

CLASSE DE MAÎTRE SUR LE DATA STORYTELLING

LA VULGARISATION DES DONNÉES

Présentation dans le cadre des journées de
philanthropie des données

IVADO & AIMIA

Véronique Dufort

LABORATOIRE D'INNOVATION URBAINE DE MONTRÉAL

6 juin 2017

Montréal 

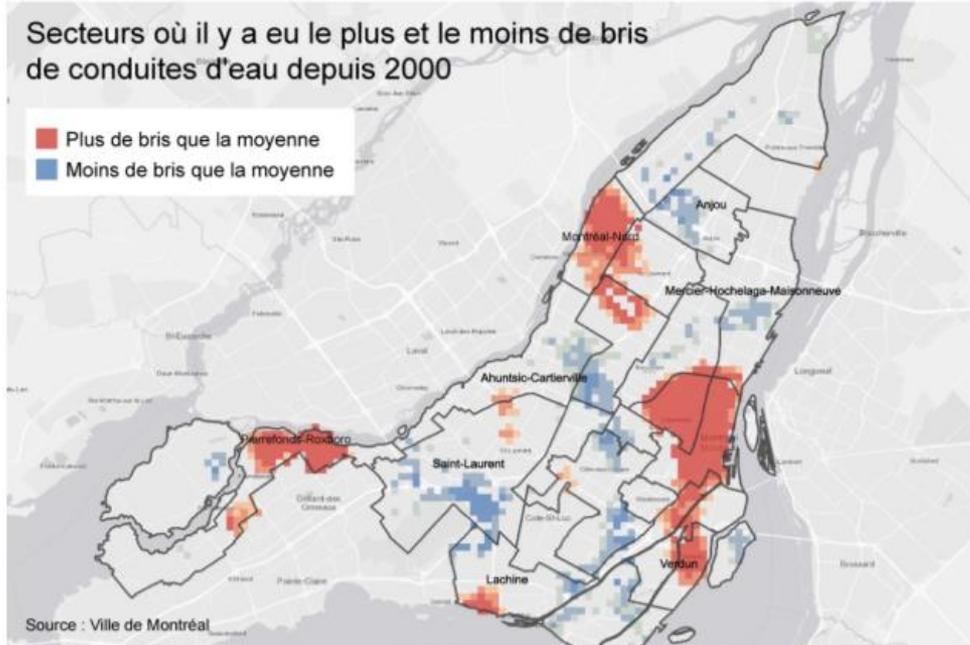
MISE EN CONTEXTE MONTRÉALAIS

- Plan d'actions Ville intelligente et numérique avec les données ouvertes comme pilier central en mai 2015
- Adoption d'une **Politique de données ouvertes** et d'une **Directive sur la gouvernance des données** en décembre 2015
 - PORTÉE ORGANISATIONNELLE
 - PRINCIPES DIRECTEURS DE L'ACCÈS AUX DONNÉES
 - LIMITES À L'OUVERTURE (CATÉGORIE D'EXCEPTION)
- Transformation du BVIN en Laboratoire d'innovation urbaine mai 2018 où la culture autour du partage des données est toujours au coeur de la démarche

POURQUOI LES DONNÉES OUVERTES?

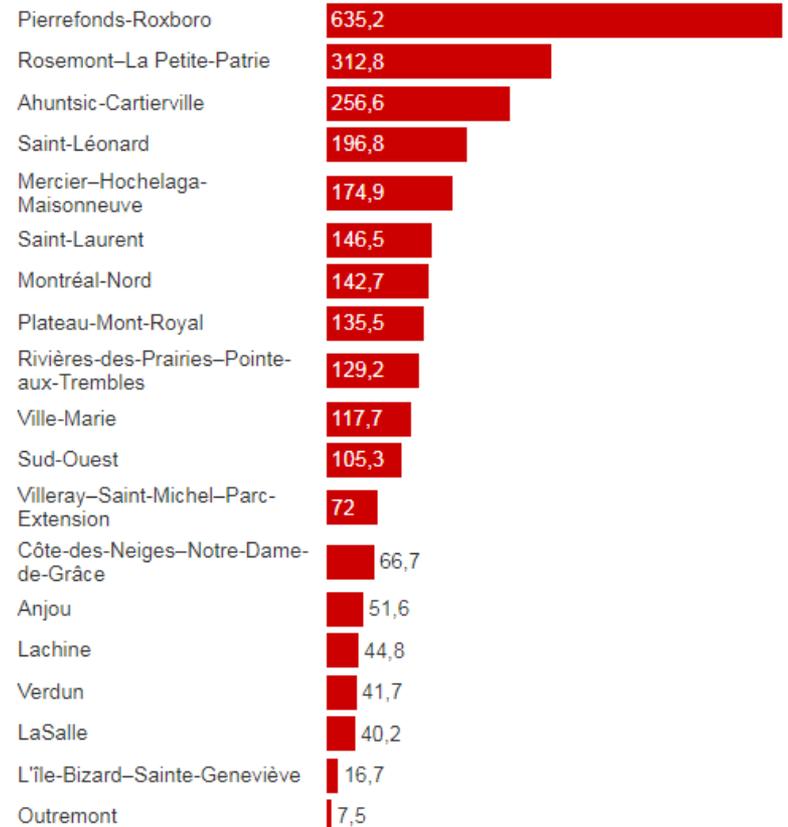
- **Efficacité organisationnelle**
 - Favoriser la collaboration entre services municipaux sur base des données
 - Encourager la collaboration multi-paliers: local (villes), régional (agglo, CMM), national (provincial, fédéral)
- **Développement économique**
 - Favoriser le développement de nouveaux outils ou de nouvelles entreprises basés sur l'économie du savoir
- **Service aux citoyens**
 - Aider les entreprises à développer des outils à valeur ajoutée pour les citoyens
- **Transparence gouvernementale**
 - Permettre aux citoyens de mieux comprendre le fonctionnement de l'organisations
 - Augmenter la confiance envers les institutions publiques grâce à une plus grande imputabilité
- **Engagement citoyen**
 - Possibilité d'utiliser les données comme base d'échange (ex. dans le cadre de consultations)

EXEMPLE DONNÉES OUVERTES SUR LES RÉPARATIONS DU RÉSEAU D'EAU POTABLE



Secteurs les plus et les moins touchés par des bris de conduites d'eau à Montréal Photo : Radio-Canada

Bris de conduites d'eau par 1000 habitants



Auteur: Roberto Rocha

<https://ici.radio-canada.ca/nouvelle/1085101/conduites-eau-montreal-bris-aqueduc-infrastructures>

CONCEPTS

Qu'est ce que la vulgarisation?

