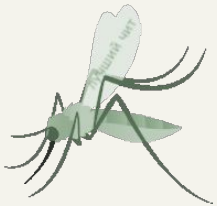
Chasing the Curve: Analyzing Dengue Outbreak Dynamics Across Geographic and Demographic Landscapes



By: Arun Palaniappan, Mubashar Khan, Taylor Whipple

Background:

Dengue is the **most common mosquito-borne disease globally** (Centers for Disease Control and Prevention, 2024) and remains a pressing public health concern throughout both North and South America. Rising case counts and widening geographic spread indicate the need to better understand the drivers behind its transmission and outcomes.

While numerous studies have examined the **relationship between climate change and dengue transmission** (such as Feng et al. and Morin et al.), research on **socioeconomic factors has been limited** primarily to small-scale urban studies (Zellweger et al.).

Motivation:

Our goal for this project is to conduct an analysis of **dengue transmission patterns** across **13 countries** in the Americas (10 South American nations, the United States, Canada, and Mexico). Specifically, we aim to identify the **socioeconomic** and **temporal factors** that influence dengue **incidence, severity,** and **mortality**.

Using data from the **World Bank** and the **Pan American Health Organization (PAHO)**, we will compare case counts, disease severity, and key economic indicators across countries and over time. This study will examine **temporal trends** in transmission and evaluate how **socioeconomic conditions** shape the burden and outcomes of dengue across the region.

Primary Questions:

- How do **key development indicators** (population density, poverty rates, and gross national income (GNI)) **correlate** with **dengue case counts** and **disease severity** across countries in the Americas?
- What **temporal trends** in dengue transmission can be identified that may inform more effective prevention and control strategies?

Hypotheses:

- Countries with **lower socioeconomic status** will experience **higher** dengue incidence and **more** severe cases
- Dengue cases have shown an **increasing trend over time**

Data Sources

Name	Description	Size	Access
Pan American Health Organization (PAHO)	Our **primary** dataset contains comprehensive dengue case data for countries across North and South America. The data is collected by Ministries and Institutes of Health from their respective countries. • This dataset serves as the foundation for analyzing both temporal trends and cross-country differences in dengue incidence, severity, and mortality. Time Coverage: **2013-2023** Key Columns: **Total Dengue Cases, Non-severe Cases, Severe Cases, Deaths**	Size: ~14,000 bytes (195 records) Format: CSV	PAHO Dengue Data (This website contains the same data but in a different format to the original datasource, which is no longer available)
World Bank Group	Our **secondary** dataset contains socioeconomic development indicators from the World Bank Group's 'World Development Indicators' database. The data is compiled from officially recognized international sources. • This dataset complements the dengue case data by enabling analysis of economic and demographic factors. Time Coverage: **2013-2023** Key Columns: **GNI per capita**, **Atlas method** (current US$), **Population density** (people per sq. km of land area), **Poverty headcount ratio at $2.15 a day** (2017 PPP) (% of population)	Size: ~184,000 bytes (781 rows) Format: CSV	Data Bank - World Development Indicators

Data Manipulation and Methods

<u>**Combining The Data:**</u>

To begin the analysis, we first integrated the socioeconomic and demographic indicators with the dengue case data. This involved **using Pandas to load both datasets from CSV files** into the system, creating two separate DataFrames for further processing.

<u>**World Bank Data Transformation:**</u>

For consistency across datasets, we **standardized country names** in the World Bank dataset by mapping "United States" and "Venezuela, RB" to match the naming conventions used in the PAHO data. Additionally, the World Bank dataset required **restructuring using a technique called DataFrame melting**, which involves converting year-specific columns into rows. This was accomplished using the melt() function with appropriate ID and value variables.

<u>**Dengue Data Transformation:**</u>

In the dengue dataset, we **removed leading and trailing whitespace** using .str.strip() and **renamed columns** for improved readability. We then **merged the datasets** using Pandas' merge() function, performing a **left join** on the **country** and **year** variables.

