Empowerment through data visualisations

How a new medium can benefit the world at large

Luca Liechti
Human-IST Institute
University of Fribourg, Switzerland
jan.liechti@students.unibe.ch

ABSTRACT

Information visualisations are becoming a new mass medium. We are interested in what aspects of a visualisation make it fit to empower individuals. We argue that there are two main aspects to this: Being consciously non-manipulative, and crafting visualisations so that they leave a lasting impression. We suggest giving user full interactivity, using open data, and combining different datasets. We illustrate these points with *Spies in the skies*, a data visualisation about the flight routes of FBI and DHS[1] aircraft.

KEYWORDS

Data Visualisation, Information Visualisation, Empowerment, Open Data, Open Source

1 INTRODUCTION

Information visualisation has been ascribed the potential to become a new mass medium[2]. As with all mass media, the question is then: How does it influence the public discourse, who uses it to push which agenda, and can it be used to benefit the world at large?

More concretely, we are interested in how an information visualisation can be considered *empowering*. According to Merriam-Webster, someone is empowered if they have "the power, right, or authority to do something"[3]. We use the term in this very broad definition.

There has been some research focusing on how visualisations can be persuasive [7] or even just memorable [1]. But critically, when we say 'empowerment', we do not mean 'persuasion of what the author thinks is empowering', but *enabling viewers to form an own and relevant opinion*. Underlying our arguments is the assumption that even though modern societies are incredibly complex and there is simply an overwhelming amount of data around, individuals can still empower themselves. Finding the right information plays a crucial role here, just as much as knowing there is important information to be found. We conceive empowerment always as self-empowerment: an individual cannot

passively become empowered. Thus, having the option to decide at every point of a process, in this case of using a visualisation, is a key to empowerment.

This paper is structured as follows: First, we will quickly discuss related work. Then, we will propose and discuss some properties that make a visualisation empowering. After that, we present a case study of a visualisation that embraces the properties we have defined. We conclude with open research questions and a short summary.

2 RELATED WORK

There is not a whole lot of academic literature on the more critical aspects of information visualisation. In this section, we summarise the results of two papers on the subject [4, 5]. Dörk et al. [5] list four aspects that make a given visualisation *critical* in the sense that it can both explain and transform society:

Disclosure Things to be disclosed can be selection or partitioning of data, origin of data, or various design decisions determining representation and interaction.

Plurality More than one facet of a phenomenon should be visualised.

Contingency Closely related to plurality, this requires a multitude of conclusions to be drawn from a visualisation. Our re-formulation of this principle is that the dataset should just be large enough, as reality is very rarely a one-sided affair.

Empowerment Visualisations generally empower their creators to let their voices be heard, but this should be extended to the readers, enabling them to critically question the visualisation and use it to tell their own story. There should be a shift from awareness to action, and a connection between the visualisation and civic engagement.

D'Ignazio and Klein [4], building on Dörk et al.'s work, extend these points. Coming from a feminist point of view, they encourage us to *rethink binaries* (visualise non-binary categories as such), to *consider context* (give extra awareness to the different ways the creation of a visualisation can influence it at any stage), and to *make labor visible* (by explicitly

---

[1]Department of Homeland Security

[2]http://edition.cnn.com/2011/OPINION/04/19/sexy.data/

[3]https://www.merriam-webster.com/dictionary/empowerment

writing about e.g. data provenane), among others.

These aspects in their entirety are concerned with different layers of data visualisations. Some deal with rather abstract mindsets that authors should have, others with specific qualities of the visualisation and its framing, and even others with states of mind of recipients.

We are most interested in the visualisation itself. Reasoning on this level is less speculative, and can be better illustrated with real-world examples. In the following section, we briefly discuss the points mentioned here, and then add some of our own.

