

Introduction

Le présent document constitue une synthèse structurée des travaux pratiques réalisés dans le cadre du cours de **Statistiques Exploratoires Spatiales (SES)**. À l'intersection de la géographie numérique, de la science des données et de l'analyse territoriale, ce programme a permis d'explorer des problématiques de développement critiques à travers plusieurs contextes géographiques, notamment au Bénin, au Burundi, en République Démocratique du Congo et au Niger.

L'objectif de cette série de travaux était de mobiliser des outils de pointe — tels que Python, R, JavaScript et la plateforme Google Earth Engine — pour transformer des données géospatiales brutes en indicateurs décisionnels. Du diagnostic de l'accès aux services de santé à l'estimation des terres arables par imagerie satellitaire, chaque projet a été l'occasion d'appliquer des méthodes d'analyse spatiale rigoureuses (statistiques zonales, classification urbaine/rurale, calculs d'indices spectraux) pour répondre aux enjeux des Objectifs de Développement Durable (ODD).

TP2 : Analyse spatiale des infrastructures essentielles au Bénin

1. Contexte

Ce TP vise à analyser la répartition géographique des infrastructures essentielles au Bénin afin d'évaluer l'accessibilité des populations aux services de base.

L'étude mobilise des données géospatiales ouvertes pour examiner les infrastructures de santé, d'éducation, de transport ainsi que leur relation avec la distribution de la population. L'objectif principal est d'identifier les zones sous-desservies nécessitant des interventions prioritaires en matière d'aménagement du territoire.

2. Objectifs du TP

Les objectifs de ce projet étaient :

- Cartographier les infrastructures essentielles
- Analyser leur distribution spatiale
- Mesurer l'accessibilité des localités aux services
- Identifier les disparités territoriales entre départements
- Produire des visualisations statiques et interactives

3. Langage et environnement technique

Le langage de programmation qui a été utilisé pour faire ce TP est python. Pour le faire, nous avons travaillé de manière modulaire, c'est-à-dire, de manière à avoir plusieurs scripts différents qui exécutent des tâches différentes. Tous ces scripts seront reliés par un script principale “run_all.py” qui sera l'unique script à exécuter.

4. Sources des données

Nous avons besoin de trois types de données différentes pour faire notre travail :

- Données d'infrastructures

Le premier type de données correspond aux localisations des infrastructures au Bénin. Pour cela, nous avons utilisé OpenStreetMap. Nous avons travaillé avec des fichiers Shapefile (.shp) contenant les données suivantes : hôpitaux, cliniques, pharmacies, réseau routier, hydrographie, aires protégées.

- Données administratives

Pour travailler sur des zones spécifiques du pays, il nous fallait avoir les découpages administratifs sous plusieurs niveaux. Pour cela, nous utilisons geoBoundaries avec la découpe ADM0 (niveau national), ADM1 (12 départements) et ADM2 (77 communes).

- Données de population

Enfin, les données sur la population pour une année donnée était nécessaire. Nous avons choisi de considérer la découpe par pixel (1 km et 100m) proposé par WorldPop en 2024.

5. Traitements effectués

Le projet est structuré en plusieurs étapes :

- Le chargement des données
- Le prétraitement
- L'analyse de proximité (buffers de 100m, identification de zones couvertes...)
- Analyse descriptive

