

**Master de Géographie Parcours Géomatique, géodécisionnel,
géomarketing et multimédia (G2M)**

Projet tutoré

**Division sociale de l'espace résidentiel
dans la
ZONE MÉTROPOLITAINE DE MEXICO**



Enora Mahieux, Gérard In , Marion Lacroix, Yasmine Khalef

**Supervisé par
Jean-François Valette**

2022 - 2023

Table des matières

INTRODUCTION	3
1. PRESENTATION DES DONNÉES	4
1.1 Données spatiales.....	4
1.2 Données socio-démographiques.....	5
1.3 Données « Criminalité »	5
a. La base de données « Crime ».....	6
b. Les données gouvernementales.....	6
2. CADRES GÉOGRAPHIQUE ET CHRONOLOGIQUE DE L'ÉTUDE	6
3. TECHNOLOGIES EMPLOYÉES	8
4. EXPLOITATION DES DONNÉES : CHOIX DES VARIABLES	9
4.1 Données socio-démographiques.....	9
4.1.1 Définition de la précarité.....	9
4.1.2 Élaboration d'un indice de précarité	9
4.2 Données de criminalité.....	11
5. ANALYSES.....	13
5.1 ANALYSE DESCRIPTIVE DE LA CRIMINALITÉ	13
5.2 EXPLORATION DES DONNEES CRIMINALITE.....	16
5.2.1 Matrice de corrélation, corrélogramme.....	16
5.2.2 P-values et tests de significativité	17
5.2.3 Analyse en composantes principales (ACP).....	18
5.2.4 Classification ascendante hiérarchique (CAH).....	20
5.3 EXPLORATION DES DONNEES SOCIO-DEMOGRAPHIQUES.....	23
5.3.1 Étude de corrélation.....	23
5.3.2 Analyse en Composantes Principales (ACP)	24
5.3.3 Classification Ascendante Hiérarchique (CAH).....	26
5.4 RAPPORT ENTRE LA PRECARITE ET LA CRIMINALITE	30
CONCLUSION	34
Bibliographie.....	36
TABLE DES FIGURES	37

INTRODUCTION

Mégapole macro-céphalique par ses dimensions, métropole d'ordres national et mondial concentrant les fonctions de commandement et d'organisation économiques et logistiques du pays, également siège des institutions politiques, et pôle majeur en termes d'offre culturelle et universitaire du Mexique dont elle est la capitale, Mexico dispose d'un poids écrasant sur son territoire national.

Au sens administratif du terme, Ciudad de Mexico correspond en réalité uniquement au « District Fédéral », siège des pouvoirs fédéraux, et compte 9 millions d'habitants.

Mais son agglomération dépasse largement cette limite en s'étendant vers les Etats fédérés limitrophes de Mexico et d'Hidalgo, pour former la « Zone Urbaine de la Vallée de Mexico » : la ZMVM.

De fait, au sein d'un système urbain plutôt déséquilibré en sa faveur, l'aire métropolitaine de Mexico, avec ses 22 millions d'habitants, pèse 18% de la population du pays, et représente 25% du PIB mexicain [1].

Or en dépit de cette concentration de l'activité et de son dynamisme économique qui, outre ses orientations sociétales et politiques progressistes, la distinguent du reste du pays, la ZMVM n'en demeure pas moins un miroir du Mexique, le reflet d'une société inégalitaire, traversée par de fortes tensions sociales : en 2000, plus de 60% des actifs de l'agglomération gagnaient 200 euros par mois [2] ; et en 2010, 40% de sa population ne bénéficiait d'aucune couverture sociale, malgré les programmes sociaux étatiques [3]. Dans certaines municipalités périphériques, cette proportion s'élève à plus des deux tiers de la population [4].

La diversité sociale du Mexique et ses écarts importants de richesse s'y incarnent ainsi spatialement dans une forte différenciation entre ses beaux quartiers de l'ouest, et ses périphéries populaires de l'est et du nord.

Si l'agglomération reste épargnée par les règlements de compte entre groupes de narcotrafiquants qui endeuillent le pays depuis plusieurs décennies, l'insécurité y est élevée malgré l'importante présence policière. Son contexte demeure marqué par une forte délinquance : selon les chiffres du bureau du Procureur général de justice du District Fédéral, le taux de délits rapportés aux autorités a quasiment doublé entre 1990 et 1998, passant de 1412 à 2951 pour 100 000 habitants [2], et l'on y recense 65000 vols et 2000 homicides par an (dont 40% ne sont pas déclarés) [1].

Aussi, le diagnostic des enjeux de précarité et de criminalité, leur inscription dans l'espace de la ZMVM, l'analyse de la structure spatiale de ces phénomènes peuvent offrir diverses perspectives d'intérêt : à travers quelle implantation géographique se manifestent ces phénomènes ? Pour quelles formes et quelles concentrations de la précarité et de la criminalité ? Quelles sont les caractéristiques des populations résidentes concernées ?

Peut-on, en somme, établir et démontrer, pour le périmètre de l'aire métropolitaine de Mexico, un lien entre conditions de précarité et faits criminels ? C'est l'objet de la présente étude.

1. PRESENTATION DES DONNÉES

La conduite de notre projet s'est appuyée sur trois jeux de données :

1.1 Données spatiales

Il s'agit des shapefiles correspondant aux différents échelons du découpage administratif en vigueur au Mexique (*localidad, municipio...*), et aux unités statistiques (*ageb, manzana*) élaborées par l'*Instituto Nacional de Estadística y Geografía* (l'Institut national de statistiques et de géographie du Mexique), ou INEGI.

Créé en 1983, l'INEGI est un organisme public autonome chargé de réglementer et de coordonner le système national d'information statistique et géographique. Il réalise tous les dix ans des recensements de la population et publie mensuellement des statistiques économiques.

Pour rappel, notre aire d'étude est la ZMVM, unité statistique de l'INEGI. Indisponible sur le site de l'institut, nous avons construit cette couche à partir des échelons précités.

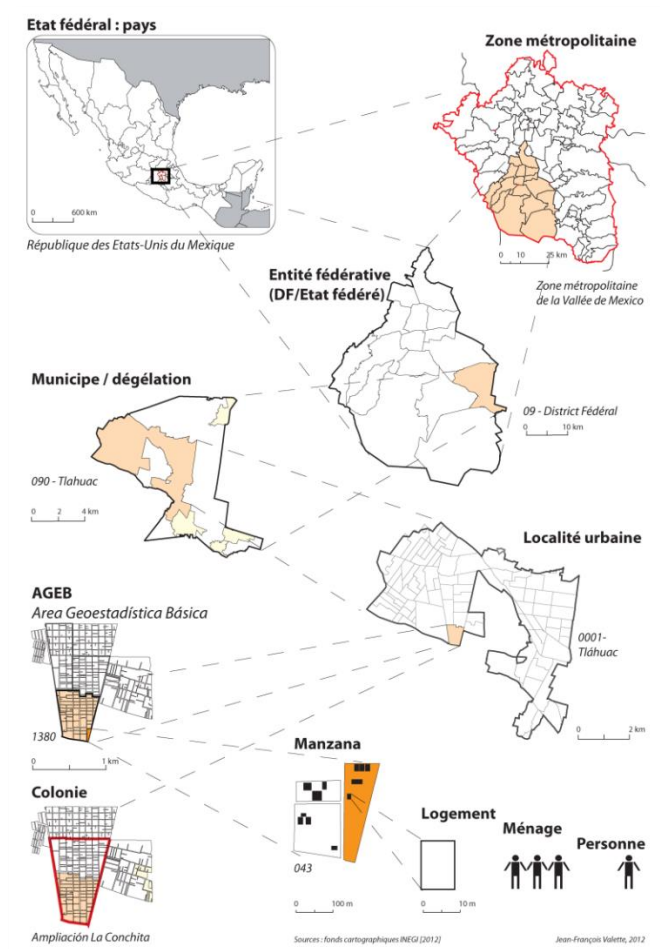


Figure 1 - Mailles géostatistiques INEGI (réal. : Jean-François Valette)

1.2 Données socio-démographiques

Elles se répartissent en 9 fichiers Excel, mis à disposition par ce même INEGI, et sont issues du recensement décennal de 2020. Elles sont regroupées au sein de deux bases :

- La base « Personas » et « Viviendas », dite « échantillon détaillé », contient des informations sur les personnes et les logements (correspondant aux ménages) ; une clé fait le lien entre les individus constitutifs d'un ménage.
- La base dite « agrégée » fournit les mêmes informations, mais transformées et agrégées aux différents échelons géographiques, et par tranche d'âge.

D'un point de vue structurel, chaque observation y est décrite par un volume conséquent de variables (près d'une centaine pour chacun des recensements) regroupées au sein d'ensembles thématiques :

- Caractéristiques du logement occupé (niveau d'équipement, accès aux TIC, composition du foyer et structure des revenus...),
- Des personnes (niveau d'étude, activité, travail, mobilités, couverture sociale, statut marital...),
- Alimentation,
- Finances,
- Religion,
- Etc.

1.3 Données « Criminalité »

Enfin, les données relatives aux faits criminels...

Notre réflexion sur ce matériau, dans le cadre du Mexique, s'est appuyée sur le site [elcri.men](#) [5], réalisé par M. Diego Valle-Jones.

Le data scientist mexicain y croise les données du *Secretariado Ejecutivo del Sistema Nacional de Seguridad Publica* – Secrétariat Exécutif du Système National de Sécurité Publique - (SESNSP), et les données « Homicide » de l'INEGI sur la totalité des municipes du Mexique, selon une typologie de crimes, et à travers une présentation statistique de leur évolution. L'intérêt de sa démarche réside dans le recul critique qu'il adopte vis-à-vis des données officielles.

Il a inspiré nos recherches bibliographiques et nous a orientés vers les données que nous évoquerons à la rubrique « Exploitation des données de criminalité : choix des variables ».

Mais auparavant, posons cette brève définition de la criminalité dans le Droit français : elle est l'ensemble des actes illégaux, délictueux et criminels, commis dans un milieu donné, à une époque donnée, et correspond à la transgression des normes juridiques d'un système social [6].

La notion de crime, et donc de criminalité, relèvent par conséquent d'une norme à la fois objective et subjective. Aussi, afin d'éviter les généralisations abusives relevant de notre propre système normatif

et culturel, et suivant les recommandations de M. Valette, nous avons choisi d'exploiter les deux jeux suivants :

a. La base de données « Crime »

Établie par M. Valette, elle recense, sur les trois dernières décennies, le nombre annuel d'homicides et de féminicides, ainsi que la population totale par genre et par municipalité. Elle prend sa source à l'INEGI et se présente au format Excel.

b. Les données gouvernementales

Elles proviennent du SESNSP, sous forme de fichiers Excel annuels, consolidés lors de leur mise à jour [7].

La fréquence des crimes y est renseignée mensuellement, pour chaque municipalité de l'agglomération, et leur description répond à une typologie détaillée. Ainsi, 40 types de crime sont recensés, et chacun d'eux y est catégorisé par :

- La nature du préjudice subi : atteinte aux biens, atteinte à l'intégrité corporelle, violences familiales, etc.
- Son type : homicide, séquestration, extorsion, braquage, vol, droit commun, etc.
- Un sous-type : volontaire/involontaire, lieu du crime, etc.
- Sa modalité : arme blanche/arme à feu, avec/sans violence, véhiculé ou non, etc.

Nos analyses se sont préalablement appuyées sur cette catégorisation.

2. CADRES GÉOGRAPHIQUE ET CHRONOLOGIQUE DE L'ÉTUDE

Elle porte sur la Zone métropolitaine de la vallée de Mexico, pour l'année 2020 (dernier recensement en date établi par l'INEGI), à l'échelle des Municipales (équivalent des communes françaises en termes de superficie).

Ce choix repose sur la disponibilité des données, cet échelon étant le plus fin et le plus documenté dans les données criminelles à notre disposition.

Composée de trois entités fédératives – Mexico, le District Fédéral et Hidalgo -, la ZMVM compte 76 municipes.

Précisons enfin que les données « Crime » utilisées sont celles de 2019, soit un an avant l'épidémie de Covid, afin de disposer d'une photographie qui ne soit pas altérée par le caractère exceptionnel de la situation sanitaire de 2020.

