

Statistiques et Cartographie - TP1

Consignes : Le tableau de données et de travail `StatCarto_TP1.xlsx` est à télécharger sur Moodle ; vous ne serez pas notés sur le contenu final du fichier mais sur votre copie uniquement, sur laquelle il faudra bien noter tous les résultats demandés.

Exercice : Connaissance des Risques d’Inondation

Ce travail se base sur les données d’enquête que vous avez saisies lors du TD2. On propose de s’intéresser à la connaissance du risque d’inondation par les sondés. Sur 185 grilles saisies, 182 sont exploitables en terme de localisation précise de la commune des sondés.

Contexte (3pts) → *Discutez de problématiques possibles à partir des données que vous avez recueillies et saisies ; quels peuvent être les apports et limitations d’une approche quantitative, et comment peut-elle être complémentaire d’une approche qualitative ? Pensez vous que la distinction quantitatif-qualitatif soit pertinente en géographie contemporaine ?*

Données (2pts) → *Commentez la distribution spatiale des lieux de résidence des sondés (Document 1). La quantité de 182 réponses vous semble-t-elle suffisante pour tirer des conclusions statistiques robustes ? Pourquoi ?*

Statistiques générales (5pts) L’onglet `Donnees` donne l’ensemble des scores donnant le niveau de connaissance du phénomène de crue (nombre entier entre 1 et 5).

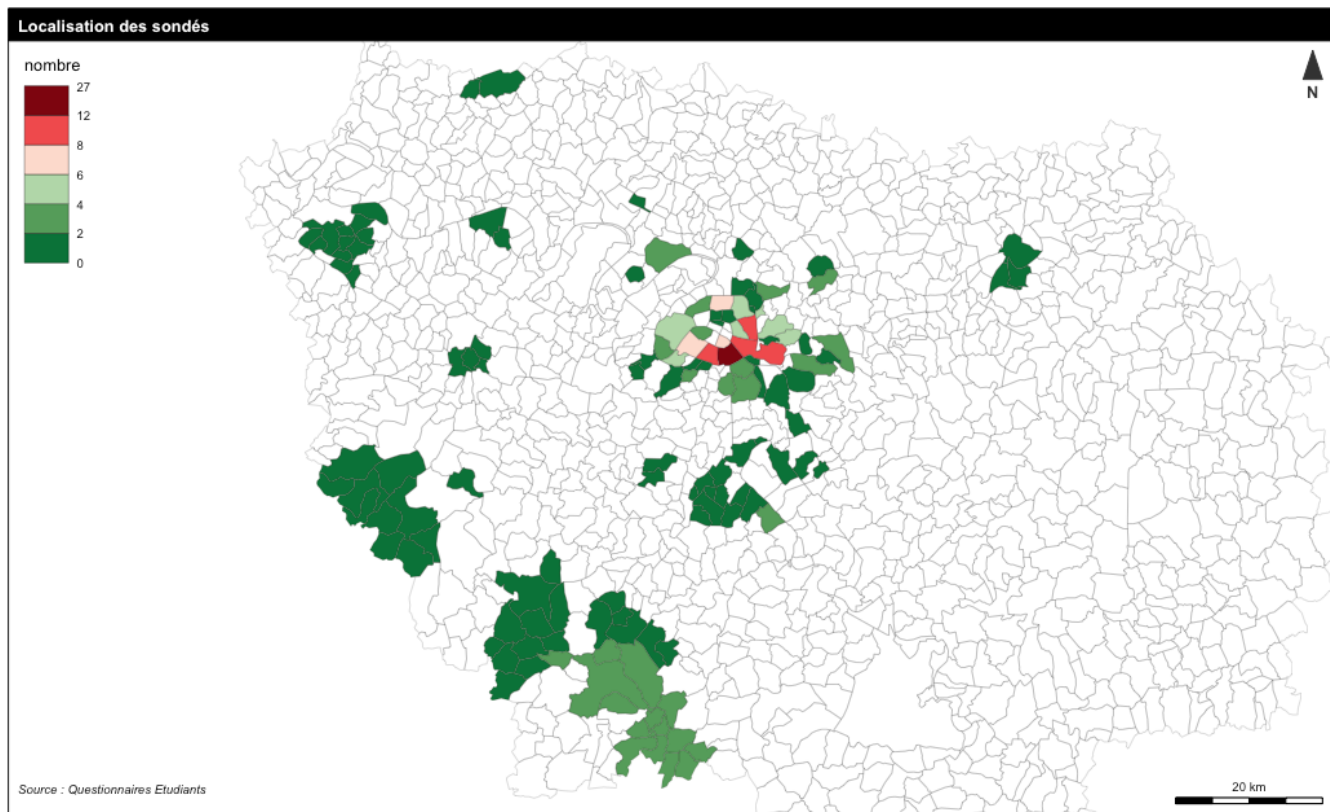
→ *Tracez le diagramme de distribution du niveau de connaissance (à reproduire sur votre copie) ; calculez moyenne et écart-type¹ de l’échantillon. Commentez la forme de la distribution et les valeurs obtenues.*

