



HAL
open science

L'ouverture des données, une opportunité pour la recherche sur les transports et la mobilité

Patrice Aknin, Etienne Come, Latifa Oukhellou

► To cite this version:

Patrice Aknin, Etienne Come, Latifa Oukhellou. L'ouverture des données, une opportunité pour la recherche sur les transports et la mobilité. *Transports urbains : mobilité, réseaux, territoires*, 2018, 132, pp.21-27. 10.3917/turb.132.0021 . hal-01897482

HAL Id: hal-01897482

<https://hal.science/hal-01897482>

Submitted on 6 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L'ouverture des données, une opportunité pour la recherche sur les transports et la mobilité

Auteurs : Patrice Aknin¹, Etienne Côme², Latifa Oukhellou²

¹ IRT SystemX, Paris-Saclay, 8 avenue de la Vauve, 91120 Palaiseau

² IFSTTAR, COSYS/GRETTIA, 14-20 Bd de Newton, 77427 Champs-sur-Marne

1. Les données pour et par la mobilité

Les systèmes de mobilité sont des déterminants des sociétés modernes, tout particulièrement en contexte urbain qui impose ses contraintes de massification, mais également en zone moins dense où la diversité des solutions prévaut. La manière dont ces systèmes sont utilisés par les usagers (dans le temps et dans l'espace), la complémentarité ou la concurrence des modes, leurs qualités intrinsèques et leur fiabilité, sont des sujets fortement investigués en raison des enjeux sociétaux qu'ils sous tendent, comme le droit à la mobilité, l'inclusion sociale, la transition énergétique et la maîtrise des impacts environnementaux liés au transport. Les pouvoirs publics jouent évidemment un rôle important, pour impulser des pratiques de mobilités durables visant en particulier à diminuer l'usage des voitures individuelles et à développer l'usage des transports en commun (TC) seuls ou en combinaison avec d'autres modes de mobilité tels que la marche, le vélo, l'autopartage, le covoiturage... La multimodalité, rendue possible par une offre élargie de systèmes de transport disponibles pour un même déplacement, est une piste importante d'évolution des politiques de mobilité durable. Une seconde piste, non moins importante, est la pratique de l'intermodalité où différents modes de transport sont utilisés successivement au cours d'un même déplacement.

Du point de vue de l'utilisateur, les choix modaux sont, outre les déterminants économiques (tarification, prix du carburant, prix du stationnement), pour partie dépendants des choix réalisés en termes d'urbanisme : répartition spatiale des densités de population et d'emploi à l'échelle métropolitaine, aménagement de la voirie et offre de transport à l'échelle locale.

L'analyse et l'observation des mobilités s'appuie traditionnellement sur les enquêtes de mobilité (Enquête Ménages Déplacements (EMD), enquêtes de type origine-destination, enquête cordon...) Zmud *et al.* 2013, Cottrill *et al.* 2013, TRB 2012. De telles enquêtes présentent plusieurs avantages : elles couvrent tous les modes de transport et tous les motifs de déplacement sont renseignés, elles contiennent des méta informations sur les enquêtés (genre, catégorie socio-professionnelle ...). En revanche, elles sont coûteuses et de ce fait ne sont réalisées qu'avec des fréquences assez faibles (typiquement 5 ou 10 ans entre deux enquêtes publiques), ce qui ne permet pas un suivi très proche des évolutions en cours et des politiques publiques visant à les infléchir.

La compréhension du fonctionnement des systèmes de mobilité s'est donc historiquement développée principalement par le biais de ces enquêtes déclaratives et ce n'est qu'assez récemment que des alternatives crédibles basées sur des approches dites « data » ont pu voir le jour. Le déploiement rapide des systèmes de géo-localisation et de communication (téléphones cellulaires, localisation satellitaire, internet mobile, billettique, véhicule sonde ...), l'introduction de l'internet des objets et le stockage facilité des données, sont autant de facteurs technologiques qui ont conduit à l'accroissement exponentiel des données collectées. Ces données massives viennent ainsi compléter les données préexistantes liées à la planification et l'exploitation des transports et de la mobilité.

Dans le domaine des transports, l'ouverture et la réutilisation des données s'inscrit dans un contexte politique et social qui promeut aujourd'hui les données au rang de « bien commun ». Ainsi l'ambitieux rapport Jutand de 2015 (Jutand 2015), résultat du travail du comité du débat national sur l'ouverture des données publiques dans le domaine des transports, a été jusqu'à recommander au gouvernement de considérer des données non-publiques comme des données « d'intérêt général », c'est-à-dire des données dont l'exploitation croisée permet la création de services à forte valeur ajoutée pour la collectivité. La création de la mission Etalab en 2011 a prolongé cette dynamique. Cette mission joue un rôle important dans la coordination et le partage des données via une plateforme ouverte¹ qui héberge les données publiques et recense leurs principales réutilisations.

Il convient toutefois de bien préciser que les données Transport sont de natures diverses et que leur ouverture est parfois difficile, même pour les données relatives aux transports publics en raison de freins institutionnels ou commerciaux. Pour rendre les choses plus concrètes, détaillons les types de données disponibles et leur statut pour deux cas de systèmes de transports détaillés dans les sections suivantes, à savoir les transports collectifs de type bus et métros et les systèmes de vélos en libre service.

- Données sur l'offre théorique de transport souvent fournies au format GTFS (General Transit Feed Specification). Il s'agit d'un modèle de données défini par Google pour

¹ <https://www.data.gouv.fr/fr/>

les flux de transport. Ces données comportent la topologie du réseau et les tables horaires théoriques (les horaires planifiés). Elles sont ouvertes et mises à disposition pour réutilisation. Elles sont pour l'instant au coeur du mouvement d'ouverture comme le montre la plateforme dédiée au transport².