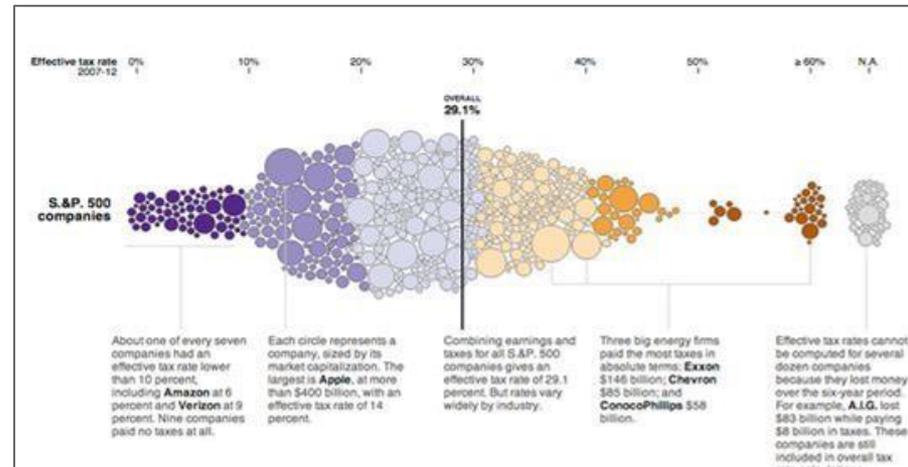
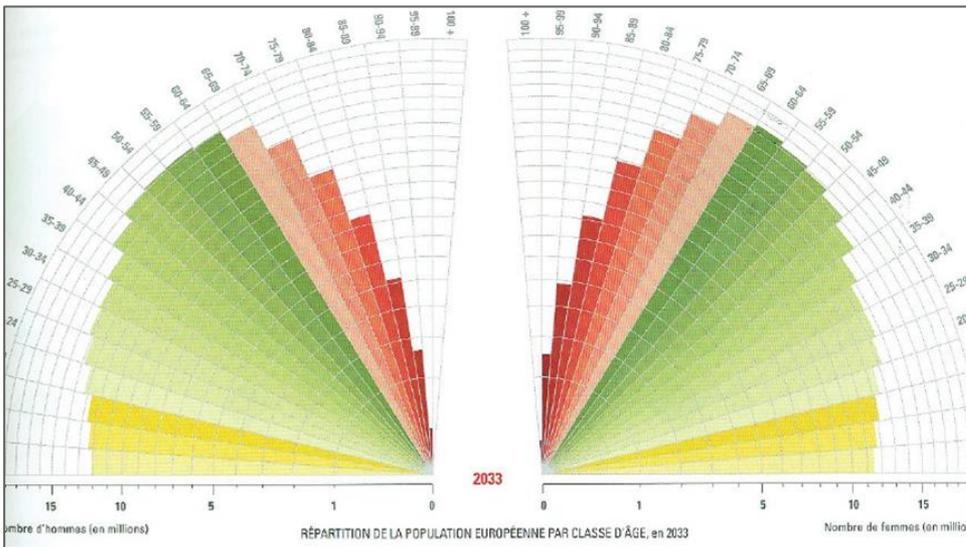
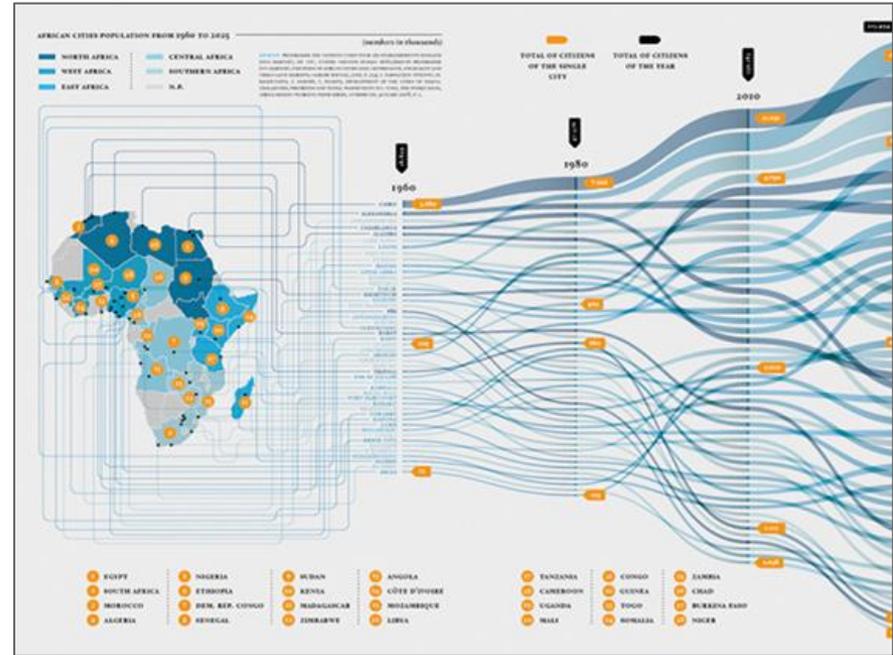
“Action de mettre à la portée du plus grand nombre, des non-spécialistes, des connaissances techniques et scientifiques.”

LAROUSSE

Contexte données:

- Pouvoir des données en 2018
- La capacité à exploiter les données pour être capable de les comprendre, de les traiter, d'en extraire de la valeur, de les visualiser et de les communiquer
- Importance de rendre accessible les données dans le contexte
- Les données comme outil de communication

EXEMPLES DIVERS - Atteinte de l'objectif d'accessibilité?



CONCEPTS

Composantes clés du “Data storytelling”:

- Données
- Visuels
- Narratif : contexte est important à expliquer!
 - La visualisations de données est un moment dans le temps alors qu’une histoire de données concerne une séquence dans le temps
 - Une histoire bien racontée **ENGAGE L’ÉMOTION**
 - Une histoire par sa nature tend vers une résolution

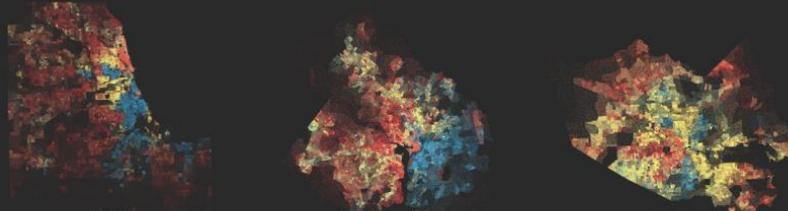
EXEMPLE

America is more diverse than ever – but still segregated

The United States is on track to be a majority-minority nation by 2044. But census data show most of our neighbors are the same race.

By [Aaron Williams](#) and [Armand Emamdjomeh](#) Updated May 10, 2018

- Black
- White
- Hispanic
- Asian/Pacific Islander
- Native American
- Multi-race and other



Legacy segregation

Diverse suburbs

Rapidly integrated

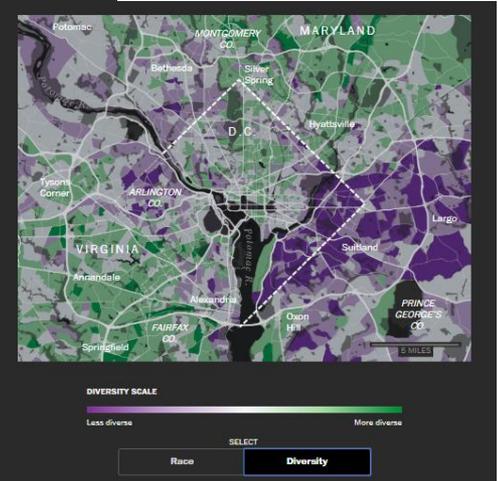
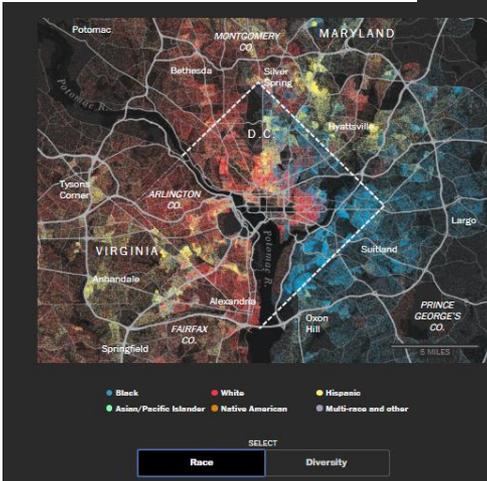


Why segregation persist

Since 1990, a majority of U.S. metro areas have seen increases in racial diversity. While this means that more Americans are living in diverse neighborhoods, the numbers are still lower than researchers anticipated.