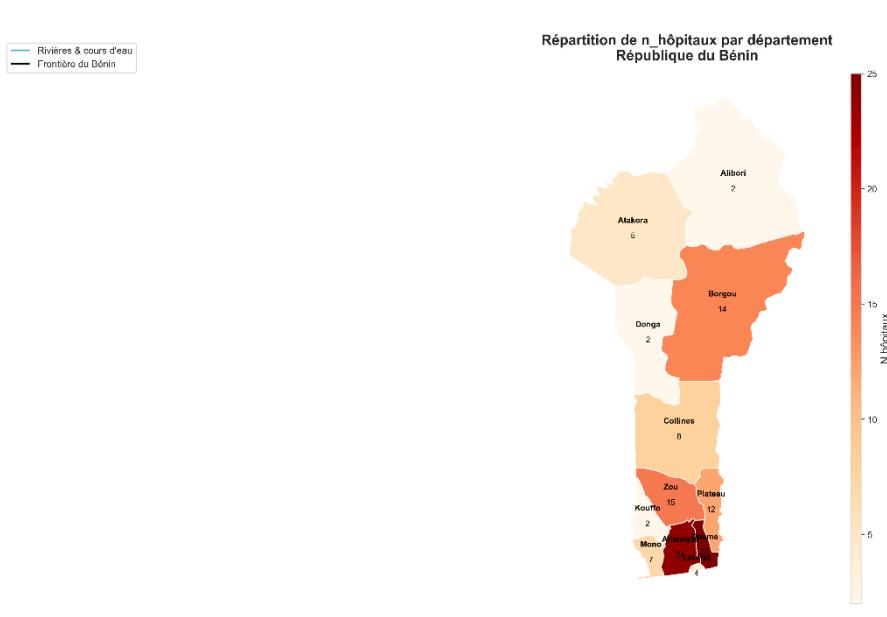
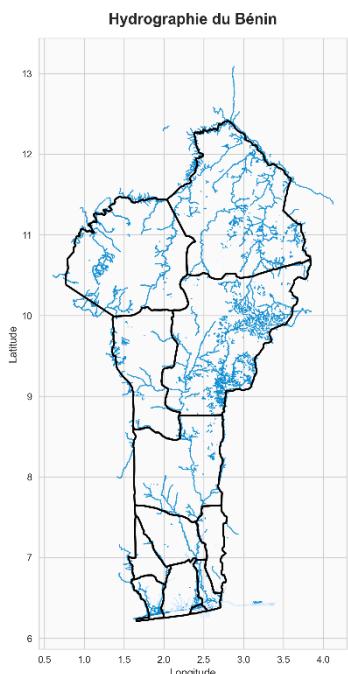
6. Résultats et outputs produits

Le projet a généré :

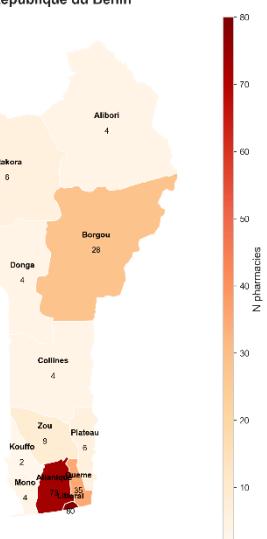
Fichiers statistiques

- Distances aux hôpitaux
- Distances aux cliniques
- Distances aux pharmacies
- Distances aux écoles
- Statistiques agrégées par département