- Pour les transports collectifs, les données de fréquentation par station ou ligne de transport agrégées (au quart d'heure par exemple) à partir des données de validation individuelle sont des données parfois ouvertes, à l'instar de ce qui est fait par Ile-de-France Mobilités (ex STIF).
- Pour les systèmes de vélos en libre service, les données ouvertes portent (en plus des informations sur les stations comme la taille, la localisation...) sur le nombre de vélos disponibles à chaque borne.
- Les données individuelles et anonymisées relatives aux trajets eux-mêmes s'avèrent très riches d'enseignement mais leur accès n'est possible aujourd'hui que pour des besoins de recherche et nécessite la signature de convention de mise à disposition des données entre les partenaires qui stipule notamment les objectifs visés par la recherche ainsi qu'une déclaration à la CNIL.

Beaucoup d'autres données peuvent être recueillies comme

- les requêtes d'itinéraires qui peuvent renseigner sur la demande de transport,
- les remontées d'incidents pour l'aide à l'exploitation et à la maintenance,
- les informations aux voyageurs ou dualement les informations produites par les voyageurs et qu'ils poussent dans les réseaux sociaux
- ...

Enfin, l'usage des données peut déborder largement le cadre pour lequel l'utilisateur les avait produites ou fait produire. Prenons par exemple le cas des données de comptage d'énergie à bord des trains. Il s'agit sur le réseau ferré français d'une mesure de la puissance consommée ou reinjectée à la caténaire, géolocalisée, disponible toutes les 5 minutes 24h sur 24, pour tous les trains. Voici un cas exemplaire d'une donnée générée pour les besoins de facturation objective d'un exploitant ferroviaire qui entend régler sa facture énergétique au gestionnaire d'infrastructure au kwh près. Mais que peuvent nous dire ces données ? Voici une liste non exhaustive de service que cette donnée peut rendre :

- Qualifier le type de conduite du point de vue énergétique sur un trajet donné, privilégier les modes d'éco-conduite des trains
- Connaitre la position des trains et en déduire les retards au plan de transport
- Localiser les trains sur les voies annexes, le temps utile en circulation, le temps de stationnement, et le temps passé en centre de maintenance

² <https://transport.data.gouv.fr/>).

- Estimer la charge embarquée dans les trains (fret ou voyageur) en comparant la puissance absorbée sur des trajets similaire
- Faciliter la mise au point de plan de transport durable en synchronisant les phases d'accélération de certains trains pendant que d'autres freinent sur des zones proches rentabilisant ainsi la réinjection de courant à la caténaire
- Détecter automatiquement certaines défaillances des infrastructures électriques
- ...

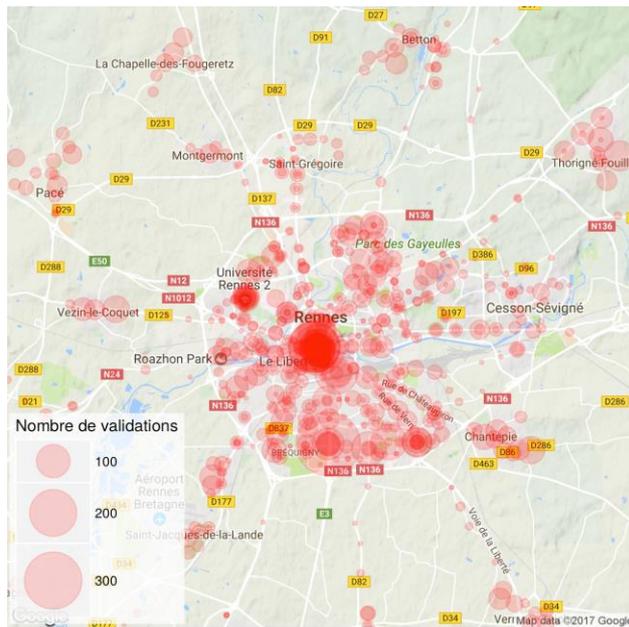
L'exploitation de ces données nécessite la mise en œuvre de méthodes avancées d'analyse et requiert bien souvent le croisement de plusieurs sources de données hétérogènes. Nous allons considérer dans ce qui suit deux cas d'usage, celui du système de vélos en libre service de Paris et celui des transports collectifs de la ville de Rennes. L'objectif est de montrer le potentiel de ces données pour apporter un nouvel éclairage sur l'usage des systèmes de mobilité partagée et une meilleure connaissance des usagers dans le cas des transports collectifs.

