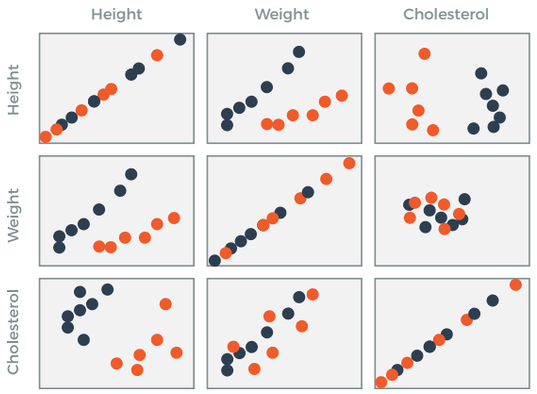
Eamonn Maguire, DPhil

Principles of Data Visualization - Lecture 2
iCSC, CERN
March 2017

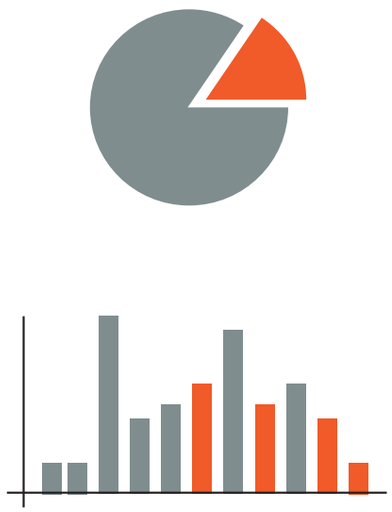
And that was just to represent a low number of dimensions

What happens when we have a high number of dimensions?

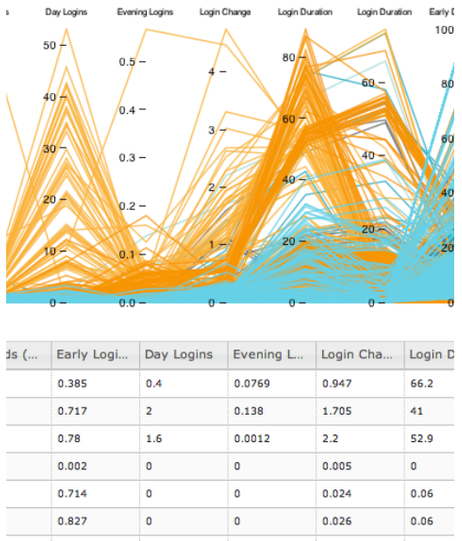
Multidimensional Visualization



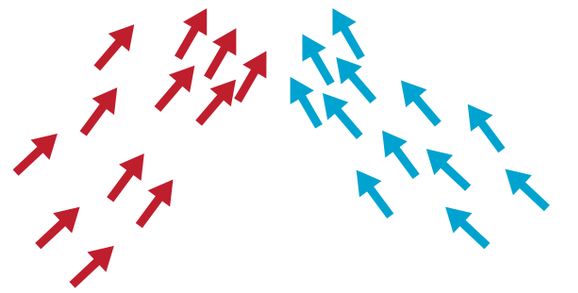
Scatter Plot Matrices



Linked Plots



Parallel Coordinates



Temperature - Colour

Wind direction - Orientation

Wind Speed - Proximity

Location - Position

Glyphs

Multidimensional Visualization

Scatter Plot Matrices

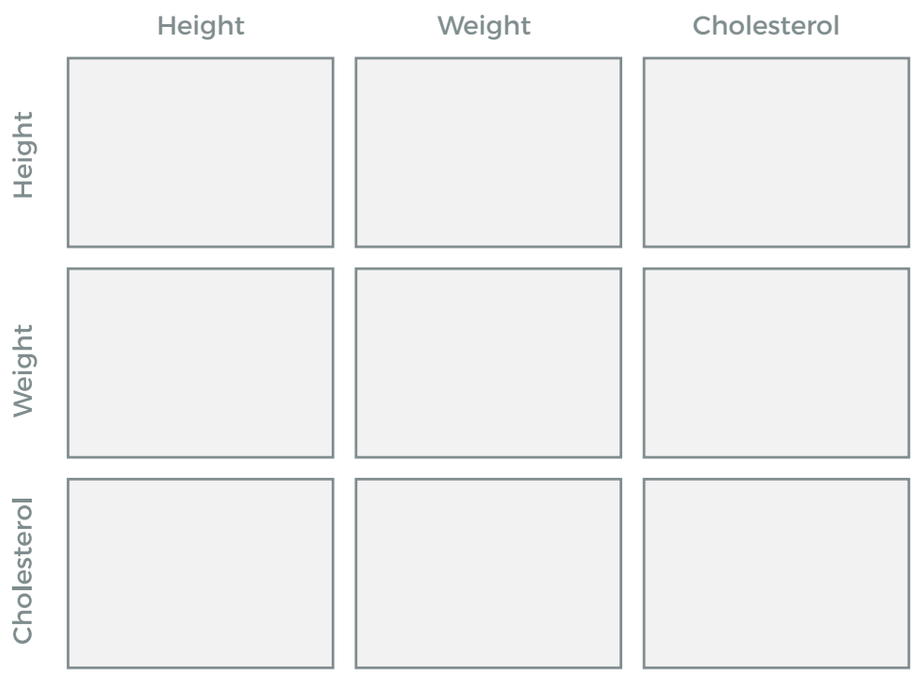
Name	Height	Weight	Chol
John	1.76	63	4.5
Mike	1.79	70	4.15
Jim	1.61	60	6.7
Francois	1.84	90	5.03
	...		

Multidimensional Visualization

Scatter Plot Matrices

Name	Height	Weight	Chol
John	1.76	63	4.5
Mike	1.79	70	4.15
Jim	1.61	60	6.7
Francois	1.84	90	5.03

...



Multidimensional Visualization

Scatter Plot Matrices

Name	Height	Weight	Chol
John	1.76	63	4.5
Mike	1.79	70	4.15
Jim	1.61	60	6.7
Francois	1.84	90	5.03

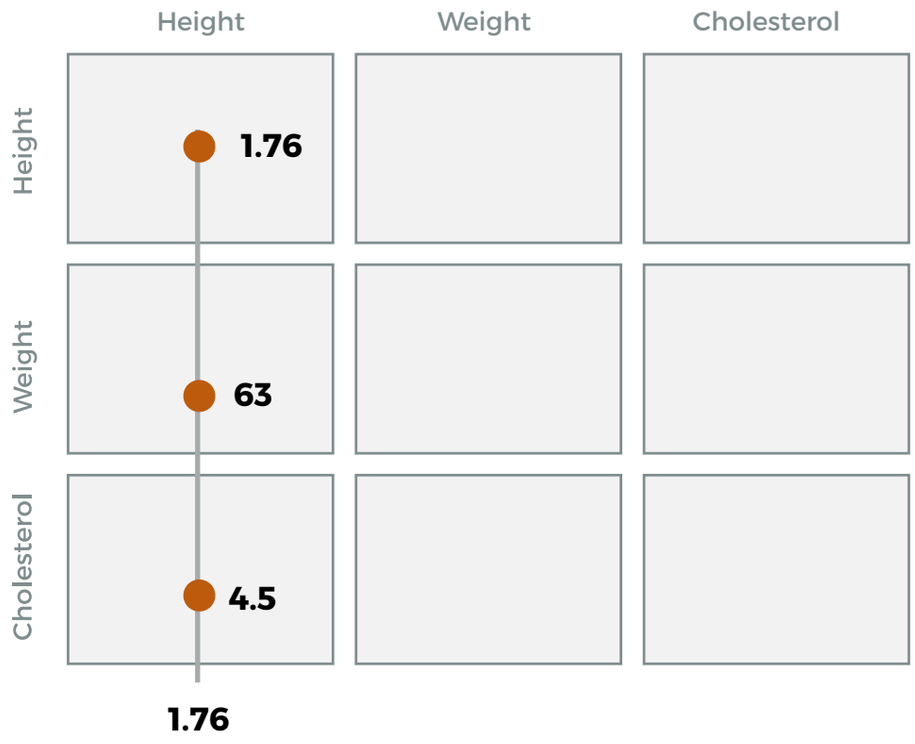
...

Multidimensional Visualization

Scatter Plot Matrices

Name	Height	Weight	Chol
John	1.76	63	4.5
Mike	1.79	70	4.15
Jim	1.61	60	6.7
Francois	1.84	90	5.03

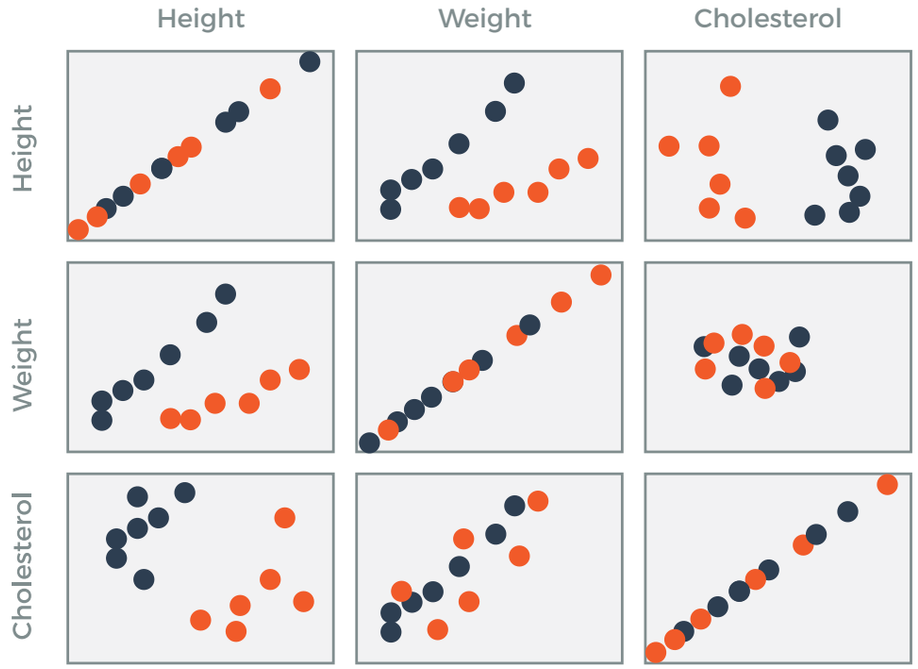
...



Multidimensional Visualization

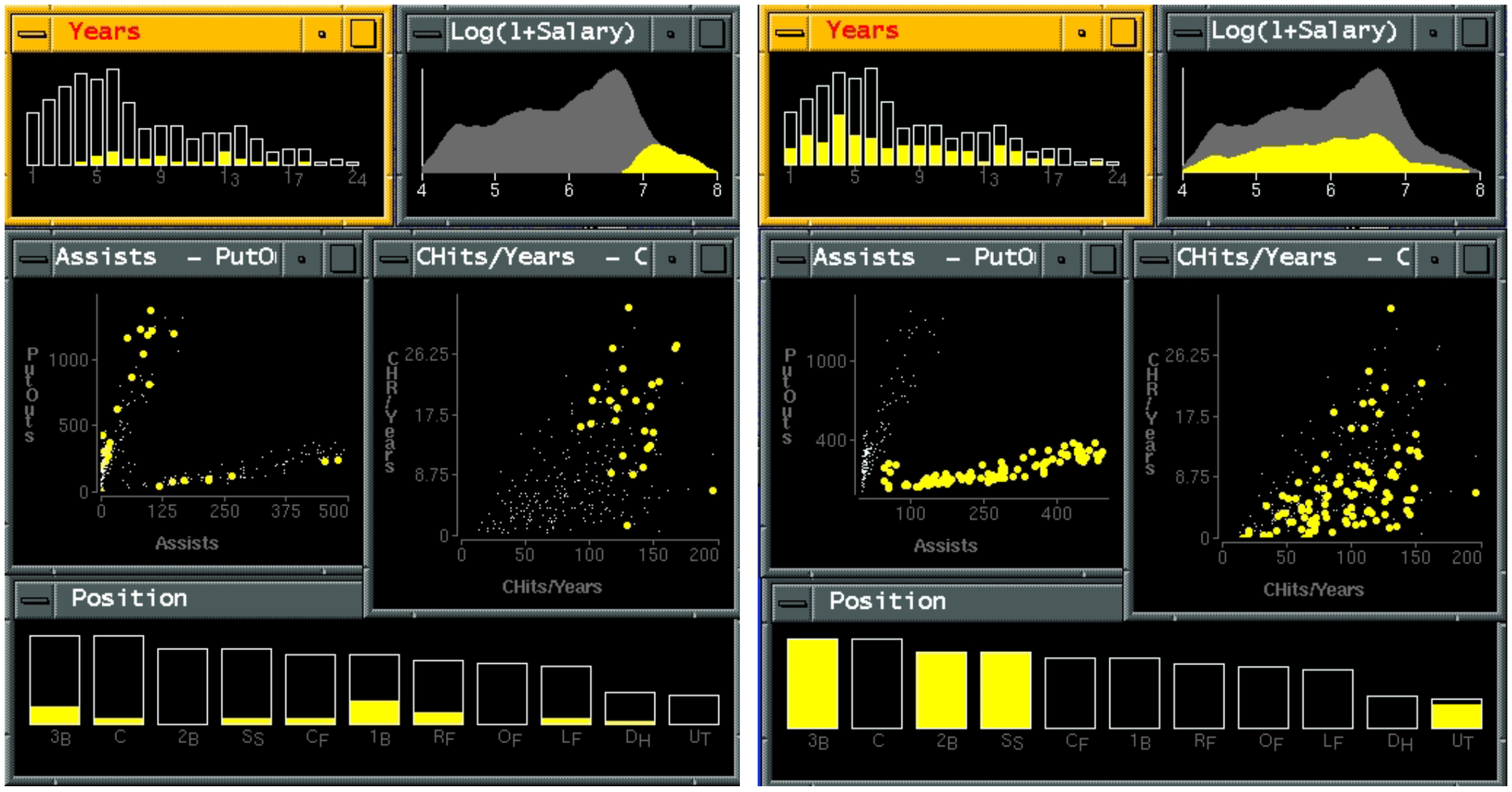
Scatter Plot Matrices

Name	Height	Weight	Chol
John	1.76	63	4.5
Mike	1.79	70	4.15
Jim	1.61	60	6.7



Multidimensional Visualization

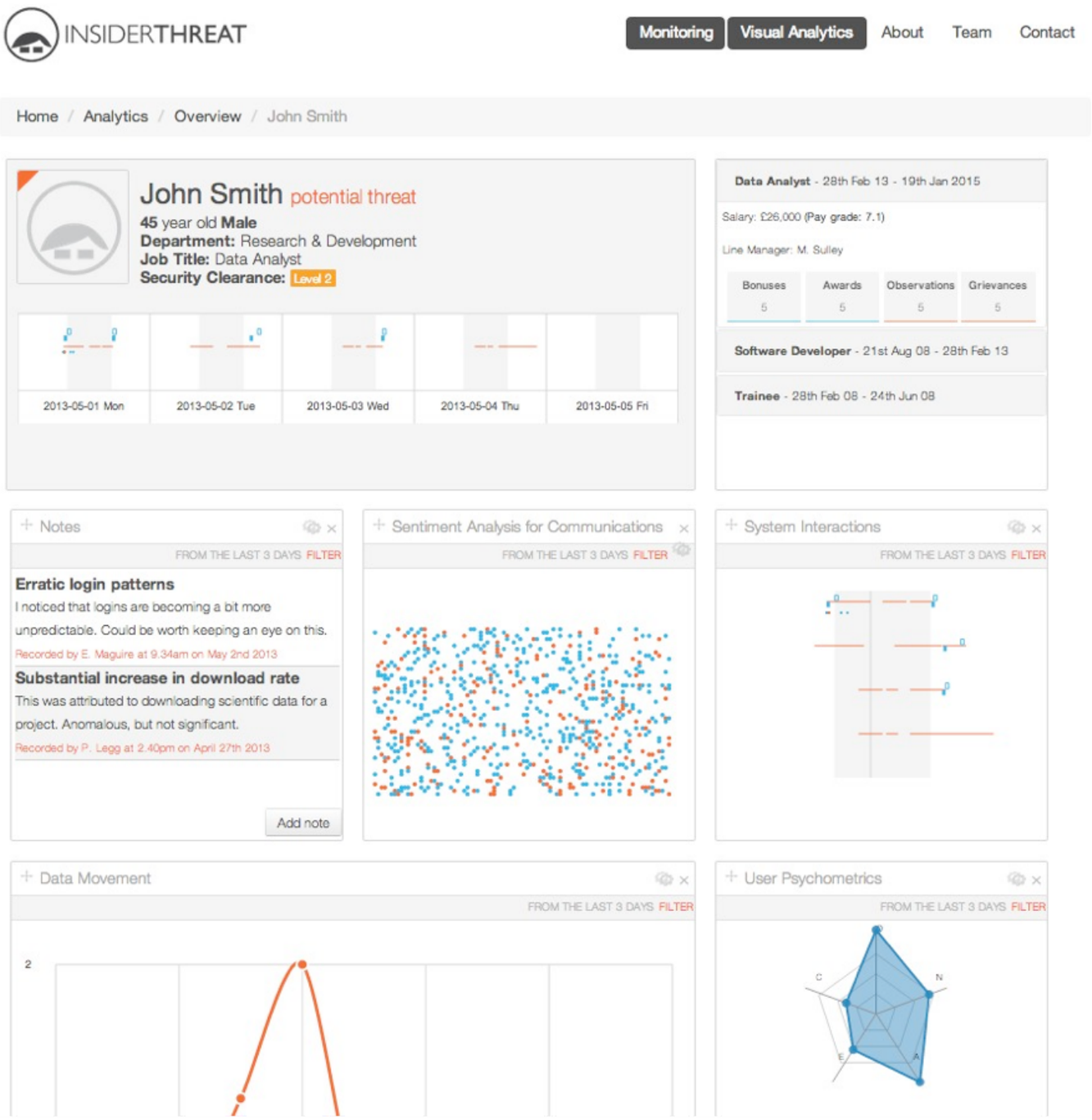
Linked Plots



Visual Exploration of Large Structured Datasets. Wills. Proc. New Techniques and Trends in Statistics (NTTS), pp. 237–246. IOS Press, 1995.

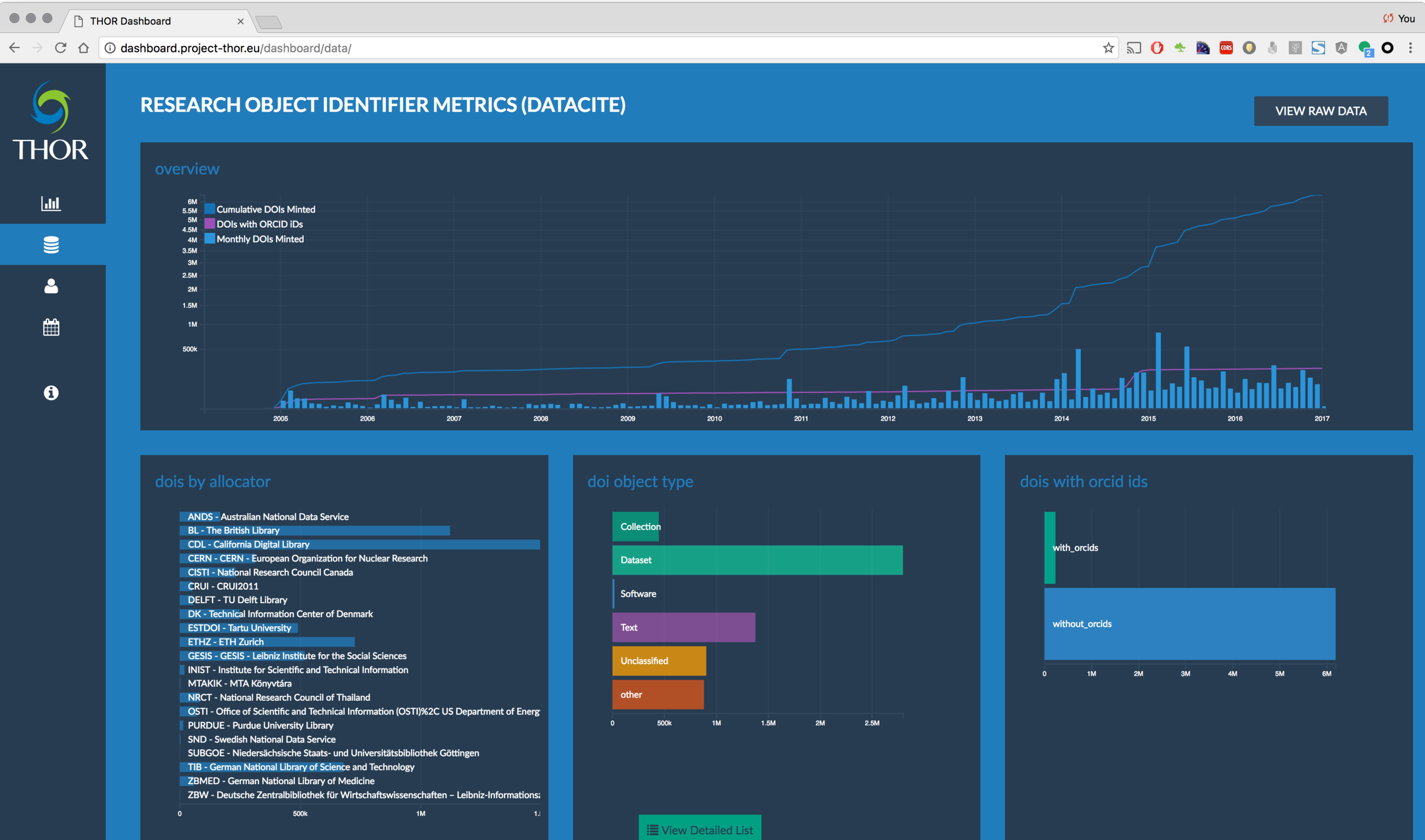
Multidimensional Visualization

When one visualization won't cut it...



