



# Eamonn Maguire

Principles of Data Visualization II  
University of Liege, Belgium  
27th April 2018

## We have to be careful when mapping data to the visual world

Some visual channels are more effective for some data types over others.

Some data has a natural mapping that our brains expect given certain types of data

There are many visual tricks that can be observed due to how the visual system works

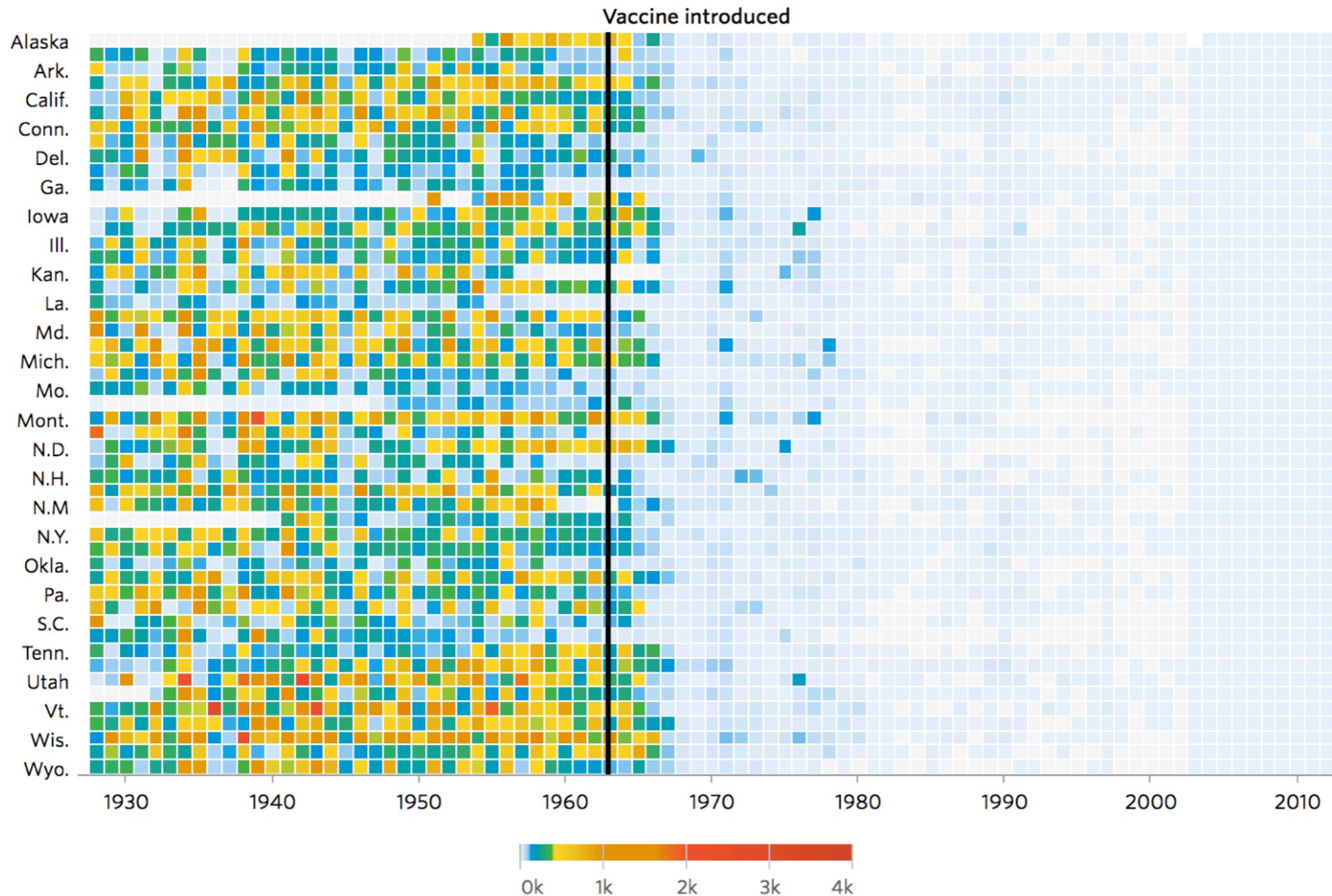
We don't see in 3D, and we have difficulties interpreting information on the Z-axis.

Colour

# Color

The simplest, yet most abused of all visual encodings.

## Measles



<http://graphics.wsj.com/infectious-diseases-and-vaccines/>

# Color

The simplest, yet most abused of all visual encodings.

SANFORD AND SELNICK

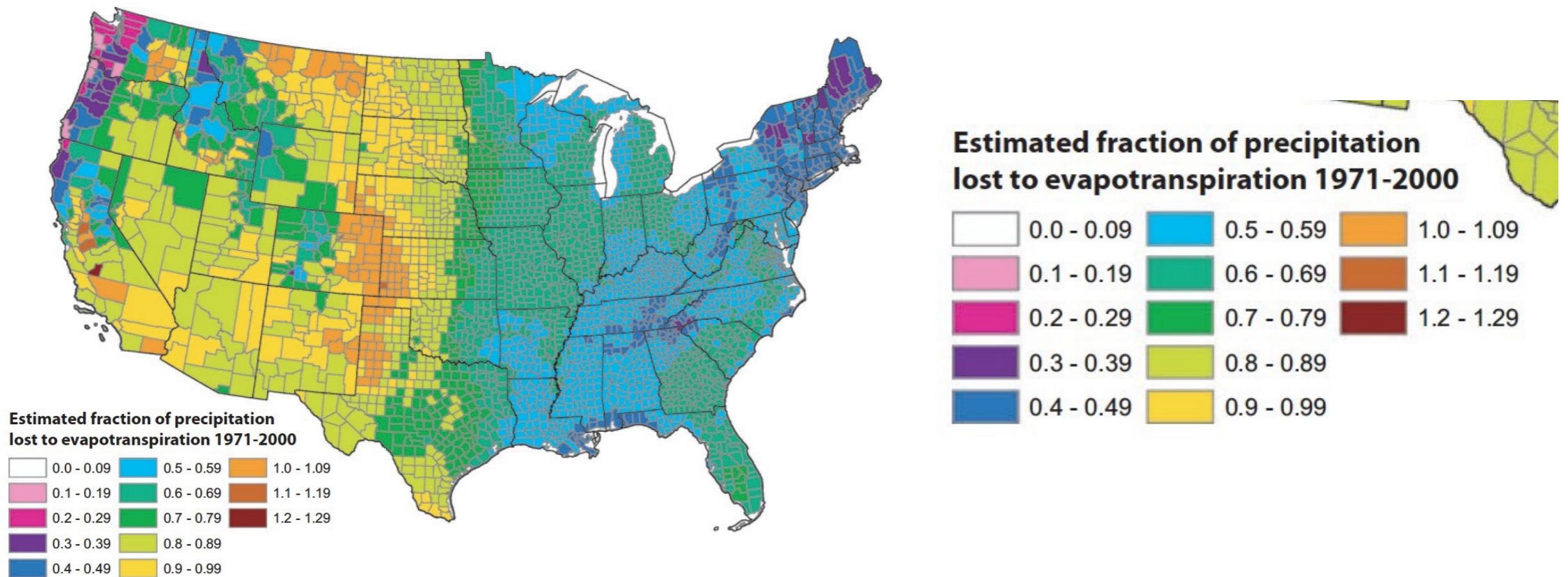
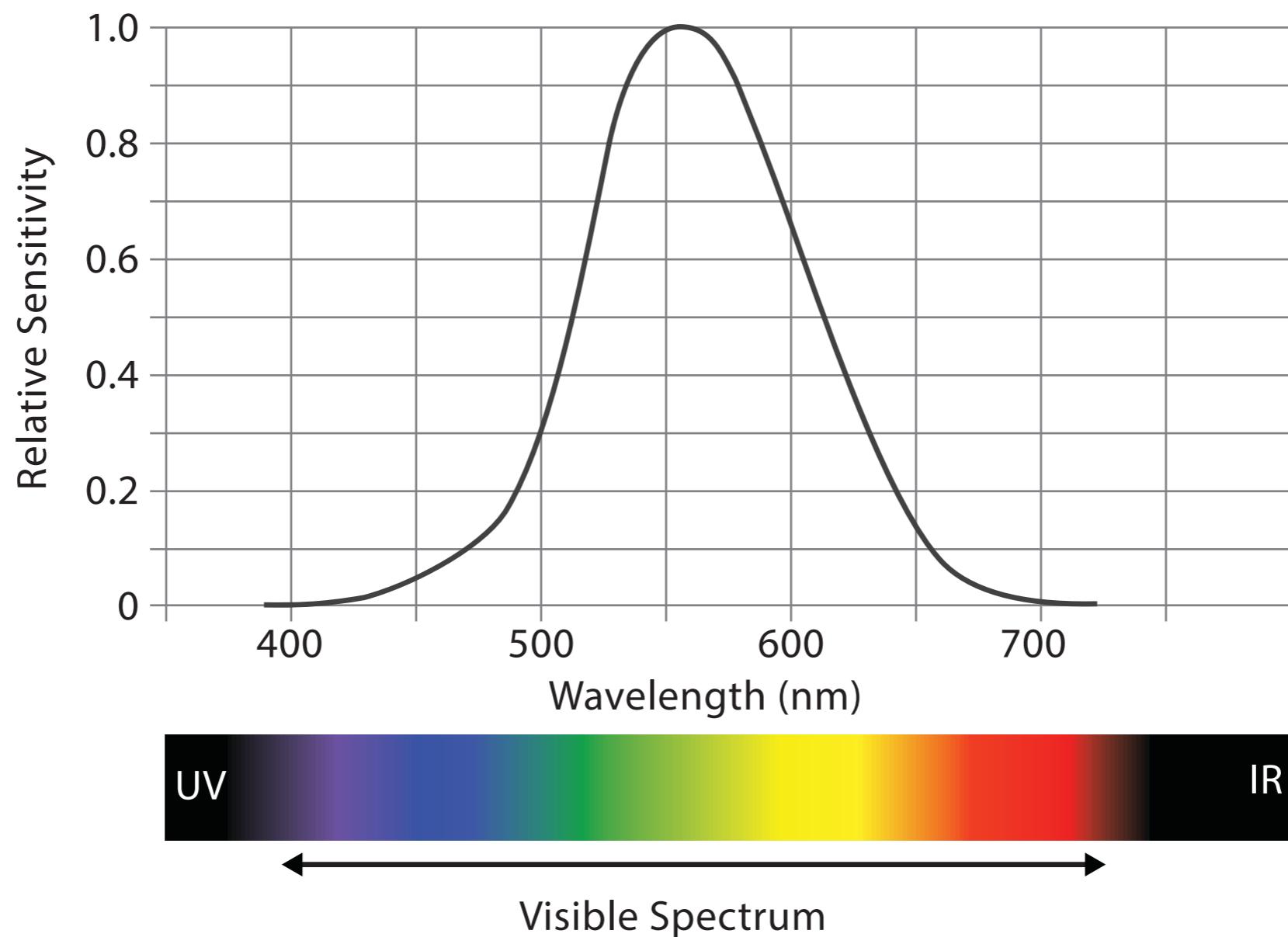


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation ( $P$ ) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of  $ET/P$  were made first at the 800-m resolution of the PRISM climate data. The mean values for the counties (shown) were then calculated by averaging the 800-m values within each county. Areas with fractions  $>1$  are agricultural counties that either import surface water or mine deep groundwater.

The problem is that a smooth step in a value does not equate to a smooth colour transition...

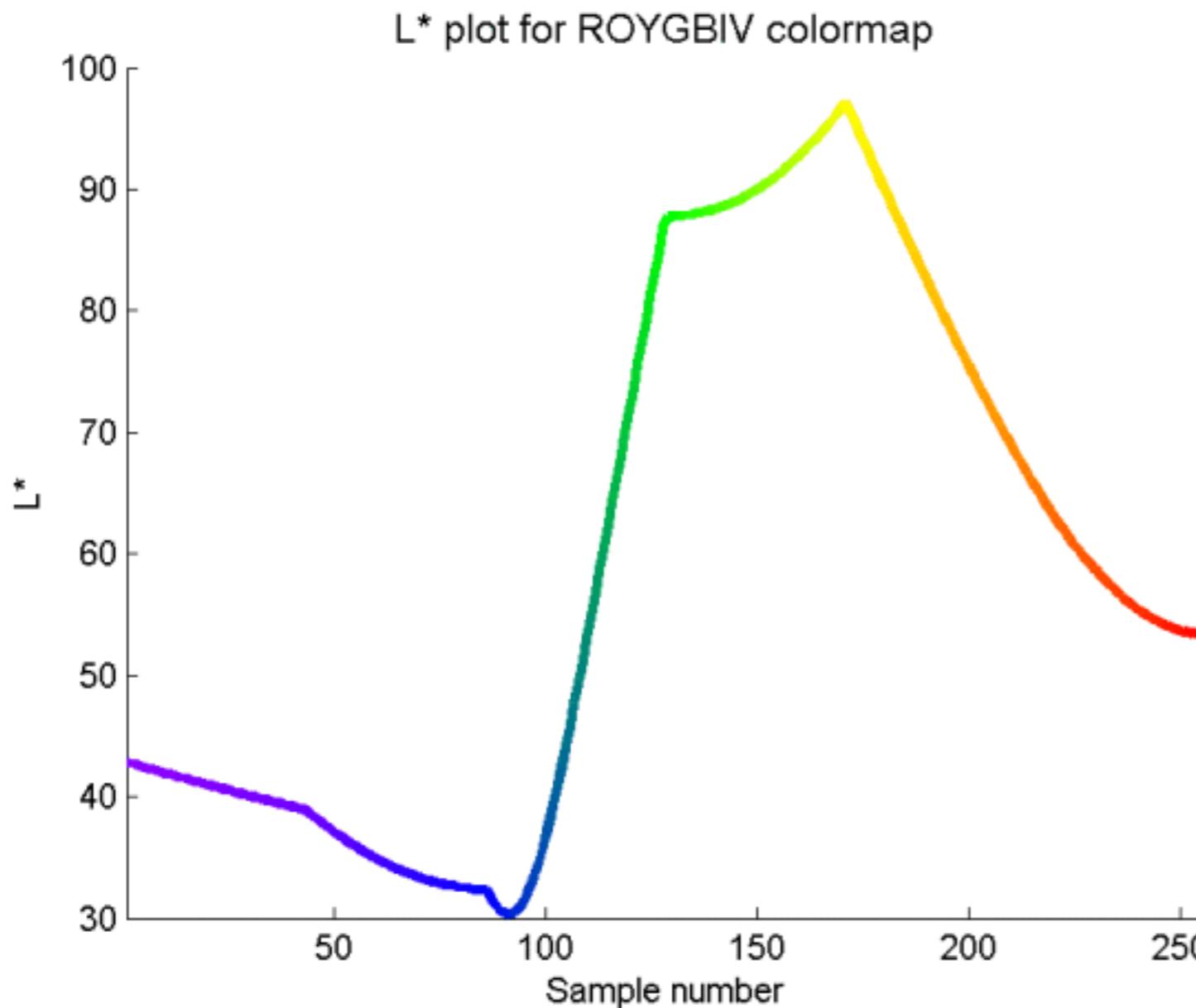
# Color

Additionally, colour is not equally binned in reality. We perceive colours differently due to an increased sensitivity to the yellow part of the spectrum...



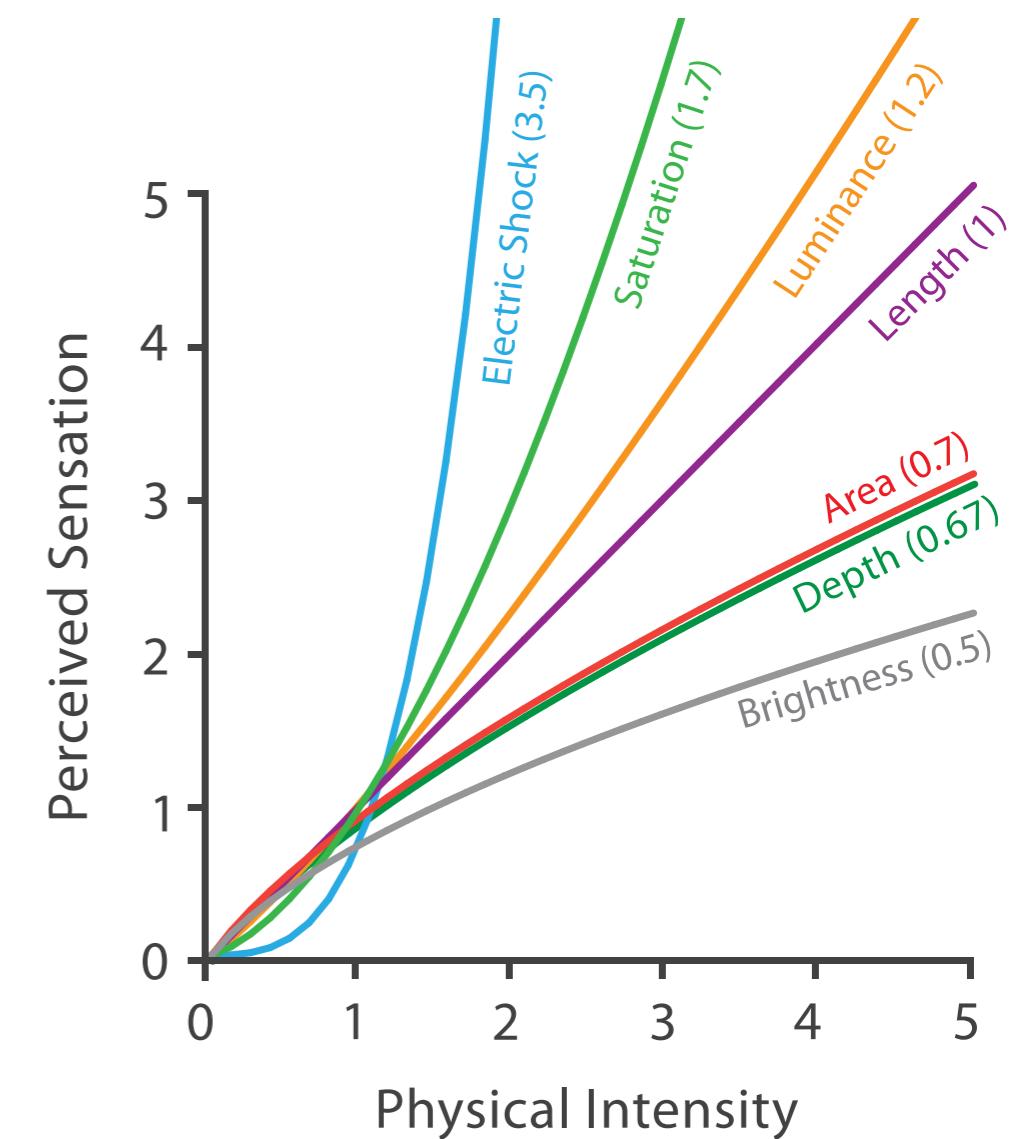
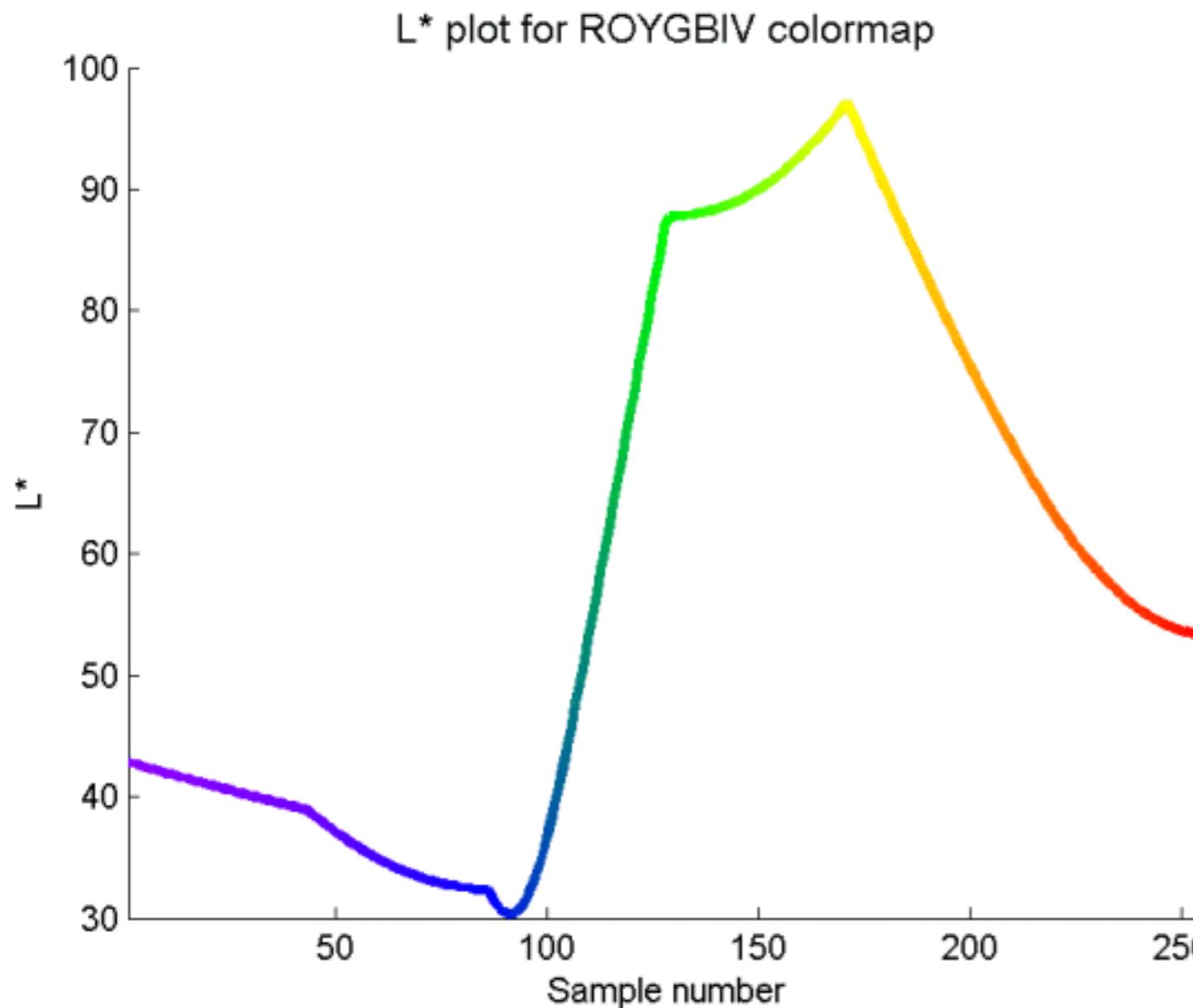
# Color

Luminosity is also not stable across the colours, meaning some colours will pop out more than others... and not always intentionally.



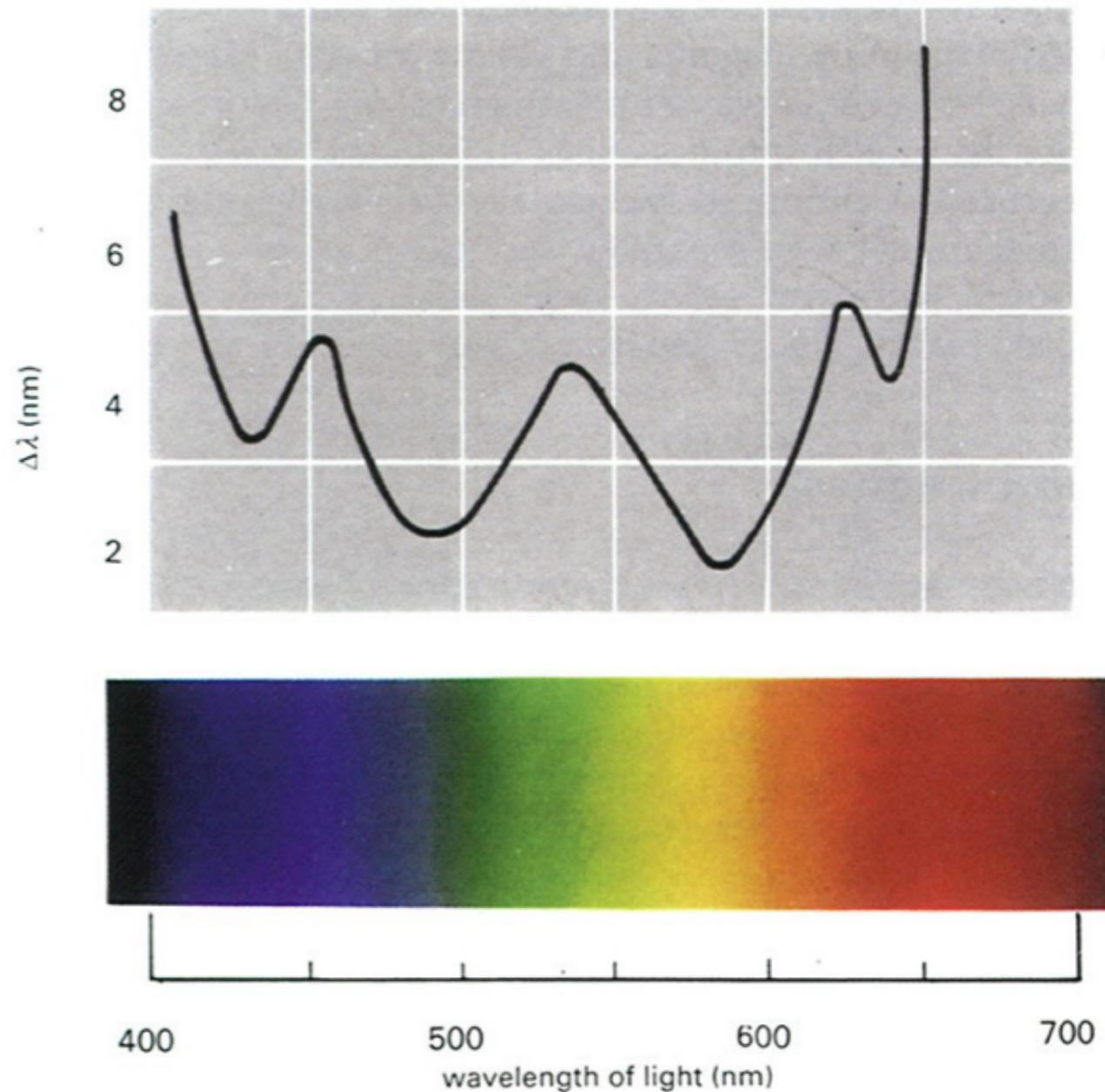
# Color

Luminosity is also not stable across the colours, meaning some colours will pop out more than others... and not always intentionally.



# Color

And how we perceive changes in hue is also very different.



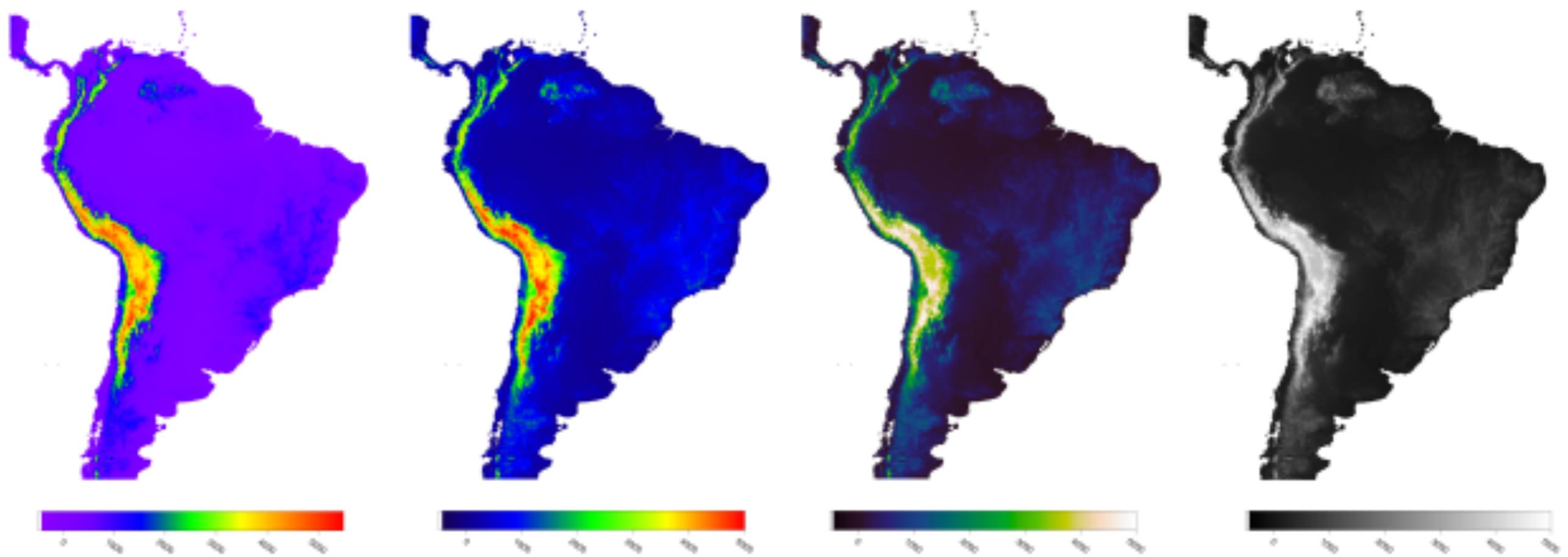
Copyright: Richard L. Gregory - Eye and Brain - Princeton University Press - used with permission

Gregory compared the wavelength of light with the smallest observable difference in hue (expressed as wavelength difference)

**Is there a colour palette for scientific visualization  
that works?**

# Color

## HSL linear L rainbow palette



<https://mycarta.wordpress.com/2012/10/06/the-rainbow-is-deadlong-live-the-rainbow-part-3/>

Kindlmann, G., Reinhard, E. and Creem, S., 2002, Face-based Luminance Matching for Perceptual Colormap Generation, IEEE Proceedings of the conference on Visualization '02

# Color

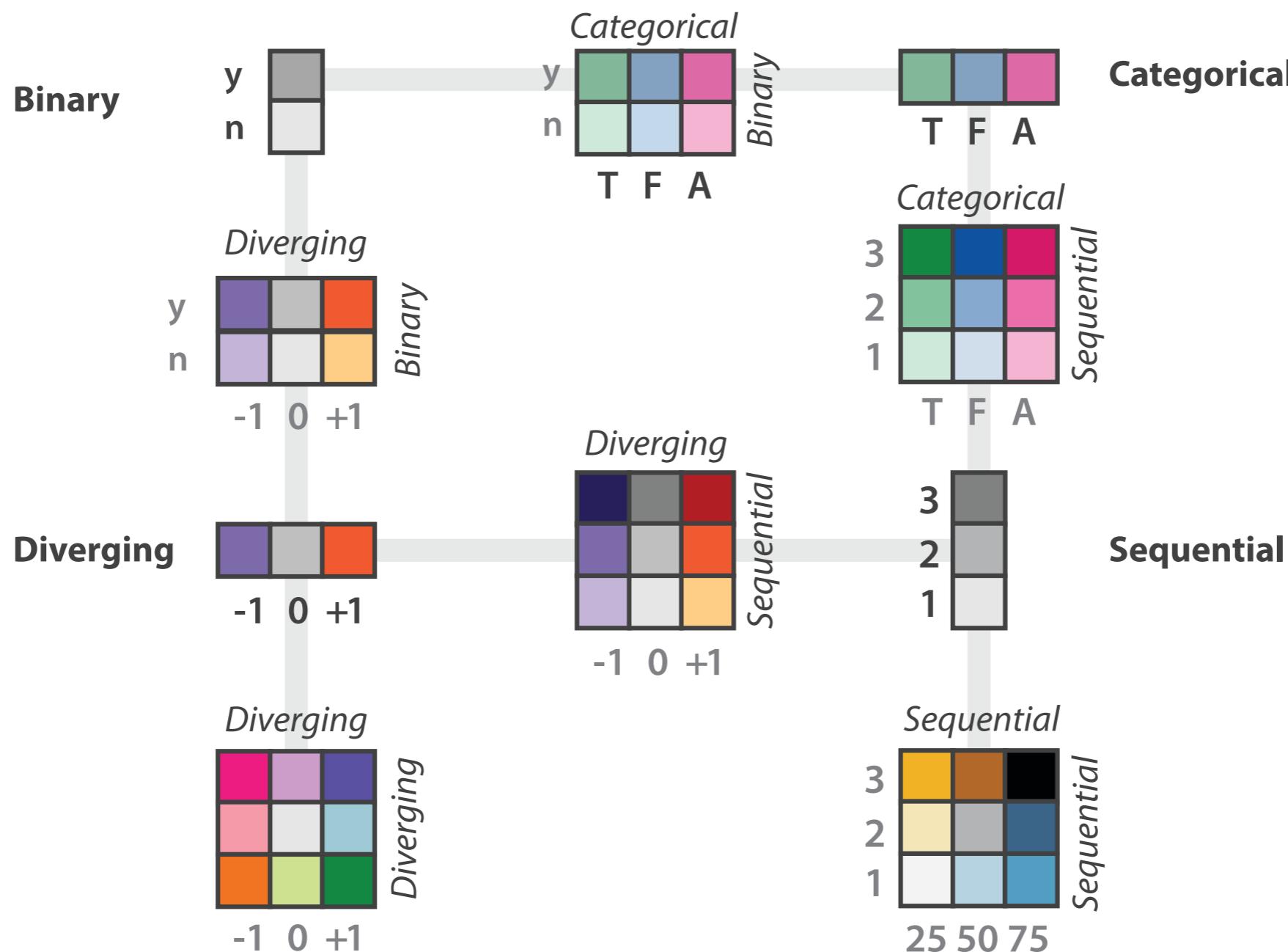
## HSL linear L rainbow palette



These are available in matplotlib and therefore in seaboard etc, so there's no excuse :)

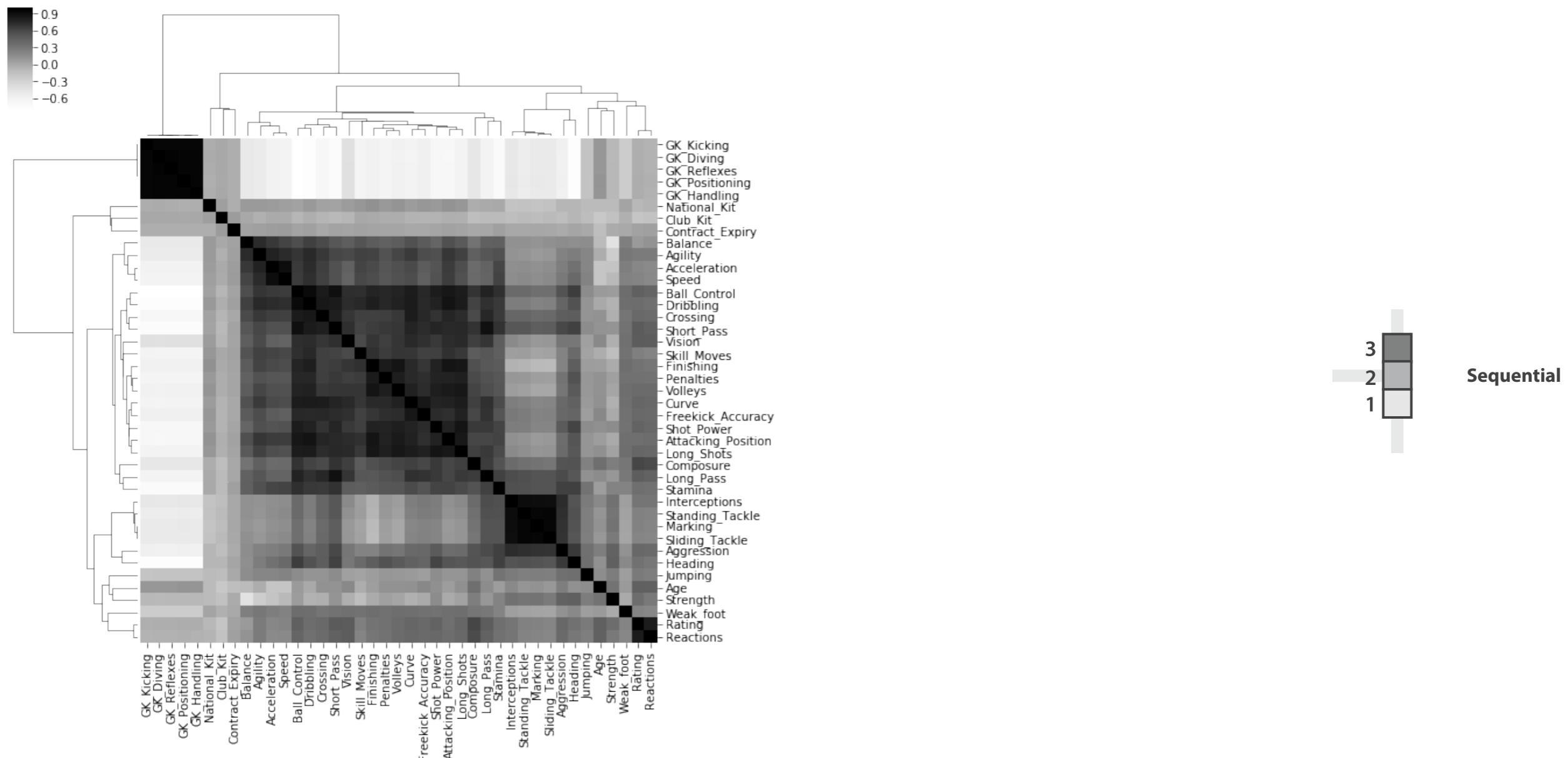
# Color

There are also lots of default colour maps that can be applied to particular data types.