3 ADDITIONAL PARAMETERS OF EMPOWERMENT

The points of *plurality, contingency,* and *non-binaries* are characterizing what is going on *in* the data and their representation. Being non-binary imposes some rather austere restrictions on the graphical tools at our disposal; the commandments of pluralism and contingency forbid us to base our work on a small or overly skewed dataset. On the other hand, the points of *disclosure* and *labor visibility* are concerned with where our data come from, who has gathered them, and what we as authors communicate about them. This communication is external to the visualisation itself and *frames* the visualisation rather than being a part of it.

Our suggestions fall in two categories: First, how can we ensure to present viewers an unbiased, non-manipulative visualisation? The first two points deal with this aspect. Second, how can we *really* catch the viewer's attention? This is discussed in the third point.

Keep it reader-driven

Segel and Heer, in their seminal paper on narrative data visualisation [8], establish a fundamental axis on which each data visualisation can be placed. That spectrum is between author-driven and reader-driven visualisation. A completely author-driven visualisation would be a video with no option for the viewer to interact with. A completely reader-driven visualisation would be one where nothing happens without user input, and where the user has at all times the full interaction options at her disposal. One often employed pattern is the Martini glass structure, where the first few exploratory steps in a visualisation are guided, and then the options for the user are growing more and more.[4]

In the context of the empowering aspect of data visualisations, it is hard to find an argument for anything different than full exploratory freedom for the reader. This is related to Dörk et al.'s plurality principle. However, a plurality of aspects of a given topic is also possible in the exploratory part of a martini glass structured visualisation. Leaving the guided part out, however, does not shape the viewer's image of the visualisation as much, and makes a more open and unbiased experience possible.

If the creator of a visualisation absolutely wants to focus the viewer's attention on certain details or outcomes, it is from the perspective of open-ended user experience much better to put that in the framing of the visualisation, e.g. the article accompanying it. Like this, the reader can decide for herself if she would like to double-check these facts, or focus directly on the part that interests her the most.

The case for open data

In an essay by Janssen et al. [6], the authors suggest that open data have an enormous potential for society. However, according to the authors, one of the big myths surrounding the topic is that publicizing data will automatically yield benefits.

Interestingly enough, neither Dörk et al. nor D'Ignazio and Klein are mentioning open data at any point in their papers. There are however many connections between open data and empowerment. We shed a light on some aspects here.

We find it favourably to use open datasets as the basis of the visualisation. There are basically two other options: Either buy a dataset, or create it yourself.

Buying can clearly be dismissed, because as a visualisation author, you are depending on the owner of the dataset in many aspects: The creation and curation of the dataset may not be known, leaving the door open to numerous technical problems. Also, the owner may have *willfully* tampered with it. On top of that, one is most likely forbidden to give the public access to the dataset, making it impossible to double-check any visualisation.

But why is using an open dataset better than creating your own? Clearly, if a certain dataset simply does not exist, there is no alternative to generating it oneself. However, with the abundance of government- and community-created datasets, there are a few points speaking to their advantage. First, doing something useful and empowering with these data is what they were gathered for in the first place. As Janssen et al. would say: Publicizing data is not automatically beneficial, but if users interact with it, it is increasingly hard to argue against that. In that sense, doing good work with it is "giving back" to the community and increasing the incentive to publish even more data openly. Second, using data from the public domain may demonstrate to readers the possibilities there are to gather, contribute to, retrieve, and interact with open data. Lastly, and most importantly, we make the point that if an open dataset is used, then the

---

[4]Interestingly, this approach—specifically studied as a means to find out how to better empower data visualisation consumers—has been shown *not* to engage the viewer more than an open, fully reader-driven, exploratory visualisation [2].

data have not been collected with a specific visualisation in mind. It is much harder to push an agenda that way, and it is impossible to subconsciously let a potential outcome interfere with the gathering of data.

In conclusion: open data is better than proprietary source. But what if there is no data at all? In case one has to gather the data oneself, we argue for releasing it to the public domain. This on one hand strengthens the open data community, on the other it allows users to critically dissect your visualisation, looking both for honest mistakes as well as malicious tampering.

