

Projet BIXI - Partie 1: Analyse exploratoire

MATH60604 - Modélisation statistique

James Roy, Alfred Assal, Samuel Croteau, Abdoul Wassi Badirou

2023-09-21

Contents

1	Introduction	1
2	Exploration des données	1
2.1	Exploration univariée	2
2.2	Exploration multivariée	24
	Conclusion	34

1 Introduction

Dans le cadre du cours de modélisation statistique nous devons, à travers un projet, mettre en pratique les connaissances théoriques acquises. Le projet porte sur les activités de BIXI Montréal qui gère un système de vélopartage dans la grande région de Montréal. L'objectif du projet est d'identifier les facteurs qui influencent l'utilisation des Bixi, la durée des déplacements et les revenus qui en découlent. Le projet est divisé en 4 parties. La première porte sur l'analyse exploratoire et les autres consistent à explorer différents modèles à savoir: des modèles de régression linéaire, modèles linéaires généralisés et des modèles linéaires mixtes. Un jeu de données est mis à notre disposition. Il est important de rappeler que les données utilisées sont un sous-ensemble purement aléatoire des données complètes, coligées par l'entreprise.

La présente analyse porte sur la première partie du projet: faire une analyse exploratoire des données pour mieux comprendre le contexte et modèle d'affaires.

Plan Pour y arriver, nous analysons d'abord les variables individuellement. Ensuite à travers une analyse mutivairée, nous explorons les relations entre les différentes variables par rapport aux variables cibles. Les faits saillants de nos résultats sont ensuite résumés dans la conclusion.

2 Exploration des données

La table 1 nous montre à titre informatif les 6 premières observations du jeu de données. Nous y retrouvons toutes les variables décrites dans l'énoncé.

La visualisation de la structure compacte des données nous montre que le jeu de données est constitué de 10000 observations et 14 variables.

Table 1: 6 premières observations du jeu de données

station	mm	dd	wday	mem	holiday	dur	avg	rev	n_AM	n_PM	n_tot	temp
10	4	24	Saturday	1	0	917.88333	29.60914	NA	3	28	31	14.1
10	5	24	Monday	1	1	1123.41667	26.12597	NA	2	39	43	14.4
10	6	3	Thursday	1	0	115.03333	16.43333	NA	2	3	7	17.5
10	6	18	Friday	0	0	56.68333	28.34167	11.0025	1	1	2	21.2
10	7	21	Wednesday	1	0	285.68333	16.80490	NA	5	10	17	20.0
10	7	22	Thursday	1	0	382.48333	18.21349	NA	3	15	21	19.6

2.1 Exploration univariée

2.1.1 Les variables qualitatives

Rajouter ici le tableau sommaire des variables qualitatives

2.1.1.1 Station La variable station est une variable qualitative qui représente les différentes stations de départ des trajets en BIXI. Nous l'avons encodé sous forme de variable catégorielle nominale car il n'y a pas de relation d'ordre entre les différentes stations. Le jeu de données couvre 792 stations Bixi.

La figure 1 montre le nombre de stations en fonction du nombre de leur observations dans les données. On constate qu'il n'y a pas le même nombre d'observations pour toutes les stations. Donc pour certaines stations, il y a des observations manquantes, soit pour une journée, soit pour un mois, soit pour une catégorie de membre. Cela est due au fait que le jeu de donnée que nous utilisons est un sous-ensemble aléatoire des données réelles.

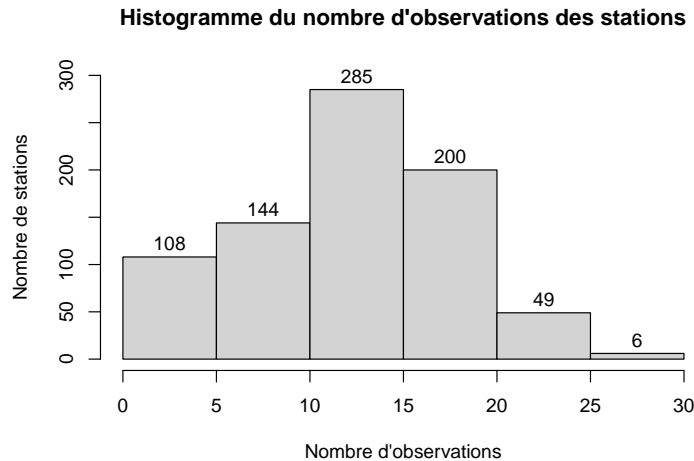


Figure 1: Histogramme du nombre d'observations des stations

La figure 2 nous montre que pour une station, nous avons entre 2 et 8 mois dans l'année pour lesquels nous avons des observations. Est-ce que l'absence d'observation pour les autres mois de l'année pour chaque station est due à la méthode de génération des sous-ensembles des données ou simplement au fait qu'il n'y a aucune activité à ces période.

2.1.1.2 Mois Il s'agit d'une variable catégorielle ordonnée car les mois arrivent dans un ordre donné. La saison Bixi s'étend d'avril (4) à novembre (11) comme on peut le voir à la figure 3. En dehors des mois

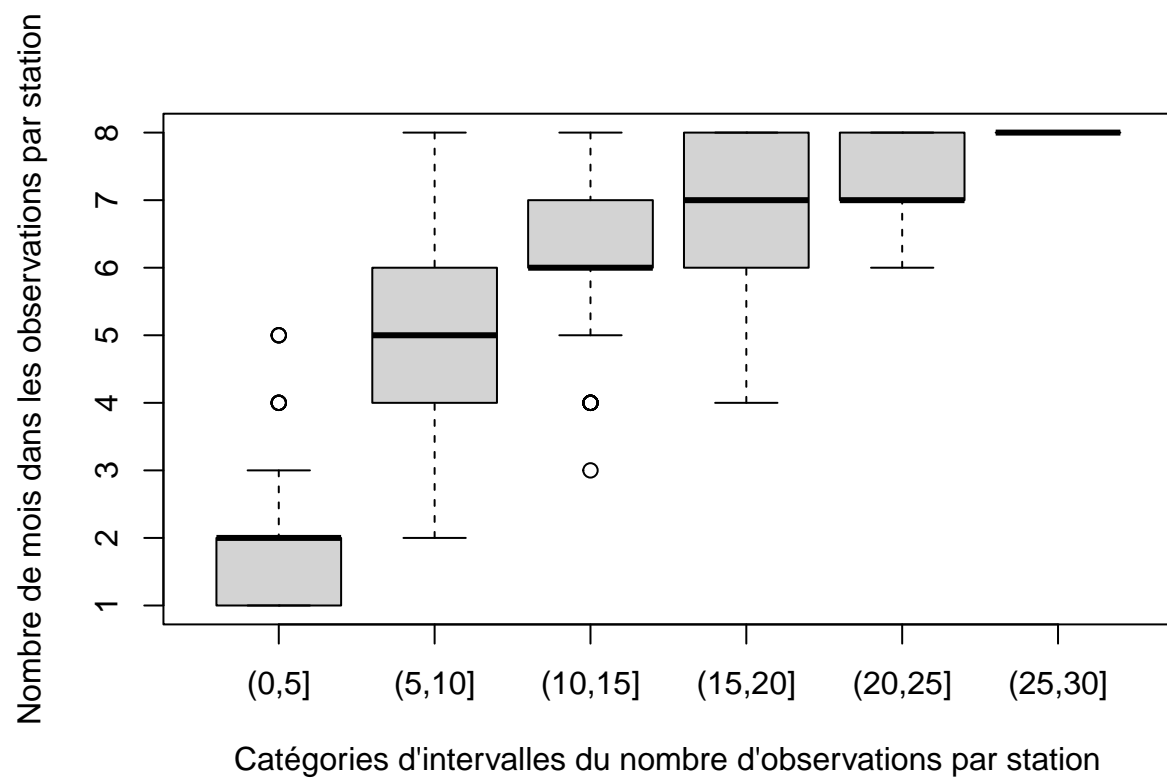


Figure 2: Boîte à moustache du nombre de mois dans les observations par station en fonction des différentes catégories de nombre d'observations par station

d'avril et novembre, nous avons approximativement le même nombre d'observations par mois. Il y a environ 2 fois moins d'observations pour les mois d'avril et novembre. C'est cohérent avec le fait que la saison s'étend de la mi-avril à la mi-novembre. De ce point de vue, notre sous-ensemble de jeu de donnée est représentatif de la saison réelle.

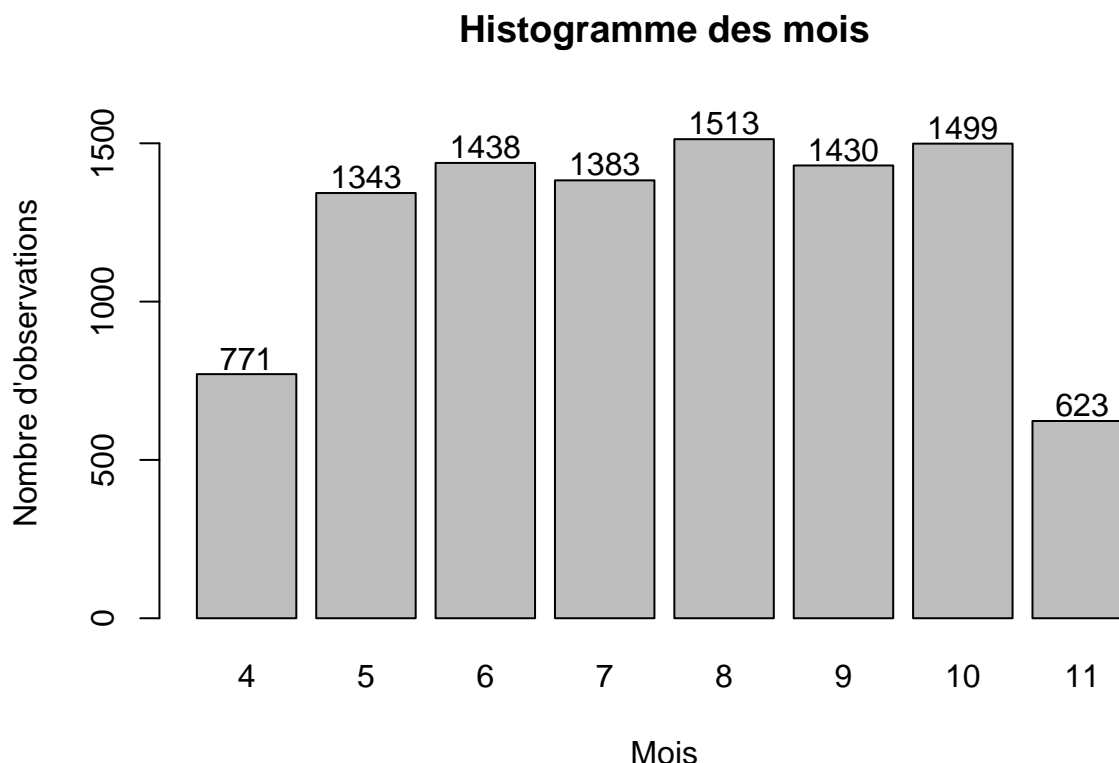


Figure 3: Histogramme des mois

La figure 4 nous montre que les mois d'avril et novembre (avril plus que novembre) sont ceux où il y a le moins de revenu, les bixi sont les moins utilisés et les utilisateurs font de plus courtes distances. Les mois d'août et septembre génèrent le plus de revenu. L'utilisation des bixi et la durée des trajets est plus importante de juin à septembre. En gros, les périodes creuses de l'année sont avril et novembre (surtout avril). Et les périodes pointent de l'année sont entre juin et septembre (surtout septembre).

Étant donné que le revenu des membres est manquant, faisons-nous le même constat pour les non-membres uniquement? La figure 5 nous montre que le constat est identique et plus évident. Par ailleurs on observe l'absence des deux valeurs les plus extrêmes pour le nombre de total de déplacement en septembre. Cela suggère donc que le plus grand nombre de déplacement journalier a été fait par des usagers membres en septembre. Mais le plus long déplacement a été fait par des usagers non-membres en août. On se demande donc si les usagers membres ont plus tendance à effectuer de nombreux trajets mais de courte distance et les usagers non-membres moins de trajet mais plus long. Nous y répondrons dans la section 2.1.1.5.

2.1.1.3 Jour Il s'agit d'une variable catégorielle ordonnée. Les jours s'étendent du 1 au 31 en fonction des mois. Dans nos données, le premier jour de la saison est le 9 avril et le dernier jour de la saison est le 15 novembre. Or sur le site de Bixi, la saison commence le 15 avril. Il nous manque donc dans nos données la première semaine réelle de la saison.

