

A dramatic, high-contrast image of a stormy sky with dark, swirling clouds and a bright light source breaking through at the top, creating a silhouette effect.

CEDRIC CAUDRON

ETUDE DES TORNADES AUX USA ENTRE 1950 ET 2022

Pour l'évaluation du cours NFE212 par Mr Simon Vershilde

ANNEE 2023

The logo for le cnam Hauts-de-France, featuring the text 'le cnam' in a bold, white, sans-serif font above 'Hauts-de-France' in a smaller, white, sans-serif font, all contained within a red square.

le cnam
Hauts-de-France

Table des matières

- INTRODUCTION 1
- 1 LES ORAGES ET TORNADES 2
 - 1.1 LES ORAGES 2
 - 1.2 LES TORNADES..... 3
- 2 : LES DONNEES..... 4
 - 2.1 LES DONNEES SUR LES TORNADES 4
 - 2.1.1 Choix du jeu de données 4
 - 2.1.2 Préparation des données 5
 - 2.2 LES DONNEES SUR LES ETATS..... 7
 - 2.2.1 Choix du jeu de données 7
 - 2.2.2 Préparation des données 7
 - 2.3 LES DONNEES SUR LA DENSITE DE POPULATION..... 8
 - 2.3.1 Choix du jeu de données 8
 - 2.3.2 Préparation des données 9
 - 2.4 CREATION DU MODELE DE DONNEES 10
- 3 : LES VISUELS 11
 - 3.1 WIREFRAME 11
 - 3.2 PRESENTATION GENERALES DES TORNADES AUX USA 12
 - 3.2.1 Anomalie ou réalité pour les EF0 ? 13
 - 3.2.2 Pourquoi moins de tornades à l’ouest ? 14
 - 3.2.3 Situation des EF5 16
 - 3.3 ANALYSE DES CARACTERISTIQUES TAILLE ET DISTANCE PARCOURUE 17
 - 3.4 ANALYSE DES MOMENTS D’APPARITION (ANNEES, MOIS, HEURES)..... 18
 - 3.5 ANALYSE DES PERTES MATERIELLES ET AGRICOLES 19
 - 3.6 ANALYSE DES DOMMAGES HUMAIN DES TORNADES..... 21
 - 3.7 ANALYSE DES DOMMAGES HUMAIN DES TORNADES..... 23
 - 3.8 BONUS 25
 - 3.9 LA DERNIERE 26
- 4 : CONCLUSION 27

Introduction

Les tornades...

Un phénomène naturel tellement impressionnant... Et dévastateur...

Tellement impressionnant et dévastateur qu'il a fait justement l'objet de nombreuses études ! Et la formation de ce phénomène ne comporte quasiment plus aucun secret.

Les USA sont le pays comportant le plus grand nombre de tornades répertoriées au monde. C'est sûrement pour cela qu'il y a eu de nombreuses études sur le phénomène sachant l'effet dévastateur qu'il peut avoir (et notamment le fait que les USA sont la première puissance économique mondiale avec l'impact qu'elles peuvent engendrer).

Avec toutes ces études les tornades sont forcément plus prévisibles au moins sur le territoire américain ? Et bien non !

Les tornades, même si l'on comprend leur mécanisme restent un phénomène extrêmement imprévisible (tout comme les orages dont elles sont issues), et malgré les progrès elles font toujours autant (je suppose) de dégâts, de victimes (blessées ou décès...)

Ce document, ainsi que la visualisation Power Bi associée **n'a pas pour but** d'essayer d'améliorer la détection ni même de comprendre mieux la formation de ce phénomène.

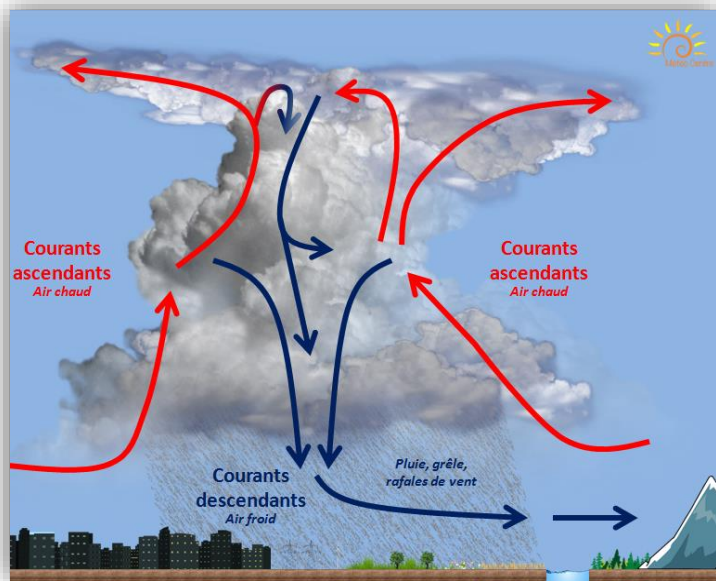
Néanmoins, je vais essayer de vous proposer une vision d'ensemble, décennies par décennies (avec quelques incursions dans une granularité plus fine), des tornades ayant eu lieu sur le sol américain, ainsi que de montrer quelques conséquences que cela engendre, et ce entre 1950 et 2019.

1 LES ORAGES ET TORNADES

1.1 Les orages

Avant toute chose, il faut comprendre, même de manière très générale le phénomène des orages.

Les orages sont avant tout un conflit de masse d'air. Un air chaud qui, la plupart du temps, est installé depuis un moment. Et une masse d'air plus froide qui arrive.



Je passe sur les détails, sur la formation de la foudre, sur pourquoi le cumulonimbus a une forme d'enclume (spoiler c'est à cause de la troposphère), je passe sur la formation des averses (du coup également pourquoi il pleut sans orages ou pourquoi il y a de la grêle) et bien d'autres choses.

Car mon R.O.I. pour ce sujet ne le permet pas malheureusement... 😊

Voir <https://www.meteo-centre.fr/blog/les-orages/> pour plus d'infos.

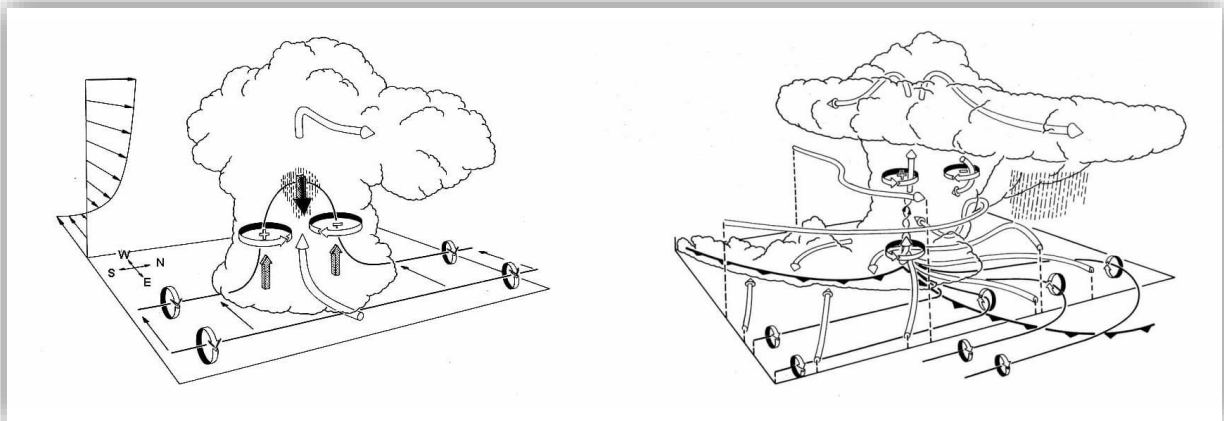
1.2 Les tornades

Nous avons notre beau cumulonimbus. Que lui faut-il pour donner naissance à une tornade ?

