Visualisation de l'information Projet

Rapport final

Groupe 12

Julien Albert Youri Derlet Christian Nazili W.

17 décembre 2018

Table des matières

1	Introduction	1				
2	Domaine	1				
3	Analyse des données (What)	1				
4	4 Pitch (Why)					
5	Utilisateurs (Who)	3				
6	Définition des tâches (How)	4				
	6.1 Les visualisations (tâches, actions et cibles)	4				
	6.2 Prototypage low-fi	5				
7	Prototypage hi-fi	5				
	7.1 Tâche 1 : découvrir la tendance	5				
	7.2 Tâche 2 : découvrir la concentration	5				
	7.3 Tâche 3 : découvrir la distribution	7				
	7.4 Tâche 4 : agir	7				
8	Considérations générales	8				
9	Conclusion	8				

1 Introduction

Sommes-nous seuls dans l'univers? Qui ne s'est jamais posé cette question? Personne. Certains pensent même que nous avons déjà reçu des visites d'êtres provenant d'un autre système solaire de la Voie Lactée semblable au nôtre.

Le but principal de ce projet est de réaliser une visualisation à partir de témoignages concernant les apparitions d'objets volants non identifiés (OVNIs) dans le monde. La visualisation est à réaliser en JavaScript en utilisant la bibliothèque d3js. Ce rapport a pour but de présenter la visualisation réalisée en utilisant le cadre d'analyse en trois parties vu au cours : le What, Why and How.

Ainsi, ce rapport se découpe comme suit. Une première section décrit le domaine, c'est-à-dire le contexte et les objectifs dans lesquels s'inscrit notre visualisation. La deuxième expose une analyse des données mises à notre disposition (What) en décrivant notamment leurs caractéristiques. Ensuite, nous définissons les tâches à réaliser (Why). Cette section présente successivement l'idée générale de la visualisation (le pitch), les utilisateurs potentiels (Who) et la définition proprement dite des tâches. Finalement, nous proposons un premier prototype basse fidélité et un prototype haute fidélité matérialisant notre proposition. Nous terminons par quelques considérations générales sur la démarche qui a guidée l'élaboration de la visualisation et une conclusion.

2 Domaine

Pour ce travail, nous avons décidé de nous mettre dans la peau d'un lobby américain ayant pour objectif de promouvoir l'augmentation du budget de la défense. Dans la perspective des élections de mi-mandat, notre mission est de favoriser chez l'électeur l'émergence d'un état d'esprit susceptible de le pousser à choisir des candidats promouvant l'augmentation du budget de la défense.

Notre objectif est de créer une visualisation afin de promouvoir la création d'une composante spatiale pour l'armée US ("Space Force"), avec un financement adéquat. Cet outil doit utiliser les nombreuses observations d'OVNI recensées dans le jeu de données à notre disposition pour montrer l'existence d'une activité extraterrestre importante. Il doit également mettre l'accent sur le fait que nous ignorons les intentions des extraterrestres et qu'il convient donc de prendre des mesures de précaution adéquates. Nos cibles sont principalement des électeurs peu éduqués et/ou potentiellement sensibles aux discours complotistes.

3 Analyse des données (What)

Les données dont nous disposons pour notre travail concernent des témoignages recueillis sur une plateforme internet et centralisés sur le site Kaggle ¹. Ceux-ci concernent des objets volants non identifiés (OVNI)
observés dans le monde entier. Nous disposons de trois jeux de données au format json contenant 500,
3000 et 80332 observations. Ces fichiers sont nommés respectivement : ufo_500, ufo_3000 et ufo_all. Chacun de ces fichiers comporte les mêmes caractéristiques. Celles-ci sont reprises et expliquées dans le Tableau 1.

Il s'agit de jeux de données de type "table" encodés au format json où chaque entrée est une observation caractérisée par les attributs décrits ci-dessus. Chaque entrée peut également être considérée comme un point (via ses latitudes et ses longitudes). Ces jeux sont donc également de type géométrique.