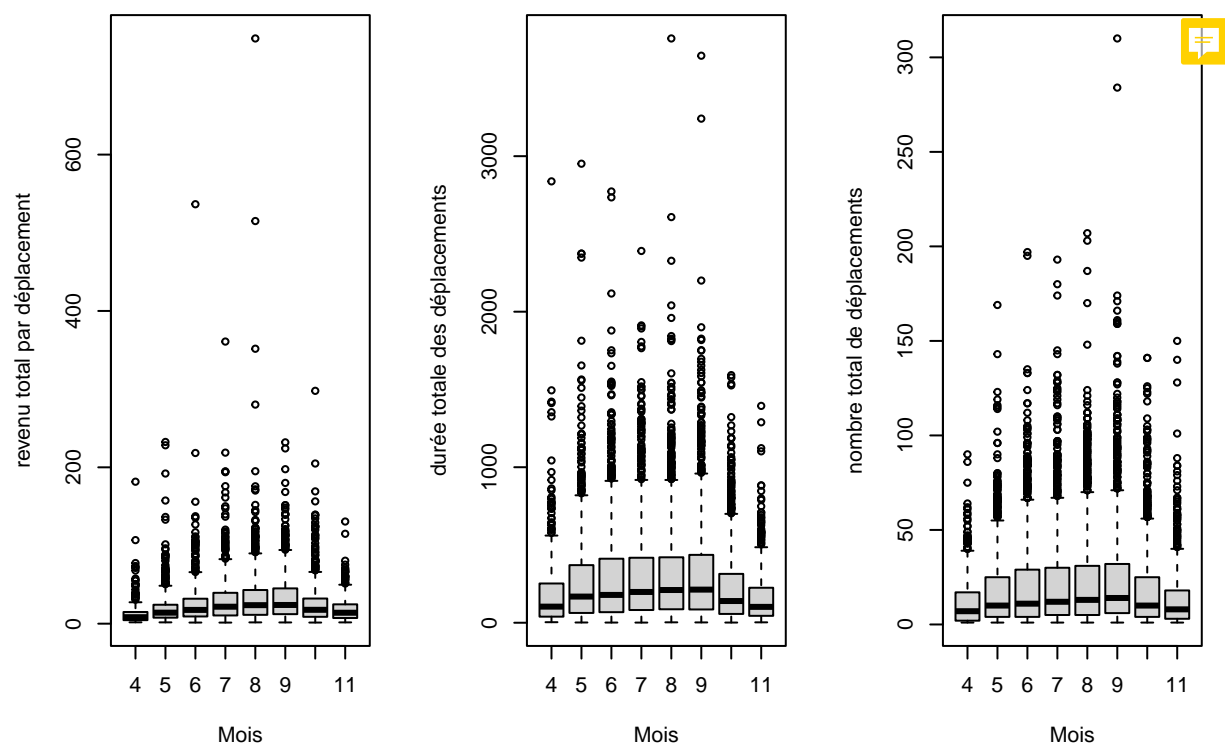


Figure 4: Boîtes à moustaches des revenus, durée et nombre totale de déplacement en fonction du mois

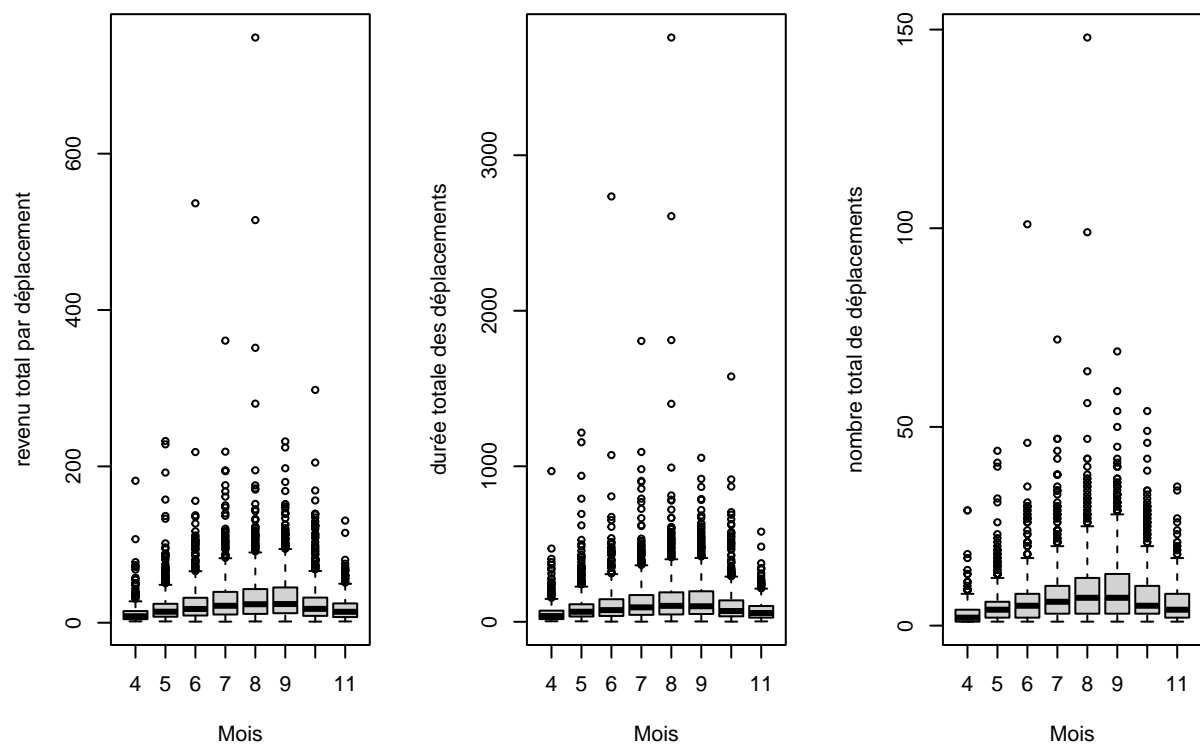


Figure 5: Boîtes à moustaches des revenus, durée et nombre totale de déplacement en fonction du mois pour les non-membre uniquement

Il y a environ 2 fois moins d'observations pour la journée du 31 par rapport aux autres mois. La distribution des autres jours semble approximativement uniforme.

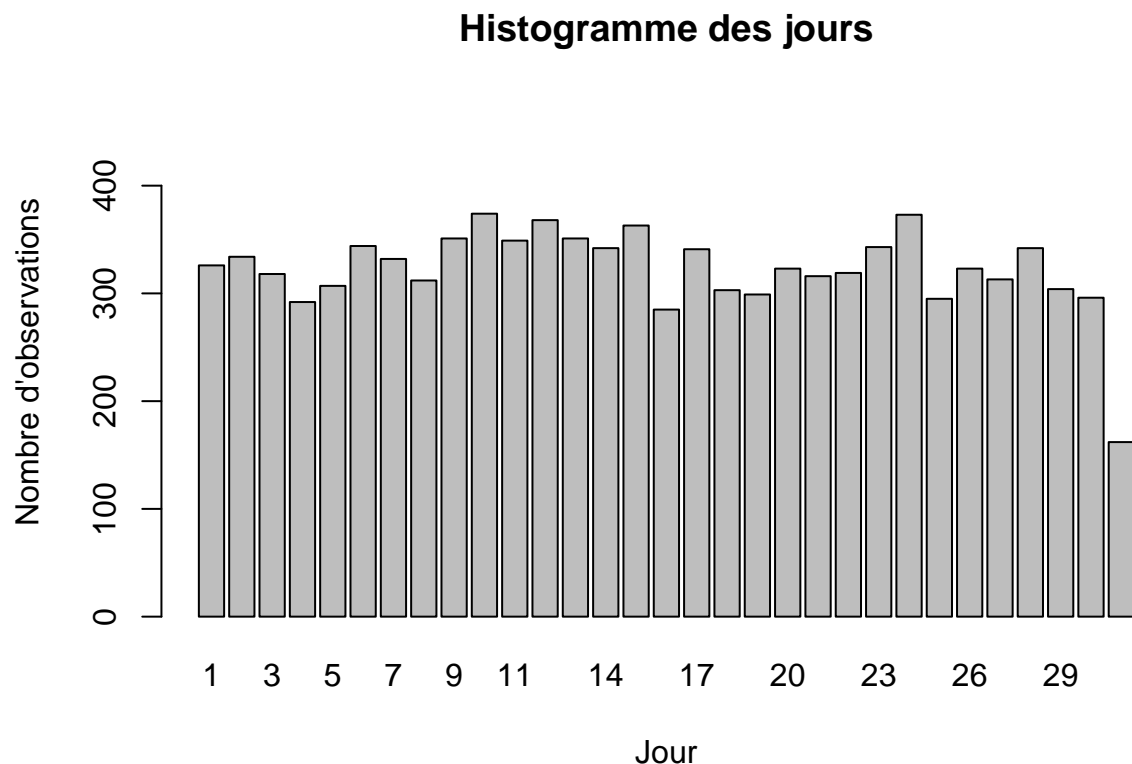


Figure 6: Histogramme des jours

Les figures 7 et 8 nous montrent que la journée du 24, une journée dans la dernière semaine de chaque mois, semble être la plus lucrative en revenu et la plus achalandée, aussi bien en terme de nombre de déplacement que de durée.

2.1.1.4 Jour de semaine Il s'agit d'une variable catégorielle ordonnée. Mais le choix du premier jour est relatif. La figure 9 nous montre que la distribution des jours de semaine est approximativement uniforme car nous avons environ le même nombre d'observations pour les différents jours de la semaine.

La figure 10 nous montre que les revenus générés sont plus importants en fin de semaine (Samedi et dimanche). Les déplacements d'une journée qui ont généré le plus de revenu sont également ceux qui ont été le plus long. Par contre, ils ne sont pas ceux qui ont totalisé le plus grand nombre de déplacement. On peut donc se demander si les déplacements les plus lucratifs sont les longs. La figure 11 confirme cette évidence. En effet, étant donné que la tarification est liée à la durée d'utilisation du bixi pour les non-membres, on peut s'attendre à avoir une forte corrélation entre le revenu et la durée des déplacements. On observe également qu'il y a un plus grand nombre de déplacement du mardi au vendredi. Les déplacements les plus importants ont eu lieu le jeudi. En fait, comme remarqué aux sections 2.1.1.2 et 2.1.1.3, il s'agit des déplacements des jeudis 2 et 9 septembre.

2.1.1.5 Membres Est-ce que les usagers membres ont plus tendance à effectuer de nombreux trajets mais de courte distance et les usagers non-membres, moins de trajet mais plus long? NON

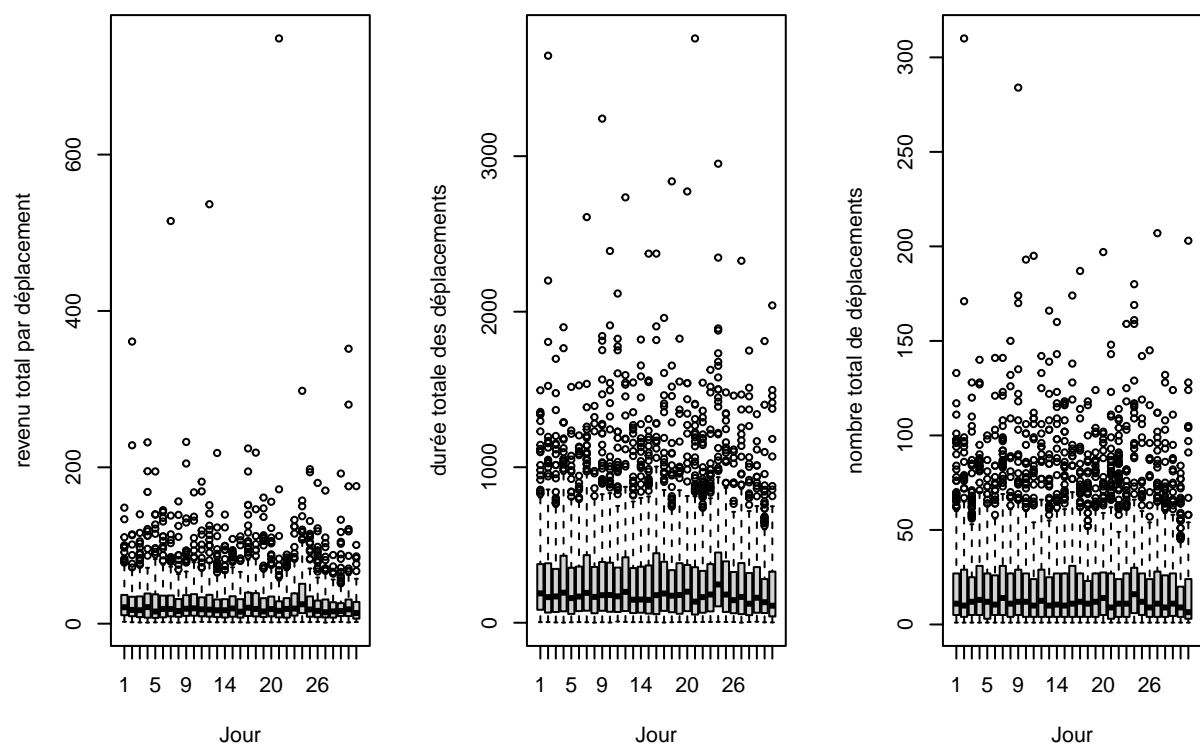


Figure 7: Boîtes à moustaches des revenus, durée et nombre totale de déplacement en fonction du jour.

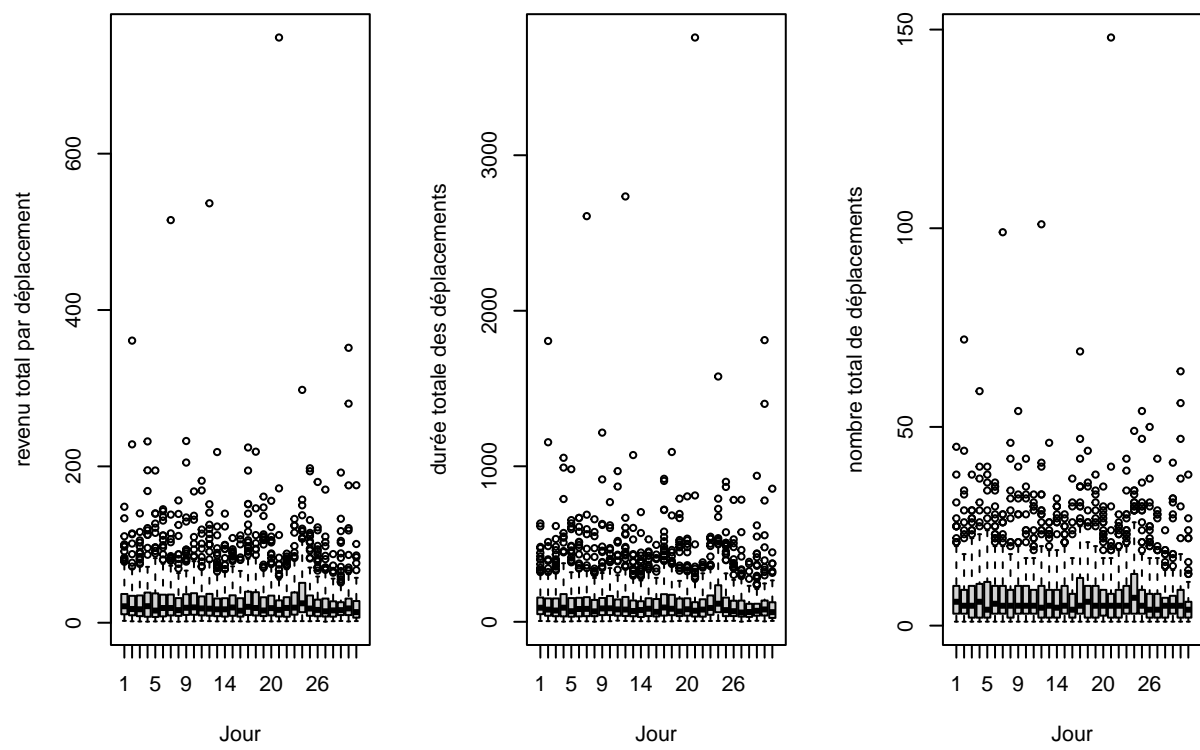


Figure 8: Boîtes à moustaches des revenus, durée et nombre totale de déplacement en fonction du jour pour les non-membre uniquement.

