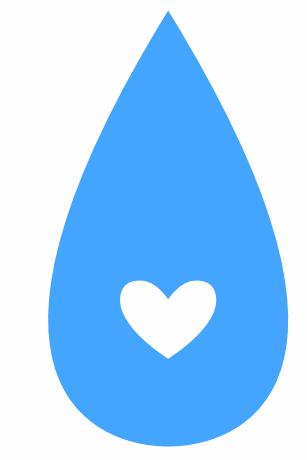




PROJET 8 : FAITES UNE ÉTUDE SUR L'EAU POTABLE



Réalisé par Yu Ting HUANG

OPENCLASSROOMS



En tant que Data Analyst dans l'ONG DWFA, notre mission consiste à réaliser un dashboard présentant une vue globale de l'accès à l'eau potable dans le monde.

Cela doit permettre d'identifier les pays qui rencontrent des difficultés d'accès à l'eau potable.

C'est parmi ceux-ci que seront ciblés les pays à aider en fonction du domaine d'expertise qui aura été financé.

Les indicateurs:

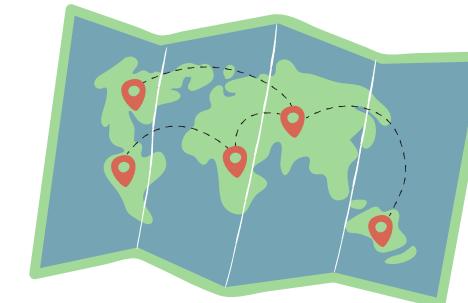
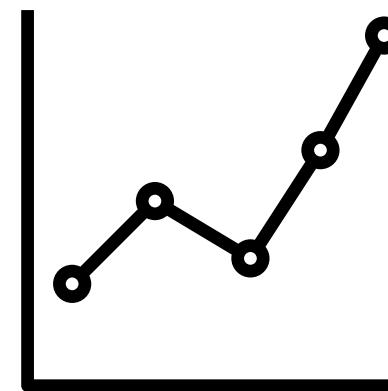
- Population**
- Taux d'urbanisation**
- Taux de mortalité due à l'eau insalubre**
- Stabilité politique**





BLUEPRINT

Vue mondiale

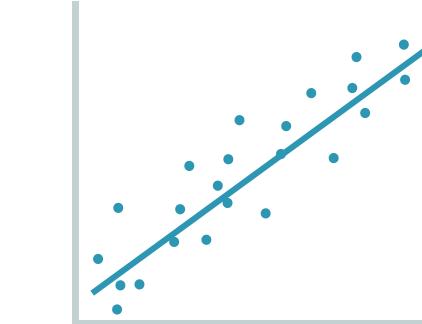
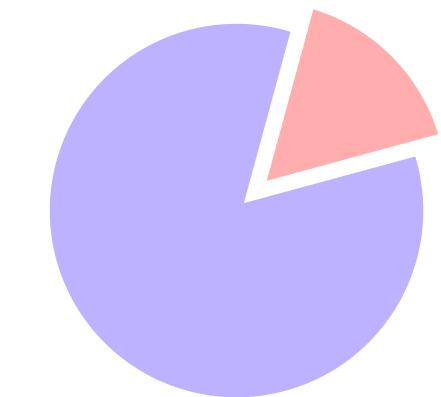
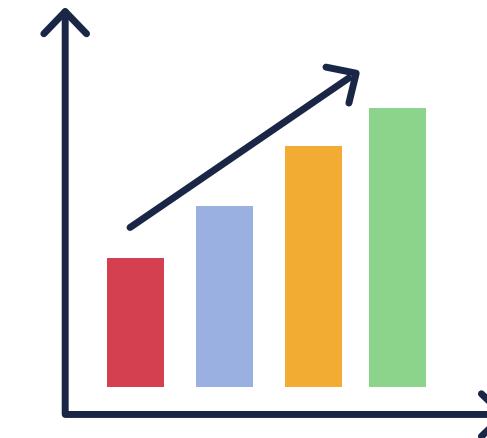
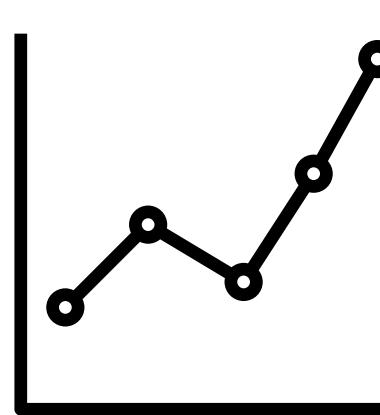