2. Analyse des comportements de mobilité des usagers des Transports en Commun grâce aux données billettiques

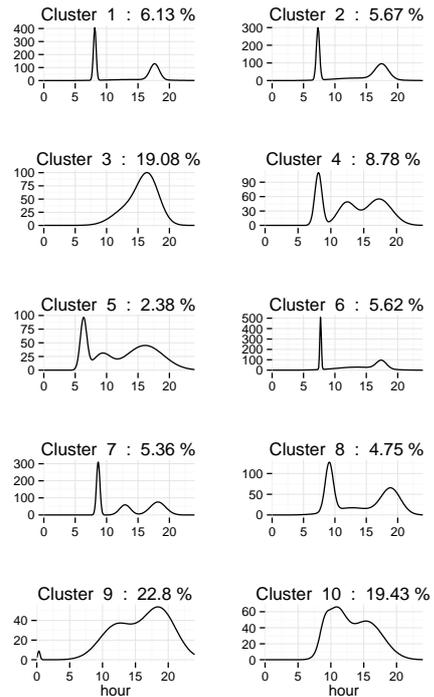
Dans le cadre d'un projet collaboratif financé par le programme de recherche et d'innovation dans les transports terrestres (Projet Mobilletic, 2013-2016), un partenariat avec Keolis Rennes a été établi pour la mise à disposition de données individuelles anonymisées de trajets bus et métro du réseau STAR. Une question centrale pour ce projet est liée à la valorisation de ces données massives (environ 250 000 validations/jour) en vue de mieux connaître les usagers des transports publics et leurs habitudes temporelles. L'objectif est de dépasser la vision « jour moyen » de la semaine qu'offre les enquêtes, même si les données billettiques n'adressent qu'environ 13% de la mobilité quotidienne dans le cas rennais.

IDENTIFICATION AUTOMATIQUE DES ROUTINES TEMPORELLES DES USAGERS DE TRANSPORT EN COMMUN PAR FOUILLE DE DONNÉES

Pour l'analyse des pratiques de mobilité des individus sur un réseau de transport, l'application de méthodes d'analyse de données spécifiques sur données billettiques peuvent renseigner les habitudes des usagers. Des modèles statistiques ont été développés afin d'identifier, à partir des données brutes et anonymisées, des groupes d'usagers ayant des habitudes similaires concernant l'utilisation des transports en commun (Briand. *et al.* 2016). Ceci a permis de regrouper les usagers réguliers (environ 10 000) en 10 groupes en fonction de leurs heures de fréquentation des transports en commun. La méthode proposée permet de s'affranchir des habituelles limitations relatives au découpage en blocs horaires qui conduisent à des effets frontière arbitraires.



(a)



(b)

Figure 1 : (a) Nombre de validations enregistrées sur le réseau rennais (bus et métro) le mardi 3 juin 2014 entre 7h et 8h. (b) Profils temporels journaliers pour l'ensemble des 10 groupes d'utilisateurs identifiés (seul le mardi est présenté sur cette figure). Les usagers sont affectés à un groupe au sens de leurs habitudes temporelles d'usage des transports en commun.

Les profils journaliers types obtenus sont reportés sur la Figure 1. On constate qu'ils diffèrent par le nombre et la localisation temporelle de leurs pics ainsi que leur forme plus ou moins diffuse. Une première catégorie de groupes d'utilisateurs peut ainsi être identifiée (groupes 1, 2, 6 et 8). Ces classes présentent des profils d'activité à deux pics avec des spécificités propres à chacune d'entre elles : la classe 8 correspond à une activité plus importante dans l'après-midi, le pic du matin de la classe 2 a lieu plus tôt, tandis que la classe 6 possède une plus grande variance. Un deuxième sous-ensemble de groupes, dont le profil temporel présente trois pics, peut aussi être identifié (classes 4, 5 et 7). En plus des usages du matin et de fin d'après-midi, les passagers appartenant à ces classes utilisent les transports collectifs pour la pause déjeuner. Là encore, des différences dans l'heure et la forme des pics peuvent être notées. Enfin, la dernière catégorie de classes, qui regroupe une grande partie des usagers (environ 60%), englobe des profils correspondant à des usages diffus. Les classes 9 et 10 ont des profils diffus symétriques, tandis que le profil de la classe 3 présente un seul pic l'après-midi. Ces résultats mettent en exergue la diversité de la demande dans les transports en commun qui se caractérise par des profils d'activités à la fois diffus (60% des usagers) et plus « concentrés » temporellement (pic du matin, parfois pic du midi et le pic du soir).

Ces données permettent d'enrichir la connaissance sur les usagers et de voir notamment qu'une grande majorité d'entre eux n'a pas de routine de déplacement pendulaire, contrairement à une idée répandue. Les routines temporelles ainsi identifiées caractérisent finement la demande, et peuvent servir d'entrées aux modèles de simulation ; elles peuvent

également être utilisées par les gestionnaires urbains et les opérateurs de transport pour des opérations de planification.

3. Le système de vélos en libre service à Paris

Sur les systèmes de Vélos en Libre Service (VLS), deux types de données sont généralement disponibles et exploitables. 1) On peut avoir accès aux données agrégées représentant le stock de vélos et d'emplacements libres des stations à différents instants. 2) Ou alors, on dispose de données d'entrée et de sortie du réseau VLS : pour chaque déplacement, l'heure et la station de départ, ainsi que l'heure et la station de remise du vélo sont disponibles. Le type de données mis à disposition conditionne évidemment les analyses que l'on peut mener par la suite. Nous détaillerons ici deux exemples de réutilisation de ces données.

Open data et analyse des systèmes VLS

Le développement de l'opendata a conduit à l'ouverture de données temps réel de stocks de vélos pour un nombre important de systèmes VLS. Saisissant cette opportunité, la plateforme VLSSTATS a été conçue pour accéder aux données générées par une vingtaine de systèmes VLS comme ceux de Paris, Lyon ou New-York afin de les analyser à différentes échelles spatiales et temporelles³. L'originalité de ce travail (retenu parmi les 10 meilleures réutilisations d'open data par Datagouv.fr en 2014) est de « donner du sens » à ces données à l'aide d'une vision historique, nécessaire pour mieux comprendre, analyser et au final améliorer de tels services.

