



Mockup

Exo : Achalandage et ponctualité des lignes de bus 9 et 22 de la couronne Nord de Montréal

INF8808 - Visualisation de données

Adam Prévost - 1947205
Armelle Jézéquel - 2098157
Clara Serruau - 2164678
Jules Lefebvre - 1847158
Julien Dupuis - 1960997

23 mars 2022

Table des matières

Mise en contexte	2
1.1 Contexte	2
1.2 Objectif	2
1.3 Public cible	2
Jeu de données	2
2.1 Métadonnées	2
2.2 Données	3
Questions cibles	5
Mockup	6
4.1 Visualisation 1	6
4.2 Visualisation 2	8
4.3 Visualisation 3	10

1. Mise en contexte

1.1 Contexte

Le transport en commun est une activité essentielle et répandue dans la grande région de Montréal. Dans cette région, au moins une personne sur cinq utilise le transport en commun régulièrement selon *Statistique Canada*. La compagnie qui s'occupe du bon fonctionnement du réseau est *Exo*. Le domaine des transports en commun est en continuelle modernisation, où se côtoient anciennes et nouvelles infrastructures. Nous porterons notre attention sur le réseau des autobus, pour lesquels de plus en plus de données sont recueillies et analysées (achalandage, ponctualité de chaque autobus, etc.) Elles peuvent, entre autres, servir à offrir des services de meilleure qualité à la population de Montréal et des alentours.

1.2 Objectif

L'objectif d'*Exo* est d'utiliser les données d'achalandage (nombre de personnes à bord) et de ponctualité (temps de retard ou d'avance) récoltées par GPS à chaque arrêt pour rendre compte de la qualité du service. Ces données sont usuellement étudiées séparément, mais on souhaite les étudier conjointement afin de rendre compte d'une autre réalité du réseau. Plus précisément, la qualité de service est évaluée en ciblant les lignes, les voyages, les arrêts ou les tronçons les plus problématiques. Le but de la visualisation est d'aider à prendre des décisions correctives en changeant les horaires ou les trajets par exemple. La visualisation doit donc aussi permettre d'identifier des tendances si un problème est généralisé.

1.3 Public cible

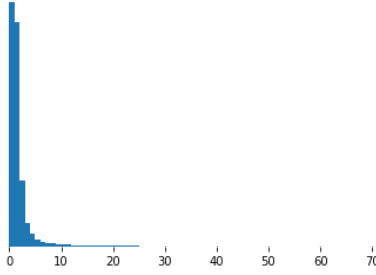
La visualisation est principalement dédiée aux employés d'*Exo* qui planifient les opérations de transport en commun. Ils utiliseront celle-ci pour avoir une idée de la qualité de service et prendre des décisions éclairées. La visualisation ne sera pas à destination des usagers du réseau de transport, donc les utilisateurs de la visualisation auront déjà de l'expertise dans le domaine.

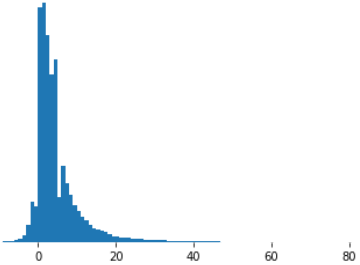
2. Jeu de données

2.1 Métadonnées

Les données sont fournies par *Exo*, et ont été récoltées de septembre à décembre 2021. Les données sont récoltées par des capteurs présents dans les bus : GPS, capteur de détection de personne montante, etc. Les données sont sujettes à la défaillance des capteurs et des bus. Certains bus peuvent avoir eu des problèmes durant le voyage et certains arrêts ne sont pas marqués (faute d'usager montant ou descendant). Le fichier à traiter comprend 206 076 lignes de données.

