

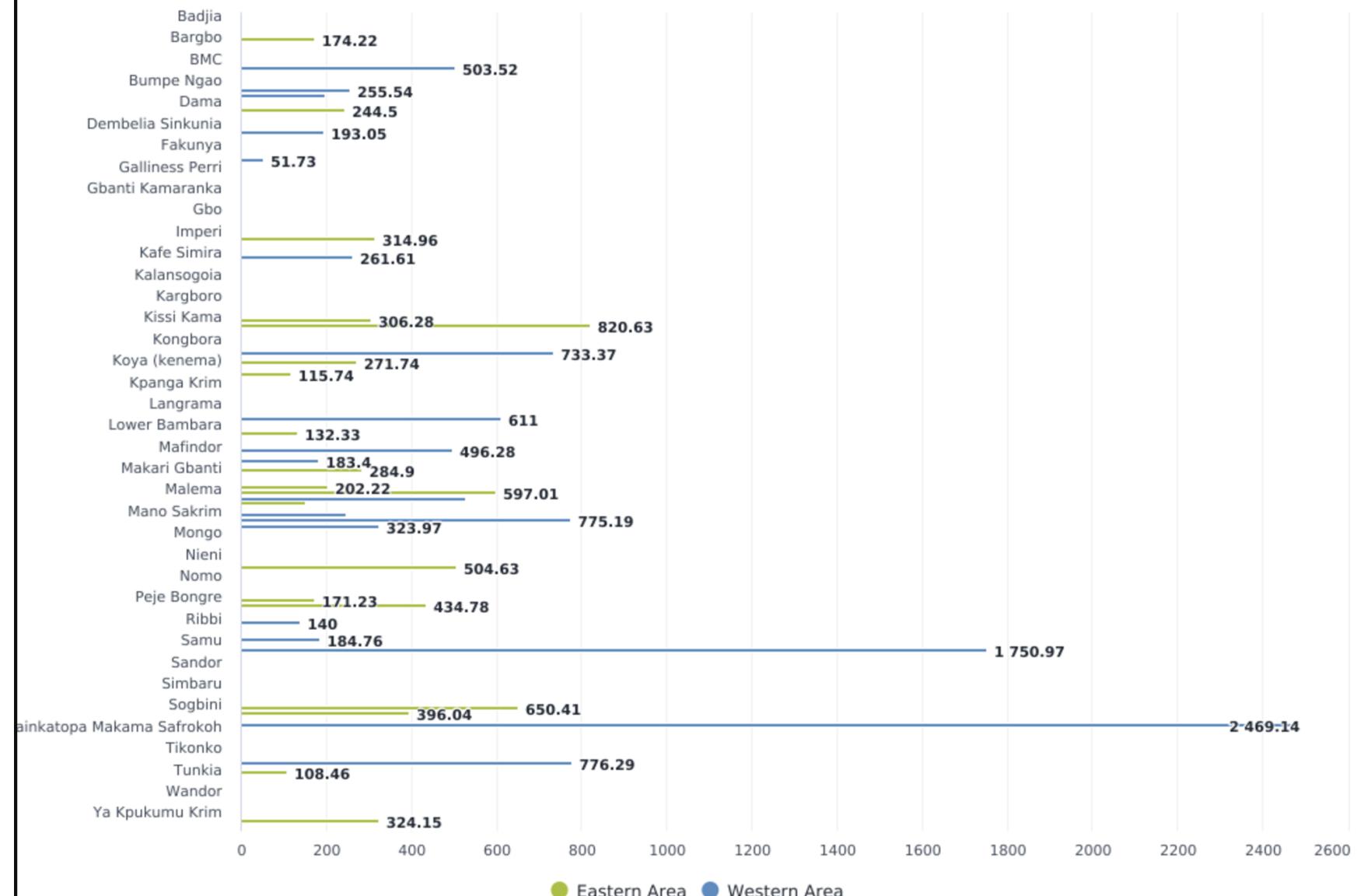
Group 4 DHIS2 Insights and Analysis

By: Harsh Patel, Kalp Patel, Deepishka
Pemmasani, Surya Tejaswi Mallidi, Poorva
Reddy Vanga, Pallavi Vandana, and
Bhavana Bethi



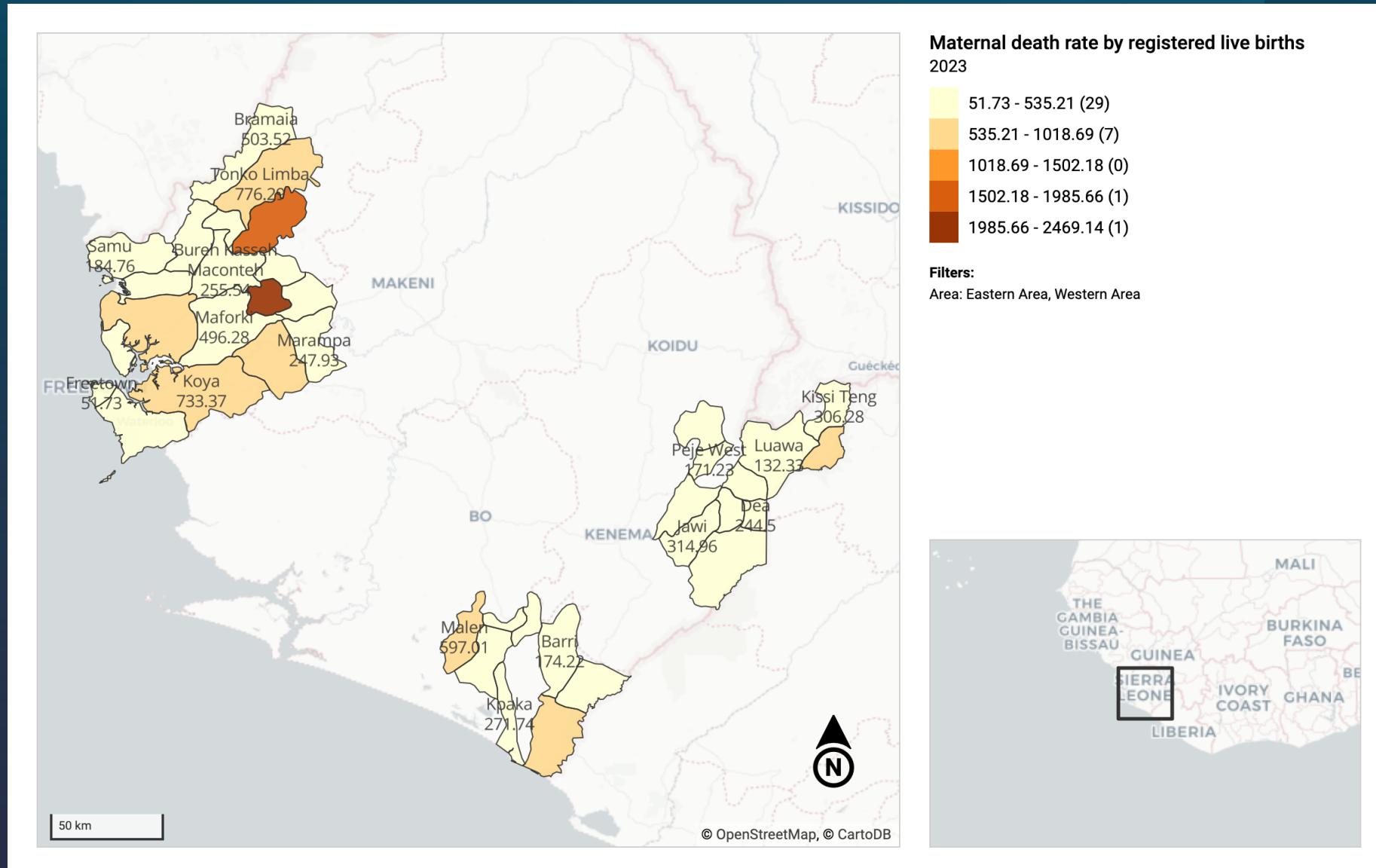
For 2023, Compare the Maternal death rate by registered live birth for the eastern-Chiefdoms to the western-Chiefdoms. Name the chiefdom with the highest maternal death rate. Name its surrounding chiefdoms with low maternal death rate.

2023 - Maternal death rate by registered live births



- Indicator : Maternal death rate by registered live birth
- Fixed period - yearly - 2023
- Sierra Leone - level = Chiefdom
- Eastern area and Western area
- Column graph
- Series = Area, Category = Organization, Filter = Period and data

Chiefdoms with the Highest Maternal Death Rate by live births is Tainkatopa Makama Safrokoh (Western Chiefdom).
Value = 2469.14



Analysis

Within Western
Chiefdoms, Tainkatopa
Makama Safrokoh has
the highest Maternal
Death Rate with value =
2469.14

1. Bureh Kasseh
Maconteh: 255.54 per
registered live birth

2. Maforki: 496.28 per
registered live birth

3. Buya Romende:
198.41 per registered
live birth

4. Dibia: 193. 05
registered live birth

5. Marampa: 247.93 per
registered live birth

- For the 2023 and 2024 monthly data for the above indicator, using 3-month and 5-month moving averages, can you predict the values of the indicator in January 2025 for the chiefdom with the highest maternal death rate? Which is the better of the two models when compared to the actual values of 2024?

After following the steps mentioned in slide 1, modify period to include monthly data for 2024 and add period to rows and select Pivot table and download CSV.

In R, **select organisationunitname =**
Tainkatopa Makama Safrokoh

Calculate 3-month and 5-month moving averages on interpolated values

Missing values imputed with **linear interpolation**.

Calculate MAE and RMSE for both models

Predict values for January 2025.

Describe the indicators (with all the internal details) that were used to answer Q1.

Indicator name: Maternal death rate by registered live births

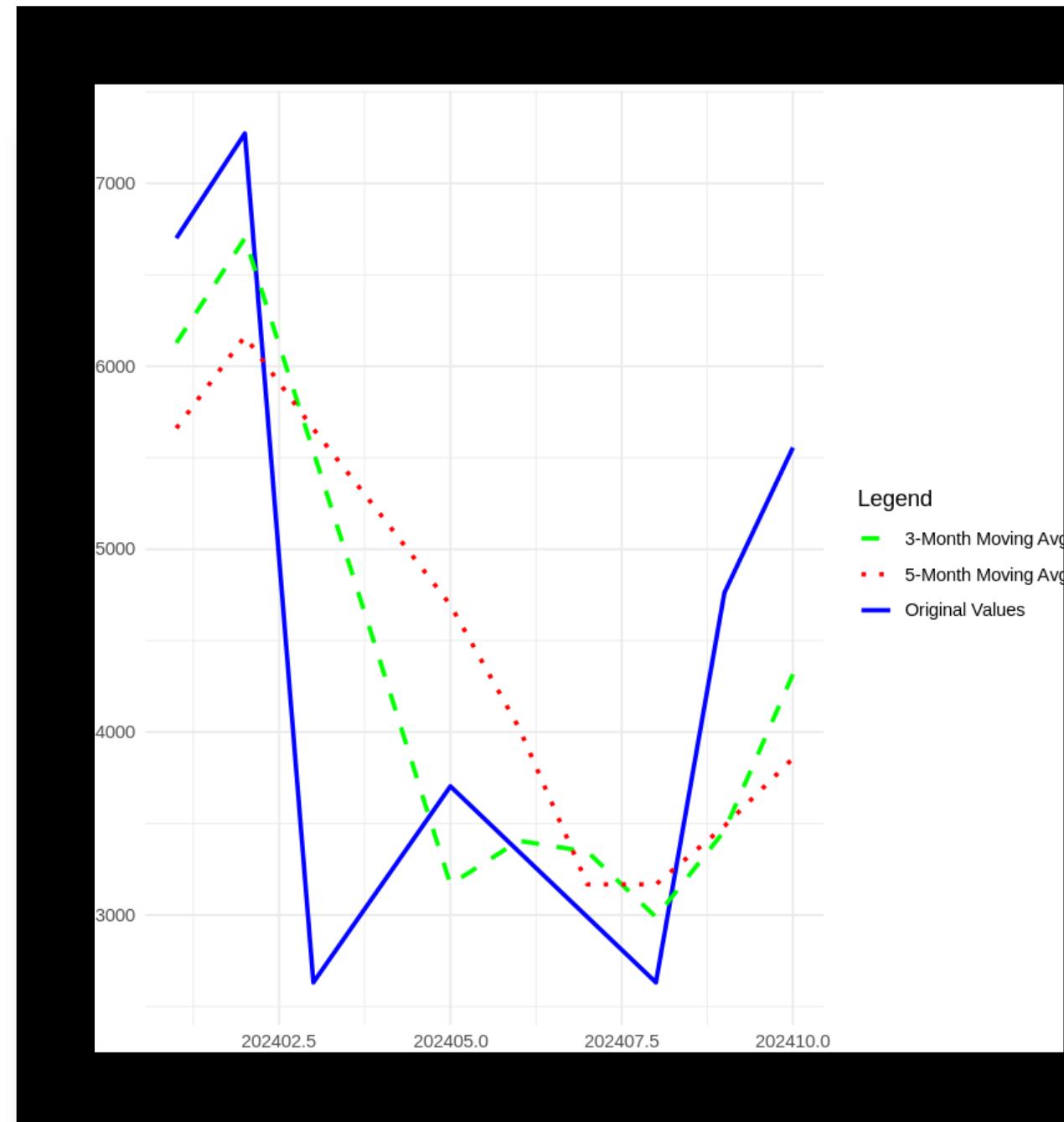
- Description: Ratio of Maternal deaths by 100000 registered live births.
- Numerator: Maternal Death
 - Description: The annual number of female deaths from any cause related to or aggravated by pregnancy or its management (excluding accidental or incidental causes) during pregnancy and childbirth or within 42 days of termination of pregnancy, irrespective of the duration and site of the pregnancy.
 - Denominator: Registered live births
 - Description: The complete expulsion or extraction from its mother a product of human conception, irrespective of the duration of pregnancy, that, after such expulsion or extraction, breathes or shows any other evidence of life such as beating of the heart, pulsation of the umbilical cord or definite.
 - Indicator type: Per hundred thousand

Interpretation

MAE_3M	RMSE_3M	MAE_5M	RMSE_5M
1151.571	1437.611	1439.564	1641.697

Predicted values for
January 2025

3-Month Moving Average	4316.347
5-Month Moving Average	3856.864



- What can you say about the BCG coverage rate of children less than 1yr around the beaches of the capital city of Sierra Leone based on data for the last 18 months? Present a tabular representation of this data for all the health centers in the western area.

