**L’épidémie de choléra à Londres septembre 1854 : l’apport de l’analyse spatiale**

**Sommaire**

[Objectif 1](#_Toc211441764)

[Données à utiliser 1](#_Toc211441765)

[1. Localiser et décrire pour poser une hypothèse : Cartographier des morts de choléra en septembre 1854. Réponse à la question « Où ? » 1](#_Toc211441766)

[2. Cartographier des morts de choléra en septembre 1854 : réponse à la question « Où et comment ? » 2](#_Toc211441767)

[3. Repérer la structure spatiale des morts du choléra : réponse à la question « Comment et pourquoi ? » 4](#_Toc211441768)

[4. Expliquer et fournir une aide à la décision 7](#_Toc211441773)

# Objectif

Mobiliser des outils d’analyse spatiale pour comprendre la diffusion d’une épidémie. Il s’agit de suivre la même démarche que le médecin John Snow a entreprise il y a 160 ans pour comprendre les causes de la diffusion du choléra à Londres, mais avec l'aide de moyens et méthodes modernes. On considère souvent l’analyse de Dr Snow comme la première véritable analyse spatiale.

*Inspiré de l’analyse de Dr Snow https://www.gislounge.com/john-snows-cholera-map-gis-data/*

*A voir également : http://www.theguardian.com/news/datablog/2013/mar/15/john-snow-cholera-map*

# Données à utiliser

1. Télécharger les données à utiliser à partir du moodle du cours. Le dossier Data\_Origine contient :

* Deaths\_Pumps.xlsx : la localisation des morts de choléra pendant l’épidémie de septembre 1854 et la localisation des pompes à eau dans le quartier de Soho en 1854
* Snow\_1854.tif – raster géoréférencé à partir de la carte de Soho fait par John Snow en 1854
* ReseauRuesSoho.shp – réseau de rue digitalisé à partir de la carte de Soho fait par John Snow

1. Dans le dossier DrSnow\_cholera\_Geonum créez un Geopackage que vous nommez DrSnow. Vous enregistrerez l’ensemble des fichiers produites pendant la séance dans le DrSnow.gpkg.

# Localiser et décrire pour poser une hypothèse : Cartographier des morts de choléra en septembre 1854. Réponse à la question « Où ? »

*Au milieu du 19ème siècle, il y a eu une grande épidémie de choléra en Europe. Le choléra est une maladie infectieuse, due à une bactérie, dont on n'a pas su pendant longtemps comment elle se transmettait. Lors d'une épidémie de choléra dans le centre de Londres, le médecin John Snow a cherché la source de cette bactérie (l’épidémie est partie du quartier de Soho où habitait Snow). Pour pouvoir faire des hypothèses sur la source de l’épidémie Dr Snow a dressé une carte représentant la répartition des morts dans le quartier de Soho qui était le plus atteint par l’épidémie.*

Vous allez commencer par cartographier des données décrivant le nombre de mort à chaque immeuble (adresse) du quartier de Soho.

1. Ouvrez le fichier *Deaths\_Pompes.xlsx,* puis préparez les données pour pouvoir les cartographier*.*
2. Quelle variable visuelle utiliserez-vous pour cartographier la répartition du nombre de morts à chaque adresse ? Pourquoi ?

1. Cartographiez la répartition du nombre de morts à chaque adresse (si vous avez des difficultés, et **seulement** à ce moment, consultez le guide)
2. Combien de personnes sont mortes victimes de l'épidémie de choléra ? Quel outil utilisez-vous pour répondre à cette question ? Dans le tableau résultats, à quoi correspond Compte ?

# Cartographier des morts de choléra en septembre 1854 : réponse à la question « Où et comment ? »

Outre la cartographie thématique, il existe différentes méthodes qui permettent de visualiser les zones de concentration d’un phénomène pour lequel on dispose de données vectorielles (points, polygones ou lignes).

On peut notamment représenter la répartition des objets (ici les cas de choléra) sous forme d’une *surface continue* de données lissées. L’intérêt majeur de ce type de traitement spatial des données est de faciliter la lecture d’une distribution spatiale en mettant mieux en évidence les zones à forte concentration.

**Encadré 1 : Méthode de lissage**

Dans QGIS il existe différents outils ou extensions qui permettent d’effectuer un lissage.

L’outil *Carte de chaleur (Estimation par noyau)* (Kernel Density en anglais) permet d’effectuer ce lissage. Vous le trouverez dans la boîte à outils interpolation[[1]](#footnote-1).

