DeCART Summer School

Data Visualization

Day 2

5 August 2019

Nils Gehlenborg, PhD

Learning Goals

- What is data visualization, how does it work, and what are the principles behind it? (Day 1)
- How can we design and evaluate visualizations? How can we visualize tabular and network data? (Day 2)
- How can we visualize time-series and genomic data? (Day 3)
- How to use Altair https://altair-viz.github.io) to make interactive visualizations in Python.

Review Day 1

What is data visualization?

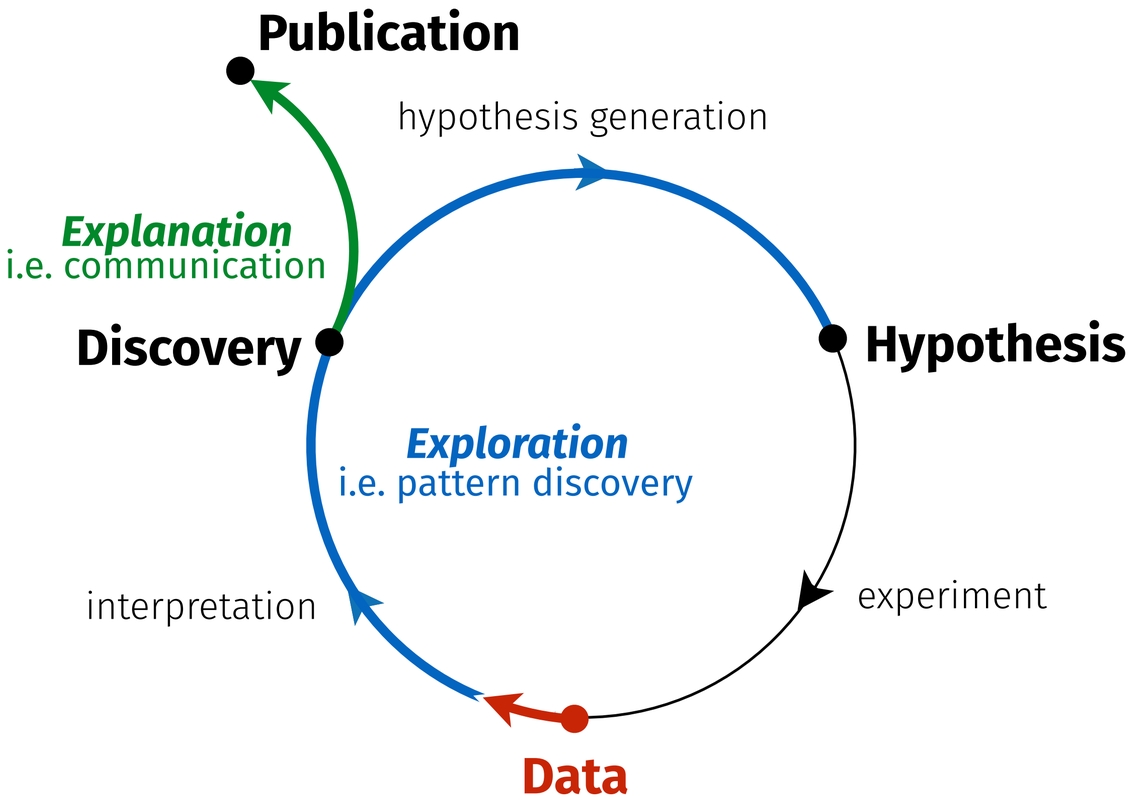
Computer-based visualization systems provide visual representations of datasets intended to help **people** carry out some **task** **more effectively**.

— *Tamara Munzner*

Visualization Use Cases

Exploration
Confirmation
Explanation

Discovery Process



How does visualization work?

How does visualization work?

Visualization uses perception to free up cognition.

How does visualization work?

Visualization uses perception to free up cognition.

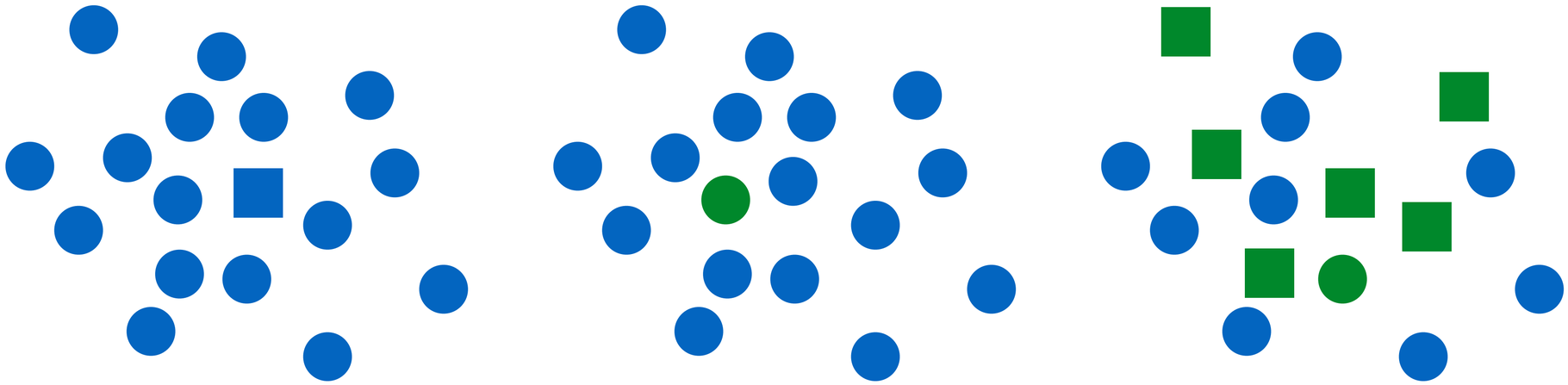
Visualization is an external cognitive aid
and augments working memory.

How does visualization work?

$$\begin{array}{r} \mathbf{453 \ \ x \ \ 862} \ \ \mathbf{=} \ \ \mathbf{?} \\ \hline 906 \\ +\ 27{,}180 \\ +\ 362{,}400 \\ \hline 390{,}486 \end{array}$$

Preattentive Processing

Preattentive Processing



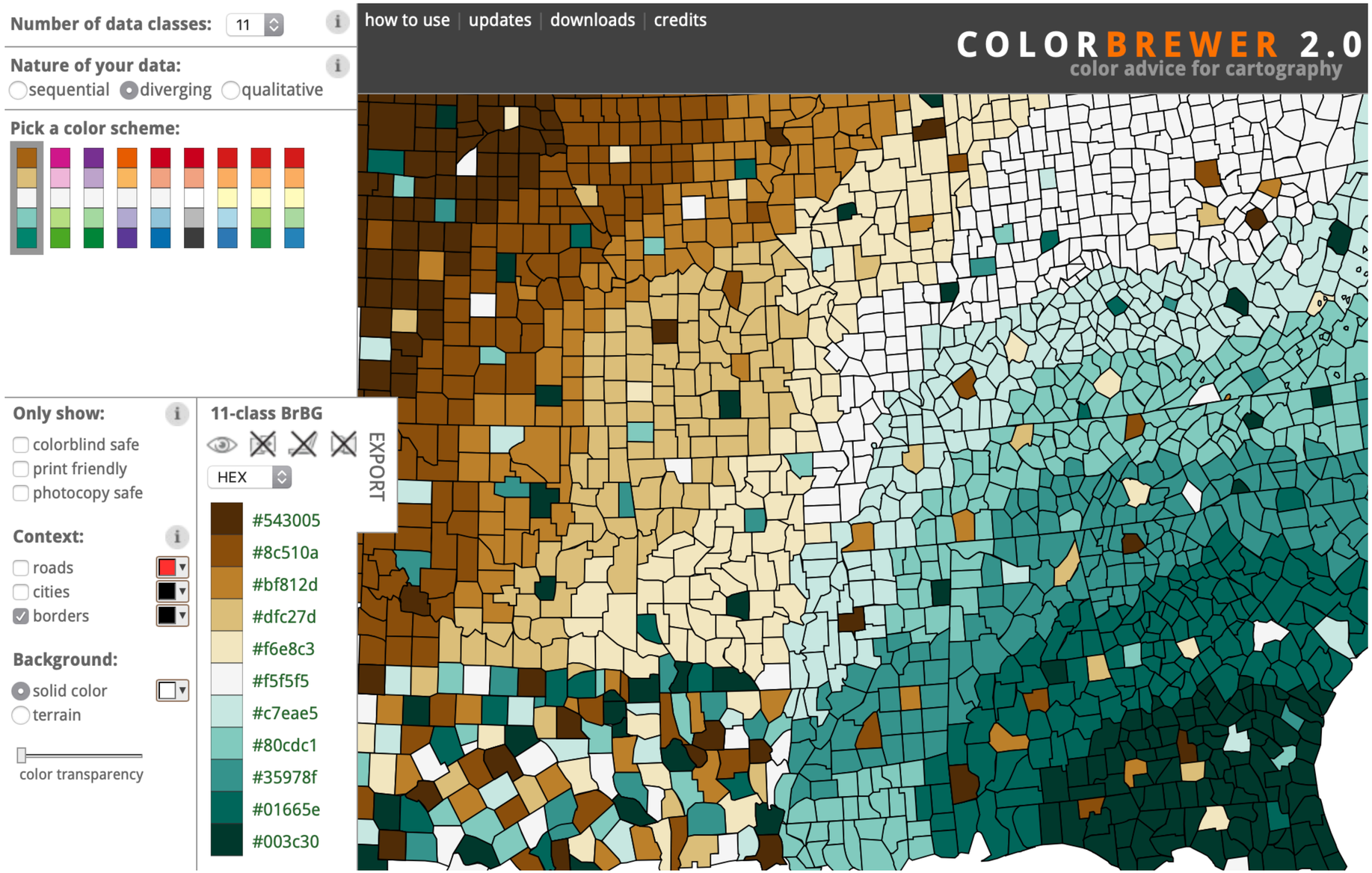
Shape: Yes! Color: Yes! **Shape + Color: No!**

- visual properties that can be perceived in less than 250 ms
- no sequential scanning of the image required, unlike text or numbers
- examples for other visual properties that can be processed preattentively: orientation, curvature, direction of motion, size and others

after Healey et al. 1993, Healey and Enns 1999

Color Maps

Diverging, Sequential, Qualitative



http://colorbrewer2.org

Exercise: Redesign

Euler Diagram

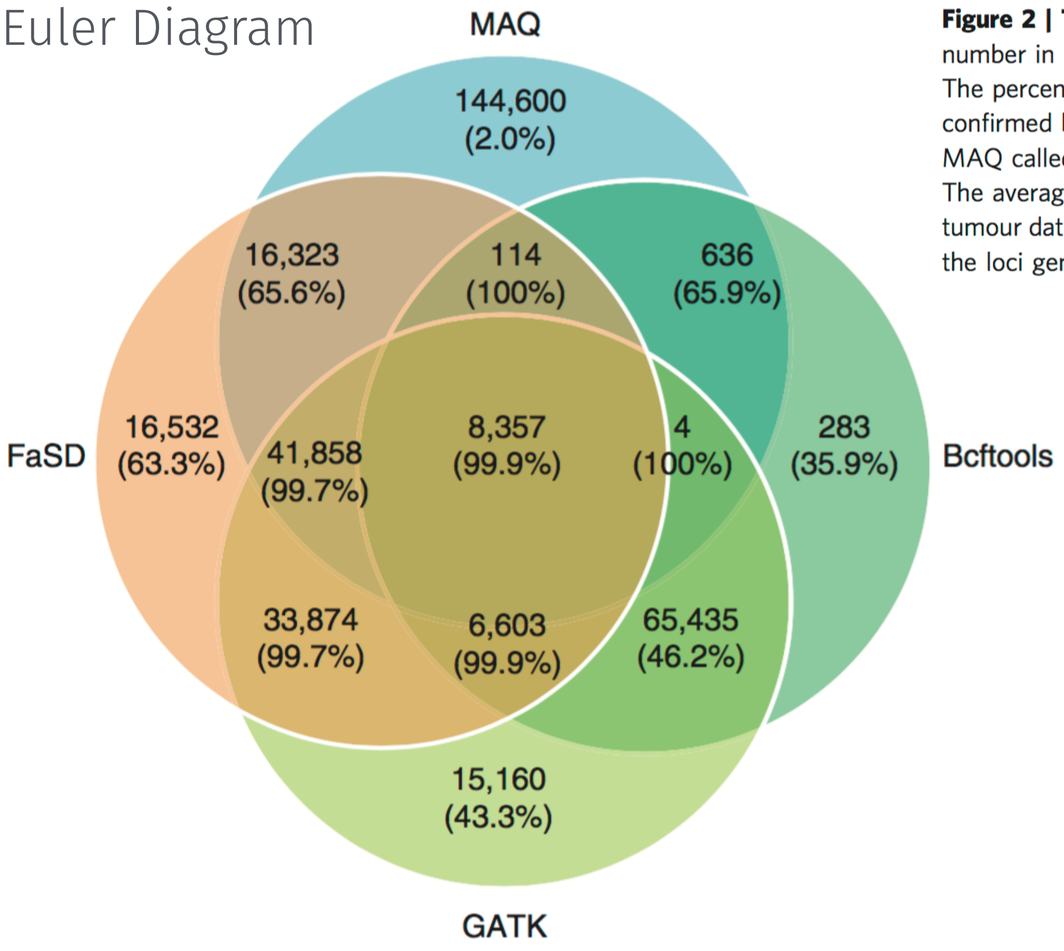


Figure 2 | The Venn diagram of SNPs detected by different tools. The number in each cell is the number of SNPs in the corresponding category. The percentage under the number is the proportion of SNPs that were confirmed by the Affymetrix SNP array. The FaSD, GATK, Bcftools and MAQ called 123661, 171291, 81432 and 211892 SNPs in total, respectively. The average depth of this data set was 10 ×. The figure is based on the tumour data set and Bowtie was used as aligner, and statistics are based on the loci genotyped by the Affymetrix SNP array.

F Xu et al., Nature Communications 3, 1258 doi:10.1038/ncomms2256

Exercise: Redesign

Euler Diagram

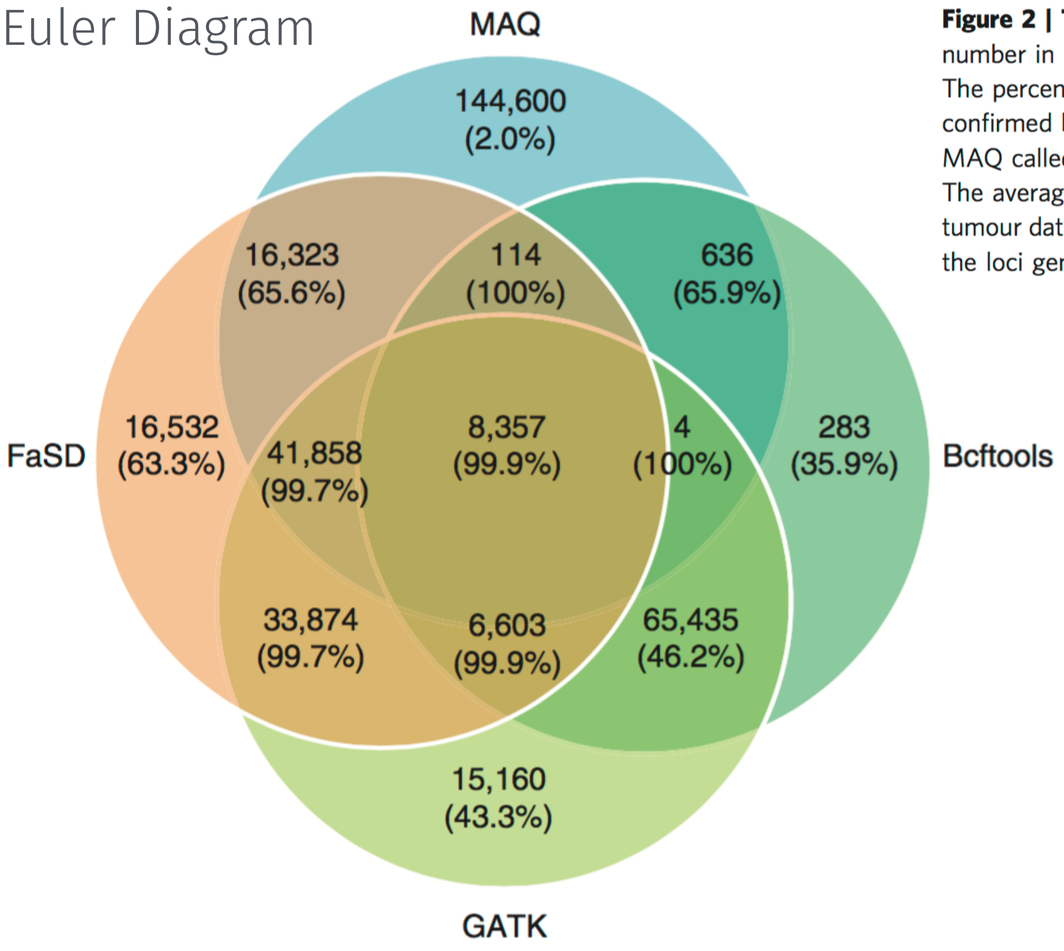
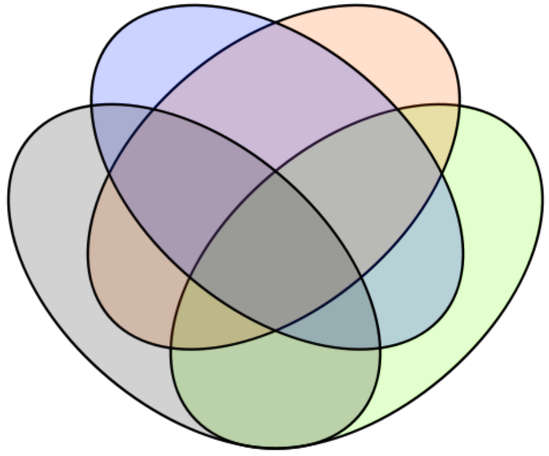


Figure 2 | The Venn diagram of SNPs detected by different tools. The number in each cell is the number of SNPs in the corresponding category. The percentage under the number is the proportion of SNPs that were confirmed by the Affymetrix SNP array. The FaSD, GATK, Bcftools and MAQ called 123661, 171291, 81432 and 211892 SNPs in total, respectively. The average depth of this data set was 10 ×. The figure is based on the tumour data set and Bowtie was used as aligner, and statistics are based on the loci genotyped by the Affymetrix SNP array.

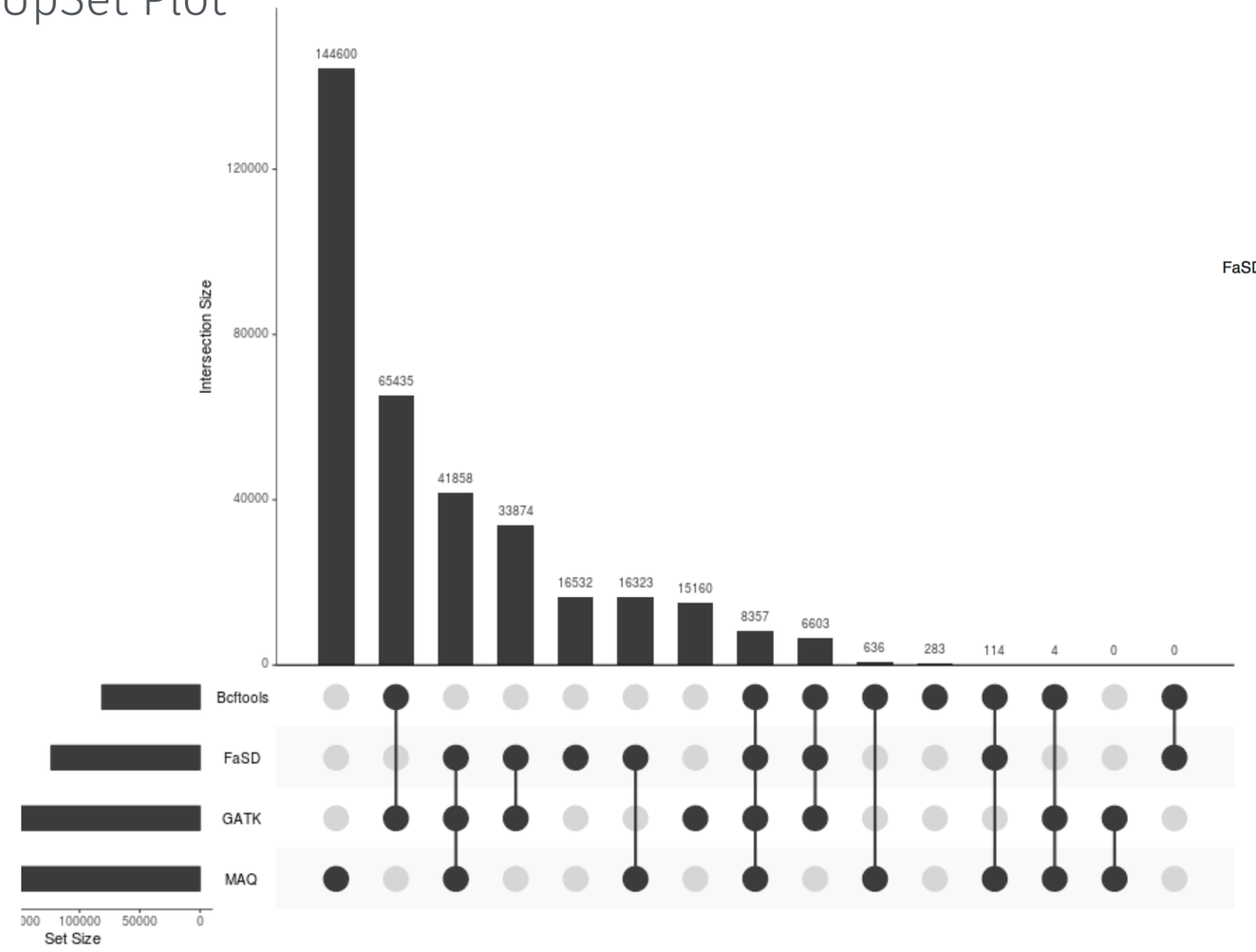
Venn Diagram



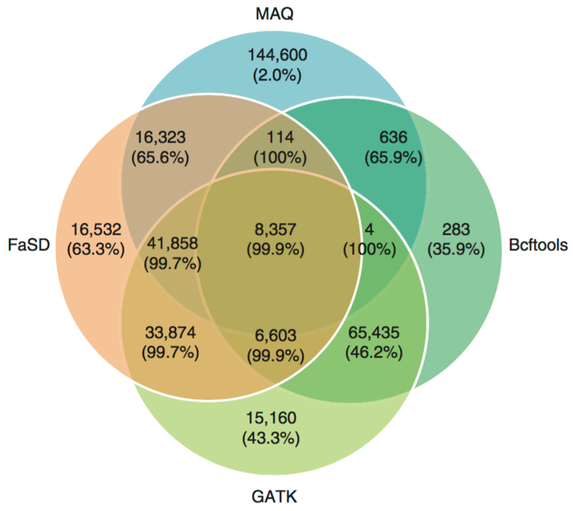
F Xu et al., Nature Communications 3, 1258 doi:10.1038/ncomms2256

Exercise: Redesign

UpSet Plot



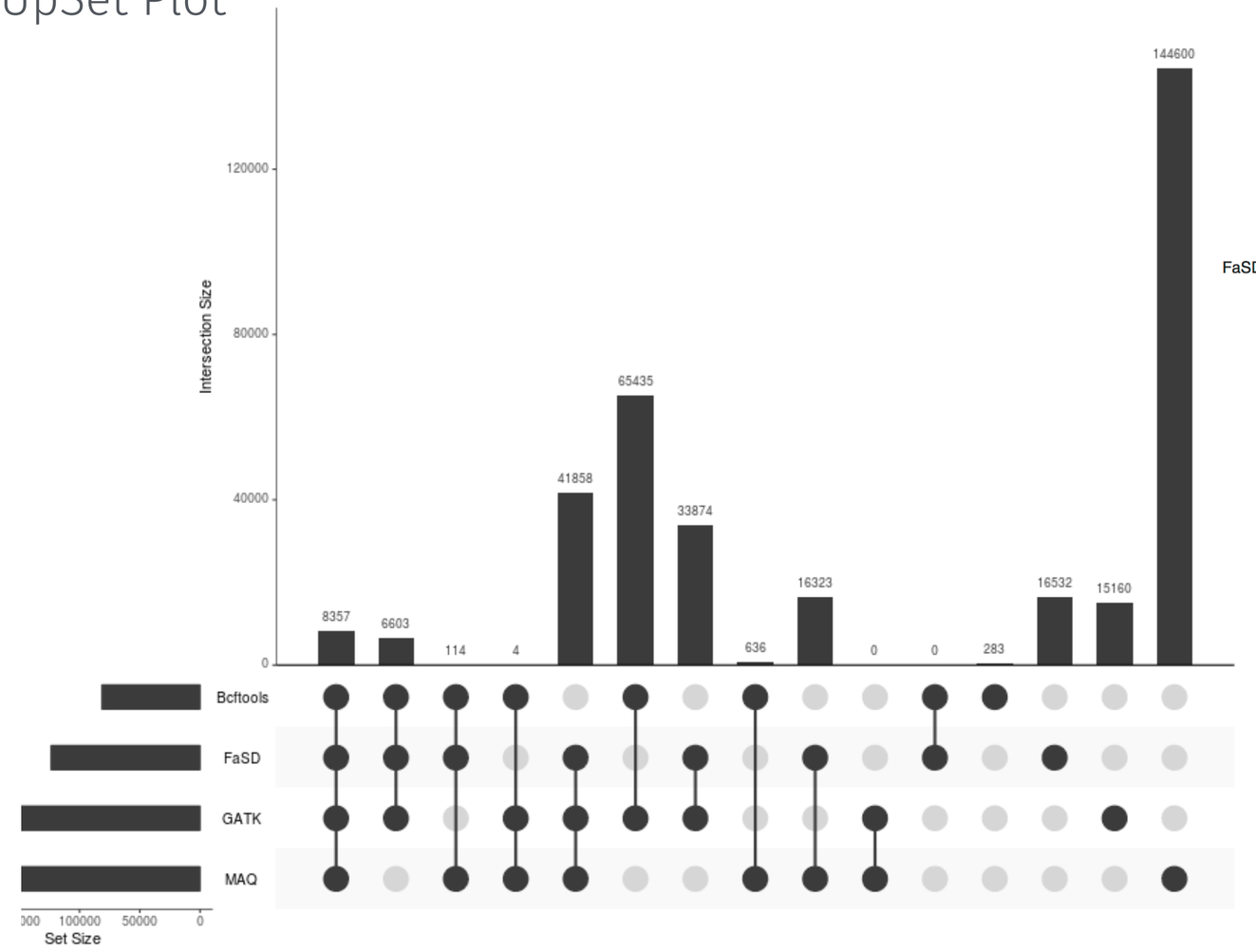
Euler Diagram



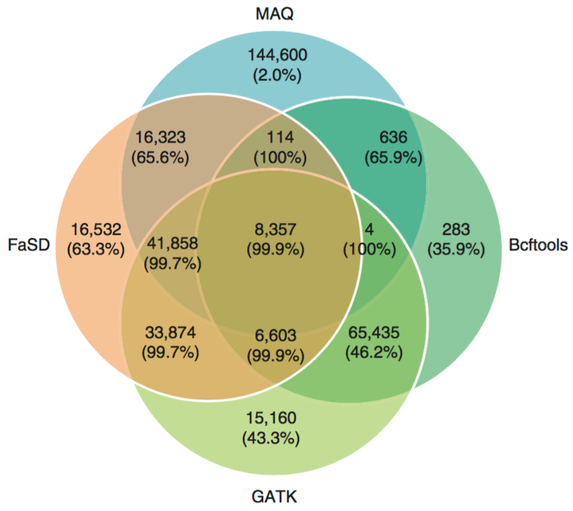
https://cran.rstudio.com/web/packages/UpSetR/ & https://gehlenborglab.shinyapps.io/upsetr/,

Exercise: Redesign

UpSet Plot



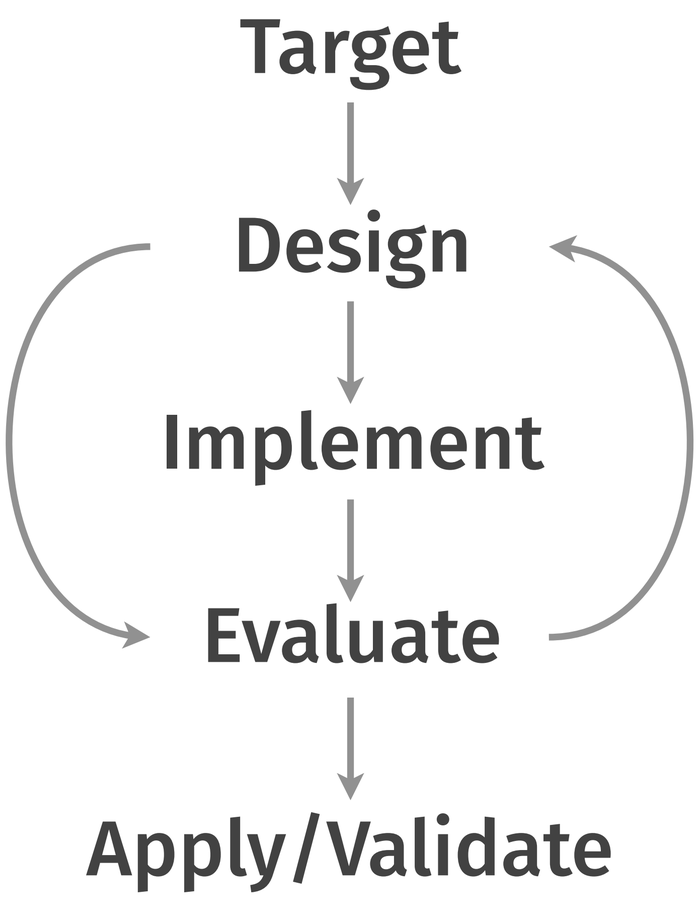
Euler Diagram



https://cran.rstudio.com/web/packages/UpSetR/ & https://gehlenborglab.shinyapps.io/upsetr/,

Design Process

User-Centered Participatory Design



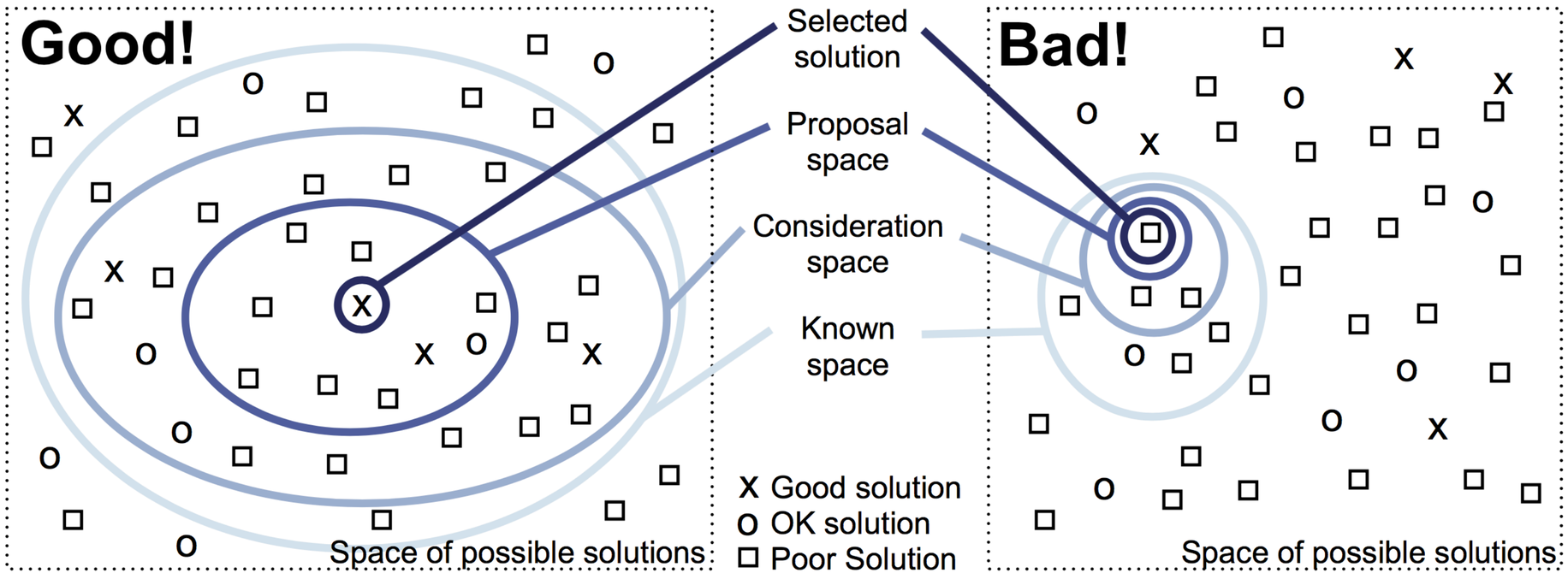
user-centered design

usability engineering

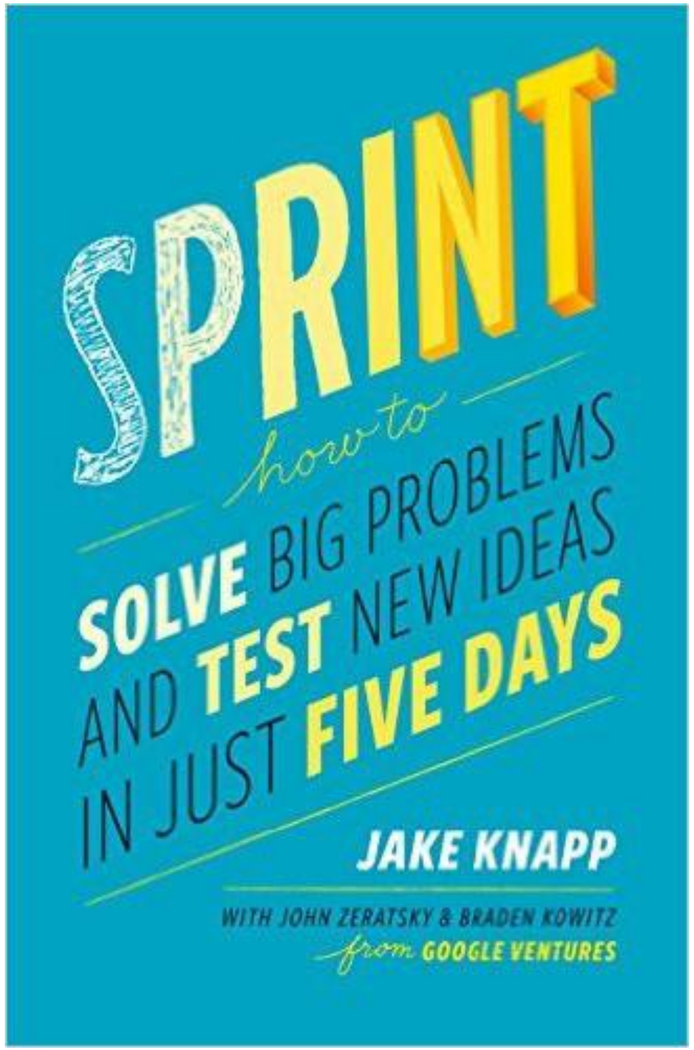
participatory design

adapted from Miriah Meyer, University of Utah and Cydney Nielsen, University of British Columbia

Goal of the Design Process

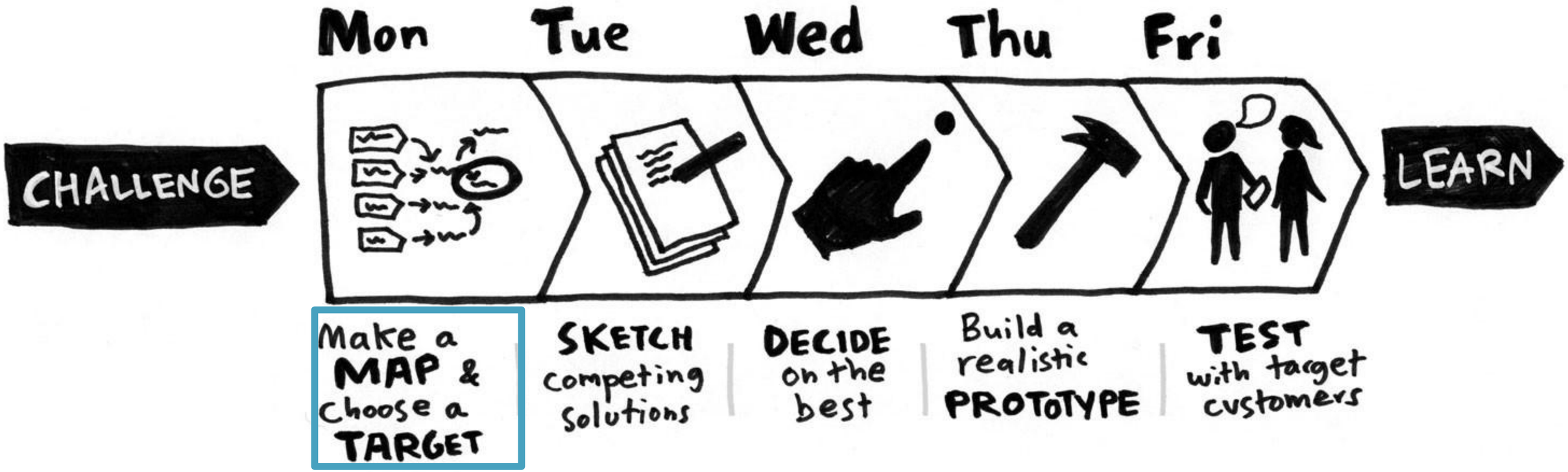


Design Sprint



Knapp, Jake, John Zeratsky, and Braden Kowitz. Sprint: How to solve big problems and test new ideas in just five days. Simon and Schuster, 2016.

