

Population & Sociétés

Analyse des flux de vélos publics à New York

Joséphine Boyer

Le réseau Citibike propose aux New-Yorkais des vélos en libre-service. Afin d'aider les gestionnaires du réseau, nous proposons ici une étude de l'utilisation des vélos par les usagers. Cela permettra aux techniciens de pouvoir intervenir plus efficacement sur le parc afin de garantir une fluidité dans la prise et la remise d'un vélo par les usagers.

Les données dont on dispose contiennent tous les trajets effectués au mois d'avril 2017 et proviennent du site internet <https://www.citybikenyc.com>. Les analyses ont été réalisées avec R.

1. Utilisation sur une journée

En premier lieu, nous étudions la variation de la quantité de vélos utilisés sur une journée. Pour cela, nous avons extrait des données la date de début et le jour de la semaine de chaque trajet. La date de début étant exprimée en format POSIX, il est possible d'en extraire l'heure du trajet. Il a ensuite suffi de

créer des histogrammes classant les trajets effectués par heure et par jour, visibles sur la figure 1 ci-dessous. Nous avons ajouté un histogramme (en vert) qui fait la somme de l'utilisation journalière des vélos, tous les jours de la semaine confondus.

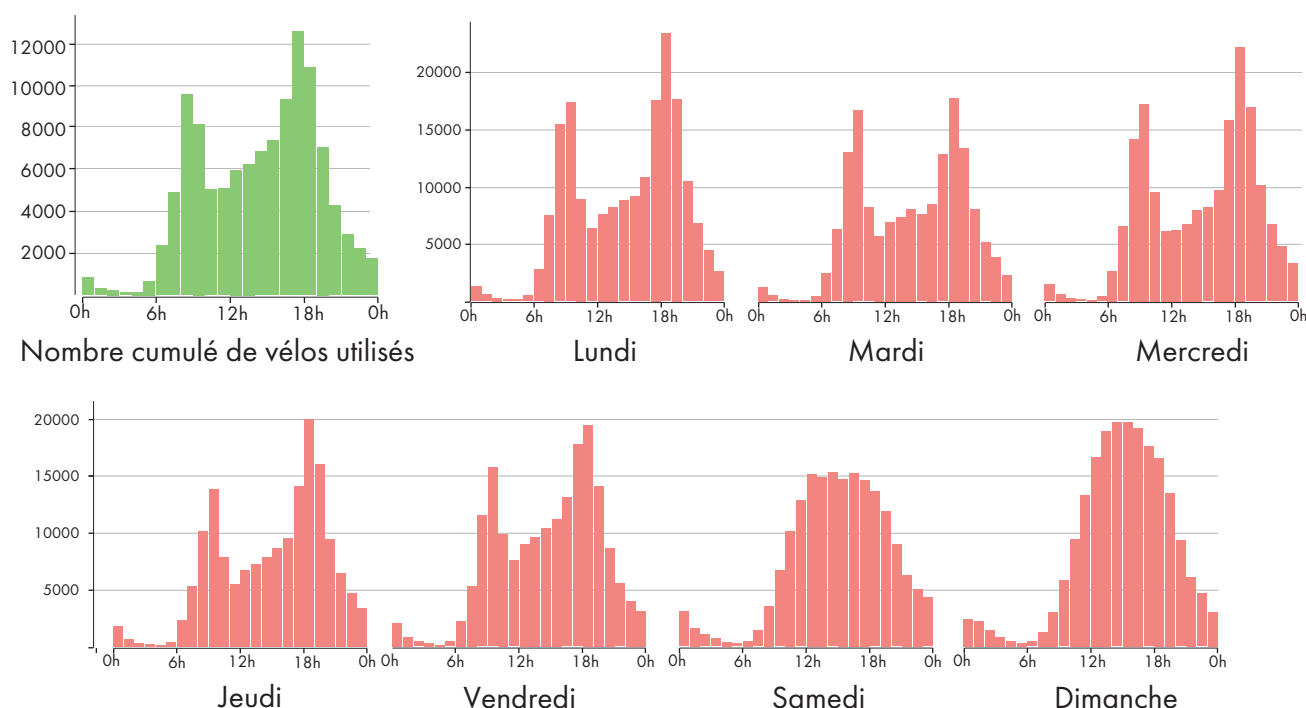


Figure 1 - Nombre de vélos en circulation pendant un mois selon l'heure de la journée