“There have been declines [in segregation], but they just haven’t been as fast as we would expect,” said Kyle Crowder, a professor of sociology at the University of Washington. He and Maria Krysan, a professor of sociology at the University of Illinois at Chicago, recently published their research on residential stratification in a book titled “Cycle of Segregation.”

In the book, they argue that although segregation is generally decreasing, factors such as social networks and communities play a large role in keeping segregation embedded in American life.



OBJECTIF ET PUBLIC CIBLE

- Bien connaître l'audience pour trouver le ton juste
 - Créer des personas
 - Éviter d'être trop technique
 - Alléger en fonction de l'audience visé
- Déterminer l'objectif (réflexion stratégique)
 - Quel est le mandat?
 - Quel enjeux on souhaite adresser?
 - Quel message on veut passer? Quel impact on veut avoir?

BUT: Faire comprendre ou convaincre l'auditoire. Présenter les données pour mieux mémoriser, **influencer**, engager ou emmener des changements.

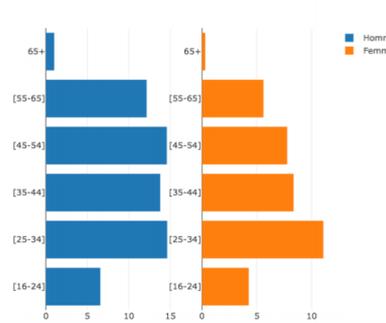
→ **EMPOWERMENT** de l'AUDIENCE --> mobiliser pour prendre action!

EXEMPLE DE PRÉSENTATION DE DONNÉES AVEC DES BUTS DIFFÉRENTS - MTL TRAJET (1/2)

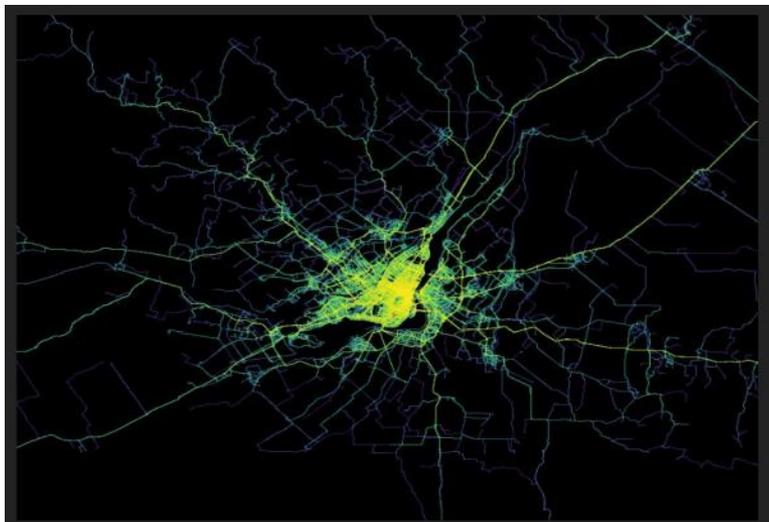
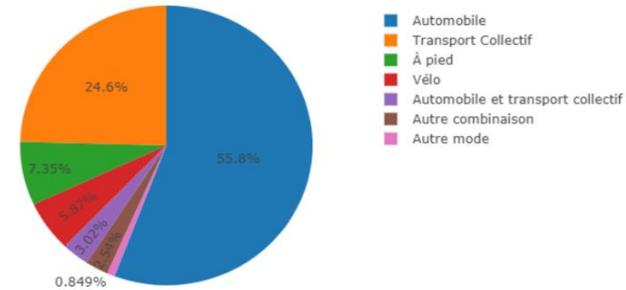
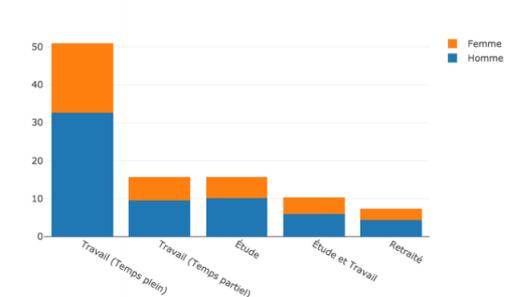
Exemple VILLE



Distribution des usagers par sexe et age (%)



Distribution des usagers par occupation (%)

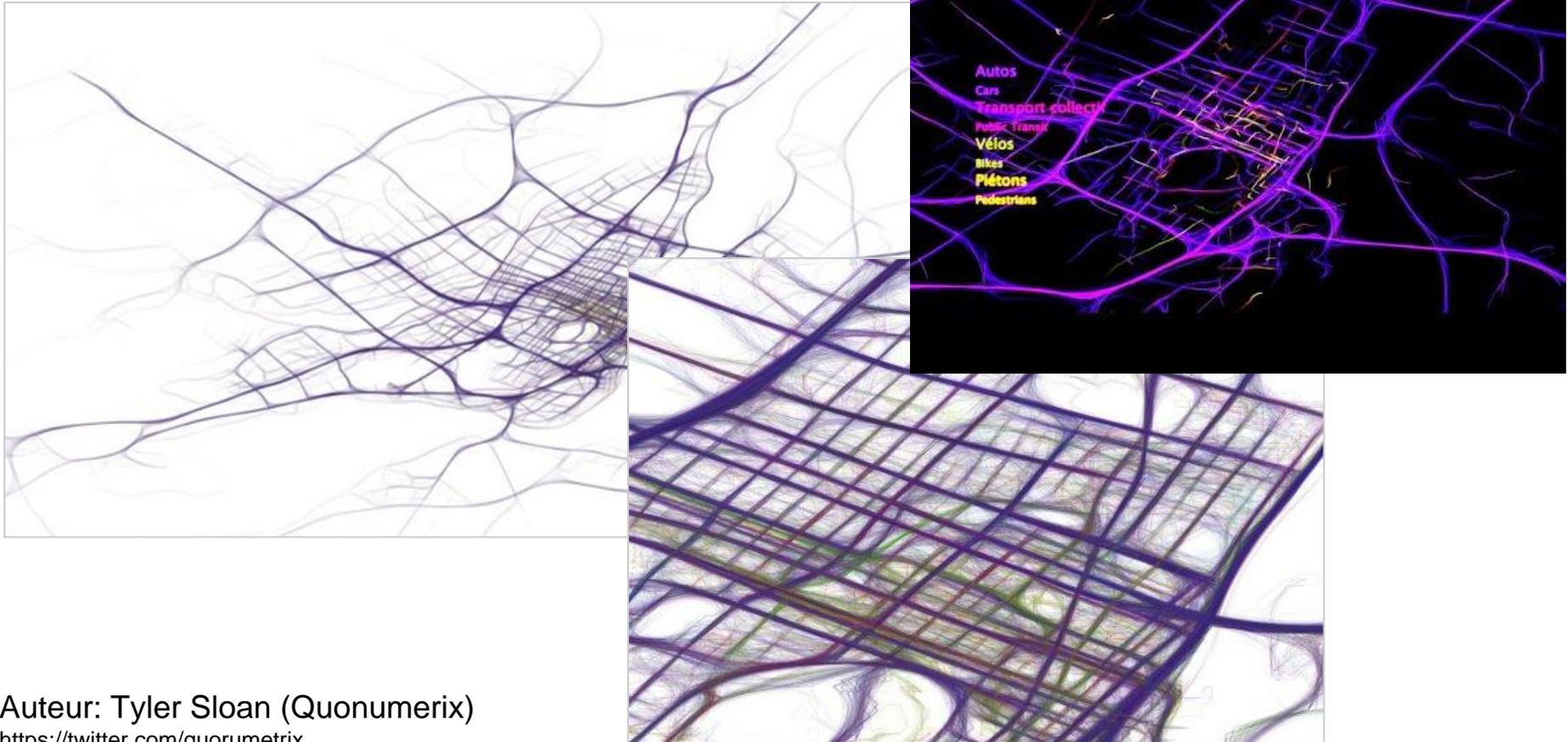


▸ L'avenue Mont-Royal ressort fortement comme un axe piéton par rapport aux rues avoisinantes.

