

INSTITUT FRANCAIS
DES SCIENCES
ET TECHNOLOGIES
DES TRANSPORTS,
DE L'AMENAGEMENT
ET DES RESEAUX

Fouille de données billettiques pour l'analyse des comportements de mobilité

Latifa Oukhellou, Etienne Côme

24 Avril 2017
Restitution Ville de Rennes



IFSTTAR

Traces numériques



de mobilité



Des données particulières

Les données billetiques

Des données simple



- identifiant anonymisé de l'utilisateur (3 mois)
- type de titre de transport
- station de la validation (+ligne et direction pour les bus)
- jour et heure de la validation

Open-Data ? (ex STIF données agrégées)

Des données simple



- identifiant anonymisé de l'utilisateur (3 mois)
- type de titre de transport
- station de la validation (+ligne et direction pour les bus)
- jour et heure de la validation

Une données à part l'identifiant usager

Un jeu de données riche



- identifiant anonymisé de l'utilisateur (3 mois)
- type de titre de transport
- station de la validation (+ligne et direction pour les bus)
- jour et heure de la validation

2 ans de profondeur temporelle

Plan

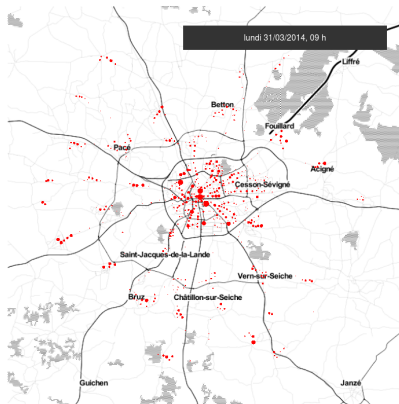
Que faire sans identifiant ?

- Analyse des volumes entrants
- Clustering des jours
→ détection d'outliers, prévision moyen terme,...
- Clustering des stations
→ analyse de l'offre, prévision moyen terme,...

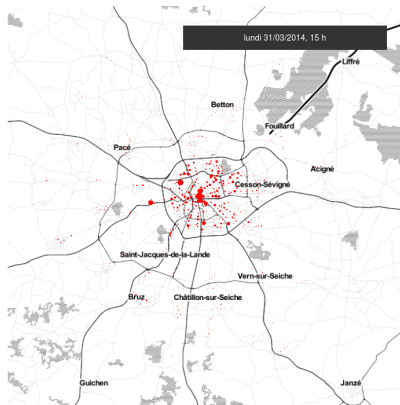
Que permet de faire l'identifiant usager ?

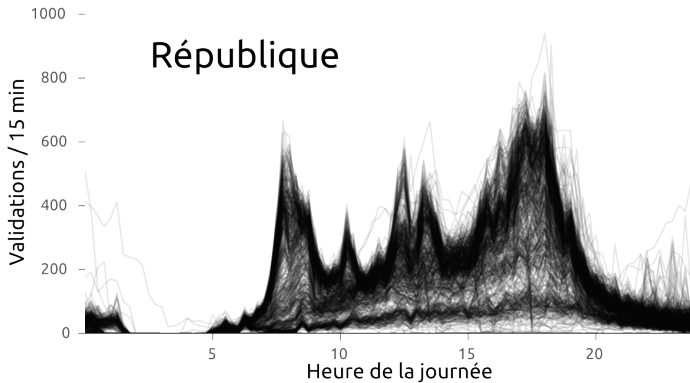
- Reconstruction des destinations
→ analyse des correspondances, matrice OD dynamique,...
- Prévision court terme des OD
- Clustering des usagers