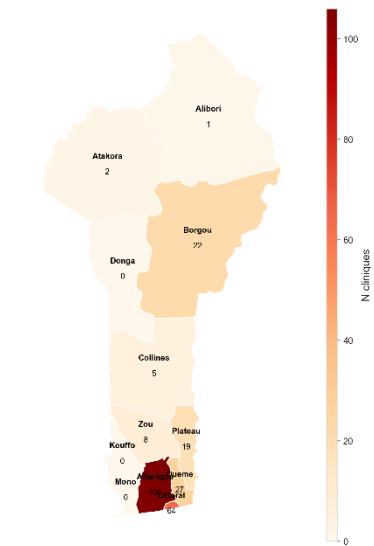
Cartes statiques



Répartition de n_pharmacies par département
République du Bénin



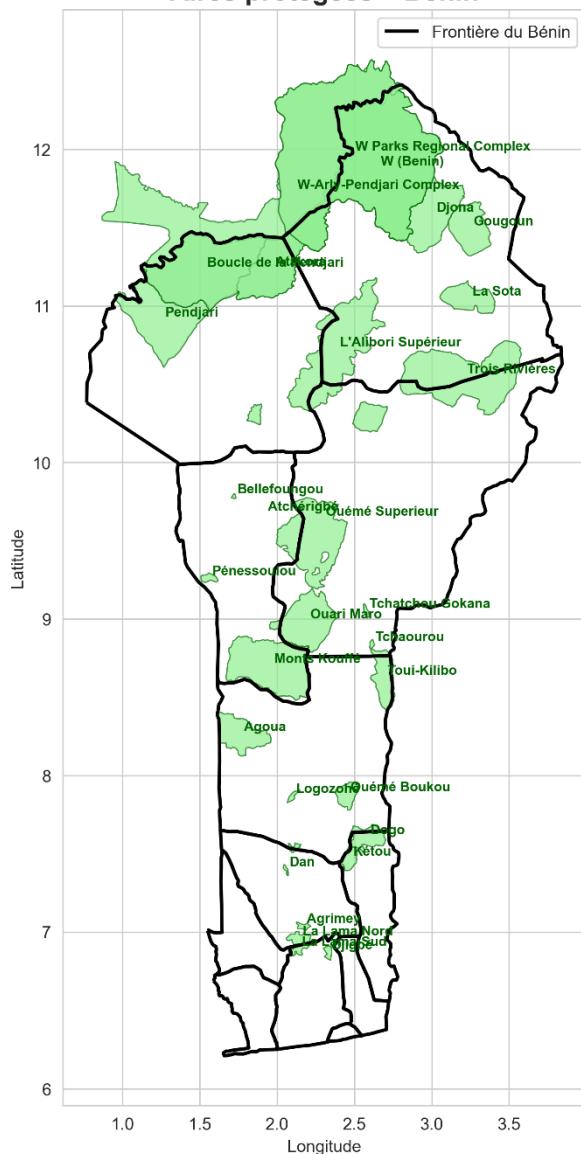
Répartition de n_cliniques par département
République du Bénin



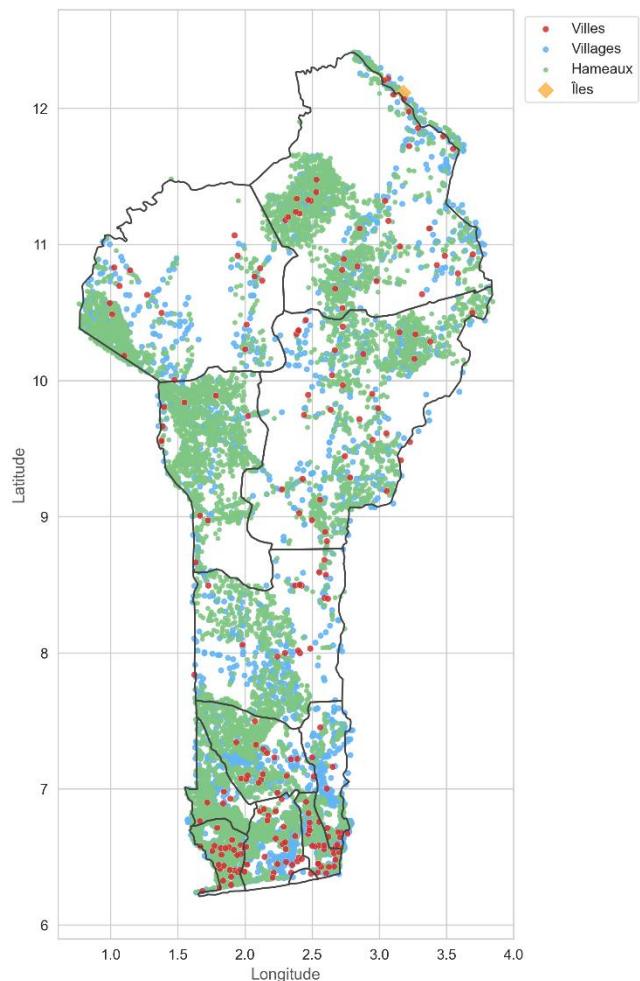
Source : OSM + geoBoundaries

Source : OSM + geoBoundaries

Aires protégées – Bénin



Localités – Bénin



Cartes interactives (HTML)

- Cartes zoomables avec clustering
- Popups informatifs
- Visualisation dynamique de l'accessibilité

8. Résultats principaux

Les analyses ont permis :

- D'identifier des disparités spatiales importantes entre départements
- De mettre en évidence des zones à faible accessibilité
- De quantifier les distances moyennes aux services essentiels
- De proposer des zones prioritaires pour de futures implantations

9. Compétences mobilisées

- Manipulation de données vectorielles et raster
- Programmation Python
- Analyse d'accessibilité spatiale
- Agrégation statistique par unité administrative
- Production de cartes statiques et interactives
- Structuration d'un projet d'analyse spatiale modulaire

TP3 : Analyse géospatiale de la couverture sanitaire, des zones protégées et des corridors de transport au Bénin

1. Contexte

Ce travail vise à analyser de manière intégrée plusieurs dimensions territoriales du Bénin :

- L'accessibilité aux infrastructures sanitaires
- La population desservie
- Les corridors de transport
- Les zones protégées
- La répartition spatiale de la population

L'analyse est centrée sur l'accessibilité, la population desservie, les zones d'influence, et la répartition spatiale.