Columns – (Period): April 2023 – December 2023 (9 months) + January 2024 – September 2024 (9 months). Which makes the total of 18 months.

Filter – (Data): BCG Coverage <1 year.

Rows – (Organisation Unit): Selected 31 health facilities in Freetown (Western Area) by facility level and Western Area group.

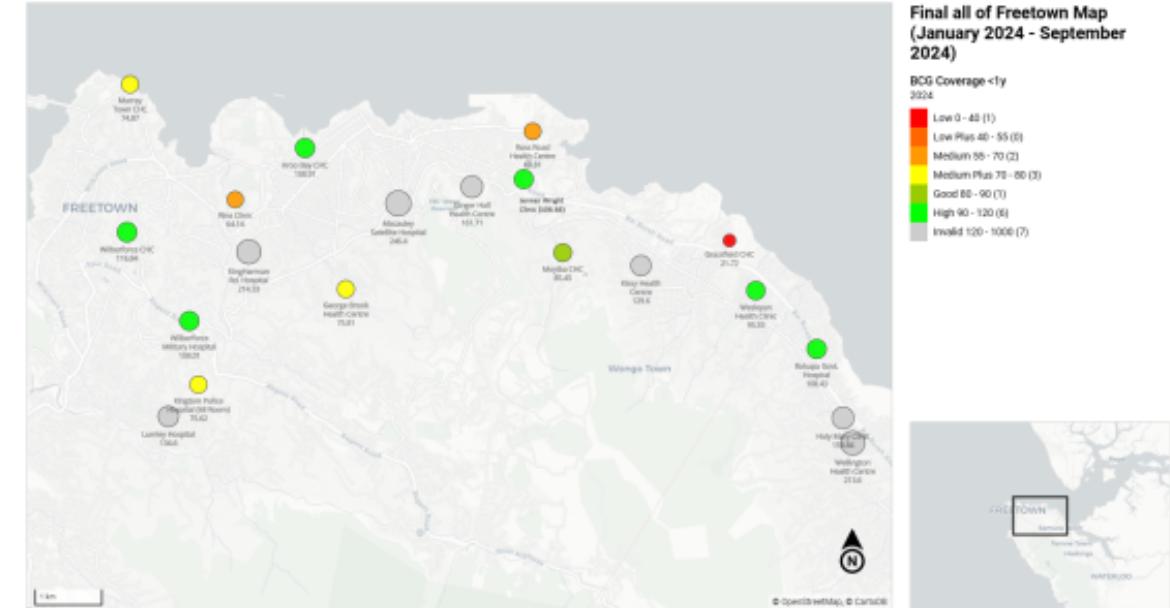
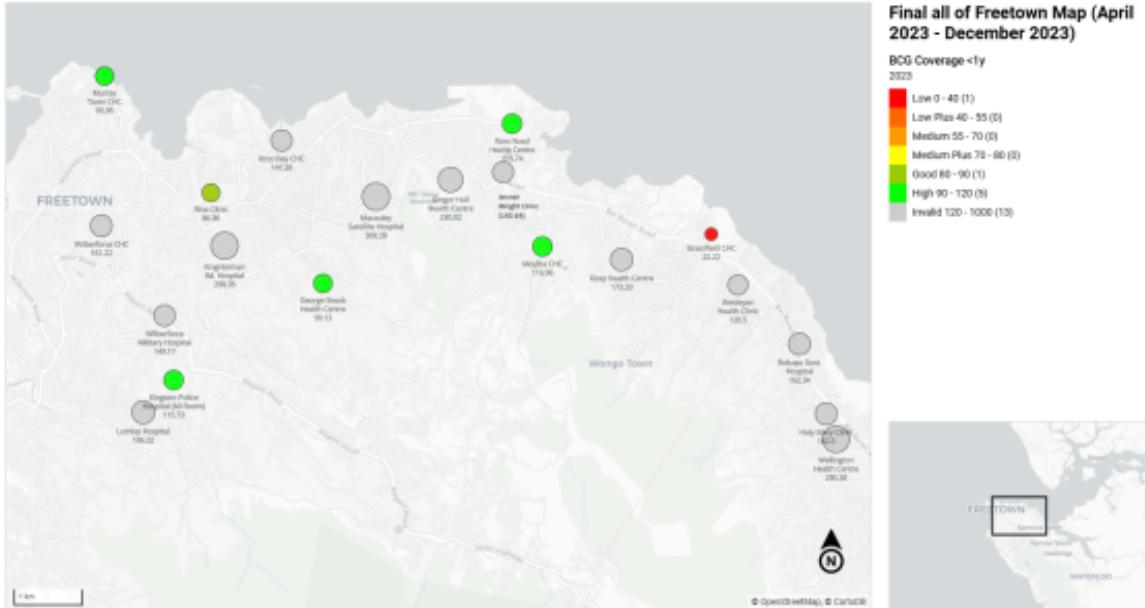
Exclusions: Removed health centers with no data; excluded November & December 2024 due to missing data.

Outcome: Generated a pivot table showing BCG coverage for children under 1 year in Freetown's beaches over the past 18 months.

Group 4 - All of Freetown (April 2023 - September 2024)

	BCG Coverage <1y																		
	April 2023	May 2023	June 2023	July 2023	August 2023	September 2023	October 2023	November 2023	December 2023	January 2024	February 2024	March 2024	April 2024	May 2024	June 2024	July 2024	August 2024	September 2024	
Blessed Mokaba clinic	258.03	236.57	221.82	168.66		201.45	197.14	217.29		269.74	292.92	244.05		231.2	216.79	164.84		196.2	
Calaba town CHC	133.9	178.54	215.73	107.99	110.87	26.78	158.38	187.46	51.83	153.52	144.53	126.76	174.64	174.64	211.02	105.63	108.45	26.2	
Family Clinic	122	55.56	136.35	131.95	111.12	136.35	131.95	129.17	118.06	108.89	138.23	115.7		54.45	133.62	129.31	108.89	133.3	
George Brook Health Centre	97.23	106.98	99.89	100.53	97.96	79.91	116	130.52	12.89	134.89	101.07	92.03	108.12	104.63	97.7	98.33	95.81	78.2	
Ginger Hall Health Centre	204.4	160.72	163.95	284.35	241.08	289.57	280.23	289.57	203.99	195.56	252.16	193.55	189.58	157.26	160.42	278.23	235.89	283.3	
Grassfield CHC	117.07										64.12	87.23	110.75						
Holy Mary Clinic	305.01	295.17	254.18		131.19	152.51					290.32	258.62	290.32	266.67	290.32	250		129.03	1
Jenner Wright Clinic	144.38	136.23	180.47	139.72	127.5	144.38	171.16	135.36	104.79	136.51	145.93	136.51	153.41	133.1	176.33	136.51	124.57	141.1	
KingHarman Rd. Hospital	301.23	315.81	291.19	291.51	291.51	366.5	242.93	431.76	272.08	185.16	304.51	284.86	284.54	308.6	284.54	284.86	284.86	358.2	
Kington Police Hospital (MI Room)	219.32		137.07	132.65	145.92		145.92	137.07	172.45	168.2	124.48	207.02			133.7	129.39	142.33		
Kissy Health Centre	167.8	183.92	200.24	207.24	161.49	162.23	243.13	166.87	77.15	175.47	175.38	158.8	153.22	179.86	195.83	202.67	157.93	158.2	
Kroo Bay CHC	146.9	156.03	155.86	147.36	149.1	157.65	131.76	154.07	142.16	128.75	144.88	138.92	148.8	152.47	152.3	144	145.7	154.1	
Lumley Hospital	126.29	211.69	234.53	159.31	183.32	182.66	248.79	166.88	163.68	179.17	189.25	119.45	182.94	206.9	229.23	155.71	179.17	178.2	
Macauley Satellite Hospital	298.95	379.22	432.26	379.22	367.49	145.43	195.47	242.39	234.57	397.57	269.7	282.89	339.72	370.81	422.67	370.81	359.34	142.1	
Marie Stopes Clinic (Abedeen R)	332.73	288.97		294.47		174.89		247.41	213.29	290.65	277.62	314.88	268.36	282.58		287.96		171.1	
Moyiba CHC	163.63		144.38	93.15	121.09	161.22	139.72	120.32	93.15	127.53	97.38	154.86	141.2		141.2	91.1	118.43	157.1	
Murray Town CHC	81.1	92.13	130.46	119.43	92.13	130.46	81.9	95.2	54.6	116.74	96.27	76.72	58.59	90.06	127.53	116.74	90.06	127.1	
Ola During Clinic	285.72	437.44	640.99	699.31	566.18	533.65	582.27	668.2	430.12		357.86	270.39	453.85	427.76	626.81	683.85	553.66	521.1	
Philip Street Clinic	258.35	215.29	365.99	236.12	187.51	387.52	395.85	394.7		197.37	385.59	245.01		210.98	358.67	231.4	183.76	379.1	
Rina Clinic	46.21	107.33	55.45	80.5	80.5	78.56	89.44	92.42	67.08	118.18	121.65	43.77	94.98	105.05	54.28	78.79	78.79	76.1	
Rokupa Govt. Hospital	102.94	123.6	194.44	158.65	132.82	224.94	230.6	228.75	154.96	140.64	113.72	97.37	137.88	120.81	190.05	155.07	129.82	219.1	
Ross Road Health Centre	100.86	93.85	108.62	82.59	103.24	118.32	215.86	106.68	215.86	119.21	117.63	95.37		91.7	106.13	80.7	100.87	115.1	
SLIMS Clinic	240.7	239.32	224.22	207.41	204.22	237.4	232.94	207.73	232.94	240.48	260.4	227.99		234.23	219.45	203	199.88	232.1	
SLRCS (Freetown) Clinic	133.22	88.21	70.12	122.14	115.35	56.09	101.78	70.12	135.71	179.61	113.77	126.39	137.48	86.48	68.74	119.74	113.09	54.1	
St Anthony clinic	145.38	121.68	88.41	95.06	93.16	86.44	83.65	88.41	53.23	143.11	87.42	137.54		118.95	86.42	92.93	91.07	84.1	
Stella Maries Clinic	168.66	310.12	309.21	299.24	244.83	309.21	206.75	252.99	206.75	186.47	227.8	159.83	192.68	303.68	302.79	293.02	239.75	302.1	
UMC (Urban Centre) Hospital	89.14	40.37	45.71	47	37.05	29.71	60.83	58.29	52.53	70.82	95.93	84.33		39.46	44.69	45.95	36.22	29.1	
Wellington Health Centre	286.64	329.4	322.47	242.72	173.37	250.81	225.38	250.81	381.41	288.21	376.05	271.26	350.37	322.12	315.33	237.35	169.53	245.1	
Wesleyan Health Clinic	183	118.07	54.9	112.16	76.74	79.3		134.2	141.68	225.09	160.41	173.15	155.07	115.43	53.68	109.66	75.03	77.1	
Wilberforce CHC	129.89	179.57	139.16	170.59	179.57	139.16	134.68	185.55	179.57	175.73	140.89	123.01	172.51	175.73	136.19	166.95	175.73	136.1	

BCG Maps 2023 vs 2024



Challenge: The pivot table alone didn't provide a clear visual of which health centers around the beaches were affected.

Solution: Created 2 maps using the pivot table data to visualize BCG coverage for children under 1 year.

Map 1: BCG coverage for April 2023 – December 2023 (9 months)

Map 2: BCG coverage for January 2024 – September 2024 (9 months)

Note: DHIS2 did not allow for specific start and end dates due to unknown data recording days. Hence the reasoning for creating 2 maps.

BCG Coverage Rate (2023 vs 2024):
BCG coverage for children under 1 year old at Western Area health centers decreased between 2023 and 2024. The largest decreases were observed at Wilberforce CHC, Kissy Health Centre, and Rokupa Govt. Hospital.

