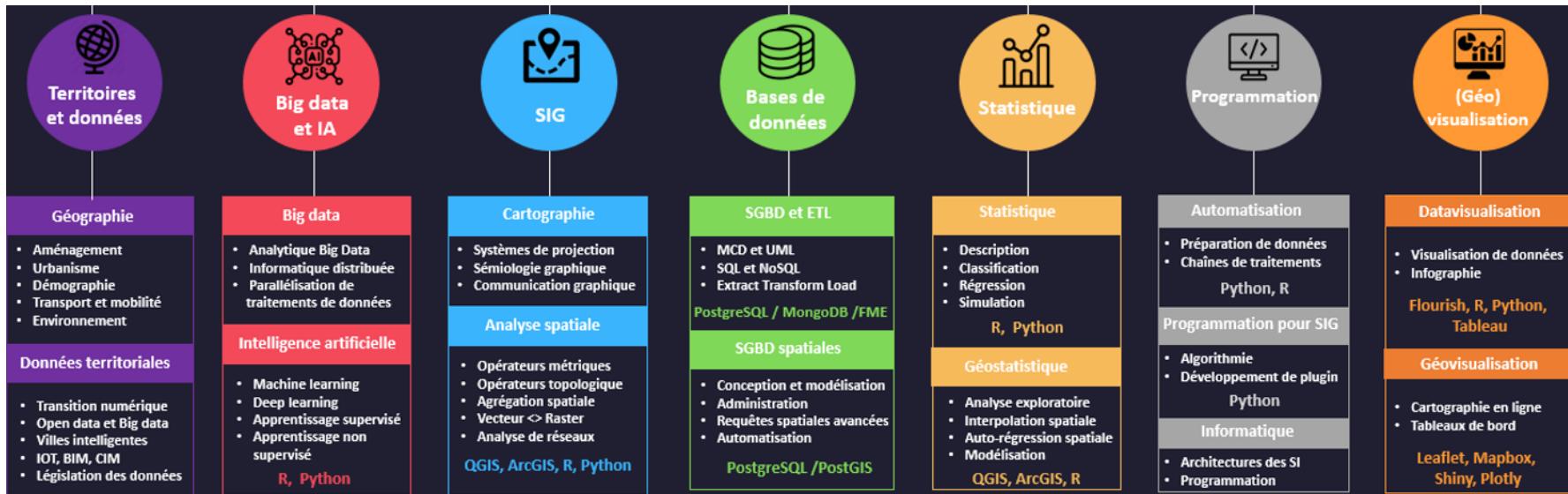


Science des données spatiales & Agences d'urbanisme

Outils, méthodes, compétences et enjeux



Science des données spatiales

- Domaine interdisciplinaire qui combine les principes de la géomatique et de la science des données
 - Elle vise à **traiter automatiquement** et **extraire des connaissances** et des informations à partir de multiples données pour comprendre et résoudre des problèmes liés à l'espace, à l'environnement et à d'autres domaines connexes
- Assemblage de méthodes, de savoirs et de techniques combinant préparation et exploration de données, analyse spatiale, statistique, cartographie, programmation et intelligence artificielle

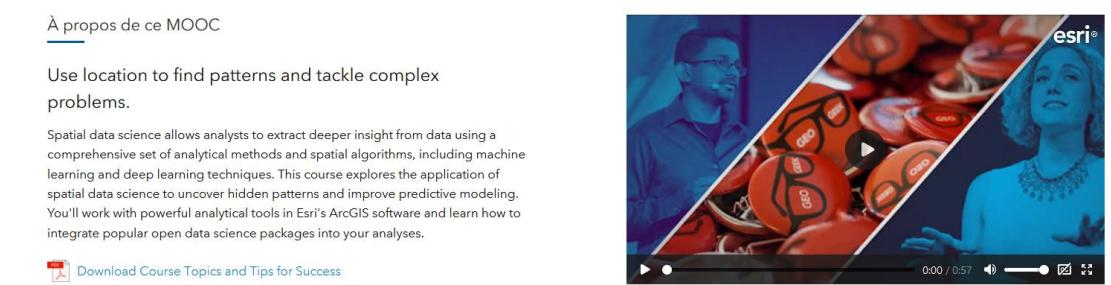


Science des données spatiales

- La science des données spatiales est aujourd’hui mise en avant par de plus en plus d’entreprises spécialisées dans le domaine du géospatial

The screenshot shows a blue-themed page from the CARTO website. At the top, there's a navigation bar with links for Platform, Solutions, Read, Learn, and Connect. Below this, a section titled "DEFINITION" features the question "What is Spatial Data Science?". A detailed text definition follows: "Spatial Data Science (SDS) is a subset of Data Science that focuses on the special characteristics of spatial data, using modeling to know where and why things happen." At the bottom left is a red button labeled "Watch video →".

The screenshot shows a MOOC listing for "Spatial Data Science: The New Frontier in Analytics" on the Esri Academy website. The page has a red header with the Esri logo and navigation links for Produits, Secteurs d'activité, Assistance et services, Récits, À propos, Catalogue (which is underlined), Certification, Mon académie, and Aide. The main content area features the course title, a "S'inscrire" button, and social sharing icons. It also mentions "6 Semaines (2-3 heures d'étude par semaine)" and "Gratuit".



Science des données spatiales

- Côté académique, on assiste à l'émergence de centres de recherche dédiés, à des colloqués dédiés et à l'ouverture de nombreuses formations universitaires spécifiques

 UNIVERSITY OF LIVERPOOL

[Study with Liverpool](#) ▾ [Our research](#) ▾ [About us](#) ▾ [Search](#) [A-Z](#) [Sign in: Staff / Students](#)

Geographic Data Science Lab

[Geographic Data Science Lab](#)

- [About](#)
- [Study with us](#)
- [Work with us](#)
- [Our People](#)
- [Research](#)
- [News](#)
- [Contact](#)

Part of the School of Environmental Sciences

Geographic Data Science Lab

A centre of excellence for research and teaching, solving grand human challenges through geographic data science, working at the intersection between data science and geography.

[Find out more >](#)

 University of BRISTOL

[Study at Bristol](#)

Current students | Current staff | Alumni [Search](#)

[Home](#) > Study at Bristol > Postgraduate study > Faculty of Science > MSc Geographic Data Science and Spatial Analytics

MSc Geographic Data Science and Spatial Analytics

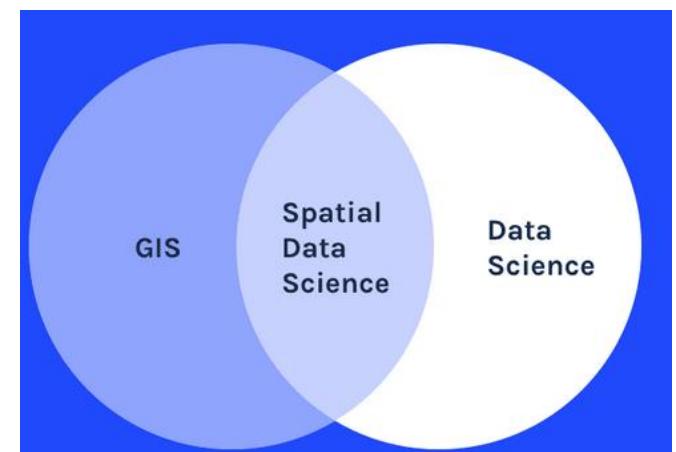
[Apply now →](#) [Book online events →](#)

[Overview](#) [Structure](#) [Entry requirements](#) [Fees and funding](#) [Careers](#) [How to apply](#) [Contact](#)



Science des données spatiales

- En France, la géomatique et la science des données restent encore éloignées et peu de passerelles existent entre les communautés
 - Du côté de la science des données, la dimension spatiale est essentiellement envisagée par le prisme de la cartographie (représentation)
 - Du côté de la géomatique, la science des données n'est pas encore réellement intégrée aux développements actuels et reste envisagée comme un domaine très axé sur les statistiques et l'intelligence artificielle



Science des données spatiales

- Il existe des synergies à développer entre ces deux mondes, pour à la fois repenser la géomatique à l'heure des données massives et enrichir la science des données d'une réelle dimension géographique
 - D'un côté, la géomatique peut bénéficier des avancées de la science des données pour le traitement et l'analyse de gros volumes de données spatiales ou l'utilisation des méthodes d'intelligence artificielle
 - De l'autre, la science des données peut bénéficier des apports de la géomatique pour l'exploration, l'analyse spatiale et la représentation des données spatialisées

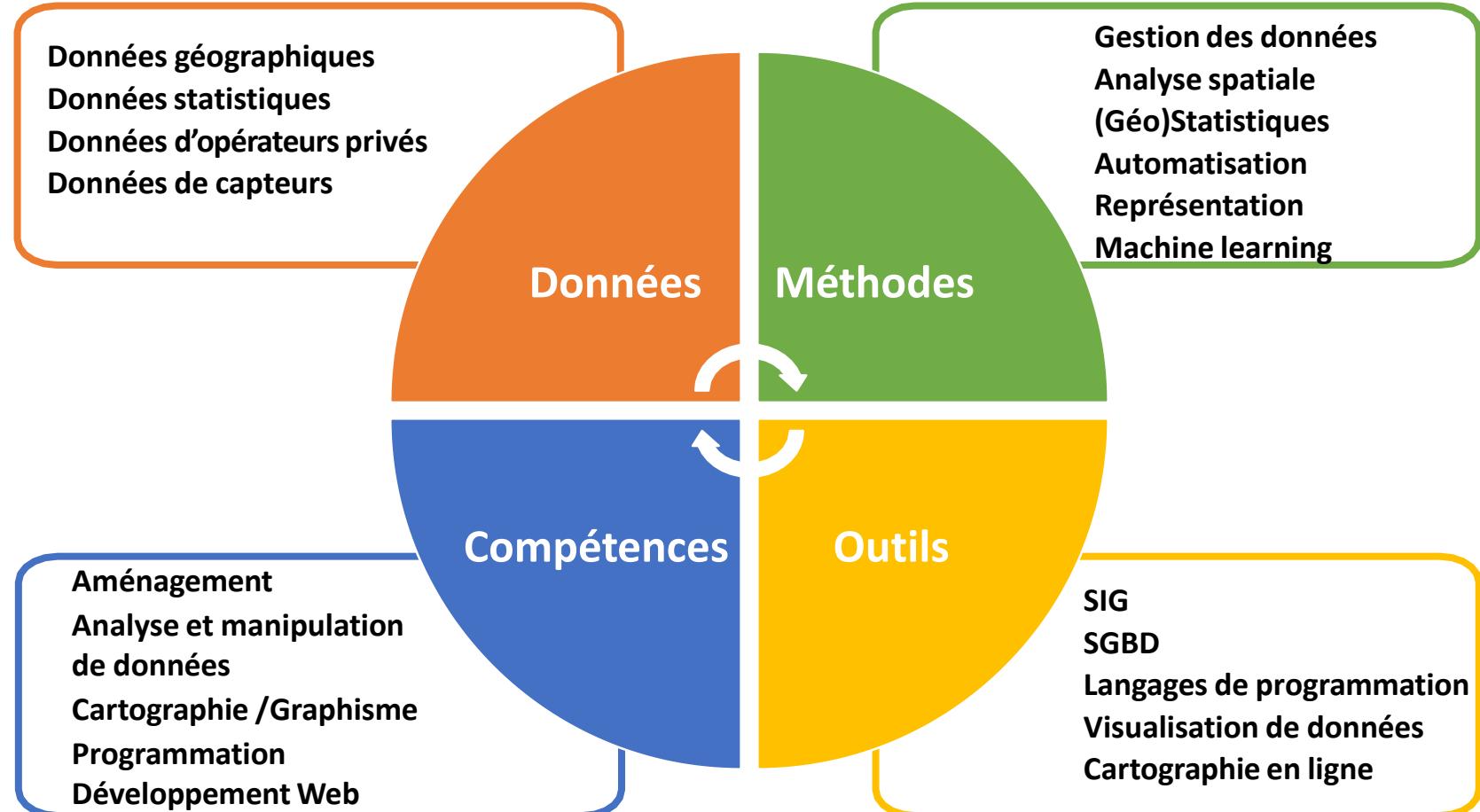
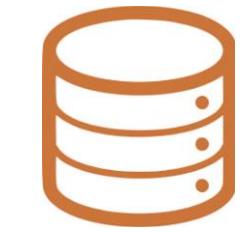




Alors Hype ou réalité opérationnelle ?

