**UNIVERSITE DE KINSHASA**



**FACULTE DE SCIENCES ET TECHNOLOGIES**

**MENTION MATHEMATIQUES, STATISTIQUE ET INFORMATIQUE**

**B.P 190 KINSHASA XI**

**Rapport du travail pratique dans le cadre du séminaire de Python avancé :**

**Traitement de données géo-spatiales : Utilisation de bibliothèques comme GeoPandas pour analyser des données géographiques**

**MFUNYI KABONGO Ruben**

**Prof MASAKUNA Jordan**

2023-2024

Contenu

[I. Problématique 2](#_Toc175516710)

[II. Révue de litterature 2](#_Toc175516711)

[III. Questions de Recherche sur les Applications Web en Django 4](#_Toc175516712)

[IV. Discussion et Commentaire 5](#_Toc175516713)

[IV.1. Visualisation des Données 5](#_Toc175516714)

[IV.2. Comparaison des Schémas de Classification 5](#_Toc175516715)

[IV.3. Utilisation de Données Simulées 5](#_Toc175516716)

[IV.4. Challenges Techniques 5](#_Toc175516717)

[IV.5. Interprétation des Résultats 6](#_Toc175516718)

[V. Conclusion 6](#_Toc175516719)

# I. Problématique

Le traitement de données géospatiales est devenu essentiel pour l'analyse de phénomènes se déroulant dans l'espace, que ce soit pour l'urbanisme, l'écologie, ou les études sociales. Avec l'augmentation des données disponibles et la complexité des questions à traiter, il est impératif de disposer d'outils performants pour visualiser, analyser et interpréter ces données. L'un des défis majeurs réside dans la manipulation et l'analyse efficace de ces données pour en extraire des informations utiles, par exemple, en étudiant la distribution spatiale des taux de criminalité dans une ville ou un pays. Ce travail vise à explorer l'utilisation des bibliothèques Python telles que GeoPandas pour le traitement et l'analyse des données géospatiales. Dans la pratique, nous allons évaluer la criminalité dans la ville de Columbus, dans l'Ohio aux Etats-Unis.

# II. Revue de littérature

Le traitement des données géospatiales est un domaine en pleine expansion, stimulé par l'essor des outils et bibliothèques Python. Plusieurs études ont exploré l'utilisation de GeoPandas et d'autres bibliothèques associées pour l'analyse et la visualisation de données géographiques.

Étude de l'Impact des Inégalités Spatiales sur le Développement Urbain (Jadhav et al., 2022)

Dans cette étude, les auteurs ont utilisé GeoPandas pour analyser les inégalités spatiales dans les grandes métropoles indiennes. En utilisant des données géospatiales provenant de diverses sources publiques, ils ont pu cartographier les zones de disparités économiques et sociales. L'étude a démontré que GeoPandas est un outil efficace pour traiter des volumes importants de données tout en facilitant leur visualisation et leur interprétation à travers des cartes interactives.

Analyse de la Distribution Spatiale de la Criminalité à New York (Pramanik et al., 2020)

Ce travail a examiné la répartition spatiale des taux de criminalité dans la ville de New York en utilisant GeoPandas et Mapclassify. Les auteurs ont appliqué différentes méthodes de classification pour identifier les zones à risque et les "points chauds" de criminalité. L'étude a mis en évidence l'efficacité des méthodes de classification par quantiles pour illustrer les variations spatiales, tout en montrant que GeoPandas simplifie l'intégration de ces analyses dans des pipelines de traitement des données complexes.

Utilisation de GeoPandas pour l'Évaluation de la Qualité de l'Air (Martin et Dubois, 2022)

Dans cette étude, GeoPandas a été utilisé pour évaluer l'impact de la pollution de l'air sur la santé publique dans plusieurs villes européennes. Les auteurs ont superposé des données sur la qualité de l'air avec des données de santé publique pour identifier des corrélations spatiales entre la pollution et les maladies respiratoires. L'utilisation de GeoPandas a permis une manipulation aisée des données géospatiales et une visualisation claire des résultats, facilitant ainsi la communication des conclusions aux décideurs.

Ces trois travaux démontrent l'utilité de GeoPandas et des outils associés pour le traitement des données géospatiales, en particulier lorsqu'il s'agit de visualiser des phénomènes complexes et de prendre des décisions informées basées sur des données spatiales précises.