Multidimensional Visualization

With dc.js, crossfilter, and d3.js



Multidimensional Visualization

With dc.js, crossfilter, and d3.js

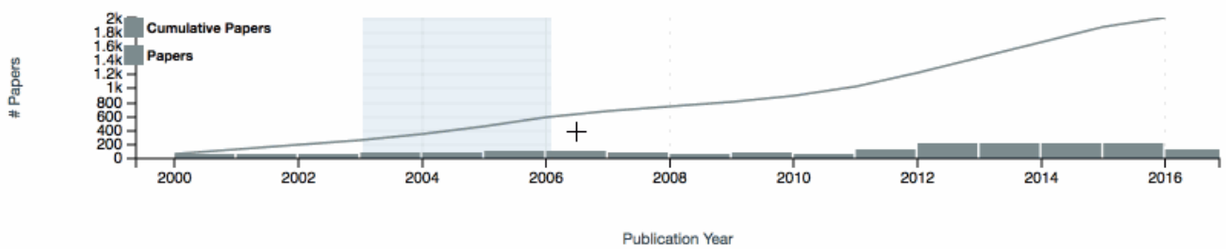
Cite Summary

Download as CSV

Paper Overview

2,017 out of 2414 Papers 794 with no citations not shown

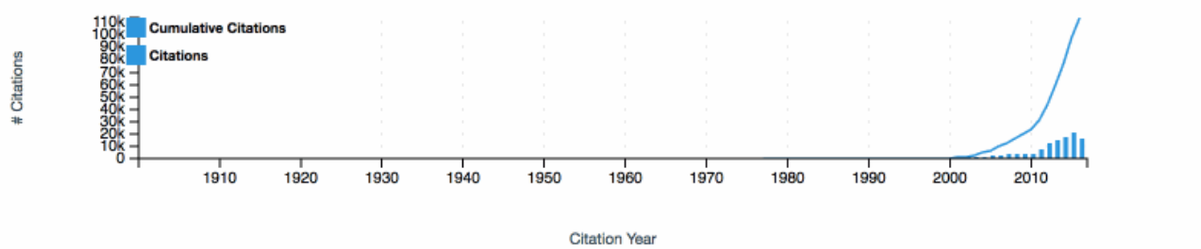
Reset All Filters



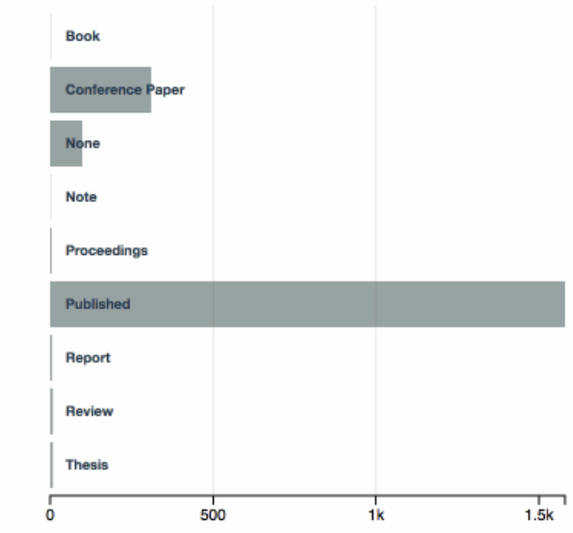
Citation Overview

114,029 out of 114,029 Citations

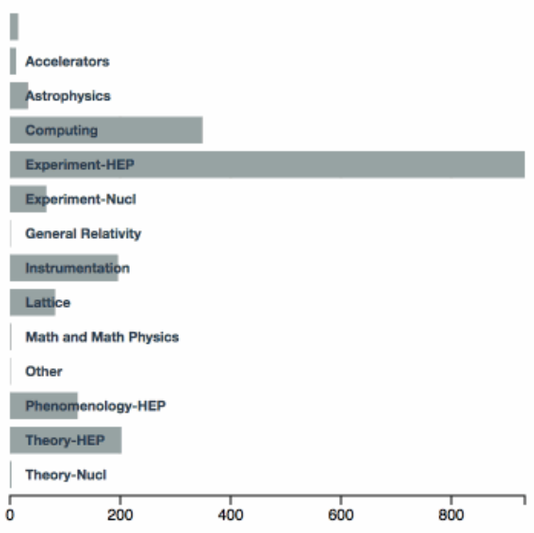
Reset All Filters



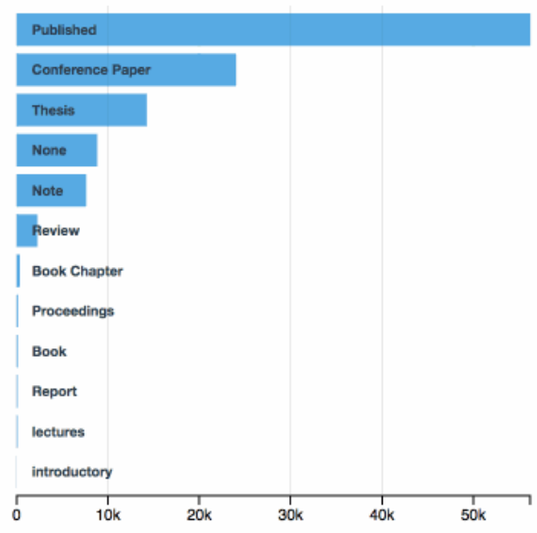
Paper Type



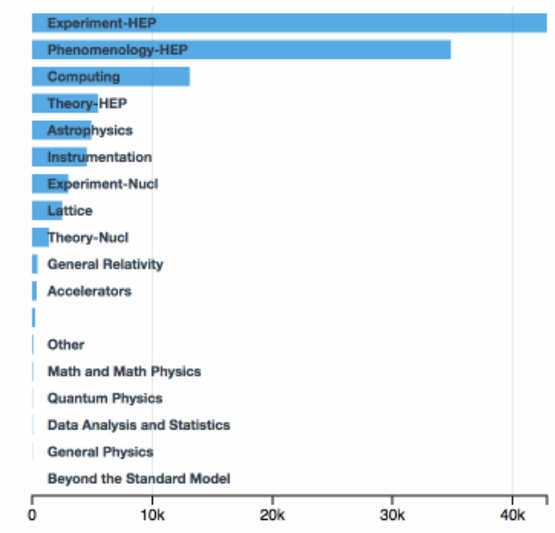
Paper Subject Area



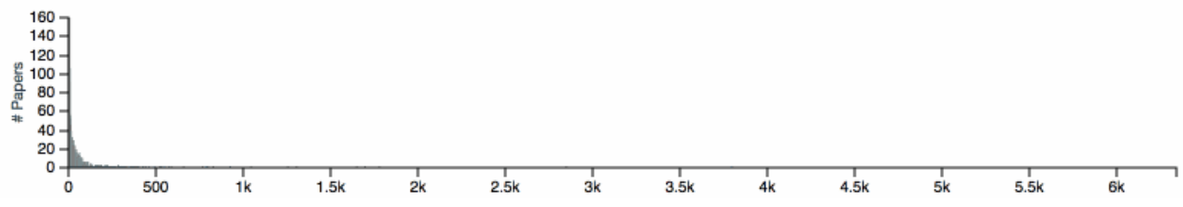
Citation Types



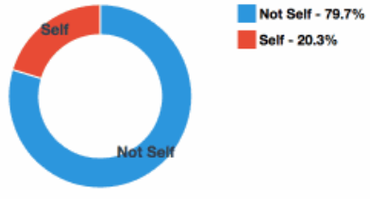
Citation Subject Area



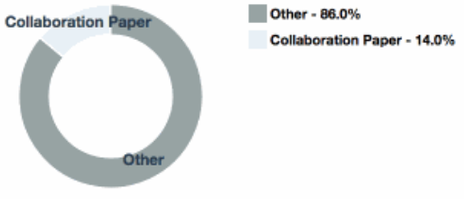
Paper Citation Counts



Self Citations



Collaboration Papers



Multidimensional Visualization

With dc.js, crossfilter, and d3.js

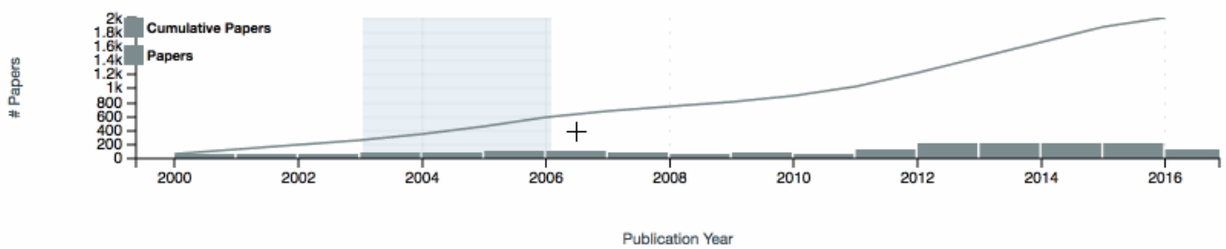
Cite Summary

Download as CSV

Paper Overview

2,017 out of 2414 Papers 794 with no citations not shown

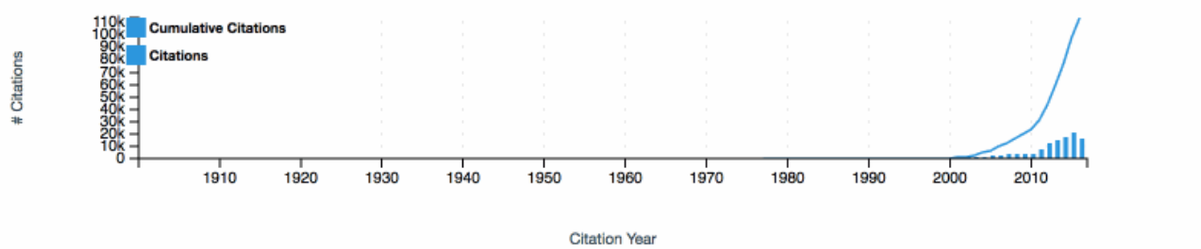
Reset All Filters



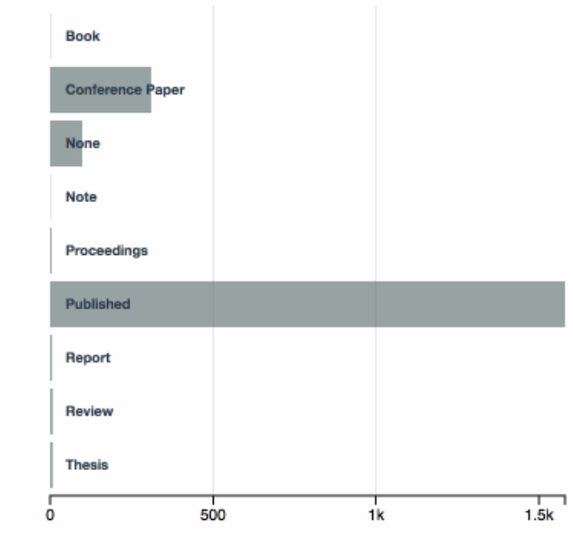
Citation Overview

114,029 out of 114,029 Citations

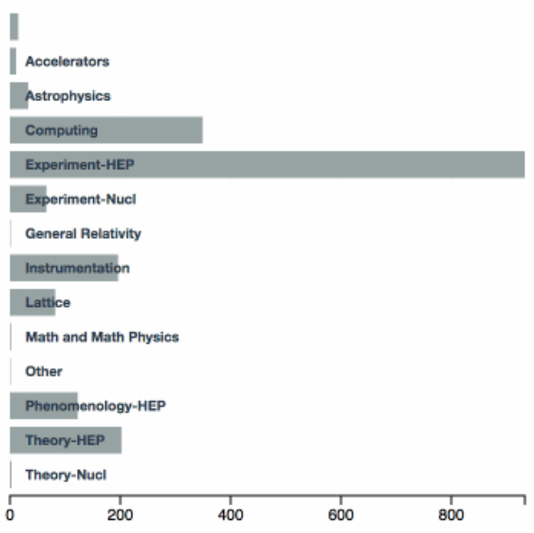
Reset All Filters



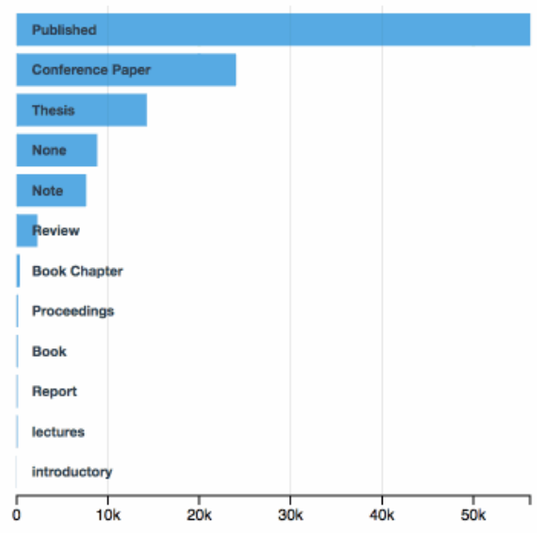
Paper Type



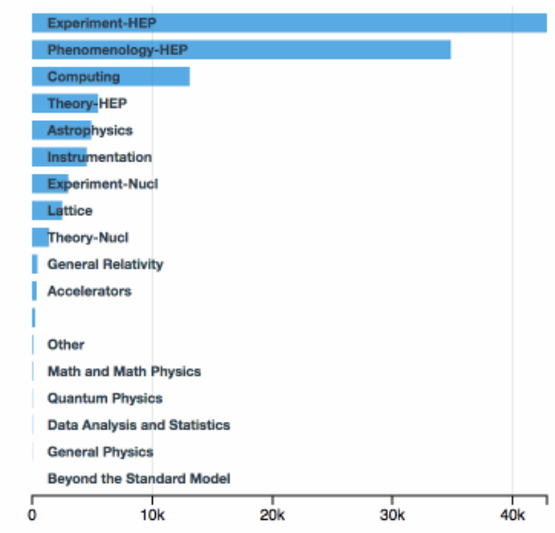
Paper Subject Area



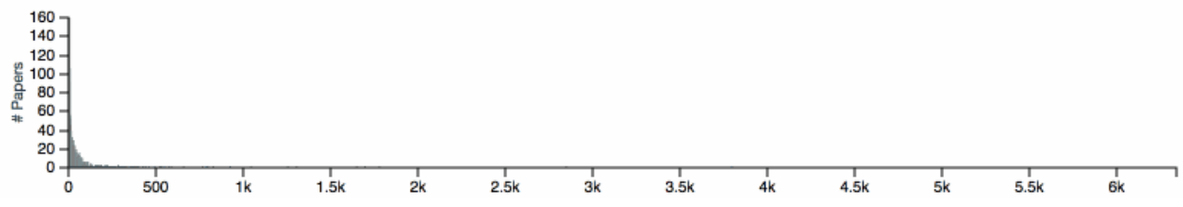
Citation Types



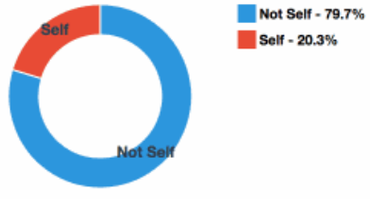
Citation Subject Area



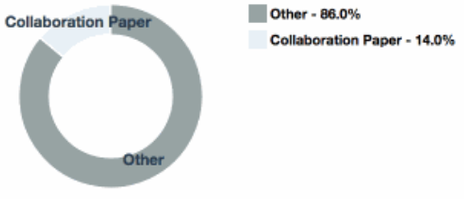
Paper Citation Counts



Self Citations

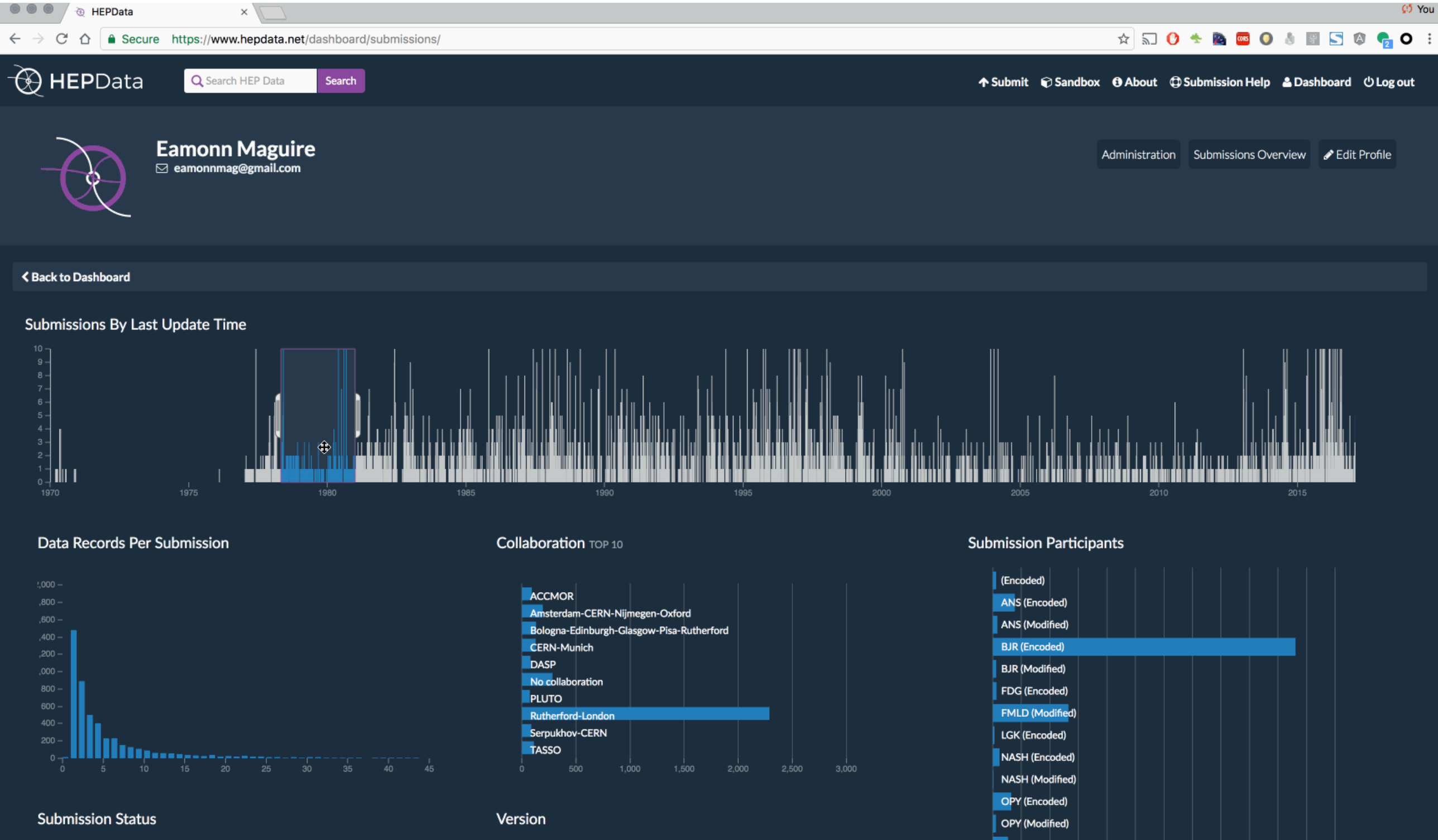


Collaboration Papers



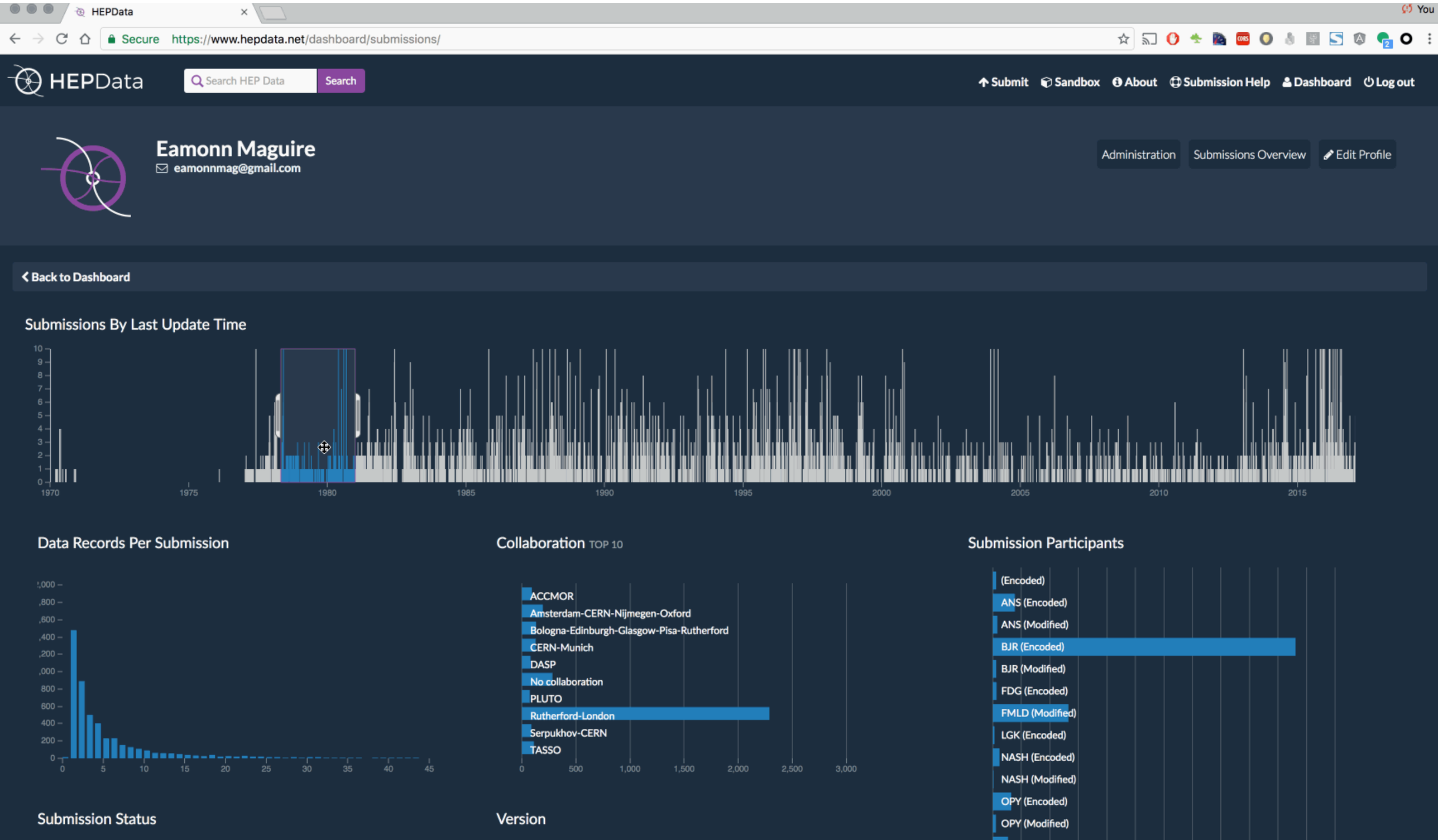
Multidimensional Visualization

With dc.js, crossfilter, and d3.js



Multidimensional Visualization

With dc.js, crossfilter, and d3.js



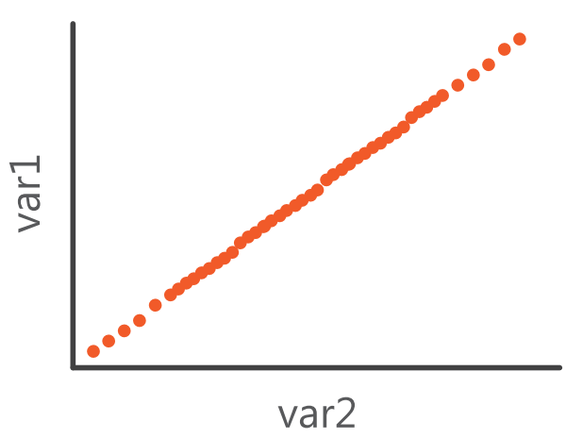
My Tutorial on Creating Dashboard Visualizations

