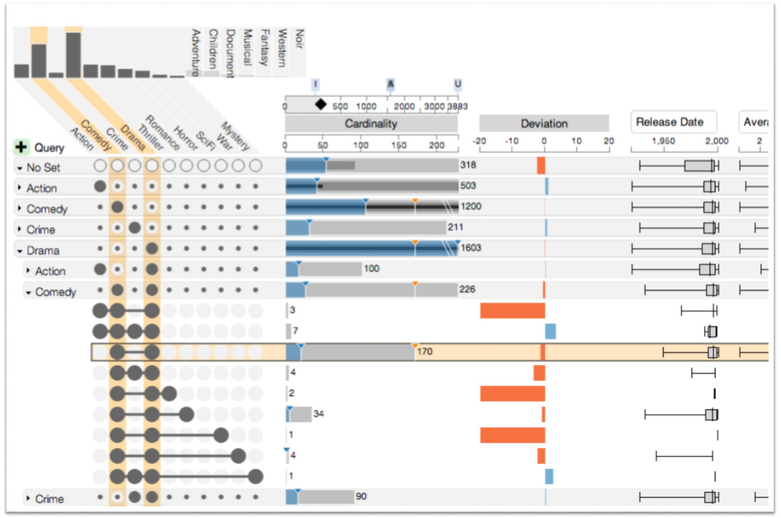
JKU JOHANNES KEPLER UNIVERSITÄT LINZ

HARVARD School of Engineering and Applied Sciences

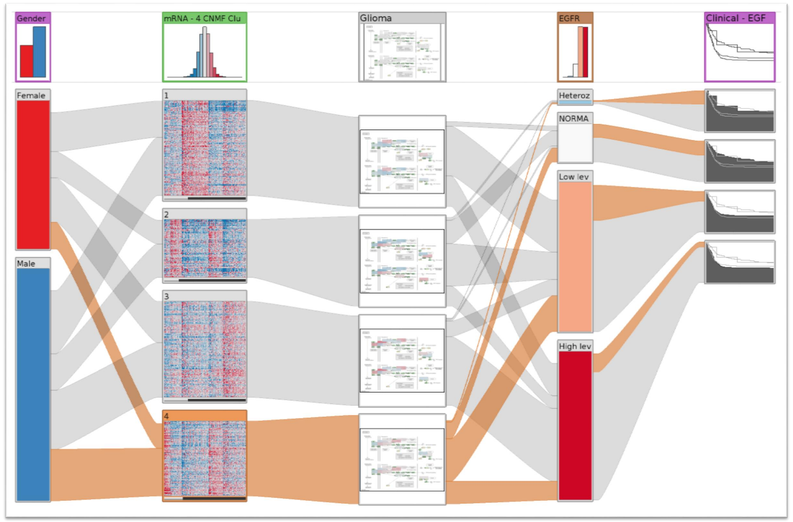
HARVARD MEDICAL SCHOOL

NOVARTIS INSTITUTES FOR BIOMEDICAL RESEARCH

Content

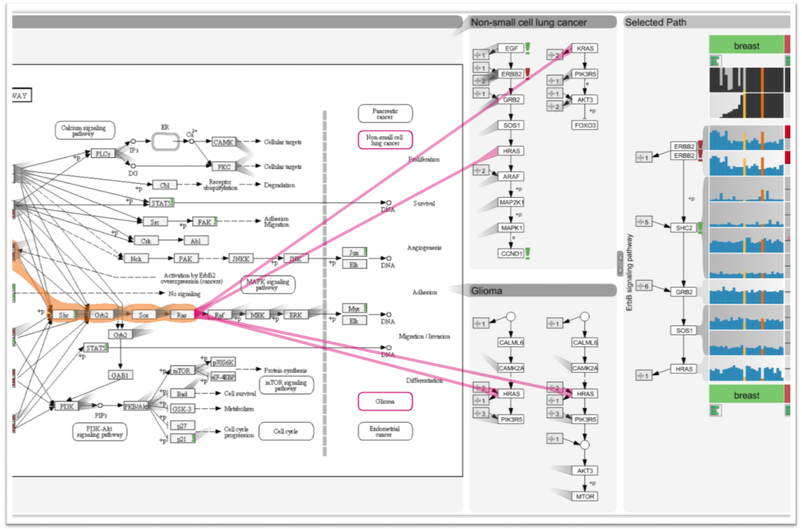


Set Visualization



Heterogeneous Datasets

Cancer Subtype Analysis



Multivariate Networks

Pathway Analysis

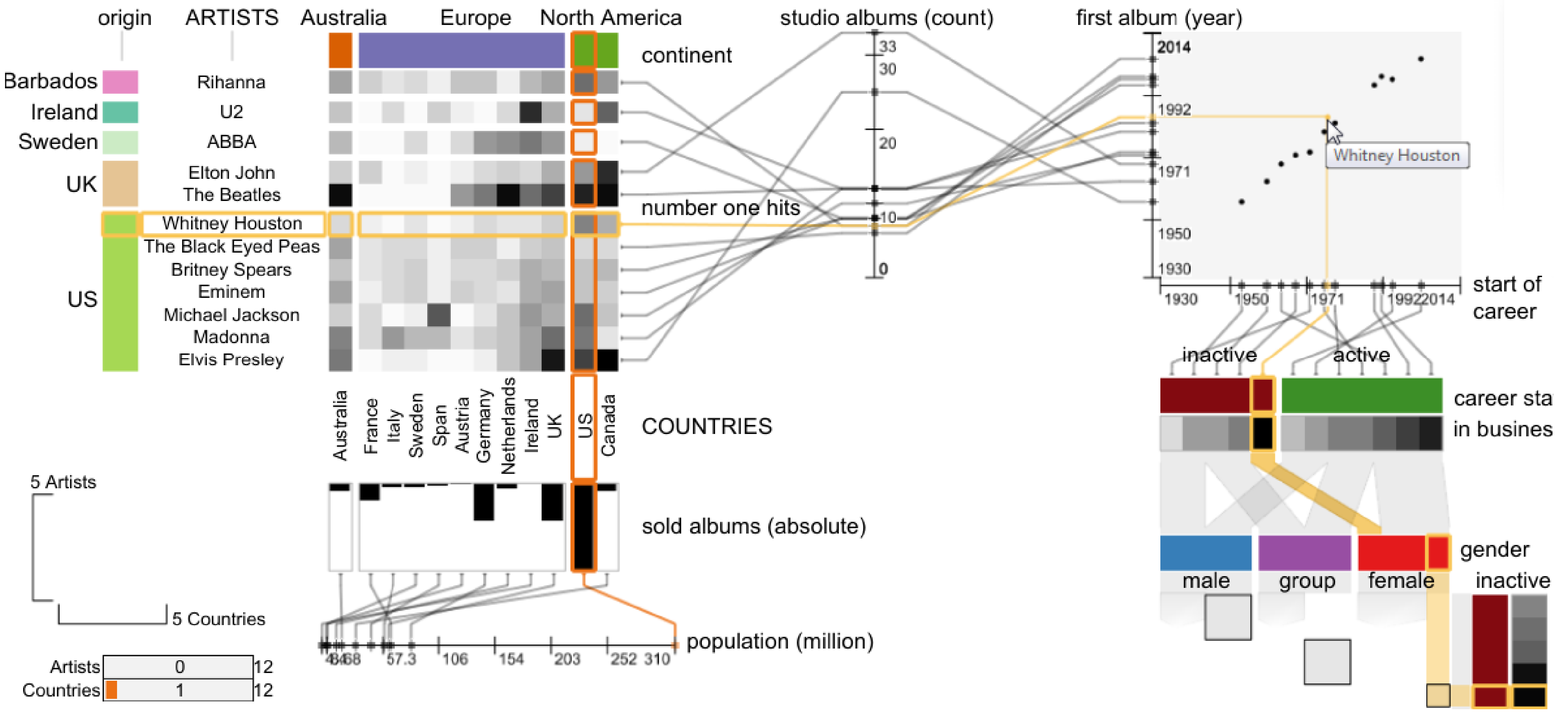
Subset Visualization

Domino: Extracting, Comparing, and Manipulating Subsets across Multiple Tabular Datasets

http://domino.caleydo.org

Samuel Gratzl, Nils Gehlenborg, Alexander Lex, Hanspeter Pfister and Marc Streit

IEEE TVCG (InfoVis '14)
HONORABLE MENTION AWARD



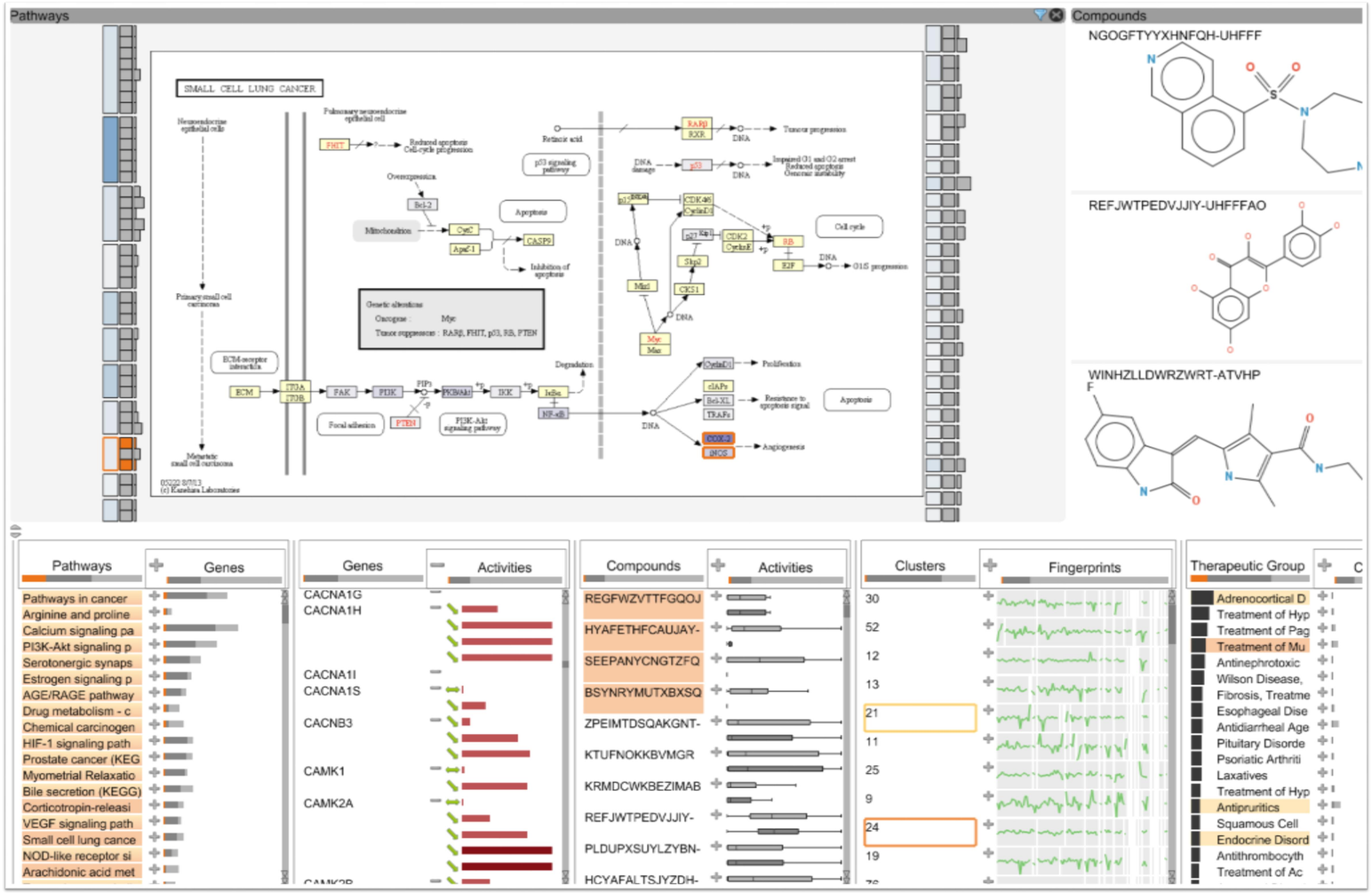
Multi-Relational Data
Drug Discovery

ConTour: Data-Driven Exploration of Multi-Relational Datasets for Drug Discovery

http://contour.caleydo.org

Christian Partl, Alexander Lex, Marc Streit, Hendrik Strobelt, Anne-Mai Wasserman, H. Pfister, and Dieter Schmalstieg

IEEE TVCG (VAST '14)



Sequence Visualization

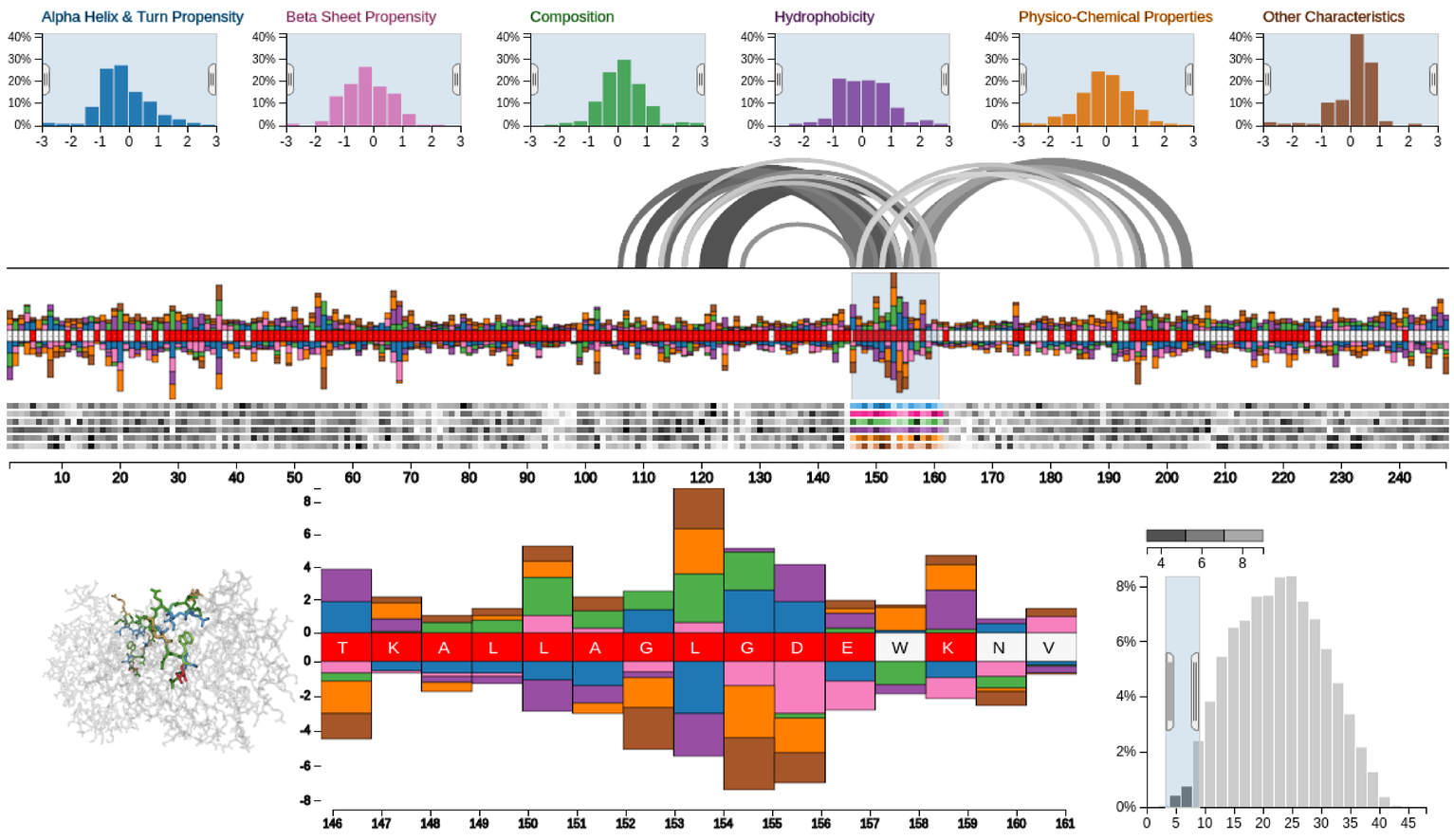
Protein Mutations

Mu-8: Visualizing Differences between Proteins and their Families

http://mu-8.com

Johnathan D Mercer, Balaji Pandian, Alexander Lex*, Nicolas Bonneel, and Hanspeter Pfister

BMC Proceedings 2014



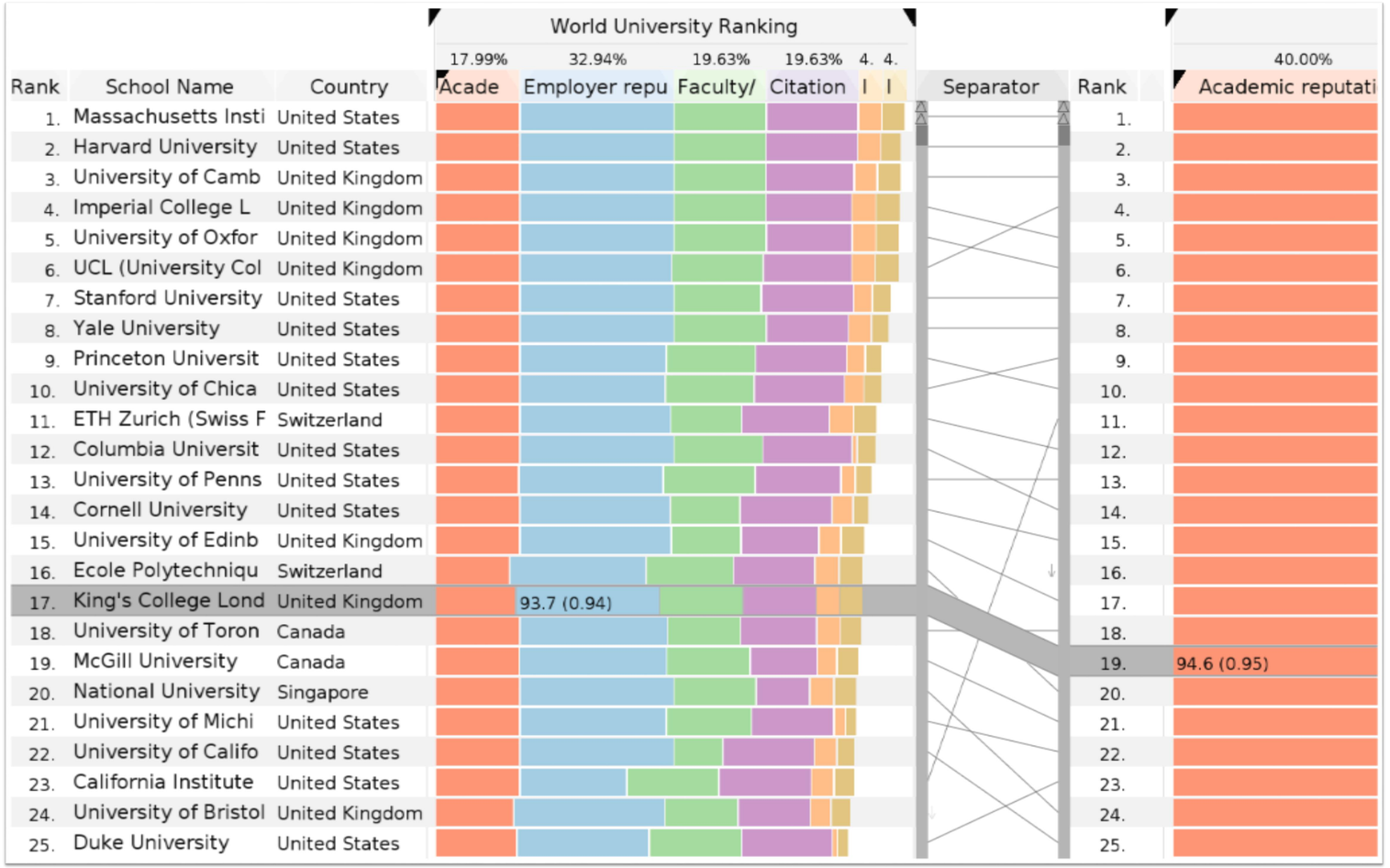
Multi-Attribute Rankings

LineUp: Visual Analysis of Multi-Attribute Rankings

http://lineup.caleydo.org

Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister, and Marc Streit

IEEE TVCG (InfoVis '13)
BEST PAPER AWARD

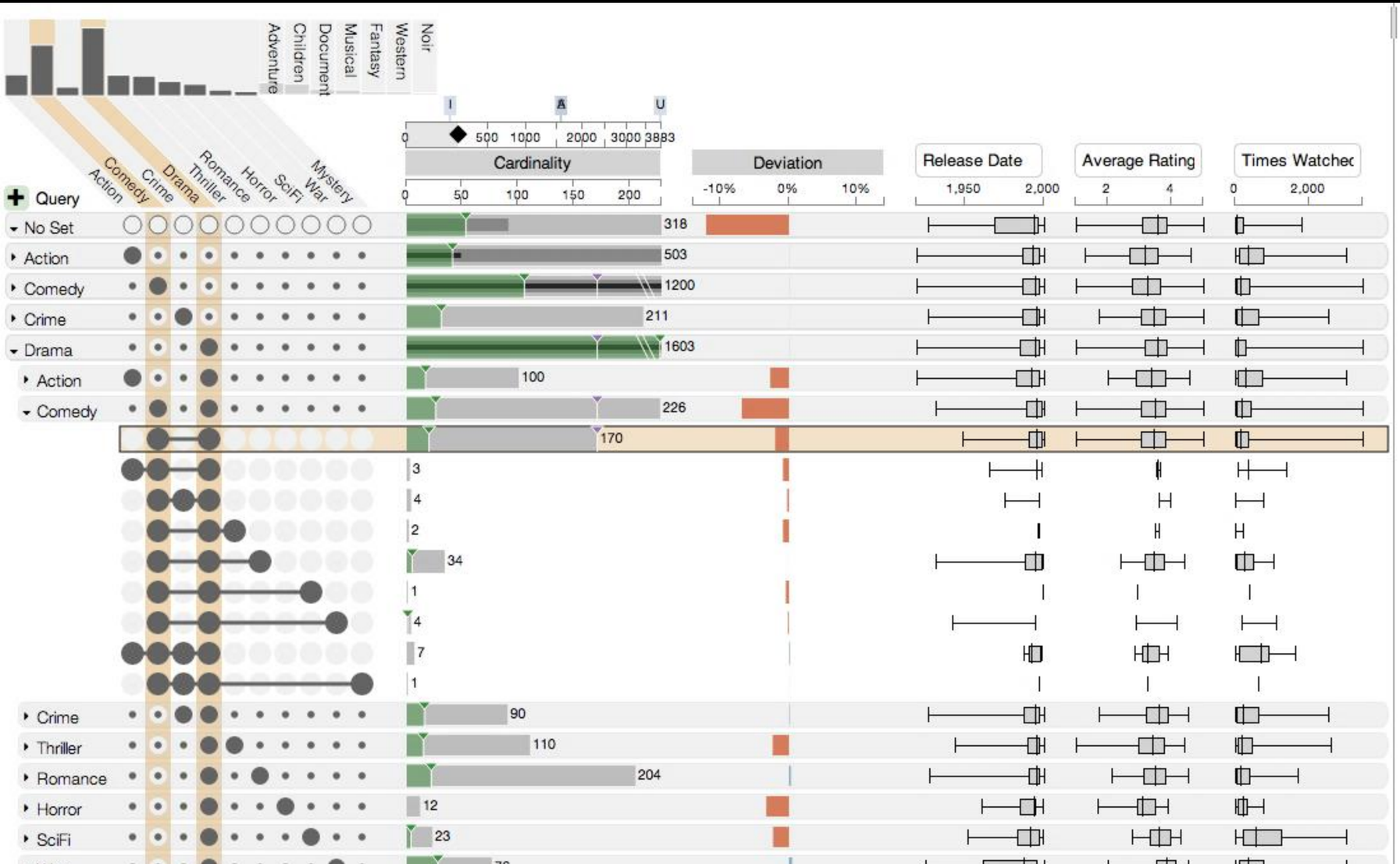


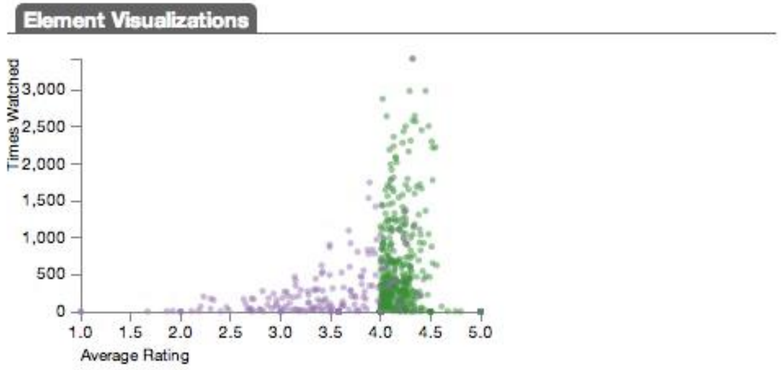
Alexander Lex, Nils Gehlenborg, Hendrik Strobelt,
Romain Vuillemot, and Hanspeter Pfister

UpSet

InfoVis'14

Visualizing Intersecting Sets

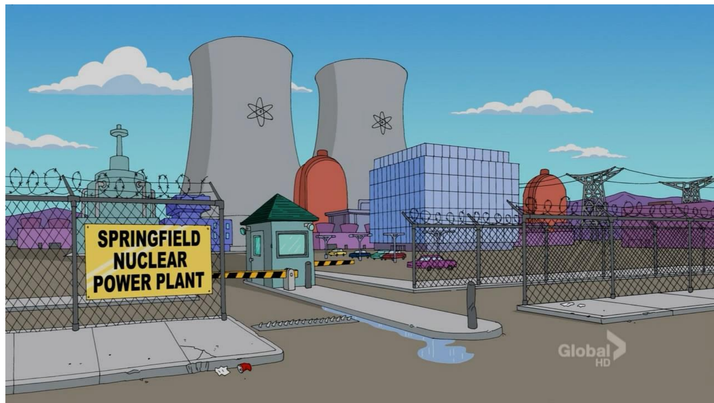




Name	Release Date	Average Rating	Times Watched	Set Count
Toy Story (1995)	1995	4.15	2077	2
Sense and Sensibility (1995)	1995	4.03	835	2
Persuasion (1995)	1995	4.06	179	1
City of Lost Children, The (1995)	1995	4.06	403	2

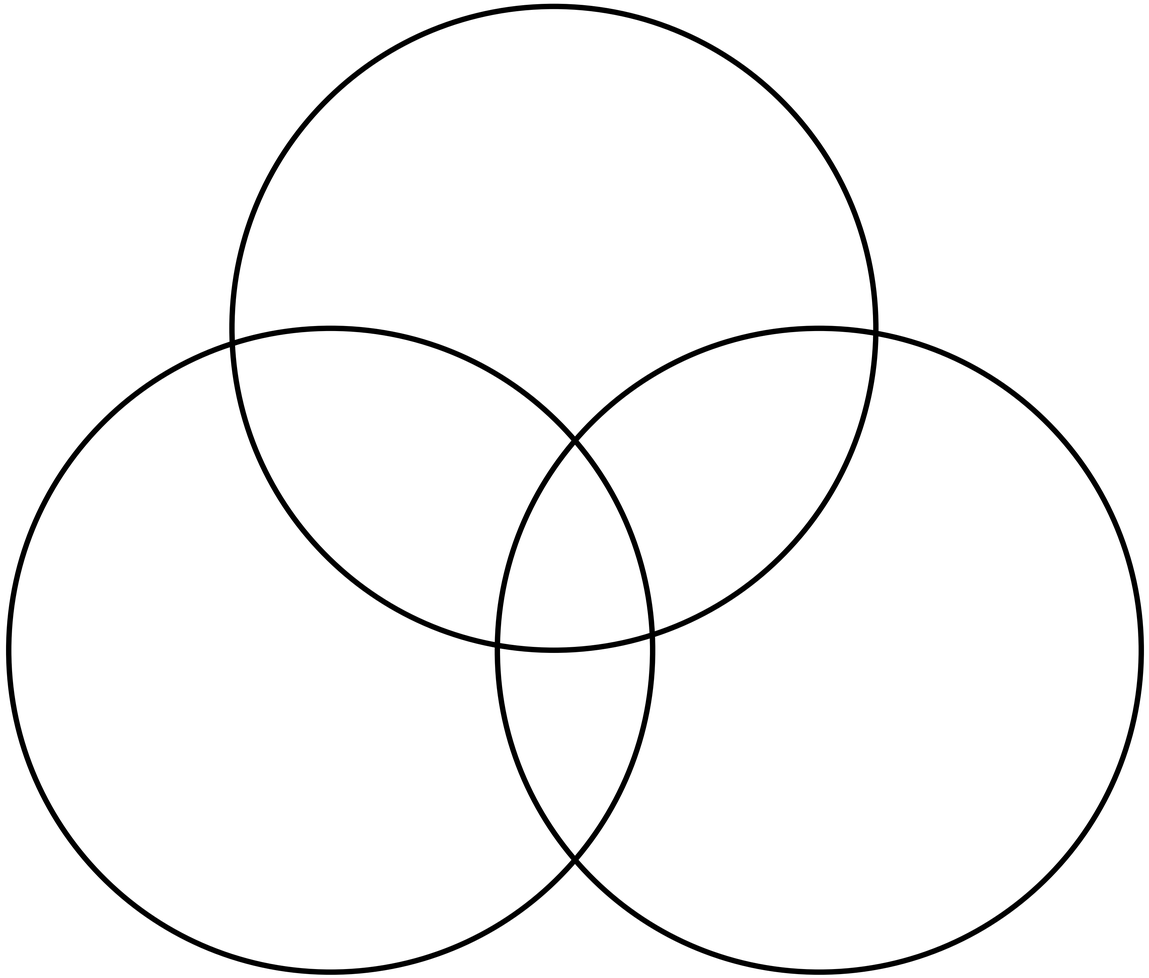

SPRINGFIELD
ELEMENTARY
SCHOOL


SPRINGFIELD
ELEMENTARY
SCHOOL


SPRINGFIELD
NUCLEAR
POWER PLANT
Global
HD

Evil

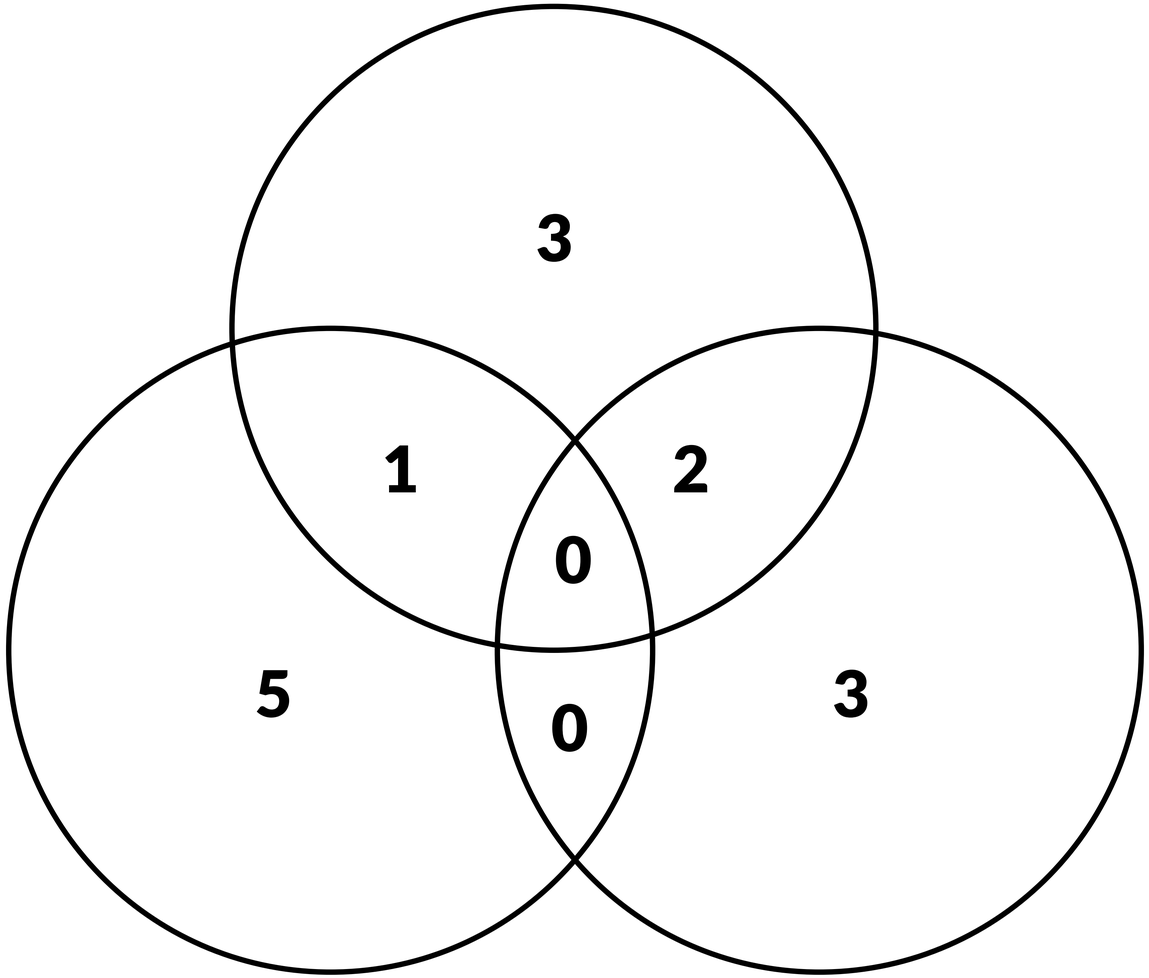

MATT GROENING




SPRINGFIELD ELEMENTARY SCHOOL


SPRINGFIELD NUCLEAR POWER PLANT
Global HD


MATT GROENING

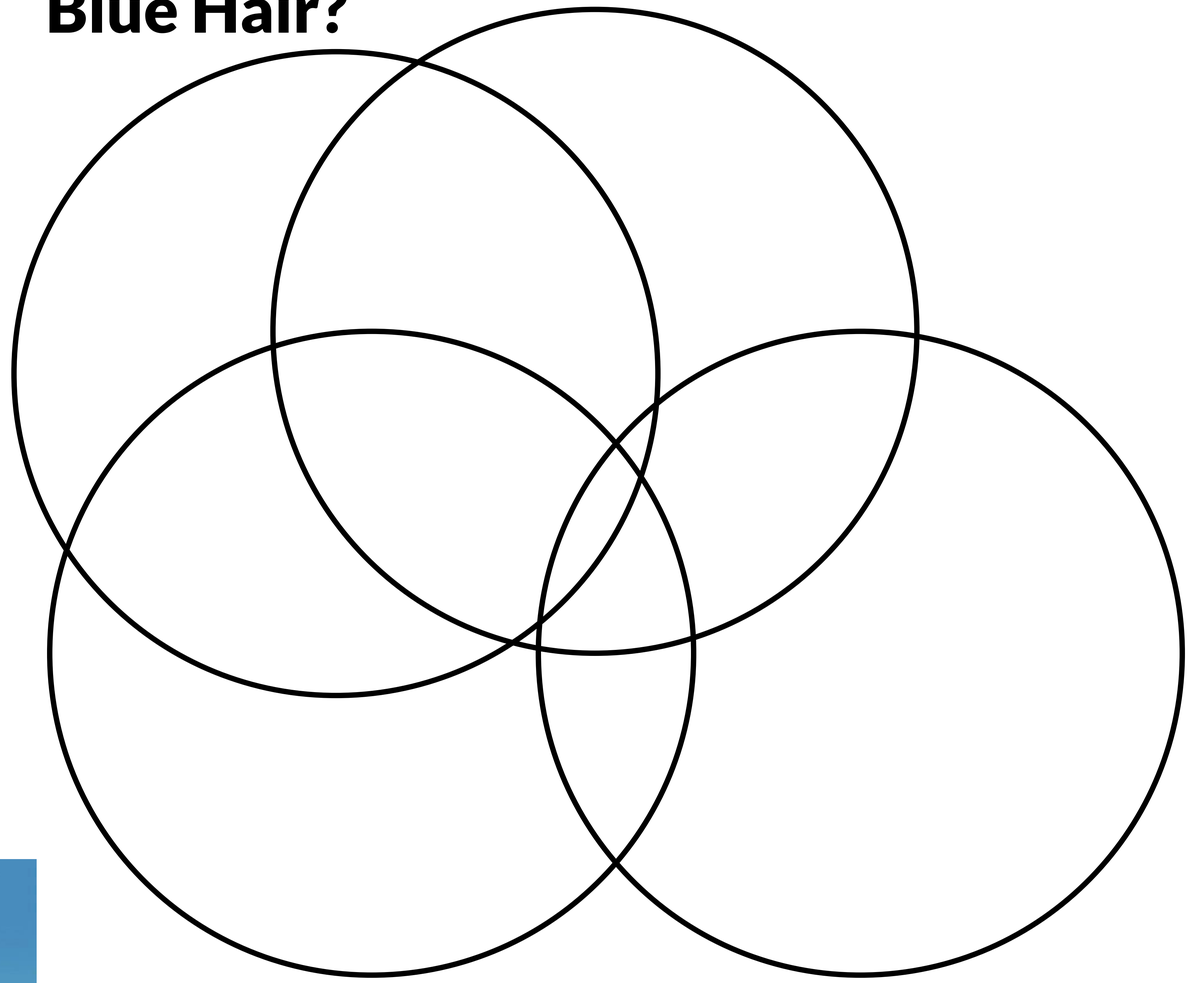

Evil
3
1
2
0
5
0
3


SPRINGFIELD ELEMENTARY SCHOOL


SPRINGFIELD NUCLEAR POWER PLANT
Global HD

Evil

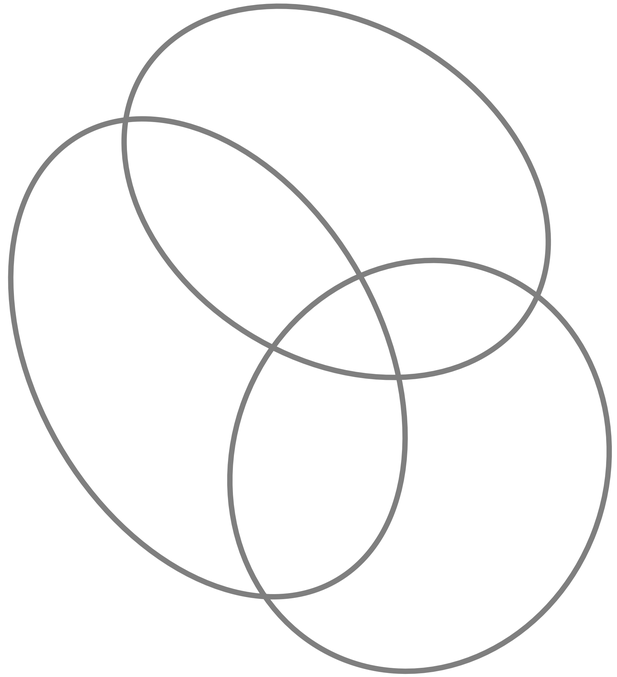
Blue Hair?



