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

THE UNIVERSITY OF UTAH

Driving Scientific Discovery with Interactive Visual Data Analysis

visualization design lab

SCI
www.sci.utah.edu

CURRICULUM VITAE

http://alexander-lex.net

PhD from Graz University of Technology, 2008–2012

PostDoc at Harvard University, 2012–2015

Assistant Professor of Computer Science at SCI Institute, University of Utah, since 2015

NSF CAREER Award, 2018
> $ 2 mio USD funding for my lab from NSF, NIH, DoD, Industry
Currently funding 6 PhD students, 1 SE

Applied for Early Tenure in Fall 2019
Publish in Data Vis, Bioinformatics, and HCI
8 Best Paper Awards or Honorable Mentions

visualization
design lab

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

visualization

The purpose of computing is insight,

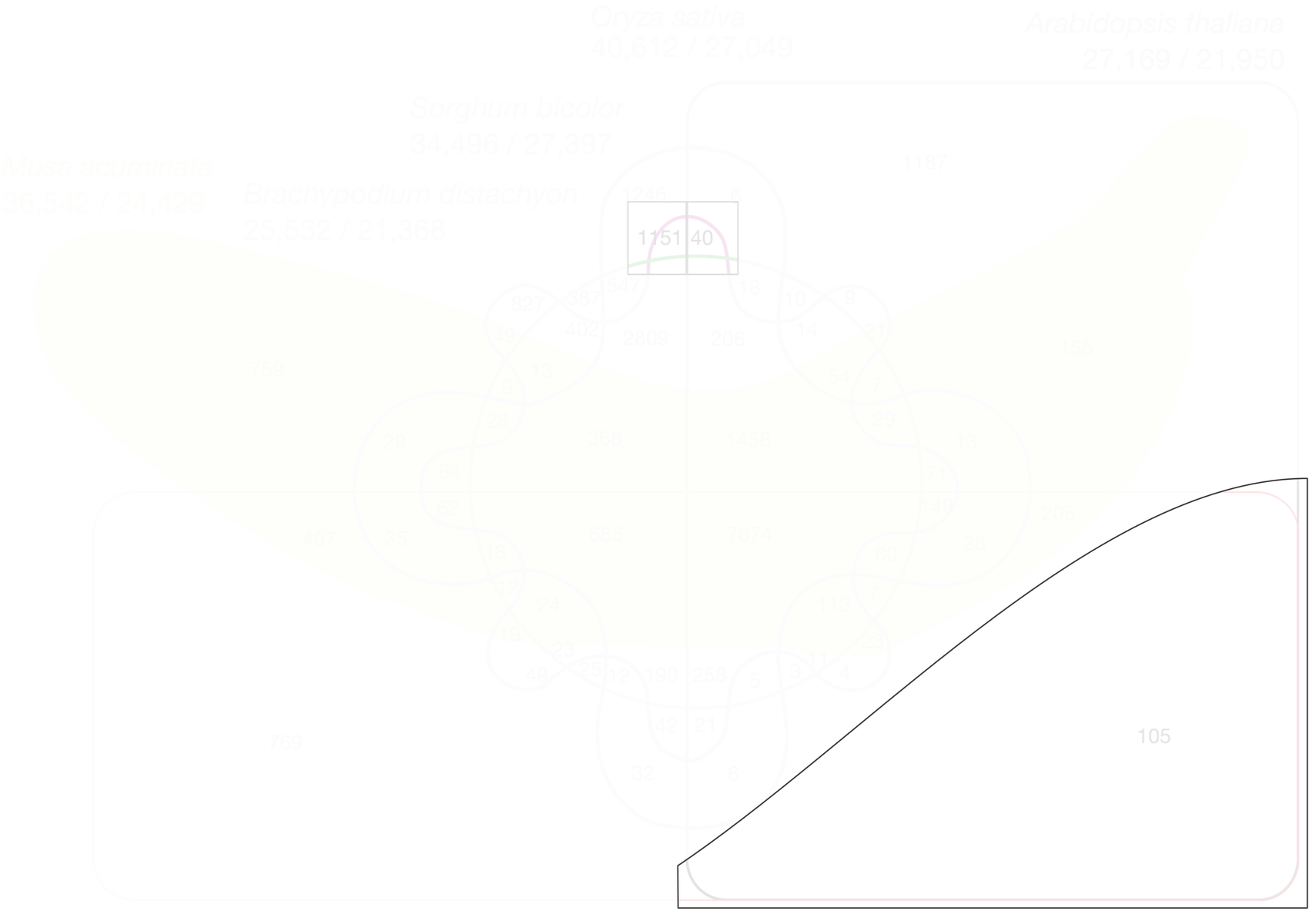
not numbers.

pictures

[Card, Mackinlay, Shneiderman]

[Richard Wesley Hamming]

Banana	*M. acuminata*
Date	*P. dactylifera*
Cress	*Arabidopsis thaliana*
Rice	*Oryza sativa*
Sorghum	*Sorghum bicolor*
Brome	*Brachypodium distachyon*



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

GOOD DATA VISUALIZATION

... makes data accessible

... combines strengths of humans and computers

... enables insight

... communicates

CAN'T WE JUST 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
							12.5
							5.56
							7.91
							6.89

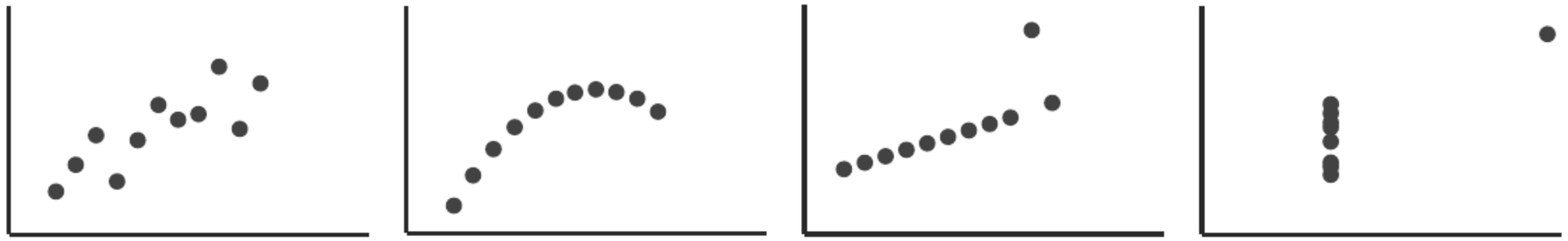
Mean x: 9 y: 7.50

Variance x: 11 y: 4.122

Correlation x - y: 0.816

Linear regression: y = 3.00 + 0.500x

ANSCOMBE'S QUARTETT

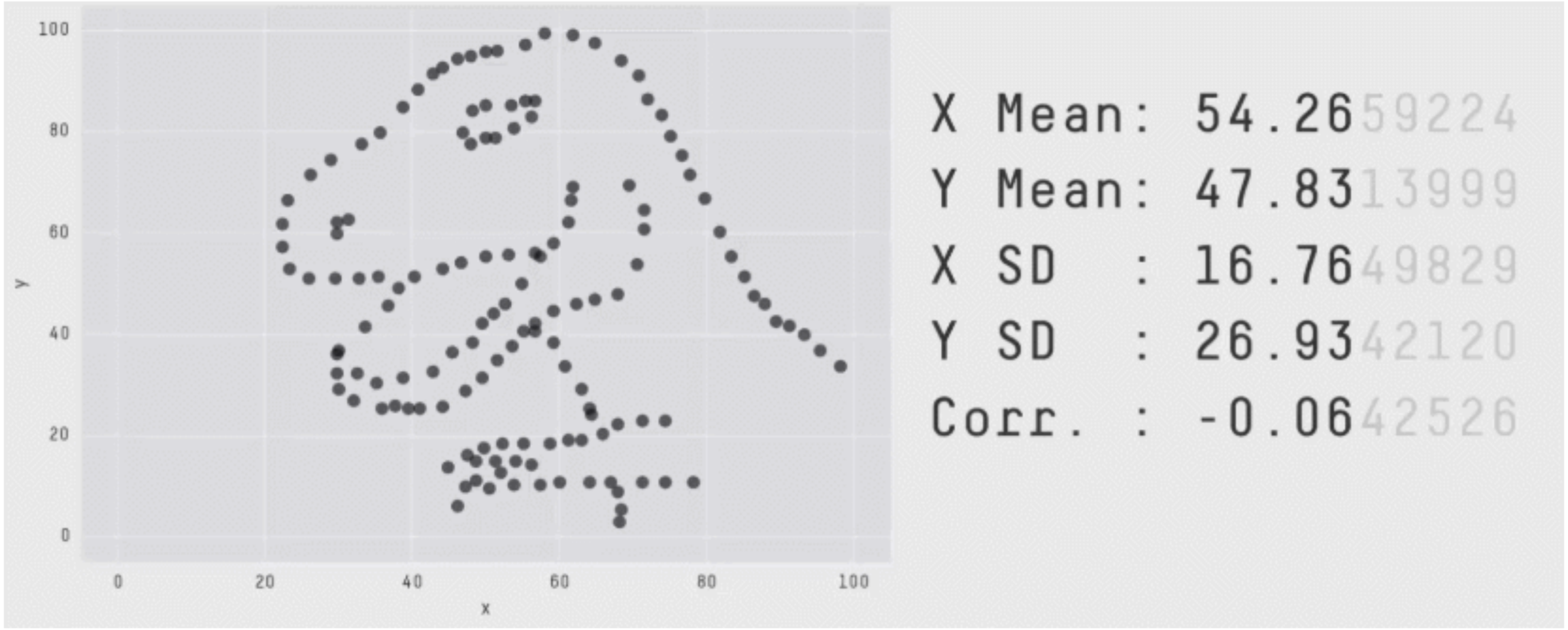


Mean x: 9 y: 7.50

Variance x: 11 y: 4.122

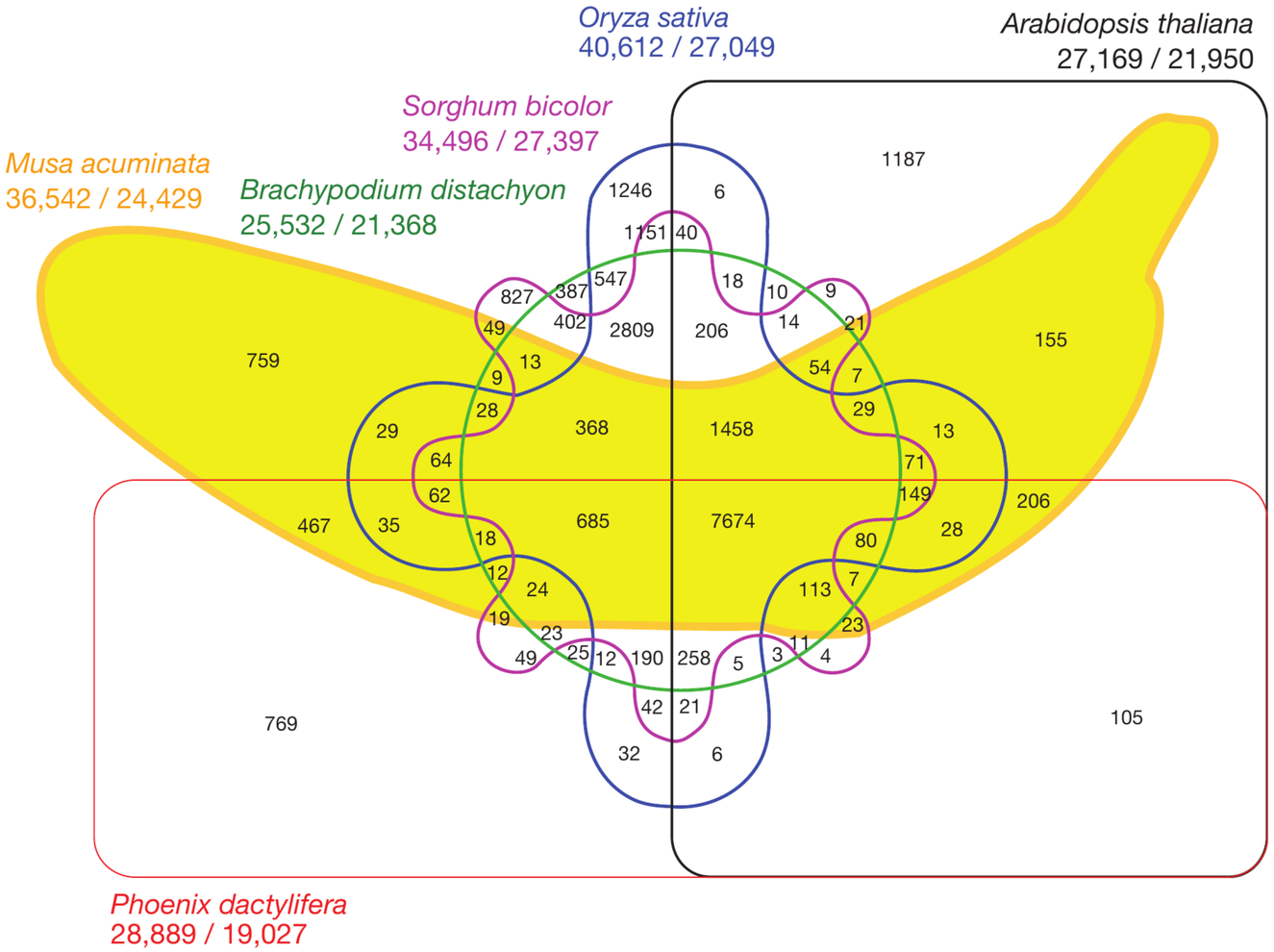
Correlation x – y: 0.816

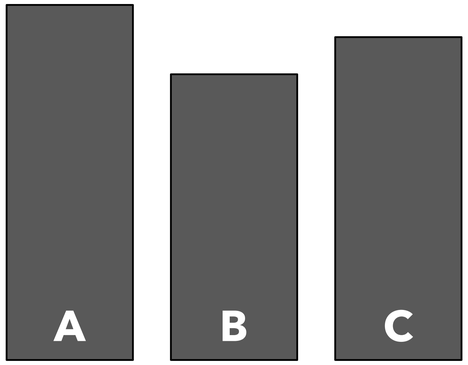
Linear regression: y = 3.00 + 0.500x



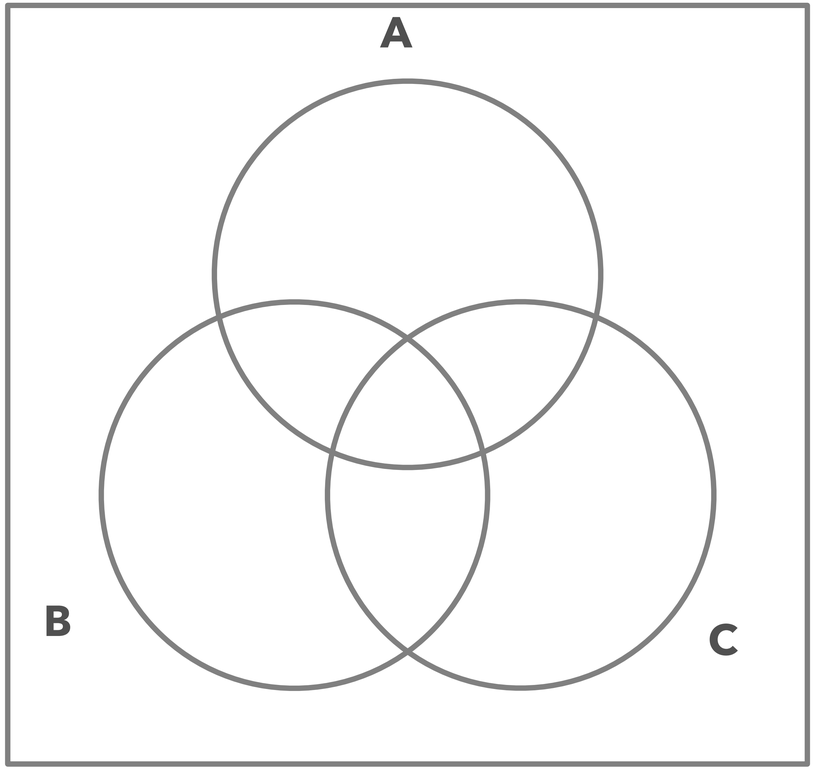
Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing, CHI 2017, Justin Matejka, George Fitzmaurice

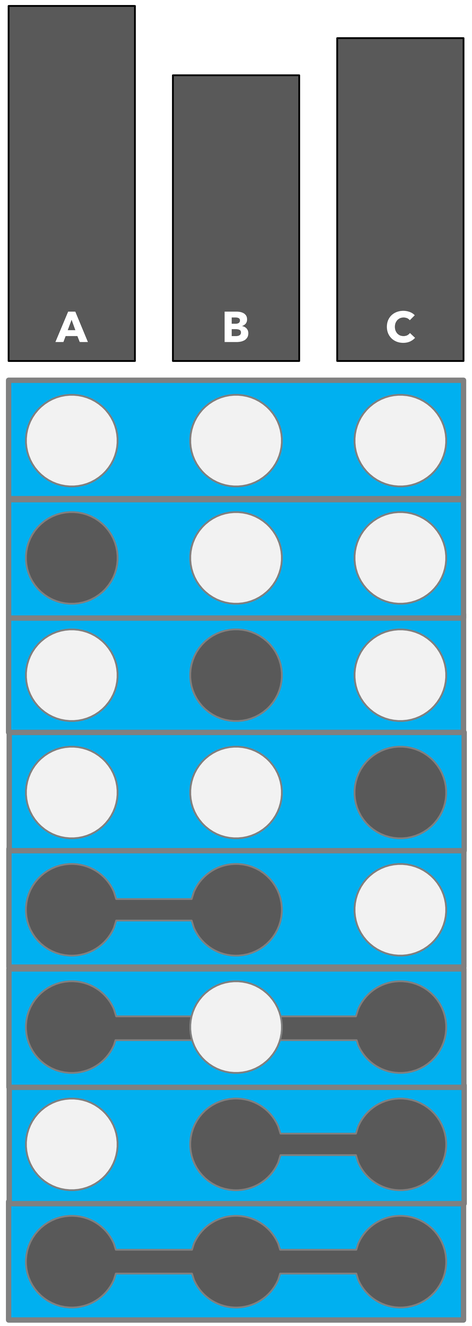
SO CAN WE DO BETTER?