Rôle de la proximité et de l’accessibilité aux zones inondables (5pts) On cherche à savoir si la distance du domicile à une zone inondable influence la connaissance des risques, autrement dit si les populations les plus concernées sont aussi informées. On suppose que l’ignorance est distribuée uniformément dans les populations et donc dans l’espace, et on s’intéresse aux réponses satisfaisant un certain niveau de connaissance (au moins 3). On a calculé pour chaque commune la distance à une zone inondable, et classé les réponses dans deux classes pour simplifier (“proche”, si la distance est inférieure à 2km, “lointain” si elle est supérieure), dans l’onglet `Distance`².

→ *Calculez moyenne et écart-type pour chacune des classes (aide : utiliser la fonction “trier” pour ordonner par classe). Proposez une interprétation sur l’effet de la distance.*

¹la formule vue en cours est implémentée par la fonction `ECARTTYPE` d’Excel®, tandis que la fonction `ECARTTYPEP` implémente l’estimateur non-biaisé de la variance. Le biais intrinsèque à l’estimateur standard est dû à l’utilisation de la moyenne estimée au sein de l’estimateur de variance. Il tend vers 0 quand la taille de l’échantillon tend vers l’infini. Dans notre cas, il est de l’ordre relatif du millième, vous pouvez utiliser la fonction de votre choix sans impact significatif sur le résultat.

²Le document 3 cartographie les distances par commune.



Document 1: Nombre de réponses au sondage par communes (code postal).

La distance brute ne prend pas en compte l'intensité des inondations potentielles (voir surfaces inondables en document 2). Pour cela, on définit une mesure "d'accessibilité aux zones inondables"³ qu'il faut interpréter comme une quantité d'inondation potentielle dans un voisinage géographique de la commune. L'onglet *Accessibilite* donne ces valeurs et une classification simplifiée ("haute" et "basse").

→ De même, calculez moyenne et écart-type pour chaque classe. Interprétez et comparez aux résultats pour la distance.

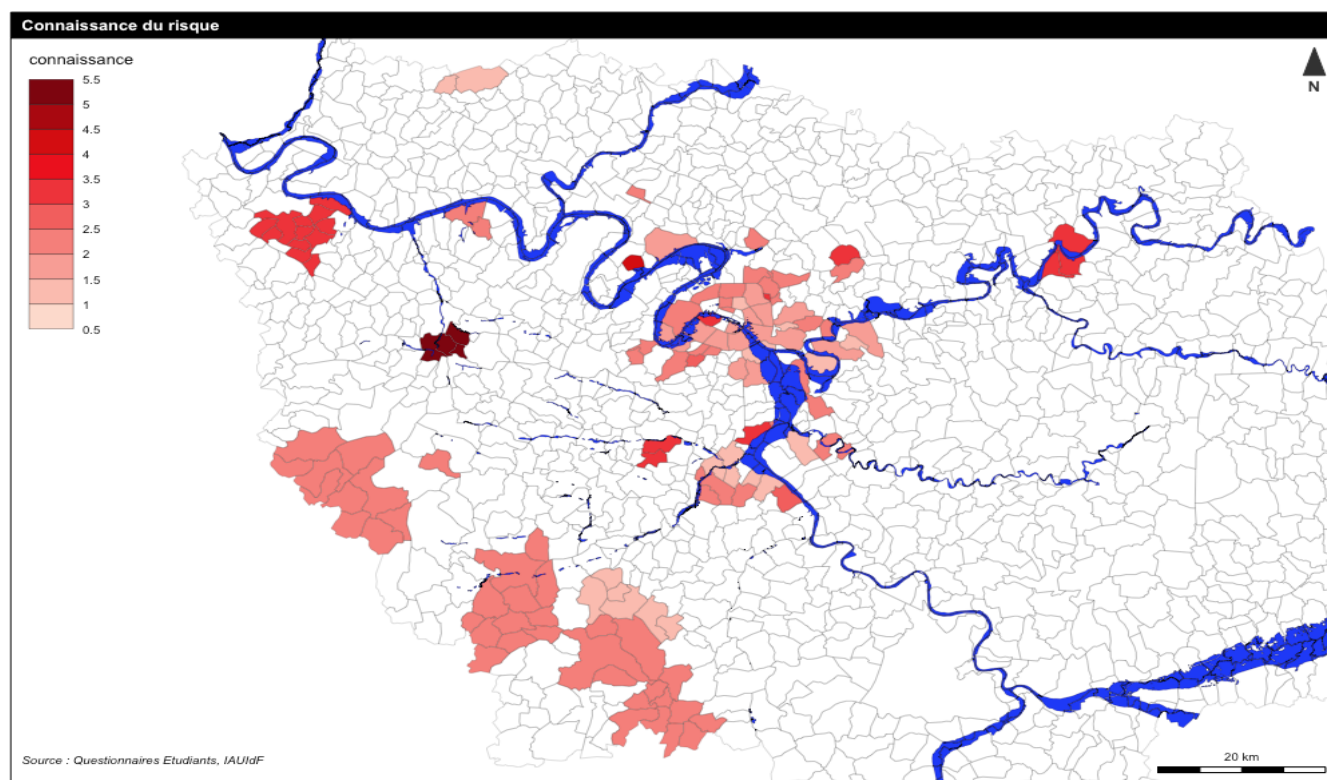
Point de vue cartographique (2pts) → Commentez la synthèse cartographique du niveau de connaissance en document 2 ; est-il possible de tirer de conclusions générales à partir de la carte uniquement ? Dans quelle mesure la cartographie et l'analyse statistique sont-elles complémentaires ?