2.2 Données

Variable	Description	Data type	Data values / range
Date	Date du voyage	date :YYYY-M M-DD	1ier septembre 2021 au 1ier décembre 2021
Secteur	Le secteur de la métropole dans lequel évolue la ligne de bus	string	une seule valeur : CITLA
Ligne	Numéro de la ligne de bus	int	{9; 22}
Voyage	Numéro du voyage effectué dans la journée (1 = 1ier bus de la journée)	int	range numérique : [1,126] range textuelles : {69_F1-01; 7_F1-01; ...} (voyages lors des jours fériés, à ignorer car non représentatifs)
Arret_code	Code de l'arrêt de bus	int	range : [80400, 84967] 231 valeurs différentes)
Arret_nom	Nom de l'arrêt de bus	string	{du Curé-Labelle / Saint-Georges; Saint-Georges / Ouimet; du Curé-Labelle / 104e Avenue Est; ...} 188 valeurs différentes
Montants	Personnes qui embarque	int	min : 0.0 max : 69.0 moyenne: 1.53994157495293 médiane : 1.0 
Etat_ponctualité	Discrétisation de <i>Minutes_ecart_planifie</i> : [-11;-1] : <i>Avance</i> [0;4] : <i>Ponctuel</i> (sauf dans le cas du dernier arrêt	string	{Ponctuel, Retard, Avance}

	de la ligne : [-11;4] : <i>Ponctuel</i> [5;80] : <i>Retard</i>		
Minutes_ecart_planifie	Minutes entre le moment de passage réel du bus et le prévu <i>Réel - Prévu</i>	int	min : -11.0 max : 80.0 moyenne: 4.426027290902385 médiane : 3.0 
Séquence_arret	Ordonne les arrêts pour une ligne/direction	int	Ligne 9 (Métro Montmorency) : [1; 74] Ligne 9 (Lafontaine Via Gare Saint-Jérôme) : [1; 79] Ligne 22 (Gare Sainte-Thérèse) : [1; 43] Ligne 22 (Bois-des-Filion Via Georges IV) : [1; 41]
Girouette	Nom de la direction du bus	string	9 : {Lafontaine Via Gare Saint-Jérôme; Métro Montmorency} 22 : {Bois-des-Filion Via Georges IV; Gare Sainte-Thérèse}
Direction	Direction du bus	string	{sens opposé; sens}
Periode_horaire	Période de la journée durant laquelle a eu lieu le voyage	string	{AM; PM; interpointe; Soir/Nuit/Matin}
Arret_Latitude	Latitude de l'arrêt de bus	float	range : [45,55828; 45,804104]
Arret_Longitude	Longitude de l'arrêt de bus	float	range : [-74,011985; -73,720406]

Précisions sur les données :

- **Secteur** : le code CITLA correspond à la couronne nord de Montréal, plus précisément au secteur Laurentides.
- **Arret_code** , **Arret_nom** : Il y a plus de codes que de noms d'arrêt, car un arrêt peut avoir un même nom dans les 2 sens, mais avoir un code différent pour chacun.
- **Minutes_ecart_planifie**, **Etat_ponctualité** : la seconde colonne est une discrétisation de la première. Un bus est considéré ponctuel s'il a jusqu'à 4 minutes de retard. Au-delà,

il est considéré comme en retard. Pour l'avance, un bus peut être considéré comme ponctuel s'il a de l'avance et qu'il s'agit du dernier arrêt, ou qu'il a moins de 4 minutes de retard et moins d'une minute d'avance. Enfin, un bus est considéré comme en avance s'il a au moins une minute d'avance.

[-11;-1] : Avance

[0;4] : Ponctuel (sauf dans le cas du dernier arrêt de la ligne : [-11;4] : Ponctuel)

[5;80] : Retard

- **Montant** : nombre fourni par un compteur de passagers dans l'entrée de l'autobus.
- **Voyage** : Certains voyages ont un code textuel. Ils correspondent à des trajets réalisés lors d'un jour férié. Ces données ne sont pas pertinentes pour l'analyse du reste de l'année, le client nous a demandé de les ignorer.
- **Séquence_arrêt** : La ligne 9 a 79 arrêts, mais l'arrêt Station Montmorency est parfois l'arrêt 1, parfois l'arrêt 79 dans la direction Lafontaine Via Gare Saint-Jérôme (est-ce une erreur, ou un cas particulier ?)