Design Sprint

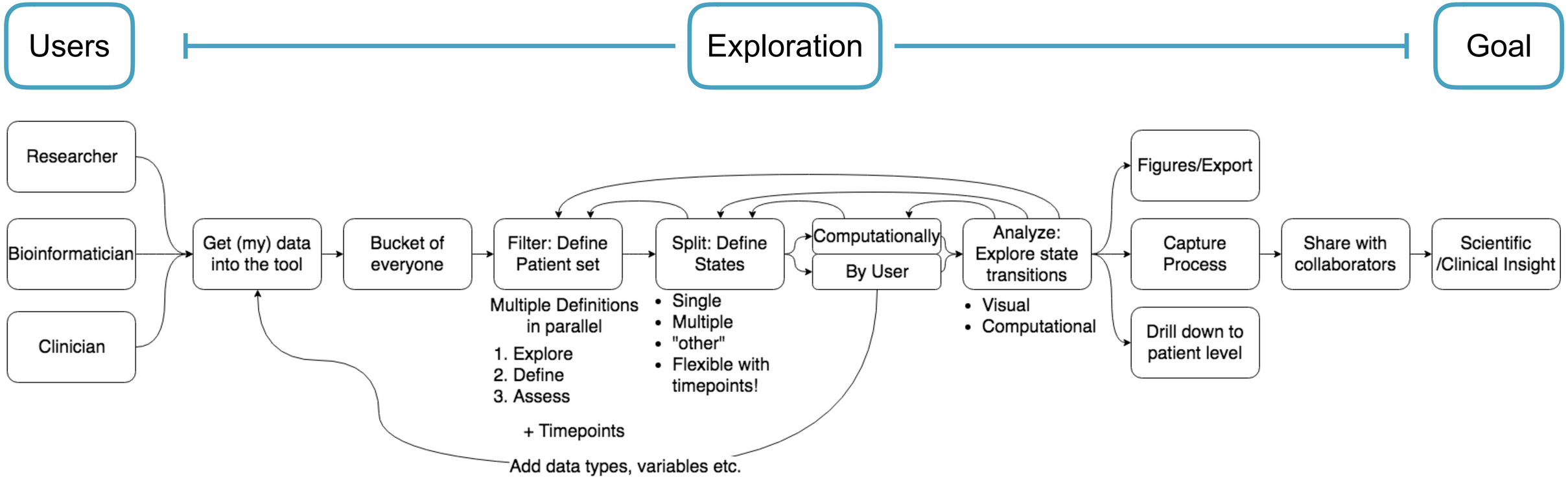


Defining the problem

- Goal: “Develop the ‘go-to’ visualization approach for longitudinal cancer genomics data through an agile framework, that will have measurable technical and scientific impact.”

- Pessimistic questions: e.g. “Will we be able to integrate a useful version of this tool into the cBio portal?”

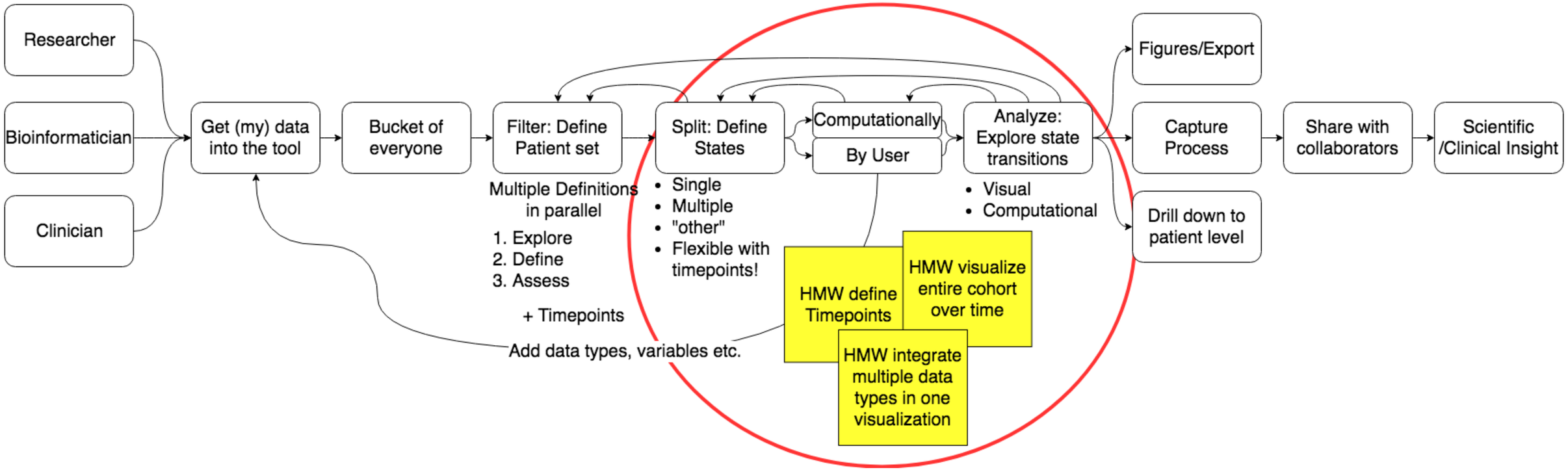
Monday - Map


Users
Exploration
Goal
Researcher
Bioinformatician
Clinician
Get (my) data into the tool
Bucket of everyone
Filter: Define Patient set
Multiple Definitions in parallel
1. Explore
2. Define
3. Assess
+ Timepoints
Split: Define States
• Single
• Multiple
• "other"
• Flexible with timepoints!
Computationally
By User
Analyze: Explore state transitions
• Visual
• Computational
Figures/Export
Capture Process
Drill down to patient level
Share with collaborators
Scientific /Clinical Insight
Add data types, variables etc.

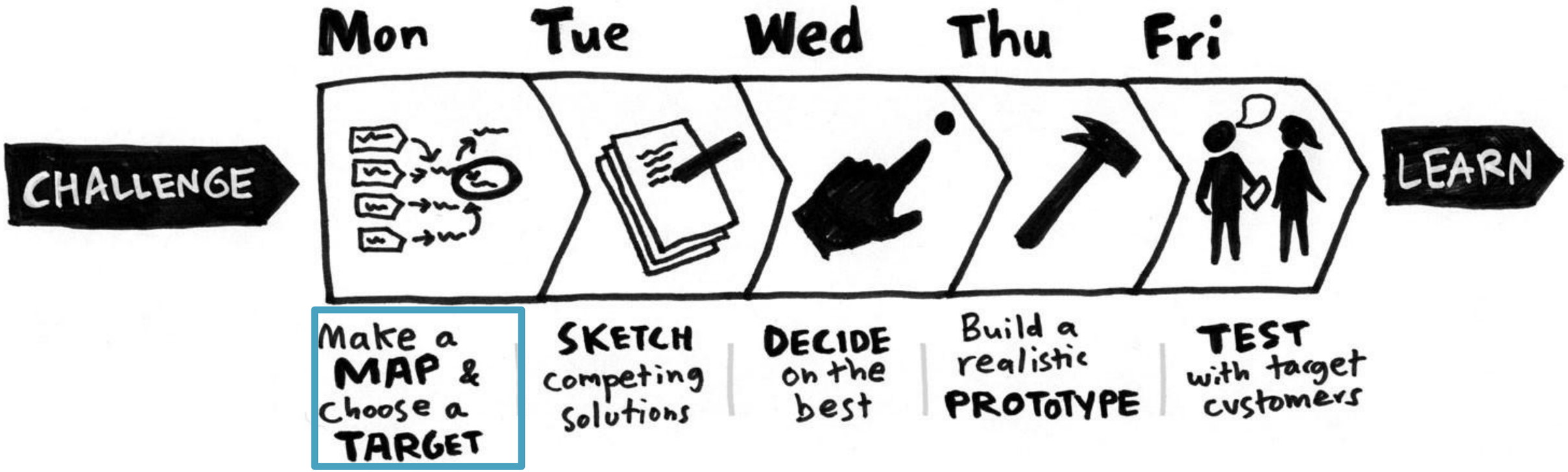
Monday - User Interviews

- 4 cancer researchers, 20 minutes each
- Unstructured interviews
- Notes in “How might we?” format

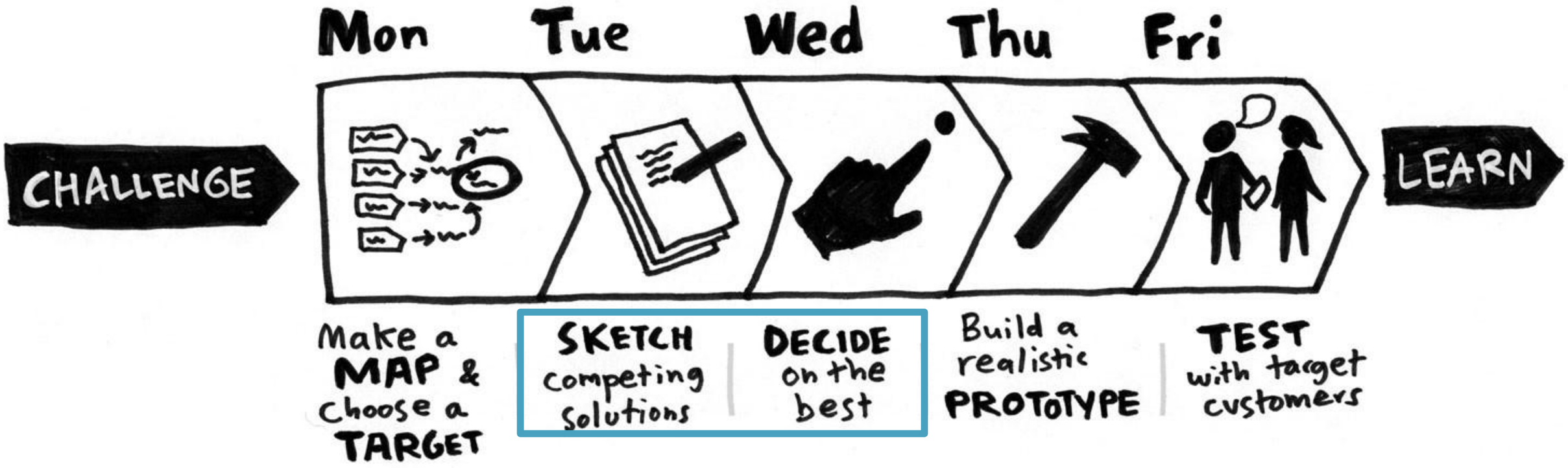
“How might we?”


Researcher
Bioinformatician
Clinician
Get (my) data into the tool
Bucket of everyone
Filter: Define Patient set
Split: Define States
Computationally
By User
Analyze: Explore state transitions
Figures/Export
Capture Process
Share with collaborators
Scientific /Clinical Insight
Drill down to patient level
Multiple Definitions in parallel
1. Explore
2. Define
3. Assess
+ Timepoints
Add data types, variables etc.
• Single
• Multiple
• "other"
• Flexible with timepoints!
• Visual
• Computational
HMW define Timepoints
HMW visualize entire cohort over time
HMW integrate multiple data types in one visualization

Design Sprint

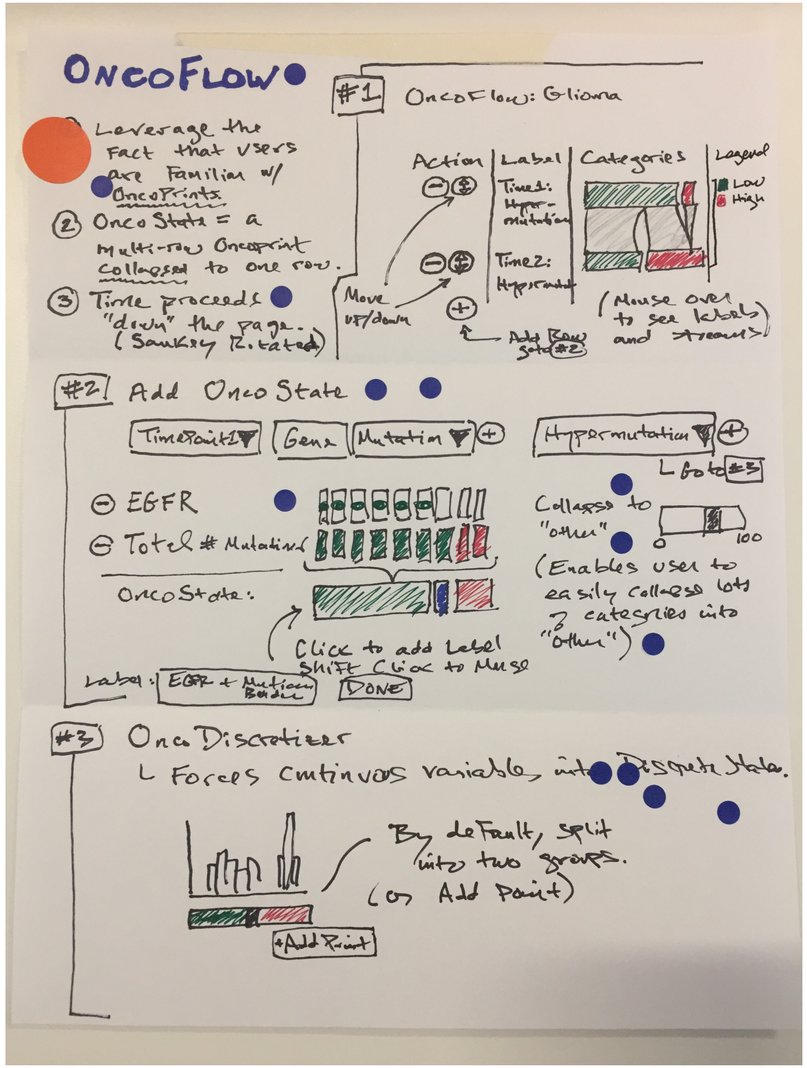


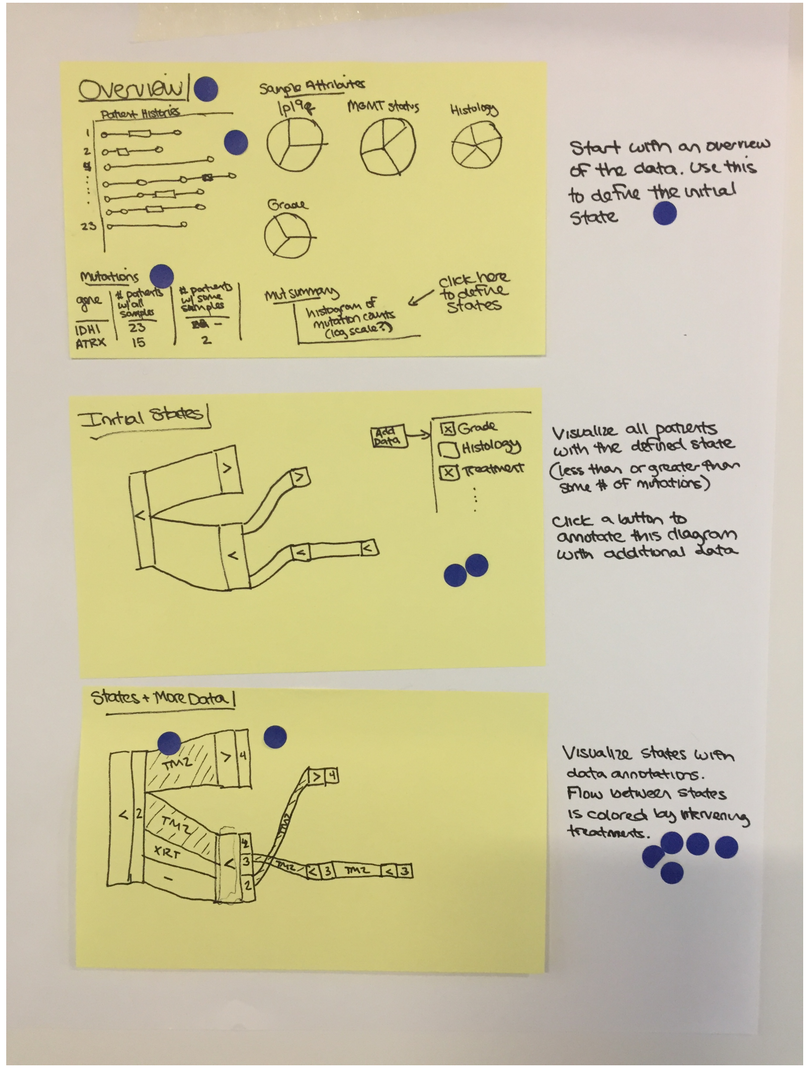
Design Sprint



Tuesday/Wednesday

Tuesday/Wednesday

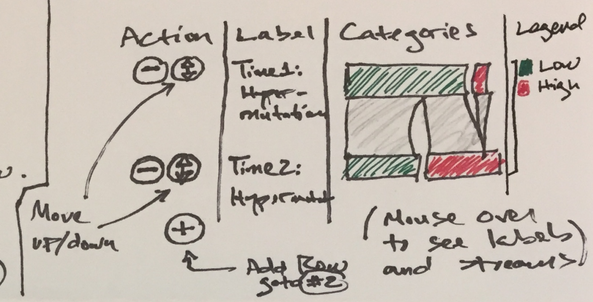




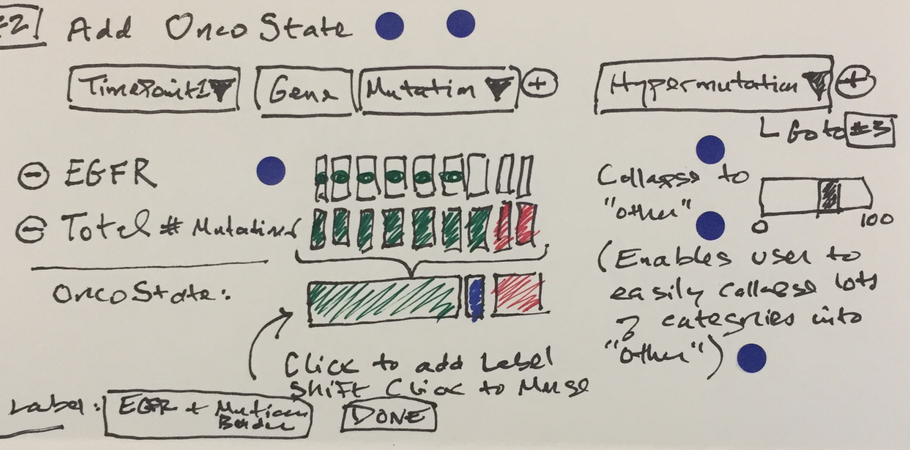
OncoFlow

1. Leverage the fact that users are familiar w/ OncoPrints.
2. OncoState = a multi-row Oncoprint collapsed to one row.
3. Time proceeds "down" the page. (Sankey Rotated)

#1 OncoFlow: Glioma

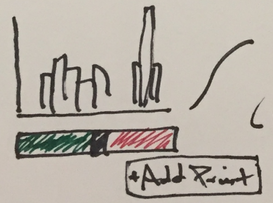


#2 Add OncoState



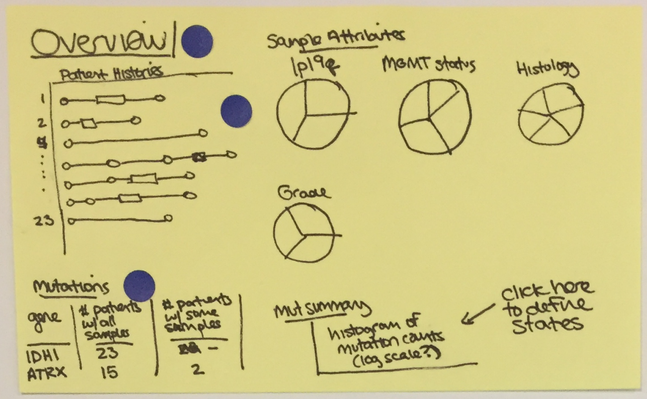
#3 OncoDiscretizer

↳ Forces continuous variables into Discrete States.



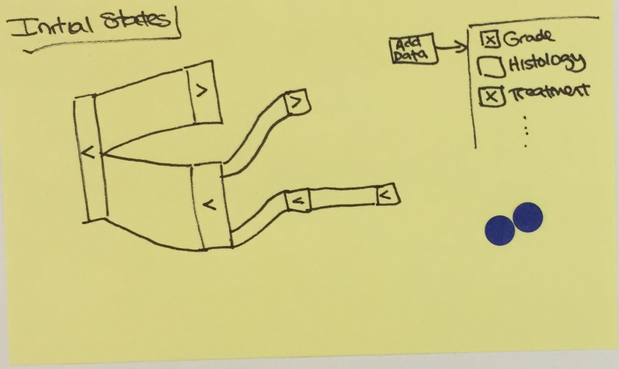
By default, split into two groups. (or Add Point)

Overview



Start with an overview of the data. Use this to define the initial state

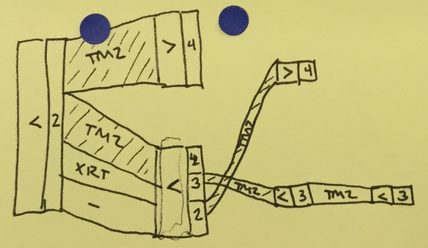
Initial States



Visualize all patients with the defined state (less than or greater than some # of mutations)

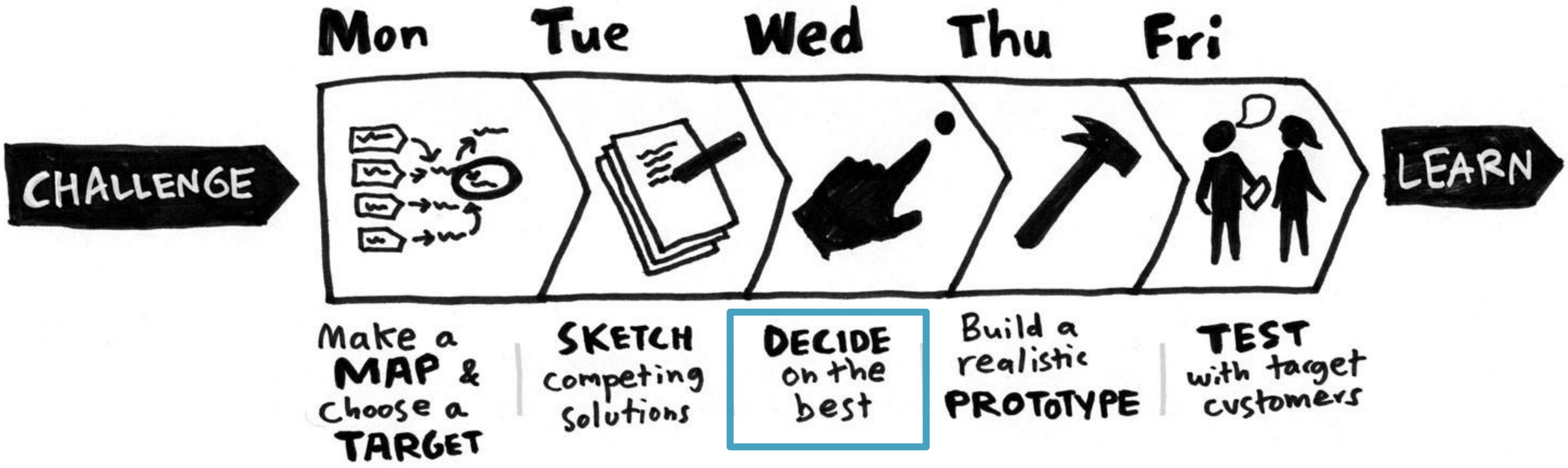
Click a button to annotate this diagram with additional data

States + More Data



Visualize states with data annotations. Flow between states is colored by intervening treatments.

Design Sprint



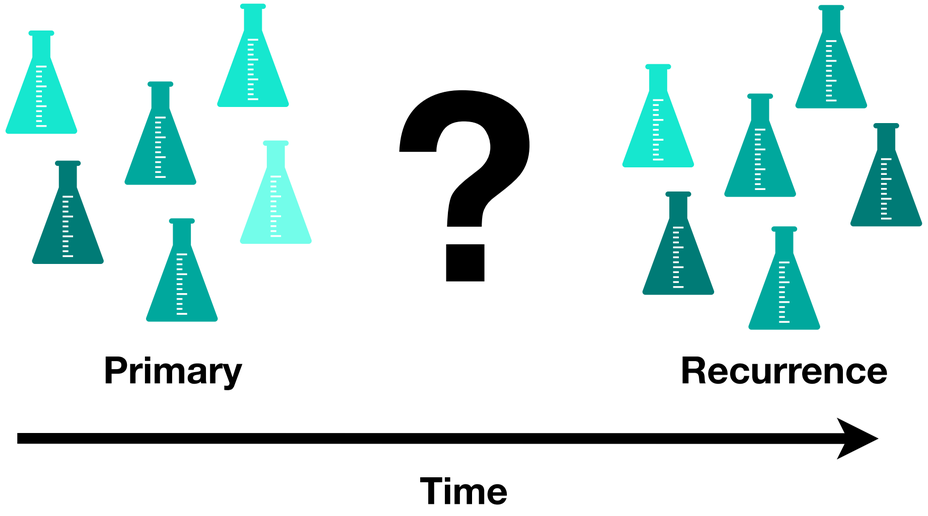
Test Study

- Study by Johnson et al., 2014, *Science*, 343(6167), 189–193.
- 23 patients with Glioma
- Samples of initial tumors and one or multiple recurrences
- Data: overall mutation count, mutations, neoplasm histologic grade, treatments,...

Many patients receiving the chemotherapy Temozolomide (TMZ) develop hypermutation and a high grade in the recurrence.

Your Goal

Use **OncoFlow** to identify any interesting differences between primary and recurrence samples



Launch OncoFlow

OncoFlow

Prototype demo

Signed in as Marco Polo

Select Feature:

Clinical Features

- Grade
- Treatment
- ...
- ...
- ...

Genomic Features

- Mutation Count
- IDH1
- ...
- ...
- ...

Timepoint

Patients (n = 23)

Heatmap | Flow

Summary

1 Grade

2 Grade

Prototype demo

Select Feature:

Clinical Features

- Grade
- Treatment
- ...
- ...
- ...

Genomic Features

- Mutation Count
- IDH1
- ...
- ...
- ...

Timepoint

Patients (n = 23)

Heatmap | Flow

Summary

1

Grade

#Mut

2

Grade

#Mut

You just added a continuous variable. Do you want to discretize it? Click the Cog Wheel

Prototype demo

Select Feature:

Clinical Features

- Grade
- Treatment
- ...
- ...
- ...

Genomic Features

- Mutation Count
- IDH1
- ...
- ...
- ...

Timepoint

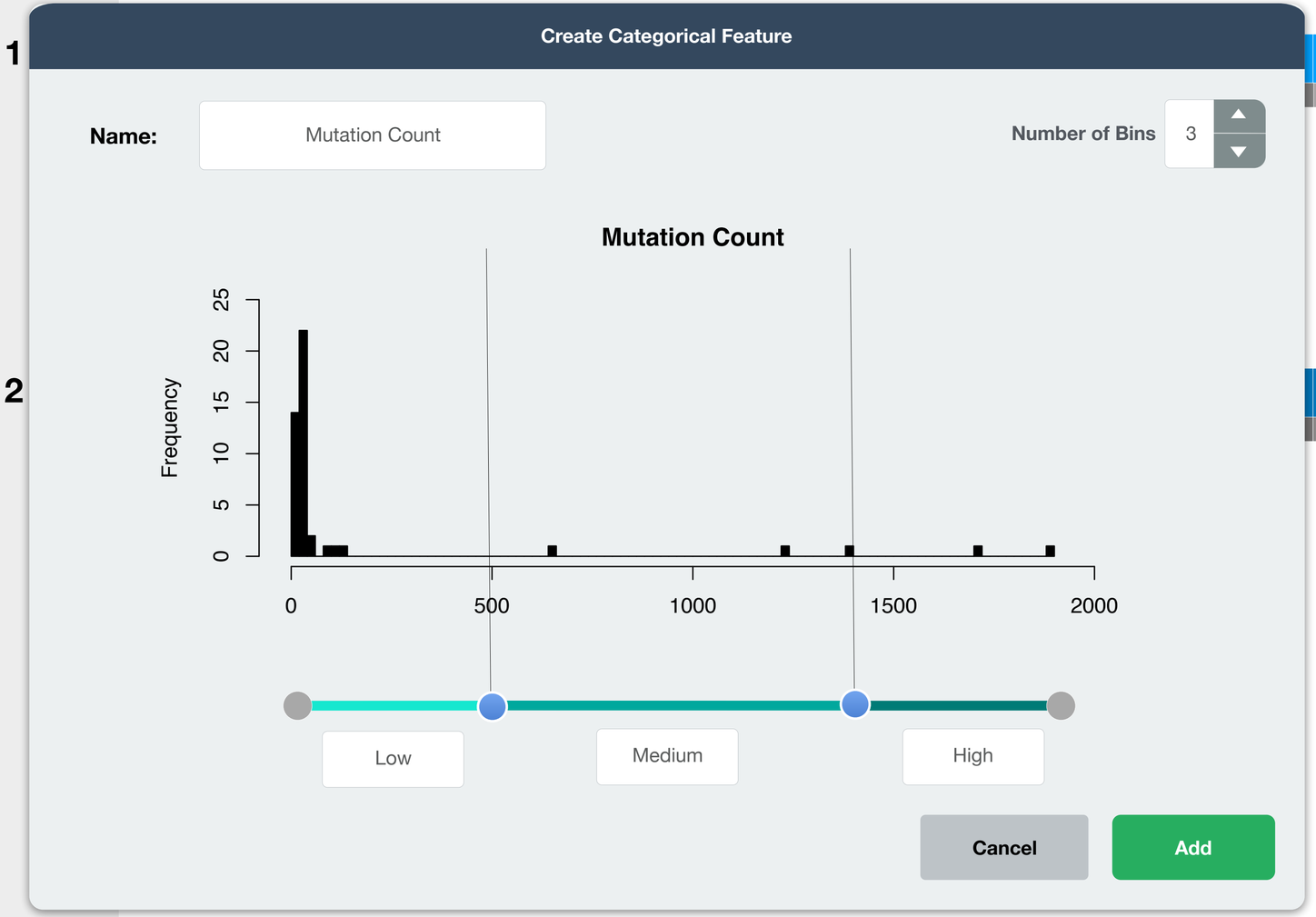
1

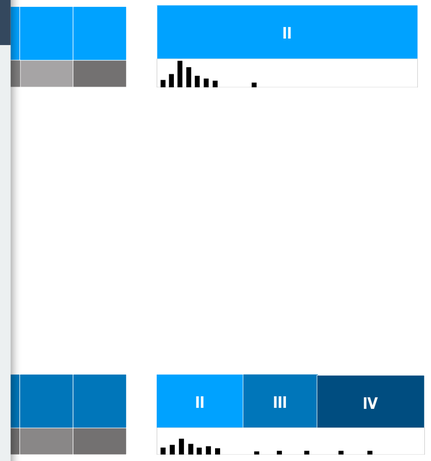
2

Patients (n = 23)

Heatmap | Flow

Summary





OncoFlow

Prototype demo

Signed in as Marco Polo

Select Feature:

Clinical Features

- Grade
- Treatment
- ...
- ...
- ...

Genomic Features

- Mutation Count
- IDH1
- ...
- ...
- ...

Timepoint

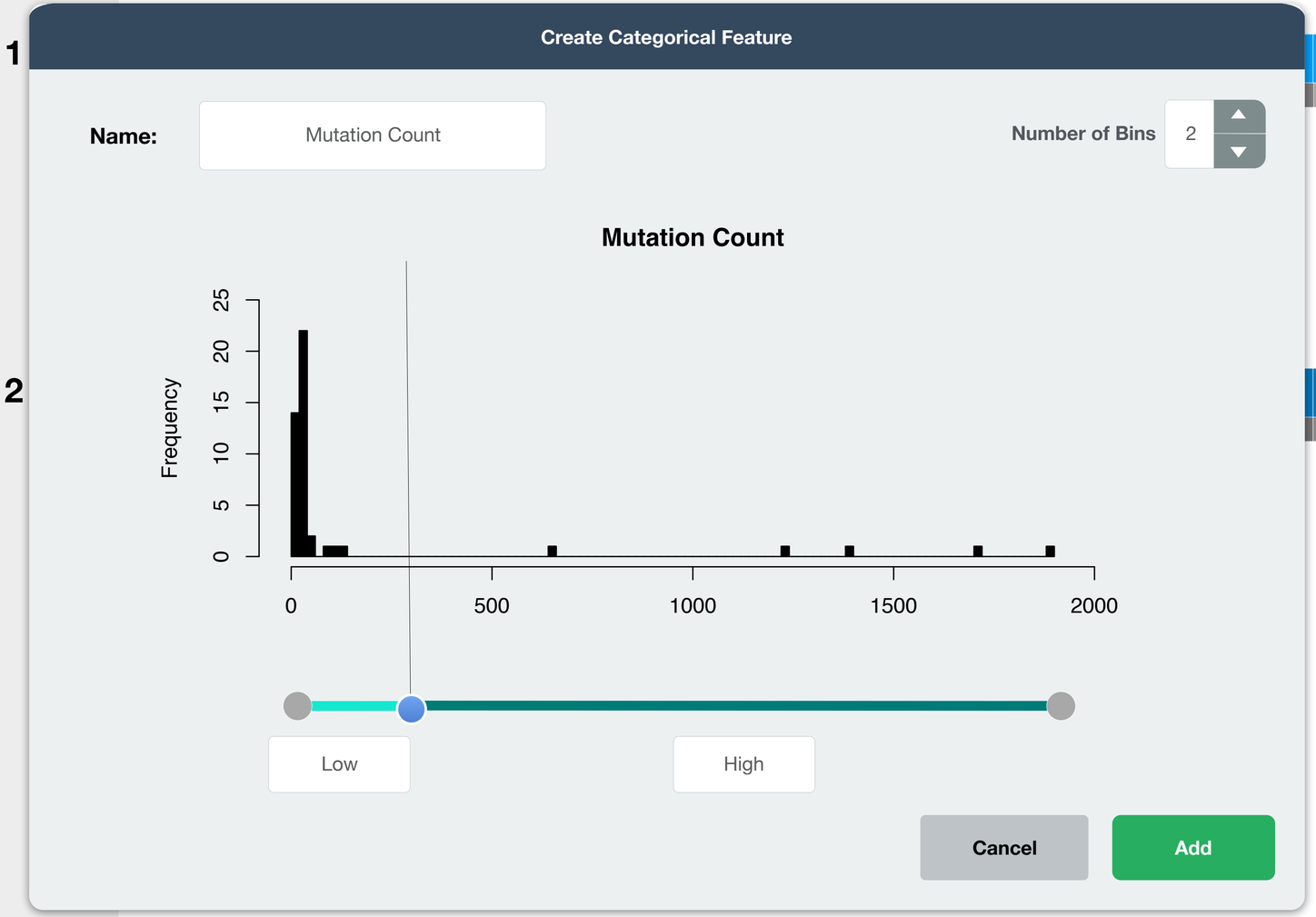
1

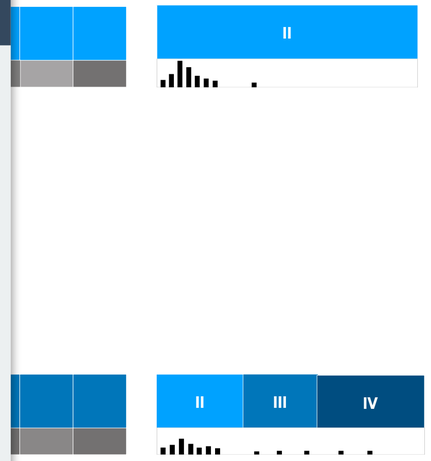
2

Patients (n = 23)

Heatmap | Flow

Summary





OncoFlow

Prototype demo

Signed in as Marco Polo

Select Feature:

Clinical Features

- Grade
- Treatment
- ...
- ...
- ...

Genomic Features

- Mutation Count
- IDH1
- ...
- ...
- ...

Timepoint

Patients (n = 23)

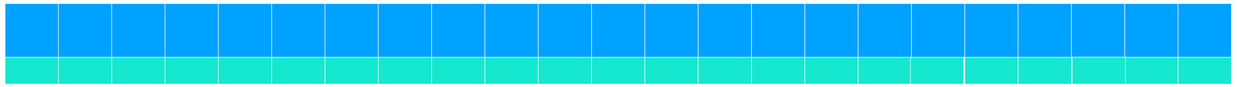
Heatmap | Flow

Summary

1

Grade

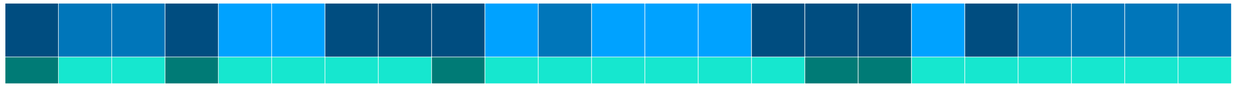
#Mut



2

Grade

#Mut



Prototype demo

Select Feature:

Clinical Features

- Grade
- Treatment
- ...
- ...
- ...