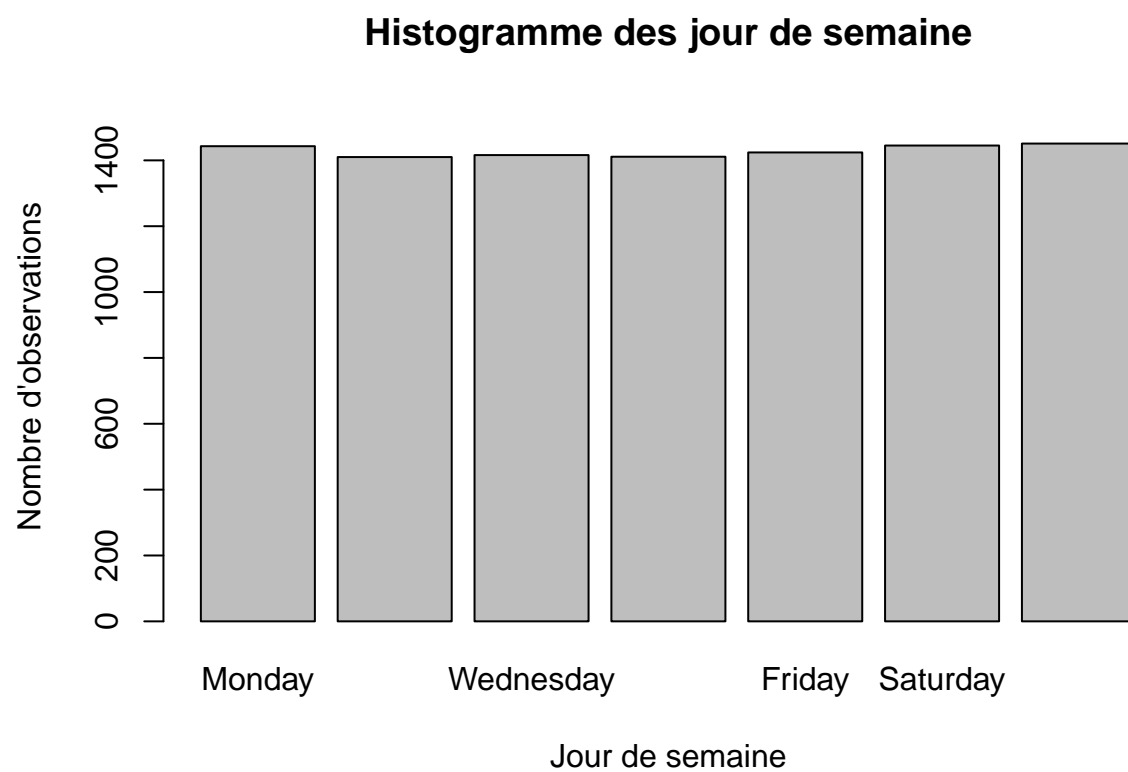


Figure 9: Histogramme des jour de semaine

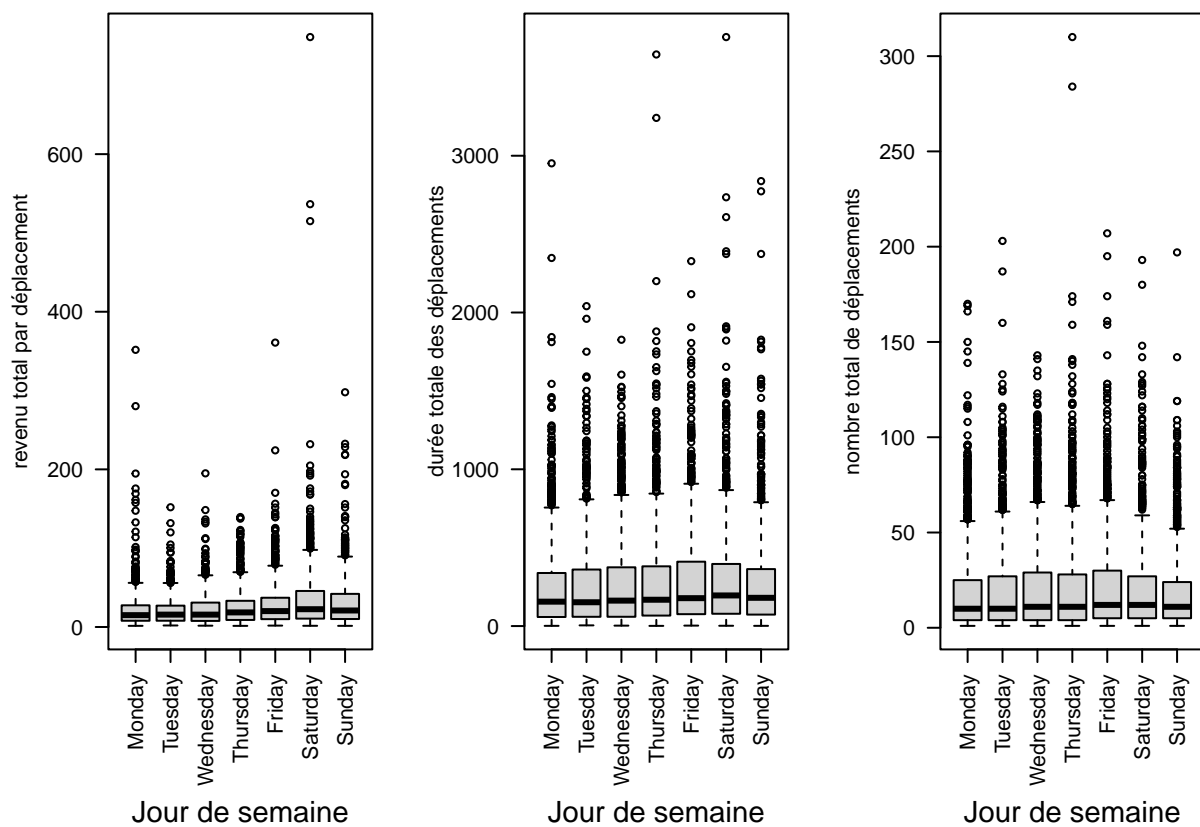


Figure 10: Boîtes à moustaches des revenus, durée et nombre totale de déplacement en fonction des jours de semaine.

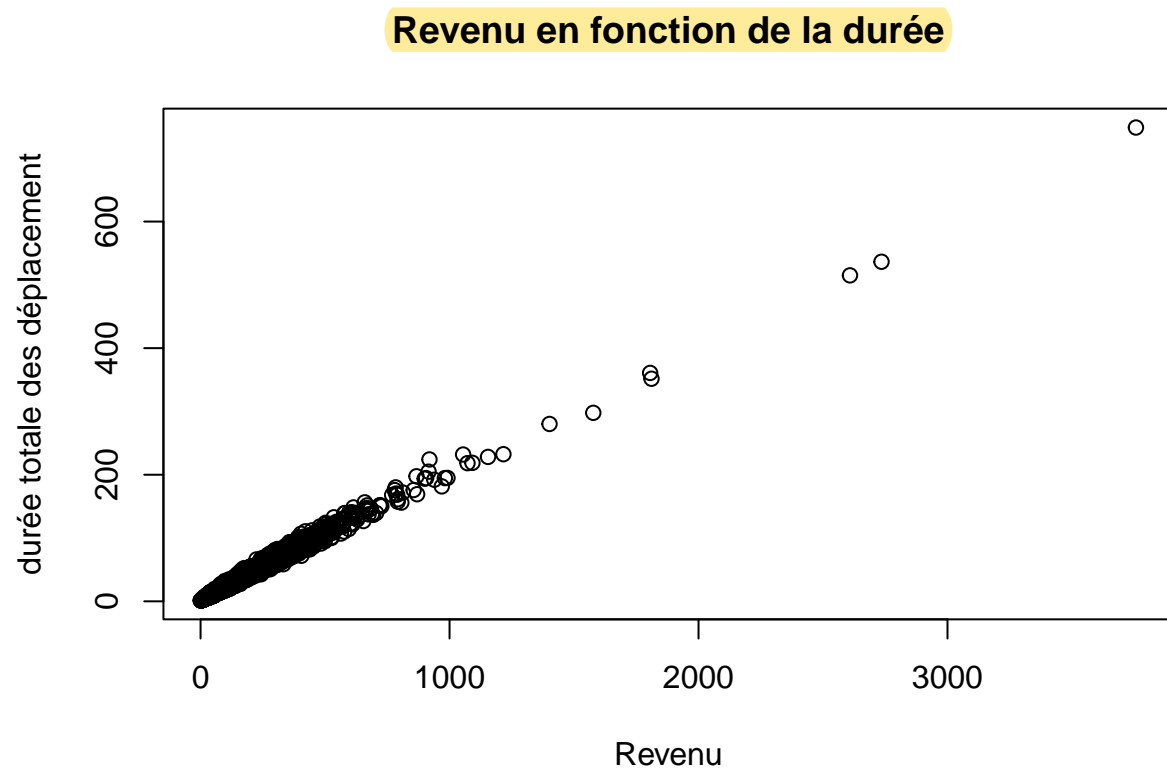


Figure 11: Revenu en fonction de la durée totale des déplacements

Il s'agit d'une variable catégorielle nominale. La figure 12 nous montre qu'il y a légèrement un peu plus de membre que de non-membre.

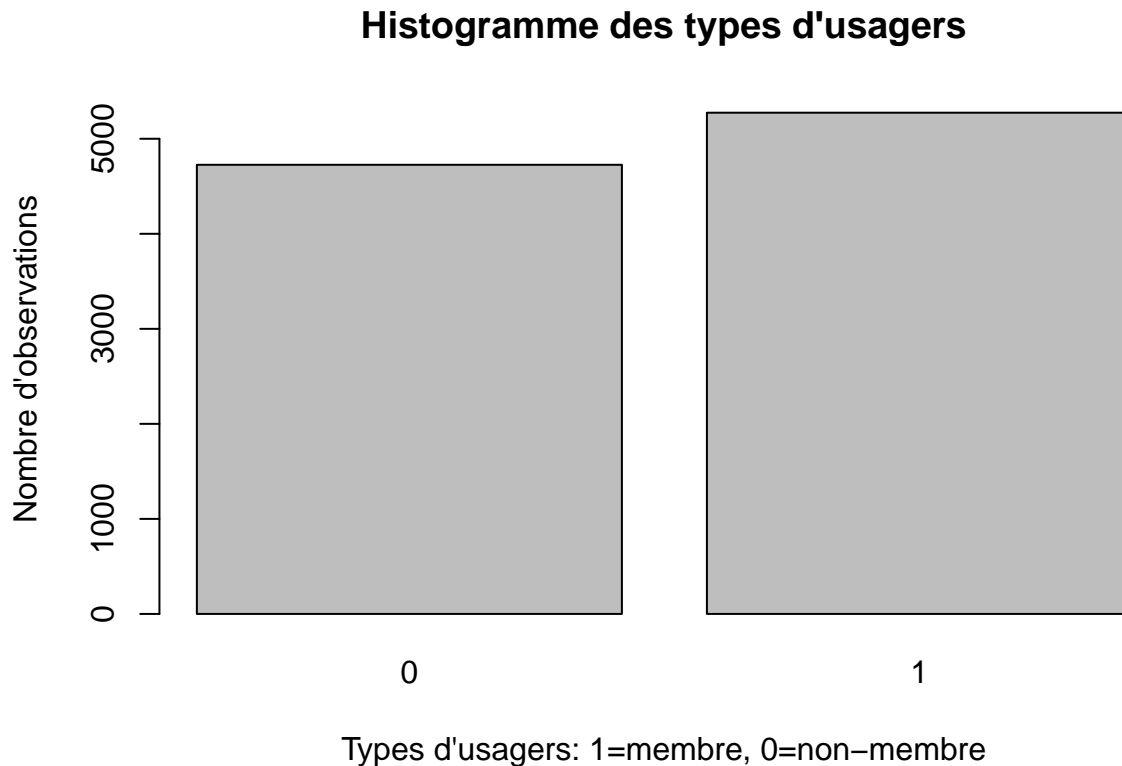


Figure 12: Histogramme des types d'utilisateur

La figure 13 nous montre que les usagers membres ont tendance à utiliser plus souvent et plus longtemps les bixi que les non-membre.

2.1.1.6 Jour férié Les observations sur des jours fériés représentent environ 2%. Nous devons plus tard y accorder une attention particulière dans la modélisation afin d'éviter que cette catégorie soit négligée. En effet, la figure 14 nous montre que les revenus générés sont un peu plus importants les jours fériés que les autres jours. La durée des déplacements est aussi un peu plus importante les jours fériés que les autres jours. Mais il n'y a pas une différence remarquable pour le nombre de déplacements. Probablement que ce comportement reflète des usages différents. Les jours fériés les usagers utilisent peut-être les bixi pour des promenades contrairement aux autres jours où ils s'en servent pour aller à une destination précise.