3. Questions cibles

	Questions	Priorité (/3)	
	Aspect temporel		
1	Quels moments de la journée sont les plus problématiques?	☆☆☆	1
2	Quel est le retard moyen?	☆☆☆	2-3
3	Quel est l'achalandage moyen?	☆☆☆	3
4	Est-ce qu'un autobus peut rattraper son retard après un arrêt problématique?	☆☆	2
5	Quels problèmes sont récurrents et lesquels sont ponctuels?	☆☆	
6	Existe-t-il des relations entre le retard et le nombre de passagers qui embarquent à chaque arrêt?	☆☆	2
7	Existe-t-il des tendances temporelles particulières dans l'échantillon donné? (impact de la rentrée scolaire, jour de la semaine problématique, Halloween, etc.)	☆☆	

	Aspect géographique		
8	Quel est le retard relatif gagné ou perdu à chaque arrêt?	☆ ☆ ☆	2
9	Quel est le nombre de passagers relatif gagné ou perdu à chaque arrêt?	☆ ☆ ☆	2
10	Quels sont les arrêts ou les tronçons les plus problématiques?	☆ ☆ ☆	2
11	Est-ce qu'un sens du trajet est plus ou moins problématique que l'autre?	☆ ☆	3
12	Quelle est la qualité de service sur les lignes en fonction de l'arrêt ou la zone géographique?	☆ ☆	1
13	Existe-il des contraintes géographiques particulières qui occasionnent des retards?	☆	
14	Existe-t-il des lignes problématiques?	☆	1

Modifications apportées suite à la remise du plan de projet et aux commentaires du client:

1. Suppression de la question «Existe-t-il des tendances temporelles particulières dans l'échantillon donné? (impact de la rentrée scolaire, jour de la semaine problématique, Halloween, etc.)». L'échantillon de données est trop petit (environ 3 mois) pour trouver des liens significatifs. De plus, les données importantes à évaluer sont la tendance normale de trajet, pas les jours exceptionnels.
2. Suppression de la question «Existe-il des contraintes géographiques particulières qui occasionnent des retards?». Les contraintes géographiques ne sont pas adéquatement quantifiables, il faudrait étudier donc le trajet de chaque ligne individuellement, en dehors des données fournies.

4. Mockup

4.1 Visualisation 1

La première visualisation réponds au questions suivantes :

1	Quels moments de la journée sont les plus problématiques?
---	---

10	Quels sont les arrêts ou les tronçons les plus problématiques?
11	Est-ce qu'un sens du trajet est plus ou moins problématique que l'autre?
12	Quelle est la qualité de service sur les lignes en fonction de l'arrêt ou la zone géographique?
14	Existe-t-il des lignes problématiques?

Description

Cette première visualisation sera une carte de chaleur concernant la ligne (9 ou 22) et la direction (aller ou retour) sélectionnées.

Il y a 3 modes dans cette visualisation : le mode par défaut, qui permet de visualiser un indicateur combinant la ponctualité et l'achalandage, puis un mode permet de visualiser uniquement la ponctualité, et un uniquement l'achalandage. L'indicateur combinant les 2 métriques est : nombre de montants ponctuel / nombre de montants total. Pour changer de mode, il faudra utiliser un bouton à côté de la carte.

Dans cette carte de chaleur en 2D, il y a 2 axes : l'axe vertical correspond au numéro de trajet. Cette donnée est familière aux planificateurs d'Exo, il n'est pas besoin de la traduire en heure. L'axe horizontal correspond aux arrêts. Mettre tous les noms d'arrêts chargera trop la vis, donc on propose de mettre juste le premier et le dernier. Pour calculer les moyennes de chaque carré, on fera une moyenne sur tous les jours, pour l'arrêt et le numéro de trajet concerné.

Cette visualisation permet de mettre en valeur le fait que certains arrêts ou certains trajets sont plus problématiques que d'autres. Dans ce contexte, "problématique" signifie que beaucoup de personnes sont impactés par l'absence de ponctualité du bus.

