***REPUBLIQUE DU SENEGAL***

***Un peuple-Un but-Une foi***

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

***Ministère de l’économie, des finances et du plan***

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

***Agence nationale de la Statistique et de la Démographie***

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

***Ecole nationale de la Statistique et de l’Analyse économique Pierre Ndiaye***

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

***Bureau des Statistiques de l’Amicale des Etudiants et Stagiaires de l’ENSAE***

**Hackaton Stat en action**

**Vers une Analyse Proactive des Conflits : Cartographie et Clustering des Hotspots de Tension en Afrique de l’Ouest**

**Rédigé par :**

**Dieng Samba**

**Fogwoung Djoufack Sarah-Laure**

**Niass Ahmadou**

*Elèves ingénieurs statisticiens économistes*

*3e année cycle long.*

Avril 2025

# Avant-propos

L’École nationale de la Statistique et de l’Analyse économique Pierre Ndiaye (ENSAE) est un établissement d’enseignement supérieur rattaché à l’Agence nationale de la Statistique et de la Démographie (ANSD). Elle est membre du Réseau des Écoles de Statistiques africaines (RESA), avec l’École nationale de la Statistique et d’Économie Appliquée (ENSEA) d’Abidjan, l’Institut sous-régional de la Statistique et d’Économie appliquée (ISSEA) de Yaoundé et l’École nationale d’Économie appliquée et de Management (ENEAM) de Cotonou.

La formation délivrée par l’ENSAE comprend plusieurs filières : les Ingénieurs de Travaux Statistiques (ITS), une formation de quatre ans achevée en 2023 ; les Analystes Statisticiens (AS), un programme de trois ans lancé en 2020 ; les Ingénieurs Statisticiens Économistes (ISE) en cycle court sur trois ans et en cycle long sur cinq ans, ce dernier ayant également commencé en 2020. S'ajoutent à ces programmes le **Master Aide à la Décision et Évaluation des Politiques publiques (ADEPP)** et le **Master en Statistiques agricoles**.

L’ENSAE dispense des cours théoriques de modélisation et de méthodes statistiques. Les formations comprennent également un volet pratique régi par les stages, les enquêtes pédagogiques, et techniques d’utilisation de logiciels statistiques.

Ainsi, en tant qu’élèves Ingénieurs Statisticiens Economistes en troisième année de formation, ce rapport est rédigé dans le cadre du cours de python avancé de vingt heures. Il porte sur **l’analyse de la situation financière des petites et moyennes entreprises au Cameroun** **en 2022** à l’aide du langage Python.

C’EST NECESSAIRE ???

# Sommaire

### ****Introduction****

1. **1.1 Contexte et Problématique**
   * Présentation générale de la situation en Afrique de l'Ouest (évolutions socio-politiques, risques de conflits, instabilité, etc.).
   * Importance de prévenir et anticiper les conflits dans cette région.
   * Pourquoi l’analyse des conflits est un enjeu majeur pour les acteurs politiques, les ONG, et les chercheurs.
2. **1.2 Objectifs du Projet**
   * **Objectif général :** Développer une plateforme interactive permettant d'identifier, visualiser, et analyser les zones à haut risque de conflits en Afrique de l'Ouest en utilisant des données historiques.
   * **Objectifs spécifiques :**
     + Collecter et préparer les données de conflits et socio-économiques (ACLED, données socio-économiques).
     + Analyser les événements de conflits en identifiant des patterns et des hotspots.
     + Créer une carte interactive et une plateforme qui facilitent la compréhension des dynamiques régionales.
3. **1.3 Organisation du Rapport**
   * Bref aperçu de la structure du rapport pour que le lecteur sache à quoi s'attendre.

### ****Chapitre 1 : Méthodologie****

1. **2.1 Collecte des Données**
   * **Données des conflits :** Description des données ACLED, période couverte, types d'événements collectés, etc.
   * **Données socio-économiques :** Description des données socio-économiques utilisées (PIB, chômage, pauvreté, etc.), sources et période.
2. **2.2 Préparation et Nettoyage des Données**
   * Processus de nettoyage des données : gestion des valeurs manquantes, vérification de la cohérence temporelle et géographique.
   * Fusion des données : mise en place d’une base de données unifiée contenant à la fois les événements de conflit et les indicateurs socio-économiques.
3. **2.3 Méthodes d'Analyse**
   * **Analyse descriptive des événements :** Description de la distribution des événements par année, région, type, etc.
   * **Cartographie des événements :** Présentation des outils et techniques utilisés pour cartographier les événements (ex : cartographie dynamique avec des outils comme Dash, React, etc.).
   * **Clustering pour identifier les hotspots :** Méthodes de clustering utilisées (DBSCAN, K-Means), critères et paramètres définis, résultats attendus.
4. **2.4 Développement de la Plateforme Interactive**
   * **Structure de la plateforme :** Description du backend, du frontend, et de la base de données.
   * **Fonctionnalités principales :** Carte interactive des hotspots, analyse des événements, statistiques régionales, etc.