Genomic Features

- Mutation Count
- IDH1
- ...
- ...
- ...

Timepoint

Patients (n = 23)

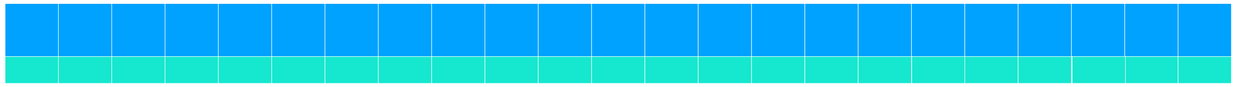
Heatmap | Flow

Summary

1

Grade

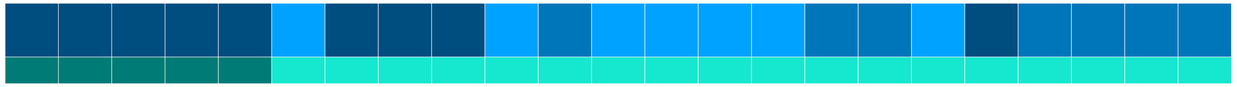
#Mut



2

Grade

#Mut



Prototype demo

Select Feature:

Clinical Features

- Grade
- Treatment
- ...
- ...
- ...

Genomic Features

- Mutation Count
- IDH1
- ...
- ...
- ...

Timepoint

Patients (n = 23)

Heatmap | Flow

Summary

1

Grade

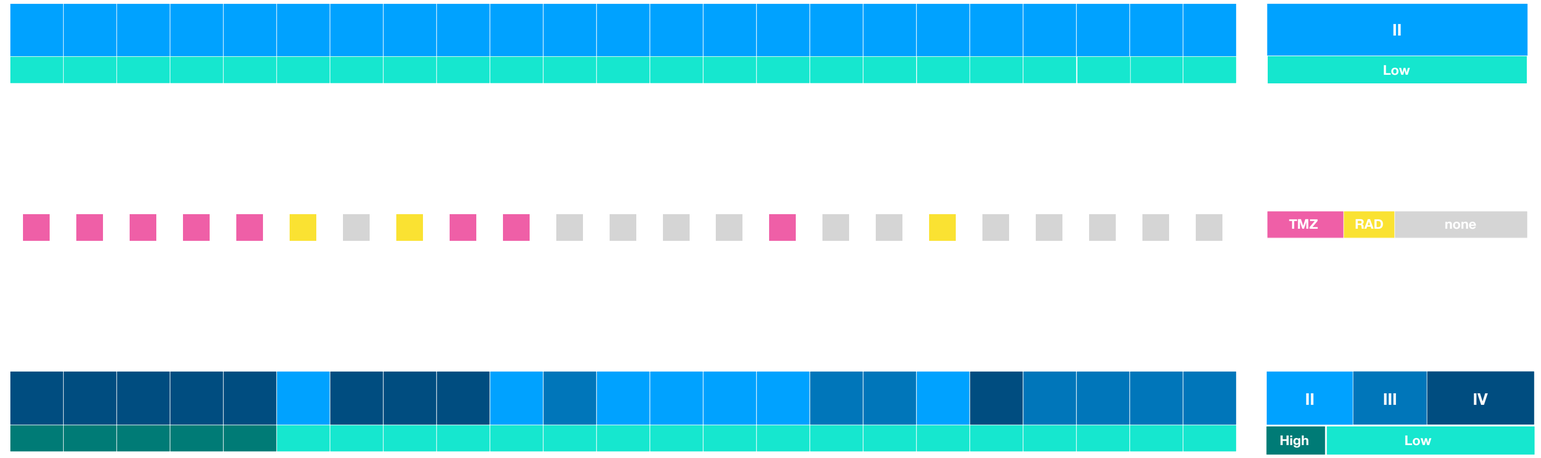
#Mut

Treatment

2

Grade

#Mut



OncoFlow

Prototype demo

Signed in as Marco Polo

Select Feature:

Clinical Features

- Grade
- Treatment
- ...
- ...
- ...

Genomic Features

- Mutation Count
- IDH1
- ...
- ...
- ...

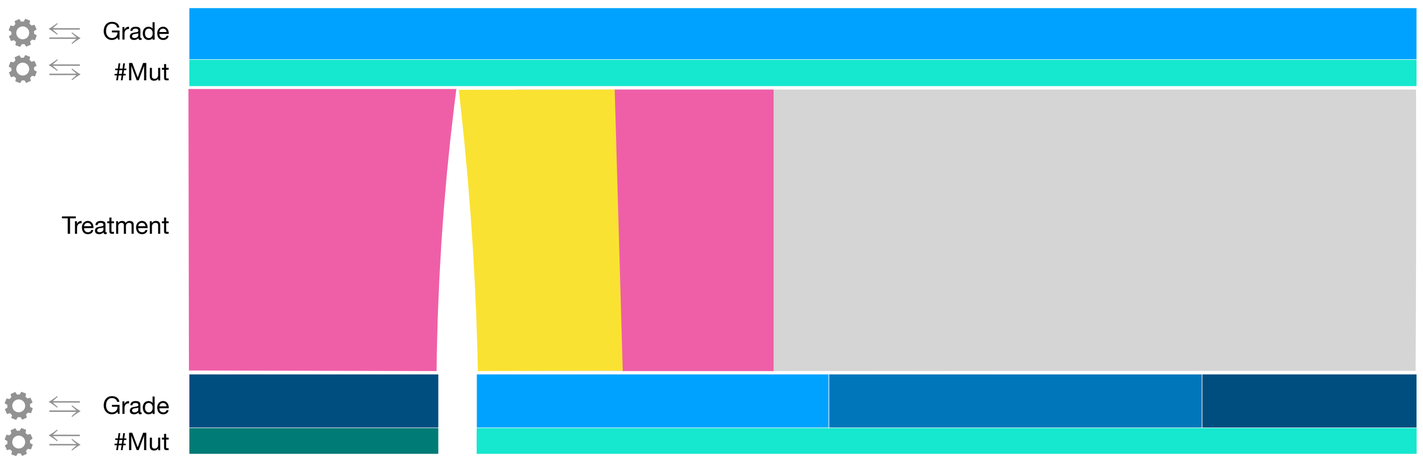
Timepoint

1

2

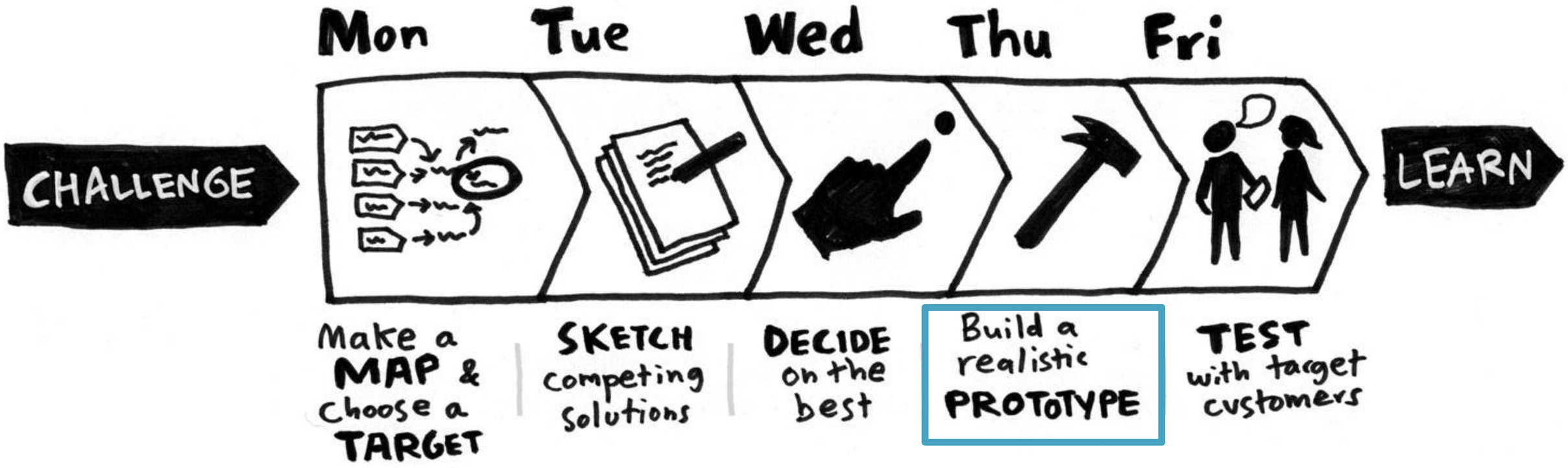
Patients (n = 23)

Heatmap | Flow



Summary

Design Sprint



Knapp, Jake, John Zeratsky, and Braden Kowitz. Sprint: How to solve big problems and test new ideas in just five days. Simon and Schuster, 2016.

User Feedback

Select Feature:

Clinical Features

- Grade
- Treatment
- ...
- ...
- ...

Genomic Features

- Mutation Count
- IDH1
- ...
- ...
- ...

+ liked sidebar

+ adding rows intuitive

Timepoint

Patients (n = 23)

Heatmap | Flow

Summary

1

Grade

#Mut

II

- can't find button

- what is this?

- same patient?

2

Grade

#Mut

II III IV

- meaning of buttons

You just added a continuous variable. Do you want to discretize it? Click the Cog Wheel

- is discretize the right expression?

+ understands patient alignment

+ understands legend

OncoFlow

Prototype demo

Signed in as Marco Polo

Select Feature:

Clinical Features

- Grade
- Treatment
- ...
- ...
- ...

Genomic Features

- Mutation Count
- IDH1
- ...
- ...
- ...

Timepoint

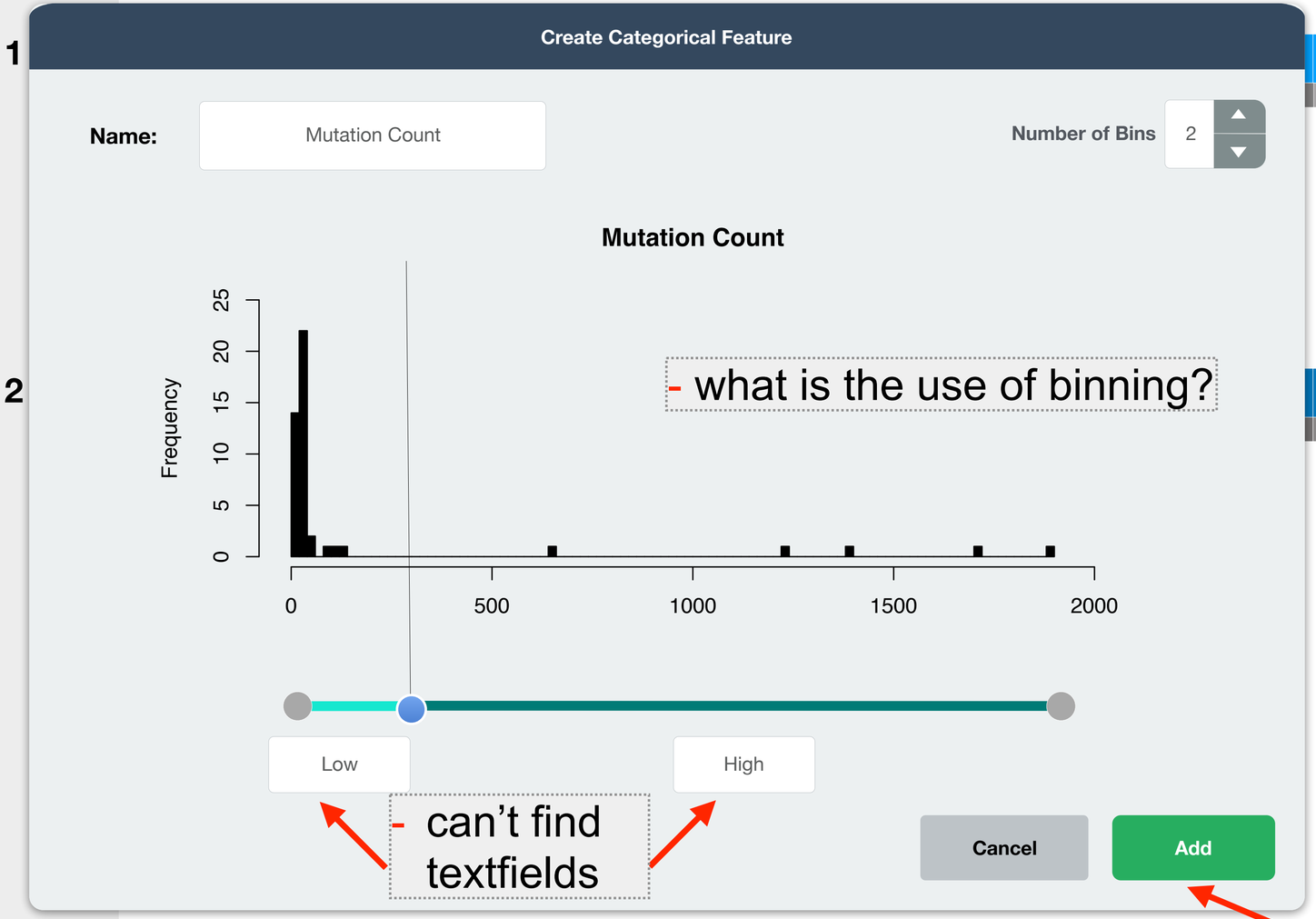
1

2

Patients (n = 23)

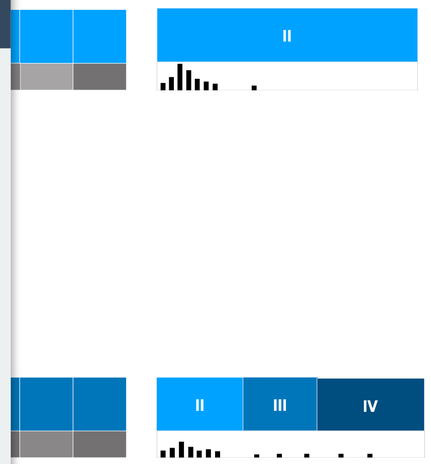
Heatmap | Flow

Summary



- what is the use of binning?

- can't find textfields



+ gets binning

- is "add" the right word

Prototype demo

Select Feature:

Clinical Features

- Grade
- Treatment
- ...
- ...
- ...

Genomic Features

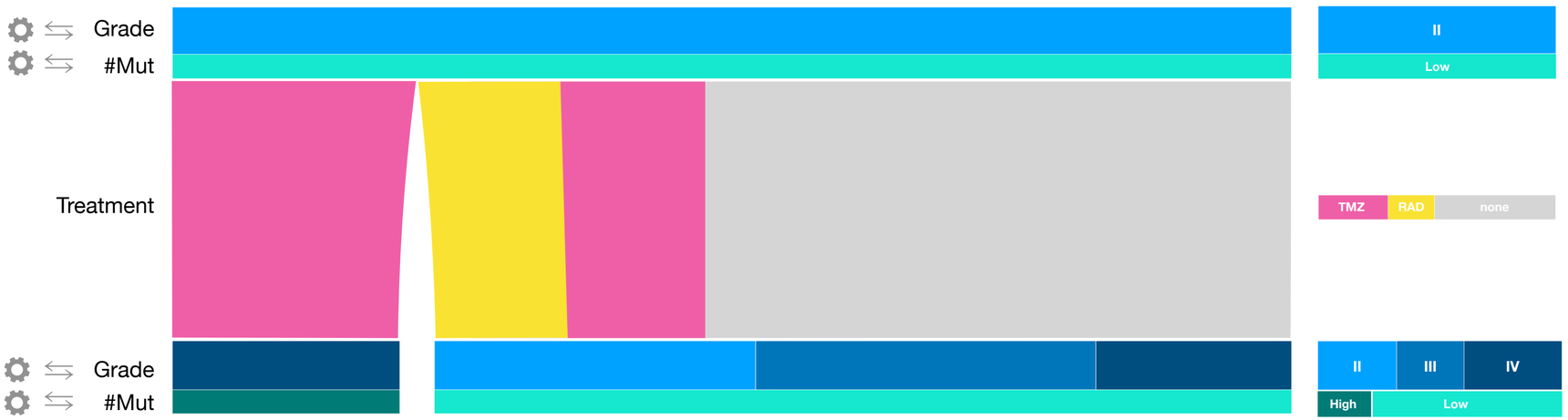
- Mutation Count
- IDH1
- ...
- ...
- ...

Timepoint

Patients (n = 23)

Heatmap | **Flow**

Summary



- looks out of sync with heat map

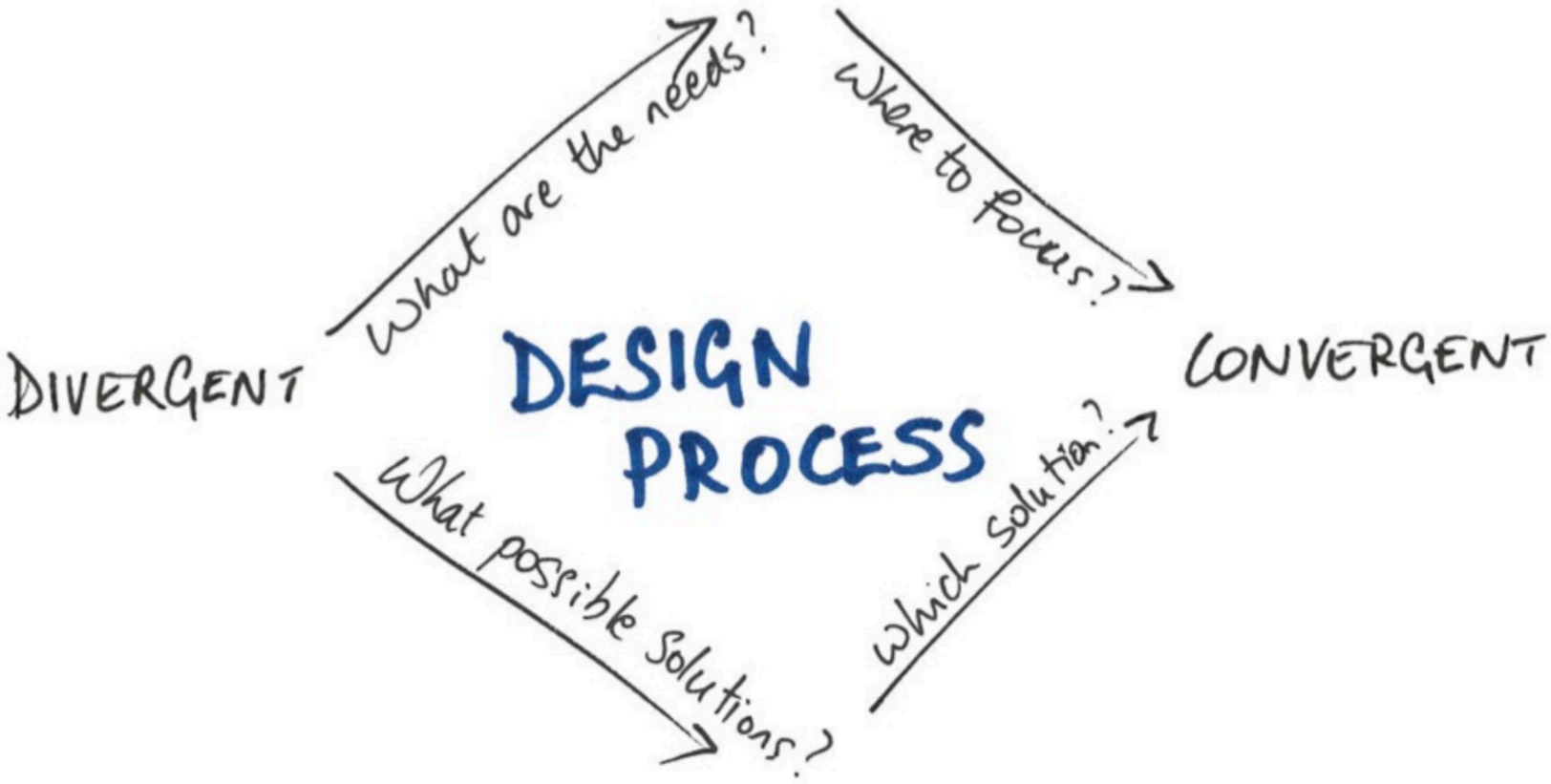
- what are the flows? Where is the single patient?

- does not add any insight

Demo

http://oncothreads.gehlenborglab.org

Five Design-Sheet Methodology



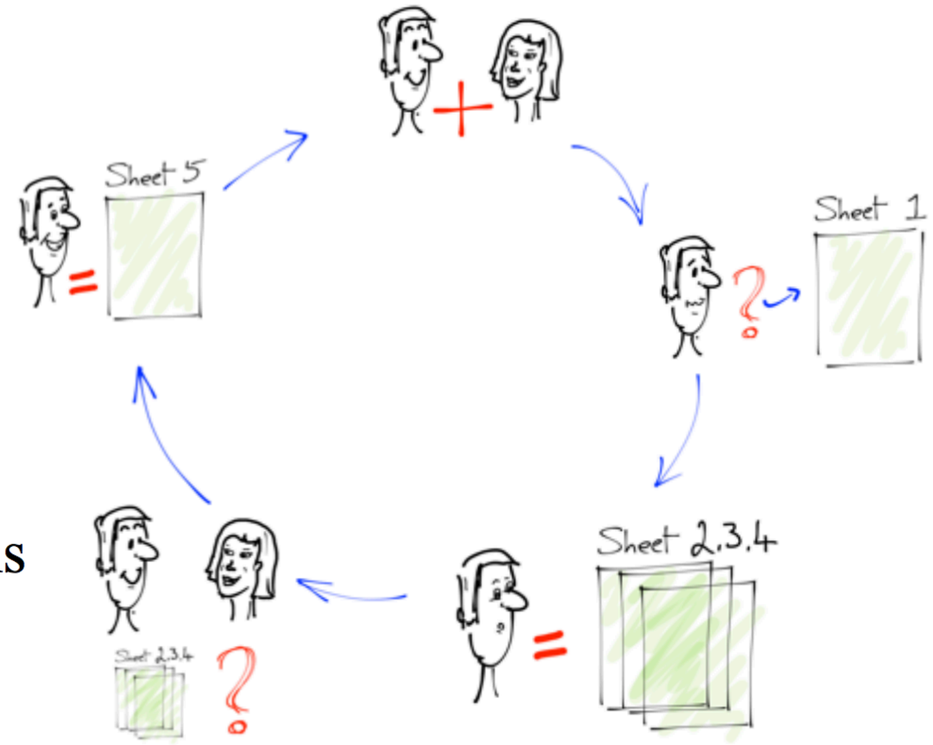
http://fds.design

Five Design-Sheet Methodology

Five sheets

Five parts per sheet

Five parts to the process

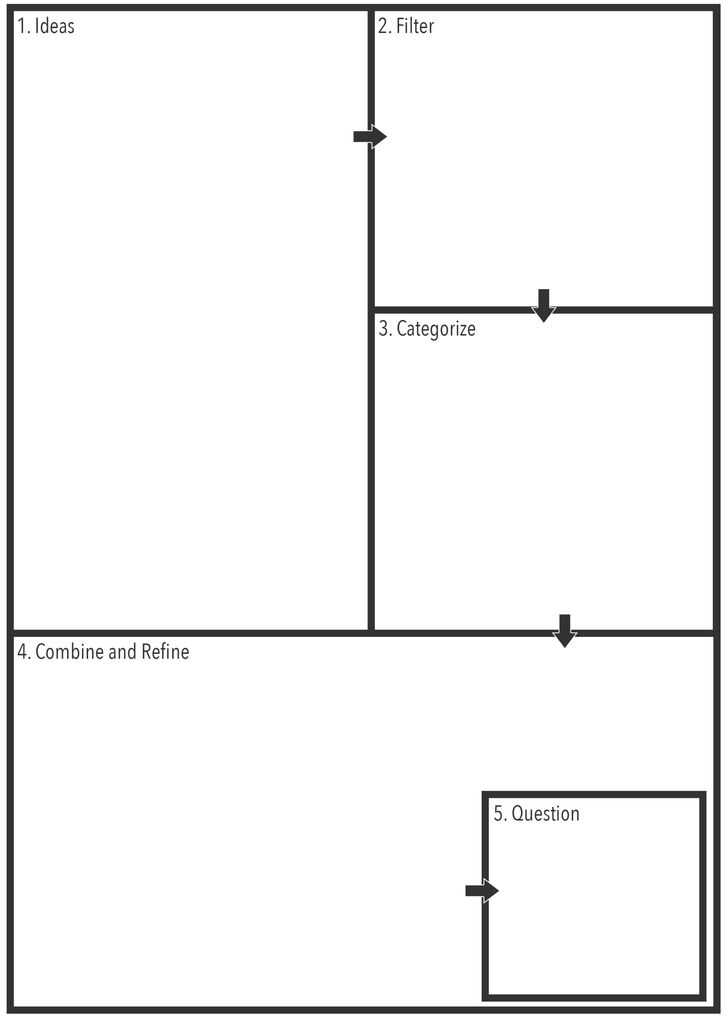


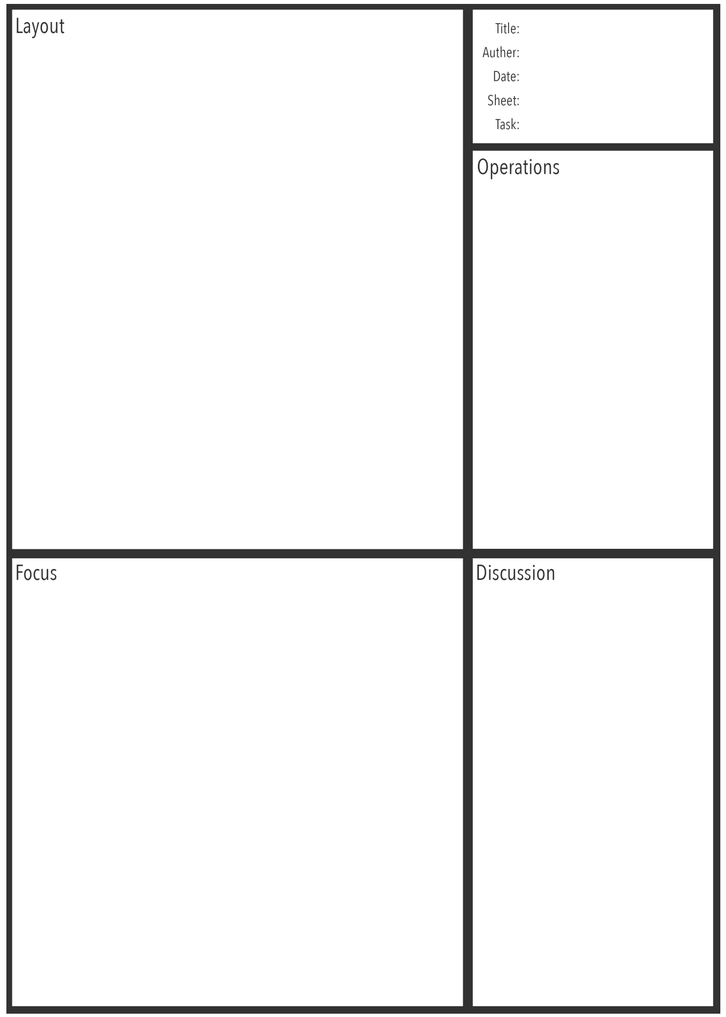
Sheet 1 = Ideas

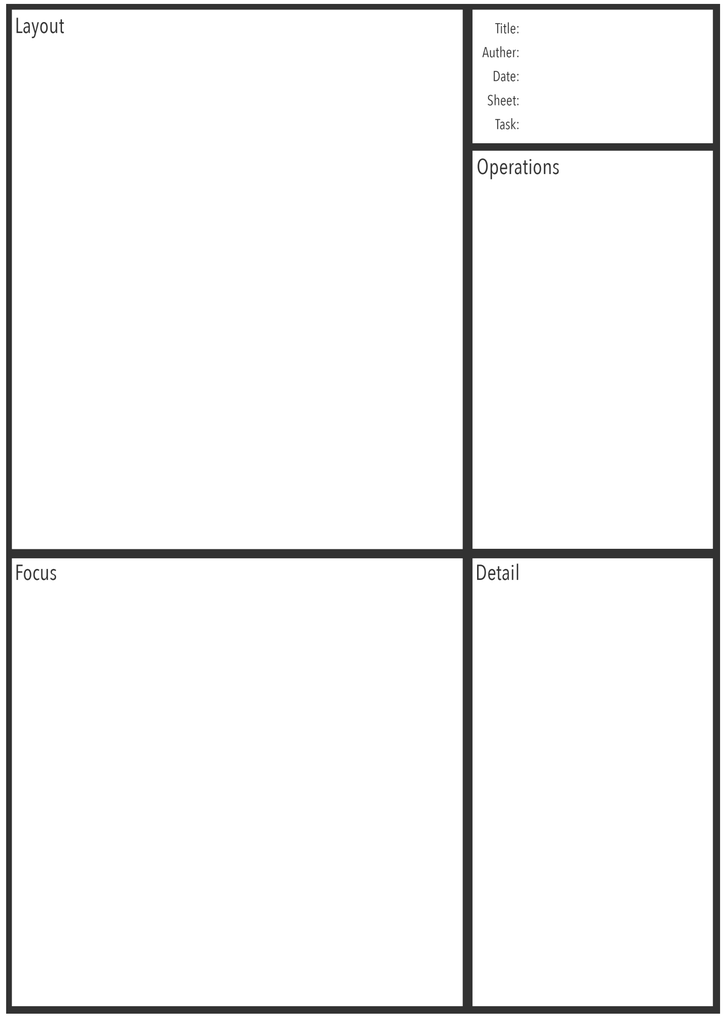
Sheets 2,3,4 = Main designs

Sheet 5 = Realization

Five Design-Sheet Methodology







http://fds.design

Five Design-Sheet Methodology

The Five Design Sheet Methodology

Jonathan Roberts, Chris Headleand, Panos Ritsos

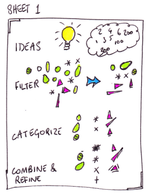
Bangor University.

The Five Design Sheet Methodology was first published in 2011; it has now been used for over 5 years by students, company engagement and for website design. In this study we are investigating how the FdS has been used and how it has been beneficial.

The FdS consists of 5 sheets. Each sheet consists of 5 parts, which aids the user to remember what structure to follow and what information to place in each sheet.

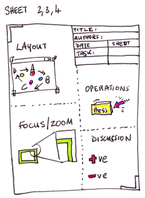
Sheet 1.
The Ideas sheet, that allows users to think about the different design ideas, and to consider many possible outcomes.

Ideas,
Filter,
Categorize,
Combine/Refine,
Question.



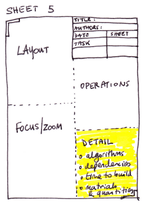
Sheets 2, 3, 4
Three design sheets; these are alternative and potential design outcomes, and they sketch what the proposed system would look like.

Layout,
Focus/zoom,
Info, Operations,
Discussion

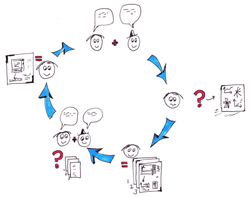


Sheet 5.
The final realization sheet that would be implemented.

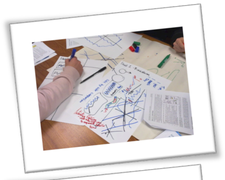
Layout,
Focus/zoom,
Info,
Operations,
Detail



Roberts, J. C. (2011). The Five Design-Sheet (FdS) approach for Sketching Information Visualization Designs. In *Eurographics 2011-Education Papers* (pp. 27-41). The Eurographics Association.



The FdS is part of a process (from the top), meeting with the client, thinking of design ideas, creating design sheets (2,3,4), discussing with client, and implementing system, before meeting with client and starting the cycle again, to refine the system.



"Lo-Fi prototyping works because it effectively educates developers to have a concern for usability and formative evaluation, and because it maximizes the number of times you get to refine your design before you must commit to code."
Marc Rettig, CACM, 37:4, 1994

http://fds.design

Design Exercise: Flu Dashboard

- **Audience 1: Public Health Researchers (2 teams)**
  - Visualization literacy?
  - Data literacy?
  - Tasks?
- **Audience 2: Travelers (2 teams)**
  - Visualization literacy?
  - Data literacy?
  - Tasks?
- **Use Five Design-Sheet Methodology (45 minutes)**

Design Exercise: Flu Dashboard

- **Audience 1: Public Health Researchers (2 teams)**
  - Visualization literacy? high
  - Data literacy? high
  - Tasks?
    - Need up-to-date data
    - See how patterns change week by week
    - Identify outbreaks based on patterns in the data: visually and algorithmically
    - Check effectiveness of flu vaccines (data: low, medium, high)
    - information about strains (data: percentage per strain, adding to 100)

Design Exercise: Flu Dashboard

- **Audience 1: Public Health Researchers (2 teams)**
  - compare current year to previous years
  - compare geographic locations to each other: by country
  - export data and figures
  - see age information and death information (per case age in years, per country # of flu deaths)
  - incorporate population size of each country and number of cases tested vs confirmed (pop. per country, per country per week # cases tested)
  - number of hospitalizations (per country per week # of cases)
  - number of doses of flu meds in stock (per country per week # doses)

Design Exercise: Flu Dashboard

- **Audience 2: Travelers (2 teams)**
  - Visualization literacy? low
  - Data literacy? low
  - Tasks?
    - which countries have the highest/lowest number of cases? (over time)
    - during what time of the year do the cases peak?
    - how does the number of cases change over the course of a year?
    - what will the situation be like at time point X? (time range X->Y)?
    - do I need to get vaccinated before I travel? (recommendation: yes/no)

Design Exercise: Flu Dashboard

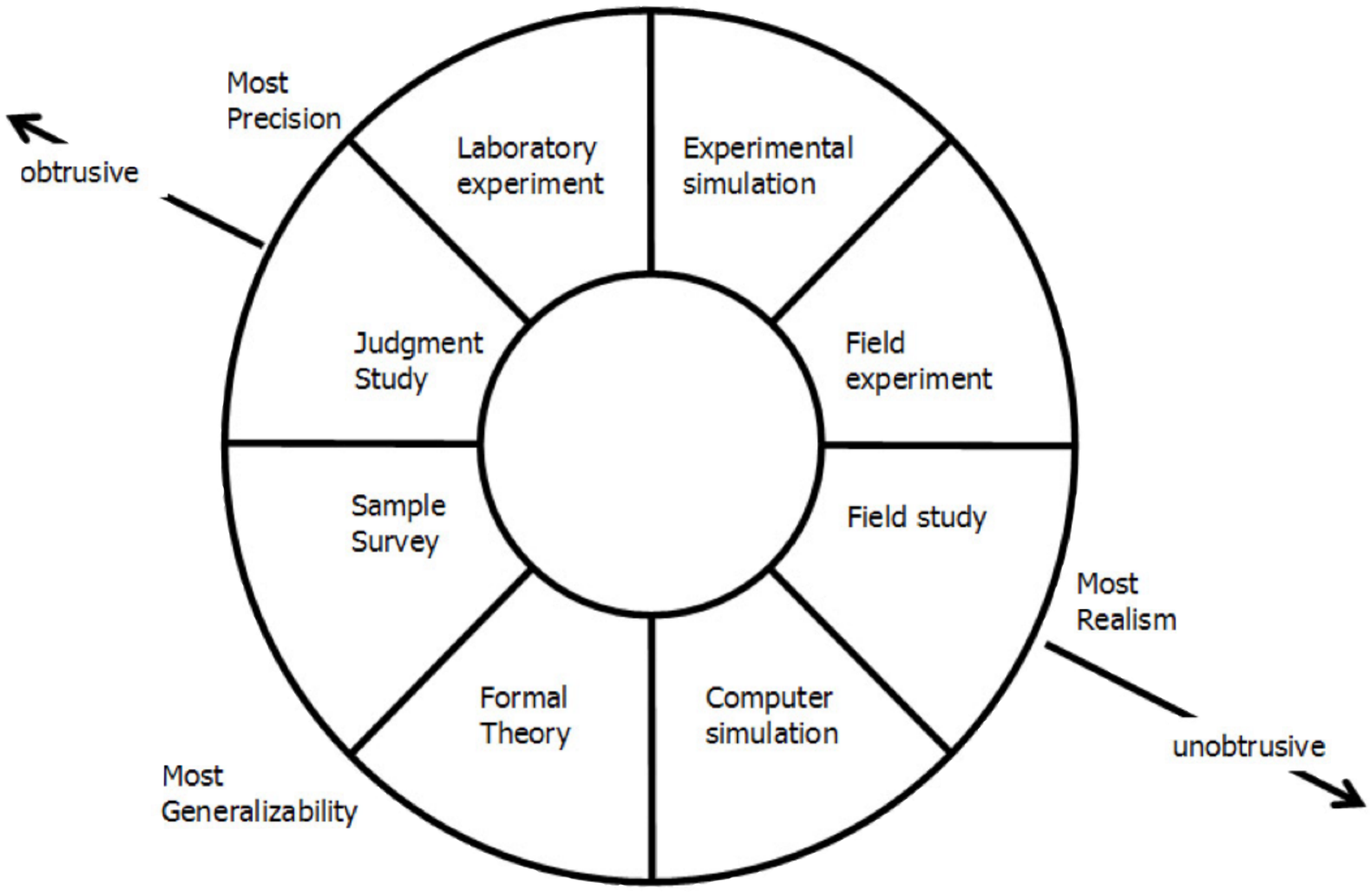
- **Audience 2: Travelers (2 teams)**
  - what strains are currently in circulation and does my vaccine (?) cover that? (data: percentage per strain, adding to 100; yes/no for vaccine)
  - What is my risk of having severe complications if I get the flu? (low, medium, high)
  - What medical care would be available in the country? (low/medium/high per country)
  - How “bad” is the most common strain in that country? (per strain: not bad, bad, very bad)
  - Are there any travel restrictions for that country? (yes/no)

Creative Visualization-Opportunities Workshops

- **http://vdl.sci.utah.edu/CVOWorkshops/**
  - Developed by researchers at SCI (University of Utah)
  - Specifically for visualization design

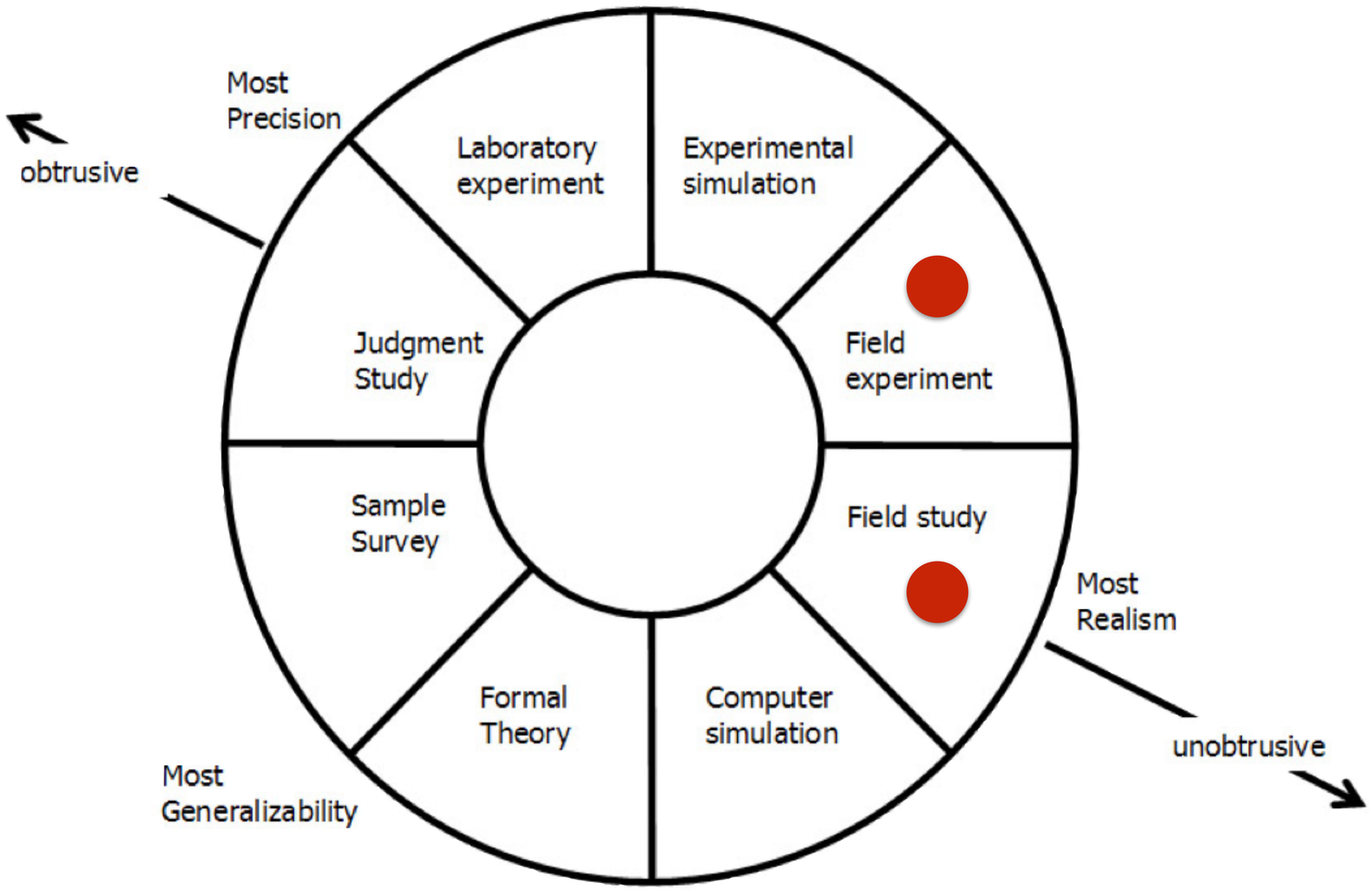
Evaluation

Validation Techniques



Carpendale 2008, Evaluating Information Visualizations

Validation Techniques



Carpendale 2008, Evaluating Information Visualizations

Qualitative Usability Testing

- **Field study**
  - Intended to be unobtrusive
  - Conducted in the actual situation (a user with their own analysis tasks and data)
  - Watching someone use your tool can be very informative; you may be surprised!