^{1.} https://www.kaggle.com/NUFORC/ufo-sightings

Caractéristique	Sémantique	Type	Catégorie	Remarques
			(Dataviz)	
date posted	date où l'observation	string	ordinale	non utilisée
	a été reportée			
datetime	date de l'observation	string (au for-	ordinal	1, 24 et 694 erreurs de
	(temps GMT)	mat datetime)		conversion
duration	durée de l'observation	string	quantitative sé-	non utilisée
(hours/min)	(en heures/minutes)		quentielle	
duration (se-	durée de l'observation	integer	quantitative sé-	trois valeurs non en-
conds)	(en secondes)		quentielle	tières dans ufo_all
shape	forme de l'objet ob-	string	catégorielle	21, 23 et 30 formes
	servé			différentes, présence de
				valeurs vide mais pas
				de doublon ni de valeur
				erronée
city	ville où a eu lieu l'ob-	string	catégorielle	pas de valeur vide
	servation			
comments	commentairesde l'ob-	string		valeurs vides dans
	servateur			ufo_3000 et ufo_all ,
				casse variable
country	pays où a eu lieu l'ob-	string	catégorielle	liste de valeurs : "au",
	servation			"ca", "gb", "us", "de"
				et "" (valeur vide)
state	état où a eu lieu l'ob-	string	catégorielle	valeurs vides
	servation			
latitude	latitude où a eu lieu	float	quantitative di-	une valeur non-réelle
	l'observation		vergente	dans ufo_all
longitude	longitude où a eu lieu	float	quantitatif	
	l'observation		divergente	

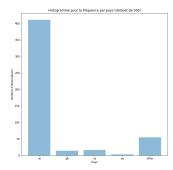
TABLE 1 – Synthèse des différentes caractéristiques présentes dans chaque jeu de données.

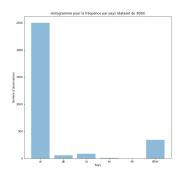
Concernant les données géographiques, nous constatons une surreprésentation des observations pour les États-Unis. En effet, environ 80% des données concerne ce pays, comme les barcharts de la Figure 1 le montrent. Nous notons également que l'Australie, le Canada, la Grande-Bretagne et l'Allemagne sont également représentés. Par contre, les autres pays sont regroupés dans une catégorie *other*. De manière pratique, nous suivons les conseils du professeur et utiliserons uniquement les latitudes et les longitudes. Les attributs pays, état et ville seront recalculés si nécessaire. Cette décision est également motivée par la moindre qualité de ces caractéristiques dans les jeux de données.

Les jeux de données proposés contiennent des observations datant des années 1940 jusqu'aux années 2010. Comme le montre la Figure 2, il y a une augmentation du nombre d'observations dans le temps. Elles augmentent faiblement, de manière linéaire, à partir des années 40 jusqu'au milieu des années 90 pour augmenter exponentiellement par la suite. Nous envisageons ces données temporelles uniquement de manière ordinale.

4 Pitch (Why)

Comme nous venons de l'observer, le nombre d'apparitions ne cesse de croître depuis que les données sont récoltées. Serait-ce un signe d'une présence avérée d'êtres venus d'ailleurs dans l'univers? Devons-nous craindre une colonisation à venir?





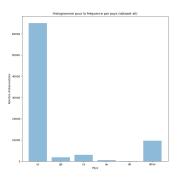
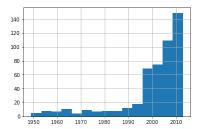
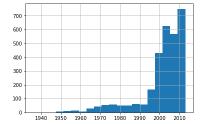


FIGURE 1 – Chaque barchart représente la distribution des observations partout dans le monde. Celui de gauche concerne le premier jeu de données à savoir ufo_500, le second ufo_3000 et celui de droite ufo_all.





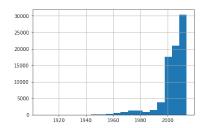


FIGURE 2 – Chaque histogramme décrit l'évolution du nombre d'observations au cours des années 1940 à 2010, en séquençant l'axe temporaire en 20 bâtons. L'histogramme de gauche concerne le premier jeu de données à savoir ufo_500 , le second ufo_3000 et celui de droite ufo_all . https://github.com/JulienAlbert/unamur-dataviz.git

De même, bon nombre de ces informations collectées concernent les États-Unis. Cela signifierait que l'invasion concerne principalement cette région, où la technologie est en pleine expansion. Ou alors cela indique que l'armée américaine créée de nouveaux outils et les tests sur son territoire... Où sont ces lieux?

Dans un même temps, Donald Trump décide de créer une sixième branche de l'armée : une *Space Force*. Cette armée permettrait de surveiller le trafic spatial et nous protéger d'envahisseurs. Ce 18 juin, il a déjà signé une directive de politique générale sur la gestion du trafic dans l'espace [1].

Pour toutes ces raisons, nous avons décidé de créer une visualisation afin d'influencer le vote des électeurs. Celle-ci sera destinée à des utilisateurs naïfs américains. Elle aura pour but de leur faire peur afin de les inciter à soutenir le programme républicain de leur président. Présentons dans la section suivante, deux personnages types.

