

La visualisation des données

un regard sur son passé, son présent et son futur

Clément Levallois

Version 1.0, 2017-31-07

Table of Contents

1. Où va la visualisation des données?	1
2. Avant la dataviz	1
a. Infoviz.....	1
b. Infographie	2
c. Une autre communauté: la business intelligence	3
d. Et GIS.	3
e. La scène composée par infovis, infographie, BI et SIG	3
3. L'émergence de dataviz	3
a. Individuals possessing an unusually broad mix of skills:	4
b. Twitter based communication around the "#dataviz" hashtag	5
c. A tight knit group across the US and Europe.	5
d. A couple of emblematic projects	5
e. Two aspects where data visualization epitomizes its value: maps and networks.	6
f. If we were looking for 2 defining traits of dataviz	7
4. 2014-2015: The stabilization of dataviz	8
5. 2015 onwards: where is dataviz going?	9
The end	10

1. Où va la visualisation des données?

par <https://www.clementlevallois.net> [Clement Levallois], adapté d'une présentation faite à [DataStorm 2ème édition](#), Lisbonne, 15 juillet 2015.

La bonne chose à propos des champs en science computationnelle est qu'ils évoluent si vite. Tous les six mois environ, un nouveau venu arrive en apprentissage automatique, le développement d'applications mobiles ou l'exploration de texte. Il est donc difficile de rester à l'avant-garde de ces domaines, et parfois nous pourrions même perdre de vue la signification et les objectifs de ce qui nous intéressait dans le domaine, il y a seulement quelques années.

Je profite donc de cette conférence pour réfléchir à la visualisation de données, qui est un domaine si jeune. J'aimerais explorer la façon dont la visualisation de données a évolué, pourquoi il fallait qu'elle émerge, où elle se situe aujourd'hui, et je vais essayer d'imaginer où elle évoluera dans les années à venir.

Je suis plus un observateur qu'un participant dans ce domaine.



Mon point de vue est celui de quelqu'un qui est venu à la visualisation de données vers 2009 grâce à des visualisations de réseau avec [Gephi](#), obtenant mes informations principalement des discussions, liens et podcasts partagés par les visualiseurs de données que je suis et avec qui j'interagis sur Twitter.

Je sens qu'au cours des six dernières années, la **dataviz** a évolué de manière significative: elle est apparue et a cristallisé en un sujet distinct, a vécu à travers un âge d'or, et nous sommes aujourd'hui ailleurs. Commençons par le début.

2. Avant la dataviz

Avant la dataviz, il y avait quelques champs de pratique établis traitant des graphiques et des données. Je mentionnerai 4 d'entre eux mais il y en a sûrement d'autres importants :

infoviz, infographie, Business intelligence, et SIG.

a. Infoviz

Infoviz est la "visualisation d'information" et est le nom d'un domaine académique concerné par:

l'étude de représentations visuelles (interactives) de données abstraites pour renforcer la cognition humaine.

La force de ce domaine est qu'il est un banc d'essai pour de nombreuses hypothèses que vous

auriez sur la façon dont les visualisations peuvent être efficaces, en termes de «renforcement de la cognition humaine»:

- Comment fonctionnent les couleurs?
- Qu'en est-il des différentes échelles?
- Quelle est la performance des utilisateurs pour résoudre des problèmes donnés lors de la lecture du graphique représenté sous forme de matrice, par rapport aux graphiques représentés en tant que réseaux? Les données longitudinales doivent-elles être présentées sous forme de tranches de temps ou de films d'animation pour une meilleure compréhension?

La recherche en visualisation d'information vise à fournir des réponses solides et empiriques à ces questions importantes.