Evil

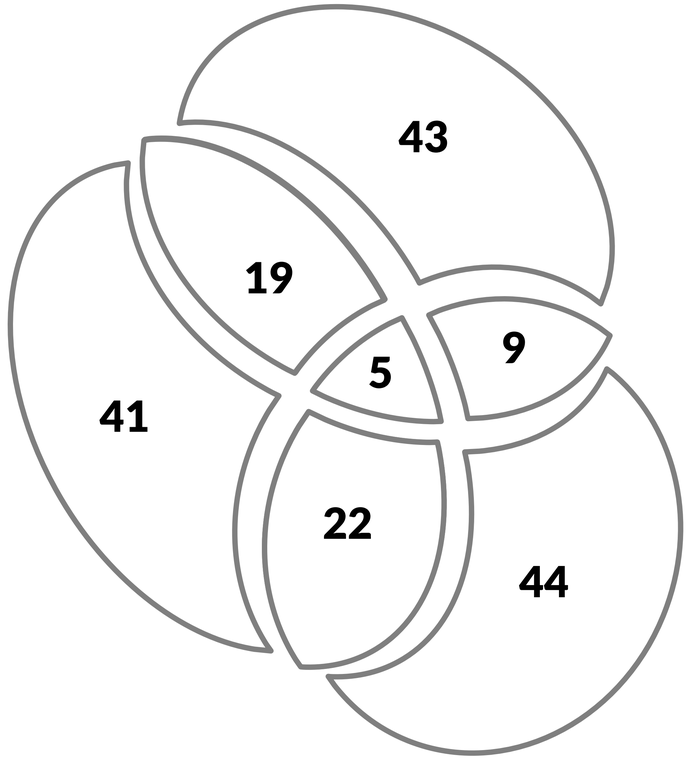
Blue Hair?

Duff Beer?

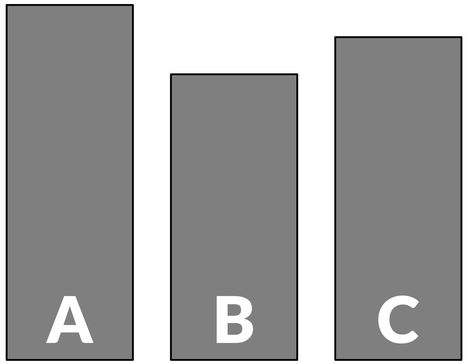


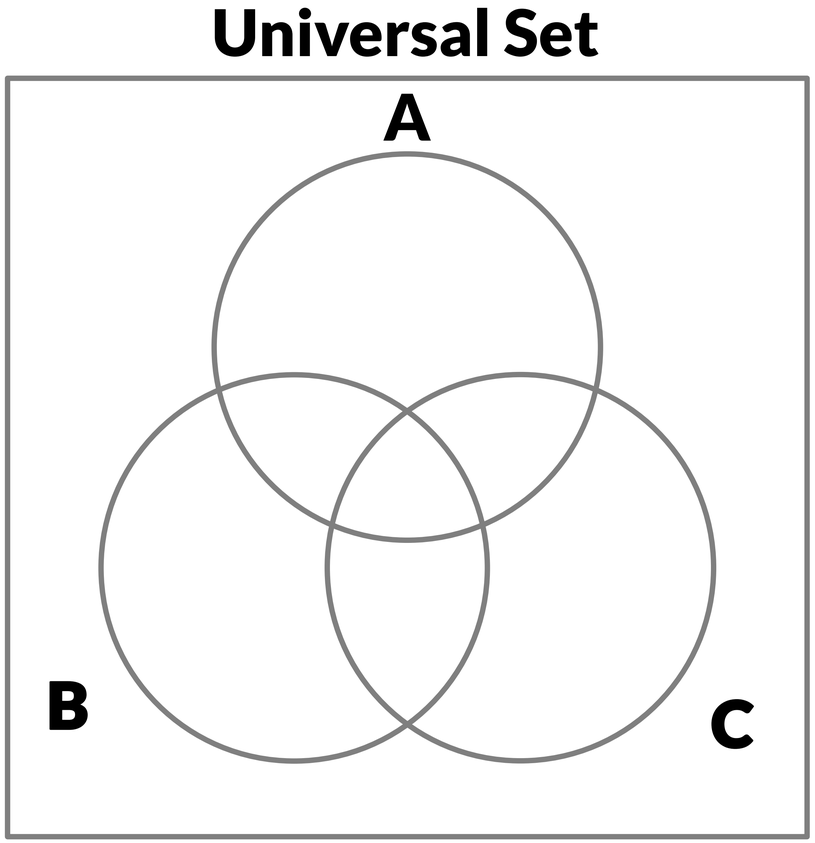
[created with EulerAPE]

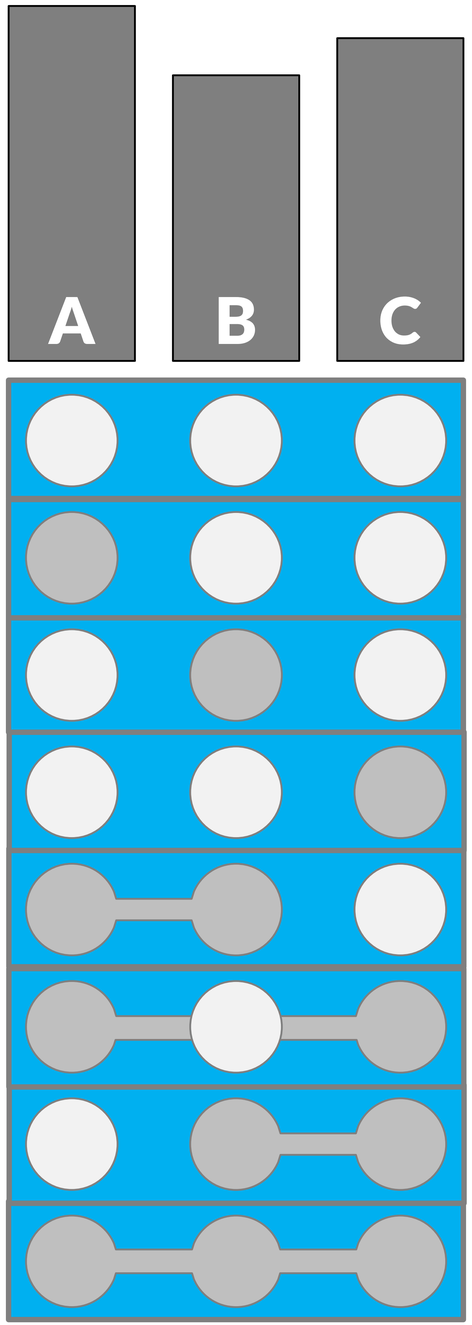
22 ? 19



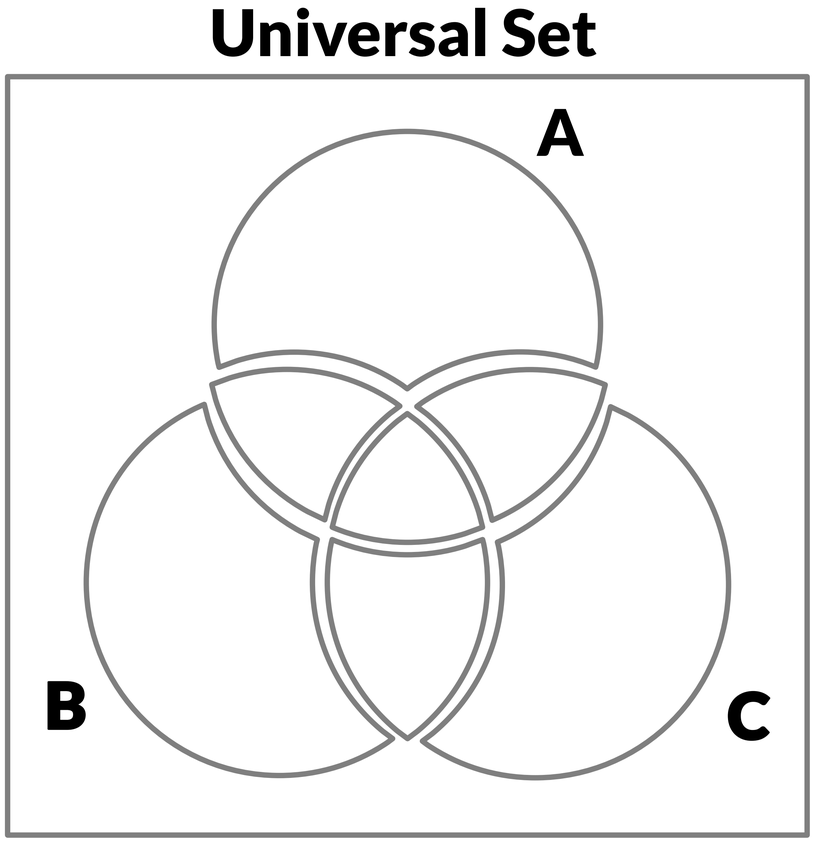
[created with EulerAPE]

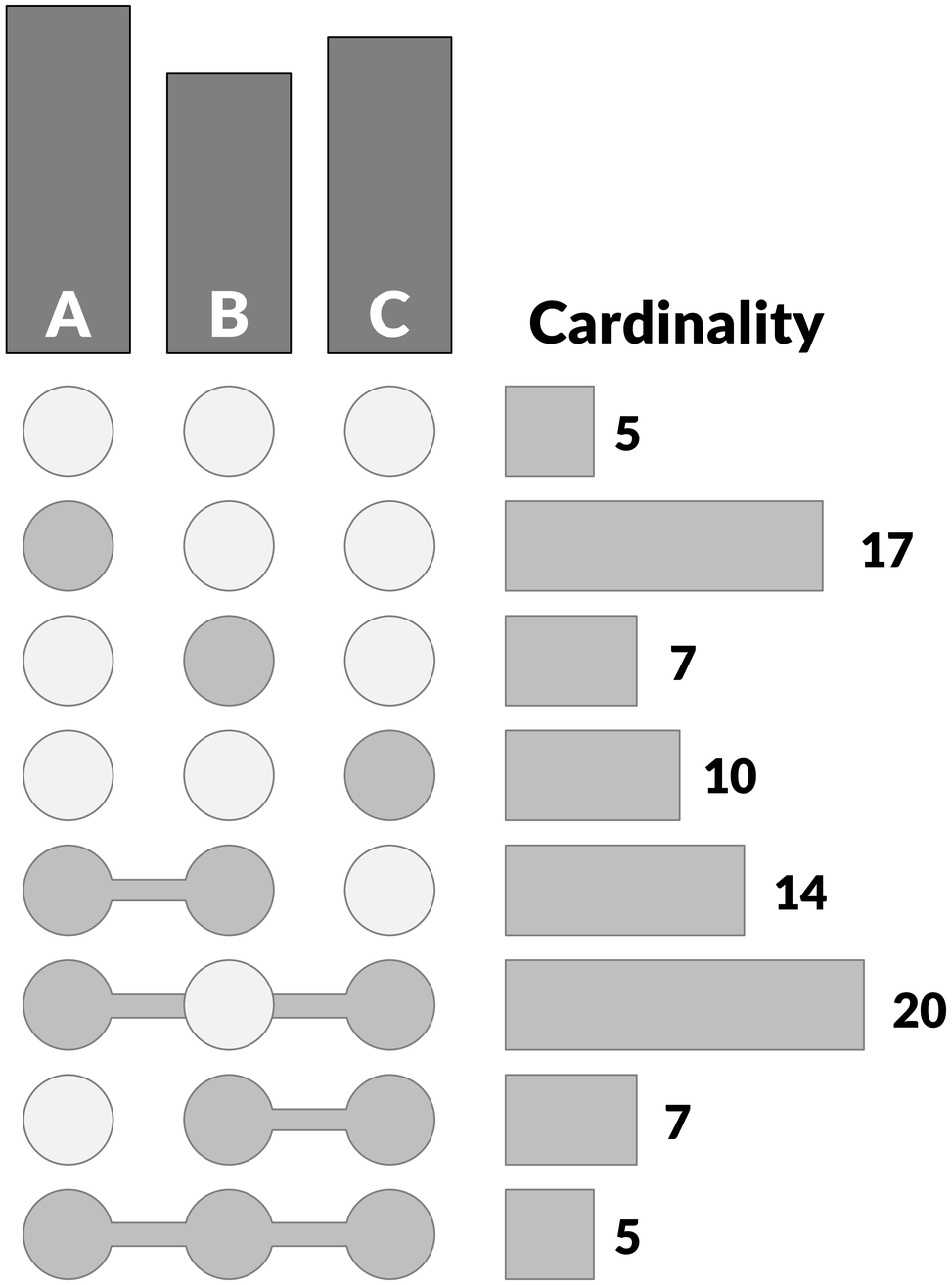

A
B
C

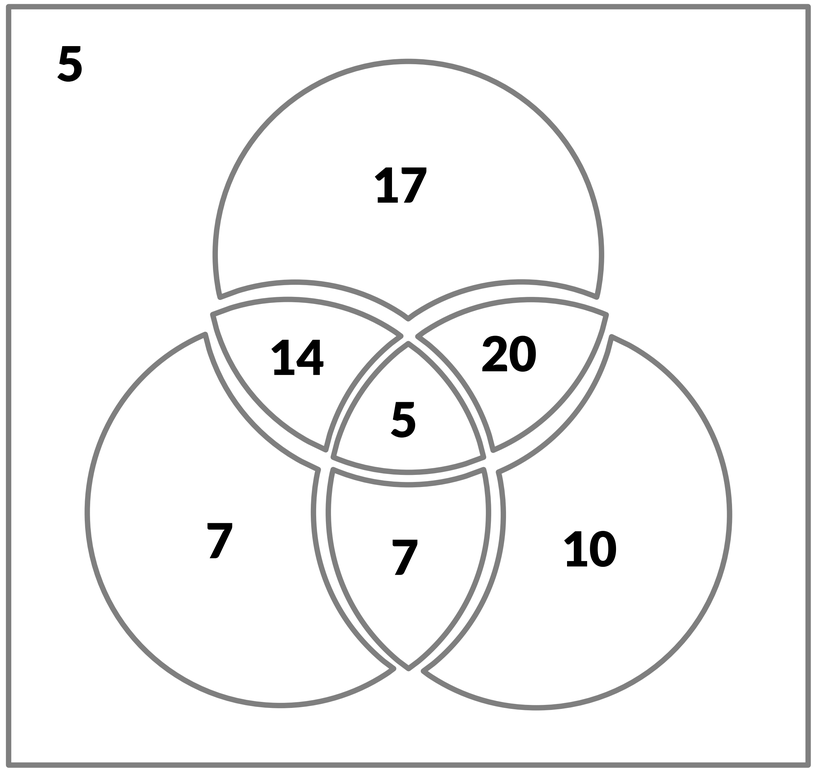

Universal Set
A
B
C


A
B
C

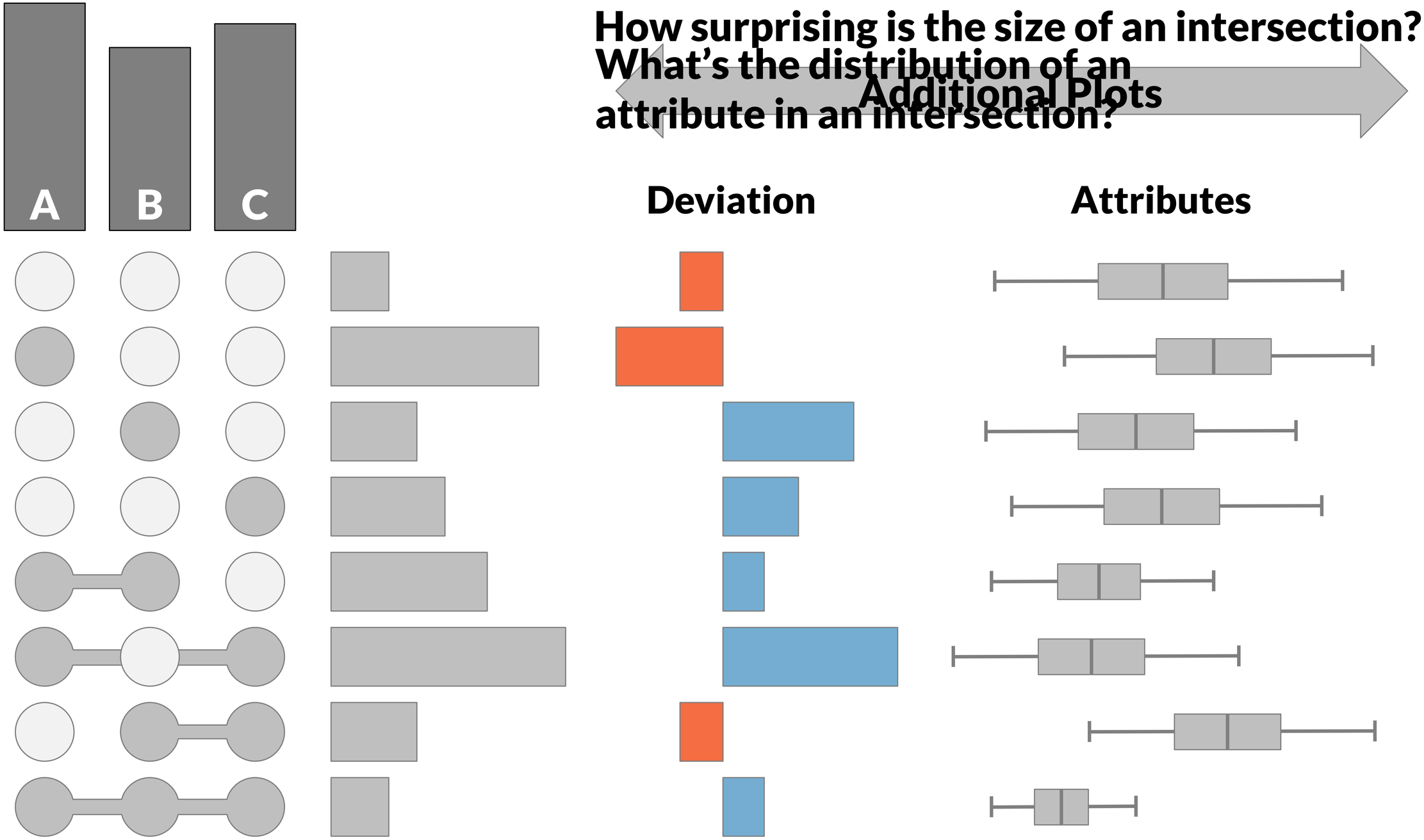

Must
Must Not

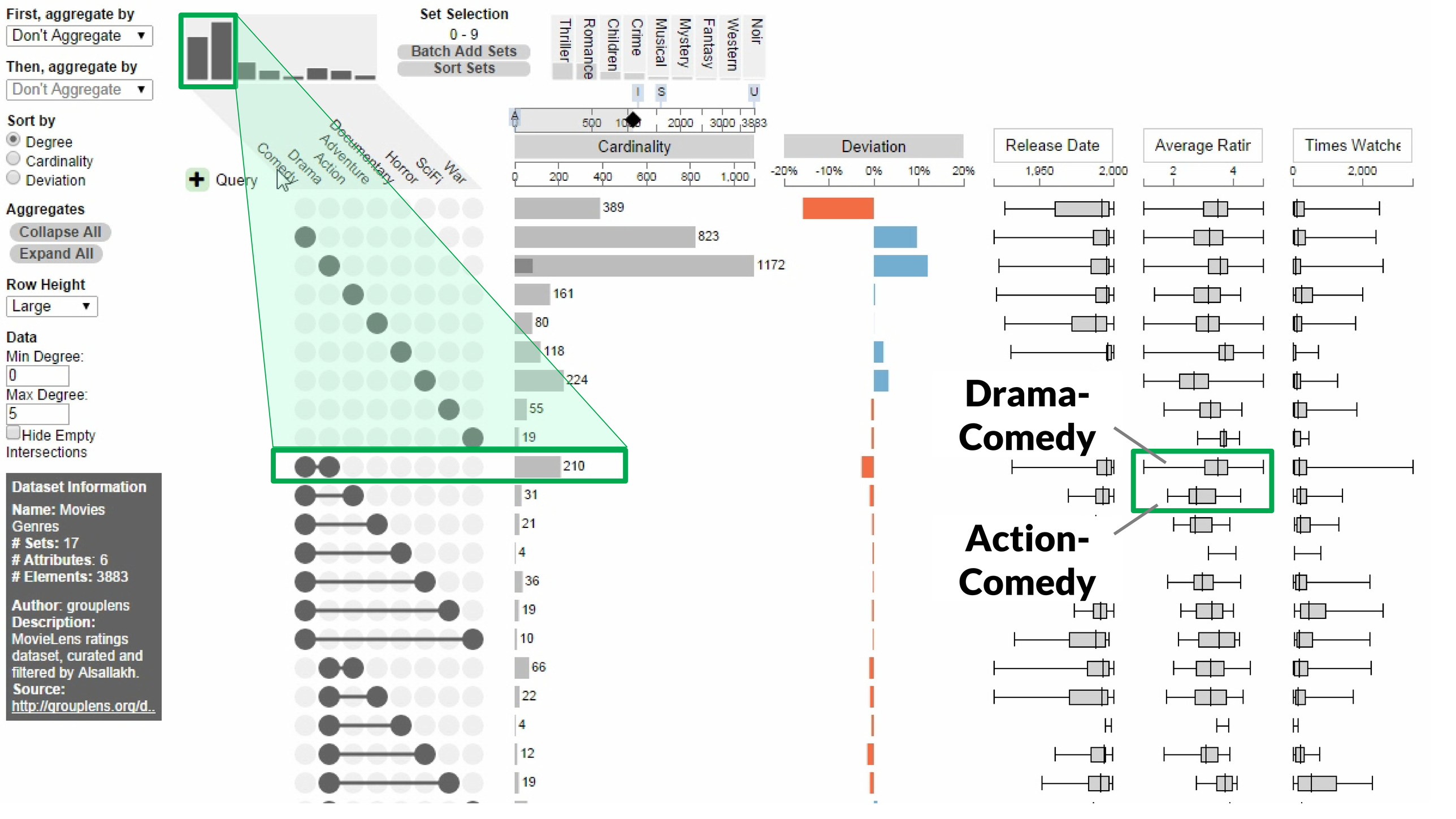

Universal Set
A
B
C


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

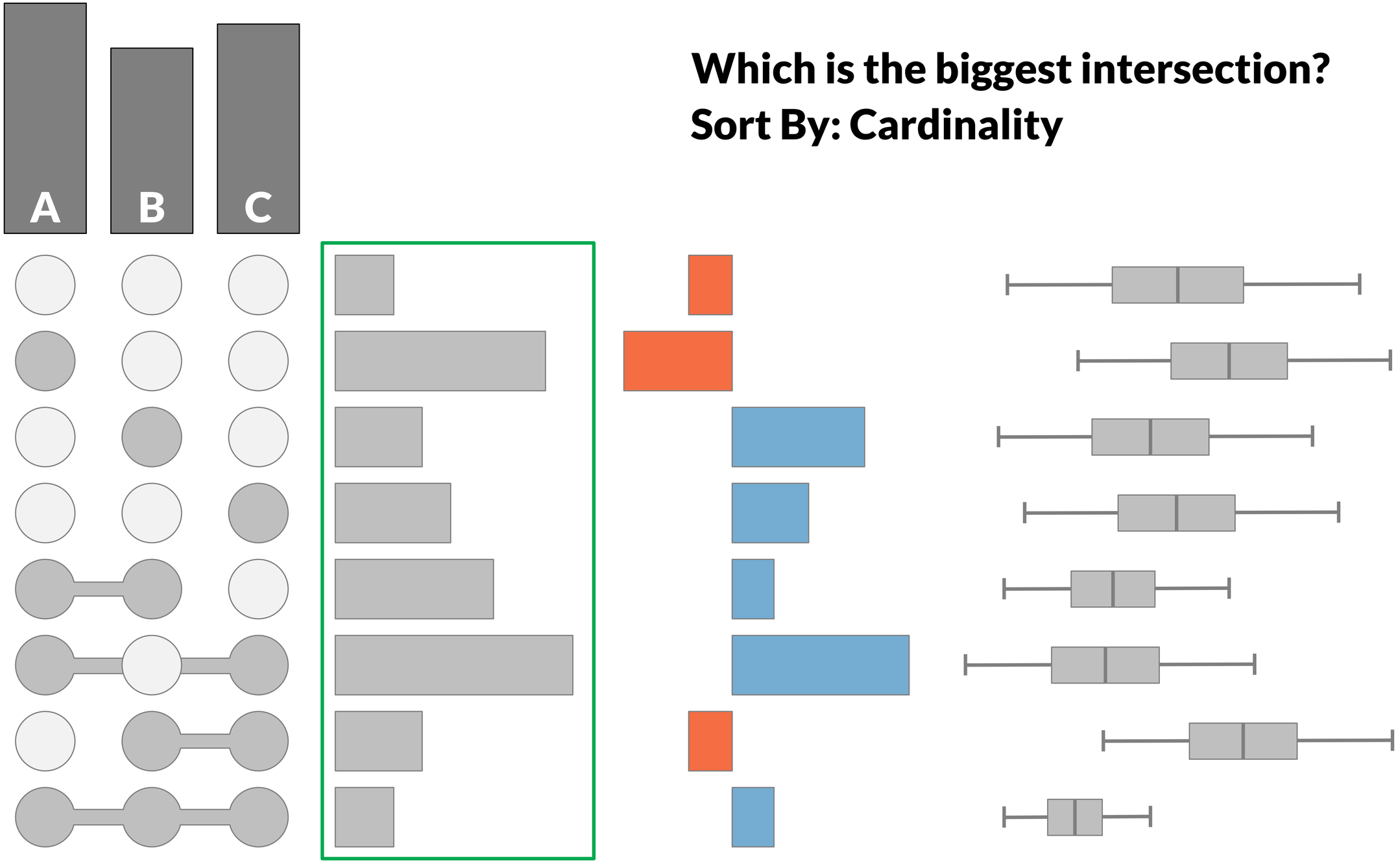

5
17
14
20
5
7
7
10

Plotting Attributes

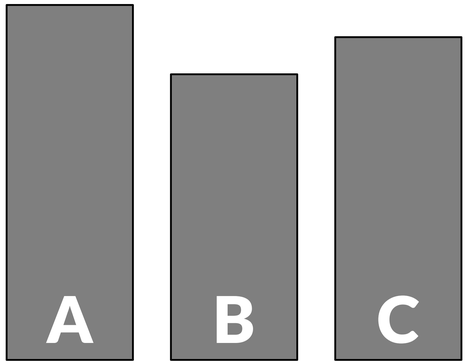

How surprising is the size of an intersection?
What's the distribution of an attribute in an intersection?
Additional Plots
A
B
C
Deviation
Attributes


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
389
823
1172
161
80
118
224
55
19
210
31
21
4
36
19
10
66
22
4
12
19
Drama-Comedy
Action-Comedy

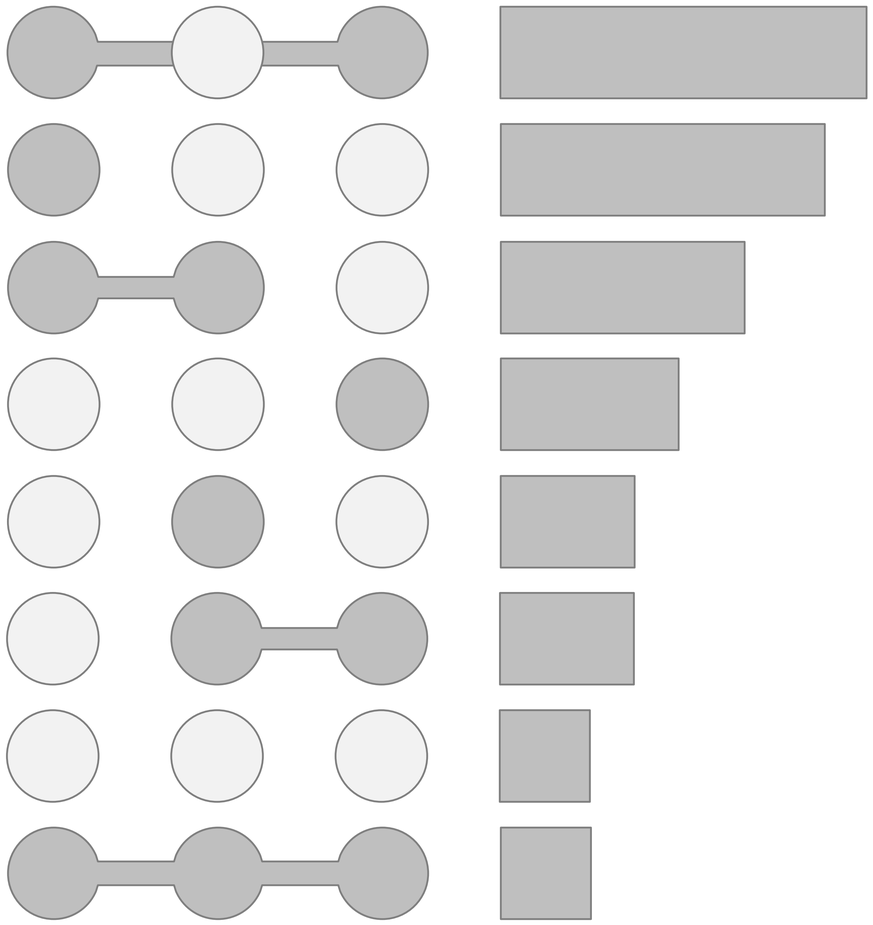
Sorting

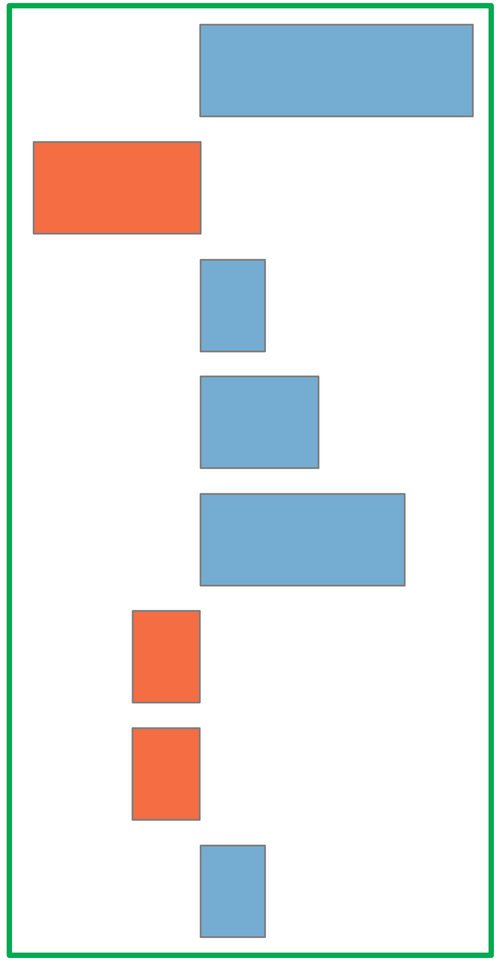


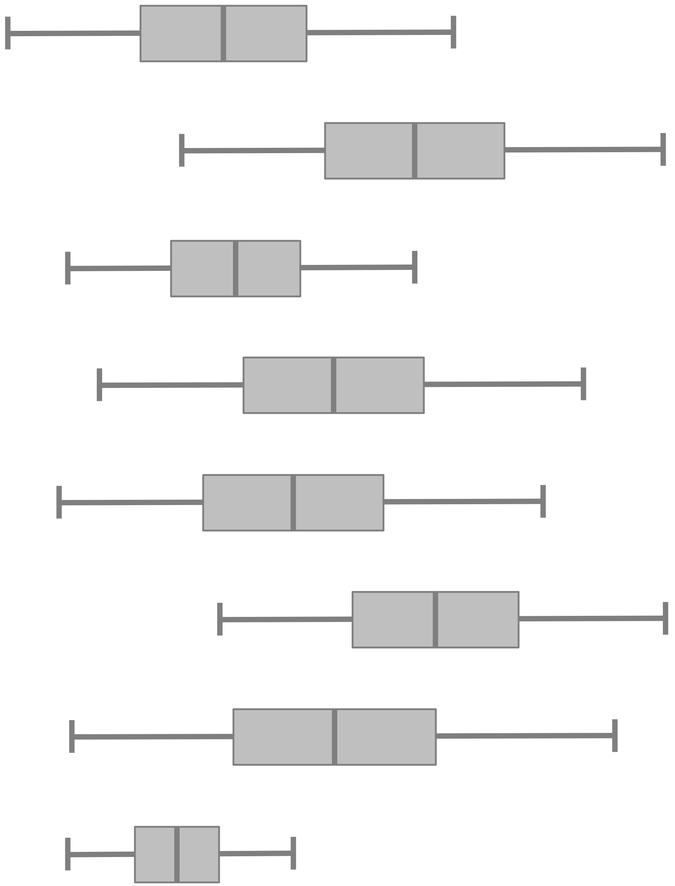
Which is the biggest intersection?
Sort By: Cardinality