2. Objectifs du TP

Les objectifs principaux étaient :

- Mesurer la population desservie par les hôpitaux et pharmacies
- Différencier l'accessibilité en milieu urbain et rural
- Construire des corridors de transport stratégiques
- Quantifier la population vivant dans les zones protégées
- Produire des cartes statiques et interactives
- Réaliser une analyse territoriale multicritères

3. Langage et environnement technique

Le langage utilisé était python dans l'environnement Jupyter Notebook.

4. Sources des données

a) Données administratives

- **Source :** geoBoundaries
- Limites nationales et départementales du Bénin

b) Données d'infrastructures

- **Source :** OpenStreetMap
- Infrastructures sanitaires (hôpitaux, pharmacies)
- Réseau routier
- Réseau ferroviaire
- Localités (village, town, hamlet...)
- Aires protégées

c) Données de population

- **Source :** WorldPop
- Année : 2024
- Résolution : 100 m et 1 km

5. Méthodologie et traitements effectués

5.1 Chargement et harmonisation

- Importation des données vectorielles et raster

5.2 Classification urbain / rural

- Extraction des types de localités OSM
- Calcul de densité dans un rayon de 2 km
- Attribution d'un statut urbain ou rural aux infrastructures

Cette étape permet d'introduire une différenciation territoriale pertinente dans l'analyse d'accès.

5.3 Buffers différenciés selon le milieu

Infrastructure Urbain Rural

Hôpital 5 km 15 km

Pharmacie 2 km 7 km

Les buffers ont été construits en projection métrique (UTM), puis reprojetés pour extraction de population.

5.4 Estimation de la population desservie

Pour chaque infrastructure :

- Construction du buffer
- Extraction de la population via statistiques zonales
- Agrégation des résultats par zone (urbain/rural)

Cette étape permet de quantifier concrètement la couverture sanitaire.

5.5 Construction des corridors de transport

Un corridor stratégique a été défini à partir :

- Routes principales (trunk, primary, secondary)
- Voie ferrée

Étapes :

- Filtrage des infrastructures de transport
- Buffer de 2 km
- Fusion routes + rails
- Extraction :
 - Population dans le corridor
 - Localités traversées

5.6 Analyse des zones protégées

- Calcul de la superficie protégée par département
- Extraction de la population vivant dans les aires protégées
- Cartographie choroplète par pourcentage de population protégée

6. Méthodes d'analyse spatiale mobilisées

- Analyse par buffers multiples
- Analyse différenciée urbain / rural
- Statistiques zonales raster–vecteur
- Agrégation spatiale départementale
- Analyse de corridors
- Cartographie thématique et interactive

Ce TP mobilise une approche plus avancée d'analyse spatiale comparativement au TP2, notamment par l'intégration raster–vecteur.

7. Résultats et outputs produits

Résultats analytiques

- Population desservie par hôpital
- Population desservie par pharmacie
- Population dans les corridors de transport
- Population vivant en zone protégée
- Nombre de localités dans les corridors

Cartes statiques

- Buffers sanitaires
- Zones protégées
- Corridors de transport
- Choroplèthes départementales

Cartes interactives

- Buffers dynamiques
- Affichage de population au survol

- Différenciation urbain / rural
- Superposition des corridors

8. Résultats principaux

L’analyse a permis :

- D’identifier des écarts importants de couverture sanitaire entre zones urbaines et rurales
- De quantifier précisément la population desservie
- De mettre en évidence l’importance stratégique des corridors de transport
- D’évaluer l’impact spatial des zones protégées sur la population

9. Limites du projet

- Dépendance à la qualité des données OpenStreetMap
- Hypothèses simplificatrices pour la classification urbain/rural
- Résolution raster limitée à 100 m
- Données ferroviaires peu détaillées

10. Compétences mobilisées

- Intégration données raster–vecteur
- Analyse multi-échelle
- Construction de buffers différenciés
- Analyse de corridors
- Statistiques zonales
- Production de cartes interactives avancées
- Structuration d’un projet géospatial complet

TP4 : Identification des terres arables au Burundi

1. Contexte

Ce travail pratique s'inscrit dans le cadre du cours de **Statistiques Exploratoires Spatiales** et porte sur l'identification des terres arables au Burundi.

