

Colloque
**POINT DE MIRE SUR LES
TRAVAUX DE LA DIRECTION
GÉNÉRALE DU LABORATOIRE
DES CHAUSSÉES DU MTQ**

19 Novembre 2019

La contribution de la science des données aux analyses de suivi de performance

Anne-Marie Boulé, statisticienne

Direction générale du laboratoire des chaussées

Ministère des Transports du Québec

Plan de la présentation

- Contexte du projet
- Méthodes de la science des données
 - *Clustering*
 - Arbres de classification
 - Forêts aléatoires
- Conditions de succès
- Ce qui reste à travailler

Contexte du projet

Suivi de performance

- Comparaison de travaux routiers réalisés

Sélection de travaux pour le groupe 1

Sélectionné	No. d'enregistrement	Année des travaux	RTSS	Direction	Voie	Ch. début	Ch. fin	Longueur (km)	Intervention	Axe budgétaire	Coût total travaux (\$)	Superf. totale travaux (m2)	Coût intervention (\$)	Date ouverture trafic	Type de liant	RSSCE	Nr
<input checked="" type="checkbox"/>	8808133002	2013	00050-04-013-000C	Toutes	Toutes	4,545	14,827	10,282	15-Cor.BB(20)+CU(60)	Conservation					Bitume PG 64-34	✓	88
<input checked="" type="checkbox"/>	8808133002														Bitume PG 64-34	✓	88
<input checked="" type="checkbox"/>	8406133004													30/09/2013	Bitume PG 70-28	✓	84
<input checked="" type="checkbox"/>	8406133004													30/09/2013	Bitume PG 70-28	✓	84
<input checked="" type="checkbox"/>	8406133002													30/09/2013	Bitume PG 70-28	✓	84
<input checked="" type="checkbox"/>	8406133002	2013	00040-02-033-000G	2 (chain. décroiss.)	Toutes	2,140	3,411	2,271	19-Plan(50)+CU(50)	Conservation	771 000 \$			30/09/2013	Bitume PG 70-28	✓	84

Sélection de travaux pour le groupe 2

Sélectionné	No. d'enregistrement	Année des travaux	RTSS	Direction	Voie	Ch. début	Ch. fin	Longueur (km)	Intervention	Axe budgétaire	Coût total travaux (\$)	Superf. totale travaux (m2)	Coût intervention (\$)	Date ouverture trafic	Type de liant	RSS
<input type="checkbox"/>	85721431001	2014	00013-02-020-000G	2 (chain. décroiss.)	Toutes	0,855	1,303	0,448	72-Cor.BB+CU HP sur ch.BC	Conservation	274 413 \$	6 779	147 097 \$	23/06/2014	Bitume PG 70-28	✓
<input type="checkbox"/>	8606133203														Bitume PG 70-28	✓
<input type="checkbox"/>	8606133204														Bitume PG 70-28	✓
<input type="checkbox"/>	8606133205														Bitume PG 70-28	✓
<input type="checkbox"/>	85721432001												3 907 \$	23/06/2014	Bitume PG 70-28	✓
<input type="checkbox"/>	85721432001	2014	00013-02-010-000G	2 (chain. décroiss.)	Toutes	2,905	3,115	0,210	74-Plan +CU_HP sur ch.Mixte	Conservation	396 915 \$	14 498	326 907 \$	23/06/2014	Bitume PG 70-28	✓
<input type="checkbox"/>	85721432001	2014	00013-02-020-000D	1 (chain. croissant)	Toutes	0,213	0,608	0,395	74-Plan +CU_HP sur ch.Mixte	Conservation	396 915 \$	14 498	326 907 \$	23/06/2014	Bitume PG 70-28	✓

Contexte du projet

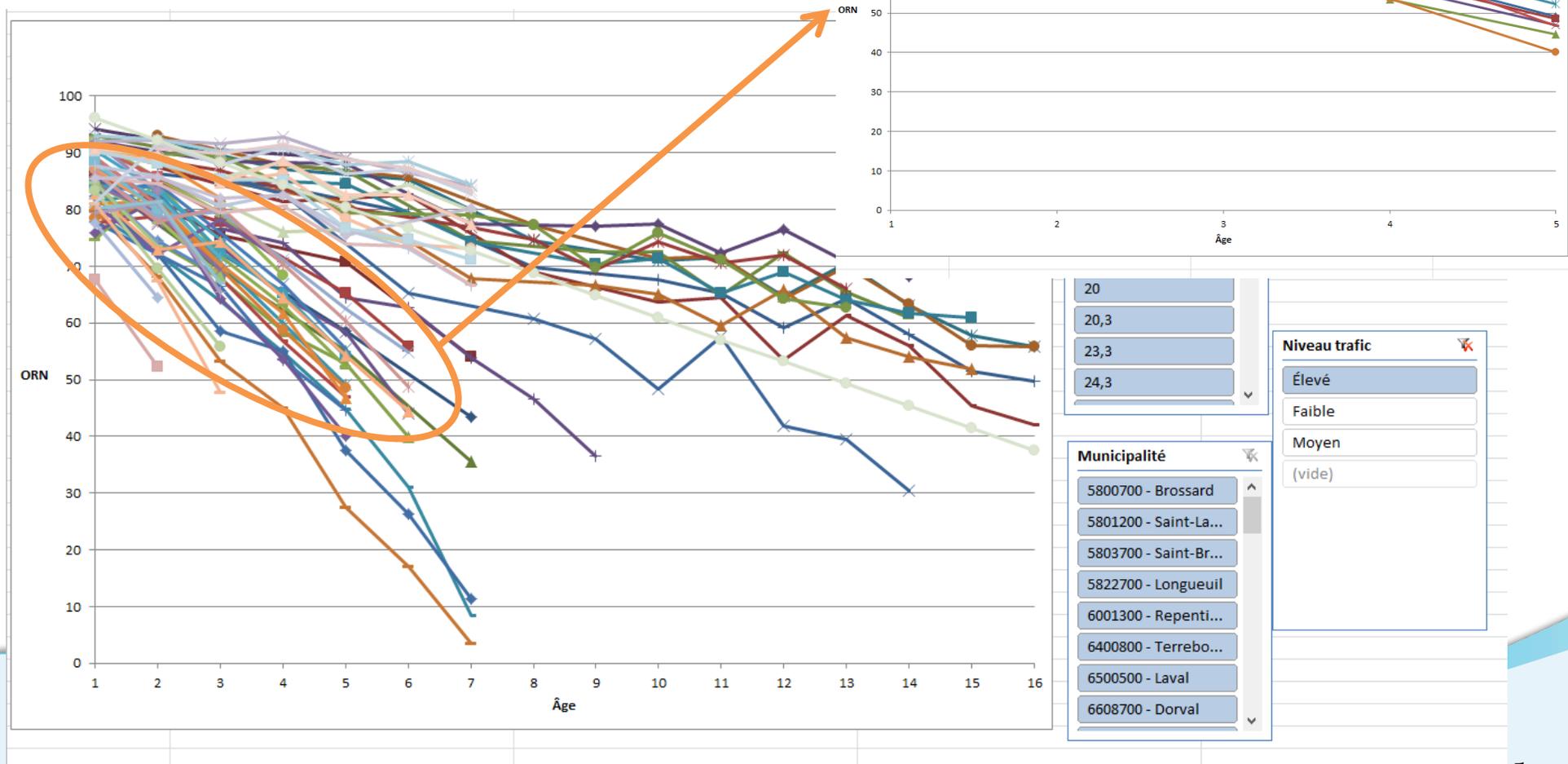
Analyses pour expliquer la performance



- Performance de plusieurs travaux
- À partir d'un nuage de lignes, recherche manuelle de groupes de chaussées au comportement semblable (selon un IP), avec des filtres
 - Long
 - Inefficace

Contexte du projet

EXEMPLE



Contexte du projet

Objectifs

Développer des méthodes pour:

- **Automatiser l'identification des groupes** de chaussées au comportement semblable selon IP
- **Rechercher des variables** expliquant le comportement de ces groupes
- Valider la **robustesse** des groupes