Data Quality Improvement: The 2023 map showed a large "Invalid 120-1000" category, indicating unreliable data. In 2024, more detailed coverage categories were used, suggesting improved data collection and measurement processes between the two years. This improved granularity and reliability of the 2024 data compared to 2023 makes it a more valuable source of information for understanding the vaccination coverage trends in the region.

- Give the underlying data elements/categories/category combinations that were used for analysis in Q.3. Perform correlation analysis using values of the data elements for all the health centers in the Western area, and interpret the analysis.

Secondary Pivot Table

- Expansion:** Created a secondary pivot table with the same dimensions, but included additional data from the Rural Western Area.
- Selected Health Centers:** Added 10 health centers from the Rural Western Area: Goderitch Health Centre, Hastings Health Centre, Lakka/Ogoo Farm CHC, Mutual Faith Clinic, Newtown CHC, Regent (RWA) CHC, Songo CHC, Tombo CHC, Waterloo CHC, and York CHC, all with data from April 2023 – September 2024.
- Purpose:** This final pivot table was used for analysis in question 4. Because this data includes all of the health centers in the western area.

Columns Rows Filter

Group 4 - (All of the Western Area) BCG Coverage <1yr - Final Table

	BCG Coverage <1y																		
	April 2023	May 2023	June 2023	July 2023	August 2023	September 2023	October 2023	November 2023	December 2023	January 2024	February 2024	March 2024	April 2024	May 2024	June 2024	July 2024	August 2024	September 2024	
Blessed Mokaba clinic	258.03	236.57	221.82	168.66		201.45	197.14	217.29		269.74	292.92	244.05		231.2	216.79	164.84		196.88	
Calabia town CHC	133.9	178.54	215.73	107.99	110.87	26.78	158.38	187.46	51.83	153.52	144.53	126.76	174.64	174.64	211.02	105.63	108.45	26.2	
Family Clinic	122	55.56	136.35	131.95	111.12	136.35	131.95	129.17	118.06	108.89	138.23	115.7	54.45	133.62	129.31	108.89	133.62	78.16	
George Brook Health Centre	97.23	106.98	99.89	100.53	97.96	79.91	116	130.52	128.89	134.89	101.07	92.03	108.12	104.63	97.7	98.33	95.81		
Ginger Hall Health Centre	204.4	160.72	163.95	284.35	241.08	289.57	280.23	289.57	195.99	195.56	252.16	193.55	189.58	157.26	160.42	278.23	235.89	283.33	
Goderich Health Centre	110.91	210.45	73.94	107.33	151.53	173.97	96.81	139.18	71.55	214.08	140.82	104.98	148.89	205.84	72.32	104.98	148.21	170.16	
Hastings Health Centre	130.78	158.2	197.89	316.39	149.87	172.07	99.91	132.5	149.87	154.71	135.78	123.77	159.87	154.71	193.51	309.42	146.57	168.28	
Holy Mary Clinic	305.01	295.17	254.18		131.19	152.51				290.32	258.62	290.32	266.67	290.32	250		129.03	150	
Jenner Wright Clinic	144.38	136.23	180.47	139.72	127.5	144.38	171.16	135.36	104.79	136.51	145.93	136.51	153.41	133.1	176.33	136.51	124.57	141.06	
KingHarman Rd. Hospital	301.23	315.81	291.19	291.51	291.51	366.5	242.93	431.76	272.08	185.16	304.51	284.86	284.54	308.6	284.54	284.86	284.86	358.13	
Kington Police Hospital (Mi Room)	219.32		137.07	132.65	145.92	145.92	137.07	172.45	168.2	124.48	207.02				133.7	129.39	142.33		
Kissy Health Centre	167.8	183.92	200.24	207.24	161.49	162.23	243.13	166.87	77.15	175.47	175.38	158.8	153.22	179.86	195.83	202.67	157.93	158.66	
Kroo Bay CHC	146.9	156.03	155.86	147.36	149.1	157.65	131.76	154.07	142.16	128.75	144.88	138.92	148.8	152.47	152.3	144	145.7	154.05	
Lakka/Ogoo Farm CHC	122.75	151.18	137.62	107.99	97.19	104.15	107.99	66.95	122.39	147.62	142.77	115.98	152.54	147.62	134.38	105.44	94.9	101.69	
Lumley Hospital	126.29	211.69	234.53	159.31	183.32	182.66	248.79	166.88	163.68	179.17	189.25	119.45	182.94	206.9	229.23	155.71	179.17	178.53	
Macauley Satellite Hospital	298.95	379.22	432.26	379.22	367.49	145.43	195.47	242.39	234.57	397.57	269.7	282.89	339.72	370.81	422.67	370.81	359.34	142.21	
Marie Stoops Clinic (Abedeen R)	332.73	288.97		294.47		174.89		247.41	213.29	290.65	277.62	314.88	268.36	282.58		287.96	171.03		
Moyiba CHC	163.63	144.38	93.15	121.09		161.22	139.72	120.32	93.15	127.3	97.38	154.86	141.2		141.2	91.1	118.43	157.67	
Murray Town CHC	81.1	92.13	130.46	119.43	92.13	130.46	81.9	95.2	54.6	116.74	96.27	76.72	58.59	90.06	127.53	116.74	90.06	127.53	
Mutual Faith Clinic	228.03	331.01	456.06	165.51	154.47	85.51	148.96	114.02	137.92	216.04	173.2	216.04		324.06	446.48	162.03	151.23	83.72	
Newton CHC	85.37	92.94	156.5	89.5	79.17	99.59	92.94	106.71	96.38	114.38	100.69	80.74	107.76	90.83	152.95	87.47	77.37	97.33	
Ola During Clinic	285.72	437.44	640.99	699.31	566.18	533.65	582.27	668.2	430.12		357.86	270.39	453.85	427.76	626.81	683.85	553.66	521.05	
Philip Street Clinic	258.35	215.29	365.99	236.12	187.51	387.52	395.85	394.7		197.37	385.59	245.01		210.98	358.67	231.4	183.76	379.77	
Regent (RWA) CHC	88.19	135.13	146.98	106.68	192.03	124.94	106.68	95.54	113.79	97.54	141.5	83.6	136.79	132.37	143.98	104.5	188.11	122.39	
Rina Clinic	46.21	107.33	55.45	80.5	80.5	78.56	89.44	92.42	67.08	118.18	121.65	43.77	94.98	105.05	54.28	78.79	78.79	76.89	
Rokupa Govt. Hospital	102.94	123.6	194.44	158.65	132.82	224.94	230.6		228.75	154.96	140.64	113.72	97.37	137.88	120.81	190.05	155.07	129.82	
Ross Road Health Center	100.86	93.85	108.62	82.59	103.24	118.32	215.86	106.68	215.86	119.21	117.63	95.37		91.7	106.13	80.7	100.87	115.5	
SLIMS Clinic	240.7	239.32	224.22	207.41	204.22	237.4	232.94	207.73	232.94	240.48	260.4	227.99		234.23	219.45	203	199.88	232.36	
SLRCS (Freetown) Clinic	133.22	88.21	70.12	122.14	115.35	56.09	101.78	70.12	135.71	179.61	113.77	126.39	137.48	86.48	68.74	119.74	113.09	54.99	
Songo CHC	192.13	99.16	256.17	278.89	170.43	128.08	120.85	131.29	120.85	254.25	181.19	181.61	312.77	96.86	250.21	272.41	166.47	125.11	

Data Preparation and Analysis

- **Data Source:** Downloaded “BCG Coverage All of Western Area.csv” for analysis, including both Freetown and Rural Western Area health centers.
- **Data Preparation:** Handled missing values through linear interpolation. Estimating the missing values between known data points.
 - **Normality Check:** Used Shapiro-Wilk test; data did not follow a normal distribution.
 - **Linearity Check:** Performed visual inspection of linearity. It wasn't normally distributed.
 - **Correlation Method:** Recommended Spearman's rank correlation due to non-normal distribution and potential outliers. Spearman's is robust to outliers, as it relies on rank order rather than raw values, ensuring accurate insights despite extreme values.
- **Outliers:** Retained in the analysis as Spearman's correlation is less affected by outliers and reflects overall trends. Even with presence of outliers, Spearman's correlation can still provide reliable insights into the general relationship between variables.

determine_correlation_method(data)			
NA	5.087890e-03	2.070461e-02	
X23.Jun	X23.Jul	X23.Aug	
3.482289e-04	1.674094e-06	3.820334e-06	
X23.Sep	X23.Oct	X23.Nov	
2.901991e-04	3.462827e-06	8.491564e-07	
X23.Dec	X24.Jan	X24.Feb	
2.492212e-03	3.413077e-03	9.677950e-05	
X24.Mar	X24.Apr	X24.May	
5.069856e-03	9.054834e-04	2.036112e-02	
X24.Jun	X24.Jul	X24.Aug	
3.497736e-04	1.684687e-06	3.890847e-06	
X24.Sep			
2.892824e-04			
outliers detected using IQR method:			
organisationunitname	X23.Apr	X23.May	
NA	0	1	
X23.Jun	X23.Jul	X23.Aug	
3	1	2	
X23.Sep	X23.Oct	X23.Nov	
3	2	3	
X23.Dec	X24.Jan	X24.Feb	
2	1	0	
X24.Mar	X24.Apr	X24.May	
0	1	1	
X24.Jun	X24.Jul	X24.Aug	
3	1	2	
X24.Sep			
3			

Visual inspection of linearity is being displayed...

Recommendation: Use Spearman's rank correlation, as the data may be non-nor

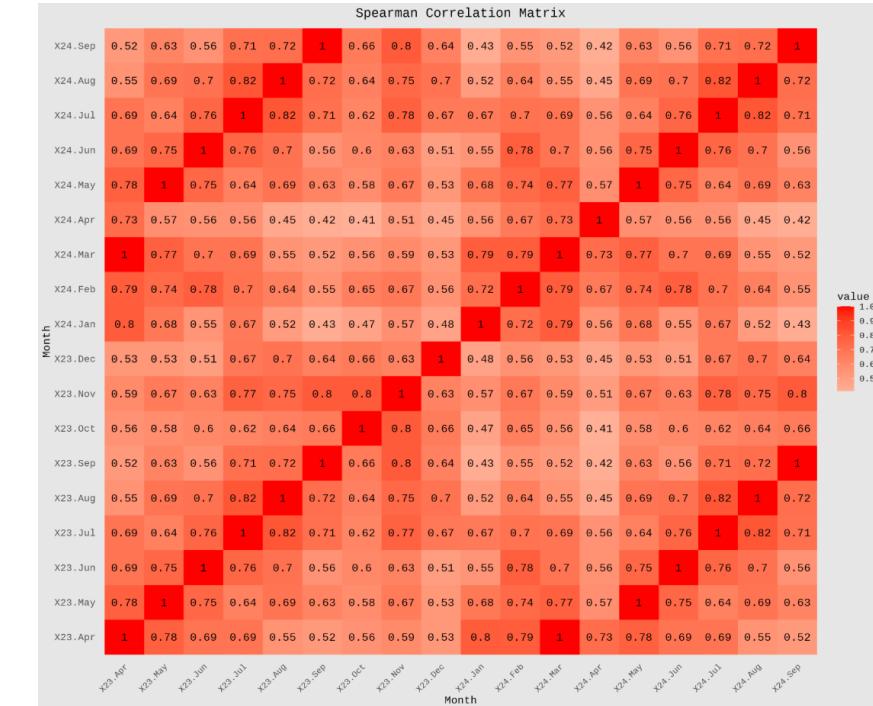
Data Analysis Summary

Correlation Matrix Insights:

- Strong positive correlations between most health centers indicate similar BCG coverage trends.
- Some negative correlations suggest inverse patterns between certain health centers.
- The Spearman p-value ($< 2.2e-16$) is highly significant, showing that the correlations are unlikely to occur by chance.
- Overall, BCG coverage in the Western Area health centers tends to move together, with underlying factors influencing the trends.

Spearman's Correlation Results:

- Strong positive correlation ($p = 0.956$) between 2023 and 2024 BCG coverage.
- Suggests consistent vaccination efforts across health centers in Western Area, with minimal variation over time.
- Statistically significant p-value supports the reliability of the findings.



