PROJECT REFLECTION

Lalitha Pranathi Pulavarthy

FA25-B585-Biomedical Analytics

Saptarshi Purkayastha

13 Dec 2025

1. Background

Before starting my group project, I explored few resources indicating that Sierra Leone has some of the worst maternal and child health indicators globally [1]. My group project presents a multi-level geospatial and predictive analysis of maternal and child health indicators in Sierra Leone, using data from the District Health Information Software 2 (DHIS2) platform [2]. One observation became clear early in this project. National averages can hide as much as they reveal. I realized that understanding health system performance requires looking at where targets are met, where they are missed and asking why. Our project examined these patterns across three health domains, including maternal health, child immunization, and child malnutrition, by tracing data through the organizational hierarchy of districts, chiefdoms, and facilities.

2. Introduction

The analysis addressed ten questions spanning different analytical approaches. I grouped related analyses by health domain. Questions 1 and 2 examined Antenatal Care Intermittent Preventive Treatment (ANC IPT) coverage trends and forecasting. Question 8 analyzed the geographic distribution of births attended by skilled health personnel. These three questions are grouped together as they all address maternal care during pregnancy and delivery. Questions 4 through 7 focused on childhood immunization, including national measles coverage, drill-down analysis to identify underperforming districts, chiefdoms and facilities, correlation analysis between dropout rates and coverage. Questions 3a and 3b addressed child malnutrition, with geographic mapping and a convolutional neural network approach to predict district-level malnutrition from choropleth maps. Question 9 integrated these analyses into a dashboard, and Question 10 examined data quality through validation rules. *See appendix for questions.*

3. Maternal Health Indicators

The analysis of maternal health indicators in this section is informed by established theoretical frameworks and methodological principles. Theoretically, ANC IPT coverage is rooted in WHO recommendations for preventing malaria during pregnancy, a leading cause of maternal morbidity in endemic regions [3]. In 2021, 13.3 million pregnancies in WHO African region were exposed to malaria infection, contributing to 11% of neonatal deaths and 20% of stillbirths [4]. Sierra Leone, located in West Africa with high malaria transmission, faces similar challenges. To address this burden, WHO recommends administering Intermittent Preventive Treatment in pregnancy (IPTp) using sulfadoxine-pyrimethamine to pregnant women in areas with moderate to high malaria transmission, beginning in the second trimester and continuing each scheduled antenatal care visit [4]. A multi-country study indicates that IPT in pregnancy decreases the incidence of low birth weight by 29, neonatal mortality by 31% and severe maternal anemia by 38% [4]. In DHIS2, the above described IPTp is monitored through ANC IPT coverage indicators. See Fig. 1 for ANC IPT 1 and ANC IPT 2 coverage indicator description from DHIS2.

Beyond antenatal interventions, skilled attendance during childbirth represents another critical component of the maternal health continuum. The same multi-country study documented skilled birth attendance ranging from 40% in Madagascar to 86% in DR Congo [4]. A recent analysis of Demographic and Health Survey data across 26 sub-Saharan African countries found that 58.4% women attended four or more ANC visits, while 11.2% received no ANC at all [5]. Hence, the connection between antenatal care and births by skilled attendants is well-established as women who attend ANC are more likely to deliver with a skilled attendant because ANC visits provide opportunities for birth preparedness counseling and encourage facility-based deliveries. A skilled health personnel is a professional trained and certified to provide safe pregnancy, delivery, and newborn care. They include Midwives, State Enrolled Community Health Nurse (SECHN), Community Health Officer (CHO), and Maternal and Child Health (MCH) Aides. For DHIS2 description of births attended by skilled health personnel see Fig. 2 below.

Name	ANC IPT 1 Coverage
Numerator description	IPT 1st dose total given
Numerator expression	`IPT 1st dose given at PHU Fixed+IPT 1st dose given at PHU Outreach+IPT 1st dose given by TBA Fixed+IPT 1st dose given by TBA Outreach`
Denominator description	ANC 1st visit total
Denominator expression	`ANC 1st visit Fixed+ANC 1st visit Outreach`
Annualized	No
Indicator type	Per cent, 100
Group membership	• ANC • Reproductive Health

Name	ANC IPT 2 Coverage
Numerator description	IPT 2 doses given total
Numerator expression	`IPT 2nd dose given at PHU Fixed+IPT 2nd dose given at PHU Outreach+IPT 2nd dose given by TBA Fixed+IPT 2nd dose given by TBA Outreach`
Denominator description	ANC 1st visit total
Denominator expression	`ANC 2nd visit Fixed+ANC 2nd visit Outreach`
Annualized	No
Indicator type	Per cent, 100
Group membership	• ANC • Reproductive Health

Figure 1. ANC IPT 1 and ANC IPT 2 coverage

Name	Births attended by skilled health personnel
Numerator description	Live births attended by skilled health personnel
Numerator expression	`Live births Midwives, Female + Live births Midwives, Male + Still births Midwives, Female + Still births Midwives, Male + Live births SECHN, Female + Live births SECHN, Male + Still births SECHN, Female + Still births SECHN, Male + Live births CHO, Female + Live births CHO, Male + Still births CHO, Female + Still births CHO, Male + Live births MCH Aides, Female + Live births MCH Aides, Male + Still births MCH Aides, Female + Still births MCH Aides, Male`
Denominator description	Null
Denominator expression	`1`

Figure 2. DHIS2 description of Births attended by skilled health personnel

Methodologically, the section employs time series analysis that are widely applied in epidemiological forecasting including linear regression and simple exponential smoothing methods to examine coverage and generate forecasts that were taught in lecture four and lecture five of INFO-B585 Fall 2025 course [6].

3.1 ANC IPT Coverage Trendlines

The data processing for question 1 analysis involved extracting ANC IPT 1 and ANC IPT 2 coverage indicators at national-level for Sierra Leone across an eighteen-month period from January 2024 through June 2025. The question required examining both indicators simultaneously to identify whether and when their trendlines would intersect, which would indicate convergence in coverage between first and second IPT doses. The analysis proceeded in two stages. First, DHIS2's built-in linear trendline feature was applied within the Data Visualizer as seen in Figure 3 below.

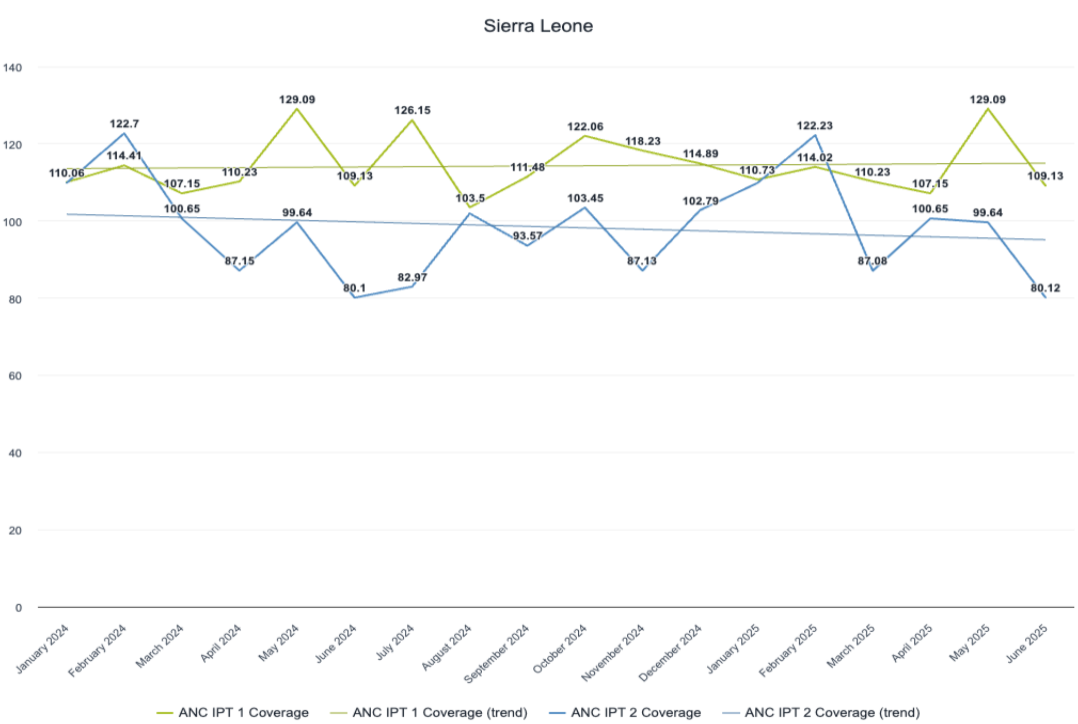

Figure 3. Trendlines for ANC IP1 and ANC IPT 2 coverage

Second, to address the question of when the trendlines might meet in the future, the data was exported to R where linear regression using ordinary least squares method along with a 6-month forecast as illustrated in Figure 4 below.