[exemple infovis] | *example-infovis.jpg*

Le problème est que cette emphase sur les tests d'hypothèses peut être préjudiciable à l'avancement sur un autre objectif fondamental: créer des visualisations **engageantes pour leur public**, largement adoptées et utilisées dans un monde si plein d'informations (ou de bruit). Cette attention par les individus devient la plus rare des ressources. Comme illustré par la figure ci-dessus, et si je devais être un peu dur vis à vis de l'infoviz, je dirais que vous avez besoin d'un doctorat pour la comprendre : les interfaces qu'ils conçoivent ne sont pas assez engageantes, pas implémentées sur les plateformes que nous utilisons et l'information affichée améliore à peine la cognition humaine pour les non-spécialistes.

b. Infographie

On pourrait dire que l'infographie est un peu le contraire d'infovis: les agences de communication font à peu près ce qu'elles veulent pour attirer l'attention de leurs lecteurs, au détriment de la véracité et de la fiabilité des données qu'elles invoquent. L'exemple ci-dessous montre à quel point une infographie peut être colorée et accrocheuse, mais totalement inefficace pour transmettre des informations.

[exemple infographics] | *exemple-infographics.png*

Bien sûr, il existe d'excellentes infographies et Alberto Cairo, professeur et journaliste de profession, nous rappelle dans son livre [The Functional Art](#) que l'infographie soigneusement réalisée est un excellent moyen de transmettre des informations complexes dans une quantité limitée d'espace. Mais je crois comprendre que ce n'est pas dans le contrat de base de l'infographie d'avoir une relation univoque les données, il y a la permission d'**illustrer** les données. Le lecteur doit faire davantage confiance au media qui publie l'infographie que dans le cas d'une infovis: selon qu'il s'agisse d'un journal établi avec une bonne équipe graphique ou d'une agence de communication faisant du travail rapide et baclé, on peut faire confiance aux infographies ou se faire induire en erreur.

c. Une autre communauté: la business intelligence

[example bi] | *example-bi.png*

La mission en BI consiste essentiellement à réaliser des visualisations «au niveau excel» en termes de reporting et de suivi des données métiers. Rien d'extraordinaire ici: bar charts, camemberts (souvent en 3D comme dans l'illustration ci-dessus, ce qui est mauvais), graphiques linéaires et barres de progression assemblés dans des tableaux de bord, vendus par des entreprises plus expérimentées dans le développement logiciel qu'en graphisme.

d. Et GIS.

image

formatée / gis.jpg [align = "center"]

Geographical Information Systems (GIS) peut avoir une réclamation pour la plus longue tradition dans la visualisation des données. C'est après tout leur travail pour dessiner des cartes, qui est données géolocalisées. Il se pourrait que cette longue tradition fût aussi une malédiction: parce qu'ils ont développé ces logiciels de bureau largement utilisés dans les années 1990, les années 2000 et encore aujourd'hui, ils étaient ancrés dans des technologies qui ne pouvaient pas être facilement adaptées lorsque les technologies web devenaient plus riches. des façons plus attrayantes de dessiner des cartes et de projeter des superpositions de données sur celles-ci.

e. La scène composée par infovis, infographie, BI et SIG

La scène est donc la suivante: les scientifiques dans le domaine de la "visualisation de l'information" dans leur coin étant les gardiens du temple des "visualisations correctes", mais ils ont du mal à trouver un public pour ces graphismes.

Infographie dans le coin opposé, qui ont accès à des foules de lecteurs tous les jours dans les pages de journaux et de brochures marketing, mais avec le sentiment qu'ils ne montrent pas vraiment les données - ils éditorialisent beaucoup, pour le meilleur ou pour le pire.

Dans l'un des deux autres domaines, nous avons une intelligence commerciale qui est un peu méprisée en raison de la simplicité de leurs graphiques qui ne rend pas justice à la richesse des données, mais enviée parce qu'ils ont accès à des données pertinentes, coûteuses et percutantes. .

Et SIG qui travaille avec des données d'une manière qui est universellement comprise et jugée pertinente (cartes), mais avec un degré d'innovation de ce domaine qui reste assez faible.

3. L'émergence de dataviz

Quelque chose s'est passé autour de 2008 et 2009, ce qui a changé ce statu quo. Un certain nombre de bibliothèques de graphiques et de dessins javascript ont été publiées:

- [RaphaelJS](#) (08/08/08)
- le [Javascript Infovis Toolkit](#) (2009)
- [Protovis](#) (2009)
- [Processing.js](#) (2010)
- et [D3](#) (2011), désormais le framework le plus performant pour dataviz avec les technologies web.