2.1.2 Les variables quantitatives

Le tableau 2 présente un sommaire descriptif des variables qualitatives. En moyenne la durée totale des trajets qui partent d'une station par jour est d'environ 5h. Par ailleurs, il y a trois journées qui totalisent des déplacements supérieurs à 50h. Deux d'entre eux sont associés à la station 256, les jeudi 2 et 9 septembre. L'autre est associé à la station 470, le samedi 21 août. Ces durées extrêmes sont tout de même surprenantes. Celles associées à la station 256 paraissent moins aberrantes tout de même, puisque le nombre de déplacements de ces journées est aussi extrême tant en AM qu'en PM. Cela suggère qu'il y a probablement eu ces deux jeudis consécutifs des événements dans le secteur qui pourraient expliquer ces utilisations extrêmes. De plus,

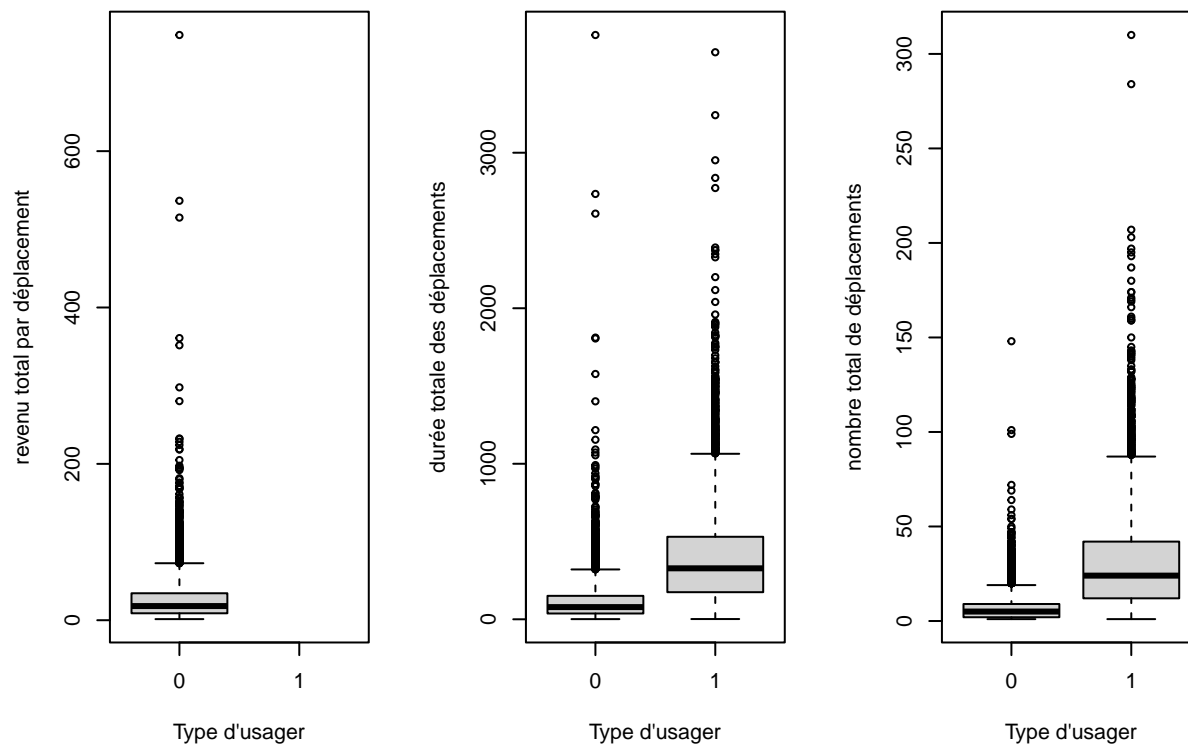


Figure 13: Boîtes à moustaches des revenus, durée et nombre totale de déplacement en fonction du type d'utilisateur.

Table 2: Sommaire descriptives des variables qualitatives

	dur	avg	rev	n_AM	n_PM	n_tot	temp
	Min. : 1.117	Min. : 1.117	Min. : 1.418	Min. : 0.000	Min. : 0.00	Min. : 1.00	Min. : 0.00
	1st Qu.: 66.092	1st Qu.:11.285	1st Qu.: 8.883	1st Qu.: 1.000	1st Qu.: 2.00	1st Qu.: 4.00	1st Qu.:1.00
	Median : 169.875	Median :13.839	Median : 18.111	Median : 2.000	Median : 6.00	Median : 11.00	Median : 1.00
	Mean : 267.795	Mean :15.404	Mean : 27.337	Mean : 4.309	Mean : 11.42	Mean : 19.72	Mean :16.00
	3rd Qu.: 375.179	3rd Qu.:17.717	3rd Qu.: 34.491	3rd Qu.: 6.000	3rd Qu.: 16.00	3rd Qu.: 27.00	3rd Qu.:2.00
	Max. :3756.900	Max. :58.817	Max. :748.535	Max. :63.000	Max. :165.00	Max. :310.00	Max. :28.00
			NA's :5274				

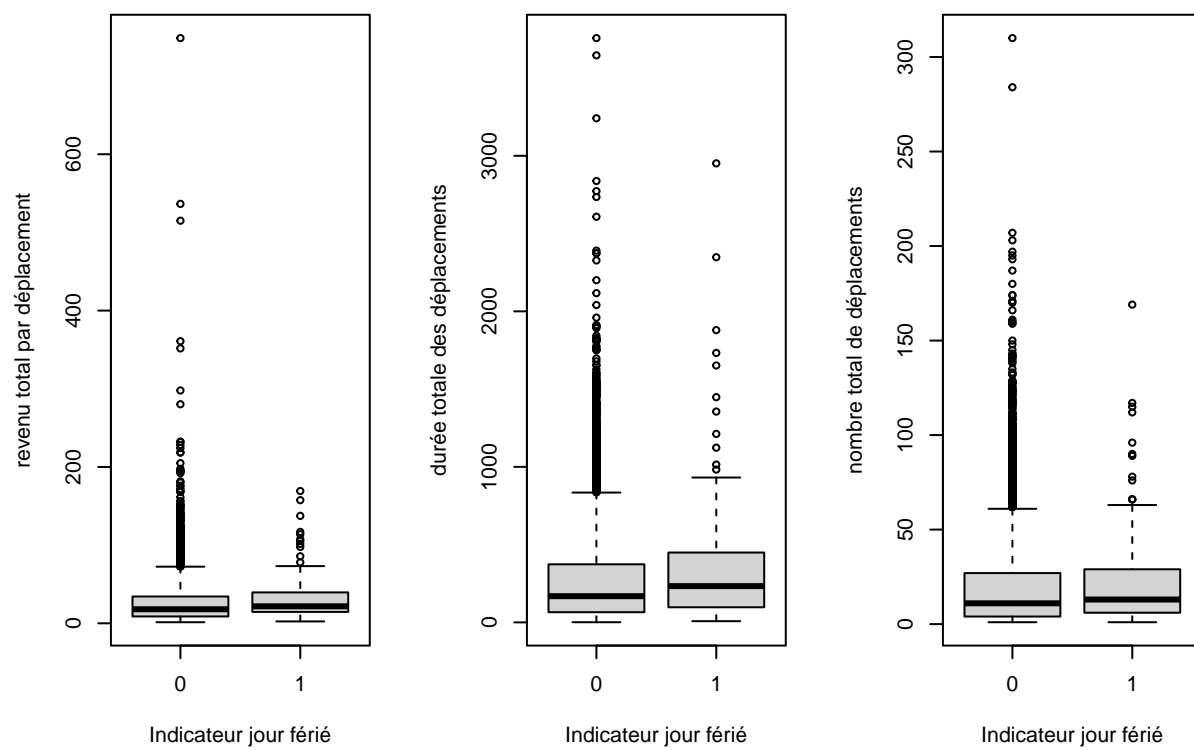


Figure 14: Boîtes à moustaches des revenus, durée et nombre totale de déplacement en fonction des jours fériés.

ils sont associés à des usagers membres. Cependant, la durée totale associée à la station 470 est la plus extrême et la plus surprenante. En effet, le nombre de déplacement associé n'est pas le plus extrême en plus c'est associé à des usagers non-membres qui ont un profil de consommation de courte durée. Ce déplacement peut s'expliquer si un utilisateur non-membre prend un bixi et le conserve chez lui plutôt que de le déposer à une station.

Toutes les variables quantitatives ont des valeurs extrêmes sauf la température moyenne quotidienne. Elle présente également une distribution quasiment symétrique. Les autres ont une distribution qui s'étale à droite. Une transformation sera nécessaire éventuellement dans la modélisation pour les rendre symétrique.

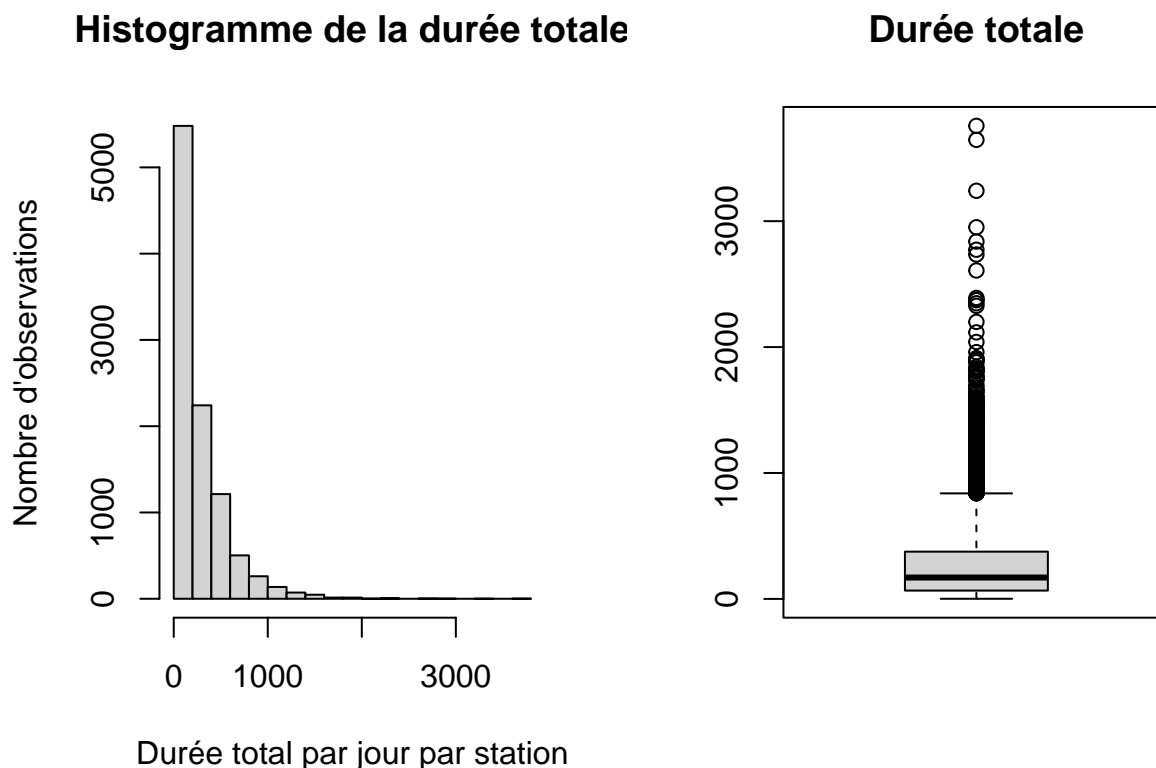


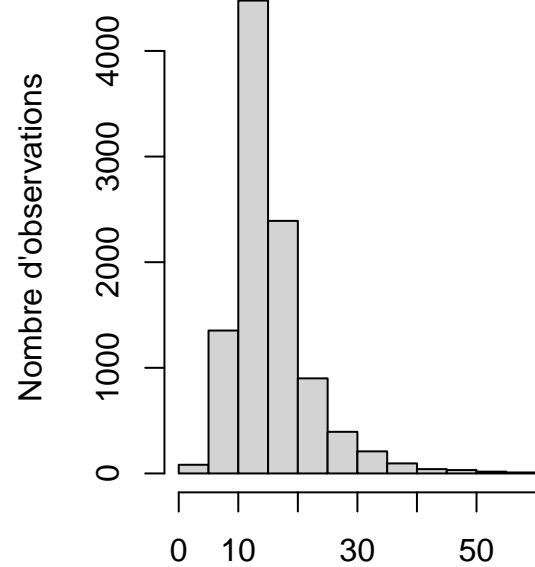
Figure 15: Histogramme et boîte à moustache de la durée totale

2.2 Exploration multivariée

Avant de débiter l'exploration multivariée des variables cibles, il est important d'avoir une vue d'ensemble des relations entre les variables. Pour ce faire, une matrice de corrélation comme illustrer dans la figure 23 est une représentation simple qui permet de visualiser les tendances entre chaque variable. Par exemple, il est possible de voir que la corrélation entre **duret** **rev** est de 0.99. De plus, les relations de **duret** de **rev** vont se comporter de façon quasi identique avec les autres variables. Ainsi, il est possible d'en représenter qu'une seule des deux lors de l'analyse multivariée. Cette problématique sera abordée plus en détail dans la section 2.2.3.

En ce qui concerne la matrice de corrélation, deux variables ont été enlevées à des fins pratiques. La première est **wday** (les journées de la semaine) et la deuxième est **mem** (membre de bixi ou non). Par contre, il est logique d'affirmer par déduction que ces deux variables restent tout de même intéressantes à croiser avec nos variables d'intérêt.

