

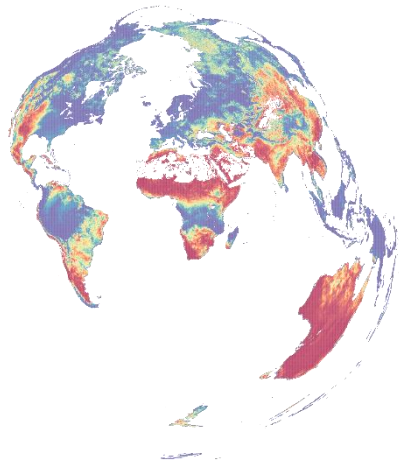
ANALYSE GÉOSPATIALE AVEC R : DE ZERO À HÉROS

Dr. Yélognissè Frédi Agbohessou

07/08/2024

Le matériel du cours est disponible à l'adresse suivante

<https://fredi-agbohessou.github.io/courses/course1/geospatialR.rar>



QUI SUIS-JE?

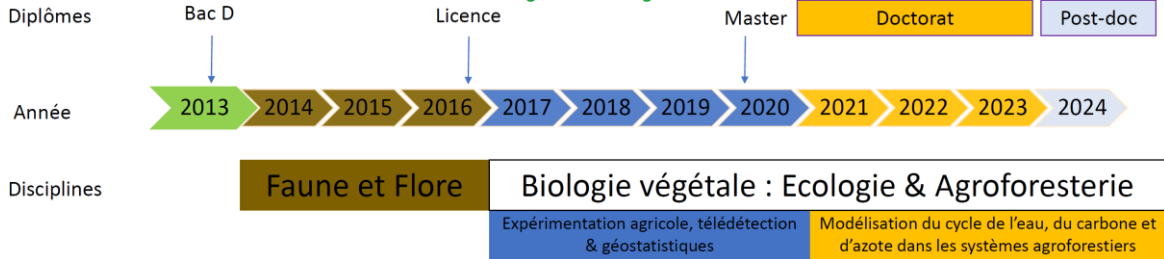


fredi.agbohessou@cirad.fr

@yagbohessou

<https://fredi-agbohessou.github.io>

Dr. Yélognissè Frédi Agbohessou



- Master : Effets du *Faidherbia albida* (Del.) A. Chev. sur la productivité aérienne et souterraine de l'arachide (*Arachis hypogaea* L.) : estimation du rendement et du Land-Equivalent-Ratio partiel de l'arachide (LER_cp) par drone
- Doctorat : Modélisation du bilan de gaz à effet de serre des systèmes agroforestiers du sahel : de la parcelle au paysage
- Post-doctorat : Modélisation de l'impact des stress climatiques sur le fonctionnement des hévéas (*Hevea brasiliensis*) en Asie du sud est.
- Plusieurs années d'expérience dans la modélisation mécaniste appliquée à des données réelles
- Plusieurs publications d'article en comité de lecture dans des revues internationales
- Depuis 2020 : 9 formations certifiantes sur les sciences de données, la programmation (R, Python, Julia) et la télédétection (Udemy & datacamp)

PLAN

- CHAPITRE 0 : Préparation et Pré-requis
- CHAPITRE 1 : Rappel sur les systèmes d'information géographique (SIG) et les données géospatiales
- CHAPITRE 2 : Manipulations des données Vectorielles
- CHAPITRE 3 : Manipulations des données raster
- CHAPITRE BONUS A : Interpolation des données raster à partir d'observations ponctuelles
- CHAPITRE BONUS B : Comment gérer un fichier NetCDF ?
- CHAPITRE BONUS C : Manipulations Avancées des Données Vectorielles

CHAPITRE 0 : PRÉPARATION ET PRÉ-REQUIS

QUEL EST L'OBJET DE CE COURS?

- Apprendre à utiliser le langage R pour extraire, manipuler, visualiser et analyser des données géospatiales issues des bases de données environnementales mondiales telles que ERA5, MODIS...

POURQUOI CE COURS ?

