Malaria incidence prediction in Afghanistan

Xingyan

4/28/2022

```r
library(rgdal)
library(raster)
library(ggplot2)
library(spatstat)
library(plotrix)
library(fields)
library(leaflet)
library(maptools)
library(RColorBrewer)
library(lattice)
library(geoR)
library(plotrix)
library(car)
library(sp)

# Moran's I and spatial dependencies
library(spdep) # Spatial Dependence: Weighting Schemes, Statistics and Models
library(ape) # Analyses of Phylogenetics and Evolution
library(pgirmess) # Data Analysis in Ecology

# Libraries for point processes
library(spatstat)
library(splancs) # K-function
library(smacpod) # Spatial scanning statistic
```

Introduction

Malaria is one of the significant public health concerns around the world. It is a severe febrile illness caused by a parasite that commonly infects a specific type of mosquito which feeds on humans1. Common symptoms of Malaria include fevers and shaking chills. If left untreated, it can progress to severe illness and even death within 24 hours. Today, nearly half of the world's population was at risk of Malaria. People under high risk of Malaria include infants, children under 5, pregnant women, patients with HIV/AIDS, and migrant workers with low immunity.

Afghanistan has the world's third-highest malaria burden. The incidence rate of confirmed malaria cases was 8 per 1,000, and the slide positivity rate was 17.64%. A total number of 123 districts are at high risk of malaria endemic. Temperature and precipitation are significant predictors of malaria incidence. While temperature influences the life cycle of the parasite and mosquito, precipitation provides a place for mosquito fertilization and breeding 4. Understanding the spatial pattern of malaria outbreaks is necessary for future prevention, given

Malaria's environmental mediated transmission nature. This analysis focuses on the relationship between average yearly temperature and precipitation and the malaria prevalence rate in Afghanistan and aims to predict the prevalence of Malaria based on these two geographic characteristics.

Given the environmental reservoir of Malaria, this analysis is based on the hypothesis that higher prevalence will be clustered in the study area rather than evenly spread.

Methods

The dataset is part of a national malaria indicator assessment conducted in 2008, including age range, the number of individuals who tested positive, and the number of individuals tested in 157 areas in Afghanistan. The location of each region is geo-referenced. Prevalence was calculated by dividing the number of individuals testing positive by the number of individuals tested.

This analysis features a Moran's I statistic to detect local and global clustering. Both statistical tests were performed on malaria prevalence in all 157 areas. Moran's I statistic used the logit function to transform prevalence. In addition to detecting local and global clustering, a spatial regression analysis was performed using a Generalized Linear Model and a Generalized Additive Model with a binomial logit link function to examine the relationship between average yearly temperature and precipitation and malaria prevalence. The environmental covariates were obtained from WorldClim's Bioclimatic variables. The mean annual temperature is coded as BIO1, and annual precipitation is coded as BIO12.

Exploration

Import datasets

```
AFG_malaria_data <- read.csv("MalariaAfghanistan.csv")
AFG_Adm_1 <- raster::getData("GADM", country="AFG", level=1)
#set the projection to WGS84
proj4string(AFG_Adm_1) <- CRS('+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84
=0,0,0 ')
```

```
summary(AFG_malaria_data)
```

```
##        X                id           citation1              latitude
##  Min.   :  1    Min.   :803299   Length:157          Min.   :30.17
##  1st Qu.: 40    1st Qu.:806879   Class :character    1st Qu.:34.28
##  Median : 79    Median :812131   Mode  :character    Median :34.54
##  Mean   : 79    Mean   :812259                       Mean   :34.65
##  3rd Qu.:118    3rd Qu.:816608                       3rd Qu.:35.92
##  Max.   :157    Max.   :822023                       Max.   :38.32
##    longitude        lower.age        upper.age         pf.pos
##  Min.   :61.06   Min.   : 0.000   Min.   :35.00   Min.   :0.00000
##  1st Qu.:64.91   1st Qu.: 1.000   1st Qu.:57.00   1st Qu.:0.00000
##  Median :68.47   Median : 2.000   Median :65.00   Median :0.00000
##  Mean   :67.35   Mean   : 5.382   Mean   :64.21   Mean   :0.08917
##  3rd Qu.:70.51   3rd Qu.:12.000   3rd Qu.:70.00   3rd Qu.:0.00000
##  Max.   :71.47   Max.   :22.000   Max.   :99.00   Max.   :3.00000
##     examined
##  Min.   :  4.00
##  1st Qu.: 20.00
##  Median : 31.00
##  Mean   : 42.76
##  3rd Qu.: 52.00
##  Max.   :244.00
```

Prevalence of Malaria in Afghanistan

Visualizing Malaria prevalence with different colorations gives an overview of possible clustering of higher prevalence in the eastern side of Afghanistan.

```
#Get the prevalence rate of malaria
AFG_malaria_data$prevalence = AFG_malaria_data$pf.pos/AFG_malaria_data$examined

#Explore if there is any clustering.
#We can see that malaria cases may cluster in the western side of Afghanistan
pal = colorNumeric("Oranges", AFG_malaria_data$prevalence)
leaflet(AFG_malaria_data) %>% addTiles() %>% addCircleMarkers(~longitude, ~latitude, fil
lOpacity=1,
                                        fillColor= ~pal(prevalenc
e), radius=~prevalence*200, stroke=TRUE, weight=1) %>%
  addLegend(pal = pal, values = ~prevalence)
```

Leaflet (https://leafletjs.com) | © OpenStreetMap (http://openstreetmap.org) contributors, CC-BY-SA
(http://creativecommons.org/licenses/by-sa/2.0/)

```
#Generate the histgram and explore the distribution of prevalence rate of malaria
hist(AFG_malaria_data$prevalence, xlab = "Prevalence", main = "")
```
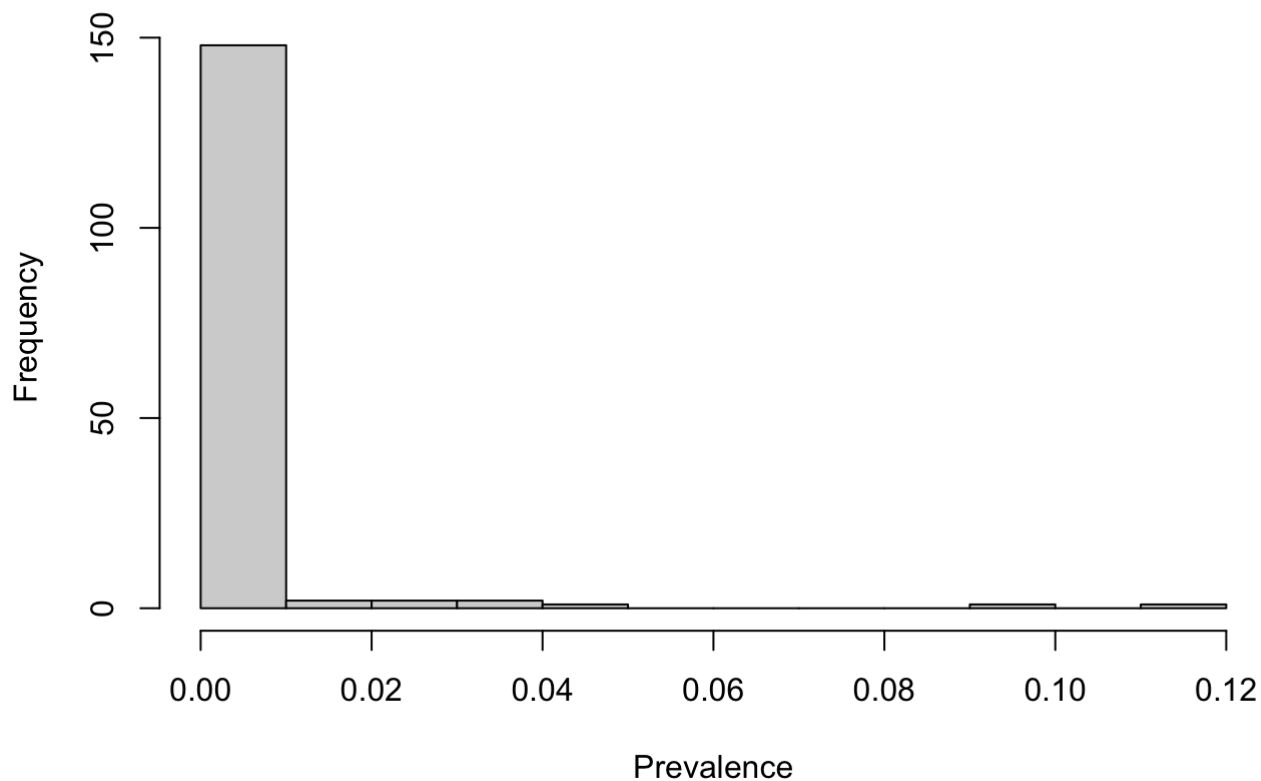


```
#The distribution of prevalence rate is left skewed
#Use the logit transformation
AFG_malaria_data$log_odds <- logit(AFG_malaria_data$prevalence + 0.00001)
hist(AFG_malaria_data$log_odds, xlab = "Log odds", main = "")
```
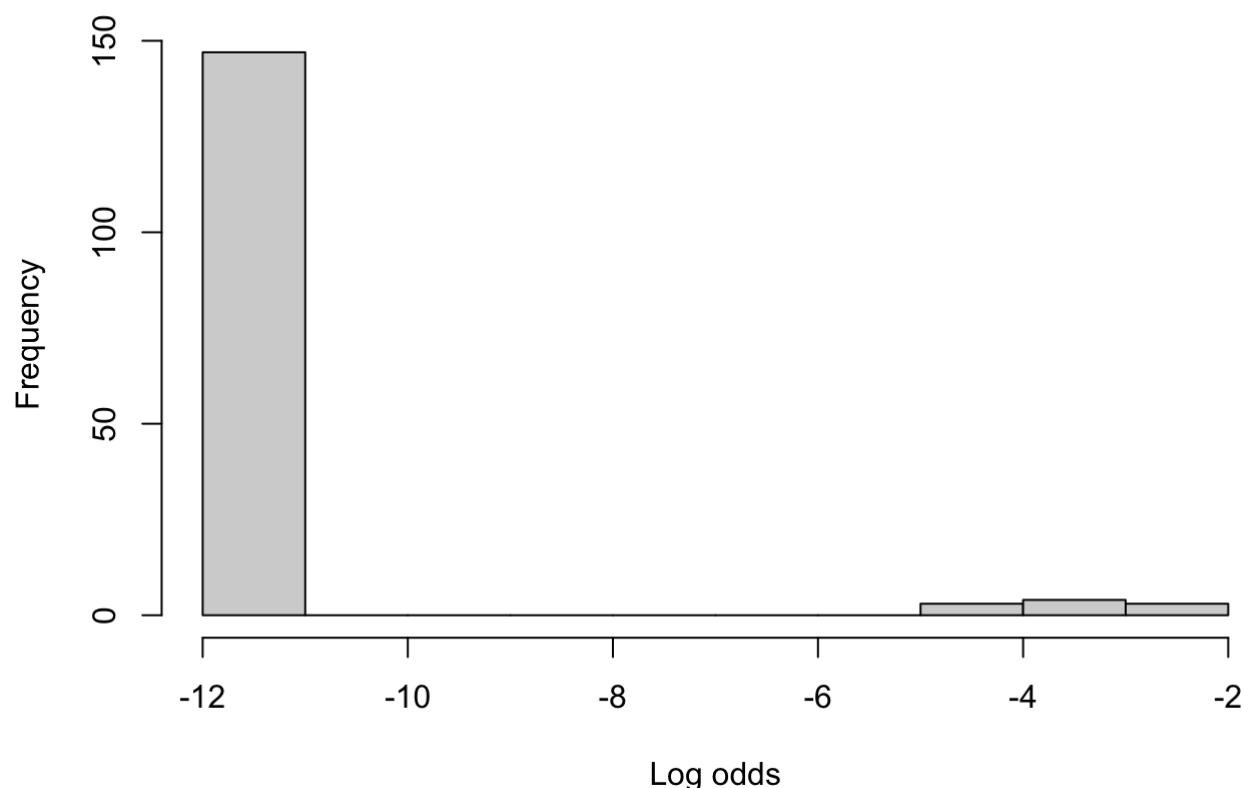
Global spatial autocorrelation