# Color

Here I'm showing the correlation between football player attributes.  
Is the choice of colour map helping this comparison?



```
import seaborn as sns  
sns.clustermap(fifa.corr(), cmap='Greys')
```

# Color

Here I'm showing the correlation between football player attributes.  
Is the choice of colour map helping this comparison?

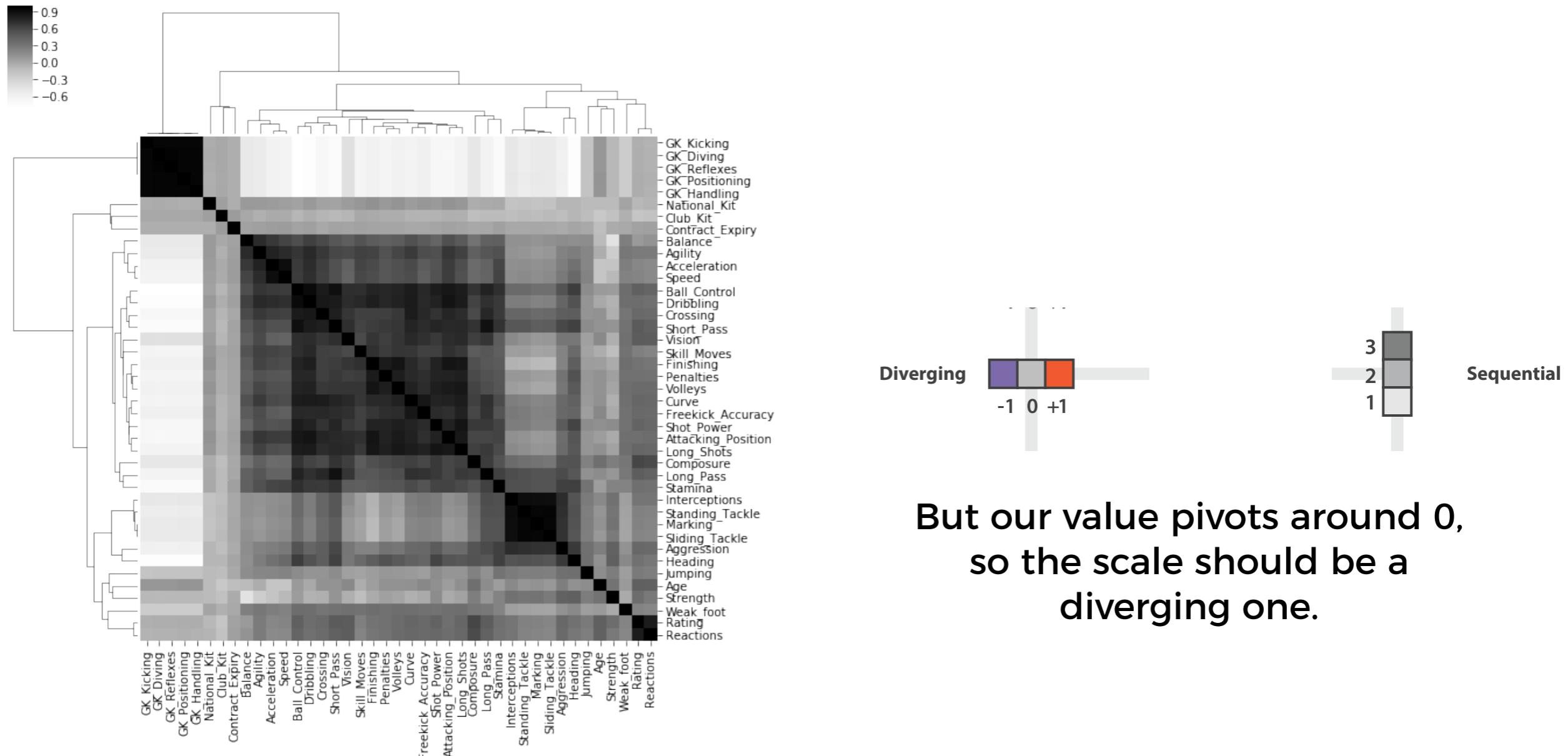


But our value pivots around 0,  
so the scale should be a  
diverging one.

```
import seaborn as sns
sns.clustermap(fifa.corr(), cmap='Greys')
```

# Color

Here I'm showing the correlation between football player attributes.  
Is the choice of colour map helping this comparison?

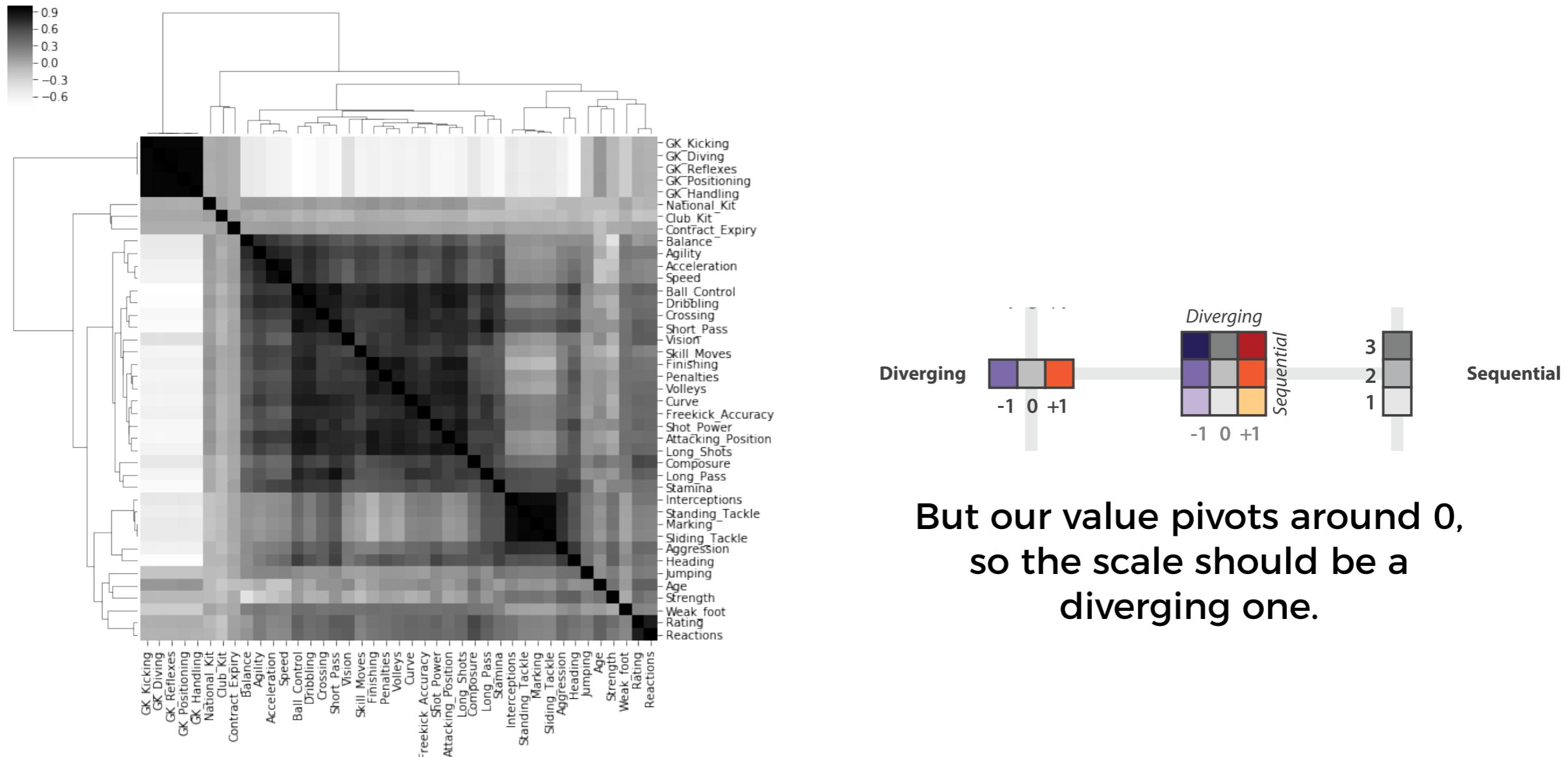


But our value pivots around 0,  
so the scale should be a  
diverging one.

```
import seaborn as sns
sns.clustermap(fifa.corr(), cmap='Greys')
```

# Color

Here I'm showing the correlation between football player attributes.  
Is the choice of colour map helping this comparison?

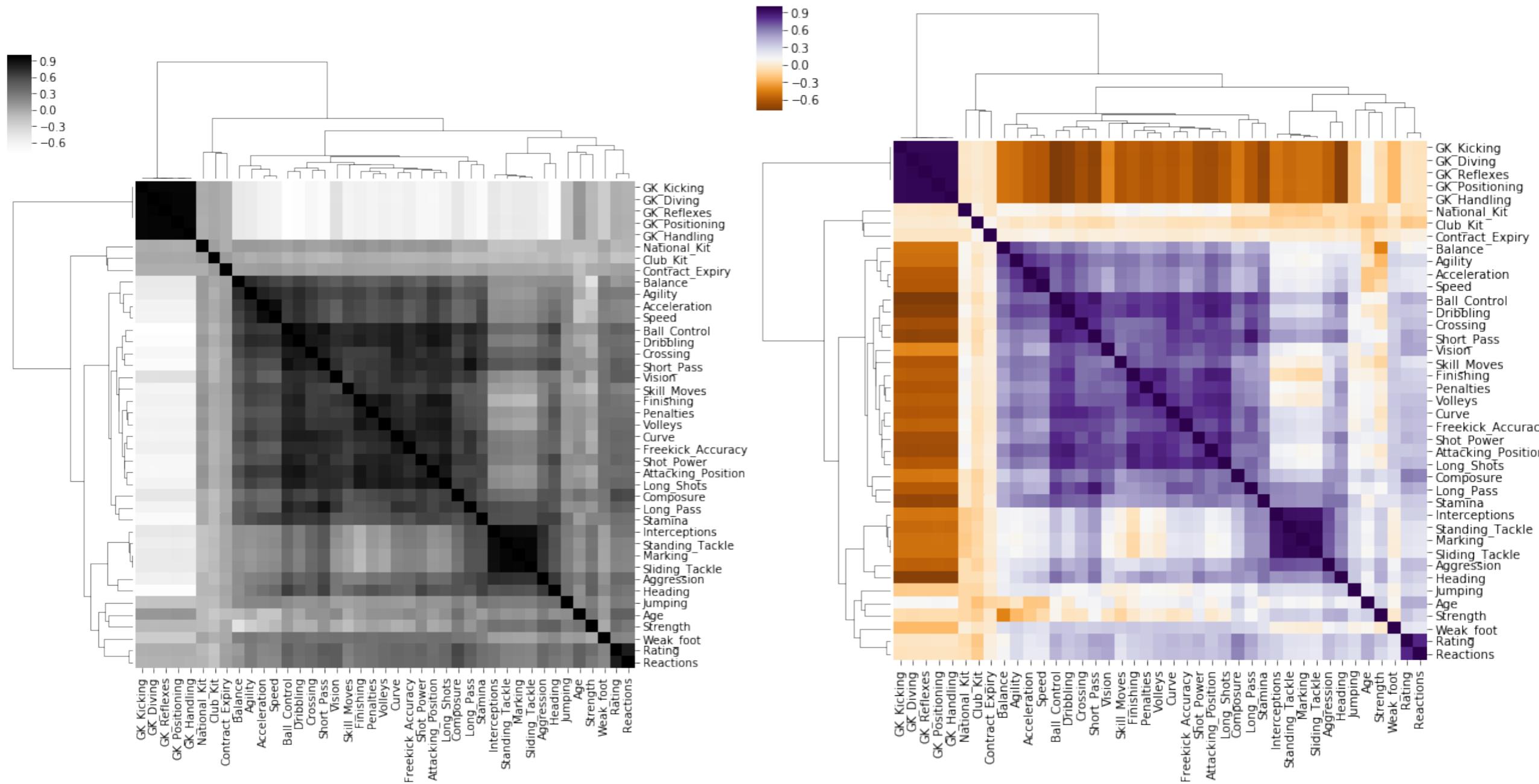


But our value pivots around 0,  
so the scale should be a  
diverging one.

```
import seaborn as sns
sns.clustermap(fifa.corr(), cmap='Greys')
```

# Color

Here I'm showing the correlation between football player attributes.  
Is the choice of colour map helping this comparison?

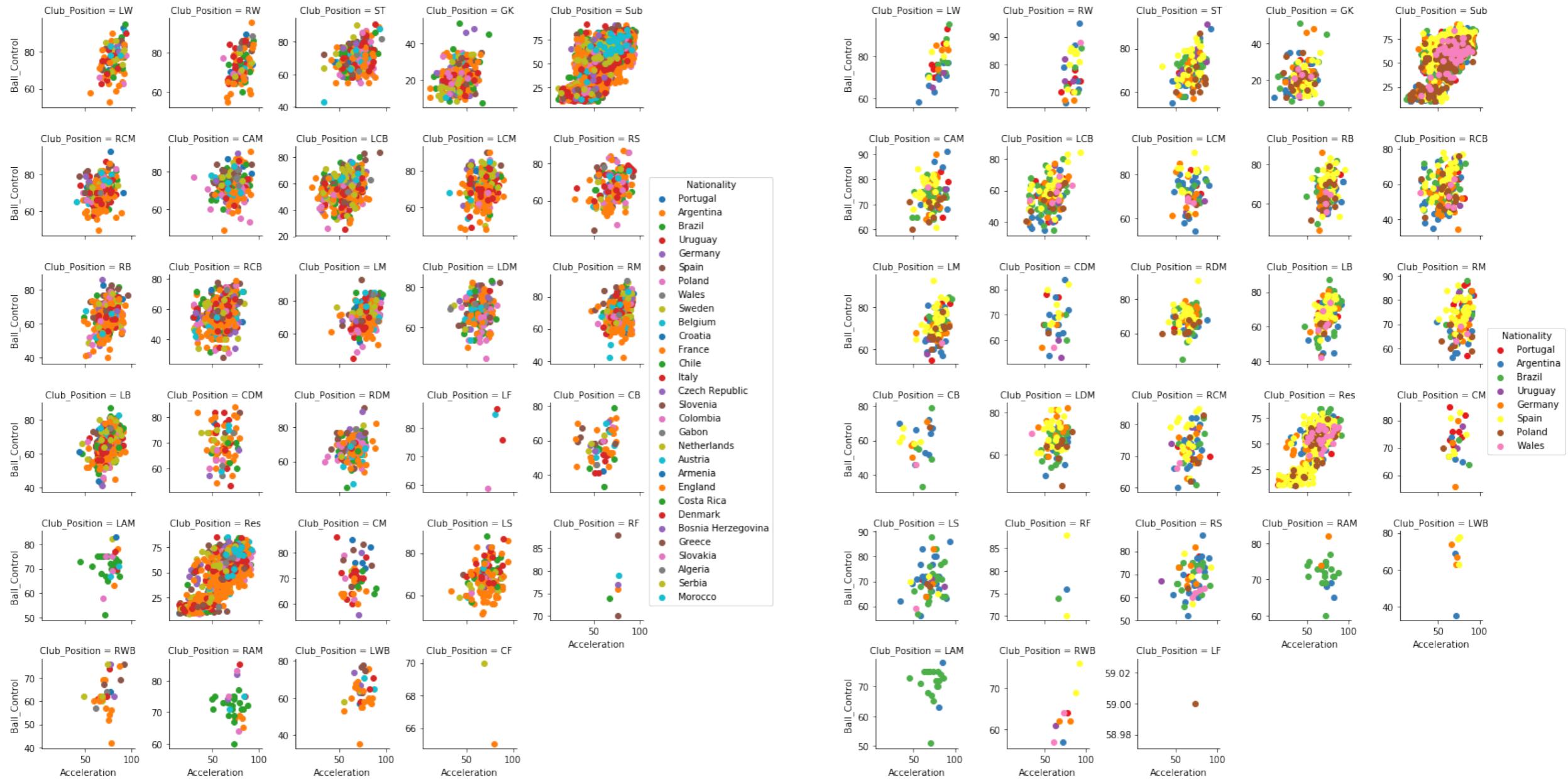


```
import seaborn as sns  
sns.clustermap(fifa.corr(), cmap='PuOr')
```

# Color

You also don't want to have too many colours.

Too many colours means that users have to remember what a colour means. So a max of around 8 categories in a plot is recommended, otherwise the 'distance' between colours becomes too small.



# Color

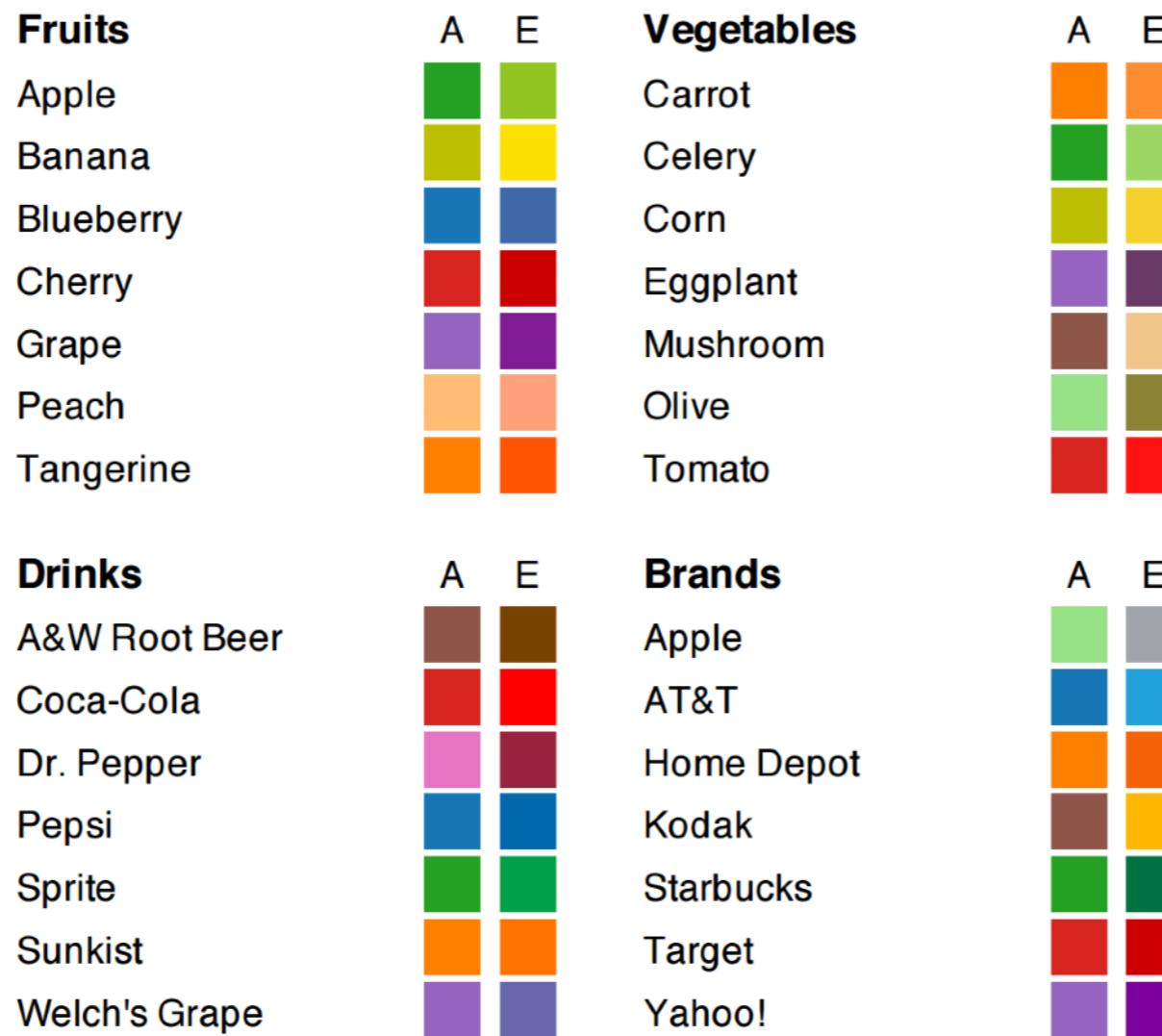
**Semantic relevance**

Or just consistency

When there are many colours for example, we find it difficult to remember abstract associations.

# Color

## What are semantically resonant colours?



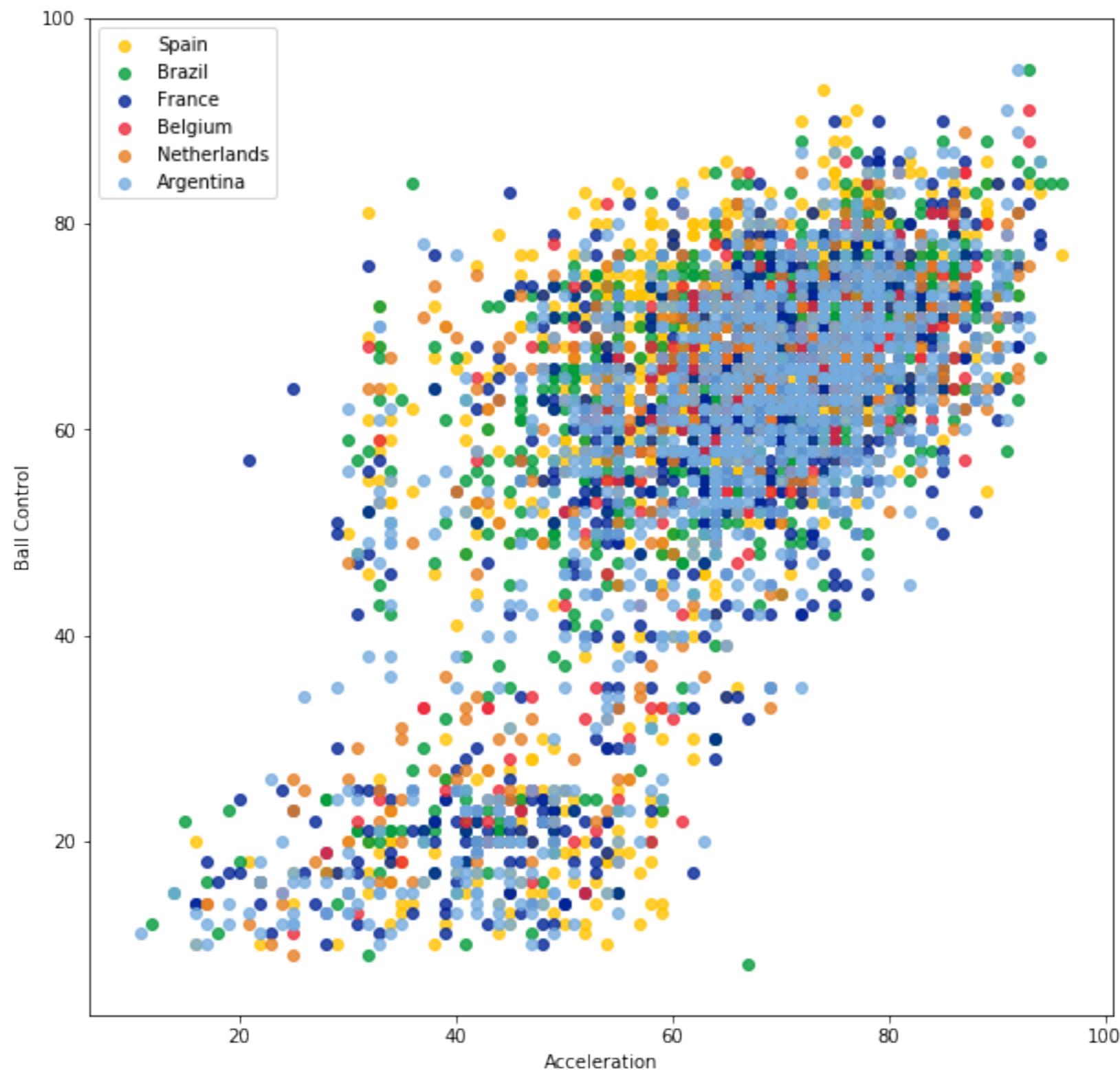
**Figure 6:** Color assignments for categorical values in Experiment 1. (A = Algorithm, E = Expert)

Selecting Semantically-Resonant Colors for Data Visualization

Sharon Lin, Julie Fortuna, Chinmay Kulkarni, Maureen Stone, Jeffrey Heer  
Computer Graphics Forum (Proc. EuroVis), 2013

# Color

What are semantically resonant colours?



# Color

Semantic colouring is a good idea in theory, but there are limited areas where this really works.

But, if you are going to use colour, try to think how you can make it easier for users to decode the colour to the category without constantly having to look up a legend. That way, the decoding time is less.

Saving time...

## We have to be careful when mapping data to the visual world

Some visual channels are more effective for some data types over others.

Some data has a natural mapping that our brains expect given certain types of data

There are many visual tricks that can be observed due to how the visual system works

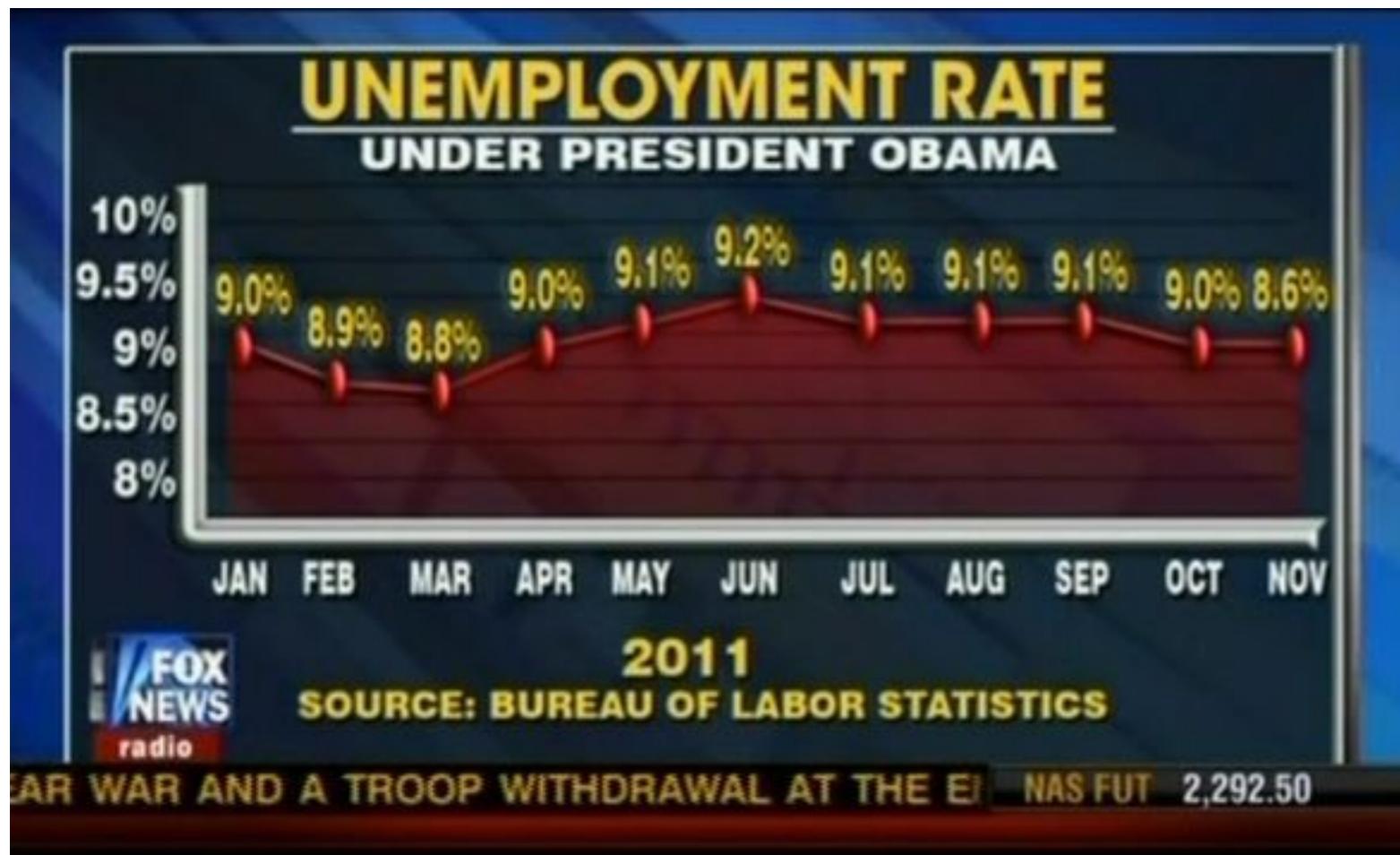
We don't see in 3D, and we have difficulties interpreting information on the Z-axis.