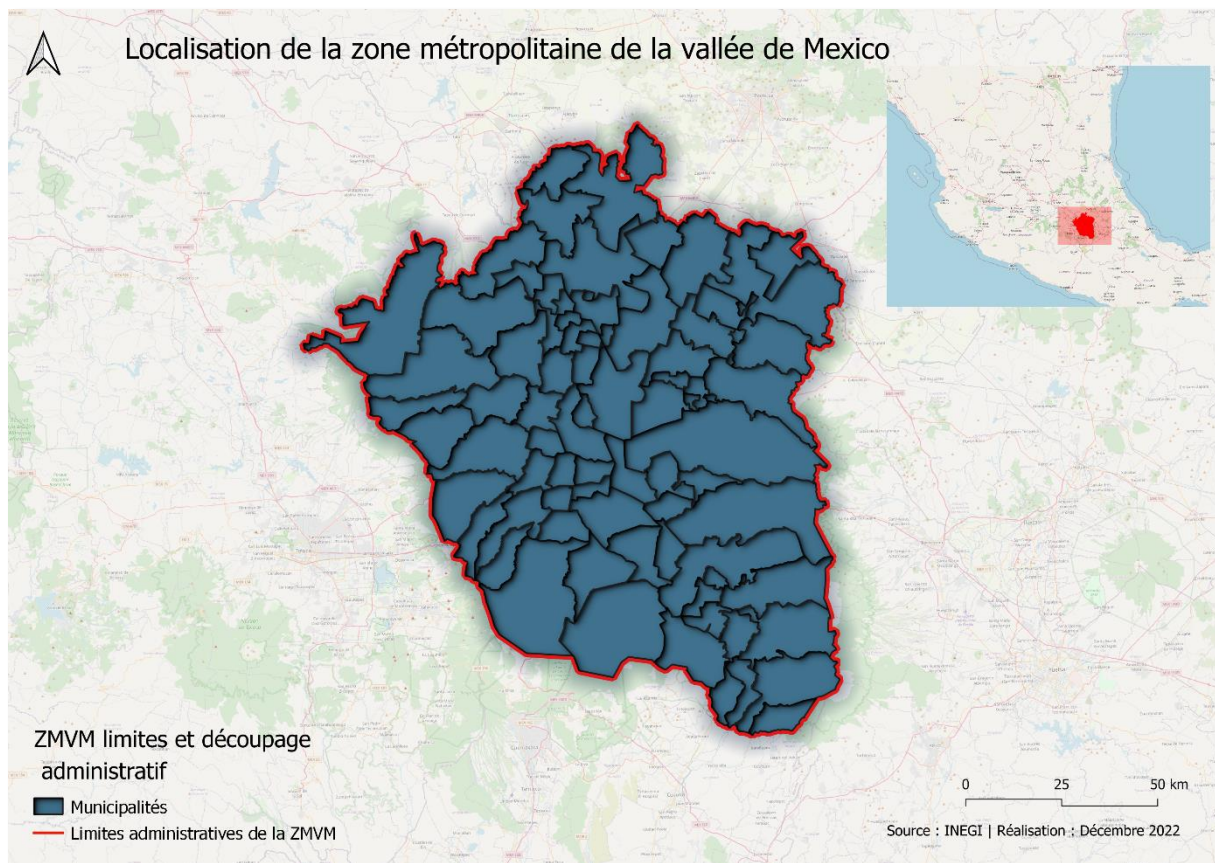


Figure 2 - Zone Métropolitaine de la Vallée de Mexico

3. TECHNOLOGIES EMPLOYÉES

Notre travail couvrant à la fois les aspects de documentation et de développement, de récupération et de traitement de données, de production et de représentation de l'information, dans le cadre d'une gestion de projet, nous nous sommes appuyés sur les technologies suivantes :

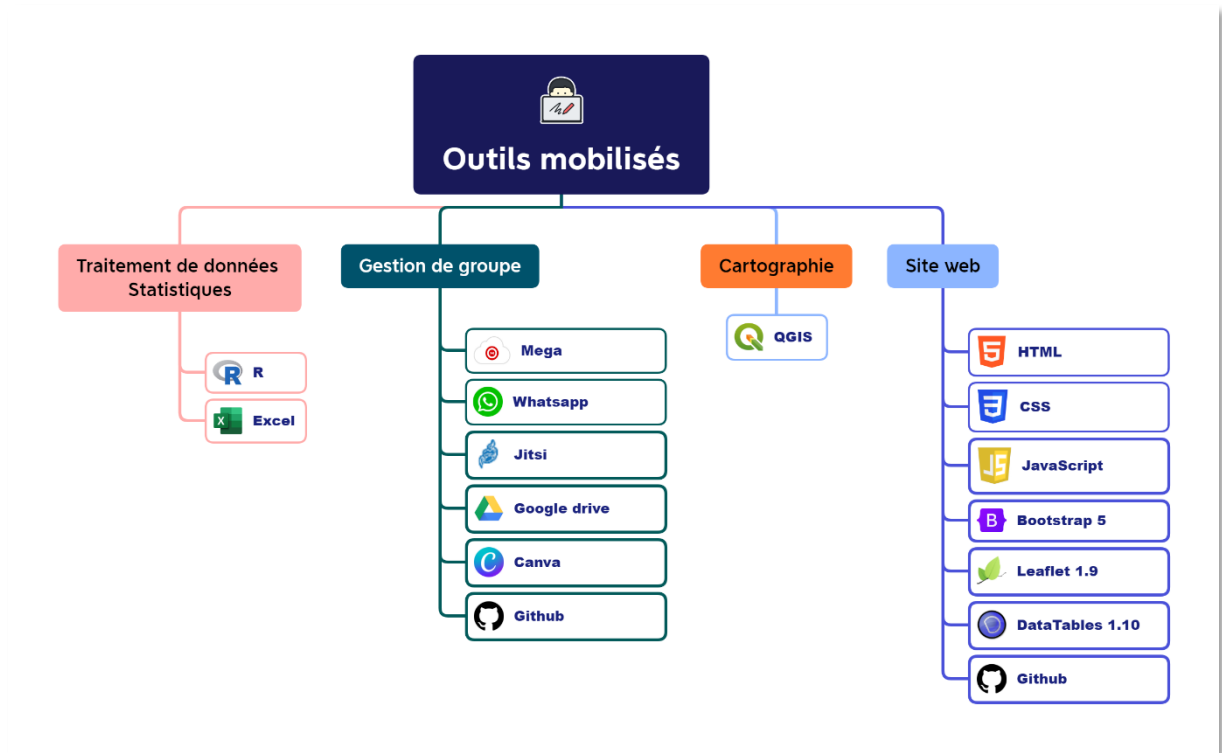


Figure 3 - Diagramme de présentation des technologies mobilisées dans notre projet

4. EXPLOITATION DES DONNÉES : CHOIX DES VARIABLES

4.1 Données socio-démographiques

Nous avons choisi dans l'exploration des données socio-démographiques de nous focaliser sur la géographie de la précarité. Ce choix prend sa source dans nos recherches sur la criminalité, qui indiquent un lien entre précarité et criminalité.

Ainsi, préalablement au choix de nos variables dans les bases étendues et agrégées, on s'est attaché à définir notre thématique de précarité.

4.1.1 Définition de la précarité

« État, caractère de ce qui est précaire. » [8] [9]

Au regard de cette définition peu éclairante, il nous a paru opportun d'approcher le terme par trois synonymes (proposés par le Larousse et le Petit Robert), plus expressifs :

- L'**instabilité**, qui s'oppose à la stabilité d'une situation. On entend ici que la pauvreté ne peut être considérée comme un critère de précarité à elle seule. Suivant cette idée, un individu disposant d'un revenu faible mais stable, bénéficie d'une situation moins précaire qu'un individu disposant d'un salaire plus élevé, mais soumis à des aléas de santé, financiers, etc. ;
- La **fragilité** désigne l'incapacité ou la faible capacité d'un individu à résister à cet aléa. Ici, cette fragilité peut être d'ordre économique comme sanitaire ;
- L'**incertitude**, enfin, est liée à cette notion de fragilité. Outre la difficulté de répondre à une situation difficile, elle convoque l'incapacité d'anticiper les événements et donc de savoir comment y répondre lorsque ceux-ci surviennent.

Dans son article « Construire un indicateur de précarité : les étapes d'une démarche empirique », où la notion ne comptait pour synonymes que l'instabilité et la fragilité, André Villeneuve notait en 1984 : « Les deux sens sont bien distincts, mais non opposés : l'instabilité implique des hauts et des bas, et les bas sont a priori des occasions de fragilité » [10].

De fait, la précarité ne repose pas uniquement sur la pauvreté. Celle-ci en est une composante, mais ne l'explique pas entièrement. En conséquence, elle ne constituera pas le facteur déterminant de notre étude.

4.1.2 Élaboration d'un indice de précarité

C'est donc à partir de cette conception de la précarité que nous avons envisagé la construction d'un indice (de cette précarité), sur la base des données socio-démographiques de l'INEGI.

La littérature consultée fait état de biais et de malfaçons que nous avons essayé d'éviter dans notre démarche, inédite pour les membres de notre groupe.

Afin de ne pas s'enfermer dans une thématique de la précarité en particulier, notre travail a donc consisté, à partir des 350 variables de l'INEGI, à composer un bouquet cohérent de variables issues de diverses thématiques, avec pour objectif d'en étudier l'éventuelle porosité/contagiosité.

Aussi, les critères retenus pour sélectionner les variables les plus représentatives possibles ont reposé sur :

- Une fréquence significative (ce qui est le cas des variables de l'INEGI) ;
- Une corrélation forte avec d'autres variables issues de chacune des thématiques des données fournies par l'Institut.

Enfin, nous avons décidé de conserver uniquement les données que nous estimions objectives. Ce critère exclut de fait certaines variables qui, de prime abord, semblaient intéressantes. Par exemple : « les interrogés considèrent-ils manger à leur faim ou non ? ».

À partir de la documentation fournie par l'organisme, nous avons donc traduit puis recensé l'ensemble des thématiques des données INEGI.

Nos variables choisies parmi ces thématiques pour l'indice de précarité sont regroupées dans le schéma ci-dessous.

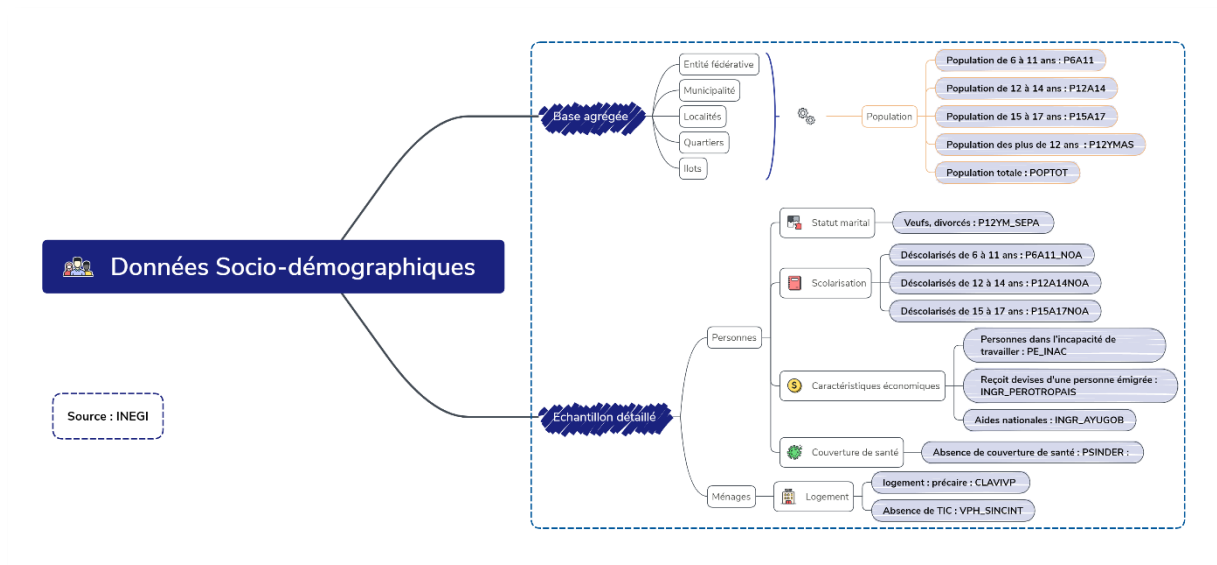


Figure 4 - Diagramme de choix des variables socio-démographiques

4.2 Données de criminalité

Nous nous sommes affranchis de la granularité très fine des données publiées par le gouvernement mexicain (préjudice > type > sous-type > modalité) en nous arrêtant aux types seuls pour nous concentrer sur une palette plus générique.

Ainsi, après une étude de leur distribution, nous avons choisi de conserver tout type dont la part est supérieure à 5% du total de la criminalité.

Outre ces variables, et après discussion avec M. Valette, nous avons retenu des données ayant une part moindre, mais socialement et culturellement intégrées à la notion de criminalité au Mexique.

Nous obtenons donc des variables hétérogènes dans leur sens et leur niveau de violence (contravention vs homicide, par exemple), et sur lesquelles nous n'avons pas opéré de distinction. Certains résultats étant particulièrement significatifs : le vol, par exemple, représente 43,5% du total de la criminalité.

Crimes recensés par le gouvernement Mexicain	Somme	Part
Extorsion	3553	0,5%
Féminicide	213	0,0%
Homicide	6115	0,9%
Blessures	76786	11,9%
Autres crimes de Common Law	74929	11,6%
Vol	281585	43,5%
Enlèvement	416	0,1%
Violence domestique	40643	6,3%

Figure 5 - Tableau de fréquence des variables de criminalité sélectionnées

Nos variables de travail sont :

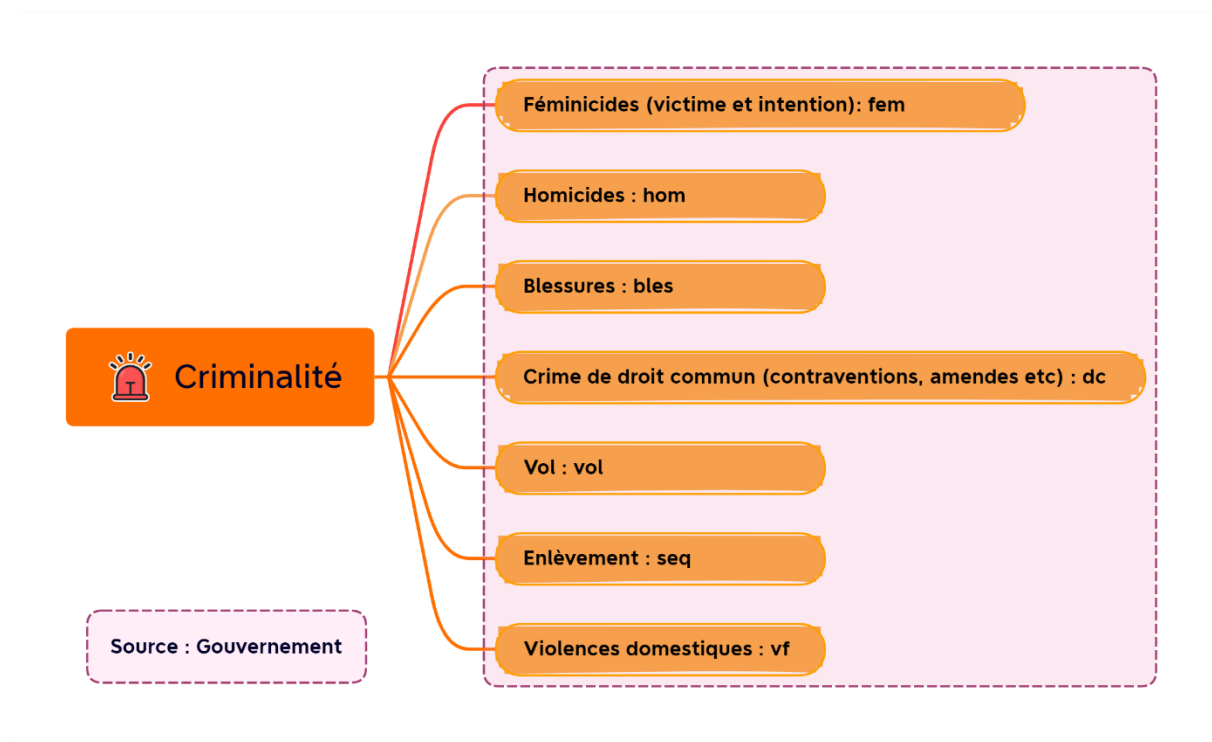


Figure 6 - Diagramme des variables de criminalité sélectionnées

5. ANALYSES

5.1 ANALYSE DESCRIPTIVE DE LA CRIMINALITÉ

Avant toute étude plus approfondie du territoire, nous décidons de nous pencher sur des analyses univariées relativement simples de la criminalité. L'objectif est de la présenter avant d'y revenir plus en détail par la suite.