https://thor-project.github.io/dashboard-tutorial/

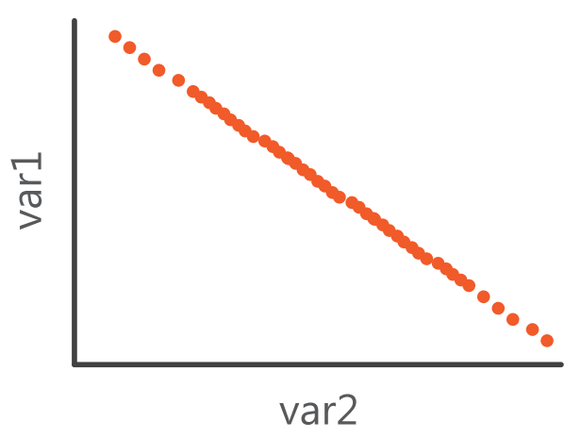
Multidimensional Visualization

Parallel Coordinate Plots

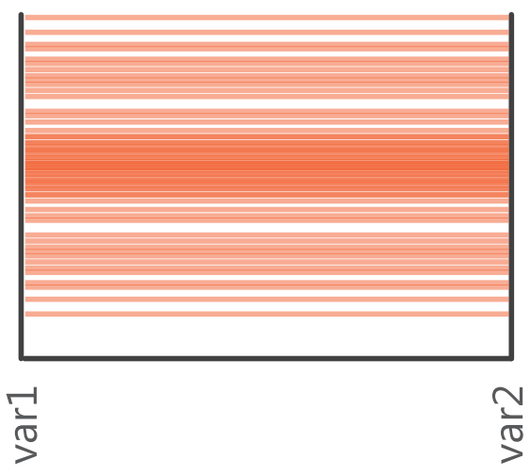
Positive Correlation

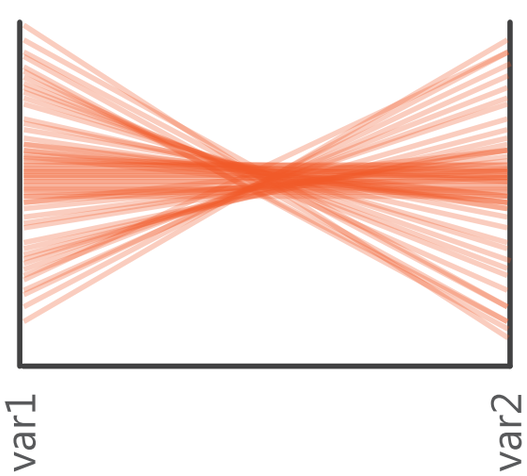


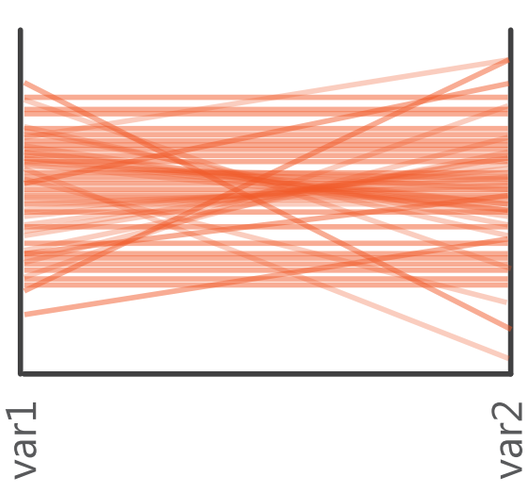
Negative (inverse) Correlation



No Correlation

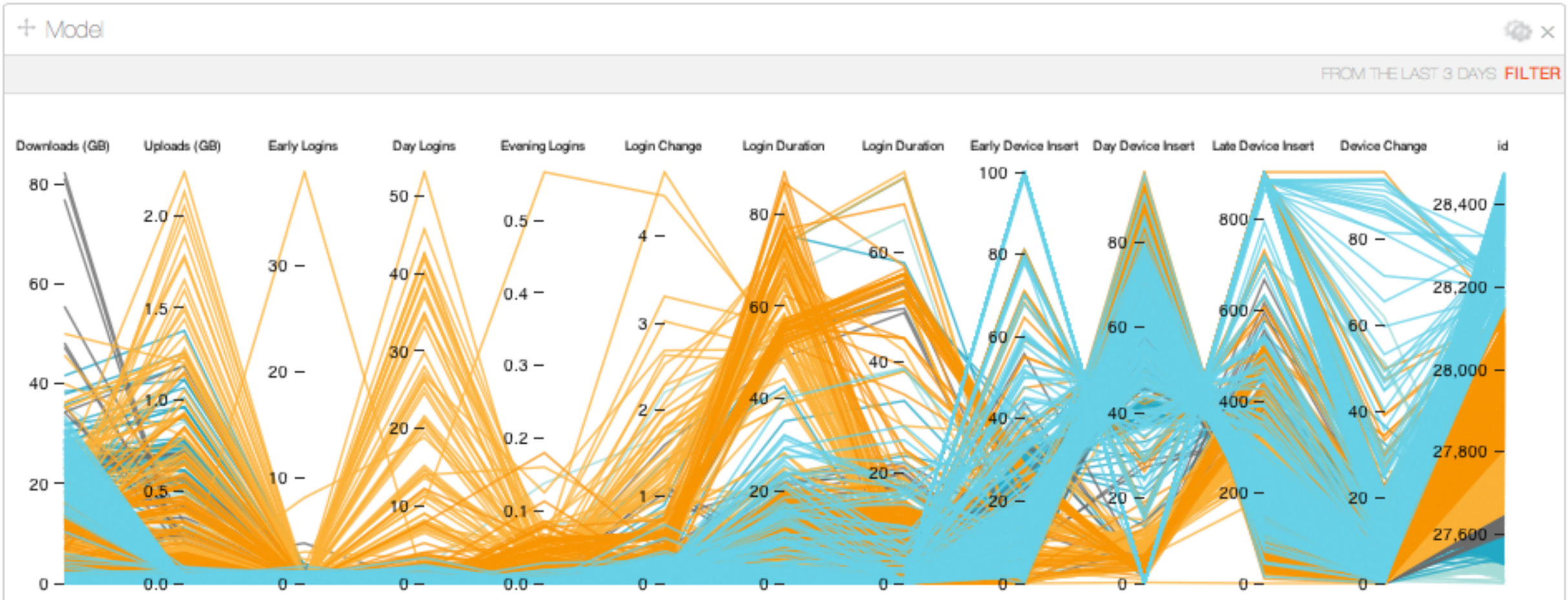






Multidimensional Visualization

Parallel Coordinate Plots

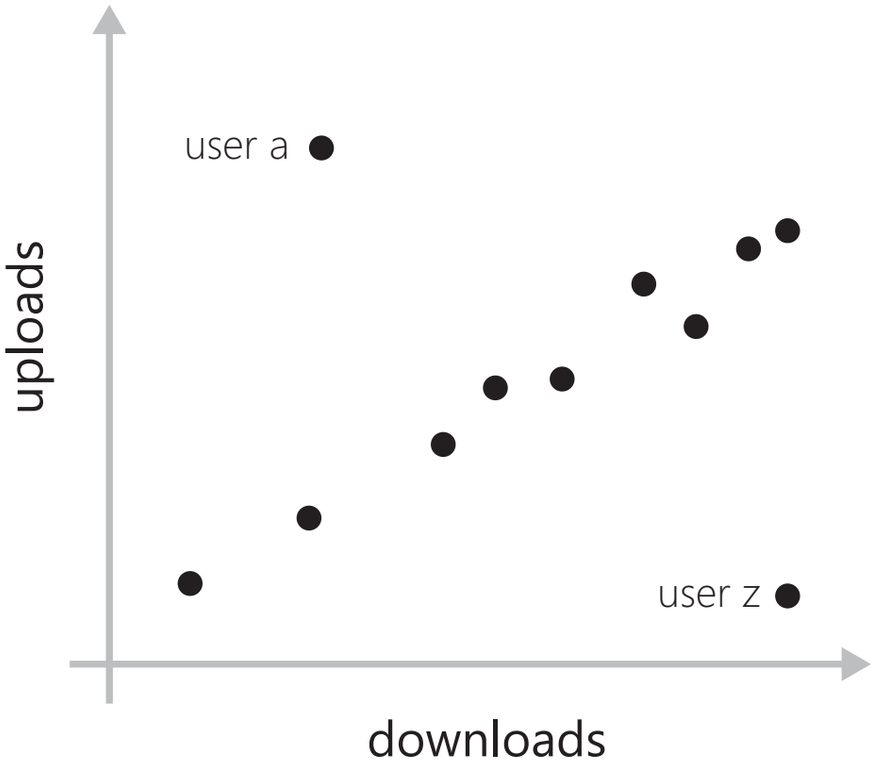


group	Downloads...	Uploads (...	Early Logi...	Day Logins	Evening L...	Login Cha...	Login Dura...	Login Dura...	Early Devi...	Day Devic...	Late Devic
Communicati...	19.9	0.285	0.385	0.4	0.0769	0.947	66.2	65.8	1.4	7.4	357
Communicati...	35.8	0.5	0.717	2	0.138	1.705	41	39	5.1	7.4	358
Communicati...	23.5	0.88	0.78	1.6	0.0012	2.2	52.9	51.3	2.6	12.8	329
Communicati...	0.28	0.004	0.002	0	0	0.005	0	0	99.48	0.24	876
Communicati...	0.85	0.024	0.714	0	0	0.024	0.06	0.06	81.11	15.87	717
Communicati...	0.85	0.024	0.827	0	0	0.026	0.06	0.06	81.11	15.87	717
Communicati...	0.85	0.024	0.011	0	0	0.024	0.06	0.06	81.11	17.94	717

Parallel Coordinate Plots

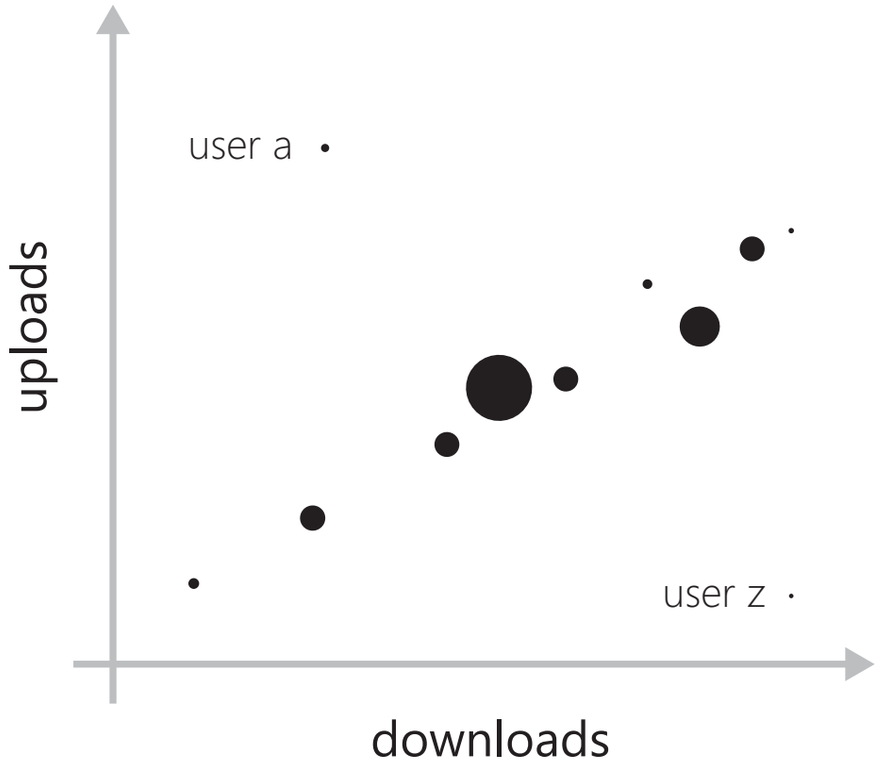
Lets take an example where we have many variables to display...
Each user is represented by a circle

2 Dimensions



3 Dimensions

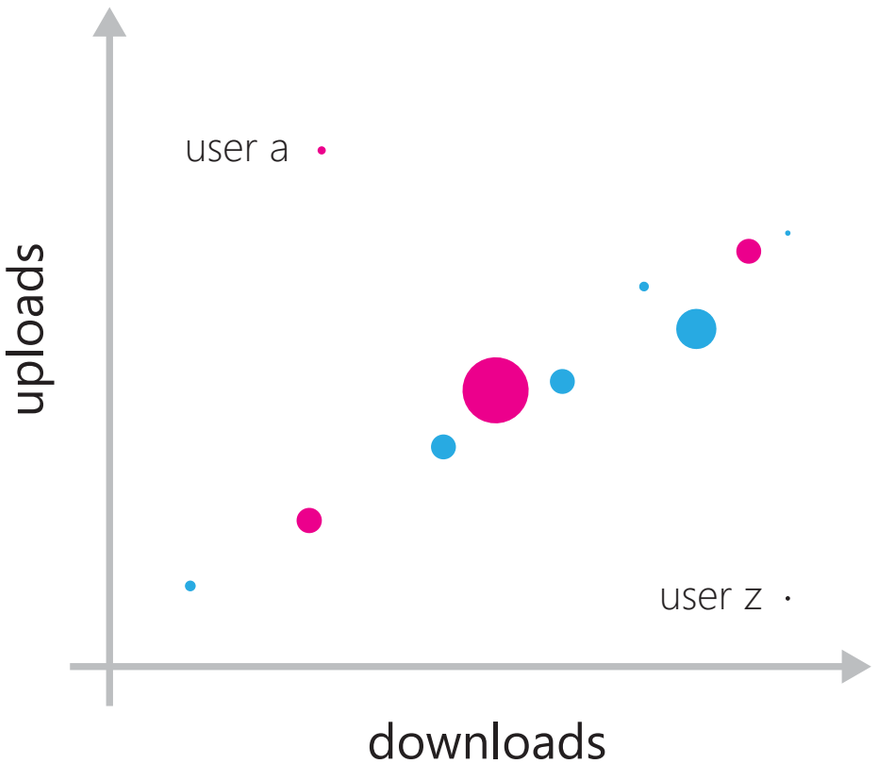
Size indicates number of logins per day



Parallel Coordinate Plots

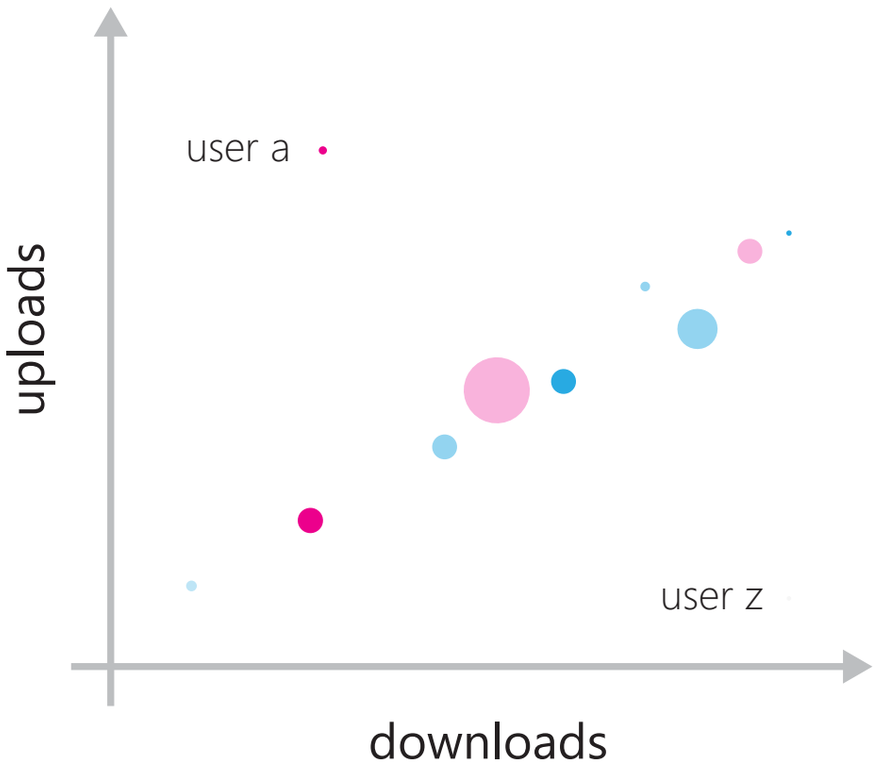
4 Dimensions

Color indicates users department

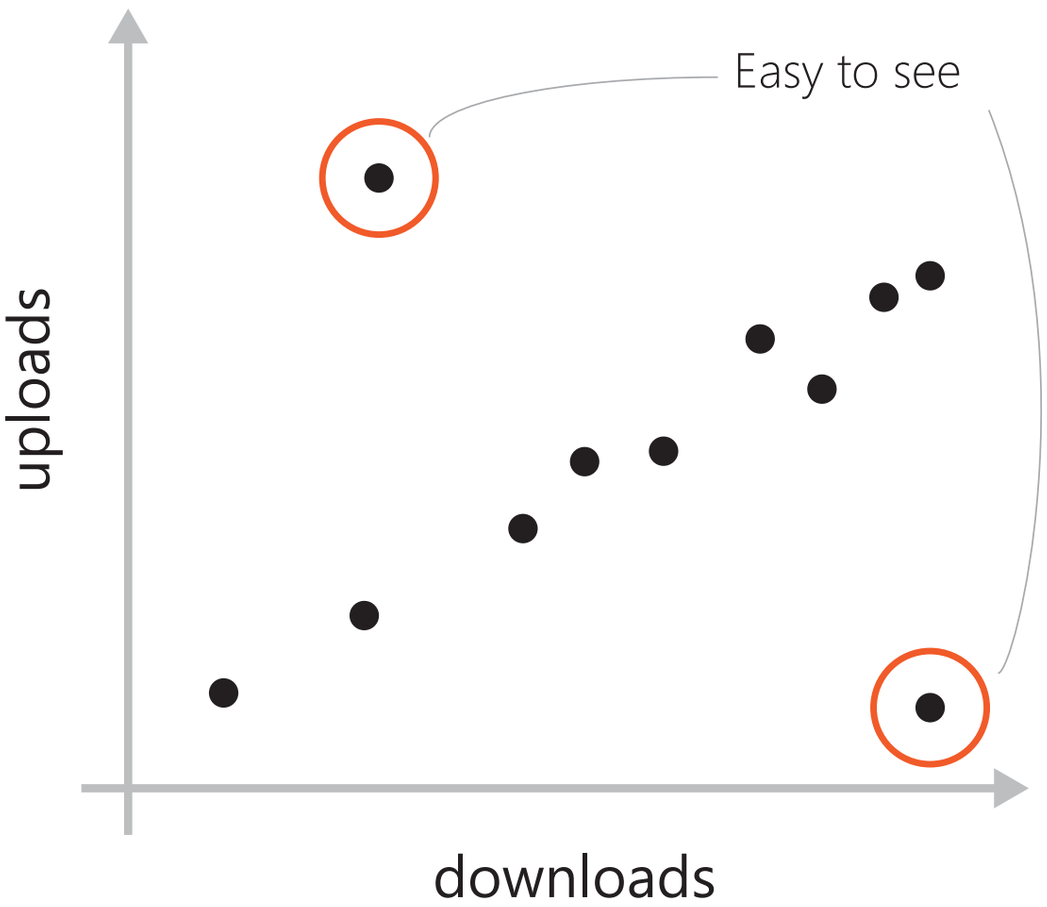


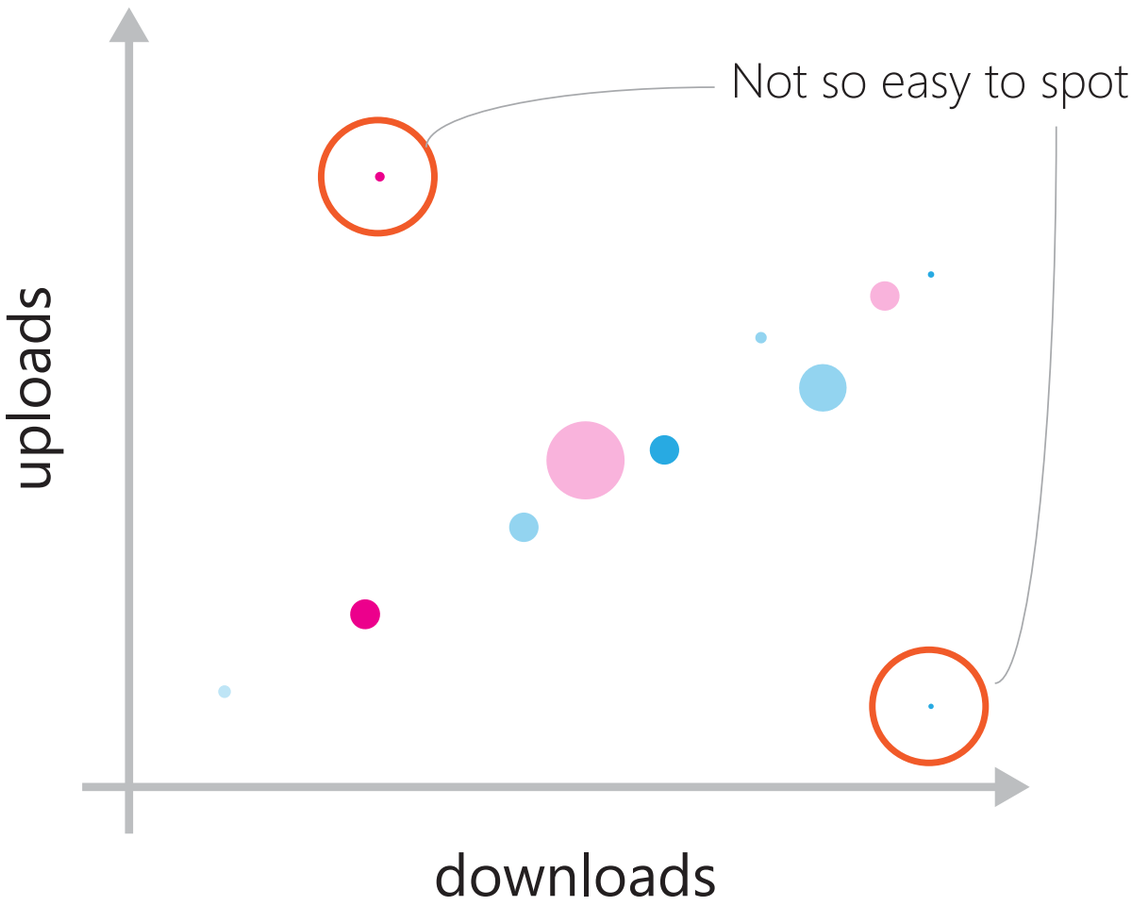
5 Dimensions

Transparency indicates consistency in logins



As we get to higher levels of dimensions, we'll have problems. Our choice of visual encoding will affect the visual availability of each dimension to the user.



