République du Sénégal



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



Projet de conception d'une application interactive pour la visualisation de l'évolution des conflits au sein de l'Union Économique et Monétaire Ouest-Africaine

Vers une analyse proactive des conflits : Cartographie et classification des zones de tension au sein de l'UEMOA

Rédigé par :

Dieng Samba Fogwoung Djoufack Sarah-Laure Niass Ahmadou

Elèves ingénieurs statisticiens économistes 3e année cycle long.

Avril 2025

Sommaire

Sommaire	2
Table des figures	3
Introduction	4
I. Contexte et état des lieux	4
II. Problématique	5
III. Objectifs du projet	5
IV. Justification du projet	6
Méthodologie	7
I. Source de données	7
II. Nettoyage des bases de données	7
III. Présentation des variables	10
IV. Fonctionnalités de l'application	14
Présentation des membres de l'équipe	24
Glossaire	27
Références bibliographiques	28

Table des figures

Figure 1: Exploration interactive-Carte des évènements	15
Figure 2: Exploration interactive-Filtres de sélection	15
Figure 3: Exploration interactive-Indicateurs clés pour la sélection	16
Figure 4: Exploration interactive-Détail et téléchargement des évènements filtrés	16
Figure 5:Statistiques descriptives – Répartition par pays	17
Figure 6: Statistiques descriptives – Répartition par type de violence	18
Figure 7: Statistiques descriptives - Évolution temporelle des événements	19
Figure 8: Statistiques descriptives - Focus sur les violences contre civils	20
Figure 9: Analyses avancées – Lissage-ACP-Clustering	22
Figure 10: Analyses avancées: Sankey Diagram	23

Introduction

I. Contexte et état des lieux

Depuis les années 1990, l'Union Économique et Monétaire Ouest-Africaine (UEMOA) a été secoué par de multiples crises politiques, sociales et sécuritaires. L'UEMOA, dont la mission initiale était de renforcer l'intégration économique et monétaire entre ses États membres, fait face aujourd'hui à une instabilité croissante qui menace non seulement la cohésion régionale, mais également la capacité des États à maintenir la paix et la sécurité.

Les crises se sont multipliées au fil des décennies. Des rébellions touarègues au Mali dans les années 1990 aux attaques terroristes récurrentes dans la région sahélienne, en passant par les coups d'État successifs au Burkina Faso, au Mali et au Niger, l'UEMOA est devenue une zone où les conflits prennent des formes multiples et se multiplient à un rythme alarmant. Selon les données du Uppsala Conflict Data Program (UCDP), le nombre de conflits a considérablement augmenté, passant de 12 conflits en 1989 à plus de 900 conflits ayant débutés en 2023, soit une multiplication par 75 en seulement 34 ans. Cette évolution témoigne de l'intensification des tensions et de la nécessité de comprendre les dynamiques conflictuelles.

La création de l'Alliance des États du Sahel (AES), réunissant le Mali, le Burkina Faso et le Niger, marque une rupture significative avec les institutions régionales traditionnelles, telles que la CEDEAO et l'UEMOA. Ce nouvel ensemble politique, porté par des régimes militaires, s'inscrit dans un contexte de méfiance vis-à-vis des mécanismes régionaux jugés inefficaces pour faire face aux crises sécuritaires (The Guardian. (2024, July 8). *Ecowas warns of 'disintegration' as juntas split from west African bloc*). Cette recomposition du paysage politique ouest-africain appelle à une compréhension approfondie de l'évolution des conflits : **où éclatent-ils ? Quand ? Quels**

types de violences ? Quelles zones sont les plus touchées ? Et surtout, quelles tendances peuton identifier sur le long terme ?

II. Problématique

Dans un contexte marqué par la montée de l'instabilité et les recompositions politiques régionales, comment visualiser, analyser et mieux comprendre l'évolution des conflits dans l'espace UEMOA depuis 1989 ?

