

Red Foxes in BC: an Analysis of the Spatial Point Process

Nowshaba Durrani, Ricky Heinrich, Viji Rajagopalan

2023-04-09

Introduction

For our Data 589 project, we have selected Red Fox (Scientific Name - *Vulpes Vulpes*) to do the analysis. In the GBIF database they have approximately, 610,958+ georeferences records for this species around the world, however for this project we have selected to do the analysis of the occurrence of Red Fox in BC only. So with the above function we have fetched the information for British Columbia only in 127 columns and 242 number of entries.

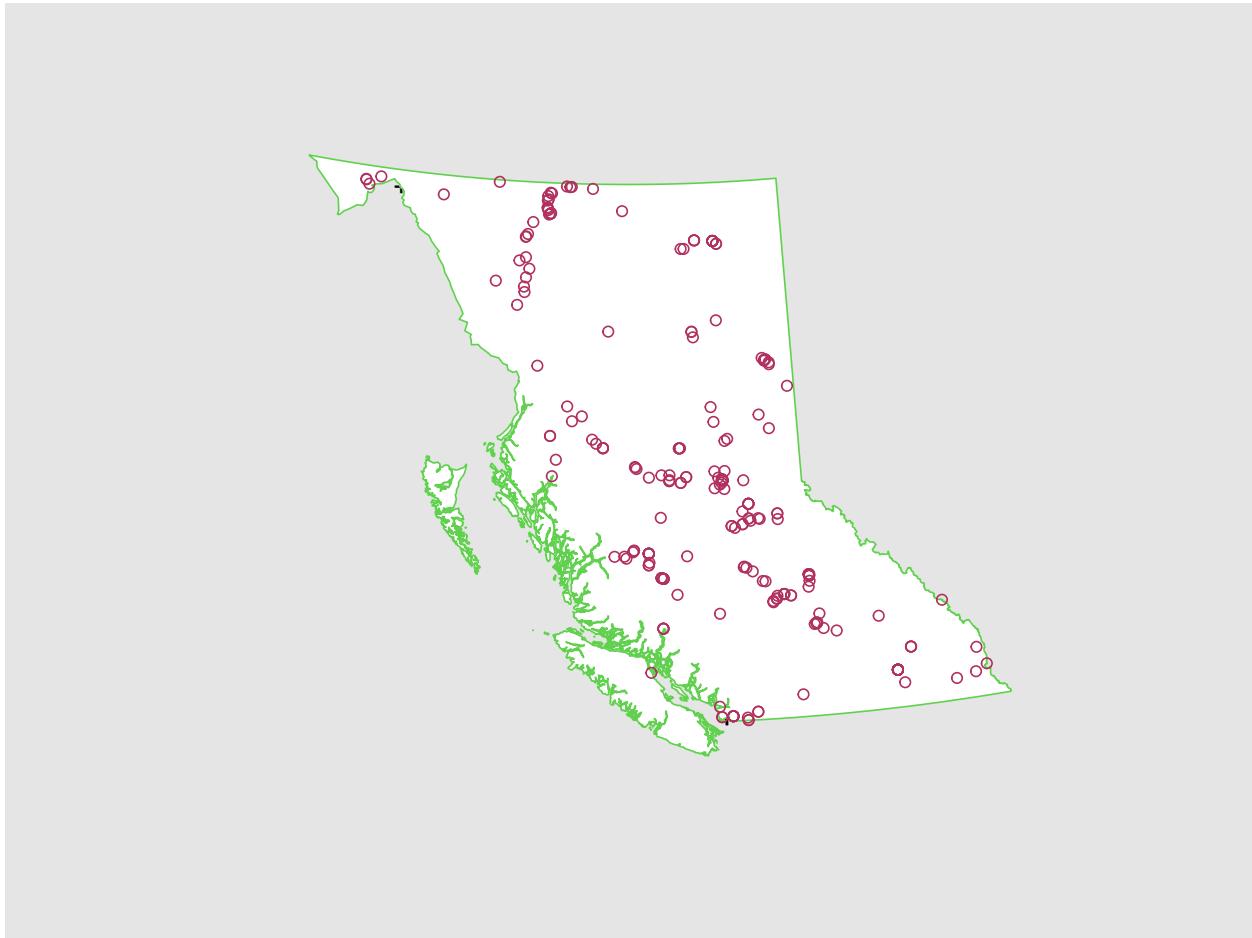


Figure 1: Occurrence of Red Foxes in BC

Here we have plotted all the occurrences of Red Fox in the BC region and we can see that the species are scattered in the region specially in the upper and middle part of the province. Now we will be exploring what

is contributing to the occurrences of the species in the specific places based on various factors like elevation, close to water bodies, forests, human habitats, etc.