Create a correlogram to explore Moran's I over different spatial lags.

```
maxDist<-max(dist(cbind(AFG_malaria_data$longitude, AFG_malaria_data$latitude)))
maxDist
```

```
## [1] 12.02951
```

Results

The Moran's I statistic had a statistically significant p-value of 0, which indicates that global clustering does occur in the study area. After computing Moran's coefficients on distance classes, we can see significant positive clustering at less than one decimal degree

```
xy=cbind(AFG_malaria_data$longitude, AFG_malaria_data$latitude)
pgi.cor <- correlog(coords=xy, z=AFG_malaria_data$log_odds, method="Moran", nbclass=10)
plot(pgi.cor) # statistically significant values (p<0.05) are plotted in red
```

Moran I statistic = f(distance classes)

distance classes
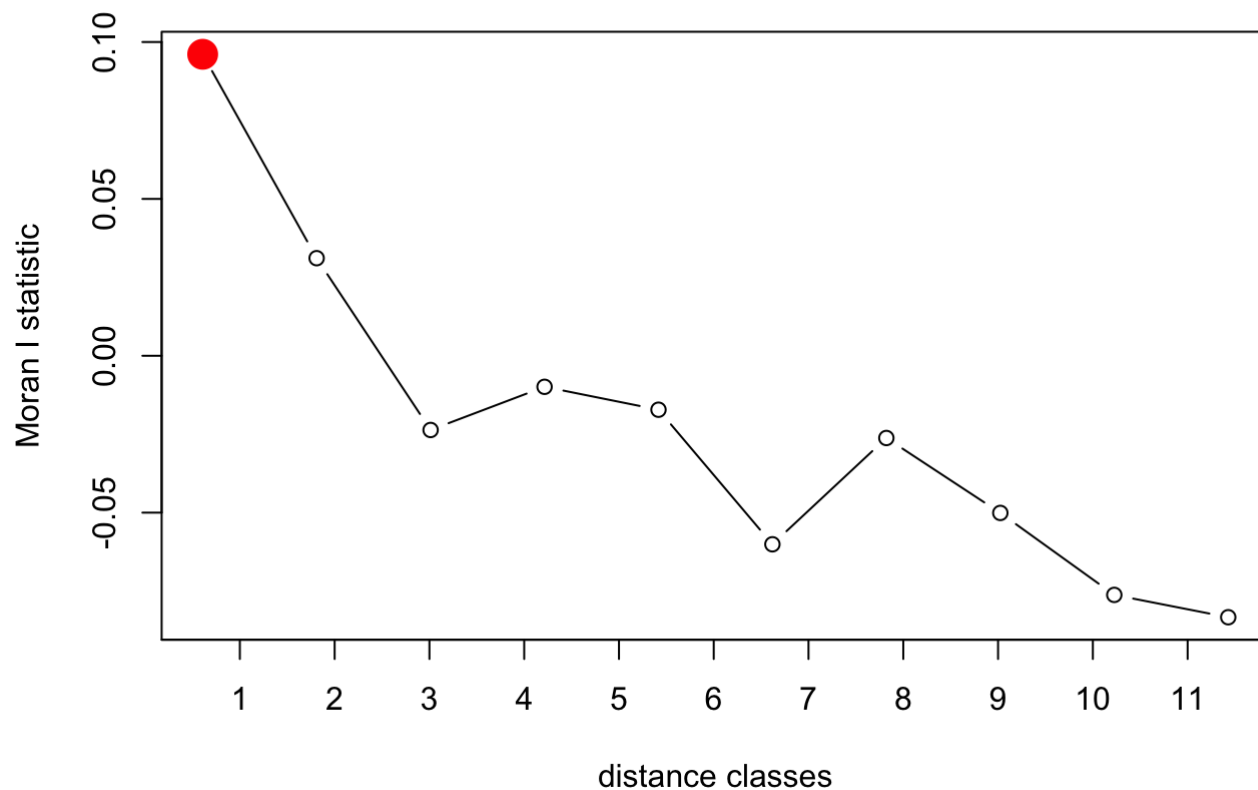
```
pgi.cor# distclass is midpoint for the bin
```

```
## Moran I statistic
##       dist.class         coef      p.value    n
##  [1,]  0.6077247  0.096091678 3.013879e-05 4052
##  [2,]  1.8100182  0.031109049 1.897891e-01 1706
##  [3,]  3.0123112 -0.023654969 7.926363e-01 4692
##  [4,]  4.2146042 -0.009892164 5.544380e-01 2988
##  [5,]  5.4168972 -0.017196321 7.067318e-01 4618
##  [6,]  6.6191902 -0.060087641 9.521840e-01 2410
##  [7,]  7.8214832 -0.026210259 6.994483e-01 1596
##  [8,]  9.0237762 -0.050090857 8.854457e-01 1968
##  [9,] 10.2260692 -0.076213203 7.888801e-01  396
## [10,] 11.4283622 -0.083333333 5.476178e-01   66
```

```r
coords<-coordinates(xy) # set spatial coordinates to create a spatial object
IDs<-row.names(as.data.frame(coords))

Neigh_nb<-knn2nb(knearneigh(coords, k=1, longlat = TRUE), row.names=IDs)     # using the
"spdep" package
# assigns at least one neighbor to each and calculates the distances between
dsts<-unlist(nbdists(Neigh_nb,coords)) # returns the distance between nearest neighbors
for each point
summary(dsts)
```

```
##      Min.  1st Qu.    Median      Mean  3rd Qu.      Max.
## 0.006578 0.027006 0.049318 0.104121 0.118081 0.776602
```

```r
max_1nn<-max(dsts)
max_1nn
```

```
## [1] 0.7766024
```

```r
# We create different neighbor structures based upon distance
Neigh_kd1<-dnearneigh(coords,d1=0, d2=max_1nn, row.names=IDs)   # neighbors within maxim
um distance
Neigh_kd2<-dnearneigh(coords,d1=0, d2=2*max_1nn, row.names=IDs) # neighbors within 2X ma
ximum distance

nb_1<-list(d1=Neigh_kd1, d2=Neigh_kd2) # list of neighbor structures
sapply(nb_1, function(x) is.symmetric.nb(x, verbose=F, force=T))
```

```
##   d1   d2
## TRUE TRUE
```

```r
#To run a spatial test for clustering, I assign weights to the neighbor list. I will use
the neighbor structure with all neighbors within the maximum neighbor distance between a
ny two points.
weights<-nb2listw(Neigh_kd1, style="W")   # row standardized binary weights, using minim
um distance for one neighbor
weights
```

```
## Characteristics of weights list object:
## Neighbour list object:
## Number of regions: 157
## Number of nonzero links: 3344
## Percentage nonzero weights: 13.56647
## Average number of links: 21.29936
##
## Weights style: W
## Weights constants summary:
##     n    nn  S0        S1        S2
## W 157 24649 157 36.01849 634.1175
```

```
set.seed(1234)
bperm<-moran.mc(AFG_malaria_data$log_odds , listw=weights,nsim=999)
bperm
```

```
##
##  Monte-Carlo simulation of Moran I
##
## data:  AFG_malaria_data$log_odds
## weights: weights
## number of simulations + 1: 1000
##
## statistic = 0.1156, observed rank = 984, p-value = 0.016
## alternative hypothesis: greater
```

```
par(mfrow=c(1,1), mar= c(5, 4, 4, 2))
hist(bperm$res, freq=T, breaks=20, xlab="Simulated Moran's I")
abline(v=0.15, col="red")
```