III. Objectifs du projet

Le projet a pour objectif de développer une application interactive offrant une exploration des conflits au sein de l'UEMOA entre 1989 et 2023. Cette application vise à faciliter l'analyse, la visualisation et l'interprétation des dynamiques complexes des événements violents, en permettant une compréhension détaillée et ciblée des enjeux. Plus précisément, les objectifs spécifiques sont les suivants:

- Proposer une exploration interactive personnalisée des données, permettant à l'utilisateur de filtrer les événements selon divers critères tels que le pays, le type de violence, la période et le seuil de mortalité, afin d'obtenir une vue d'ensemble ciblée des conflits.
- Fournir des statistiques descriptives visuelles, permettant une analyse comparative des événements de violence par pays, type de violence et impact sur les civils, tout en offrant une présentation claire de l'évolution temporelle des conflits.
- Comparer les tendances entre les pays membres afin de dégager de mettre en évidence les tendances spécifiques à chacun.
- Intégrer des analyses avancées, telles que l'Analyse en Composantes Principales (ACP) et le clustering, afin d'identifier des patterns, des regroupements et des relations entre les différentes variables des conflits.
- Fournir des visualisations temporelles et géographiques détaillées, incluant des cartes interactives et des heatmaps, pour aider à identifier les tendances saisonnières et les zones les plus touchées par la violence.

Ainsi, cette application offrira une ressource efficace pour la prise de décision, en fournissant une vue d'ensemble claire et détaillée des dynamiques des conflits dans l'espace UEMOA. Elle facilitera l'identification des zones et des périodes les plus touchées, contribuant ainsi à une gestion mieux informée des enjeux liés aux conflits.

IV. Justification du projet

En Afrique de l'Ouest, et plus particulièrement au sein de l'UEMOA, les données sur les conflits violents proviennent de sources multiples (rapports d'ONG, bases internationales telles ACLED ou UCDP, médias, etc.) et se présentent sous des formats variés. Cette dispersion rend leur agrégation et leur exploitation difficiles, ce qui limite la compréhension des dynamiques conflictuelles et complique la prise de décision pour prévenir et gérer efficacement les crises.

La récente formation de l'Alliance des États du Sahel (AES) – consécutive au départ de certains pays de la CEDEAO – illustre les conséquences d'une gestion insuffisante des conflits dans la sous-région. Les acteurs se sont retrouvés privés d'un suivi clair et détaillé des événements, ce qui a contribué à un climat de méfiance et à la recherche d'alternatives régionales.

Le projet répond à ces enjeux en proposant une application interactive qui :

- Centralise et rend accessibles les données relatives aux conflits violents de 1989 à 2023, afin d'offrir une source unique et structurée d'informations aux chercheurs, journalistes et décideurs.
- Facilite l'analyse et la visualisation des tendances et des dynamiques des conflits grâce à des outils interactifs (cartes, graphiques, indicateurs, heatmaps), permettant d'identifier rapidement les zones et périodes à risque.
- Intègre des analyses avancées (ACP, clustering) pour révéler les patterns et les relations entre les variables, afin d'éclairer la compréhension des dynamiques conflictuelles et d'anticiper d'éventuelles évolutions.
- Soutient la prise de décision en fournissant des visualisations claires et actualisées, essentielles pour orienter les stratégies de prévention et limiter les risques d'une mauvaise gestion des crises.

Ainsi, cette application vise à offrir une ressource fiable pour une meilleure gestion des conflits dans l'UEMOA, contribuant à prévenir les ruptures institutionnelles et à instaurer une approche plus proactive et coordonnée dans la région.

Méthodologie

I. Source de données

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 des bases de données des huit (08) pays de l'UEMOA, 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 des bases, la fusion en une seule 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.

II. Nettoyage des bases de données

1. Problèmes rencontrés avec la version initiale des bases

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.

2. Etapes de nettoyage des bases de données

Ce processus de nettoyage a été appliqué de manière identique à l'ensemble des pays. Les étapes sont les suivantes :

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 :

La deuxième étape 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.

3. 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 et 4455 observations.

III. 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 :

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.

4 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 État et groupe armé,
- 2 : Conflit entre groupes armés,
- 3 : Violence 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.

IV. Fonctionnalités de l'application

1. 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 <code>valueBox()</code>, 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.

