



Etude sur l'eau potable

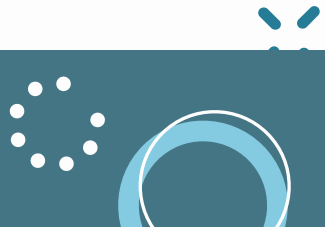
Présentée par Hayette Bouzouita,
Data Analyst en mission dans
l'ONG DWFA (Drinking Water For All)



Drink Water For All



- ONG ayant pour ambition de donner de l'**eau potable à tous** le monde.
- **3 domaines d'expertise :**
 1. Création de services d'accès à l'eau potable ;
 2. Modernisation de services d'accès à l'eau déjà existants ;
 3. Consulting auprès d'administrations/gouvernements à propos des politiques d'accès à l'eau.



Sommaire

1. Présentation du projet
2. Choix de l'outil de visualisation
3. Traitement des données
4. Blueprint et Mockup
5. Dashboard



Projet



- **Contexte** : DWFA a fait une demande de financement auprès d'un bailleur de fonds en présentant les 3 domaines d'expertise. Le bailleur choisira un domaine d'expertise. Le financement permettra d'investir dans le domaine d'expertise choisi dans un pays non déterminé pour l'instant.
- **But** : Réaliser un **tableau de bord** présentant une vue globale de l'accès à l'eau potable dans le monde pour aider à choisir **le pays à cibler** dès que le bailleur aura donné une réponse sur le domaine qu'il souhaite financer.



Pourquoi Tableau Public ?

The logo for Tableau Public, featuring the word "tableau" in blue with three small red and orange plus signs above the "u", followed by the word "public" in blue.

Outil de visualisation

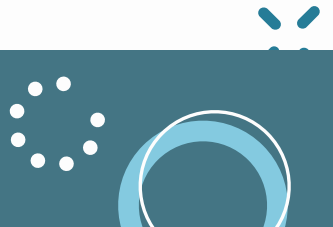
+ Points positifs :

- Recommandé par OC
- Logiciel appartenant à Tableau (il y a d'autres logiciels chez eux)
- Un des plus utilisés pour la visualisation de données
- Permet, entre autres, de créer et diffuser des tableaux de bord
- Disponible en app de bureau ou en app web
- Gratuit sans limite de durée

- Points négatifs : Les dashboard sont hébergés sur les serveurs de l'entreprise Tableau, donc accessible publiquement par n'importe qui -> Une alternative serait "Tableau Serveur"



Il existe d'autres outils de visualisation comme Power Bi, Knime...



Pipeline de Prétraitement des données

Fichiers CSV bruts

Nettoyage & transformation
(Python via JupyterNotebook)

Jointure, relations, champs calculés
(Tableau Public)

Tables	Variables	Type	Unité	Exemples de valeurs
Population	Country	Chaîne de caractères		Afghanistan
	Year	Entier		2000
	Granularity	Chaîne de caractères		Total
	Population	Décimal	1000 persons	20779.953
Source : http://www.fao.org/faostat/en/#data/PS Granularity : Total / Urban / Rural / Male / Female Year : 2000 > 2018				
BasicAndSafelyManagedDrinkingWaterServices	Country	Chaîne de caractères		Armenia
	Year	Entier		2000
	Granularity	Chaîne de caractères		Total
	Population using at least basic drinking-water services (%)	Décimal	%	95.13351
	Population using safely managed drinking-water services (%)	Décimal	%	29.59756
	Source : https://apps.who.int/gho/data/node.main.WSHWATER?lang=en Granularity : Total / Urban / Rural Year : 2000 > 2017			
MortalityRateAttributedToWater	Country	Chaîne de caractères		Angola
	Year	Entier		2016
	Granularity	Chaîne de caractères		Total
	Mortality rate attributed to exposure to unsafe WASH services	Décimal		48.83467
	WASH deaths	Décimal		14065.2
Source : https://apps.who.int/gho/data/view.main.SDGWSHB00392v?lang=en Granularity : Total / Male / Female Year : 2016				
PoliticalStability	Country	Chaîne de caractères		Albania
	Year	Entier		2000
	Granularity	Chaîne de caractères		Total
	Political_Stability	Décimal	index	-0.54
	Source : http://www.fao.org/faostat/en/#data/PS Granularity : Total Year : 2000 > 2018			
Region	Region	Chaîne de caractères		Europe
	Country	Chaîne de caractères		Albania
Region : ['Europe', 'Western Pacific', 'Eastern Mediterranean', 'South-East Asia', 'Africa', 'Americas']				

```
import pandas as pd

[8]: population = pd.read_csv('Population.csv')
      mortality = pd.read_csv('MortalityRateAttributedToWater.csv')
      political_stability = pd.read_csv('PoliticalStability.csv')
      water_service = pd.read_csv('BasicAndSafelyManagedDrinkingWaterServices.csv')
      region = pd.read_csv('RegionCountry.csv')
```