A l'aide de l'outil et à partir des données agrégées historisées, on peut procéder à la comparaison de statistiques mensuelles calculées sur différents systèmes VLS telles que la taille du réseau (nombre de stations, nombre de bornes et nombres de vélos), l'usage du réseau quantifié par le nombre d'heures d'usage des vélos par jour, ou encore le rendement du système en estimant le nombre d'heures d'usage par jour et par vélo. D'autres indicateurs relatifs au taux de défaillance ou de saturation du système sont également estimés. Le premier renseigne sur le nombre de bornes hors service tandis que le second indique les pourcentages de stations pleines et de stations vides, ces deux derniers indicateurs étant directement corrélés aux principales difficultés rencontrées par les usagers de ce système de mobilité partagée.

L'outil de visualisation développé offre plusieurs façons d'explorer ces données. On peut survoler une ville pour avoir des détails sur le fonctionnement du système VLS ou choisir une station particulière pour avoir une vision plus fine. On peut par exemple étudier la dernière semaine de données d'une ville. L'interface offre la possibilité de choisir une des statistiques pour visualiser son évolution temporelle pour l'ensemble du système et sa répartition spatiale. Elle offre également la possibilité de mener une analyse à l'échelle d'une station pour connaître son état et pour avoir des détails complémentaires sur son fonctionnement. La figure 2 illustre un exemple de résultat fourni par cette application.

³ <http://vlsstats.ifsttar.fr/>



Enfin, naviguez dans l'historique en utilisant les flèches gauche et droite de votre clavier ou en utilisant les boutons ci dessous:

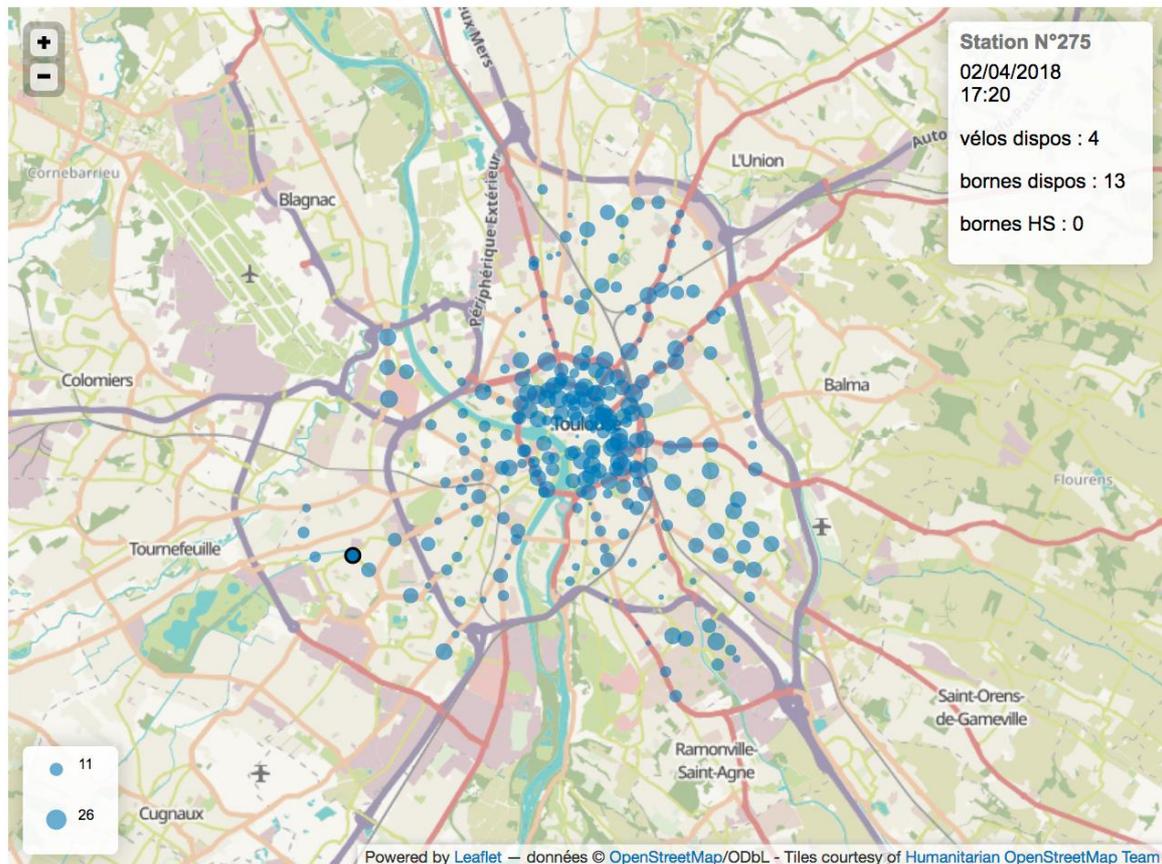
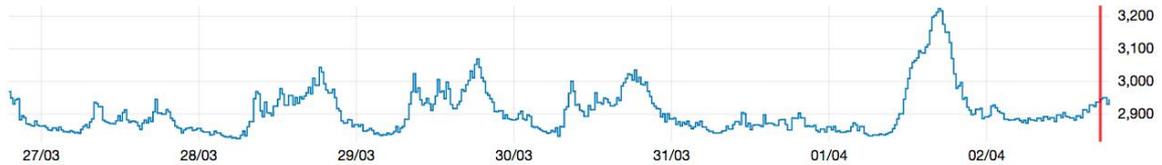
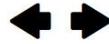


