

FACILITER L'ANALYSE DE DONNÉES SPATIO-TEMPORELLES MASSIVES

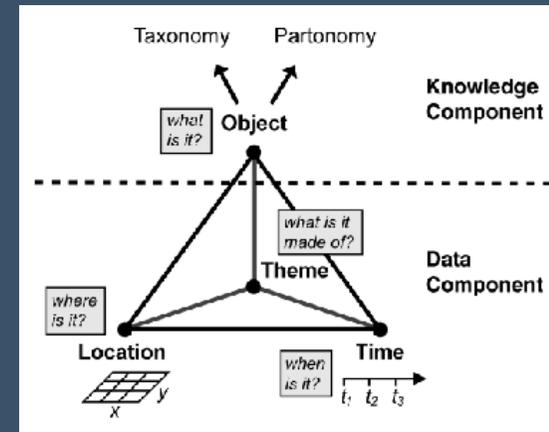
Plaidoyer illustré pour le recours
à l'exploration interactive.

Robin CURA
robin.cura@parisgeo.cnrs.fr

Doctorant en Géographie
Université Paris 1
UMR Géographie-cités

Contexte et Objectifs

- Définition des termes
 - Données spatio-temporelles :
 - « *The multiple dimensions of the Data Component can be divided into three categories: location (position in the three spatial dimensions), time (position along a time line), and theme (a set of observations, measurements, or attribute values associated with a particular location and time).* » (Peuquet, 2002)
 - Données ici issues de capteurs
 - Données massives
 - Exploration interactive



Contexte et Objectifs

- Définition des termes
 - Données spatio-temporelles
 - Données massives :
 - « *Big data is a term for data sets that are so large or complex that traditional data processing application software is inadequate to deal with them. Challenges include capture, storage, analysis, data curation, search, sharing, transfer, visualization, querying, updating and information privacy.* » (Wikipedia)
 - Volume / Vélocité / Variété
 - Exploration interactive

Contexte et Objectifs

- Définition des termes
 - Données spatio-temporelles
 - Données massives
 - Exploration interactive :
 - *Exploratory Data Analysis (Tukey, 1977)*
 - *Graphical Data Analysis (Unwin, 2015)*
 - *Visual Analytics (Keim et al., 2008)*
 - *(geo)Visual Analytics (ibid; MacEachren, 2017)*

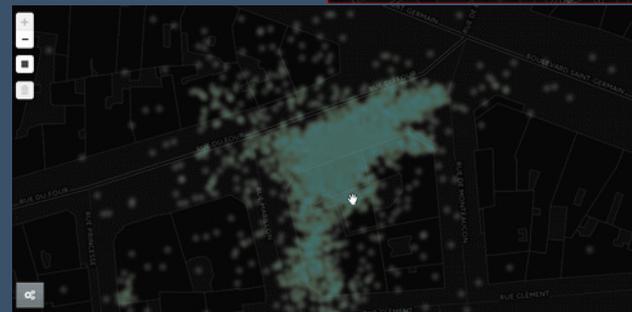
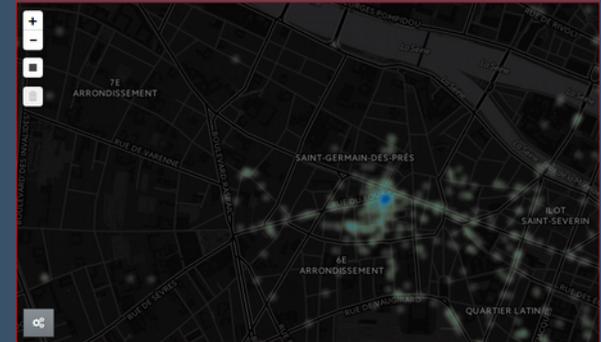
« This paper presents an argument for a “human-in-the-loop” (geo)Visual Analytics approach to leveraging big (geo)data as a means to understand and enable experience of place as a dynamic construct. [...] The approach contrasts with those that rely exclusively on computational methods to produce information and generate answers. (geo)Visual Analytics (gVA) is presented as both a science and set of methods/tools that are focused specifically on support of human analytical reasoning with big, heterogeneous, dynamically changing, and often ‘messy’ data that include geographical components. »