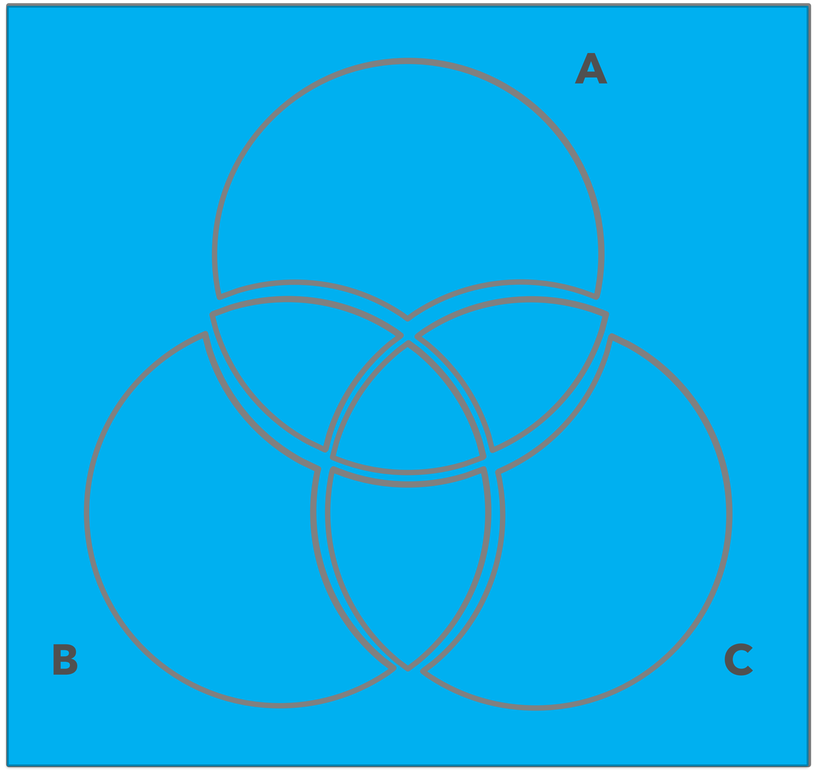


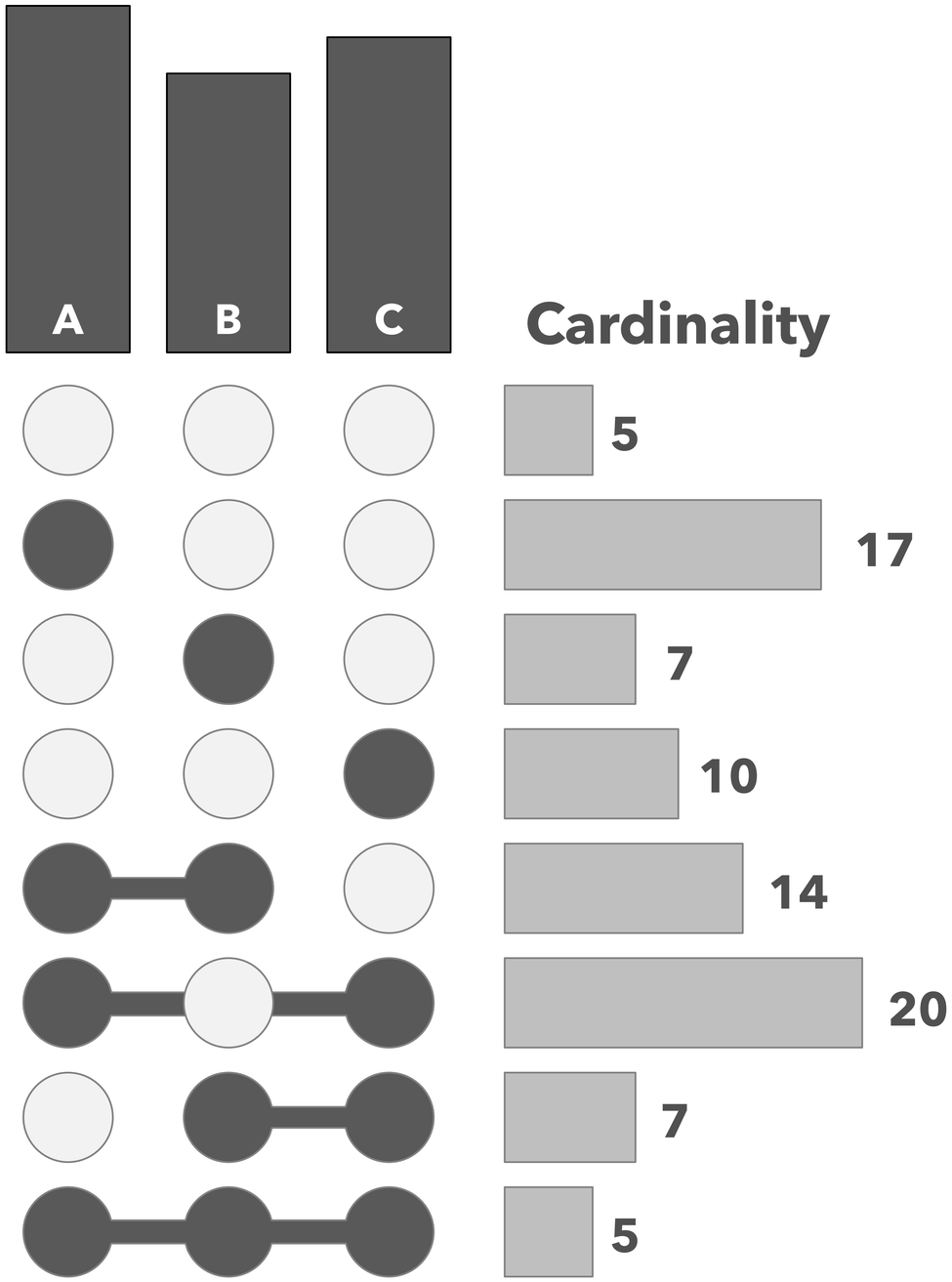
Universal Set

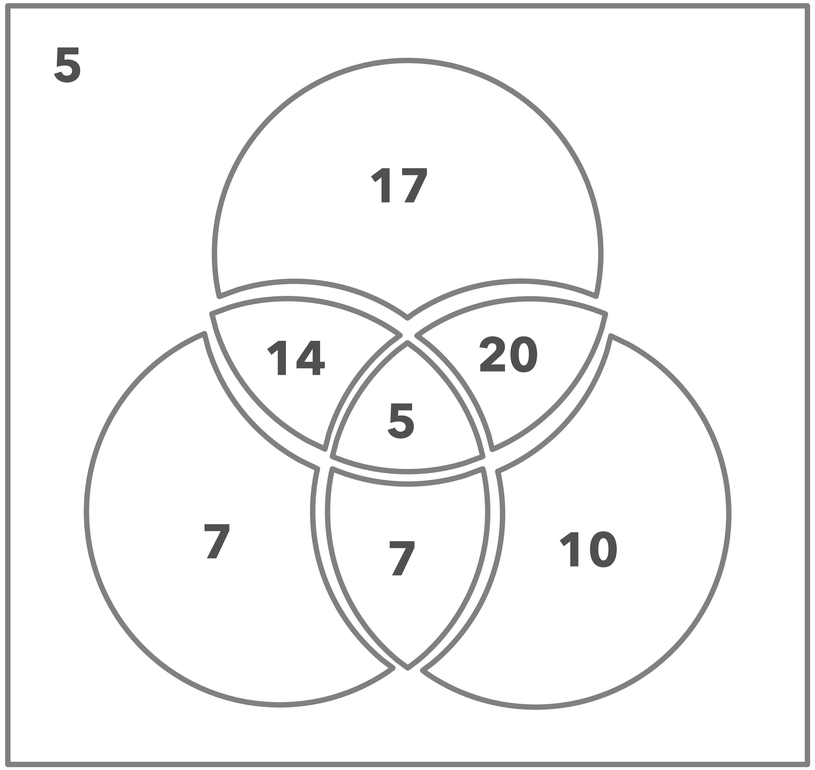




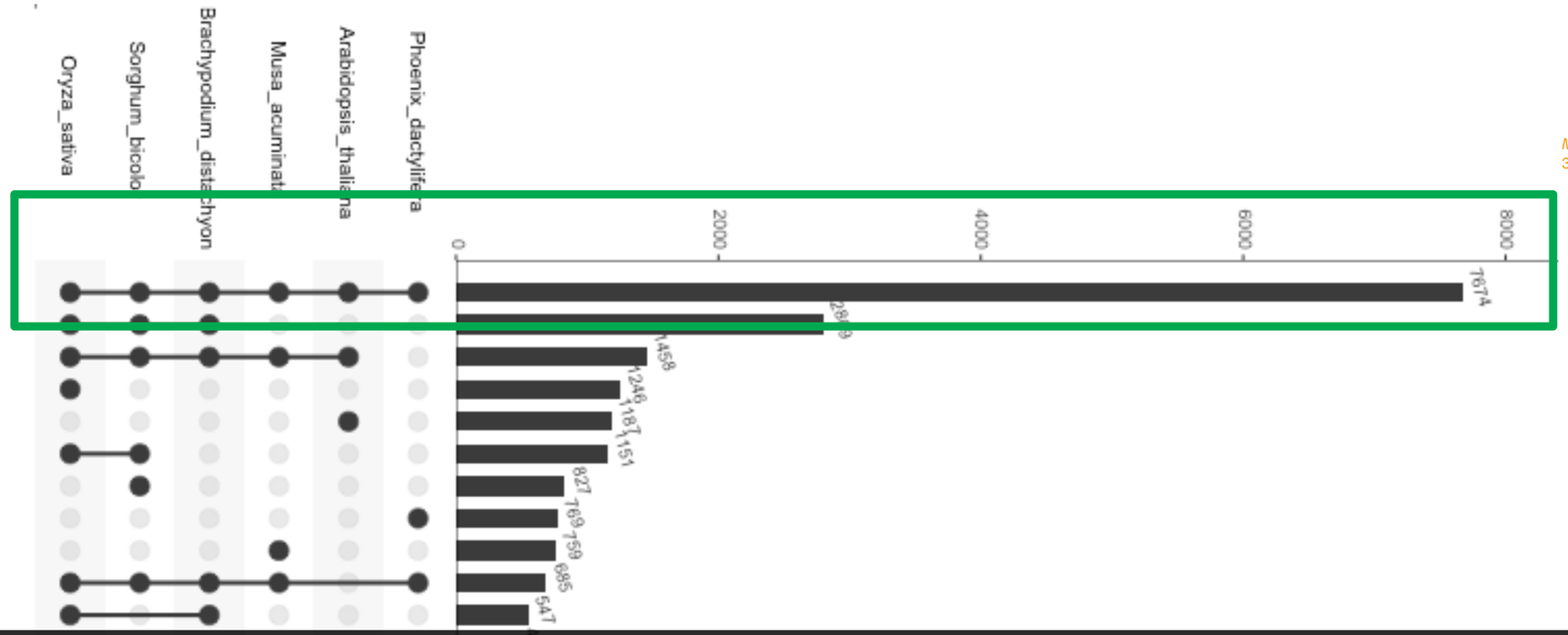
Universal Set

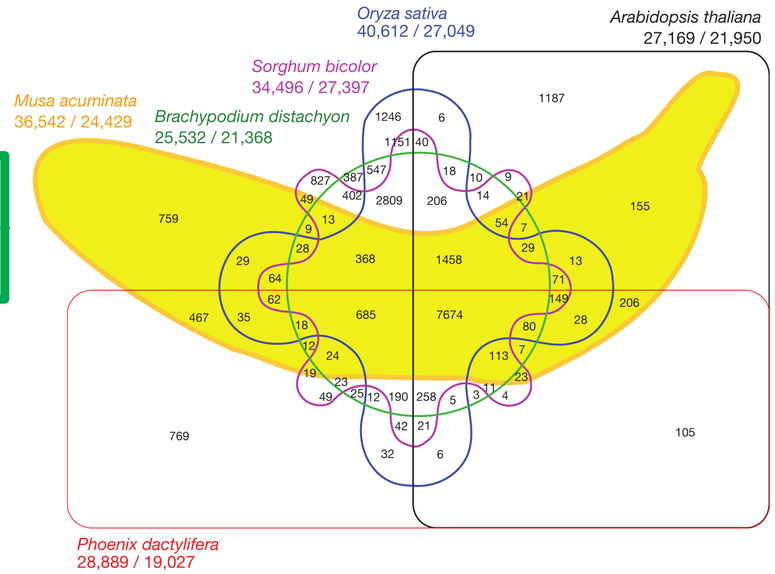



A
B
C
Cardinality
5
17
7
10
14
20
7
5


5
17
14
20
5
7
7
10

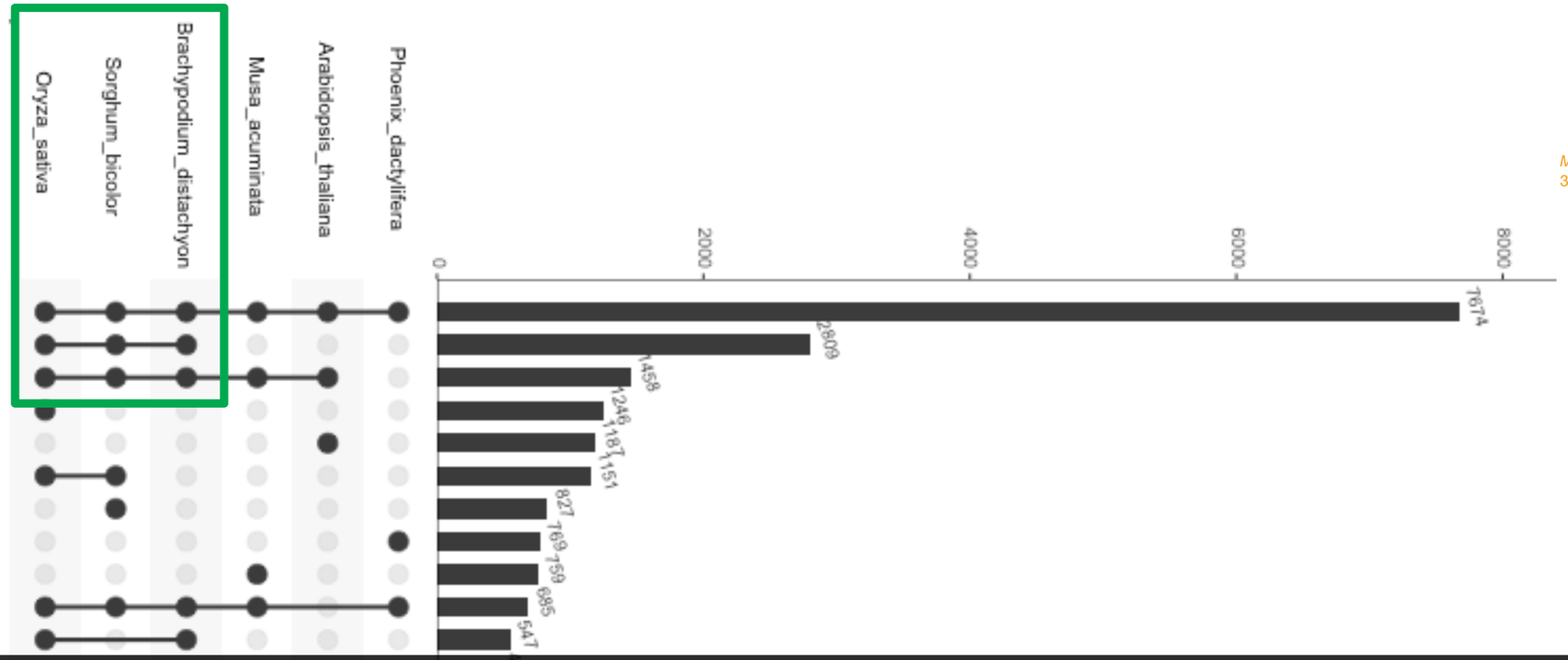
THE BANANA CHART REDESIGNED: UPSET

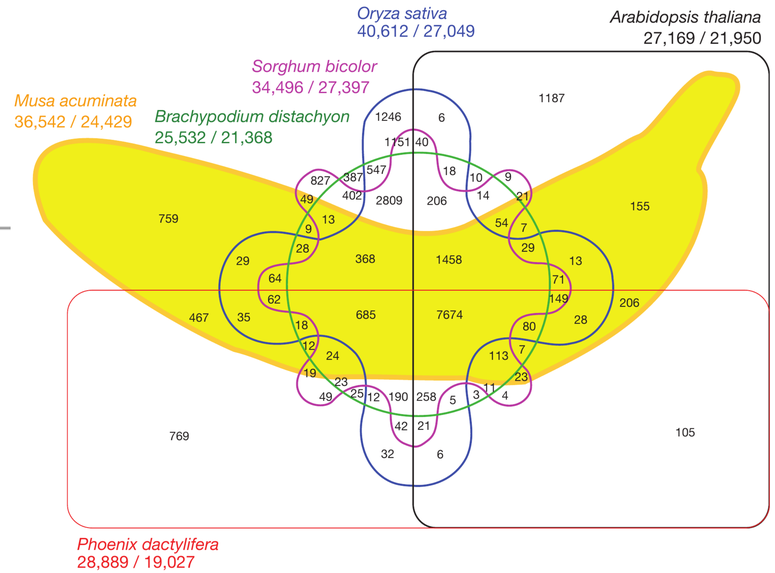




Largest Intersection Includes All Sets

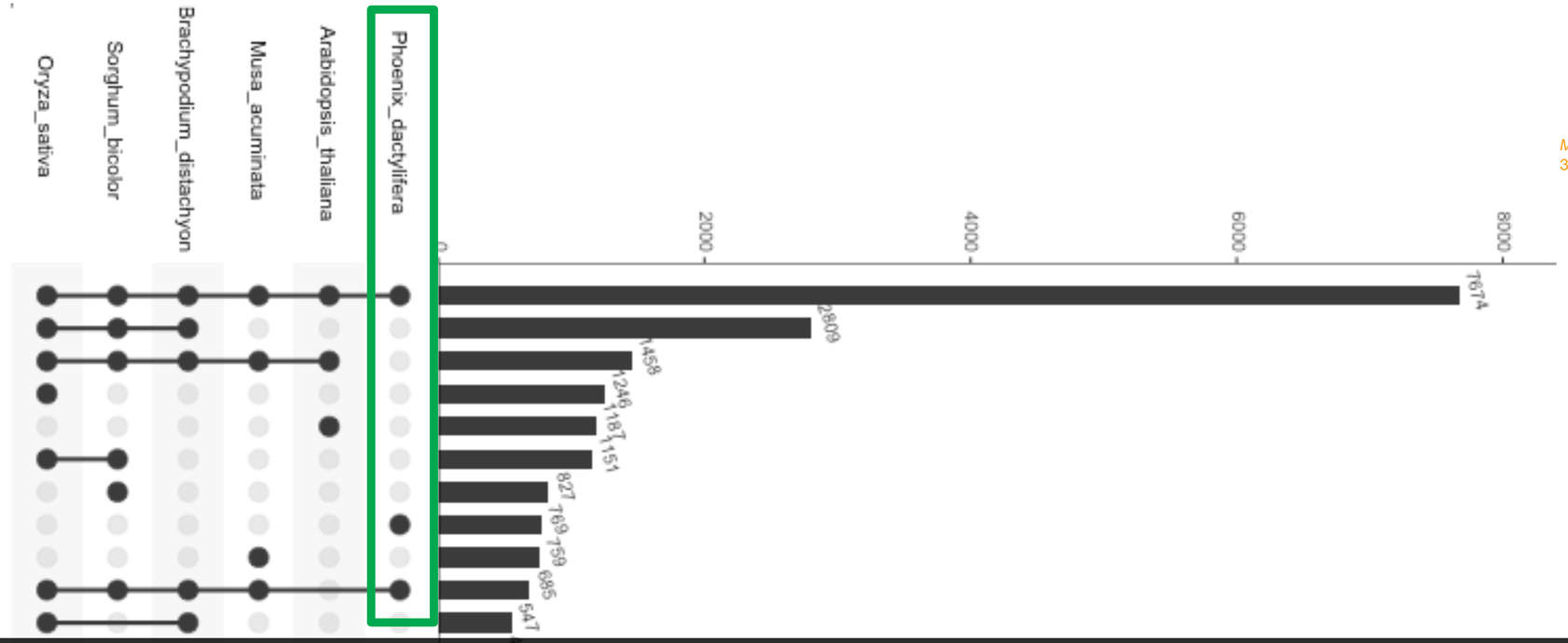
THE BANANA CHART REDESIGNED: UPSET

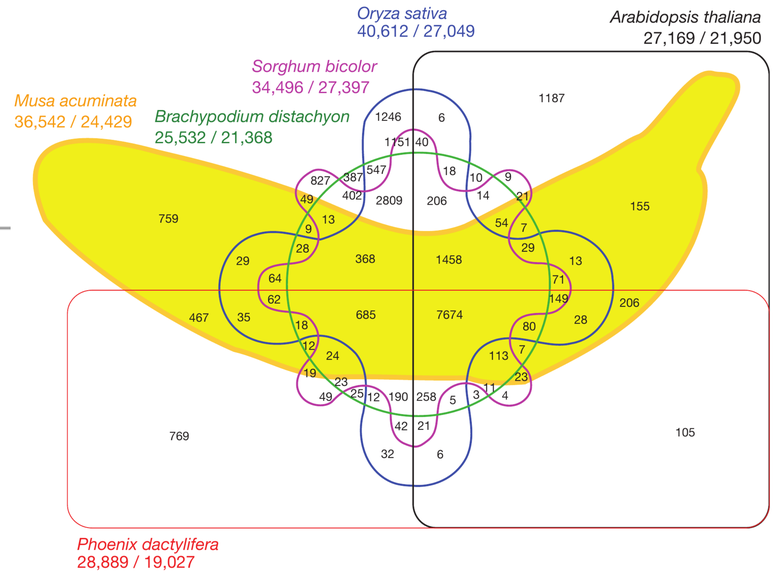




Three Leftmost Species Are Most Similar

THE BANANA CHART REDESIGNED: UPSET





Rightmost species is most different

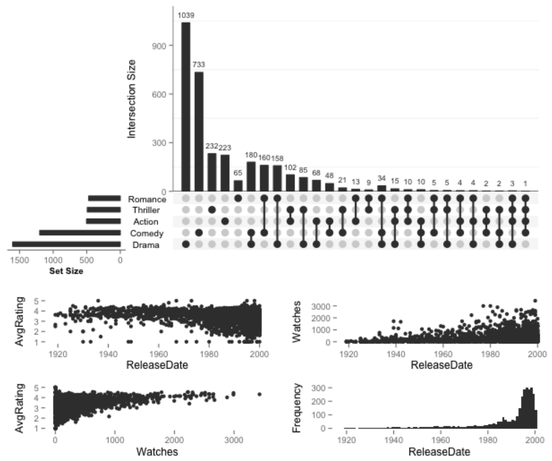
http://vcg.github.io/upset/

UPSET

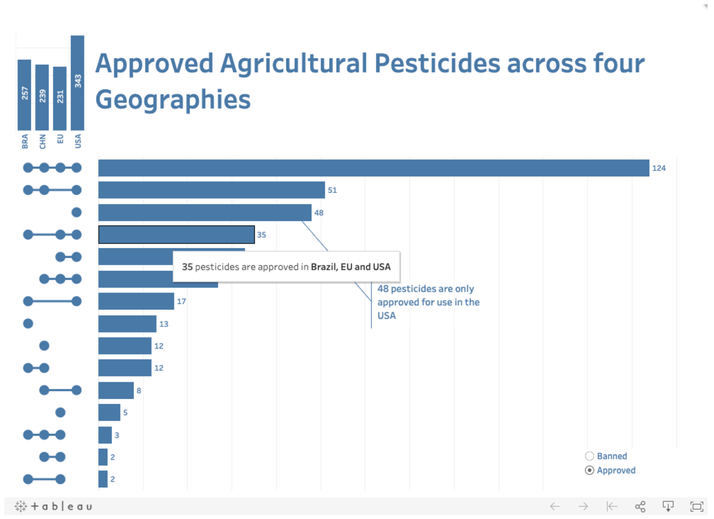
The canonical way to show set data with > 3 sets

Second-most cited VIS paper of the last decade

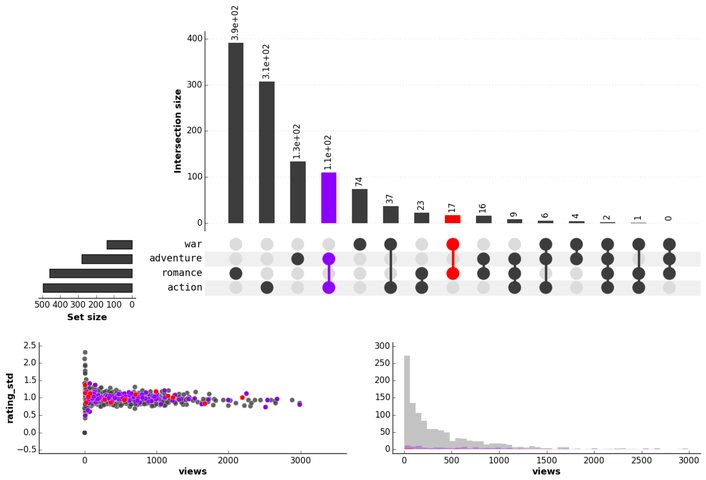
Multiple implementations in various languages



R



Tableau



Python

DATA SCIENCE

WHAT IS DATA SCIENCE?

A data scientist is a statistician who lives in San Fransisco.

Data science is statistics on a Mac.

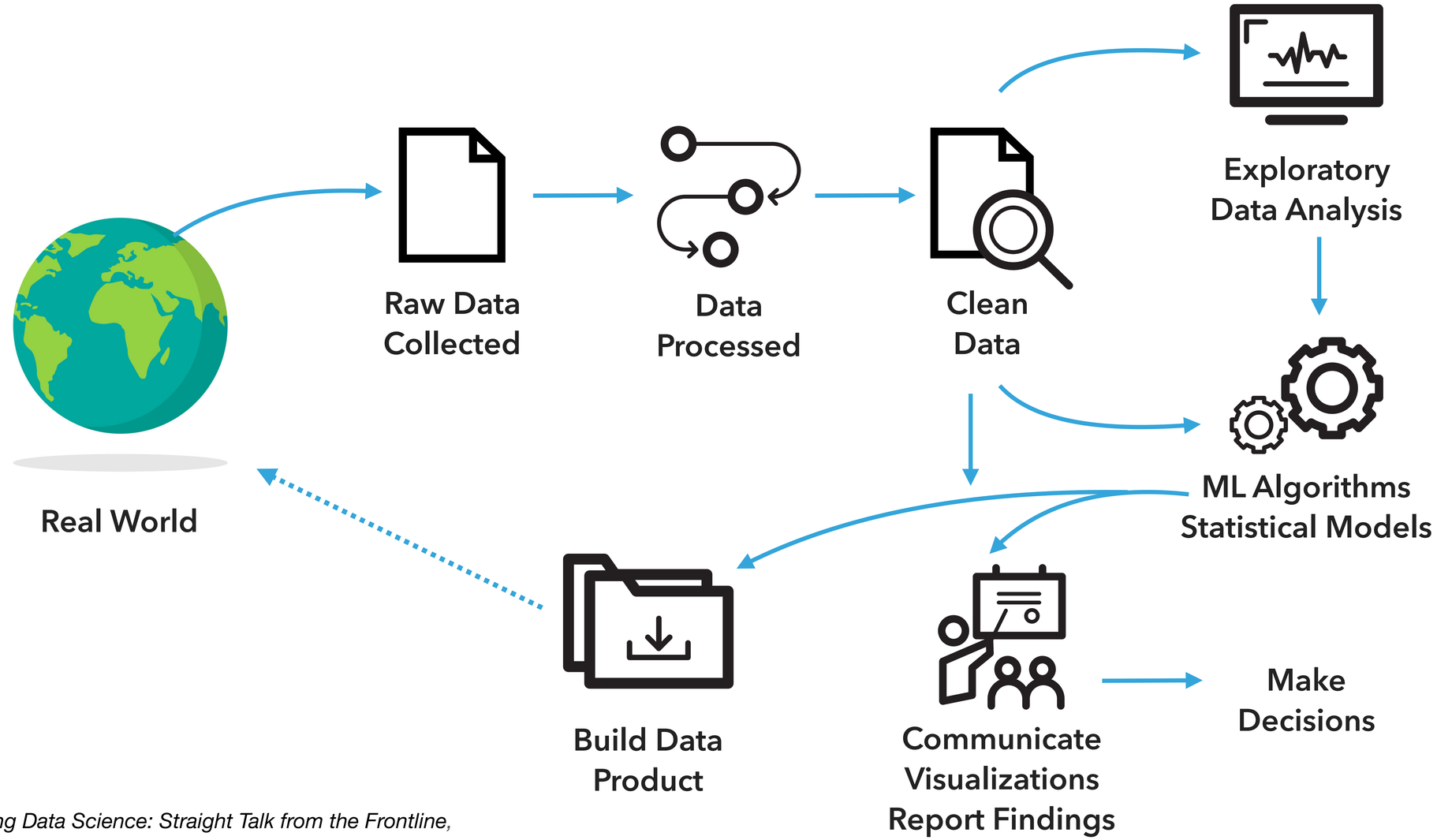
A data scientist is someone who is better at statistics than any software engineer and better at software engineering than any statistician.

WHAT IS DATA SCIENCE?

Data science is a multi-disciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data.[Wikipedia]

What's the Role of Visualization?

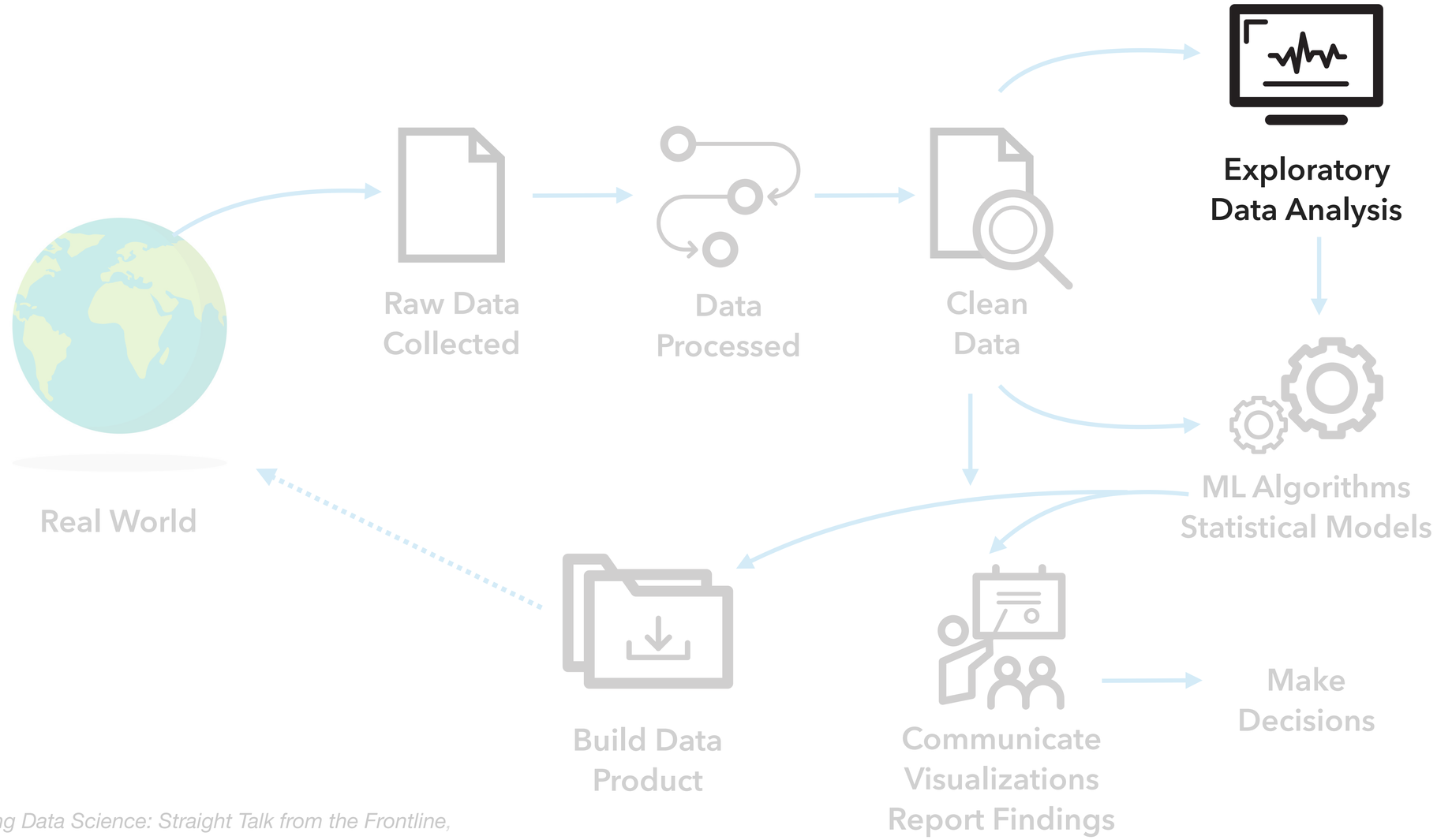
What's the Role of Visualization in Data Science?



Adapted from Doing Data Science: Straight Talk from the Frontline,

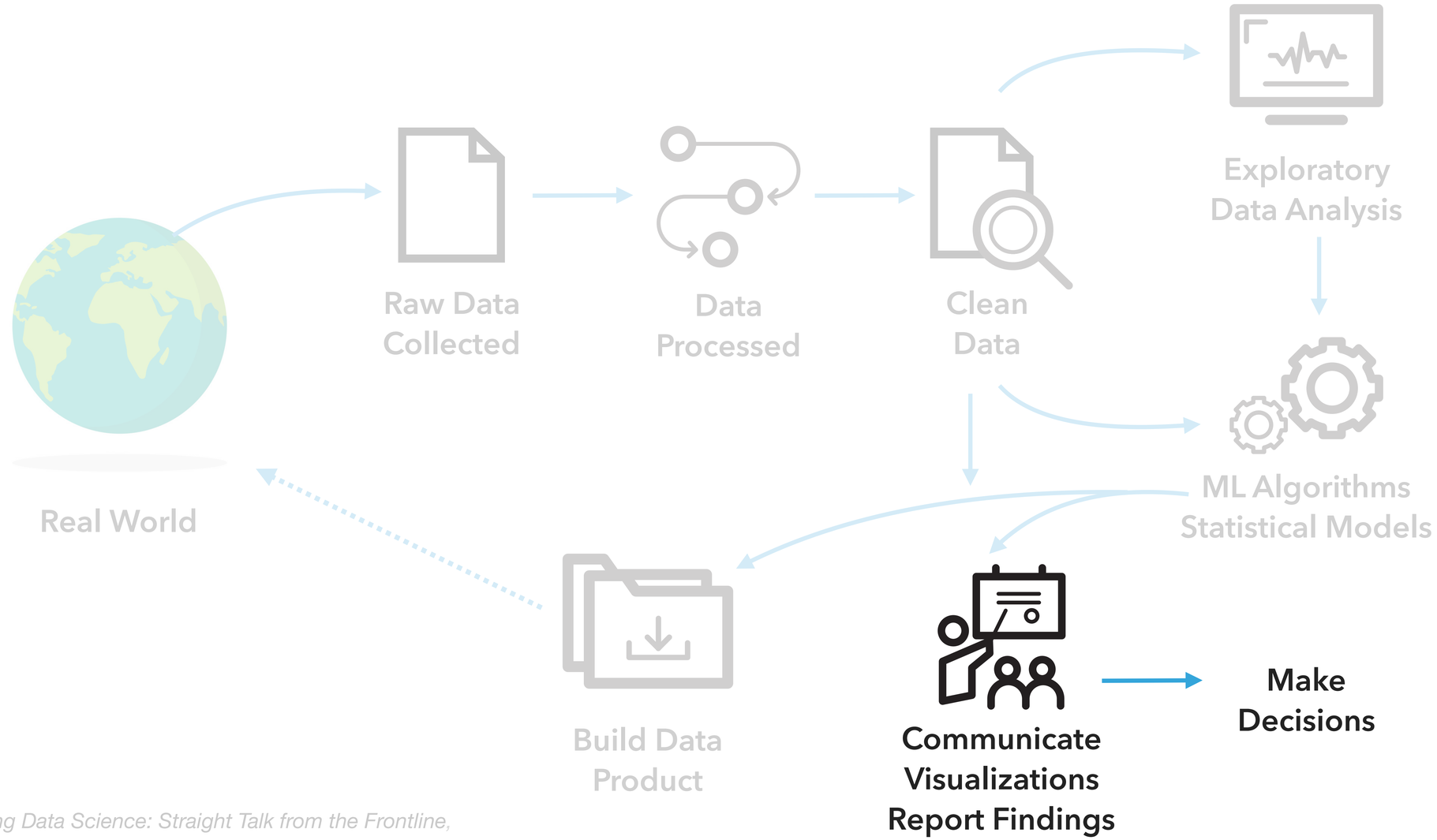
[O'Neil and Schutt, 2013]

What's the Role of Visualization in Data Science?



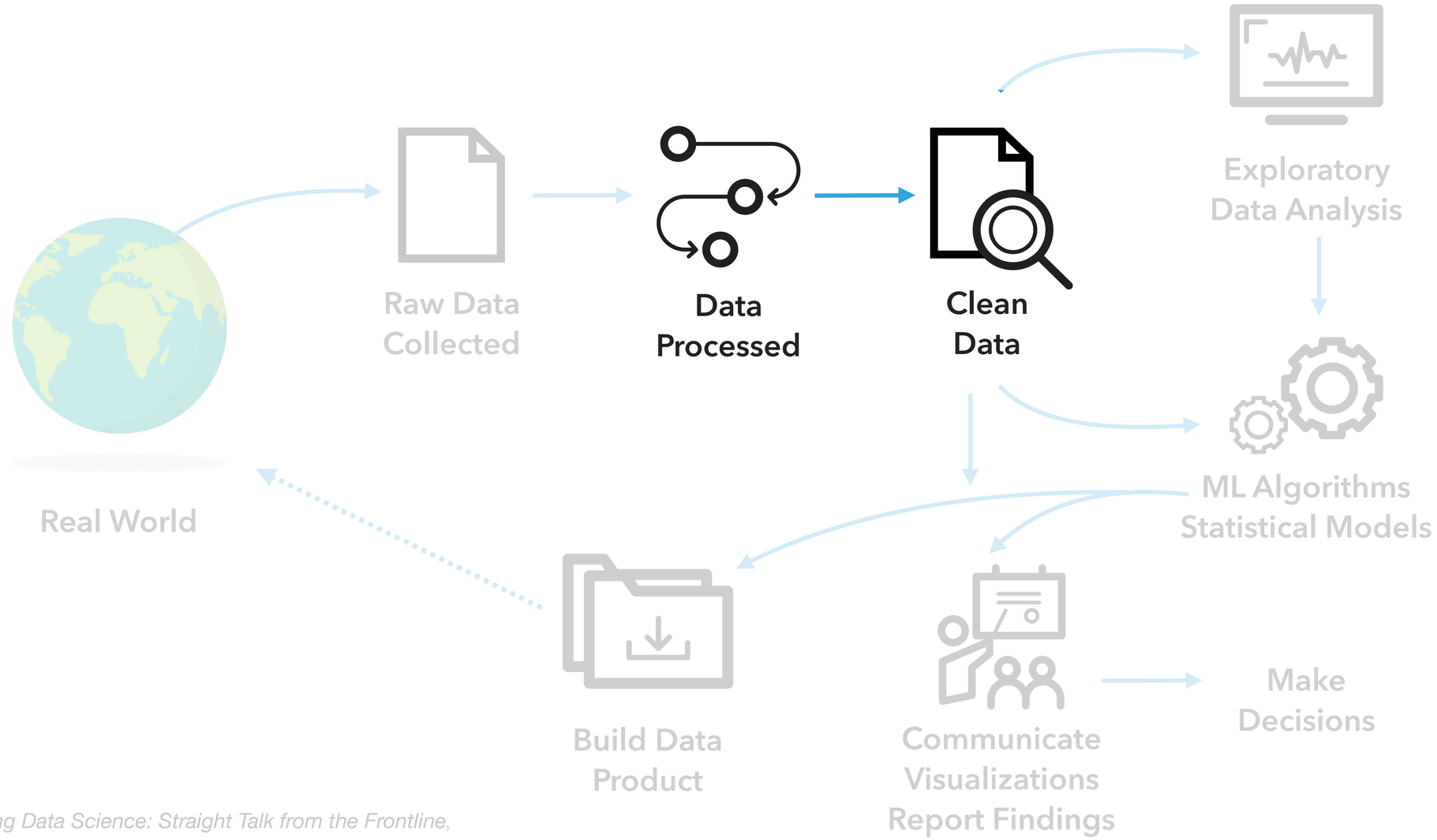
Adapted from Doing Data Science: Straight Talk from the Frontline, [O'Neil and Schutt, 2013]

What's the Role of Visualization in Data Science?



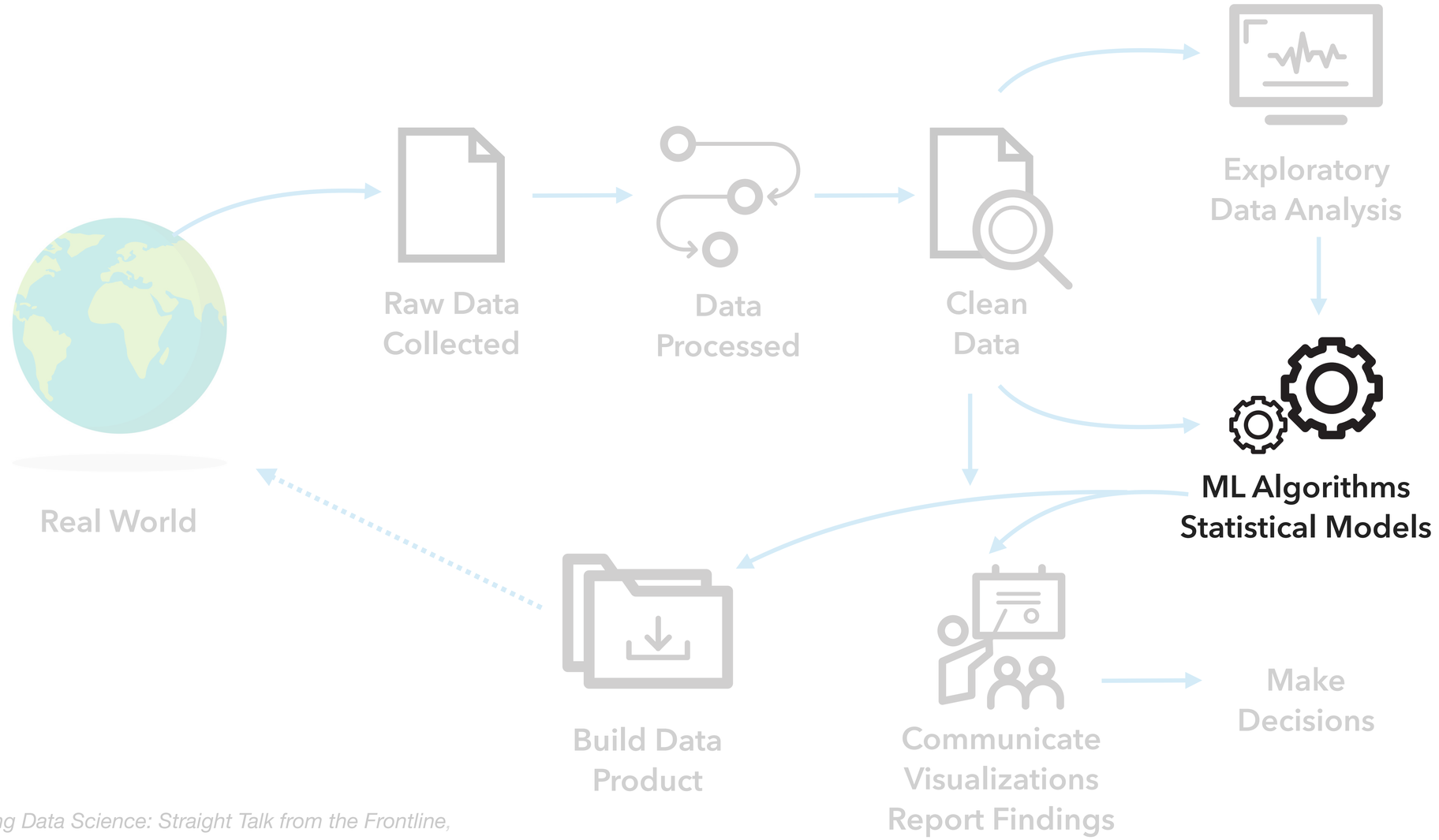
Adapted from Doing Data Science: Straight Talk from the Frontline, [O'Neil and Schutt, 2013]

What's the Role of Visualization in Data Science?



Adapted from Doing Data Science: Straight Talk from the Frontline, [O'Neil and Schutt, 2013]

What's the Role of Visualization in Data Science?



Adapted from Doing Data Science: Straight Talk from the Frontline, [O'Neil and Schutt, 2013]