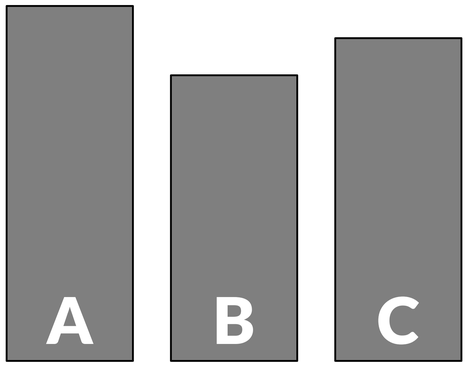


Which is the most ‘surprising’ intersection?
Sort By: Deviation

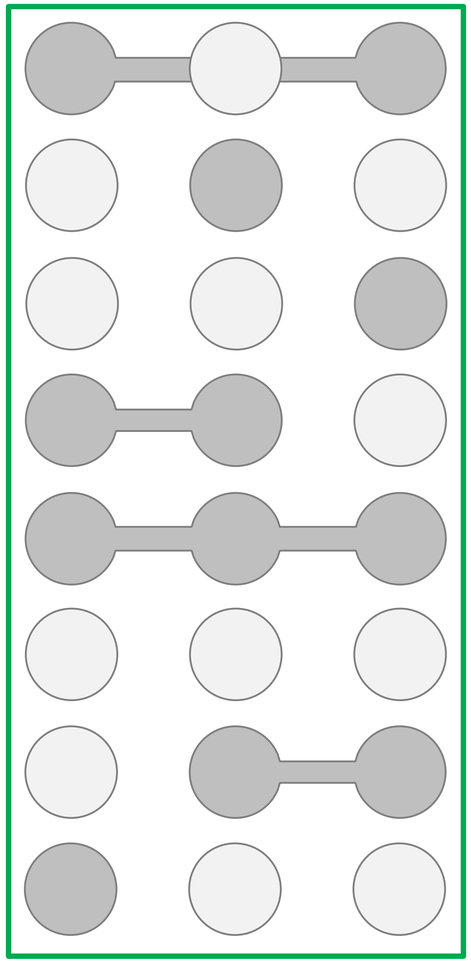


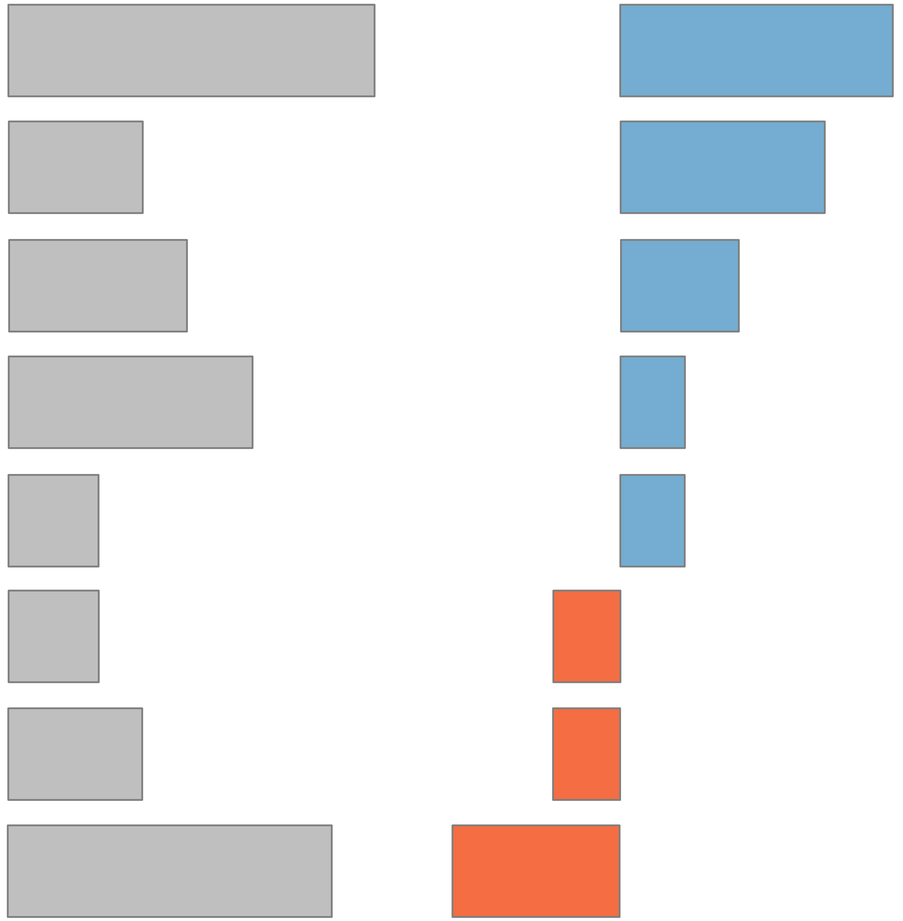


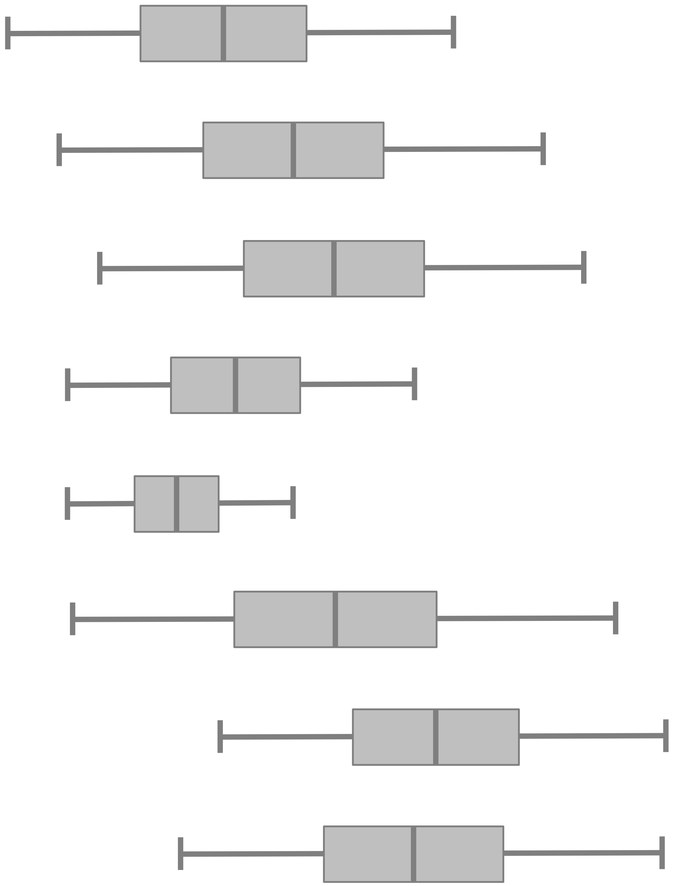


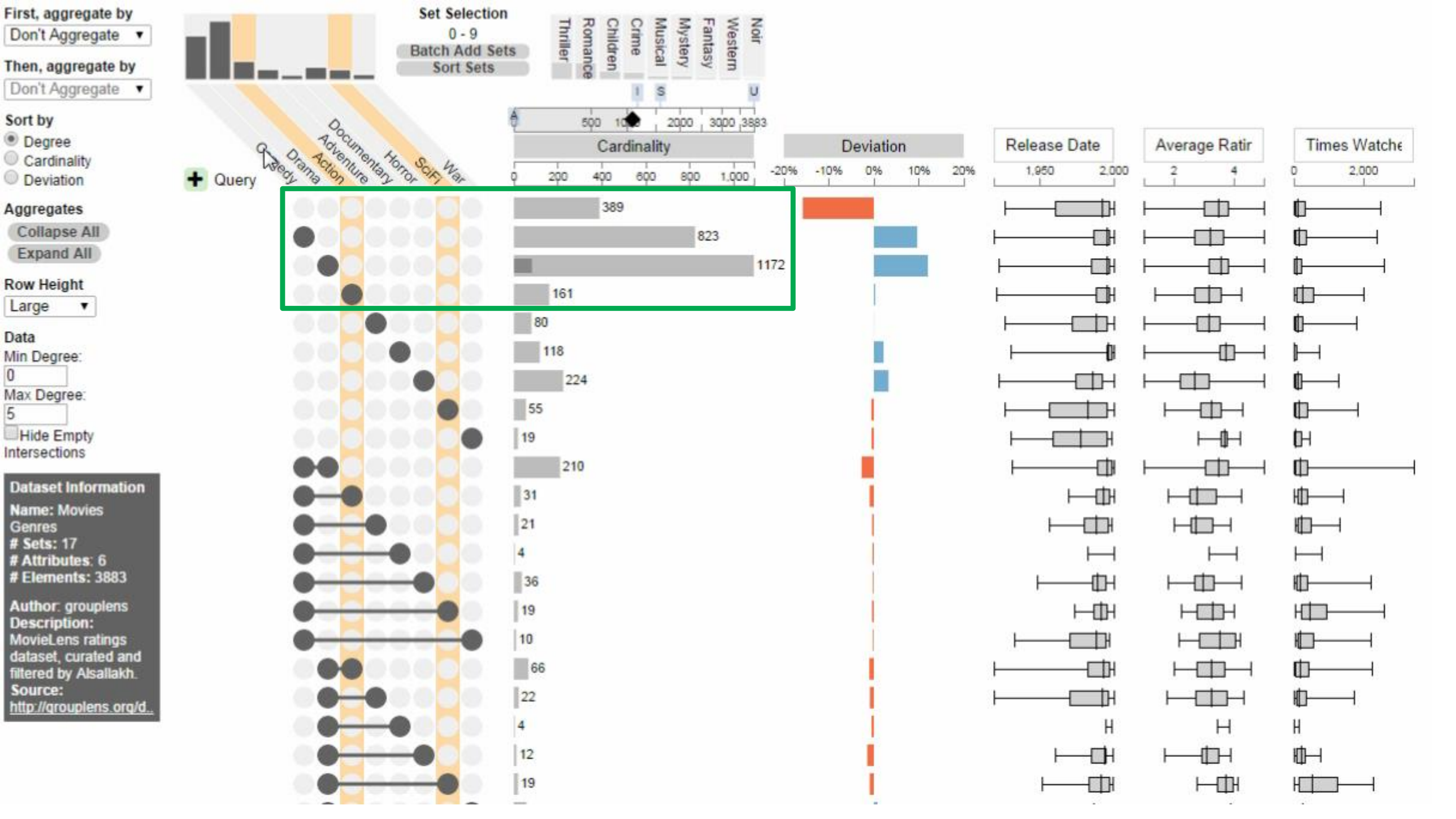


What are the properties of the intersections involving ‘A’?
Sort By: Set

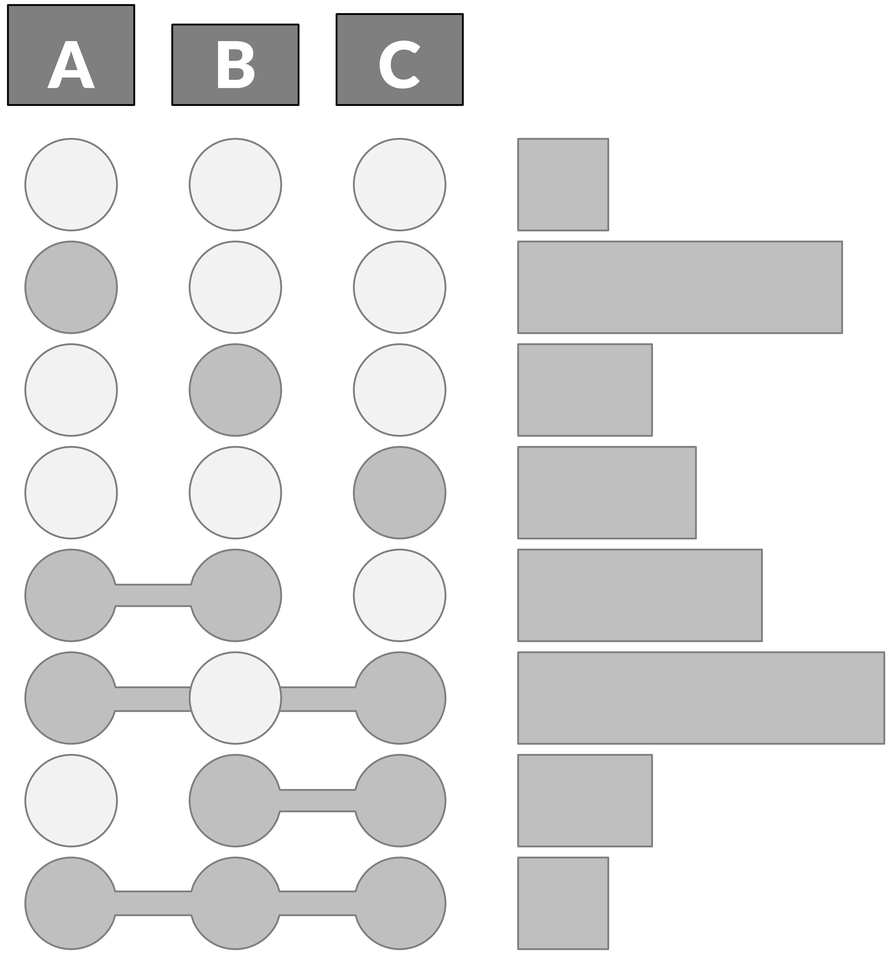




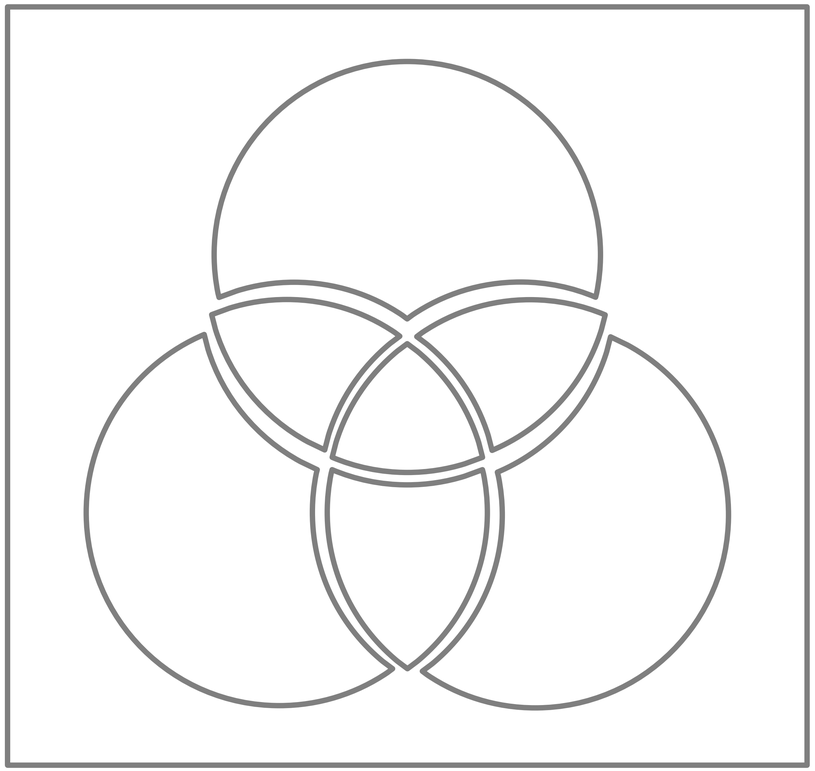


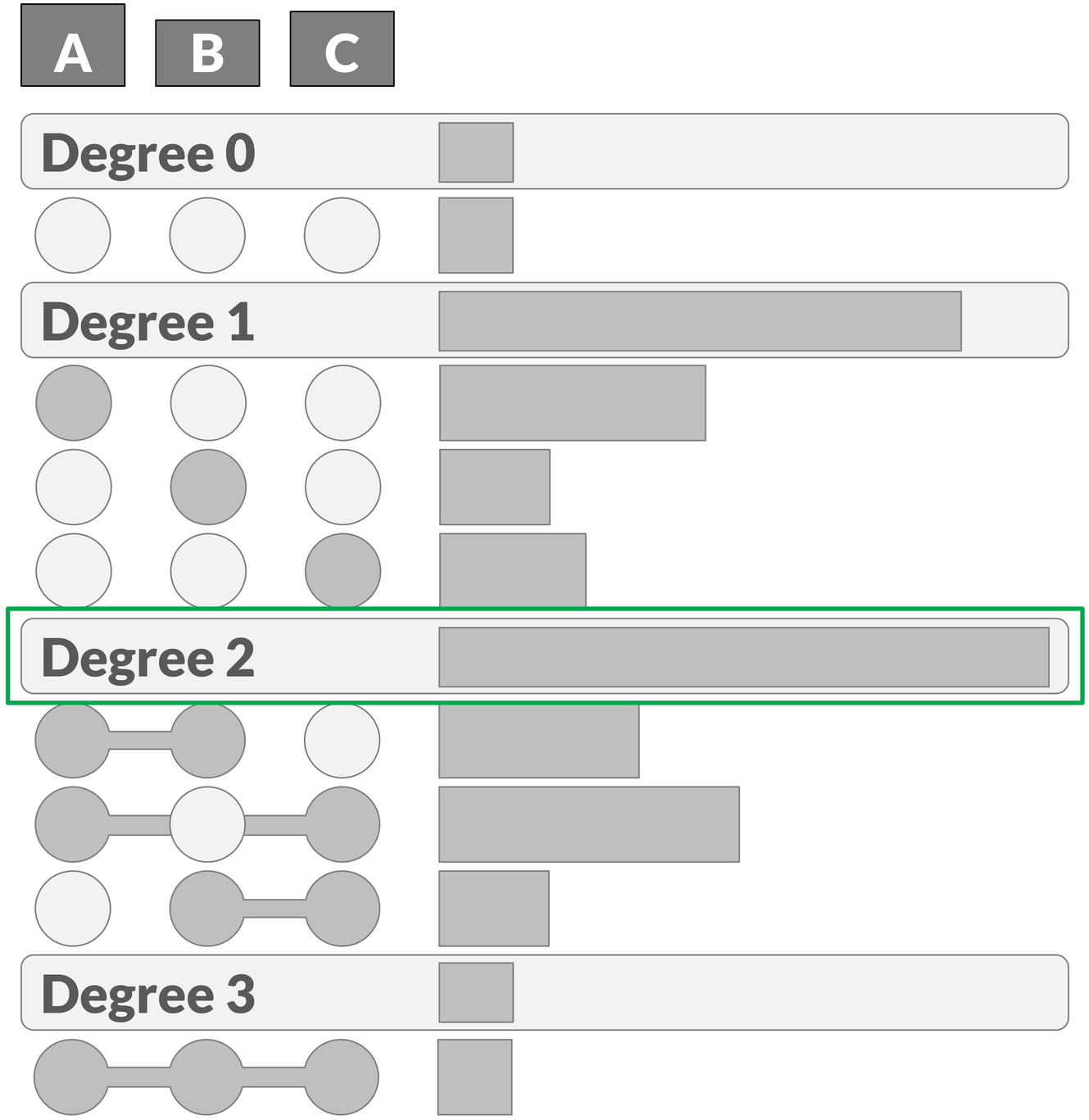

First, aggregate by
Don't Aggregate
Then, aggregate by
Don't Aggregate
Sort by
Degree
Cardinality
Deviation
Aggregates
Collapse All
Expand All
Row Height
Large
Data
Min Degree:
0
Max Degree:
5
Hide Empty Intersections
Dataset Information
Name: Movies Genres
Sets: 17
Attributes: 6
Elements: 3883
Author: 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

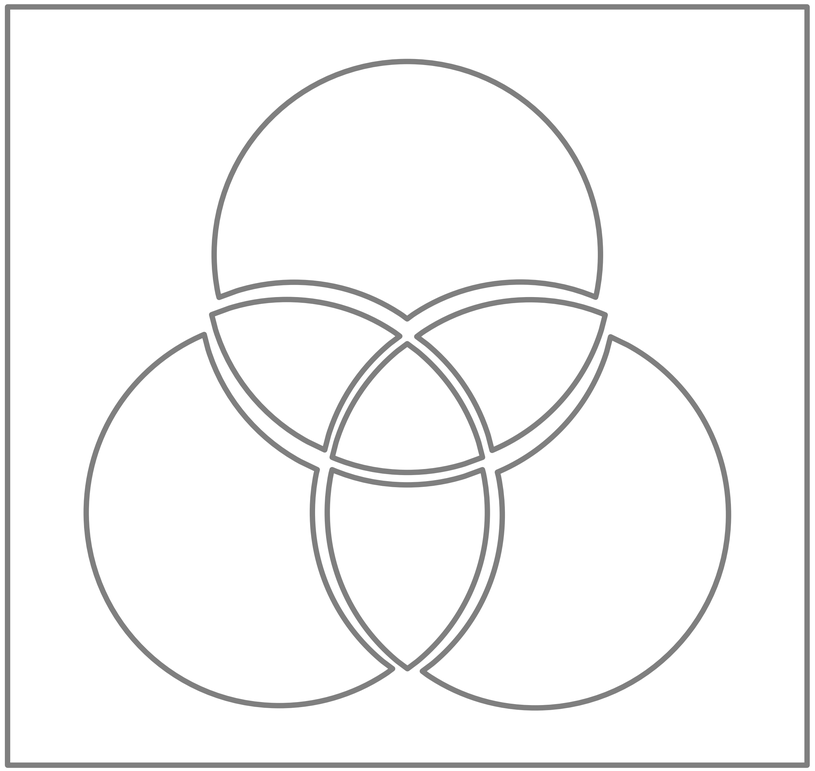


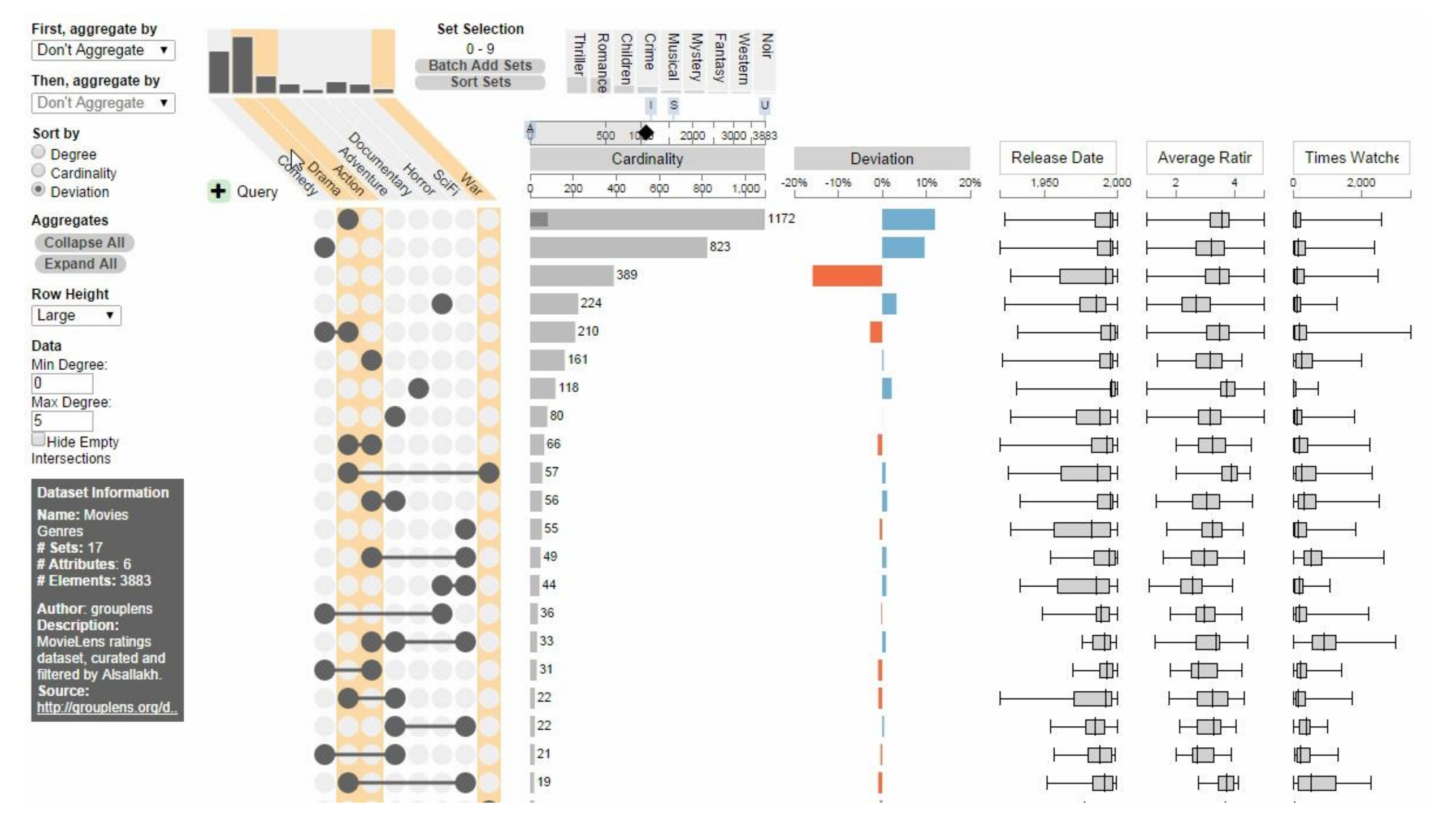
Are most elements in only one set?
Aggregate By: Degree





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




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
1172
823
389
224
210
161
118
80
66
57
56
55
49
44
36
33
31
22
22
21
19

A B C

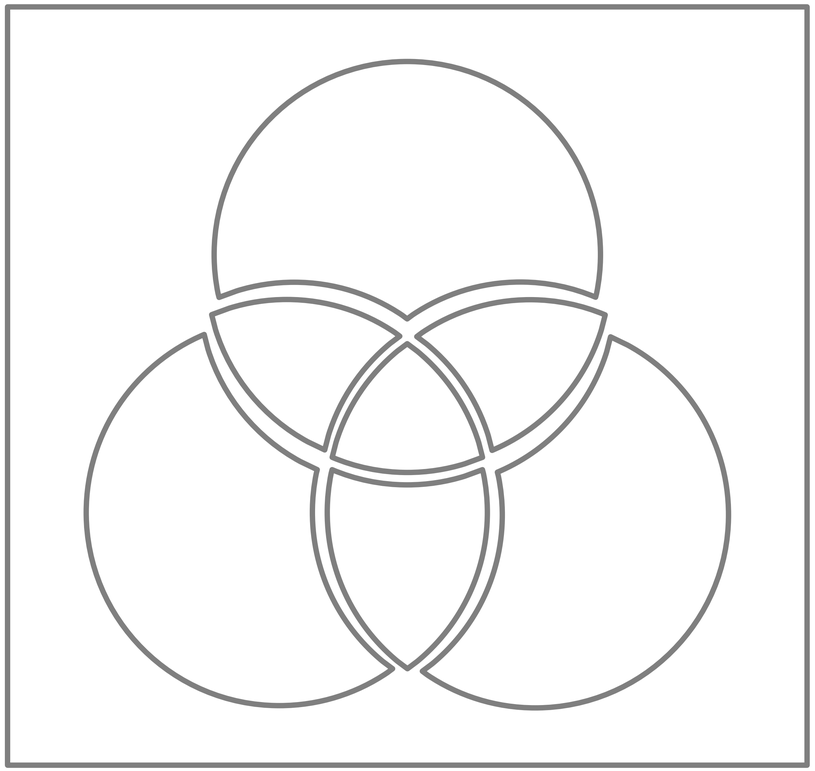
How are the elementsof ‘B’ distributed?
Aggregate By: Set

Degree 0

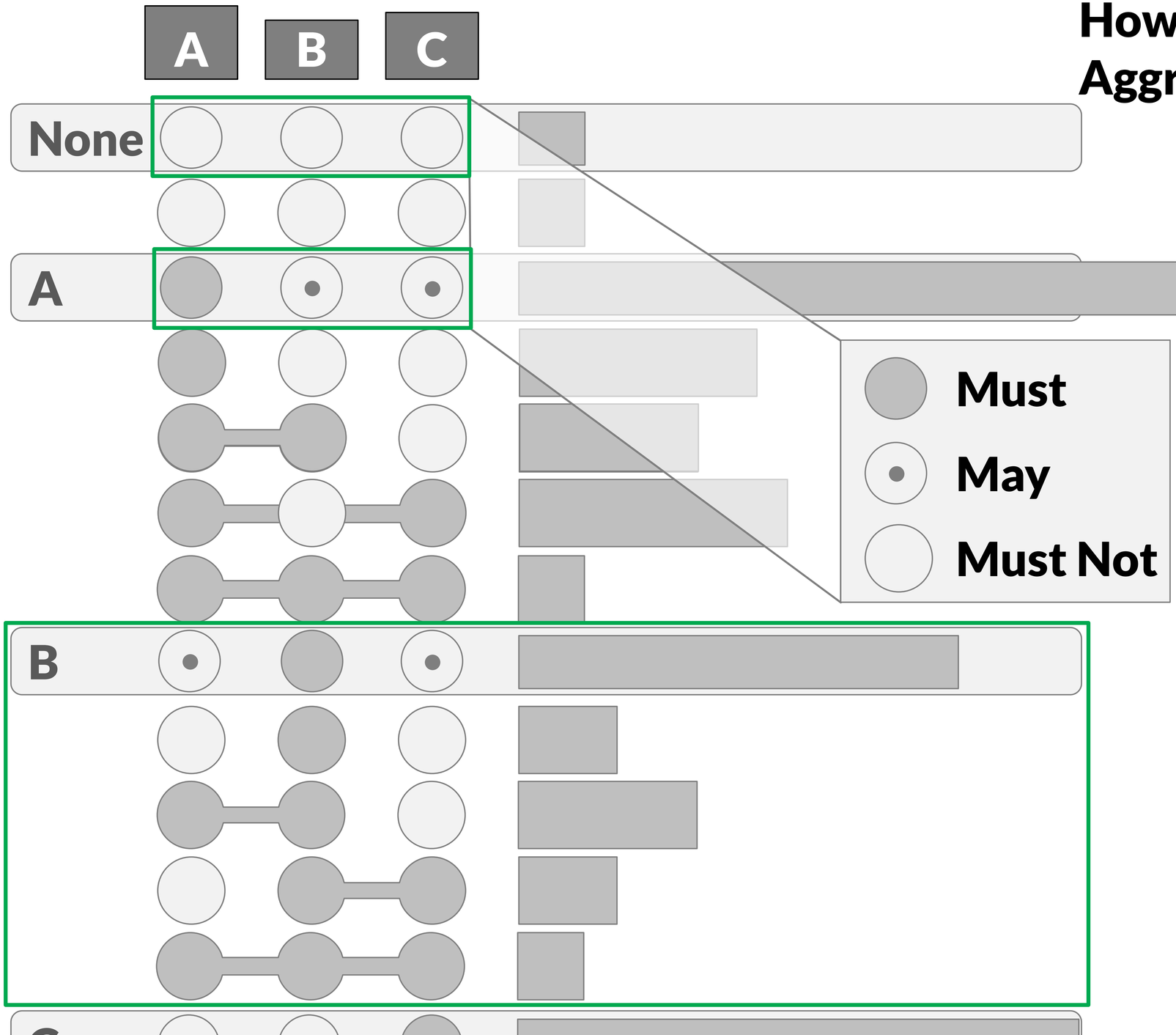
Degree 1

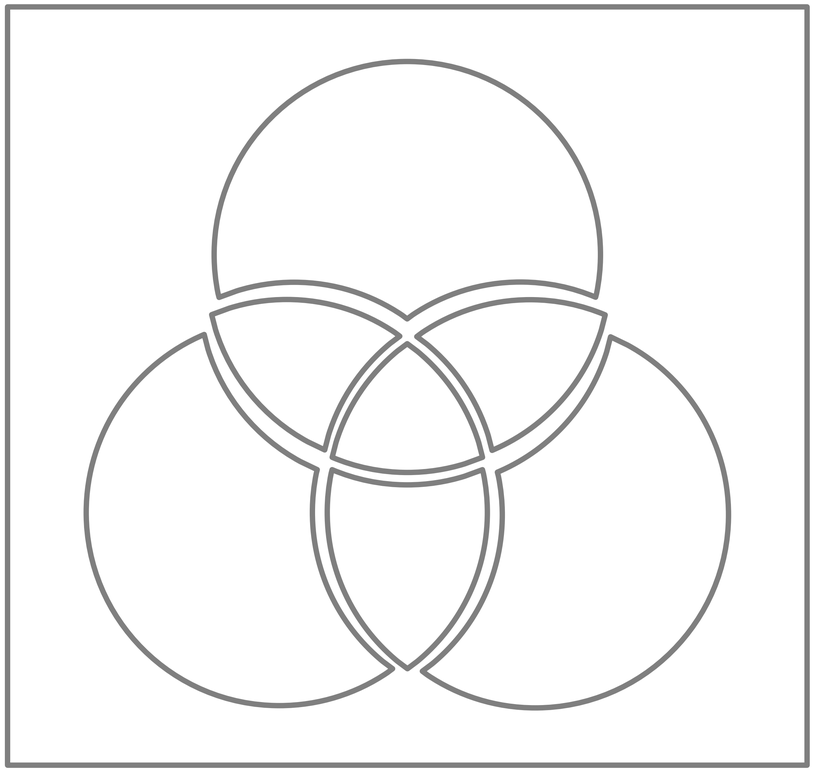
Degree 2

Degree 3

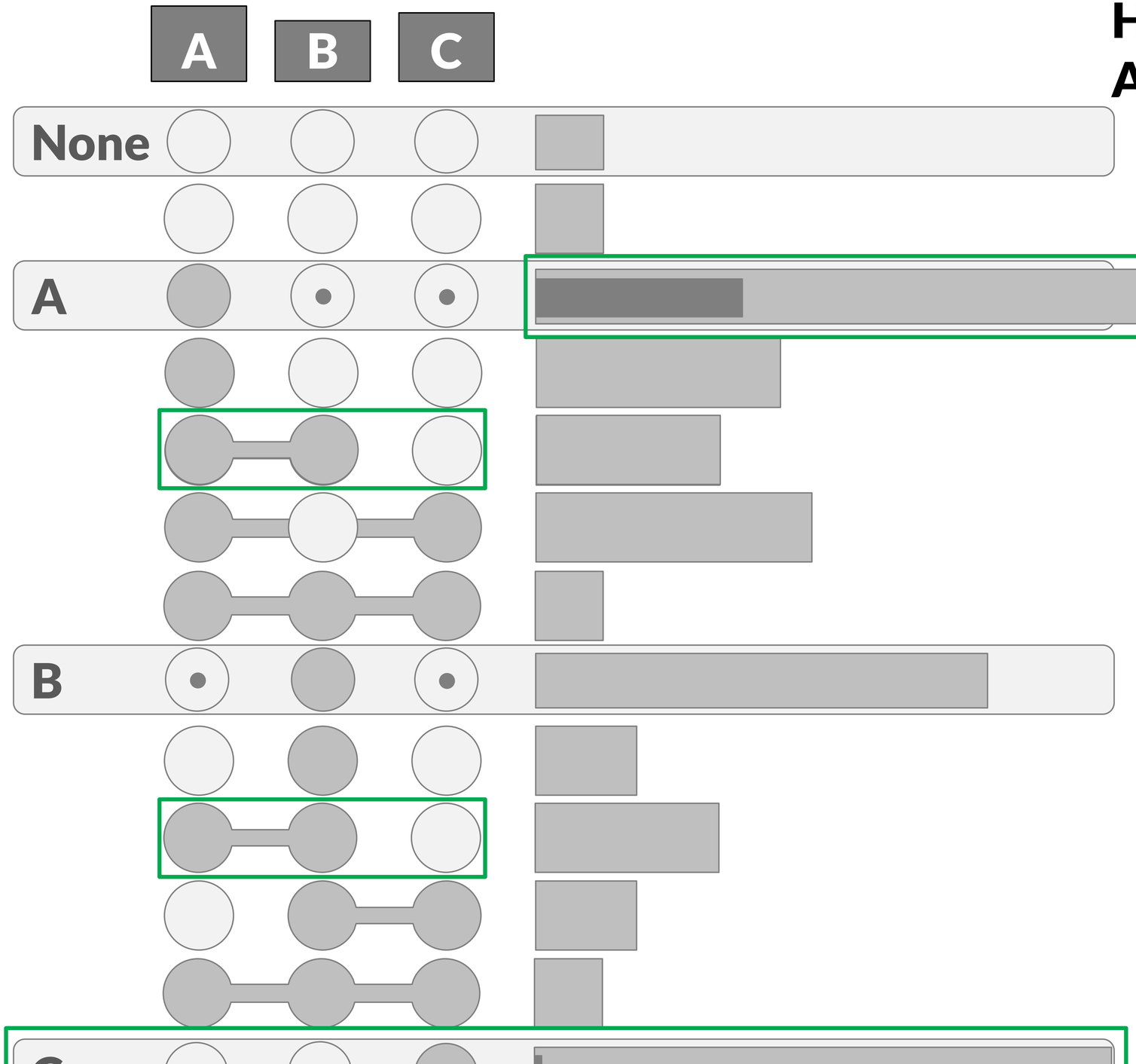


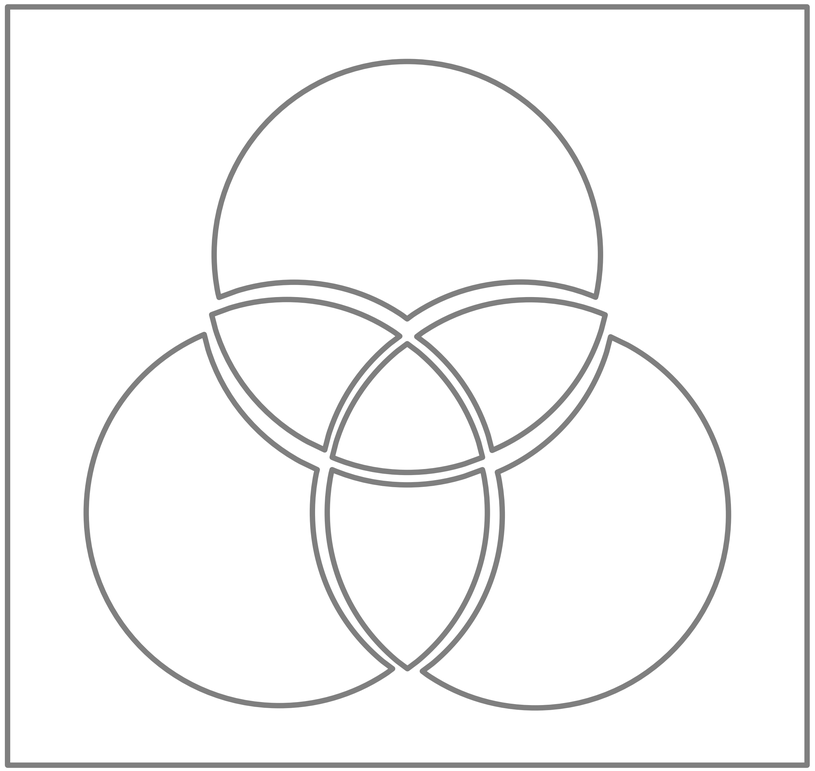
How are the items of ‘B’ distributed?
Aggregate By: Set

