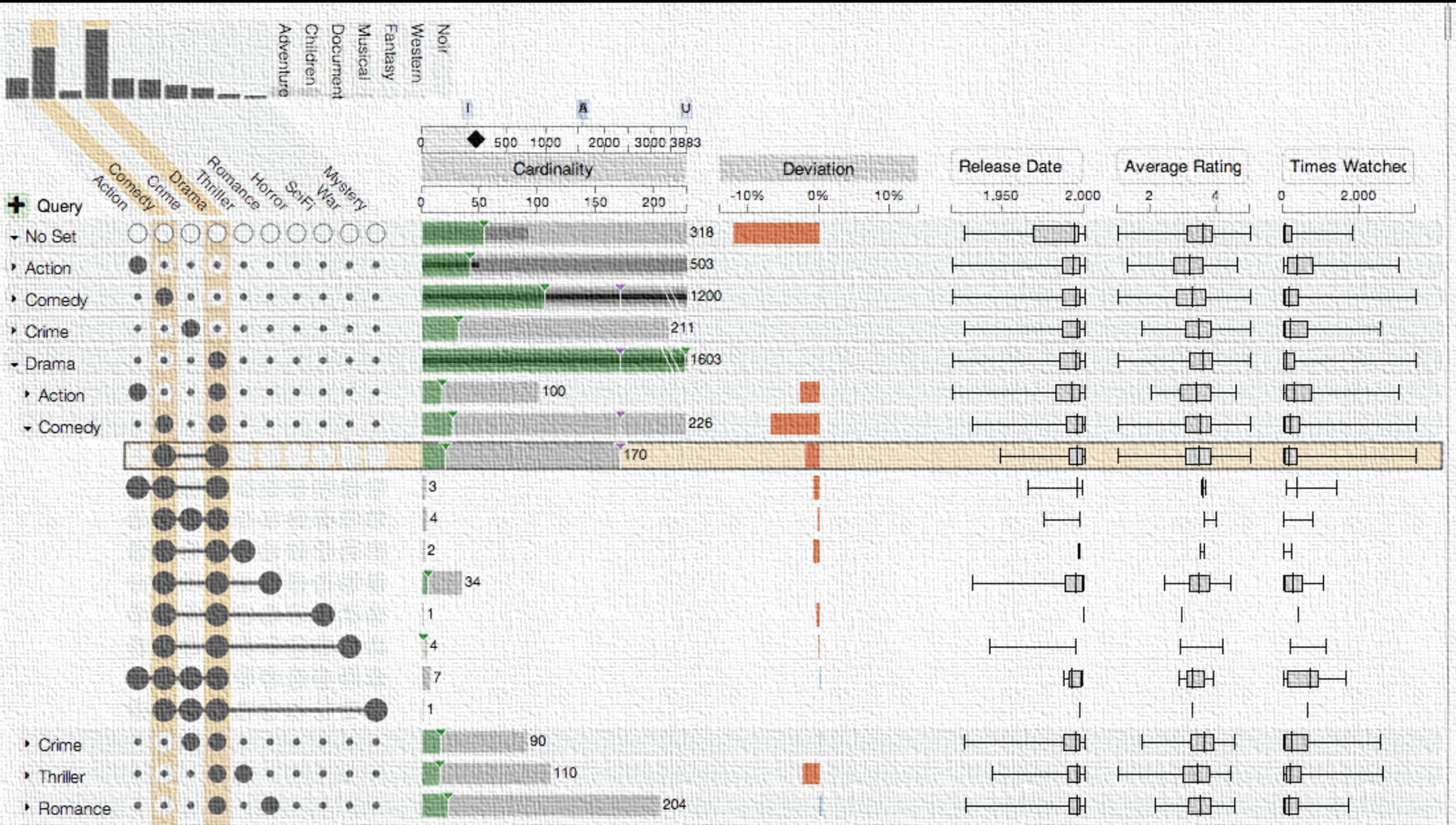
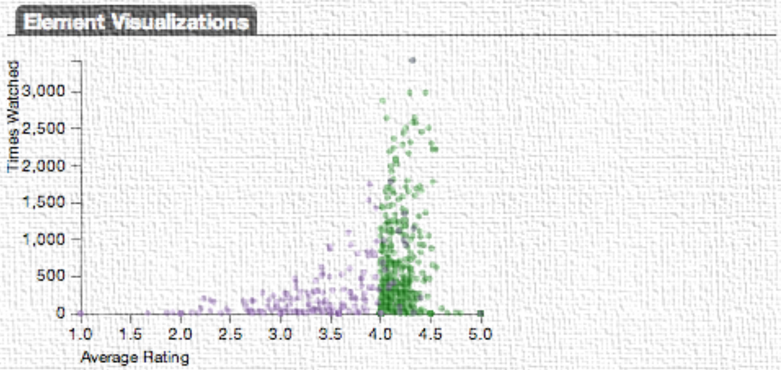
Alexander Lex
@alexander_lex
http://alexander-lex.com

April 23, 2015

HARVARD
School of Engineering
and Applied Sciences

UpSet: Visualization of 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

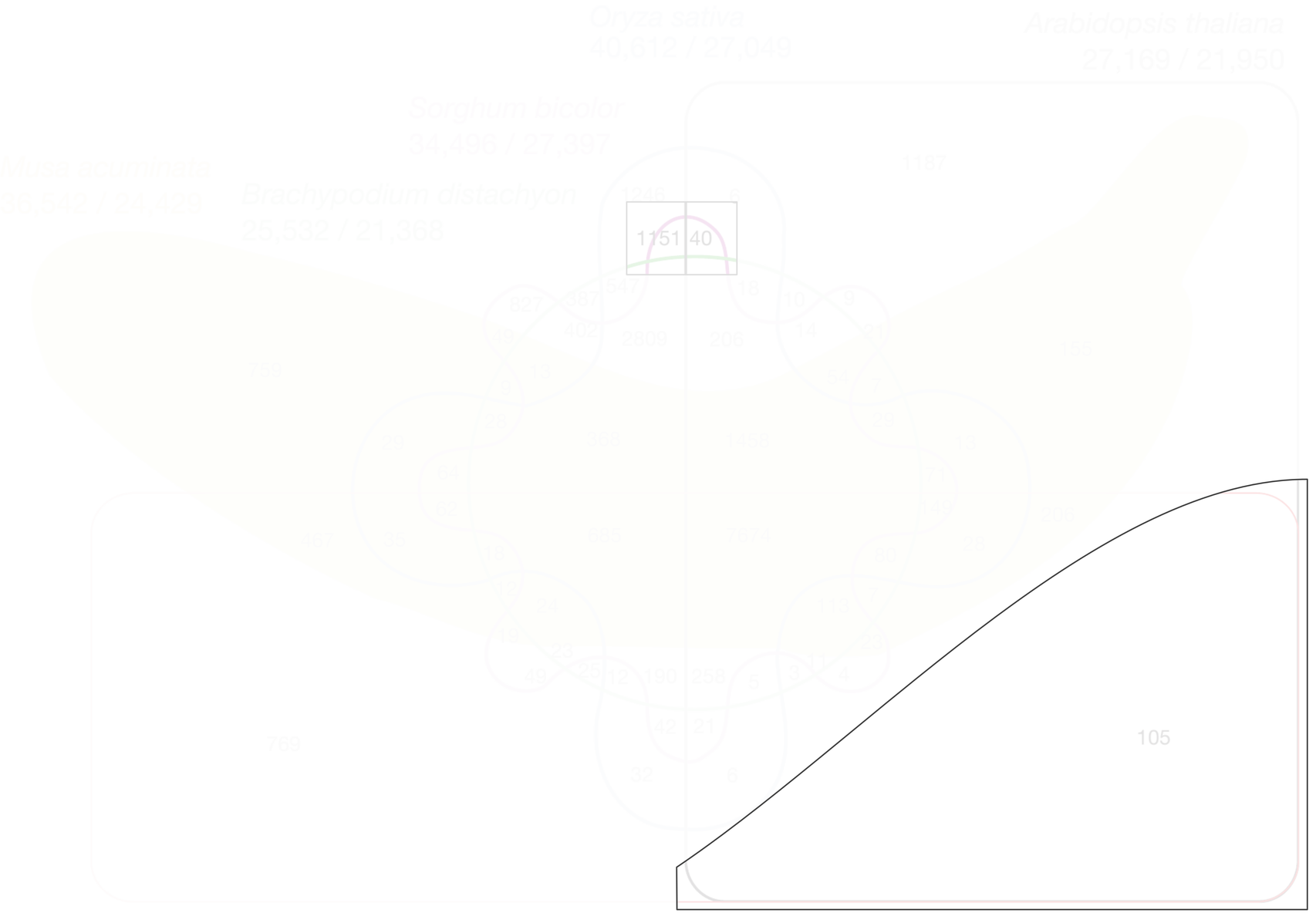
visualization *pictures*

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

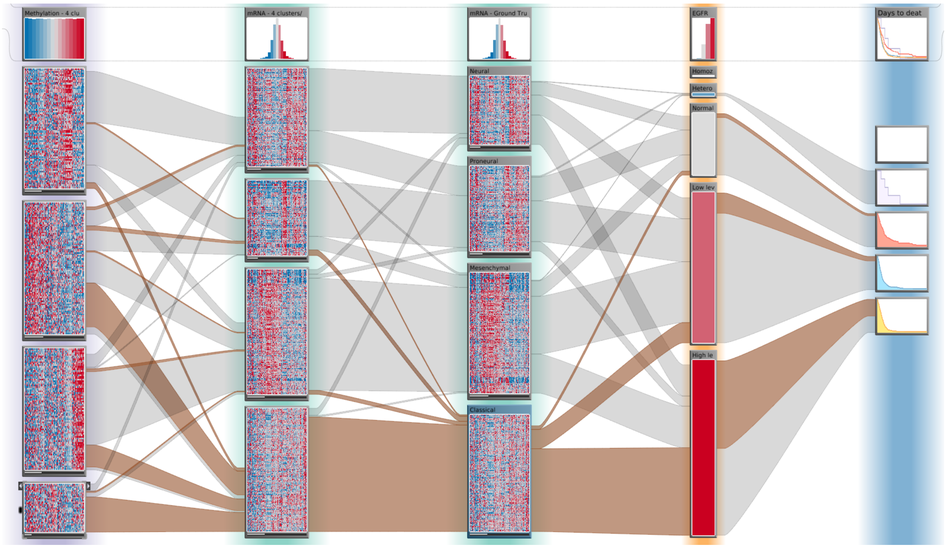
… makes data accessible

… combines strengths of
humans and computers

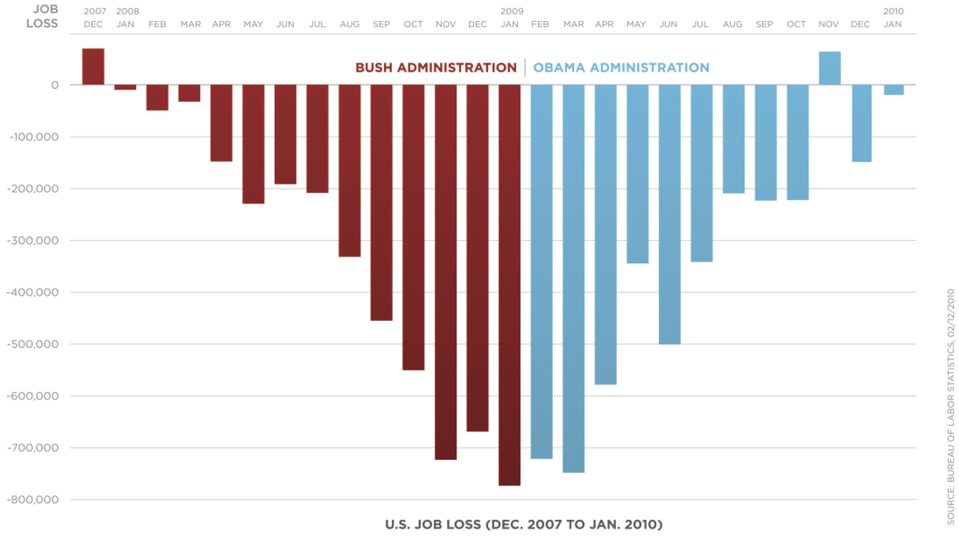
… enables insight

… communicates

Purpose of Visualization



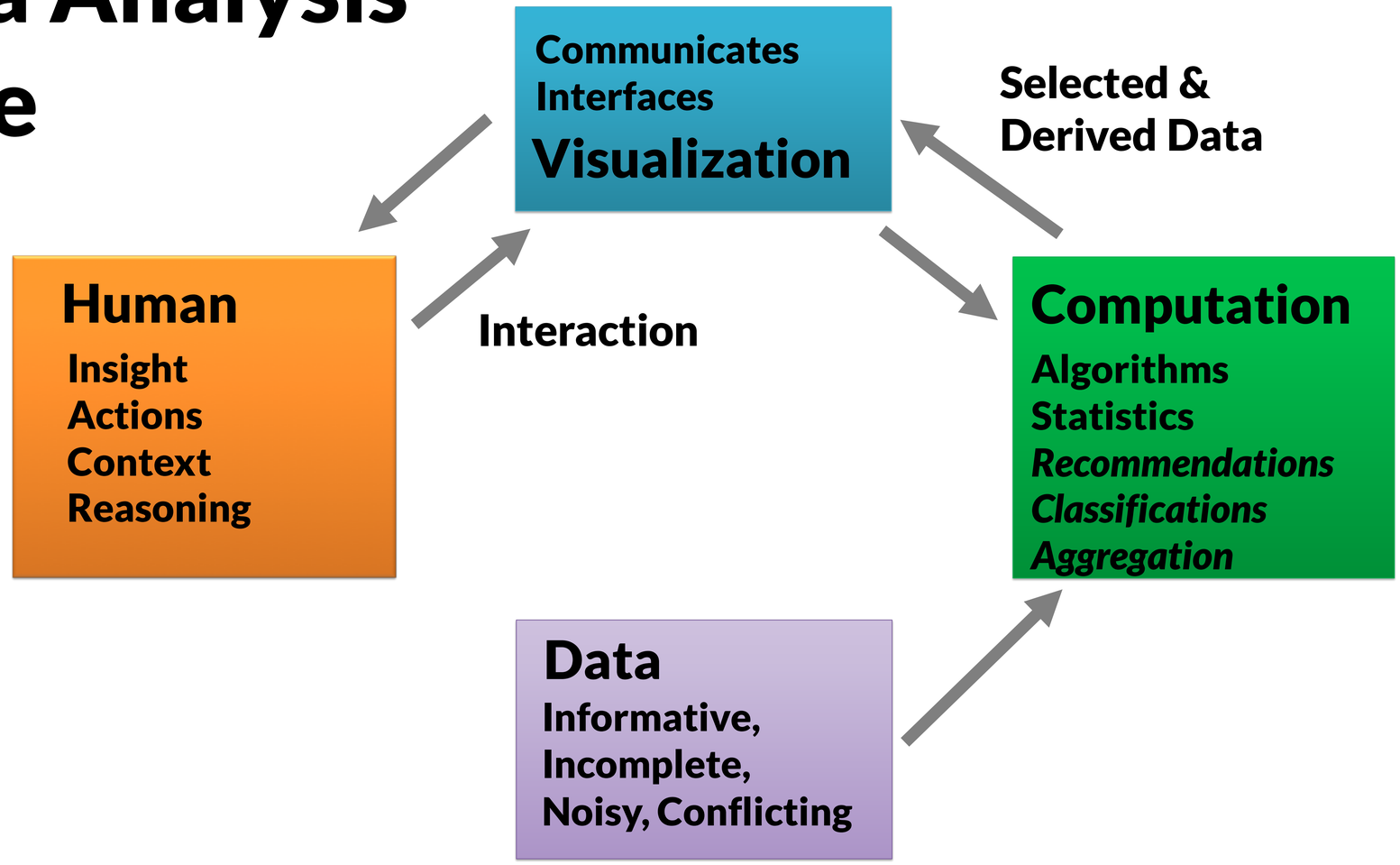
[Obama Administration]



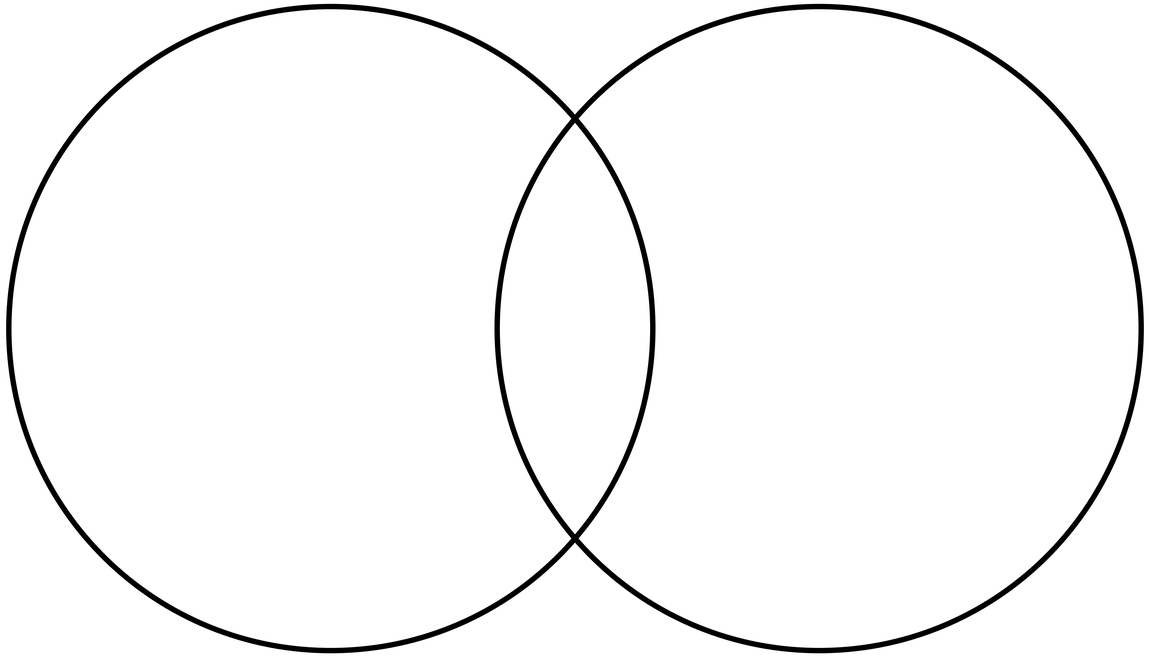
Open Exploration ⟷ **Communication**

Interacting with Data

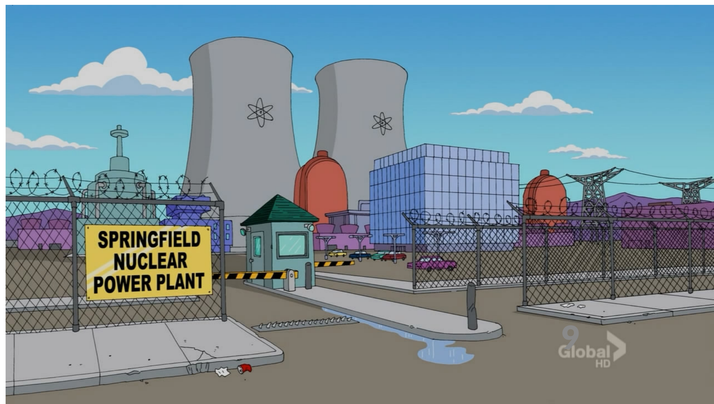
The Future of Data Analysis is (also) Interactive