Colour

Scales

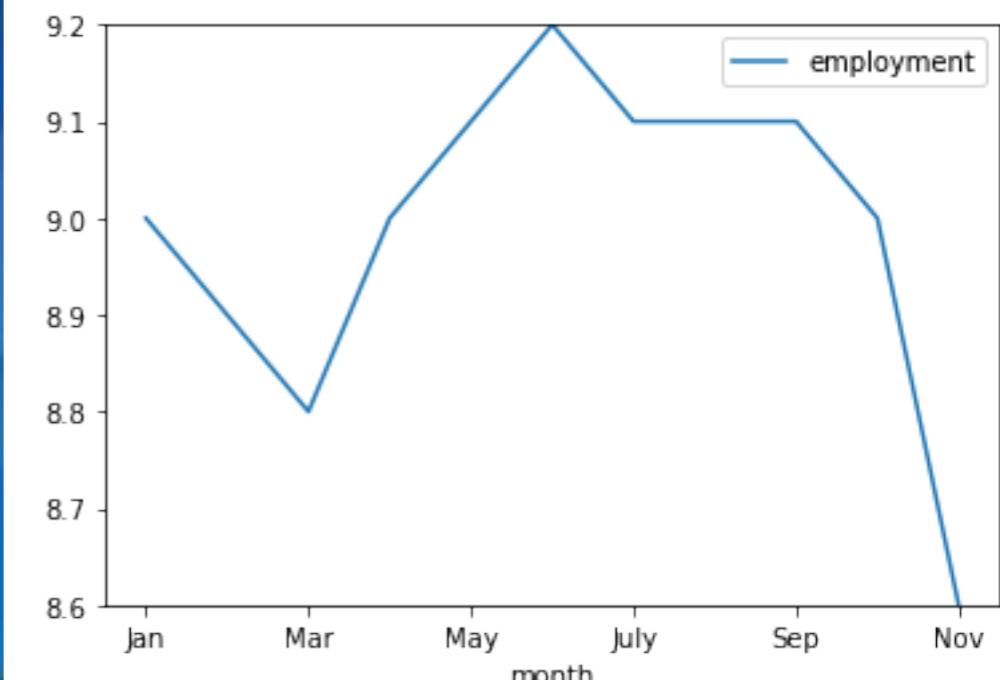
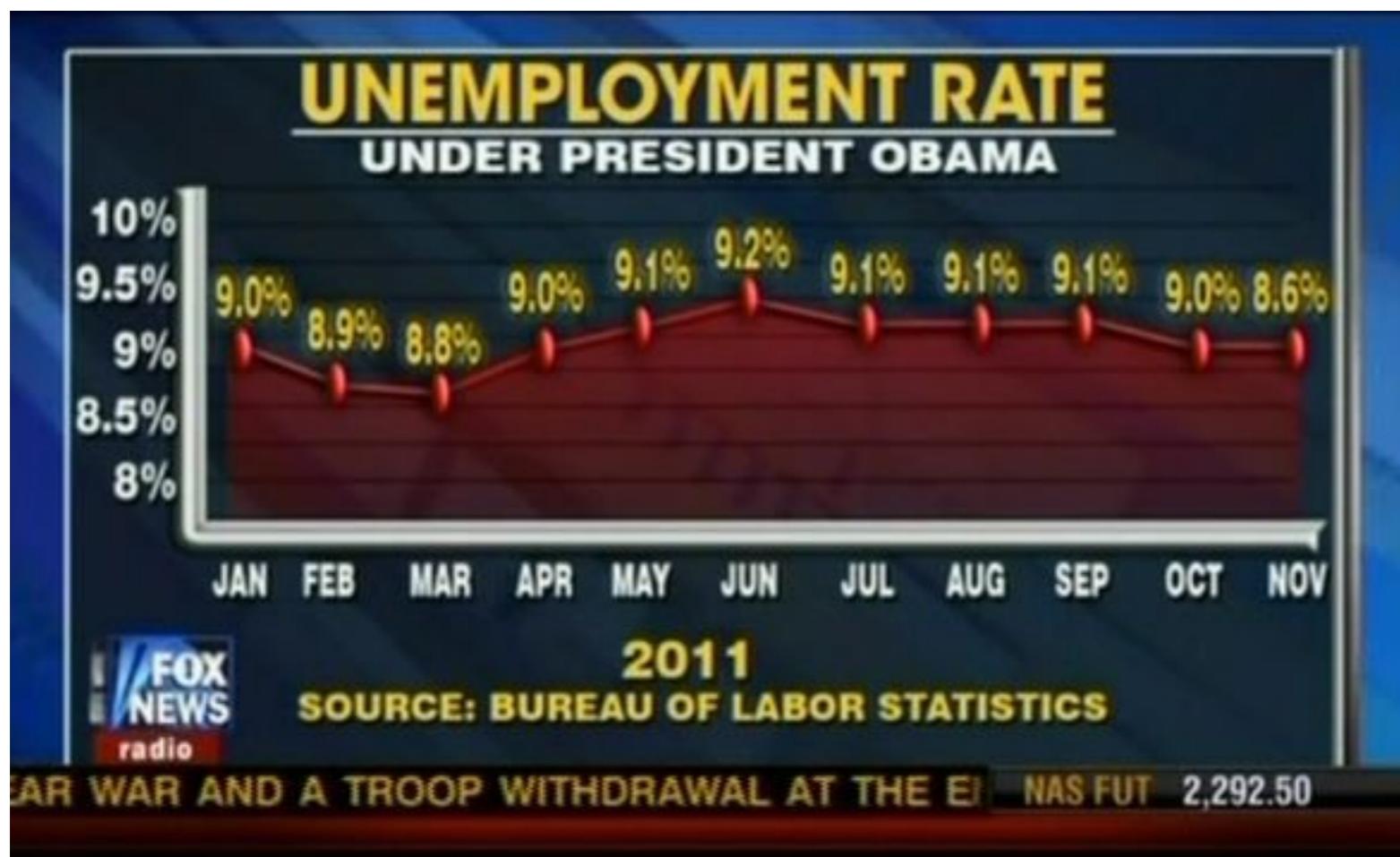
## Scales

Be aware of traps in visualizing data, when creating or reading.  
Especially with scale.



## Scales

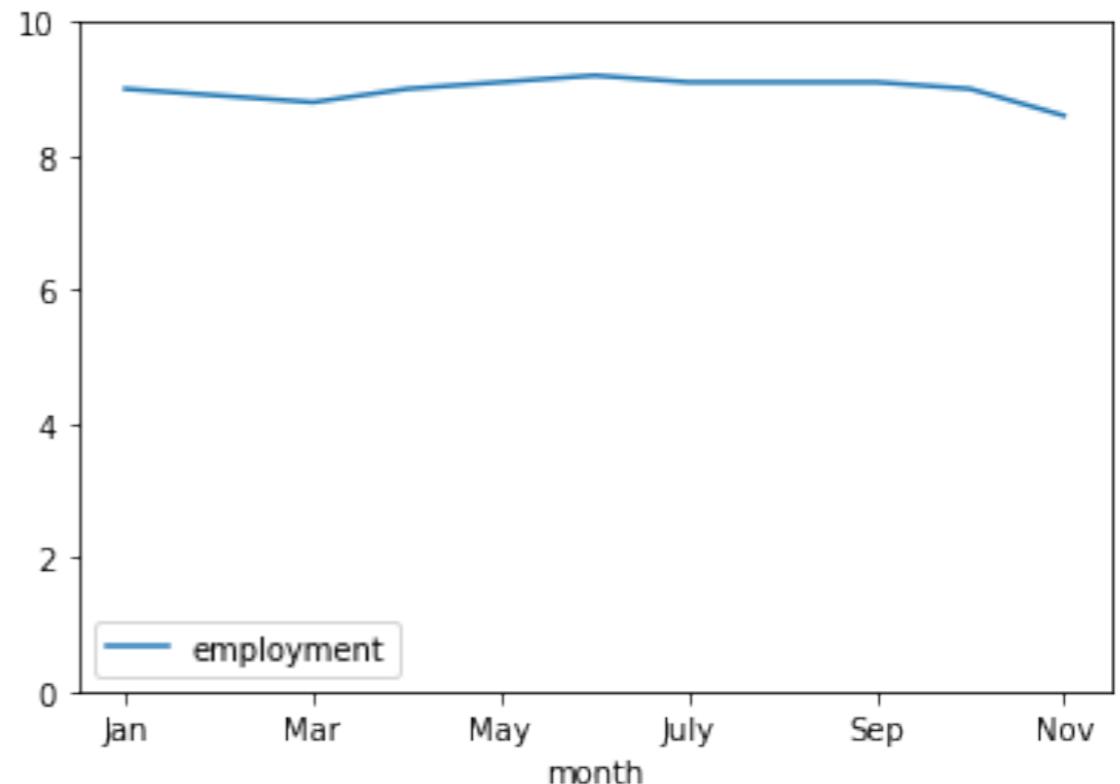
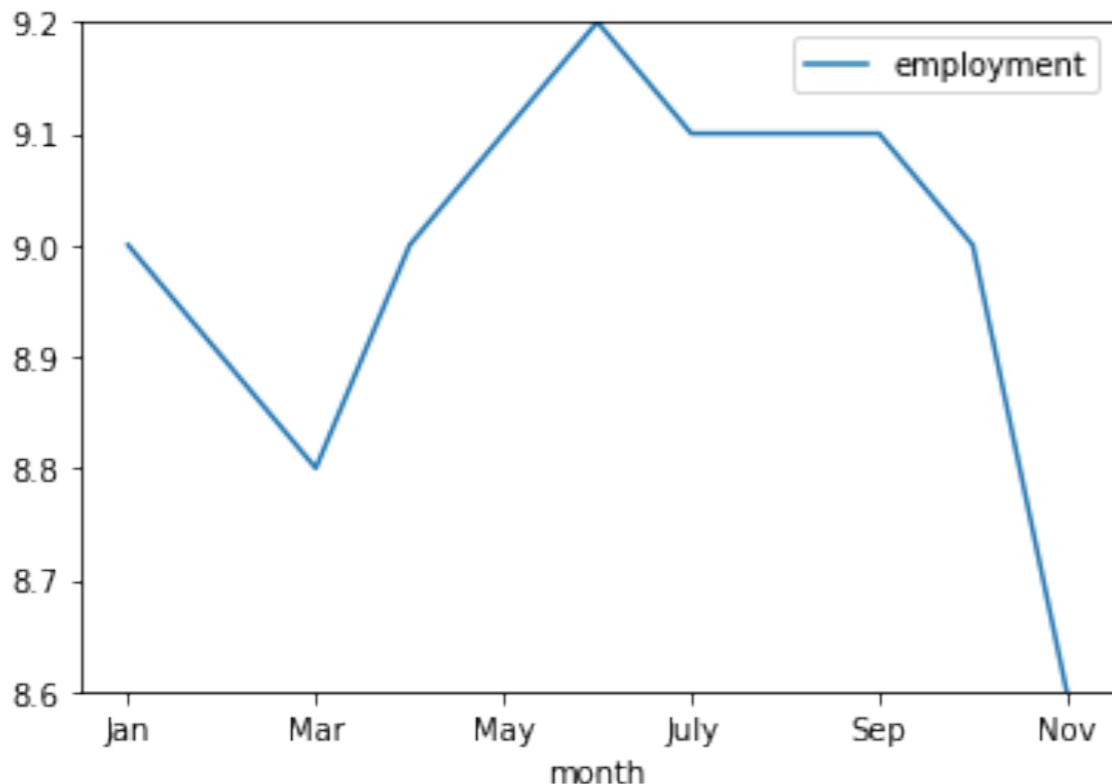
Be aware of traps in visualizing data, when creating or reading.  
Especially with scale.



## Scales

But even this is not good in theory.

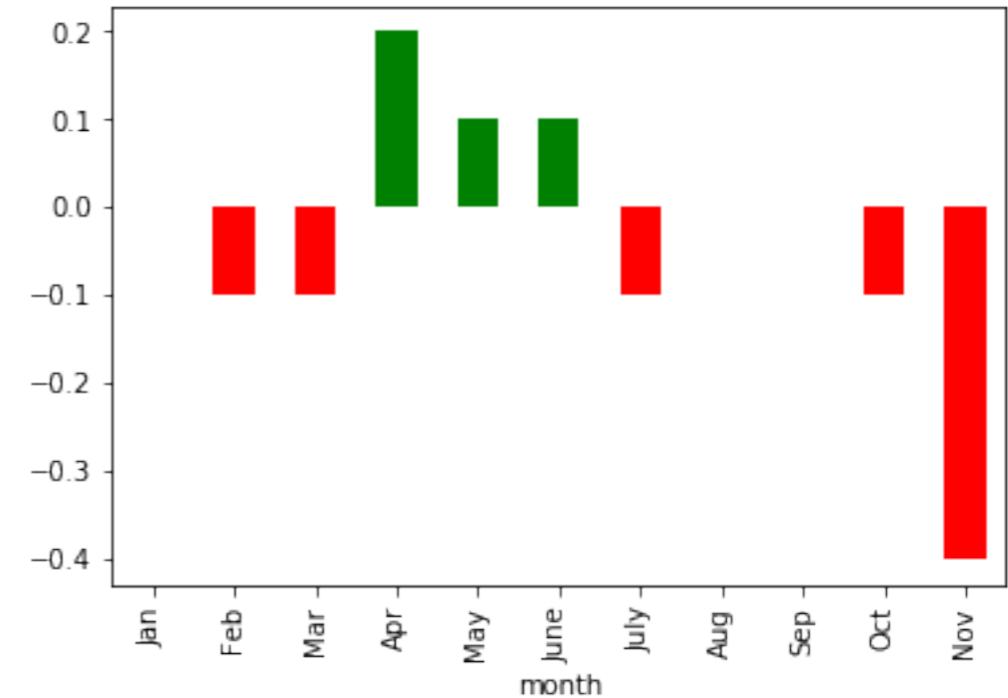
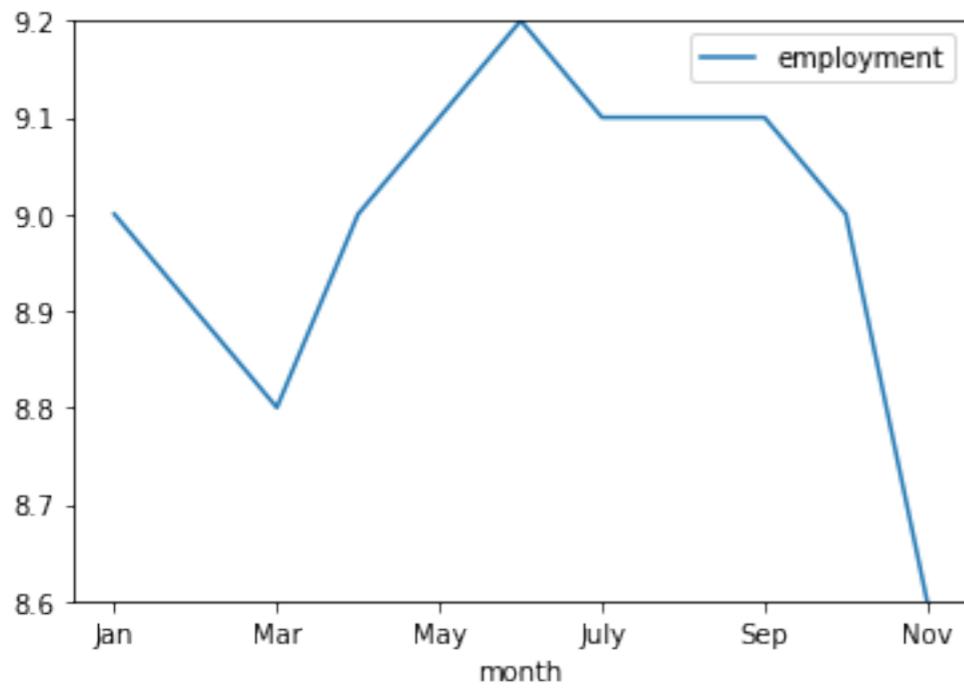
By truncating the y axis, we are magnifying the effect as well.



But having zero for the y axis makes it difficult to see change too.

# Scales

So, maybe we should think about other ways of showing change.



**So far, we've only seen how to represent a low number of dimensions**

**What happens when we have a high number of dimensions?**

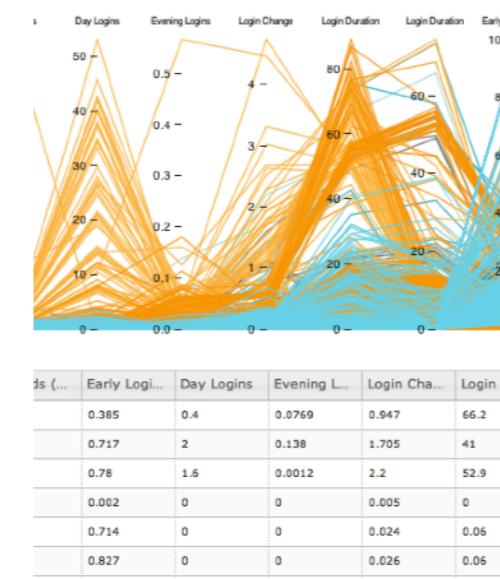
# Multidimensional Visualizations



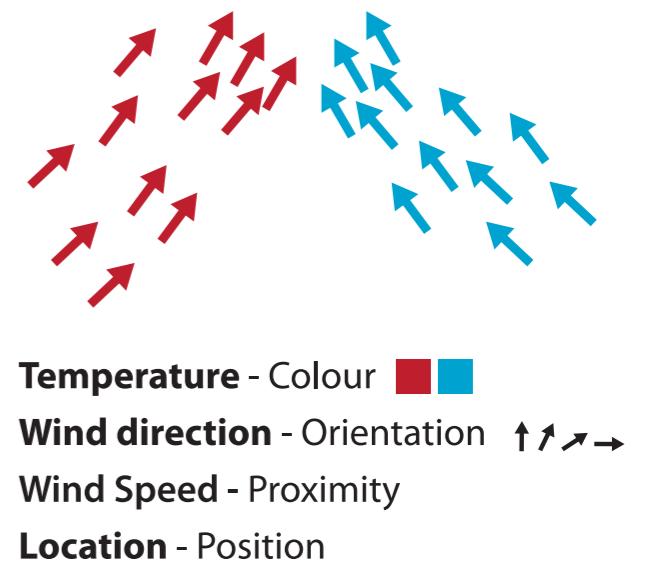
# Scatter Plot Matrices



# Linked Plots



# Parallel Coordinates



# 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

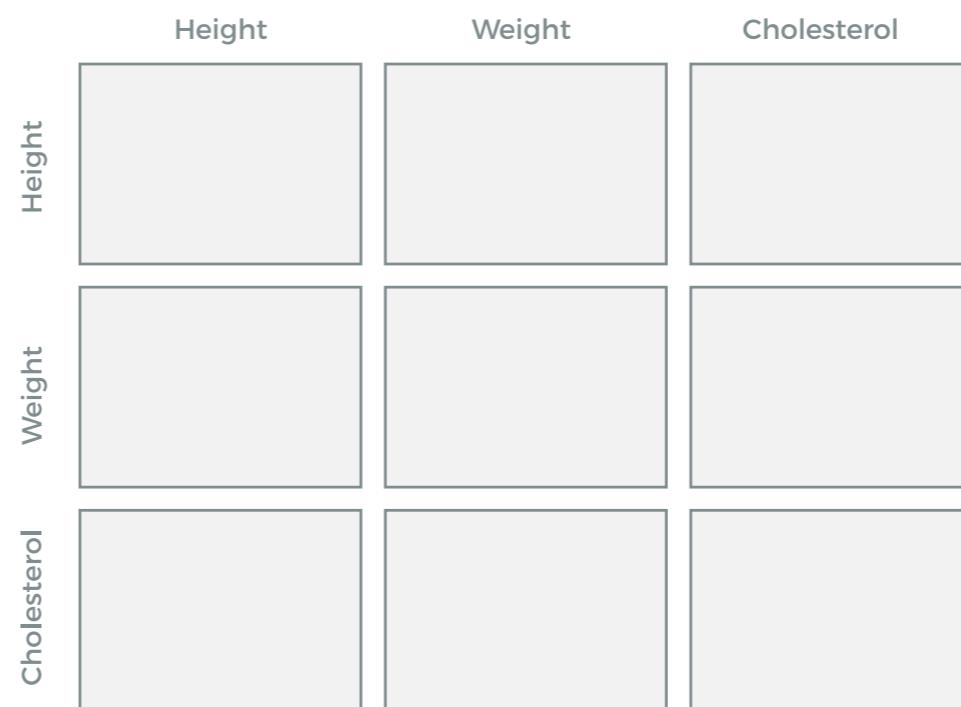
...

# 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

...

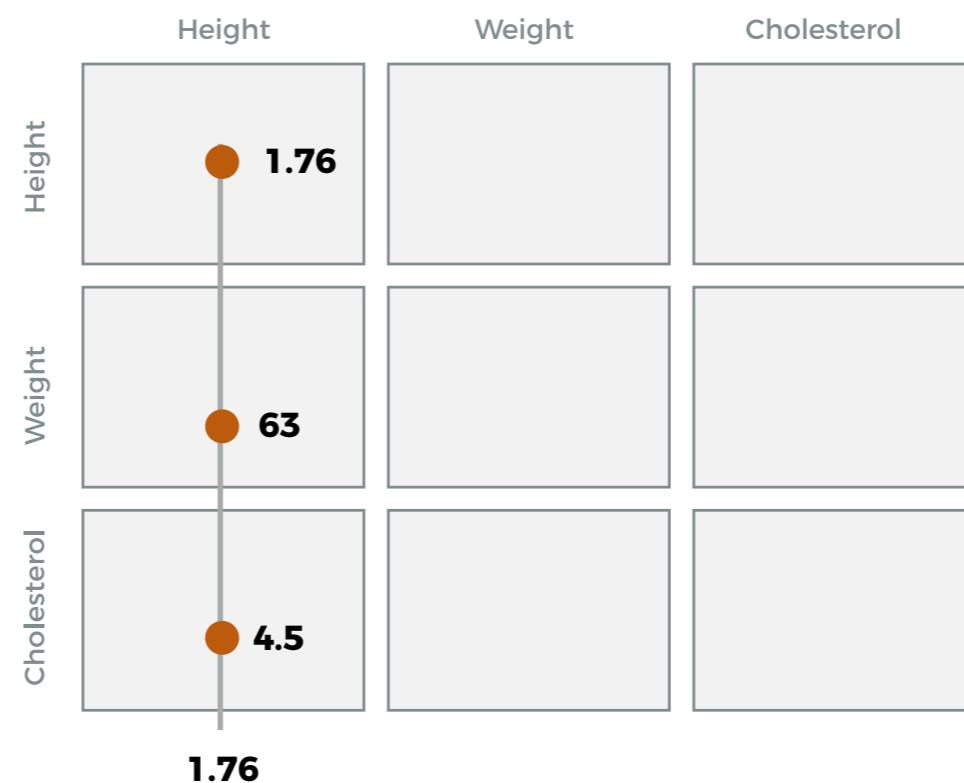


# Multidimensional Visualization

## 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

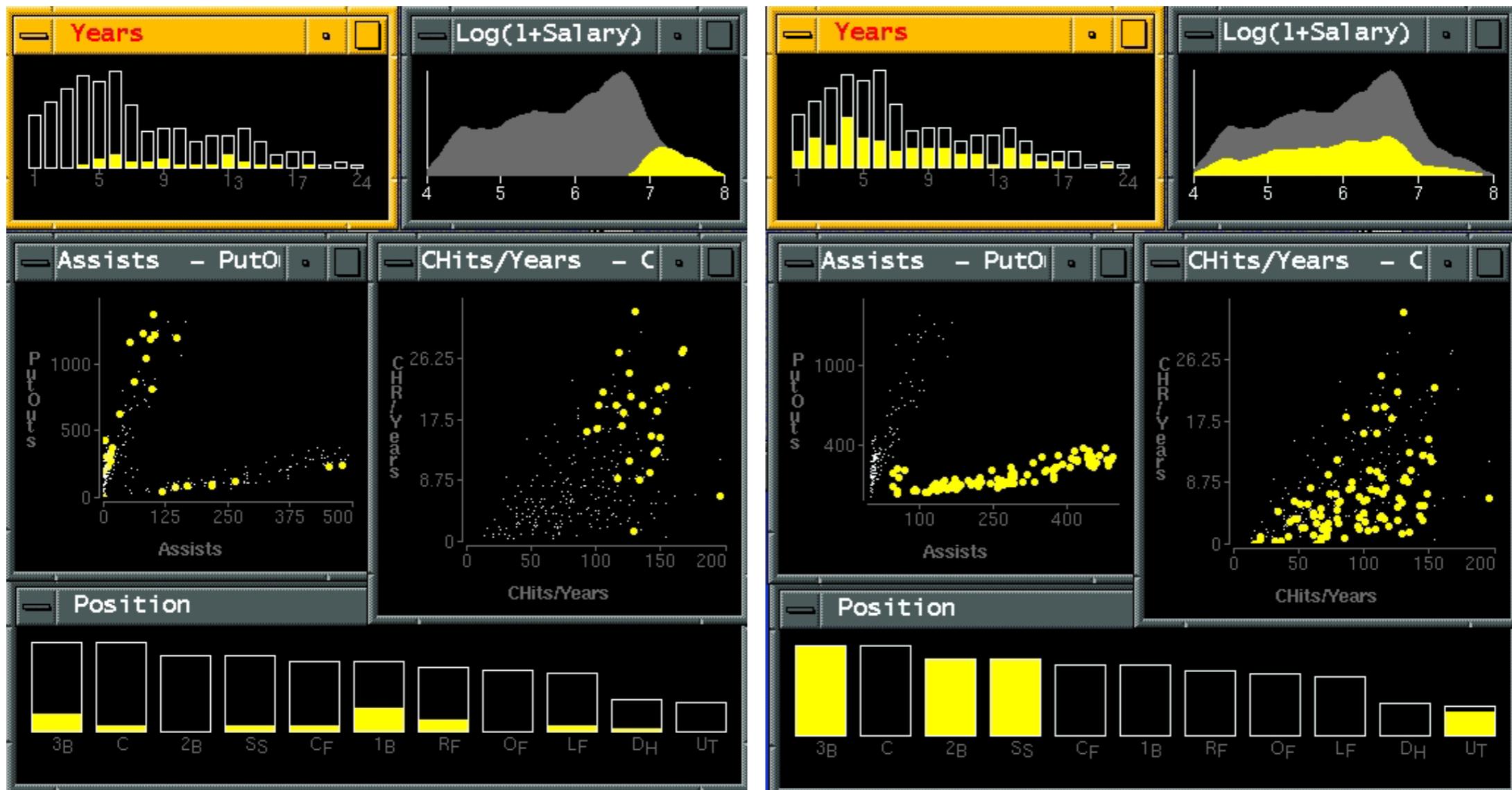
## Scatter Plot Matrices

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



# Multidimensional Visualization

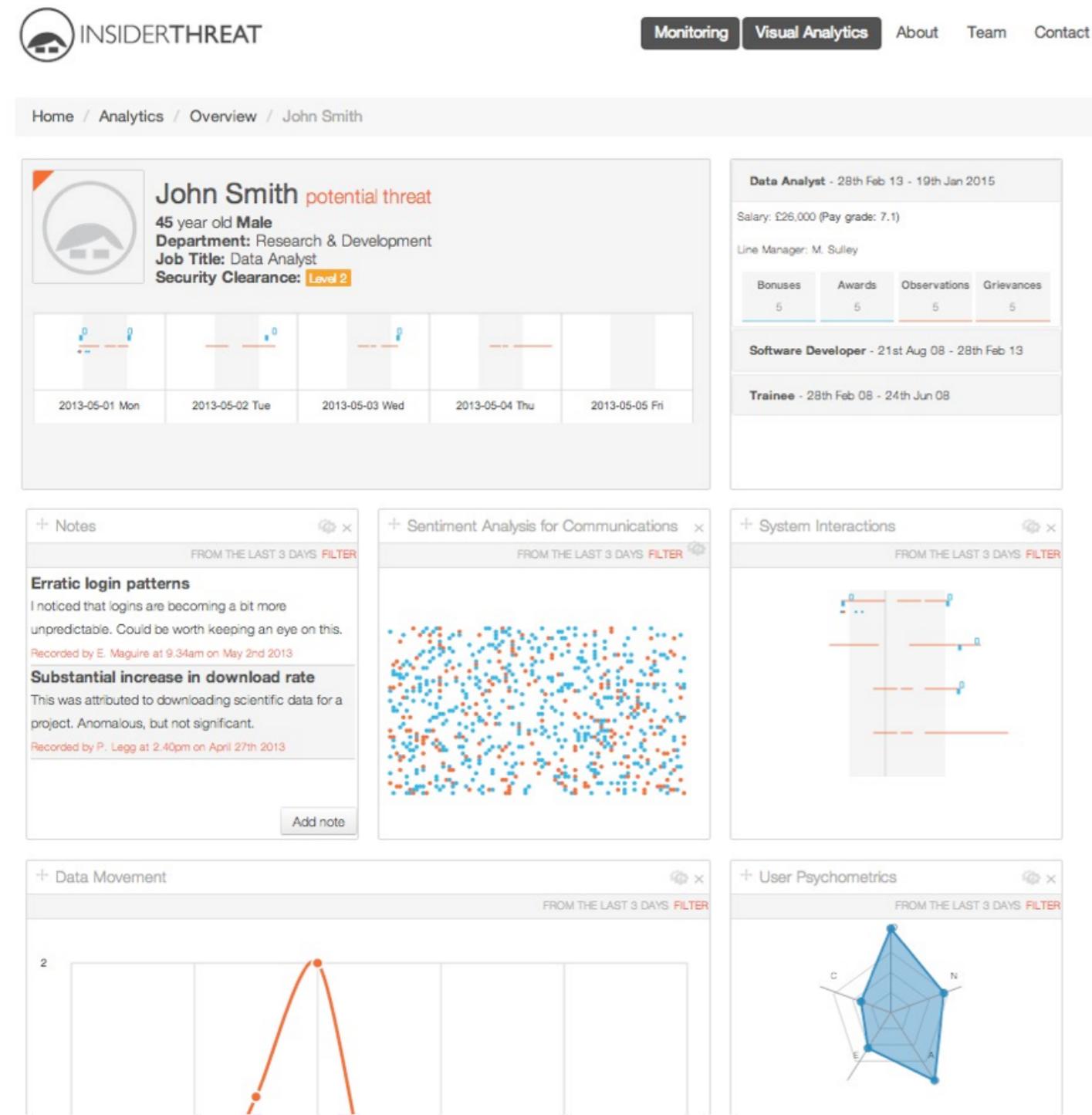
## Linked Plots



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

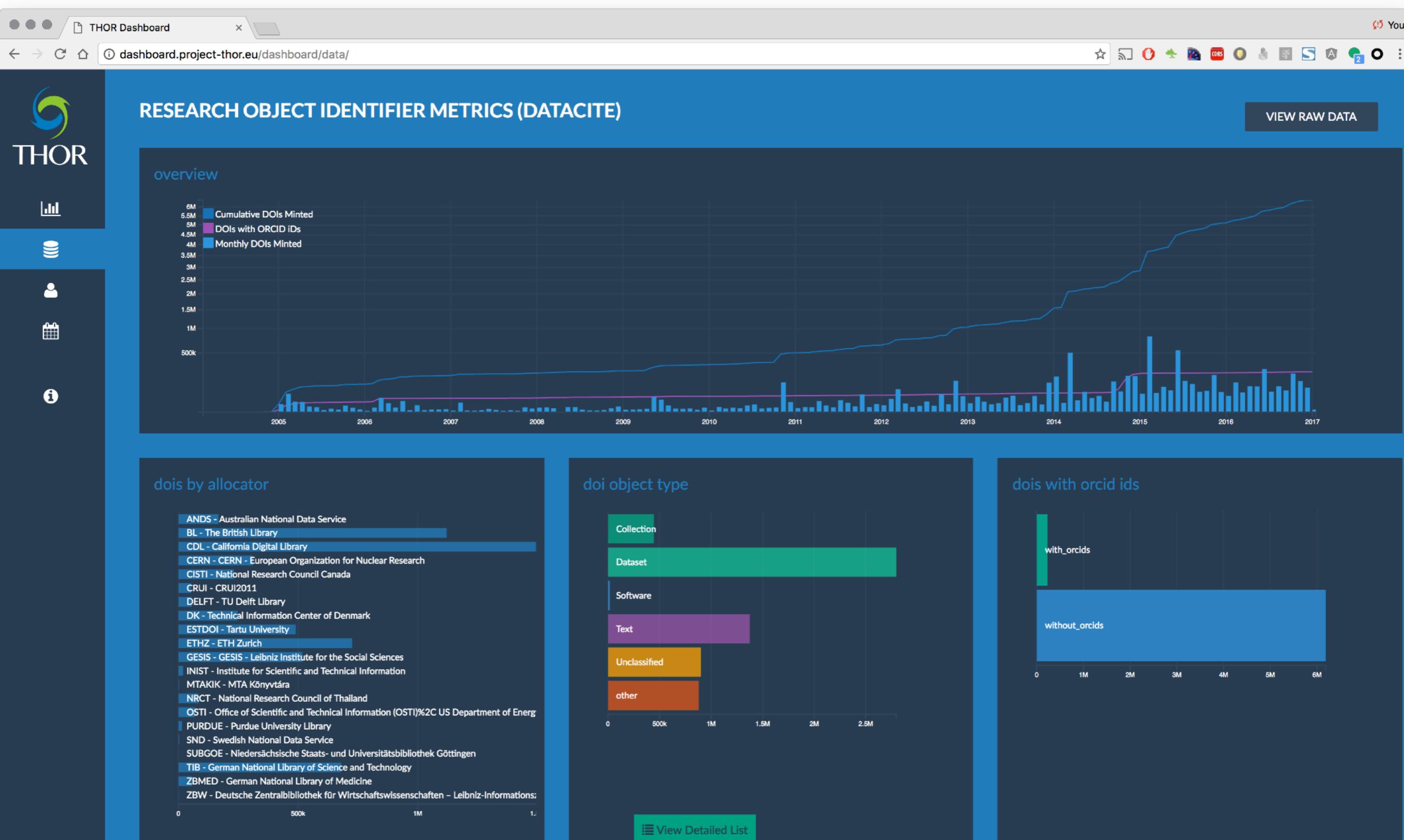
# Multidimensional Visualization

When one visualization won't cut it...