Interactions

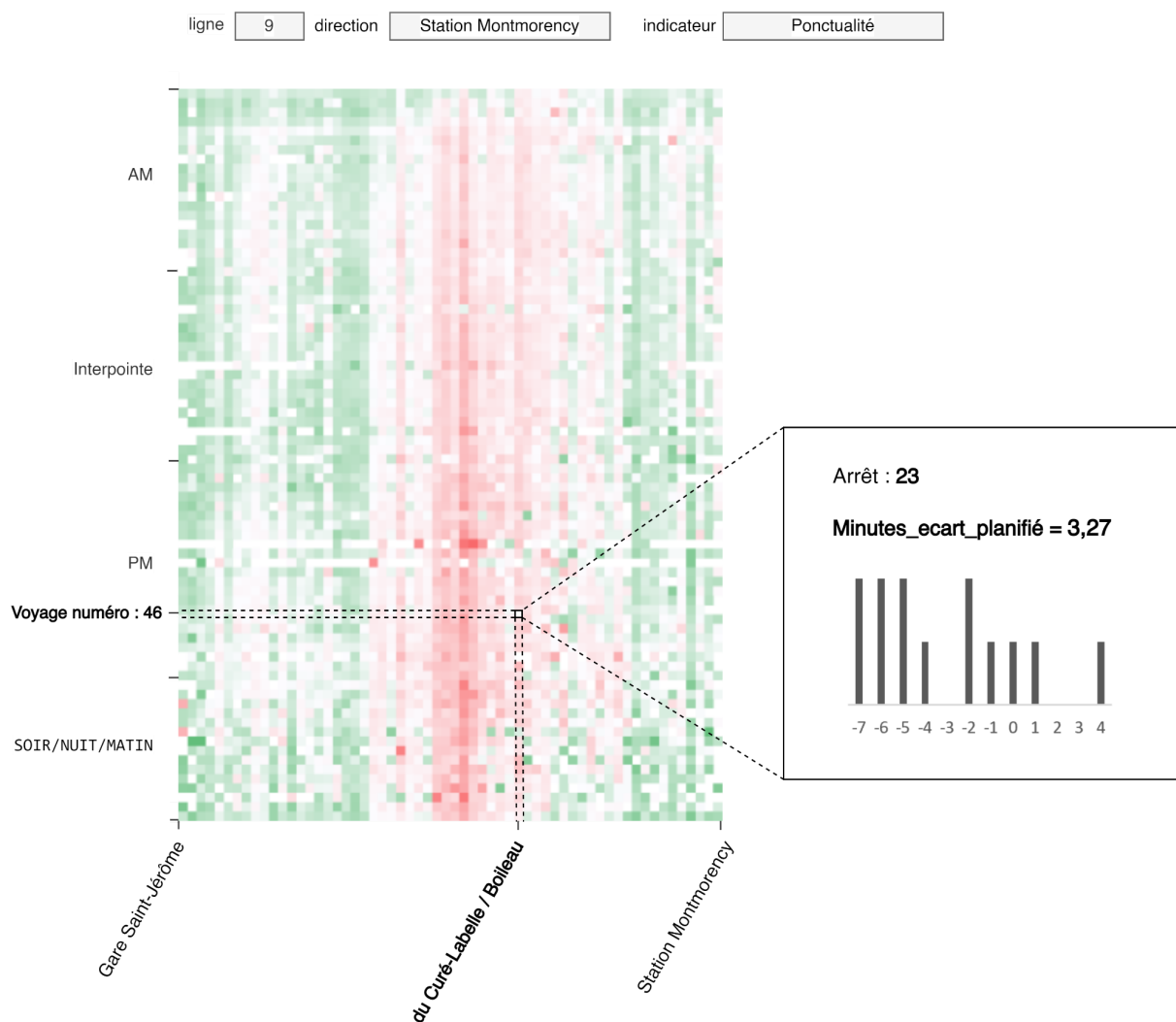
Dans cette visualisation, il y a principalement 2 interactions possibles pour l'utilisateur.

La première est le changement de mode : comme décrit plus haut, en cliquant sur un bouton, l'utilisateur peut passer du mode par défaut, au mode ponctualité, au mode achalandage.

L'autre fonctionnalité est le survol : quand l'utilisateur passe la souris sur un rectangle de la carte de chaleur, un tooltip s'affichera. Le tooltip renseigne les informations suivantes:

- la moyenne de la valeur visualisé
- Arrêt: numéro de l'arrêt
- Nom de l'arrêt
- numéro du trajet
- Histogramme de la valeur visualisé (x: valeur visualisé, y: nombre de journées)

Dans le mode par défaut, la valeur visualisée est l'indicateur de ponctualité. Dans le mode achalandage, c'est l'achalandage, c'est-à-dire le nombre de personnes montées à cet arrêt. Dans le mode retard, c'est les minutes d'écart, c'est-à-dire le temps de retard enregistré par rapport au temps supposé.



