BIOS 6640-R Project

Piper Williams

10/26/2018

Background

Malaria is a parasitic blood disease that is most commonly transmitted via the *Anopheles* mosquito species.[1] Other forms of malaria parasite transmission include congenital transmission and transfusion transmission.[1] There are four malaria parasites known to infect humans: *Plasomdium falciparum, vivax, ovale,* and *malariae.*[1] Symptoms associated with malaria begin to manifest anywhere from seven to thirty days after the initial exposure and include fever, headache, increased heart rate, and various gastrointestinal problems.1 If an infected individual is not diagnosed and treated promptly, malaria has the potential to be fatal. It should be noted that the most vulnerable population to malaria infection is children under five years of age, as this population is less likley to have acquired immunity. While the global incidence of malaria has decreased by approximately 18% from 2010 to 2016, malaria remains a major public health issue for numerous countries.[2]

In the year 2016, there were approximately 216 million reported cases of malaria and 450,000 malaria-related deaths worldwide.[3] In the same year, 91% of all malaria-related deaths were concentrated in the World Health Organization (WHO) African region, which includes the country of Mozambique.[3] In Mozambique, approximately four to six million malaria cases are reported each year. While this disease is known to be largely preventable and curable, it continues to be an epidemic in Mozambique and other countries in sub-Saharan Africa.[4] Allotting preventative resources to highly vunerable areas, which include insecticide-treated bednets and indoor residual spraying, educating citizens in these regions about the importance of using these resources, and malaria surveillance are each critical components of malaria prevention and possible elimination.[5] To ensure that malaria prevention resources are distributed and implemented effectively, the trends in malaria incidence must be better understood.

In this report, the spatial and temporal trends in malaria incidence in children under five and their relationship to specific climatic conditions in Mozambique were further explored. Increased rainfall and increased ambient temperatures have previously been shown to be associated with increased malaria transmission, as these climate conditions provide an optimal environment for the malaria parasite to be transmitted to the *Anopheles* mosquito and, ultimately, to humans.[1] However, the time at which these climatic conditions are associated with malaria exposure remains unknown. Increased rainfall creates a larger number of collections of water, which serve as prime breeding sites for the mosquito species.[1] Increased ambient temperatures shortens the malaria parasite's growth cycle, which increases the probability of transmission from an infected mosquito mother to her offspring.[1] In uncovering these spatio-temporal relationships and their association with climatic trends, there is an increased potential to prevent malaria transmission in Mozambique and other countries around the world.

Description of Research Problem and Data

The ultiamte goal of this report was to uncover underlying trends in malaria incidence in Mozambique. In particular, the visuals presented aimed to shed light on geographic and seasonal trends in malaria incidence over time and their association with trends in certain climatic conditions as well. As discussed previously, increased rainfall and temperature are strongly associated with increased malaria transmission. However, the lag times of rainfall and temperature that are most associated with malaria incidence are not currently known. The relationships between these two climatic conditions and malaria incidence were extensively explored here. In better understanding these trends and their associations with one another, preventative resources could be allocated more effectively to vulnerable areas in Mozambique and potentially decrease malaria transmission.

The final data set used in subsequent visualizations consisted of weekly data for the 142 districts of Mozambique from 2010 to 2017. The first two weeks of data for 2010 were missing. Also, data was collected only for the first 9 weeks of 2017. Due to the lack of data for this year, I chose to exclude 2017 data in this report. Variables measured included weekly total rainfall (mm), average rainfall (mm), average temperature (Celcius), relative humidity (%), saturation vapor pressure deficit (mmHg), surface barometric pressure (hPa), total district population, proportion of total population under five years of age, number of malaria cases under five reported, etc. The outcome-of-interest was cases per 1000 in children under five, which was calculated as follows:

$$CPT - U5 = [\frac{C - U5}{(Population_{total} * U5_{weight})}] * 1000$$

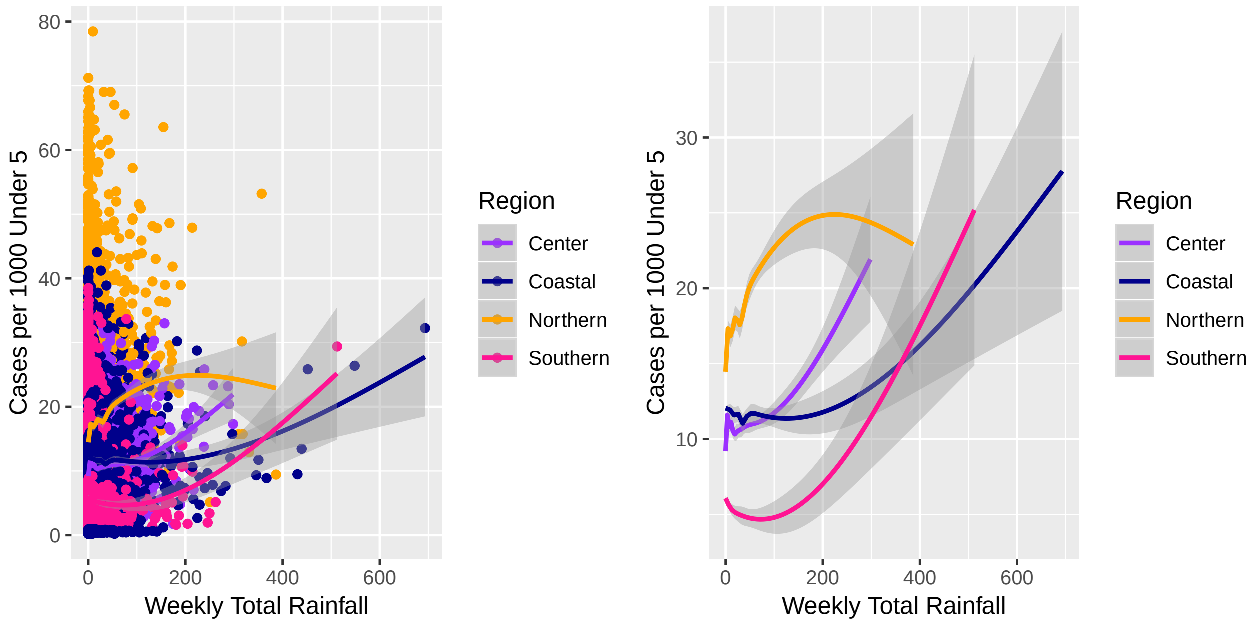
$CPT - U5$ is cases per 1000 under 5, $C - U5$ is total number of cases under 5, $Population_{total}$ is the total population of the district, and $U5_{weight}$ is the proportion of the total population under five years of age. A proportion of cases was used as the outcome rather than simply the number of malaria cases in children under five to control against the fact that certain districts have a larger total population than others (thus, more malaria cases simply due to a larger population in that district).

