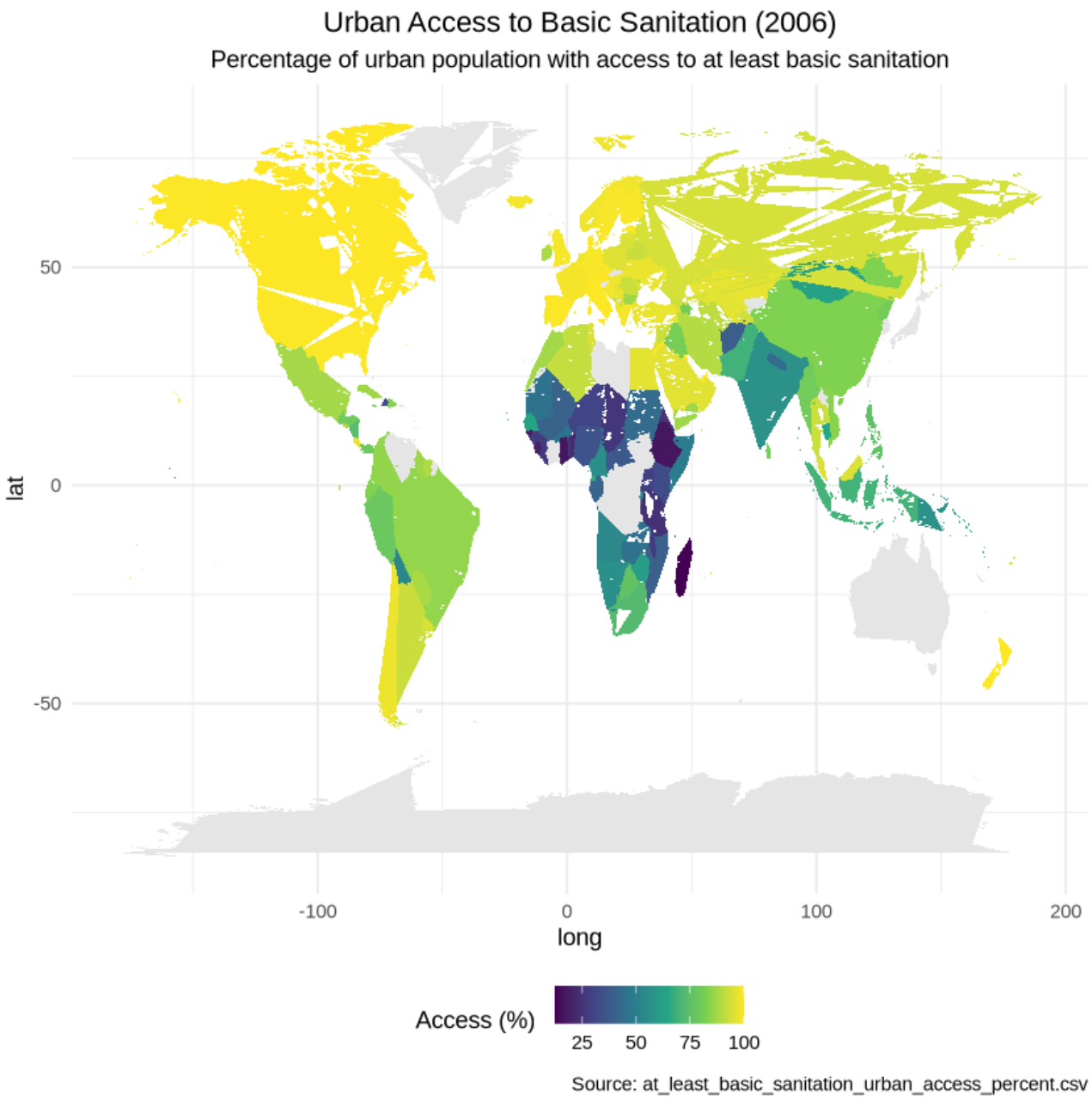
WORKSHOP 2 | AI-Assisted Data Analysis Practice with Julius

Name: Eric Ebenezer Boham || 19th May 2025.

Indicators chosen: Basic sanitation and Malaria Cases || **Year**: 2006 (recent year available in malaria csv)