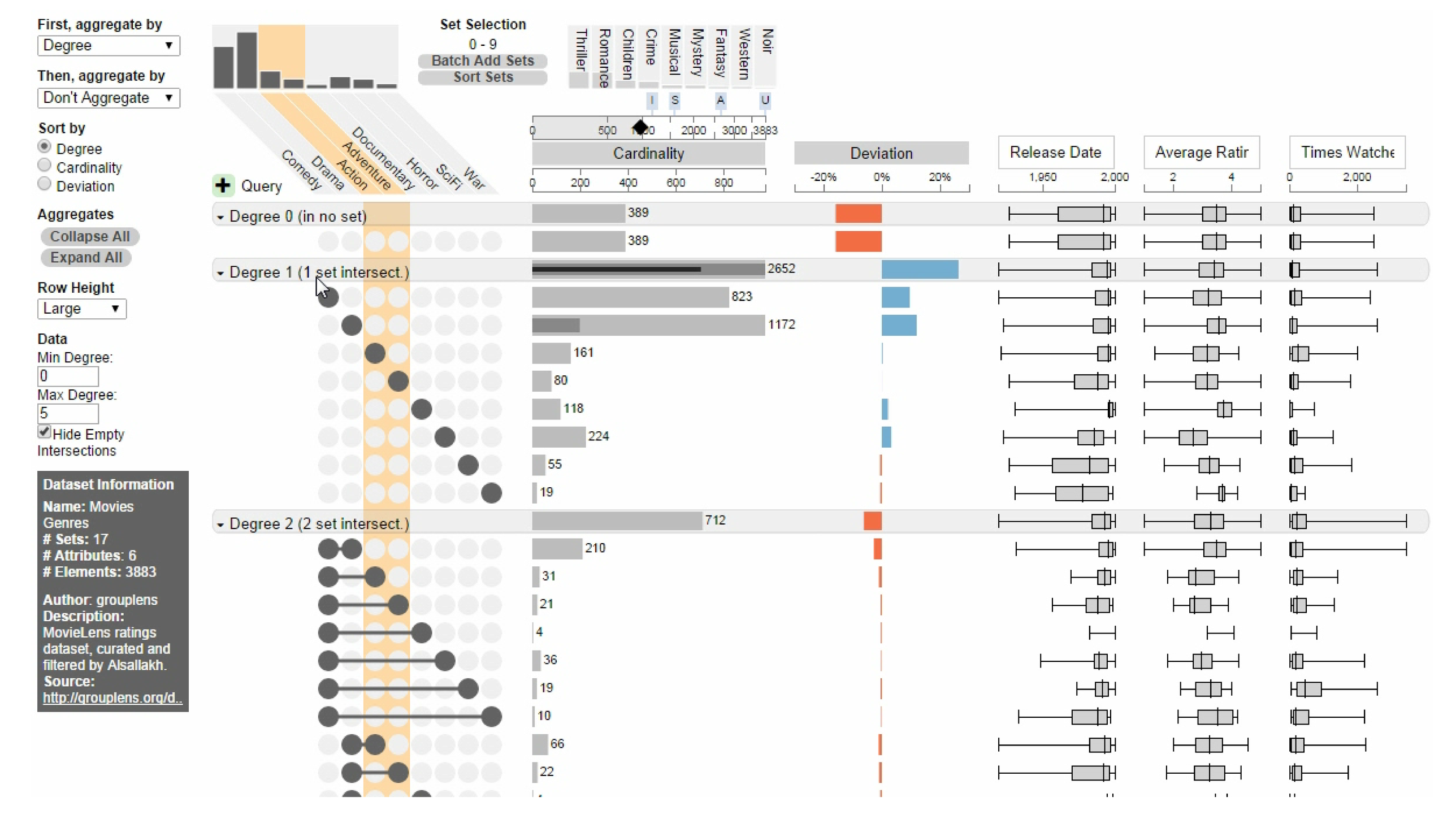




How are the items of ‘B’ distributed?
Aggregate By: Set



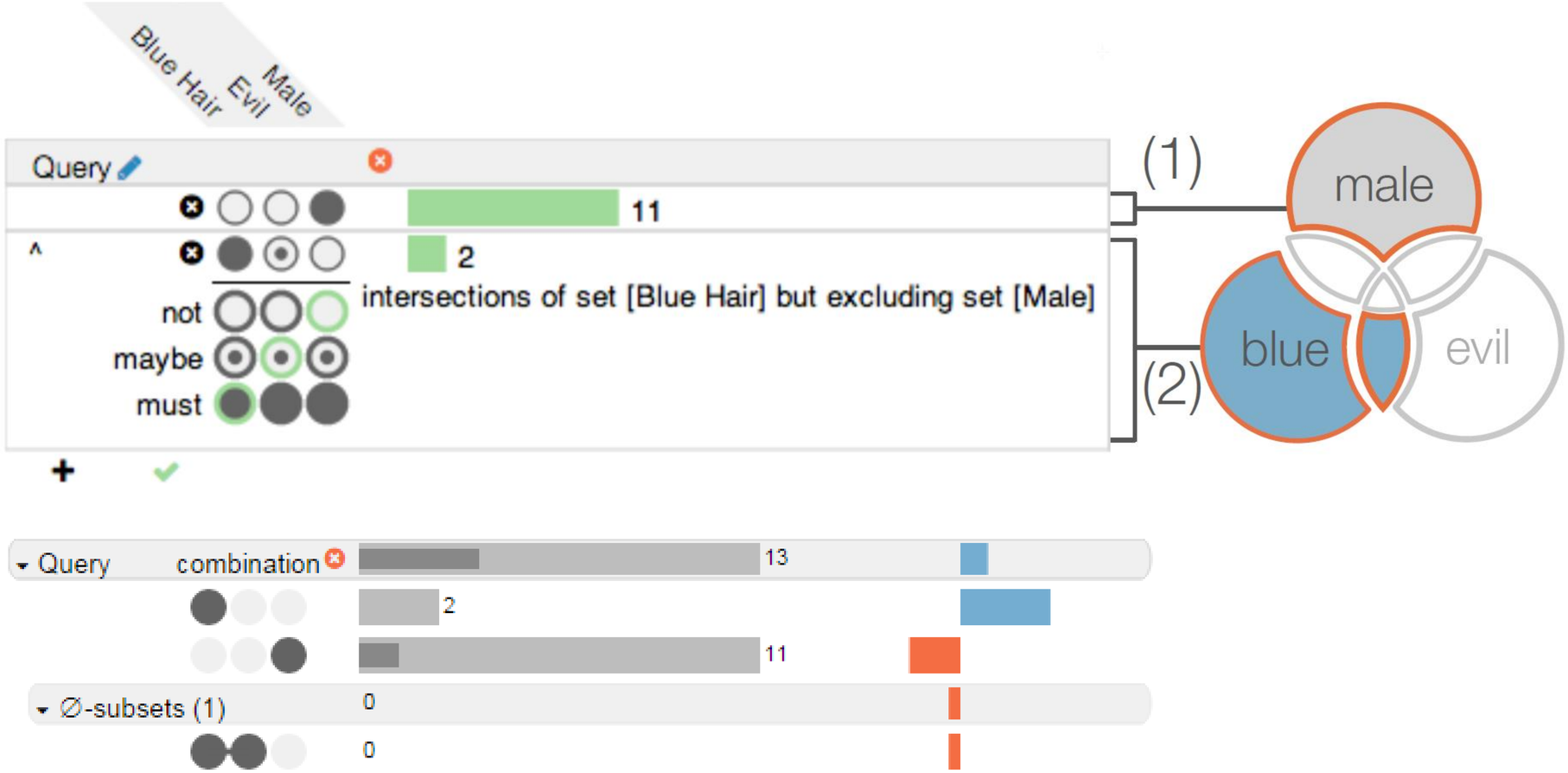


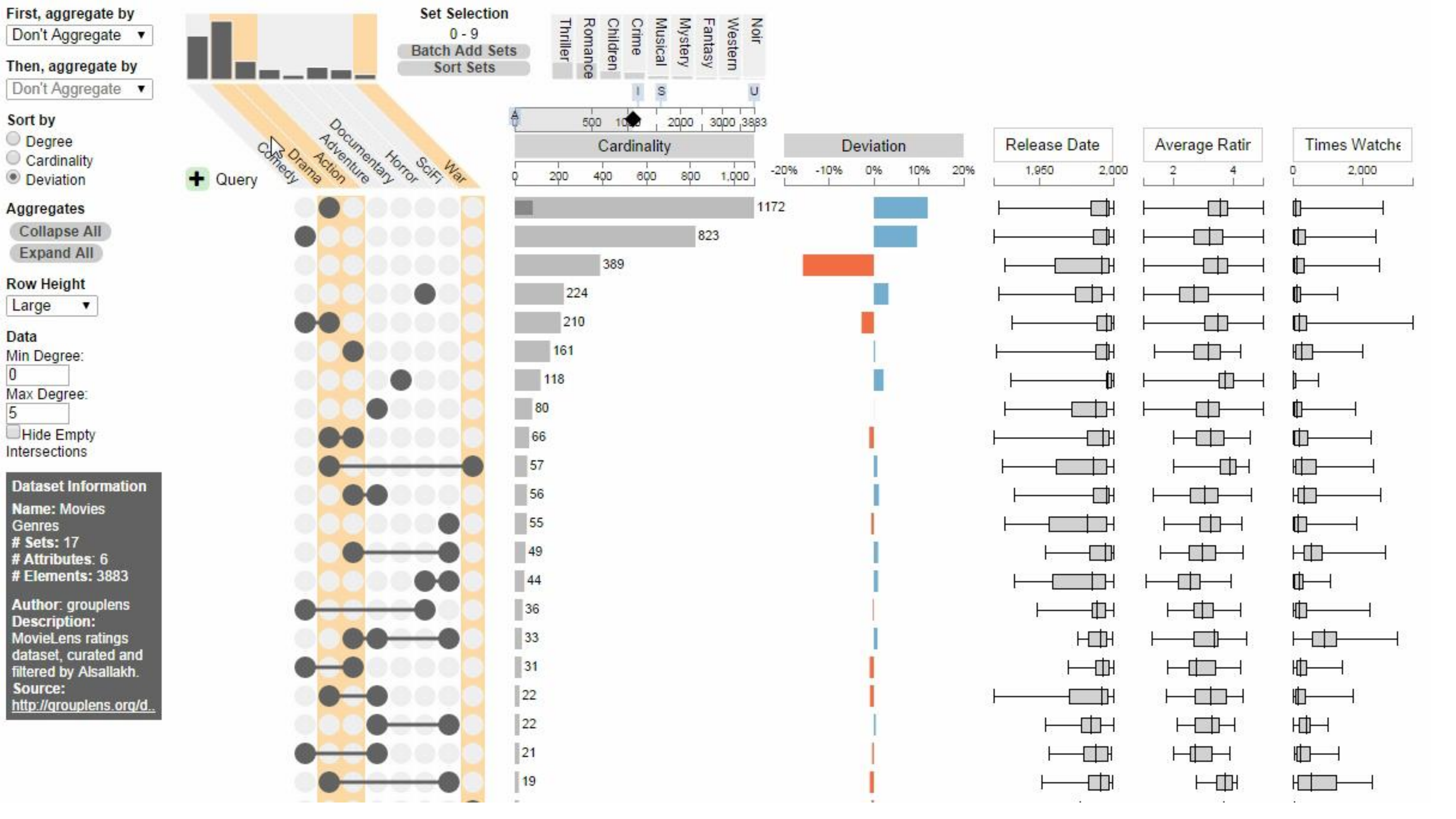

First, aggregate by
Degree
Then, aggregate by
Don't Aggregate
Sort by
Degree
Cardinality
Deviation
Aggregates
Collapse All
Expand All
Row Height
Large
Data
Min Degree:
0
Max Degree:
5
Hide Empty Intersections
Dataset Information
Name: Movies Genres
Sets: 17
Attributes: 6
Elements: 3883
Author: grouplens
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

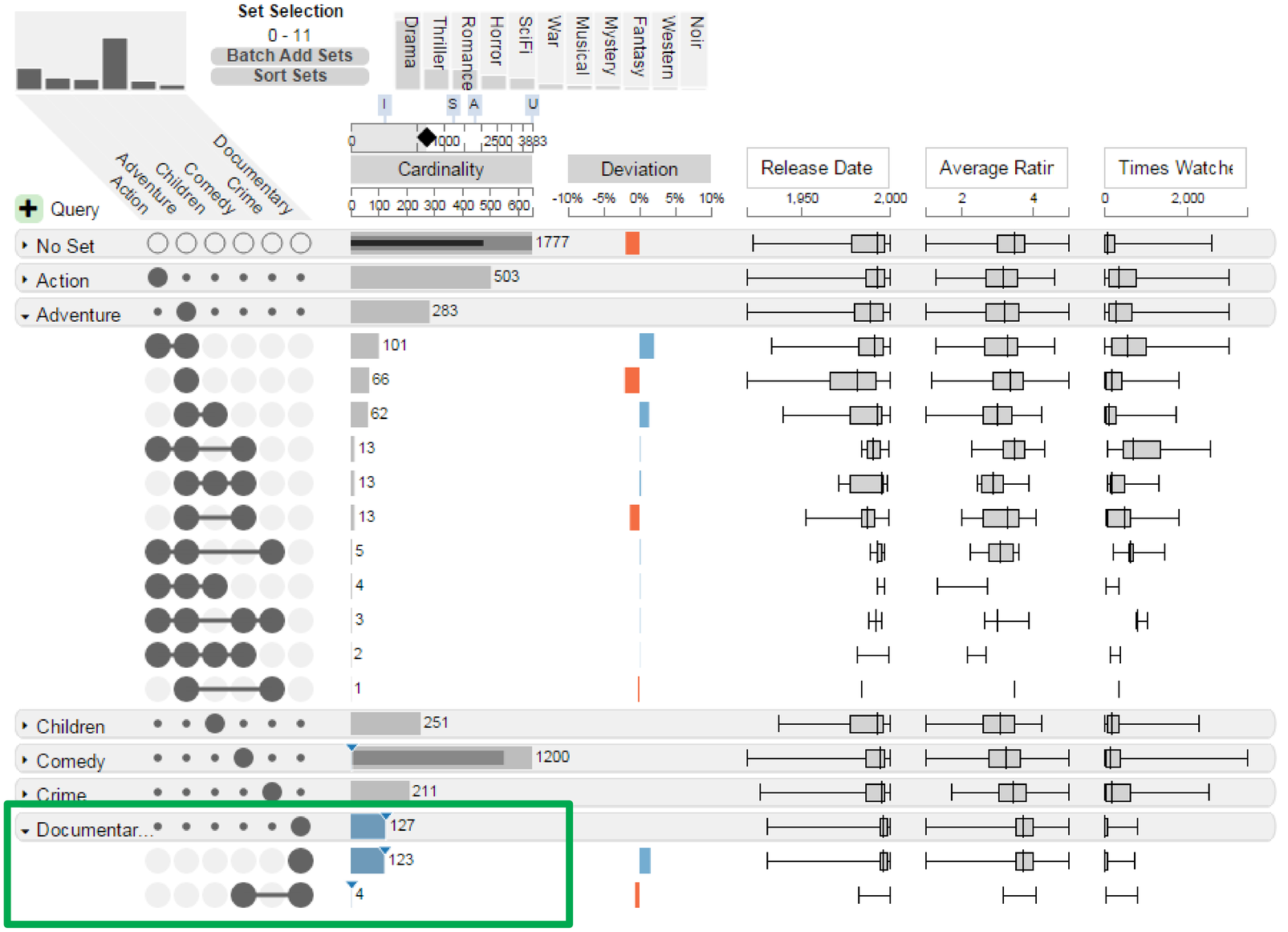
Which Simpsons characters are either male but not evil and don't have blue hair, or are female and have blue hair?

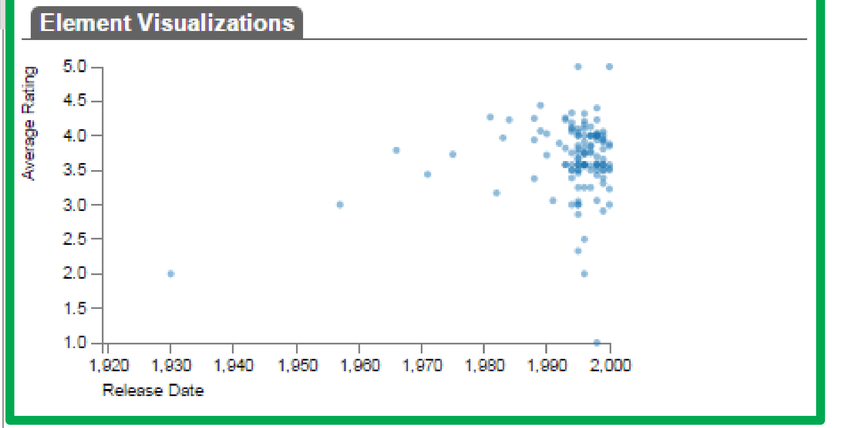
Answer with Queries




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
1172
823
389
224
210
161
118
80
66
57
56
55
49
44
36
33
31
22
22
21
19

Exploring Attributes

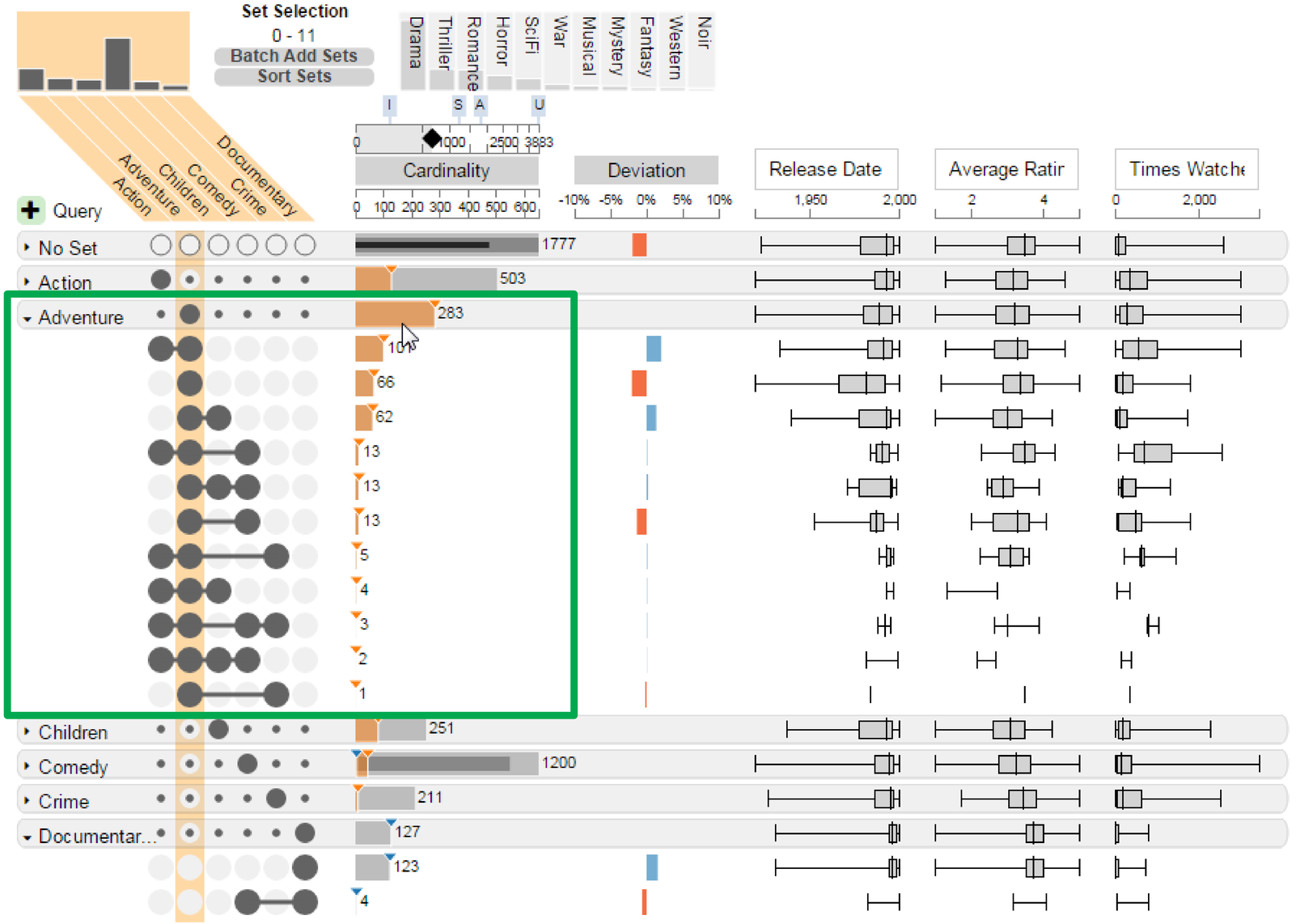




Query Results

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 the attribute of two intersections compare?



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 the attribute of two intersections compare?

Applications

Molecular Biology

Social Network Analysis

Economics

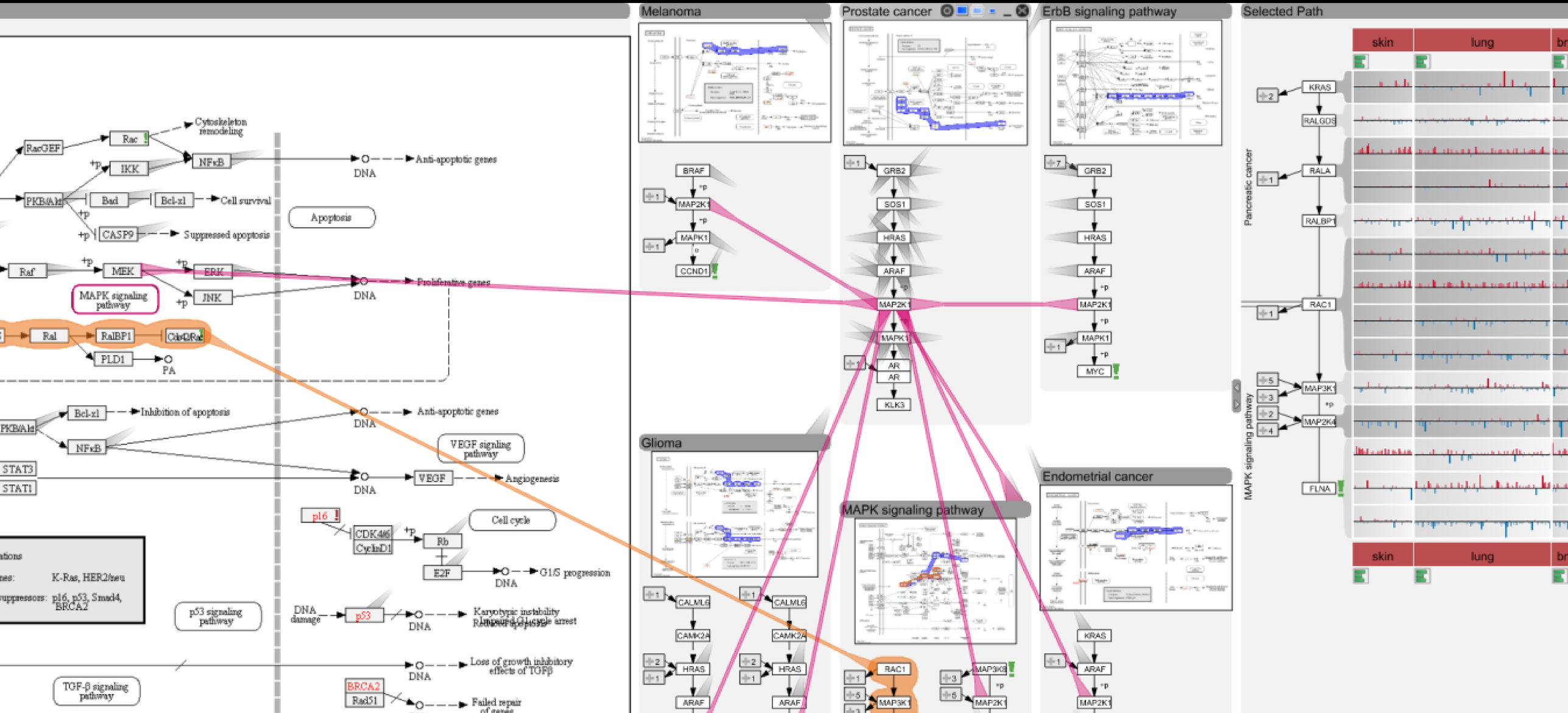
Drug Discovery

….

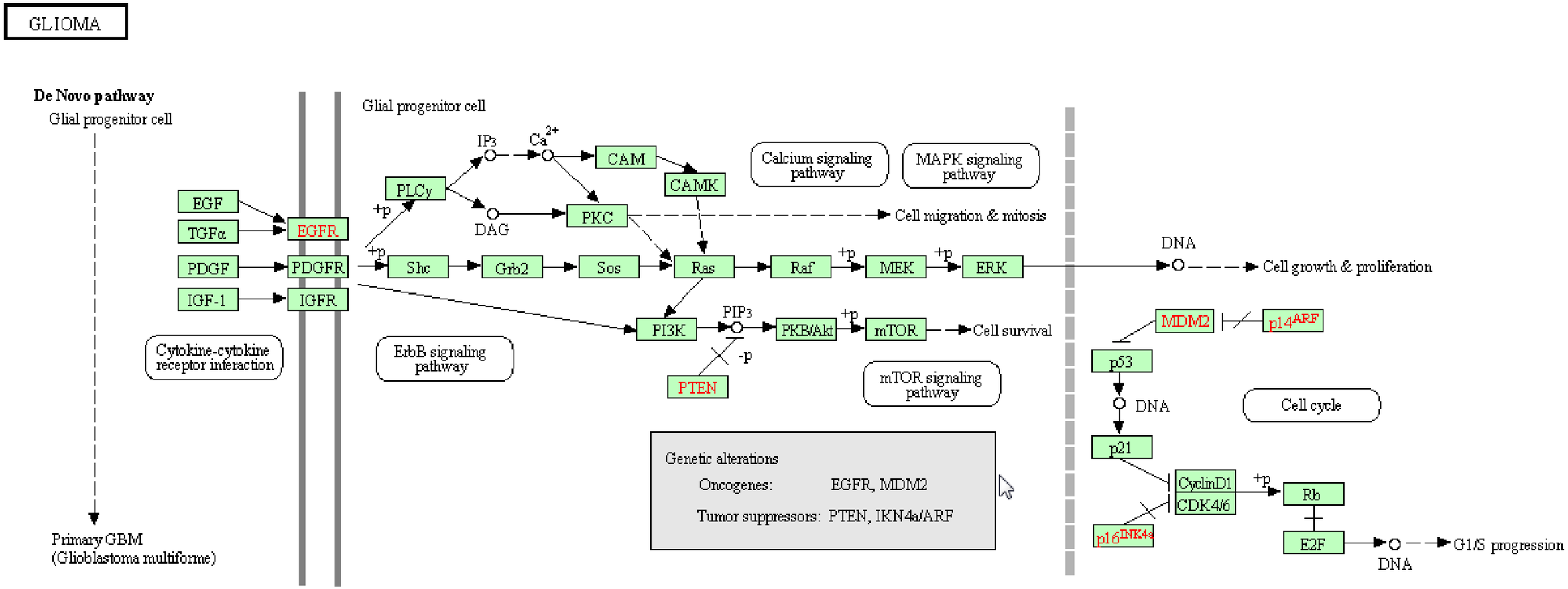
http://vcg.github.io/upset

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

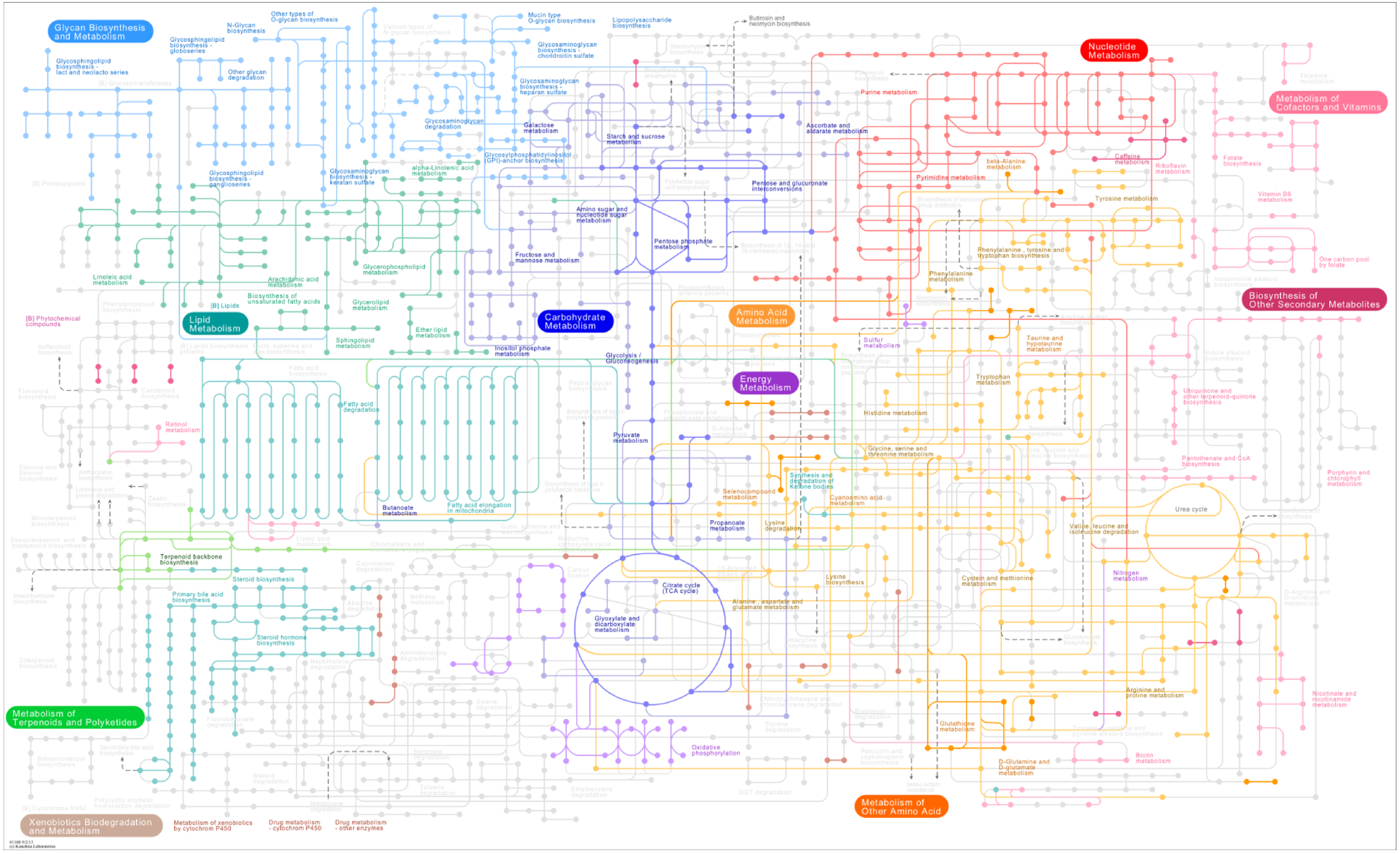
Pathways – Entourage



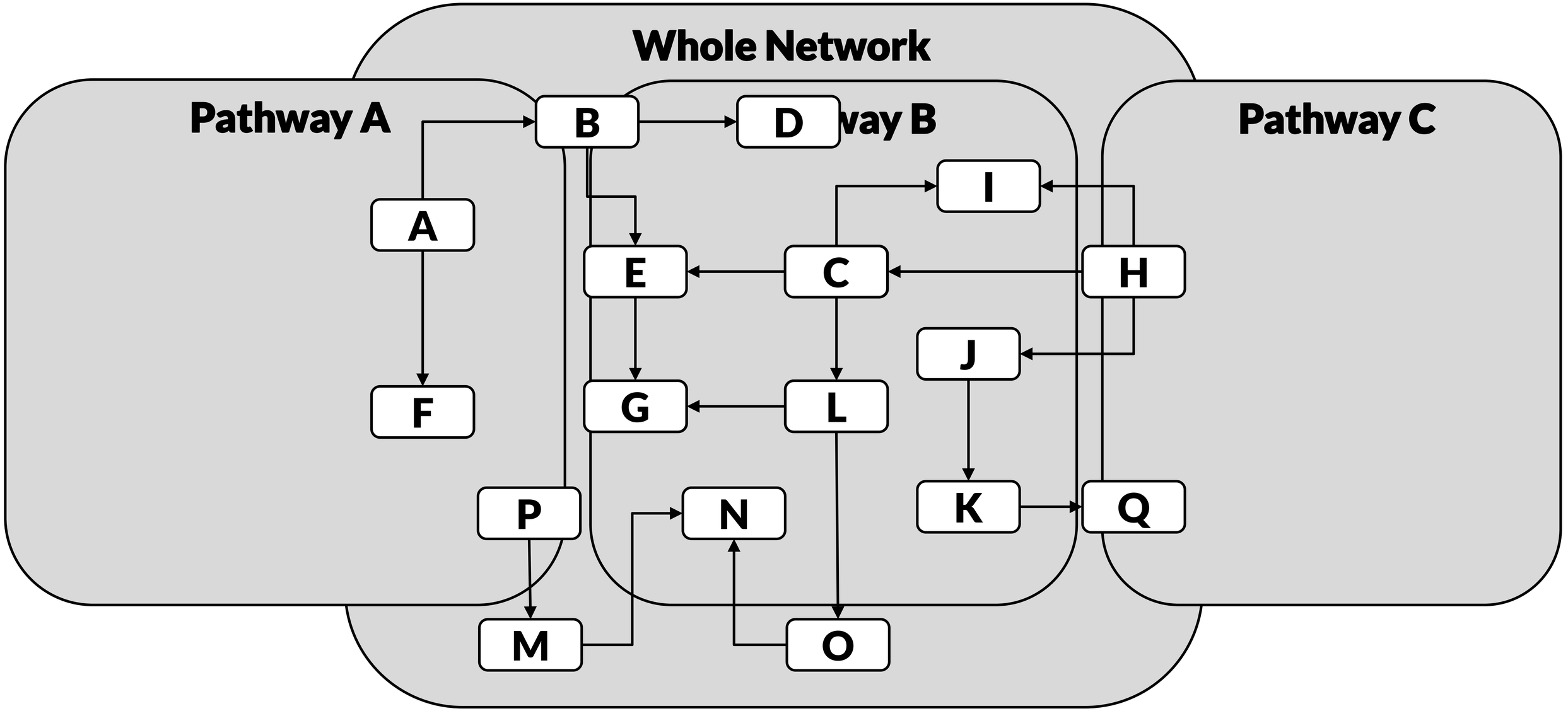
A Pathway


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
PLCγ
IP3
Ca2+
DAG
CAM
CAMK
PKC
Calcium signaling pathway
MAPK signaling pathway
Cell migration & mitosis
+p
Shc
Grb2
Sos
Ras
Raf
MEK
ERK
DNA
Cell growth & proliferation
ErbB signaling pathway
PI3K
PIP3
PKB/Akt
mTOR
Cell survival
PTEN
-p
mTOR signaling pathway
Genetic alterations
Oncogenes: EGFR, MDM2
Tumor suppressors: PTEN, IKN4a/ARF
MDM2
p14ARF
p53
DNA
Cell cycle
p21
CyclinD1
CDK4/6
Rb
p16INK4a
E2F
G1/S progression
DNA