<u>**Clean and Transform The Data:**</u>

The merged dataset presents several data quality challenges, including:

- <u>Inconsistent missing value representation</u> ('..', empty string, 'nan')
- <u>Incorrect data type recognition</u> (object instead of number)

To address these issues, we implemented a cleaning process across all columns except **'Year'** and **'Country'**. Values were converted to strings, comma separators were removed, and all variations of missing values were standardized to **NumPy's NaN** using .replace(). We then converted all numeric columns to the appropriate data types using pd.to_numeric() with the **errors='coerce'** parameter.

To preserve sample size and maintain data integrity, we filled **NaN values** with **country-specific means**, calculated using .groupby('Country').transform() along with a lambda function that computes the average for each metric within each country across all years. **Countries with entirely missing data** (such as **Venezuela** and **Guyana**) were **dropped** to avoid skewing the analysis.

Foundational Analysis

Dengue Data Overview:

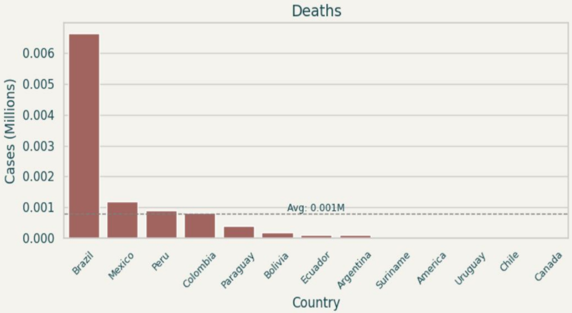
The initial analysis examined dengue fever data from **13 countries over the period 2013–2023**, comprising **143 total observations.** Some initial data points are presented below:

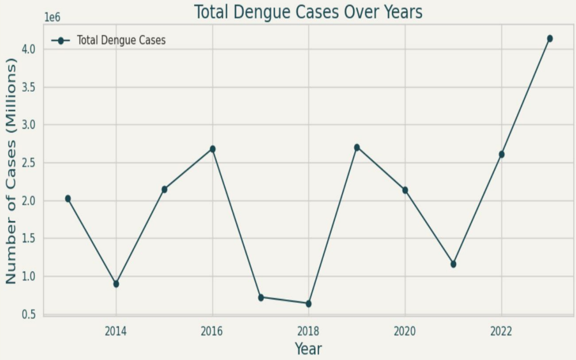
Metric	Value
Total dengue cases	**21,868,710**
Total deaths	**10,365**
Death proportion	**0.0467**

The vast majority of cases (**99.6%**) were **non-severe**, with only a small proportion classified as severe and requiring medical attention. The **severity distribution varied considerably by country**, with some reporting a notably higher share of severe cases.

The analysis revealed significant variation in dengue burden across both countries and time. **Brazil** reported the highest total number of cases, with **over 17 million**, followed by **Mexico** with approximately **1.6 million.**

Mortality data also varied substantially. Brazil recorded the highest number of dengue-related deaths at **6,641**, followed by Mexico with **1,181**, and Peru with **909**. In contrast, several countries, including the **United States**, **Uruguay**, and **Chile**, reported **zero** deaths during the study period.





Over time, dengue incidence showed considerable year-to-year fluctuation. The highest number of cases occurred in **2023**, with **over 4 million**, while **2018** had the lowest, at around **630,000.** Other notable years included **2013**, **2015**, **2016**, **2019**, and **2020**, each recording over 2 million cases.

Economic Analysis: Income Classification

Classification Thresholds:

The World Bank income classification system employs the following categories:

- Low income: GNI per capita below $1,135
- Lower middle income: GNI per capita $1,135 to $4,465
- Upper middle income: GNI per capita $4,465 to $13,845
- High income: GNI per capita above $13,845

Application:

The economic classification was implemented using a Python function that **applies World Bank income thresholds to GNI per capita data.** The `classify_income_world_bank()` function uses conditional statements to categorize countries into the four income levels. The function handles missing data by returning 'Unknown' and was applied to the dataset using pandas' `.apply()` method. This automated approach successfully classified all 143 observations, enabling analysis of dengue patterns across different economic development levels.

Data Distribution:

Among the 143 observations in the merged dataset, the distribution across income levels was:

- Upper middle income: **90 observations (62.9%)**
- High income: **41 observations (28.7%)**
- Lower middle income: **12 observations (8.4%)**
- Low income: 0 observations

GNI per capita values were calculated as country-year specific, acknowledging that economic conditions can change over the study period. This approach **allows for countries to potentially shift between income categories during the 2013-2023 timeframe**, providing a more accurate representation of economic conditions at the time dengue cases were recorded.

