

# Apresentação

TEMA:

**Compreendendo as consequências sanitárias dos índices sociais**

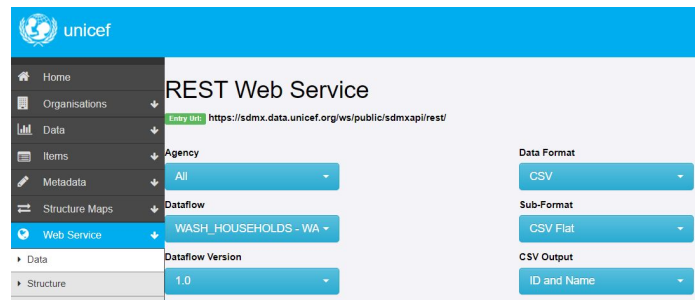
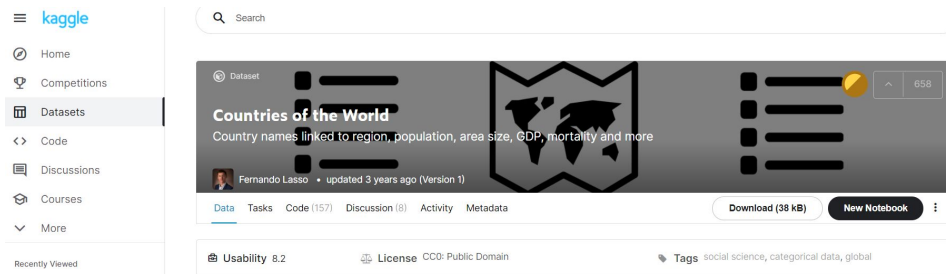
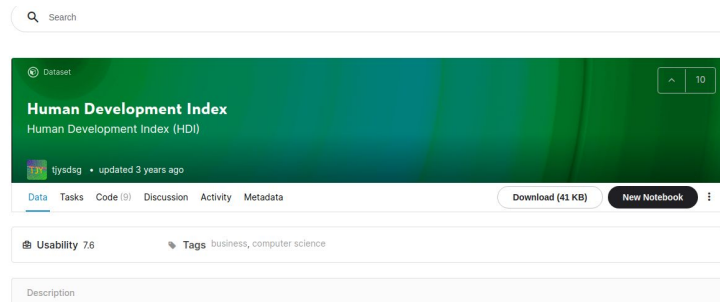
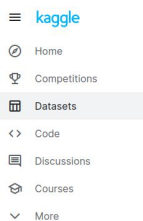
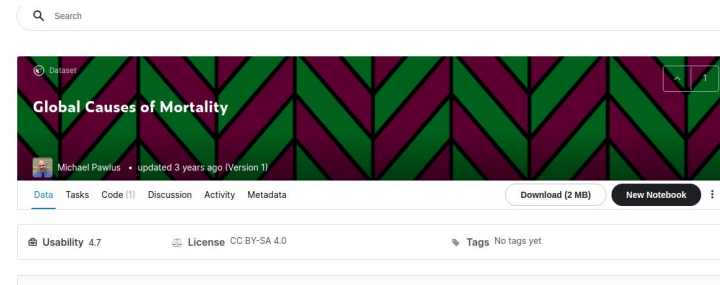
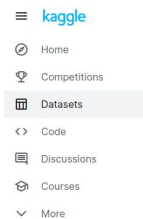
# Datasets seleccionados

