

Analyse spatiale avancée

Etude spatiale de la criminalité à San-Francisco



Nathan FOLMER
Master 2 Geomatique
2022/2023

Professeur encadrant:
Jean Pierre Chéry

Sommaire:

<i>Introduction:</i>	3
<i>Ressources utilisées:</i>	4
<i>Démarche:</i>	4
<i>Analyse du semis de point:</i>	5
<i>Mesure de l'autocorrélation spatiale</i>	6
<i>Mesure des variations locales de la criminalité</i>	8
<i>Conclusion:</i>	10
<i>Bibliographie</i>	11

Introduction:

À travers ce rapport, l'objectif est de construire une analyse spatiale de la criminalité à San Francisco. C'est à dire de comprendre les causes de la distribution et de la concentration du crime dans cette ville. La criminalité se définissant ici comme l'ensemble des actes illégaux, délictueux et criminels. Dans le cas de cette étude, la criminalité fera référence aux vols à main armée.

Sous l'angle de la criminalité écologique, l'intérêt ne sera pas de comprendre pourquoi un délinquant commet un délit mais plutôt de discerner et évaluer quelles sont les caractéristiques physiques et social du territoire dans lequel l'acte criminel est commis.

L'objectif de ce projet est donc d'une part de comprendre à quelles échelles il est possible de travailler pour évaluer le crime. Est-ce à l'échelle de la ville, du quartier, ou plutôt de l'îlot? D'autre part, il serait intéressant de déterminer les variables utiles à l'examen des espaces de concentrations de la criminalité. Multiples sont les déterminants capables de les qualifier. Certains sont dits physiques lorsqu'ils portent leurs attentions sur les variations temporelles/ climatiques. D'autres sont dit social en évaluant par exemple le mode d'urbanisme, ou bien des paramètres socio-économique.

À partir de multiples lectures citées en dernière partie, cette analyse spatiale se construira à partir d'une théorie dite de « désorganisation ». Elle a été développée par l'école de Chicago et expliquerait l'influence que pourrait avoir un quartier sur une personne et notamment dans une probabilité que cette dernière commette un crime. Selon cette théorie, le taux de criminalité est intimement lié aux caractéristiques propres des quartiers concernées.

Selon Merton(1953): « *l'enrichissement personnel prôné comme le but à atteindre par la société américaine favoriserait la délinquance en ce sens où en raison des inégalités socio-économiques de départ, les plus pauvres seraient incités à recourir à des moyens illégaux, en l'occurrence la délinquance, pour parvenir à leurs fins.* » Ainsi pour ce travail, il serait pertinent de disposer de certaines variables socio-économiques comme le revenu médian des ménages, ou bien le taux de chômage pour mesurer les écarts entre quartiers voire une ségrégation spatiale qui pourrait potentiellement jouer un rôle sur le taux de criminalité car moteur d'une désorganisation social.

Pour Hillier et Sahbaz(2009), une mixité fonctionnelle de l'espace bâti concentrerait plus de délits que dans des espaces mono-fonctionnelles: « *Les populations aisées ont tendance à se représenter l'espace urbain comme potentiellement dangereux: notamment lorsque les écarts de revenus y sont importants* ». Dans un sentiment renforcé de rester entre personnes de mêmes classes sociales, des quartiers clos comme on en connaît aux USA permettraient de se protéger de la misère et aussi de la criminalité. Ces espaces privés, très souvent résidentiels avec peu d'activités mélangés donnerait lieu à un meilleur contrôle social/surveillance.

Ainsi l'étude de la distribution et de la concentration du crime à San Francisco prendra forme autour de deux hypothèses:

1- San Francisco connaît une forme de concentration spatiale des vols à main armée qui coïncide spatialement avec des inégalités socio-économiques représentées ici par le revenu médian par ménage. Les quartiers avec des faibles revenues sont plus enclin à avoir une forte criminalité.

