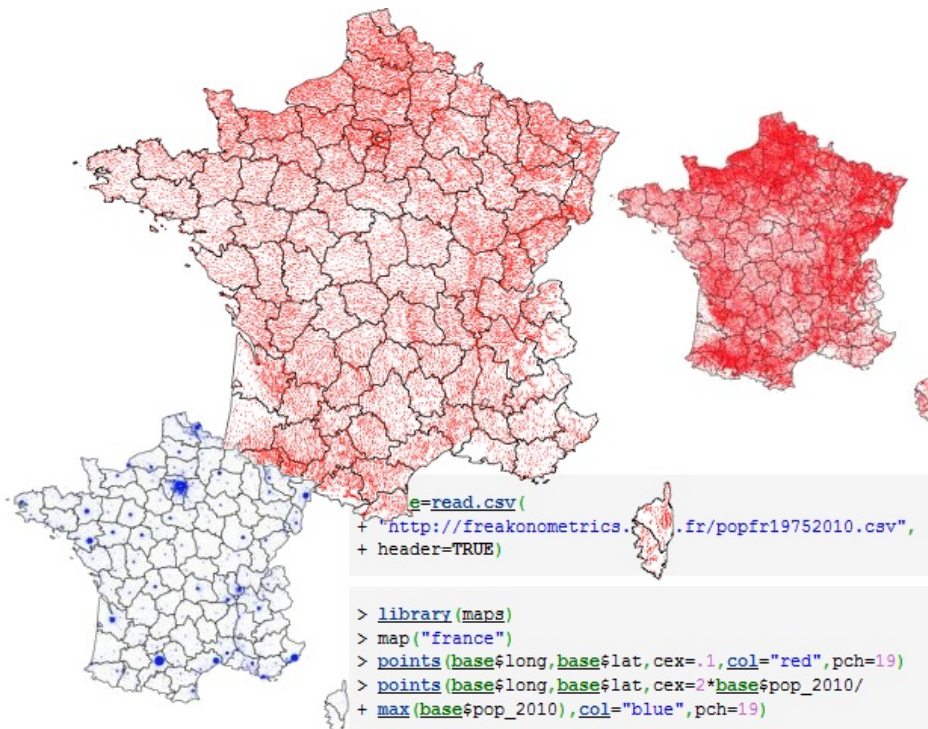


LES DATA EN FORME

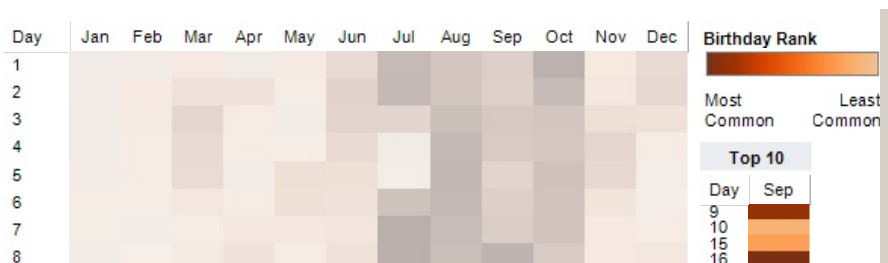
LE 29 MAI 2012 PAULE D'ATHA

Des tornades, de l'eau, du vent et la bonne dose habituelle de représentations chatoyantes de la densité humaine et de démographie colorée : c'est la nouvelle fournée de veille hebdomadaire par les journalistes de données de la rédaction d'*Owini*, à lire et à partager.

A priori, le premier lien de notre veille hebdomadaire ne mange pas de pain numérique, ne pulvérise aucune rétine, n'estomaque pas le lecteur en mal d'*awesomeness*. Mais il est rudement pratique. Grand merci donc à **Arthur Charpentier** – plus connu (mais sans doute pas assez) via son blog **Freakonometrics** – et à **Ewen Gallic**, un jeune Rennais étudiant à Montréal (c'est dire s'il patauge dans le bon Open Data) qu'il relaie cette semaine. Ce **précieux jeu de données** [en] ("*dataset*" dans la langue de **Simon Rogers**) est des plus utiles pour les data-journalistes que nous sommes, puisqu'il rassemble en un fichier CSV de 24 Mo les coordonnées GPS des 36 000 communes françaises (API Google Maps et **GeoHack**) et la population de ces communes entre 1975 et 2010 (Insee et travail d'estimation de population personnel par rapport aux données de l'Insee pour prendre en compte les fusions et/ou les scissions de communes durant ces 35 ans). Avis aux amateurs, donc !



Puisqu'on est lancés sur la démographie, relevons également cette semaine "l'augmentation" – six ans après – d'un travail de journalisme de (base de) données effectué par Amitabh Chandra de l'université de Harvard, et initialement publié par le *New York Times* en 2006 sous la forme un peu austère d'un tableau HTML : "**Votre anniversaire est-il populaire ?**" [en], basé sur le nombre de bébés étasuniens nés entre 1973 et 1999, où l'on apprend que le 16 septembre est la date la plus courante pour la naissance des petits américains. Ce boulot, un peu rustique, a été repris récemment par le data-journaliste de NPR **Matts Stiles** sous la forme d'**une visualisation un peu plus sexy** [en]. Visualisation statique toutefois, figée dans l'illustrator, elle-même rapidement reprise par **Andy Kriebel**, spécialiste de #dataviz, qui lui a apporté un peu d'interactivité grâce à la célèbre plateforme **Tableau**. Pour un résultat probant.