# Multidimensional Visualization

With dc.js, crossfilter, and d3.js



# Multidimensional Visualization

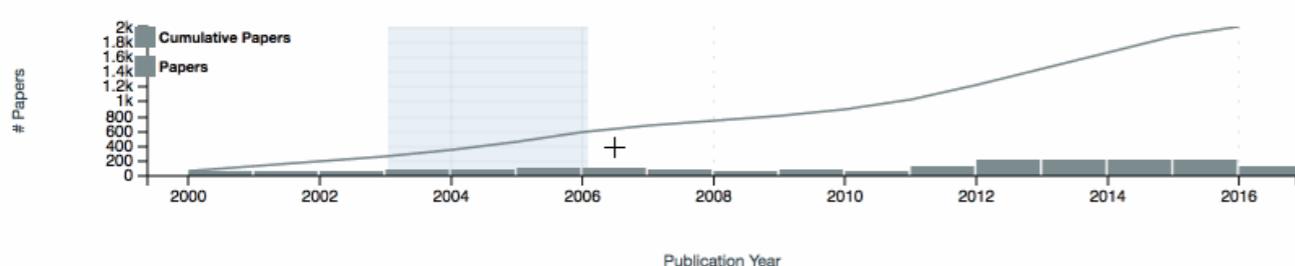
With dc.js, crossfilter, and d3.js

## Cite Summary

[Download as CSV](#)

### Paper Overview

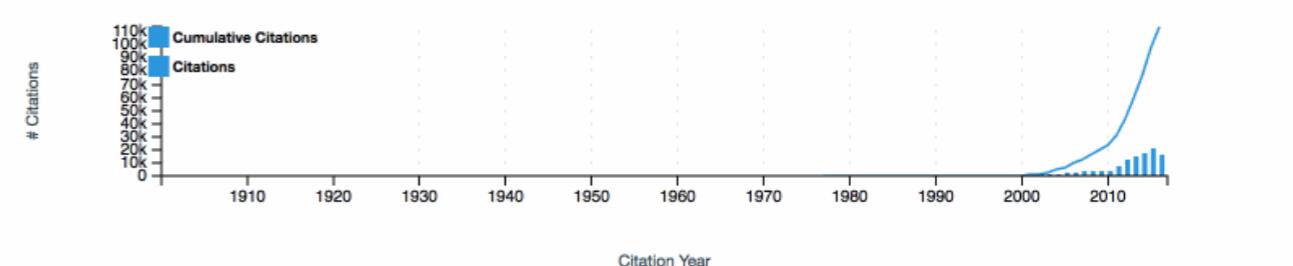
2,017 out of 2,414 Papers 794 with no citations not shown



Reset All Filters

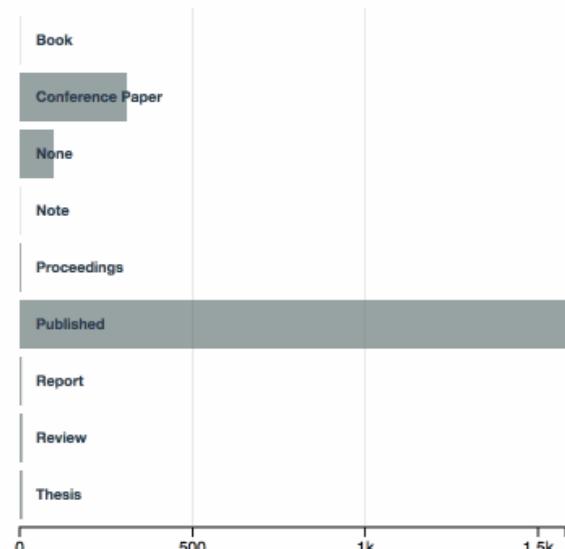
### Citation Overview

114,029 out of 114,029 Citations

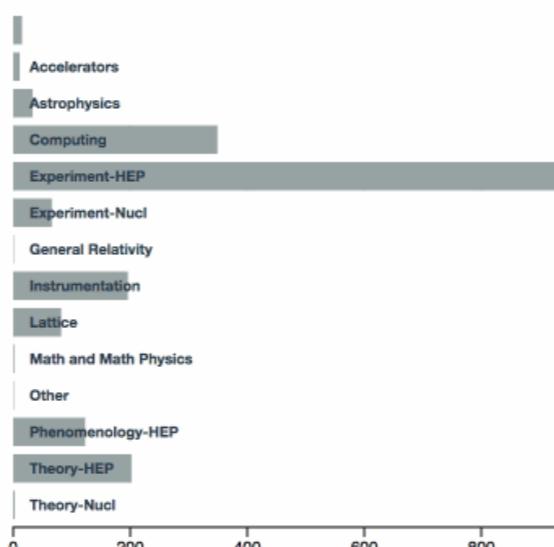


Reset All Filters

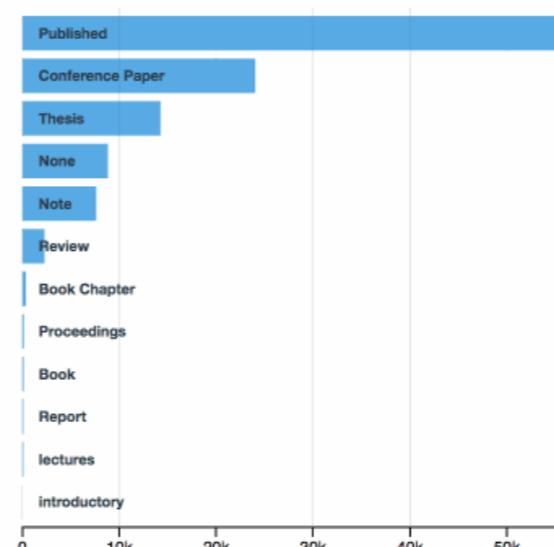
### Paper Type



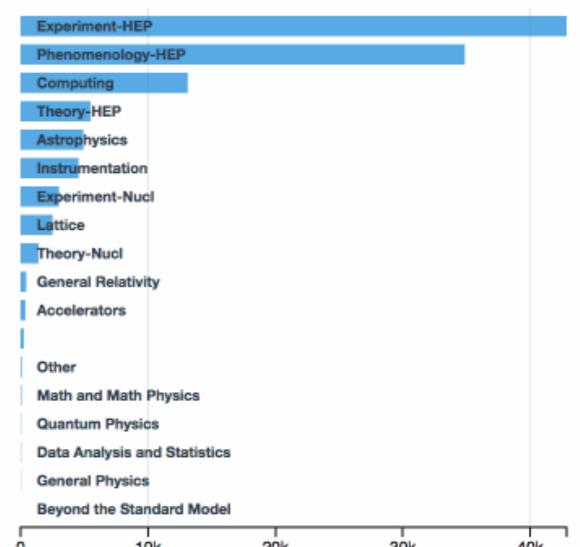
### Paper Subject Area



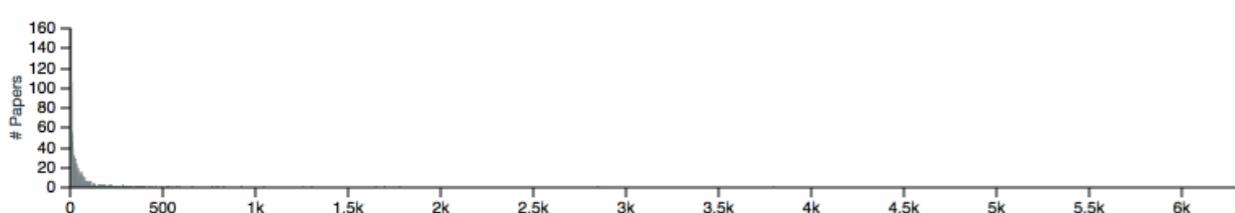
### Citation Types



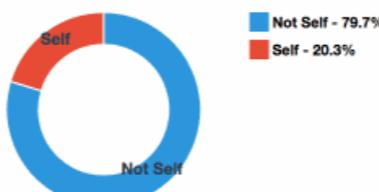
### Citation Subject Area



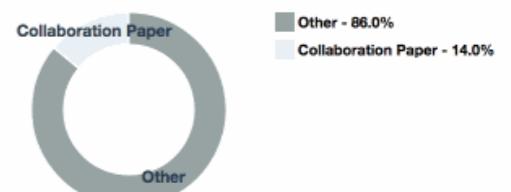
### Paper Citation Counts



### Self Citations



### Collaboration Papers



# Multidimensional Visualization

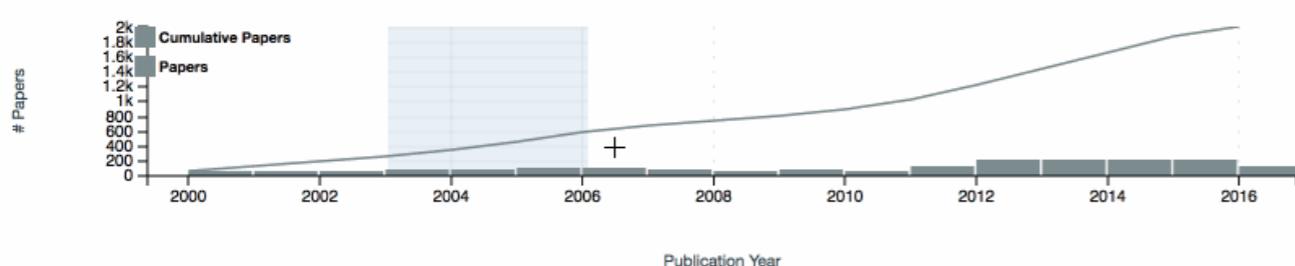
With dc.js, crossfilter, and d3.js

## Cite Summary

[Download as CSV](#)

### Paper Overview

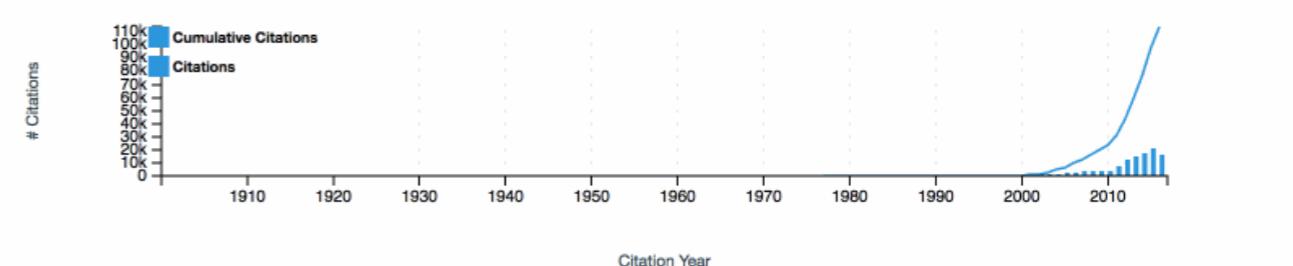
2,017 out of 2,414 Papers 794 with no citations not shown



Reset All Filters

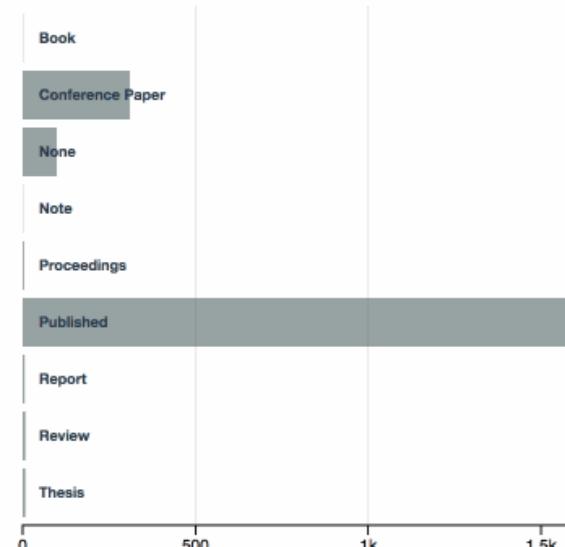
### Citation Overview

114,029 out of 114,029 Citations

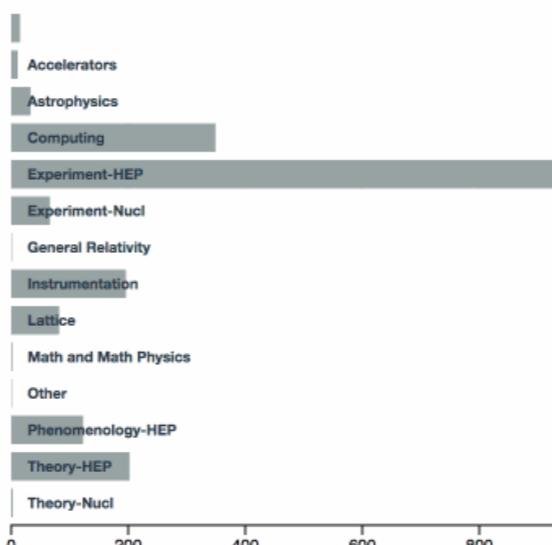


Reset All Filters

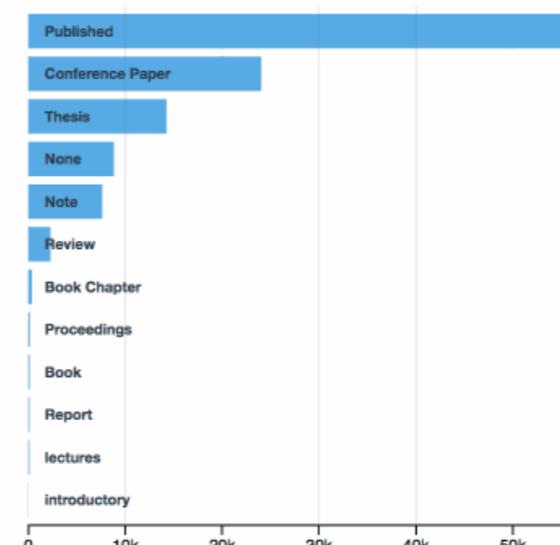
### Paper Type



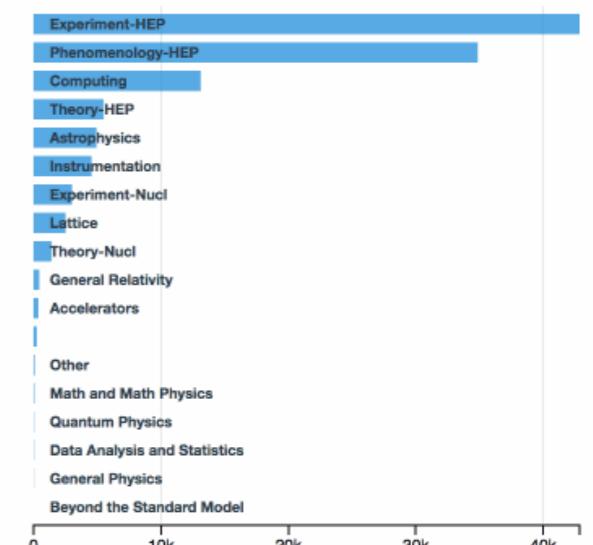
### Paper Subject Area



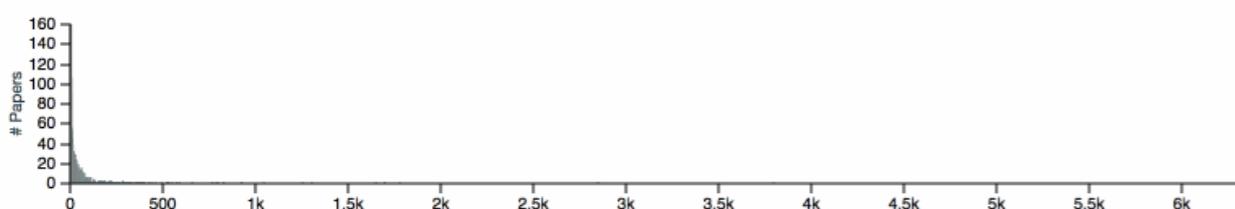
### Citation Types



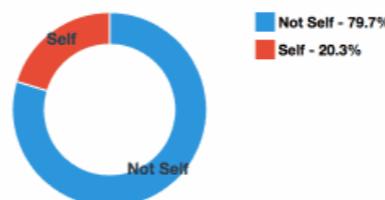
### Citation Subject Area



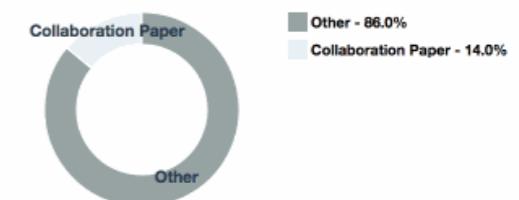
### Paper Citation Counts



### Self Citations

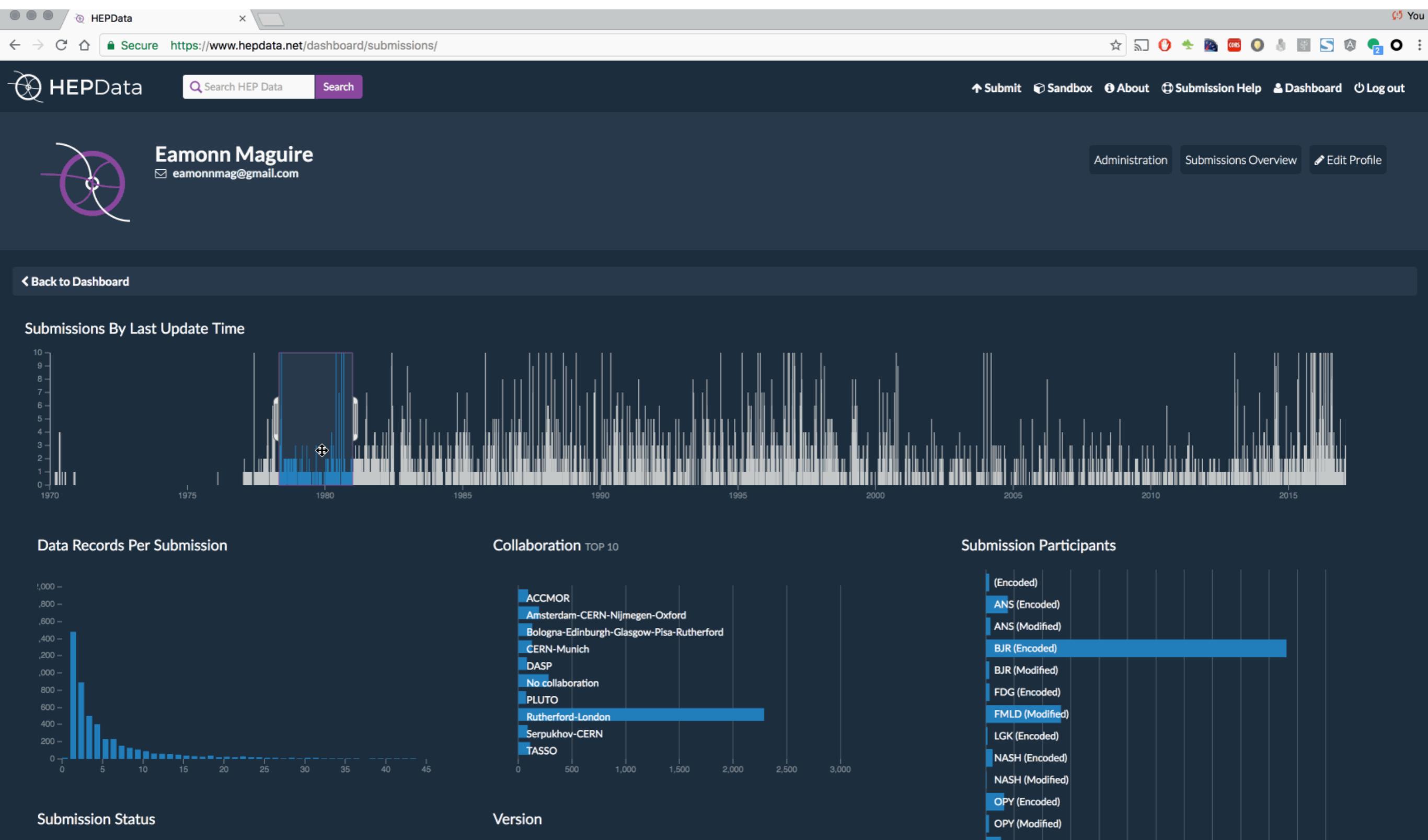


### Collaboration Papers



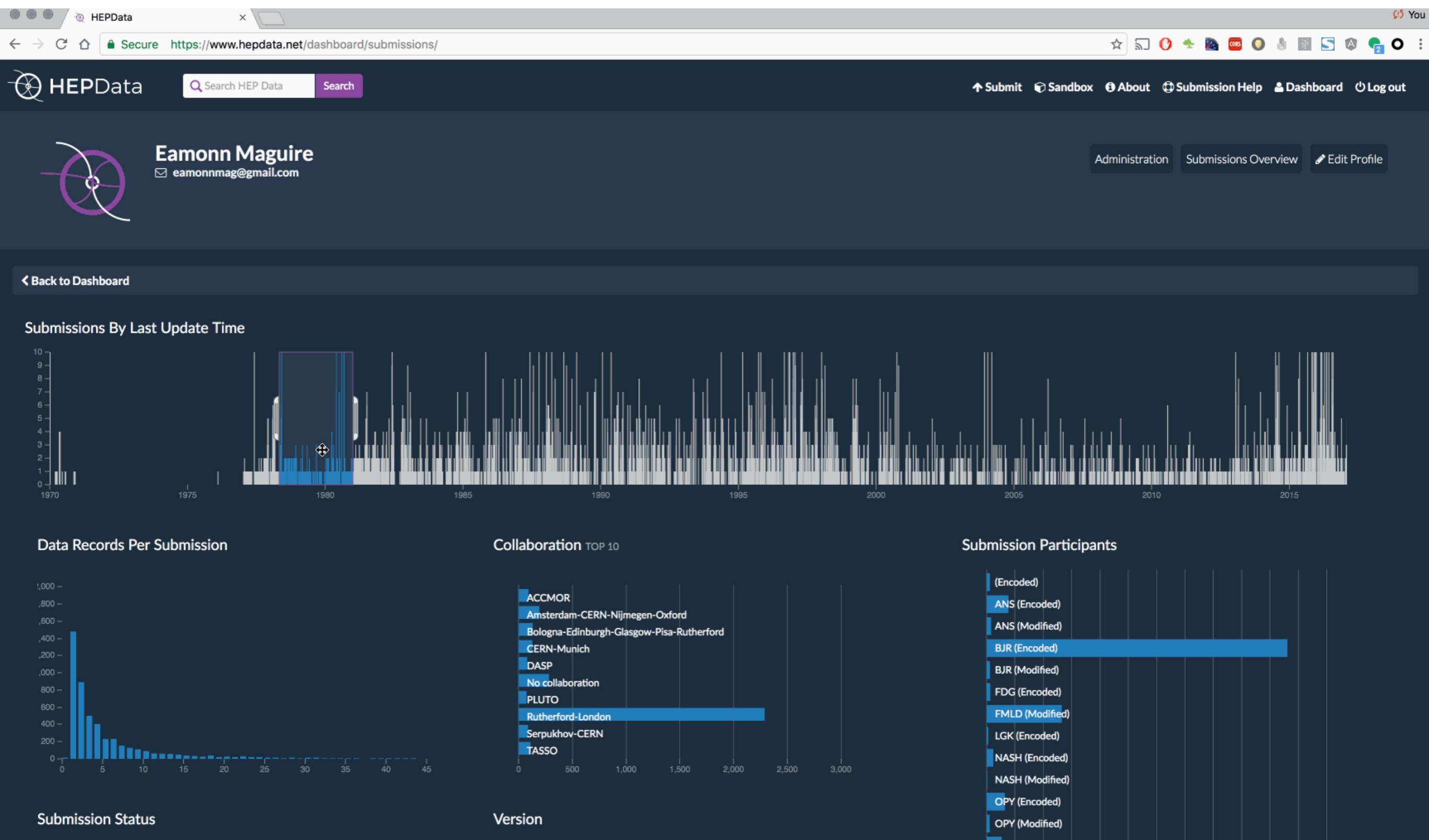
# Multidimensional Visualization

With dc.js, crossfilter, and d3.js



# Multidimensional Visualization

With dc.js, crossfilter, and d3.js



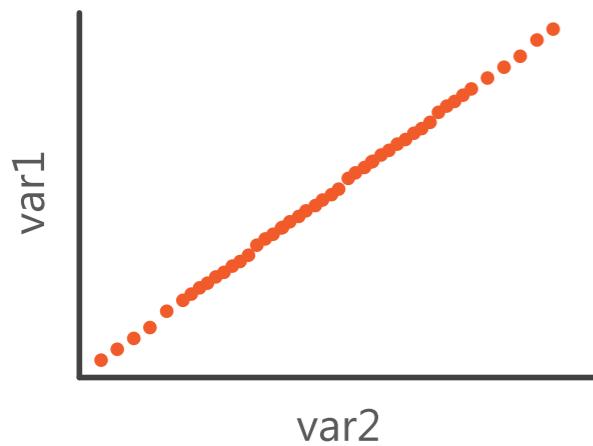
**My Tutorial on Creating Dashboard Visualizations**

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

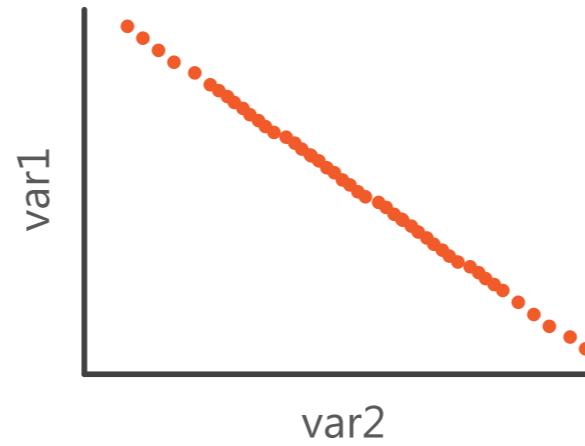
# Multidimensional Visualization

## Parallel Coordinate Plots

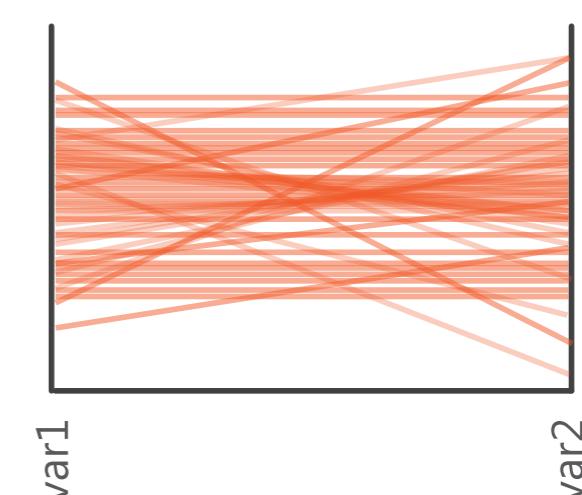
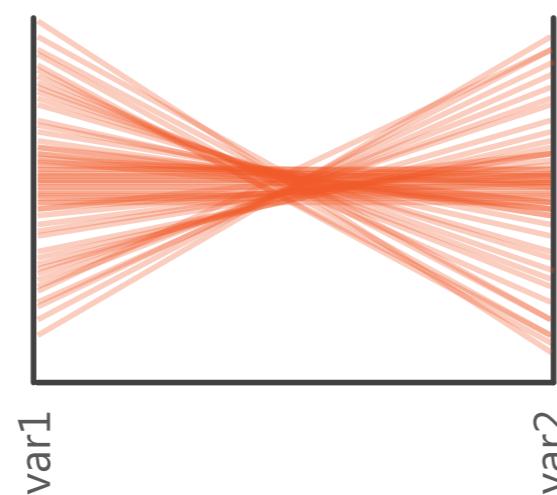
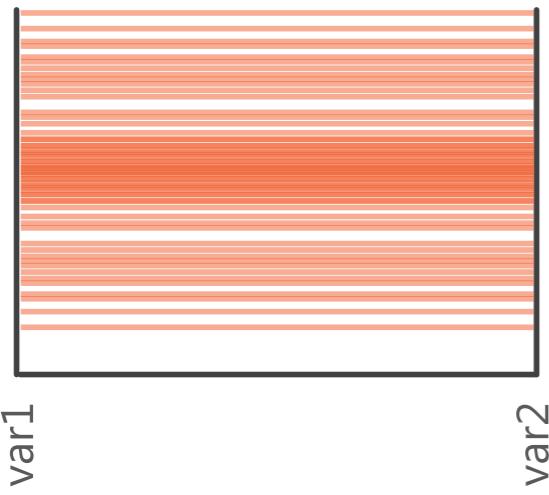
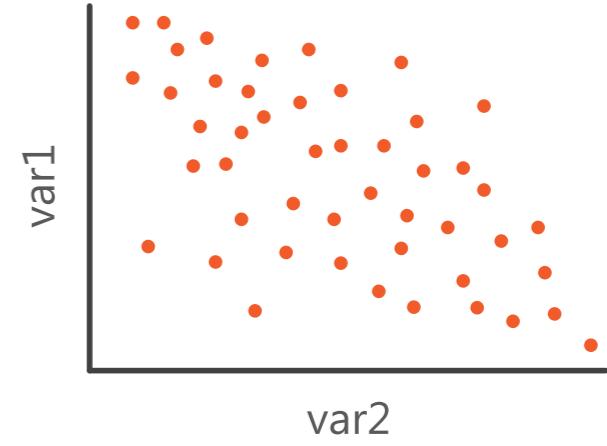
Positive Correlation



Negative (inverse) Correlation

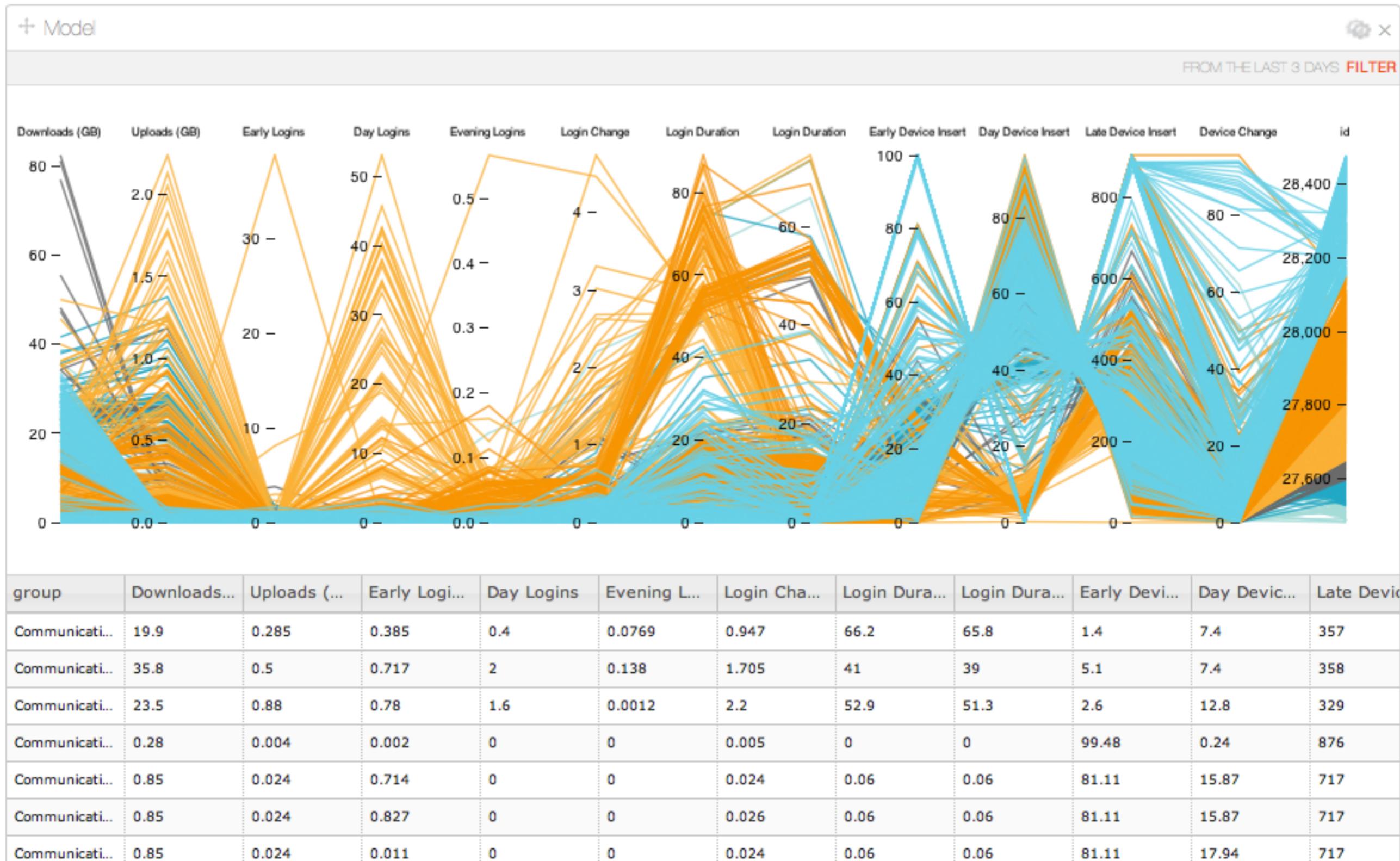


No Correlation



# Multidimensional Visualization

# Parallel Coordinate Plots

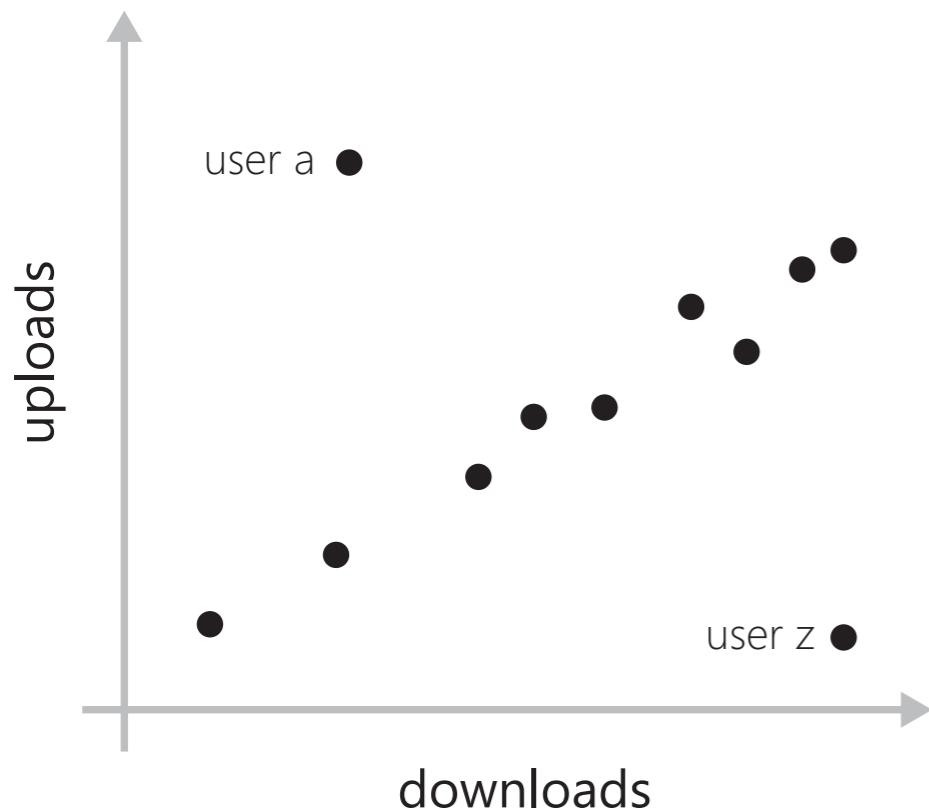


# Parallel Coordinate Plots

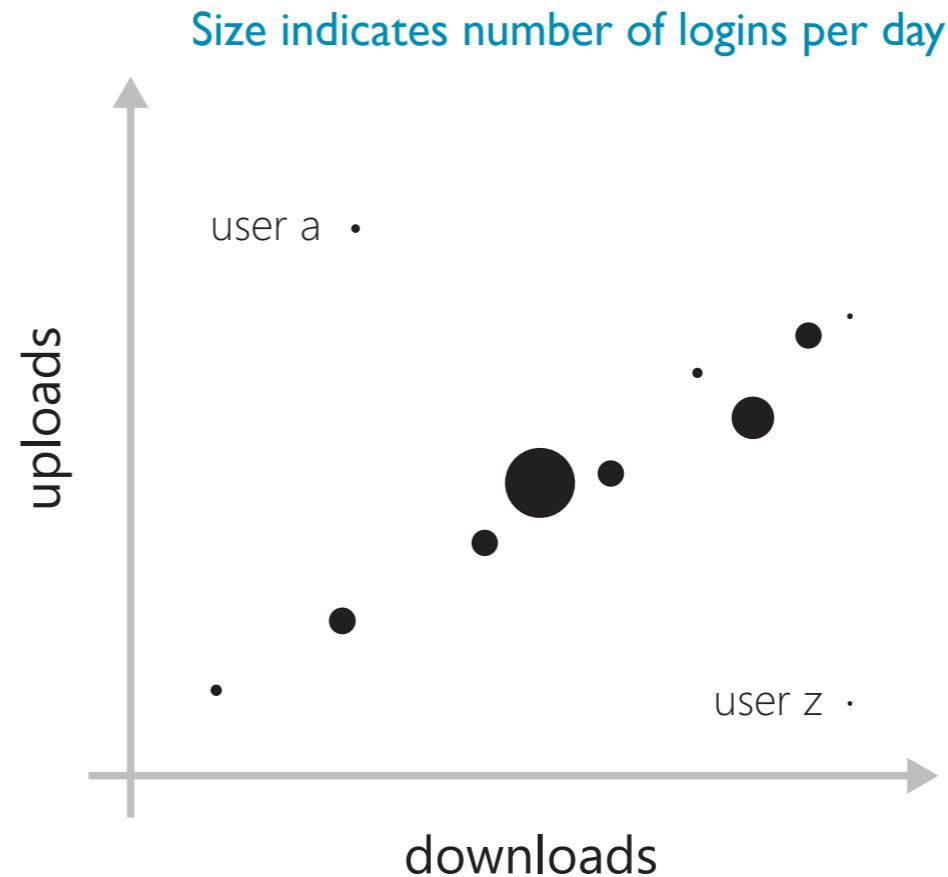
Lets take an example where we have many variables to display...

