Blue Pulse

ISSS608 Group 9: Fan Jing, Lai Meihao , Lin Xinyue

Issues & Problems

Access to safe water and sanitation remains one of the most pressing global health challenges. As of 2022, **2.2 billion** people lacked safely managed drinking water, and **3.5 billion** lacked access to safe sanitation. These deficiencies are critical contributors to the spread of waterborne diseases such as diarrheal, typhoid, and hepatitis, particularly in regions with underdeveloped infrastructure and fragile public health systems. Despite international efforts, progress toward **Sustainable Development Goal 6 (SDG 6)** — ensuring universal access to clean water and sanitation — has been uneven and insufficient, requiring urgent, data-informed action to address widening global disparities.

Motivation

Complex relationships between water infrastructure and disease vary across time and regions, making quantification difficult. Most studies lack longitudinal, cross-country analyses accounting for temporal trends and regional effects. Our approach combines **interactive data visualization** with **robust panel data modelling** to reveal hidden patterns and generate actionable insights supporting global health and sustainability initiatives.

This research fills a critical knowledge gap through comprehensive analysis of temporal and regional variations, informing more effective policy interventions and resource allocation strategies.

Data Source

The project examined datasets from the **World Bank's World Development Indicators (WDI)** and the **Institute for Health Metrics and Evaluation (IHME)**. The WDI provides data on water, sanitation, and socioeconomic indicators, while IHME contributes disease incidence and mortality data, specifically focusing on waterborne diseases such as Diarrhea, Hepatitis A and Typhoid. The merged dataset covers **140 countries** between **Year 2006 and Year 2021** and consists of **2,240 observations** and **13 variables**.

Approach

Exploratory Data Analysis (EDA)

- Employed **box plots** and **ggbetweenstats** to explore **regional disparities** and **temporal patterns** in water infrastructure and health outcomes.
- Focused on visualizing the **distribution** of indicators by region and year to identify inequality hotspots and outliers.

Panel Linear Modeling (PLM):

- Applied five econometric models: **Pooled OLS**, **Fixed Effects**, **Random Effects**, **Fixed Effects with Time**, and **Dynamic Panel Models** to evaluate the **causal impact** of water factors on disease outcomes.
- Used **Hausman tests** and **serial correlation diagnostics** to select robust models accounting for **country-level and time-specific variation**.

Confirmatory Data Analysis (CDA):

- Designed **animated bubble charts** to illustrate dynamic **relationships between water indicators and disease rates** across regions from 2006 to 2021.
- Developed **correlation bar plots** to assess the **strength and direction** of associations between water variables (e.g., sanitation, open defecation) and diseases (e.g., diarrhea, typhoid).
- Enabled interactive filtering by **region**, **year**, and **indicator type** for targeted insights.

Insight

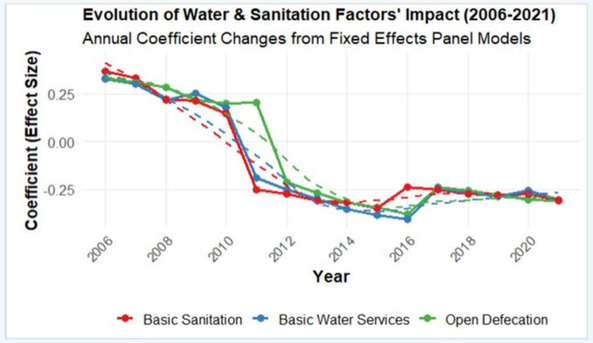
Determinants of Waterborne Disease Incidence

- **Open Defecation** consistently exhibits strong **positive correlation** with **diarrheal and typhoid incidence**, highlighting poor sanitation as a key risk factor.

Water Indicator	Disease	Correlation (r)
open_defecation	diarrhea	0.67
open_defecation	typhoid	0.6

Water Indicator	Disease	Correlation (r)
drinking_water	diarrhea	-0.74
sanitation	diarrhea	-0.81

- **Access to basic water services** (drinking water and sanitation) shows **significant negative correlation** with disease burden, confirming their protective role.