- **Compréhension des Écosystèmes** : Permet d'analyser et de visualiser les données spatiales pour mieux comprendre la distribution et les interactions des espèces dans différents écosystèmes.
- **Gestion des Données** : Apprend à manipuler des bases de données géospatiales, cruciales pour la gestion des ressources naturelles et la conservation de la biodiversité.
- **Modélisation Spatiale** : Facilite la création de modèles spatiaux pour prédire les changements environnementaux, tels que la répartition future des espèces sous l'effet du changement climatique.
- **Analyse de la Fragmentation de l'Habitat** : Aide à évaluer l'impact des activités humaines sur les habitats naturels, permettant de prendre des décisions éclairées pour la conservation.

POURQUOI CE COURS ?

- **Suivi et Surveillance** : Utilise des techniques géospatiales pour surveiller les changements dans l'utilisation des terres, la déforestation et autres indicateurs environnementaux.
- **Outils Statistiques Avancés** : Offre des compétences en analyse statistique avancée, indispensables pour interpréter des données écologiques complexes.
- **Interopérabilité** : R est compatible avec divers formats de données géospatiales et peut être intégré avec d'autres outils et logiciels, élargissant ainsi les possibilités d'analyse.
- **Accessibilité et coût, communauté et ressources, compétitivité sur le marché du travail...**

QU'EST-CE QUE CE COURS VOUS APPORTERA ?

- Comprendre les fondamentaux des données raster et vectorielles.
- Apprendre à importer, manipuler et exporter des données géospatiales dans R.
- Réaliser des opérations spatiales telles que les superpositions, les intersections, le buffering et le découpage.
- Analyser et visualiser des données géospatiales en utilisant les puissants packages de R.
- Appliquer des techniques d'analyse géospatiale à des problématiques réelles.

POURQUOI FAIRE UNE ANALYSE GÉOSPATIALE AVEC R AU LIEU DE QGIS ?

Critères	R	QGIS
Automatisation	Automatisation facile via des scripts reproductibles	Moins automatisé, nécessite souvent des actions manuelles
Capacités statistiques	Puissant pour les analyses statistiques avancées	Moins adapté aux analyses statistiques complexes
Intégration des données	Intégration facile de différents types de données	Principalement centré sur les données géospatiales
Personnalisation des visualisations	Très personnalisable avec des outils comme ggplot2	Personnalisation possible, mais moins flexible que R
Interface utilisateur	Nécessite des compétences en codage (pas d'interface graphique)	Interface graphique intuitive, facile à utiliser
Manipulation de grands volumes de données	Peut être lent avec de grandes données géospatiales	Optimisé pour les fichiers volumineux, rapide
Support de la communauté	Large communauté avec des ressources pour la reproductibilité scientifique	Large communauté et nombreux plugins
Édition des données	Moins adapté à l'édition de données géospatiales	Idéal pour l'édition et la gestion des couches géospatiales
Performance	Moins performant avec des données volumineuses	Performant pour le traitement rapide des données

PRÉ-REQUIS

- Une expérience préalable avec R est requise
- Une bonne connaissance des statistiques de base et de la visualisation des données est utile mais pas obligatoire

MATÉRIEL :

- R and Rstudio

CHAPITRE 1 : RAPPEL SUR LES SYSTÈMES D'INFORMATION GÉOGRAPHIQUE (SIG) ET LES DONNÉES GÉOSPATIALES

QU'EST-CE QU'UN SIG?

- Un système d'information géographique (S. I. G) est une technologie servant à créer, gérer, analyser et cartographier tout type de données (*esri.com*).
- C'est un outil qui collecte, stocke, traite et analyse des données géographiques. Il permet de visualiser ces données sous forme de couches superposées sur une carte géoréférencée, facilitant leur extraction, tri et traitement selon les besoins de l'utilisateur.
- Le SIG associe les données à une carte et intègre des données de localisation (endroit où se trouvent les choses) à toutes sortes d'informations descriptives (état des choses à cet endroit) (*esri.com*).

