Keynote
Dealing with Data 2019
University of Edinburgh, Jan 2020

Dealing with Data

Benjamin Bach
http://benjbach.me
@benjbach

Keynote
Dealing with Data 2019
University of Edinburgh, Jan 2020

Dealing **vis** Data

Benjamin Bach
http://benjbach.me
@benjbach

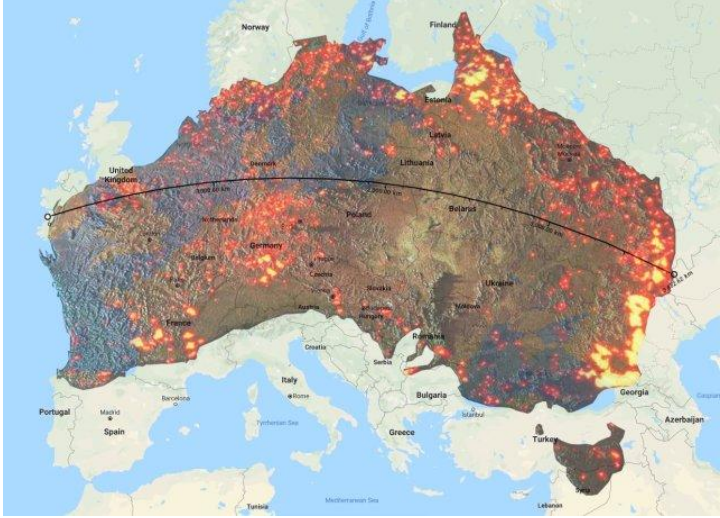
Data

↘ **Visual representations**

↘ **Information**

↘ **Insights**

↘ **Decisions + Actions**



Anthony Hearsey

Collection

Cleaning

Exploration

Transformation

Analysis

Presentation

Data Visualization

Human-Computer Interaction

Data Visualization

Human-Computer Interaction

Challenges

- How to visualize **complex data**?
- How do we best leverage human **perception**?
- How do we provide for powerful **interaction**?
- How to **communicate** with visualizations?
- How to build efficient **tools**?
- How to make visualizations **understandable**?

https://**visualinteractivedata**.github.io

visual | interactive
data | School of Informatics, Institut for Design Informatics, University of Edinburgh

design informatics | informatics

Latest

12/19+++**Visual+interactive data** contributes towards **four** papers for CHI 2020! Congrats everyone!

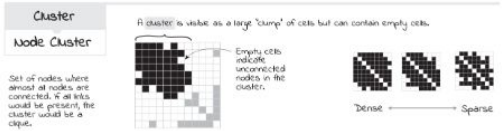
10/19+++**Ryan Bowler** joins as a PhD student, working on Temporal Design.

10/19+++**Mashael Hamad Alkadi** joins as a PhD student, working on Learning Analytics and understanding visualizations.

09/19+++Tobias Kauer joins as a PhD student, working on Discursive Visualizations.

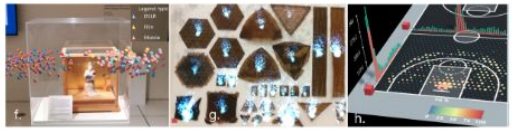
07/19+++Our survey on geographic networks goes online.

Selected Projects (more)



Data Visualization Cheat Sheets are aimed at supporting learning and teaching of data visualization techinques and investigate ways to better teach data visualization in general.

Data Comics combine visual language and narrative patterns from traditional comics to tell expressive stories with data and visaulization. We explore this exciting medium and develop workshops and guiding material.



The Edinburgh VisHub is an university-wide lab for data visualization equipment to foster research, collaboration, learning, and outreach. **DXR** and **IATK** are two toolkits for creating immersive data visualizations in augmented and virtual reality. Check how to participate in our upcoming workshop at CHI 2020.

People

Benjamin Bach, Lecturer (Assistant Prof.) in Design Informatics and Visualization at the University of Edinburgh. His research designs and investigates interactive information visualization interfaces to help people explore, communicate, and understand data across media such as screens, mixed reality, paper, and physicalizations.
Google Scholar

Zezhong Wang, PhD student working on *Methods for Creating and Teaching Data Comics*.
Personal website
Google Scholar

Tobias Kauer, PhD student working on *Discursive Data Visualization*.
Personal website
Google Scholar

Mashael Hamad Alkadi, PhD student wr
Learning Analytics.

John Harper-Lee workir
Google Scholar

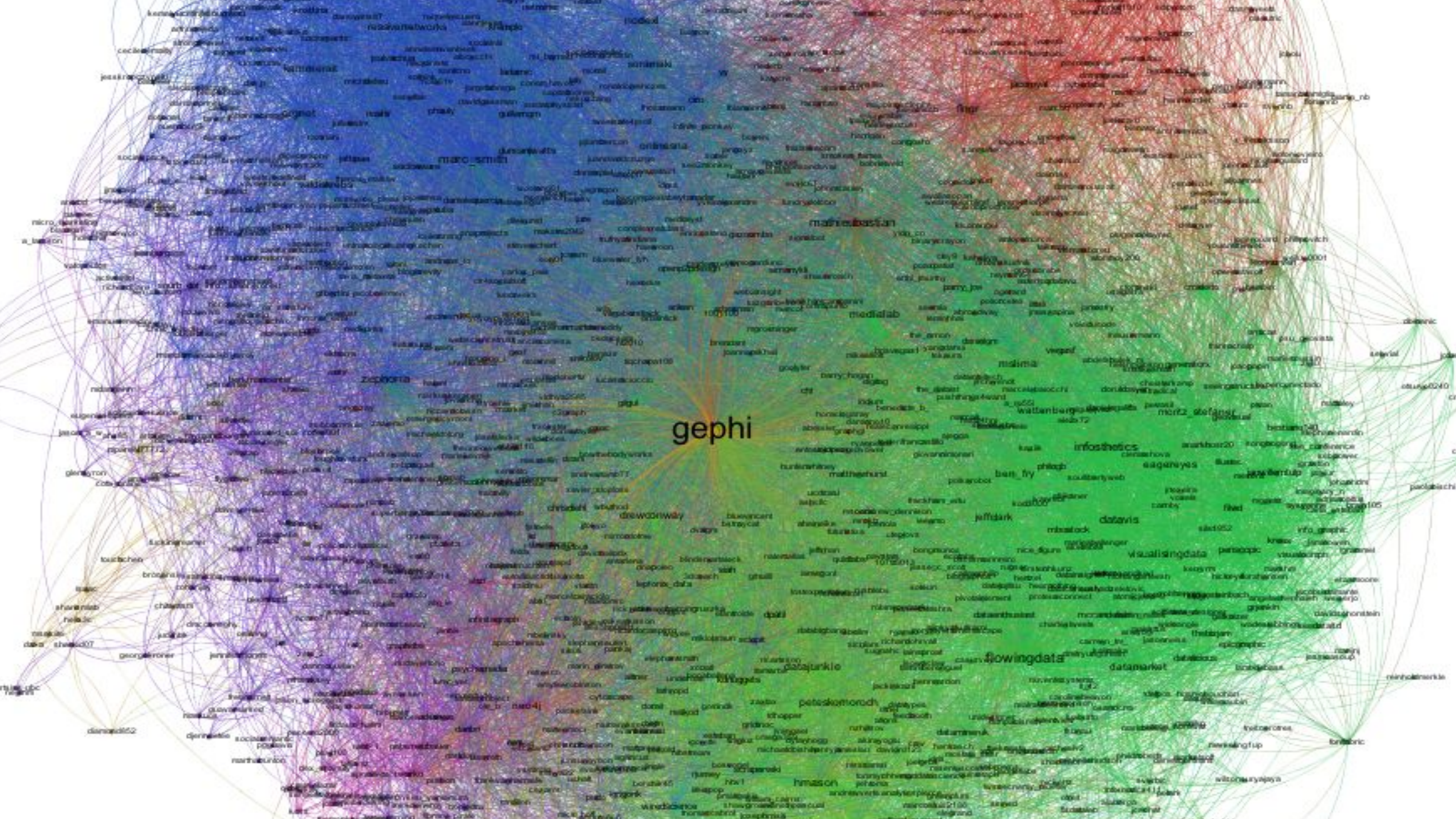
Ryan B
way
esearching
e complexities
puter-Interactions.

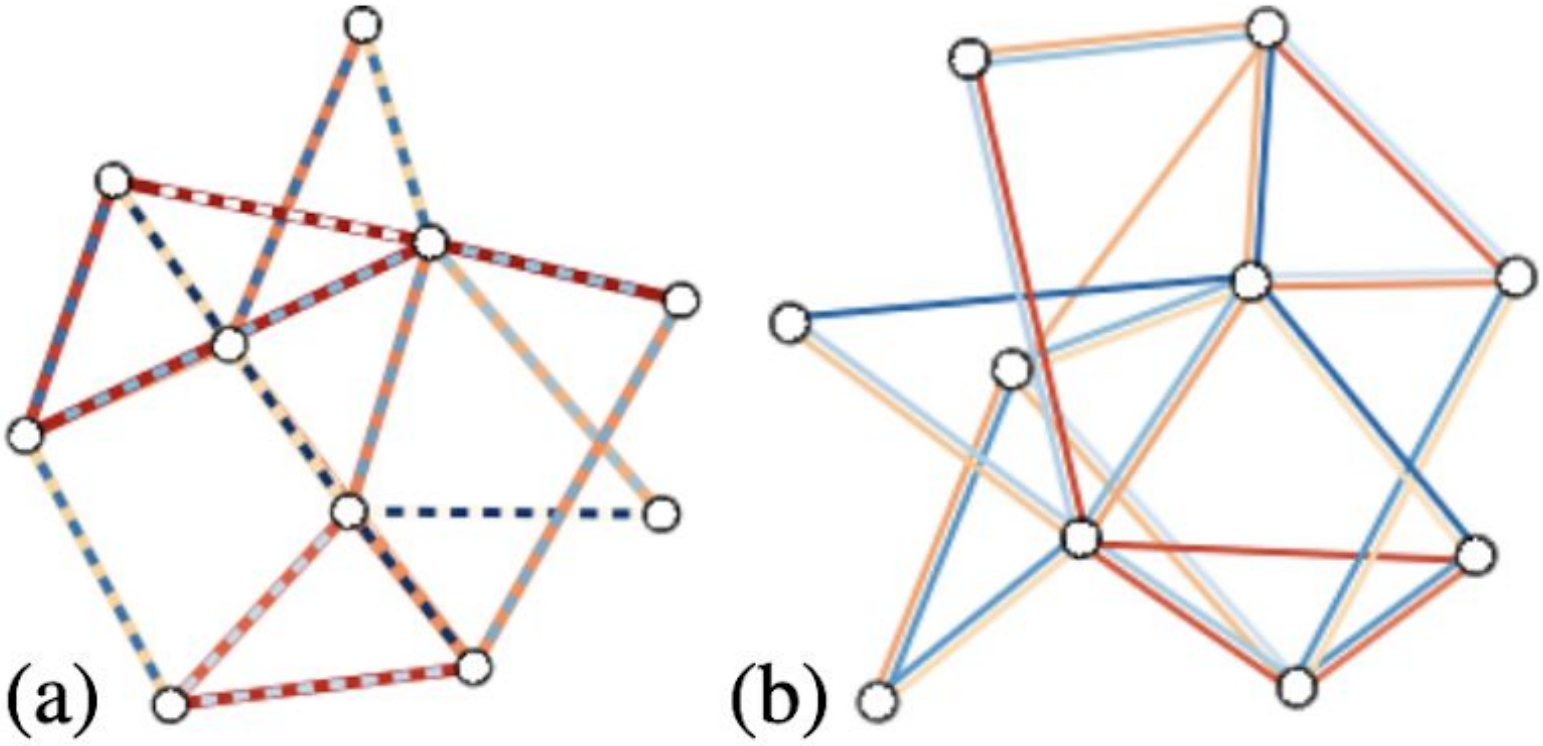
Hiring!