Histogram of bperm$res

Simulated Moran's I
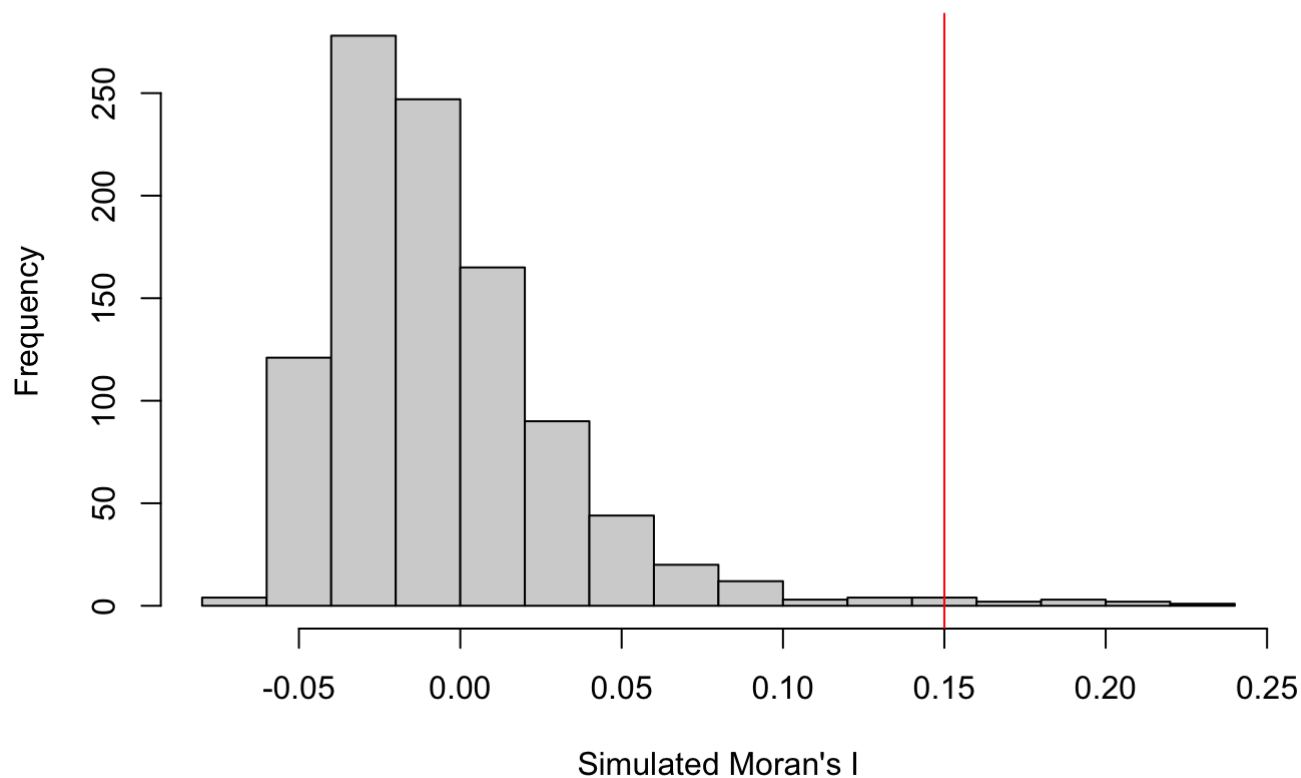
Local Clustering Analysis

```
I <-localmoran(AFG_malaria_data$log_odds, weights)
Coef<-printCoefmat(data.frame(I[IDs,], row.names=row.names(coords),
                              check.names=FALSE))
```

```
##                  Ii         E.Ii      Var.Ii    Z.Ii Pr(z != E(Ii))
##   [1,] -0.10875501 -0.00043157  0.00111224 -3.2481      0.0011620 **
##   [2,] -0.09185515 -0.00043157  0.00118601 -2.6547      0.0079380 **
##   [3,]  0.06732496 -0.00043157  0.01660411  0.5258      0.5990080
##   [4,]  0.06732496 -0.00043157  0.00410733  1.0572      0.2904047
##   [5,] -0.12230455 -0.00043157  0.00104488 -3.7703      0.0001631 ***
##   [6,]  0.06732496 -0.00043157  0.00280897  1.2784      0.2010968
##   [7,]  0.06732496 -0.00043157  0.00280897  1.2784      0.2010968
##   [8,]  0.06732496 -0.00043157  0.00280897  1.2784      0.2010968
##   [9,] -0.12651854 -0.00043157  0.00107781 -3.8406      0.0001227 ***
##  [10,]  0.06732496 -0.00043157  0.00410733  1.0572      0.2904047
##  [11,]  0.06732496 -0.00043157  0.00410733  1.0572      0.2904047
##  [12,]  0.06732496 -0.00043157  0.00575980  0.8928      0.3719716
##  [13,]  0.06732496 -0.00043157  0.00443192  1.0178      0.3087813
##  [14,] -0.11826988 -0.00043157  0.00101335 -3.7017      0.0002141 ***
##  [15,]  0.06732496 -0.00043157  0.00410733  1.0572      0.2904047
##  [16,] -0.11826988 -0.00043157  0.00101335 -3.7017      0.0002141 ***
##  [17,] -0.09573760 -0.00043157  0.00122559 -2.7224      0.0064815 **
##  [18,]  1.41585916 -0.09460465  0.20120896  3.3673      0.0007590 ***
##  [19,] -0.11826988 -0.00043157  0.00101335 -3.7017      0.0002141 ***
##  [20,] -0.15273066 -0.00043157  0.01660411 -1.1819      0.2372359
##  [21,]  0.06732496 -0.00043157  0.02228446  0.4539      0.6499084
##  [22,]  0.06732496 -0.00043157  0.00808358  0.7536      0.4510807
##  [23,] -0.10875501 -0.00043157  0.00111224 -3.2481      0.0011620 **
##  [24,] -0.16901154 -0.00043157  0.00122559 -4.8154      1.469e-06 ***
##  [25,] -0.11826988 -0.00043157  0.00101335 -3.7017      0.0002141 ***
##  [26,] -0.10871954 -0.00043157  0.01319590 -0.9427      0.3458486
##  [27,] -0.10875501 -0.00043157  0.00111224 -3.2481      0.0011620 **
##  [28,]  0.06732496 -0.00043157  0.01319590  0.5898      0.5553002
##  [29,] -0.11284990 -0.00043157  0.00114826 -3.3175      0.0009081 ***
##  [30,] -0.09981416 -0.00043157  0.00126716 -2.7919      0.0052404 **
##  [31,]  2.11424680 -0.09419364  0.24240350  4.4856      7.272e-06 ***
##  [32,]  0.06732496 -0.00043157  0.00410733  1.0572      0.2904047
##  [33,]  0.06732496 -0.00043157  0.01660411  0.5258      0.5990080
##  [34,]  0.06732496 -0.00043157  0.00808358  0.7536      0.4510807
##  [35,]  0.06732496 -0.00043157  0.01660411  0.5258      0.5990080
##  [36,]  0.06732496 -0.00043157  0.00280897  1.2784      0.2010968
##  [37,] -0.16338448 -0.00043157  0.00118601 -4.7317      2.226e-06 ***
##  [38,]  0.06732496 -0.00043157  0.01660411  0.5258      0.5990080
##  [39,]  0.06732496 -0.00043157  0.06772729  0.2604      0.7945884
##  [40,]  0.06732496 -0.00043157  0.00637947  0.8483      0.3962608
##  [41,]  0.06732496 -0.00043157  0.00410733  1.0572      0.2904047
##  [42,]  0.06732496 -0.00043157  0.00637947  0.8483      0.3962608
##  [43,]  0.06732496 -0.00043157  0.02228446  0.4539      0.6499084
##  [44,]  0.06732496 -0.00043157  0.00713685  0.8020      0.4225280
##  [45,]  0.06732496 -0.00043157  0.00410733  1.0572      0.2904047
##  [46,] -0.10875501 -0.00043157  0.00111224 -3.2481      0.0011620 **
##  [47,]  0.06732496 -0.00043157  0.00280897  1.2784      0.2010968
##  [48,]  0.06732496 -0.00043157  0.00480645  0.9773      0.3284085
##  [49,]  0.06732496 -0.00043157  0.00443192  1.0178      0.3087813
##  [50,] -0.24524912 -0.00043157  0.00176190 -5.8325      5.461e-09 ***
##  [51,]  3.64397078 -0.07703980  1.17636179  3.4308      0.0006019 ***
```

```
##  [52,] -0.11826988 -0.00043157  0.00101335 -3.7017  0.0002141 ***
##  [53,]  0.06732496 -0.00043157  0.06772729  0.2604  0.7945884
##  [54,]  0.06732496 -0.00043157  0.00280897  1.2784  0.2010968
##  [55,]  0.06732496 -0.00043157  0.00297126  1.2430  0.2138577
##  [56,]  0.06732496 -0.00043157  0.00410733  1.0572  0.2904047
##  [57,]  0.06732496 -0.00043157  0.01660411  0.5258  0.5990080
##  [58,]  0.06732496 -0.00043157  0.00410733  1.0572  0.2904047
##  [59,]  0.06732496 -0.00043157  0.00410733  1.0572  0.2904047
##  [60,]  0.06732496 -0.00043157  0.00808358  0.7536  0.4510807
##  [61,]  0.06732496 -0.00043157  0.00443192  1.0178  0.3087813
##  [62,]  0.06732496 -0.00043157  0.00280897  1.2784  0.2010968
##  [63,]  0.06732496 -0.00043157  0.00266142  1.3134  0.1890512
##  [64,]  0.06732496 -0.00043157  0.00280897  1.2784  0.2010968
##  [65,]  0.06732496 -0.00043157  0.00410733  1.0572  0.2904047
##  [66,]  0.06732496 -0.00043157  0.00410733  1.0572  0.2904047
##  [67,] -0.09573760 -0.00043157  0.00122559 -2.7224  0.0064815 **
##  [68,]  0.06732496 -0.00043157  0.01092376  0.6483  0.5168014
##  [69,]  0.06732496 -0.00043157  0.00266142  1.3134  0.1890512
##  [70,]  0.06732496 -0.00043157  0.02228446  0.4539  0.6499084
##  [71,] -0.11826988 -0.00043157  0.00101335 -3.7017  0.0002141 ***
##  [72,] -0.11826988 -0.00043157  0.00101335 -3.7017  0.0002141 ***
##  [73,]  0.06732496 -0.00043157  0.01660411  0.5258  0.5990080
##  [74,]  1.60353827 -0.13161906  0.26848848  3.3487  0.0008119 ***
##  [75,]  0.06732496 -0.00043157  0.00280897  1.2784  0.2010968
##  [76,]  0.06732496 -0.00043157  0.00637947  0.8483  0.3962608
##  [77,]  0.06732496 -0.00043157  0.00410733  1.0572  0.2904047
##  [78,]  0.06732496 -0.00043157  0.00713685  0.8020  0.4225280
##  [79,] -0.23883575 -0.00043157  0.00199749 -5.3342  9.595e-08 ***
##  [80,]  0.06732496 -0.00043157  0.01092376  0.6483  0.5168014
##  [81,] -0.11826988 -0.00043157  0.00101335 -3.7017  0.0002141 ***
##  [82,]  0.06732496 -0.00043157  0.00280897  1.2784  0.2010968
##  [83,]  0.06732496 -0.00043157  0.00808358  0.7536  0.4510807
##  [84,]  1.60756286 -0.08698905  0.18100520  3.9830  6.805e-05 ***
##  [85,] -0.81289753 -0.06291742  3.04569835 -0.4297  0.6673844
##  [86,]  0.06732496 -0.00043157  0.01319590  0.5898  0.5553002
##  [87,]  