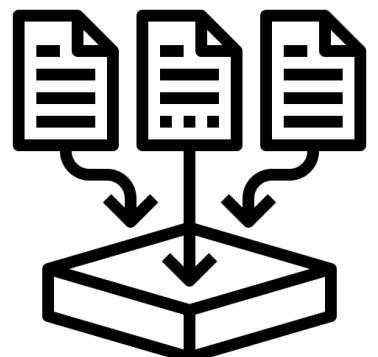
- Approche très tendance, très moderne, très à la mode
 - Une nouvelle étiquette pour caractériser et identifier la géomatique au sein d'autres secteurs
 - S'inscrire dans les logiques et les dynamiques actuelles autour de la « data »
 - La Spatial data science est morte vive l' Urban data science !
- Au-delà d'un effet de mode une vraie évolution de la géomatique
 - La révolution des données
 - De nouvelles méthodes à mobiliser
 - L'évolution des outils
 - Des attentes du monde opérationnel en pleine évolution

Repenser la géomatique





- Trouver les « bonnes » et les « nouvelles » données territoriales
- Travailler avec de nouveaux formats de données
- Travailler avec des données issues de nouvelles sources
- Récupérer des données via des APIS
- Récupérer et archiver des données en temps réel
- Récupérer des données via le WebScraping
- Travailler sur des millésimes
- Intégrer les données au sein des BD/SGBD/SI





- Structurer des données multi-sources
 - Archiver / Documenter
- Combiner des données hétérogènes
 - Structuration / Granularité / Temporalité
- Stocker et manipuler des données volumineuses
 - Indexer / Paralléliser
- Interroger les données
 - Filtrer, trier, regrouper
- Préparer les données
 - Convertir / Spatialiser / Nettoyer / Corriger / Enrichir / Restructurer
- Mettre en qualité ses données
 - Précision / temporalité / sémantique



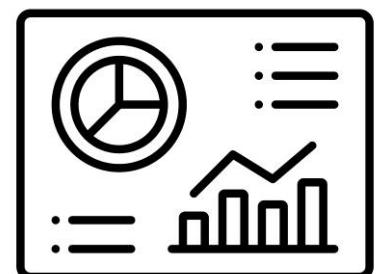


- Analyse spatiale
 - Répartition / Distance / Agrégation / Analyse multicritères
- Analyses statistiques
 - Univarié / Multivarié (ACP, ACM, AFC, CAH, clustering)
- Analyses géostatistiques
 - Interpolation / Autocorrélation / Lissage
- Modélisation
- Machine learning
 - Approches supervisées et non supervisées





- Cartographie statique
 - Cartographie thématique / Print / Numérique / Communication
- Visualisation de données
 - Dataviz statique ou dynamique
- Cartographie en ligne
- Tableaux de bord
- Data Storytelling
- Animation

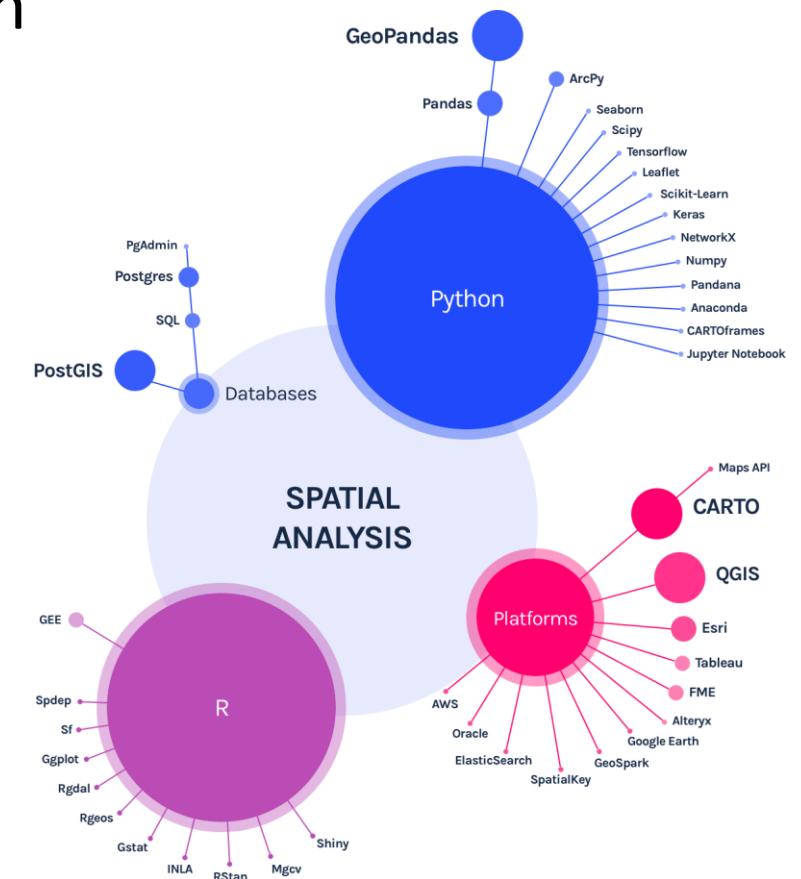


Science des données spatiales



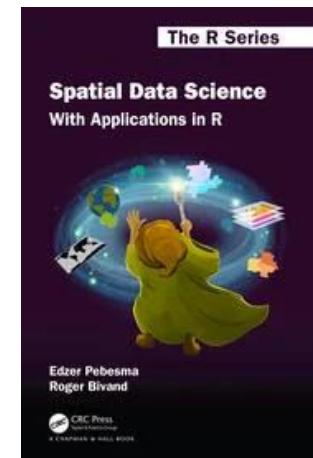
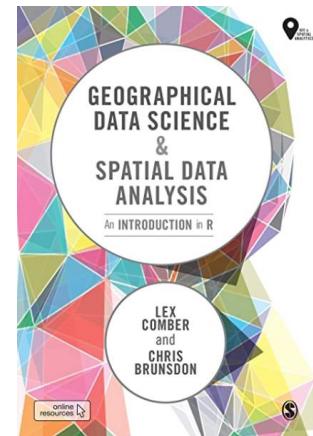
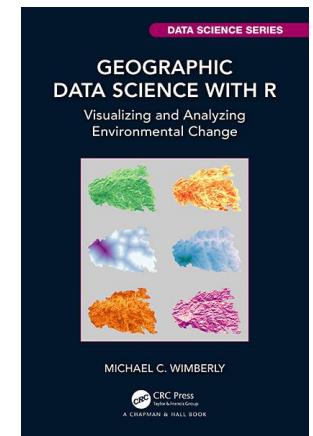
Le choix des armes

- Au centre de la GDS se trouve la question des outils et plus spécifiquement des langages de programmation
- Faire de la géomatique en codant
 - De l'interface graphique à la console
 - Du clic bouton aux lignes de code
 - Des boites de dialogues à la syntaxe
 - Des boites à outils aux librairies
- Automatisation et reproductibilité
- Puissance, stabilité et évolutivité



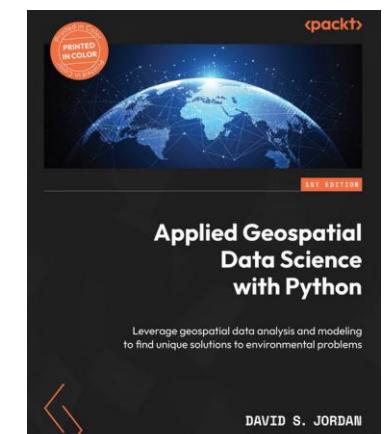
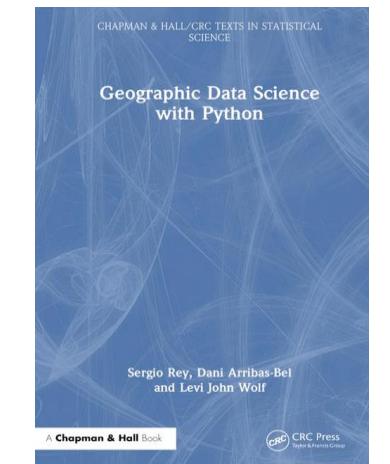
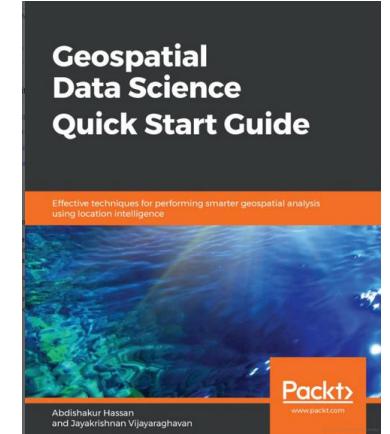


- R est un langage de programmation orienté statistiques et la science des données
- Il propose de nombreux packages dédiés aux données géographiques
 - Importation/exportation de données (*sp, sf, raster, rgdal*)
 - Analyse spatiale (*tidyverse, spatial, rgeos, lwgeom*)
 - Géostatistique (*gstat, spatstat, geoR*)
 - Cartographie et géovisualisation (*maps, ggplot2, tmap, mapview*)
 - Ressources et communauté SIG très active (CRAN)
- Très utilisé par les chercheurs



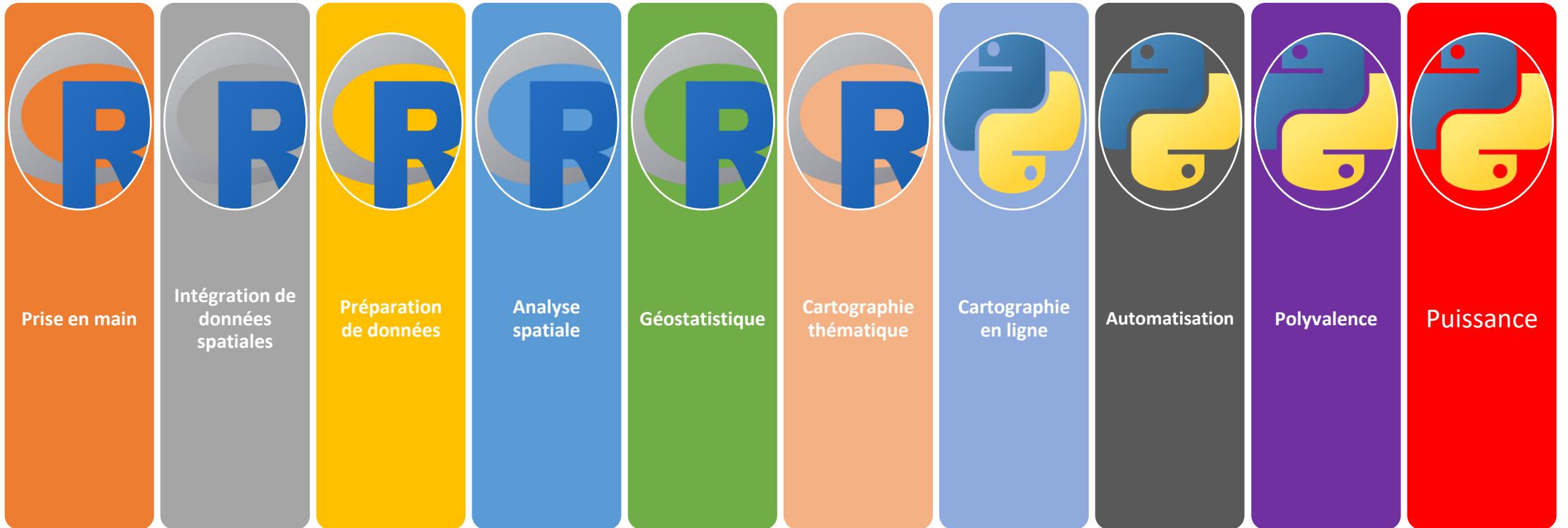


- R est un langage de programmation plus populaire et plus polyvalent utilisé dans de nombreux domaines
- Il propose de nombreux packages dédiés aux données géographiques
 - Importation/exportation de données (*Pandas, Fiona*)
 - Analyse spatiale (*Numpy, Shapely, GeoPandas, Rastersio*)
 - Géostatistique (*scikit-learn, PyKrig, PySAL*)
 - Cartographie et géovisualisation (*Matplotlib, Cartopy, Folium, Pydeck*)
 - Ressources et communauté SIG moins active
- Très utilisé par les acteurs opérationnels (privés et publics)





Petite synthèse



Pourquoi pas les deux !