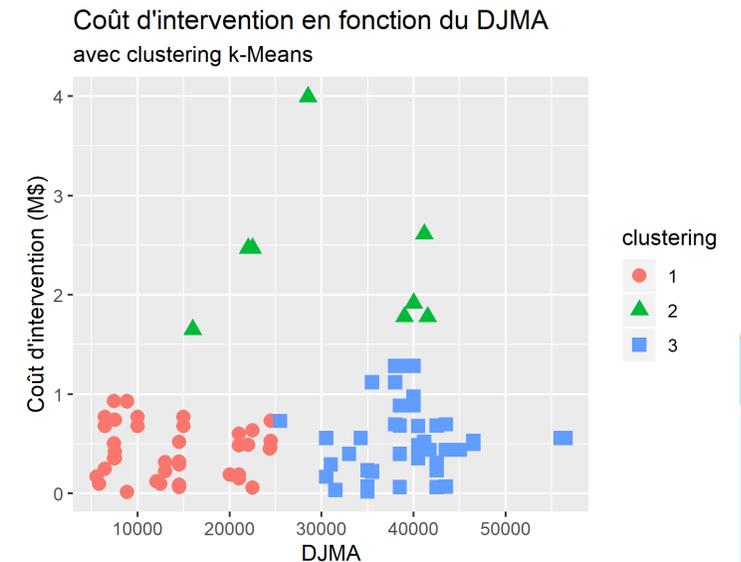
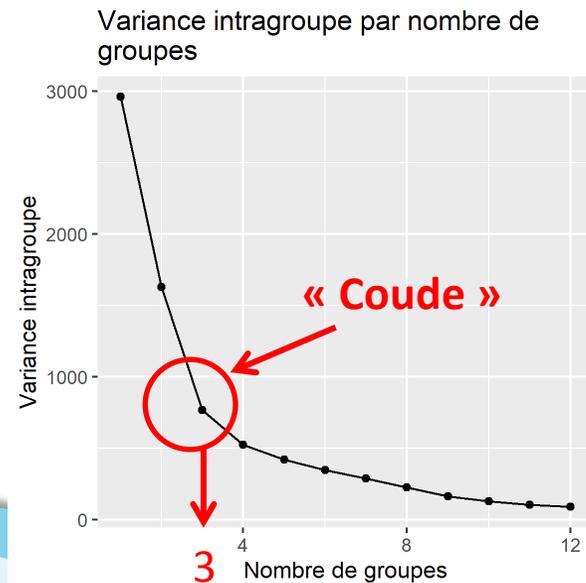
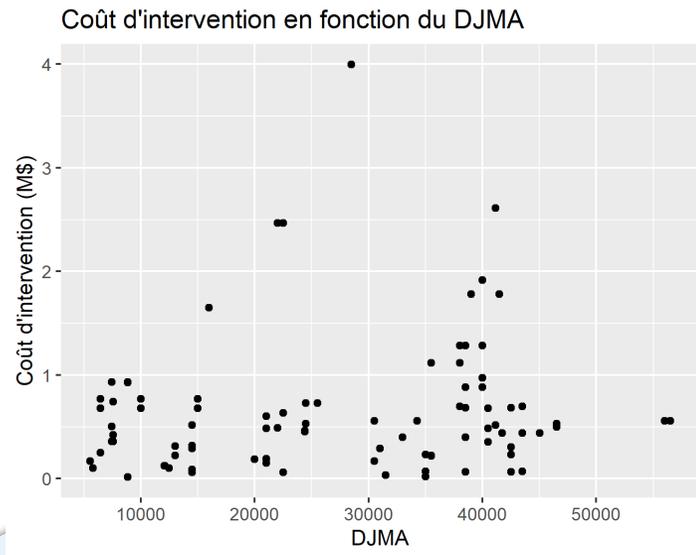
CLUSTERING

ARBRES DE
CLASSIFICATION

FORÊTS
ALÉATOIRES

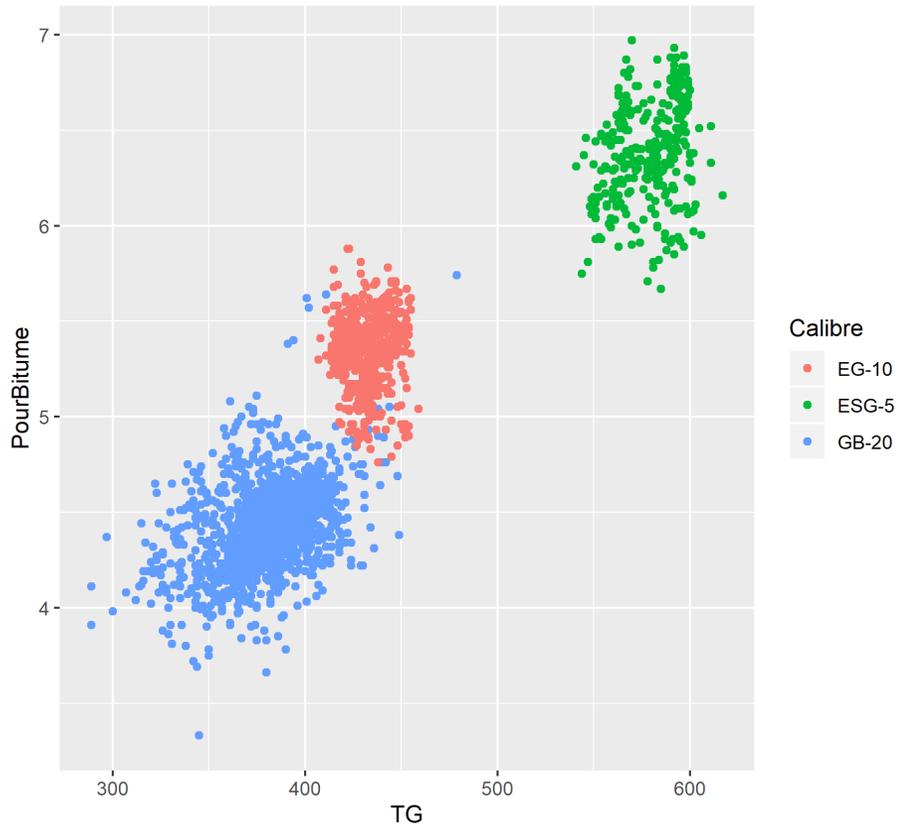
Clustering

- Algorithme qui forme des groupes de données
- À partir d'une ou plusieurs variables (dimensions)
- Détermination du nombre de groupes
- Basé sur un critère à optimiser ou à stabiliser

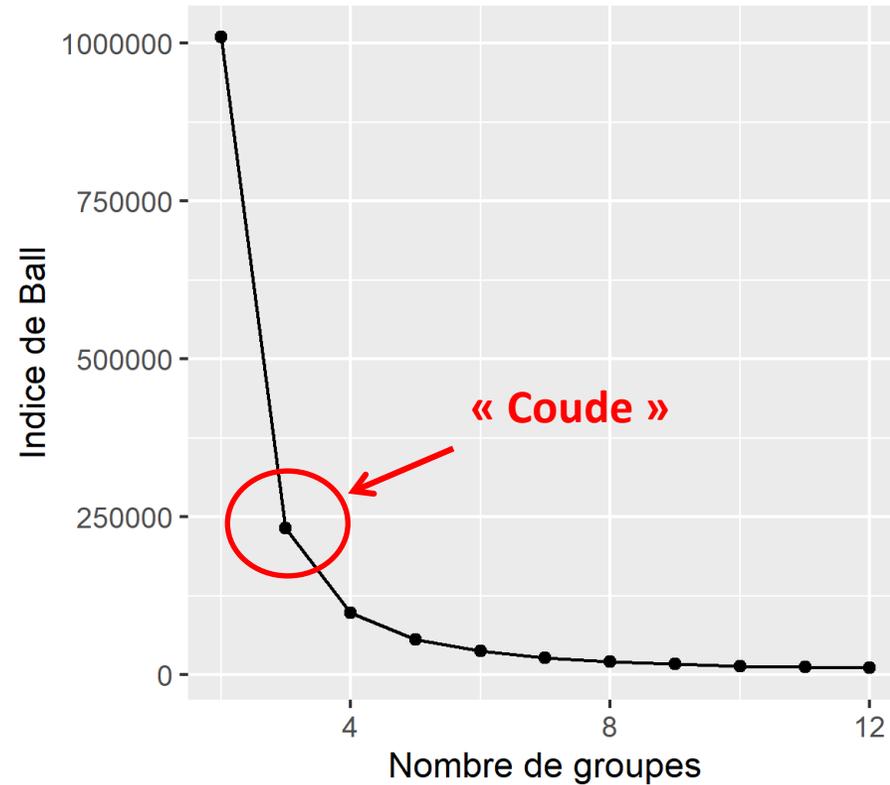


Clustering (exemples)