Combine datasets

A canonical definition of data visualisations describes one of their aims as "amplify[ing] cognition" [3]. Dörk et al. argue that visualisations are "not so much to gain 'analytic insight', but rather to get a *heightened recognition* of an issue, awareness about an online community's shared resources, or even reflection about oneself."[5] [5]. In the same way, D'Ignazio and Klein argue that visualisations do not only affect us cognitively, but also on a visceral and emotional level[4]. This is consistent with results by Pandey et al. [7], who studied how visualisations can be persuasive. The most often cited reason for changing an attitude by participants was that they were *struck by evidence,* a formulation that highlights the not-only-rational ways in which visualisations work. In short, they can completely uproot everyday cognition, in the sense of providing intellectual and/or emotional stimuli that are so novel to the viewer that the act of consciously exploring a visualisation is on its own so disruptive that some beliefs are immediately questioned. This questioning is a fertile ground for empowerment.

One way how quite unforeseen facts can be brought to life is by simply combining datasets in the right way. We argued that using datasets that are gathered by someone else is better than gathering your own data because of subconscious skew. For the same reason, it is even better to combine different datasets, so as to make any skew present in any of the used datasets as insignificant as possible.

4   CASE STUDY

In order to give an example of the points we are adding to the discussion, we will discuss the visualisation *Spies in the Skies*, published by BuzzFeed News on April 6th, 2016[6].

The *Spies in the Skies* visualisation

Spies in the Skies contains flight tracking information of aircraft operated by the FBI and DHS in the United States.

Tracked routes are shown on a map. FBI planes are visualised in red, and DHS planes in blue. As viewer, we can navigate to any point in time within the recorded time frame, which spans over a bit more than four months. Flight routes are painted in very thin lines. The actual discrete recorded positions are shown as dots on the map and are only visible if we navigate to the date on which they were recorded.
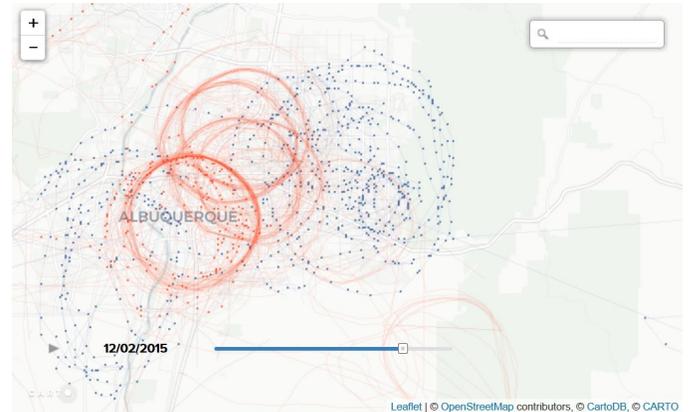

Figure 1: Albuquerque NM.

On the map in figure 1, we see the city of Albuquerque and the recorded air traffic in its vicinity. We can see that on December 2nd, 2015, a DHS plane flew in somewhat circular shapes over it. There is some FBI activity on that day as well, although the red non-dotted circles indicate that on some other day (or days), there was much more intense FBI activity in this area. Almost all the recorded FBI data belongs to flights that circled many times around a specific location, while it is unclear if the same can be said about the DHS activity on December 2nd.

Let us now share three insights we gained from using it (together with Google), before analysing some of the points proposed above in order to see in what ways it is an explicitly empowering visualisation.

Georgetown KY, 8.10.2015 After a threatening vandalism is found on a bathroom wall at Eastern Kentucky University[7], a DHS aircraft circles the city multiple times.

Inwood NY, 21.9.2015 A demonstration for more affordable housing by a mere 30 participants[8] drew an FBI aircraft to circle the site multiple times.