Figure 2: Exploration interactive-Filtres de sélection

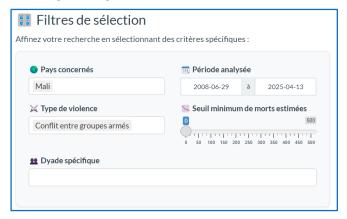


Figure 1: Exploration interactive-Carte des évènements

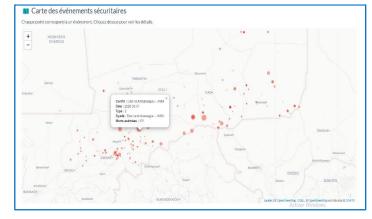
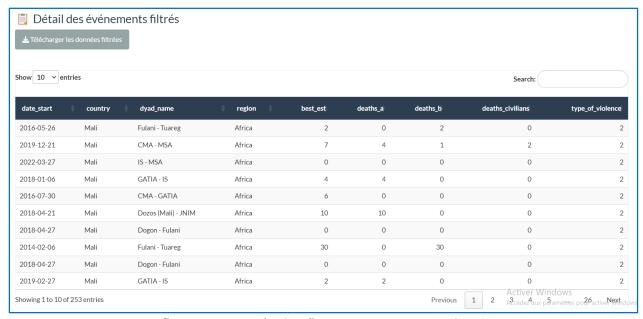


Figure 3: Exploration interactive-Indicateurs clés pour la sélection



Source: Uppsala Conflict Data Program (UCDP).

Figure 4: Exploration interactive-Détail et téléchargement des évènements filtrés



<u>Source</u>: Uppsala Conflict Data Program (UCDP)

2. 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 <code>ggplot()</code>, et les **variables** associées incluent <code>country</code> et <code>best_est</code>.

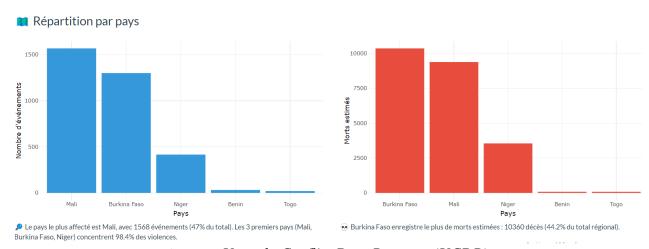


Figure 5:Statistiques descriptives – Répartition par pays

Source: Uppsala Conflict Data Program (UCDP).

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.

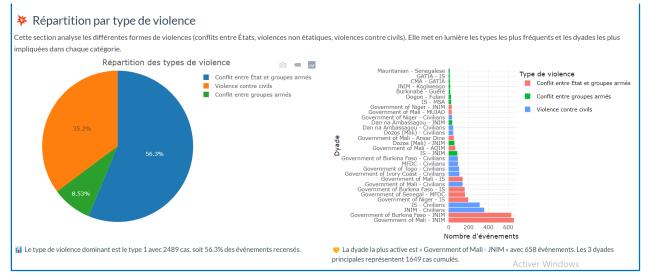


Figure 6: Statistiques descriptives – Répartition par type de violence

Source: Uppsala Conflict Data Program (UCDP).

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.

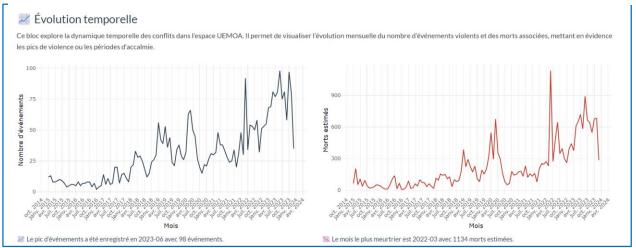


Figure 7: Statistiques descriptives - Évolution temporelle des événements

Source: Uppsala Conflict Data Program (UCDP).

♣ 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.

Figure 8: Statistiques descriptives - Focus sur les violences contre civils

Source: Uppsala Conflict Data Program (UCDP).

3. 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.

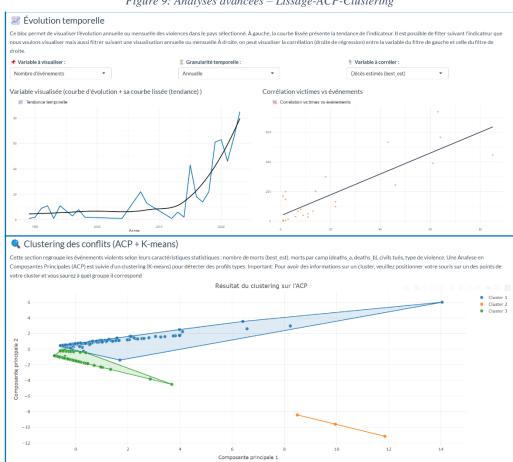


Figure 9: Analyses avancées – Lissage-ACP-Clustering

<u>Source</u>: Uppsala Conflict Data Program (UCDP)

Government of Nigeria

Figure 10: Analyses avancées: Sankey Diagram

Source: Uppsala Conflict Data Program (UCDP)

Présentation des membres 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 au sein de l'UEMOA. 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.