Avec le décollage des mobiles sans les plugins Flash et Java (rappelez-vous: l'iPhone était sorti en 2007 et ne supportait pas Flash), la popularité décroissante du plugin Java même sur les navigateurs de bureau, vous voyez en 3 ans un grand shift: unification des frameworks de visualisation sur le web en utilisant javascript.

Le web devient de plus en plus une plate-forme en soi (plus populaire que le lancement de logiciels de bureau), avec la sortie de Google Chrome en 2008 - Javascript et CSS deviennent beaucoup moins cassés que lorsque Internet Explorer était dominant. Pour quel impact?

Il a brouillé les cartes: avec Java est venu un moyen très rigide de concevoir des interfaces: les fenêtres, les menus et même les polices avaient un aspect Java dans le navigateur.

Avec Flash, vous aviez un solide historique d'interaction et de compétences en conception, mais vous pouviez utiliser Flash sans codage, de sorte que les conceptions créées avec Flash pouvaient rester assez déconnectées des jeux de données qu'elles représentaient.

Tout ce qui est devenu jeté dans le melting-pot de Javascript où tout le monde a dû désapprendre leur cadre et apprendre sur une terre vierge.

La visualisation des données n'était pas la progéniture naturelle de l'un des 4 champs que j'ai mentionnés, il est apparu en dehors d'eux.

Cela a amené de nombreux nouveaux venus à s'essayer à ces nouveaux outils, libérés des habitudes et des conventions des quatre domaines que nous avons vus.

Ces nouveaux venus qui ont créé dataviz avaient une manière différente de regarder les choses, un outil différent, et différentes façons de fonctionner en groupe. Cette communauté est remarquable à plusieurs égards:

a. Individuals possessing an unusually broad mix of skills:

Coding skills for the preparation of the data (Python or R for example), skills in javascript and other scripting language for visual design (ActionScript, Processing), a knowledge of the rules of design and a feel for esthetics, and creativity. That is what you need to create this:

[mta] | *mta.jpg*

(live url: <http://www.mta.me>) (by Alexander Chen, a Creative Director at Google Creative Lab)

b. Twitter based communication around the "#dataviz" hashtag

In this community, people evaluate each other's works, shared their latest realization chat about past and upcoming conferences but more importantly exchange info about new frameworks and resources.

[dataviz communities] | *dataviz-communities.jpg*

(live url: <http://neoformix.com/2012/DataVisFieldSubGroups.html>)

c. A tight knit group across the US and Europe.

I identify (this is a non exclusive list of course) [Santiago Ortiz](#), [Jerome Cukier](#), [Jer Thorp](#), [Gregor Aisch](#), [Jan Willem Tulp](#), [Lynn Cherny](#), [Nathan Yau](#) from Flowing Data, [Kim Rees](#) from Periscope, [Moritz Stefaner](#), with a couple of established academics like [Enrico Bertini](#), [Scott Murray](#), [Jon Schwabish](#), [Alberto Cairo](#), and in relation with teams at the Guardian and the NYT, and [Andy Kirk](#) at VisualisingData as an evangelist and instructor.

They were particularly active in spreading news about dataviz and sharing their critical insights which contributed shaping boundaries for the field.

This is a personal and of course biased observation, a systematic investigation reveals a different picture (see above, and below, which is a zoom on the group where I think we would find most people self identifying as dataviz specialists):

[dataviz group] | *dataviz-group.jpg*

(live url: http://neoformix.com/2012/DataVisField1000_Group2.pdf)

d. A couple of emblematic projects

i. OECD Better Life Index by Moritz Stefaner et al

Not infovis, not infographics, just dataviz: simplicity, interaction, access to the data.

[oecd better life index] | *oecd-better-life-index.jpg*

(live url: <http://www.oecdbetterlifeindex.org/>)

ii. The "Ghost Counties" visualization by Jan Willem Tulp

It shows that a marriage is possible between creativity and esthetics on one hand, and cold hard

data on the other hand (foreclosures per county in the US).