Analyse des volumes entrant



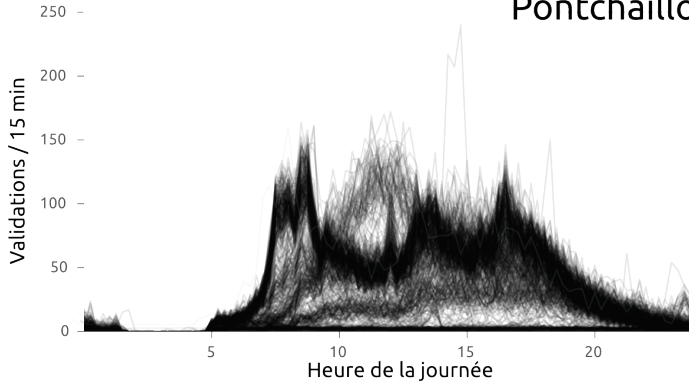
Analyse des volumes entrant



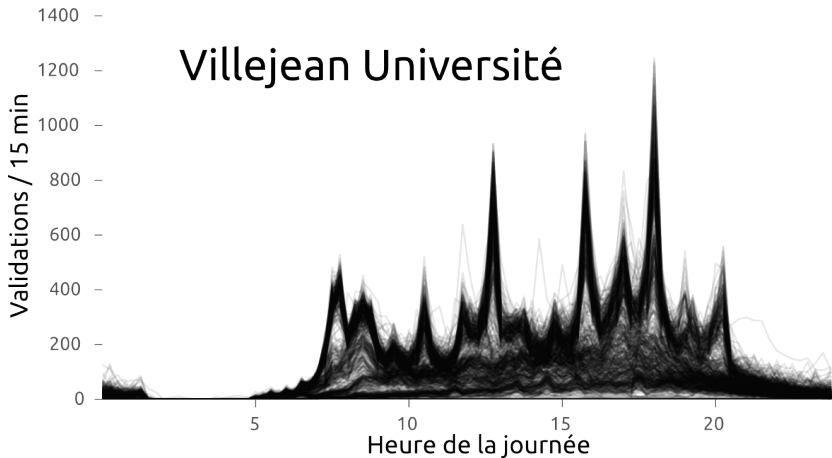


Des profils qui varient dans l'espace et dans le temps

Pontchaillou

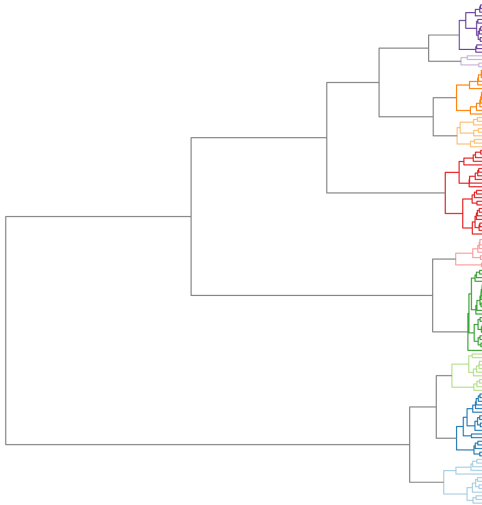


Des profils qui varient dans l'espace et dans le temps

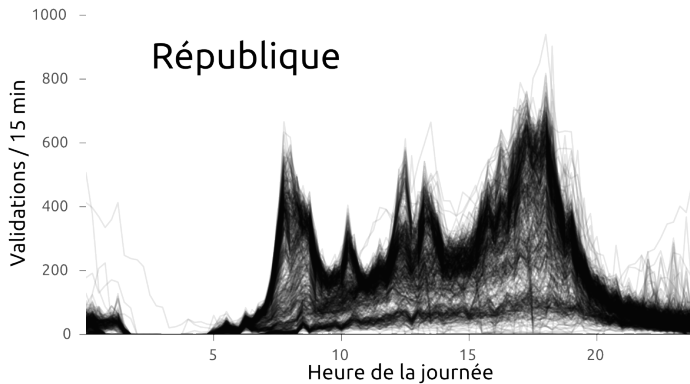


Des profils qui varient dans l'espace et dans le temps

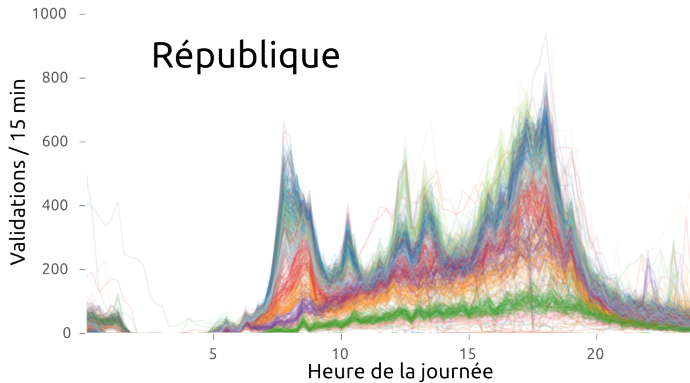
Une variété inter journalière visible (CAH)