### ****Chapitre 2 : Présentation de l’Équipe et de ses Membres****

1. **3.1 Composition de l’Équipe**
   * Description de la structure de l'équipe (nombre de membres, rôles).
2. **3.2 Profils des Membres de l’Équipe**
   * Présentation des membres de l'équipe avec leur parcours et leur expertise :
     + **Membre 1 :** Rôle, expérience, contributions spécifiques au projet.
     + **Membre 2 :** Rôle, expertise technique, responsabilités.
     + **Membre 3 :** Rôle, compétences, impact sur la mise en œuvre.

### ****Chapitre 3 : Livrables Attendus et Impact du Projet****

1. **4.1 Livrables du Projet**
   * **Livrable principal :** Plateforme interactive fonctionnelle pour l’analyse des événements de conflits et des hotspots en Afrique de l’Ouest.
   * **Autres livrables :** Rapport d'analyse, cartes interactives, documentation technique.
2. **4.2 Impact et Bénéfices**
   * **Impact sur les acteurs locaux :** Aide à la prévention des conflits pour les gouvernements, ONG, chercheurs.
   * **Bénéfices pour la communauté internationale :** Sensibilisation aux dynamiques conflictuelles et meilleure préparation pour intervenir.
   * **Opportunités de développement futur :** Possibilité d'étendre le projet à d'autres régions, d'ajouter de nouvelles fonctionnalités, etc.

### ****Conclusion et Perspectives****

1. **5.1 Résumé des Objectifs et Résultats Attendues**
   * Récapitulation des objectifs et des livrables attendus.
2. **5.2 Perspectives d’Amélioration et Développement**
   * Discussion sur les évolutions possibles du projet à moyen et long terme, notamment en termes de données supplémentaires, de fonctionnalités, ou d'intégration avec d'autres systèmes de gestion de crises.
3. **5.3 Conclusion Finale**
   * Synthèse des points clés et des attentes du projet. Importance de l’initiative pour la région et le rôle crucial de l’analyse des conflits.

### ****Annexes****

* **Annexe 1 : Documentation technique**
* **Annexe 2 : Sources de données et références**
* **Annexe 3 : Graphiques et Cartes interactives**

PENSER A DEFINIR LES TERMES COMPLIQUES DU TITRES

PLAN POUR L’INTRODUCTION

* Voir d’abord si on peut définir ce qu’on entend par conflits
* Puis montrer que depuis les années 2000, le niveau de conflits ne cesse pas de croitre
* Parler des systèmes de conflits dans la région
* Donner des statistiques à chaque fois do not forget
* Parler de l’AES qui a quitter encore en réponse aux bouleversements politiques récents, comme le coup d'État au Niger, et aux menaces d'intervention militaire de la CEDEAO et ont décidé de s'organiser pour une défense mutuelle face aux pressions externes. SI LA SUITE LOGIQUE Y ES
* OBJECTIF GENREAL DU PROJET : Développer une plateforme interactive pour identifier, visualiser et analyser les zones à haut risque de conflits en Afrique de l’Ouest.

OBJECTIF SPECIFIQUES :  **Collecte et préparation des données** : Recueillir les données historiques sur les événements liés aux conflits (provenant notamment d'ACLED) couvrant la période 2000-2022, ainsi que les indicateurs socio-économiques pertinents (PIB régional, taux de chômage, taux de pauvreté, accès à l’éducation, etc.). Ces données seront nettoyées et fusionnées pour constituer une base solide pour l'analyse.

 **Analyse descriptive des événements** : Étudier la distribution des événements dans le temps et l'espace afin d'identifier les tendances, les pics de violence et les caractéristiques spécifiques de chaque type d'événement (manifestations, grèves, conflits, etc.).

 **Identification des hotspots d'instabilité** : Appliquer des techniques de clustering, telles que DBSCAN et K-Means, pour détecter de manière objective les zones géographiques où les événements se concentrent, permettant ainsi de mettre en évidence des "hotspots" de conflits.

 **Développement de la plateforme interactive** : Concevoir et déployer une interface web dynamique permettant de visualiser et d'explorer les données et les résultats des analyses (cartographie des hotspots, graphiques temporels, etc.), afin de faciliter la compréhension des dynamiques conflictuelles par les décideurs et les chercheurs.

 **Documentation et rapport d’analyse** : Produire un rapport détaillé décrivant la méthodologie, les analyses réalisées, les résultats obtenus et leur interprétation, ainsi qu'une documentation technique pour l'utilisation et la maintenance de la plateforme.

* PRESENTATION DU PLAN

# Sommaire

# Liste des graphique/tableaux/encadrés

# Résumé

C’EST NECESSAIRE ?

