

« TimeLineEDB », application web d'exploration interactive de données de géolocalisation

par : [Robin Cura](#)

Mots-clés : concours géovisualisation et cartographies dynamiques, données issues de smartphones, données spatio-temporelles, FIG2016, géovisualisation

Site officiel de l'application : <https://analytics.huma-num.fr/Robin.Cura/TimeLineEDB/>

Contexte

D'après le dernier *Baromètre du Numérique* (Arcep, 2016), près de deux Français sur trois possèdent un smartphone, et parmi eux, une forte part l'utilise fréquemment pour « chercher un restaurant, un bar, un musée ou un magasin à partir d'une application [où ils sont géolocalisés] ». Les principaux systèmes d'exploitation mobiles (Apple iOS et Google Android), après accord de l'utilisateur¹, collectent de manière systématique un historique des localisations enregistrées sur le smartphone. Cela produit à terme, pour chaque utilisateur de ces systèmes, des données spatio-temporelles massives. Elles peuvent être consultées sur les smartphones dans tous les cas, mais seul Android permet leur visualisation externe ainsi que leur export, via le service [Google Timeline](#).

Dans la lignée des pratiques de « *quantified self* », nous avons créé un outil d'exploration de ces données, TimeLine Exploratory DashBoard ([TimeLineEDB](#)), afin de permettre à chacun de mesurer l'ampleur de ces informations collectées, et, au moyen d'une interface d'analyse intuitive, de rendre compte de la forte capacité intrusive que ces données peuvent conférer. Sur le plan de la géovisualisation, nous souhaitons ici montrer que face à une masse de données, les méthodes de représentation et d'interrogation les plus simples peuvent suffire à extraire de précieuses informations et synthèses des caractéristiques et comportements de mobilité d'un individu.

Données d'entrée : historique de localisation de Google Timeline

Les données sont issues des capteurs des smartphones Android. Selon le mode de localisation choisi par l'utilisateur, elles peuvent être collectées depuis le récepteur GPS inclus (forte précision, environ 5-10 m), depuis les réseaux wi-fi alentour (précision de 20-50 m en zone dense), ou encore *via* la triangulation des antennes relais proches (faible précision, environ 200-500 m en zone dense). Ces localisations sont collectées automatiquement de deux manières différentes. En cas d'inactivité du smartphone, elles sont enregistrées toutes les 5 minutes environ, et seront envoyées sur les serveurs de Google dès qu'un transfert de données sera possible. En cas d'activité (utilisation du smartphone, réception d'un appel, d'un SMS, navigation sur internet, etc.), cet enregistrement est bien plus fréquent, de l'ordre de la minute. En situation d'utilisation active de la géolocalisation (navigation GPS par exemple), cette fréquence peut descendre à 3-5 secondes.

Pour un utilisateur donné, cela constitue donc rapidement de nombreux « points » (environ 100 000 par an en faible utilisation du smartphone). La qualité sémantique de ces données est particulièrement faible, chacun de ces points n'étant composé que de la localisation, du temps associé et d'une indication peu fiable de la qualité de géolocalisation.

Exploration interactive des données avec TimeLineEDB

Face à cette masse de données peu compatible avec les méthodologies d'analyse spatio-temporelles les plus fréquentes (la *time-geography* et les algorithmes de *clustering* spatio-temporels sont peu performants en présence de données massives), nous considérons qu'une bonne compréhension peut s'acquérir par une exploration visuelle, au moyen de filtrages dynamiques et interactifs tels que pratiqués dans le champ des « *geovisual analytics* » (Keim *et al.*, 2008 ; Andrienko *et al.*, 2010). Les SIG, de par leur formalisme de requête (souvent SQL), se prêtent peu à de l'exploration interactive, et nous avons donc créé un outil ad hoc d'interrogation et de géovisualisation de ces données de localisation, sous forme d'application web. Cela la rend accessible depuis un simple navigateur internet récent, l'ouvrant dès lors à un large public.

Cette application, TimeLineEDB (**figure 1**), s'appuie sur des technologies libres² organisées autour du langage d'analyse de données R (R Core Team, 2016). Elle fait en particulier un large usage du *package* « shiny » (Chang *et al.*, 2017), qui permet la création d'interfaces graphiques dynamiques web.



Figure 1. Copie écran de l'interface TimeLineEDB.

