

Séminaire PUCA : Modèles et Big Data

L. Oukhellou

Université Paris-Est, IFSTTAR, Marne-la-vallée, France

4 Mai 2017, Champs-sur-marne

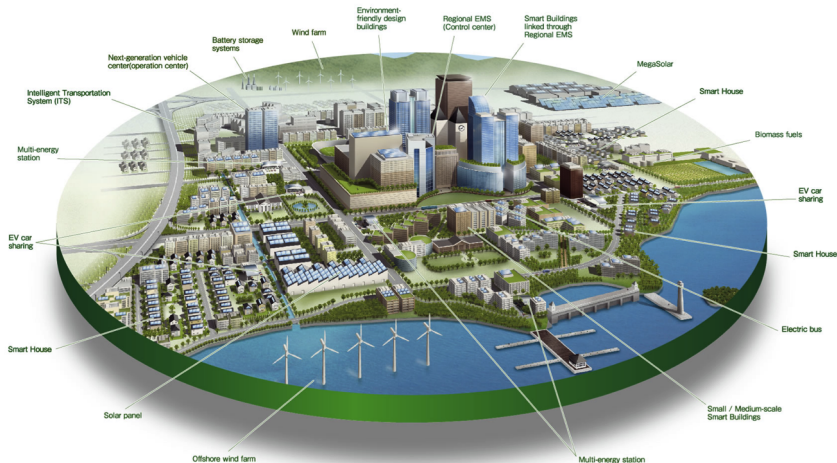
Big Data, Contexte

- Capacités de stockage
- Coûts de stockage divisé par 100 000 en 30 ans (1Moctets : 700\$ en 1981, 1 centime en 2013)
- Augmentation de la quantité d'information collectée : 5 exabytes jusqu'en 2013, même quantité en 2 jours

→ **Big Data (Volume, Vélocité (données multi-sources), Véracité)**



Energy and water domains : smart meters



Quelles données pour quels objectifs

Catégorisation

- Données dédiées (ex : capteurs intelligents ou smart meters)
- ou Usage secondaire des données (GPS, WIFI, billettiques, réseaux sociaux,)

Objectifs

- Monitoring de systèmes urbains
- Outils d'aide à la décision
- **Valorisation des données avec des challenges scientifiques**
(*business intelligence*) ex IVADO au Canada

Transport

Big Data versus Enquêtes (complémentarité-substitution)

- Données d'enquêtes (5 à 10 ans), tous modes, 1 semaine jours ouvrés, méta-données
- Traces numériques,

Données Transport (offre et demande)

- Données de l'offre (GTFS, General Transit Feed specification)
- Traces numériques sur la demande
 - ▶ Traces WIFI, GPS
 - ▶ **Données billettiques**

→ Données massives, longitudinales, spatio-temporelles **mais** manquantes, pas de métadonnées (anonymisation)

→ Enrichissement et traitements avancés de données

Au début du siècle ...

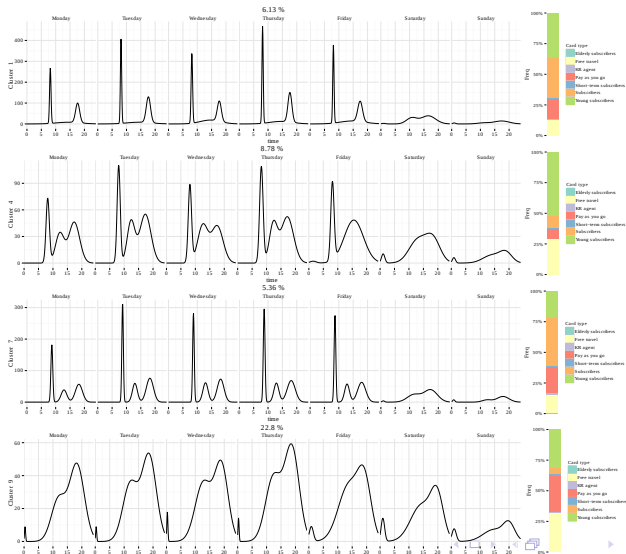


Données billettiques et Mobilité : Tryptique "Usager, Réseau, Territoires"

- **Usager** : comportements de mobilité
- **Réseau** : modéliser et prédire les flux des passagers (plusieurs horizons temporelles)
- **Territoires** : mesurer l'intensité urbaine