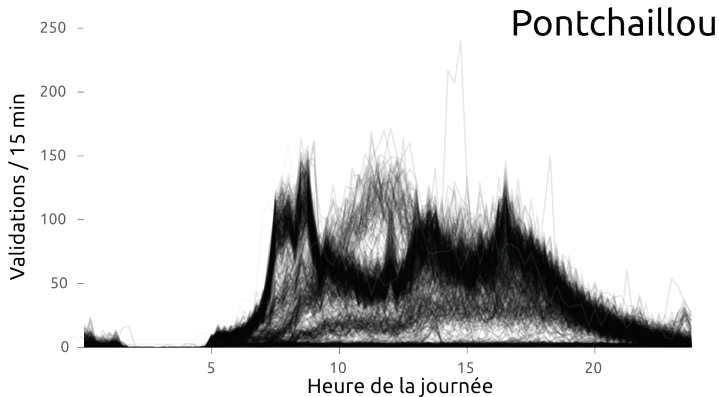
Une variété inter journalière visible (CAH)



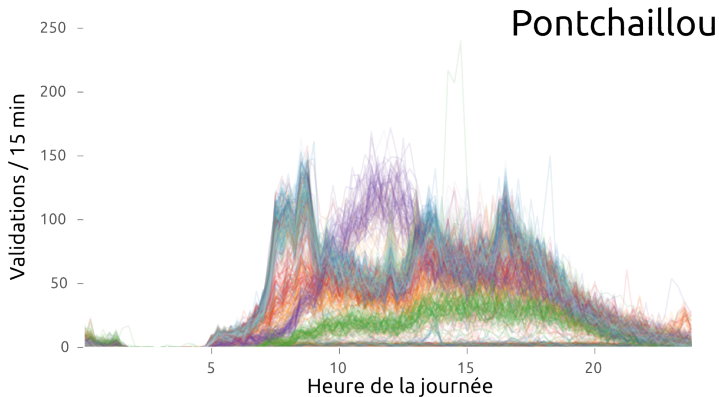
Une variété inter journalière visible (CAH)



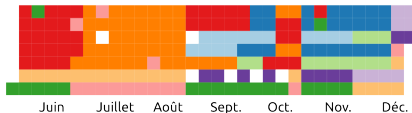
Une variété inter journalière visible (CAH)



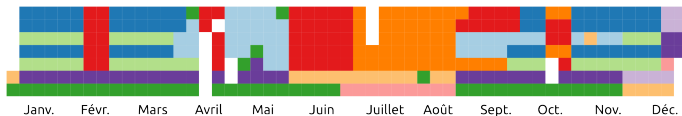
Une variété inter journalière visible (CAH)



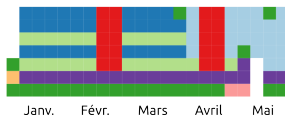
Lundi
Mardi
Mercredi
Jeudi
Vendredi
Samedi
Dimanche



2014

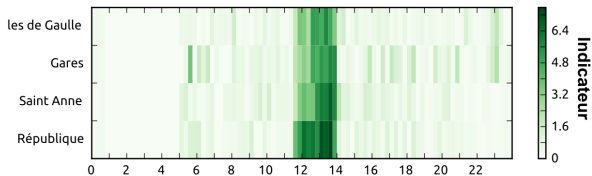


2015



2016

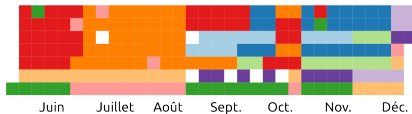
Qui s'explique majoritairement par des effets calendaires



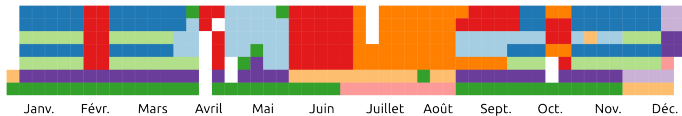
Qui peuvent être exploités pour détecter des outliers