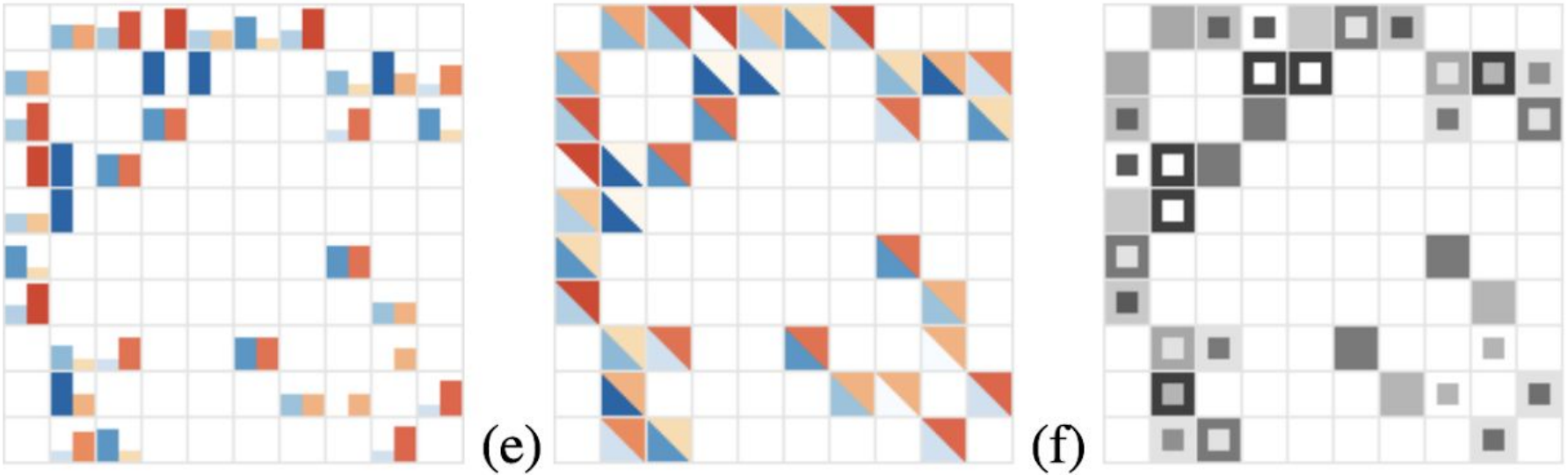
design informatics

THE UNIVERSITY of EDINBURGH informatics

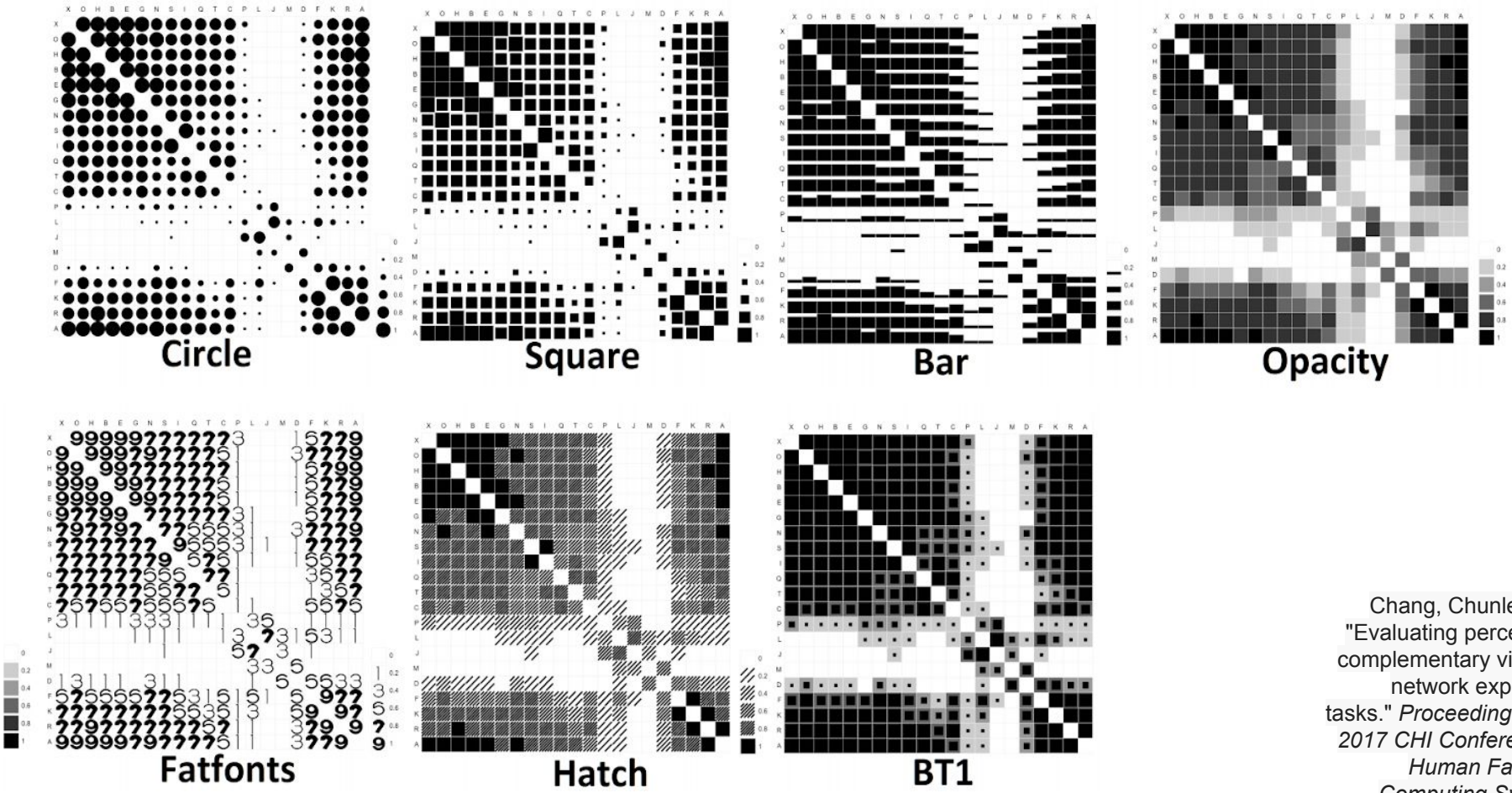
Complex Data


gephi
marc_smith
infosthetics
flowingdata
datavis
visualisingdata
ben_fry
eagereyes
drewconway
zephoria
medialab
hmason
peteskomoroch
datajunkie
datamarket
jeffclark


(a)
(b)

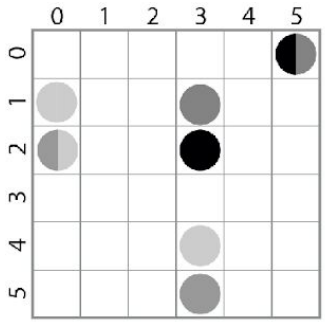


Alper, Basak, et al. "Weighted graph comparison techniques for brain connectivity analysis." *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2013.

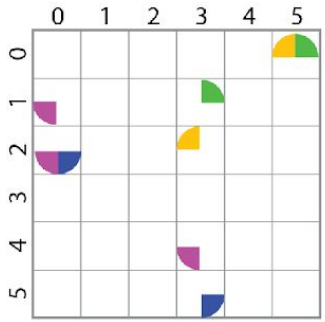


Chang, Chunlei, et al. "Evaluating perceptually complementary views for network exploration tasks." *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2017.

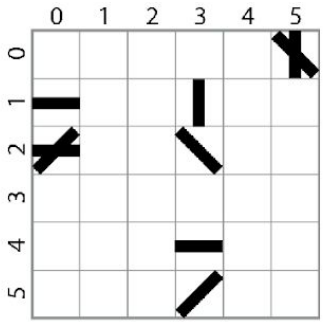
Chang, Chunlei, et al. "Evaluating perceptually complementary views for network exploration tasks." *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2017.

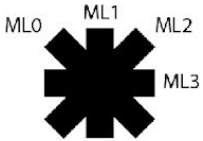


OPACITY + PIE CHART

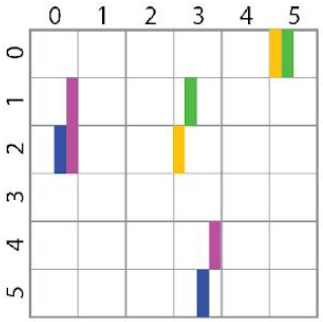


COLOUR + PIE CHART

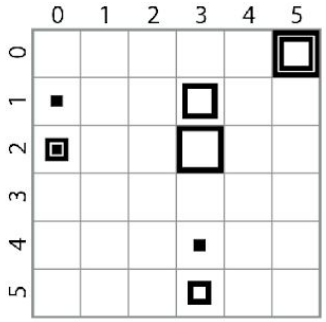




ORIENTATION



COLOUR + POSITION



SIZE

GEOGRAPHIC NETWORK VISUALISATION

About • Poster • Read the abstract

Select filters:
61 techniques (show all)

Geography Representation

Map	Distorted Map	Abstract

Network Representation

Abstract Nodes & Explicit Edges	Abstract Nodes & Abstract Edges
Explicit Nodes & Explicit Edges	Explicit Nodes & Abstract Edges

Integration

Geography as Basis	Balanced	Network as Basis

Interaction

No Interaction	Optional Interaction
Required Interaction	Interaction Technique

Origin-Destination Flow Maps in Immersive Environments

Yang, Y.; Dwyer, T.; Jenny, B.; Marriott, K.; Cordeil, M.; Chen, H. (2019) [DOI Link]

map explicit-explicit base-geo required-interaction

Revealing Patterns and Trends of Mass Mobility Through Spatial and Temporal Abstraction of Origin-Destination Movement Data

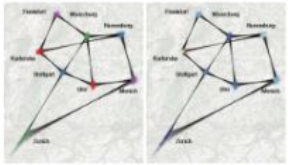
Andrienko, G.; Andrienko, N.; Fuchs, G.; Wood, J. (2017) [DOI Link]

map abstract-abstract base-geo optional-interaction

Visual Abstraction of Large Scale Geospatial Origin-Destination Movement Data

Zhou, Z.; Meng, L.; Tang, C.; Zhao, Y.; Guo, Z.; Hu, M.; Chen, W. (2019) [DOI Link]

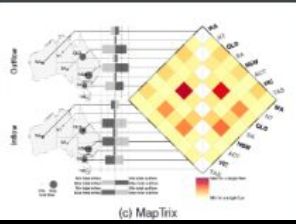
map abstract-abstract balanced required-interaction



Probabilistic Graph Layout for Uncertain Network Visualization

Schulz, C.; Nocaj, A.; Goertler, J.; Deussen, O.; Brandes, U.; Weiskopf, D. (2017) [DOI Link]

map explicit-explicit base-geo no-interaction



Animated Edge Textures in Node-Link Diagrams: A Design Space and Initial Evaluation

Romat, Hugo; Appert, Caroline; Bach, Benjamin; Henry-Riche, Nathalie; Pietriga, Emmanuel (2018) [DOI Link]

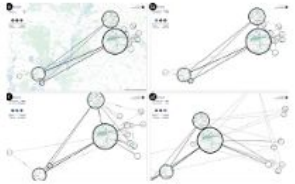
map explicit-explicit base-geo no-interaction