Données billettiques de Rennes, Données Vélib'

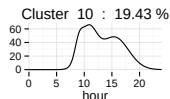
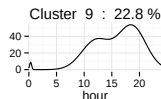
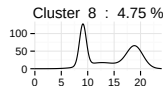
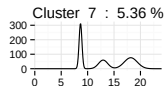
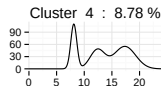
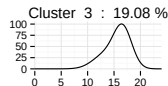
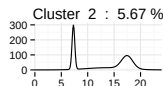
Clusters' presentation : Temporal activity profiles and card type



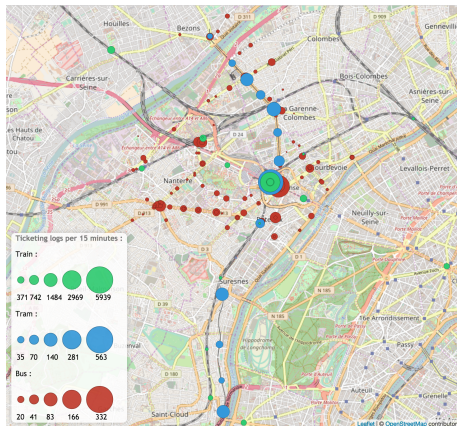
All clusters' Tuesday activity

Mobility Patterns

- Very regular, commuters
- Regular, almost no week-end and night activity
- More diffuse patterns, inverted patterns



Passenger flow forecasting with Smart Card Data (La défense)



Short and long term passenger flow prevision with Smart Card Data (La défense)

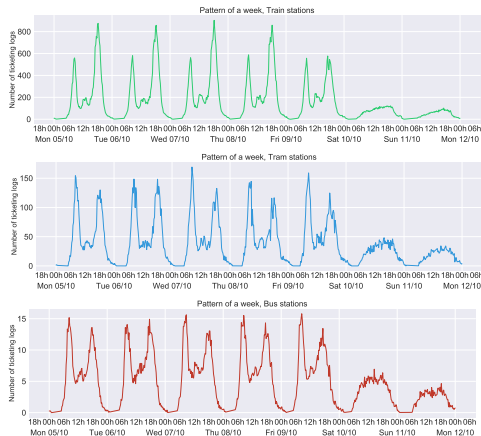
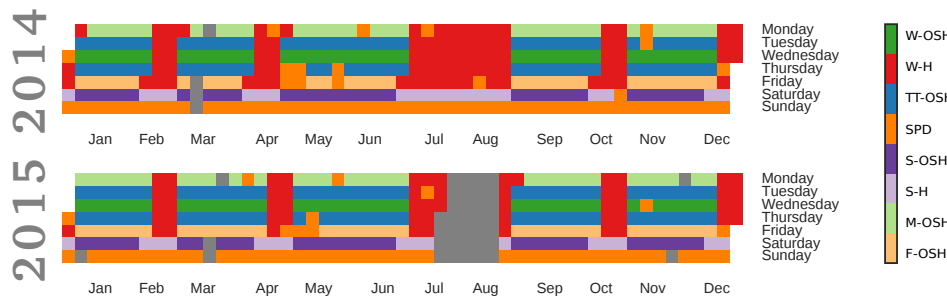


Figure: Weekly patterns of the number of ticketing logs per transportation mode (railway, Tram and Bus)

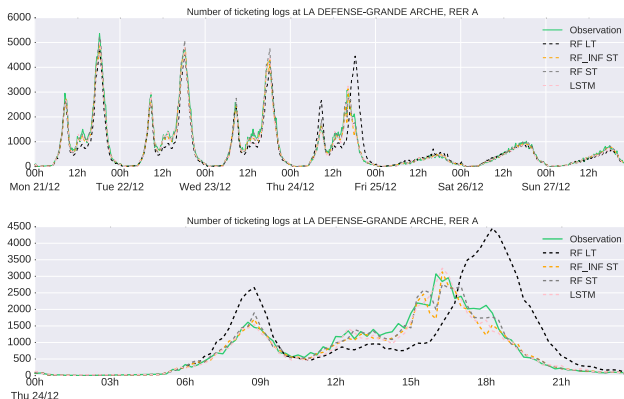
Long term passenger flow forecasting with Smart Card Data