Pourcentage de bitume selon le total granulométrique
Données originales

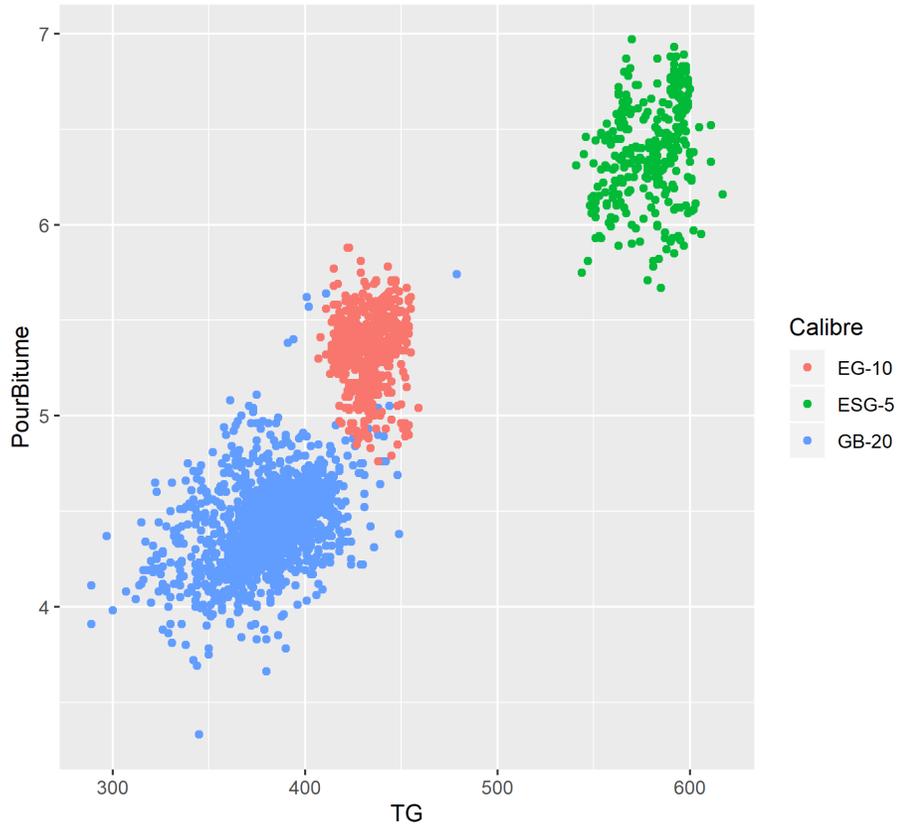


Indice de Ball par nombre de groupes, clustering K-Means

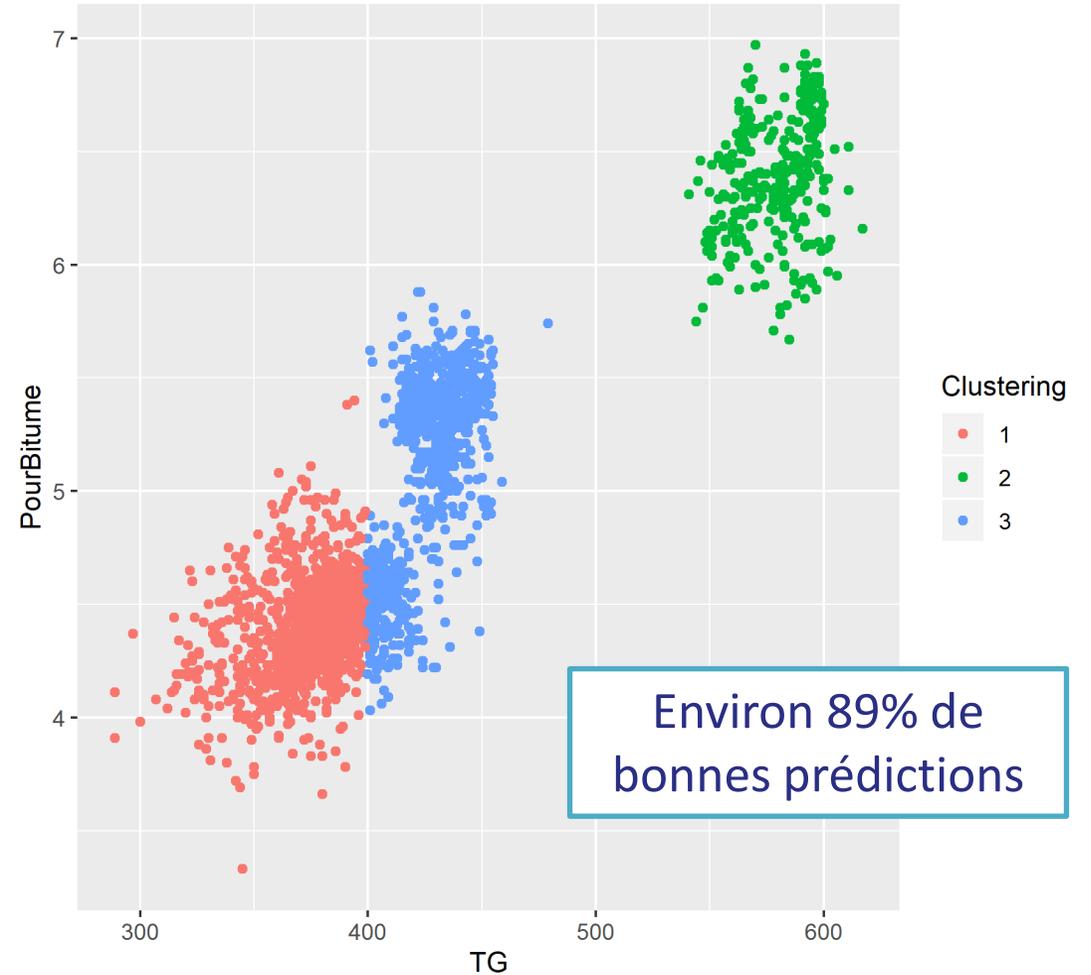


Clustering (exemples)

Pourcentage de bitume selon le total granulométrique
Données originales



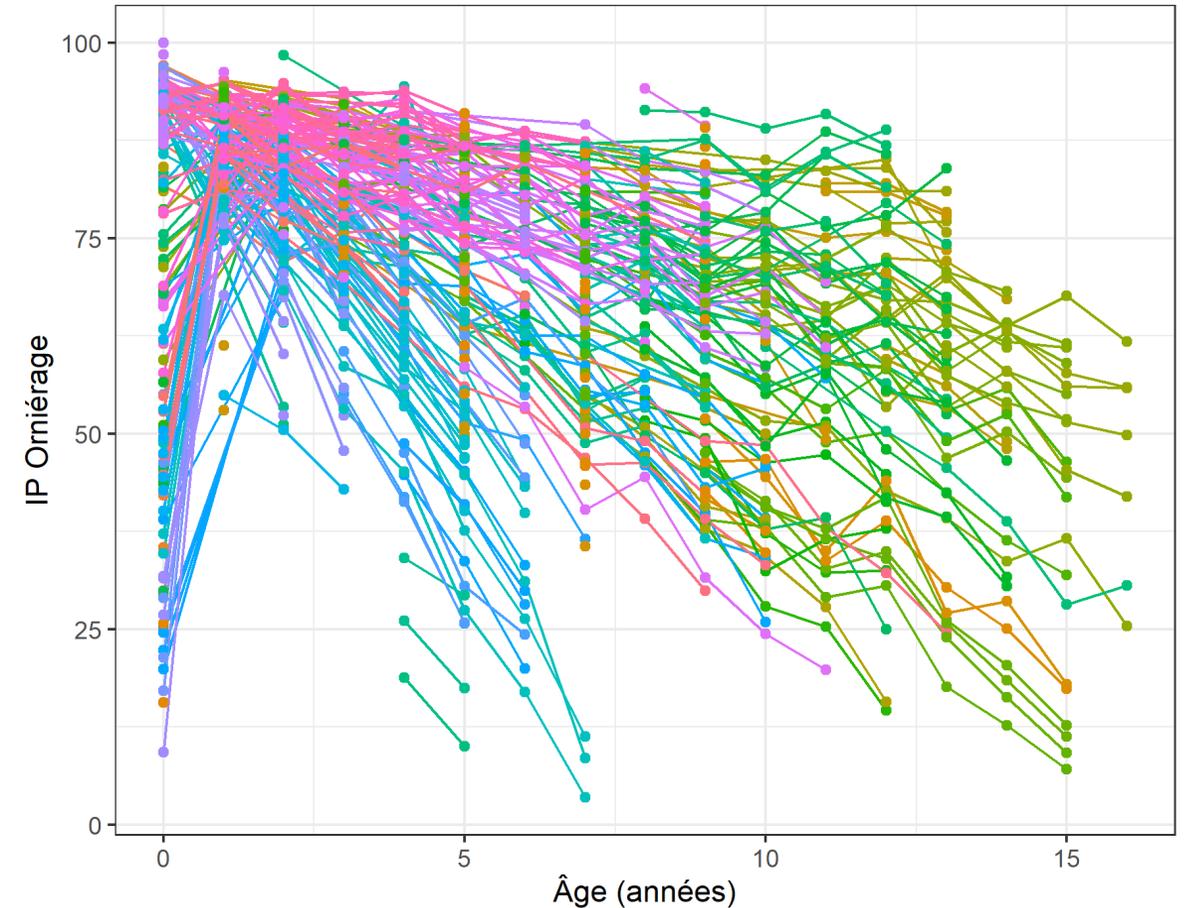
Pourcentage de bitume selon le total granulométrique
Clustering k-Means, sachant 3 groupes