2- L'occupation du sol joue un rôle sur la distribution des crimes dans la ville. Il y a moins de criminalité dans les quartiers résidentiels et plus de crimes dans les quartiers qui concentre à la fois des activités économiques, des bureaux mais aussi des logements.

Ressources utilisées:

La ville de San Francisco met à disposition un portail de données libres et gratuites offrant l'opportunité de récupérer des informations diverses et variées. Ainsi, le fond de carte de la ville, les quartiers, l'occupation du sol (couche de bâtiments) et les données de crimes ont été téléchargés et utilisés grâce à leur format vectoriel. Ces dernières informations forment donc des points localisés dans l'espace et dans le temps. La donnée date de l'année 2011. Une première sélection a été faite afin d'extraire les crimes: « vols à main armée » parmi l'ensemble des 34000 crimes commis cette année là.

Les données sur la population, et autres variables socio-économiques ont été extrait de statisticalatlas.com et ont été manuellement ajoutées aux couches vectorielles à travers la création de nouveaux paramètres.

Démarche:

Dans un premier temps, il serait intéressant d'analyser le semis de point représentant dans l'espace la répartition des vols à mains armés. Pour cela, la forme du semis sera définie à travers deux méthodes. La construction de l'indice de dispersion R nous permettra de prouver que les points de criminalités sont dispersés ou alors à l'inverse, tendent à se concentrer.

Dans un deuxième temps, l'idée sera de caractériser la ressemblance d'entités voisines grâce à des indicateurs d'autocorrélation. En l'occurrence, il y sera comparé le nombre de vols à mains armées et de constater ou non si les quartiers contigus se voient similaires au vu de l'intensité du crime. Une cartographie de "hot spots" et "cold spots" y sera présentée. À cette même étape, l'intérêt sera aussi de mettre en lumière les discontinuités spatiales autour de la mesure du revenu médian par foyer et de comprendre s'il existe un rapport avec les résultats obtenus par l'autocorrélation. Cette démarche nous permettra d'examiner peut-être que les quartiers avec une forte concentration des vols à mains armées forment aussi des espaces de ruptures socio-économiques avec les autres quartiers qui l'entourent.

Enfin dans un troisième temps, toujours dans une dynamique de comprendre la distribution du crime à San Francisco, l'analyse à partir des limites administratives des quartiers sera dépassée et nous tenterons de comprendre s'il peut y avoir une délimitation plus fine des différentes intensités du crime. À cette même étape, une étude du bâti y sera menée afin de vérifier si au sein même des quartiers, le crime tend à se produire plus dans des espaces multi-fonctionnels.

Analyse du semis de point:

Afin de caractériser la forme d'un semis de points, deux méthodes s'offrent à nous.

La première est celle de la méthode des quadrats et la seconde est celle de la méthode du plus proche voisin. La confrontation des deux résultats de ces versions nous permettra de confirmer la forme du semis. Dans la méthode des quadrats, la première étape consiste à former un carroyage régulier sur la zone d'étude (*figure 1*). Ici, une grille couvrant l'entièreté de la ville de San Francisco a été mise en place. Chaque carreau a une largeur et une longueur de 500m. Le résultat forme une grille de 691 polygones. L'idée maintenant est de dénombrer les points (825) représentant les vols à mains armées localisés dans chacun des carreaux. Ceci nous permet de récupérer la densité moyenne de point par surface:

D (densité) = N (points)/ M (nombre de mailles) = 1,19. L'indice de concentration IC se calcule par la variance de cette densité divisée par la densité: $IC = 17,3 / 1,19 = 14,53$

L'indice étant largement supérieur à 1, la distribution est concentrée selon cette méthode.