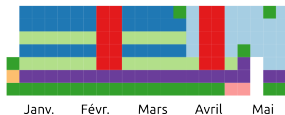
Lundi
Mardi
Mercredi
Jeudi
Vendredi
Samedi
Dimanche



2014



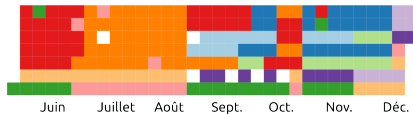
2015



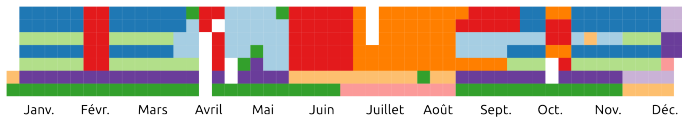
2016

Oufaire des prévisions à moyen terme

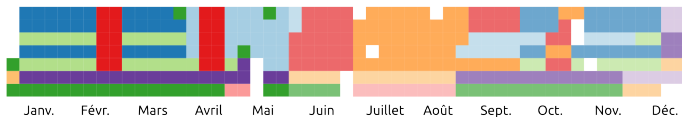
Lundi
Mardi
Mercredi
Jeudi
Vendredi
Samedi
Dimanche



2014



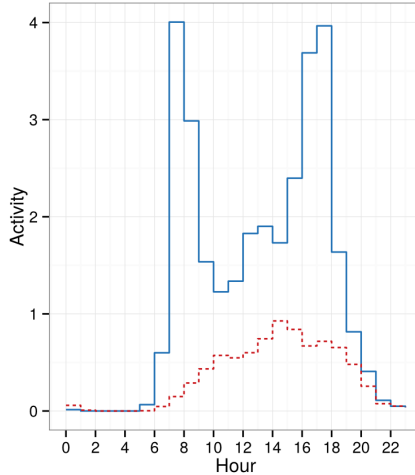
2015



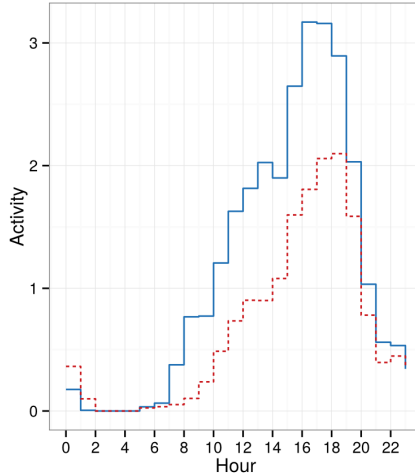
2016

Ou faire des prévisions à moyen terme

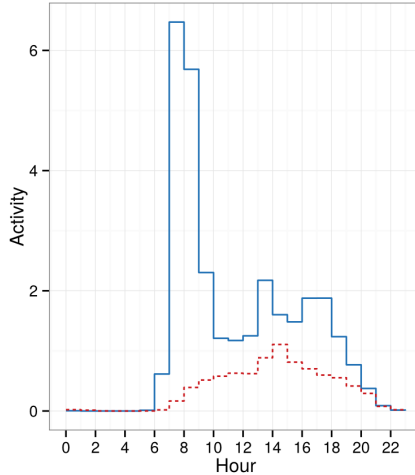
Une variabilité également spatiale



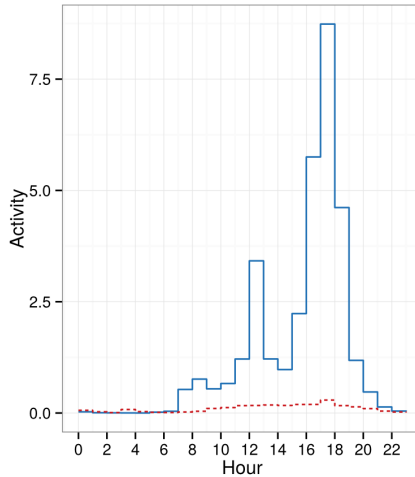