Je vais encore une fois rester simple car ce n'est pas le but du projet. Mais il lui faut :

- Grosse différence entre la masse d'air chaude et la masse d'air froide (ce n'est pas obligatoire, cela va de façon simplissime déterminer (ou non) la violence de l'orage)
- D'un cisaillement des vents dans l'atmosphère (ça c'est un paramètre quasi obligatoire)

Je reste très simple mais au final ce sont les deux paramètres principaux pour donner, du moins, une tornade significative. Je tiens à préciser que le terme « mini tornade » utilisé à tort et à travers pendant des années par les médias n'a aucun sens météorologique... Il se passe tellement de phénomènes venteux ou autres sous les orages mais cela ne peut en aucun cas les qualifier de tornade. Et une tornade a une définition bien précise.



Pour en savoir plus : <https://www.keraunos.org/recherche/comprendre-les-orages-pedagogie-vulgarisation/tornades-trombes-tubas/comment-se-forme-une-tornade.html>

2 : LES DONNEES

2.1 Les données sur les tornades

2.1.1 Choix du jeu de données

J'ai récupéré les données officielles de la NOAA (National Oceanic and Atmospheric Administration), qui est l'agence américaine de l'étude de l'océan et de l'atmosphère.

Il y a deux jeux de données disponibles. Un très précis avec des informations sur les comtés et états traversés. C'est le premier que j'ai tenté d'utiliser mais, au vu du but de ce document, il n'était pas nécessaire d'avoir un tel niveau de granularité (et d'ailleurs il était beaucoup plus complexe à utiliser).

J'ai donc opté pour l'utilisation du deuxième jeu de données, qui ne contiennent pas ces informations précises. Il est beaucoup plus facile à utiliser car dans celui-ci chaque ligne

U.S. TORNADOES* (1950-2022) *Read format description document!	
Change Log	
Last Update: 25 April 2023	
1950-2022_all_tornadoes.csv (7.4 mb)	Raw database dump includes all state and continuing county segments.
1950-2022_actual_tornadoes.csv (7.3 mb)	Single tracks. No state segments or continuing county info (e.g. sg="1").

correspond à une et une seule tornade.

On peut retrouver ce jeu de donnée sur le site officiel de la NOAA à cette adresse :

https://www.spc.noaa.gov/wcm/data/1950-2022_actual_tornadoes.csv

2.1.2 Préparation des données

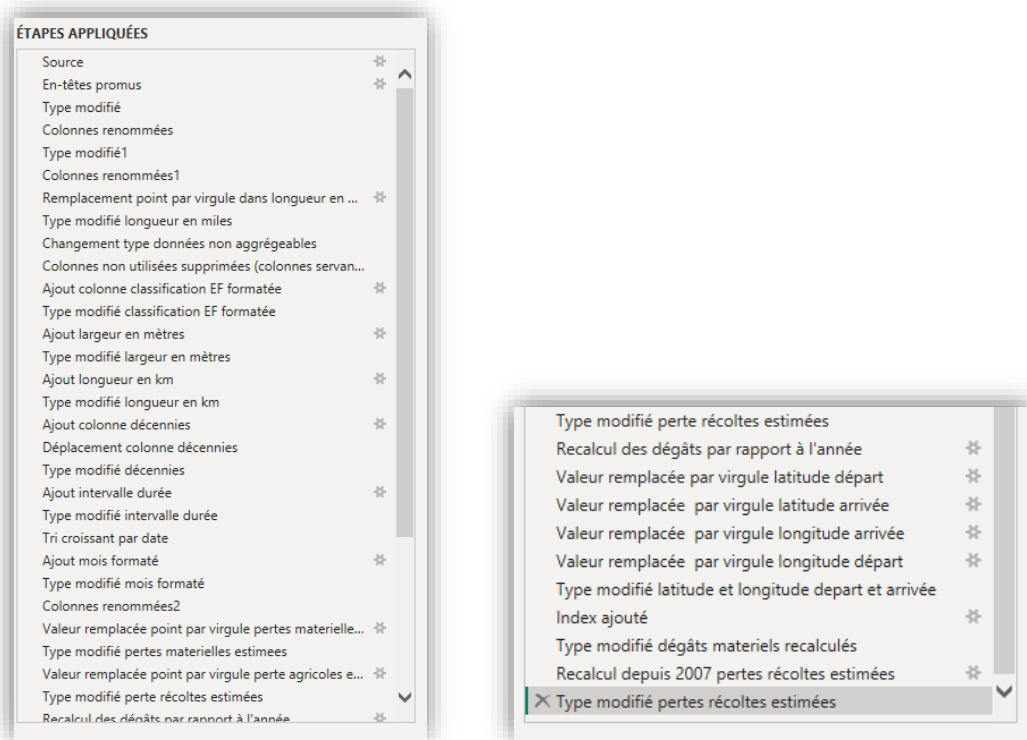
Je ne vais pas entrer dans le détail de chaque transformation que j'ai effectué, mais il y a eu un gros travail dessus, autant pour une meilleure lisibilité des rapports, que pour une harmonisation des données comme je vais l'expliquer ci-dessous.

La première chose à faire a été d'étudier la documentation du jeu de données. Afin de comprendre à quoi chaque colonne se référait. De plus, je me suis rendu compte que la documentation n'était pas totalement à jour ce qui m'aura valu quelques recherches afin de comprendre le fonctionnement de quelques colonnes (car notamment pour les dégâts matériels, au fil du temps, l'information a changé de format, ce qui m'aura valu de faire des transformations conditionnelles, et avant cela une analyse poussée des données car je n'avais pas toujours la date à laquelle ces changements sont intervenus).

La documentation (fournie) peut se trouver également sur le site officiel de la NOAA à cette adresse :

https://www.spc.noaa.gov/wcm/data/SPC_severe_database_description.pdf

Ci-dessous, la liste des étapes de transformation, que j'ai renommées pour une meilleure maintenabilité de l'ensemble (ça m'a d'ailleurs sauvé la vie...).



Ainsi, j'ai fait différentes transformations, comme la conversion des miles en kilomètres, des yards en mètres, ainsi que des calculs conditionnels comme pour les pertes matérielles estimées, qui sera le seul que je détaillerai ci-dessous.

```
#"Type modifié perte récoltes estimées",  
"dégâts matériels estimés recalculés",  
each if Number.FromText([année]) < 1996 then null  
else if Number.FromText([année]) < 2016 then [pertes matérielles estimées]*1000000  
else [pertes matérielles estimées])
```

J'ai donc créé une nouvelle colonne, dans laquelle tous les dégâts avant 1996 je ne les prend pas en compte. En effet je me suis limité à une étude à partir de 1996 pour cette donnée précise. Car avant cette année, la donnée correspond à une fourchette de dégât estimé. J'aurais pu l'utiliser, mais par simplicité j'ai préféré la mettre de côté.

A partir de 1996, la donnée représente le nombre de millions de dollars de dégât estimé. Ainsi, pour harmoniser l'ensemble, je multiplie par 1.000.000 pour avoir la donnée globale.

Et à partir de 2016, c'est le montant en lui-même qui est indiqué directement dans la colonne. Ainsi il ne faut pas faire de transformation.

Et c'est pour trouver cette date de 2016 que j'ai dû faire une analyse poussée des données car elle n'est pas présente dans la documentation. Il y est juste indiqué que « plus tard ce serait le montant total dans la colonne » mais sans aucune date. J'ai donc dû la déterminer moi-même.

A savoir que j'ai dû effectuer à peu près le même raisonnement pour les pertes agricoles estimées.

De plus, un autre point important que je souhaite préciser. Chaque ligne du fichier correspond bien à une tornade différente. Mais il y a des doublons dans le numéro de tornade tout de même. Ceci s'explique également par un changement de méthode au fil des années pour l'attribution du numéro. Ainsi, j'ai créé un index au cas où, et sinon je compte bien le nombre de ligne sur le numéro de tornade et non un compte distinct.