Methods

Briefly describe the data and what variables are included. Provide a detailed description of the analytical workflow that was applied to the data, citing any relevant literature and statistical packages employed. There should be enough information that anyone can reproduce the workflow if they had access to the data. Length: As long as necessary.

The data comes from the Global Biodiversity Information Facility (GBIF) databases. We used the package `rgbif` to access the ‘*Vulpes Vulpes*’ data from R directly, sorting by instances occurring in BC. We’ve extracted the longitude and latitude data from this, and converted it appropriately using the `sp` package.

Our second source of data contained the BC Window object, as well as possible covariate data: elevation, forest cover, Human Footprint Index (HFI) and distance to water.

We used the package `spatstat` to build a `ppp` object with the converted coordinates of the Red Fox locations from the GBIF data and the window from our second data source.

To conduct first moment analysis, we used functions from the aforementioned `spatstat` package. We did a quadrat test as well as hotspot analysis to gain insight into the homogeneity assumption of the point process.

For second moment analysis, we looked into Ripley’s K-function and pair correlation function using functions from `spatstat`. This provides us with insight into possible clustering tendencies of the point process.

Next we looked into the relationship of the intensity with each covariate.

First Moment Analysis

Quadratcount

We start with investigating whether the occurrence of red foxes in BC seems homogeneous, as it will inform our steps to define the intensity. We have conducted a quadrat test of homogeneity with both 5 x 5 and 10 x 10 quadrats. These quadrats are shown in Figure 2, where we can visually tell that the intensity in each quadrats are not the same. The quadrat test for both the 5 x 5 and the 10 x 10 quadrats provide a p-value of 2.2e-16, confirming that the varied intensities are not due to chance alone, but rather due to an inhomogeneous point process.

Hot spot analysis

As the next step, we investigate analyze for any hot spots in the occurrences of red foxes. In Figure 3, we can see that hotspots appear scattered around the province.

2nd Moment Analysis

Ripley’s K function

Ripley’s K-function provides information on whether there are significant deviations from independence between points. Taking into account that the intensity of red fox occurrences appear inhomogeneous, we can see in Figure 4 that there is some evidence of clustering up to a certain distance, as the black line, indicating the observed data, is separate from the 95% confidence bands of the values expected with no clustering. This suggests that the relationship between points may be due to effects between points rather than relationship with covariates. [CHECK THIS CLAIM ?]

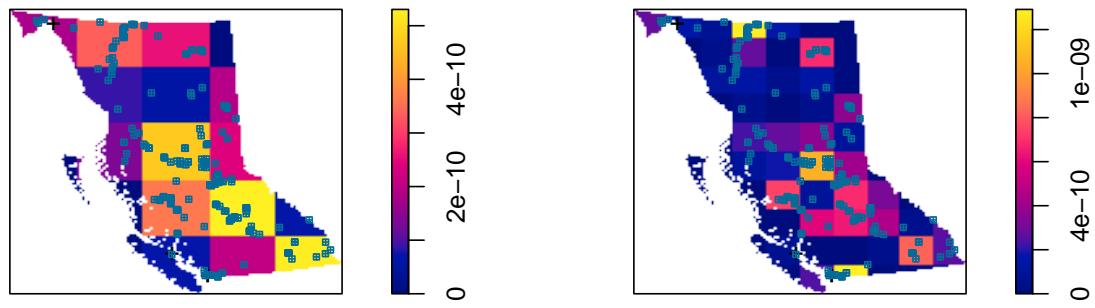


Figure 2: Intensity of Quadrat counts of Red Fox occurrences, left 5x5, right 10x10