Les cartographies qui suivent ont toutes été réalisées à la fois sur des données de stocks et de taux (pour 100 000 habitants). La discrétisation choisie est celle de Jenks, qui ne permet donc pas de comparaison entre les cartes mais reflète relativement bien les jeux de données.

Toutes les criminalités se répartissent de la même manière : les parts les plus importantes de tous les crimes se situent dans le cœur de la ZMVM, les franges de la zone étant d'apparence moins touchées.

Certains types de criminalité, tels que les crimes de droit commun, les homicides et féminicides ou les séquestrations s'étendent davantage que d'autres comme les violences familiales qui se concentrent dans le centre.

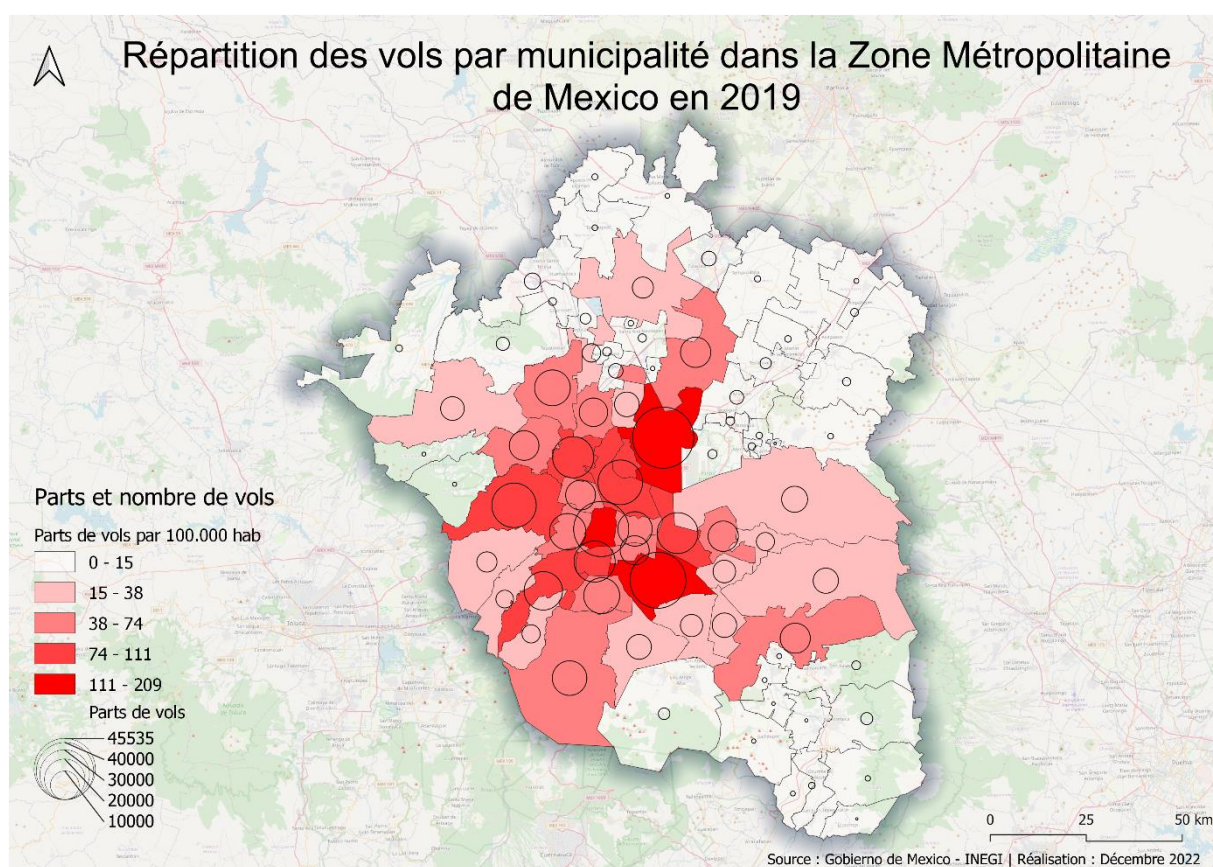


Figure 7 - Carte des vols

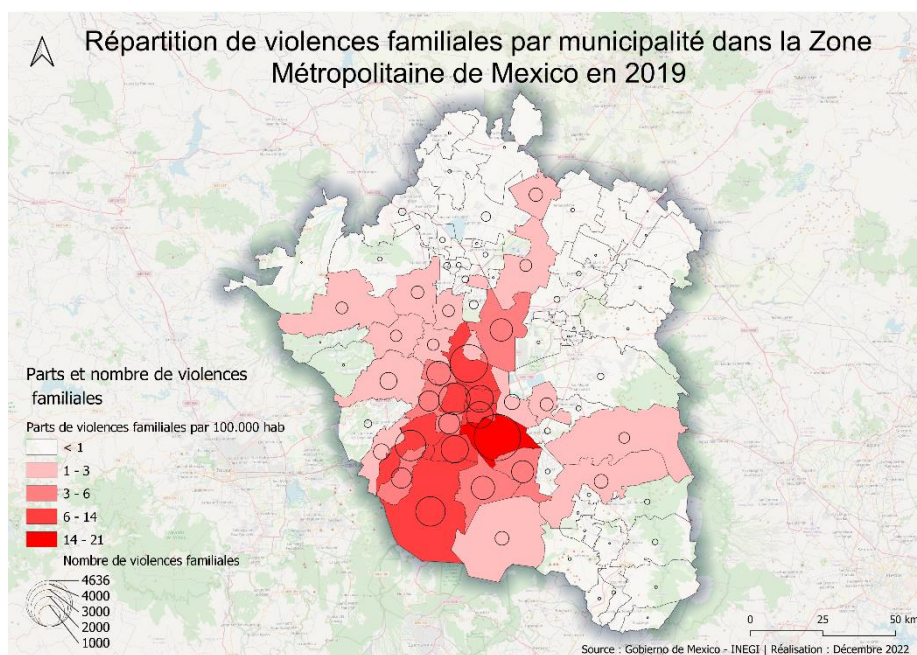


Figure 8 - Carte des violences familiales

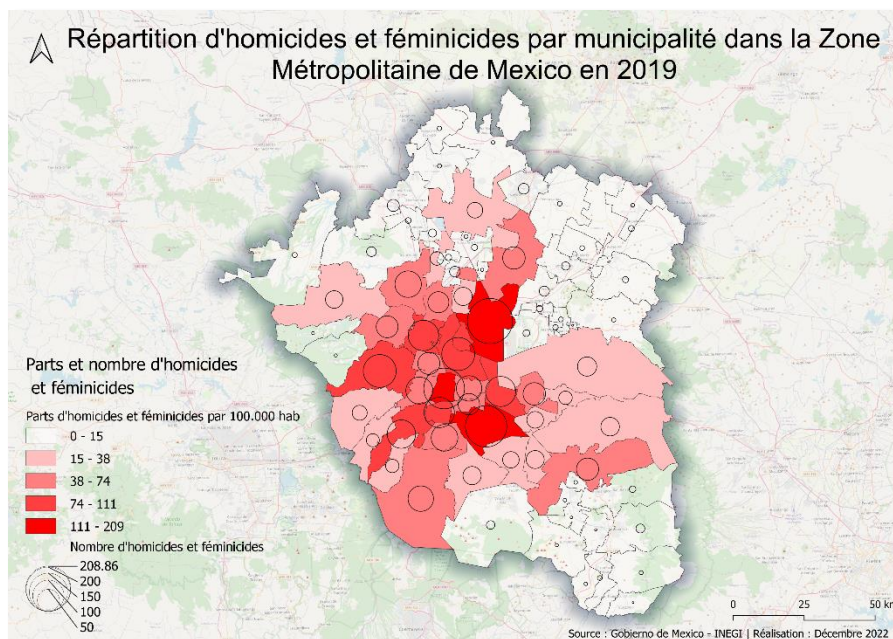


Figure 9 - Carte des homicides et féminicides

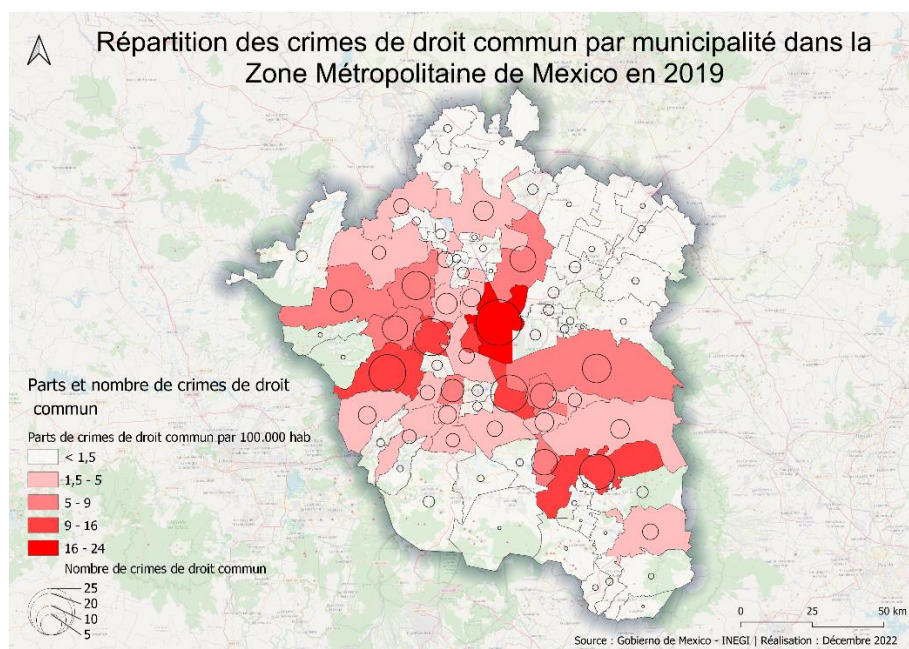


Figure 10 - Carte des délits de droit commun

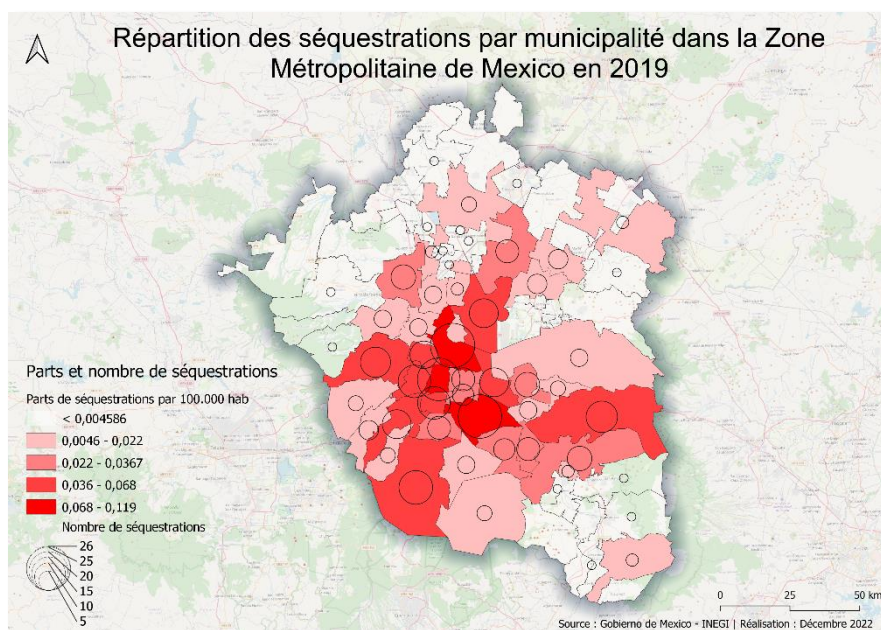


Figure 11 - Carte des séquestrations

5.2 EXPLORATION DES DONNEES CRIMINALITE

5.2.1 Matrice de corrélation, corrélogramme

Notre hypothèse est qu'il existe un lien entre les variables de criminalité sélectionnées.

Nous nous sommes appuyés, pour l'évaluation de la dépendance entre ces variables, sur une matrice de corrélation. Cette table contient les coefficients de corrélation entre les différentes paires possibles.