Results

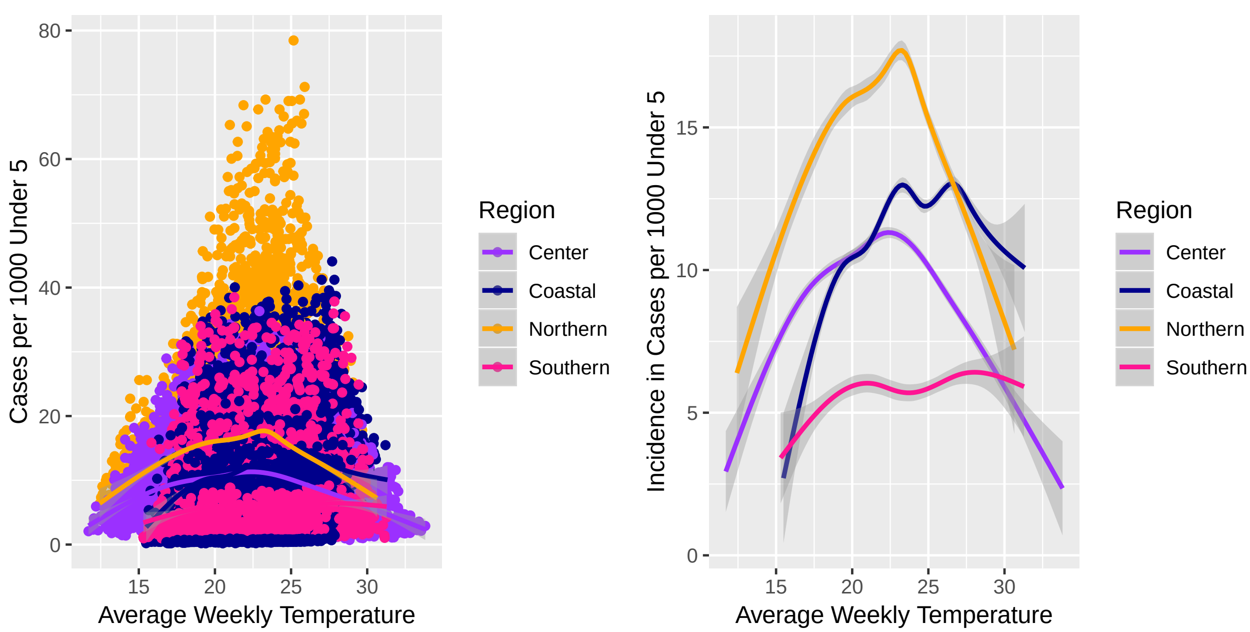
Initially-Observed Data Trends

While there was no formal statistical modeling conducted in this report, exploratory data analyses were carried out to better visualize the outcome-of-interest (cases per 1000 under 5, CPT-U5) and its relationship with other variables-of-interest. First, it was discovered that CPT-U5 was postively skewed (see Appendix). A log-transformation as well as square-root transformation of the outcome were implemented to see if this would change the overall distribution of the outcome. A square-root transformed outcome seemed to follow a relatively normal distribution (see Appendix).

The relationships between CPT-U5 and weekly total rainfall (mm) as well as average temperature (Celcius) are visualized below. These trends were also evaluated separately for each of the four regions in Mozambique.



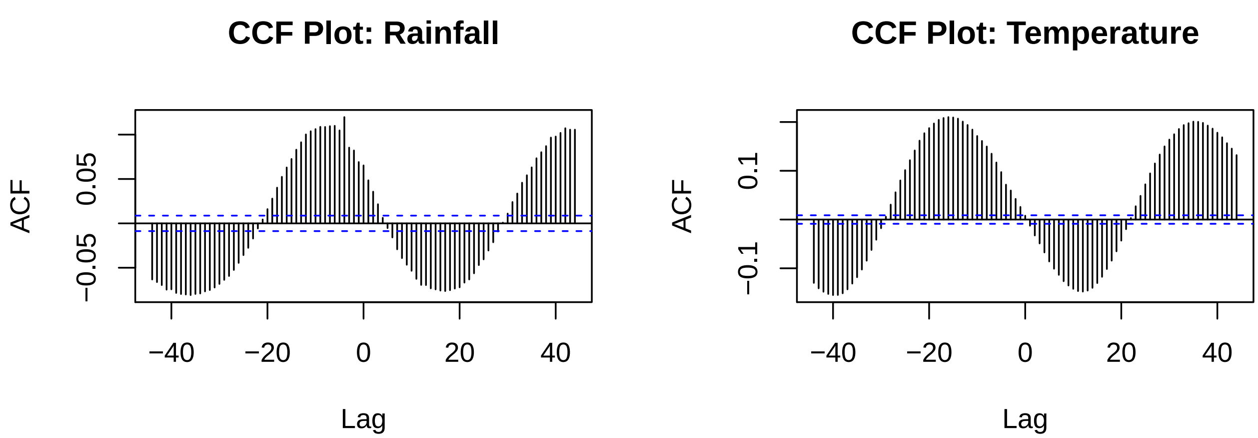
From a first glance, the smoothing splines above show that there is a non-linear relationship between rainfall and CPT-U5. The graphs also indicate that, overall, the northern region of Mozambique tended to have a higher proportion of CPT-U5, while the southern region tended to have a lower porportion of CPT-U5. While there appears to be a somewhat quadratic trend between CPT-U5 and rainfall, it is important to note that this could be due to the fact that the right tails of the smoothing splines are heavily influenced by the small number of extremely large rainfall data points. It is well known that cubic smoothing splines can be heavily influenced by "outliers"/extreme observations in the data, and one should be cautious when interpretting smoothing splines if extreme data values are present.



The plots above indicate a potential quadratic, maybe even quartic relationship between average temperature and CPT-U5. Similarly to the previous graphs, this indicates that, overall, the nothern region tends to have a higher proportion of CPT-U5, while the southern region tends to have a lower proportion of CPT-U5. Also, these plots show that there is a "sweet spot" for malaria incidence in terms of temperature. As discussed previously, it is true that malaria transmission tends to increase with an increase in temperature, but this is only up to a certain point. One can imagine that as temperature continues to increase, there may be a threshold at which malaria parasites cannot survive. Thus, this explains why we see a drop-off in the incidence of malaria beyond a certain temperature.

Optimal Lag of Total Rainfall and Average Temperature

It is known that there is a strong association between malaria incidence and climatic conditions, such as rainfall and temperature. However, the optimal lag time between these climatic conditions and malaria incidence is not currently well known. The cross-correlation function is helpful in determining which lags of the climatic conditions may be useful predictors of CPT-U5. It is important to note that a negative lag indicates that there is a correlation between the climatic condition at a time prior to t and malaria incidence at time t. A positive lag indicates that there is a correlation between the climatic condition at a time after t and malaria incidence at time t. Below are the cross-correlation function plots for weekly total rainfall and weekly average temperature. Also included below is a table that summizes the lag assocated with the largest auto-correlation function ($|ACF|$) value.