Visualization =

Human Data Interaction

RESEARCH AREAS

TECHNICAL CONTRIBUTIONS

Novel Visualization Techniques

Visualization Process Innovations

Data Wrangling Methods

DOMAIN DRIVEN TECHNIQUES

Tailored Methods and Systems for High Impact Science Problems

EMPIRICAL & THEORETICAL WORK

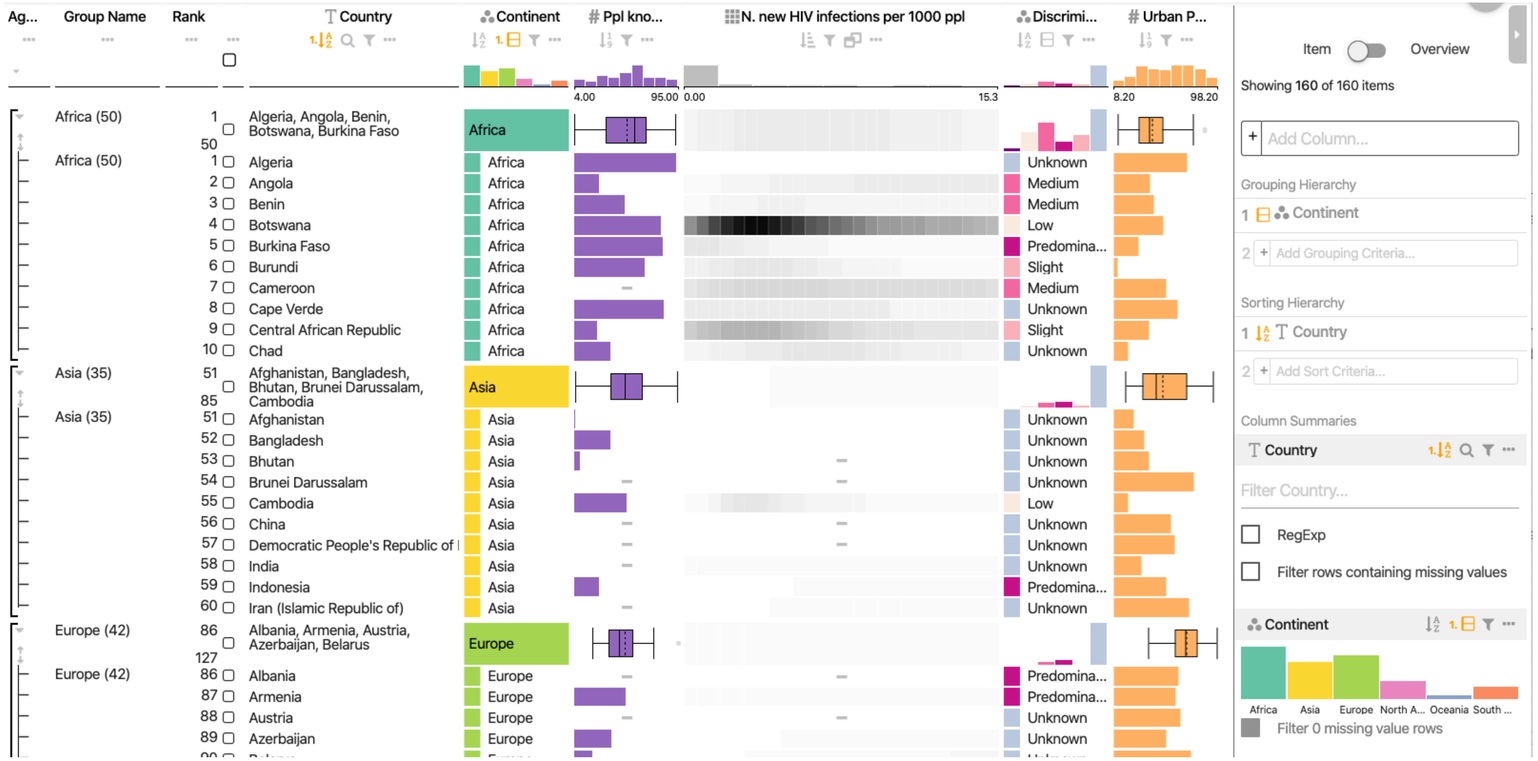
Evaluation Methodology

Design Spaces / Taxonomies

TECHNICAL CONTRIBUTIONS

Tabular Data

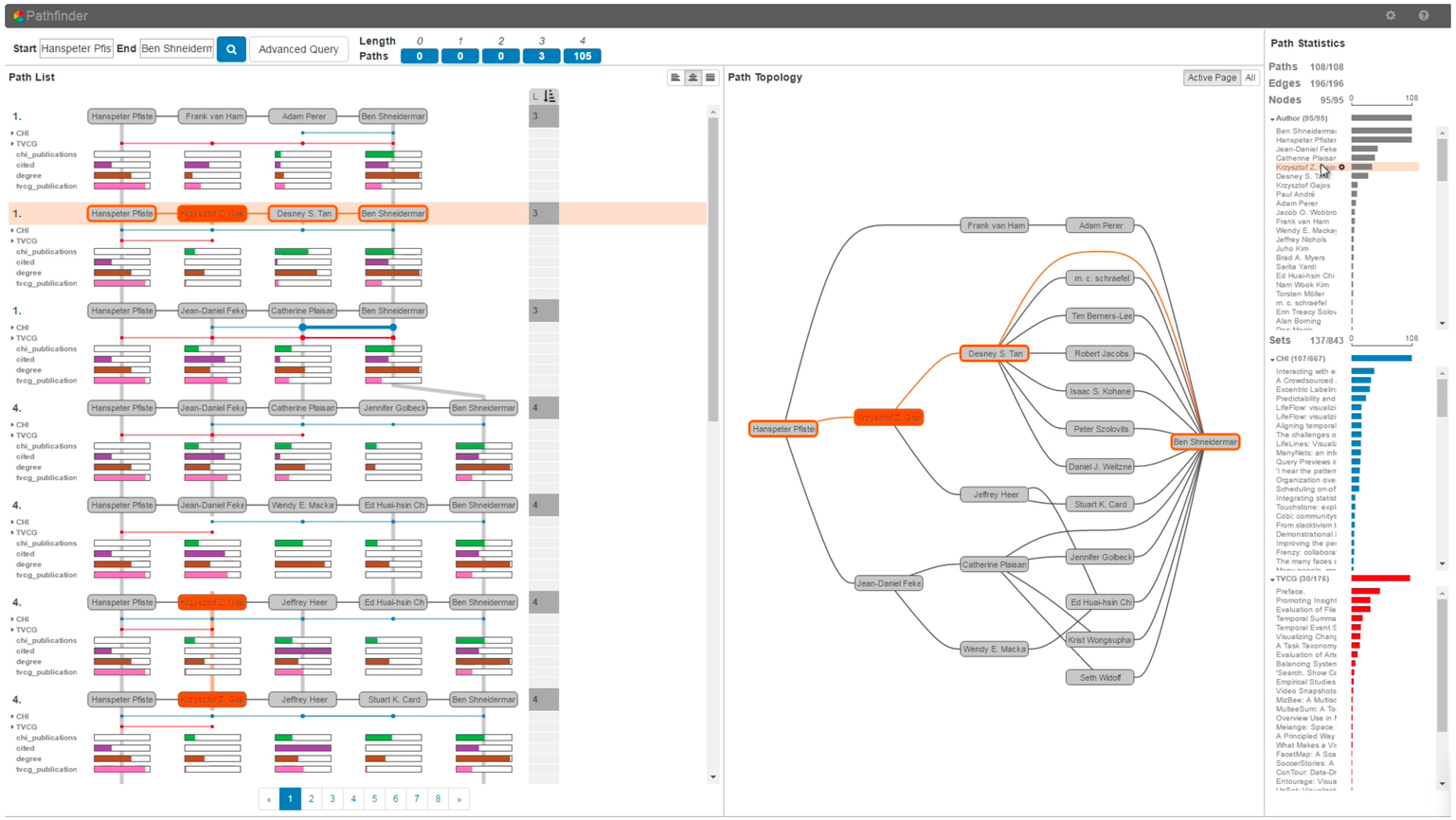
Novel Visualization Techniques



TECHNICAL CONTRIBUTIONS

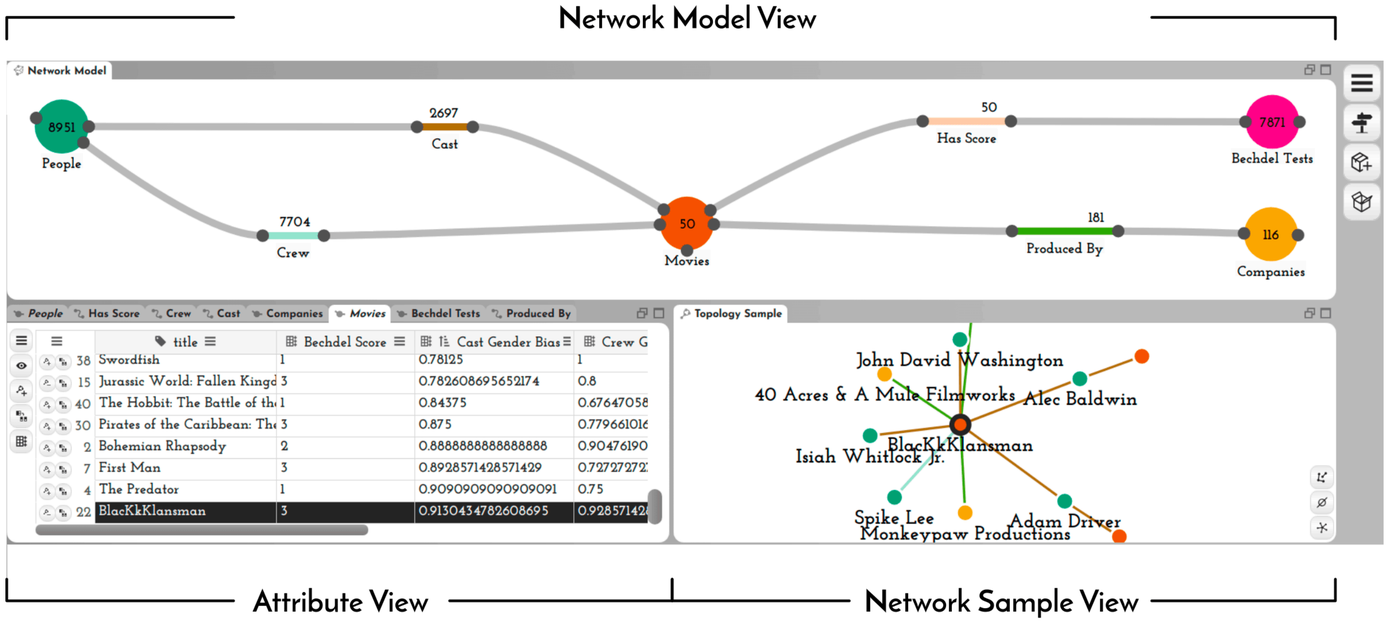
Multivariate Networks

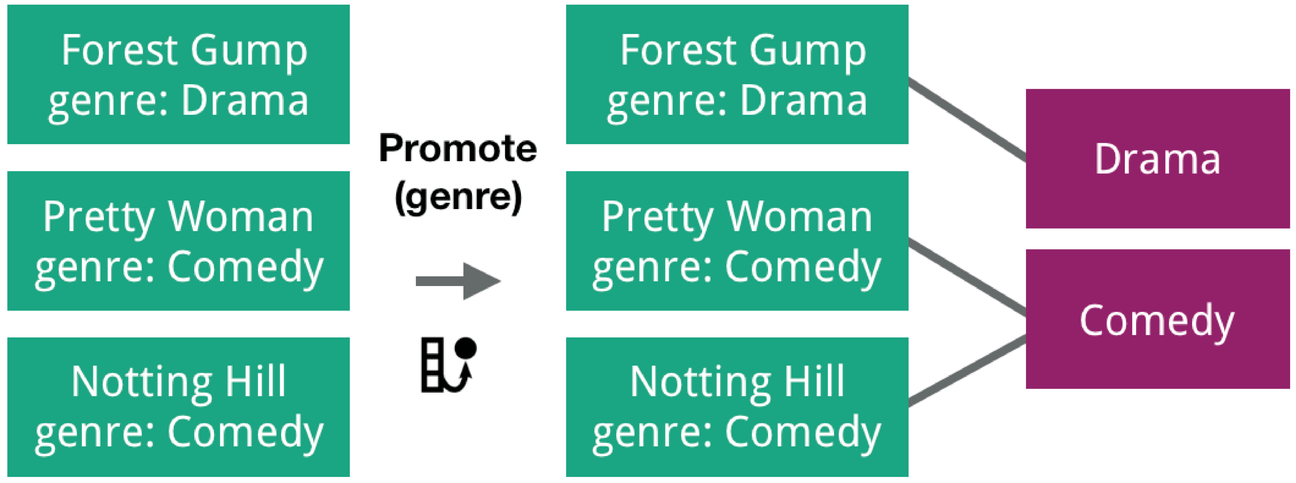
Novel Visualization Techniques



Reshaping Networks

Data Wrangling Methods





TECHNICAL CONTRIBUTIONS

Data Wrangling Methods

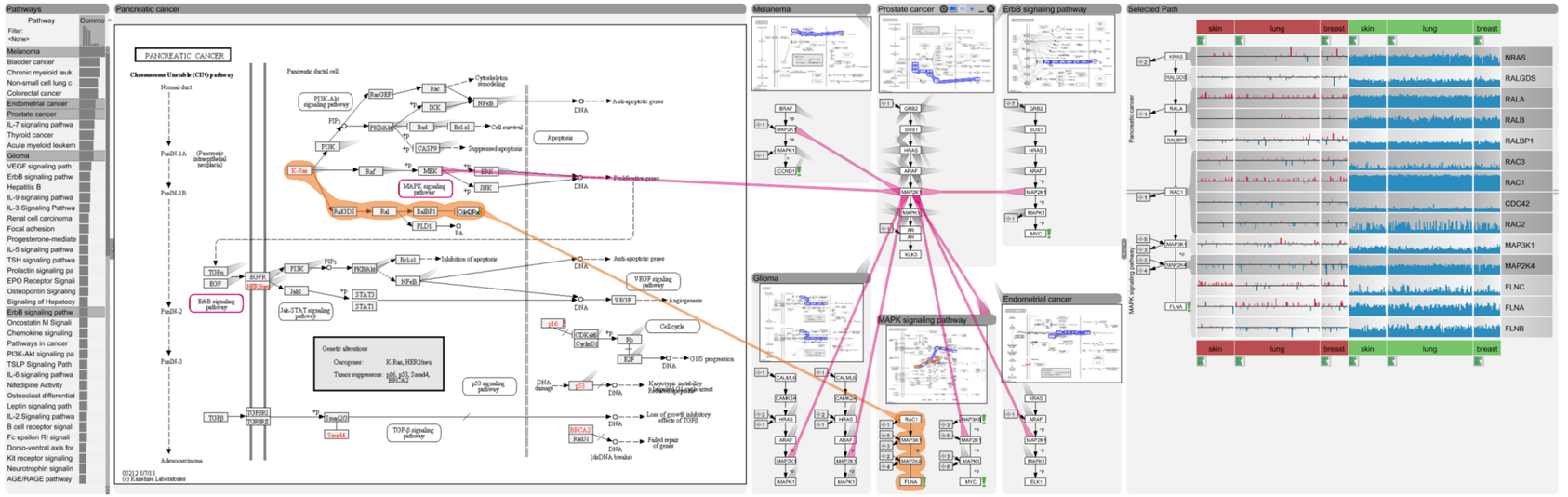
DOMAIN DRIVEN TECHNIQUES

Tailored Methods and Systems for High Impact Science Problems

DOMAIN DRIVEN TECHNIQUES

Pathways and Omics Data

Tailored Methods and Systems for High Impact Science Problems



DOMAIN DRIVEN TECHNIQUES

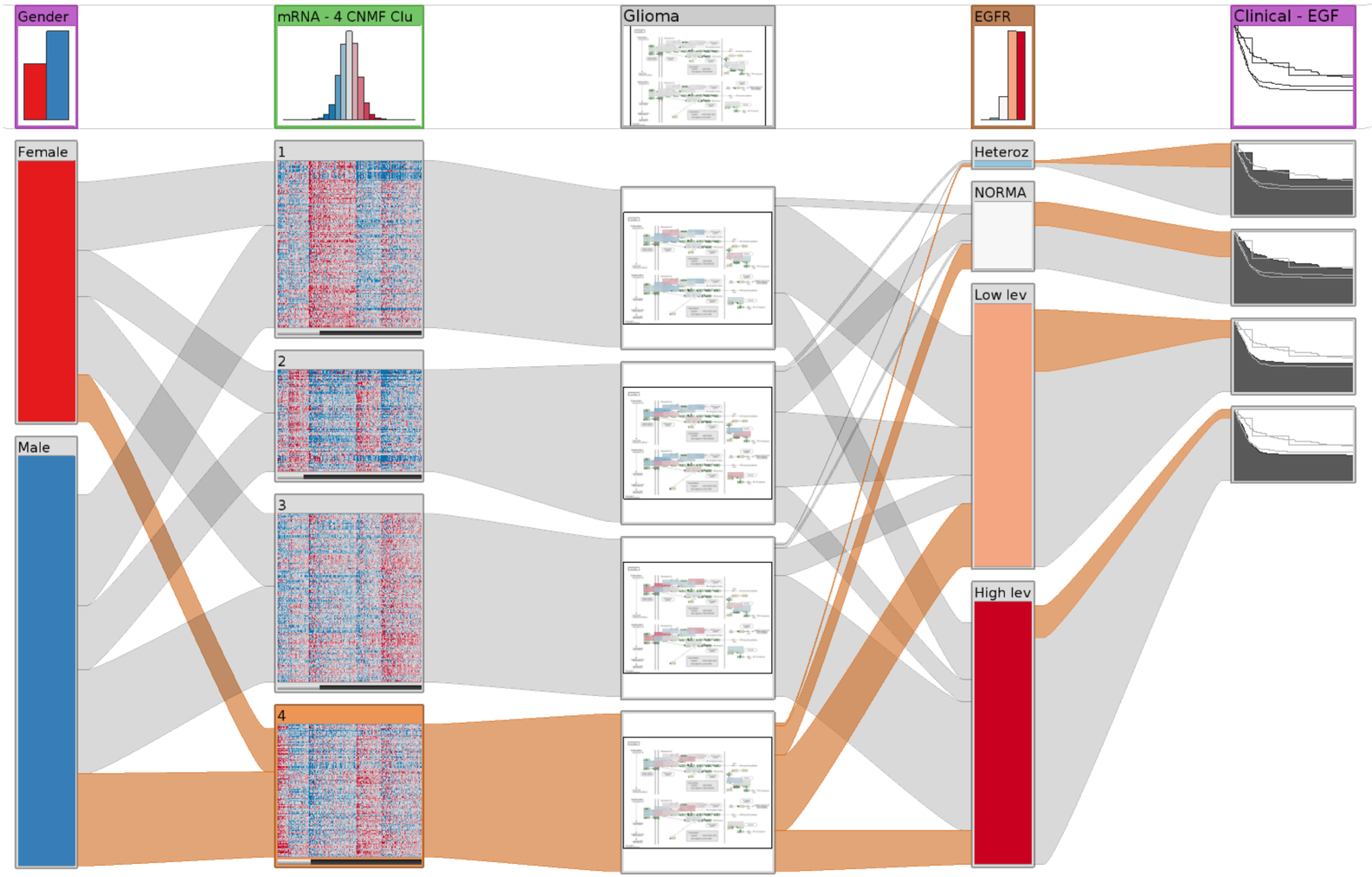
Phylogenies

Tailored Methods and Systems for High Impact Science Problems

DOMAIN DRIVEN TECHNIQUES

Genomic Data

Tailored Methods and Systems for High Impact Science Problems



DOMAIN DRIVEN TECHNIQUES

Tailored Methods and Systems for High Impact Science Problems

EMPIRICAL & THEORETICAL WORK

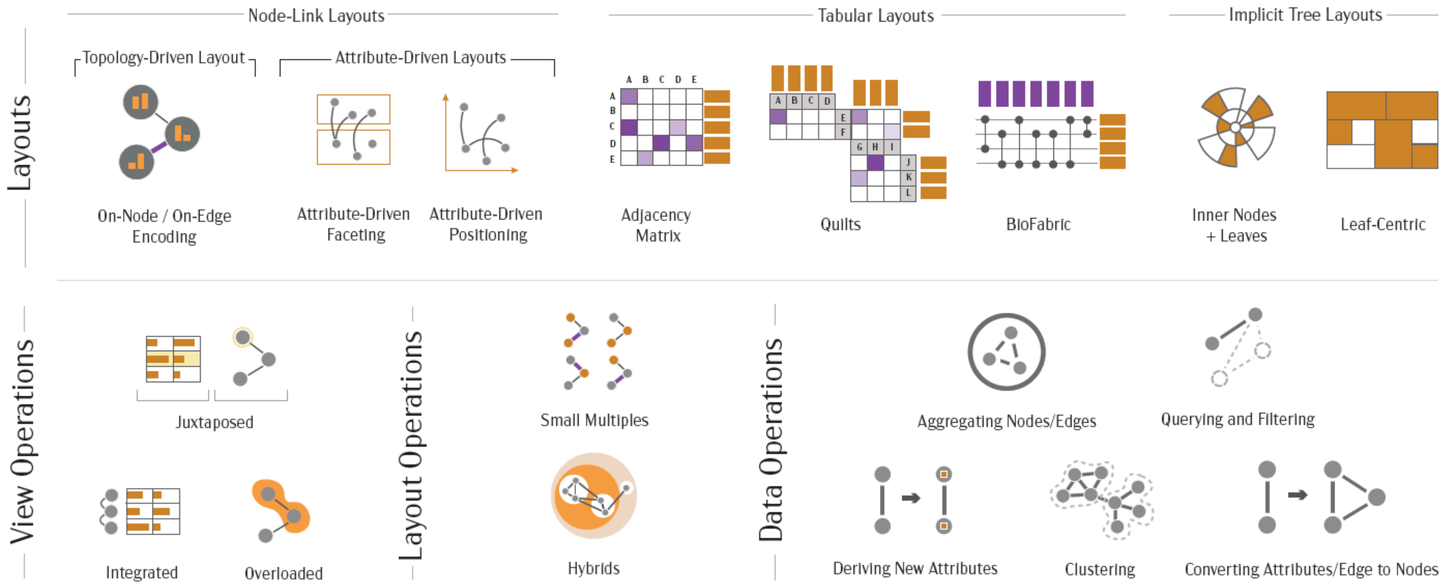
Evaluation Methodology

Design Spaces / Taxonomies

EMPIRICAL & THEORETICAL WORK

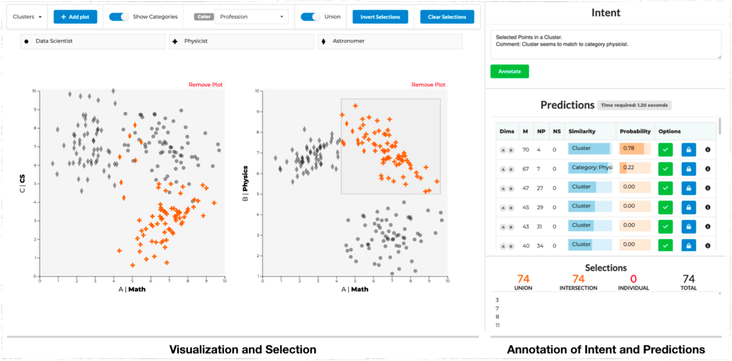
Evaluation Methodology

Design Spaces / Taxonomies



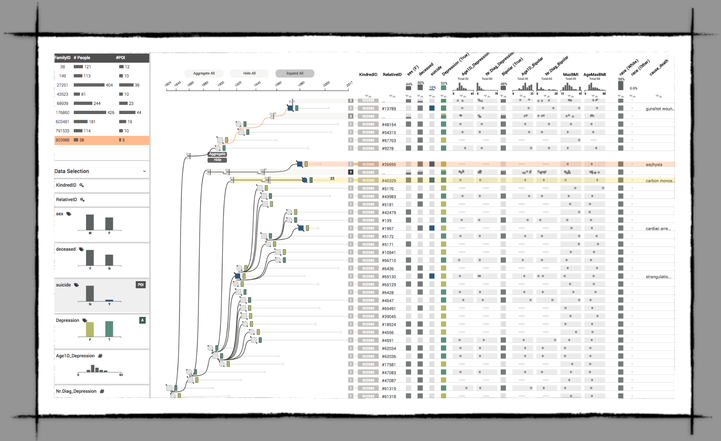
TECHNICAL CONTRIBUTIONS

Literate Visualization



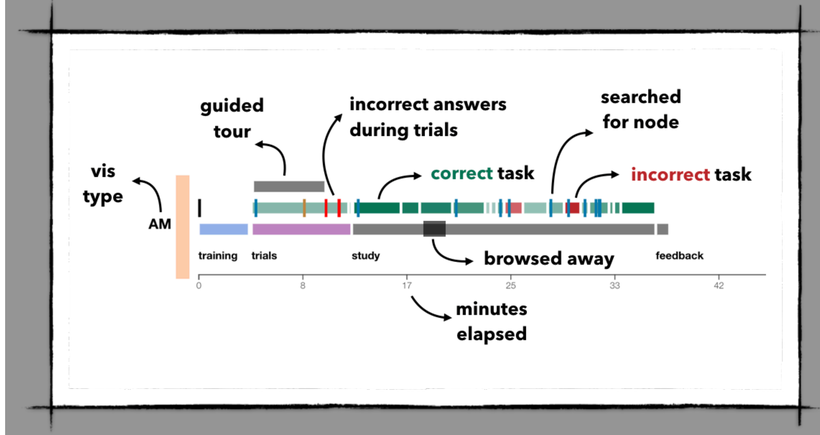
DOMAIN DRIVEN TECHNIQUES

Clinical Genealogies

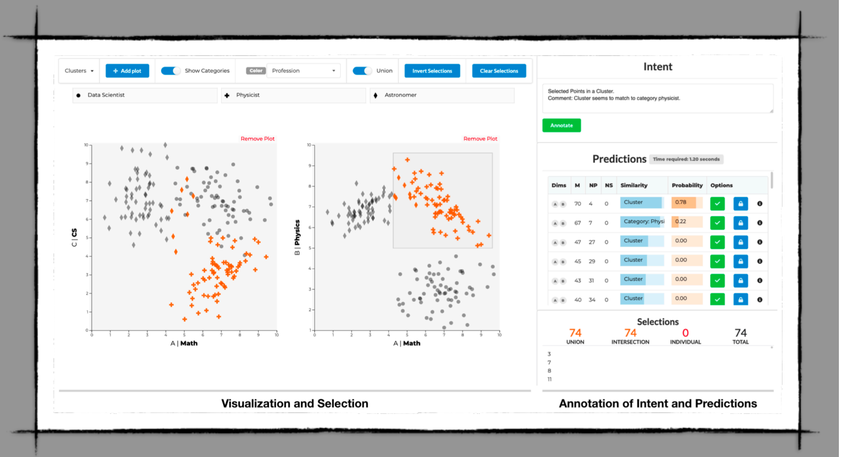


EMPIRICAL & THEORETICAL WORK

Evaluating Complex Systems



TECHNICAL CONTRIBUTIONS



LITERATE VISUALIZATION:

MAKING VISUAL ANALYSIS SESSIONS REPRODUCIBLE AND REUSABLE

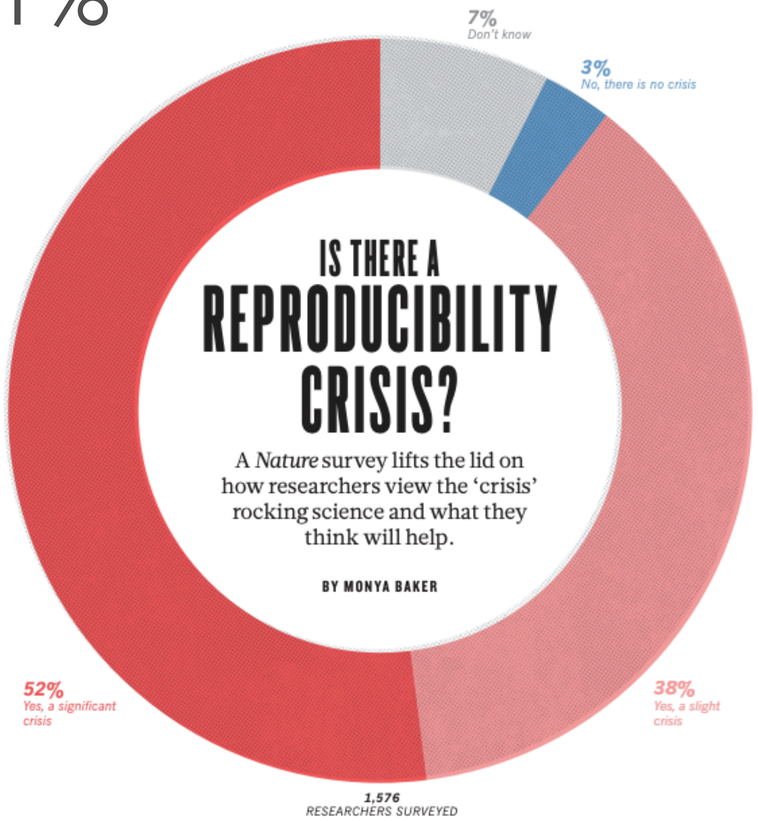
Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Nicola Cosgrove, Marc Streit

REPRODUCIBILITY CRISIS IN SCIENCE

Reproducibility Rates

Psychology: 40%

Cancer Biology: 11%



[Baker, Penny, 2016]
[Open Science Collaboration, 2016]
[Begley, Ellis, 2012]

WHY IS THERE A REPRODUCIBILITY CRISIS?

Perverse incentives (publish or perish)

Bias for “flashy” results

Problems in data analysis

Lack of data sharing

Lack of sharing the analysis process

…

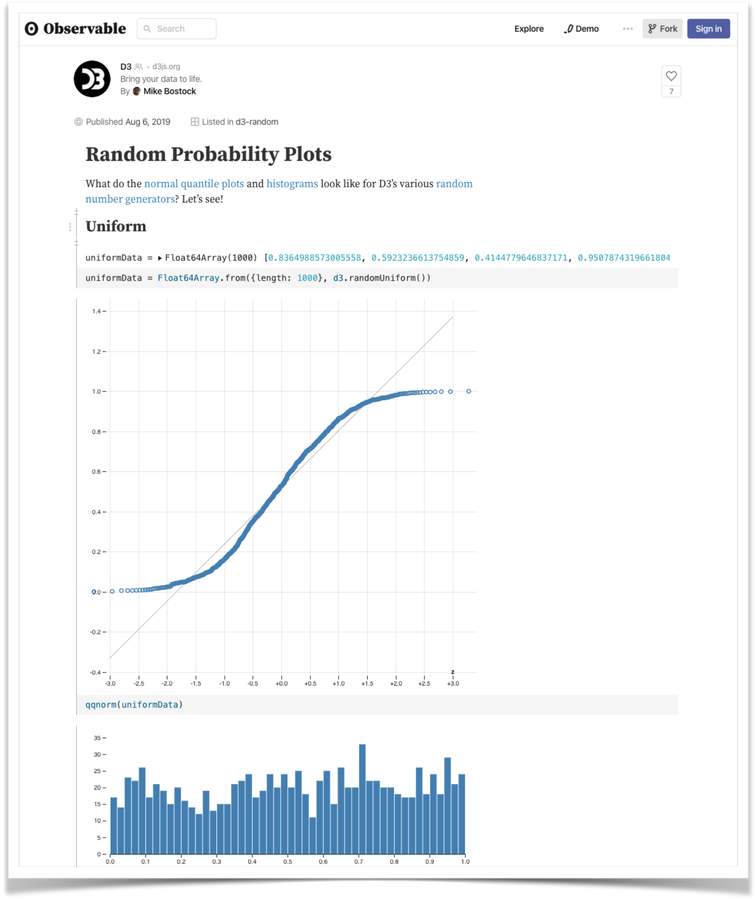
LITERATE PROGRAMMING

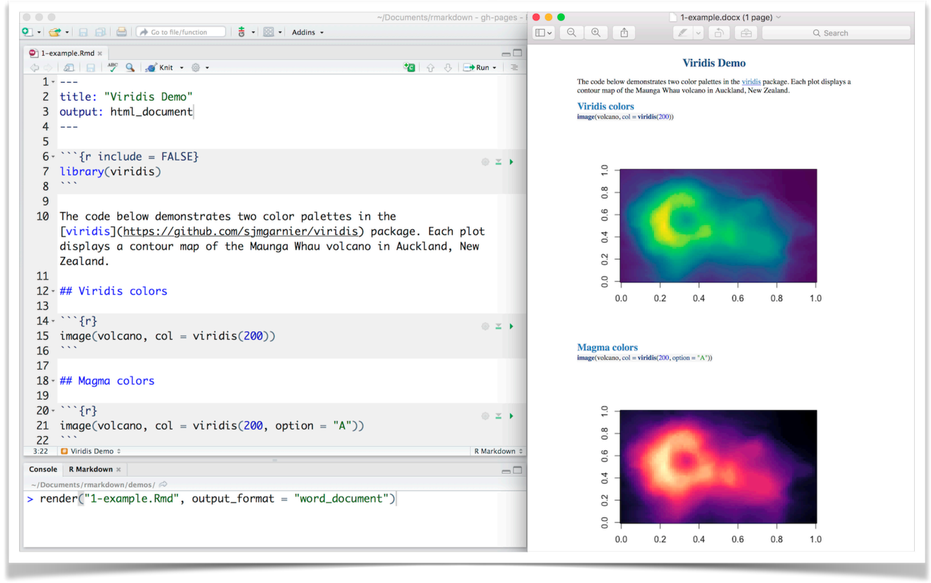
Explain the why and how using any means necessary!

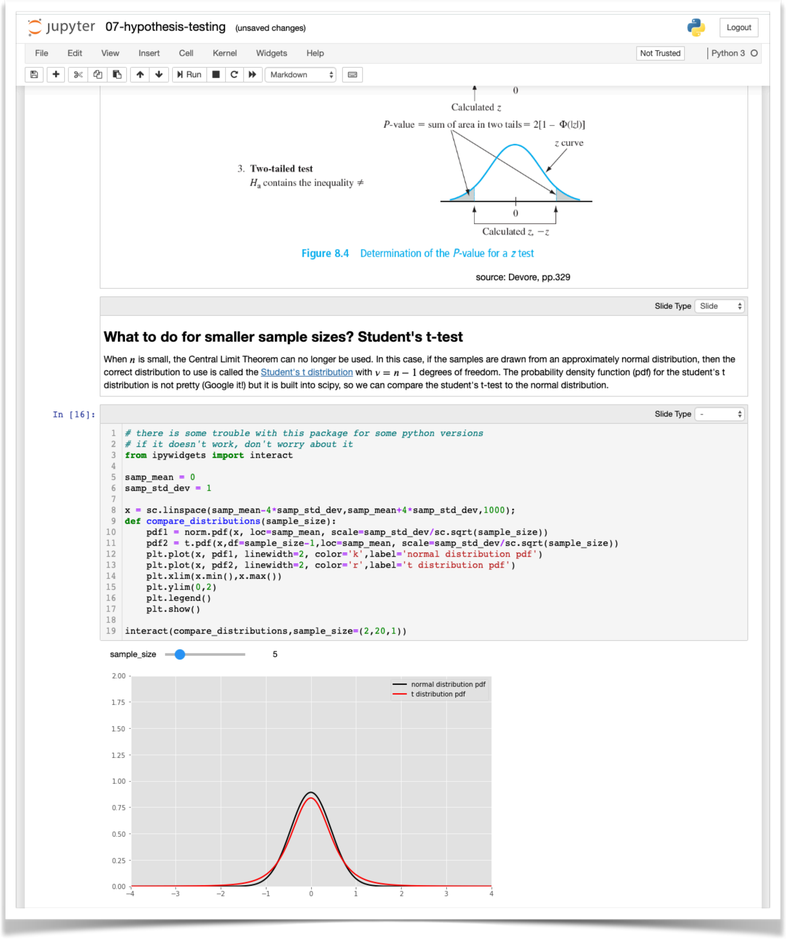
- Text
- Images / Visualizations
- Formulas
- Videos
- Links
- Code

[Donald E. Knuth, 1984]

LITERATE PROGRAMMING IN THE WILD







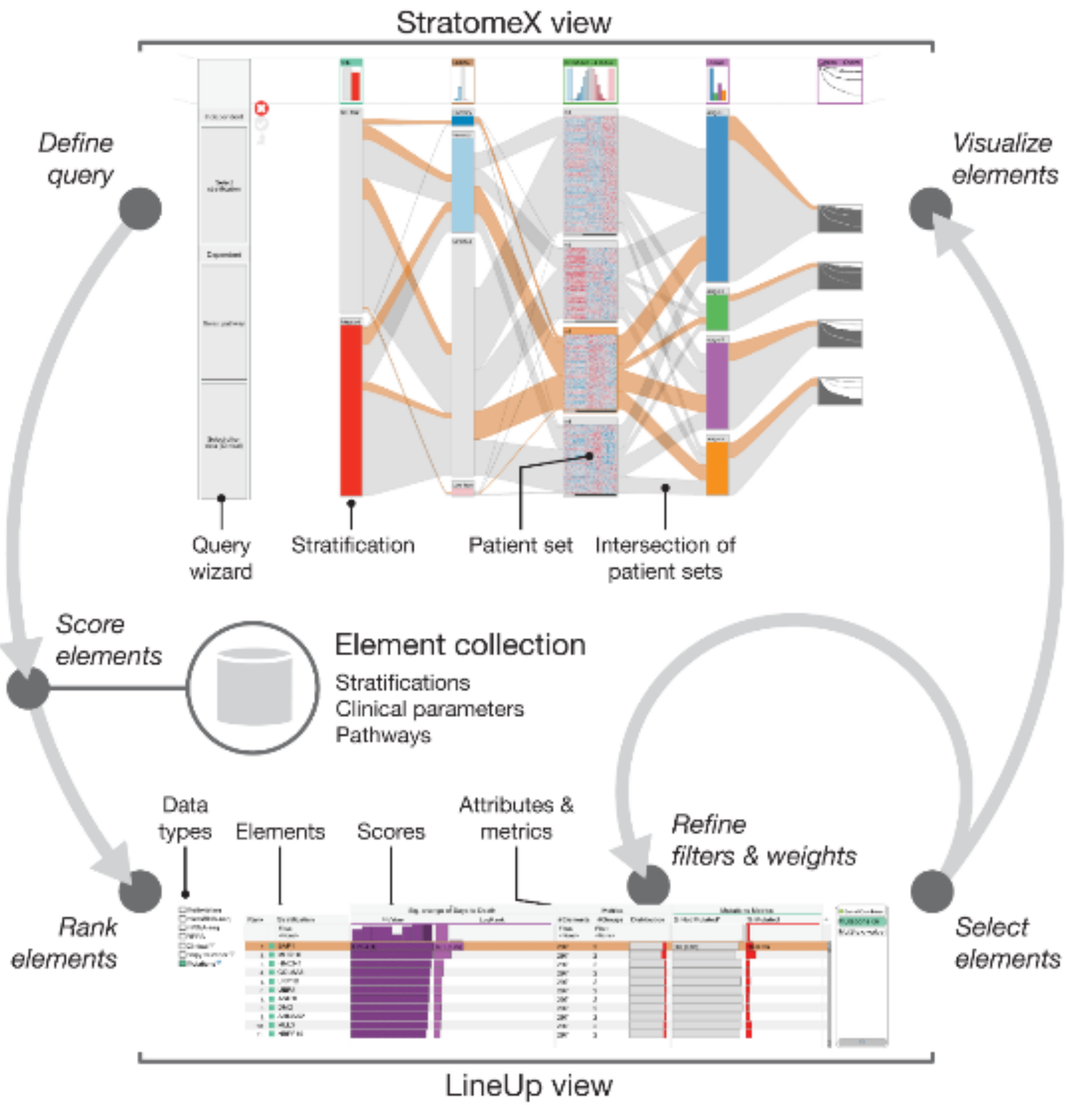
Observable | **R Markdown** | **Jupyter Notebooks**

LITERATE VISUALIZATION

Idea: make the process of an interactive, visual analysis session well reasoned and documented

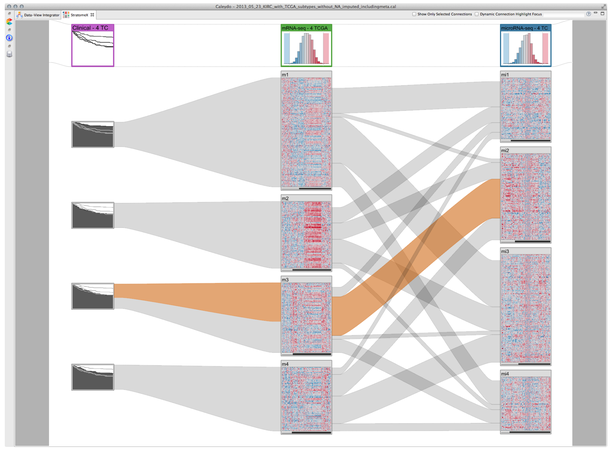
Current State:
no record of what was done,
let alone why

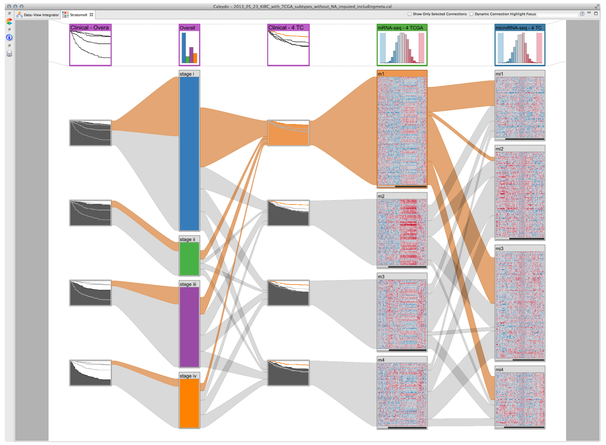
A MANUAL ATTEMPT AT LITERATE VISUALIZATION