---

[5]Emphasis mine.
[6]https://www.buzzfeed.com/peteraldhous/spies-in-the-skies

[7]https://edition.cnn.com/2015/10/07/us/eastern-kentucky-university-threat/index.html
[8]https://www.nytimes.com/2015/09/22/nyregion/new-york-zoning-plan-requires-more-affordable-homes.html

Stockton AZ, 18.9.2015 In preparation of a massive crackdown on street gangs[9], several FBI aircraft circle the area around Stockton AZ in the preceding days.

Analysis

Before we analyse some of the points made above, let us start with stating the obvious. The FBI is a secret service, and the DHS works to secure the United States' borders, so we are dealing with sensitive information in this visualisation. In fact, we will see some of the efforts made by the creators to bring this information to light in the following. That alone does not make a visualisation empowering, yet it is also hard to argue that having information at hand that other people are working hard to keep from you is not empowering in itself. We will now discuss what it is concretely that makes this visualisation such a good example for empowerment.

Disclosure. Here, there is nothing left to be desired. The creators are disclosing where their data come from, so it can be checked by any viewer. Furthermore, the complete code used to transform and visualize the data is published as well and linked in the article[10]. This allows for the tightest of scrutinies to be applied by anyone willing to do so. At the same time, it is an instance of what D'Ignazio and Klein call *make work visible.*

Plurality and contingency. Plurality might not strictly be given. We only see the flight patterns of these two agencies, and not the reasons behind the flights. Thus, only one side of the story is shown. It is however easy to find additional information for at least some of these patterns, as we have seen above. *Contingency* is perfectly given. It is possible to draw widely varying conclusions from this visualisation. Consider the flight patterns over Los Angeles and over the "Big Sky" states (see appendix A). While in the Los Angeles area, some plane or helicopter seems to be in the air on practically any given day, the three states of Montana, Idaho, and Wyoming had had next to zero visits from these two agencies[11]. Thus, depending on what aspect of the visualisation the viewer focuses on, she might reach totally opposing conclusions.

Completely reader-driven. This is clearly given. You have an option to play the visualisation from start to beginning, but the flow of time is the only thing guiding this sequence—there is, especially, no pre-arranged martini glass-like intro

sequence. Other than that, as a viewer, you are free to zoom in or out at any point, and select any point in time within the recorded time frame. All hints and discoveries that are given about the data are in the framing article, not in the visualisation itself[12].

Combining open source datasets. This is the real strength of this visualisation. Underlying it are two completely different datasets, both in the public domain. The first is a piece of investigative journalism itself: On June 2nd, 2015, the Associated Press released information linking more than 50 planes to fake companies used as a front by secret services[13]. For this article, BuzzFeed News drew from other sources and extended that initial list. The second dataset is the massive open source data repository of FlightRadar24, a Swedish for-profit company that is collecting and visualising flight information, both gathered with their own hardware as well as crowdsourced hardware. Combining these two datasets produces a cognitive artifact that is quite unlike anything else and can hardly have been expected by the creators of any of the datasets.

5 FURTHER RESEARCH

In this work, we have rather speculatively argued for empowering data visualisation to embrace a few practices. These are of course mere conjecture until empirically proven. Another open question is, frankly, whether 'empowering data visualisations' are a thing at all. It would be a gigantic work to investigate the concrete influence of some visualisations.

6 CONCLUSIONS

We have brought forward some points as to what makes a data visualisation empowering. We then demonstrated these points on a concrete visualisation, *Spies in the Skies.* This visualisation is empowering for many reasons. It provides the reader with unbiased information on a topic on which information is hard to come by and which is relevant in important areas, such as civil rights. By combining two datasets created completely independent of each other, it leverages the full potential of visualisations by presenting truly unanticipated results. It is quite plausible that a big part of a viewer's knowledge about secret services is overturned by consuming this visualisation. That, in turn, opens up a space for critical evaluation and civil action.

---

[9]http://eastcountytoday.net/stockton-police-take-down-two-notorious-street-gangs/
[10]http://buzzfeednews.github.io/2016-04-federal-surveillance-planes/analysis.html
[11]It is important to note here that the article itself surmises that is almost certainly underestimates the flight traffic, as there may well be other, secret planes and helicopters in the air.