TimeLineEDB est organisée autour de deux catégories d'outils :

- les outils d'affichage et d'interrogation temporelle (**figure 1, A à C**) : ces graphiques affichent la fréquence temporelle des points de la base de données. Le premier montre la fréquence journalière (plages horaires), le second la fréquence hebdomadaire (jours de la semaine) et le dernier la fréquence annuelle (mois de l'année). Une sélection graphique peut être réalisée suivant sur ces critères temporels et la carte est alors mise à jour en mettant en évidence les localisations correspondant au pas de temps sélectionné. Cette démarche permet d'explorer les lieux pratiqués par l'utilisateur selon différentes temporalités. Ces filtres temporels s'appuient sur la cyclicité de la dimension temporelle et permettent d'identifier les rythmes propres de l'utilisateur ;
- la carte interactive (**figure 1, D**) : cette carte dynamique affiche les points sous forme de carte de chaleur. Elle est mise à jour lors des sélections temporelles. Elle permet aussi d'effectuer des sélections de localisations sur la carte, et cette sélection entraîne une mise à jour des graphiques, montrant les usages temporels des lieux. L'utilisation de *heatmaps* cherche à révéler la logique spatiale des déplacements de l'utilisateur et une régularité dans la fréquentation des lieux. Elle

permet de décrire les espaces de vie de l'individu.

Ces outils permettent, par sélections croisées et reportées, d'explorer le jeu de données exemple qui s'affiche à l'ouverture de l'application, mais aussi d'explorer les données de l'utilisateur, si tant est que Google en possède sur lui. Pour cela, toute l'application est documentée par un système de tutoriel intégré, qui permet de comprendre les informations présentées, la manière d'interagir avec celles-ci, et guide l'utilisateur dans les phases de récupération, d'export et d'intégration de ses propres données depuis le service [Google Takeout](#).

Exemple d'utilisation

Nous présentons ici un exemple d'exploration interactive, sur le jeu de données d'un utilisateur quelconque. L'utilisation de l'outil étant intrinsèquement dynamique, nous avons choisi d'en réaliser une vidéo (**figure 2**). Notons bien qu'il ne s'agit que d'une exploration à valeur illustrative, chaque utilisateur étant invité à procéder par sérendipité pour extraire des informations de l'exploration de ses données.

<http://mappemonde.mgm.fr/wp-content/uploads/temp/TimeLineEDB.mp4>

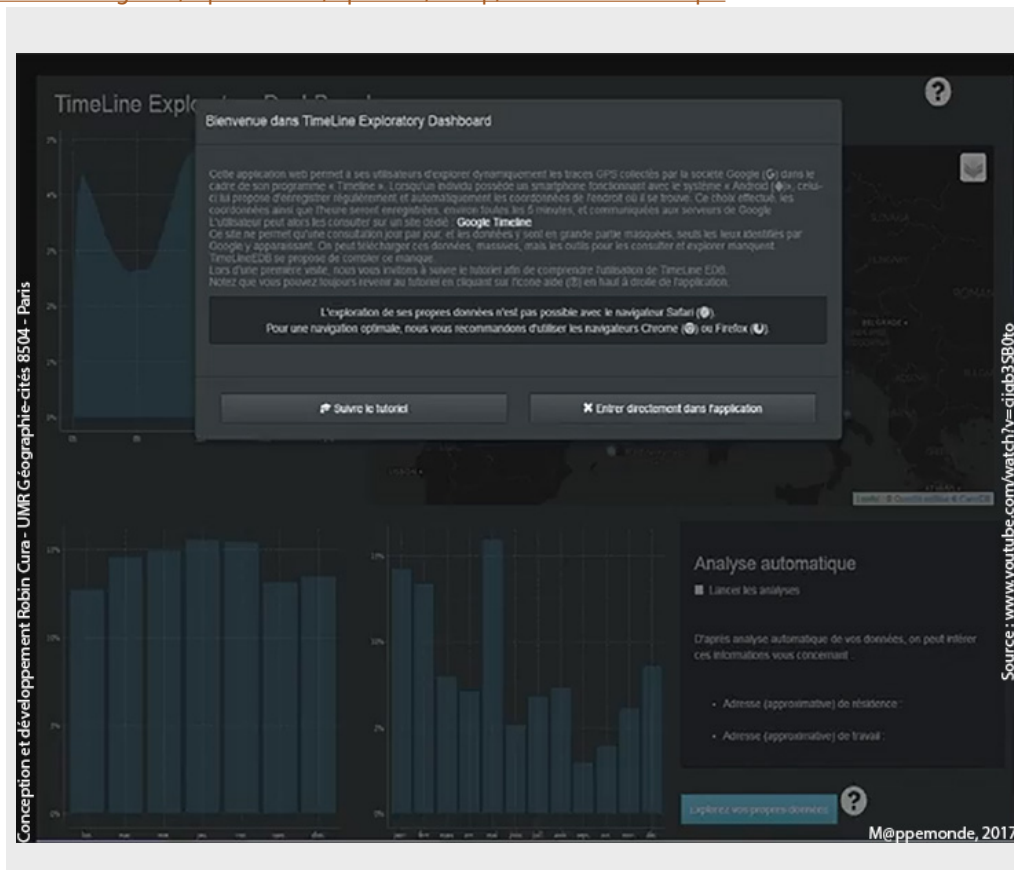


Figure 2. Tutoriel de l'utilisation de TimeLineEDB (vidéo).