Module-based visualization of large-scale graph network data

Li, Chenhui; Baciu, George; Wang, Yunzhe (2017) [DOI Link]

map abstract-explicit balanced required-interaction

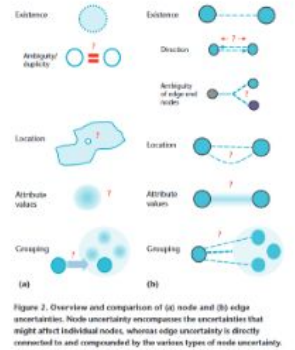
FFTEB: Edge bundling of huge graphs by the Fast Fourier Transform



Shifted Maps: Revealing spatio-temporal topologies in movement data

Otten, Heike; Hildebrand, Lennart; Nagel, Till; Dörk, Marian; Müller, Boris (2018) [DOI Link]

map abstract-explicit balanced required-interaction



Typology of Uncertainty in Static Geolocated Graphs for Visualization

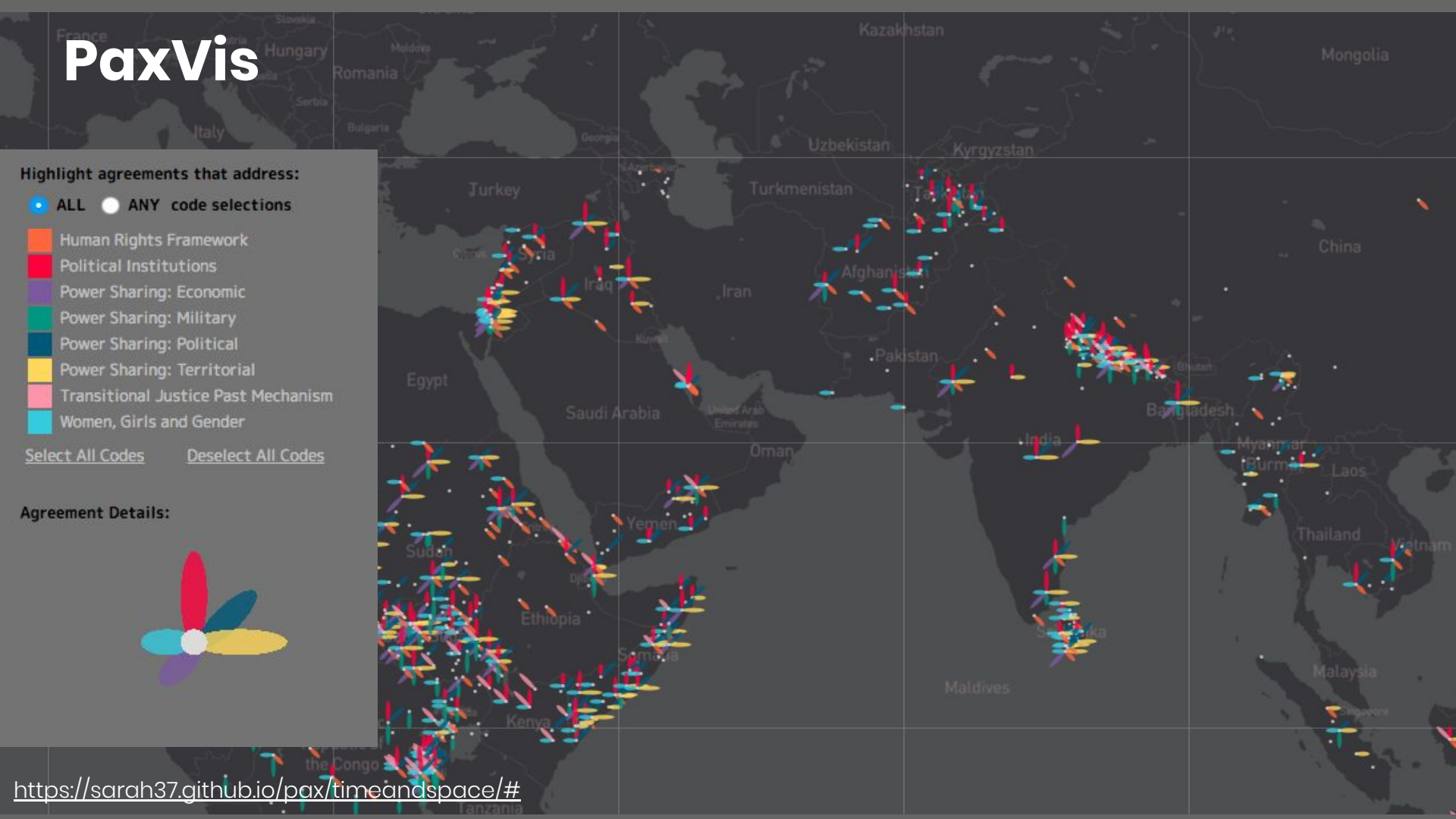
Landesberger, T. von; Bremm, S.; Wunderlich, M. (2017) [DOI Link]