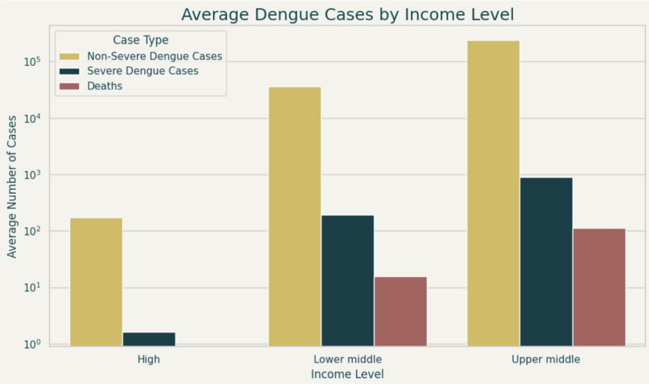
This classification framework enables analysis of how economic development levels correlate with dengue incidence rates, case severity, and mortality outcomes to manage dengue outbreaks effectively.

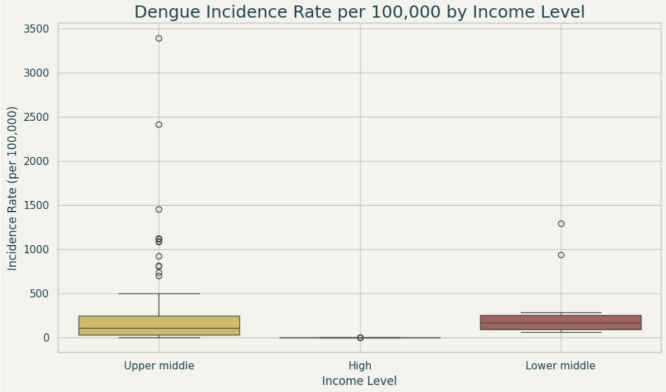
Economic Analysis: Dengue and Income Level

The analysis of dengue cases by income level reveals some clear patterns. Across all income groups, **non-severe** dengue cases constitute the **vast majority,** with severe cases and deaths consistently representing a lower portion.

Case Type	Relative Frequency (per Income Level)
Non-Severe	**Highest**
Severe	**Moderate**
Deaths	**Lowest**

Notably, **upper middle-income** countries experience the **highest average** number of dengue cases overall. This suggests that these countries face a greater burden of dengue compared to others. Lower middle-income countries fall somewhere in between, while high-income countries report the lowest average number of cases.





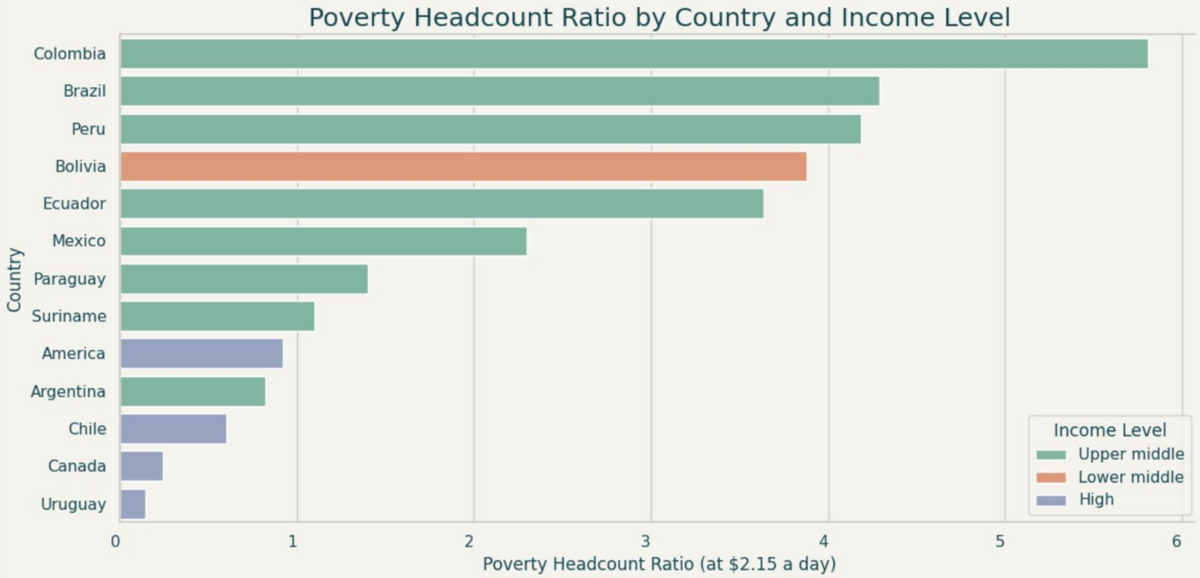
When examining dengue incidence rates relative to population size, a slightly different picture emerges. The median incidence rate is actually **higher** in **lower middle-income** countries compared to upper middle-income countries, while **high-income** countries show much **lower** median rates. This indicates that although upper middle-income countries have a higher total number of cases, the frequency or intensity of dengue outbreaks relative to their population may be greater in lower middle-income countries.