# Abstract

# Introduction

# Présentation de l’équipe

Issus de la classe **ISEP3 de l’ENSAE Dakar**, unis par une volonté commune de mettre en pratique leurs connaissances académiques pour contribuer à la prévention des conflits en Afrique de l’Ouest. En effet, ce projet s’inscrit dans la continuité des enseignements en en **statistiques exploratoires spatiales** et en **développement web** au premier semestre de cette troisième année. Ces cours nous ont initiés aux outils d’analyse géographique et aux principes de conception d’applications interactives. Par ailleurs, notre expérience avec **R Shiny**, acquise lors du cours de traitements statistiques avec R de notre deuxième année, a été déterminante pour structurer la plateforme. Ce projet incarne ainsi la concrétisation de nos apprentissages théoriques, appliqués à un enjeu sociétal majeur.

1. Profil des membres

* Samba DIENG : Responsable de la Collecte des Données et des Analyses Avancées

Samba DIENG a assuré la collecte et l'intégration des bases de données, ce qui a constitué la première étape cruciale de notre projet. Il a ensuite appliqué des méthodes de regroupement avancées, notamment l'Analyse en Composantes Principales (ACP) et le clustering, afin de détecter des tendances et identifier des profils communs dans les données. Son expertise s'est matérialisée par la réalisation de visualisations graphiques complexes et intuitives qui offrent une lecture claire des résultats. Ces analyses avancées permettent d’informer efficacement les décideurs grâce à des représentations visuelles précises, transformant ainsi des données brutes en insights stratégiques.

* Ahmadou NIASS : Responsable du Développement de l’Exploration Interactive

Passionné par les interfaces utilisateur intuitives, Ahmadou NIASS a conçu l’interface interactive qui permet aux utilisateurs d’explorer les données de manière dynamique et personnalisée. Il a développé une série de filtres interactifs permettant de sélectionner un ou plusieurs pays, de définir une plage de dates, de choisir différents types de violence, de fixer un seuil minimum de morts estimées et même de sélectionner des dyades spécifiques. Grâce à son expertise, il a intégré une carte interactive qui géolocalise chaque événement, où la taille et la couleur des marqueurs varient selon le nombre de morts, avec des popups informatifs sur le conflit, la date, le type de violence, la dyade et les estimations de morts. En complément, il a implémenté des indicateurs synthétiques qui résument le nombre total d’événements, le total de morts estimées et la proportion de civils tués, ainsi qu’un tableau interactif permettant de consulter et d’exporter l’ensemble des données. Ce travail a permis d’offrir une expérience de navigation fluide et ergonomique à tous les utilisateurs, tout en assurant une visualisation claire et précise des informations.

* Fogwoung Djoufack Sarah-Laure : Responsable du Nettoyage des Données et de la Documentation

Sarah-Laure a pris en charge le nettoyage et la préparation de la base de données, garantissant ainsi la qualité et la fiabilité des informations exploitées par l’équipe. Son rôle a été crucial pour s’assurer que les données brutes, issues de multiples sources, soient correctement traitées et prêtes à être analysées. En parallèle, elle a élaboré la documentation du projet avec soin, en détaillant la source des données, en décrivant précisément les variables utilisées et en expliquant l’ensemble des fonctionnalités de la plateforme. Elle a également réalisé des analyses descriptives qui permettent de contextualiser les résultats, en illustrant par exemple la répartition des événements, l’évolution temporelle et d’autres tendances pertinentes. Son travail a permis de rendre le projet non seulement techniquement solide, mais aussi accessible et compréhensible, tant pour les utilisateurs que pour toute personne souhaitant en apprendre davantage sur notre démarche.

La réussite de ce projet repose sur la complémentarité de nos compétences et sur une approche collaborative soigneusement organisée en fonction de nos profils respectifs. Nous avons adopté une méthodologie agile, nous réunissant chaque semaine pour réajuster nos priorités et partager nos avancées, tandis que l'utilisation d'outils tels que GitHub a assuré un versionnement rigoureux du code. Chaque membre a apporté une contribution singulière : l'analyse géospatiale a transformé des données brutes en cartes parlantes, et le développement web a rendu ces informations accessibles à tous grâce à une interface ergonomique et intuitive.

# SOURCE DE DONNEES

Le **Uppsala Conflict Data Program (UCDP)** est un programme de recherche internationalement reconnu, basé à l’Université d’Uppsala en Suède. Il constitue l’une des bases de données les plus complètes et les plus fréquemment utilisées pour l’étude des conflits armés à l’échelle mondiale. Depuis les années 1980, le UCDP collecte, vérifie et structure des informations relatives aux événements violents impliquant des États, des groupes armés ou des acteurs non étatiques.

Dans le cadre de ce projet, les données exploitées proviennent spécifiquement de la **base de données UCDP GED** (*Georeferenced Event Dataset*), qui se distingue par la précision géographique des événements qu’elle recense. Chaque observation correspond à un événement conflictuel unique, géolocalisé, avec des métadonnées complètes telles que la date, le lieu, les acteurs impliqués, le type de conflit et les pertes humaines estimées.

