

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	2
b. Infographie	2
c. Une autre communauté: la business intelligence	4
d. Et les SIG	4
e. La scène composée par infovis, infographie, BI et SIG	5
3. L'émergence de dataviz	5
a. Des professionnels avec un éventail large de compétences	6
b. Une communauté basée sur Twitter autour du hashtag "#dataviz"	7
c. Un groupe très soudé aux États-Unis et en Europe.....	7
d. Quelques projets emblématiques	8
e. Deux aspects où la visualisation de données incarne sa valeur: les cartes et les réseaux.....	10
f. Si nous recherchions 2 traits déterminants de dataviz.....	12
4. 2014-2015: La stabilisation de la dataviz	13
5. À partir de 2015 : où va la dataviz?	16
Des vidéos sur le sujet	18
Pour aller plus loin	18



1. Où va la visualisation des données?

par [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:

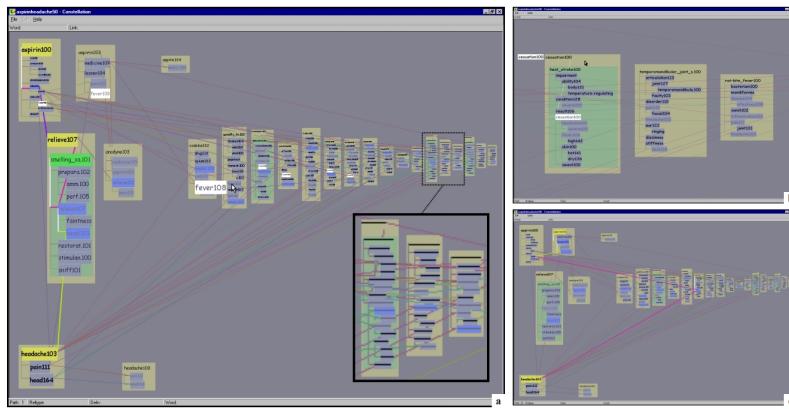
L'étude de la représentation visuelle de données, principalement abstraites, sur une interface graphique

— entrée Wikipedia "Information visualisation", https://fr.wikipedia.org/wiki/Visualisation_d%27informations

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.

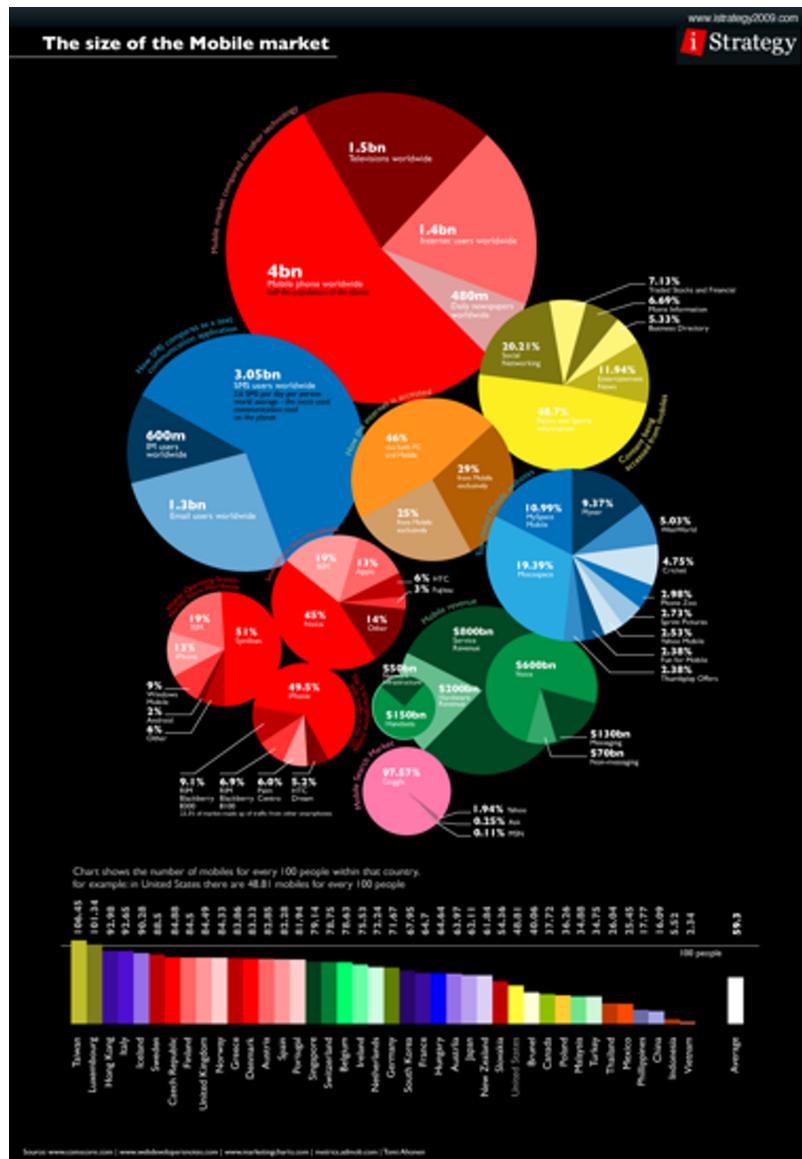


Le problème est que cette emphase sur les tests d'hypothèses peut être préjudiciable à l'avancement sur un autre objectif fondamental: créer des visualisations **engageantes pour leur public**, largement adoptées et utilisées dans un monde si plein d'informations (ou de bruit). Cette attention par les 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

