



Communiquez vos résultats

Étude:

Phénomènes Aérospatiaux Non identifiés

2021

Isabelle Barbier



1. INTRODUCTION

Les phénomènes "étranges" ou PAN (pour **P**hénomènes **A**érospatiaux **N**on identifiés) fascinent autant qu'ils interrogent. Qu'ils soient observés dans le ciel ou imprimés sur une photo, la frontière est ténue de les attribuer à des phénomènes extra-terrestres, voire à des apparitions d'aliens. Pourtant, seuls 3,3 % de ces phénomènes demeurent non identifiés après enquête, 40,3 % sont des événements probablement identifiés et 23,2 % sont parfaitement identifiés. Reste 33,2 % d'entre eux, qui n'ont pas pu être déterminés par manque de données (1).

"Les envahisseurs, des êtres étranges venus d'une autre planète. Leur destination : la Terre. Leur but : s'y établir et en faire leur univers. David Vincent les a vus. Pour lui, cela a commencé pendant une triste nuit, le long d'une route solitaire de campagne, alors qu'il cherchait un raccourci que jamais il ne trouva" (2). Cette étude n'a pas pour objectif de débusquer une présence alien mais simplement de recenser les phénomènes les plus communs et les plus rares, d'analyser leurs répartitions géographiques et temporelles, leurs fréquences et leurs durées.

Une analyse des phénomènes au niveau mondial sera proposée puis l'étude se concentrera sur les trois pays les mieux représentés du fichier à savoir les USA, le Royaume-Uni et le Canada.

Enfin, l'analyse prendra en compte les populations et des données météorologiques.

Deux modèles seront proposés. Le premier devra prédire si le phénomène peut être expliqué ou non. Le deuxième devra indiquer le pays le plus probable d'apparition du phénomène en fonction de sa durée, de sa forme et du mois.

Sommaire

1. Introduction.....	1
2. Nettoyage des données.....	3
3. Analyse des données : monde entier.....	9
4. Analyse des données : USA, Canada et Royaume-Uni.....	17
5. Étude des phénomènes en fonction du pays.....	25
6. Modélisations.....	41
7. Conclusion.....	47

Glossaire

Terme	Définition
Angular_shape	Forme angulaire du type chevron, delta, rectangle, triangle, diamant, cône, 'pyramide, dôme, hexagone, croix
Cat_UFO_shape	Catégorie de forme des phénomènes
Changing_shape	Forme changeante, inconstante
Cylinder_shape	Forme cylindrique de type cigare, cylindre
Formation_shape	Phénomènes en formation, le déplacement de chaque objet est en coordination
Hour	Heure
Month	Mois
Observation	PAN
Other_shape	Forme qui ne peut être référencée car inconnue ou du fait d'une mauvaise observation
Ovoid_shape	Forme ovoïde du type sphère, larme, œuf, rond, cercle, disque, ovale
PAN	Phénomène Aérospatial Non identifié
Phénomène	PAN
UFO_shape	Forme du PAN
UK	Royaume-Uni
United-Kingdom	Royaume-Uni
Year	Année

2. NETTOYAGE DES DONNÉES

2.1. Description du fichier

Le fichier d'étude est importé de la plateforme Kaggle(3) . Il se nomme " ufo_sighting_data " et est au format csv. Il recense l'ensemble des observations de phénomènes " étranges " rapportés de 1910 à 2014 au niveau mondial. Trois pays sont largement représentés et seront retenus ultérieurement pour l'étude : les USA, le Royaume-Uni et le Canada. Le fichier comporte 80332 observations et 11 variables.

2.1.1. Types de variables

Les variables sont toutes au format " object ".

Description des variables :

Date_time : date et heure de l'observation des phénomènes

city : ville de l'observation

state/province : état ou province de l'observation

country : pays

UFO_shape : forme du phénomène

length_of_encounter_seconds : durée du phénomène

described_duration_of_encounter : description de la durée du phénomène

description : témoignage

date_documented : date de publication du témoignage

latitude : latitude du lieu d'observation

longitude : longitude du lieu d'observation.

Les variables vont être typées en fonction de leur signification :

datetime64 : *Date_time*

float64 : *described_duration_of_encounter* , *latitude*, *longitude*

2.1.2. Traitement des valeurs manquantes

Le fichier est relativement bien renseigné mais il manque un certain nombre de valeurs sur des variables essentielles à l'étude telles que "country" et "state/province".

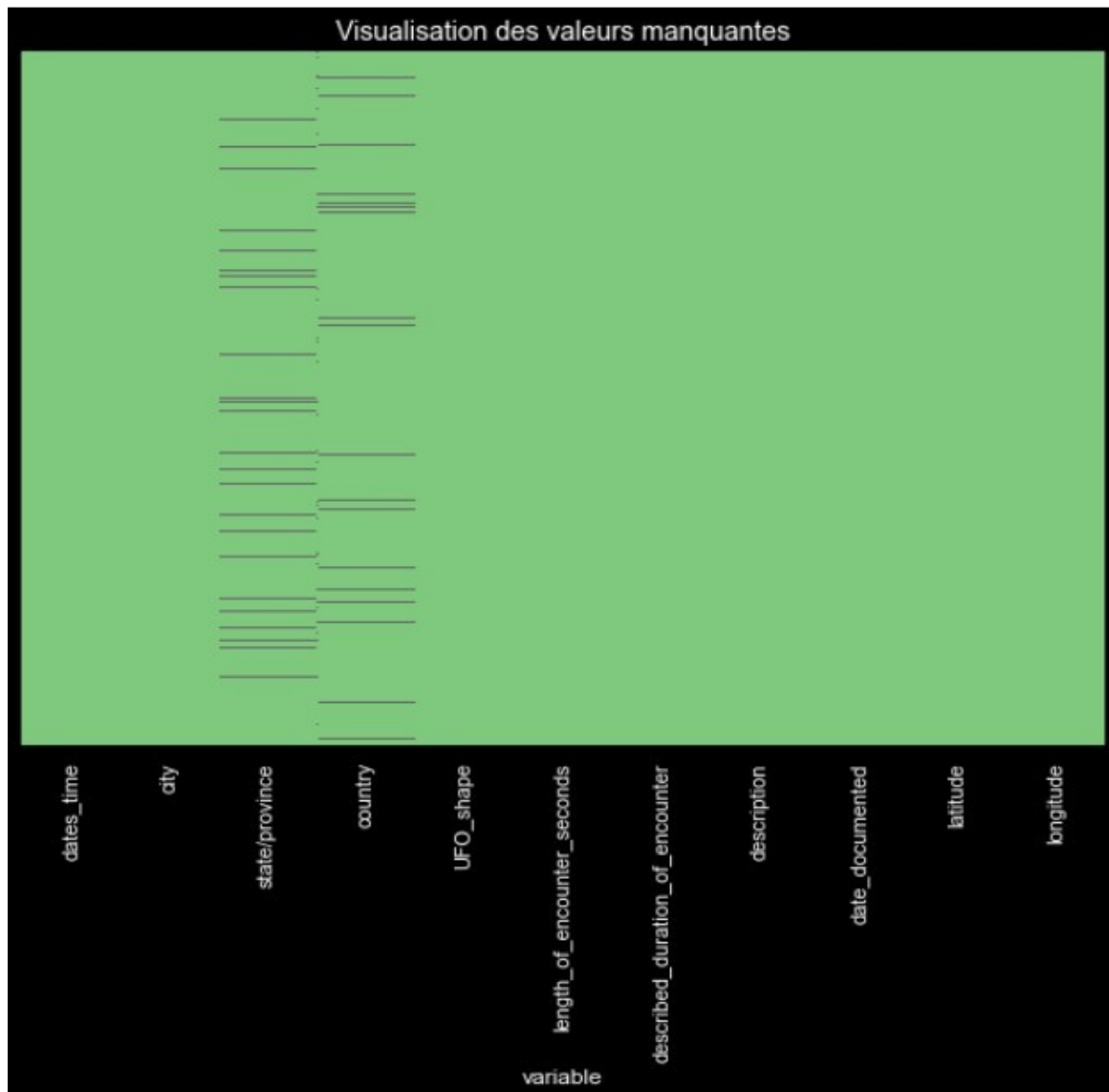


Figure 1: Valeurs manquantes par variable

Même si la proportion des valeurs renseignées est dans l'ensemble satisfaisante, les valeurs manquantes concernant ces variables doivent être implémentées pour une étude plus fine.

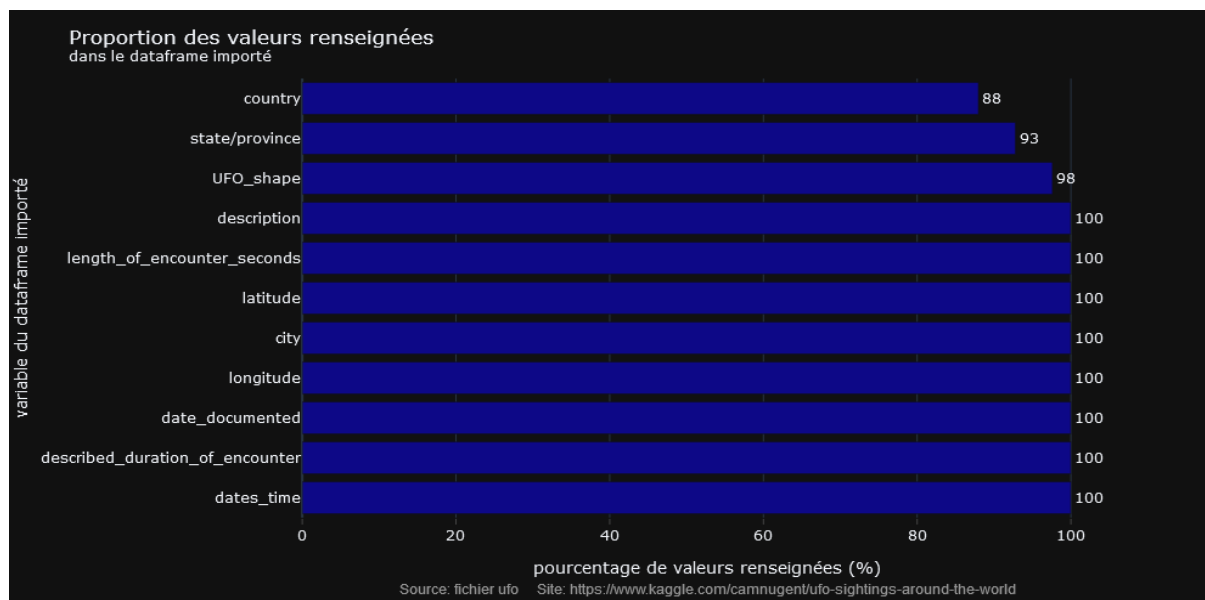


Figure 2: Des valeurs manquantes sur des variables cibles

Les données ne comportant pas la description de la forme du phénomène (variable " UFO_shape ") sont supprimés.

Implémentation des données (4):

Certaines données sont renseignées dans d'autres variables. Par exemple, des valeurs " country " sont indiquées entre parenthèses dans la variable " city ". Ces valeurs sont récupérées et placées correctement dans la variable " country ". De même, des valeurs de la variable "state/province " ne correspondent pas à la valeur correcte de la variable "country" (ce sont notamment le cas de plusieurs villes canadiennes attribuées, à tort, aux USA).

La variable " length_of_encounter_seconds " est implémentée à partir de sa valeur indiquée dans la variable " described_duration_of_encounter ".

L'unique valeur manquante de " latitude " est implémentée à partir d'une valeur récupérée sur internet(5).

Suppression des données lorsque les variables " city " et "state/province " sont toutes deux manquantes.

Suppression des colonnes redondantes (à l'instar de " described_duration_of_encounter " avec " length_of_encounter_seconds ") ou inutiles pour cette étude telles que " description " et " date_documented ".

Implémentation des provinces du Royaume-Uni :

Une liste des états/districts du Royaume-Uni est établie à partir de données de différentes sources internet(6). Chaque valeur de " city "étant reliée à une valeur de " state/province ", les valeurs manquantes sont implémentées grâce à la jointure des deux fichiers. La même méthode (création d'un fichier rassemblant les variables " city " et " state/province " sera utilisée pour l'implémentation des valeurs de la variable " state/province " manquantes.

2.2. Importation de fichiers complémentaires pour affiner l'étude

2.2.1. Population

Il est intéressant de voir s'il existe une corrélation entre le nombre d'habitants et les phénomènes(4). Pour cela, le nombre d'habitants pour les USA, le Royaume-Uni et le Canada est récupéré sur internet(7–11). Les données obtenues sont fusionnées au fichier d'étude.

2.2.2. Météo

Les phénomènes peuvent-ils être corrélés à la météo ? Pour le vérifier, nous récupérons le fichier "ufo_weather" sur la plateforme kaggle(12). Il rassemble l'ensemble des données météorologiques et observations de phénomènes " étranges " rapportées de 1970 à 2019 au niveau mondial (4). Ce fichier, comportant 18000 observations et 15 variables, sera nettoyé et traité indépendamment du fichier principal(4). Après nettoyage, il se nomme 'weather_clean' et est au format csv.

2.3. Export des fichiers

Les fichiers, ainsi nettoyés, sont tous exportés au format csv pour l'analyse. Quatre fichiers seront nécessaires à l'étude :

"ufo_clean.csv" : fichier principal nettoyé qui comprend les données mondiales.
"ufo_state_us_ca_gb_clean.csv" : fichier nettoyé qui comprend les données de nos trois pays d'étude (USA, Royaume-Uni et Canada).
"pop_us_ca_gb_clean.csv" : fichier nettoyé qui comprend, en plus des données de nos trois pays d'étude (USA, le Royaume-Uni et le Canada), le nombre d'habitants par pays. **"weather_clean.csv"** est le fichier nettoyé de données qui comprend des données météo.

Tous ces fichiers sont utilisés pour l'analyse.

3. ANALYSE DES DONNÉES : MONDE ENTIER

3.1. Étude générale des phénomènes au niveau mondial

Le fichier " ufo_clean " comprend les données mondiales des phénomènes de 1914 à 2014.

Une répartition des phénomènes inégale selon les pays.

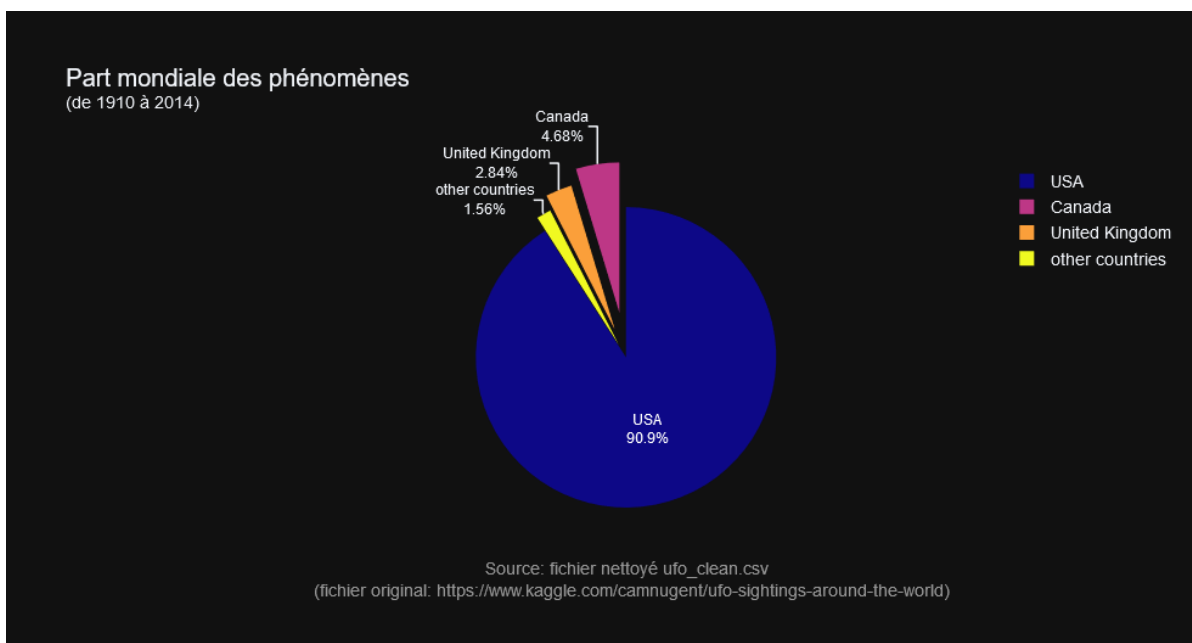


Figure 3: Les USA, le Canada et le Royaume-Uni sont les principaux pays représentés dans le fichier

La majorité des données du fichier concernent les USA (plus de 90% des données du fichier).

3.1.1. Étude des formes de phénomènes

La fréquence d'un phénomène dépend de sa forme.

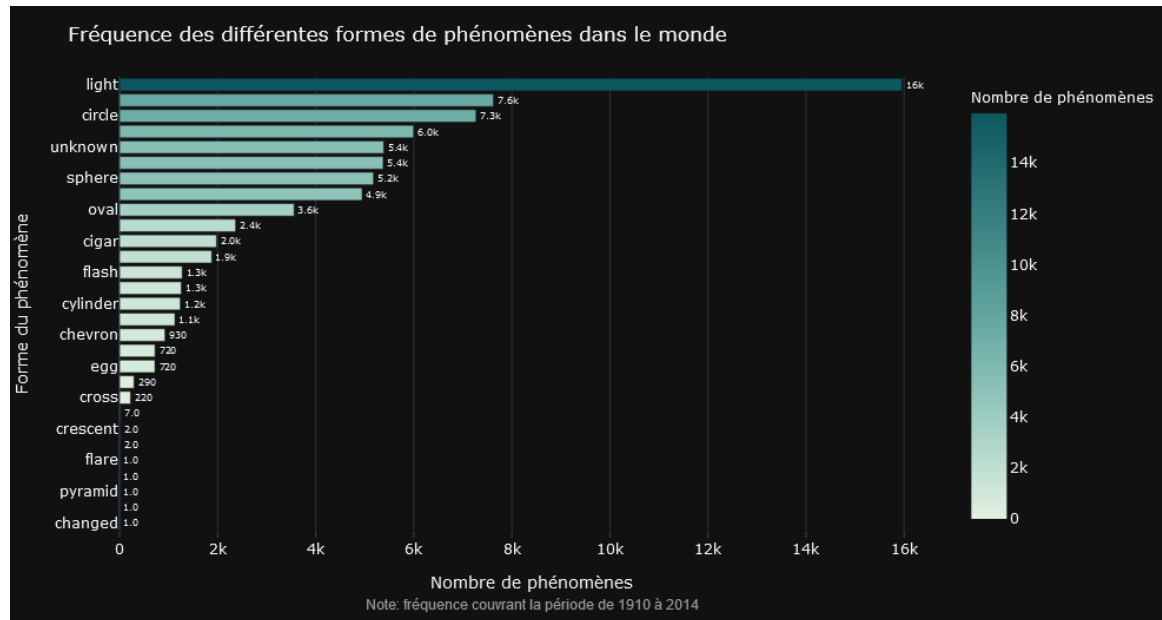


Figure 4: Les phénomènes lumineux sont majoritaires dans le monde

Les formes lumineuses sont les plus fréquentes. A contrario, les formes plus complexes comme les pyramides ou changeantes sont beaucoup plus rares. Les cônes sont moins nombreux mais plus persistants.

De nombreux phénomènes naturels sont lumineux. A l'instar des Phénomènes Transitoires Lumineux ou TLE (13), on pourrait citer les éclairs, les sprites (14). Ces événements lumineux très particuliers et souvent méconnus du grand public peuvent être référencés parmi les phénomènes atypiques. D'autres phénomènes lumineux peuvent être attribués à des objets tels des météores ou des débris spatiaux rentrant dans l'atmosphère, ou encore le reflet du soleil sur des satellites en orbite...Une multitude de phénomènes se trahissent par l'émission de lumière.

Les phénomènes se caractérisent également selon leur durée moyenne.