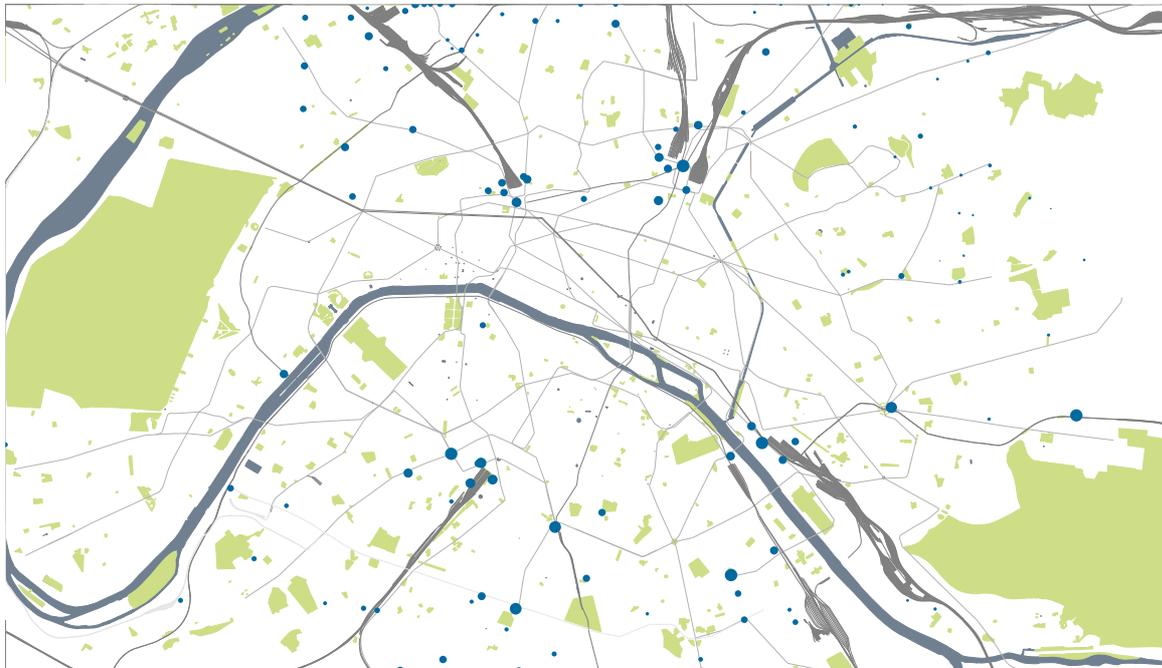
Figure 2 : Exemple de statistiques mensuelles estimées sur le réseau de vélos en libre service de la ville de Toulouse (<http://vlsstats.ifsttar.fr/>)

Données de trajets et usage du système VLS. Cas du velib'1 à Paris

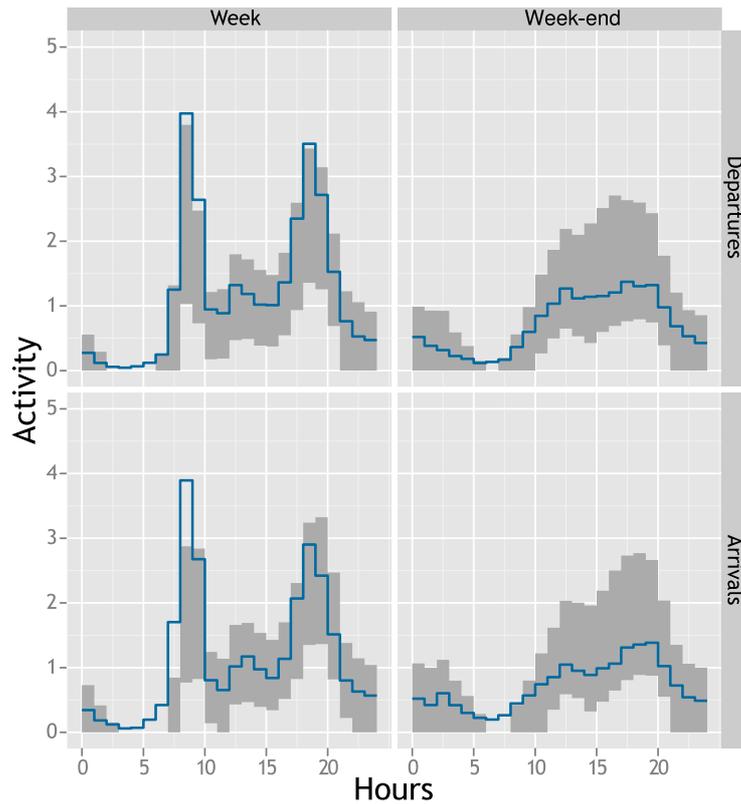
Un partenariat avec la mairie de Paris et JC Decaux a permis l'accès à une base de données de trajets Vélib' (entrées-sorties sur le réseau) à des fins de recherche. Ces données par nature plus riches que les données agrégées, permettent de mener des analyses plus avancées sur l'usage du système VLS. La base de données mise à disposition comportait l'ensemble des déplacements anonymisés sur une durée de 5 mois avec pour chacun d'entre eux la station et l'heure de départ ainsi que la station d'arrivée et l'heure de remise du vélo.

Une première analyse a consisté à mieux comprendre l'usage du vélib' à partir des données d'usage et à relier ces usages aux données socio-économiques de la ville. Nous avons proposé dans les articles Côme *et al.* 2013, Côme *et al.* 2014 un modèle de regroupement automatique des stations selon leur profil d'usage. Cette modélisation statistique (clustering) permet de partitionner de manière automatique les stations Vélib' (1200 au total) en 8 groupes (ou classes). Les stations appartenant au même groupe correspondent à des usages similaires identifiés par un profil d'usage journalier typique. Celui-ci caractérise la station par le nombre moyen de départs et de retours durant chaque heure d'une journée type de semaine et de week-end. Ces profils sont normalisés par l'activité moyenne de la station afin d'être comparable d'une station à l'autre.