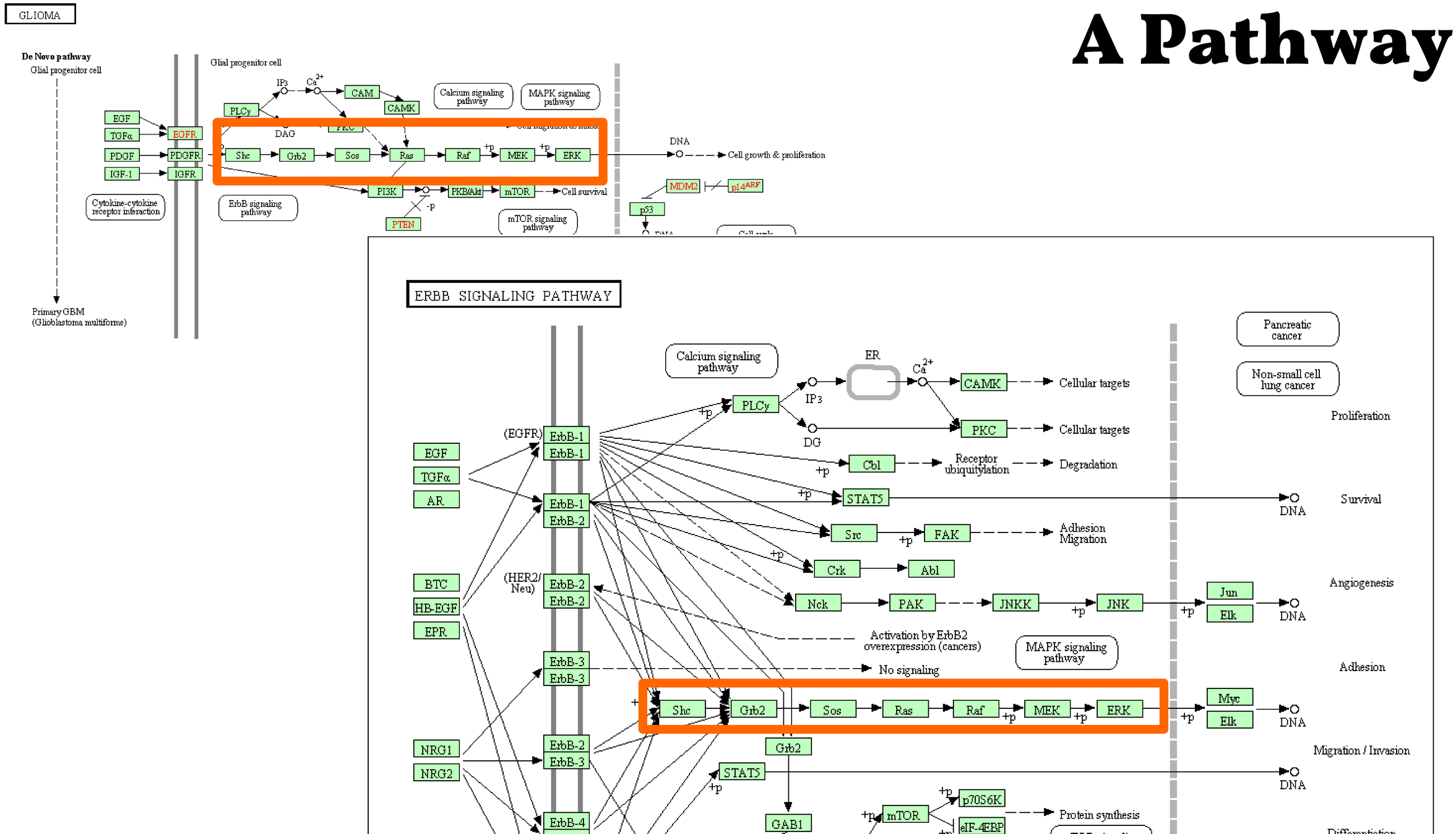
The bigger picture



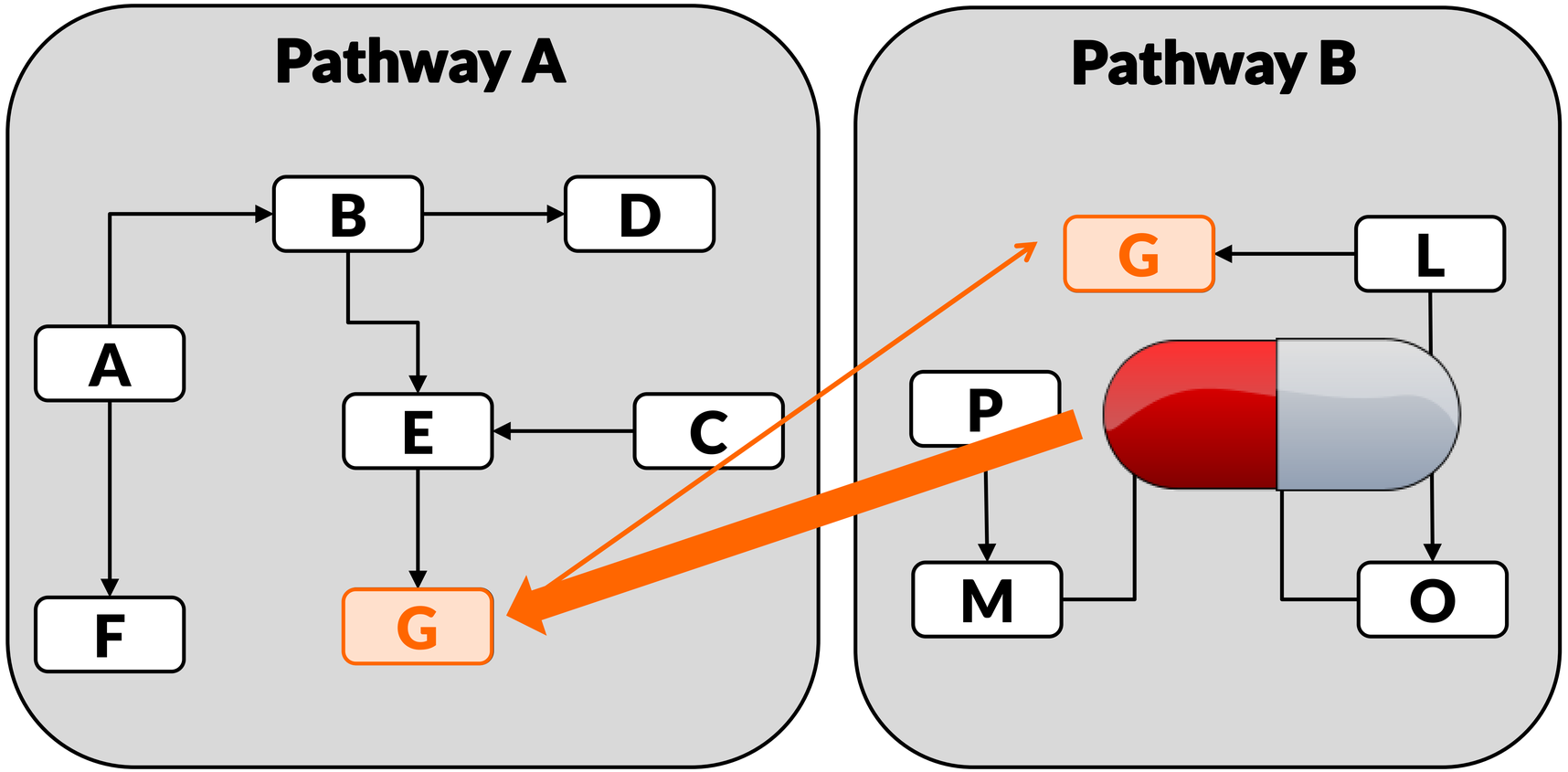
Background


Whole Network
Pathway A
Pathway B
Pathway C
A
B
D
E
C
H
I
J
F
G
L
K
Q
P
N
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+
IP3
DG
CAMK
PKC
Cellular targets
Cbl
Receptor ubiquitylation
Degradation
STAT5
Src
FAK
Adhesion Migration
Crk
Abl
Nck
PAK
JNKK
JNK
Jun
Elk
Activation by ErbB2 overexpression (cancers)
No signaling
MAPK signaling pathway
Myc
Elk
Grb2
STAT5
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
DNA

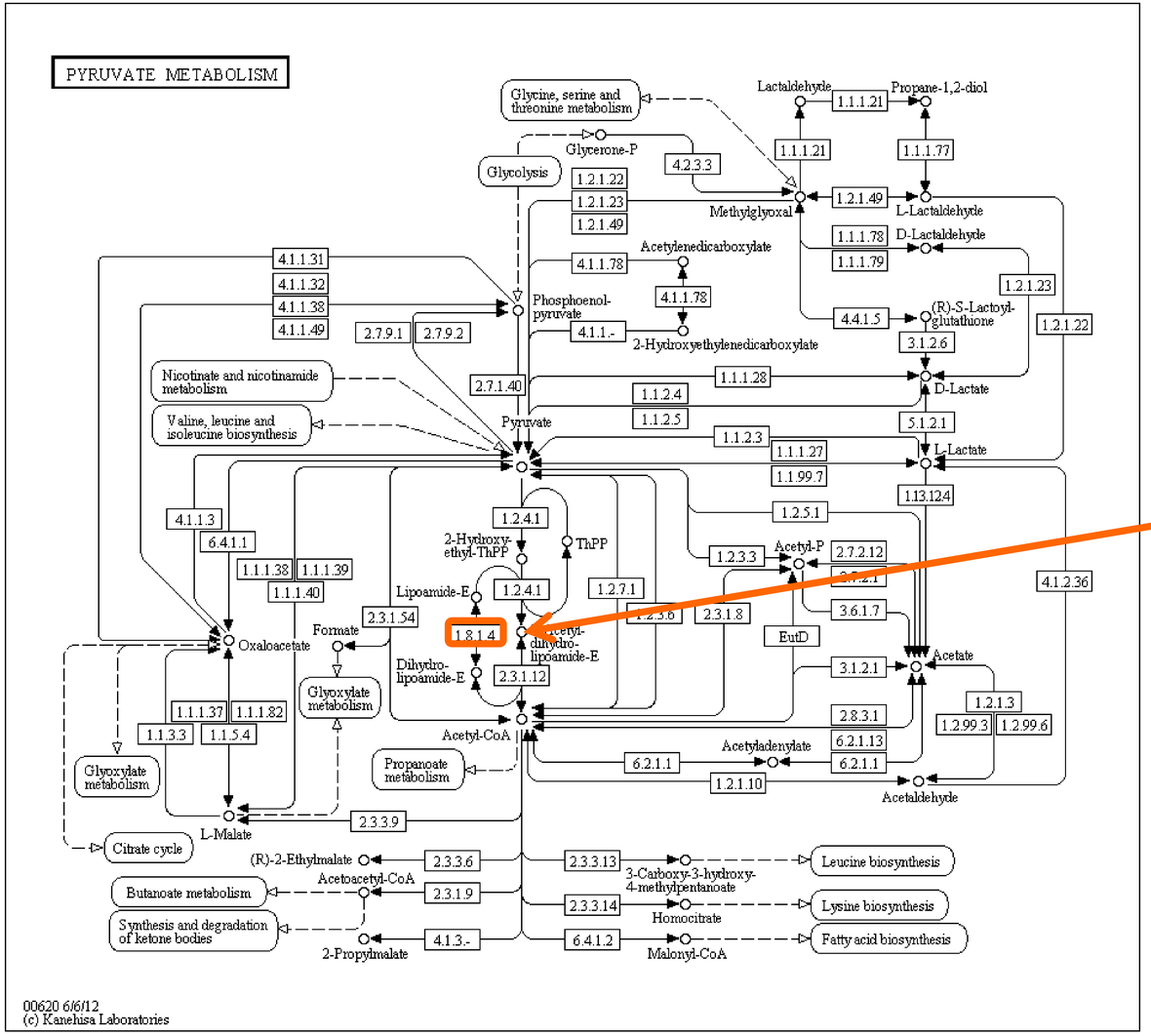
Challenges

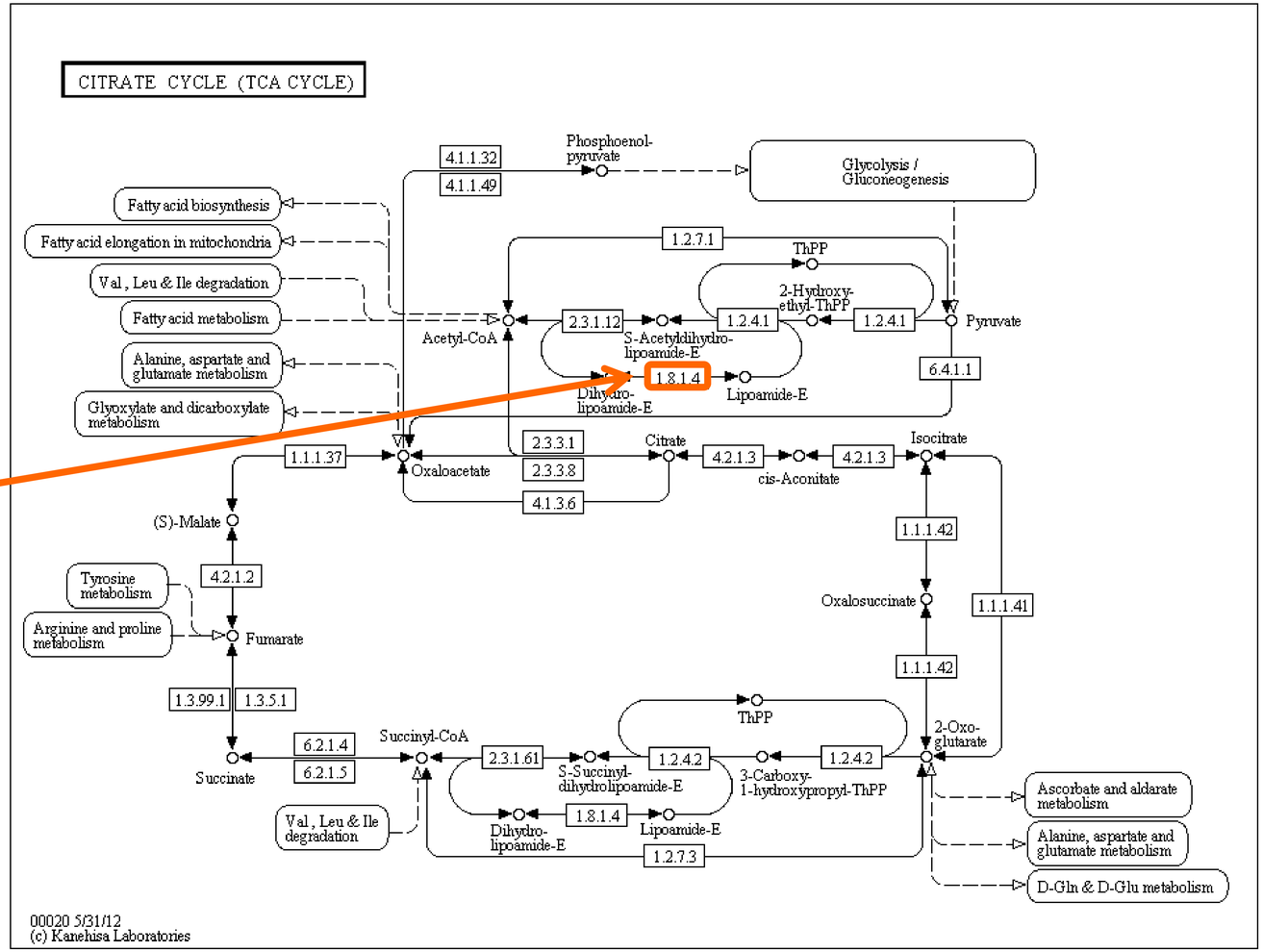


Drug side-effects

Drug repositioning

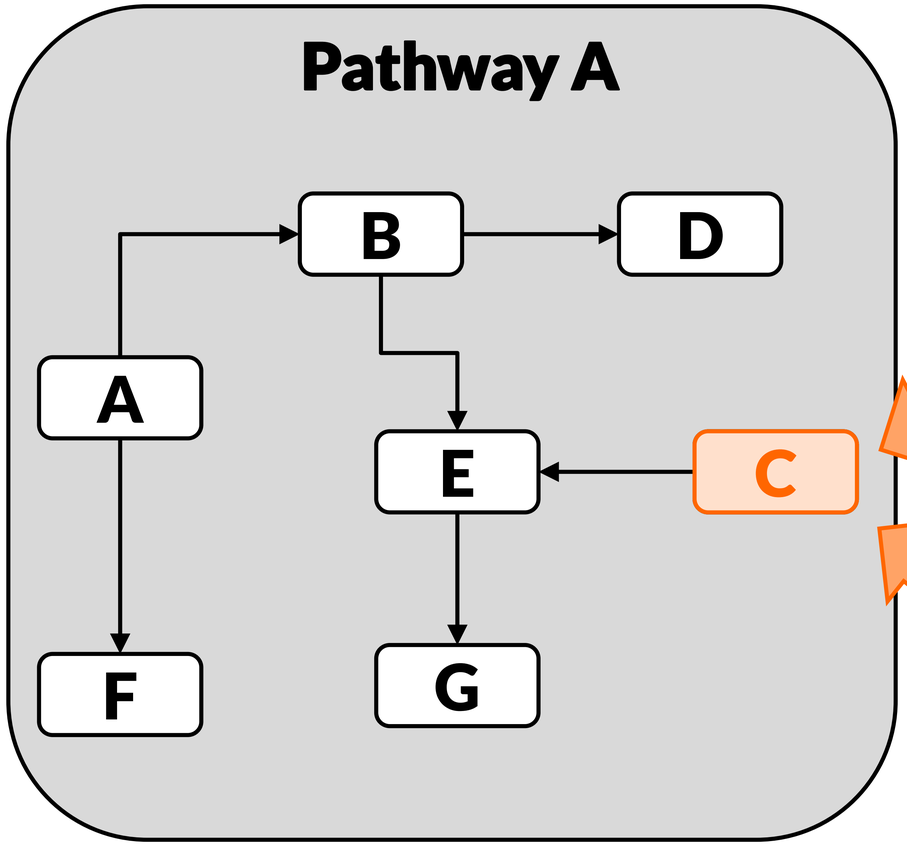
Challenges





How to visualize pathway relationships?

Challenges



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	…
…	…	…	…	

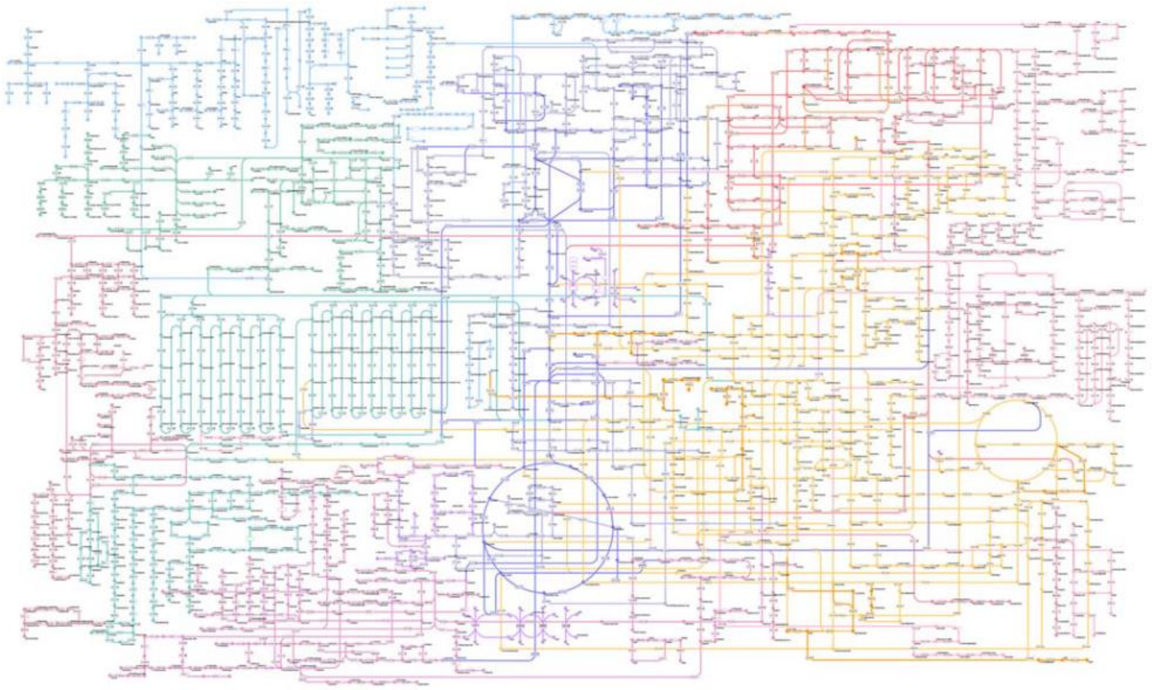
Node	Sample 1	Sample 2	Sample 3	…
A	low	low	very high	…
B	normal	low	high	…
C	high	very low	normal	…
…	…	…	…	

How to visualize experimental data on pathways?

How to visualize pathway relationships?

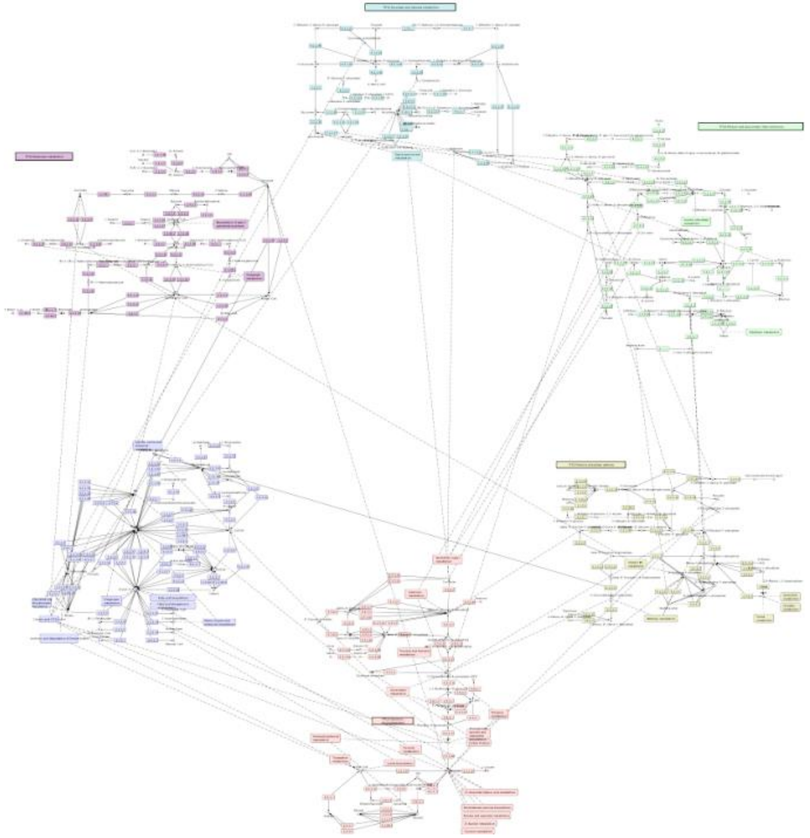
Approaches

Whole Network



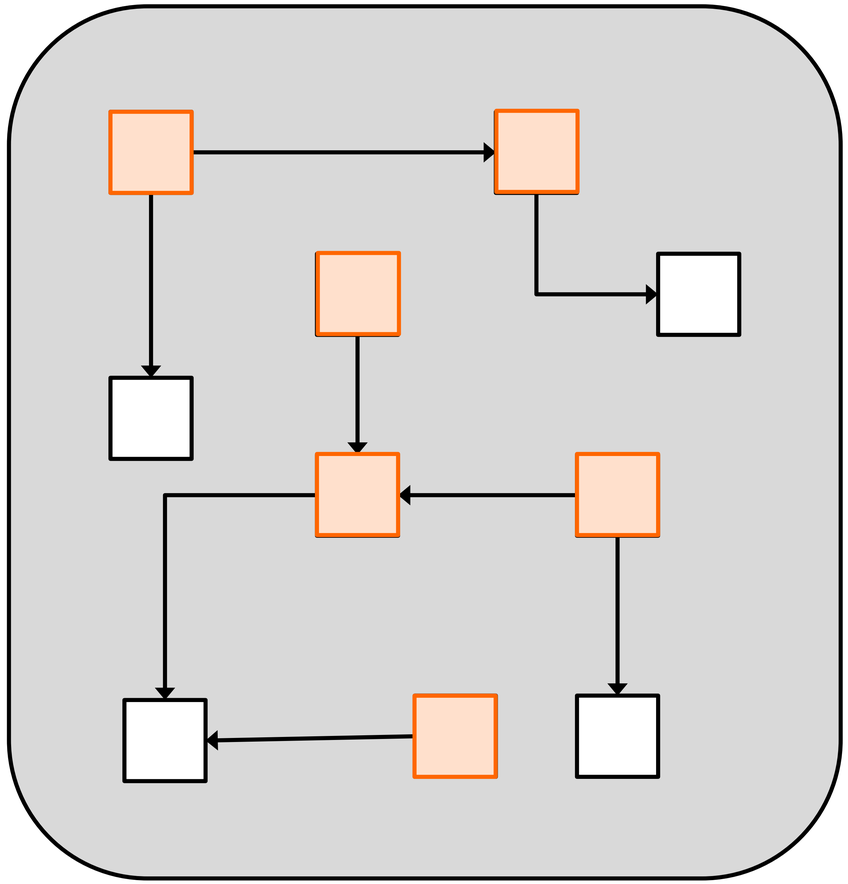
[Kono2009]

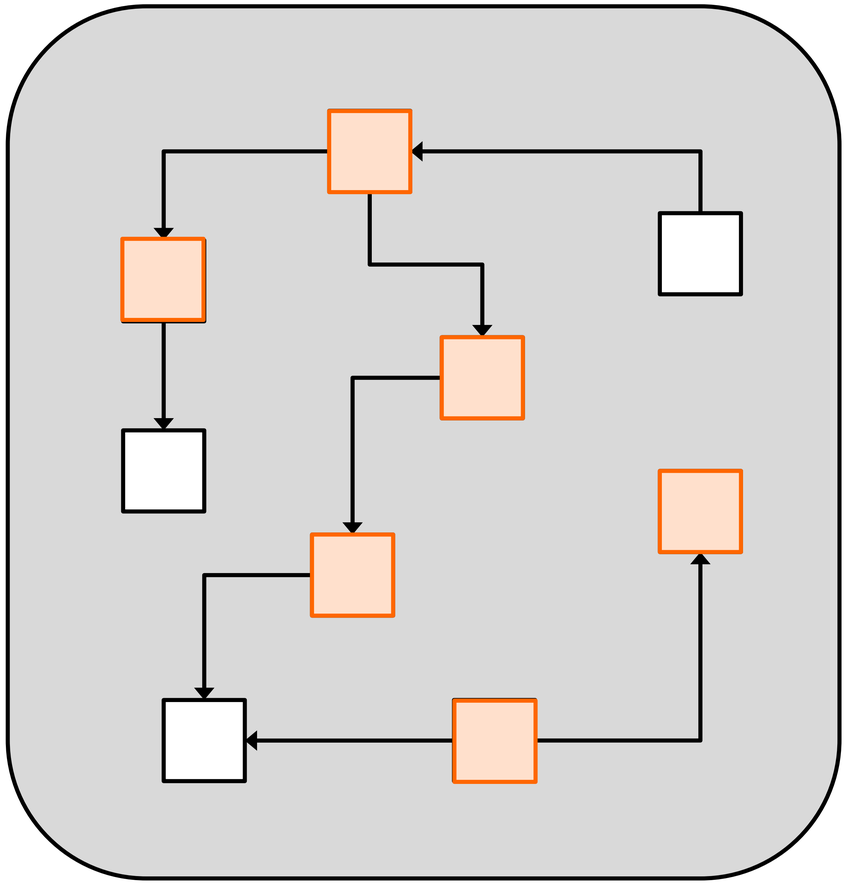
Connected Pathways



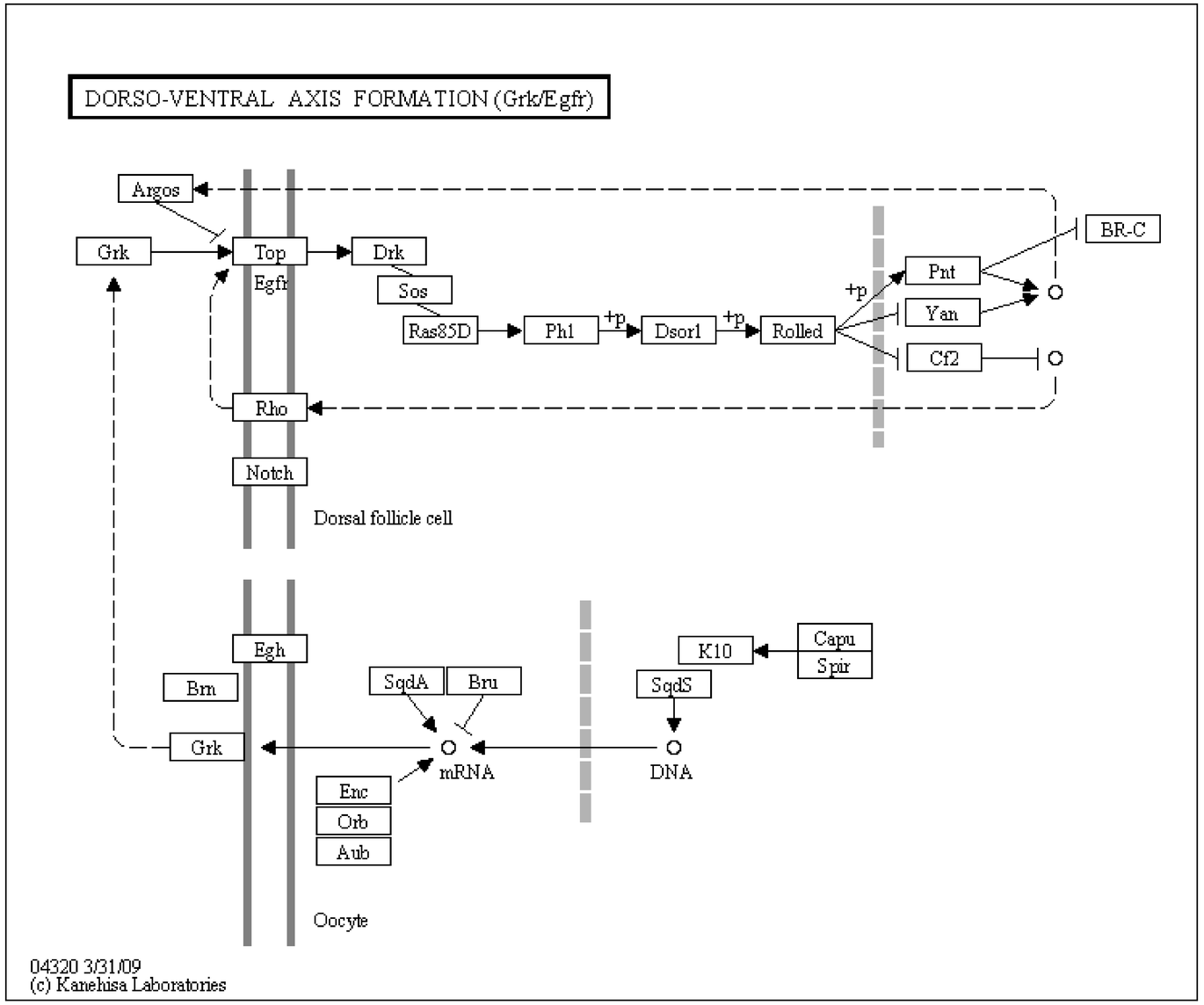
[Klukas2006]