Nous avons ensuite représenté cette matrice par un corrélogramme, afin de mettre en évidence les variables les plus corrélées. Le coefficient de corrélation est compris entre -1 et 1, d'une corrélation négative vers une corrélation positive : les valeurs se rapprochant de 1 indiquent une forte corrélation entre les variables concernées.

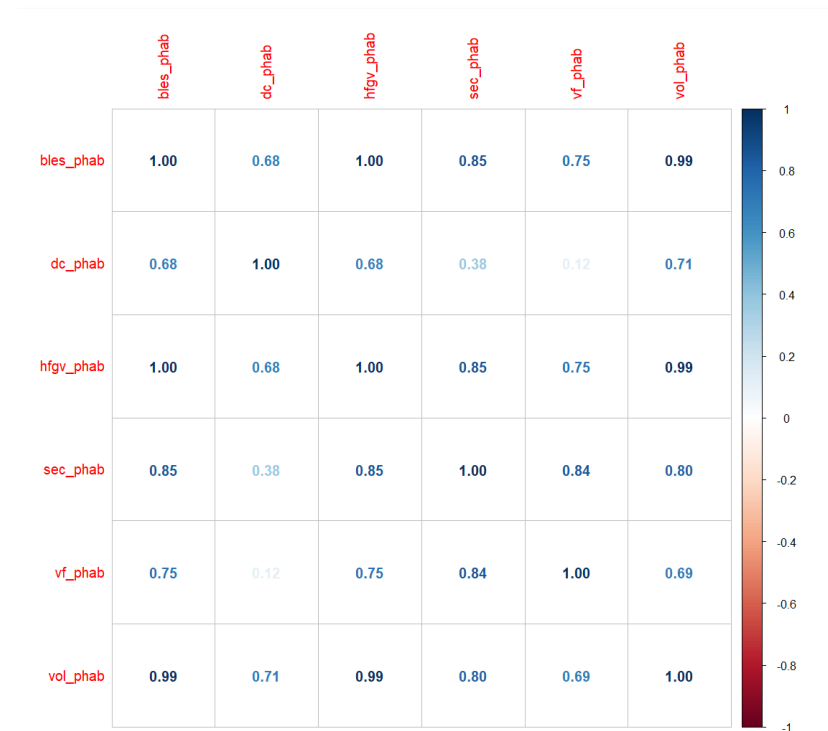


Figure 12 - Corrélogramme des variables de criminalité

Le résultat, présenté ci-dessus, fait état de fortes corrélations positives entre les variables de criminalité. Celles-ci se situent pour la plupart entre 0.68 et 1. Les variables faiblement corrélées sont les variables de séquestration et de violence familiales avec les crimes de droits commun.

Ces résultats restent dans l'ensemble assez prévisibles : on peut penser qu'un homicide comporte des blessures.

5.2.2 P-values et tests de significativité

La p-value est un indicateur du niveau de significativité d'une corrélation et nous permet de nous assurer du lien entre les corrélations observées.

p-value	Blessures	Crimes de droit commun,	Homicide/féminicide	Sequestration	Violence familiale	Vol
Blessures		0.8599	0.0000	0.2328	0.4330	0.0007
Crimes de droit commun,	0.8599		0.8602	0.1242	0.0226	0.5711
Homicide/féminicide	0.0000	0.8602		0.2326	0.4328	0.0007
Sequestration	0.2328	0.1242	0.2326		0.0111	0.4381
Violence familiale	0.4330	0.0226	0.4328	0.0111		0.6989
Vol	0.0007	0.5711	0.0007	0.4381	0.6989	

Figure 13 - P-values des corrélations

Notre test de significativité est plutôt favorable : la p-value est, dans un tiers des cas, inférieure à 0.05.

Hemiptera et Afanipera. Plac. 100

- Homicides et féminicides – Blessures
- Vol – Blessures
- Violences familiales – Crimes de droit commun
- Violences familiales – Séquestration
- Vol – Homicides et féminicides
- Vol – Homicides et féminicides

5.2.3 Analyse en composantes principales (ACP)

Puis pour explorer nos données, nous avons pratiqué une Analyse en Composantes Principales (ACP), de façon à en extraire les informations importantes et les visualiser graphiquement.

Nous avons visé à travers cette analyse, à mettre en évidence une éventuelle explication de nos variables de criminalité par certaines composantes. Les faits criminels, fortement liés en bivarié se trouvent toujours proches les uns des autres dans notre résultat. Par ailleurs, toutes nos variables sont positives et proches de 1 sur le premier plan de notre ACP. L'analyse nous semble davantage permettre d'établir un gradient de la criminalité : à la fois plus ou moins forte, plus ou moins violente, présente ou non.

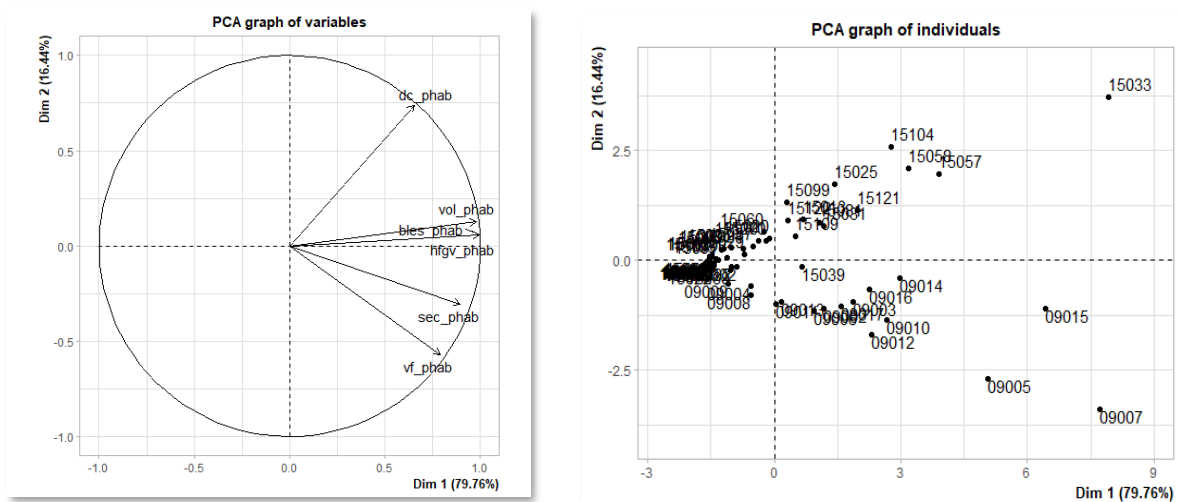


Figure 14 - Cercle de corrélation et nuage de points de l'ACP

Pour valider nos interprétations, nous calculons plusieurs autres paramètres.

Notamment la part que joue chacune des variables dans la constitution de nos 2 premiers axes. Ainsi, le 1^{er} axe est composé de toutes nos variables de manière assez équitable, à l'exception des crimes de droit commun qui sont bien mieux représentés par l'axe 2 de l'ACP. Deuxième axe qui prend également mieux en compte les violences familiales, mais négativement (-0.575).

ACP	Comp.1	Comp.2
Blessures	0.454	
Crimes de droit commun,	0.298	0.742
Homicide/féminicide	0.454	
Séquestration	0.409	-0.307
Violence familiale	0.362	-0.575
Vol	0.447	0.127

Figure 15 - Participation des variables aux composantes

On regarde également la part d'inertie des variables expliquées par nos deux premières composantes. Pour le premier axe, cette proportion est de 79.8%. Ce qui est largement suffisant pour représenter nos variables. Si l'on conserve nos deux premiers axes, cette proportion monte à 96%.

	Comp.1	Comp.2
Standard deviation	2.1875956	0.9931361
Proportion of Variance	0.7975957	0.1643865
Cumulative Proportion	0.7975957	0.9619823

Figure 16 - Explication de l'ACP par ses composantes

Enfin, on observe la corrélation entre notre premier axe de l'ACP et chacune de nos variables de criminalité. Toutes sont corrélées, la plus basse de nos variables ayant une corrélation de 0.65 et les plus hautes de 0.99.

Type de crime	Corrélation
Blessures	0.9936043
Crimes de droit commun,	0.6527556
Homicide/féminicide	0.9936174
Séquestration	0.8956955
Violence familiale	0.7918118
Vol	0.9776109

Figure 17 - Corrélation entre criminalité et Axe 1 de l'ACP

Notre hypothèse de lien entre les variables de criminalité est donc vérifiée.

5.2.4 Classification ascendante hiérarchique (CAH)

Nous avons ensuite procédé à une classification de nos données au moyen d'une Classification Ascendante Hiérarchique (CAH), appliquée aux résultats de l'ACP.

Cette méthode produit un arbre de regroupement hiérarchique, propre à identifier des groupes d'individus présentant des traits communs.

La combinaison de ces deux méthodes, ACP puis CAH, a consisté à :

- D'abord réduire les dimensions de notre jeu initial (son grand nombre de variables) au moyen d'une analyse factorielle : l'ACP ;
- Ensuite, produire des groupes stables sur la base de ses résultats, avec une CAH.

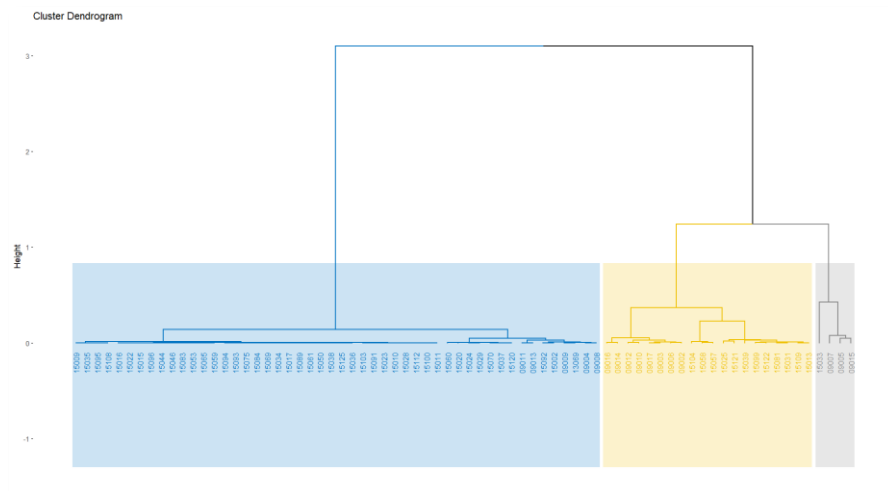


Figure 18 - Dendrogramme

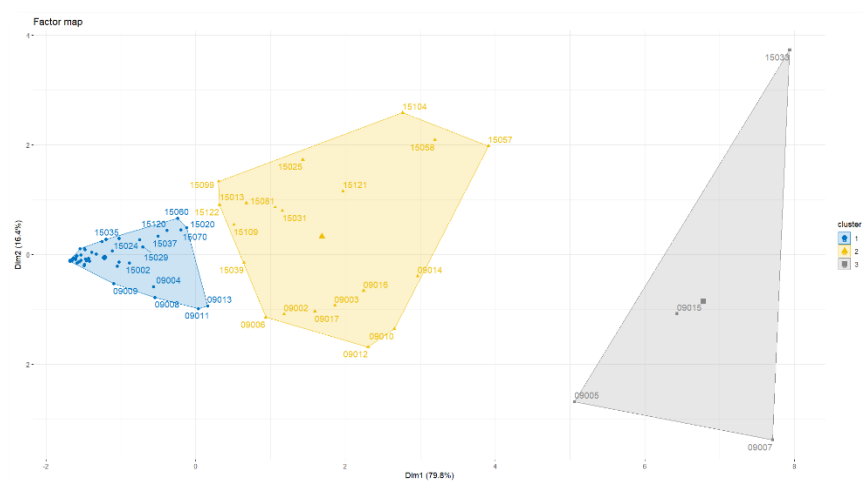


Figure 19 - Clusters de la CAH

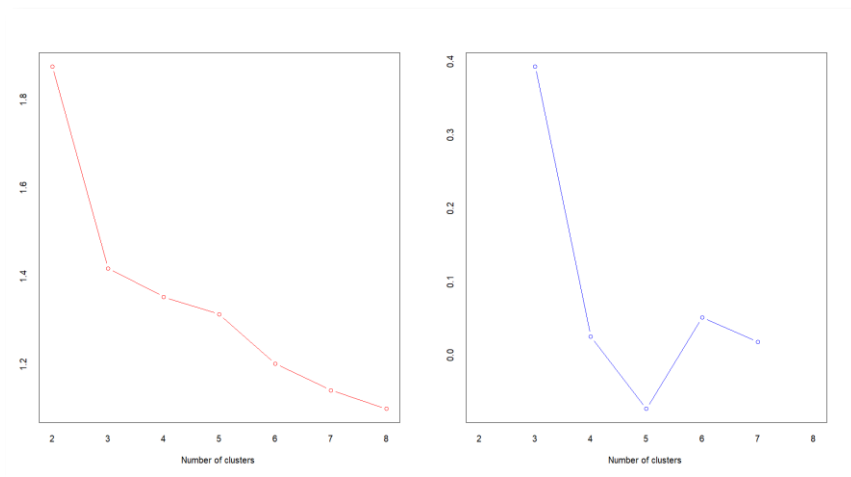


Figure 20 - Sauts d'inertie de la CAH

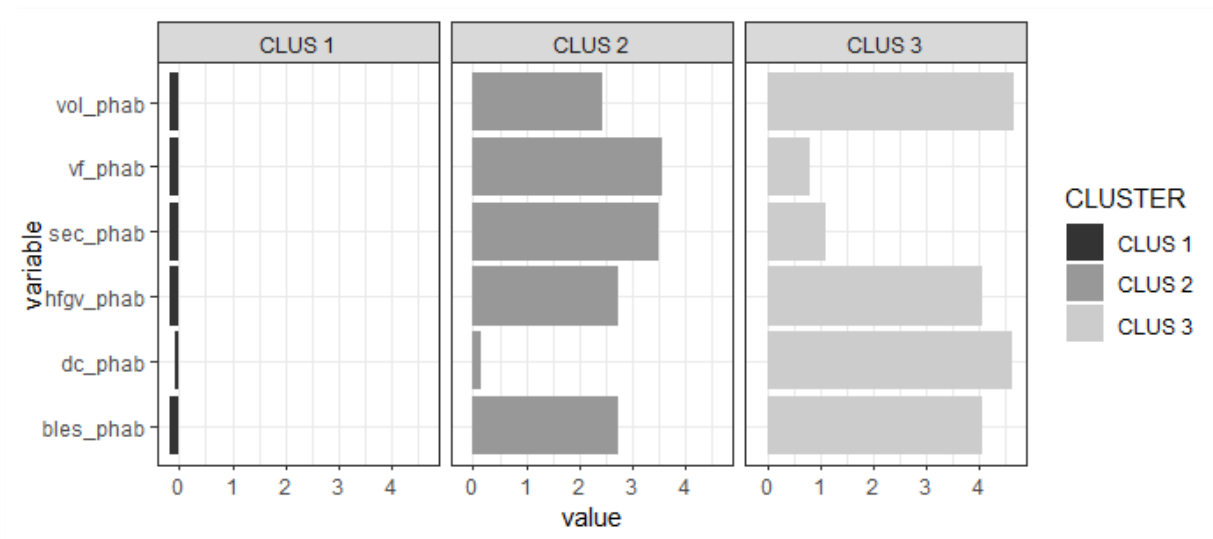


Figure 21 - Profils des clusters

Nous obtenons donc 3 classes :