La figure 3a montre deux exemples de groupes de stations : le premier dénommé « gares» correspond à un ensemble de stations Vélib' toutes localisées à proximité des gares. Les profils d'usage pour de telles stations sont présentés figure 3b. On peut noter que leur comportement est caractérisé par des pics très marqués aux heures de pointe en semaine et une utilisation relativement faible le week-end.



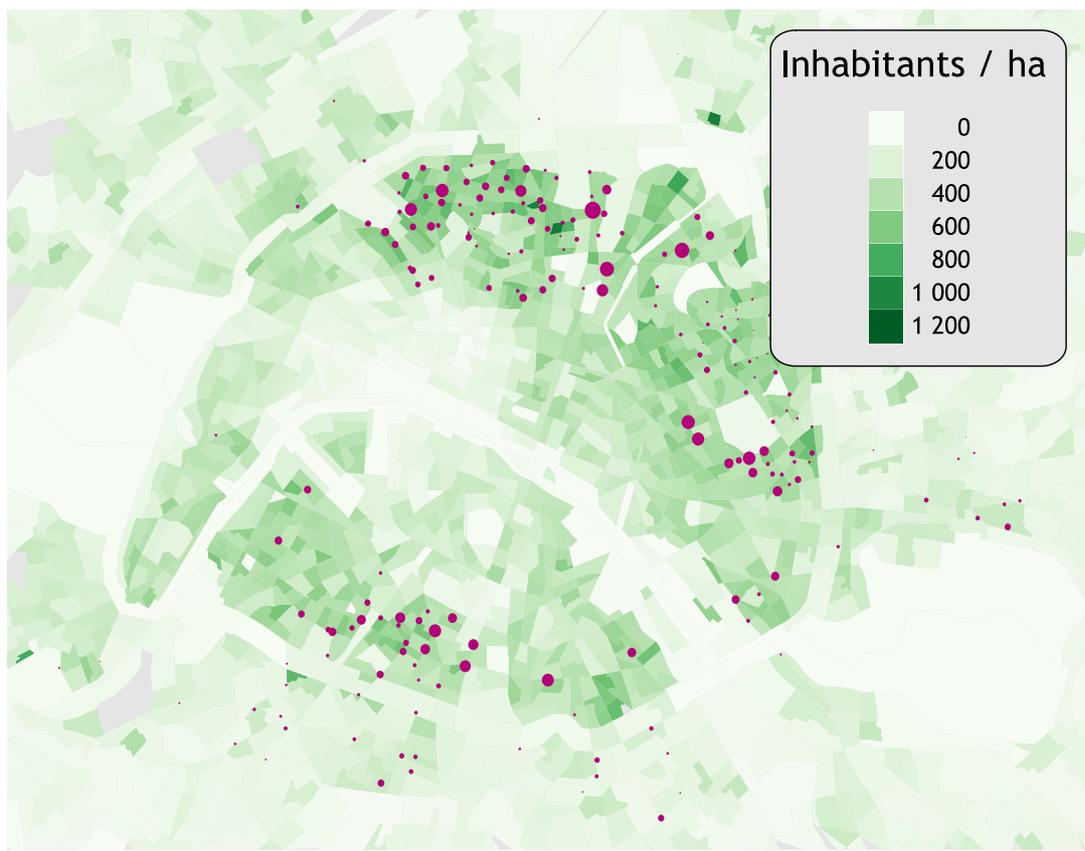
3.a



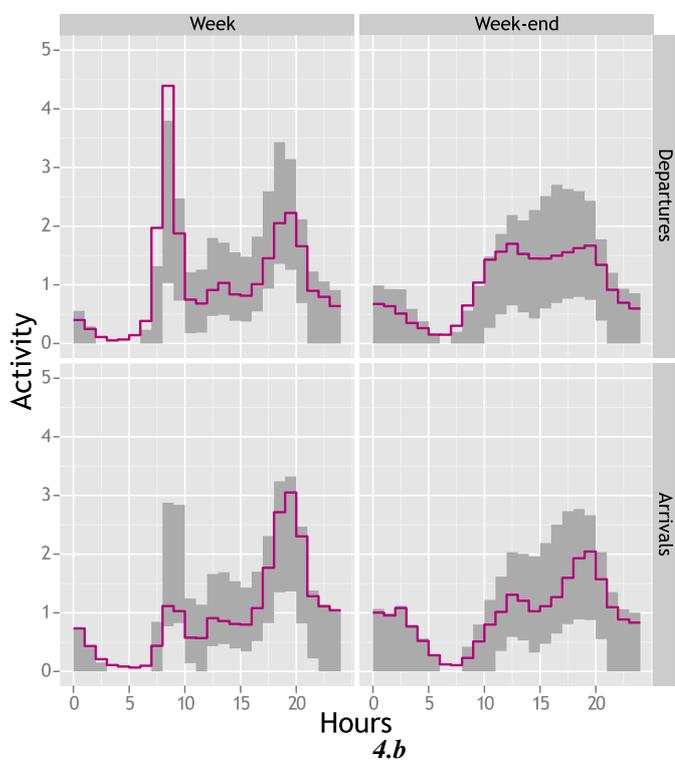
3.b

Figure 3 : Exemple d'un groupe de stations dénommé « Gares » obtenu par partitionnement automatique. Localisation des stations Vélib' de ce cluster et profils d'activité temporel de ce groupe de stations (nombre moyen de départs et de retours durant chaque heure d'une journée type de semaine et de week-end).