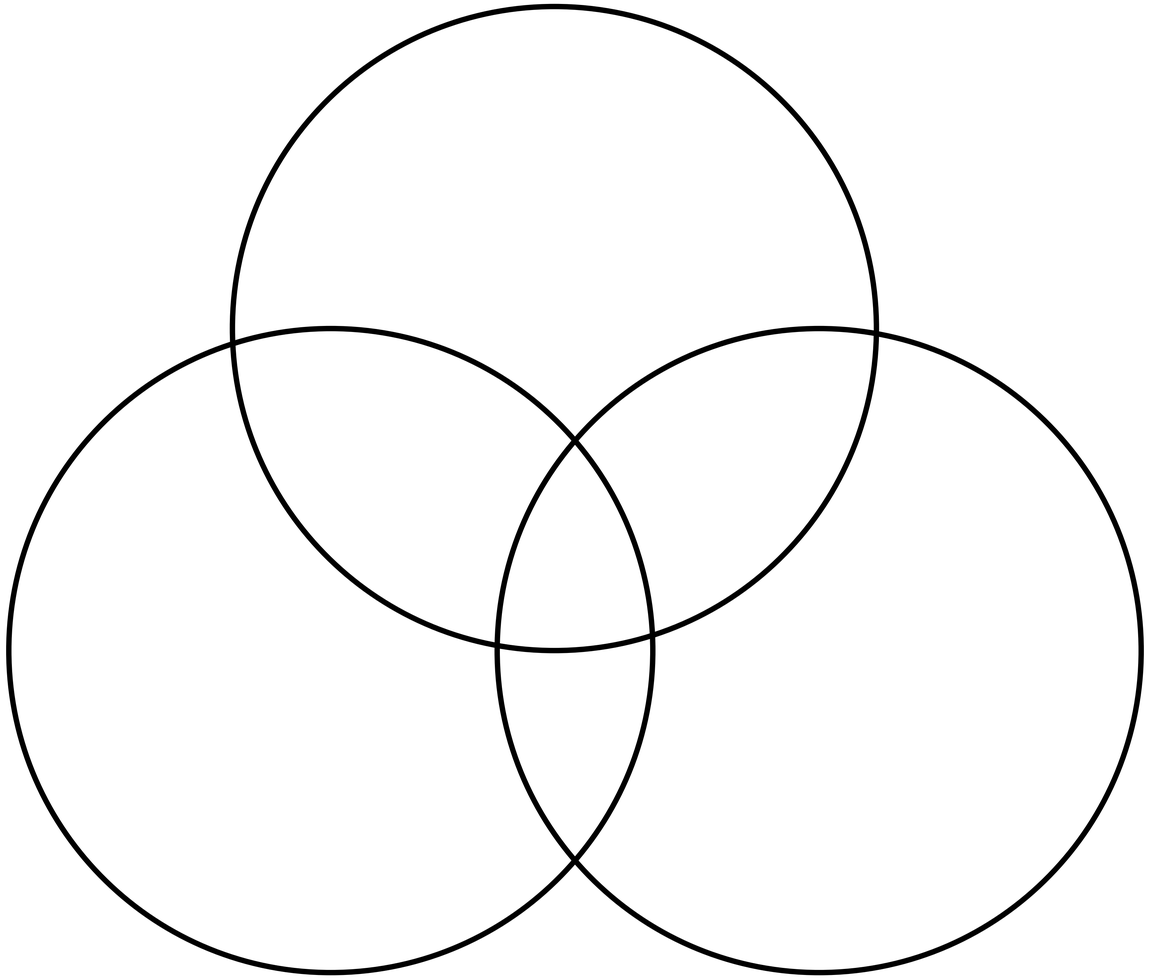
School



School **Power Plant**

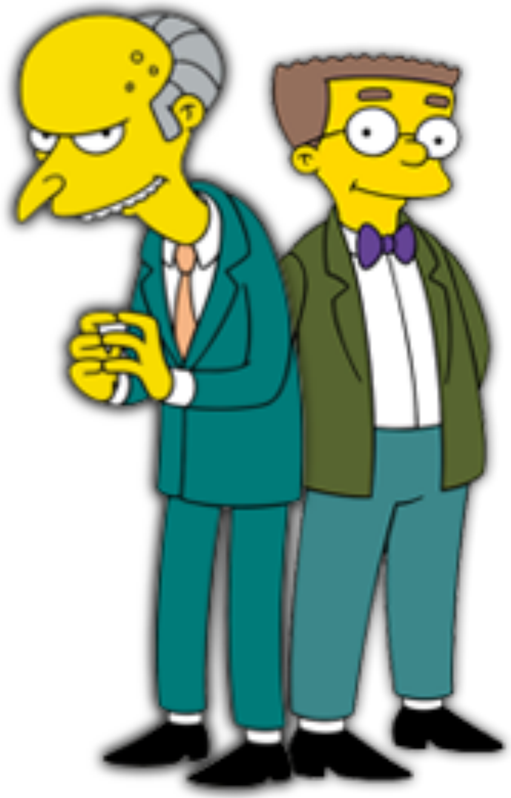


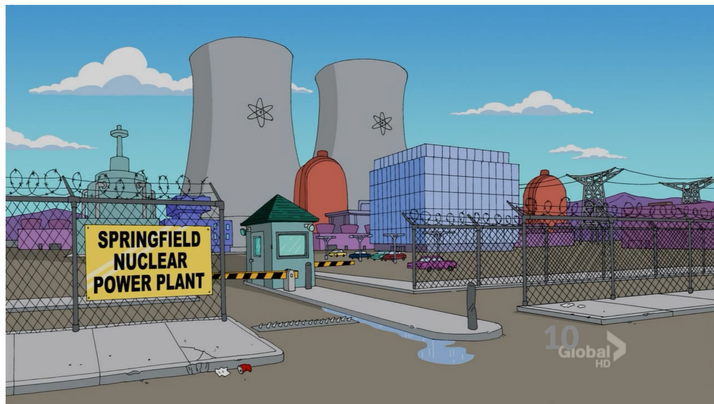
Evil



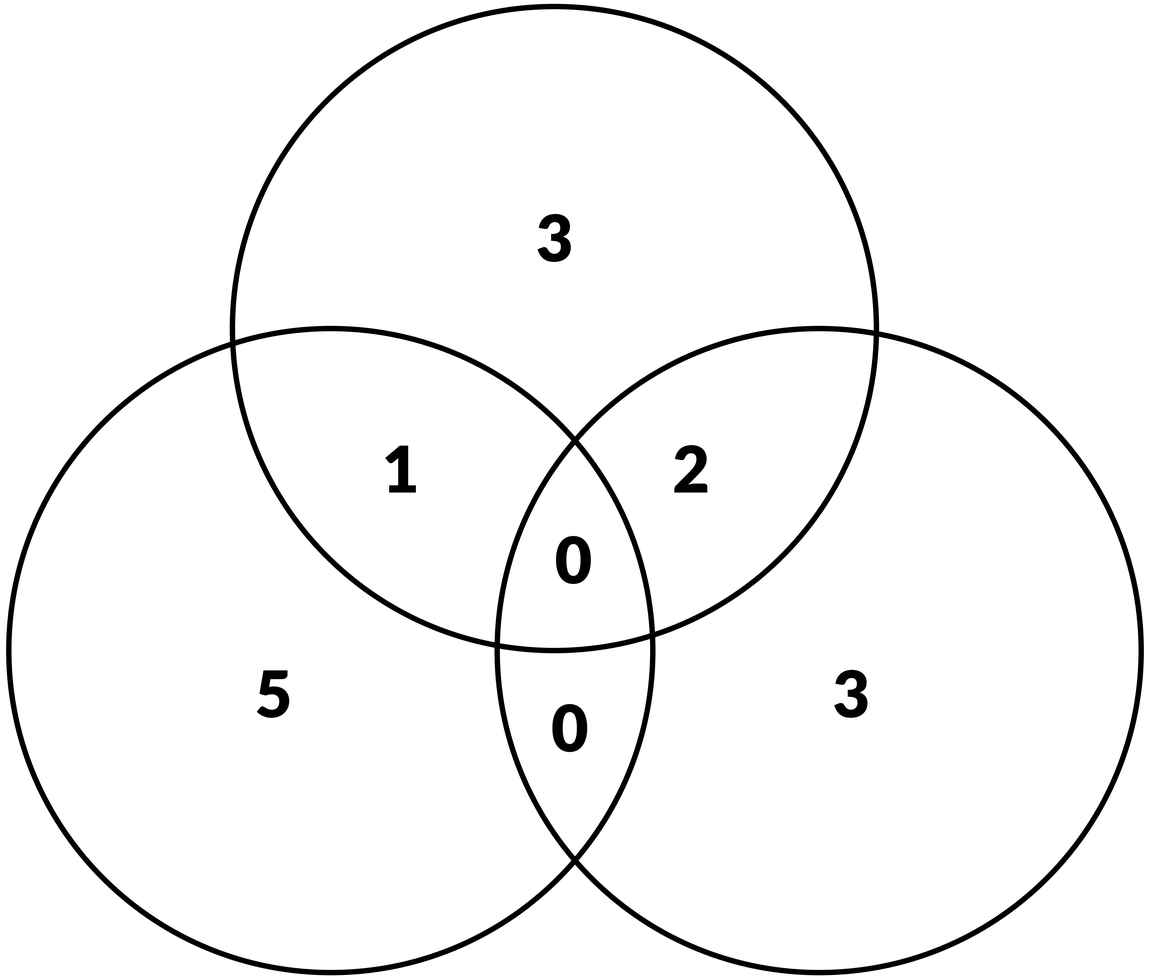
School

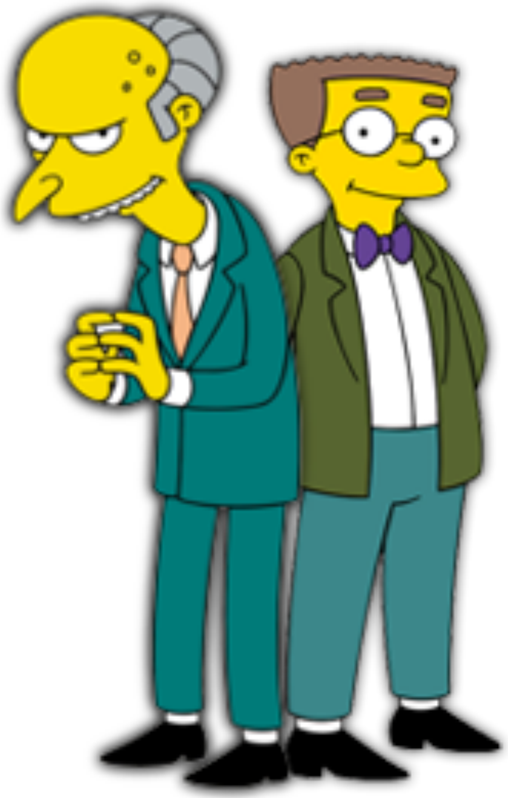
Power Plant



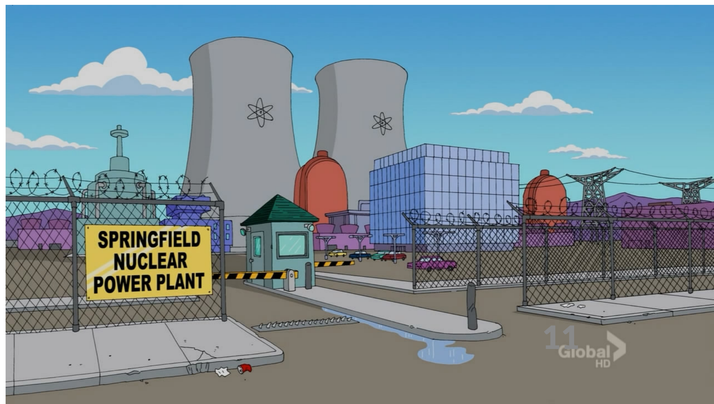



MATT GROENING


Evil
3
1
2
0
5
0
3
School
Power Plant

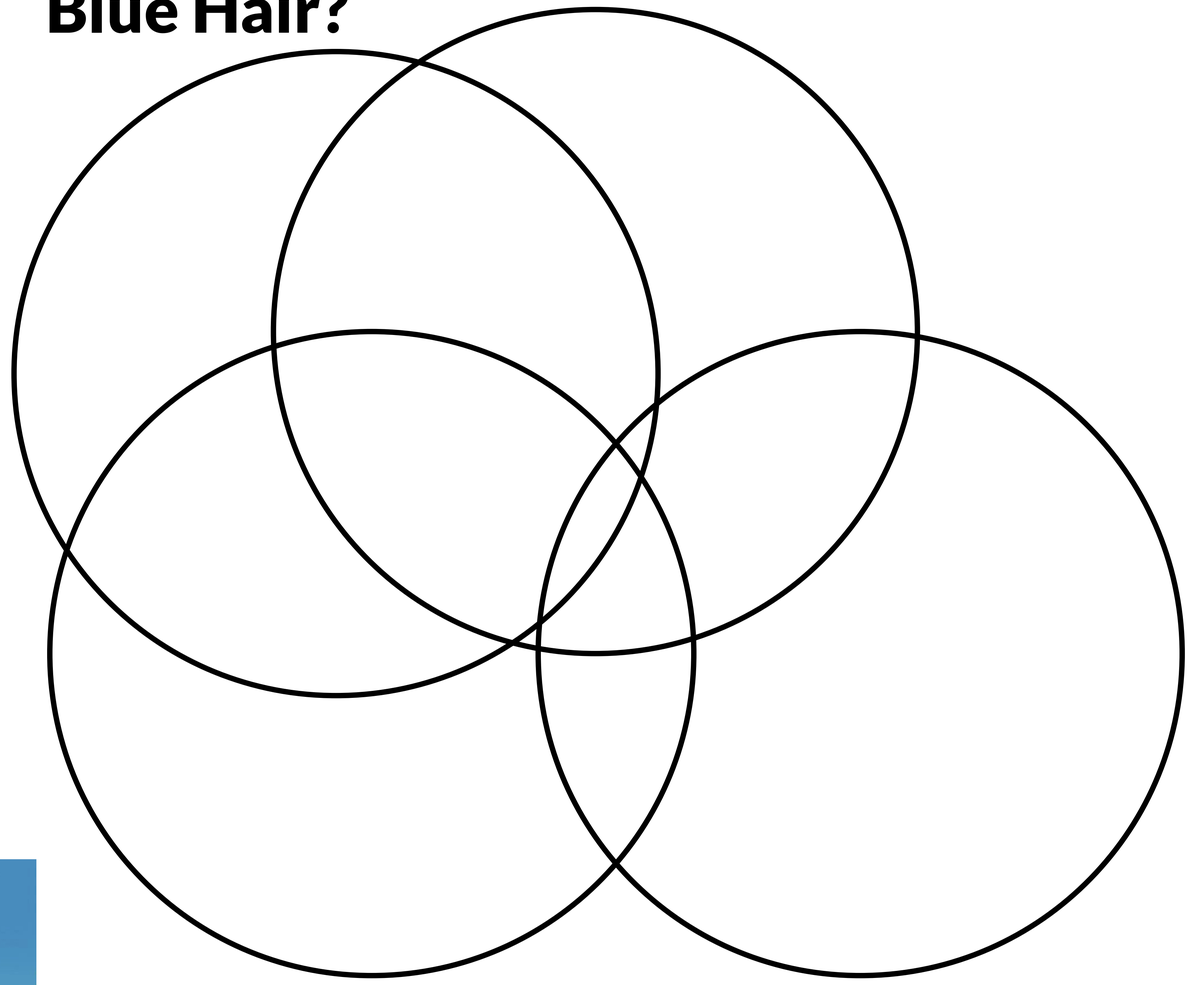


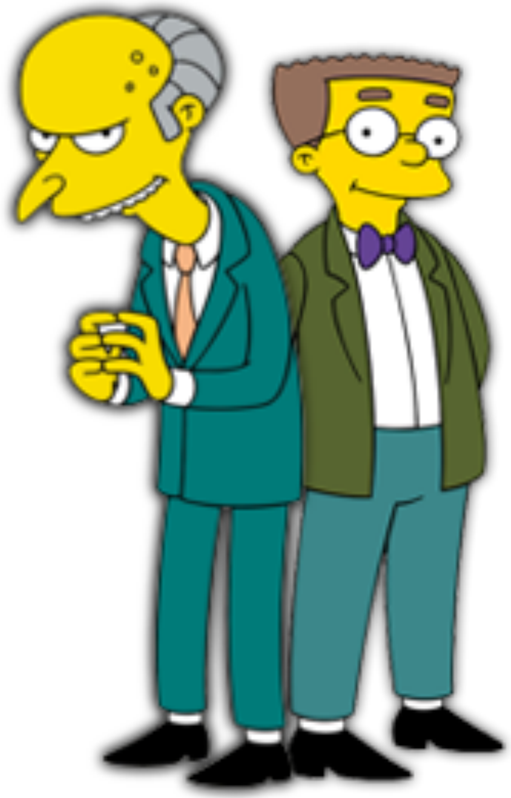

SPRINGFIELD ELEMENTARY SCHOOL


SPRINGFIELD NUCLEAR POWER PLANT

Evil

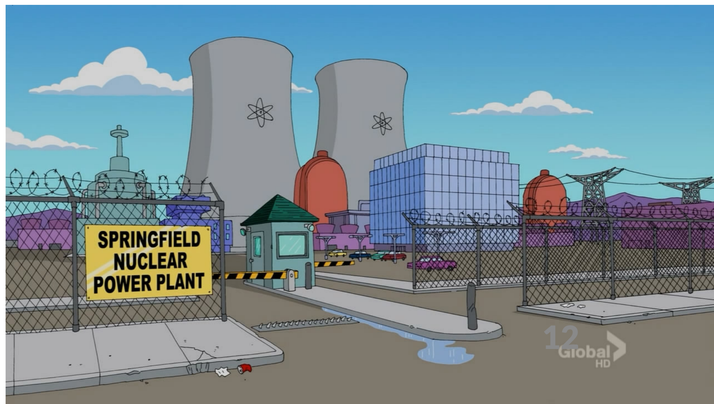
Blue Hair?





School

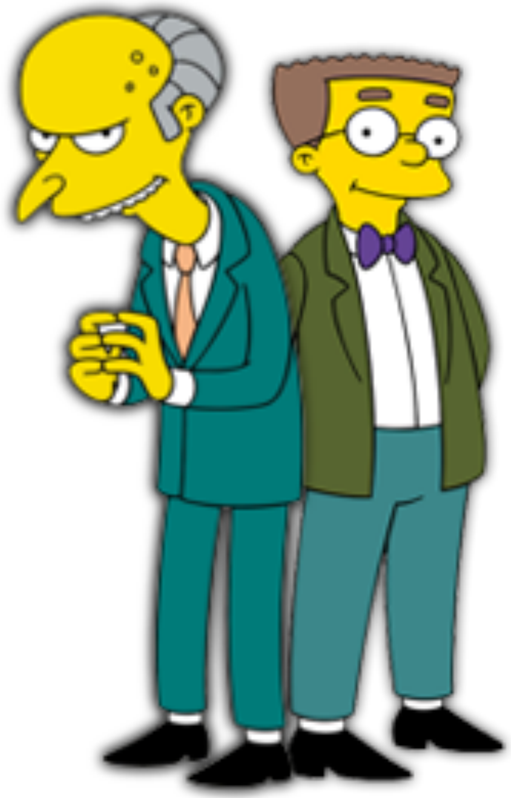
Power Plant



Blue Hair?