EXEMPLE DE PRÉSENTATION DE DONNÉES AVEC DES BUTS DIFFÉRENTS - MTL TRAJET (2/2)

Exemple Citoyen (intérêt en visualisation: “exploring the space between data science and art”)

1 month of commuting of cars, cyclists, transport and pedestrians.



Auteur: Tyler Sloan (Quonumerix)
<https://twitter.com/quorumetrix>

MISE EN FORME

Outil de communication

Lien entre des données complexes et la cible par la mise en forme

- Commencez avec une **question**
 - Commencez avec un problème convaincant, pas une explication de la technologie
 - Possibilité d'utiliser les personas dans le récit
- Terminez avec une **vision**, le message clé, une fin inspirante
- Expliquez avec des visuels, Racontez avec des mots
 - Ne pas présenter seulement les résultats
 - Utilisez des visualisations pour donner un sens aux données
- Soyez **honnête** et crédible
 - Éviter les biais
 - Partagez VOTRE expérience
- Soyez **clair et concis**
- Utilisez le potentiel de la **cartographie**, l'humain aime se situer dans l'espace!

EXAMPLE: Film Dialogue analysis, Source THE PUDDING

<https://pudding.cool/2017/03/film-dialogue/>

Lately, Hollywood has been taking so much shit for rampant sexism and racism. The prevailing theme: **white men dominate movie roles.**

But it's all rhetoric and no data, which gets us nowhere in terms of having an informed discussion. How many movies are actually about men? What changes by genre, era, or box-office revenue? What circumstances generate more diversity?

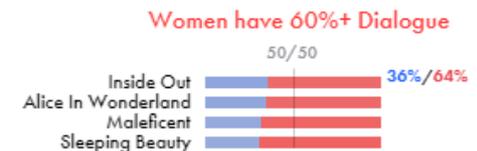
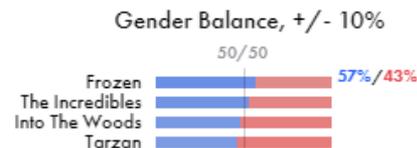
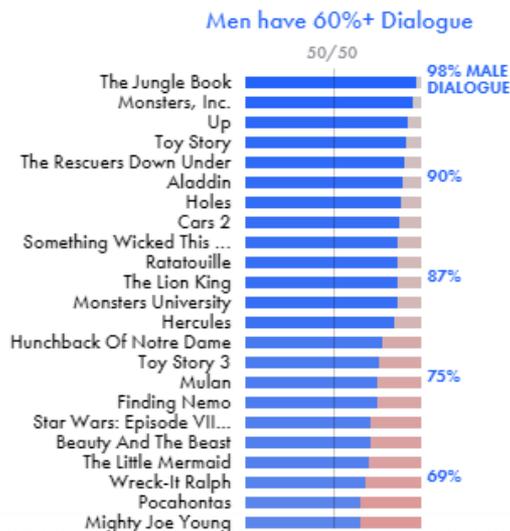
We didn't set out trying to prove anything, but rather compile *real data*. We framed it as a census rather than a study. So we Googled our way to 8,000 screenplays and matched each character's lines to an actor. From there, we compiled the number of words spoken by male and female characters across roughly 2,000 films, arguably the largest undertaking of script analysis, ever.

Let's begin by examining dialogue, by gender, for just Disney films.

Disney Screenplay Dialogue, Broken-down by Gender

2,000 Screenplays: Dialogue Broken-down by Gender

Only High-Grossing Films: Ranked in the Top 2,500 by US Box Office*



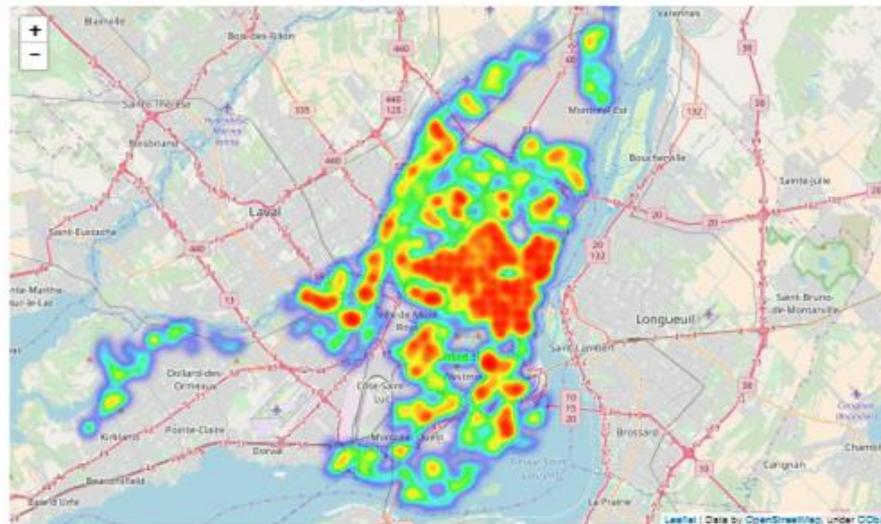
EXEMPLE: Analyses sur les services 311 (requêtes) en lien à la salubrité (Ville de Montréal)

Analyse croisée sur les punaises de lit

Nous avons mis en relation les données d'extermination de punaises de lit, disponibles en données ouvertes (voir extrait ci-dess), avec celles des DDS relatives aux punaises de lit.

nbr_extermin	date_fin	arrondissement	latitude	longitude	
0	1	2001-07-25	Saint-Laurent	45.517411	-73.696415
1	2	2001-07-27	Ahuntsic - Cartierville	45.546973	-73.659862
2	1	2011-01-18	Mercier - Hochelaga-Maisonneuve	45.542364	-73.546521
3	1	2011-01-21	Rosemont - La Petite-Patrie	45.560050	-73.576694
4	1	2011-01-25	Mercier - Hochelaga-Maisonneuve	45.507411	-73.523624

Le nombre total d'extermination de punaises de lit, de 2011 à 2016, est illustré sur la carte:



Nous avons voulu vérifier si les données du 311 sont représentatives des phénomènes observés sur le territoire. Par exemple, si les citoyens d'un arrondissement formulent beaucoup de DDS à propos des punaises de lit, est-ce symptomatique d'un problème de punaises dans cet arrondissement?