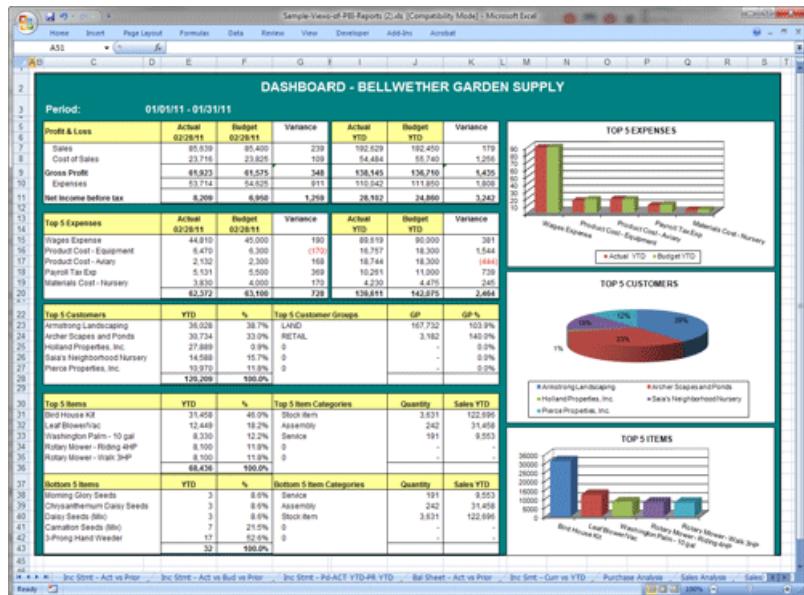
On pourrait dire que l'infographie est un peu le contraire de l'infovis : les agences de

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.



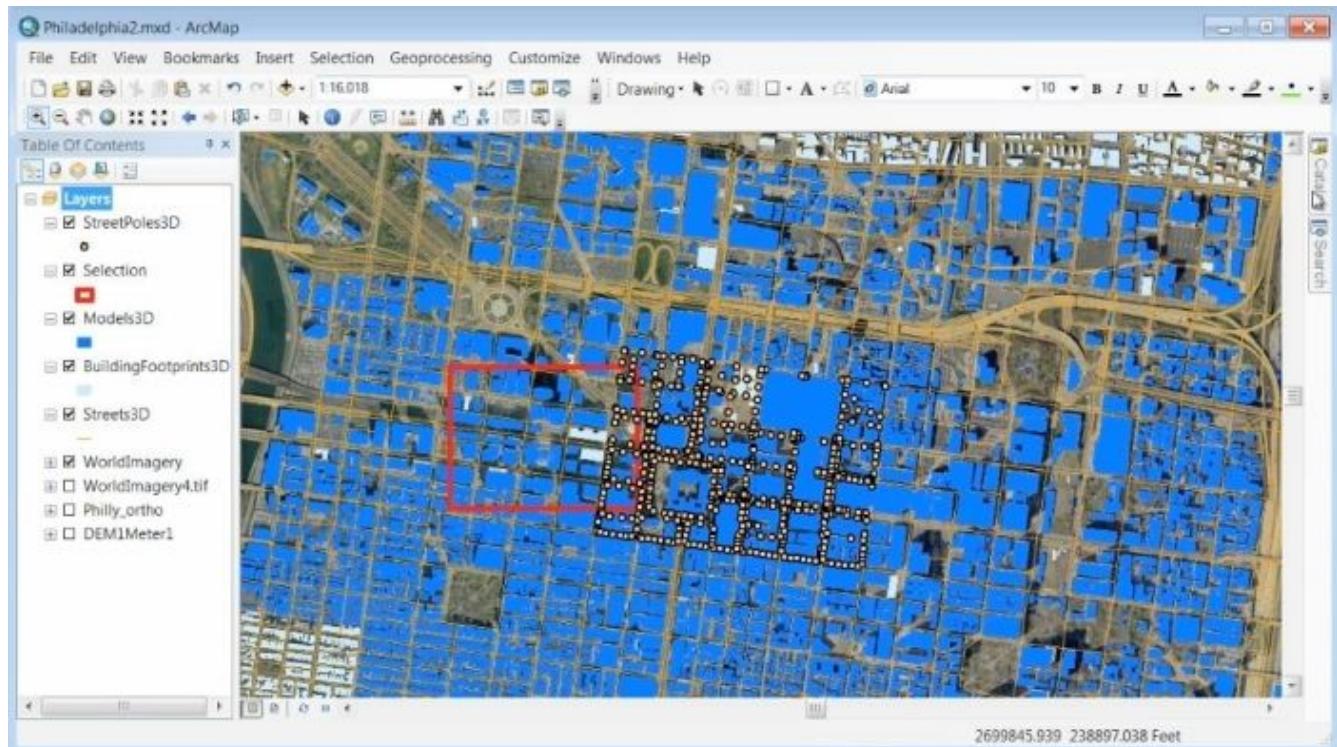
Bien sûr, il existe d'excellentes infographies et Alberto Cairo, professeur et journaliste de profession, nous rappelle dans son livre *The Functional Art* que l'infographie soigneusement réalisée est un excellent moyen de transmettre des informations complexes dans une quantité limitée d'espace. Mais 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



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 : histogrammes, camemberts (souvent en 3D comme dans l'illustration ci-dessus, ce qui est mauvais) et graphiques linéaires assemblés dans des tableaux de bord, vendus par des entreprises plus expérimentées dans le développement logiciel qu'en graphisme.

d. Et les SIG



Les Système d'information Géolocalisés (SIG) (GIS en anglais) peuvent prétendre détenir la plus longue tradition de visualisation des données. C'est après tout leur fonction première de dessiner des cartes, à partir de données géolocalisées. Il se pourrait que cette longue tradition fût aussi un handicap : parce qu'ils ont développé ces logiciels de bureau largement utilisés dans les années 1990, les années 2000 et encore aujourd'hui, les cartographies géographiques étaient ancrées dans des technologies qui ne pouvaient pas être facilement adaptées lorsque les technologies web, plus riches, sont apparues.

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.