Prévision long-terme pour la planification

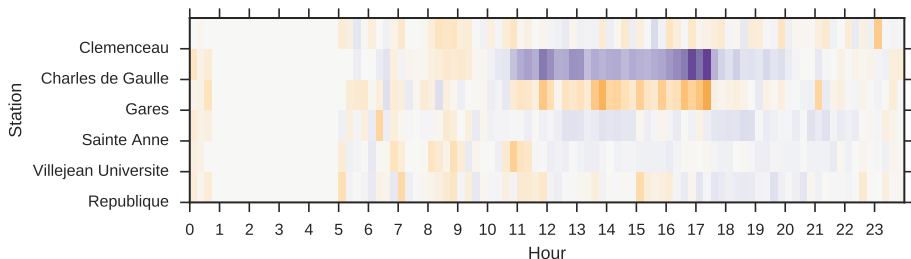


Short term passenger flow forecasting with Smart Card Data

Prévision court-terme pour l'exploitation



Event detection in a transit network



starbusmetro @starbusmetro · 26 mai 2016

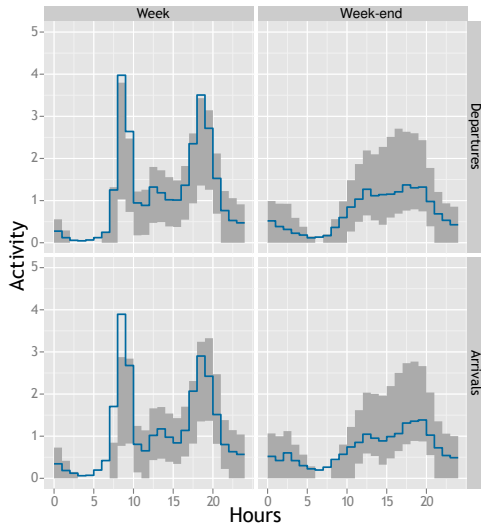
17h25-Station Charles de Gaulle de nouveau ouverte. Merci de votre patience
[#rennes](#) [#infotrafic](#)