On constate nettement que les vélos sont majoritairement utilisés en semaine entre 7h et 9h et entre 17h et 19h. Durant le week-end, la plus forte utilisation a lieu entre 11h et 20h, avec un pic en milieu d'après-midi. Les vélos sont donc très sûrement utilisés pour les aller-retours travail-domicile durant la semaine, et pour les loisirs le week-end. Cependant, ce type d'utilisation doit varier selon le type d'abonnement, les abonnés doivent utiliser les vélos pour aller à leur travail, tandis que les utilisateurs occasionnels doivent les utiliser à des fins de loisir. Nous observerons cela dans la prochaine partie.

Globalement, il y a toujours au moins 150 ($=5000/30$) trajets effectués par heure en permanence de 7h à 21h tous les jours, avec des pics d'utilisation pouvant aller jusqu'à 650 ($=20\,000/30$) trajets effectués par heure.

Comme nous pouvions nous en douter, les vélos sont les moins utilisés durant la nuit (entre minuit et 6h), c'est le moment le plus opportun pour procéder aux réparations et remplacements qui pourraient diminuer la flotte pendant quelques heures. Le nombre de trajets effectués est minimal entre 4h et 5h tous les jours.

2. Profil des utilisateurs

Afin de mieux répondre aux besoins de la clientèle, il est important de la connaître. Nous proposons ci-dessous une analyse des profils des utilisateurs basée sur les données que nous avons à notre disposition. Ces données nous donnent le genre (variable **gender**) et le type d'abonnement (variable **usertype** : abonné ou client) du cycliste pour chaque trajet effectué.

Pour avoir la répartition globale des profils d'utilisateurs, nous avons sélectionné le genre et le type d'abonnement de tous les trajets à notre disposition, puis nous avons créé un mosaic plot (visible ci-dessous sur la figure 2a). Des diagrammes circulaires montrant pour l'un la proportion d'abonnés ou de clients et pour l'autre la proportion de chaque genre précisent l'analyse (figures 2b et 2d). Ils ont été réalisés à partir d'un

bar plot projeté en coordonnées polaires. Enfin, la figure 2d montre la proportion de chaque type d'utilisateur selon le genre, grâce à un bar plot dont le paramètre 'fill' correspond au champ **usertype**.

Il faut tout de même faire attention à l'approximation faite ici : on considère que tous les cyclistes utilisent un vélo à la même fréquence. En effet, les données concernent les trajets effectués et non les utilisateurs des vélos. Nous n'avons pas de champs qui renseignent l'id de l'utilisateur qui nous permettrait de compter le nombre d'abonnés, de clients, ou de femmes par exemple. Les proportions et chiffres donnés ci-après concernent l'occupation des vélos par les profils d'utilisateurs et non la répartition des profils parmi les utilisateurs enregistrés.

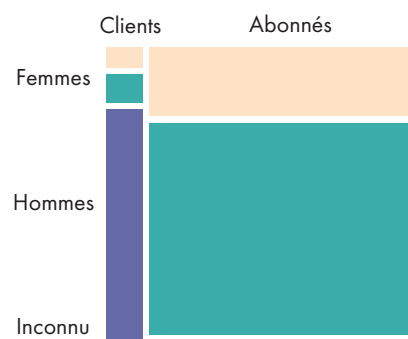


Figure 2a - Mosaic plot : proportions d'utilisateurs classées selon leur genre et leur type d'abonnement