Contexte et Objectifs

- Quel usage de l'exploration interactive aujourd'hui ?
 - Prédominance du SIG
 - Prédominance d'algorithmes de traitements avancés
 - Les « moulinettes » ont quasi disparu
 - « *And, an argument is presented that gVA applied to these rich and dynamic data offers new windows to understanding place in ways that are not provided through traditional Geographical Information System (GISystem) methods applied to structured geospatial data.* » (MacEachren, 2017)

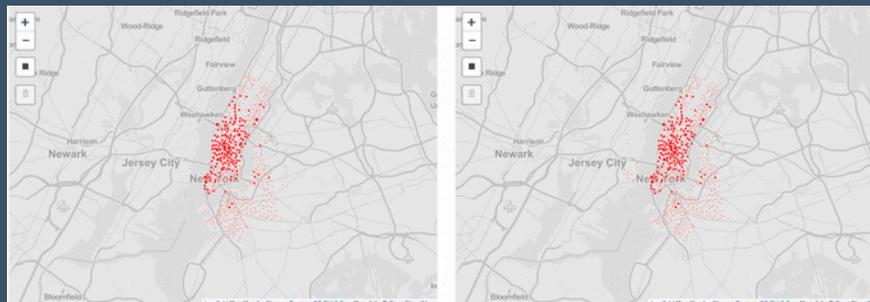
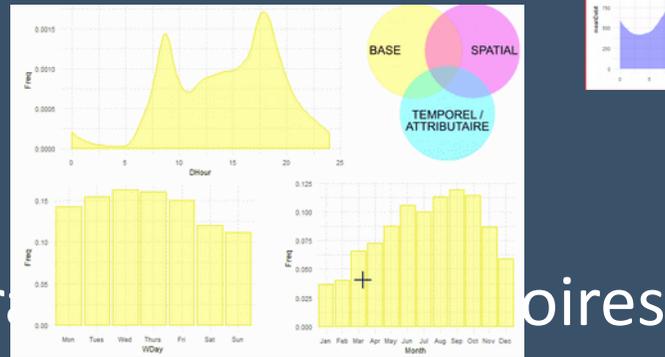
Méthodes

- Représentation
 - Tout montrer
 - Jouer avec les échelles
 - Privilégier la fluidité



Méthodes

- Interactions
 - Small-multiples
 - Linked-views
 - Brushing
 - Vers des interactions



Jeux de données

Jeux de données

Google TimeLine : Localisation via *smartphone*

Vos trajets 🔒 AUJOURD'HUI

ANNÉE ▼ MOIS ▼ JOUR ▼

181 lieux

Consultez les lieux que vous fréquentez le plus souvent et tous ceux dans lesquels vous êtes rendu grâce à l'historique des positions.

AUTRES TRAJETS

Paris
29 mars - 1 mai 2017

L'historique des positions est activé

Votre position est mise à jour par votre appareil mobile, et vous seul pouvez y accéder.

SUSPENDRE L'HISTORIQUE DES POSITIONS

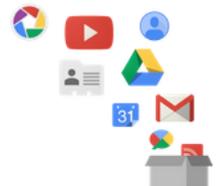
Domicile et lieu de travail

Télécharger vos données

Votre compte, vos données
Exporter une copie.

Créez une archive contenant les données relatives aux produits Google que vous utilisez.

Gérer les archives



Sélectionner les données à télécharger

Sélectionnez les produits Google à inclure dans l'archive et configurez les paramètres de chaque produit. Vous serez le seul à avoir accès à cette archive. [En savoir plus](#)

Produit Détails

Historique des positions Format JSON

Historique des positions **JSON** ▼

Suivant

Afficher plus de produits

Personnaliser le format de l'archive

Jeux de données

Google TimeLine : Localisation via

smartphone

Thème \emptyset

- Etendue : \approx France
- Type : Espace continu
- Résolution : \approx 5-20 m

- Etendue : \approx 1 an
- Type : Temps continu
- Fréquence : \approx 3-5 min

1 utilisateur

Espace

Temps



\approx 175 000 localisations

Jeux de données

Trafic: Comptages routiers à partir de capteurs permanents

Référentiel géographique pour les données trafic issues des capteurs permanents

Informations Tableau Carte Analyse Export API



Partager Intégrer Widget

Données trafic issues des capteurs permanents

Informations Tableau Analyse Export API

Ce jeu de données comprend l'ensemble des données horaires de débit et taux d'occupation remontées par les capteurs permanents de trafic installés sur le réseau renseigné de la Ville de Paris (réseau urbain et boulevard périphérique).

