Relationship Between Sanitation Access and Child Mortality

Data Analysis Report

2025-05-18

Data Analysis

Let's examine the data coverage by year for each dataset to identify the most recent year with good coverage (less than 30% missing values).

Missing Values in Sanitation Dataset (%)

year	total_countries	missing_values	pct_missing
2022	215	24	11.162791
2021	215	19	8.837209
2020	215	16	7.441860
2019	215	13	6.046512
2018	215	11	5.116279
2017	215	4	1.860465

Missing Values in Child Deaths Dataset (%)

year	total_countries	missing_values	pct_missing
2015	211	0	0
2014	211	0	0
2013	211	0	0
2012	211	0	0
2011	211	0	0
2010	211	0	0

```
## Most recent year with good coverage for sanitation data: 2022
```

```
## Most recent year with good coverage for child deaths data: 2015
```
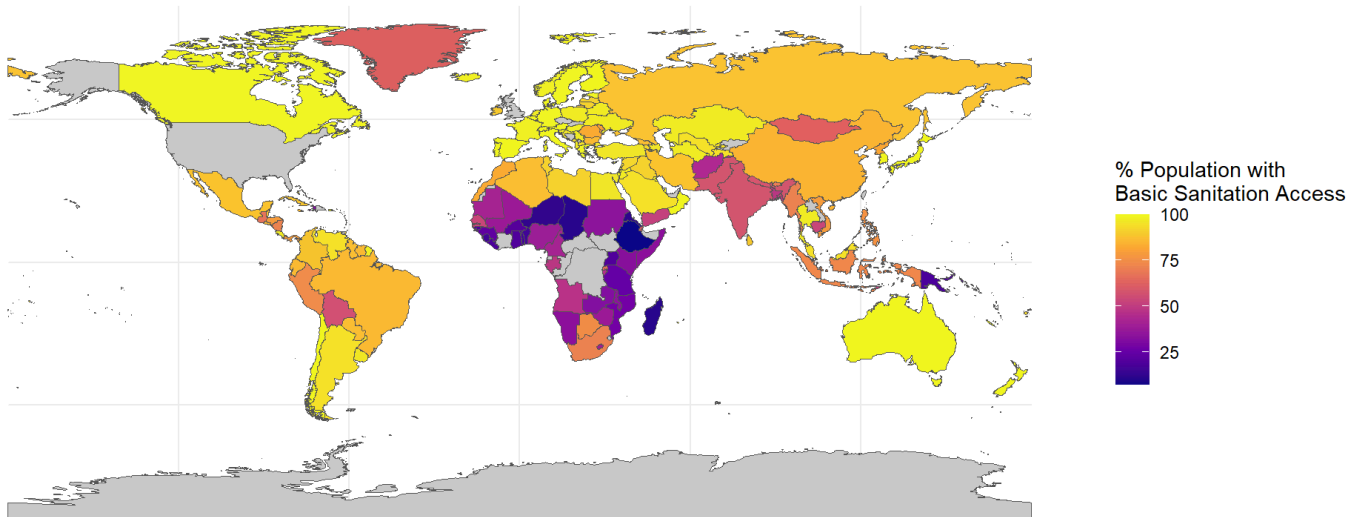
```
## Selected year for analysis: 2015
```

Data Preparation for Mapping

Now we'll prepare the data for the selected year and join it with world map data.

