

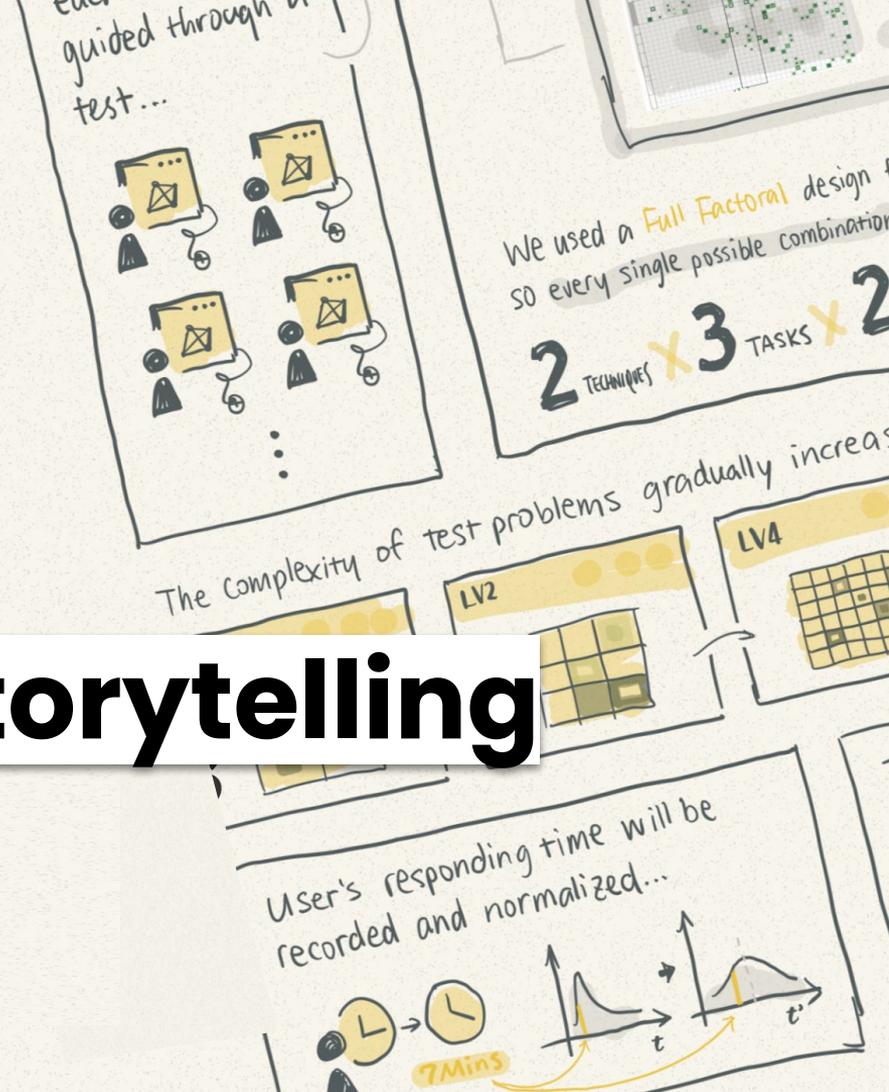
Data Comics for Data-Driven Storytelling

Benjamin Bach

University of Edinburgh



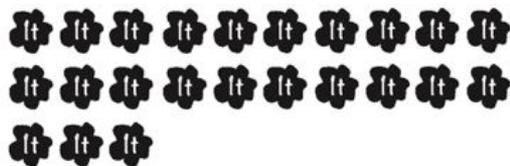
Visual+
Interactive
Data



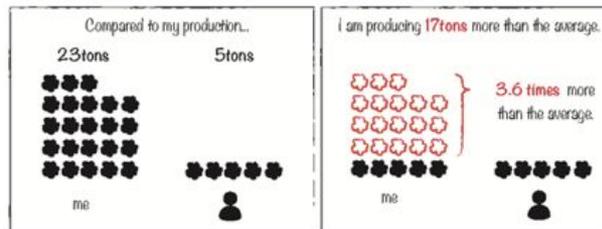
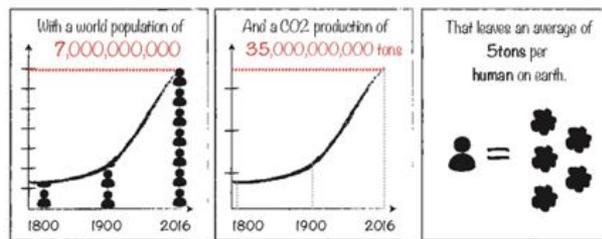
CO Footprint



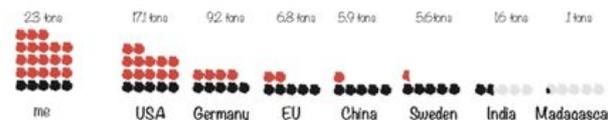
Which produced...



23 tons of CO₂.



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





Visual+
Interactive
Data



THE UNIVERSITY
of EDINBURGH



University
of Glasgow



Microsoft
Research



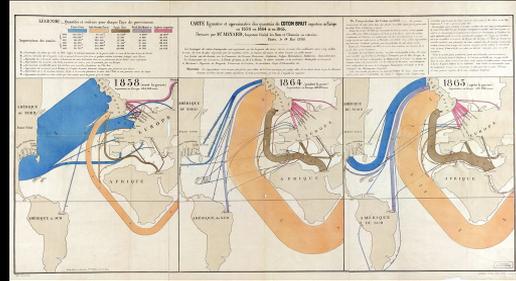
UNIVERSITY OF
TORONTO



HARVARD
UNIVERSITY

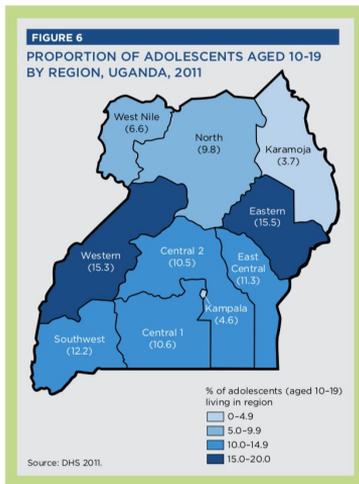


UNIVERSITY OF
CALGARY



What are data comics?

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

FIGURE 7
PERCENT OF ADOLESCENTS AGED 10-19 LIVING IN HOUSEHOLDS WITH NO IMPROVED OR WITH SHARED SANITATION, EAST AND SOUTHERN AFRICA

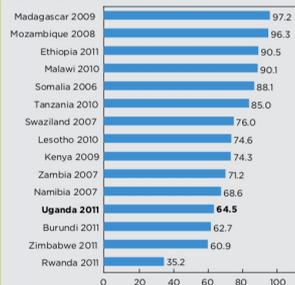


FIGURE 8
PERCENT OF ADOLESCENTS AGED 10-19 LIVING IN HOUSEHOLDS WITH NO IMPROVED WATER SOURCE, EAST AND SOUTHERN AFRICA

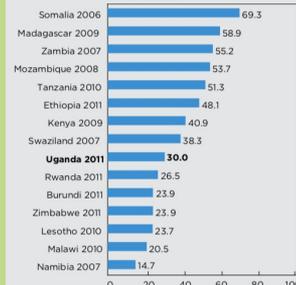
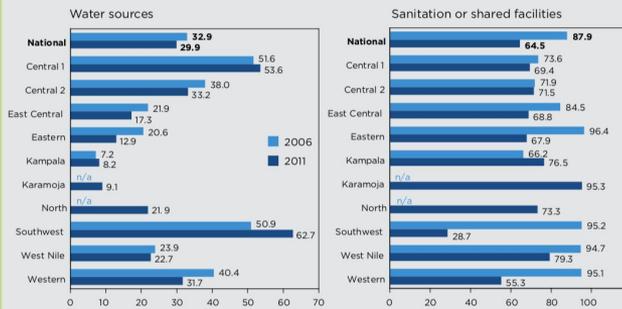
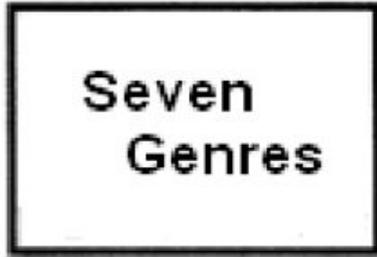


FIGURE 9
PERCENT OF ADOLESCENTS AGED 10-19 LIVING IN HOUSEHOLDS WITHOUT ACCESS TO IMPROVED WATER AND WITHOUT ACCESS TO IMPROVED OR WITH SHARED SANITATION, IN UGANDA, BY REGION, 2006 AND 2011

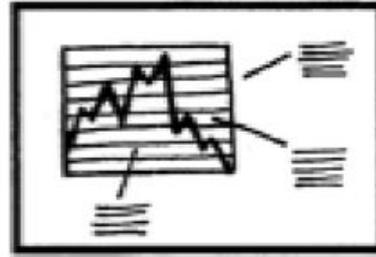


Note: Changes to the geographic boundaries were made to the North region in the 2011 DHS. The 2006 DHS North region is now divided into the North and Karamoja. For this reason, rates for 2006 are not shown for the North since it is not comparable and Karamoja was not identified in the previous survey.

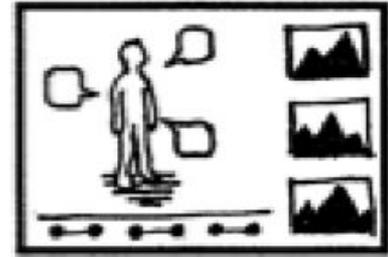
Data Comics



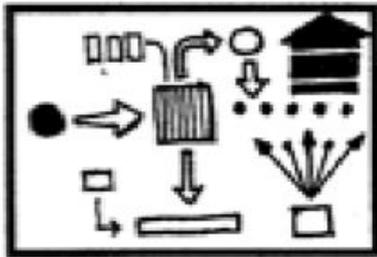
Magazine Style



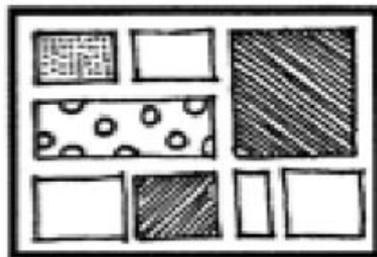
Annotated Chart



Partitioned Poster



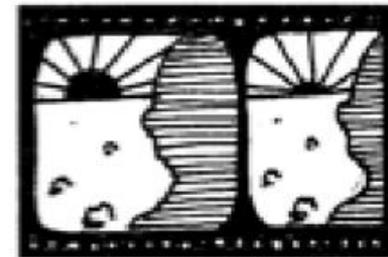
Flow Chart



Comic Strip

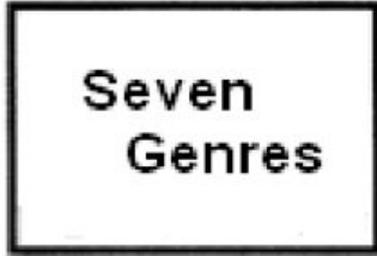


Slide Show

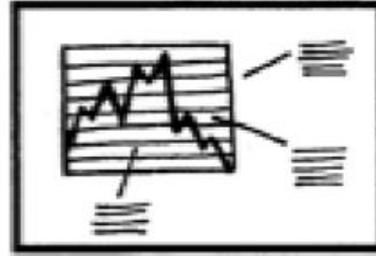


Film/Video/Animation

Data Comics



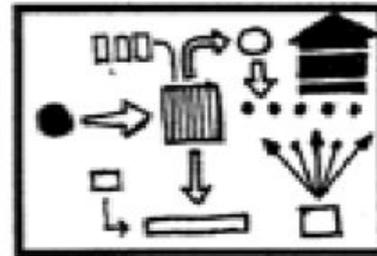
Magazine Style



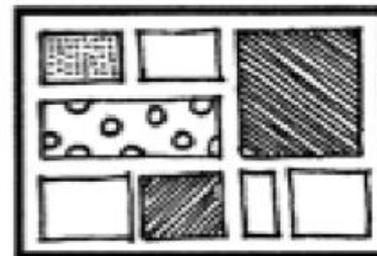
Annotated Chart



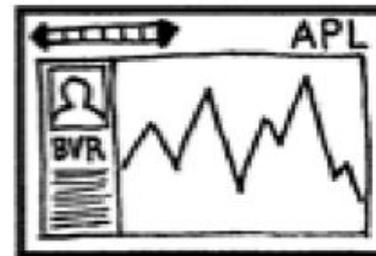
Partitioned Poster



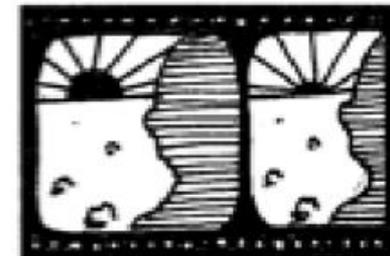
Flow Chart



Comic Strip

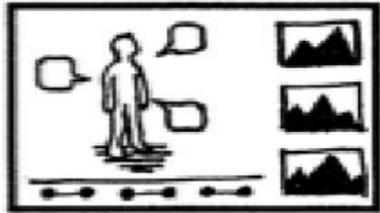


Slide Show



Film/Video/Animation

Partitioned Poster (Infographic)



Partitioned Poster

<https://www.visualcinnamon.com/portfolio/baby-spike>

GRAPHIC SCIENCE

The Average
7.3 babies born
per minute

Eat First
More births of all types
occur right after lunch

Day Shift

Births peak around 8 A.M.,
then rise again between noon
and 1 P.M. Hospitals typically
have more doctors and nurses
on hand during the morning
and fewer later in the day.

The Average
447 babies born per hour

Fewest Births
Sunday night
between 2 and 3 A.M.

Early Riser

More babies than average
are born on weekdays during
daylight hours. Fewer are
born on weekends or at night,
primarily because fewer
hospital staffers are on duty,
so women tend not to schedule
their delivery then. Despite
folklore, a full moon has no effect.