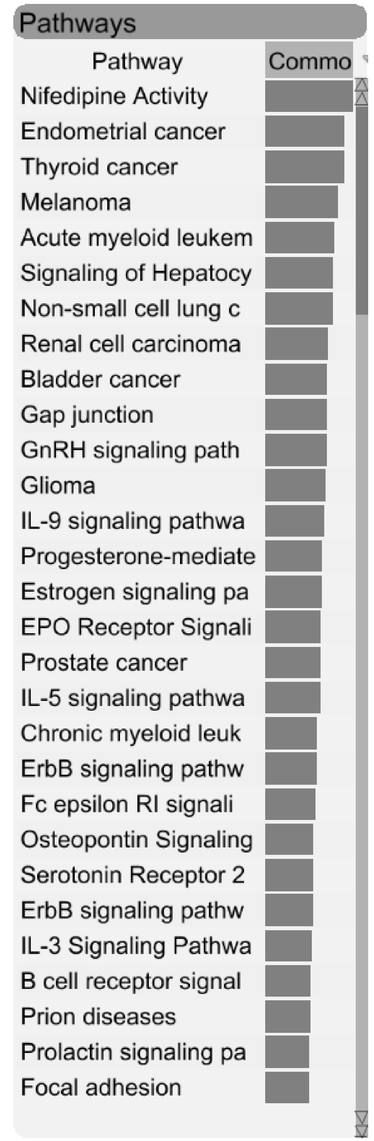
Finding Related Pathways



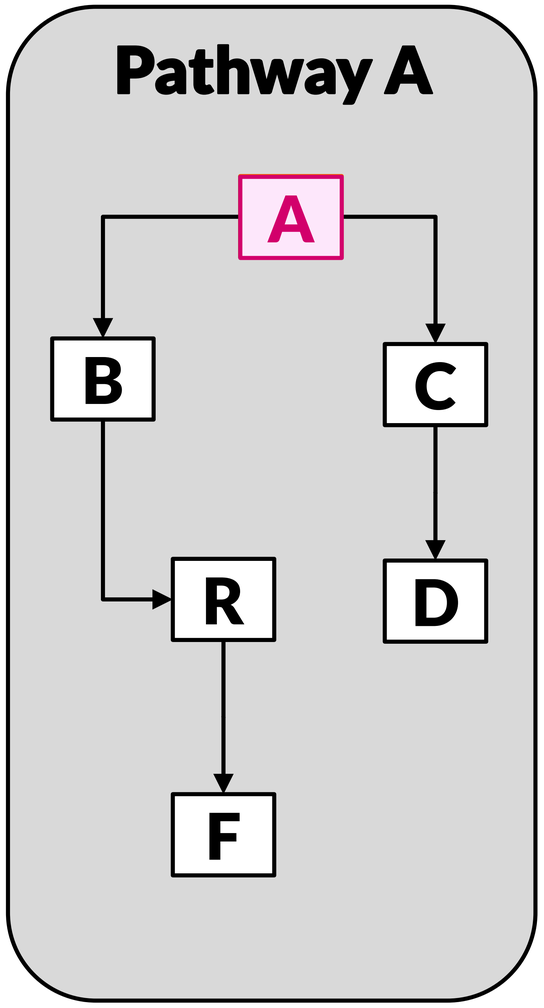


Finding Related Pathways

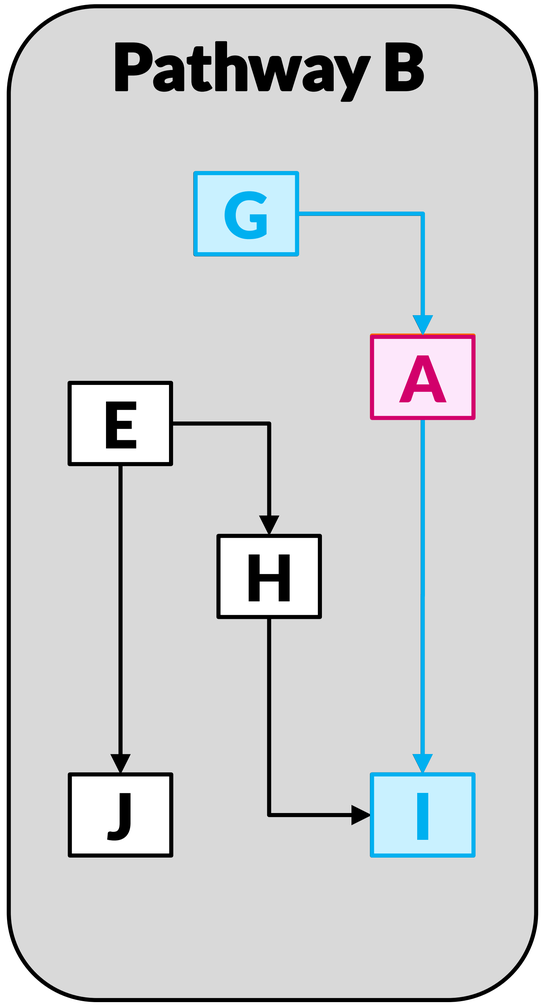




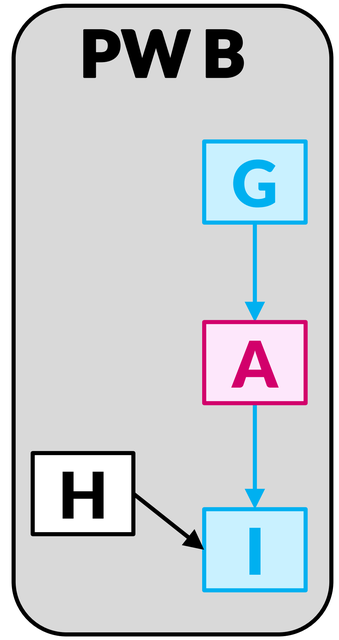
Contextual Subsets



Focus Pathway

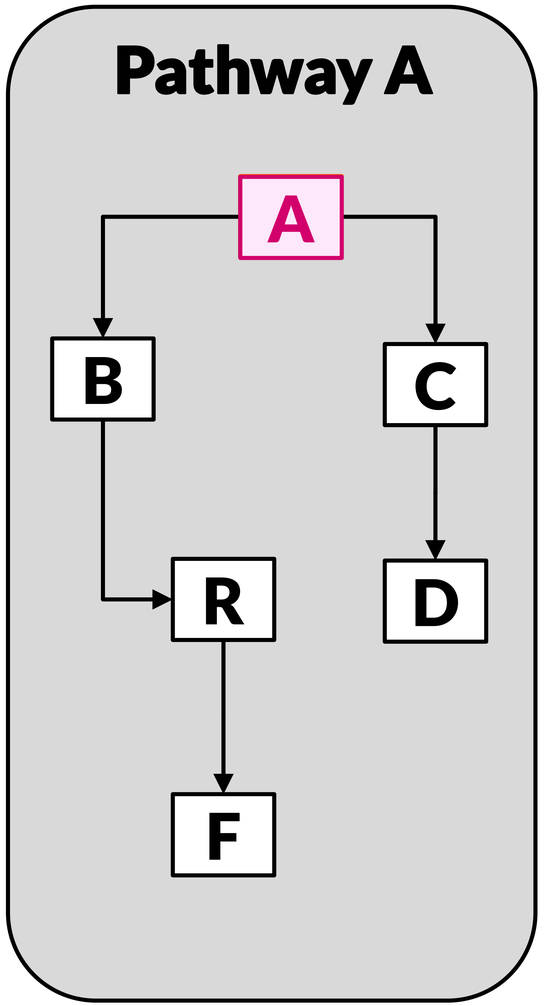


Context Pathway

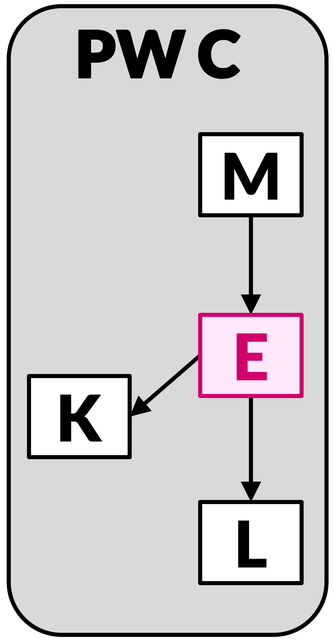


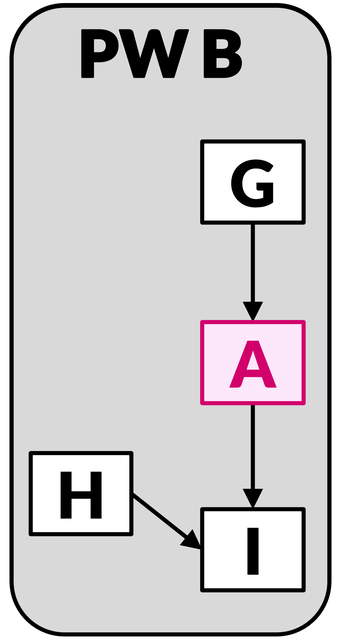
Contextual Subset

Contextual Subsets



Focus Pathway

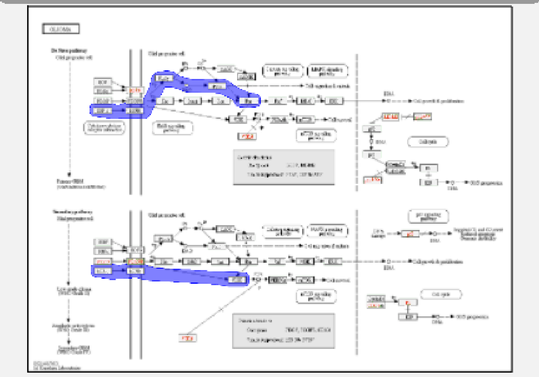


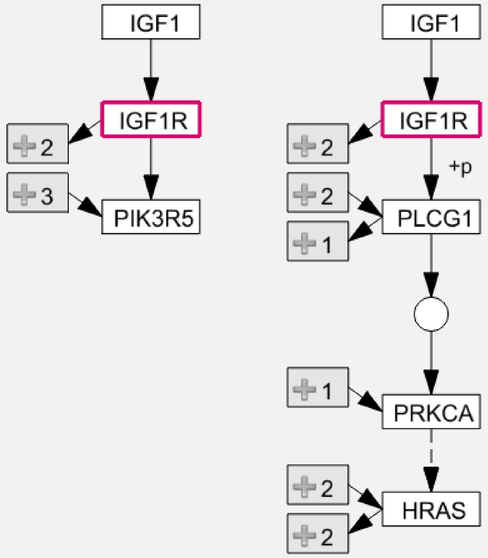


Context Pathways

Levels of Detail

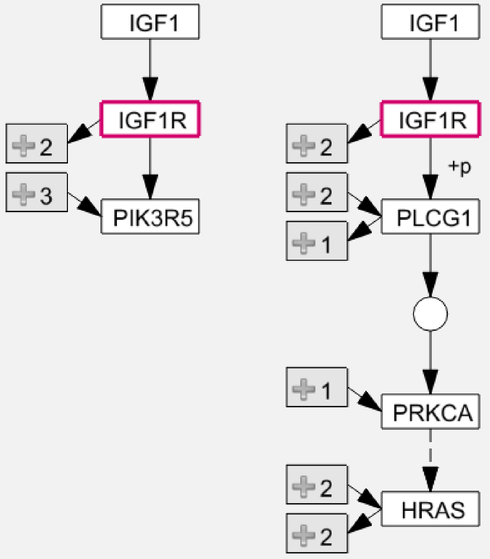
Glioma





High

Glioma

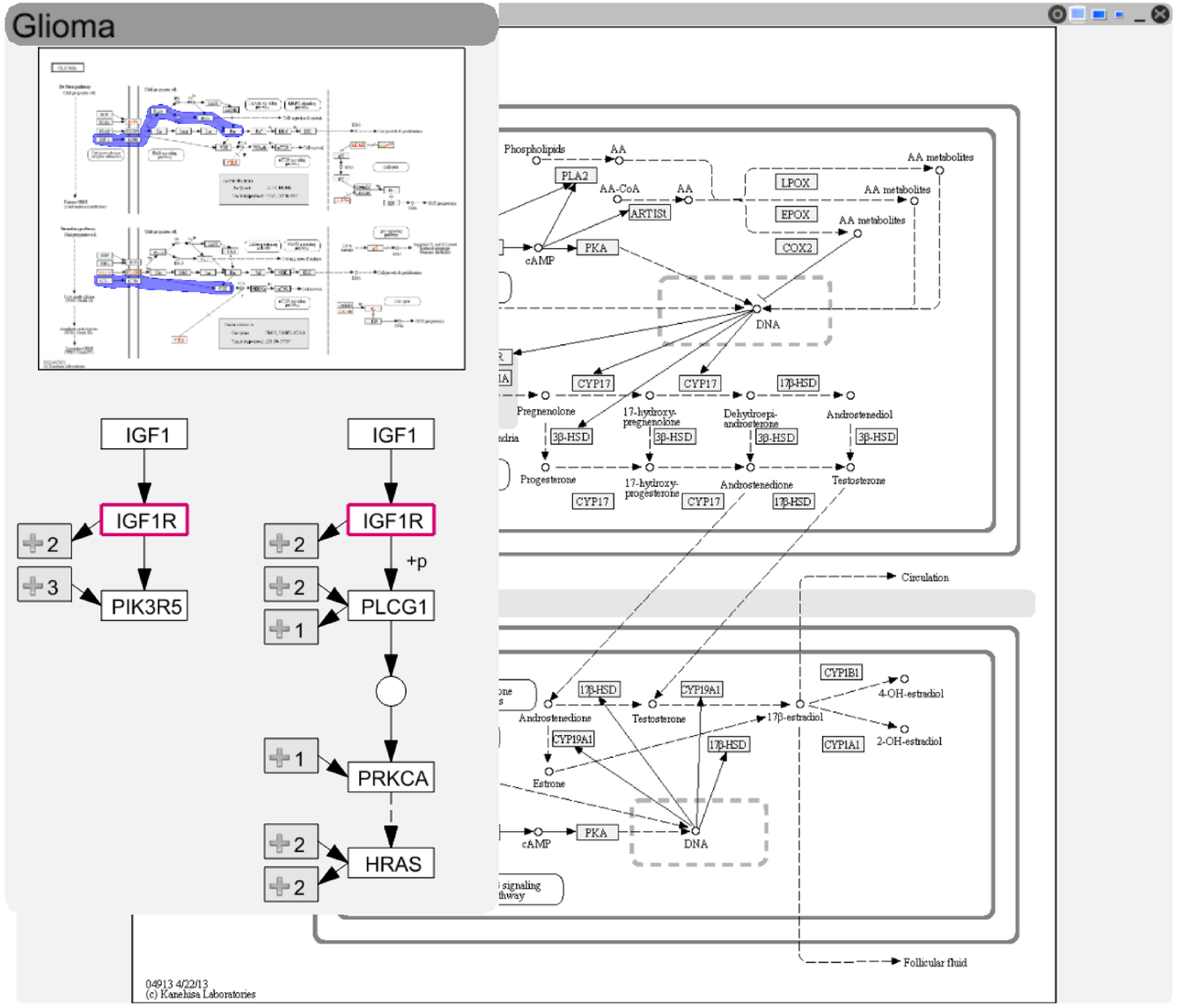


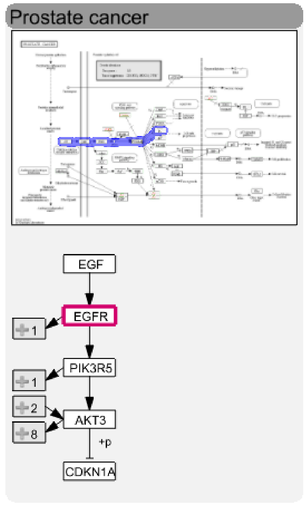
Medium

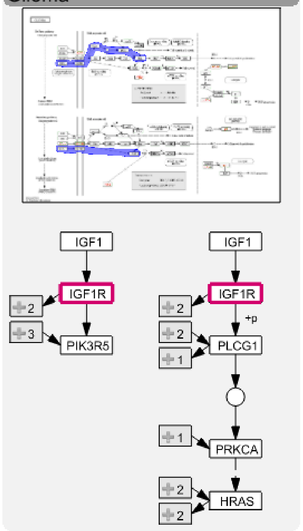
Glioma

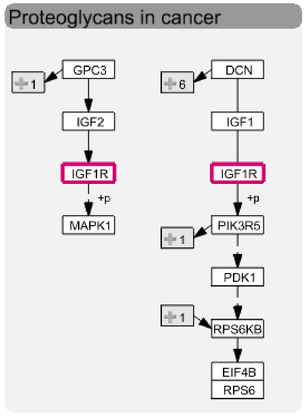
Low

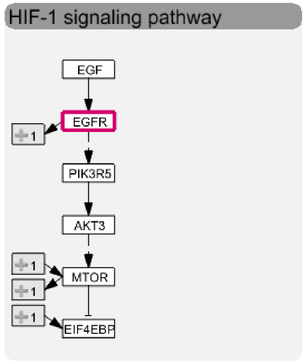
Layout


Glioma
IGF1
IGF1R
2
3
PIK3R5
IGF1
IGF1R
2
2
1
+p
PLCG1
1
PRKCA
2
2
HRAS
Phospholipids
PLA2
AA
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
Follicular fluid
04913 4/22/13
(c) Kanehisa Laboratories

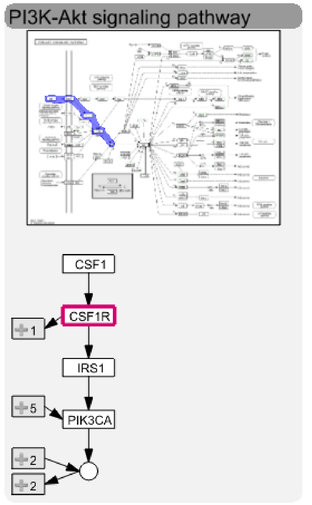

Prostate cancer
EGF
1
EGFR
1
PIK3R5
2
8
AKT3
+p
CDKN1A

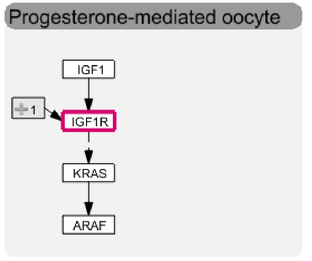

Glioma
IGF1
IGF1R
2
3
PIK3R5
IGF1
IGF1R
2
2
1
+p
PLCG1
1
PRKCA
2
2
HRAS

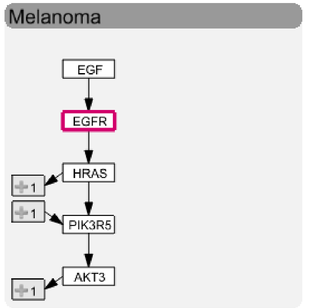

Proteoglycans in cancer
1
GPC3
IGF2
IGF1R
+p
MAPK1
6
DCN
IGF1
IGF1R
+p
1
PIK3R5
PDK1
1
RPS6KB
EIF4B
RPS6