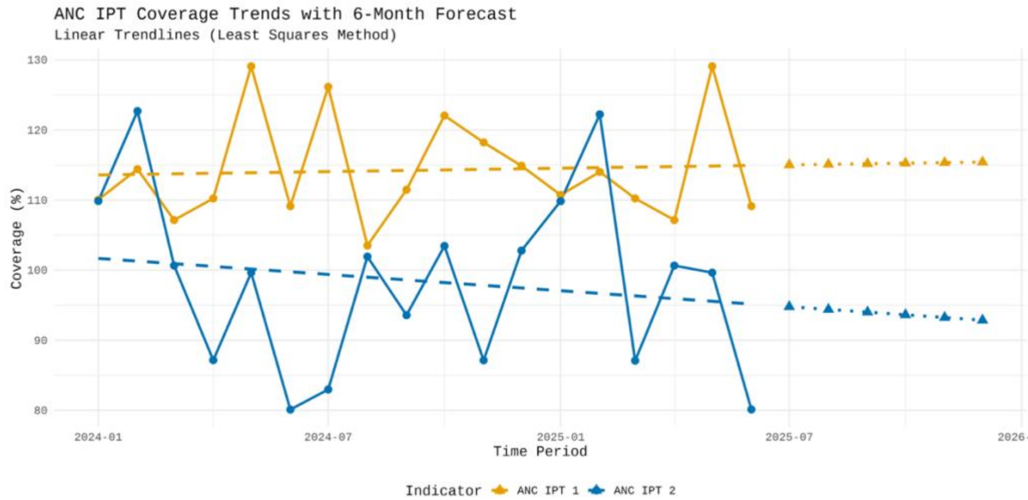

Figure 4. 6-month forecast for ANC IPT coverage

ANC IPT 1 coverage remained consistently higher than ANC IPT 2 coverage throughout the observation period, with IPT 1 fluctuating between 103% and 129% while IPT 2 exhibited greater volatility. The trendlines for ANC IPT 1 Coverage and ANC IPT 2 Coverage do not intersect within the January 2024 to June 2025 observation period, nor within the six-month forecast window extending to January 2026. This suggests worsening retention rates between doses. This divergence pattern signals that pregnant women in Sierra Leone are initiating malaria prevention through ANC but failing to complete the recommended dosing schedule at increasing rates. The research question specifically asked when the trendlines would meet, which requires extrapolating two linear equations to calculate their intersection point. For this purpose, ordinary least squares was appropriate because it produces linear equations that can be solved algebraically to determine if and when two trendlines converge. The six-month forecast horizon was selected to for a reasonable extrapolation window where linear assumptions might still hold. However, during the presentation Q&A session, I recognized important limitations when using this approach for forecasting future values. Linear regression assumes independence between observations, but this assumption may be violated in time series data if autocorrelation is present.

The model also weights all observations equally rather than prioritizing recent values, which are typically more predictive of near-term trends. Underlying variability is not being considered in this approach as mentioned in lecture five [6].

3.2 SES Forecasting

The data processing for Question 2 involved extracting ANC IPT 2 Coverage for Western Area district across a twenty-three month period from January 2024 through November 2025. The data was exported from DHIS2 and converted into a time series object in R to enable application of Single Exponential Smoothing as mentioned in lecture five [6]. SES computes a weighted average of past observations with weights declining exponentially as data becomes older [6]. In a single exponential smoothing forecast, each new forecast is based on the previous forecast plus a percentage of the difference between the actual value and that forecast, expressed as:

$$\text{New forecast} = \text{Old forecast} + \alpha(\text{Actual value} - \text{Old forecast})$$

$$\mathbf{F_t = F_{t-1} + \alpha(A_{t-1} - F_{t-1})}$$

where F_t is the forecast for period t, F_{t-1} is the forecast for period t-1, α is the smoothing constant, and A_{t-1} is the actual value in period t-1. The research question required selecting two alpha values and comparing their forecasting performance. We performed a grid search from $\alpha = 0.1$ to 0.9 to evaluate smoothing behavior. The grid search indicated that $\alpha = 0.3$ provided the most stable, low-variance smoothing, while $\alpha = 0.7$ produced the most reactive smoothing that closely follows recent changes. As shown in the lecture five examples, when $\alpha = 0.0$ the forecast never changes from the initial value, and when $\alpha = 1.0$ the forecast simply equals the previous actual value. We selected $\alpha = 0.3$ and $\alpha = 0.7$ to compare a steady versus highly responsive SES model and evaluated their performance using the last six months of data (June through November 2025). Both SES models showed variable forecasting performance across the last six months (June - November 2025), with the α = 0.3 model achieving better overall accuracy (MAE = 9.97) compared to α = 0.7 (MAE = 10.33). While the α = 0.7 model responded more quickly to recent changes, it produced more volatile forecasts with some large errors, particularly in August 2025. In contrast, the α = 0.3 model provided more consistent forecasts by smoothing out short-term fluctuations. See Figure 5. The December 2025 forecast values of 95.75% (α = 0.3) and 99.51% (α = 0.7) both fall within a reasonable range of recent observations. Overall, the smoother forecasting approach with α = 0.3 offered better short-term accuracy for this dataset, suggesting that ANC IPT 2 coverage data benefits from reducing the influence of month-to-month volatility.

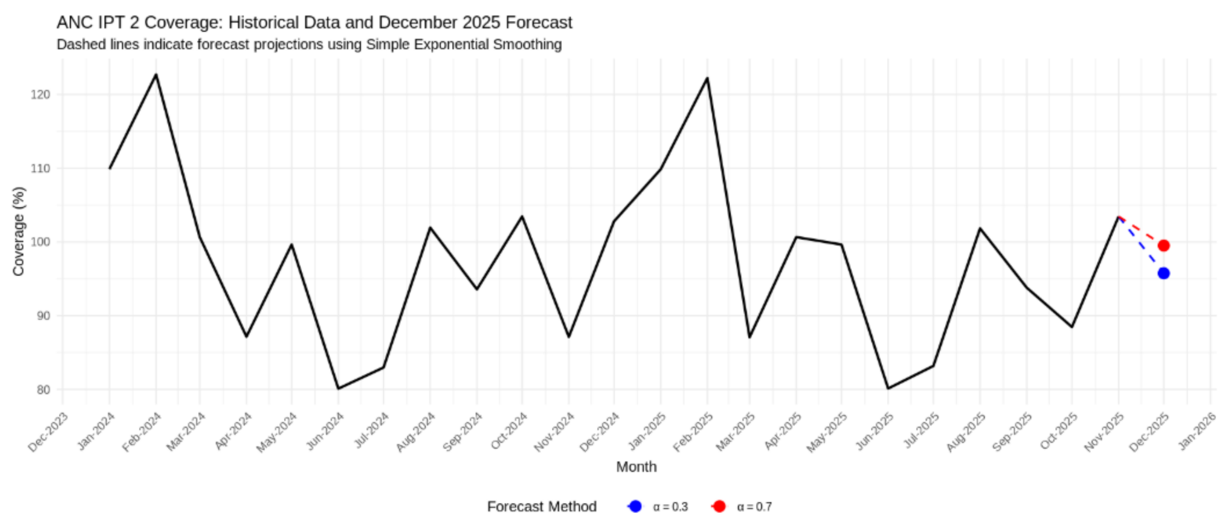

Figure 5. SES forecast for ANC IPT 2 Coverage in December 2025

The Q&A session during presentation made me realize the limitations of the approach to select alpha. The grid search approach for alpha selection has limitations. The search evaluated discrete values from 0.1 to 0.9 in increments of 0.1, meaning the true optimal alpha could lie between tested values. Additionally, the alpha that minimizes error on historical data may not perform optimally on future data if underlying patterns change. A more robust approach could employ time series cross-validation to select alpha, though for this analysis the grid search provided sufficient insight into the tradeoff between stability and responsiveness.