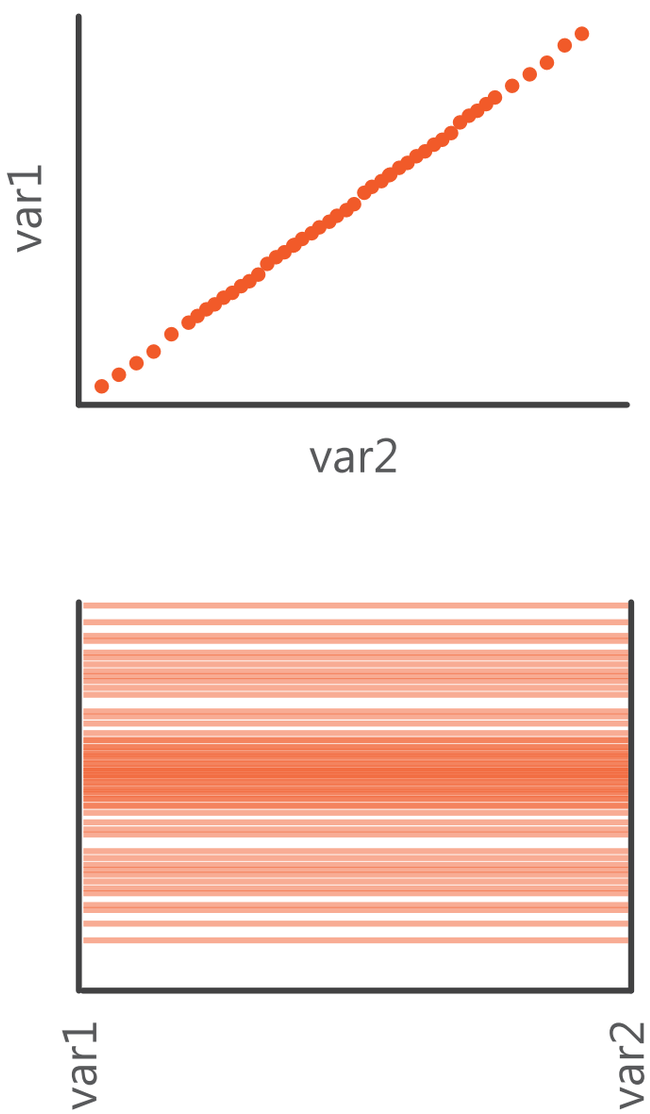


Parallel coordinates are a visualization technique employed when a large number of dimensions need to be displayed (often without a temporal element) and where each of those dimensions can be equally important in the decision making process.

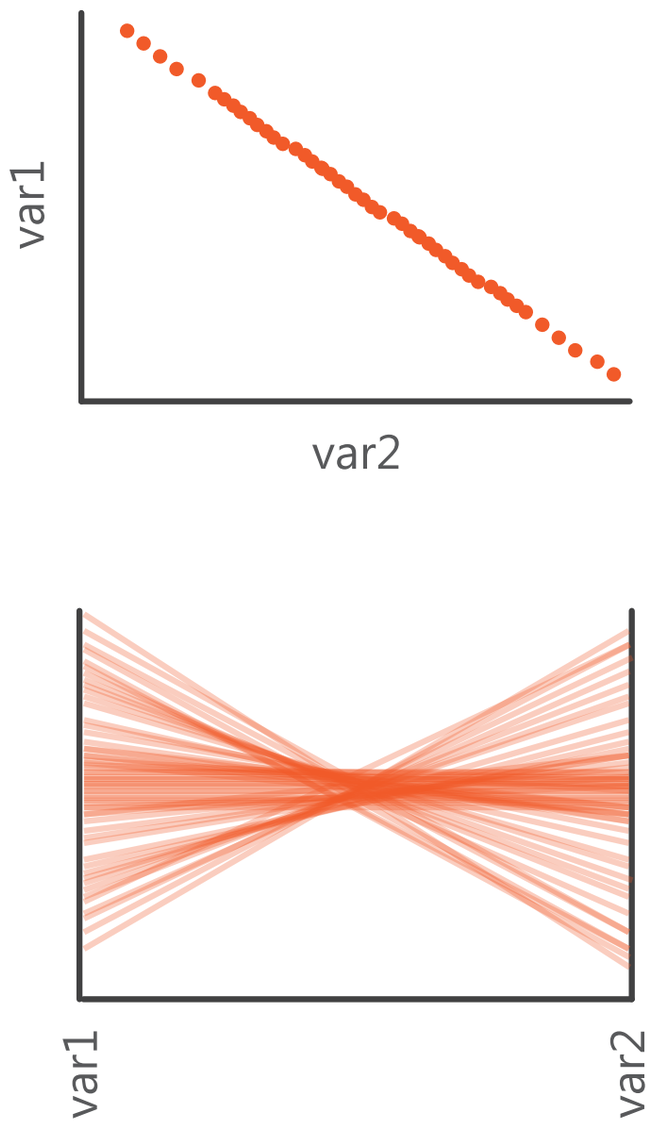
In the scatter plots here, it's easy to see **correlation** between downloads and uploads, but with the other dimensions that's difficult.