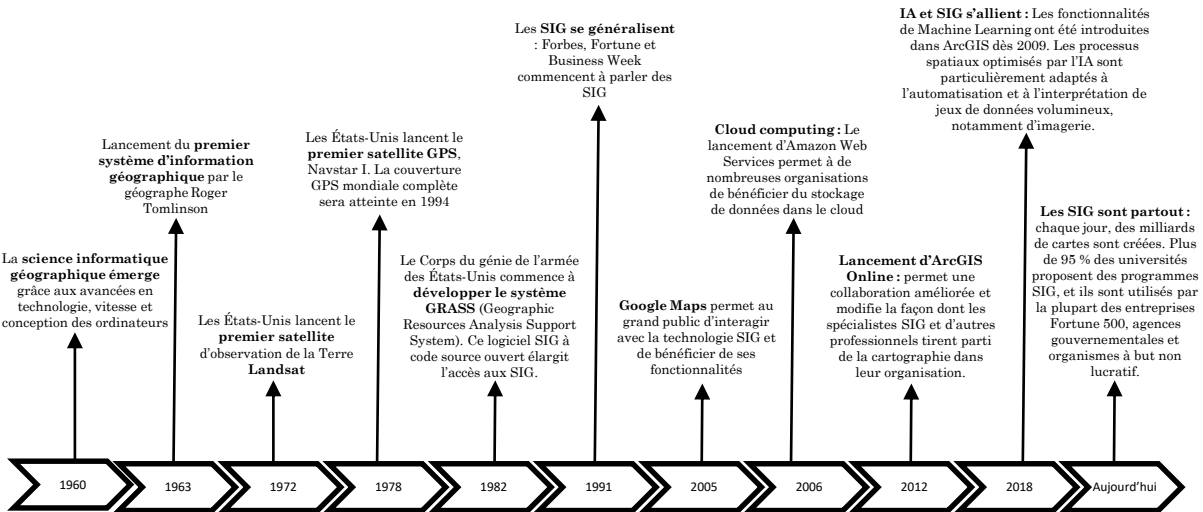


Exemples de SIG que vous connaissez déjà



R peut aussi être
utilisé comme un SIG

CHRONOLOGIE DE L'ÉVOLUTION DES SIG: QUELQUES DATES CLÉS



QUELLE EST LA FONCTION D'UN SIG ? PLUS QUE DU MAPPAGE

- **Gestion des données** : stockage, intégration et unification des données (même format)
- **Cartographie et visualisation** : clarification des données complexes, facilitation de l'analyse rapide des tendances et amélioration de la prise de décision en rendant l'information plus accessible.
- **Analyse spatiale** : générer de nouvelles informations
- **Communication** : communiquer plus rapidement des idées complexes

QUELLE EST LA FONCTION D'UN SIG ? PLUS QUE DU MAPPAGE

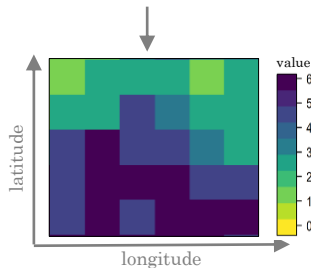
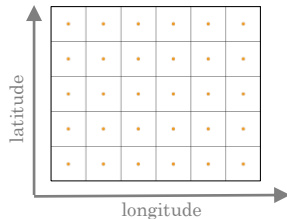
bovin	caprin	ovin	asins	camelins	equin	total
5435	7382	3881	0	0	62	16760
5466	5637	3335	0	0	6	14444
5623	6475	3551	0	0	38	15667
4856	5555	3059	0	0	6	13476
5075	5767	3062	0	0	6	13910
4765	5358	2972	0	0	4	13099
11430	10795	5658	0	0	53	27936
10888	10399	6107	0	0	63	27457
12035	10873	5873	0	0	71	28852
10225	10602	5708	0	0	70	26605
11922	11214	5638	0	0	64	28838
10061	11747	5174	0	0	96	27078
3176	3610	2230	0	0	13	9029
2549	3470	2378	0	0	12	8409
2632	3572	2290	0	0	11	8505
2807	3728	2228	0	0	12	8775
2479	3041	1986	0	0	19	7525
2298	3070	2015	0	0	15	7398
1758	3093	2038	0	0	15	6904
1499	2328	1671	0	0	9	5507