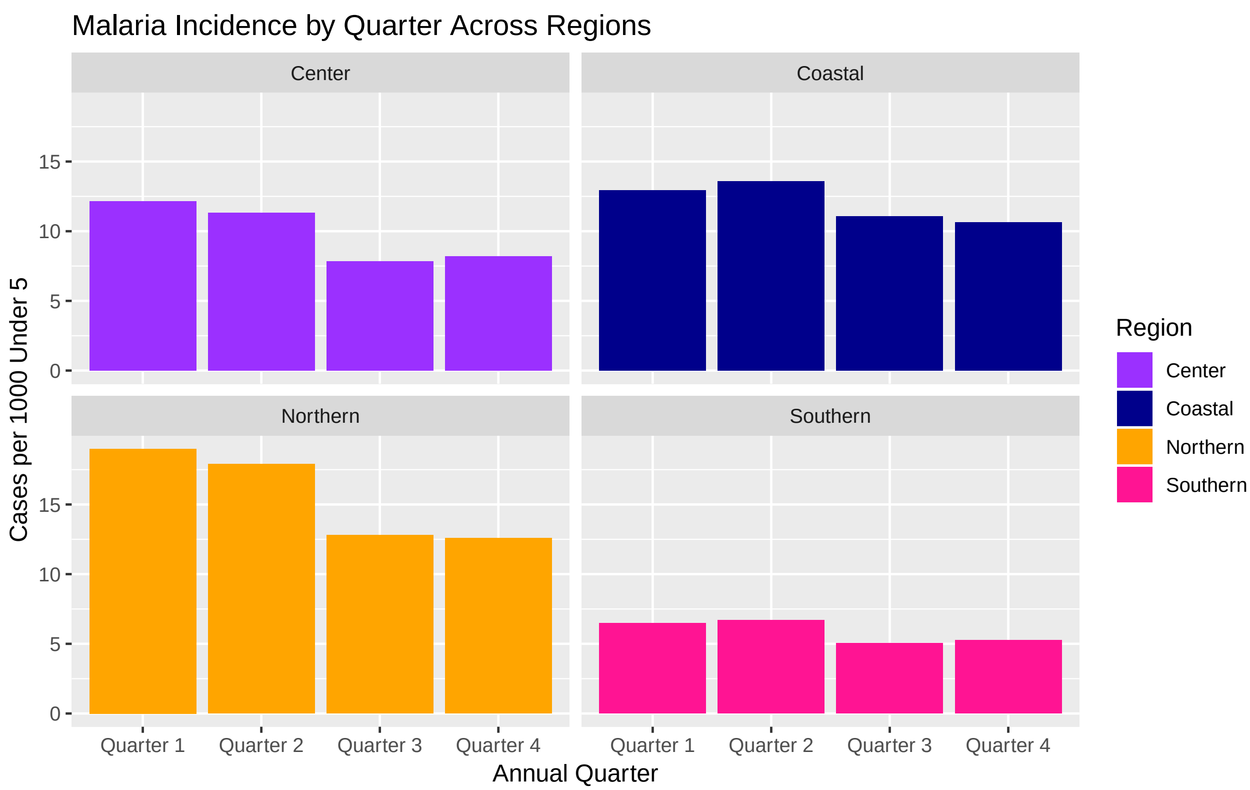


	Lag	Max. ACF
Weekly Total Rainfall	-4	0.1197
Average Weekly Temperature	-16	0.2101

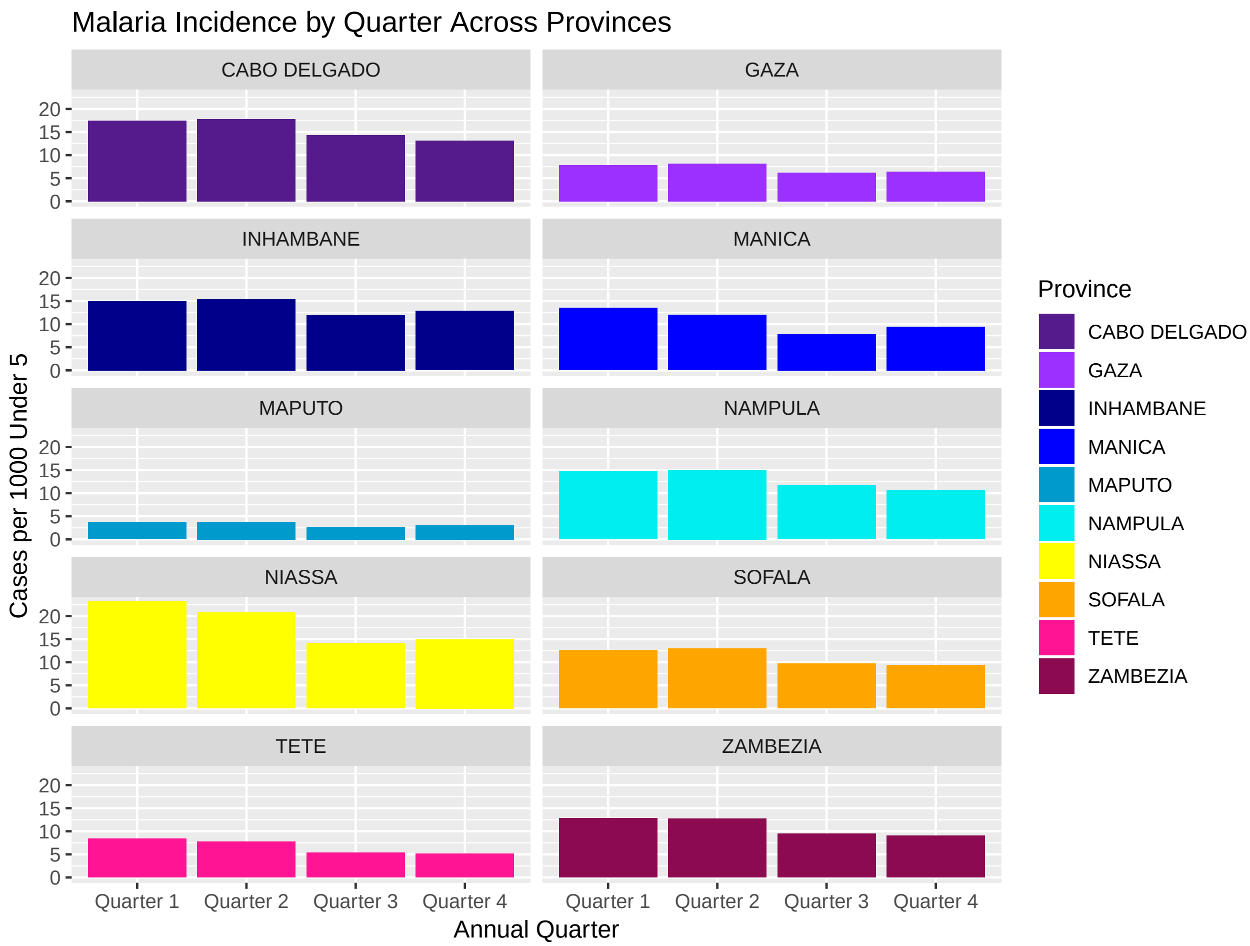
Here, we can see that the optimal lags for total rainfall and temperature were -4 and -16, respectively. This means that rainfall 4 weeks prior to t is most strongly correlated with CPT-U5 at t and average temperature 16 weeks prior to t is most strongly correlated with CPT-U5 at t. It was originally thought that rainfall and temperature lead times typically ranged from 2 to 8 weeks. While the results from the CCF Rainfall plot are consistent with this range of expected lag times, the CCF Temperature plot indicates that the lag time for temperature was much longer than originally anticipated. One key point that should be made is that the ACF values for lags proximal to the -16 lag seemed to be extremely close in value to the ACF at lag -16. So while the -16 lag had the highest $|ACF|$, this was only marginally so. Another key point that should be made is that the cross-correlation function uses the Pearson correlation method in its calculation of the ACF. It is possible that the correlation between the climatic condition variables and the outcome-of-interest are not linearly correlated. This may be why we see a difference in the actual optimal lag time for temperature here versus the anticipated lag time.