5 Utilisateurs (Who)

L'utilisateur type de notre visualisation est le citoyen américain lambda qui désire découvrir des conspirations et qui se laisse tromper facilement par des "preuves" légères et réfutables rapidement, comme celles énoncées précédemment. Notre visualisation va lui permettre d'affirmer ses craintes et son idéologie. Elle va lui montrer que son pays se fait observer par des OVNIs plus qu'ailleurs dans le monde et que ce phénomène devient de plus en plus récurrent.

Persona 1 : John a 50 ans, il est ouvrier dans une usine au Texas, est divorcé depuis peu et a deux enfants qu'il voit de temps en temps. Il va régulièrement boire un verre avec des collègues au café du coin après le boulot. Sur les conseils d'un de ses enfants, il a acheté un ordinateur récemment afin d'être mieux informé.

Persona 2 : Jim a 30 ans, il est sans emploi régulier et vit chez sa mère. Il passe beaucoup de temps sur son ordinateur. Outre les jeux vidéos, il regarde beaucoup de vidéos commentant l'actualité à la recherche de points de vue différents des médias traditionnels. Il a le sentiment que ces derniers cachent des choses...

6 Définition des tâches (How)

Dans cette section, nous allons définir comment nous allons construire notre visualisation pour satisfaire nos deux utilisateurs. Celle-ci devra répondre aux exigences de notre pitch.

6.1 Les visualisations (tâches, actions et cibles)

Les tâches nécessaires pour que notre visualisation finale atteigne son but sont les suivantes.

- 1. La première tâche de notre utilisateur est de **découvrir** l'augmentation des observations dans le temps de manière générale.
- 2. La deuxième est de **découvrir** la forte concentration des observations aux États-Unis (par rapport aux autres pays).
- 3. La troisième est d'une part que l'utilisateur **découvre** la répartition des observations dans les différents états des États-Unis, et d'autre part qu'il **recherche** les chiffres des états qui l'intéressent davantage.
- 4. (bonus) La quatrième est de pousser l'utilisateur à agir en soutenant le parti républicain.

Ces trois visualisations permettront aux utilisateurs de **découvrir** une problématique : l'invasion des extraterrestres qui est en cours et la généralisation du phénomène à tous les états américains. Comme tâche bonus, nous avons décidé d'inviter nos utilisateurs à **agir** (en soutenant le parti républicain). Nous les inviterons ainsi à consulter des sites de ce parti politique. En résumé, les actions associées sont : découvrir (et rechercher) pour agir.

Pour y arriver, nous avons besoin de visualiser ces informations. Pour se faire nous avons travaillé sur les données du jeu de données complet (ufo all). Premièrement, un line chart doit être réalisé afin de montrer la tendance à l'augmentation des observations dans le temps. Le nombre d'observations par année va permettre de montrer à l'utilisateur qu'il y a une augmentation de celles-ci. Une phrase de transition accrocheuse lui suggérera de vérifier où les apparitions ont été les plus visibles. Deuxièmement, le barchart permettra de voir la distribution du nombre d'observations par pays. Les utilisateurs découvriront que c'est aux État-Unis qu'il y en a le plus. Une autre phrase lui suggérera de voir une représentation selon les états américains. Troisièmement, les utilisateurs consulteront alors une visualisation géographique des données via une carte des États-Unis (données spatiales). Nous avons décidé de réaliser une carte découpée par états avec une couleur rouge uniforme afin que se soit les utilisateurs mêmes qui choisissent le lieu pour lequel ils désirent recevoir de l'information : en premier lieu là où ils habitent, ensuite les états voisins. Et comme la couleur est partout la même, inconsciemment cela leur induira une crainte. Quatrièmement, les utilisateurs seront invités à agir sur base des informations présentées. Ce qui est réalisé par le biais d'un texte présentant la solution à la menace avec des liens pour soutenir le parti républicain. La cible de cette tâche est l'interprétation des données que nous cherchons en l'occurrence à contraindre dans une direction qui nous est favorable. Pratiquement, l'interprétation des informations présentées est canalisée par les petits textes intermittents guidant la lecture des visualisations, mais également par la construction verticale de la page favorisant une lecture linéaire et le sentiment de se voir raconter une histoire.