Parallel Coordinate Plots

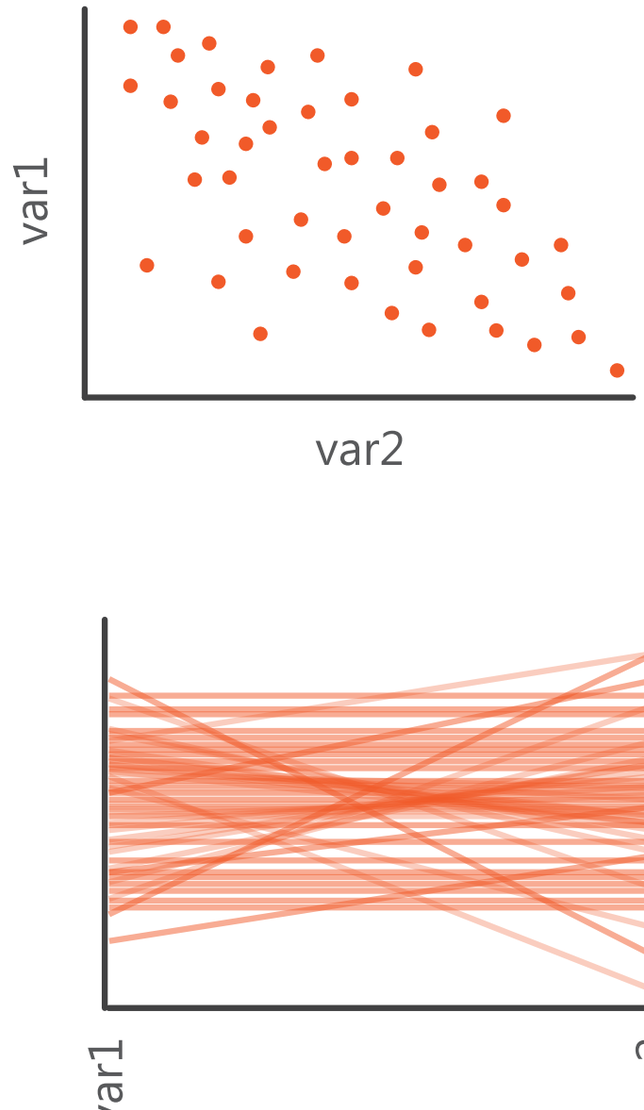
Positive Correlation



Negative (inverse) Correlation

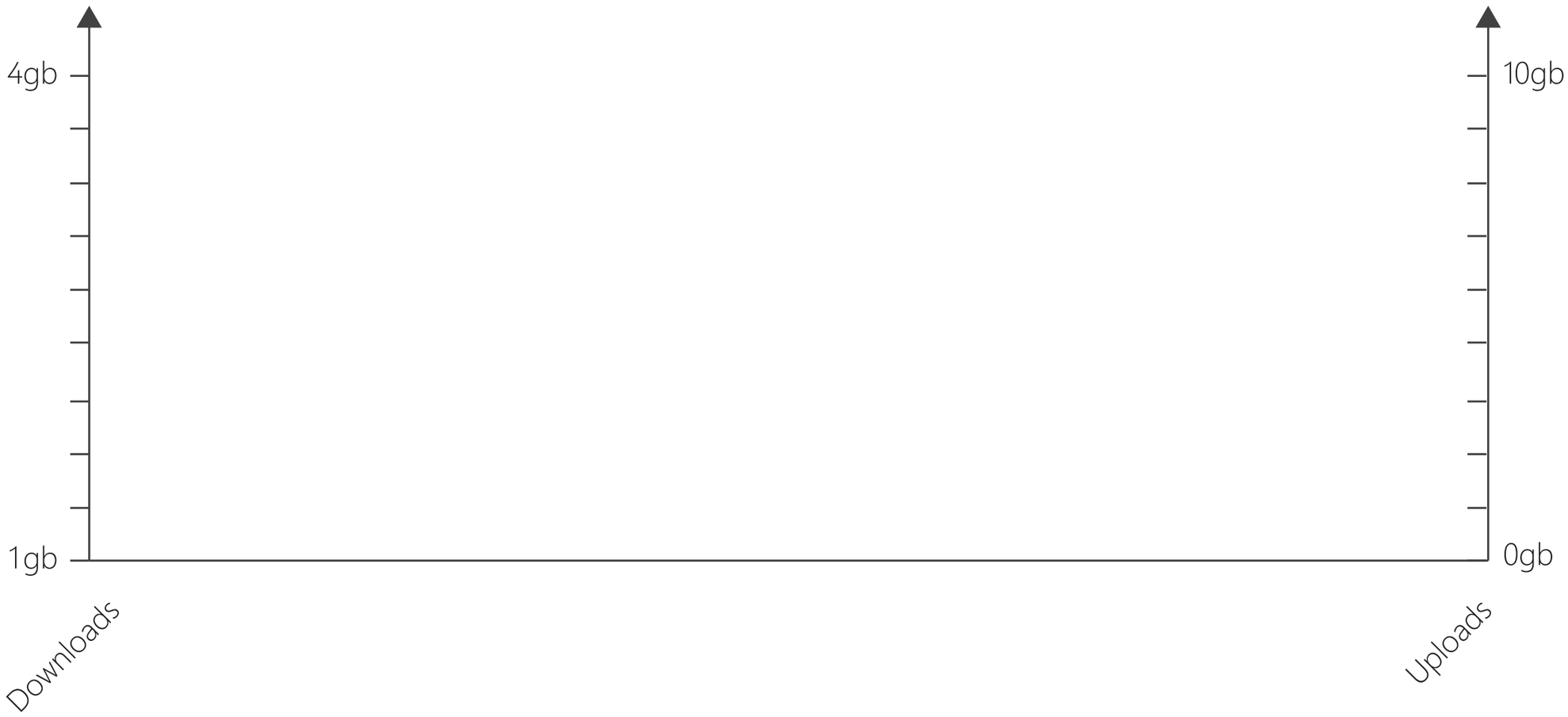


No Correlation



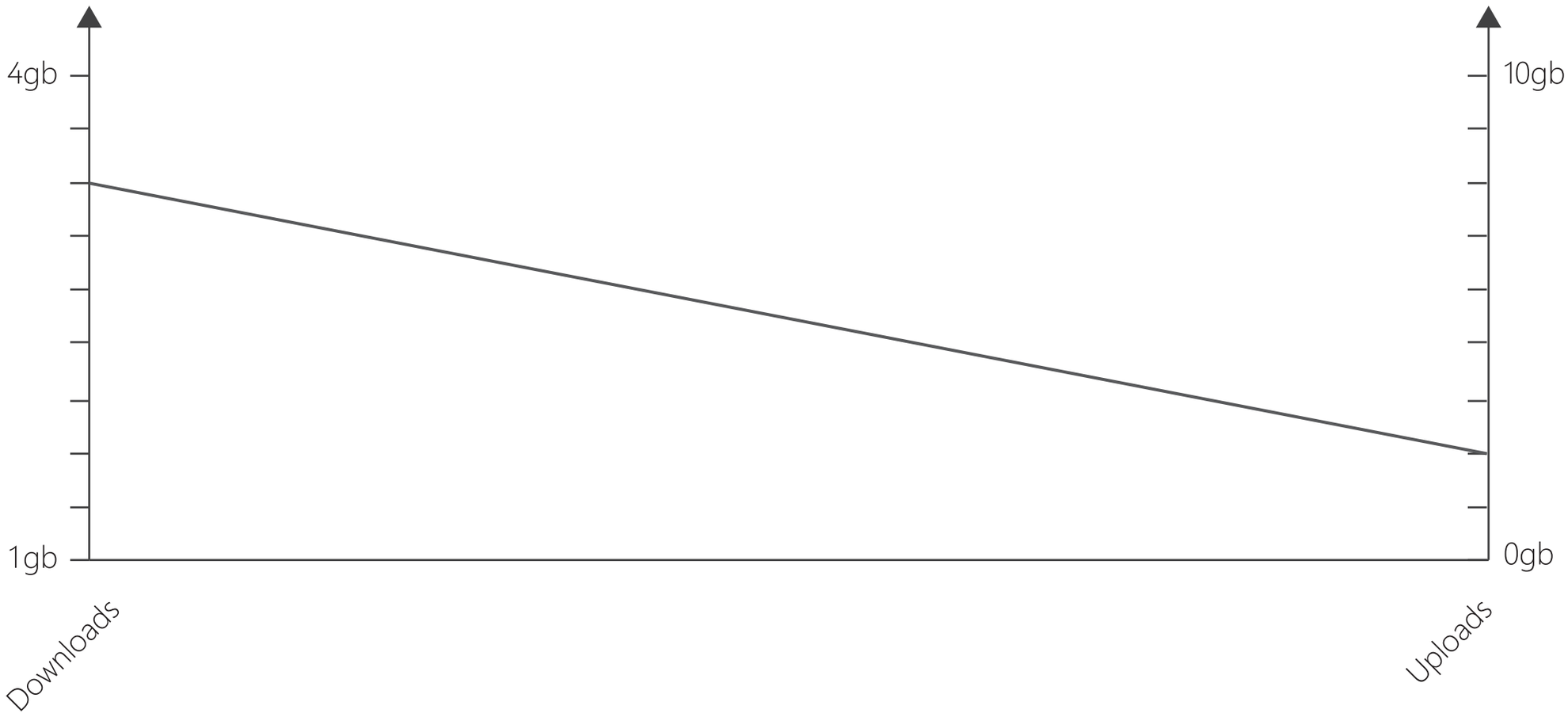
Parallel Coordinate Plots

2 Dimensions
Uploads
Downloads



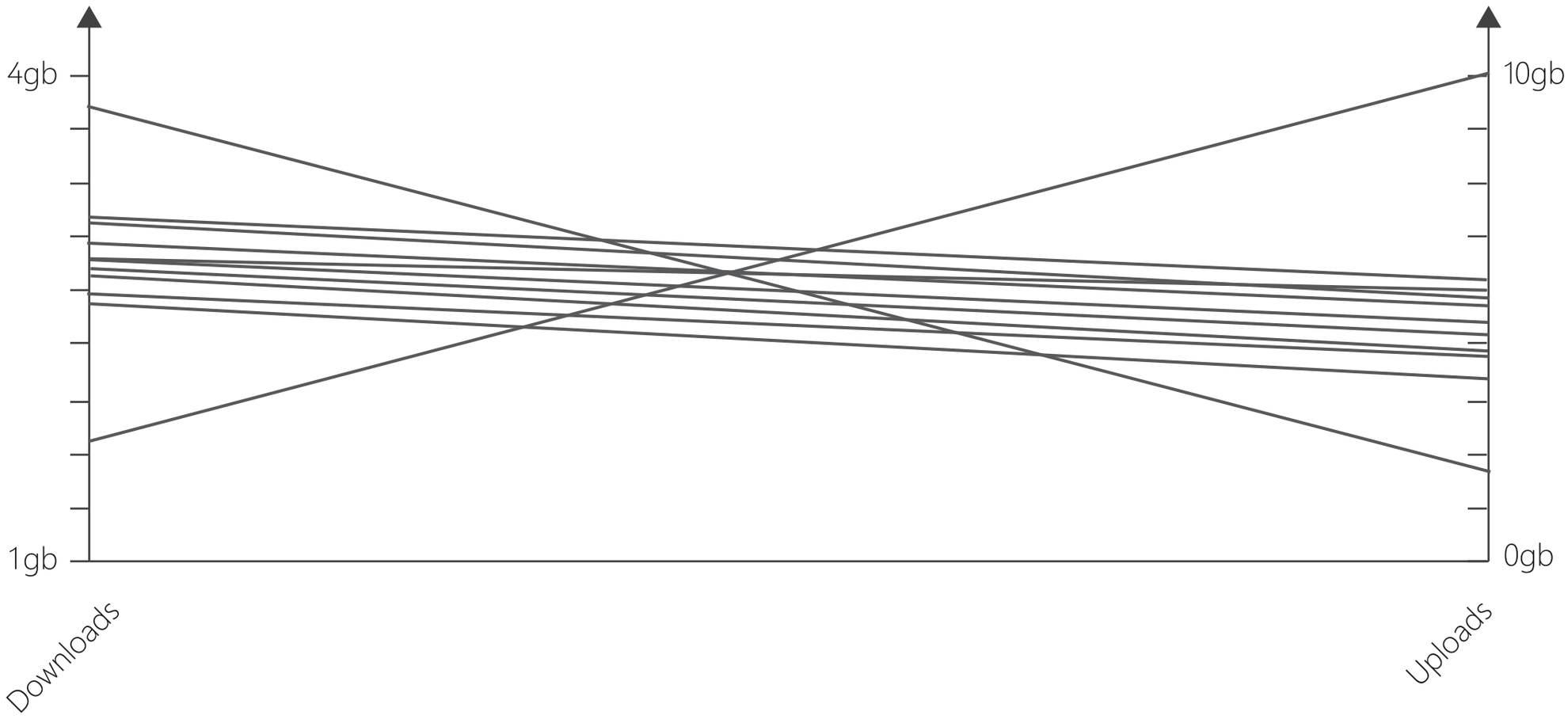
Parallel Coordinate Plots

2 Dimensions
Uploads 1 User
Downloads



Parallel Coordinate Plots

2 Variables
Uploads 11 Users
Downloads



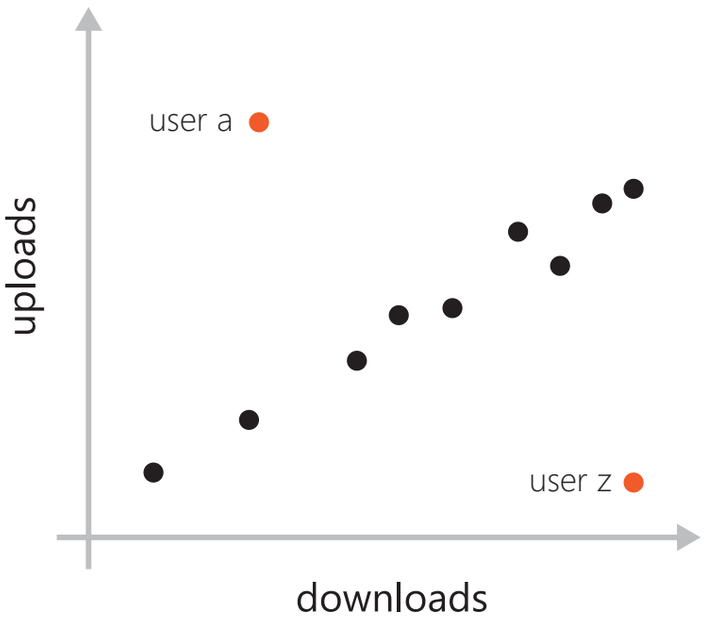
Parallel Coordinate Plots

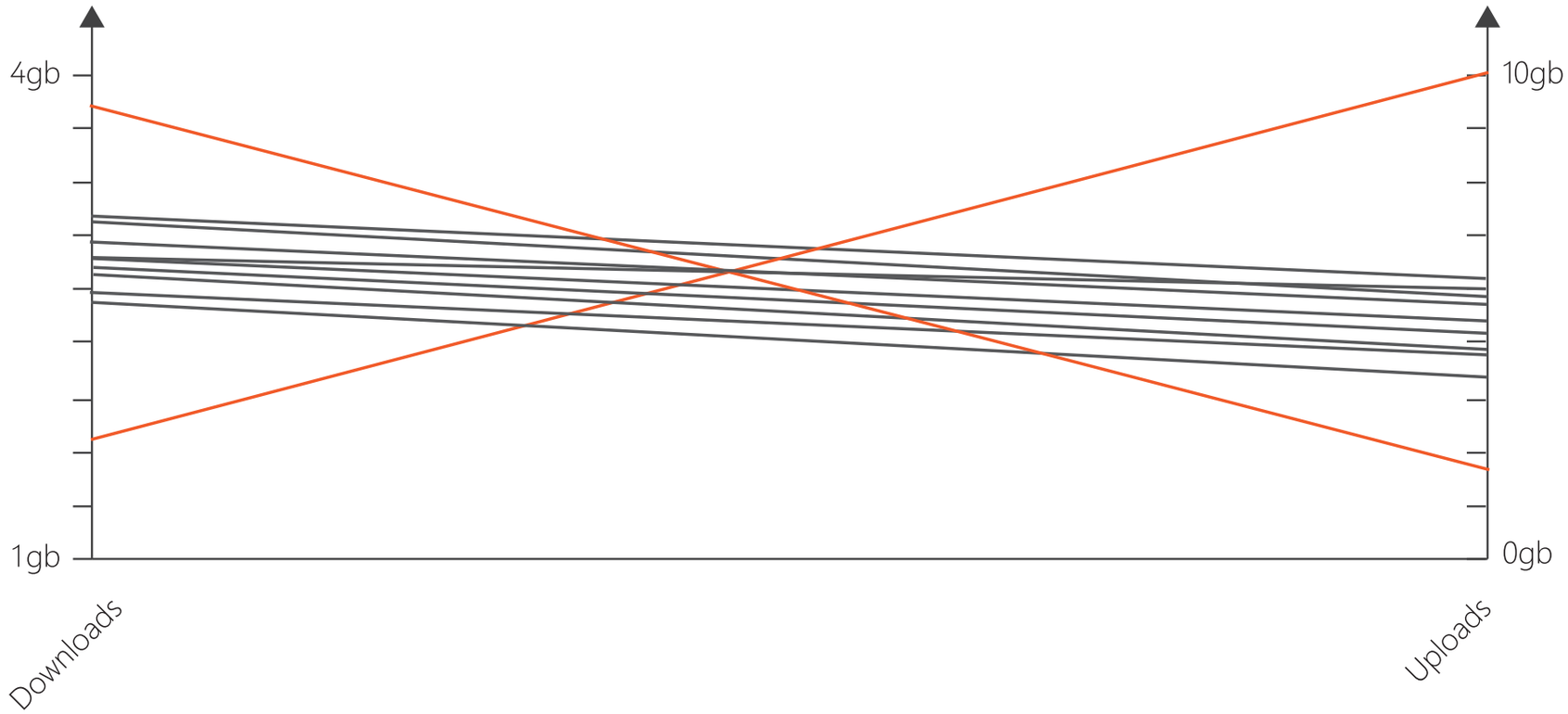
2 Dimensions

Uploads

Downloads

11 Users





Parallel Coordinate Plots

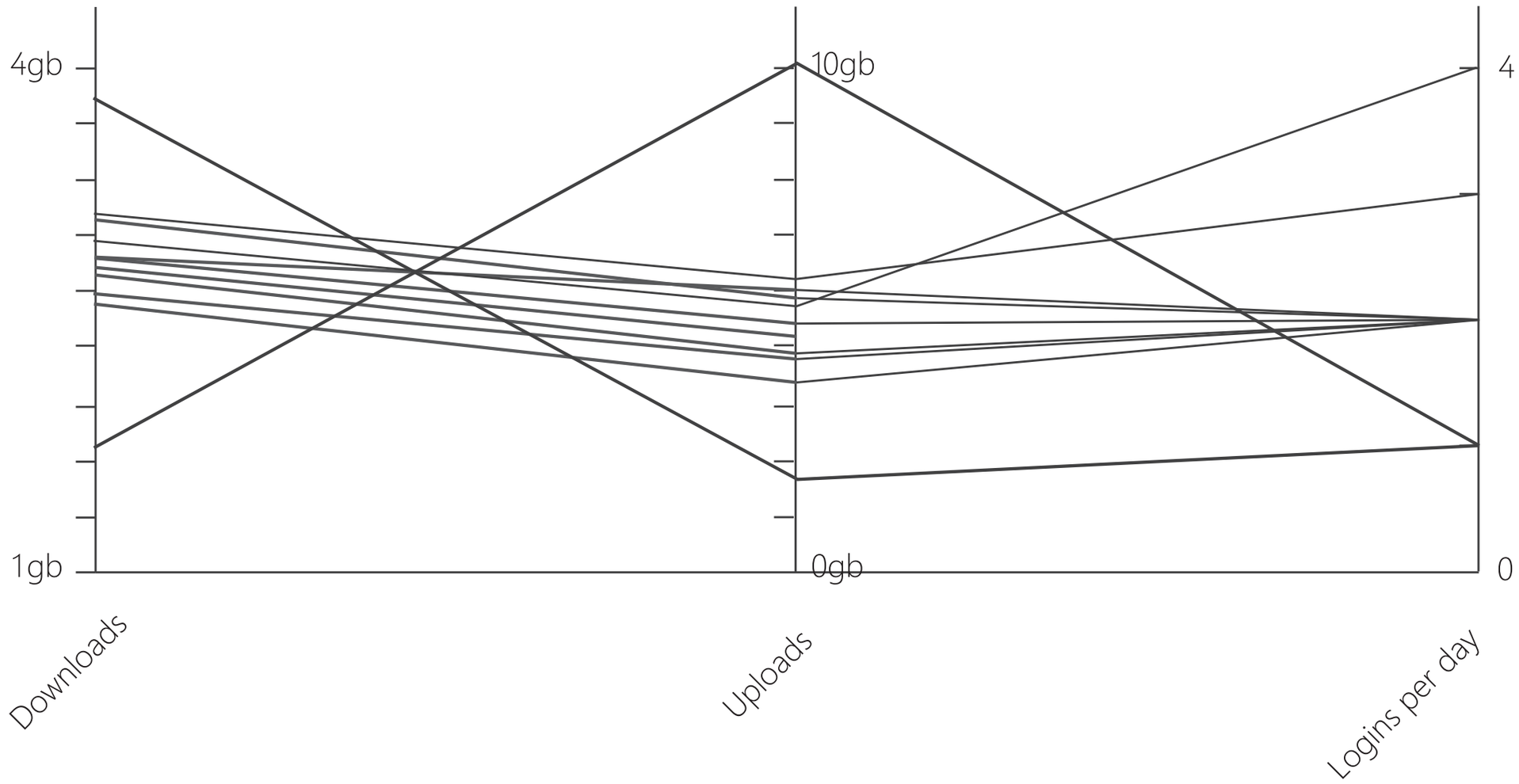
3 Dimensions

Uploads

Downloads

Logins per day

11 Users



Parallel Coordinate Plots

4 Dimensions

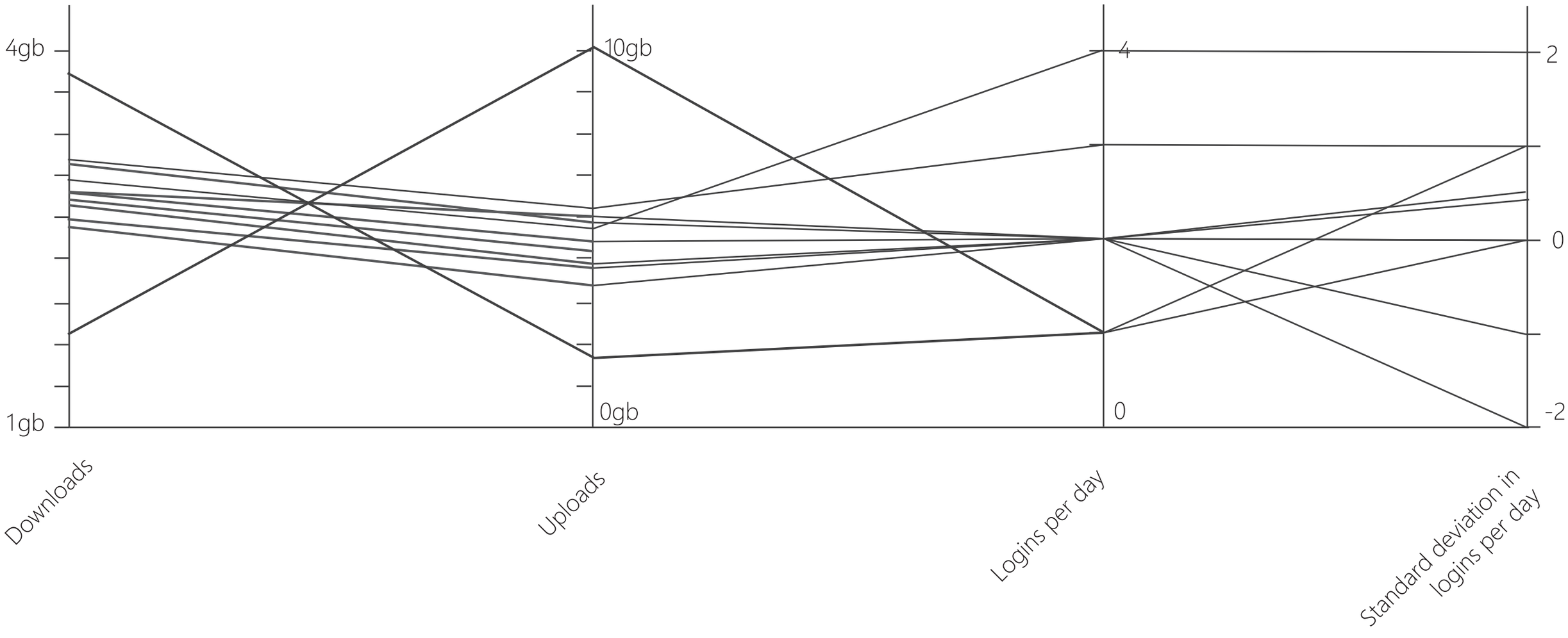
Uploads

Downloads

Logins per day

Std. deviation in logins per day

11 Users

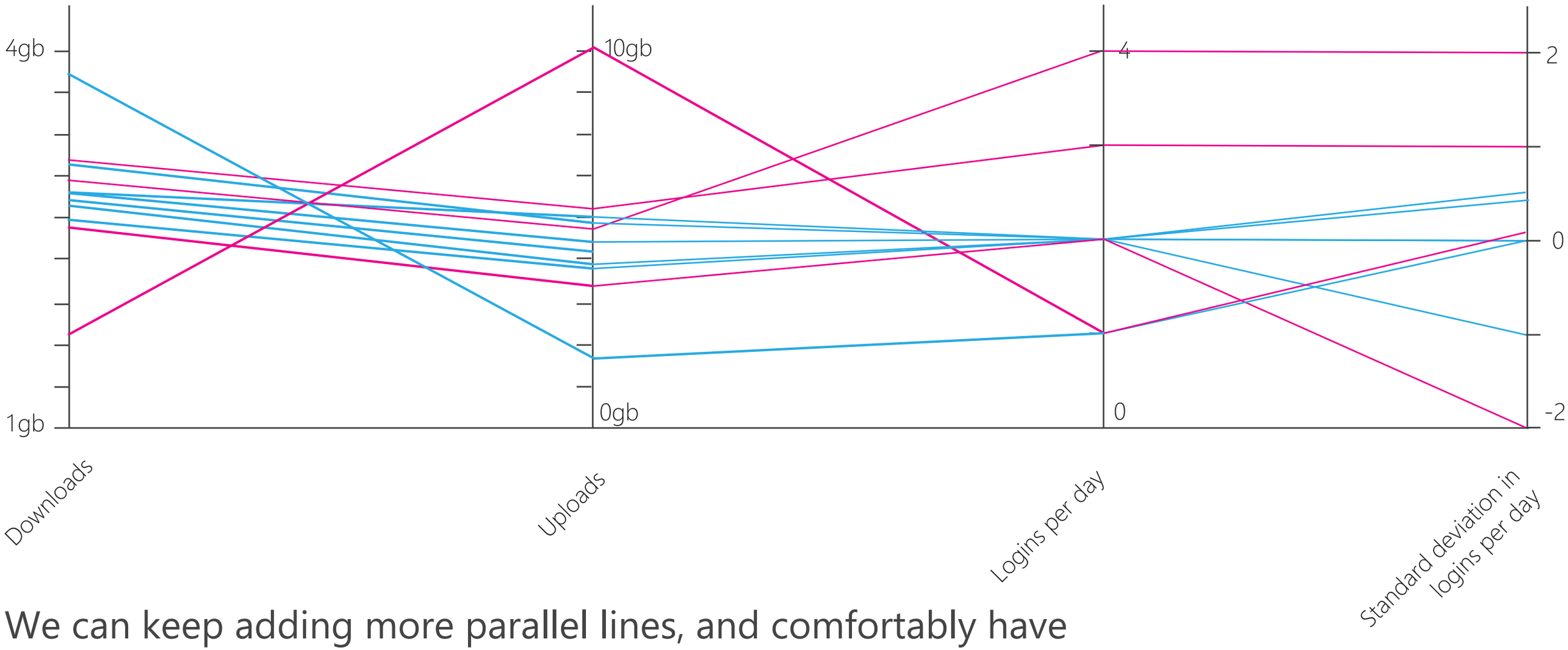


Parallel Coordinate Plots

5 Dimensions
Uploads
Downloads
Logins per day
Std. deviation in logins per day
Department

11 Users

We use color for department since it's categorical information.

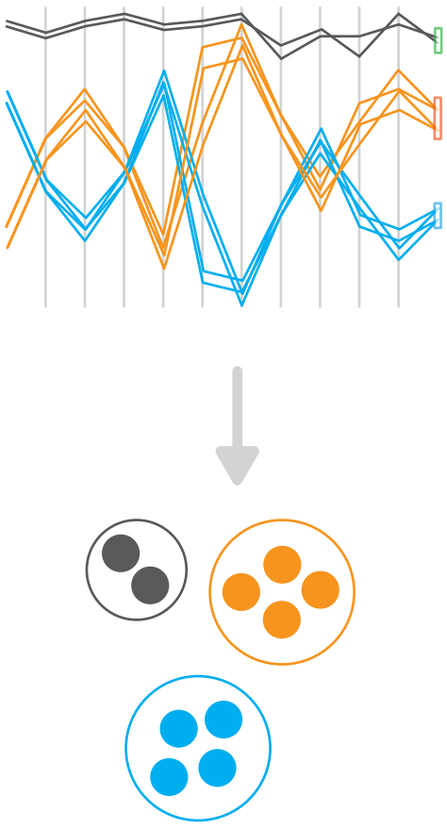


We can keep adding more parallel lines, and comfortably have around 20 dimensions for many users displayed at once.

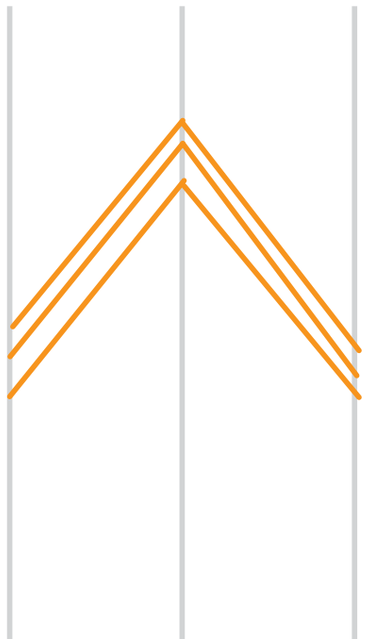
Parallel Coordinate Plots

Parallel coordinates provide an efficient way to visualize many variables, along with their associated **clusters**, **anomalies**, value **distributions** and **correlations.**

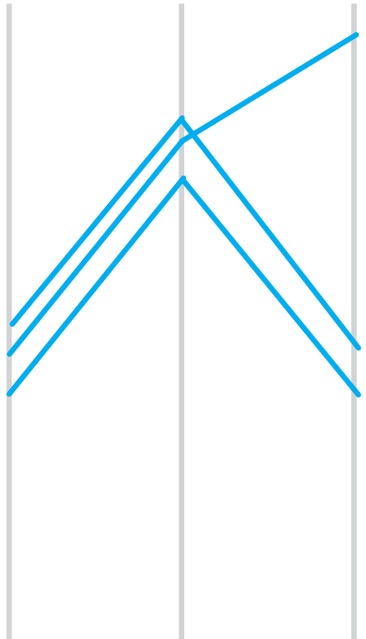
Clustering



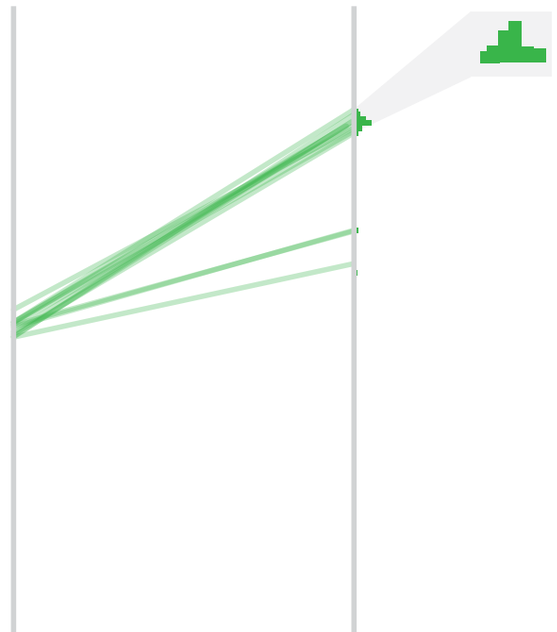
Correlation



Outlier Detection



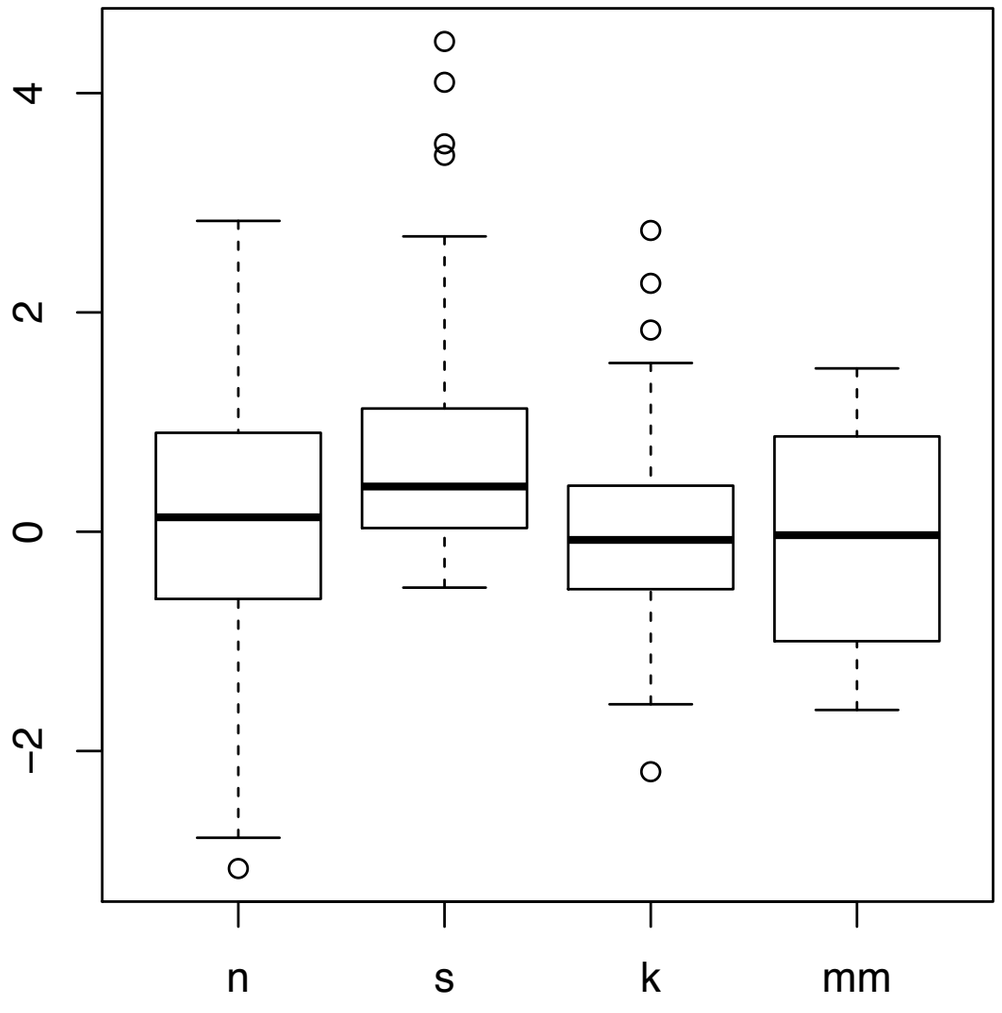
Distribution



Multidimensional Visualization

Glyphs

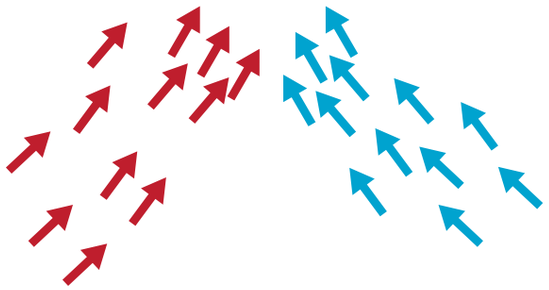
- static item aggregation
- **task**: find distribution
- **data**: table
- derived data
  - 4 quantitative attributes
    - median: central line
    - lower and upper quartile: boxes
    - lower upper fences: whiskers
  - outliers beyond fence cutoffs explicitly shown