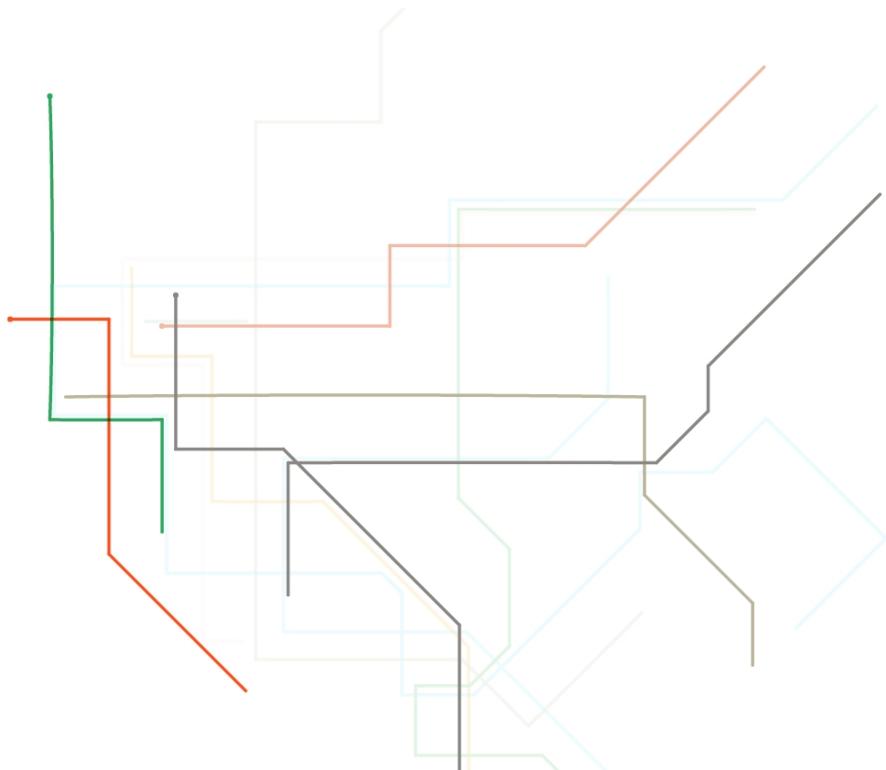
Cela a amené de nombreux nouveaux venus à s'essayer à ces nouveaux outils, libérés des habitudes et des conventions des quatre domaines que nous avons vus.

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. Des professionnels avec un éventail large de compétences

Compétences en coding pour la préparation des données (Python ou R par exemple), en javascript et autres langages de script pour la conception visuelle (ActionScript, Processing), une connaissance des règles de la conception et un sens de l'esthétique et de la créativité.

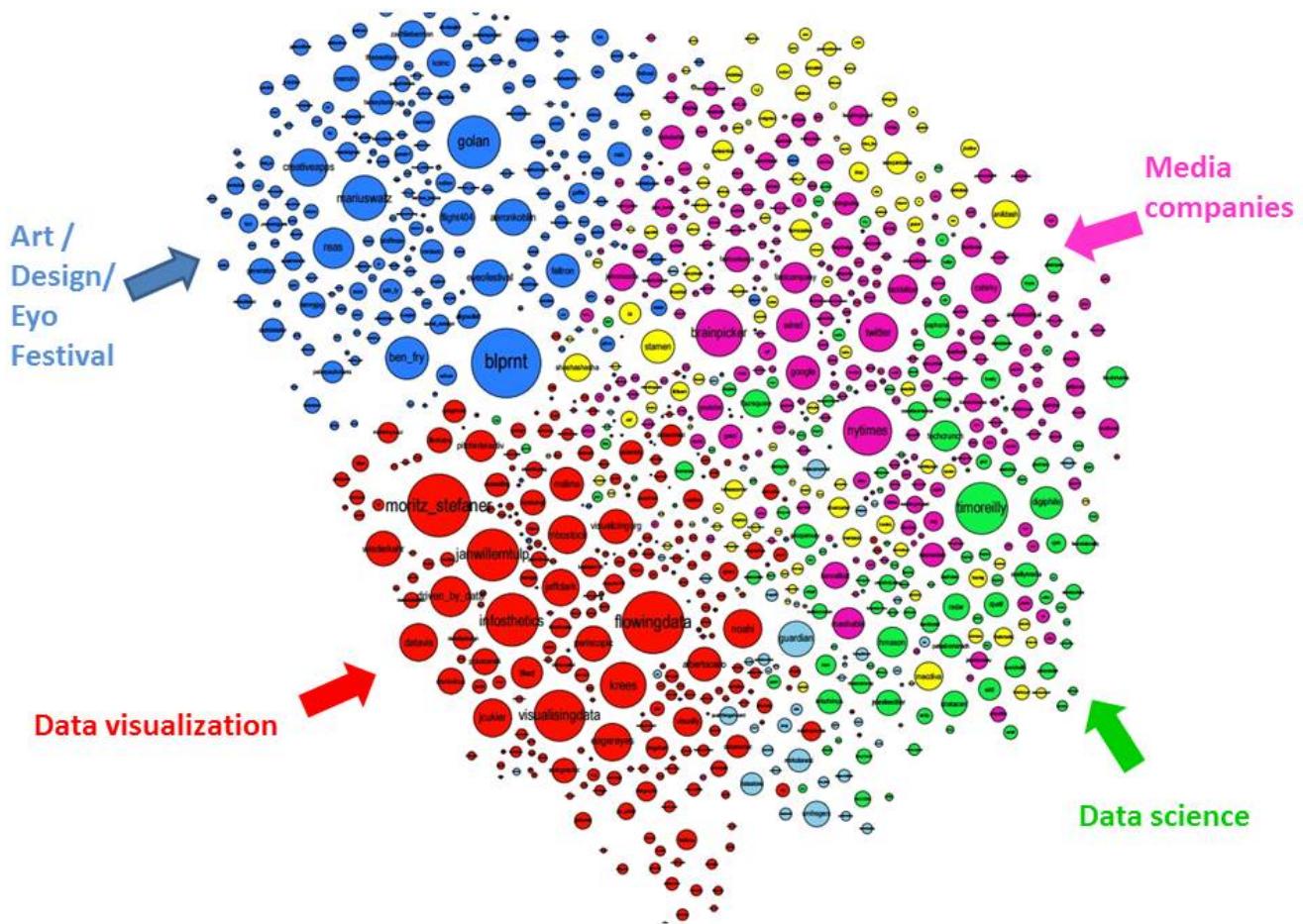
Toutes ces compétences sont nécessaires pour créer, par exemple, cette visualisation des trajets de métro à New York :



(URL : <http://www.mta.me> par Alexander Chen, directeur de la création chez Google Creative Lab)

b. Une communauté basée sur Twitter autour du hashtag "#dataviz"

Dans cette communauté, les gens évaluent les travaux de leurs collègues, partagent leurs dernières discussions sur les conférences passées et à venir, mais surtout échangent des informations sur les nouveaux développements techniques, données intéressantes et sources d'inspiration.