4.2 Visualisation 2

La deuxième visualisation réponds au questions suivantes :

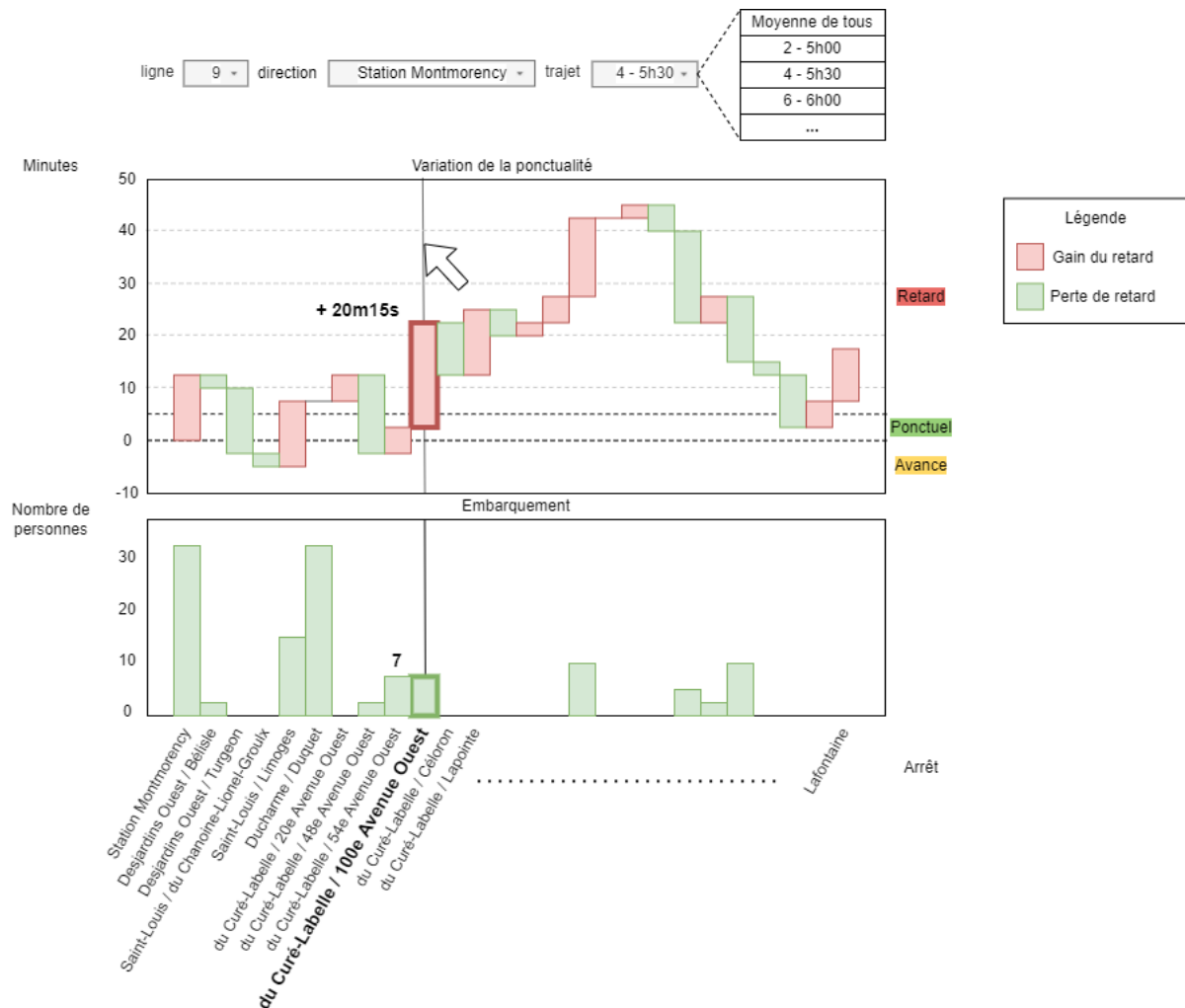
2	Quel est le retard moyen?
4	Est-ce qu'un autobus peut rattraper son retard après un arrêt problématique?
6	Existe-t-il des relations entre le retard et le nombre de passagers qui embarquent à chaque arrêt?
8	Quel est le retard relatif gagné ou perdu à chaque arrêt?
10	Quels sont les arrêts ou les tronçons les plus problématiques?