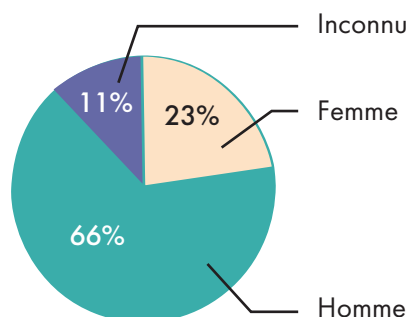


Figure 2b - Répartition des utilisateurs selon leur genre

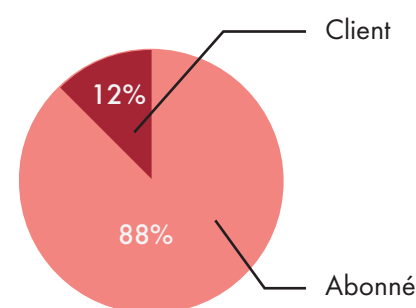


Figure 2c - Répartition des utilisateurs selon le type de leur abonnement

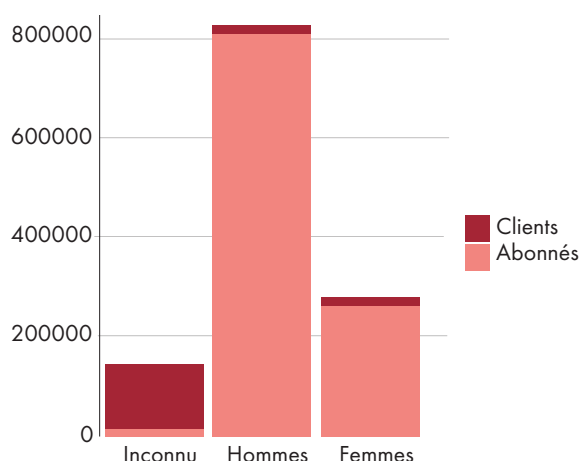


Figure 2d - Nombre de trajets effectués par genre et type d'abonnement

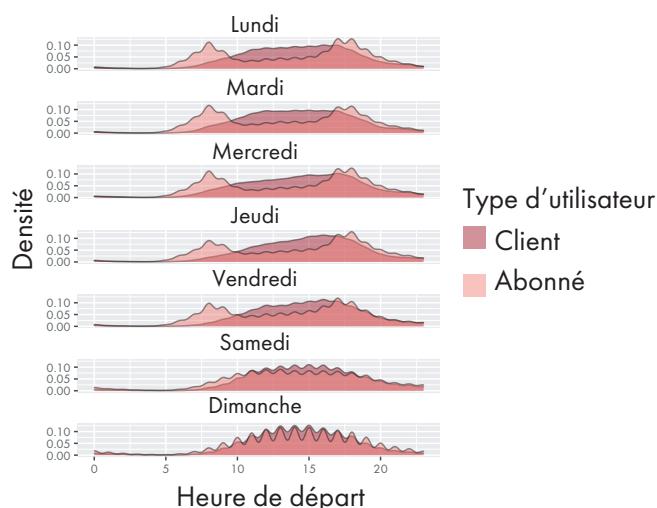


Figure 2e - Distribution horaire des trajets par type d'abonnement

L'étude conjointe des graphiques obtenus nous indique en premier lieu que la majorité des clients ne renseignent pas leur genre lors de leur souscription, car la majorité des cyclistes de genre inconnu sont des clients (figures 2a et 2c).

Ensuite, parmi les cyclistes dont on connaît le genre, la figure 2b montre clairement que la majorité des utilisateurs sont des hommes (66% des utilisateurs contre 23% de femmes). La figure 2c montre que presque 9 trajets sur 10 sont réalisés

dans le cadre d'un abonnement au service.

La figure 2d complète en donnant les proportions : les hommes réalisent à peu près trois fois plus de trajets que le sexe opposé.