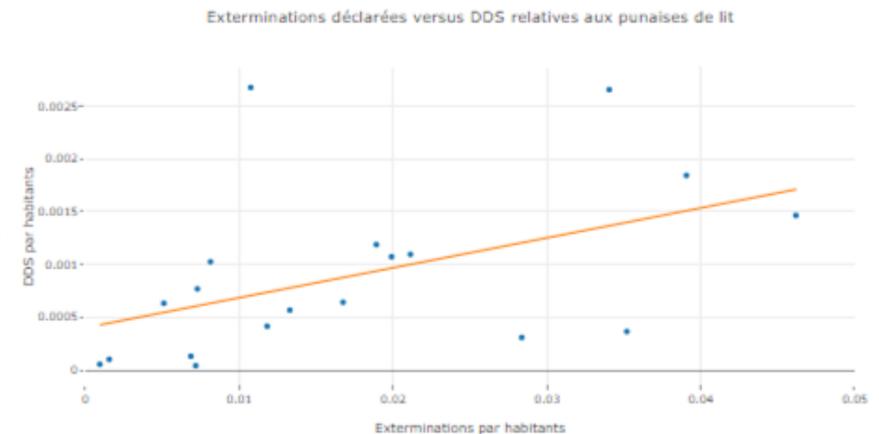
Les données DDS et d'extermination de punaises de lit sont mis en relation dans le tableau suivant. Les valeurs ont été normalisées par le nombre de citoyens de l'arrondissement.

Nous avons voulu vérifier si les données de 311 sont représentatives des phénomènes observés sur le territoire. Par exemple, si les citoyens d'un arrondissement formulent beaucoup de DDS à propos des punaises de lit, est-ce symptomatique d'un problème de punaises dans cet arrondissement?

Les données DDS et d'extermination de punaises de lit sont mis en relation dans le tableau suivant. Les valeurs ont été normalisées par le nombre de citoyens de l'arrondissement.

	arrondissement	exterminations	exterminations normalisées	DDS	DDS normalisées
17	Ville-Marie	3680	0.048183	123	0.001464
8	Le Plateau-Mont-Royal	3622	0.039068	166	0.001843
10	Rosemont - La Petite-Patrie	4717	0.035152	49	0.000366
8	Mercier - Hochelaga-Maisonneuve	4477	0.034050	349	0.002694
18	Villeray-Saint-Michel - Parc-Extension	4034	0.028364	44	0.000309
0	Ahuntsic - Cartierville	2681	0.021128	139	0.001095
8	Montréal-Nord	1669	0.019900	90	0.001073
7	Le Sud-Ouest	1353	0.018911	65	0.001166
2	Côte-des-Neiges - Notre-Dame-de-Grâce	2764	0.016748	106	0.000642
16	Saint-Léonard	1006	0.013288	43	0.000566
14	Saint-Laurent	1109	0.011818	39	0.000416
18	Verdun	711	0.010747	177	0.002675
1	Anjou	341	0.008133	43	0.001026
6	Lachine	303	0.007281	32	0.000769
10	Outremont	169	0.007171	1	0.000042
12	Rivière-des-Prairies - Pointe-aux-Trembles	730	0.006869	14	0.000132
4	LaSalle	379	0.005103	47	0.000633
11	Pierrefonds - Roxboro	106	0.001549	7	0.000102
9	L'Île-Bizard - Sainte-Geneviève	17	0.000939	1	0.000055

En mettant en relation les DDS et les exterminations normalisées pour chaque arrondissement, on obtient le nuage de points suivant:

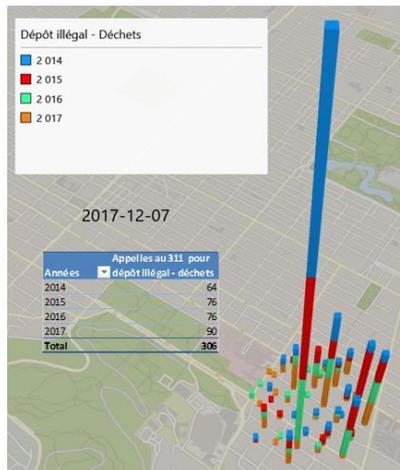


On observe une faible relation linéaire ($R^2 = 0.23$) entre les DDS et les exterminations. En d'autres mots, pour la plupart des arrondissements, plus il y a de DDS à propos des punaises de lits, plus il y a d'exterminations.

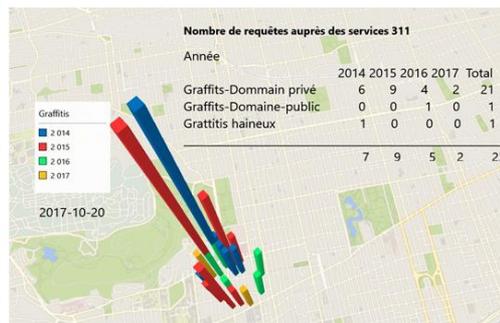
EXEMPLE: Projet citoyen : “ Qu’est-ce c’est passé ici ?”

Regardons dans les données ouvertes pour voir ce qui c’est passée ici

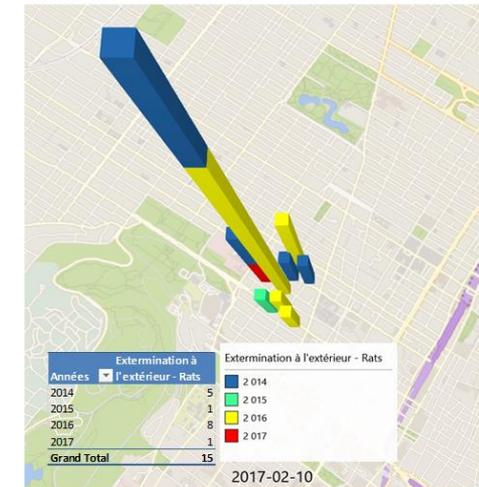
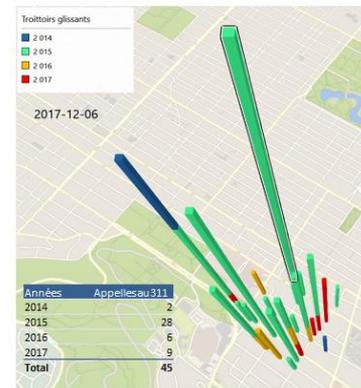
(à 300 mètres de la latitude 45.511078 et la longitude -73.574651)



Requêtes au services 311



Requêtes au services 311



SYNTHÈSE ET SIMPLICITÉ

“Je vous écris une longue lettre parce que je n'ai pas le temps d'en écrire une courte.”

Blaise Pascal

Alléger la présentation pour atteindre son objectif : un effort supplémentaire!

- Ne pas montrer l'effort d'analyse pour influencer
- “Une image vaut milles mots” pour engager l'émotion: Utiliser des visualisation pour donner un sens aux données.
- Utiliser des visualisations simples et accessibles à votre public

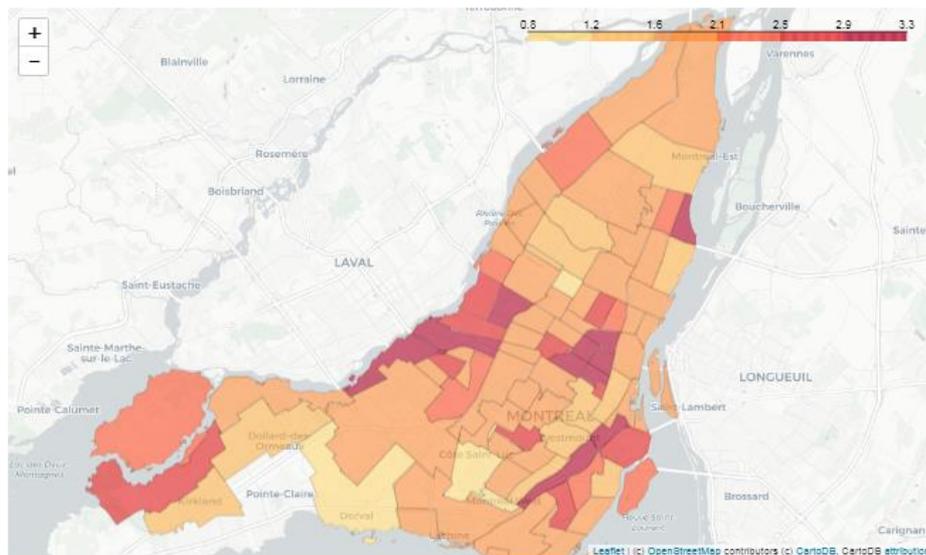
Approche de design de visualisation

- **objectif:** pourquoi la visualisation -> l'objectif doit être spécifique
- **contenu:** quoi visualiser -> seulement ce qui est nécessaire
- **structure:** comment visualiser -> Utiliser le bon support visuel pour la démo
- **formatage :** apparence et focus -> Éviter les distractions