fichier au format xlsx

```

,"bovin","caprin","ovin","asins","camelins","equin",
5435,7382,3881,0,0,62,16760,0.32,0.44,0.23,0,0,0
1,5466,5637,3335,0,0,6,14444,0.38,0.39,0.23,0,0,0
2,5623,6475,3551,0,0,18,15667,0.36,0.41,0.23,0,0,0
3,4856,5555,3059,0,0,6,13476,0.36,0.41,0.23,0,0,0
4,5075,5767,3062,0,0,6,13910,0.36,0.41,0.22,0,0,0
5,4765,5358,2972,0,0,4,13099,0.36,0.41,0.23,0,0,0
6,11430,10795,5658,0,0,53,27936,0.41,0.39,0.2,0,0
7,10888,10399,6107,0,0,63,27457,0.4,0.38,0.22,0,0
8,12035,10873,5873,0,0,71,28852,0.42,0.38,0.2,0,0
9,10225,10602,5708,0,0,70,26605,0.38,0.4,0.21,0,
1,11922,11214,5638,0,0,64,28838,0.41,0.39,0.2,0,0
1,1,10061,11747,5174,0,0,96,27078,0.37,0.43,0.19,0
1,2,3176,3610,2230,0,0,13,9029,0.35,0.4,0.25,0,0,0
1,3,2549,3470,2378,0,0,12,8409,0.3,0.41,0.28,0,0,0
    
```

fichier au format csv ou txt

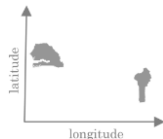
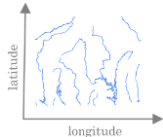
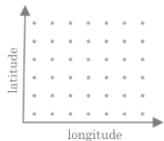


fichier raster au format tiff ou nc

PRINCIPALES COMPOSANTES DE LA STRUCTURE DES DONNÉES DANS UN SIG : **DONNÉES SPATIALES**

Données Vectorielles

- **Points** : Utilisés pour représenter des objets géographiques qui peuvent être localisés par une seule coordonnée (ex : bornes géodésiques, arbres).
- **Lignes** : Représentent des objets géographiques ayant une dimension linéaire (ex : routes, rivières).
- **Polygones** : Représentent des surfaces fermées (ex : parcelles de terrain, lacs).



PRINCIPALES COMPOSANTES DE LA STRUCTURE DES DONNÉES DANS UN SIG : **DONNÉES SPATIALES**

Données Raster

- **Images de Satellite** : Captures de la surface terrestre prises depuis l'espace.
- **Photographies Aériennes** : Images prises depuis des avions, souvent utilisées pour des analyses plus détaillées.

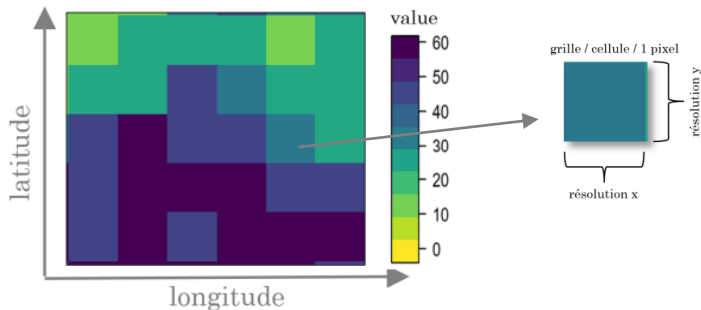


Image Google de Cotonou (Bénin) montrant le Palais des Congrès, la Place de l'Amazonie et la nouvelle Cité Ministérielle en construction.