2.2 Les données sur les états

2.2.1 Choix du jeu de données

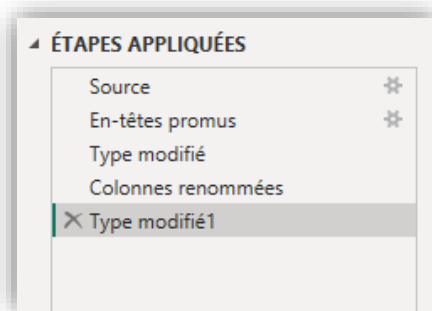
Voulant effectuer des agrégations par état, j'avais besoin d'une table qui répertorierait ceux-ci ainsi que leurs abréviations, leurs noms complets et également le code FIPS.

J'ai donc trouvé, sur le site de la NOAA, un jeu de données correspondant à mon besoin. Il est disponible à l'adresse suivante :

https://www.schooldata.com/pdfs/US_FIPS_Codes.xls

2.2.2 Préparation des données

Rien de bien compliqué dans la préparation et la transformation de ces données. J'ai juste renommé les colonnes (pour les avoir en français) et modifié le type si besoin.



2.3 Les données sur la densité de population

2.3.1 Choix du jeu de données

Je me suis également procuré un jeu de données recensant les densités de population par état et par décennie pour le territoire américain, car je me suis dit qu'il serait intéressant de recouper ceci avec les données que j'ai par rapport à la mortalité et au nombre de blessés des tornades, données que j'ai déjà dans le premier dataset par tornade.

J'ai trouvé ce dont j'avais besoin sur le site officiel du « United States Census Bureau » qui est l'organisme en charge du recensement aux USA.

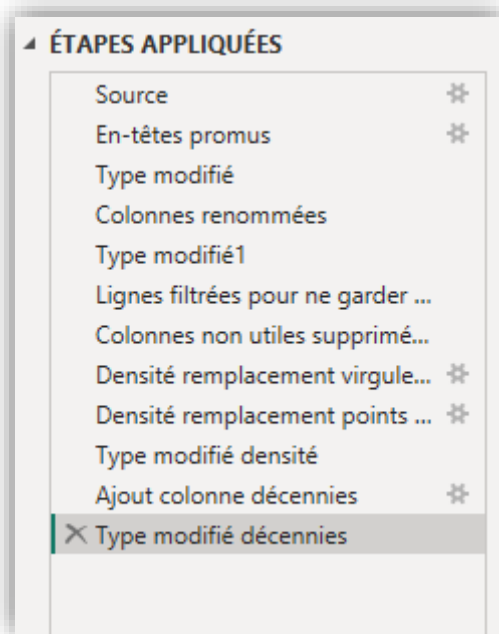
Le fichier que j'ai utilisé est disponible à l'adresse suivante :

<https://www2.census.gov/programs-surveys/decennial/2020/data/apportionment/apportionment.csv>

Il comprend notamment, et c'est ce qui m'intéresse, la densité par état (en global également) et par décennie de la population américaine, de 1910 à 2020. Il comprend également d'autres données qui ne nous intéressent pas dans le cadre de ce projet.

2.3.2 Préparation des données

La préparation des données pour ce dataset n'a pas été très poussée également.

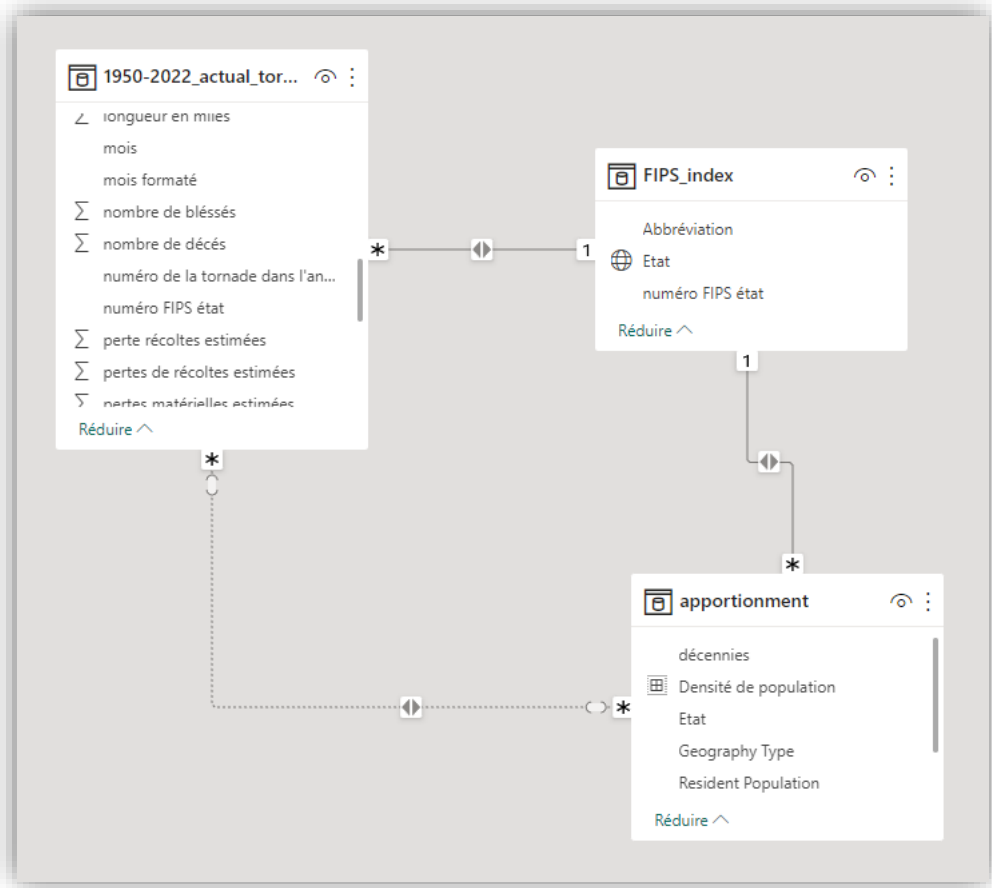


J'ai surtout supprimé les colonnes qui ne m'intéressaient pas, j'ai renommé les colonnes restantes, mais surtout, j'ai créé une colonne « décennie » qui me servira à faire le lien avec celle que j'ai déjà créée dans le dataset des tornades, afin que si l'on change de décennie, nous avons les données de densité correspondantes.

J'avoue que cette table, afin de rester cohérent avec l'ensemble, aurait mérité juste un peu plus de manipulations. Comme le renommage en français de toutes les colonnes utilisées plus tard.

2.4 Création du modèle de données

Après ce long travail primordial de préparation des données, il fut temps de créer le modèle de façon cohérente.



Ainsi, nous avons le modèle de données ci-dessus. Et je dois dire que je suis un peu perplexe. Ce n'est pas un modèle en étoile comme on l'entend (à moins de considérer la table FIPS_index comme la table de faits). De plus, j'ai dû ajouter une relation secondaire entre la table apportionnement et la table des tornades (relation décennies).

Malgré tout, dans Power Bi, ce modèle fonctionne donc je l'ai gardé. Même si, si j'avais eu plus de temps, j'aurais aimé l'améliorer.

Au niveau des relations nous trouvons donc :

- Actual_tornadoes.numéroFIPS = FIPS_index.numéroFIPS
- FIPS_index.Etat = apportionnement.Etat
- Apportionnement.décennies = actual_tornadoes.décennies (secondaire)

3 : LES VISUELS

3.1 Wireframe

Uniquement destiné à un usage technique, ce visuel pose les bases du thème que je veux utiliser pour l'ensemble du fichier. En présentation, il serait supprimé, ou tout simplement mis de côté, car il n'a pas d'intérêt dans l'analyse. Il me sert de « fourre tout » pour avancer plus vite par la suite dans la création des vrais visuels.

La raison pour laquelle je l'ai gardé ici, c'est pour montrer que j'ai tenté de créer un thème visuel pour la présentation qui soit cohérent avec celui du sujet abordé. En entreprise, j'utiliserais la charte graphique de celle-ci.

Ensuite, sur un aspect purement graphique, j'ai voulu ici créer l'atmosphère orageuse. Et dans le choix de mes couleurs, j'ai fait attention à garder le contraste qu'il faut pour que ce soit lisible, ainsi que des couleurs qui représentent toujours (plus ou moins) la même chose, et adaptées à la donnée représentée.