3.3 Lowest Births by skilled attendants

The data processing for Question 8 required careful consideration of available indicators. Two initial indicators were examined: "Births attended by skilled health personnel (estimated pregnancies)" and "Births attended by skilled health personnel (registered live births)." However, neither indicator told the complete story as both were missing stillbirths, and neither told the right story as both were calculated as percentages. To address these limitations, a derived indicator was created: "All births with skilled attendants." This derived indicator keeps all four types of skilled attendants (Midwives, State Enrolled Community Health Nurse, Community Health Officer, and Maternal and Child Health Aides), includes stillbirths, and replaces ‘estimated pregnancies’ or ‘registered live births’ denominators with 1 to create a factor data type. The data was extracted at the district level for 2024 (yearly) and visualized using DHIS2's mapping functionality with proportional symbols like explained in lecture eight [7]. See Figure 6 below.

The analysis revealed Bonthe as the district with the fewest skilled-attended births, a finding attributable to its location in the far southern coast where riverine and island terrain creates natural barriers to overland movement. The next-lowest districts, Koinadugu, Kono, and Pujehun show similar structural challenges. All three lie far from the capital and are predominantly rural, with sparse population density and limited transport infrastructure. Unlike districts surrounding Freetown, these areas lack major highways, as indicated by the absence of primary (yellow) or secondary (orange) road networks on the map. This creates long travel distances, fewer reliable transport routes, and reduced ease of reaching health facilities, particularly for skilled birth care. The combination of remoteness, challenging terrain, and inadequate transportation infrastructure appears to reduce access to skilled birth services in these areas.

The decision to create a derived indicator rather than using the existing DHIS2 indicators demonstrates the importance of critically evaluating available data elements before analysis. The original percentage-based indicators would have masked the absolute volume differences between districts, potentially misleading program planners about where the greatest number of unattended births occur. However, during the Q&A session, two minor limitations were identified. The analysis did not examine whether facility type is a statistically significant predictor of non-reporting. While fixed versus outreach is already accounted for in the indicator construction, facility type (such as Community Health Centers, Community Health Posts, or Maternal and Child Health Posts) was not tested as a predictor. A more rigorous approach would test whether certain facility types are systematically less likely to report skilled birth data. Such analysis could reveal structural reporting gaps tied to facility capacity rather than true service delivery deficits. This question is my most favorite because of the complexity and nuances involved.

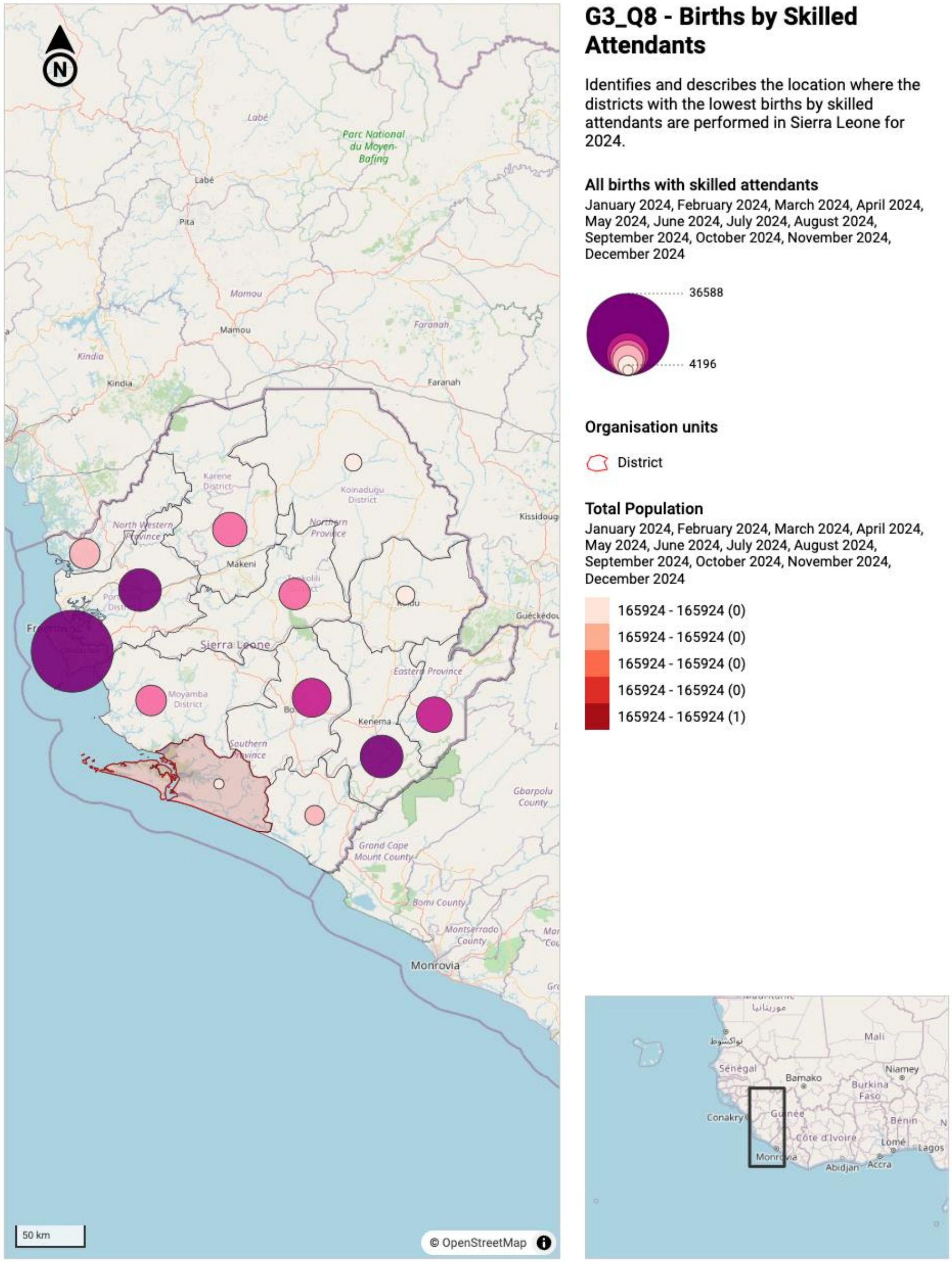

Figure 6. All births with skilled attendants showing the lowest

4. Childhood Immunization

The analysis of childhood immunization indicators in this section draws on established recommendations and recent evidence from sub-Saharan Africa. To achieve regional measles elimination objectives, coverage of 95% or higher with two doses of measles-containing vaccine is recommended [8]. Despite this target, a 2024 study examining WHO-UNICEF estimates found that regional measles second dose (MCV2) coverage in the African Region increased from only 7% in 2013 to 49% in 2023, still far below the 95% threshold and lower than the global MCV2 coverage of 74% . Sierra Leone, which introduced MCV2 in 2015, achieved MCV1 coverage of 90% and MCV2 coverage of 73% by 2023, with a dropout rate of 19% between the two doses. The study noted that most countries have failed to raise their MCV2 coverage levels near MCV1 coverage levels, indicating that a large proportion of children drop out between the first and second doses [8].

Methodologically, this section employs visualization-based analysis, hierarchical drill-down techniques, and correlation analysis. For Question 4, national-level Measles 1 coverage for children less than one year was examined across 2024 and 2025 to identify months where the 60% coverage target was missed. Questions 5 and 6 applied hierarchical drill-down analysis to identify specific districts, chiefdoms, and facilities with the lowest coverage, enabling characterization of underperforming areas and identification of trend patterns in low vaccination coverage. For Question 7, a bar chart comparison of Penta 1-Measles dropout rates and Measles coverage was created for all districts, followed by correlation analysis to examine the relationship between these two indicators.