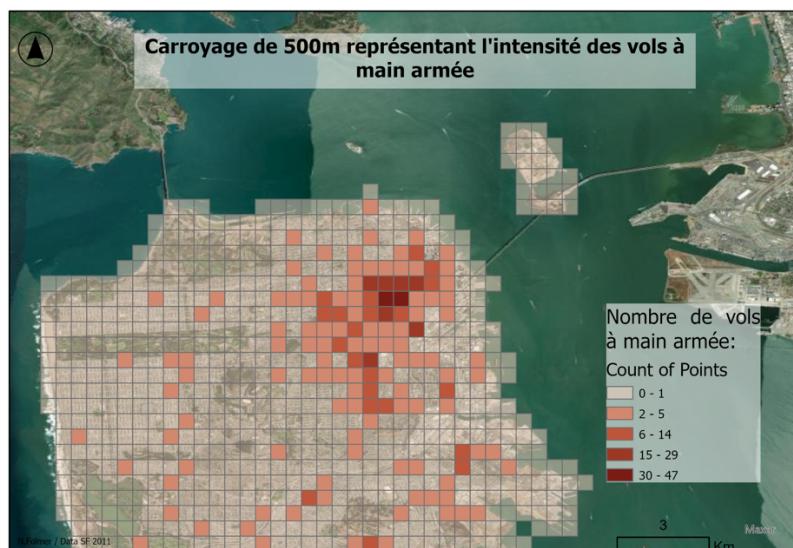


Figure: 1

Dans la méthode du plus proche voisin, on calcule pour chaque point du semis la distance minimale qui le sépare de son voisin le plus proche. Puis on forme la moyenne de ces distances au plus proche voisin. Ce résultat forme la Distance dite « observée ». La distance dite « théorique » est quant à elle déterminé à partir d'une formule:

$$\text{Distance théorique} = 0,5/\text{var(Distance observée)}.$$

Enfin l'indice de dispersion R donne le rapport entre ces deux distances et informe si le semis de point a plus tendance à se disperser (valeur supérieure à 1,5), à se concentrer (valeur proche de 0) ou bien tend vers une répartition aléatoire (environ égal à 1).

Sous Arcgis Pro, on voit que cette valeur (*Nearest Neighbor Ratio*) est égal à 0,53 ce qui indique une concentration du semis de point (*figure 2*).

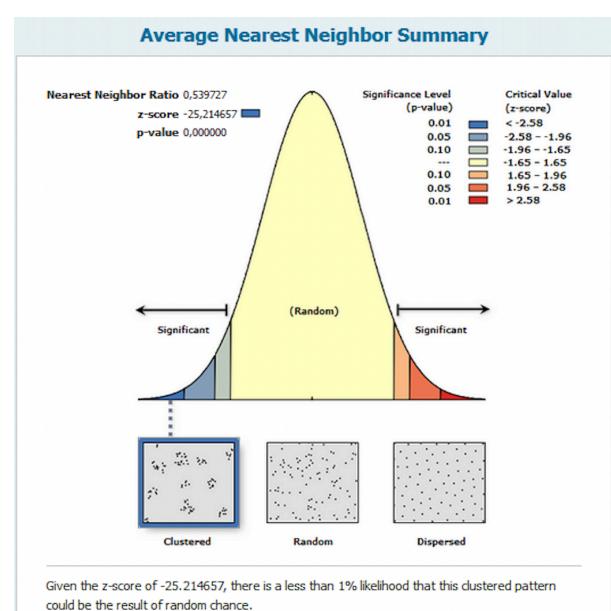


Figure: 2

Au regard des deux méthodes, les résultats sont identiques. Il y a une concentration de la criminalité à San Francisco.