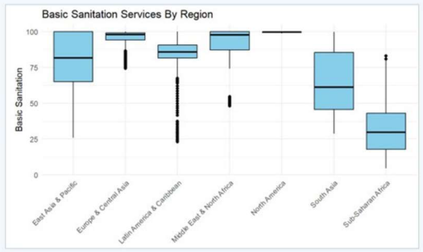
The effectiveness of **Basic Sanitation** and **Basic Water Service** in reducing disease burden evolved over time.

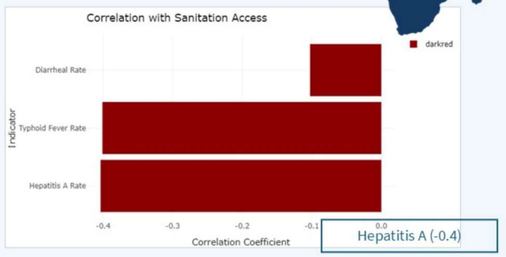
Basic sanitation and water services initially had limited impact but became significantly protective after 2010.

Open defecation consistently showed a strong negative effect, highlighting its risk.

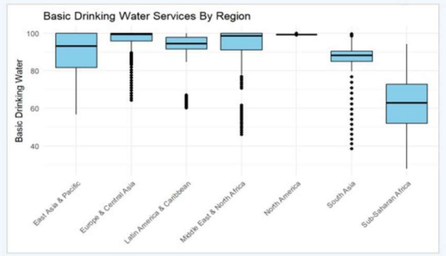
Regional Disparities in Water Infrastructure

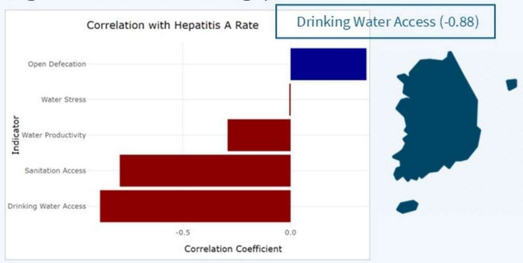
•**Sub-Saharan Africa** reports the **lowest access to sanitation**, aligning with high disease rates and strong correlation strength in regression models.





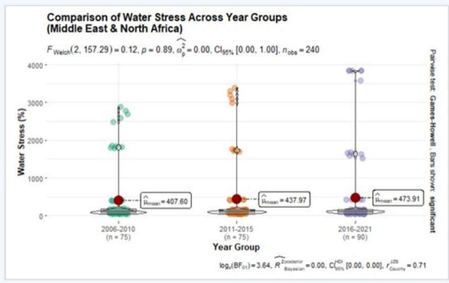
•**South Asia** demonstrates the **highest coefficient magnitudes**, particularly in relation to **drinking water and open defecation**, indicating critical infrastructure gaps.

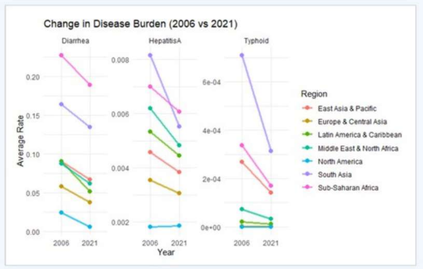




Regional Disparities in Water Infrastructure

•**Sub-Saharan Africa** reports the **lowest access to sanitation**, aligning with high disease rates and strong correlation strength in regression models.





Global health has shown remarkable progress from 2006 to 2021, with most regions experiencing significant reductions in waterborne disease rates, particularly in diarrheal diseases. The visualization reveals persistent regional disparities, with Sub-Saharan Africa and South Asia continuing to face higher disease burdens compared to North America and Europe & Central Asia due to disparities in Water Infrastructure.

Future Work

This project confirms that poor water and sanitation access is strongly linked to higher waterborne disease rates. Through EDA, CDA, and PLM, we reveal key health impacts and provide a Shiny app for interactive insights. Future work may explore disease forecasting using time-series models, climate integration (e.g., drought, rainfall), and equity-focused analyses to uncover disparities across urban–rural and socioeconomic groups.