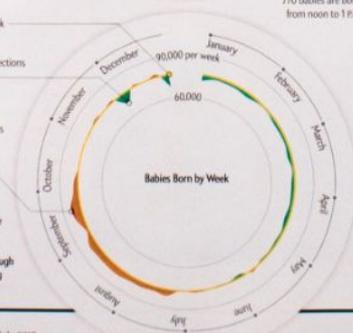
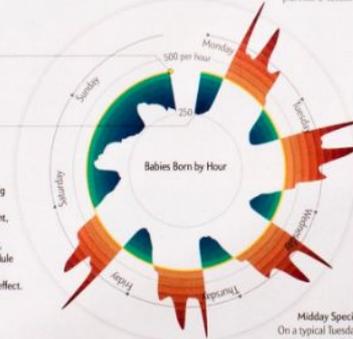
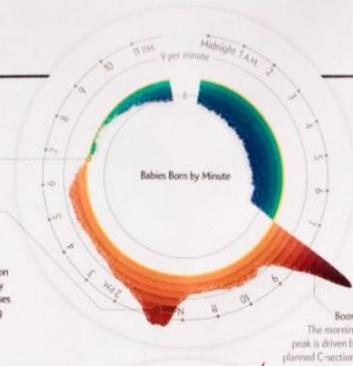
The Average
70,000 babies born per week

No, Thanks
Moms do not schedule C-sections
around Thanksgiving

Happy Holidays
Babies seem to arrive
nine months after Christmas
and New Year's Eve

SUMMER SON

Evidently, more people have
sex during colder months,
leading to more births nine
months later from July through
October, and less sex during
warmer months.



The Baby Spike

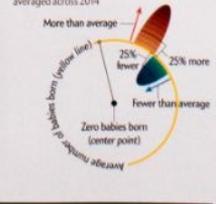
Births peak on
weekdays during
daytime work hours

Two generations ago babies were born pretty much spontaneously, around the clock. But today in the U.S., about half of all births are cesarean sections prescheduled by Mom or deliveries induced by doctors concerned about the mother's or baby's health. These medical procedures have skewed the days of the week, and hours of the day, during which those little bundles of joy arrive.

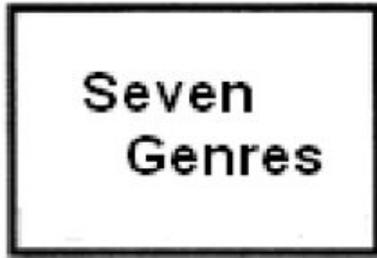
The procedures dominate because more than 98 percent of infants are born in a hospital, despite what seems to be the rising popularity of home births. Far more babies now arrive on weekdays than on weekends, most between 8 A.M. and 6 P.M. "We can't schedule spontaneous labor, obviously," says Neel Shah, a physician and professor at Harvard Medical School. "But we can schedule delivery."

—Mark Fischetti and Zan Armstrong

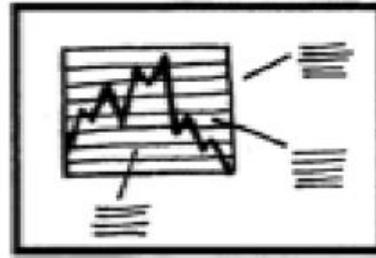
Each graph shows U.S. data averaged across 2014



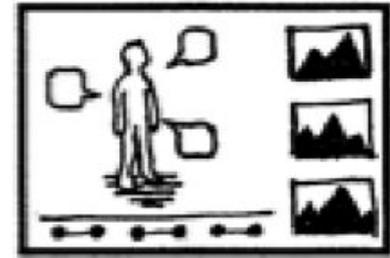
Data Comics



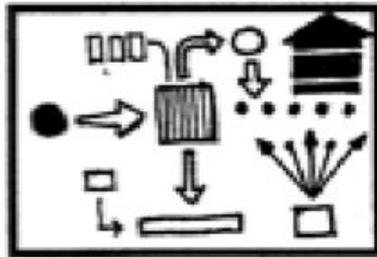
Magazine Style



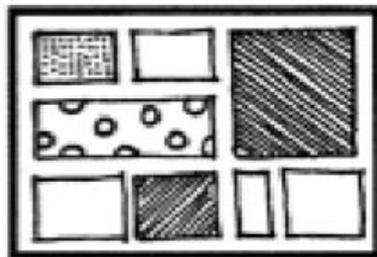
Annotated Chart



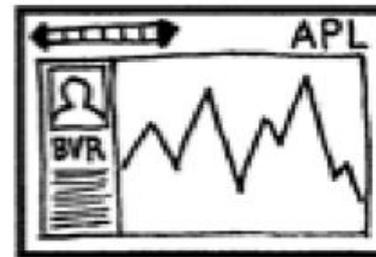
Partitioned Poster



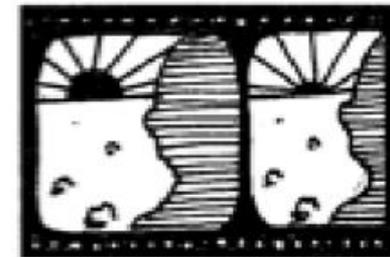
Flow Chart



Comic Strip

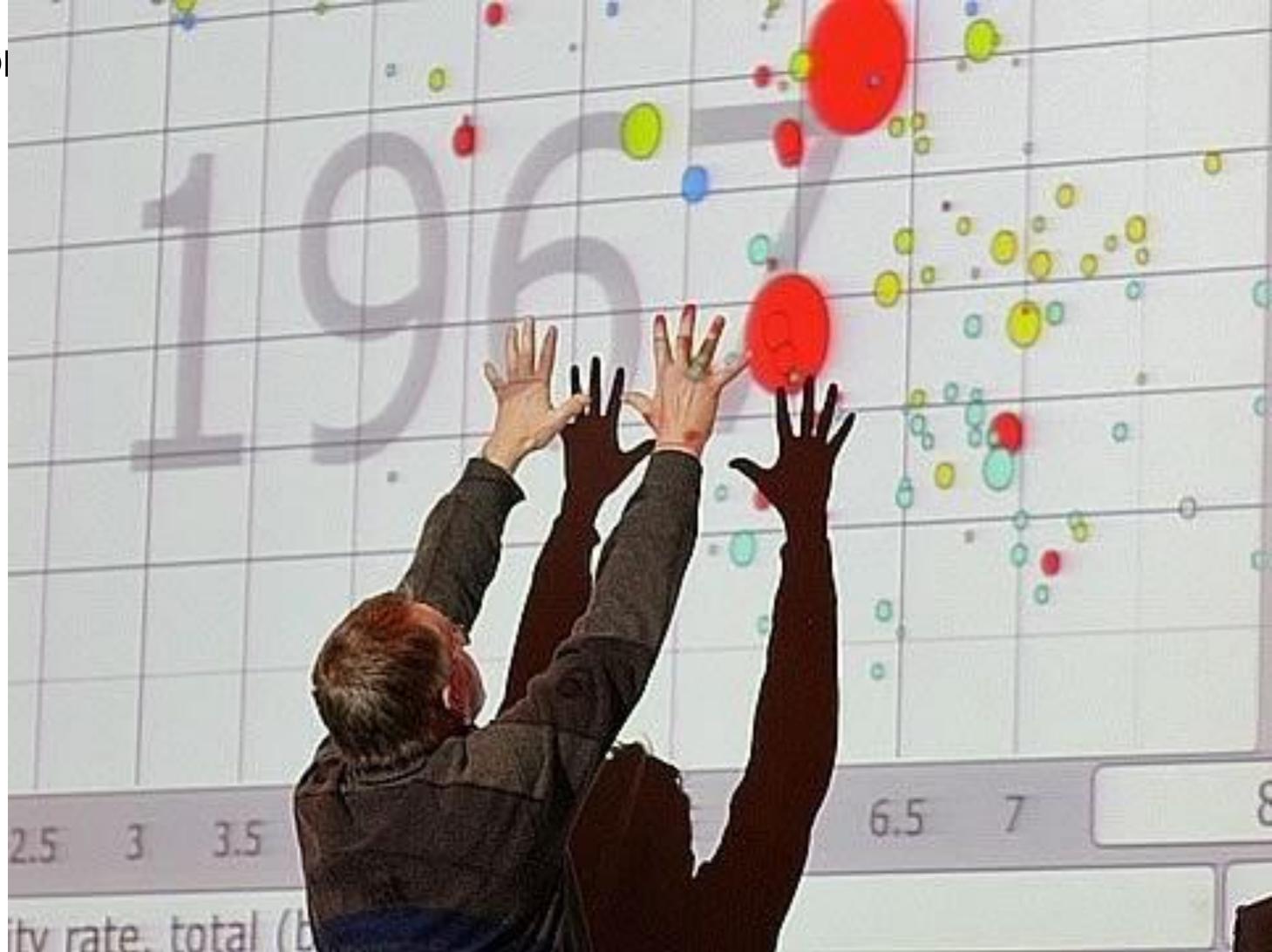


Slide Show



Film/Video/Animation

Life Presentation



https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen?language=en

Data Videos

INEQUALITY
IN
AMERICA



<https://www.youtube.com/watch?v=QPKKQnijnsM>

<https://www.youtube.com/watch?v=QPKKQnijnsM>

Space vs. Time oriented formats

Space oriented

Use space to structure information

Large and detailed images

Space == importance

Encourage exploration

Adapt to readers' pace

> Reader driven

Time oriented

Use time to structure information

Sequences of images

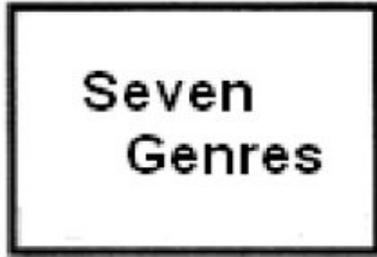
Time+order == importance

Encourage explanation

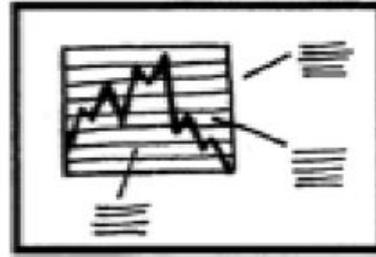
Follow author's narrative

> Author driven

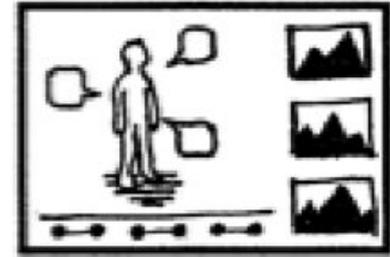
Data Comics



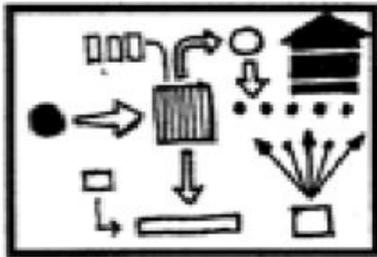
Magazine Style



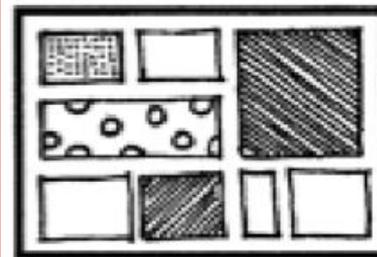
Annotated Chart



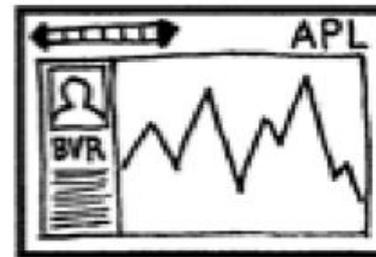
Partitioned Poster



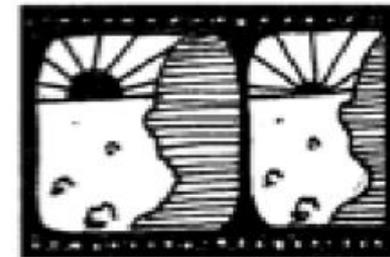
Flow Chart



Comic Strip

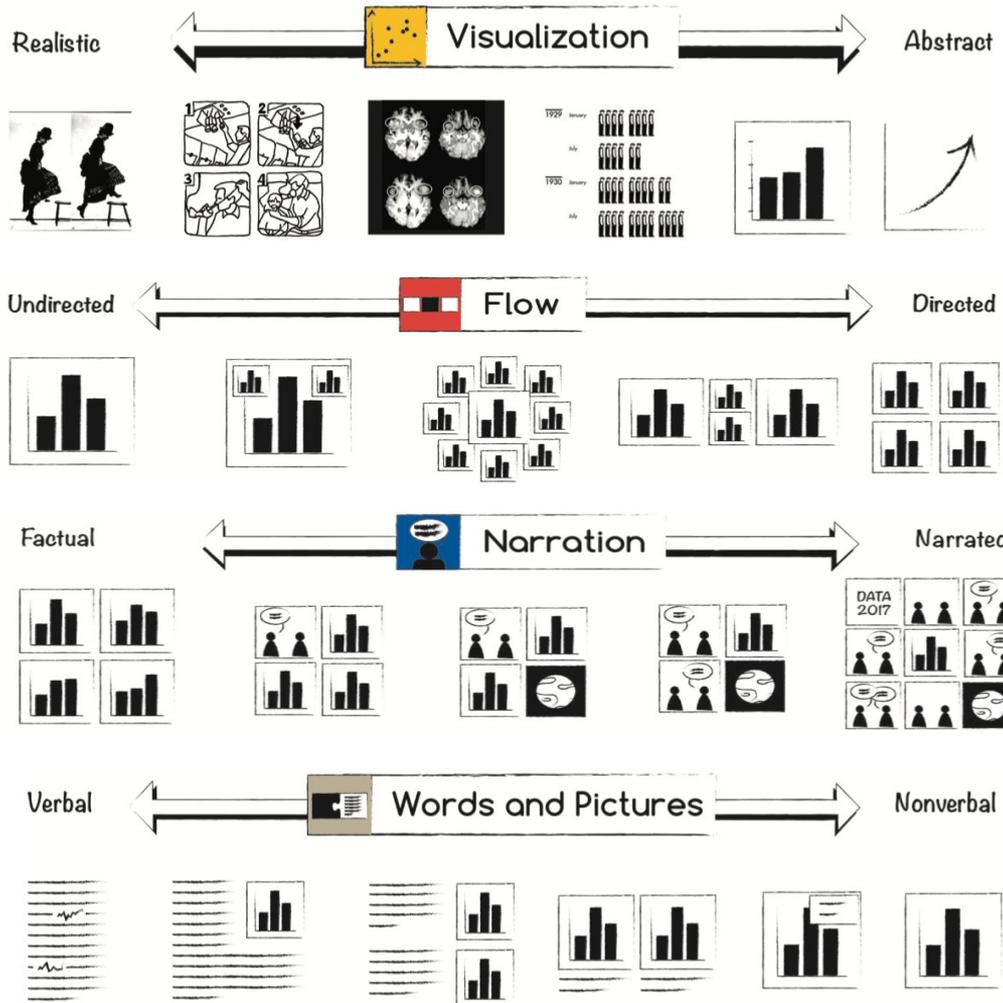


Slide Show



Film/Video/Animation

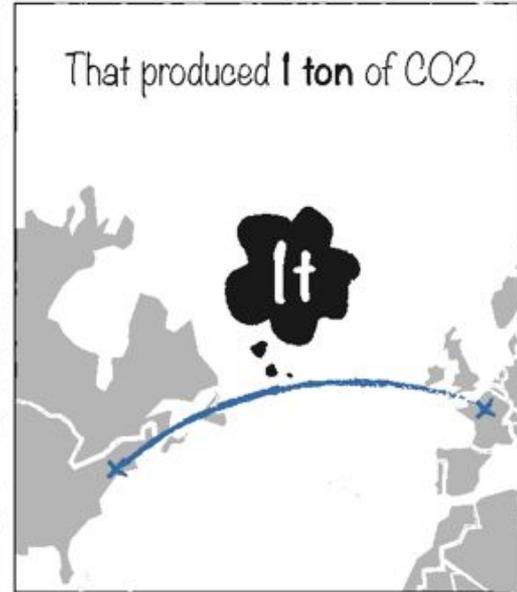
Dimensions of data comics



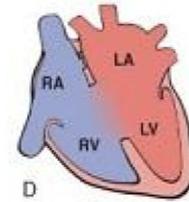
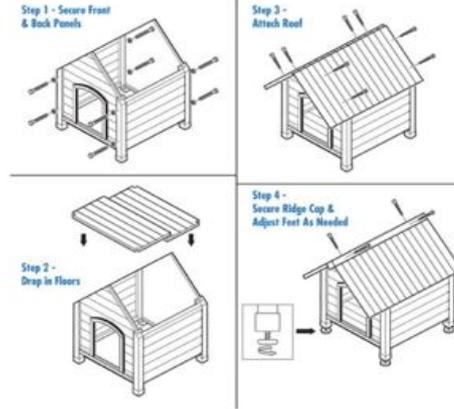
Panel (sequences)