En résumé, les cibles associées sont : tendance, distribution, données spatiales et interprétation. Voici le tableau suivant pour récapituler les visualisations avec leur cible :

Tâche	Visualisation	Cible
Découvrir l'augmentation des observations	Line chart	Tendance
Découvrir la concentration aux États-Unis	Barchart	Distribution
Découvrir la répartition dans les états	Map	Données spatiales
Agir en soutenant le parti républicain	Textes, liens et structure de la page	Interprétation

6.2 Prototypage low-fi

La Figure 3 est un petit prototype sur papier représentant une vue intégrale de notre visualisation. Il s'agit d'une page Web que l'utilisateur pourra faire défiler pour naviguer entre les trois représentations : la première est donc le line chart, la seconde est bien entendu le barchart et la dernière est une *cacahuète*, à savoir la map des États-Unis divisée en état. Le fil conducteur de la page web est de raconter une histoire via une lecture verticale de haut en bas. La navigation sera pensée afin de favoriser ce sens de lecture, et donc le caractère narratif de la visualisation (nous nous inspirons ici du *storytelling* qui est une technique de communication consistant à mettre en récit un message afin de favoriser son adoption [2]). Cette narration se voudra percutante et compréhensive pour nos utilisateurs.

7 Prototypage hi-fi

7.1 Tâche 1 : découvrir la tendance

L'objectif de la première tâche est que l'utilisateur découvre que les observations d'OVNIs augmentent. Nous utilisons ici un line chart car il s'agit d'un bon idiome pour montrer le comportement d'un attribut quantitatif (le nombre d'observations) en fonction d'un attribut ordinal (l'année) [3]. Le focus est évidemment sur la ligne démontrant cette tendance. Celle-ci est donc représentée de manière assez épaisse et en rouge afin d'attirer l'attention au maximum. Le rouge permet également de renforcer le sentiment d'importance pour l'utilisateur de l'information communiquée. Les canaux utilisés pour la ligne sont la position verticale pour le nombre d'observations et la position horizontale pour l'année. Les autres éléments graphiques (axes, graduation, titres, etc.) ont été voulu sobres et simples, par contraste avec la ligne, afin de ne pas perturber la perception de la visualisation. Ils ont juste pour but de permettre de comprendre la signification de la ligne. Sur le plan technique, cette visualisation utilise une agrégation par année du jeu de données de départ. Le code utilise la librairie D3.js et est une adaption de celui qui a été présenté par M. Antoine Clarinval.

7.2 Tâche 2 : découvrir la concentration

La deuxième tâche consistant à montrer la forte concentration des observations d'OVNIs aux États-Unis par rapport aux autres pays, l'utilisation du barchart nous semblait évidente. Il s'agit en effet d'un idiome adapté à la comparaison de quantités pour un nombre peu important de catégories [3]. Dans notre cas, différentes lignes représentent les pays via le canal positionnel horizontal et les quantités associés sont représentées en utilisant le canal positionnel vertical. Comme pour la tâche précédente, l'utilisation de la couleur permet, d'une part, d'appuyer sur le message que nous voulons faire passer, à savoir la forte concentration des observations aux États-Unis, et d'autre part, de favoriser le sentiment d'importance de l'information communiquée à l'utilisateur. Le design des autres éléments graphiques est motivé par les mêmes raisons que

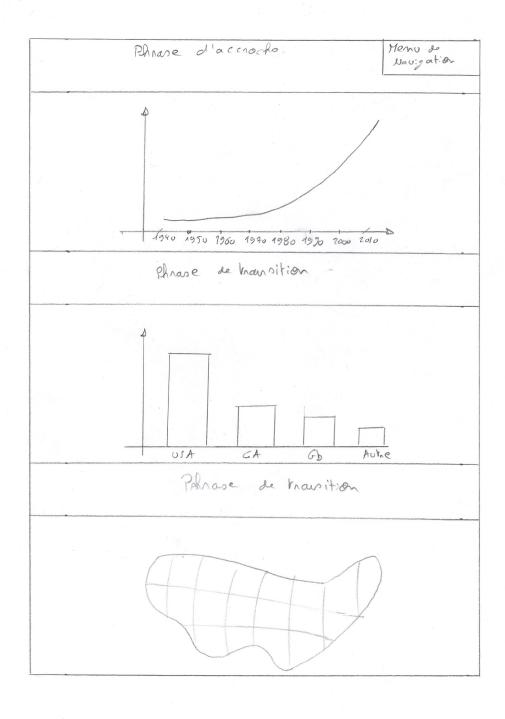


FIGURE 3 – Prototype papier.

la tâche précédente.

