

**Group F Mid-Project Update**  
**Members: Zhiqi, Fariha, Sarah, Meichen**

**Rationale for removing abundance hypothesis (Hypothesis 1) from the project proposal**

Abundance data at the subspecies level was not available to merge with our species presence dataset. Instead of focusing on abundance, we decided to work on understanding species richness and how that may influence the number of malaria cases.

**Rationale for modifying transmission effectiveness hypothesis (Hypothesis 3)**

We have modified this hypothesis to make it more testable using the available data. Hypothesis 2 targeting species richness aims to identify which subgenera are more effective for transmission. And hypothesis 3 considers whether that correlation is leading to more cases. As such we feel these hypotheses are better suited to our dataset and can still allow us to make meaningful conclusions about malaria in Africa.

**Hypothesis 1**

H<sub>A</sub>: The net latitudinal range of species in the *Anopheles* subgenus is increasing from 1898 to 2016 in Africa.

H<sub>0</sub>: The average absolute latitudinal range of species in the *Anopheles* subgenus is not changing from 1898 to 2016 in Africa.

Prediction: We predict that there may be changes to the average, minimal and maximal latitudes where the *Anopheline* malaria vector species is observed. In this case, net latitudinal range indicates an expansion in both directions from the equator. *Anopheline* mosquitoes are the main vectors for malaria transmission, and we expect that these mosquitoes will expand their range (represented by net latitudinal range), as observed by their presence in more areas across Africa. We predict this given the documented history of Malaria cases in the African continent increasing over the years (Fact Sheet about Malaria, 2022).

**Hypothesis 2**

H<sub>A</sub>: The species richness in *Anopheles* subgenera observed in African regions is decreasing from 2007 to 2016.

H<sub>0</sub>: There is no change over time in the species richness within the *Anopheles* subgenera observed in African regions.

Prediction: If we observe that mosquitoes within the *Anopheles* subgenus are expanding their range latitudinally, we could attribute this effect to climate warming. The literature supports that climate warming would have a more significant effect on temperate regions (with higher latitude) and this can influence the distribution of *Anopheline* mosquitoes (Afrane et al., 2012; Ryan et al., 2020). As such, this change in latitude can exert selection pressure on the *Anopheles* subgenus, which could lead to a decrease in species richness as one or several subgenera may be more effective malaria vectors.

### Hypothesis 3

H<sub>A</sub>: There is a positive correlation between *Anopheles* subgenus species richness at the country level and the number of reported malaria cases.

H<sub>0</sub>: There is no correlation between the *Anopheles* subgenus species richness at the regional level and the number of reported malaria cases.

Prediction: We anticipate that a decrease in *Anopheles* subgenus species richness would result in a greater number of reported malaria cases. A decrease in *Anopheles* subgenus species richness may reflect an overall decrease in the number of mosquitoes that can transmit malaria. However, this decrease in species richness may also reflect the fact that certain subgenera become dominant and contribute to an increased number of malaria cases over time.

## Dataset Description

### Malarial Mosquito Dataset

To test our first and second hypothesis, we use information acquired from the Malarial Mosquito Database on Kaggle, which is a compilation and integration of information from published malaria vector literature, from 1898 to 2016 ([link](#)). Each row represents a published study that's related to malaria vector species in the genus *Anopheles*, and the columns provide detailed information regarding the studies:

- Location where the study took place: Country, Region, Municipality, Latitude and Longitude (sourced from Google Earth, Encarta or GeoNames)
- Start year and end year of the study's species observation period
- Species complex/species/sub-species identified by the study
- Sampling methods used by the study (e.g., Animal Bait Catches, Bed net traps, etc.)<sup>[1]</sup>
- Methods used to identify complex/species/subspecies (e.g., DNA probes, Morphology, PCR)
- Other information, such as the title of the study