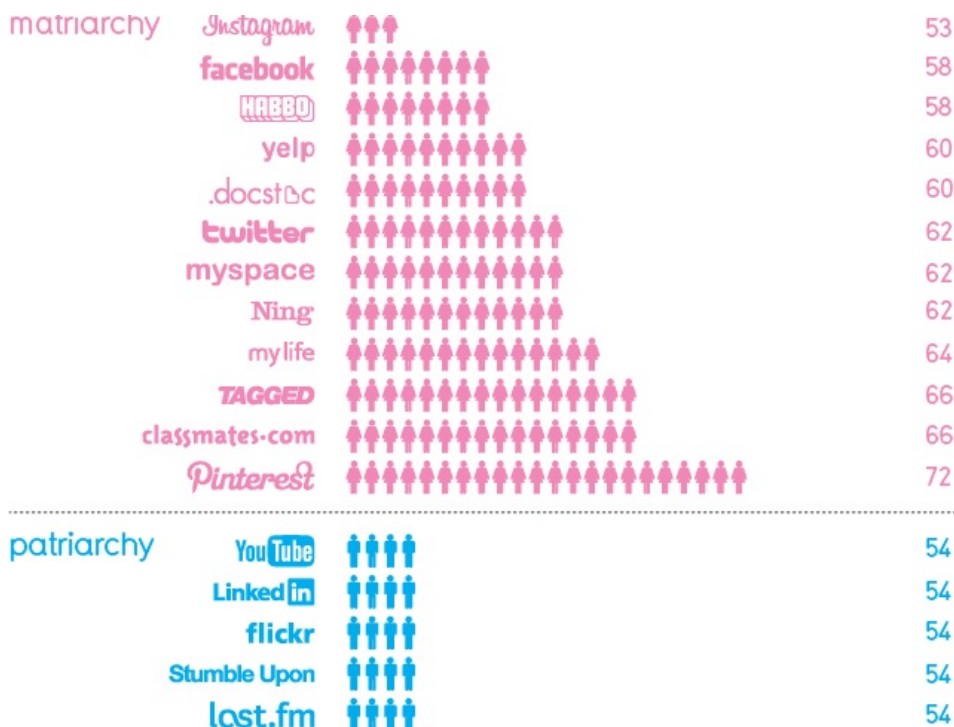




Mise-à-jour du 30 mai : Arthur Charpentier (décidément) nous indique fort justement que Freakonomics **a récemment pondu une version avec des données françaises** [en] de "Which birth dates are most common", qu'il en soit ici vivement adulé.

Mamma Mia

Autre œil mouillé jeté sur l'humanité, c'est celui du dieu vivant de la visualisation de données, **David McCandless**, qui nous pond (avec sa glorieuse équipe) en ce joli mois de mai (où les feuilles volent au vent si jolie mignonne) une infographie sobrement intitulée "**Les meufs règnent-elles ?**" – ou : l'équilibre des genres sur les réseaux sociaux. On y constate de prime abord, grâce au travail de recherche effectué par Dan Hampson et avec l'aide additionnelle de Piero Zagami et Tatjana Dubovina à la palette, que les filles sont plus présentes que les mecs sur Facebook, Twitter, Instagram ou Pinterest – pour ne citer que les plus connus – tandis que l'inverse s'applique sur Spotify, Google+, Flickr ou YouTube. Et qu'en tout, c'est presque 100 millions d'utilisatrices supplémentaires chaque mois qui se connectent sur les réseaux sociaux, et qu'elles sont davantage dans une logique d'échange que de consommation. Mais c'est une interprétation personnelle, chacun pourra jouer avec **la source** pour se faire la sienne.