Défis liés au *clustering*

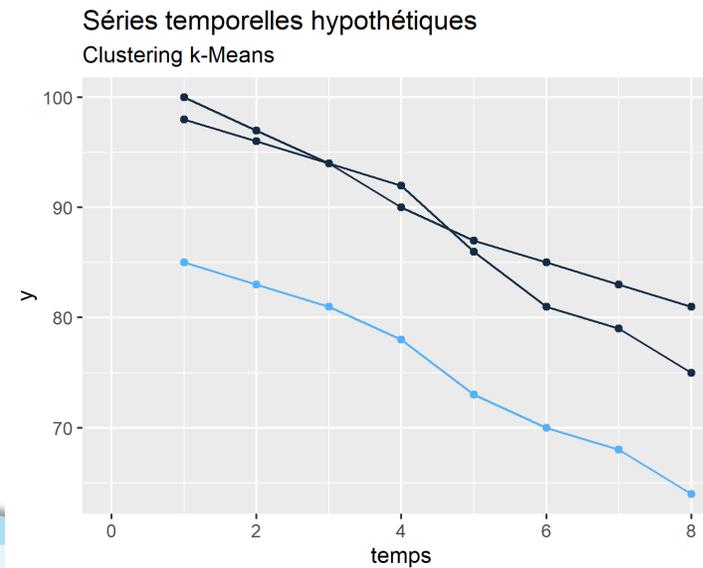
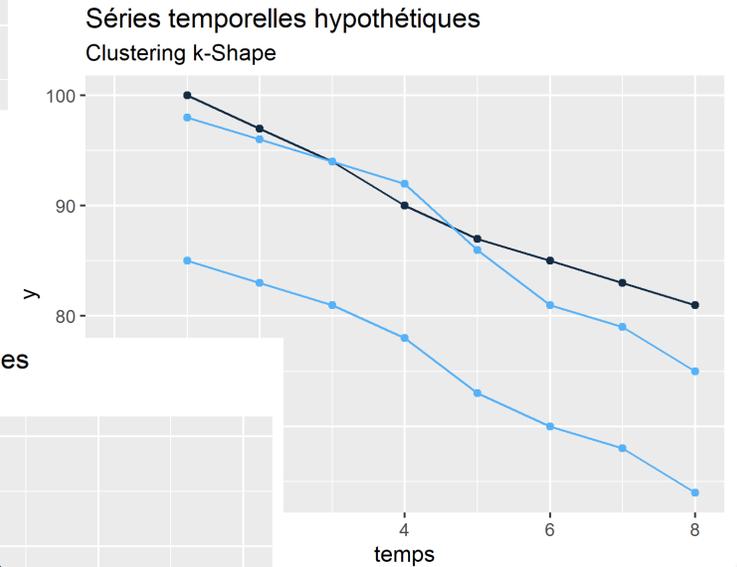
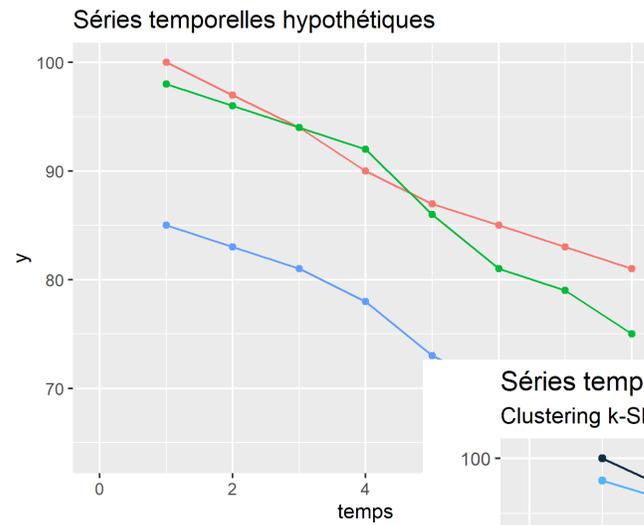
- Données temporelles
- Segments pas tous mesurés aux mêmes âges, à chaque année
- Choix du type de *clustering*