### Malaria in Africa Dataset

"Malaria in Africa" contains information regarding malaria occurrences in a number of African countries, along with some data about the preventive measures taken there ([link](#)). This dataset was sourced from the World Bank and covers all African countries over ten years, from 2007 to 2017, and has 595 observations <sup>[2]</sup>. The columns of the dataset are as follows:

- Country name: country of interest
- Year: Year, from 2007 to 2017 inclusive
- Incidence of malaria cases per 1,000 population at risk
- Number of malaria cases reported
- Columns F:X concern country demographic factors and degree of presence of specific malaria control measures

## **Dataset Manipulation**

The analysis will be performed on two datasets. The first dataset was constructed from the “Malarial Mosquito Dataset”, and includes the following columns:

- Location data: Country, Region, Subregion, Latitude and Longitude
- Time data: Start year and end year of the study’s species observation period, from 1898 to 2016
- Species occurrence data: To investigate the latitudinal distribution of the species, we restructured the dataset so that each row is broken into several rows, each of them marking the occurrence of one species, so that we may plot latitude against time while differentiating the species.
- Multiple methods: A boolean variable that identifies whether multiple detection methods were used to observe a subspecies.

We excluded the following variables as they are not of our interest:

- State of the found species (adults/larvae)
- Other species information (sibling species)

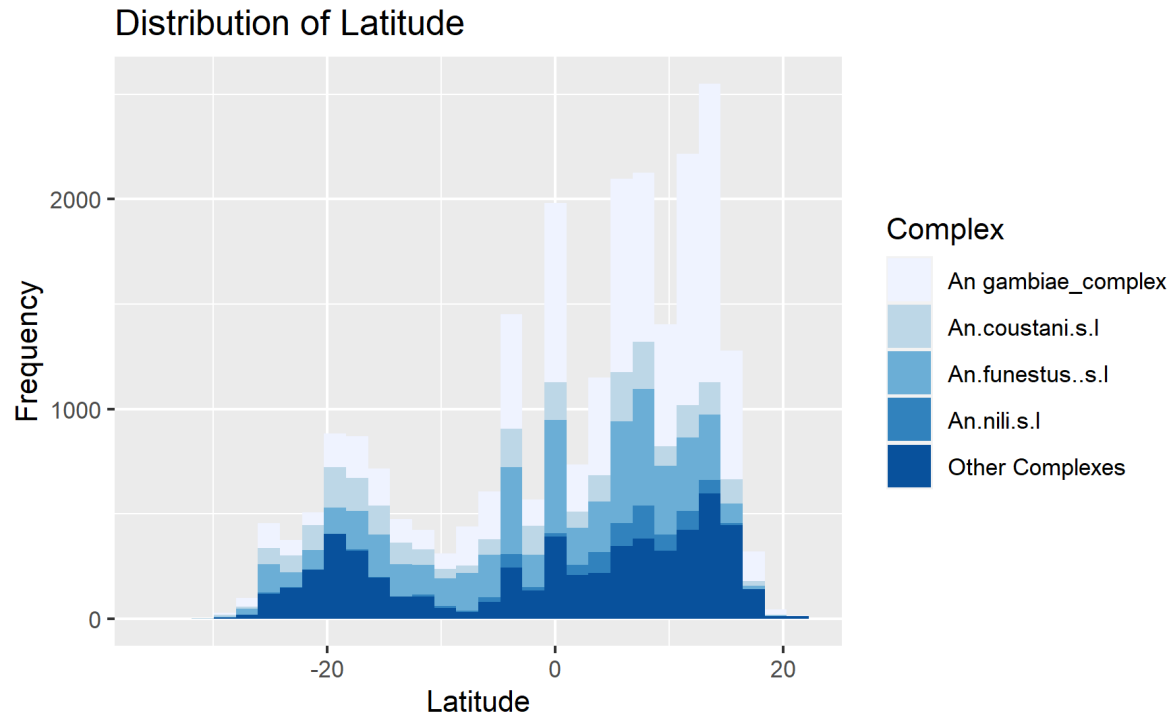
The second dataset used for analysis was constructed by merging information from the “Malarial Mosquito” and “Malaria in Africa” datasets. The datasets were merged by the country and year for which an observation took place. The merged datasets include the following columns:

- Location data: Country, latitude
- Time data: Year of observation, from 2007 to 2017
- Species richness: To measure annual species richness within a country, we calculated the number of different complexes [3] found in each country each year. We used complexes instead of species, since members of a complex are usually morphologically indistinguishable and share numerous ecological properties. Sampling error might have been reduced this way.
- Malaria incidence (the number of malaria cases per 1000 people)
- Number of malaria cases reported (confirmed by examination)
- Rurality: percentage of the country’s total population living in rural areas
- Rural population growth: percentage annual change of a country’s rural population

We excluded the information on preventive measurements due to the large amount of NAs. General dataset properties and species dictionary are in separate files.

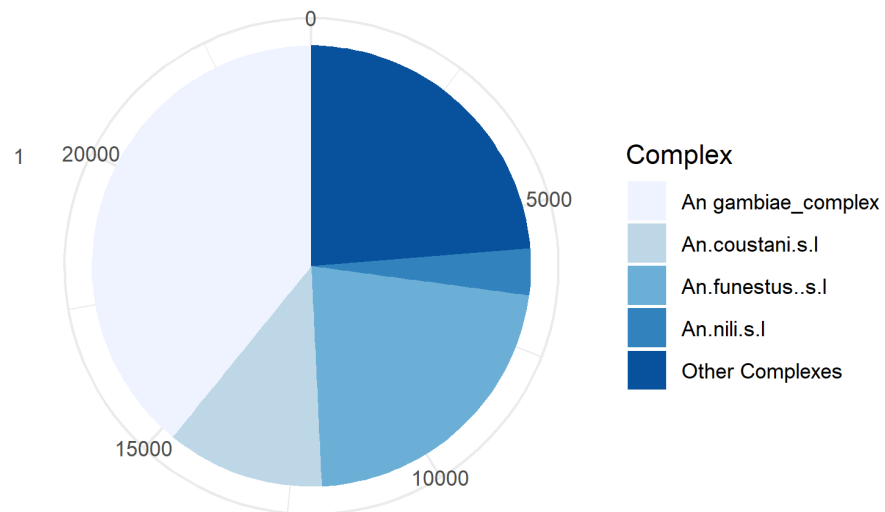
## **Observations from Exploratory Data Analysis**

- A large number of studies were conducted in the 0-20 latitudinal range.