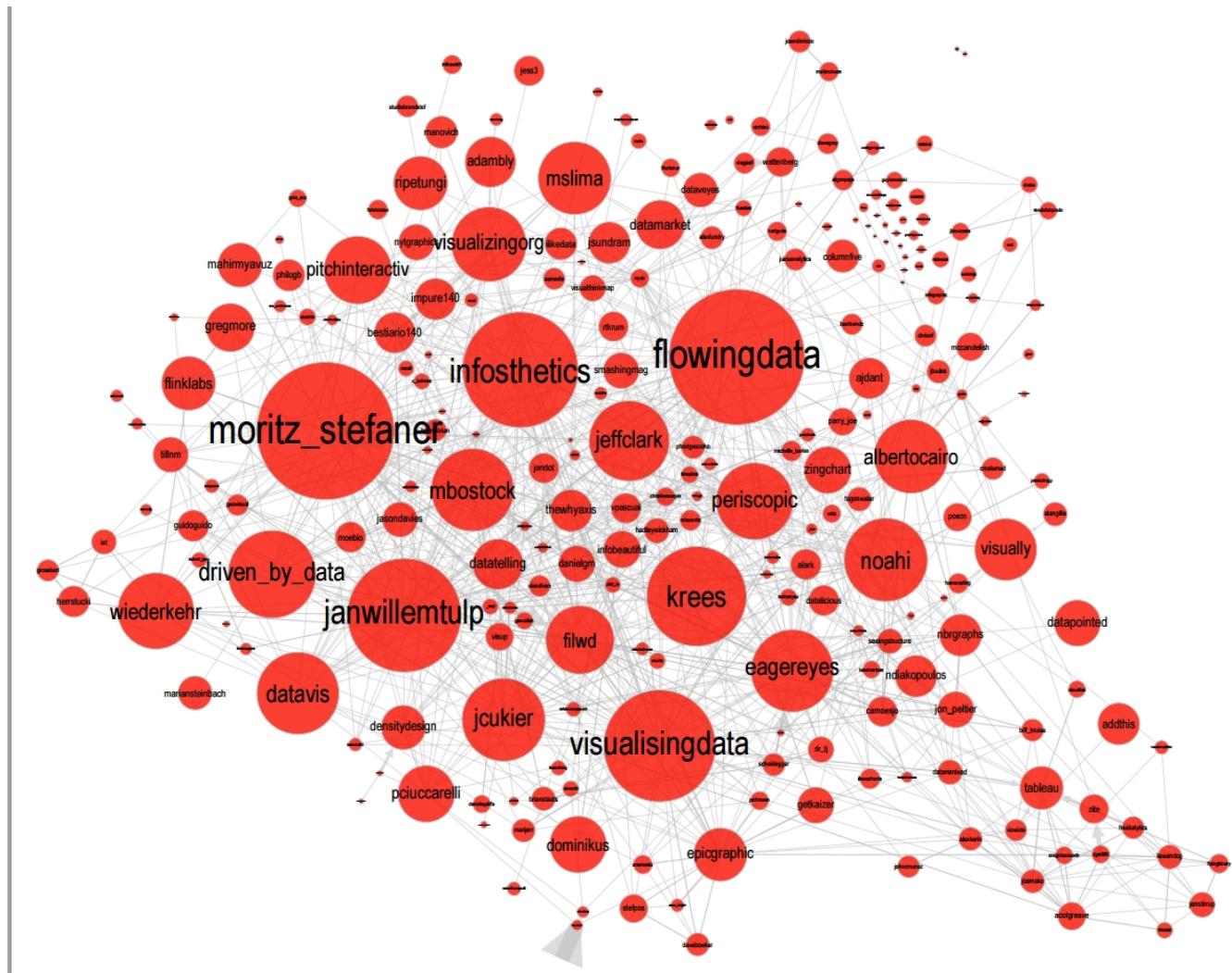
(version interactive: <http://neofromix.com/2012/DataVisFieldSubGroups.html>)

c. Un groupe très soudé aux États-Unis et en Europe.

J'identifie (il s'agit bien sûr d'une liste non exclusive) Santiago Ortiz, Jerome Cukier, Jer Thorp, Gregor Aisch, Jan Willem Tulp, Lynn Cherny, Nathan Yau de Flowing Data, Kim Rees de Periscopic, Moritz Stefaner , avec quelques universitaires établis comme Enrico Bertini, Scott Murray, Jon Schwabish, Alberto Cairo, et en relation avec les équipes du Guardian et du NYT, et Andy Kirk à VisualisingData en tant qu'évangéliste et instructeur.

Ils ont été particulièrement actifs dans la diffusion d'informations sur la dataviz. Le partage de leurs points de vue critiques a contribué à définir les limites du domaine.