Mesure de l'autocorrélation spatiale

L'autocorrélation spatiale permet de dépasser la simple analyse statistique du semis de point en intégrant une dimension spatiale de la répartition des vols à main armée. Grâce à l'indice local de Moran, il est possible d'identifier les voisins qui se ressemblent. Ceux-ci peuvent avoir similarités parce qu'ils sont proches spatialement et aussi proche par rapport à la valeur de la variable analysée. C'est l'outil « Analyse de grappes et de valeurs aberrantes » qui sera utilisé depuis Arcgis Pro afin de déterminer les différentes zones d'agrégation à partir de la classification suivante:

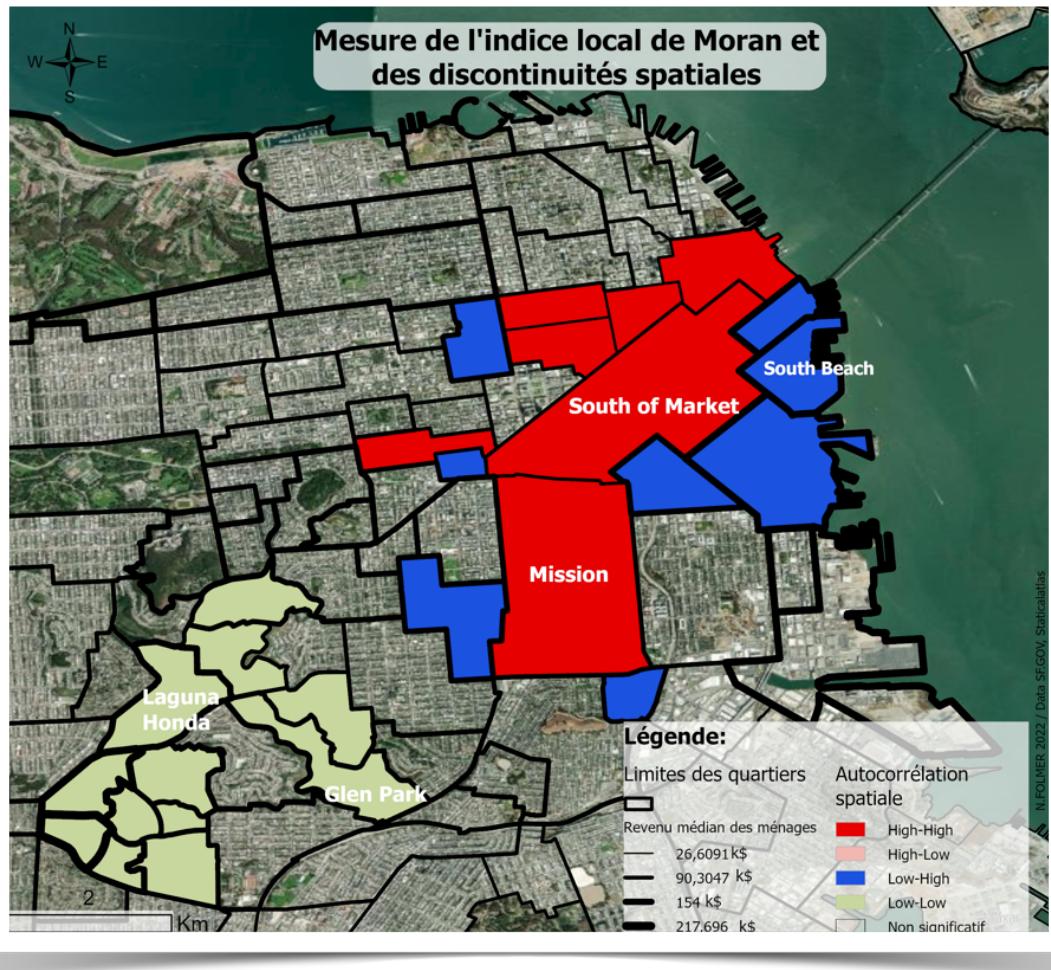
- *High-high: Le voisinage se ressemble et les valeurs sont fortes. Auto-corrélation positive*
- *High-Low: Le voisinage ne se ressemble pas mais les valeurs sont fortes. Auto-corrélation négative*
- *Low-High: Le voisinage ne se ressemble pas avec des valeurs plutôt faibles; Auto-corrélation négative*
- *Low-Low: Le voisinage se ressemble avec des valeurs faibles. Auto-corrélation positive*
- *Non significante: Correspond aux individus et leurs voisins ayant des valeurs moyennes qui ne les permettent pas de les classer réellement.*