Carpendale 2008, Evaluating Information Visualizations

Qualitative Usability Testing

- **Field experiment**
  - Design user tasks to simulate real analyses
  - Recruit group of users and arrange one-on-one sessions
  - Encourage thinking aloud and note top usability issues

Carpendale 2008, Evaluating Information Visualizations

Qualitative Usability Testing

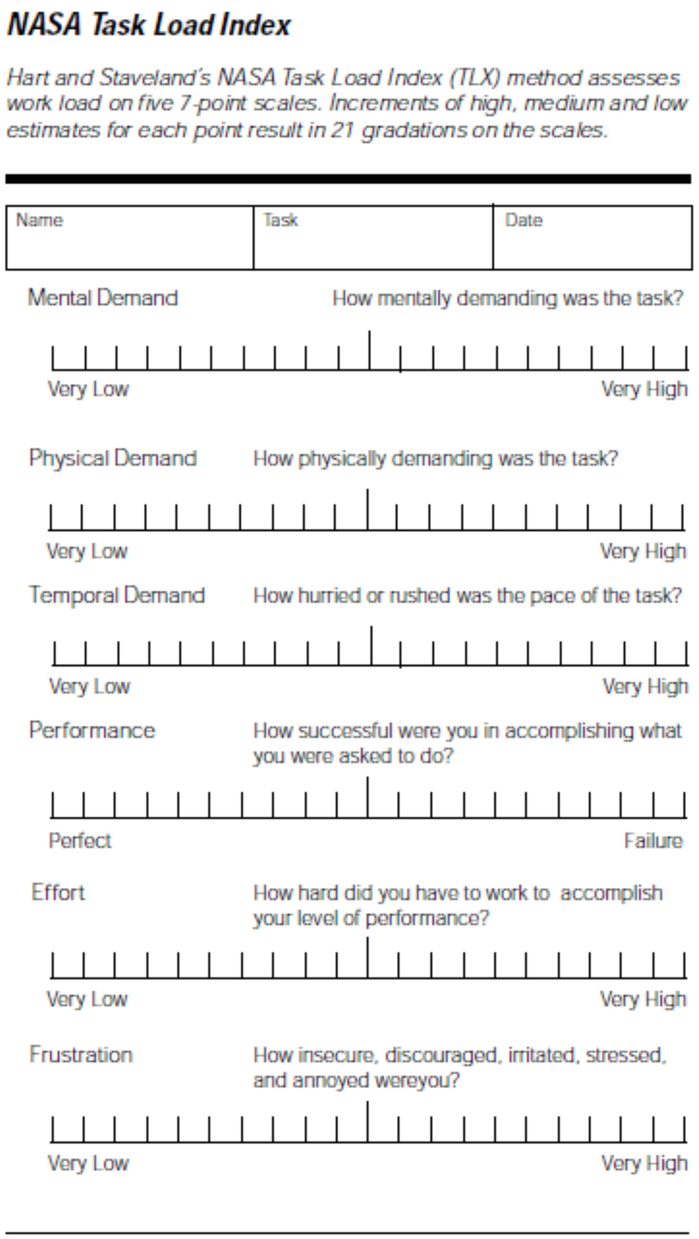
- Learn through observation
- Learn about behavior, not just opinions
- Not about proof, but about insight and context
- Test early and test often
- **Important that your users are assured that you are testing the software, not them**

collected by Francis Rowland, EMBL-EBI

Quantitative User Testing

- Error Rate
- Work Load (NASA-TLX)
- Time (see LineUp paper)
- Insights
  - **Oh boy!**

NASA TLX



NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

https://upload.wikimedia.org/wikipedia/commons/f/fc/NasaTLX.png

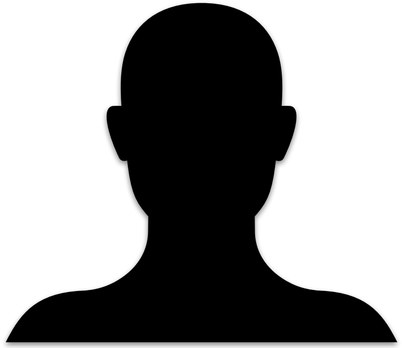
Evaluation

What is an effective visualization and what is not?

Considerations - e.g. measuring effectiveness quantitatively is extremely problematic

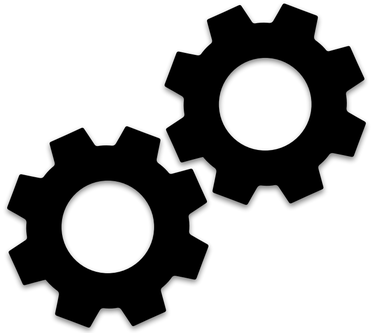
Evaluation

1



human aspect

2



system complexity

3

? vs !

outcome is question,
not answer

Evaluation

What is an effective visualization and what is not?

Considerations - e.g. measuring effectiveness quantitatively is extremely problematic

Impact - e.g. on developers (are we doing good work?), users (is this a good tool?), and reviewers (should this be published/funded?), e.g. in evaluating visual against analytical approaches

Tabular Data

Arrange Tables

Express Values

Separate, Order, Align Regions

- Separate

- Order

- Align

- *1 Key*
  List

- *2 Keys*
  Matrix

- *3 Keys*
  Volume

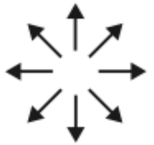
- *Many Keys*
  Recursive Subdivision

Axis Orientation

- Rectilinear

- Parallel

- Radial



Layout Density

- Dense

- Space-Filling

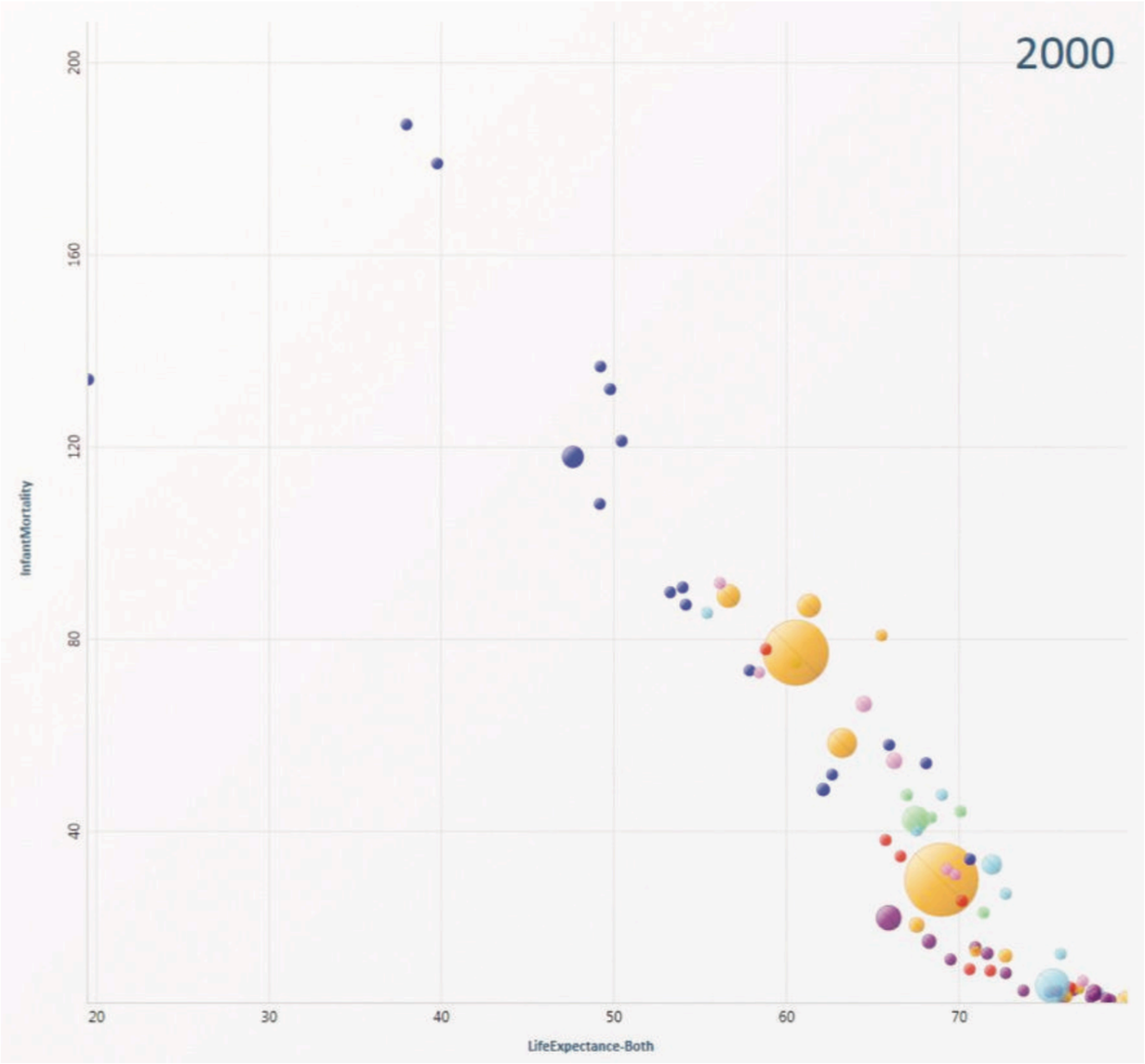


Figure 7.2. Scatterplot. Each point mark represents a country, with horizontal and vertical spatial position encoding the primary quantitative attributes of life expectancy and infant mortality. The color channel is used for the categorical country attribute and the size channel for quantitative population attribute. From [Robertson et al. 08, Figure 1c].

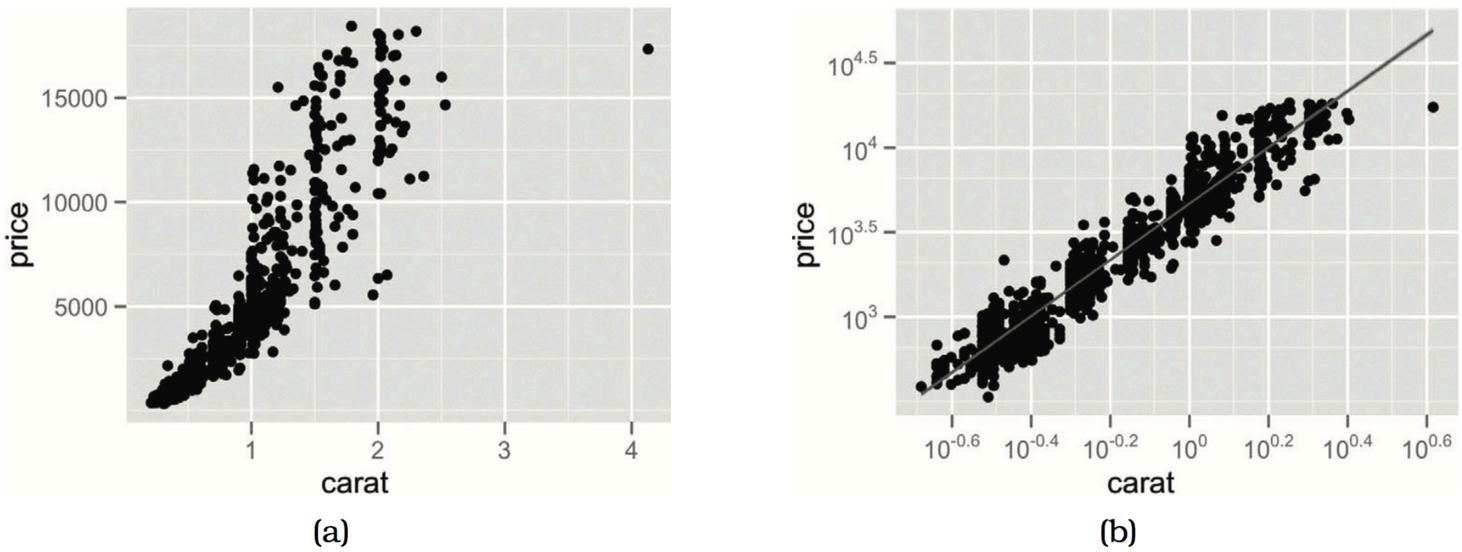
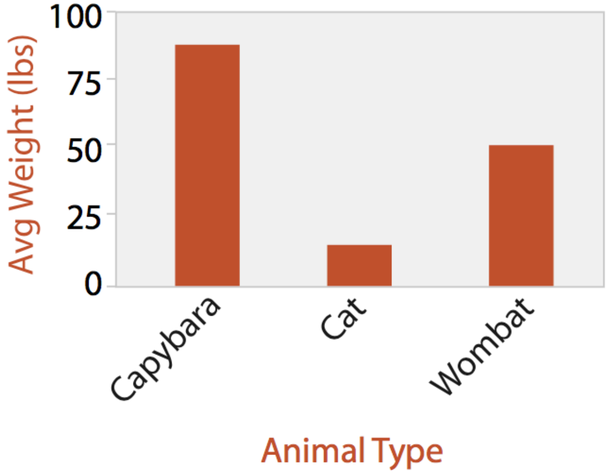
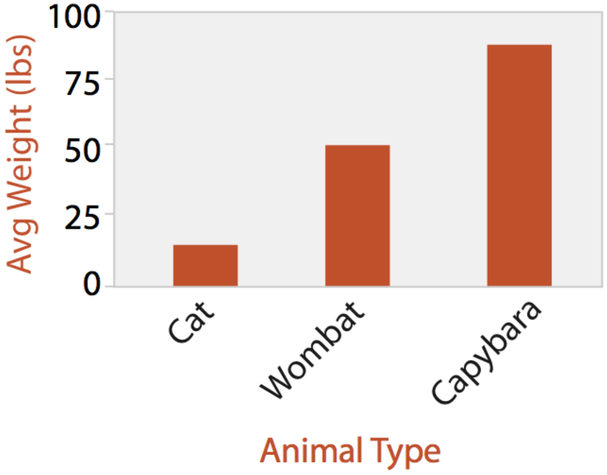


Figure 7.3. Scatterplots. (a) Original diamond price/carat data. (b) Derived log-scale attributes are highly positively correlated. From [Wickham 10, Figure 10].


100
75
50
25
0
Avg Weight (lbs)
Capybara
Cat
Wombat
Animal Type


100
75
50
25
0
Avg Weight (lbs)
Cat
Wombat
Capybara
Animal Type

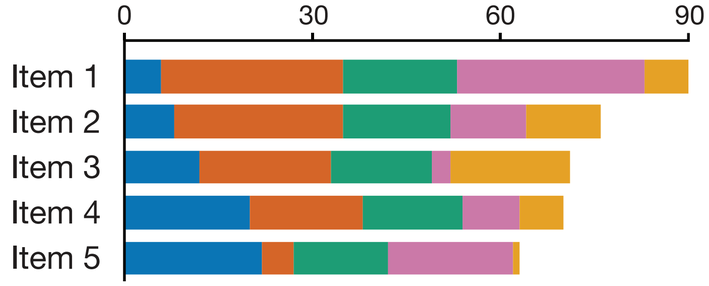
Bar Charts for Items & Categories

	1	2	3	4	5
Item 1	6	29	18	30	7
Item 2	8	27	17	12	12
Item 3	12	21	16	3	19
Item 4	20	18	16	9	7
Item 5	22	5	15	20	1

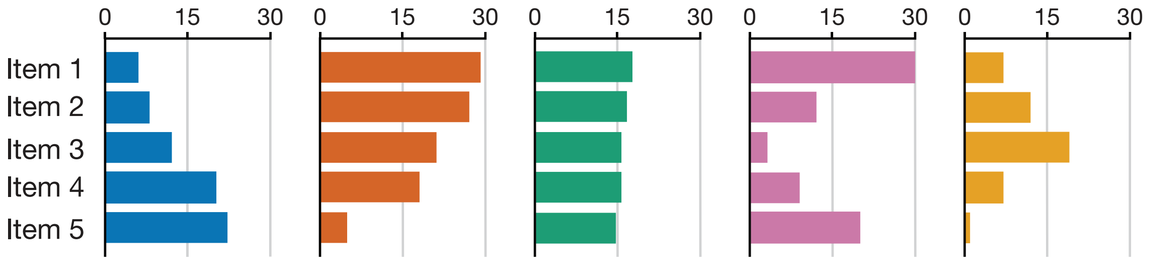
Streit and Gehlenborg, Nature Methods, 2014

Bar Charts for Items & Categories

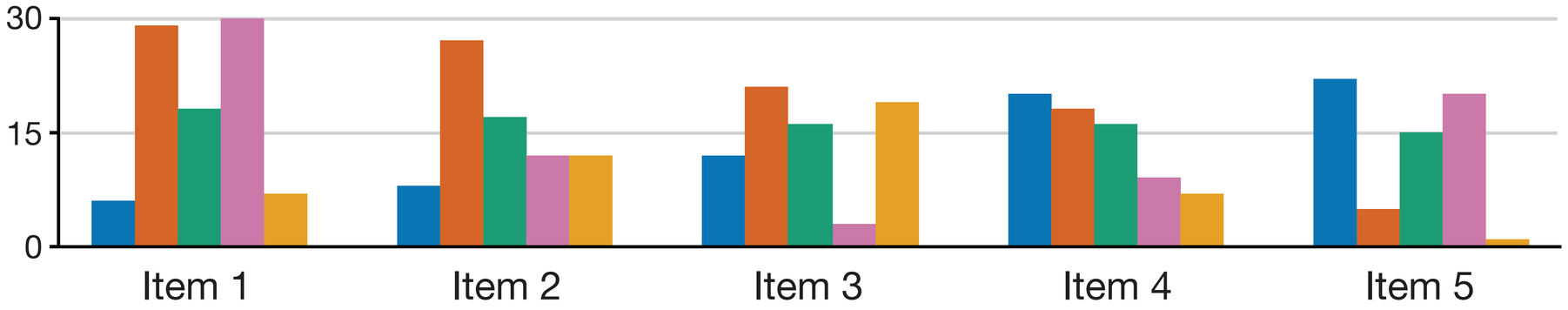
Stacked Bar Chart



Layered Bar Chart



Grouped Bar Chart



Streit and Gehlenborg, Nature Methods, 2014

Table Lens

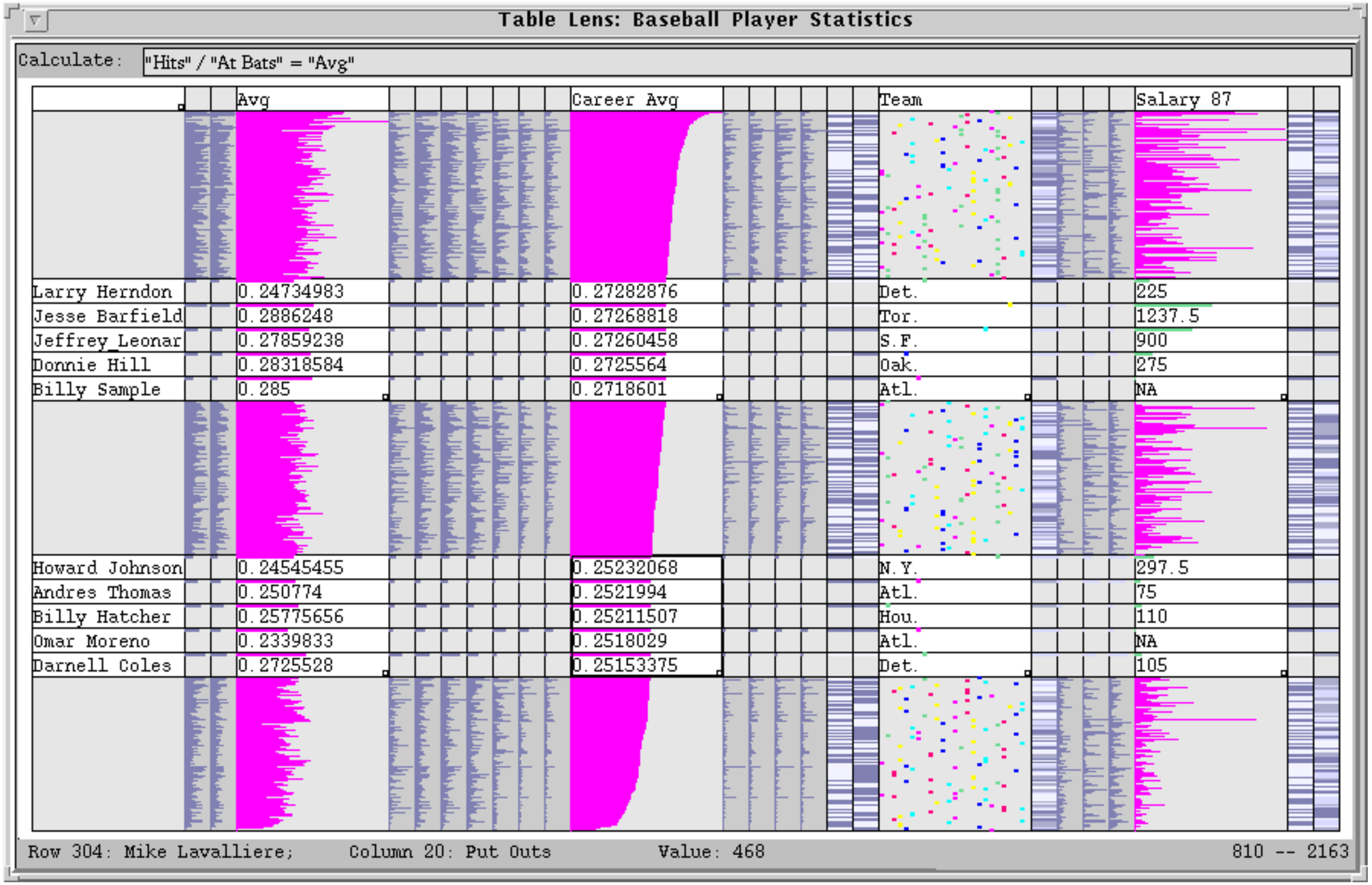


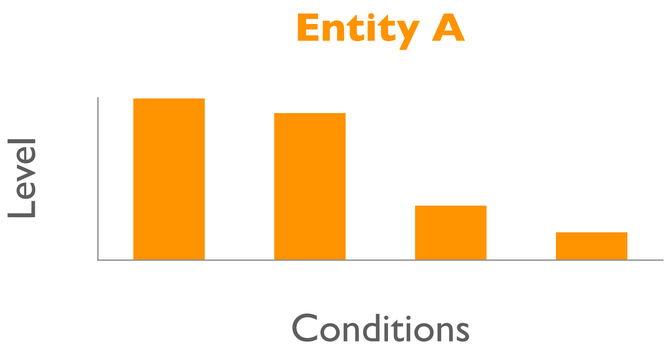
Table Lens (Rao & Card), 1994

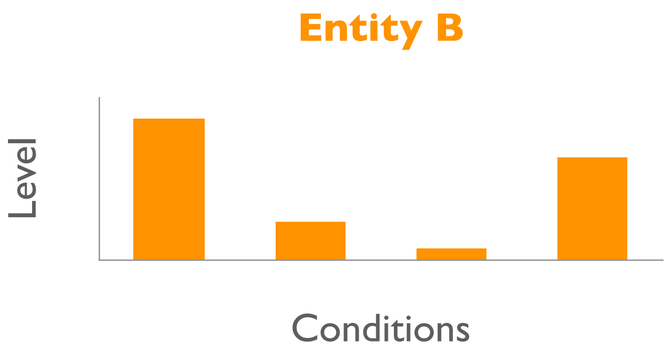
Multivariate Data

Homogeneous Tables

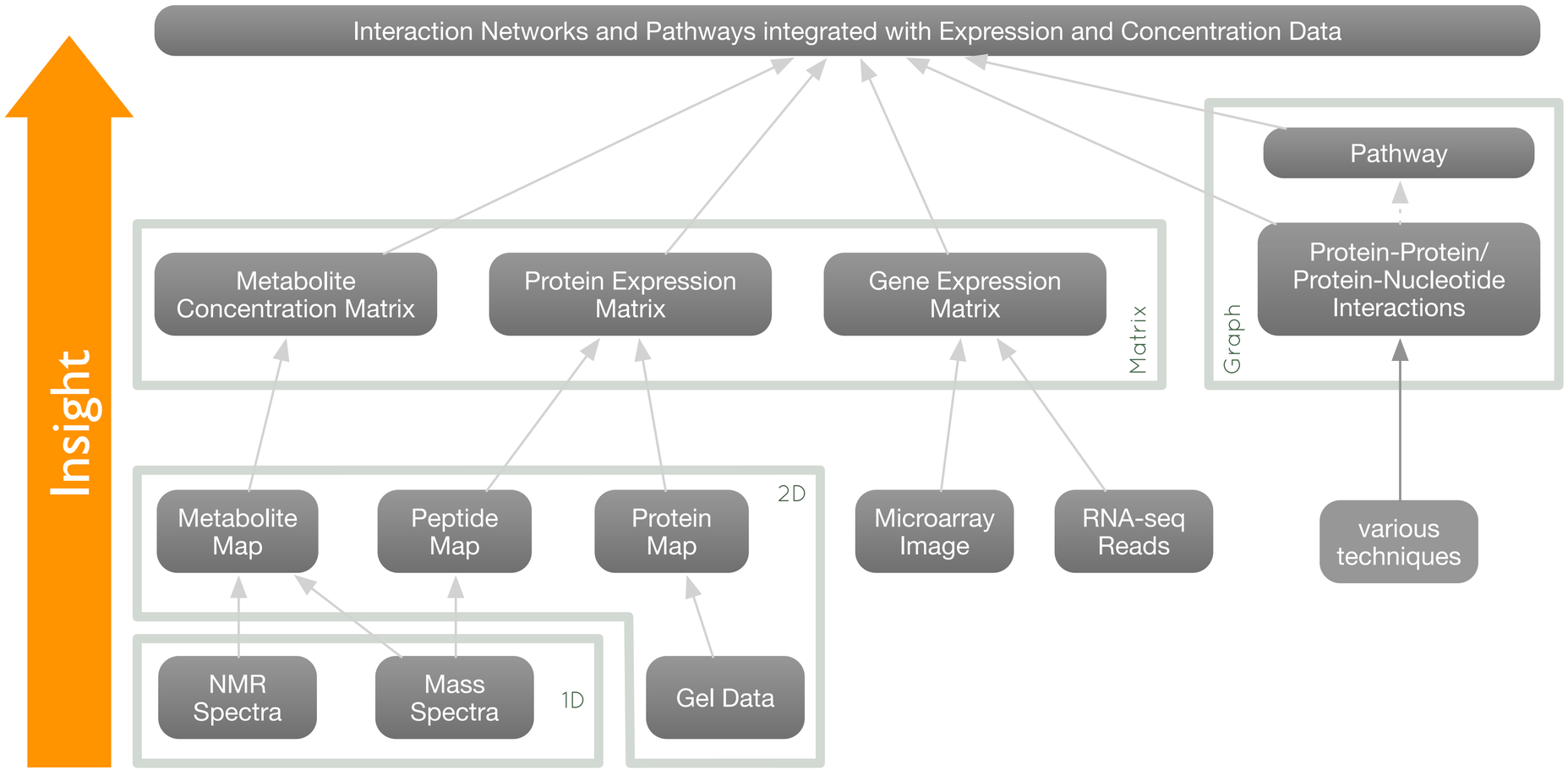
Multivariate Data

- typical “omics” data: transcriptomics, proteomics, metabolomics
- expression/concentration levels of many biological entities (transcripts, proteins, etc.) across many different conditions/time points
- entity levels measured per sample on a “genome-wide” scale
- often entities are not measured directly





Multivariate Data

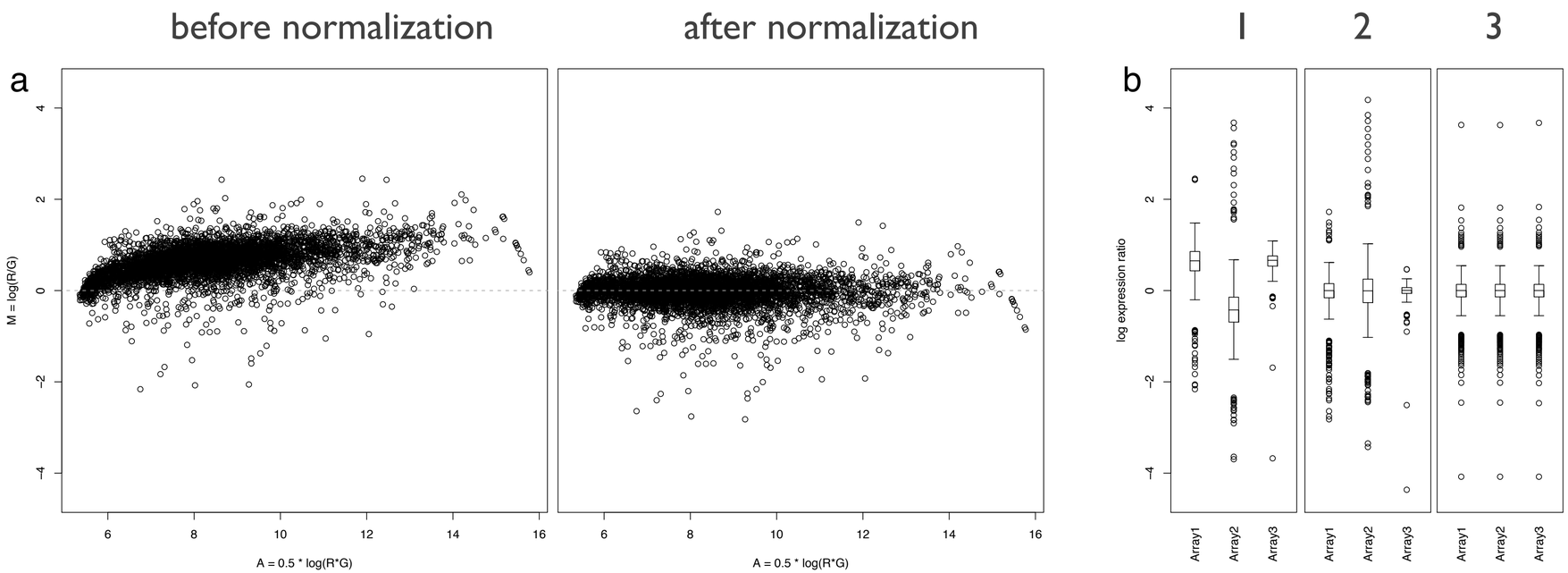


Gehlenborg et al. 2010, Nature Methods

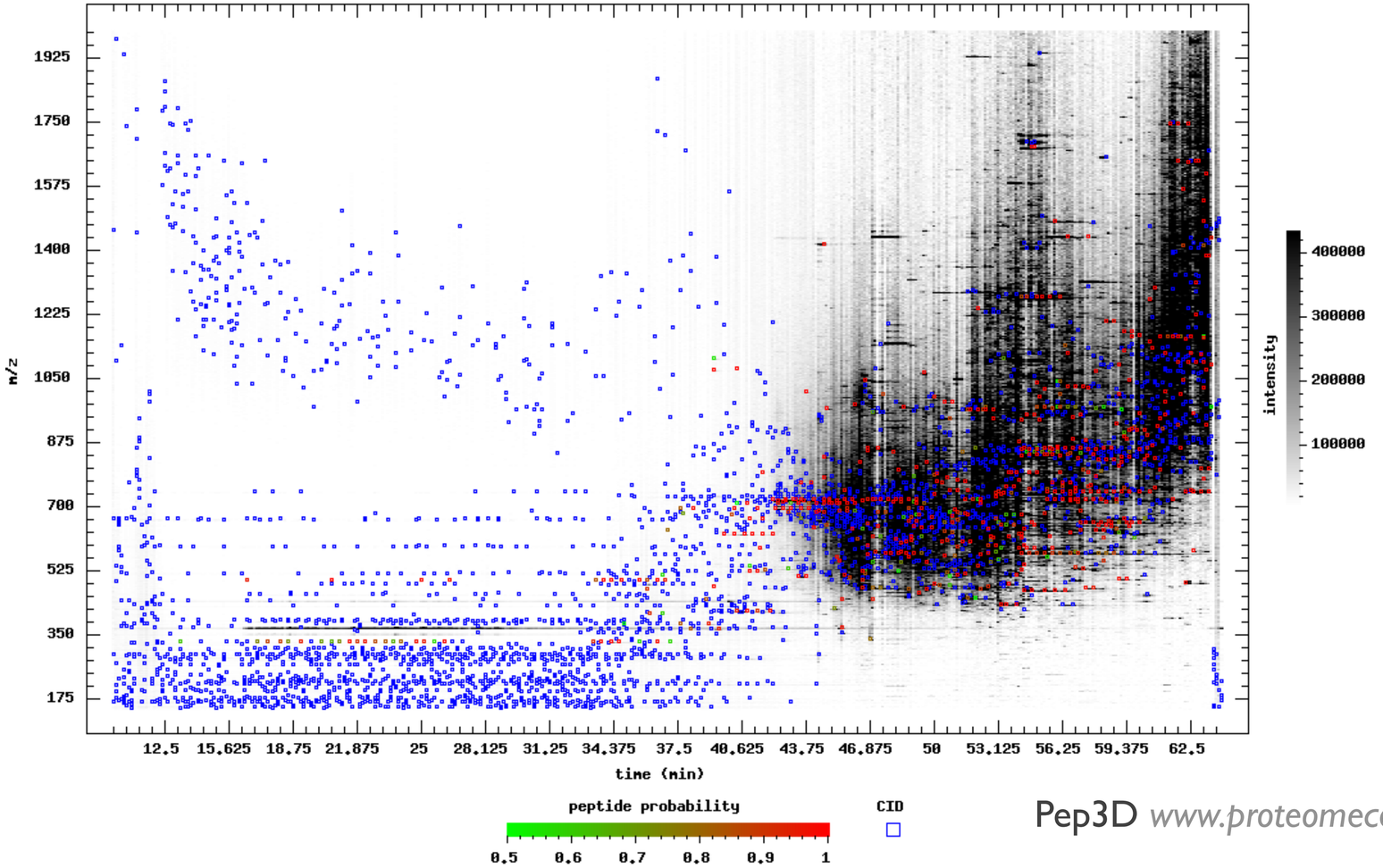
Multivariate Data: Transcriptomics

MA Plot: 1 array

Box Plot: 3 arrays



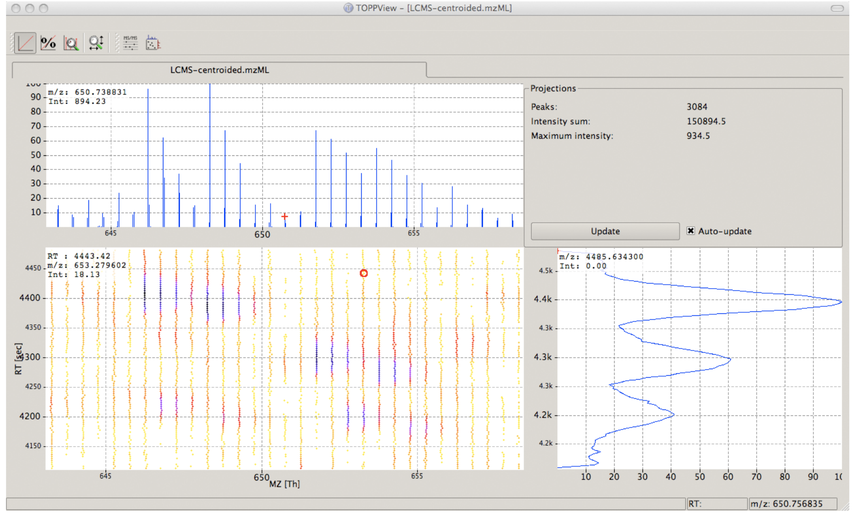
1 = before normalization
2 = after within-array normalization
3 = after between-array normalization