Spearman's rank correlation rho

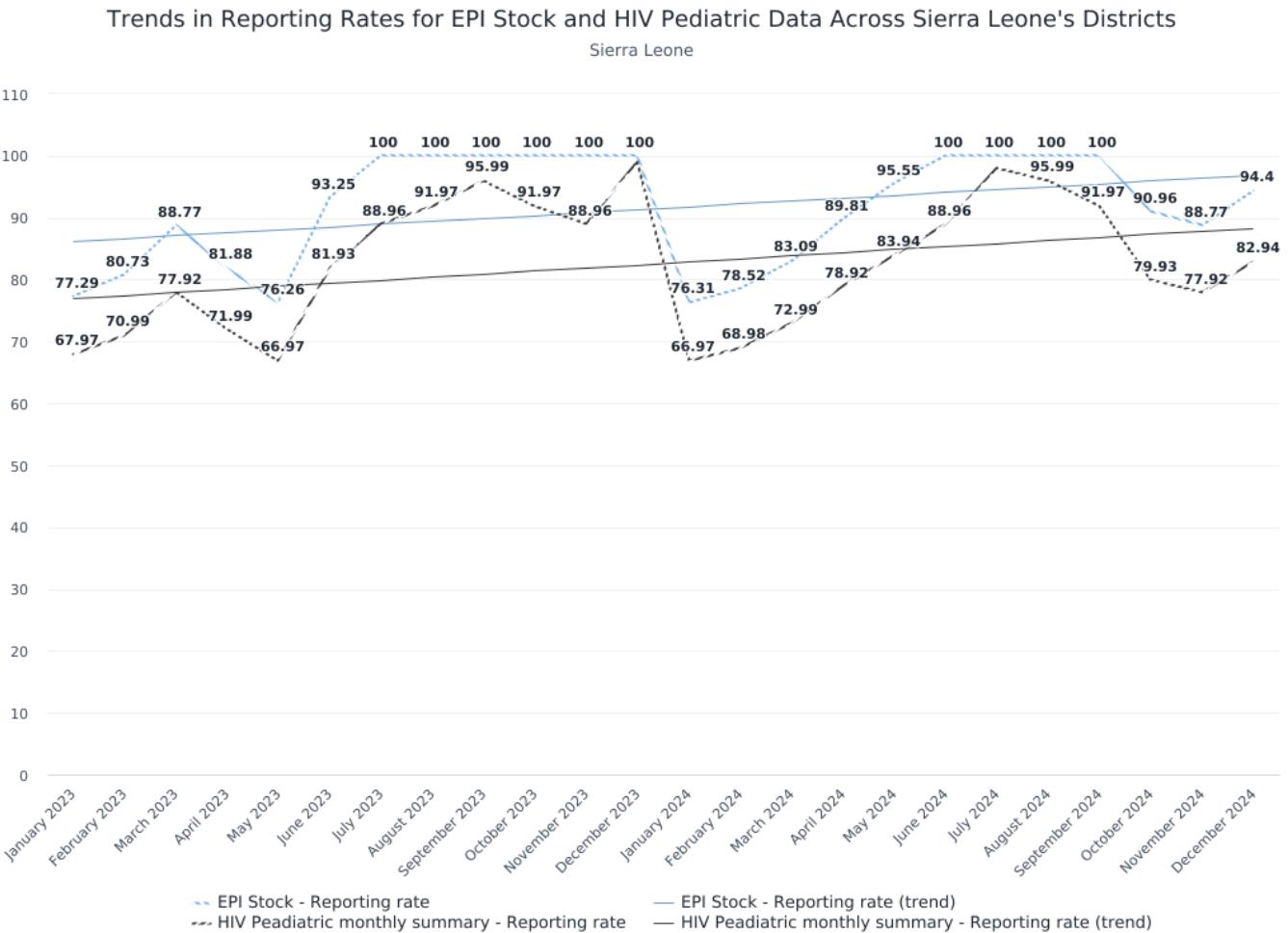
```
data: data_2023$total_2023 and data_2024$total_2024
S = 468, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.9560976

Spearman's Rank Correlation Interpretation:
Correlation coefficient (rho): 0.9560976
Interpretation of rho: Strong positive correlation
p-value: 0
The correlation is statistically significant (p-value is very small).
```

Describe the trends of Reporting rates for the following datasets in all the districts of Sierra Leone by months for available data on:

- EPI Stock
- HIV Pediatric monthly

Trends in Reporting Rates for EPI Stock and HIV Pediatric Data



EPI Stock - Reporting Rate

Upward trend throughout the period.

January 2023: 77.29%.

April to October 2023: Reached 100% consistently.

November 2023: Slight decline to 94.4% but remains high.

HIV Pediatric Monthly Summary - Reporting Rate

Started lower at 67.97% in January 2023.

Dipped to its lowest point at 66.97% in April 2023.

October 2023: Reached 100% peak.

November 2023: Declined slightly to 88.77%.

- Both datasets show positive trends over time.

Four components of time series of EPI-Reporting Rates

Trend

- Indicates long-term periodic growth and recovery.
- Upward movement in early 2023, peaking at 100% in mid-2023.
- Declines afterward, with partial recoveries in 2024.
- Upward trend seen

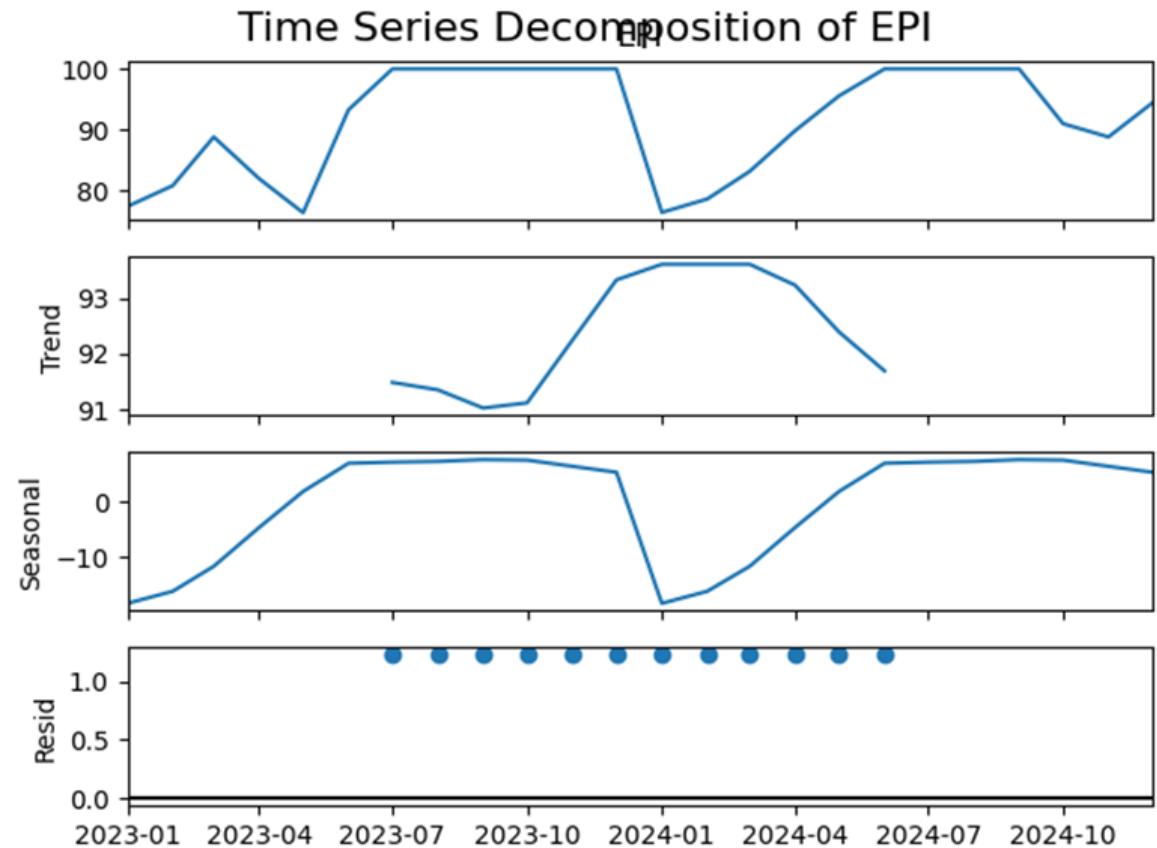
Seasonality

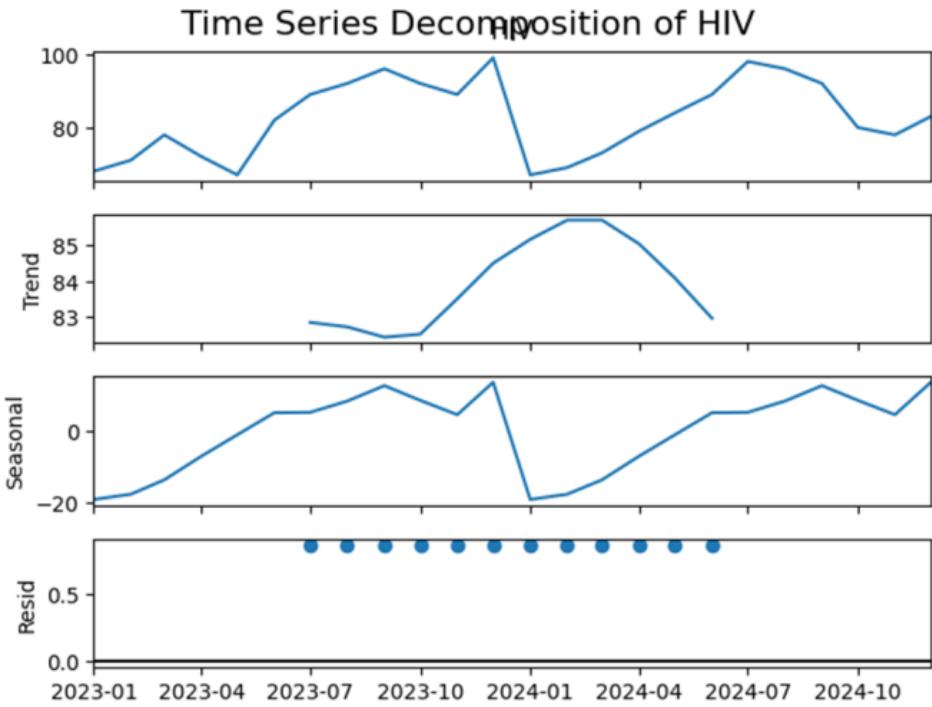
- Regular peaks and troughs suggest seasonal influence.
- High reporting rates in mid-year, lower rates toward year-end.

Cyclicity

Is not clearly distinguishable over this two-year period.

Noise: There is variability in the data, such as the dips in early 2023 and late 2024, indicates random fluctuations not explained by trend or cyclicity.





Four components of time series for HIV Pediatric Monthly :

Trend

- Long-term periodic growth and recovery observed.
- Peaks in mid-2023 and mid-2024 (~100%).
- Declines in late 2023 and late 2024, followed by partial recoveries.
- Overall upward trend

Seasonality

- Regular peaks and troughs suggest seasonal patterns.
- Peaks during mid-year; troughs at year-end.
- Likely influenced by calendar-driven factors like reporting cycles.

Cyclicity

- Cyclic patterns are absent in this data.

Noise

- Abrupt dips and recoveries, e.g., early 2024, indicate random fluctuations.

This analysis takes into account 24 months of data, which helps us understand seasonality in the two data elements.

MODEL

Data Preprocessing:

- Checked for missing values and renamed columns for consistency.
- Performed Augmented Dickey-Fuller (ADF) test: Data was non-stationary (failed to reject null hypothesis)



Model Selection:

- Narrowed down to **ARIMA**, **Holt-Winters**, and **ETS** due to clear seasonality observed in ACF plots and time decomposition.
- Ruling out **Auto ARIMA** and **Linear Holt's** models due to seasonality

Model Building:

- Split data: 80% training, 20% test.
- Built **SARIMA**, **ETS**, and **Holt-Winters** models.

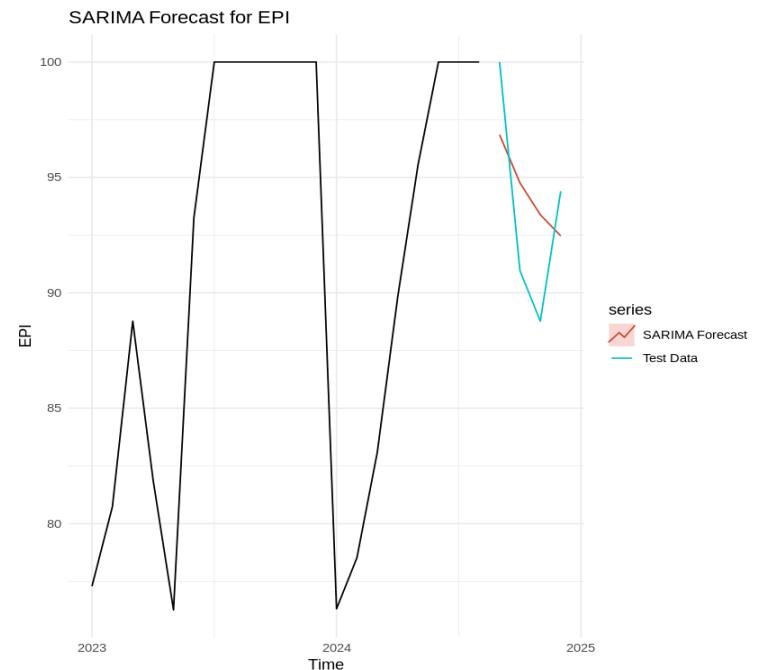
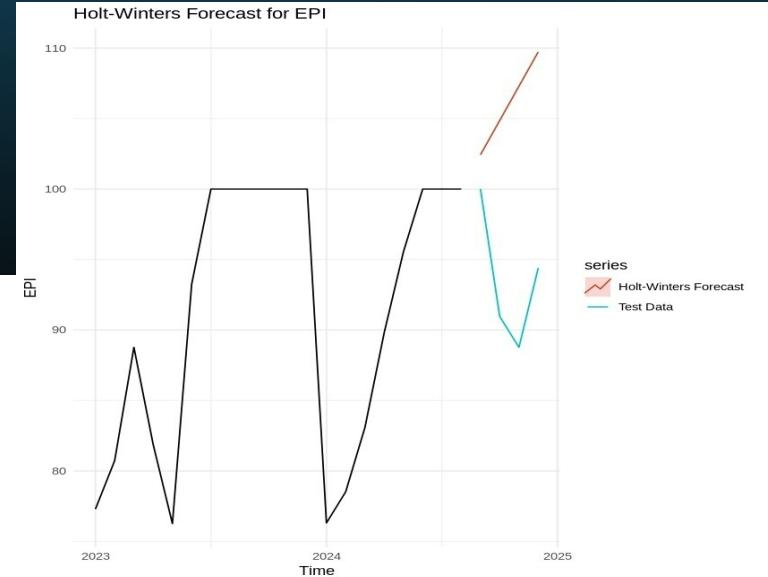


Evaluation:

- Compared models using RMSE, MAE, MAPE, and residuals (ACF).
- **SARIMA**: Best fit, handled differencing well, low ACF in residuals.
- **ETS**: Good metrics but constant point forecasts indicated poor performance for trend.
- **Holt-Winters**: Struggled with seasonality due to linear trend assumption and large errors.