L'objectif principale pour l'étude est ici de sélectionner les quartiers qui ont auto-corrélation positive avec des valeurs fortes. À partir de ces derniers, une analyse de la discontinuité spatial sera réalisée à partir de la variable du revenu médian par quartier. Celle-ci a été choisit, car représente comme le taux de chômage ou le nombre de famille vivant sous le seuil de pauvreté un des principales déterminants de la théorie de la désorganisation.

L'idée est de comprendre si les ruptures entre quartiers par rapport à cette variable économique coïncident avec des délimitations construites par l'auto-corrélation spatiale. Pour ainsi dire, se demander si des quartiers proches ayant une intensité forte du crime sont aussi semblables aux vues du revenu médian des ménages.

Concernant la méthode: la réalisation de l'indice s'est construite en associant les points de crimes à chaque quartier. Au sein d'une nouvelle couche créée, une colonne se nommant « Count of point » fait la somme des vols à mains armés dans chacune des limites administratives de San-Francisco (quartiers). Cette dernière sera par la suite utilisée en entrée pour la construction de l'indice local de Moran via l'outil "analyse de grappes" décrite précédemment. Il est nécessaire de préciser qu'un choix a été fait quant à la sélection du voisinage pour le calcul de l'indice de Moran. En l'occurrence, c'est l'ordre contiguïté de niveau 1 qui forme les interactions spatiales. Cela veut dire que l'on associe les éléments qui ont un nombre minimal de frontières à franchir: deux éléments voisins prennent la valeur de 1 alors que deux éléments non contigus ont une valeur nulle.

Afin de représenter les discontinuités spatiales du revenu médian par foyer, l'idée est de faire varier l'épaisseur des contours de quartiers en fonction des valeurs discrètes du revenu. Ainsi plus l'épaisseur est importante, plus le quartier à un revenu médian élevé.



L'analyse du semis de point nous a indiqué une concentration des vols à main armée et la mesure de l'autocorrélation spatial dans cette seconde carte nous informe que celle-ci s'opère dans le Nord Est de la ville de San Francisco. Des quartiers contigus comme « Mission » et « South of Market » en forme les exemples. L'autocorrélation y est positive, le voisinage se ressemble par des valeurs fortes. En outre, on remarque que les limites administratives y sont fines puisqu'il n'y a pas une grande variation du revenu médian des ménages. Ces quartiers en rouge sont donc similaires du point de vue de l'intensité des crimes et de ce paramètre économique. En bleu, au contraire sont représentés les quartiers qui ne connaissent que peu de vols à main armée au cours de l'année 2011. L'autocorrélation y est négative, étant donné que les voisins ne se ressemblent pas. Il est à noté que cette opposition que l'on constate entre « South Beach » et « South of Market » se constate par le revenu médian des ménages. En effet, une différence de plus de 100000 dollars/an marque une rupture franche entre ces deux quartiers cités. Elle est symbolisée ici par une forte épaisseur du contour.

Dans une situation inverse, des quartiers comme « Laguna Honda » ou « Glen Park », situés plus au sud dans la ville de San Francisco se distinguent par une faible intensité du crime. Faisant partie d'un îlot de couleur verte, ces quartiers sont contigus et semblables et ont ainsi une autocorrélation positive. Ces entités administratives ont une quantité élevée de revenus médians par ménages, allant de 154000 dollars/an à 217000 dollars/an, ce qui les rend similaires aux quartiers en bleu situés le long de la côte (South Beach). L'hypothèse 1 est donc validée car les ressemblances et dissemblances marquées par la mesure de l'indice local de Moran se voit être identique avec les discontinuités observées du revenu médian par foyer. Les quartiers comme Mission ou South of Market, plus pauvres, sont aussi les quartiers où il y a le plus de vols à main armée.