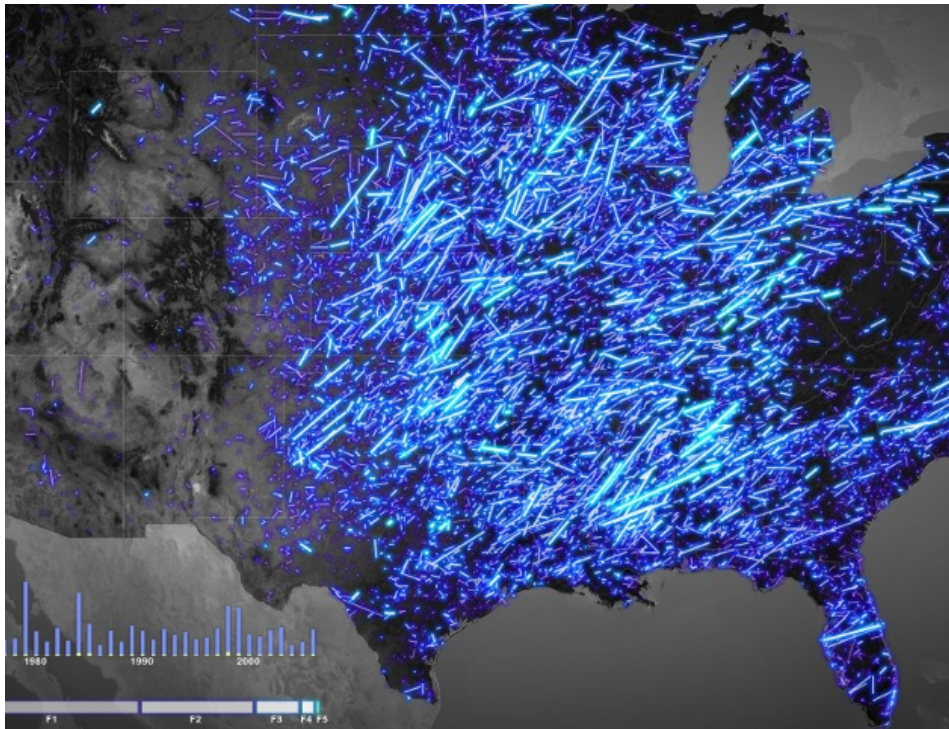
Petite entreprise ne connaît pas la crise

Le commerce en magasin ne faiblit pas en Italie malgré la morosité ambiante... grâce, notamment, aux étrangers hors-UE, pistés par l'exploitant de magasins "duty-free" Global Blue, qui a dressé la liste des pays les plus actifs chez nos voisins et en a pondu **une jolie infographie** [it] pour illustrer tout ça. À la vérité, ce ne sont ni le sujet ni les données qui auront attiré notre attention, mais bien la mise en forme, originale et bien léchée. À vérifier en mettant tout ça en plein écran.



Dans le vent

On reste dans la visualisation de données et dans les teintes de bleu(s) avec un vrai travail basé sur l'Open Data et notamment aux données récoltées de la National Oceanic and Atmospheric Administration (NOAA) **via le portail data.gov**. Ici, **John Nelson** – spécialiste en expérience utilisateur dans le civil – a catégorisé des données pistant les tornades selon leur positionnement sur **l'échelle de Fujita**, réglant ainsi l'intensité des traces les indiquant sur la carte en proportion de leur puissance au sol. Un histogramme indique également le nombre de victimes sous la forme d'une frise chronologique. Seul petit regret : le projet aurait été parfait s'il avait été un peu interactif. On se contentera donc de ce magnifique rendu.



Dans l'eau

Quand on est attentif aux questions posées par la gestion et l'utilisation de l'eau (comme le sont certainement la majorité des lecteurs, déjà sensibilisés sur *Owni* par le projet du **Prix de l'eau**), on étudiera sans doute avec plaisir cette nouvelle infographie repérée **via Flowingdata** : **"Quelles nations consomment le plus d'eau ?"** [en]. C'est à partir de données extraites par Arjen Hoekstra et Mesfin Mekonnen de l'université de Twente (Pays-Bas) que **Jen Christiansen**, directrice artistique et illustratrice au *Scientific American Magazine* a créé **cette visualisation colorée** qui met en exergue que si la population abondante est le premier facteur de variation de consommation de l'eau, une agriculture inefficace et une utilisation excessive de cette ressource naturelle pour la production de nourriture (30% de l'empreinte de l'eau aux États-Unis est issu de la consommation de viande) est également fatale.



Réflexions

Pour nos lecteurs anglophones, restent trois billets sur lesquels nous mettons le doigt – et que nous aurons sans doute bientôt l'occasion d'évoquer plus longuement :

Crowdsourced crisis mapping: how it works and why it matters, qui dresse un

panorama de la cartographie de crise et de l'utilisation du crowdsourcing, notamment du projet **Crowdmap**.

Data journalism research at Columbia aims to close data science skills gap, dans lequel l'ancienne éditrice-en-chef du *Guardian* **Emily Bell** analyse la progression du journalisme de données et la nécessaire évolution de la formation scientifique de ceux-ci. **Anyone can do it. Data journalism is the new punk**, où l'actuel patron du pôle Data dudit *Guardian* projette sa vision du métier de data-journaliste et le compare sous bien des aspects au mouvement punk. L'idée est amusante, le fond est bien plus sérieux qu'il n'y paraît.



Bien sûr, pour certains, ce ne sera jamais du journalisme. Et quelle importance ? Pendant qu'ils dissertent sur les définitions, nous autres on se mettra simplement au boulot.



Une bonne data-semaine à tou(te)s !

1 ping

Toutes les tempêtes visualisées depuis 1851 » revue du web, Just another weblog le 29 octobre 2012 - 15:18

[...] est l'oeuvre de John Nelson, dont Owni vous a déjà parlé dans les data en forme lorsque nous avons repéré son très bel ouvrage sur les tornades aux États-Unis. Le billet [...]