- CLUS 1 :
 - Représente des Municipales ayant une faible part de criminalité, et ce dans toutes les variables
- CLUS 2 :
 - Regroupe les Municipales présentant une forte proportion de criminalité, à l'exception des délits de droit commun (dc_phab)
- CLUS 3 :
 - Rassemble les Municipales comportant beaucoup de criminalité, avec une sous-représentation des violences domestiques (vf_phab) et des séquestrations (sec_phab)

Nous avons ensuite réalisé une cartographie de cette analyse :

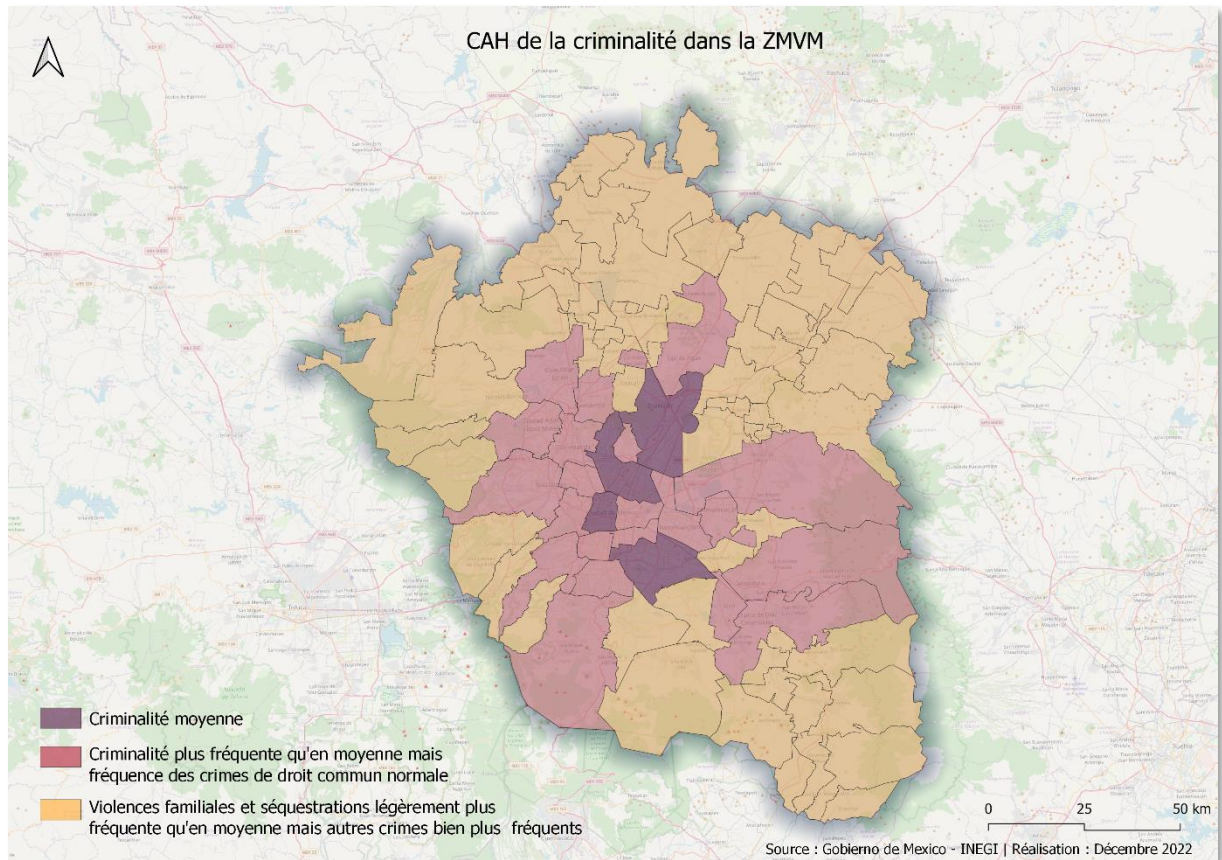


Figure 22 - Cartographie de la CAH

Nous pouvons voir se dessiner une géographie de la criminalité pour la Zone Métropolitaine de la Vallée de Mexico.

On retrouve au centre une criminalité moyenne dans toutes les catégories. En étoile se présente une criminalité forte, mais avec des crimes de droit commun moindres. Dans les franges périphériques du Nord et du Sud, on note une criminalité forte mais moins de violences familiales et de séquestration.

5.3 EXPLORATION DES DONNEES SOCIO-DEMOGRAPHIQUES

Pour l'étude de nos données socio-démographiques, nous nous sommes focalisés sur la représentation de la précarité à Mexico. Pour cela, nous souhaitons voir la relation de nos variables de précarité entre elles. La matrice de corrélation nous a permis d'écarter certaines variables qui ne présentaient pas de corrélation entre elles.

5.3.1 Étude de corrélation

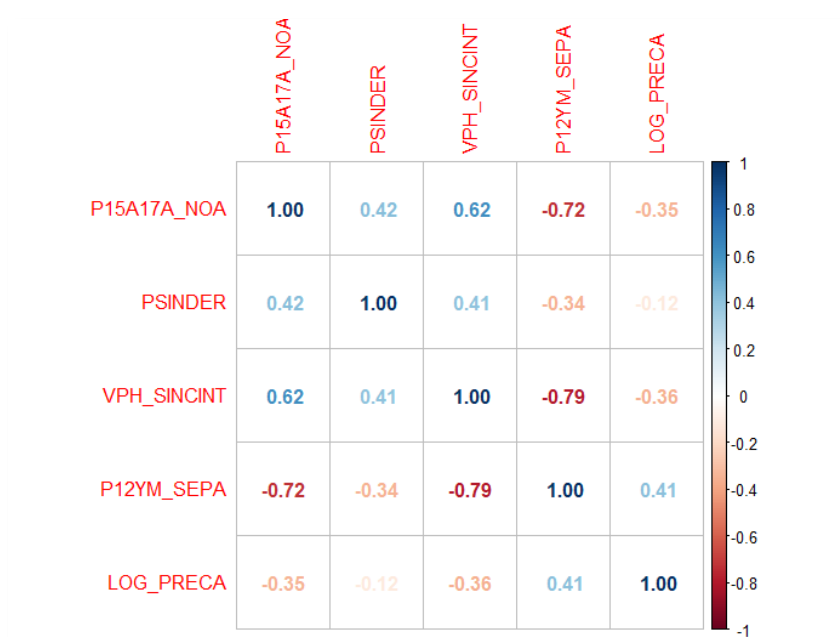


Figure 23 - Corrélogramme

Les variables sont toutes corrélées et de façon générale, fonctionnent bien entre elles.

Parmi les corrélations, la plus importante est la variable des déscolarisés de 15 à 17 ans, elle est négativement et fortement corrélée à la variable des individus séparés (veufs, divorcés, etc.). Elle est aussi corrélée à la variable des ménages ne possédant pas d'équipement de technologie de l'information.

Seuls restent les logements précaires et les personnes n'ayant pas de couverture sociale qui sont très faiblement corrélés (-0,12).

	P15A17A_NOA	PSINDER	VPH_SINCINT	P12YM_SEPA	LOG_PRECA
P15A17A_NOA		0.1412	0.0226	0.0058	0.0806
PSINDER	0.1412		0.1448	0.1316	0.2081
VPH_SINCINT	0.0226	0.1448		0.0018	0.0889
P12YM_SEPA	0.0058	0.1316	0.0018		0.1026
LOG_PRECA	0.0806	0.2081	0.0889	0.1026	

Figure 24 - P-values des variables socio-démographiques

Les résultats des p-values effectuées suivent la dynamique du corrélogramme et présentent une bonne à très bonne significativité. A nouveau, le duo logements précaires-sans couverture sociale obtient un résultat peu convaincant et qui n'est pas significatif.

5.3.2 Analyse en Composantes Principales (ACP)

Pour la création d'un indice de précarité, nous avons décidé de mettre en place une analyse en composante principale (ACP). Ses résultats, s'ils sont pertinents, nous serviront à l'établir.

Les variables sélectionnées regroupent plusieurs types de précarités, vulnérabilités ou fragilités. Aucune d'elle n'est liée à l'inverse à des particularités socio-économiques de la classe moyenne ou supérieure. Ainsi, il est fortement probable que nos résultats nous donnent davantage à voir quant aux types de précarité qu'à un réel gradient. Nous nous pencherons donc sur des questions de typologie des précarités.

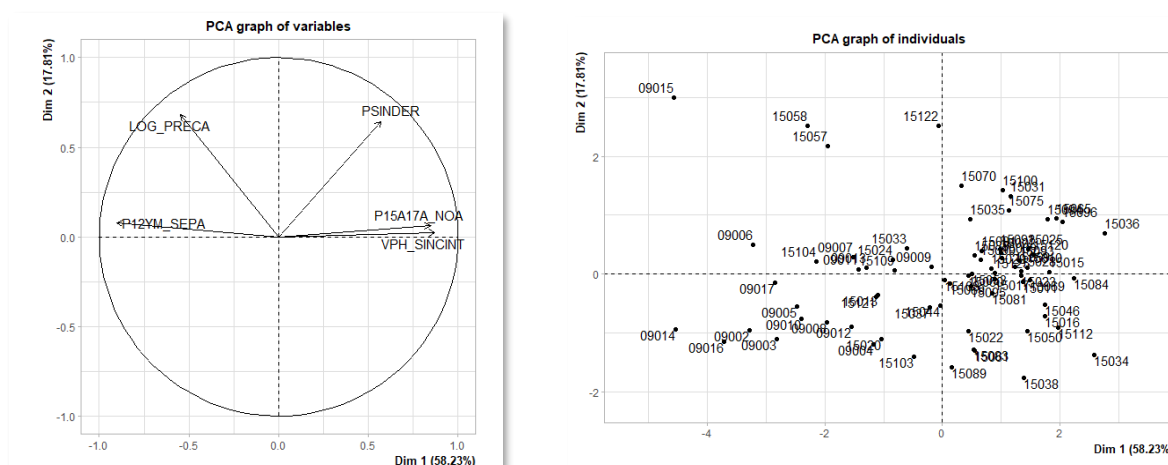


Figure 25 - Cercle et nuage de corrélation de l'ACP

Au départ, notre ACP était composée de 9 variables et nous en avons peu à peu éliminé.

Dans un premier temps, nous avons deux variables liées à la déscolarisation. Nous avons au final choisi de retirer les déscolarisés de 6-14 ans. L'école étant obligatoire jusqu'à 14 ans, cette population semblait indiquer une situation de forte précarité. Néanmoins, ils ne représentaient finalement que

5 % de notre population. Nous avons préféré conserver les déscolarisés de 15-17ans (80 % de la population concernée).

De même, nous avons deux variables liées aux transferts d'argent apportés par une personne émigrée (*remesas*) ou aides financières du gouvernement. La première a été retiré après étude car elle n'était corrélée qu'aux aides financières. Tandis que la seconde ne nous a pas paru pertinente à conserver car elle provenait de notre propre biais normatif français : le système d'aides sociales mexicain n'étant pas aussi performant ni comparable à ce que l'on peut avoir en France.

Enfin, après avoir modélisé plusieurs ACP, nous avons également décidé de soustraire les personnes dans l'incapacité de travailler (inactifs handicapés, retraités, etc.). Les valeurs obtenues n'étaient pas très significatives dès l'analyse bivariée, avec une p-value entre 0,6 et 0,8 et des corrélations relativement faibles au regard des autres. Pour finir, lors de l'ajout de cette variable à l'ACP, elle ne prenait que peu part aux composantes de notre étude.

Ces quatre variables en moins, notre ACP à 5 branches nous paraissait suffisamment solide pour la suite de l'étude. Aussi, le premier axe explique à 58 % notre modèle. Le second l'explique à 18 %. Ces deux résultats nous semblaient convenable pour poursuivre l'étude.