La **période couverte** va du 1er janvier 2013 au dernier mois complet.

Un **descriptif** plus détaillé des données est disponible dans la rubrique "Pièces jointes" de la présente page d'information. **Penser à bien lire cette fiche explicative des données avant de procéder à tout traitement.**

Le **référentiel géographique** est accessible au sein du jeu de données "Référentiel géographique pour les données trafic issues des capteurs permanents", disponible à l'URL : <https://opendata.paris.fr/explore/dataset/referent...> On soulignera que la jointure entre le référentiel géographique et les données de trafic doit se faire par l'attribut `id_arc_tra`, et uniquement celui-ci.

Identifiant du jeu de données `comptages-routiers-permanents`

Téléchargements 1 773

Thème Mobilité et Espace Public
Mot clé circulation, trafic, comptage routier, capteur trafic
Licence [Open Database License \(ODbL\)](#)
Langue Français
Modifié 22 juin 2015 11:19
Producteur Mairie de Paris/Direction de la Voirie et des Déplacements

Dernier traitement 13 mai 2017 21:44 (métadonnées)
13 mai 2017 21:44 (données)

DCAT

Créateur Direction de la Voirie et des déplacements - Service des Déplacements
Contributeur Direction de la Voirie et des déplacements - Service des Déplacements
Fréquence de mise à jour mensuelle

Pièces jointes

Cliquez pour déplier

Modèle de données

Cliquez pour déplier

Réutilisations

Cliquez pour relier

Soumettre une réutilisation

Cliquez pour déplier

Jeux de données

Trafic: Comptages routiers à partir de capteurs permanents

- Thème • Nombre de voitures / tronçon
• Taux d'occupation du tronçon

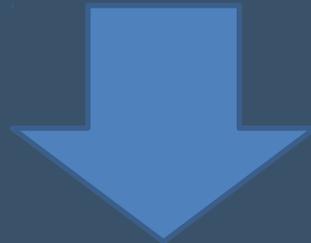
- Etendue : Paris
- Type : Polyignes
- Résolution : ≈ 3000 tronçons de rue