Sur le plan technique, nous avons cette fois utilisé **AmCharts**. C'est une librairie Javascript qui permet de faire aussi de la visualisation. Notre choix a été d'utiliser cette librairie parce qu'elle nous paraissait plus simple que D3.js. Comme nos données étaient au format json, nous avons utilisé un format json que amChart offre pour l'importation des données, la fonction s'appelle datasource. Une fois que les données sont bien chargées, nous avons utilisés **XY Chart** pour afficher notre barchart. Nous avons enlevé la couleur par défaut pour mettre des couleurs personnalisées. **AmCharts** offre exactement les mêmes fonctionnalités que D3.js pour un nombre de lignes de code nettement inférieur. **AmCharts** peut être implémenté en TypeScript/ES6, Javascript et Json. Nous avons utilisé Javascript.

7.3 Tâche 3 : découvrir la distribution

La troisième tâche est double. Dans un premier temps, il s'agit de montrer dans un premier temps que les observations sont réparties sur l'ensemble des états américains. Et dans un second temps, l'utilisateur doit pouvoir obtenir les chiffres des états pour lesquels il a un intérêt particulier (par exemple, là où il habite). L'idée sous-jacente est qu'il s'approprie davantage le message véhiculé par le biais de notre visualisation en manipulant les données présentées, avec pour but d'influencer son comportement in fine.

Pratiquement, nous utilisons un fond de carte centré sur les États-Unis et de teinte sombre (choix justifié par la thématique du jeu de données). Sur celui-ci sont dessinés des polygones rouges représentant les différents état. À noter que le choix de la couleur est motivé par les mèmes raisons que précédemment. Enfin, lorsque l'utilisateur survole un état, celui-ci passe en surbrillance (via une augmentation de l'opacité du polygone) et le nombre d'observations s'affiche. Les canaux utilisés sont les positions horizontale et verticale pour la localisation géographique et la couleur pour identifier les zones d'intérêt. Techniquement, le code utilisé est une adaptation de celui qui a été présenté par M. Antoine Clarinval (utilisant D3.js et l'API Mapbox). Les données ont été agrégées par état et intégrées dans un GeoJSON avec les polygones correspondants [4, 5].

7.4 Tâche 4 : agir

Enfin, la dernière tâche, qui englobe les tâches précédentes et permet de réaliser l'objectif principale de notre visualisation, consiste à pousser l'utilisateur à soutenir le parti républicain. L'objectif est de travailler sur l'interprétation des données présentées afin de construire une narration susceptible de favoriser le comportement souhaité. La structure de la visualisation est le premier élément sur lequel nous avons capitalisé pour y arriver. Elle est en effet construite comme une succession de séquences ayant pour but d'amener l'utilisateur à agir. Le sens de lecture est également renforcé par des liens présents dans les textes des différentes séquences afin de pouvoir accéder à la suite. Enfin, un menu a été ajouté simplement pour faciliter d'autres parcours.

Le second élément est la rédaction de textes permettant de raconter effectivement l'histoire et de contraindre l'interprétation des données afin que celle-ci appuie le message général. Tout d'abord, nous cherchons à susciter l'intérêt de l'utilisateur potentiel via le texte d'introduction. Ensuite, chaque séquence est composée d'un titre cherchant à interpeller l'utilisateur et orienter son interprétation, de la visualisation proprement dite, et enfin d'un petit texte rappelant l'interprétation suggérée et présentant des arguments jouant sur différentes cordes émotionnelles afin d'engager l'utilisateur à poursuivre l'histoire. À noter que l'enchaînement de ces séquences crée un effet de dramatisation (nous passons d'une simple question à une menace de grande ampleur). Enfin, le texte de conclusion capitalise sur cette dramatisation pour proposer à l'utilisateur la seule action susceptible de lui offrir un happy ending: soutenir le parti républicain! Graphiquement, nous avons préféré jouer uniquement sur la typographie afin d'accentuer certains éléments sans nuire à la lecture des

visualisations.

Le choix du template a été fait en fonction de l'utilisateur. Étant donné qu'il s'agit pour lui de découvrir une nouvelle information et d'agir en conséquence, nous sommes partis dans l'optique de faire une page html sur une seule structure, lui permettant de naviguer facilement. La page a été réalisé avec HTML5 et fonctionne bien avec Mozilla et Google Chrome (versions à jour). Les couleurs sont choisies pour attirer la curiosité de l'utilisateur. À noter que la couleur de fond change au fur et à mesure du défilement de la page.