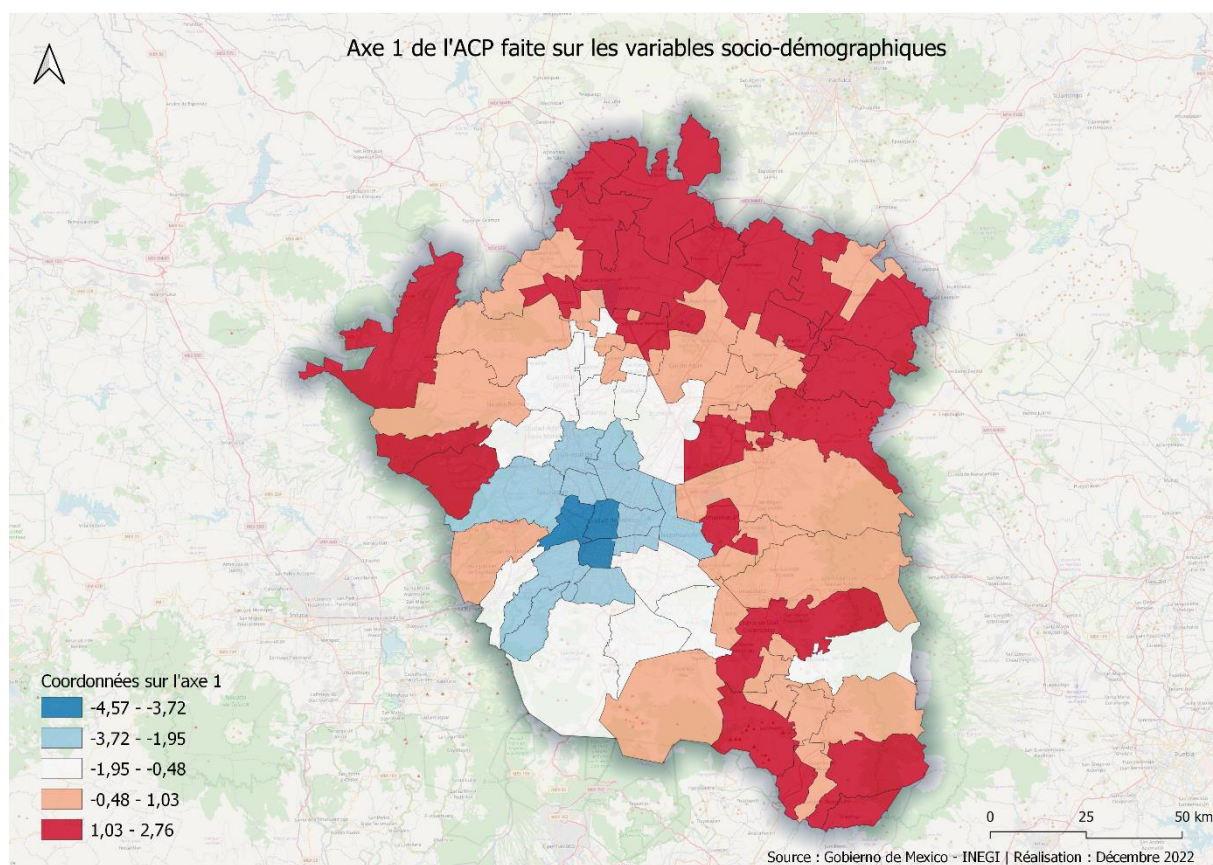


Figure 26 - Cartographie de l'ACP

Plus on tire vers le positif (rouge), plus on a des personnes sans équipements de technologie de l'information, déscolarisés et sans couverture médicale. De l'autre côté, plus on tire vers le négatif (bleu), plus on a de personnes séparées et vivant dans des logements précaires.

On voit donc se dessiner une géographie de la précarité avec au centre de Mexico les municipes tirant vers le négatif, et plus on s'en éloigne et se dirige vers les franges de la ZMVM, plus les résultats tirent vers le positif.

5.3.3 Classification Ascendante Hiérarchique (CAH)

Malgré la réalisation de l'indice via l'ACP, nous avons choisi de passer par une CAH afin d'avoir des classes plus stables.

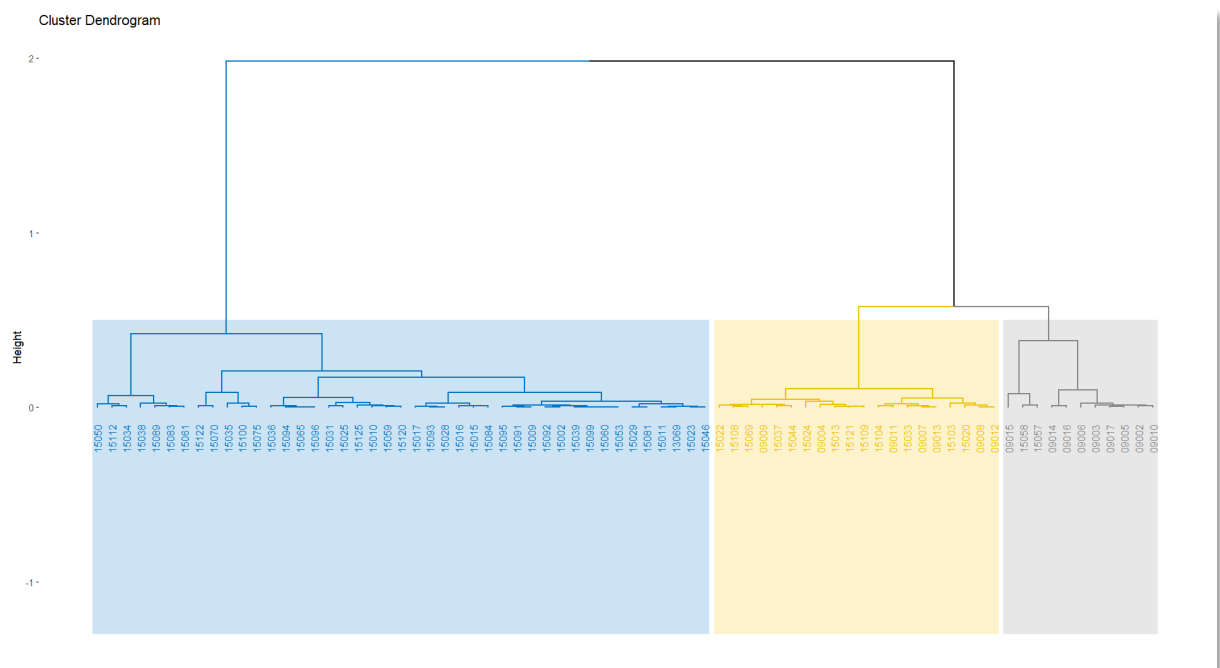
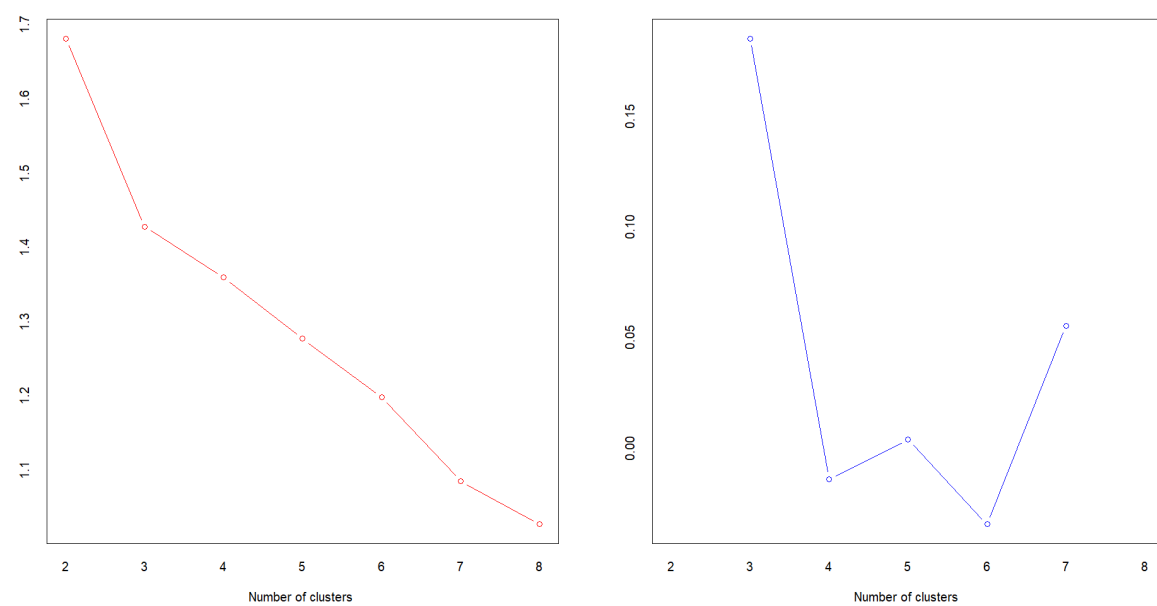
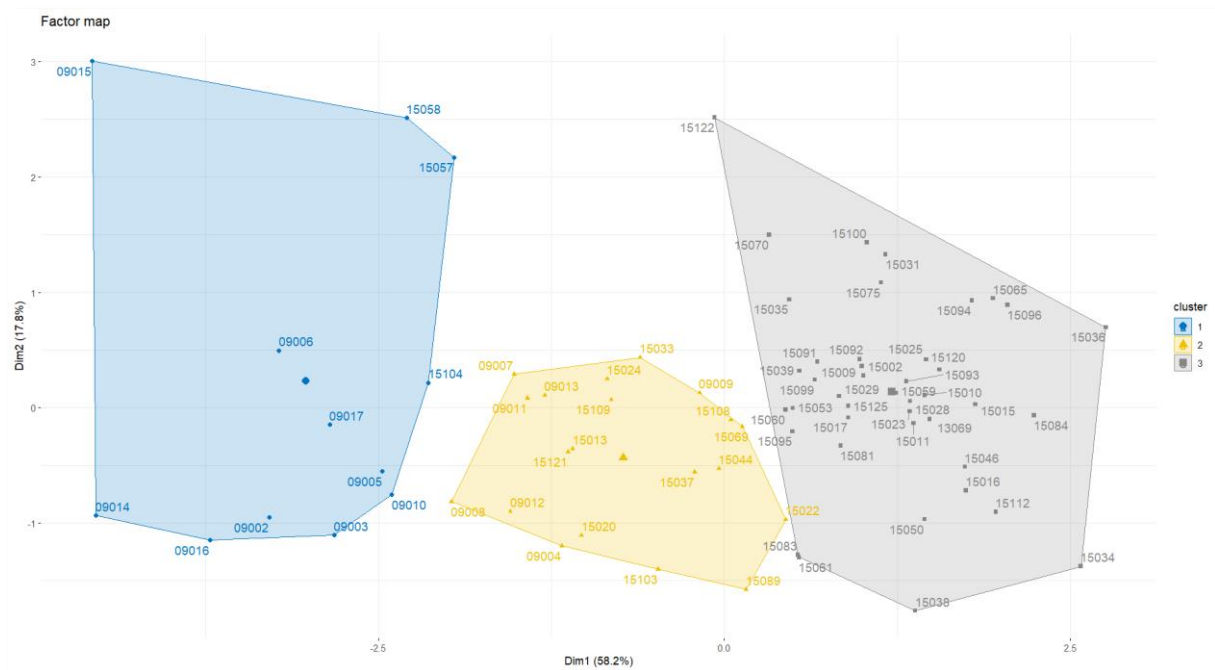


Figure 27 - Dendrogramme de la CAH



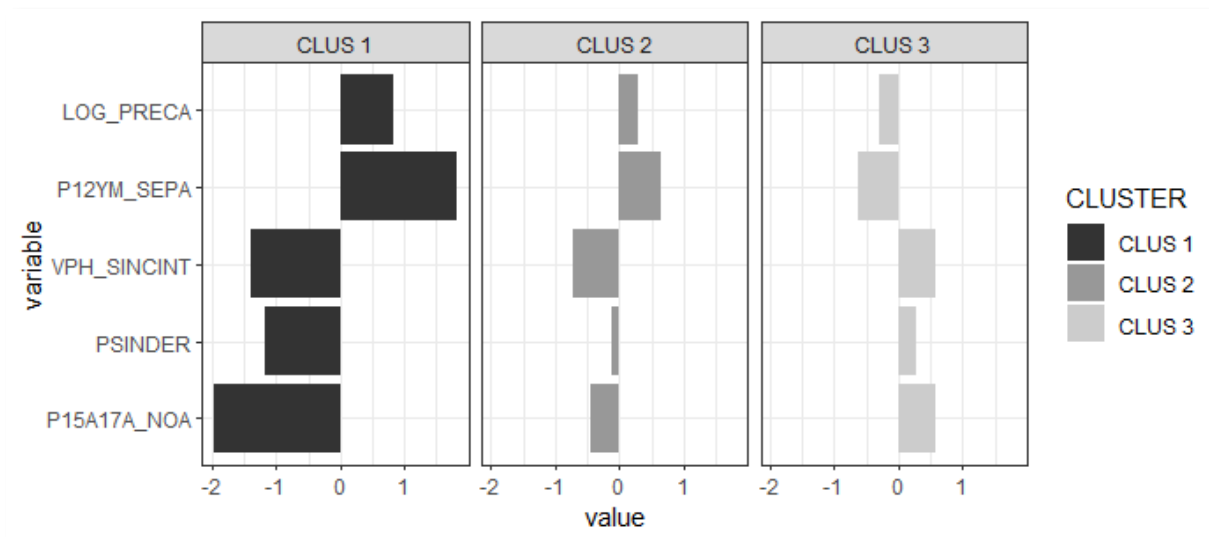


Figure 30 - Profils de clusters de la CAH

Les résultats obtenus nous ont d'abord menés vers la réalisation de 3 classes distinctes représentées dans le dendrogramme. Cependant, lors de l'étude des clusters, nous nous sommes retrouvés avec une première classe contenant des variables extrêmes. Les logements précaires et les personnes séparées y étaient sur-représentées (respectivement +0.4 et +1 ET) tandis que les autres variables y étaient sous représentées (-0.6 à -1 ET). Surtout, les deux autres classes étaient très semblables et peu pertinentes à analyser.