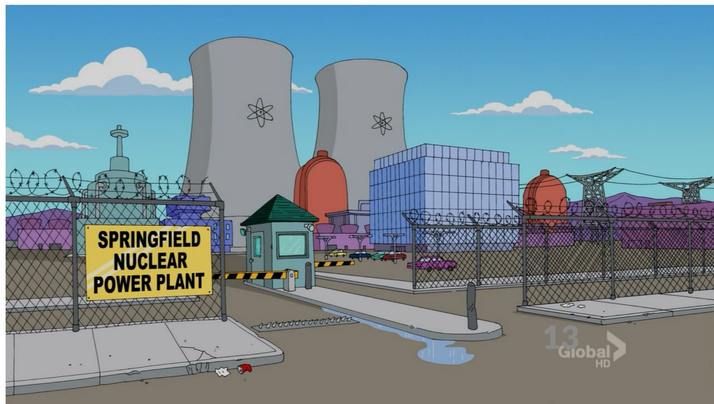
Evil

Duff Beer?

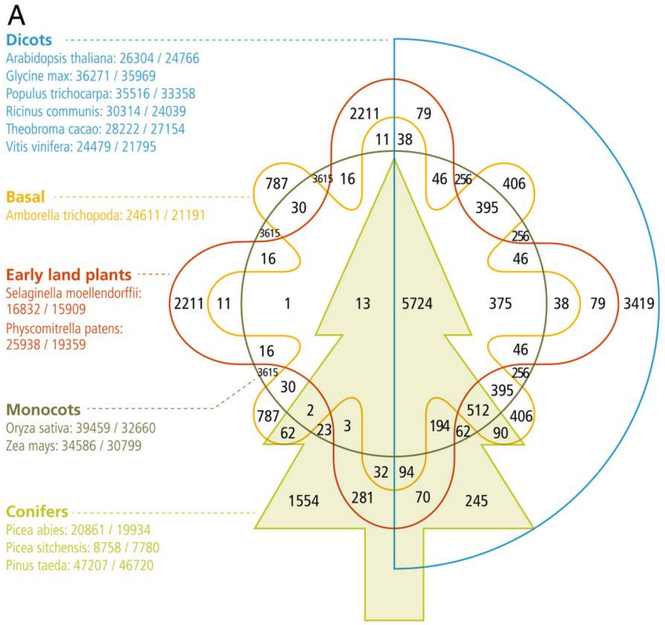


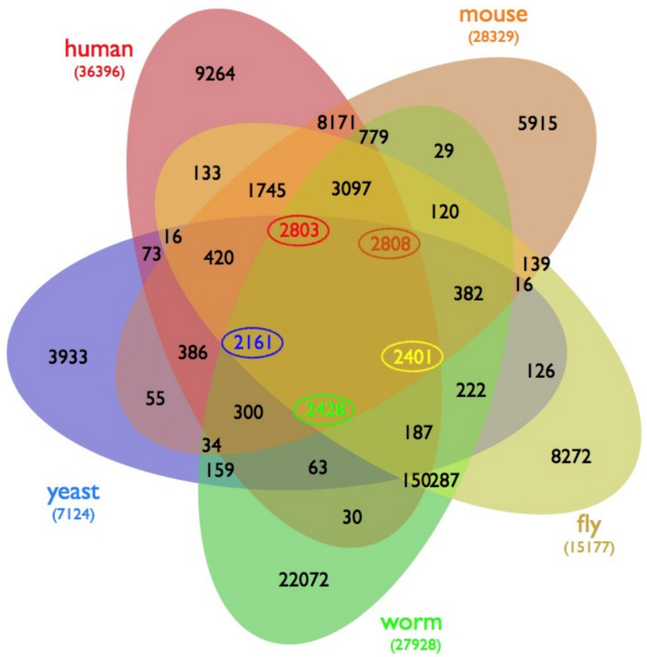
School

Power Plant

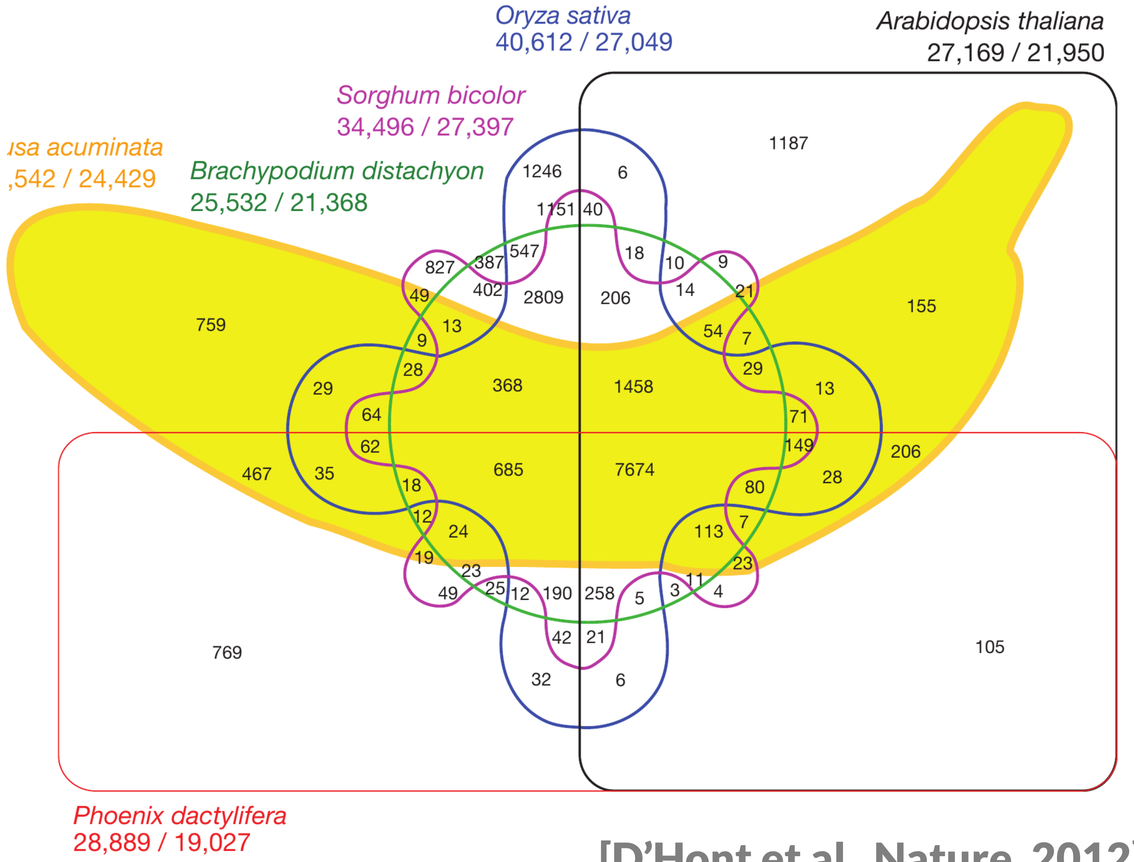


[Neale et al., BMC Genome Biology, 2014]

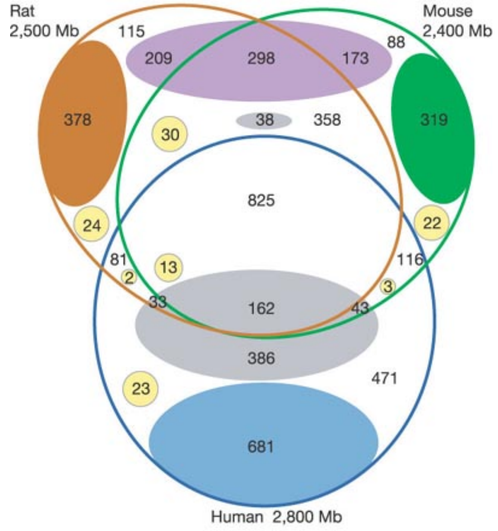


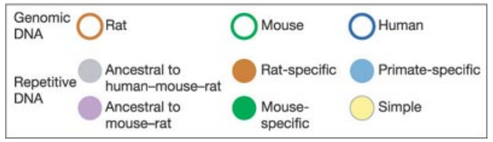


[Wiles et al., BMC Systems Biology]

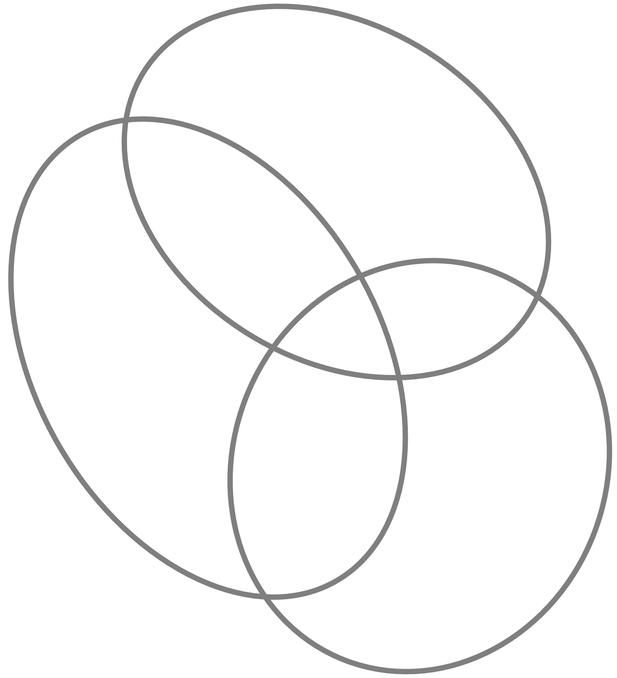


[D'Hont et al., Nature, 2012]



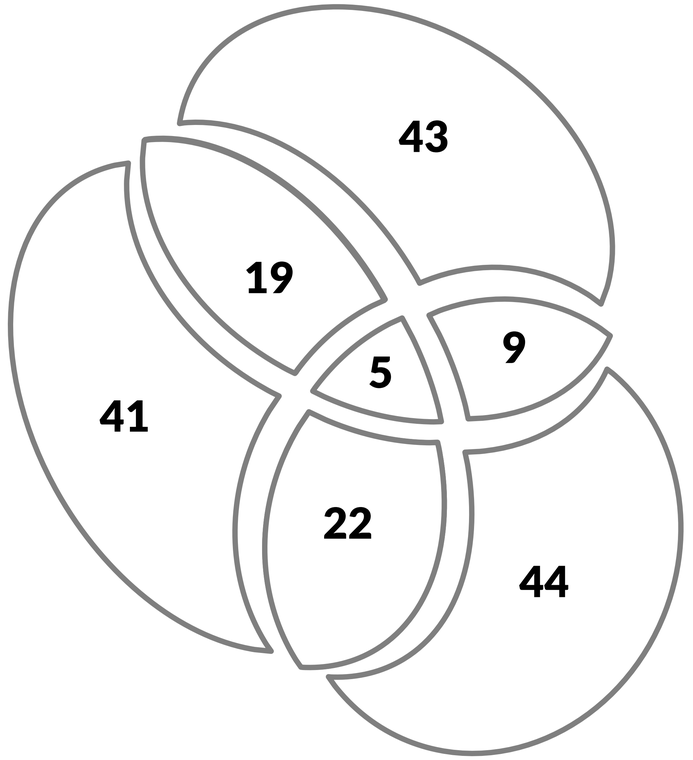


[Gibbs et al., Nature, 2004]



[created with EulerAPE]

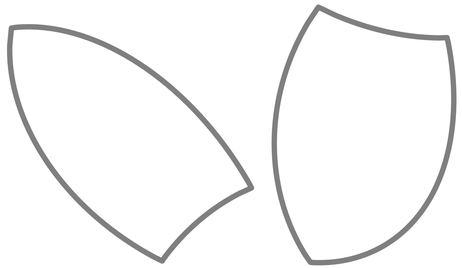
22 ? 19



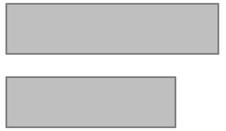
[created with EulerAPE]

Set Vis Goals

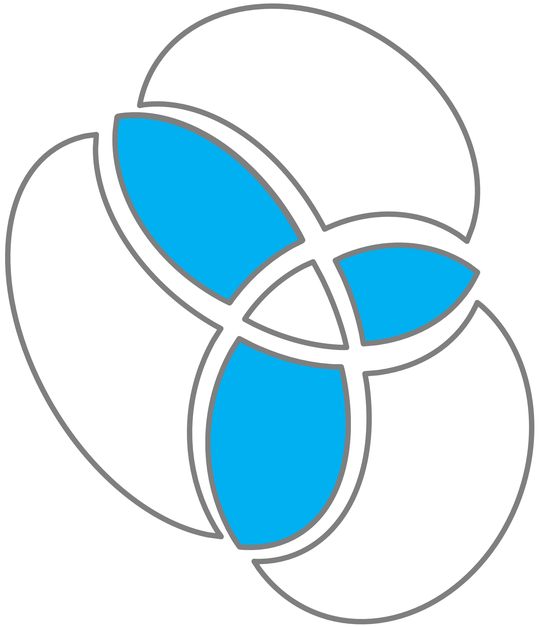
1. Efficient visual encoding



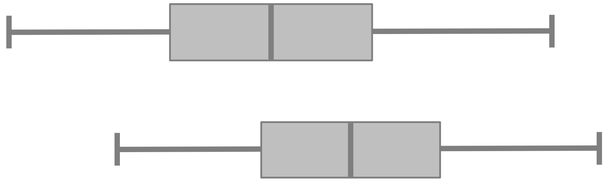
vs.



2. Creating complex slices of a dataset



3. Visualize attributes



[Movie Lens Dataset]