Choropleth Map for Sanitation Access

Access to Basic Sanitation by Country in 2015

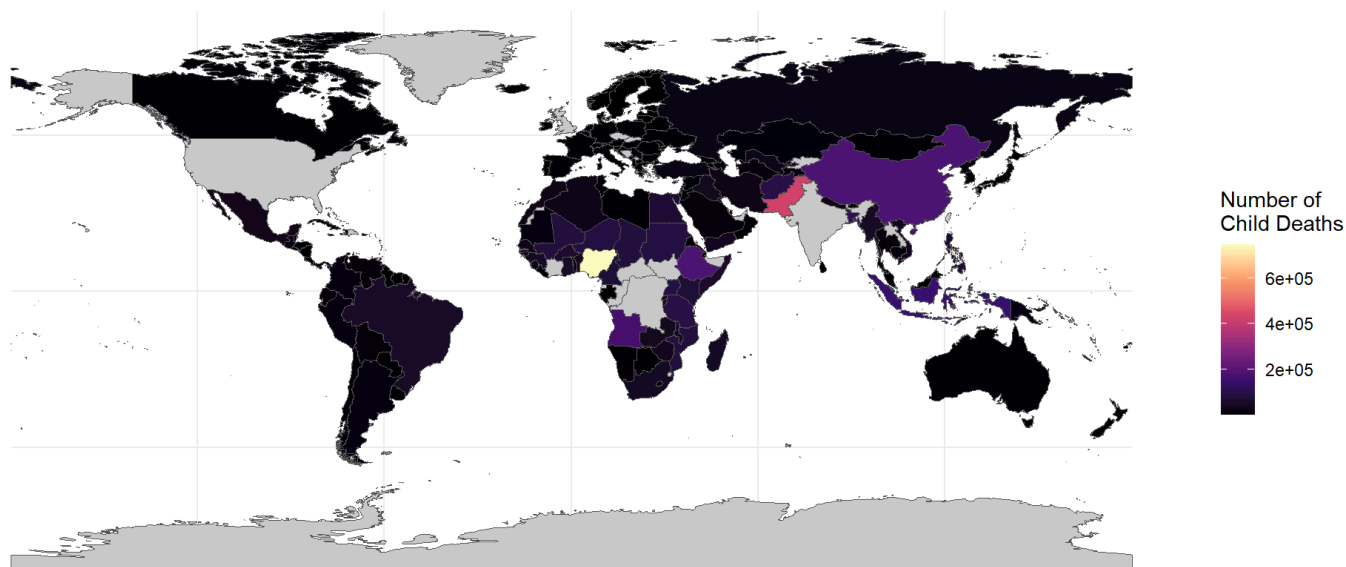Percentage of population with access to basic sanitation facilities

Data source: at_least_basic_sanitation_overall_access_percent.csv

Choropleth Map for Child Mortality

Number of Child Deaths by Country in 2015

Total number of deaths among children

Data source: number_of_child_deaths.csv

Relationship Between Sanitation Access and Child Mortality

Let's merge the data and examine the relationship between these two indicators.

Merged Data for 2015

country	sanitation_access	child_deaths
Aruba	98.6	17
Afghanistan	44.0	94300
Angola	47.4	169000
Albania	97.7	618
Andorra	100.0	2
UAE	99.1	661

Now, let's create a scatterplot to visualize the relationship between sanitation access and child deaths.