Perspectives

TimeLineEDB s'appuie sur des principes bien connus et éprouvés en géovisualisation et en géomatique, et ne prétend à ce titre pas à une quelconque innovation dans ces domaines. Ce qui nous paraît innovant tient en

deux caractéristiques principales.

Usages géographiques de la géovisualisation et de l'exploration interactive

Les méthodes d'analyse exploratoire interactives, fortement utilisées dans les années 1990, sont peu à peu tombées en désuétude, au profit de l'utilisation de SIG, d'analyses statistiques avancées et d'algorithmes de plus en plus complexes. Nous souhaitons ici montrer que, face à des jeux de données de plus en plus larges, la simple visualisation et utilisation de filtres interactifs — spatiaux comme temporels — peuvent suffire à trouver les faits saillants permettant de passer des données à une compréhension de l'espace. Ces données sont largement adaptées à de l'extraction de connaissances, ce qui nous encourage à soutenir la pertinence de l'exploration interactive sur des jeux de données massifs.

On cherche ainsi à remettre au cœur des pratiques géographiques quantitatives la géovisualisation et l'exploration interactive, souvent utilisées sur des données purement statistiques, mais plus rarement sur des données spatiales et/ou temporelles.

Usages grand public de l'information géographique et des *big-data*

L'application vise à sensibiliser le public à la masse de données privées qu'il communique, parfois inconsciemment, à de grands opérateurs privés. L'application permet ainsi à chacun d'appréhender la quantité d'informations personnelles qu'on peut y récupérer avec des traitements simples. Le volume de ces données permet d'identifier très aisément et précisément les lieux et rythmes de vie d'un individu. Nous souhaitons ici illustrer et faire expérimenter la démarche d'exploration interactive de données spatio-temporelles par une approche ludique, sous forme « d'investigation » à partir des propres données de chacun, ou du jeu de données d'exemple utilisé.

Dans une plus large mesure et pour un public plus divers, TimeLineEDB s'inscrit dans la lignée des outils de valorisation scientifique, et souhaite montrer avec pédagogie la richesse et la complexité des données spatio-temporelles et des méthodes d'analyse qu'elles requièrent, tout en illustrant l'un des apports de la géomatique à la géographie et aux sciences humaines et sociales.

Bibliographie

ANDRIENKO G., ANDRIENKO N., DEMSAR U. *et al.* (2010). "Space, time and visual analytics". *International Journal of Geographical Information Science*, vol. 24, n° 10, p. 1577-1600.

ARCEP (2016). *Baromètre du Numérique - Édition 2016*. CREDOC.

CHANG W., CHENG J., ALLAIRE J. J., XIE Y., MCPHERSON J. (2017). *shiny : Web Application Framework for R*. En ligne : <https://CRAN.R-project.org/package=shiny>

KEIM D., ANDRIENKO G., FEKETE J.-D., GÖRG C., KOHLHAMMER J., MELANON G. (2008). « Visual analytics : Definition, process, and challenges ». In KERREN A., STASKO J.T., FEKETE J.D., NORTH C. (ÉD) *Information visualization*, Berlin/Heidelberg : Springer, coll. "Lecture Notes in Computer Science" , vol. 4 950, p. 154-175.

R CORE TEAM. (2016). *R : A Language and Environment for Statistical Computing*. Vienne (Autriche) : R Foundation for Statistical Computing. En ligne : <https://www.R-project.org/>

Notes

- Dans un smartphone utilisant iOS, cet accord est implicite dès l'utilisation de la géolocalisation. Sur
1. ↑ Android, cet usage requiert une activation manuelle de l'utilisateur, qui lui est proposée, parmi de nombreuses autres options, à la première utilisation de la géolocalisation.
 2. ↑ TimeLineEDB est libre, sous licence A-GPL, et ses sources sont consultables sur un dépôt GitHub dédié : <https://github.com/RCura/TimeLineEDB>