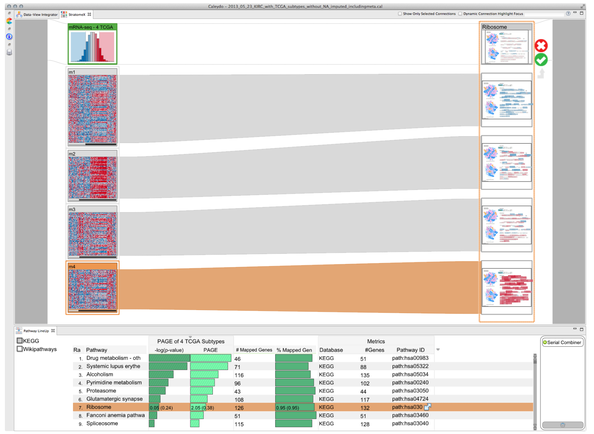


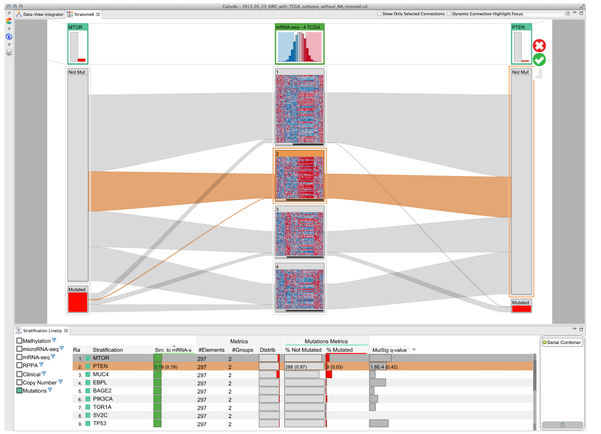
Guided Visual Exploration of Genomic Stratifications in Cancer

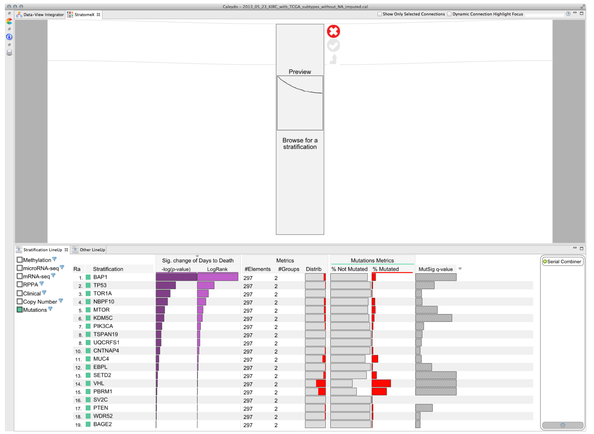
Nature Methods 11, 9 (2014), 884–885

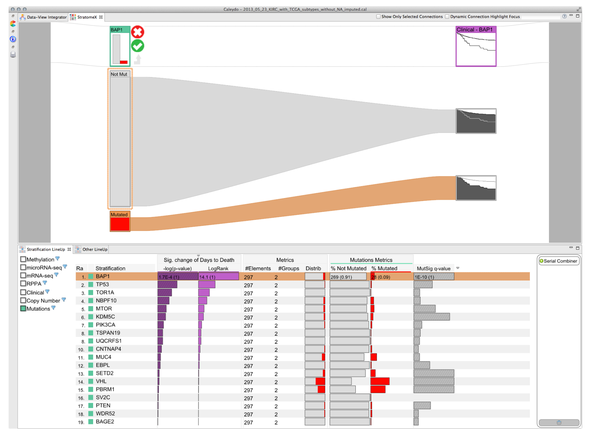


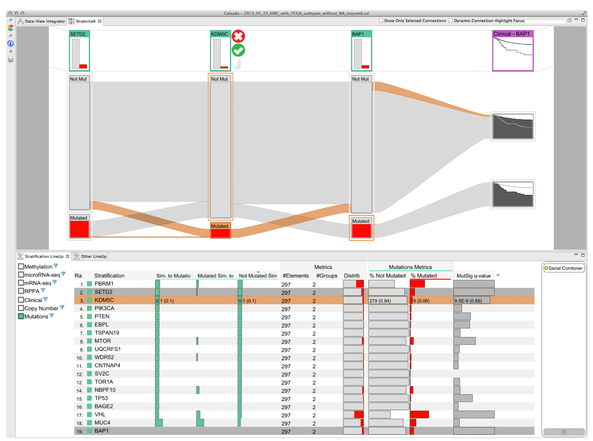


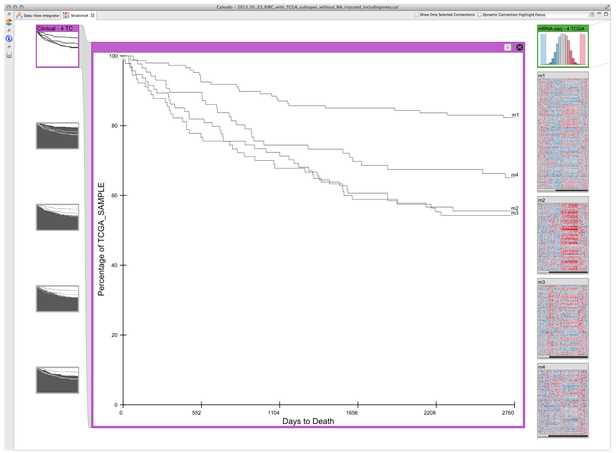






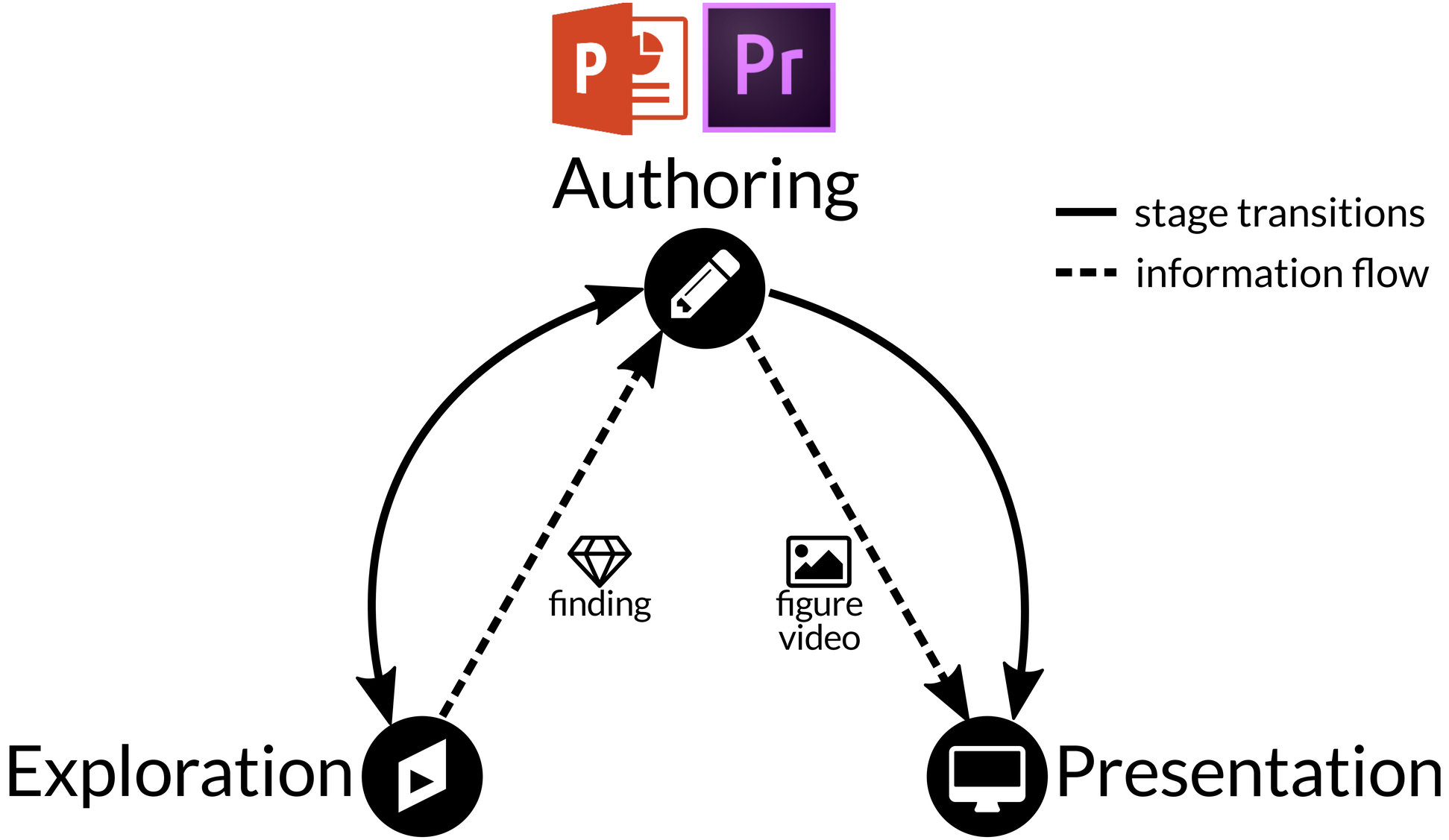


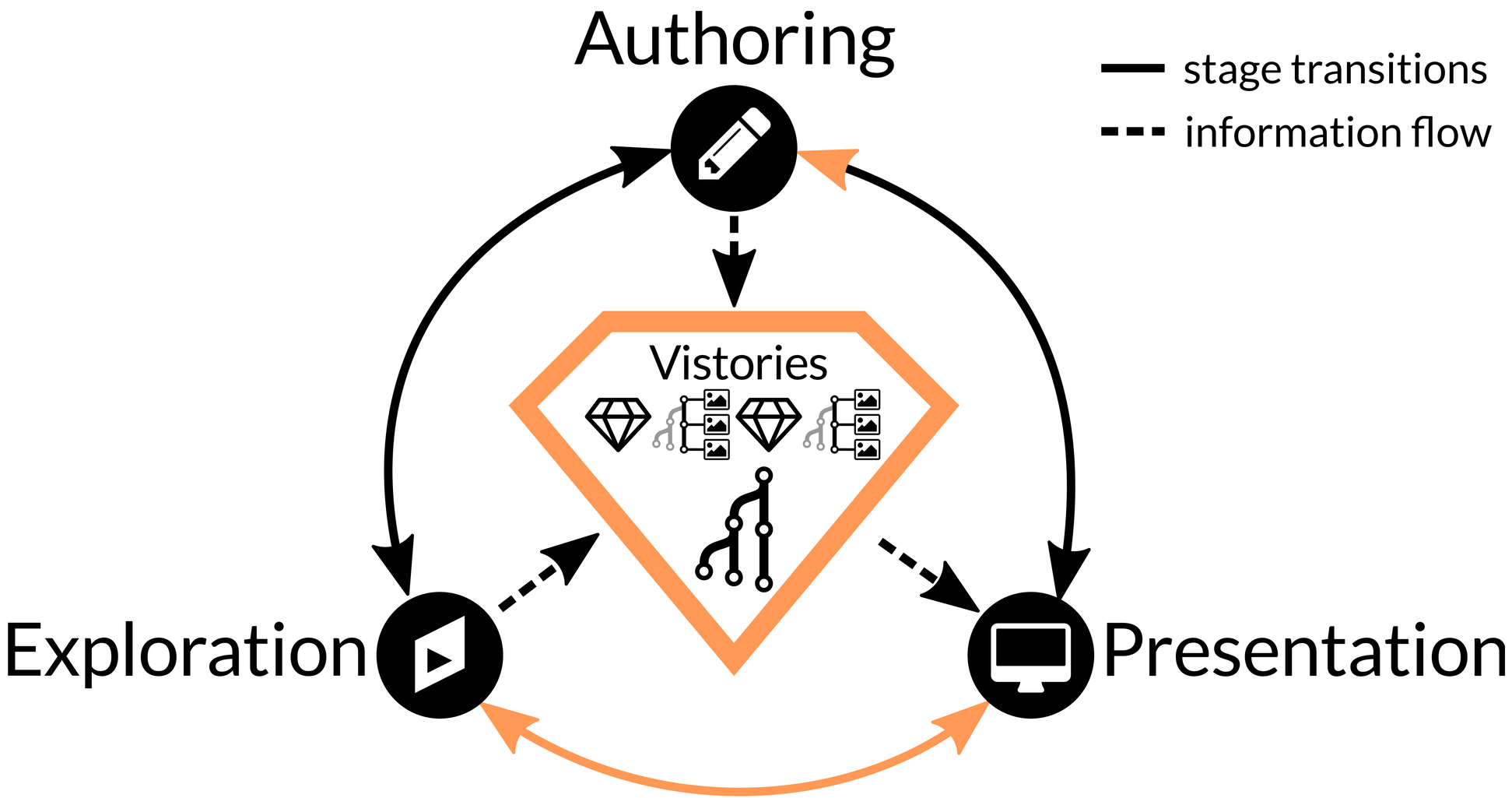


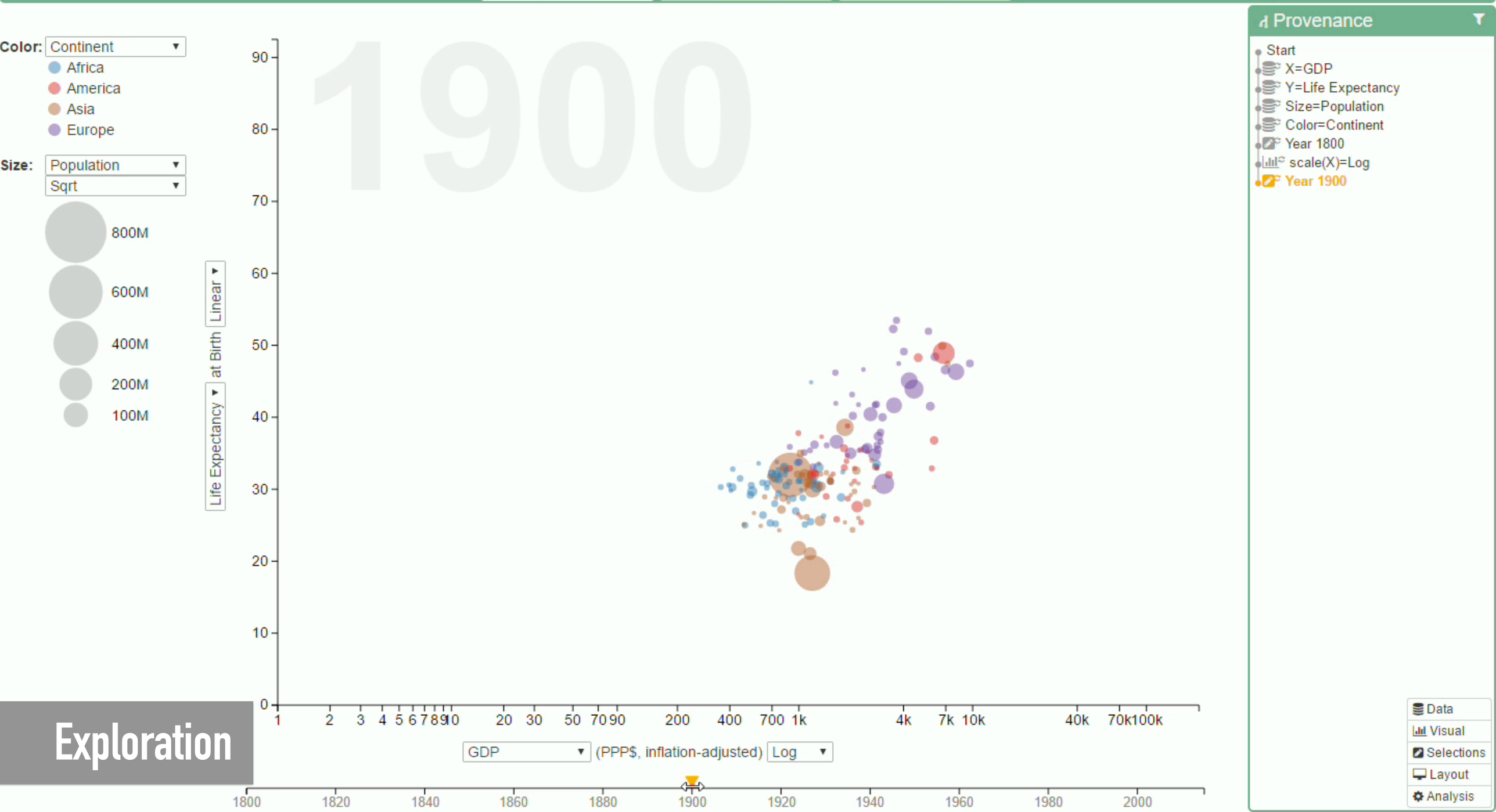


CASE STUDY

Streit et al., Guided Visual Exploration of Genomic Stratifications in Cancer, Nature Methods, 2014


Authoring
stage transitions
information flow
finding
figure
video
Exploration
Presentation


Authoring
stage transitions
information flow
Vistories
Exploration
Presentation

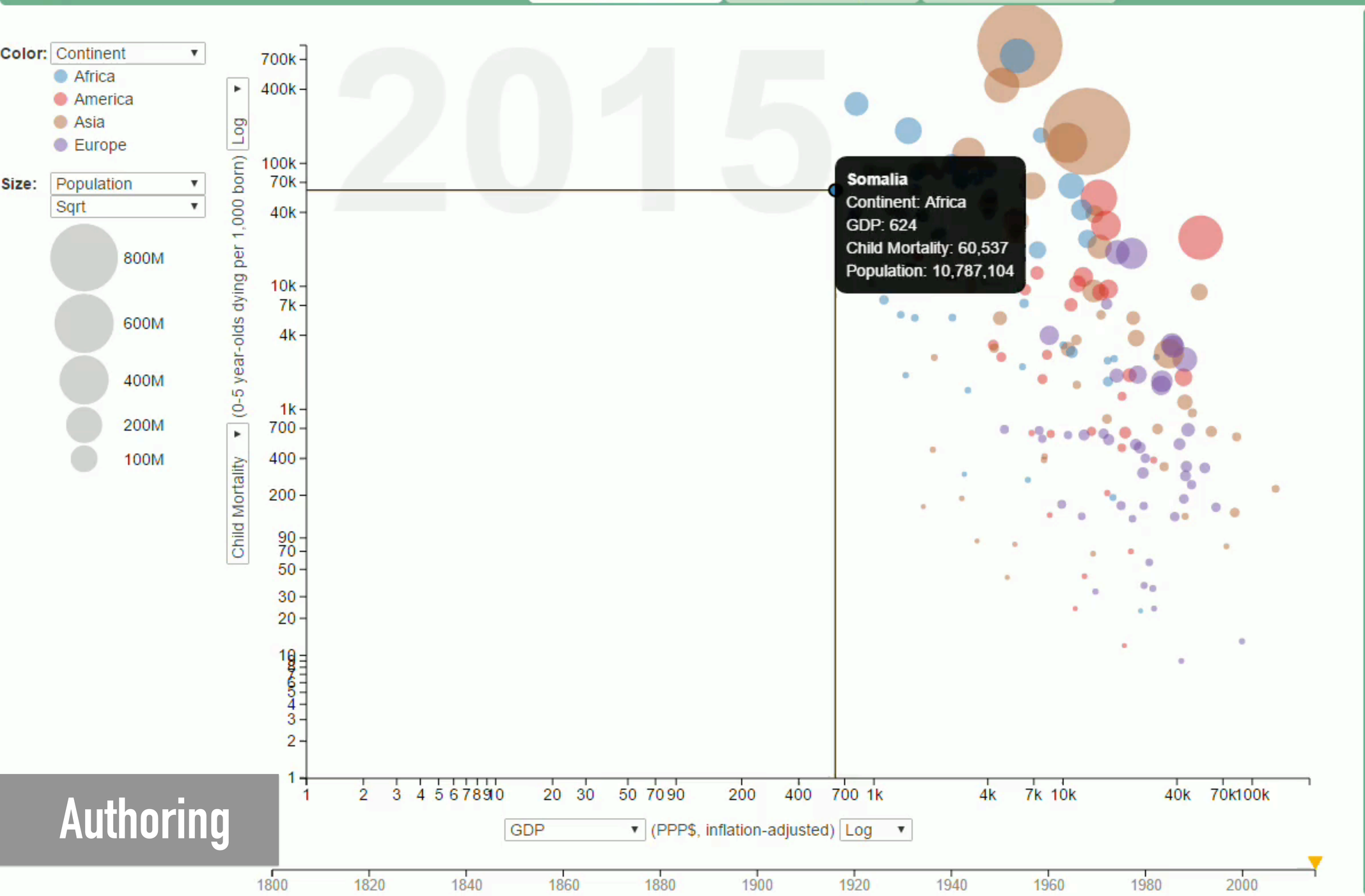


Provenance

- Start
- X=GDP
- Y=Life Expectancy
- Size=Population
- Color=Continent
- Year 1800
- scale(X)=Log
- Year 1900

Data
Visual
Selections
Layout
Analysis

Exploration

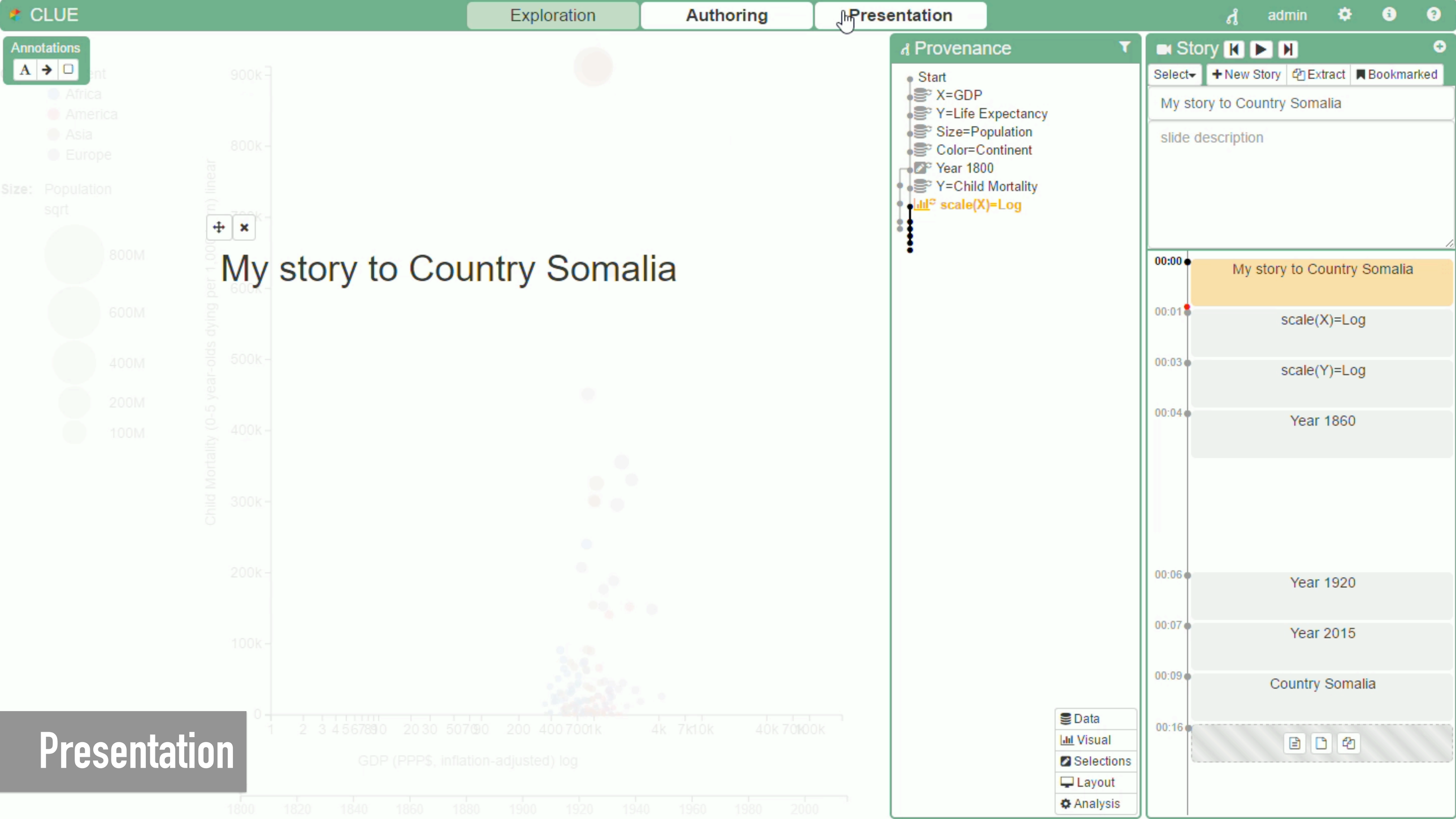


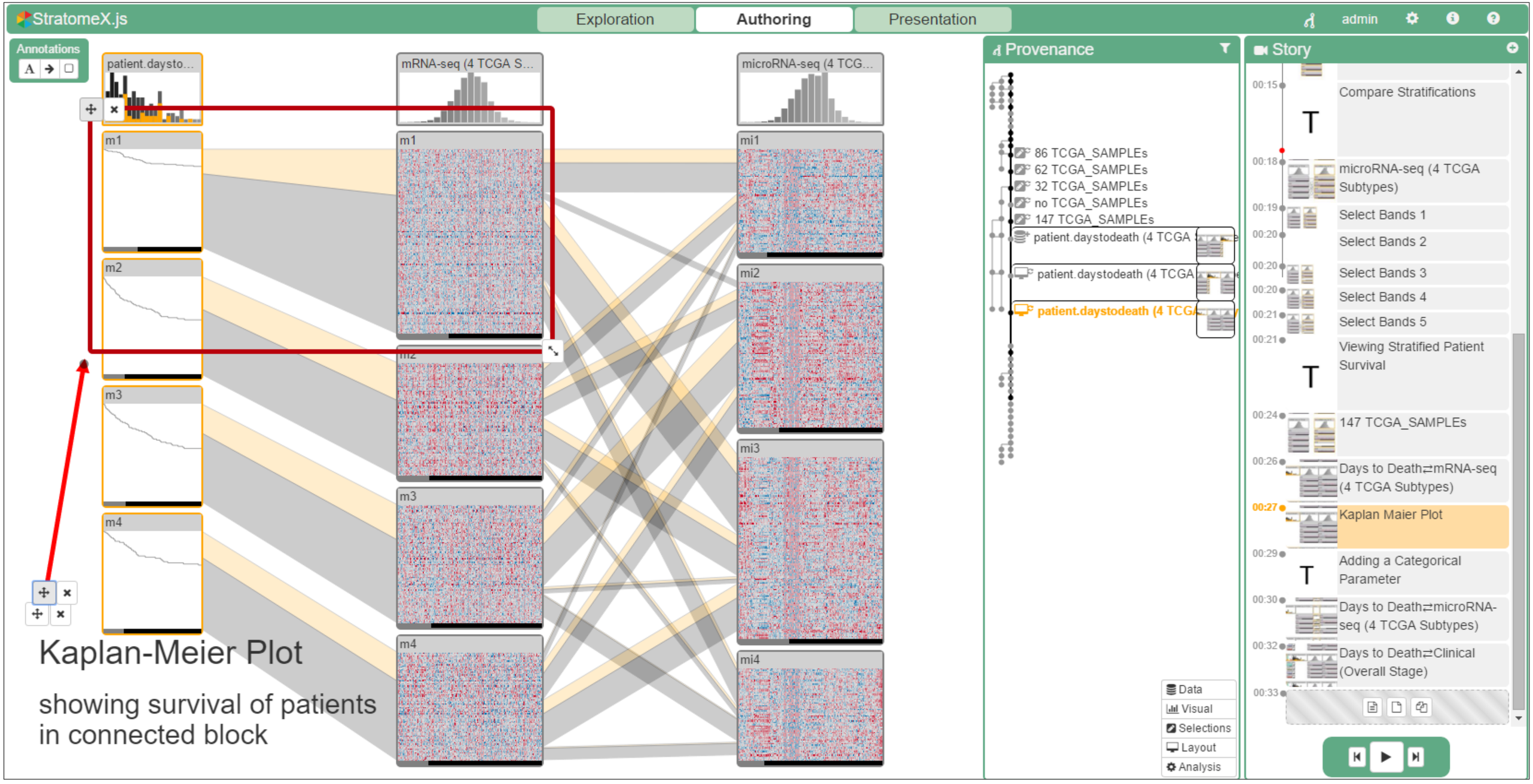
Provenance

- Year 1800
- Y=Child Mortality
- scale(X)=Log
- scale(Y)=Log
- Year 1860
- Year 1920
- Year 2015
- Country Somalia

Data
Visual
Selections
Layout
Analysis

Authoring


CLUE
Exploration
Authoring
Presentation
admin
Annotations
Provenance
Start
X=GDP
Y=Life Expectancy
Size=Population
Color=Continent
Year 1800
Y=Child Mortality
scale(X)=Log
Story
Select
New Story
Extract
Bookmarked
My story to Country Somalia
slide description
My story to Country Somalia
00:00
My story to Country Somalia
00:01
scale(X)=Log
00:03
scale(Y)=Log
00:04
Year 1860
00:06
Year 1920
00:07
Year 2015
00:09
Country Somalia
00:16
Data
Visual
Selections
Layout
Analysis
Africa
America
Asia
Europe
Size: Population
sqrt
800M
600M
400M
200M
100M
Child Mortality (0-5 year-olds dying per 1,000 born) linear
900k
800k
500k
400k
300k
200k
100k
0
1
2
3
4
5
6
7
8
9
10
20
30
50
70
90
200
400
700
1k
4k
7k
10k
40k
70k
100k
GDP (PPP$, inflation-adjusted) log
1800
1820
1840
1860
1880
1900
1920
1940
1960
1980
2000
My story to Country Somalia
Presentation



http://vistories.org/

PROGRESS; BUT PROVENANCE IS "DUMB"

We solved the WHAT,
but not the WHY

So, can we make it easier to capture the why?

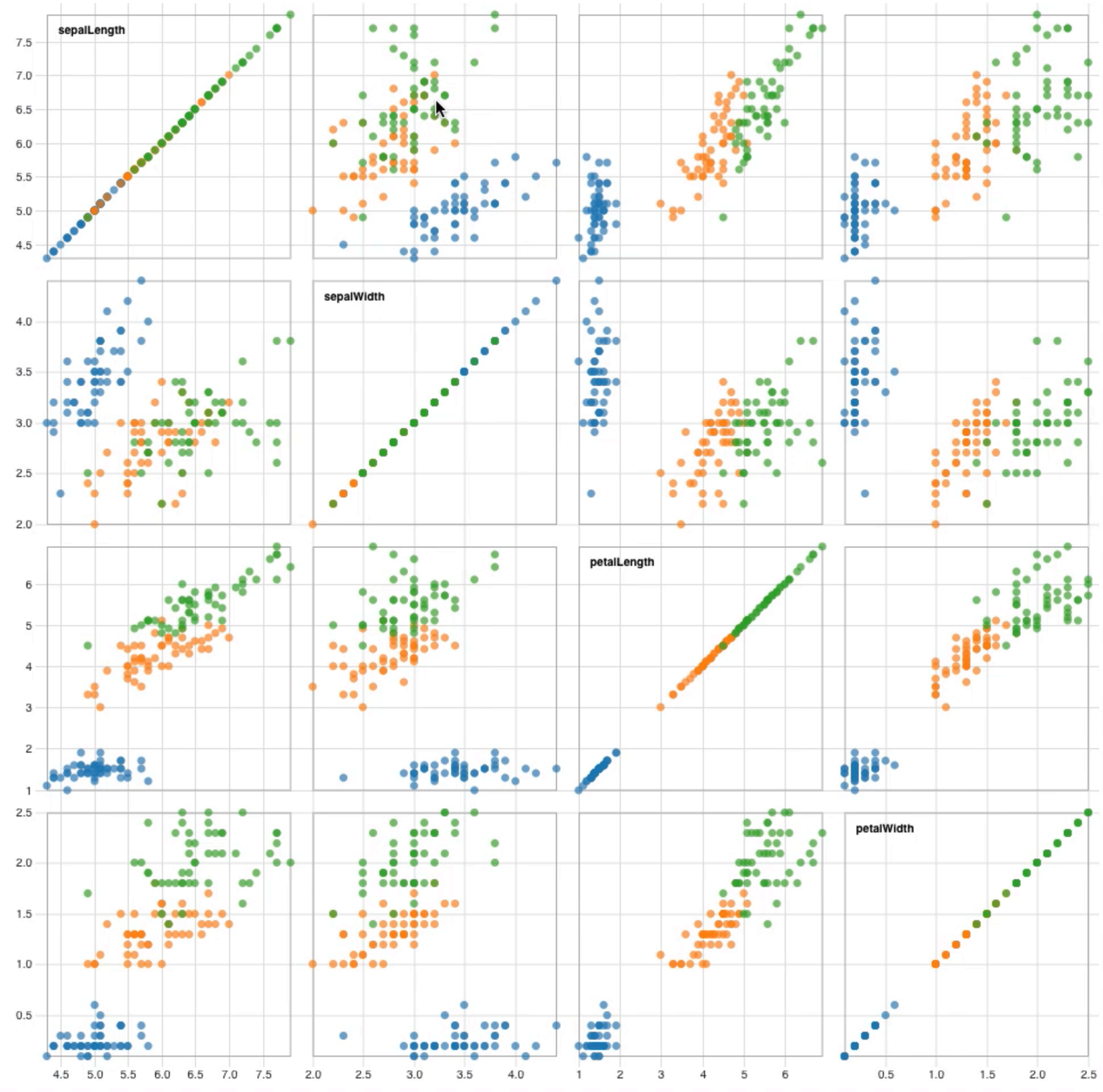
TECHNICAL CONTRIBUTIONS

CAPTURING USER INTENT

WHEN BRUSHING IN SCATTERPLOTS

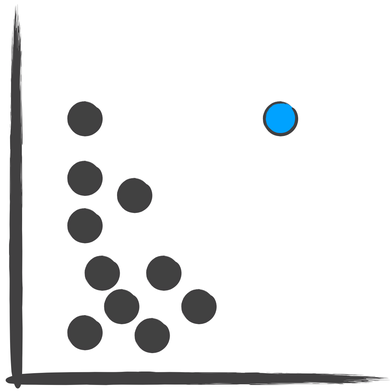
Kiran Gadhave, Jochen Görtler, Oliver Deussen, Miriah Meyer, Jeff Phillips

WHAT'S BRUSHING?

