

Cartographies animées en sciences sociales. Apports heuristiques et épistémologiques

Éric Guichard, Triangle + IXXI

1^{er} février 2022

Présentation

- MCF HdR à l'Enssib, laboratoire Triangle et IXXI.
- Spécialiste du « numérique » et de la philosophie des techniques.
- Humble culture informatique des anciens temps : Unix/Linux, Perl, cgi-bin, L^AT_EX.
- Cartographie : une découverte ancienne (1992–2000), une pratique théorisée.
Retour récent : Écoles d'été de cartographie et de visualisation,
<http://barthes.enssib.fr/ECV-2017> (et 2018, 2019), travail sur les antennes de téléphonie mobile (Inria), puis engagement covid.
- Mise en perspective : *Études Digitales*, numéro 10, sortie février 2022.

En espérant que cette intervention puisse être utile, je remercie les organisateurs de cette journée de m'avoir invité ; notamment **David Parsons** et **Laurence Besson**.

Rappel du résumé

Je décrirai deux atlas :

- celui de la téléphonie mobile
(<http://barthes.enssib.fr/sociotel/antennes.html>)
- et l'atlas animé de la covid-19
(<http://barthes.enssib.fr/coronavirus/cartes/dev/monde/> pour vous,
<http://barthes.enssib.fr/coronavirus/cartes/Rmonde> pour le grand public).

Je détaillerai certaines procédures et conclurai sur les apports de la cartographie et de l'écriture numérique aux sciences sociales et à l'épistémologie.

Problématique et sources

Un projet dans le cadre de l'IPL Better-Net, fédéré par l'Inria.

Questions initiales : les « antennes » téléphoniques sont-elles installées au « doigt mouillé » par les opérateurs, une fois prises en compte les contraintes de l'Arcep ? Ou leur distribution spatiale obéit-elle à des logiques économiques, qui favorisent ou discriminent certains segments de la société ? Nous comptons étudier deux hypothèses :

- 1 Est-ce que les espaces populaires et les zones de pauvreté sont délaissés par les opérateurs, ou au contraire favorisés ?
- 2 Est-ce que ces derniers privilégient les jeunes (par exemple de 15 à 25 ans) au motif qu'ils seraient les plus gros consommateurs de bande passante ?

Par la suite et sauf précision contraire, les mots « antenne », « émetteur » et « support » seront utilisés comme synonymes, puisqu'ils renvoient tous à une même localisation.

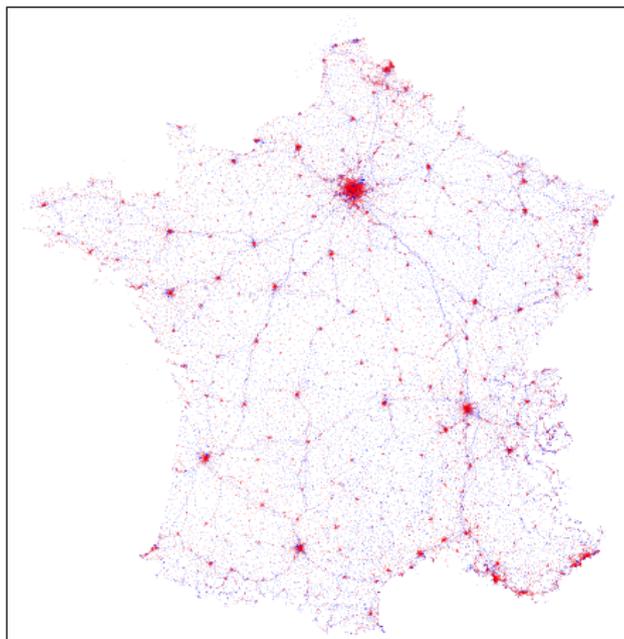
Projet

Articuler des données techniques et géographiques, relatives aux supports, antennes, émetteurs, avec d'autres, plus socio-démographiques.

Le social et le technique

Antennes en France. Commentaire à l'oral. Méthode : tikz + lua \LaTeX (merci à Jérémy Just). Pdf aussi accessible sur

<http://barthes.enssib.fr/Antoine/Dante/Antennes-carres.pdf>.