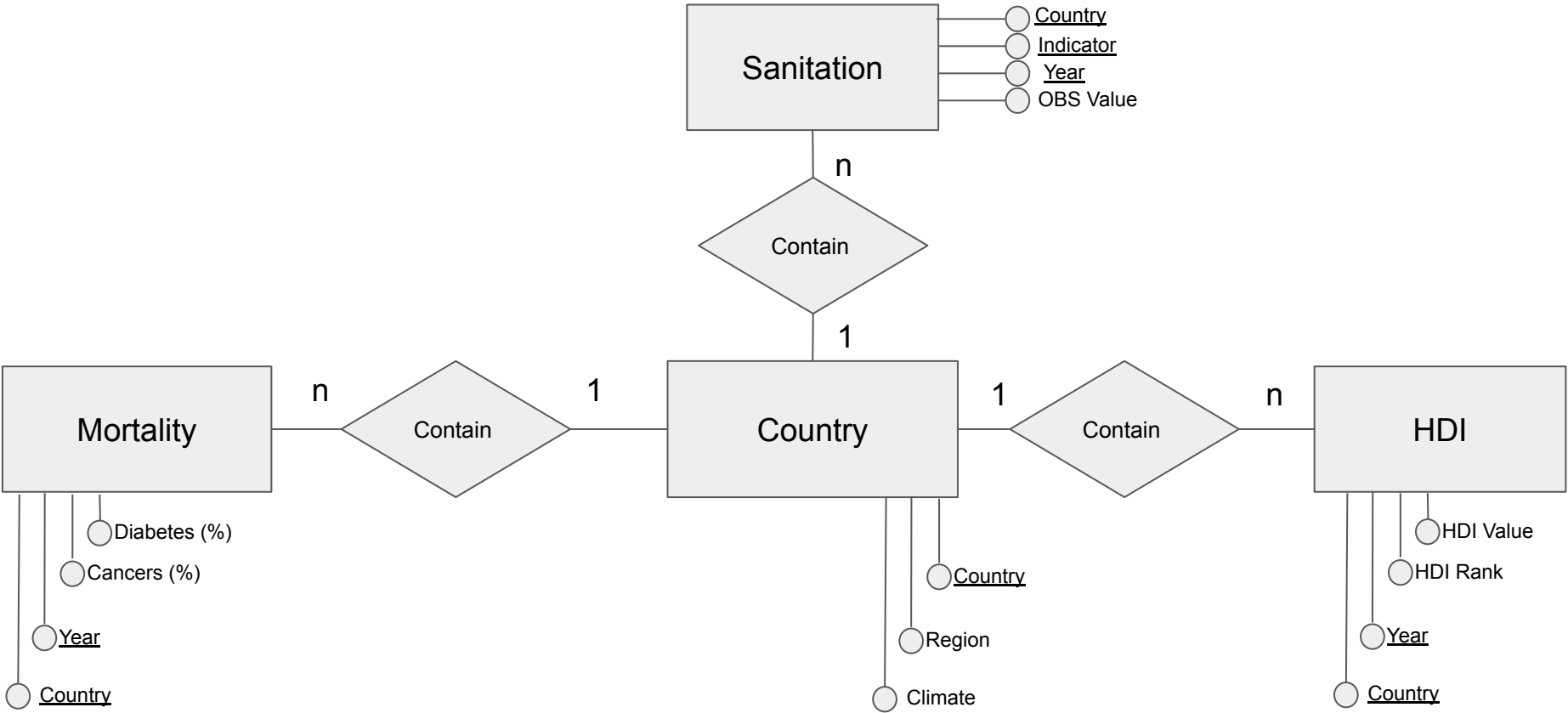
<https://www.kaggle.com/michaelpawlus/global-causes-of-mortality>

<https://www.kaggle.com/tjysdsg/human-development-index>

<https://sdmx.data.unicef.org/web-service/data.html>

<https://www.kaggle.com/fernandol/countries-of-the-world>





# Operações nos datasets

- Requisição por API;
- Alteração no formato do dataframe;
- Remoção de campos nulos;
- Uniformização do período de tempo;
- Criação das chaves estrangeiras;

HDI Rank	Country	1990	1991	1992	1993	1994	1995
168	Afghanistan						
68	Albania	0.65	0.63	0.61	0.61	0.62	0.63
85	Algeria	0.58	0.58	0.59	0.59	0.6	0.6
35	Andorra						
147	Angola						
70	Antigua and Bar						
47	Argentina	0.7	0.71	0.72	0.73	0.73	0.73
83	Armenia	0.63	0.63	0.58	0.59	0.6	0.61
3	Australia	0.87	0.87	0.87	0.87	0.88	0.88
20	Austria	0.8	0.8	0.81	0.81	0.81	0.82

## Country(Country, Region, Climate)

Country	Region	Climate
Afghanistan	ASIA (EX. NEAR E	1
Albania	EASTERN EUROP	3
Algeria	NORTHERN AFRICA	1
American Samoa	OCEANIA	2
Andorra	WESTERN EUROPE	3
Angola	SUB-SAHARAN AFRICA	
Anguilla	LATIN AMER. & CARIBBEAN	2
Antigua & Barbuda	LATIN AMER. & CARIBBEAN	2
Argentina	LATIN AMER. & CARIBBEAN	3
Armenia	C.W. OF IND. STATES	4
Aruba	LATIN AMER. & CARIBBEAN	2
Australia	OCEANIA	1
Austria	WESTERN EUROPE	3
Azerbaijan	C.W. OF IND. STATES	1
Bahamas, The	LATIN AMER. & CARIBBEAN	2
Bahrain	NEAR EAST	1
Bangladesh	ASIA (EX. NEAR EAST)	2
Barbados	LATIN AMER. & CARIBBEAN	2

*HDI*(country, year, HDI\_VALUE)