IP Orniérage selon l'âge du revêtement par localisation des travaux

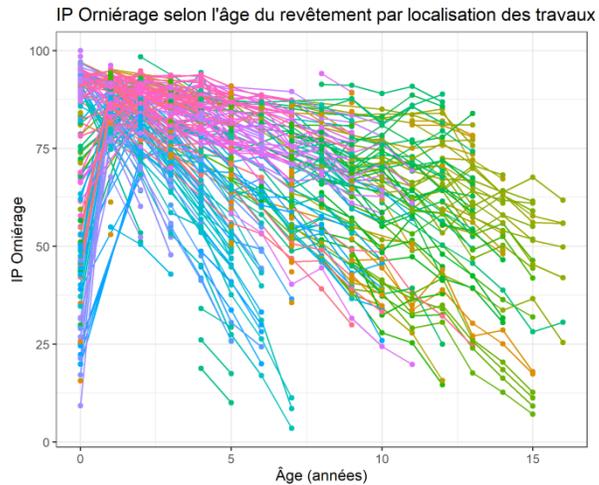


Défis liés au *clustering*

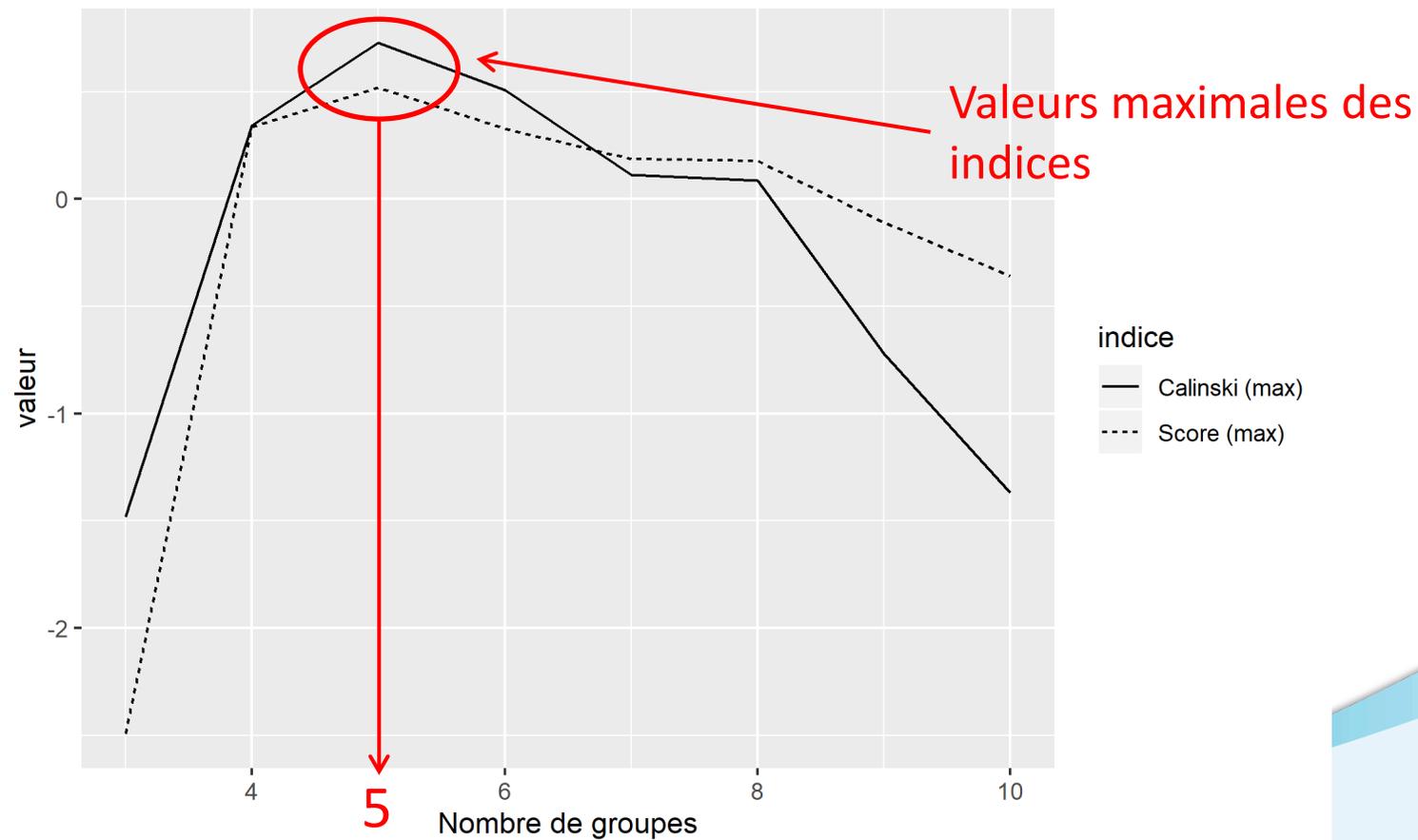
- Données temporelles
- Segments pas tous mesurés aux mêmes âges, à chaque année
- Choix du type de *clustering*
- Détermination du nombre de groupes : indices de qualité



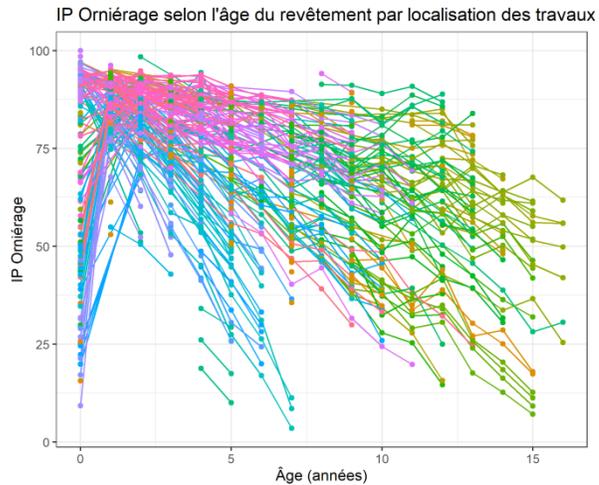
Application du *clustering*



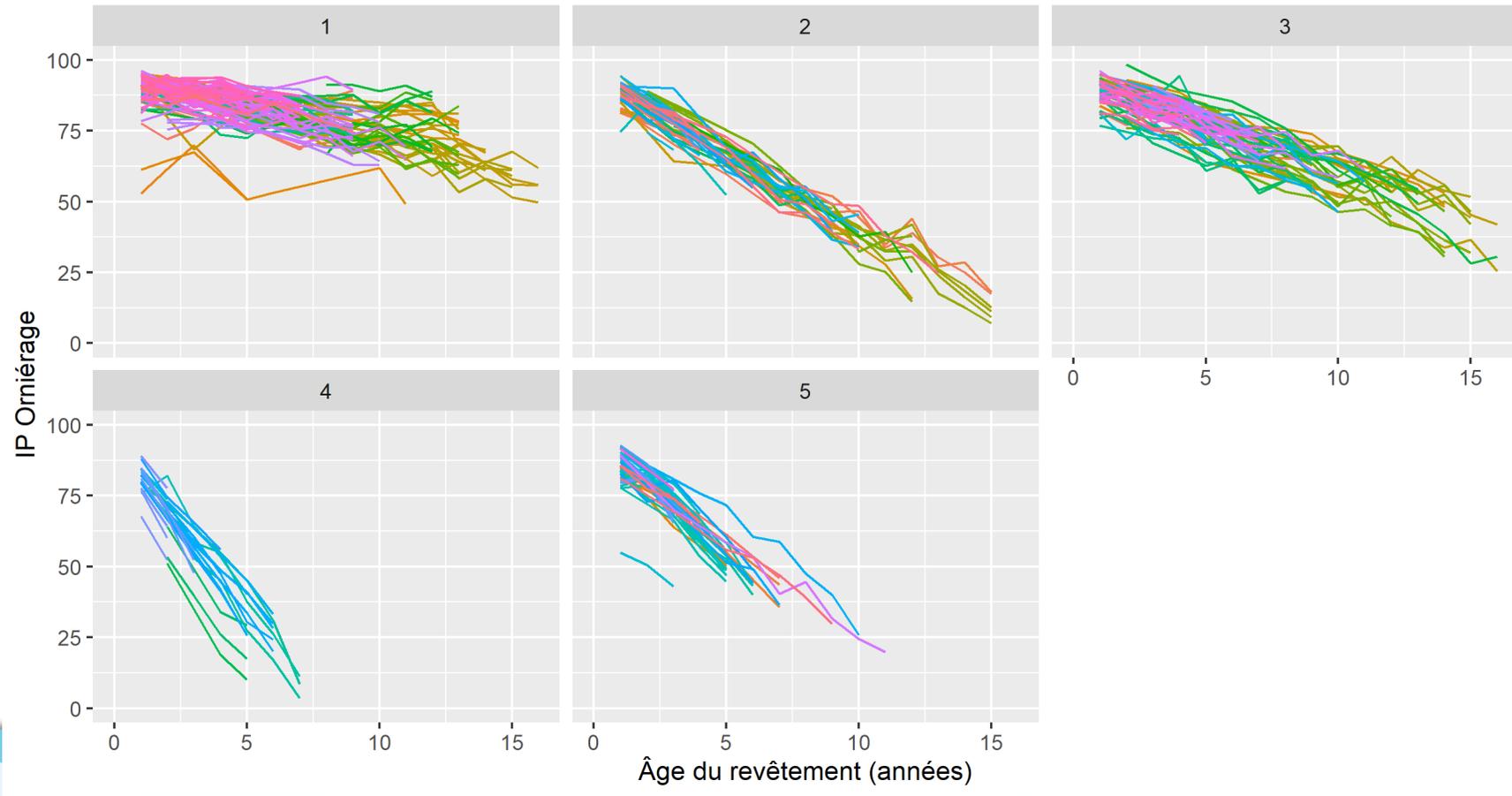
Indices de qualité du clustering k-Shape, IP Orniérage
Valeur médiane de 25 répétitions par nombre de groupes