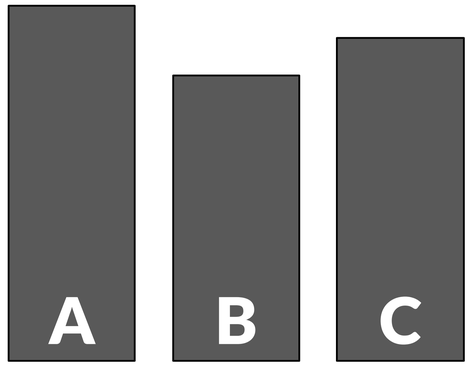
Attribute Details

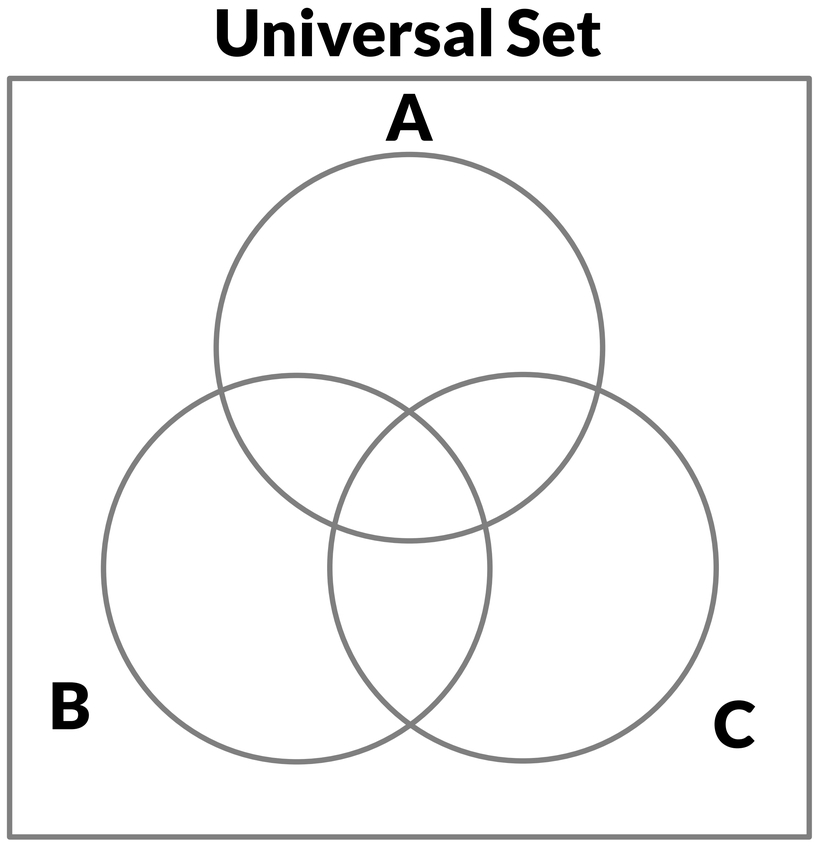
Visualizing Intersections

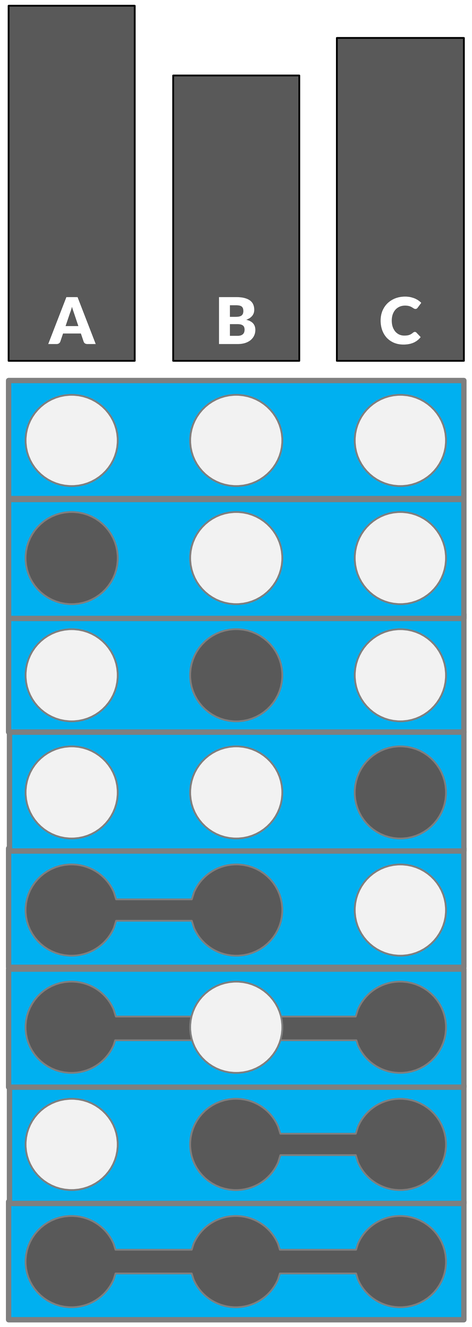
Visualizing Properties

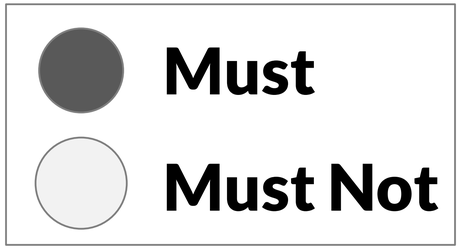
Element List & Queries

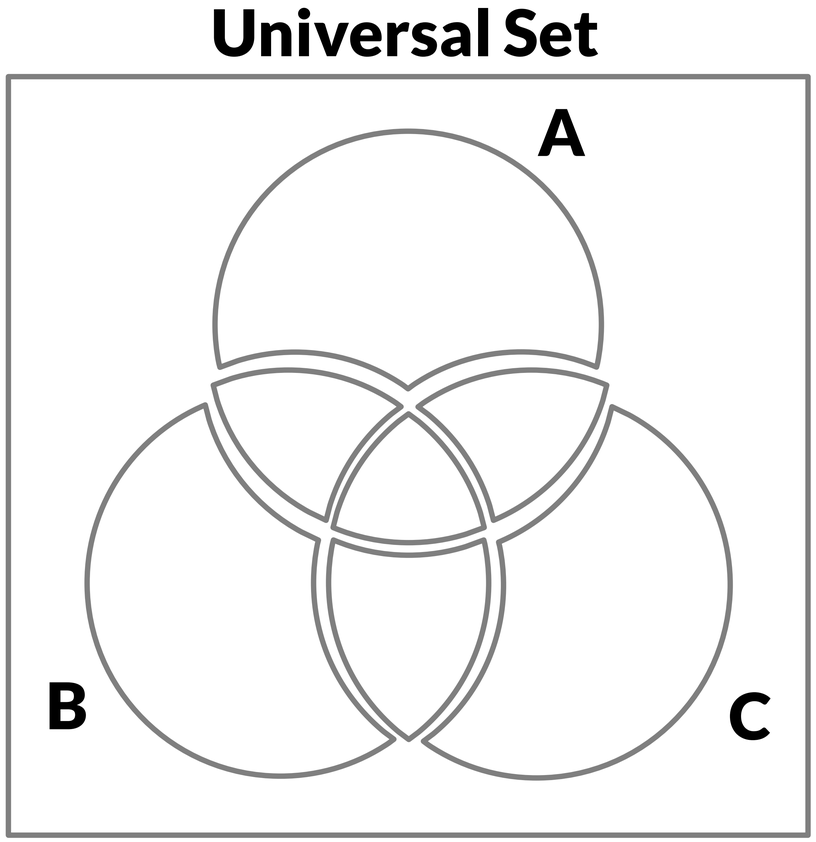
Visualizing Intersections

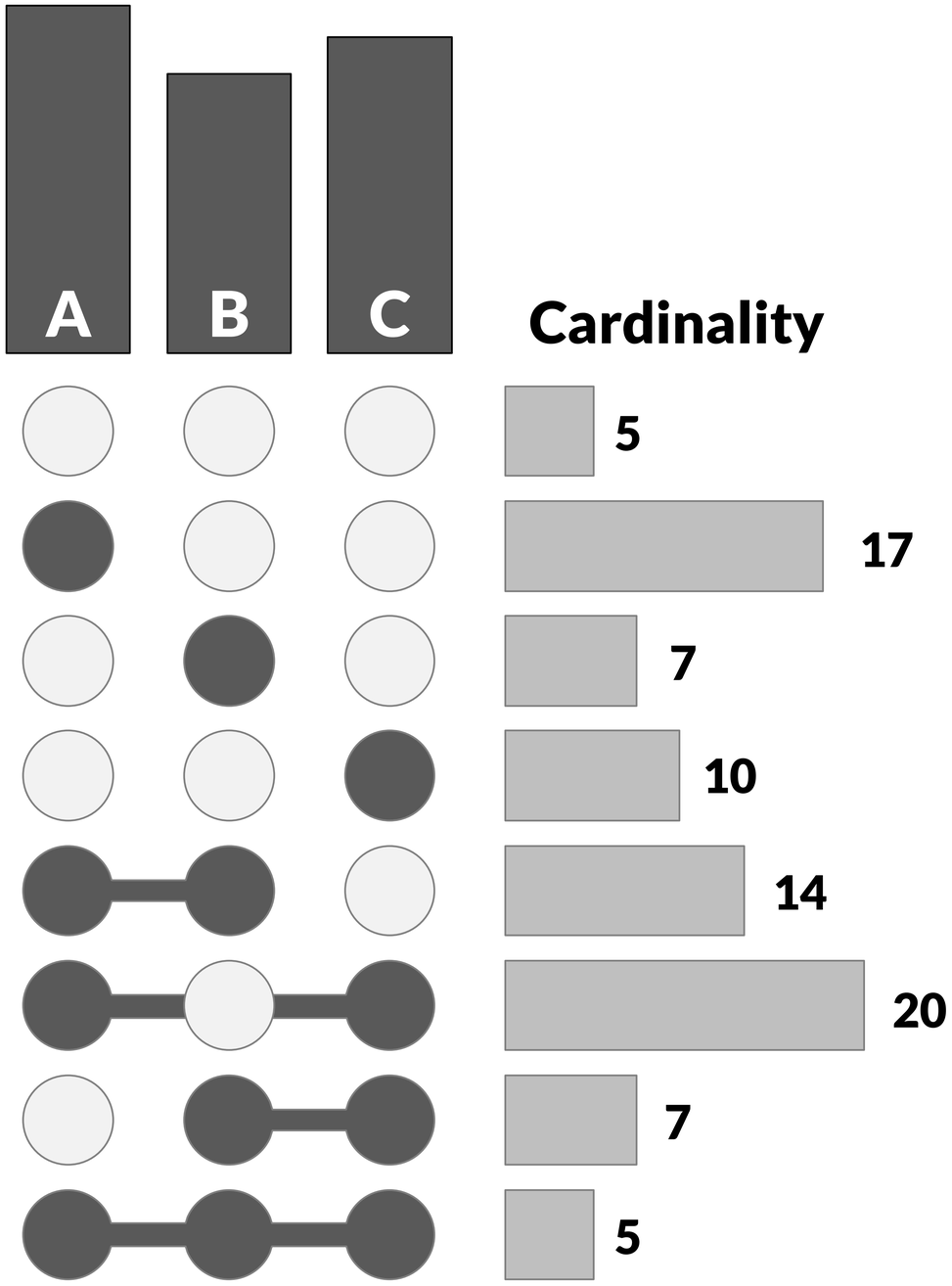

A
B
C

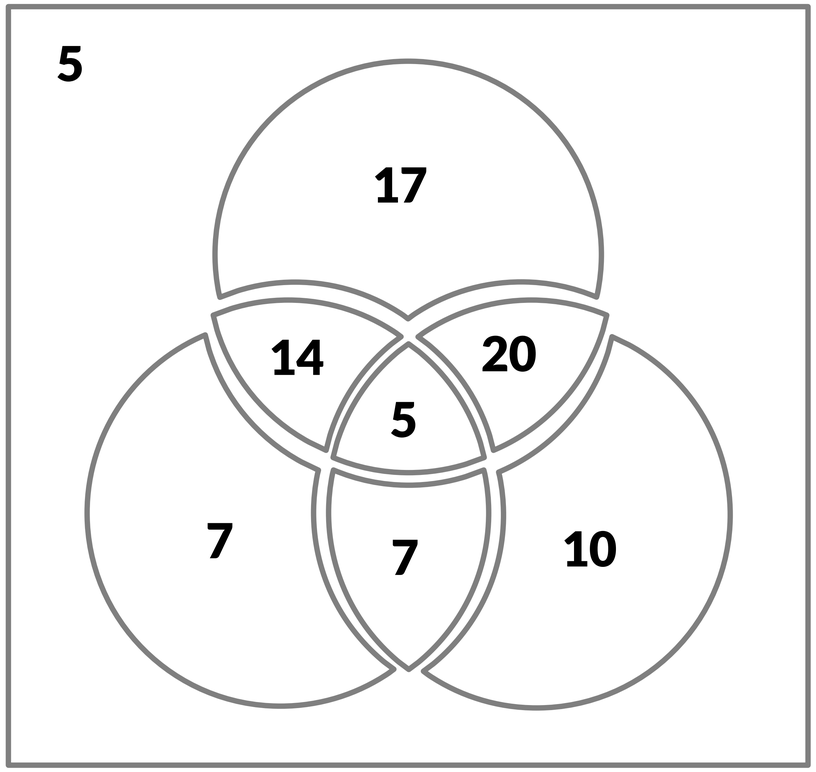

Universal Set
A
B
C


A
B
C

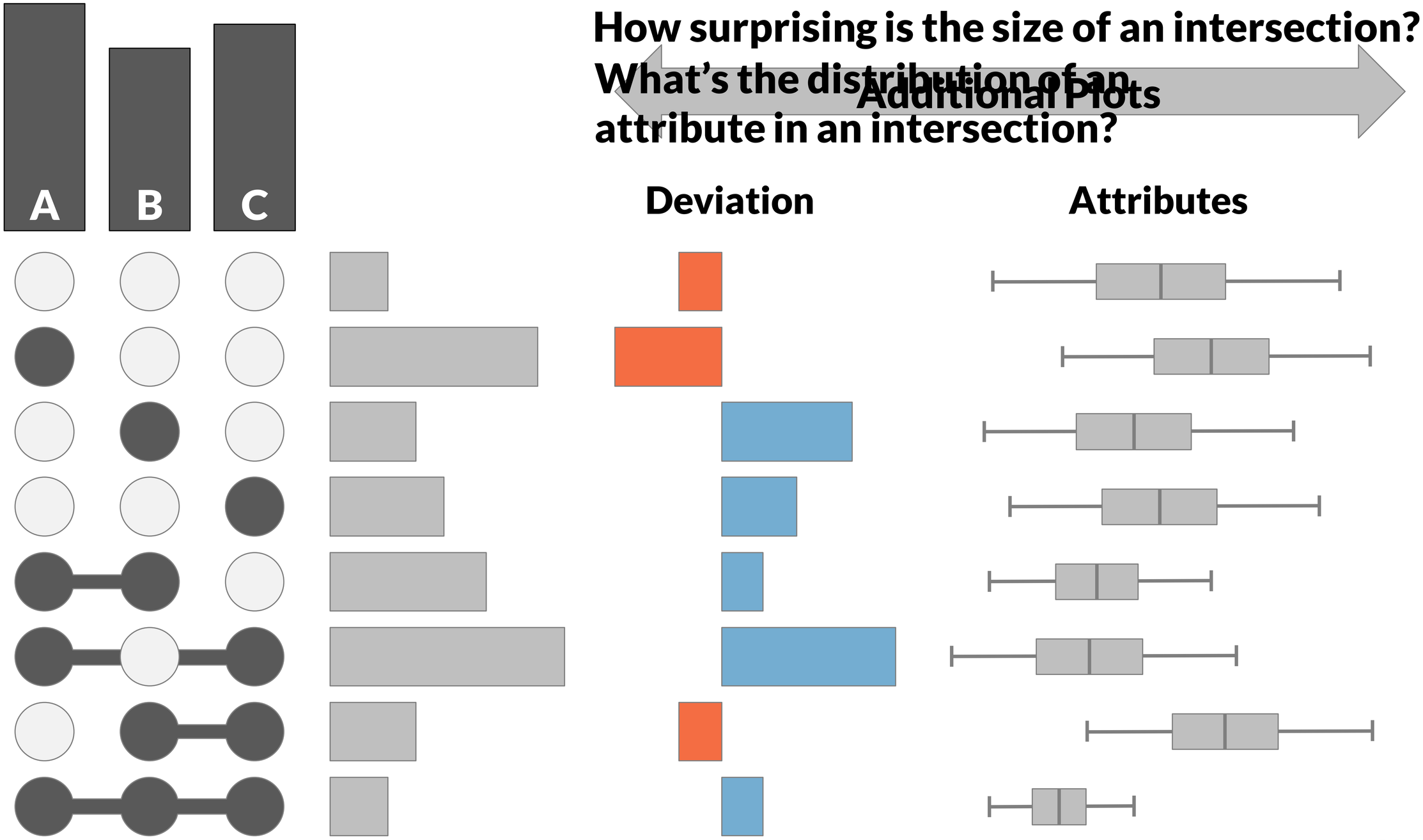

Must
Must Not

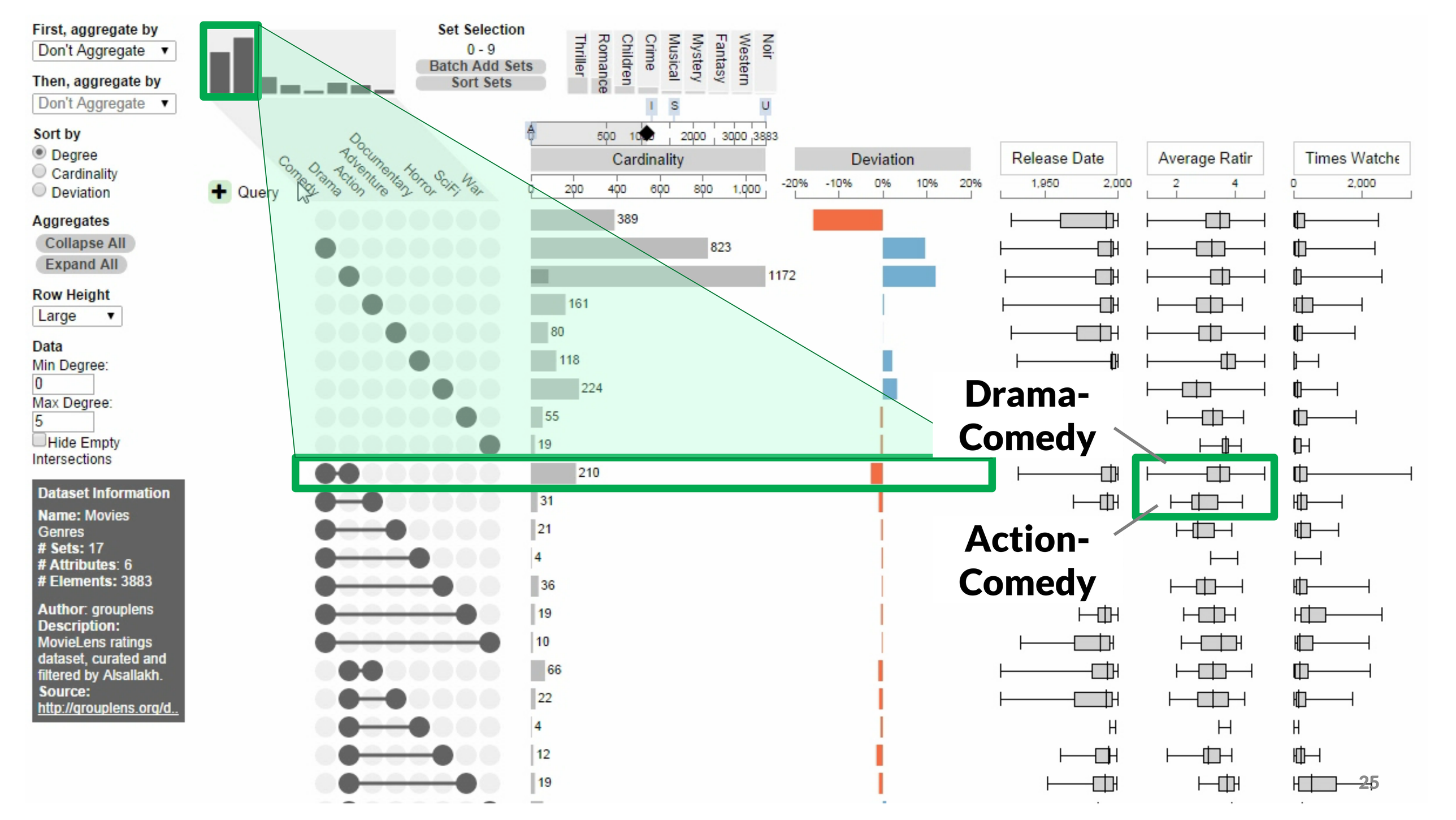

Universal Set
A
B
C


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

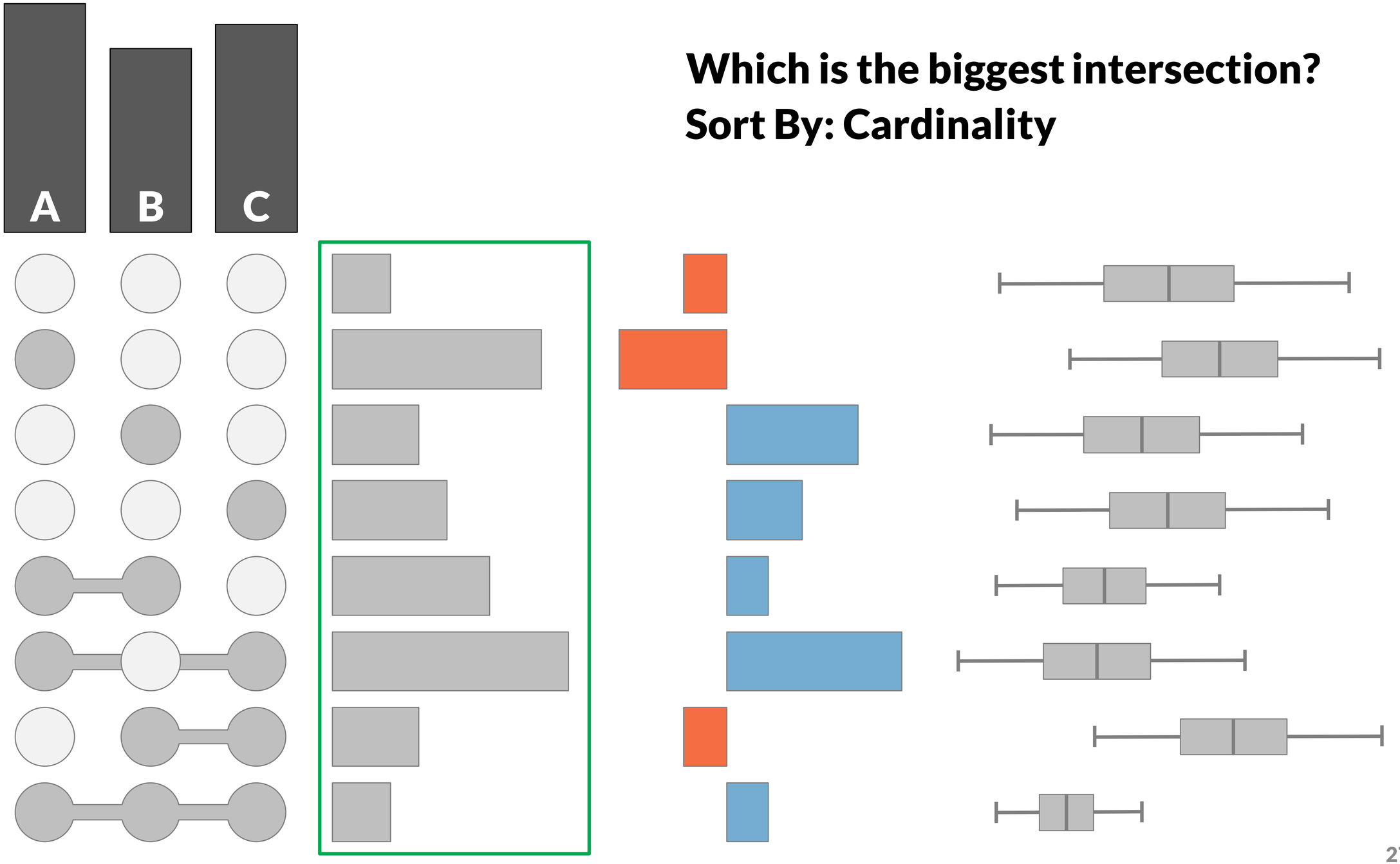

5
17
14
20
5
7
7
10

Plotting Attributes

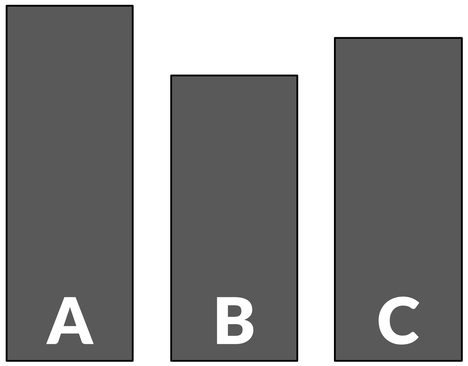

How surprising is the size of an intersection?
What's the distribution of an attribute in an intersection?
Additional Plots
A
B
C
Deviation
Attributes