map explicit-explicit base-geo no-interaction

http://geographic-networks.github.io/

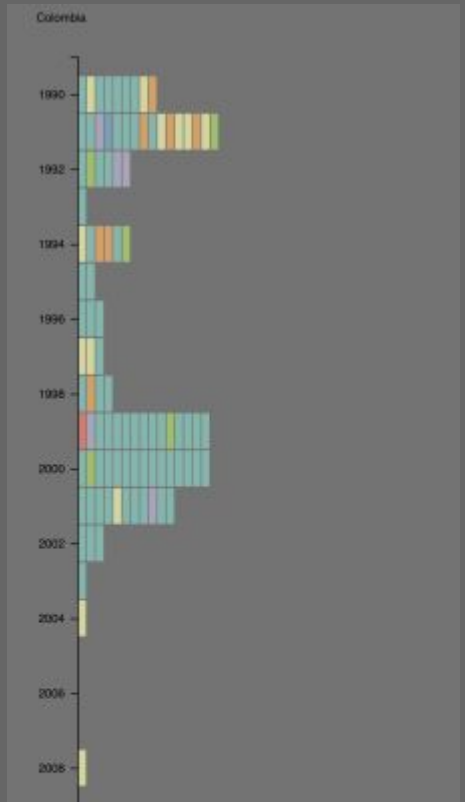
Interactive Exploration

PaxVis

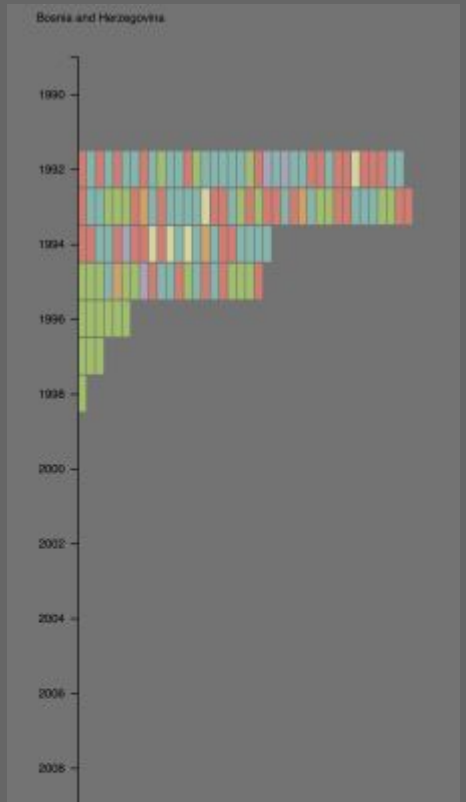


https://sarah37.github.io/pax/timeandspace/#

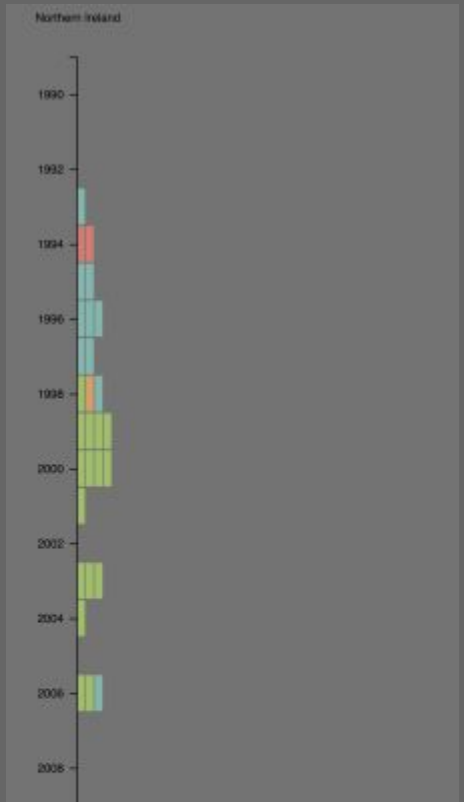
Columbia

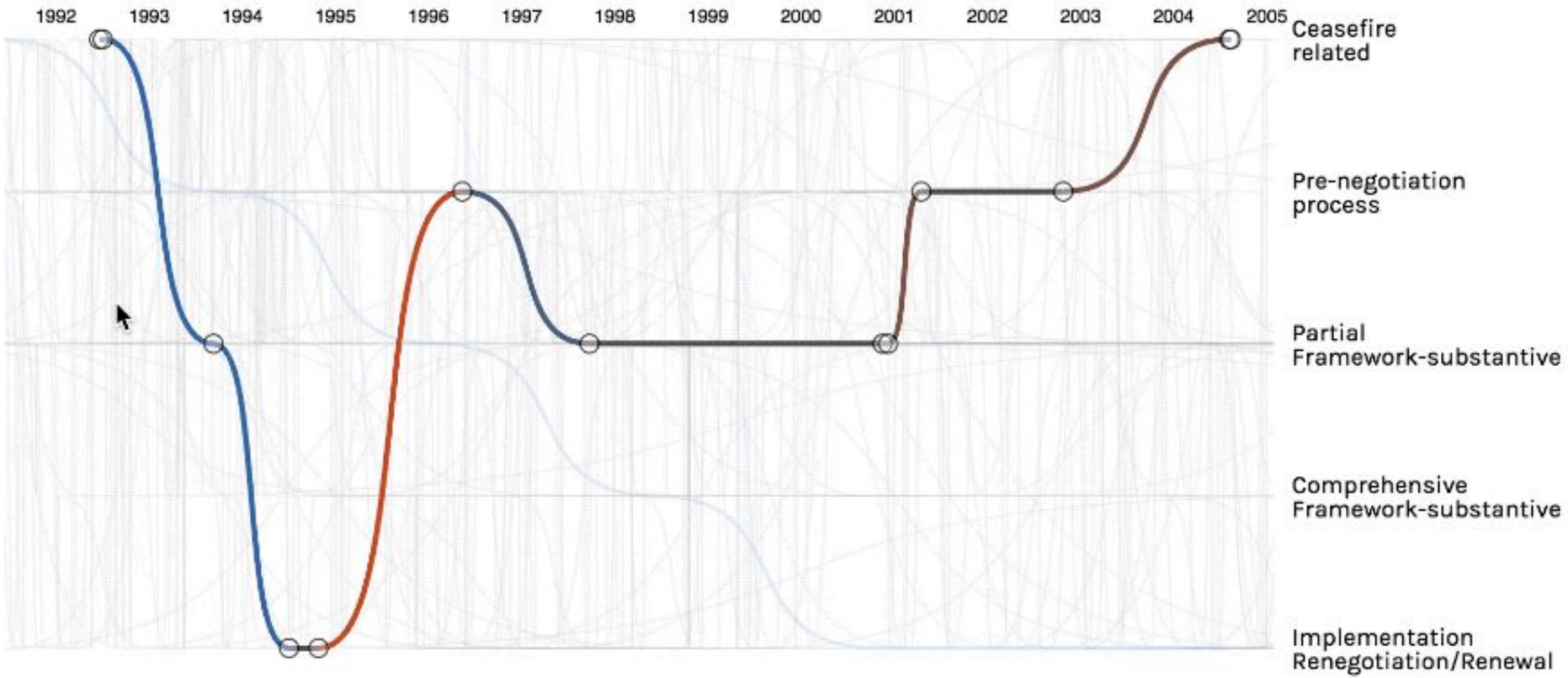


Bosnia and Herzegovina



Northern Ireland





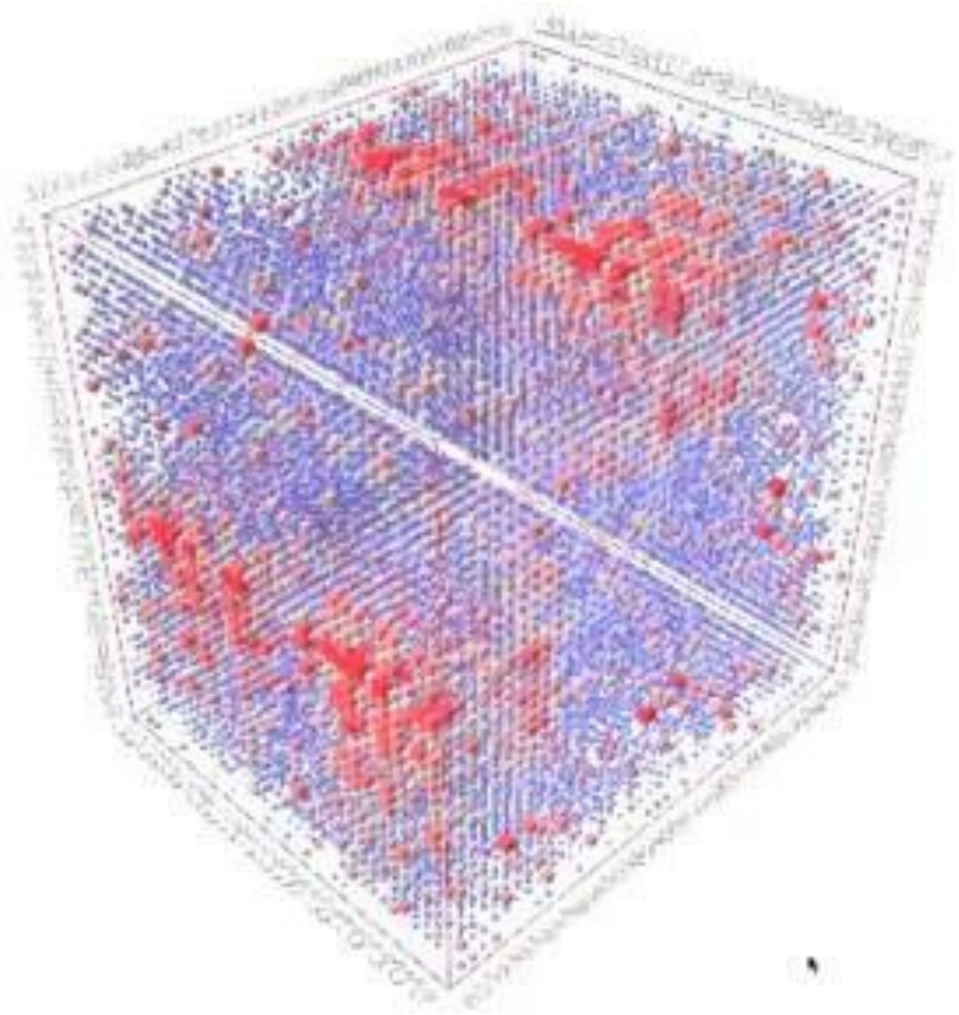
Peace Process

South Ossetia peace process

Agreement

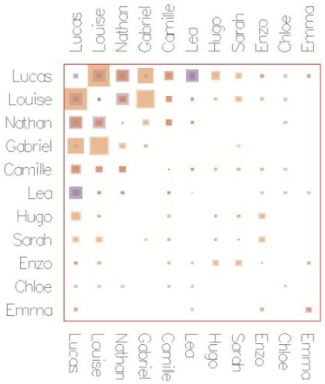
Hover circle to select agreement


Small Time Multiples

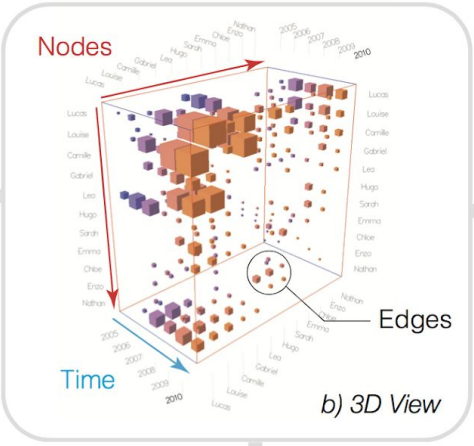


Bach, Benjamin, Emmanuel Pietriga, and Jean-Daniel Fekete. "Visualizing dynamic networks with matrix cubes." *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 2014.

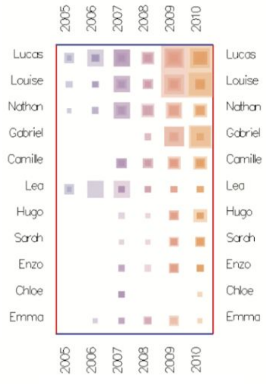
The Matrix Cube



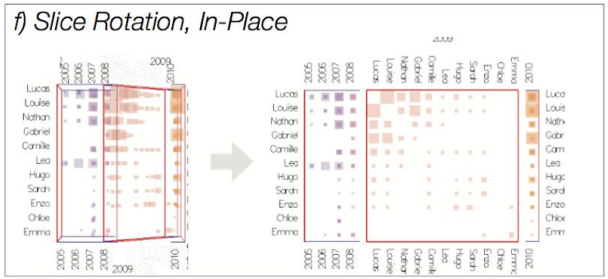
a) Rotation and Time Projection

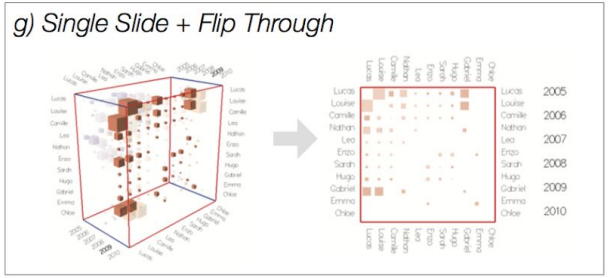


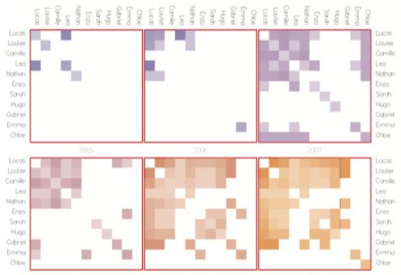
b) 3D View



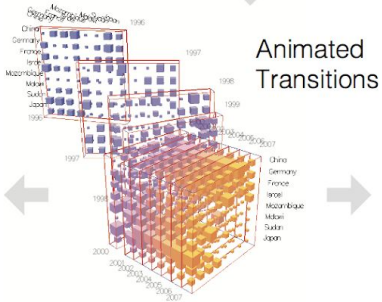
c) Rotation and Node Projection



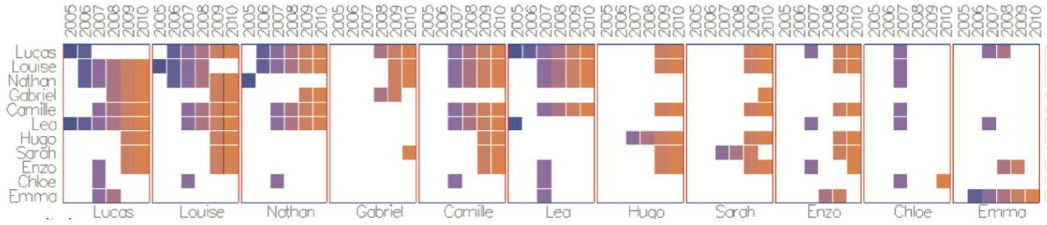




d) Time Small Multiples



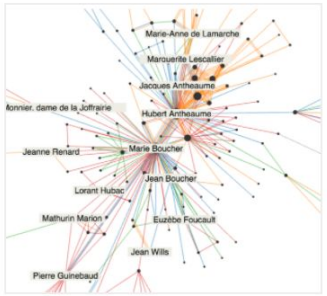
Animated Transitions

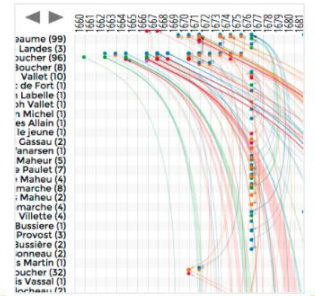


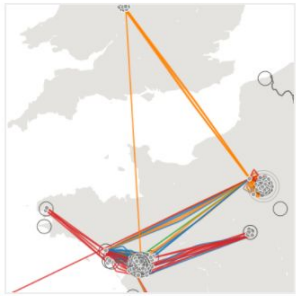
The Vistorian (beta)

https://vistorian.net

Interactive Visualizations for Dynamic and Multivariate Networks.
Free, online, and open source.





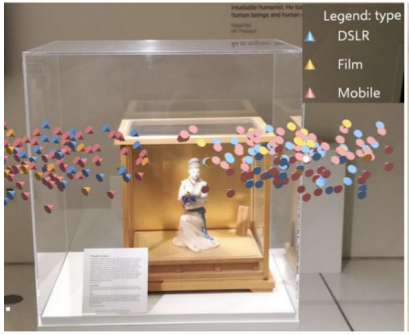


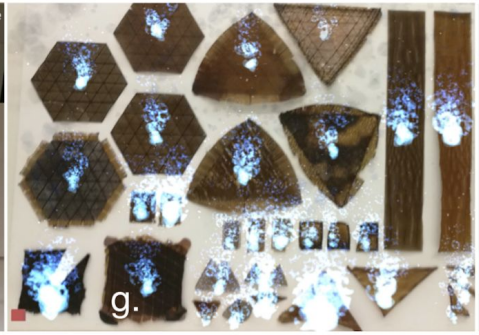
Visualizations Example Session Your Session

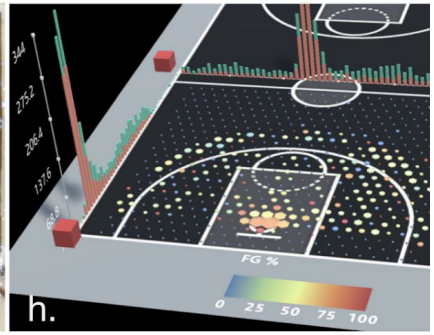
Manual Github Contact

Immersive Environments

Bach, Benjamin, et al. "The hologram in my hand: How effective is interactive exploration of 3D visualizations in immersive tangible augmented reality?." *IEEE transactions on visualization and computer graphics* 24.1 (2017): 457-467.