Multidimensional Visualization

Glyphs

Simple Glyph



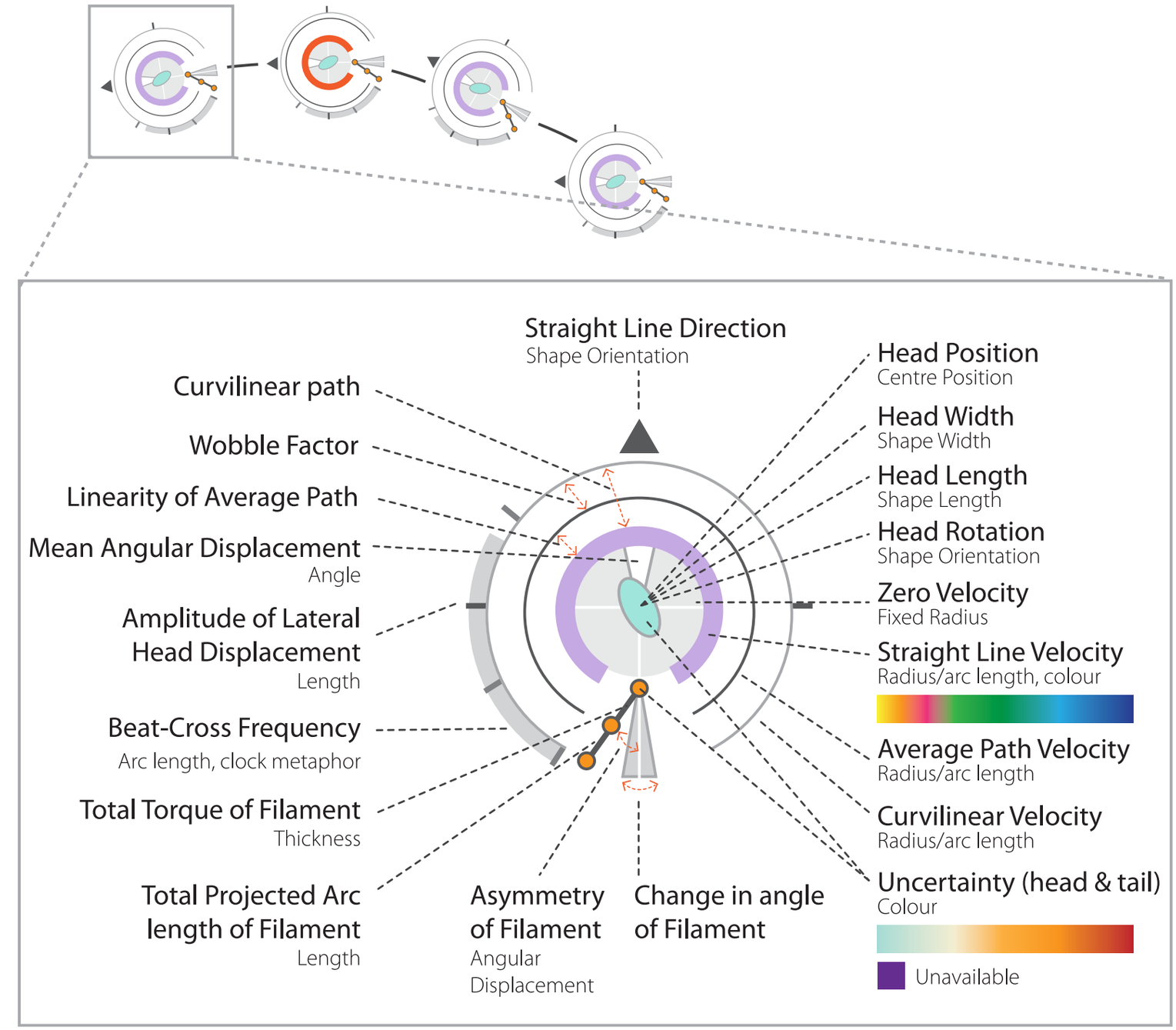
Temperature - Colour

Wind direction - Orientation

Wind Speed - Proximity

Location - Position

Complex Glyph



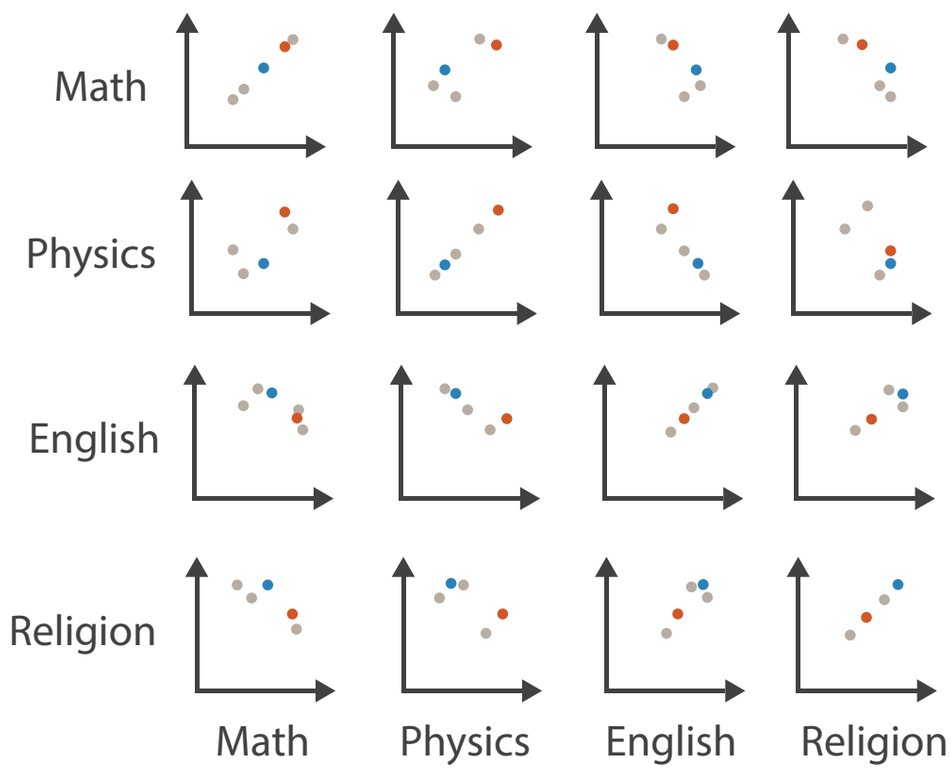
Multidimensional Visualization

A Simple Example | Student Test Results

Table

Math	Physics	English	Religion
85	95	71	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

Scatter Plot Matrix



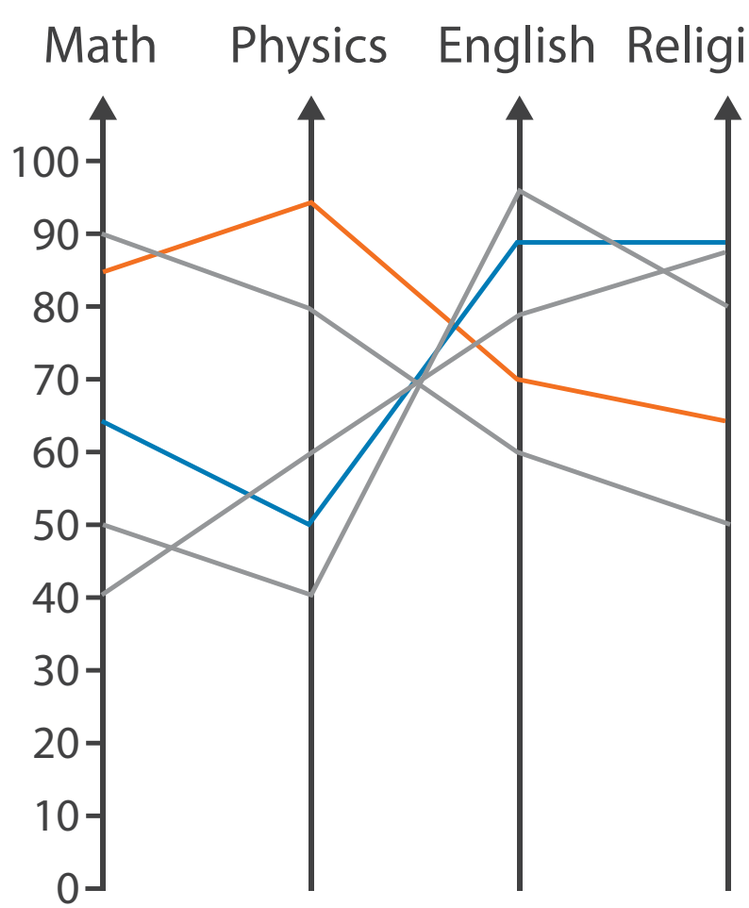
Multidimensional Visualization

A Simple Example | Student Test Results

Table

Math	Physics	English	Religion
85	95	71	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

Parallel Coordinates



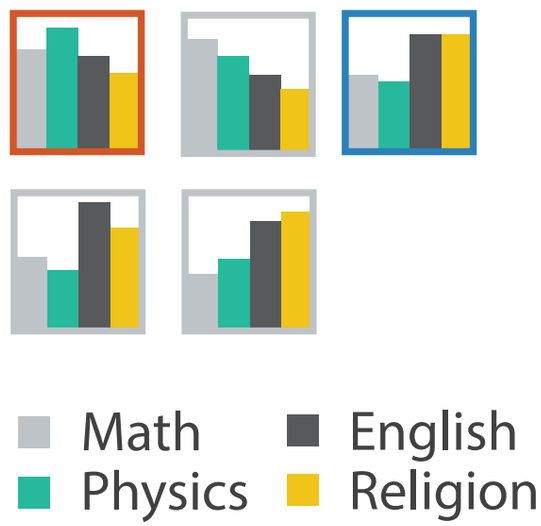
Multidimensional Visualization

A Simple Example | Student Test Results

Table

Math	Physics	English	Religion
85	95	71	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

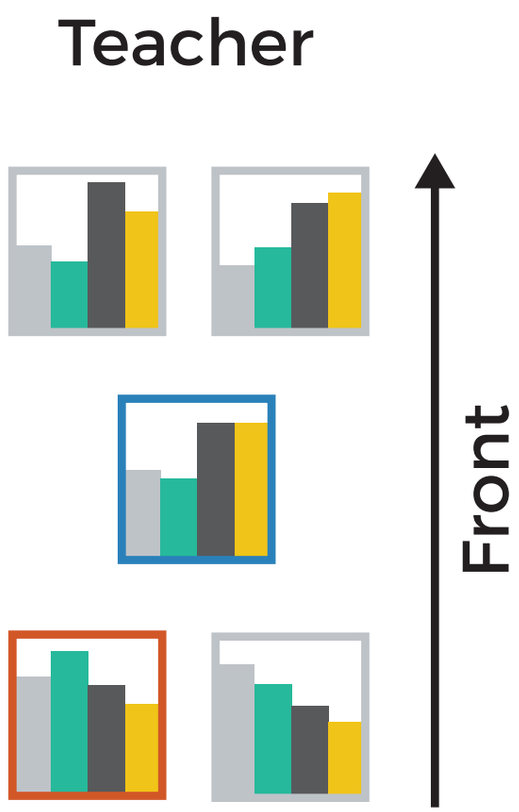
Glyph



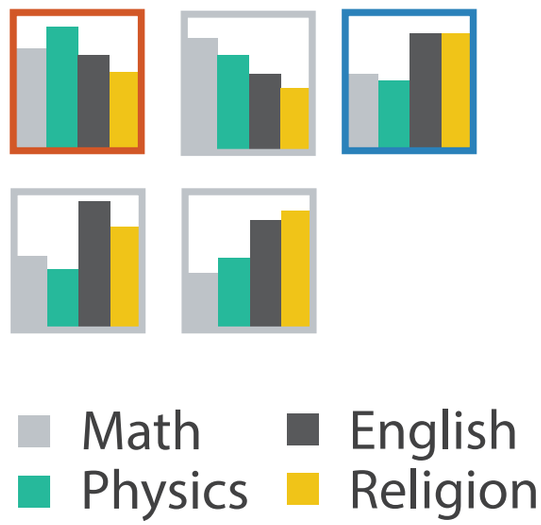
Arrange Spatially

Visualization

dent Test Results

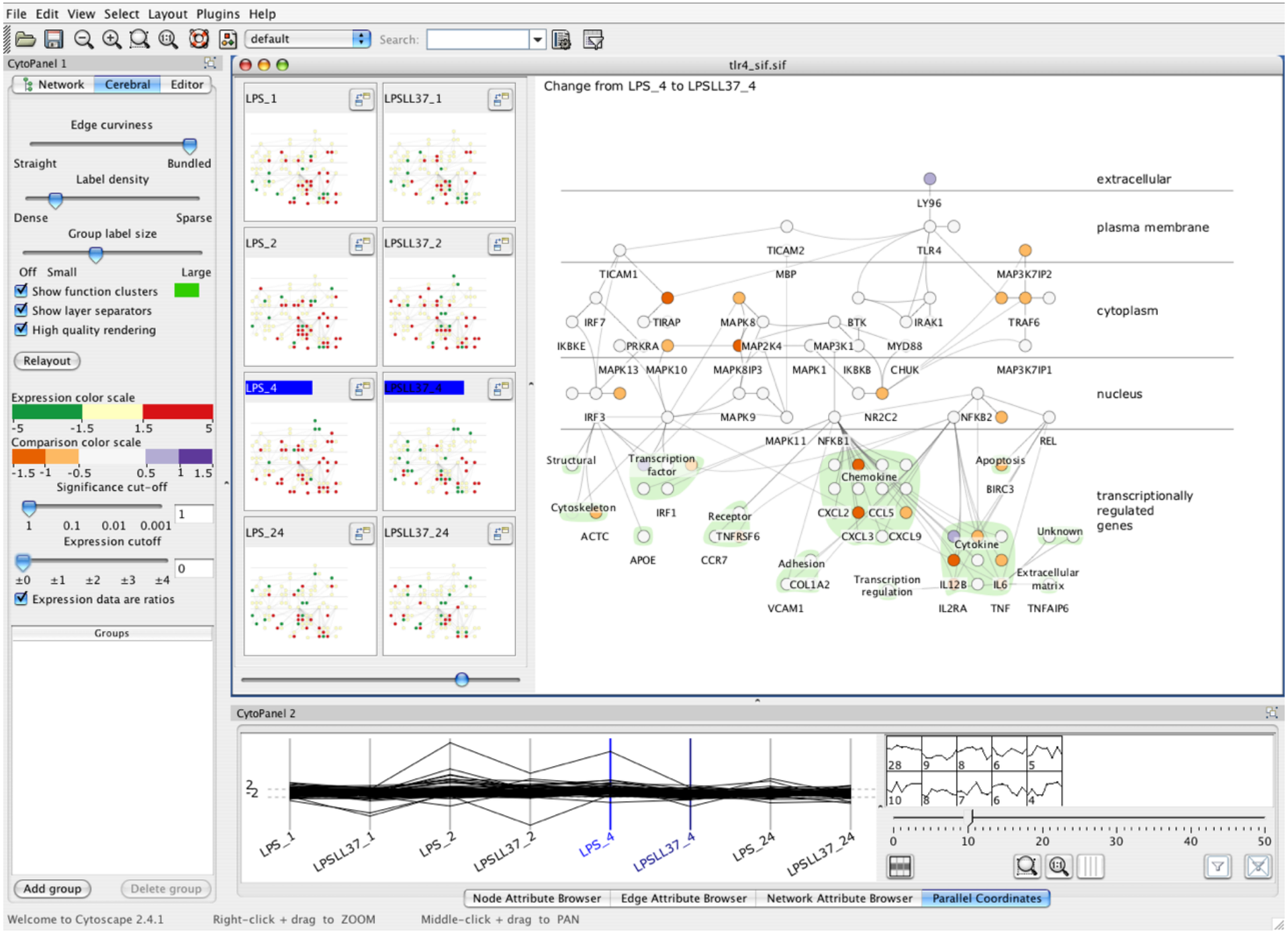


Glyph

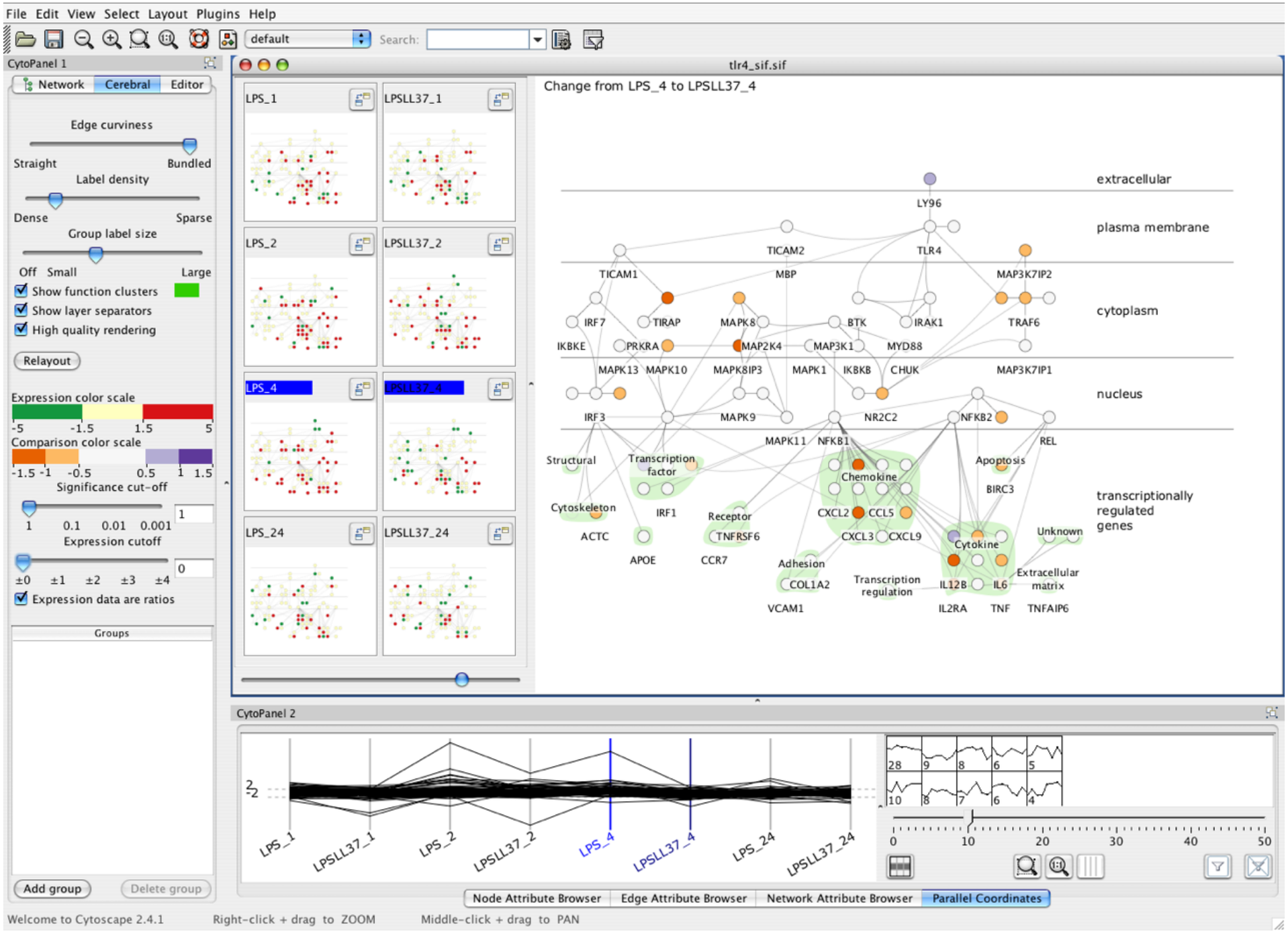


What about topological data?

Graphs/Networks

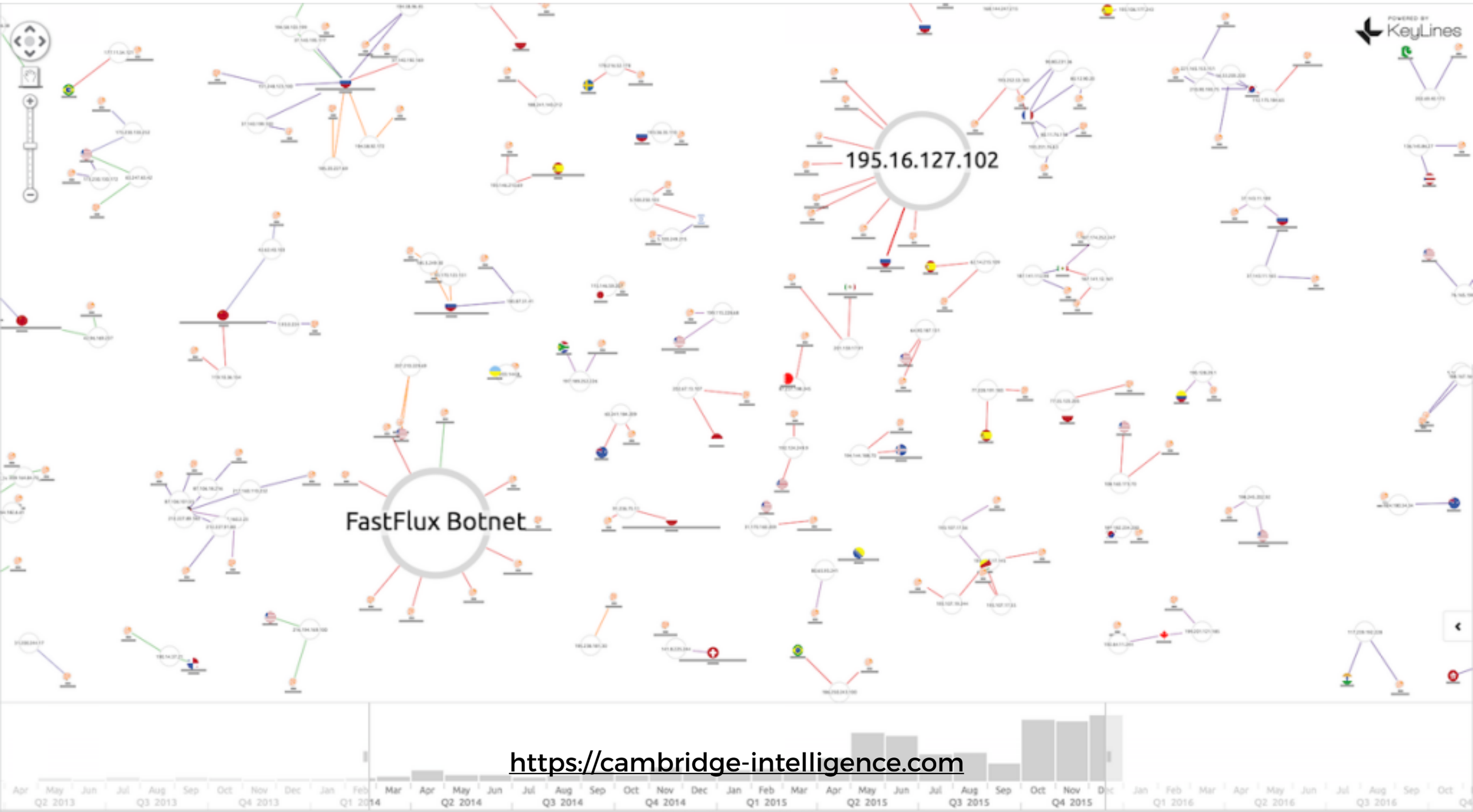


Graphs/Networks



Cerebral:Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE TVCG (Proc. InfoVis) 14(6):1253-1260, 2008.]