L'objectif est d'estimer la superficie réellement exploitable pour l'agriculture en combinant des données satellitaires, topographiques et administratives.

L'approche repose sur :

- La construction d'une **base arable potentielle**
- L'application de **masques d'exclusion physiques et légaux**
- L'agrégation statistique aux niveaux administratifs

Le projet est entièrement réalisé sur la plateforme Google Earth Engine (GEE) en JavaScript.

2. Objectifs du TP

Les objectifs principaux étaient :

- Identifier les terres potentiellement cultivables
- Exclure les zones non exploitables (pentes, eaux, zones protégées, surfaces imperméables)
- Estimer la superficie arable totale du pays
- Calculer les ratios arables par province et commune
- Produire une interface cartographique interactive
- Générer des classements statistiques territoriaux

3. Langage et environnement technique

Le langage utilisé est JavaScript (Google Earth Engine) associé à Google drive pour l'exportation.

4. Sources des données

Le projet mobilise plusieurs bases de données internationales intégrées dans GEE :

Source Rôle

a) Limites administratives

- **Source :** GADM
- Niveaux : pays, provinces (17 en 2015), communes (133 en 2015)

b) Terres cultivées

- **Source :** GFSAD
- Identification des terres agricoles existantes (2015)

c) Déforestation

- **Source :** Hansen Global Forest Change
- Forêts perdues entre 2000 et 2015

d) Zones protégées

- **Source :** World Database on Protected Areas
- Masque d'exclusion légal

e) Eaux permanentes

- **Source :** JRC Global Surface Water
- Masque d'exclusion physique (occurrence > 75%)

f) Topographie

- **Source :** Shuttle Radar Topography Mission (SRTM)
- Exclusion des pentes > 15°

g) Surfaces imperméables

- **Source :** Global Man-made Impervious Surface (GMIS)
- Exclusion zones bâties > 10%

5. Méthodologie et traitements effectués

5.1 Pré-traitement des données raster

- Conversion des couches en masques binaires
- Nettoyage des valeurs NoData
- Harmonisation des projections

5.2 Construction de la base arable potentielle

Base arable =

- Terres cultivées (GFSAD 2015)
- OU forêts récemment déboisées (Hansen GFC)

Cette union permet d'intégrer les terres agricoles actuelles et les zones récemment convertibles.

5.3 Construction des masques d'exclusion

Zone d'exclusion totale = union de :

- Eaux permanentes
- Pentes > 15°
- Zones protégées
- Surfaces imperméables

5.4 Détermination des terres arables finales

Terres arables finales =

Base arable – Zone d'exclusion totale

Cette opération permet d'identifier uniquement les surfaces :

- Potentiellement cultivables
- Non soumises à contrainte physique
- Non soumises à restriction légale

5.5 Agrégation statistique

Calcul pour chaque unité administrative :

- Superficie totale (km²)
- Superficie arable (km²)
- Ratio arable (%)

Niveaux :

- Provinces (17 en 2015)
- Communes (133 en 2015)

6. Visualisation et interactivité

6.1 Cartographie interactive

Interface GEE permettant :

- Activation/désactivation des couches
- Affichage dynamique des statistiques
- Consultation par clic :
 - Données communales
 - Données provinciales

6.2 Graphiques générés

Classements automatiques :

Par province :

- Top 5 superficie arable
- Top 5 ratio arable
- Top 5 superficie totale

Par commune :

- Top 5 superficie arable
- Top 5 ratio arable
- Top 5 superficie totale

6.3 Exports produits

- Stats_Communes.csv
- Stats_Provinces.csv

Ces exports permettent une exploitation ultérieure sous Excel, QGIS, R ou Python.