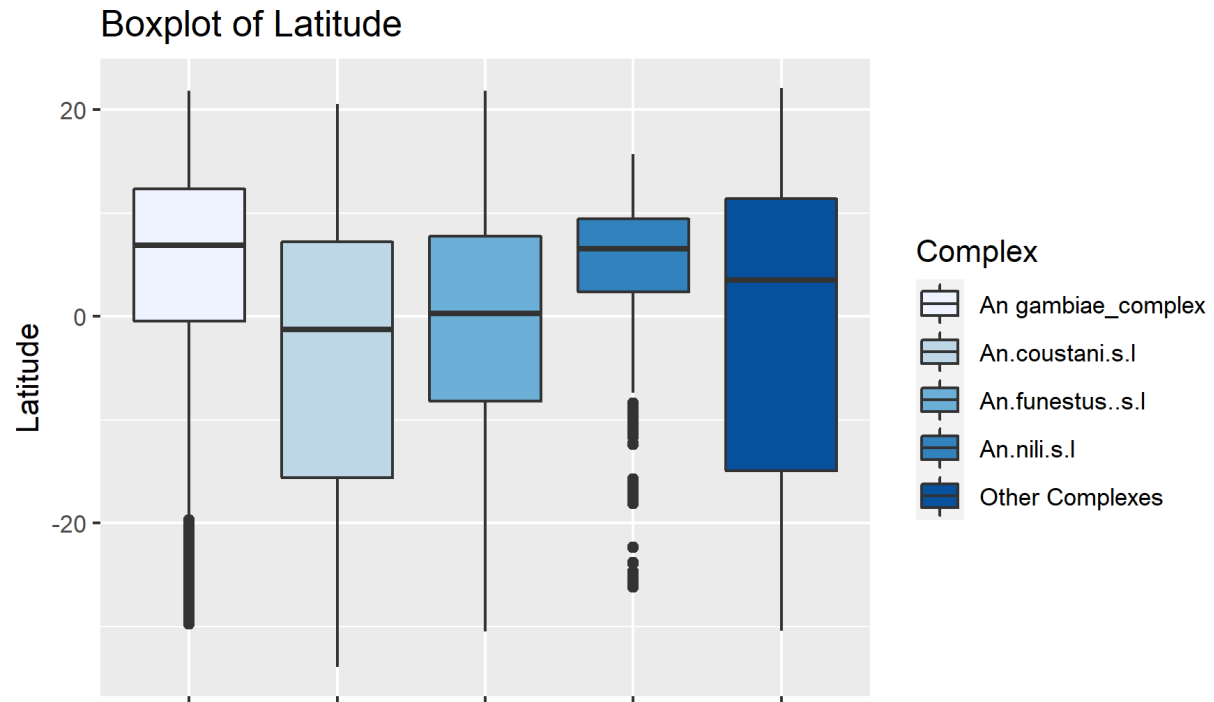


- *An. Gambiae* and *An. Funestus* are 2 very commonly found complexes.

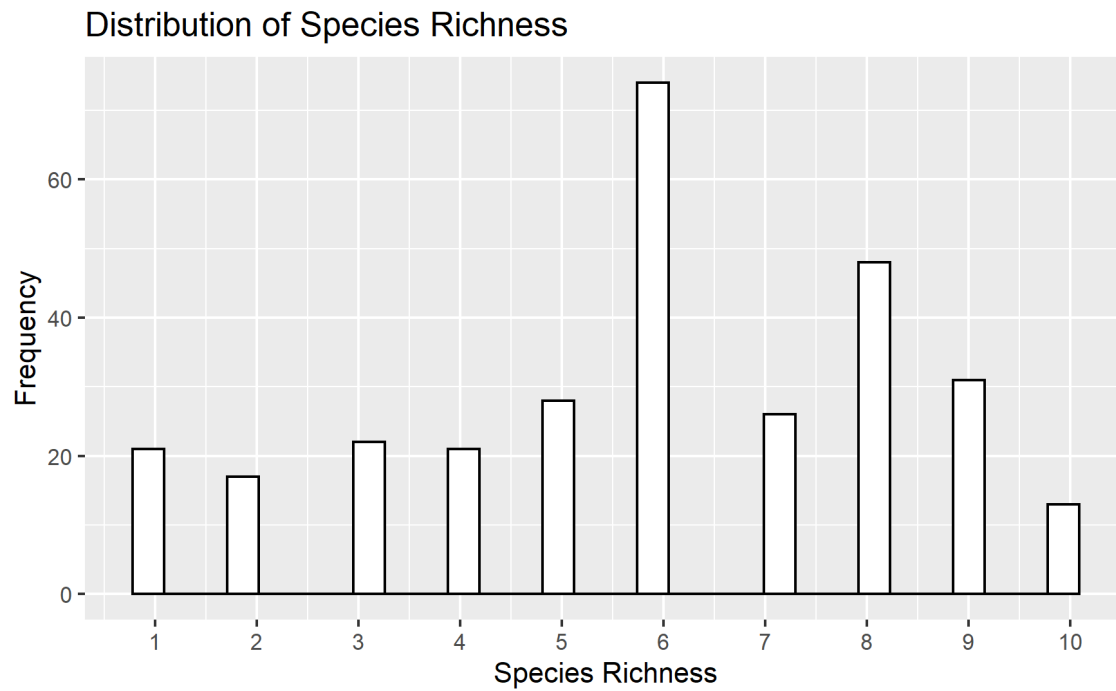
### Occurences of Complexes



- Complexes are found in different latitudinal ranges.