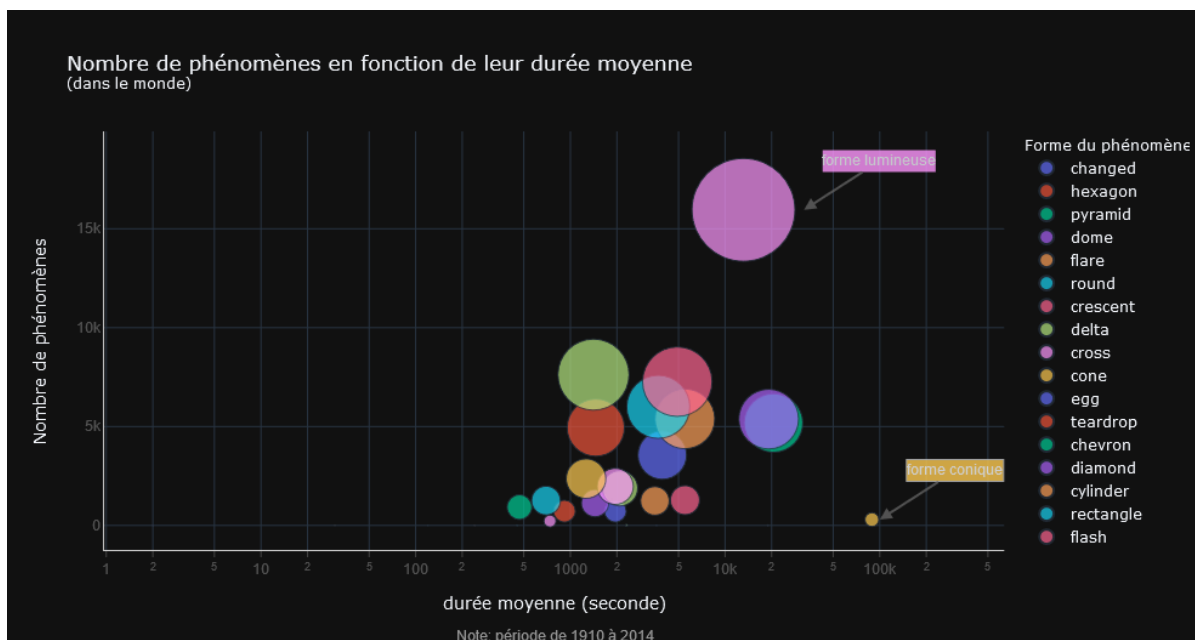


Figure 5: Les phénomènes coniques sont les plus persistants dans le monde avec une durée moyenne de plus d'une journée

Les phénomènes lumineux sont les plus observés et durent longtemps. A l'inverse , les chevrons sont les moins nombreux et les plus brefs.

Les différentes formes de phénomènes vont être regroupés selon leurs caractéristiques communes pour une meilleure analyse.(voir nouveaux groupes (15)). Ainsi, sept catégories de formes sont créées pour faciliter l'étude (lumineuses, ovoïdes, angulaires, cylindriques, en formation, changeantes et autres formes).

Ainsi catégorisés, une nouvelle étude des phénomènes est lancée.

Fréquence des phénomènes en fonction des catégories de forme.

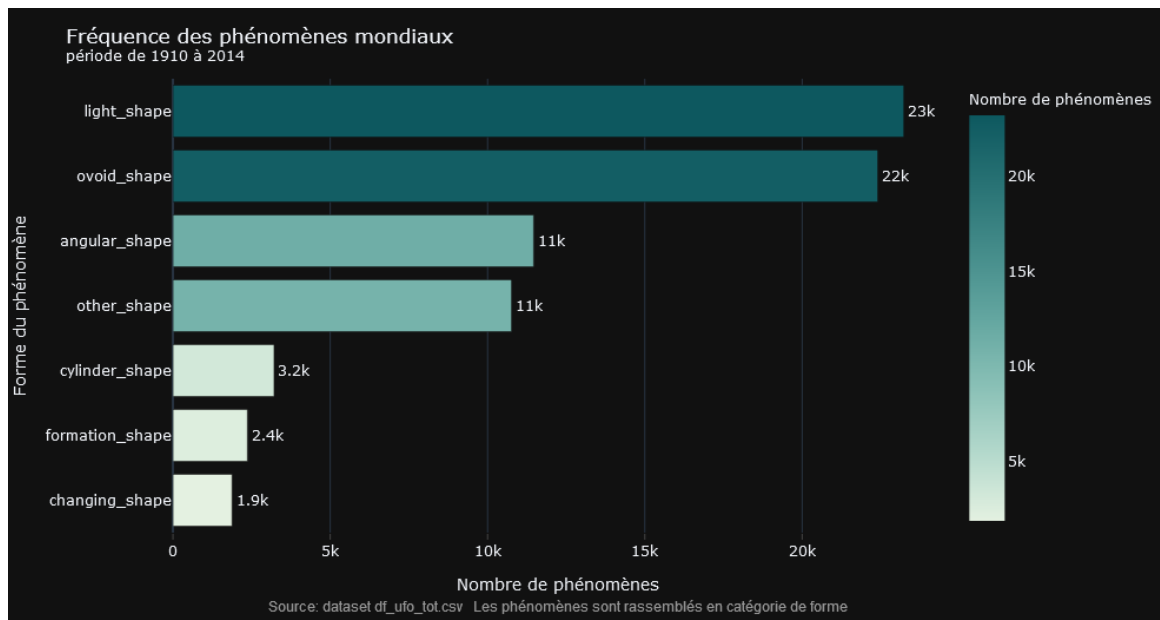


Figure 6: Les formes lumineuses et ovoïdes sont les plus communes dans le monde

Même conclusion que précédemment : les formes plus complexes telles que changeantes ou en formation sont minoritaires. Les phénomènes non répertoriés (other_shape) sont les plus persistants avec une durée moyenne de 3,5 heures.

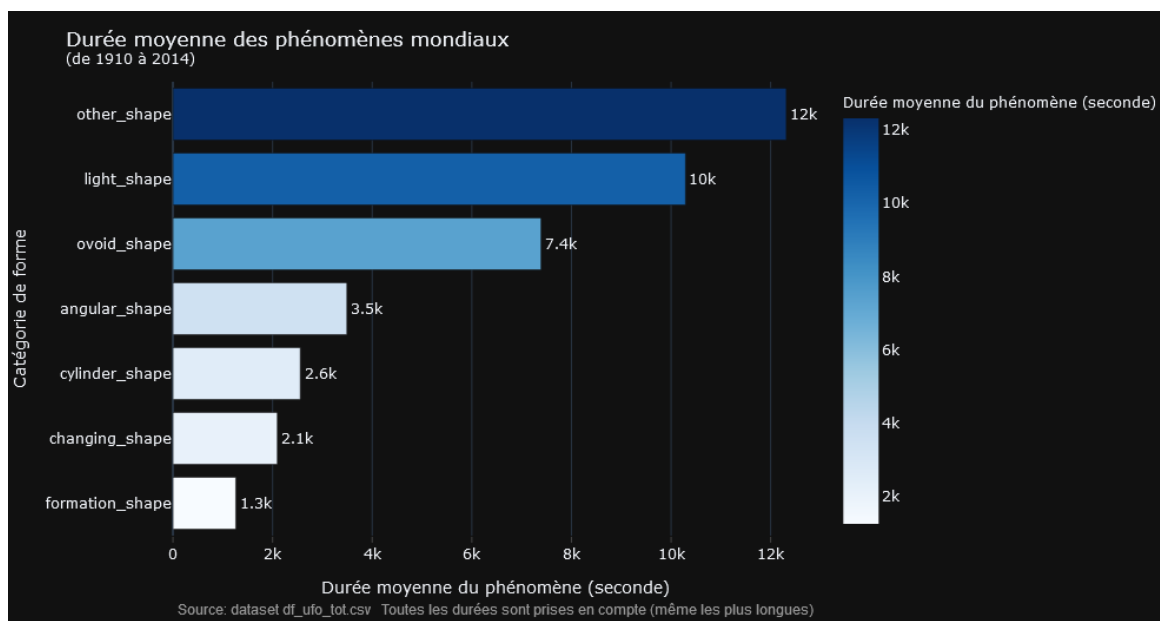


Figure 7: Les formes non identifiées sont les plus persistantes dans le monde

Les phénomènes en formation ont une durée moyenne d'environ 20 minutes. Ce sont les plus brefs et les moins fréquents. On peut résumer les caractéristiques des phénomènes avec cette représentation :

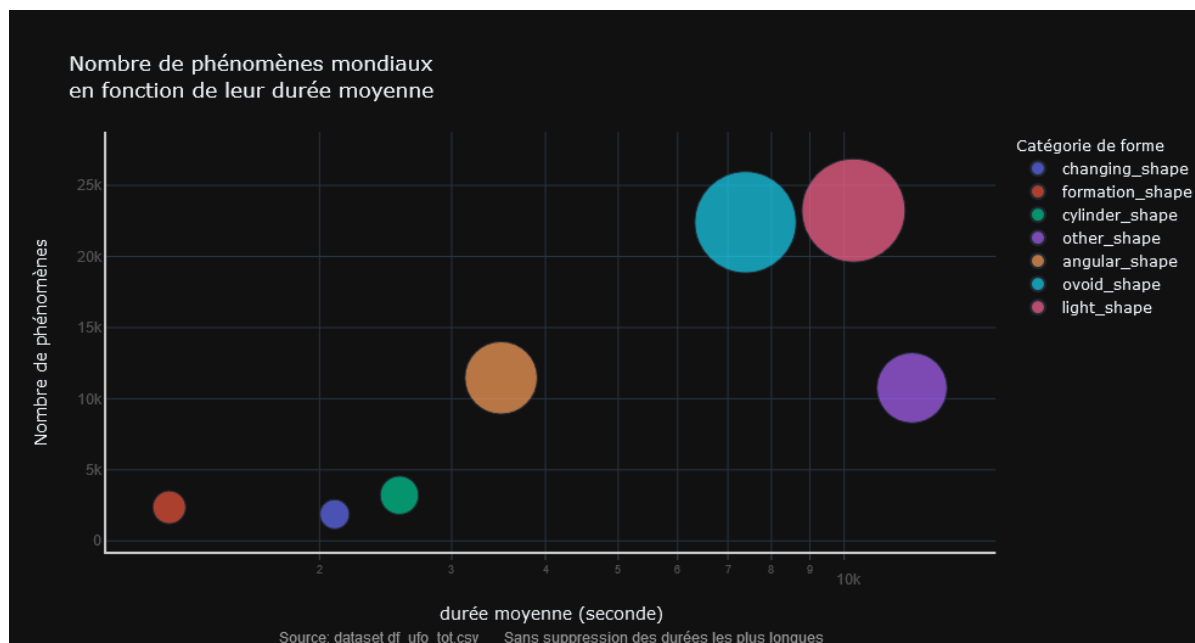


Figure 8: Les formes lumineuses et ovoïdes sont les plus répandues

3.1.2. Analyse temporelles

Des phénomènes plus nombreux selon les années.

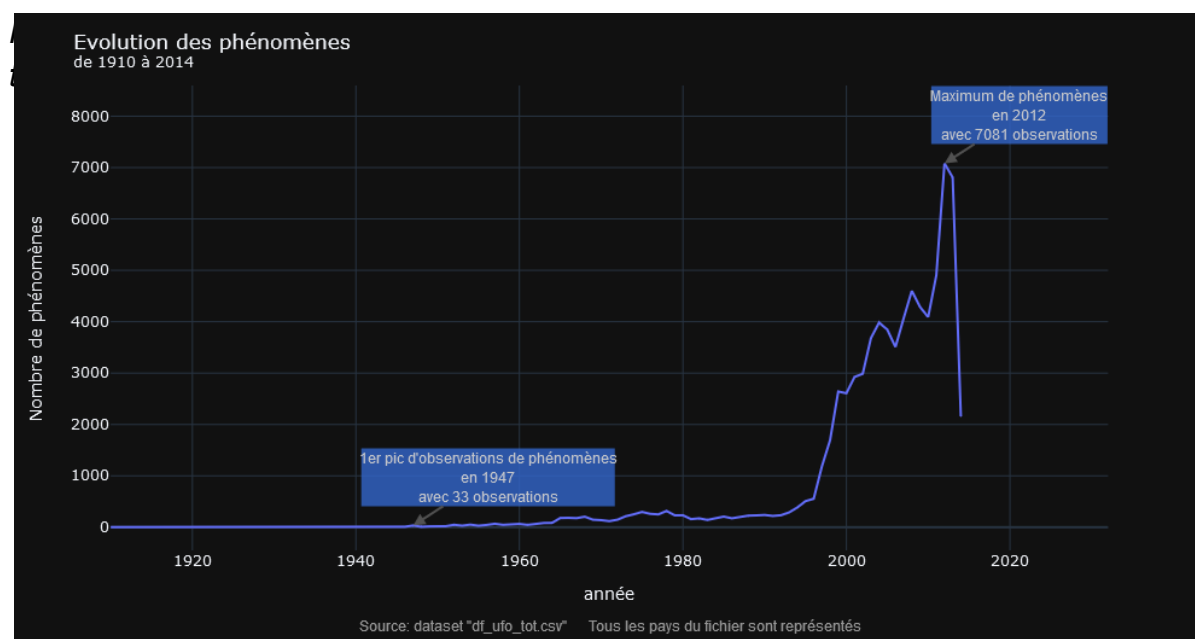


Figure 10: Evolution constante des observations mondiales

A partir de 1947, on note une progression des observations. Cette époque correspond à de fortes tensions liées à la guerre froide. Cette même année est créée par le gouvernement américain le projet Sign (16) suite à de nombreux témoignages d'observations de soucoupes volantes. D'autres projets verront le jour comme, par exemple, le Blue book project en 1952 sous l'impulsion de l'US Air Force (17). Progression constante des observations à partir des années 60. On observe un maximum en 2012, année présumée apocalyptique..., riche en événements astronomiques tel que le passage de Vénus devant le soleil (18).

Des observations saisonnières

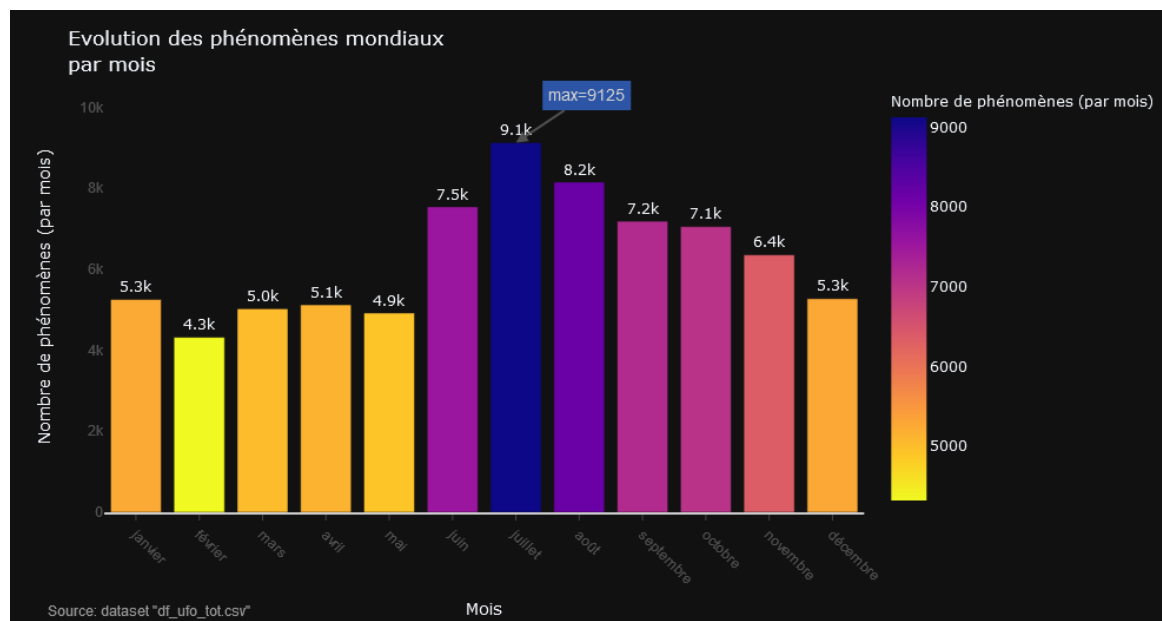


Figure 11: Les phénomènes dans le monde sont plus nombreux de juin à octobre avec un maximum pour le mois de juillet

Les observations sont plus fréquentes en été qu'en hiver avec un pic au mois de juillet. Les beaux jours combinés aux vacances scolaires sont une période propice à l'observation ce qui permet d'augmenter le nombre de signalements. De plus, de nombreux événements astronomiques telles les étoiles filantes, comètes...(19) incitent les populations à être plus attentives.

Des phénomènes plutôt nocturnes

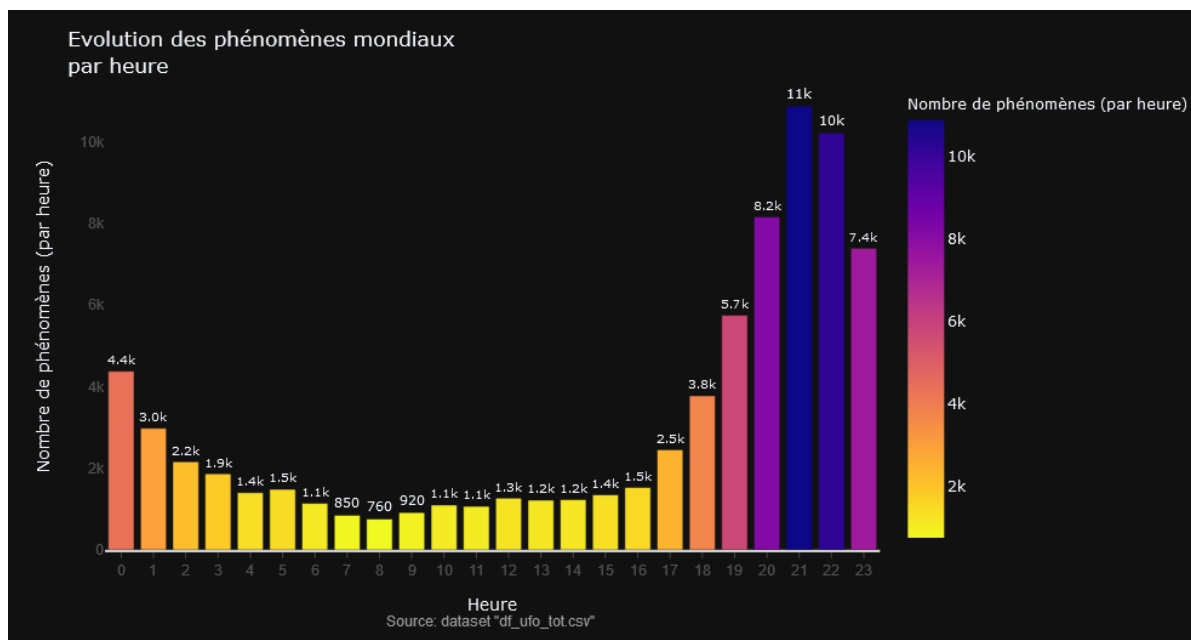
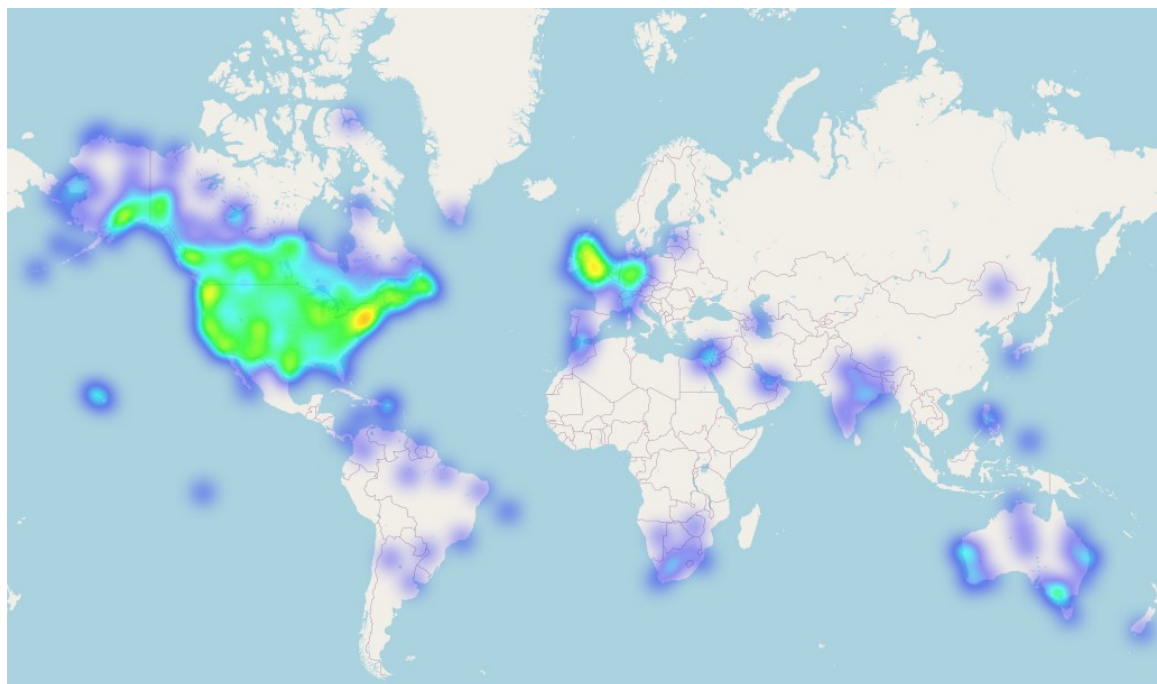


Figure 12: Les phénomènes mondiaux apparaissent davantage à partir de 19h00 avec un maximum à 21h00

On observe un maximum d'observations vers 21 heures et jusqu'à 22 heures. Le minimum se situe en début de matinée (vers 8 heures). Les phénomènes lumineux (majoritaires) sont plus repérables la nuit.