Une variabilité également spatiale



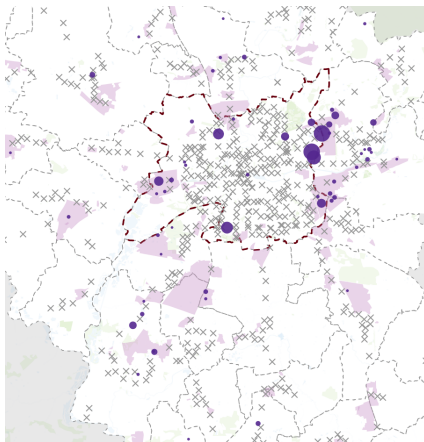
Une variabilité également spatiale



Une variabilité également spatiale



Une variabilité également spatiale

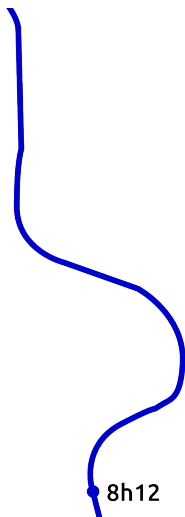


et des différences de contexte socio-économique

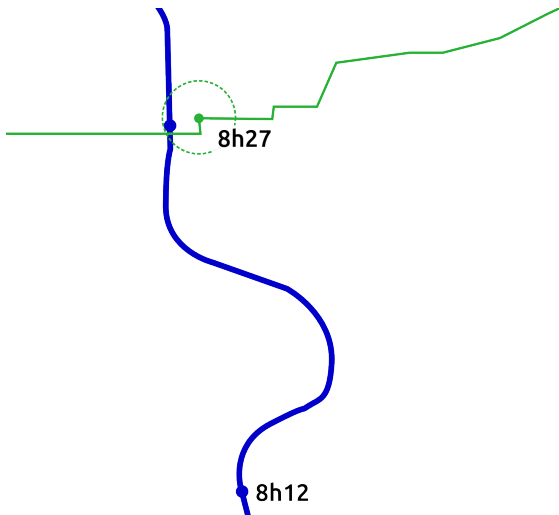
L'identifiant usager

pour enrichir les données

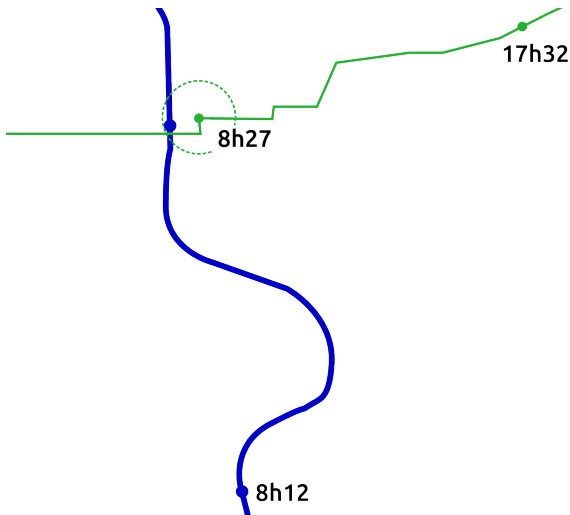
Permet de reconstituer une partie significative des trajets



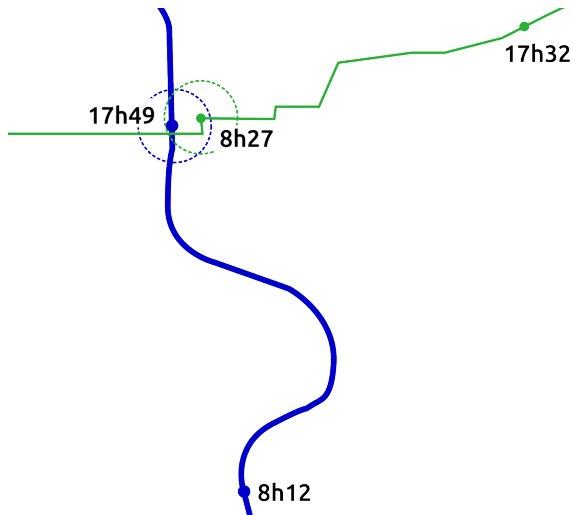
Permet de reconstituer une partie significative des trajets