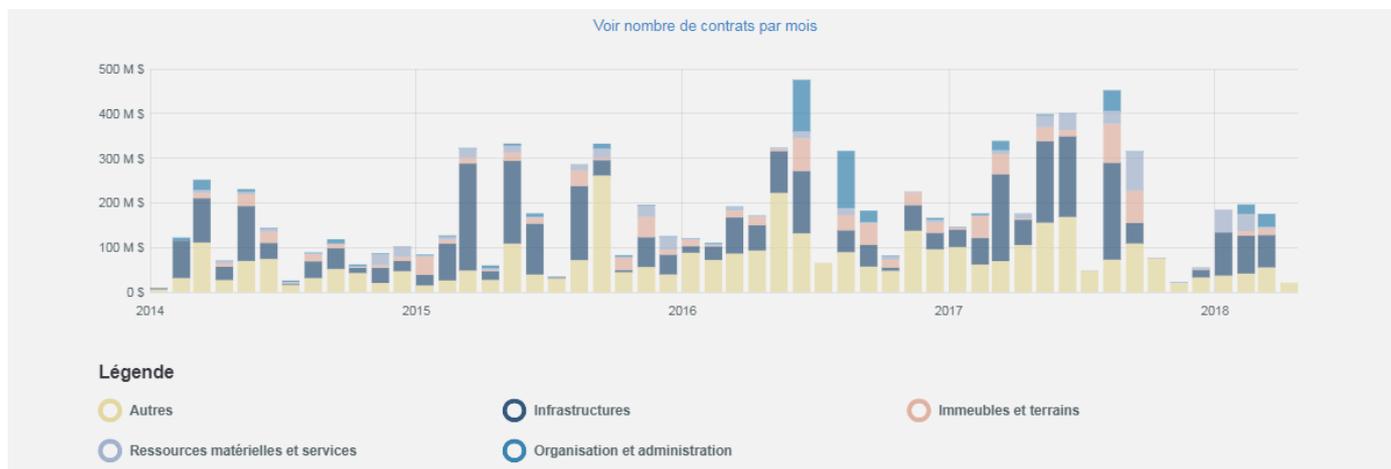
EXEMPLES:

ANALYSES SUR LES PÉRIODES DE DÉNEIGEMENT

Durée médiane des opérations (jours) pour les 4 périodes:

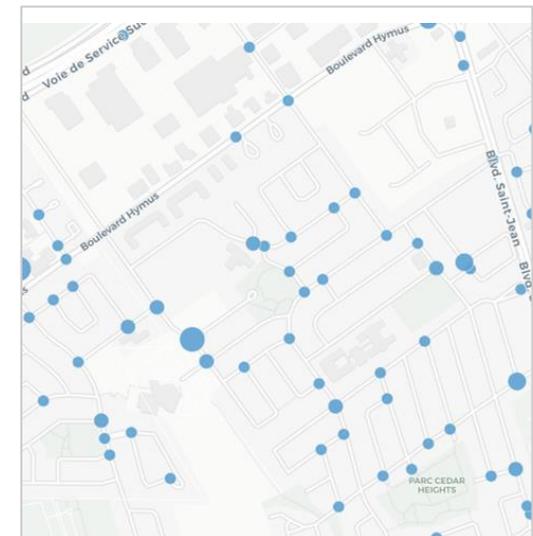
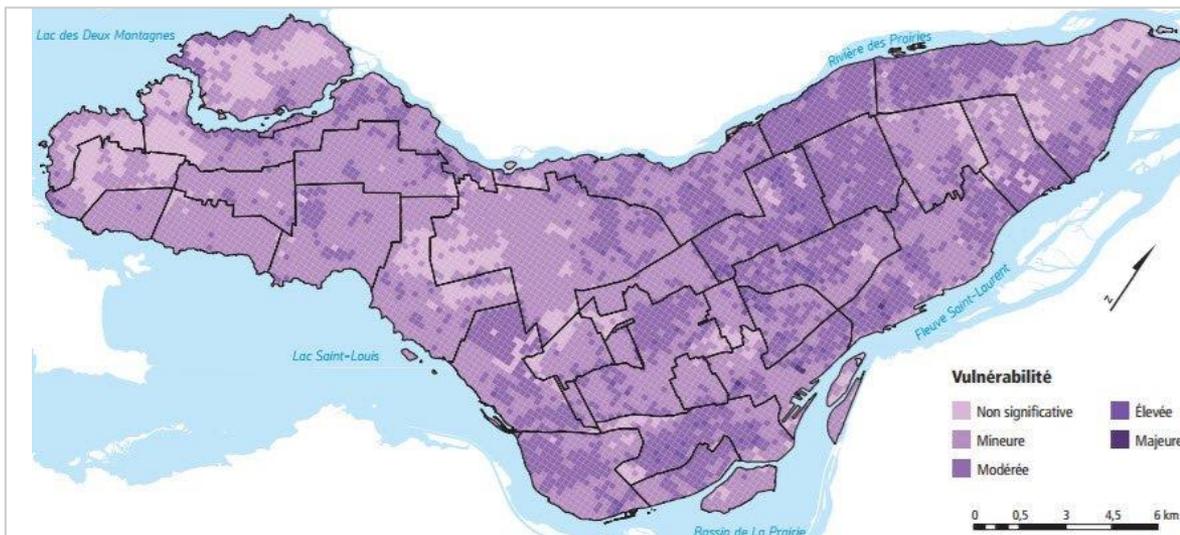
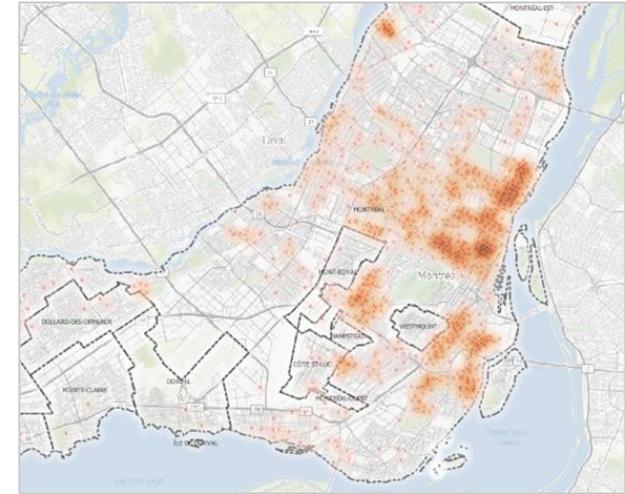


VUE SUR LES CONTRATS (VILLE DE MONTRÉAL)