Application du *clustering*

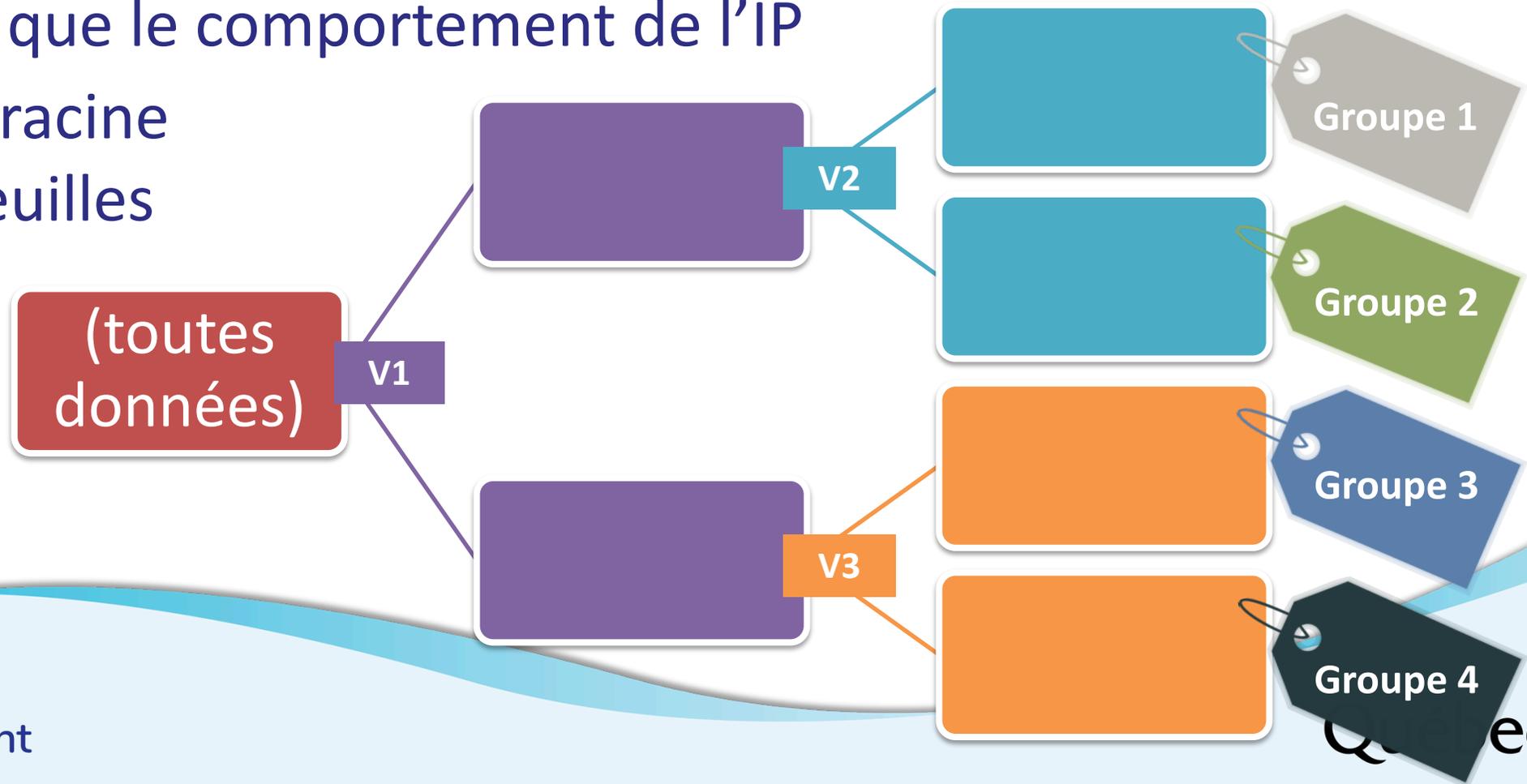


Groupes formés par k-Shape sur séries de longueurs inégales, extrapolées



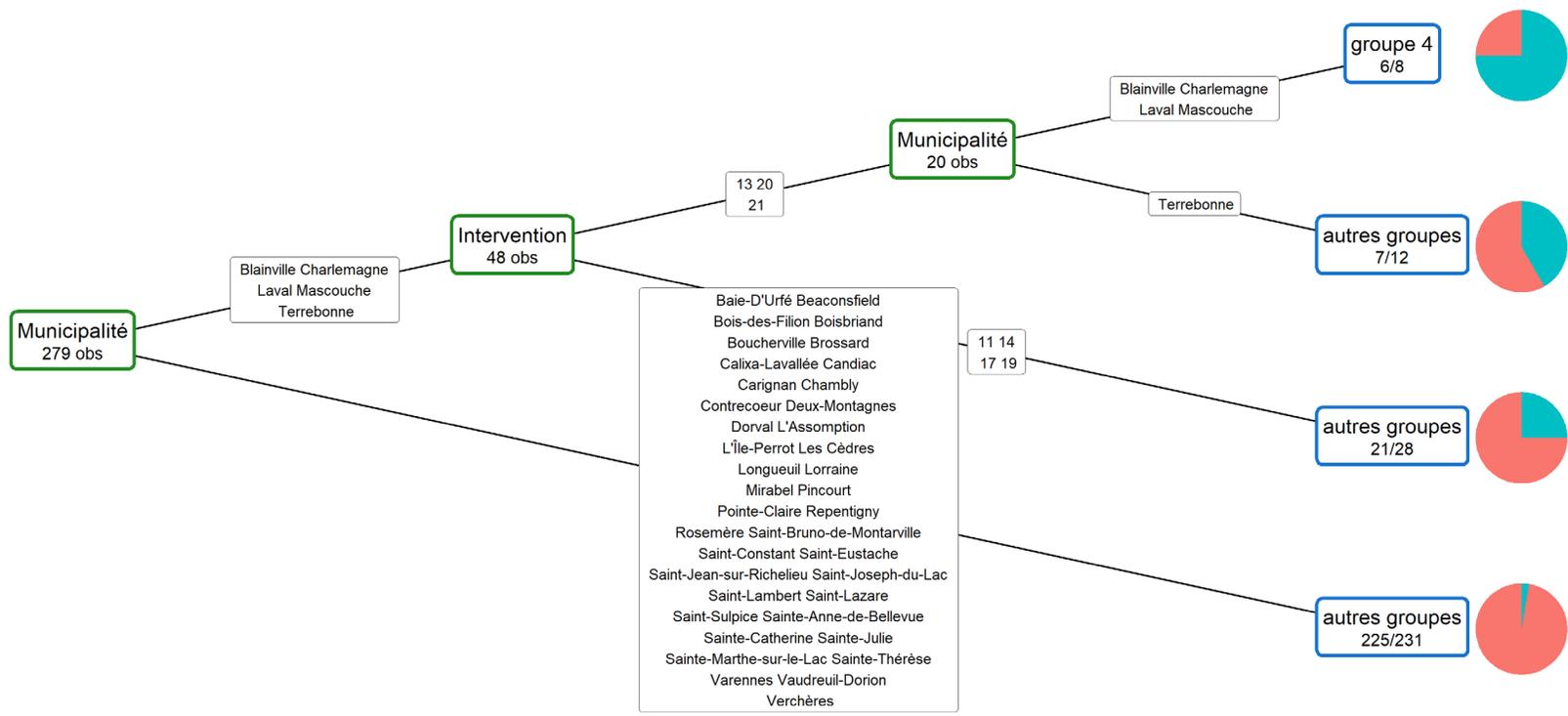
Arbres de classification

- Algorithme qui bâtit des règles pour expliquer les groupes, autre que le comportement de l'IP
- De la racine aux feuilles



Application d'un arbre de classification

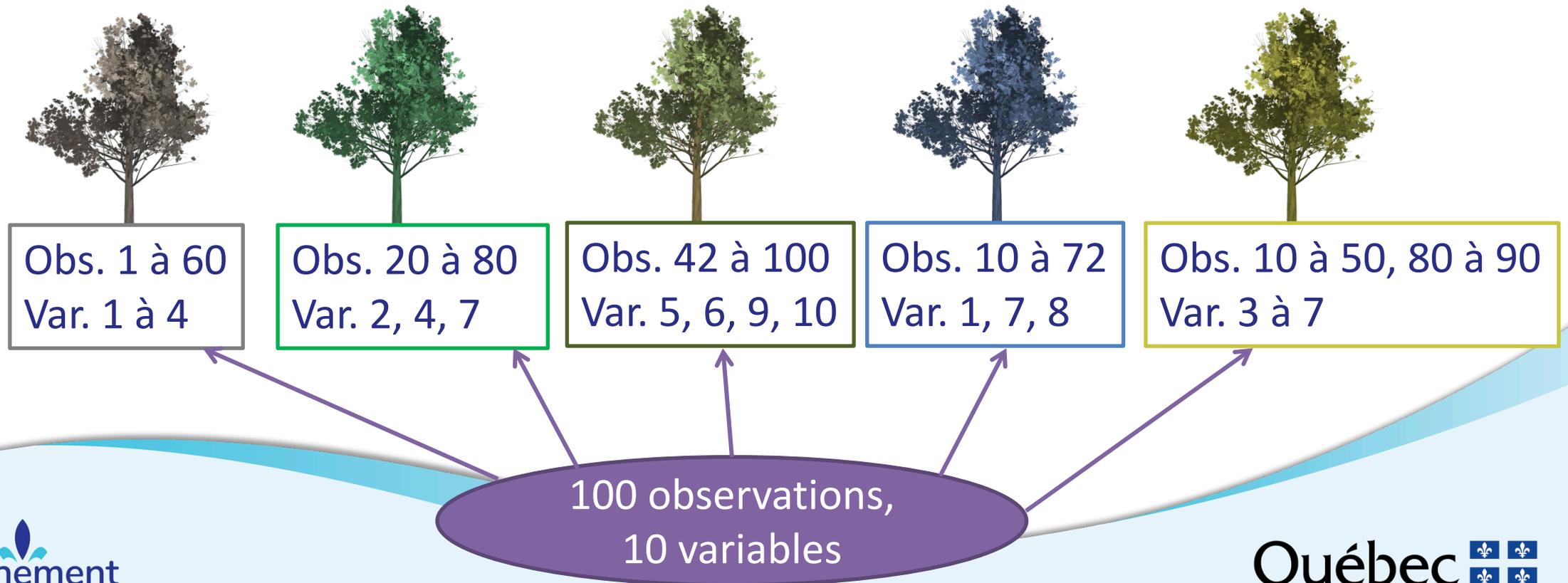
Arbre de classification pour le groupe 4