7. Résultats principaux

L'analyse a permis :

- D'estimer la superficie arable nationale
- D'identifier les provinces à fort potentiel agricole
- De comparer les performances territoriales
- De produire une cartographie décisionnelle interactive
- De quantifier précisément le ratio arable par unité administrative

8. Limites du projet

- Données administratives basées sur la configuration 2015
- Réorganisation territoriale en 2025 (5 provinces, 42 communes)
- Résolution des données raster
- Dépendance à la qualité des bases internationales

9. Compétences mobilisées

- Traitement raster massif dans GEE
- Construction de masques spatiaux
- Analyse multicritère spatiale
- Logique booléenne appliquée aux données géographiques
- Agrégation statistique administrative
- Production d'indicateurs territoriaux
- Développement d'interface cartographique interactive

TP 5 : Analyse de l'Efficacité de l'Urbanisation en RDC (ODD 11.3.1)

1. Contexte et Objectifs

Ce projet s'inscrit dans le cadre du suivi des **Objectifs de Développement Durable (ODD)** des Nations Unies, spécifiquement l'indicateur **11.3.1**. L'étude porte sur la République Démocratique du Congo (RDC) sur la période **2017-2020**.

L'objectif principal est de mesurer le ratio entre le taux de consommation des terres (également urbain) et le taux de croissance démographique afin d'évaluer si le développement des villes congolaises est "efficace" (densification) ou s'il dérive vers un étalement incontrôlé.

2. Langage et Technologies Utilisés

Le projet a été entièrement réalisé avec le langage **R**, en s'appuyant sur un écosystème de packages spécialisés pour la donnée spatiale :

- **terra** : Pour le traitement lourd des données raster (population et occupation du sol).
- **sf** : Pour la manipulation des couches vectorielles (limites administratives et urbaines).
- **exactextractr** : Pour réaliser des statistiques zonales précises croisant pixels et polygones.
- **tidyverse & ggplot2** : Pour la manipulation de données tabulaires et la production de visualisations.

3. Sources des Données

Le projet mobilise des sources de données globales de référence :

- **Démographie** : Données **WorldPop** (résolution 1 km) pour les années 2017 et 2020.
- **Occupation du sol** : **ESA WorldCover** (résolution 10 m) pour l'identification des zones bâties.
- **Limites administratives** : Base de données **GADM** pour les frontières nationales et provinciales de la RDC.

4. Traitements Effectués

La méthodologie suit rigoureusement la norme **DEGURBA** (Degree of Urbanisation) :

- **Classification spatiale** : Identification des cellules à haute densité (≥ 1500 hab/km 2) et regroupement en grappes urbaines selon les seuils de population (Centres urbains $\geq 50,000$ hab).
- **Analyse d'occupation du sol** : Mosaïquage et ré-échantillonnage des tuiles ESA pour extraire les surfaces bâties par zone urbaine.
- **Calculs statistiques** :
 - **LCR (Land Consumption Rate)** : Taux de consommation des terres.
 - **PGR (Population Growth Rate)** : Taux de croissance de la population.

- **LCRPGR** : Calcul du ratio final pour déterminer l'indice d'étalement.

5. Outputs (Résultats)

- **Données chiffrées** : Tableaux (CSV/Excel) récapitulant la population et les surfaces bâties par ville.
- **Cartographie** : Cartes de densité de population et cartes comparatives de l'évolution urbaine entre 2017 et 2020.
- **Indicateurs ODD** : Valeurs de l'indicateur 11.3.1 permettant de classer les villes selon leur mode de croissance (efficace, proportionnel ou inefficace).

TP 6 : Analyse Multi-Indices et Synergie Socio-Environnementale (Niger)

1. Contexte et Problématique

Ce projet s'inscrit dans une démarche de **statistique spatiale appliquée au développement**. L'enjeu est de pallier le manque de données environnementales localisées en utilisant l'imagerie satellitaire pour enrichir les enquêtes ménages traditionnelles. Au Niger, pays sahélien fortement dépendant des conditions climatiques, l'objectif est de caractériser précisément l'état de la biomasse et les stress hydriques durant la saison de croissance 2022 (juin à septembre) afin d'analyser leur impact sur le bien-être des populations mesuré par l'EHCVM.