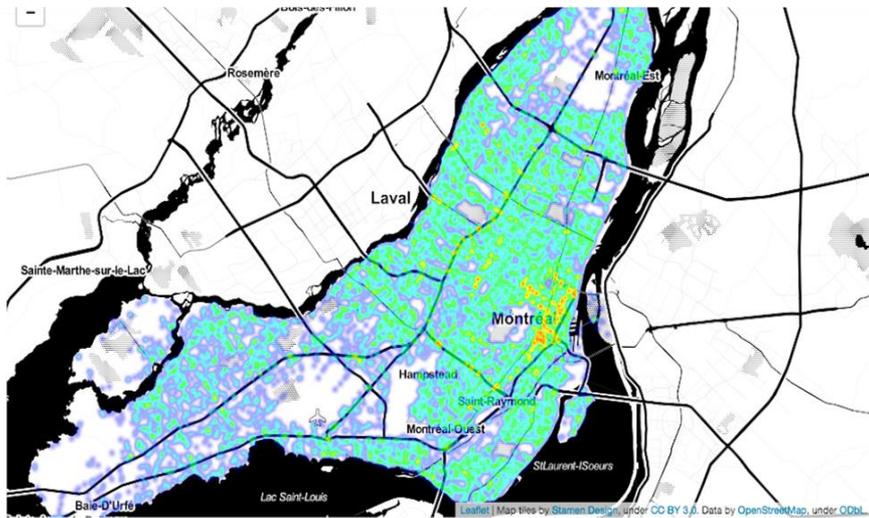
ENRICHIR LES DONNÉES

- Utilisation données ouvertes (ou autres) pour bonifier les analyses ou contextualiser
- Agréger les données géographiques sur un unité visuel adéquat

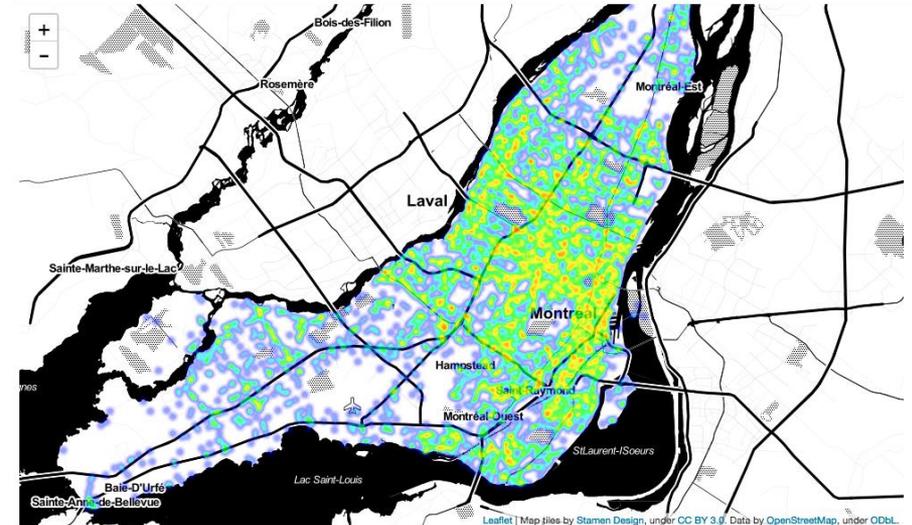


EXEMPLES - COLLISIONS ROUTIÈRES

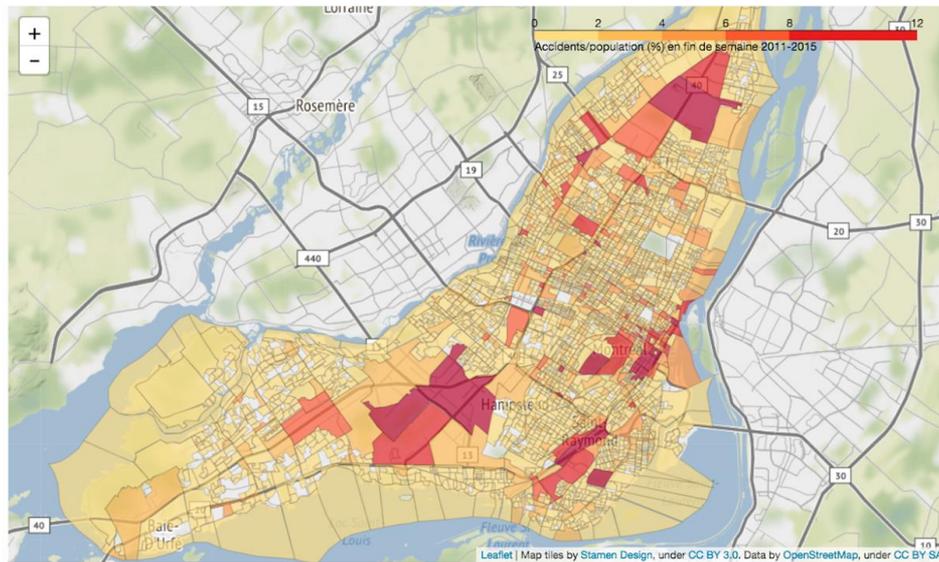
Les précipitations (neige, pluie) causent des accidents partout sur le territoire, indépendamment du trafic



Tous les accidents

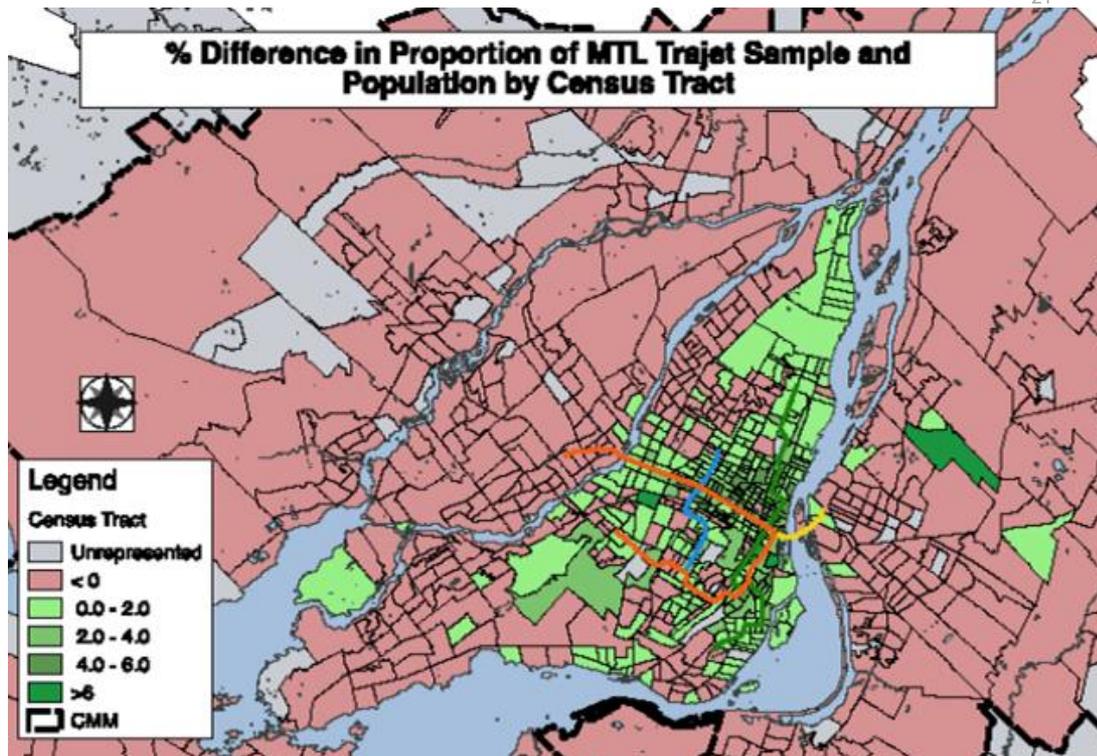
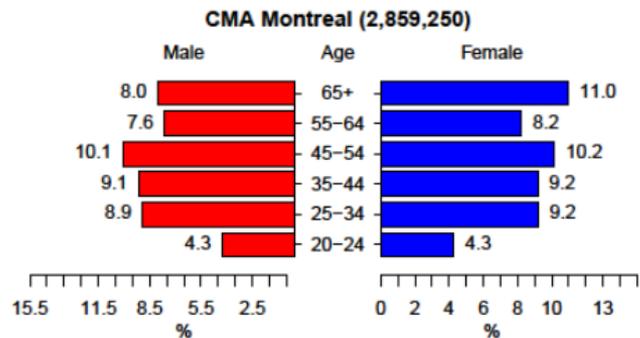
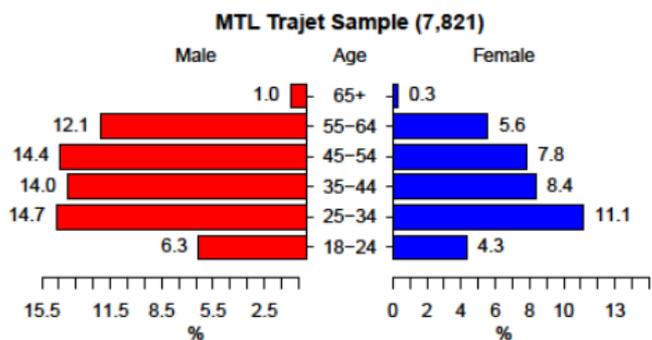