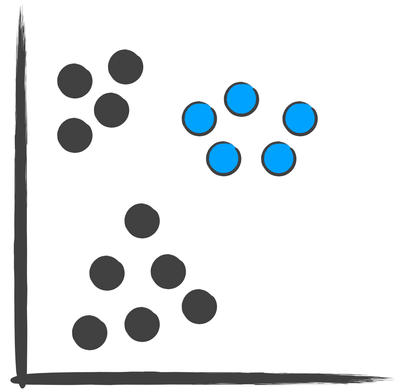


WHAT IS INTENT WHEN BRUSHING?

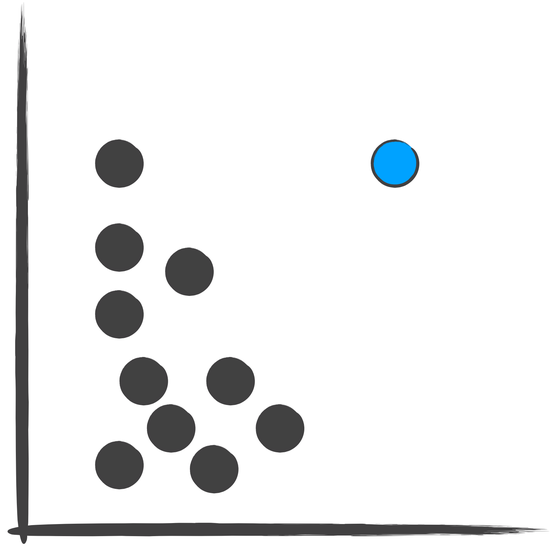
Intent is the user's reason for performing a brush with a visualization.



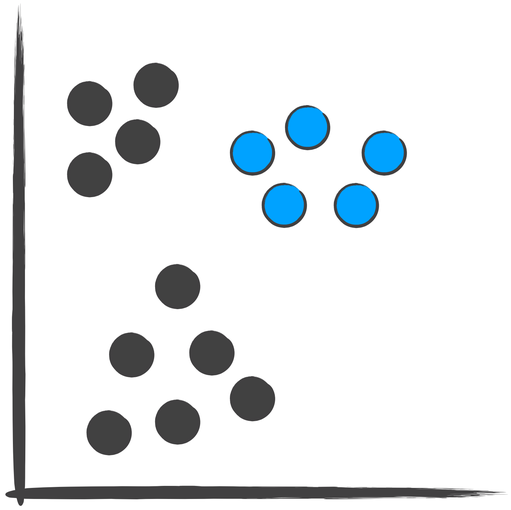
Outlier



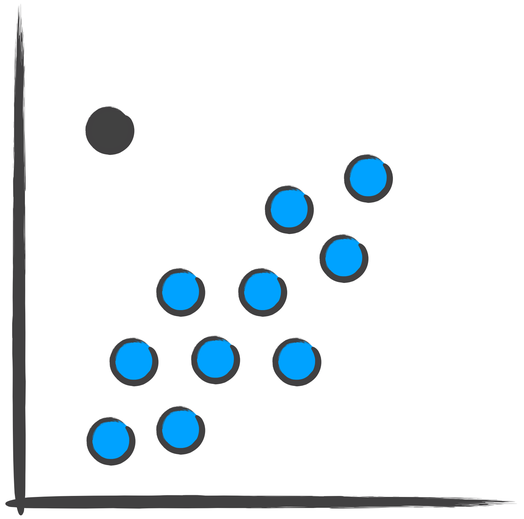
Clusters



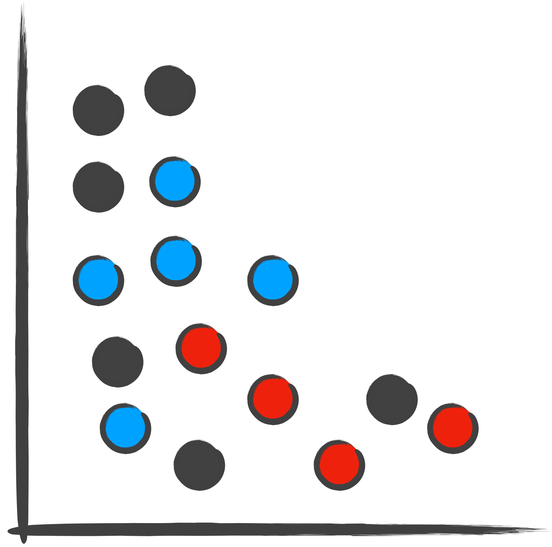
Outlier



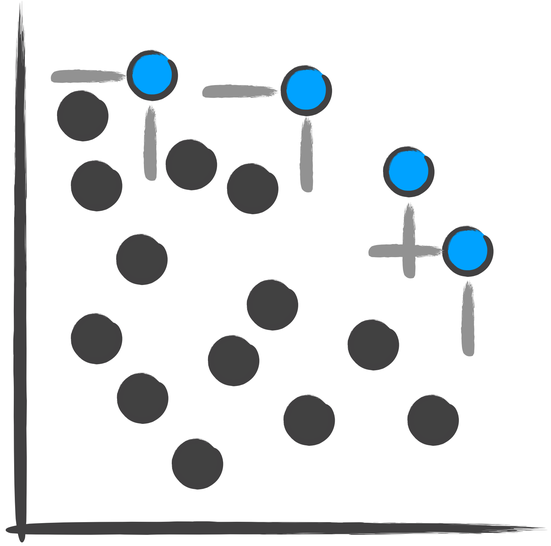
Clusters



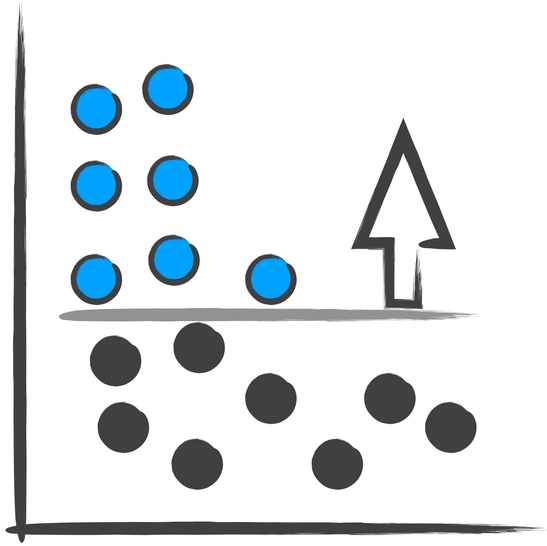
Correlation



Categories



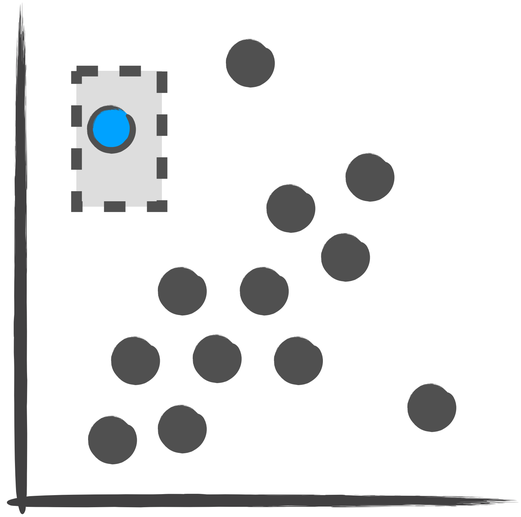
Multivariate Optimization



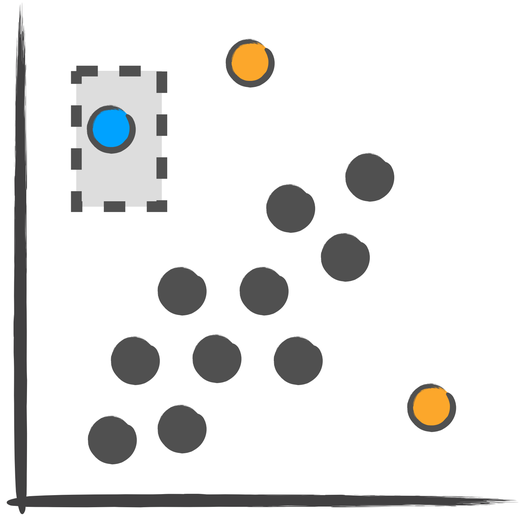
Ranges

WHY DO WE CARE?

Speed up complex selections



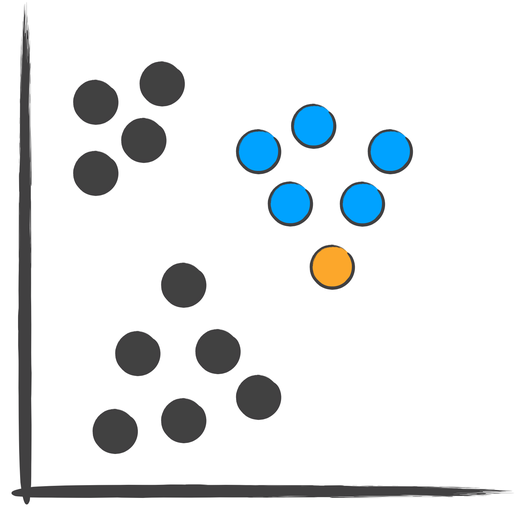
Selection



Outliers?

WHY DO WE CARE?

ID Based Selection:
Selected Elements: 7, 9, 13, 18, 22

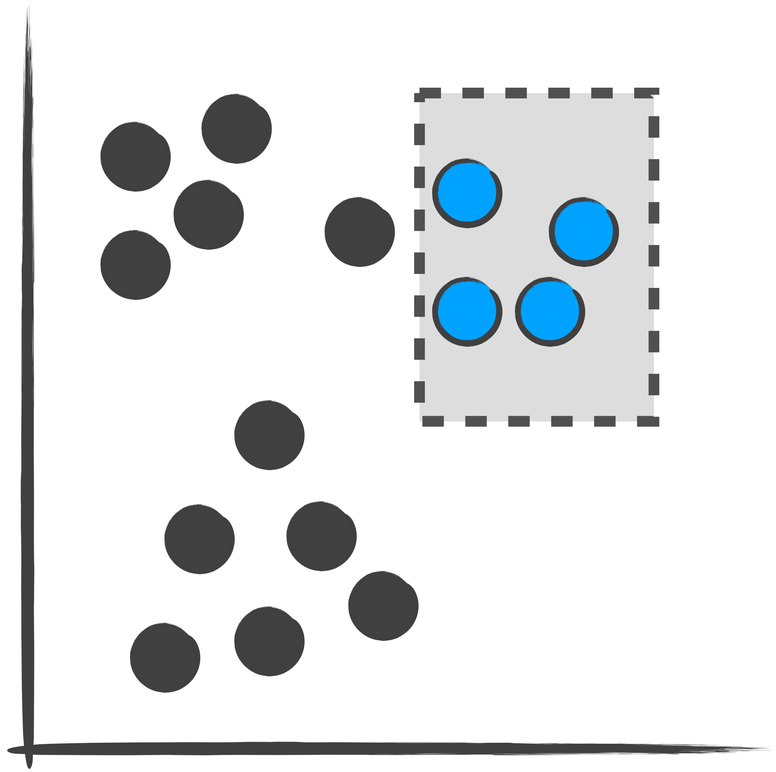


Semantic Selection:
Elements in K-Means cluster centered at [2, 3]

Meaningful, higher level concept:
improves reproducibility

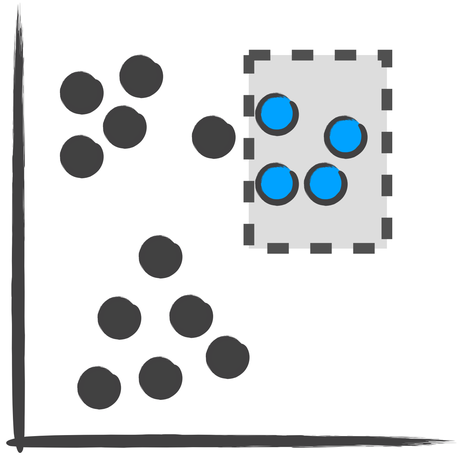
Robust to changes and updates in dataset:
enables re-usability

HOW DO WE INFER INTENT?

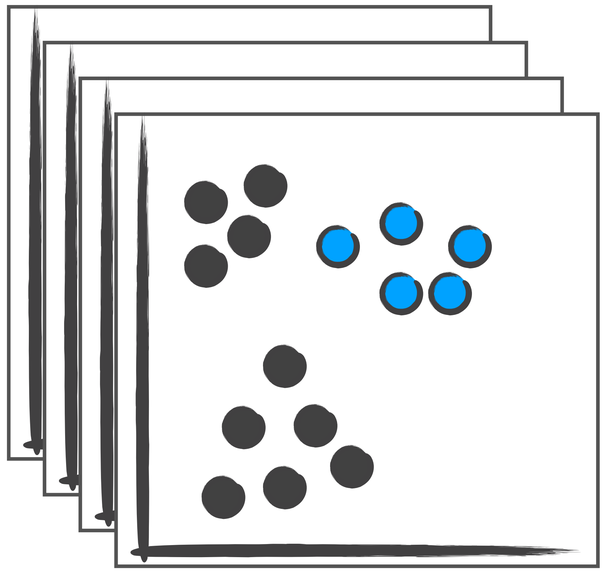


Selection

HOW DO WE INFER INTENT?



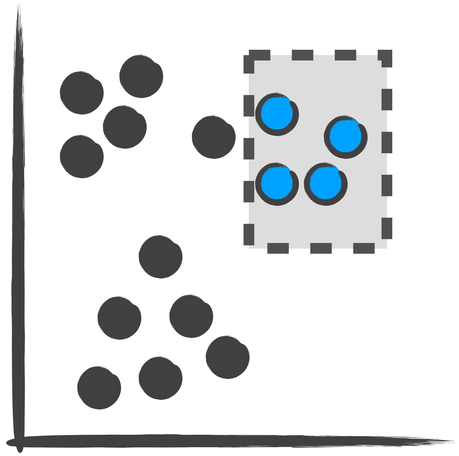
Selection



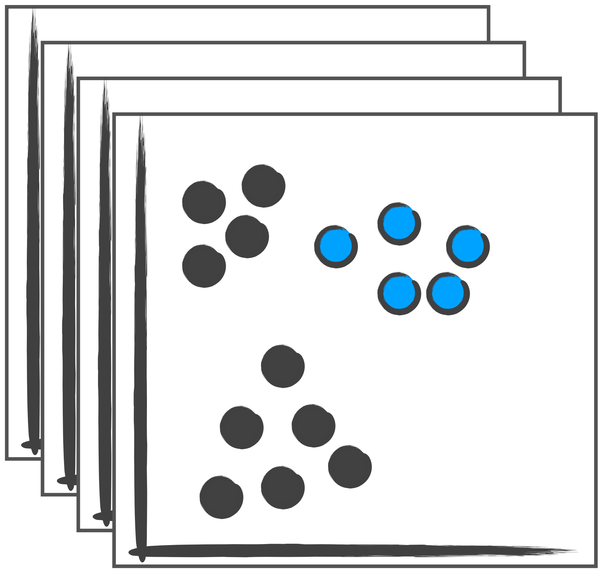
Predictions

K-Means
DBScan
Regression
Outlier Detection
Skyline
Decision Trees / Ranges
Categories

HOW DO WE INFER INTENT?



Selection



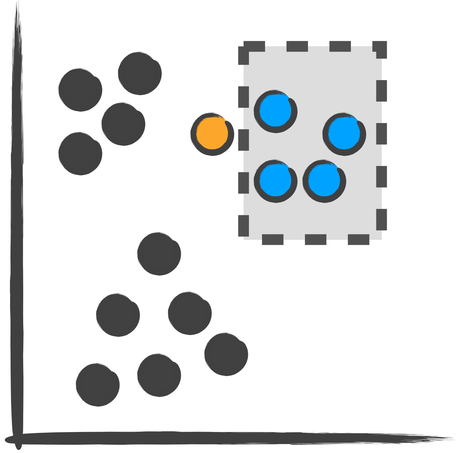
Predictions

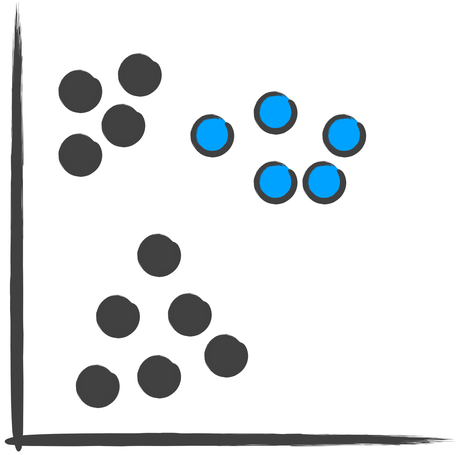
K-Means
DBScan
Regression
Outlier Detection
Skyline
Decision Trees / Ranges
Categories

1. Range
2. Cluster
3. Outlier

Ranking

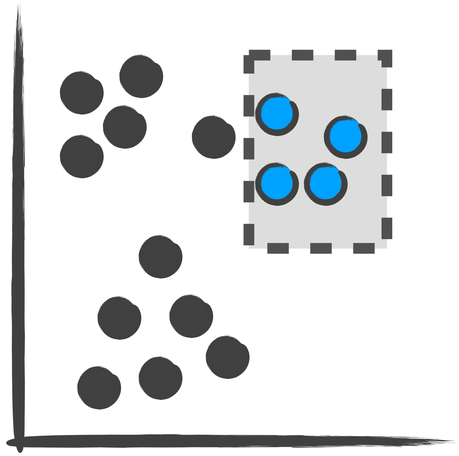
Jaccard Distance
Naive Bayes
Classifier
Heuristic
Measures



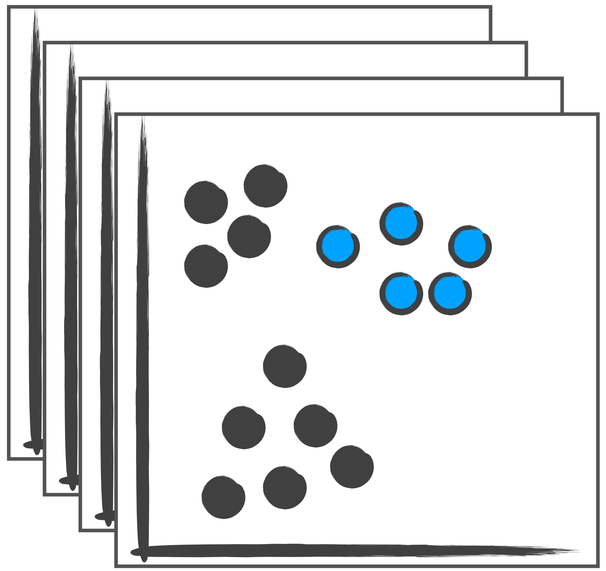


$$J(S,C) = \frac{|S \cap C|}{|S \cup C|}$$

HOW DO WE INFER INTENT?



Selection



Predictions

K-Means
DBScan
Regression
Outlier Detection
Skyline
Decision Trees / Ranges
Categories

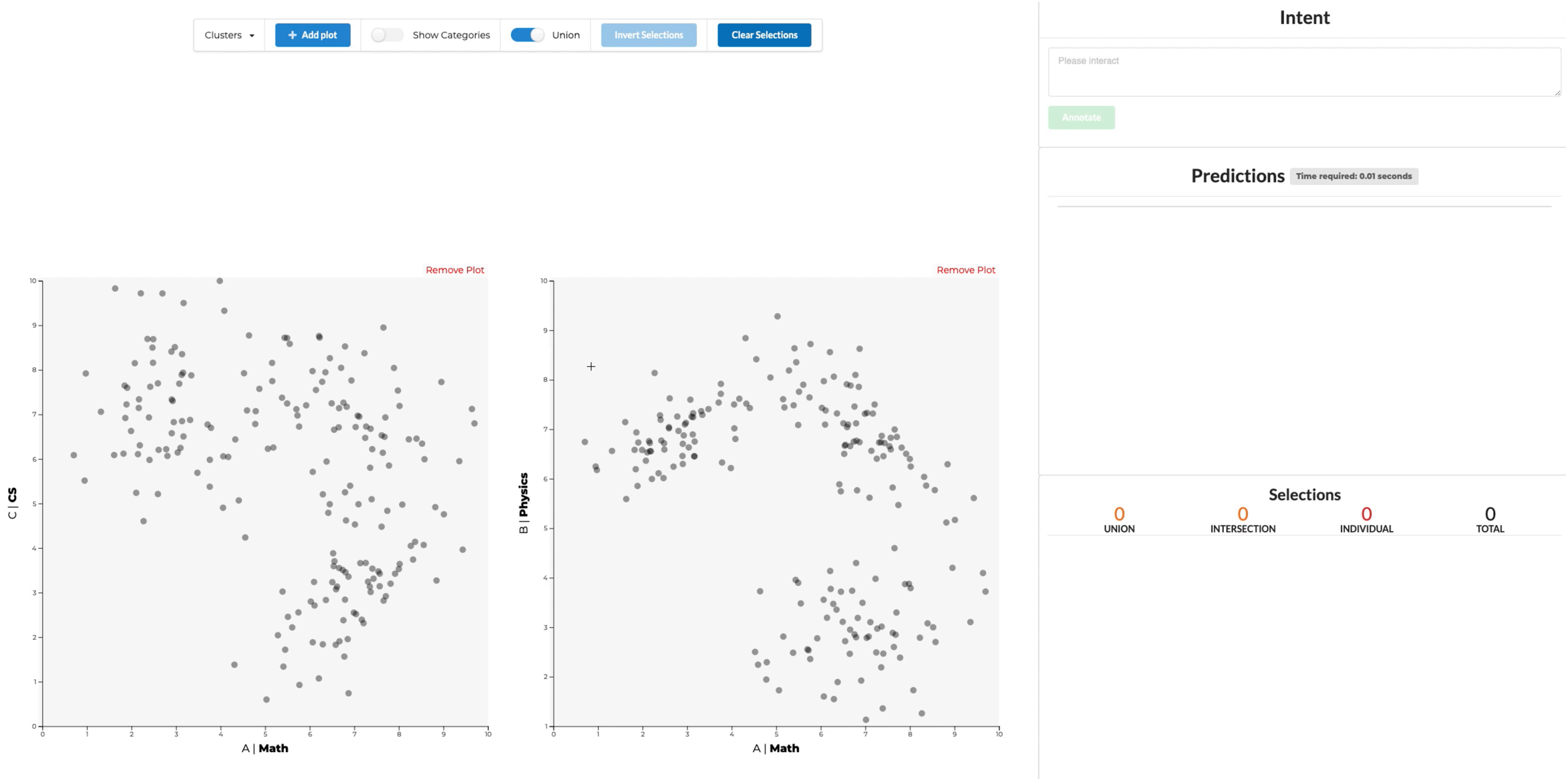
1. Range
2. Cluster
3. Outlier

I think this cluster…

Ranking

Jaccard Distance
Naive Bayes
Classifier
Heuristic
Measures

Confirming Intent & Annotation



Visualization and Selection

Annotation of Intent and Predictions

http://vdl.sci.utah.edu/predicting-intent/

TOWARDS LITERATE VISUALIZATION

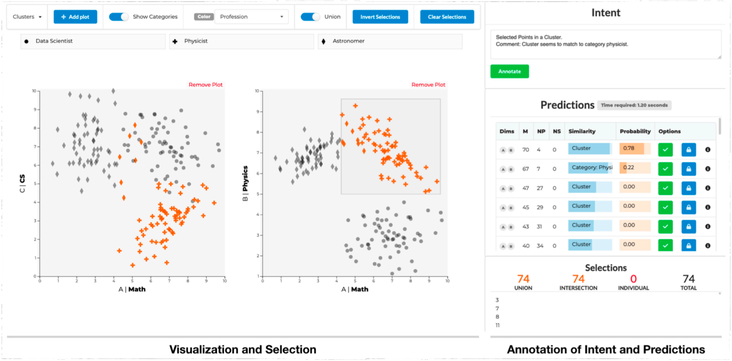
Semantic Selections & Annotations allow us to have higher-level information about events in an analysis process.

Analysts have the means to justify their choices.

Makes it easier to build analysis stories and retrieve important states.

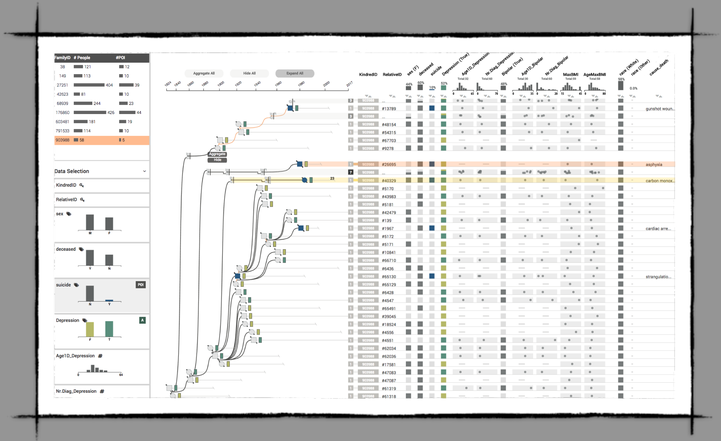
TECHNICAL CONTRIBUTIONS

Literate Visualization



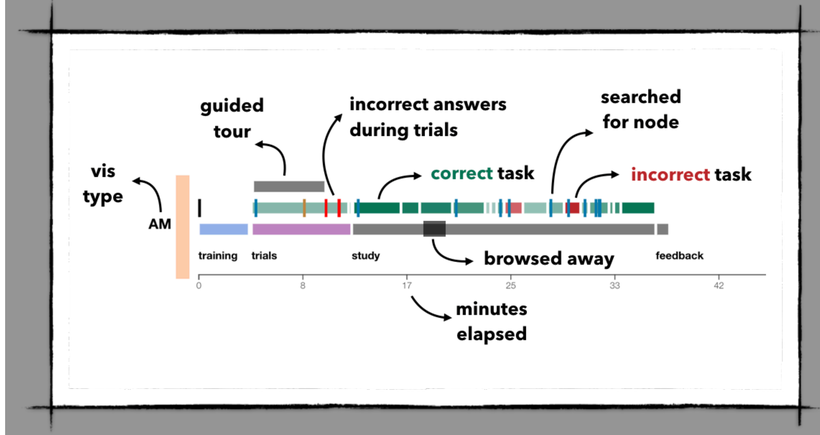
DOMAIN DRIVEN TECHNIQUES

Clinical Genealogies



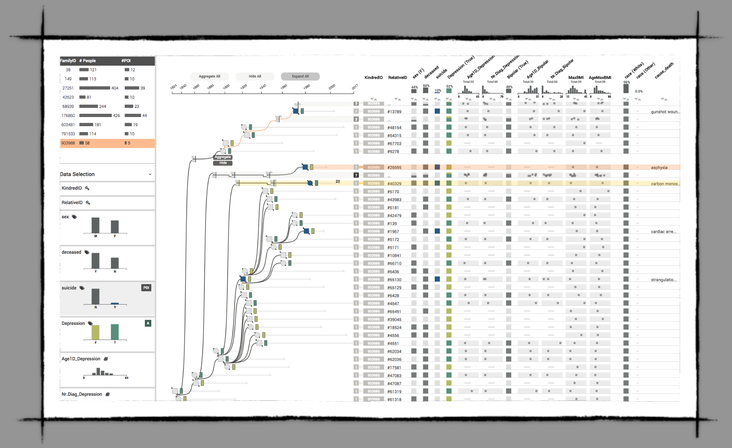
EMPIRICAL & THEORETICAL WORK

Evaluating Complex Systems



DOMAIN DRIVEN TECHNIQUES

Clinical Genealogies



LINEAGE:

VISUALIZING CLINICAL DATA IN GENEALOGY GRAPHS

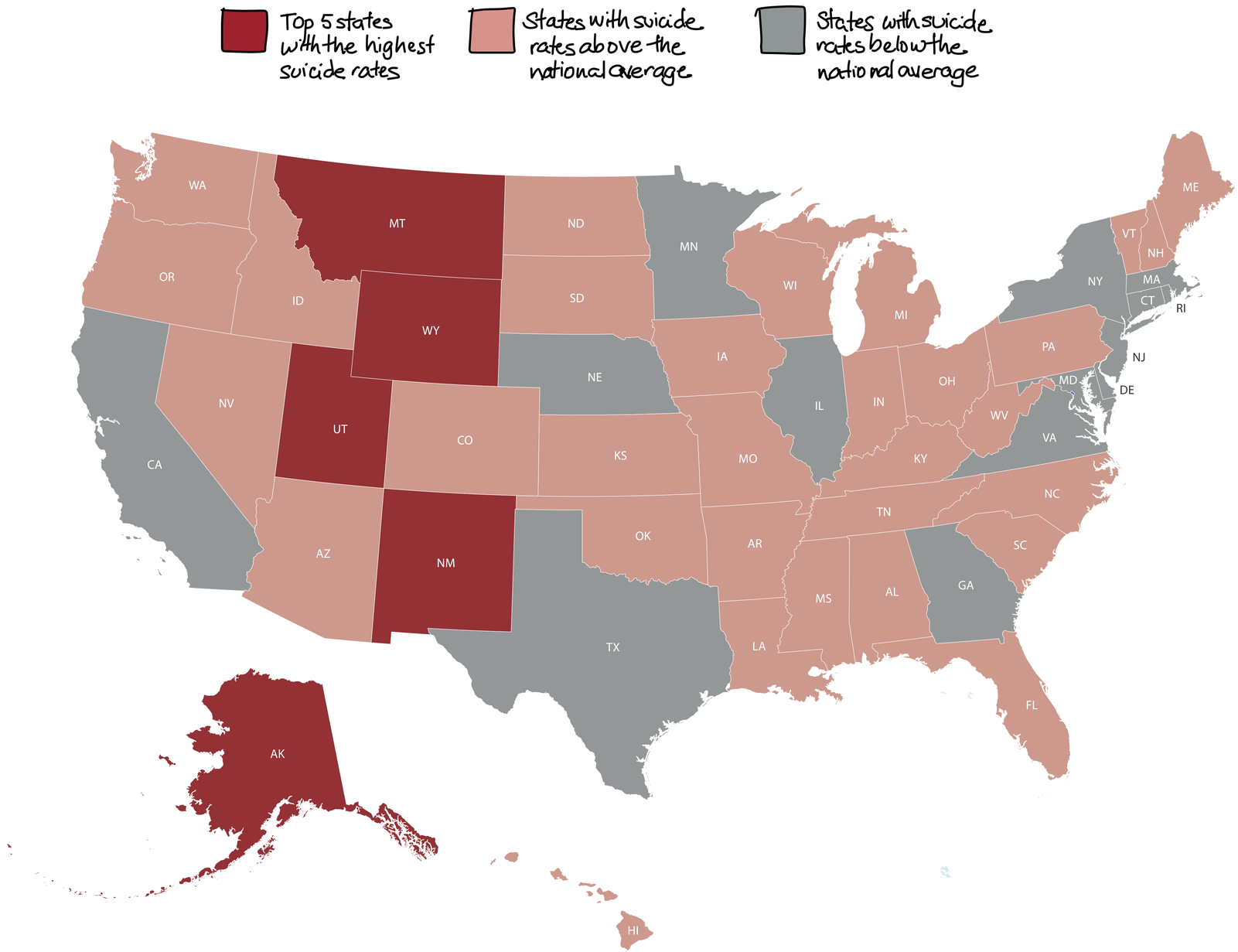
Carolina Nobre, Nils Gehlenborg, Hilary Coon, Alexander Lex


THE WORLD HEALTH
ORGANIZATION ESTIMATES
ONE PERSON DIES OF
SUICIDE EVERY 40 SECONDS

SUICIDE IS THE SECOND LEADING CAUSE OF DEATH

IN YOUTHS BETWEEN 15 AND 29 YEARS OLD

AVERAGE SUICIDE RATE BY STATE
(National average 13.26 per 100,000)

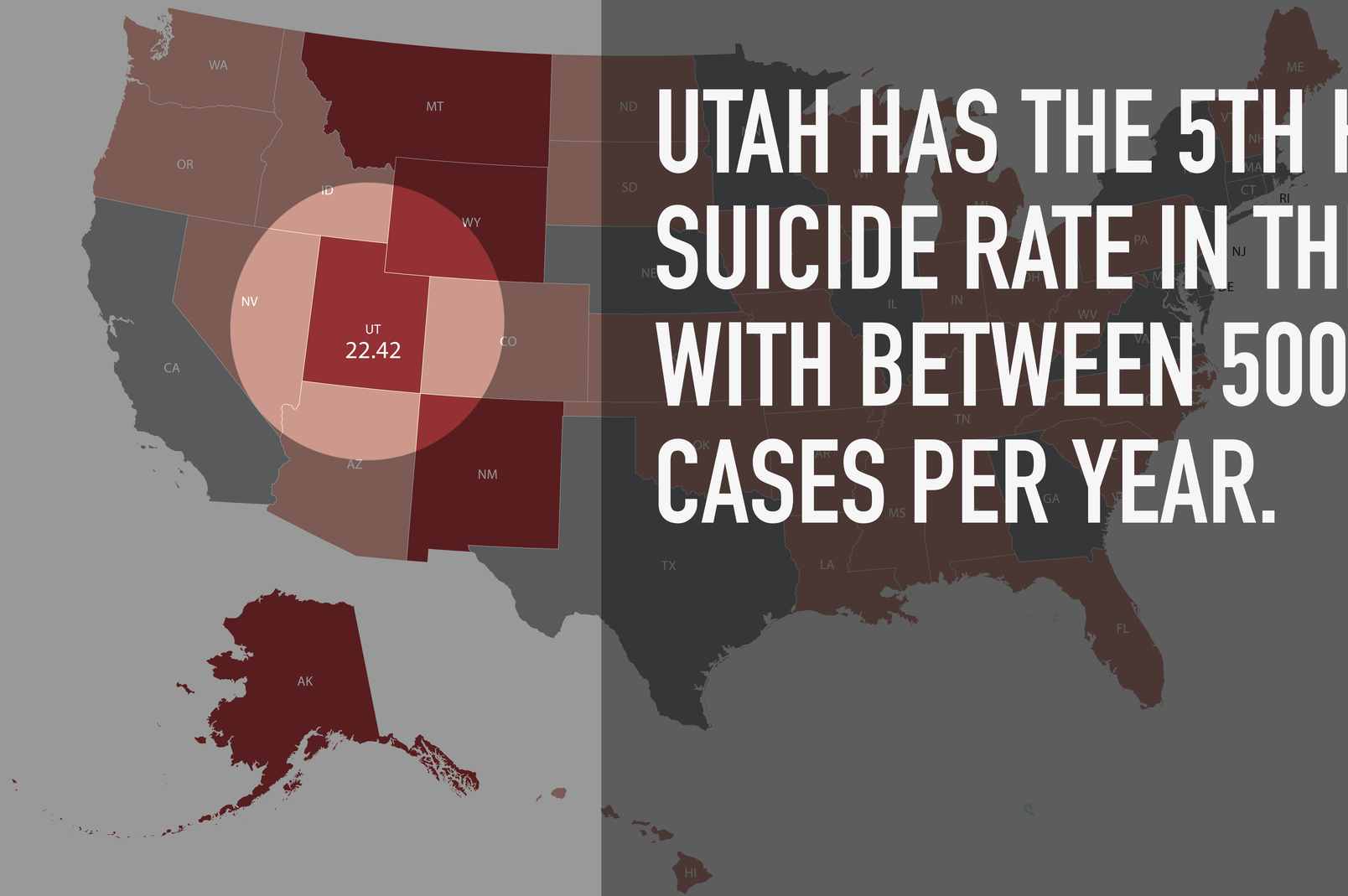


AVERAGE SUICIDE RATE BY STATE
(National average 13.26 per 100,000)

Top 5 states with the highest suicide rates

States with suicide rates above the national average

States with suicide rates below the national average



UTAH HAS THE 5TH HIGHEST SUICIDE RATE IN THE COUNTRY, WITH BETWEEN 500–600 CASES PER YEAR.