[12]For instance, the authors rightly find it interesting that while the two agencies claim that these flights are always related to a concrete case, their rate drops by around 70% over the weekends.
[13]https://www.msn.com/en-us/news/us/fbi-behind-mysterious-surveillance-aircraft-over-us-cities/ar-BBkywnm

ACKNOWLEDGEMENTS

REFERENCES

[1] Michelle A. Borkin, Azalea A. Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. 2013. What Makes a Visualization Memorable? *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2306–2315.

[2] Jeremy Boy, Jean-Daniel Fekete, and Françoise Détienne. 2015. Storytelling in Information Visualizations: Does it Engage Users to Explore Data? *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)* (2015). DOI : http://dx.doi.org/10.1145/2702123.2702452

[3] S. K. Card, J. D. Mackinlay, and B. Shneiderman. 1999. *Information visualization: Using vision to think.* Morgan Kaufmann Publishers.

[4] Catherine D'Ignazio and Lauren F. Klein. 2016. Feminist Data Visualization. *IEEE VIS* (2016).

[5] Marian Dörk, Christopher Collins, Patrick Feng, and Sheelagh Carpendale. 2013. Critical InfoVis: Exploring the Politics of Visualization. *CHI 2013 Extended Abstracts* (2013).

[6] Marijn Janssen, Yannis Charalabidis, and Anneke Zuiderwijk. 2012. Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management* 29 (2012), 258–268.

[7] Anshul Vikram Pandey, Anjali Manivannan, Oded Nov, Margaret L. Satterthwaite, and Enrico Bertini. 2014. The Persuasive Power of Data Visualization. *New York University Public Law and Legal Theory Working Papers.* Paper 474. (2014).

[8] Edward Segel and Jeffrey Heer. 2010. Narrative Visualization: Telling Stories with Data. *IEEE Trans. on Visualisation and Computer Graphics* 16, 6 (2010), 1139–1148.
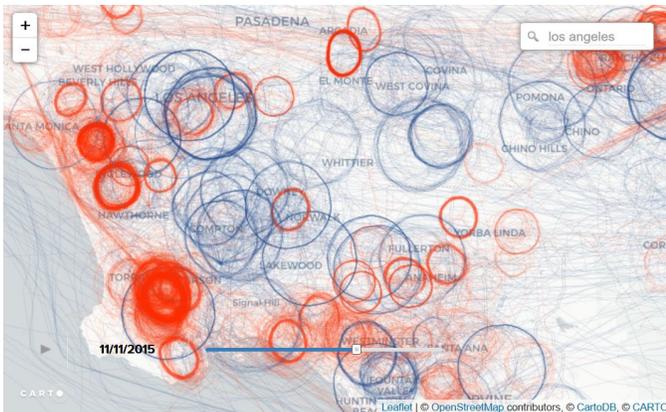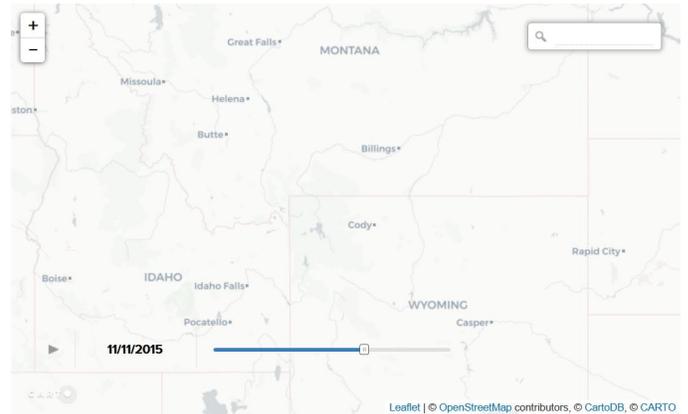
A   HIGH AND LOW AIR TRAFFIC

Figure 2: Greater Los Angeles CA.

Figure 3: Idaho, Montana, Wyoming.