- CHE: country → Country País

HDIRank ↑	country ↑	year ↑	HDI_VALUE ↑
68	Albania	1990	0.65
85	Algeria	1990	0.58
47	Argentina	1990	0.7
83	Armenia	1990	0.63
3	Australia	1990	0.87
20	Austria	1990	0.8
43	Bahrain	1990	0.75
136	Bangladesh	1990	0.39
58	Barbados	1990	0.72
17	Belgium	1990	0.81
106	Belize	1990	0.64
163	Benin	1990	0.35
101	Botswana	1990	0.58
79	Brazil	1990	0.61
51	Bulgaria	1990	0.69
185	Burundi	1990	0.3
146	Cambodia	1990	0.36
151	Cameroon	1990	0.44
12	Canada	1990	0.85
44	Chile	1990	0.7
86	China	1990	0.5

country ▲ ▼	country_code	year ▼	Cardiovascular_diseases ▼	Cancers ▼	Respiratory_diseases ▼	Diabetes ▼
Afghanistan	AFG	1990	17.61	4.03	2.11	3.83
Afghanistan	AFG	1991	17.8	4.05	2.13	3.82
Afghanistan	AFG	1992	18.39	4.17	2.21	3.9
Afghanistan	AFG	1993	18.96	4.27	2.28	3.97
Afghanistan	AFG	1994	19.09	4.26	2.31	3.97
Afghanistan	AFG	1995	19.86	4.39	2.41	4.11
Afghanistan	AFG	1996	20.41	4.48	2.49	4.21
Afghanistan	AFG	1997	20.36	4.45	2.49	4.19
Afghanistan	AFG	1998	19.62	4.26	2.41	4.04
Afghanistan	AFG	1999	21.11	4.57	2.6	4.34
Afghanistan	AFG	2000	21.43	4.64	2.64	4.41
Afghanistan	AFG	2001	21.64	4.7	2.67	4.5
Afghanistan	AFG	2002	22.26	4.86	2.74	4.67
Afghanistan	AFG	2003	22.81	5.09	2.84	4.98

*Mortality*(**country**,**year**,country\_code, Cardiovascular diseases (%), Cancers (%), Respiratory diseases (%), Diabetes (%), Dementia (%), Lower respiratory infections (%), Neonatal deaths (%), Diarrheal diseases (%), Road accidents (%), Liver disease (%), Tuberculosis (%), Kidney disease (%), Digestive diseases (%), HIV/AIDS (%), Suicide (%), Malaria (%), Homicide (%), Nutritional deficiencies (%), Meningitis (%), Protein-energy malnutrition (%), Drowning (%), Maternal deaths (%), Parkinson disease (%), Alcohol disorders (%), Intestinal infectious diseases (%), Drug disorders (%), Hepatitis (%), Fire (%), Heat-related (hot and cold exposure) (%), Natural disasters (%), Conflict (%), Terrorism (%))