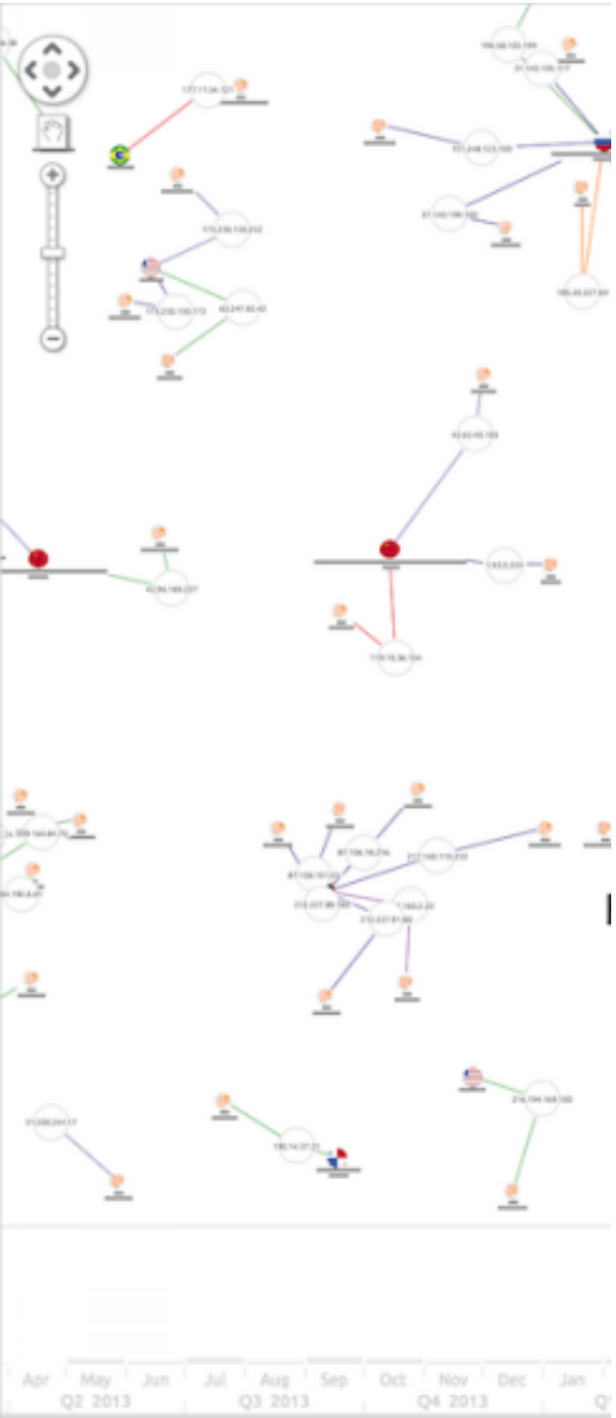
Graphs/Networks

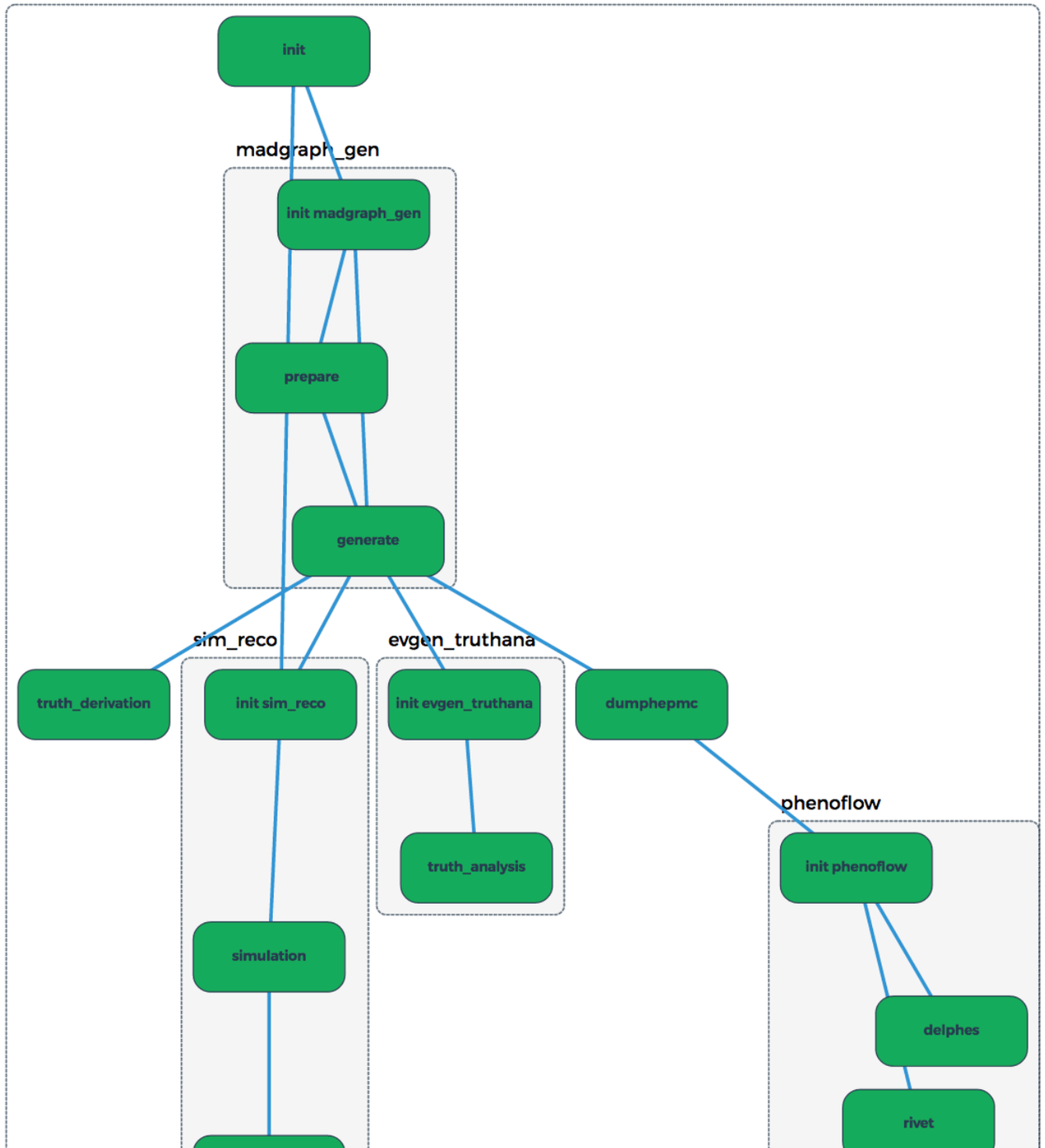


https://cambridge-intelligence.com

Graphs/Networks

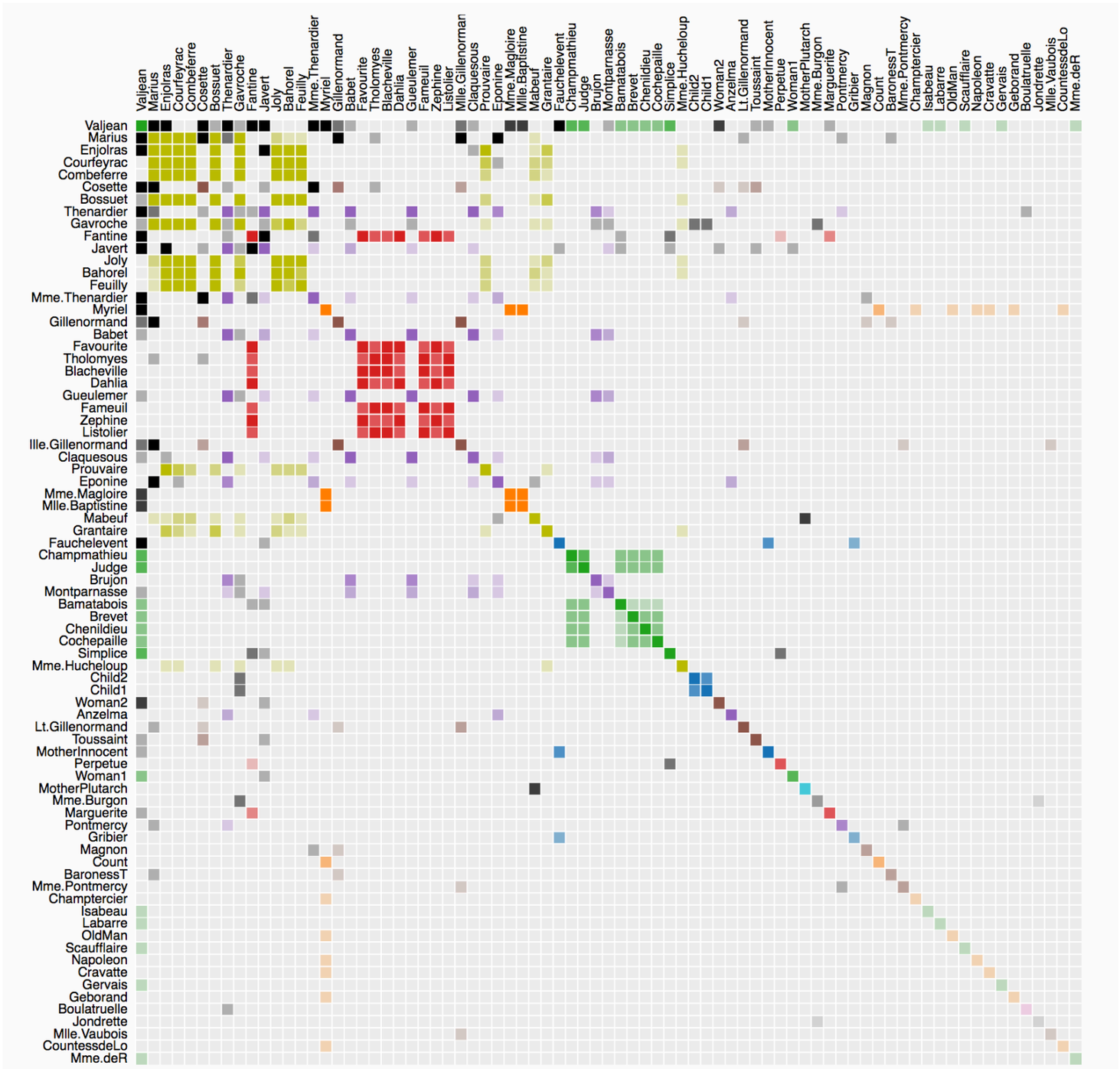
CERN
Analysis Preservation





Graphs/Networks

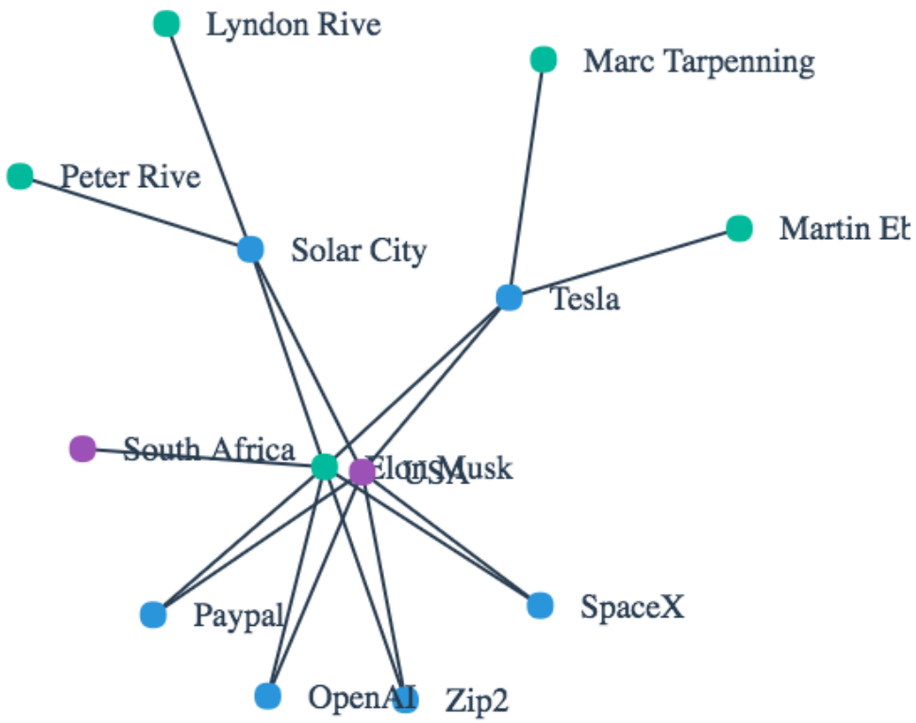
Matrix Representations



https://bost.ocks.org/mike/miserables/

Graphs/Networks

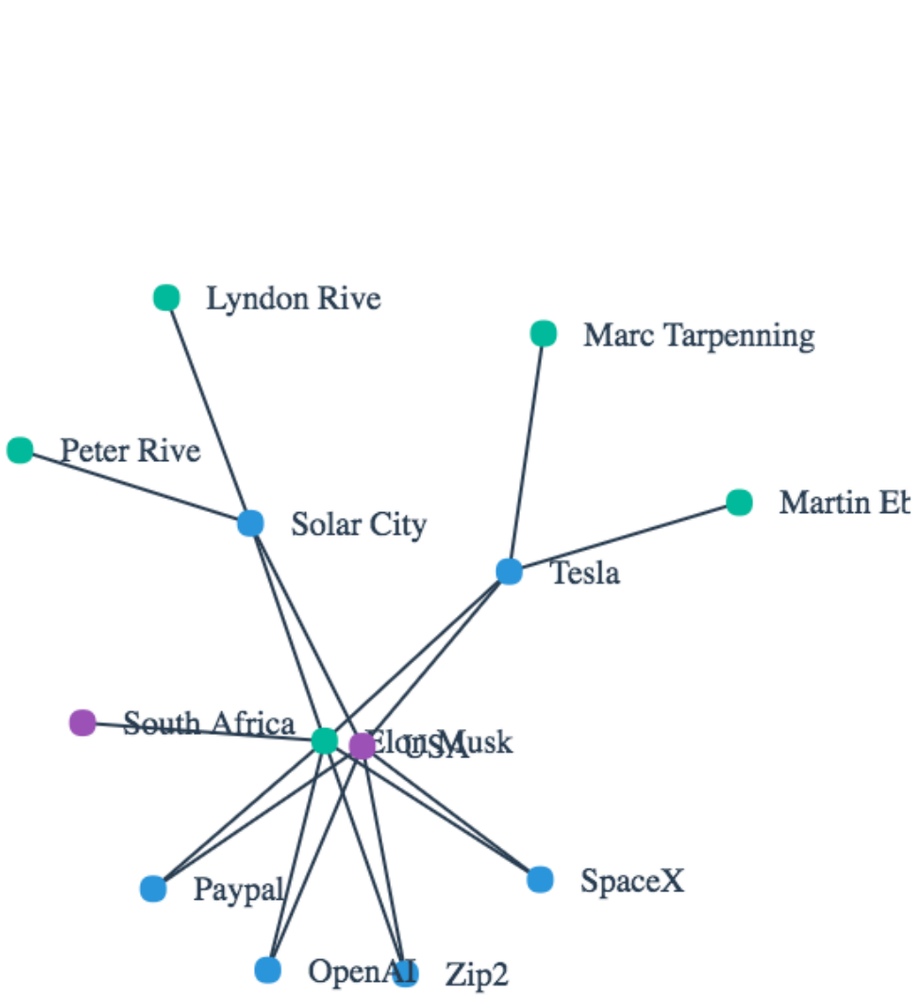
Hive Plots

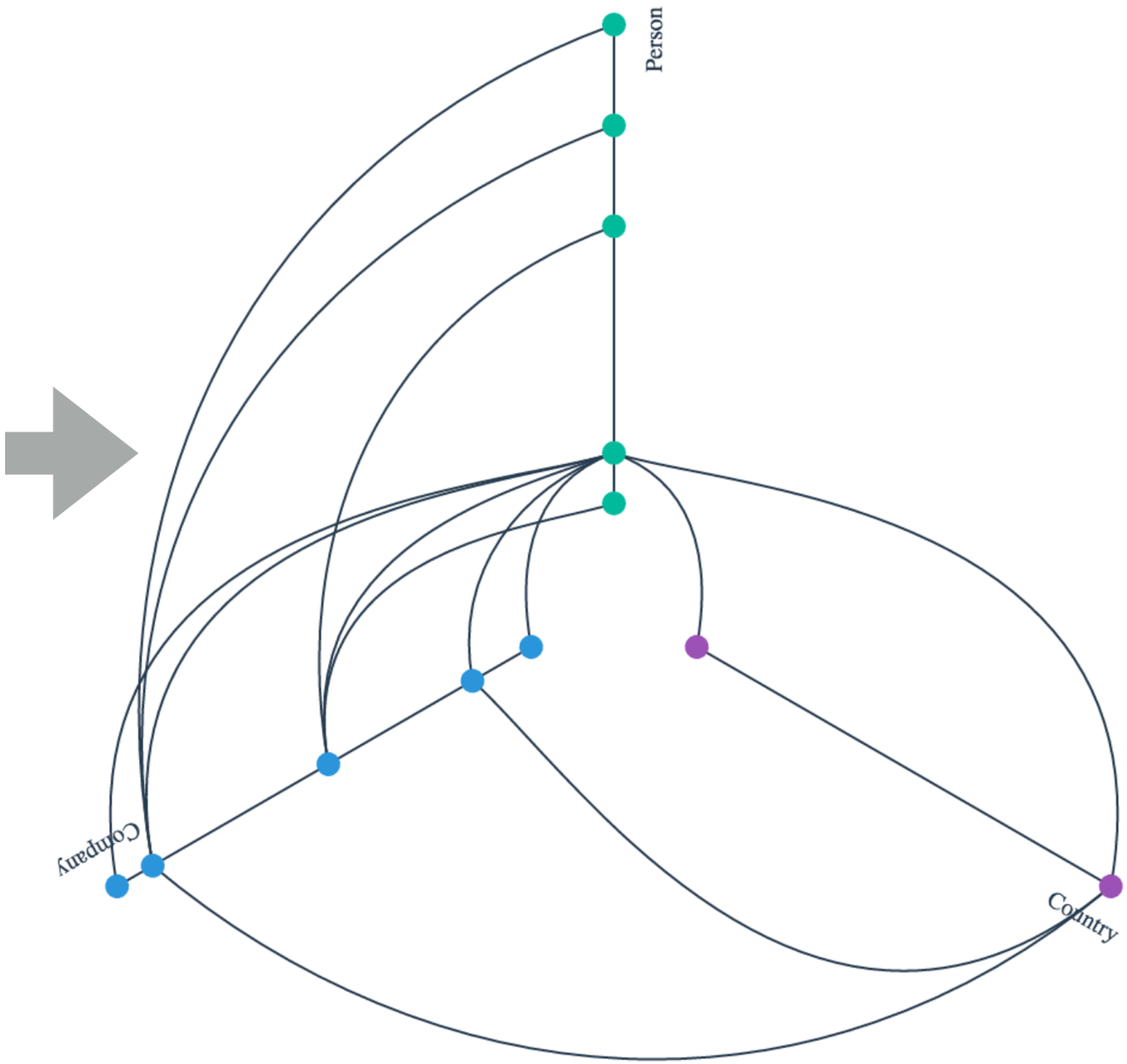


http://jsfiddle.net/7a7b5dwp/

Graphs/Networks

Hive Plots





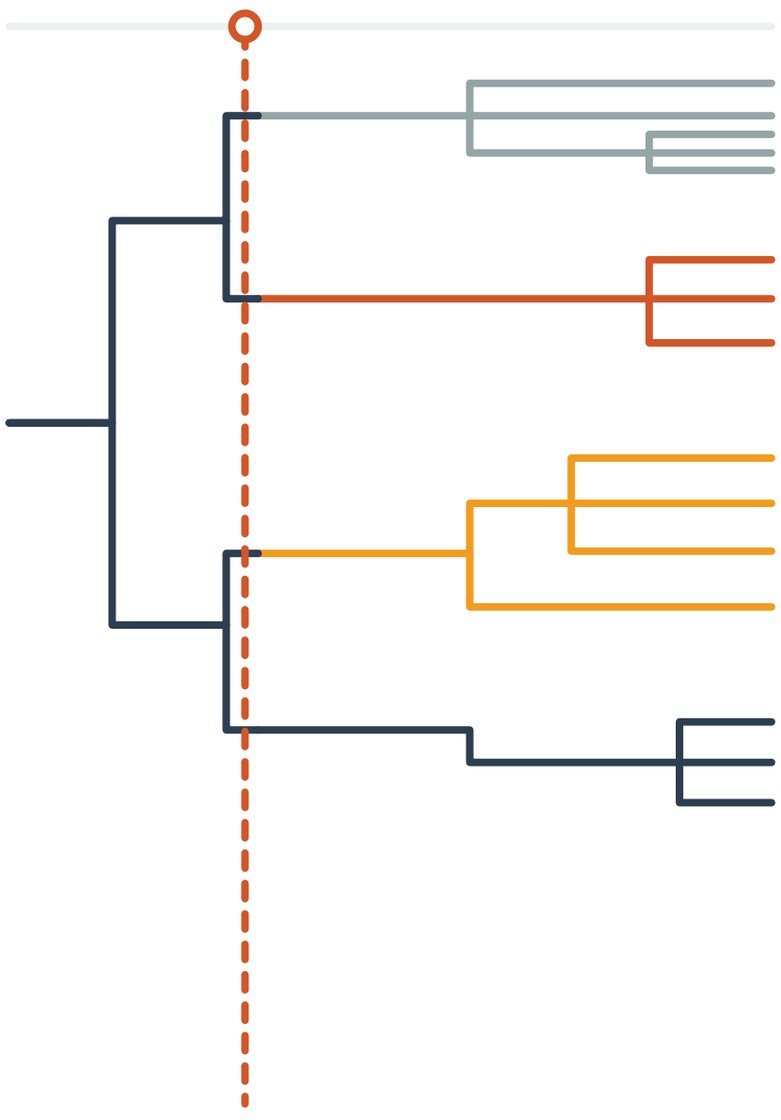
http://jsfiddle.net/7a7b5dwp/

http://jsfiddle.net/eamonnmag/vso70qnr/

Trees

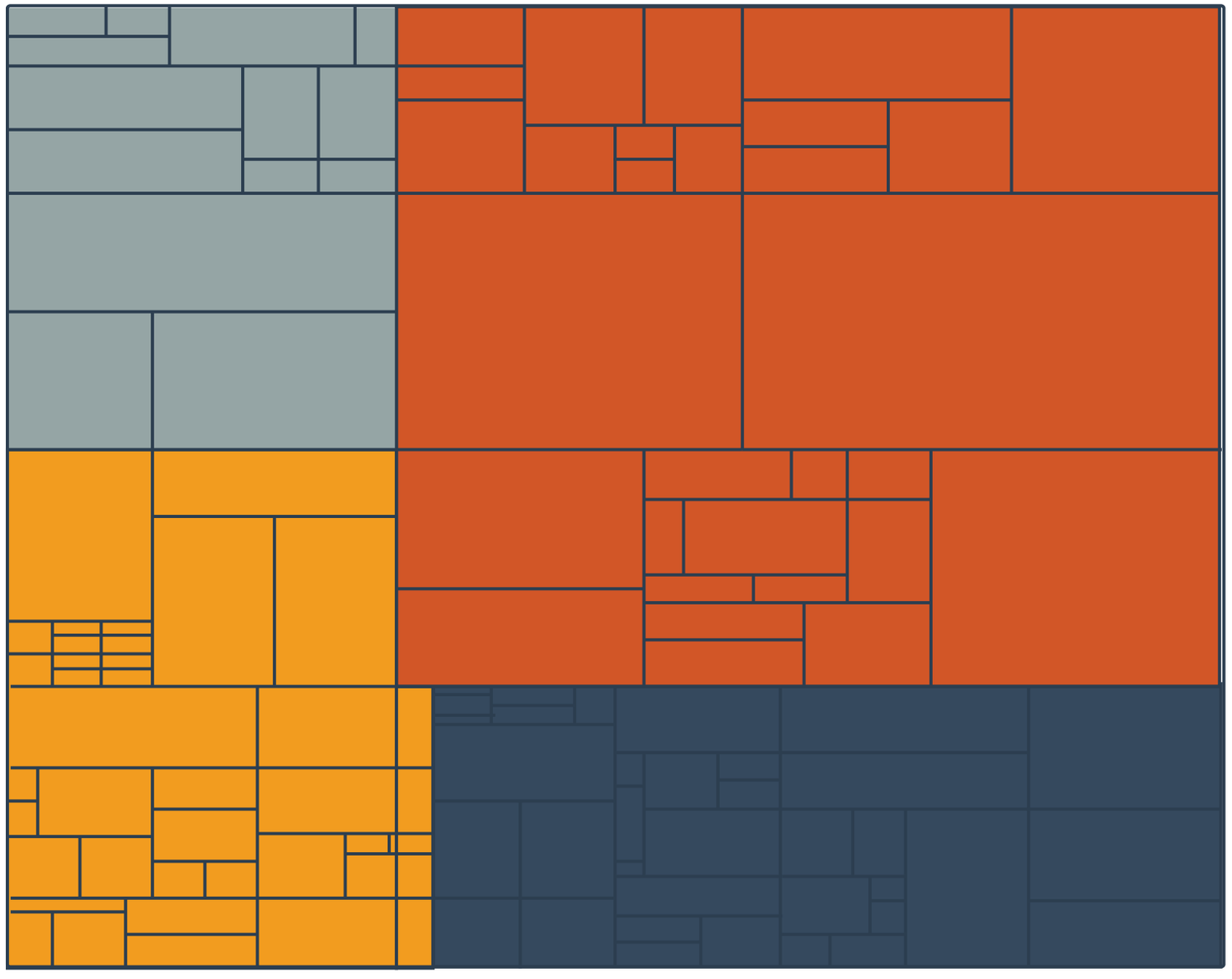
Dendrogram ?

Cut Level 0.3 4 Clusters



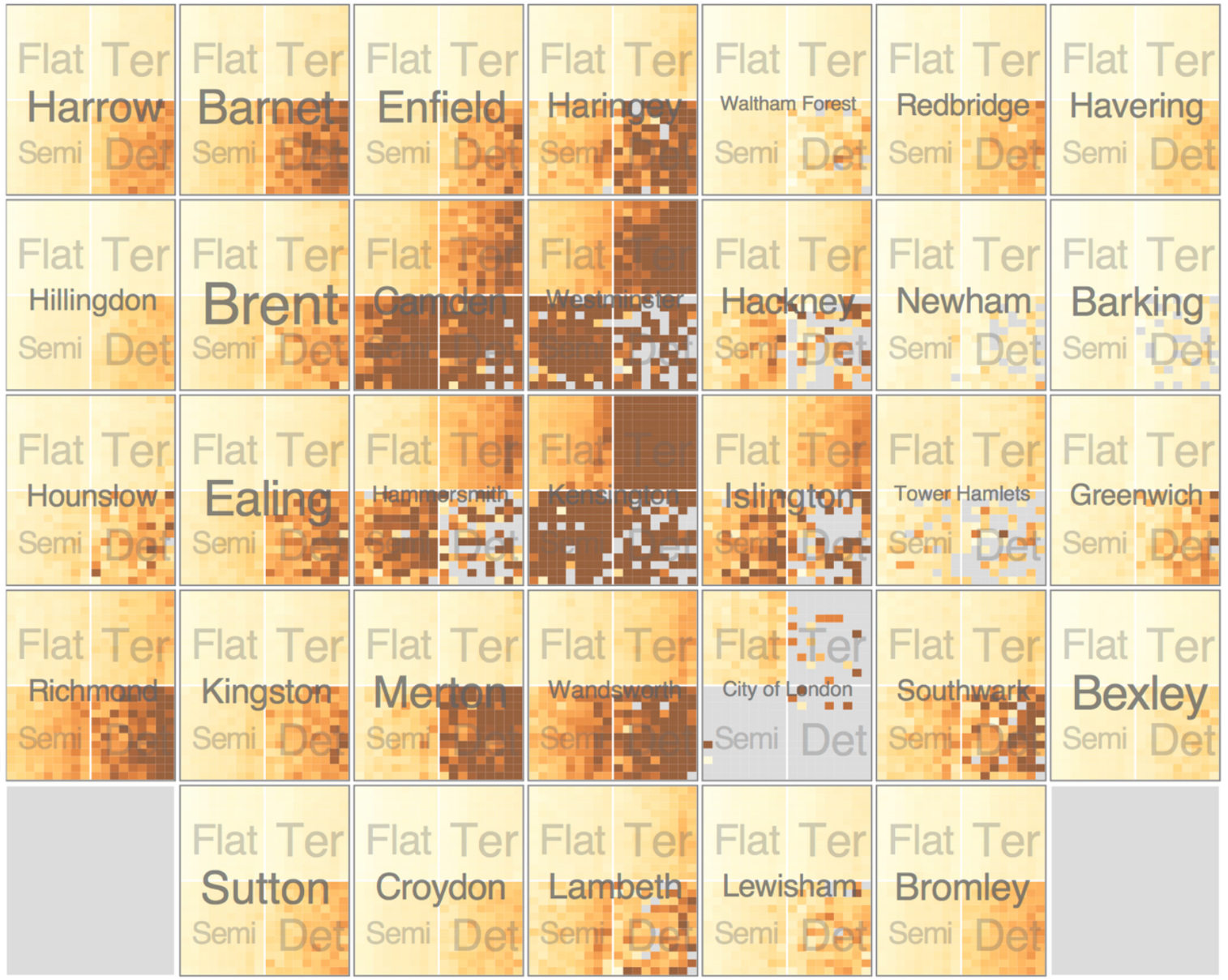
View as treemap ▼

Search ?



Treemaps Partitioning

- split by neighbourhood
- then by type
- colour by price

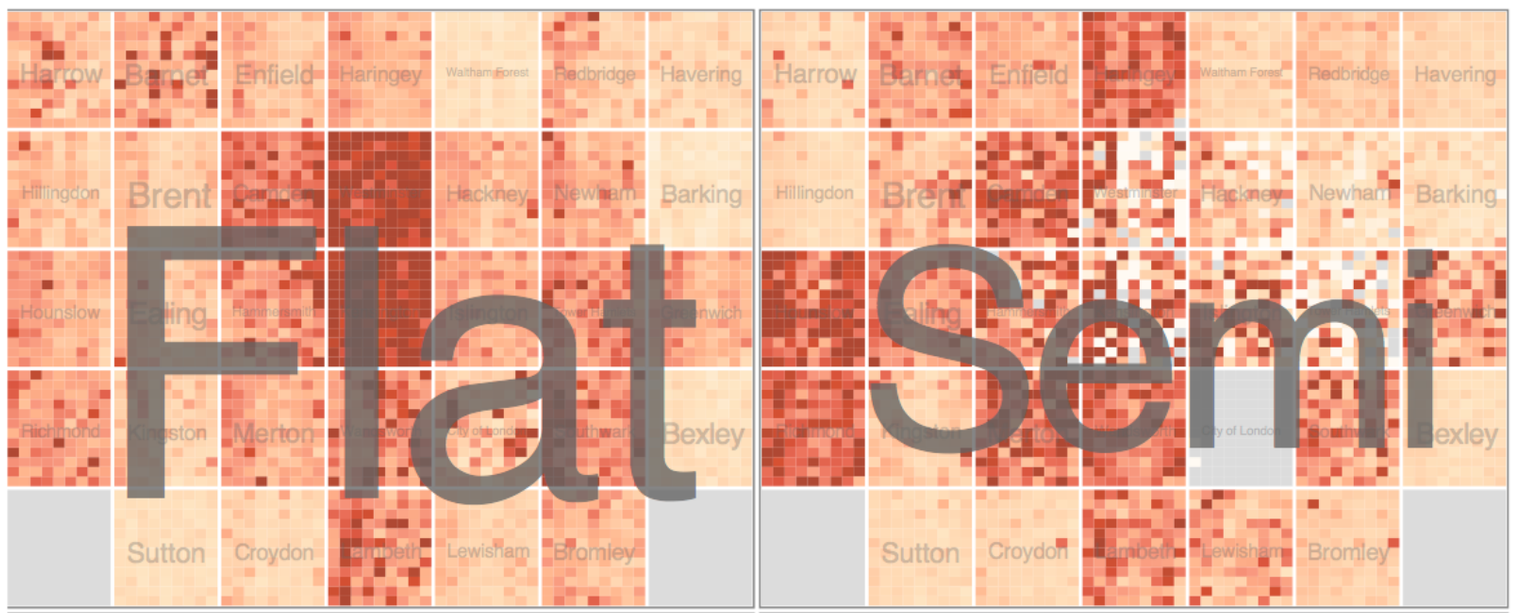


- neighbourhood patterns
  - where it's expensive
  - where you pay much more for detached type

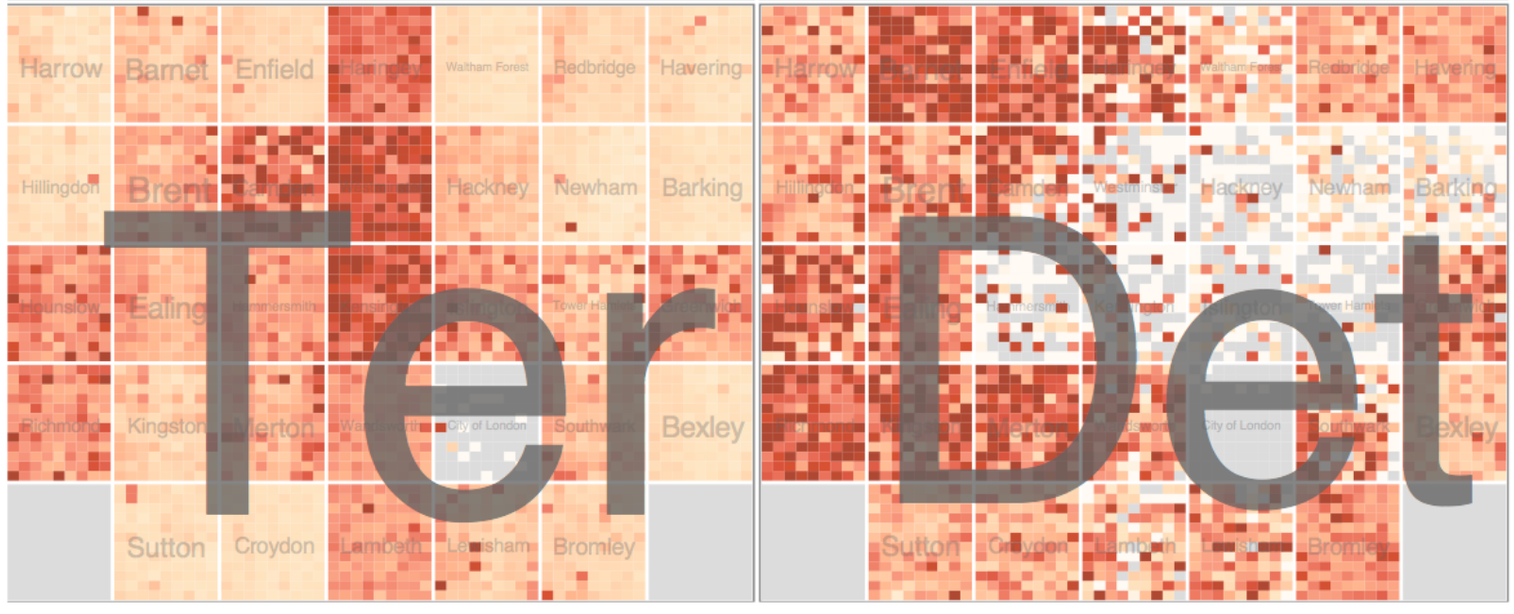
Configuring Hierarchical Layouts to Address Research Questions. Slingsby, Dykes, and Wood. IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2009) 15:6 (2009), 977–984.

Treemaps Partitioning

- switch order of splits
  - type then neighbourhood
- switch colour
  - by price variation



- type patterns
  - within specific type, which neighbourhoods are inconsistent



Configuring Hierarchical Layouts to Address Research Questions. Slingsby, Dykes, and Wood. IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2009) 15:6 (2009), 977–984.

Validation & User Testing

Validating your visualisation...

Validating your visualisation...

Domain situation
Observe target users using existing tools

Data/task abstraction

Visual encoding/interaction idiom