Histogramme de la durée moyenne



Durée moyenne par jour par station

Durée moyenne

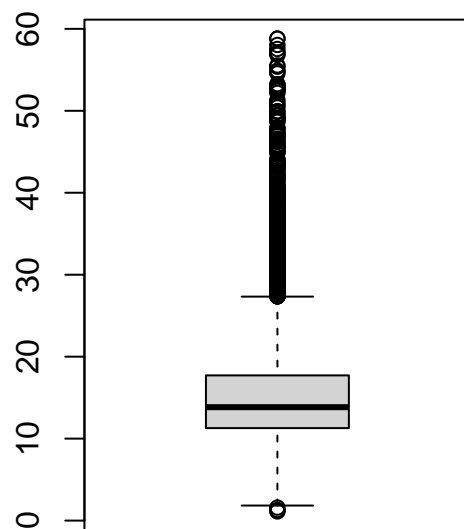


Figure 16: Histogramme et boîte à moustache de la durée moyenne

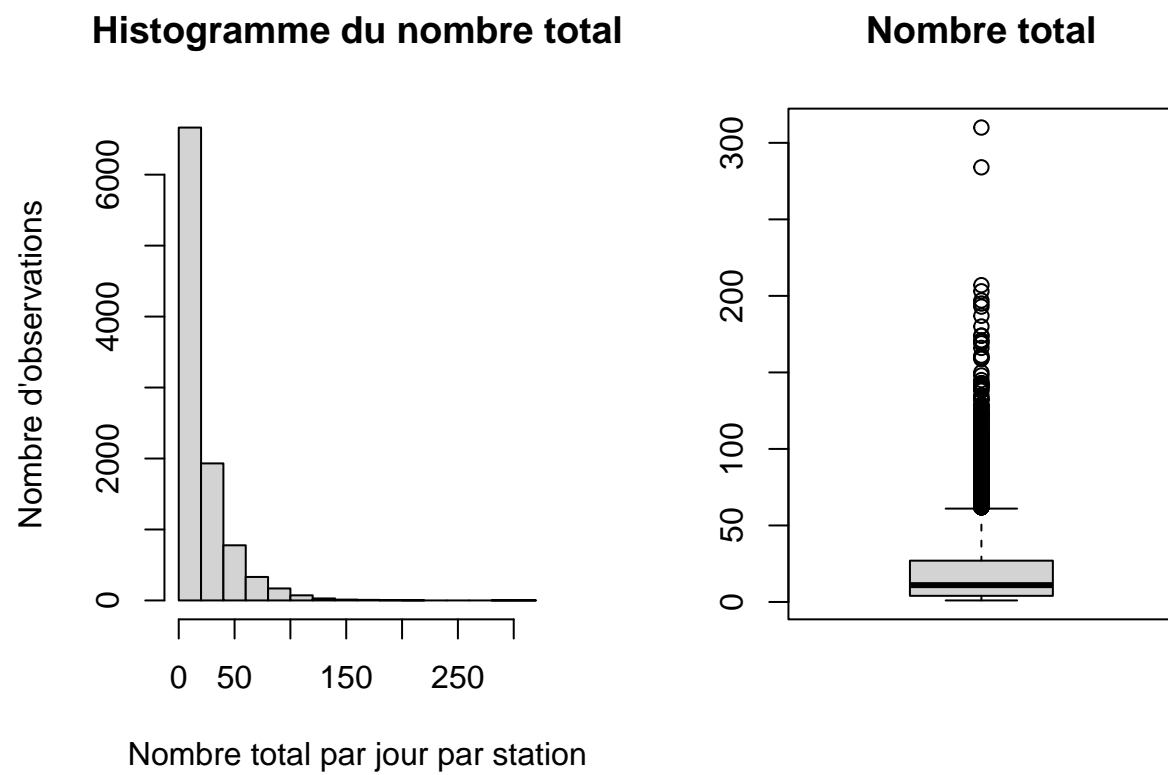


Figure 17: Histogramme et boîte à moustache du nombre total de déplacement

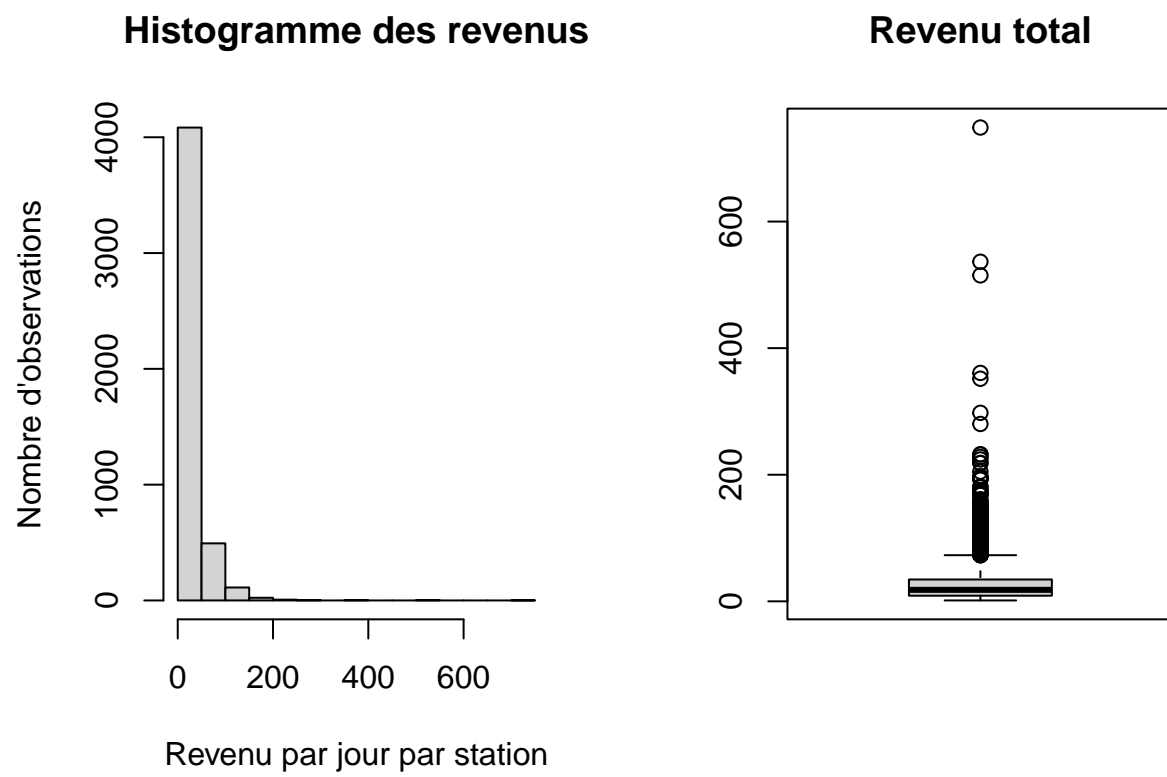
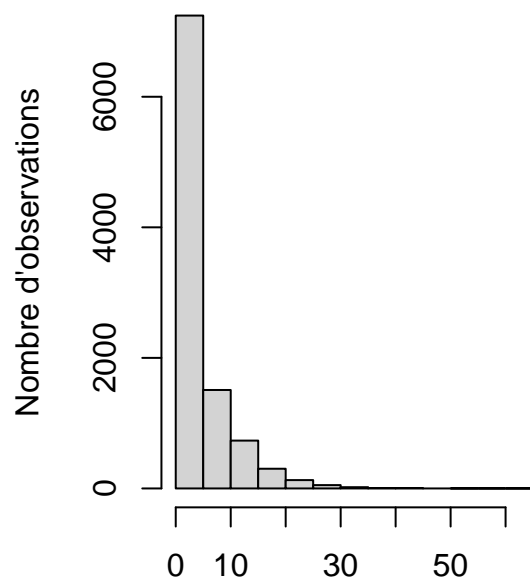
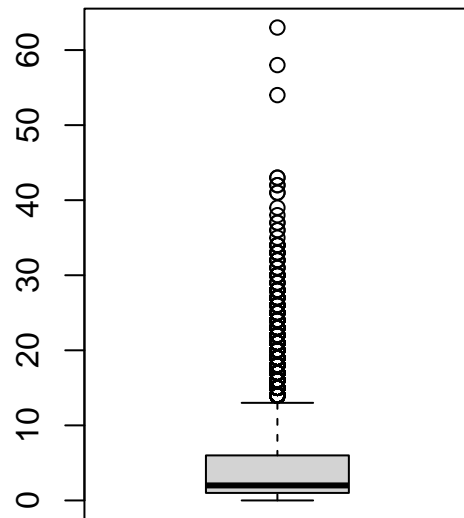


Figure 18: Histogramme et boîte à moustache des revenus

Histogramme du nombre de déplacements



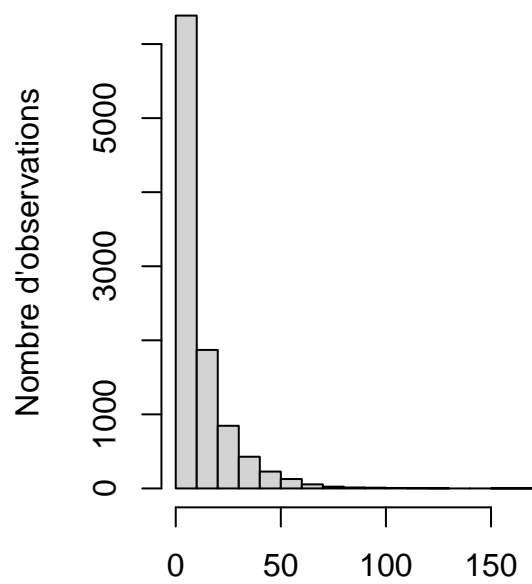
Nombre de déplacements AM



Nombre de déplacements en AM

Figure 19: Histogramme et boîte à moustache du nombre de déplacement en matinée

Histogramme du nombre de déplacements



Nombre de déplacements en PM

Nombre de déplacements PM

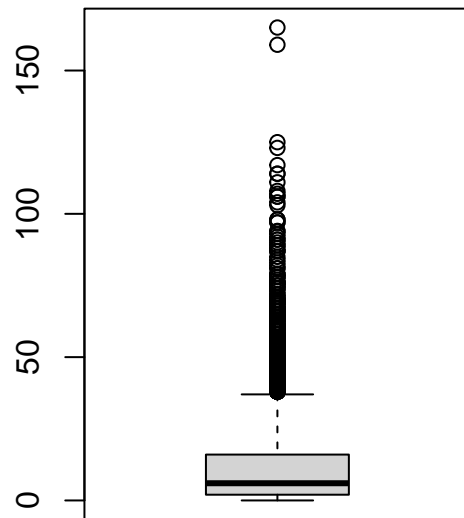


Figure 20: Histogramme et boîte à moustache du nombre de déplacement en après-midi

amme de la température moyenne (Température moyenne quotidienne

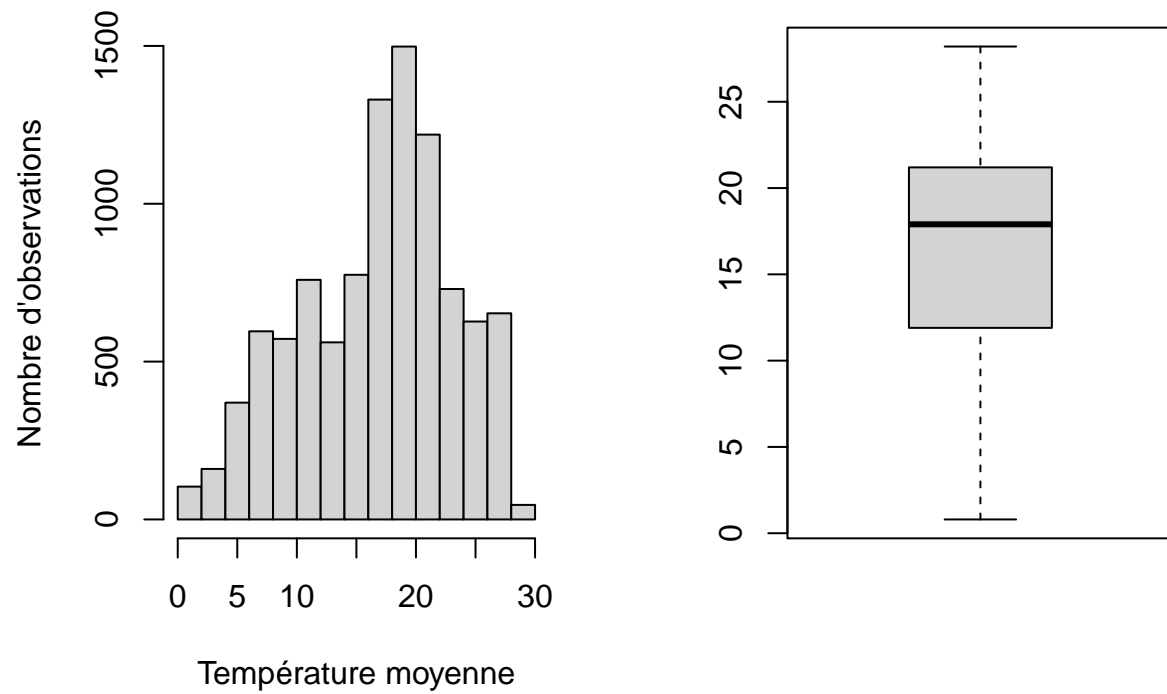
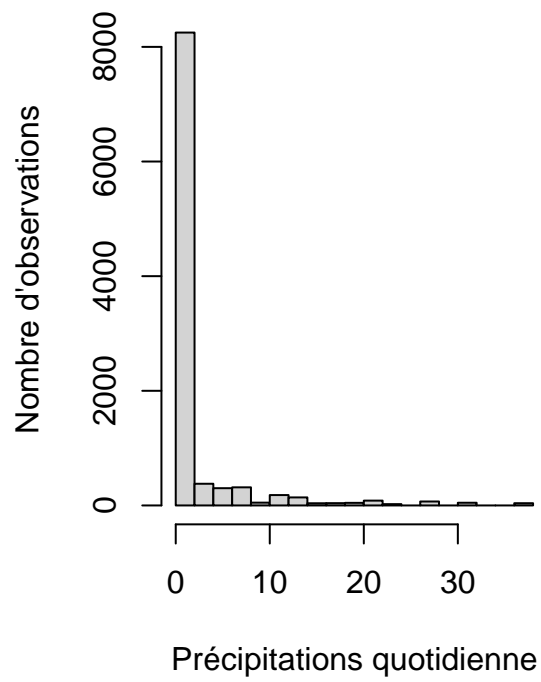


Figure 21: Histogramme et boîte à moustache de la température moyenne quotidienne

Histogramme de la quantité de précipi



Précipitations quotidienne

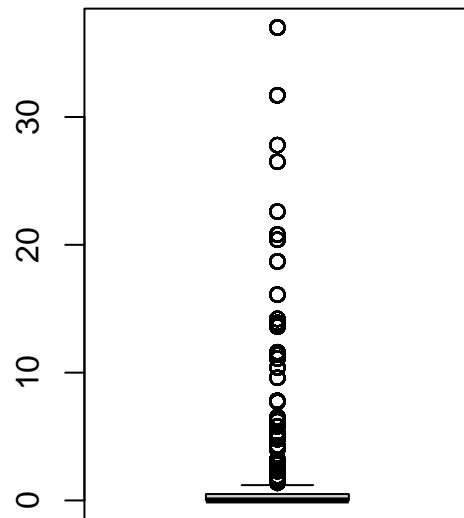


Figure 22: Histogramme et boîte à moustache de la quantité de précipitations quotidienne