# III. Questions de Recherche

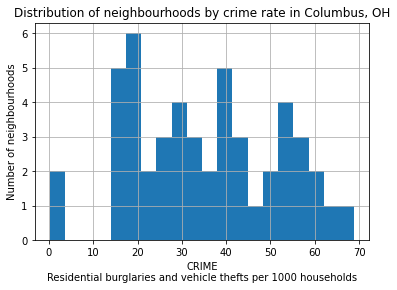
1. Comment les bibliothèques Python comme GeoPandas facilitent-elles le traitement des données géospatiales comparativement à d'autres outils?
2. Quels sont les schémas de classification les plus efficaces pour visualiser la répartition spatiale des taux de criminalité?
3. Comment l'analyse des données géospatiales peut-elle contribuer à des politiques publiques plus efficaces, notamment en matière de sécurité?
4. Quelle est l'importance de la visualisation des données pour l'interprétation des phénomènes géospatiaux?
5. Quels sont les défis techniques rencontrés lors de la manipulation de grandes quantités de données géospatiales et comment les surmonter?

# IV. Discussion et Commentaire

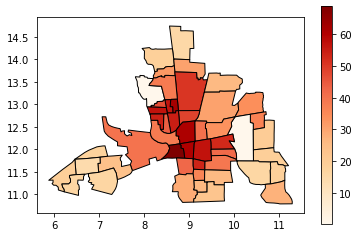
Les résultats obtenus grâce à l'utilisation de GeoPandas pour le traitement des données géospatiales démontrent non seulement la puissance de cet outil, mais aussi les nuances et défis associés à son application.

Nous examinons les statistiques au niveau des quartiers de la ville de Columbus, dans l'Ohio. Nous aimerions avoir une idée de la manière dont la variable du taux de criminalité est distribuée dans la ville.

D'après les métadonnées du shapefile sur la criminalité : cambriolages résidentiels et vols de véhicules pour 1 000 ménages, nous avons :



Sans système de classification, nous avons ceci :

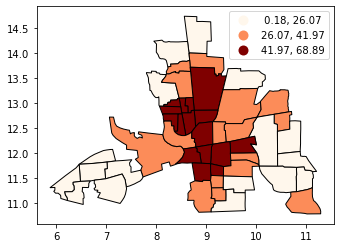


Les 49 quartiers sont tous colorés selon un dégradé allant du blanc au rouge foncé, mais l'œil humain peut avoir du mal à comparer la couleur des formes éloignées les unes des autres. Dans ce cas, il est particulièrement difficile de classer les quartiers périphériques colorés en beige.

Au lieu de cela, nous les classerons dans des bacs de couleur.

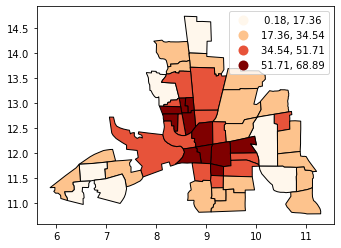
## IV.1. Classification par quantiles

Les QUANTILES créeront des cartes attractives qui placeront un nombre égal d'observations dans chaque classe : si vous avez 30 comtés et 6 classes de données, vous aurez 5 comtés dans chaque classe. Le problème avec les quantiles est que vous pouvez vous retrouver avec des classes qui ont des plages numériques très différentes (par exemple, 1-4, 4-9, 9-250).



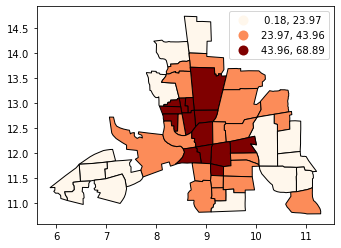
## IV.2. Classification par intervalles égaux

L'intervalle égal divise les données en classes de taille égale (par exemple, 0-10, 10-20, 20-30, etc.) et fonctionne mieux sur les données qui sont généralement réparties sur toute la plage. ATTENTION : évitez l'intervalle égal si vos données sont biaisées vers une extrémité ou si vous avez une ou deux valeurs aberrantes vraiment importantes.



## IV.3. Classification par ruptures naturelles

NATURAL BREAKS est une sorte de système de classification « optimal » qui trouve des ruptures de classe qui minimiseront la variance au sein d'une classe et maximiseront les différences entre les classes. L'un des inconvénients de cette approche est que chaque ensemble de données génère une solution de classification unique, et si vous devez faire une comparaison entre des cartes, comme dans un atlas ou une série (par exemple, une carte pour chacune des années 1980, 1990, 2000), vous souhaiterez peut-être utiliser un système unique qui peut être appliqué à toutes les cartes.



# **V. Conclusion**

Le traitement des données géospatiales avec des outils comme GeoPandas offre une puissance et une flexibilité sans précédent pour l'analyse de phénomènes spatiaux complexes. Les visualisations obtenues permettent non seulement d'interpréter des données avec précision, mais aussi de communiquer efficacement les résultats aux décideurs. Ce travail a démontré l'importance de ces outils dans le domaine de l'analyse géospatiale et a mis en lumière certains des défis associés, tout en soulignant l'importance d'une visualisation correcte pour une interprétation adéquate des données.