

Integrating Health Data for Sustainable Development

Data Integration Project
Results



Huthayfa Mutan



Experiments. Cleaning:

Multiple Vague locations such as unions, continents, summed up results for the globe. Vague Indicator Names such as ppp which was solved manually.

Missing data. Mismatching names because of different spelling.

50250	1101040	-	-	000	158	South Asia	S5	GBD Super Regions
47038	1570819	20438	34401	763	158	South Asia	S5	GBD Super Regions
53171	2045207	21974	36986	888	158	South Asia	S5	GBD Super Regions
52883	2459340	2270	3821	999	158	South Asia	S5	GBD Super Regions
35846	2741656	-	-	1105	158	South Asia	S5	GBD Super Regions
31772	2846612	-	-	1200	158	South Asia	S5	GBD Super Regions
56391	2963221	-	-	1302	4	Southeast Asia, East Asia, and Oceania	S6	GBD Super Regions
93015	2996460	-	-	1319	4	Southeast Asia, East Asia, and Oceania	S6	GBD Super Regions
33032	2962077	-	-	1309	4	Southeast Asia, East Asia, and Oceania	S6	GBD Super Regions
30664	2904780	-	-	1306	4	Southeast Asia, East Asia, and Oceania	S6	GBD Super Regions
16480	2785060	-	-	1307	4	Southeast Asia, East Asia, and Oceania	S6	GBD Super Regions
16859	2571004	-	-	1305	4	Southeast Asia, East Asia, and Oceania	S6	GBD Super Regions
35652	2540716	-	-	1375	4	Southeast Asia, East Asia, and Oceania	S6	GBD Super Regions
52006	2487947	-	-	1362	4	Southeast Asia, East Asia, and Oceania	S6	GBD Super Regions
48333	2499033	-	-	1362				
90610	2619061	-	-	1414				
30301	2781090	-	-	1482				
28166	531020	567	856	1120				



Before:

location_id	location_name	year_id	age_group_id	age_group_name	haq_index_age_type	indicator_id	indicator_name	meas
6	China	1990	27	Age-standardized	Overall	100	HAQ Index	In
6	China	1990	27	Age-standardized	Overall	297	Tuberculosis	Mortality incide ratios (f
6	China	1990	27	Age-standardized	Overall	302	Diarrheal diseases	Mortality incide ratios (f
6	China	1990	27	Age-standardized	Overall	322	Lower respiratory infections	standar death r (f
6	China	1990	27	Age-standardized	Overall	328	Upper respiratory infections	standar death r (f

```
'ghes_total_upper', 'ghes_total_ppp_mean', 'ghes_total_ppp_lower',  
'ghes_total_ppp_upper', 'ppp_total_mean', 'ppp_total_lower',  
'ppp_total_upper', 'ppp_total_ppp_mean', 'ppp_total_ppp_lower',  
'ppp_total_ppp_upper', 'oop_total_mean', 'oop_total_lower',  
'oop_total_upper', 'oop_total_ppp_mean', 'oop_total_ppp_lower',  
'oop_total_ppp_upper', 'dah_total_mean', 'dah_total_ppp_mean',  
'the_per_cap_mean', 'the_per_cap_lower', 'the_per_cap_upper',  
'the_per_cap_ppp_mean', 'the_per_cap_ppp_lower',  
'the_per_cap_ppp_upper', 'ghes_per_cap_mean', 'ghes_per_cap_lower',  
'ghes_per_cap_upper', 'ghes_per_cap_ppp_mean', 'ghes_per_cap_ppp_lower',  
'ghes_per_cap_ppp_upper', 'ppp_per_cap_mean', 'ppp_per_cap_lower',  
'ppp_per_cap_upper', 'ppp_per_cap_ppp_mean', 'ppp_per_cap_ppp_lower',  
'ppp_per_cap_ppp_upper', 'oop_per_cap_mean', 'oop_per_cap_lower',  
'oop_per_cap_upper', 'oop_per_cap_ppp_mean', 'oop_per_cap_ppp_lower',  
'oop_per_cap_ppp_upper', 'dah_per_cap_mean', 'dah_per_cap_ppp_mean',  
'ghes_per_the_mean', 'ghes_per_the_lower', 'ghes_per_the_upper',  
'ppp_per_the_mean', 'ppp_per_the_lower', 'ppp_per_the_upper',  
'oop_per_the_mean', 'oop_per_the_lower', 'oop_per_the_upper',  
'dah_per_the_mean', 'dah_per_the_lower', 'dah_per_the_upper',  
'the_per_gdp_mean', 'the_per_gdp_lower', 'the_per_gdp_upper',
```



After:

```
mysql> select recording_date, value, country.name, indicator.name from records join country on c_id= country.id join indicator on ind_id=indicator.id order by rand() limit 10;
```

recording_date	value	name	name
2000-01-01	323507	Poland	Prepaid Private Health Spending purchasing power parity
1990-01-01	0.5350983738899231	Puerto Rico	Colon and rectum cancer
2012-01-01	0	Norway	Development Assistance for Health
2019-01-01	0.0000019300800886412617	Iceland	Diarrheal diseases
2016-01-01	5853	Iceland	Total Health Spending per person
1990-01-01	0.000013302799743541982	Romania	Gallbladder and biliary diseases
2001-01-01	503826	Namibia	Prepaid Private Health Spending purchasing power parity
1990-01-01	0.3856184780597687	Zimbabwe	Testicular cancer
2002-01-01	841	Republic of Korea	Total Health Spending per person
2017-01-01	3410167	Venezuela (Bolivarian Republic of)	Out-of-pocket Health Spending

```
10 rows in set (0.11 sec)
```



Results (Show cases).

What countries have Total Health Spending Lower than 500 millions and HAQ Index HIGHER than 60.

```
mysql> select * from indicator where name='Total Health Spending' OR name='HAQ Index';
```

id	name
13	HAQ Index
34	Total Health Spending

```
mysql> select country.name from records join country on c_id =country.id where ind_id=13 and recording_date >='2019-01-01' AND value>=60.0  
intersect  
select country.name from records join country on c_id =country.id where ind_id=34 and recording_date >='2019-01-01' AND value<=50  
0000;
```

name
Maldives
Andorra
Cook Islands
Greenland
Monaco
San Marino

6 rows in set (0.03 sec)



Results (Show cases).

What is the Development Assistance for Health per person compared with Stroke Risk-standardised death rates (RSD) for germany for the year 2019.

```
mysql> select recording_date, value, country.name, indicator.name from records join country on c_id= country.id join indicator on ind_id=indicator.id where country.name='Germany' and (indicator.name='Stroke' or indicator.name='Total Health Spending per person') and recording_date>='2019_01_01';
```

recording_date	value	name	name
2019-01-01	0.00017133599612861872	Germany	Stroke
2019-01-01	5887	Germany	Total Health Spending per person

2 rows in set, 1 warning (0.01 sec)