0.06732496 -0.00043157  0.00637947  0.8483  0.3962608
##  [88,] -0.11826988 -0.00043157  0.00101335 -3.7017  0.0002141 ***
##  [89,]  0.06732496 -0.00043157  0.06772729  0.2604  0.7945884
##  [90,]  0.06732496 -0.00043157  0.01319590  0.5898  0.5553002
##  [91,]  1.56857159 -0.12393175  0.25504517  3.3514  0.0008042 ***
##  [92,] -0.12230455 -0.00043157  0.00104488 -3.7703  0.0001631 ***
##  [93,]  0.06732496 -0.00043157  0.00637947  0.8483  0.3962608
##  [94,] -0.11826988 -0.00043157  0.00101335 -3.7017  0.0002141 ***
##  [95,]  0.06732496 -0.00043157  0.01092376  0.6483  0.5168014
##  [96,]  0.06732496 -0.00043157  0.01660411  0.5258  0.5990080
##  [97,] -0.14800385 -0.00043157  0.00107781 -4.4950  6.956e-06 ***
##  [98,] -0.10875501 -0.00043157  0.00111224 -3.2481  0.0011620 **
##  [99,]  0.06732496 -0.00043157  0.02228446  0.4539  0.6499084
## [100,]  0.06732496 -0.00043157  0.00524340  0.9357  0.3494189
## [101,]  0.06732496 -0.00043157  0.01319590  0.5898  0.5553002
## [102,] -0.11826988 -0.00043157  0.00101335 -3.7017  0.0002141 ***
## [103,]  0.06732496 -0.00043157  0.00443192  1.0178  0.3087813
```

```
## [104,]   0.06732496 -0.00043157   0.00280897   1.2784       0.2010968
## [105,]   0.06732496 -0.00043157   0.00808358   0.7536       0.4510807
## [106,]   1.52860830 -0.10496742   0.22756012   3.4245       0.0006160 ***
## [107,]   0.06732496 -0.00043157   0.00266142   1.3134       0.1890512
## [108,]   0.06732496 -0.00043157   0.00575980   0.8928       0.3719716
## [109,] -0.09981416 -0.00043157   0.00126716  -2.7919       0.0052404 **
## [110,]   0.06732496 -0.00043157   0.00410733   1.0572       0.2904047
## [111,]   0.06732496 -0.00043157   0.00637947   0.8483       0.3962608
## [112,]   0.06732496 -0.00043157   0.01092376   0.6483       0.5168014
## [113,] -0.37278628 -0.00043157   0.03364517  -2.0300       0.0423568 *
## [114,]   0.06732496 -0.00043157   0.06772729   0.2604       0.7945884
## [115,]   0.06732496 -0.00043157   0.00808358   0.7536       0.4510807
## [116,]   2.65381780 -0.07851990   0.27316486   5.2278       1.715e-07 ***
## [117,]   0.06732496 -0.00043157   0.02228446   0.4539       0.6499084
## [118,]   0.06732496 -0.00043157   0.00524340   0.9357       0.3494189
## [119,] -0.11826988 -0.00043157   0.00101335  -3.7017       0.0002141 ***
## [120,]   0.06732496 -0.00043157   0.00410733   1.0572       0.2904047
## [121,]   0.06732496 -0.00043157   0.02228446   0.4539       0.6499084
## [122,]   0.06732496 -0.00043157   0.00575980   0.8928       0.3719716
## [123,] -0.12651854 -0.00043157   0.00107781  -3.8406       0.0001227 ***
## [124,]   0.06732496 -0.00043157   0.00280897   1.2784       0.2010968
## [125,]   0.06732496 -0.00043157   0.00410733   1.0572       0.2904047
## [126,]   0.06732496 -0.00043157   0.00280897   1.2784       0.2010968
## [127,] -0.10875501 -0.00043157   0.00111224  -3.2481       0.0011620 **
## [128,] -0.11826988 -0.00043157   0.00101335  -3.7017       0.0002141 ***
## [129,]   0.06732496 -0.00043157   0.01319590   0.5898       0.5553002
## [130,]   0.06732496 -0.00043157   0.01319590   0.5898       0.5553002
## [131,]   0.06732496 -0.00043157   0.00297126   1.2430       0.2138577
## [132,]   0.06732496 -0.00043157   0.00280897   1.2784       0.2010968
## [133,]   0.06732496 -0.00043157   0.00930080   0.7026       0.4823223
## [134,] -0.11826988 -0.00043157   0.00101335  -3.7017       0.0002141 ***
## [135,]   0.06732496 -0.00043157   0.01319590   0.5898       0.5553002
## [136,]   1.37745442 -0.08818675   0.18888863   3.3723       0.0007455 ***
## [137,]   0.06732496 -0.00043157   0.00410733   1.0572       0.2904047
## [138,]   0.06732496 -0.00043157   0.00410733   1.0572       0.2904047
## [139,]   0.06732496 -0.00043157   0.00280897   1.2784       0.2010968
## [140,]   0.06732496 -0.00043157   0.00410733   1.0572       0.2904047
## [141,]   0.06732496 -0.00043157   0.00280897   1.2784       0.2010968
## [142,]   0.06732496 -0.00043157   0.00280897   1.2784       0.2010968
## [143,]   0.06732496 -0.00043157   0.00382331   1.0958       0.2731665
## [144,]   0.06732496 -0.00043157   0.00443192   1.0178       0.3087813
## [145,]   0.06732496 -0.00043157   0.01092376   0.6483       0.5168014
## [146,]   0.06732496 -0.00043157   0.00443192   1.0178       0.3087813
## [147,]   0.06732496 -0.00043157   0.00808358   0.7536       0.4510807
## [148,] -0.13454580 -0.00043157   0.00098314  -4.2773       1.892e-05 ***
## [149,]   0.06732496 -0.00043157   0.02228446   0.4539       0.6499084
## [150,] -0.09185515 -0.00043157   0.00118601  -2.6547       0.0079380 **
## [151,] -0.10875501 -0.00043157   0.00111224  -3.2481       0.0011620 **
## [152,]   0.06732496 -0.00043157   0.00637947   0.8483       0.3962608
## [153,]   0.06732496 -0.00043157   0.00808358   0.7536       0.4510807
## [154,]   0.06732496 -0.00043157   0.01660411   0.5258       0.5990080
## [155,]   0.06732496 -0.00043157   0.00930080   0.7026       0.4823223
```

```
## [156,]  -0.09981416 -0.00043157  0.00126716 -2.7919        0.0052404 **
## [157,]  -0.11826988 -0.00043157  0.00101335 -3.7017        0.0002141 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Plot the spatial data against its spatially lagged values (the weighted mean of its ne
ighbors)
nci<-moran.plot(AFG_malaria_data$log_odds, listw=weights,
                xlab="Log prevalence", ylab="Spatially lagged log prev", labels=T, pch=1
6, col="grey")
text(c(3,3, -5,-5),c(0.9, -1.9,0.9,-1.9), c("High-High", "High-Low", "Low-High", "Low-Lo
w"), cex=0.8)
```
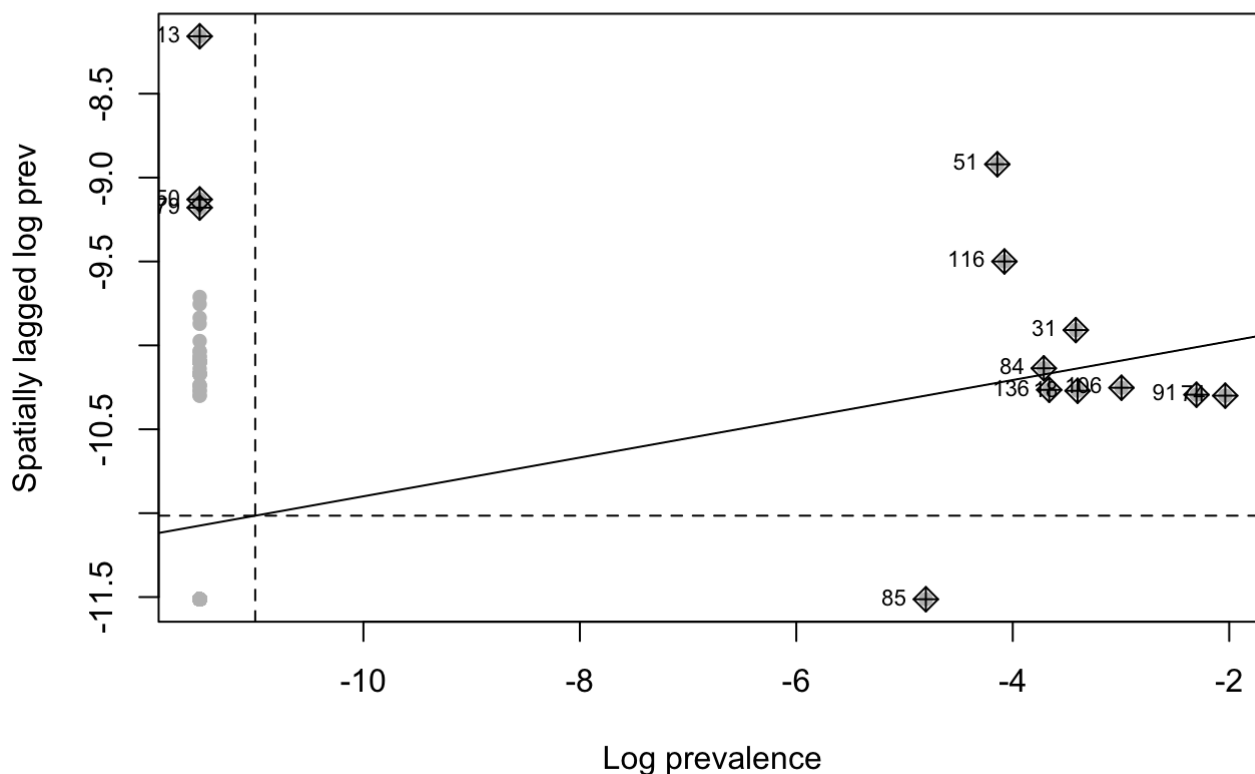


```r
# Map points that are local outliers in the plot
infl<-nci$is_inf==T # find which points are statistically significant outliers
sum(infl==T)
```

```
## [1] 13
```

```
x<-AFG_malaria_data$log_odds
lhx<-cut(x, breaks=c(min(x), mean(x), max(x)), labels=c("L", "H"), include.lowest=T)

wx<-lag(weights,AFG_malaria_data$log_odds)
lhwx<-cut(wx, breaks=c(min(wx), mean(wx), max(wx)), labels=c("L", "H"), include.lowest=
T)
lhlh<-interaction(lhx,lhwx,infl,drop=T)


names<-rep("none", length(lhlh))
names[lhlh=="L.L.TRUE"]<-"LL"
names[lhlh=="H.L.TRUE"]<-"HL"
names[lhlh=="L.H.TRUE"]<-"LH"
names[lhlh=="H.H.TRUE"]<-"HH"
```
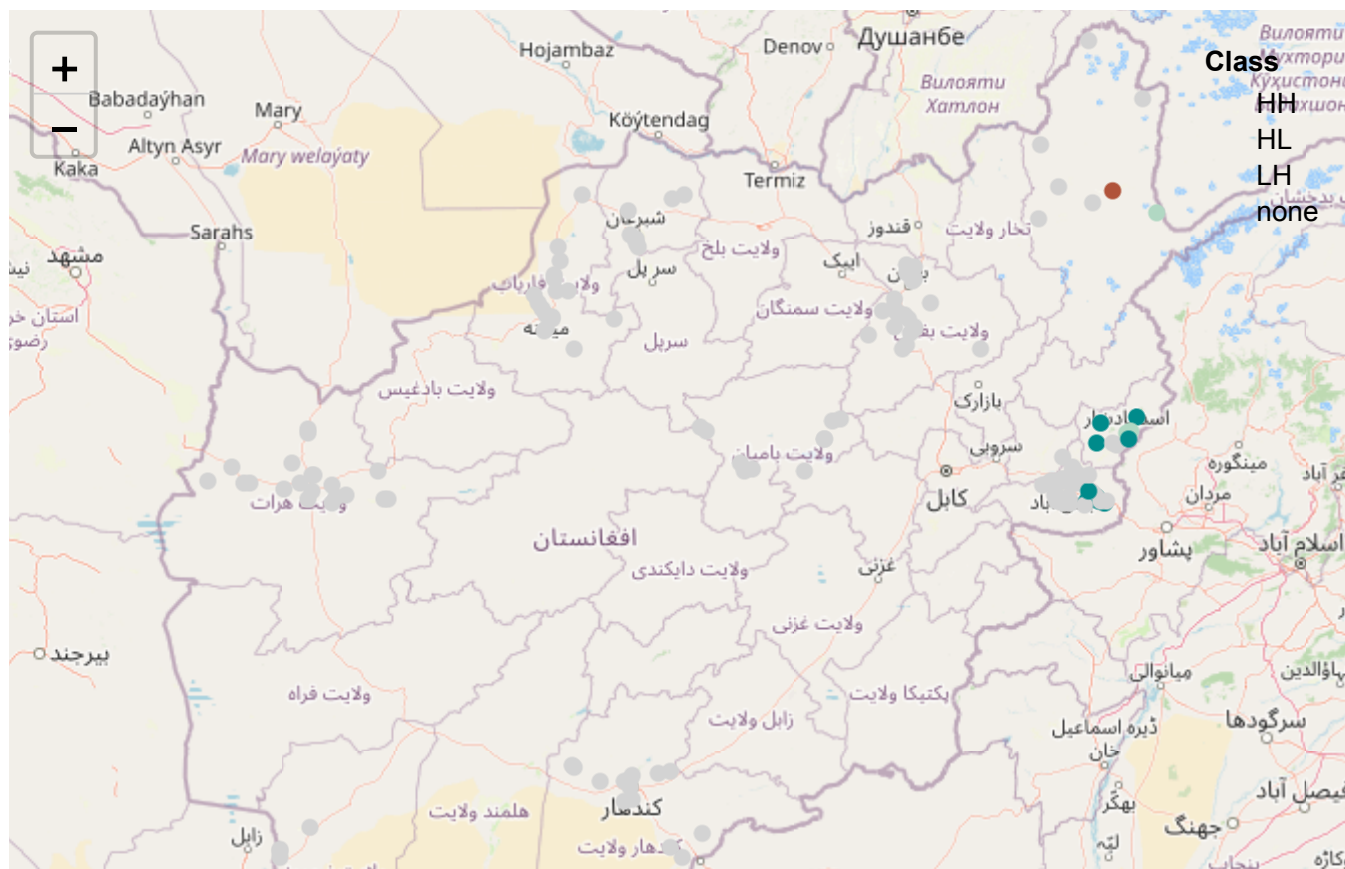
Cluster of Malaria Prevalence in Afghanistan