Ceci est une observation personnelle et bien sûr biaisée, une enquête systématique révèle une image différente (voir ci-dessus, et ci-dessous, un zoom sur le groupe où, selon moi, la plupart des gens s'identifieraient eux-mêmes en tant que spécialistes de dataviz):



(version en ligne: http://neofix.com/2012/DataVisField1000_Group2.pdf)

d. Quelques projets emblématiques

i. Le "Better Life Index" de l'OCDE par Moritz Stefaner et al

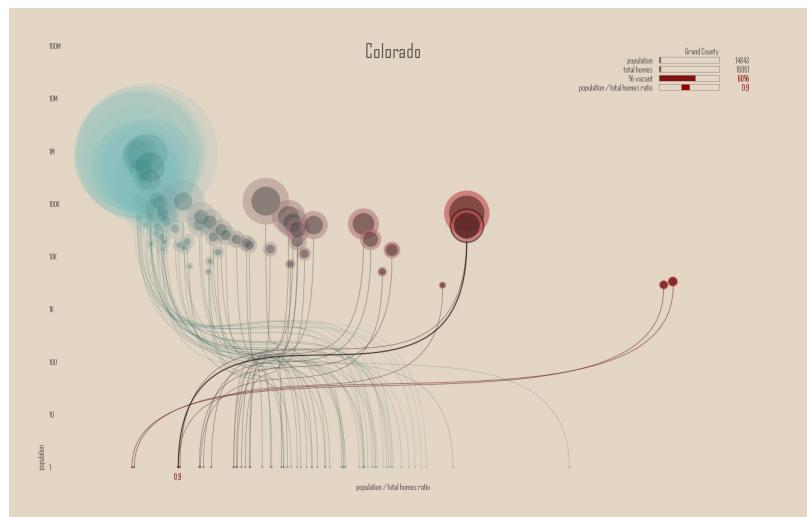
Pas de l'infovis, pas une infographie, juste une dataviz: simplicité, interaction, accès aux données.



(version en ligne: <http://www.oecdbetterlifeindex.org>)

ii. La visualisation "Ghost Counties" de Jan Willem Tulp

Tulp montre qu'un mariage est possible entre la créativité et l'esthétique, d'une part, et les données dures et froides, d'autre part (nombre d'expulsions de propriété par comté aux États-Unis).



(version en ligne, nécessite Internet Explorer et le plugin Java: <http://www.janwillemtulp.com/eyeo/>)

iii. Décès par armes à feu, une visualisation de Periscopic aux États-Unis

Cette dataviz illustre le pouvoir de la narration (à travers l'introduction), la granularité des données et l'impact.



(URL active: <http://guns.periscopic.com/?year=2013>)

L'émergence de la visualisation de données en tant qu'ensemble de praticiens et de professionnels coïncidait avec la montée en puissance de la nouvelle importance des données en tant que vecteur de valeur pour les entreprises.

La "visualisation de données" s'est positionnée comme un puissant levier pour extraire de la valeur des ensembles de données:

- la dataviz possède la rigueur nécessaire pour rendre compte de manière quantifiée des caractéristiques de données clés (ce que l'on trouverait autrement dans l'infoviz)
- la dataviz est suffisamment engageante pour permettre aux spécialistes du domaine de comprendre les enjeux du jeu de données représenté.

e. Deux aspects où la visualisation de données incarne sa valeur: les cartes et les réseaux.

i. Les cartes

La visualisation des données géolocalisées et des données de réseau a bien sûr une longue histoire avant la naissance de la visualisation de données: de nombreuses fonctions de cartographie intégrées au logiciels de SIG, ainsi que des progiciels d'analyse de réseau intégraient également des fonctions de visualisation.

Ce que la visualisation de données a apporté, c'est une visualisation percutante rendant l'engagement avec les données juste plus fort, plus puissant.

Stamen, une agence étroitement liée à la communauté de la visualisation de données, réalise ce

type de cartes:



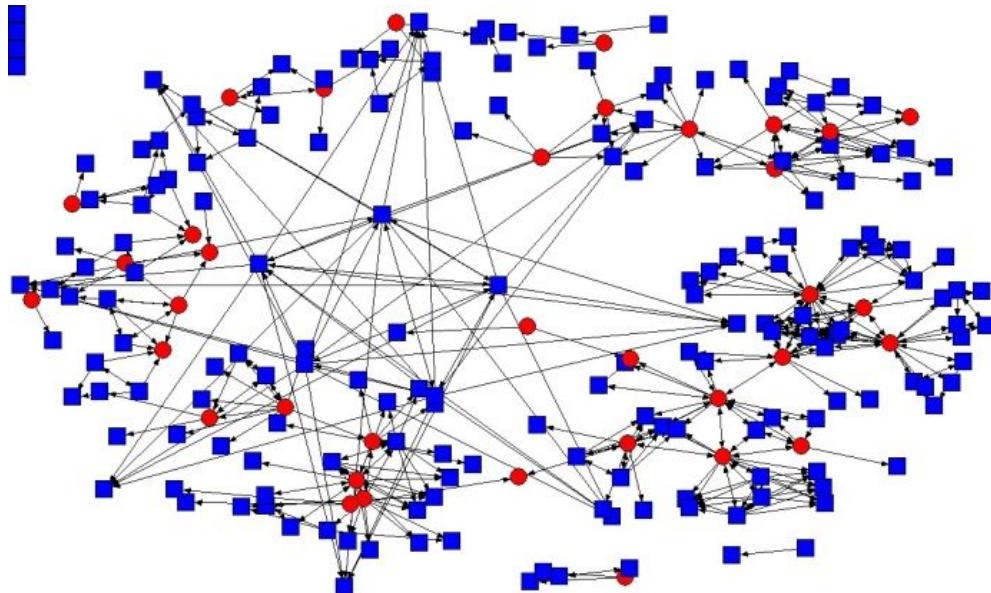
(URL active: <http://prettymaps.stamen.com/201008/#10.00/38.7250/-9.1500>)

Cette carte interactive de Stamen est assez différente de votre cartographie SIG habituelle!

Ce type de carte apporte: une interaction, une conception sur mesure et, surtout, un **engagement accru** avec l'utilisateur.

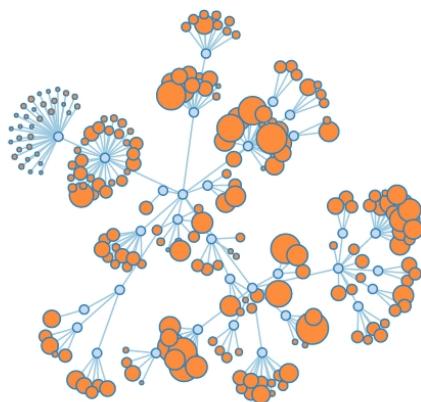
ii. Les réseaux

En termes de réseaux, un réseau typique pré-dataviz ressemblerait à ceci:



La dataviz a apporté interaction, interactions Web:

Collapsible Force Layout



(URL active: <http://bl.ocks.org/mbostock/1062288>)

Ce type de visualisation est différent pour les raisons suivantes:

- Vous pouvez explorer le viz, pas seulement le regarder.
- vous pouvez le partager - il suffit de coller l'URL.
- il peut être développé et modifié par un grand nombre de développeurs car il est écrit en javascript, langage commun du développement Web.
- il y a un sens aigu de l'esthétique et un sentiment naturel de l'utiliser.