7



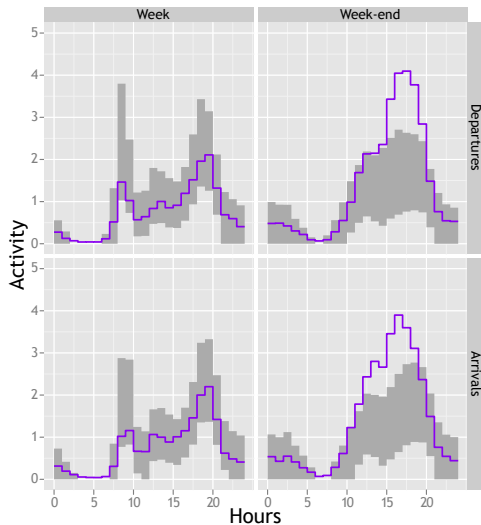
Vélib', Railway stations



Vélib', Railway stations



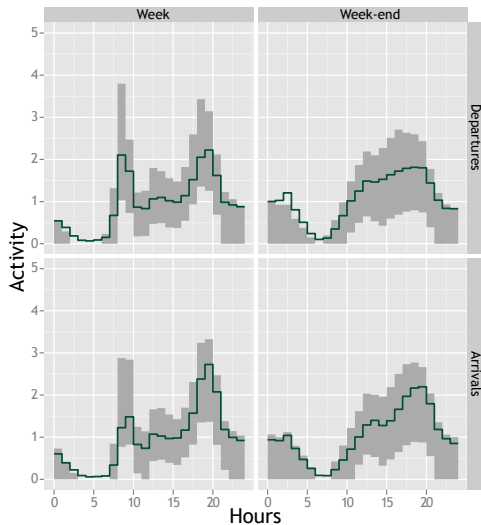
Vélib', Parks



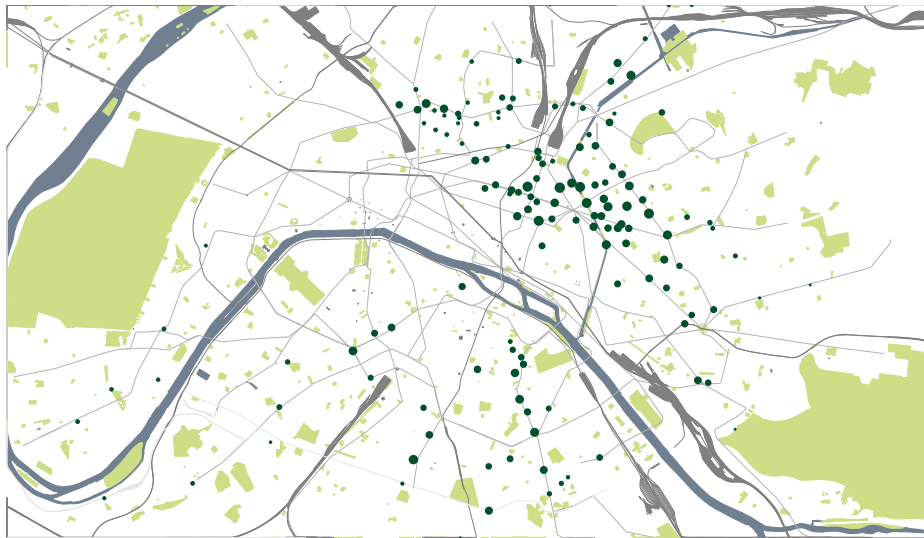
Vélib', Parks



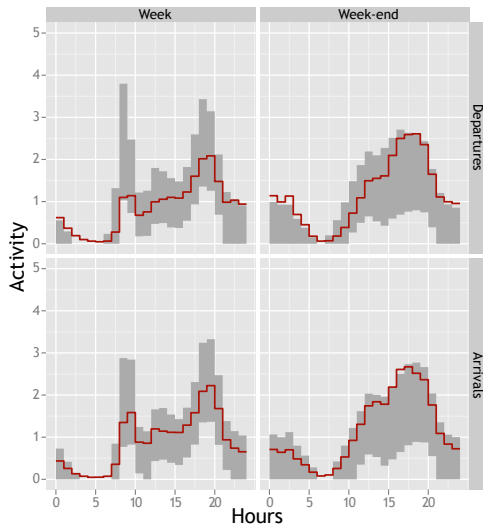
Vélib', Spare time, night