Un autre groupement de stations que nous avons dénommé « habitations » est intéressant à analyser. Comme le montre son profil d'usage journalier moyen présenté figure 4b, ce groupe de stations se caractérise par une forte utilisation le matin en semaine, et un fort pic correspondant aux retours le soir à la sortie du travail, il s'agit de mouvements pendulaires de la semaine et les stations concernées se situent bien dans des quartiers résidentiels comme le montre la figure 4a. On notera toutefois que les stations des secteurs principalement résidentiels du 16^e et 7^e arrondissements n'apparaissent pas dans ce cluster. Il semble que les habitants de ces quartiers n'utilisent pas le Vélib' pour se rendre sur leur lieu de travail.



4.a



4.b

Figure 4 : Exemple d'un groupe de stations dénommé « habitations » obtenu par partitionnement automatique (ou clustering). Localisation des stations Vélib' de ce cluster et profils d'activité temporel de ce groupe de stations (nombre moyen de départs et de retours durant chaque heure d'une journée type de semaine et de week-end).

Ce type de clustering offre la possibilité à l'exploitant d'adapter plus finement son offre station par station (dimensionnement des stations, politique d'achalandage en vélos, programme de maintenance...)

Un autre outil de visualisation⁴ a été conçu pour une exploration d'un autre corpus de données correspondant à 2 500 000 déplacements enregistrés sur l'ensemble des stations Vélib' pendant le mois d'avril 2011. La carte interactive qu'il construit représente le nombre de départs ou d'arrivées de Vélib' de chacune des stations durant la plage temporelle sélectionnée grâce aux différents menus. Le survol d'une station permet de visualiser les stations d'origines ou de destinations. Cet outil permet une exploration des données en y incorporant à la fois la dimension « clustering de stations » et la dimension « profils journaliers d'usage ». Plusieurs possibilités sont offertes à l'utilisateur notamment pour sélectionner un groupe particulier de stations correspondant à un des usages types identifiés. .

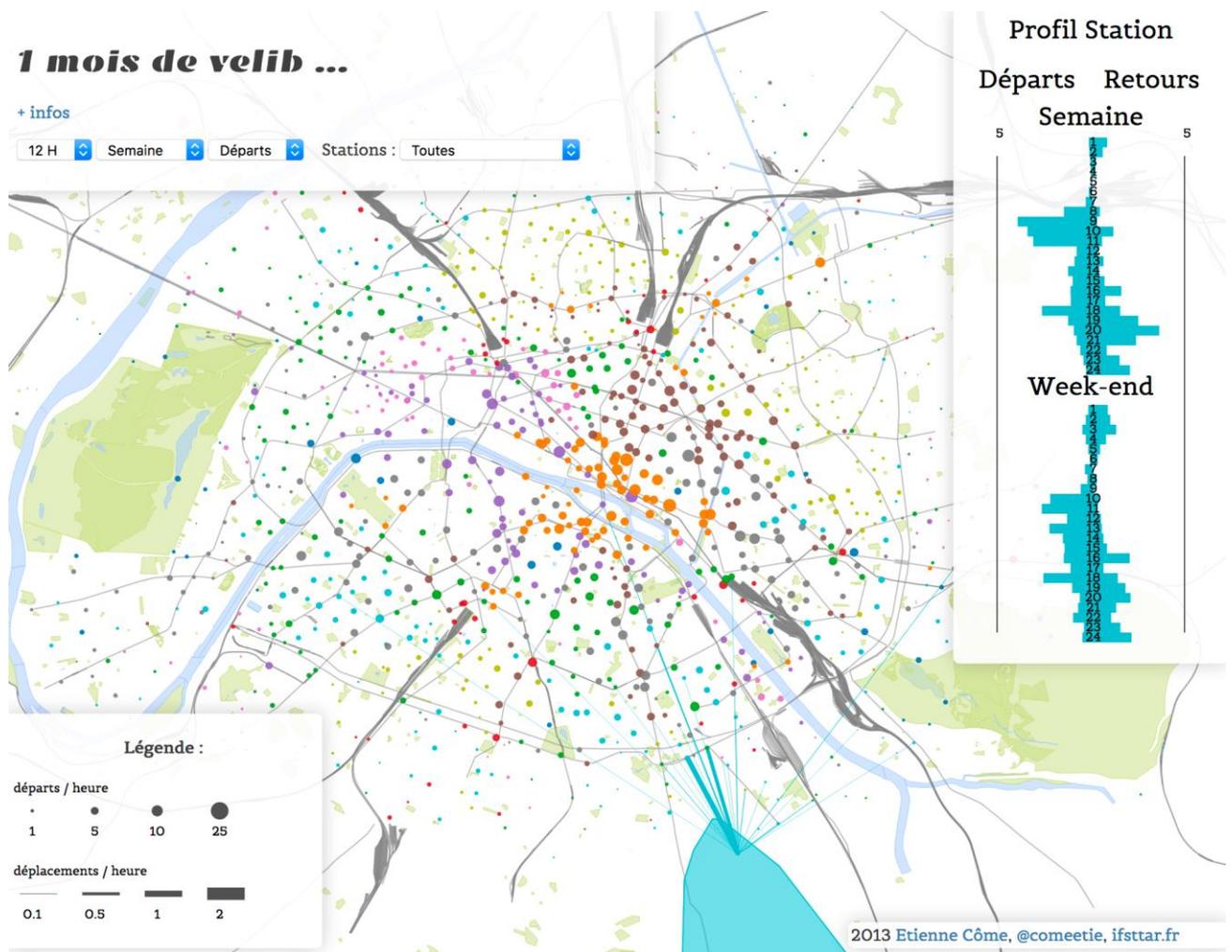


Figure 5 : Outil de visualisation pour l'exploration des données trajets du Vélib' enregistrés pendant 1 mois.
<http://www.comeetie.fr/galerie/velib/>