Hello, Quarto

Python R Julia Observable

Combine Jupyter notebooks with flexible options to produce production quality output in a wide variety of formats. Author using traditional notebook UIs or with a plain text markdown representation of notebooks.

Palmer Penguins

```
author: Norah Jones
format:
  html:
    code-tools: true
    code-fold: true
```

```
[3]: #| echo: false
import pandas as pd
df = pd.read_csv("palmer-penguins.csv")
```

Exploring the Data

See @fig-bill-sizes for an exploration of bill sizes by species.

```
[6]: #| label: fig-bill-sizes
#| fig-cap: Bill Sizes by Species
import matplotlib.pyplot as plt
import seaborn as sns
g = sns.FacetGrid(df, hue="species", height=3, aspect=3.5/1.5)
g.map(plt.scatter, "bill_length_mm", "bill_depth_mm").add_legend()
```

```
[6]: <seaborn.axisgrid.FacetGrid at 0x2946720e0>
```

Figure 1: Bill Sizes by Species

Palmer Penguins

AUTHOR
Norah Jones

PUBLISHED
March 12, 2023

Show All Code Hide All Code

Exploring the Data

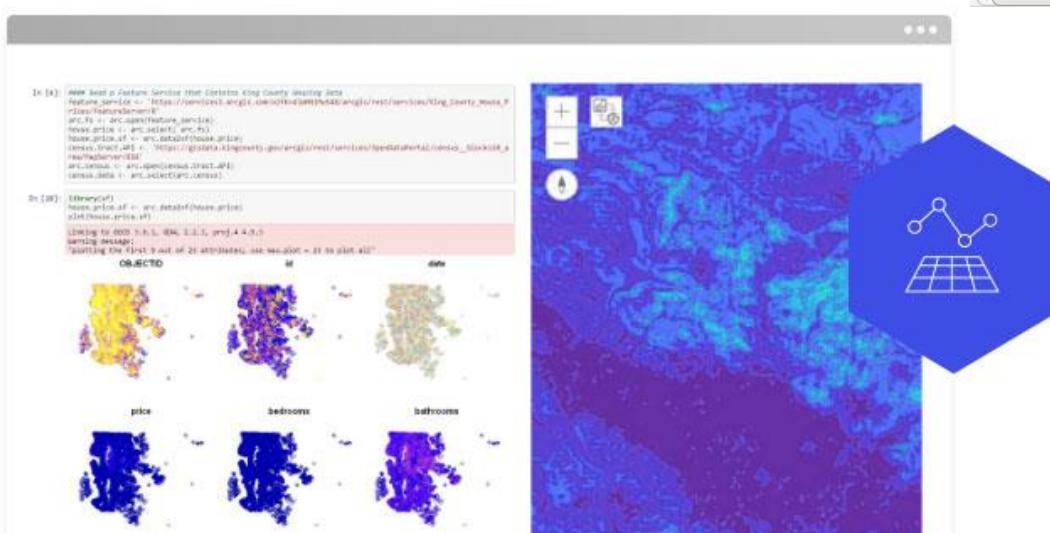
See [Figure 1](#) for an exploration of bill sizes by species.

▼ Code

```
import matplotlib.pyplot as plt
import seaborn as sns
g = sns.FacetGrid(df, hue="species", height=3, aspect=3.5/1.5)
g.map(plt.scatter, "bill_length_mm", "bill_depth_mm").add_legend()
```

Figure 1: Bill Sizes by Species

Liens avec nos logiciels SIG



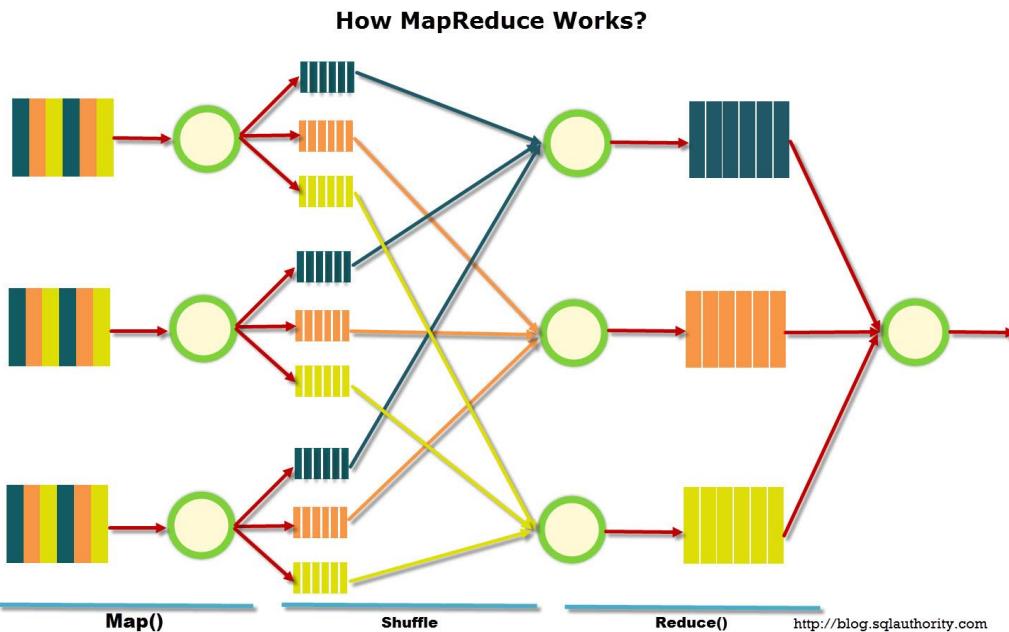
```
# my script to count the number of schools in girona
vl = QgsVectorLayer("/home/alexandre/Desktop/buildings.shp", "buildings", "ogr")
if not vl.isValid():
    print "failed to load the layer"
count = 0
for feature in vl.getFeatures():
    if feature["TYPE"] == "School":
        count += 1
print "total schools:", count
```

```
import fme
import fmeobjects
class FeatureProcessor(object):
    def __init__(self):
        self.feature_list = []
        self.total_area = 0.0
    def input(self, feature):
        self.feature_list.append(feature)
        self.total_area += feature.getGeometry().getArea()
    def close(self):
        for feature in self.feature_list:
            feature.setAttribute("total_area", self.total_area)
            self.pyoutput(feature)
    def process_group(self):
        pass
```



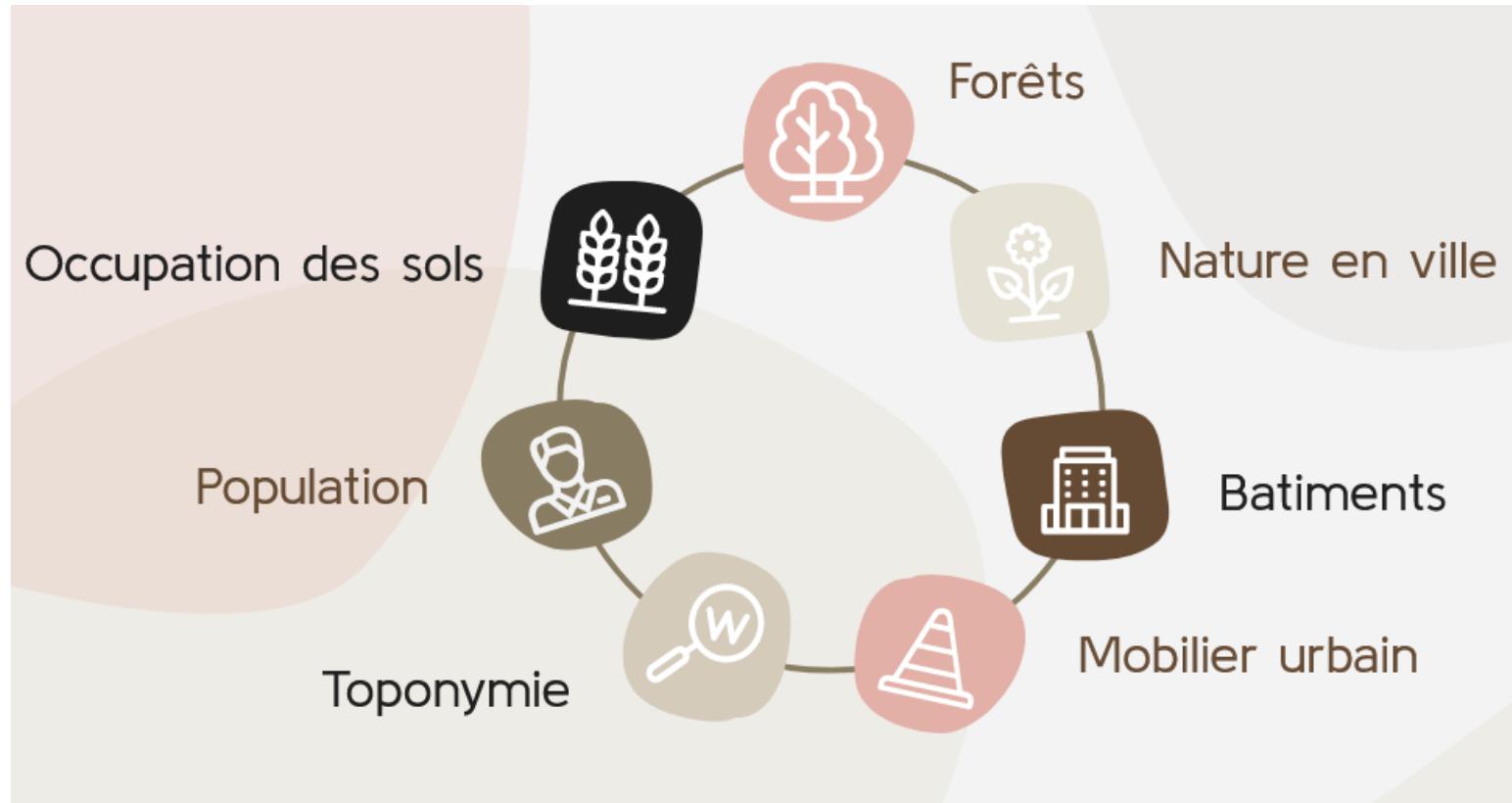
```
import arcpy
try:
    aprx = arcpy.mp.ArcGISProject("CURRENT")
    l = aprx.listLayouts()[0]
    if not l.mapSeries is None:
        ms = l.mapSeries
        if ms.enabled:
            ms = l.mapSeries
            indexLyr = ms.indexLayer
            ms.exportToPDF(r"C:\Student\ProgrammingPro\Chapter6\BellCounty\DelinquentLandProperties.pdf")
except Exception as e:
    print("Error: " + e.args[0])
```