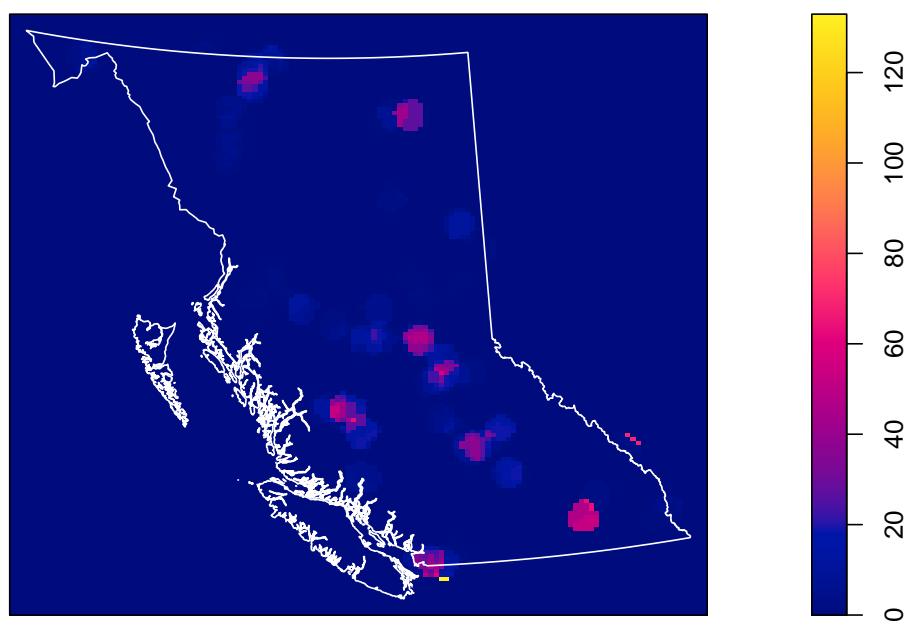


Figure 3: Hotspot of Red Foxes

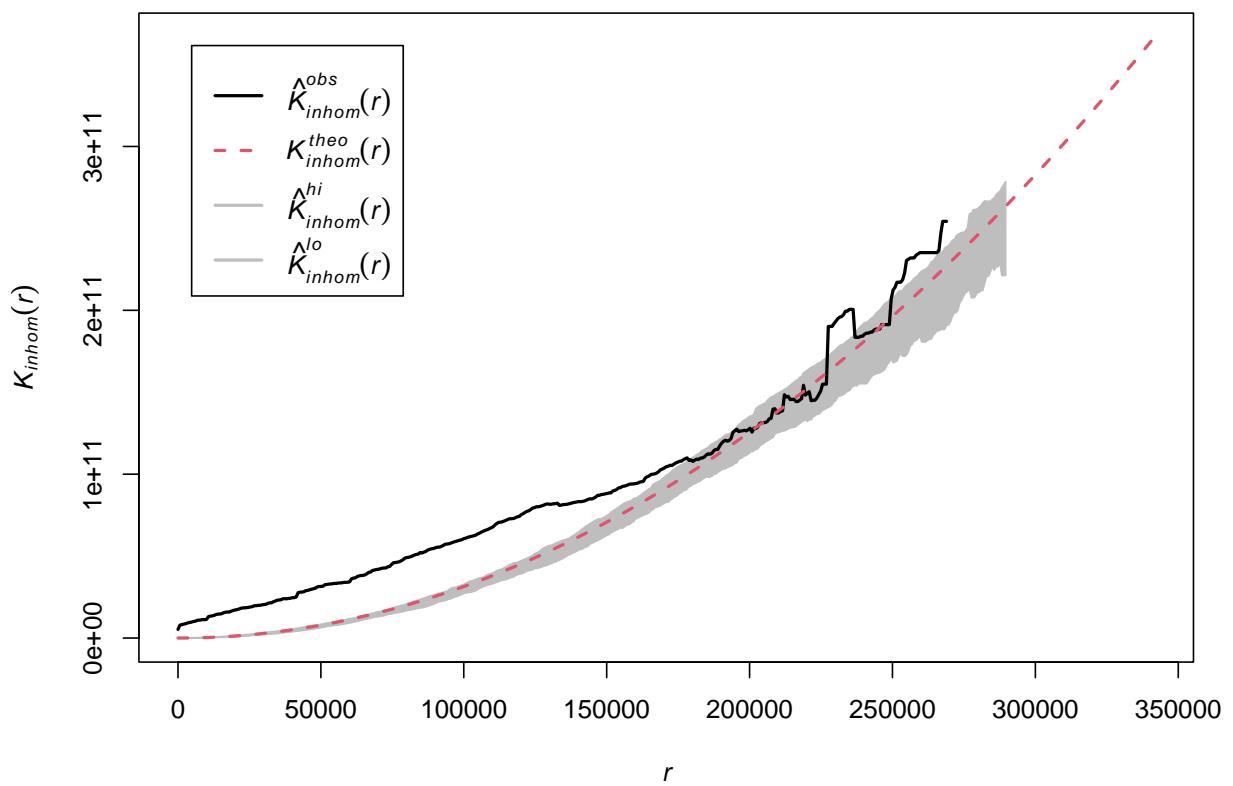


Figure 4: Ripley's K function with border correction assuming inhomogeneity

Pair correlation function

To get a sense of the distances for which clustering occurs, we used the pair correlation function. Figure 5 shows evidence for clustering at distances smaller than around 23 000m, or 23km but after that the observed values are not significantly different than those expected from a random spatial process.

```
## Generating 19 simulations by evaluating expression ...
## 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19.
##
## Done.
```