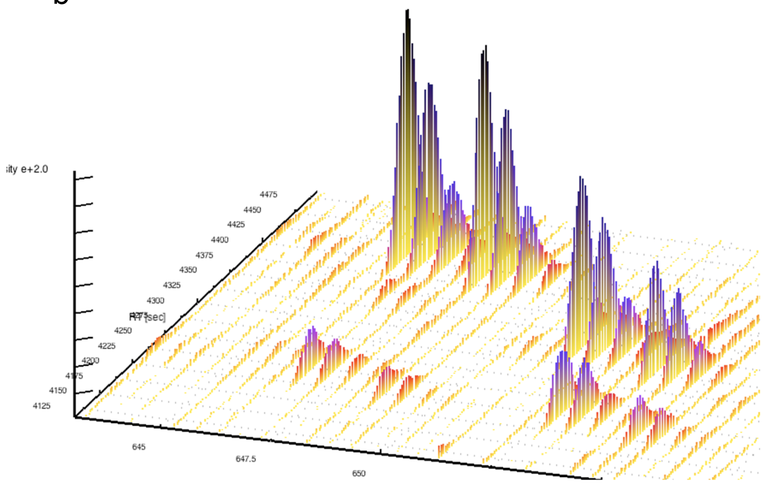
Pep3D *www.proteomecenter.org*

Multivariate Data: Proteomics

a

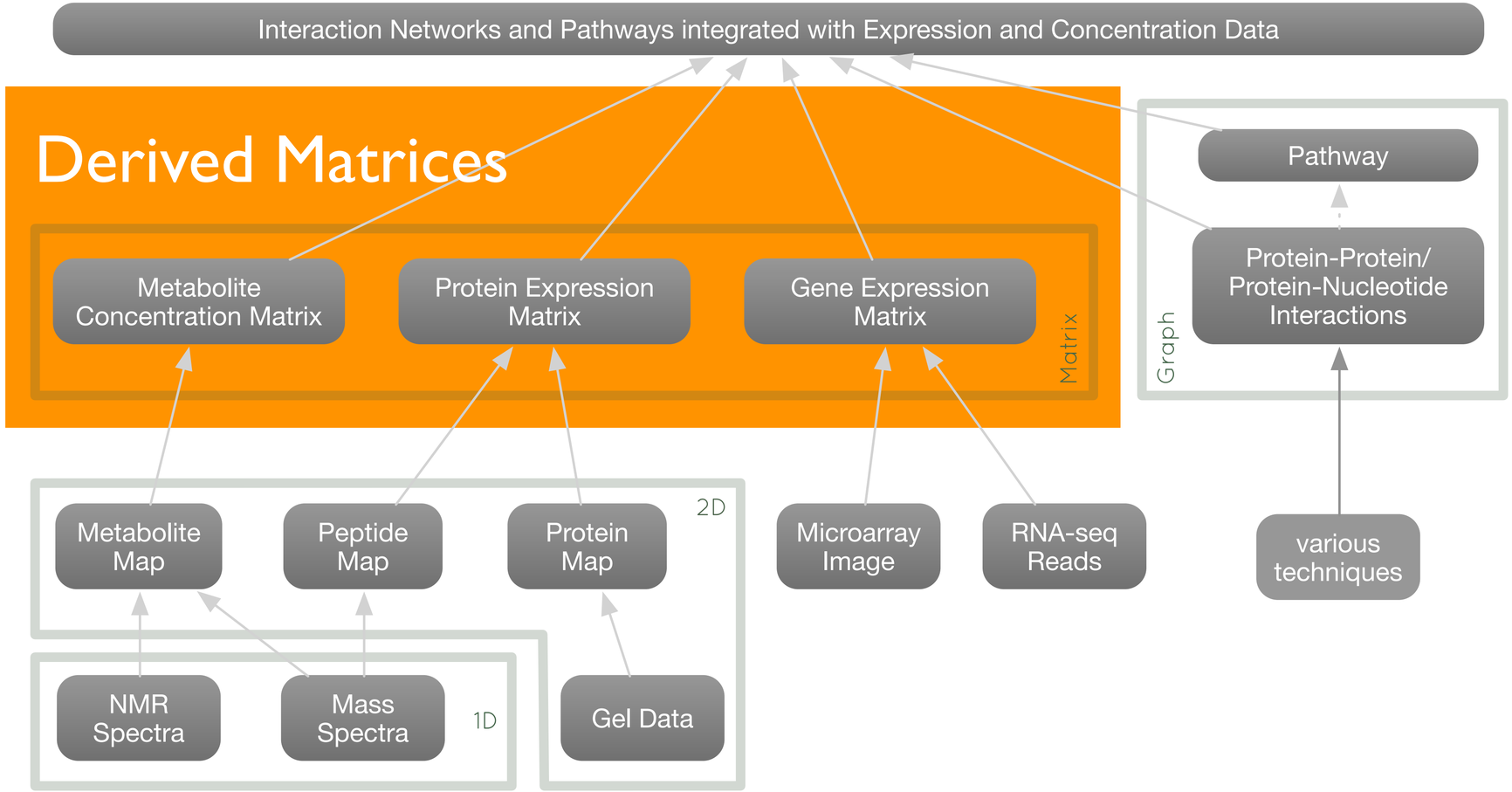


b



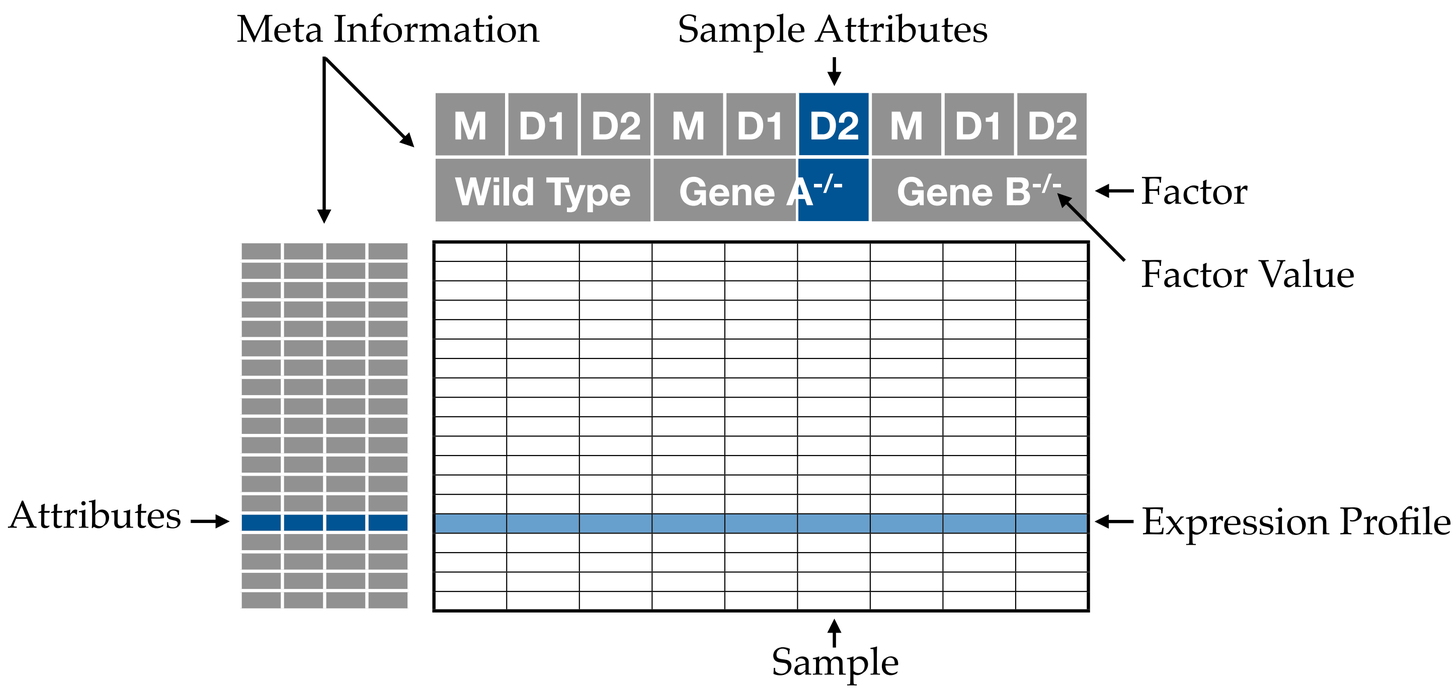
TOPPView *www.open-ms.de*

Multivariate Data: Derived Data



Multivariate Data: Derived Matrices

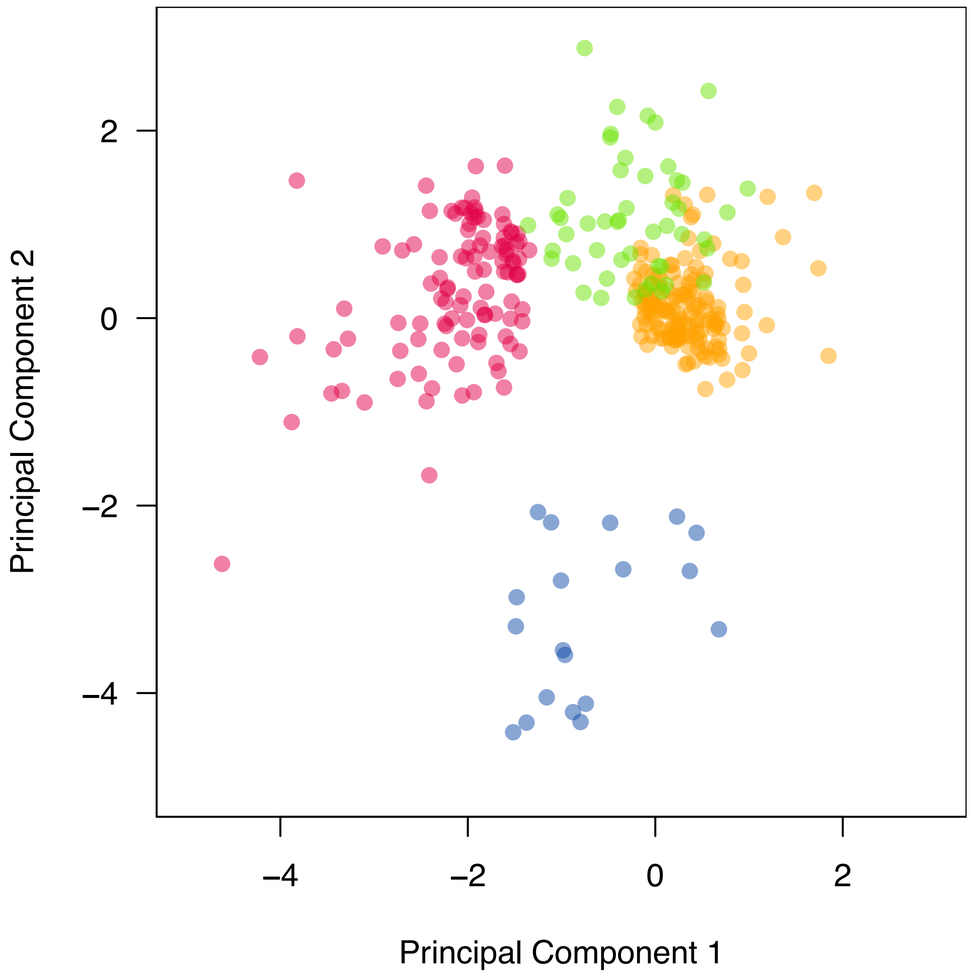
- matrices of multi-dimensional vectors
- usually abundance profiles, e.g. transcript or protein levels, metabolite concentrations



Gehlenborg et al. 2010, Nature Methods

Multivariate Data: Derived Matrices

Scatter Plot



Multivariate Data: Derived Matrices

- **Scatter Plots and Dimensionality Reduction**

  - used to visualize high-dimensional profiles as projections in lower-dimensional spaces (usually 2D, sometimes also 3D ...)

  - there is always a loss of information in the process, goal is to minimize the loss of information

  - many different algorithms: Principal Components Analysis (PCA), Multi-Dimensional Scaling (MDS), Isomap, etc.

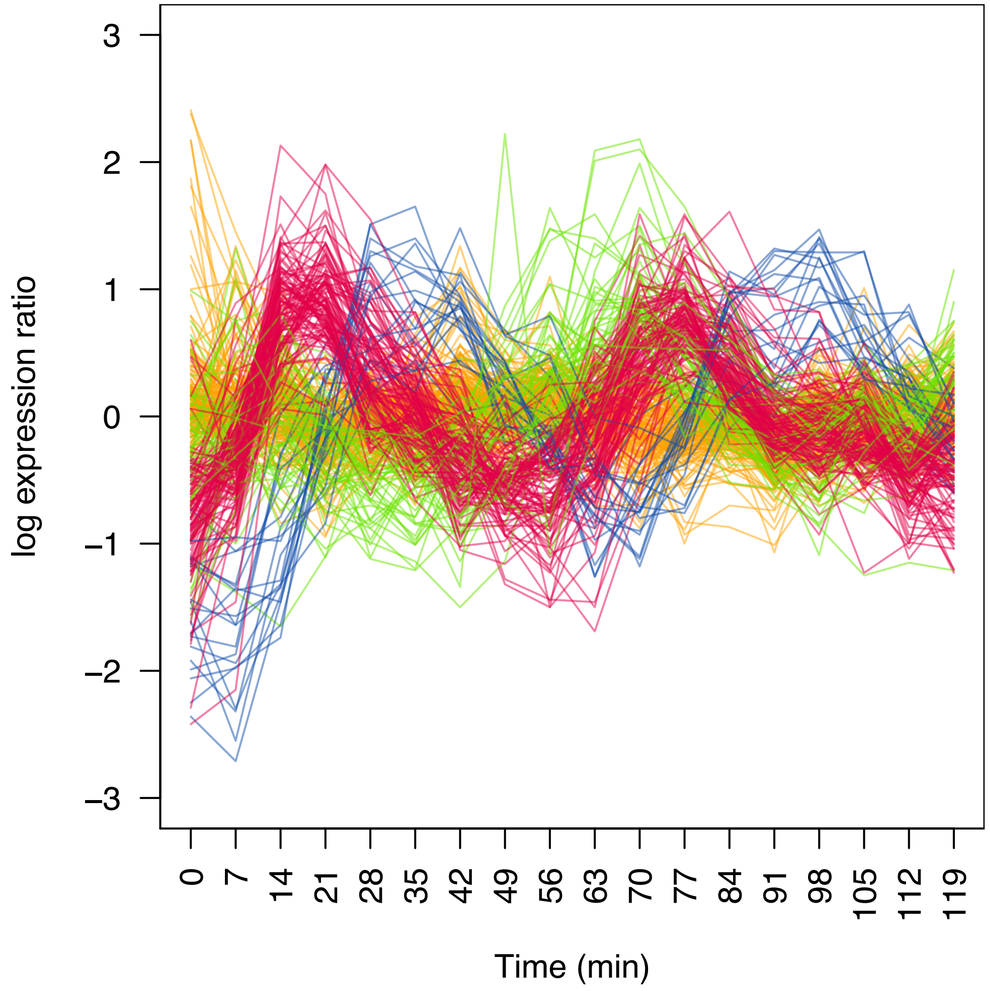
  - **Pros** - good choice to get an idea about the overall structure of the whole data set: clusters, outliers, gaps in the data

  - **Cons** - because of the dimensionality reduction the original profiles are not accessible in the visualization

Gehlenborg et al. 2010, Nature Methods

Multivariate Data: Derived Matrices

Profile Plot *a.k.a.*
Parallel Coordinates



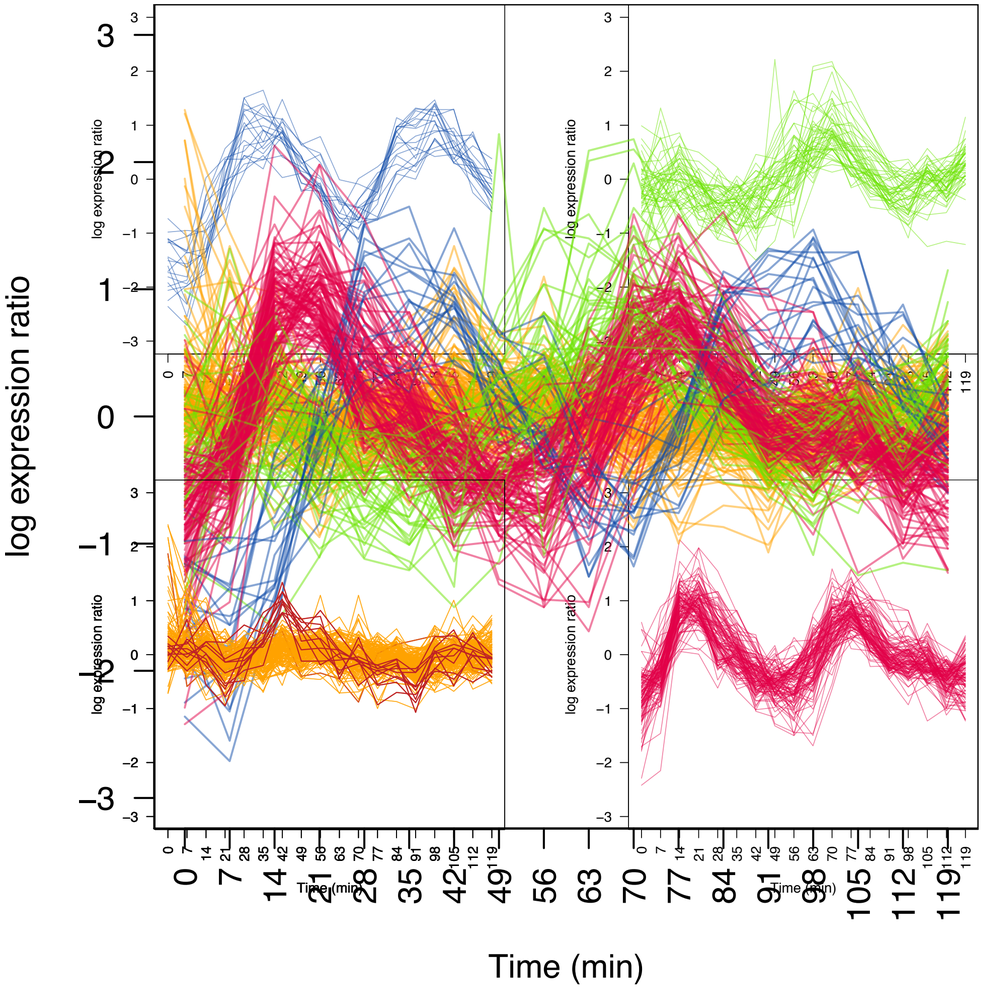
Multivariate Data: Derived Matrices

- **Profile Plot/Parallel Coordinate Plots**
  - **Pros**
    - encoding by position: profiles easy to read
    - color-coding of expression profiles (groups) very efficient
  - **Cons**
    - overplotting
    - grows horizontally with every additional sample

Gehlenborg et al. 2010, Nature Methods

Multivariate Data: Derived Matrices

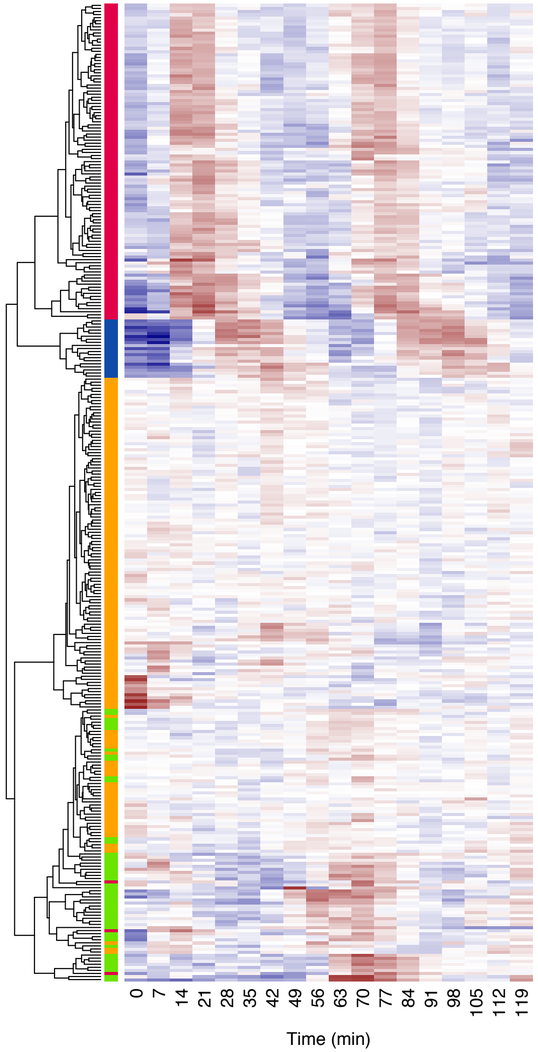
Profile Plot *a.k.a.*
Parallel Coordinates



Gehlenborg et al. 2010, Nature Methods

Multivariate Data: Derived Matrices

Heat Map with Dendrogram



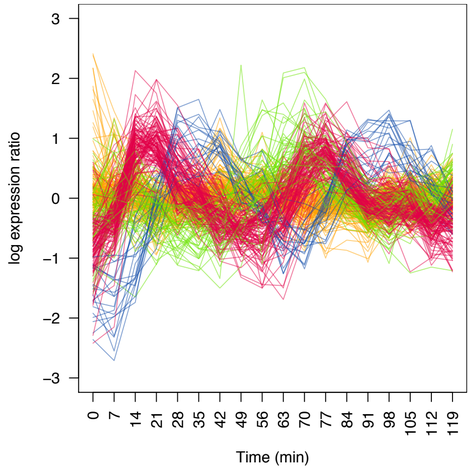
Multivariate Data: Derived Matrices

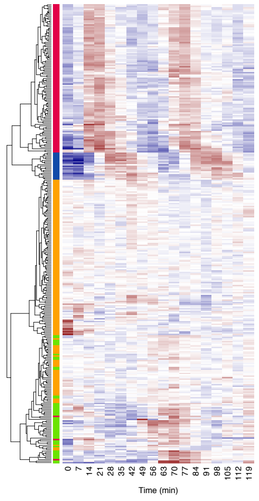
- **Heatmap**
  - **Pros**
    - no overplotting, yet a very dense information display
    - can be combined with dendrogram and additional information can be encoded in further columns or in the height of rows
  - **Cons**
    - only qualitative interpretation possible due to color coding
    - grows horizontally with every additional sample and grows vertically with every additional profile

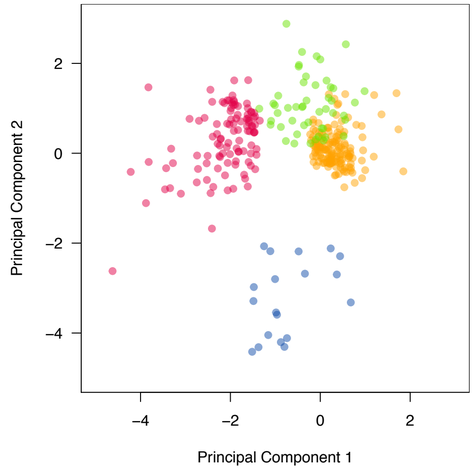
Multivariate Data: Summary

few, high-res

many, low-res

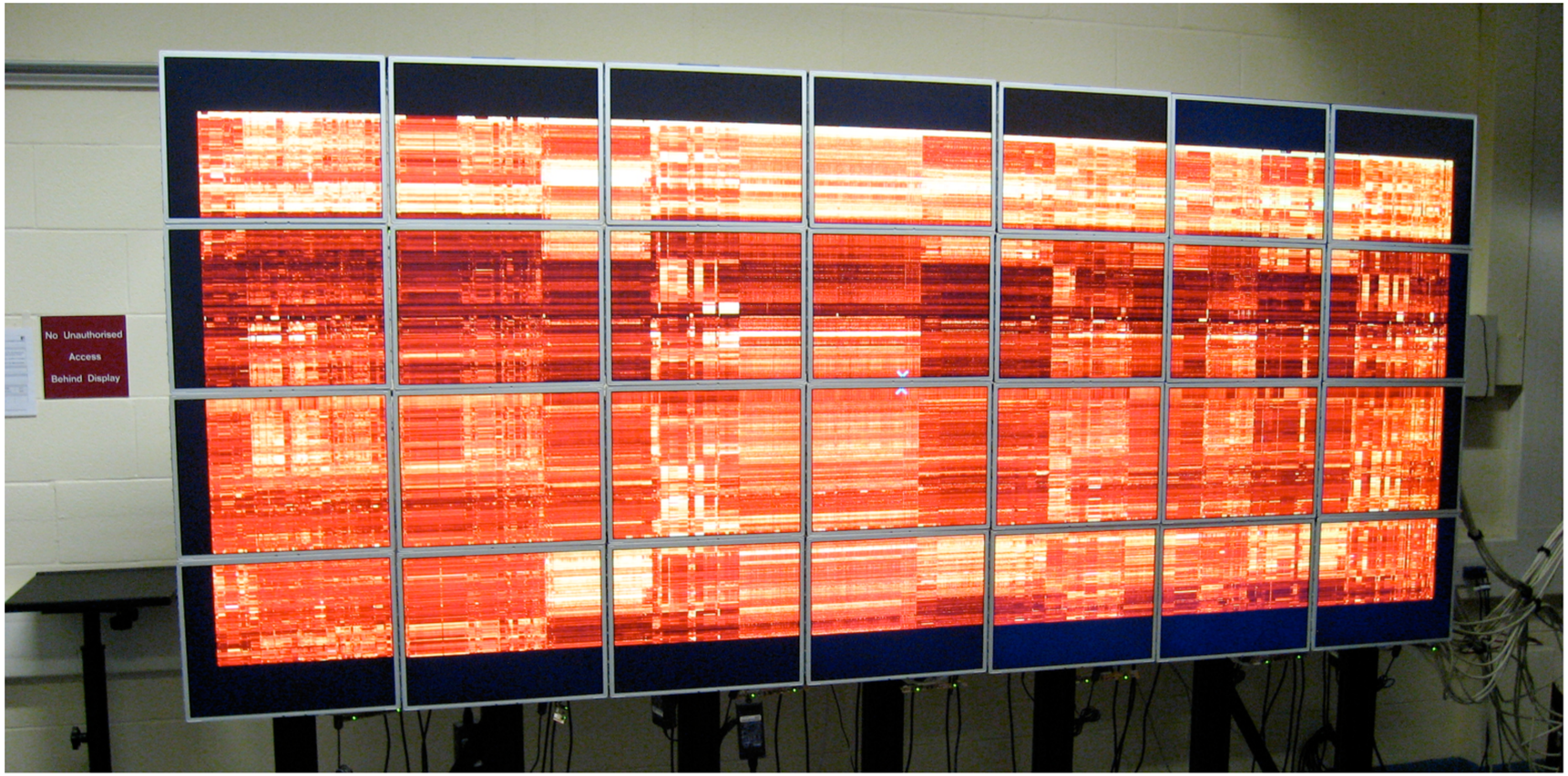






data: Lukk et al., *2010, Nature Biotechnology*

Problem: Very Large Expression Matrices

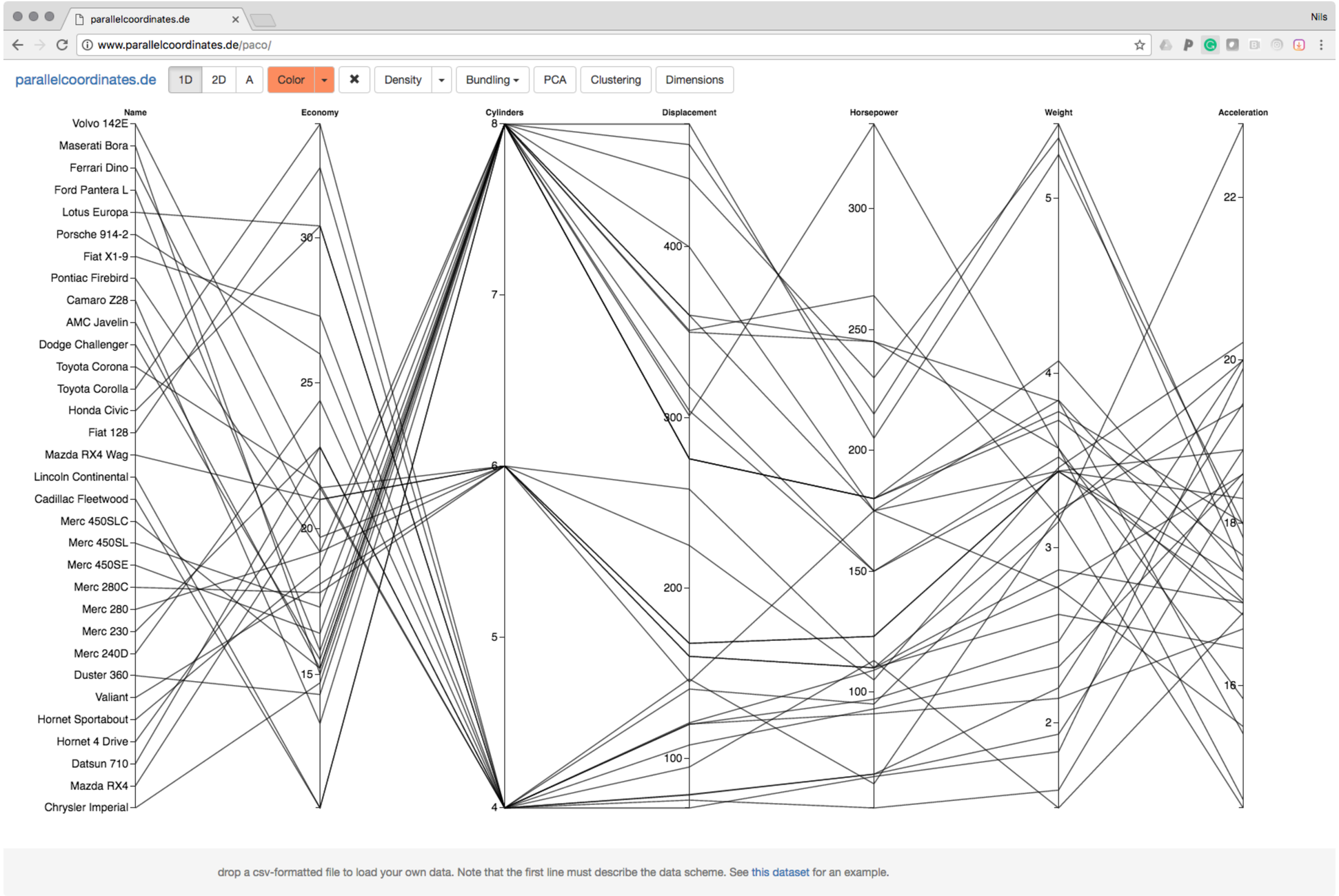


Power Wall (7x4 screens = 11,200x4,800), University of Leeds

1000 transcripts, 5372 samples

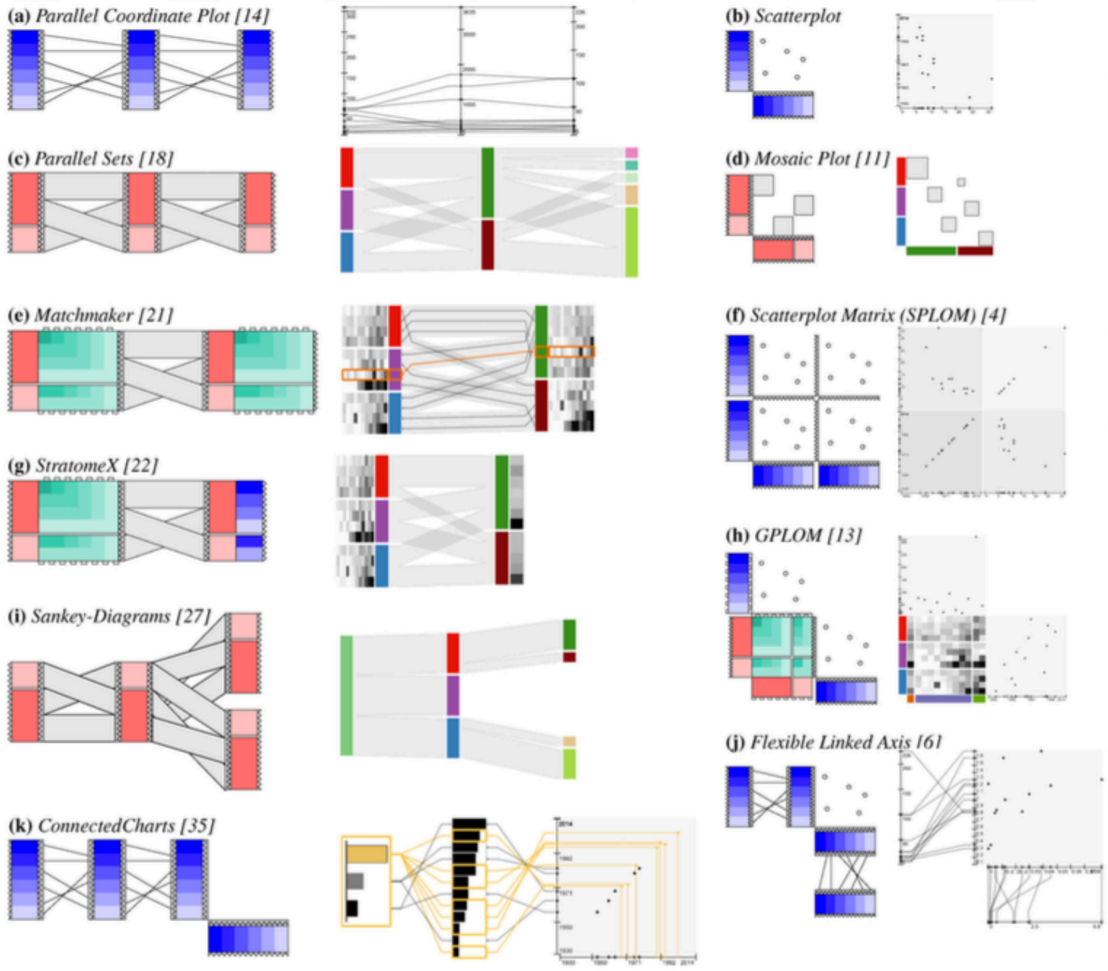
Multivariate Data

Heterogeneous Tables



http://www.parallelcoordinates.de/paco/

Domino


(a) Parallel Coordinate Plot [14]
(b) Scatterplot
(c) Parallel Sets [18]
(d) Mosaic Plot [11]
(e) Matchmaker [21]
(f) Scatterplot Matrix (SPLOM) [4]
(g) StratomeX [22]
(h) GPLOM [13]
(i) Sankey-Diagrams [27]
(j) Flexible Linked Axis [6]
(k) ConnectedCharts [35]

https://sgratzl.github.io/domino_vis2014/#/title

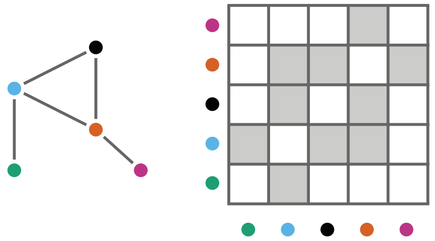
Visualization of Trees and Networks

Refresher

- What is a graph? **This is a data structure not a visual representation!**
  - Set **V** of vertices (= nodes) and set **E** of edges
  - Directed graph vs undirected graph
  - Directed acyclic graph (DAG)
- What is a tree?
  - Binary tree: two children per vertex
  - General tree: any number of children

Storing Graphs and Trees

- Use a matrix for adjacency (edges explicit)
  - edges are matrix cells



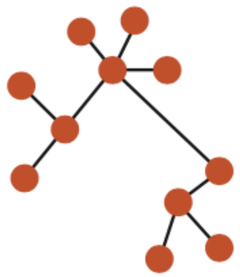
- Use a list with two fields per vertex (edges implicit)
  - incoming edges
  - outgoing edges

Arrange Networks and Trees

Node–Link Diagrams

Connection Marks

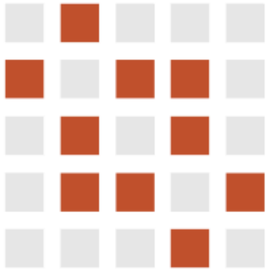
✔ NETWORKS ✔ TREES



Adjacency Matrix

Derived Table

✔ NETWORKS ✔ TREES



Enclosure

Containment Marks

✘ NETWORKS ✔ TREES

Munzer, Visualization Analysis and Design, 2014

Node-Link Diagram Approaches

Trees

(a) (b)

Figure 9.2. Node–link layouts of small trees. (a) Triangular vertical for tiny tree. From [Buchheim et al. 02, Figure 2d]. (b) Spline radial layout for small tree. From http://mbostock.github.com/d3/ex/tree.html.