2. Nettoyage

Table population

```
[12]: population.head(5)
```

	Country	Granularity	Year	Population
0	Afghanistan	Total	2000	20779.953
1	Afghanistan	Male	2000	10689.508
2	Afghanistan	Female	2000	10090.449
3	Afghanistan	Rural	2000	15657.474
4	Afghanistan	Urban	2000	4436.282

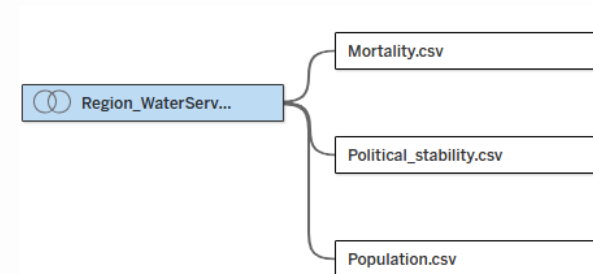


Tableau public n'offre que très peu de possibilités pour nettoyer ses données. Comme le nettoyage de données est un domaine à part entière, la société Tableau a donc développé un logiciel spécifique à la préparation des données : Tableau Prep. Il est très souvent utilisé en amont de la création d'un tableau de bord avec Tableau Public.



Tables	Variables	Type	Unité	Exemples de valeurs
Population	Country	Chaîne de caractères		Afghanistan
	Year	Entier		2000
	Granularity	Chaîne de caractères		Total
	Population	Décimal	1000 persons	20779.953
Source : http://www.fao.org/faostat/en/#data/FS Granularity : Total / Urban / Rural / Male / Female Year : 2000 > 2018				
BasicAndSafelyManagedDrinkingWaterServices	Country	Chaîne de caractères		Armenia
	Year	Entier		2000
	Granularity	Chaîne de caractères		Total
	Population using at least basic drinking-water services (%)	Décimal	%	95.13351
	Population using safely managed drinking-water services (%)	Décimal	%	29.59756
Source : https://apps.who.int/gho/data/node.main.WSHWATER?lang=en Granularity : Total / Urban / Rural Year : 2000 > 2017				
MortalityRateAttributedToWater	Country	Chaîne de caractères		Angola
	Year	Entier		2016
	Granularity	Chaîne de caractères		Total
	Mortality rate attributed to exposure to unsafe WASH services	Décimal		48.81467
	WASH deaths	Décimal		14065.2
Source : https://apps.who.int/gho/data/view.main.SDGWSHBOD392v?lang=en Granularity : Total / Male / Female Year : 2016				
PoliticalStability	Country	Chaîne de caractères		Albania
	Year	Entier		2000
	Granularity	Chaîne de caractères		Total
	Political_Stability	Décimal	index	-0.54
Source : http://www.fao.org/faostat/en/#data/FS Granularity : Total Year : 2000 > 2018				
Region	Region	Chaîne de caractères		Europe
	Country	Chaîne de caractères		Albania
Region : ['Europe', 'Western Pacific', 'Eastern Mediterranean', 'South-East Asia', 'Africa', 'Americas']				

Fichiers CSV bruts

Dictionnaire des données :
5 fichiers sources



```
import pandas as pd
```

```
[8]: population = pd.read_csv('Population.csv')
      mortality = pd.read_csv('MortalityRateAttributedToWater.csv')
      political_stability = pd.read_csv('PoliticalStability.csv')
      water_service = pd.read_csv('BasicAndSafelyManagedDrinkingWaterServices.csv')
      region = pd.read_csv('RegionCountry.csv')
```

2. Nettoyage

Table population

```
[12]: population.head(5)
```

```
[12]:
```

	Country	Granularity	Year	Population
0	Afghanistan	Total	2000	20779.953
1	Afghanistan	Male	2000	10689.508
2	Afghanistan	Female	2000	10090.449
3	Afghanistan	Rural	2000	15657.474
4	Afghanistan	Urban	2000	4436.282