Additionally, some upper middle-income countries experience particularly high incidence rates, as evidenced by several outliers well above the typical range. This variability suggests that within income groups, **individual countries can face substantially different levels of dengue risk.**

Economic Analysis: Poverty Headcount Ratio

There is a clear gradient in poverty levels across the countries analyzed, with **lower and upper middle-income countries exhibiting significantly higher poverty headcount ratios** compared to high-income countries. **Colombia, Brazil, Peru, Bolivia, and Ecuador** rank highest in poverty rates, reflecting broader economic challenges within these nations.

In contrast, **high-income countries such as Canada, Uruguay, and the United States** show the lowest poverty levels, aligning with their stronger economic standing. Interestingly, some **upper middle-income countries**, including **Mexico and Paraguay**, have **lower poverty rates than Bolivia**, a lower middle-income country, but still considerably higher than high-income peers.



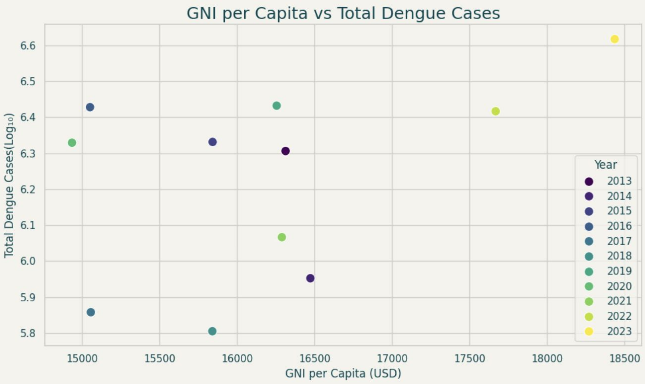
Overall, these disparities emphasize the **strong association between income classification and extreme poverty**, and provide essential context for understanding the socioeconomic conditions that may influence dengue vulnerability and public health response capacity across the region.

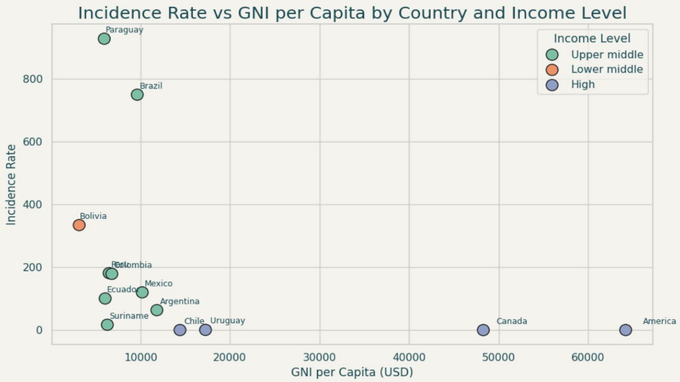
Economic Analysis: Gross National Income Level

From 2013 to 2023, both **GNI per capita** and **total dengue cases** followed **irregular trends**, with no steady increase over time. Only in the most recent years, particularly 2023, did both values reach their highest levels.