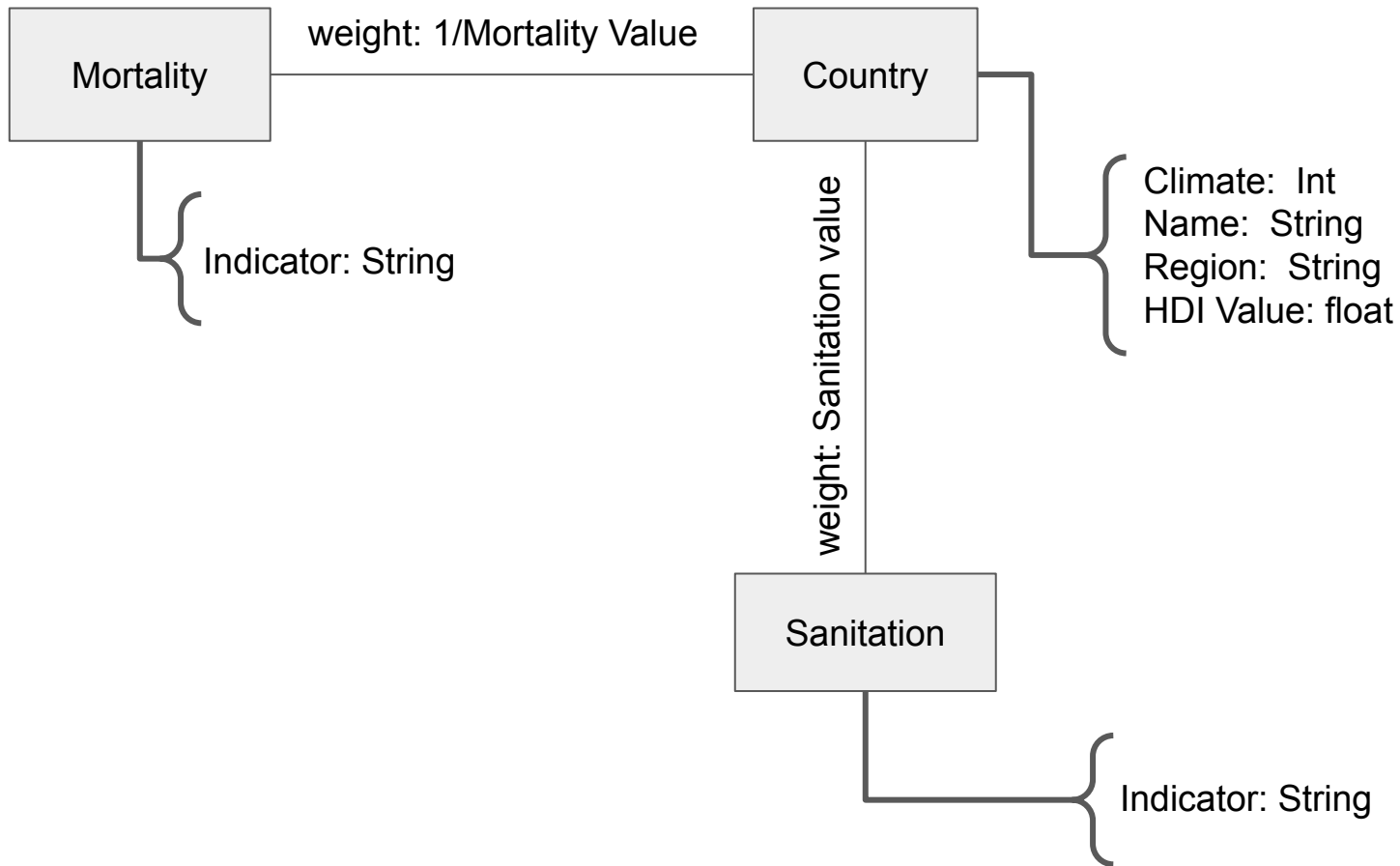
- CHE: country -> Country País

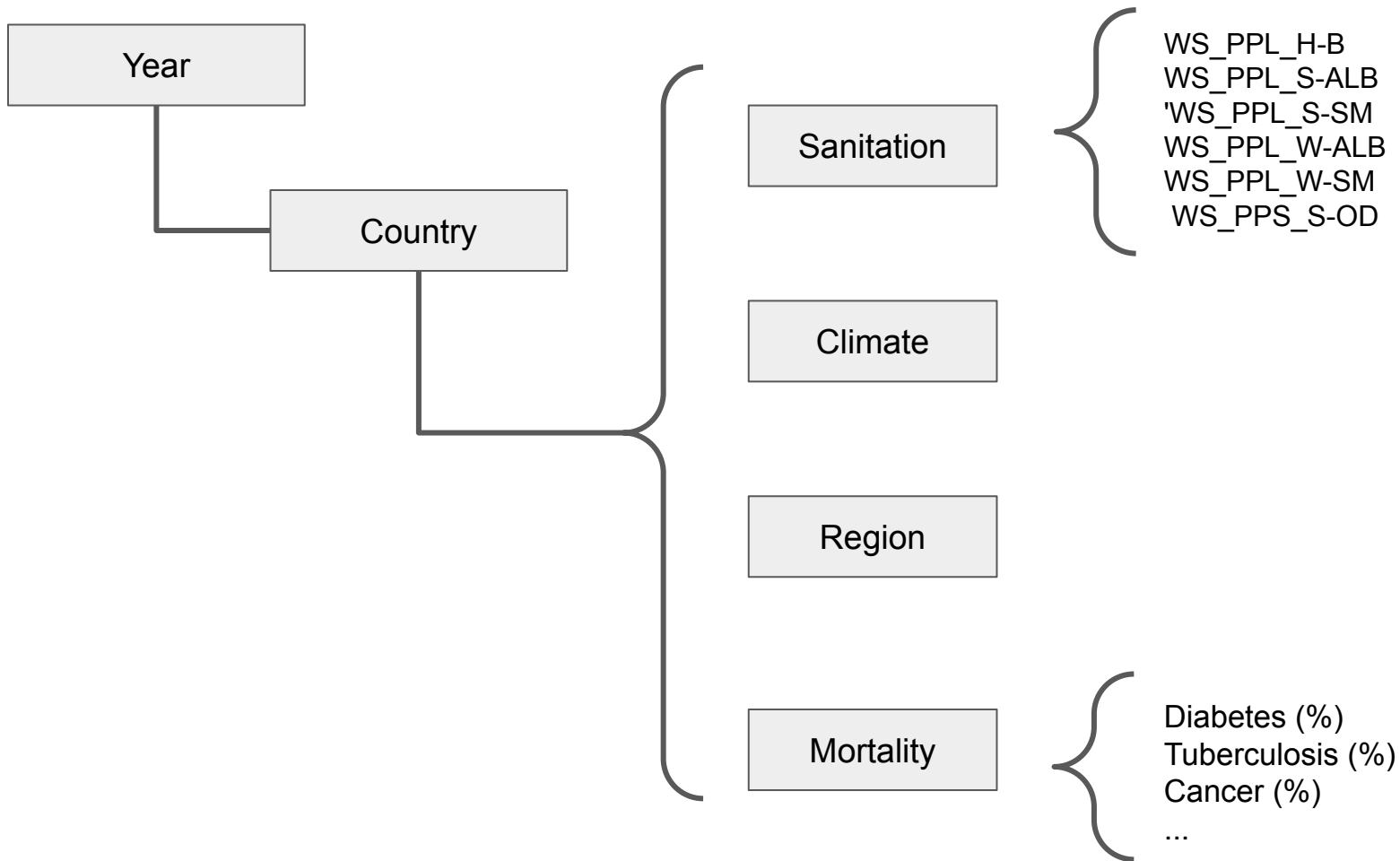
*Sanitation*(REF\_AREA, **country**, INDICATOR, Indicator, Service Type, year Unit of measure, OBS\_VALUE)