Nous avons téléchargé la version la plus récente de cette base de données directement depuis le site officiel du UCDP (https://ucdp.uu.se/). Les données sont fournies au format CSV, ce qui facilite leur traitement à l’aide d’outils statistiques et de visualisation. L’équipe de projet a ensuite procédé à un nettoyage rigoureux de la base, à la sélection de variables pertinentes pour l’analyse, et à l’adaptation du format pour un usage interactif dans une application développée sous R Shiny.

Cette base constitue une source précieuse pour l’analyse des dynamiques de conflit en Afrique de l’Ouest, car elle permet non seulement de suivre l’évolution des violences dans le temps, mais aussi de cartographier leur distribution spatiale.

# NETTOYAGE DES BASES DE DONNEES

Le processus de nettoyage des données a commencé par le téléchargement des fichiers sources pour chaque pays.

**Problèmes rencontrés avec la version initiale de la base de données :**

Initialement, les données étaient organisées dans un format où les informations étaient séparées par des virgules. Cependant, ce format a rapidement posé plusieurs problèmes lors de l’importation dans des outils comme Excel. En effet, la virgule était utilisée à la fois comme délimiteur global pour séparer les colonnes et comme séparateur interne pour certaines valeurs dans des colonnes spécifiques. Ce problème était particulièrement marqué pour les événements comportant plusieurs sources, où les sources étaient séparées par des virgules. Cette double utilisation de la virgule créait une confusion, car Excel interprétait toutes les virgules comme des délimiteurs de colonnes, risquant ainsi de mal organiser les données et de fragmenter des informations essentielles, voire de les perdre.

Un autre problème majeur était que, parfois, les informations n’étaient pas séparées correctement par des virgules. Par exemple, certaines données étaient simplement ajoutées dans la colonne suivante. Cela pouvait provoquer des colonnes vides pour certains événements ou, pire, des colonnes contenant des informations non pertinentes, non liées aux mêmes données. Cette mauvaise organisation compliquait considérablement l'importation et la gestion des données, car elle perturbait la structure attendue et risquait de fausser les analyses ultérieures.

Lors de l'importation dans Excel, la fonctionnalité "Convertir" était susceptible de mal gérer cette séparation des colonnes, ce qui aurait abouti à une mauvaise répartition des informations ou, pire encore, à l’écrasement des données contenues dans d’autres colonnes. Pour éviter ces problèmes, il a été décidé de ne pas utiliser Excel, mais de nettoyer les données directement dans R, où un meilleur contrôle pouvait être exercé sur l’importation et le traitement.

**Étapes du nettoyage des données :**

* **Regroupement des informations :** La première étape du nettoyage a consisté à regrouper toutes les informations dans une seule colonne. Pour ce faire, l'astérisque (\*) a été choisi comme délimiteur temporaire, car il n'apparaissait dans aucune des données originales. Ce choix a permis de regrouper toutes les informations sans risque de confusion avec les séparateurs de colonnes, préservant ainsi l'intégrité des données.
* **Filtrage des lignes inutiles :** L'une des premières étapes a été d’éliminer les lignes qui ne contenaient pas d’informations pertinentes. Certaines lignes étaient purement textuelles ou servaient d’en-têtes de sections, et ne représentaient pas des événements concrets. Ces lignes, contenant souvent des noms de communes ou des descriptions générales sans lien direct avec les événements, ont été filtrées. Le critère de filtrage était simple : éliminer les lignes qui ne commençaient pas par un chiffre, car elles correspondaient à des informations supplémentaires ou des titres, et non à des données d'événements.
* **Restructuration des données :** Après le filtrage, les données restantes ont été restructurées. La colonne initiale contenant des données textuelles complexes a été scindée en plusieurs colonnes pour séparer les différentes informations contenues dans chaque ligne. Le mot-clé "POINT" a servi de référence pour diviser les données en deux parties :

Avant "POINT", se trouvaient des informations supplémentaires, telles que l'identifiant de l'événement, l'année, le type de violence, etc.

Après "POINT", se trouvaient les informations géographiques essentielles, telles que les coordonnées géographiques de l'événement.