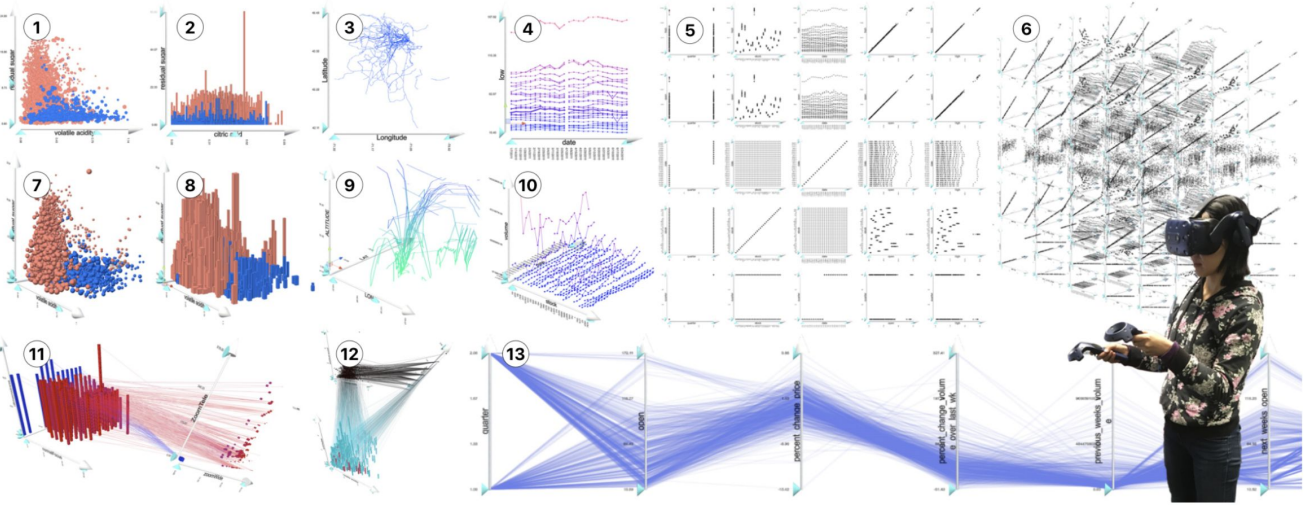




DXR

HARVARD UNIVERSITY

Sicat, Ronell, et al. "DXR: A toolkit for building immersive data visualizations." *IEEE transactions on visualization and computer graphics* 25.1 (2018): 715-725.



IATK

MONASH University

Cordeil, Maxime, et al. "IATK: An immersive analytics toolkit." *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019.

Edinburgh VisHub

A Collaborative Data Visualisation Space for Interdisciplinary Research, Teaching, and Public Engagement

Request Equipment

VisHub Labs

AR + VR
Touch + Screens
Print + Make

Navigation

Home
Goals
Projects
Activities
Get in touch

design informatics | informatics

Funded by

EPSRC
Engineering and Physical Sciences Research Council

http://edinburghvishub.github.io/

Touch, Screens, and Projection

Microsoft Surface Hub

84'' pen + touch display on movable stand, running either on battery or plug electiciy.

- 84'' 3840 x 2560 pixels display
- Intel® UHD Graphics 620
- Microsoft Surface Hub 2 Pen (active)

Link

Microsoft Surface Studio 2

28'' pen and touch display. Movable from screen-mode (vertical screen) to studio-mode (almost horizontal). Great for skeching, large screen visualizaion, collaborative touch applications.

- 28'' PixelSense™ Display with 4500 x 3000 (192 PPI) resolution, 3:2 aspect ratio, and 10 point multi-touch.
- 1TB storage
- WiFi
- 4 x USB 3.0 (one high power port), full-size SD™ card reader (SDXC) compatible, USB-C, 3.5 mm headphone jack, 1 Gigabit Ethernet port
- NVIDIA® GeForce® GTX 1060 6GB GDDR5 memory (with i7/16/1TB configuration)

Link

Microsoft Surface Book 2

13'' and 15'' laptop with detachable display and pen+touch surface. Ready for touch applications and AR.x

Specs:

- 1060 NVIDIA GeForce GTX graphics
- 16 BG RAm
- Intel Core i7 8th Gen
- 512 GB storage

Mixed Reality Meta Headset

The Meta Head-mounted display provides for stereoscopic vision, while being tethered to the computer. Calculations are performed on strong GPU hardware. Stronger graphics cards give you more performance. The field of view reasonable large (much larger than HoloLens). Good resolutions. Slow movements without lag. Interaction through simple hand-tracking. Integration possible with Leap. Special portable computer available for full mobile.

No beacons for tracking required. Environmental tracking OK but not great

Update: Unfortunately, as of early 2019, Meta has stopped working as a company however, the device is still present and working well.

Magic Leap

Magic Leap is is similar to HoloLens, but more performant with a larger field of view. Environment tracking is good. Its ~$2200 in the developer edition.

[https://www.magicleap.com]

HP Reverb

High-resolution VR head-set with controllers. Requires tethering to machine with powerful graphics card.

Link

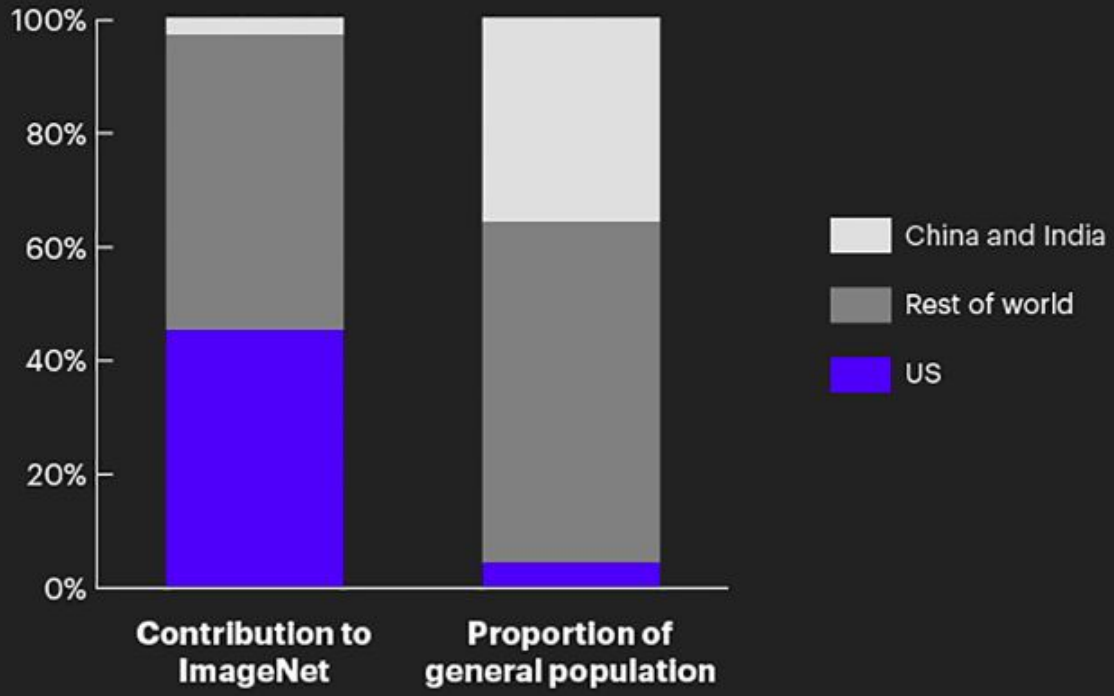
Zed Mini 2

Stereo cameras for VR head-sets,
including depth sensing, dynamic space mapping.

Understanding Machine Decisions

Even big data systems are susceptible to bias

In the ImageNew dataset, 45% of the data comes from US sources, while the US accounts for 4% of the world population



SOURCE SHANKAR 2017

Parametric Press

The Myth of the Impartial Machine

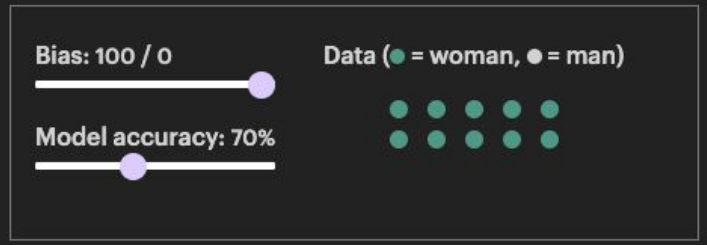
Alice Feng | Shuyan Wu

https://parametric.press/issue-01/the-myth-of-the-impartial-machine/

Algorithms can amplify bias found in data

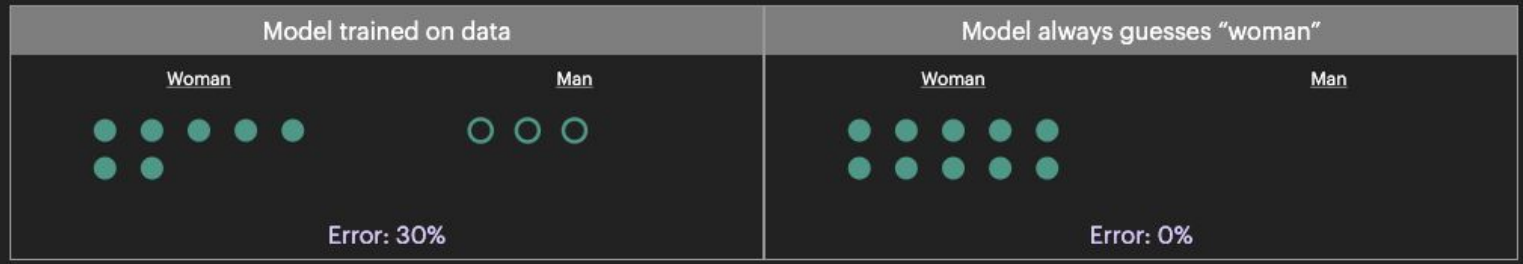
An image-classification model is being trained to identify the gender of people in cooking images. See how the composition of the data and the model accuracy influence the likelihood of amplifying bias. When is the model incentivized to predict that cooking images are always of a woman and when is it not?

Input. Adjust the sliders to set model bias and accuracy.

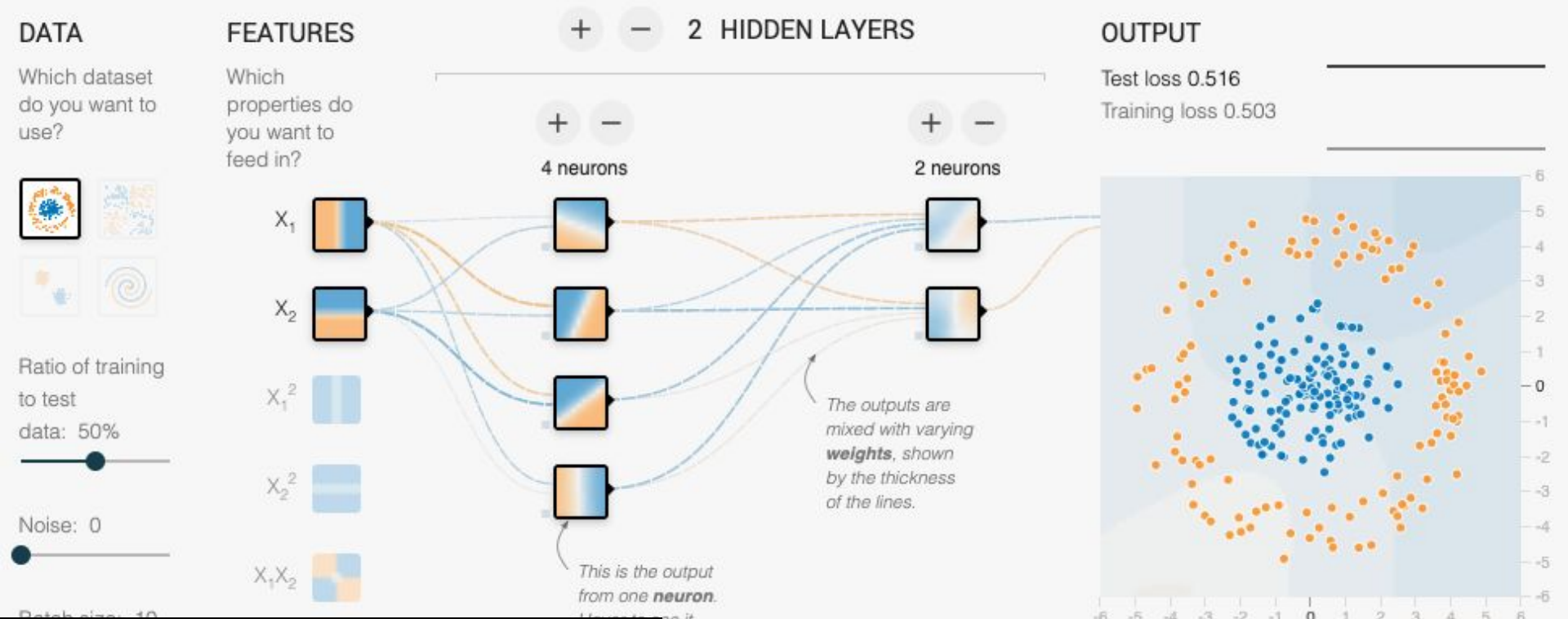


Model Predictions. Compare two models, one that is trained on the data, and one that always guesses "woman."