3.1.3. Analyse géographique

Des phénomènes répartis principalement sur le continent américain et le Royaume- Uni



Carte 1: Les phénomènes mondiaux sont répartis principalement dans l'hémisphère nord. Les zones jaunes orangées sont les plus denses

On peut se demander si le fichier initial est orienté intentionnellement sur ces deux pays ou si les observations y sont réellement plus fréquentes.

Pour le reste de l'étude, les USA, le Canada et Royaume-Uni seront les pays retenus car très bien représentés.

4. ANALYSE DES DONNÉES : USA, CANADA ET ROYAUME-UNI

Le fichier comprend 73872 observations et 9 variables. Préparation du fichier à l'étude. Création de variables temporelles (year, month, hour). Création de sept catégories de forme (20). On remarque un phénomène d'une durée de 31 ans . A l'évidence, il s'agit d'outliers (ou valeurs aberrantes) qui sont analysées.

4.1. Étude et traitement des outliers

Un phénomène d'une durée supérieure à 24h soit 86400 secondes est considéré comme outlier.

Les formes ovoïdes les plus fréquentes pour les phénomènes outliers

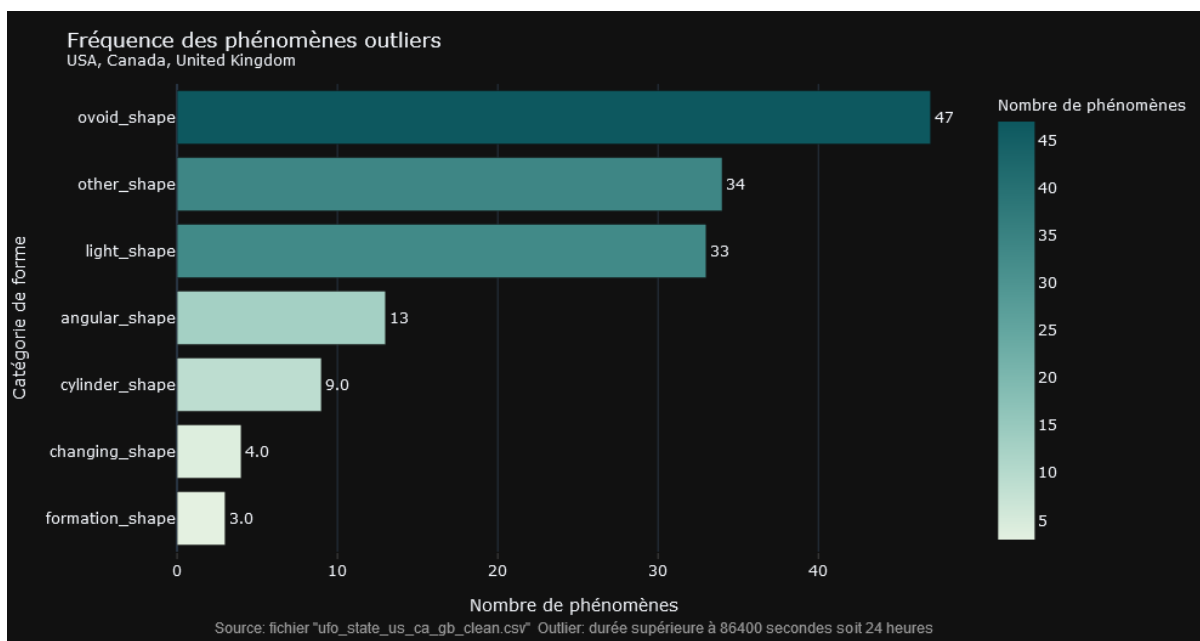


Figure 13: La catégorie de forme ovoïde est la catégorie la plus représentée pour les durées supérieures à la journée

Beaucoup de formes non identifiées correspondent à des outliers.

Les formes lumineuses les plus persistantes.

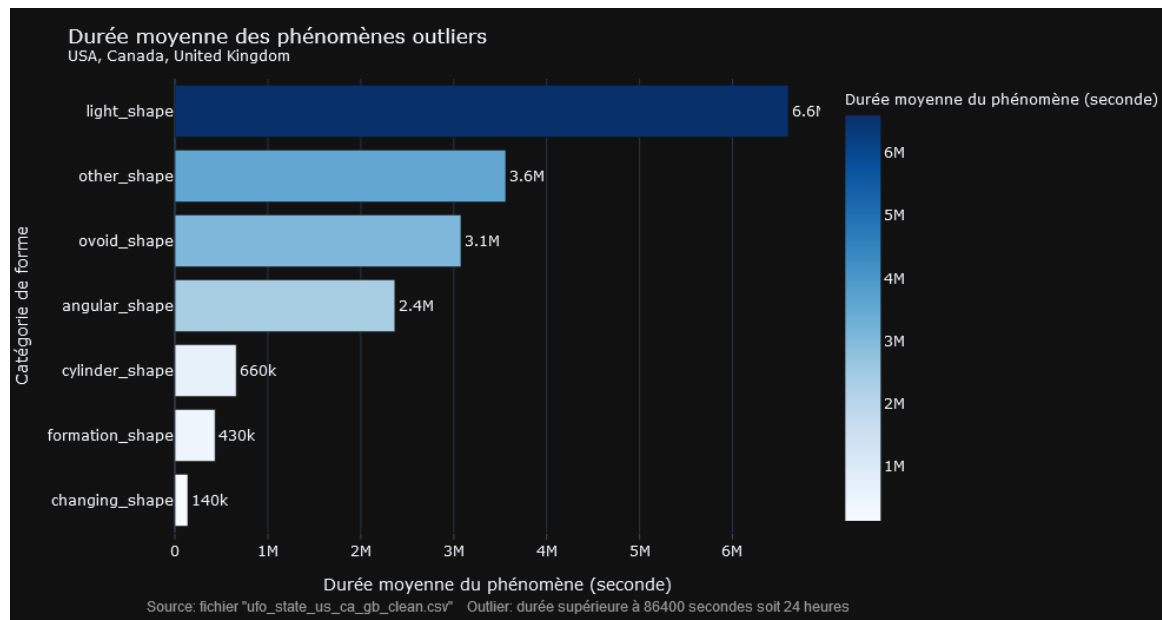


Figure 14: Pour les durées supérieures à la journée, les formes lumineuses sont celles qui sont les plus persistantes

Les phénomènes inconnus sont ensuite les plus longs. Compte tenu de la durée de plus d'une journée, il est normal que de nombreux phénomènes soient inexpliqués. Pour l'étude, seules les **durées inférieures à 30 minutes seront retenues**. Cette durée semble cohérente pour écarter toute observation mal répertoriée.

4.2. Étude et traitement sans outliers

Pour l'étude, seules les durées inférieures ou égales à 1800 secondes soit 30 minutes seront prises en compte. Le nouveau fichier (sans outliers) comporte 67274 observations et 15 variables.

4.2.1. Analyse univariée

Un test de Kolmogorov-Smirnov rejette l'hypothèse H_0 de normalité de la variable "length_of_encounter_seconds". Une transformation logarithmique de la variable en "log_length_of_encounter_seconds" ne permet pas de ne pas rejeter l'hypothèse H_0 de normalité du test. Ces variables peuvent être néanmoins considérées comme gaussiennes du fait du nombre important d'observations.(20)

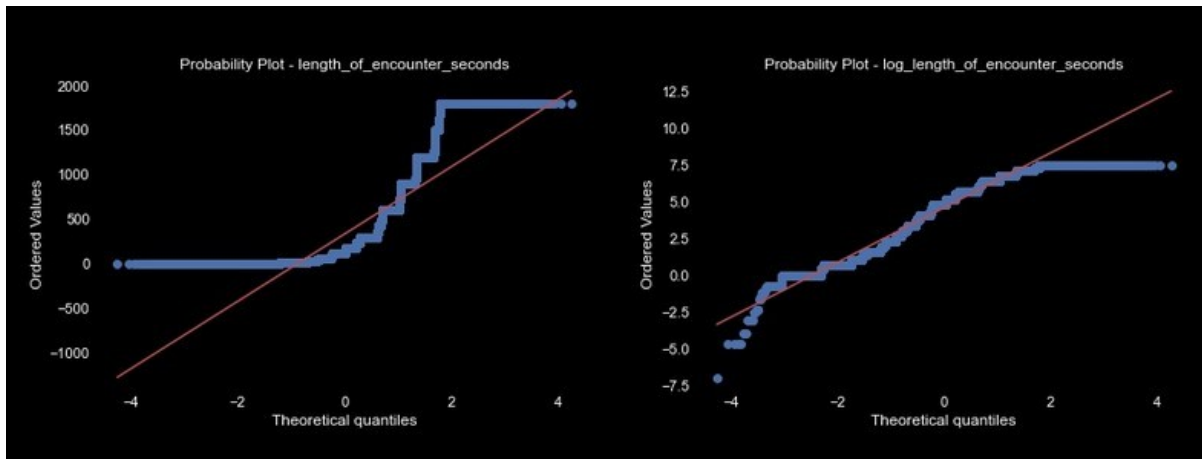


Figure 15: Des variables considérées comme gaussiennes au vu du nombre d'observations

4.2.2. Analyses bivariées

Analyse de variable comprenant les catégories de forme et la variable relative à la durée du phénomène. (20)

Un test d'ANOVA nous permettra de répondre à la question suivante : La durée d'un phénomène dépend-elle de sa catégorie de forme ?

Un test du χ^2 permettra d'établir les corrélations de la catégorie de forme des phénomènes avec les différentes variables qualitatives que sont le pays d'origine, les variables temporelles (mois, jour, année...).

Des phénomènes qui durent plus ou moins selon leur forme

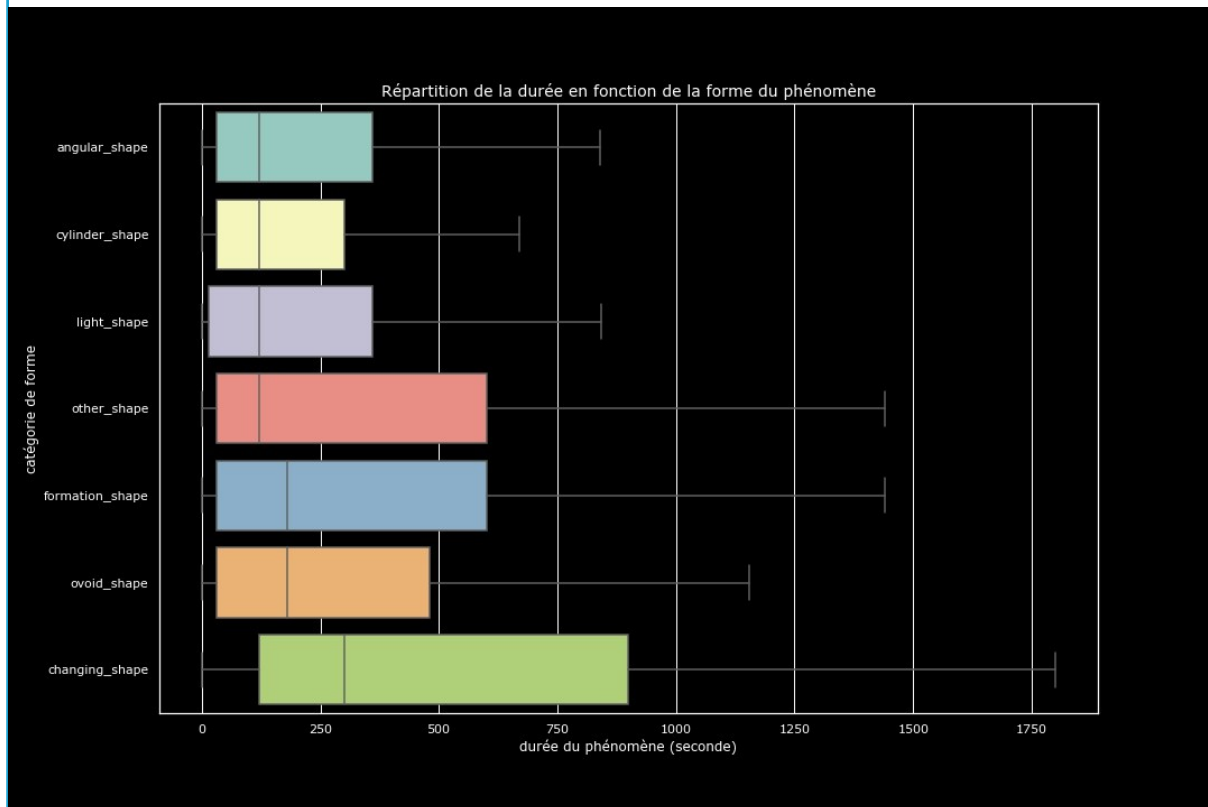


Figure 16: La distribution des durées (sans les outliers) des phénomènes différent selon leur forme

Les phénomènes cylindriques et lumineux sont des phénomènes les plus "constants" et les plus courts.

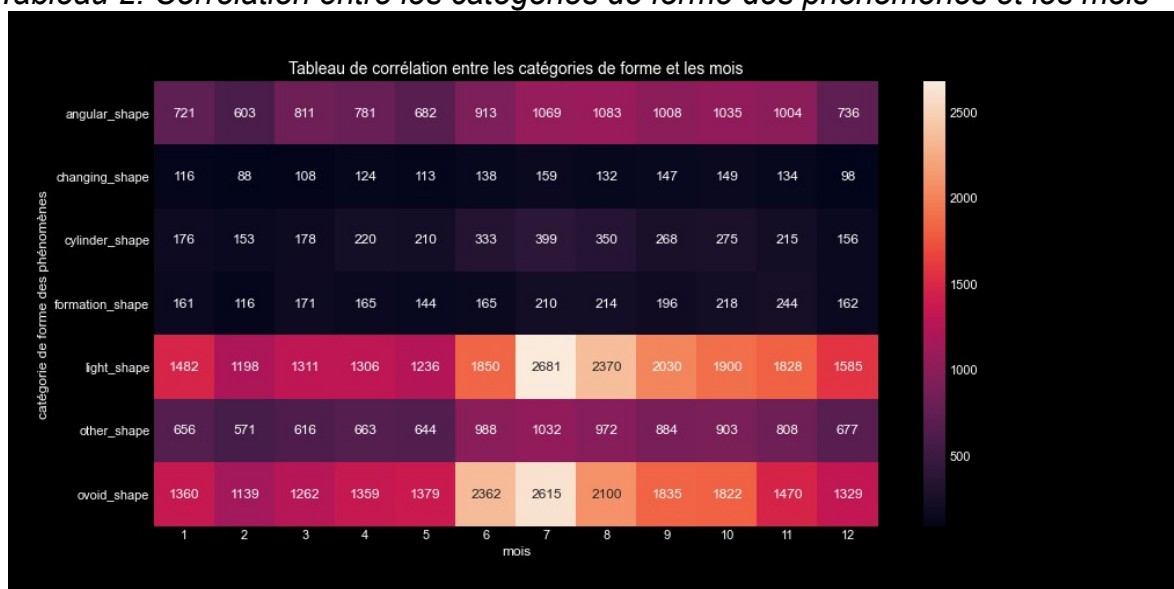
Un test d'ANOVA nous permet de conclure que la durée des phénomènes diffèrent selon leur forme et selon le pays. Un test du chi2, permet de conclure que les formes sont différentes selon le pays, la durée et la temporalité (20). L'étude des différentes corrélations est détaillée au chapitre 4.

Tableau 1: Corrélation entre les catégories de forme des phénomène et les pays



Les catégories de forme dépendent de la temporalité ici le mois.

Tableau 2: Corrélation entre les catégories de forme des phénomènes et les mois



4.2.3. Étude générale des formes des phénomènes

Les phénomènes lumineux et ovoïdes majoritaires dans les USA, le Canada et le Royaume-Uni confondus. Mais les formes changeantes sont les plus persistantes.