Munzer, Visualization Analysis and Design, 2014

Trees

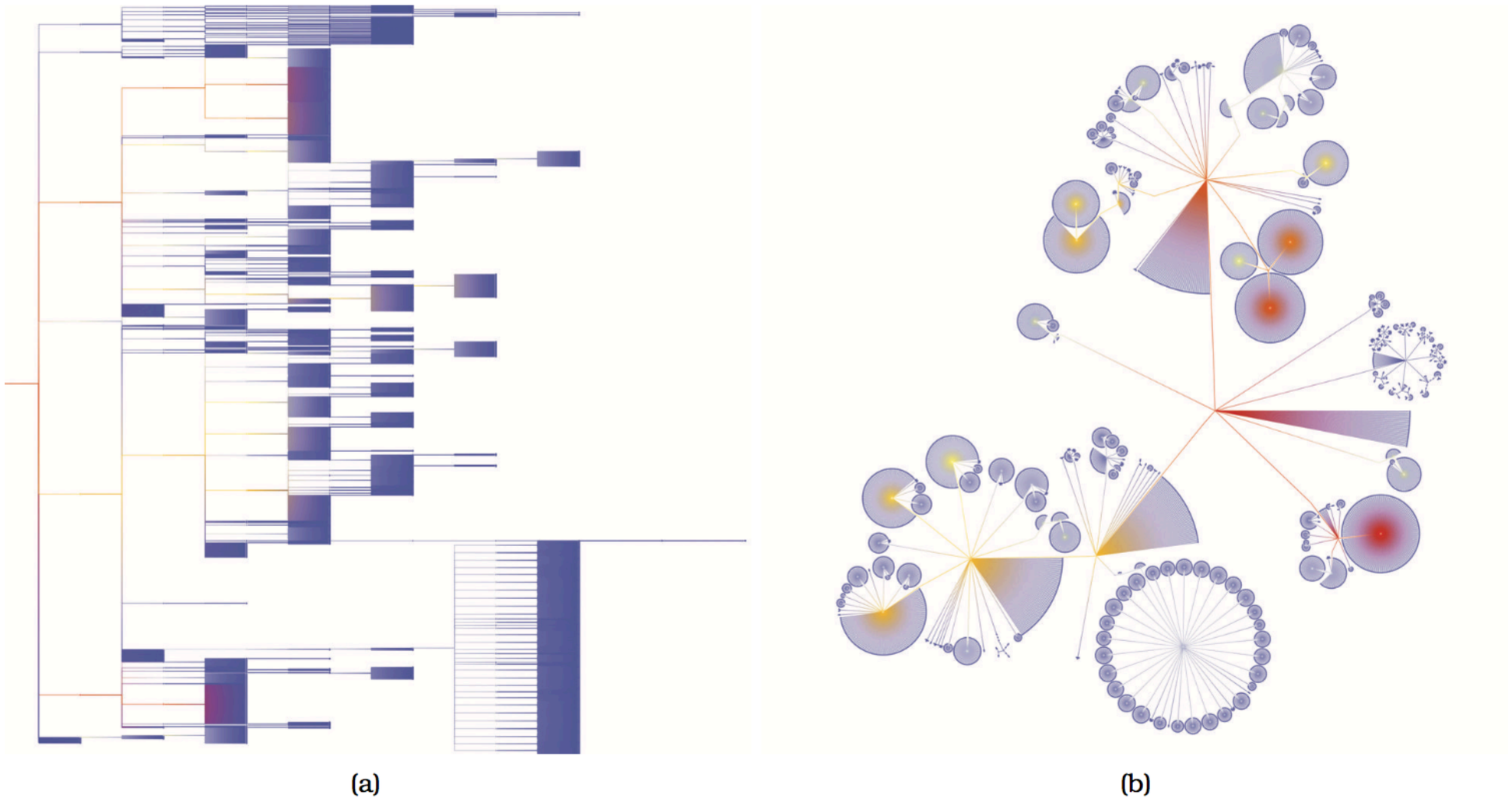


Figure 9.3. Two layouts of a 5161-node tree. (a) Rectangular horizontal node–link layout. (b) BubbleTree node–link layout.

Munzer, Visualization Analysis and Design, 2014

Graphs

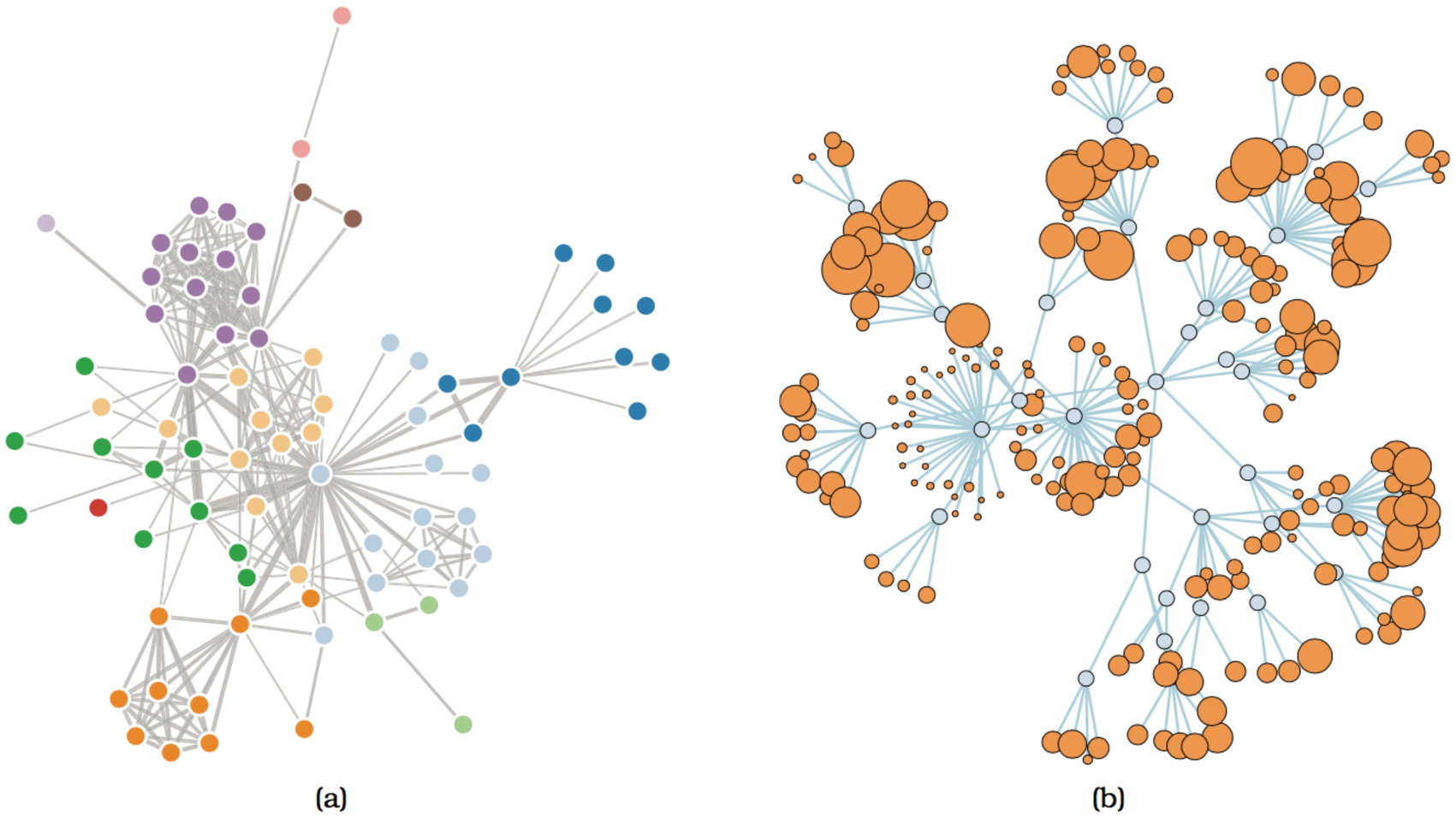
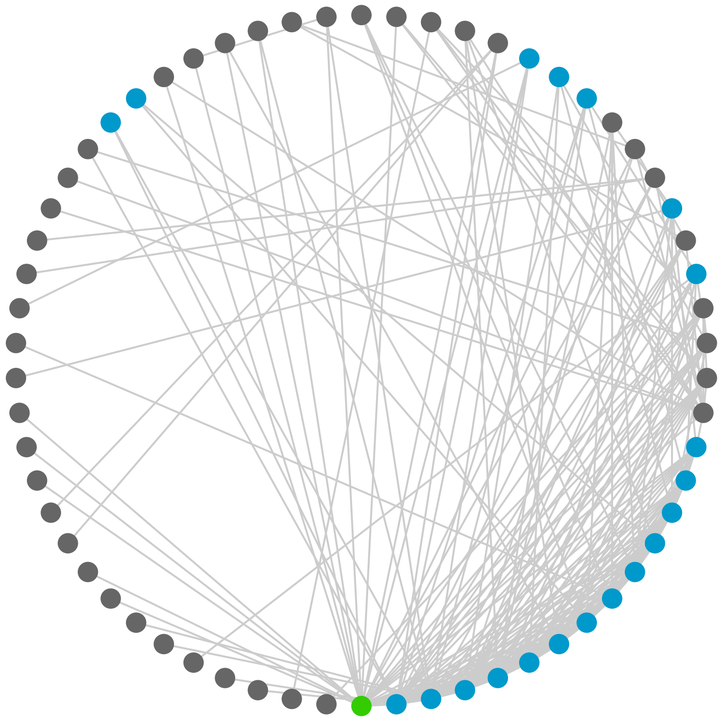


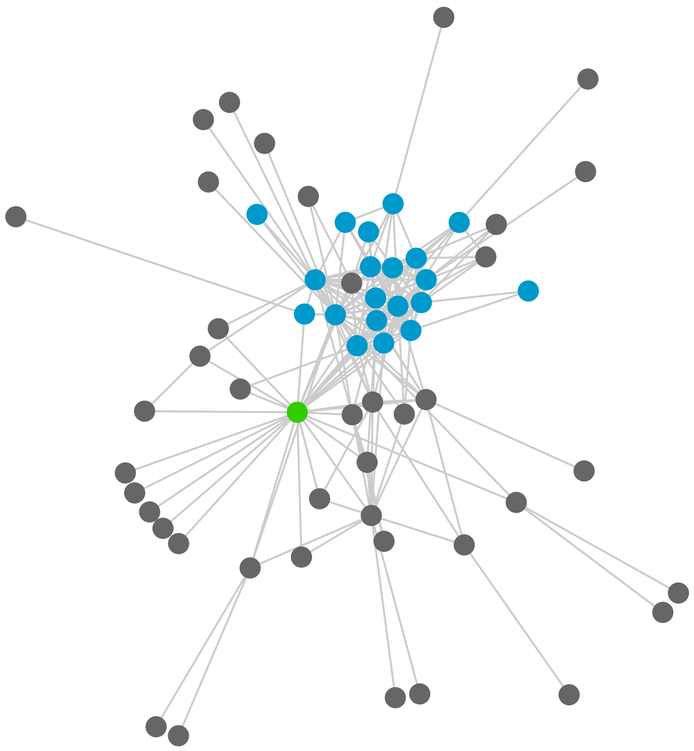
Figure 9.4. Node–link layouts of small networks. (a) Force-directed placement of small network of 75 nodes, with size coding for link attributes. (b) Larger network, with size coding for node attributes. From http://bl.ocks.org/mbostock/4062045 and http://bl.ocks.org/1062288.

Munzer, Visualization Analysis and Design, 2014

Graph Layouts

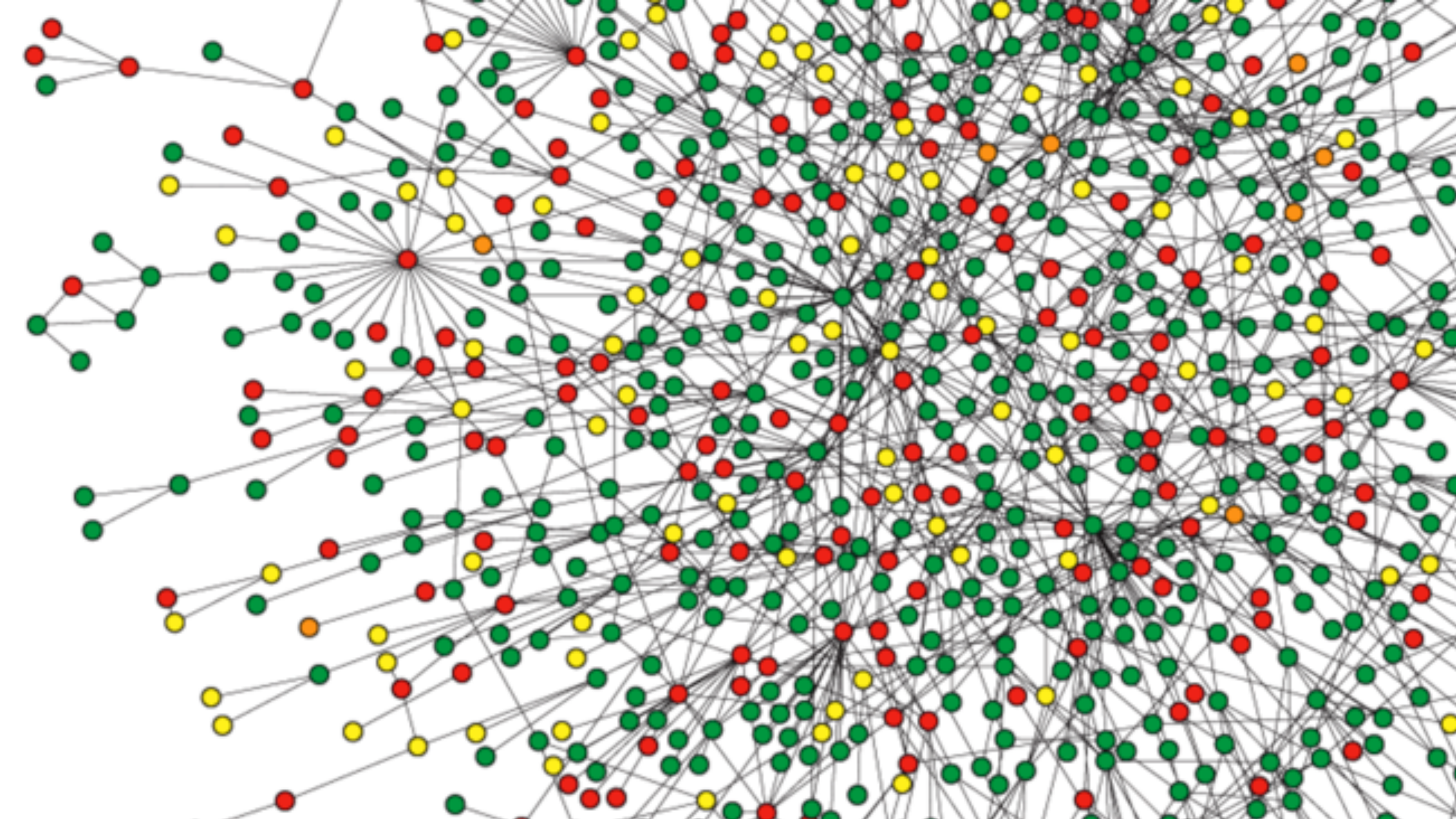


Circular Layout

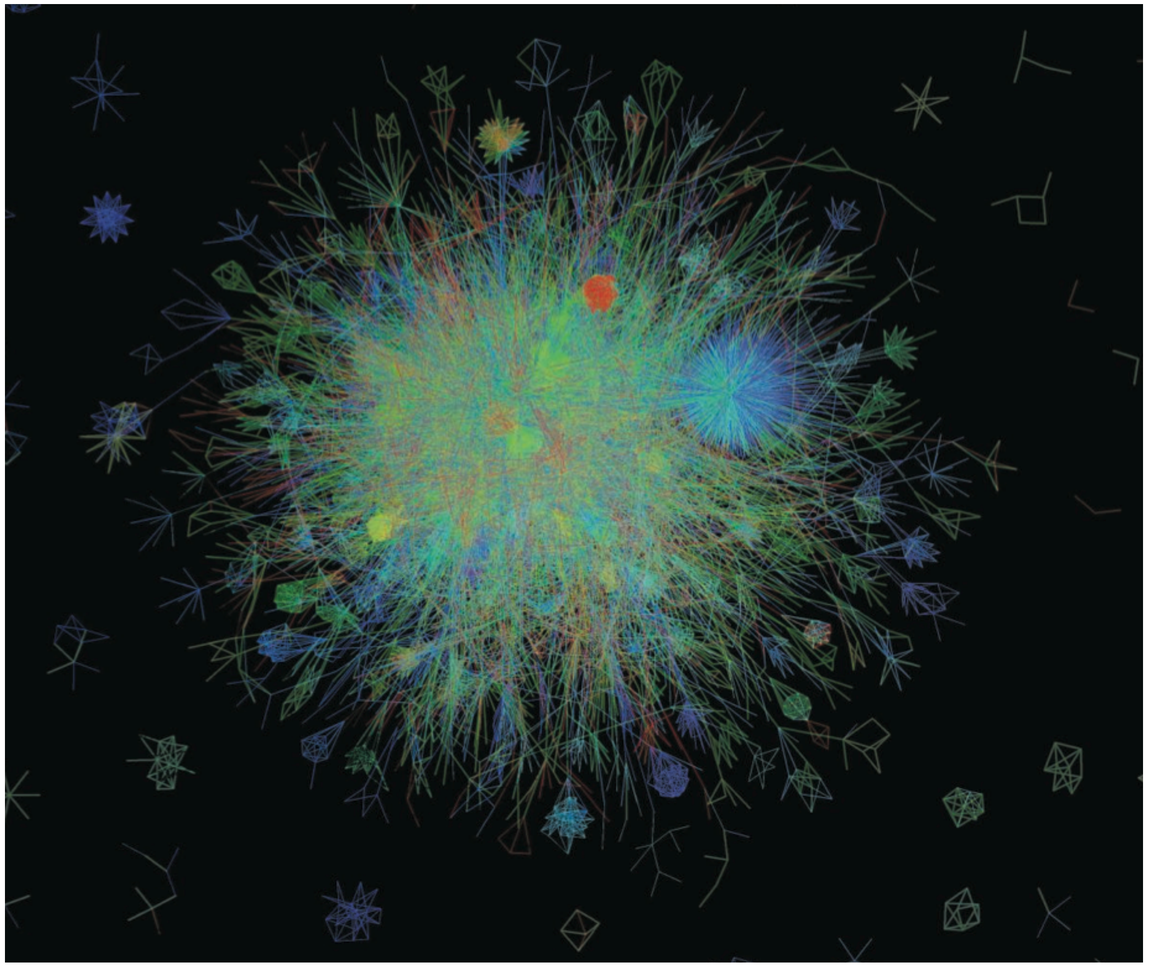


Force-directed Layout

Gehlenborg and Wong, Nature Methods, 2012



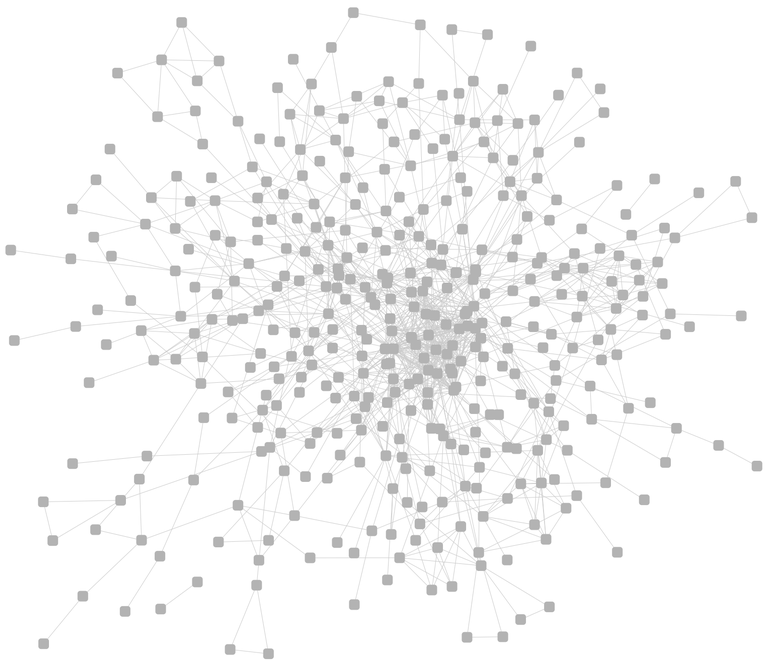
Interactions: Hairballs, Ridiculograms et al.



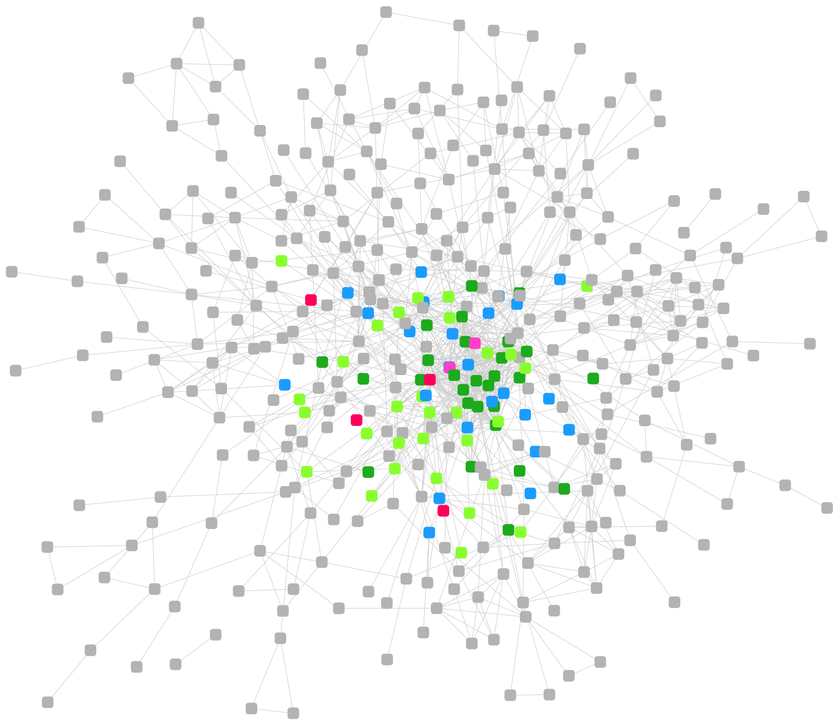
NSF Visualization Challenge 2011 Honorable Mention: AraNet