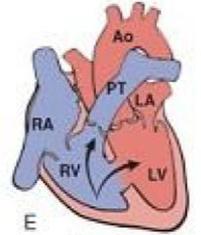
Panel (sequences)



Sequential Explanations

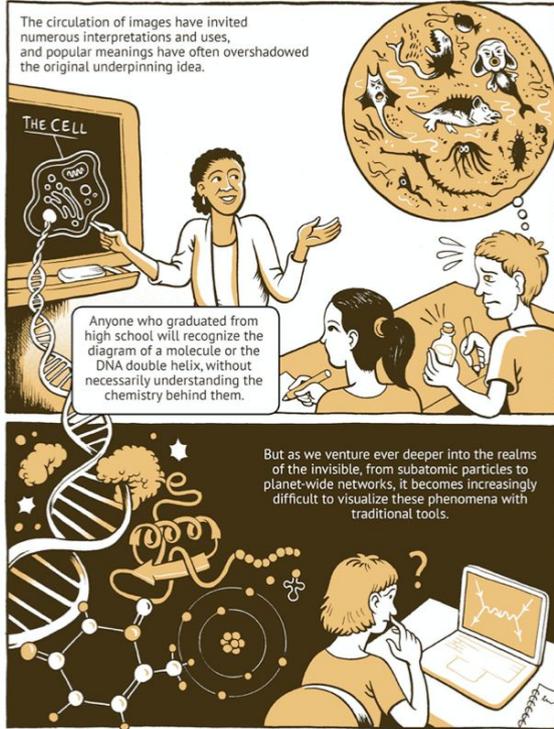


Complete Atrioventricular Canal Defect



Large VSD with Irreversible Pulmonary Hypertension

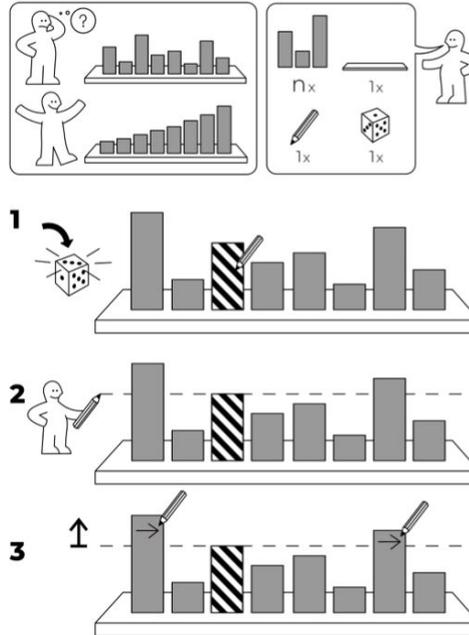
Scientific comics



Farinella, Matteo. "Of Microscopes and Metaphors: Visual Analogy as a Scientific Tool." *The Comics Grid: Journal of Comics Scholarship* 8 (2018).

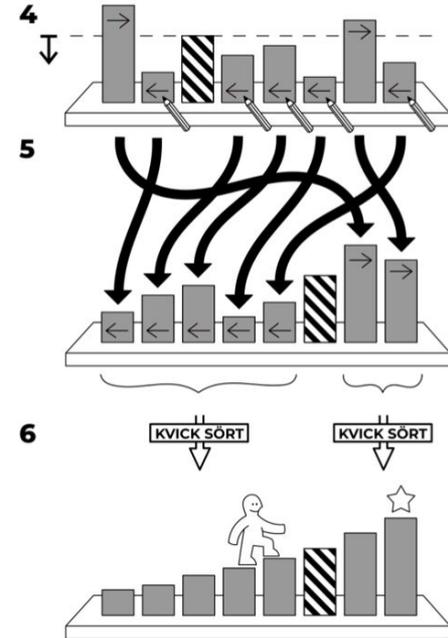
Visual instructions

KVICK SÖRT



idea-instructions.com/quick-sort/
v1.1, CC by-nc-sa 4.0

IDEA

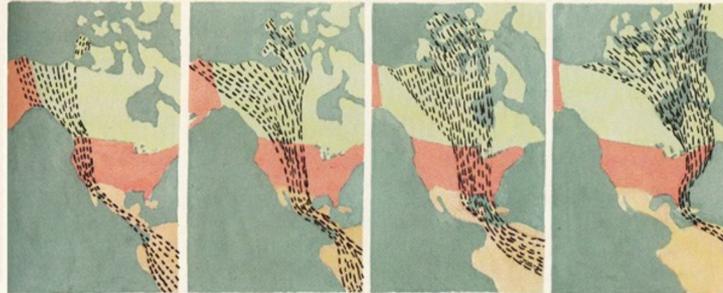


idea-instructions.com/quick-sort/

Isotype, Arnold Gantz, ~1920



Maps can show the way birds go when they fly north or south for the season.



Pacific
flyway

Central
flyway

Mississippi
flyway

Atlantic
flyway

MY LAST WEEK'S SLEEP RECORD

07.May-13.May

MY NAME IS ZEZHONG



I SLEPT 47h46min LAST WEEK



28%

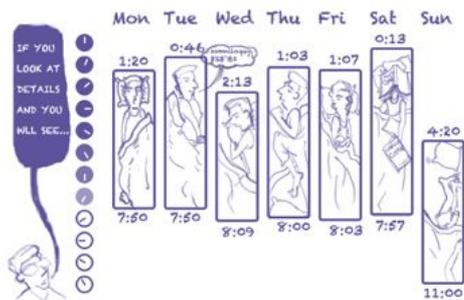


THIS WEEK I SPENT 28% OF MY TIME ON SLEEPING

THAT WAS AN AVERAGE OF 6.9H PER DAY... SOUNDS NOT BAD



BUT...



I WENT TO BED AT 1:35 ON AVERAGE



WHAT ABOUT THIS!



BUT I WENT WIMMING IN THE MORNING EVERYDAY!



WAIT!



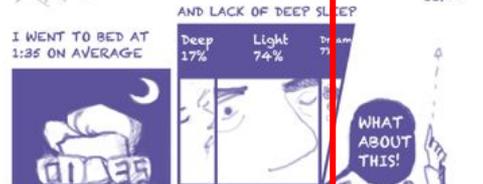
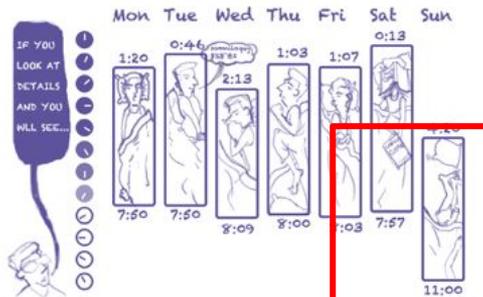
BECAUSE...



THE SWIMMING POOL OPENS AT 11 ON SUNDAY

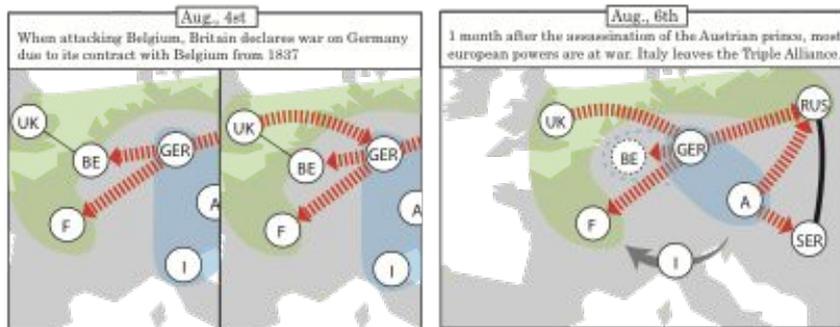
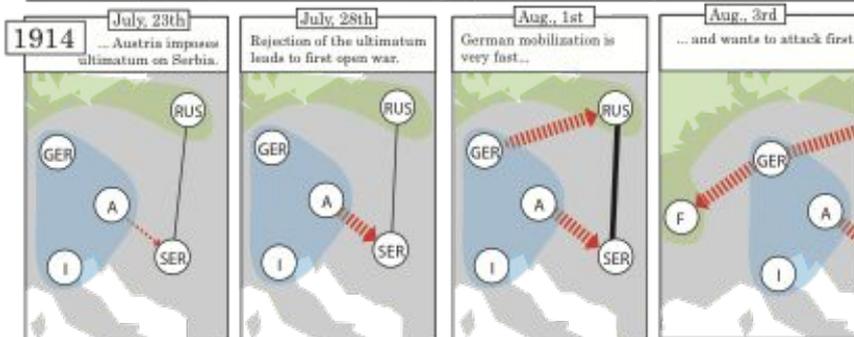
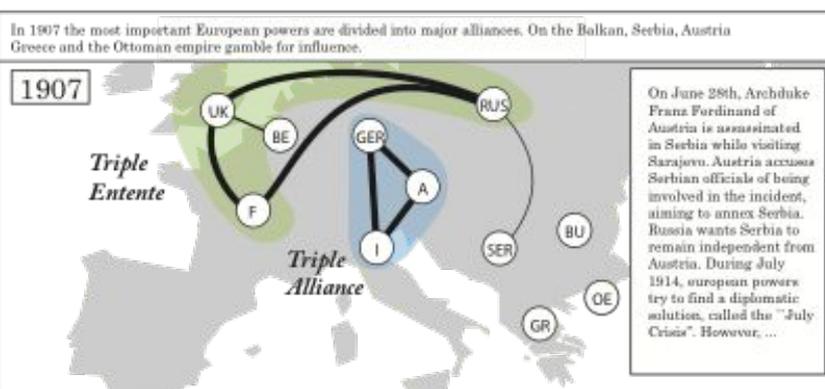
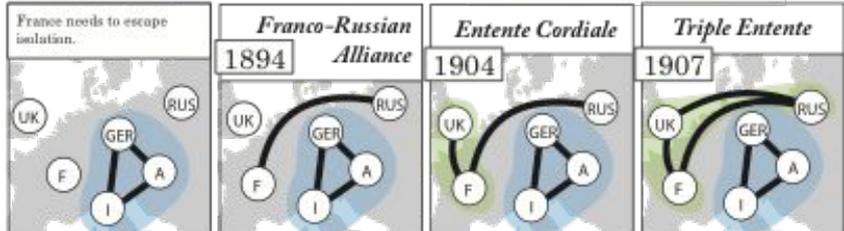
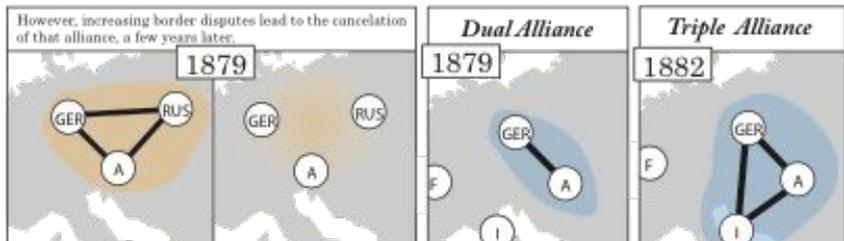
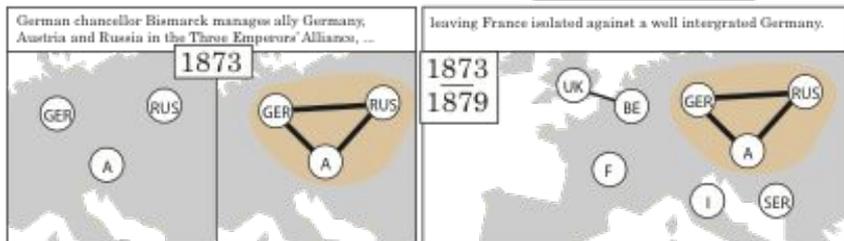
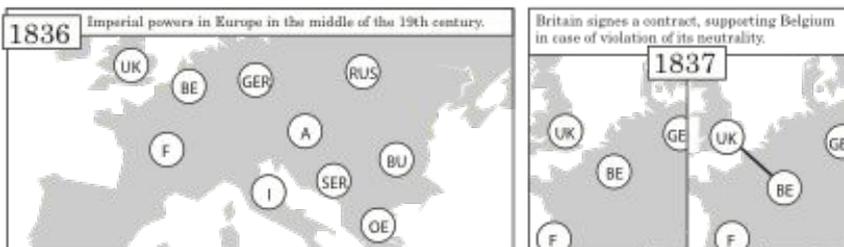
MY LAST WEEK'S SLEEP RECORD

07.May-13.May



(c) Zezhong Wang, in Bach, B., Wang, Z., Farinella, M., Murray-Rust, D. and Henry Riche, N., 2018, April. Design patterns for data comics. In *Proceedings of the 2018 chi conference on human factors in computing systems* (pp. 1-12).