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

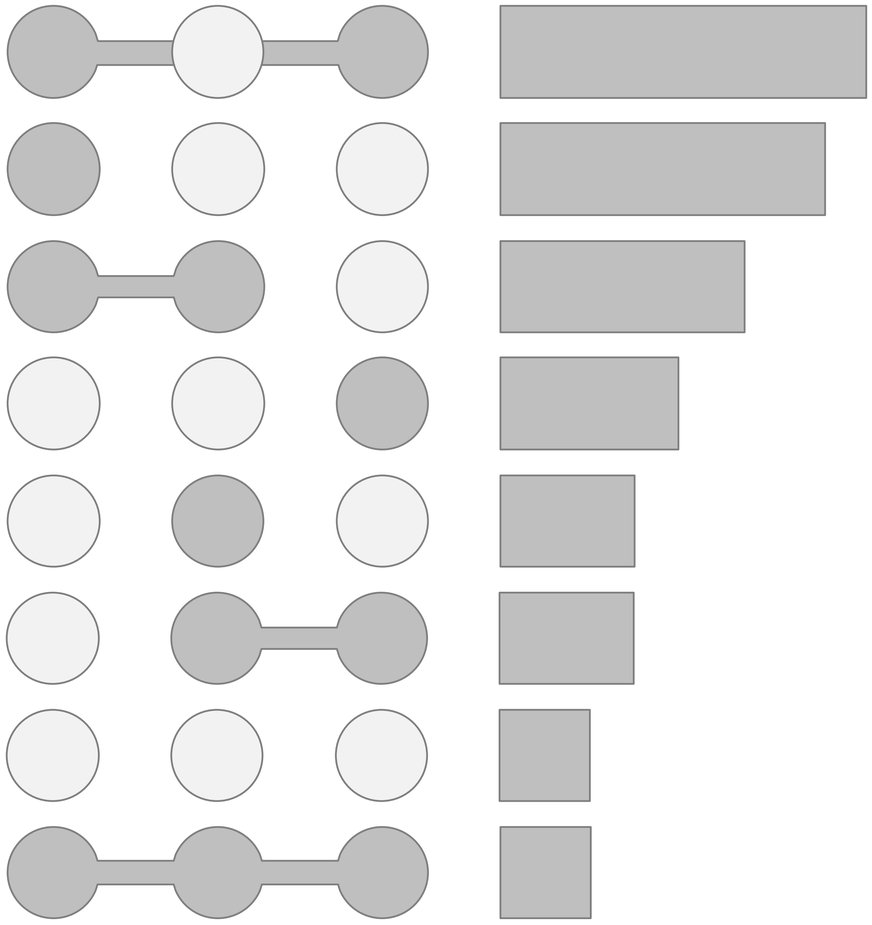
Sorting

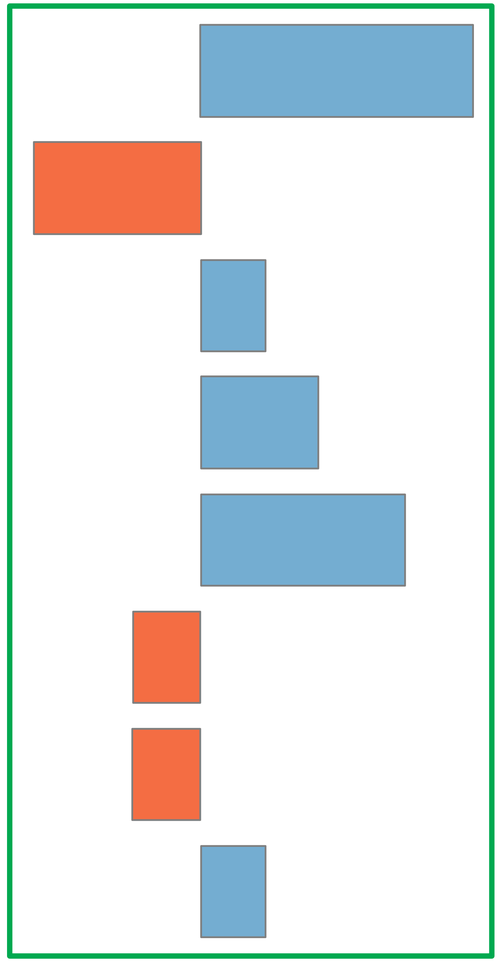


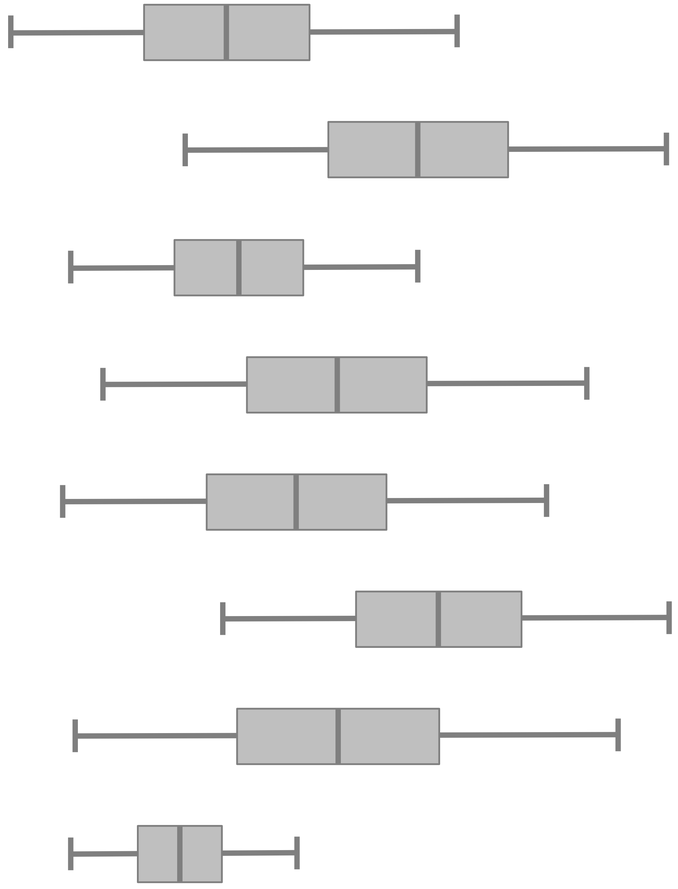
Which is the biggest intersection?
Sort By: Cardinality