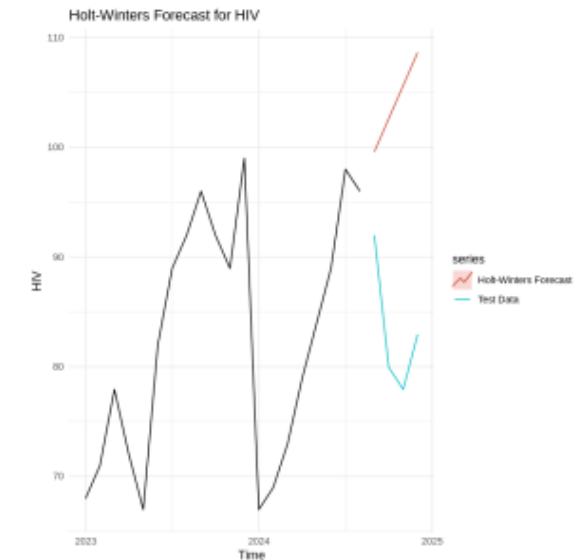
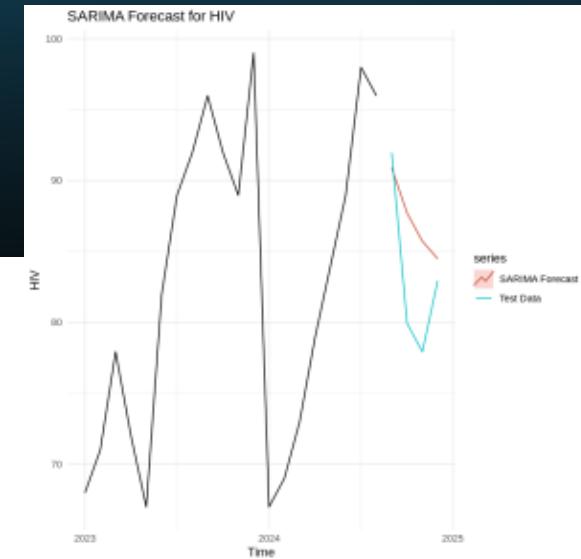
EPI STOCK REPORTING RATE MODEL

Metric	Holt Winters(Training)	Holt Winters(test)	SARIMA (Training)	SARIMA (Test)
ME	-1.8188	-12.5462	0.611	-0.8324
RMSE	7.8879	13.9396	7.2177	3.5139
MAE	5.2426	12.5462	5.6532	3.3746
MPE	-2.3975	-13.7048	-0.0069	-1.044
MAPE	6.1	13.7	6.5	3.64
MASE	0.979	2.3429	1.0557	0.6302
ACF1	0.0459	0.0746	0.1455	-0.2372
Theil's U	NA	2.6633	NA	0.5929



HIV Pediatric monthly

Metric	Holt-Winters (Training)	Holt-Winters (Test)	SARIMA (Training)	SARIMA (Test)
ME	-1.8495	-20.9295	0.6099	-4.0129
RMSE	9.3243	22.3674	8.7542	5.5897
MAE	6.0601	20.9295	6.6864	4.5488
MPE	-3.0751	-25.8045	-4.0541	-5.11
MAPE	7.8883%	25.8045%	8.4663%	5.6926%
MASE	0.9338	3.2249	1.0303	0.7009
ACF1	-0.0342	0.0842	-0.2303	-0.2356
Theil's U	NA	3.6363	NA	0.8827



Comparison and Best model:

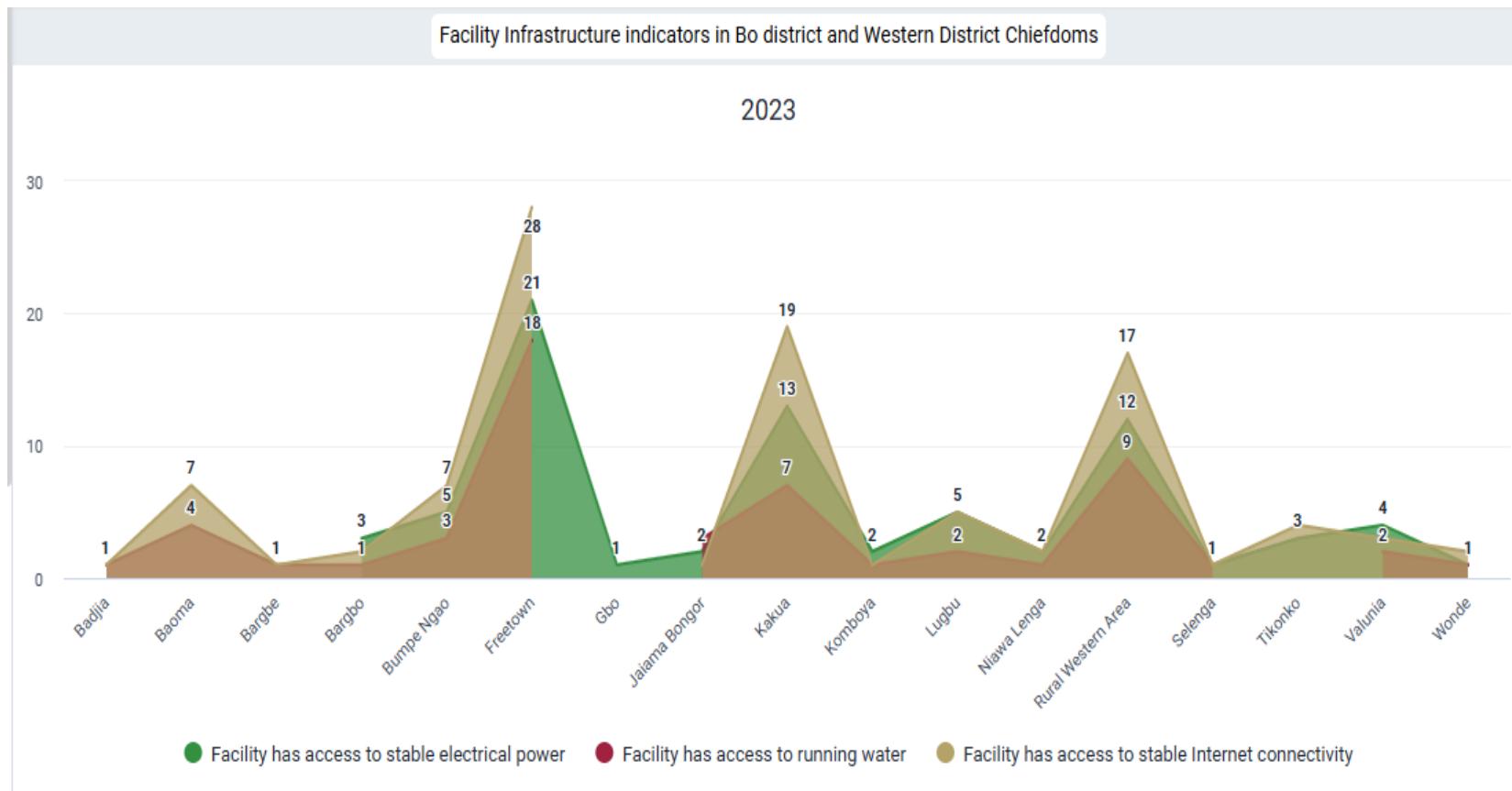
For EPI Stock reporting rates:

- **Accuracy:** The SARIMA model outperforms Holt-Winters in all key metrics for the test set, particularly in terms of RMSE, MAE, and MAPE.
- **Error Reduction:** SARIMA shows significantly lower errors, suggesting it provides more reliable forecasts for the EPI stock.

For HIV Pediatric monthly:

- **Accuracy:** The SARIMA model shows a clear advantage with lower RMSE, MAE, and MAPE values for the test set.
- **Error Reduction:** SARIMA's substantially lower errors suggest it is more accurate and reliable for HIV forecasting.

- Compare the 3 facility infrastructure indicators for the last year, between the chiefdoms inside Bo district and Western Area districts. Use an area chart to display the visualization. What are other better ways to visualize this and why?



Area chart comparing facility access to stable electrical power, running water, and stable Internet connectivity across chiefdoms of Bo district and Western Area district in 2023. Freetown has the highest access levels, while many regions have minimal or no access to these facilities.

Data: "Indicator"

"Public facilities with access to Internet connectivity"

- o "Public facilities with access to running water"

- o "Public facilities with access to stable electric power"

- Period:

- o Last year

- Organization Unit:

- o Chiefdoms of Bo and Western Area

- o Level: Chiefdoms

- o Group: District

- Visualization:

- o Area Chart

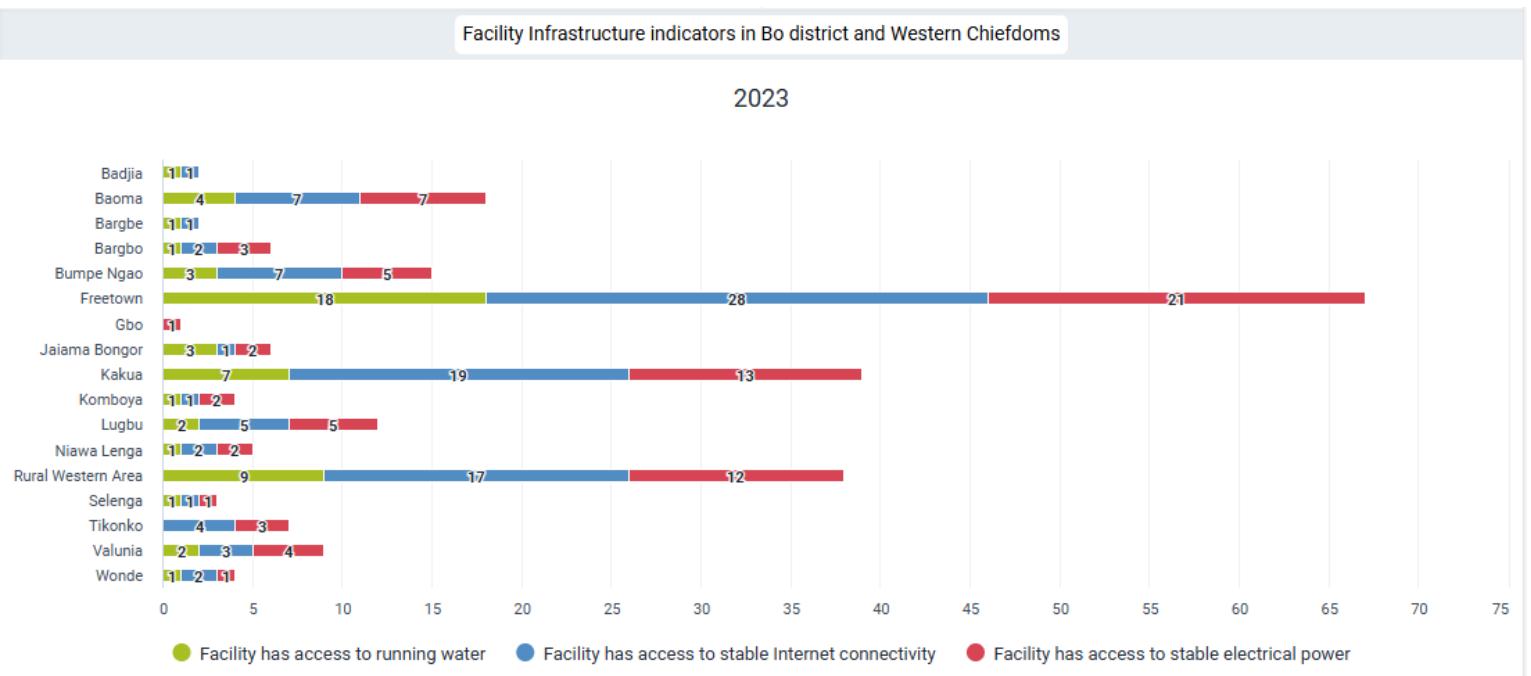
- Configure:

- o Series: Organizational Unit

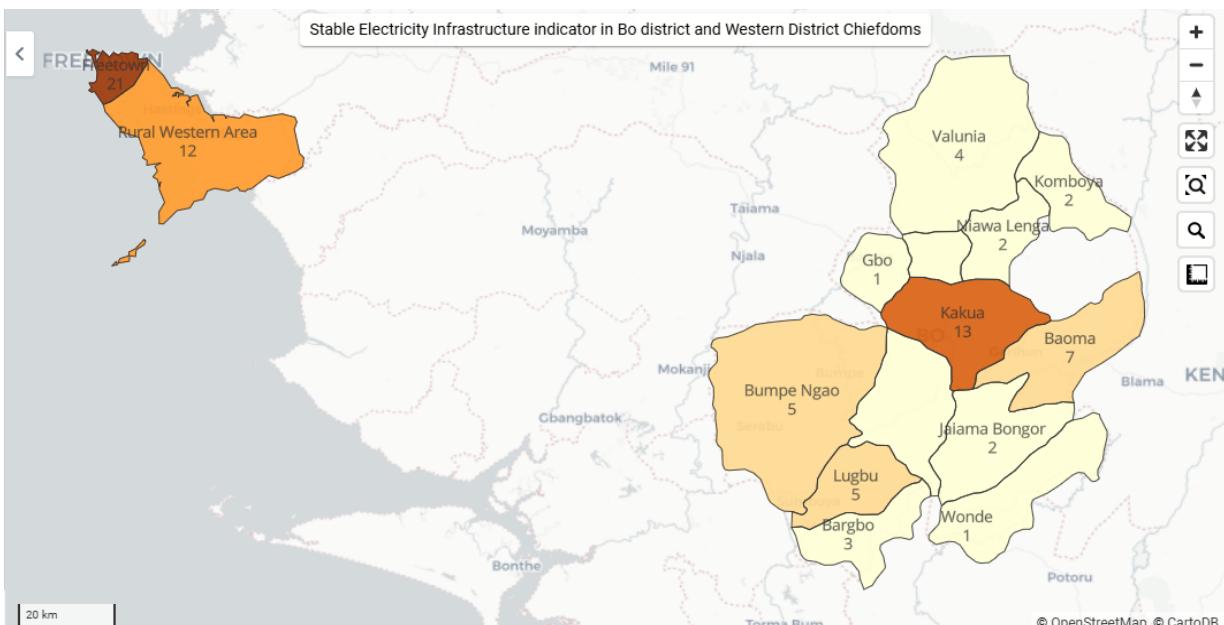
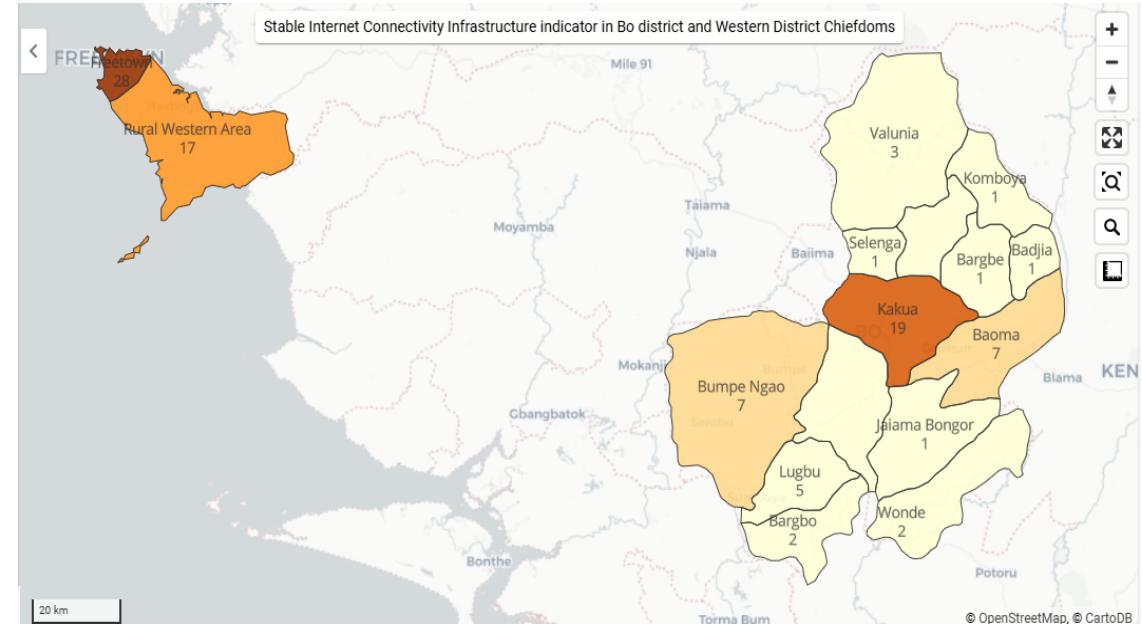
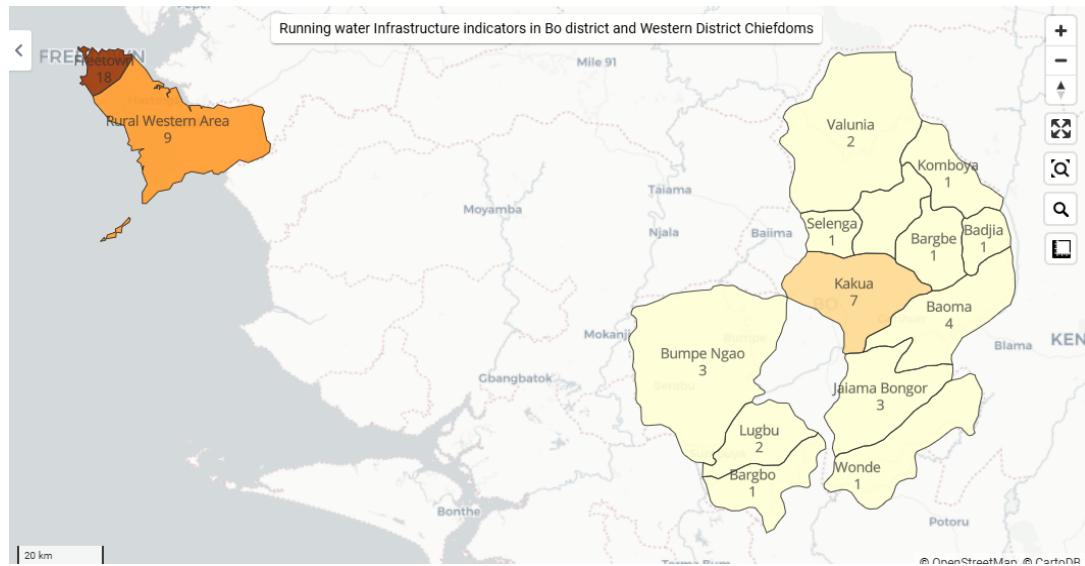
- o Category: Data

- o Filter: Period

Facility Infrastructure indicators in Bo district and Western Chiefdoms



- This horizontal stacked bar chart, illustrates the distribution of infrastructure access across various locations. Each bar is segmented into three categories:
 - **Facility access to running water** (green segment).
 - **Facility access to stable Internet connectivity** (blue segment).
 - **Facility access to stable electrical power** (red segment).
- Key Details:
 - **Y-Axis:** Lists the locations being analyzed (e.g., Badjia, Baoma, Freetown, etc.)
 - **X-Axis:** Represents the number of facilities with access to the mentioned infrastructure types



Three thematic maps, each representing a different facility infrastructure indicator in Bo and Western Area districts

- Access to Running Water:** Regions with darker shades have better access to running water, while lighter shades indicate limited access.
- Access to Stable Internet Connectivity :** Displays regional variations in Internet access, with darker regions showing higher connectivity levels.
- Access to Stable Electrical Power :** Highlights access to electrical power, with darker shades indicating regions with better infrastructure compared to those with lighter shades, which are underserved.

Analysis

- **Freetown:**
 - Most facilities have access to all three types of infrastructure, especially Internet connectivity (28 facilities), running water (18 facilities), and electrical power (21 facilities). This highlights Freetown as the most developed region in terms of infrastructure.
 - Regions with Moderate Access:
 - Kakua: Facilities show significant access to Internet connectivity (19) but lower levels for running water and electric power.
- **Rural Western Area:**
 - Exhibits moderate access to all three infrastructure types, with Internet connectivity (17) being the highest.
 - Regions with Limited Access:
 - Locations such as Badjia, Selenga, Wonde, and Bargbo have very low counts across all three categories, suggesting severe infrastructural challenges.
 - Infrastructure Gaps:
 - Stable electrical power (red segment) is frequently the least available across locations.
 - Access to running water (green segment) is also limited in several regions, indicating a critical area for improvement.

Prepare a table with all the new cases (data elements) for each month of 2023 and 2024 by the district. By using ANOVA, test the hypothesis that the occurrence of Malaria (using treated as a conduit) is predictable from the occurrence of other

Procedure:

vector-borne diseases

In the "Data Visualizer" App in DHIS2

- Data:**

- Data Item:** "Data Elements"

- Schistosomiasis new
 - Yellow fever new
 - Onchocerciasis new
 - Malaria treated new =

- Calculation:**

- Malaria treated at PHU with ACT < 24 hrs new +
 - Malaria treated at PHU with ACT > 24 hrs new +
 - Malaria treated at PHU without ACT < 24 hrs new +
 - Malaria treated at PHU without ACT > 24 hrs new +
 - Malaria treated in community with ACT < 24 hrs new +
 - Malaria treated in community with ACT > 24 hrs new +

- Period:** Fixed Period – Monthly – January 2023 -December 2024

- Organization Unit:** Sierra Leone

- Visualization:** Pivot Table

	District groups - District levels			
	Onchocerciasis new	Schistosomiasis new	Yellow fever new	Total new malaria cases
January 2023	102	596	41	67 594
February 2023	48	550	35	54 534
March 2023	94	701	36	85 118
April 2023	89	742	4	61 934
May 2023	84	785	39	173 908
June 2023	73	760	18	146 831
July 2023	82	435	42	146 611
August 2023	123	416	134	167 883
September 2023	58	539	86	158 752
October 2023	67	361	7	115 895
November 2023	123	474	33	133 574
December 2023	109	383	8	86 392
January 2024	102	596	41	67 555
February 2024	48	550	35	54 513
March 2024	89	742	4	61 934
April 2024	94	701	36	85 118
May 2024	84	785	39	173 908
June 2024	73	760	18	146 831
July 2024	82	435	42	146 611
August 2024	123	416	134	167 883
September 2024	58	539	86	158 752
October 2024	123	474	33	133 574
November 2024	67	361	7	115 895
December 2024	109	383	8	86 392

Hypotheses:

H0: vector diseases are independent from malaria.

Ha: vector diseases case number can predict number of malaria.

Test Assumptions:

Observations are independent

Normal distribution of residuals

Homogeneity of variance

Run ANOVA:

Interpretation of ANOVA results:

- Since yellow fever is significant, the null hypothesis can be rejected.
- This suggests that there is an association between the number of new malaria cases when considering this disease.

```
malaria_vect_dat <- read.csv("Question 7 part A final.csv")
names(malaria_vect_dat)
# Then modify the ANOVA code accordingly:
anov_mod <- aov(`Total.new.malaria.cases` ~
  `Schistosomiasis.new` + `Yellow.fever.new` +
  `Onchocerciasis.new`, data = malaria_vect_dat)

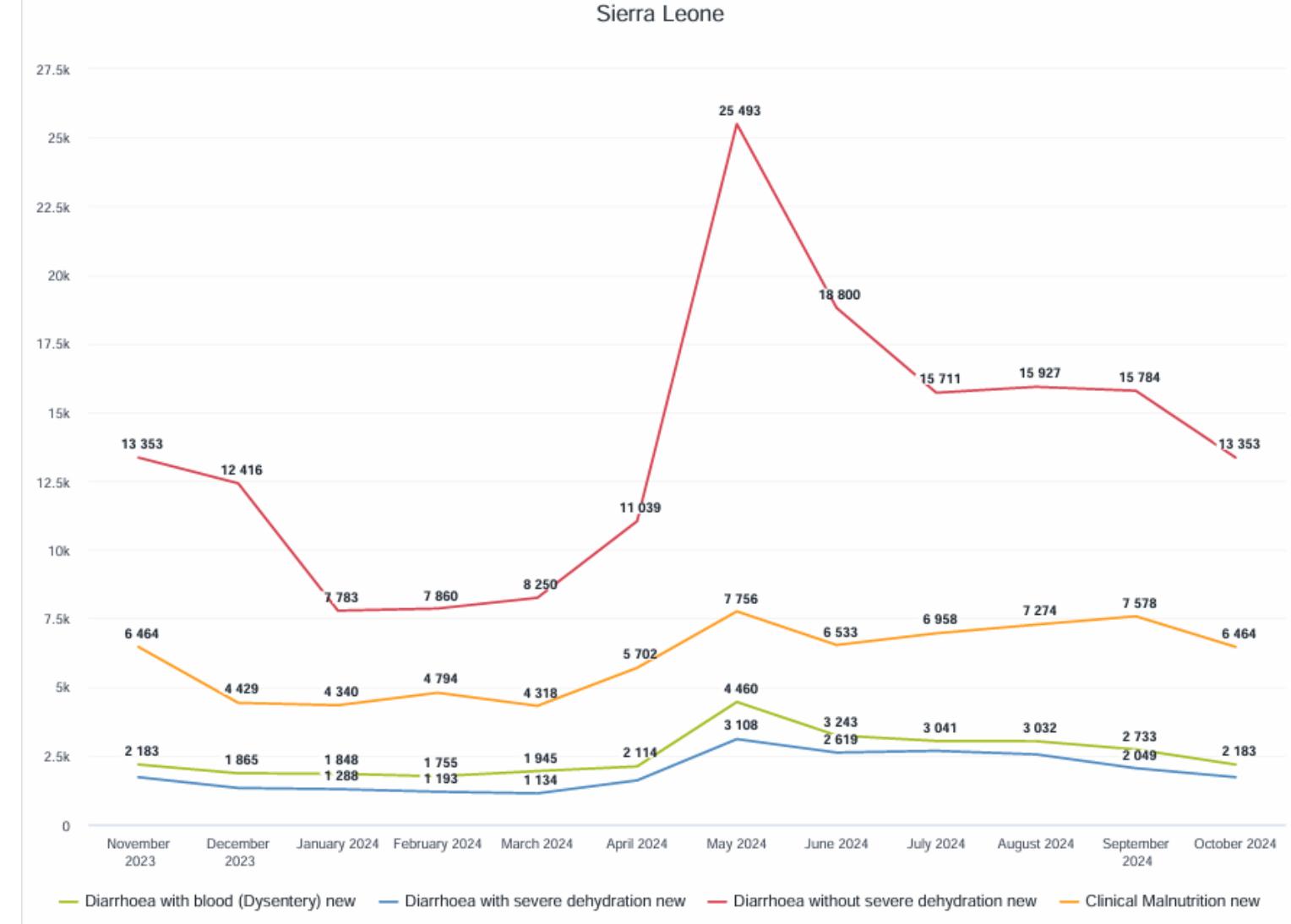
summary(anov_mod)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Schistosomiasis.new	1	1.008e+08	100762331	2.464	0.121
Yellow.fever.new	1	9.067e+08	906731167	22.171	1.27e-05 ***
Onchocerciasis.new	1	1.902e+07	19022291	0.465	0.498
Residuals	68	2.781e+09	40896614		

Signif. codes:	0	***	0.001	**	0.01 *
	.	0.05	.	0.1	' '
	1				

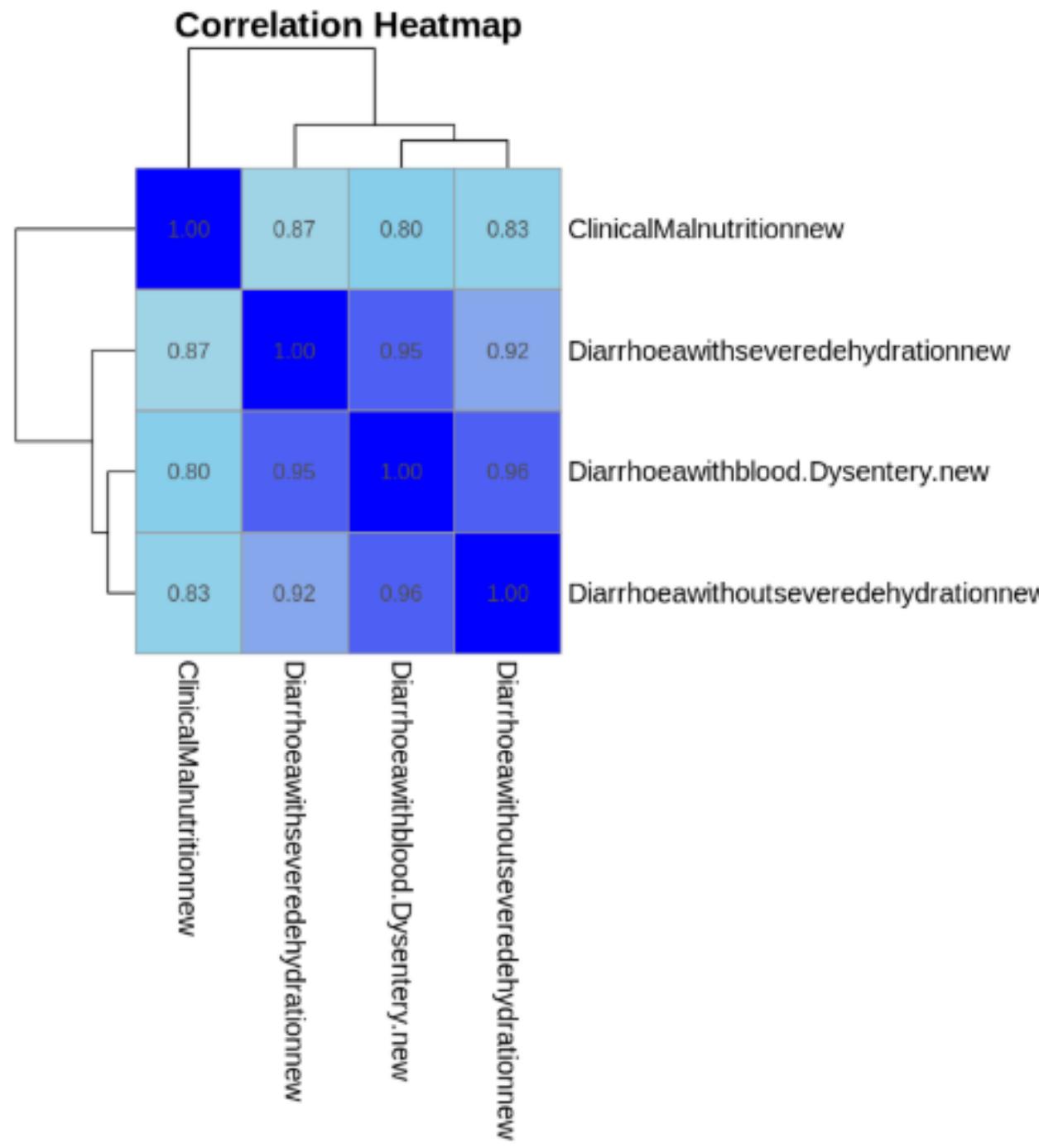
- Create a line chart of new cases of Clinical malnutrition, Diarrhoea with blood (dysentery), diarrhea with severe dehydration, diarrhea without severe dehydration - for the last 1 year in all of Sierra Leone. Describe the correlation that you may see between these diagnoses with the appropriate chart.

There is a correlation between all cases of diarrhea and clinical malnutrition. The number of cases slightly decreases in December, gradually increases over the following months, peaks in May, and then begins to decline steadily.



Correlation matrix

- All variables have strong positive correlations with one another.
- The strongest correlation is between Diarrhoeawithblood(Dysentery) and Diarrhoeawithoutseveredehydration (0.96)
- It is deal for quantifying relationships between multiple variables, providing precise numerical measures of association.



Sentiment analysis of team opinions
on the two visualizations, including
preferences for specific chart types.

We used the reticulate package to perform sentimental analysis of the team's opinions

```
[1] install.packages('reticulate')
install.packages('tidyverse')
install.packages('DT')
library(reticulate)
library(tidyverse)
library(DT)

→ Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

also installing the dependencies 'RcppTOML', 'here', 'png'

Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

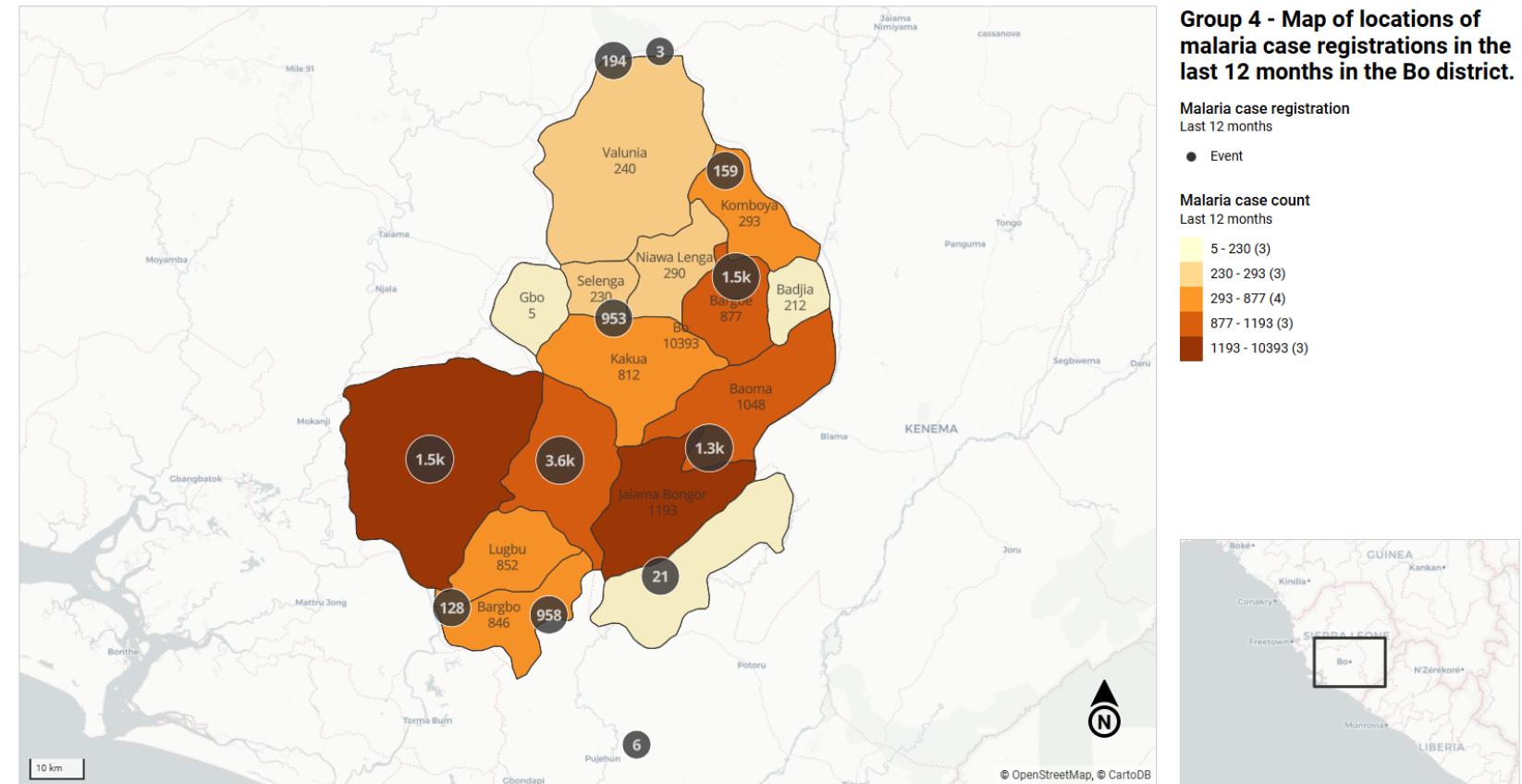
also installing the dependencies 'lazyeval', 'crosstalk'
```

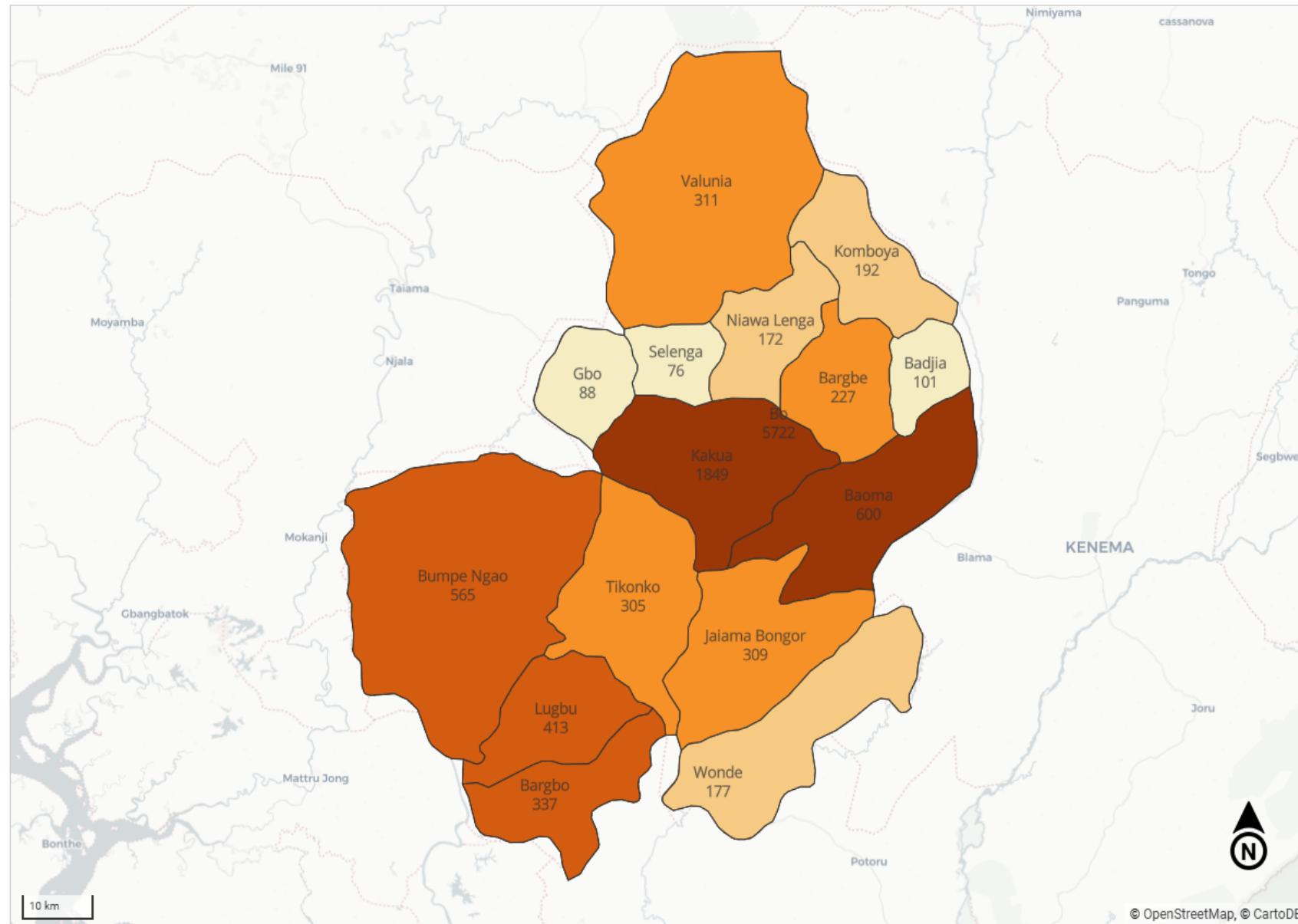
Cont.....

Question 3	
Positives	11
Negatives	10
Average Score	0.938
Conclusion: The map preference for the group members is predominantly positive with an average confidence score of 0.938	

Question 8	
Positives	15
Negatives	6
Average Score	0.949
Conclusion: The line chart preference for the group members is predominantly positive with an average confidence score of 0.949.	

Create a map of locations of malaria case registrations in the last 12 months in the Bo district. 2. Create another map of Inpatient morbidity and mortality over the last 12 months in Bo district. Using these maps as images, and a clustering algorithm, describe the clusters and their relationship within a map or between the two maps using any nearness metric.





map of Inpatient morbidity and mortality over the last 12 months in Bo district

map of Inpatient morbidity and mortality over the last 12 months in Bo district.

Inpatient morbidity and mortality Age in years Last 12 months

76 - 172 (3)
172 - 227 (3)
227 - 337 (4)
337 - 600 (3)
600 - 5722 (3)
No data