4.1 Measles Coverage Visualization Analysis

The data processing for Question 4 analysis involved extracting Measles Coverage <1y indicator at national-level for Sierra Leone across a twenty-four month period from January 2024 through December 2025. The indicator is calculated as the sum of measles doses given through fixed and outreach services divided by the total population under one year. The question required examining monthly coverage against the 60% performance target to identify months where the target was missed. As mentioned in lecture 8, a line chart was selected as the most appropriate visualization for this analysis because line graphs are optimal for displaying continuous data over time and revealing temporal patterns, trends, and fluctuations [7].

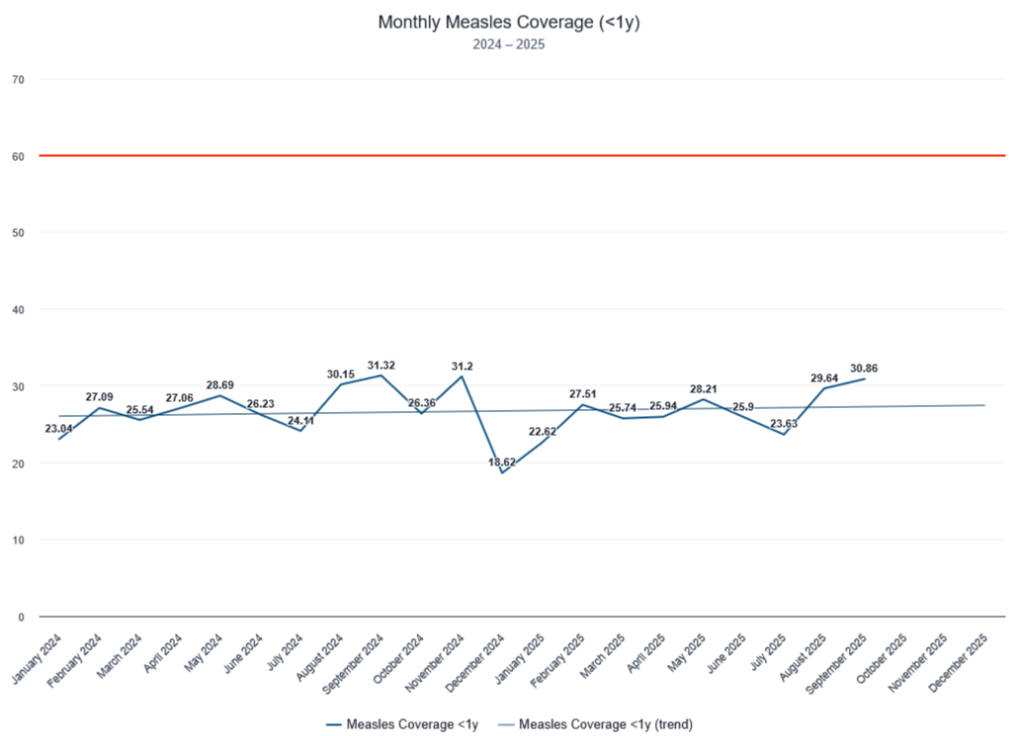

Figure 7. Monthly measles coverage from 2024-2025

The analysis was conducted using DHIS2's Data Visualizer. The line chart was created with a reference line at 60% to enable visual comparison against the target. As shown in Figure 7, national measles coverage remained consistently below the 60% target for all months in the observation period, fluctuating between approximately 18% and 32%. The lowest coverage occurred in December 2024 at 18.62%, while the highest was recorded in September 2024 at 31.32%. The persistent gap between observed coverage (18-32%) and the 60% target represents a substantial shortfall of approximately 28-42 percentage points. This pattern reflects systemic national underperformance rather than isolated monthly declines, suggesting challenges in measles vaccination uptake and/or data completeness across the country. The flat trendline visible in Figure 7 confirms that coverage showed no meaningful improvement over the two-year period. This finding warranted further drill-down analysis to identify specific geographic areas contributing to the national underperformance, which is addressed in the following sections.

4.2 Drill-Down Analysis of Underperforming Areas

The data processing for Question 5 required hierarchical drill-down analysis to identify specific districts, chiefdoms, and facilities that missed the 60% coverage target. Similar to the approach used in Question 8 for skilled birth attendance, a derived indicator was created: "Measles Doses <1y." This derived indicator keeps both types of locations where children under one year received doses (fixed and outreach) and replaces the "Total population < 1 year" denominator with 1 to create a factor data type. The data was extracted at district, chiefdom, and facility levels for 2024-2025 and visualized using DHIS2's mapping functionality with thematic layers. See Figure 8 for the district bar chart and Figure 9 for the drill-down map.

At the district level, all 13 districts with data fell below the 60% expected threshold. As shown in Figure 8, Western Area achieved the highest coverage at 55.6%, still below the target, while Kono recorded the lowest at 10.09%. Most districts fell between 10-26%, highlighting a widespread national gap rather than isolated low-performing areas. Drilling down to the chiefdom level, the map legend reveals that 129 chiefdoms fell below expected (0-60%), only 4 achieved good coverage (60-90%), none reached above target (90-101%), 1 was out of range, and 18 chiefdoms reported no data. At the facility level for 2024, 51 facilities fell below expected, 51 achieved good coverage, 11 were above target, 31 were out of range, and 457 facilities reported no data.