Each user is represented by a circle

2 Dimensions



3 Dimensions

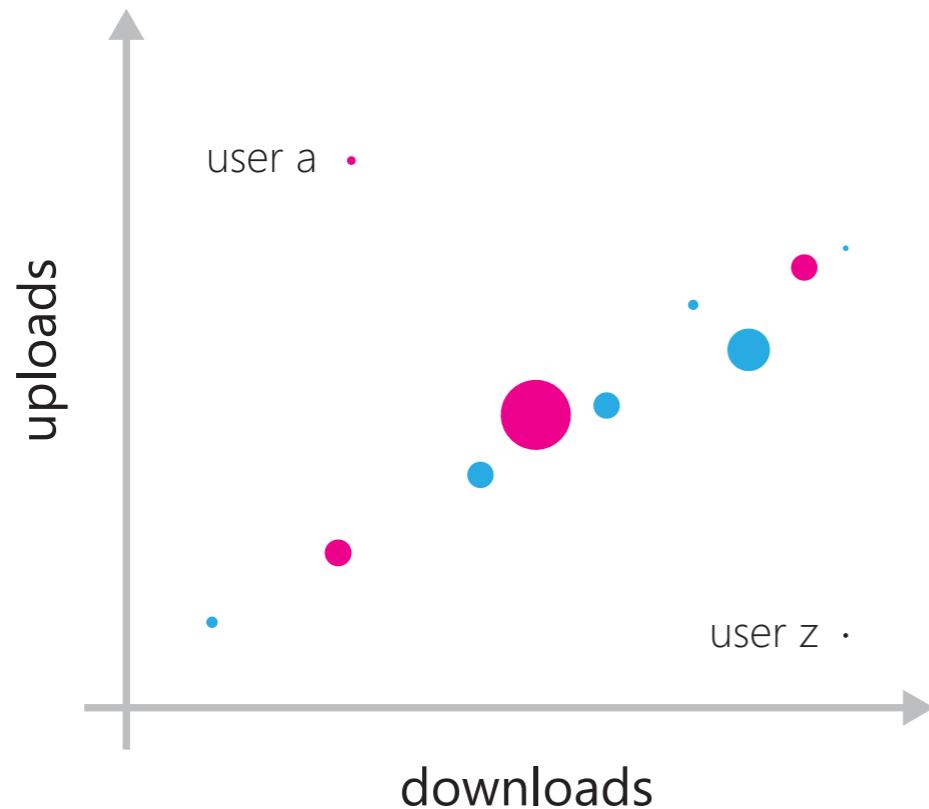


Size indicates number of logins per day

# Parallel Coordinate Plots

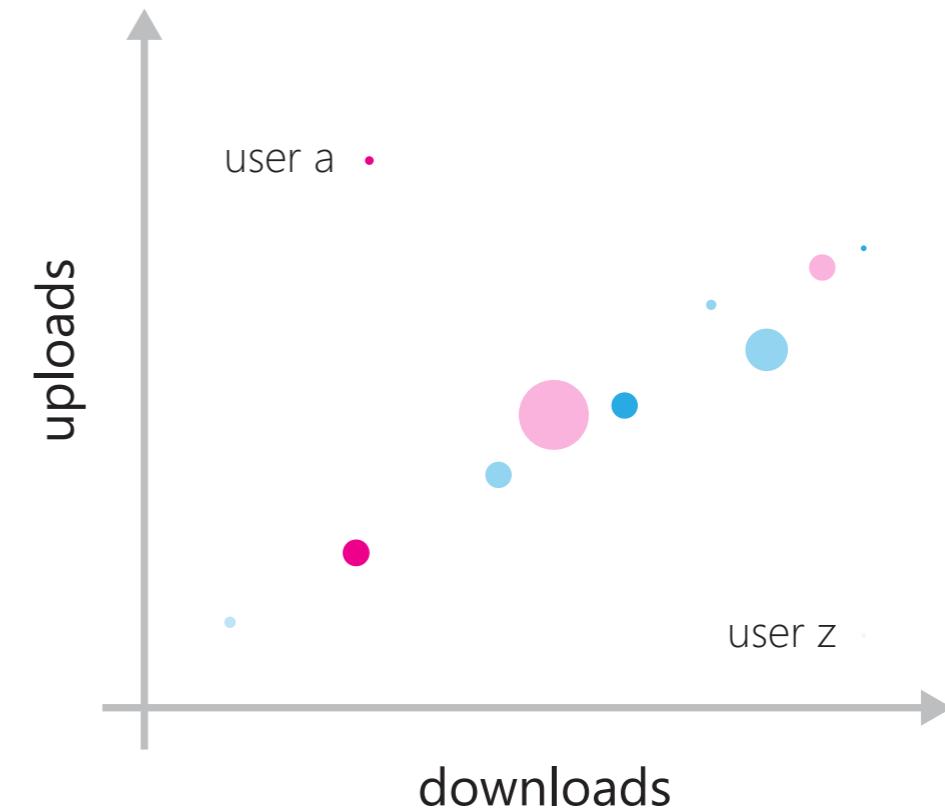
4 Dimensions

Color indicates users department



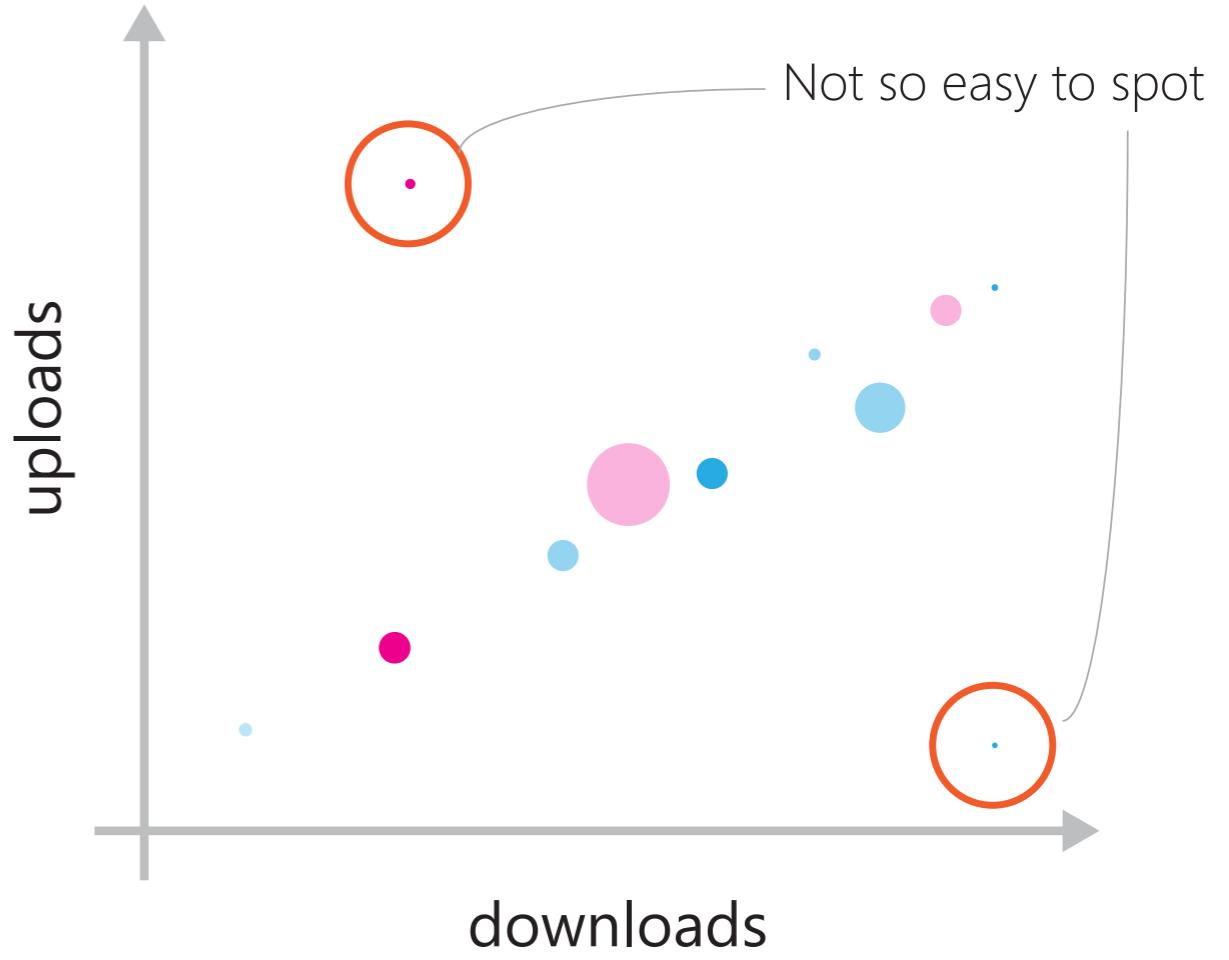
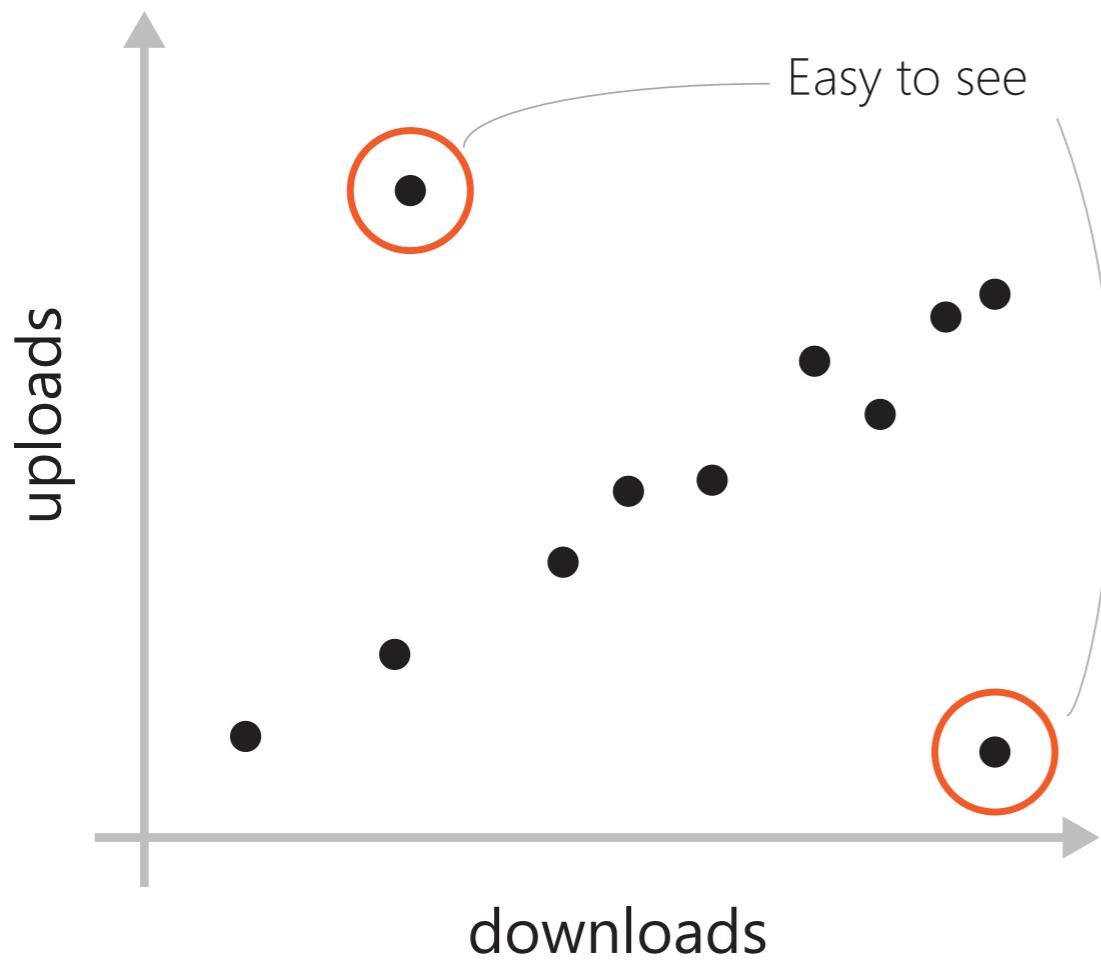
5 Dimensions

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 Coordinate Plots

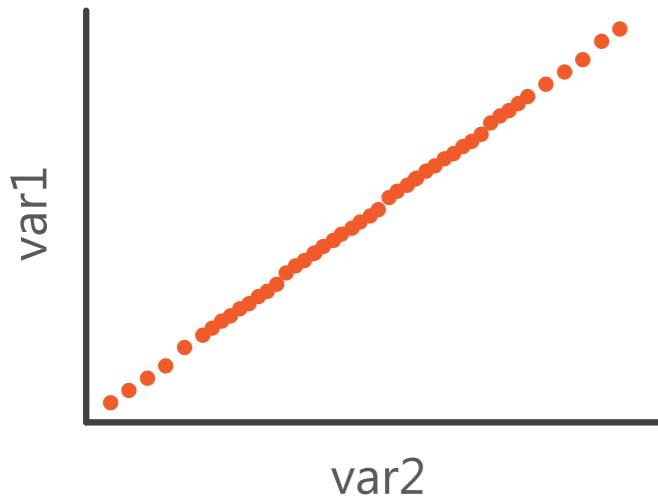


Parallel coordinates are a visualization technique employed when a large number of dimensions need to be displayed (often without a temporal element) and where each of those dimensions can be equally important in the decision making process.

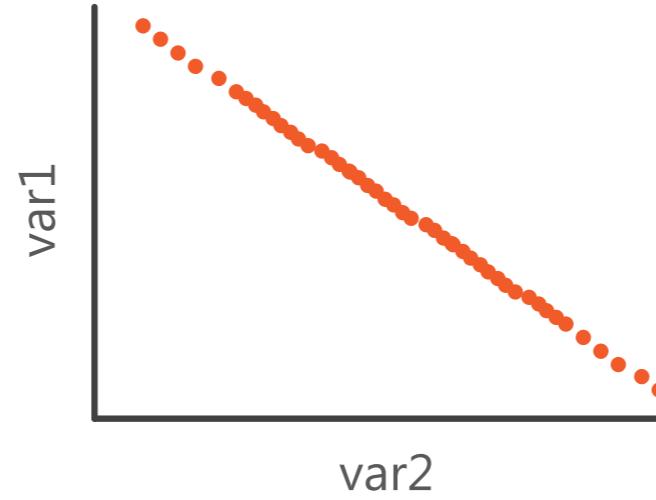
In the scatter plots here, it's easy to see **correlation** between downloads and uploads, but with the other dimensions that's difficult.

# Parallel Coordinate Plots

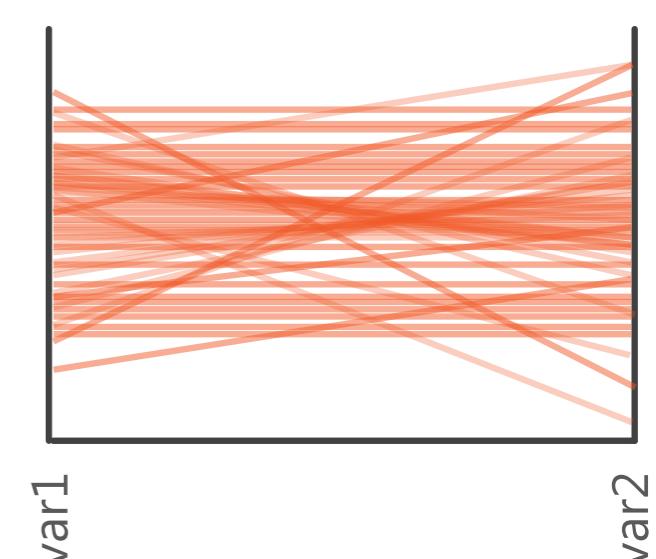
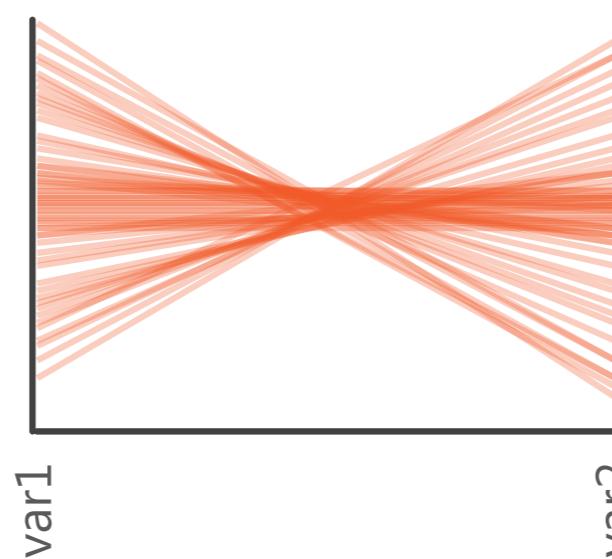
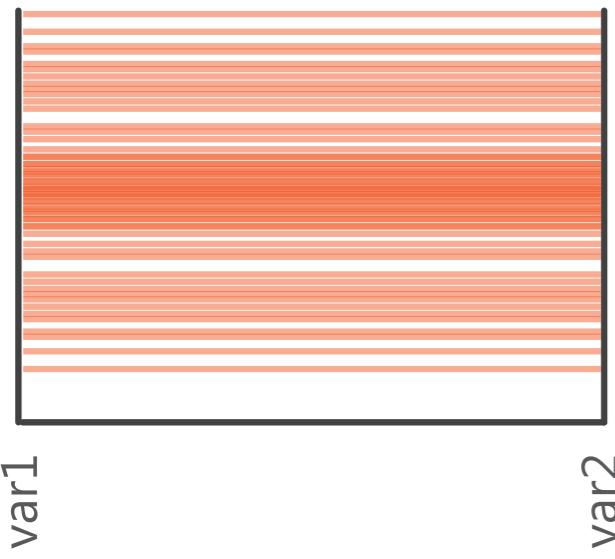
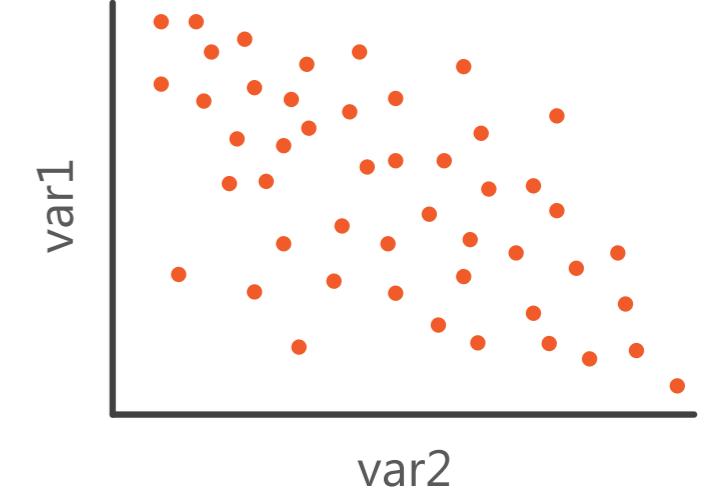
Positive Correlation



Negative (inverse) Correlation



No Correlation



# Parallel Coordinate Plots

2 Dimensions

[Uploads](#)

[Downloads](#)



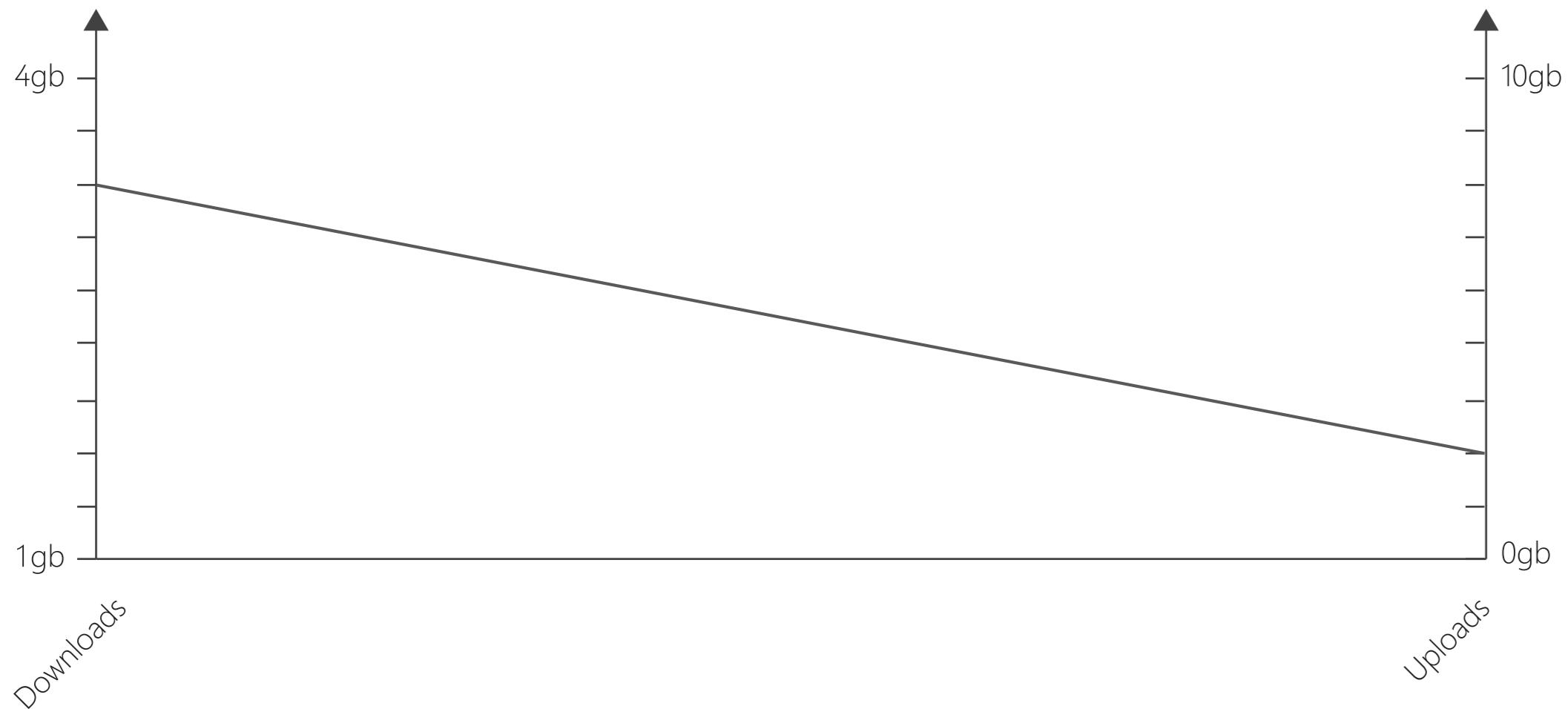
# Parallel Coordinate Plots

2 Dimensions

Uploads

1 User

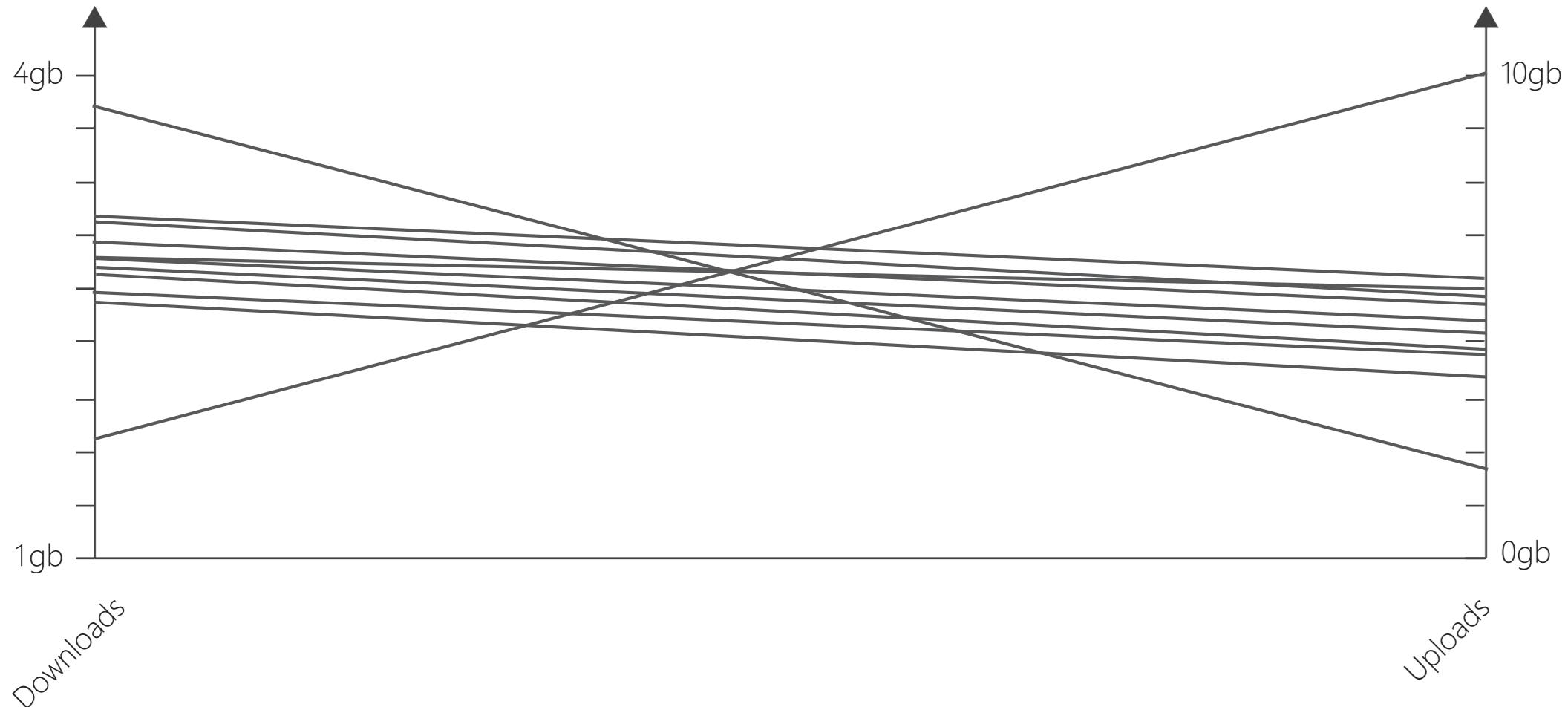
Downloads



# Parallel Coordinate Plots

2 Variables

Uploads      11 Users  
Downloads

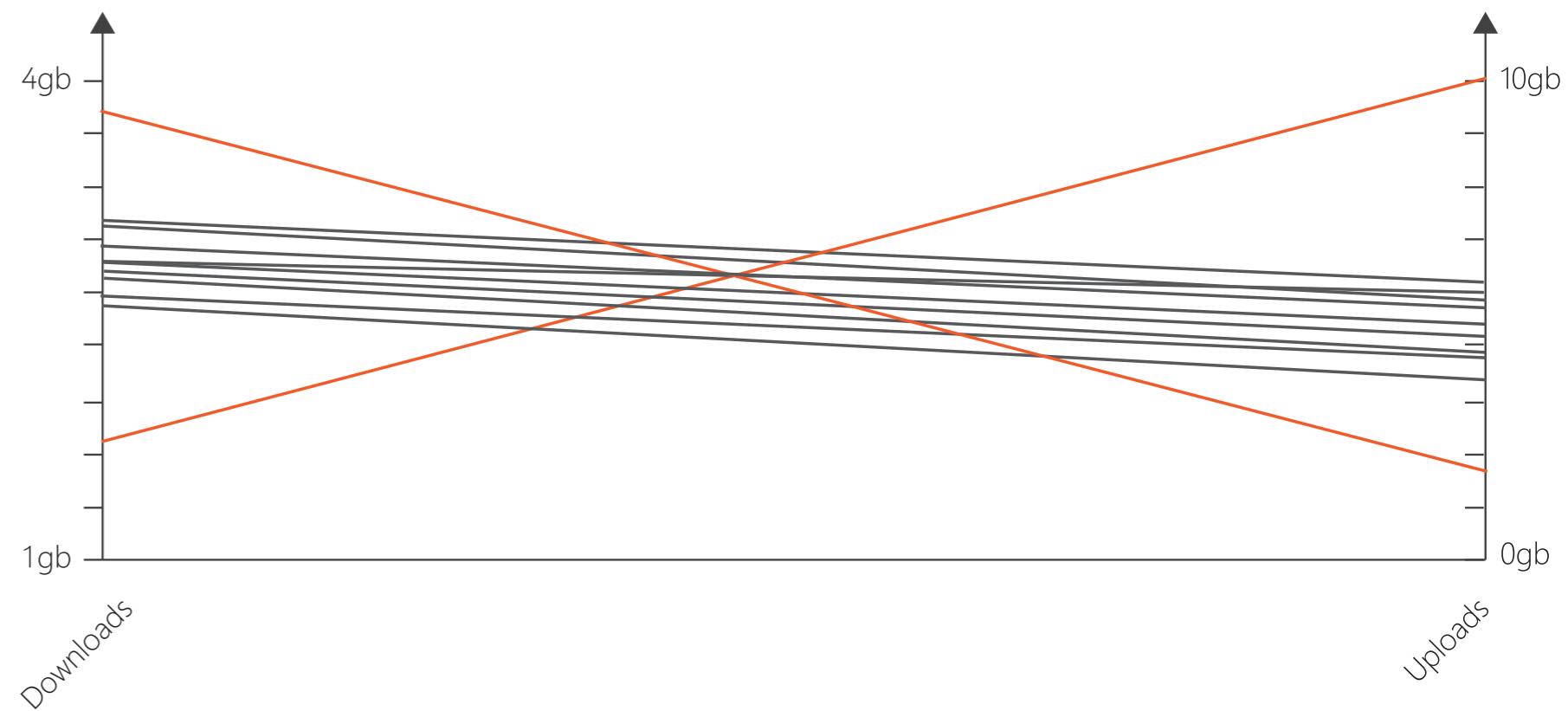
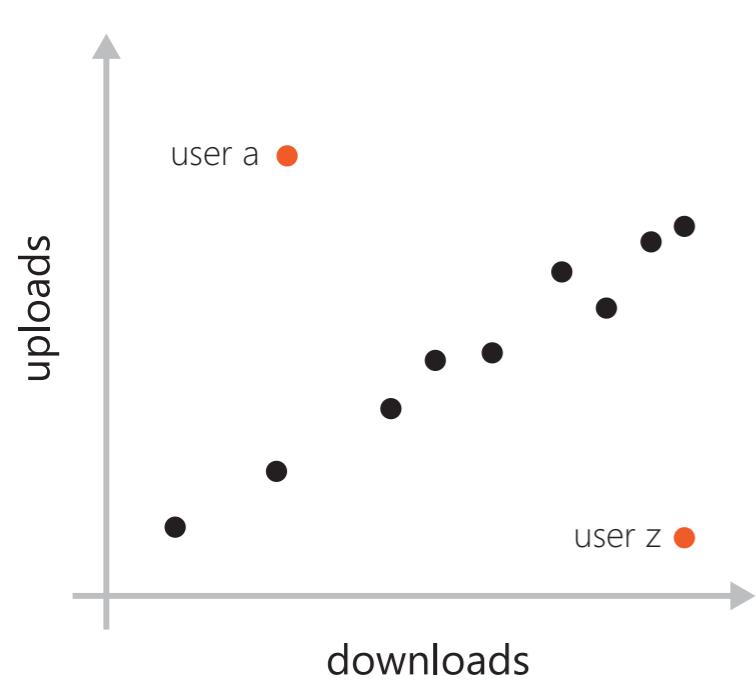