Transformation de la table Water Service

Avant Pivot (format large)

Country	Year	Granularity	Taux Eau Basique (%)	Taux Eau Sûre (%)
A	2000	Total	70	40
B	2000	Total	55	20
C	2000	Total	80	65

Après Pivot (format long)

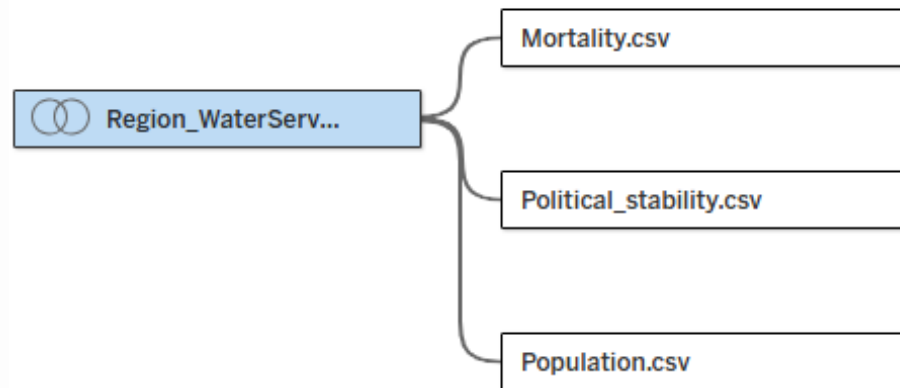
Country	Year	Granularity	Type Eau	Taux (%)
A	2000	Total	Basique	70
A	2000	Total	Sûre	40
B	2000	Total	Basique	55
B	2000	Total	Sûre	20
C	2000	Total	Basique	80
C	2000	Total	Sûre	65

Nettoyage & transformation (Python via Jupyter Notebook)

- ✓ **Nettoyage** : Supprimer colonnes inutiles, Comblér une année manquante...
- ✓ **Transformation** : Conversion de population en millions, pivot sur le type d'eau...
- ✓ **Vérification** : Nombre de population mondiale, correspondance noms des pays...



Region_WaterService.csv est composé de 2 tables. ⓘ



Jointure, relations, champs calculés
(Tableau Public)

✓ Jointure :

Region <-> WaterService

Inner Join via Country pour associer chaque pays à sa région.
Comme les pays sont identiques dans les 2 tables, il n'y a pas de perte de lignes. Résultat : 1 nouvelle table fusionnée.

✓ Relations :

R_WS <-> Mortality (Clé : Country)

R_WS <-> Political_Stability (Clé : Country + Year)

R_WS <-> Population (Clé : Country + Year + Granularity)

✓ Champs calculés :

Taux de croissance démographique, Efficacité politique, Score critique, ...



Jointure : fusion physique de 2 tables (obtention d'une table finale) => fusion FIXE des données

Relation : garde les tables séparées, éviter une "super table" pleine de valeurs nulles, ne fusionne que lors de la requête si besoin => fusion DYNAMIQUE



Blueprint

Blueprint - Etude d'accès à l'eau potable

Le tableau ci-dessous reprend les détails essentiels à suivre pour l'élaboration du dashboard.

Besoin utilisateurs	Question Type	Indicateurs /KPI	Données nécessaires (+ Agrégation)	Visualisation	Vue (+Sélecteur)
3 vues : Mondiale Continentale Nationale	Quel est le nombre de population d'un pays, d'un continent et dans le monde ?	ALL	Table Région Agrégation des indicateurs au <u>niv</u> mondial et continental	3 vues ayant des scopes différents Cohérence dans les données Palette Bleu en réf à l'eau	Création des 3 vues Sélecteurs pour choisir les continents et les pays
Nombre de population	Combien y a-t-il d'habitant (en zone totale, rurale et urbaine) en Europe ?	Nb population + évolution dans le temps	Table Population	Line plot <u>Stacked barplot</u> Chiffre clef en %	Vue Mondiale Continentale Nationale
Le nombre de mort à cause de l'eau insalubre	Combien y a-t-il de mort dû à l'eau insalubre dans le monde ?	Nb mort (à cause de l'eau)	Table <u>Mortality</u> (2016)	Chiffre Clef	Vue mondiale
Le taux de mortalité dû à l'eau insalubre	Quel est le taux de mortalité dû à l'eau insalubre en Angola ?	Taux de mortalité (à cause de l'eau insalubre)	Table <u>Mortality</u> (2016)	Carte Chiffre Clef	Vue Mondiale et Continentale Nationale
La part d'habitant ayant accès à l'eau potable	En Arménie, quelle est la part d'habitant qui a accès à l'eau potable ?	La part d'habitant ayant accès à l'eau potable (en %) + évolution dans le temps	Table <u>WaterServices</u>	Line plot <u>Grouped barplot</u> Chiffre Clef	Vue mondiale Continentale Nationale