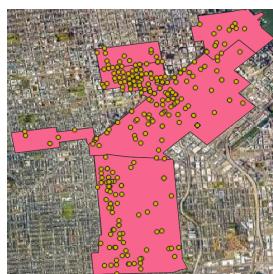
Mesure des variations locales de la criminalité

L'analyse de l'autocorrélation ne nous permet pas de visualiser les variations spatiales des vols à mains armés au sein même des quartiers grandement affectés par ces crimes. C'est donc ce que nous allons essayer d'analyser dans cette partie. L'acte de violence n'est pas borné par une frontière administrative comme pourrait montrer la carte précédente.

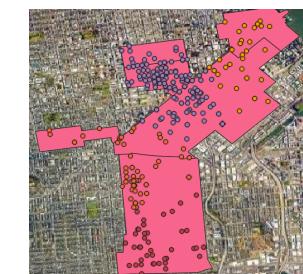
Concernant la démarche, il sera réalisé premièrement une interpolation spatiale à partir des quartiers High-High afin de démontrer qu'il y a variation de l'intensité du crime dans ces espaces. Puis pour chaque champ continu, nous examinerons le type de bâtiment. D'après l'hypothèse 2 énoncée en introduction, il sera intéressant de quantifier le nombre de bâtiments résidentiels par rapport aux nombres de bâtiments ayant de multiples fonctions. C'est-à-dire utilisés à la fois par les commerces et/ou par les bureaux et/ou par les services et/ou par l'éducation et/ou par la santé et par les logements.

En voici la méthode:

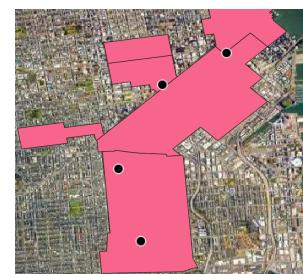
Premièrement, il faut extraire les points de crime dans les quartiers « High-High » (figure 3). Puis 4 différents clusters regrouperont l'ensemble des point de ce semi. Un partitionnement grâce à K-means nous permet de les mettre en valeur autour de ces 4 optimums locaux (figure 4). Enfin on créer un centre de gravité par clusters. On y associe ensuite dans un nouveau champ, la somme des crimes par clusters (figure 5).



Fiaure:3



Fiaure: 4



Fiaure: 5

Deuxièmement, on réalise une interpolation IDW à partir de ces 4 centres de gravité (figure 6). Cela nous permet de générer des champs continus à partir de données discrètes. L'outil demande une transformation des vecteurs en raster. Le résultat forme différentes classes de valeurs qui représentent l'intensité du crime (figure 7). Dans une démarche d'estimer le nombre de bâtiment par champs, il nous faut polygoniser ce travail. (Raster vers vecteur). Grâce à la calculatrice de champs du mode vecteur (figure), il est possible d'extraire chaque classe créée par l'interpolation (figure). Ainsi, on retrouve 5 catégories représentées par 5 couches vecteur différentes: (figure 8)

Couche 1: moins de 55 crimes | Couche 2: entre 55 et 71 crimes | Couche 3: entre 71 et 92 crimes | Couche 4: entre 92 et 138 crimes | Couche 5: plus de 138 crimes



Figure: 6



Figure:7



Figure: 8

Troisièmement, il faut extraire le bâti pour chacune de ces 5 couches (figure 9). Grâce à l'extension GroupStat sur QGIS, il est possible de faire la somme d'un type de bâtiment par champ continu. En l'occurrence, pour chaque classe de valeur, deux champs ont été créés, un pour représenter la somme des bâtiments de type résidentiels et un autre pour la somme des bâtiments de type usages multiples (logement+établissements scolaires+commerces+services...). Ce résultat est illustré dans la carte finale par un histogramme.