Interactions: Manual Layout

Step 1



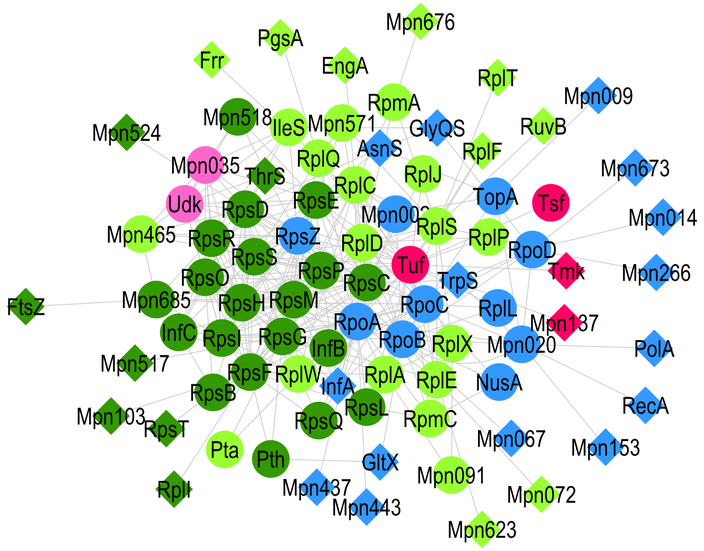
Step 2



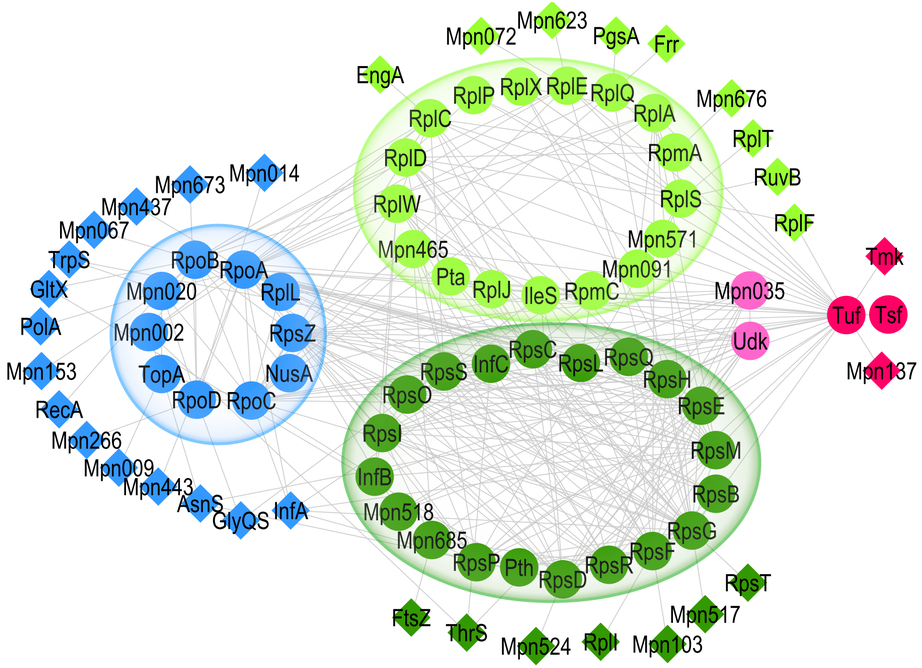
Gehlenborg et al. 2010, Nature Methods

Interactions: Manual Layout

Step 3



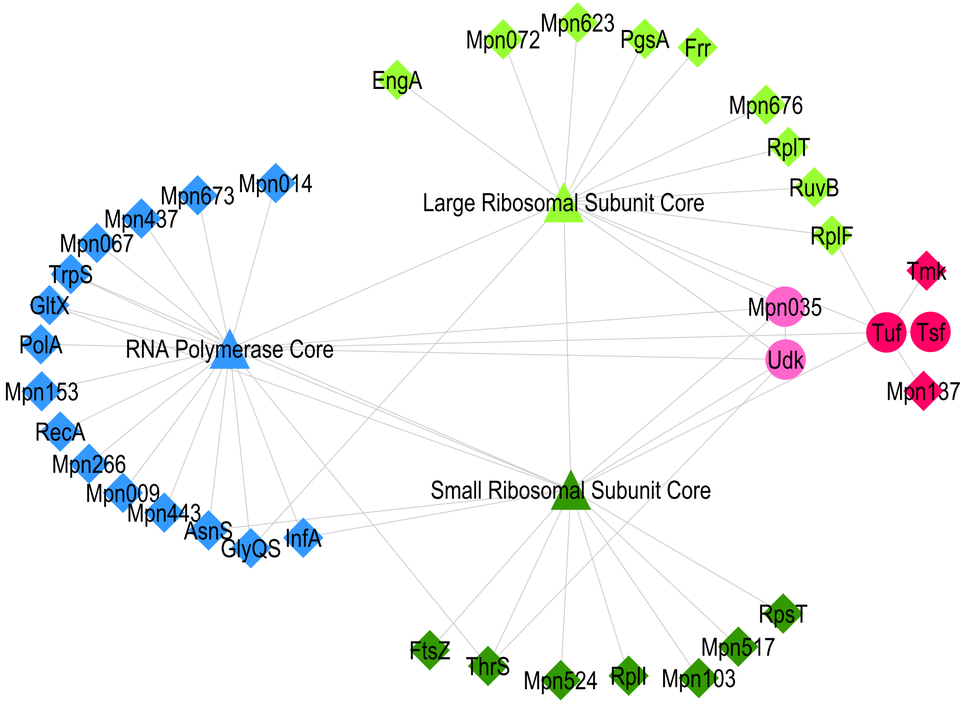
Step 4



Gehlenborg et al. 2010, Nature Methods

Interactions: Manual Layout

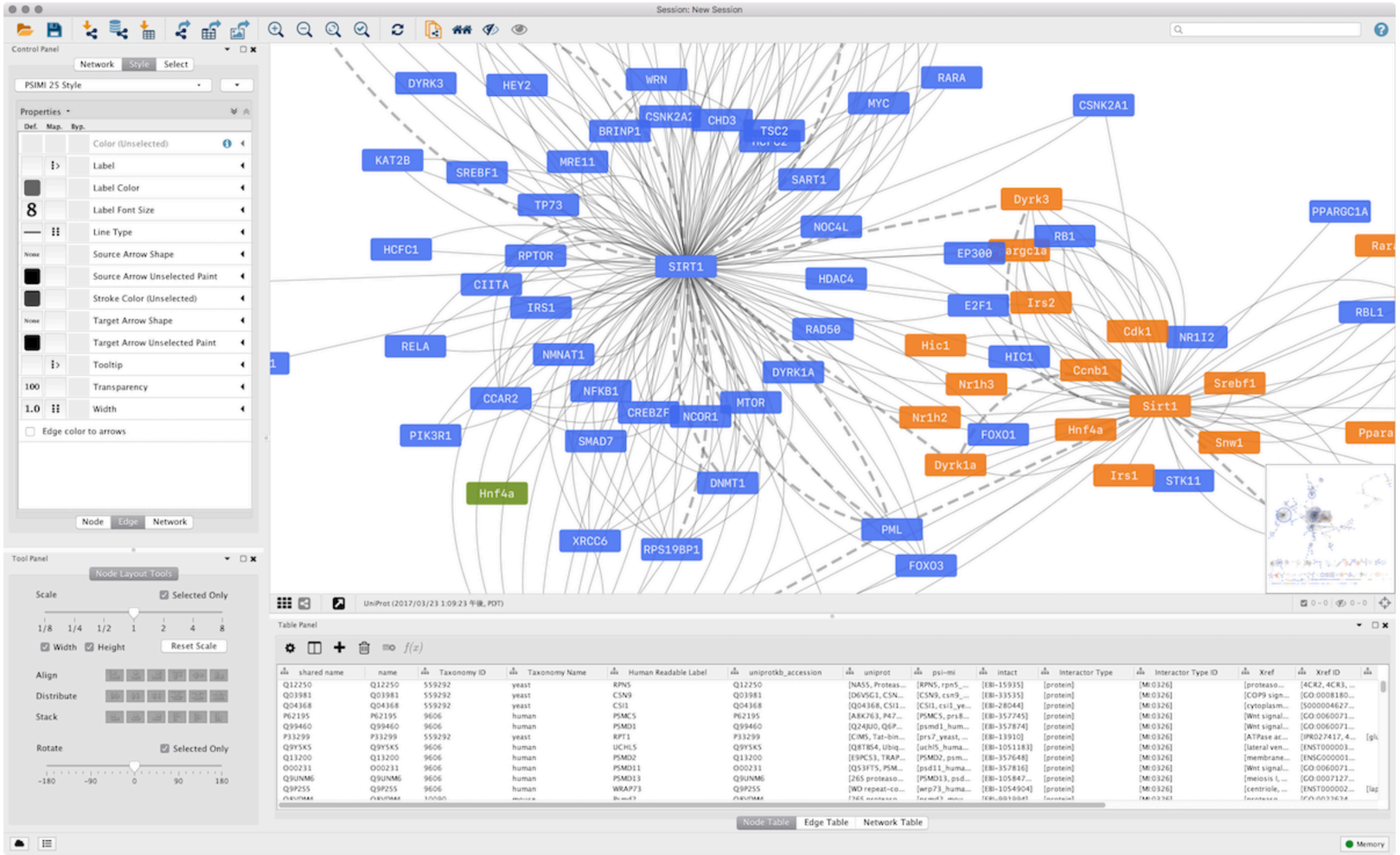
Step 5



Gehlenborg et al. 2010, Nature Methods

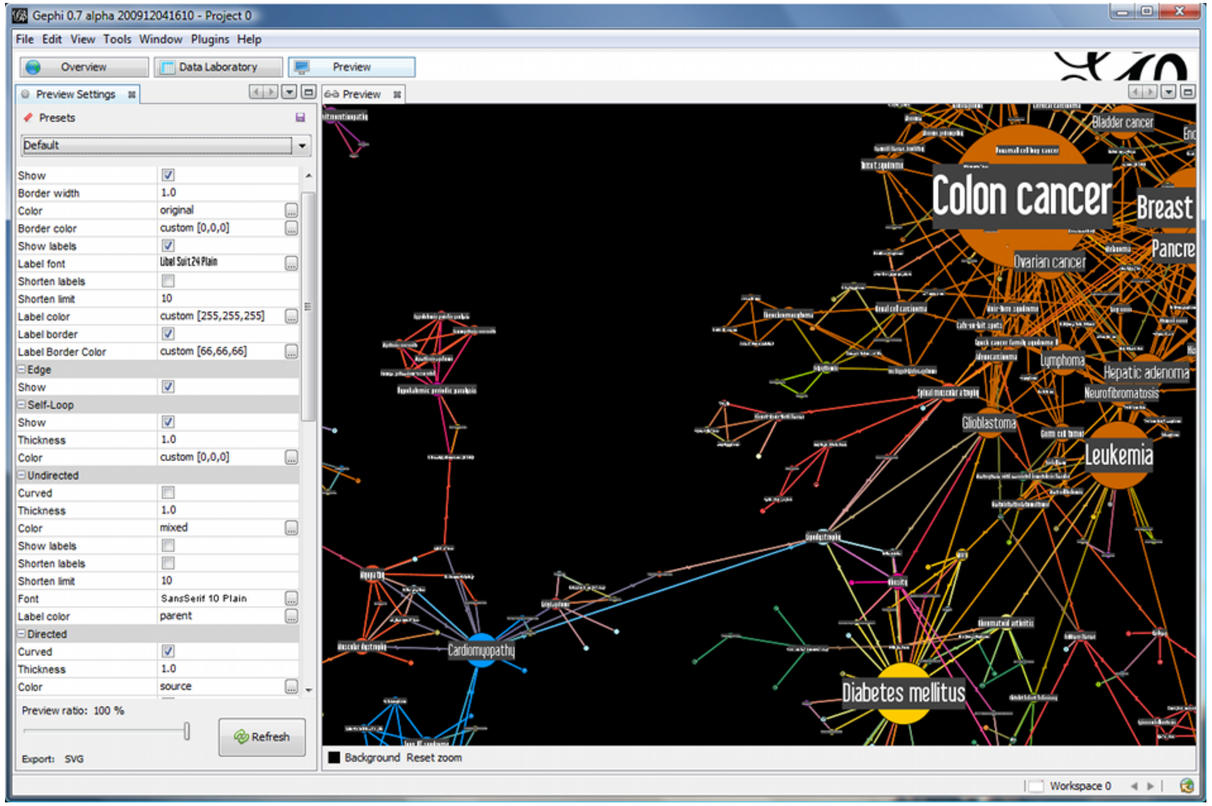
Working with Graphs

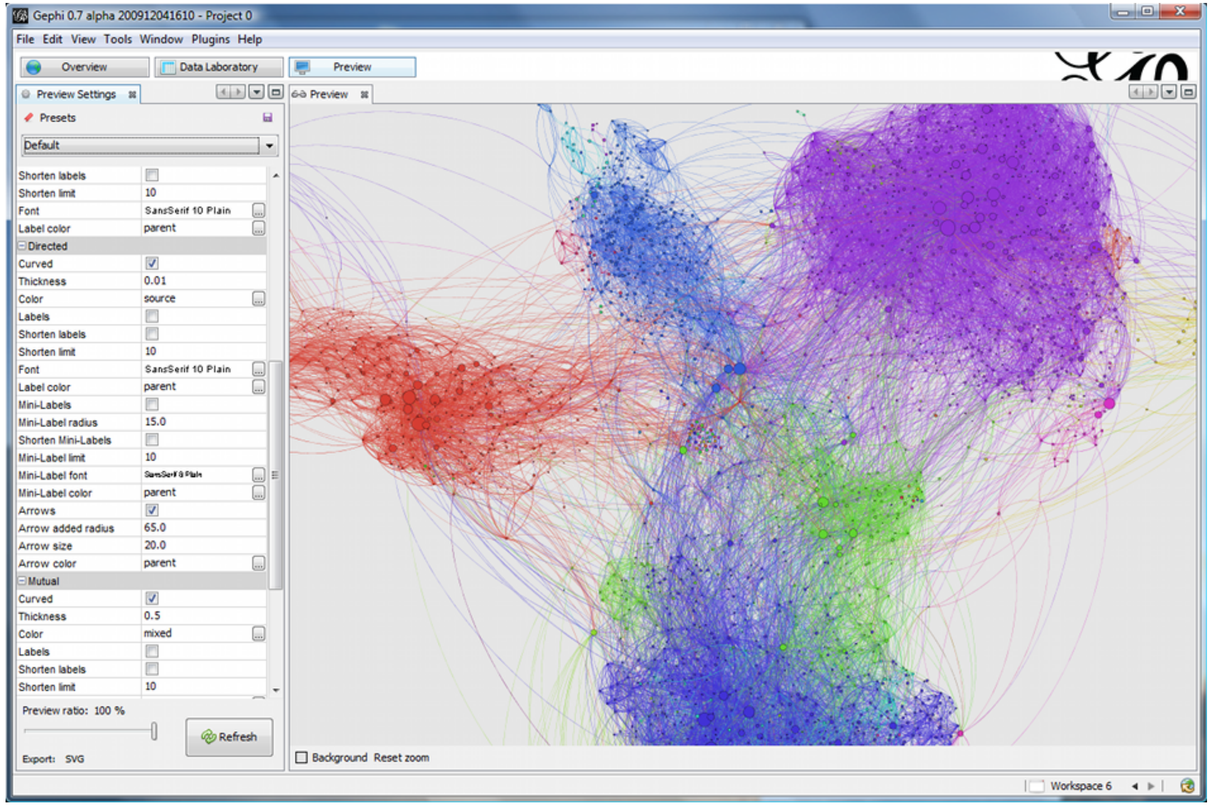
Cytoscape



http://www.cytoscape.org

Gephi

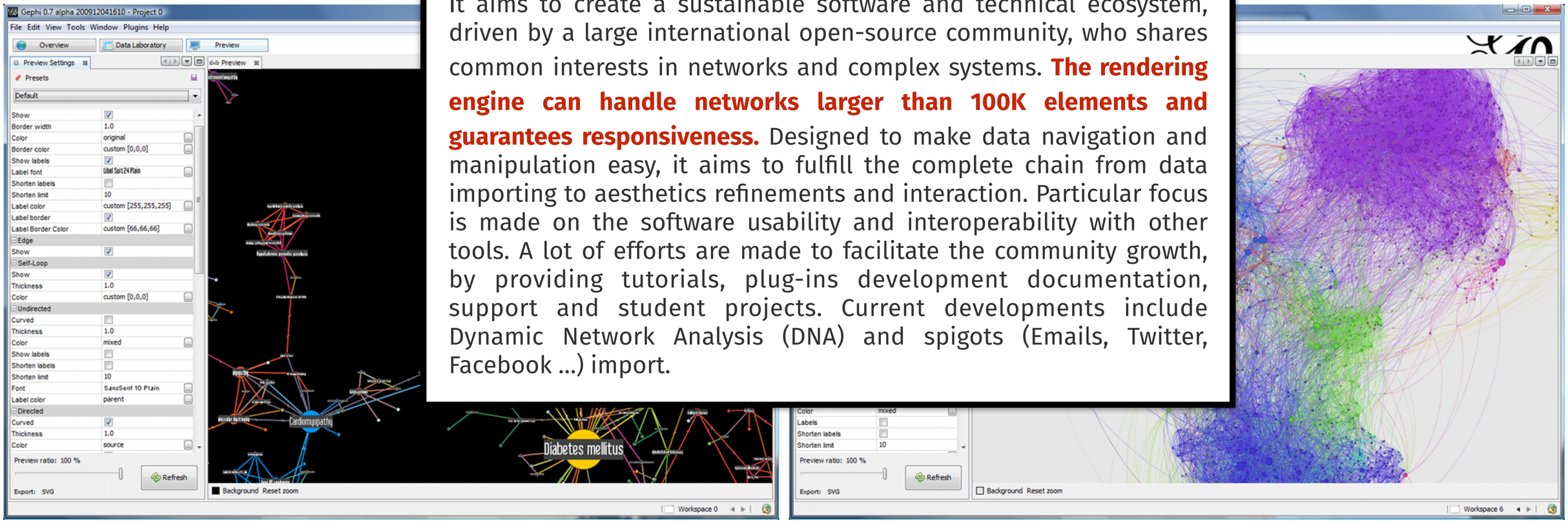




http://www.gephi.org

Gephi

Abstract: Gephi is a new open-source network visualization platform. It aims to create a sustainable software and technical ecosystem, driven by a large international open-source community, who shares common interests in networks and complex systems. **The rendering engine can handle networks larger than 100K elements and guarantees responsiveness.** Designed to make data navigation and manipulation easy, it aims to fulfill the complete chain from data importing to aesthetics refinements and interaction. Particular focus is made on the software usability and interoperability with other tools. A lot of efforts are made to facilitate the community growth, by providing tutorials, plug-ins development documentation, support and student projects. Current developments include Dynamic Network Analysis (DNA) and spigots (Emails, Twitter, Facebook ...) import.

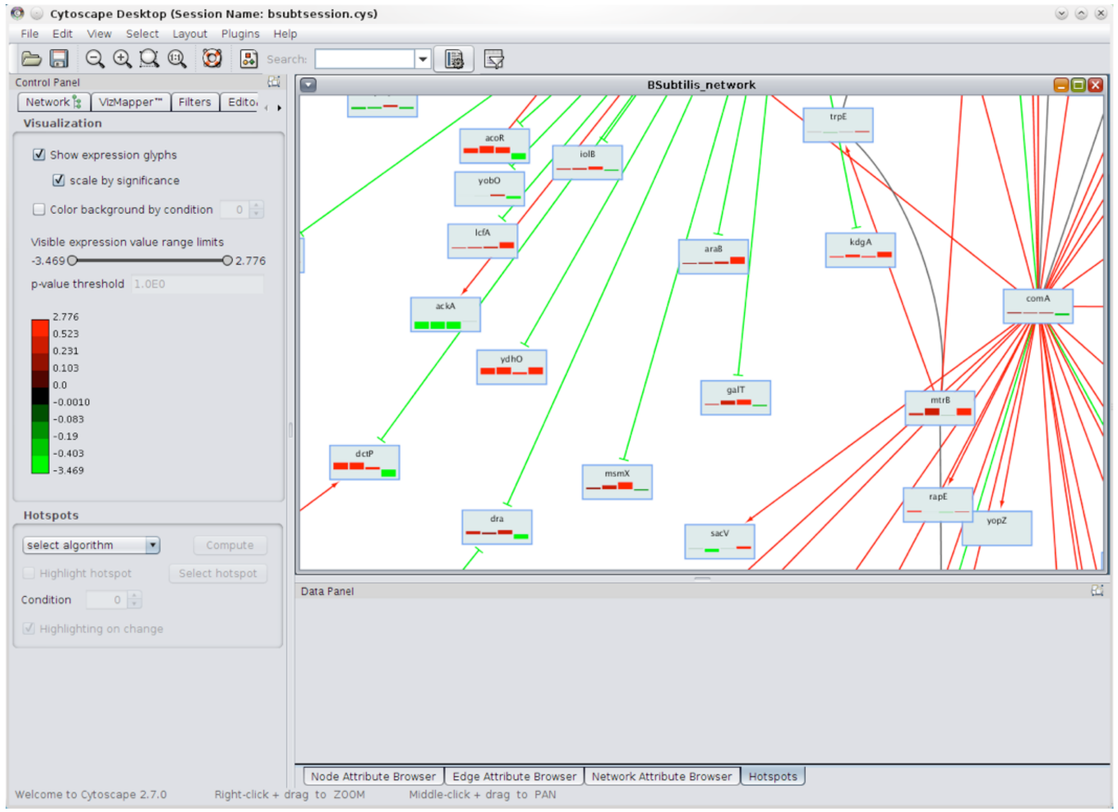


http://www.gephi.org

Interactions: Gene Regulatory Networks

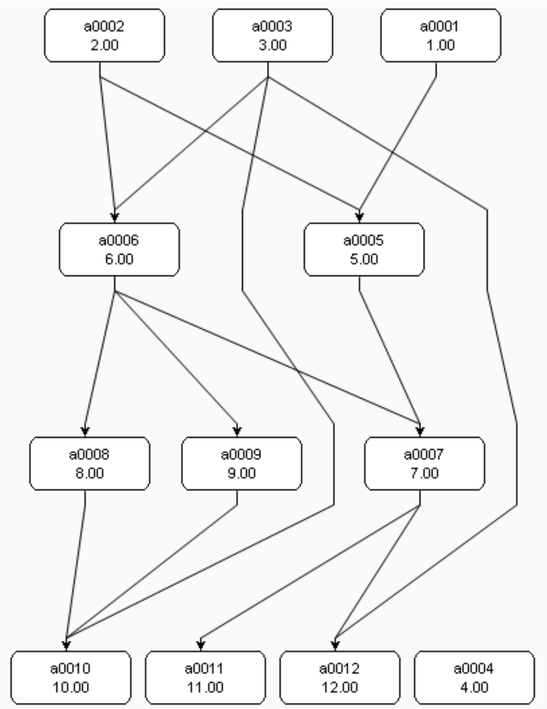
SpotXplore

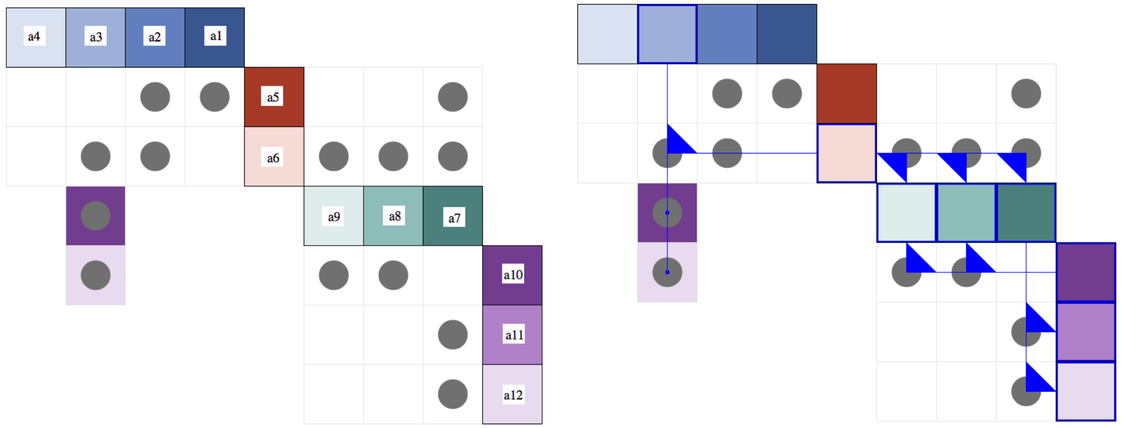
- maps expression profiles onto regulatory network
- statistics can be visualized
- interaction
- highlight subnetworks
- Cytoscape plugin



Westenberg et al. 2008, Comp Graph Forum & Westenberg et al. 2010, Bioinformatics

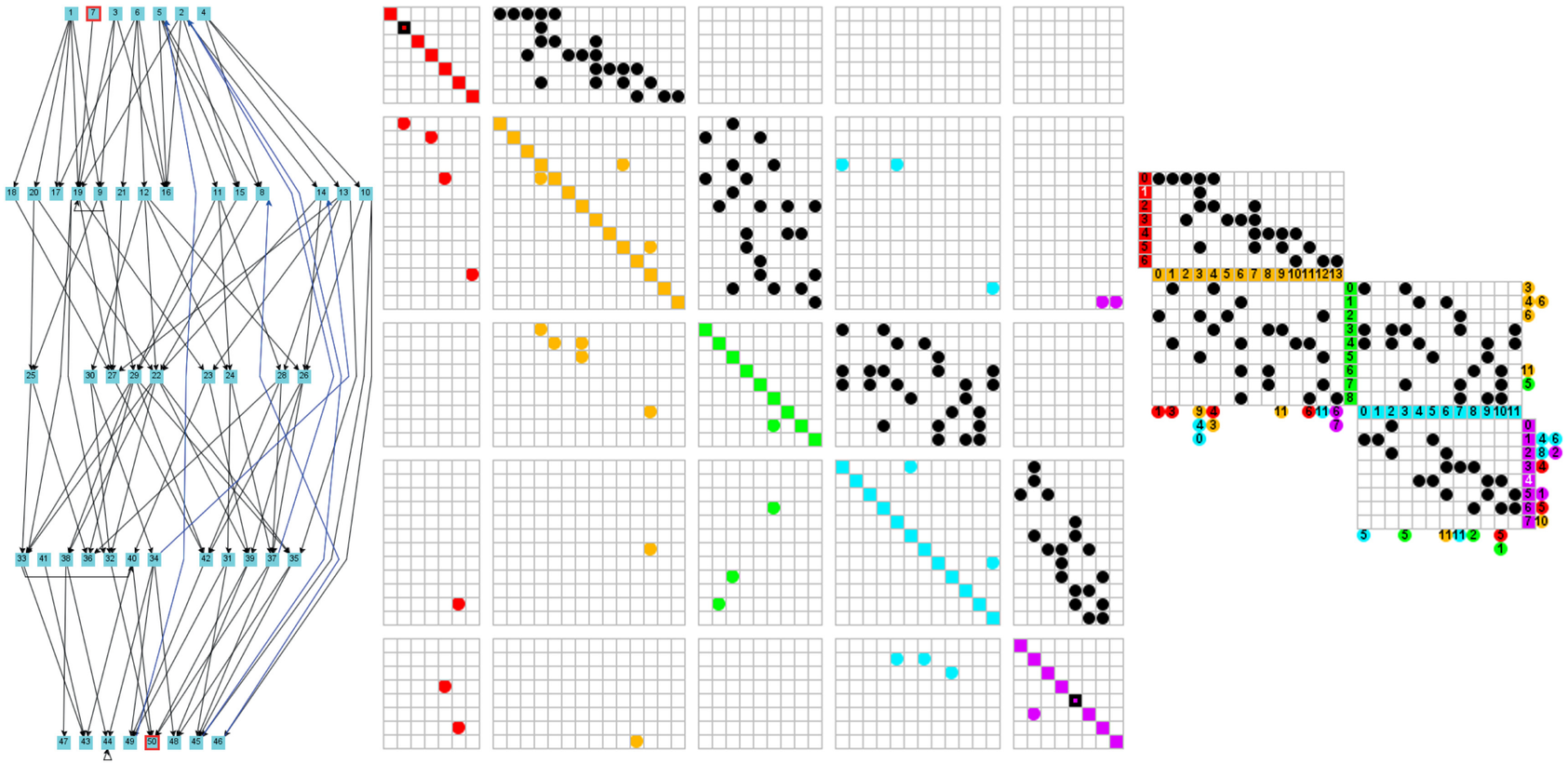
Interactions: Quilts





Watson et al., Proceedings of InfoVis 2007 (Poster)

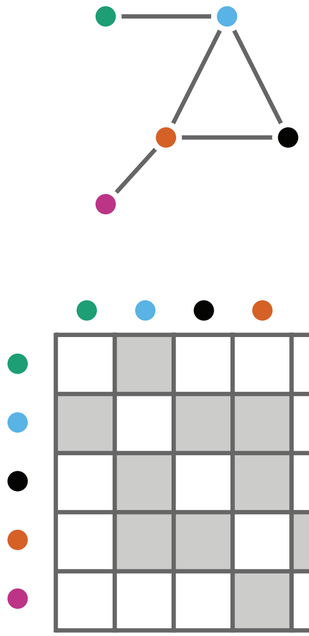
Interactions: Quilts

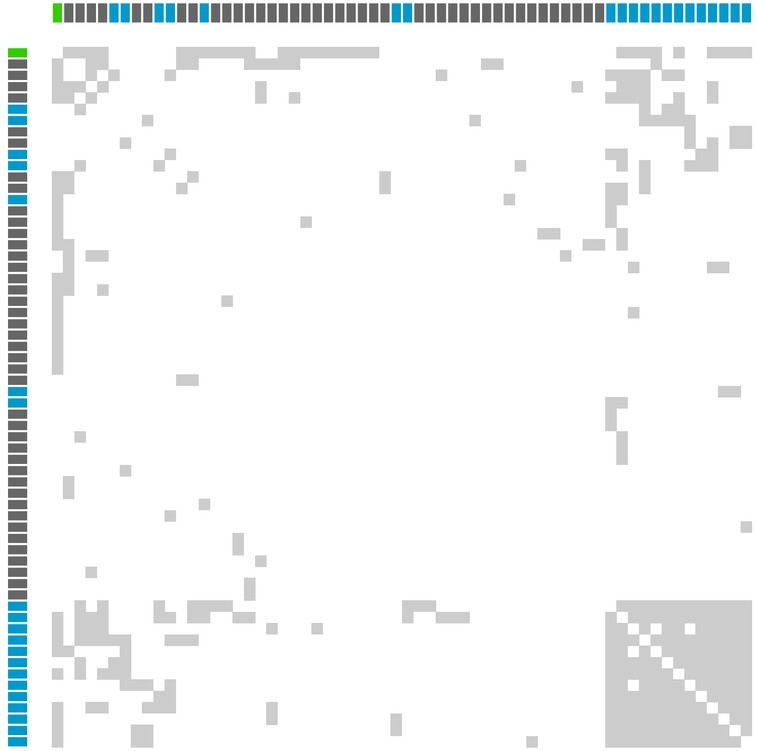


Bae and Watson, Proceedings of InfoVis 2011

Adjacency Matrix Approaches

Adjacency Matrix





Gehlenborg and Wong, Nature Methods, 2012

Adjacency Matrix

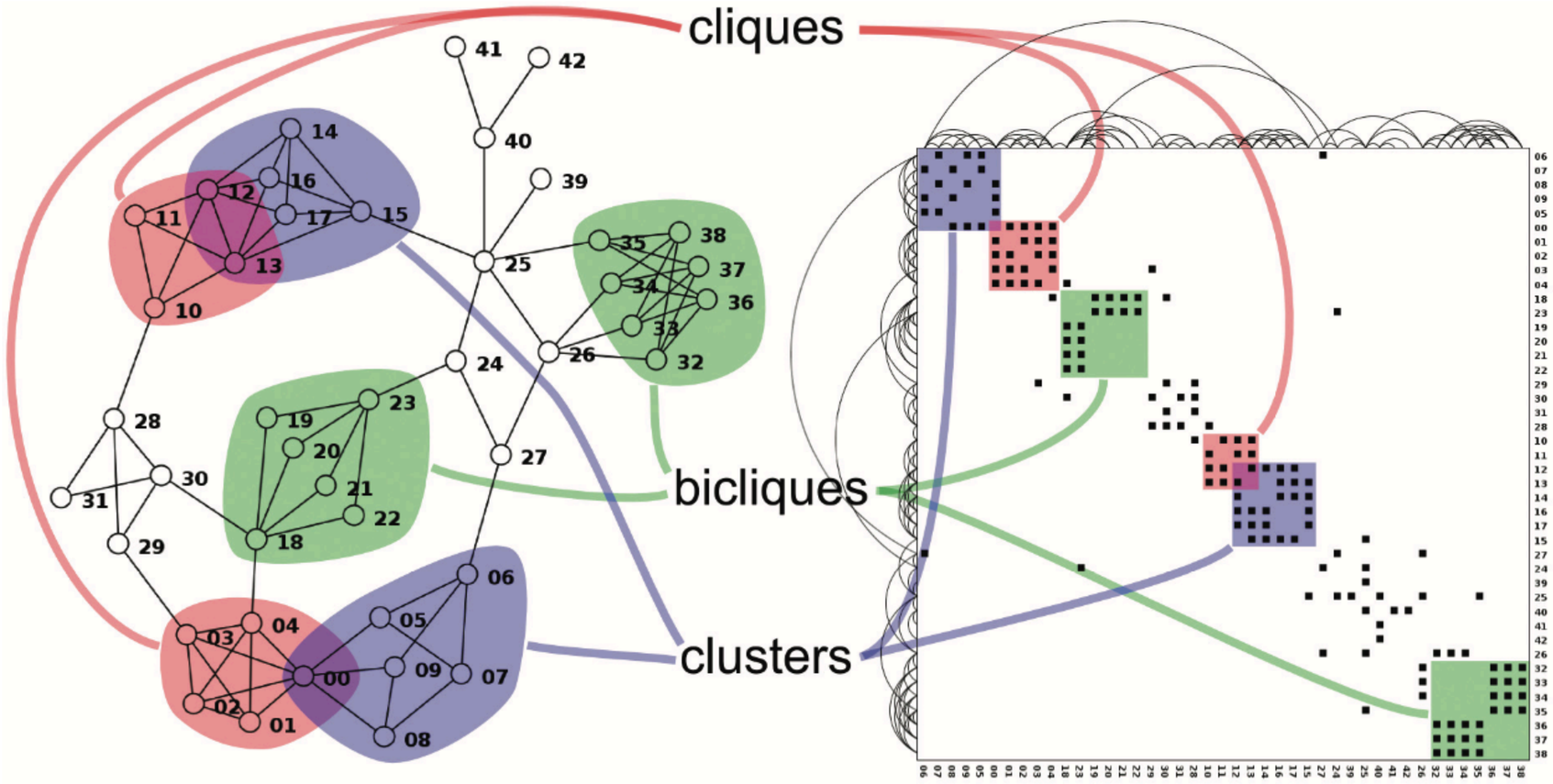
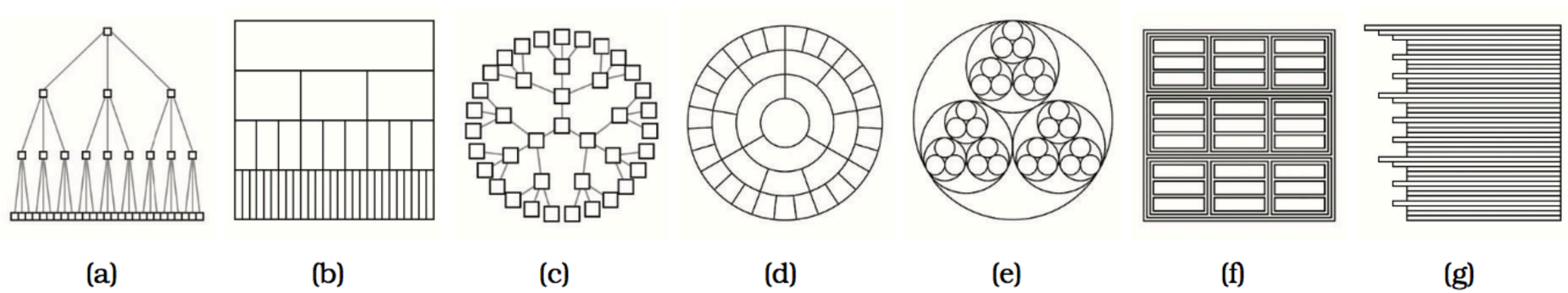


Figure 9.7. Characteristic patterns in matrix views and node–link views: both can show cliques and clusters clearly. From [McGuffin 12, Figure 6].

Nested Representations

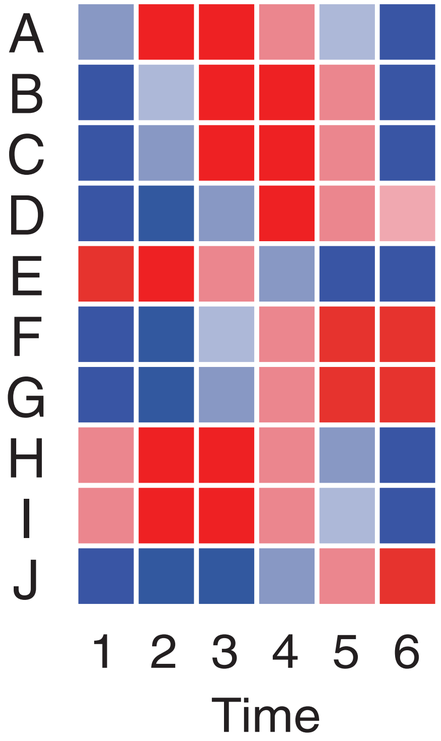


Munzer, Visualization Analysis and Design, 2014

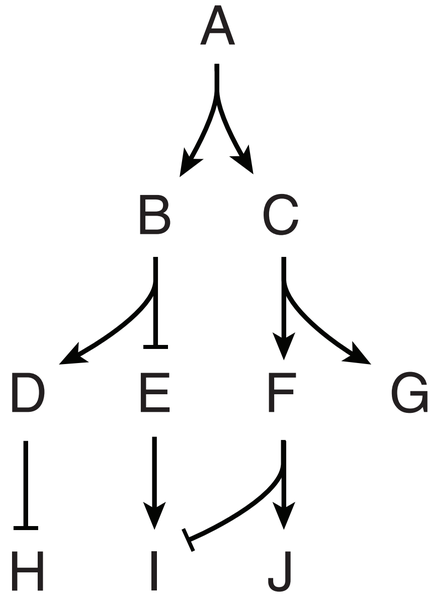
VIEW Map Bubbles
Standard and Poor's 500 index stocks categorized by sectors and industries. Size represents market cap.
Share map
MAP FILTER
S&P 500
World
Full
Exchange Traded Funds
1 Month Performar
Quick search ticker
A Agilent Technologies, Inc.
AA Alcoa Corporation
AAL American Airlines Group
AAP Advance Auto Parts, Inc.
AAPL Apple Inc.
ABBV AbbVie Inc.
ABC AmerisourceBergen
ABT Abbott Laboratories
ACN Accenture plc
ADBE Adobe Systems
TECHNOLOGY
INTERNET INFORMATION PROVIDERS
GOOGL -2.82%
FB +0.19%
BUSINESS SOFTWARE & SERVICES
MSFT +1.27%
TELECOM SERVICES - DOMESTIC
T
VZ
SEMICONDUCTOR - BRO
INTC +1.28%
AVGO -6.33%
TXN -3.22%
APPLICATION SOFTWA
ORCL +3.69%
NETWORKING & CO
CSCO -3.56%
COMMUNIC
QCOM -8.90%
INFORMATION TECHN
IBM -2.88%
ACN -5.75%
SEMICOND
NVDA -4.39%
SERVICES
ENTERTAINMENT - DIVERSIFI
CMCSA +0.10%
DIS +0.65%
CATALOG & MAIL ORDER H
AMZN +5.13%
DISCOUNT, VARIETY STORES
WMT +3.85%
HOME IMPROVEM
HD -0.42%
LOW
BUSINESS SERVI
PCLN
RAILROADS
UNP +1.15%
CSX -2.99%
NSC -5.54%
AIR DELIVERY &
UPS -2.43%
FDX -2.65%
CATV SYSTEMS
CHTR +3.06%
NFLX +0.45%
RESTAURANTS
MCD +2.71%
DRUG STO
WBA -3.31%
SPECIALT
SBUX +6.10%
DEPART
TJX -1.52%
APPARE
ROST
LODGI
MAR
SPECIALTY
ULTA
DRUGS
MCK -2.19%
CAH
AUTO P
AZO -4.89
AAP
REGION
SYY +1.90%
MAJOR AI
DAL -2.66
UAL
AAL
RESORTS
CCL +1.30
LUV +4.97%
GROCER
KR
CONSUMER GOODS
ELECTRONIC EQUIPMENT
AAPL +2.02%
BEVERAGES - SOFT DRI
KO +2.31%
PEP +3.90%
PERSONAL PRODUCTS
PG
CL +0.47%
KMB -0.33%
EL +0.17
PROCESSED &
GIS -3.97%
K
SJM
MJN
AUTO MAN
GM -8.17%
F -10.52
CONFECTI
MDLZ +1.65%
HSY
AUTO
JCI +0.62%
CIGARETTES
PM +3.66%
MO -4.22%
RAI
TEXTILE - APP
NKE -2.98%
FOOD - MAJOR
KHC +0.95%
PACKAGI
BEVER
STZ
NWL -3.69%
TSN +2.54%
CLX
HRL
BF-B
TEXTILE -
CLEANING
ECL
ADM
HAS
BEVER
BASIC MATERIALS
MAJOR INTEGRATED OIL
XOM +2.44%
CVX +1.50%
INDEPENDENT
COP +10.01%
EOG +3.24%
OXY +3.20%
APC +1.18%
PXD +4.25%
HES
DVN
APA
OIL & GAS EQU
SLB
HAL -0.94%
FTI
CHEMICALS - M
DOW -2.07%
PX
APD +0.20
AGRICULTU
DD
MON +2.50%
OIL & G
PSX -1.44%
SPECIALT
LYB -2.53
PPG +3.27
OIL & GAS PI
KMI +2.95%
HEALTHCARE
DRUG MANUFACTURERS - M
JNJ -1.30%
PFE -0.88%
MRK -1.79%
ABBV -1.00%
BMY -5.85%
LLY +2.12%
BIOTECHNOLOGY
AMGN -9.48%
CELG +0.38%
GILD -2.29%
BIIB -8.05%
REGN -6.33%
HEALTH CARE PLANS
UNH -2.07%
CVS -2.42%
AET -1.80%
ESRX -0.45%
ANTM +0.76%
CI
MEDICAL APPLIANCES
MDT -2.42%
ABT -2.35%
SYK -0.11%
BSX +0.41%
ISRG
ZBH
EW
DRUGS - GENERI
AGN -0.68%
ZTS
MYL -11.06
MEDICAL INSTR
BDX +0.13%
BAX
MEDICAL L
TMO -3.71%
HCA +5.40%
DVA
UHS
INDUSTRIAL GOODS
DIVERSIFIED MACHINERY
GE
MMM -0.33%
DHR -1.12%
ITW -1.30%
ETN +3.37%
CMI -2.63%
HON -2.04%
ROP -2.59%
ROK -1.46
IR
AEROSPACE/DEFENSE PRODUC
BA
LMT +0.19%
GD -1.36%
UTX +0.37%
RTN -1.46%
FARM &
CAT +3.27%
DE
AEROSPAC
NOC -1.81%
GENER
WY +3.81
WASTE M
WM -0.93
RSG -0.06
INDUS
EMR -0.93%
PH
UTILITIES
ELECTRIC UTILITIES
NEE -0.05%
DUK +2.36%
SO -0.44%
D
EIX +3.73
PPL +3.71
ED +2.83
PCG +3.57%
XEL +3.81%
FE
ETR
AEP +3.04%
WEC
AEE
CMS
DTE
SCG
LNT
PNW
AES
DIVERSIFIED UTILI
EXC +1.54
PEG
CNP
ES
OKE
AWK
FINANCIAL
MONEY CENTER BANKS
JPM -6.14%
BAC -10.54%
WFC -9.60%
C
PNC -7.67%
STI -7.93%
PROPERTY & CASUALTY INSURANCE
BRK-B -5.51%
AIG -4.31%
TRV -0.95
ALL -1.47%
PGR -1.21%
XL
CREDIT SERVICES
V
MA
AXP -3.98%
PYPL +0.93%
COF -10.17%
DFS
SYF
EFX
TSS
INVESTMENT BR
GS -8.87%
MS -11.09%
SCHW -10.01%
CME -6.29%
ASSET MANAGE
BLK -0.48%
BK -2.84%
STT -1.31%
BEN -3.36%
LIFE INSURANC
MET -3.56%
PRU -6.51%
REGIONAL - MIDW
USB -7.33%
FITB
KEY
REGIONAL
BBT -9.05%
RF
REIT - RETAIL
SPG +3.62%
GGP
O
KIM
REIT - I
PSA +2.26
PLD
EXR
REIT - R
AVB +5.73
EQR
ESS
REIT -
AMT +7.87%
VNO
ACCIDE
AFL +2.10
REGION
CFG -9.30
ZION
HCN
VTR
INSURA
MMC -1.60%
AON +0.21
IVZ
PFG -1.71
LNC -6.80
TMK
DIVERSIFI
ICE -0.63%
BXP
REIT -
REIT - R
REIT - I
REIT -
REIT -
REGION
REGION
REGION

Graphs and Multivariate Data

Interactions: And Multivariate Data?



+



= ?

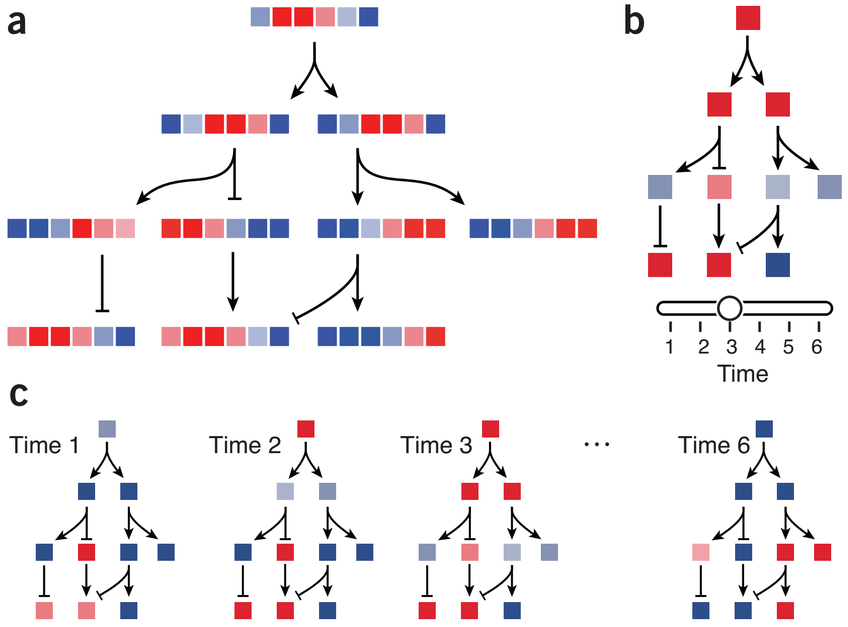
Gehlenborg and Wong, Nature Methods, 2012

Interactions: And Multivariate Data!?

(a) **Complex glyphs** = multiple values per node?

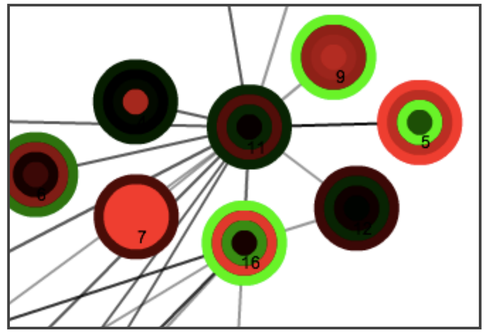
(b) **Animation** = one value per node, one network shown at a time?

(c) **Small multiples** = one value per node, all networks shown

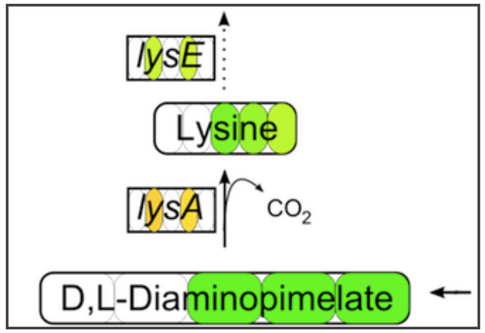


Gehlenborg and Wong, Nature Methods, 2012

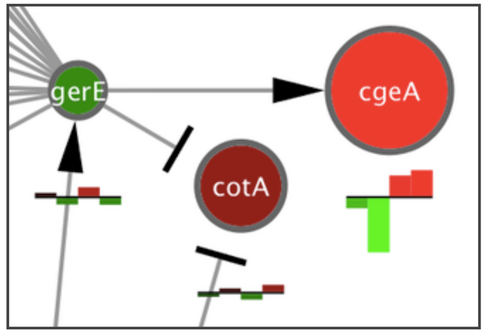
Interactions: And Multivariate Data!