[ghost counties screenshot] | *ghost-counties-screenshot.jpg*

(live url, needs Internet Explorer and the Java plugin: <http://www.janwillemtulip.com/eyeo/>)

iii. U.S. Gun Deaths by Periscopic

It illustrates the power of story telling (through the intro), granularity of the data, and impact.

[gun deaths] | *gun-deaths.jpg*

(live url: <http://guns.periscopic.com/?year=2013>)

The emergence of data visualisation as a set of practice and professionals was coinciding with the surge in the new importance of data as a driver of value for business.

"Data visualization" became positioned as one powerful lever to extract value from datasets: it possesses both the rigor needed to report objectively on key data features, that you'd find otherwise in information visualisation, and the power to be engaging with the domain specialists or the managers in charge of finding insights in the data.

e. Two aspects where data visualization epitomizes its value: maps and networks.

i. Maps

Visualization of geolocalized data and of network data has of course a long history before the birth of data visualization: many software integrated mapping functions from Geographical Information Systems, and network analysis packages also had visualization add-ons.

What data visualization brought was impactful visualizations making engagement with data just stronger, more powerful.

Stamen, an agency with strong ties in the data visualization community, does this kind of maps:

[stamen viz] | *stamen-viz.jpg*

(live url: <http://prettymaps.stamen.com/201008/#10.00/38.7250/-9.1500>)

This interactive map by Stamen is quite different from your usual GIS mapping! What this kind of map brings is: interaction, custom-made design, and most of all enhanced **engagement** with the viewers.

ii. Networks

In terms of networks, a pre-dataviz typical network would look like:

[ucinet] | *formatted/ucinet.jpg*

Dataviz brought interaction, web-based interactions:

[d3 force layout] | *d3-force-layout.jpg*

(live url: <http://bl.ocks.org/mbostock/1062288>)

This type of visualization is different because:

- you can explore the viz, not just stare at it.
- you can share it - just paste the url.
- it can be developed and modified by a large pool of developers because it is written in javascript, which is the common language of web development.
- there is a strong sense of aesthetics and natural feeling using it.

→ it will encourage curiosity, exploration, and just increase 10 folds the time spent on it by the viewers.

f. If we were looking for 2 defining traits of dataviz

i. Data is for the viewer to see and play with

There is the assumption that the visualization should not provide you with flat and unverifiable conclusions: it should show the data in a transparent, verifiable form.

Of course there is a narrative and an editorialization of how the data is presented, **but** it always remains possible for the viewer to challenge this editorial view because the data is here for anyone to explore and interact with. This represents a fundamental break with infographics, which can hide the underlying data by design, or show it with strong bias by carelessness and still be "OK" by pre-dataviz standards. It is also a break with infovis, where data is indeed there but you might not be enticed to engage with it.

ii. Custom made, creative act

Because we are in the browser there is no click and point solutions for the visualization of the data.

This departs strongly from GIS where "custom" maps could be done by selecting options in a menu, and also a big change from dashboards in business intelligence where you could drag and drop charts to build a visualization.

The sense of esthetics and the particularity of the datasets makes of each dataviz a craftwork.

One of the best examples of a creative and simple design is this one by Hint.fm:

[windmap] | *formatted/windmap.jpg*

(live url: <http://hint.fm/wind/>)

(live url for a worldwide version: <http://earth.nullschool.net/>)

4. 2014-2015: The stabilization of dataviz

Anyhow, industrialization in dataviz came in rapidly, with Tableau becoming the leader for general purpose viz, dashboards reinvented themselves in dataviz-style with Bime, Qlik, Palantir to name a few.

[logos bi] | *logos-bi.png*

Dataviz became integrated into the business discourse on big data: the Harvard Business Review features in 2012 a blog section on data visualization where Jer Thorp contributed to set perspectives straight on data,

[jer thorp] | *jer-thorp.jpg*

(live url: <https://hbr.org/2012/11/data-humans-and-the-new-oil/>)

Nielsen, the leader of market data and market research, worked on its corporate identity to include data visualization, with data-driven visuals custom made by Jan Willem Tulp:

[nielsen viz] | *nielsen-viz.jpg*

Since 2012 or so, General Electric partners with Fathom, the agency founded by Ben Fry (co-creator of Processing!) to build visualizations relative to their corporate identity, with some impressive realizations:

[ge] | *formatted/ge.jpg*

(live url: <http://visualization.geblogs.com/visualization/powering/>)

And in 2015, you know dataviz has fully stabilized when you see a panel on dataviz with Chelsea Clinton:

(live url: <https://www.youtube.com/watch?v=YFrmQDCpgxs> - the panel is with Ben Fry).