→ cela encouragera la curiosité, l'exploration et augmentera de 10 fois le temps passé par l'utilisateur.

f. Si nous recherchions 2 traits déterminants de dataviz

i. L'utilisateur de la dataviz peut manipuler et accéder aux données qui lui sont présentées

La visualisation ne doit pas vous fournir des conclusions "imposées" et invérifiables: elle devrait montrer les données sous une forme transparente et vous donner les moyens de vérifier par vous-même.

Bien sûr, il y a une narration et une éditorialisation de la manière dont les données sont présentées, mais il reste toujours possible pour la personne qui regarde de contester cette vue éditoriale, car on peut explorer et interagir avec les données - et peut-être aboutir à des conclusions différentes. Cela représente une rupture fondamentale avec l'infographie, qui peut masquer les données sous-jacentes de leur conception, ou les montrer avec un fort biais d'inattention, tout en restant "OK"

selon les normes habituelles de ce type de représentations. La dataviz est aussi une rupture avec infovis, où les données sont bel et bien là mais vous n'êtes peut-être pas tentés de les explorer, car la présentation est abscon.

ii. Fait sur mesure, acte créatif

Parce que nous sommes dans le navigateur web, il n'existe aucune solution "click and point" pour la visualisation des données.

Cela diffère fortement des SIG où des cartes "personnalisées" pouvaient être créées en sélectionnant des options dans un menu, ainsi qu'un grand changement par rapport aux tableaux de bord de la veille stratégique, dans lesquels vous pouviez faire glisser des graphiques pour créer une visualisation.

Le sens de l'esthétique et la particularité des jeux de données font de chaque dataviz un travail artisanal.

L'un des meilleurs exemples de conception simple et créative est celui de Hint.fm:



(version en ligne : <http://hint.fm/wind/>)

(une version mondiale de cette visualisation : <http://earth.nullschool.net/>)

4. 2014-2015: La stabilisation de la dataviz

Quoi qu'il en soit, l'industrialisation de la dataviz est arrivée rapidement, Tableau étant devenu le leader des tableaux de bord (ou "dashboards"). Les dashboards se sont réinventés à la manière de la dataviz, avec notamment Bime, Qlik, Palantir, PowerBi.



La dataviz a été intégrée au discours commercial sur le Big Data: le Harvard Business Review propose en 2012 une section de blog sur la visualisation de données dans laquelle Jer Thorp "Thorp, Jer" a contribué à définir des perspectives claires en matière de données :



Jer Thorp is the co-founder of the Office For Creative Research, a multi-disciplinary research group exploring new modes of engagement with data. He is and is an adjunct Professor in New York University's ITP program. From 2010-2012, he was the Data Artist in Residence at *The New York Times*.

≡ MENU

Harvard
Business
Review

CUSTOMERS

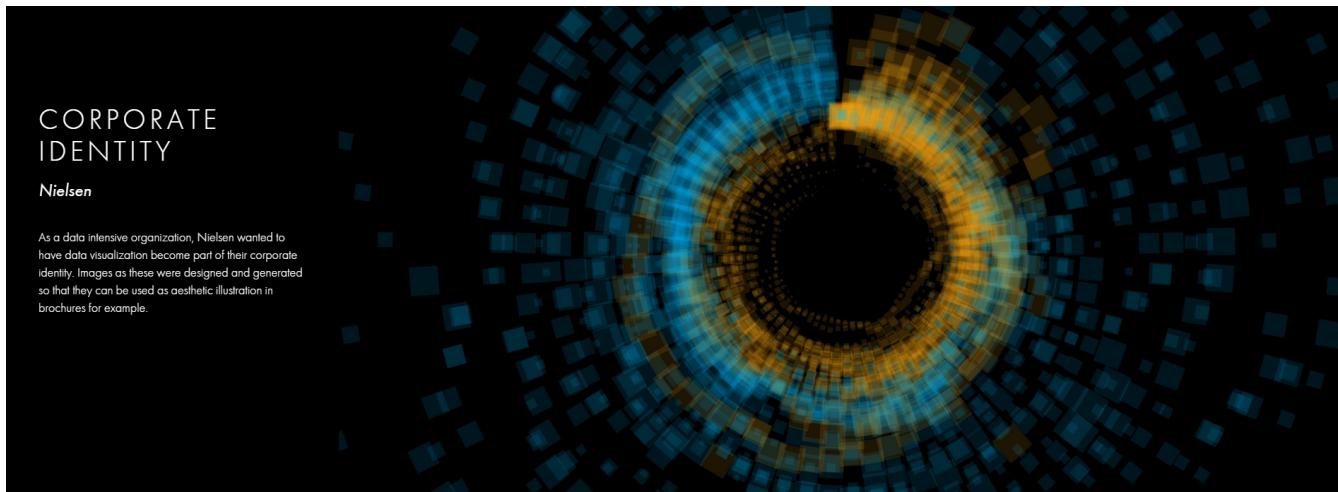
Big Data Is Not the New Oil

by [Jer Thorp](#)