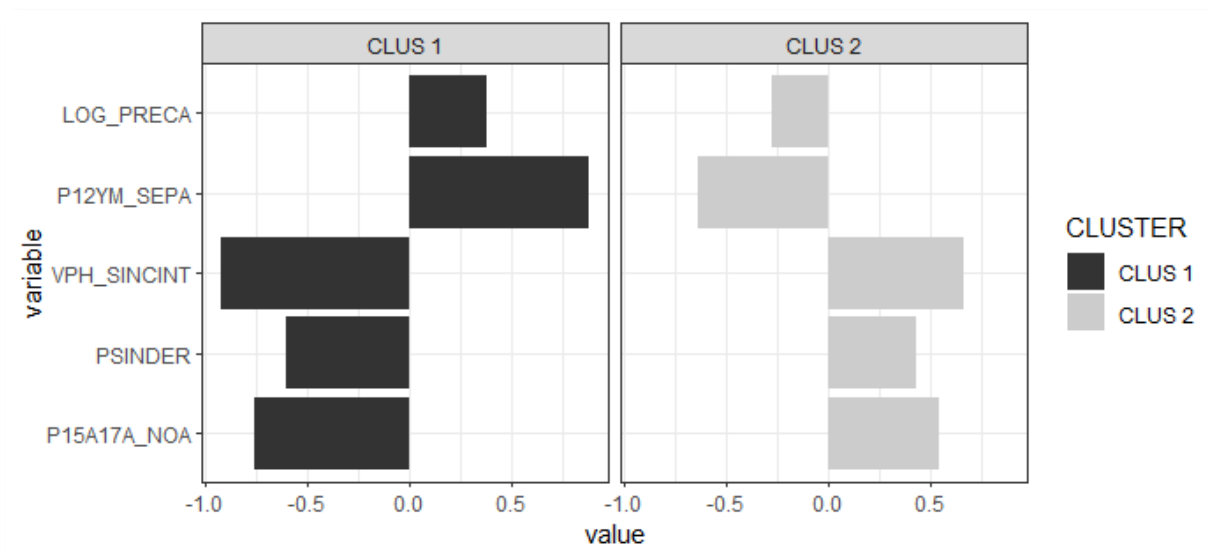


Figure 31 - Profils des deux nouveaux clusters de la CAH

Ainsi, nous avons changé de méthode de classification, passant d'une classification en *complete*, qui maximisait l'écart inter-classe, à une classification suivant le modèle de Ward (l'un des plus fréquents

utilisés dans le clustering). Ce choix nous a permis de descendre à deux classes plus satisfaisantes au lieu de trois mal proportionnés.

Ces nouveaux résultats, bien que peu fins, nous ont permis d'obtenir deux profils en miroir, dont le résultat est la carte suivante.

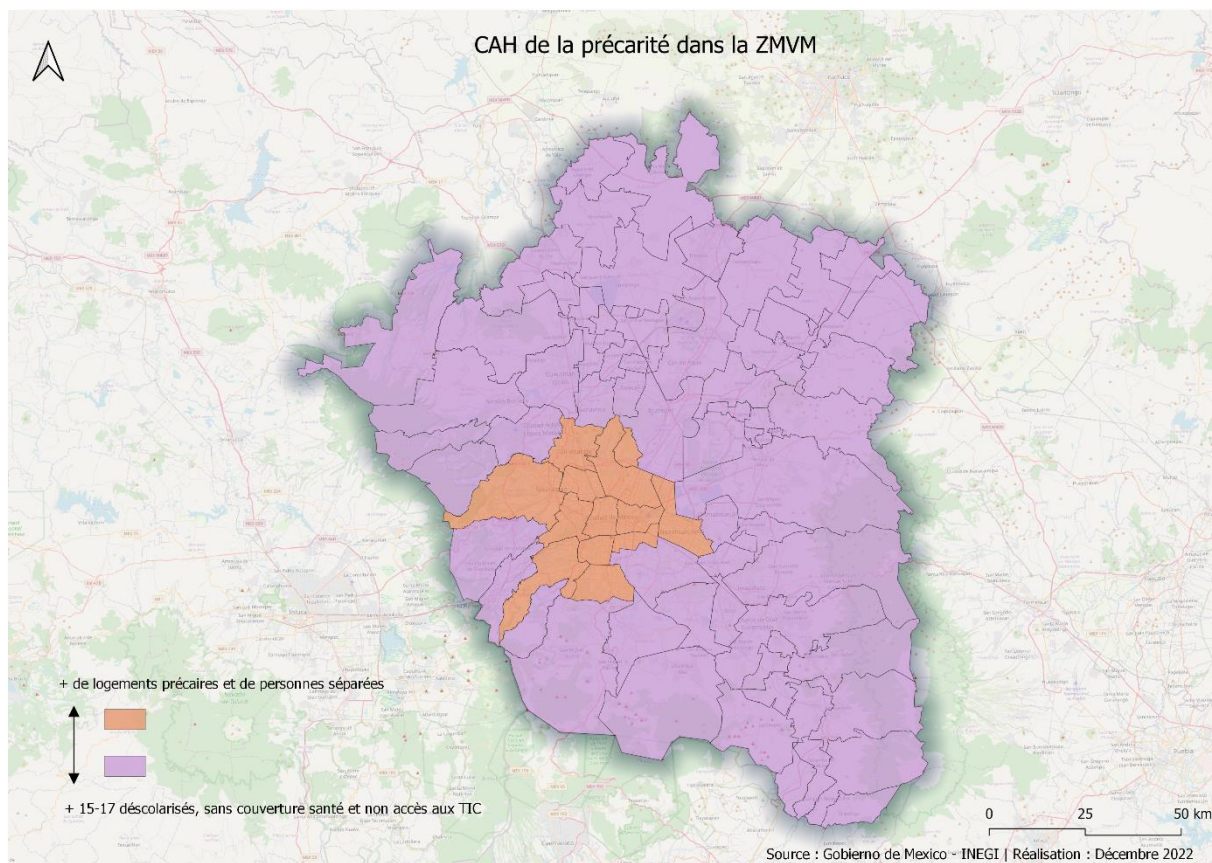


Figure 32 - CAH de la précarité

On voit nettement la répartition entre deux situations de précarité :

- l'une est davantage liée à la précarité du logement et à la situation maritale (séparés) d'un côté. On retrouve principalement cette classe dans le cœur de la ZMVM
- l'autre représente les déscolarisés, sans couverture de santé et sans accès aux TIC. Cette classe est présente dans les franges de la ZMVM

Cette CAH reprend pour beaucoup les résultats obtenus lors de l'ACP, mais nous permet de définir plus précisément une typologie des précarités.

5.4 RAPPORT ENTRE LA PRECARITE ET LA CRIMINALITE

Nous faisons l'hypothèse d'un lien de dépendance entre les variables de criminalité et l'Axe 1 de l'ACP sur la précarité.

Nous calculons déjà des coefficients de corrélation linéaire (Pearson) pour estimer s'il y a un lien de dépendance.

crim_phab	coefficient_corr
sec_phab	-0.6858001
bles_phab	-0.6810913
hfgv_phab	-0.681084
vol_phab	-0.655541
vf_phab	-0.6536845
dc_phab	-0.2678527

Figure 33 - Coefficients de corrélation

Seuls les crimes de droit commun ne semblent pas corrélés à la précarité (coef < 0,65). Tous les autres liens sont significatifs, mais nous choisissons de nous focaliser sur les 2 variables les plus corrélées à la précarité : les séquestrations (avec -0,685) et les blessures (-0,681).

On voit donc des liens suffisamment significatifs, linéaires et négatifs d'ordre moyen entre les variables (quand la précarité diminue, les séquestrations et les délits avec blessures diminuent également).

On cherche alors à formaliser cette relation de dépendance en calculant leur régression linéaire.

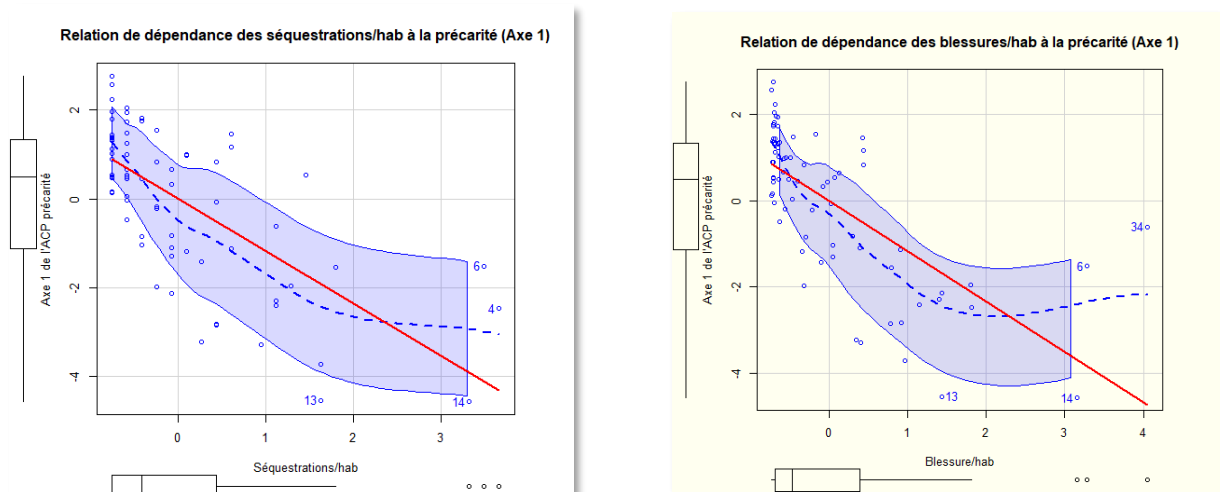


Figure 34 - Graphiques de régression linéaire

Puis nous estimons à l'aide d'un coefficient de détermination (R2) la qualité de la prédiction de la régression linéaire calculée.

```
Call:
lm(formula = y ~ x, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-2.90360 -0.74638  0.07999  0.87928  2.59685

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.373e-16  1.464e-01   0.000      1
x          -1.178e+00  1.473e-01  -7.996 1.56e-11 ***
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.259 on 72 degrees of freedom
Multiple R-squared:  0.4703, Adjusted R-squared:  0.463
F-statistic: 63.93 on 1 and 72 DF, p-value: 1.562e-11
```

Figure 35 - Résumé de régression linéaire « séquestration »

On voit ici que le R2 pour les séquestrations est de 0.463: c'est à dire que 46% de la variation des séquestrations sont expliquées par la composante 1 de l'ACP sur la précarité.

Le coefficient de régression associé à la précarité est égal à -1.178e+00, soit -1. Cela signifie que lorsque la précarité diminue de 1, les séquestrations diminuent également de 1.

La p-value associée à ce coefficient est de 1.56e-11, soit très proche de 0. Cela signifie que l'effet de la précarité sur les séquestrations est statistiquement significatif.

```
[1] NA

lm(formula = y ~ x, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-2.8690 -0.6758  0.0610  0.6416  4.1322

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.406e-17  1.472e-01   0.000      1
x          -1.170e+00  1.482e-01  -7.893 2.43e-11 ***
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.267 on 72 degrees of freedom
Multiple R-squared:  0.4639, Adjusted R-squared:  0.4564
F-statistic: 62.3 on 1 and 72 DF, p-value: 2.428e-11
```

Figure 36 - Résumé de la régression linéaire "blessures"

On voit ici que le R2 pour les délits avec blessures est de 0.4564: c'est à dire que 46% de la variation des blessures est expliquée par la composante 1 de l'ACP sur la précarité.

Le coefficient de régression associé à la précarité est égal à -1.170e+00, soit -1. Cela signifie que lorsque la précarité diminue de 1, les blessures diminuent également de 1.

La p-value associée à ce coefficient est de 2.43e-11, soit très proche de 0. Cela signifie que l'effet de la précarité sur les délits avec blessures est statistiquement significatif.

On voit donc que nos modèles expliquent tous les deux 46% des valeurs pour les séquestrations et les délits avec blessures, ce qui est satisfaisant mais reste assez peu pour expliquer les variables. Néanmoins il nous semblait intéressant de pousser notre étude jusqu'à l'étude des résidus.