Antennes : construction des obtenues

- Circulations dans cinq fichiers « pesant » au total 243 Mo et décrivant selon plus de 40 variables (souvent absconses) les supports, stations (« virtuelles » ; ex. : parc militaire), bandes, antennes et émetteurs pour repérer les quelques informations qui nous intéressaient :
- les latitude, longitude, code postal, technologie, opérateur et date de mise en fonction d'un émetteur.
- Restaient à supprimer les émetteurs inutiles pour notre étude.
- Perl ou python.

De premiers projets de cartographie classique, vite abandonnés :



Figure – Présence de bouées-support en pleine mer du Nord. Copie d'écran d'une carte réalisée avec QGis.

Socio-démographie de la France

- L'Insee a découpé la France en carrés habités de 200 mètres de côté. Nous nous sommes limités à l'Hexagone et à la Corse, décrits par 2 278 216 tels carrés.
- Chaque carré précise le nombre exact de ses habitants et ses coordonnées. D'autres variables sont récupérables, via une autre base, qui garantit l'anonymat : 698 659 rectangles (unions des carrés précités) de plus de 11 ménages.
- Parmi les 25 variables proposées, nous avons retenu les nombres d'individus et de ménages, leur taille moyenne, le nombre de ménages « pauvres » (tels que définis par l'Insee : déclarant un revenu fiscal de moins de 7500 Euros) et les nombres de personnes de moins de 15 ans, de 15 à 25 ans, d'adultes et de plus de 65 ans. Ces données furent extrapolées aux carrés habités qu'ils contiennent, sachant que
 - dans les zones peuplées, un rectangle se confond souvent avec un carré
 - la « procédure de découpage en rectangles ne tient pas compte de la continuité du bâti » (sic Insee).

Note La base *open data* fournie par l'Insee est *étrange* au sens où les formats de données ne sont pas « libres », mais propriétaires. Un Go de données → 50 Mo.
Essentiellement du perl.

Carrés Insee : premier exemple

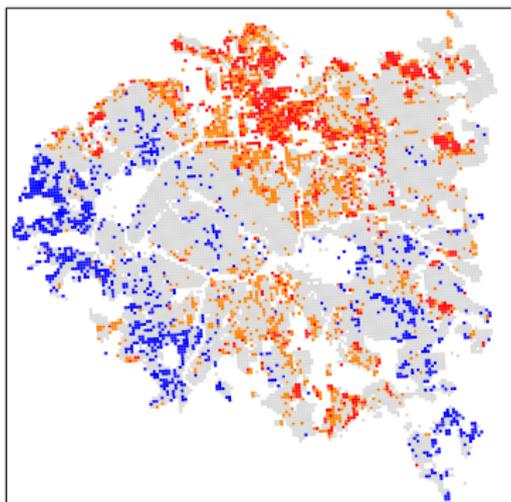


Figure – Dpts 75, 92 , 93, 94. En orange : carrés dont plus de 25% des ménages sont très pauvres, mais dont la moyenne fiscale est supérieure à 9000 Euros par ménage. En rouge : carrés pauvres avec beaucoup de ménages très pauvres. En bleu : carrés de moyenne fiscale supérieure ou égale à 27 000 Euros par ménage; on repère très peu d'îlots parisiens. . . En gris : autres carrés. 12 542 carrés Insee, dont 1742 en orange, 940 en rouge, 1130 en bleu, 8728 en gris. Carte produite avec Tikz et luaL^AT_EX via perl.

Travail d'appariement entre antennes et carrés Insee

- Lambert 2 étendu vs WGS-84. Aïe !
- → Bibliothèques python, 130 heures de calcul.
- Appartenance d'un point (54 000 antennes) à l'un des 895 767 carrés de plus de 10 habitants : autres bibliothèques python, autant de temps. . .
Travail à l'origine de la première carte des 54 000 antennes en rouge et bleu.
- Des « données » enfin appareillées. Des *données* pas données du tout.

Résultats

- Pauvres proches des antennes.

	Moyenne	Médiane
Ménages pauvres	532m	317m
Ménages non pauvres	735m	452m
Tous les ménages	698m	422m

Table – Moyennes et médianes des distances ménage-antenne. Merci à Anthony Busson et à Thomas Begin.

- En étudiant les antennes, on apprend beaucoup sur la sociologie de la France.
- Limites du big data : trois millions de Français manquent (zones quasiment vides).
- Des riches qui disparaissent. Cf. carte de la région parisienne.
- Au plan de la méthode : confort des scripts, de \LaTeX et tikz.
- L'atlas...