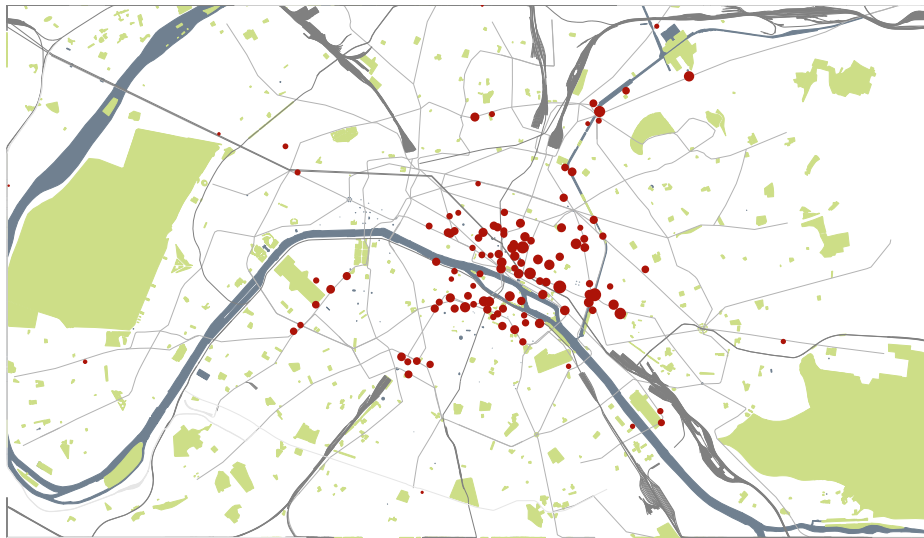
Vélib', Spare time, night



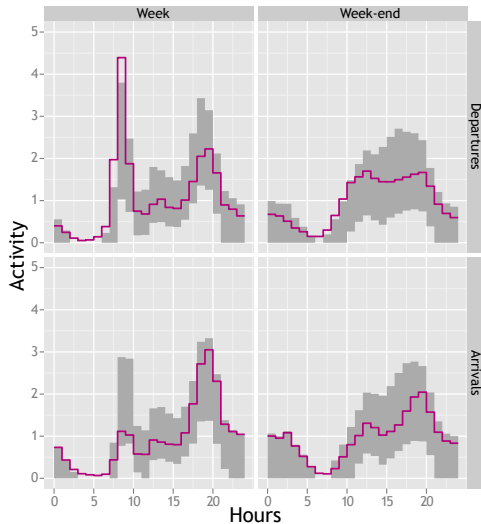
Vélib', Spare time, night and week-end



Vélib', Spare time, night and week-end

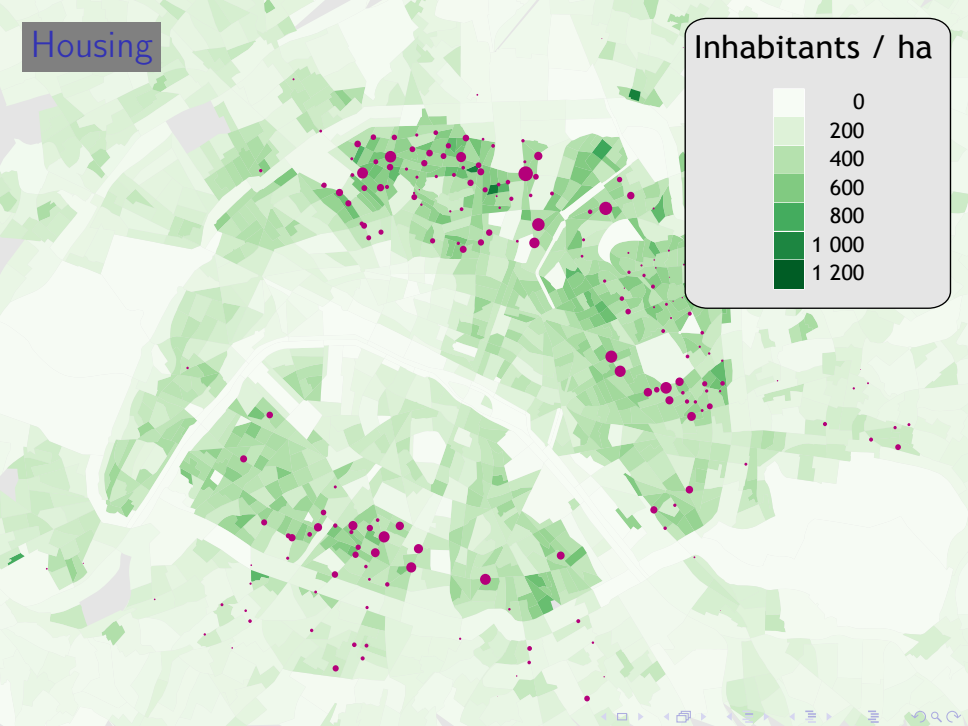
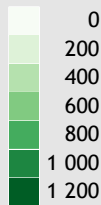