* **Extraction et séparation des coordonnées géographiques :** Les quatre dernières valeurs de la partie après "POINT" ont été extraites pour isoler les coordonnées géographiques (latitude et longitude). Ces coordonnées ont été ensuite séparées en deux colonnes distinctes : latitude et longitude.
* **Élimination des lignes sans données géographiques :** Les lignes sans coordonnées géographiques ont été supprimées, car elles étaient inutilisables pour l’analyse spatiale. Cela a permis de se concentrer sur les événements pertinents et exploitables.
* **Séparation des autres informations :** La colonne contenant les informations après "POINT" a été éclatée en plusieurs colonnes supplémentaires, créant des variables telles que "geom\_wkt", "priogrid\_gid", "country", "region", et d’autres informations liées aux événements. Cette étape a permis de transformer une chaîne de texte complexe en un tableau structuré et plus facile à analyser.
* **Vérification des doublons :** Après restructuration, une vérification des doublons a été effectuée sur les variables clés : latitude, longitude, id et geom\_wkt, qui servent d’identifiants uniques pour chaque événement. Aucun doublon n’a été trouvé, confirmant que chaque ligne représentait bien un événement unique.
* **Conversion des variables numériques :** Certaines variables, telles que les identifiants (id, priogrid\_gid) et les chiffres associés aux événements (par exemple, "deaths\_a", "deaths\_b"), ont été converties en format numérique pour permettre des calculs et analyses statistiques ultérieurs.
* **Suppression des colonnes inutiles :** Enfin, les colonnes qui ne contenaient pas d'informations pertinentes pour l'analyse ont été supprimées. Celles-ci étaient soit redondantes, soit non nécessaires au traitement des données.

**Fusion des bases de données :**

Une fois les étapes de nettoyage et de structuration effectuées pour chaque pays, les bases de données ont été fusionnées en une seule base consolidée. Cette base unique a été utilisée pour effectuer une analyse globale des événements, permettant ainsi une exploitation complète et cohérente des données. Elle comprend un total de trente-deux (32) variables, qui sont les suivantes :

"active\_year", "annee\_end", "annee\_start", "best\_est", "code\_status", "conflict\_dset\_id", "conflict\_name", "conflict\_new\_id", "country", "country\_id", "date\_end", "date\_prec", "date\_start", "deaths\_a", "deaths\_b", "deaths\_civilians", "deaths\_unknown", "dyad\_dset\_id", "dyad\_name", "dyad\_new\_id", "geom\_wkt", "high\_est", "id", "latitude", "low\_est", "longitude", "mois\_end", "mois\_start", "priogrid\_gid", "relid", "region", "type\_of\_violence".

# Présentation des variables

Dans cette section, nous présentons les variables utilisées dans la base de données, chacune décrite de manière détaillée. Ces variables sont essentielles pour l'analyse des événements. Classées par ordre alphabétique, nous avons :

#### ****active\_year****

La variable **active\_year** est un booléen qui indique si l’année de l’événement est active dans la base de données. Elle permet de filtrer les événements par année, en vérifiant si l'événement est pris en compte dans les analyses pour cette année-là.

#### ****best\_est****

La **best\_est** est l’estimation centrale du nombre de morts, représentant la moyenne la plus probable. Cette estimation est la plus fiable disponible à partir des données disponibles, fournissant une indication du nombre de victimes le plus probable dans un événement donné.

#### ****code\_status****

La variable **code\_status** indique le statut de l’inclusion de l’événement dans la base de données. Elle peut prendre deux valeurs :

* "Clear" : l’événement remplit tous les critères UCDP,
* "Organisation" : l’événement remplit tous les critères sauf ceux relatifs aux acteurs locaux.

#### ****conflict\_dset\_id / conflict\_new\_id****

Les variables **conflict\_dset\_id** et **conflict\_new\_id** sont des identifiants numériques qui permettent de regrouper les événements dans un même conflit. Ces identifiants sont utilisés pour analyser les dynamiques des conflits en lien avec les événements associés.

#### ****conflict\_name****

La variable **conflict\_name** contient le nom du conflit, par exemple "Bénin : JNIM vs Civils". Elle permet de situer l’événement dans le contexte spécifique du conflit et d’identifier rapidement de quel conflit il s’agit.

#### ****country****

La variable **country** correspond au nom du pays où l’événement a eu lieu. Elle permet d'identifier rapidement le contexte géographique de l’événement, par exemple "Bénin".

#### ****country\_id****

Le **country\_id** est un code numérique attribué à chaque pays selon la codification de Gleditsch & Ward. Ce code permet une identification unique des pays et est utile pour effectuer des analyses statistiques et des jointures avec d'autres bases de données.

#### ****date\_prec****

La **date\_prec** fournit la précision de la date de l’événement, codée de 0 à 5 :

* 1 signifie que la date est exactement connue,
* 5 signifie qu’il n’y a aucune information sur la date, et celle-ci est approximative.

#### ****deaths\_a****

La variable **deaths\_a** représente le nombre estimé de morts pour "Side A" (généralement l’agresseur ou la force d’intervention dans le conflit). Elle permet d'évaluer les pertes humaines subies par l'une des parties.

#### ****deaths\_b****

Similaire à **deaths\_a**, **deaths\_b** correspond au nombre estimé de morts pour "Side B" (groupe victime ou opposant dans le conflit). Elle permet de comparer les pertes humaines entre les différentes parties impliquées.

#### ****deaths\_civilians****

La variable **deaths\_civilians** indique le nombre de civils tués dans un événement. Cette donnée est cruciale pour comprendre l'impact humanitaire des événements, en particulier dans les conflits où les civils sont souvent les principales victimes.