Description

La visualisation est composée de 2 diagrammes à bandes interagissant entre eux. Les 2 diagrammes sont affichés ensemble pour permettre de comparer et/ou trouver des liens entre la ponctualité et l'achalandage, alors ils partagent le même axe horizontal, soit les arrêts. L'utilisateur doit spécifier la ligne ainsi que la direction. Par défaut, les données affichées sont la moyenne à chaque arrêt, mais on peut aussi spécifier d'utiliser la moyenne de la semaine, de la fin de semaine ou même choisir un trajet en particulier. Le premier diagramme à pour but plus spécifique d'indiquer la variation de retard moyen à chaque arrêt. De plus, il est possible de comprendre si un autobus peut ou non rattraper son retard suite à un arrêt problématique. Il sera aussi possible de définir le retard relatif gagné ou perdu à chaque arrêt et ainsi déterminer que les arrêts sont les plus problématiques selon le trajet. Le second diagramme à pour but d'afficher le nombre de personnes qui embarquent en moyenne à chaque arrêt. Avec cette présentation, il sera possible de définir si un lien existe entre le retard d'un autobus à un arrêt et le nombre de passagers qui embarquent à cet arrêt.

Interactions

Dans cette visualisation, plusieurs interactions seront possibles d'être effectuées par l'utilisateur. Tout d'abord, celui-ci pourra sélectionner plusieurs paramètres changeant les données présentées. Il pourra sélectionner la ligne d'autobus, la direction ainsi que le trajet pour lesquels il souhaite observer les données. Par la suite, l'utilisateur pourra passer son curseur au-dessus d'une ligne en particulier. Cela permettra de focaliser sur cette ligne et d'afficher le gain ou le retard précis de cet arrêt, ainsi que le nombre précis de personnes qui embarquent à cet arrêt. Cela permettra de faire un lien direct entre les deux données.



4.3 Visualisation 3

La troisième visualisation réponds au questions suivantes :

2	Quel est le retard moyen?
3	Quel est l'achalandage moyen?
11	Est-ce qu'un sens du trajet est plus ou moins problématique que l'autre?

