

RÉPUBLIQUE DU SÉNÉGAL

Un peuple- un But- une Foi



Agence nationale de la Statistique et de la Démographie



ANSD
Agence Nationale de
la Statistique et de la Démographie



Ecole nationale de la Statistique et de l'Analyse économique Pierre Ndiaye

Cours de Statistiques exploratoires spatiales

Pays : Burkina Faso

Réalisé par :

Alioune Abdou Salam KANE

Khadidiatou DIAKHATE

Ange Emilson Rayan RAHERINASOLO

Awa DIAW

Sous la supervision de :

M. Aboubacre HEMA

ISE1-Cycle long – Novembre 2024

Table des matières

Introduction	1
Résumé des Travaux Pratiques (TP)	1
Logiciel R.....	1
TP1	1
TP2	1
GEE with Python.....	2
TP3	2
TP4	2
Python.....	3
TP5	3
TP6	3
TP7	3
Autre.....	3
TP8	3
TP9	4
Conclusion.....	4

Introduction

Ce cours a pour objectif de présenter la statistique descriptive spatiale, la cartographie avec R, entre autres. Il s'est déroulé sous une forme très interactive et pratique. Les séances étaient principalement organisées autour de travaux pratiques (TP), réalisés en groupes de 4 étudiants, afin de :

- favoriser le travail collaboratif, où chaque membre du groupe pouvait apporter une contribution unique selon ses forces (analyse, présentation, gestion des données).
- approfondir les compétences techniques, en manipulant directement des données spatiales issues de sources réelles ou simulées.
- préparer des présentations en vue de restituer les résultats des analyses, un exercice qui combinait pédagogie et professionnalisme.

Les travaux pratiques ont été réalisés en utilisant des données provenant de quatre pays : le Cameroun, le Sénégal, le Burkina Faso et Madagascar. Cinq logiciels ont été mobilisés au cours de ces séances. Ce document présente un résumé des différentes activités réalisées dans le cadre des TP.

Résumé des Travaux Pratiques (TP)

Logiciel R

TP1

Dans ce TP, il s'agissait d'importer des shapefiles (données de types vecteurs) et des rasters à travers les différents logiciels. Par la suite la visualisation devrait être améliorée.

Les bases de données utilisées sont : `bfa_admbnda_adm{0, 1, 2, 3}_igb_20200323.shp`

Le pays ou `adm0` est subdivisé en régions administratives et territoriales (`adm1`), elles-mêmes divisées administrativement en provinces ou en départements (`adm2`), et territorialement en communes urbaines et rurales (`adm3`).

Le TP nous a permis de nous souvenir de l'exposé que nous avons suivi en ISEP2 sur le traitement de données géospatiales.

Ce TP a également été l'occasion de découvrir un nouveau package sur R « `cowplot` » qui permet d'extraire la légende et d'afficher la carte sans légende.

Avec les autres packages que nous connaissions déjà, nous avons pu améliorer la visualisation de nos vecteurs et de rasters en ajoutant des éléments de carte comme la légende, la flèche du nord, l'échelle, et les labels. Spécifiquement pour les rasters nous avons utilisé une carte interactive.

Ce TP a jeté les bases pour une compréhension des outils d'analyse géospatiale car nous y avons revu les notions de bases comme centroïdes, raster, vecteur, latitude, longitude, etc.

TP2

L'objectif était d'utiliser les techniques de manipulation et de visualisation spatiale et de calculer des statistiques descriptives. Le TP était axé autour de deux analyses. Une première

pour les données vectorielles (les mêmes que pour le TP1) et la seconde pour les rasters (Abson-dev Statistique-Exploratoire-Spatiale main TP2-data_Malaria_Burkina.tiff)

- Pour les vecteurs

Concernant les statistiques, le nombre de géométries par niveau administratif a été calculé avec `nrow()`. Suite à la remarque d'une camarade, il a fallu améliorer ce code en supprimant les doublons avec `nrow(distinct())`.

La projection et le système de référence de coordonnées (CRS) ont été identifiés avec `st_crs()`. L'étendue est calculée avec `st_bbox()`, les centroides avec `st_centroid()`, l'aire avec `st_area()` et le périmètre avec `st_length()`.

- Pour les rasters

Nous avons chargé les rasters sous forme de stack selon « `rasters <- stack(list.files())` », et chacune des 20 images a été visualisée individuellement.

Nous avons ensuite calculé quelques statistiques selon « `statistique (calc(rasters, fun = la statistique, na.rm = TRUE))` »

Le professeur a souligné l'importance de rester prudent face aux codes, d'effectuer des recherches sur les codes sources pour assurer la pertinence des analyses (l'écart-type biaisé ou non par exemple).