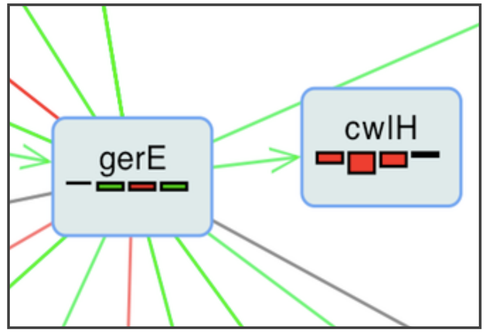
Lichen



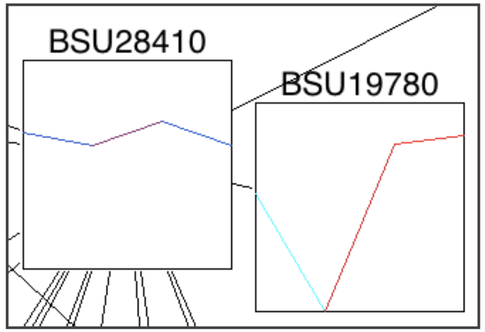
Prometra



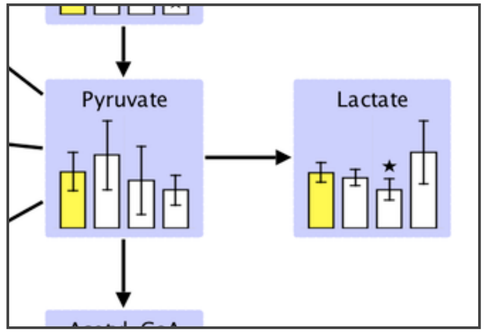
VistaClara (Cytoscape)



GENeVis



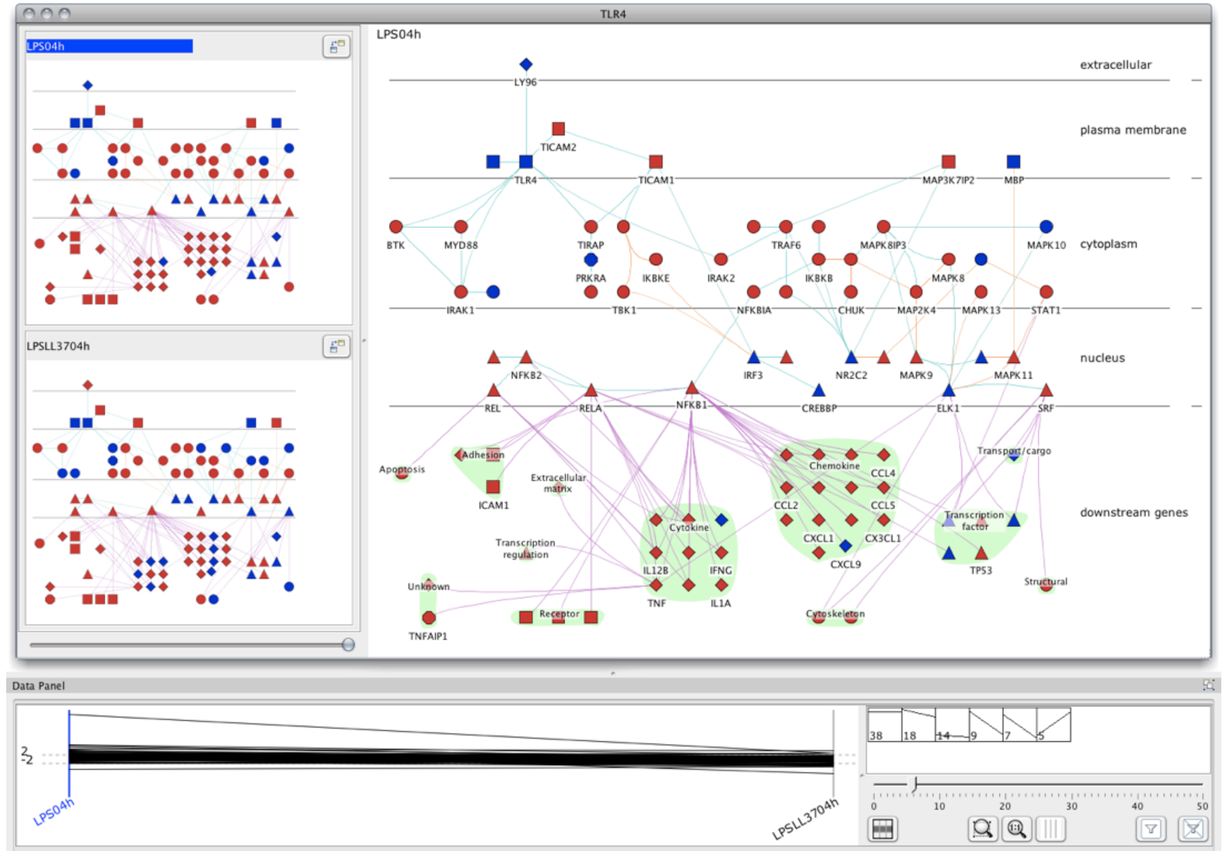
VisANT



VANTED

Gehlenborg et al., Nature Methods, 2010

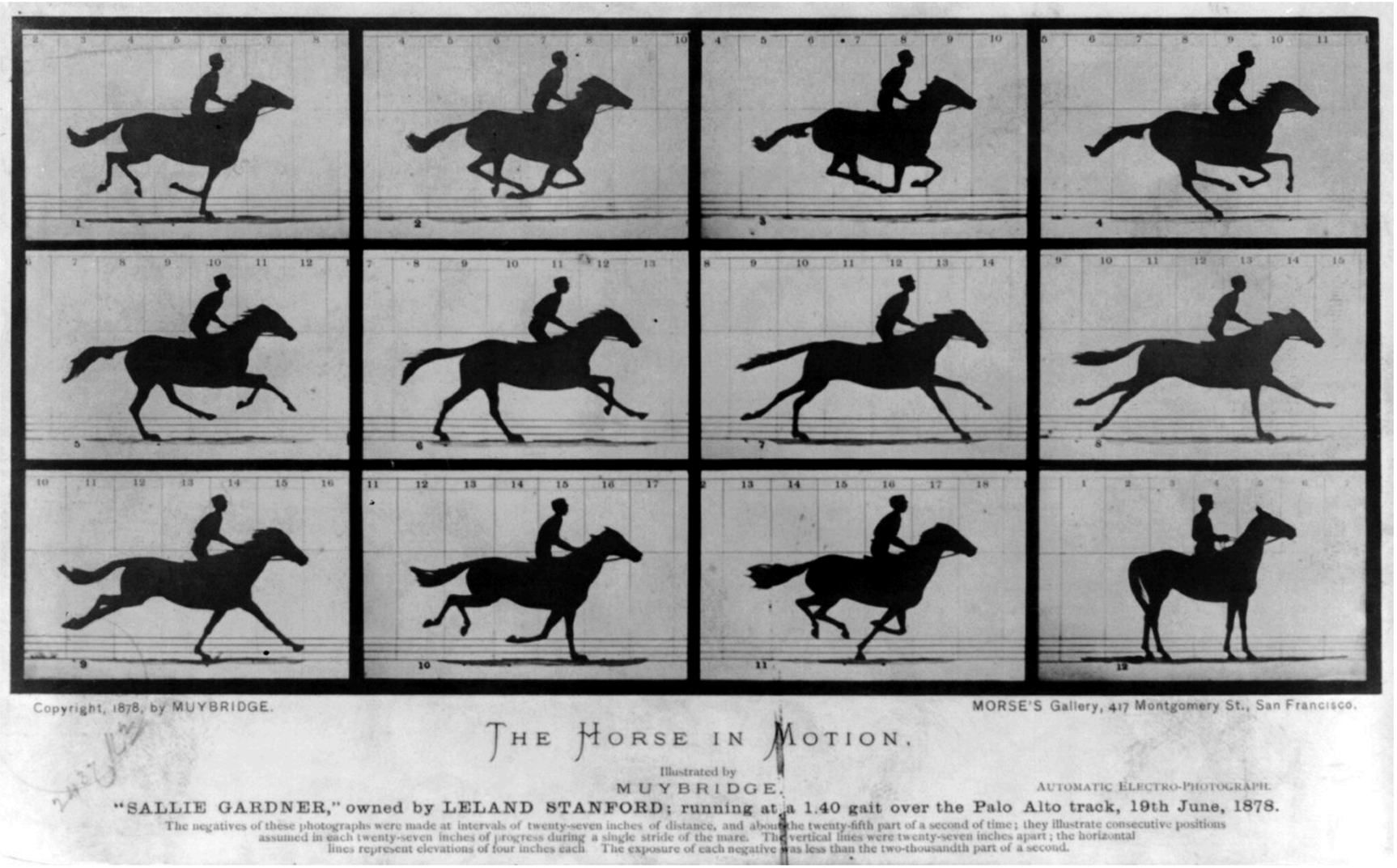
Interactions: And Multivariate Data!



Cerebral (Cytoscape plugin)

Gehlenborg et al. 2010, Nature Methods

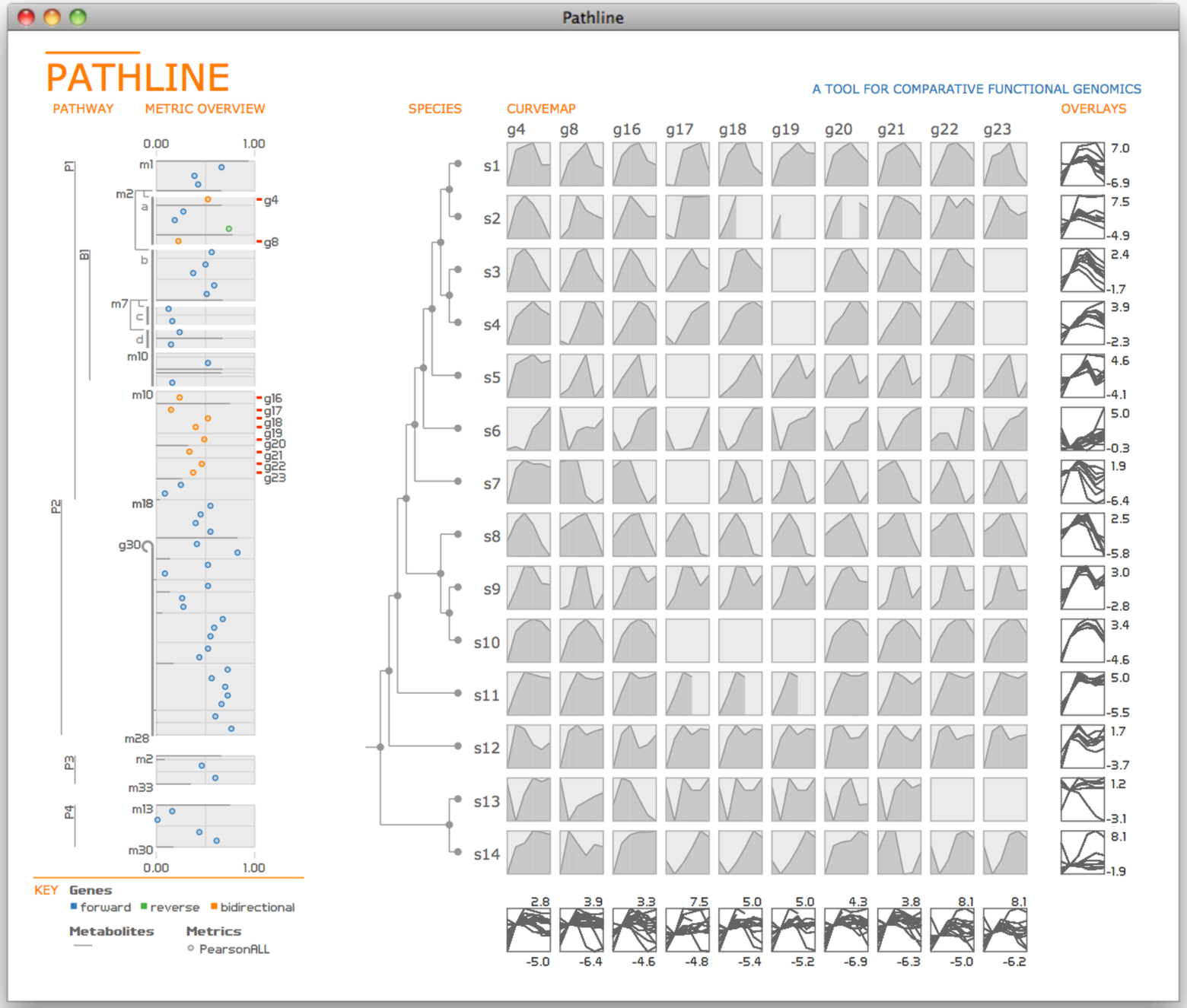
Small Multiples



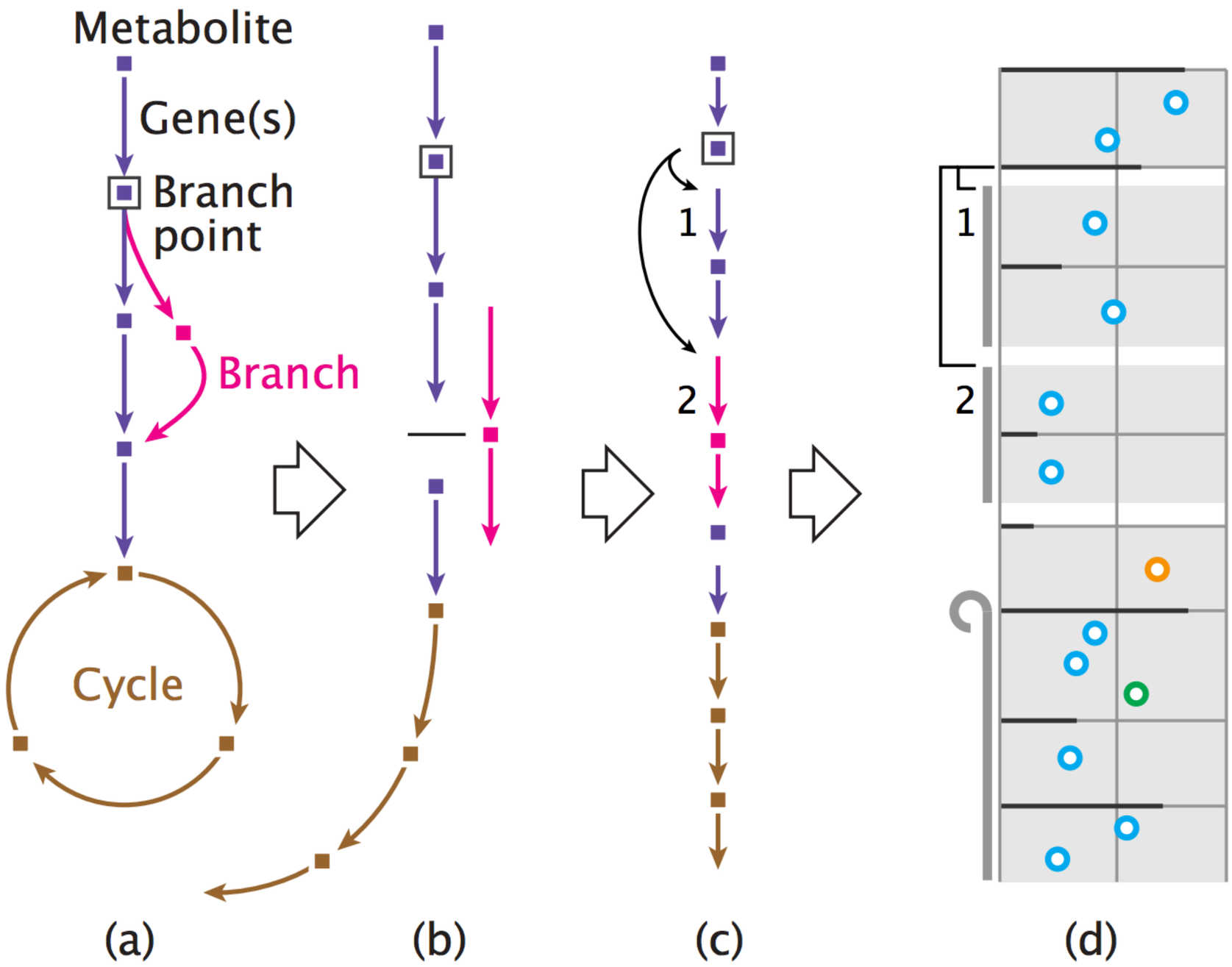
Small Multiples

http://www.nytimes.com/interactive/2012/07/20/us/drought-footprint.html

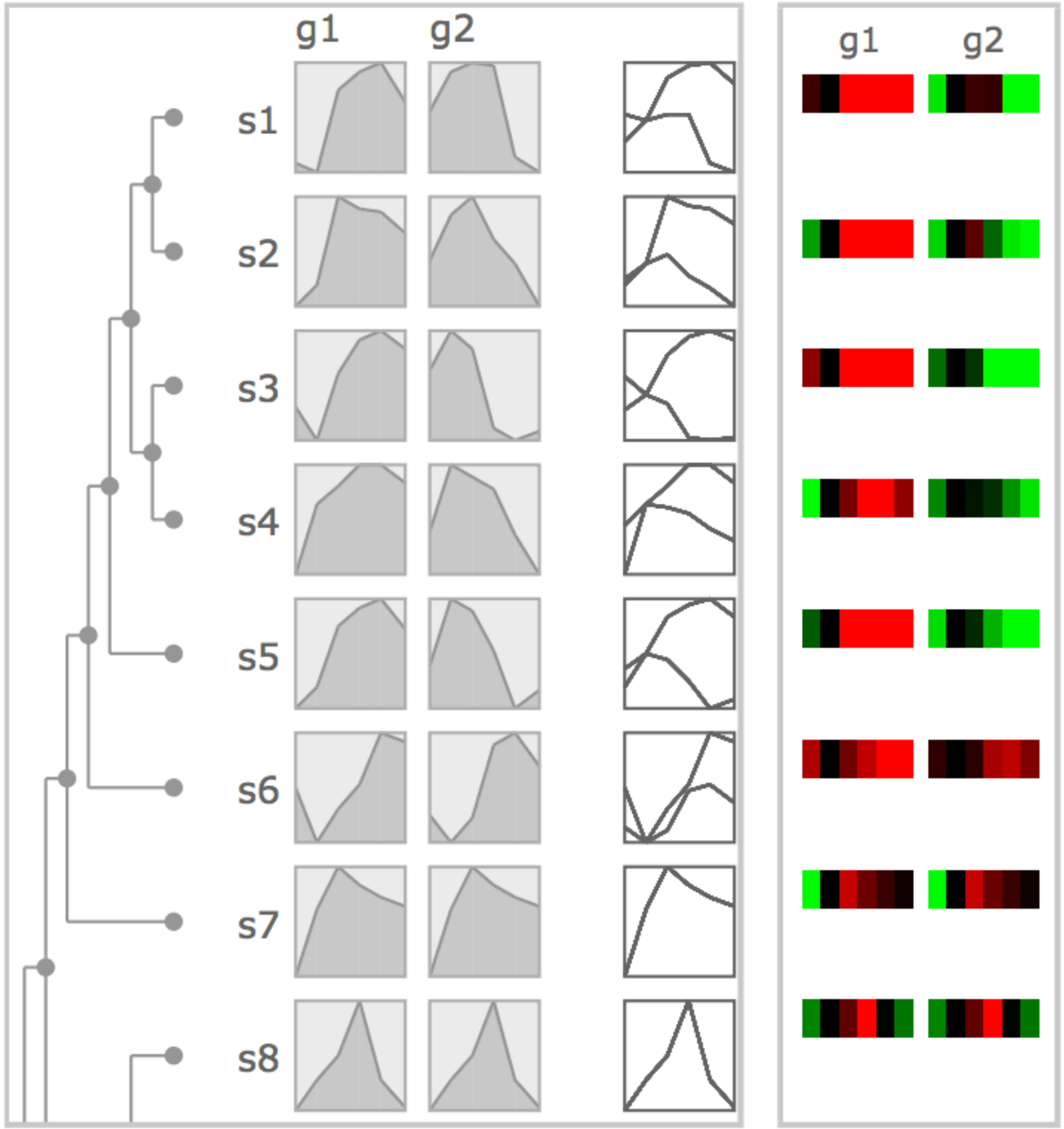
Juxtaposed Views



Meyer et al., EuroVis, 2010



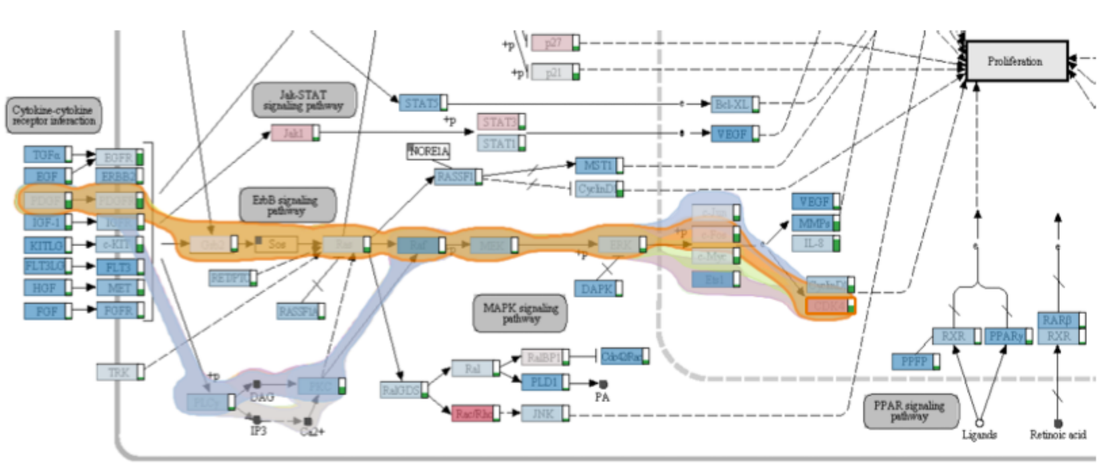
Meyer et al., EuroVis, 2010

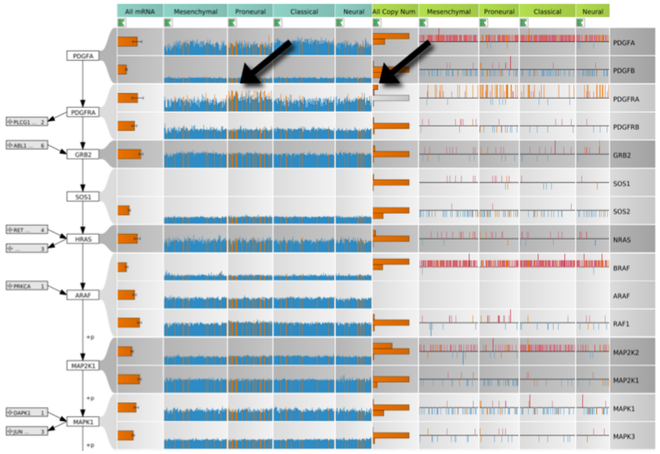


(a) (b)

Meyer et al., EuroVis, 2010

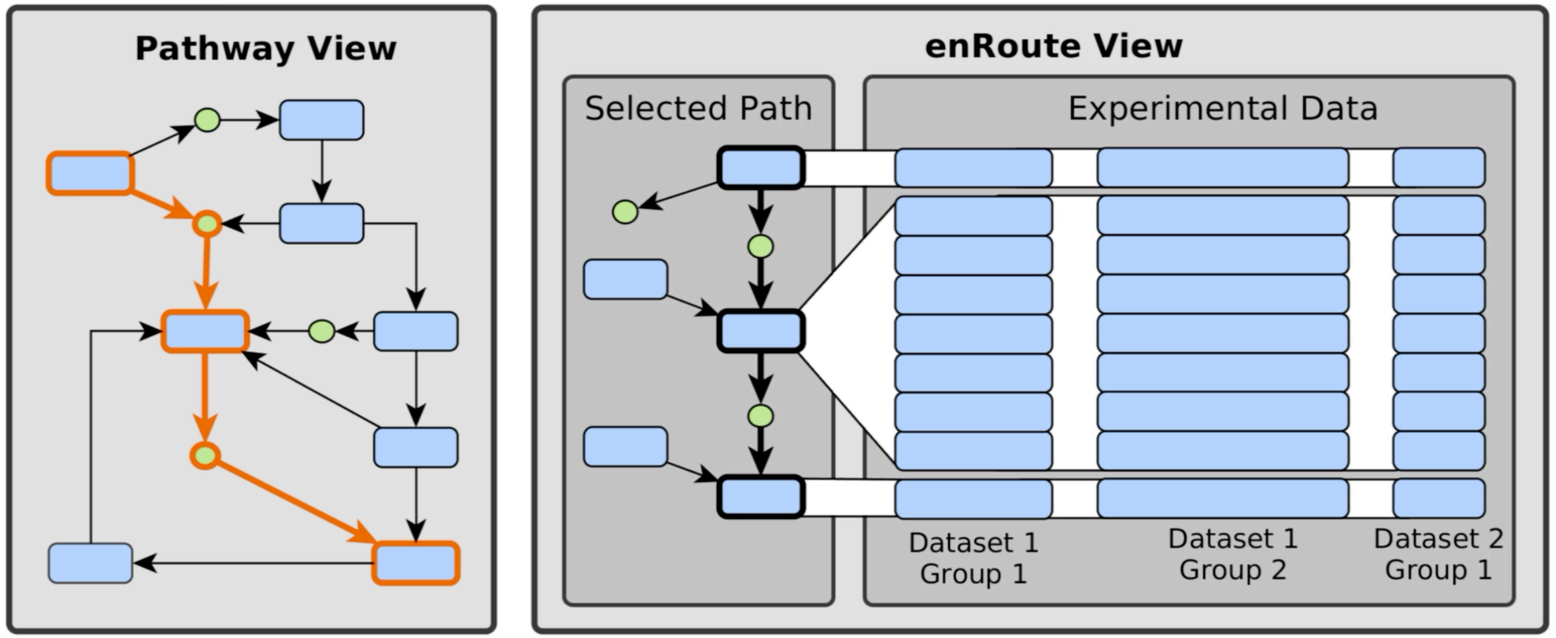
Interactions: And Multivariate Data!





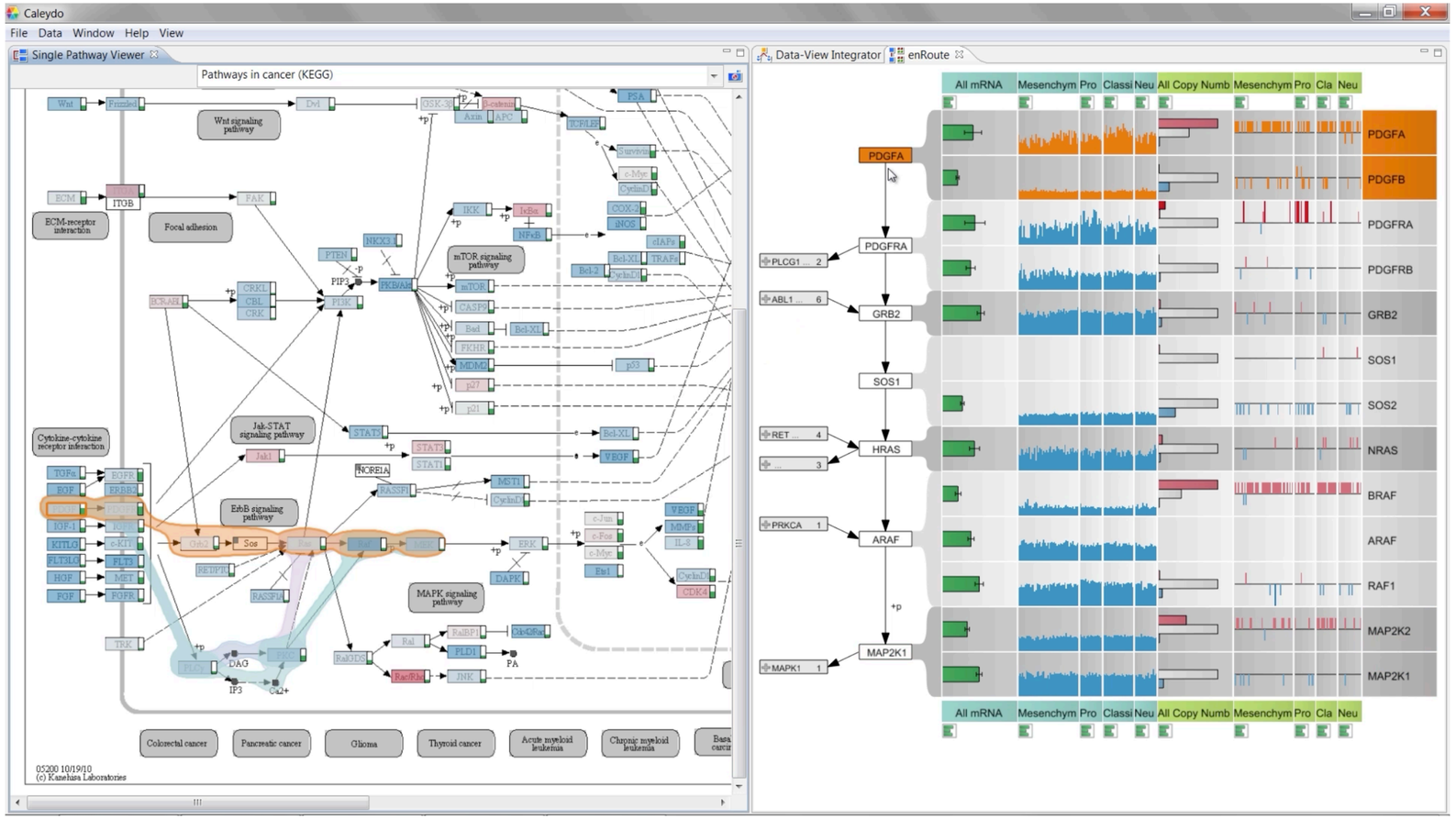
Partl *et al.*, enRoute: Dynamic Path Extraction from Biological Pathway Maps, *Proc. BioVis*, 2012, http://enroute.caleydo.org

EnRoute



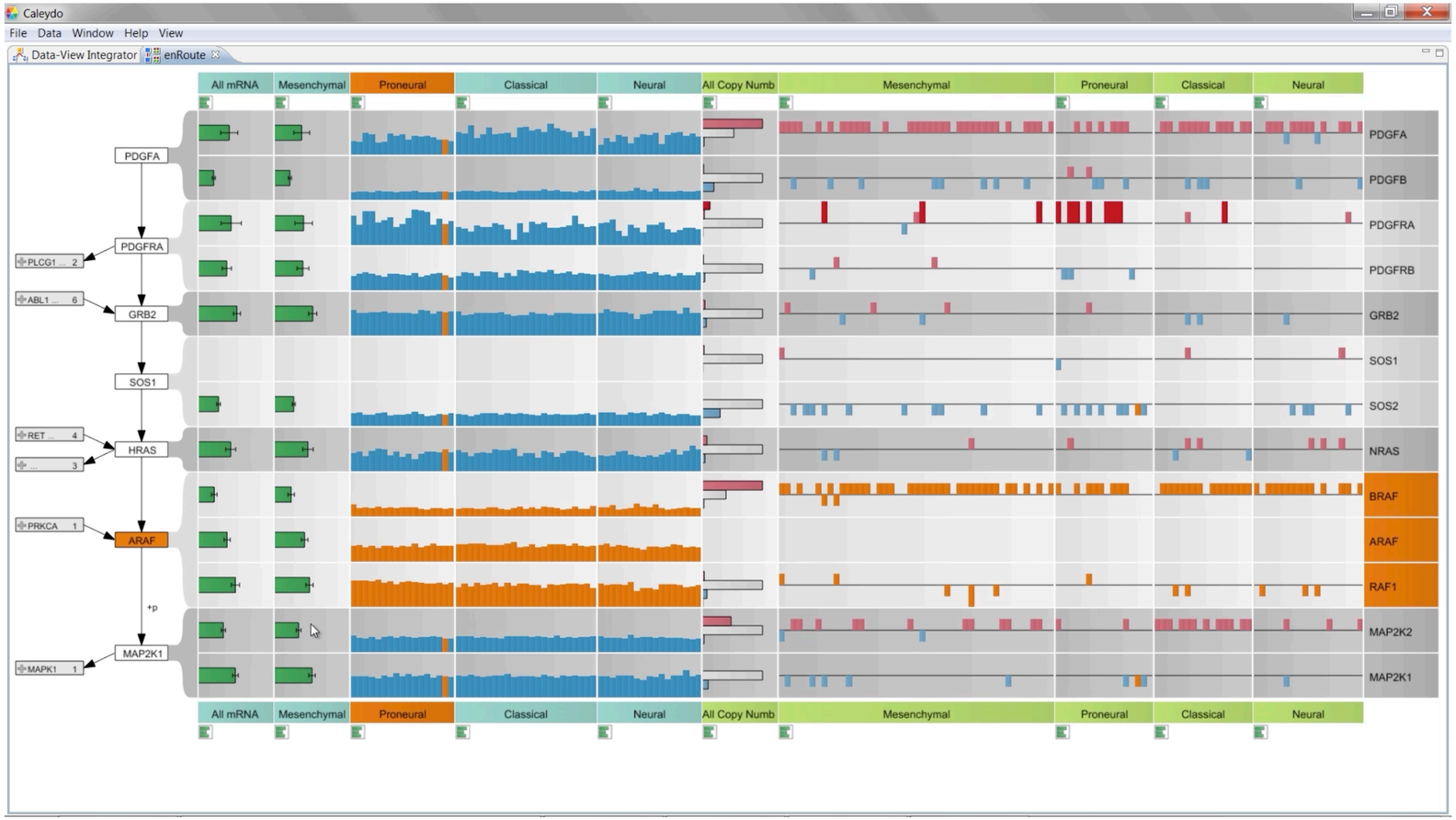
Partl *et al.*, enRoute: Dynamic Path Extraction from Biological Pathway Maps, *Proc. BioVis*, 2012, http://enroute.caleydo.org

EnRoute



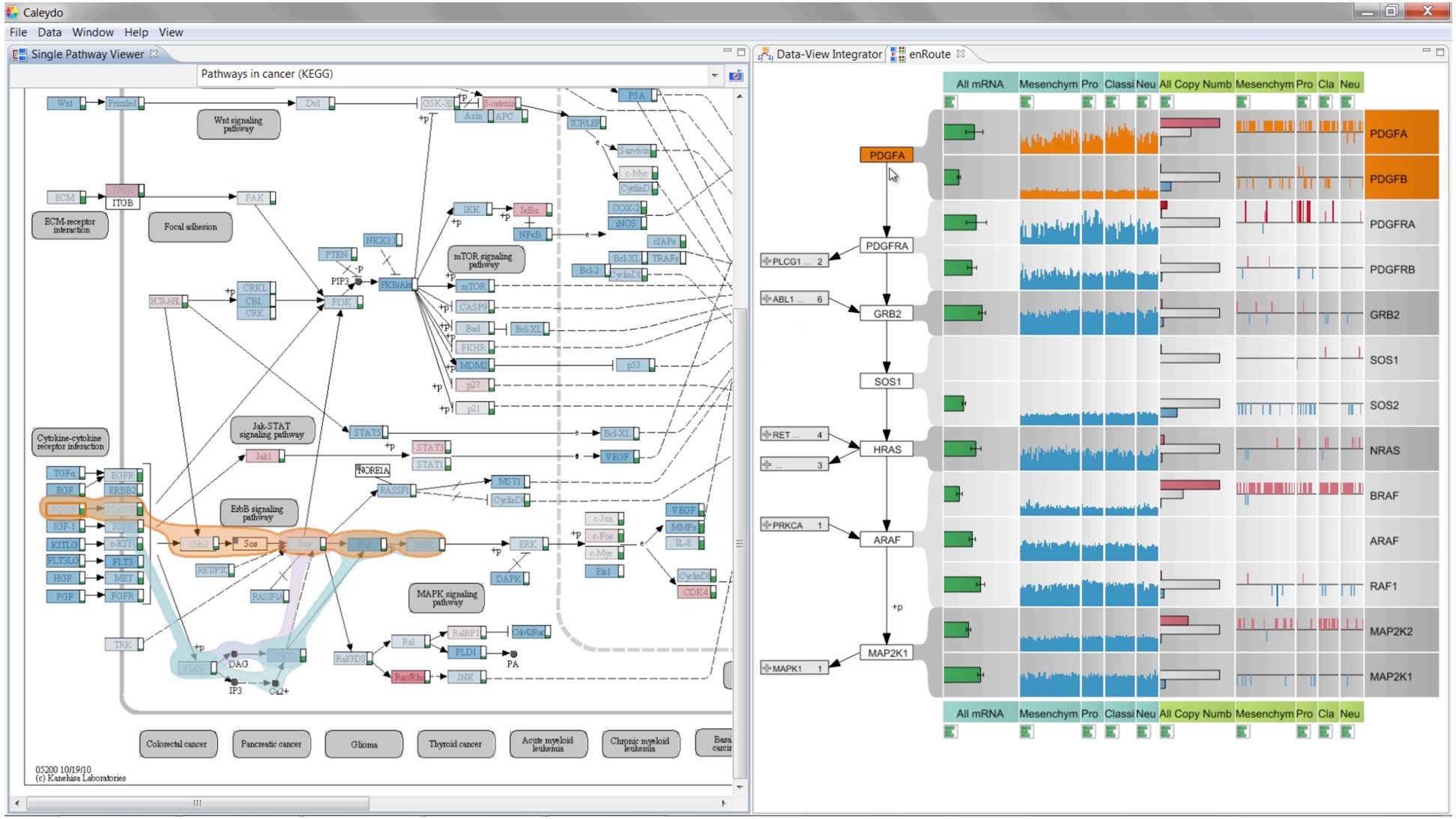
Partl *et al.*, enRoute: Dynamic Path Extraction from Biological Pathway Maps, *Proc. BioVis*, 2012, http://enroute.caleydo.org

EnRoute



Partl *et al.*, enRoute: Dynamic Path Extraction from Biological Pathway Maps, *Proc. BioVis*, 2012, http://enroute.caleydo.org

Exploration of Heterogeneous Data in Network Context



Partl et al., IEEE BioVis, 2012