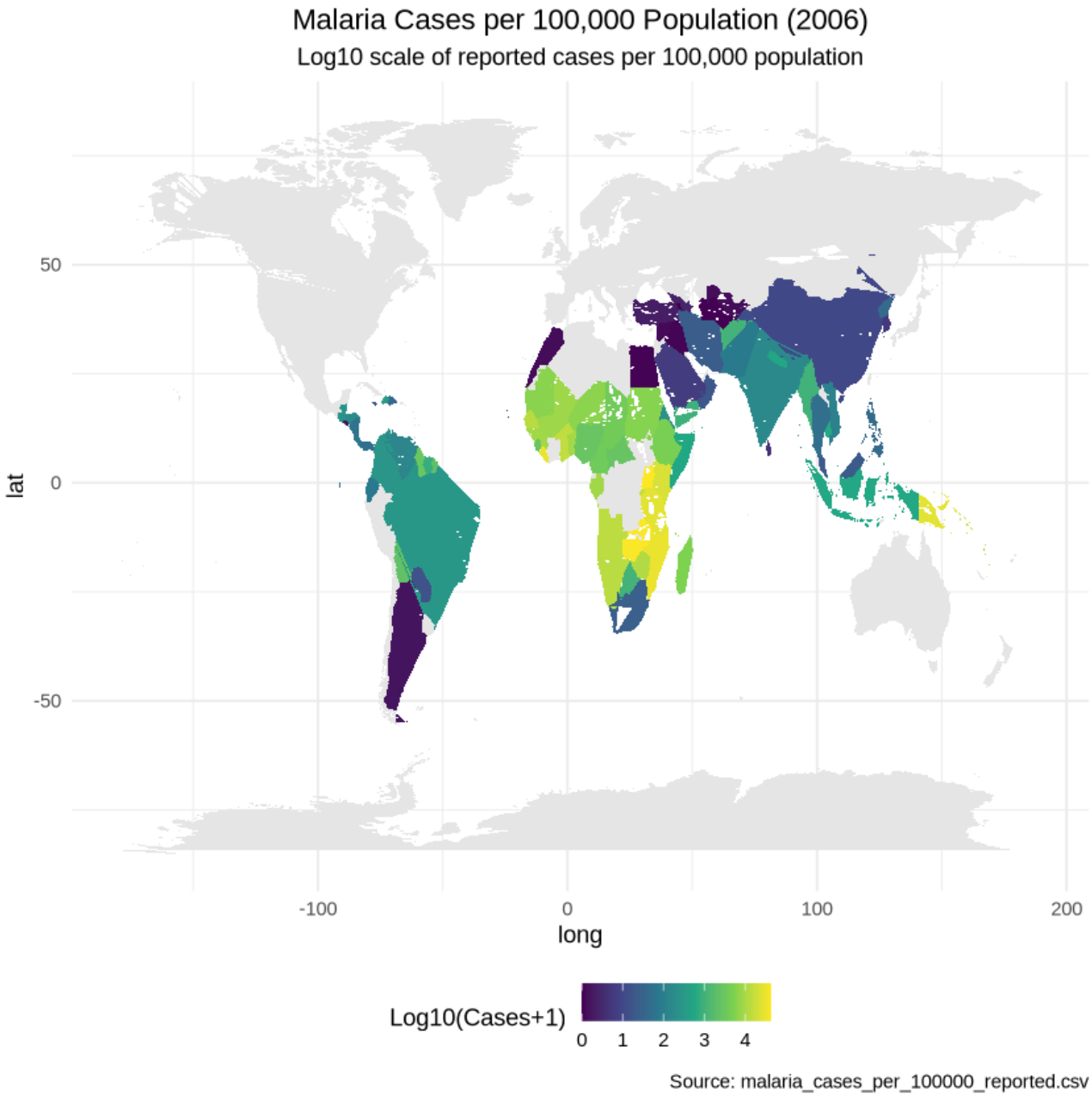
1. **World map visualization for Basic Sanitation:**



The map above shows significant global disparities in urban sanitation access. Developed regions like North America, Western Europe, and Australia show high access rates (90-100%). In contrast, many African countries and parts of South Asia display much lower access rates, often below 60%.