GEE with Python

TP3

Dans ce TP, il s'agissait de calculer le taux moyen de paludisme dans le pays par niveau administratif en utilisant les données du TP2 et d'exporter les résultats sur un fichier .csv.

Nous passons désormais de R à Google Earth Engine (GEE) pour les analyses spatiales.

Le dernier groupe avait présenté une introduction sur Google Earth Engine (GEE) ainsi que ses principales fonctionnalités. Grâce à leur assistance, la transition s'est déroulée de manière fluide, notamment pour établir la connexion à GEE. Cela a impliqué la création d'un compte Google associé à Earth Engine, suivie des étapes de création et d'activation d'un environnement (conda), de l'installation, de la configuration, de l'authentification via `ee.Authenticate()`, ainsi que de l'initialisation avec `ee.Initialize()`.

Les principaux challenges rencontrés dans cette transition ont été liés à la découverte d'un outil nouveau. Google Earth Engine with Python, bien que puissant, nécessitait une adaptation rapide, car nous étions déjà familiers avec R, un logiciel que nous trouvions plus convivial et intuitif.

TP4

Dans ce TP, il s'agissait de calculer les moyennes et les écarts types du taux de malaria entre 2000 et 2022, à partir du fichier raster `202406_Global_Pf_Parasite_Rate_BFA_2000.tiff`, puis exporter les résultats au format csv.

Après quoi il fallait représenter la situation de chaque pixel pour les années 2020, 2021, et 2022 en utilisant une classification en trois catégories.

Le principal challenge rencontré résidait dans l'affichage de la légende avec les trois couleurs. Pour contourner ce problème, nous avons eu recours à Html.

Python

TP5

Ce TP a été le plus challengeant pour le groupe Burkina.

L'utilisation de nombreuses fonctions pour différentes étapes a rendu le processus parfois lourd et compliqué. Pour la visualisation du raster de population, l'argument interpolation de la commande « `plt.imshow(pop_data, cmap="viridis", interpolation="none")` » s'est avéré salvateur. Grâce à cette ligne, il a été possible de visualiser le raster, ce qui n'était pas possible au départ car le graphique apparaissait entièrement en une seule couleur.

Un autre point important était également la constante 111 320. Elle permet de convertir les tailles des pixels exprimées en degrés en mètres suivant : $1\text{ degré} = 111\,320\text{m}$. Cette constante varie légèrement en fonction de notre position par rapport à l'équateur.

Lors des présentations, le prof avait beaucoup insisté sur la vérification de nos outputs. Pour ce TP5, il avait recommandé de vérifier que la somme des parties correspond bien au tout. **TOUJOURS VÉRIFIER.**

TP6

Le fichier `Points_data.csv` a servi de base pour cet exercice, qui portait principalement sur la rasterisation.

Une nouvelle bibliothèque a été utilisée : `shapely`. Cette bibliothèque a permis de créer et de manipuler des objets géométriques, notamment des points.

Dès le départ, seuls les points relatifs au Burkina Faso ont été sélectionnés.

Le nombre de points a été compté à différents niveaux administratifs et conformément aux recommandations du professeur, nous avons systématiquement vérifié nos résultats.

Le TP a été enrichi avec deux nouvelles tâches le redimensionnement de la taille des pixels (passage de la résolution de 100 m à 5 km) et la représentation graphique (évolution du nombre de points de 1997 à 2024).

TP7

Le TP 7 était consacré à l'exploration du package `sdmApp`. Il s'agira de parcourir le package, ses objectifs, ses fonctions, les types de données à utiliser, les résultats, ses limites, etc.

Autre

TP8

Il s'agit de faire un résumé des TPs que nous avons réalisés.

TP9

Il s'agit de rechercher des bases de données sur l'Agriculture, le changement climatique, la santé et l'éducation.

Conclusion

Ce cours a permis d'explorer des bases de données variées et d'améliorer nos compétences en codage, avec un fort accent sur la compréhension des scripts et une diversité d'outils comme GEE, Python et R. Le travail en groupe a favorisé la collaboration, tandis que les cas pratiques et les présentations ont renforcé l'apprentissage actif et l'autonomie. Cette expérience a nous a incité à éviter une approche mécanique, d'être des « apprentis sorciers », pour reprendre les termes du Professeur. Bien que nous n'ayons utilisé que trois logiciels sur les cinq prévus, les supports pédagogiques mis à disposition, notamment sur GitHub nous ont été d'une grande aide et nous permettront de réviser de temps à autre.