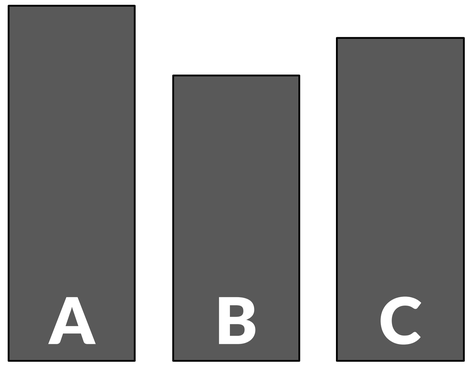


Which is the most ‘surprising’ intersection?
Sort By: Deviation

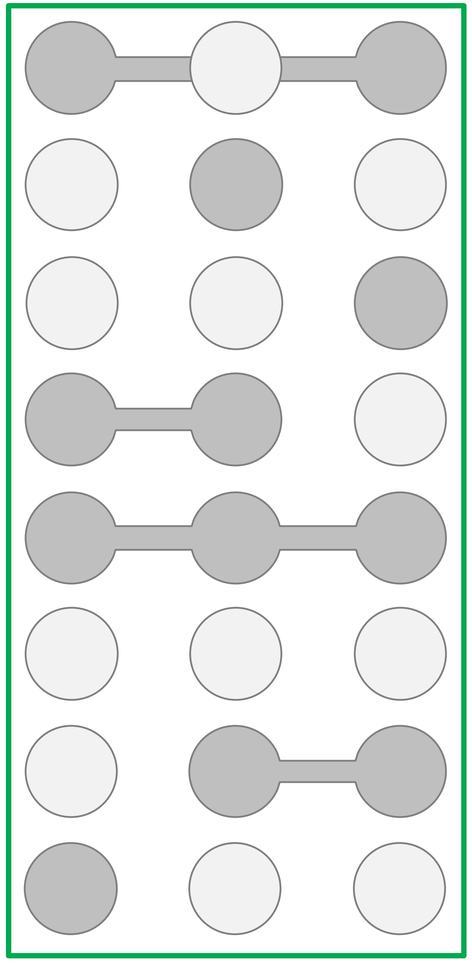


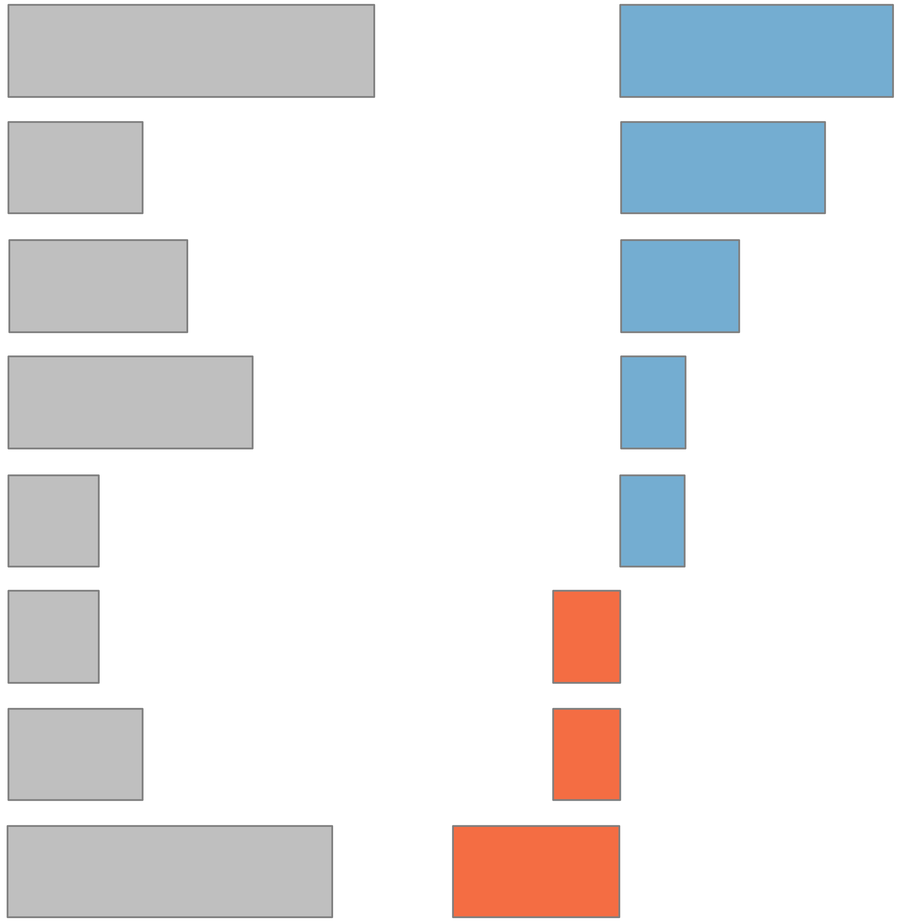


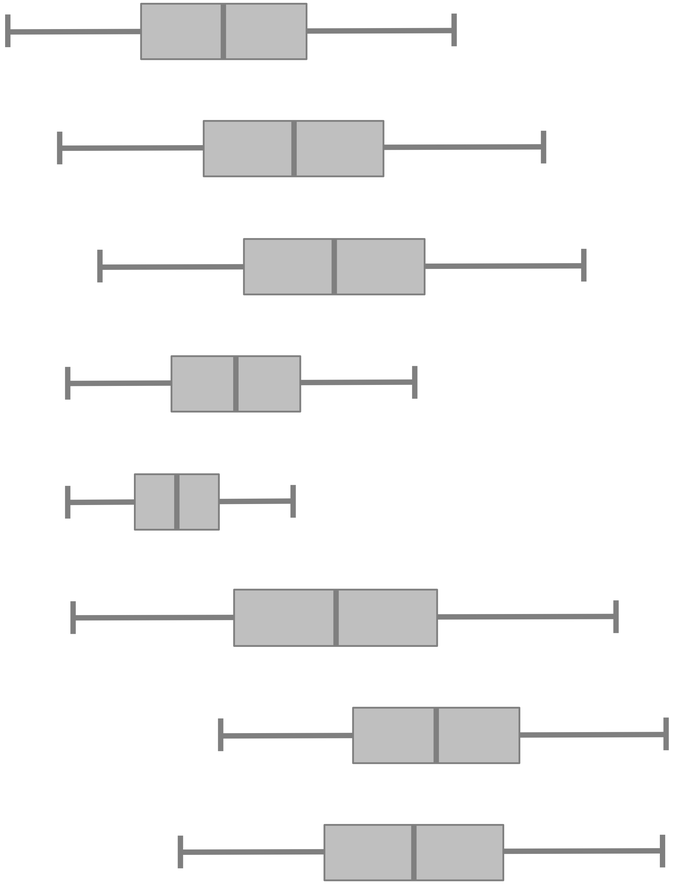


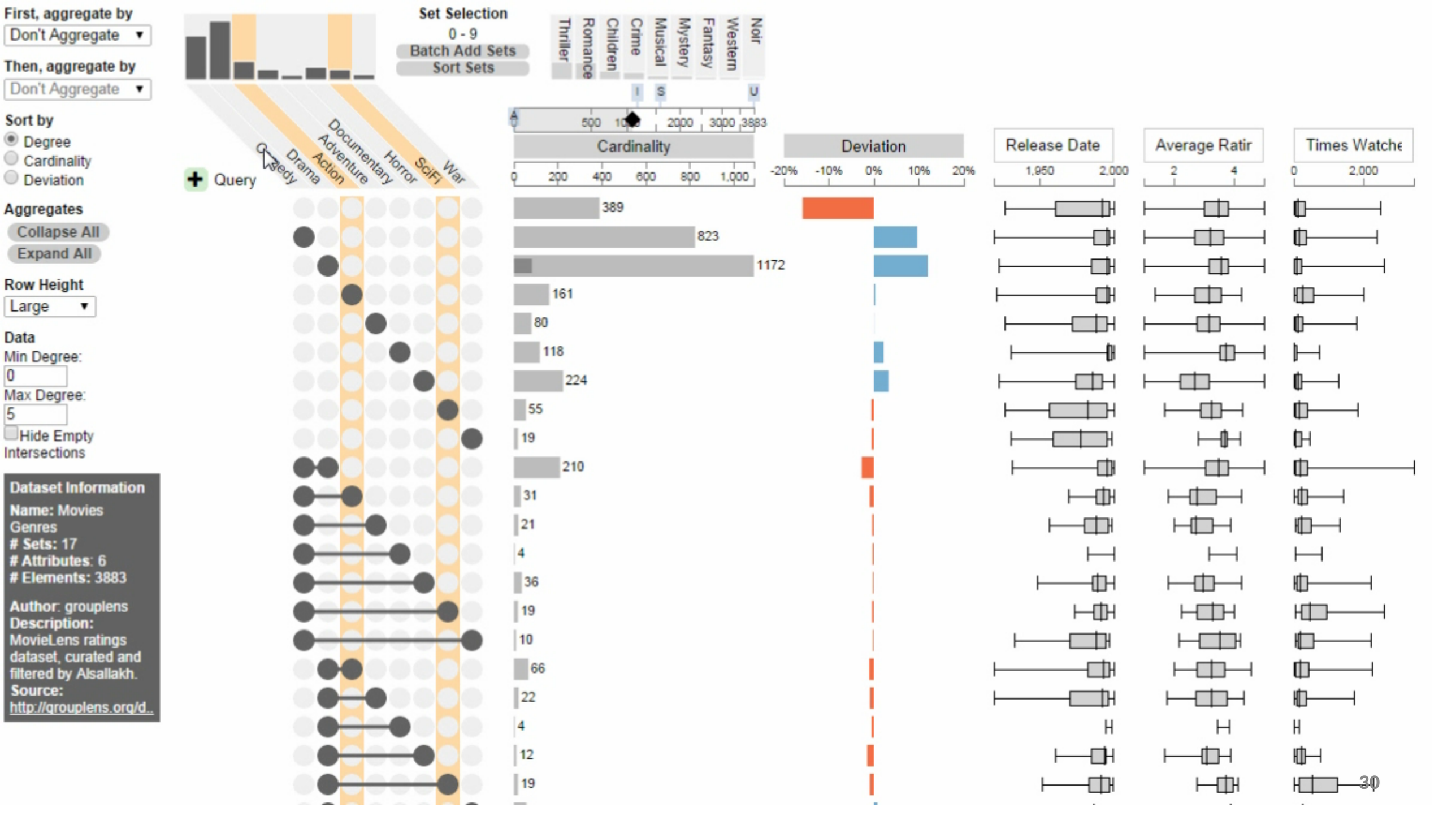


What are the properties of the intersections involving ‘A’?
Sort By: Set

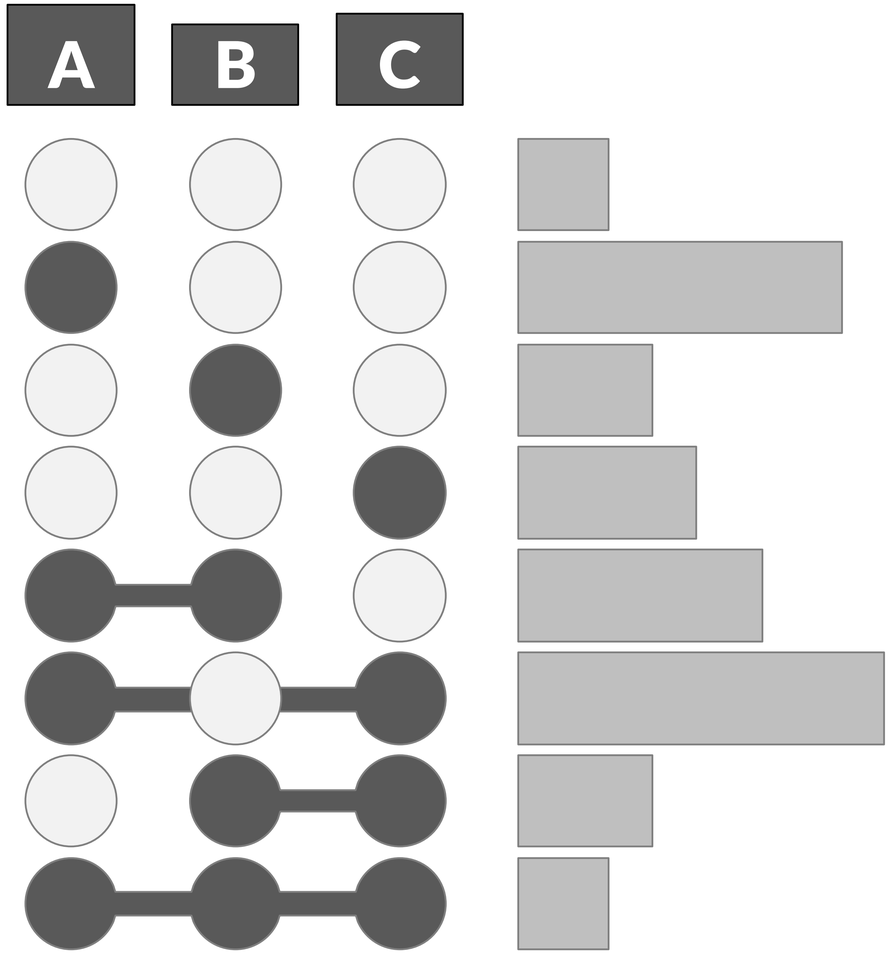




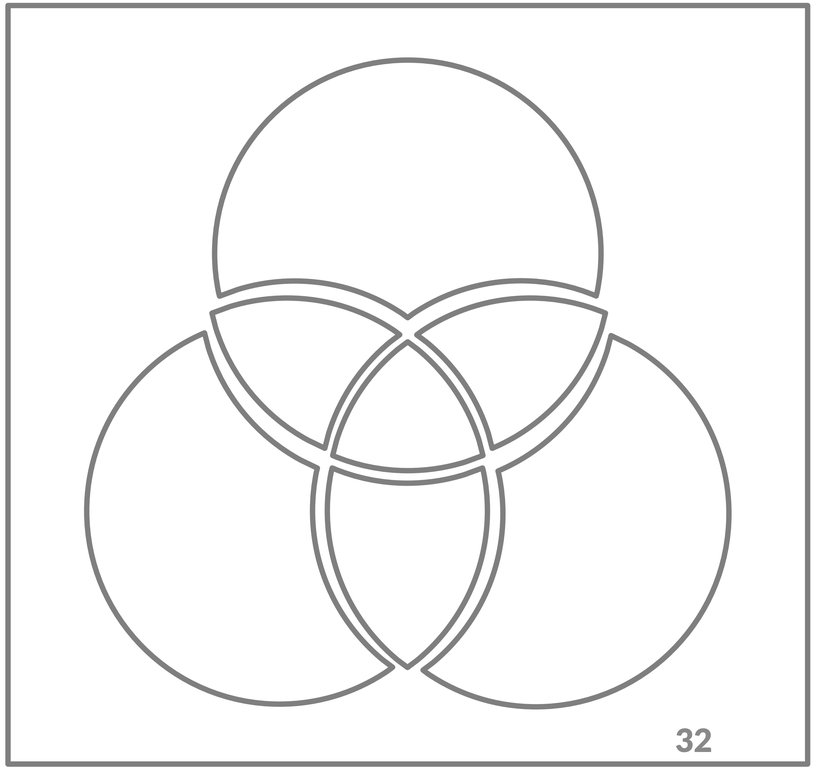


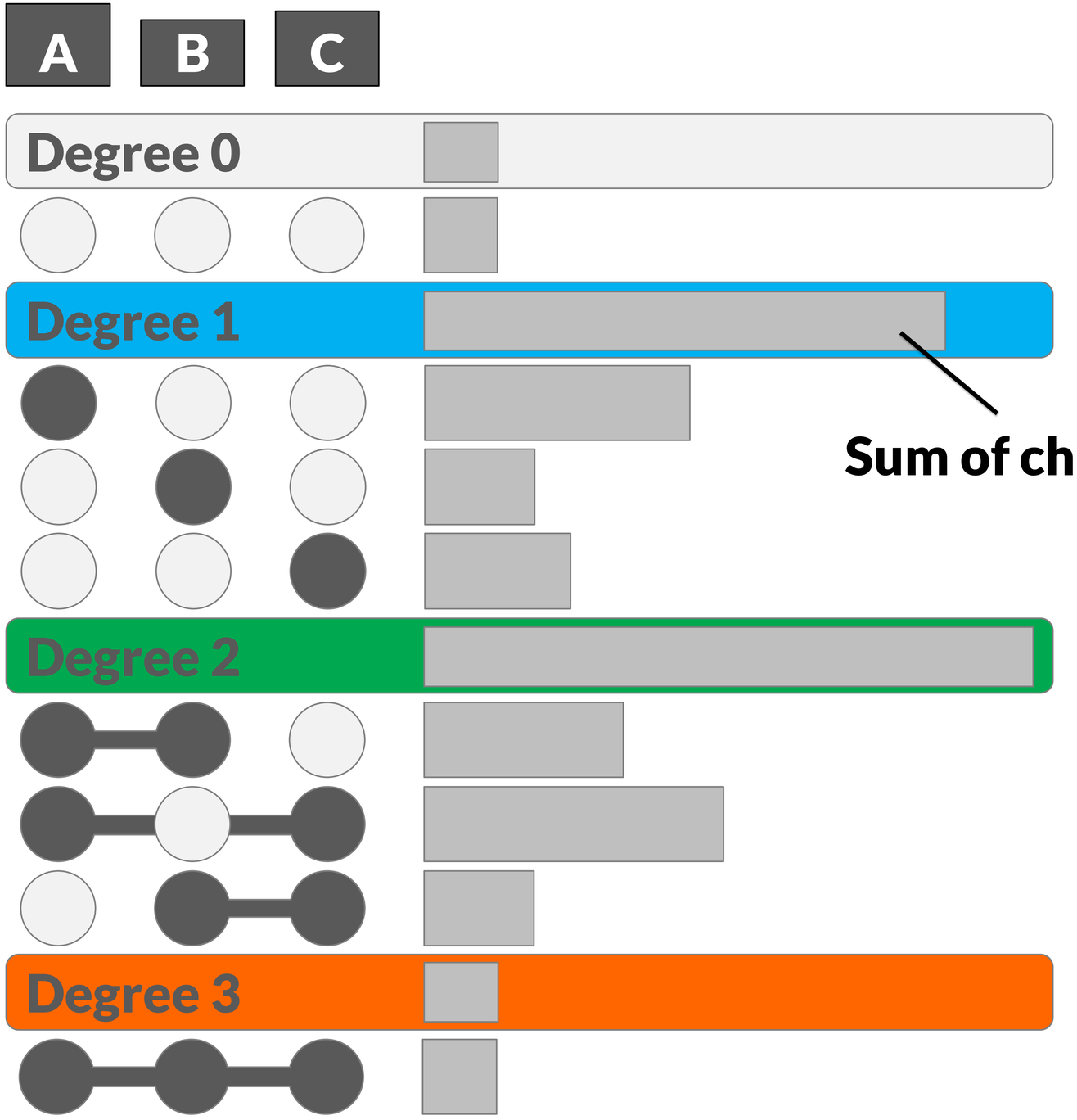

First, aggregate by
Don't Aggregate
Then, aggregate by
Don't Aggregate
Sort by
Degree
Cardinality
Deviation
Aggregates
Collapse All
Expand All
Row Height
Large
Data
Min Degree:
0
Max Degree:
5
Hide Empty Intersections
Dataset Information
Name: Movies Genres
Sets: 17
Attributes: 6
Elements: 3883
Author: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



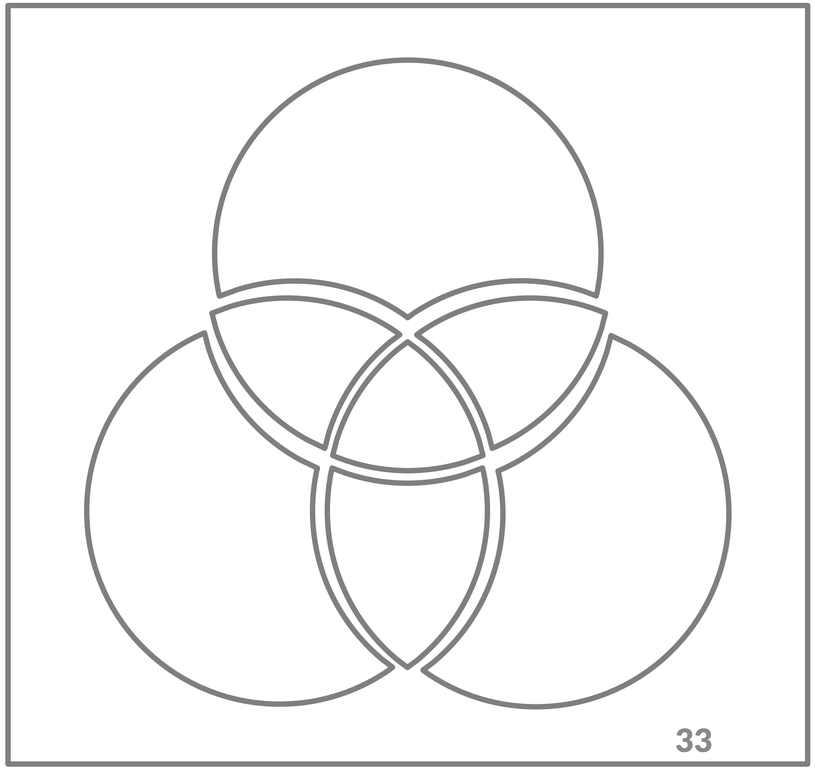
Are many items shared between two sets?
Aggregate By: Degree



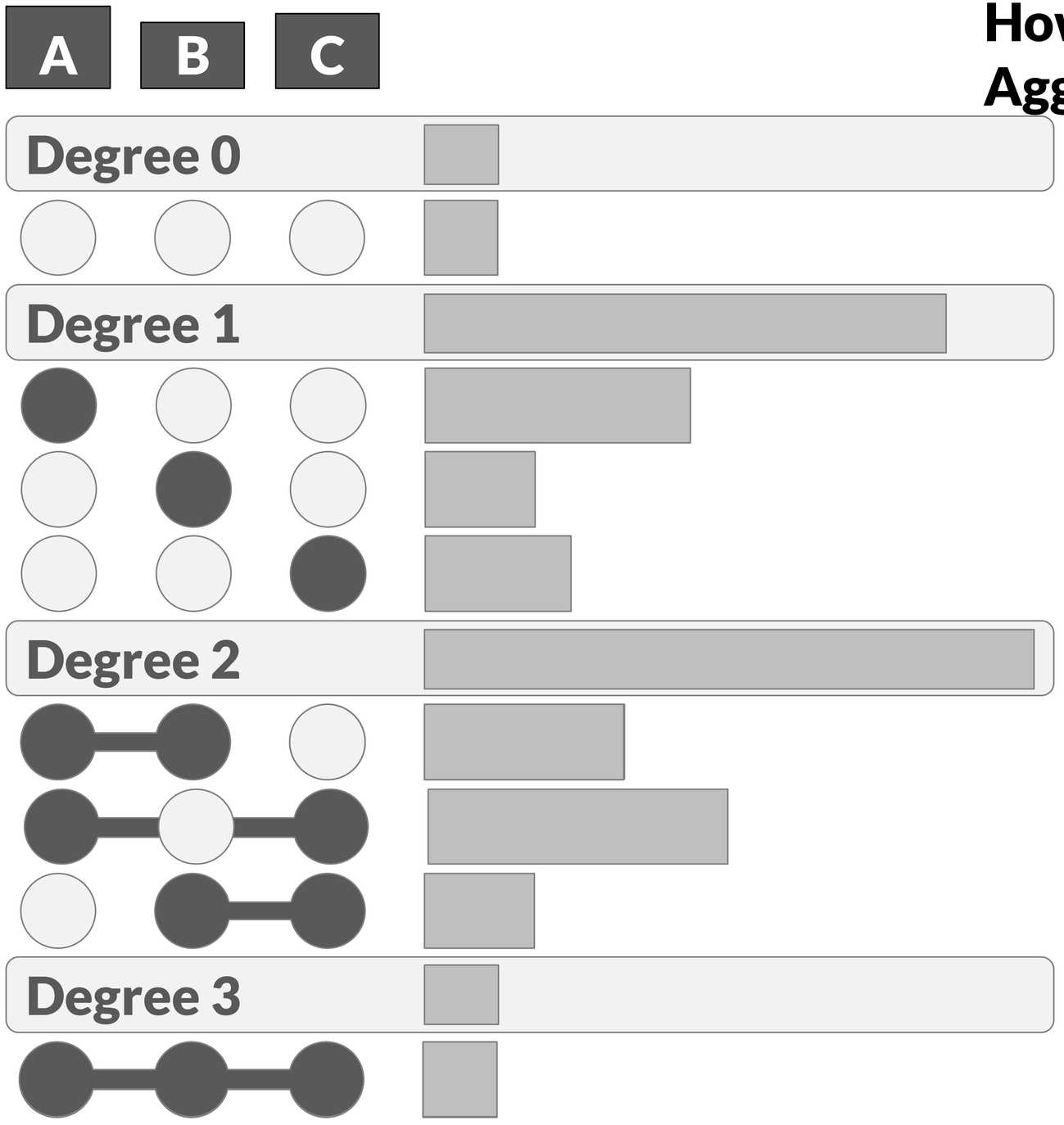


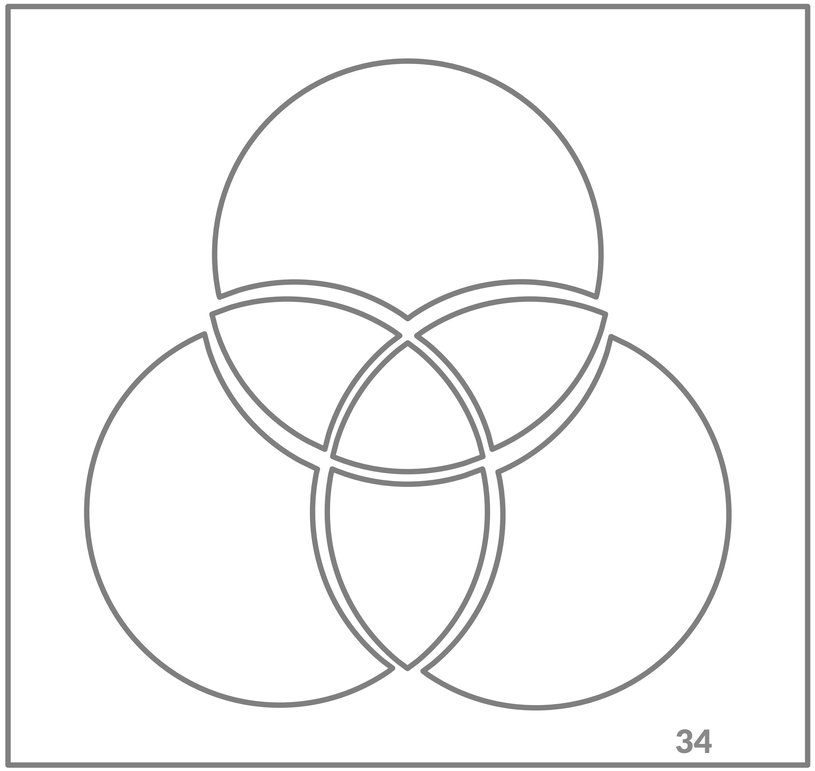
Are many items shared between two sets?
Aggregate By: Degree

Sum of children

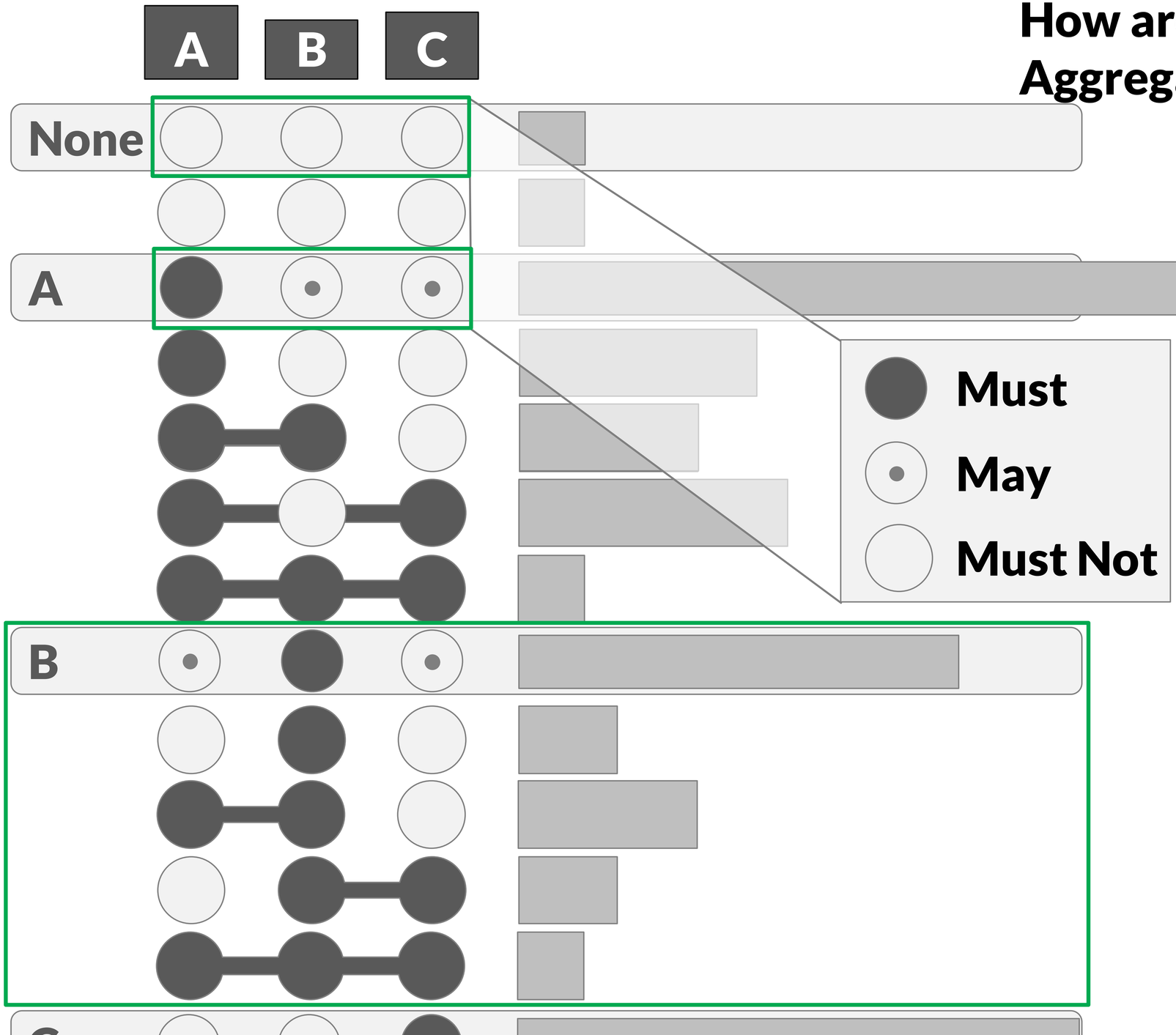


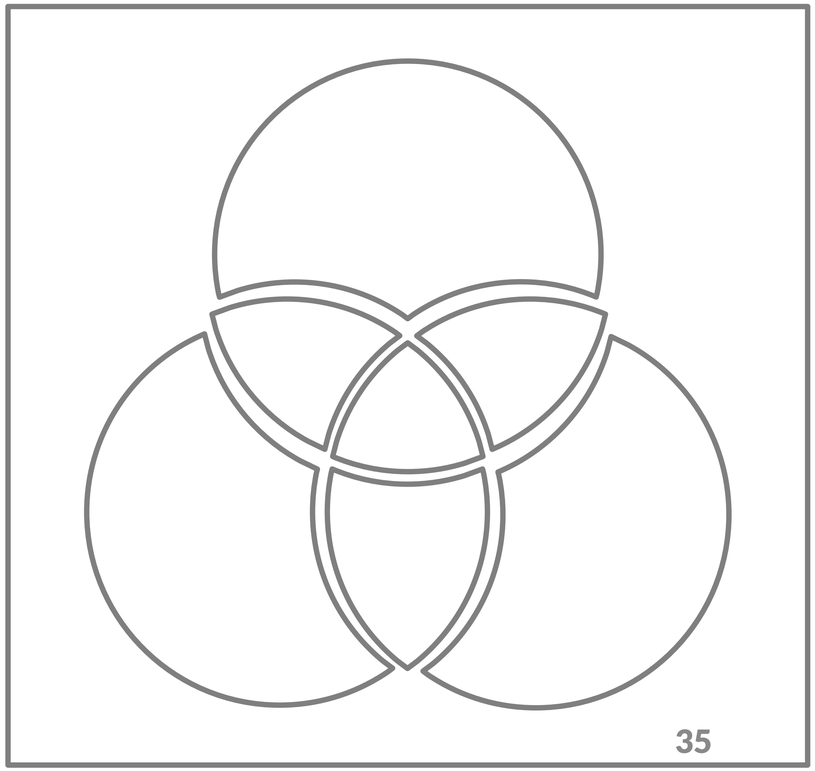
How are the elements of ‘B’ distributed?
Aggregate By: Set