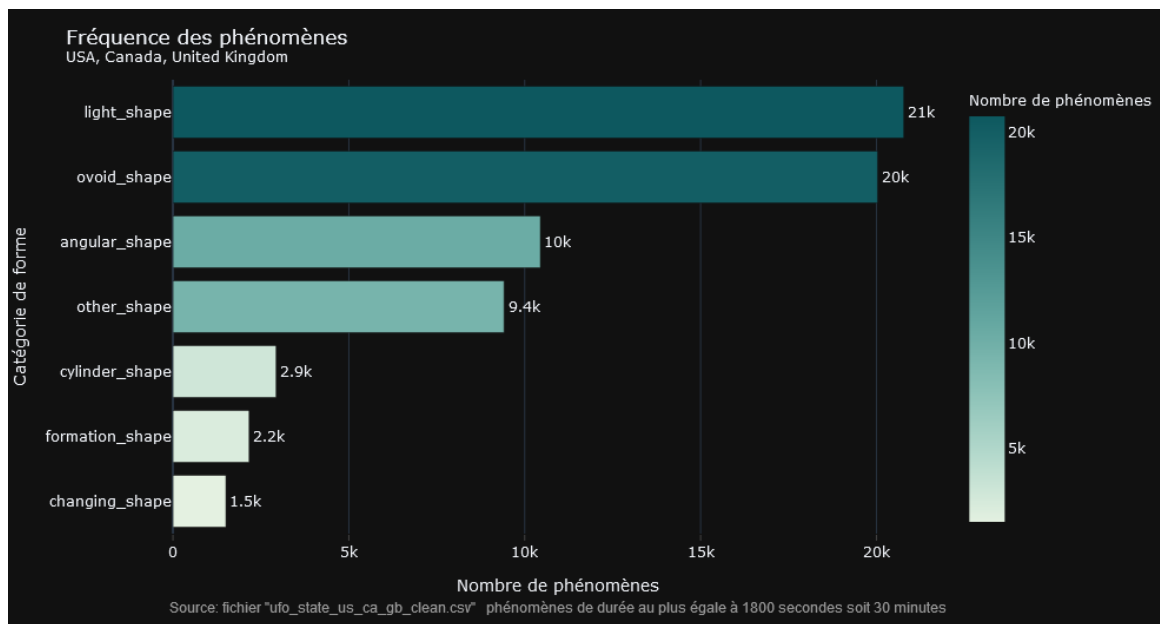


Figure 17: Les formes lumineuses et ovoïdes sont majoritaires pour les 3 pays

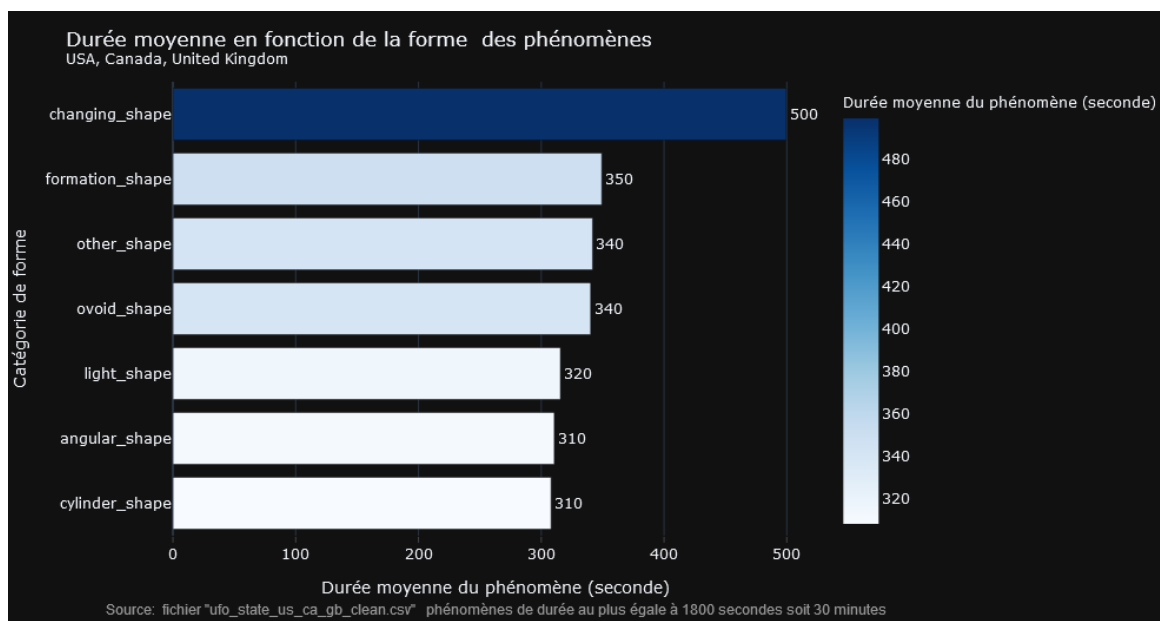


Figure 18: Seules les formes changeantes sont plus longues

Après découpage des durées en intervalles.

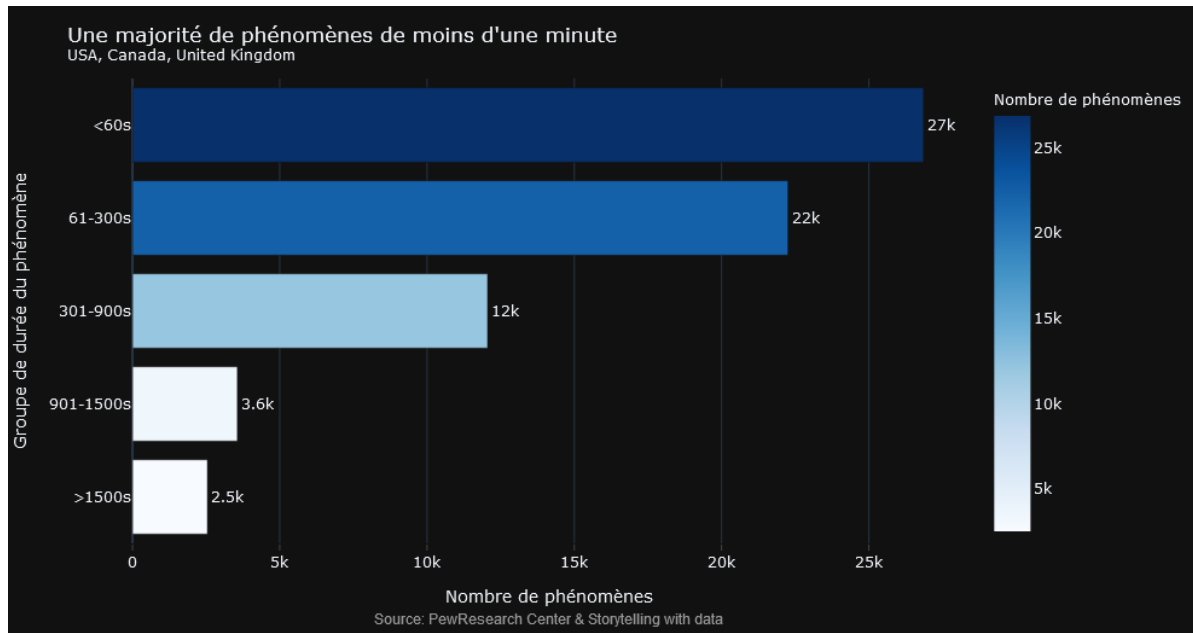


Figure 19: Une majorité de phénomènes de moins d'une minute

Les phénomènes durent en majorité moins de 60 secondes. En minorité, nous retrouvons les phénomènes les plus longs (d'une durée supérieure à 25 minutes).

5. ÉTUDE DES PHÉNOMÈNES EN FONCTION DU PAYS

5.1. Étude des phénomènes

Les fréquences des formes de phénomènes diffèrent selon les pays.

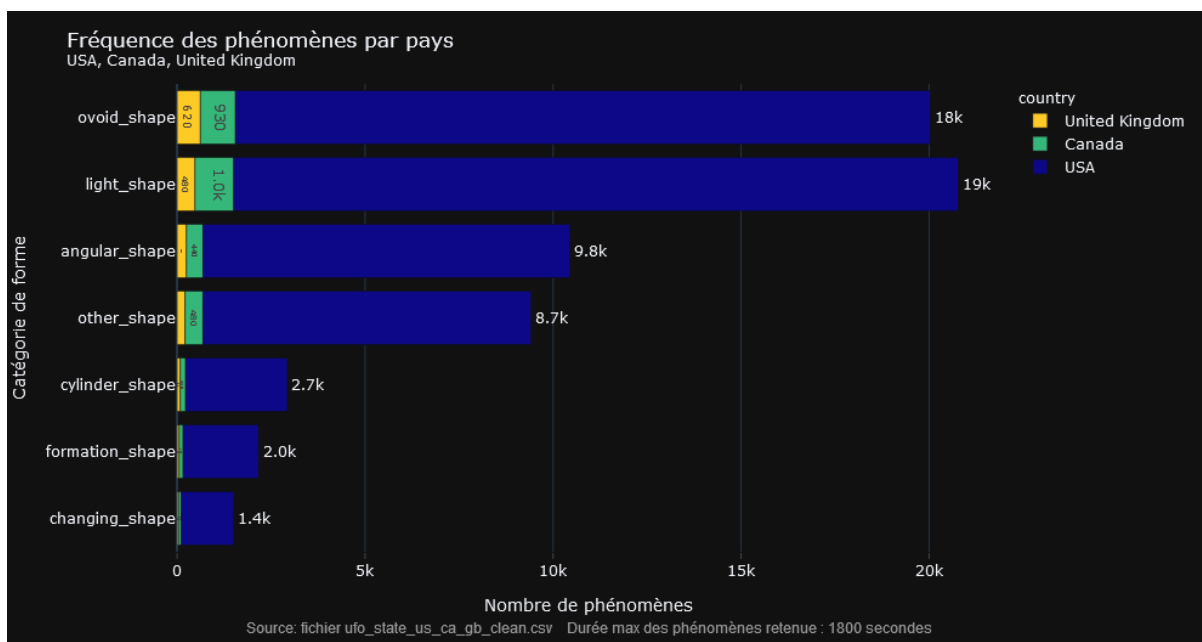


Figure 20: Les formes ovoïdes et lumineuses sont les mieux représentées aux USA, au Canada et au Royaume-Uni

Les phénomènes ovoïdes sont plus fréquents au Royaume-Uni. Les phénomènes lumineux sont majoritaires aux USA et au Canada.

La durée des phénomènes est variable selon le pays

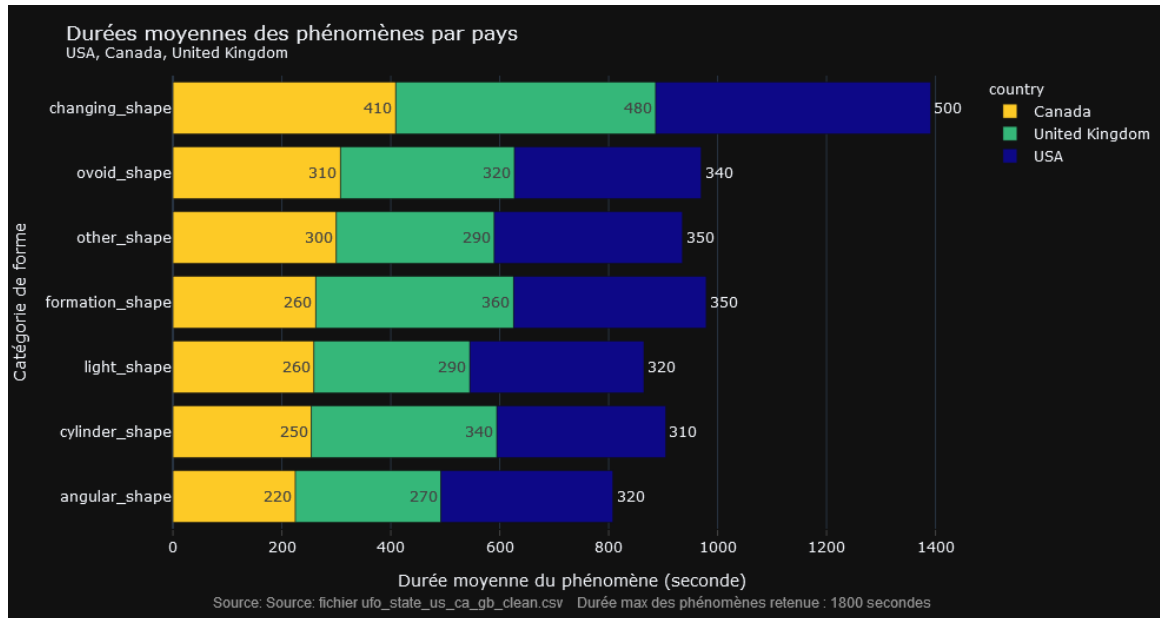


Figure 21: Les durées des phénomènes dépendent de leur forme et du pays

Les phénomènes sont plus courts au Canada.

5.1.1. Analyse temporelle

1995, l'année où le nombre d' observations devient important.

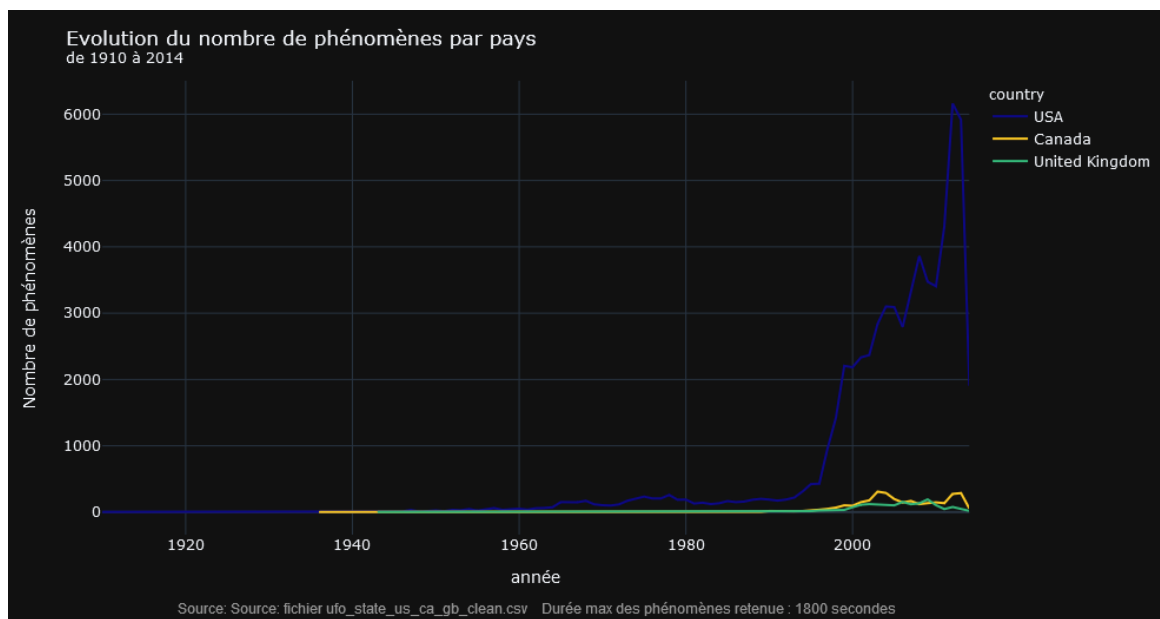


Figure 22: Evolution du nombre de phénomènes plus importante aux USA

Il faudra attendre les années 2000 pour remarquer un frémissement des observations au Canada et au Royaume-Uni puis 2003 pour un accroissement plus important. En 2012, on note une forte augmentation des observations dans les 3 pays. Recrudescence des observations pour le Canada en 2003 et pour le Royaume-Uni en 2009. Les phénomènes se répartissent en fonction du temps de la même manière quelque soit le pays concerné.

Des observations saisonnières

Afin de comparer équitablement l'apparition des phénomènes dans les trois pays d'étude, une nouvelle variable "prop_ufo %" est créée. Elle correspond à la proportion d'un phénomène donné sur l'ensemble des phénomènes recensés dans un pays donné (20).

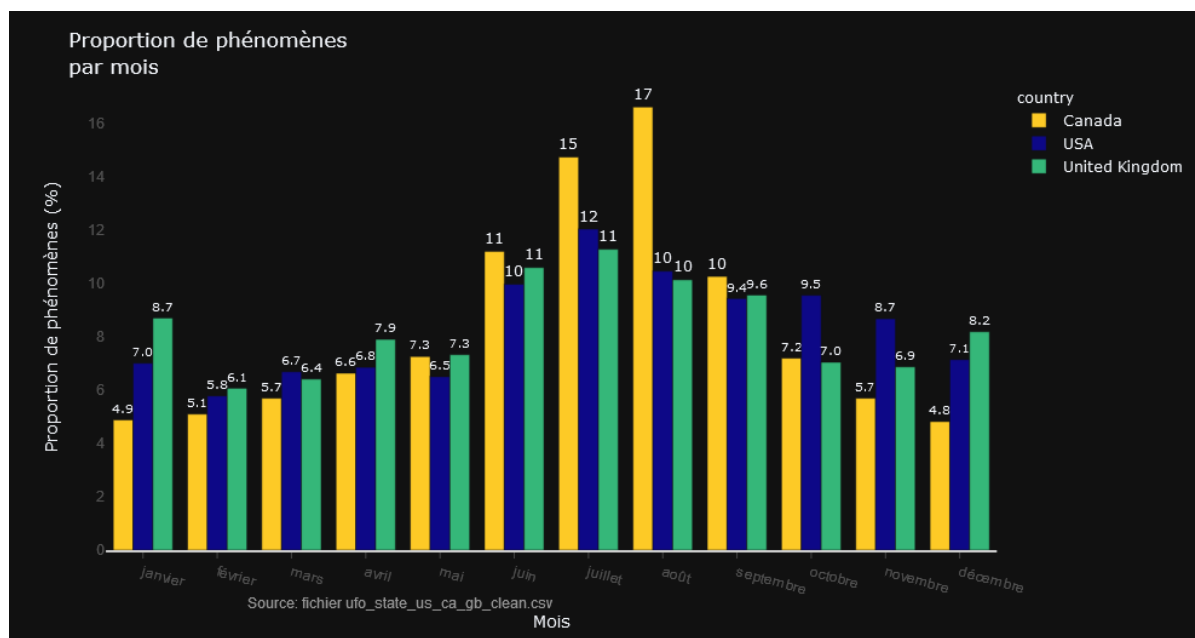


Figure 23: La proportion de phénomènes est la plus importante au Canada durant l'été. L'hiver est plus favorable aux phénomènes anglo-saxons

Comme la tendance mondiale, les USA, le Canada et le Royaume-Uni voient leur nombre d'observation augmenter les mois d'été. Des observations majoritaires en fin de semaine pour les trois pays

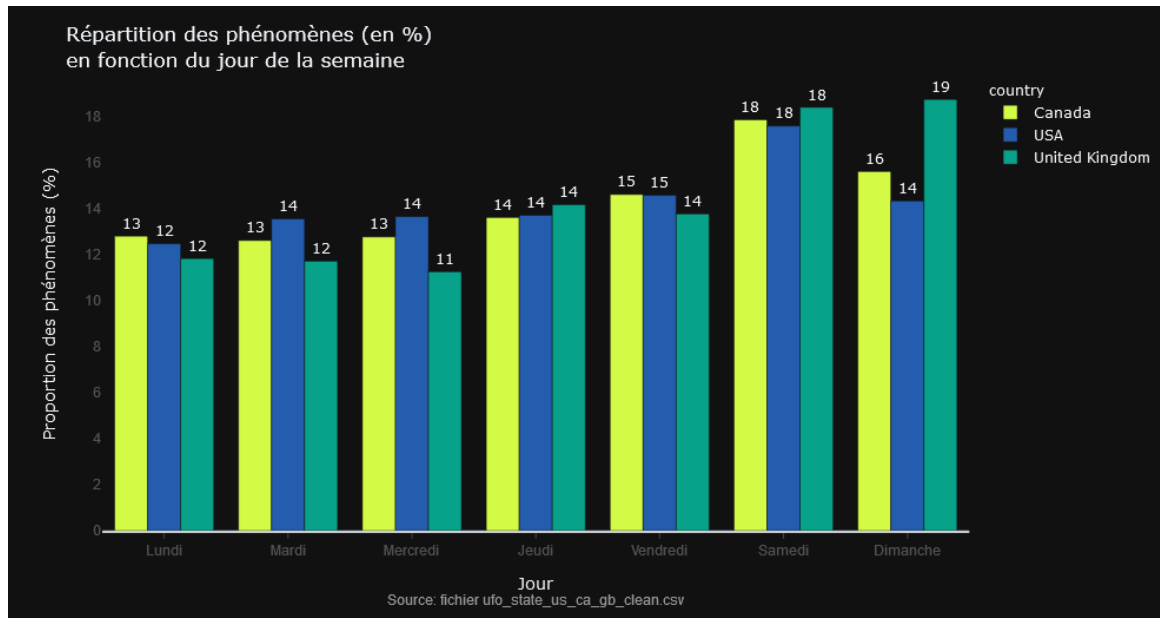


Figure 24: Le samedi et le dimanche favorables aux phénomènes

Les sorties du week-end sont plus propices à l'observation du ciel.