European Alliances before World War I (1836-1914)

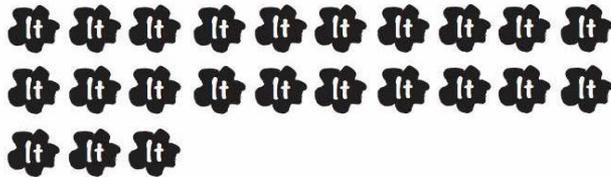


(c) Benjamin Bach, in Bach, B., Kerracher, N., Hall, K. W., Carpendale, S., Kennedy, J., & Henry Riche, N. (2016, May). Telling stories about dynamic networks with graph comics. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 3670-3682). ACM.

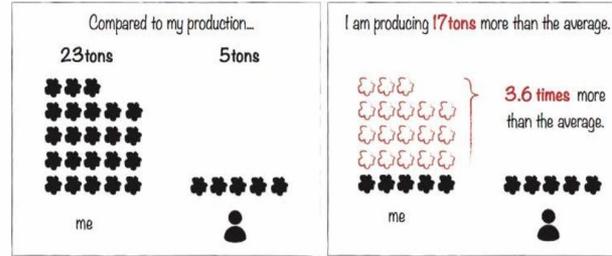
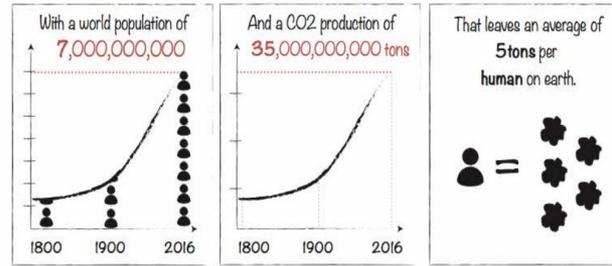
CO₂ Footprint



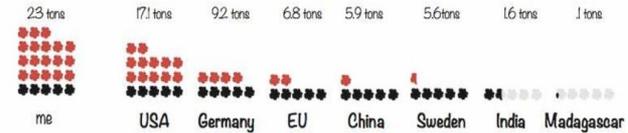
Which produced...



23 tons of CO₂.



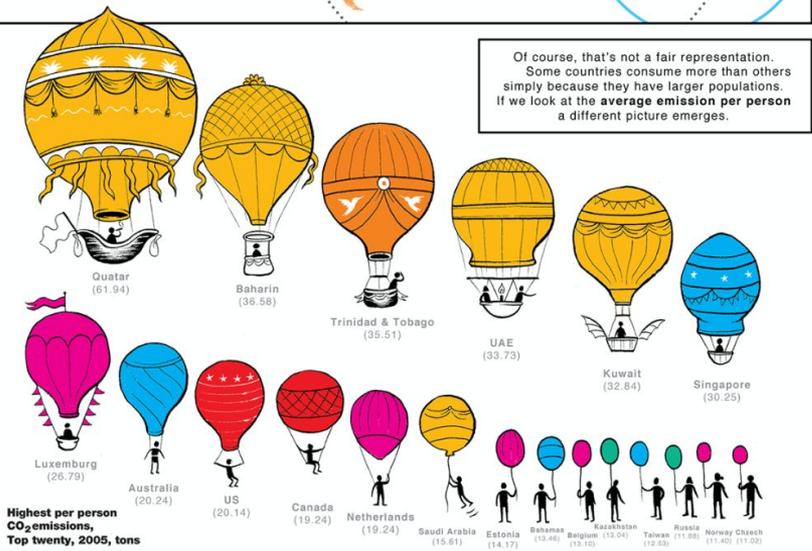
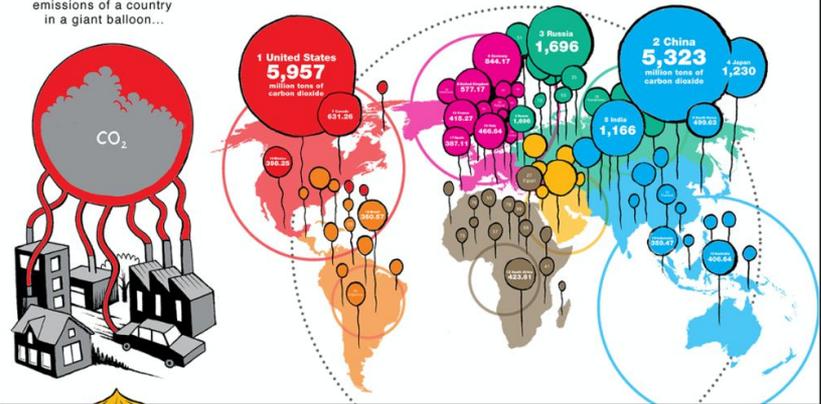
Thus, my travels in 2016 alone produced more CO₂ than the average person in the most countries:



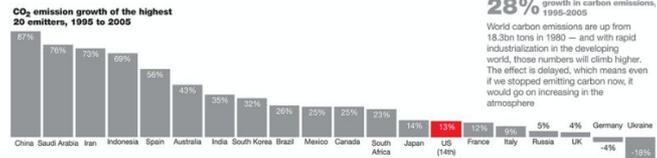
Hot spots – the carbon atlas

Imagine if we could capture all the annual CO₂ emissions of a country in a giant balloon...

...this is what the world would look like:



Moreover, some countries are taking active steps to curb their CO₂ emissions, while others are raising rapidly:



World total 28.19bn tons of CO₂
28% growth in carbon emissions, 1995-2005

World carbon emissions are up from 18.3bn tons in 1990 – and with rapid industrialization in the developing world, those numbers will climb higher. The effect is delayed, which means even if we stopped emitting carbon now, it would go on increasing in the atmosphere



North America 6.99bn tons of CO₂
14% growth in carbon emissions, 1995-2005

The US as a major producer of greenhouse gases has been reluctant to accept that man-made climate change even existed – and refused to accept the Kyoto protocol. But freak weather events and an avalanche of scientific evidence have forced it to rethink its position



Europe 4.67bn tons of CO₂
9% growth in carbon emissions, 1995-2005

For the first time, there is hard scientific evidence of climate change affecting Europe, said the Intergovernmental Panel on Climate Change recently. Freak weather events, such as the heatwaves of 2003, will become ever more common



Eurasia 2.58bn tons of CO₂
4% growth in carbon emissions, 1995-2005

Russia's carbon emissions dropped from 583 million metric tons of carbon in 1992 to 485 million metric tons in 1995, due to its then-deteriorating economic situation. Now, the energy giant may make clean up trading carbon credits



Asia & Oceania 10.36bn tons of CO₂
58% growth in carbon emissions, 1995-2005

Rapid industrialization combined with greater numbers of people moving to cities has provoked a huge rise in carbon emissions – with China rapidly moving to become the world's greatest carbon emitter in the next two years – some scientists say this has happened already



Central & South America 1.10bn tons of CO₂
29% growth in carbon emissions, 1995-2005

Increased freak weather events mean the IPCC is concerned South America will be hard-hit by climate change. Agriculture, water supplies and the unique natural habitat could be affected by a temperature increase of up to 4C by the end of the century



Africa 1.04bn tons of CO₂
28% growth in carbon emissions, 1995-2005

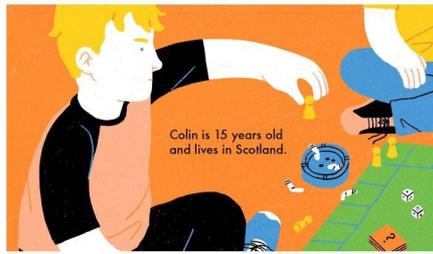
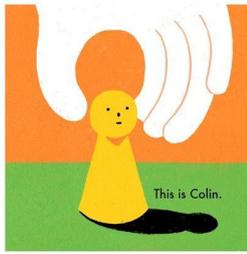
Its carbon emissions may be small but this is the continent most vulnerable to the effects of climate change, hitting food and water supplies, causing coastal flooding and an increase in tropical disease such as malaria – as well as destroying parts of the ecosystem



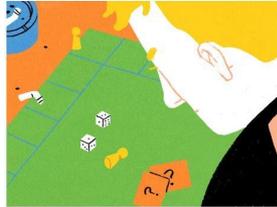
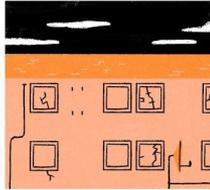
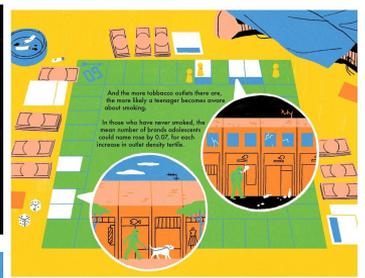
Middle East 1.45bn tons of CO₂
62% growth in carbon emissions, 1995-2005

The region is a major contributor to global greenhouse gas emissions, through an oil and gas industry which produces over 30 percent of world oil supply and over 10 percent of its gas

(c) Matteo Farinella, in Wang, Z., Wang, S., Farinella, M., Murray-Rust, D., Henry Riche, N. and Bach, B., 2019, May. Comparing effectiveness and engagement of data comics and infographics. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

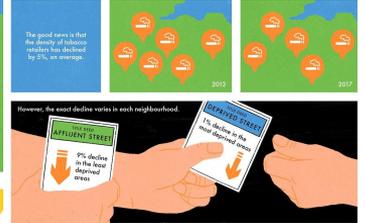


Even though virtually all of Colin's friends know about the serious health impacts of smoking, some of them still believe that it helps them cope with life or to relax.



These assumptions can lead to disastrous results.

people die every day in Scotland from smoking-related illnesses



9% of Colin's peer group smoke. He's been pressured to try before, but didn't really like it.

Young smokers like Colin's friends are very common.

66% of smokers take up smoking ...before the age of 18.



That's cheap property!
I'm getting it!

I am influenced by my surroundings.

And so on it!

Teenagers like Colin who live in areas with the most tobacco retailers are 47% more likely to smoke than those living in areas with the fewest retailers.

And with that comes a whole host of short-term health effects:

- Chance: ADDICTION GO BACK 3 SPACES
- REDUCED LUNG FUNCTION IMPAIRED LUNG GROWTH LOSE A TURN
- EARLY CARDIOVASCULAR DAMAGE
- SEVERE WHEEZING PAY £50 FOR ASTHMA MEDICATION



In 2012, deprived areas had 70% more places to buy tobacco than in more affluent ones.

Area	RENT	Number of Shops
TITLE DEED DEPRIVED STREET	£6	With 1 Tobacco shop
TITLE DEED DEPRIVED STREET	£ 30	With 2 Tobacco shops
TITLE DEED DEPRIVED STREET	£ 90	With 3 Tobacco shops
TITLE DEED DEPRIVED STREET	£ 150	With 4 Tobacco shops
TITLE DEED DEPRIVED STREET	£ 210	With CIGAR SHOP
TITLE DEED AFFLUENT STREET	£ 28	
TITLE DEED AFFLUENT STREET	£ 150	
TITLE DEED AFFLUENT STREET	£ 150	

So, in summary:

TITLE DEED DEPRIVED STREET	TITLE DEED AFFLUENT STREET
• More tobacco retailers	• Fewer tobacco retailers
• Smaller decrease of retailers	• Larger decrease of retailers
• Teenagers living here are more likely to smoke	• Teenagers living here are less likely to smoke

It is therefore no surprise that people living in deprived neighbourhoods are 4 times more likely to smoke than those living in more affluent ones.

When you live in an area with a substantial number of tobacco retailers, you are more likely to smoke, whether or not you became a smoker. The reason for this is that the presence of all the factors and take control of your own destiny. Your fate is not to be decided by a roll of the dice!

GUYS LET'S JUST HAVE A CHAT ABOUT
MASS EXTINCTIONS.

Cruxator Fabryasae

THERE'S GONNA BE A LOT OF GRAPHS AND TEXT AND SHIT SO JUST BE PREPARED TO FEEL LIKE THIS.

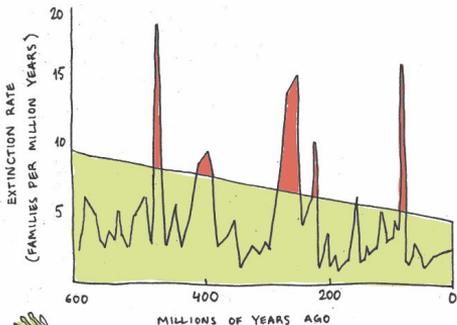


SERIOUSLY IF YOU DON'T NEED THIS IN YOUR LIFE YOU CAN SKIP TO PAGE 3.

LET'S START WITH A HISTORY LESSON.

HERE COMES THE FIRST GRAPH.

Source: University of California Museum of Paleontology's Understanding Evolution (<http://evolution.berkeley.edu>)



IT'S A BIT HARD TO READ BUT I'M GONNA TRY AND BREAK IT DOWN TO YOU. IT'S ABOUT MASS EXTINCTIONS IN THE PAST 600 MILLION YEARS.

THE Y-AXIS SHOWS US THE AMOUNT OF DEATH BASICALLY.

THE X-AXIS IS THE TIMELINE.

WE'RE TALKING LONG TIME PERIODS. 600 000 000 - 0 YEARS AGO.

THE GREEN AREA IS BACKGROUND EXTINCTION. IT'S THE RATE AT WHICH SPECIES DIE OFF NORMALLY, LIKE IT'S NO BIG DEAL.



THE BIG DEALS ARE THE SPIKES. I'VE MADE THEM RED SO YOU CAN FIND THEM MORE EASILY. THOSE ARE THE MASS EXTINCTION EVENTS.