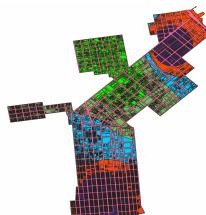
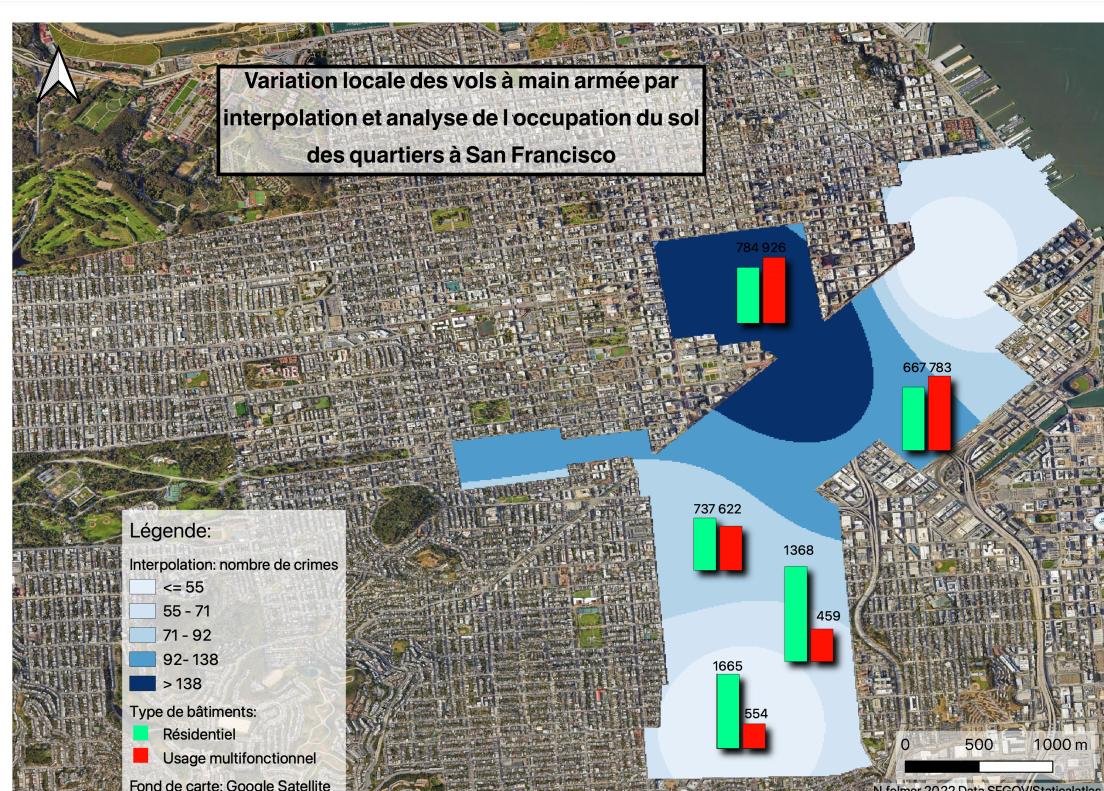


Figure: 9

L'interpolation spatiale nous permet ainsi de comprendre que les vols à main armée se concentrent dans une partie du quartier de « Tenderloin » ici représenté en couleur bleu foncé. En s'éloignant de cette zone où il y a plus de 138 crimes par an, on remarque une baisse progressive de l'intensité du crime. Par exemple, au sud du quartier « Mission », il n'y a plus « que » 55 vols avec main armée, ce qui est 2 fois moins élevé qu'à Tenderloin. En analysant l'occupation du sol. En voie que plus il y a de crimes, plus l'usage des bâtiments dans les quartiers varie. En effet, on constate cette dynamique dans les zones de 92 à 138 crimes par an et les zones de plus de 138 crimes par an. À Tenderloin, il y a 926 bâtiments multi-fonctionnels pour 764 bâtiments de type uniquement résidentiels. À l'inverse, moins il y a de mixité des usages, moins il y a de crimes. Ce phénomène s'observe lorsque l'on se dirige vers le sud de Mission. Il y a 1665 bâtiments résidentiels pour seulement 554 bâtiments multifonctions. Ce qui est 3 fois moins. L'hypothèse 2 est donc vérifié grâce à cette carte. Il aurait été intéressant de voir si cette dynamique s'observe aussi dans les quartiers low-low. C'est-à-dire les quartiers contigus dans lesquels on dénombre très peu de crimes.