Néanmoins je précise que je suis daltonien, et que s'il y a des choses qui sont incohérentes à ce niveau, je m'en excuse par avance. (en entreprise je demanderais conseil à des collègues pour pallier cette difficulté...)



3.2 Présentation générales des tornades aux USA

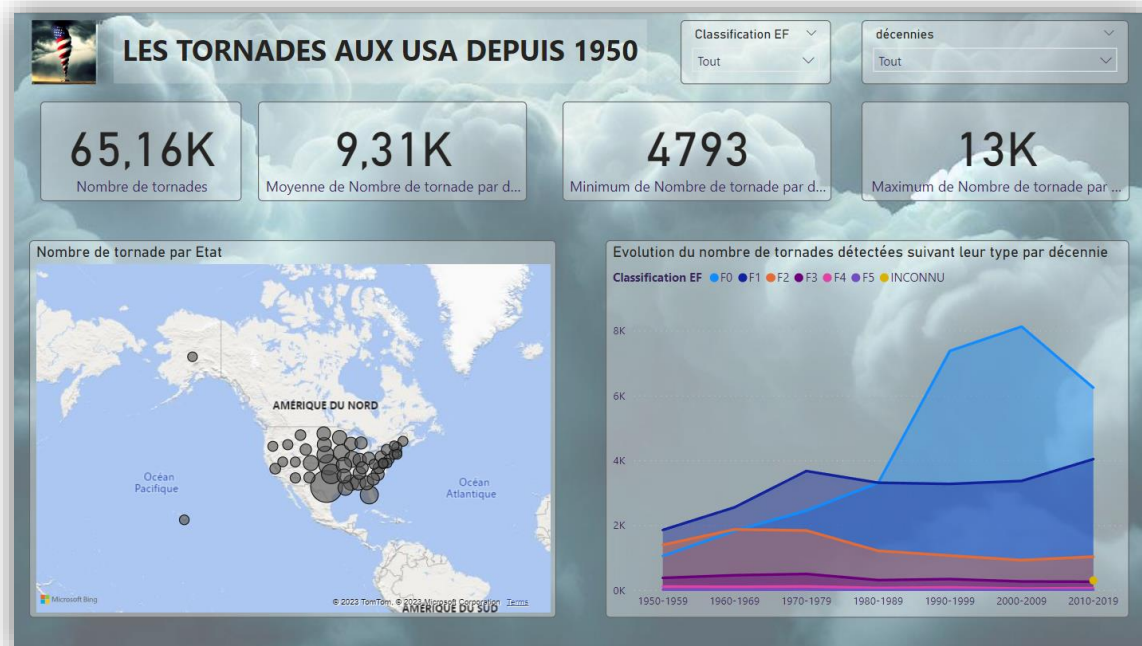
Visuel simple et général présentant un résumé des données au plus haut niveau de granularité.

On peut tout de même filtrer par classification EF (Echelle de Fujita représentant l'intensité des tornades) et par décennies.

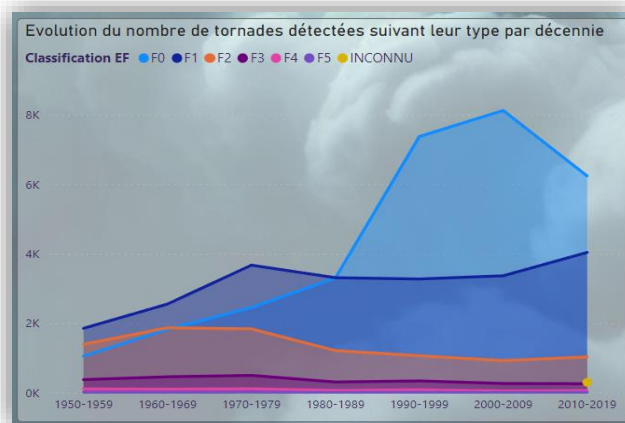
Pour ce visuel j'ai volontairement retiré la décennie 2020-2029 pour la raison évidente qu'elle n'est pas terminée donc très peu de données sur cette période.

Il y a trois choses très intéressantes à relever dans l'analyse de ce visuel :

- Forte augmentation des tornades EF0 à partir des années 1990
- Beaucoup moins de tornades à l'ouest des USA
- Si l'on filtre sur les EF5, forte concentration au centre et sud est des USA



3.2.1 Anomalie ou réalité pour les EF0 ?



Sur ce visuel, on note de façon flagrante une forte augmentation du nombre de tornades dans la décennie 90, alors que les autres restent relativement stables voire un peu à la baisse.

Il s'avère qu'après recherches, les outils de détection ont été améliorés justement dans les années 90.

Elles ne sont donc pas plus nombreuses, mais sont mieux détectées, ce qui explique cette augmentation.

Les années les plus tornadiques aux États-Unis restent celles de 2004 avec 1 813 tornades, 2008 avec 1 692 tornades et 2011 avec 1 691 tornades, **selon les données de la NOAA**. Et même si l'on constate une augmentation entre les années 1950, début des relevés officiels, et les années 1990, celle-ci est en grande partie liée à l'amélioration des outils de détection.

Source :

<https://www.futura-sciences.com/planete/actualites/meteorologie-y-t-il-plus-plus-tornades-104325/>

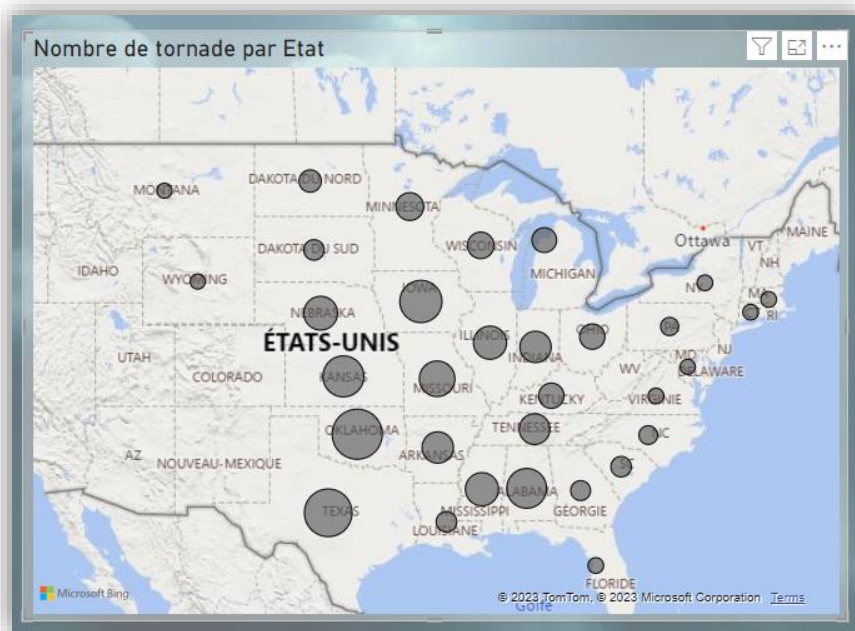
Ce que l'article ne dit pas en revanche, c'est que l'augmentation est surtout sur les tornades EF0, donc les moins puissantes (et par conséquent les plus difficiles à détecter).

J'aurais pu, en outre, vérifier comme l'article le fait, que le réchauffement climatique n'a pas d'influence sur le nombre de tornades.

3.2.2 Pourquoi moins de tornades à l'ouest ?

On peut remarquer que les tornades se concentrent surtout au centre et au sud est des USA. Mais beaucoup moins à l'ouest. Pourquoi ?

Sur le visuel ci-dessous, j'ai volontairement filtré sur les tornades les plus puissantes (EF4 et EF5) et on voit qu'il n'y a en pas du tout à l'ouest.



Si l'on regarde la carte de plus près, on peut voir le relief. Et il s'avère qu'il n'y en a pas à l'ouest des rocheuses, la chaîne de montagne la plus haute des USA !