Le lissage renforce la visibilité des zones de concentration et élimine des épiphénomènes. Il permet donc de souligner la structure spatiale du phénomène étudié.

Avec l’outil *Carte de chaleur* dans QGIS les valeurs affichées sur la carte résultat (en mode raster) sont calculées :

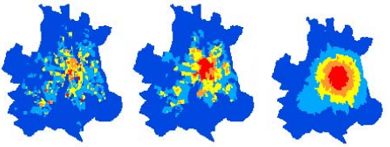
- soit en fonction du nombre de points (par exemple, le nombre de fermes touchées par la grippe aviaire) : si une ferme est entourée de nombreuses fermes touchées par la grippe aviaire, la cellule correspondant à cette ferme dans le raster résultat verra sa valeur renforcée. A l'inverse, une ferme touchée par la grippe aviaire isolée dans l’espace géographique (non entourée d’autres fermes touchées par la grippe aviaire) n’apparaîtra pas comme une ferme touchée par la grippe aviaire (la cellule du raster aura une valeur nulle ou très faible).

- soit en fonction d’un champ de la table attributaire (par exemple, le nombre de volailles touchées par la grippe aviaire dans une ferme). Dans ce cas, si une ferme avec un grand nombre de volailles touchées par la grippe aviaire est entourée de fermes également très fortement atteintes par le la grippe aviaire (nombre de volailles touchées est élevé), la cellule correspondant à cette ferme dans le raster, ainsi que les cellules voisines, auront une valeur renforcée.

Il est nécessaire de définir un rayon de lissage qui délimite le périmètre à l’intérieur duquel les entités sont utilisées pour calculer la valeur affectée à chaque cellule. Le choix du rayon influence le niveau de généralisation du résultat, plus le rayon de lissage est important, plus le lissage est important (exemple cartes ci-dessous) :

**Lissage de la population à Toulouse en 1999 avec 3 rayons de lissage différents**

50 m 150 m 1000 m



Source : Certu, 2005

1. A votre avis dans le cas des morts de choléra est-il plus intéressant de réaliser le lissage à partir des entités ou à partir d’un des champs, précisez le cas échéant lequel ? Expliquez pourquoi.

1. Réalisez la carte lissée de la répartition des morts dans le quartier de Soho en choisissant un rayon de 50 m. Choisissez. Choisissez une résolution du résultat plus grande que celle proposée défaut
   1. Quel est le mode de représentation des données produites ?
   2. A quoi correspondent les valeurs des cellules à votre avis ?

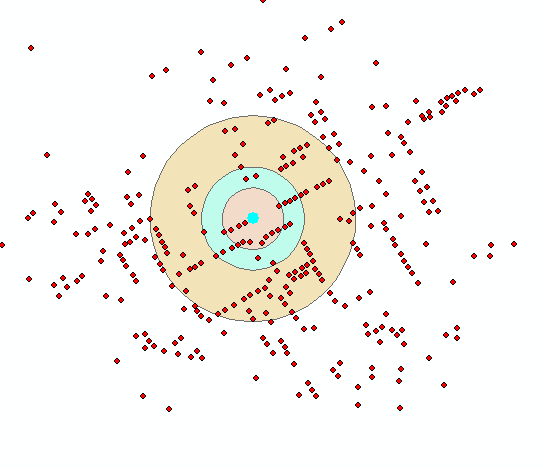
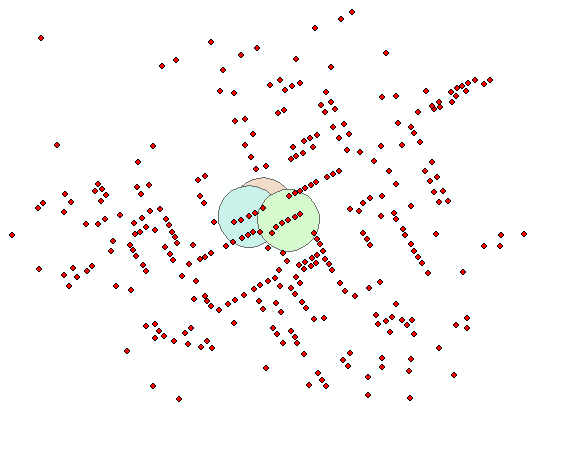
* 1. Quelle distance est utilisée pour caractériser ce rayon de recherche et à quelle famille de distance appartient-elle ?