# Parallel Coordinate Plots

2 Dimensions

Uploads      11 Users

Downloads



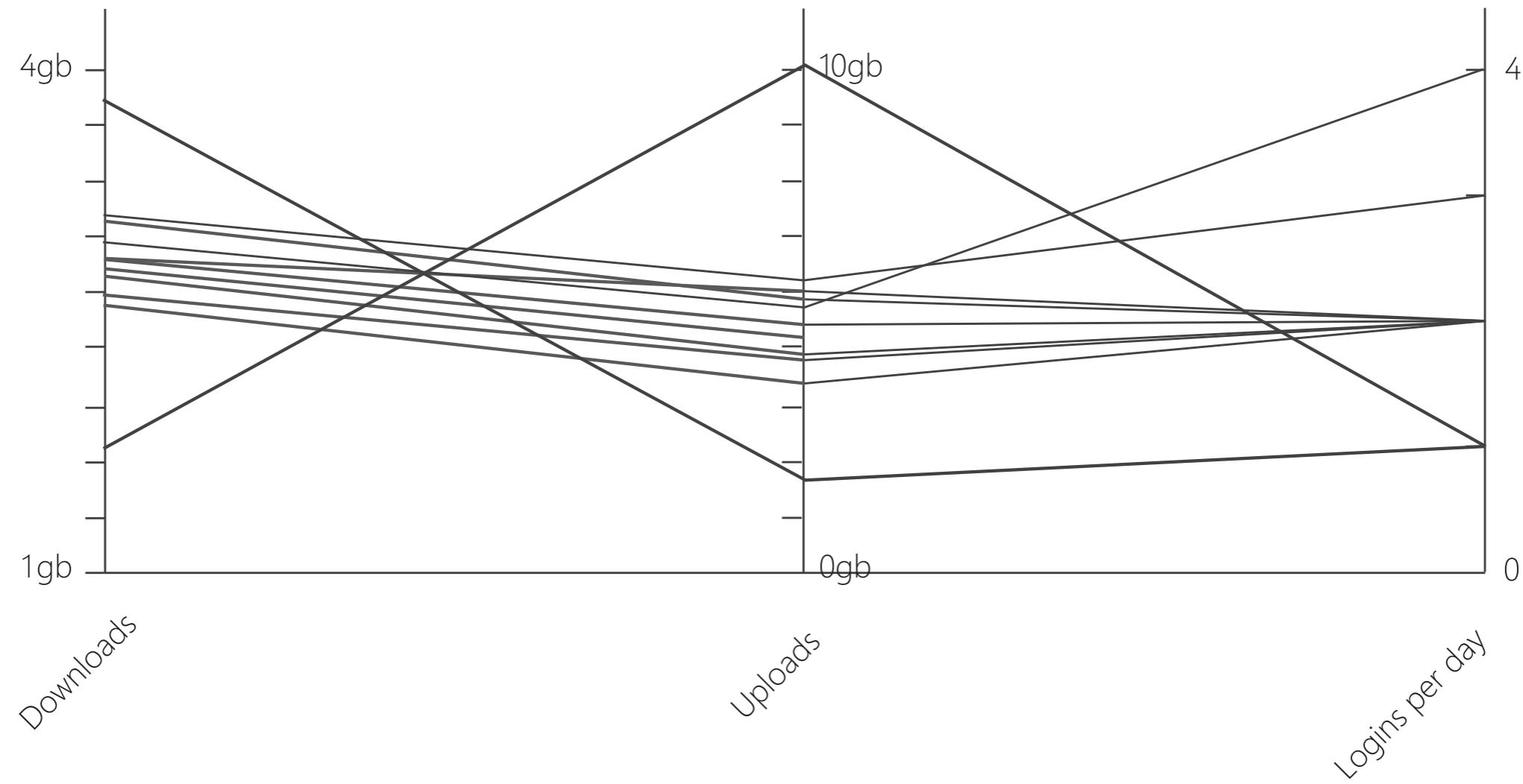
# Parallel Coordinate Plots

3 Dimensions

Uploads      11 Users

Downloads

Logins per day



# Parallel Coordinate Plots

4 Dimensions

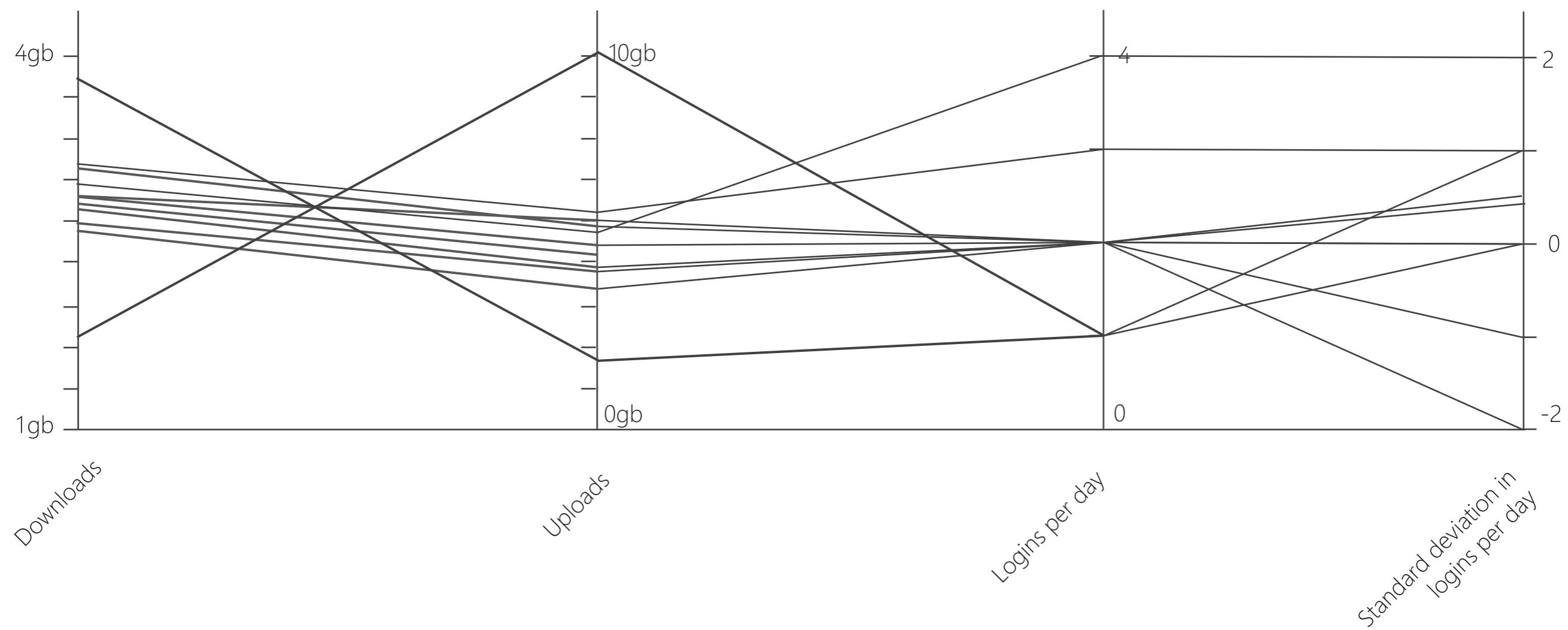
Uploads

11 Users

Downloads

Logins per day

Std. deviation in logins per day



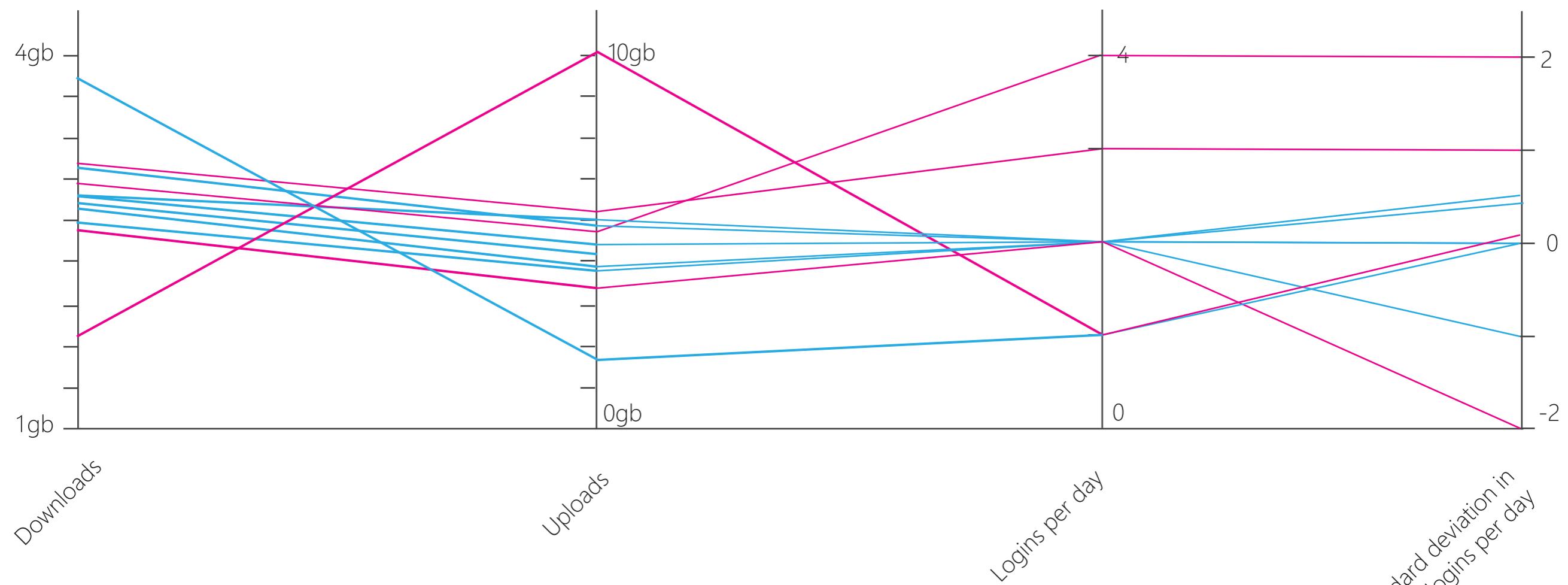
# Parallel Coordinate Plots

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.

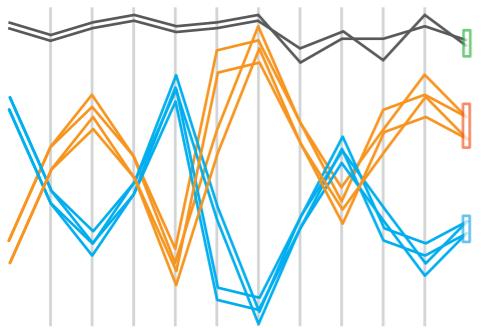


We can keep adding more parallel lines, and comfortably have around 20 dimensions for many users displayed at once.

# Parallel Coordinate Plots

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

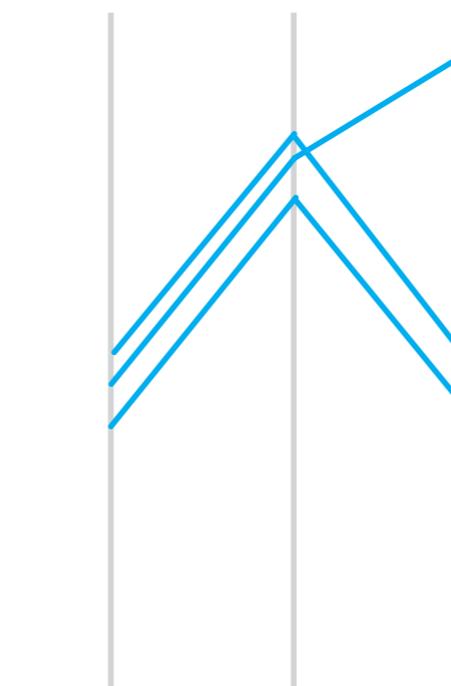
Clustering



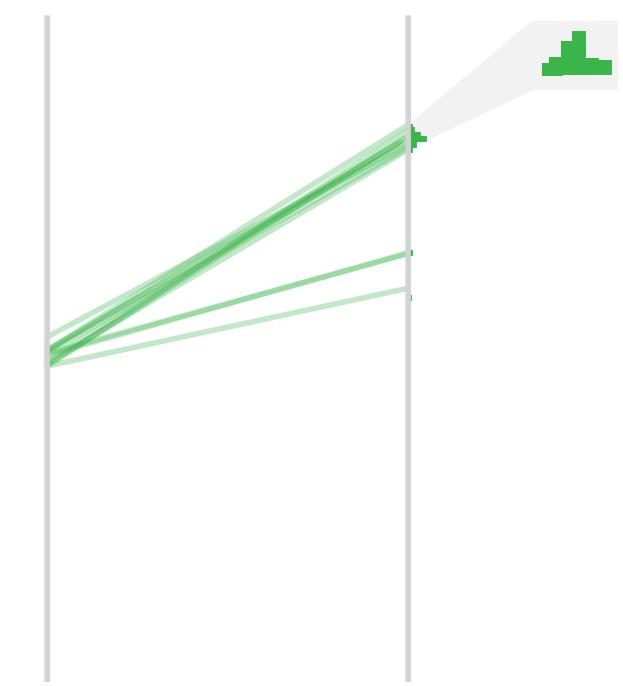
Correlation



Outlier Detection



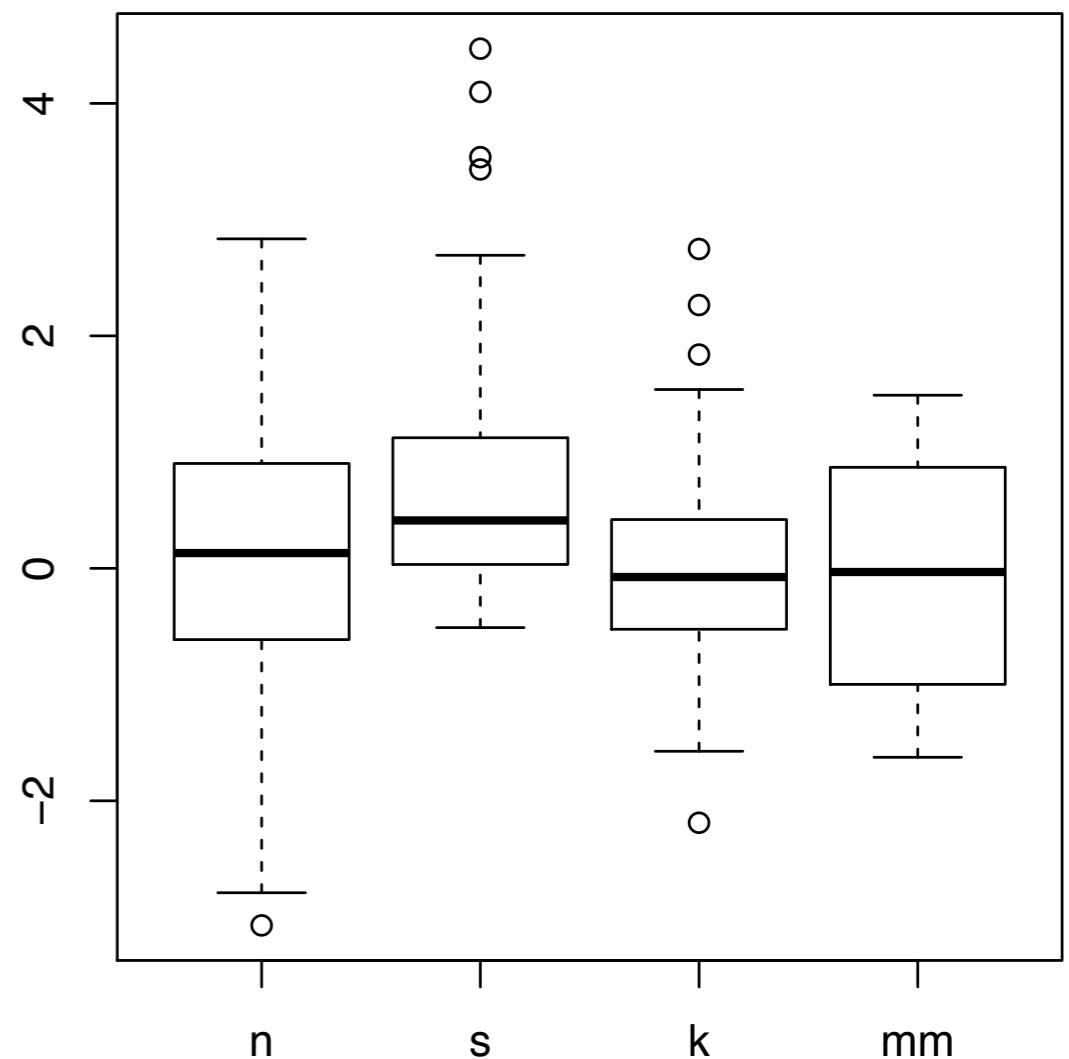
Distribution



# Multidimensional Visualization

## Glyphs

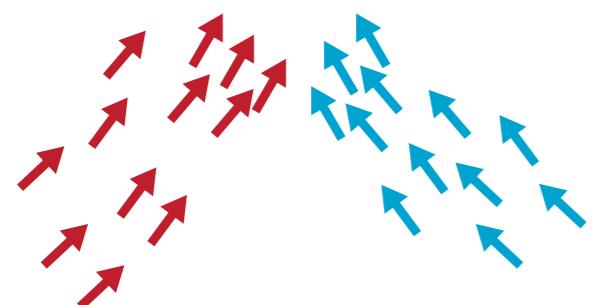
- static item aggregation
- **task:** find distribution
- **data:** table
- derived data
  - 4 quantitative attributes
    - median: central line
    - lower and upper quartile: boxes
    - lower upper fences: whiskers
  - outliers beyond fence cutoffs explicitly shown



# Multidimensional Visualization

## Glyphs

### Simple Glyph



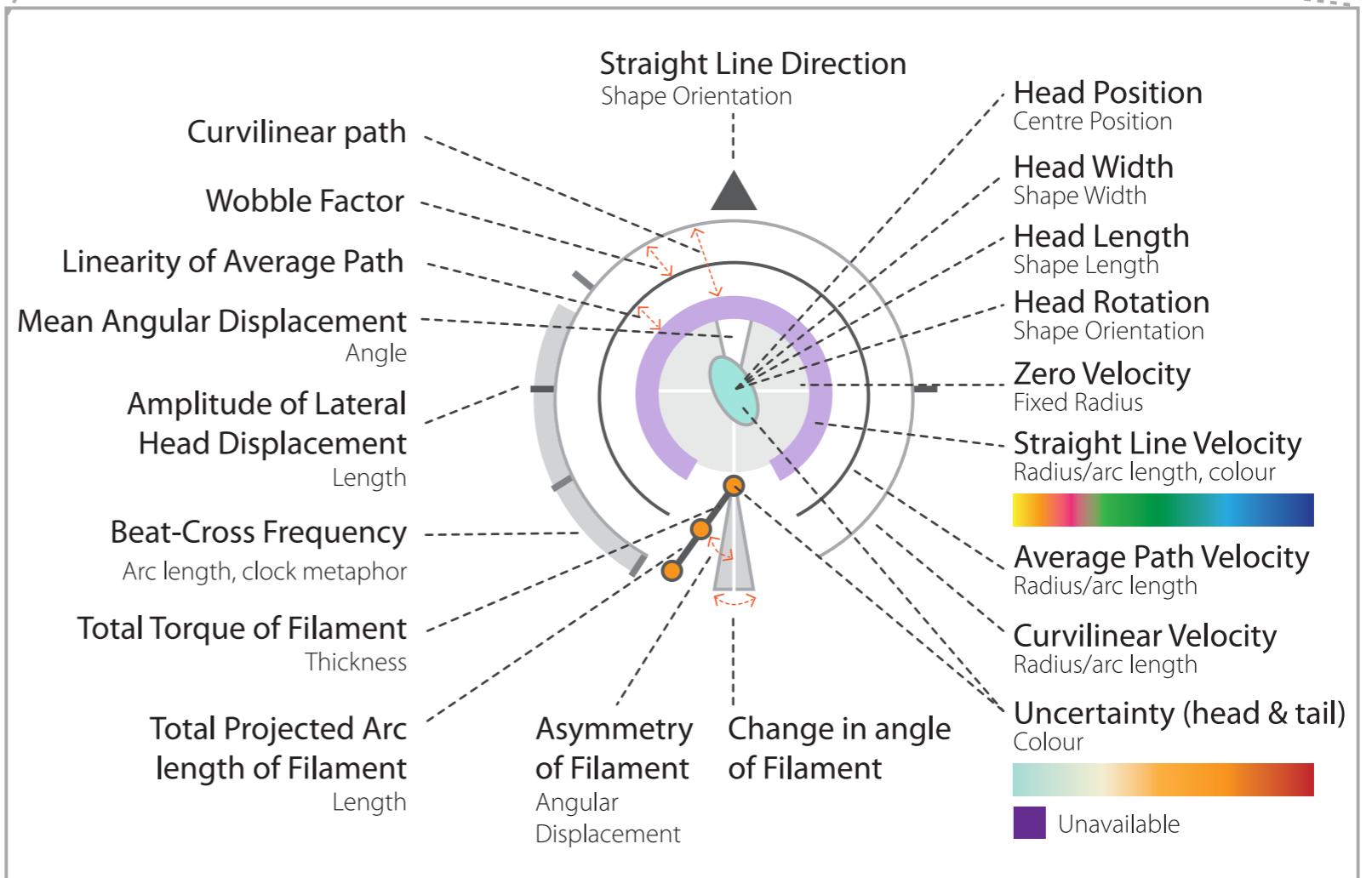
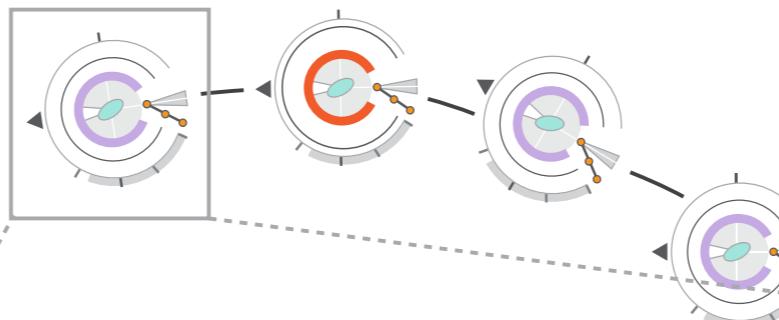
Temperature - Colour ■ ■

Wind direction - Orientation ↑↑→→

Wind Speed - Proximity

Location - Position

### Complex Glyph



# Multidimensional Visualization

## Glyphs | Example

When evaluating the impact of a publication, we generally look at the citation count.

This can be useful, but it doesn't tell us how impactful that publication was within its area.

Can we provide a way to summarise the impact of a publication in an intuitive way?

# Multidimensional Visualization

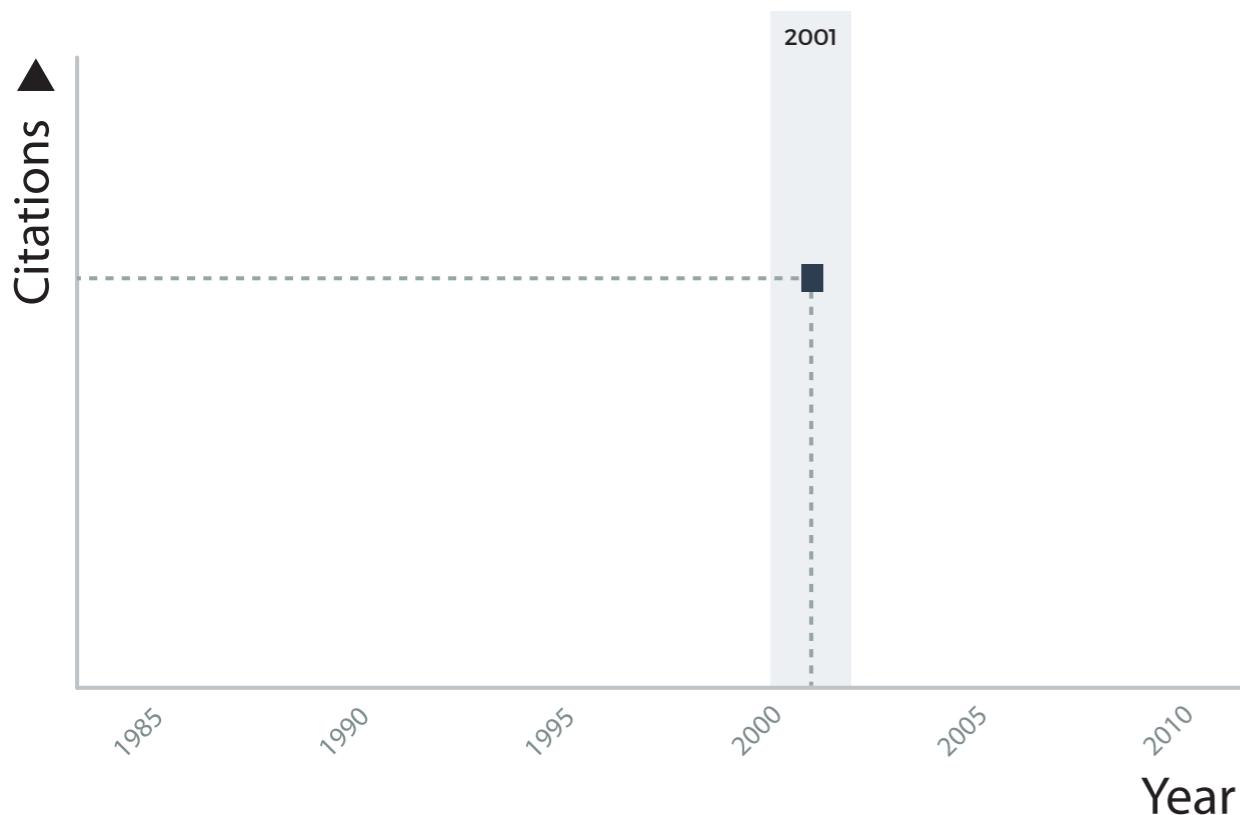
## Glyphs | Example

We wished to create a design that could be repurposed for a number of scenarios:

- 1) in a detailed view;
- 2) as a glyph; and
- 3) in a summary graph for an author or research field.