Quarterly Malaria Incidence Across Mozambique Regions and Provinces

Rather than aggregating over all time, the graphs below show malaria incidence in CPT-U5 for each annual quarter. Visualizing the data as such will give a better sense of the dynamic trend in malaria incidence over time. Also, in breaking up data visualizations into annual quarters, the temporal trends in the climatic conditions and their relationship with malaria incidence over time will hopefully become more apparent as well. For the graphs below and for the remaining data visualizations, the data will be broken up via annual quarters. Quarter 1 was defined as weeks 1-13, quarter 2 as weeks 14-26, quarter 3 as weeks 27-39, and quarter 4 as weeks 40+. For the year 2015, there were 53 weeks of data recorded, and for the purposes of this report, week 53 was chosen to be included in quarter 4. The bar graphs below show the average CPT-U5 across regions (bar graph 1) and the average CPT-U5 across provinces (bar graph 2).



The graphs show that, on average, the northern region had the highest average CPT-U5 (average CPT-U5: 15.57) while the southern region had the lowest average CPT-U5 (average CPT-U5: 5.87). Across all four regions, annual quarters 1 and 2 seemed to have a higher average CPT-U5 in comparison to quarters 3 and 4.

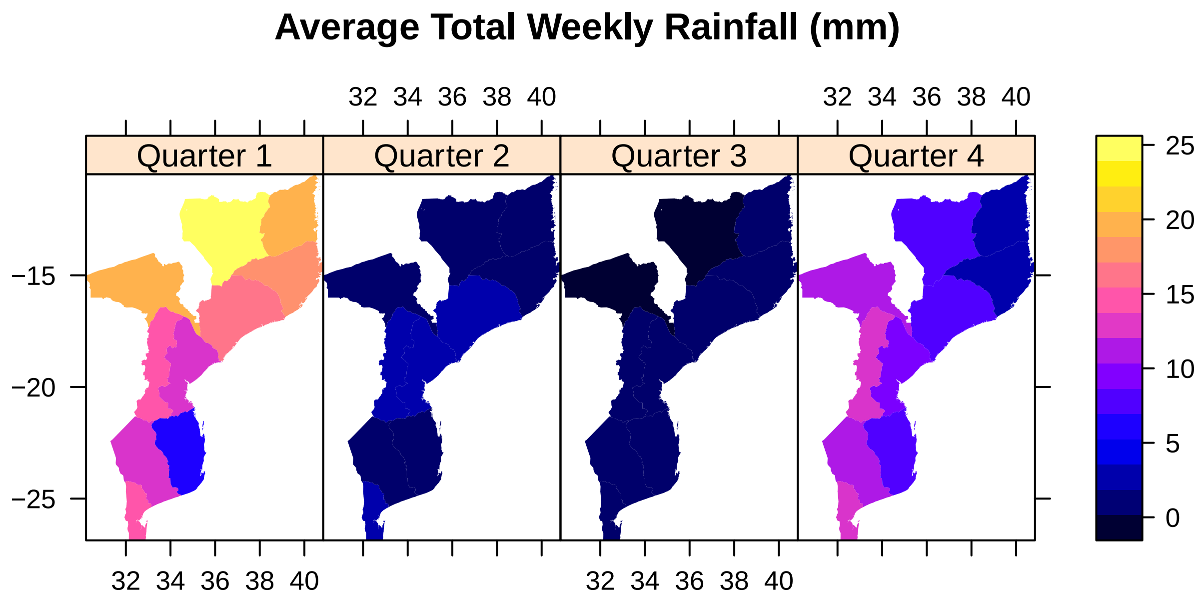


Bar graph 2 was stratified by provinces in Mozambique rather than regions to get a better sense of which provinces are contributing more to the regional patterns observed in the previous graph. Here, we can see that the Niassa province had the highest recorded average CPT-U5 and the Maputo province the lowest recorded average CPT-U5. The Niassa province is located in the northern region of Mozambique, while the Maputo province is located in the southern region of Mozambique. Again, annual quarters 1 and 2 seemed to have a higher average CPT-U5 in comparison to quarters 3 and 4 at the provincial level, similarly to what was observed at the regional level.

Two observations become exceedingly obvious just from the bar graphs alone. First, malaria incidence is geogrpahically dynamic. Second, malaria incidence is temporally dynamic.

Quarterly Total Rainfall and Temperature Across Mozambique Provinces

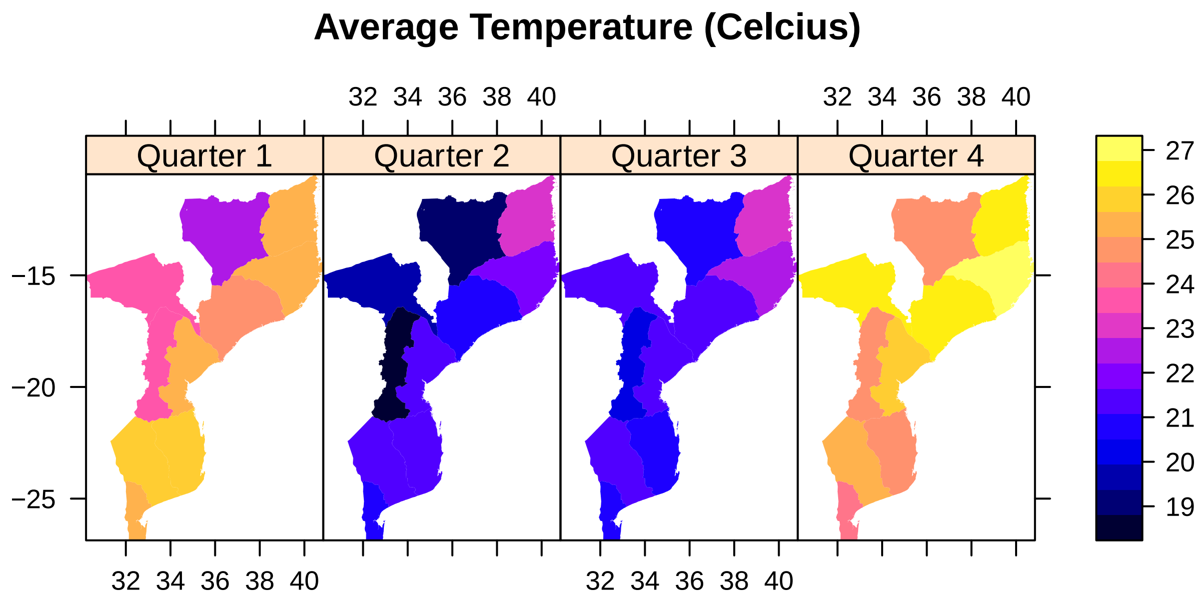
Now that the optimal lag times for total rainfall and temperature have been determined, it would be beneficial to understand the regional and temporal differences in total rainfall and temperature across Mozambique. Below are provincial maps of Mozambique that show average total rainfall and average temperature by annual quarters.



There are a number of conclusions we can draw from these map visualizations. First, rainfall seemed to be higher for quarters 1 and 4 (highest for quarter 1). This is consistent with the lag time for rainfall that is most associated with malaria incidence. The optimal lag time for rainfall was found to be 4 weeks prior to malaria incidence. Thus, it makes sense that the highest total rainfall occurs in quarter 1. We previously saw that quarters 1 and 2 have a higher proportion of cases per 1000 under 5. Thus, it is expected that rainfall would be highest during quarter 1. Also, there seems to be almost zero rainfall on average in quarter 3. These maps indicate that rainfall varies across provinces and also varies across annual quarters.