**Explication cartographique des rayons de différentes distances :**

Rayon de **100 m**

**98** points voisins sont utilisés dans le calcul du nombre de morts lissé

Rayons de 30 m pour trois points. En réalité le calcul du nombre de morts lissé (donc la moyenne) est effectué pour chaque cellule dans le raster en sortie, à partir de ces voisins (déterminé en fonction du rayon de lissage choisi (cf figure à gauche)

Rayon de **50 m**

**25** points voisins sont utilisés dans le calcul du nombre de morts lissé

Rayon de **30 m**

**11** points voisins sont utilisés dans le calcul du nombre de morts lissé

Point de référence dans l’exemple

1. Améliorez le rendu de la carte lissée (couleurs, transparence, lisibilité) – à superposer à la carte de Dr Snow
2. Refaites l’analyse en faisant varier le rayon de recherche à 70 m, puis à 100 m. Qu’observez-vous ?

1. Proposez une rédaction de légende qui aide le lecteur à facilement comprendre la carte et modifiez la légende (remplacez les chiffres).

1. Commentez la répartition spatiale des victimes de l’épidémie dans le quartier de Soho.

.

1. Au milieu du 19e siècle la théorie communément acceptée était la transmission du choléra par l’air, pourtant d’après l’observation de la répartition des morts John Snow a déduit que l'air n’était pas en cause. Arrivez-vous à la même conclusion ? Argumentez.

Aujourd’hui on sait que le vecteur de transmission du choléra est l’eau (et le contact physique entre personnes), mais c’est la carte de la répartition des morts de choléra qui a donnée l’idée à Snow de vérifier l’hypothèse de l’eau comme vecteur de transmission.

*Dans cette première étape vous avez déterminé la localisation des objets étudiés (les morts) et vous vous êtes posé la question de la répartition géographique du phénomène. En effet, la première question que Dr Snow s’est posée était « où ? ». Il a répondu par l’étude de la carte des victimes de choléra. Puis il s’est posé une deuxième question, « où et comment ? ». J. Snow s’est ainsi intéressé à la localisation relative des victimes (les unes par rapport aux autres) et a cherché à déterminer comment les morts se sont répartis à l’intérieur du quartier (concentration).*

# Repérer la structure spatiale des morts du choléra : réponse à la question « Comment et pourquoi ? »

*Avec l’hypothèse de l’eau comme vecteur de transmission et dans l’objectif de comprendre la cause de l’épidémie, ainsi que les causes de sa diffusion par la suite, Dr Snow a localisé très précisément sur la carte précédente les* ***pompes à eau*** *dans le quartier de Soho.*

D’après l’observation de la carte, Dr Snow s’est demandé s’il n’y avait pas une concentration plus importante de cas de choléra autour de certaines pompes à eau. Il n’avait pas à l’époque les moyens pour vérifier de façon scientifique cette deuxième hypothèse. Aujourd’hui vous avez ces moyens et pour analyser la structure spatiale des victimes autour des différentes pompes à eau, vous allez utiliser différentes méthodes d’analyse spatiale pour vérifier si certaines pompes à eau pourraient être les sources de contamination.

## Identifier les pompes à eau les plus proches de chaque adresse de mort avec une distance mathématique

Réalisez un schéma de traitements de l’ensemble de la démarche de cette partie et insérez le dans le document.

1. Créez un jeu de données vecteur (dans le geopackage) des pompes à eau.
2. Vous allez commencer par utiliser la méthode des polygones de Thiessen. Quel est le principe de cette méthode ? Que définit-elle ? Sur quelle distance se base-t-elle ?

1. Quelle est l’hypothèse sous-jacente à cette méthode en ce qui concerne l’utilisation des pompes à eau par les habitants. Quelle est la principale limite de cette hypothèse ?

1. Créez des polygones de Thiessen en fonction des pompes. Testez une fois sans zone tampon, puis une deuxième fois avec une zone tampon de 10 %. Comparez les deux résultats. Selon vous, quel est le résultat le plus pertinent ? Pourquoi ?

1. Affichez sur chaque polygone la localisation des morts et les noms des pompes à eau. Que constatez-vous grâce aux polygones de Thiessen ?

1. Quel pourcentage des décès de choléra se sont produits dans chaque zone d’influence des pompes telles qu’elles ont été définies par les polygones de Thiessen ?

## Identifier les pompes à eau les plus proches de chaque adresse de mort en utilisant la distance par le réseau viaire

L’objectif est ici de déterminer quel est la pompe la plus proche de chaque lieu dans le quartier de Soho, en tenant compte d’un déplacement par les rues (= distance empirique) et non par un déplacement à vol d’oiseau (= distance euclidienne) comme dans l’étape précédente. Il s’agit donc à nouveau d‘identifier les zones de proximité (zones potentielles d’influence) de chacune des pompes, mais cette fois-ci en se fondant sur une distance empirique qui est la distance kilométrique sur le réseau de rues.