Conclusion:

1. Concentration des vols à main armée: La méthode des quadrats complété par la méthode du plus proche voisin nous a permis de vérifier que le semis de point à une forme concentrée
2. Définition par autocorrélation des quartiers voisins marqués par une forte criminalité. La concentration des crimes s'opèrent dans les quartiers High-High tels que Mission ou South of Market. Les classifications réalisées par l'autocorrélation spatiale coïncide avec les fractures socio-économiques des quartiers selon l'analyse du revenu médian par foyer. Si l'on considère ce déterminant comme un facteur principale dans la théorie de la désorganisation alors l'hypothèse 1 est validée. Les caractéristiques sociales et économiques des quartiers peuvent influencer l'individu dans une probabilité que celui-ci commette un crime. Dans ce cas là, moins le quartier à un revenu médian par foyer élevé, plus le nombre de vols à main armé aura tendance à augmenter.
3. Le crime ne s'arrête pas aux frontières des quartiers et on observe une variation du nombre de vols à main armée à l'intérieur même de ces entités administratives. Ce constat a été possible grâce à l'interpolation spatiale. De plus, l'usage des bâtiments influence le nombre de crimes commis. Plus l'usage est mixé, plus il y a de vols à main armée. Le phénomène inverse se produit dans les zones plus résidentielles. L'hypothèse 2 est validée.

LIMITES:

Les discontinuités spatiales observées à travers le revenu médian ne suffisent peut être pas pour affirmer une corrélation simple des inégalités de revenu avec la classification réalisée par l'indice local de Moran. Il faudait mesurer les ruptures spatiales aussi à partir de paramètres tel que le taux de chômage, où le nombre de familles vivant sous le seuil de pauvreté par quartier pour affiner les résultats de l'hypothèse 1. Du point de vue de la technicité, il est important de préciser que les résultats peuvent évoluer en fonction du type d'interpolation choisi (IDW ou TIN), du type de relations de voisinage dans l'autocorrélation (par la distance, l'ordre de contiguïté, les voisins proches). Ainsi les résultats produits ne sont que le fruit d'un parti pris.

Bibliographie

<https://www.thpanorama.com/blog/cultura-general/desorganizacin-social-teora-formas-y-ejemplos.html>

<https://tel.archives-ouvertes.fr/tel-01955264/document>

<http://julienas.ipt.univ-paris8.fr/vgodard/pub/geomarke/memoires/07-08/boudiaf.pdf>

Pour l'autocorrélation spatiale :

<https://www.erudit.org/en/journals/crimino/1994-v27-n1-crimino937/017345ar.pdf>

<https://www.cairn.info/revue-espace-geographique-2010-1-page-51.htm>

<https://fr.readkong.com/page/la-concentration-spatiale-relative-de-la-criminalite-et-son-1847382>

<https://oasis.irstea.fr/wp-content/uploads/2013/10/10-Autocorrelation.pdf>

DONNÉES:

<https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-Historical-2003/tmnfyvry> / Data San-Franciso

Source: <https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-2018-to-Present/wg3w-h783>

Occupation du sol:

<https://data.sfgov.org/browse?tags=shapefile>

<https://data.sfgov.org/Geographic-Locations-and-Boundaries/SF-Find-Neighborhoods/pty2-tcw4>

Household income: <https://statisticalatlas.com/place/California/San-Francisco/Household-Income#data-map/neighborhood>