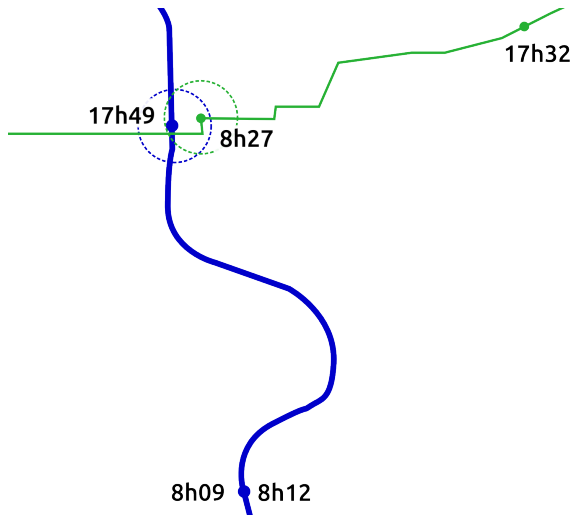
Permet de reconstituer une partie significative des trajets



Permet de reconstituer une partie significative des trajets



Permet de reconstituer une partie significative des trajets



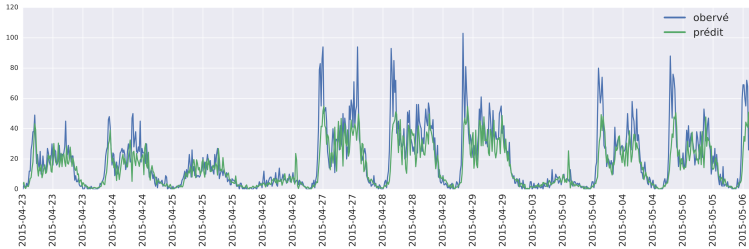
Enrichissement des données

- Utilisation de l'origine suivante pour inférer la destination
- Utilisation de seuils spatiaux et temporels
- Détection des correspondances
- Nécessite de garder les identifiants 48h

72% de destinations reconstruites

- Permet des agrégations par OD
- Analyse des Pôles d'échange (C. Richer)
- Matrice OD dynamique ou serpent de charge

Prédiction d'OD court terme



Prédiction d'OD court terme

- Prédiction des flux sur les 15min suivantes
- Possibilité de comparer différents modèles
- Profondeur temporelle de la base de donnée importante

Modèle	Apprent.	Val.	Test
Calendaire	11.6	12.35	8.86
Var	4.34	5.56	5.88
LSTM (métro)	4.01	4.94	5
LSTM (métro + bus)	2.73	4.52	4.71

Erreur (MSE) estimée pour différents modèle pour la prévision à court terme d'OD.

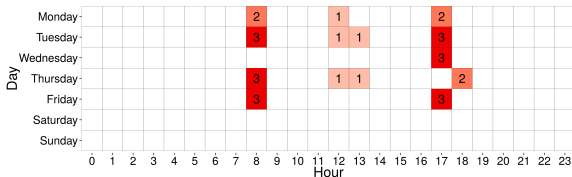
Clustering usagers

pour une analyse centrée sur l'utilisateur

Objectifs

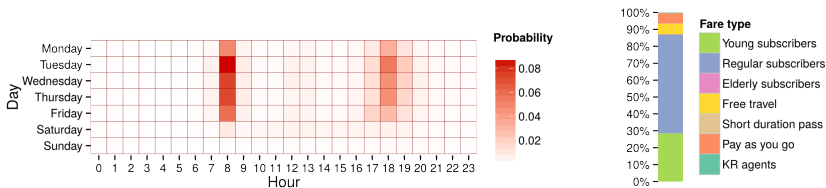
- Etudier les habitudes temporelle des usagers
- Regarder comment se décompose la charge
- Mieux connaître les usagers