La stabilité politique du pays	Comment est la stabilité politique en Albanie ?	Stabilité politique (indice) + évolution dans le temps	Table <u>PoliticalStability</u> (voir l'indice min et max, le neutre pour interpréter)	Line plot Diagramme en barre horizontale Chiffre clef	Vue Mondiale Continentale Nationale + Filtre
Indicateur pour domaine 1 : Création de service	Quelle est la corrélation entre l'urbanisation et l'accès à l'eau potable, en Arménie ?	Taux d'accès à l'eau potable (total) Taux de population urbaine	Table <u>WaterServices</u> Table population	Graphique : Nuage de points	Vue nationale Accès au domaine 1 via sélecteur sur les 3 domaines
Indicateur pour domaine 2 : Modernisation des services	Comment est la qualité des services en Arménie (zone totale, urbaine et rurale) ?	Taux de services (d'infrastructures) "basiques" Taux d'infrastructures de qualité	Table <u>WaterServices</u>	Graphique : Nuage de points	Vue nationale Accès au domaine 2 via sélecteur sur les 3 domaines
Indicateur pour domaine 3 : Consulting	Comment les conditions politiques en Albanie influencent-elles la mise en œuvre des politiques d'accès à l'eau potable ?	Efficacité de la politique gouvernementale d'accès à l'eau Stabilité politique	Table <u>Mortality</u> + Table <u>WaterServices</u> <i>politique efficace = Taux de mortalité faible + bon accès des habitants aux services d'eau potable</i> Table <u>PoliticalStability</u>	Graphique : Nuage de points	Vue nationale Accès au domaine 3 via sélecteur sur les 3 domaines (Filtre pour exclure les pays trop instable ?)



Blueprint = "Cahier des charges", fil conducteur, Traduire les besoins en une liste d'indicateurs quantifiables



Mockup

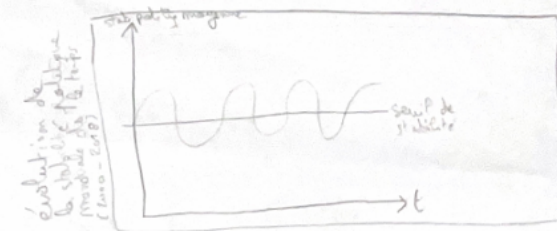
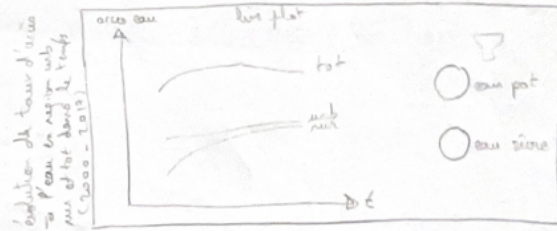
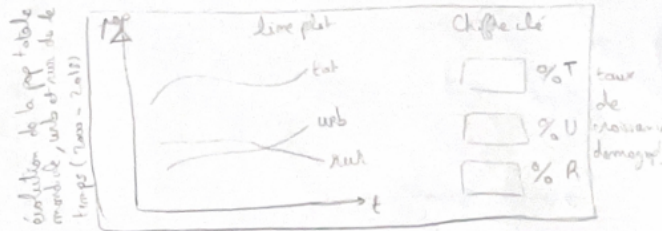
VOE MONDIALE : Vue d'ensemble des enjeux mondiaux liés à l'eau insalubre et à l'accès à l'eau potable