Evolution en 3 axes:

- **population,**
- **stabilité politique,**
- **accès à l'eau potable basique
(rural/ urbain/total)**

Besoin utilisateurs	Mesures spécifiques à utiliser	Visualisation	Page/Onglet/Vue*
Voir l'évolution de la population rurale vs la population urbaine et population totale	Population rurale, urbaine et population totale	Line plot	Vue mondiale
Comprendre la stabilité politique dans le monde	Agrégation stabilité politique par continent (par la moyenne)	Line plot Barplot	Vue mondiale/ Vue continentale
Voir l'évolution de la part d'habitants ruraux ayant accès à l'eau potable vs la population urbaine et totale.	Agrégation « population utilisant au moins des services d'eau potable de base » en multipliant l'agréation population de chaque pays et diviser par la population mondiale	Line plot	Vue mondiale

Vue régionale / continentale



Besoin utilisateurs	Mesures spécifiques à utiliser	Visualisation	Page/Onglet/Vue*
Voir la répartition de population rurale par continent vs la population urbaine	Population rurale et population urbaine	Stacked barplot	Vue continentale
Comprendre la stabilité politique dans le monde	Agrégation stabilité politique par continent (par la moyenne)	Line plot Barplot	Vue mondiale/ Vue continentale
Voir la part d'habitants ayant accès à l'eau potable par continent	Agrégation « population utilisant au moins des services d'eau potable de base » en multipliant l'agrégation population de chaque pays et diviser par la population mondiale (par continent)	Barplot	Vue continentale

Vue d'ensemble des régions:

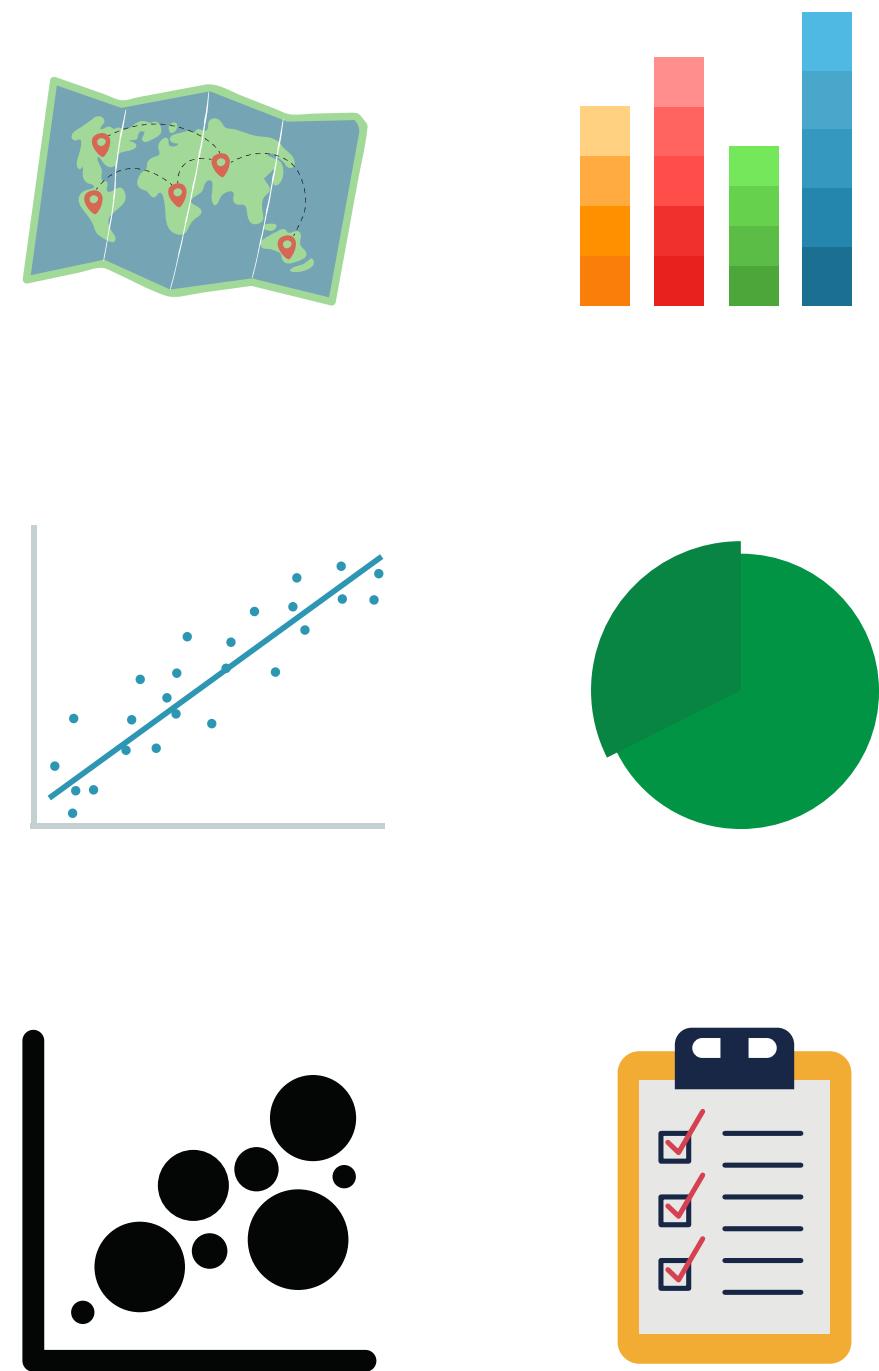
- **Répartition rural/urbain**
- **Stabilité politique,**
- **Accès à l'eau potable basique / de qualité**
- **Taux de mortalité**