Avant l'atlas, retour sur la rareté de la richesse

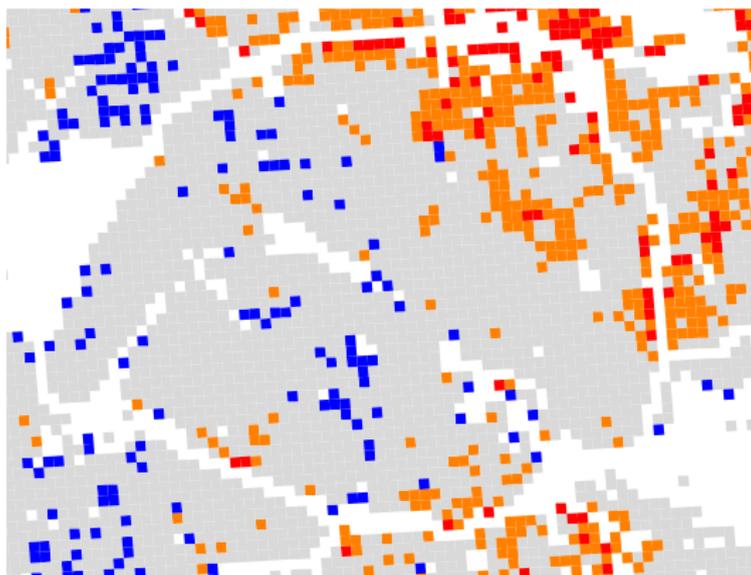


Figure – Carrés habités Insee à Paris. On repère facilement la Seine, le boulevard périphérique, les bois de Vincennes et de Boulogne (en blanc).

En bleu : (trop rares ?) carrés de moyenne fiscale supérieure ou égale à 27 000 Euros par ménage. En orange et rouge : carrés dont plus de 25% des ménages ont de très bas revenus. Carte : EG.

L'atlas

- Merci à Antoine Chemardin.
- Interactif : variables, départements, type antennes, etc.
- Zoom, fonds, mémoire URL, infos (clic ou passage souris).
- Méthode : Mapbox avec OSM et tout le travail antécédent.
- Adresse pour commentaire oral :
<http://barthes.enssib.fr/sociotel/antennes.html>
- Exemple...

Près de la Doua

Copie d'écran de <http://barthes.enssib.fr/sociotel/antennes.html#45.77025107540601,4.867455508873036,14.485020386505902,69XXX>, Pauvres



Des journaux, automatiques et instructifs

Face au flou relatif à la pandémie en février 2020, s'informer au mieux et tenter de produire des visualisations qui facilitent l'analyse.

Scripts perl (environ 1000 lignes) + une crontab : production automatisés de fichiers L^AT_EX eux aussi automatiquement compilés et mis en ligne à 7 heures du matin.

Des journaux et graphiques plus lisibles en 2020 qu'aujourd'hui. Néanmoins : <http://barthes.enssib.fr/coronavirus/Graphes-Covid19-du-jour.pdf> (pour les morts) et <http://barthes.enssib.fr/coronavirus/Graphes-Covid19-confirmes-du-jour.pdf>, qui reste ces temps-ci particulièrement explicite.

Bref, une intelligence du quotidien, explicitement *artificielle*.

Méthode et bilan : crontab + wget + perl de l'ancien temps + tikz presque oublié = résultats honorables (sourires).

Résultats

L'Espagne oubliée.