#### ****deaths\_unknown****

La variable **deaths\_unknown** représente le nombre de morts dont l’origine est incertaine, souvent lorsque l'on ne peut pas attribuer les victimes à un camp spécifique ou à des civils. Elle reflète l'incertitude dans les données disponibles.

#### ****dyad\_dset\_id / dyad\_new\_id****

Les variables **dyad\_dset\_id** et **dyad\_new\_id** sont des identifiants uniques représentant les dyades, c’est-à-dire les paires d’acteurs impliqués dans un événement. Elles permettent d'analyser les relations entre les différents acteurs d’un conflit.

#### ****dyad\_name****

La variable **dyad\_name** désigne le nom de la dyade d'acteurs impliqués dans l’événement, par exemple "JNIM - Civilians". Elle aide à identifier les groupes en conflit et à mieux comprendre les dynamiques de violence entre ces groupes.

#### ****event\_clarity****

La variable **event\_clarity** indique le niveau de précision dans le codage de l’événement. Elle est codée sur une échelle de 1 à 2, où :

* 1 indique un événement bien défini avec une source claire,
* 2 indique un événement moins précis, souvent dû à une information agrégée ou incertaine.

#### ****geom\_wkt****

La variable **geom\_wkt** contient les coordonnées géographiques au format WKT (Well-Known Text), représentant le point d’un événement spécifique. Par exemple, "POINT (2.046 9.314)" indique une localisation précise sur la carte. Ce format est utilisé pour l'intégration dans des systèmes SIG pour une représentation géospatiale des événements.

#### ****high\_est / low\_est****

Les variables **high\_est** et **low\_est** fournissent respectivement les estimations haute et basse du nombre de morts dans un événement. Ces valeurs aident à comprendre les marges d’incertitude autour des estimations centrales de la violence.

#### ****latitude / longitude****

Les variables **latitude** et **longitude** fournissent les coordonnées géographiques exactes de l’événement sous forme décimale. Ces données permettent une représentation cartographique précise des événements et sont souvent utilisées avec **geom\_wkt** pour les analyses géospatiales.

#### ****priogrid\_gid****

La variable **priogrid\_gid** fait référence à un code unique de la cellule PRIO-GRID où l’événement a eu lieu. Cette cellule représente une unité géographique standardisée de 0,5°x0,5°, facilitant les analyses spatiales agrégées et les comparaisons avec d’autres bases de données, comme l’UCDP-GED.

#### ****region****

La variable **region** fait référence à la région géographique selon les classifications UCDP/PRIO, telle que "Western Africa". Elle permet d’analyser les événements sur une échelle régionale et d’observer les tendances en matière de violence dans différentes zones géographiques.

#### ****type\_of\_violence****

La variable **type\_of\_violence** est un code numérique qui décrit le type de violence impliqué dans l’événement. Les valeurs typiques incluent :

* 1 : Conflit armé entre États,
* 2 : Conflit entre un État et un groupe organisé,
* 3 : Violence unilatérale contre les civils.

#### ****year\_start / year\_end****

Ces variables indiquent respectivement l’année de début (**annee\_start**) et de fin (**annee\_end**) de l’événement. Elles permettent de situer l’événement dans le temps et d’analyser les périodes de conflit.

# Exploration interactive

L'onglet **« Exploration interactive »** de l'application permet une analyse approfondie et personnalisée des événements, en offrant une grande flexibilité dans le choix des critères de filtrage. L'utilisateur peut affiner ses recherches en sélectionnant plusieurs pays, une plage de dates, le type de violence (conflit entre États, violence non étatique, ou violence contre civils), un seuil minimum de morts estimées, et une dyade spécifique (groupe A contre groupe B). Ces critères offrent une exploration des données sur mesure, selon les intérêts de l'utilisateur.

Pour faciliter cette interaction, plusieurs composants de **Shiny** ont été intégrés, tels que selectInput() pour la sélection multiple des pays et des types de violence, dateRangeInput() pour spécifier la période recherchée, et sliderInput() pour définir un seuil de morts estimées. Ces filtres sont parfaitement intégrés dans l'interface et réagissent instantanément aux ajustements effectués par l'utilisateur, actualisant ainsi la carte, les indicateurs et le tableau des événements en temps réel.

La **carte interactive** constitue un outil central de cette fonctionnalité. Une fois les filtres appliqués, les événements sont géolocalisés sur la carte, chaque point représentant un événement violent. La couleur et la taille de chaque point sont proportionnelles au nombre de morts estimées, permettant une visualisation intuitive de la gravité des événements. Cette carte est générée à l'aide du package **Leaflet**, en utilisant la fonction addCircleMarkers() pour afficher les événements. Lorsqu'un utilisateur clique sur un point, une info-bulle apparaît avec des informations détaillées sur l'événement, telles que le nom du conflit, la date, le type de violence, la dyade impliquée, et l'estimation des morts, grâce à la fonction popup() de Leaflet.