- Etendue : 1 an
- Type : Agrégé
- Fréquence : 1/h

Espace

n rues

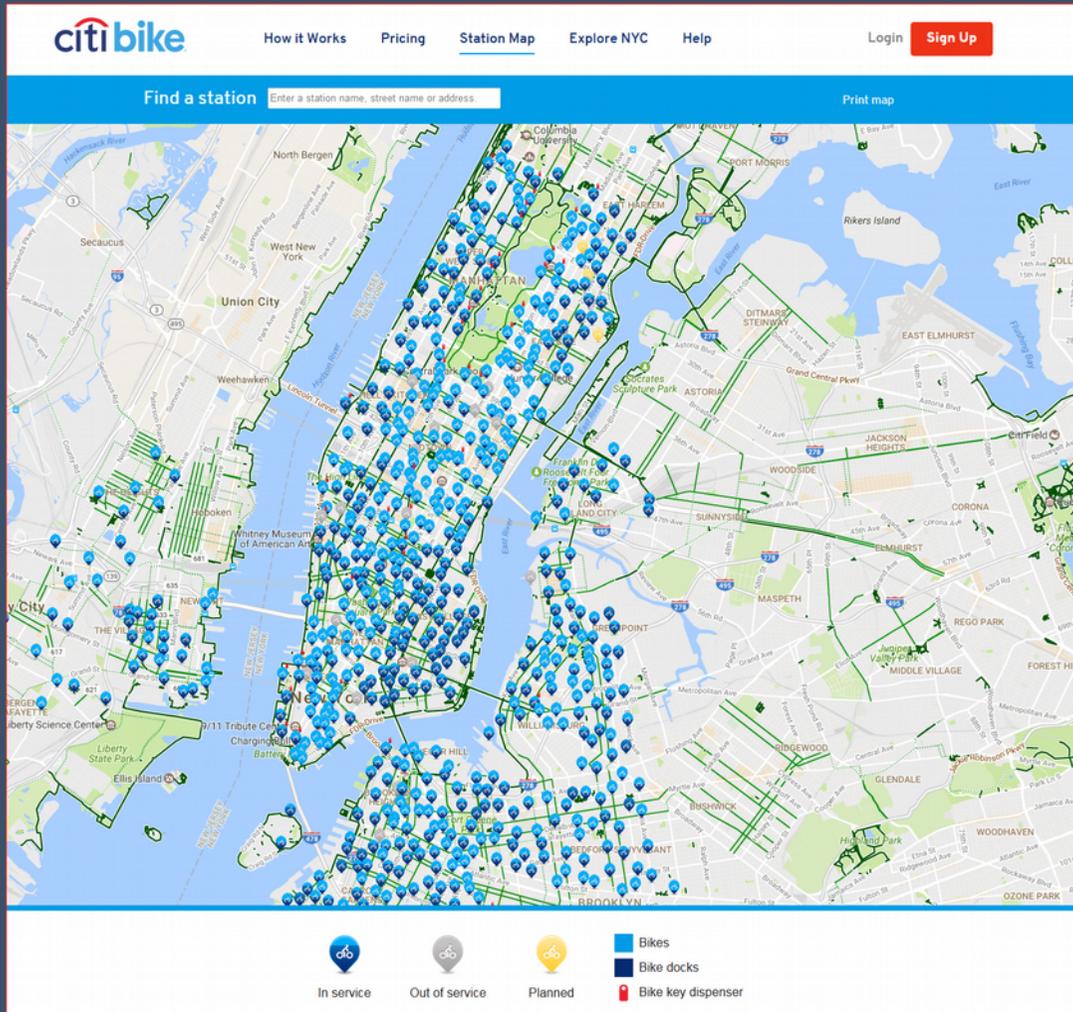
Temps



$\approx 25\ 000\ 000$ comptages

Jeux de données

CitiBike : Trajets en vélos libres services



The screenshot shows the CitiBike website's 'System Data' page. At the top, there is a navigation bar with the CitiBike logo and links for 'How it Works', 'Pricing', 'Station Map', 'Explore NYC', and 'Help'. The main content area is titled 'System Data' and contains the following text:
System Data
Where do Citi Bikers ride? When do they ride? How far do they go? Which stations are most popular? What days of the week are most rides taken on? We've heard all of these questions and more from you, and we're happy to provide the data to help you discover the answers to these questions and more. We invite developers, engineers, statisticians, artists, academics and other interested members of the public to use the data we provide for analysis, development, visualization and whatever else moves you.
This data is provided according to the [NYCBS Data Use Policy](#).
Citi Bike Trip Histories
We publish [downloadable files of Citi Bike trip data](#). The data includes:

- Trip Duration (seconds)
- Start Time and Date
- Stop Time and Date
- Start Station Name
- End Station Name
- Station ID
- Station Lat/Long
- Bike ID
- User Type (Customer = 24-hour pass or 7-day pass user; Subscriber = Annual Member)
- Gender (Zero=unknown; 1=male; 2=female)
- Year of Birth

This data has been processed to remove trips that are taken by staff as they service and inspect the system, trips that are taken to/from any of our "test" stations (which we were using more in June and July 2013), and any trips that were below 60 seconds in length (potentially false starts or users trying to re-dock a bike to ensure it's secure).