Next, during quarter 1, the northern region had the largest average total weekly rainfall compared to the other regions In fact, the Niassa province, which is in the northern region, experienced approximately 23.94 mm of weekly total rainfall, on average. The province that experiences the second largest weekly total rainfall during this quarter was Cabo Delgado, another province in the northern region, with an average weekly total rainfall of 19.58 mm.

There is one interesting pattern that should be discussed. It might have been expected that the northern region/provinces would have higher rainfall year-round due to the fact that this area has a higher level of malaria transmission compared to other regions/provinces. However, in quarters 2-4, it is clear that this is not the case. For example, in quarter 4, it seems that the central region, even some portion of the southern region, experiences more rainfall in comparison to the northern region. This indicates that the northern region may experience more inconsistent and extreme climatic patterns in comparison to the other regions. Map plots showing average quarterly total rainfall across *districts* can be found in the appendix.

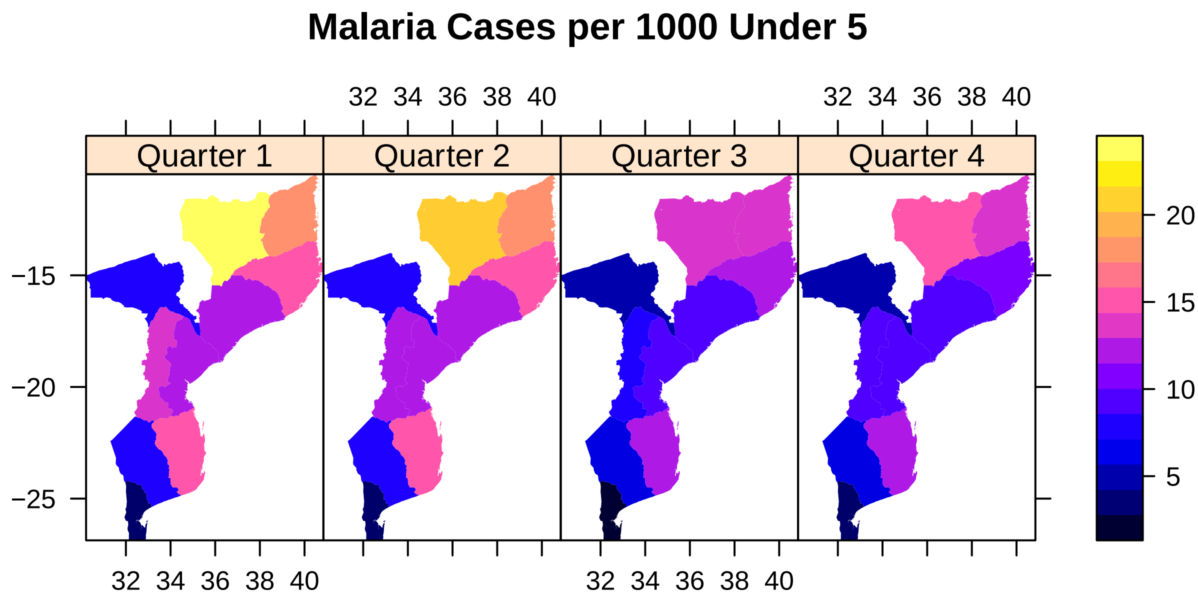


The plots above show average temperature across provinces per annual quarter. It appears that quarter 4 had the highest average temperature compared to quarters 1-3. This is consistent with the lag time of temperature that is most associated with malaria incidence, which was found to be 16 weeks prior to malaria incidence. We would expect the highest temperatures to be in quarter 4 based on this information, and this seemed to be the case. Quarter 1 seemed to have the second highest average temperature across provinces.

Unlike rainfall, the northern region/provinces generally seemed to have higher temperature year-round in comparison to the other regions/provinces. In particular, the two provinces that seemed to have the highest average temperatures year round were Nampula and Cabo Delgado. It is also worth noting that the temperature across all provinces and across all quarters only varies by about 8 degrees Celcius. Map plots showing average quarterly temperature across *districts* can be found in the appendix.

Quarterly Malaria Incidence Across Mozambique Provinces

Now we want to understand if malaria incidence is clustered in a particular region/province in Mozambique. As shown previously, we would expect the northern region to have a higher malaria incidence compared to the other regions/provinces.



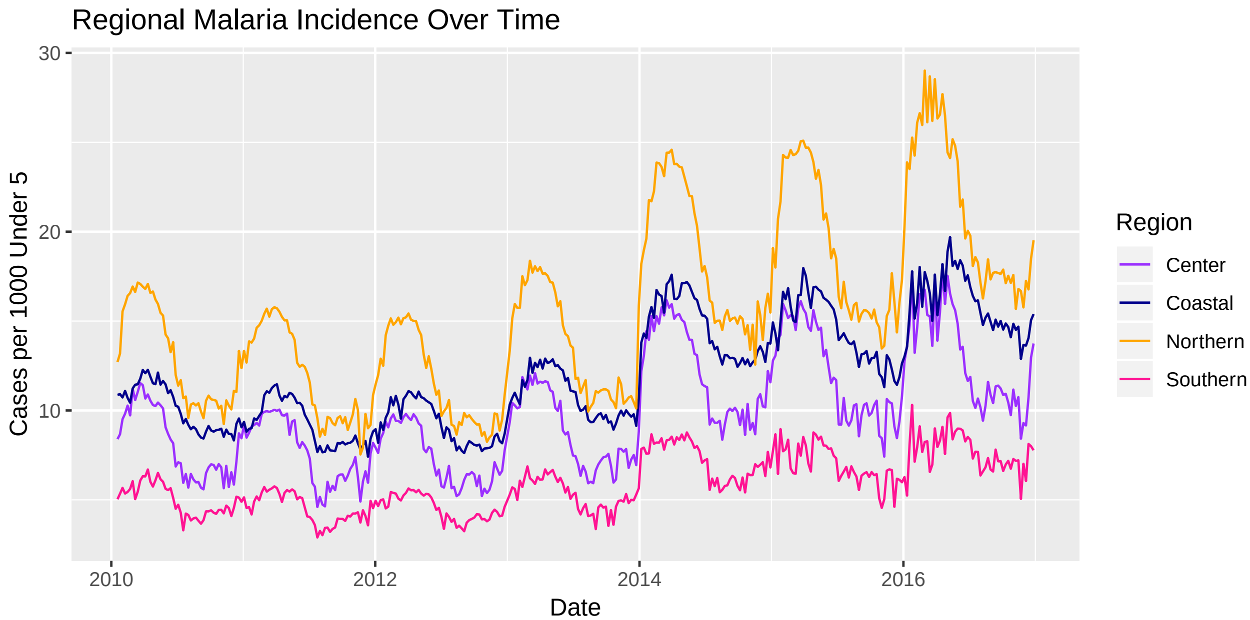
Across quarters and across provinces, it appears that the Niassa and the Cabo Delgado provinces seemed to have the highest average malaria incidence in CPT-U5. Also, malaria incidence seemed to be increased in annual quarters 1 and 2. This is consistent with the previous bar graph visualizations that indicated the northern region had higher malaria incidence compared to the southern, central, and coastal regions and that quarters 1 and 2 had higher malaria incidence.