Autres variables (2pts) → En correspondance à des problématiques proposées en première question, discutez d'autres variables ou aspects⁴ qui seraient pertinents à prendre en compte.

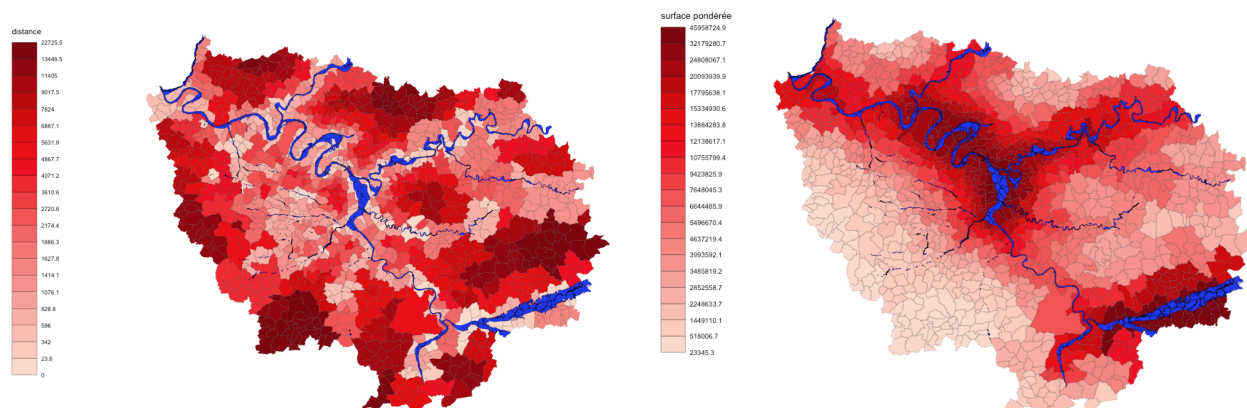
Données Synthétiques (1pts) → Discutez de l'intérêt d'extrapoler des données manquantes (communes sans répondant) ou de simuler des données synthétiques (réponses ayant la même structure statistique) à partir du jeu de données disponibles.

³qui correspond à la moyenne des surfaces des zones inondables pondérées par la distance à la commune, donnée pour la commune i , avec A_j surfaces des zones inondables, $Z_i = \sum_j A_j \cdot \exp(-d_{ij}/d_0)$ avec ici $d_0 = 5\text{km}$. Elle est également cartographiée en document 3.

⁴Qui peuvent être dans le questionnaire initial ou non ; souvent un premier questionnaire pilote est administré et le questionnaire est modifié selon les réponses obtenues, pour adapter ou préciser des points.



Document 2: Niveau moyen de connaissance du risque agrégé par code postal ; zones inondables (Plus Hautes Eaux Connues).



Document 3: (Gauche) Distance des communes aux zones inondables ; (Droite) Surface accessible de zone inondable.

Exercice Facultatif

Consigne : *Exercice à traiter si vous avez fini le premier largement en avance et facilement.*

La question de la “robustesse statistique” d’une estimation, par exemple dans notre cas d’une moyenne ou d’un écart-type, est essentielle pour la validation des interprétations qu’on peut en tirer. On se propose d’illustrer cette notion dans un cas simple, sur un jeu de données