[→ Download Citi Bike trip history data](#)

Citi Bike Daily Ridership and Membership Data
Explore [Citi Bike daily ridership and membership data](#) from January 2017 onwards. Click on the link at the bottom right of the graph to download the data as a csv file.
Here are data sets for previous quarters:

- 2013 Data: [Launch - Sep; Oct - Dec](#)
- 2014 Data: [Jan - Mar; Apr - Jun; Jul - Sep; Oct - Dec](#)
- 2015 Data: [Jan - Mar; Apr - Jun; Jul - Sep; Oct - Dec](#)
- 2016 Data: [Jan - Mar; Apr - Jun; Jul - Sep; Oct - Dec](#)

Jeux de données

CitiBike : Trajets en vélos libres services

- Destination
- Durée

Thème

- Genre de l'utilisateur
- Âge de l'utilisateur

- Etendue : \approx New-York
- Type : Stations CitiBike
- Résolution : \approx 650 stations

- Etendue : 1 an
- Type : Temps \approx continu
- Fréquence : \approx 25 / min

n trajets

Espace

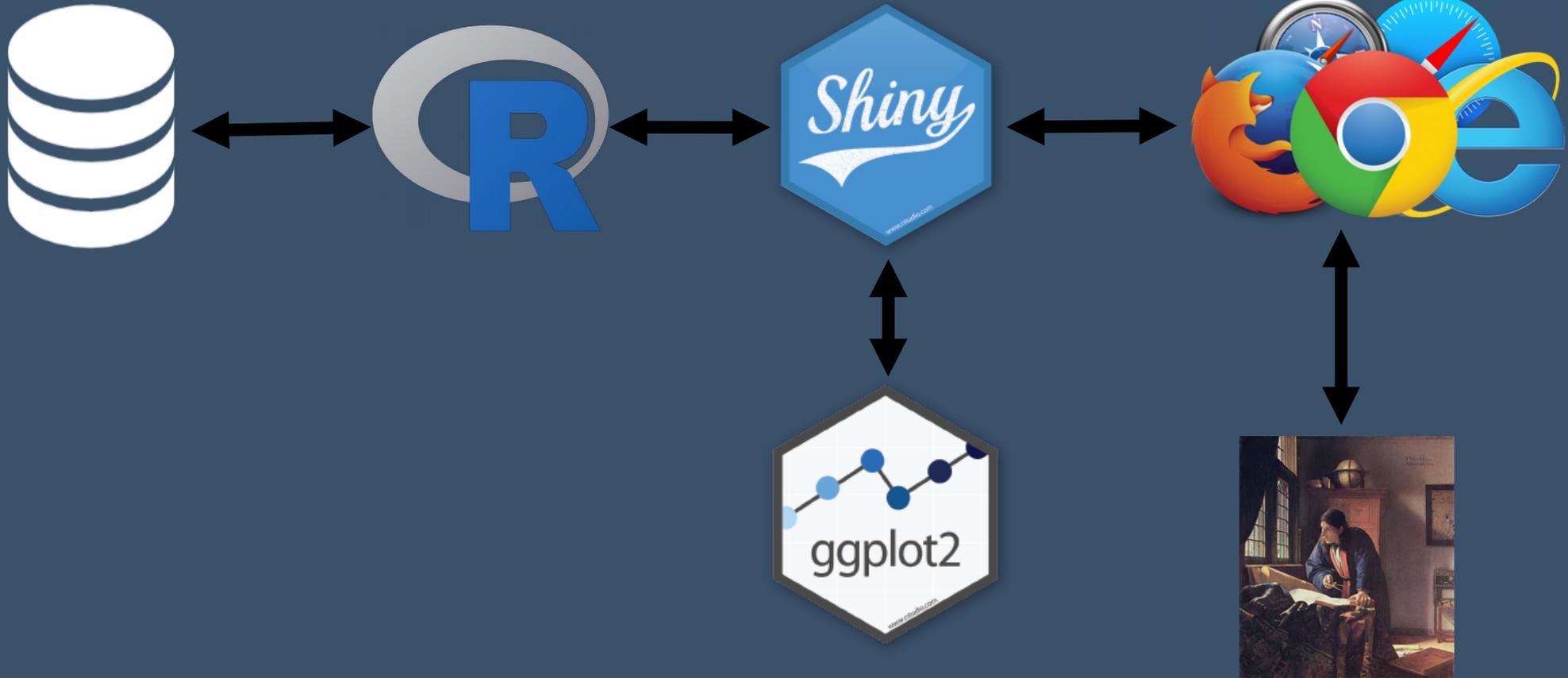
Temps



\approx 13 500 000 trajets

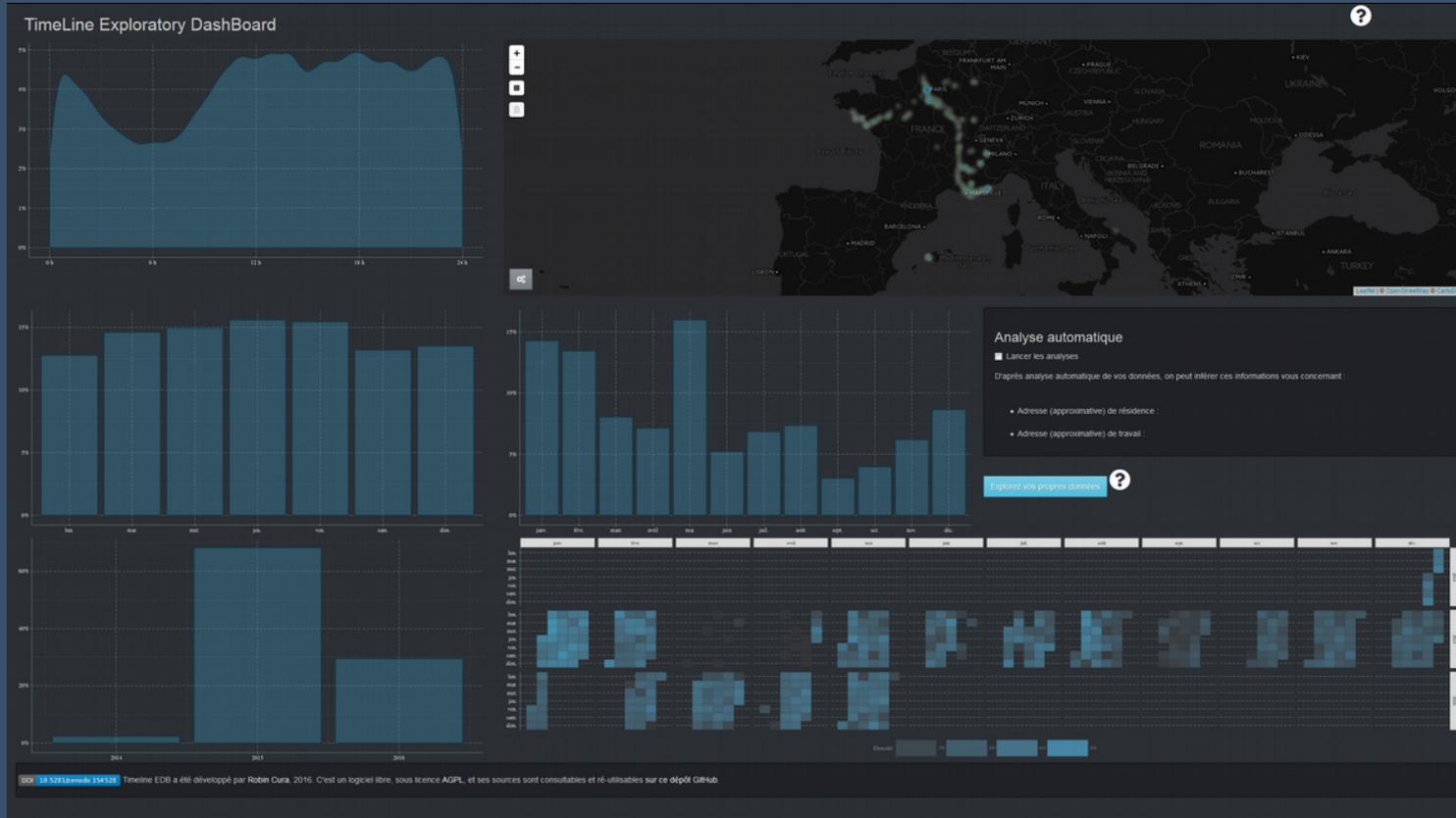
Applications

Leaflet 



Applications

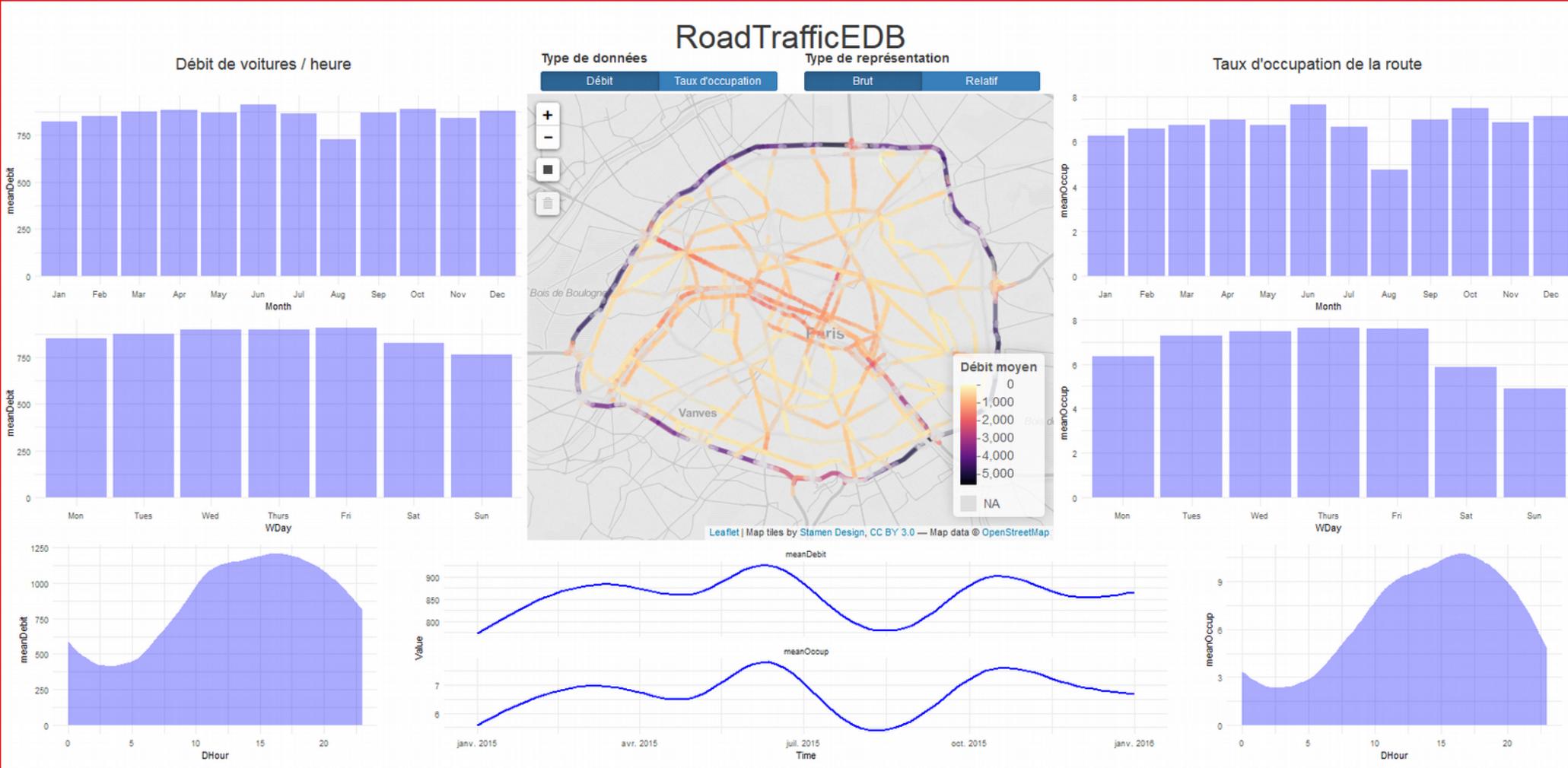
TimeLineEDB



Cura R. (2017), « "TimeLineEDB", application web d'exploration interactive de données de géolocalisation », M@ppemonde, vol. 120, 2015/4.