Géomatique et Big Data

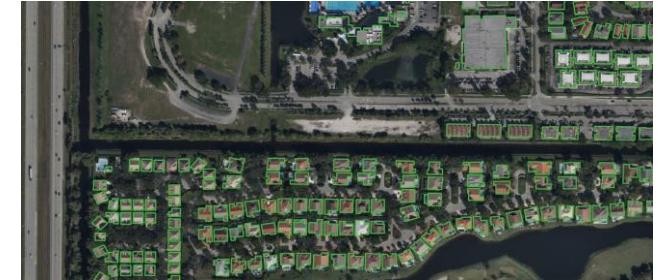


Logiciel	Hadoop GIS	GIS tool for Hadoop	SpatialHadoop	Spatial Spark	Impala & ISP-MC
Créateur	Université d'Emory	ESRI	Université du Minnesota	Apache Spark	Simin You
Projet actif	Non	Oui	Oui	Maintenu mais pas actif	Non
Fonctions spatiales	Index Distance au plus proche voisin Jointure Agrégations spatiales 10 traitements spatiaux	Index Distance au plus proche voisin Jointure Agrégations spatiales. 86 traitements spatiaux	Index Distance au plus proche voisin Jointure Agrégations spatiales 88 traitements spatiaux	Index Distance au plus proche voisin Jointure Agrégations spatiales	Index Jointure
Niveau de documentation	Moyen	Elevé	Faible	Moyen	Faible

Géomatique et IA en mode raster



Apprentissage supervisé : Détection, extraction, segmentation et classification d'objets



Highway interchange signs



Polar bear sign



Géomatique et IA en mode vecteur

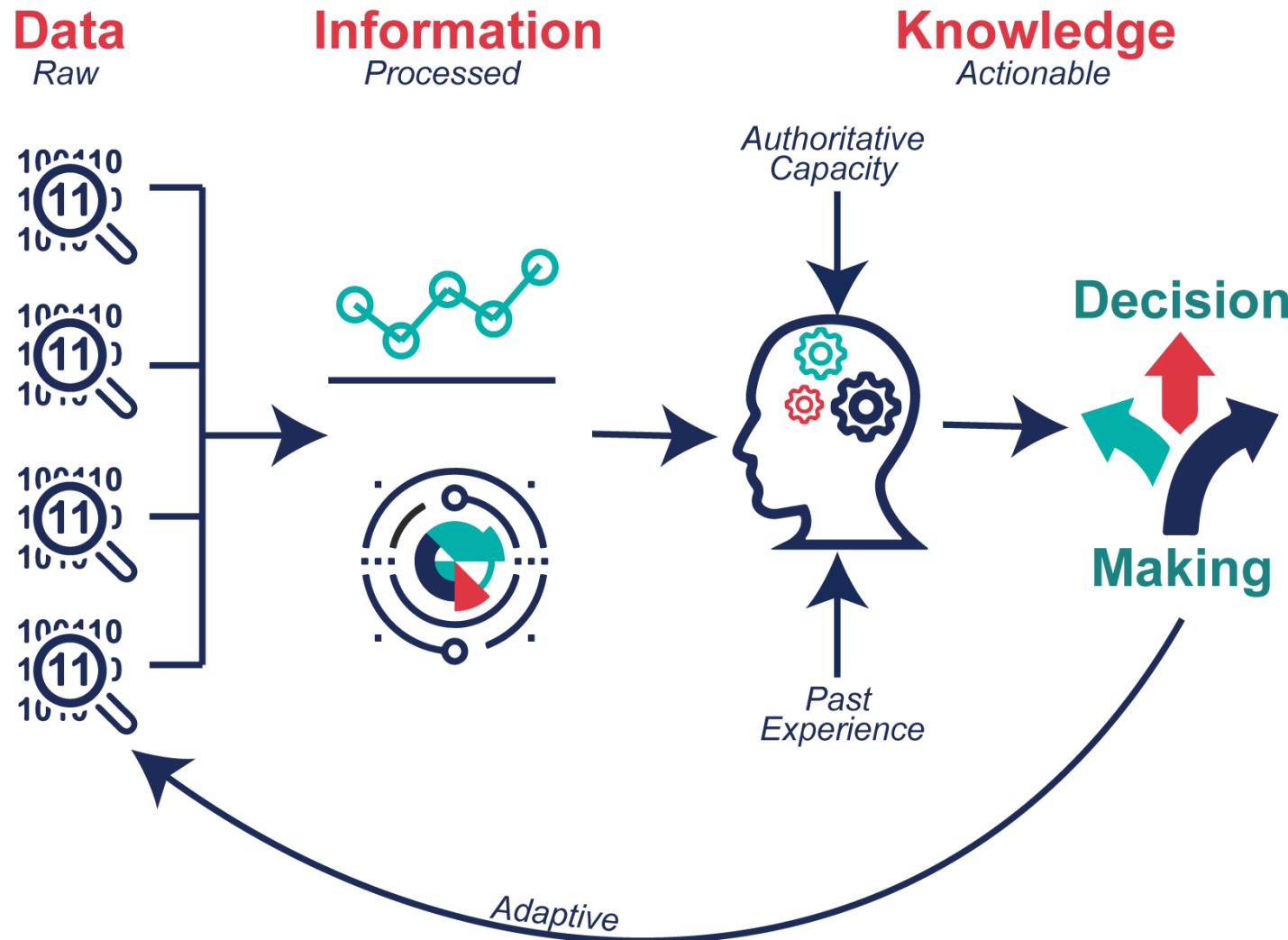
- Apprentissage non supervisé (le modèle ne connaît pas les réponses et les prédit)
- Mobilisation de méthodes statistiques et géostatistiques appliquées aux données géographiques
 - Détection d'anomalies
 - Classification non supervisée (K-means)
 - Réduction de dimension (ACP, ACM)
 - Clustering spatial (CAH, DBSCAN)
 - Régression géographiquement pondérée
 - Krigeage bayésien empirique

	Conventional Machine Learning	Machine Learning for Spatial Analysis
Regression	<ul style="list-style-type: none">• Linear regression• Decision Tree regression• Random Forest regression• Gradient Boosting regression• Support Vector Machine regression• K Nearest Neighbor	<ul style="list-style-type: none">• Empirical Bayesian Kriging• EBK with independent variables• Ordinary Least Squares (OLS) Regression• Geographically Weighted Regression (GWR)• Areal Interpolation
	<ul style="list-style-type: none">• Decision Tree classification• Random Forest classification• Support Vector Machine classification• Maximum Likelihood	<ul style="list-style-type: none">• Decision Tree classification• Random Forest classification• Support Vector Machine classification• Maximum Likelihood
	<ul style="list-style-type: none">• Hyerarchical clustering• K-means• PAM• DBSCAN	<ul style="list-style-type: none">• Multivariate clustering• Spatially Constrained Multivariate Clustering• Hot Spot Analysis• Density-based clustering• Image Segmentation• Space Time Pattern Mining

Données spatiales = données spéciales

- Les données spatiales sont souvent traitées de manière inappropriée ou même ignorées dans l'apprentissage automatique
 - Par rapport à d'autres ensembles de données, comme les données de séries chronologiques, l'intégration des données spatiales dans les algorithmes d'apprentissage automatique est encore limitée
- La composante spatiale comme spécificité
 - Dépendance spatiale (proximité), échelles, MAUP, unité d'observation
- La composante sémantique comme complexité
 - Des attributs hétérogènes relatifs à la localisation
- La modélisation de l'espace comme enjeu
 - 2D/3D, vecteur/raster, discrète/continue
- Complexité de l'espace et des territoires comme contexte d'analyse

Des données à la décision



La question des contexte d'usage

pour quoi ?

- ◆ communiquer
- ◆ connaître
- ◆ percevoir
- ◆ comprendre
- ◆ généraliser
- ◆ présager
- ◆ décider

pour qui ?

- ◆ humains
- ◆ machines

Diagnostic, gestion,
planification, modélisation,
concertation, pilotage,
évaluation...

MÉTHODES DE VENTILATION DE POPULATION

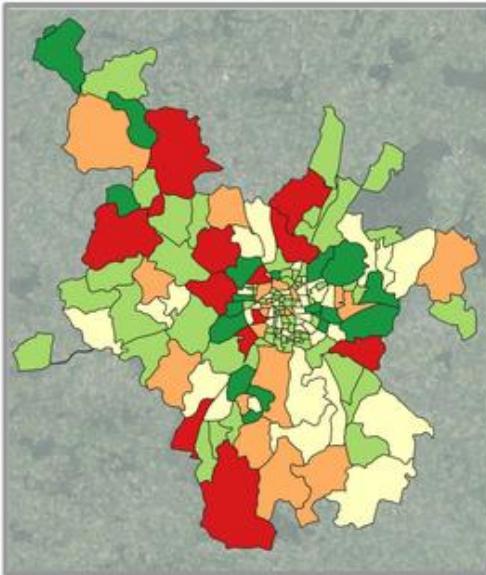


447 429 hab

43 communes

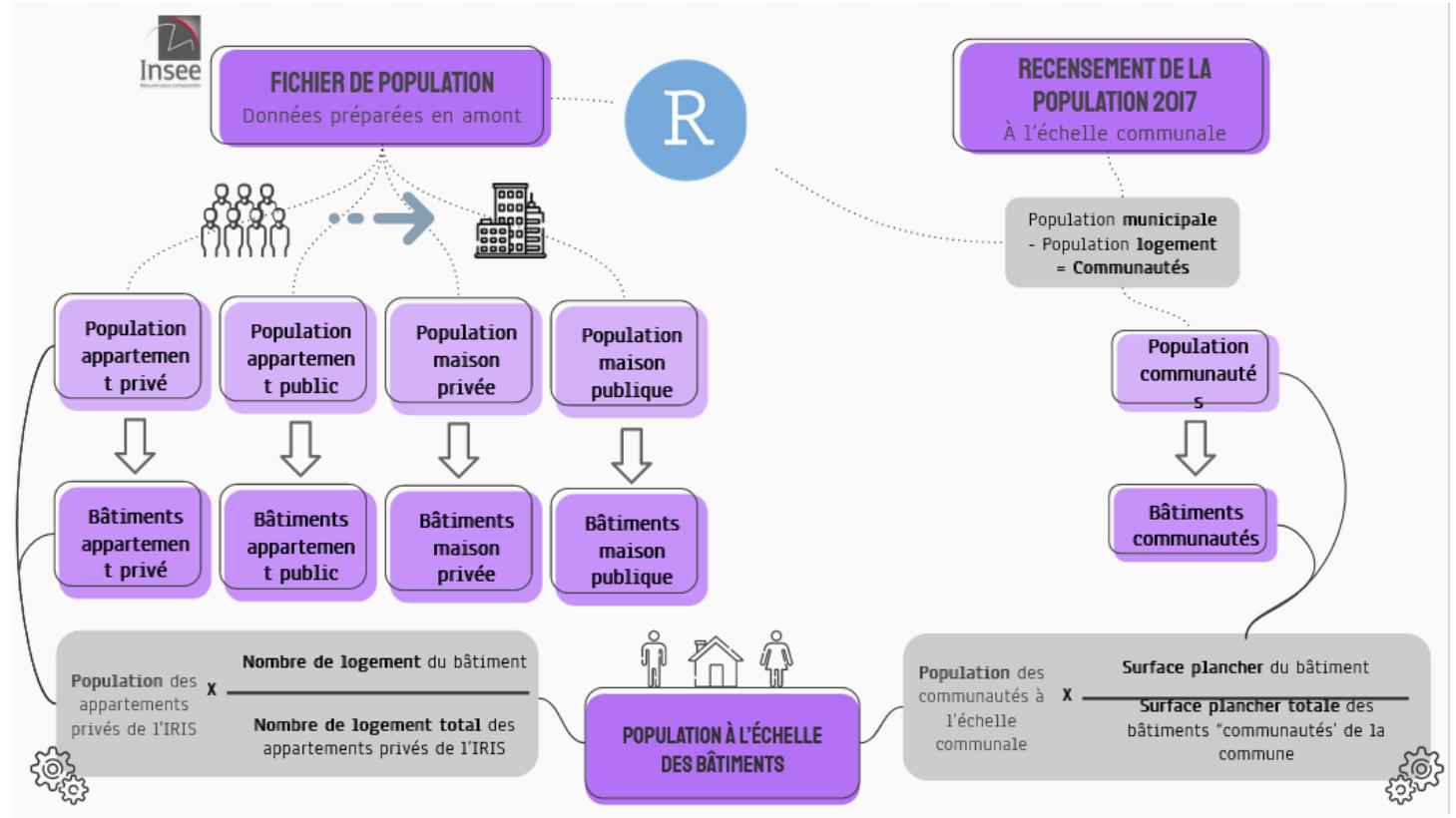
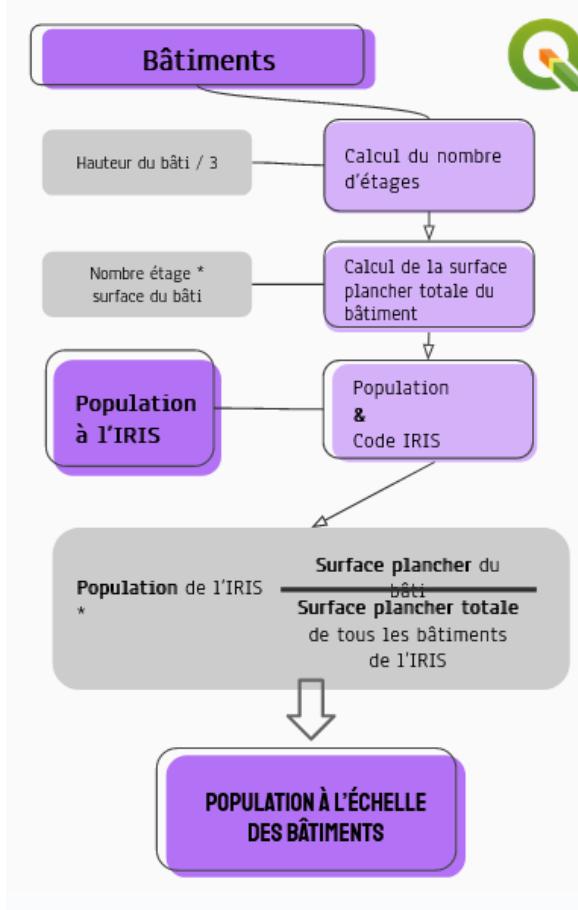
Échelle 7