Méthode



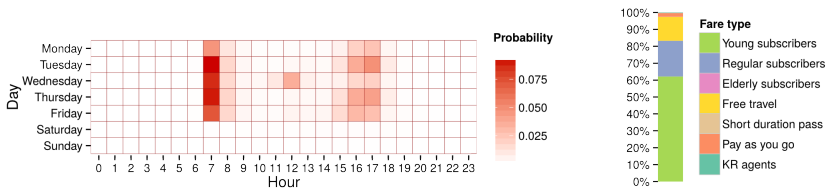
- Profil d'utilisation
- Clustering

Un usage pendulaire



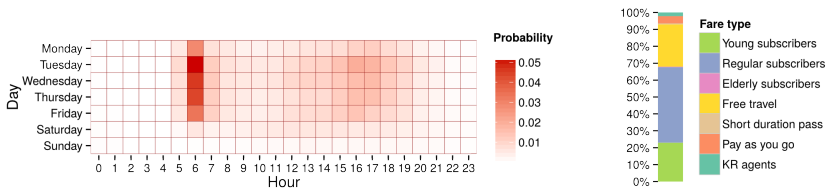
Profil moyen d'un cluster 4.55% du jeu de données

Un usage pendulaire



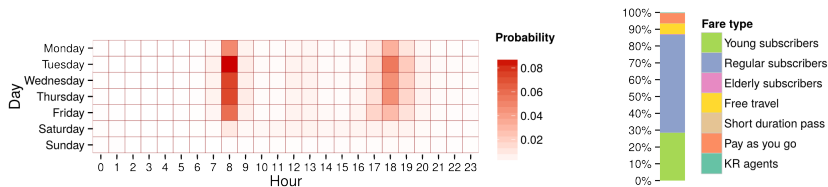
Profil moyen d'un cluster 12.54% du jeu de données

Un usage pendulaire



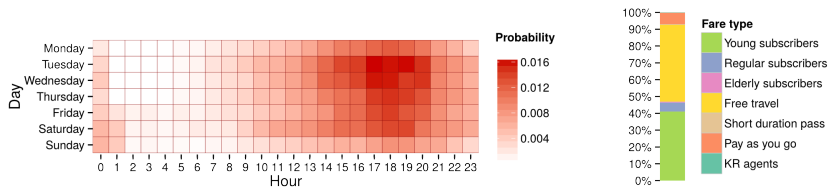
Profil moyen d'un cluster 3.6% du jeu de données

Mais pas que



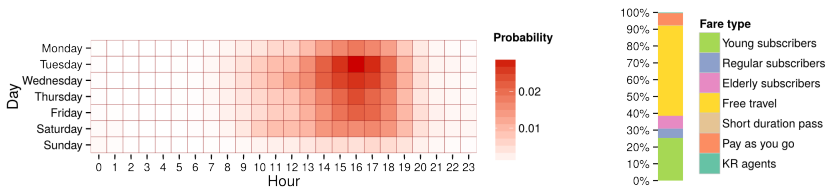
Profil moyen d'un cluster 15.13% du jeu de données