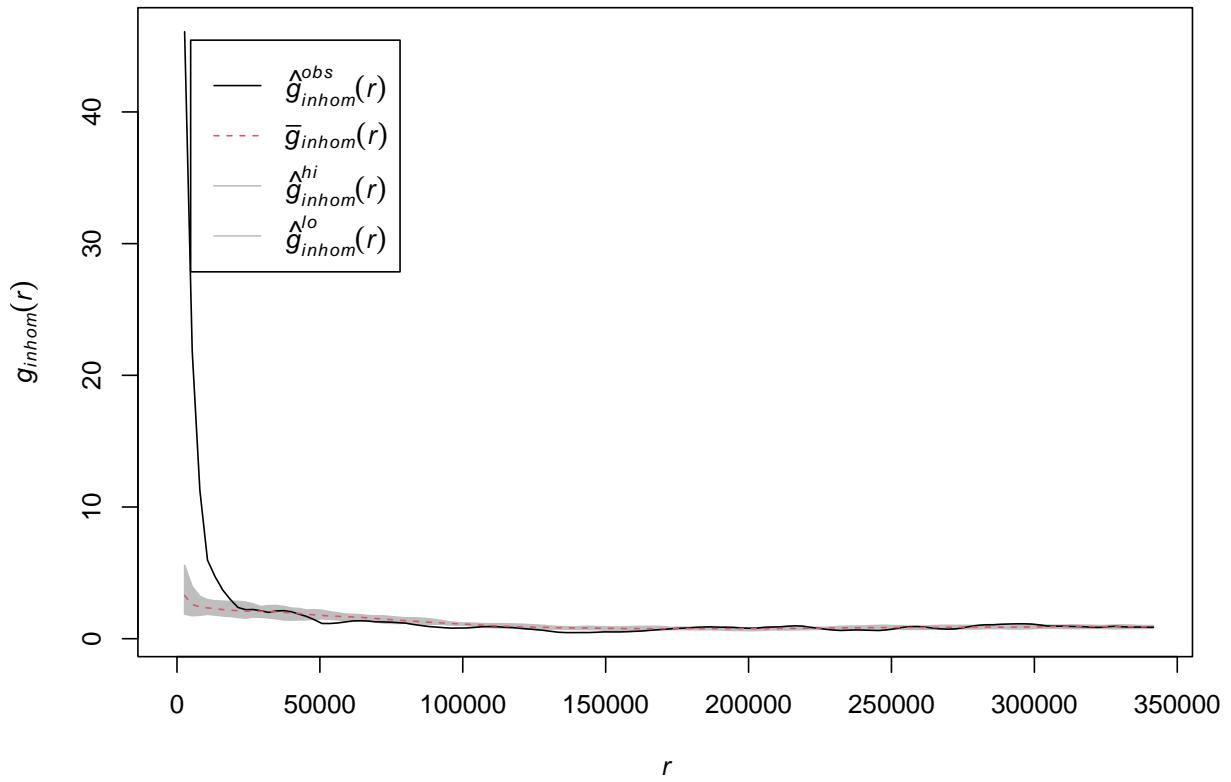


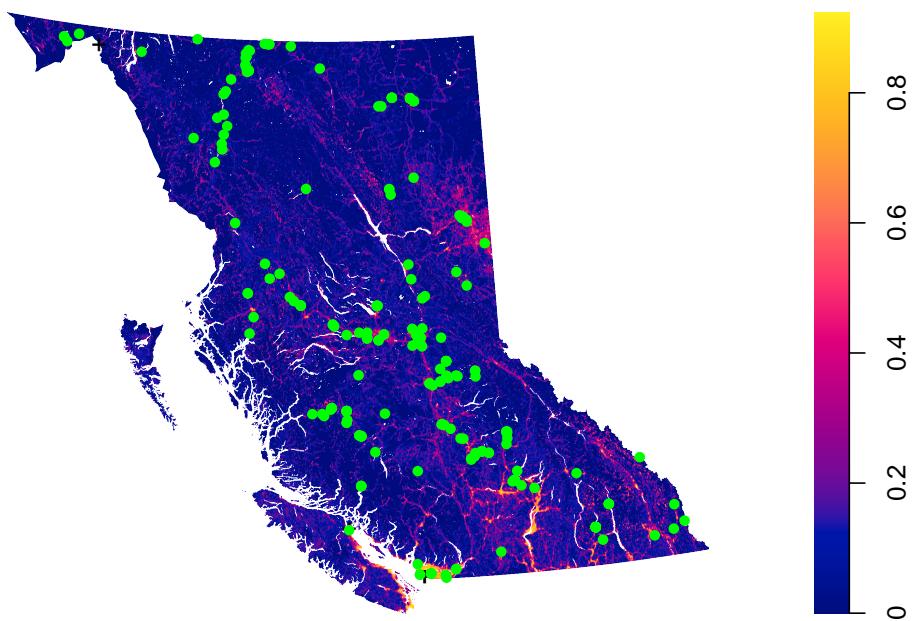
Figure 5: Pair correlation function assuming inhomogeneity