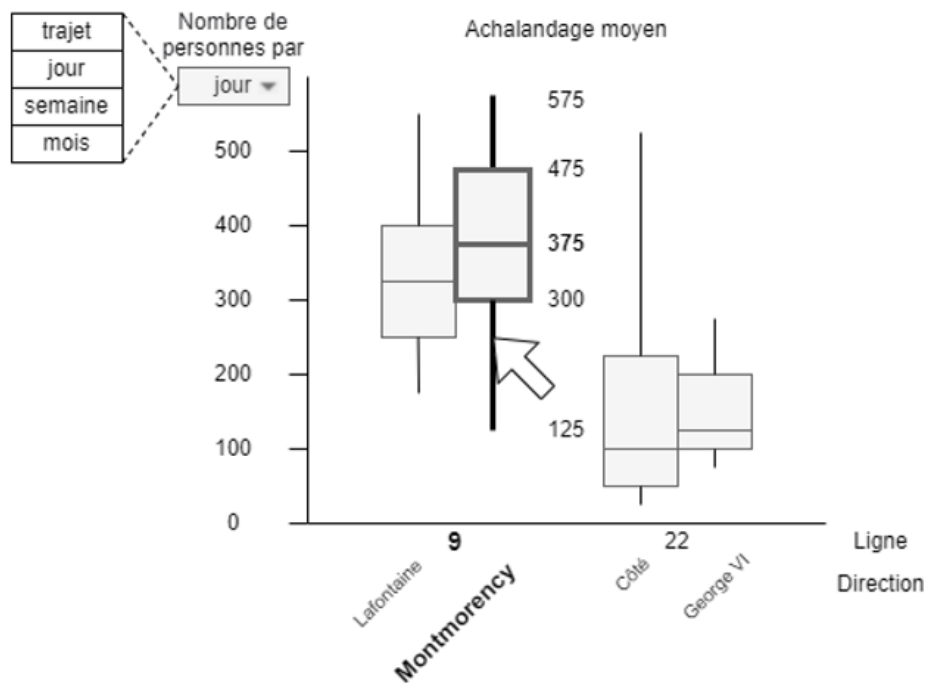
Description

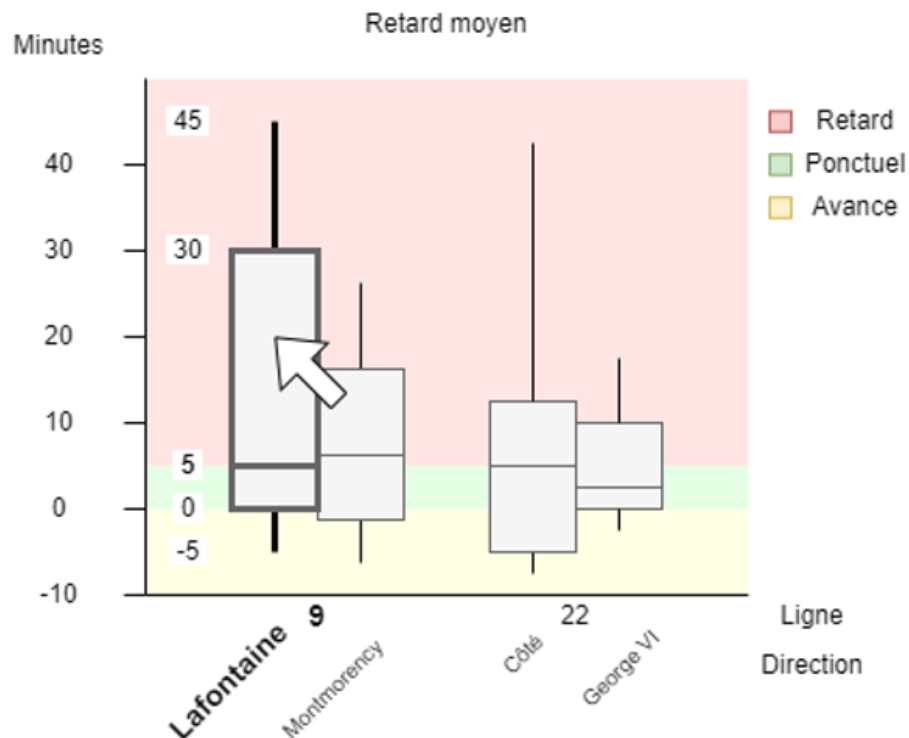
La visualisation est composée de 2 diagrammes de quantile groupé, un pour l'achalandage et l'autre pour la ponctualité. Chaque groupe représente une ligne et chaque groupe est composé des 2 directions possibles. Chaque élément est une bougie de quantile (0, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$ & 1) qui donne une bonne idée de la médiane et de la distribution des données. Cette visualisation sert à donner un aperçu de très haut niveau des lignes. Puisqu'il y a peu de lignes, les 2

diagrammes seront affichés l'un à côté de l'autre (horizontalement), pas un en dessous de l'autre.

Interactions

Lorsque l'utilisateur passe sa souris sur une bougie, celle-ci est mise en gras, ainsi que le numéro de la ligne et la direction sur l'abscisse. Aussi, la valeur des quantiles est affichée à côté de leur barre correspondante. La médiane (Q2) est mise en gras. Dans la visualisation d'achalandage, il est possible de sélectionner la durée (nombre de personnes par trajet/jour/semaine/mois en moyenne).





4.4 Style & arrangement

Style

Le style utilisé sera le même que Exo, c'est à dire utiliser la même police (Myriad) et les même couleurs (blanc, noir, turquoise #1F96A4, pourpre #8B304E).

Arrangement

Les visualisations sont arrangées dans un simple tableau de bord qu'on peut faire défiler verticalement. Il y a deux sections: une section «Survol» qui montre la visualisation 3 et une autre section «Exploration» qui montre la visualisation 1 puis 2. La section «Exploration» contient 2 menus déroulant permettant de sélectionner la ligne et la direction. La sélection met à jour les 2 visualisations avec les données correspondantes.