Enfin, la figure 2e nous indique que le type d'abonnement de l'utilisateur affectera l'utilisation du service, car les abonnés utilisent le vélo durant des périodes très réduites le matin et le soir, ce qui en fait un enjeu pour les gestionnaires.

3. Comment combler les déséquilibres ?

Les utilisateurs utilisent les vélos pour se rendre à des endroits généralement fréquentés, et il est courant que les vélos s'accumulent si personne ne vient les déplacer aux endroits où ils manquent. Dans cette partie, nous proposons une étude spatiale qui devra permettre de mieux connaître les lieux de surcharge et de déficit, et à quel moment de la journée ils sont les plus forts.

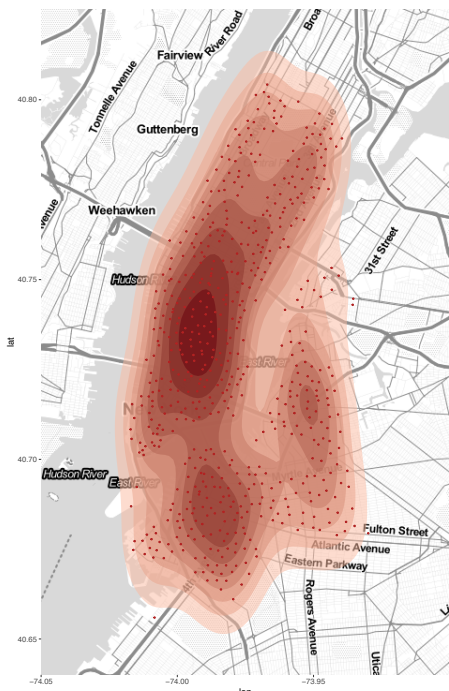


Figure 3a - Densité des stations CityBike à New York City

Nous avons premièrement réalisé une carte de densité, en représentant toutes les stations de vélos en libre service sur une carte. Ce sont les points rouges de la figure 3a. À partir de ce semis de points, nous avons ensuite utilisé la fonction `stat_density_2d` de la librairie `ggplot2` afin de calculer et faire figurer la densité des stations. Nous remarquons sur ce résultat de ce calcul (figure 3a) que les stations sont très concentrées au centre de Manhattan, et à Brooklyn dans une moindre mesure (en rouge foncé). Ce devrait donc être dans ces endroits-ci que les interventions auront le plus lieu.

Ensuite, nous voulions connaître les stations où le plus de vélos arrivent sans que les utilisateurs ne les reprennent, ou celles où tous les vélos sont tous utilisés sans que d'autres n'arrivent.

Pour cela, nous avons compté pour chaque station le nombre de vélos qui en partent et le nombre de vélos qui y arrivent sur tout le mois d'avril. Nous faisons ensuite la différence entre le nombre d'arrivées et le nombre de départs pour chaque station.

Nous obtenons les résultats visibles sur les figures 3b et 3c, nous avons choisi de ne représenter que les 50 stations les plus en déficit ou en surcharge. Les cercles verts indiquent que plus de vélos sont arrivés (la station est en surcharge), et ceux en rouge indiquent que la station est en déficit de vélos (plus de vélos sont arrivés à cette station qu'il ne sont repartis). La taille des cercles est proportionnelle au nombre de vélos en défaut ou en surplus. Nous avons choisi de séparer l'analyse des jours de la semaine de ceux du week-end car, comme vu dans la partie 1, les habitudes d'utilisation des vélos varient beaucoup entre ces deux périodes.