Groupe détecté ■ autres groupes ■ groupe 4

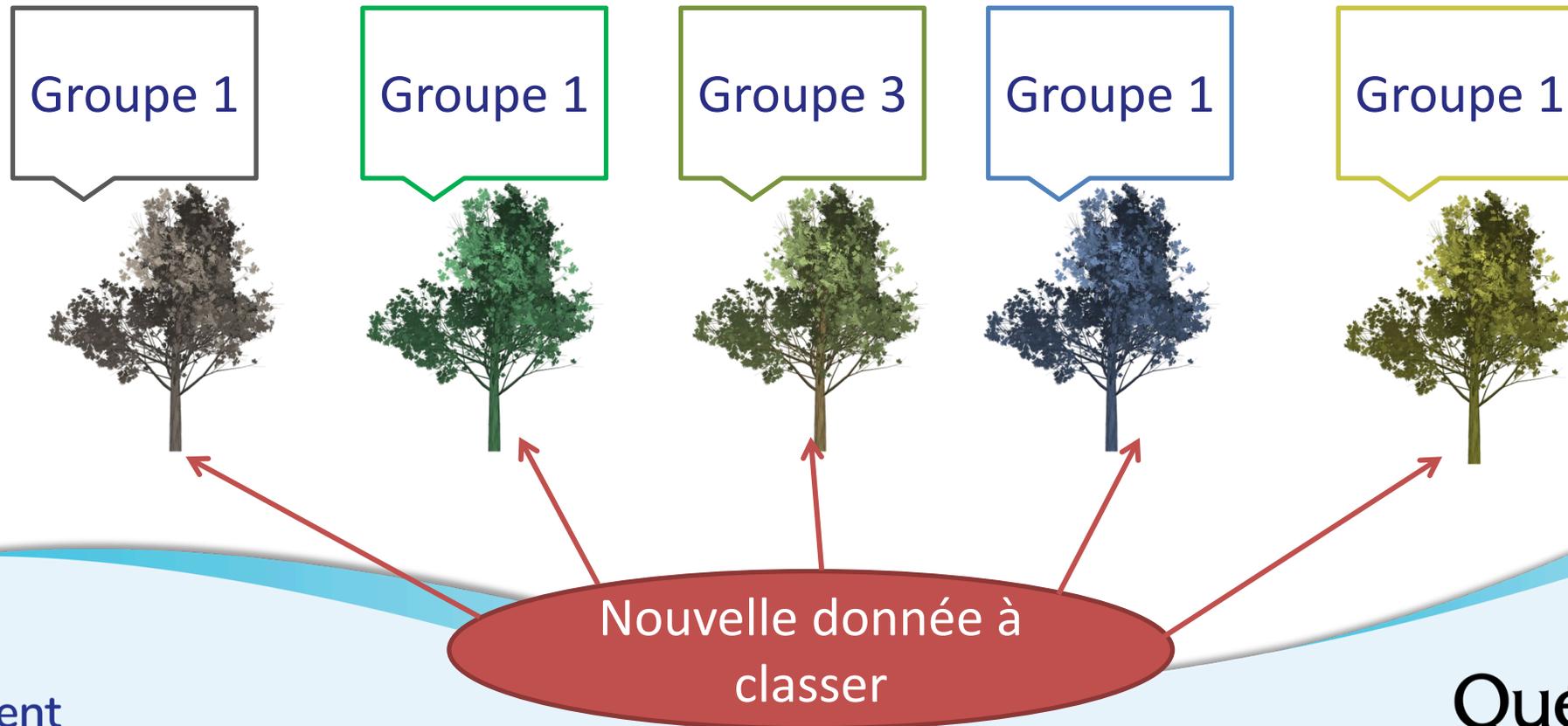
Forêts aléatoires

- Forêts : ensemble d'arbres **formés** à partir de sous-ensembles de données et de variables



Forêts aléatoires

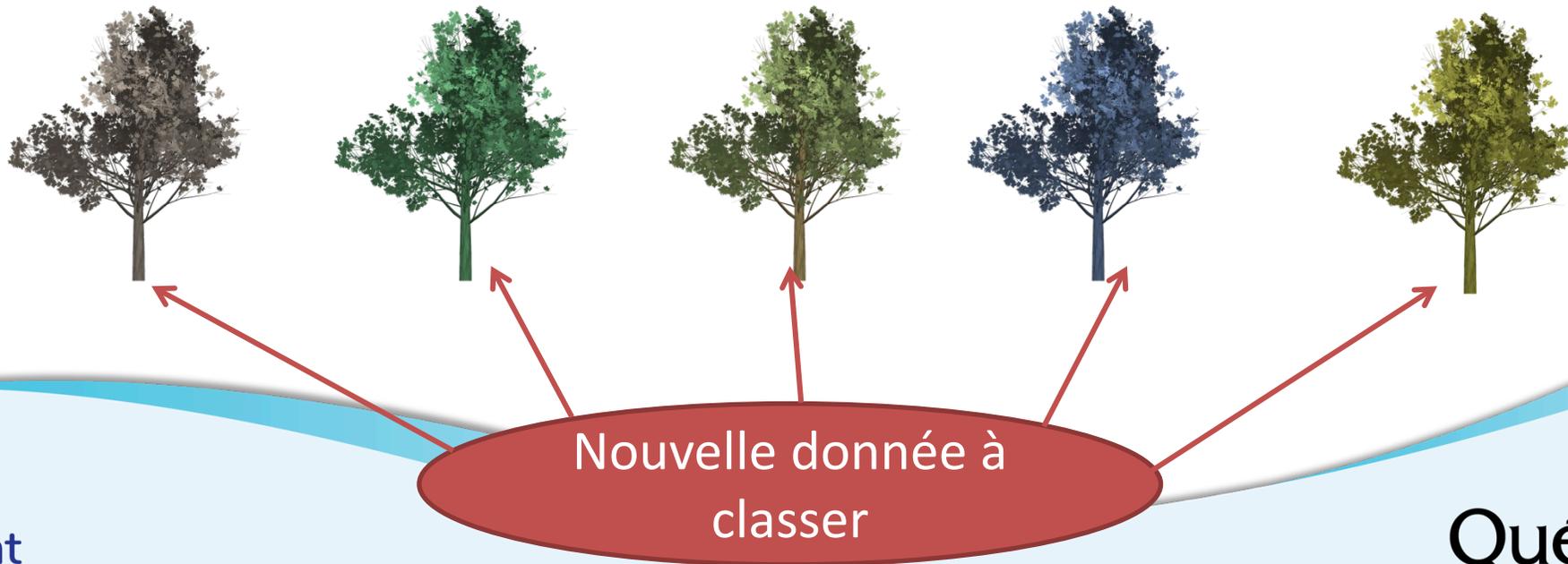
- Prédiction à partir d'une forêt : chaque arbre « vote »
- La majorité l'emporte



Forêts aléatoires

- Prédiction à partir d'une forêt : chaque arbre « vote »
- La majorité l'emporte

La nouvelle donnée est prédite « groupe 1 »