(○ ○ = incorrect prediction)



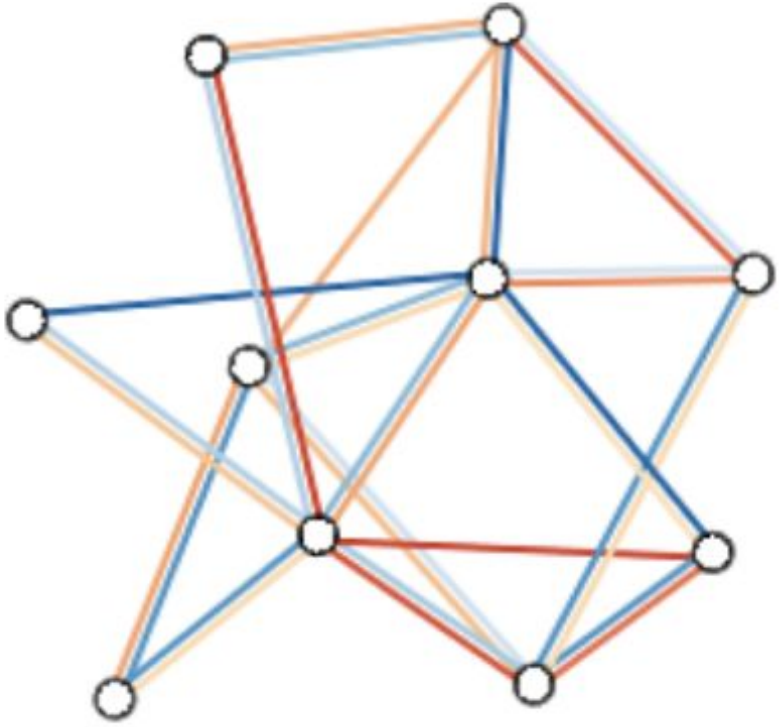
Result. Model **WOULD** be incentivized to amplify bias.

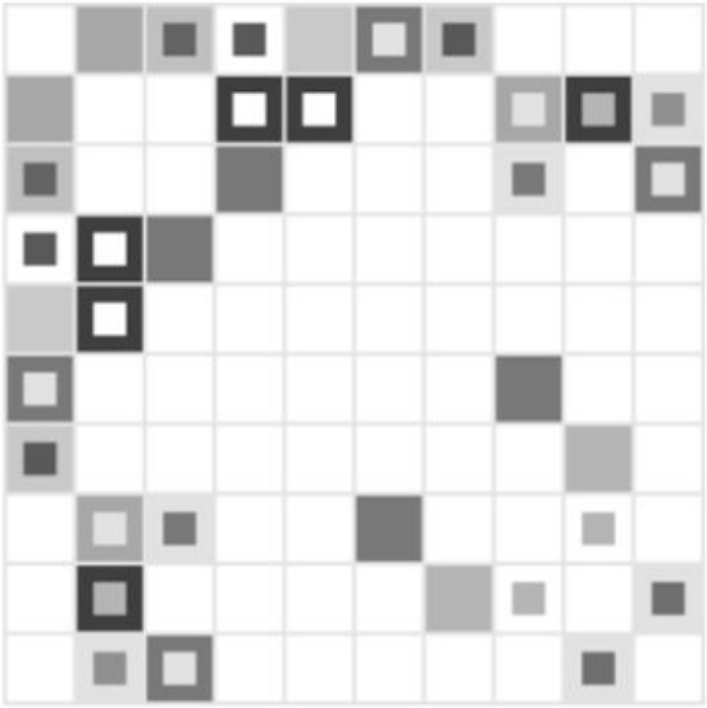


Tensorflow Playground

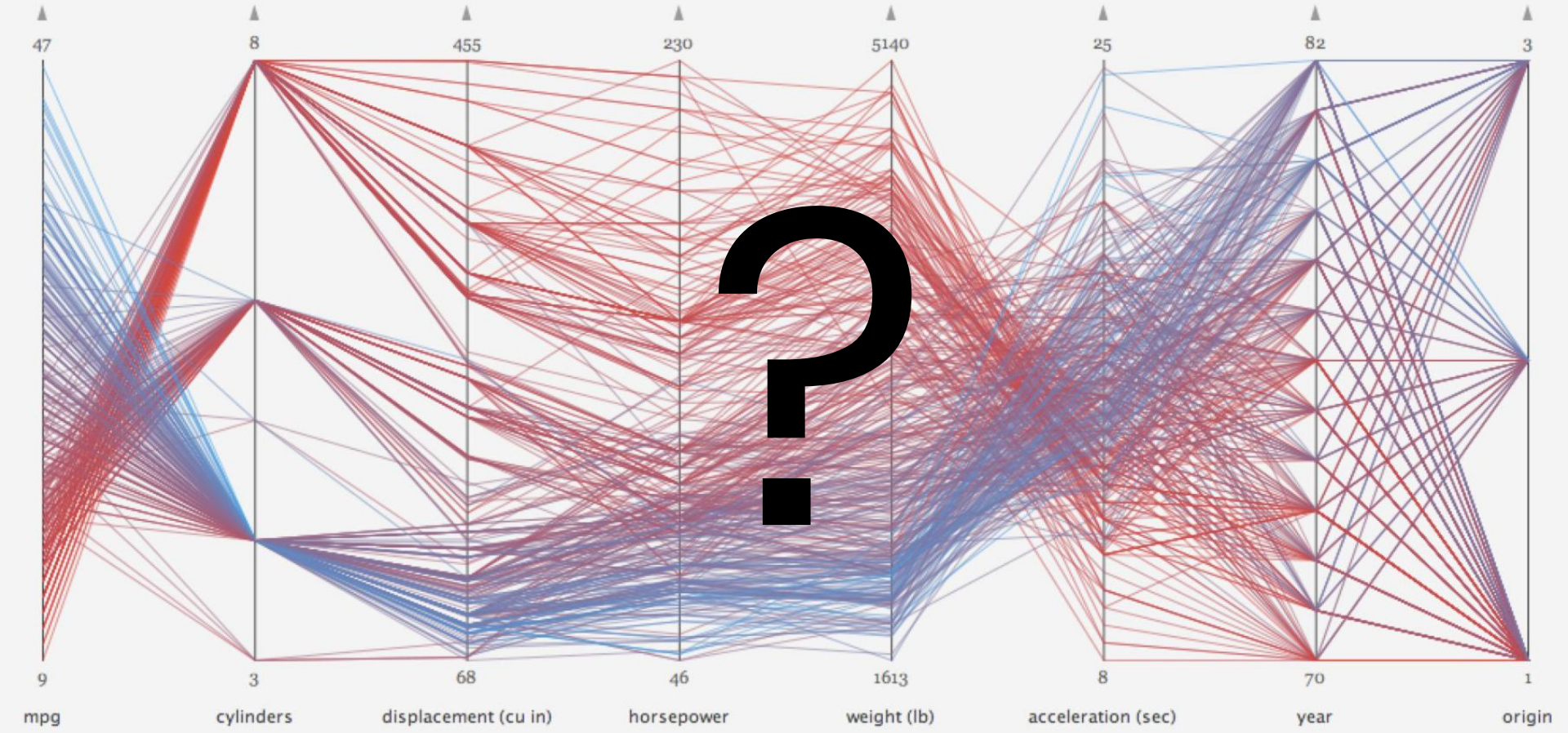
Martin Wattenberg | Fernanda Viegas

http://playground.tensorflow.org/

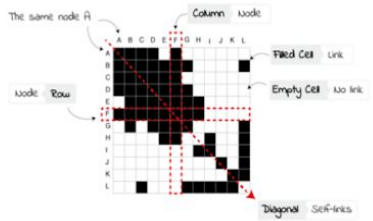




Supporting Learning and Use


47
8
455
230
5140
25
82
3
?
9
3
68
46
1613
8
70
1
mpg
cylinders
displacement (cu in)
horsepower
weight (lb)
acceleration (sec)
year
origin

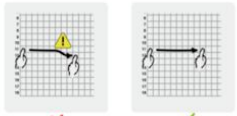
Adjacency Matrix — Anatomy



Adjacency Matrix — Pitfalls

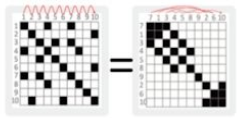
Follow a row

Be careful when following a row (or column), to not change into a neighboring one!



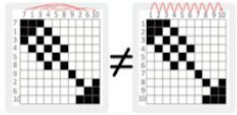
Row and column order

The order of rows and columns can matters. Sometimes it you don't see any visual pattern, they may exist after reordering.



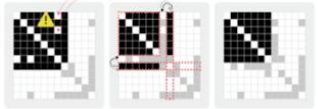
Matrices here look different, but present the same data sets.

Matrices here look similar, but present different data sets!



Cliques and ordering

Cliques can appear smaller than they are, sometimes. Empty cells within a cluster prevent us from seeing cliques.



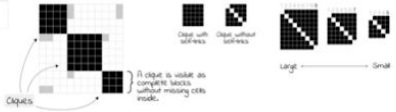
Reordering can make a large clique appearing

Adjacency Matrix — Visual Patterns

Block

Node Clique

Set of nodes where every node is connected to every other node.

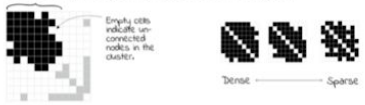


Cluster

Node Cluster

Set of nodes where almost all nodes are connected. If all links would be present, the cluster would be a clique.

A cluster is visible as a large "clump" of cells but can contain empty cells.

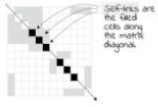


Diagonals

Self Links

Self-links are links that connect a node to itself.

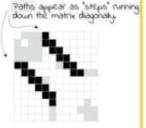
Examples include self-citations in citation networks.



Stairs

Paths

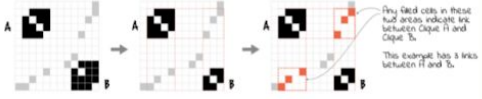
A set of nodes so that there is a set of connections that lead from the first to the last node in that set.



Off-diagonal cells

Connectors

Connectors indicate links between two cliques or clusters (A and B).

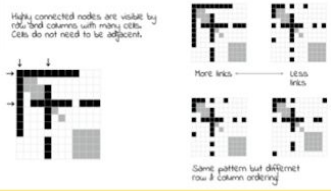


Dense row / column

Hub nodes

Highly connected nodes are visible by row and columns with many cells. Cells do not need to be adjacent.