Mais pas que



Profil moyen d'un cluster 6.44% du jeu de données

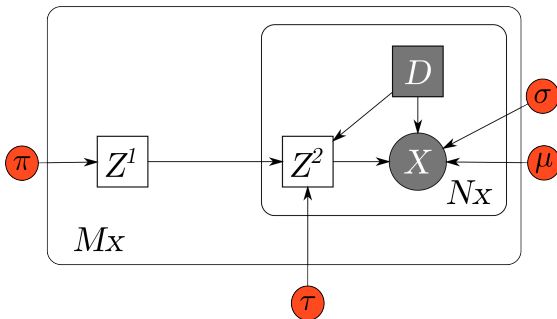
Mais pas que



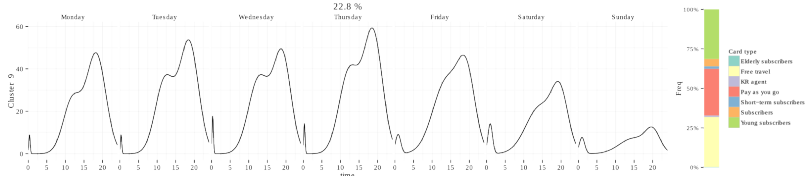
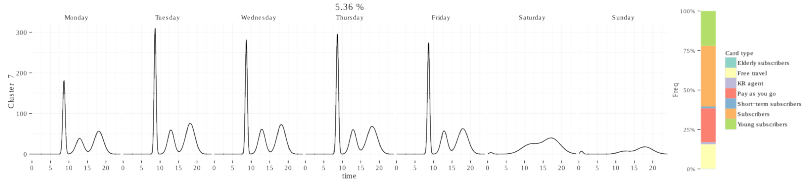
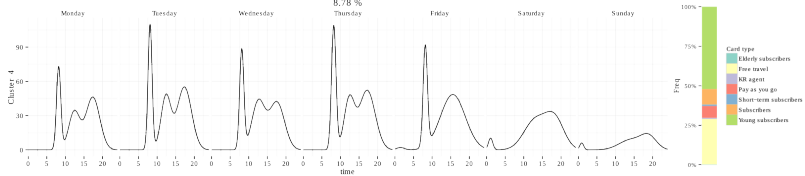
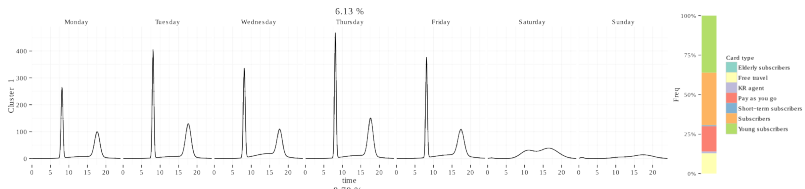
Profil moyen d'un cluster 8.64% du jeu de données

Méthode

Prise en compte des temps exacts de validation
(suppression des comptages / heure)



Modèle génératif pour le clustering d'utilisateurs en temps continu



Analyse du comportement horaire des usagers

- Des sous populations assez bien différenciées
- Permet de sortir de l'usager moyen
- Mieux comprendre les besoins

Conclusion

- Des données riches qui permettent d'étudier la variabilité temporelle et spatiale de la demande
- Des travaux exploratoires montrant différentes caractéristiques de la demande
- Des informations sur les usagers pertinentes (pour la reconstruction des destination ou plus)

Perspectives

- Poursuite des travaux sur la prévision moyen terme/court terme
- Poursuite des travaux sur la détection et la caractérisation d'anomalies
- Mise en oeuvre sur des réseaux plus complexe (ex Paris)

Merci pour votre attention.

et à toute l'équipe du projet
Latifa Oukhellou, Mohamed El Marhsi
Anne Sarah Briand, Florian Toqué
Cyprien Richet, Nicolas Coulombel
Keolis Rennes

IFSTAR - GRETTIA

Cité Descartes

14-20 boulevard Newton

77447 Marne-la-Vallée Cedex 2 - FRANCE

Mail: etienne.come@ifsttar.fr

Tel. : +33 (0)1 81 66 87 11

Site : <http://www.grettia.ifsttar.fr>

Publications

- El Mahrsi, M., Briand, A. S., Côme, E., & Oukhellou, L. *"Utilité des données billettiques pour l'analyse des mobilités urbaines: le cas rennais"* Données urbaines, Economica, 2015, 11p.
- El Mahrsi, M., Côme, E., Oukhellou, L. & Verleysen, M. *"Clustering Smart Card Data for Urban Mobility Analysis"* IEEE Transactions on Intelligent Transportation Subsystems (Volume: PP, Issue: 99), pp 1 – 17, 2016.
- Briand, A. S., Côme, E., El Mahrsi, M., & Oukhellou, L. *"A mixture model clustering approach for temporal passenger pattern characterization in public transport"*, International Journal of Data Science and Analytics, Avril 2016, Volume 1, issue 1, pp 37-50.
- Toqué, F., Côme, E., El Mahrsi, M., Oukhellou, L. *"Forecasting Dynamic Public Transport Origin-Destination Matrices with Long-Short Term Memory Recurrent Neural Networks"*. In Proceedings of IEEE 19th International Conference on Intelligent Transportation Systems , Rio de Janeiro , BRESIL 2016.
- Briand, A. S., Côme, E., Trépanier, M., & Oukhellou, L. *"Analysing year-to-year changes in public transport passenger behaviour using smart card data"* Transportation Research Part C : Special Issue on Smartcard data (accepté)

Articles en cours de soumission :

- Briand, A. S., Côme, E., & Oukhellou, L. *"Anomaly detection and characterization in Smart-Card data"* (en cours de soumission)