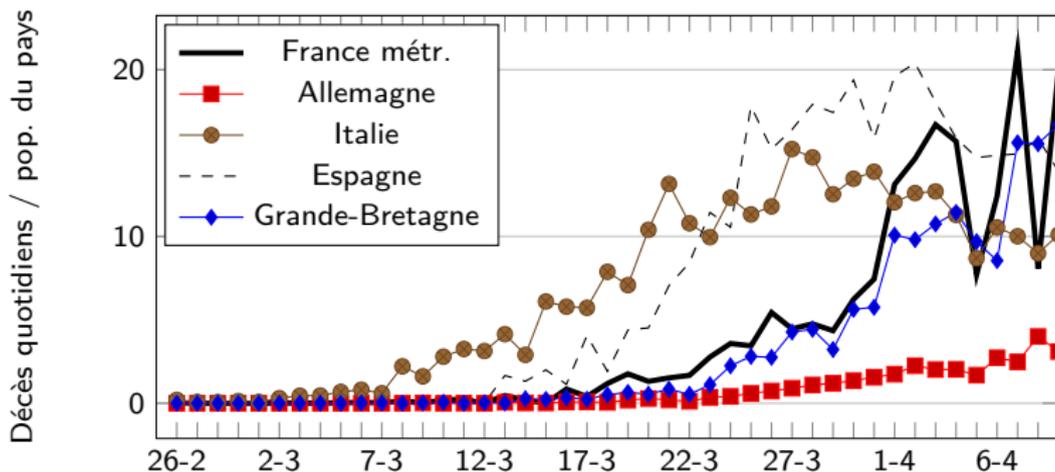


Figure – Proportion du nombre de morts par jour pour un million d'habitants dans les pays considérés. L'accroissement soudain des statistiques françaises le 2 avril (+900 décès Ehpad) met la France au niveau moyen de l'Espagne, qui dépassait déjà l'Italie vers le 25 mars. Ces 3 pays et la Grande-Bretagne seront au même niveau le 11 avril 2020.

Les morts négatifs

Ils sont 31 en Allemagne le 11 avril 2020, 2 puis 217 en France les 16 et 19 mai, 31 en Italie le 2 juin et... 1211 le 1^{er} mai en Espagne; une telle situation s'est aussi rencontrée aux Pays-Bas, au Danemark, en Tchéquie, au Japon, etc.

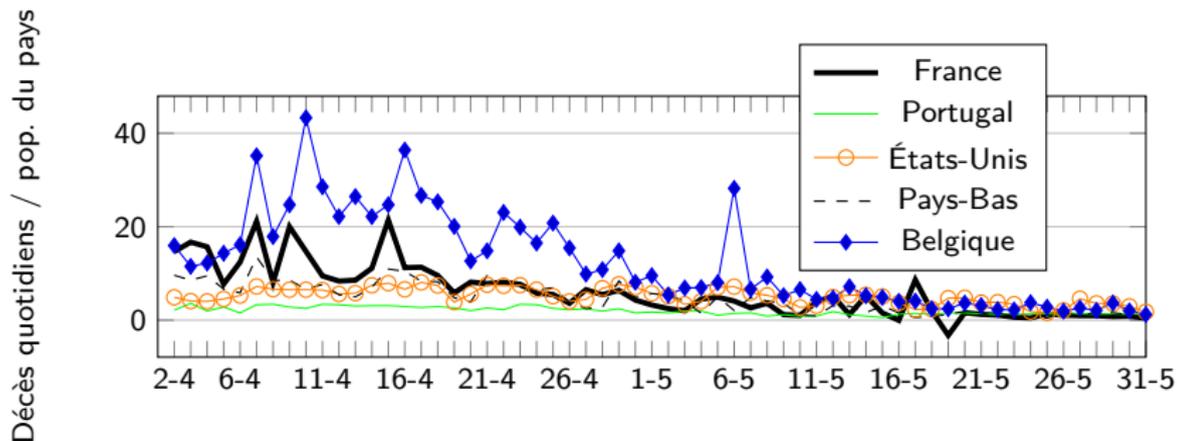


Figure – Proportion du nombre de morts par jour pour un million d'habitants dans les pays considérés. Sélection des graphiques pour du 2 avril 2020 au 31 mai. On repère les dents de scie des courbes et les morts négatifs français du 19 mai.