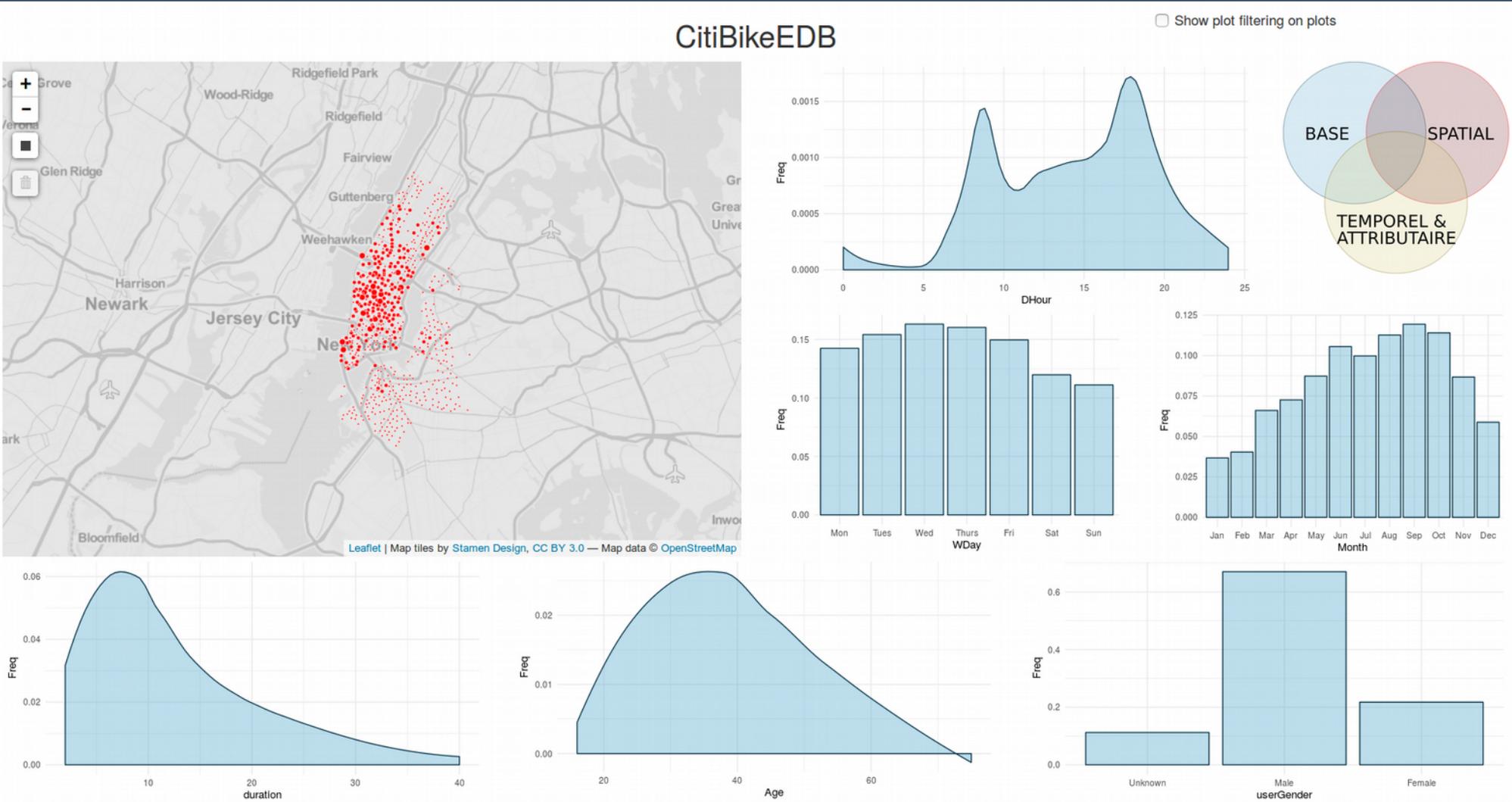
Applications

RoadTrafficEDB



Applications

CitiBikeEDB



Quand faire appel à l'exploration interactive ?

- Pour comprendre le jeu de données
 - Découverte par l'exploration
- Pour valider le jeu de données
 - Visualiser pour trouver les anomalies
- Pour tester une hypothèse
 - Explorer les résidus / spécificités
- Pour découvrir de nouvelles hypothèses
 - Exploration interactive comme catalyseur de sérendipité
- Pour valoriser une étude
 - Ouverture au grand public par approche ludique

MERCI

Robin CURA
robin.cura@parisgeo.cnrs.fr

Doctorant en Géographie
Université Paris 1
UMR Géographie-cités