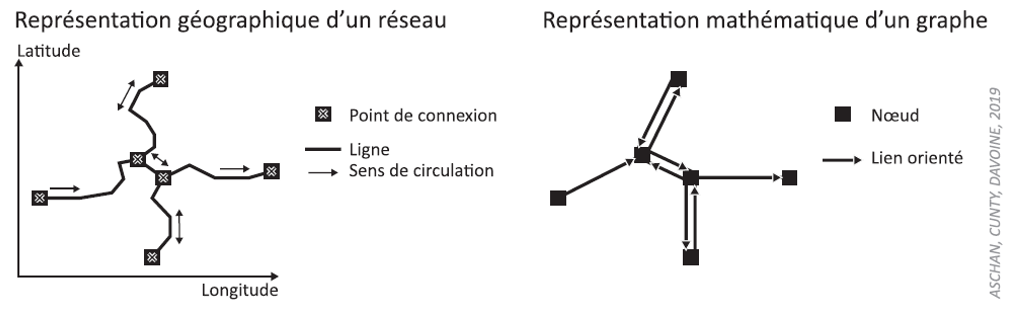
**Réalisez un schéma de traitements de l’ensemble de la démarche de cette partie et insérez le dans le document.**

### Préparer des jeux de données

1. Vérifiez que le jeu de données d’entités linéaires est structuré sous forme d’un réseau

Pour effectuer des analyses de réseau sur un jeu de données, il est important que les entités linéaires qui représentent les tronçons (liens) du réseau soient bien « coupés » à chaque point de connexion (nœuds) du réseau (figure 1). Vérifiez que c’est bien le cas pour ReseauRuesSoho.shp en sélectionnant quelques tronçons, un à un.

**Figure 1** : Représentation géographique et mathématique



***Commentaire figure 1 :*** *La structuration des données sous la forme d’un réseau implique de définir des relations topologiques qui lient les entités points (nœuds) et les entités lignes (liens) entre elles. On simplifie alors le réseau sous la forme d’un objet mathématique appelé graphe ou réseau logique. Les liens peuvent être orientés (figure 1), ce qui signifie qu’ils ont des nœuds de départ et d’arrivée selon un sens de circulation. Cette information est indispensable pour pouvoir ensuite calculer la distance sur le réseau.*

1. Pour pouvoir faire une analyse de l’accessibilité sur un réseau il est indispensable de connecter au réseau les lieux pour lesquels les distances seront calculées. Ces lieux sont appelés des sommets.
2. A quoi correspondent les sommets ici ?
3. Est-ce que les pompes sont connectés au réseau ? Expliquez pourquoi.

1. Dans QGIS, le module V.net (vecteur network de Grass) permet de relier les sommets au point le plus proche sur ce réseau. Ouvrez l’outil et sélectionnez les paramètres suivants :
   1. Opération à effectuer : *Connect*
   2. « Threashold for connecting centers to the network » : il s’agit ici de spécifier le seuil de distance maximal à utiliser pour connecter les pompes au tronçon du réseau de rue la plus proche. Quelle distance proposez-vous ici ?
   3. Les autres paramètres n’ont pas besoin d’être modifies.
   4. Enregistrez le fichier résultat.
2. Quel est le résultat graphique de cette opération ?

### Déterminer la pompe la plus proche de chaque tronçon de route et donc l’« aire d’attraction » de chaque pompe en fonction d’une distance empirique

1. Vous allez affecter chaque tronçon du réseau à la pompe la plus proche.
   1. Quel outil de « vecteur network de Grass » utiliserez-vous pour identifier pour chaque tronçon du réseau de rues quelle est la pompe la plus proche selon la distance calculée à partir du réseau ?
   2. Effectuez ce traitement en veillant à bien choisir les données en entrée.
   3. Observez le résultat (objets graphiques et table attributaire) et expliquez-le.

* 1. A quoi correspond le champ *cat* crée ?

* 1. Affichez les aires d’attraction des pompes sur la carte. Quelle variable visuelle choisissez-vous et pourquoi ?

* 1. Améliorer la lisibilité de la carte.

1. Quelle est l’hypothèse sous-jacente sur les pratiques de déplacement des habitants de Soho pour s’alimenter en eau ?

1. Comment pourrait-on affiner la modélisation des tronçons des rues pour ensuite pouvoir calculer une distance coût au lieu d’une simple distance kilométrique (ce n’est pas possible ici mais c’est une réflexion dans l’absolu) ?