Vélib', Housing

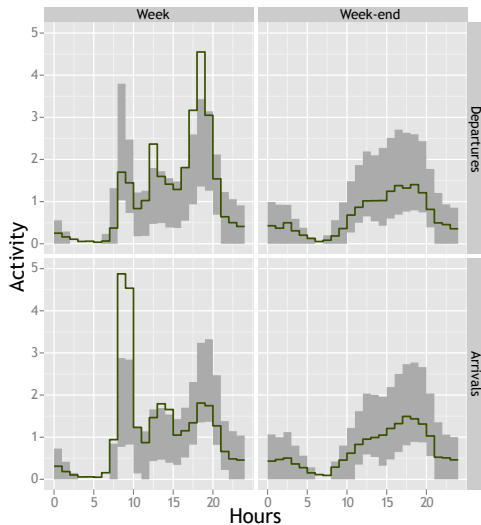


Housing

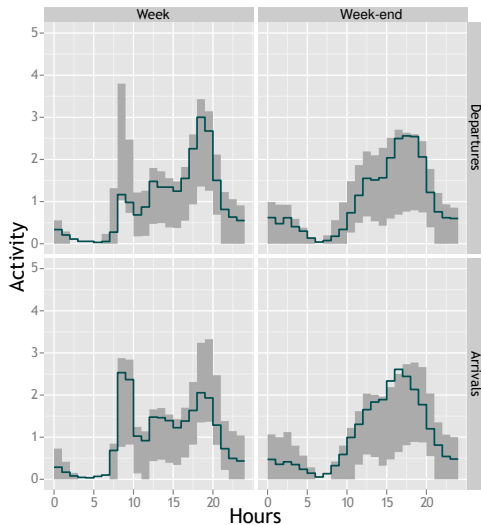
Inhabitants / ha



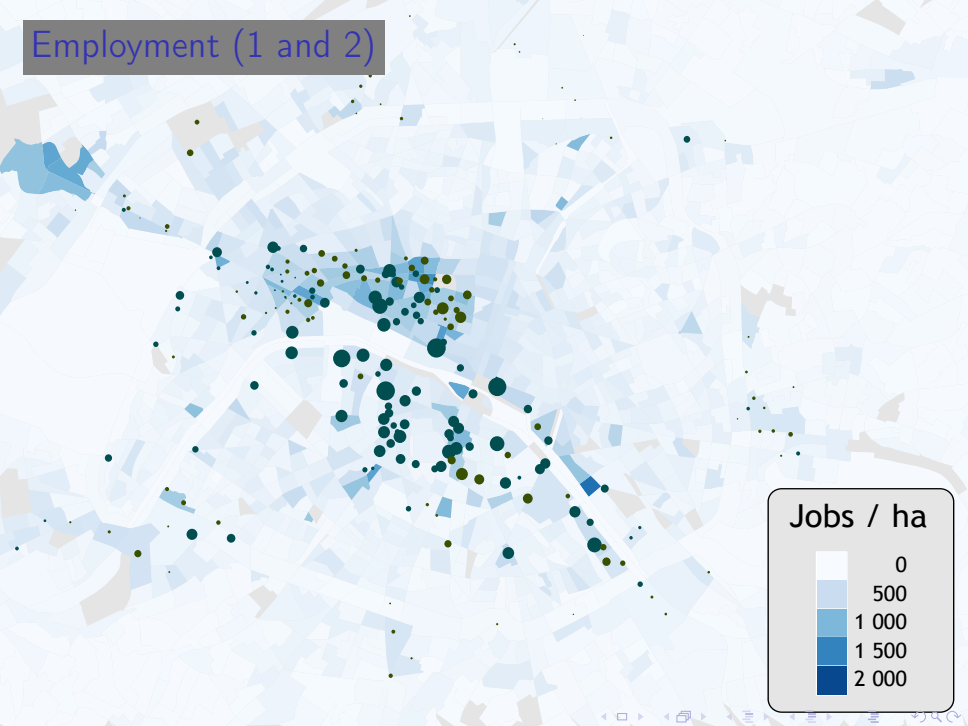
Vélib', Employment (1)



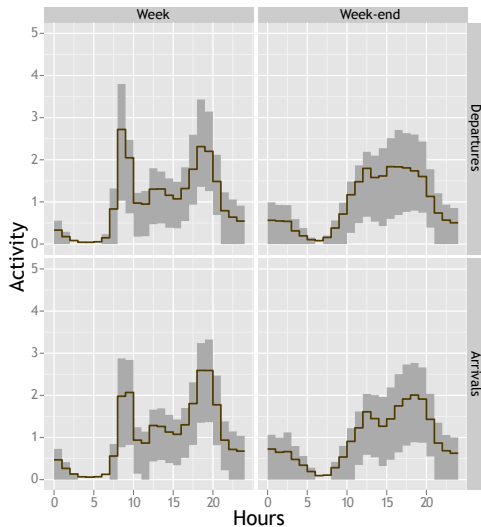
Vélib', Employment (2)



Employment (1 and 2)



Vélib', Mixed usage



Vélib', Mixed usage



Les "Big Data" pour la création de service à haute valeur ajoutée

- Outils de monitoring de systèmes urbains
- Croisement de plusieurs sources de données
- Traitement et visualisation avancés de données (Data Analytics et Data Vis), fonctionnement de systèmes urbains (analyse predictive, offre et demande)

Questions adjacentes :

- Protection vie privée, éthiques , création de services pour les citoyens et la communauté
- Accès données : Open Data, données plus riches