Des phénomènes plus nocturnes que diurnes

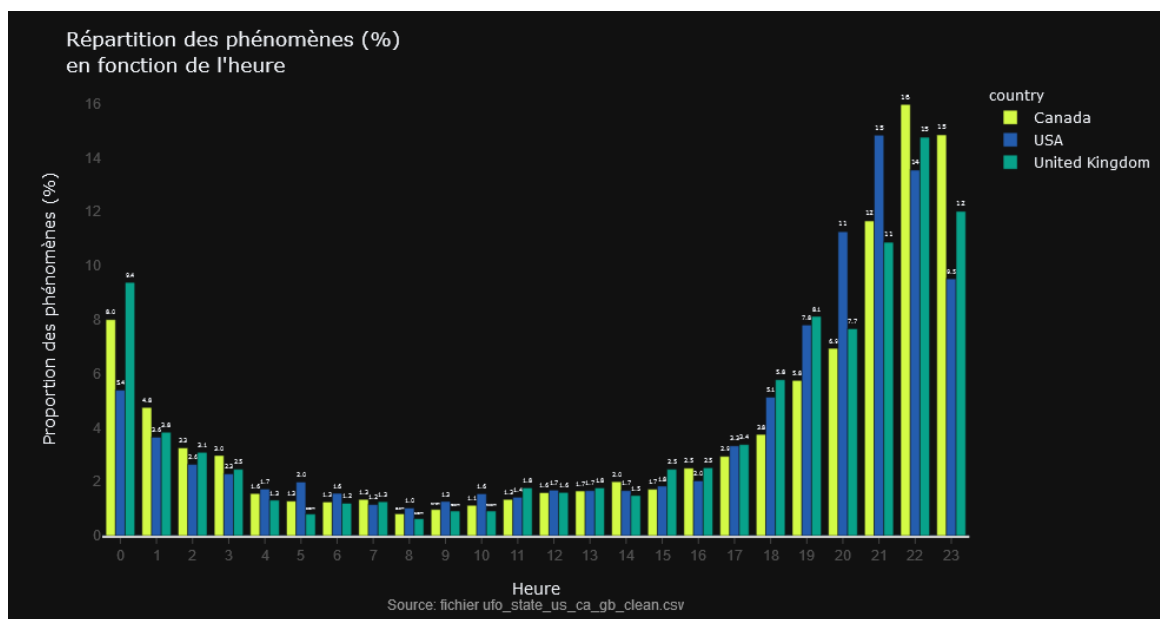


Figure 25: Répartition journalière des phénomènes quelque soit le pays considéré

Avec un minimum d'observations entre 7h00 et 16h00, puis un maximum entre 21h00 et 22h00, la répartition des phénomènes suit la même tendance quelque soit le pays.

La proportion des formes des phénomènes diffère selon les années.

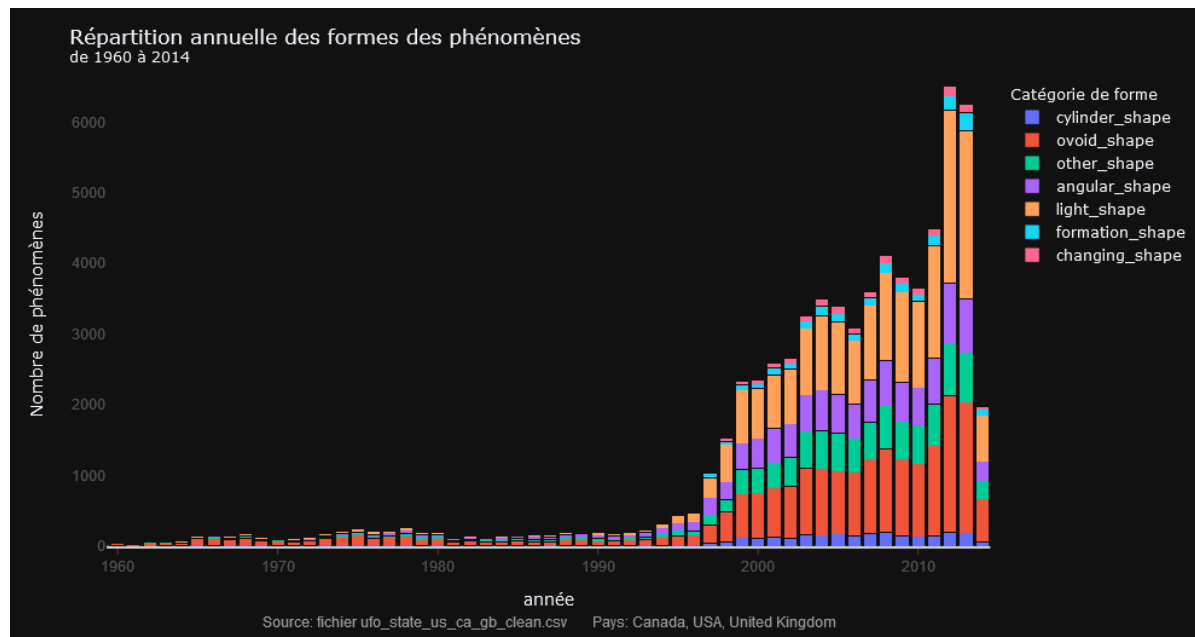


Figure 26: Accroissement du nombre de forme des phénomènes en particulier pour les formes ovoïdes et lumineuses

Si durant les années 60, les formes ovoïdes prévalent, à partir de 1995, les phénomènes lumineux ou cylindriques lui volent la vedette.

De nouvelles technologies sont mises en place et sont sources de phénomènes lumineux. Par exemple, en 1997, le système Iridium est placé en orbite terrestre basse (21). Les satellites le composant sont souvent sources de flash lumineux, les rayons du soleil se reflétant dans les satellites.

Les phénomènes inexplicables deviennent de plus en plus fréquents.

Un excellent site internet recense toutes ces "méprises du ciel" (15).

Les mois d'été favorisent les phénomènes lumineux et les ovoïdes.

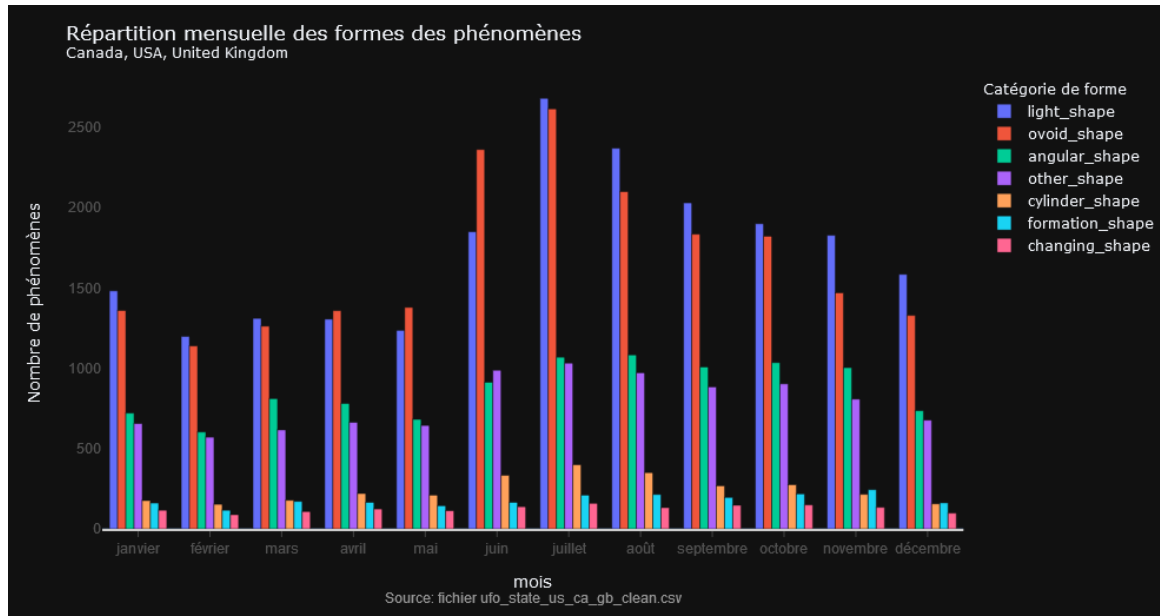


Figure 27: On observe la même distribution des catégories de forme des phénomènes mais leur nombre diffère selon le mois

On pourrait alors conclure que l'éclairement a un impact sur la forme des phénomènes...

Les formes lumineuses et ovoïdes plus fréquentes à partir de 17h00.

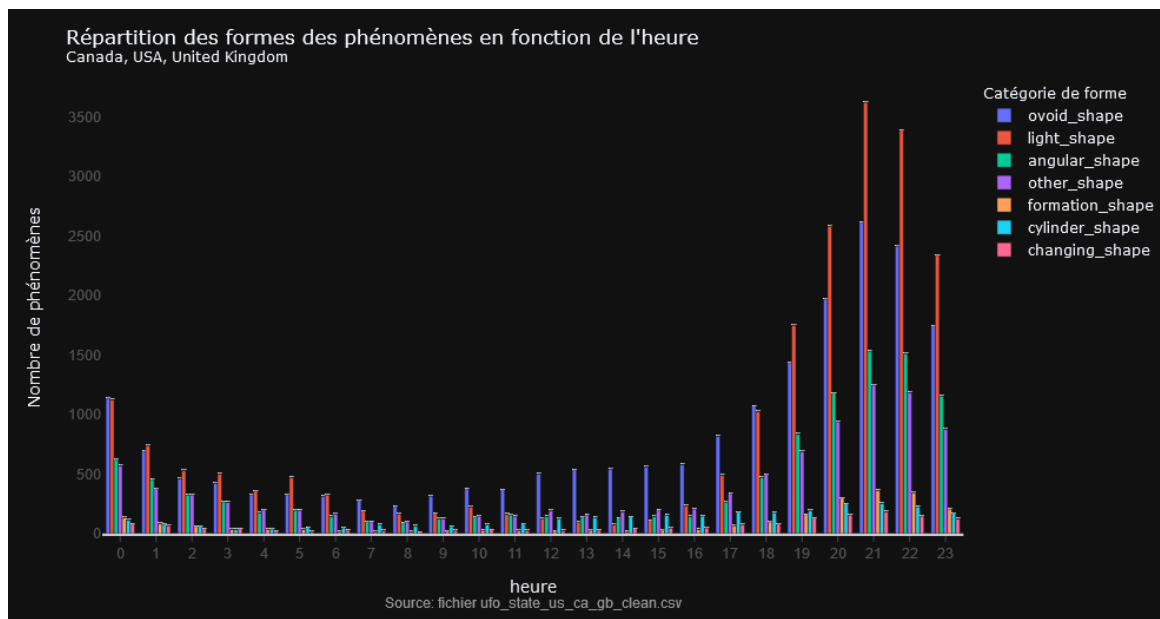


Figure 28: Le nombre de phénomènes s'accroît en soirée

Elles sont également prépondérantes le samedi

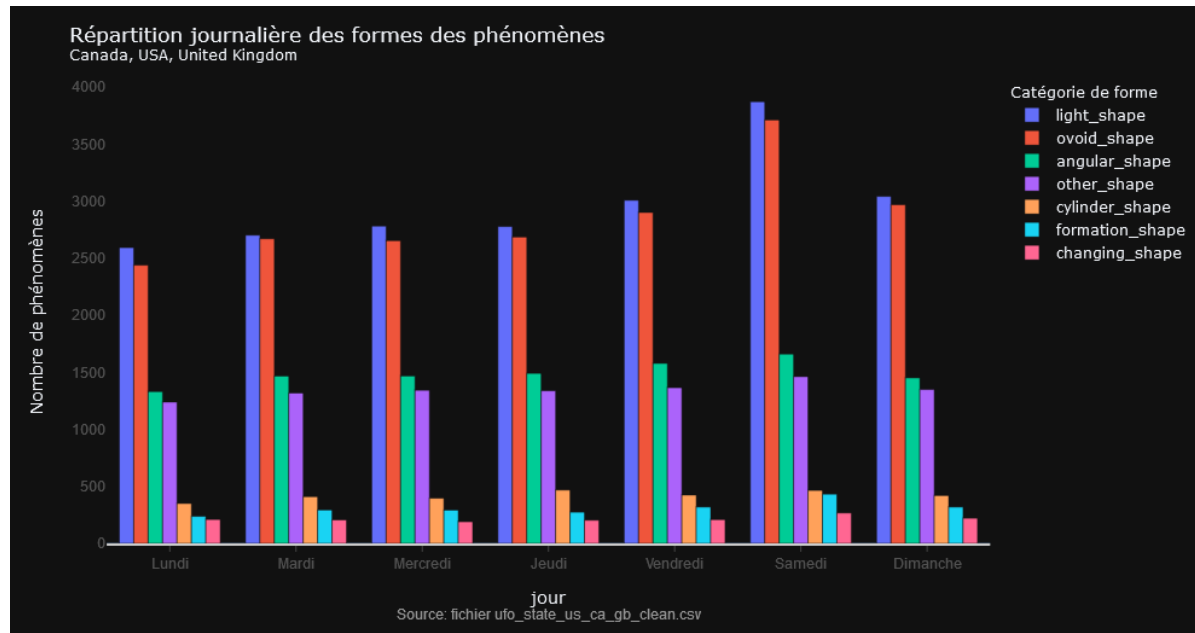


Figure 29: La plupart des formes des phénomènes n'évoluent pas selon le jour. Seules les formes ovoïdes et lumineuses semblent plus nombreuses le samedi

5.2. Influence de la population sur les phénomènes

Il n'y pas d'observation sans témoin. Les zones les plus peuplées vont naturellement correspondre au nombre d'observation le plus élevé. Il faut donc établir un ratio du nombre d'observations par rapport au nombre d'habitants pour établir une étude plus fine des observations. Pour cela, le fichier contenant le nombre d'habitants par état/province pour les USA, le Canada et le Royaume-Uni est importé. Il se nomme pop_us_ca_gb_clean.csv. Il est fusionné au fichier d'étude. Ainsi, la variable correspondant à la proportion d'observations pour 100 000 habitants est introduite. Elle sera utilisée dans les analyses (23).

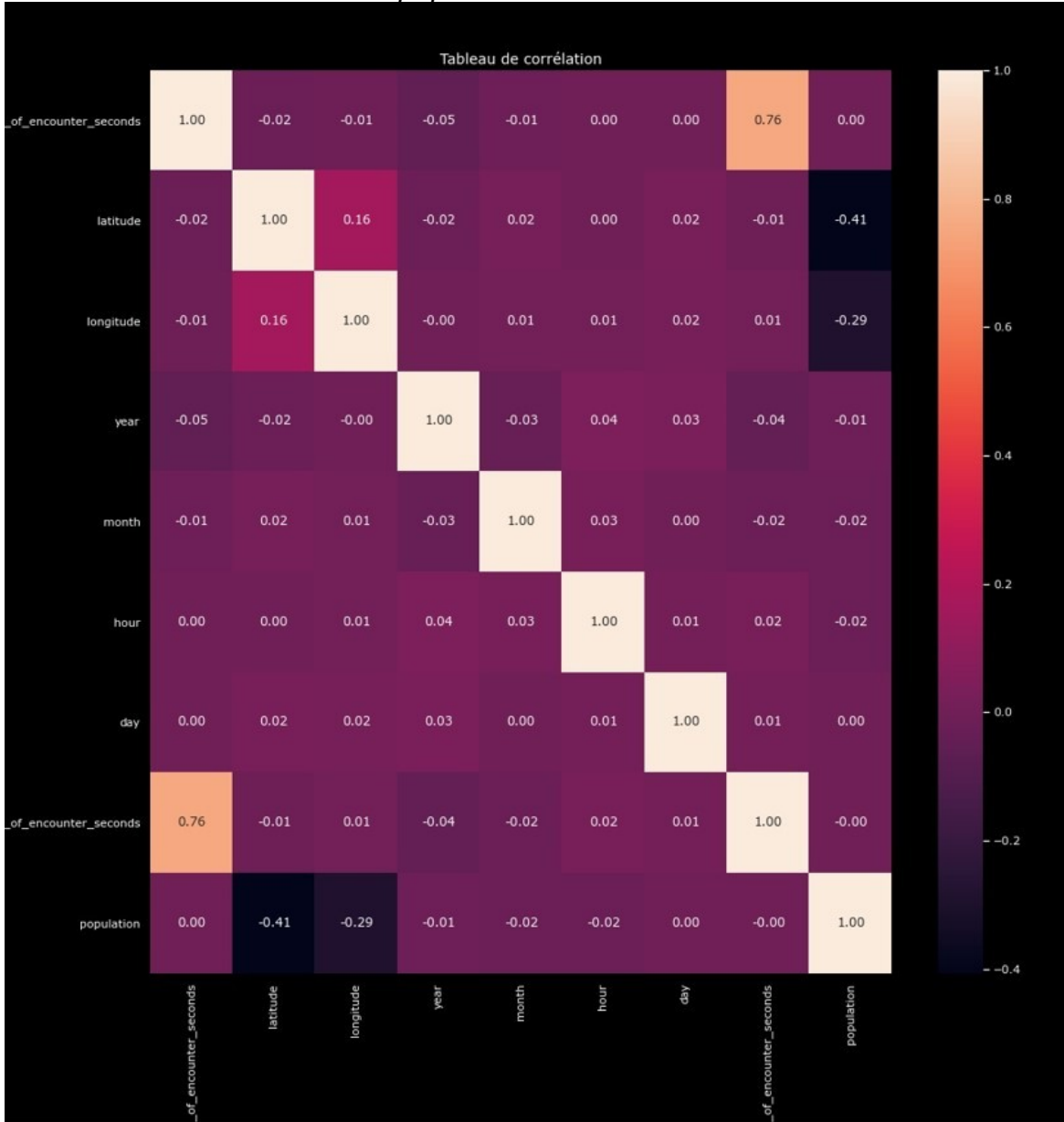
5.2.1. Analyse univariée

Selon le test de Kolomogorov-Sminov, aucune variable ne suit une loi normale. Compte tenu du nombre d'observations, nous pouvons tout de même estimer ces variables comme gaussiennes.

5.2.2. Analyse bivariable

Un test d'ANOVA nous confirme que le nombre d'habitants influe sur les phénomènes. Un test de corrélation selon Pearson confirme une relation entre la latitude et le nombre d'habitants.

Tableau 3: Corrélation entre la population et la latitude



5.2.3. Étude des phénomènes en incluant l'impact de la population

Disparité du nombre de phénomènes selon l'État concerné.

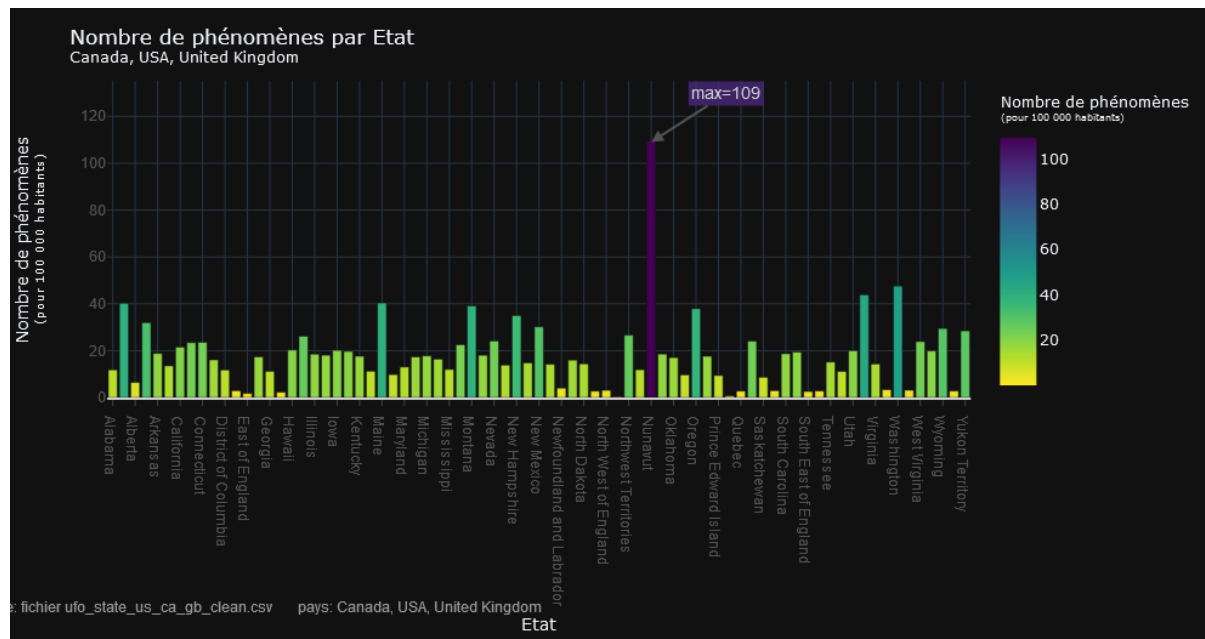


Figure 30: Le nombre de phénomènes (pour 1000 000 habitants) varie en fonction de l'État considéré avec un maximum pour le territoire de Nunavut

On remarque un nombre d'observations important pour le Nunavut.