1. Quel pourcentage des décès de choléra se sont produits dans la zone d’influence de la pompe de Broad Street (quand l’aire d’influence est identifiée par le chemin le plus court à chaque pompe par les rues) ?

1. Est-ce que les résultats obtenus par les polygones de Thiessen (distance euclidienne) et par la méthode prenant en compte la distance réelle par les rues entre le domicile et chaque pompe convergent ? Pour répondre à la question, comparez les deux cartes et les parts de morts dans les aires d’attraction identifiées avec les deux méthodes.

1. Ces résultats appuient-ils la supposition que le point de départ de l'épidémie de choléra est de l'eau contaminée ? Expliquez.

1. Que pouvez-vous en conclure ?

Dans cette étape vous avez comparé et analysé conjointement les répartitions de deux phénomènes dans l’objectif d’expliquer une localisation par d’autres localisations. Il s’agissait donc de déceler en quoi la localisation des pompes à eau a pu avoir une influence sur la diffusion de l’épidémie de choléra.

# Expliquer et fournir une aide à la décision

Affichez sur la carte le nom des compagnies de distribution des eaux pour chaque pompe ? Que remarquez-vous ?

***Explication****: A l’époque, la distribution de l’eau à Londres était faite par plusieurs compagnies. Il y avait à la fois des puits publics peu profonds où les habitants venaient chercher leur eau et des conduites qui puisaient l’eau de la Tamise. Le puits de Broad Street, exploité par « Southwark and Vauxhall Water Company » était un puits public où l’eau provenait de la Tamise à l’aval de Londres. En outre le puits de Broad Street était creusé à un mètre d’une fosse d’aisance (pas d’égout à l’époque). La fosse d’aisance était vraisemblablement infectée par le Vibrio cholerae.*

***Aide à la décision****: C'est le 7 septembre que, suite aux observations de Snow, le puit fut mis hors service (on a enlevé la poignée de la pompe !). Le 11 du même mois, l'épidémie disparaissait pratiquement. Malgré l'évidence, les autorités médicales ne voulurent pas croire que les latrines pouvaient être à l'origine de la contamination. Même si un fervent partisan du docteur Snow, le pasteur Whitehead, découvrit que certaines de celle-ci fuyaient réellement dans le puit qui alimentait le puit, ce n'est qu'en 1866, alors que le choléra frappa à nouveau Londres, que la théorie de Snow fut acceptée, quelques années après sa mort.*

# Mettez en page une carte au choix que vous avez créé pendant ce TD et insérez la carte ici.

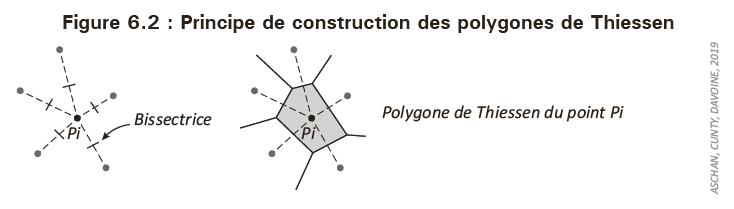
**Déposez sur moodle ce document de TD complété par les réponses aux questions et les deux schémas de traitements avant le 15/11/2025.**

**Annexe 1 :** Les polygones de Thiessen

(extrait de l’ouvrage Aschan C., Cunty C., Davoine PA., 2019, Les systèmes d’information géographique, Armand Colin, collection Cursus)

Une méthode simple pour définir les aires d’attraction respectives de chaque pompe, modélisées sous la forme d’un semis de points, est basée sur la construction de polygones de Thiessen qui utilisent la distance euclidienne.

Les polygones de Thiessen sont créés à partir d’un maillage de semis de points, en joignant les points les plus proches par des lignes pour effectuer une triangulation (appelée triangulation de Delaunay), puis en identifiant les bissectrices qui constituent les arcs des polygones (figure 6.2).



Les polygones de Thiessen divisent donc la surface couverte par le semis de points en polygones, chaque point se trouvant à l’intérieur d’un polygone et d’un seul, et chaque polygone contenant un point et un seul. Ainsi, tout emplacement dans un polygone de Thiessen est plus proche du point localisé dans ce polygone que de tout autre point du semis.

1. NB. Il ne s’agit pas d’une interpolation, mais d’un lissage des données (méthode différente de l’interpolation) [↑](#footnote-ref-1)