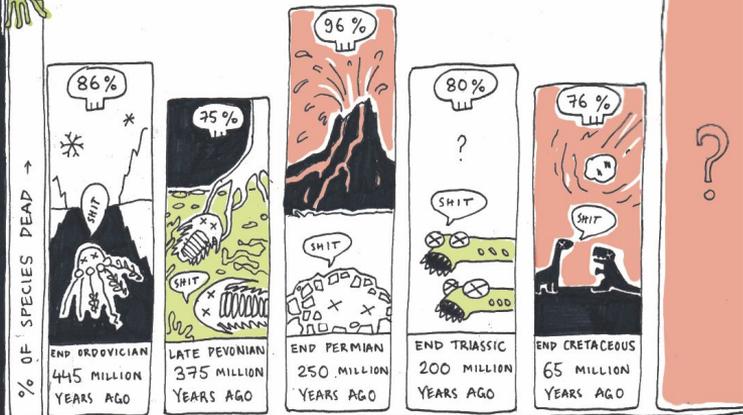
THAT'S WHEN MORE THAN 75% OF ALL SPECIES ON EARTH HAVE DIED OFF.

SO THAT'S HAPPENED 5 TIMES HERE.



LOOK HERE'S MY ATTEMPT TO VISUALISE THE BIG EXTINCTION EVENTS. YES I MADE IT MYSELF. AND I'M A FROG.

THERE ARE NUMBERS HERE THAT AREN'T FROM THE PREVIOUS GRAPH. DON'T FREAK OUT. I USED A DIFFERENT SOURCE: Viviane Richter: "The Big Five Mass Extinctions" <https://cosmosmagz.me.com/palaontology/big-five-extinctions> ESSENTIALLY IT'S THE SAME THING THOUGH.



IT'S A DATA COMIC INSIDE A DATA COMIC! HAHHAHA

I'M VERY EXCITED ABOUT DATA COMICS

ANYWAY

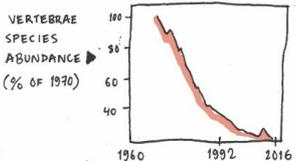
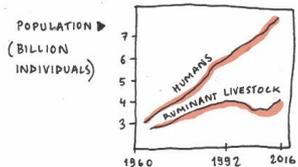
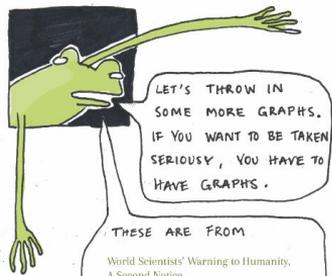
IMAGINE IF THIS PORTION OF LIFE ON EARTH JUST DIED OFF. THAT'S BAD, RIGHT ???

WELL, A LOT OF SCIENTISTS SAY THAT WE'RE NOW ENTERING THE 6TH MASS EXTINCTION.

SEE THAT OMINOUS PANEL THERE? THAT'S WHERE WE ARE.

AND GUESS WHAT'S CAUSING IT THIS TIME? (TURN THE PAGE TO FIND OUT)

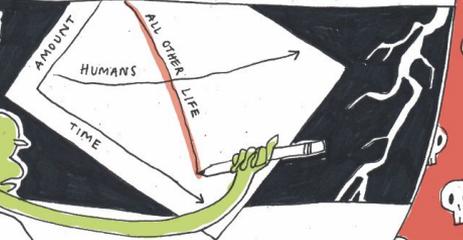
THAT'S RIGHT. THE HUMANS.



THESE ARE FROM World Scientists' Warning to Humanity, A Second Notice, <https://doi.org/10.1093/biosci/bix125>

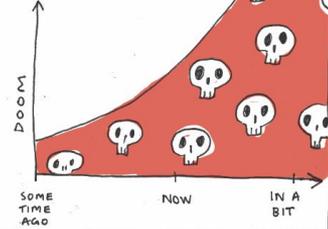
IT WAS PUBLISHED IN 2017. SO IT'S PRETTY RECENT DATA.

YOU SEE WHERE THIS IS GOING?

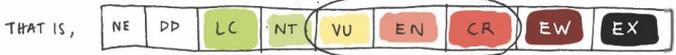


THE CURRENT EXTINCTION RATE IS ESTIMATED TO BE FROM **10 TO 1000 TIMES HIGHER** THAN IT SHOULD BE (THAT REFERS TO THE BACKGROUND EXTINCTION RATE). MAKING ACCURATE ESTIMATES IS TRICKY...

... BUT THIS IS THE GENERAL CONSENSUS.



IUCN (INTERNATIONAL UNION FOR CONSERVATION OF NATURE) HAS DATA ON SPECIES THAT ARE **THREATENED** WITH EXTINCTION RIGHT NOW.

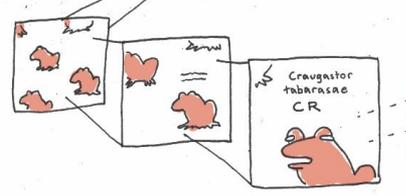


SPECIES CLASSIFIED AS VULNERABLE, ENDANGERED OR CRITICALLY ENDANGERED.



THIS IS REAL SHIT.

WAIT - ZOOM IN ON THE AMPHIBIANS

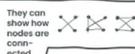


SOURCES: The IUCN Red List of Threatened Species, <https://www.iucnredlist.org/>
 University of California Museum of Paleontology's Understanding Evolution, <http://evolution.berkeley.edu>
 Viviane Bichler: The Big Five Mass Extinctions, <https://cosmosmagazine.com/paleontology/big-five-extinctions>
 William J. Ripple, Christopher Wolf, Thomas M. Newsome, Mauro Galetti, Mohammed Alangir, Eileen Crist, Mahmoud I. Mahmoud, William F. Laurance, 15,364 scientists' signatures from 184 countries: World Scientists' Warning to Humanity: A Second Notice, *BioScience*, Volume 67, Issue 12, 1 December 2017, Pages 1026-1028, <https://doi.org/10.1093/biosci/bix125>

Explaining data analysis <https://statscomics.github.io>

I Context, Motivation & Problem Study

1 This is a **WEIGHTED GRAPH**.

2 They can show how nodes are connected...  ...and how **strong** each connection is.

3 

4 In the real world, they are often not static, but can **change over time**. 

5 So the question arises... 

6 How can you properly visualize a **change** between two graphs? What is a good way to visualize them? 

7 We examined multiple visualizations...  ...and carried out a study to compare some of them. 

II Tasks & Conditions

8 Though there are many existing visualization solutions, most are based on 2 techniques: **node-link diagrams** and **adjacency matrices**.

9  Weight changes indicated by capacity of links

10  Weights are indicated by shade of cells in grid

To test which one performs better, we designed three tasks. Tasks like these are frequently used in domains like brain connectivity analysis:

11

12 **TASKS**

-- TREND -- Assess the weight change of a node's connections

1 Identify all connections to the highlighted node.

2 Assess the change in weight for each connection.

3 Estimate the aggregated change from all these connections.

4 Select an option from 2 to 5.

-- CONNECTIVITY -- Assess the connectivity of common neighbors of two nodes

1 Find the common neighbors, meaning the nodes that are connected to both of the highlighted nodes.

2 Among them, find out how many are present in both graphs.

3 Estimate the change in each region.

4 Estimate the change in the region with the highest edge weight change.

5 Click on the region with the highest edge weight change.

13

14 Identify the region with most changes

1 Create each region and observe the graph change.

2 Estimate the change in each region.

3 Estimate the change in the region with the highest edge weight change.

4 Click on the region with the highest edge weight change.

15 Additionally, we arranged different sizes & densities for the datasets we created.

16 **-- SMALL --** **-- LARGE --** **-- SPARSE --** **-- DENSE --**

III Hypotheses

For each task, we measured performance as number of correct trails. We sought to verify the following hypotheses:

17 **H1** For the trend task, **Matrices** outperform **node-link diagrams** for dense networks.

18 **H2** For the connectivity task, **Matrices** do NOT outperform **node-link diagrams**.

19 For the region task, **Matrices** always outperform **node-link diagrams**.

20 Overall, we expect **node-link diagrams** to decrease in performance for dense datasets.

IV Study Setup, Data Collection & Data Transformation

21 We recruited 11 participants to complete the study... 

22 Where we estimate their performance on understanding the weighted graph changes. 

23 We used a full factorial design for the test problems, so every single possible combination was present.

24 Also, every participant was doing all conditions. 

25 The complexity of test problems gradually increases in the test... 

26 **L1.1** **L1.2** **L1.3** **L1.4** **L1.5**

27 **L2.1** **L2.2** **L2.3** **L2.4** **L2.5**

28 **L3.1** **L3.2** **L3.3** **L3.4** **L3.5**

29 A user's responding time will be recorded and log-transformed... 

30 At the end, they will also be asked to give answer on which technique they prefer. 

V Results

Result shows that **Adjacency matrices** perform better for visualizing weighted graph changes under most conditions... 

31 Which technique let users understand changes more accurately? 

32 Which technique let users understand changes more quickly? 

33 Which technique do users prefer? 

And people like them, too! 

VI Hypotheses Evaluation

34 We expected **node-link diagrams** to decrease in performance for dense datasets. This is verified true, but we didn't expect it will decrease for large datasets as well. 

35 **H1** & **H2** We thought **matrices** would outperform **node-link diagrams** for trend (H1) and region (H2) tasks on dense datasets... Turns out this applies to all tasks across all datasets! 

36 We thought for the connectivity task, **matrices** should NOT outperform **node-link diagrams**... But well, it did. 

H4 Supported

H1, H3 Supported

H2 Unsupported

Wang, Z., Ritchie, J., Zhou, J., Chevalier, F. and Bach, B., 2020. Data Comics for Reporting Controlled User Studies in Human-Computer Interaction. *IEEE Transactions on Visualization and Computer Graphics*, 27(2), pp.967-977.

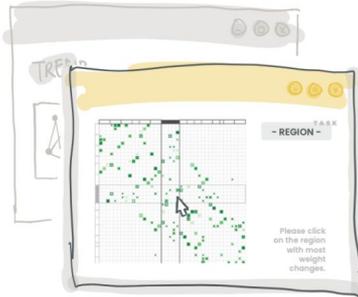
Explaining data analysis

Data Collection & Data Transformation

We recruited **11** participants to complete the study...



Where we estimate their performance on understanding the weighted graph changes.



We used a **full factorial** design for the test problems, so every single possible combination was present.

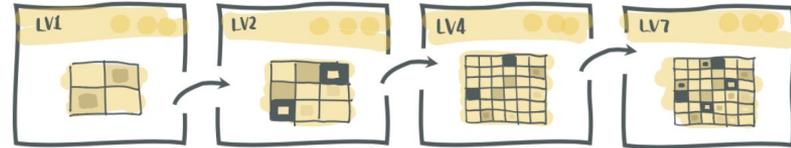
$$2^{\text{TECHNIQUES}} \times 3^{\text{TASKS}} \times 2^{\text{SIZES}} \times 2^{\text{DEFINITIONS}}$$

Also, every participant was doing all conditions.

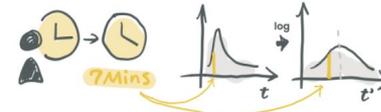
(within-subject study)



The complexity of test problems gradually increases in the test...



A user's responding time will be recorded and log-transformed...



At the end, they will also be asked to give answer on which *technique* they prefer.



Explaining data analysis

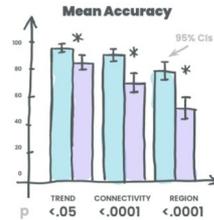
V

Results

Result shows that  Adjacency matrices perform better for visualizing weighted graph changes under most conditions...

Which technique let users understand changes more accurately?

RESULT RM-ANOVA

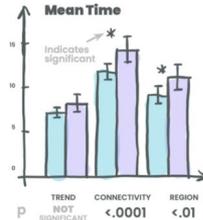


WINNER



Which technique let users understand changes more quickly?

RESULT RM-ANOVA



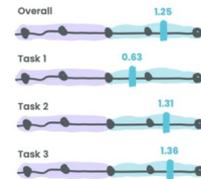
WINNER



And people like them, too! ❤️

Which technique do users prefer?

RESULT



Likert Scale

WINNER

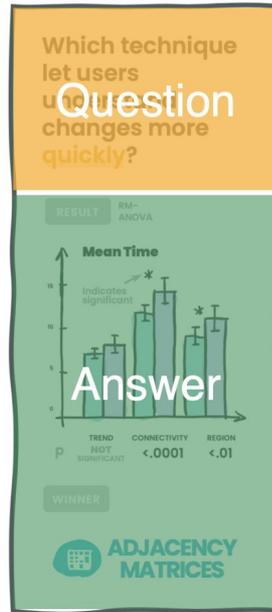
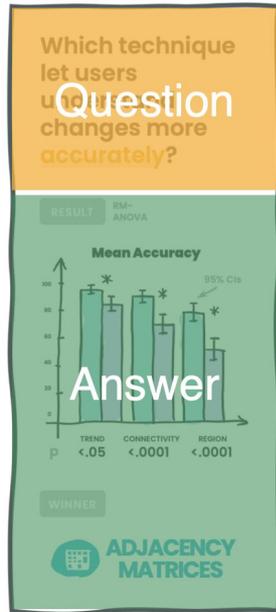


Explaining data analysis

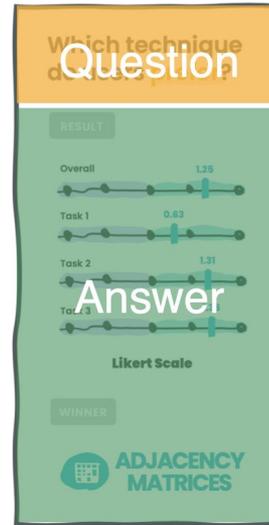
V

Results

Result shows that  Adjacency matrices perform better for visualizing weighted graph changes under most conditions...



And people like them, too! ❤️



Explaining data analysis

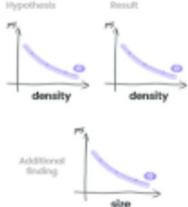
Hypotheses Evaluation

H4

We expected **node-link diagrams** to decrease in performance for **dense datasets**.

This is verified true, but we didn't expect it will decrease for **large datasets** as well.

34



The figure contains three line graphs. The top-left graph is labeled 'Hypothesis' and shows a downward-sloping curve of performance (P) versus density. The top-right graph is labeled 'Result' and shows a similar downward-sloping curve. The bottom graph is labeled 'Additional finding' and shows a downward-sloping curve of performance (P) versus size.

✓

H4 Supported

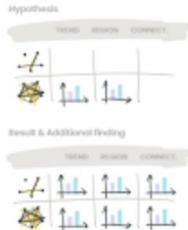
With additional findings

H1 & H3

We thought **matrices** would outperform **node-link diagrams** for **Trend (H1)** and **Region (H3)** tasks on **dense datasets**...