Étude spécifique du Nunavut

Le Nunavut est un territoire de plus de 2 millions de km² et se situe dans le nord canadien. Il ne compte que 39 536 habitants. Une base militaire canadienne et une station météorologique (Eureka) y sont implantées (24). On y remarque une très forte concentration des phénomènes.



Figure 31: De nombreux phénomènes sur le territoire Nunavut

Fréquence des phénomènes au Nunavut

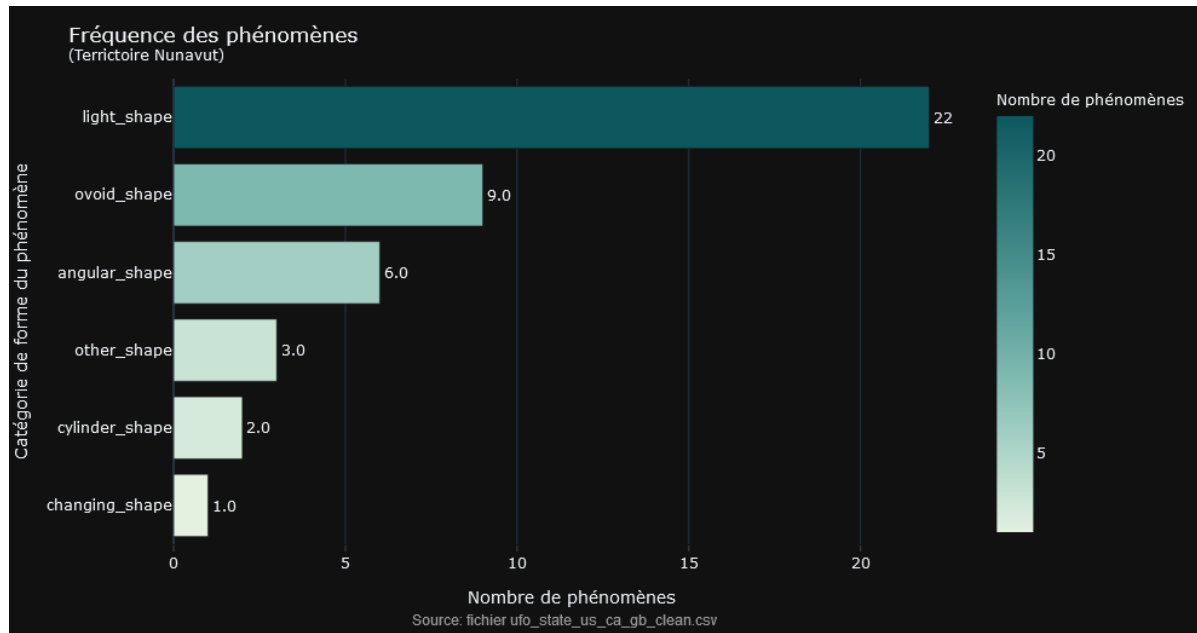


Figure 32: Les formes lumineuses prédominent sur le territoire Nunavut

Les phénomènes lumineux sont les plus fréquents. Le territoire étant proche du cercle arctique, de nombreux phénomènes lumineux naturels telles que les aurores boréales, parhélies(25) s'y manifestent.

Durée des phénomènes au Nunavut

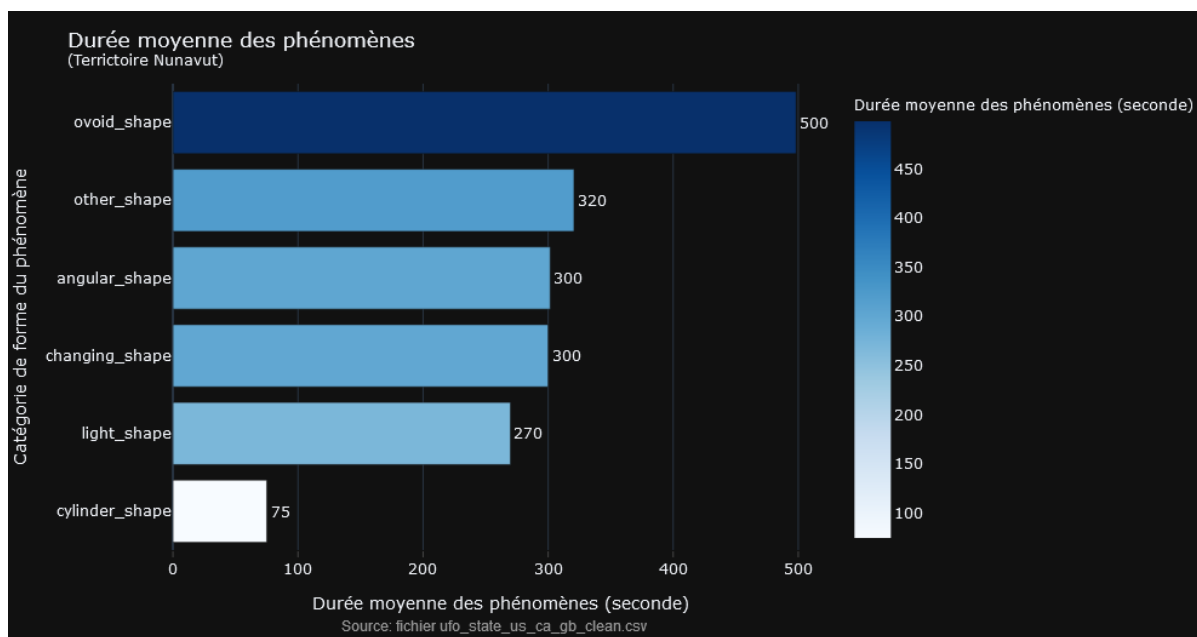
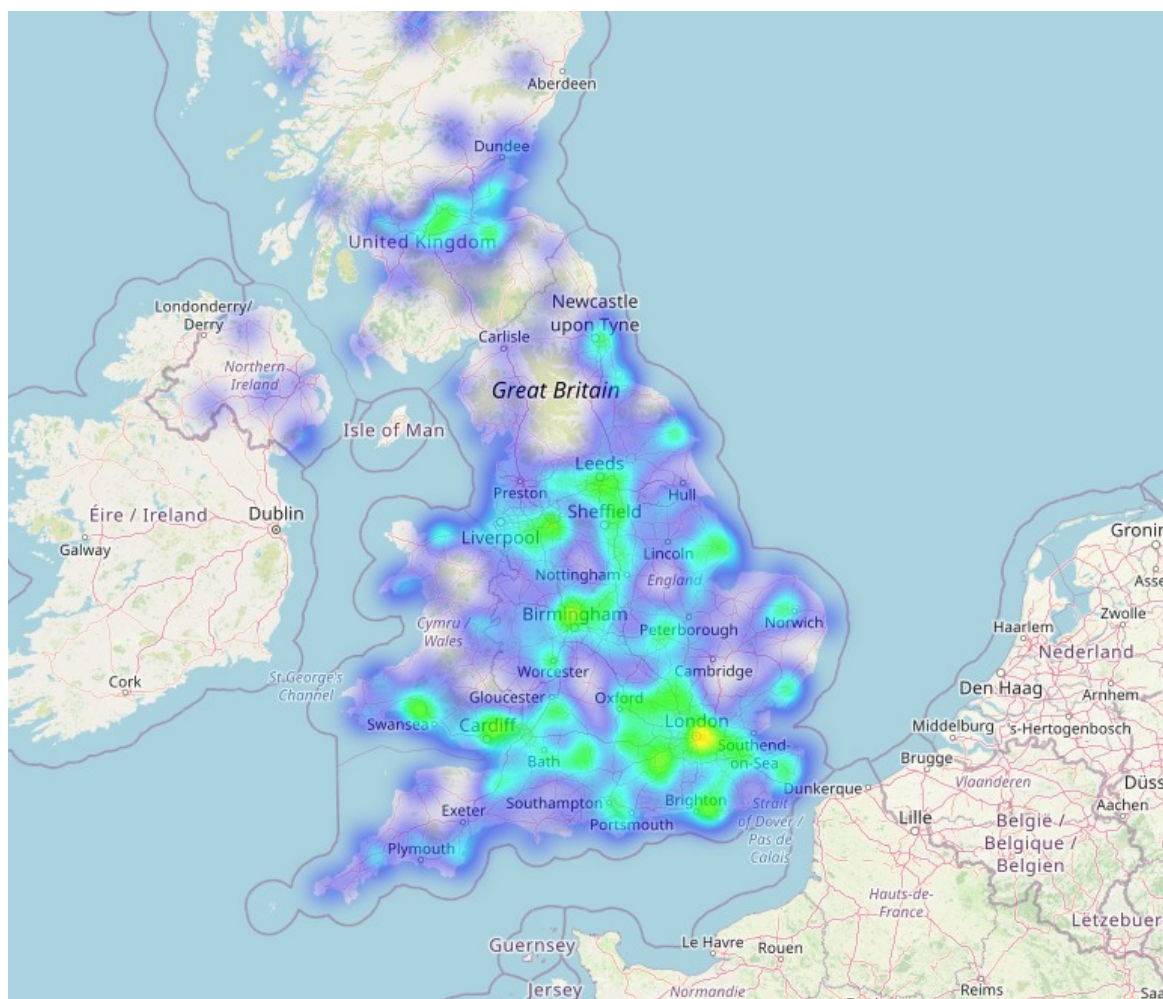


Figure 33: Les formes ovoïdes de plus de 8 minutes sont les plus persistantes

Les phénomènes ovoïdes sont les plus persistants. Souvent ce sont des phénomènes météorologiques comme des nuages lenticulaires (16). Les formes cylindriques sont les plus fugaces.

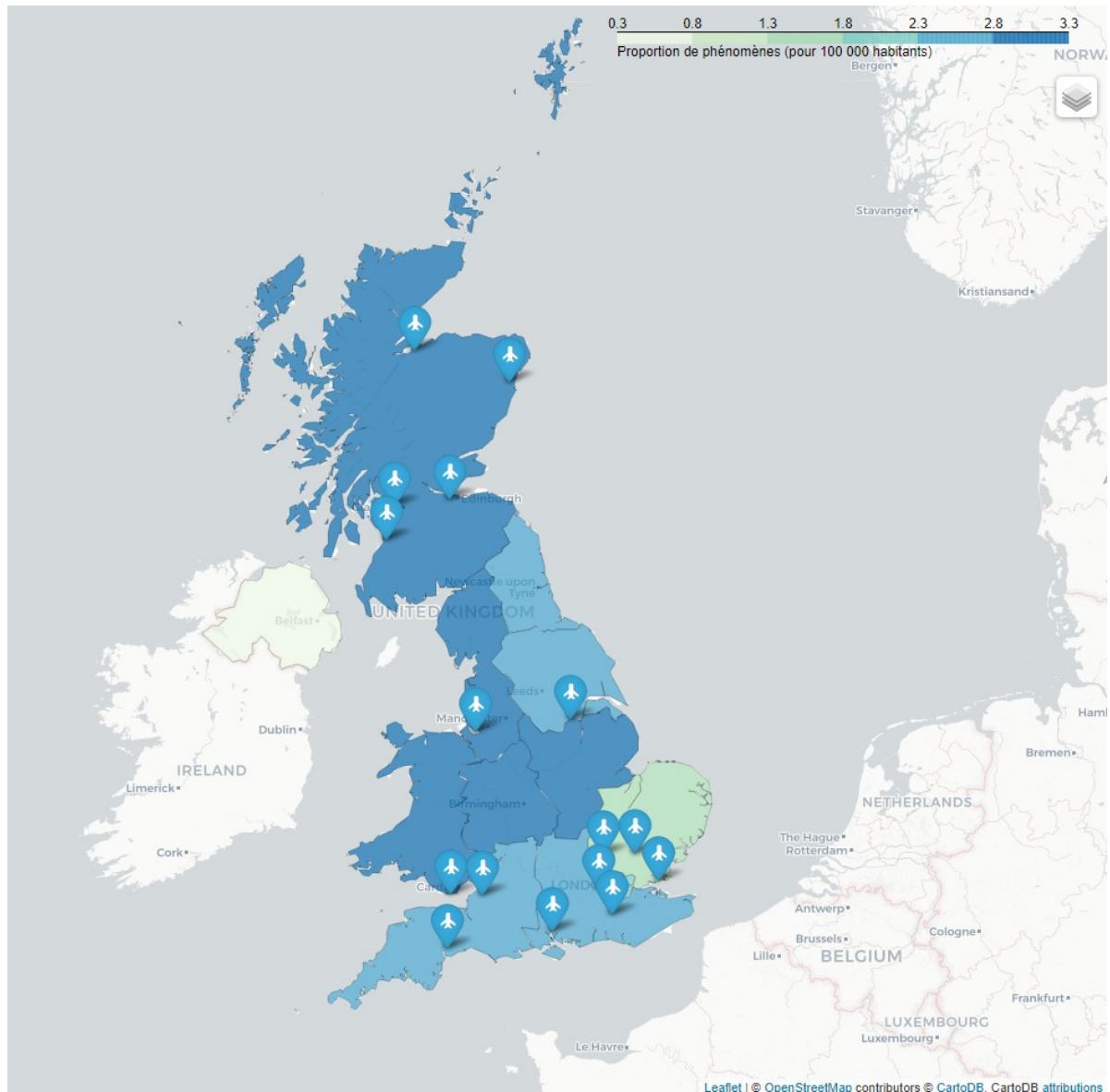
Étude des phénomènes aux USA, Royaume-Uni et Canada

Des phénomènes, en général, concentrés autour des grandes villes.



Carte 2: Exemple de répartition des phénomènes. Ici, au Royaume-Uni

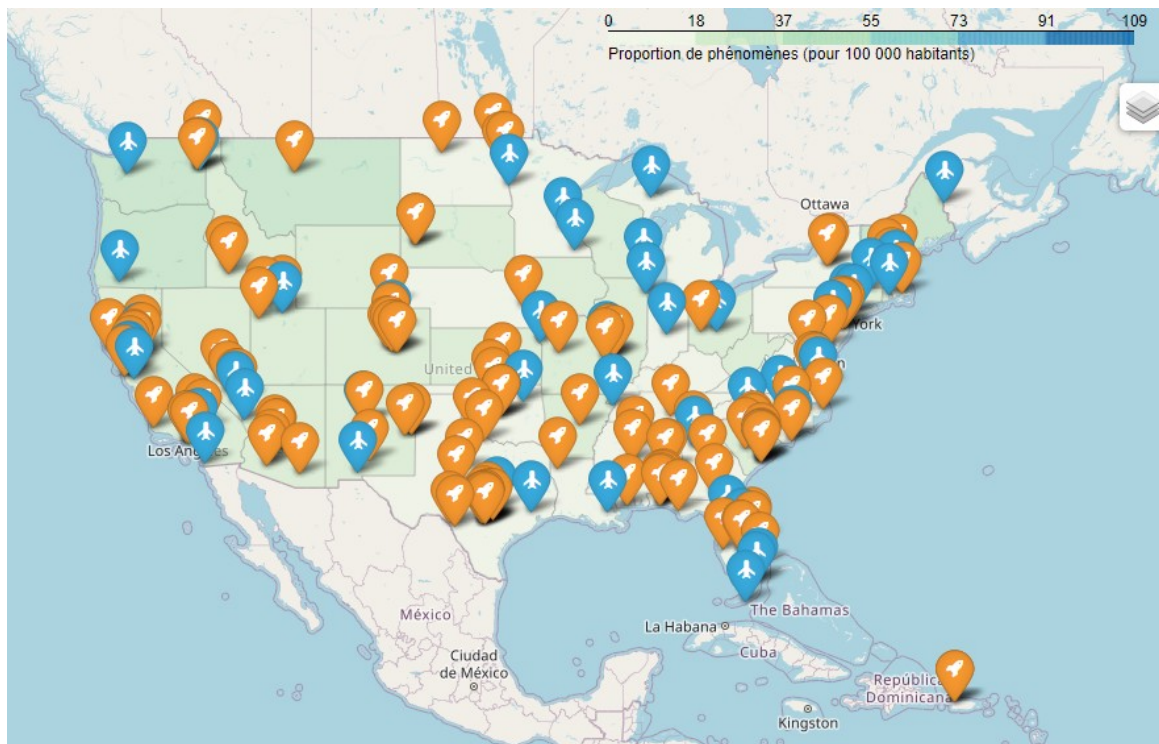
Ces phénomènes peuvent, pour certains, être liés à l'activité humaine.



Carte 3: Les phénomènes (pour 100 000 habitants) sont répandus sur la côte ouest du Royaume-Uni

Les lignes aériennes sont source de phénomènes atmosphériques étonnants. Leurs feux de signalisation ou leurs traînées de condensation aussi appelées contrails (27) génèrent de phénomènes étranges bien que parfaitement explicables par la science

Les bases militaires peuvent être responsables de phénomènes avec, par exemple, le lancement de fusées éclairantes ou de leurres. Prenons par exemple les USA et regroupons les zones d'observations de phénomènes, les aéroports et les sites de l'armée (spécifiquement l'armée de l'air) .



Carte 4: Des phénomènes plus répandus à l'Ouest des USA qu'à l'Est

Les sites comprenant à la fois des sites de l'armée de l'air et des aéroports ne semblent pas correspondre à un nombre de phénomènes plus importants que sur les sites qui en sont dépourvus. Mais il faudrait une étude plus approfondie pour s'en assurer.

5.3. Influence de la météo sur les phénomènes

Le fichier initial est le fichier au format csv nommé " ufo_weather ". Il est téléchargé sur la plateforme Kaggle. Nettoyé, il est renommé " weather_clean ". Il est au format csv. Il comporte 17998 observations et 13 variables. Pour l'analyse temporelle, les variables " year ", "month", "day" et "hour" sont créées. Le fichier recense des observations de 1970 à 2019.