The local Moran's I statistic resulted in a statistically significant p-value, indicating that there is a local clustering.

```
AFG_malaria_localM<-as.data.frame(cbind(xy,names))
colnames(AFG_malaria_localM)<-c("longitude", "latitude", "names")
AFG_malaria_localM[c("longitude", "latitude")] <- lapply( AFG_malaria_localM[c("longitud
e", "latitude")], function(x) as.numeric(as.character(x)) )
factpal <- colorFactor(c( "cyan4","coral4","coral","cyan","lightgrey"), names)
leaflet(AFG_malaria_localM) %>% addTiles() %>% addCircleMarkers(~longitude, ~latitude, f
illOpacity=1,
                                                    color= ~factpal(names), r
adius=4, stroke=TRUE, weight=1) %>%
  addLegend(pal = factpal, values = ~names, title="Class")
```

Leaflet (https://leafletjs.com) | © OpenStreetMap (http://openstreetmap.org) contributors, CC-BY-SA (http://creativecommons.org/licenses/by-sa/2.0/)
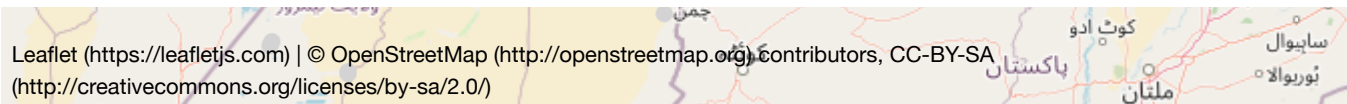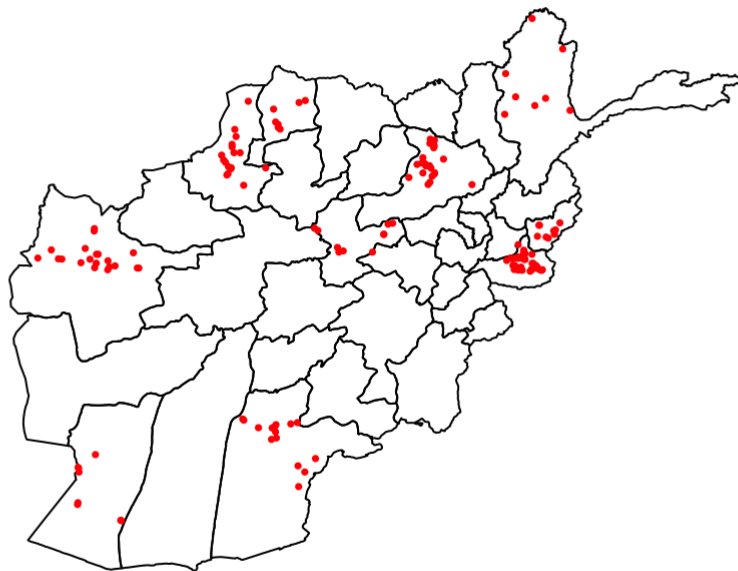
Non-spatial and Spatial Analysis