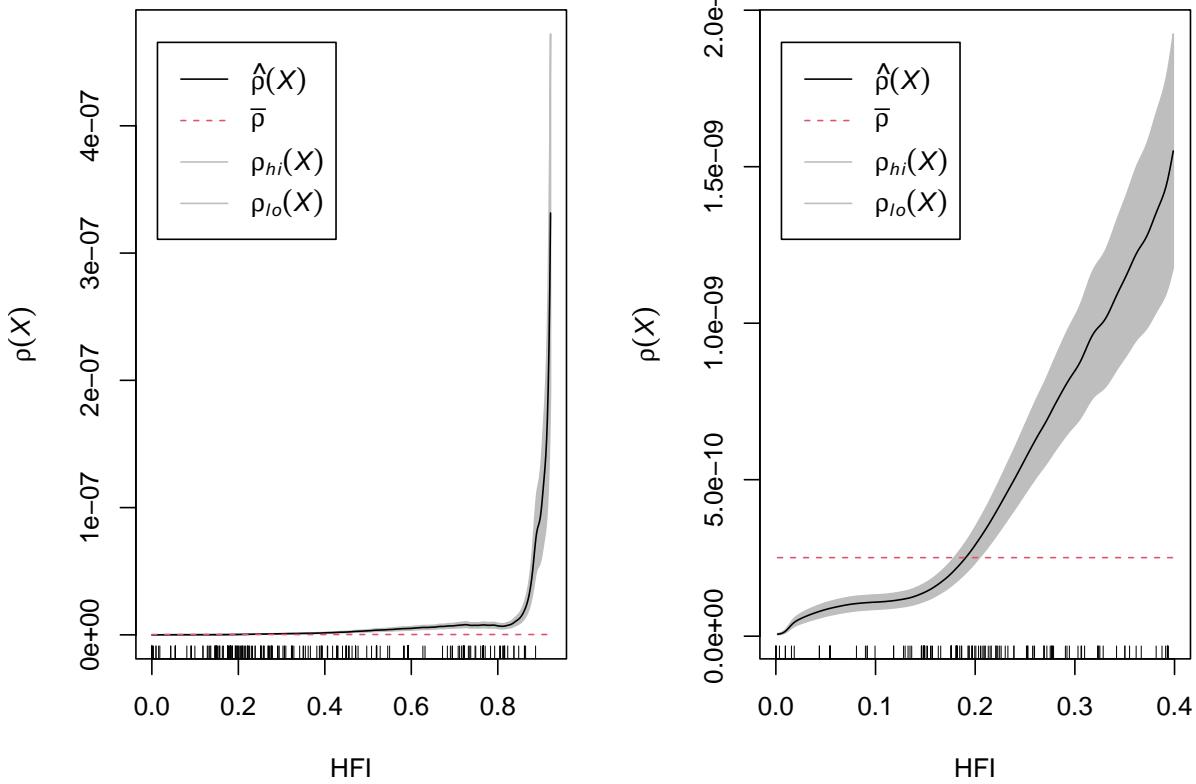
Relationship with Covariates

Our data includes 4 covariates we can explore: the elevation, the forest cover, the human footprint inventory (HFI), and the distance to water. Given our research questions, we will start with investigating the HFI and the forest cover.

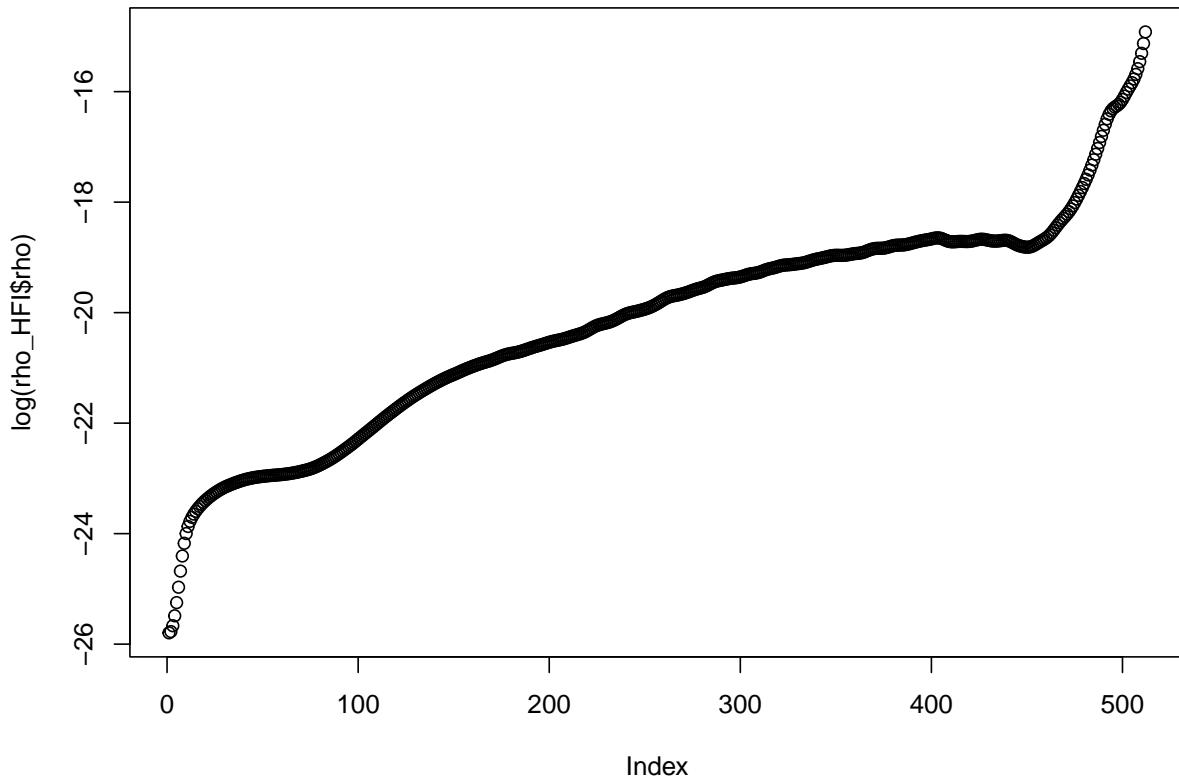
HFI

Forest





In the first figure, we could be fooled into thinking that there is no relationship up to around $HFI = 0.4$, until which it seems like an exponential relationship. However, zooming in from $HFI \in [0, 1]$ to $HFI \in [0, 0.4]$, we see that the confidence bands don't intersect at all with the red line, which is the expected value given no relationship. This relationship appears non-linear and possibly exponential, where the greatest intensity of observed red foxes occurs at high HFIs. This relationship was expected, as our dataset is not exhaustive but rather is crowdsourced, and naturally foxes are more likely to be noticed by humans in spaces with higher HFIs.