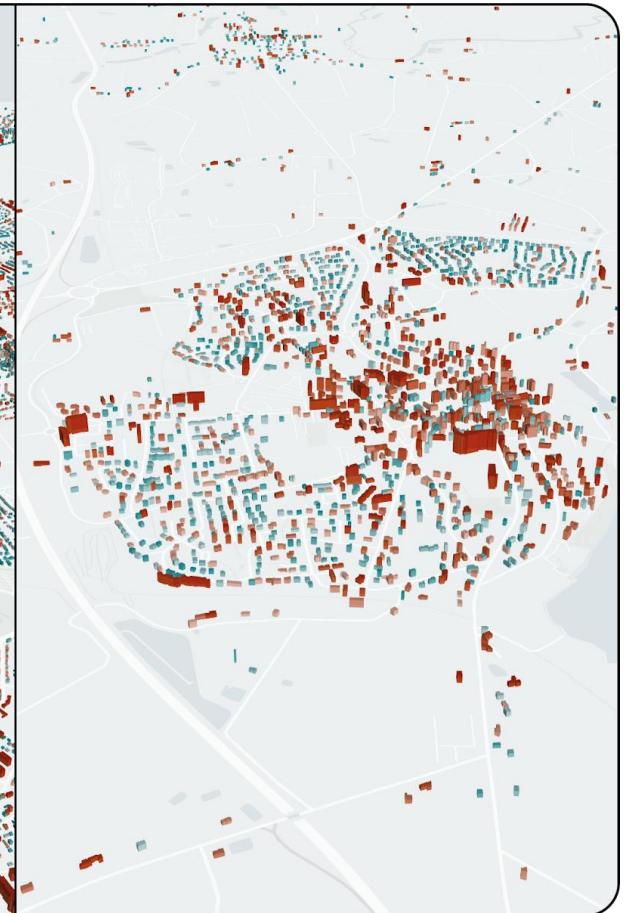
Population à l'IRIS



Population aux bâtiments







“CLASSIQUE”



- Rapide
- Simple dans les traitements
- Corrélation avec les données initiales



- Répartition de la population en fonction de la caractéristique physique du bâti
- Ne tient pas compte de la réalité des logements

“ESTIMATION”



- Personnalisation et modulation des facteurs de pondération



- Utilisation de moyennes nationales, éloignées des réalités locales
- Environnement QGIS moyennement approprié

“AVANCÉE”



- Documentation, reproductibilité et automatisation
- Plus de fonctionnalités
- Stabilité



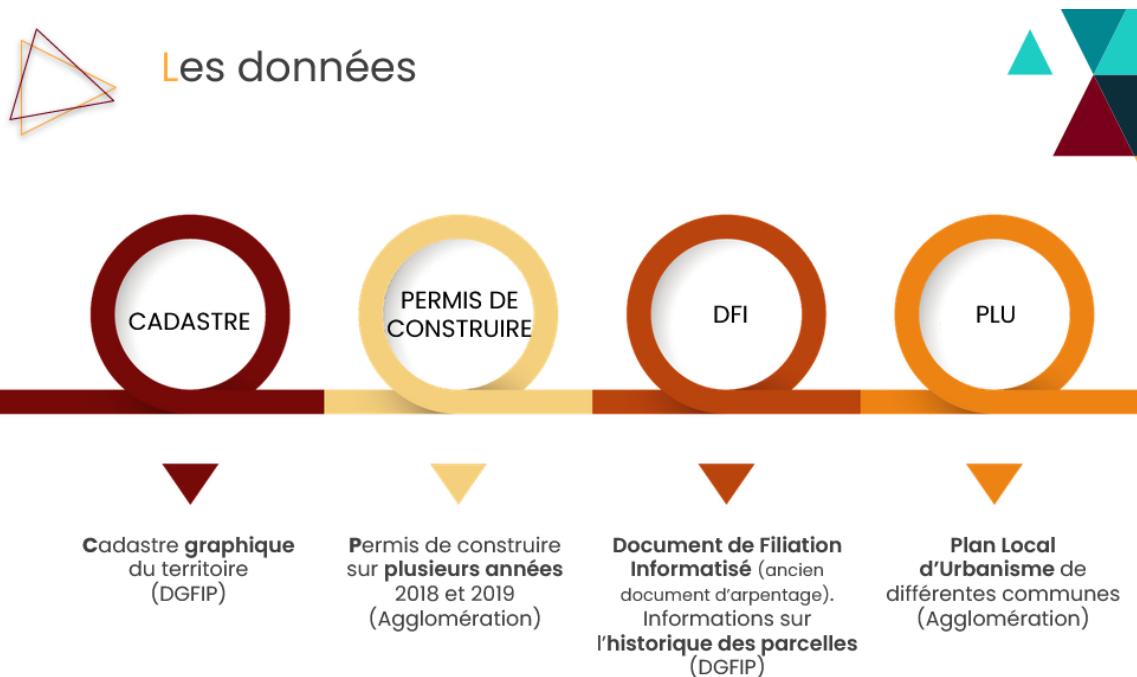
- Complexité technique
- Lenteur
- Mauvaise prise en compte des cas particuliers

HISTORISATION DES DONNÉES PARCELLAIRES & DES AUTORISATIONS D'URBANISME DU PAYS DE LORIENT

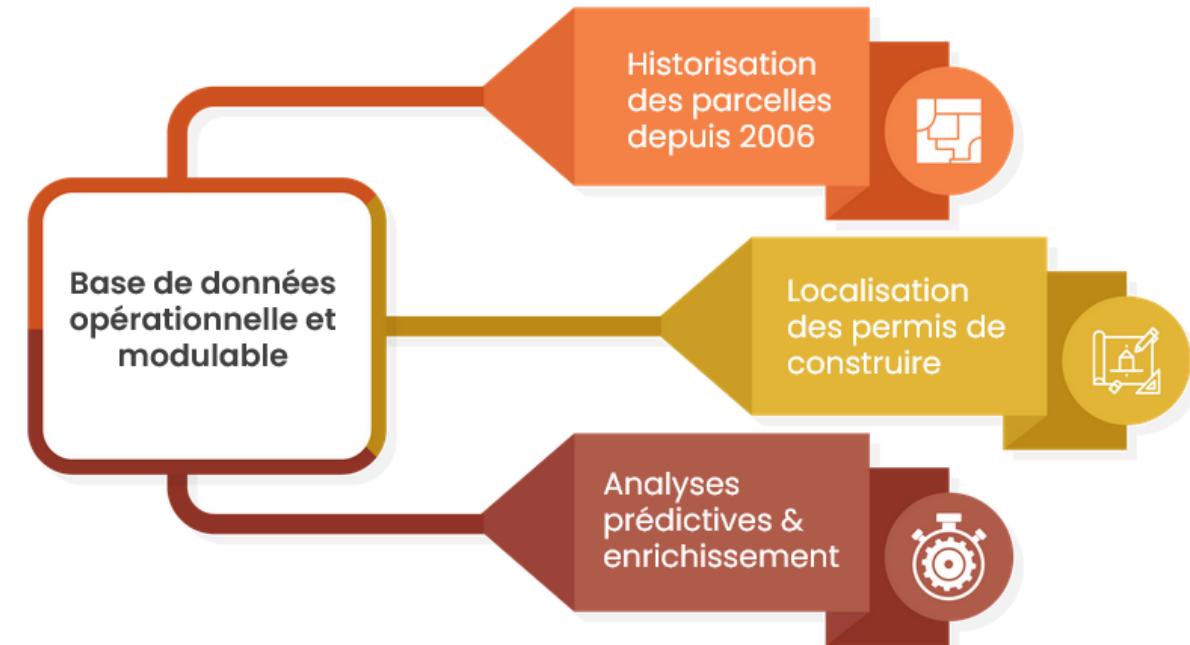
Vincent BRE
Marianna LEBANSAIS
Marion PEREZ
Jessy REPLUMAZ