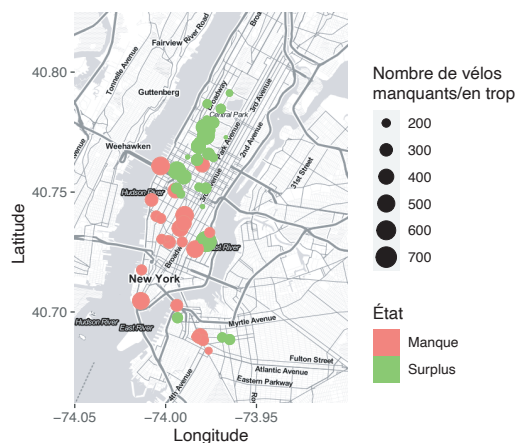


Figure 3b - Déséquilibres des stations en semaine

Pour les 50 stations les plus déséquilibrées

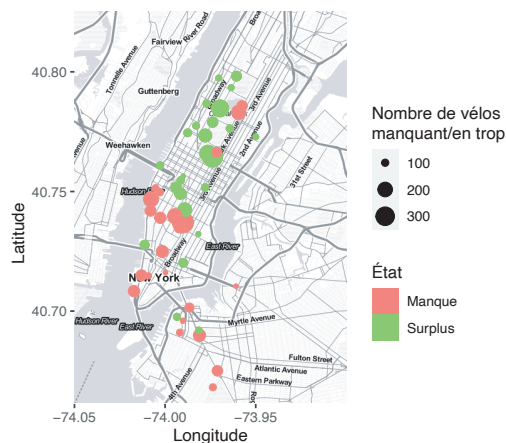


Figure 3c - Déséquilibres des stations le week-end

Pour les 50 stations les plus déséquilibrées

La première information qui ressort des figures 3b et 3c est que généralement, les stations sont déficitaires dans le centre ville de Manhattan, et ont trop de vélos au Sud de Central Park et dans l'Upper East Side. Ensuite, la situation représentée cumule tous les trajets du mois, et certaines stations cumulent au cours de ce mois jusqu'à 700 vélos

manquants ou en surcharge. Cette constatation nous indique combien le travail des gestionnaires est précieux pour garder toutes les stations utilisables par les New Yorkais, car cette simulation ne prend pas en compte leur travail. Si l'on prend en moyenne 300 vélos à déplacer pour une station sur un mois, cela représente 10 vélos tous les jours en moyenne.

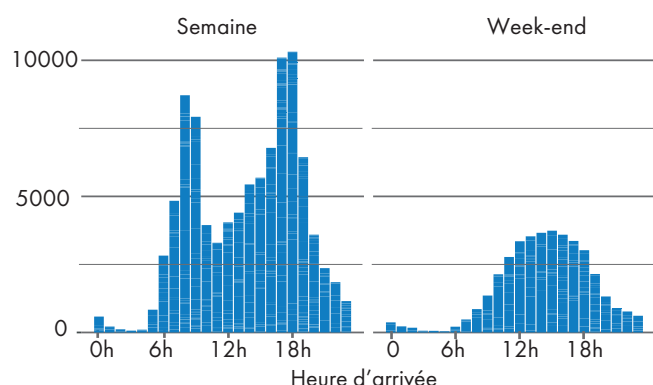


Figure 3d - Nombre de vélos remis dans les 20 stations les plus surchargées selon l'heure de la journée

Afin de déterminer les heures où les stations sont les plus surchargées, nous avons sélectionné les 20 stations les plus surchargées et les 20 les plus en déficit, et avons affiché leur activité sur une journée. La figure 3d montre ainsi pour les jours en semaine et en week-end séparément, le nombre de vélos remis aux stations. La figure 3e montre les vélos pris au cours de la journée dans les 20 stations qui sont les plus déficitaires. Les deux figures étant à la même échelle, on s'aperçoit que les stations déficitaires le sont plus que les stations surchargées : il