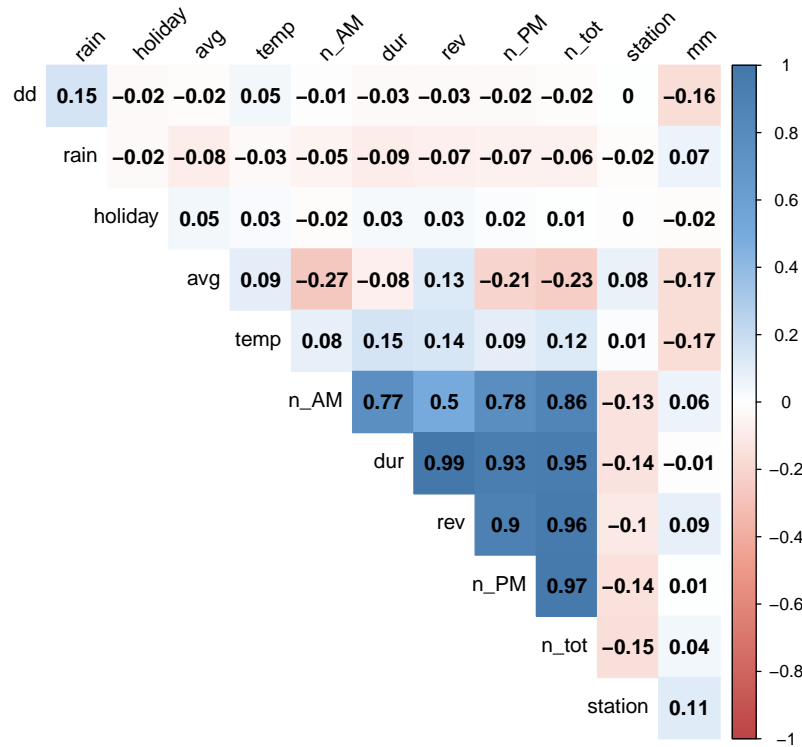


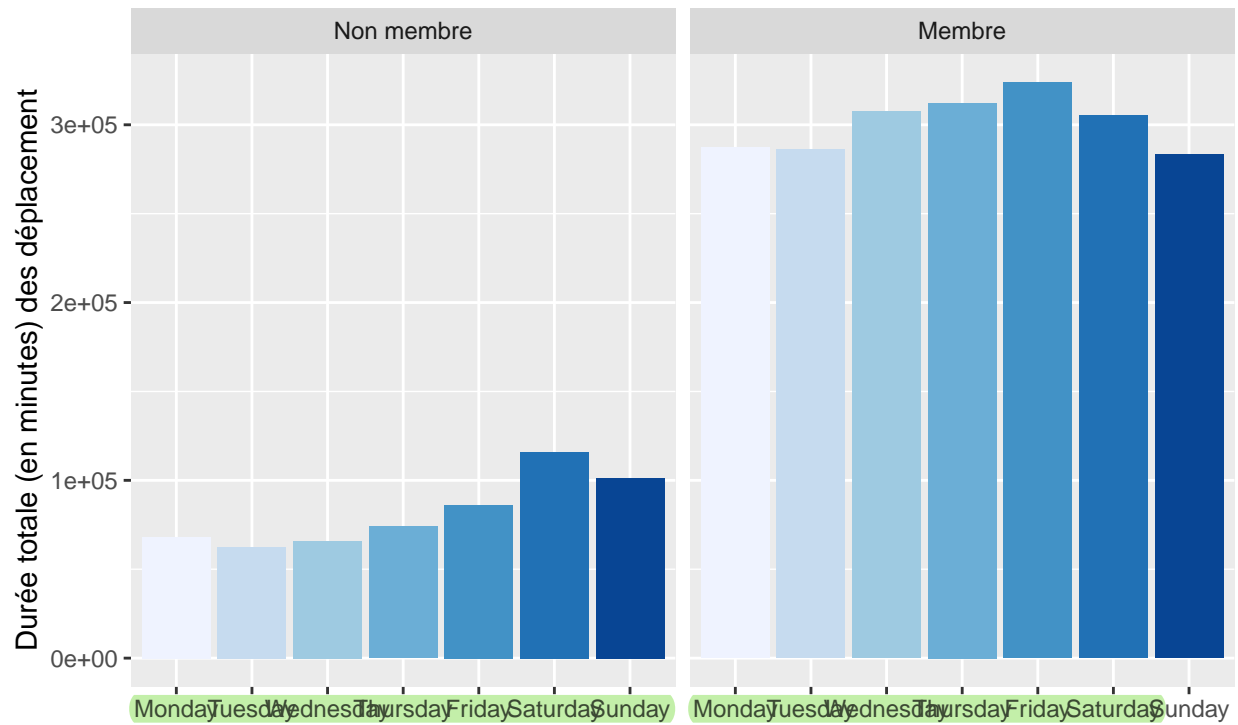
Figure 23: Matrice de corrl lation

2.2.1 Exploration des relations avec la dur e des trajets

En croisant la variable `dur` avec la variable `wday` et `mem`, nous pouvons obtenir le figure ?? . D’abord, on remarque que les membres utilisent beaucoup plus les Bixi (au niveau de la dur e) que les non-membres, et ce, peu importe le jour de la semaine. Les types d’utilisation semblent aussi changer, par exemple, les non-membres utilisent relativement plus les Bixi les fin de semaines que les membres. En effet, les membres semblent avoir une utilisation croissante la semaine et d croissante la fin de semaine. Naturellement, on pourrait penser que c’est parce qu’ils utilisent le Bixi comme un moyen de transport r gulier pour aller au travail ou   l’ cole par exemple.

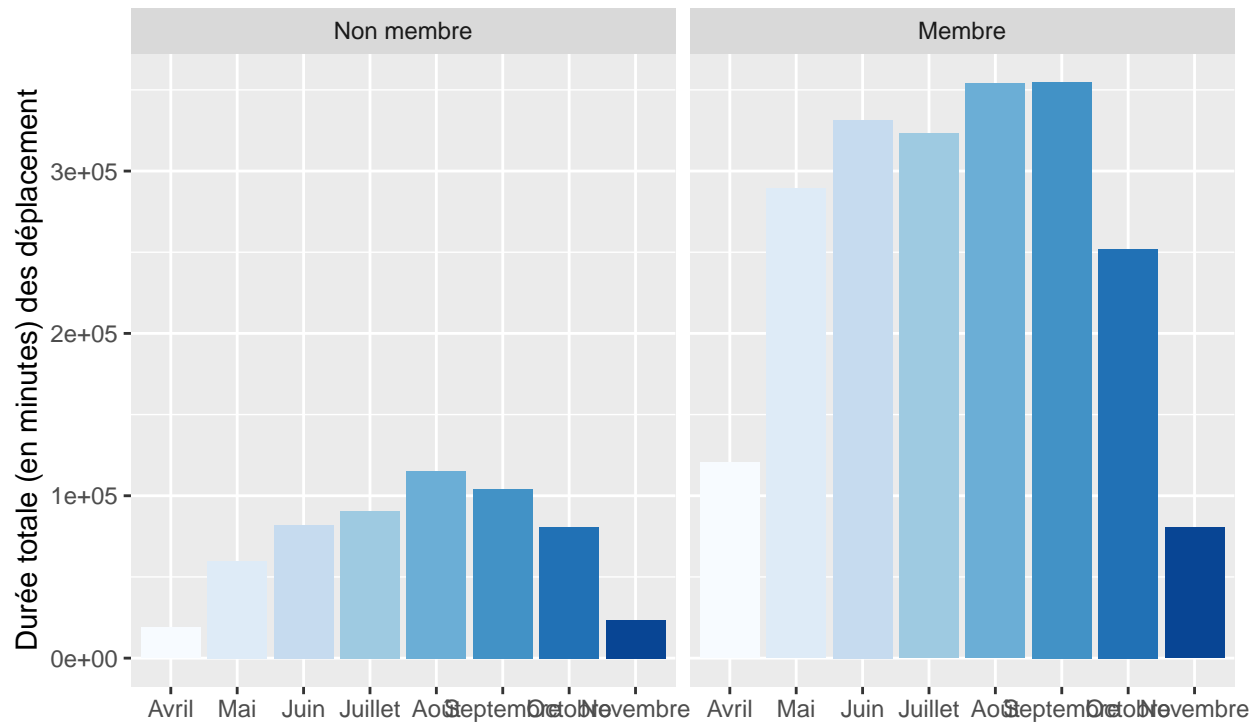
D’un point de vue d’affaire, il pourrait  tre int ressant pour l’entreprise Bixi de faire des accords avec les gros employeurs de Montr al pour augmenter ses clients et acc l rer la transition vers des moyens de transport plus doux. Dans ce m me ordre d’id e, la localisation des Bixi proche des milieux de travail et des milieux scolaires devient une avenue int ressante si l’entreprise a pour but d’augmenter son nombre d’abonn s mensuel.

Durée totale des déplacements par jours selon les membres et les non membres



En croisant la variable `dur` avec la variable `mm` et `mem`, nous pouvons obtenir le figure ?? . Plusieurs analyses sont pertinentes à partir de ce graphique. D'abord, les membres utilisent beaucoup plus les Bixi (au niveau de la durée) que les non-membres, et ce, peu importe le mois de l'année. D'une façon similaire, l'utilisation semble atteindre son pic vers août pour les deux types de clients.

Durée totale des déplacements par jours selon les membres et les non membres



En croisant la variable `dur` avec la variable `temp` et `mem`, nous pouvons obtenir le figure ?? . Naturellement, on voit une relation positive entre la température totale moyenne et la durée totale de déplacement. La relation semble toutefois ralentir voir stagner vers les 20°C. Cela peut pourrir venir du fait que l'utilisation des Bixi est plus élevée en été (donc les mois avec une température plus élevée). Ainsi, ce graphique donne globalement les mêmes informations que le précédent étant donné que la température est fortement corrélée avec les mois.

Au niveau des déplacements moyens (avec la variable `avg`), il semble que la durée moyenne des déplacements diminue après les 15°C.

Au niveau de la corrélation, elle est de 0.53 pour les non-membres et de 0.67 pour les membres pour le déplacement moyen total et de 0.26 pour les non-membres et de 0.34 pour les membres pour le déplacement moyen.

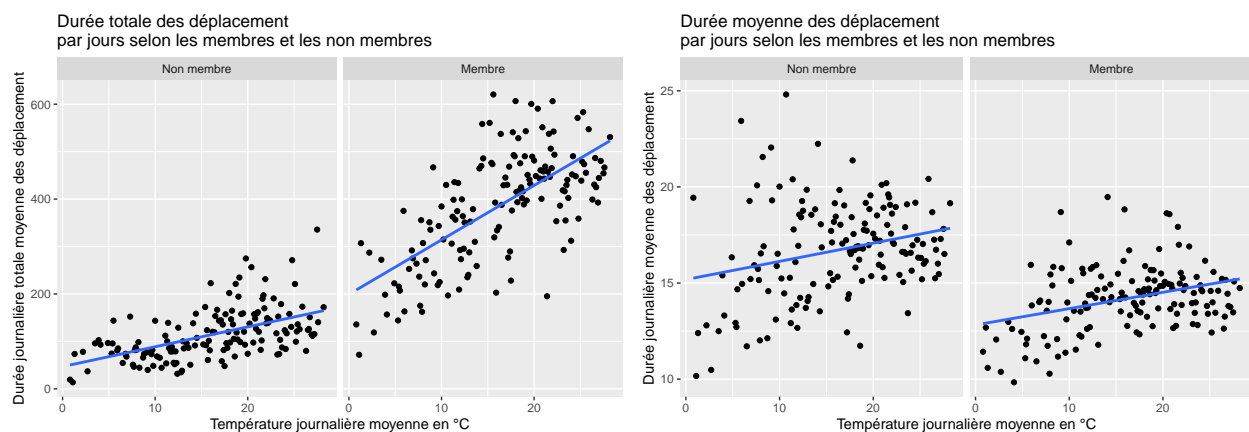


Table 3: Détail sur la durée selon les abonnements

Membre	Durée moyenne	Durée totale	Durée totale relative
Non-membre	16.73821	572884	121.2196
Membre	14.20860	2105066	399.1403

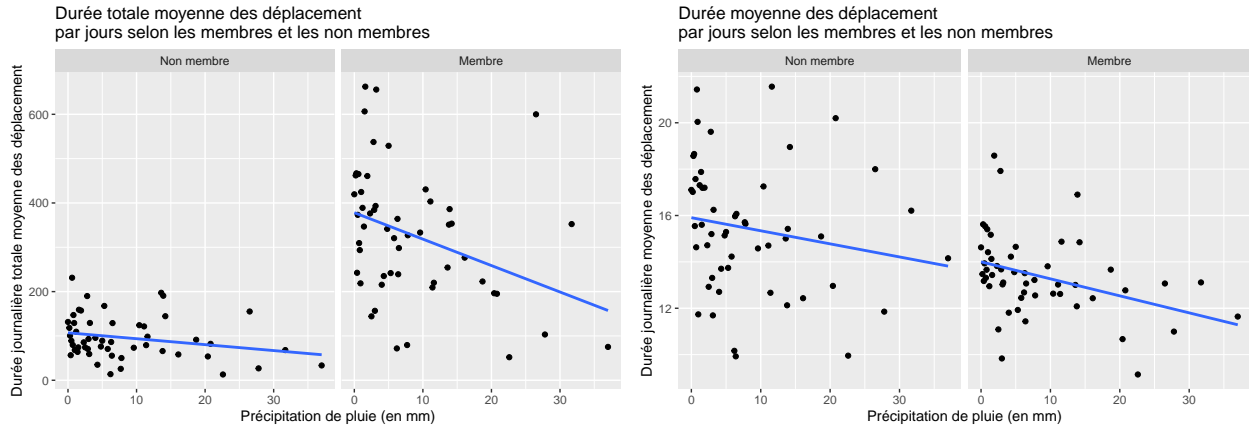
Table 4: Détail sur la durée selon les fériés

Jour	Durée moyenne	Durée totale	Durée totale relative
Non férié	15.35440	2597155.75	266.2384
Férié	17.38251	80794.15	329.7720

En croisant la variable **dur** avec la variable **rain** et **mem**, nous pouvons obtenir le figure ?? . Pour les non-membres l'effet de la pluie sur la durée totale moyenne n'est pas autant significative que pour les membres. Cela est assez étonnant, car il semble que les membres devraient avoir un usage régulier et ainsi moins sensible aux facteurs exogènes comme la pluie.