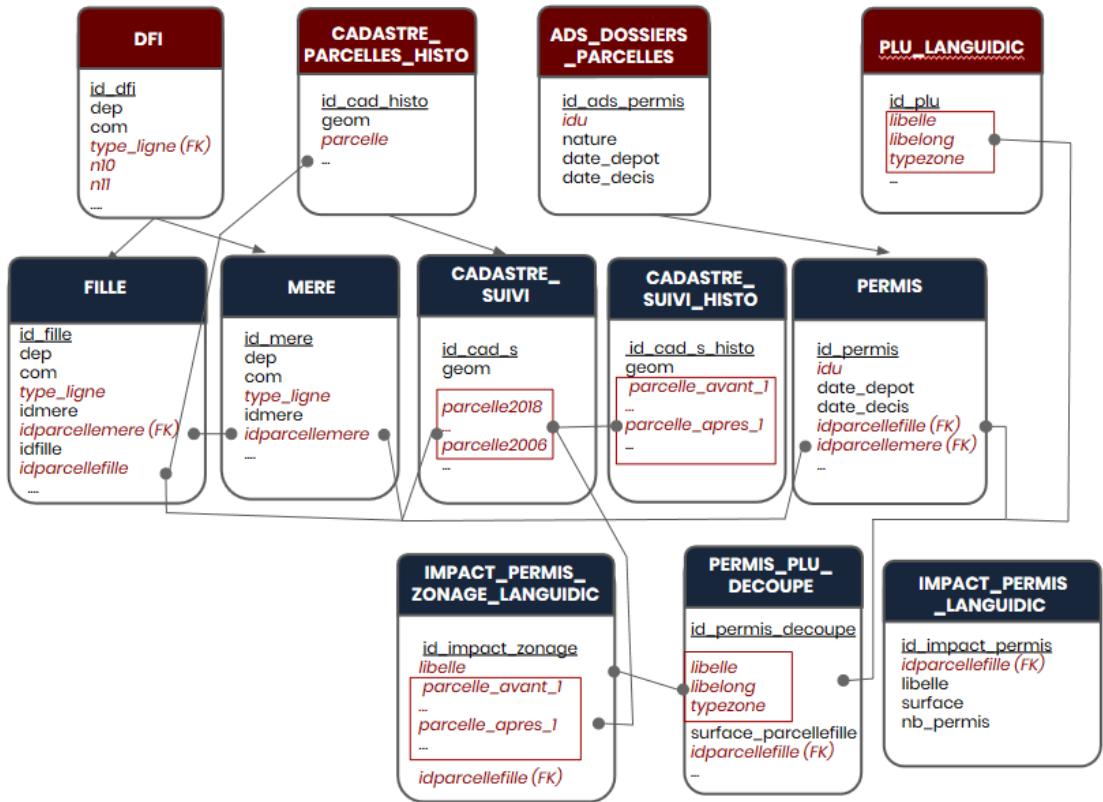


Les données



Méthodologie & Déroulement



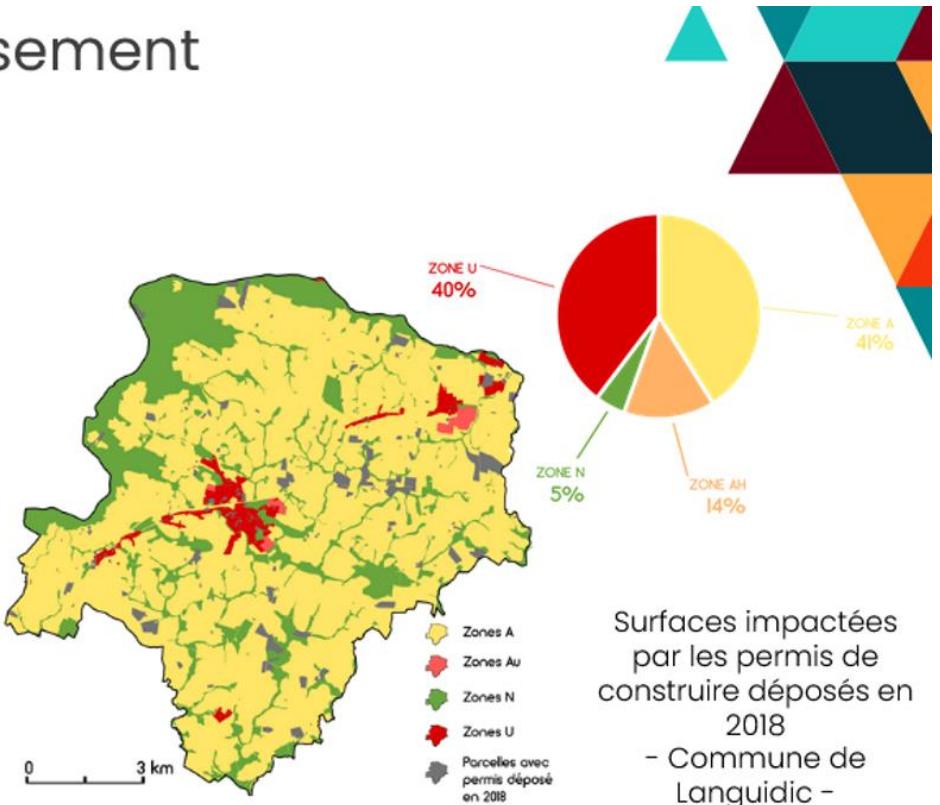


```

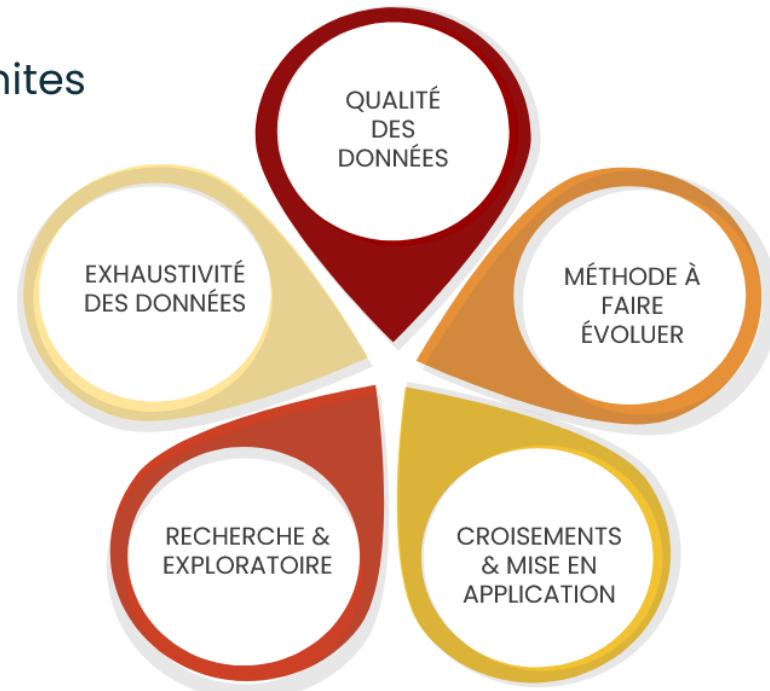
-- Nom du schéma : a
-- Nom de la table dfi : dfi_lorient_agglo
-- Nom de la table parcelles mères : mere_lorient_agglo
----- Crédation du fichier parcelles filles à l'échelle de Lorient Agglomération à partir de dfi_lorient_agglo et d
-- Crédation d'une nouvelle table
CREATE TABLE a.fille_lorient_agglo AS
-- Crédation d'une première couche temporaire avec une commande with
-- With nom_table as (Select ... from ...)
-- Une couche que je peux mobiliser dans la suite de mes traitements sans la voir physiquement dans notre base de donné
-- Sélection de l'ensemble de mes champs de la table mere_lorient_agglo
-- Sélection de l'ensemble de mes champs de la table dfi_lorient_agglo correspondant à des identifiants de parcelles fi
-- Pour éviter de recopier tous les champs à la main : dans PGadmin, clic droit sur une table puis Scripts puis Script
WITH merejoin AS (SELECT mere_lorient_agglo.*, n10, n11, n12, n13, n14, n15, n16, n17,
n18, n19, n20, n21, n22, n23, n24, n25, n26, n27, n28, n29, n30,
n31, n32, n33, n34, n35, n36, n37, n38, n39, n40, n41, n42, n43,
n44, n45, n46, n47, n48, n49, n50, n51, n52, n53, n54, n55, n56,
n57, n58, n59, n60, n61, n62, n63, n64, n65, n66, n67, n68, n69,
n70, n71, n72, n73, n74, n75, n76, n77, n78, n79, n80, n81, n82,
n83, n84, n85, n86, n87, n88, n89, n90, n91, n92, n93, n94, n95,
n96, n97, n98, n99, n100
FROM a.mere_lorient_agglo
-- JOINTURE : LEFT JOIN nom_schema.nom_table ON nom_table.champ_commun = nom_table.champ_commun
LEFT JOIN a.dfi_lorient_agglo ON
mere_lorient_agglo.idfille = dfi_lorient_agglo.dfi
-- CONDITIONS : n10 peut correspondre à des identifiants de parcelles mère et de parcelles fille donc on veille à ne pa
-- Les parcelles mères ont plusieurs champs commun avec les parcelles filles, il faut que les champs num_lot, nat_dfi
WHERE mere_lorient_agglo.idmere != dfi_lorient_agglo.n10 AND mere_lorient_agglo.num_lot = dfi_lorient_agglo.num_lot
AND mere_lorient_agglo.nat_dfi = dfi_lorient_agglo.nat_dfi AND mere_lorient_agglo.date_valid = dfi_lorient_agglo.date_v
-- Etape 2 : cross table lateral -> les parcelles filles sont listées en ligne entre n10 et n177, on les veut dans une
-- Crédation d'une 2e couche temporaire
-- Je sélectionne tous les champs de ma table temporaire précédente
-- Je donne un nom à une table : exemple d ici en sélectionnant tous les champs avec *
fille AS (SELECT merejoin.gid, merejoin.dep, merejoin.com, merejoin.section, merejoin.dfi, merejoin.nat_dfi, merejoin.d
merejoin.idparcelle AS idparcellemere, d.*)
FROM merejoin
-- Tous les champs listés après VALUES iront dans le champ d(idfille)

```

Prédicatif & Enrichissement



Limites



L'INSTITUT PARIS REGION

Estimation des populations à évacuer face aux risques d'inondation par débordement en Île-de-France

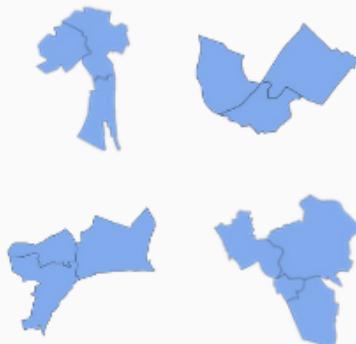


Nina Bouchain, Victor Charpentier, Hugo Dupont, Anthony Laumain, Gurvan Pluen

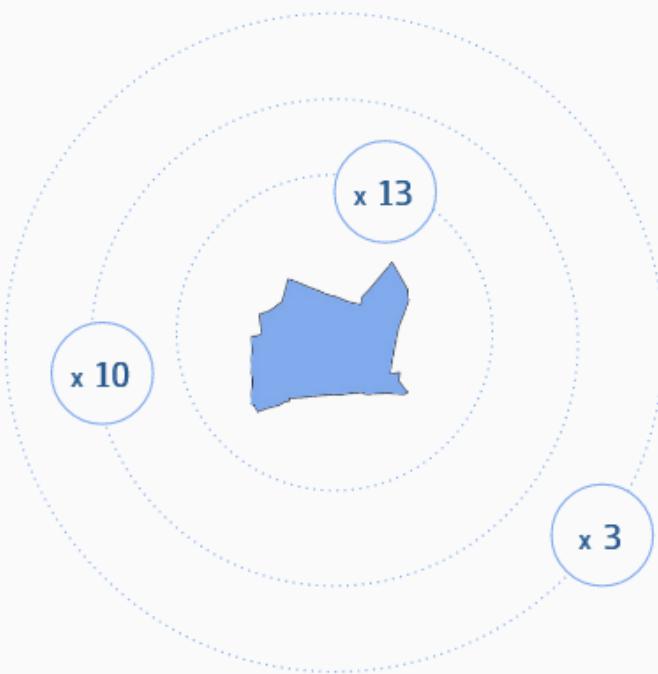
POURQUOI AUTOMATISER ?