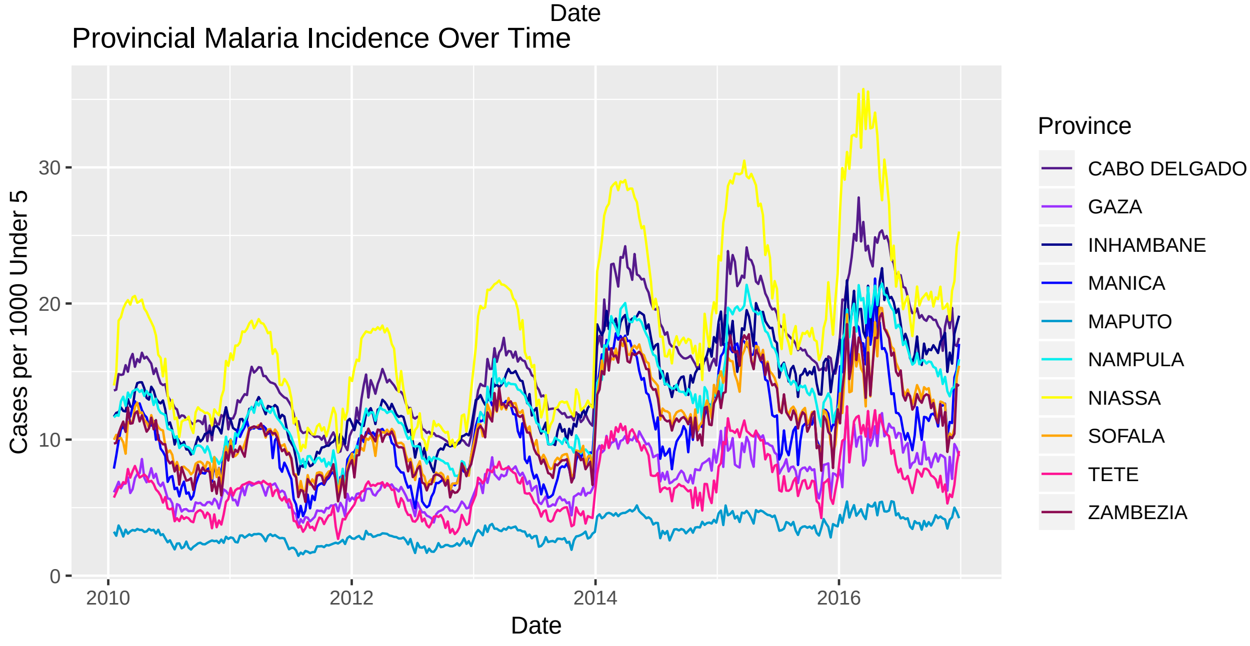
The higher proportion of CPT-U5 in the northern region overlapped with the patterns we see for total rainfall and temperature. Generally speaking, the provinces with increased malaria incidence tended to have higher rainfall and temperature. The particular provinces I am refering to are the Niassa, Nampula, and Cabo Delgado provinces. However, this wasn't necessarily across the whole year. While the northern region/provinces seemed to have the highest average temperature over the whole year, it was discovered that rainfall was not consistently high or low for all provinces throughout the year. Thus, it can be concluded that the northern region may have more extreme, but annually inconsistent, climatic patterns, which likely provide the optimal enviroment for the malaria parasite to grow and be transmitted.

Another interesting detail regarding malaria incidence was discovered. In the appendix of this report is a map plot that shows quarterly malaria incidence in CPT-U5 across *districts*. The map indicates that there is one district that seems to have an unusually high CPT-U5 malaria incidence across quarters. This district is the Meluco district in the Cabo Delgado province (average CPT-U5 is approximately 40). Another district that seems to have a high CPT-U5 malaria incidence is the Nipepe district in the Niassa province. Both of these provinces are in the northern region of Mozambique. These districts are likely contributing the most to the high malaria incidence observed in the northern region. Again, this district-level map plot can be found in the appendix of the report.

Regional and Provincial Malaria Incidence Over Time

Finally, we want to see how Malaria is changing over *all* time, not just by quarter. As mentioned, due to the lack of data available for the year 2017, this year was left out of the visualizations/analyses for this report. While the data did not provide exact dates, we can still visualize the trends in malaria transmission weekly from 2010 to 2016. These time series graphs are shown below, in which one plot stratified CPT-U5 across regions and the other across provinces.

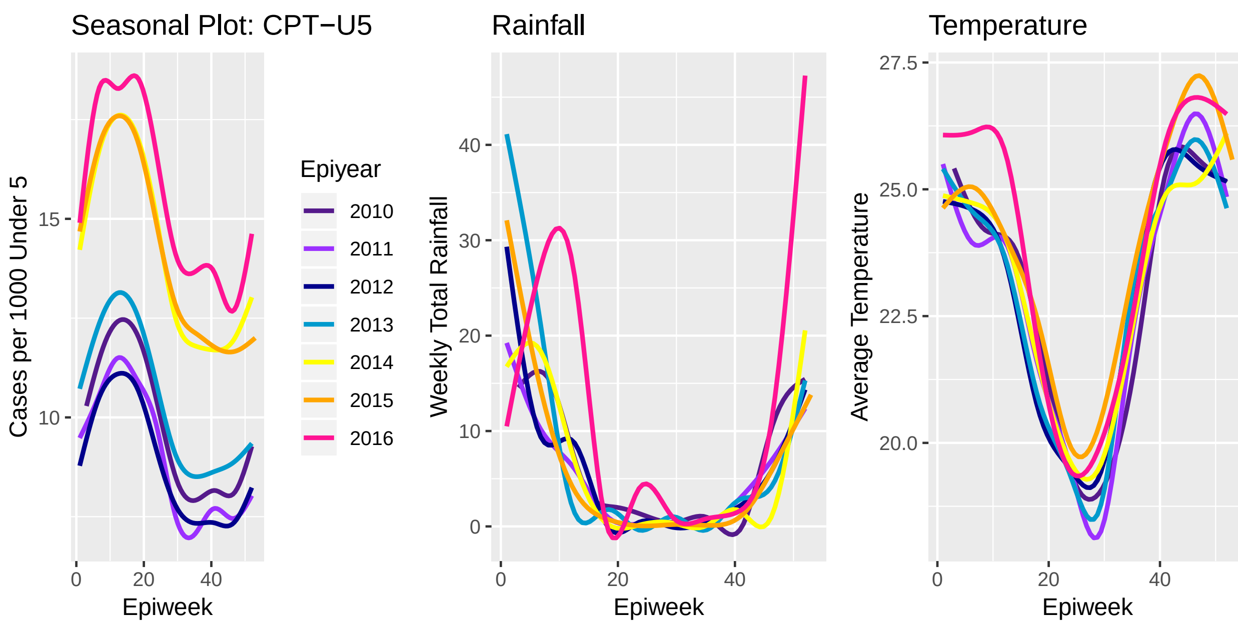




These time series plots show that there seemed to be a steady increase in malaria incidence in CPT-U5 from 2010 to 2016 across regions and/or provinces. These graphs also indicate that malaria incidencemay have increased slightly more over time in the northern region/provinces and less over time in the southern region/provinces. One other key observation is that the fluctuation in CPT-U5 in the northern region is much greater than that in the southern region. Previously we discovered that the northern region may have more extreme fluctuations in climatic conditions over time, and this may be translating into a greater fluctuation (more extreme sinusoidal pattern) in malaria incidence compared to the other regions in Mozambique.