Justify design with respect to alternatives

Algorithm
Measure system time/memory
Analyze computational complexity

Analyze results qualitatively

Measure human time with lab experiment (*lab study*)

Observe target users after deployment (*field study*)

Measure adoption

The elephant in the room... scaling up

“Visualization can surprise you, but doesn't scale well. Modelling scales well, but can't surprise you.”

Hadley Wickham

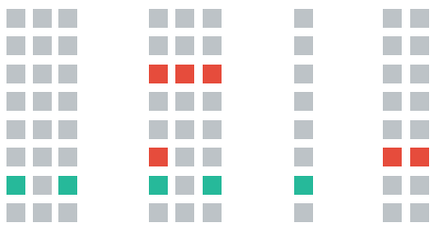
The elephant in the room... scaling up

One of the biggest challenges in HEP, Biology, chemistry, and business is scale.

Our screens have a limited number of pixels. And our data is often much larger.

You can do two things to get around this.

Solutions



Reduce the problem space

Try and get users to focus queries as soon as possible to reduce the amount of data to be visualised.

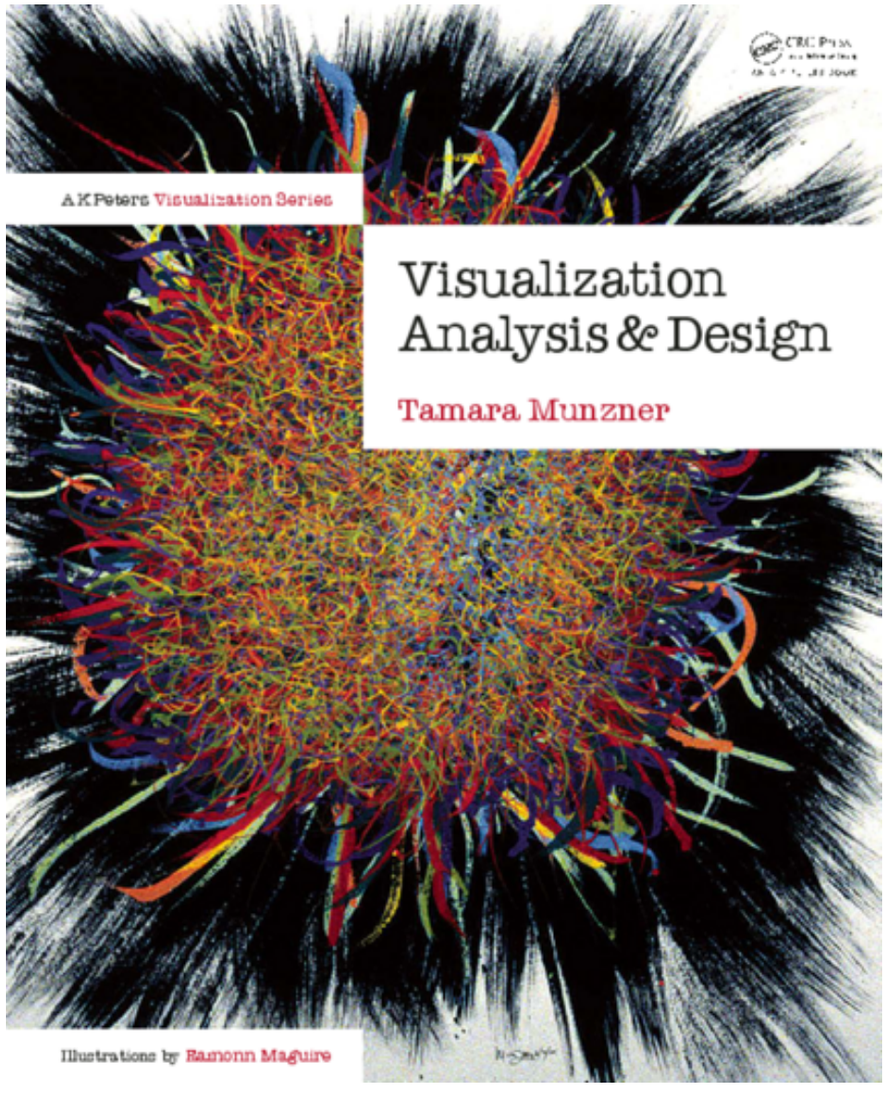
Provide ways to aggregate information in to meaningful overview visualisations...

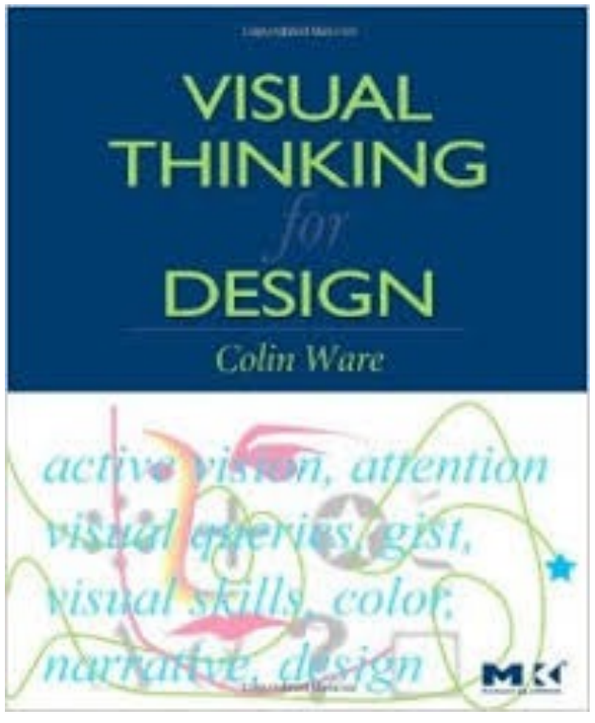
GPU usage and WebGL

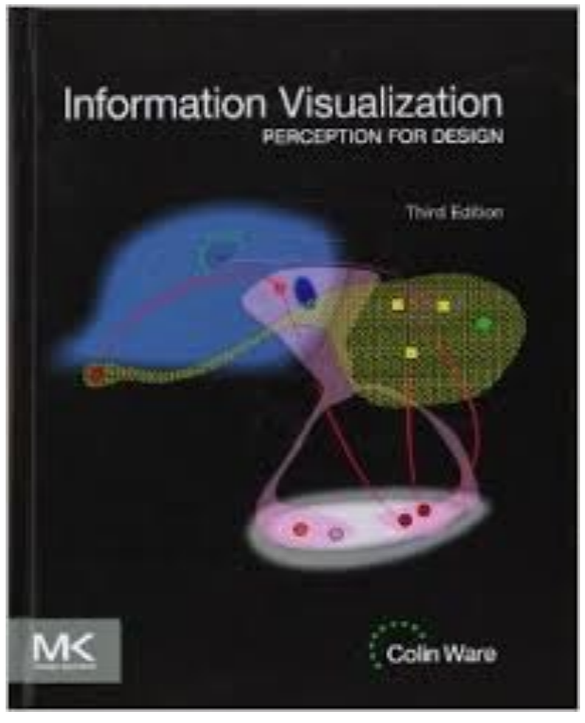
As done in imMens from Jeffrey Heer's lab in Washington. Renders billions of data points at 50 fps in the browser.

Not yet compatible across all browsers, and not everyone has a dedicated GPU.

More??







Visualization Analysis and Design.
Munzner. A K Peters Visualization Series, CRC Press, Visualization Series, 2014.

Further Links

Tutorial on D3

http://antarctic-design.co.uk/biovis-workshop15/

Tutorial on Dashboard Visualizations

https://thor-project.github.io/dashboard-tutorial/

Visualization Survey Sites

Set Visualization - http://www.cvast.tuwien.ac.at/SetViz

Time Series Visualization - http://survey.timeviz.net/

Parallel Coordinates Visualization

Periodic Table of Visualizations

Data Vis Catalogue

Questions

@antarcticdesign
eamonnmag@gmail.com