Reproductibilité

13 communes



10 scénarios



3 échelles d'analyse



Bâtiments



Grille de 250m

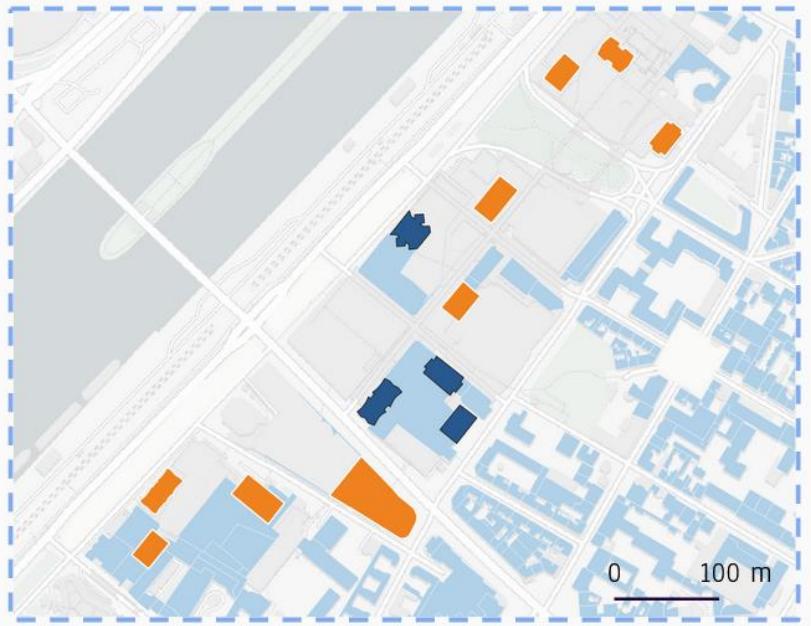


Communes



Paris - XV^e arrondissement

LE NETTOYAGE DES DOUBLONS



IGH
IGH

2 cas de figures :



densibati
contient
IGH

VS



densibati
dans
IGH

MAILLES COMMUNES

Agrégation

Fusion des couches IGH et
bâtiments

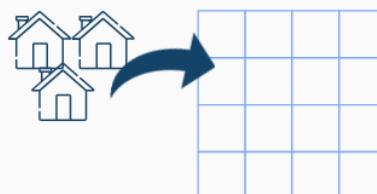


26 535
polygones



Boucle selon le
nombre de scénario

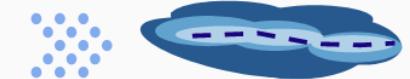
PROCESSUS D'AGRÉGATION DES DONNÉES

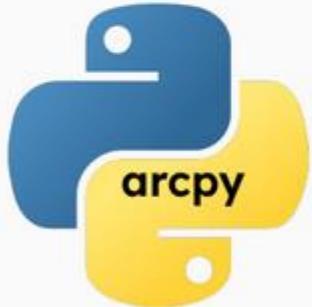


Somme de la population par zone
Somme de la population concernée
Nombre de bâtiments
Zonage

Zone 1
Zone 2
Zone 3

Emprise des zonages





~30 min



Script_IRP_en_libre.ipynb ☆

Fichier Modifier Affichage Insérer Exécution Outils Aide Dernière modification effectuée le 15 mai

+ Code + Texte

[]

Carte de Paris 15ème au scénario de crue 90

Classement de zone

- IGH
- Zone 1
- Zone 2
- Zone 3

Zone 1: 41978 hab
Zone 2: 47438 hab
Zone 3: 32139 hab

Carte de Paris 15ème au scénario de crue 100

Classement de zone

- IGH
- Zone 1
- Zone 2
- Zone 3

Zone 1: 52232 hab
Zone 2: 39663 hab
Zone 3: 30400 hab

▼ Export des couches principales finales

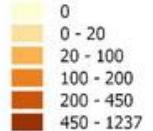
```
[ ] # Export des couches principales en gpk
grilles_final.to_file("CouchesFin.gpkg", layer='grilles_final', driver="GPKG")
communes_final.to_file("CouchesFin.gpkg", layer='communes_final', driver="GPKG")
IGH.to_file("CouchesFin.gpkg", layer='IGH_final', driver="GPKG")
Bati_PopIGH.to_file("CouchesFin.gpkg", layer='densibati_final', driver="GPKG")
emprise_zone_final.to_file("CouchesFin.gpkg", layer='emprise_zone_final', driver="GPKG")
```

PARIS 15ÈME (INSEE: 75115)

230981 HABITANTS - SECTEUR SEINE PARIS - SCÉNARIO R115



POPULATION PAR BÂTIMENT



EMPRISE DU ZONAGE

- Zone 1 - priorité
- Zone 2 - moyenne
- Zone 3 - faible

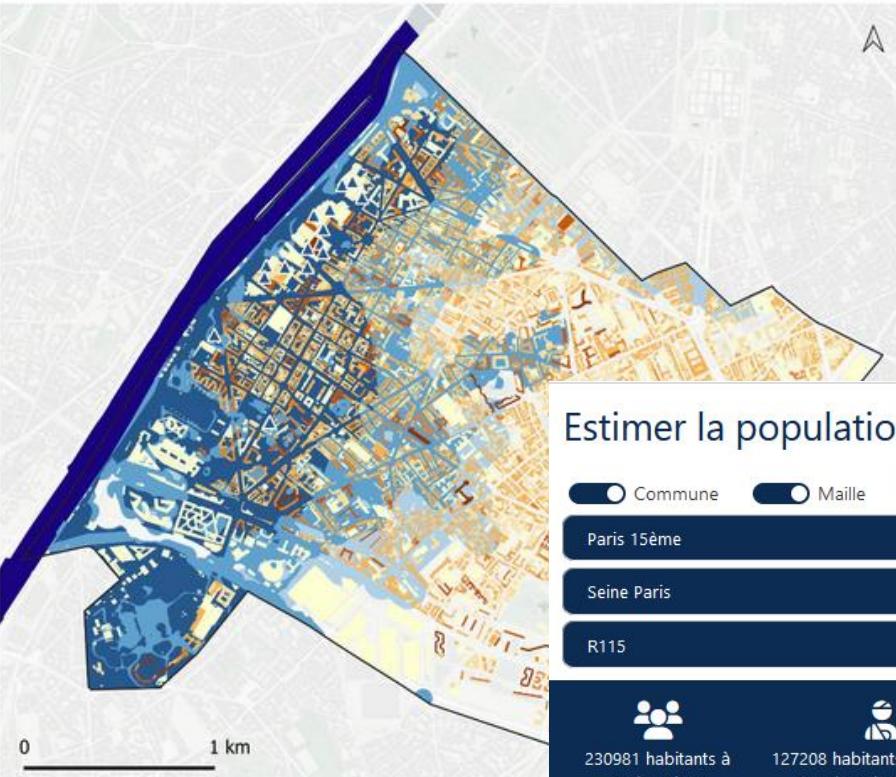
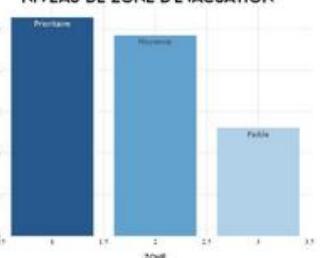
IMMUEBLE DE GRANDE HAUTEUR

- IGH en Zone 1
- IGH en Zone 3

127208 HABITANTS CONCERNÉS

8.47 KM² DE SUPERFICIE

LA POPULATION À ÉVACUER PAR NIVEAU DE ZONE D'ÉVACUATION



Estimer la population à évacuer lors de scénarios de crues en Ile-de-France

Commune Mairie IGH

Paris 15ème

Seine Paris

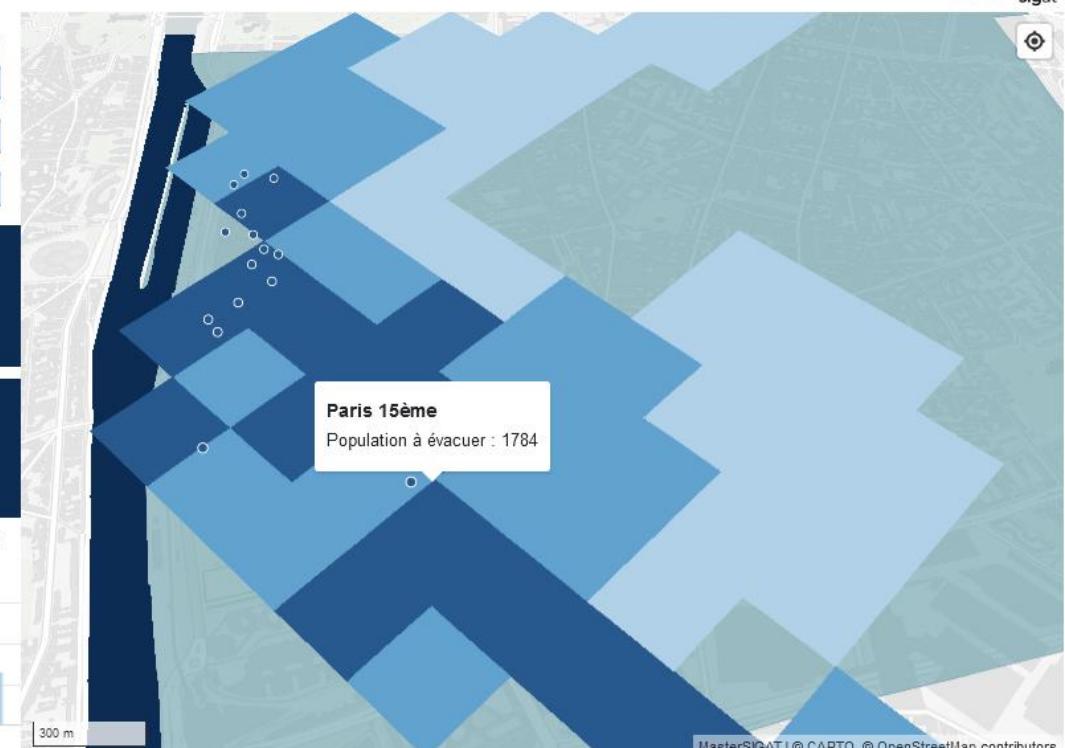
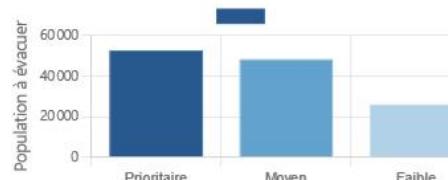
R115