Tous les accidents lors de précipitations de neige ou pluie



Nombre d'accidents normalisés par le nombre de citoyens de l'aire de diffusion (Statistique Canada)

EXEMPLES MTL TRAJET

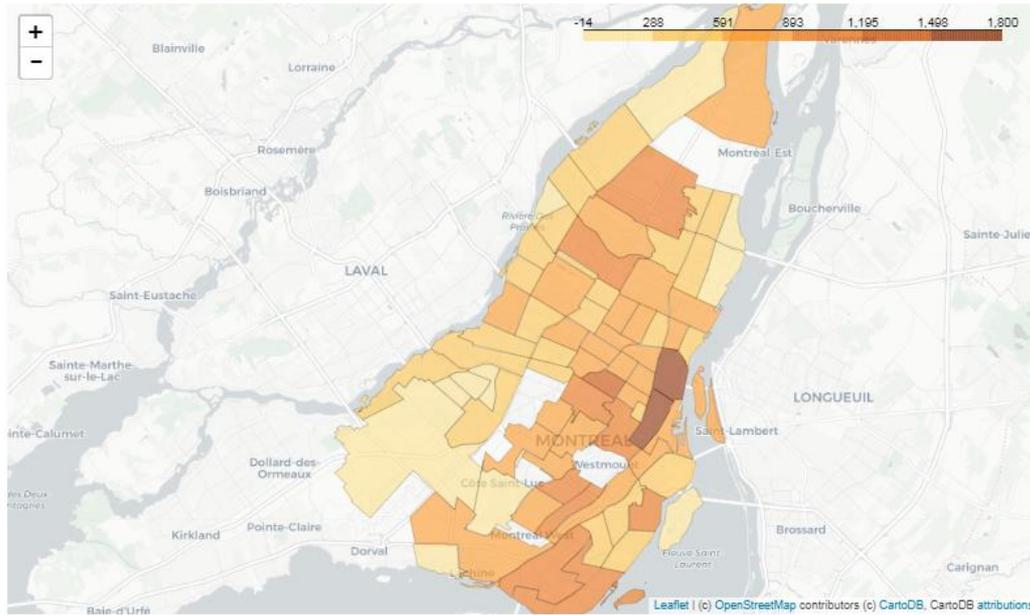


Mode	MTL Trajet	OD2013	MTL Trajet (%)*	OD2013 (%)*
Vélo	743	26364	11,1	2,2
Voiture	2886	802239	43,1	68,4
Autre combinaisons	116	2833	1,7	0,2
Autre unique	40	4731	0,6	0,4
TC**	2242	240575	33,5	20,5
TC + Voiture	371	43334	5,5	3,7
Marche	297	53543	4,4	4,6

* Les pourcentages peuvent ne pas s'additionner à 100% en raison de l'erreur d'arrondissement.

** TC: Transport en Commun

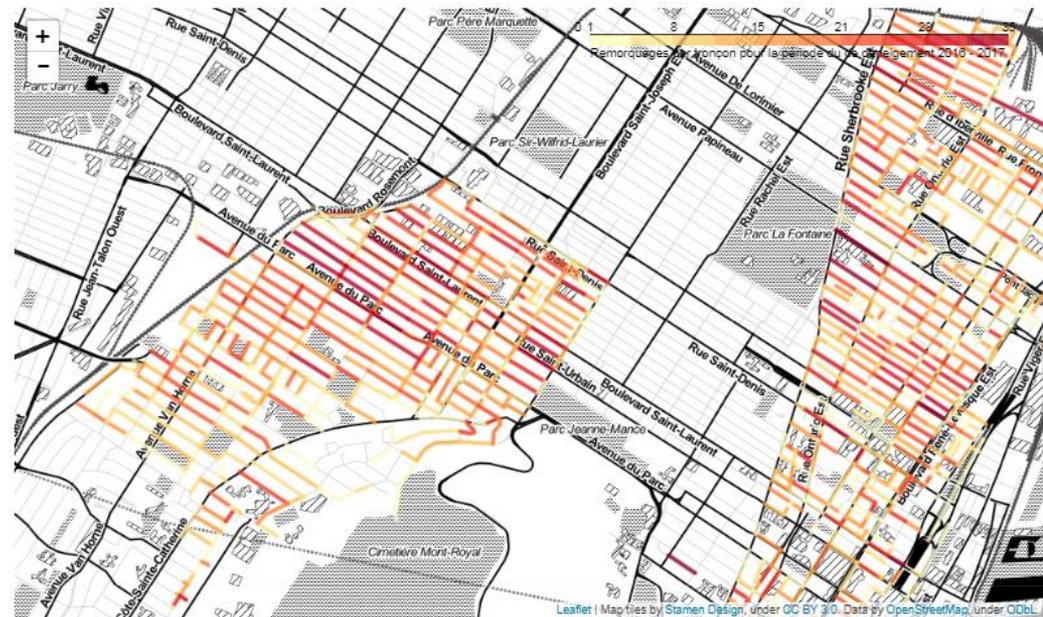
- ▶ MTL Trajet vs enquête OD 2013
- ▶ Utilisation de la voiture moins commun dans le cas de MTL Trajet
- ▶ Le vélo et le transport en commun sont plus élevés dans le cas de MTL Trajet
- ▶ Distance moyenne domicile-travail:
 - ▶ MTL Trajet: 9,331 km
 - ▶ OD 2013: 10,868km



EXEMPLES ANALYSE DU REMORQUAGE À MONTRÉAL

Les 10 quartiers comptant le plus de remorquages:

Détails des tronçons des 5 quartiers comptant le plus de remorquages ainsi que les 10 tronçons les plus occupés:



MERCI DE VOTRE ATTENTION!

Véronique Dufort, ing
Conseillère - Données ouvertes
Laboratoire d'innovation urbaine
Ville de Montréal
veronique.dufort@ville.montreal.qc.ca