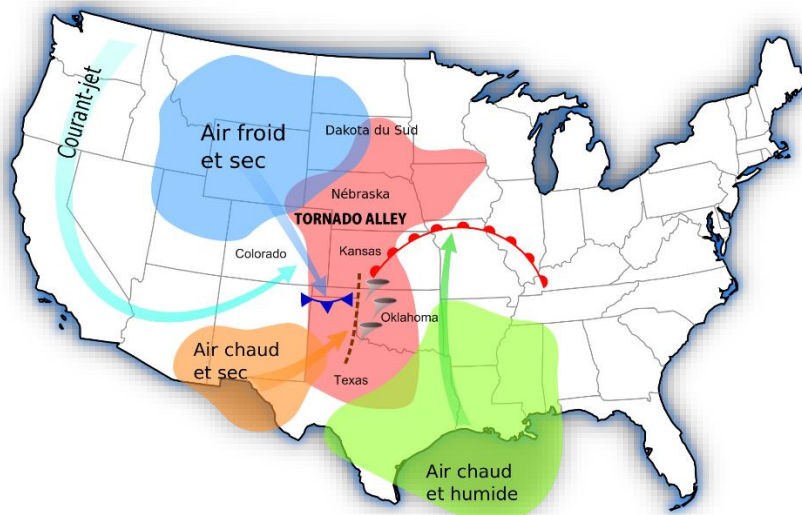
La raison est que les conditions climatiques (revoir première partie) pour la formation des tornades est un conflit de masses d'air. Cela signifie que la différence de masse d'air à l'ouest des rocheuses n'est pas assez importante pour créer des tornades d'envergure.

Ceci s'explique, simplement, par le fait qu'à un certain moment (que nous verrons plus tard), il y a un air froid et sec qui arrive du nord (Canada), et un air chaud et humide (golfe du Mexique) qui entrent en conflit. De plus, à l'ouest existent les Appalaches (moins hautes par contre), une autre chaîne montagneuse, ce qui amplifie le phénomène dit de couloir et provoque l'affrontement de plein fouet de ces masses d'air. Ce phénomène donne naissance à la Tornado Alley !

PROJET TORNADES USA – CAUDRON CEDRIC

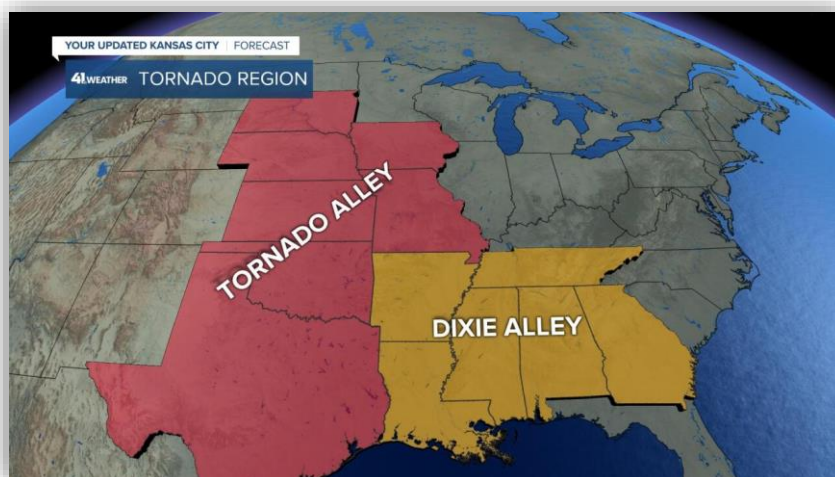
Par contre, on peut apercevoir, même si ce n'est pas flagrant ici, une zone comprenant l'Alabama où les tornades violentes sévissent quand même. Et ceci, après recherche, m'a fait découvrir l'existence de la Dixie Alley. Une autre zone propice aux tornades que je ne connaissais pas et beaucoup moins connue aux USA.

J'en reparlerai dans la partie suivante.



Source :

https://fr.wikipedia.org/wiki/Tornado_Alley



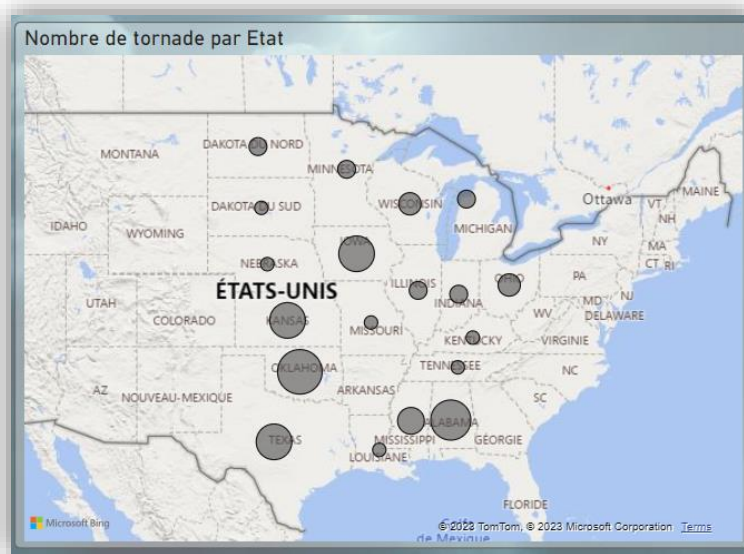
Source :

<https://www.kshb.com/weather/weather-special/tornado-alley-dixie-alley-last-2-severe-weather-seasons-compared-to-20-year-average>

3.2.3 Situation des EF5

Les tornades EF5 sont les plus puissantes existantes jusqu'à présent. Elles sont très rares (et heureusement), car ce sont des monstres qui ne laissent rien sur leur passage. J'en reparlerai plus tard quand j'analyserai les dégâts humains.

Ici nous allons nous concentrer sur leur localisation depuis 1950.



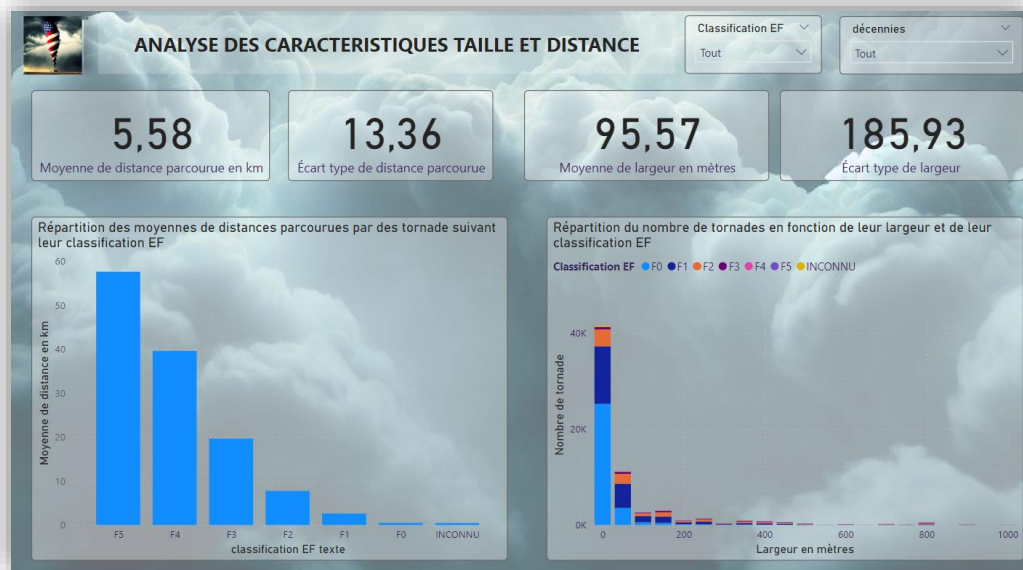
Nous voyons donc sur ce visuel que les EF5 sont vraiment concentrées au centre (plutôt sud) des USA, à l'exception de l'Alabama et le Mississippi.

Dans tous les cas, et cela confirme la thèse de la partie précédente, aucune ne se forme de part et d'autre des Rocheuses d'un côté, des Appalaches de l'autre.

Même si les EF5 ne sont pas fréquentes (59 depuis 1950), leur particularité d'être les plus puissantes, et donc d'avoir des conditions extrêmes pour se former, confirme les régions où elles se forment, et donc l'existence de la Tornado Alley et de la Dixie Alley (pour cette dernière vraiment j'étais étonné au vu de ces visuels et c'est après recherche que j'ai appris son existence).