230981 habitants à Paris 15ème 127208 habitants à évacuer pour le scénario R115

Niveau de priorité d'évacuation

MAILLES Faible Moyen Prioritaire
IGH Aucun Faible Prioritaire

Population à évacuer en fonction du degré de priorité

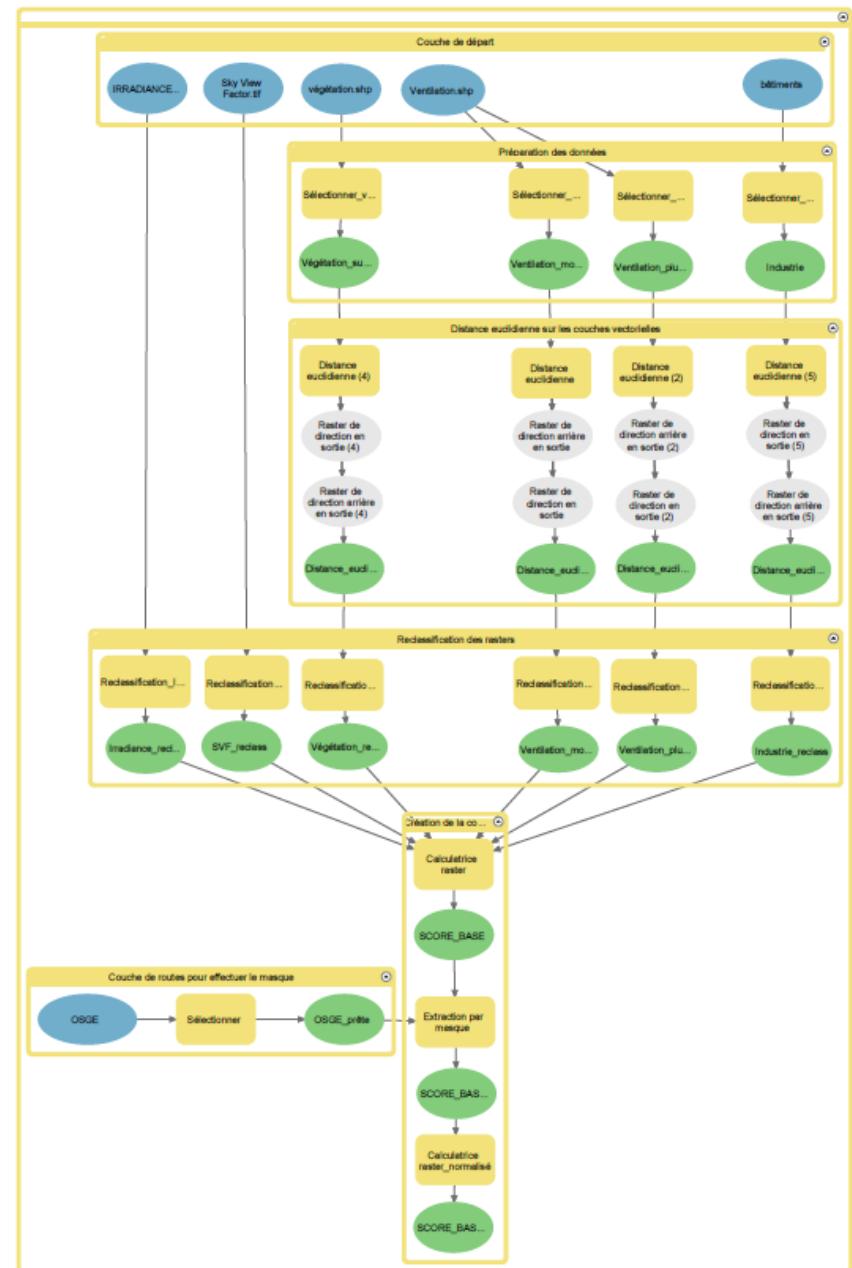
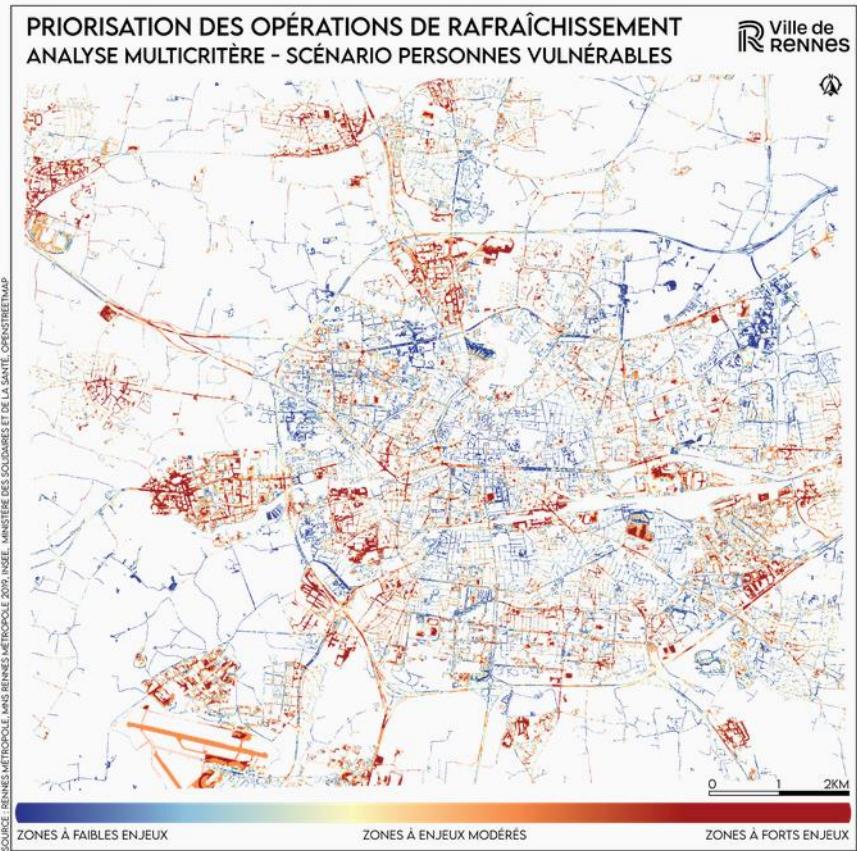
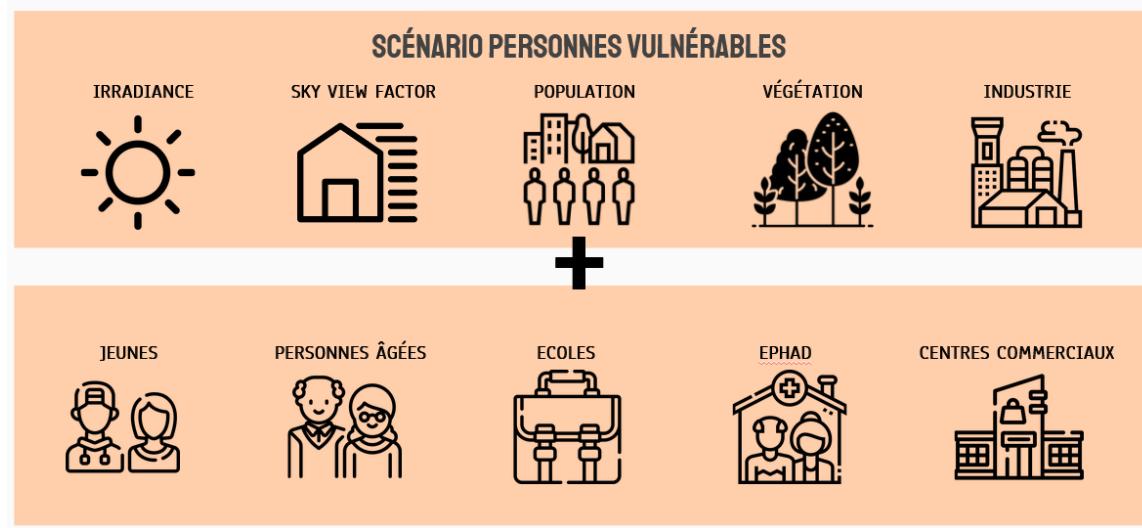


VILLE DE RENNES

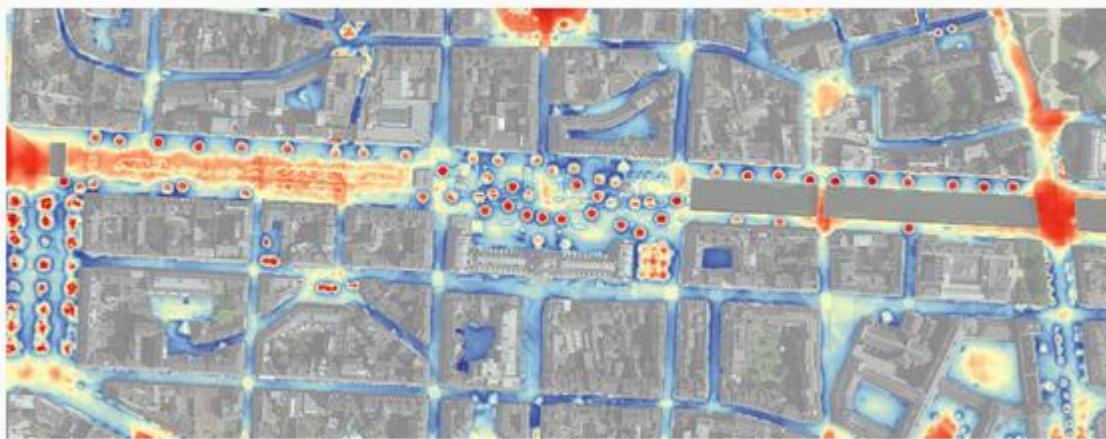
Détection des zones vulnérables aux fortes chaleurs et propositions d'une méthodologie d'identification des zones prioritaires à la plantation d'arbres.



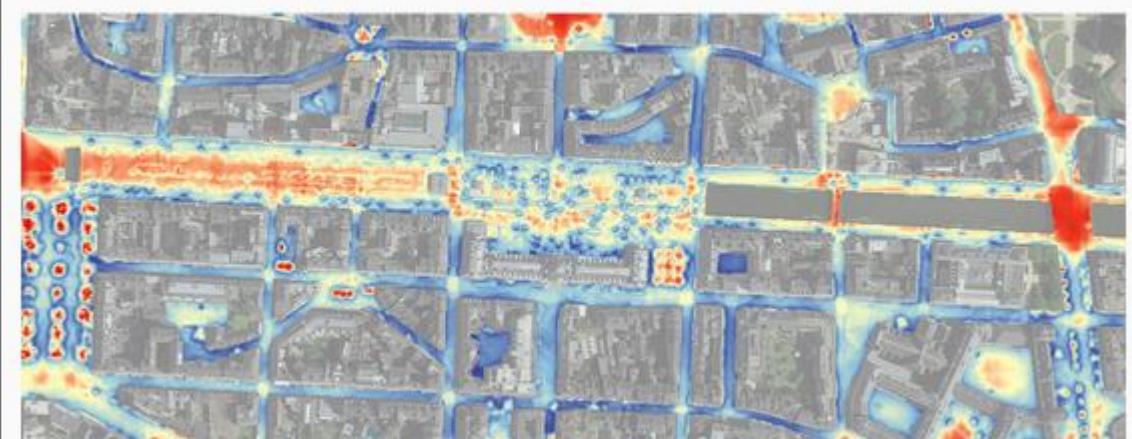
EMMANUELLE KERJOUAN, NINA BOUCHAIN, MATHIS BILLAUD, QUENTIN DELABARRE



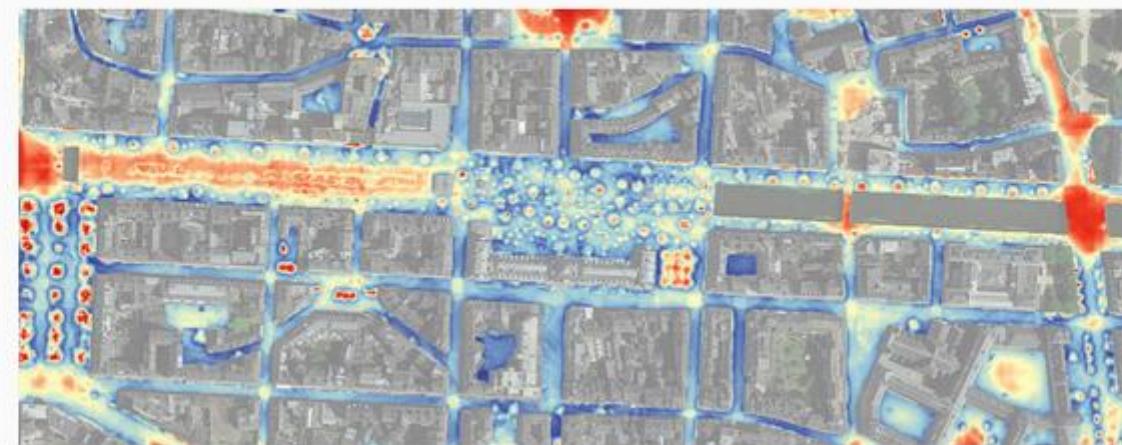
Scénario 1 : 50 arbres



Scénario 2 : 125 arbres



Scénario 3 : 200 arbres

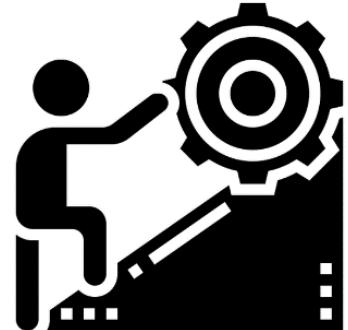




Le géomètre au centre de la SDG

- Les questions de départ
 - Objectifs, usages, contextes, utilisateurs
- Liens avec les problématiques des territoires
 - Aménagement, urbanisme, mobilité, environnement
- Réflexions sur les données
 - Sources, précision, exhaustivité, disponibilité, MAJ
- Réflexion sur les méthodes
 - Agrégation / Echelles / Sémiologie graphique / Proximité / Réseaux
- Le choix des outils
 - Logiciels SIG, langages de programmation, ETL, SGBD...
- L'automatisation comme levier
 - Automatisation et documentation de scripts (reproductibilité)
- La valorisation des analyses comme enjeu
 - Combiner cartes et dataviz pour donner à voir de multiples manières

Géomaticien & Data-scientist



- Le géomaticien expert des données géographiques
 - Spécialiste des données spatiales
 - Maîtrise des méthodes d'analyse spatiale
 - Maîtrise des questions de cartographie
 - Arrimage au problématiques territoriales
- Le Data-scientist expert du traitement de données
 - Spécialiste du traitement automatique de données massives
 - Expertise dans l'analyse statistique et la programmation
 - Maîtrise du datamining ou du machine learning
- De nouvelles complémentarités à développer
 - Le géomaticien n'est pas un data scientist
 - Le data scientist n'est pas un géomaticien

Discussion



- Quelle place pour la SDS dans vos structures ?
- Quels besoins méthodologiques ?
- Quels choix techniques ?
- Quelles compétences recherchées ?
- Quelle vision à moyen terme ?
- Une place pour le machine learning dans vos structures ?
- ...