Turns out this applies to **all tasks across all datasets!**

35



The figure shows two rows of diagrams. The top row is labeled 'Hypothesis' and shows a matrix diagram and a node-link diagram with arrows pointing to a bar chart. The bottom row is labeled 'Result & Additional finding' and shows the same matrix and node-link diagrams with arrows pointing to a bar chart.

✓

H1, H3 Supported

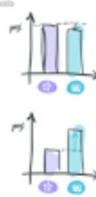
With additional findings

H2

We thought for the **Connectivity** task, **matrices** should NOT outperform **node-link diagrams**...

But well, it did.

36



The figure shows two bar charts. The top chart is labeled 'Hypothesis' and shows two bars, one for matrices (M) and one for node-link diagrams (NL), with the NL bar being taller. The bottom chart is labeled 'Result' and shows the same two bars, but the M bar is taller than the NL bar.

✗

H2 Unsupported

Explaining data analysis <https://statscomics.github.io>

I Context, Motivation & Problem Study

This is a **WEIGHTED GRAPH**.

They can show how nodes are connected... and how **strong** each connection is.

In the real world, they are often not static, but can **change over time**.

So the question arises...

How can you properly visualize a **change** between two graphs? What is a good way to visualize them?

We examined multiple visualizations...

...and carried out a study to compare some of them.

II Tasks & Conditions

Though there are many existing visualization solutions, most are based on 2 techniques:

8 **node-link diagrams** and **adjacency matrices**.

9 Weight changes indicated by capacity of links

10 Weights are indicated by shade of cells in grid

To test which one performs better, we designed three **tasks**. Tasks like these are frequently used in domains like brain connectivity analysis:

11

TASKS

12 **Assess the weight change of a node's connections**

13 **Assess the connectivity of common neighbors of two nodes**

14 **Identify the region with most changes**

15 Additionally, we arranged different sizes & densities for the datasets we created.

16

III Hypotheses

For each task, we measured performance as number of correct trails. We sought to verify the following hypotheses:

17 **H1** For the trend task, **Matrices** outperform **node-link diagrams** for dense networks.

18 **H2** For the connectivity task, **Matrices** do NOT outperform **node-link diagrams**.

19 For the region task, **Matrices** always outperform **node-link diagrams**.

20 Overall, we expect **node-link diagrams** to decrease in performance for dense datasets.

IV Study Setup, Data Collection & Data Transformation

21 We recruited 11 participants to complete the study...

22 Where we estimate their performance on understanding the weighted graph changes.

23 We used a full factorial design for the test problems, so every single possible combination was present.

24 Also, every participant was doing all conditions.

25 The complexity of test problems gradually increases in the test...

26

27

28

29 A user's responding time will be recorded and log-transformed...

30 At the end, they will also be asked to give answer on which technique they prefer.

V Results

Result shows that **Adjacency matrices** perform better for visualizing weighted graph changes under most conditions...

31 Which technique let users understand changes more accurately?

32 Which technique let users understand changes more quickly?

33 And people like them, too! ♥

Which technique do users prefer?

VI Hypotheses Evaluation

34 We expected **node-link diagrams** to decrease in performance for dense datasets. This is verified true, but we didn't expect it will decrease for large datasets as well.

35 **H1 & H2** We thought **matrices** would outperform **node-link diagrams** for trend (H1) and region (H2) tasks on dense datasets... Turns out this applies to all tasks across all datasets!

36 **H2** We thought for the connectivity task, **matrices** should NOT outperform **node-link diagrams**... But well, it did.

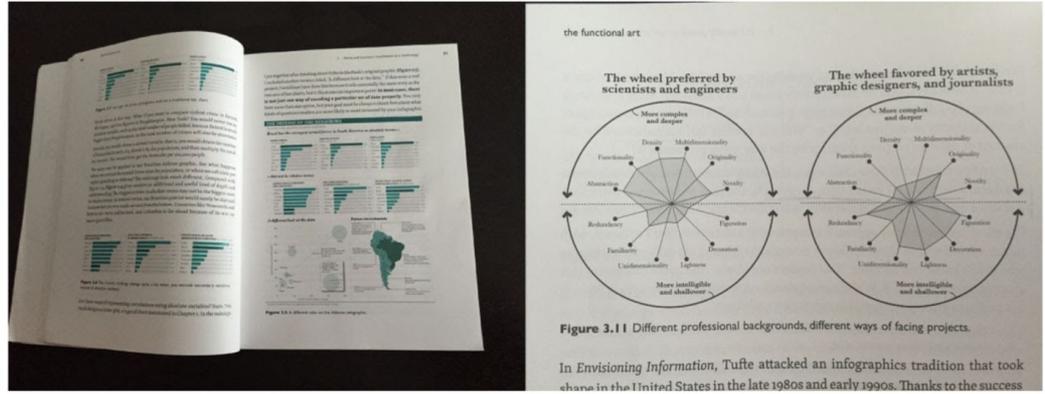
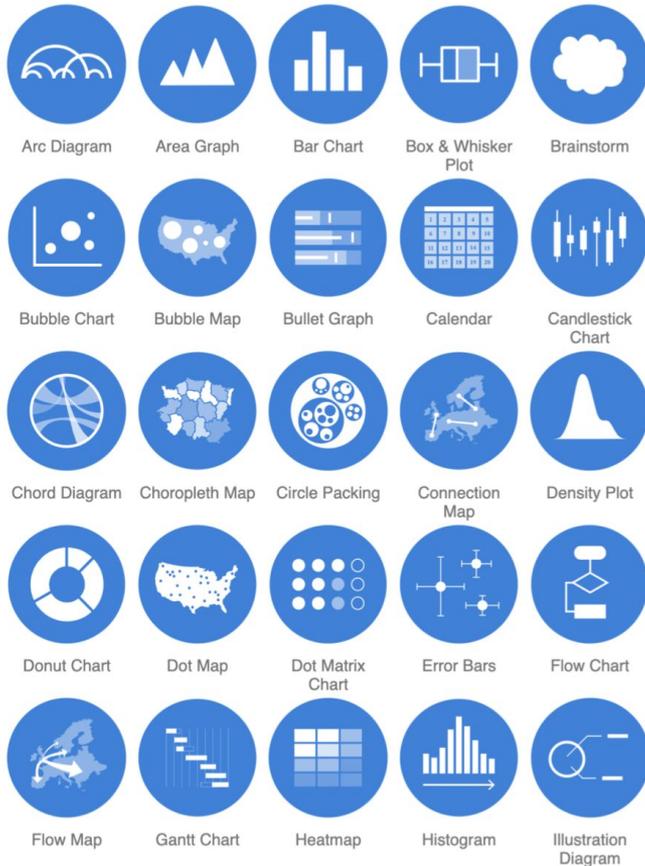
H4 Supported

H1, H3 Supported

H2 Unsupported

Wang, Z., Ritchie, J., Zhou, J., Chevalier, F. and Bach, B., 2020. Data Comics for Reporting Controlled User Studies in Human-Computer Interaction. *IEEE Transactions on Visualization and Computer Graphics*, 27(2), pp.967-977.

Find support....



The Functional Art by Alberto Cairo



Visual Language for Designers by Connie Malamed

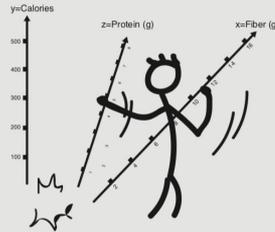
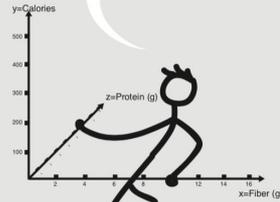
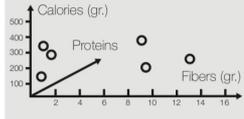
Visualization Cheat Sheets



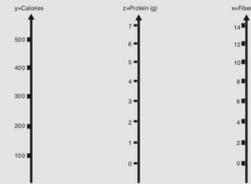
Parallel Coordinates

Construction

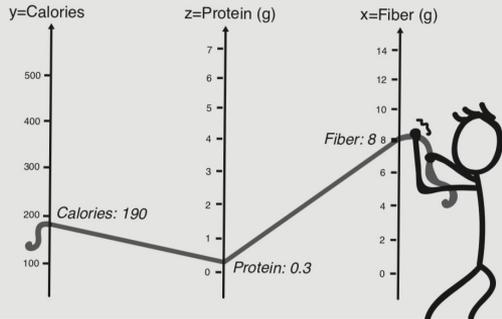
Let's change the arrangement of the axes...



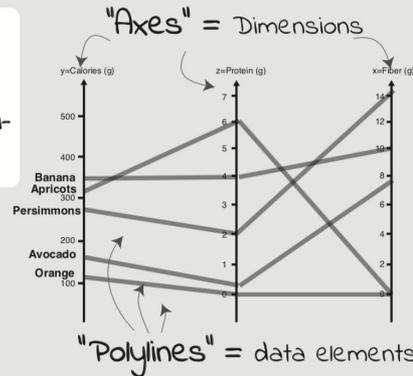
...like so :



we now connect the values for each fruit with a line.



And obtain our Parallel Coordinates Plot (abbreviated: PCP).



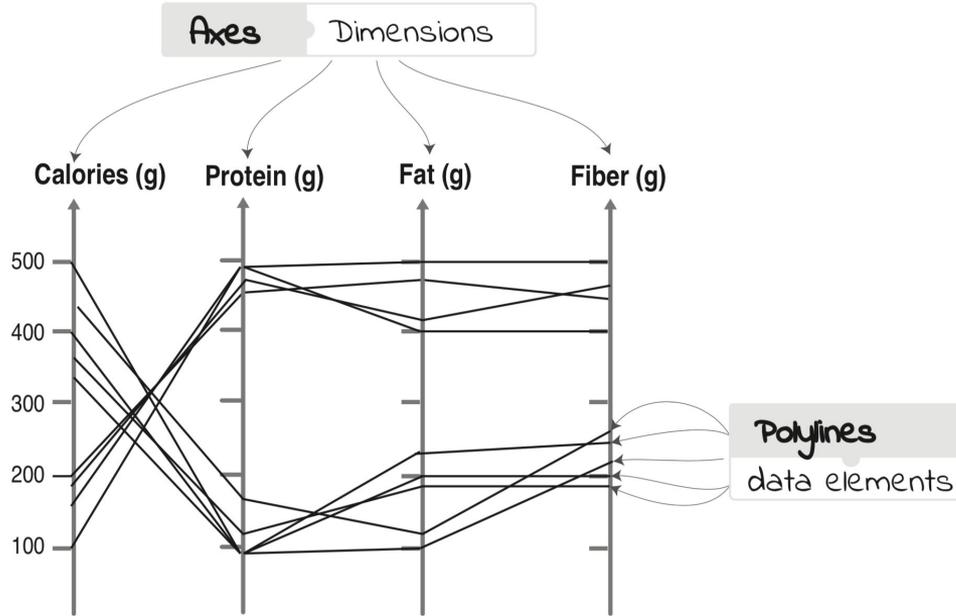
Wang, Z., Sundin, L., Murray-Rust, D. and Bach, B., 2020, April. Cheat Sheets for Data Visualization Techniques. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

Visualization Cheat Sheets



Parallel
Coordinates

Anatomy



<https://visualizationcheatsheets.github.io>

Wang, Z., Sundin, L., Murray-Rust, D. and Bach, B., 2020, April. Cheat Sheets for Data Visualization Techniques. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

Visualization Cheat Sheets



Parallel Coordinates

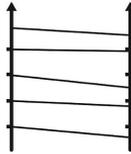
Visual Patterns

Parallel lines

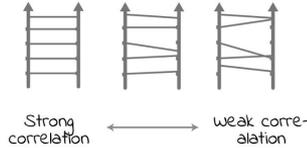
Positive Correlation

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

Correlations are not causalities!



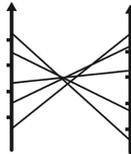
A correlation is visible through rather parallel polylines between two axes.



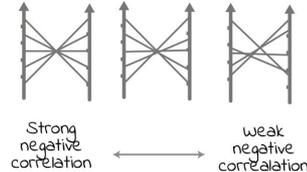
Crossing lines

Negative Correlation

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



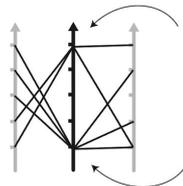
An inverse correlation is visible through lots of crossing polylines between two axes.



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.

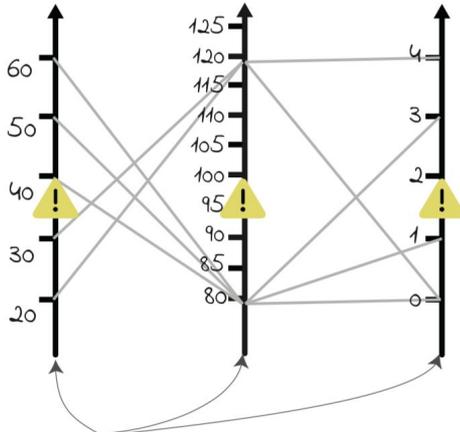
Visualization Cheat Sheets



Parallel Coordinates

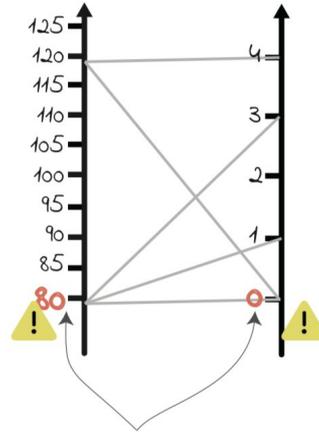
Pitfalls

Axis scales



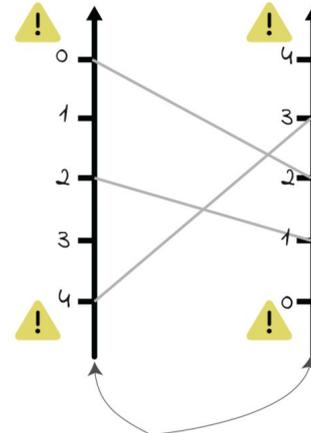
Different dimensions usually have different scales and units.

Truncated axes



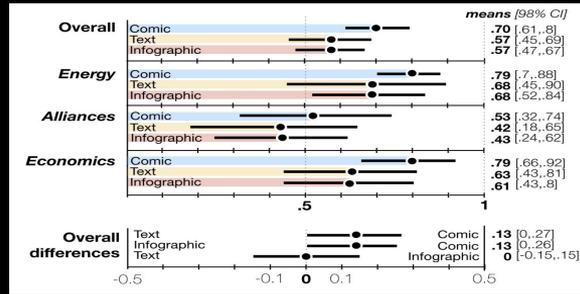
values on axes can start from values other than '0'.