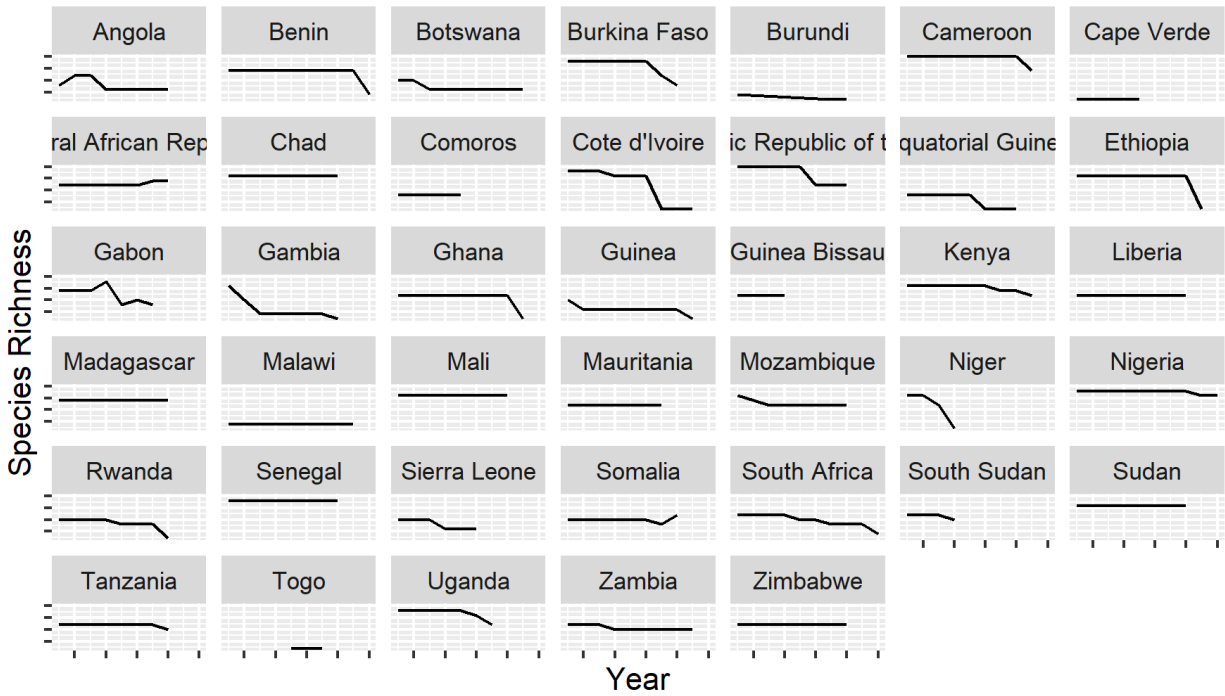


- An intermediate species richness value is more commonly observed.