However, previous spikes in dengue cases, such as in 2016 and 2019, occurred during years with notably lower GNI. Similarly, 2022 saw one of the highest GNI per capita values but did not show a corresponding drop in dengue burden.

These patterns suggest that **there is no clear relationship between national income levels and total dengue cases over time**,





At the country level, **the United States and Canada show a large gap in GNI per capita compared to all other countries in the dataset**, standing out within the high-income group. Both report **very low dengue incidence rates**, which may reflect stronger health systems and more effective prevention strategies.

In contrast, **Brazil and Paraguay**, with significantly lower GNI, report the **highest incidence rates**, indicating a heavier disease burden relative to their economic capacity. **Bolivia**, with the lowest GNI, shows a moderate incidence rate, exceeding some upper middle-income countries such as Mexico or Argentina. While there is a general trend where **higher GNI is associated with lower dengue incidence**, the variation within income groups highlights the role of additional contextual factors beyond income alone.

Correlation Analysis

GNI per capita shows a **moderate negative correlation with poverty rate** (-0.45), confirming that higher-income countries tend to have lower levels of extreme poverty. Its weak negative correlations with dengue-related variables suggest that **higher income** alone is **not strongly predictive** of reduced dengue burden.

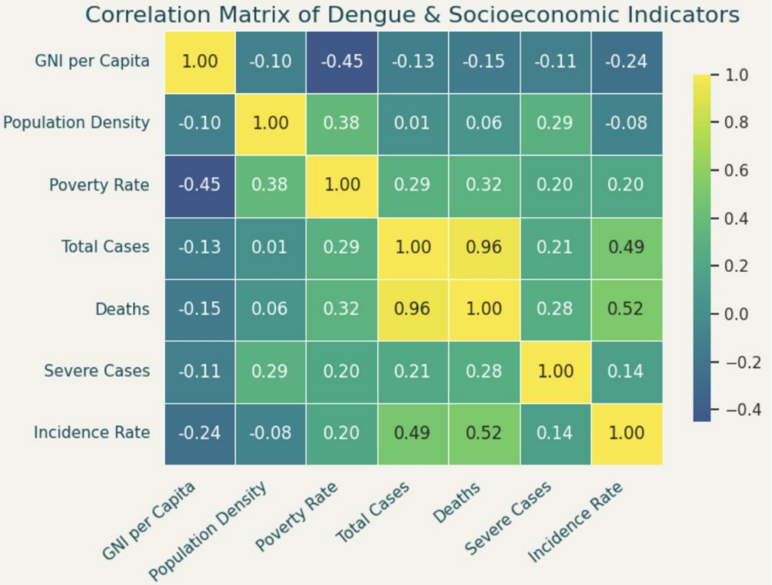
Poverty rate shows **positive correlations** (0.2 to 0.4) with all dengue-related variables, indicating that **higher poverty levels are associated with greater dengue burden**, including more total cases, deaths, and higher incidence rates.

Population density is weakly correlated with most variables but has a **moderate positive correlation with poverty rate** (0.38) and a smaller one with severe dengue cases (0.29). This suggests that **densely populated, lower-income areas** may face **increased risk** of severe dengue transmission.

Among health outcomes, **total dengue cases correlate strongly with deaths** (0.96) and **moderately** with **incidence rate** (0.49), showing that **higher case counts tend to drive both mortality and population-level impact**. Deaths also show a moderate positive correlation with incidence rate (0.52), reinforcing this link.

Severe cases display generally weak correlations with all variables, suggesting that **severity may be influenced by other clinical or environmental factors not captured here**.

Overall, the matrix points to **poverty** — not income or population density — **as the most consistent socioeconomic factor linked to dengue burden**, highlighting the importance of addressing economic vulnerability in disease prevention and control strategies.



Conclusion

Development indicators and dengue burden

While poverty shows some relationship with higher dengue burden, the overall correlations between socioeconomic factors (poverty, GNI, and population density) and dengue outcomes are relatively weak or inconsistent. This suggests that development context matters, but **no single indicator can fully explain patterns of transmission or severity**, pointing to the importance of localized, multifactor analysis.

Temporal patterns in dengue transmission

Dengue trends over time are **irregular and outbreak-driven**, with significant spikes in specific years like 2016, 2019, and 2023. These patterns highlight the need for **flexible and responsive public health strategies**, rather than reliance on steady trends or economic development as predictors.