Les errances administratives

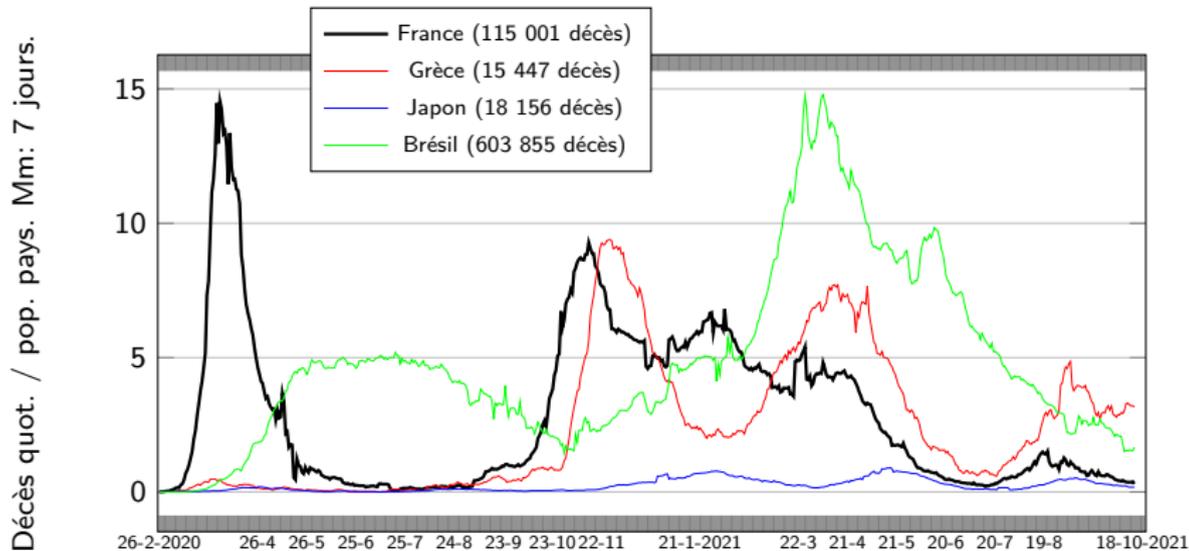


Figure – Moyenne mobile sur 7 jours de la proportion de décès quotidiens par million d'habitants. Les moyennes mobiles ont l'avantage de masquer les morts négatifs, mais le caractère erratique des courbes lissées signale *a fortiori* celui des données initiales.

Un modèle

- Rectifier de mauvaises données.
- Construire un modèle conforme à la théorie, qui intègre l'évolution du $R(t)$ à partir des confirmés 6 jours auparavant.
- Produire une estimation robuste du taux de reproduction pour un territoire (pays, département).
- Aller vers la prévision.
- Article :
<https://hal.inria.fr/hal-02921836/file/journal.pone.0237901.pdf>
- Un travail collectif.

Des premières cartes animées

- Morts : <http://barthes.enssib.fr/coronavirus/cartes/Morts-Monde-carte-animee-du-jour.svg>.
- Confirmés : <http://barthes.enssib.fr/coronavirus/cartes/Confirmes-Monde-carte-animee-du-jour.svg>.
- R : <http://barthes.enssib.fr/coronavirus/cartes/R-Monde-carte-animee-du-jour.svg>.
- Uniquement Perl + svg + crontab
- Un travail complexe de nettoyage et rectification des sources (îles, pays, Russie, Antarctique, etc.).
- Problèmes : arrêt sur image impossible, svg non optimisé, taille croissante des fichiers (6 Mo en juin 2020, 35 aujourd'hui).
- Design : une gradation de couleurs complexe (tentative continuité), mais des informations instructives (triangles de morts, etc.).

Versions 2 et 3 des cartes avec js

- Même fond, un travail sur les couleurs des caissons, une interactivité affinée.
- Retour aux compétences d'A. Chemardin.
- Un soutien du Cnrs.
- <http://barthes.enssib.fr/coronavirus/cartes/Rmonde/>
- Développement en cours :
<http://barthes.enssib.fr/coronavirus/cartes/dev/monde/monde.html>

Apports heuristiques de la cartographie, épistémologie

- Apports heuristiques : cf. cartes de Paris. Disparition artificielle des classes sociales. Mais la notion de classe sociale est elle-même *artificielle*.
Donc orientation vers ce qu'est l'écriture du monde, sans jugement de valeur.
 - La carte explicite le cheminement de notre pensée : circulation entre virtuel et actuel, pour comprendre un réel (somme des deux) toujours confus.
La carte nous rappelle aussi qu'il n'y a pas qu'un seul monde (ou qu'une seule vérité). Elle est un entre-deux stimulant car interrogatif entre les charlatans déistes et les charlatans donnéistes.
 - Cartes et graphiques objectivent notre incapacité d'humains à objectiver le monde, sauf par petites touches, par approximations successives.
- [extraits de « Ouvertures », E. Guichard, *Études Digitales*, n° 10, Garnier.]

Conclusion

Pour plus d'infos, cf.

- divers articles sur <http://barthes.enssib.fr/articles/#Guichard>,
- Le numéro 10 d'*Études Digitales* à paraître.
- Les introductions aux écoles d'été de cartographie et de visualisation :
<http://barthes.enssib.fr/ECV-2019>, <http://barthes.enssib.fr/ECV-2018>,
<http://barthes.enssib.fr/ECV-2017>.

Merci.

Eric.Guichard@enssib.fr

Eric.Guichard@ens-lyon.fr

<http://barthes.enssib.fr>