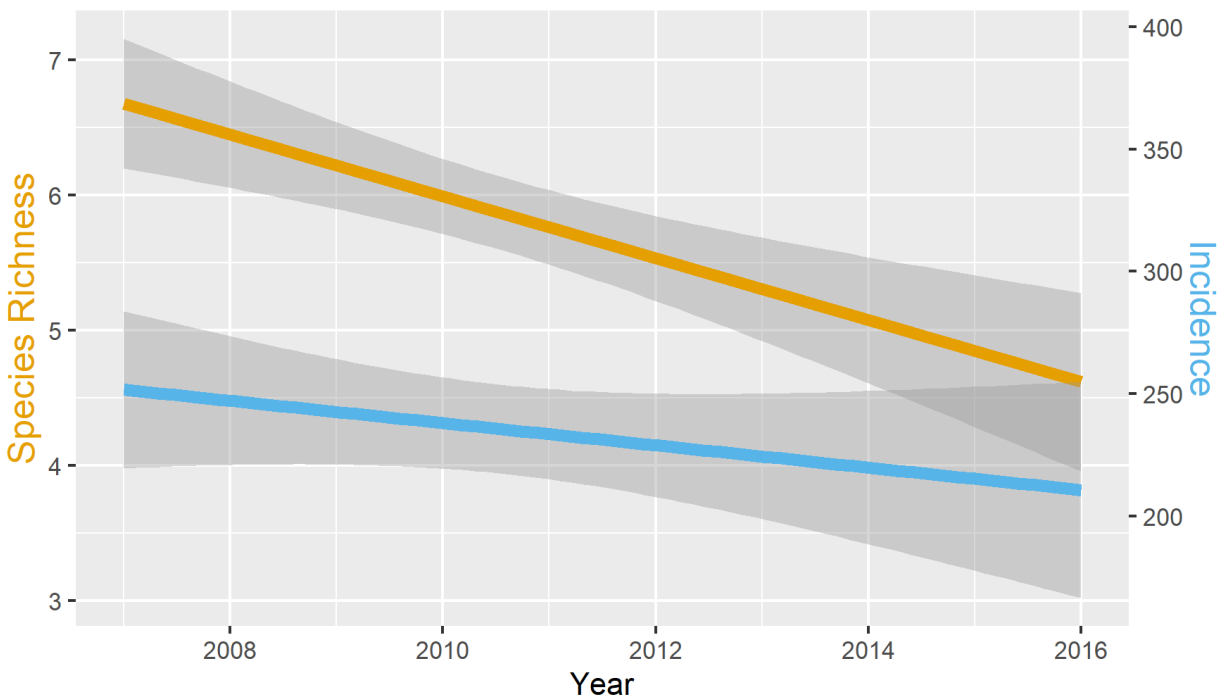
- Species richness showed a decreasing trend over time.