Identifies and incorporates two relevant spatial covariates, existing (e.g. worldclim) or generated (e.g. land cover, distance to something)

```
raster::plot(AFG_Adm_1)
points(AFG_malaria_data$longitude, AFG_malaria_data$latitude,
       pch = 16, ylab = "Latitude", xlab="Longitude", col="red", cex=.5)
```



```
bioclim_layers <- raster::getData('worldclim', var='bio', res=0.5, lon= 67.7100, lat=33.9391)
```

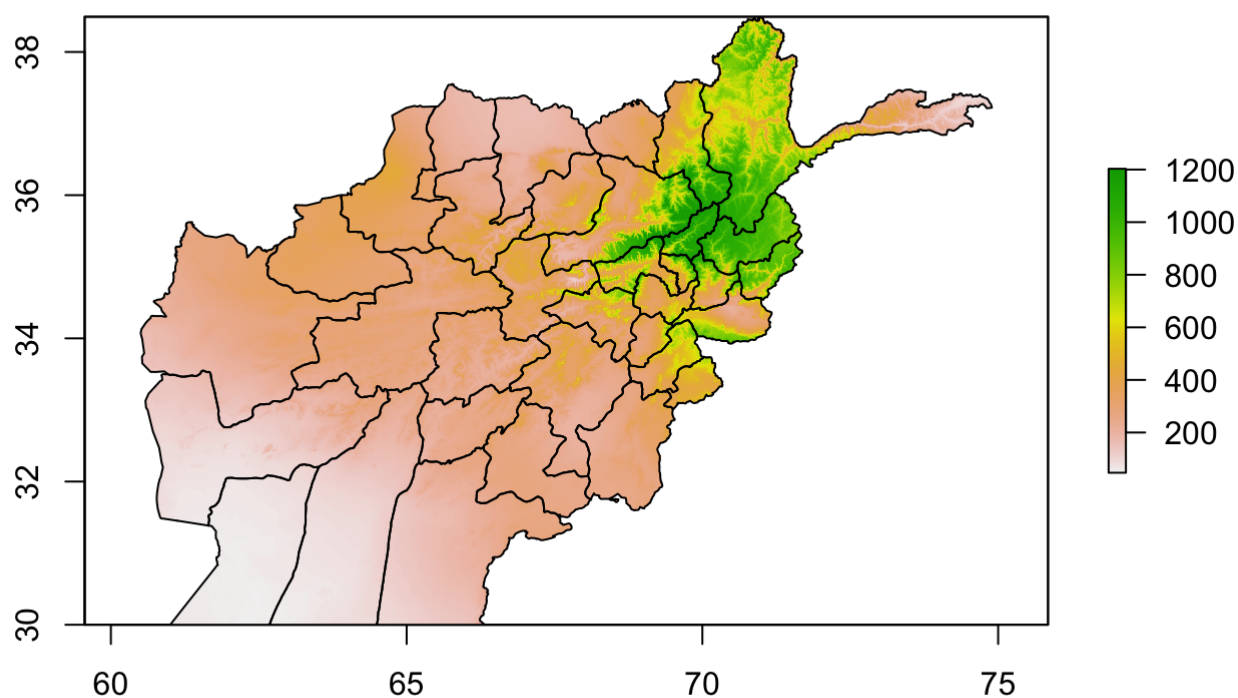***Mean annual temperature & Mean annual precipitation in Afghanistan***

The regression considered the relationship between temperature and precipitation and malaria prevalence. We can see that the eastern side of Afghanistan has high temperature and high precipitation.

```
bioclim_layers_AFG  <- crop(bioclim_layers, AFG_Adm_1)
bioclim_layers_AFG <- raster::mask(x = bioclim_layers_AFG, mask = AFG_Adm_1)
plot(bioclim_layers_AFG[[1]])
```
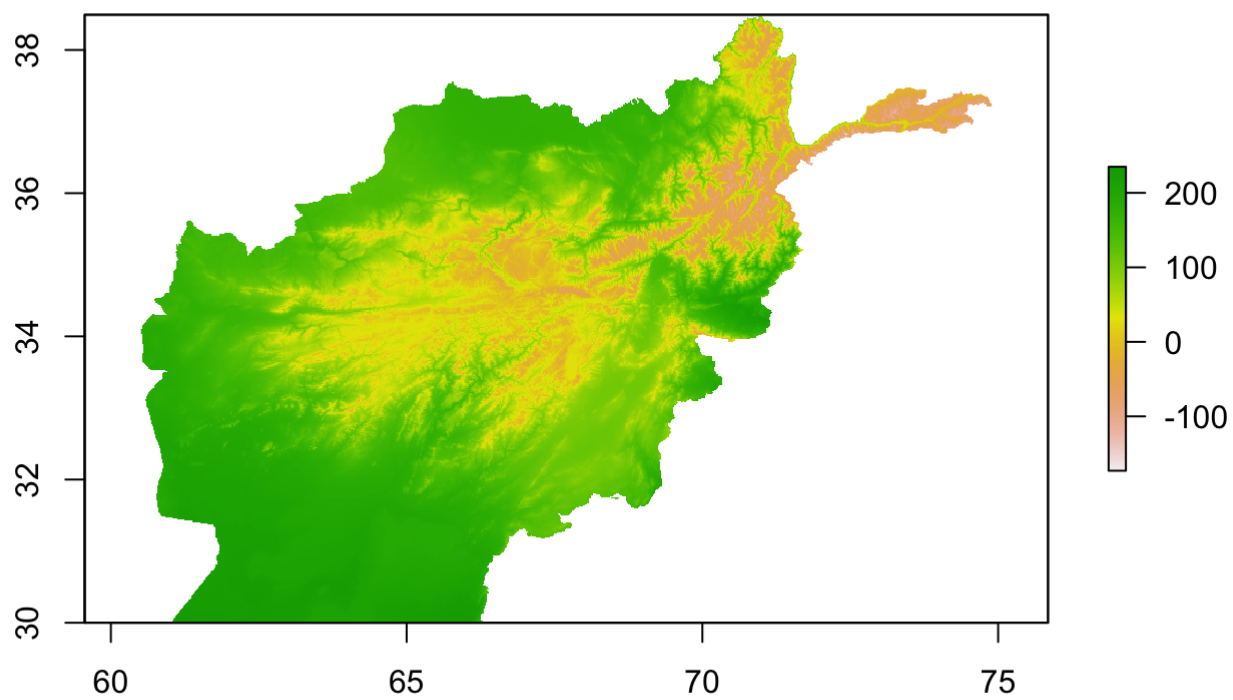
```
plot(bioclim_layers_AFG[[12]])
lines(AFG_Adm_1)
```

```
AFG_malaria_data$bioclim1 <- extract(bioclim_layers_AFG[[1]], AFG_malaria_data[,c("longi
tude", "latitude")])
AFG_malaria_data$bioclim12 <- extract(bioclim_layers_AFG[[12]], AFG_malaria_data[,c("lon
gitude", "latitude")])
```

Results

The non-spatial regression considered the relationship between temperature and precipitation and malaria prevalence, which showed a significant correlation between temperature and precipitation and malaria prevalence.

```
prev_afg_non_spatial <- spaMM::fitme(cbind(pf.pos, examined - pf.pos) ~ bioclim1 + biocl
im12, data=AFG_malaria_data, family=binomial())
```

```
## Registered S3 methods overwritten by 'registry':
##    method                  from
##    print.registry_field  proxy
##    print.registry_entry  proxy
```

```
summary(prev_afg_non_spatial)
```

```
## formula: cbind(pf.pos, examined - pf.pos) ~ bioclim1 + bioclim12
## Estimation of fixed effects by ML.
## family: binomial( link = logit )
##   ------------ Fixed effects (beta) ------------
##               Estimate Cond. SE t-value
## (Intercept) -17.867843 3.416960  -5.229
## bioclim1      0.048127 0.014811   3.249
## bioclim12     0.007918 0.001817   4.357
##   ------------- Likelihood values  -------------
##                          logLik
## p(h)   (Likelihood): -35.35321
```

```
# Compute correlogram of the residuals
nbc <- 10
cor_r <- pgirmess::correlog(coords = AFG_malaria_data[,c("longitude", "latitude")],
                        z = residuals(prev_afg_non_spatial),
                        method="Moran", nbclass=nbc)
# Take a look
cor_r
```

```
## Moran I statistic
##         dist.class          coef   p.value    n
##  [1,]   0.6077247 -0.0080244846 0.5257502 4052
##  [2,]   1.8100182 -0.0389367459 0.7658498 1706
##  [3,]   3.0123112 -0.0140477355 0.6434704 4692
##  [4,]   4.2146042  0.0019701293 0.3686273 2988
##  [5,]   5.4168972 -0.0011398542 0.3907536 4618
##  [6,]   6.6191902  0.0114732636 0.2814695 2410
##  [7,]   7.8214832 -0.0080503053 0.5066862 1596
##  [8,]   9.0237762 -0.0008544983 0.4067767 1968
##  [9,] 10.2260692 -0.0059158896 0.4307750  396
## [10,] 11.4283622 -0.1426155094 0.8740172   66
```