Limitations:

This analysis has several limitations that should be taken into account when interpreting the findings.

- **Disease surveillance and reporting systems** vary widely across countries, with lower-income nations often having less comprehensive dengue monitoring capabilities.
- Although Venezuela and Guyana were initially included in the study, **insufficient data availability** led to their exclusion, which somewhat narrows the geographic coverage of the analysis.

These data gaps highlight broader challenges of conducting thorough regional disease surveillance and suggest that the true burden of dengue in the Americas may be underestimated and underreported.

Next steps

Future analysis should aim to integrate climate and environmental variables such as rainfall and temperature, with these socioeconomic factors to explore additional avenues of dengue transmission. It will also be important to investigate subnational data to identify disparities within countries that may not be visible at the national level. Conducting regression and spatial analyses can also be used to uncover key drivers of outbreak intensity and geographic spread.

Statement of Work

Arun contributed to the project by completing the data merging and manipulation portion of the code along with the corresponding section of the final report. Taylor focused on the analysis portion of the code while also assisting in the creation of the final report. Mubashar contributed by developing data visualizations, supporting the analysis of results, and assisting in the creation of the final report. Overall, the collaboration was effective with each team member taking ownership of distinct project components while supporting shared deliverables like the final report.

To improve collaboration, the team could benefit from establishing more structured check-in meetings and deadlines throughout the project timeline to ensure a better flow of work between different components.

Sources and Citations

Centers for Disease Control and Prevention. (2024). *Dengue outbreaks, 2024*. https://www.cdc.gov/dengue/outbreaks/2024/index.html (accessed 5/13/25)

Feng, F., Ma, Y., Qin, P., Zhao, Y., Liu, Z., Wang, W., & Cheng, B. (2024). Temperature-driven dengue transmission in a changing climate: Patterns, trends, and future projections. *GeoHealth*, 8, e2024GH001059. https://doi.org/10.1029/2024GH001059

Luchikom. (n.d.). *Mosquito GIF* [Animated image]. Tenor. https://tenor.com/view/luchikom-gif-14677406624236508679

Morin, C. W., Comrie, A. C., & Ernst, K. (2013). Climate and dengue transmission: Evidence and implications. *Environmental Health Perspectives*, 121(11-12), 1264-1272. https://doi.org/10.1289/ehp.1306556

Pan American Health Organization. (n.d.). *PAHO logo* [Image]. https://lh6.googleusercontent.com/proxy/-eAepmFm2F-dT7m3oYySAgGFfAo-53PF9VRDoDcXJfA1sRF-gQ4Er8bVAD1gdLHzrJqnDkeuBWaFoWH4k_jIsDOjRNvPmee83zNDikkJAE8K6Li3os4NI0AaoaNTF6unNhQ7LCGQKVIiBZJfpfOfo7-6pQxO1g=s1793

World Bank. (2024, July 1). World Bank country classifications by income level for 2024-2025. *World Bank Blogs*. https://blogs.worldbank.org/en/opendata/world-bank-country-classifications-by-income-level-for-2024-2025

World Bank Group. (n.d.). *World Bank Group logo* [Image]. https://lh6.googleusercontent.com/proxy/c11m96zOhKnB_mTiFmhA1PaGUXhtIBbyNDlDDp7ocm33ZHFk6A0qpawybHHBWria4FkcCupH5aFUf4_s0DqfXm8u1VcmGhSjqm1WMYwbVte-cVb6kj3WmjvZSUOqA5WWHU61sw3V-W3nQqYUMFE7qk8M2kUan5_3OhTwwXaECz3oekX4GMkByvxyxhfXa6TLLA=s1280

Zellweger, R. M., Cano, J., Mangeas, M., Taglioni, F., Mercier, A., Despinoy, M., Menkès, C. E., Dupont-Rouzeyrol, M., Nikolay, B., & Teurlai, M. (2017). Socioeconomic and environmental determinants of dengue transmission in an urban setting: An ecological study in Nouméa, New Caledonia. *PLoS Neglected Tropical Diseases*, 11(4), e0005471. https://doi.org/10.1371/journal.pntd.0005471