Filter sur l'année

NB total de décès dû à l'eau insalubre (2016)
chiffre clé

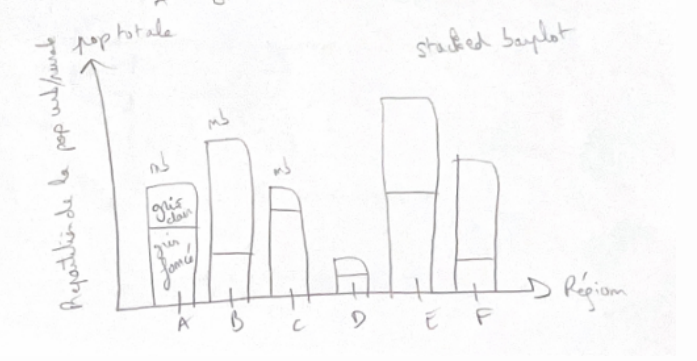
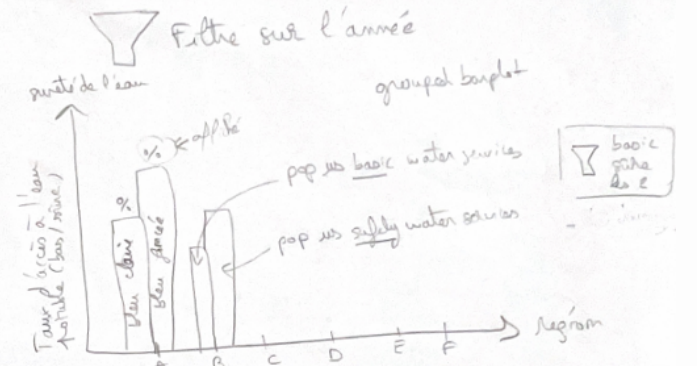
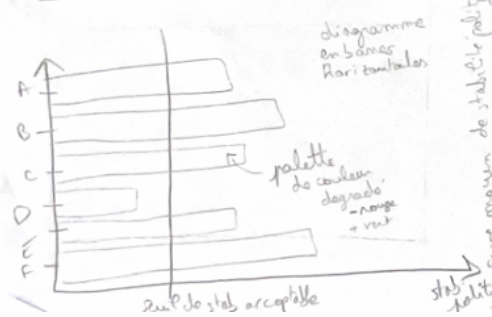
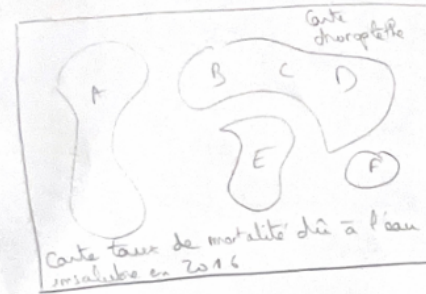


- Palette de couleurs rouge/bleu (situations critiques et non critiques)
- Accès à la vue continentale ou nationale
- Filter "clé sur un continent" filtrant les autres graph



Vue Continentale : Comparaison des différents continents pour identifier les zones les plus critiques et prioriser les interventions.

Filter sur la région

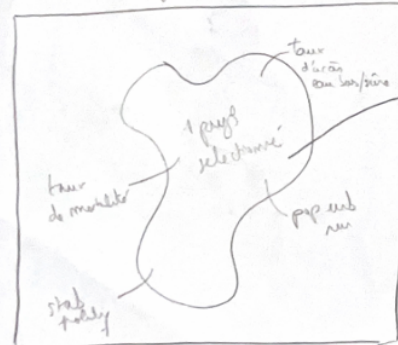
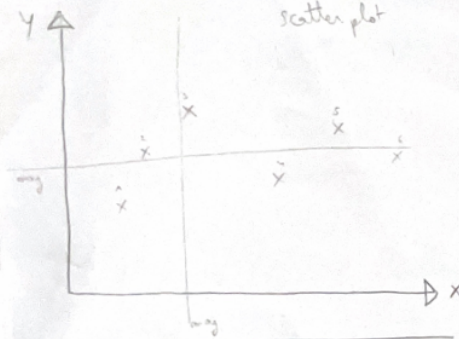


Mockup = "brouillon du dashboard", pas forcément représenter 100% du dashboard, il n'a pas besoin d'être exhaustif. Faire apparaître les visualisations, sélecteurs, indicateurs, filtres.



Une matricielle: fournir des données spécifiques à un pays afin de mieux comprendre les besoins et les opportunités d'interventions

Domaine 1, 2, 3



Pays	X	Y	stabilité politique	Taux accès eau et

possibilité de cliquer sur le pays pour le sélectionner

Tableau dynamique des pays filter

Norm du pays sélectionné	
bp tot, urb, rur	
Nb décès	
Taux mortalité	
Taux accès eau (bars et sure)	
stab politique	

info en clicke clié via info bulle



Dashboard / Lien vers l'histoire Tableau

https://public.tableau.com/views/HB_17523196534350/Histoire1?:language=fr-FR&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