Au niveau des déplacements moyens (avec la variable **avg**), l'effet est assez similaire entre les membres et les non-membres, les déplacements sont en moyennes plus courts avec la pluie.

Au niveau de la corrélation entre la pluie et la durée totale moyenne des déplacements, elle est de -0.36 pour les membres et de -0.24 les non-membres. Pour la durée moyenne, les corrélations sont de -0.38 pour les membres et de -0.18 pour les non-membres.



En croisant la variable **dur** et **avg** avec **mem**, nous pouvons obtenir le tableau 3.

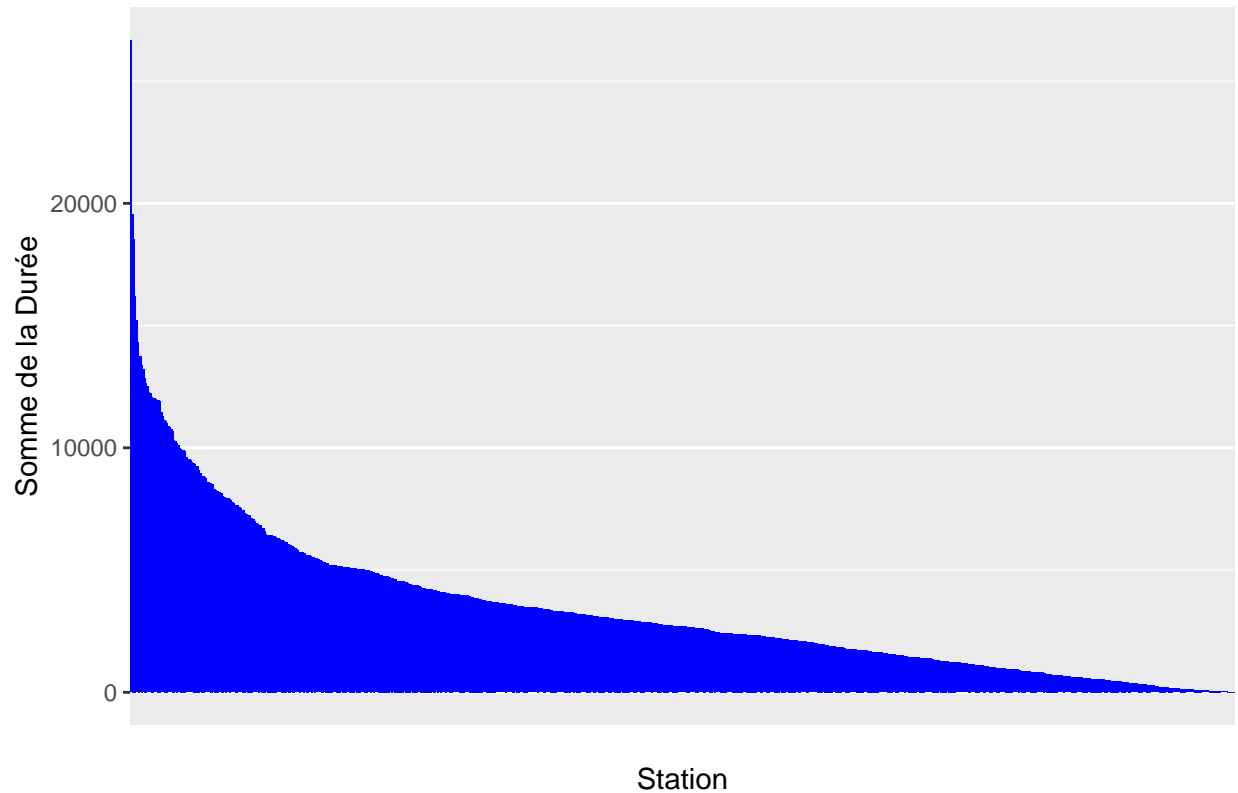
La durée moyenne des trajets est assez similaire entre les membres et les non-membres. En revanche, les membres ont une durée totale de trajets beaucoup plus haute que les non-membres, ce qui rejoint ce que nous avons observé précédemment. Il apparait aussi que les membres sont responsables de la majeure partie des déplacements (en durée). Ainsi, selon la rentabilité du programme d'adhésion (donnée non-disponible), il pourrait être opportun d'attirer des nouveaux membres avec des promotions et des campagnes publicitaires.

En croisant la variable **dur** et **avg** avec **holiday**, nous pouvons obtenir le tableau 4.

La durée moyenne des trajets est légèrement supérieure pour les jours fériés. Aussi, la durée totale relative des trajets est supérieure pendant les jours fériés. Cela s'explique par le fait que les usagers ont plus de temps disponible pour utiliser les Bixi. Il serait aussi avantageux pour Bixi d'augmenter la disponibilité des vélos lors des jours fériés.

Selon le graphique ??, il est possible de modéliser les stations par leur durée totale de déplacements de la saison. On voit que la plupart des stations sont sous le seuil des 10 000 minutes.

Stations Classées en Ordre Décroissant par Somme de Durée



En fonction de la durée totale durant la saison, le top 3 des stations sont :

- de la Commune / Place Jacques-Cartier (Vieux-Montréal)
- de la Commune / St-Sulpice (Vieux-Montréal)
- Métro Mont-Royal (Rivard / du Mont-Royal) (Plateau Mont-Royal)

Ainsi, on remarque que les deux stations les plus populaires en terme de durée sont situées dans le Vieux-Montréal. Il est possible d'attribuer cela aux touristes.

2.2.2 Exploration des relations avec l'utilisation des Bixi

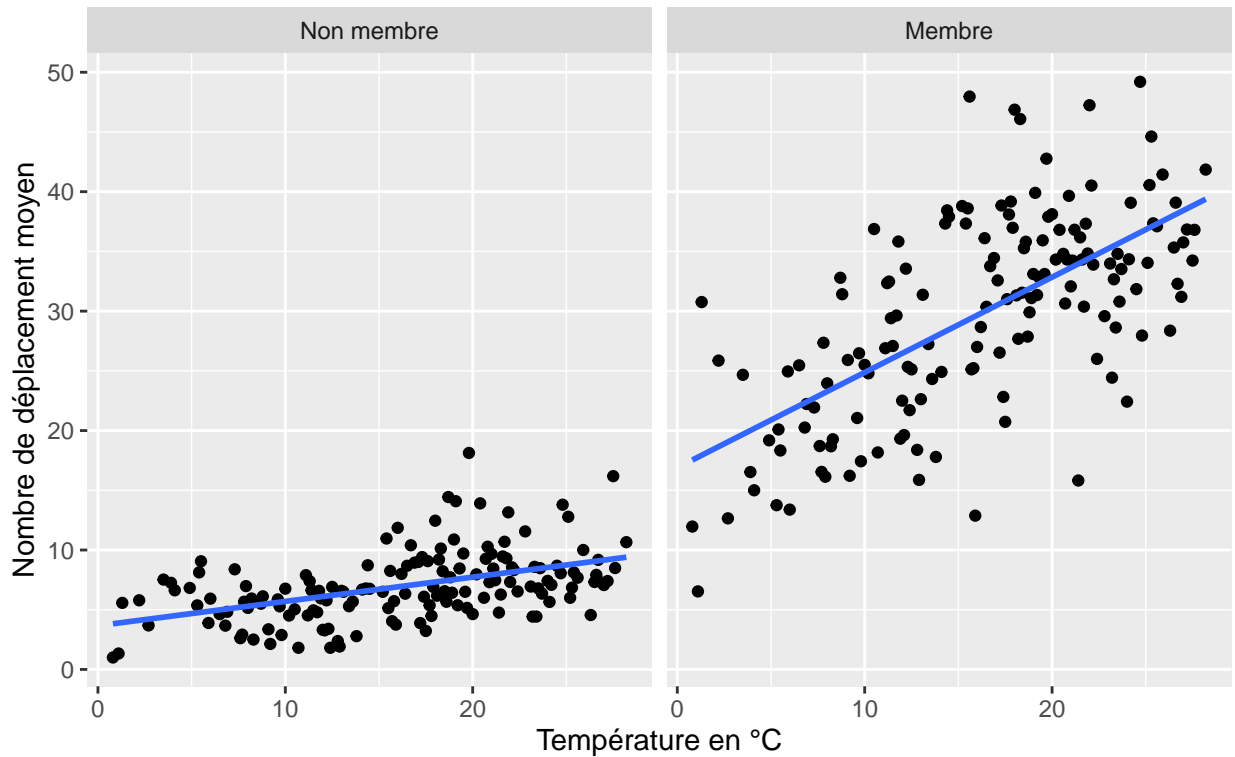
Dans la prochaine section, nous allons explorer les relations entre l'utilisation des Bixis, soit la variable `n_tot` (ainsi que `n_AM` et `n_PM` dans certains cas) et diverses variables de notre base de données.

La figure ?? montre la relation entre le nombre de déplacement total moyen par jour par rapport à la température pour les membres et les non-membres.

Table 5: Température moyenne par mois

Mois	Température moyenne (°C)
4	9.94
5	14.62
6	21.29
7	20.43
8	23.61
9	17.37
10	12.22
11	5.46

Nombre moyen de déplacements par rapport à la température

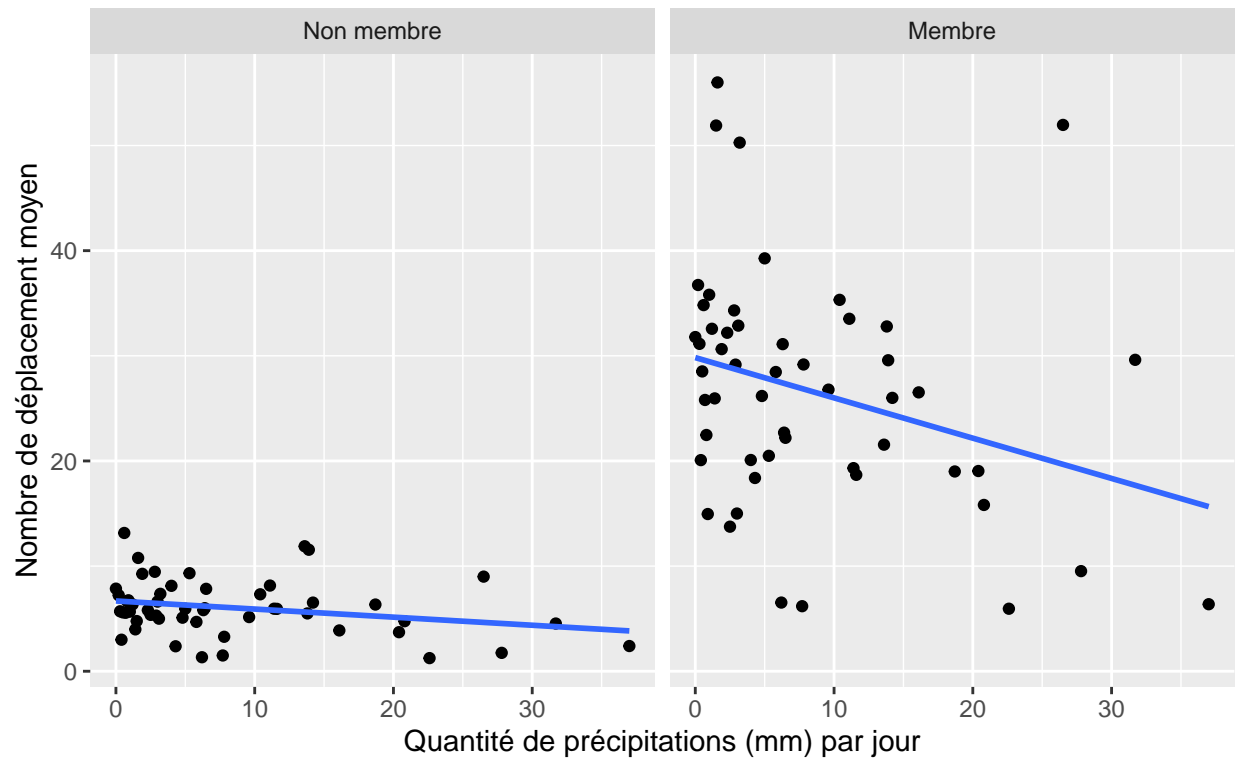


D'après ce graphique, il est possible de constater qu'il semble exister une relation linéaire entre le nombre de déplacement moyen effectué en Bixi et la température, que ce soit pour un membre et un non-membre. Dans les deux cas, on observe que le nombre de déplacement moyen a tendance à augmenter avec la température. Une forte corrélation positive de 0.65 pour les membres et une corrélation moyenne de 0.48 pour les non-membres.

D'un point de vue d'affaires, si l'entreprise Bixi désirerait par exemple effectuer une mise à jour logiciel ou effectuer de la maintenance sur certains vélos et que ceci nécessiterait de récupérer une certaine quantité de Bixi, il faudrait que l'entreprise entreprenne ces changements pendant les journées où il fait plus froid, ou durant les périodes plus froides soit durant les mois d'avril, mai, octobre ou novembre comme le montre le tableau afin d'éviter de causer des problèmes de manque de vélos pour les usagers.

Par la suite, nous avons observé la relation entre le nombre de déplacement total moyen par jour d'après la quantité de précipitation tombés dans une journée. La figure ?? montre cette relation.