American Journal of Epidemiology

Vol. 181, No. 5
DOI: 10.1093/aje/kwu341
Advance Access publication: February 10, 2015

Original Contribution

Acute Air Pollution Exposure and Risk of Suicide Completion

er, Hilary Coon, Douglas Gray, Phillip Wilson, nshaw

rtment of Psychiatry, School of Medicine, University of Utah, 650 Komas Drive, Suite 206, @hsc.utah.edu).

for publication August 11, 2014.

OPEN

Citation: Transl Psychiatry (2013) **3**, e325; doi:10.1038/tp.2013.100

www.nature.com/tp

ORIGINAL ARTICLE

Genetic risk factors in two Utah pedigrees at high risk for suicide

H Coon[1], T Darlington[1], R Pimentel[2], KR Smith[2,3], CD Huff[4], H Hu[4], L Jerominski[1], J Hansen[1], M Klein[5], WB Callor[6], J Byrd[6], A Bakian[1], SE Crowell[1,7], WM McMahon[1], V Rajamanickam[8], NJ Camp[8], E McGlade[1,9], D Yurgelun-Todd[1,9], T Grey[6] and D Gray[1,9]

We have used unique population-based data re
over twice that expected from demographically
two high-risk pedigrees. In the first of these (pe
death was 30.95. In the second (pedigree 5), 7/
decedents in pedigree 12 and nine in pedigree
analyzed using the Variant Annotation, Analysis,
functional impact of the DNA variation, aggrega
prioritized variants that were: (1) shared across
(3) ⩽ 5% in genotyping data from 398 other Ut
from 1358 controls and/or in dbSNP. Results inc
FAM38A and *HRCT1* for pedigree 5). Other gene

The Role of Social Isolation in Suicide

Deborah L. Trout M.A.

First published: Spring 1980 | https://doi-org.ezproxy.lib.utah.edu/10.1111/j.1943-278X.1980.tb00693.x | Cited by: 95

The author wishes to thank Dr. Charles Neuringer for his assistance with the preparation of this manuscript.

MOTIVATION & DATA

Understand **Complex Conditions**

Discover **Genetic Risk Factors**

Dataset:

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

Based on **Utah Population Database**

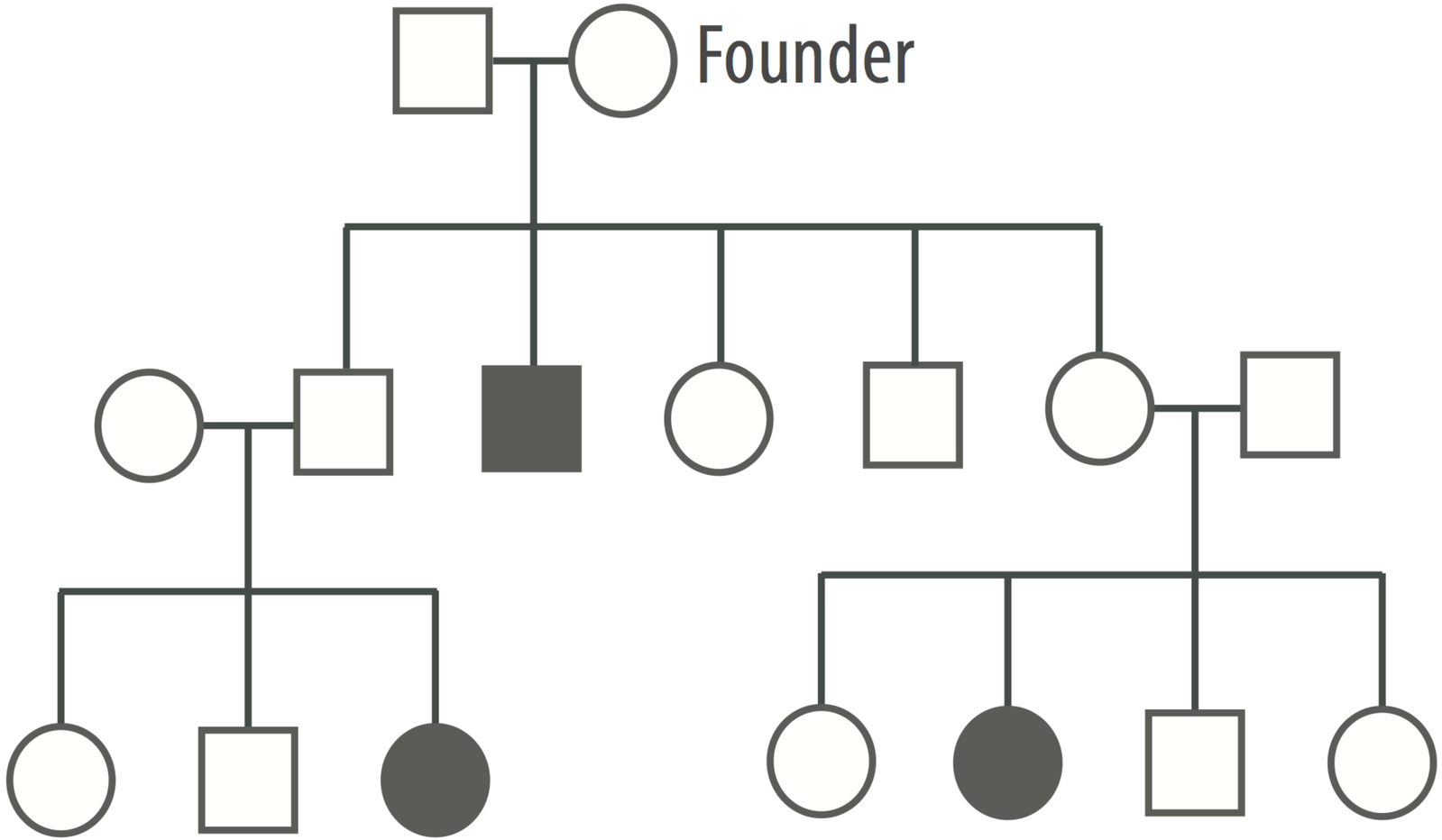
SPECIFIC GOALS

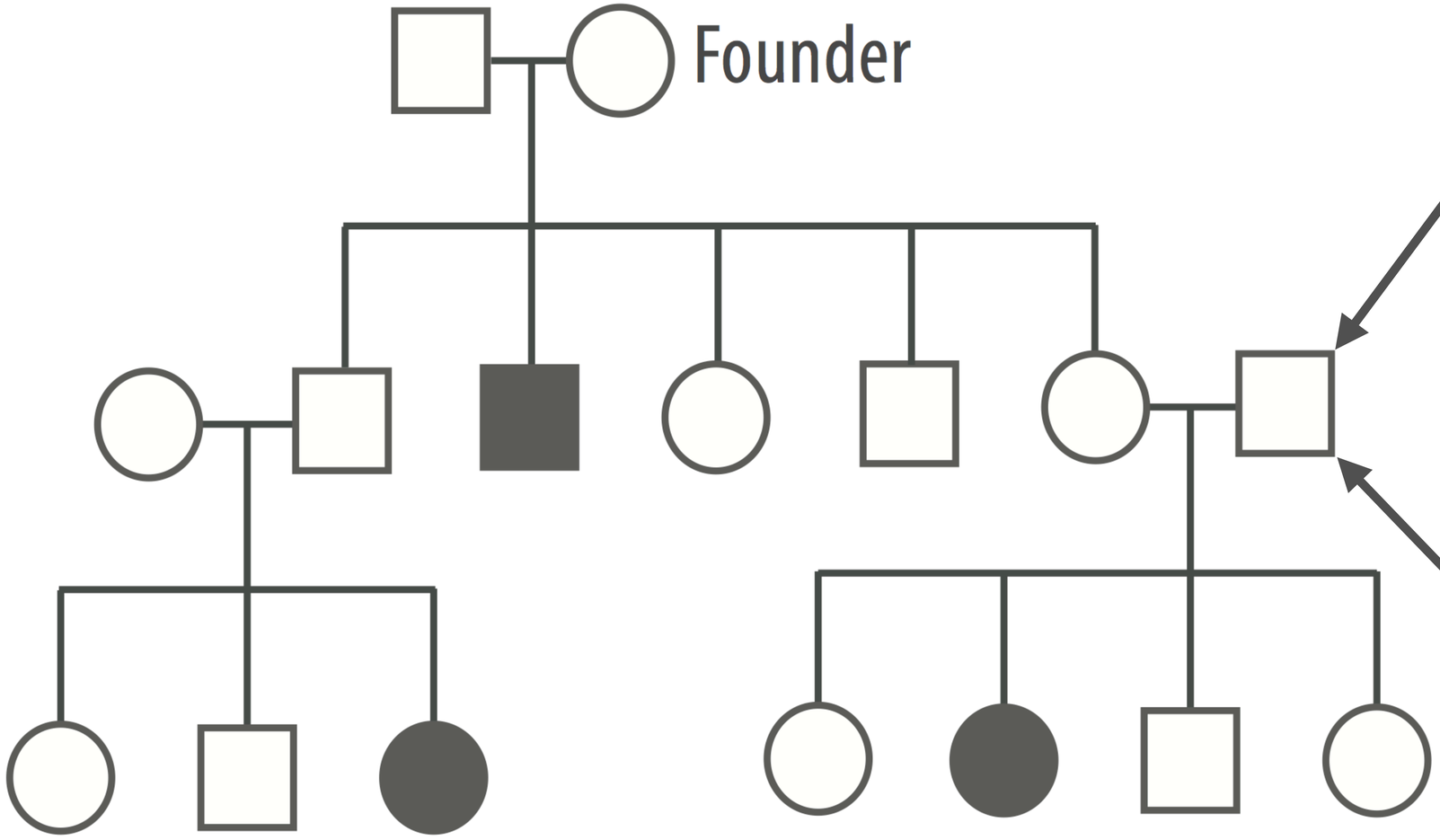
Find familial cases that also have an “interesting” phenotype

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

Prioritize those cases for analysis of shared genomic sequences

Proofread the Data!


Founder



Age

Sex

Race

Bipolar

Depression

Asthma

Obesity

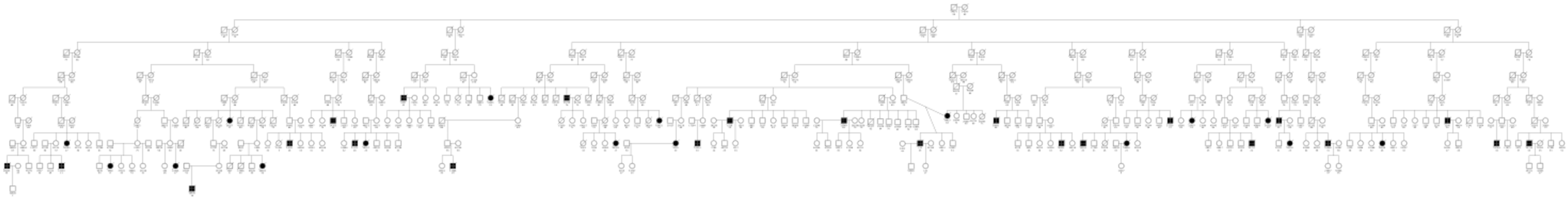
Schizophrenia

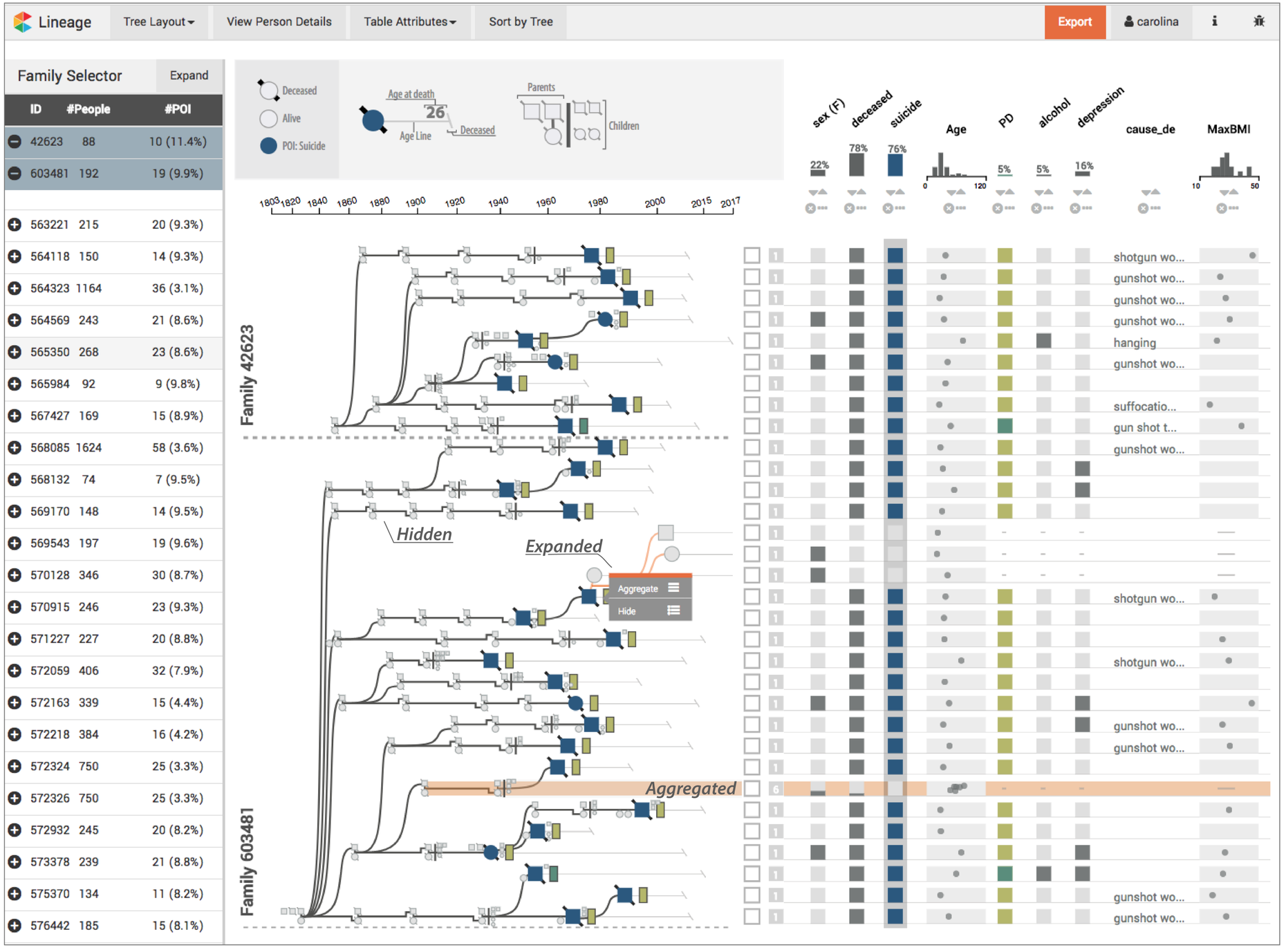
Cause of Death

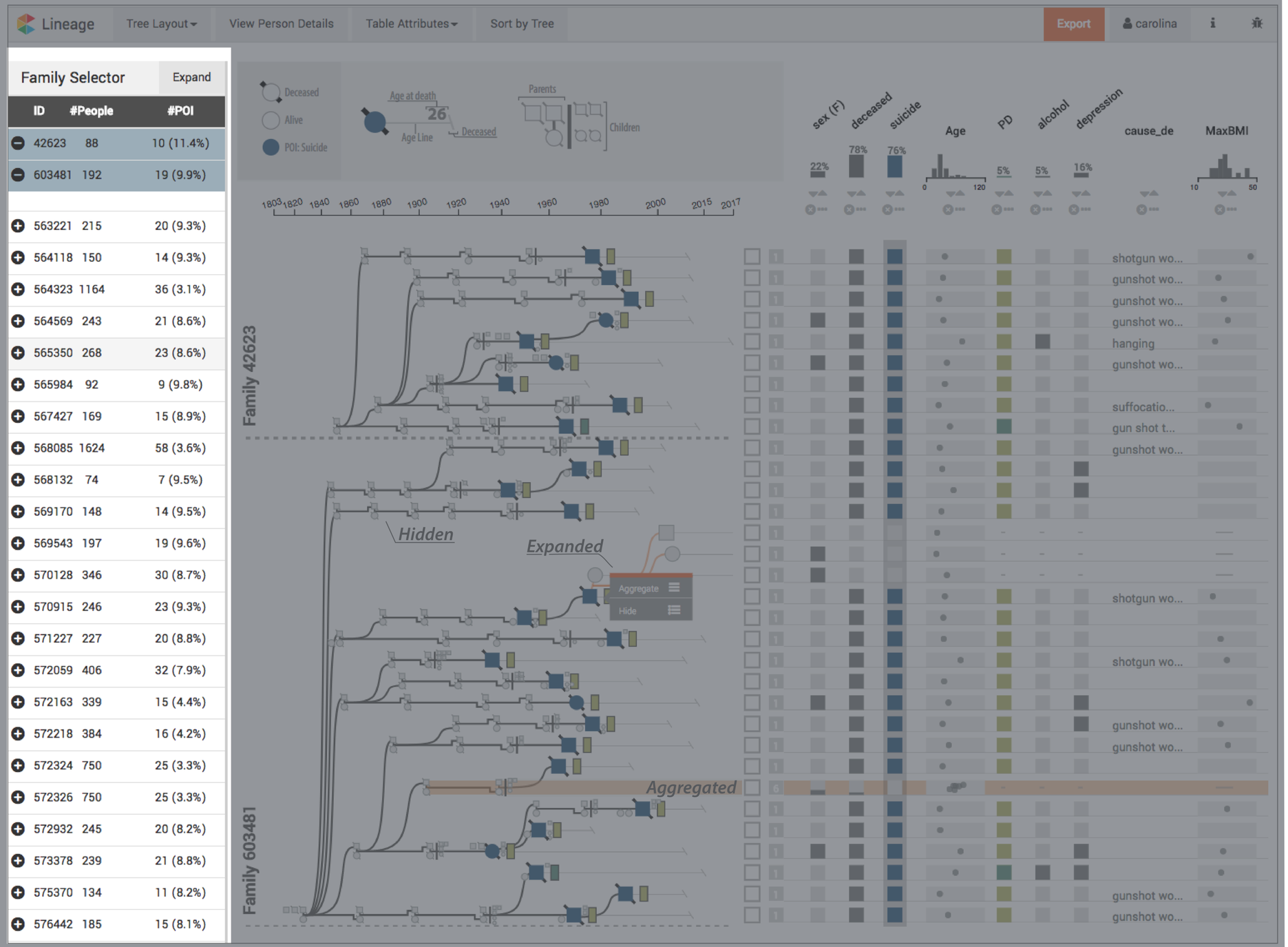
Weapon Used

...

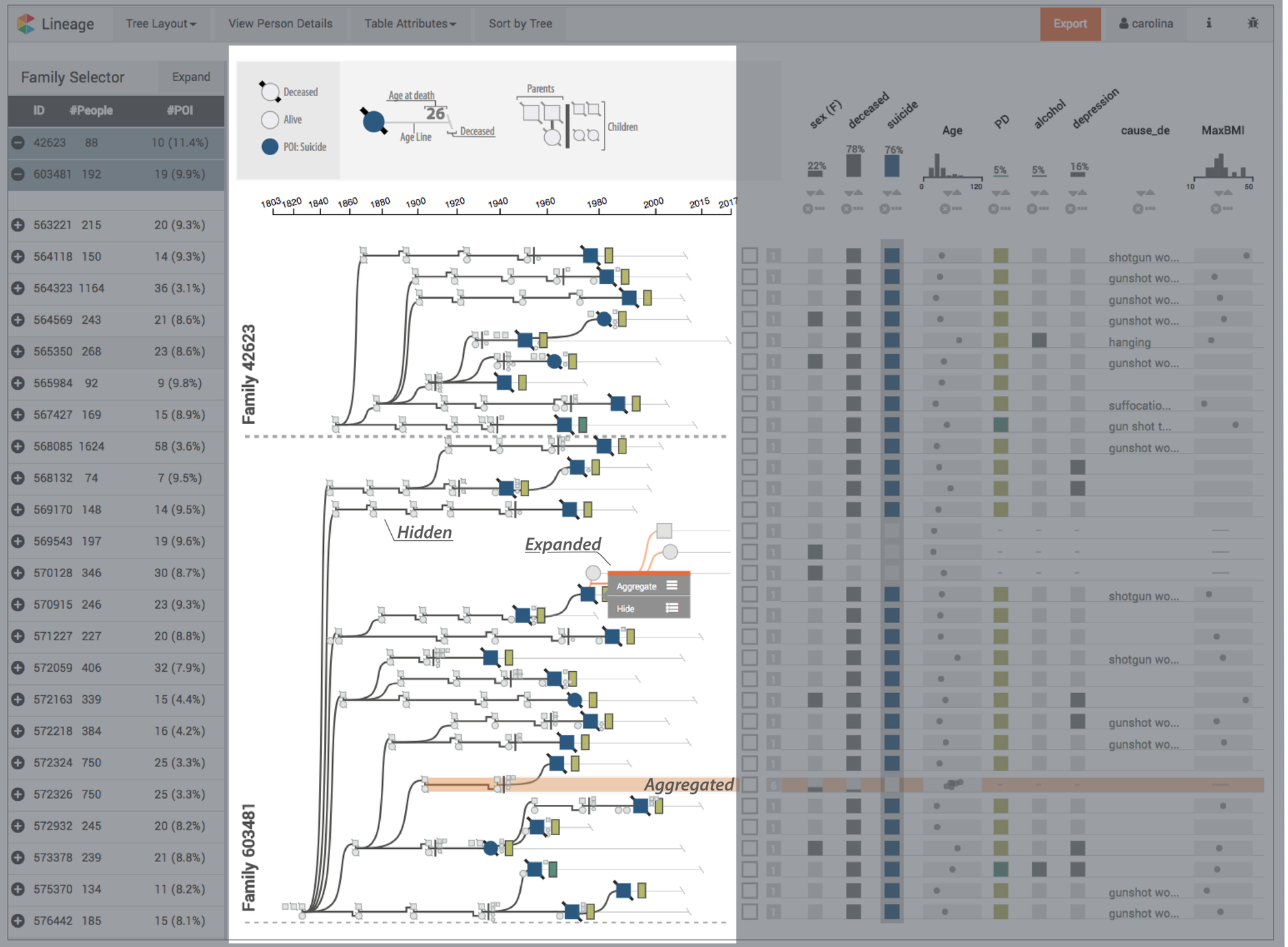
GENEALOGY WITH ~400 MEMBERS RENDERED WITH PROGENY




Lineage
Tree Layout
View Person Details
Table Attributes
Sort by Tree
Export
carolina
Family Selector
Expand
ID #People #POI
42623 88 10 (11.4%)
603481 192 19 (9.9%)
563221 215 20 (9.3%)
564118 150 14 (9.3%)
564323 1164 36 (3.1%)
564569 243 21 (8.6%)
565350 268 23 (8.6%)
565984 92 9 (9.8%)
567427 169 15 (8.9%)
568085 1624 58 (3.6%)
568132 74 7 (9.5%)
569170 148 14 (9.5%)
569543 197 19 (9.6%)
570128 346 30 (8.7%)
570915 246 23 (9.3%)
571227 227 20 (8.8%)
572059 406 32 (7.9%)
572163 339 15 (4.4%)
572218 384 16 (4.2%)
572324 750 25 (3.3%)
572326 750 25 (3.3%)
572932 245 20 (8.2%)
573378 239 21 (8.8%)
575370 134 11 (8.2%)
576442 185 15 (8.1%)
Deceased
Alive
POI: Suicide
Age at death
26
Age Line
Deceased
Parents
Children
sex (F)
deceased
suicide
Age
PD
alcohol
depression
cause_de
MaxBMI
22%
78%
76%
5%
5%
16%
0
120
10
50
1803 1820 1840 1860 1880 1900 1920 1940 1960 1980 2000 2015 2017
Family 42623
Family 603481
Hidden
Expanded
Aggregate
Hide
Aggregated
shotgun wo...
gunshot wo...
gunshot wo...
gunshot wo...
hanging
gunshot wo...
suffocatio...
gun shot t...
gunshot wo...
shotgun wo...
shotgun wo...
gunshot wo...
gunshot wo...
gunshot wo...
gunshot wo...

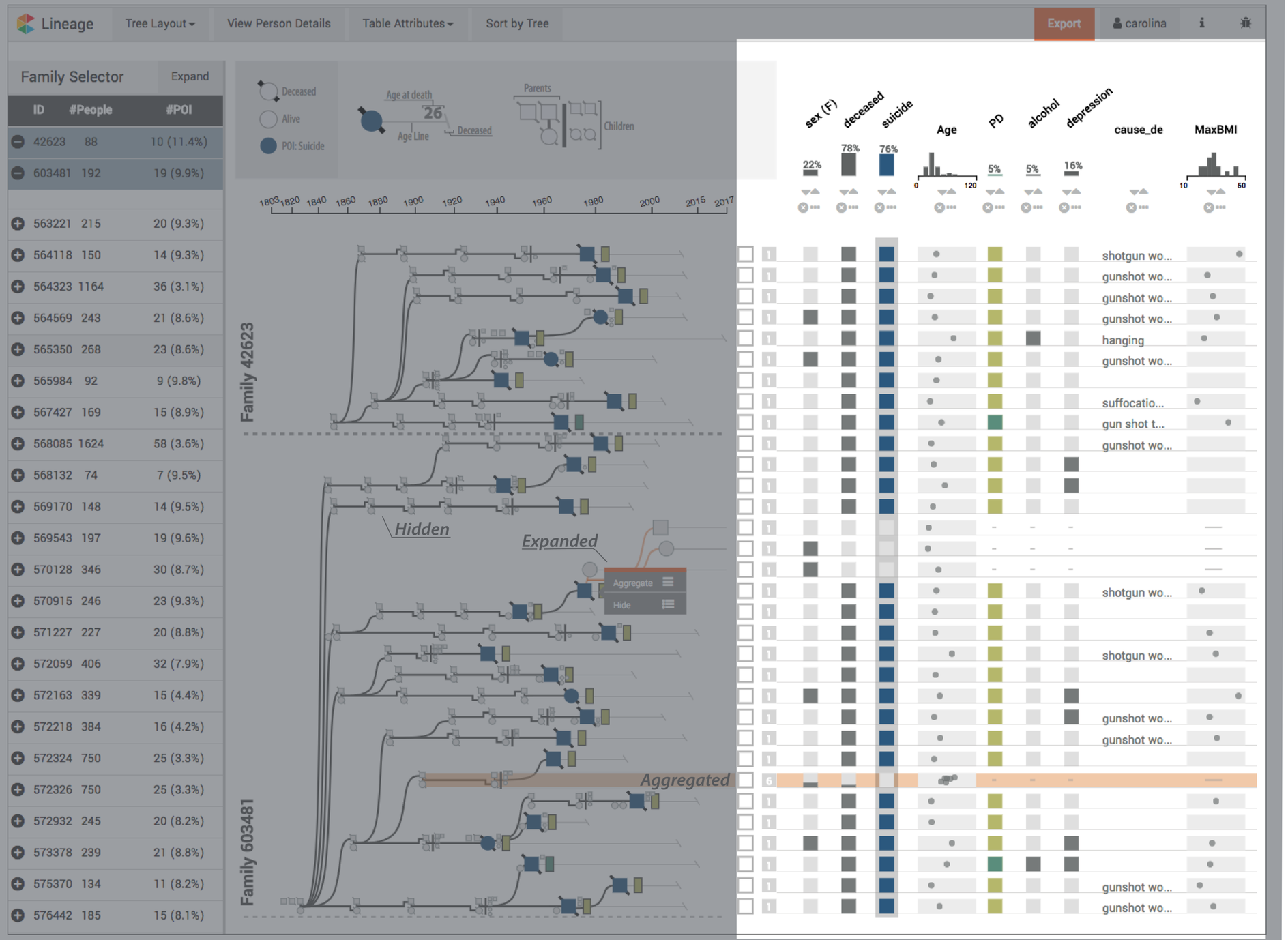


Family Selector



Family Selector

Pedigree Visualization

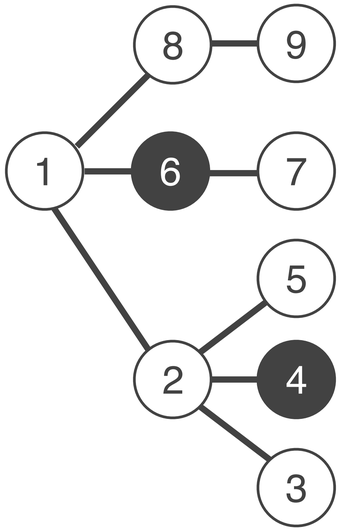


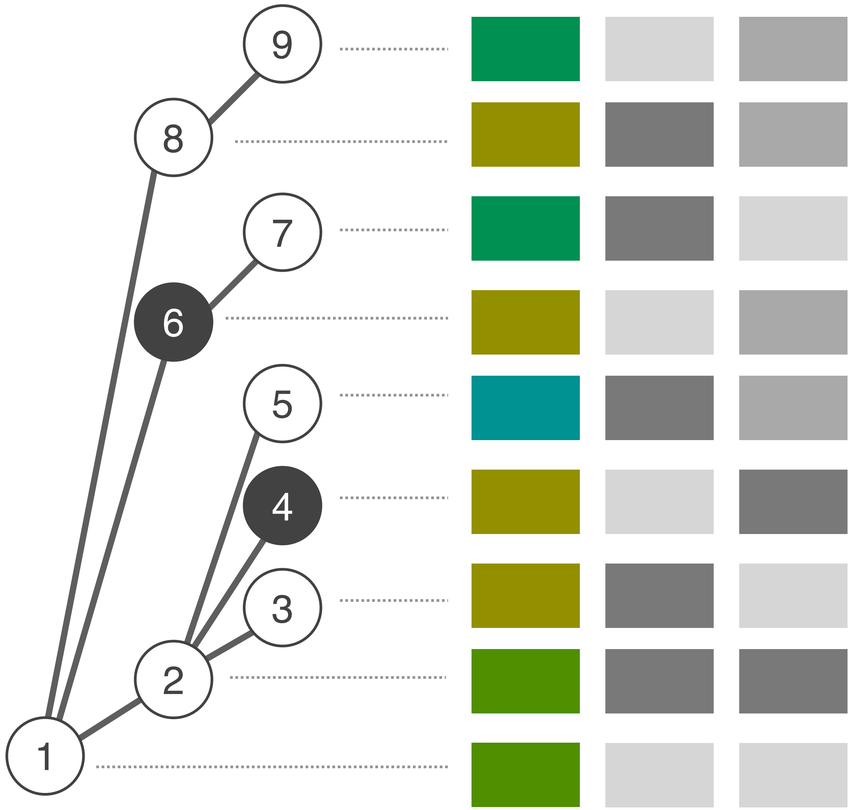
Family Selector

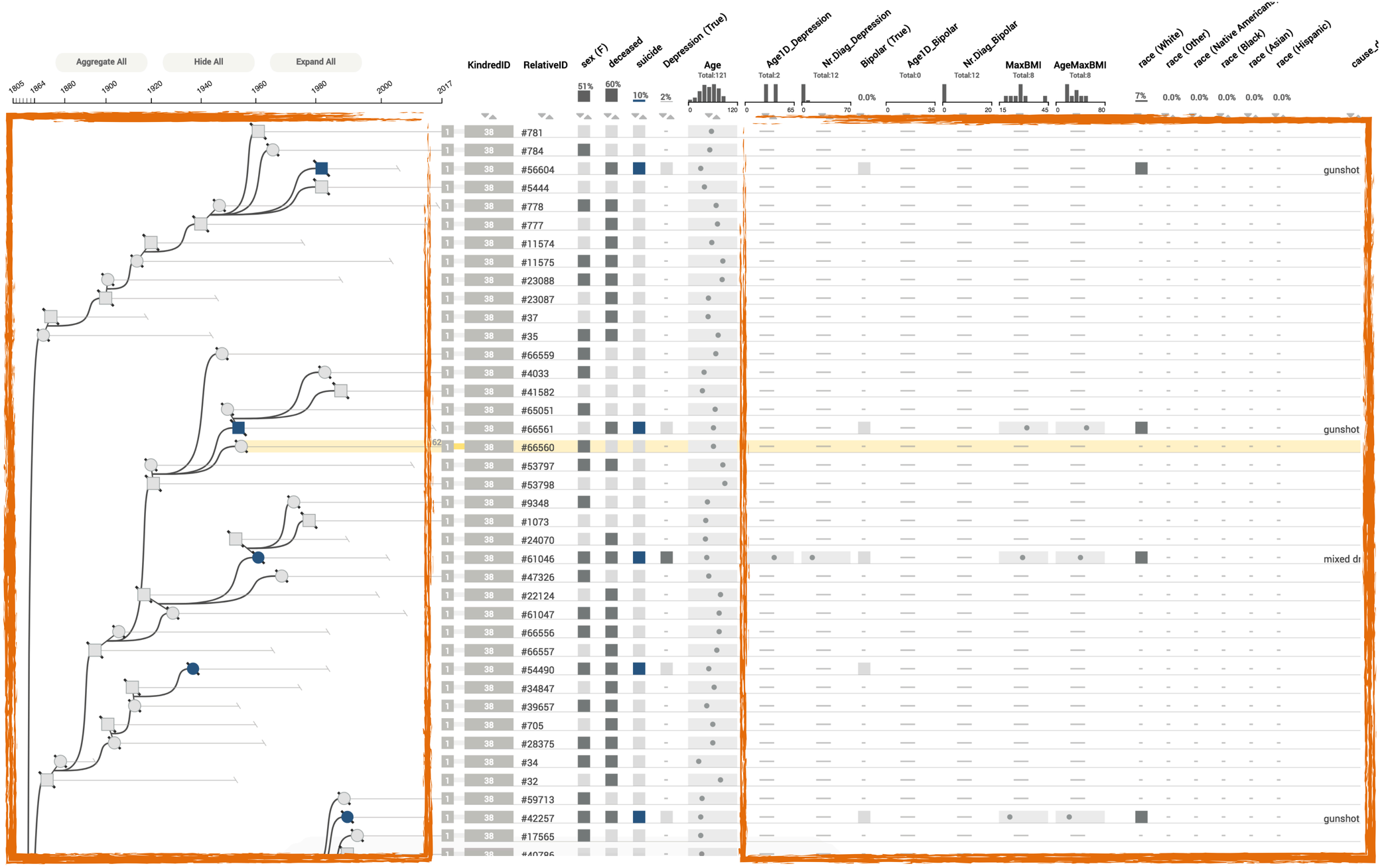
Pedigree Visualization

Attribute Table

LINEARIZING

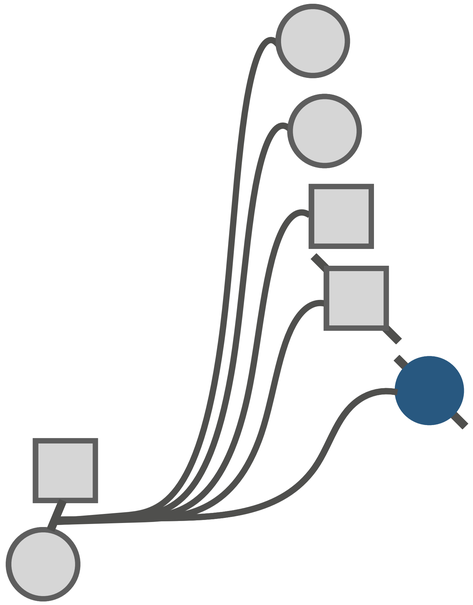


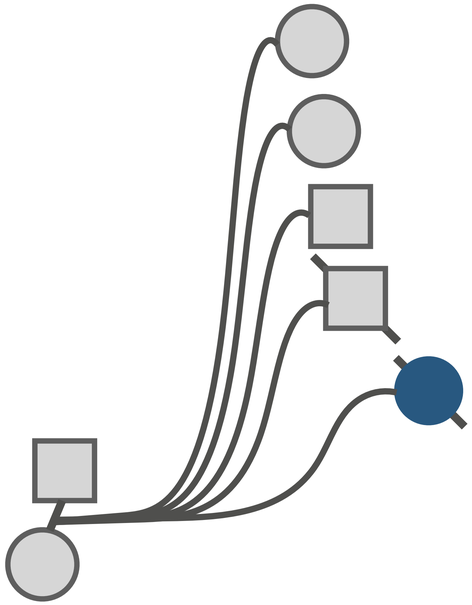


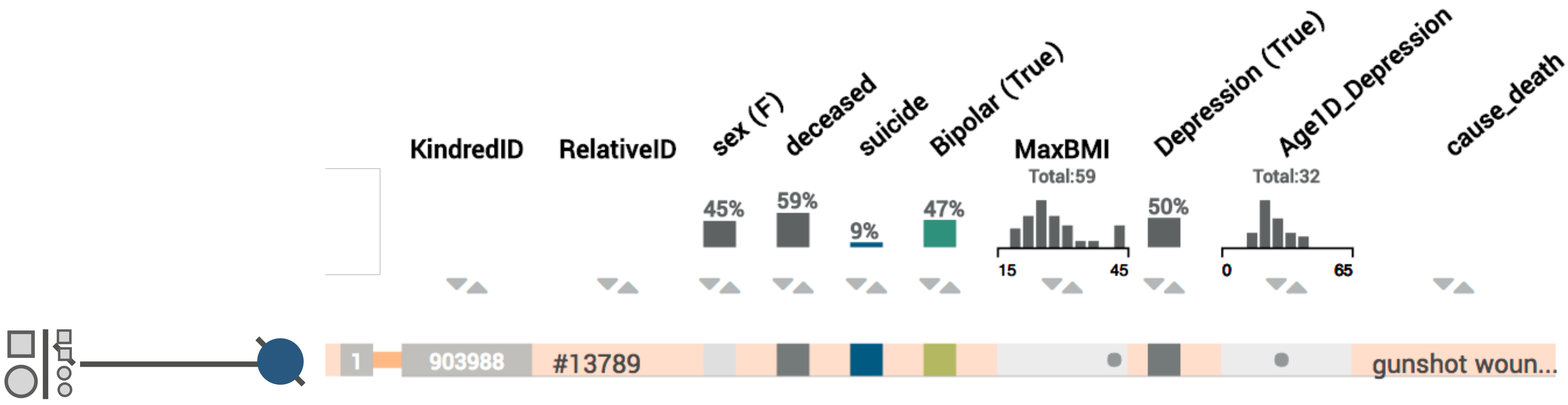


Can’t show many people

Lots of missing data



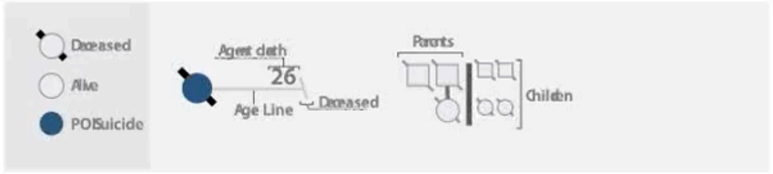


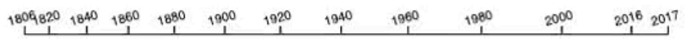

KindredID
RelativeID
sex (F)
deceased
suicide
Bipolar (True)
MaxBMI
Total:59
Depression (True)
Age1D_Depression
Total:32
cause_death
45%
59%
9%
47%
50%
15
45
0
65
1
903988
#13789
gunshot woun...