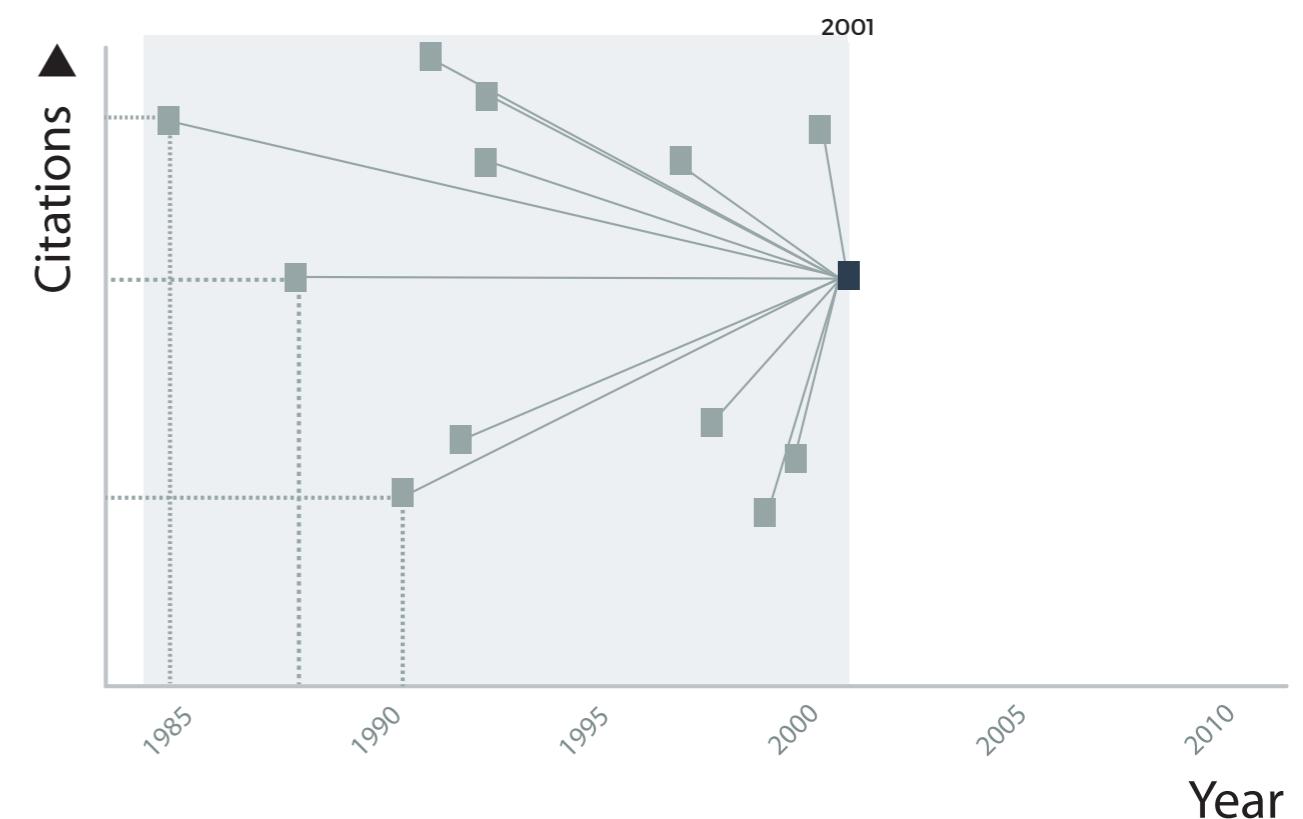
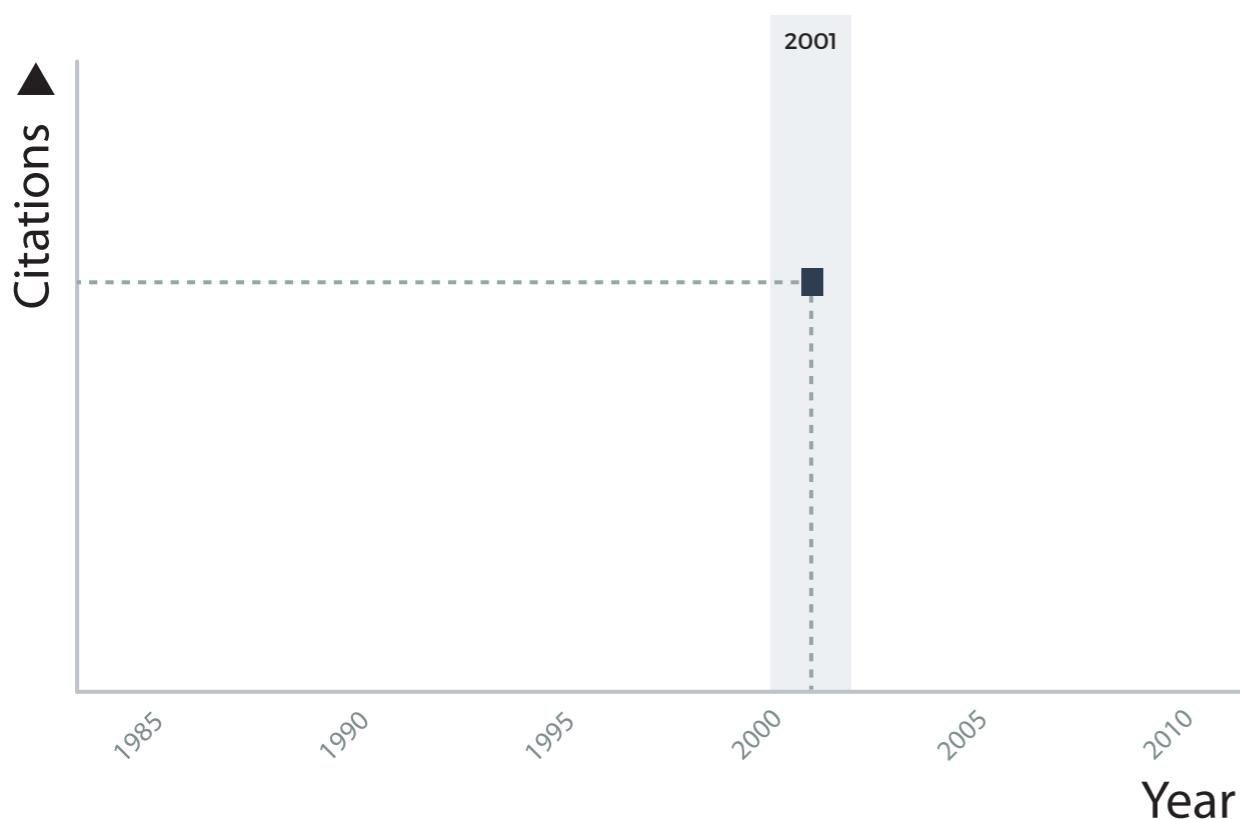
# Multidimensional Visualization

## Glyphs | Example



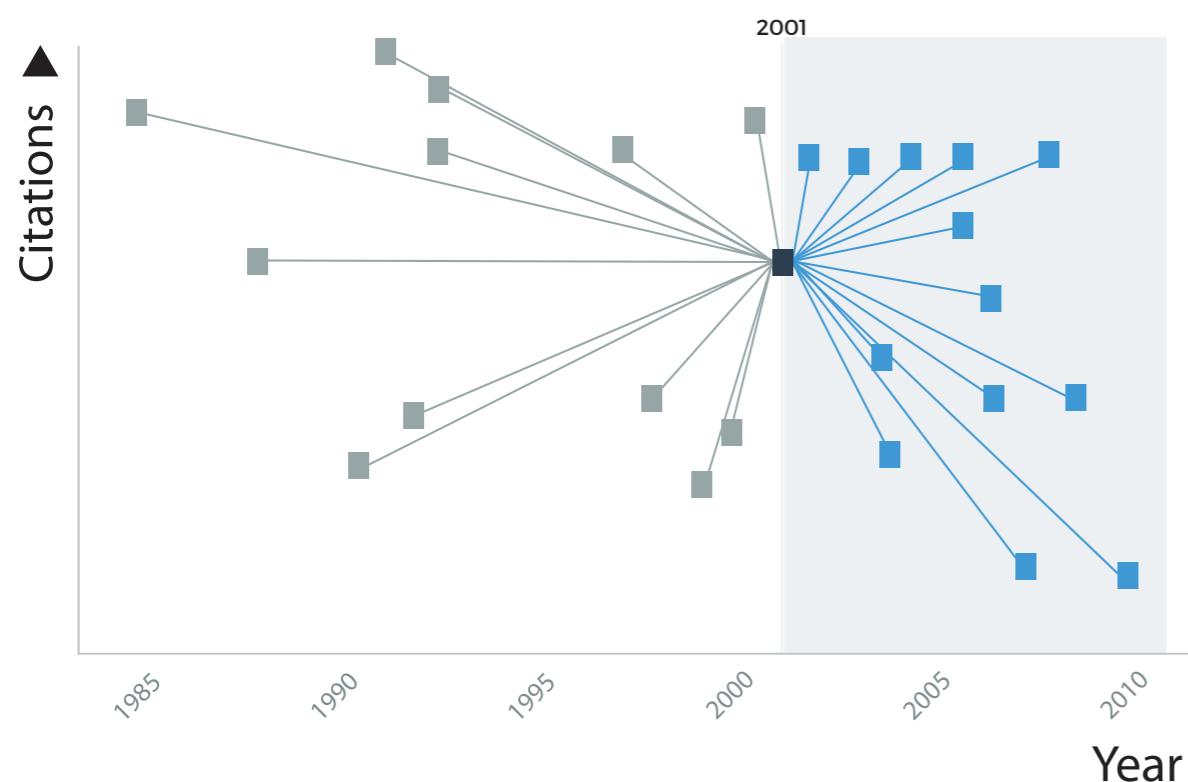
# Multidimensional Visualization

## Glyphs | Example



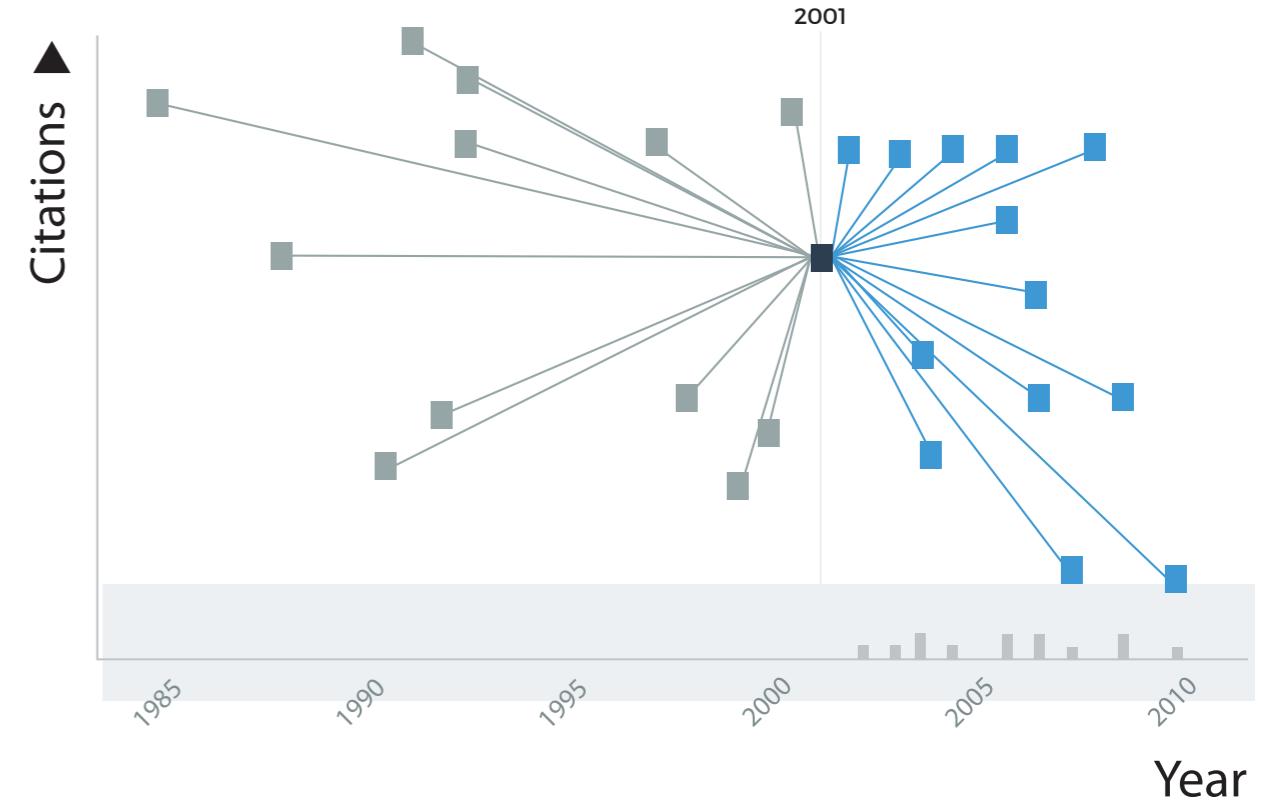
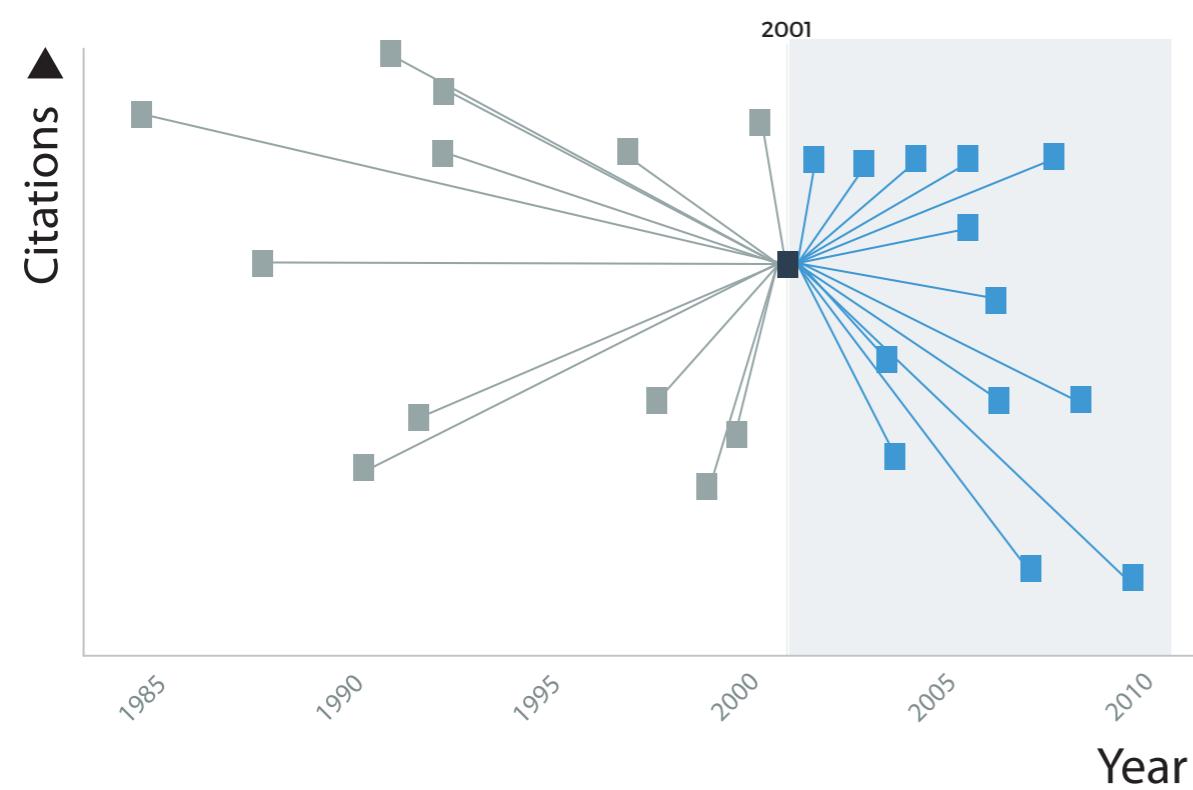
# Multidimensional Visualization

## Glyphs | Example



# Multidimensional Visualization

## Glyphs | Example

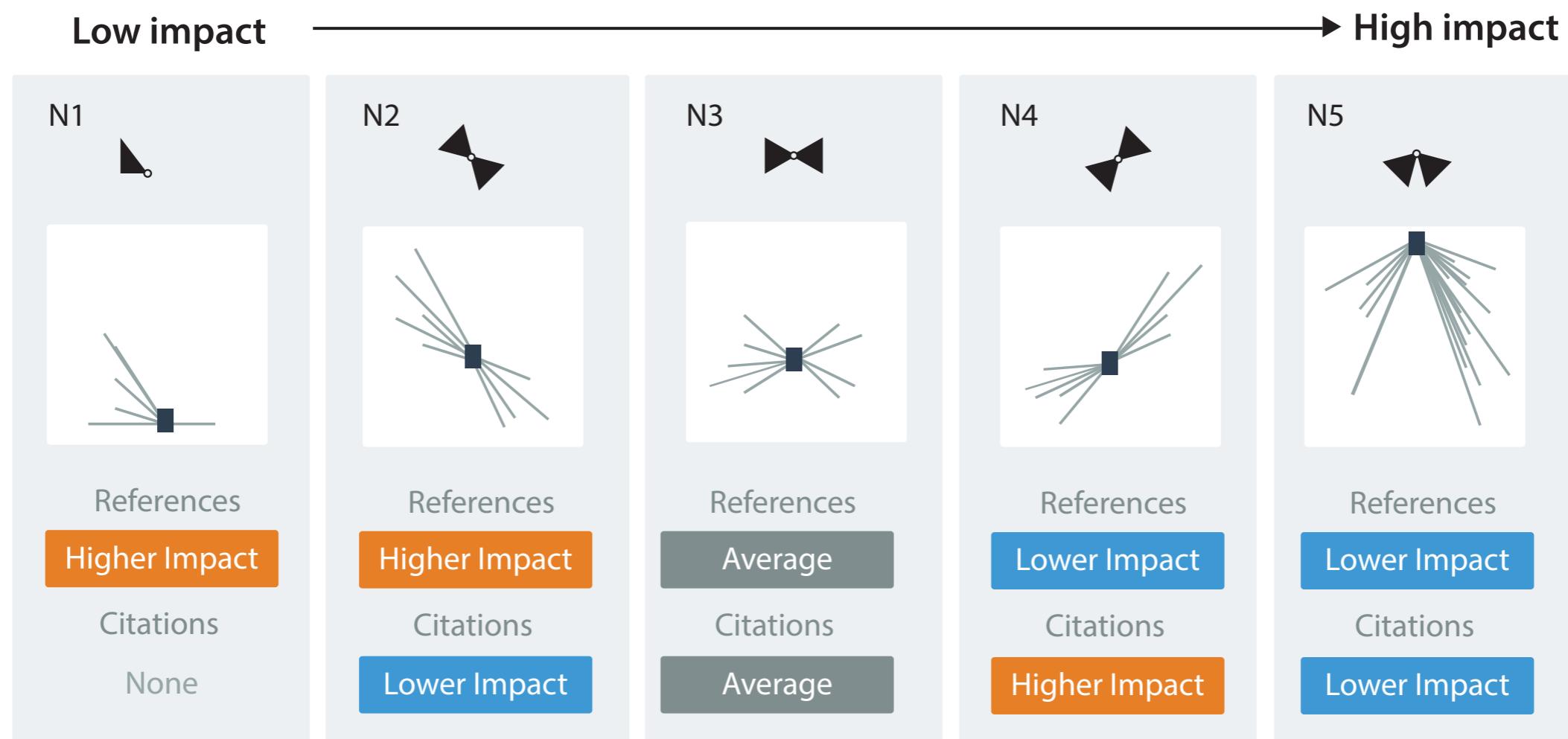


# Multidimensional Visualization

## Glyphs | Example

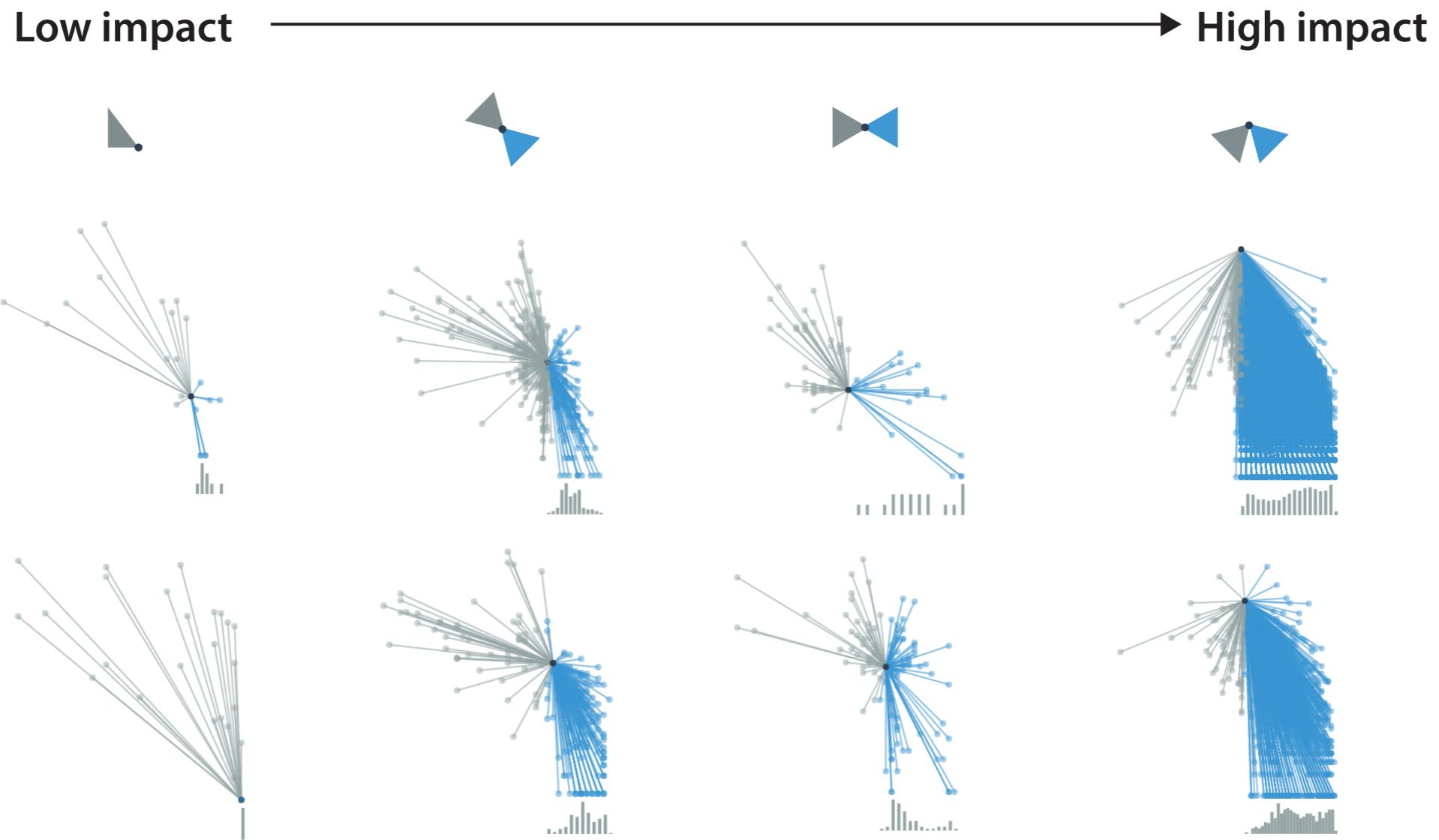
# Multidimensional Visualization

## Glyphs | Example



# Multidimensional Visualization

## Glyphs | Example



E. Maguire, J. Martin Montull, and G. Louppe, Visualization of Publication Impact,  
In Proceedings of EuroVis 2016, Short Paper (2016)

<http://inspirehep.github.io/impact-graphs/>

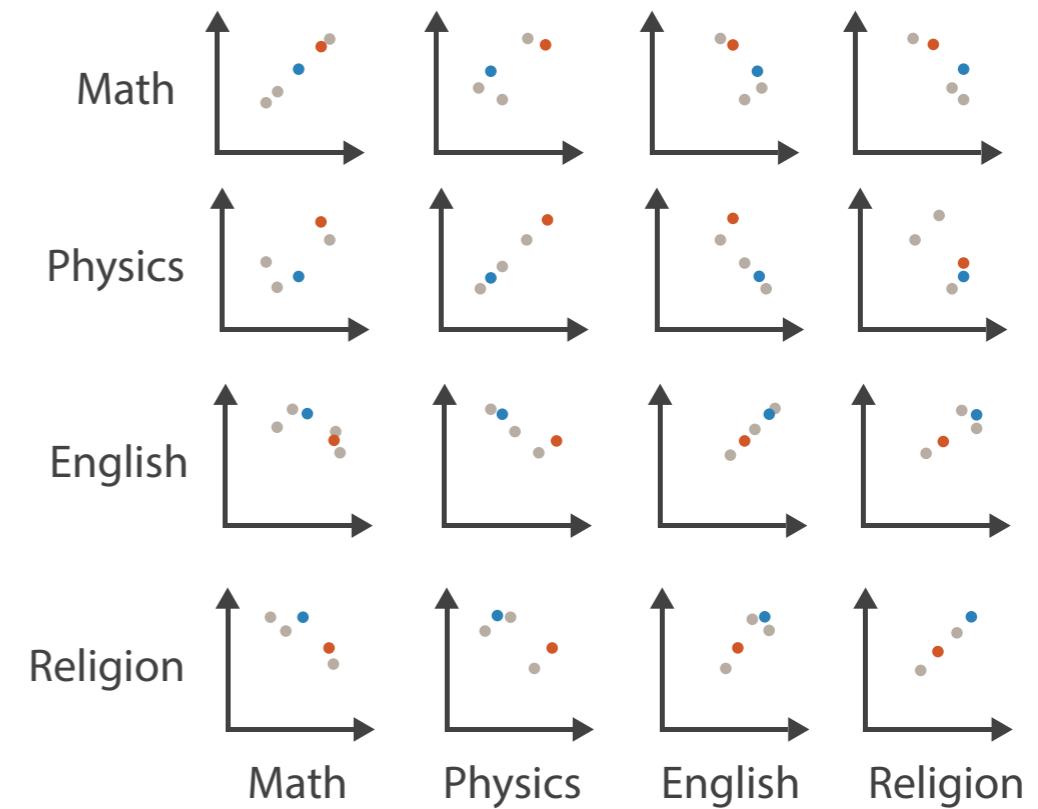
# Multidimensional Visualization

## 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



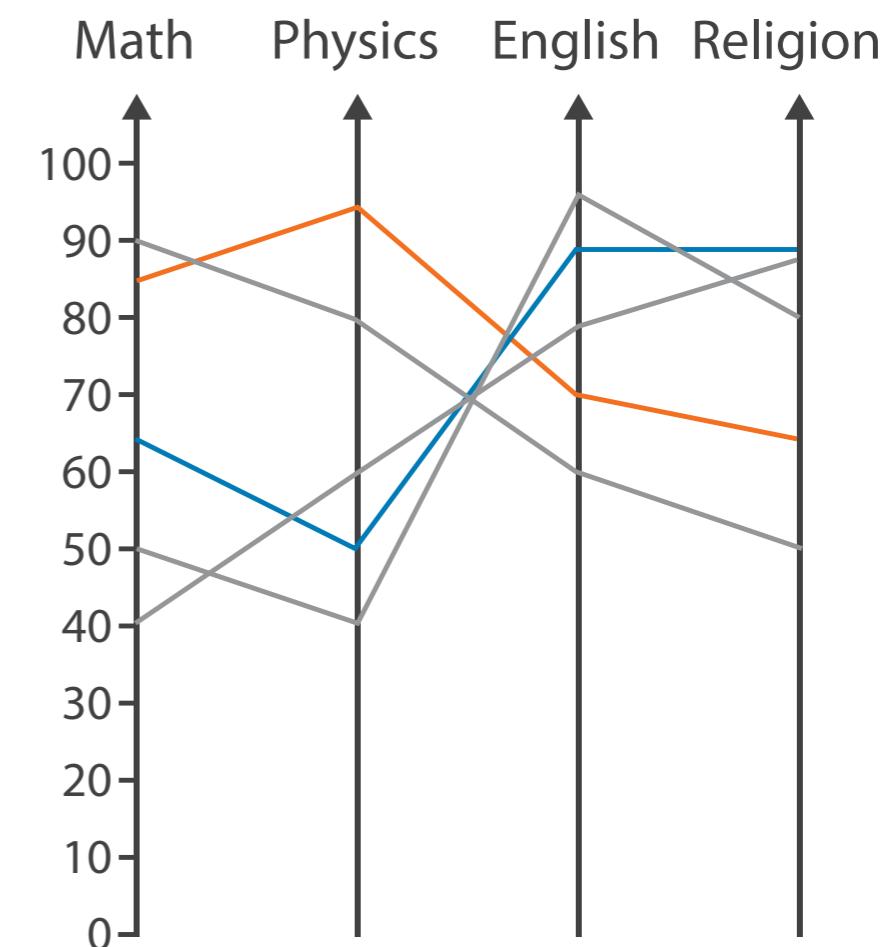
# Multidimensional Visualization

## 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



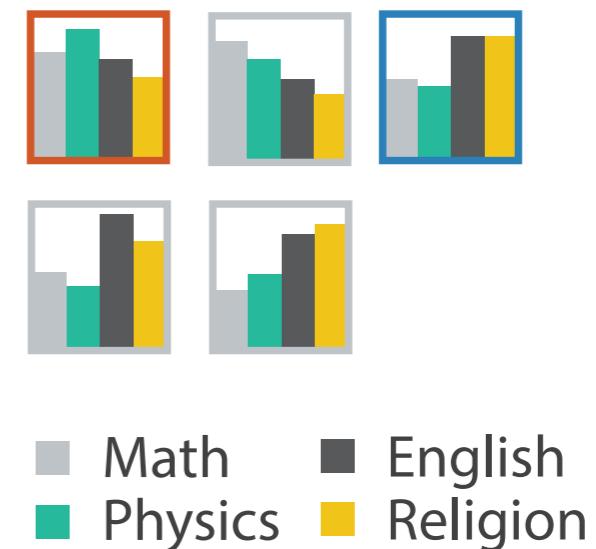
# Multidimensional Visualization

## A Simple Example | Student Test Results

Table

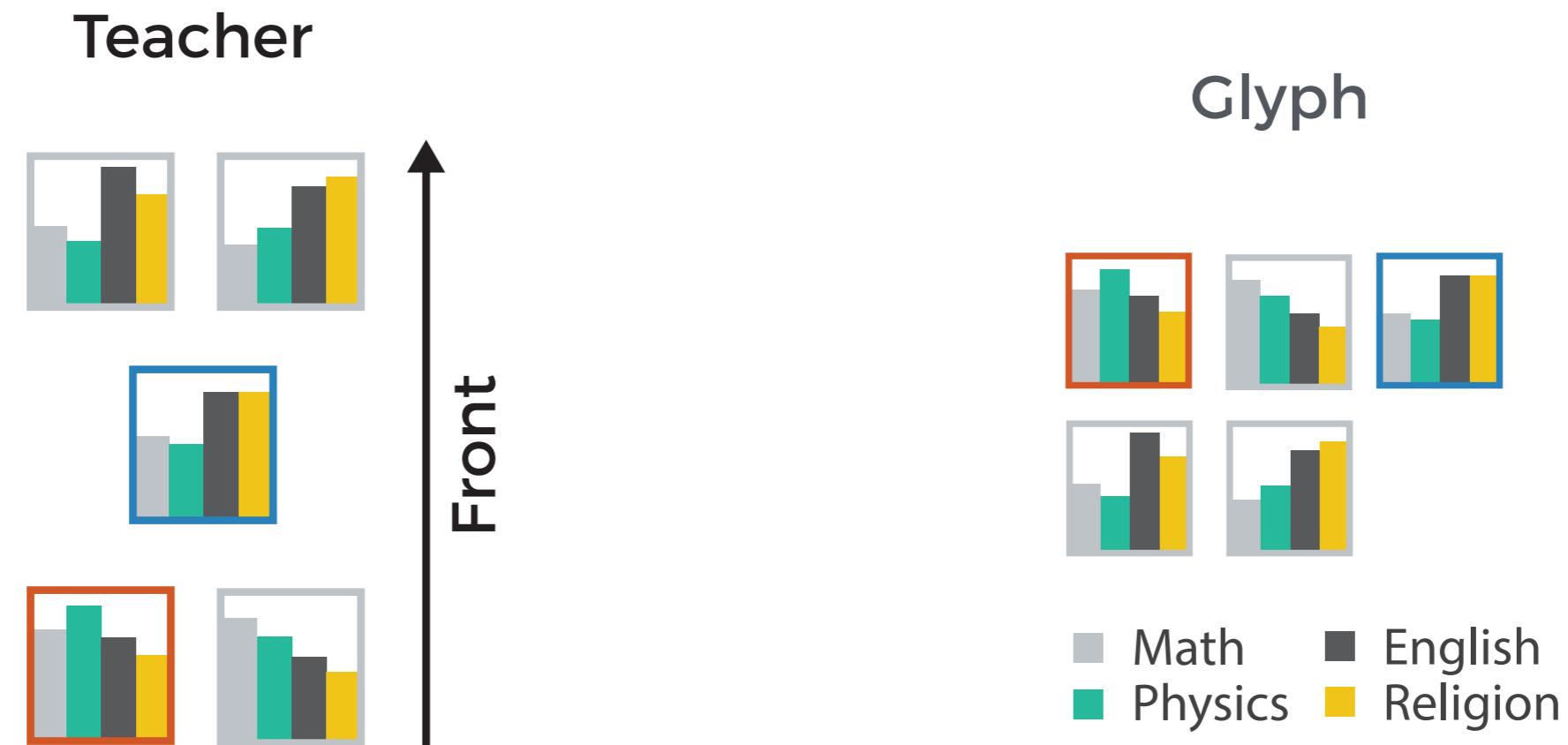
Math	Physics	English	Religion
85	95	71	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

Glyph



Arrange Spatially

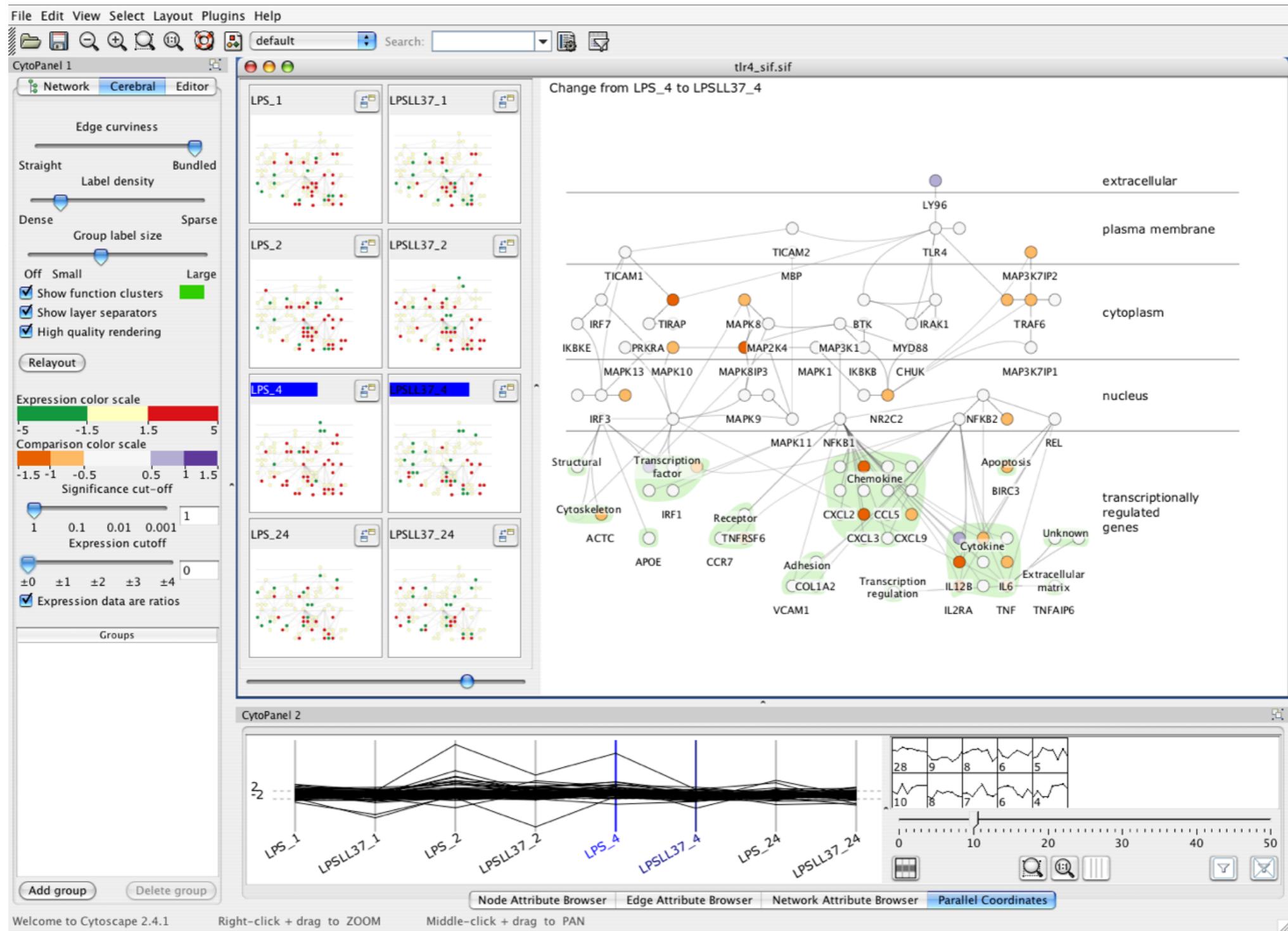
# on it Test Results



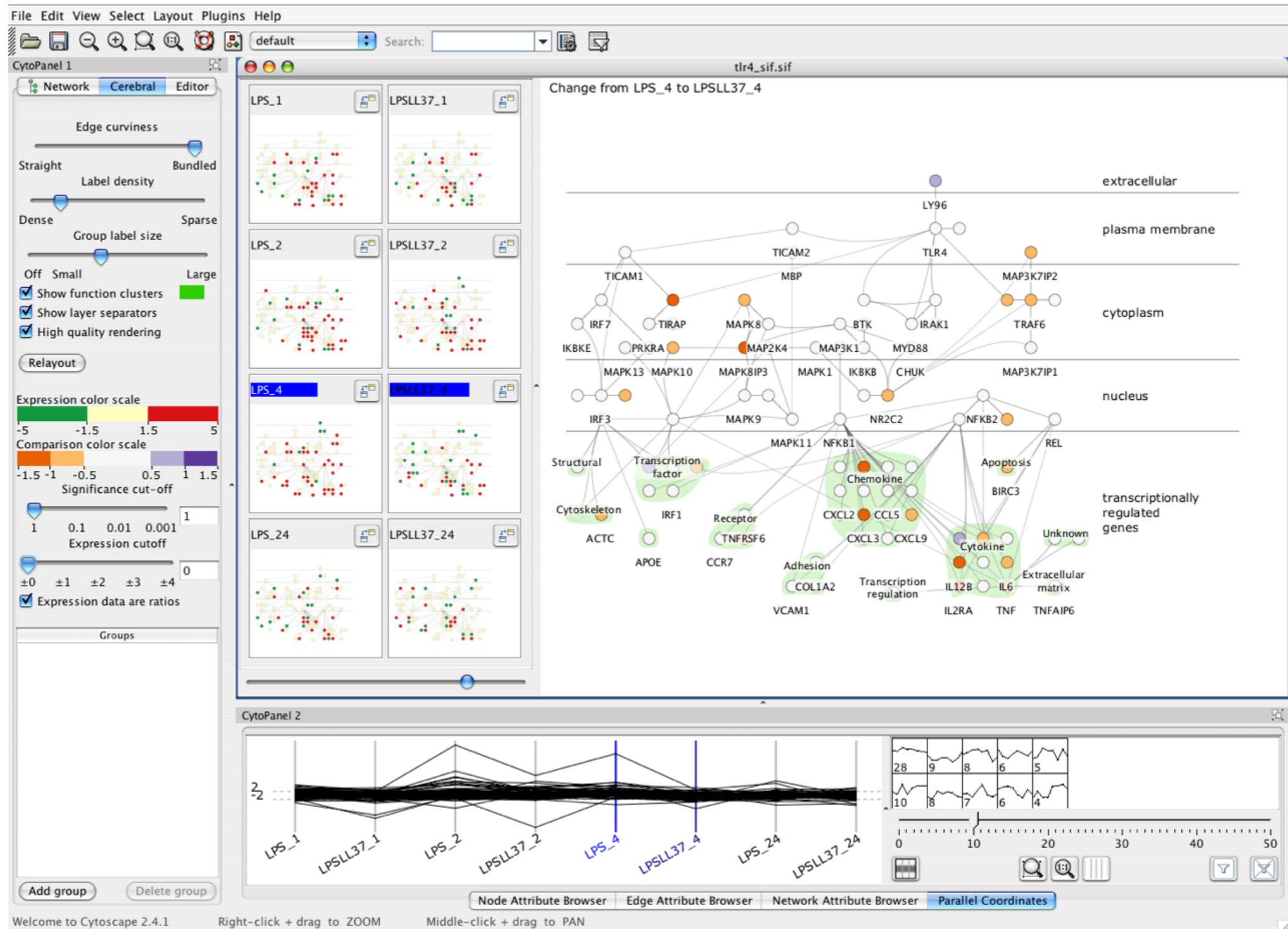
## What about topological data?