There is one other factor that may be influencing the patterns we observed in the time series plots. The capital of Mozambique is Maputo, which is located in the southern region of the country. One reason there were less observed fluctuations and also less of an increase in malaria incidence in CPT-U5 in the souther region/provinces may be because those in the southern region are closer to the capital, which is highly urbanized compared to the rest of the country. Thus, the citizens in this region may have more access to health care resources that reduce malaria transmission.

Another to show if malaria incidence is increasing over time is via seasonal plots, shown below. Seasonal plots for rainfall and temperature are also shown.



Here, we can again see that malaria incidence increased over time from 2010 to 2016. Also, these seasonal plots indicate that there were rainfall and temperature increases over time at the beginning of the year and towards the end of the year. Particularly in 2016, there was a drastic increase in rainfall and temperature at the beginning and the end of that year.

Conclusions

Based on the results presented, it can be concluded that malaria incidence clearly has a geographical and temporal trend. In particular, there seemed to be larger malaria incidence in CPT-U5 in the northern region (Niassa, Cabo Delgado, and Nampula provinces) of Mozambique and smaller malaria incidence in the southern region (particularly for the Maputo province). Malaria incidence was higher in annual quarters 1 and 2 in comparison to quarters 3 and 4. Also, from 2010 to 2016, malaria incidence seemed to increase across all regions/provinces over time.

Malaria incidence was also strongly associated with climatic conditions, such as weekly total rainfall and average weekly temperature. It was found that total rainfall with a 4-week *lead* time (-4-week lag time) was most strongly associated with malaria incidence, while average temperature with a 16-week lead time (-16 week lag time) was most strong associated with malaria incidence. Typically, higher rainfall and higher temperature, within a certain limit, were associated with an increase in malaria incidence in CPT-U5. Overall, higher rainfall and higher temperatures were observed in the northern region of Mozambique. However, the northern region had unusually high weekly total rainfall in quarter 1, but did not seem to maintain a high rainfall throughout the year. In fact, in quarter 4, it appeared that the central and even portions of the southern region had higher weekly total rainfall compared to the northern region. This indicated that the northern region may have more erratic and extreme climate patterns, which may give rise to a more optimal environment for malaria transmission.

There are a few key findings that were unexpected in the exploration of this data. First, there were two districts found that had an unusually high malaria incidence across all four annual quarters: the Meluco district in the Cabo Delgado province and the Nipepe district in the Niassa province. The reasons why this may be occuring should be further explored in future research. Second, the time series plots and seasonal plots indicate that malaria incidence in CPT-U5 is on the rise. It was also discovered that rainfall and temperature in annual quarters 1 and 4 increased from 2010 to 2016. The relationship between these climatic conditions may be somewhat sensitive, in that small changes in the climatic conditions could result in potentially drastic increases in malaria incidence.

From the results presented in this report, it is apparent that malaria transmission preventative resources, such as bednets, indoor residual spraying, and educating citizens on the importance of these resources, should be more effectively allocated to those in the northern region. These results may also call attention to the fact that those in the northern region may lack access to necessary health care resources. It was shown that the Maputo province overall had the lowest malaria incidence. While geography and climate drive malaria incdicence, urbanization drives it as well. Those living in or proximal to more urbanized areas are likely to have more available health care access, which is also a potential factor to consider along with other spatial, temporal, and climatic factors.

The next steps I would recommend would be to find a way to formally model this data to more formally explain the trends observed in this data. A linear mixed model with a random intercept for district was fit, the outcome was CPT-U5, and the covariates included in the model were 1.) 4-week leading total rainfall, 2.) 16-week leading temperature, 3.) annual quarter, 4.) Epiyear, and 5.) Mozambique region. The code for this model can be found in the appendix. From this model, the northern region, on average, had the highest CPT-U5 ($t = 4.56$, $p < 0.0001$). For every one-year increase, CPT-U5 increases on average by 1.33 CPT-U5 ($t = 176.49$, $p < 0.0001$), etc. This model was fit to get an idea of what to expect. However, in the future, it is highly recommended that more time is dedicated to determining what longitudinal model and covariance structure best fits the data. Also, the relationhships between the variables of interest should be further explored to account for potential non-linear, more complex trends between the variables-of interest.

Acknowledgements

Data and spatial mapping plots and corresponding code was provided by Dr. Katie Colborn from the Department of Biostatistics at the University of Colorado Anschutz Medical Campus.

Reproducible Research

All materials for this project can be found on my BIOS6640 GitHub repository using the following URL: https://github.com/piperwilliams/BIOS6640/tree/master/RProject. The exact code used to generate this report can be found using the following URL: https://github.com/piperwilliams/BIOS6640/blob/master/RProject/Code/Final%20Report%20Code/2018-10-23%20Appendix.Rmd.

References

1. Parasites - Malaria. Centers for Disease Control and Prevention. https://www.cdc.gov/parasites/malaria/index.html. Published October 9, 2018. Accessed October 26, 2018.

2. Key points: World malaria report 2017. World Health Organization. https://www.who.int/malaria/media/world-malaria-report-2017/en/#Malaria elimination. Published December 14, 2017. Accessed October 26, 2018.

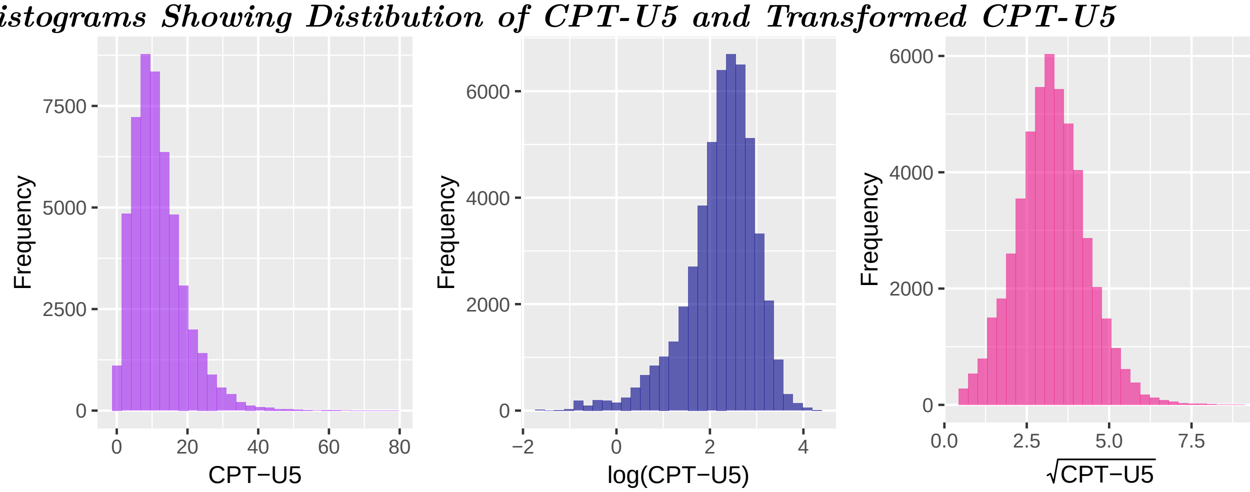
3. Malaria. World Health Organization. http://www.who.int/malaria/en/. Published October 5, 2018. Accessed October 27, 2018.

4. Preventable, Curable and Claiming Lives: Malaria. North Star Alliance. http://www.northstar-alliance.org/preventable-curable-and-claiming-lives-malaria/. Published April 25, 2013. Accessed October 27, 2018.