On va donc tester les écarts à notre modèle de prédiction, en regardant si les différences entre les valeurs observées et les valeurs prédites (les résidus) s'écartent de manière significative de notre modèle. Ces résidus sont récupérés dans un tableau puis cartographiés :

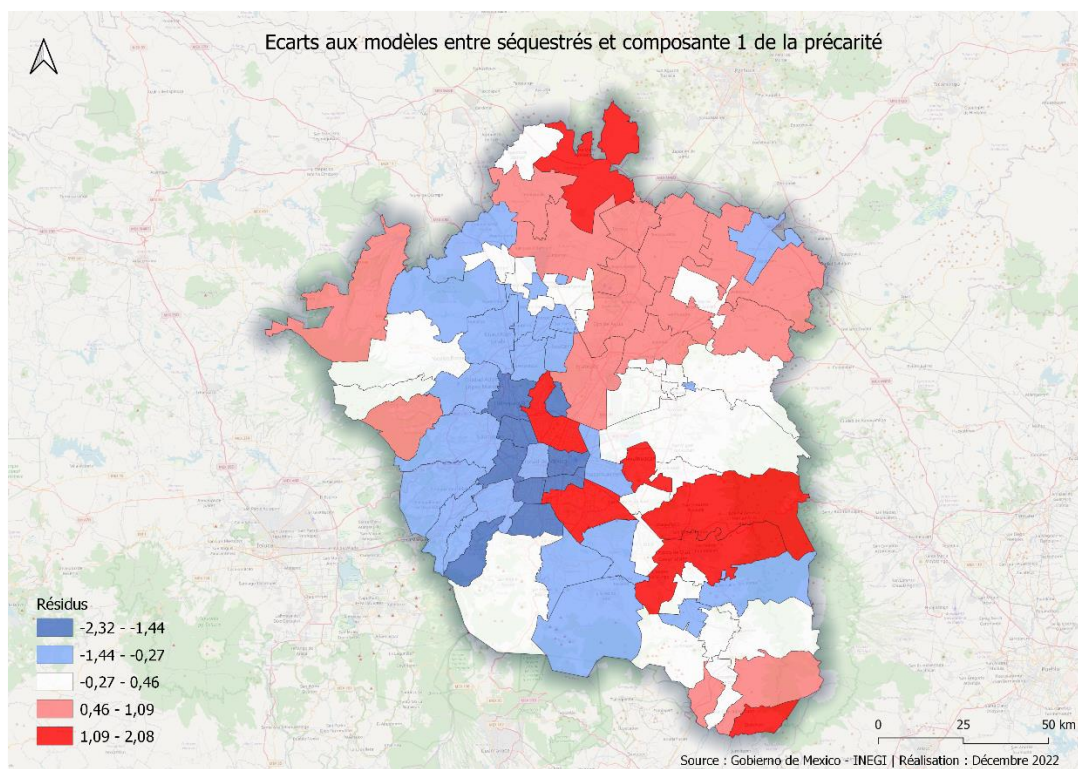


Figure 37 - Cartographie des résidus "Séquestrés"

Sur la partie Ouest, il y a eu moins de séquestrations/enlèvements que ce qui était prévu. Pour la partie Nord-Est, il y en a plus que ce que le modèle prédisait. De plus certains municipes aléatoirement répartis sur la ZMVM (rouge vif sur la carte) sont nettement supérieur aux taux attendus.

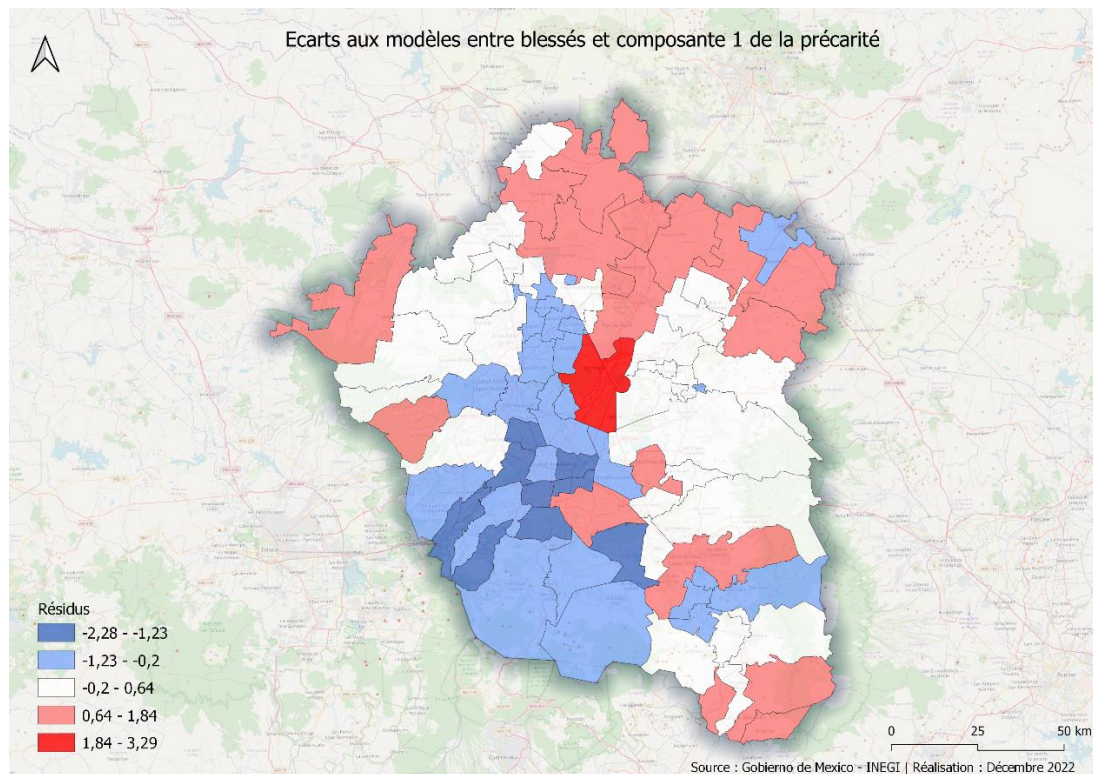


Figure 38 - Cartographie des résidus "Blessures"

Notre modèle prédit d'avantage de blessés dans le Sud/Sud-Ouest de la ZMVM qu'il n'y en a en eu réalité. Et inversement dans le Nord-Ouest. De nombreuses valeurs sont proches de 0 à l'Ouest et à l'Est, notre modèle semblant plus exact sur ces municipes.

CONCLUSION

Pour étudier la criminalité, nous avons récupéré des données du gouvernement et de l'INEGI. Parmi elles, nous avons pu extraire des variables traduisant des faits criminels à la fois très nombreux mais aussi culturellement et socialement significatifs. Avec une ACP, nous avons d'abord réduit les dimensions de notre jeu de données initial. Puis grâce à une CAH, nous avons produit une classification. Ainsi dans ces variables, pourtant très hétérogènes, nous avons pu établir des corrélations et vérifier notre hypothèse de base en calculant des classes de criminalité. Cette typologie décrit une géographie de la criminalité dans la ZVMV.

En parallèle, nous avons dissocié la précarité de la pauvreté en la définissant comme une situation d'instabilité, de fragilité et d'incertitude. A partir des 300 variables socio-démographiques de l'INEGI, nous avons composé un bouquet cohérent et corrélé de variables, issus de différentes thématiques et nous permettant d'en étudier une éventuelle porosité ou contagiosité. Par la nature de notre étude, nous avons pu identifier des types de précarité plutôt qu'un gradient. A l'issue d'une ACP, nous avons pu produire un axe de composition qui expliquait jusqu'à 58 % de notre modèle. Nous avons pu cartographier ce phénomène, qui présente un centre marqué par une précarité issue de logements précaires et d'une situation maritale fragile. Dans les franges extérieures, nous voyons apparaître une précarité dues à un faible taux d'équipement technologique, la déscolarisation adolescente et l'absence de couverture médicale. Afin de définir des classes de façon plus stables, nous avons réalisé une CAH. Cette classification nous a permis d'obtenir deux profils en miroir de la précarité, qui vérifient la spatialisation découverte avec l'ACP.

Enfin nous avons posé l'hypothèse d'un lien entre précarité et criminalité. Mis à part les crimes de droit commun, l'ensemble des variables de criminalité sont significativement et négativement corrélées à la précarité. Nous avons décidé de formaliser cette relation de dépendance sur les variables les plus corrélées : les séquestrations et les blessures. Cependant, bien que significatif, la qualité de notre modèle n'est pas optimale : seulement 46 % de la variation des séquestrations et des blessures sont expliquées par la composante 1 de notre ACP sur la précarité. Malgré cela nous avons décidé de modéliser puis cartographier les écarts à notre modèle, ce qui nous a permis de mettre en lumière, même si de façon relative, des phénomènes de spatialisation.

Ainsi, on peut conclure que la précarité, dans ses définitions d'instabilité, de fragilité et d'incertitude, peut expliquer à moitié les phénomènes de criminalité au sein de la métropole de Mexico pour l'année 2020.

D'un point de vue critique, notre étude ne se positionne qu'à une échelle géographique, celle des municipales. Il pourrait être intéressant, afin d'éviter toute erreur écologique, de travailler les calculs socio-démographiques à des échelles plus fines. Cependant ce n'est pas possible sur la criminalité car les données ne sont disponibles qu'à cette échelle.

De plus, on ne peut séparer l'analyse cartographique d'un territoire sans une connaissance approfondie des relations et enjeux de pouvoir sur celui-ci. N'étant pas notre cas, cela constitue une limite dans notre étude.

Enfin dans une poursuite de notre étude, il pourrait être intéressant de récupérer de plus nombreuses variables socio-démographiques qui ne pesaient pas dans le sens de la précarité afin de nuancer nos résultats sur celle-ci et potentiellement créer un gradient.

Nous réalisons aussi avec le recul qu'en terme de technologie, nous aurions apprécié et facilité notre travail en intégrant directement nos données sur PostgreSQL.

Le projet ainsi mené a été particulièrement instructif et fécond. D'abord par sa dimension transversale, puisqu'il a mobilisé et articulé les savoirs acquis au long de notre cursus, diverses technologies, compétences et techniques au service de la thématique que nous nous sommes choisie et des livrables attendus. À cet égard, il relève à la fois d'une forme de recherche, mais aussi d'une activité de développement et de production.

Par ailleurs, le caractère massif et dense des volumes de données manipulées, en langue étrangère, nous a incité à porter un regard critique indispensable sur nos pratiques et les choix méthodologiques envisagés quant à nos traitements.

Enfin, le cadre de cette démarche réflexive, la géographie de l'aire métropolitaine de Mexico, a constitué l'opportunité d'aborder et de découvrir les spécificités d'un territoire, ainsi qu'une thématique qui nous étaient inconnues.

Nous remercions vivement M. Jean-François Valette, notre tuteur, pour son soutien et ses conseils tout au long de ce projet, ainsi que l'ensemble de l'équipe pédagogique du Master G2M de Paris 8 pour leurs enseignements.

Bibliographie

- [1] F. Johansson, «Mexico : pouvoir ou contre-pouvoir d'une métropole émergente ?,» chez *Festival de Géopolitique de Grenoble*, Grenoble, 2017.
- [2] M. Guerrien, «Transformation et fragmentation des espaces urbains. Le cas de la zone métropolitaine du bassin de Mexico,» *L'espace géographique*, pp. 336-352, Avril 2004.
- [3] A. Ribardière et J.-F. Valette, «La pauvreté urbaine à Mexico : une approche de la diversité des colonies populaires,» *EchoGéo*, n° 130, 2014.
- [4] A. Ribardière et B. Tallet, *Atlas Mégapoles : Mexico*, Autrement, Éd., 2011.
- [5] D. Valle-Jones, «<https://elcri.men/>,» 2022. [En ligne]. Available: <https://elcri.men/>.
- [6] Collectif, «<https://fr.wikipedia.org/wiki/Criminalit%C3%A9>,» 12 08 2022. [En ligne]. Available: <https://fr.wikipedia.org/wiki/Criminalit%C3%A9>.
- [7] Gobierno de Mexico, «Datos Abiertos de Incidencia Delictiva,» 2022. [En ligne]. Available: <https://www.gob.mx/sesnsp/acciones-y-programas/datos-abiertos-de-incidencia-delictiva>.
- [8] Larousse, «Définitions : précarité - Dictionnaire de français Larousse,» 2022. [En ligne]. Available: <https://www.larousse.fr/dictionnaires/francais/pr%C3%A9carit%C3%A9/63303>.
- [9] Le Robert, «précarité - Définitions, synonymes, conjugaison, exemples,» 2022. [En ligne]. Available: <https://dictionnaire.lerobert.com/definition/precarite>.
- [10] A. Villeneuve, « Construire un indicateur de précarité : les étapes d'une démarche empirique,» *Économie et Statistique*, pp. 93-105, Juillet-Août 1984.

TABLE DES FIGURES

Figure 1 - Mailles géostatistiques INEGI (réal. : Jean-François Valette)	4
Figure 2 - Zone Métropolitaine de la Vallée de Mexico	7
Figure 3 - Diagramme de présentation des technologies mobilisées dans notre projet	8
Figure 4 - Diagramme de choix des variables socio-démographiques	10
Figure 5 - Tableau de fréquence des variables de criminalité sélectionnées	11
Figure 6 - Diagramme des variables de criminalité sélectionnées	12
Figure 7 - Carte des vols	13
Figure 8 - Carte des violences familiales	14
Figure 9 - Carte des homicides et féminicides	14
Figure 10 - Carte des délits de droit commun	15
Figure 11 - Carte des séquestrations	15
Figure 12 - Corrélogramme des variables de criminalité	16
Figure 13 - P-values des corrélations	17
Figure 14 - Cercle de corrélation et nuage de points de l'ACP	18
Figure 15 - Participation des variables aux composantes	19
Figure 16 - Explication de l'ACP par ses composantes	19
Figure 17 - Corrélation entre criminalité et Axe 1 de l'ACP	19
Figure 18 - Dendogramme	20
Figure 19 - Clusters de la CAH	20
Figure 20 - Sauts d'inertie de la CAH	21
Figure 21 - Profils des clusters	21
Figure 22 - Cartographie de la CAH	22
Figure 23 - Corrélogramme	23
Figure 24 - P-values des variables socio-démographiques	24
Figure 25 - Cercle et nuage de corrélation de l'ACP	24
Figure 26 - Cartographie de l'ACP	25
Figure 27 - Dendogramme de la CAH	26
Figure 28 - Clusters de la CAH	27
Figure 29 - Sauts d'inertie de la CAH	27
Figure 30 - Profils de clusters de la CAH	28
Figure 31 - Profils des deux nouveaux clusters de la CAH	28
Figure 32 - CAH de la précarité	29
Figure 33 - Coefficients de corrélation	30
Figure 34 - Graphiques de régression linéaire	30
Figure 35 - Résumé de régression linéaire « séquestration »	31
Figure 36 - Résumé de la régression linéaire "blessures"	31
Figure 37 - Cartographie des résidus "Séquestrés"	32
Figure 38 - Cartographie des résidus "Blessures"	33