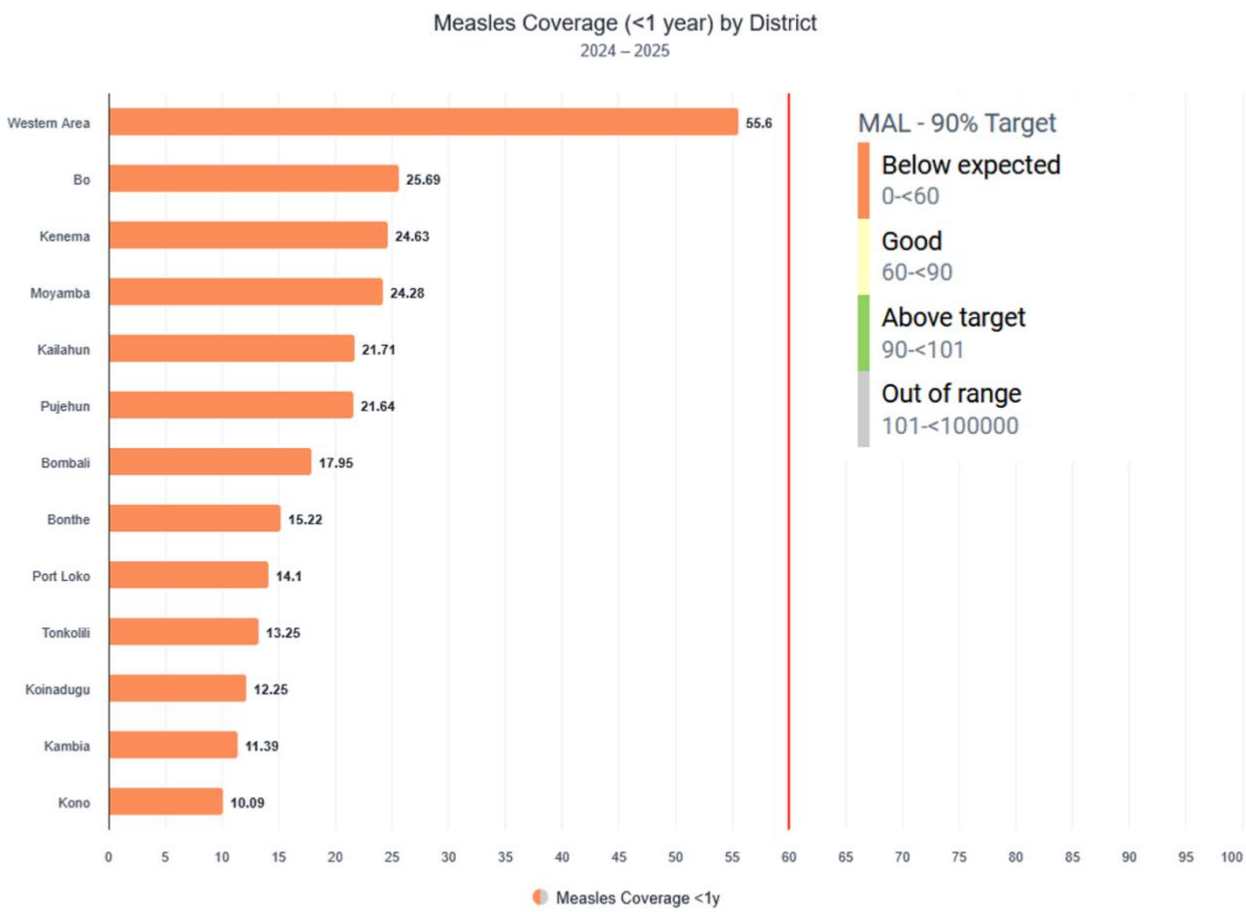

Figure 8. District horizontal bar chart for measles coverage

The analysis revealed important characteristics of underperforming areas. The lowest-performing districts were Kono (10.09%), Kambia (11.39%), and Koinadugu (12.25%). The map shows that measles coverage remains below the 60% target across nearly all chiefdoms, with the lowest-performing areas also containing the highest concentration of "no-data" facilities. The clustering of non-reporting facilities in remote, hard-to-reach regions suggests that the appearance of low coverage is driven by both true service gaps and significant under-reporting in these areas.

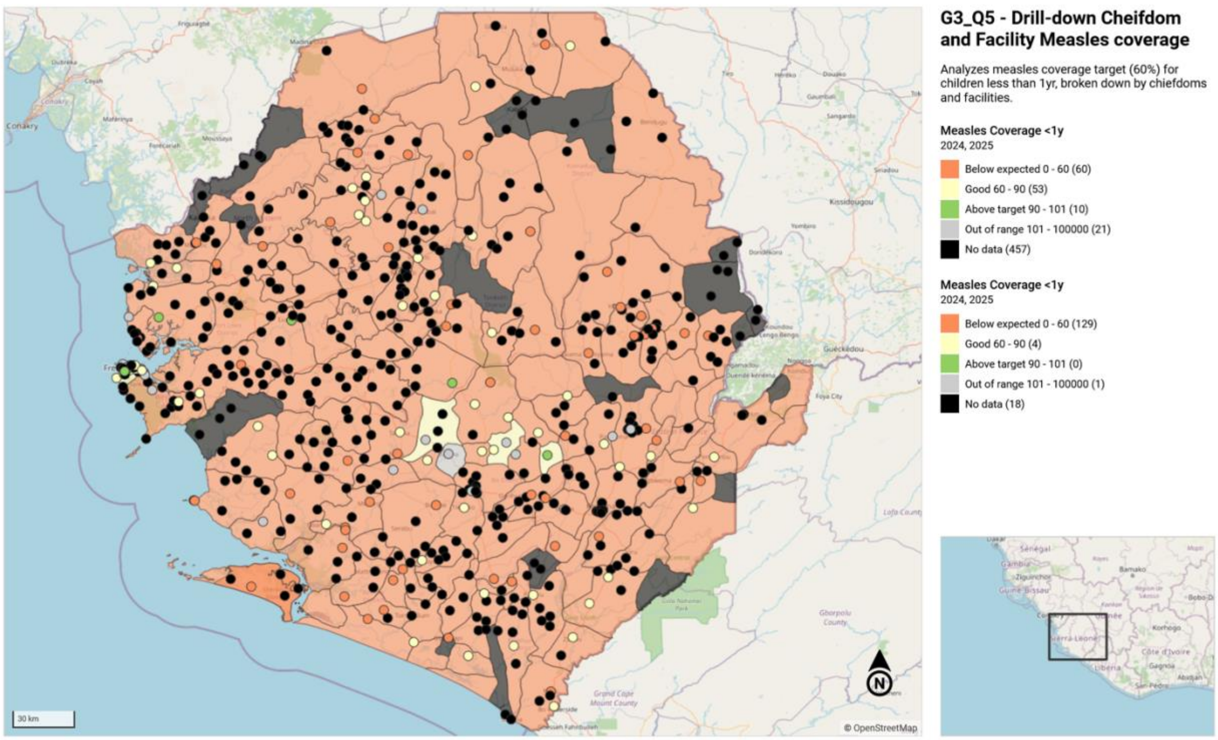

Figure 9 Drill-down map (better visualized on dashboard)

Additionally, many chiefdoms and facilities received substantial measles vaccine stock but recorded zero or no doses administered, suggesting data quality and reporting gaps rather than an actual absence of immunization services. These reporting lapses are concentrated in rural and hard-to-reach areas, indicating that low coverage values may reflect limited data visibility rather than low vaccination activity. See Figure 10 below.

Chiefdoms

2024, 2025					
	Stock PHU dispensed Measles	Measles Doses <1y	Total population < 1 year	Measles Coverage <1y - Calc	Measles Coverage <1y
Bramaia	104		1 284.98	0	
Gbinleh Dixion	208		961.99	0	
Bendu Cha	228	62	123.5	8.1	[illegible]
Mafindor	244		250.5	0	
Toli	302		206	0	
Dema	318	92	215	18.6	21.37
Gbane Kandor	318		229.5	0	
Sittia	330	206	375.5	35.15	27.39
Mambolo	334	395	1 760.48	7.1	11.2
Kpanga Krim	342		177	0	
Gbo	344	241	118	91.53	101.98
Selenga	350	216	227.5	0	47.41
Langrama	382		187	0	

Facilities

2024, 2025					
	Stock PHU dispensed Measles	Measles Doses <1y	Total population < 1 year	Measles Coverage <1y - Calc	Measles Coverage <1y
Wellington Health Centre	4 674	3 990	1 375.48	38.39	144.8
Kissy Health Centre	3 630	2 483	1 328.98	35.97	93.2
Matotoka CHC	3 320	182	237.5	[illegible]	[illegible]
PMO Clinetown	2 880		2 474.47	0	
Kroo Bay CHC	2 802	2 090	687.99	[illegible]	[illegible]
Ginger Hall Health Centre	2 240	1 447	578.49	139.16	124.
George Brook Health Centre	2 076	1 219	924.99	8	65.
St Anthony clinic	1 976	1 708	1 254.48	13.39	67.9
Jenner Wright Clinic	1 912	1 292	682.99	42.17	94.4
Lunsar CHC	1 896	913	454.49	147.86	100.
Tombo CHC	1 728	814	677.49	107.46	59.9
Masuba MCHP	1 722		1 393.98	0	
UFC Magburaka	1 720		777.49	0	

Figure 10. Measles vaccine stock

One key takeaway during the Q&A session was the 457 facilities with no data may reflect differences in service delivery modality. Fixed facilities are more likely to have consistent reporting systems, while outreach services (mobile or community-based delivery points) may lack the infrastructure for regular data submission. The non-reporting facilities may be delivering vaccinations without capturing the data in DHIS2. A more rigorous approach should have disaggregated coverage by delivery modality to determine whether the "no data" pattern reflects true service gaps or systematic under-reporting from outreach services.

4.3 Trend Analysis of Low Measles Coverage

Question 6 required describing the type of trend observed in low measles vaccination coverage. Based on the findings from Questions 4 and 5, three distinct but interrelated trends were identified. The analysis revealed three interconnected patterns. Temporally, measles coverage remained consistently inadequate throughout 2024-2025, never approaching the 60% target despite minor month-to-month variation and a marginally upward trajectory, pointing to entrenched structural barriers rather than episodic setbacks. Geographically, the drill-down revealed that poor coverage spans virtually all districts and chiefdoms, with the weakest performers also exhibiting the greatest proportion of facilities lacking data entirely. This pattern suggests that location-related factors significantly influence vaccination outcomes. From a data quality standpoint, numerous facilities that received considerable vaccine supplies reported administering none, exposing systematic under-reporting that artificially suppresses coverage figures and masks actual service delivery. Collectively, the temporal line graph, district-level bar chart, and geographic drill-down offer mutually reinforcing views, illuminating the timing of coverage shortfalls, their spatial distribution, and how reporting deficiencies and access barriers jointly account for the national gap.