Representing trees and graphs...

# Graphs/Networks

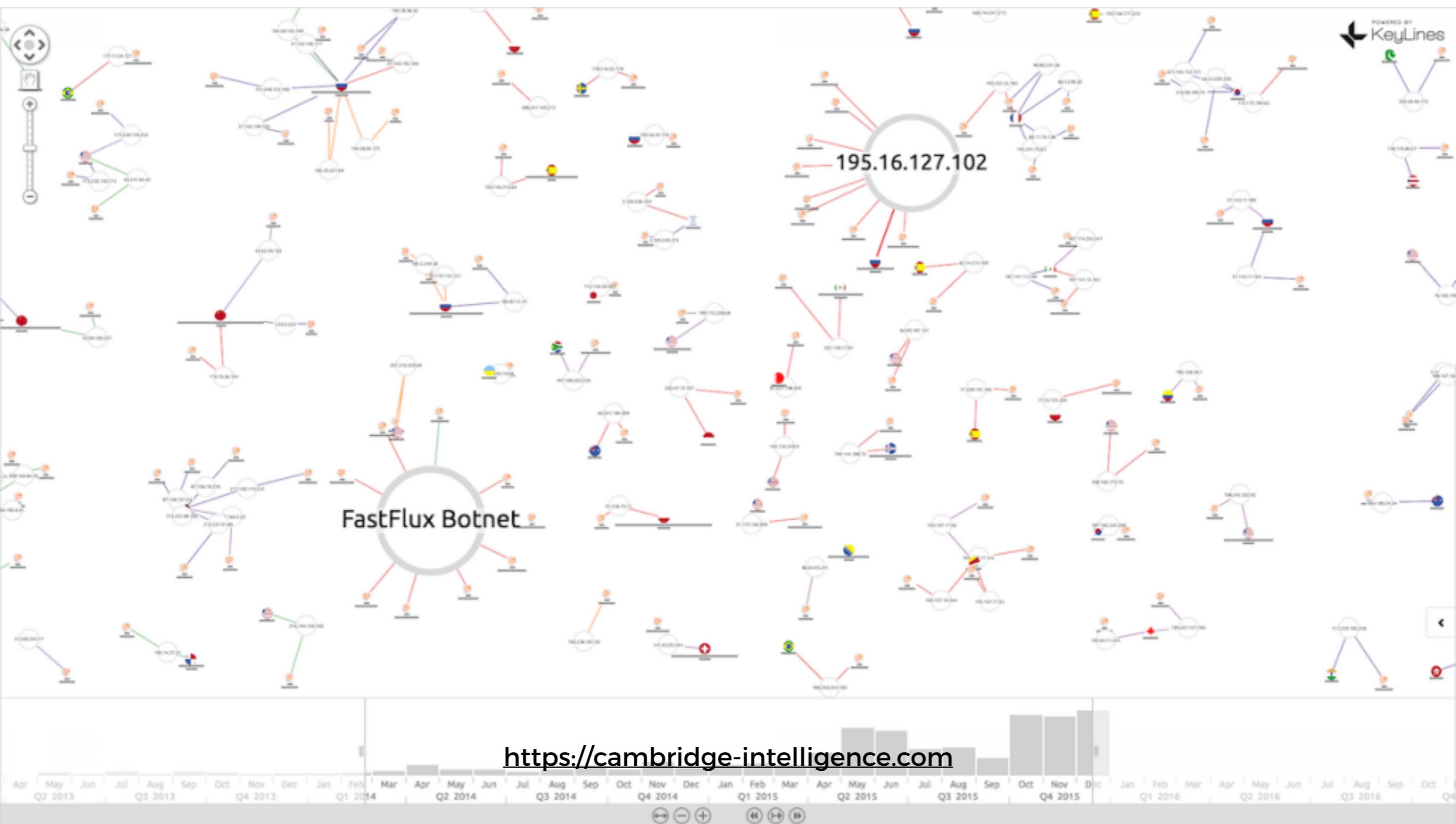


# Graphs/Networks



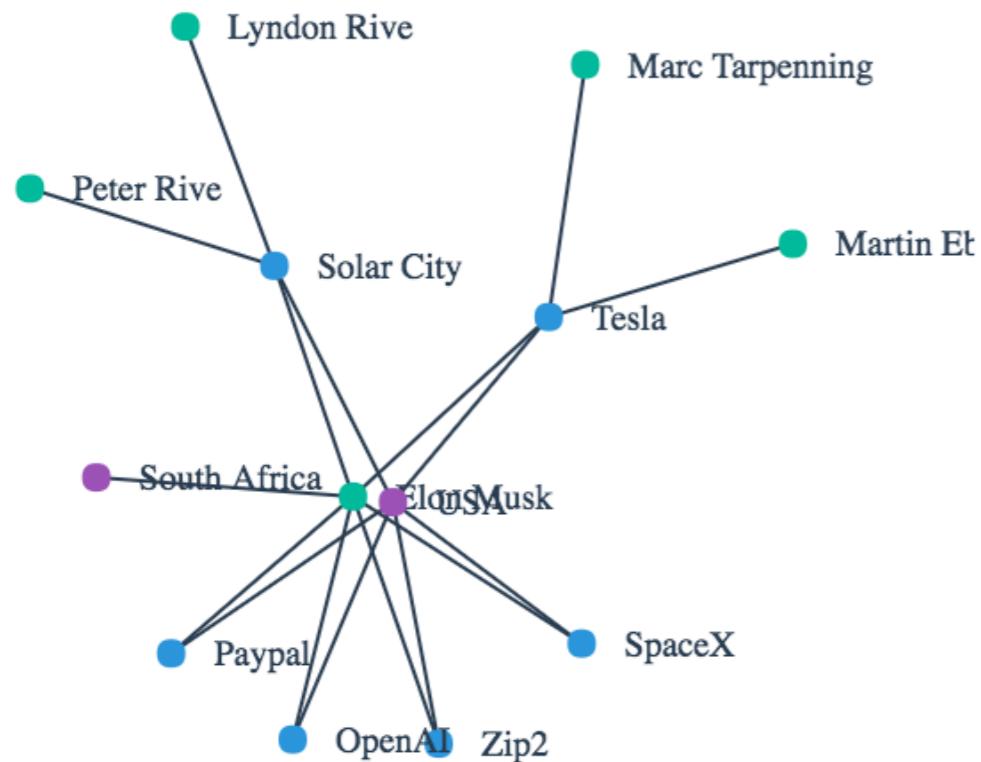
*Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE TVCG (Proc. InfoVis) 14(6):1253-1260, 2008.]*

# Graphs/Networks



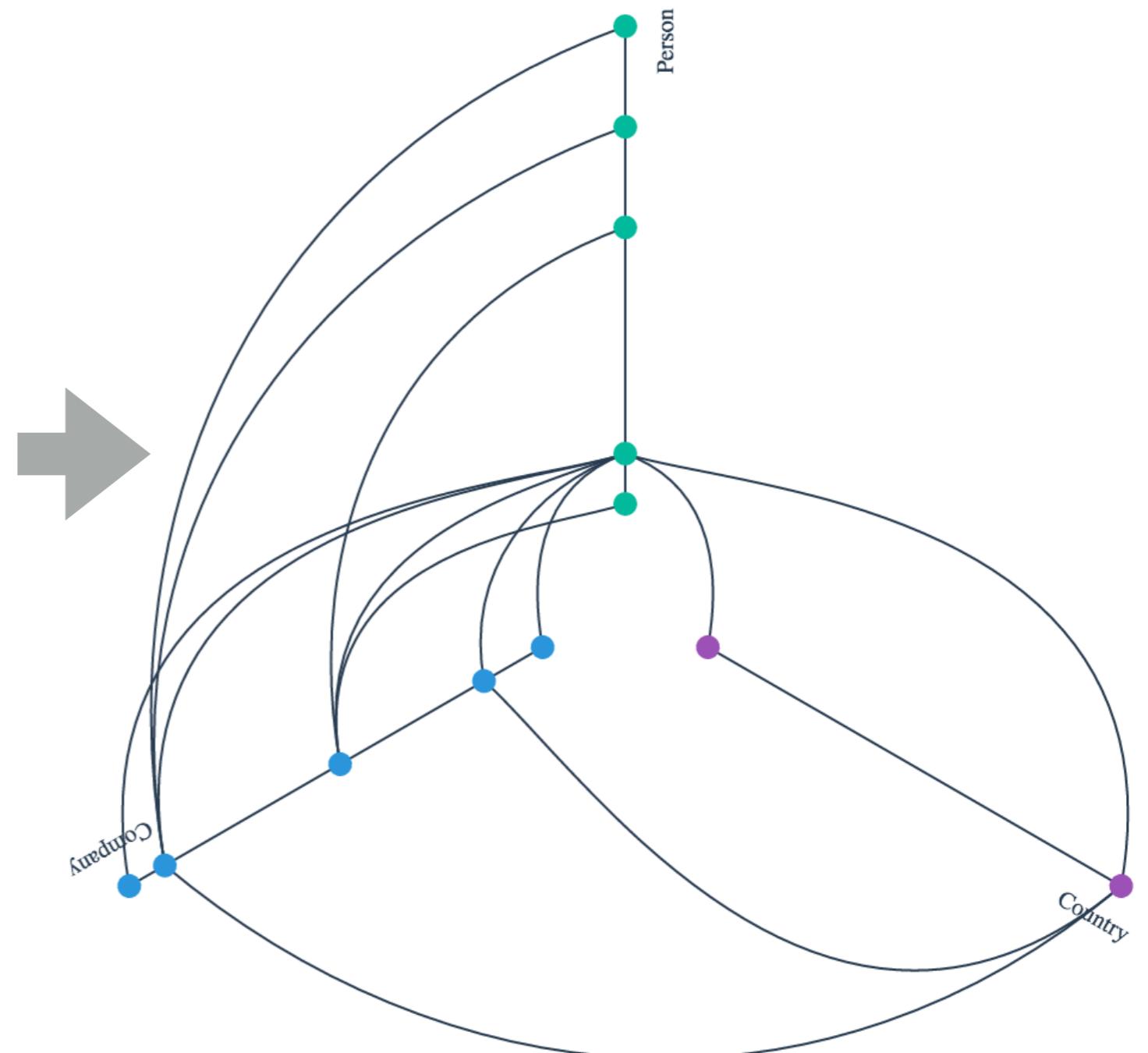
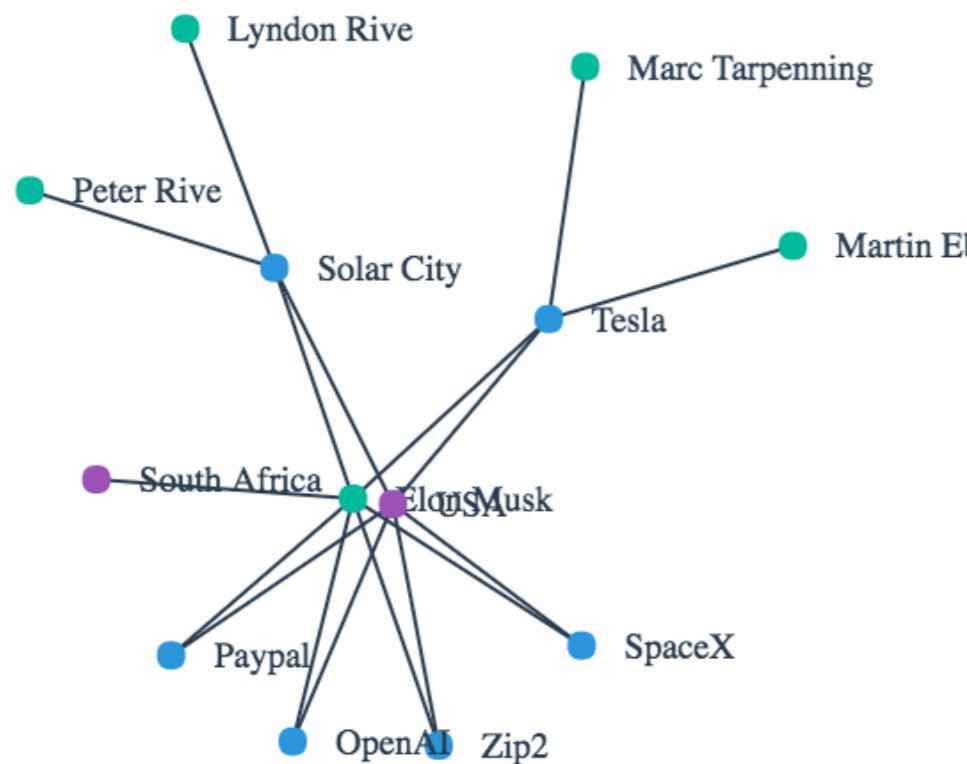
# Graphs/Networks

## Hive Plots



# Graphs/Networks

## Hive Plots

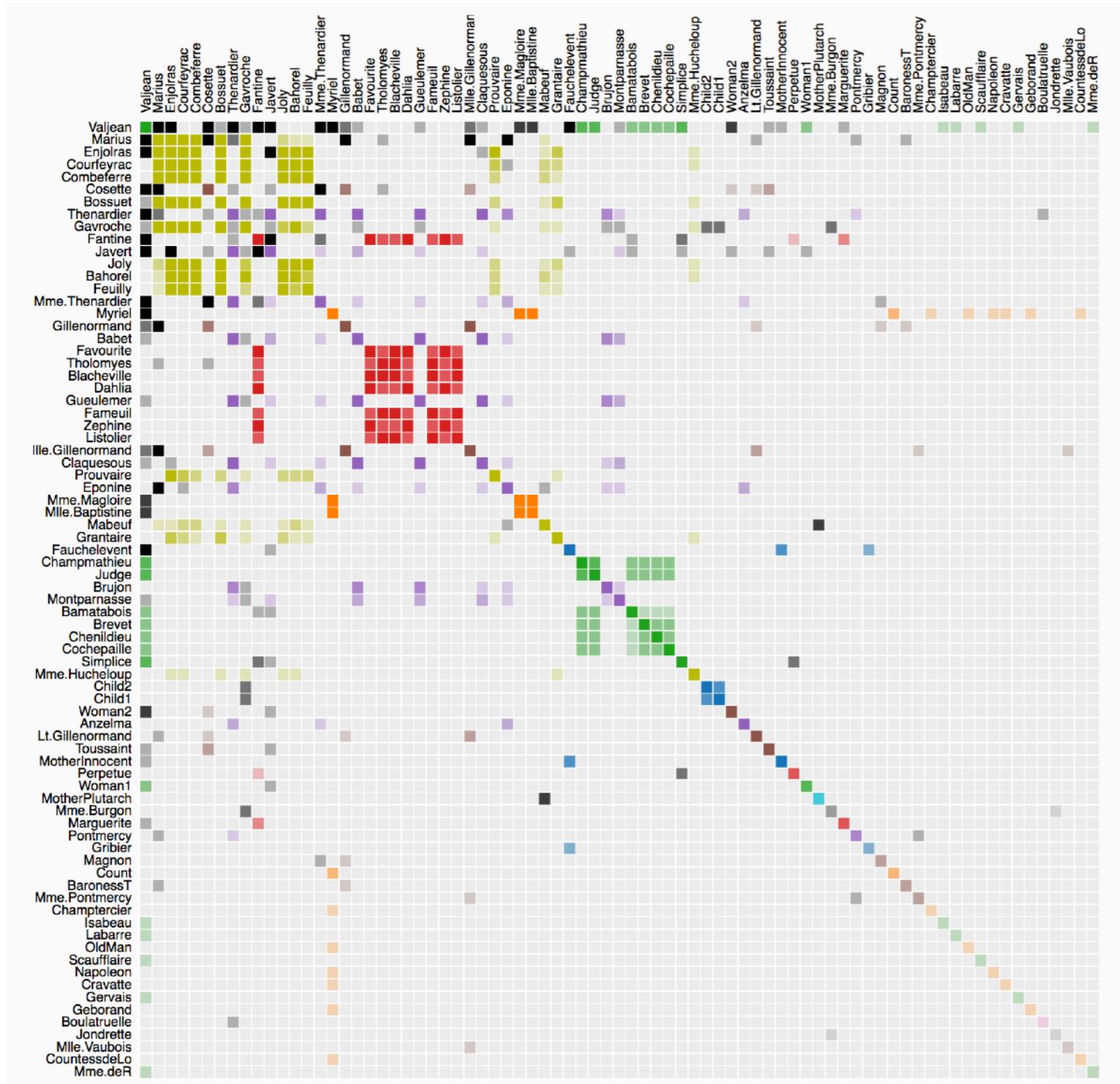


<http://jsfiddle.net/7a7b5dwp/>

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

# Graphs/Networks

## Matrix Representations



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

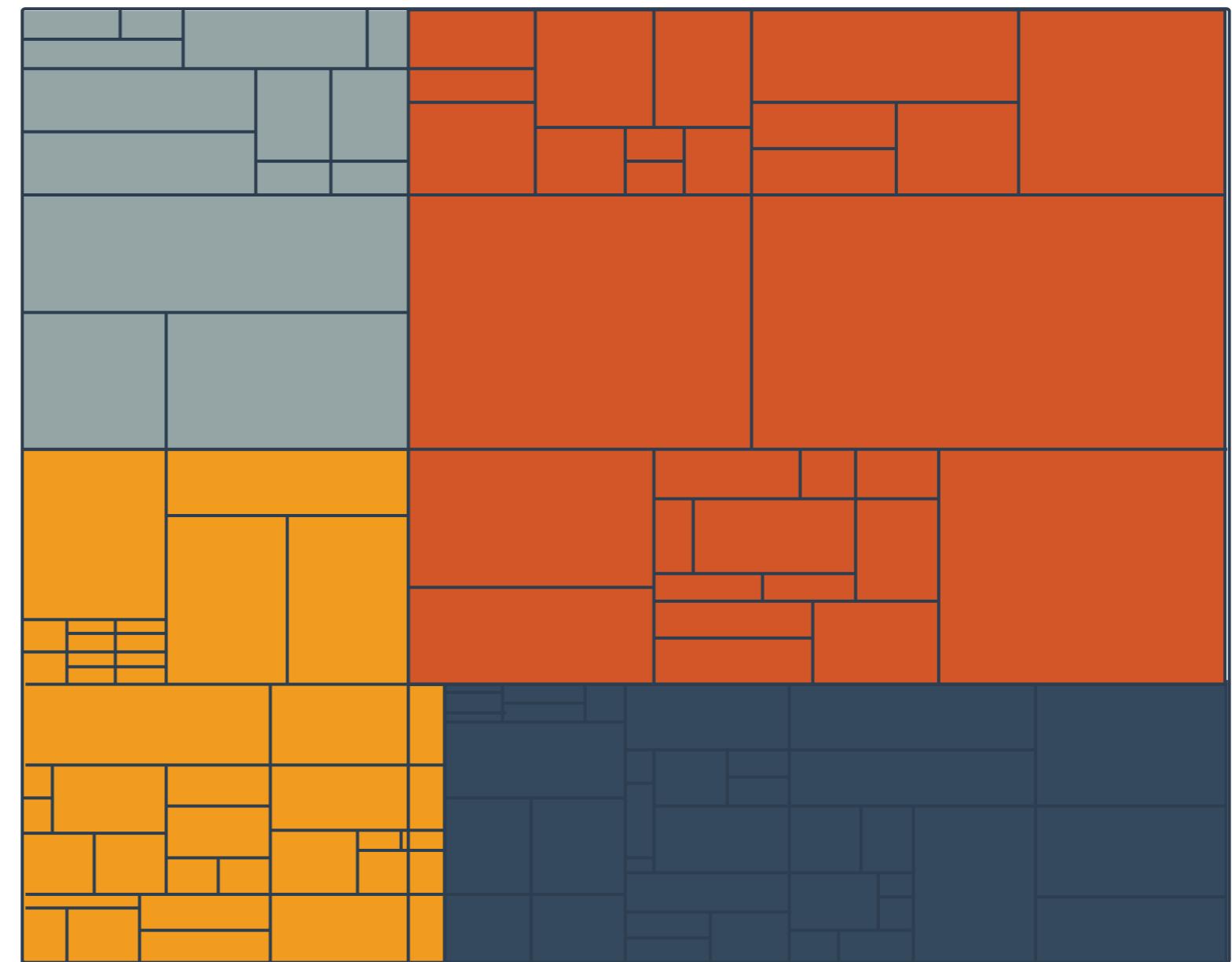
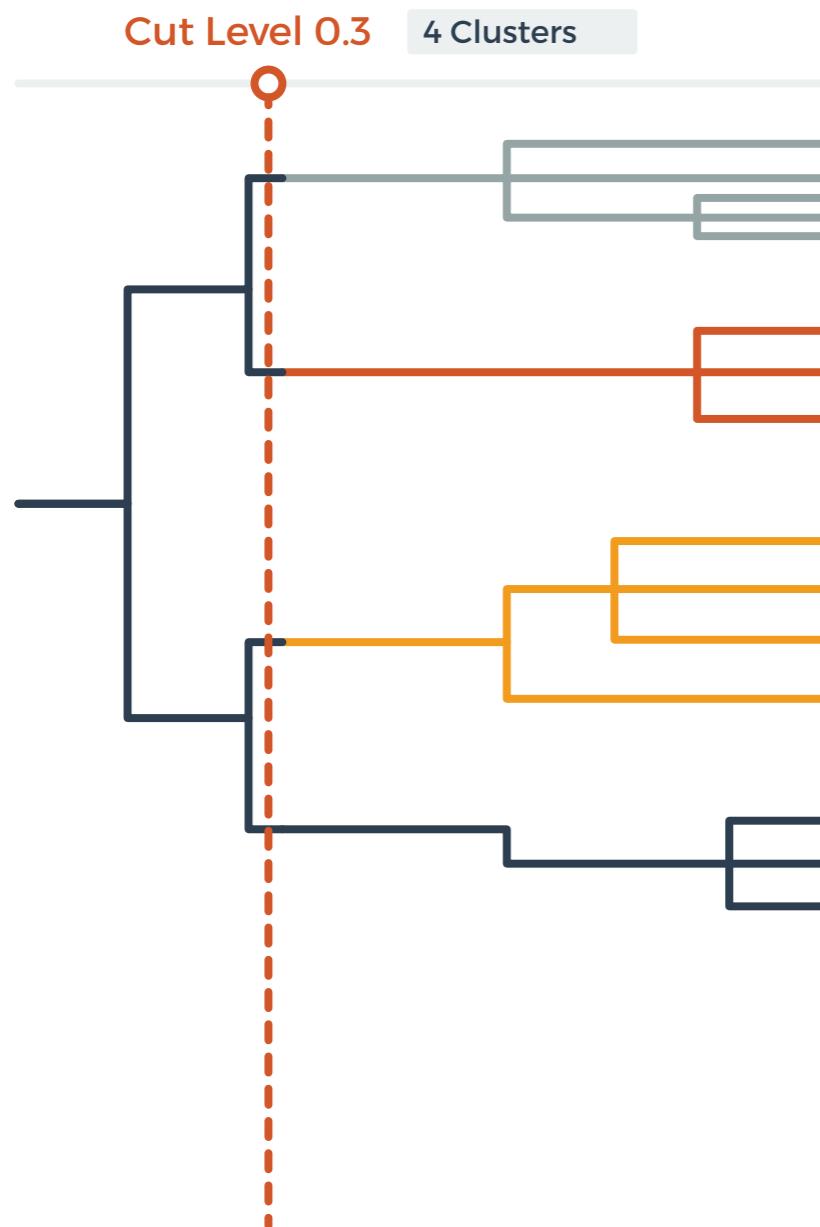
# Trees

Dendrogram

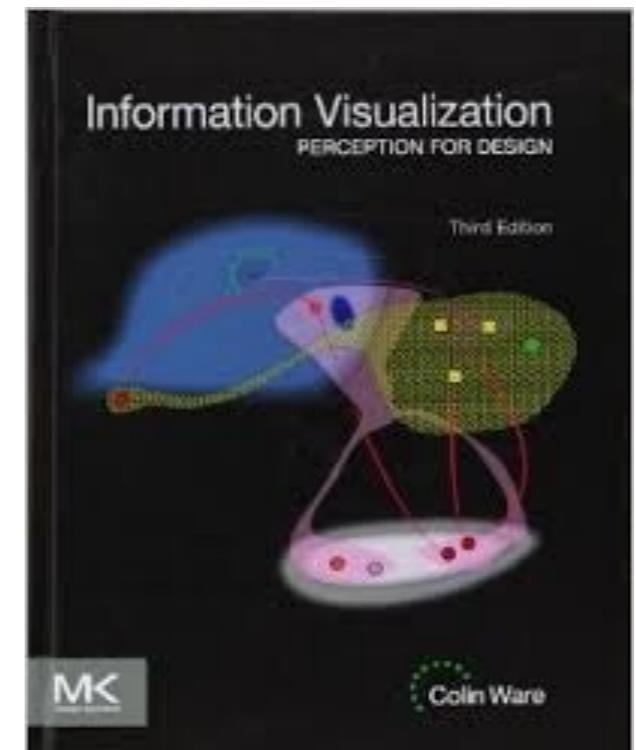
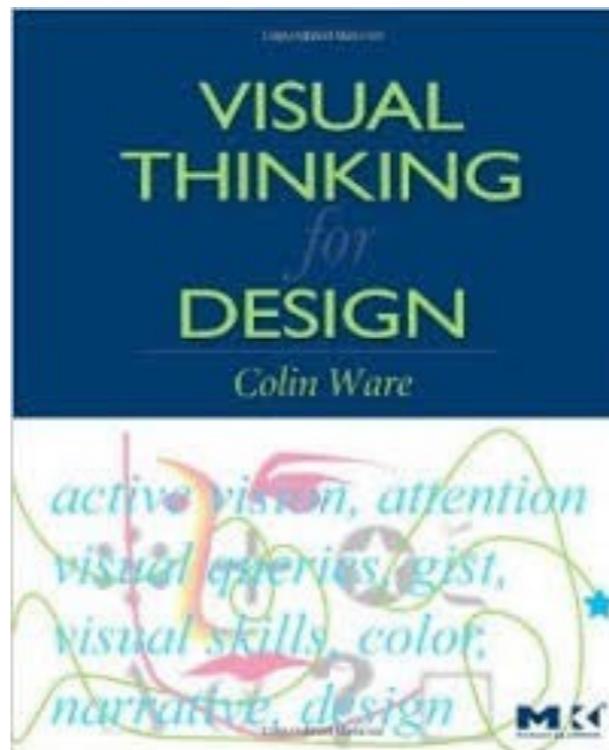
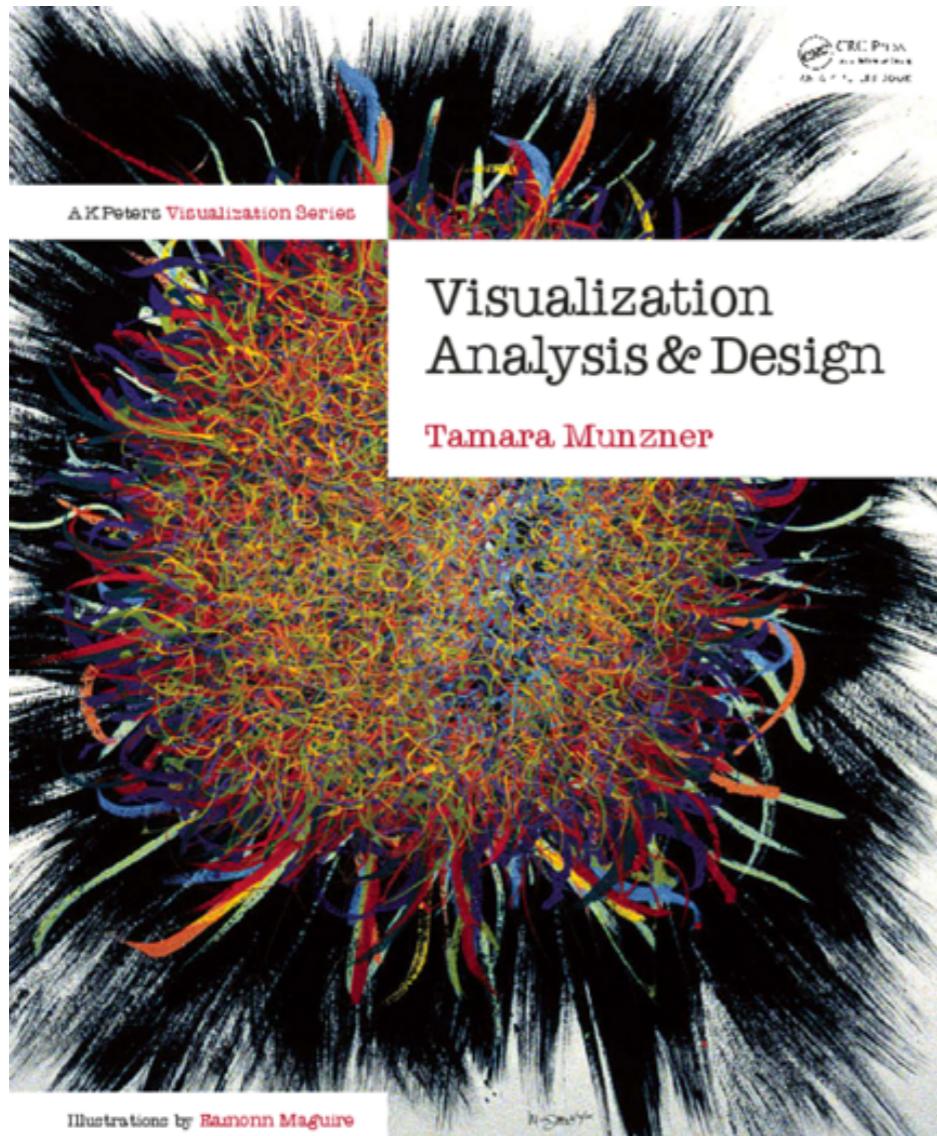


View as treemap ▾

Search



# More??



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

# Further Links

## Tutorials

**D3** <http://antarctic-design.co.uk/biovis-workshop15/>

**Dashboards** <https://thor-project.github.io/dashboard-tutorial/>

**Visualizing in Python** <https://github.com/eamonnmag/CERN-CSC-2017>

## Visualization Sites

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

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

<http://flowingdata.com/>

[Data Vis Catalogue](#)

## Python Data Vis Tools

[Pandas Data Vis](#)

Matplotlib

Seaborn



# Questions

@antarcticdesign  
[eamonnmag@gmail.com](mailto:eamonnmag@gmail.com)

# Additional

## Scales

[https://github.com/eamonnmag/vis-course-practical/  
blob/master/Visualize.ipynb](https://github.com/eamonnmag/vis-course-practical/blob/master/Visualize.ipynb)