- CHE: country -> Country País

REF_AREA	country	INDICATOR	Indicator	Service Type	year	Unit Of Measure	OBS_VALUE
AFG	Afghanistan	WS_PPL_H-B	Proportion of po	Hygiene	2005	%	36.97
AFG	Afghanistan	WS_PPL_H-B	Proportion of po	Hygiene	2006	%	37.04
AFG	Afghanistan	WS_PPL_H-B	Proportion of po	Hygiene	2007	%	37.11
AFG	Afghanistan	WS_PPL_H-B	Proportion of po	Hygiene	2008	%	37.18
AFG	Afghanistan	WS_PPL_H-B	Proportion of po	Hygiene	2009	%	37.25
AFG	Afghanistan	WS_PPL_H-B	Proportion of po	Hygiene	2010	%	37.32
AFG	Afghanistan	WS_PPL_H-B	Proportion of po	Hygiene	2011	%	37.4
AFG	Afghanistan	WS_PPL_H-B	Proportion of po	Hygiene	2012	%	37.47
AFG	Afghanistan	WS_PPL_H-B	Proportion of po	Hygiene	2013	%	37.54
AFG	Afghanistan	WS_PPL_H-B	Proportion of po	Hygiene	2014	%	37.62
AFG	Afghanistan	WS_PPL_H-B	Proportion of po	Hygiene	2015	%	37.69
AFG	Afghanistan	WS_PPL_H-B	Proportion of po	Hygiene	2016	%	37.77



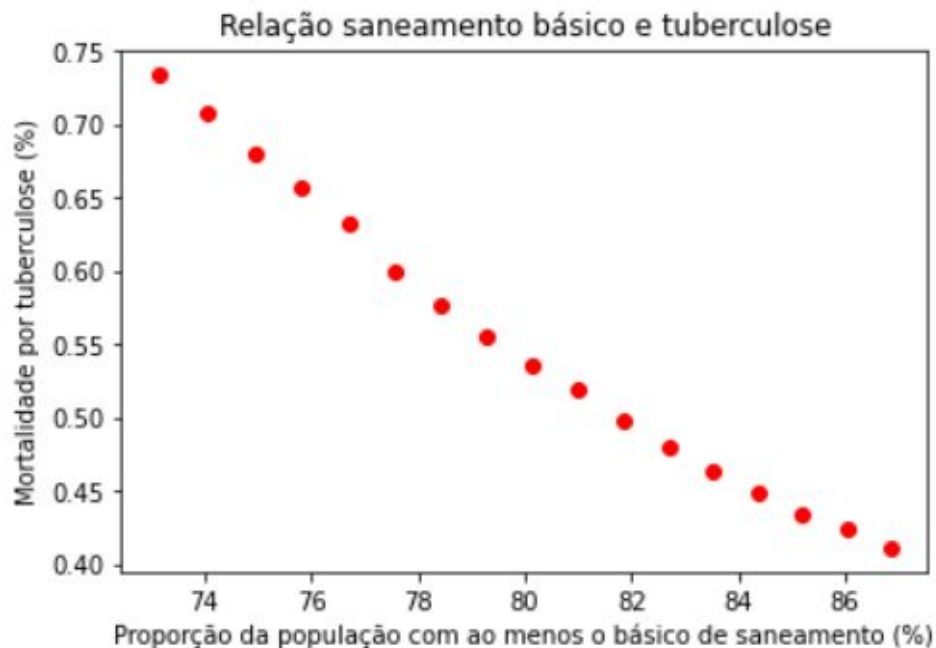




No recorte Brasil, é possível observar uma variação na mortalidade por tuberculose considerando a variação na proporção de pessoas com acesso a saneamento básico?

```
SELECT Mortality.country, Mortality.year, Mortality.Tuberculosis, S.OBS_VALUE FROM Mortality, Sanitation S
WHERE Mortality.COUNTRY = 'Brazil' AND Mortality.year = S.year AND S.country = 'Brazil' AND S.INDICATOR = 'WS_PPL_S-ALB'
```

▲index	COUNTRY	YEAR	TUBERCULOSIS	OBS_VALUE
0	Brazil	2000	0.734414167497	73.128799
1	Brazil	2001	0.707039713601	74.037903
2	Brazil	2002	0.679596443843	74.9279790000
3	Brazil	2003	0.657221048391	75.812637
4	Brazil	2004	0.632360159549	76.691536
5	Brazil	2005	0.599185072744	77.564789
6	Brazil	2006	0.576569943864	78.432907
7	Brazil	2007	0.555266378806	79.295975
8	Brazil	2008	0.535525311371	80.154083
9	Brazil	2009	0.518831636095	81.006523
10	Brazil	2010	0.498235926443	81.854202
11	Brazil	2011	0.479311192946	82.7005839999
12	Brazil	2012	0.462910759821	83.542229
13	Brazil	2013	0.448489565065	84.378517
14	Brazil	2014	0.435053705663	85.21064
15	Brazil	2015	0.424342422579	86.037964
16	Brazil	2016	0.410627158929	86.8602449999



Qual país adquiriu maior melhora relativa no seu IDH, ao analisar o período de 1990 a 2016?