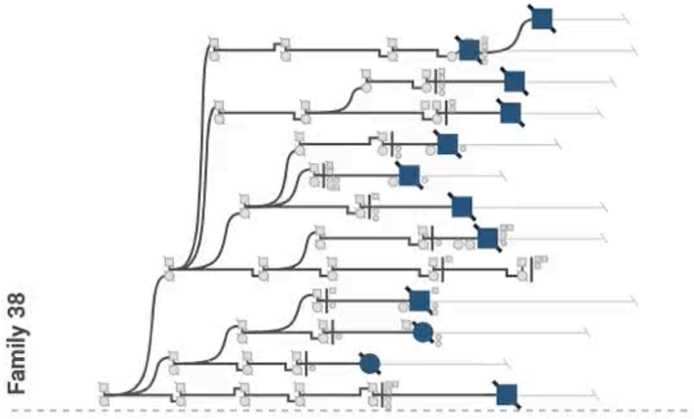
Lineage | Tree Layout | View Person Details | Table Attributes | Sort by Tree

Family Selector | Expand

ID	#People	#POI
38	129	12 (9.3%)
38	129	12 (9.3%)
149	122	10 (8.2%)
212	368	35 (9.5%)
234	145	13 (9%)
331	389	19 (4.9%)
490	277	25 (9%)
709	285	28 (9.8%)
1253	245	23 (9.4%)
1474	237	18 (7.6%)
1627	171	17 (9.9%)
1787	768	29 (3.8%)
1843	218	18 (8.3%)
1881	196	19 (9.7%)
2053	322	31 (9.6%)
2082	209	21 (10%)
2117	190	17 (8.9%)
2194	164	13 (7.9%)
2208	162	13 (8%)
2563	157	15 (9.6%)
2749	230	19 (8.3%)
2902	224	17 (7.6%)
3222	191	20 (10.5%)
3841	313	15 (4.8%)
3933	198	8 (4%)







KindredI	Relative	sex (F)	deceased	suicide	Age	LabID	bipolar spectrum ill...	anxiety-non-trauma	alcohol	PD	psychosis	depression	cause_de
		17%	100%	100%	0 120		0.0%	0.0%	8%	17%	0.0%	17%	
1	38	#18561684				#89343							gunshot wo...
1	38	#18561686				#93380							gunshot wo...
1	38	#15638474											gunshot wo...
1	38	#15512738											asphyxia
1	38	#16324365											
1	38	#15188353											
1	38	#28016042											carbon mon...
1	38	#21982847											asphyxia
1	38	#1307826				#96031							gunshot wo...
1	38	#26200382											mixed drug...
1	38	#16881308											
1	38	#15312962				#48318							gunshot wo...

USAGE & FUTURE WORK

Currently used by team of Psychiatry researchers on a daily basis

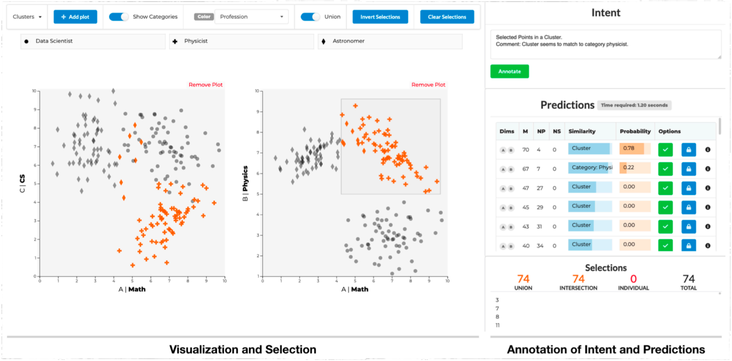
Widespread interest from other labs working with UPDB data

Integration of other data types

Geospatial, Environmental, Genomic

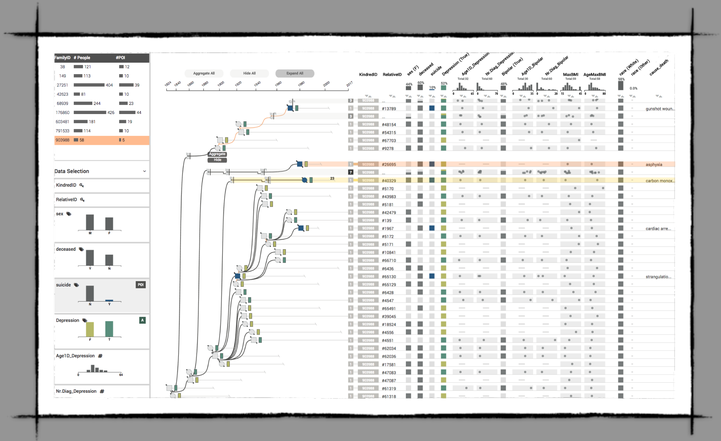
TECHNICAL CONTRIBUTIONS

Literate Visualization



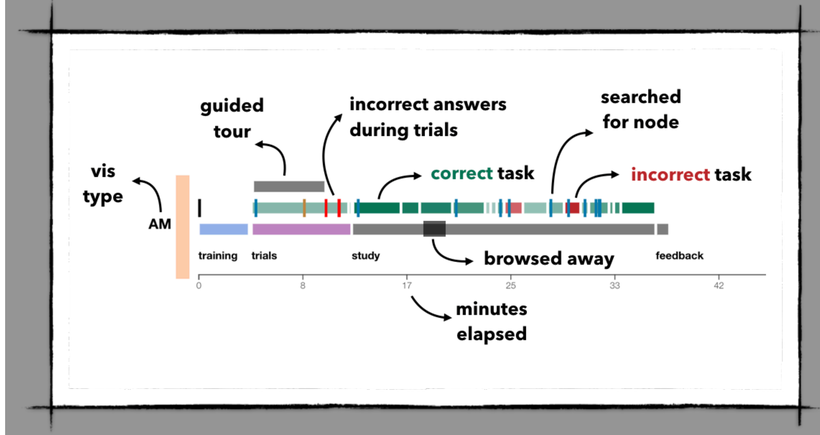
DOMAIN DRIVEN TECHNIQUES

Clinical Genealogies



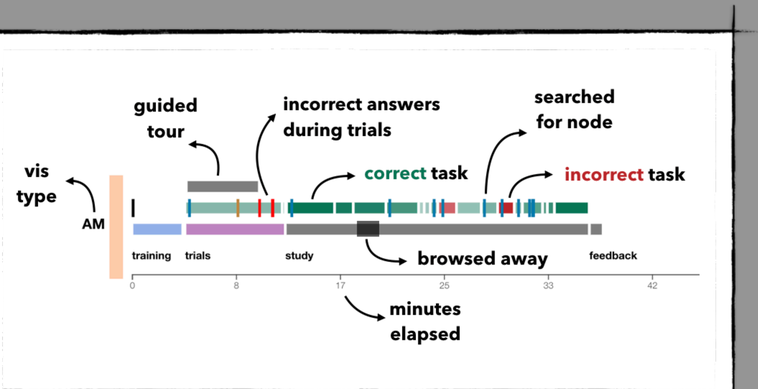
EMPIRICAL & THEORETICAL WORK

Evaluating Complex Systems



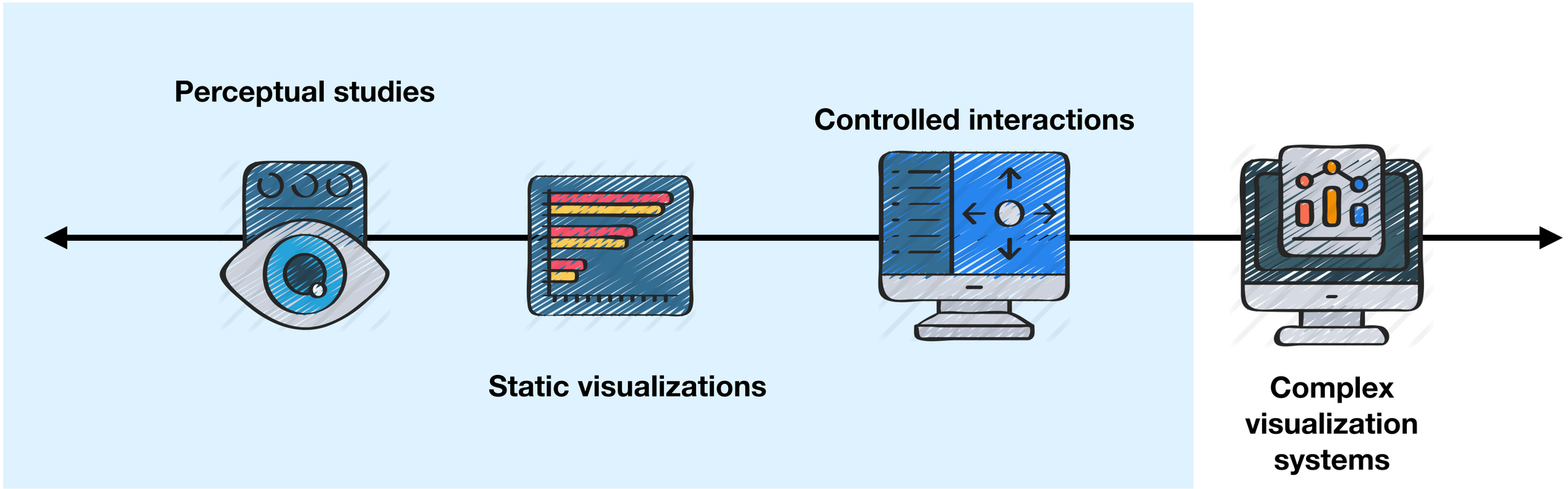
EMPIRICAL & THEORETICAL WORK

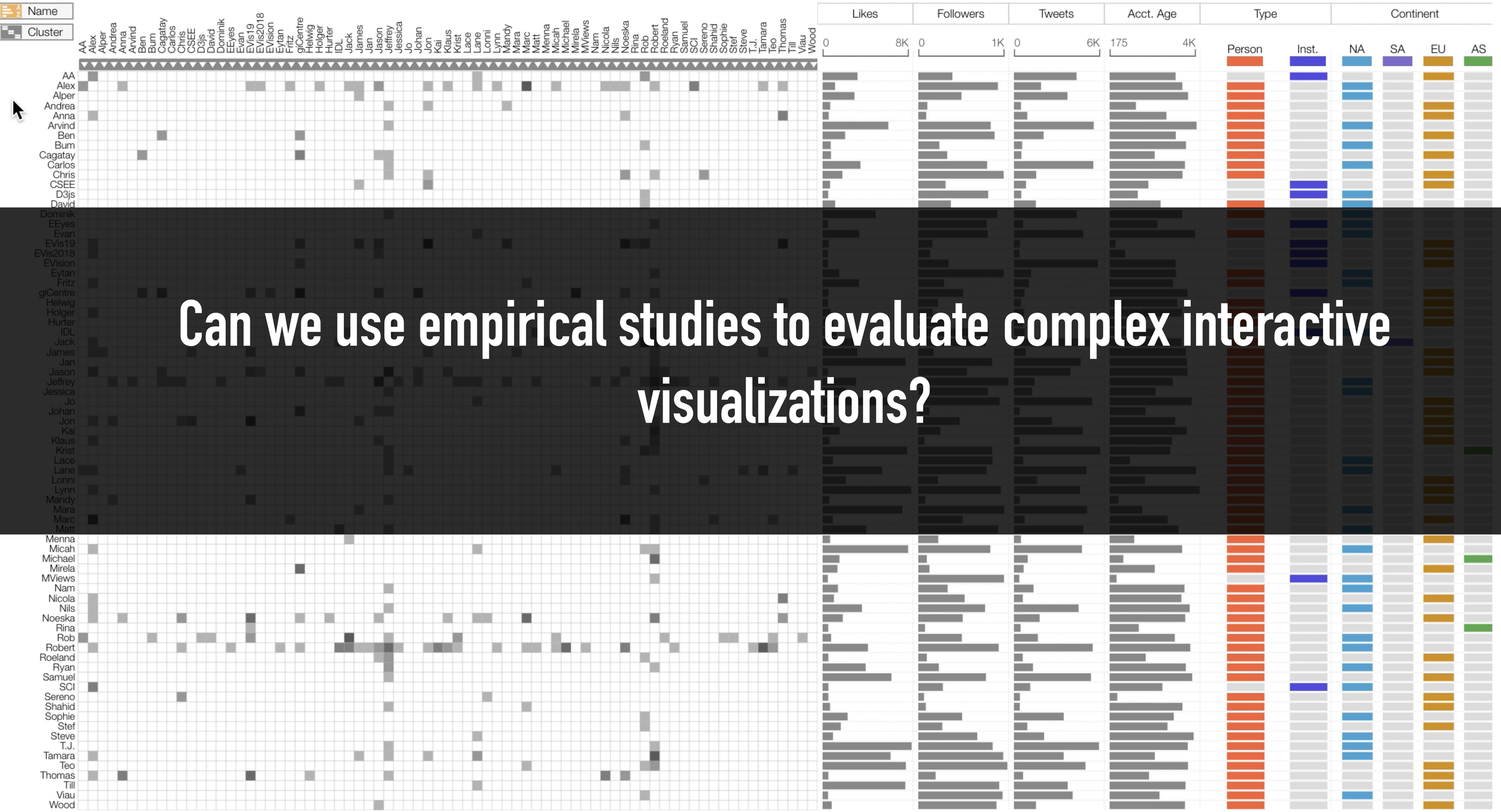
Evaluating Complex Systems



Empirically Evaluating Complex Interactive Visualization Techniques

Carolina Nobre, Dylan Wootton, Lane Harrison


Perceptual studies
Controlled interactions
Static visualizations
Complex visualization systems



Can we use empirical studies to evaluate complex interactive visualizations?

EUROVIS 2019
R. S. Laramee, S. Oeltze, and M. Sedlmair
(Guest Editors)

Volume 38 (2019), Number 3
STAR – State of The Art Report

The State of the Art in Visualizing Multivariate Networks

C. Nobre[1], M. Meyer[1], M. Streit[2], and A. Lex[1]

[1]University of Utah, Utah, USA
[2]Johannes Kepler University Linz, Austria

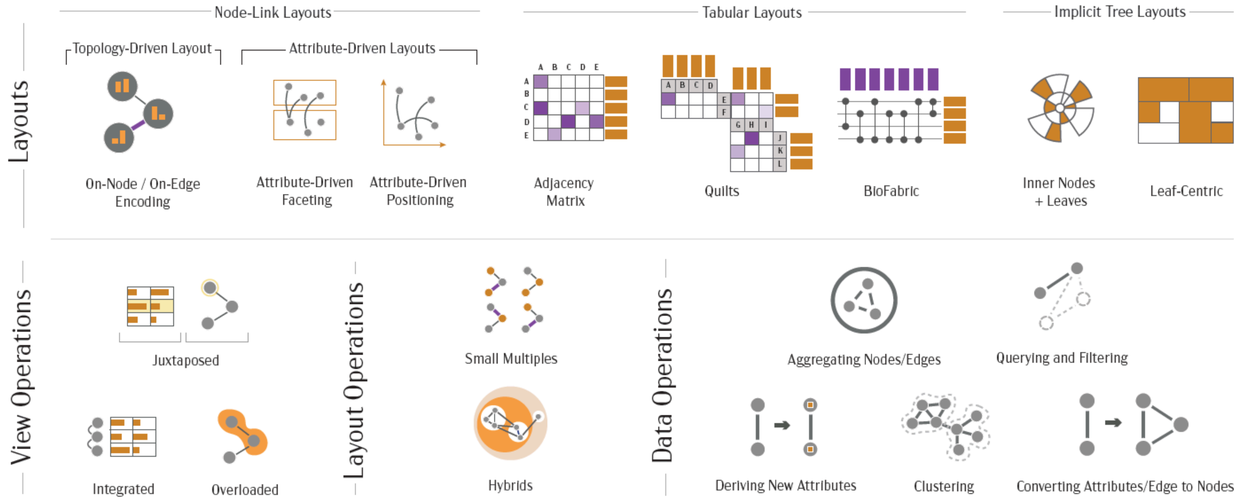


Figure 1: *A typology of operations and layouts used in multivariate network visualization.* ***Layouts*** *describe the fundamental choices for encoding multivariate networks.* ***View Operations*** *capture how topology and attribute focused visualizations can be combined.* ***Layout Operations*** *are applied to basic layouts to create specific visualization techniques.* ***Data Operations*** *are used to transform a network or derive attributes before visualizations. The colors reflect node attributes (orange), edge attributes (purple), and topology (grey).*

Abstract

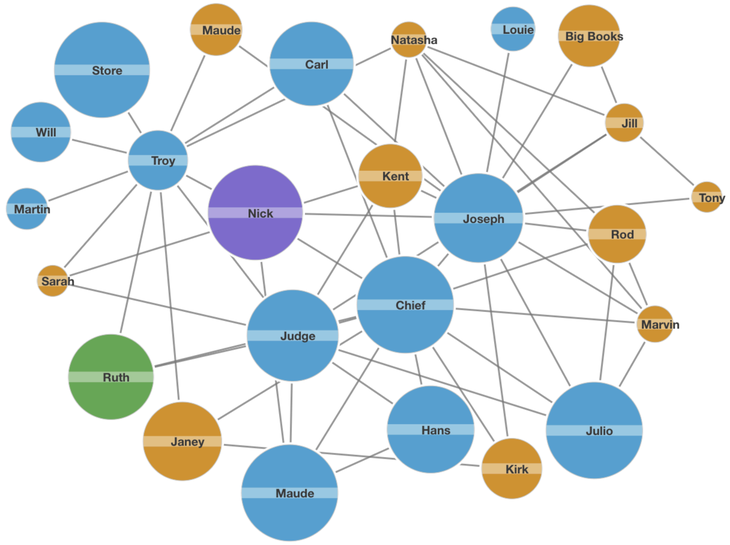
Multivariate networks are made up of nodes and their relationships (links), but also data about those nodes and links as attributes. Most real-world networks are associated with several attributes, and many analysis tasks depend on analyzing both, relations[illegible] Visualization [illegible] of the net[illegible] [illegible] and classify techniques along four axes: layouts, view operations, layout operations, and data operations. We also provide an analysis of tasks specific to multivariate networks and give recommendations for which technique to use in which scenario. Finally, we survey application areas and evaluation methodologies.

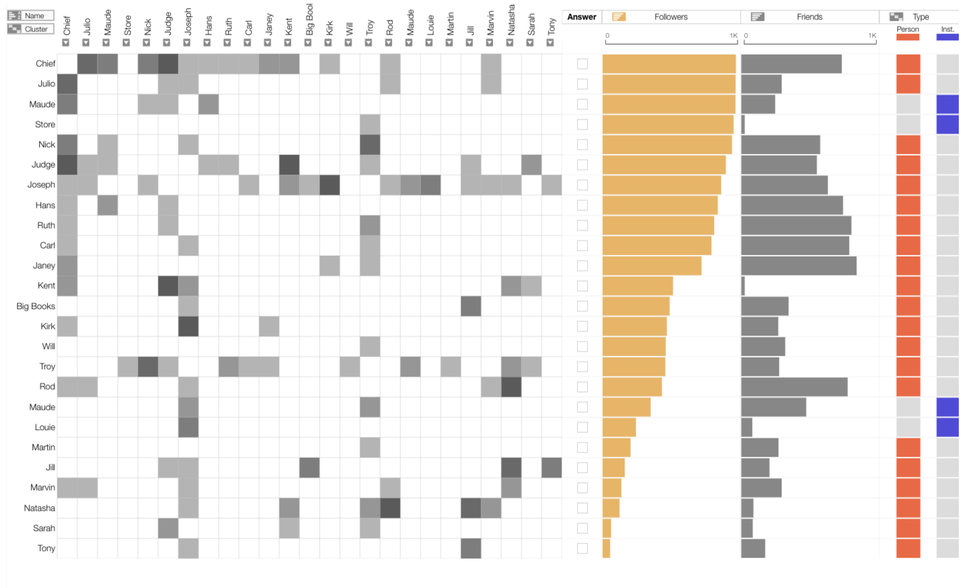
Which is better for which task?

CHALLENGE

CONFOUNDERS

HOW CAN WE MAKE SURE THAT WHAT WE TEST IS WHAT WE CARE ABOUT?





SOLUTION

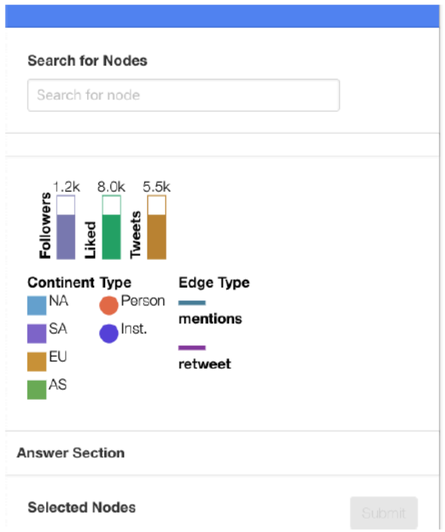
CONFOUNDERS

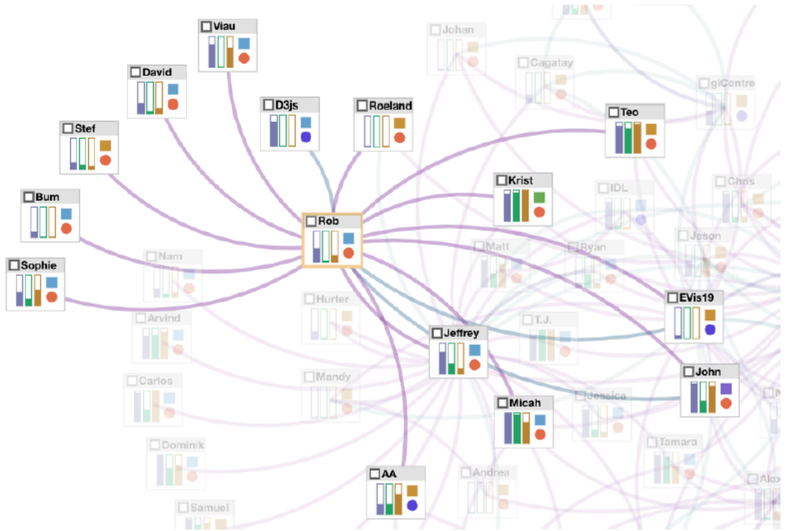
DESIGN BASED ON EXISTING GUIDELINES AND KNOWLEDGE

VALIDATED & REFINE DESIGN BASED ON EXPERT HEURISTIC EVALUATION

Please rate relative to conceivable design alternatives but assuming a node-link diagram as given.

Nested Bars/Glyphs





26. **Embedded bar charts are well suited to encode multiple numerical attributes.**
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly Disagree	◯	◯	◯	◯	◯	◯	◯	Strongly Agree

27. **Embedded colored glyphs are well suited to encode multiple categorical attributes.**
Mark only one oval.

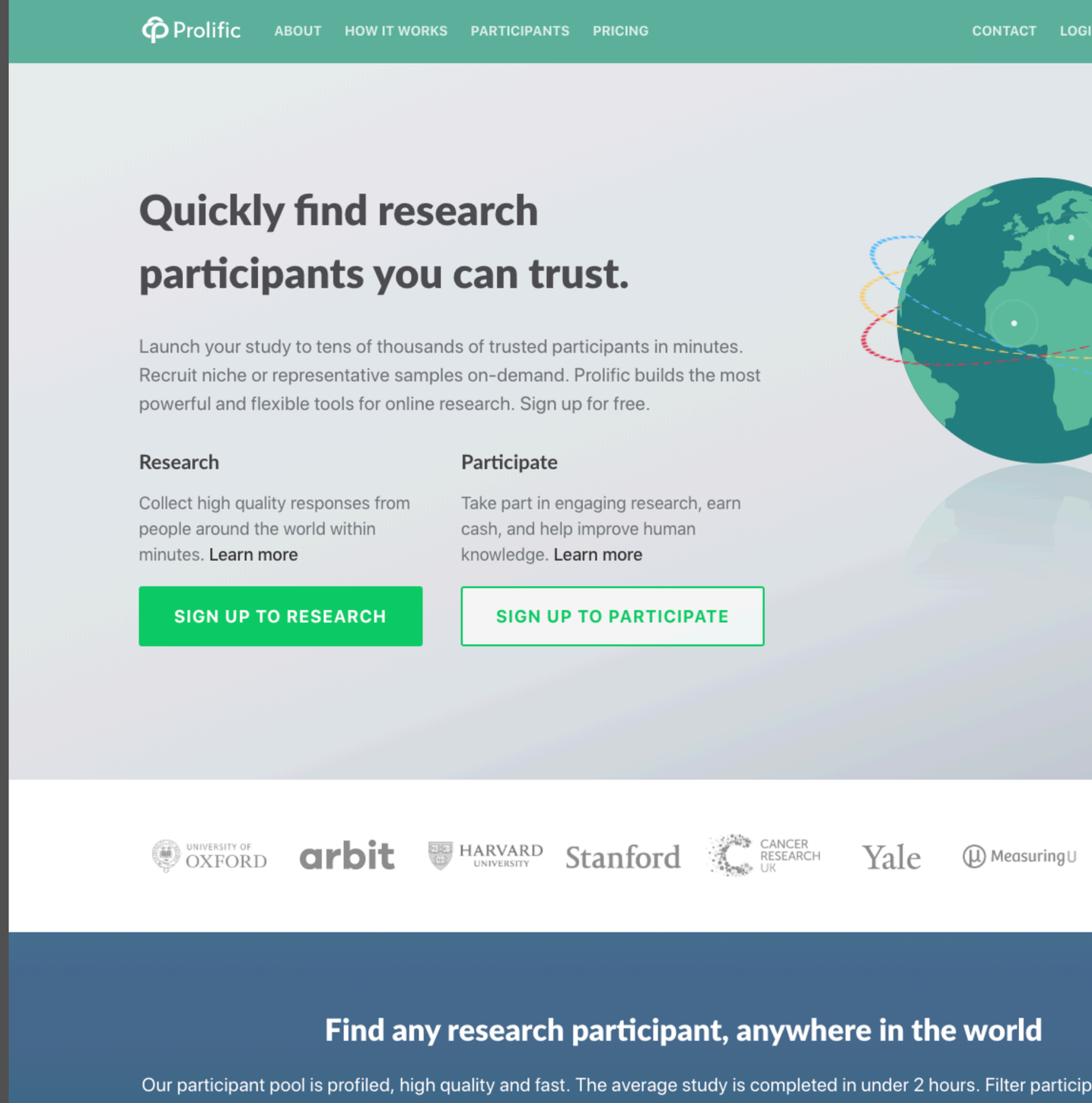
	1	2	3	4	5	6	7	
Strongly Disagree	◯	◯	◯	◯	◯	◯	◯	Strongly Agree

CHALLENGE

SCALE

NEED STATISTICAL POWER

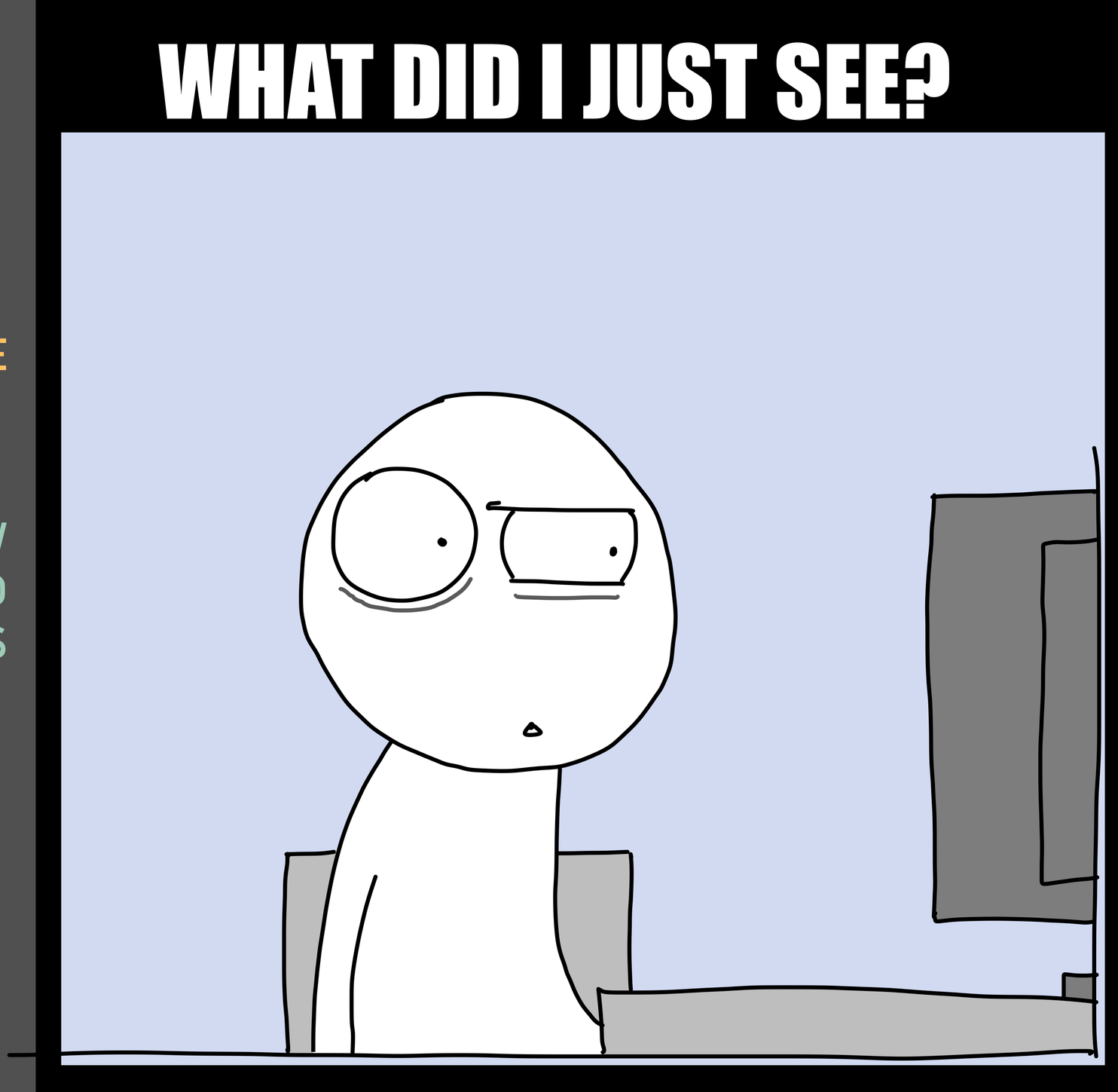
HOW CAN WE DO THIS IN A CROWDSOURCED SETTING?



