

Introduction aux données spatiales avec Geopandas

Lino Galiana
2024-05-27

1. Introduction

1.1 Quelle différence avec des données traditionnelles ?

Les chapitres précédents ont permis de découvrir la manière dont des données structurées peuvent être valorisées grâce à la librairie **Pandas**. Nous allons maintenant découvrir l'analyse de données plus complexes, à savoir les données spatiales. Ces dernières sont une sophistication des données tabulaires puisqu'en plus de partager les propriétés de celles-ci (données aplaties dans une structure de colonnes et de lignes), elles comportent une dimension géographique supplémentaire. Celle-ci est plus ou moins complexe selon la nature des données: cela peut être des points (coordonnées de localisation en deux dimensions), des lignes (une suite de points), des lignes directionnelles (la même structure précédemment mais avec une direction), des polygones (un ensemble de points)... Cette diversité des objets géographiques vise à permettre des systèmes d'information et de représentation de nombreux objets géographiques.

Par la suite, nous entendrons par "*données spatiales*" l'ensemble des données qui portent sur les caractéristiques géographiques des objets (localisation, contours, liens). Les caractéristiques géographiques des objets sont décrites à l'aide d'un **système de coordonnées**. Celles-ci permettent de représenter l'objet géographique dans un espace euclidien à deux dimensions (x, y) . Le passage de l'espace réel (la Terre, qui est une sphère en trois dimensions) à l'espace plan se fait grâce à un **système de projection**.

1.2 Structure des données spatiales

Les données spatiales rassemblent classiquement deux types de données :

1. des **données géographiques** (ou géométries) : objets géométriques tels que des points, des vecteurs, des polygones, ou des maillages (*raster*). Exemple: la forme de chaque commune, les coordonnées d'un bâtiment;
2. des **données attributaires** (ou attributs) : des mesures et des caractéristiques associées aux objets géométriques. Exemple: la population de chaque commune, le nombre de fenêtres et le nombre d'étages d'un bâtiment.

Les données spatiales sont fréquemment traitées à l'aide d'un système d'information géographique (SIG), c'est-à-dire un système d'information capable de stocker, d'organiser et de présenter des données alphanumériques spatialement référencées