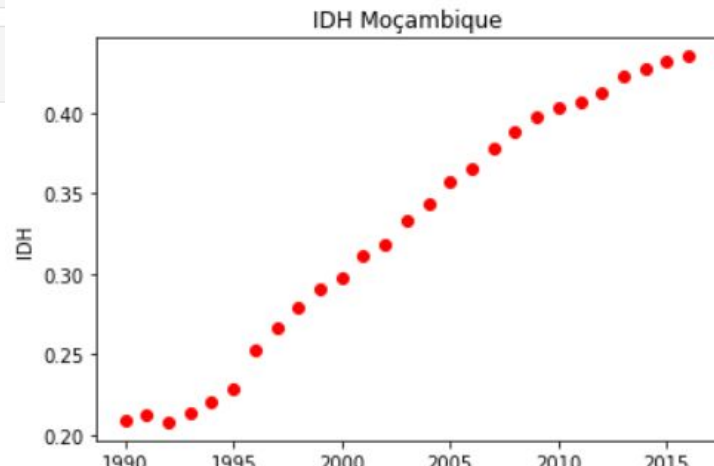
```
[26]: CREATE VIEW IDH2016 AS  
      SELECT country, year, HDI_VALUE FROM IDH  
      WHERE year = 2016
```

```
[27]: CREATE VIEW IDH1990 AS  
      SELECT country, year, HDI_VALUE FROM IDH  
      WHERE year = 1990
```

```
[28]: CREATE VIEW variacaoIDH AS  
      SELECT IDH2016.country, IDH2016.HDI_VALUE/IDH1990.HDI_VALUE as variacao FROM IDH2016, IDH1990  
      WHERE IDH2016.country = IDH1990.country
```

```
[29]: SELECT country, variacao FROM variacaoIDH  
      WHERE variacao = (SELECT MAX(variacao) FROM variacaoIDH)
```

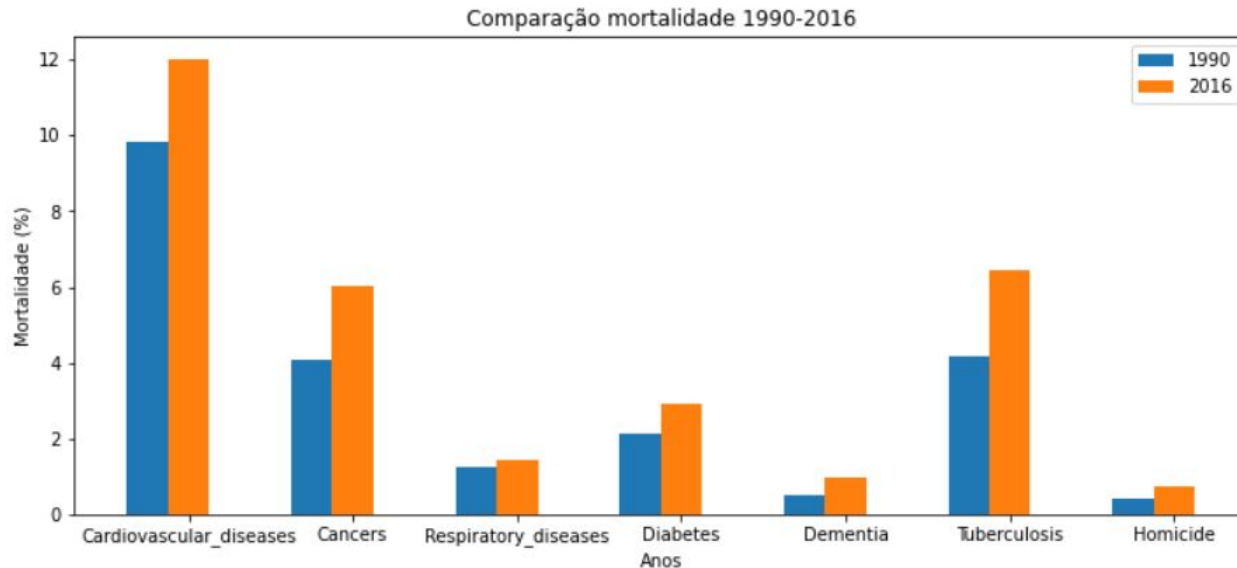
index	Key	Value
0	COUNTRY	Mozambique
1	VARIACAO	2.0813397129186604