Bien que la visualisation ait été inspirée pas plusieurs pages [6, 7], nous avons décidé de partir d'une page html complètement vide pour ensuite commencer à créer, afficher et importer les éléments que nous voulons afin de bénéficier de davantage de liberté. Pratiquement, la plupart des assets ont été téléchargés sur site HTML5 [8], puis modifiés pour avoir ce que nous avons mis dans notre prototype. Finalement, le code a été segmenté afin de faciliter sa lisibilité et le debugging.

8 Considérations générales

L'orientation choisie pour la visualisation consiste à prendre le contre-pied de l'honnêteté intellectuelle qui devrait guider un travail de ce genre. Nous souhaitons par ce biais réfléchir à la question de la manipulation de l'information et aux motivations qui peuvent sous-tendre des actions de ce genre.

Outre la question de la véracité de ce qui a été observé (et de leurs multiples explications possibles), nous constatons que ces données peuvent être interprétées de manières fort différentes. Par exemple, l'augmentation des observations dans le temps peut être expliquée comme suit. Après la création d'une plateforme de récolte de données, celle-ci compte d'abord peu d'utilisateurs, puis de nombreux utilisateurs contribuent, par la suite, à cette centralisation de l'information. De plus, le récolte d'informations sans internet est plus difficile qu'avec. Ainsi depuis qu'internet est en expansion, il est plus facile de récolter ce type de données aussi. Dans le même ordre d'idée, le fait que le site internet qui recense les observations soit américain est sans doute une explication plus plausible de la concentration des observations aux États-Unis. Et pour finir, le nombre d'observation par état est sans doute proportionnel au nombre d'habitant de cet état et/ou au type de personnes présent dans celui-ci. Par exemple, un état où bon nombre d'habitants aiment les théories du complot comportera normalement plus d'observations d'OVNIs.

Au-delà de l'importance d'être vigilant face à des outils qui ne sont pas simplement des présentations de données mais qui sont de véritables discours. Nous pensons qu'il est important de réfléchir également aux intérêts et motivations sous-jacentes à la création de ceux-ci. Cette réflexion passe obligatoirement par une meilleure connaissance du contexte et des intervenants du domaine d'intérêt. Par exemple, dans notre cas, la création d'une *Space Force* permettrait d'avoir le monopole sur les objets stellaires lancés dans l'espace. Elle protégerait les astronautes et satellites américains d'attaques provenant d'autres nations de la terre [1].

9 Conclusion

La visualisation que nous avons réalisée répond entièrement à nos utilisateurs naïfs qui sont à la recherche d'informations cachées et nouvelles. Nous avons réalisé une histoire linéaire (storytelling) afin de prendre le contre-pied de l'honnêteté intellectuelle en "manipulant" les données. Nous avons également pu prendre conscience de l'impact que peut avoir une visualisation de données, notamment sur l'interprétation. De plus, la mise en évidence du lien entre transmission d'une information et l'action qui peut en découler nous a beaucoup

interpellés. Enfin, de manière générale, nous avons apprécié ce projet pour la réflexion qu'il a suscité. Nous avons d'ailleurs clairement privilégié cette aspect au détriment de l'enjeu technique moins intéressant selon nous.

Références

- [1] NATIONALGEOGRAPHIC, Would a U.S. Space Force Be Legal? Get the Facts., 9 août 2018. https://news.nationalgeographic.com/2018/06/space-force-trump-legal-military-role-satellites-science/, dernière consultation le 1 novembre 2018.
- [2] WIKIPÉDIA, Storytelling (technique). https://fr.wikipedia.org/wiki/Storytelling_(technique), dernière consultation le 16 décembre 2018.
- [3] T. Munzner, Visualization Analysis and Design. AK Peters Visualization Series, CRC Press, 2015.
- [4] MSHAFRIR, US states in JSON form. https://gist.github.com/mshafrir/2646763, dernière consultation le 16 décembre 2018.
- [5] PUBLICAMUNDI, us-states.json. https://github.com/PublicaMundi/MappingAPI/blob/master/data/geojson/us-states.json, dernière consultation le 16 décembre 2018.
- [6] BOOTSTRAP, Bootstrap Themes. https://themes.getbootstrap.com/, dernière consultation le 16 décembre 2018.
- [7] TEMPLATED, Templated. https://templated.co/, dernière consultation le 16 décembre 2018.
- [8] W3C, w3schools.com. https://www.w3schools.com/, dernière consultation le 16 décembre 2018.