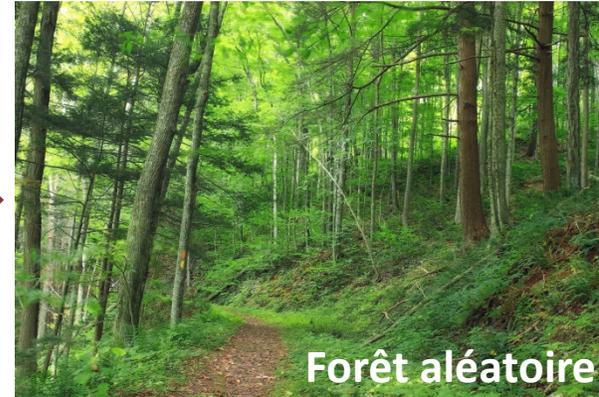
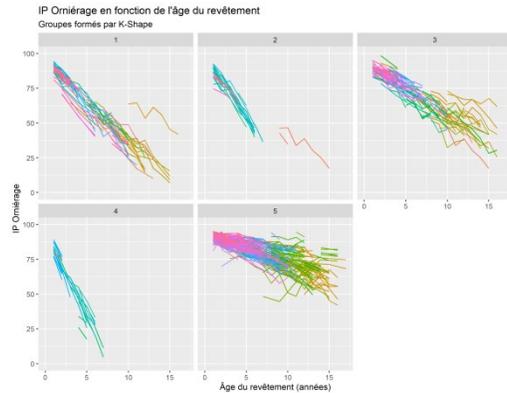
Forêts aléatoires

- + Meilleure prédiction qu'un seul arbre
- « Boîte noire », règles cachées

Pour le projet :

- Permet de voir quels groupes peuvent être bien prédits par les variables explicatives présentes

Application des forêts aléatoires



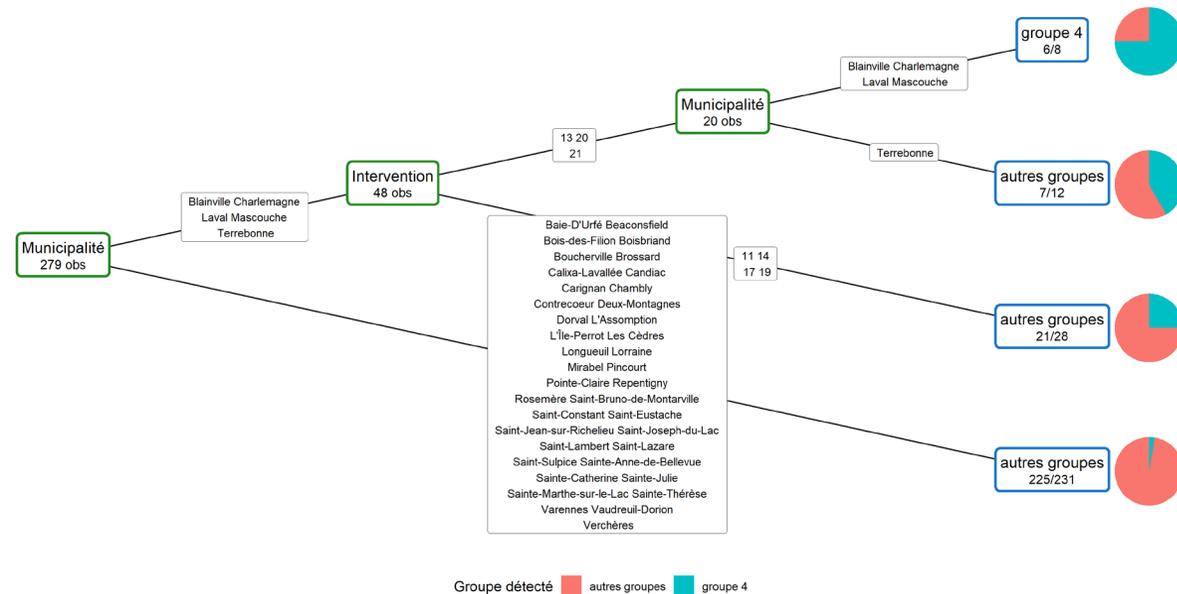
Résultats de la forêt		Groupe prédit					Bonnes prédictions
		1	2	3	4	5	
Vrai groupe	1	77	17	10	2	19	62%
	2	16	31	2	5	1	56%
	3	12	3	70	3	53	50%
	4	1	2	0	40	0	93%
	5	2	0	30	1	218	87%

Les groupes 4 et 5 se distinguent des autres, par leur comportement de l'orniérage **et** par leurs autres caractéristiques

Conclusion de l'exemple

- Les données des groupes 4 et 5 se distinguent des autres données par
 - Municipalité
 - Intervention
- Les autres groupes ne sont pas explicables par d'autres facteurs que le comportement

Arbre de classification pour le groupe 4



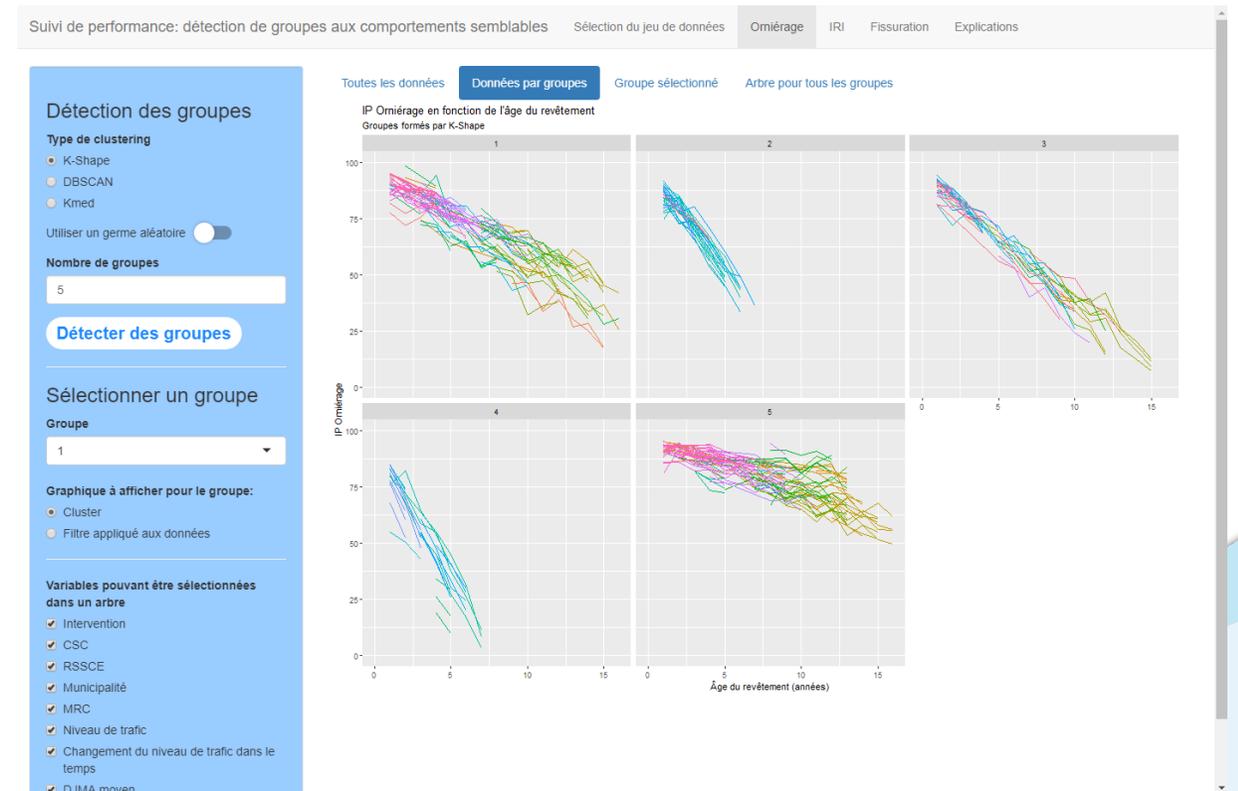
Conditions de succès

- Qualité des données
- Jeu de données bien ciblé selon les besoins
 - Bien choisir des variables en lien avec l'IP étudié
- Éviter les variables explicatives avec des valeurs manquantes



Ce qui reste à travailler

- Ajout de variables provenant d'autres systèmes
 - Variables corrélées
- Implémentation pour faire un outil accessible
 - Avec macros Excel?
 - Avec le système de suivi de performance?



MERCI!