4.4 Correlation Analysis: Dropout Rates and Measles Coverage

The data processing for Question 7 involved extracting two indicators at the district level for the period January 2024 to June 2025: Dropout rate (Penta 1 to Measles) and Measles Coverage less than 1 year. The dropout rate measures the percentage of children who received the first dose of pentavalent vaccine but did not complete their immunization schedule through measles vaccination. A grouped bar chart was constructed to compare dropout rates and measles coverage across all districts in Sierra Leone for the eighteen-month period. As illustrated in Figure 11, the visualization demonstrates considerable district-level variability, though it does not reveal a clear or consistent inverse pattern between the two measures. To quantify this relationship statistically, Pearson correlation analysis was conducted. The resulting coefficient of negative 0.295, displayed in the correlation heatmap in Figure 12, reflects a weak negative correlation. Interpreted practically, this finding indicates that districts experiencing higher dropout rates show a modest tendency toward lower measles coverage, though lacks strength.

The implication is that although efforts to minimize dropout between vaccine doses may yield incremental coverage gains, addressing wider systemic and operational challenges is essential for meaningful progress in immunization outcomes.

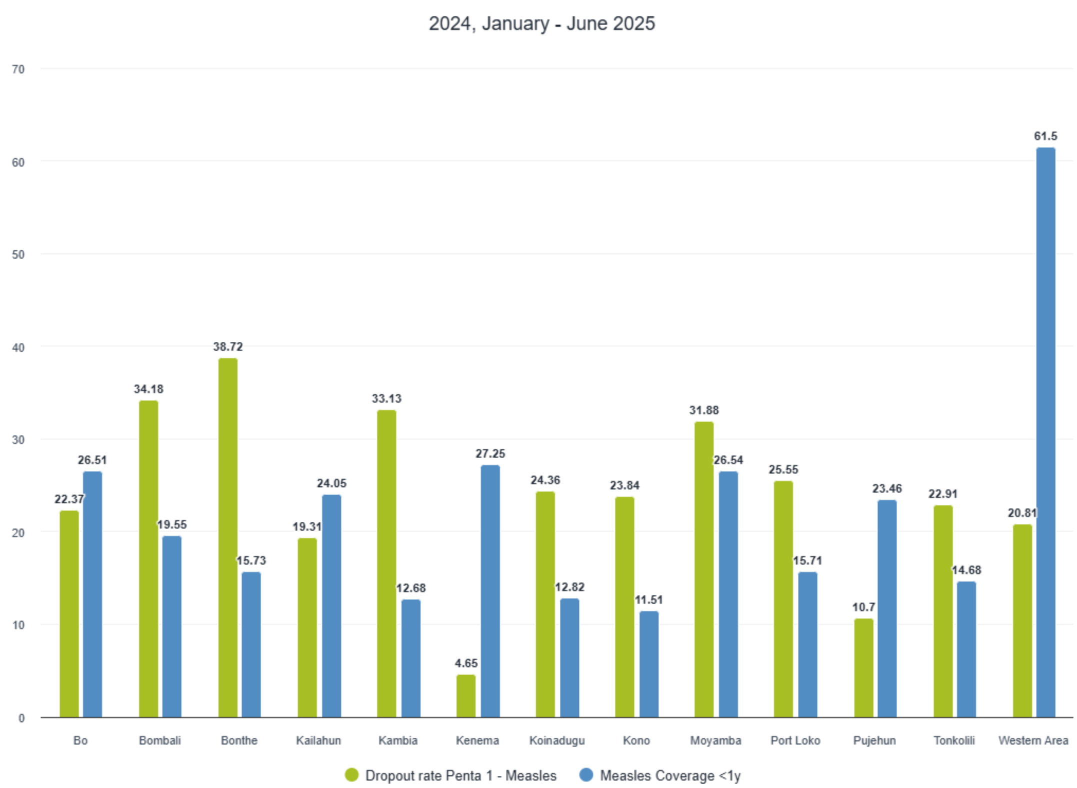

Figure 11. Barchart comparing Dropout Penta 1-Measles and Measles Coverage <1y by district

However, a critical limitation of this analysis must be acknowledged which I realized during the Q&A session. In lecture 5, we discussed assumptions before doing the correlation analysis. The Pearson correlation coefficient assumes that the relationship between variables is linear and that both variables are normally distributed. Additionally, Pearson correlation is sensitive to outliers and assumes homoscedasticity. We applied Pearson correlation directly without first testing these assumptions. A more rigorous approach would have included diagnostic steps such as examining scatterplots to assess linearity, conducting normality tests (such as Shapiro-Wilk), checking for influential outliers, and verifying whether the variance in measles coverage remains constant across different levels of dropout rates. If these assumptions were violated, alternative methods such as Spearman's rank correlation (which does not assume normality or linearity) might have been more appropriate. This oversight represents a methodological weakness in the analysis, and future work should incorporate proper assumption testing before selecting correlation methods.

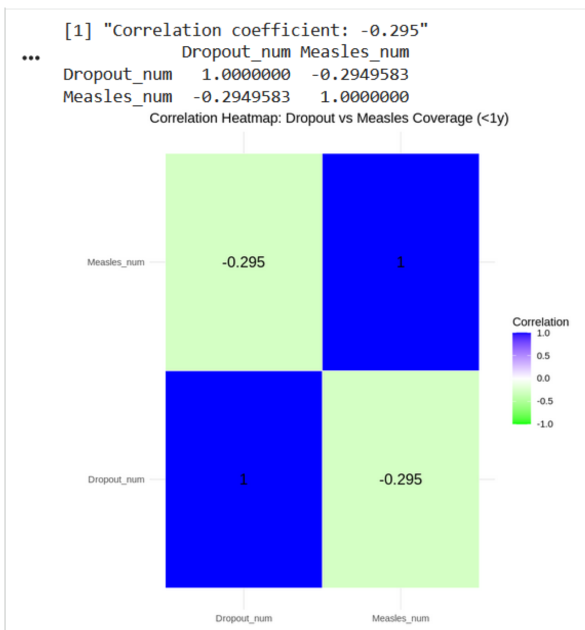

Figure 12. Correlation analysis - Heatmap

5. Child Malnutrition

The analysis of child malnutrition indicators draws on established frameworks for assessing nutritional status in children under five. Weight-for-height (wasting) is a key indicator of acute malnutrition, with WHO classification defining moderate malnutrition as below negative two standard deviations and severe malnutrition as below negative three standard deviations from the median. A 2021 study analyzing the 2019 Sierra Leone Demographic and Health Survey found that stunting prevalence was 31.6% in rural areas and 24.0% in urban areas [9]. The study noted that in Sierra Leone, more than 3 million people are estimated to lack access to sufficient food, and consequently, chronic undernutrition is widespread. Furthermore, stark differences were observed between rural and urban areas in factors associated with childhood stunting, which were attributed to nutrition adequacy, availability of healthcare services, and socioeconomic differences in rural and urban areas in Sierra Leone. These findings underscore the importance of examining malnutrition at the sub-national level.

Methodologically, this section employs geographic visualization and machine learning approaches. Question 3a used thematic mapping to identify chiefdoms with the lowest malnutrition rates and those near the capital with high malnutrition rates. Question 3b applied a convolutional neural network (CNN) to predict district-level malnutrition rates from choropleth map images, representing an innovative application of computer vision to health indicator prediction.

5.1 Geographic Distribution of Child Malnutrition

The analysis of Question 3a examined malnutrition distribution across chiefdoms in Sierra Leone for the years 2024 and 2025. We chose a choropleth map as discussed in lecture 8 [7]. Since the question did not specify a severity level, a derived indicator was developed by merging two existing measures: children with weight-for-height below 70% (representing severe cases) and those between 70 to 79% (representing moderate cases). This aggregated metric, termed "Weight for height less than 79% rate," offers a fuller picture of the acute malnutrition situation and prevents undercounting regions where moderate cases predominate while severe instances remain limited.