Axes order



values on axes can be either descending or ascending.



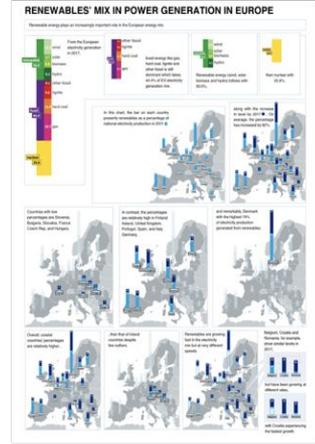
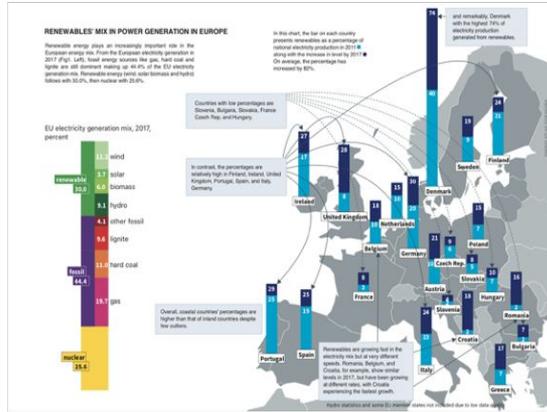
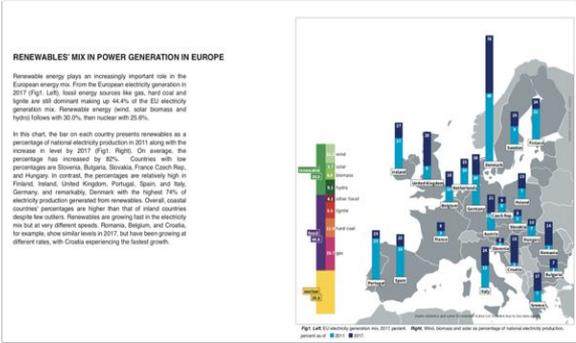


Are data comics effective?

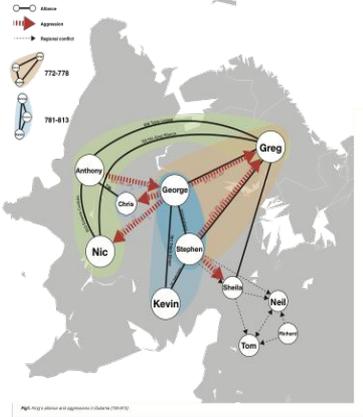
Illustrated Text

Infographics

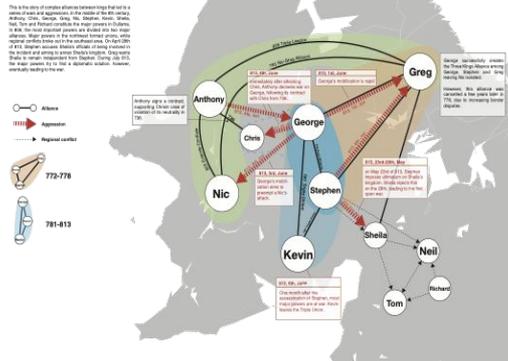
Data comics



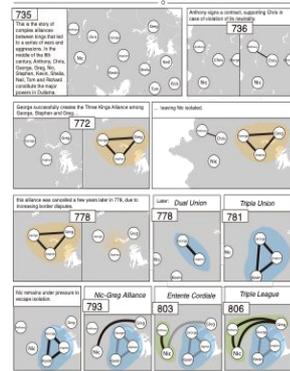
Kings' Alliances and Aggressions in Dullama (735-813)



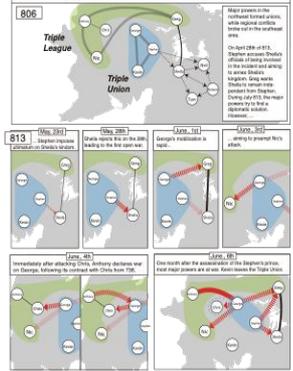
Kings' Alliances and Aggressions in Dullama (735-813)



Kings' Alliances and Aggressions in Dullama (735-813)



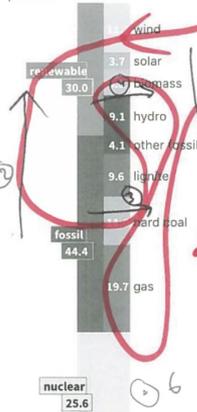
In 806, leaving the powers divided into two major alliances



RENEWABLES' MIX IN POWER GENERATION IN EUROPE

Renewable energy plays an increasingly important role in the European energy mix. From the European electricity generation in 2017 (Fig. 1), fossil energy sources like gas, hard coal and lignite are still dominant making up 44.4% of the EU electricity generation mix. Renewable energy (wind, solar biomass and hydro) follows with 30.0%, then nuclear with 25.6%.

EU electricity generation mix, 2017, percent



Countries with low percentages are Slovenia, Bulgaria, Slovakia, France, Czech Rep. and Hungary.

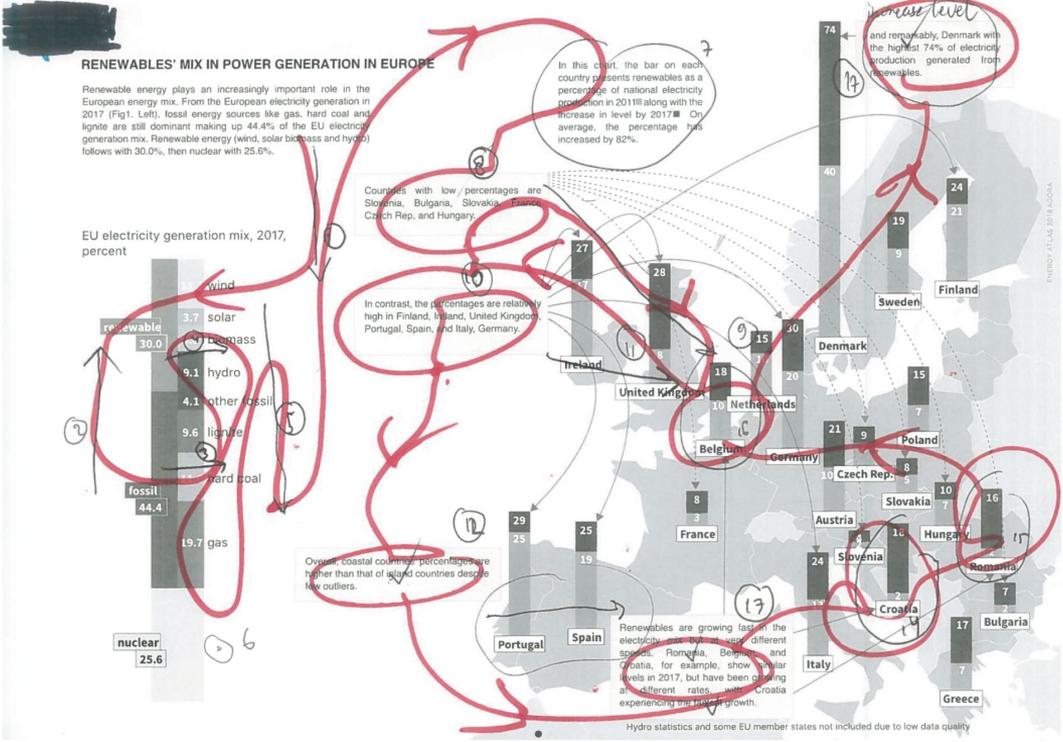
In contrast, the percentages are relatively high in Finland, Ireland, United Kingdom, Portugal, Spain, and Italy, Germany.

Overall, coastal countries' percentages are higher than that of inland countries despite few outliers.

In this chart, the bar on each country presents renewables as a percentage of national electricity production in 2018 along with the increase in level by 2017. On average, the percentage has increased by 82%.

Renewables are growing fast in the electricity mix across different spaces. Romania, Bulgaria and Croatia, for example, show similar levels in 2017, but have been climbing at different rates, with Croatia experiencing the most growth.

Hydro statistics and some EU member states not included due to low data quality



EUROSTAT/IAS BY BAUWA

Effectiveness of data comics

Pros:

Visual overview

Guide readers

Adapt to reader pace

Space == narration

Include illustrations besides vis

Memorable layout == narration

Can be shared easily

Effectiveness of data comics

Cons:

Careful with detailed images -> size

Avoid repetition -> highlight changes

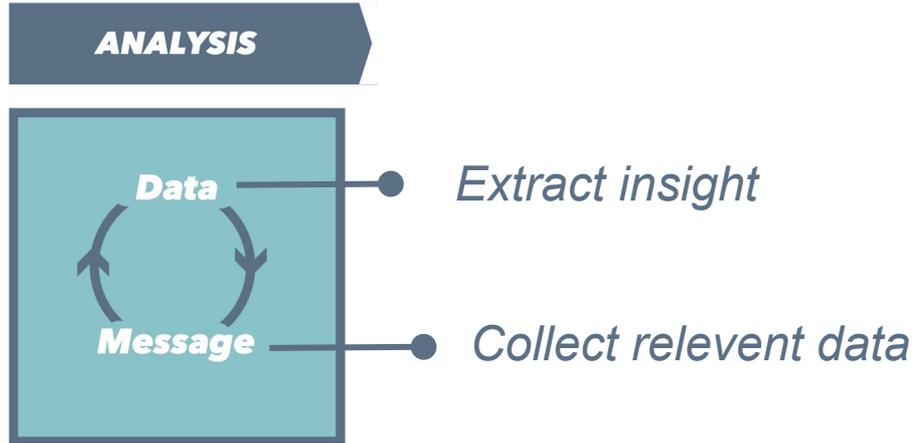
Require space

Non-interactive

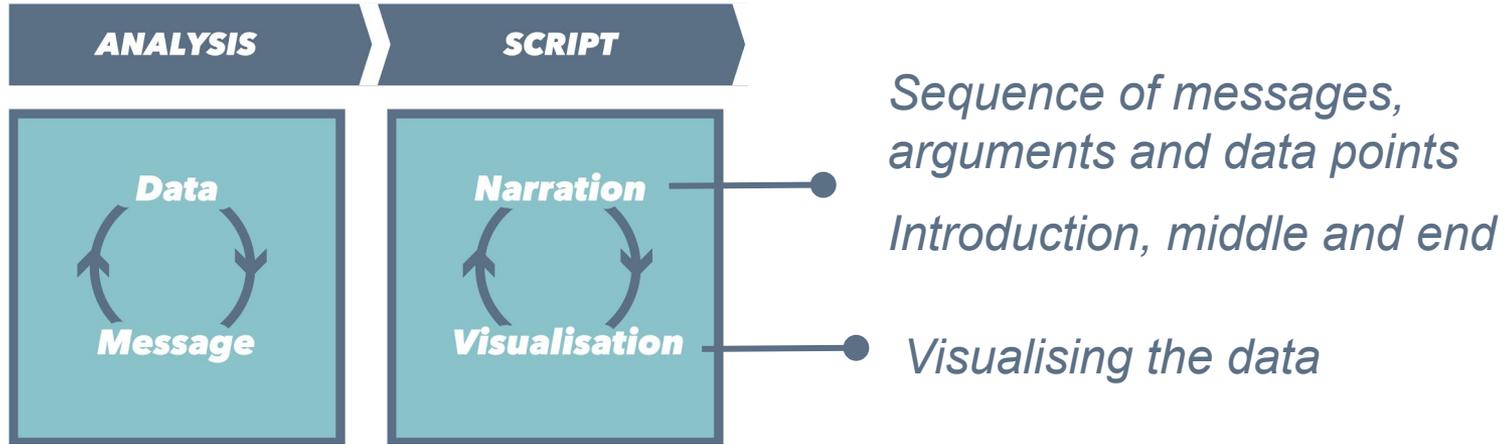


Creating data comics

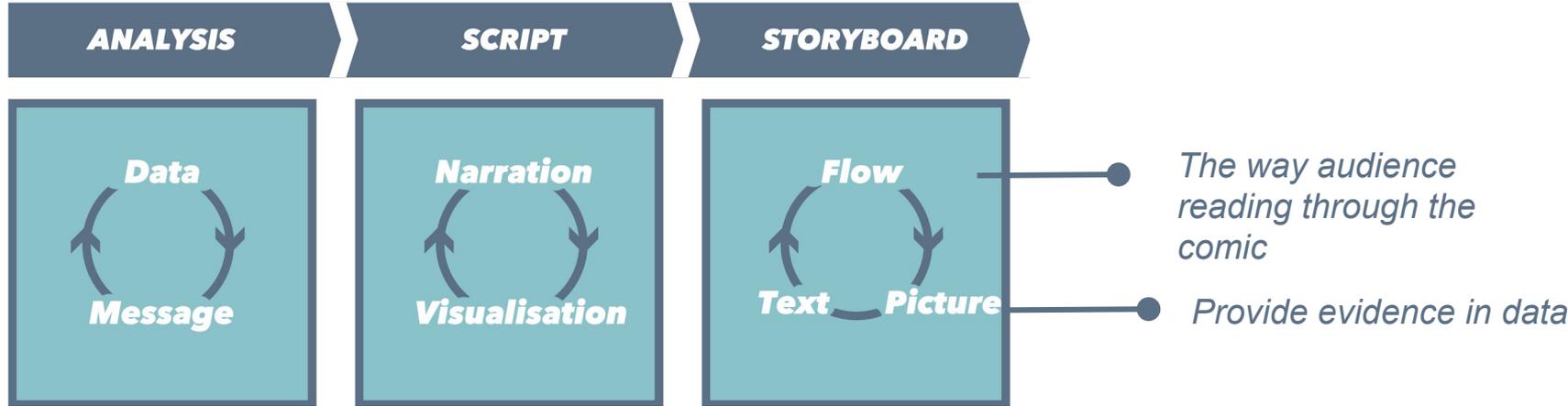
Creation process



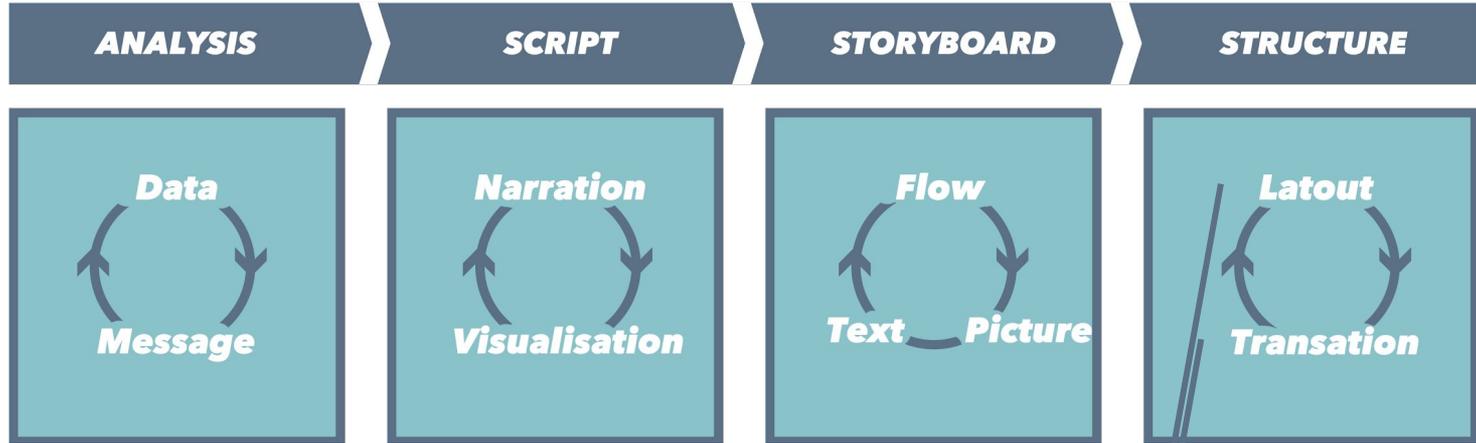
Creation process



Creation process



Creation process



● placement of panels / contents in each panel

● Connections between information

graphic design

storytelling

Drawing & illustration

storytelling

Skills

visualization
design

data analysis

creating
visualizations

Slovene Co-authorship (Version 3)

In 1960, 3 authors form unconnected clusters.