Results

To decide if spatial effect should be included in the model, we should take a look at spatial autocorrelation in the residuals. A residual spatial autocorrelation is observed in the following graph. Therefore, we should add spatial effect into the model.

```
# Plot correlogram
correlograms <- as.data.frame(cor_r)
correlograms$variable <- "residuals_glm"

ggplot(subset(correlograms, variable=="residuals_glm"), aes(dist.class, coef)) +
  geom_hline(yintercept = 0, col="grey") +
  geom_line(col="steelblue") +
  geom_point(col="steelblue") +
  xlab("distance") +
  ylab("Moran's coefficient")+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black"))
```
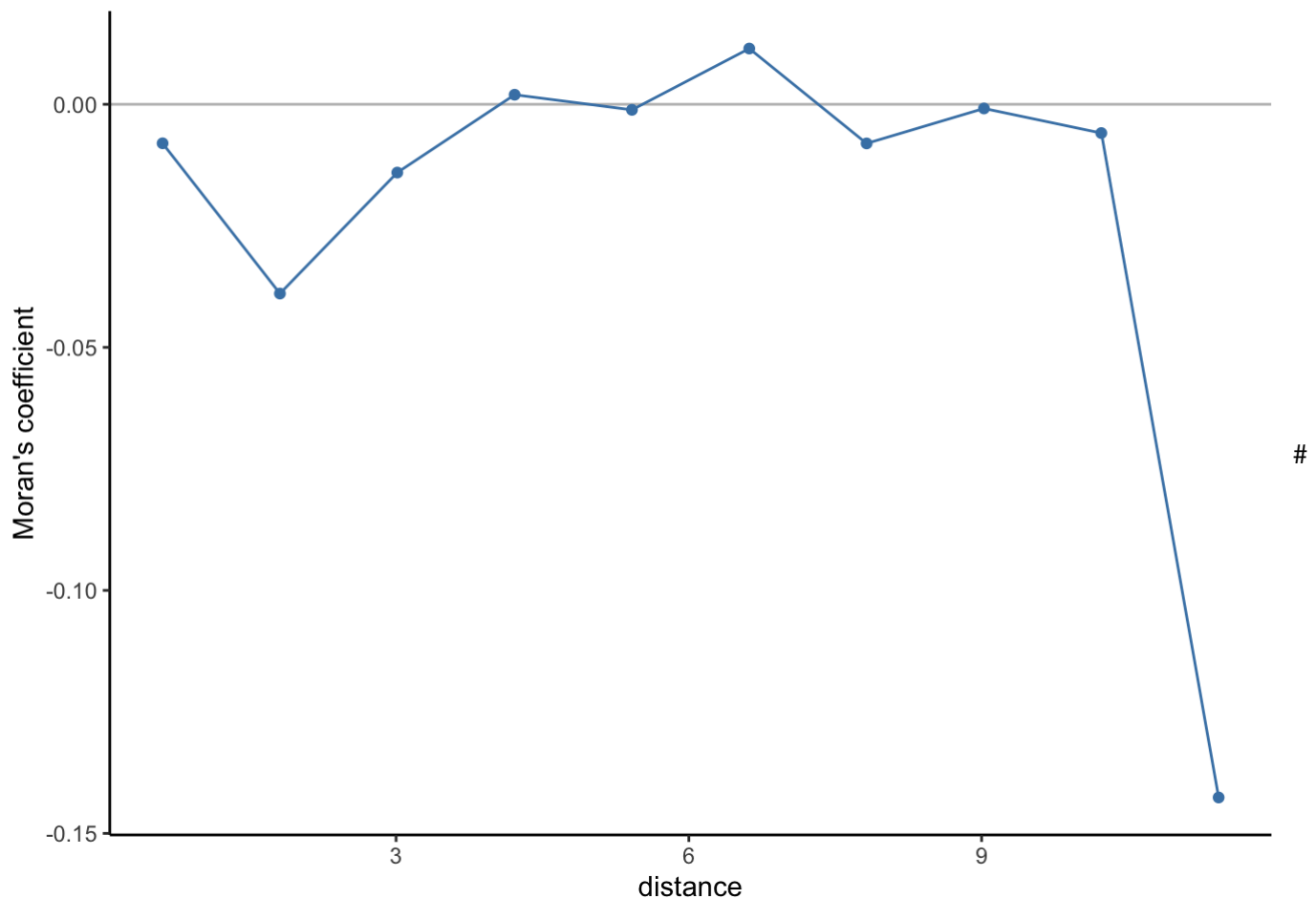

Predictions

```
prev_afg <- spaMM::fitme(cbind(pf.pos, examined - pf.pos) ~ bioclim1 + bioclim12 + Mater
n(1|latitude+longitude), data=AFG_malaria_data, family=binomial())
prev_afg
```

```
## formula: cbind(pf.pos, examined - pf.pos) ~ bioclim1 + bioclim12 + Matern(1 |
##     latitude + longitude)
## Estimation of corrPars and lambda by Laplace ML approximation (p_v).
## Estimation of fixed effects by Laplace ML approximation (p_v).
## Estimation of lambda by 'outer' ML, maximizing p_v.
## family: binomial( link = logit )
##   ------------ Fixed effects (beta) ------------
##               Estimate Cond. SE t-value
## (Intercept) -19.644945 5.510759  -3.565
## bioclim1      0.049010 0.022967   2.134
## bioclim12     0.009568 0.003423   2.795
##   -------------- Random effects --------------
## Family: gaussian( link = identity )
##                   --- Correlation parameters:
##      1.nu       1.rho
## 0.02744058 6.78699619
##           --- Variance parameters ('lambda'):
## lambda = var(u) for u ~ Gaussian;
##    latitude .  :  2.634
## # of obs: 157; # of groups: latitude ., 157
##   ------------- Likelihood values  -------------
##                          logLik
## p_v(h) (marginal L): -32.96642
```

Generate a quick scatter plot of observed vs fitted values

It seems that the predictions are accurate as the differences between observed and fitted values are small.

```
plot(AFG_malaria_data$prevalence, predict(prev_afg))
abline(0,1)
```
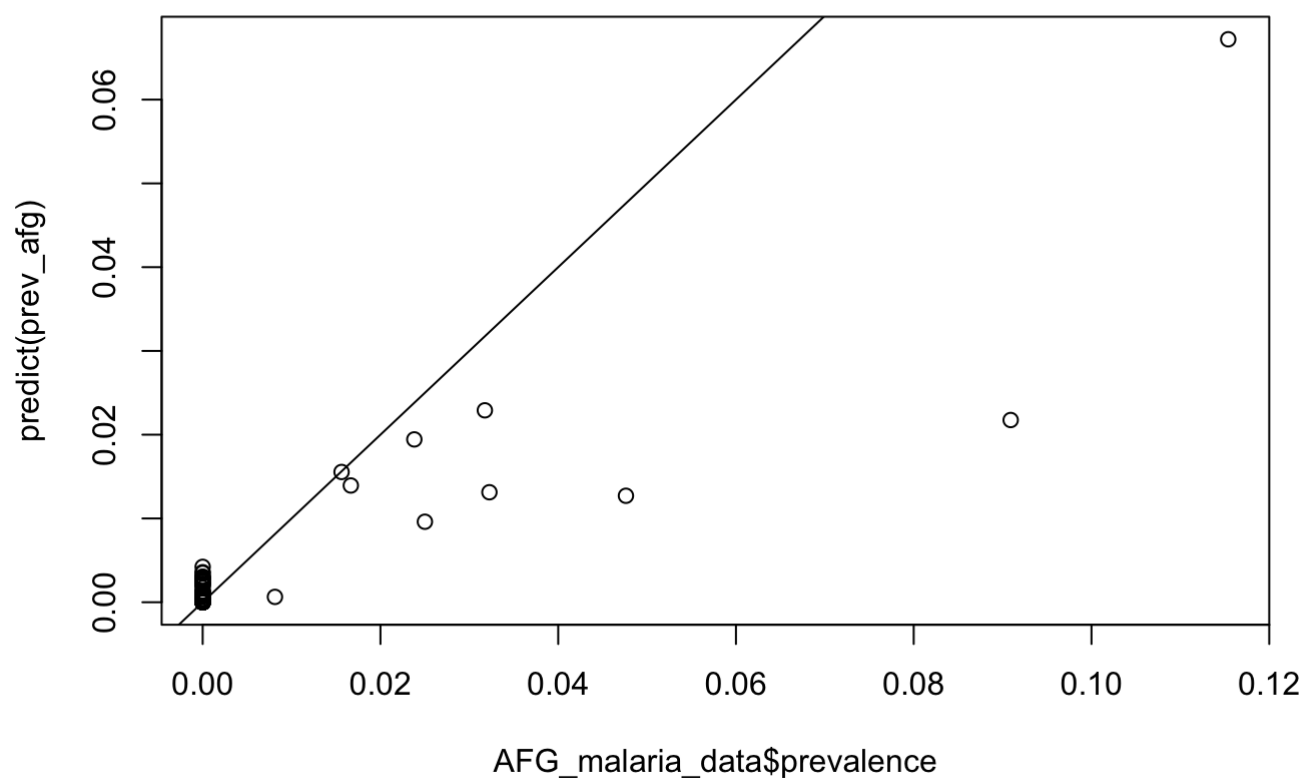
Five-fold validation using the mean squared error as performance metric

```r
cv_afg <- function(data, spamm_formula, ix_test_list) {
  mse <- c()
  for (i in 1:length(ix_test_list)) {
    test_set <- data[ix_test_list[[i]], ]
    train_set <- data[(1:300)[-c(ix_test_list[[i]])], ]
    model <- spaMM::fitme(spamm_formula, data=train_set, family=binomial())
    model_prediction <- predict(model, newdata=test_set, type="response")[,1]
    mse[i] <- mean((model_prediction * test_set$examined  - test_set$pf.pos)^2)
  }
  return(mean(mse))
}
```

```r
ix = caret::createFolds(AFG_malaria_data$pf.pos, k = 5)
```

```r
layer_names <- c("bioclim1", "bioclim12")
formula_kern <- "cbind(pf.pos, examined - pf.pos) ~ Matern(1|latitude+longitude)"
formula_model <- paste(c(formula_kern, layer_names), collapse = " + ")
scores <- c(cv_afg(AFG_malaria_data, as.formula(formula_model), ix))
# Simpler model
num_covariates <- length(layer_names)
max_covariates <- num_covariates - 1
indices <- 1:num_covariates

board <- data.frame(MSE = tail(scores, n=1), Covariates = paste(indices, collapse =
","))
while (num_covariates > max_covariates) {
  scores_iter <- c()
  ix_subsets <- gtools::combinations(n=num_covariates, r=num_covariates-1, v=indices)
  for (i in 1:nrow(ix_subsets)) {
    cov_subset <- layer_names[ix_subsets[i,]]
    formula_model <- paste(c(formula_kern, cov_subset), collapse = " + ")
    scores_iter <- c(scores_iter, cv_afg(AFG_malaria_data, as.formula(formula_model), i
x))
  }
  best <- which.min(scores_iter)
  indices <- ix_subsets[best, ]
  scores <- c(scores, scores_iter[best])
  num_covariates <- length(indices)
  if (diff(tail(scores, n=2)) < 0 & max_covariates >= 2) {
    max_covariates <- max_covariates - 1
  }
  board <- rbind(board,
                 data.frame(MSE = tail(scores, n=1),
                            Covariates = paste(indices, collapse = ",")))
}
```