## Qual o impacto no perfil de mortalidade desse país?

```
SELECT * FROM Mortality
WHERE country = 'Mozambique' AND (year = 1990 OR year = 2016)
```

index	COUNTRY	YEAR	TUBERCULOSIS	SUICIDE	DRUG_DISORDERS	CARDIOVASCULAR_DISEASES	CANCERS	RESPIRATORY_DISEASES	DIABETES
0	Mozambique	1990	4.18142016473	0.470153516193	0.0104040812326	9.82256395035	4.05454359206	1.23457069429	2.14094540122999
1	Mozambique	2016	6.44495114798	0.712077433621	0.021332767209	12.0064291034	6.0076731574	1.42683901687	2.91392113291999



Os países que apresentam os melhores valores de IDH também apresentam bons resultados de mortalidade por suicídio?

```
CREATE VIEW TOPIDH AS
  SELECT country, HDI_value FROM IDH
  WHERE year = 2016
  ORDER BY HDI_value desc
```

```
CREATE VIEW TOP20IDH AS
  SELECT top 20 * FROM TOPIDH
```

```
CREATE VIEW TOPIDHSUICIDE AS
  SELECT TOP20IDH.country, TOP20IDH.HDI_VALUE, M.Suicide FROM TOP20IDH, Mortality M
  WHERE TOP20IDH.country = M.country AND M.year = 2016
```

```
SELECT * FROM TOPIDHSUICIDE
```

Index	COUNTRY	HDI_VALUE	SUICIDE
0	Norway	0.951	1.41073156923
1	Switzerland	0.943	1.94299444911
2	Australia	0.938	1.70649288109
3	Germany	0.934	1.3076615286
4	Ireland	0.934	1.55120696759
5	Iceland	0.9329999999	1.84655139845
6	Sweden	0.932	1.51507437238
7	Singapore	0.93	2.05320326094
8	Denmark	0.9279999999	1.38250103649
9	Netherlands	0.9279999999	1.32617716312
10	Canada	0.922	1.67447173525
11	United States	0.922	1.681086748499
12	United Kingdom	0.92	0.947717670613
13	Finland	0.9179999999	1.751568858959
14	Belgium	0.915	1.99478307501
15	New Zealand	0.915	1.64632186446
16	Japan	0.907	2.1606983055
17	Austria	0.9059999999	1.77664812507
18	Luxembourg	0.904	1.65795190111

média de suicídio: 1.389

```
SELECT AVG(Suicide) FROM Mortality
WHERE year = 2016
```

1.3891157886018581

index	COUNTRY	YEAR	DRUG_DISORDERS	HDI_VALUE
0	Brazil	1990	0.0540147688616999	0.611
1	Brazil	1991	0.056973800628	0.615
2	Brazil	1992	0.0584368645607	0.622
3	Brazil	1993	0.0606730365852	0.63
4	Brazil	1994	0.0631076924469	0.64
5	Brazil	1995	0.0665801786959	0.648
6	Brazil	1996	0.0741369585648999	0.6559999999
7	Brazil	1997	0.0799974460038	0.664
8	Brazil	1998	0.0867016933315	0.67
9	Brazil	1999	0.0884403513283	0.6759999999
10	Brazil	2000	0.0914565189416	0.684
11	Brazil	2001	0.0933113351112999	0.691
12	Brazil	2002	0.0942144840213	0.698
13	Brazil	2003	0.0943179338117	0.695
14	Brazil	2004	0.0939113375087	0.698
15	Brazil	2005	0.0945105477643	0.7
16	Brazil	2006	0.0945174737654	0.7020000000
17	Brazil	2007	0.0967569376343	0.705
18	Brazil	2008	0.098973502574	0.716
19	Brazil	2009	0.103198793446	0.718
20	Brazil	2010	0.107692130427	0.727
21	Brazil	2011	0.112770686101999	0.731
22	Brazil	2012	0.118187770215	0.736
23	Brazil	2013	0.122541226442	0.748
24	Brazil	2014	0.125885195041	0.752
25	Brazil	2015	0.127735910097	0.757
26	Brazil	2016	0.128573553494	0.758

Qual a evolução das mortes devido a problemas com drogas no Brasil? Há alguma relação com o IDH?

```
SELECT M.country, M.year, M.Drug_disorders, IDH.HDI_VALUE FROM Mortality M, IDH
WHERE M.country = 'Brazil' AND IDH.country = 'Brazil' AND IDH.year = M.year
```