⁴ <http://www.comeetie.fr/galerie/velib/>

Nous pouvons observer au travers de cet exemple le potentiel d'exploration et d'analyse qu'offrent ces nouvelles sources de données pour mieux cerner les déterminants de la mobilité à l'aide d'un mode de transport en particulier. De telles analyses permettent de mieux répondre aux questions sous-jacentes telles que Où, Quand, et Pourquoi les utilisateurs de tels systèmes se déplacent.

4. Conclusion

Les nouvelles données associées aux algorithmes issues du domaine en ébullition des datasciences ouvrent de nombreuses perspectives pour une connaissance fine de la mobilité, une analyse du fonctionnement des réseaux de transport en commun et des systèmes de mobilité partagée. La puissance publique encourage la diffusion et l'ouverture des données d'usage au travers notamment de la législation sur l'open data.

Ces différentes données placent le problème de la modélisation des transports dans un contexte méthodologique nouveau pour la qualification des mobilités et l'observation de la multimodalité et l'intermodalité en particulier. La fenêtre d'observation des déplacements est importante et quasi continue ; les informations recueillies sont réparties le long du réseau. Ces données possèdent des avantages intrinsèques intéressants comme l'exhaustivité, une finesse spatiale et temporelle, l'absence de biais de réponse que l'on peut rencontrer dans les données d'enquête.

Il ne faut cependant pas négliger les obstacles techniques, organisationnels et institutionnels inhérents à l'utilisation de telles données :

- Mauvaise préparation des données, données bruitées ou manquantes, volumétrie
- Mise en base de données utilisant des technologies non pérennes ou non scalables
- Entre l'usager, le producteur technique de la donnée, l'exploitant du système de transport, l'autorité organisatrice, les responsables territoriaux... le jeu d'acteurs se complexifie rendant difficile certains arbitrages par exemple sur la propriété des données et leur droit d'usage
- Le manque de transparence sur la finalité des traitements
- ...

Mais ces données permettent à l'évidence la création de nouveaux services à l'usager et/ou à l'exploitant. Une politique d'ouverture des données (publiques et non publiques) encore plus ambitieuse est encouragée dans les nouvelles orientations de la puissance publique comme en témoignent les récents rapports sur l'Intelligence Artificielle (IA) (rapport de la mission Villani⁵, rapport de l'Académie des Technologies⁶, rapport de l'administrateur général des données⁷) où le secteur du transport et de la mobilité est systématiquement cité comme un secteur porteur d'innovation de rupture.

Références

⁵ https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf

⁶ http://academie-technologies-prod.s3.amazonaws.com/2018/04/06/13/49/30/183/Rapport_IA_DEF.pdf

⁷ https://www.etalab.gouv.fr/wp-content/uploads/2018/04/RapportAGD_2016-2017_web.pdf

J. Zmud, M. Lee-Gosselin, M. Munizaga, J-A. Carrasco “Transport survey methods. Best practice for Decision Making”, Esmeral Group Publishing Limited (2013)

C-D. Cottrill, C. Pereira Francisco, F. Zhao, I. Ferreira Dias, H-B. Lim, M. Ben Akiva, P.C. Zegras, “The Future Mobility Survey: Experiences in developing a smartphone based travel survey in Singapore” (2013).

TRB – « Travel Survey Manual » (2012) <http://www.travelsurveymanual.org/Table-of-Contents.html>

F. Jutand, Rapport Jutand : L’ouverture des données de transport, Mars 2015.

A-S. Briand, [E. Côme](#), [M-K. El Mahrsi](#), [L. Oukhellou](#), A mixture model clustering approach for temporal passenger pattern characterization in public transport. [International Journal on Data Science and Analytics 1\(1\)](#): 37-50 (2016)

E. Côme, L. Oukhellou, Model-based count series clustering for bike sharing system usage mining: A case study with the vélib system of Paris, *ACM Trans. Intell. Syst. Technol.* 5 (3) (2014) 39:1–39:21. doi:10.1145/2560188.

E. Côme, N-A. Randriamanamihaga, L. Oukhellou, P. Aknin. Spatio-temporal Analysis of Dynamic Origin-Destination Data Using Latent Dirichlet Allocation: Application to Vélib' Bike Sharing System of Paris. *TRB 93rd Annual meeting*, Jan 2014, France. TRANSPORTATION RESEARCH BOARD, 19p, 2014.

Renouveau de l’Intelligence Artificielle et de l’apprentissage automatique, rapport de l’académie des technologies, Mars 2018, http://academie-technologies-prod.s3.amazonaws.com/2018/04/06/13/49/30/183/Rapport_IA_DEF.pdf

Cédric Villani, Donner un sens à l’intelligence artificielle, pour une stratégie nationale et européenne, 2018.

https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf

La donnée comme infrastructure essentielle, 2016-2017, La documentation française, https://www.etalab.gouv.fr/wp-content/uploads/2018/04/RapportAGD_2016-2017_web.pdf