

Exercise 1

In this exercise, you will use the `tz_malaria` data-set from the `RiskMap` package. Run the code below to load the data-set.

```
library(RiskMap)
data("tz_malaria")
```

For more information on the data see the help page, which can be accessed by entering `?tz_malaria`.

1. Using the empirical logit of malaria prevalence (`Pf` is the variable of the cases and `Ex` is the number of examined individuals) and suitable graphical tools, explore the relationship between the empirical logit and the following variables:
 - Temperature
 - Precipitation
 - EVI (Enhanced Vegetation Index)
 - NTL (Night-Time Lights) Are the relationships linear? How would you enter this variable when using a logistic regression?
2. Fit a binomial GLM to model malaria prevalence as a quadratic function in temperature and then answer the following questions.
 - Write a function to calculate the peak temperature at which malaria prevalence is highest.
 - (Optional) Perform parametric bootstrapping to estimate the 95% confidence interval for the peak temperature. Print the peak temperature estimate and its confidence interval from the bootstrap.
 - Fit a linear spline model to estimate the change point in temperature where the relationship with malaria prevalence changes. Based on the AIC, between the linear spline and quadratic transformation of temperature, which provides a better fit to the data?
3. Extend the linear spline model that includes temperature to also include EVI as a predictor and fit a generalized linear mixed-effects model (GLMM) with a random intercept for `cluster.number`. Is there evidence of over-dispersion?
 - Load the raster files for temperature and EVI using the following code.

```
# Read and project raster files
r_temp <- rast("path/to/local/Tanzania_Annual_LST_2015.tif")
r_temp <- terra::project(r_temp, "EPSG:32736")
r_evi <- rast("path/to/local/Tanzania_Annual_EVI_2015.tif")
r_evi <- terra::project(r_evi, "EPSG:32736")
```

4. Using the fitted GLMM, create a raster of the predicted malaria prevalence. To do this, consider proceeding through the following steps: 1) load the boundaries for Tanzania; 2) create a predictive grid of the desired resolution; 3) extract the covariates of temperature and EVI over the grid; 4) predict malaria prevalence using the fitted GLMM over the grid; 5) convert the predictions into a raster and visualize the results. What patterns do you observe in the predicted prevalence?