De plus, concernant les EF5, j'émet une hypothèse personnelle (qu'il serait bon de vérifier), c'est que les données les concernant sont fiables (même les plus vieilles), ce qui sur les autres catégories on peut émettre des doutes. Car ce sont des phénomènes, qui, par leur puissance indescriptible, ne peut pas passer inaperçu, et ce même en pleine nature !

3.3 Analyse des caractéristiques taille et distance parcourue



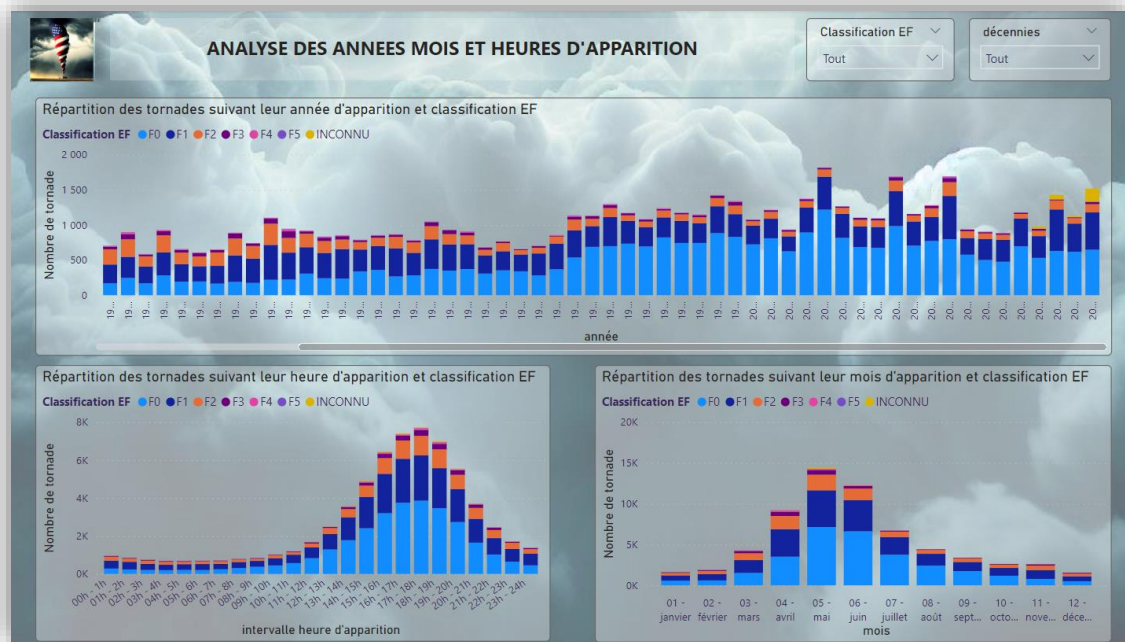
Dans ce visuel, on peut voir la répartition des tornades suivant leur distance parcourue, leur largeur, le tout suivant leur classification EF.

Sans grande surprise, les moyennes de distances parcourues sont dépendantes de la classification de la tornade (et donc de la puissance que le cumulonimbus a emmagasinée). On voit d'ailleurs que l'écart type n'est pas très important donc que les valeurs se situent toutes plus ou moins autour de la moyenne.

Concernant la largeur du tuba (donc la surface touchant terre de la base de la tornade), nous remarquons que beaucoup sont entre 0 et 50m de largeur, avec une moyenne de 95m environ, mais un écart type beaucoup plus important. Donc il y a une forte disparité de largeur dans les tornades. D'ailleurs, pour améliorer les visuels, j'ai restreint les graphiques avec une valeur limite que ce soit pour la distance parcourue ou pour la largeur, et ce dans un objectif de lisibilité.

Dernier point intéressant en filtrant sur la classification EF, on s'aperçoit qu'il y a une corrélation entre la puissance et la largeur de la tornade. Plus elle est puissante, plus elle est susceptible d'être large. J'aurais aimé implémenter un coefficient de corrélation (comme celui de Pearson) pour démontrer cela de façon plus scientifique, mais je n'ai pas eu le temps.

3.4 Analyse des moments d'apparition (années, mois, heures)



Dans ce visuel, j'analyse les données temporelles d'apparition des tornades.

Ainsi, sur le graphique du haut, on peut confirmer ce que j'ai dit plus haut qu'à partir des années 1990/1991 il y a beaucoup plus de F0 détectées.

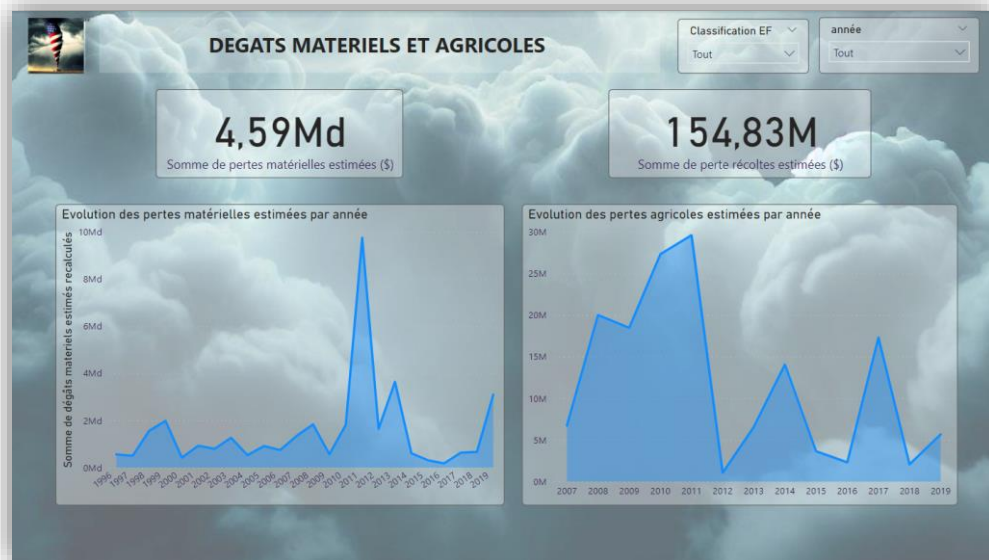
De plus, et si on se réfère à l'article de futura-sciences évoqué plus haut, on peut se rendre compte que j'arrive à la même conclusion qu'eux sur les années avec le plus de tornades répertoriées (2004, 2008 et 2011). Ouf car ils ont utilisé la même source de données !

Ensuite on peut également remarquer que les tornades se déclenchent surtout sur une période d'avril à juin, et entre 15h et 21h. Ce qui est logique au vu des conditions climatiques nécessaires à leur formation que j'ai déjà évoquées plus haut.

Au niveau de l'heure, cela correspond aux heures les plus chaudes de la journée, ce qui est également logique. Il y aura plus d'énergie si la masse d'air froide arrive pendant que l'air chaud est à son paroxysme dans la journée.

3.5 Analyse des pertes matérielles et agricoles

Ici je me concentre sur les années 1996 à 2019 pour les pertes matérielles (pour des raisons évoquées dans la première partie), et de 2007 à 2019 pour les pertes agricoles car les données commencent à cette date. Donc attention à l'échelle sur les graphiques (d'autant plus que d'un côté on est en milliards de dollars et de l'autre en millions).



De plus, je me dois d'attirer l'attention sur le fait que ce ne sont que des pertes estimées ! Ces données sont parmi les moins fiables du dataset (je reviendrai dessus plus tard), mais c'est juste pour donner une idée de l'enjeu économique des tornades pour les USA dont j'ai un peu parlé en introduction.

Les tornades coutent très chers. C'est la première conclusion à avoir ici.

La deuxième conclusion est qu'en 2011 il y a eu un énorme pic au niveau des dégâts matériels. Pourquoi ??