The indicator was displayed through a choropleth map, where chiefdoms were color-coded based on their malnutrition rates according to the legend. An underlying Google-maps-style layer helped pinpoint Freetown, the capital, enabling assessment of how near or far chiefdoms were from this urban center. This visualization strategy permitted interpretation of malnutrition trends in light of geographic features, topography, infrastructure quality, and service accessibility, moving beyond simple numerical rankings. The map analysis revealed that Malema chiefdom exhibited the lowest malnutrition rate at roughly 2.4%, situated inland and away from the coastal capital area. Conversely, multiple chiefdoms adjacent to Freetown, such as Kaffu Bullom (around 33.21%) and Lokomasama (around 34.77%), displayed considerably elevated malnutrition rates despite being geographically close to the capital. This observation demonstrates that mere proximity to the capital provides no guarantee of nutritional protection. When examined within a wider geographic framework, elevated malnutrition rates appear in chiefdoms characterized by difficult terrain, inadequate road systems, and weak connections to major urban centers, even when situated near Freetown [11]. In coastal areas like Kaffu Bullom and Lokomasama, climate-related flooding compounds these difficulties by damaging infrastructure, undermining food systems, and restricting service access [12]. Collectively, the findings suggest that malnutrition patterns in Sierra Leone result not simply from distance to the capital, but from the interplay of infrastructure quality, terrain characteristics, and accessibility factors. See Figure 13 below.

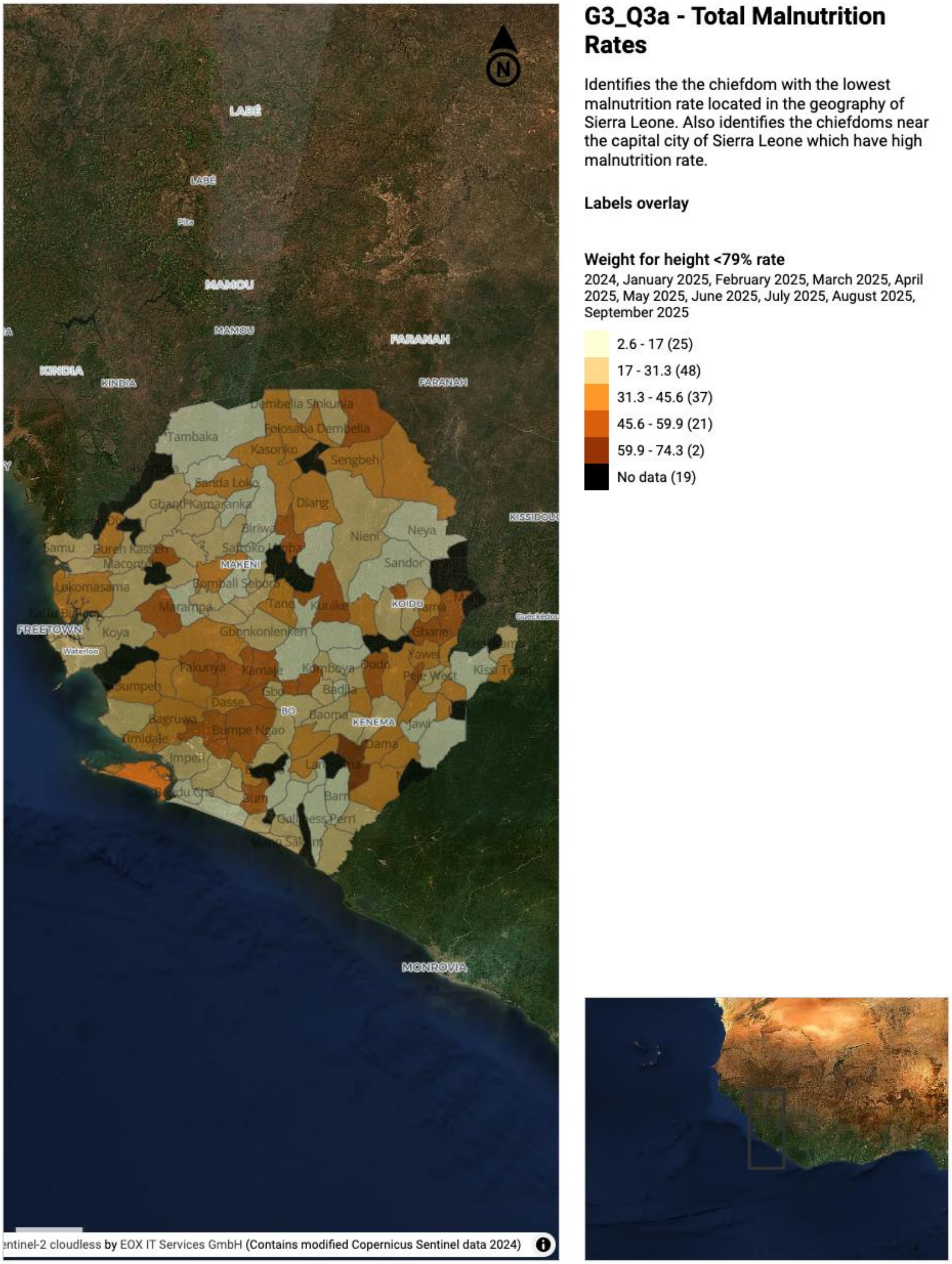

Figure 13. Child malnutrition rate

5.2 CNN-Based Prediction of District Malnutrition

Question 3b explored whether convolutional neural networks could predict malnutrition rates for a specific district by analyzing spatial patterns in choropleth maps. We followed the CNN basics taught in lecture 13. The dataset comprised 21 monthly choropleth maps spanning January 2024 through September 2025. District segmentation was performed by selecting a representative pixel within each district and applying a region-growing algorithm to generate consistent district masks across all images. See Figure 14 below. A critical challenge was the inconsistent legends across maps, with varying ranges and cutoff values. To address this, custom bins were defined based on the actual underlying malnutrition rate data rather than the inconsistent map legends. The full distribution across all districts and months, including range, quartiles, and histogram, informed the determination of bin count and cutoff points. Once finalized, bins were mapped to a standardized yellow to orange to red color palette and applied to the segmented districts, creating a consistent categorical representation suitable for model training. To prevent information leakage from future to past, the 21 maps were split chronologically: the first 14 months for training, months 15 through 20 for testing, and September 2025 held out entirely for final prediction. All images underwent standardization, including resizing to uniform dimensions, pixel value scaling to a 0 to 1 range, RGB format conversion, and stacking into the tensor structure required for CNNs. Labels focused exclusively on Bo district's bin for each month. Bo's malnutrition rate was mapped into the custom bins, converted to integer class indices, and transformed into one-hot encodings for training. For the holdout month, the bin index was retained to enable comparison between model predictions and true values. The model architecture employed ResNet50 pretrained on ImageNet as a frozen convolutional feature extractor, with weights remaining unchanged during training. A custom classification head was added, consisting of a global average pooling layer, a 128-unit dense layer with ReLU activation, dropout for regularization, and a softmax output layer matching the number of bins. The model was compiled using the Adam optimizer and categorical cross-entropy loss. Training involved only the classification head while ResNet50 remained frozen, running for 30 epochs with a batch size of 2, using the training set for learning and the test set for validation. Two models were trained to evaluate different prediction scenarios. The unmasked model retained Bo district's visibility across all maps, allowing the classifier direct access to Bo's pixel information. The masked model removed Bo's region entirely, replacing it with a neutral color, forcing the model to infer Bo's malnutrition bin solely from spatial patterns in surrounding districts. This masked approach better reflects the research question's intent: predicting Bo's rate when the map does not display Bo's value. Comparing both models reveals the extent to which predictions depend on Bo's own visual features versus broader regional patterns. For September 2025, Bo district's true malnutrition rate was 4.79%, falling into bin 4 to less than 5. Both the unmasked and masked models predicted bin 5 to less than 6, with the unmasked model assigning a probability of 0.421 to this bin and the masked model assigning 0.445. Both models achieved an F1 score of 0.667. Despite one model having access to Bo's pixel data and the other relying solely on surrounding districts, they produced identical predictions and performance metrics. This convergence likely resulted from several factors: the small test set provided insufficient data to distinguish model behaviors, Bo predominantly appeared in the same bin throughout the dataset leading both models to learn the majority class, spatial patterns in surrounding districts were sufficiently strong for the masked model to infer Bo's value accurately, and with ResNet50 frozen and limited data, both models learned nearly identical classifiers.