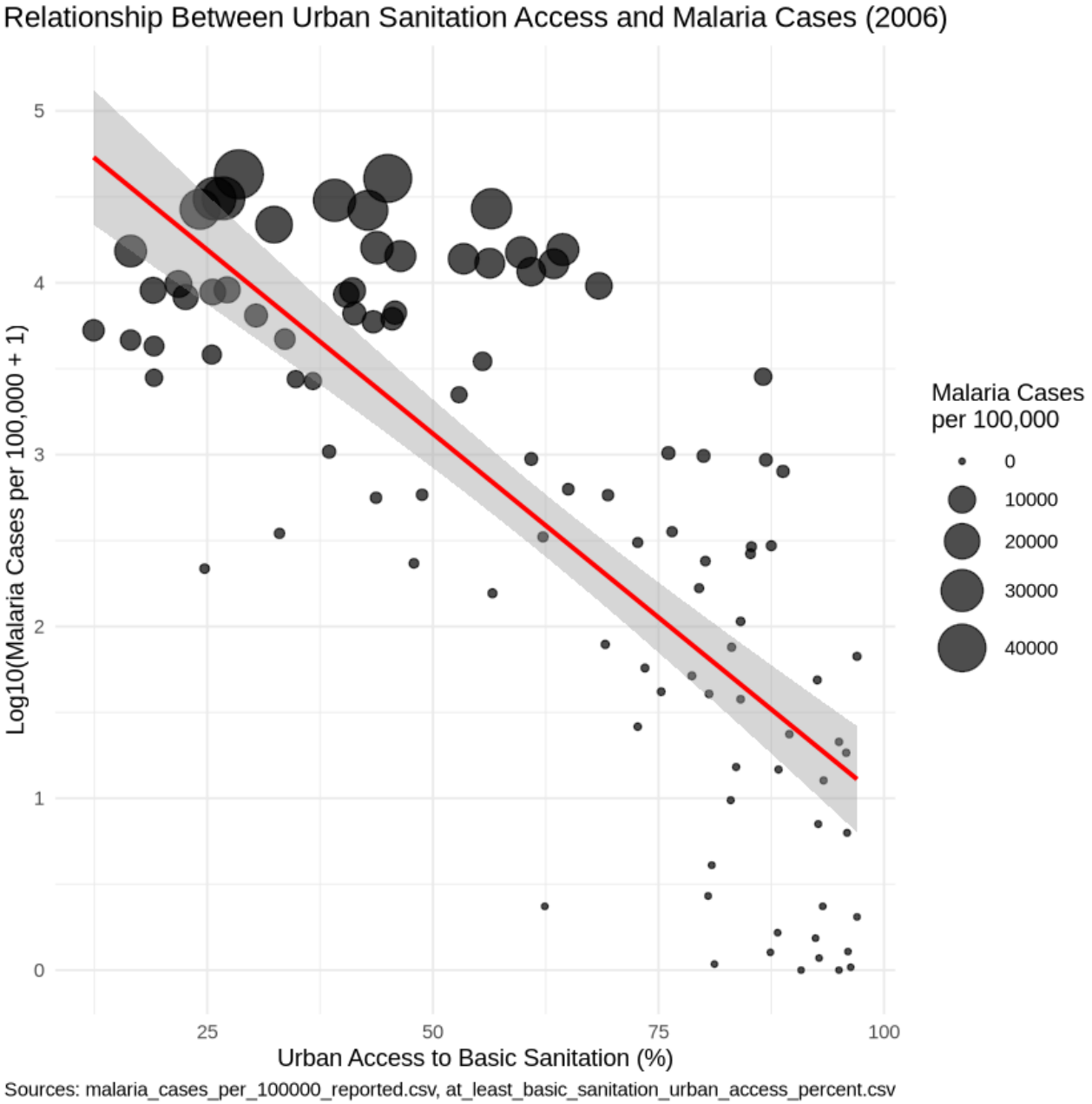
This pattern reflects the economic development divide, with lower-income countries struggling to provide adequate sanitation infrastructure in urban areas.

2. **World map visualization for Malaria Cases per 100,000:**



The map of malaria cases per 100,000 population (displayed on a log scale), reveals that malaria burden is heavily concentrated in tropical and subtropical regions, particularly in sub-Saharan Africa. Countries in central and western Africa show the highest case rates, while regions like North America, Europe, and most of East Asia report very few or no cases. This distribution aligns with the geographic range of malaria-carrying mosquitoes, which thrive in warm, humid climates.

3. **Scatterplot to visualize the relationship between urban sanitation access and malaria cases:**



4. **Description and interpretation of the relationship between the two indicators:**

The scatterplot reveals a strong negative correlation between urban sanitation access and malaria cases:

- **Strong Negative Correlation**: The Pearson correlation coefficient (r = -0.77) indicates a strong negative relationship between these variables. As urban sanitation access increases, malaria cases tend to decrease significantly.
- **Statistical Significance**: The p-value (< 2.2e-16) shows this relationship is highly statistically significant, not due to random chance.

- **Pattern Visualization**: The downward sloping red trend line clearly illustrates this negative relationship. Countries with sanitation access above 80% generally have very low malaria rates, while those with access below 60% often face much higher malaria burdens.
- **Potential Implications**: While correlation does not prove causation, this relationship suggests that improving urban sanitation infrastructure might be an important component of malaria control strategies. Better sanitation may reduce standing water where mosquitoes breed and improve overall public health conditions.
- **Outliers**: Some countries deviate from the trend line, suggesting that other factors (climate, healthcare systems, and malaria control programs) also play important roles in determining malaria prevalence.

This analysis supports the hypothesis that public health infrastructure investments, particularly in sanitation, may contribute to reducing the burden of vector-borne diseases like malaria.

5. **Short Reflection:**

Ease: Using Julius was quite easy, especially following the prompts that had been given in the workshop instructions.

Difficulty: For a couple of times (both yesterday and today), the "container" crushed and Julius stopped running. I had to rerun it when that happened. Today, it was able to proceed and give the output above.

What was surprising: I had to change the dataset from malaria deaths to malaria cases, as Julius surprisingly struggled to plot the world map for the data of malaria deaths. Julius could not install some packages for this: eg. 'sf' package. It eventually used another packaged to run the code, but it gave a blank image as the output.

In the end, I tried the whole process again today, this time using the data for malaria cases. It still could not use the 'sf package' but it was able to use the maps package and this time gave the output above.

There were also some parts of the process where it switched to python (when it was having the difficulty above), and then switched back to R to continue.

What I learnt: I learnt that AI tools can assist well with data analysis, from coding, to verifying the code and editing the challenges faced, to the interpretation of the outputs in a very short time.

I think this is incredibly useful for someone like me with little knowledge in data analysis. In that, it ceases to be a complex field for me, and it encourages me to be able to learn the basics and interact with the data I have.