Il faudrait une étude bien plus approfondie sur cette année en particulier pour donner une réponse fiable. Néanmoins, j'ai deux pistes plausibles (et peut être concomitantes) :

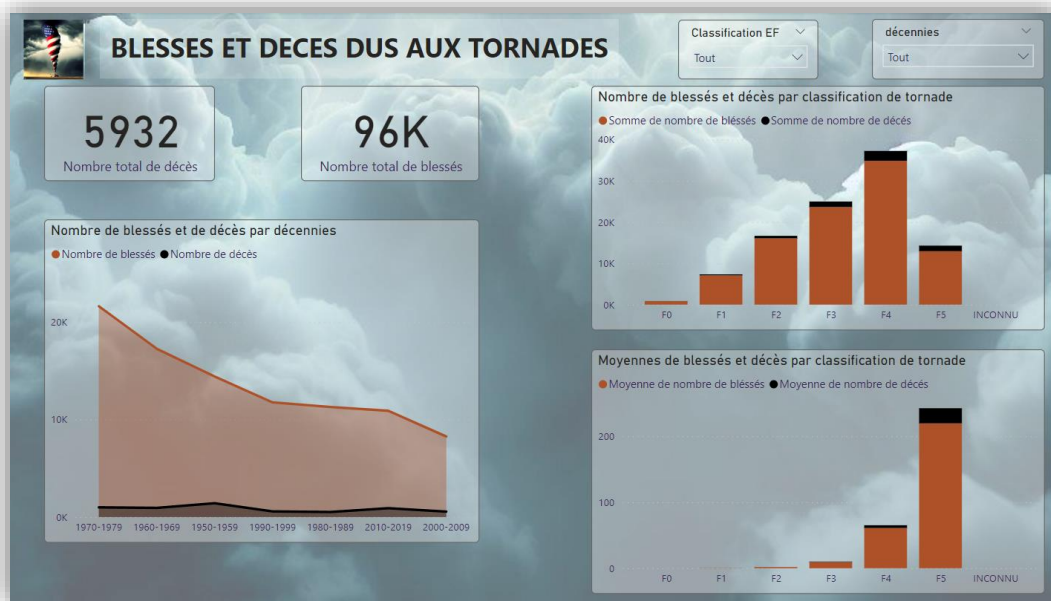
- Il y a eu plus de tornades en 2011. Chose que l'on a déjà observé plus haut, c'est la 3^{ème} année avec le plus grand nombre de tornades depuis 1950.
- Les tornades se sont concentrées dans des zones plus urbanisées (plus grande densité ?), ce qui expliquerait pourquoi les dégâts sont plus clairement plus importants alors que ce n'est pas l'année la plus prolifique en nombre de tornade. Ainsi la localisation est très importante pour évaluer ce paramètre.

Un élément de réponse pourrait être dans cet article de keraunos (observatoire français des tornades et orages violents (quasiment mon site de chevet lol)). Car il y a eu déjà fin avril un enchainement très exceptionnel de tornades dont plusieurs EF5 aux alentours de l'Alabama (donc dans la Dixie Alley). Ensuite, comme j'ai dit, il faudrait vraiment approfondir le sujet pour être sûr. Mais le ROI ne serait peut-être pas acceptable pour le moment dans le cadre de ce sujet de cours (mais personnellement ça m'intéresse!) .

Lien vers l'article :

<https://www.keraunos.org/actualites/fil-infos/2016/avril/tornades-outbreak-avril-2011-tuscaloosa-alabama-mississippi-usa>

3.6 Analyse des dommages humain des tornades



Nous arrivons maintenant au sujet le plus dramatique quant aux tornades. Les dommages humains qu’elles peuvent occasionner...

Ce visuel, volontairement très général, montre l’évolution du nombre de blesses et de décès occasionnés par les tornades aux USA depuis 1950.

Plusieurs points intéressants s’en dégagent.

Sur la courbe générale, on peut voir que le nombre de décès reste relativement stable. Mais que le nombre de blessés baisse énormément au fil du temps ! (Alors que la population aux USA et voir plus tard la densité de population également, je n’ai pas exploité ce recoupement en tant que tel, mais ça pourrait être intéressant !) L’éternelle question est donc « pourquoi » ?

Un élément de réponse, déjà abordé précédemment, est l’amélioration des moyens de détection ! En effet, on voit que cela baisse beaucoup jusqu’aux années 1990 où, comme mentionné plus haut, de nouveaux moyens de détections ont été mis en œuvre.

Il peut y avoir une multitude de paramètres expliquant la baisse des blessés, et ceux malgré les trois années particulièrement actives en tornades qu’on a vu plus haut, avec la forte augmentation de dégât matériels en 2011 !

Juste des idées comme ça :

- Amélioration de la détection comme déjà mentionné ?
- Meilleure communication du gouvernement (ou des états concernés car aux USA ceci je pense est de la responsabilité des états et non du gouvernement fédéral) et sensibilisation de la population ?
- Des tornades passant dans des endroits moins peuplés ? (même si je n’y crois pas trop à cette hypothèse)
- Ou alors un grand nombre de paramètres qui mériteraient une étude bien plus approfondie, avec des données au niveau de granularité que je n’ai pas, et qui demanderaient un temps incroyablement long à analyser...

Le deuxième visuel (en haut à droite) représente lui le nombre de blessés et décès bruts en fonction de la classification de la tornade. Déjà on se dit qu’il y a un problème. Ok les EF2,3,4 sont plus meurtrières que les EF5, mais les EF5 sont beaucoup moins nombreuses !

Et c’est la raison d’être du dernier visuel. La moyenne des blessés et décès par catégorie de tornade.

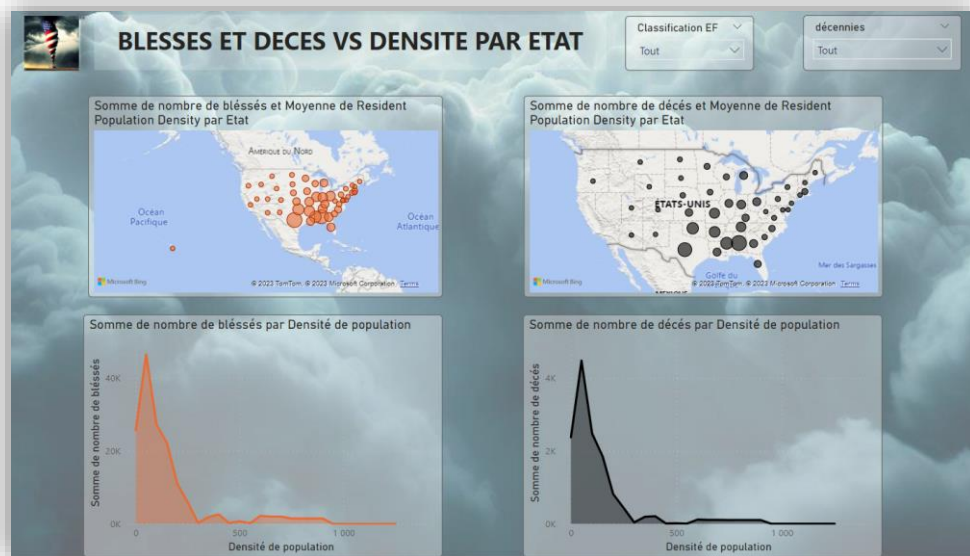
Et là on se rend compte, que les EF5 sont des monstres à qui rien ne résiste ! Quelque soit la décennie sélectionnée, c’est la championne en termes de dégât humain ! D’où mon « heureusement » qu’elles soient rares...

D’ailleurs à mettre en relation avec le visuel précédent sur les dégâts matériels (ce qui suppose qu’en 2011 elles sont passées dans des zones plus fortement peuplées), car à elles seules les EF5 ont provoqué plus de 4 milliards de dollars de dégâts estimés.

3.7 Analyse des dommages humain des tornades

Et c'est ainsi, au terme de ma réflexion, que j'arrive au dernier visuel (hors bonus) de cette présentation. Vérifier s'il y a une corrélation entre densité de population et dommages humains provoqués par les tornades.

Attention ! La densité de population est présente mais il faut survoler les bulles sur les cartes ! Ceci dû aux limitations de visualisation cartographiques de Power Bi...