HIF-1 signaling pathway
EGF
1
EGFR
PIK3R5
AKT3
1
1
1
MTOR
EIF4EBP

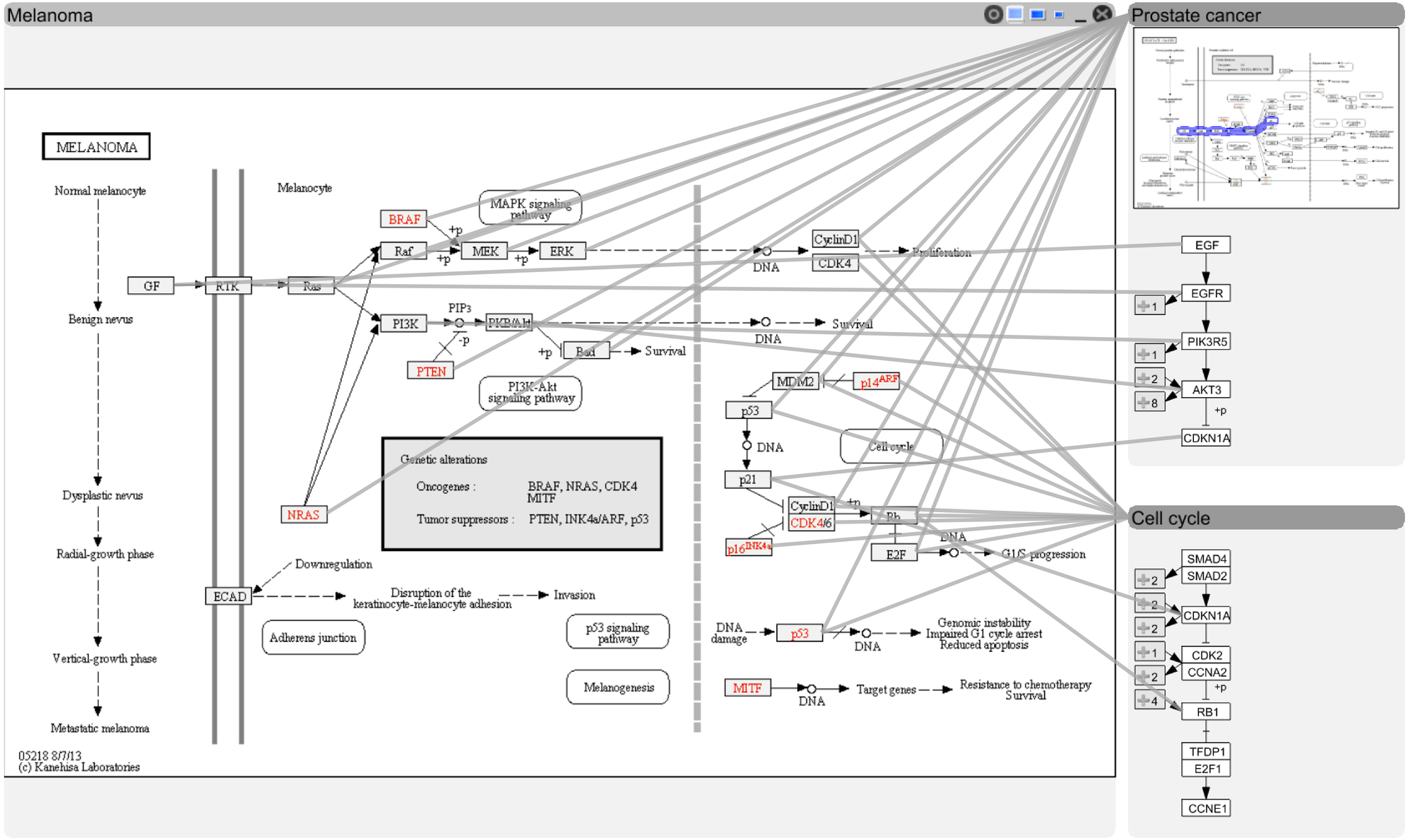

Oocyte meiosis
IGF1
1
IGF1R


PI3K-Akt signaling pathway
CSF1
1
CSF1R
IRS1
5
PIK3CA
2
2

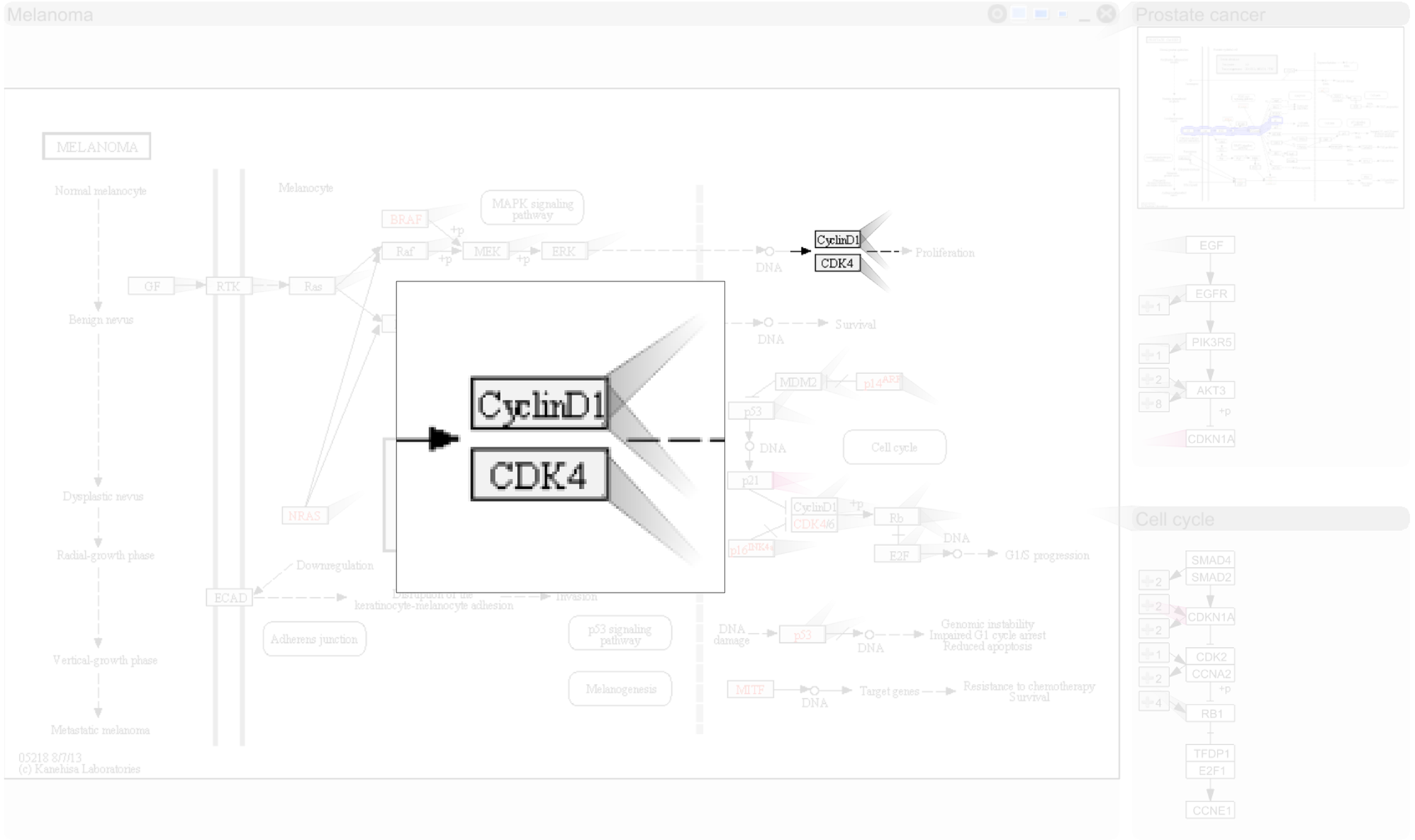

Progesterone-mediated oocyte
IGF1
1
IGF1R
KRAS
ARAF


Melanoma
EGF
EGFR
1
HRAS
1
PIK3R5
1
AKT3

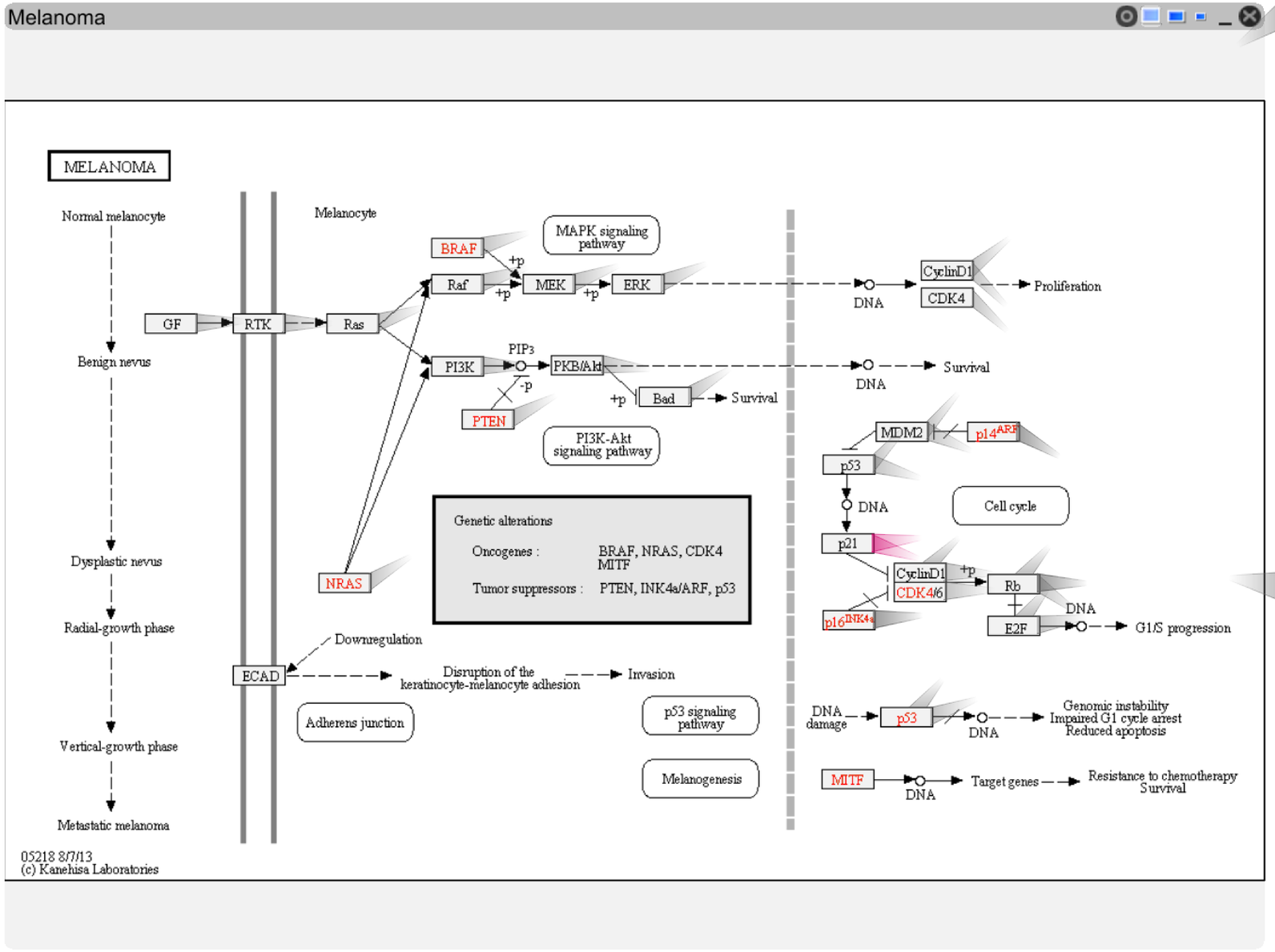
Visualizing Relationships

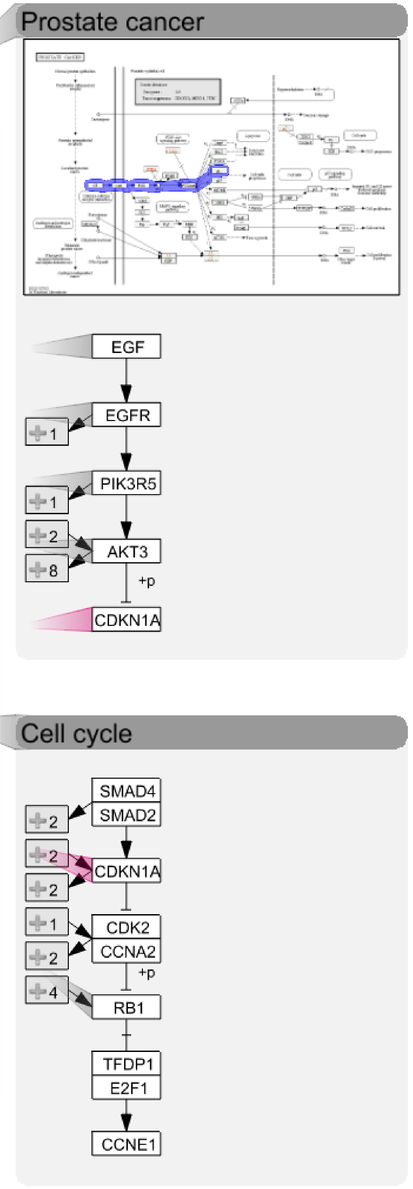


Visualizing Relationships

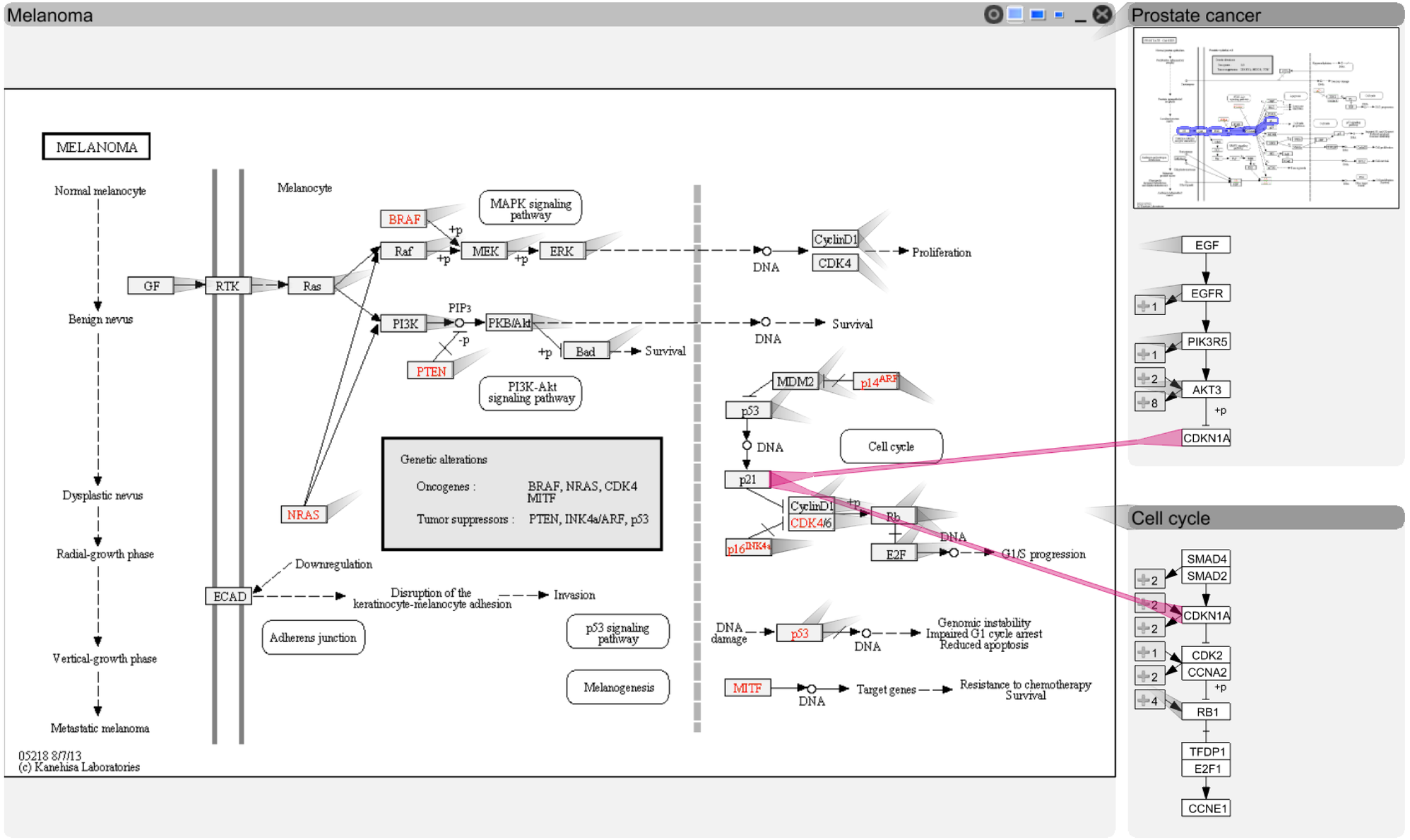


Visualizing Relationships

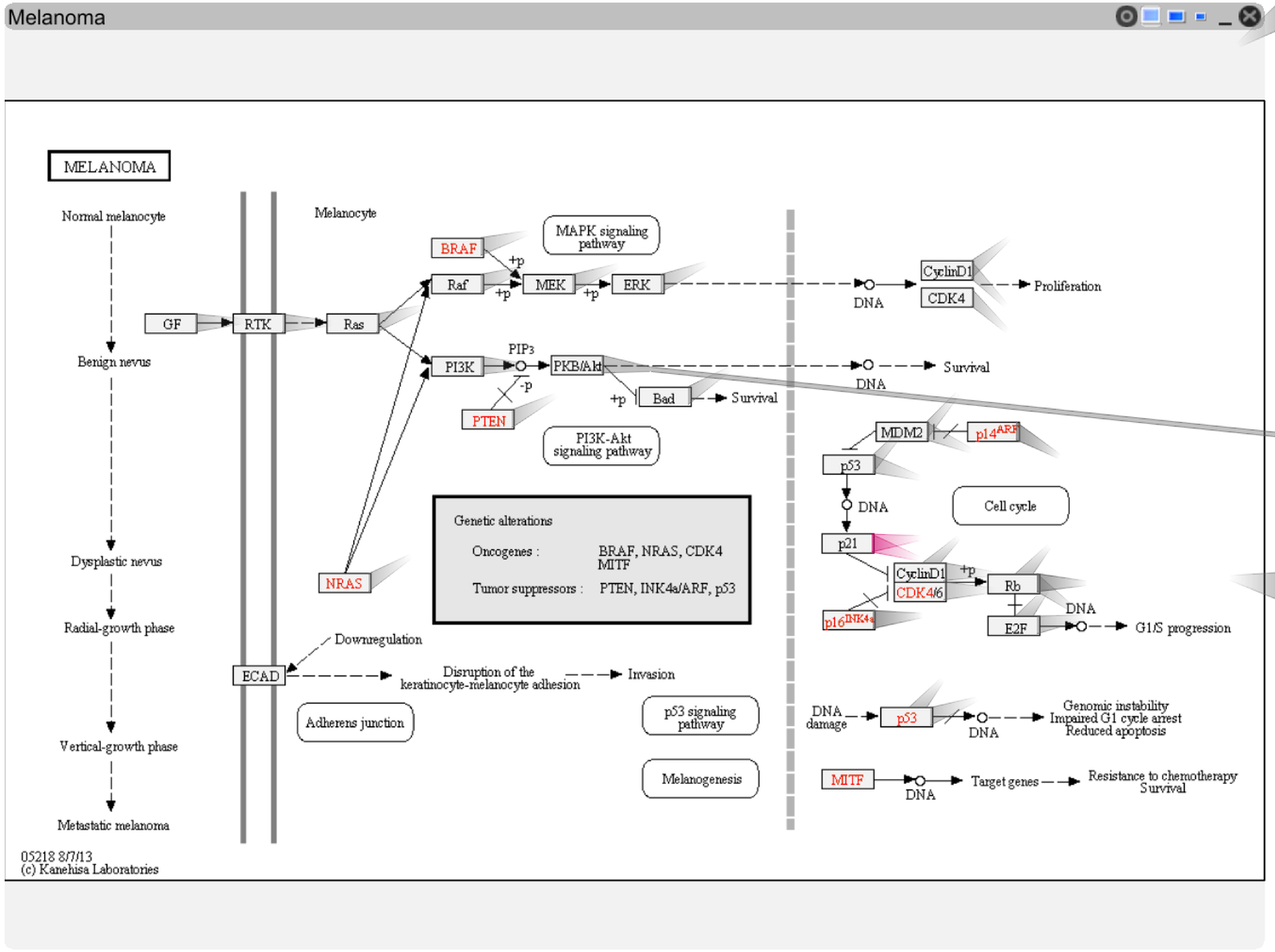


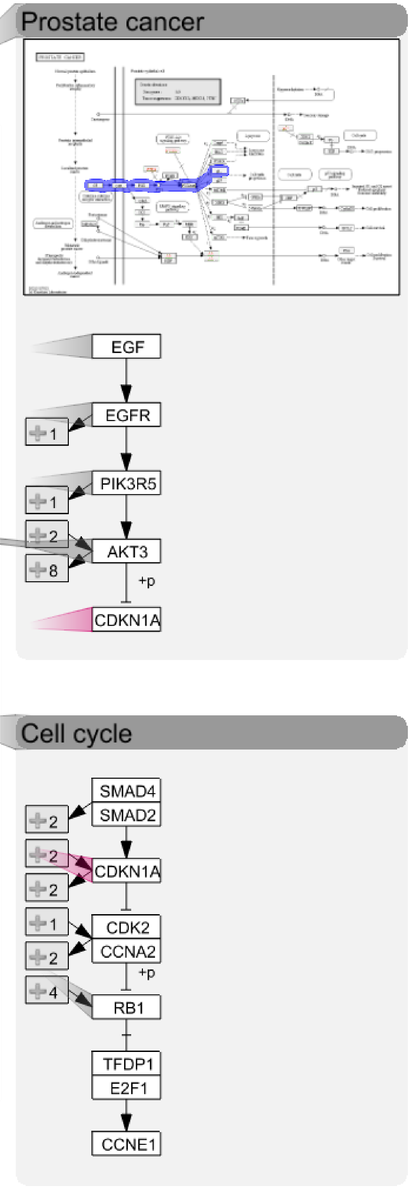


Visualizing Relationships

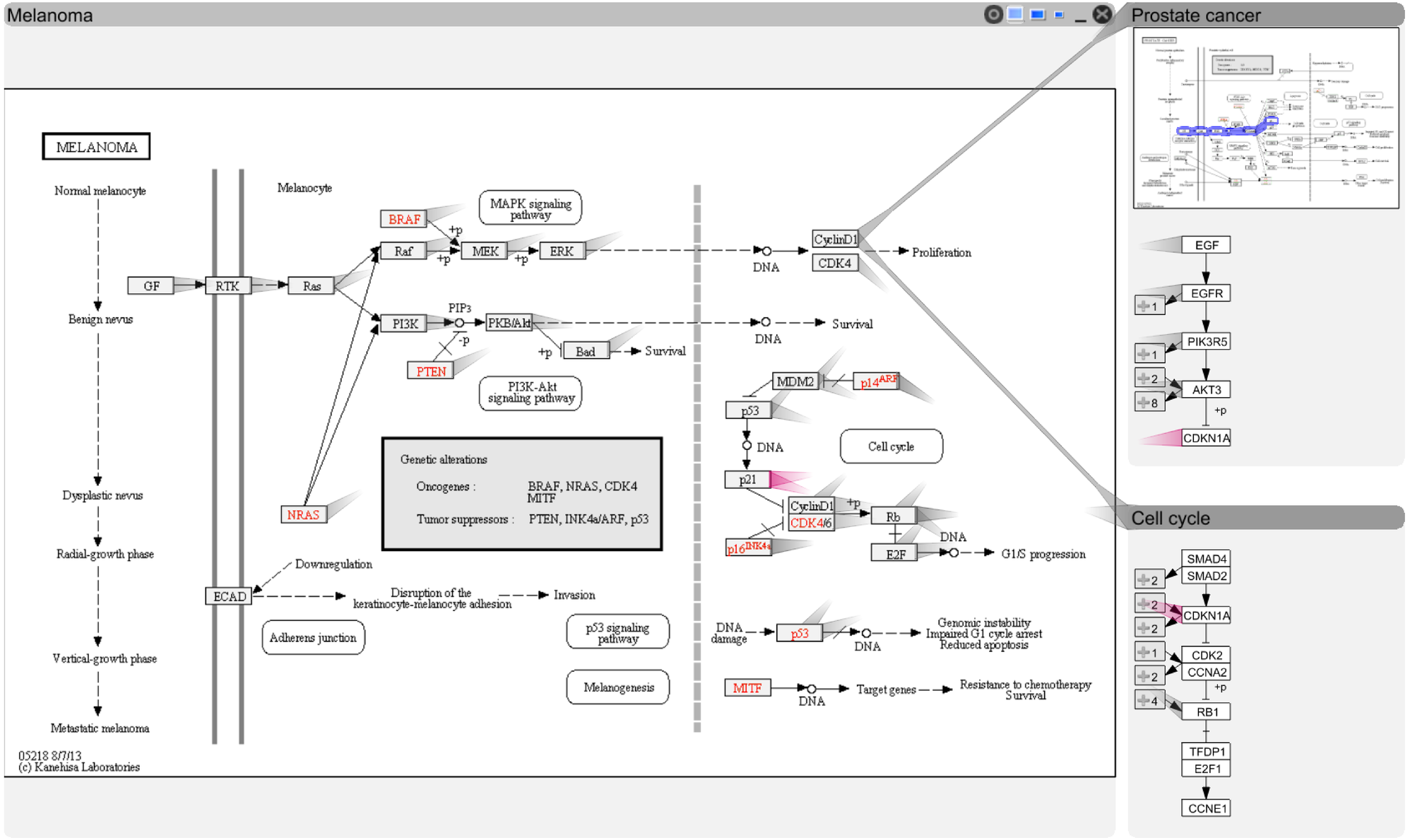


Visualizing Relationships





Visualizing Relationships



How to visualize experimental data on pathways?

Experi-mental Data and Pathways

enRoute

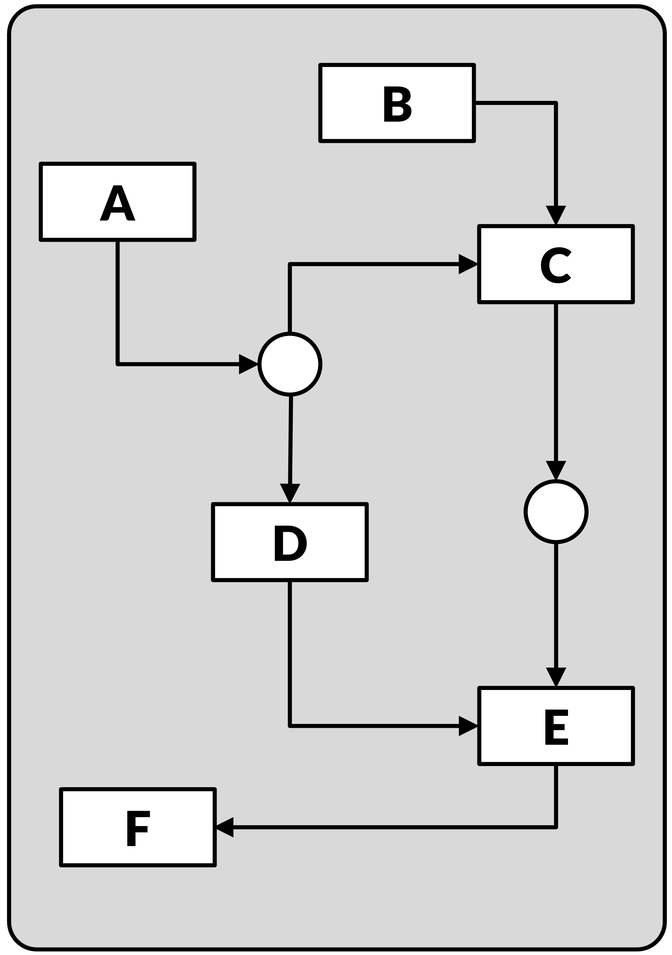
[Partl, BioVis ‘12]

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
B	2.8	1.8
C	3.1	-2.2
D	-3	-2.8
E	0.5	0.3
F	0.3	0.3

Challenge: Data Scale & Heterogeneity

Large number of experiments

- Large datasets have more than 500 experiments

Multiple groups/conditions

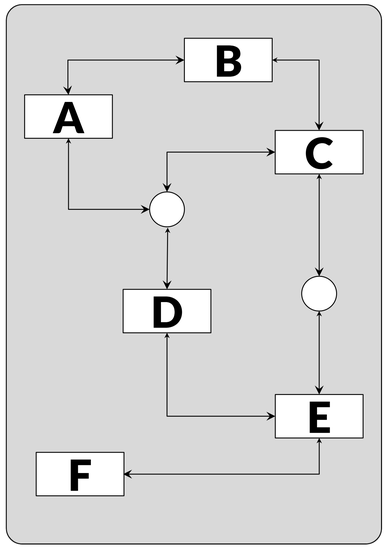
Different types of data, require different visualization techniques

Challenge: Supporting Multiple Tasks

Two central tasks:

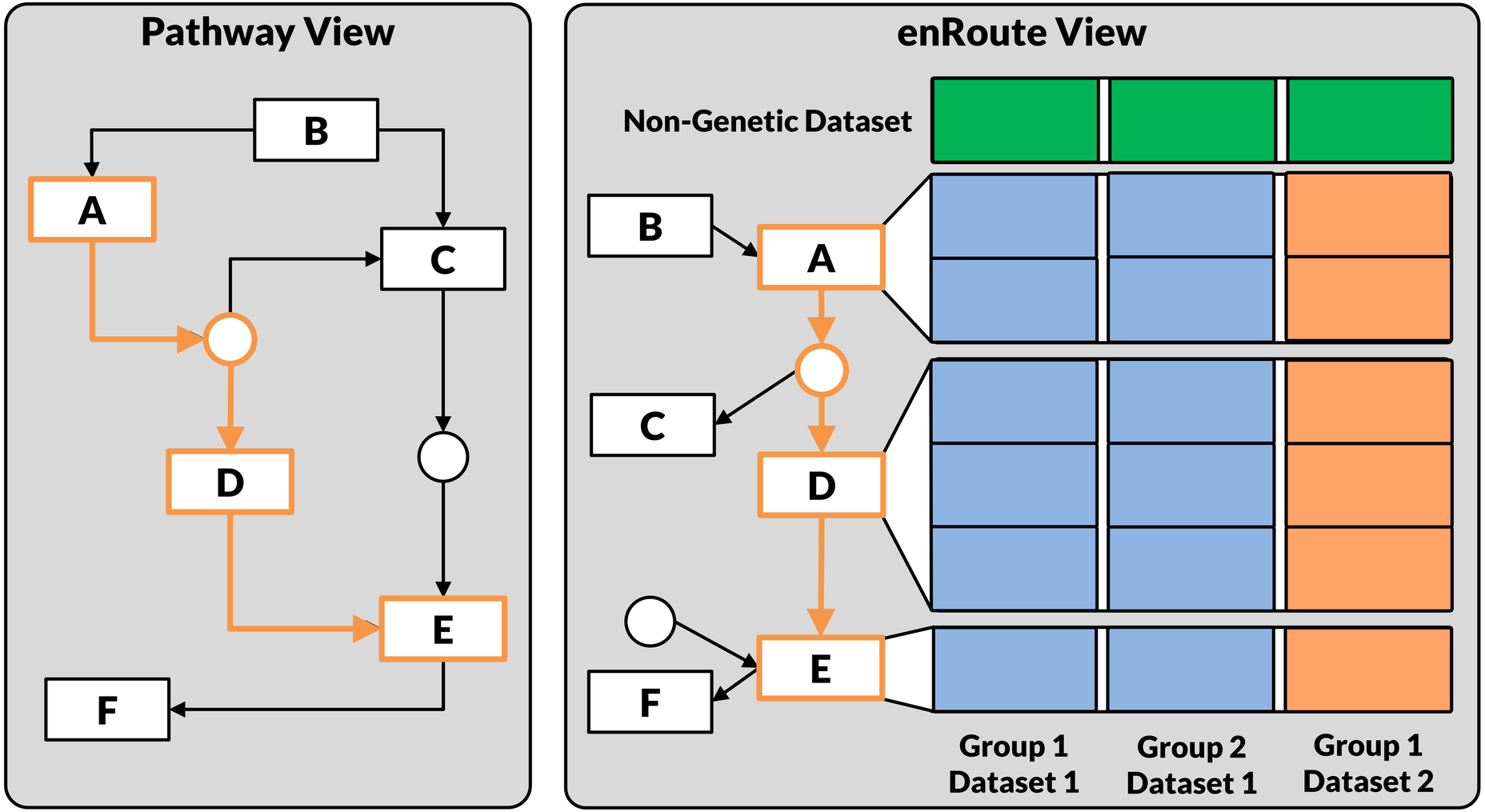
- Explore topology of pathway
- Explore the attributes of the nodes (experimental data)

Need to support both!

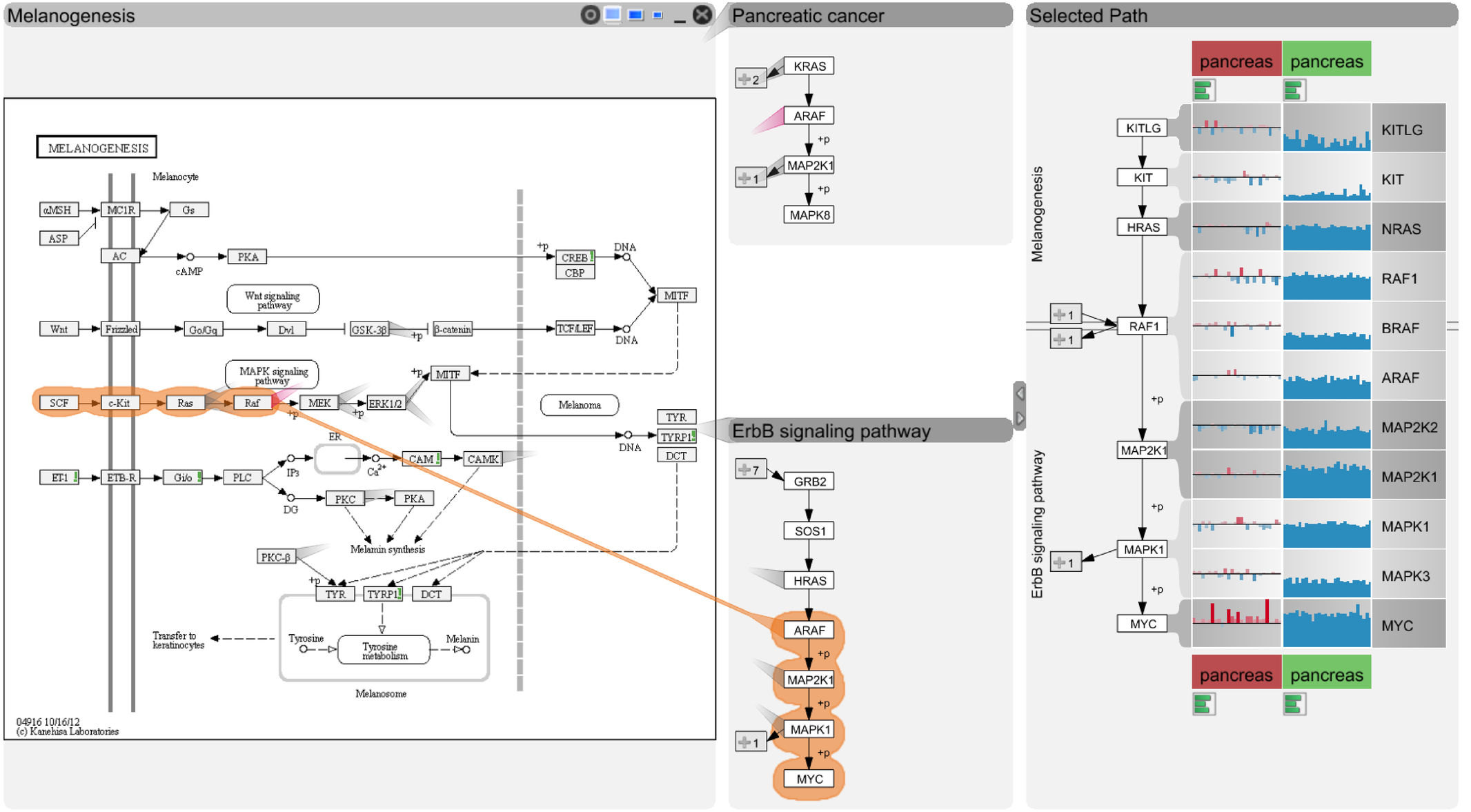


	Sample 1	Sample 2	Sample 3
Gene 1	1	1.1	0.4
Gene 2	2	0.5	1.2
Gene 3	1.4	0.2	0.5
Gene 4	0.3	0.5	0.7

Concept


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

enRoute


Melanogenesis
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

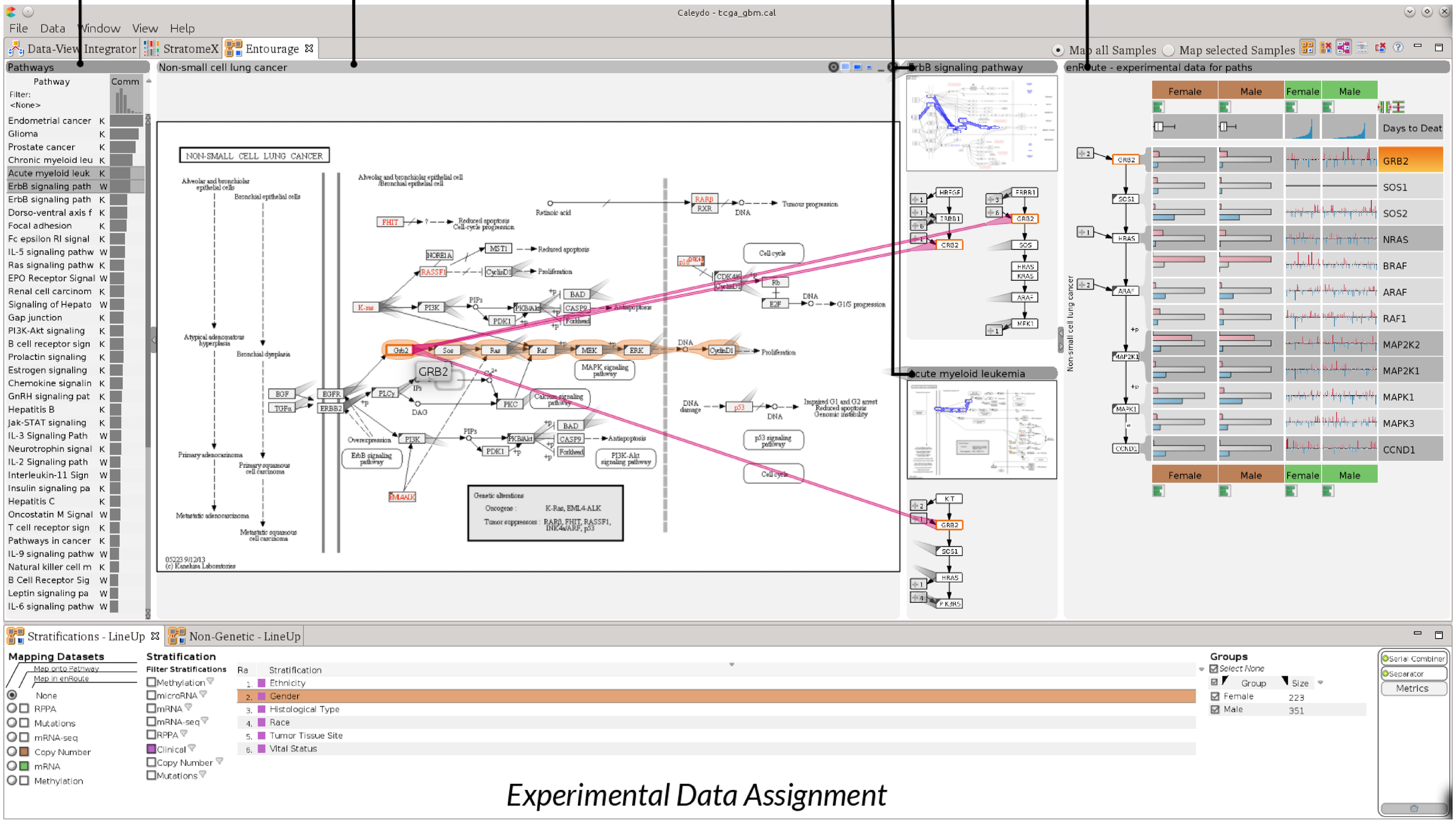
Pathways

Pathway

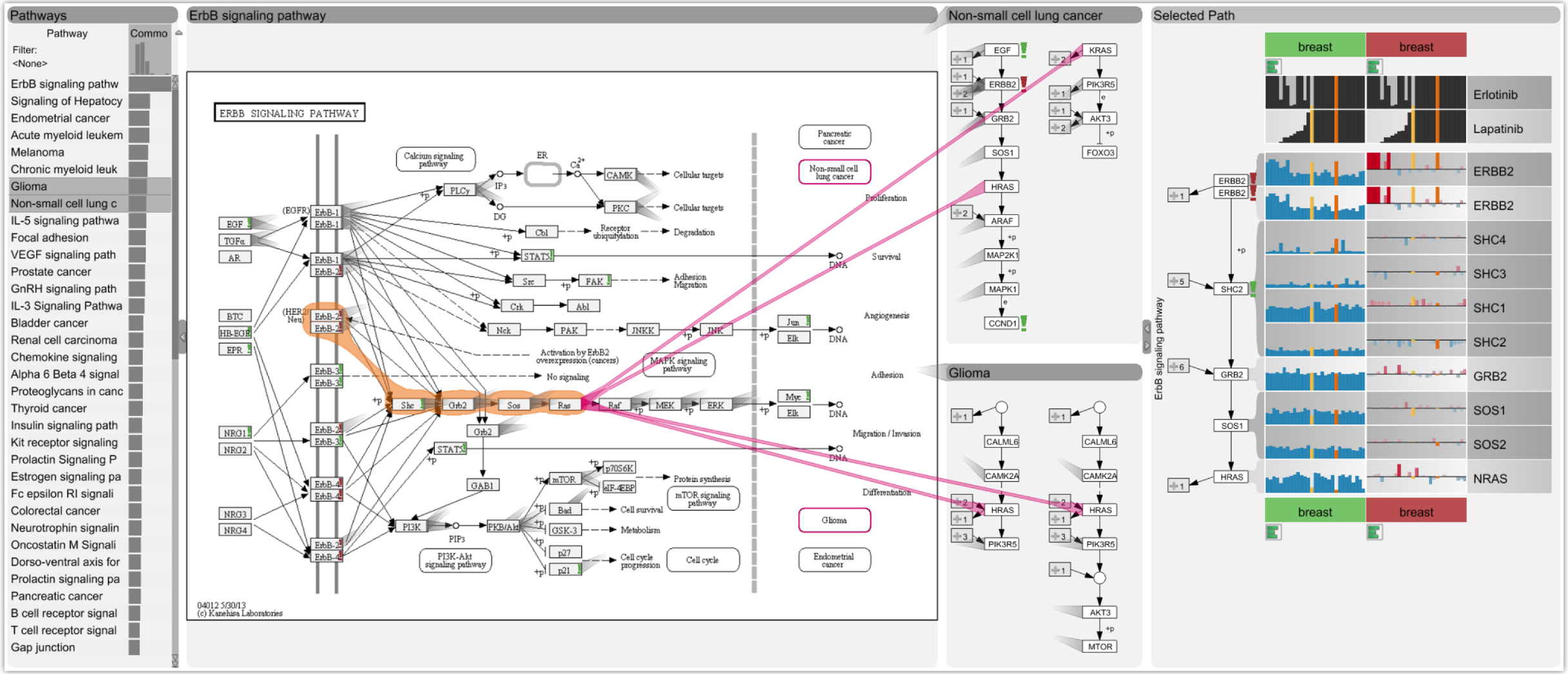
Filter:
<None>

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

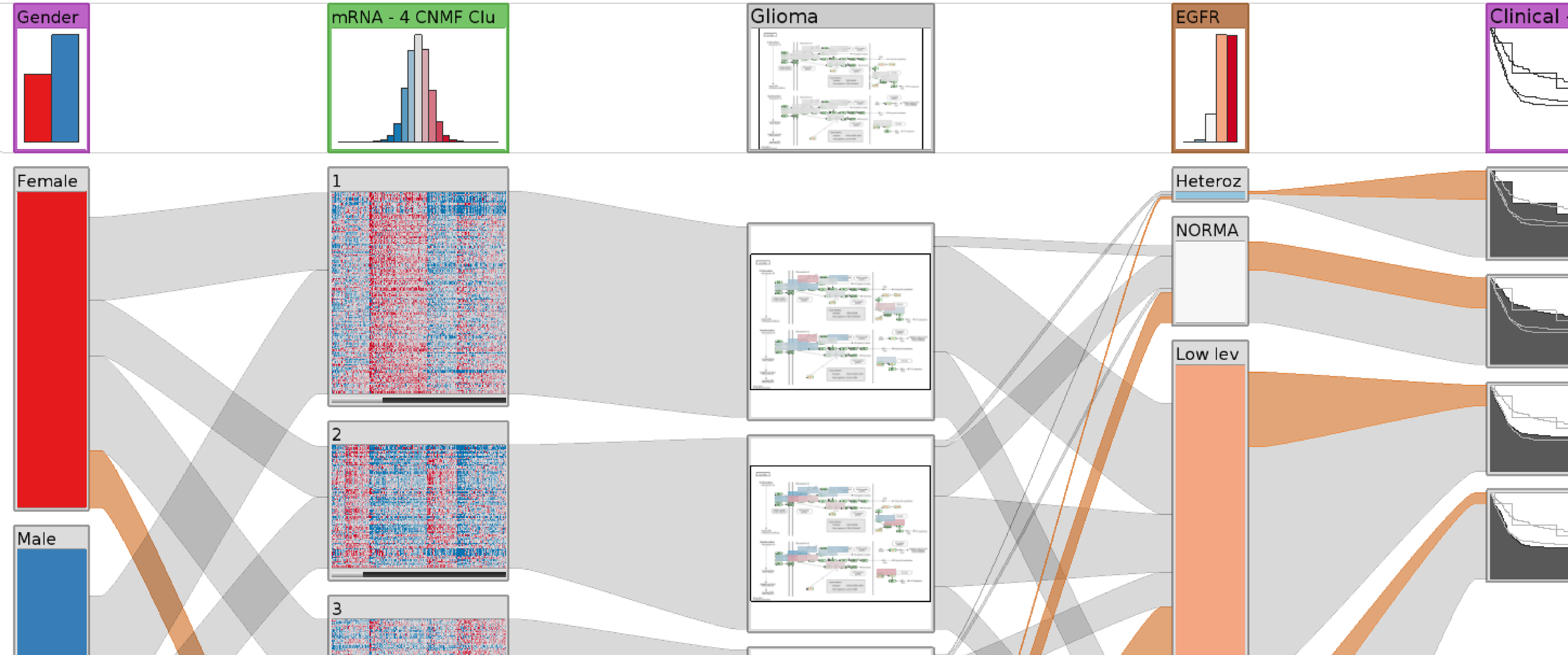


More Information:

http://entourage.caleydo.org

[Lex, EuroVis '12]
3rd Best Paper Award
[Streit, Nature Methods '14]
To appear

Heterogeneous Data Cancer Subtypes



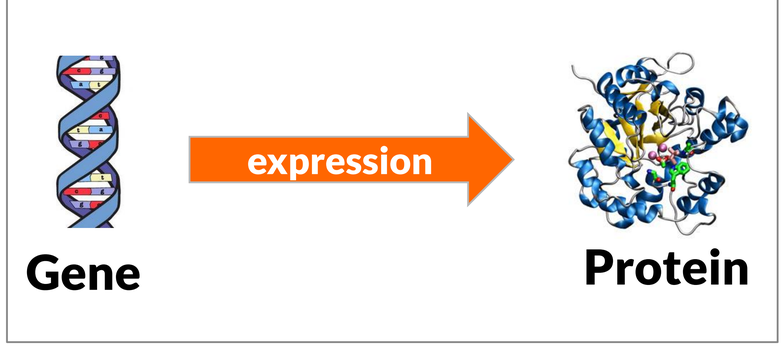
Cancer Subtypes

Cancer is not homogeneous

- different histology
- different molecular alterations

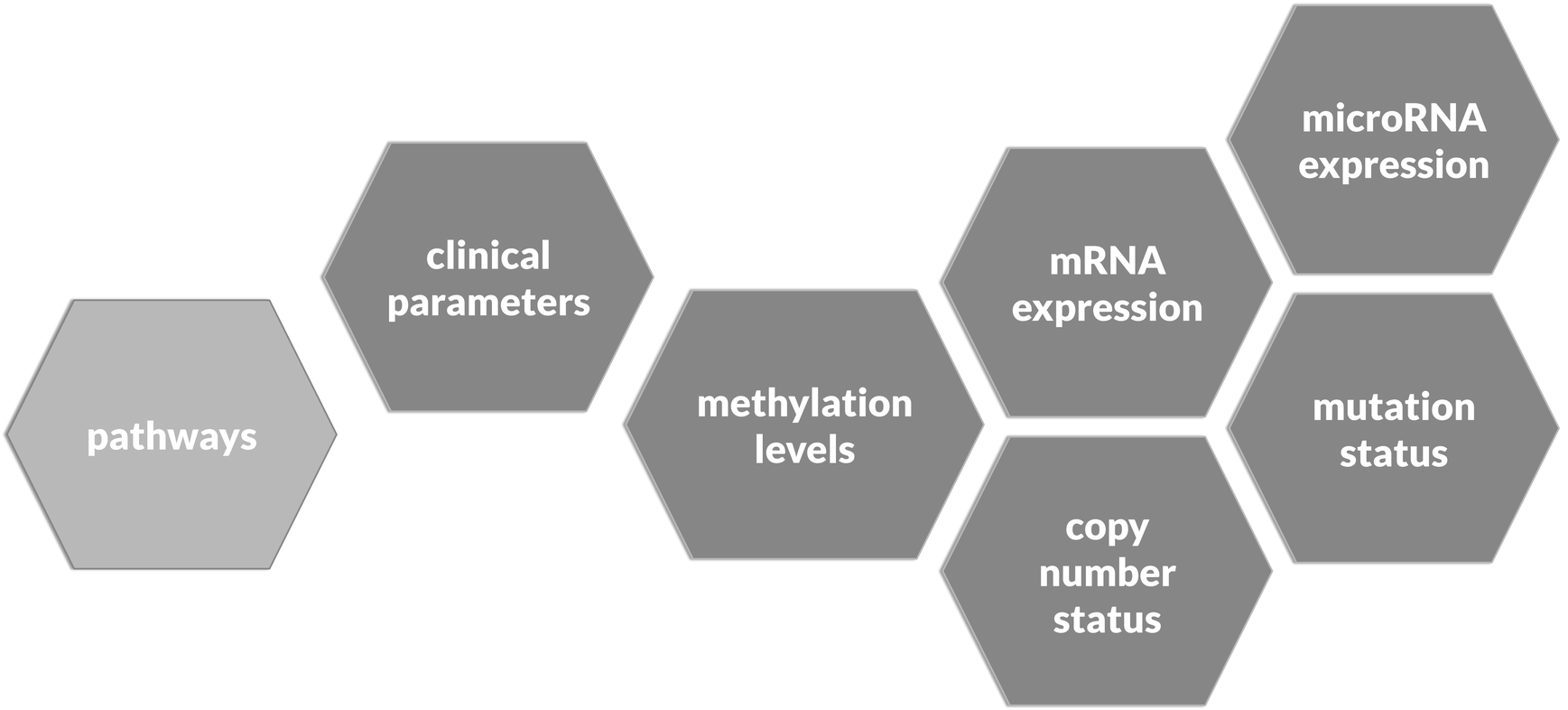
Subtypes have serious implications

- different treatment for subtypes
- prognosis varies between subtypes



The Cancer Genome Atlas

Understanding genomics to improve cancer care



StratomeX

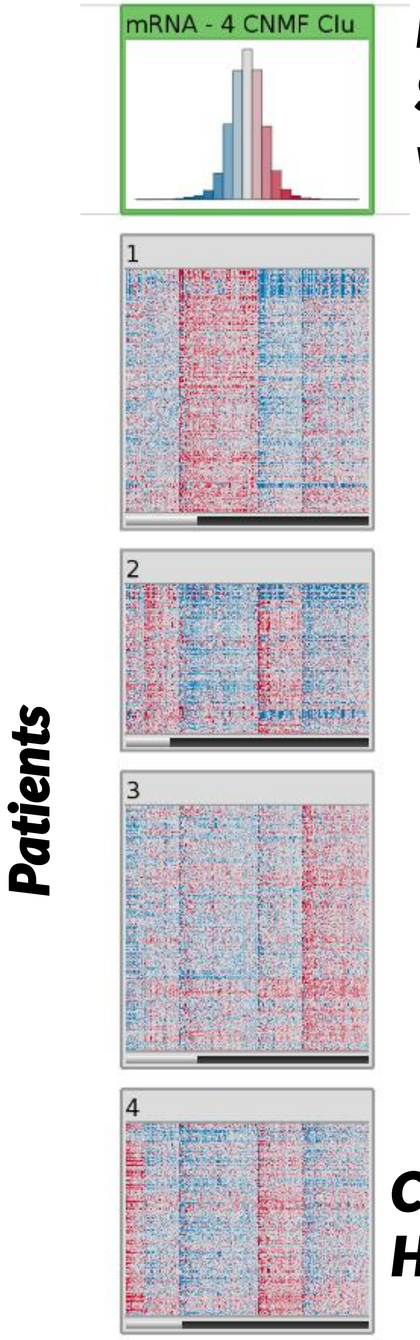
visualizes…

- … the relationships between multiple heterogeneous datasets
- … the data within the datasets
- … alternative clusterings & groupings
- … the effect of groupings on clinical parameters & biological processes