Voir l'accès à l'eau potable selon type de services (basique / qualité) par continent	Agrégation des valeurs de l'accès à l'eau potable basique /qualité par continent (par la moyenne)	Barplot	Vue continentale
Le taux de mortalité par continent à cause de l'eau insalubre	Agrégation du taux de mortalité WASH en multipliant l'agrégation population de chaque	Bar plot	Vue continentale

Vue nationale

Indicateurs pour les 3 domaines d'expertise:

- **Creation de services**
 - **Nuage de points**
- **Modernisation de services**
 - **Nuage de points**
 - **Taux services basique**
 - **Taux services de qualité**
- **Consulting:**
 - **Bubble chart**
 - **taux de mortalité**
 - **taux d'accès à l'eau potable**
 - **stabilité politique**



Besoin utilisateurs	Mesures spécifiques à utiliser	Visualisation	Page/Onglet/Vue*
Voir la répartition de population rurale par part vs la population urbaine	Population rurale et population urbaine	Pie chart	Vue nationale / choix du pays
Comprendre la stabilité politique dans le monde	Agrégation stabilité politique par pays (par la moyenne)	Carte du monde	Vue mondiale / Vue nationale
Le taux de mortalité par sexe par pays à cause de l'eau insalubre	Agrégation taux de mortalité WASH (par la moyenne)	Bar plot	Vue nationale / choix du pays
Indicateur : Domaine 1 (création de services)	Le taux d'accès à l'eau potable et le taux de la population urbaine.	Scatter plot	Vue nationale / choix du pays
Indicateur : Domaine 2 (Modernisation des services)	Le taux de services basiques et le taux de services de qualité	Scatter plot	Vue nationale / choix du pays
Indicateur : Domaine 3 (Consulting)	Politique efficace (le taux de mortalité + le taux de services de qualité + la stabilité politique)	bubble chart (pour 3 variables)	Vue nationale/ choix du pays
Voir les chiffres clés de pays qui sont sélectionnés par l'utilisateur	Population, taux de mortalité, nombre de morts, accès basique, accès de qualité, stabilité politique et taux urbain.population	Tableau	Vue nationale / choix du pays



LE PRÉTRAITEMENT DES DONNÉES

> 46 territoires problématiques sont identifiés

*SUPPRESSION DE 42 TERRITOIRES CAR:

- APPARTENANCE A AUTRE PAYS (EX: GUADELOUPE (FRANCE), PUERTO RICO (USA) ..)
- ABSENCE DE DONNÉES POUR ANALYSE
- TERRITOIRE EN CONFLIT, MANQUE DE DONNÉES (WESTERN SAHARA)

```
#拿掉france屬地
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "French Guyana")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "French Polynesia")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Saint-Martin (French part)")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Guadeloupe")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Martinique")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Channel Islands")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Mayotte")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "New Caledonia")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Saint Barthélemy")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Saint Pierre and Miquelon")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Wallis and Futuna Islands")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Réunion")],axis=0,inplace=True)

✓ 0.1s
```

```
#刪除有錢國家
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Liechtenstein")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Netherlands Antilles (former)")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Sint Maarten (Dutch part)")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Sudan (former)")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Tokelau")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Western Sahara")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Serbia and Montenegro")],axis=0,inplace=True)

✓ 0.5s
```

```
# 拿掉英國 britsh屬地
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Anguilla")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Bermuda")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "British Virgin Islands")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Cayman Islands")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Falkland Islands (Malvinas)")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Gibraltar")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Isle of Man")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Montserrat")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Saint Helena, Ascension and Tristan da Cunha")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "Turks and Caicos Islands")],axis=0,inplace=True)

✓ 0.6s
```

* AJOUT DE PAYS (DANS LE FICHIER "PAYS" REGION)

- GREENLAND, MACEDOINE (RENOMMÉ "REPUBLIC OF MACEDONIA"), TAIWAN, PALESTINE

```
# Corrige le nom de TW
df_region_popu.replace('China, Taiwan Province of','Taiwan',inplace=True)
# Ajoute la region de Taiwan dans la colonne de Region
mask= df_region_popu['Country'] == 'Taiwan'
df_region_popu.iloc[(mask).values,0] = 'Western Pacific'
✓ 0.5s
```

```
# 然後把palestine加上region
mask= df_region_popu['Country'] == 'Palestine'
df_region_popu.iloc[(mask).values,0] = 'Eastern Mediterranean'
✓ 0.2s
```

```
#幫greenland加上region
mask= df_region_popu['Country'] == 'Greenland'
df_region_popu.iloc[(mask).values,0] = 'Americas'
✓ 0.3s
```

* RENOMMAGE DE CERTAINS PAYS QUI ONT PLUSIEURS NOMS

- EX: NORTH MACEDONIA, MACEDONIA, REPUBLIC OF MACEDONIA

```
# Corrige le nom de North macedoine
df_region_popu.replace('North Macedonia','Republic of North Macedonia',inplace=True)
# Ajoute la region de North macedoine
mask= df_region_popu['Country'] == 'Republic of North Macedonia'
df_region_popu.iloc[(mask).values,0] = 'Europe'
✓ 0.3s
```



* CAS PARTICULIER DE LA CHINE:

- SUPPRESSION DE HONKONGSAR , MACAU SAR , CHINA (MAINLAND)
- REGROUPEMENT SOUS LE NOM DE CHINA

```
# Enleve HK,MC,CH
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "China, Macao SAR")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "China, Hong Kong SAR")],axis=0,inplace=True)
df_region_popu.drop(df_region_popu.index[(df_region_popu["Country"] == "China, mainland")],axis=0,inplace=True)
```

* CORRECTIONS DE REGIONS DE CERTAINS PAYS:

- ISRAEL ETAIT EN "EUROPE" AU LIEU DE "EASTERN MEDITERRANEAN"
- NORTH KOREA
- EN SOUTH-EAST ASIA AU LIEU DE WESTERN PACIFIC

```
#發現有錯誤,所以先把israel改回eastern·mediterranean
mask= df_region_popu['Country'] == 'Israel'
df_region_popu.iloc[(mask).values,0] = 'Eastern·Mediterranean'
df_region_popu[(mask)].head(2)
' 0.4s
```

```
mask= Region['Country'] == "Democratic People's Republic of Korea"
Region.iloc[(mask).values,0] = 'Western Pacific'
Region[(mask)]
0.4s
```

*AJOUT DE LA COLONNE TAUX D'URBANISATION

- POUR UTILISATION DANS TABLEAU

```
urban=Population['Granularité'] == 'Urban'  
df_urban=Population[urban]  
df_urban  
✓ 0.9s
```