PRINCIPALES COMPOSANTES DE LA STRUCTURE DES DONNÉES DANS UN SIG : **DONNÉES SPATIALES**

Données Raster

- **Image raster avec des informations géoréférencées**



La couleur renseigne sur la valeur (de la donnée) assignée au pixel (ici, environ 35).

Les données peuvent être « géolocalisées » en utilisant les valeurs de coordonnée x-y

PRINCIPALES COMPOSANTES DE LA STRUCTURE DES DONNÉES DANS UN SIG : **DONNÉES ATTRIBUTAIRES**

Caractéristiques des objets géographiques représentés par les données spatiales. Elles sont souvent stockées dans des tables de base de données et sont liées aux données spatiales par des identifiants uniques.

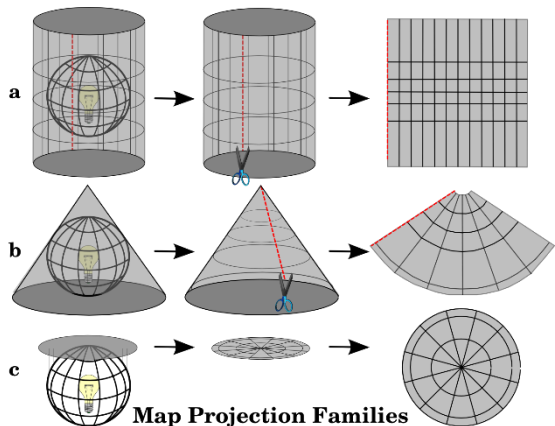
fid	id	left	top	right	bottom	glc_tot_px_count	_grassland_px_cov
33	33	-17,5301717	14,8	-17,4301717	14,7000000000...	144	1
34	34	-17,5301717	14,7	-17,4301717	14,6	144	0
82	82	-17,4301717	14,9	-17,330171699...	14,8	144	0
83	83	-17,4301717	14,8	-17,330171699...	14,7000000000...	144	0
84	84	-17,4301717	14,7	-17,330171699...	14,6	144	0
132	132	-17,3301717	14,9	-17,2301717	14,8	144	4
133	133	-17,3301717	14,8	-17,2301717	14,7000000000...	144	9
134	134	-17,3301717	14,7	-17,2301717	14,6	144	0
181	181	-17,2301717	15	-17,1301717	14,9	144	0
182	182	-17,2301717	14,9	-17,1301717	14,8	144	27
183	183	-17,2301717	14,8	-17,1301717	14,7000000000...	144	6
184	184	-17,2301717	14,7	-17,1301717	14,6	144	11
185	185	-17,2301717	14,6	-17,1301717	14,5	144	4
230	230	-17,130171700...	15,1	-17,0301717	15	144	0
231	231	-17,130171700...	15	-17,0301717	14,9	144	5
232	232	-17,130171700...	14,9	-17,0301717	14,8	144	19

PRINCIPALES COMPOSANTES DE LA STRUCTURE DES DONNÉES DANS UN SIG : **SYSTÈME DE RÉFÉRENCE SPATIALE**

Un SIG utilise des systèmes de référence de coordonnées pour projeter des emplacements terrestres tridimensionnels sur une carte en deux dimensions.

Pour plus d'informations

https://docs.qgis.org/3.4/en/docs/gentle_gis_introduction/



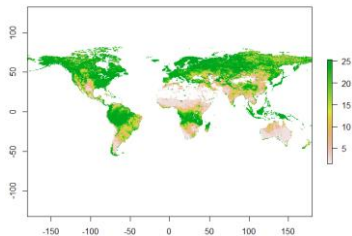
3 familles de projections cartographiques : a) projections Cylindriques, b) Projections Coniques ou c) Projections Planaires (ou Azimutales)

PRINCIPALES COMPOSANTES DE LA STRUCTURE DES DONNÉES DANS UN SIG : MÉTADONNÉES

Informations qui décrivent les données du SIG, incluant leur origine, leur date de création, leur précision, et d'autres informations contextuelles nécessaires pour comprendre et utiliser correctement les données géographiques.