5. Colborn KL, Giorgi E, Monaghan AJ, et al. Spatio-temporal modelling of weekly malaria incidence in children under 5 for early epidemic detection in Mozambique. Scientific reports. 2018;8(1):9238.

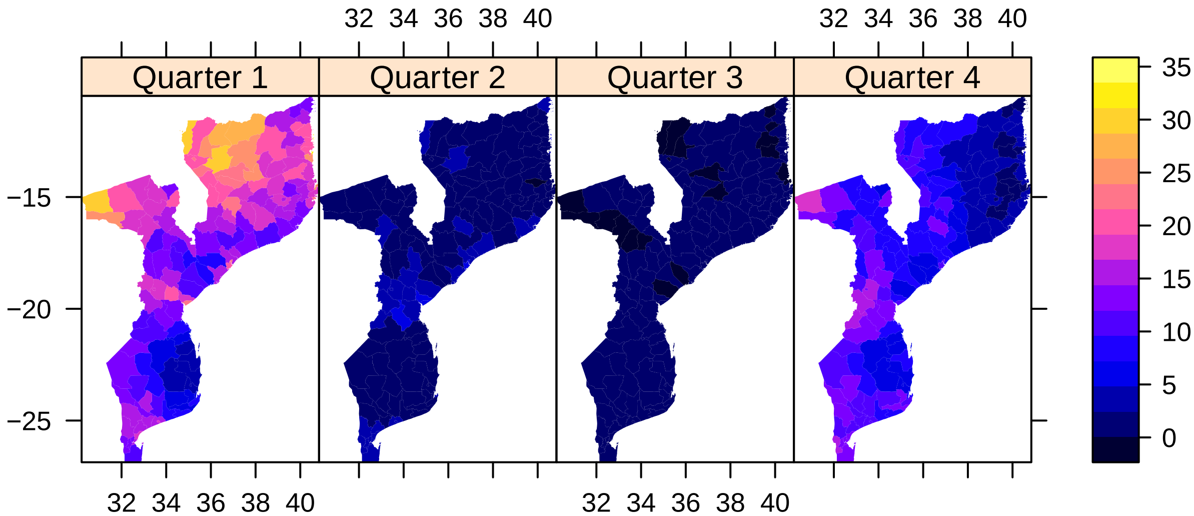
Appendix

Histograms Showing Distibution of CPT-U5 and Transformed CPT-U5

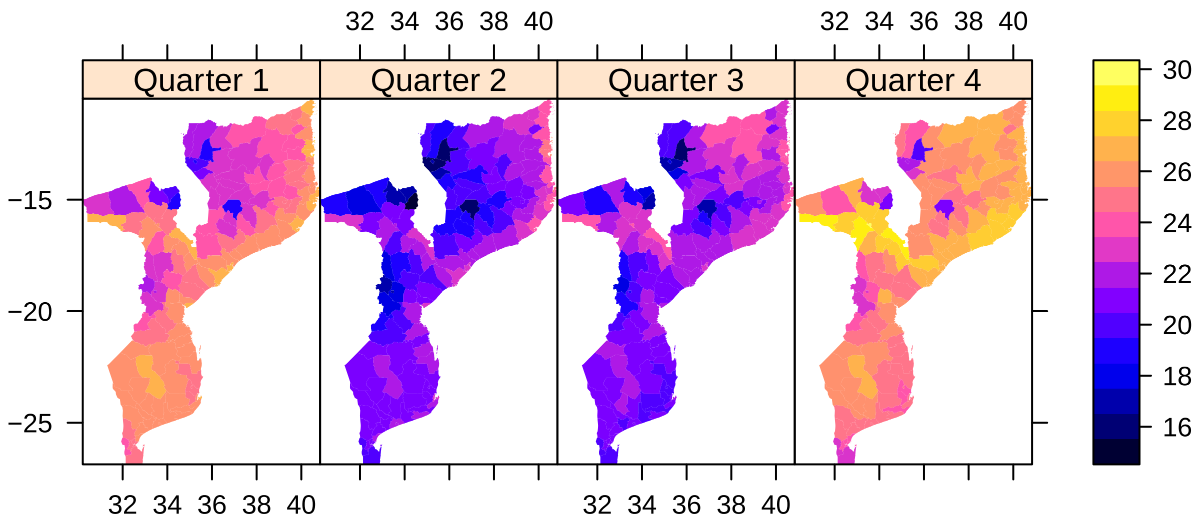


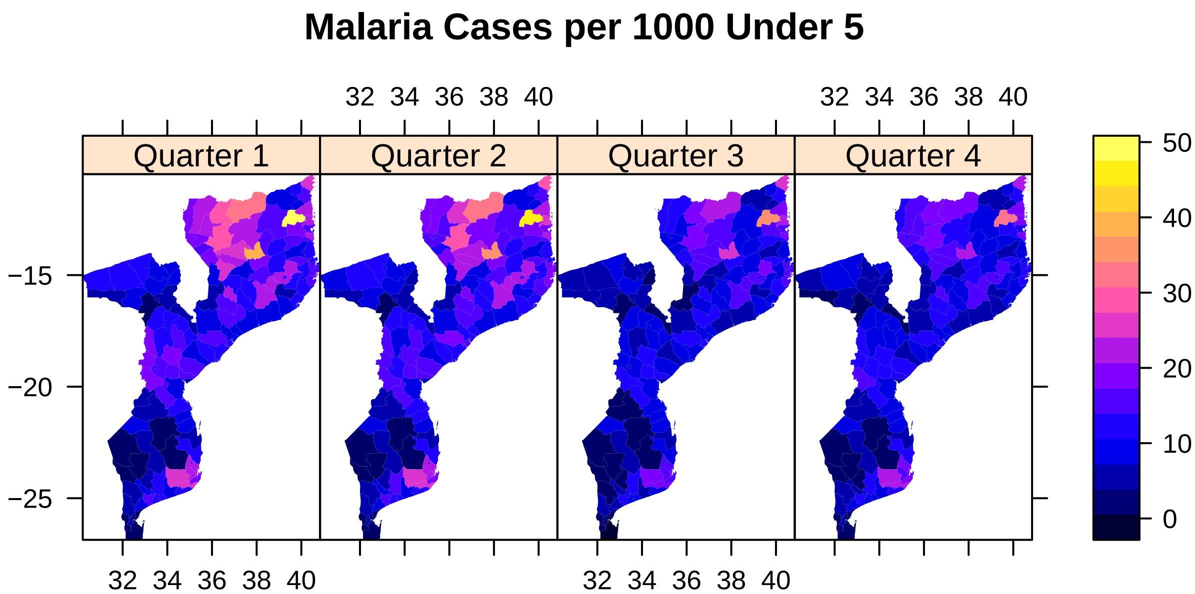
Spatial Maps of Quarterly Rainfall, Temperature, and CPT-U5 Across Districts

Average Total Weekly Rainfall (mm)



Average Temperature (Celcius)





Hypothetical Statistical Model

```
lagdata <- data %>%
  group_by(DISTCODE, Epiyear) %>%
  mutate(rainlead4 = lead(rainTot, 4)) %>%
  mutate(templead16 = lead(tavg, 16))

lagdata <- na.omit(lagdata)

model <- lme(cptu5 ~ rainlead4 + templead16 + Quarter + Epiyear + Region,
             data = lagdata, random = (~ 1 | DISTCODE)) # LMM with random slope
anova(model) # overall F
summary(model) # details of model, bad model fit
```