Highly connected nodes are nodes with many connections.



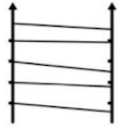
Parallel Coordinates — Visual Patterns

Parallel lines

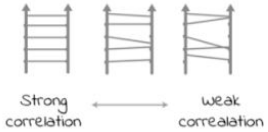
Positive Correlation

Correlations indicate that high values in on data dimensions co-occur with high values in another dimension.

Correlations are not causations!



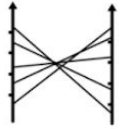
A correlation is visible through rather parallel polylines between two axes.



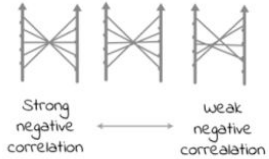
Crossing lines

Negative Correlation

Inverse correlations indicate that high values in one data dimension co-occur with low values in another data dimension.



An inverse correlation is visible through lots of crossing polylines between two axes.



Converging lines

Groups

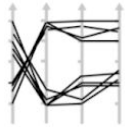
Groups indicate many elements with the same value or similar values.

Groups are visible by many lines intersecting an axis at the same position.

Grouped lines

Clusters

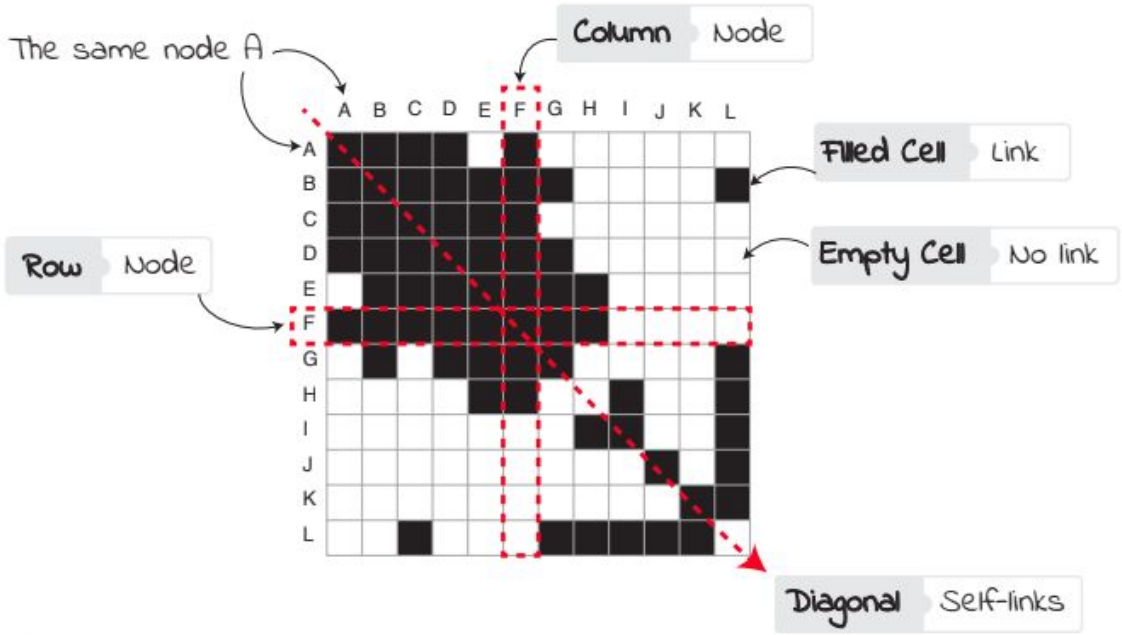
Clusters indicate data elements with similar values across several dimensions.



Clusters are visible as polylines 'following' each other across several axes, resulting in bundles.

Adjacency Matrix

Anatomy



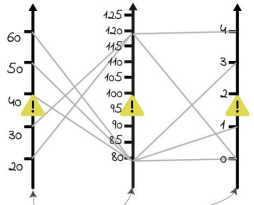
Cheat Sheets for Visualization Techniques

visualizationcheatcheets.github.io

Parallel Coordinates

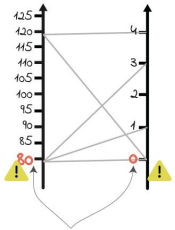
Pitfalls

Axis scales



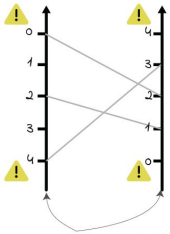
Different dimensions usually have different scales and units.

Truncated axes



Values on axes can start form values other than '0'.

Axes order



Values on axes can be either decending or acending.

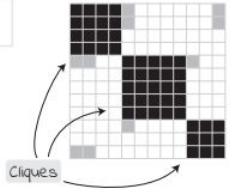
Adjacency Matrix

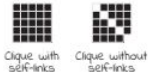
Visual Patterns

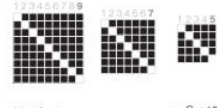
Block

Node Clique

Set of nodes where every node is connected to every other node.





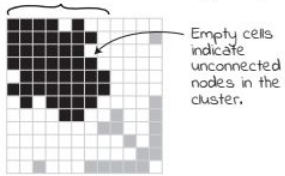


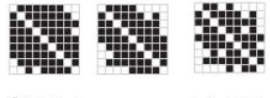
Cluster

Node Cluster

Set of nodes where almost all nodes are connected. If all links would be present, the cluster would be a clique.

A cluster is visible as a large "clump" of cells but can contain empty cells.



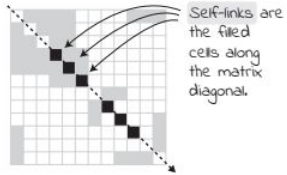


Diagonals

Self Links

Self links are links that connect a node to itself.

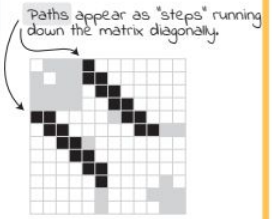
Examples include self-citations in citation networks.



Stairs

Paths

A set of nodes so that there is a set of connections that lead from the first to the last node in that set.



Cheat Sheets for Visualization Techniques

visualizationcheatcheets.github.io

https://visualizationcheatsheets.github.io/

Cheat Sheets for Visualization Techniques

visualizationcheatsheets.github.io

Home
Download all (PDF+PNG), 76MB
Download guidelines (PDF), 8MB

Paper

Cheat Sheets for Data Visualization Techniques: Zezhong Wang, Lovisa Sundin, Dave Murray-Rust, Benjamin Bach, *ACM Conference on Human Factors in Computing Systems (CHI), 2020*

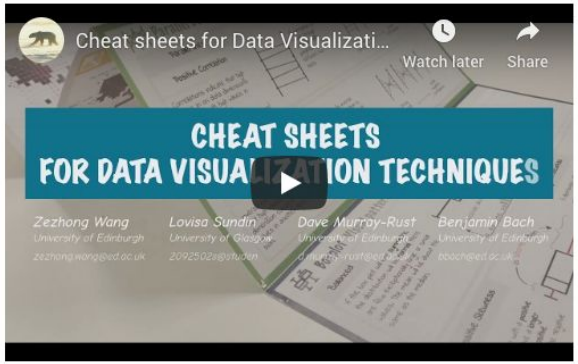
By Type

Anatomy | Introduction | Construction
Visual Pattern | Pitfalls
Well-known Relative | False Friends

By Visualization

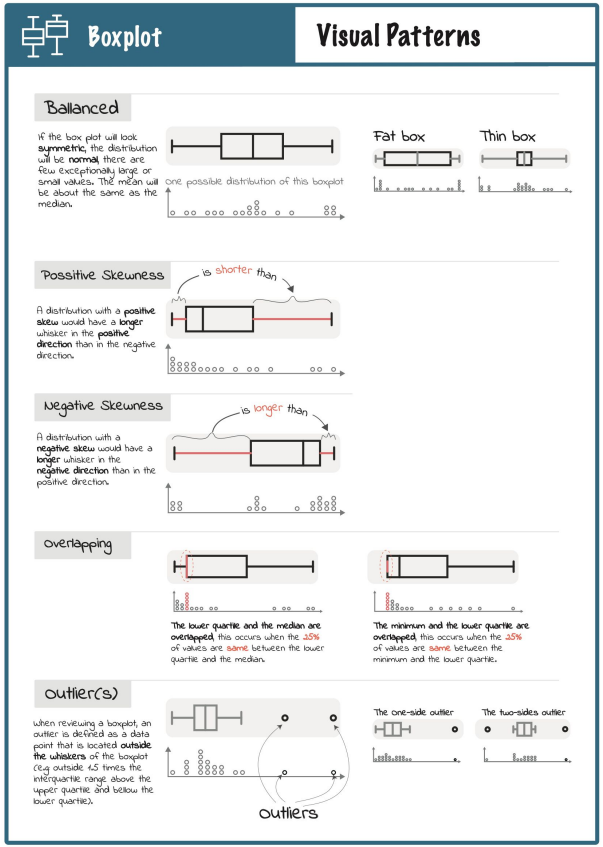
- Boxplots
- Confluence Graphs
- Adjacency Matrix
- Parallel Coordinates
- Time Curve
- Treemap

Cheat Sheets for Visualization Techniques



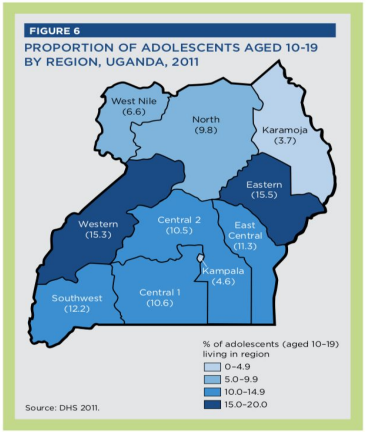
Cheat sheets are sets of *concise graphical and textual explanations*, inspired by infographics, data comics, and cheat sheets in other domains. Cheat sheets aim to support learning, teaching, and the regular use of both common and novel visualization techniques in a variety of contexts. To design cheat sheets for visualization techniques, we describe six components of a cheat sheet: anatomy, build-up, visual patterns, pitfalls, false-friends, variations. We present examples for several visualization techniques, created through an iterative design process which involved data science and visualization teachers, visual designers and students. In a qualitative and iterative user study, we gather subjective feedback from participants, show readability and usefulness of our cheat sheets, and iterated on their design. We bring this together as a design methodology, with a comprehensive design framework to easily create cheat sheets for additional visualizations.

By Type



Visual Storytelling

physical, social, political, and economic structures of a region can place residents at varying risks for vulnerability. Areas susceptible to violence or natural disaster pose clear threats to individuals. An individual's environment also affects his or her development and behavioral choices. Resources available in the physical and social environments create the contexts within which decisions are made about health, education, and employment. Political and social environments also dictate whether resources are accessible to all adolescents. An examination of the residential distribution of adolescents provides a baseline for comparing geographical patterns of vulnerability. Within Uganda, by type of residence, the majority of adolescents (87 percent) live in rural versus urban areas. Figure 6 shows the distribution of adolescents aged 10 to 19 living in Uganda. Regional distributions show Karamoja contains only four percent of the adolescent population. Kampala with a much denser population contains 4.6 percent of the population. The Eastern and Western regions contain the largest proportions of the adolescent population.