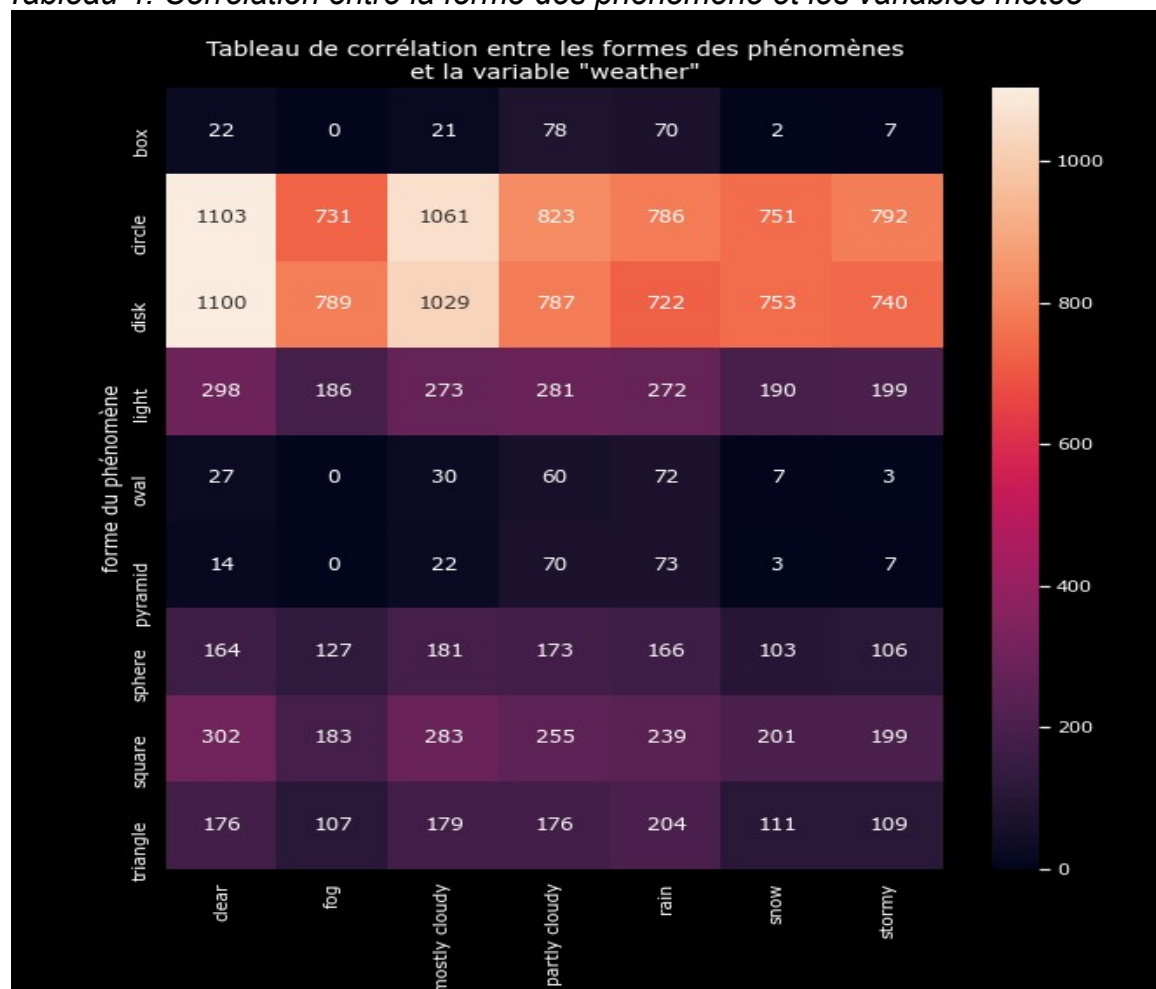
5.3.1. Analyse univariée

La variable "sighting" ne comporte qu'une unique valeur. Elle est supprimée du fichier. La variable "UFO_shape" comporte des items différents du fichier de l'étude principal. Il y a donc huit formes de phénomènes différentes. La variable "weather" comporte 7 items différents. Comme précédemment, compte tenu du nombre élevé d'observations, les variables peuvent être considérées comme suivant une loi normale.

5.3.2. Analyses bivariées

Un test d'ANOVA nous indique que la météo n'a pas d'influence sur le nombre de phénomènes observés. En revanche, un test de chi2 nous indique que la météo influe sur la forme des phénomènes mais pas sur la temporalité (28).

Tableau 4: Corrélation entre la forme des phénomène et les variables météo



Les formes lumineuses et carrées sont mieux perçues par temps clair

Les formes complexes telles qu'en forme de boîte, pyramide et ovales sont rares par temps orageux ou enneigés.

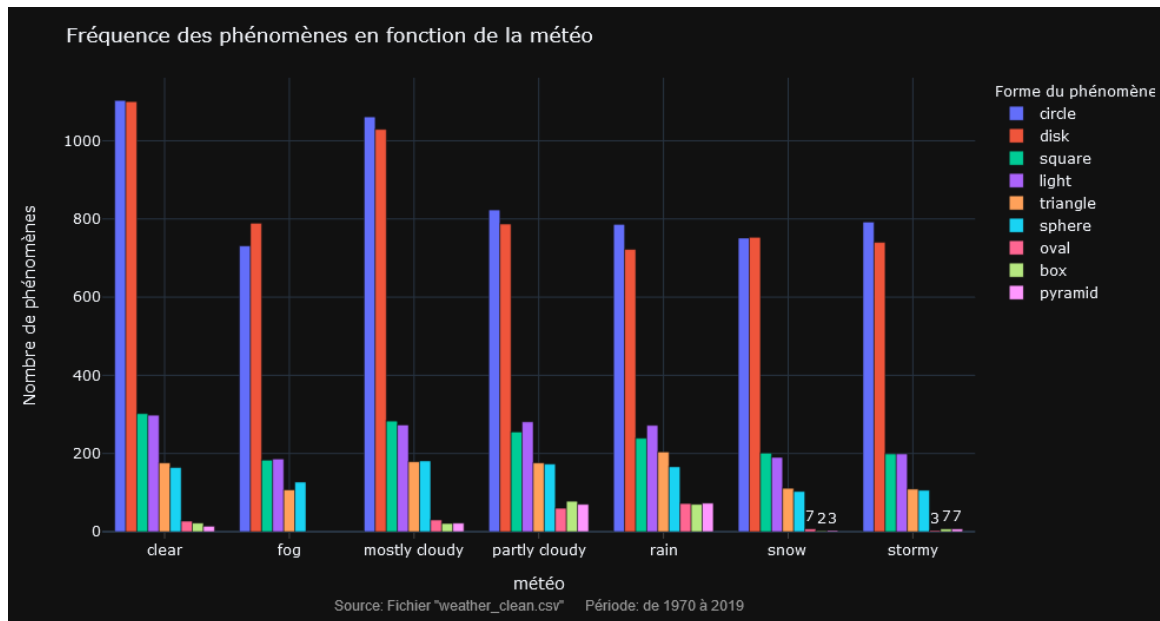


Figure 34: Les phénomènes circulaires sont majoritaires quelque soit la météo

Les temps froids et donc potentiellement clairs favorisent les cristaux de glace qui peuvent induire des phénomènes étranges. On pourrait citer les halos lunaires, les piliers de lumières(29).

Les temps pluvieux ou de brouillard favorisent les structures du type Spectre de Brocken(30).

La météo influe sur les formes des phénomènes.

6. MODÉLISATIONS

6.1. Modèle permettant d'identifier les phénomènes expliqués

Fichier utilisé : "df_ufo_weather.csv".

La variable "researchOutcome" répertorie 3 résultats possibles pour l'explication d'un phénomène : "explained", "probable" et "unexplained".

Pour l'élaboration du modèle, "probable" et "unexplained" feront référence à un phénomène inexpliqué. Nous avons neuf formes différentes de phénomènes et sept météos différentes.

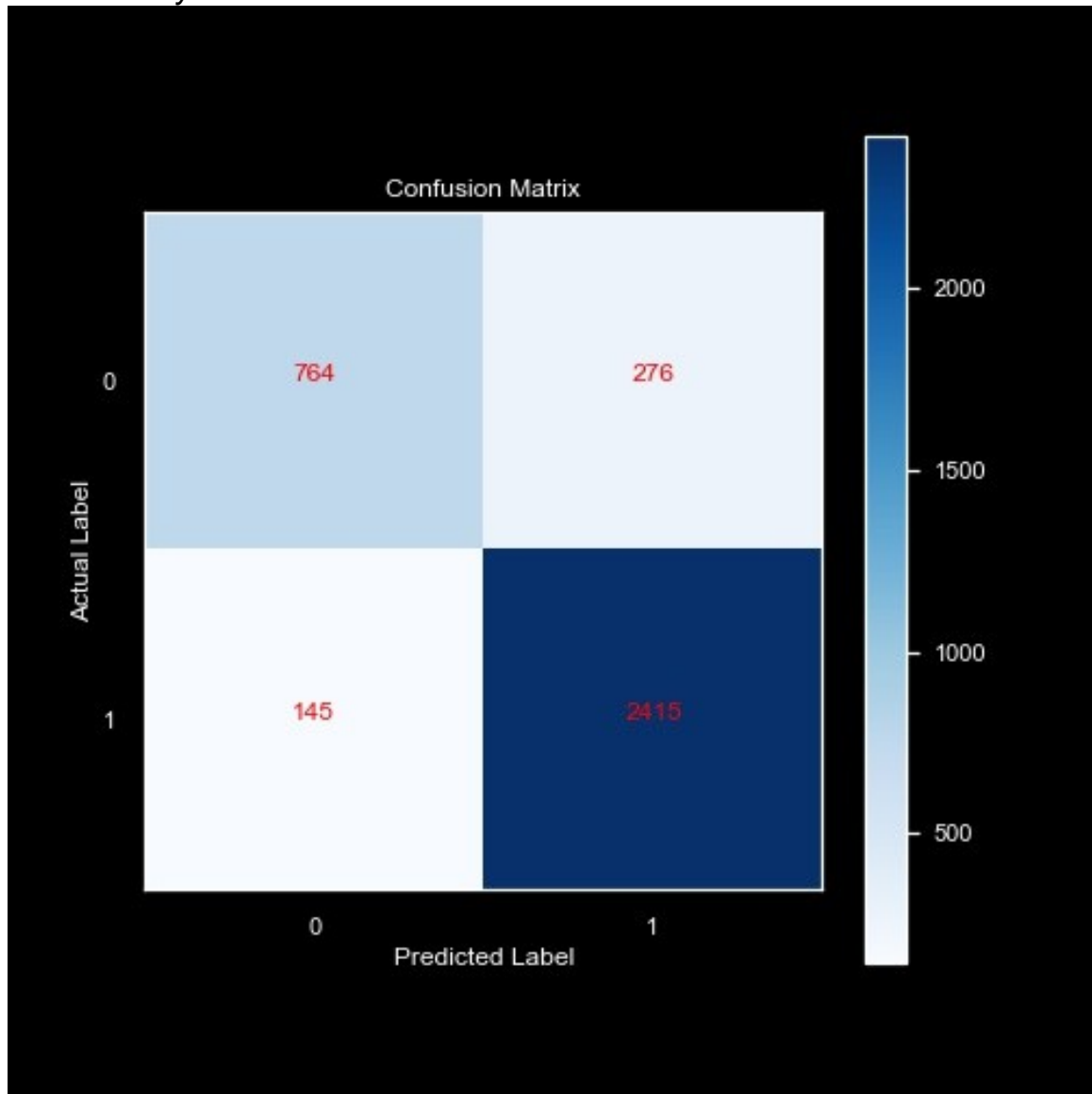
Les données sont standardisées et splittées (données d'entraînement 80 % et 20 % de données test) (31).

Le modèle est un modèle logistique.

La précision du modèle pour repérer un phénomène inexpliqué est de 85% contre 90% pour reconnaître un phénomène expliqué.

Son accuracy est de 88 % sur les données test.

*Tableau 5: Matrice de confusion du modèle logistique de phénomène expliqué.
Son accuracy est de 88 %*



Courbe ROC du modèle logistique

L'aire de la courbe ROC est à 84%.

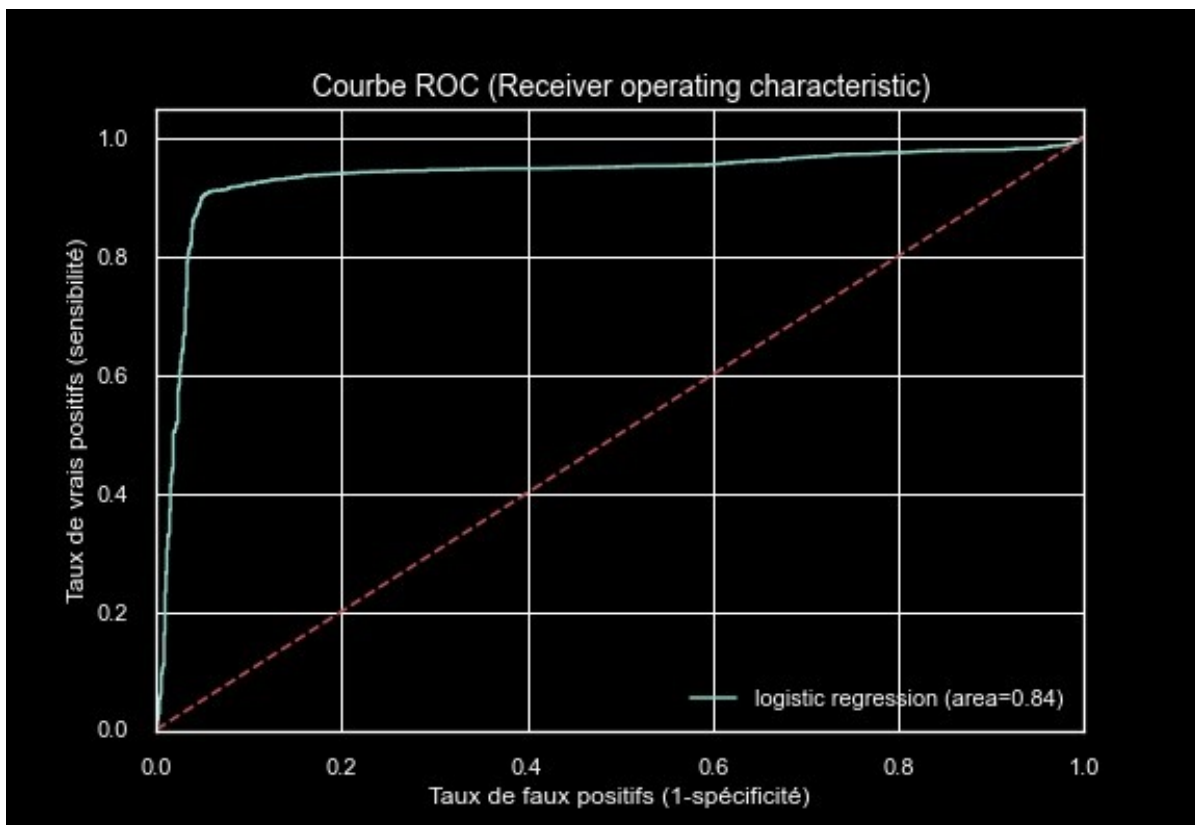


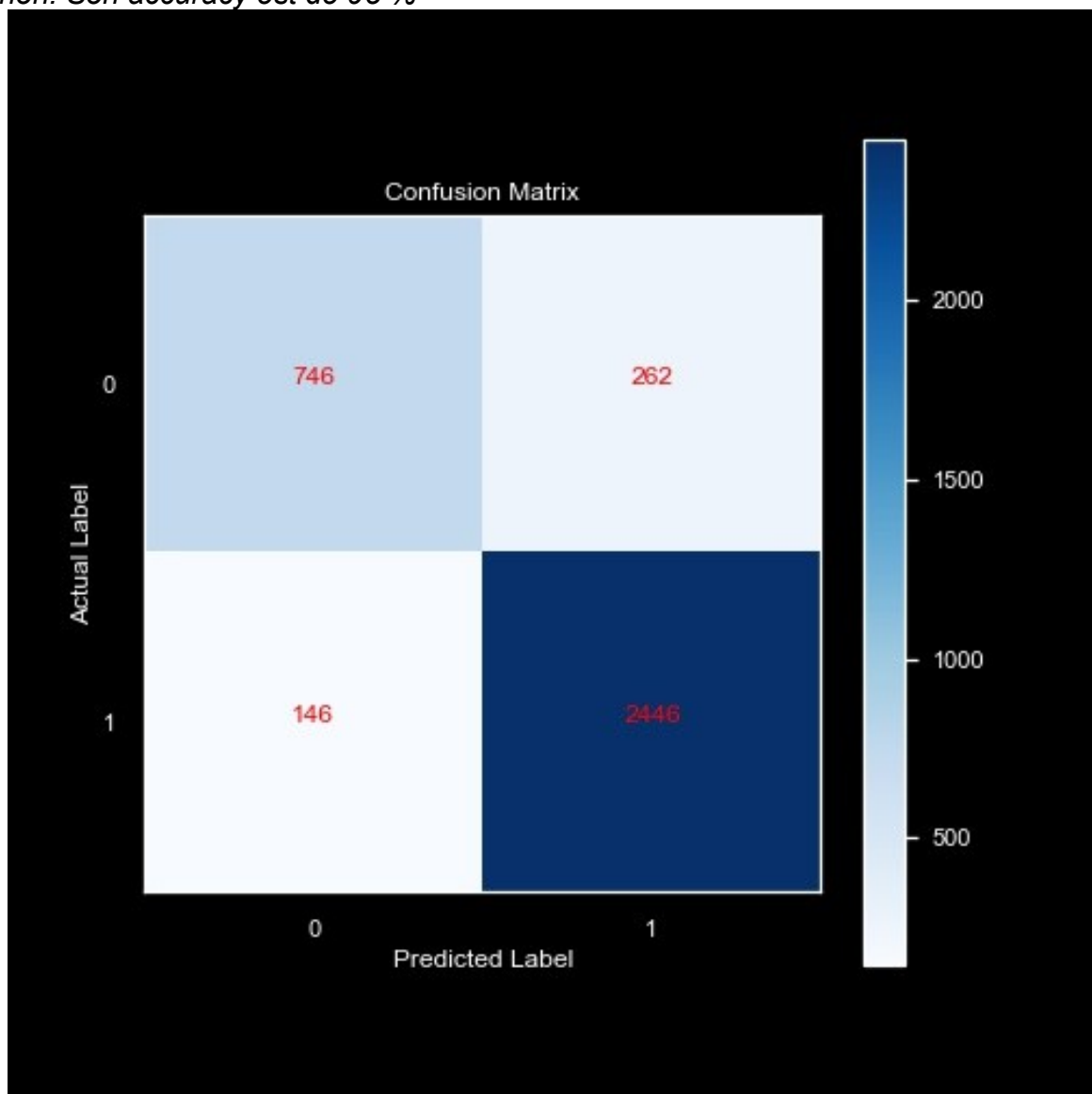
Figure 35: Courbe ROC du modèle logistique phénomène expliqué

Le modèle est un modèle de type RandomForest

Ce modèle se révèle meilleur que le modèle logistique. En effet, sa précision et son accuracy sur les données test sont plus élevées.

Son accuracy est de 95 % sur les données test.

Tableau 6: Matrice de confusion du modèle randomForest phénomène expliqué ou non. Son accuracy est de 95 %



Courbe ROC du modèle RandomForest

L'aire de la courbe ROC est à 95 % contre 84 % pour le modèle logistique. Cela confirme la meilleure qualité du modèle.

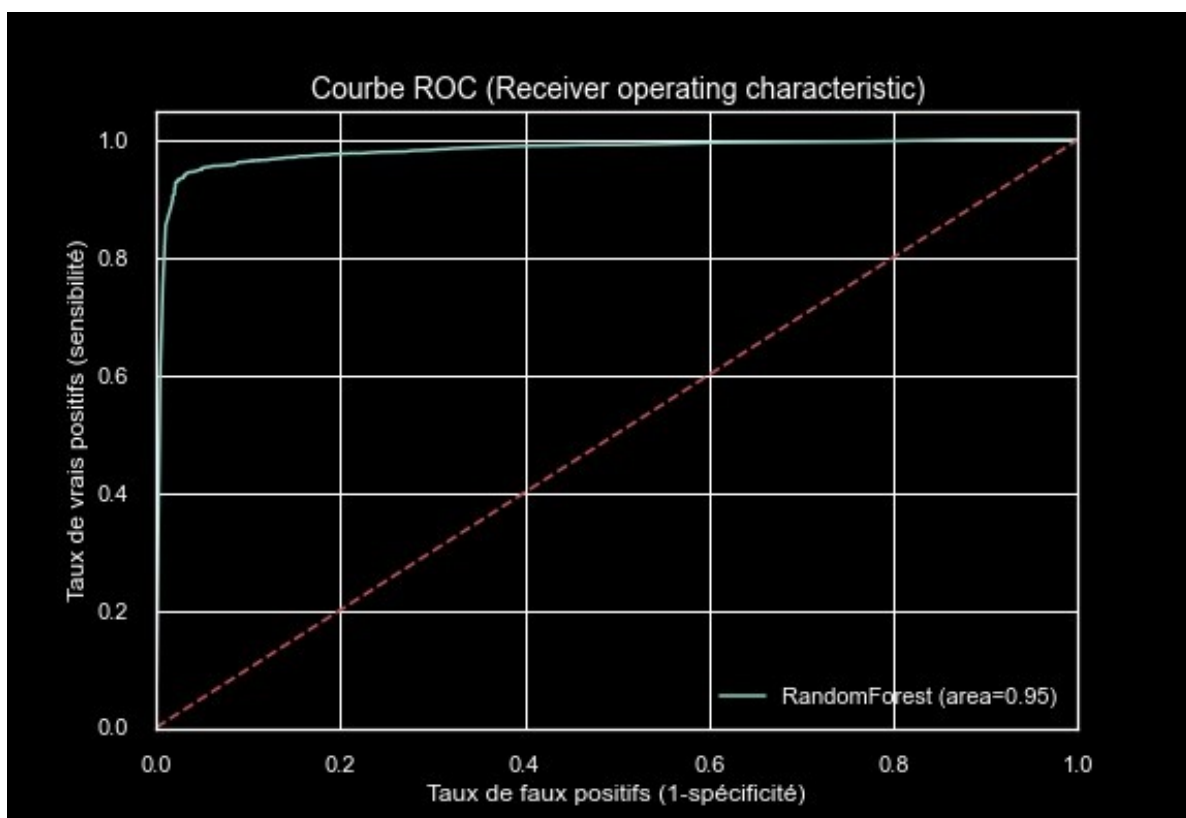


Figure 36: Courbe ROC du modèle RandomForest phénomène expliqué

6.2. Modèle permettant de prédire en fonction de la forme, de la durée du phénomène, dans quel pays le phénomène sera le plus observé

Fichier de modélisation : df_ufo_pop_tot

Le modèle est un modèle de classification des k plus proches voisins (K-NN) (31). Les paramètres du modèle sont le logarithme de la durée du phénomène, le mois d'apparition du phénomène, la catégorie de forme et le modèle d'entraînement. Par exemple, le phénomène dure 4,90 seconde (soit un log de 0,69), il apparaît au mois de septembre sous forme lumineuse. Ce phénomène se déroule, conformément au modèle, aux USA. Ce modèle a une accuracy de 92 %.