Quais regiões apresentaram aumento nas taxas de suicídio dentre o período de 1990 a 2016?

```
[118]: CREATE VIEW DADOSSUICIDIO AS
        SELECT C.Region, M.year, AVG(M.Suicide) SuicidioMedio FROM Mortality M, Countries C
        WHERE M.country = C.country
        GROUP BY (C.Region, M.year)
        ORDER BY C.Region
```

```
[119]: SELECT * FROM DADOSSUICIDIO
```

```
SELECT D1.Region FROM DADOSSUICIDIO D1
WHERE D1.year = 2016 AND D1.SUICIDIOMEDIO > (SELECT D2.SUICIDIOMEDIO FROM DADOSSUICIDIO D2 WHERE D2.year = 1990 AND D1.Region = D2.Region)
```

index	REGION
0	C.W. OF IND. STATES
1	LATIN AMER. & CARIB
2	NEAR EAST
3	NORTHERN AFRICA
4	SUB-SAHARAN AFRICA

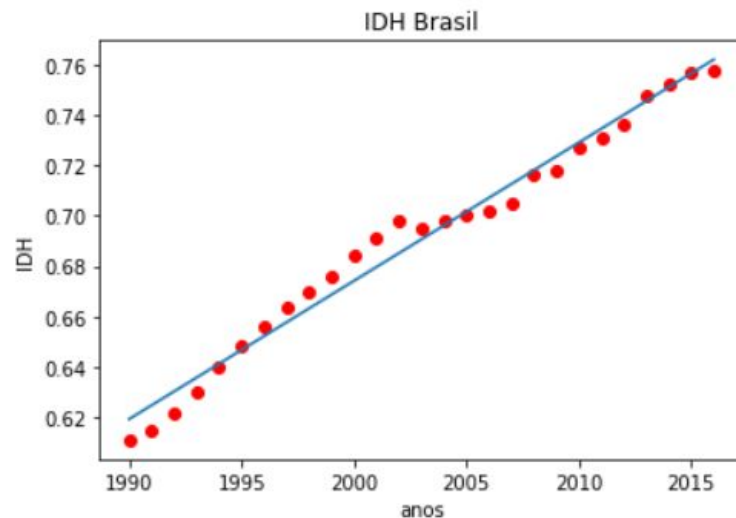
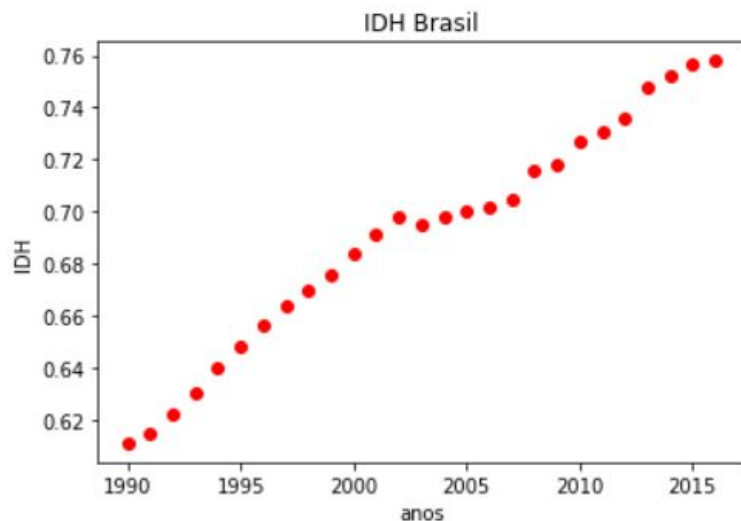
É possível determinar uma relação linear entre IDH e dados de saneamento básico/higiene com os índices de mortalidade por tuberculose? Se sim, qual desses indicadores apresenta maior peso?

$X = [x_1, x_2, x_3] = [\text{IDH}, \text{proporção de população com ao menos o básico de saneamento}, \text{proporção da população com ao menos o básico de água tratada}]$

$Y = \text{Mortalidade por tuberculose}$

```
Coeficientes: [-4.43060801 -2.30905226 -1.97700818]  
Intersecção: 8.172652185629348  
Variância: 0.586365783950338
```

# Previsão de dados de IDH de Brasil



	year	IDH_real	IDH_previsto	erro_relativo
0	2017	0.761	0.768	0.009198
1	2018	0.762	0.773	0.014436
2	2019	0.765	0.784	0.024837

A partir dos dados de saneamento básico, é possível agrupar os países em clusters baseado nas semelhanças de suas estatísticas, em determinado ano?

cluster 2:

Proportion of population using at least basic sanitation services: 36.485626

Proportion of population using at least basic drinking water services: 63.718285

Proportion of population practising open defecation 14.294533

cluster 3:

Proportion of population using at least basic sanitation services: 97.724113

Proportion of population using at least basic drinking water services: 99.187973

Proportion of population practising open defecation 0.0

cluster 0:

Proportion of population using at least basic sanitation services: 75.90220249999999

Proportion of population using at least basic drinking water services: 92.4654195

Proportion of population practising open defecation 5.62377785

cluster 1:

Proportion of population using at least basic sanitation services: 26.4891825

Proportion of population using at least basic drinking water services: 68.68840800000001

Proportion of population practising open defecation 43.966120000000004