So until 2012 and 2013 I'd say that we were in the golden age of #dataviz in terms of discoveries and charting new paths: excited comments on new productions by the NYT, debates around the goals of #dataviz: is it a way to tell stories? To open new worlds? To educate? New connections made with new comers, new agencies, people meeting for the first time in conferences after exchanging on Twitter for years, new positions, big clients... And in 2015, things seem to have stabilized and normalized.

The energy has changed. The conversation on Twitter has slowed down a lot. The sense of being pioneers has eroded, because time has passed and because we have indeed tried and explored many low hanging fruits. Many individuals are now engaged in more industrial, long term projects.

So that's not bad news: dataviz is now mainstream and well established, people are less obliged to enter free competitions and work on long personal projects at weekends and nights to get their name out, that's good. But I miss a bit the excitement of the previous years when you had one framework or one big personal project published per month, and when you had all these big shots chatting on Twitter about the upcoming developments for dataviz.

5. 2015 onwards: where is dataviz going?

So... where is dataviz going? As I said, you have this first exciting phase that passed, and we are now in a stage where processes for the creation of dataviz are more industrialized, commodified, stabilized.

This means that innovation will find other places to erupt. Why? Because the landscape of technologies keeps changing, and creative minds will seize the opportunity to play and explore these opportunities in places where no "client" is yet waiting for them. To illustrate possible paths, I like to give the example of the career of [Seb Lee-Delisle](#), who defined himself as a creative coder and now as a digital artist.

I follow his work on Twitter since about 2009. He is not at the heart of the "dataviz" network and does not define himself in regards to this label, but you'd find him on Jeff Clark's map of dataviz in 2012 nonetheless (see map above).

- he was using Adobe Flash as one of his main technologies until 2009, contributing to [PaperVision3D](#), a framework to build 3D games and animations in the Flash Player.
- He plays a bit with [Adobe Flex](#) in 2009,
- in 2010, Flash is definitely behind so he moves to HTML5 technologies, using and teaching [animated graphics in HTML5 + Javascript](#)
- in 2012, he does the lunar trail project: <http://seb.ly/work/lunar-trails/>
- in 2013, he does pixelpyros: <http://pixelpyros.org/>
- in 2014/2015, he launches workshops on "Stuff that talk to the Internets": <http://seb.ly/st4i->

This path, and similar paths followed by others, suggest that:

- The computer screen and even the screen of the mobile phone is becoming less hegemonic as the medium where data can be visualized. Objects, sculptures, buildings, furniture... this is the next frontier to be explored. Not just mapping data on a flat surface, but maybe even [actual construction of data objects by Moritz Stefaner](#).
- Interaction is taking place in richer environments than we are used to (desktop or mobile), interactions with the user become more diverse. Not just the hand and the click of the mouse, but the whole body. Not one individual facing an object, but possibly a crowd, possibly moving, possibly gesturing.
- And "data" is in the process of getting an even larger meaning. When you move away from the screen and start connecting to a variety of objects and sensors, and with a variety of people, data takes still other forms: real time measurements from the external physical environment, from the internal (body) environment, from local or distant social interactions as they unfold, all while staying connected to the APIs we are already familiar with... the mix can be bring impactful results.

So, if visualizing data from the Twitter API was the cliché of #dataviz in 2010 - 2015, the next cliché could be the instantaneous 3D printing of data generated from the connected objects and bodies in a home or a workspace. This is just my vision for dataviz, I'd be happy to discuss it with you now!

Thank you!

The end

Find references for this lesson, and other lessons, [here](#).

[round portrait mini 150][align="center", role="right"]

This course is made by Clement Levallois.

Discover my other courses in data / tech for business: <https://www.clementlevallois.net>

Or get in touch via Twitter: [@seinecle](#)