Stratifying Patients

Subtypes are identified by stratifying datasets, e.g.,

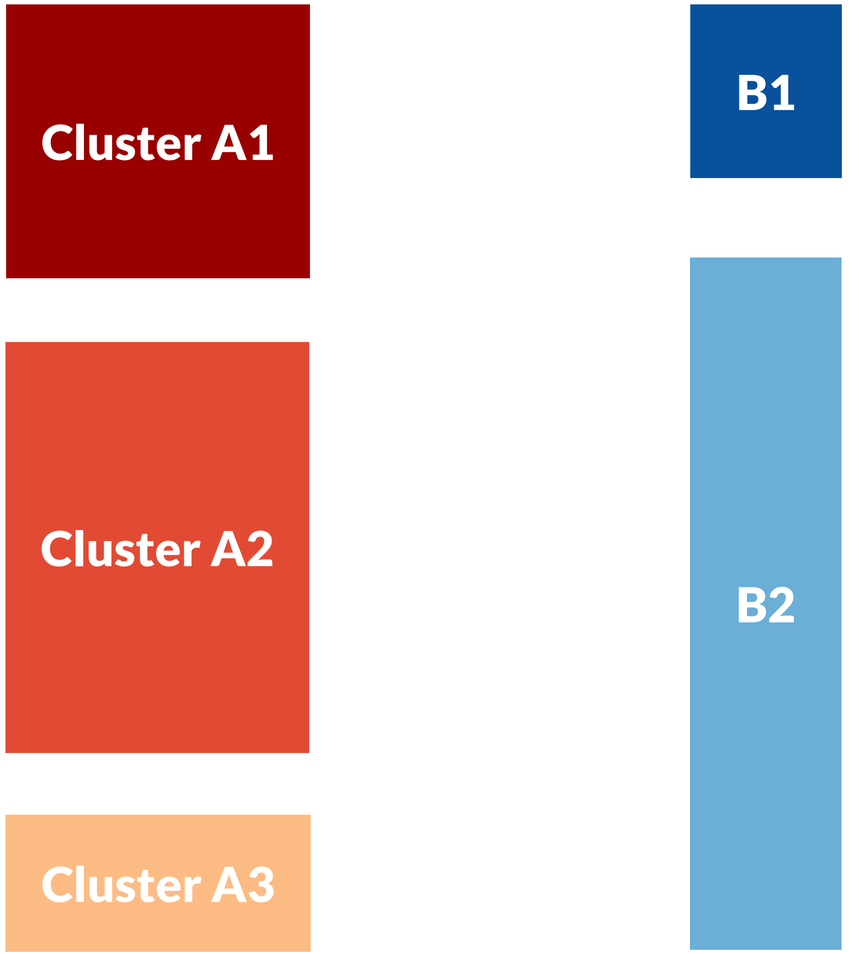
- based on an expression pattern
- a mutation status
- a copy number alteration
- a combination of these



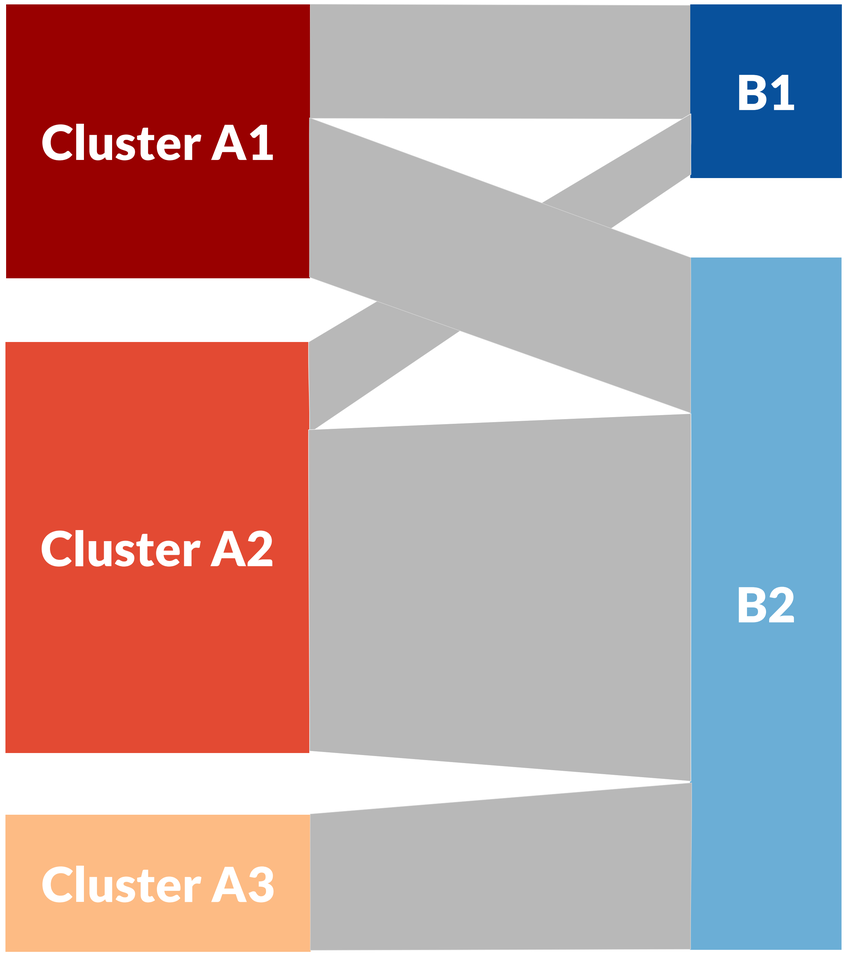
Header / Summary of whole Stratification

Candidate Subtype / Heat Map

Comparing Stratifications



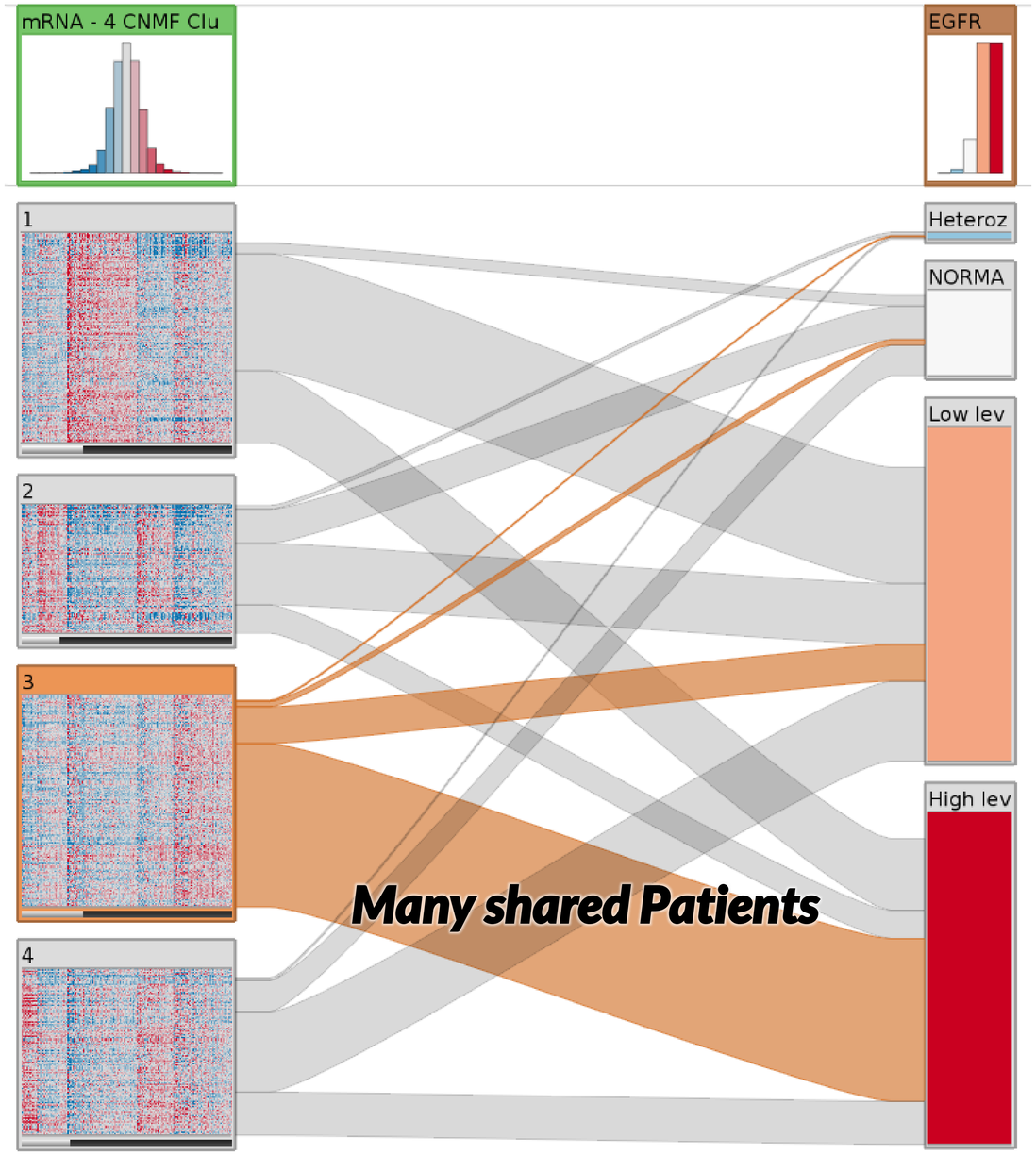
Comparing Stratifications



Tabular
e.g., mRNA

Categorical,
e.g., mutation status

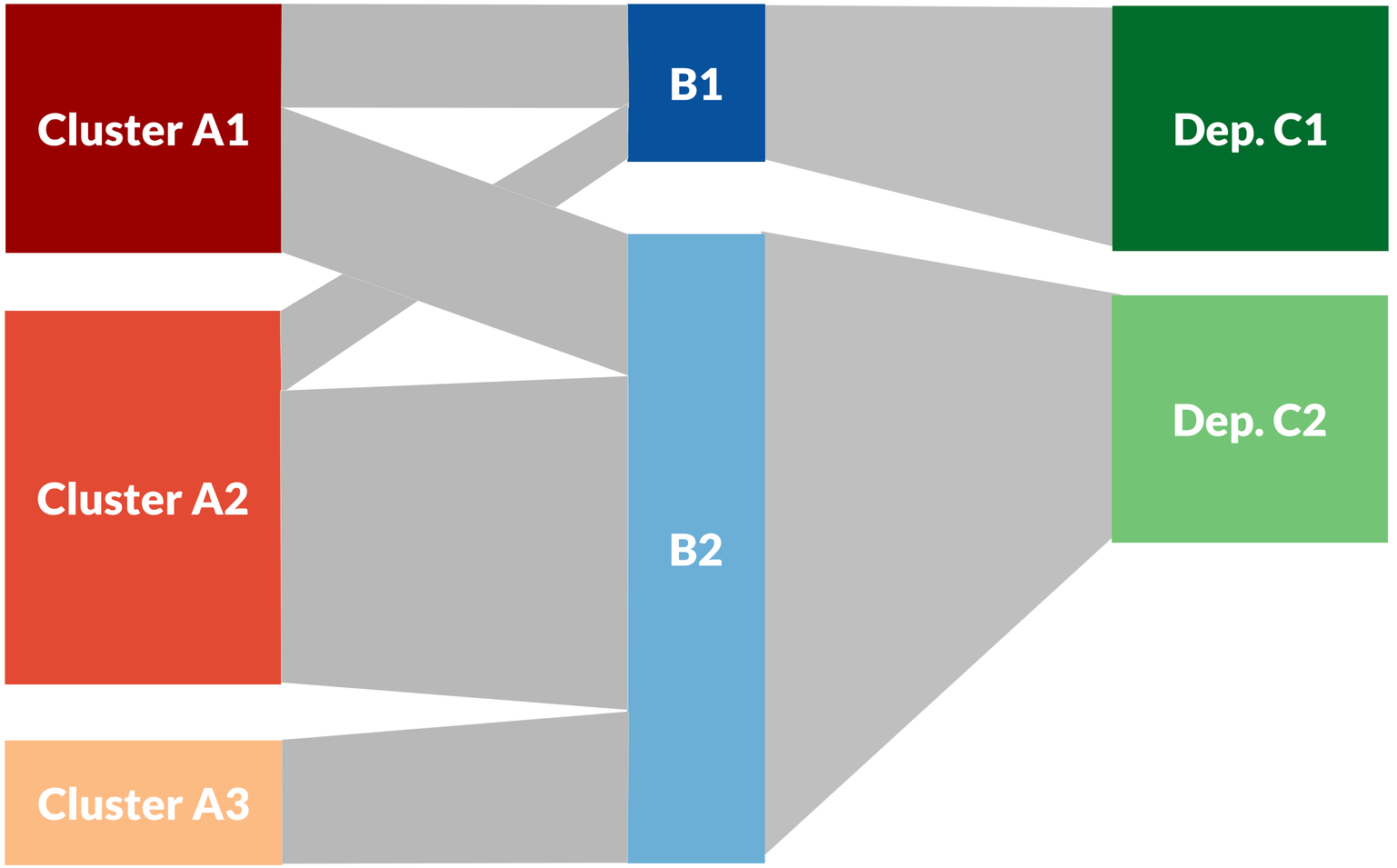
[Parallel Sets, Kosara]

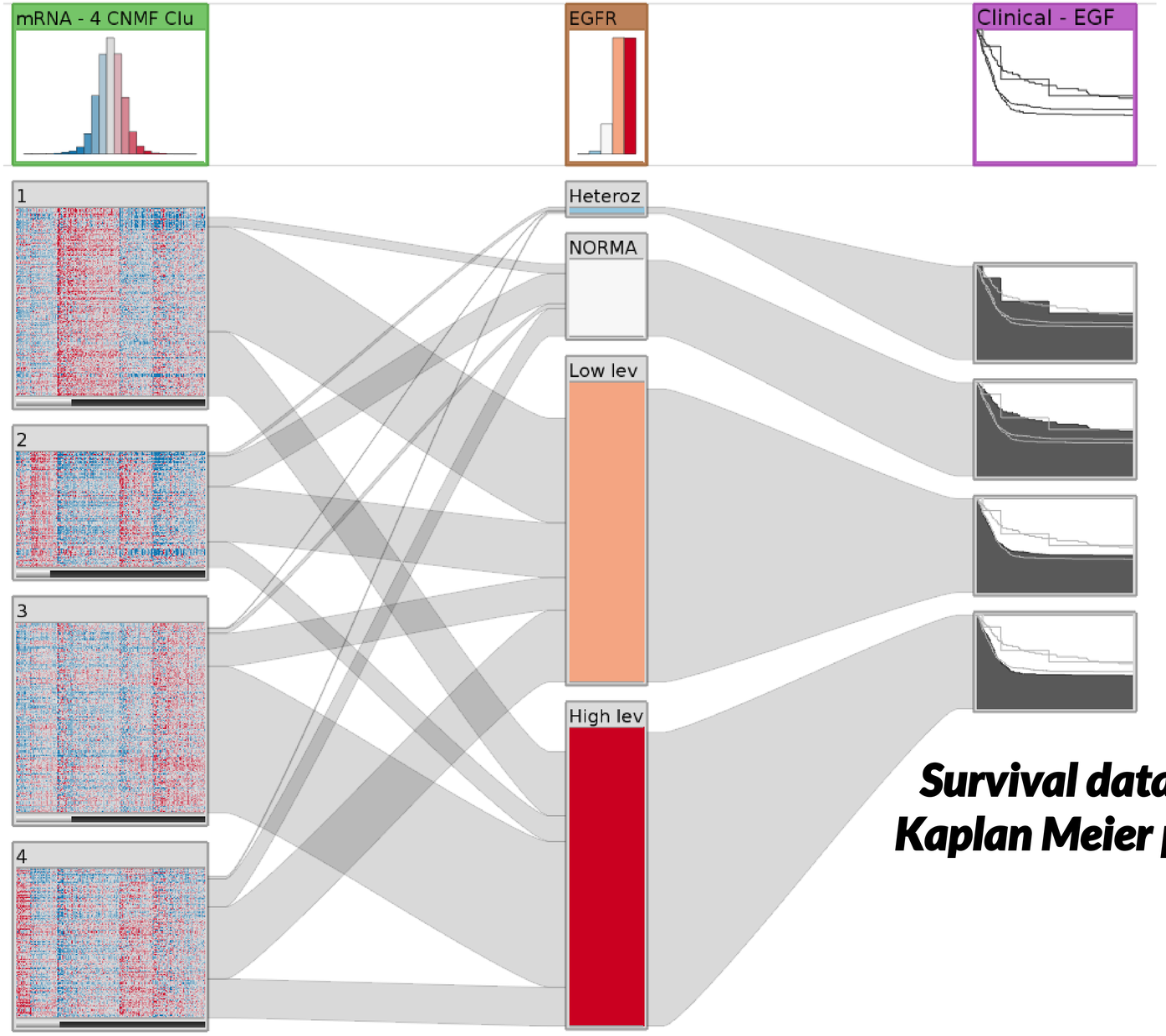


Clustering of mRNA Data

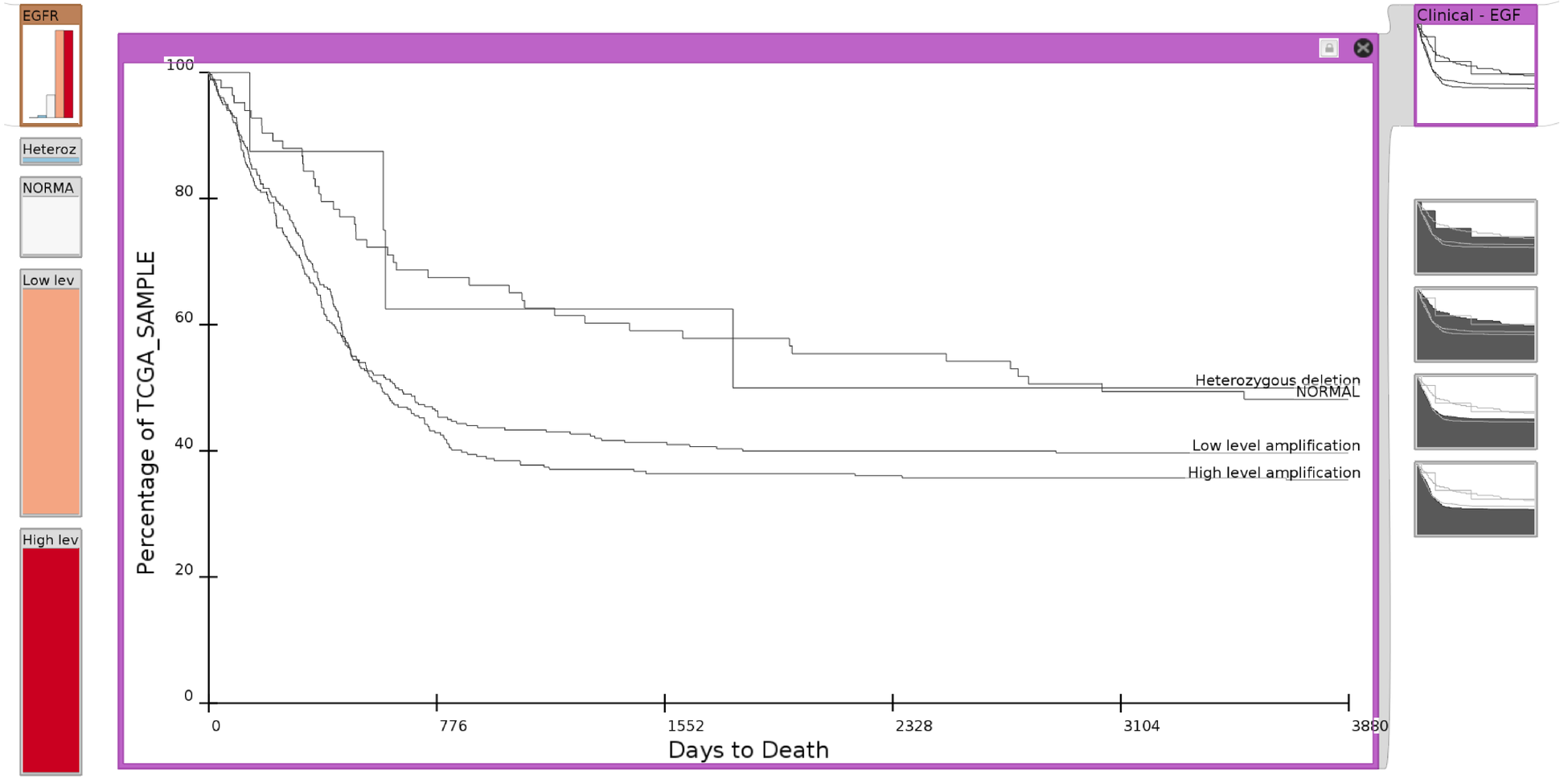
Stratification on Copy Number Status

Other Data – Same Stratification

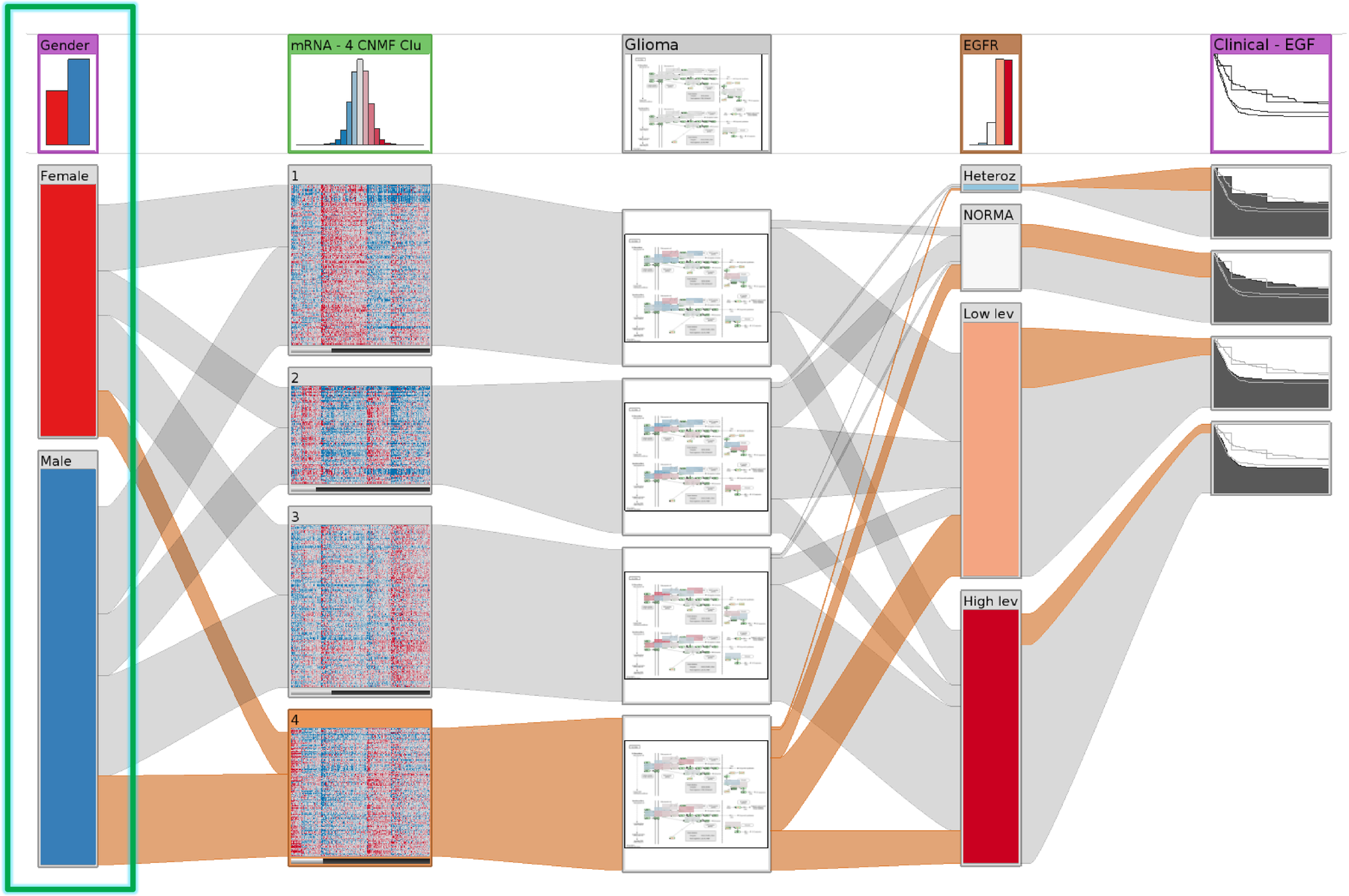




Survival data in Kaplan Meier plots

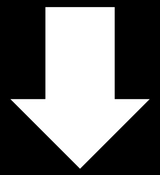

EGFR
Heteroz
NORMA
Low lev
High lev
Clinical - EGF
100
80
60
40
20
0
Percentage of TCGA_SAMPLE
Heterozygous deletion
NORMAL
Low level amplification
High level amplification
0
776
1552
2328
3104
3880
Days to Death

Detail View



Stratification based on clinical variable (gender)

Knowledge Driven Approach



Data Driven Approach

Finding Relevant Stratifications

~ 10 datasets

~ 15 clusterings per matrix

~ 15,000 stratifications for copy number & mutations

~ 500 pathways

~ 20 clinical variables

Calculate scores for matches

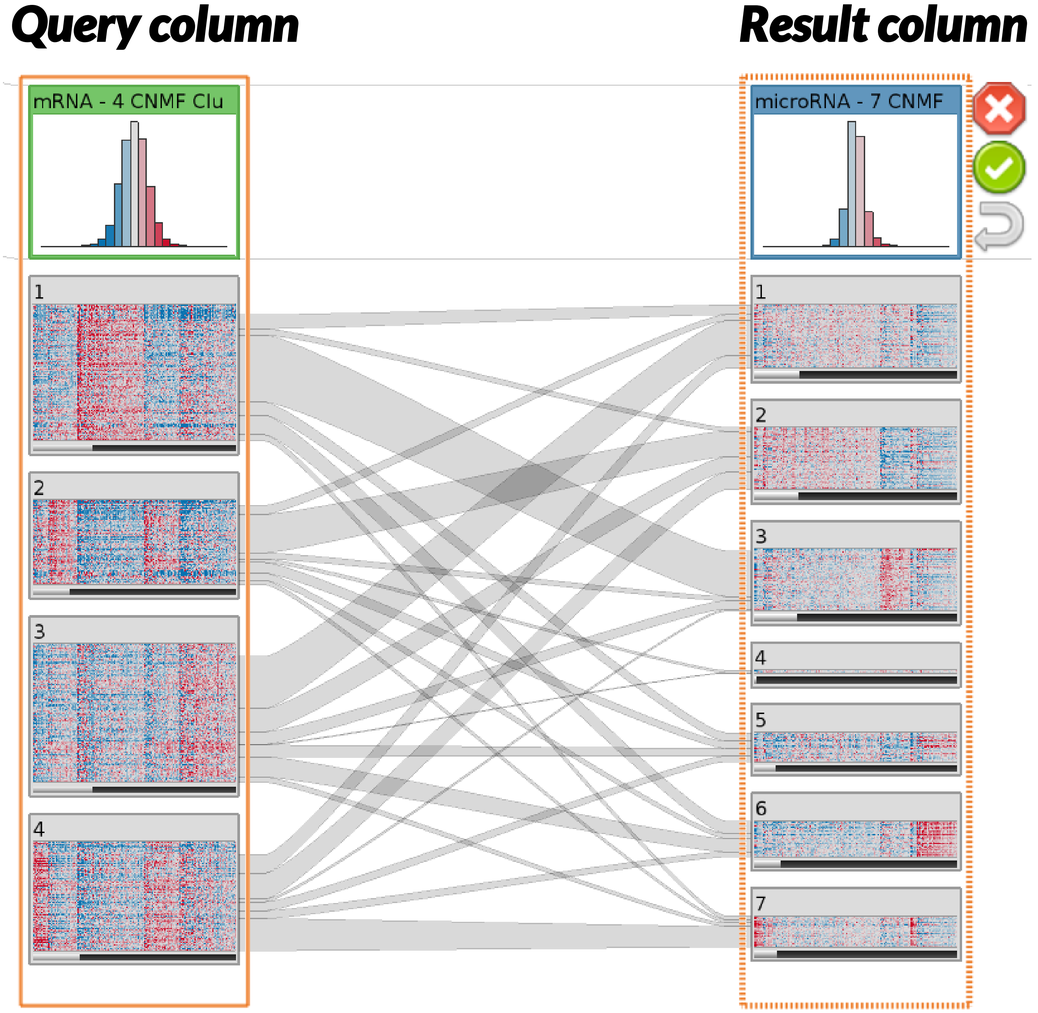
Rank the results

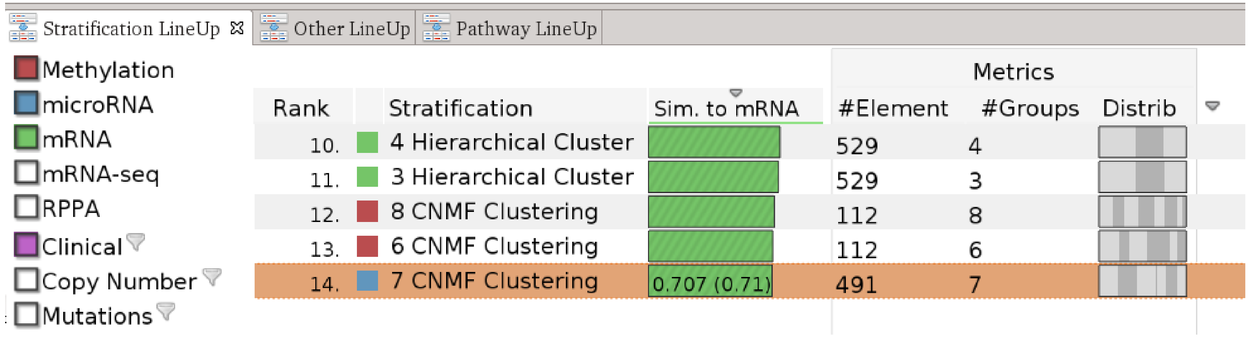
Integrating Computation

Data Visualization

Algorithms

Algorithm Output Visualization





Ranked Stratifications

File Data Window View Help

Tool Bar

Selection Info

Dataset Info

Project: TCGA GBM Package

Dataset: mRNA

Tcga_samples: 528

Genes: 12042

Processing Info

Dataset Stats

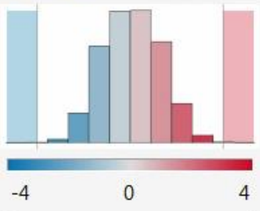
Meta Data

Persp.: 3

Tcga_samples: 156 (29.55%)

Genes: 1500 (12.46%)

Histogram



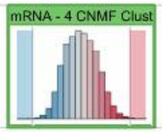
Colormap

Data-View Integrator

StratomeX

Show Only Selected Connections

Dynamic Connection Highlight Focus



More Information:

http://stratomex.caleydo.org

CALEYDO

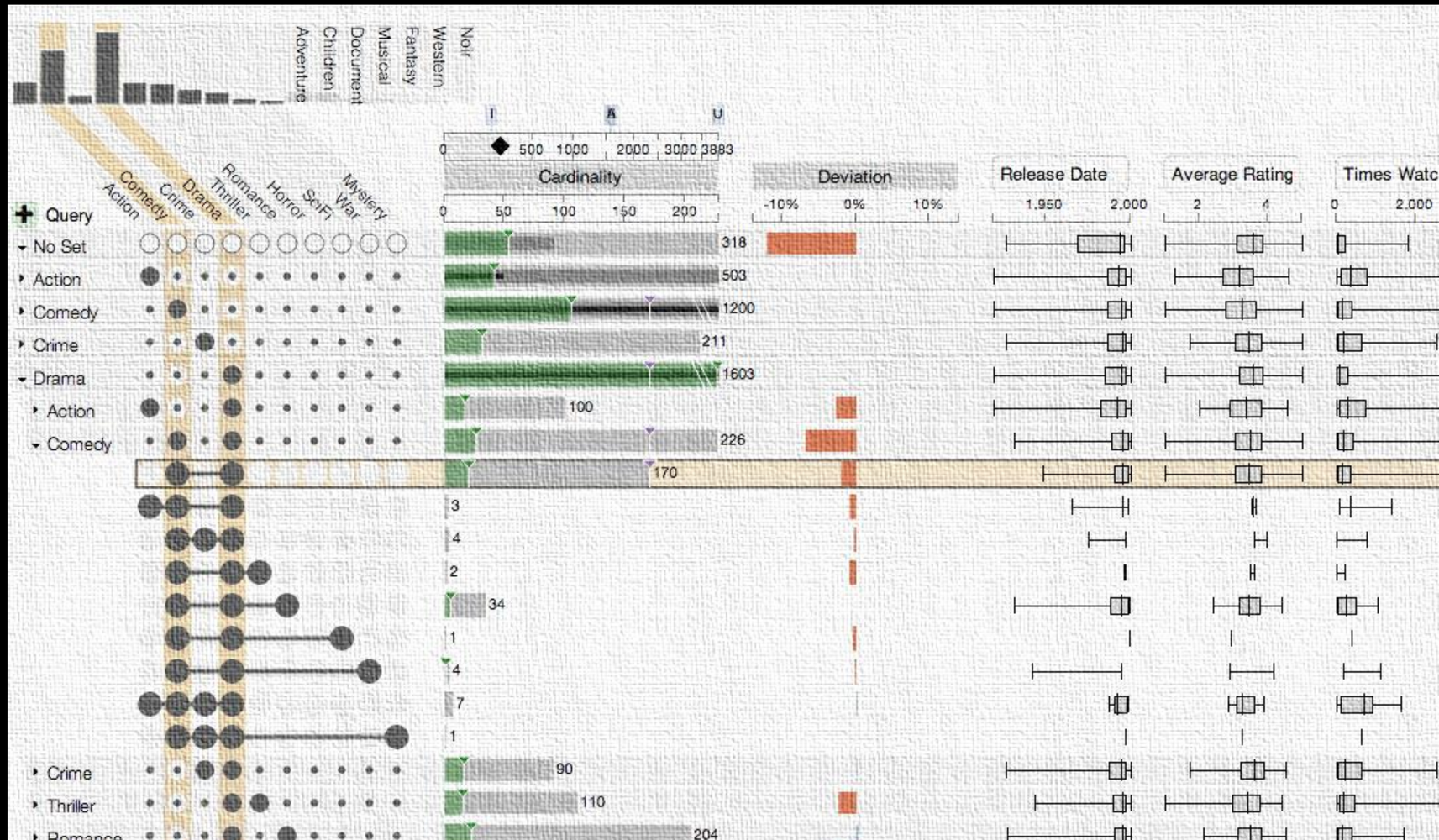
http://caleydo.org

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

HARVARD
School of Engineering
and Applied Sciences

Interactive Visual Data Analysis

?



Credits:

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