Household factors influencing vulnerability

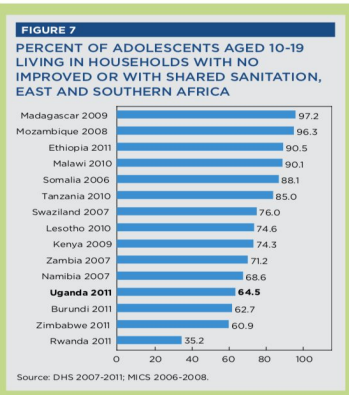
Household-level factors have direct impacts on the well-being of adolescents. Households are the primary setting where adolescents live and engage in activities. For this reason, the household environment and the people who live there have significant impacts on the lives of adolescents. Physical conditions of the home influence the health of residents. Family structures and demographic characteristics of household members affect the knowledge, decisions, behaviors and interactions in the environment of the adolescent.

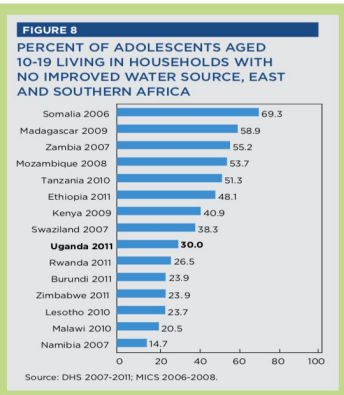
Access to improved water sources and sanitation

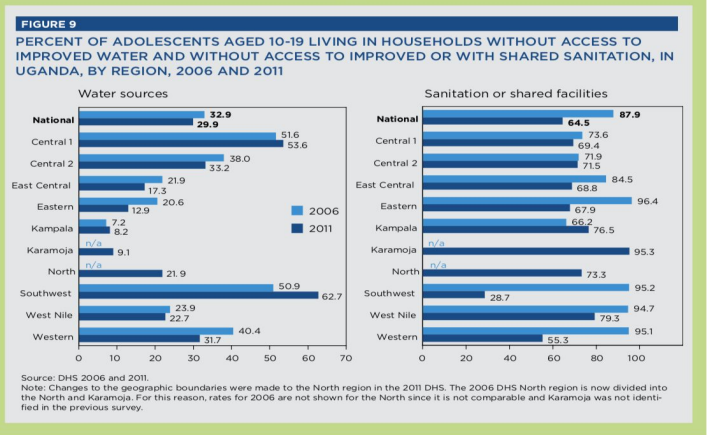
Unsafe water, inadequate sanitation, and poor hygiene are among the five leading risk factors responsible for one quarter of all deaths in the world (WHO 2009). Unsafe water supplies and inadequate sanitation in homes increase exposure to water-borne diseases and can cause diarrhea. Ensuring access to clean water sources and sanitation is key to maintaining hygiene and health. Improved water sources are those that either naturally protect water from contamination or are constructed to do so. These include piped water, public taps, standpipes, boreholes, tube wells, protected wells and springs, and rainwater collection. Improved sanitation includes constructs and systems that prevent fecal contamination. These include flush or pour toilets, ventilated pit latrines, pit latrines with slabs, and composting toilets (UNICEF 2013b).

Housing conditions across East and Southern Africa are largely in need of improvement, and lack of improved sanitation varies by country. In nearly all of East and Southern Africa, over half of adolescents either do not have improved sanitation or share facilities with other households. Conditions are worst in Madagascar and Mozambique where fewer than four percent of adolescents live in households with improved sanitation that is not shared (Figure 7). Rwanda has the lowest proportion of adolescents affected—35 percent—which is still unacceptably high. Lack of access to improved water sources affects lower proportions but is still a problem in the region. In five countries, fewer than half of adolescents have access to improved water sources (Figure 8). Water conditions are best in Namibia, where only 15 percent of adolescents have no access to improved water.

In Uganda, overall access to improved water and sanitation increased by a small but significant percentage between 2006 and 2011 (Figure 9). In 2006, 33 percent of adolescents had no access to improved water; in 2011, it is 30 percent. The proportion of adolescents without access to improved





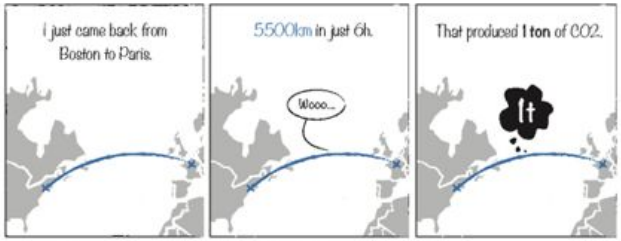


Bach, Benjamin, et al. "Design patterns for data comics." *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018.

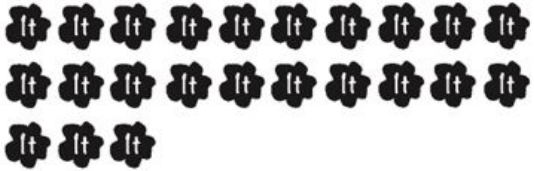
Wang, Zezhong, et al. "Comparing Effectiveness and Engagement of Data Comics and Infographics." *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 2019.

Bach, Benjamin, et al. "The emerging genre of data comics." *IEEE computer graphics and applications* 37.3 (2017): 6-13.

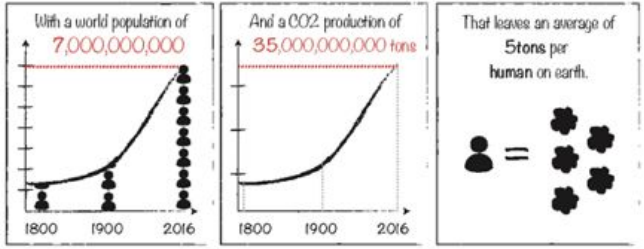
CO Footprint

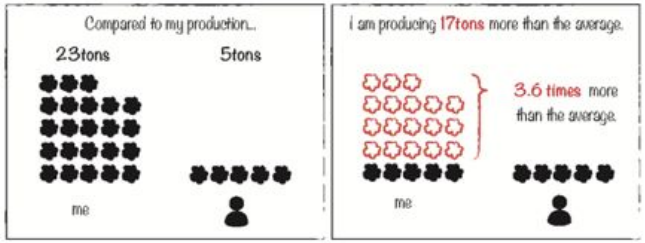


Which produced...

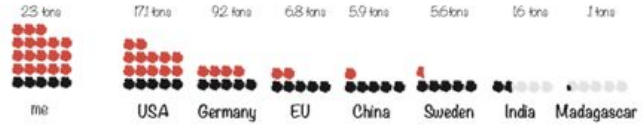


23 tons of CO2.

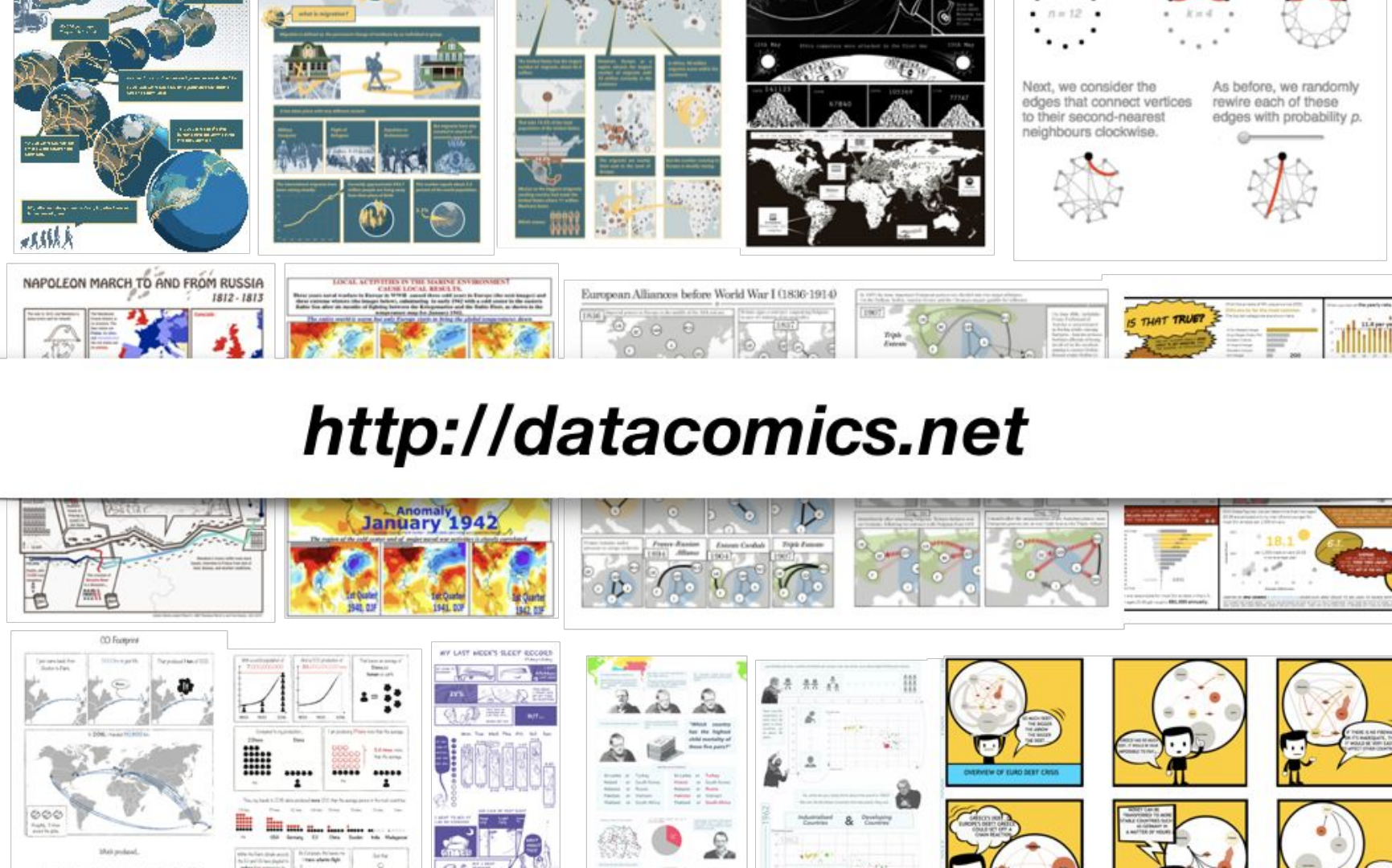




Thus, my travels in 2016 alone produced **more** CO2 than the average person in the most countries:



Wang, Zezhong, Harvey Dingwall, and Benjamin Bach. "Teaching Data Visualization and Storytelling with Data Comic Workshops." *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 2019.



http://datacomics.net

Challenges

- How to visualize **complex data**?
- How do we best leverage human **perception**?
- How do we provide for powerful **interaction**?
- How to **communicate** with visualizations?
- How to build efficient **tools**?
- How to make visualizations **understandable**?

Edinburgh Data Visualization Meetup

Edinburgh, United Kingdom
705 members · Public group
Organized by **Brendan Hill** and **2 others**

Share:

About Events Members Photos Discussions More

Manage group Create event

What we're about

Anyone interested in the process and product of more effectively and ingeniously communicating the meanings inherent in data by visual or other...

Read more

Upcoming events (1)

See all

THU, JAN 30, 6:00 PM

Edinburgh Data Visualization Meetup 12

Cirrus Logic

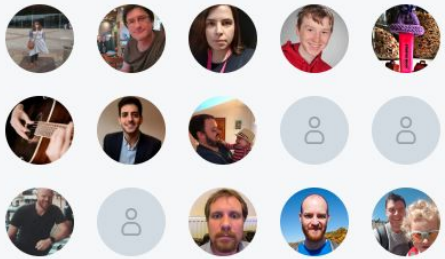
I See What You Mean - two ways vision helps us understand speech Gordon McLeod & Ben Hopson of Cirrus Logic (www.cirrus.com) Speaker identification

Organizers

Brendan Hill and **2 others**
Message

Members (705)

See all



https://
visualinteractivedata.
github.io

Benjamin Bach
http://benjbach.me
@benjbach