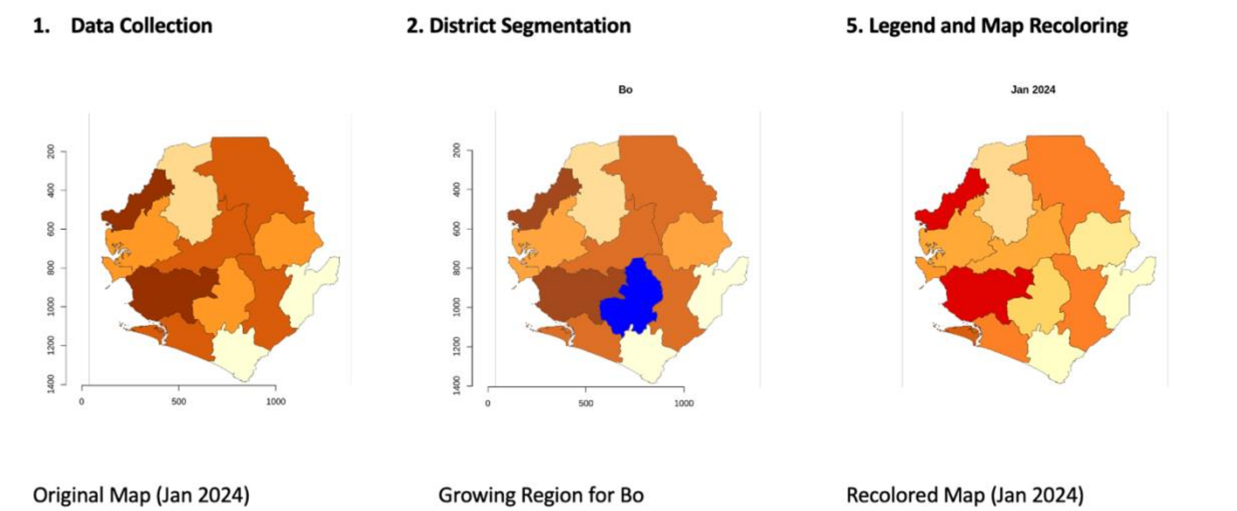

Figure 14 Comprehensive maps

7. Data Quality Assessment and Validation Rules

Question 10 examined validation mechanisms for ensuring data integrity and identified quality challenges within the dataset.

7.1 Validation Rules

Three validation checks were applied to verify data accuracy:

Birth Recording Consistency Check

The first validation verified that when live births and stillbirths are combined, they match the total birth count. Any discrepancy between these values suggests either data entry mistakes or incomplete facility records. This check matters because birth totals form the baseline for calculating numerous maternal and child health metrics throughout the project.

Vaccine Inventory Logic Check

The second validation confirmed that facilities could not report administering more vaccine doses than their available inventory. When dispensed amounts exceed stock levels, it reveals problems in supply tracking, documentation errors, or timing mismatches in reporting cycles. Analysis showed some facilities documented substantial inventory yet recorded zero immunizations, while others exhibited discrepancies between stock records and service provision.

Age-Group Population Logic Check

The third validation tested whether the count of infants (under age 1) remained below the count of young children (under age 5). This demographic relationship must always hold. Violations point to flawed population data, incorrect service area definitions, or faulty denominator calculations. Such errors produce nonsensical coverage rates above 100% or unexplained month-to-month swings.

7.2 Data Quality Issues

Analysis uncovered five significant data quality challenges:

Vaccination Reporting Gaps for Infants

Numerous facilities documented having vaccine supplies without recording any doses given to children under one year old. This pattern artificially deflated coverage calculations at both district and chiefdom levels, particularly skewing measles immunization metrics.

Unreliable Catchment Population Data

Certain facilities showed implausibly small population figures (sometimes below one person), generating coverage percentages well above 100% or causing dramatic monthly variations. These denominator problems undermined the accuracy of all coverage-based indicators.

Service Delivery Without Data Submission

Geographic areas without reported data, displayed as black regions on maps, predominantly appeared in isolated locations. These reporting voids created substantial information gaps, especially evident when examining geographic variation in service performance.

Anomalous High-Volume Reports

Select facilities, primarily in urban settings, reported dose counts far exceeding reasonable expectations for their service areas, suggesting either duplicate submissions or selection of incorrect data fields during reporting.

Geographic Imbalance in Data Completeness

Urban health centers reliably submitted comprehensive datasets, whereas facilities in rural, coastal, and geographically challenging chiefdoms exhibited persistent reporting deficiencies. This systematic geographic pattern in data availability compromised the validity of urban-rural comparisons and potentially concealed true access barriers in remote region

8. Conclusion

The project and reflection of the project helped my thinking evolve to find nuances that I could not previously notice. This analysis of maternal and child health indicators in Sierra Leone using DHIS2 data revealed complex patterns in service delivery, geographic disparities, and systemic data quality challenges that collectively inform health policy priorities.

NOTE: For question 9, a dashboard was created and not attached here as the quality of the whole image was being corrupted.

References

[1] Q. Sserwanja, L. M. Mutisya, L. Nuwabaine, K. Kamara, R. K. Mutebi, and M. W. Musaba, "Continuum of maternal and newborn health in Sierra Leone: a 2019 national survey," *Arch Public Health*, vol. 80, no. 1, p. 186, Aug. 2022, doi: 10.1186/s13690-022-00946-8.

[2] L. P. Pulavarthy, E. Babb, R. Prater, H. R. Karingula, and Y. Liu, "Maternal and Child Health Indicators in Sierra Leone: A Multi-Level Geospatial and Predictive Analysis," Luddy School of Informatics, Computing and Engineering, Dec. 11, 2025.

[3] E. Roman *et al.*, "Determinants of uptake of intermittent preventive treatment during pregnancy: a review," *Malar J*, vol. 18, no. 1, p. 372, Dec. 2019, doi: 10.1186/s12936-019-3004-7.

[4] R. González *et al.*, "The impact of community delivery of intermittent preventive treatment of malaria in pregnancy on its coverage in four sub-Saharan African countries (Democratic Republic of the Congo, Madagascar, Mozambique, and Nigeria): a quasi-experimental multicentre evaluation," *The Lancet Global Health*, vol. 11, no. 4, pp. e566–e574, Apr. 2023, doi: 10.1016/S2214-109X(23)00051-7.

[5] B. A. Ayele, E. Holliday, and C. Chojenta, "Determinants of antenatal care service utilisation in sub-Saharan Africa: an analysis of demographic and health surveys data (2015–2022)," *Arch Public Health*, vol. 83, no. 1, p. 189, July 2025, doi: 10.1186/s13690-025-01608-1.

[6] S. Purkayastha, "INFO-B585 Fall 2025 Lecture 5 Biomedical Analytics," Luddy School of Informatics, Computing and Engineering, Sept. 25, 2025.

[7] S. Purkayastha, "INFO-B585 FALL 2025 Lecture 8 Biomedical Analytics," Oct. 16, 2025.

[8] B. G. Masresha, M. E. Shibeshi, G. B. Grant, C. Hatcher, and C. S. Wiysonge, "Progress with the Second Dose Measles Vaccine Introduction and Coverage in the WHO African Region," *Vaccines*, vol. 12, no. 9, p. 1069, Sept. 2024, doi: 10.3390/vaccines12091069.

[9] Q. Sserwanja, K. Kamara, L. M. Mutisya, M. W. Musaba, and S. Ziaei, "Rural and Urban Correlates of Stunting Among Under-Five Children in Sierra Leone: A 2019 Nationwide Cross-Sectional Survey," *Nutr Metab Insights*, vol. 14, p. 11786388211047056, Jan. 2021, doi: 10.1177/11786388211047056.