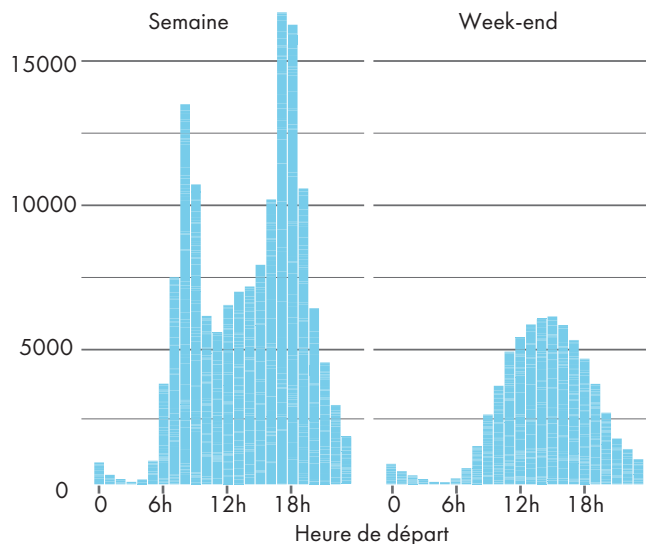


Figure 3e - Nombre de vélos pris dans les 20 stations les plus en déficit selon l'heure de la journée

manque plus de vélos qu'il n'y en a en trop. Cela est sûrement dû au fait que l'absence de vélo à une station empêche l'utilisateur d'utiliser le service, tandis qu'un surplus à une station mènera simplement l'usager à se diriger vers une autre station qui peut accueillir son vélo. Enfin, les stations sont sollicitées durant les heures de commutation domicile-travail en semaine et en milieu de journée le week-end, c'est à ces heures là qu'il faut intervenir pour réguler le stationnement des vélos en station.

4. Où et quand contrôler les vélos ?

Intéressons-nous désormais à la flotte de vélos. La fonction count nous indique qu'il y a **9 952** vélos dans la flotte, répartis entre **609** stations. Cette même fonction peut compter le nombre de fois que les vélos sont utilisés durant le mois* : 10 vélos ne sont utilisés qu'une seule fois, 13 le seront deux fois, et 7 le seront trois fois. La figure 4a nous indique l'identifiant des 20 vélos les plus utilisés, ce sont eux qui devront être contrôlés en priorité. Une analyse de ce tableau complet montre que les vélos dont l'identifiant est grand sont plus utilisés que les autres, sûrement car ils sont plus récents.

*via la formule suivante : count(bike, bikeid)

bikeid	nb_utilisations
28706	311
28757	294
25998	289
28397	288
28592	288
26361	282
28239	279
28286	279
26604	278
28887	278

bikeid	nb_utilisations
27832	277
25814	276
29412	276
27729	275
28215	273
28447	273
29005	273
25271	272
28420	272

Figure 4a - Identifiants des vélos les plus utilisés et nombre d'utilisations

Enfin, afin d'améliorer l'efficacité de la maintenance, une analyse de graphe nous permet de déterminer les 5 stations qui sont les plus centrales au réseau et qui verront passer le plus de vélos. Pour cela, on simule un graphe, dont les noeuds sont les stations et les arrêtes sont les trajets entre stations. Nous obtenons suite au calcul de betweenness (centralité d'intermédierité, ie nombre de plus courts chemins qui passent par cette station) la figure 4b, qui nous donne l'identifiant des stations où intervenir pour réparer un maximum de vélos, car elles ont plus de chance de croiser un grand nombre de trajets.

name	BETWEENNESS
282	1289.0000
3141	1204.0000
3435	1142.9949
460	1126.0000
3142	1102.0000
3112	1056.0000
501	1048.0000
3016	1048.0000
454	1004.9565
3243	972.0000

Figure 4b - Les 10 stations dont la betweenness est la plus forte

Conclusion : L'ensemble des analyses menées dans ce travail permettent de mieux définir les priorités d'intervention pour les services de gestion. Pour fluidifier le trafic des vélos, ils doivent répartir les vélos entre les stations juste avant les moments d'utilisation intense de ces vélos, pour ne pas avoir trop de vélos inutilisés lors de leur transport entre stations. Concernant la maintenance des vélos et des bornes de station, le meilleur créneau est pendant la nuit et le matin très tôt, car le service est alors peu utilisé.