	Pays	Granularité	Année	Population
4	Albania	Urban	2000	1303137.0
9	Albania	Urban	2001	1324985.0
14	Albania	Urban	2002	1356800.0

```
def mult_1000(row):  
    m1 = (df_total['Pays'] == row['Pays'])  
    m2 = (df_total['Année'] == row['Année'])  
    tmp = df_total.iloc[(m1 & m2).values, 3].values[0] # 3 is col index for Population  
    row['TotalPopu'] = tmp  
  
    m3 = (df_urban['Pays'] == row['Pays'])  
    m4 = (df_urban['Année'] == row['Année'])  
    tmp2 = df_urban.iloc[(m3 & m4).values, 3].values[0] # 3 is col index for Population  
    row['TotalUrban'] = tmp2  
  
    row['Taux Urbain Popu.'] = row['TotalUrban'] / row['TotalPopu']  
    return row  
  
tmp3 = Population.apply(mult_1000, axis=1)
```

✓ 28.1s

total=Population['Granularité'] == 'Total'
df_total =Population[total]
df_total
✓ 0.5s

	Pays	Granularité	Année	Population
0	Albania	Total	2000	3129243.0
5	Albania	Total	2001	3129704.0
10	Albania	Total	2002	3126187.0
15	Albania	Total	2003	3118023.0
20	Albania	Total	2004	3104892.0

```
Population1 = tmp3  
Population1  
✓ 0.1s
```

	Pays	Granularité	Année	Population	TotalPopu	TotalUrban	Taux Urbain Popu.
0	Albania	Total	2000	3129243.0	3129243.0	1303137.0	0.416438
1	Albania	Male	2000	1583506.0	3129243.0	1303137.0	0.416438
2	Albania	Female	2000	1545740.0	3129243.0	1303137.0	0.416438
3	Albania	Rural	2000	1818833.0	3129243.0	1303137.0	0.416438
4	Albania	Urban	2000	1303137.0	3129243.0	1303137.0	0.416438
...
20115	Palestine	Total	2018	4862979.0	4862979.0	3848420.0	0.416438
20116	Palestine	Male	2018	2466617.0	4862979.0	3848420.0	0.416438
20117	Palestine	Female	2018	2396361.0	4862979.0	3848420.0	0.416438
20118	Palestine	Rural	2018	1204356.0	4862979.0	3848420.0	0.416438
20119	Palestine	Urban	2018	3848420.0	4862979.0	3848420.0	0.416438

*SUPPRESSION DE 6 TERRITOIRES (LE FICHIER POLITIQUE) :

IL S'AGIT ICI DES ÉTATS TERRITOIRES RETIRÉS DU FICHIER POPULATION.

```
df_nan = df_rp_politique.drop_duplicates(subset=['Country'])  
df_nan  
#df_nan.shape  
✓ 0.2s
```

	Region	Country	Granularity_x	Year_x	Population	Year_y	Political_Stability_x	Granularity_y	Year	Political_Stability_y	Granularity
5621536	Nan	American Samoa	Nan	Nan	Nan	2004.0	0.74	Total	2004.0	0.74	Total
5621761	Nan	Bermuda	Nan	Nan	Nan	2000.0	0.71	Total	2000.0	0.71	Total
5622085	Nan	China, Hong Kong SAR	Nan	Nan	Nan	2000.0	0.93	Total	2000.0	0.93	Total
5622409	Nan	China, Macao SAR	Nan	Nan	Nan	2000.0	0.49	Total	2000.0	0.49	Total
5622733	Nan	China, mainland	Nan	Nan	Nan	2000.0	-0.21	Total	2000.0	-0.21	Total
5623057	Nan	China, Taiwan Province of	Nan	Nan	Nan	2000.0	0.54	Total	2000.0	0.54	Total
5623381	Nan	North Macedonia	Nan	Nan	Nan	2000.0	-0.62	Total	2000.0	-0.62	Total
5623705	Nan	Puerto Rico	Nan	Nan	Nan	2000.0	0.26	Total	2000.0	0.26	Total



POURQUOI TABLEAU ?

1

UN OUTILS
PUISSANT QUI
PERMET DE
TRANSFORMER
RAPIDEMENT LES
DONNÉES EN
INFORMATIONS
EXPLOITABLES

2

FAIRE DES
GRAPHIQUES
FACILEMENT AVEC
GLISSEZ COLLER
(EN GÉNÉRAL)

3

RESULTAT
PROFESSIONNEL