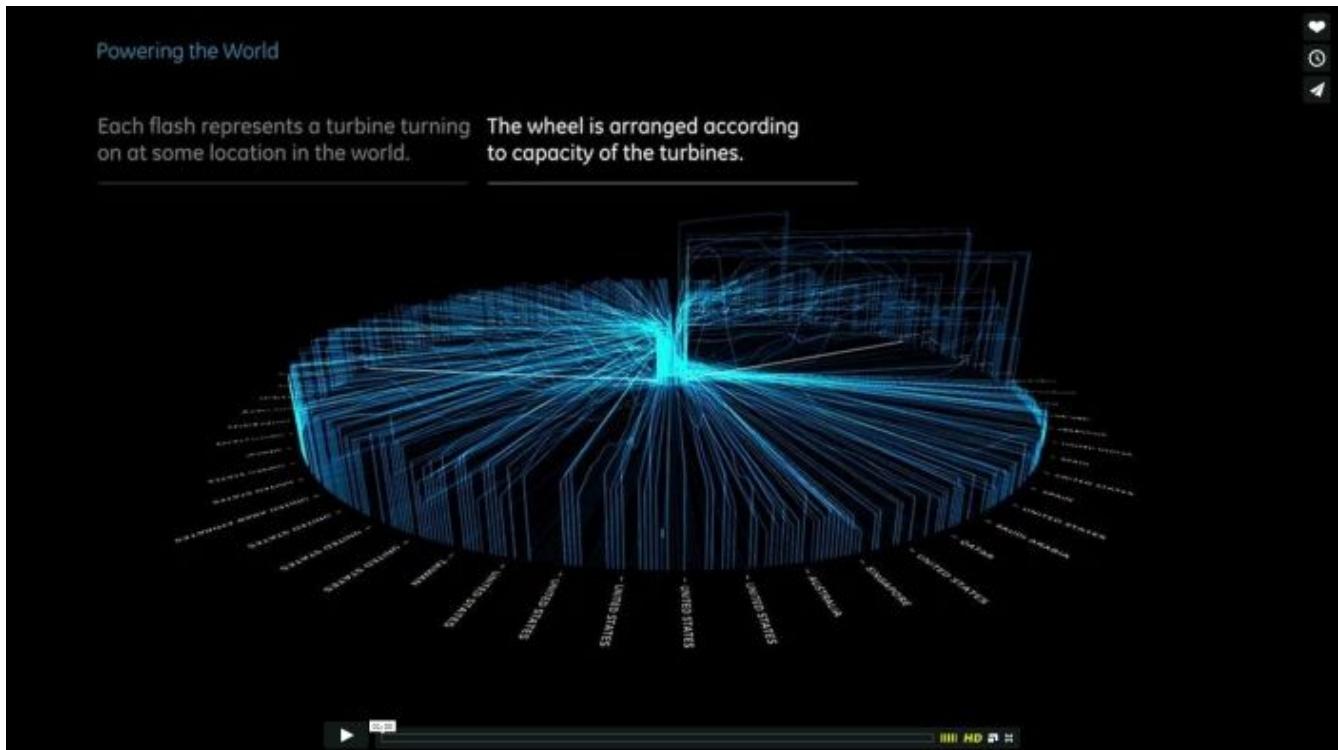
NOVEMBER 30, 2012

(version en ligne: <https://hbr.org/2012/11/data-humans-and-the-new-oil/>)

Nielsen, le leader des données et études de marché, a travaillé sur son identité visuelle pour inclure la visualisation de données, avec des visuels pilotés par les données, conçus par Jan Willem Tulp :



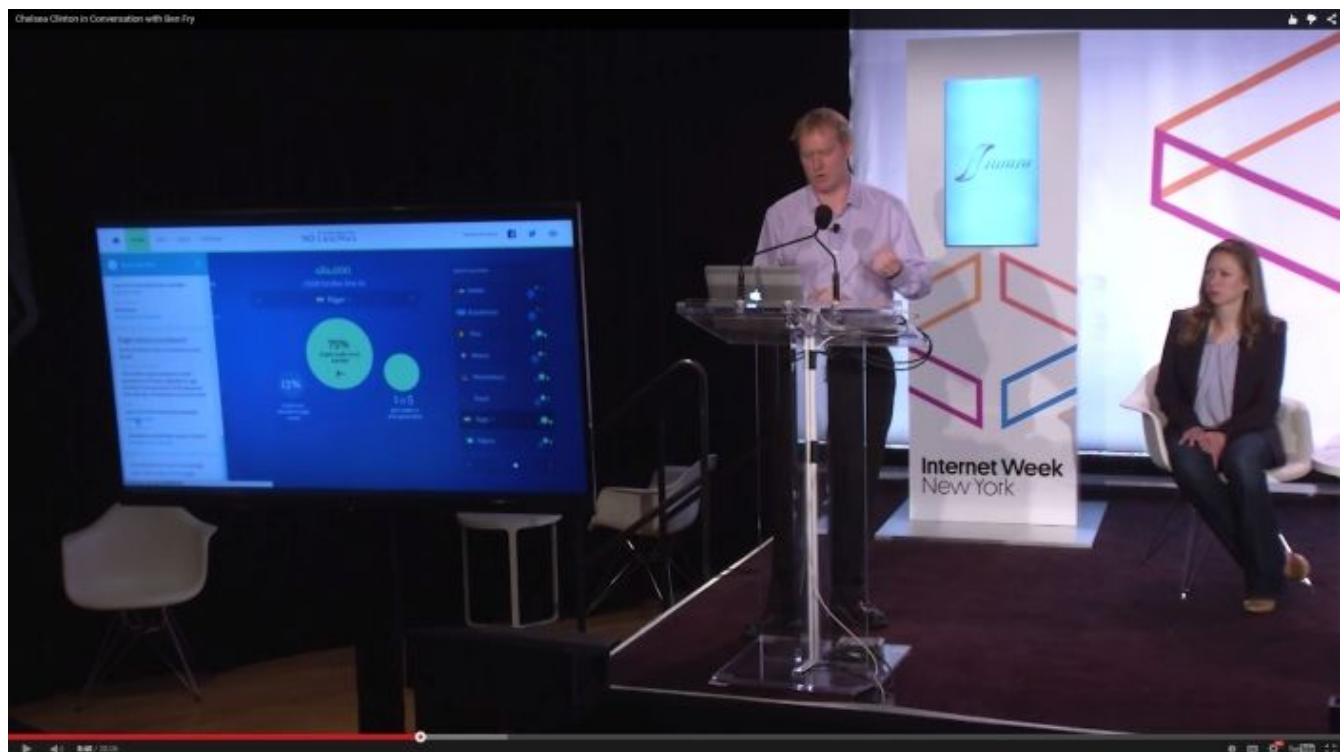
Depuis 2012 environ, General Electric s'associe à Fathom, l'agence créée par Ben Fry (co-créateur de Processing!) pour créer des visualisations en rapport à leur identité visuelle, avec quelques réalisations impressionnantes :



(URL en direct: <https://fathom.info/notebook/2124/>)

Et en 2015, vous savez que la dataviz est complètement stabilisée lorsque vous voyez un panel avec

Chelsea Clinton :



(version en ligne : <https://www.youtube.com/watch?v=YFrmQDCpgxs> - le panel est avec Ben Fry).