The University of Ljubljana starts a biology program bringing together groups of physicists and biologists.

This collaborator cluster expands as biologists and physicists work more closely.

Groups of chemists and statisticians start to team up with this group.

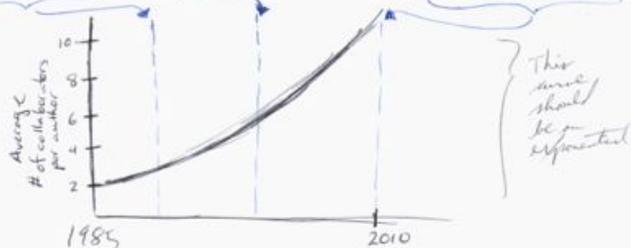
These groups are incorporated into the growing cluster.

While the cluster grows, the overall network remains unconnected as new universities and departments are formed.

25% of network

Yugoslavia collapses. People are free to travel and associate.

2010



SLOVENE COLL. II

6

As this happens, the network grows by attracting new members joining the scientific community...

... and the diameter drops constant

This trend continues, until, on the collapse of Y in 1990, after which

that year, the network loses many members

and its diameter grows again

The largest cluster falls into smaller clusters...

which remain only loosely connected

Data Comics Design Patterns

Narrative



Exposé



Multiple-explanations

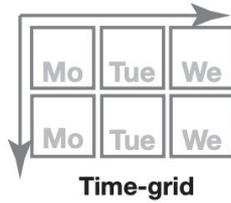


Question & answer

Temporal



Temporal sequence

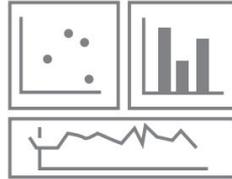


Time-grid



Time-nesting

Faceting

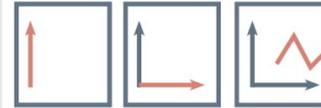


Multiple facets



Contrast

Visual Encoding



Build-up



Legend

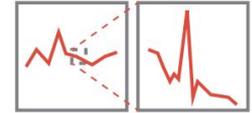


Annotated transition

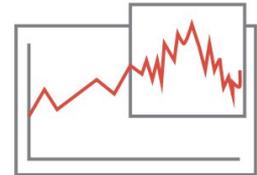
Granular



Zoom



Cut-out



Lens

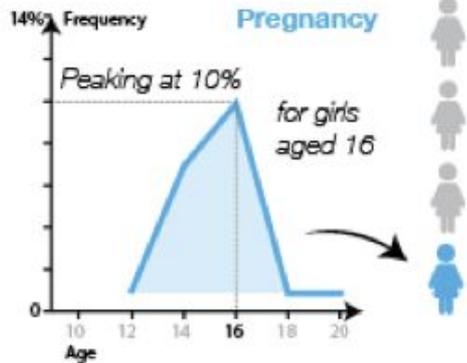
Zoom



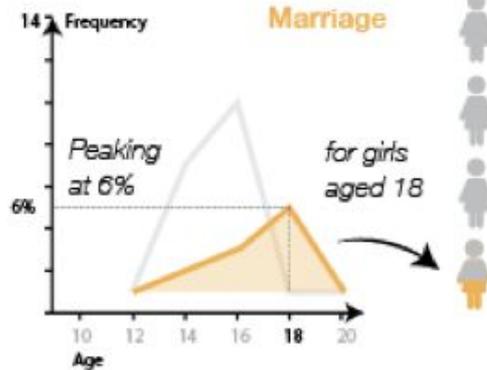
down from 127 in 2010

Multiple Explanation

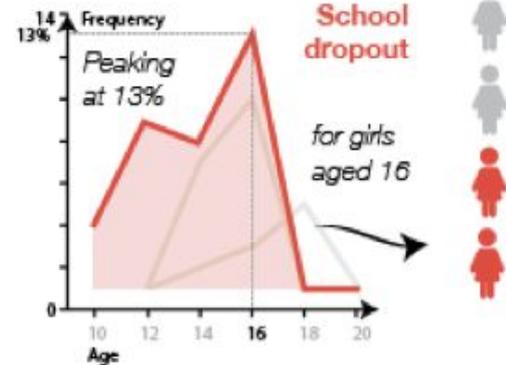
She has a one-in-four risk of becoming pregnant during adolescence,



is at high risk of being engaged in early marriage,



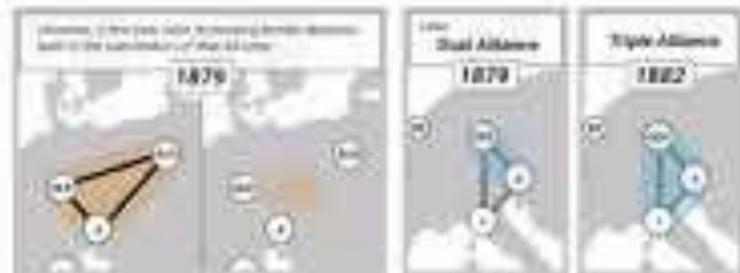
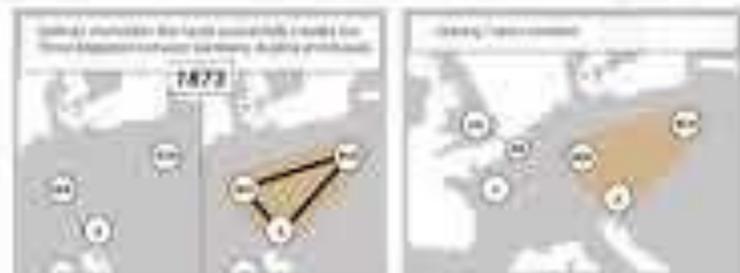
and will likely drop out of school before reaching secondary level.



Temporal Change



European Alliances before World War I



Replace



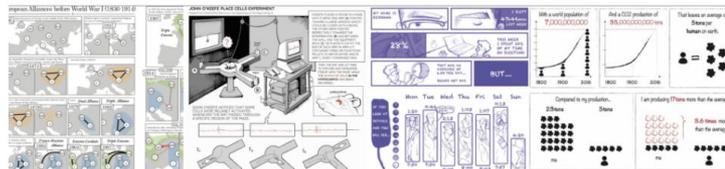
This operation will replace a point with new points after doing "click" or "mouseover" on the "element"

Documentation

[https://www.khanacademy.org/a/...](#)



Interactive Data Comics



- Home
- Publication
- Examples
- Tutorial
- Get Started
- Documentation
- Tips
- Online editor
- About



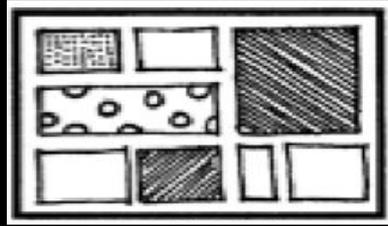
Data comics are a way of effectively communicating with data through data visualizations. They are inspired by the visual language of comics. This project adds interactivity to data comics to **support exploration, explanation, and engagement**.

Interactions are specified using a [JSON](#) specification and which can be rendered using our [online editor](#).

Interactions include:

- Highlight

<https://interactivedatacomics.github.io/>



Wrapping up

Data Comics

- Combining time and space oriented storytelling
- Familiar to many people
- Easily accessible through many media
- Widely applicable
- Effective for breaking down complexity
- Huge design space: expression, style, ...
- Design patterns to help creation
- Data-driven storytelling *is* complicated

Data Comics

- Combining time and space oriented storytelling
- Familiar to many people
- Easily accessible through many media
- Widely applicable
- Effective for breaking down complexity
- Huge design space: expression, style, ...
- Design patterns to help creation
- Data-driven storytelling **is** complicated

Data Comics

- Combining time and space oriented storytelling
- Familiar to many people
- Easily accessible through many media
- Widely applicable
- Effective for breaking down complexity
- Huge design space: expression, style, ...
- Design patterns to help creation
- Data-driven storytelling **is** complicated

Data Comics

- Combining time and space oriented storytelling
- Familiar to many people
- Easily accessible through many media
- Widely applicable
- Effective for breaking down complexity
- Huge design space: expression, style, ...
- Design patterns to help creation
- Data-driven storytelling **is** complicated

Data Comics

- Combining time and space oriented storytelling
- Familiar to many people
- Easily accessible through many media
- Widely applicable
- Effective for breaking down complexity
- Huge design space: expression, style, ...
- Design patterns to help creation
- Data-driven storytelling **is** complicated

Data Comics

- Combining time and space oriented storytelling
- Familiar to many people
- Easily accessible through many media
- Widely applicable
- Effective for breaking down complexity
- Huge design space: expression, style, ...
- Design patterns to help creation
- Data-driven storytelling **is** complicated

Data Comics

- Combining time and space oriented storytelling
- Familiar to many people
- Easily accessible through many media
- Widely applicable
- Effective for breaking down complexity
- Huge design space: expression, style, ...
- Design patterns to help creation
- Data-driven storytelling *is* complicated

Data Comics

- Combining time and space oriented storytelling
- Familiar to many people
- Easily accessible through many media
- Widely applicable
- Effective for breaking down complexity
- Huge design space: expression, style, ...
- Design patterns to help creation
- Data-driven storytelling *is* complicated



The Emerging Genre of Data Comics

Benjamin Bach, Harvard University
Nahale Henry Riche, Microsoft Research
Stephanie Gerstense, University of Calgary
Eva-Maria Pieter, Harvard University

Four Essential Components of Data Comics

Napkin Comic

Zezhong Wang

The majority is accumulating in landfills and the environment at large, where it will outlive the next several generations of human civilization.

80-100 years to degrade

400-1000 years to degrade

United States of Plastic

The Guardian

The rich drive up the mean far above what the typical income is. Therefore, statisticians instead let the median represent the average.

While the mean is the mid-point that splits the total amount of income into equal halves...

Mean

Median

... the median splits the population into equal halves, even though above-median people have a much higher income in total.

Child Poverty

Lovisa Sundin

Narration

Words and Pictures

...smokers like Colin's friends are very common.

66% of smokers take up smoking before the age of 18.

Smoking

Terri Po

...ORD NUMBERS OF WOMEN ARE WINNING FOR OFFICE... AND NO THEY'RE WINNING PRIMARIES.

CURRENT & POSSIBLE % WOMEN IN STATE LEGISLATURES

State	Current %	Possible %
Alabama	38	50
Arizona	34	50
Arkansas	38	50
California	33	50
Colorado	33	50
Connecticut	25	50
Delaware	36	50

The Political Future is Female

Olivia Walch - the NIB

China's rise is so significant because of its sheer enormity. If you were to talk to a random person on Earth, you would be statistically likely to speak to a native Chinese speaker.

Estimated number of native language speakers in 2015 (millions)

Language	Estimated number of native language speakers in 2015 (millions)
Chinese	1,300
Spanish	450
English	350
Hindi	300
Arabic	250
Portuguese	200
Bengali	150
Russian	100
Japanese	100
Other	100

The Future Sounds Chinese

Josh Kramer - The NIB

Further Reading

McCloud, Scott, and A. D. Manning. "Understanding comics: The invisible art." *IEEE Transactions on Professional Communications* 41.1 (1998): 66-69.

Farinella, M., 2018. The potential of comics in science communication. *Journal of science communication*, 17(01), pp.Y01-01.

Zhao, Z., Marr, R. and Elmqvist, N., 2015. Data comics: Sequential

Bach, B., Riche, N.H., Carpendale, S. and Pfister, H., 2017. The emerging genre of data comics. *IEEE computer graphics and applications*, 37(3), pp.6-13.

Bach, B., Wang, Z., Farinella, M., Murray-Rust, D. and Henry Riche, N., 2018, April. Design patterns for data comics. In *Proceedings of the 2018 chi conference on human factors in computing systems* (pp. 1-12).

Wang, Z., Wang, S., Farinella, M., Murray-Rust, D., Henry Riche, N. and Bach, B., 2019, May. Comparing Effectiveness and Engagement of Data Comics and Infographics. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

Wang, Z., Sundin, L., Murray-Rust, D. and Bach, B., 2020, April. Cheat Sheets for Data Visualization Techniques. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

Wang, Z., Ritchie, J., Zhou, J., Chevalier, F. and Bach, B., 2020. Data Comics for Reporting Controlled User Studies in Human-Computer Interaction.



Zezhong
Wang



Jacob
Ritchie



Jingtao
Zhou



Fanny
Chevalier

<http://visualinteractivedata.github.io>

Data Comics for Data-Driven Storytelling

Benjamin Bach

University of Edinburgh