CHALLENGE

NOVICE USERS

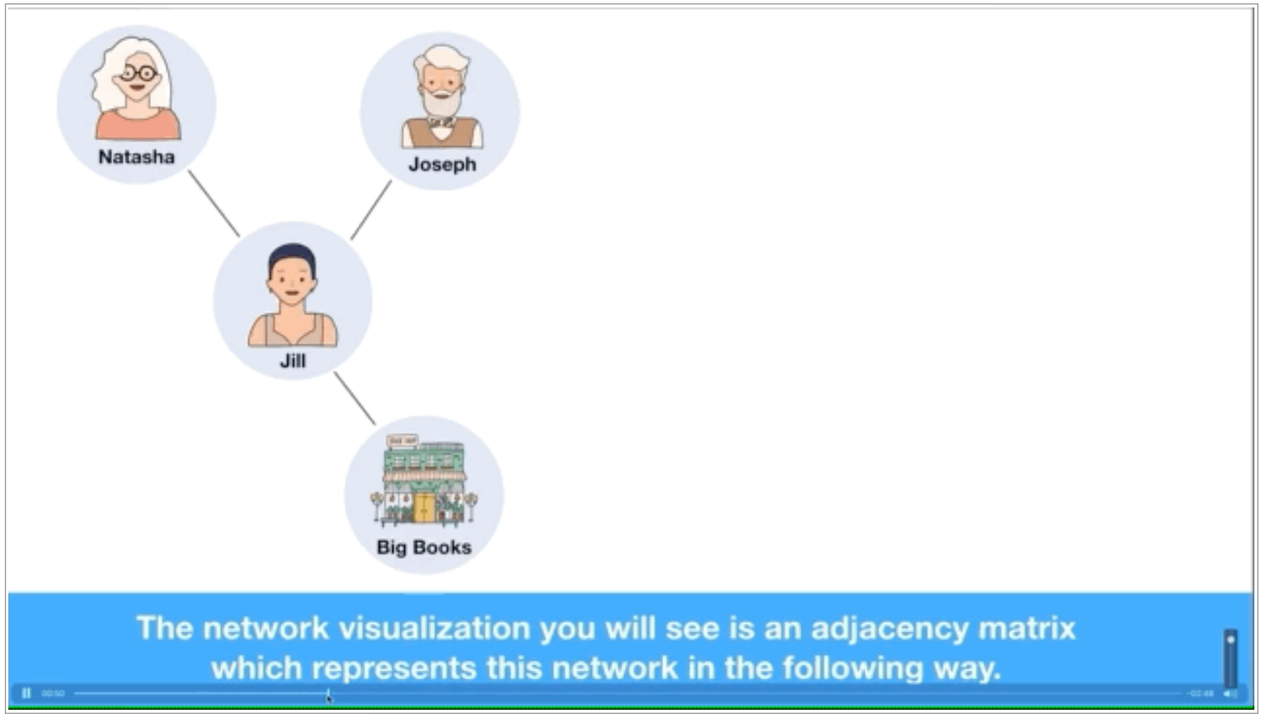
NOVICE USERS DON'T KNOW ABOUT ADVANCED VISUALIZATIONS



SOLUTION

NOVICE USERS

TRAINING CAN GIVE USERS THE EXPERTISE NECESSARY TO COMPLETE THE TASKS.



INCENTIVES

HOW CAN WE GET USERS
TO TRY HARD

AND

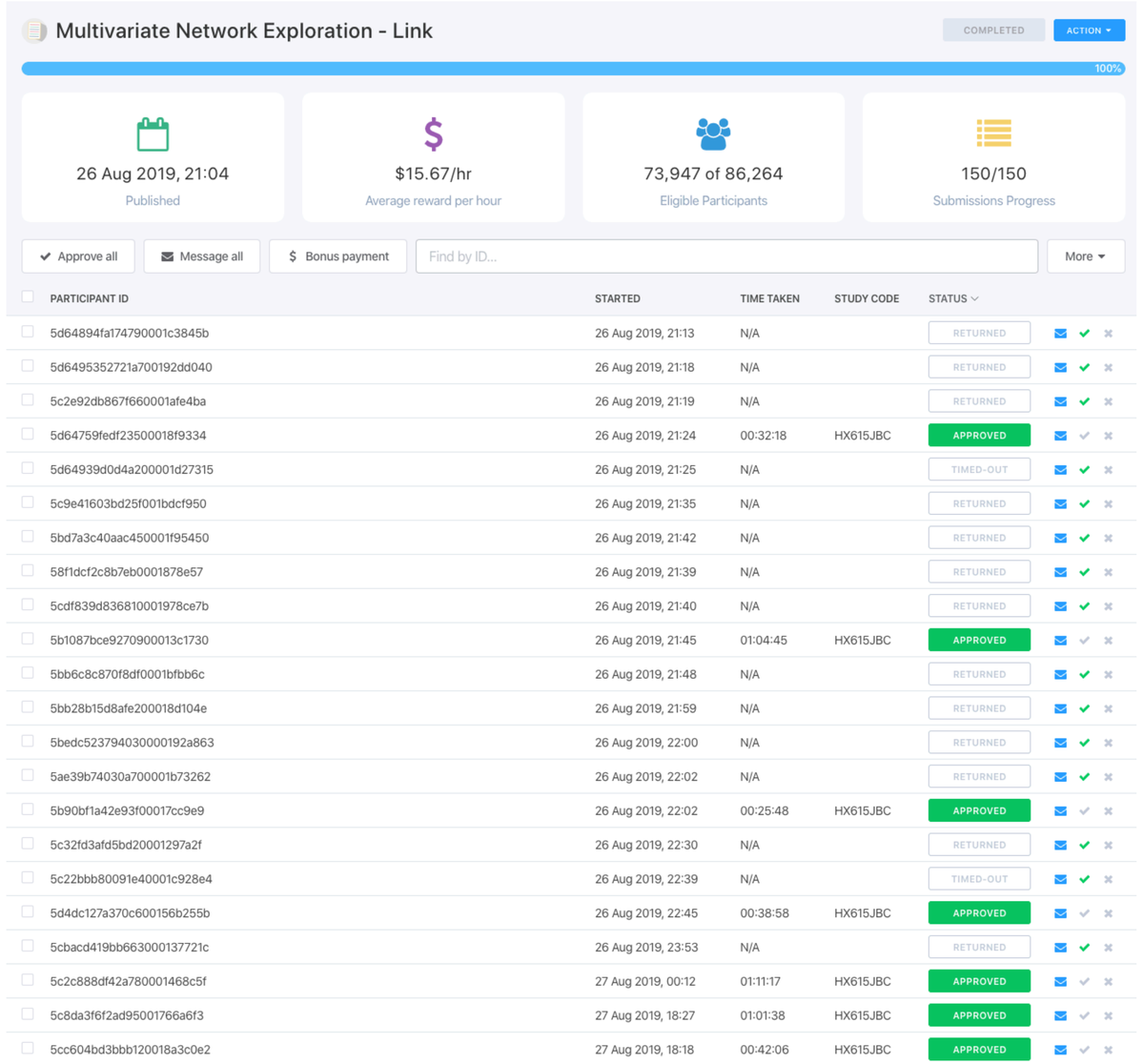
TO PARTICIPATE IN AN EXPERIMENT
THAT TAKES ~1H

[Alfrey Davilla]

INCENTIVES

AN INTERESTING PROBLEM

MONEY



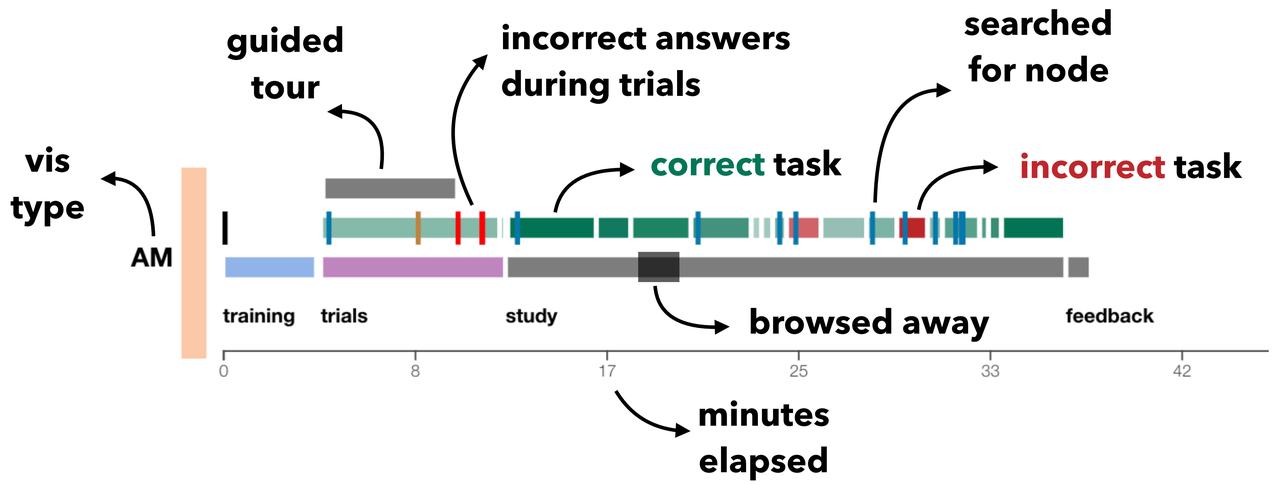
~ $ 4,500 in 2h

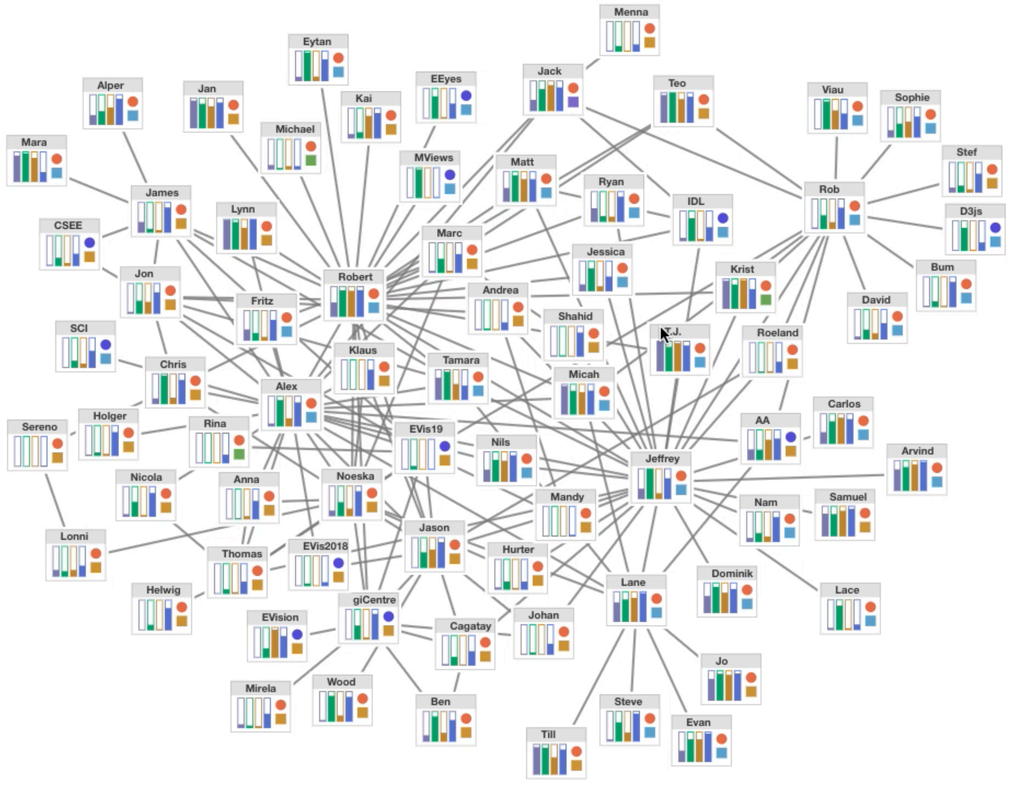
VALIDATION

HOW CAN WE MAKE SURE THIS ALL WORKS?

DETAILED PROVENANCE TRACKING

MULTIPLE PILOTS



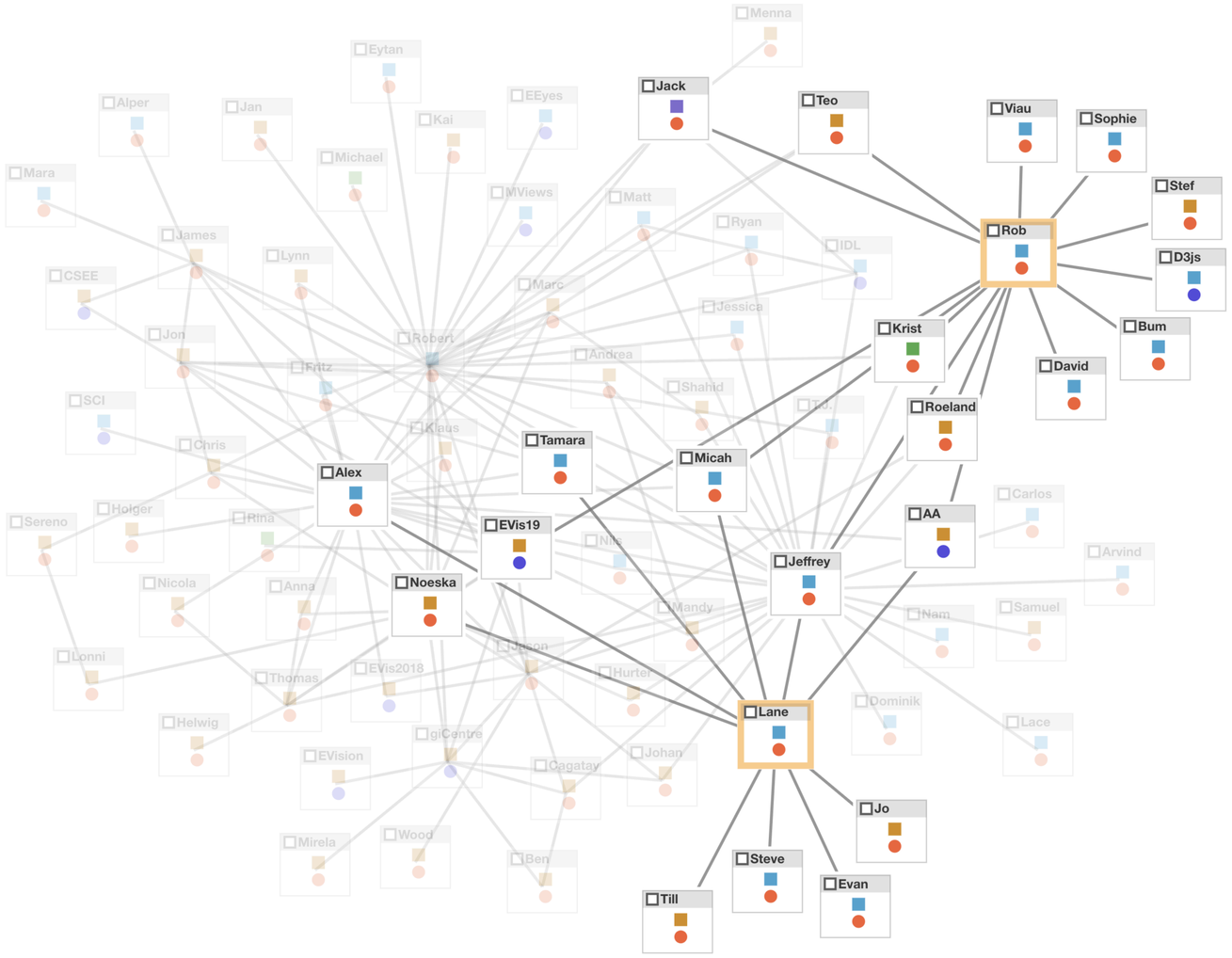

Menna
Eytan
Alper
Jan
Kai
EEyes
Jack
Teo
Viau
Sophie
Mara
Michael
MViews
Matt
Stef
Ryan
IDL
Rob
D3js
James
Lynn
CSEE
Marc
Jessica
Krist
Bum
Jon
Robert
Andrea
David
Fritz
Shahid
T.J.
SCI
Klaus
Roeland
Chris
Tamara
Micah
Alex
Carlos
Holger
Rina
EVis19
AA
Arvind
Sereno
Nils
Jeffrey
Nicola
Anna
Noeska
Samuel
Mandy
Nam
Jason
Lonni
Thomas
EVis2018
Hurter
Dominik
Helwig
Lane
giCentre
Lace
EVision
Johan
Cagatay
Jo
Mirela
Wood
Ben
Steve
Till
Evan


Name
Cluster
Likes
Followers
Tweets
Acct. Age
Type
Continent
0
8K
1K
6K
175
4K
Person
Inst.
NA
SA
EU
AS

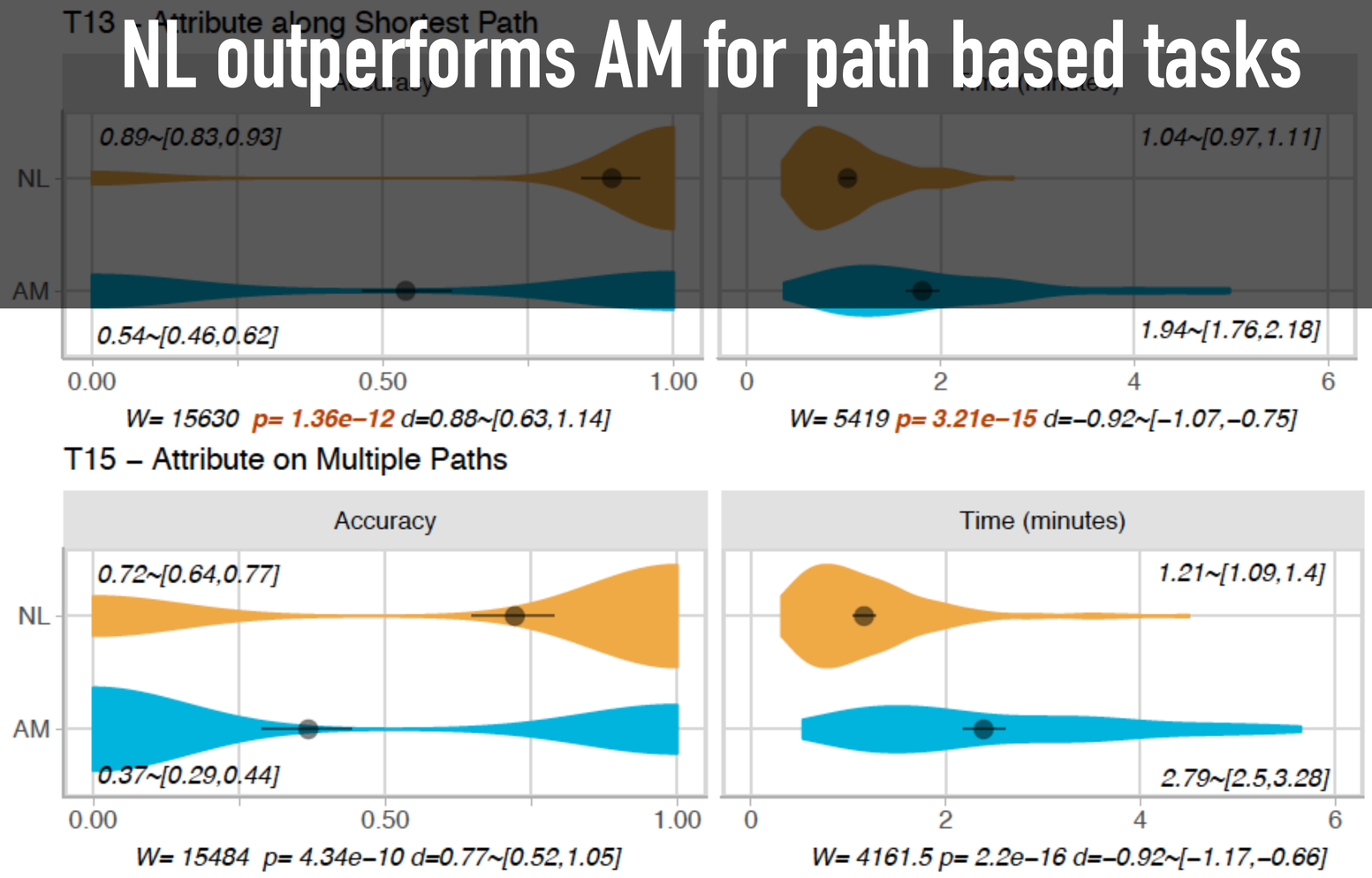
SELECTED RESULTS

PATHS

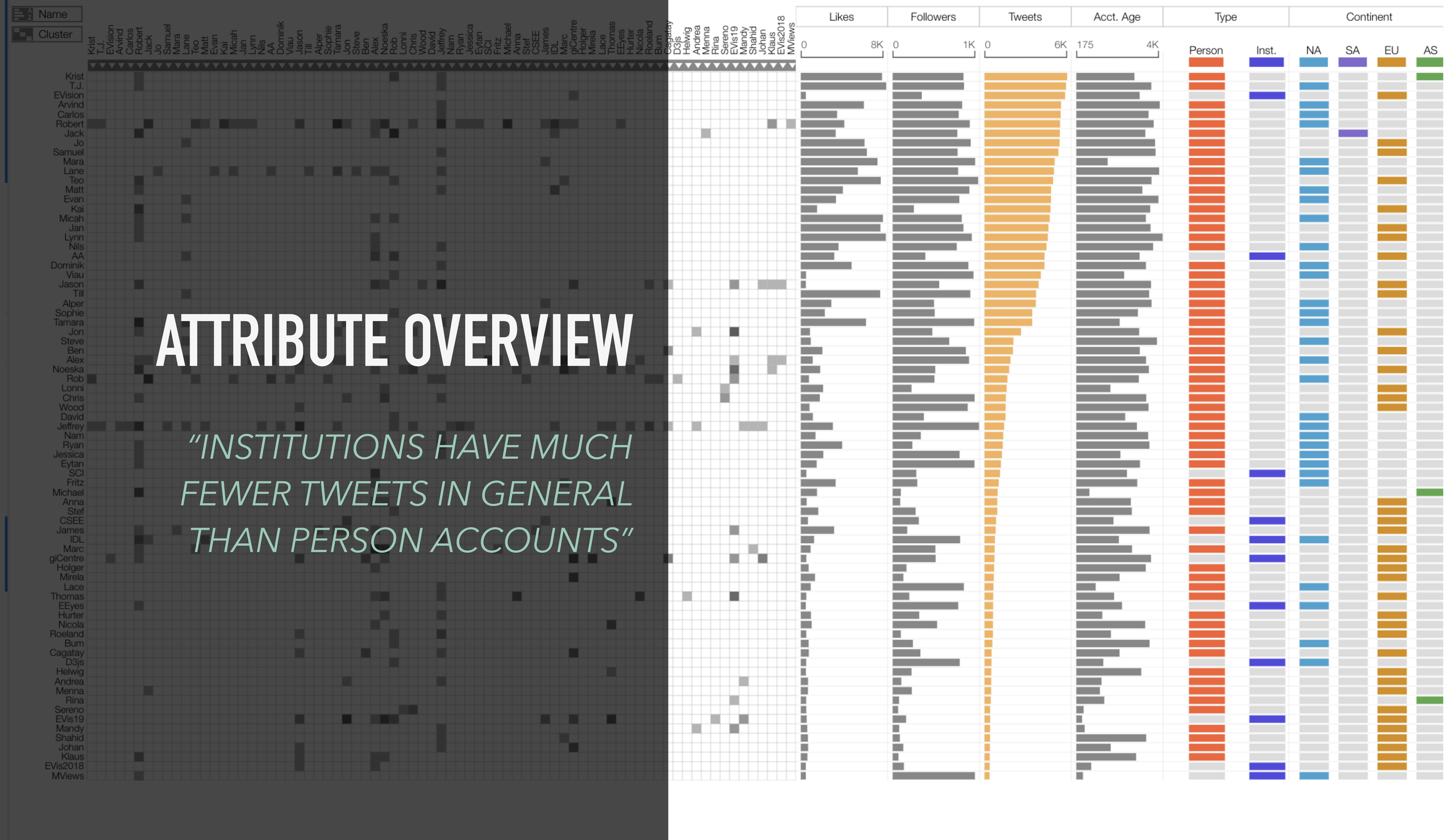
Is NL or AM better for Path Tasks?



NL outperforms AM for path based tasks

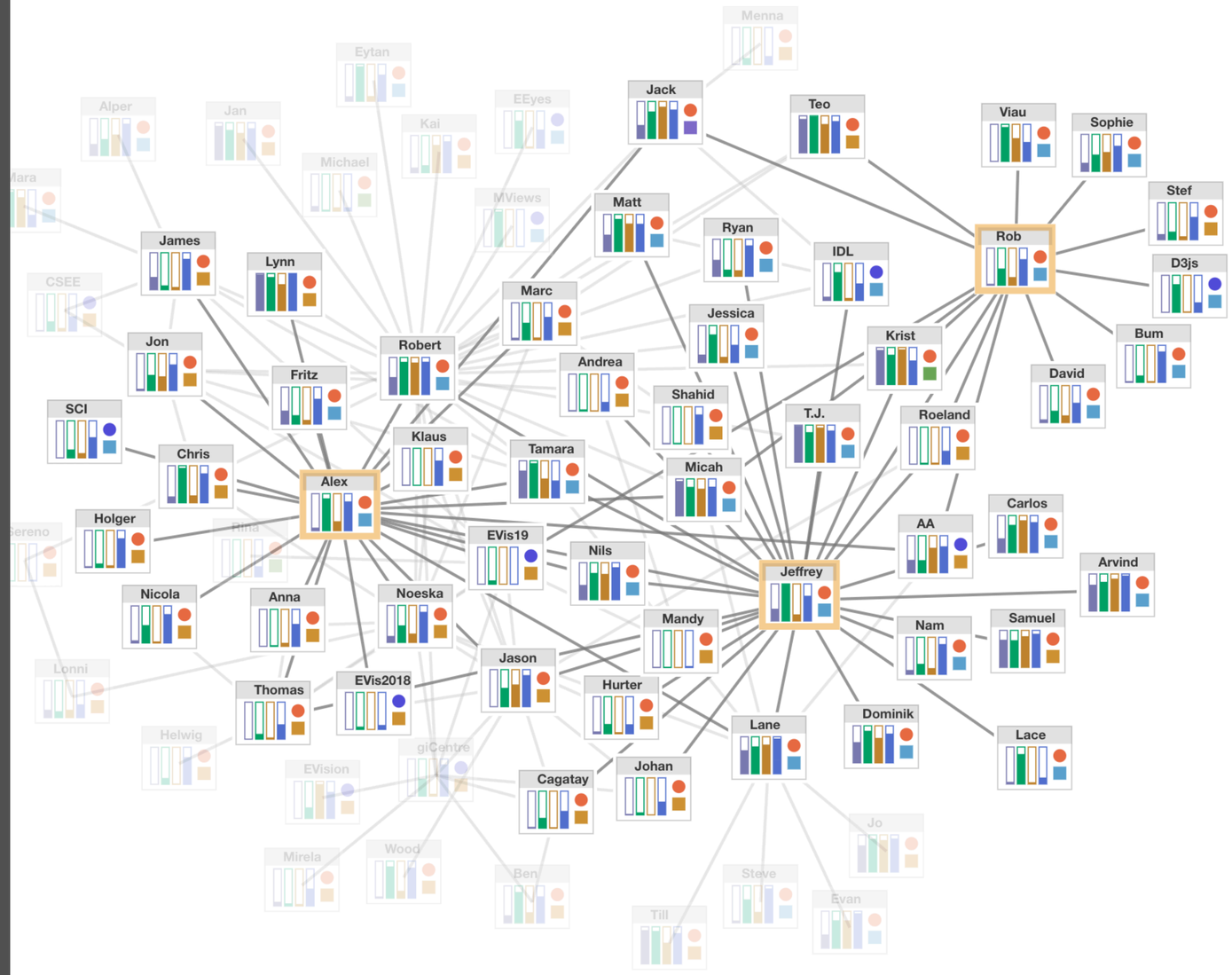


What types of insight do NL and AM representations support?


Name
Cluster
Likes
Followers
Tweets
Acct. Age
Type
Continent
0
8K
0
1K
0
6K
175
4K
Person
Inst.
NA
SA
EU
AS
Krist
T.J.
EVision
Arvind
Carlos
Robert
Jack
Jo
Samuel
Mara
Lane
Teo
Matt
Evan
Kai
Micah
Jan
Lynn
Nils
AA
Dominik
Viau
Jason
Till
Alper
Sophie
Tamara
Jon
Steve
Ben
Alex
Noeska
Rob
Lonni
Chris
Wood
David
Jeffrey
Nam
Ryan
Jessica
Eytan
SCI
Fritz
Michael
Anna
Stef
CSEE
James
IDL
Marc
giCentre
Holger
Mirela
Lace
Thomas
EEyes
Hurter
Nicola
Roeland
Bum
Cagatay
D3js
Helwig
Andrea
Menna
Rina
Sereno
EVis19
Mandy
Shahid
Johan
Klaus
EVis2018
MViews
ATTRIBUTE OVERVIEW
"INSTITUTIONS HAVE MUCH FEWER TWEETS IN GENERAL THAN PERSON ACCOUNTS"

TOPOLOGY-ATTRIBUTE

"IT DOES SEEM A BIT ODD THAT JEFFREY ALEX AND ROB HAVE SUCH LARGE NETWORKS WITH THEIR LOWER THAN AVERAGE TWEETING."



RECAP

CAN WE DO QUANTITATIVE EVALUATION WITH COMPLEX SYSTEMS?

Yes We Can!

- Picking the right techniques
- Evidence-based design
- Design validation
- Careful training
- Good compensation

Pushing the boundary of what can be evaluated using crowdsourcing

WHAT'S NEXT?

TECHNICAL CONTRIBUTIONS

Novel Visualization Techniques

Visualization Process Innovations

Data Wrangling Methods

DOMAIN DRIVEN TECHNIQUES

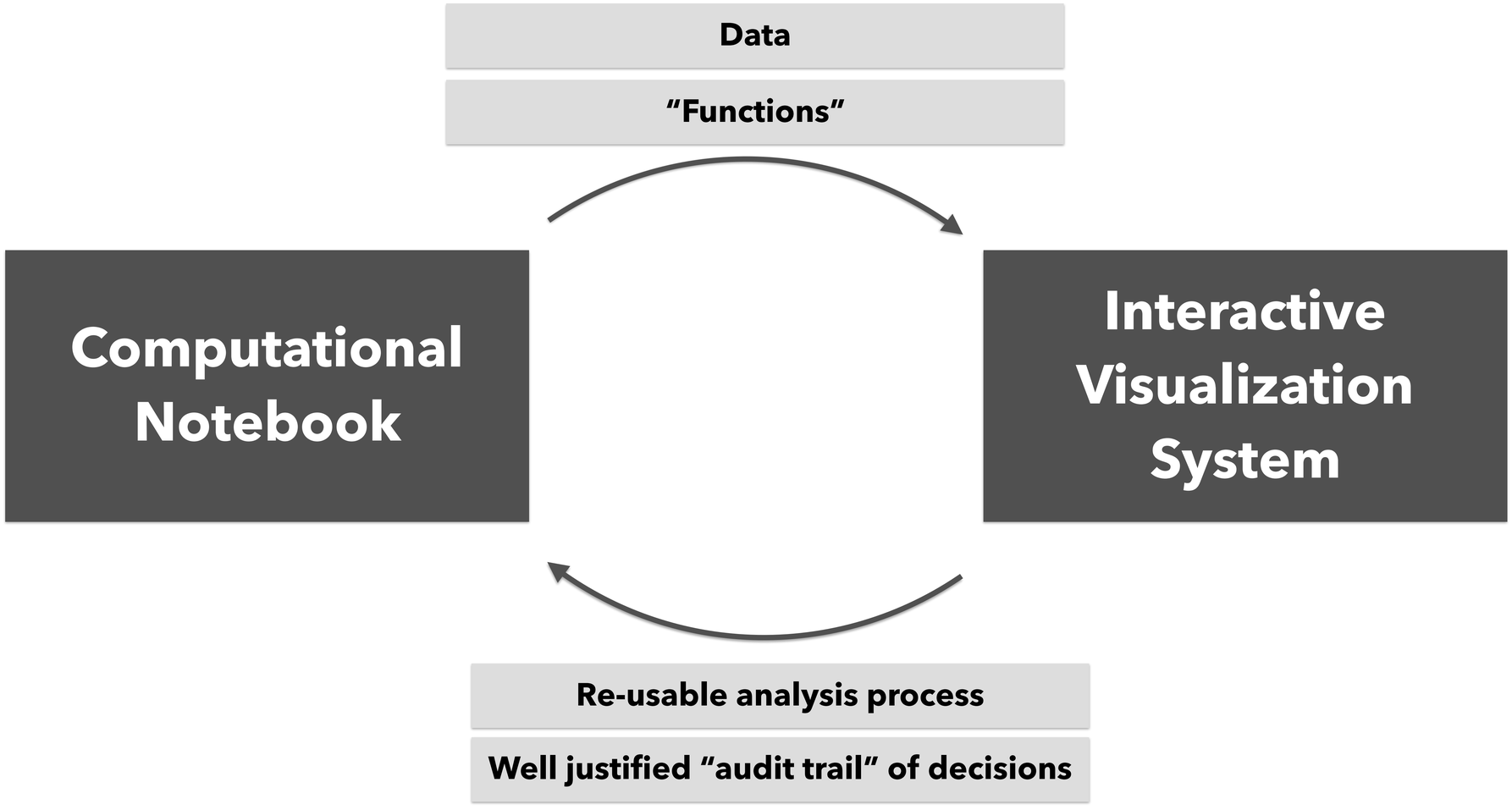
Tailored Methods and Systems for High Impact Science Problems

EMPIRICAL & THEORETICAL WORK

Evaluation Methodology

Design Spaces / Taxonomies

Vision: Make Transitions between Computational Work and Interactive Visualization Seamless



Keep Collaborating

Science Drives Visualization Research

Visualization Research Enables Science

Needed: Methods to transition from prototypes to robust tools

EMPIRICAL & THEORETICAL WORK

A lot of vis design uses best practices with weak empircal foundations

Need to continue to develop strategies to validate complex methods

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

Thanks to: Carolina Nobre, Kiran Gadhave, Jen Rogers, Haihan Lin, Dylan Wootton, Jochen Görtler, Oliver Deussen, Miriah Meyer, Jeff Phillips, Samuel Gratzl, Holger Stitz, Marc Streit, Nils Gehlenborg, Hilary Coon, Lane Harrison, Hendrik Strobelt, Romain Vuillemot, Hanspeter Pfister, and many Others!

visualization
design lab