Ce visuel n'aura pas eu les résultats que j'attendais, mais je devine pourquoi.

En effet, on remarque que les principaux dommages humains des tornades arrivent dans des zones qui ne sont pas très denses en matière de population.

Pourquoi ?

Je pense (c'est une interprétation personnelle) que cela vient du fait qu'une tornade est un phénomène très localisé !

En effet, la EF5 la plus puissante du monde pourrait ne faire aucune victime si elle reste en plein champs ! Alors qu'une tornade d'une moindre puissance pourrait en faire beaucoup si elle se déclenche en zone urbanisée.

Par contre une chose qui est sûre, corroborée par mes chiffres précédents, est que si une EF5 (ou même EF4 dans une moindre mesure) passe sur une zone urbanisée, elle fera des dégâts humains de toute façon ! Sauf si un moyen de prévention fiable est découvert et toute la population évacuée à temps.

Ce qu'il me faudrait pour évaluer tout cela de façon plus précise, ce serait une granularité de la densité de population bien plus fine (minimum au niveau du comté). D'utiliser le fichier de tornade le plus complet, avec toutes les problématiques qu'il engendre.

Car là je suis resté, autant au niveau des tornades qu'au niveau de la densité de population à la granularité « Etat », ce qui ne convient pas pour un phénomène aussi localisé.

Ainsi, ce visuel n'est pas probant et son analyse devrait rester de côté en l'état.

Ce visuel est ma déception de ce rapport, mais au final je sais pourquoi. Et encore une fois, il faudrait une étude bien plus approfondie, ce qui ne rentre pas dans le cadre de mon ROI pour ce projet (et sous réserve de trouver les données d'ailleurs...).

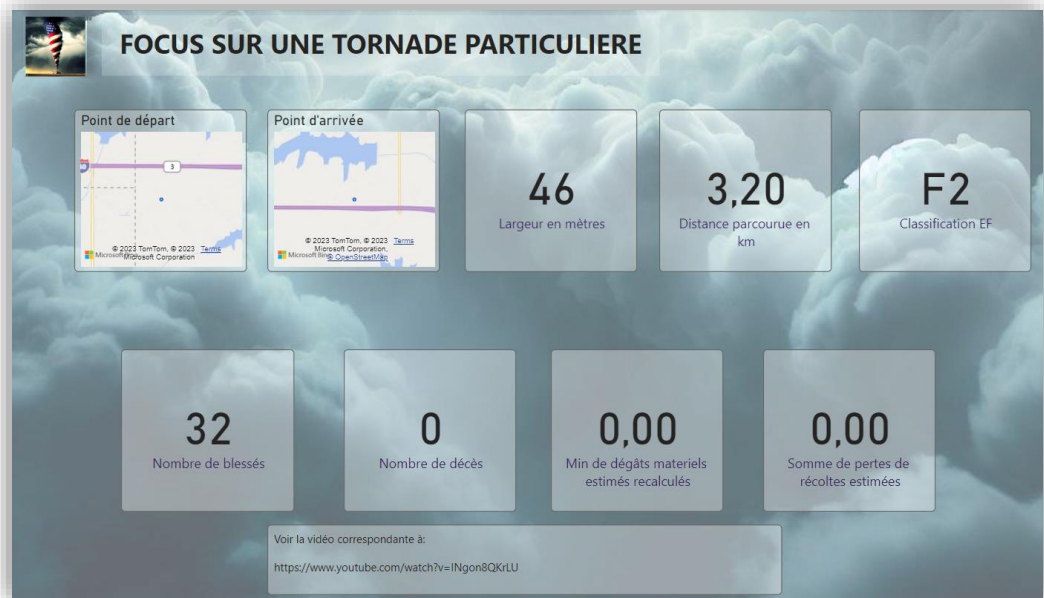
3.8 Bonus

Pour cet avant dernier visuel, j'ai voulu faire une excursion au niveau de granularité le plus fin ! Et le rattacher avec la réalité ! Car analyser des données, c'est bien, mais elles doivent être ancrées dans la réalité car cela n'aurait aucun sens (selon moi).

Ce visuel a deux buts pour moi :

- Vérifier la véracité des données en ma possession
- Voir jusqu'où, si j'avais le temps, je pouvais pousser l'analyse

Ainsi, j'ai fait l'étude d'une tornade en particulier, en relation avec une vidéo YouTube la filmant !



Avec les données en description de la vidéo, je n'ai pas eu de mal à retrouver la tornade en question. Elle fait d'ailleurs partie des tornades de 2011 (mais pas de celles fin avril qui furent exceptionnelles).

Survenue près d'Oklahoma City (entre Oklahoma City et Shawnee, autour de l'autoroute 3, celle où les youtubeurs roulent et la filment), les données que j'ai ressorties sur cette tornade sont cohérentes avec ce que l'on peut visualiser, à un détail près...

En effet, les données que j'ai sur elle mentionnent 0\$ en dégât matériels. Or sur la vidéo, on peut clairement voir qu'elle a fait des dégâts (notamment le camion qui est « pulvérisé » d'après les dires d'un protagoniste de la vidéo.

C'est ceci qui m'a mis en garde contre la véracité de la donnée concernant les dégâts matériels, et sur lesquels j'ai mis en garde plus haut.

Il peut y avoir énormément de causes à cet état de fait :

- Déjà la mise en garde dans la documentation comme quoi une entrée de 0 ne veut pas dire 0\$
- Le fait que le ou les propriétaires des biens n'auraient pas déclarés leur pertes...
- Le fait que la NOAA n'aurait pas répertorié ces pertes pour une raison ou une autre.
- Ou plein d'autres raisons... Pour cela également que c'est une **estimation** des pertes.

Mais dans tous les cas, cela m'aura permis de vérifier que, même pour la granularité la plus fine, on peut retrouver les informations pour une tornade particulière dans mon jeu de données. (et aussi prouver que ce que les youtubeurs en question ne sont pas des affabulateurs).

Et encore une fois, c'est un peu le leitmotiv de ce rapport, pour chacune des données une étude plus approfondie serait nécessaire afin d'analyser tout ce qu'il peut apporter !

3.9 La dernière

Je vais passer extrêmement rapidement sur ce dernier visuel. J'ai juste voulu donner à l'utilisateur la possibilité de voir les caractéristiques principales de n'importe quelle tornade dans la base. Il aurait fallu que je la travaille plus car pas très ergonomique au niveau de la sélection. De plus j'aurais aimé restreindre la sélection à une tornade particulière sans possibilité de sélectionner une date mais apparemment Power Bi ne le permet pas.

4 : CONCLUSION

Ainsi se termine ce rapport sur mon analyse des tornades aux USA entre 1950 et 2019.

Je pense avoir réussi à ressortir des informations clés sur la formation de celles-ci, tout en restant général (trop à mon gout), ainsi que des recoupements qui sont pertinents (ou du moins s'ils ne le sont pas j'explique ce qu'il me faudrait pour qu'ils le soient).

Au final, ce projet, bien qu'ambitieux et difficile (je suis sorti de ma zone de confort pour le réaliser, et si je pense avoir réussi c'est car je suis passionné par le sujet), m'aura également permis de me concentrer sur le principal (merci de m'avoir martelé le fameux ROI), et ce même s'il a finalement totalement changé de direction par rapport à mon intention première (bien trop ambitieuse qui était de recouper des données météorologiques pour montrer dans quelles conditions se forment les tornades).

Dans tous les cas, ce projet m'aura permis de me replonger dans ce phénomène qui me fascine, m'aura en plus permis d'apprendre de nouvelles choses ! Et clairement c'est pour moi le principal, bien au-delà de la note qu'il m'apportera (qui j'espère ne sera pas trop mauvaise tout de même lol)

PS : désolé pour la forme de ce document. Je voulais à la base faire une table des illustrations et regrouper toutes les sources dans une bibliographie, mais clairement il m'a demandé tant de temps (en une semaine car j'ai tout repris de zéro) que franchement je n'ai plus le courage de le faire. Merci de votre compréhension.