Pour offrir un résumé rapide et visuel, des **indicateurs clés** sont affichés en haut de la page. Ces indicateurs montrent le nombre total d'événements filtrés, le nombre total de morts estimées, et la proportion de morts civils par rapport aux morts totales. Ces valeurs permettent à l'utilisateur de saisir rapidement l'ampleur des événements en fonction des critères choisis. Ces indicateurs sont créés avec le package **shinydashboard** et la fonction valueBox(), qui affiche les informations sous forme de boîtes visuelles.

Enfin, un **tableau interactif** présente les données détaillées des événements filtrés, avec des colonnes telles que la date, le pays, la dyade, la région, le nombre de morts, le nombre de victimes civiles et le type de violence. Ce tableau est entièrement interactif, permettant à l'utilisateur de trier, filtrer, et rechercher des événements spécifiques. Il est généré grâce au package **DT** et à la fonction datatable(), qui intègre des fonctionnalités de tri, de recherche, et d'exportation. Un bouton d'exportation permet également de télécharger les données filtrées au format **CSV** pour une analyse plus approfondie.

Ainsi, l'onglet **Exploration interactive** combine de manière cohérente plusieurs packages de **Shiny**, créant une expérience utilisateur fluide et dynamique. Les outils d'exploration des données permettent aux utilisateurs d'obtenir des insights personnalisés et d'effectuer une analyse détaillée des événements à l'échelle globale.

# Statistiques descriptives

L'onglet **Statistiques descriptives** de l'application permet une exploration visuelle et agrégée des événements sécuritaires, offrant une vue d'ensemble claire et dynamique des données. L’objectif principal de cet onglet est de permettre à l'utilisateur d'analyser les données sous plusieurs angles : la fréquence des événements, la répartition par type de violence, l'évolution temporelle des incidents et l'impact spécifique sur les populations civiles. L'onglet utilise des graphiques interactifs pour présenter les données de manière intuitive et accessible, permettant ainsi une compréhension approfondie des tendances et des relations entre les variables.

#### ****Répartition par pays****

Le premier bloc de l'onglet **Statistiques descriptives** offre une vue d'ensemble des événements par pays, permettant à l'utilisateur de comparer les nations en termes de fréquence et de gravité des événements violents. Pour cela, deux graphiques à barres sont utilisés. Le premier montre le **nombre d’événements par pays**, et le second représente le **nombre total de morts estimées (best\_est) par pays**. Ces graphiques permettent à l'utilisateur de repérer rapidement quels pays sont les plus affectés, tant en termes de fréquence que de gravité des événements. Les **composants** utilisés pour ces visualisations sont ggplot(), et les **variables** associées incluent country et best\_est.

#### ****Répartition par type de violence****

Le second bloc se concentre sur la répartition des événements en fonction du type de violence. Ce bloc propose deux types de visualisation. Le premier est un **diagramme circulaire** (pie chart) qui montre la proportion des événements selon les trois types de violence : conflit entre États, conflit entre un État et un groupe organisé, et violence contre les civils. Le deuxième graphique est un **barplot** des types de violence par dyade, affichant les 10 dyades les plus fréquentes. Ces graphiques permettent à l'utilisateur de mieux comprendre quel type de violence est le plus courant et comment les dyades influent sur la répartition des événements violents. Les visualisations sont générées à l'aide de plotly pour le pie chart et ggplot pour le barplot. Les **variables** prises en compte sont type\_of\_violence et dyad\_name.

#### ****Évolution temporelle des événements****

Le troisième bloc met en évidence l'évolution des événements au fil du temps. Ce bloc comprend deux graphiques en ligne (line charts) interactifs. Le premier montre l’évolution du **nombre d’événements dans le temps**, et le second illustre l’évolution du **nombre total de morts estimées**. Ces graphiques permettent à l'utilisateur de détecter des périodes de crise intense ou des phases d’accalmie, en analysant les fluctuations temporelles des événements violents. L’objectif est de permettre une analyse longitudinale des données pour repérer des tendances ou des anomalies dans le temps. Ces graphiques sont créés à l’aide des packages plotly et ggplot, avec les **variables** date\_start et best\_est.

#### ****Focus sur la violence contre les civils****

Enfin, le dernier bloc se concentre spécifiquement sur les **violences contre les civils**, un aspect crucial pour évaluer l'impact humanitaire des conflits. Ce bloc propose une carte interactive montrant la géolocalisation des événements de violence contre les civils (identifiés par type\_of\_violence == 3), ainsi qu’une **timeline** dédiée aux événements ayant entraîné des morts civiles. Ces visualisations permettent à l'utilisateur de visualiser l’étendue géographique et temporelle des violences contre les civils, et d’analyser l’impact de ces événements sur les populations non combattantes. La carte est générée grâce au package leaflet, tandis que la timeline est produite avec ggplot. Les **variables** utilisées dans ce bloc sont deaths\_civilians, type\_of\_violence, latitude, et longitude.