Présentation de 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.

Présentation de Ahmadou NIASS, responsable du développement du frontend et 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.

Présentation de Sarah-Laure FOGWOUNG DJOUFACK, responsable du nettoyage des données, de la documentation et des statistiques descriptives



Sarah-Laure a assuré le nettoyage et la préparation de la base de données, garantissant ainsi la qualité et la fiabilité des informations exploitées par l'équipe. Elle a joué un rôle essentiel en traitant minutieusement les données brutes issues de sources multiples pour les rendre prêtes à l'analyse. Parallèlement, elle a rédigé la documentation du projet de manière détaillée, en précisant la source des données, en décrivant précisément les variables utilisées et en expliquant l'ensemble des fonctionnalités de la plateforme. Son travail a non seulement

renforcé la robustesse technique du projet, mais il a également facilité sa compréhension tant pour les utilisateurs que pour toute personne souhaitant mieux appréhender notre démarche.

La réussite de ce projet repose sur la synergie de nos compétences et une approche collaborative organisée selon nos expertises respectives. Nous avons adopté une méthodologie structurée, avec des réunions hebdomadaires pour ajuster nos priorités et partager nos avancées. L'utilisation d'outils tels que GitHub a permis un versionnement rigoureux du code et une meilleure coordination. Chaque membre a apporté une contribution spécifique : l'analyse géospatiale a permis de transformer des données brutes en cartes claires et informatives, tandis que le développement web a abouti à une interface ergonomique et intuitive, rendant ces informations facilement accessibles à tous.

Glossaire

- ❖ Conflit (dans le cadre de l'UCDP) : une incompatibilité contestée portant sur le gouvernement et/ou le territoire, impliquant l'usage de la force armée entre deux parties, et entraînant au minimum 25 morts liées aux combats au cours d'une année civile.
- ❖ Dyade: paire d'acteurs impliqués directement dans un événement de violence. Elle désigne le couple "agresseur-victime" ou "belligérants", par exemple un État et un groupe armé, ou deux groupes armés opposés.
- Estimation (low_est, best_est, high_est): Trois niveaux d'estimation du nombre de morts:
 - *low est* : estimation minimale,
 - *best_est* : estimation la plus probable,
 - *high_est* : estimation maximale.
- ❖ Evenement : L'incidence de l'usage de la force armée par un acteur organisé contre un autre acteur organisé, ou contre des civils, entraînant au moins un décès direct, dans l'une des catégories d'estimation (basse, moyenne ou haute), à un lieu spécifique et pendant une durée temporelle définie.
- ❖ Sankey Diagram : un type de graphique de flux utilisé pour représenter des transferts ou des mouvements de quantités entre différentes catégories ou entités. Dans ce type de diagramme, la largeur des flèches (ou des bandes) est proportionnelle à la quantité de flux ou de valeur qu'elles représentent.

Références bibliographiques

- Lindberg Bromley, S. (2017). Introducing the UCDP Peacemakers at Risk dataset, sub-Saharan Africa 1989–2009. Journal of Peace Research.
- Lindberg Bromley, S., & Greek, M. (2016). *UCDP Peacemakers at Risk (PAR) Codebook and User Guide*. Version 1.1-2016. Uppsala University, Department of Peace and Conflict Research.
- **The Guardian**. (8 juillet 2024). *Ecowas warns of 'disintegration' as juntas split from west African bloc*.
- Secrétariat du Club du Sahel et de l'Afrique de l'Ouest / OCDE. (2024). Notes ouestafricaines : Des réseaux de conflit fragmentés en Afrique du Nord et de l'Ouest. Organisation de Coopération et de Développement Économiques (OCDE).
- Uppsala Conflict Data Program (UCDP). (2021). UCDP/PRIO Armed Conflict Dataset
 Codebook, Version 21.1. Uppsala University, Department of Peace and Conflict Research
 & Peace Research Institute Oslo (PRIO), Centre for the Study of Civil Wars.