2. Stack Technique & Écosystème

Le projet repose sur une architecture de traitement hybride performante :

- **Google Earth Engine (GEE)** : Utilisation de l'API JavaScript pour le calcul distribué sur des pétaoctets de données satellitaires sans téléchargement local.
- **Capteurs Optiques (Sentinel-2 L2A)** : Exploitation des bandes spectrales de 10m à 20m de résolution pour la précision fine.
- **Capteurs Thermiques (MODIS MOD11A2)** : Utilisation du produit de température de surface (LST) à 1km de résolution.
- **Environnement Statistique** : R, Python ou Stata pour le nettoyage des micro-données de l'enquête EHCVM et la réalisation des jointures finales.

3. Sources et Nature des Données

Le projet réalise une fusion entre données de télédétection et données administratives/sociales :

- **Imagerie** : Collections COPERNICUS/S2_SR_HARMONIZED (Sentinel-2) et MODIS/061/MOD11A2.
- **Vecteurs** : Limites administratives hiérarchisées (ADM0 à ADM2) provenant de GADM ou FAO GAUL.
- **Socio-économie** : Micro-données de l'**Enquête Harmonisée sur le Bien-être des Ménages (EHCVM) 2021**, converties du format .dta (Stata) vers le .csv pour l'interopérabilité.

4. Méthodologie et Traitements Avancés

La rigueur du traitement garantit la qualité des indicateurs produits :

- **Prétraitement & Filtrage** : Application d'un masque de nuages (seuil < 20%) et calcul d'un composite médian pour éliminer les artefacts atmosphériques sur la période de juin à septembre.
- **Ingénierie des Indices Spectraux** :

- **Végétation** : Calcul du **NDVI** (vigueur de la végétation) et de l'**EVI** (plus robuste aux effets atmosphériques et aux sols nus).
- **Eau & Humidité** : Calcul du **NDWI** (détection des plans d'eau) et du **NDMI** (stress hydrique des plantes via le SWIR).
- **Indice MDVI** : Calcul d'un indice spécifique supplémentaire pour affiner l'analyse de la biomasse.
- **Analyse Thermique** : Conversion des valeurs numériques MODIS en degrés Celsius pour obtenir une cartographie réelle de la chaleur au sol.
- **Agrégation Spatiale (Zonal Statistics)** : Utilisation de réducteurs GEE (`ee.Reducer.mean()`) pour transformer les pixels en moyennes statistiques par région (8) et département (63).

5. Outputs et Produits Livrables

Le projet génère une suite complète de données prêtes à l'emploi :

- **Rasters Thématiques** : 14 fichiers GeoTIFF haute résolution couvrant l'ensemble des indices calculés et les bandes spectrales brutes (B2, B3, B4, B8, B11).
- **Datasets Alphanumériques** : Tableaux de bord départementaux et régionaux contenant les moyennes environnementales par unité administrative.
- **Template de Fusion EHCVM** : Un fichier pivot (`Niger_Template_EHCVM_2022.csv`) conçu spécifiquement pour servir de clé de jointure avec les données de l'enquête ménage.
- **Application Interactive** : Un script GEE permettant la visualisation dynamique des indices sur l'ensemble du territoire nigérien.

Conclusion

En conclusion, l'ensemble de ces travaux pratiques a permis de consolider une expertise complète en analyse spatiale et en télédétection appliquée. La diversité des outils utilisés a démontré l'importance de l'interopérabilité technologique : la puissance du calcul distribué avec GEE pour le traitement de données massives (raster), la flexibilité de **python** pour la structuration modulaire de projets, et la précision statistique de **R** pour l'analyse des dynamiques urbaines.

Au-delà de la maîtrise technique, ce parcours a mis en lumière la valeur ajoutée de la donnée géographique pour pallier le manque de données traditionnelles dans les pays du Sud. En fusionnant des micro-données d'enquêtes ménages avec des indices biophysiques (NDVI, LST, NDMI), nous avons pu produire des analyses multicritères fines, capables d'orienter les politiques publiques en matière d'aménagement du territoire, de sécurité alimentaire et de couverture sanitaire. Ces compétences constituent aujourd'hui un levier essentiel pour une gestion territoriale fondée sur les preuves et la donnée.