7. CONCLUSION

Il apparaît que les caractéristiques des phénomènes dépendent de leur localisation. En effet, les événements de forme ovoïde sont majoritaires au Royaume-Uni tandis que ce sont les phénomènes lumineux qui prédominent aux USA et au Canada. La météo influe sur les formes des phénomènes avec notamment la réflexion des rayons solaires sur les cristaux de glace, par exemple, formant ainsi des halos ou des projections lumineuses atypiques et intrigantes. Certains territoires, à l'instar du Nunavut, du fait de sa proximité avec l'Arctique, en sont un bel exemple avec un nombre de phénomènes important en proportion du nombre d'habitants.

Les phénomènes sont plus identifiés durant l'été qu'en hiver. Peut-être est-ce l'effet des vacanciers flânant la tête dans les étoiles en quête d'une pluie d'étoiles filantes qui auront mal interprétés des sky-laser, ou des feux d'artifices estivaux.

Ces événements semblent préférer les jours de week-end en particulier le samedi et surtout ce sont des événements plutôt nocturnes. Ici aussi, nous pouvons nous demander si ce ne sont pas des événements festifs mal interprétés qui sont à l'origine du signalement. Par exemple, les lanternes volantes sont fréquemment interprétées comme un débarquement alien (32).

La technologie humaine intervient également. Si dans les années 50-60, les phénomènes les plus reconnus sont en forme de soucoupes volantes (ovoïde), à partir des années 2000 ce sont les phénomènes lumineux qui prédominent. C'est à partir de ces années qu'émergent de nouvelles technologies telles que les satellites, les lasers ...qui peuvent induire des phénomènes lumineux étonnants. Le déploiement de satellites du projet Starlink a d'ailleurs suscité beaucoup d'interrogations (33) et le devenir de ces satellites en tant que débris encore plus (28). Il est donc à prévoir encore plus de nombreux phénomènes lumineux "énigmatiques" à l'avenir. Une très belle animation nous en donne un avant-goût (29).

Il reste néanmoins une part de phénomènes inexpliqués (3,3 % selon le GEIPAN) qui laisse le champ libre à toute interprétation. Libre à chacun de débattre autour du paradoxe de Fermi ou de méditer (rêver ?) sur la déclaration de John Ratcliffe, ancien directeur du renseignement national des États-Unis, " *Des mouvements difficiles à reproduire, pour lesquels nous n'avons pas la technologie ou qui se déplacent à des vitesses dépassant le mur du son sans bang sonique.* " (36)

Table des matières

1. Introduction.....	1
2. Nettoyage des données.....	3
2.1. Description du fichier.....	3
2.1.1. Types de variables.....	3
2.1.2. Traitement des valeurs manquantes.....	4
2.2. Importation de fichiers complémentaires pour affiner l'étude.....	6
2.2.1. Population.....	6
2.2.2. Météo.....	6
2.3. Export des fichiers.....	6
3. Analyse des données : monde entier.....	9
3.1. Étude générale des phénomènes au niveau mondial.....	9
3.1.1. Étude des formes de phénomènes.....	10
3.1.2. Analyse temporelles.....	13
3.1.3. Analyse géographique.....	15
4. Analyse des données : USA, Canada et Royaume-Uni.....	17
4.1. Étude et traitement des outliers.....	17
4.2. Étude et traitement sans outliers.....	18
4.2.1. Analyse univariée.....	19
4.2.2. Analyses bivariées.....	19
4.2.3. Étude générale des formes des phénomènes.....	22
5. Étude des phénomènes en fonction du pays.....	25
5.1. Étude des phénomènes.....	25
5.1.1. Analyse temporelle.....	26
5.2. Influence de la population sur les phénomènes.....	31

5.2.1. Analyse univariée.....	31
5.2.2. Analyse bivariée.....	32
5.2.3. Étude des phénomènes en incluant l’impact de la population.....	33
5.3. Influence de la météo sur les phénomènes.....	38
5.3.1. Analyse univariée.....	39
5.3.2. Analyses bivariées.....	39
6. Modélisations.....	41
6.1. Modèle permettant d’ identifier les phénomènes expliqués.....	41
6.2. Modèle permettant de prédire en fonction de la forme, de la durée du phénomène, dans quel pays le phénomène sera le plus observé.....	45
7. Conclusion.....	47

Index des figures

Figure 1: Valeurs manquantes par variable.....	4
Figure 2: Des valeurs manquantes sur des variables cibles.....	5
Figure 3: Les USA, le Canada et le Royaume-Uni sont les principaux pays représentés dans le fichier.....	9
Figure 4: Les phénomènes lumineux sont majoritaires dans le monde.....	10
Figure 5: Les phénomènes coniques sont les plus persistants dans le monde avec une durée moyenne de plus d'une journée.....	11
Figure 6: Les formes lumineuses et ovoïdes sont les plus communes dans le monde.....	12
Figure 7: Les formes non identifiées sont les plus persistantes dans le monde.....	12
Figure 8: Les formes lumineuses et ovoïdes sont les plus répandues.....	13
Figure 9: Les phénomènes sont en constante évolution à partir de 1995 pour atteindre un maximum en 2012.....	13
Figure 10: Evolution constante des observations mondiales.....	13
Figure 11: Les phénomènes dans le monde sont plus nombreux de juin à octobre avec un maximum pour le mois de juillet.....	14
Figure 12: Les phénomènes mondiaux apparaissent davantage à partir de 19h00 avec un maximum à 21h00.....	15
Figure 13: La catégorie de forme ovoïde est la catégorie la plus représentée pour les durées supérieures à la journée.....	17
Figure 14: Pour les durées supérieures à la journée, les formes lumineuses sont celles qui sont les plus persistantes.....	18

Figure 15: Des variables considérées comme gaussiennes au vu du nombre d'observations.....	19
Figure 16: La distribution des durées (sans les outliers) des phénomènes diffèrent selon leur forme	20
Figure 17: Les formes lumineuses et ovoïdes sont majoritaires pour les 3 pays.....	22
Figure 18: Seules les formes changeantes sont plus longues.....	22
Figure 19: Une majorité de phénomènes de moins d'une minute.....	23
Figure 20: Les formes ovoïdes et lumineuses sont les mieux représentées aux USA, au Canada et au Royaume-Uni.....	25
Figure 21: Les durées des phénomènes <i>dépendent</i> de leur forme et du pays.....	26
Figure 22: Evolution du nombre de phénomènes plus importante aux USA.....	26
Figure 23: La proportion de phénomènes est la plus importante au Canada durant l'été. L'hiver est plus favorable aux phénomènes anglo-saxons.....	27
Figure 24: Le samedi et le dimanche <i>favorables aux phénomènes</i>	28
Figure 25: Répartition journalière des phénomènes quelque soit le pays considéré.....	28
Figure 26: Accroissement du nombre de forme des phénomènes en particulier pour les formes ovoïdes et lumineuses.....	29
Figure 27: On observe la même distribution des catégories de forme des phénomènes mais leur nombre diffèrent selon le mois.....	30
Figure 28: Le nombre de phénomènes s'accroît en soirée.....	30
Figure 29: La plupart des formes des phénomènes n'évoluent pas selon le jour. Seules les formes ovoïdes et lumineuses semblent plus nombreuses le samedi.....	31
Figure 30: Le nombre de phénomènes (pour 1000 000 habitants) varie en fonction de l'État considéré avec un maximum pour le territoire de Nunavut.....	33

Figure 31: De nombreux phénomènes sur le territoire Nunavut.....	34
Figure 32: Les formes lumineuses prédominent sur le territoire Nunavut.....	35
Figure 33: Les formes ovoïdes de plus de 8 minutes sont les plus persistantes.....	35
Figure 34: Les phénomènes circulaires sont majoritaires quelque soit la météo.....	40
Figure 35: Courbe ROC du modèle logistique phénomène expliqué.....	43
Figure 36: Courbe ROC du modèle RandomForest phénomène expliqué.....	45

Index des tableaux

Tableau 1: Corrélacion entre les catégories de forme des phénomène et les pays.....	21
Tableau 2: Corrélacion entre les catégories de forme des phénomènes et les mois.....	21
Tableau 3: Corrélacion entre la population et la latitude.....	32
Tableau 4: Corrélacion entre la forme des phénomène et les variables météo.....	39
Tableau 5: Matrice de confusion du modèle logistique de phénomène expliqué. Son accuracy est de 88 %.....	42
Tableau 6: Matrice de confusion du modèle randomForest phénomène expliqué ou non. Son accuracy est de 95 %.....	44

Index des cartes

<i>Carte 1</i> : Les phénomènes mondiaux sont répartis principalement dans l'hémisphère nord. Les zones jaunes orangées sont les plus denses.....	16
<i>Carte 2</i> : Exemple de répartition des phénomènes. Ici, au Royaume-Uni.....	36
<i>Carte 3</i> : Les phénomènes (pour 100 000 habitants) sont répandus sur la côte ouest du Royaume-Uni.....	37
<i>Carte 4</i> : Des phénomènes plus répandus à l'Ouest des USA qu'à l'Est.....	38

Références

1. Recherche de cas | GEIPAN [Internet]. GEIPAN. [cité 3 déc 2021]. Disponible sur: https://www.cnes-geipan.fr/fr/recherche/cas?field_agregation_index_value=forme+ovoide
2. Benoît Lagane. « Les Envahisseurs » : David Vincent les a vus... Mais ça dit quoi ? [Internet]. france inter. 2019 [cité 6 déc 2021]. Disponible sur: <https://www.franceinter.fr/emissions/l-effet-miroir/l-effet-miroir-02-novembre-2019>
3. Cam Nugent. UFO Sightings around the world [Internet]. Kaggle. 2017 [cité 1 sept 2021]. Disponible sur: <https://kaggle.com/camnugent/ufo-sightings-around-the-world>
4. Isabelle Barbier. P8_01_nettoyage_donnees - Jupyter Notebook [Internet]. Jupyter Notebook. 2021 [cité 9 déc 2021]. Disponible sur: http://localhost:8888/notebooks/Documents/P8_communiquez_vos_resultats/P8_01_nettoyage_donnees.ipynb
5. Latitude.to. GPS coordinates of Mescalero, United States. Latitude: 33.1737 Longitude: -105.6073 [Internet]. Latitude.to, maps, geolocated articles, latitude longitude coordinate conversion. 2021 [cité 11 sept 2021]. Disponible sur: <https://latitude.to/articles-by-country/us/united-states/13108/mescalero>
6. Postcodebyaddress. Complete list of all post towns in the UK - Postcode by address [Internet]. Postcodebyaddress. 2021 [cité 1 déc 2021]. Disponible sur: <https://postcodebyaddress.co.uk/post-towns>
7. World Population Review. Canadian Provinces Population 2021 [Internet]. World Population Review. 2021 [cité 1 déc 2021]. Disponible sur: <https://worldpopulationreview.com/canadian-provinces>
8. Neil Park. Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland - Office for National Statistics [Internet]. Office for National Statistics. 2015 [cité 1 déc 2021]. Disponible sur: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland>
9. Population of Canada by province and territory. In: Wikipedia [Internet]. 2021 [cité 1 déc 2021]. Disponible sur: https://en.wikipedia.org/w/index.php?title=Population_of_Canada_by_province_and_territory&oldid=1055573914

10. Population of England and Wales [Internet]. Population of England and Wales - Gov.uk Ethnicity Facts and Figures. 2021 [cité 1 déc 2021]. Disponible sur: <https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/population-of-england-and-wales/latest#download-the-data>
11. Bureau UC. State Population Totals and Components of Change: 2010-2019 [Internet]. Census.gov. [cité 1 déc 2021]. Disponible sur: <https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html>
12. Dwarika Teli. UFO datasets, the truth is out there... [Internet]. Kaggle. 2020 [cité 1 déc 2021]. Disponible sur: <https://kaggle.com/dwarika/ufo-datasets-the-truth-is-out-there>
13. Percer le secret de certains phénomènes associés aux orages - Lancement en novembre du satellite français Taranis | CNRS [Internet]. cnrs. [cité 5 déc 2021]. Disponible sur: <https://www.cnrs.fr/fr/percer-le-secret-de-certains-phenomenes-associes-aux-orages-lancement-en-novembre-du-satellite>
14. Touja J. Des photos de sprites rares et spectaculaires / Agences-Spatiales [Internet]. Agences-Spatiales. 2019 [cité 5 déc 2021]. Disponible sur: <https://www.agences-spatiales.fr/photos-de-sprites-rares/>
15. Isabelle Barbier. P8_01_EDA_monde_entier - Jupyter Notebook [Internet]. Jupyter Notebook. 2021 [cité 9 déc 2021]. Disponible sur: http://localhost:8888/notebooks/Documents/P8_communiquez_vos_resultats/P8_01_EDA_monde_entier.ipynb
16. Project Sign | American UFO panel | Britannica [Internet]. [cité 3 déc 2021]. Disponible sur: <https://www.britannica.com/topic/Project-Sign>
17. Project BLUE BOOK - Unidentified Flying Objects [Internet]. National Archives. 2016 [cité 3 déc 2021]. Disponible sur: <https://www.archives.gov/research/military/air-force/ufo>
18. Drupal 7. Transit de Vénus [Internet]. espace-sciences.org. 2012 [cité 3 déc 2021]. Disponible sur: <https://www.espace-sciences.org/planetarium/article/transit-de-venus>
19. ALI PRA et KAK, À 12:00 M 28 J 2021. Astronomie : le calendrier des événements célestes à ne pas manquer cette année 2021 [Internet]. CNET France. [cité 3 déc 2021]. Disponible sur: <https://www.cnetfrance.fr/news/astronomie-calendrier-evenements-annee-39926485.htm>

20. Isabelle Barbier. P8_01_EDA_usa_uk_canada - Jupyter Notebook [Internet]. Jupyter Notebook. 2021 [cité 7 déc 2021]. Disponible sur: http://localhost:8888/notebooks/Documents/P8_communiquez_vos_resultats/P8_01_EDA_usa_uk_canada.ipynb
21. Iridium (téléphonie par satellite). In: Wikipédia [Internet]. 2021 [cité 4 déc 2021]. Disponible sur: [https://fr.wikipedia.org/w/index.php?title=Iridium_\(t%C3%A9l%C3%A9phonie_par_satellite\)&oldid=186782397](https://fr.wikipedia.org/w/index.php?title=Iridium_(t%C3%A9l%C3%A9phonie_par_satellite)&oldid=186782397)
22. Méprises-du-ciel.fr [Internet]. Méprises-du-ciel.fr. [cité 4 déc 2021]. Disponible sur: <https://meprises-du-ciel.fr/>
23. Isabelle Barbier. P8_01_EDA_population - Jupyter Notebook [Internet]. Jupyter Notebook. 2021 [cité 9 déc 2021]. Disponible sur: http://localhost:8888/notebooks/Documents/P8_communiquez_vos_resultats/P8_01_EDA_population.ipynb
24. Peter Kikkert. Nunavut | l'Encyclopédie Canadienne [Internet]. L'Encyclopédie Canadienne. 2007 [cité 4 déc 2021]. Disponible sur: <https://www.thecanadianencyclopedia.ca/fr/article/nunavut>
25. Parhélie. In: Wikipédia [Internet]. 2021 [cité 5 déc 2021]. Disponible sur: <https://fr.wikipedia.org/w/index.php?title=Parh%C3%A9lie&oldid=186891434>
26. Altocumulus lenticularis. In: Wikipédia [Internet]. 2021 [cité 4 déc 2021]. Disponible sur: https://fr.wikipedia.org/w/index.php?title=Altocumulus_lenticularis&oldid=182495479
27. epa. Aircraft Contrails Factsheet [Internet]. 2000 [cité 4 déc 2021]. Disponible sur: https://www.faa.gov/regulations_policies/policy_guidance/envir_policy/media/contrails.pdf
28. Isabelle Barbier. P8_01_EDA_meteo - Jupyter Notebook [Internet]. Jupyter Notebook. 2021 [cité 9 déc 2021]. Disponible sur: http://localhost:8888/notebooks/Documents/P8_communiquez_vos_resultats/P8_01_EDA_meteo.ipynb
29. Elodie C. Des piliers de lumière surréalistes capturés au Canada [Internet]. GOLEM13.FR. 2018 [cité 5 déc 2021]. Disponible sur: <http://golem13.fr/piliers-lumiere-surrealistes-canada/>

30. Elie Belorizky. L'étonnant spectre de Brocken [Internet]. Encyclopédie de l'environnement. [cité 5 déc 2021]. Disponible sur: <https://www.encyclopedie-environnement.org/zoom/spectre-de-brocken/>
31. Isabelle Barbier. P8_01_modelisations - Jupyter Notebook [Internet]. Jupyter Notebook. 2021 [cité 9 déc 2021]. Disponible sur: http://localhost:8888/notebooks/Documents/P8_communiquez_vos_resultats/P8_01_modelisations.ipynb#Mod%C3%A9lisations
32. Fabrice Valery. Les OVNI dans le ciel de Villemur-sur-Tarn étaient en fait... des lanternes chinoises ! [Internet]. France 3 Occitanie. [cité 6 déc 2021]. Disponible sur: <https://france3-regions.francetvinfo.fr/occitanie/haute-garonne/les-ovni-dans-le-ciel-de-villemur-sur-tarn-etaient-en-fait-des-lanternes-chinoises-259325.html>
33. Catherine Lioult. Des points lumineux qui se déplacent dans le ciel de la Côte d'Azur ? C'est Starlink et on vous dit tout [Internet]. France 3 Provence-Alpes-Côte d'Azur. 2021 [cité 6 déc 2021]. Disponible sur: <https://france3-regions.francetvinfo.fr/provence-alpes-cote-d-azur/alpes-maritimes/des-points-lumineux-qui-se-deplacent-dans-le-ciel-de-la-cote-d-azur-c-est-starlink-et-on-vous-dit-tout-1951663.html>
34. VIDEO. La pollution des satellites Starlink inquiète les astronomes [Internet]. Franceinfo. 2021 [cité 6 déc 2021]. Disponible sur: https://www.francetvinfo.fr/monde/usa/video-la-pollution-des-satellites-starlink-inquiete-les-astronomes_4788479.html
35. Stuff in Space [Internet]. stuffin.space. [cité 6 déc 2021]. Disponible sur: <http://stuffin.space/>
36. Voici ce que nous savons du rapport américain sur les visiteurs extraterrestres [Internet]. BBC News Afrique. 2021 [cité 6 déc 2021]. Disponible sur: <https://www.bbc.com/afrique/monde-57579091>