<https://earth.gsfc.nasa.gov/hydro/data/nasa-usda-global-soil-moisture-data>

NASA-USDA SMAP Global Soil Moisture Data (2019/12/31-2020/01/02)



NASA-USDA Enhanced SMAP Global Soil Moisture Data

The NASA-USDA Enhanced SMAP Global soil moisture data provides soil moisture information across the globe at 10-km spatial resolution. This data set includes: surface and subsurface soil moisture (mm), soil moisture profile (%), surface and subsurface soil moisture anomalies (-). The data set is generated by integrating satellite-derived Soil Moisture Active Passive (SMAP) soil moisture observations into the modified two-layer Palmer model using a 1-D Ensemble Kalman Filter (EnKF) data assimilation approach. The Palmer model is driven by precipitation and temperature observations provided by the United States (U.S.) Air Force 557th Weather Wing (formerly known as the U.S. Air Force Weather Agency, AFWA). The current SMAP data archive spans from March 31, 2015 and provides multiple soil moisture products developed using different algorithms. Here, we used the Level 3 (L3) SMAP product generated using the single channel algorithm (SCA) and distributed at a 9-km spatial resolution. Soil moisture anomalies were computed from the climatology of the day of interest. The climatology were estimated based on the full data record of the SMAP satellite observation and the 31 day centered moving window approach. The assimilation of the SMAP soil moisture observations help improve the model-based soil moisture predictions particularly over poorly instrumented areas of the world that lack good quality precipitation data. This dataset was developed by the Hydrological Science Laboratory at NASA's Goddard Space Flight Center in cooperation with USDA Foreign Agricultural Services and USDA Hydrology and Remote Sensing Lab.

Data availability:

01 April 2015 – present

3-days composites

Spatial coverage:

180°W – 180°E, 90°N-60°S

Data access and file formats:

NASA-USDA Enhanced SMAP Global Soil Moisture Data are available in both Gridded Binary (grb2), and Georeferenced Tagged Image File Format (GeoTiff). The binary format of the data can be accessed using the NASA Global Inventory Modeling and Mapping Studies (GIMMS) system. The tiff format of the data can be obtained using the Google Earth Engine.

Métadonnées sous forme de fichier pdf ou html

PRINCIPALES COMPOSANTES DE LA STRUCTURE DES DONNÉES DANS UN SIG : **FORMATS DE DONNÉES**

Les données SIG peuvent être stockées dans divers formats, chacun ayant ses avantages et inconvénients selon les besoins :

- **Shapefile (.shp)** : Format de fichier vectoriel développé par ESRI.
- **GeoJSON** : Format basé sur JSON pour encoder diverses structures de données géographiques.
- **KML (Keyhole Markup Language)** : Utilisé principalement avec Google Earth.
- **TIFF/GeoTIFF** : Format d'image raster avec des informations géoréférencées.
- **NetCDF** (Network Common Data Form) (.cn)

CONFIGURER VOTRE ENVIRONNEMENT R POUR L'ANALYSE GÉOSPATIALE

- Installations de R et Rstudio
- Création d'un projet R
- Installation des packages pour l'analyse et la visualisation des données géospatiales (raster, sf, ggplot, rgdal ...)

CHAPITRE 2 : MANIPULATIONS DES DONNÉES VECTORIELLES

Le contenu de ce chapitre est disponibles dans le fichier « [note_geospatialeR.html](#) »

CHAPITRE 3 : Manipulations des données raster

Le contenu de ce chapitre est disponibles dans le fichier « `note_geospatialeR.html` »

CHAPITRE BONUS A : INTERPOLATION DES DONNÉES RASTER À PARTIR D'OBSERVATIONS PONCTUELLES

Le contenu de ce chapitre est disponibles dans le fichier « `note_geospatialeR.html` »

CHAPITRE BONUS B : COMMENT GÉRER UN FICHER NETCDF ?