Donc, jusqu'en 2012 et 2013, je dirais que nous étions à l'âge d'or de la #dataviz en termes de découvertes et de nouvelles voies: commentaires enthousiastes sur les nouvelles productions du NYT, débats autour des objectifs de #dataviz: est-ce un moyen de raconter des histoires? Ouvrir de nouveaux mondes? Pour éduquer? Nouvelles connexions nouées avec de nouveaux arrivants, nouvelles agences, personnes se rencontrant pour la première fois lors de conférences après avoir échangé sur Twitter pendant des années, nouveaux postes, gros clients ... Et en 2015, les choses semblent s'être stabilisées et normalisées.

L'énergie a changé. La conversation sur Twitter a beaucoup ralenti. Le sentiment d'être des pionniers s'est érodé, car le temps a passé et parce que nous avons en effet essayé et exploré de nombreux fruits à portée de main. De nombreuses personnes sont désormais engagées dans des projets plus industriels à long terme.

Ce n'est donc pas une mauvaise nouvelle : la dataviz est maintenant bien établie et bien établie, les gens sont moins obligés de participer à des compétitions gratuites et de travailler sur de longs projets personnels les week-ends et les nuits pour se faire connaître, c'est bien. Mais l'excitation des années précédentes me manque un peu, quand vous aviez une techno ou un grand projet personnel publié par mois et quand vous aviez toutes ces grandes discussions sur Twitter à propos des développements à venir pour la dataviz.

5. À partir de 2015 : où va la dataviz?

Comme je l'ai dit, la première phase passionnante est passée et nous en sommes maintenant au stade où les processus de création de dataviz sont plus industrialisés, commercialisés et stabilisés.

Cela signifie que l'innovation trouvera d'autres endroits pour faire surface. Pourquoi? Parce que le paysage des technologies ne cesse de changer et que les esprits créatifs saisiront l'occasion de jouer et d'explorer ces opportunités là où aucun «client» ne les attend encore. Pour illustrer les pistes possibles, j'aime donner l'exemple de la carrière de [Seb Lee-Delisle](#), qui s'est défini comme un codeur créatif et maintenant comme un artiste numérique.

Je suis son travail sur Twitter depuis environ 2009. Il n'est pas au cœur du réseau "dataviz" et ne se définit pas par rapport à cette étiquette, mais vous le trouverez néanmoins sur la carte de la dataviz de Jeff Clark en 2012 (voir la carte ci-dessus).

- Il utilisait Adobe Flash comme l'une de ses principales technologies jusqu'en 2009, en contribuant à [PaperVision3D](#), un framework permettant de créer des animations et des jeux 3D dans Flash Player.
- Il joue un peu avec [Adobe Flex](#) en 2009,
- en 2010, Flash est définitivement derrière, il passe donc aux technologies HTML5, utilisant et enseignant [les visualisations interactives in HTML5 + Javascript](#)
- en 2012, il réalise un projet de "sentier lunaire": <http://seb.ly/work/lunar-trails/>
- en 2013, il continue avec les installations mais en plus grand: [pixelpyros](#) est un feu d'artifice interactif sur écran géant.
- En 2014/2015, il démocratise ce qu'il a appris et lance des ateliers sur "Des choses qui parlent aux internets": <http://seb.ly/st4i-stuff-that-talks-to-the-interwebs/>

Ce chemin et des chemins similaires empruntés par d'autres suggèrent que:

- L'écran d'ordinateur et même l'écran du téléphone mobile deviennent moins hégémoniques que le support où les données peuvent être visualisées. Objets, sculptures, bâtiments, meubles ... c'est la prochaine frontière à explorer. Non seulement mapper des données sur une surface plane, mais peut-être même en 3D: voir [cette installation](#) par Moritz Stefaner.
- Les interactions ont lieu dans des environnements plus riches que ceux auxquels nous sommes habitués (ordinateurs de bureau ou mobiles), les interactions avec l'utilisateur se diversifiant. Pas seulement la main et le clic de la souris, mais tout le corps. Pas une seule personne face à un objet, mais éventuellement une foule, éventuellement en mouvement, éventuellement en faisant des gestes.
- Et "les données" sont en train de prendre un sens encore plus grand. Lorsque vous vous éloignez de l'écran et commencez à vous connecter à une variété d'objets et de capteurs et à une variété de personnes, les données prennent encore d'autres formes: mesures en temps réel de l'environnement physique externe, de l'environnement interne (du corps), de l'environnement local, ou des interactions sociales lointaines au fur et à mesure de leur évolution, tout en restant connectés aux APIs avec lesquelles nous sommes déjà familiarisés ... le mélange peut donner des résultats percutants.

Ainsi, si la visualisation des données de l'API Twitter était le cliché de #dataviz en 2010 - 2015, le prochain cliché pourrait être l'impression 3D instantanée de données générées à partir d'objets et de corps connectés dans une maison ou un espace de travail. Ceci est juste ma vision pour la dataviz, je serais heureux d'en discuter avec vous maintenant!

Je vous remercie!

Publié en 2015

Des vidéos sur le sujet

- La data visualization : https://youtu.be/TlOW52_TPrE

Pour aller plus loin

Retrouvez le site complet [ici](#).



Ce cours est développé par Clement Levallois.

Découvrez mes autres cours : <https://www.clementlevallois.net>

Ou contactez-moi via Twitter : [@seinecle](#)