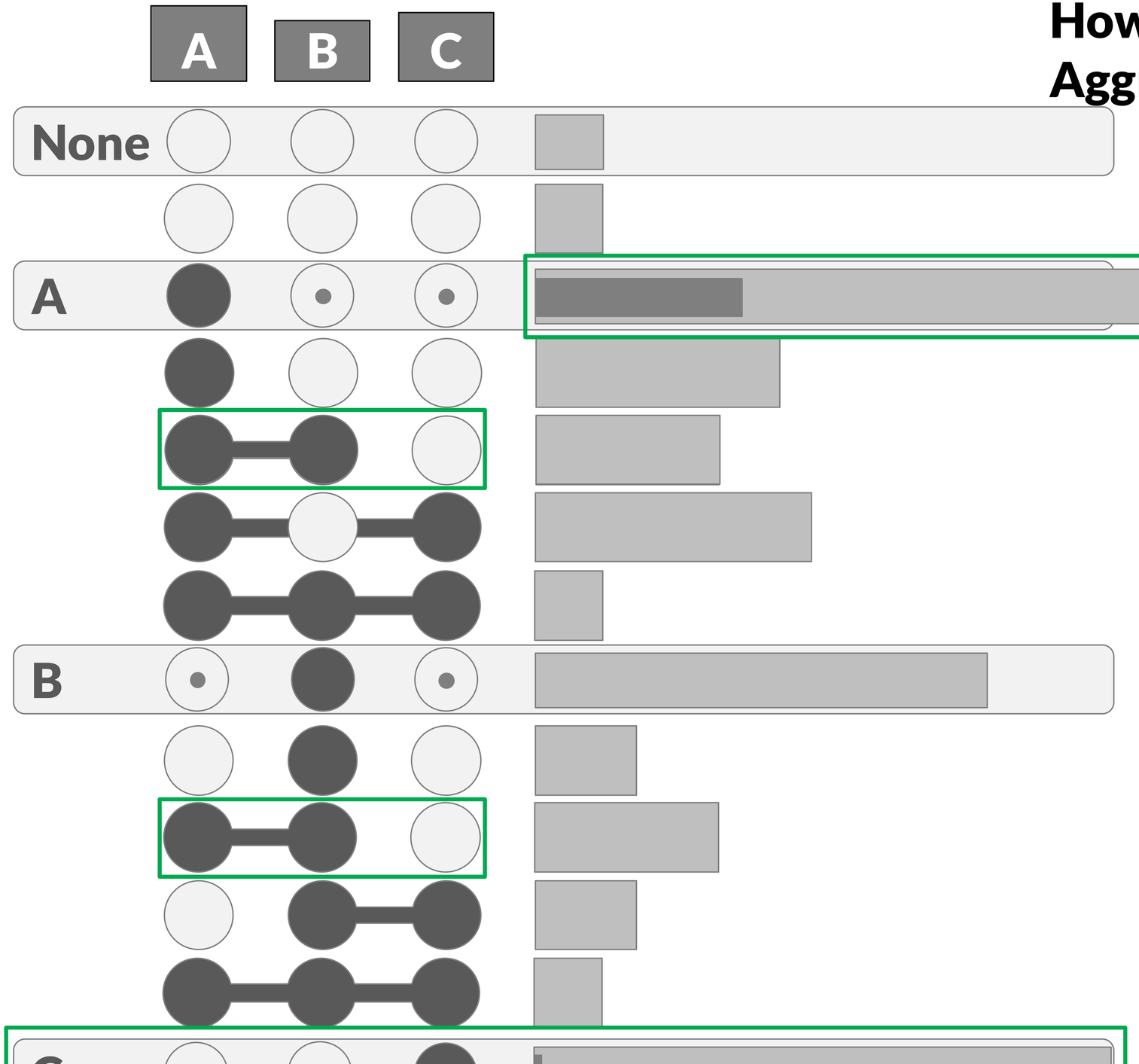


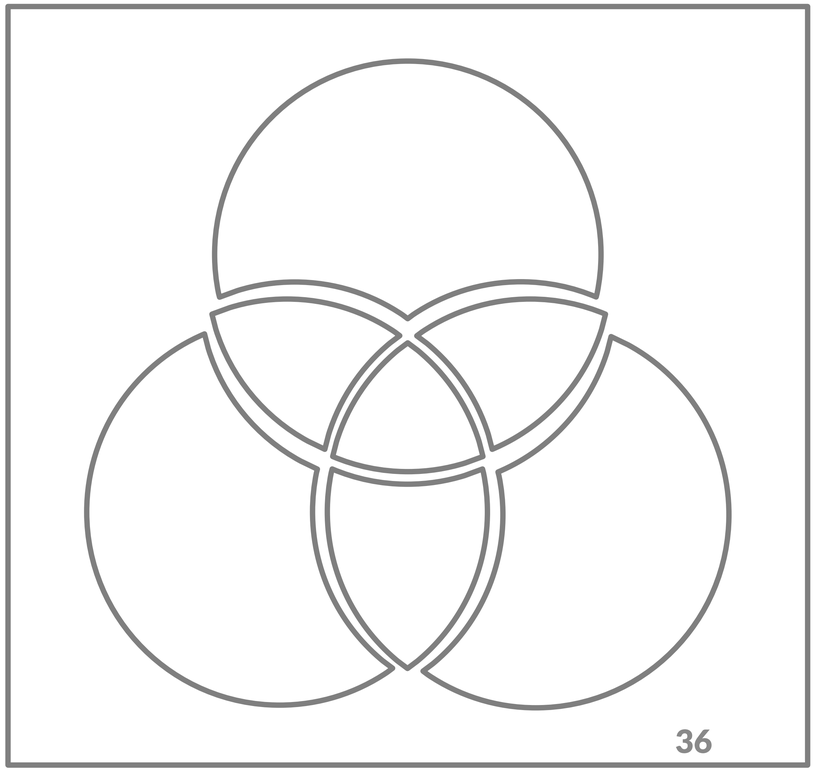
How are the elements of ‘B’ distributed?
Aggregate By: Set

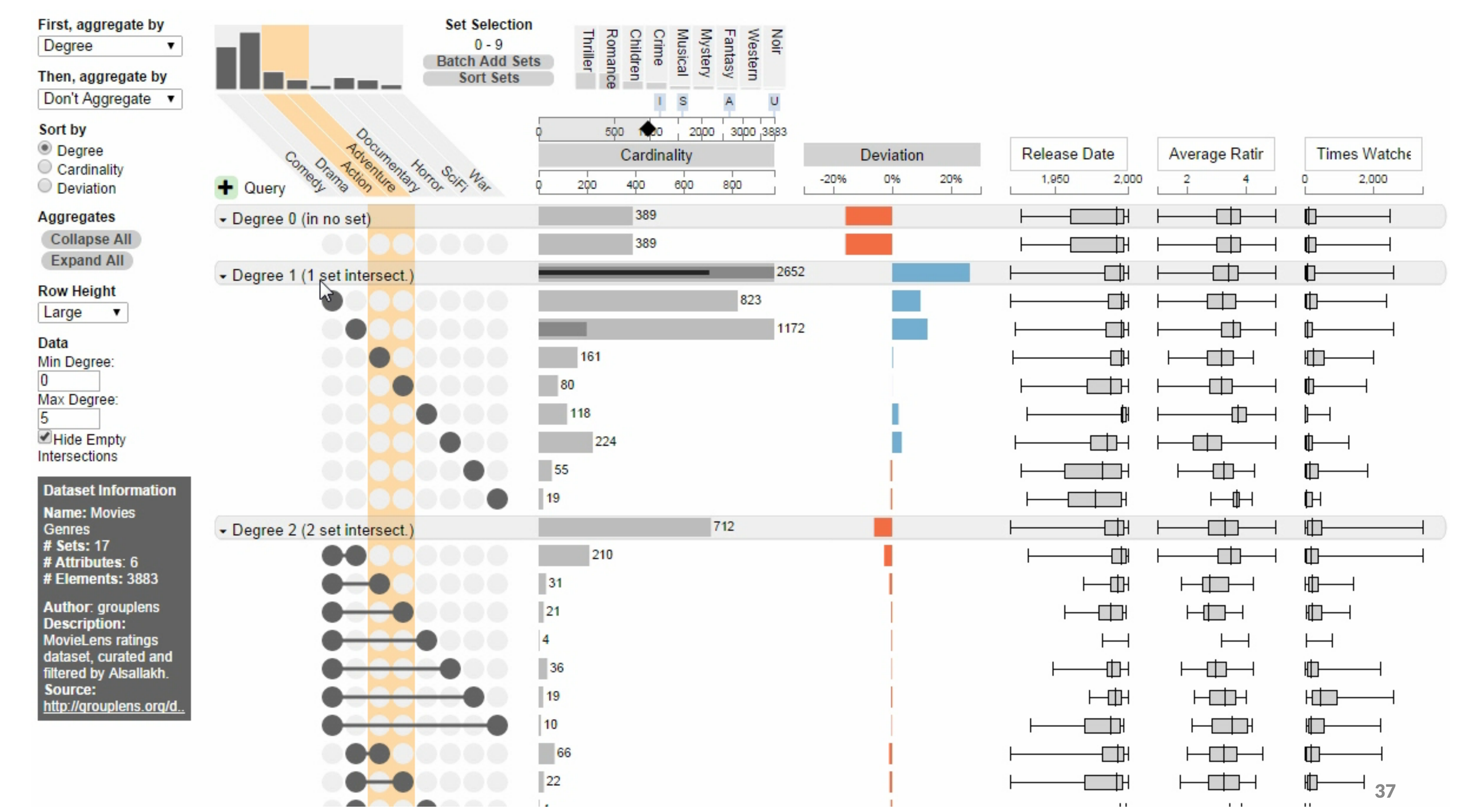




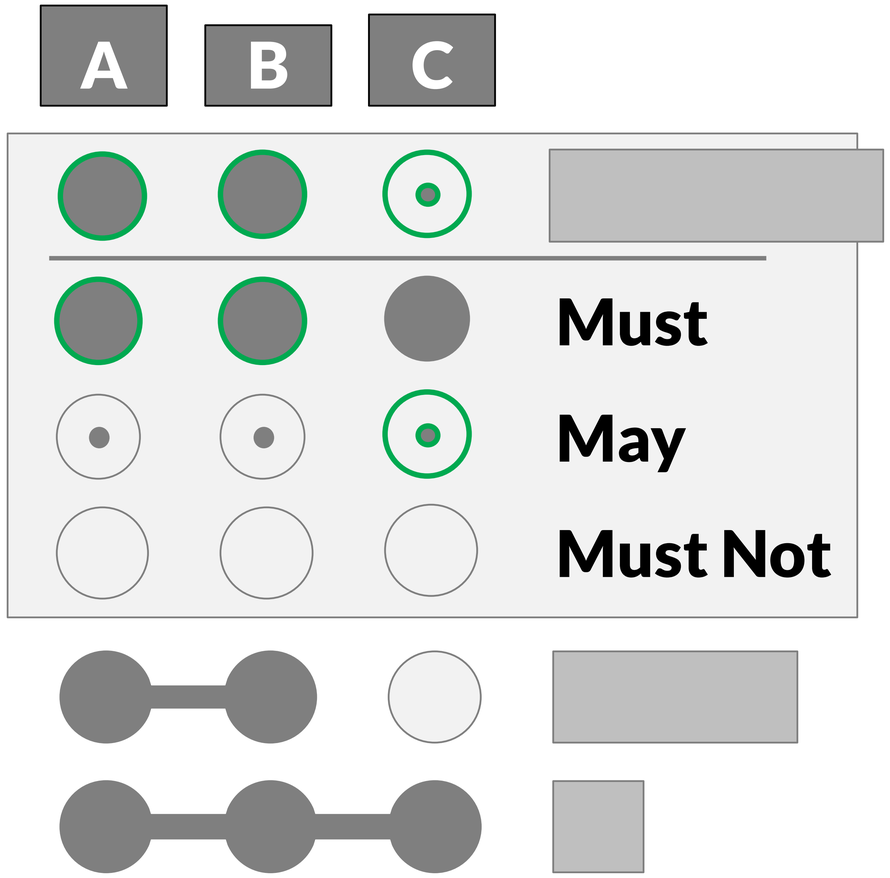
How are the elements of ‘B’ distributed?
Aggregate By: Set

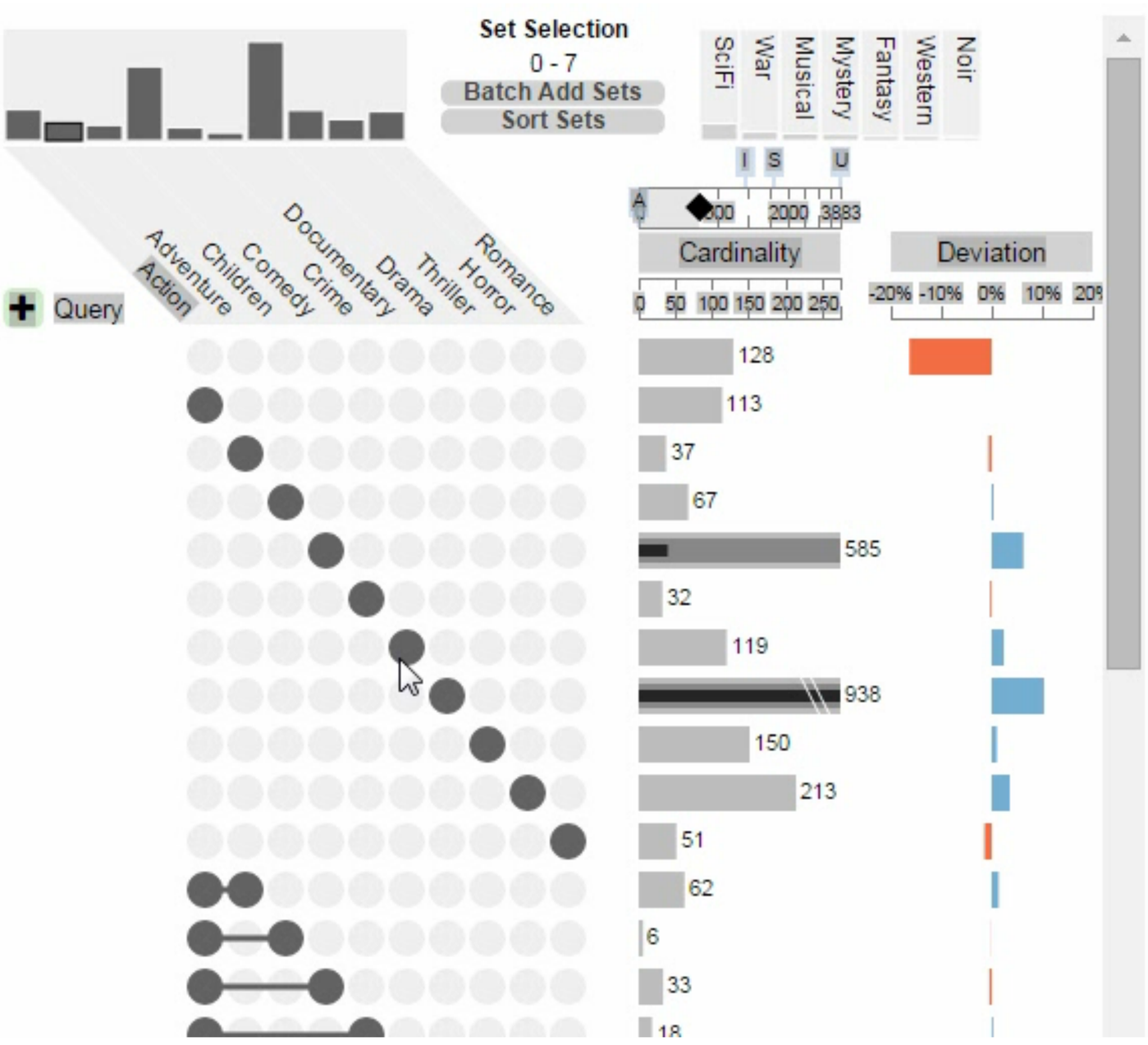





First, aggregate by
Degree
Then, aggregate by
Don't Aggregate
Sort by
Degree
Cardinality
Deviation
Aggregates
Collapse All
Expand All
Row Height
Large
Data
Min Degree:
0
Max Degree:
5
Hide Empty Intersections
Dataset Information
Name: Movies Genres
Sets: 17
Attributes: 6
Elements: 3883
Author: grouplens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