Relationship Between Sanitation Access and Child Deaths in 2015

Pearson correlation: r = -0.359

Data sources: at_least_basic_sanitation_overall_access_percent.csv and number_of_child_deaths.csv
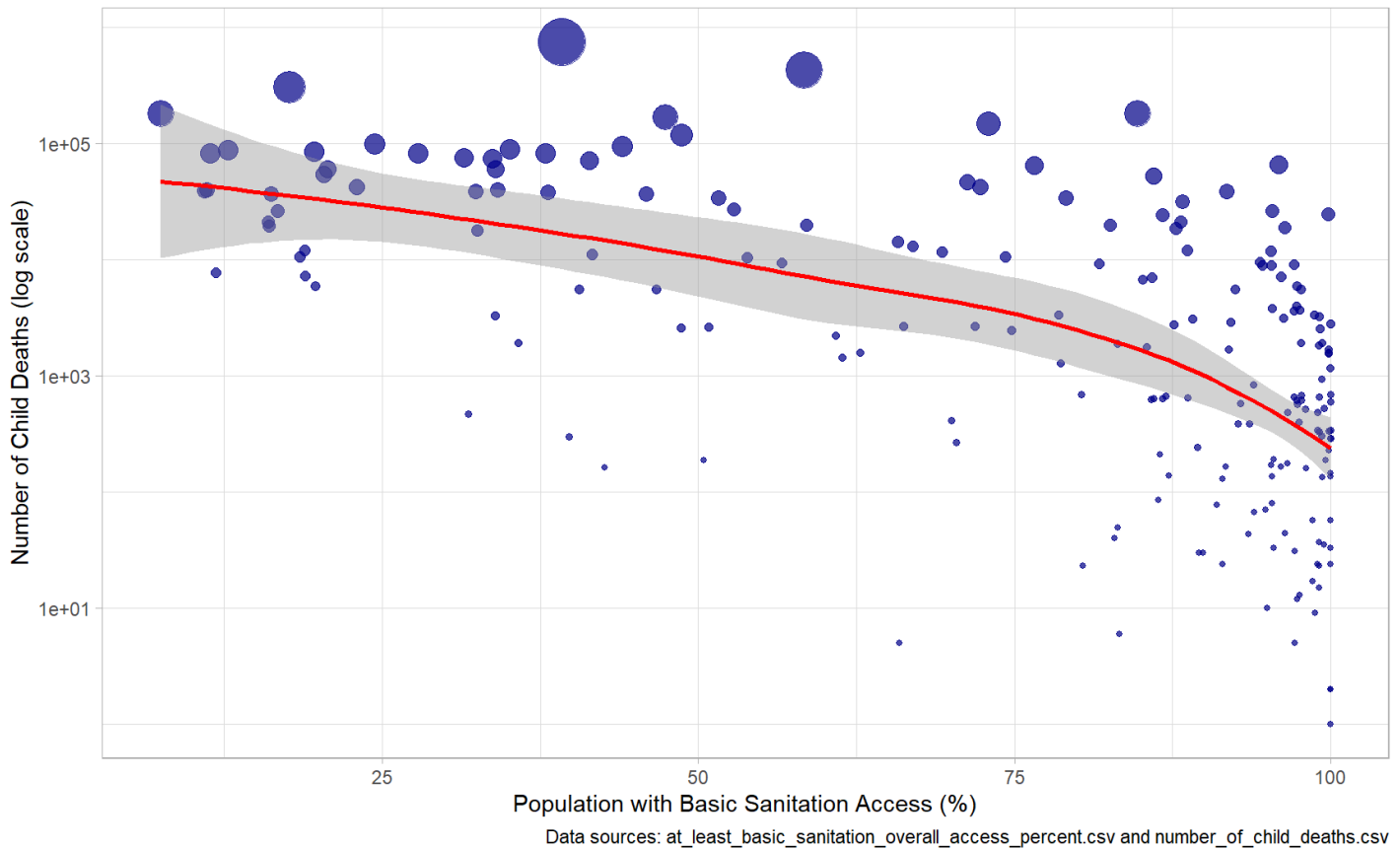
Interpretation

The scatterplot reveals a relationship between access to basic sanitation facilities and the number of child deaths.

The Pearson correlation coefficient (r = -0.359) indicates a moderate negative correlation between these variables.

The logarithmic scale for child deaths was used because the values span several orders of magnitude. The pattern suggests that improvements in basic sanitation access could be associated with reductions in child mortality.

Key observations: 1. Countries with less than 50% sanitation access generally experience higher child mortality 2. The relationship appears non-linear, with the effect of sanitation access potentially stronger at lower levels 3. The R-squared value of the regression model is 0.354, suggesting that approximately 35.4% of the variation in log-transformed child mortality can be explained by sanitation access

This analysis highlights the potential importance of improving sanitation infrastructure as part of efforts to reduce child mortality globally.

Reflections

This project looks seamless to me because of my previouse experience with R and VS Code. I look forward to working more with this. Although will providing my prompt, I made some omissions and I had to stop the code from running, edited my prompt and started it again.