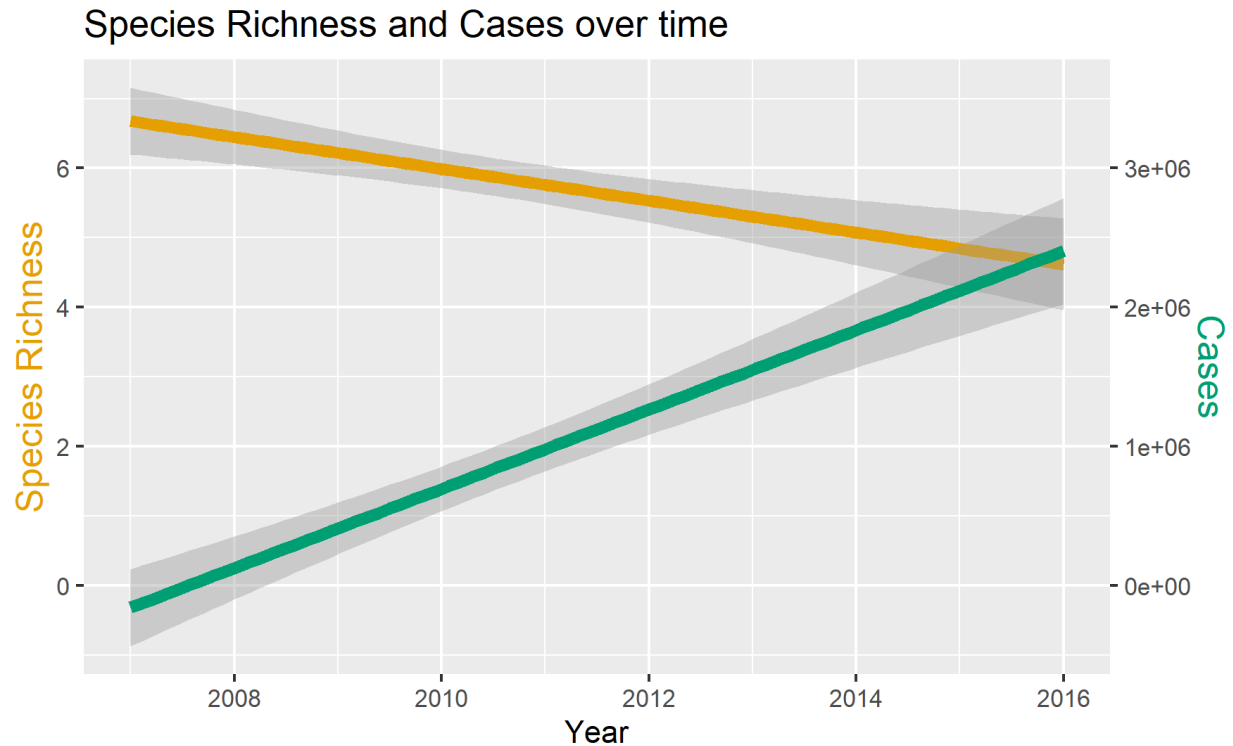
## Species Richness over time



- Malaria incidence in general seems to be decreasing, but the number of reported malaria cases is increasing.

## Species Richness and Incidence over time





## Data Analysis Plan

### Hypothesis 1

Paired t-tests will be used to test hypothesis 1. The observed latitudes at the earliest and the latest time points for each species will be compared by these tests, and the corresponding p-values will provide evidence towards or against the null hypothesis (as the paired t-tests will examine whether the mean difference between pairs of measurements is zero or not).

### Hypothesis 2

A mixed effect model will be used to test hypothesis 2. “Species Richness” will be the response variable, and “time” (Year) will be included as a fixed effect. “Country” will be included as a random effect, and we allow the slope to vary between different countries.

Model output should help us to determine whether the correlation between Species Richness and Time is statistically significant.

### Hypothesis 3

Hypothesis 3 will also be tested with a mixed effect model. In this case, the number of reported Malaria cases will be the response variable, and Species Richness will be included as a fixed effect. Country and Time will be included as random effects and we allow the slope to vary among different countries. Rurality (represented by Rural Percentage and/or Rural Growth

Percentage) might also be included as random effects, to better account for the differences between countries.

We will likely fit multiple models (and test more predictor variables such as preventive measures) and evaluate them to find a final model that best represents the scenario. Output from that model should provide insights into the significance of the correlation between Species Richness and the number of reported Malaria cases.

---

[1] A detailed list of these methods can be found in the accompanying key of this dataset.

[2] “The data on the incidence of malaria, malaria cases reported, and preventive treatments against malaria have been retrieved from the world bank open data source.”

[3] Informal taxonomic level (below sub-genus and above species). Detailed taxonomic structure listed in the species dictionary.

## References

Afrane, Y. A., Githeko, A. K., & Yan, G. (2012). The ecology of Anopheles mosquitoes under climate change: Case studies from the effects of deforestation in East African highlands. *Annals of the New York Academy of Sciences*, 1249, 204–210.  
<https://doi.org/10.1111/j.1749-6632.2011.06432.x>

*Fact sheet about malaria*. (2022, July 26). World Health Organization.  
<https://www.who.int/news-room/fact-sheets/detail/malaria>

Ryan, S. J., Lippi, C. A., & Zermoglio, F. (2020). Shifting transmission risk for malaria in Africa with climate change: A framework for planning and intervention. *Malaria Journal*, 19(1), 170.  
<https://doi.org/10.1186/s12936-020-03224-6>