Le contenu de ce chapitre est disponibles dans le fichier « `note_geospatialeR.html` »

• CHAPITRE BONUS C : Manipulations Avancées des Données Vectorielles

Le contenu de ce chapitre est disponibles dans le fichier « `note_geospatialeR.html` »

EXERCICE

EXERCICE A : MANIPULATIONS DES DONNÉES VECTORIELLES

Utilise les shapefiles provenant de la base *The Humanitarian Data Exchange* (<https://data.humdata.org/m/dataset>) pour :

- Tracer une carte des pays africains ayant une population comprise entre 10 000 000 et 20 000 000 (shapefile : ``geospatialR/shapefiles/africa_shp/Africa_Countries.shp``)
- Représenter, sous forme de graphique et de carte, les types de commodités les plus fréquents au Bénin et au Nigeria (shapefile : ``hotosm_ben_points_of_interest_points_shp.shp`` ou ``hotosm_nga_points_of_interest_points_shp.shp``)
- Réaliser une carte montrant les zones situées à plus de 10 km des types de commodités les plus fréquents au Bénin ou au Nigeria (shapefile adm1 dans ``point_of_interest/Nigeria ou Benin``)
- BONUS : Utilise le shapefile géospatial ``R/shapefiles/bio_adm2/BIO_ADM2.shp`` pour créer une carte montrant les 100 régions des pays sahéliens ayant les productions moyennes de biomasse les plus élevées en 2019 (ou, pour aller plus loin, les 100 régions ayant les productions moyennes de biomasse les plus élevées sur la période 2000-2019).

EXERCICE B : MANIPULATIONS DES DONNÉES RASTER

- Utilise le raster `geospatialR/rasters/biomass_shl/BiomassValue2019.tif` provenant de la base `The Humanitarian Data Exchange` (<https://data.humdata.org/m/dataset>) pour tracer une carte de la production de biomasse annuelle en 2019 dans les pays du Sahel.
- Affiche les limites ainsi que les noms des pays du Sahel sur la cartes
- Affiche sur la carte la production de biomasse moyenne par pays en 2019.
- Extrait la région du Sénégal ou du Nigéria (1ere carte), augmente la résolution de la carte (2eme carte) et diminue ensuite la résolution de la carte (3eme carte).

EXERCICE B : MANIPULATIONS DES DONNÉES RASTER

BONUS :

- Crée un RasterBrick de NDVI à partir des fichiers rasters se trouvant dans le dossier `geospatialR/rasters/NDVI/` avec `lapply` ou `for loop`.
- Sélectionne et affiche uniquement les 4 premières couches.
- Calcule les valeurs moyennes, maximales et minimales de NDVI pour les quatre premières couches.
- Trace une série temporelle du NDVI (toutes les couches, pas seulement les 4 premières) à un point situé dans ta région d'origine en précisant les coordonnées du point (exemple : $x = 15.5$, $y = 14$; $x =$ longitude et $y =$ latitude en UTM).

NB : Le NDVI (Normalized Difference Vegetation Index) est un indicateur utilisé en télédétection pour mesurer la "végétation verte" ou la santé des plantes. Il est calculé à partir de l'absorption de la lumière dans certaines longueurs d'onde par les plantes, en particulier dans le spectre visible et l'infrarouge proche. Des valeurs proches de 1 indiquent une végétation dense et saine. Des valeurs proches de 0 représentent une couverture végétale très faible ou inexistante (comme le sol nu ou l'eau).

MERCI POUR VOTRE ATTENTION

Souvenez-vous, la nature est le meilleur laboratoire pour comprendre la complexité des systèmes écologiques, et les outils géospatiaux nous permettent de révéler les secrets qu'elle nous murmure !

RÉFÉRENCES

- <https://www.esri.com/fr-fr/what-is-gis/overview>
- https://bookdown.org/nicohahn/making_maps_with_r5/docs/introduction.html
- <https://data.humdata.org/>
- <https://hub.arcgis.com/search?collection=Dataset&type=Shapefile>