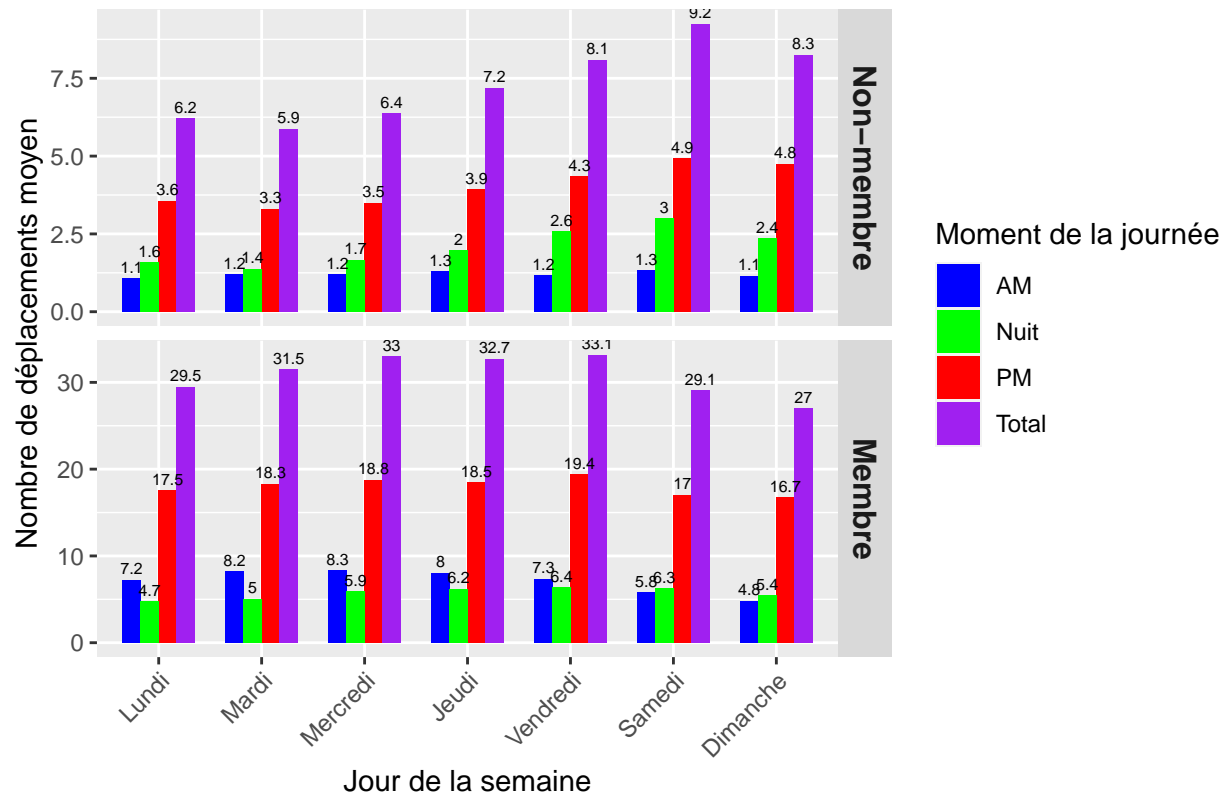
Nombre moyen de déplacements par rapport à la quantité de précipitations



D'après ce graphique, il semble y avoir une faible corrélation entre le nombre de déplacement et la quantité de précipitation dans une journée. En effet, la corrélation est de -0.26 pour les non-membres et -0.30 pour les membres.

La figure ?? montre la relation entre le nombre de déplacement moyen pour différents moment de la journée et le nombre total de déplacement d'après la journée.

Nombre de déplacements moyen en Bixi d'après le jour de la semaine



Pour les membres, on remarque qu'en moyenne, il y a plus de déplacements en après-midi qu'en avant-midi et durant la nuit. De façon générale, le nombre de déplacements en avant-midi est semblable à celui de la nuit. Il est aussi possible de remarquer que le nombre de déplacement total est très semblable pour tous les jours de la semaine et est un peu plus faible pour les jours de la fin de semaine. Finalement, on peut tirer de cela que les membres utilisent plus les Bixis durant la semaine pour aller au travail par exemple et que plus de membres les utilisent l'après-midi.

Pour les non-membres, il y a aussi plus de déplacements en après-midi qu'en avant-midi et durant la nuit. Puis, l'utilisation en avant-midi et la nuit est très semblable pour tous les jours aussi. Il est aussi possible de remarquer que le nombre de déplacement total est très semblable pour tous les jours. Cependant, contrairement aux membres, les jours de fin de semaine sont ceux avec le plus de déplacements. Finalement, on peut donc supposer que les non-membres sont plus portés à utiliser les Bixis les fins de semaines pour des loisirs et moins pour le travail la semaine.

En moyenne, le nombre de déplacement total pour les membres est plus élevé à 30.8 par jour et à 7.32 par jour pour les non-membres.

D'un point de vue d'affaires, comme les membres utilisent plus les Bixis la semaine et qu'il y a plus de déplacements pour les membres que les non-membres. Il serait judicieux de toujours s'assurer d'une bonne disponibilité de Bixi la semaine et ce surtout en après-midi. Si l'entreprise Bixi devait récupérer une certaine quantité de vélo pour la maintenance par exemple, nous suggérons qu'elle procède durant les avant-midi et/ou les fins de semaine.

Le tableau 6 montre la relation entre le nombre de déplacement et si la journée est férié ou non.

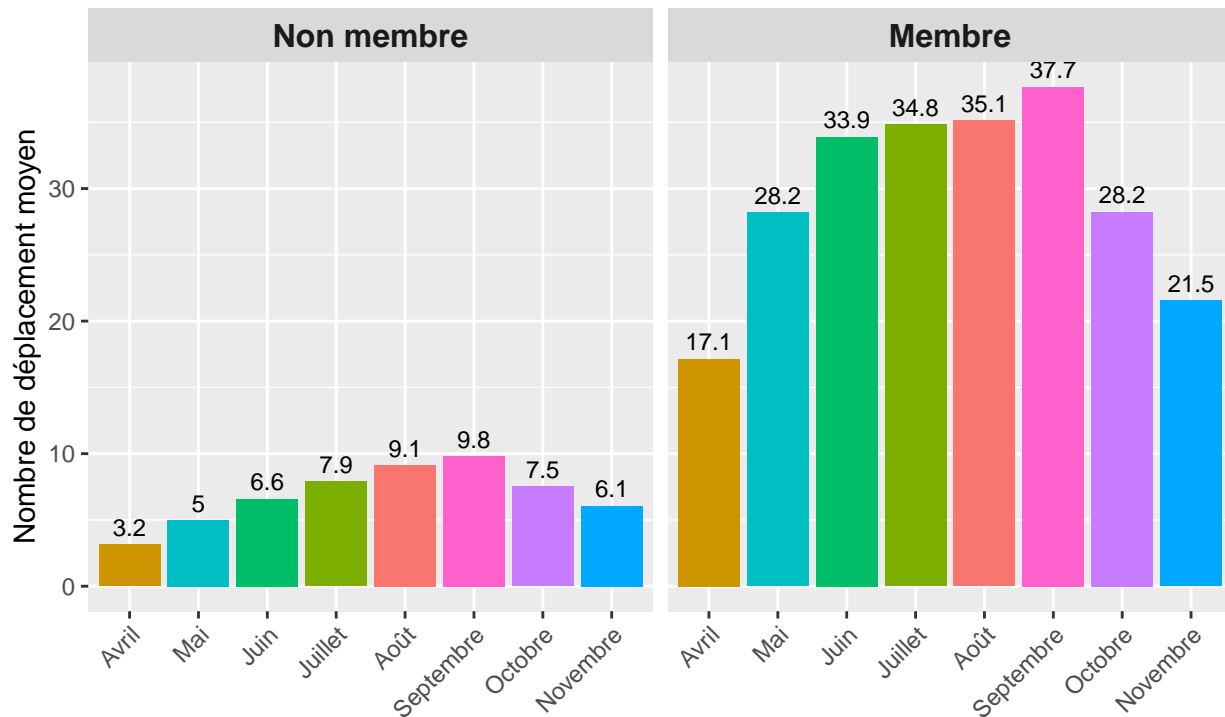
Ce tableau montre qu'en moyenne il y a un peu plus de déplacements durant les jours fériés, mais on parle ici d'un peu plus qu'un seul déplacement. Ce qui ne semble pas assez significatif pour conclure que le nombre de déplacement est affecté par le fait qu'il s'agit d'un jour férié ou non.

La figure ?? représente la relation entre l'utilisation moyenne des Bixis et le mois de l'année.

Table 6: Détail sur le nombre de déplacement selon les fériés

Jour	Nb. déplacement moyen	Nb. déplacement total
Non férié	19.69103	192086
Férié	20.91020	5123

Nombre de déplacement moyen par jours selon les membres et les non membres



On remarque que l'allure des deux graphiques est très semblable entre les membres et les non-membres. On parle ici d'une très forte corrélation de 0.86. On peut alors dire que le comportement des membres et des non-membres est très semblable par rapport au mois.

Par la suite, en comparant ces données avec les données du tableau @ref(table:graph_temp_vs_mois) illustrant la moyenne des températures par mois, on obtient une très forte corrélation d'environ 0.85 pour les membres et une forte corrélation d'environ 0.59 pour les non-membres. Nous pouvons donc faire l'hypothèse que la relation entre le nombre de déplacement moyen par mois est dépendante de la température de ce mois. Les gens ont tendance à plus utiliser le Bixi lors des mois plus chauds que les mois plus froids (avril, octobre et novembre).

Cependant, on remarque que dans les deux cas (membre et non-membre), le mois avec le plus de déplacement moyen est le mois de septembre qui pourtant n'est pas le mois le plus chaud. Notre hypothèse est qu'il s'agit de la période de la rentrée scolaire et du retour au travail pour plusieurs, alors plusieurs optent pour un transport plus écologique comme le Bixi à ce moment de l'année avant de tomber dans la période plus froide de l'année avec les mois d'octobre et novembre.

D'un point de vue d'affaires, l'entreprise Bixi devrait prioriser et s'assurer de pouvoir mettre à disposition une grande quantité de Bixi pour les périodes plus chaudes de l'année ainsi qu'en septembre.

Par la suite, nous avons identifié les stations qui génèrent le plus de déplacements pour les membres et les

Table 7: Détail sur les stations les plus utilisées pour les membres

Stations	Emplacement	Arrondissement	Nb. de déplacement total
256	Métro Mont-Royal (Rivard / du Mont-Royal)	Plateau Mont-Royal	1329
316	Prince-Arthur / du Parc	Plateau Mont-Royal	1189
169	De Bréboeuf / du Mont-Royal	Plateau Mont-Royal	1164
179	Messier / Du Mont-Royal	Plateau Mont-Royal	1136
167	Boyer / du Mont-Royal	Plateau Mont-Royal	1105

Table 8: Détail sur les stations les plus utilisées pour les non-membres

Stations	Emplacement	Arrondissement	Nb. de déplacement total
470	de la Commune / Place Jacques-Cartier	Le Vieux Montréal	652
336	Duluth / St-Laurent	Plateau Mont-Royal	373
256	Métro Mont-Royal (Rivard / du Mont-Royal)	Plateau Mont-Royal	320
762	St-Dominique / Napoléon	Plateau Mont-Royal	275
469	de la Commune / St-Sulpice	Le Vieux Montréal	270

non-membres. Les deux tableaux 8 présentent le nombre de déplacements total des 5 stations avec le plus de déplacements pour les membres et les non-membres ainsi que l'arrondissement de Montréal où se situe ces stations.

On remarque que l'arrondissement avec le plus de stations dans le top 5 est la Plateau Mont-Royal et Le Vieux-Montréal (pour deux stations). D'un point de vue d'affaires, il serait pertinent de prioriser ces arrondissements pour l'ajout de nouveaux Bixis par exemple pour s'assurer qu'il n'en manque jamais et toujours en avoir une grande quantité en fonction dans ces arrondissements. Puis, aussi d'effectuer plus de maintenance préventive des installations de ces arrondissements, car celles-ci génèrent beaucoup de d'attraction et il ne faudrait pas qu'elles soient hors service.

À ajouter dans Conclusion (Conclusion Sam partie n_tot)

Pour conclure, on peut constater que les périodes d'utilisation les plus faibles de Bixis sont: les journées plus froides, les mois plus froids (avril, octobre et novembre), la fin de semaine et l'avant-midi. Puis les périodes avec le plus de déplacements sont les journées plus chaudes, les mois plus chaud (les autres mois à l'étude), durant la semaine et en après-midi. De plus, les secteurs avec les stations qui génèrent le plus de déplacements sont le Plateau Mont-Royal et le Vieux-Montréal. Finalement, l'entreprise devraient donc s'assurer de toujours avoir le plus de Bixis possibles disponibles pour les périodes fortes, procéder aux maintenances lors des périodes plus faibles et s'assurer du bon état des stations et des vélos dans les arrondissements avec les stations les plus en demande.

2.2.3 Exploration des relations avec le revenu

Dans cette section, il sera principalement question de la variable revenu **rev**, puisque c'est la variable qui met en relation le revenu généré par les non-membres avec les autres variables. De ce fait, d'un point de vue entrepreneurial, c'est une variable d'intérêt capital, puisqu'elle participe au calcul de profit de BIXI. Cependant, puisque l'encodage des données du revenu n'est offert que pour les non-membres et que le revenu est calculé à partir de la durée du trajet, la corrélation entre ces deux dernières est relativement forte comme on peut le voir dans le tableau 23. Autrement dit, l'entreprise peut constater qu'une augmentation de la durée des trajets **dur** revient principalement à une augmentation du revenu **rev**. Conséquemment, visualiser la relation entre le revenu et les journées de la semaine **wday**, le mois **mm**, la température **temp** et les précipitations **rain** reviennent à visualiser la même chose que dans la sections de l'exploration des relations avec la durée des trajets. Dès lors, il reste néanmoins intéressant d'observer comment le revenu se comporte lors des journées fériées **holiday**.

Table 9: Détails entre les journées fériées et le revenu rapporté par les non-membres

	Revenu moyen	Revenu total
journée normale	27.17594	125498.480
Journée fériée	34.20486	3694.125

Dans le cas de la relation entre les journées fériées et le revenu, il est possible d'apercevoir qu'en moyenne les non-membres utilisent plus bixi lors des journées fériées, puisque le revenu moyen est plus élevé lors de ces journées (9). Cela étant dit, il y a un parallèle à faire avec la figure illustrée plus haut ?? où l'on peut voir que les non-membres ont plus tendance à se déplacer en bixi la fin de semaine. Par conséquent, on peut soutenir l'hypothèse que l'utilisation de bixi par les non-membres est liée à des fins plus récréatives.

Du point de vue de l'entreprise, la localisation des stations de bixi pour les non-membres devient alors cruciale. Autrement dit, placer des stations de bixi proches des lieux culturels et les lieux de divertissement pourraient ainsi favoriser le revenu de l'entreprise.

Conclusion