```

# Add custom colors for clusters
cluster_colors <- c("red", "blue", "green", "yellow", "orange", "purple")

# Add legend
legend("topright",
       legend= paste("Cluster", 1:length(unique(labels(apres))),
                     col=cluster_colors[1:length(unique(labels(apres)))],
                     pch=19))

# Add region labels
text(data$x, data$y,
      labels=data$region,
      pos=4,
      cex=0.7)

# Print cluster information
print("Clustering Results:")

# Install required packages
install.packages("apcluster")
library(apcluster)

# Create combined data with both malaria cases and mortality, along with coord:
data <- data.frame(
  region = c("Badjia", "Baoma", "Bargbe", "Bargbo", "Bumpe Ngao", "Gbo",
            "Jaialma Bongor", "Kakua", "Komboya", "Lugbu", "Niawa Lenga",
            "Selenga", "Tikonko", "Valunia", "Wonde"),
  # Average malaria cases
  malaria = c(17.5, 86.5, 73.5, 70.5, 208.3, 1.0, 98.8, 67.7, 24.4, 71.8, 26
  # Mortality values
  mortality = c(101, 600, 227, 337, 565, 88, 309, 1849, 192, 413, 172, 76, 30
  # Coordinates
  x = c(1000, 1200, 950, 700, 400, 600, 900, 900, 950, 600, 800, 700, 700, 800),
  y = c(800, 600, 750, 400, 500, 800, 500, 600, 900, 400, 850, 750, 500, 1000)
)

# Create matrix for clustering
combined_matrix <- as.matrix(data[, c("malaria", "mortality", "x", "y")])

# Create similarity matrix
similarity_matrix <- negDistMat(r=2, combined_matrix)

# Perform clustering
apres <- apcluster(similarity_matrix, details=TRUE)

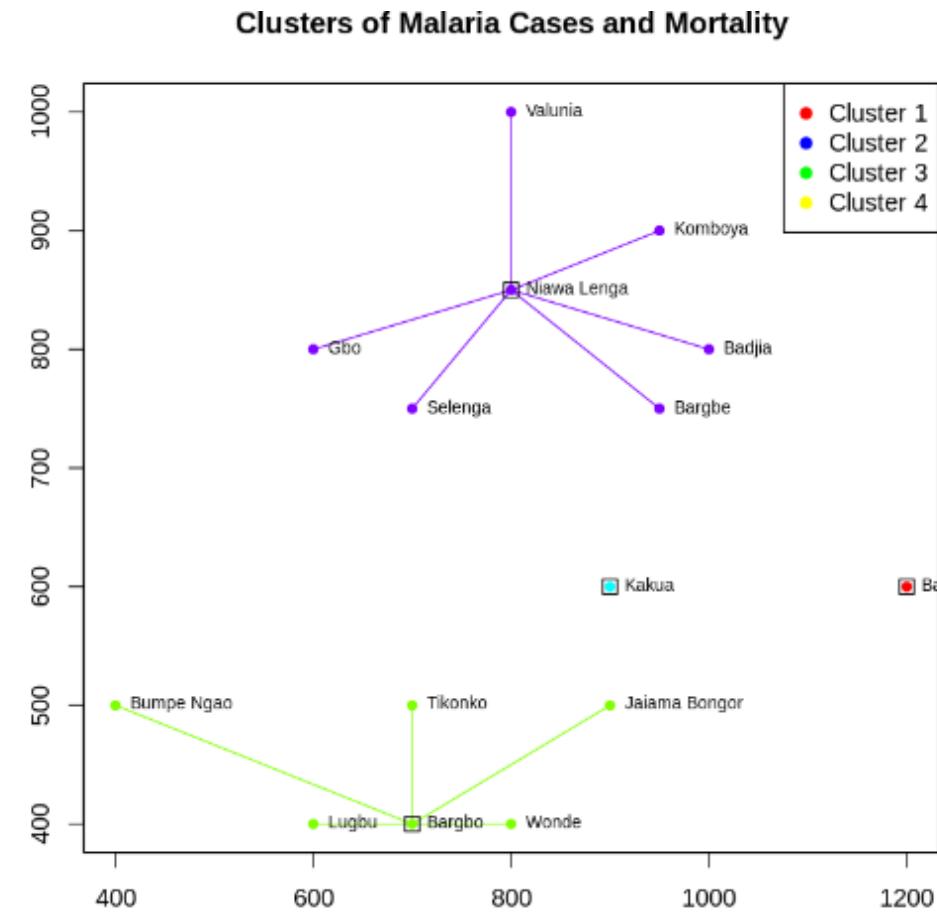
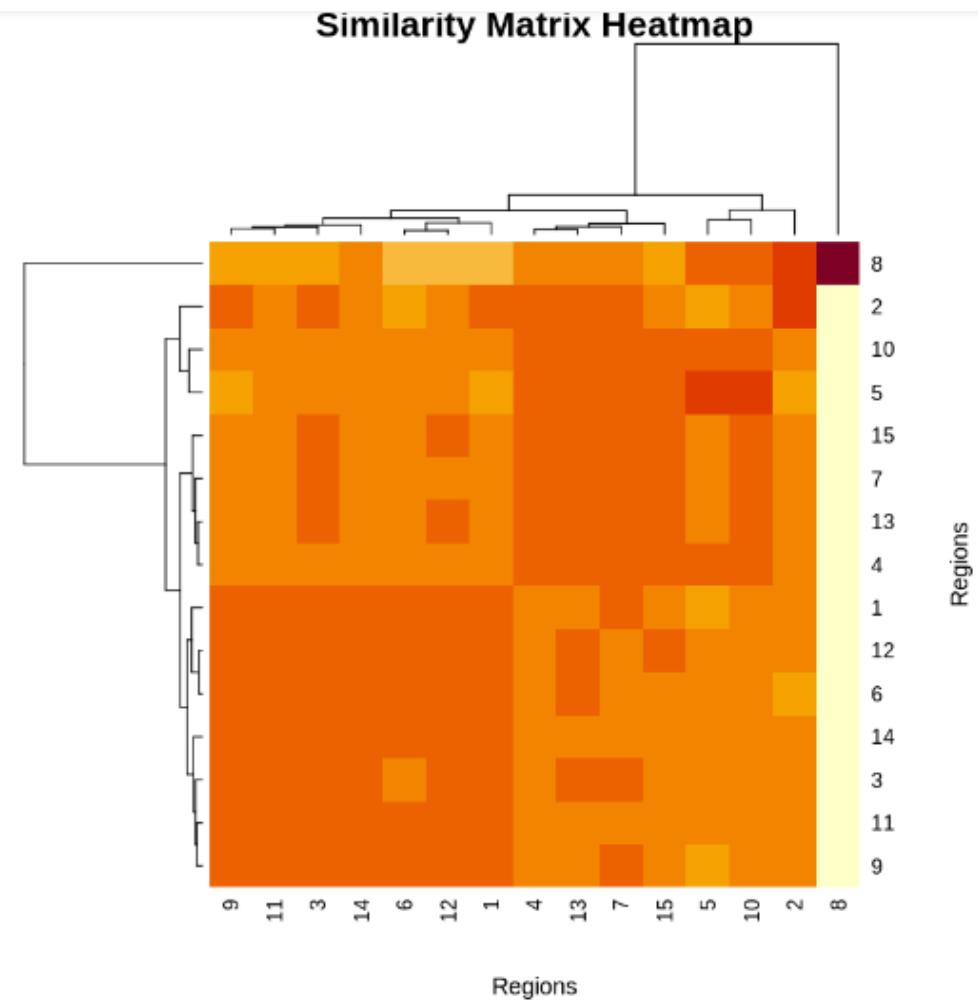
```

```

# Print cluster compositions
clusters <- factor(labels(apres))
for(i in 1:length(unique(clusters))) {
  cat("\nCluster", i, "contains:\n")
  cluster_regions <- data$region[clusters == i]
  cluster_malaria <- data$malaria[clusters == i]
  cluster_mortality <- data$mortality[clusters == i]
  for(j in 1:length(cluster_regions)) {
    cat(sprintf("%s: Malaria=% .1f, Mortality=%d\n",
                cluster_regions[j],
                cluster_malaria[j],
                cluster_mortality[j]))
  }
}

# Plot heatmap of similarity matrix
heatmap(similarity_matrix,
        main="Similarity Matrix Heatmap",
        xlab="Regions",
        ylab="Regions")

```



Interpretation s of clustering analysis:

Cluster Identification

- 4 distinct clusters identified using apcluster in R
- Clusters reveal both geographic and health-metric patterns
- Single exemplars (Baoma, Kakua) represent unique risk areas

Critical Hotspots

- Kakua: Highest mortality (1849) despite moderate malaria cases
- Bumpe Ngao: Highest malaria cases (208.3)
- Northern regions show consistently better health outcomes

Key Findings

- Weak correlation (0.334) between malaria cases and mortality
- Geographic proximity influences cluster formation
- Northern cluster practices could provide valuable lessons

Implications

- Targeted interventions needed for different clusters
- Resource allocation should consider both geographic and health metrics
- Success factors in northern regions warrant further investigation
- This clustering analysis enables evidence-based decision-making for healthcare resource distribution and intervention strategies.

Thank you
Questions?

Appendix

- Question 1b

<https://colab.research.google.com/drive/1GVMQqkvVkjFWS6UzQWgMd7LR84jDohVj?usp=sharing>

- Question 4:

<https://colab.research.google.com/drive/1A8OSSAFvofNI2viGkFoDuLYrXgCC1p5H?usp=sharing>

- Question 5:

<https://colab.research.google.com/drive/17weDzu0nXxd9770XOaE7Vn7h4ZtAGP34?usp=sharing>

- Question 7 and Question 10: <https://colab.research.google.com/drive/1iUOpHqs-Moi4eTgSSHb24ci9Zs546pe?usp=sharing>

- Question 8:

<https://colab.research.google.com/drive/1uAAoKqSvoTbcqQbrSue2Vskr73AdESuP?usp=sharing>

Cont....

- Question 9:
[https://colab.research.google.com/drive/1eON7lETmEcjwnGzTR55z8JAJZl1rOFT2?
usp=sharing](https://colab.research.google.com/drive/1eON7lETmEcjwnGzTR55z8JAJZl1rOFT2?usp=sharing)
- Xian, X., Wang, L., Wu, X., Tang, X., Zhai, X., Yu, R., Qu, L., & Ye, M. (2023). Comparison of SARIMA model, Holt-winters model and ETS model in predicting the incidence of foodborne disease. *BMC Infectious Diseases*, 23(1).
<https://doi.org/10.1186/s12879-023-08799-4>