En résumé, l'onglet **Statistiques descriptives** offre une exploration visuelle et interactive des données liées aux événements sécuritaires. À travers des graphiques clairs et dynamiques, cet onglet permet de mieux comprendre les tendances des conflits en fonction de différents critères tels que la répartition par pays, le type de violence, l’évolution temporelle des événements, et l'impact sur les civils.

# Analyses avancées

L'onglet Analyses avancées de l'application est conçu pour explorer les relations multivariées, les profils statistiques et offrir des visualisations complexes des données. Il s'adresse à des utilisateurs souhaitant analyser plus en profondeur les dynamiques sous-jacentes des conflits et des violences à travers des techniques de réduction de dimension, de clustering, de visualisation des flux et d'analyse temporelle. L’objectif est d’identifier des patterns, des groupes similaires et d’obtenir une vue détaillée des relations complexes entre les variables. L'onglet se compose de quatre blocs analytiques distincts qui permettent de répondre à ces objectifs.

### ACP (Analyse en Composantes Principales)

Le premier bloc permet d’examiner les relations entre plusieurs variables à l’aide de l’**Analyse en Composantes Principales (ACP)**. Cette méthode est utilisée pour réduire la dimensionnalité des données tout en préservant un maximum d’information. L’ACP permet de visualiser comment les pays, les régions ou les dyades se positionnent en fonction des variables telles que best\_est (estimation des morts), deaths\_a, deaths\_b, deaths\_civilians et type\_of\_violence. L'ACP a été réalisée en standardisant les données, suivie d'une application de l'algorithme de **prcomp** pour la réduction de dimension. Cette analyse est cruciale pour découvrir des groupes ou profils communs parmi les différents acteurs et zones géographiques affectées par des violences. L'objectif ici est de repérer des similarités dans les profils de violence, facilitant l’identification de groupes de pays ou de conflits qui partagent des caractéristiques similaires.

### Clustering

Le second bloc repose sur l’algorithme de **K-means clustering**, qui permet de regrouper les pays, régions, ou dyades en fonction de leur similarité selon les variables précédentes. Avant de procéder au clustering, les données sont d'abord **standardisées** en utilisant la fonction scale pour garantir que chaque variable contribue de manière égale à l'algorithme. Ensuite, le K-means est appliqué avec le nombre de clusters déterminé à l’aide de la méthode du **coudé** (Elbow method), grâce à la fonction fviz\_nbclust de la bibliothèque factoextra. Le résultat est une segmentation des conflits, permettant une analyse plus précise des zones les plus touchées. Cette approche permet de classifier les zones géographiques ou les dyades selon leur gravité, en identifiant des groupes similaires en termes de mortalité ou de type de violence.

### Sankey Diagram

Le troisième bloc présente un **Sankey diagram**, qui est utilisé pour visualiser les flux entre les types de violence, les dyades, et le nombre de morts. Cette visualisation est générée à l’aide de la fonction sankeyNetwork du package networkD3. Le diagramme illustre les interactions entre les acteurs (source et cible) et les flux de violence associés à chaque dyade, en indiquant le nombre total de morts estimées pour chaque connexion. Les données sont d’abord agrégées pour obtenir la somme des morts pour chaque dyade. Ce diagramme permet de comprendre les enchaînements entre les différents acteurs et les types de violences au sein des conflits, et il offre une vision claire des dynamiques complexes entre les acteurs de chaque conflit. Cette méthode est idéale pour explorer comment les différents types de violences sont liés aux dyades et à l’intensité des morts.

### Heatmap temporelle

Le dernier bloc propose une **carte de chaleur (heatmap)** des événements dans le temps, permettant d’identifier les pics ou cycles de conflits en fonction de la période de l'année. Cette heatmap est générée en utilisant la fonction ggplot2, où les événements sont agrégés par mois ou par année pour déterminer le nombre total de morts ou d’événements. Les données sont filtrées pour un pays donné, comme le Mali ou le Burkina Faso, et une agrégation est réalisée en utilisant la fonction summarise du package dplyr pour obtenir le nombre d’événements et la somme des morts estimées pour chaque mois. L'analyste peut ainsi repérer les périodes où les conflits sont les plus intenses ou les plus fréquents, ce qui peut aider à détecter des schémas saisonniers ou des déclencheurs spécifiques. Cette analyse permet d’obtenir des insights précieux sur l’évolution temporelle des conflits dans ces zones.

L'onglet **Analyses avancées** permet ainsi de réaliser des analyses statistiques sophistiquées, de segmenter les zones géographiques ou les dyades en fonction de leur gravité, de comprendre les flux de violence entre les acteurs, et d’identifier des tendances temporelles importantes. Ces analyses sont essentielles pour une compréhension complète des dynamiques des conflits et peuvent fournir des informations cruciales pour la gestion des crises et l’interprétation des données.