Results

The best model with lowest (MSE) includes both bioclim1(temperature) and bioclime12 (precipitation).

```r
print(board)
```

```
##         MSE Covariates
## 1 0.1642782        1,2
## 2 0.1534849          2
```

Predictions on Malaria Prevalence in Afghanistan

```
# Create an empty raster with the same extent and resolution as the bioclimatic layers
latitude_raster <- longitude_raster <-raster(nrows = nrow(bioclim_layers_AFG[[1]]),
                                    ncols = ncol(bioclim_layers_AFG[[1]]),
                                          ext = extent(bioclim_layers_AFG
[[1]]))

# Change the values to be latitude and longitude respectively
longitude_raster[] <- coordinates(longitude_raster)[,1]
latitude_raster[] <- coordinates(latitude_raster)[,2]

# Now create a final prediction stack of the 4 variables we need
pred_stack <- stack(bioclim_layers_AFG[[c(1,12)]],
                    longitude_raster,
                    latitude_raster)

# Rename to ensure the names of the raster layers in the stack match those used in the m
odel
names(pred_stack) <- c("bioclim1", "bioclim12", "longitude", "latitude")
plot(pred_stack)
```
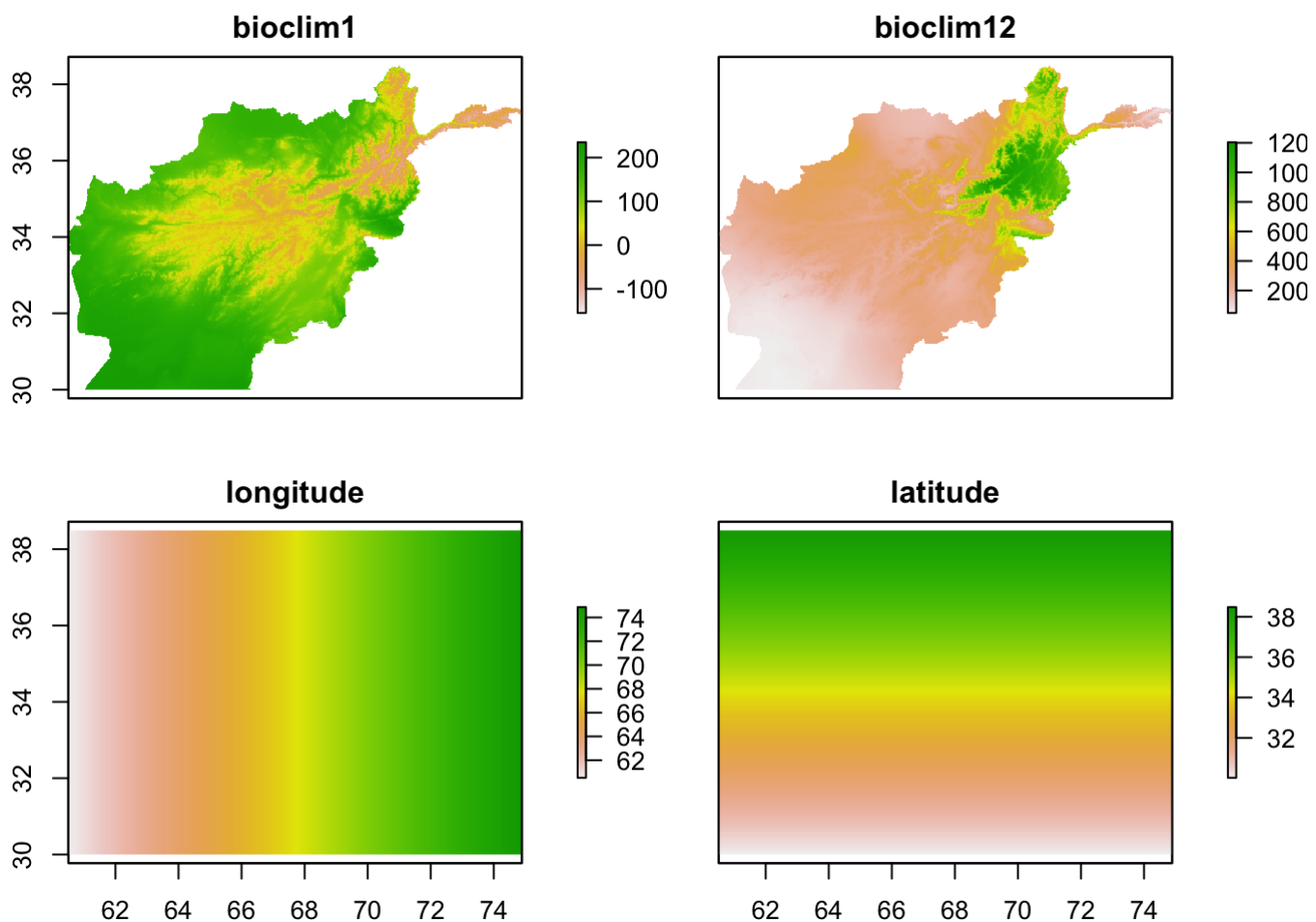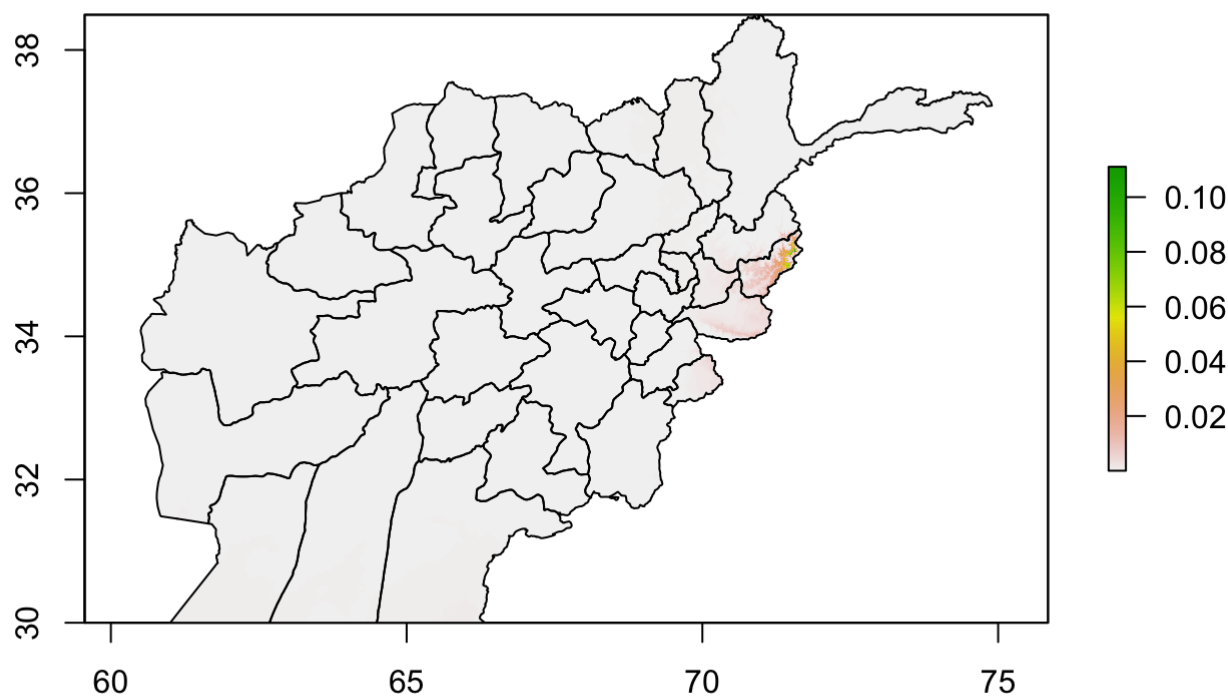


```
predicted_prevalence_raster <- predict(pred_stack, prev_afg)
plot(predicted_prevalence_raster)
lines(AFG_Adm_1)
```

Discussion

Both global and local clustering analyses support that malaria is an environmentally mediated disease. Geospatial characteristics like temperature and precipitation could impact the malaria risk. Increased temperature and precipitation are associated with greater odds of disease. This is intuitive given that malaria is commonly transmitted via the bites of mosquitos, whose life cycle and habitats are influenced by temperature and precipitation. The study could help international development institutions and the government of Afghanistan better understand the spatial distribution and clustering of malaria endemic and how geospatial characteristics are associated with the disease. International development institutions and the government of Afghanistan should pay close attention to those areas at high risk of malaria and implement public health interventions coping with climate change.