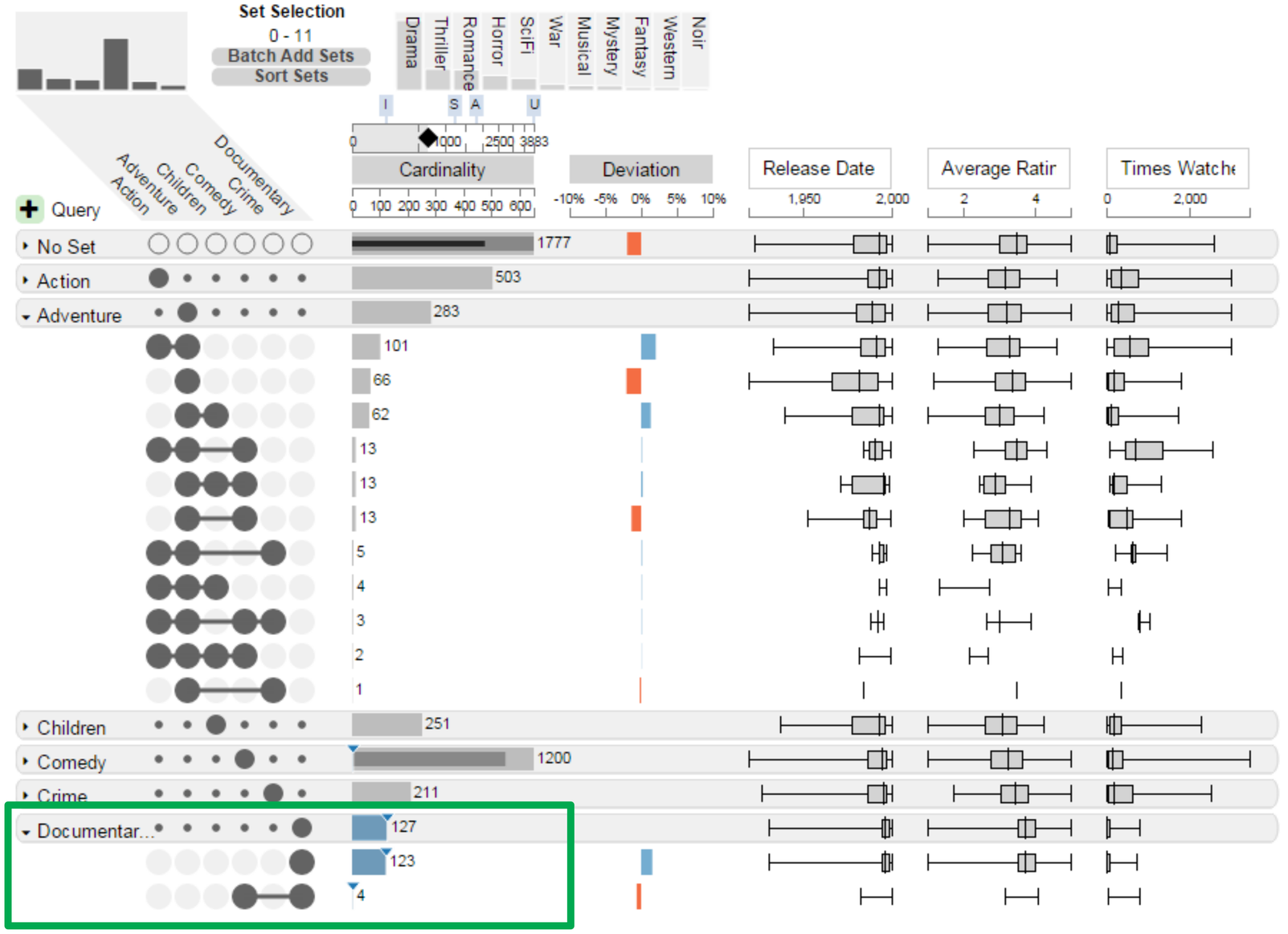
Query Filters

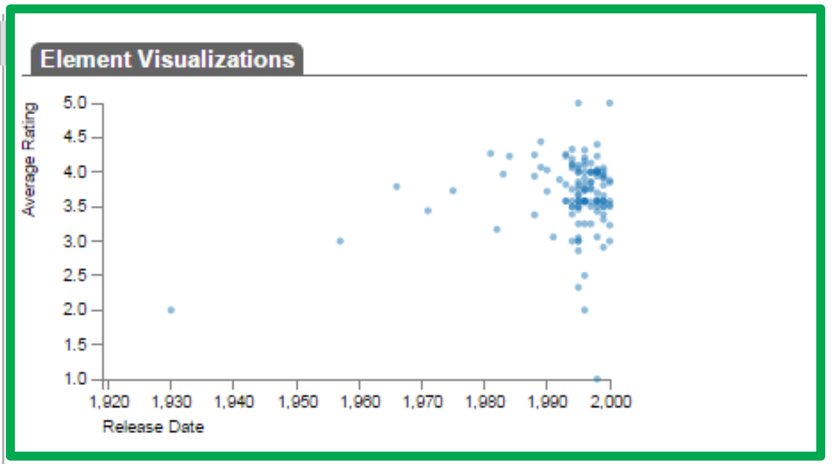
No active query.

Query Results

No active query.

Elements & Attributes

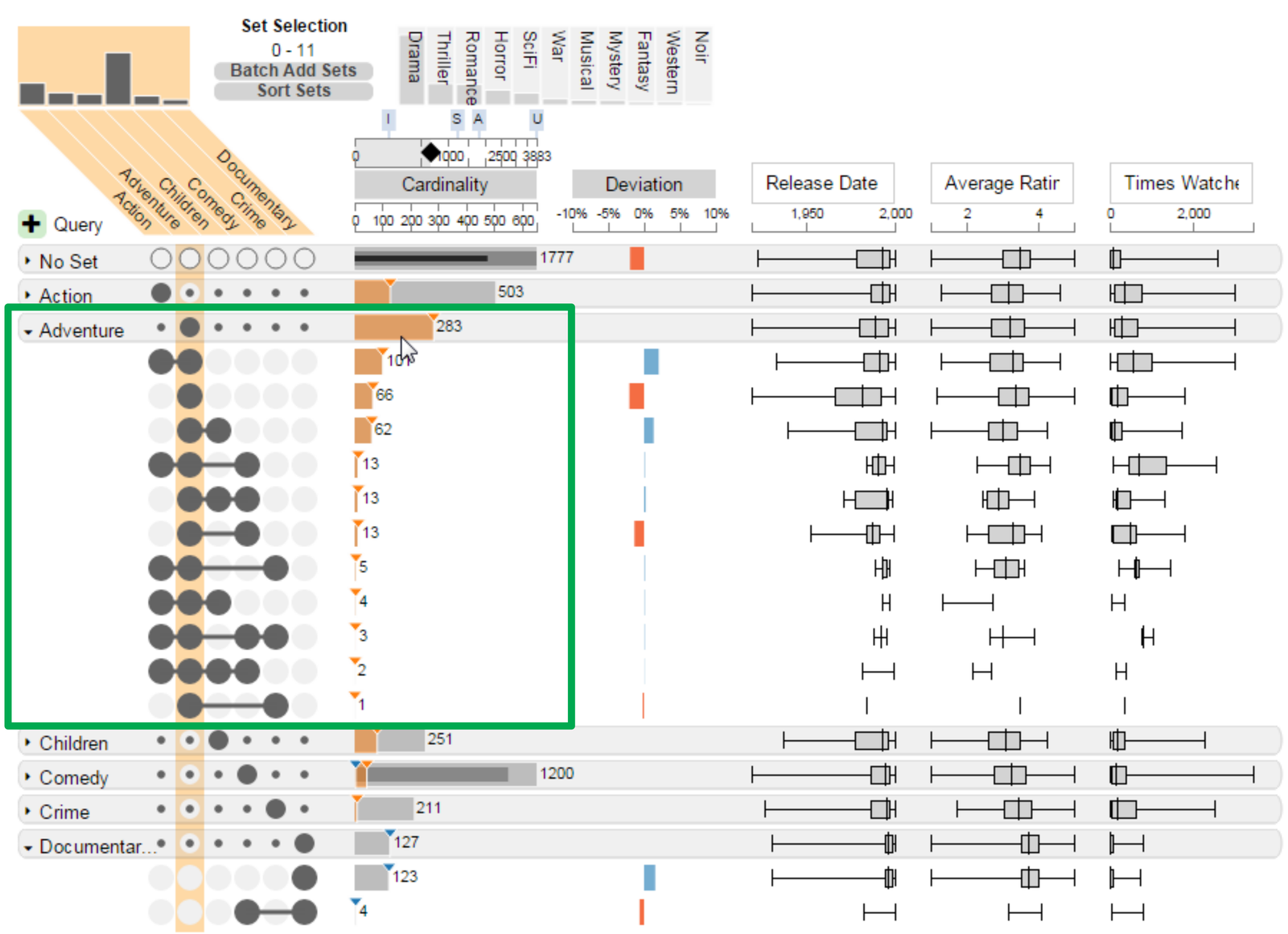




Query Results

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?

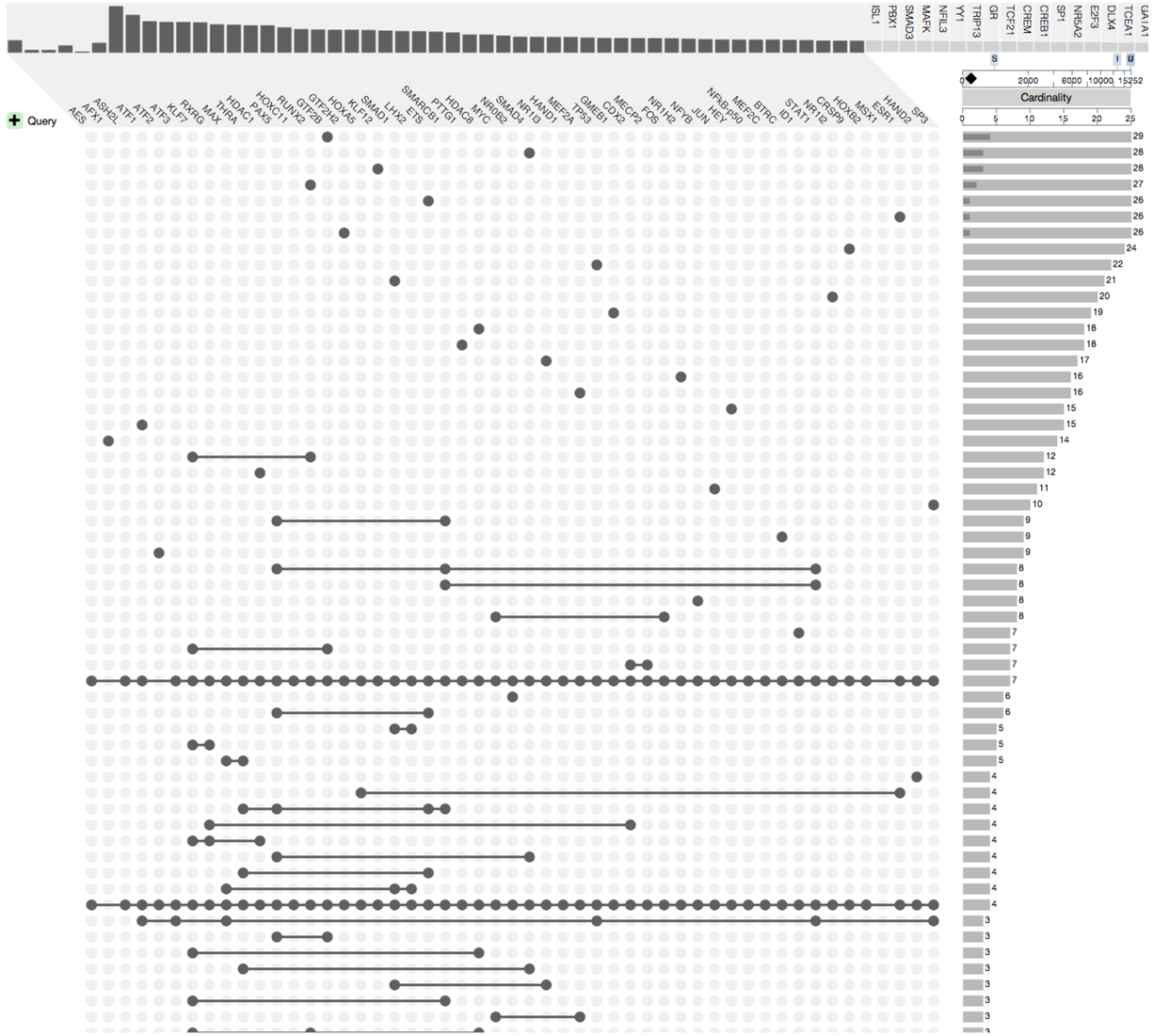
Scalability

Comfortable: ~15 sets

Possible: ~40 sets

Scales with the number of non-empty intersections

- **Most datasets are sparse**



Applications

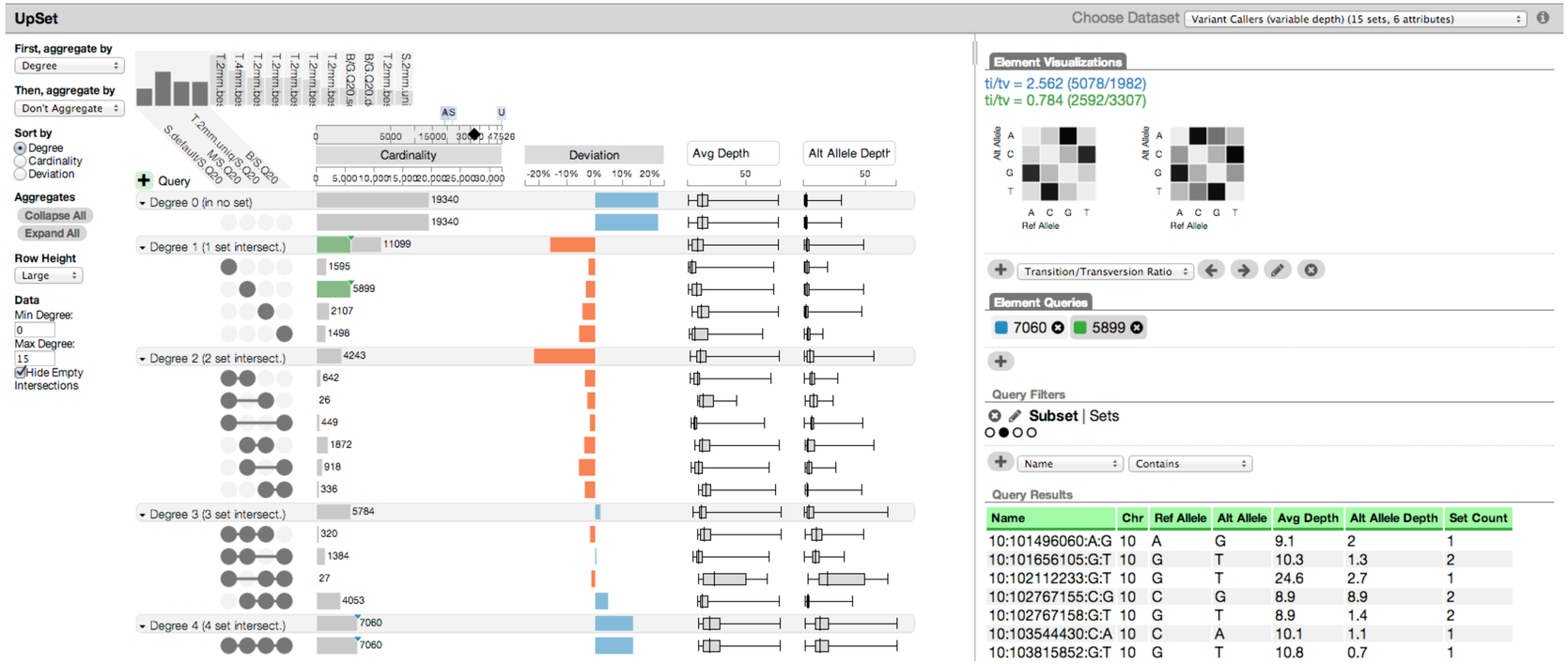
Applications

Genetics

Pharmacology

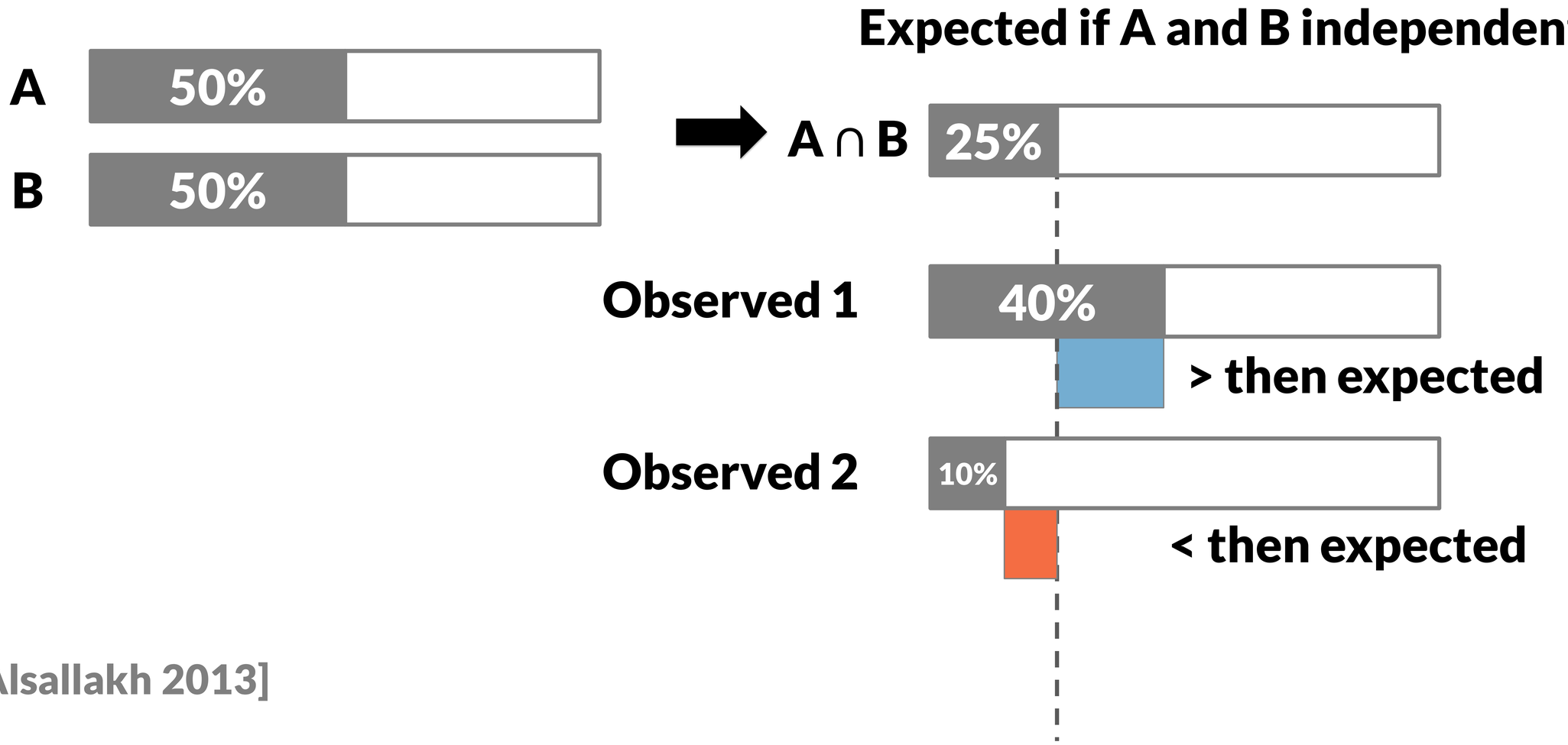
Economics

Social Networks



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

Deviation Measure



[Alsallakh 2013]

http://vcg.github.io/upset

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

April 23, 2015

HARVARD
School of Engineering
and Applied Sciences

UpSet: Visualization of Intersecting Sets

?

Credits:

Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, Anne Mai Wasserman, Hanspeter Pfister