If we plot the log of the rho, we get a line that could be reasonably interpreted as linear.

simple model

```
## Nonstationary Poisson process
## Fitted to point pattern dataset 'parks_ppp'
##
## Log intensity: ~HFI
##
## Fitted trend coefficients:
## (Intercept)      HFI
## -23.31421      5.98177
##
##             Estimate      S.E.    CI95.lo    CI95.hi   Ztest      Zval
## (Intercept) -23.31421 0.1058336 -23.521645 -23.106785 *** -220.29132
## HFI         5.98177 0.2148471  5.560677  6.402862 ***  27.84199
## Problem:
## Values of the covariate 'HFI' were NA or undefined at 0.56% (12 out of 2137)
## of the quadrature points

## Nonstationary Poisson process
## Fitted to point pattern dataset 'parks_ppp'
##
```

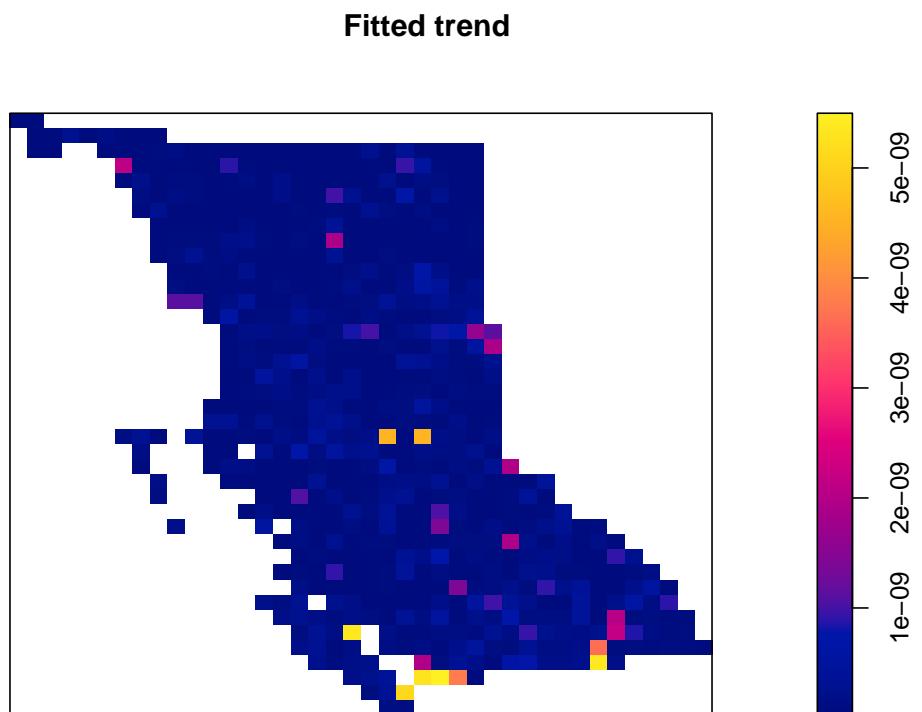
```

## Log intensity: ~HFI + exp(HFI)
##
## Fitted trend coefficients:
## (Intercept)      HFI      exp(HFI)
## -13.71641    22.17067   -10.35893
##
##             Estimate     S.E.  CI95.lo  CI95.hi Ztest      Zval
## (Intercept) -13.71641 1.404309 -16.46881 -10.964015 *** -9.767372
## HFI         22.17067 2.382441  17.50117  26.840165 ***  9.305861
## exp(HFI)    -10.35893 1.522442 -13.34286 - 7.374997 *** -6.804153
## Problem:
## Values of the covariate 'HFI' were NA or undefined at 0.56% (12 out of 2137)
## of the quadrature points

## [1] 10468.83

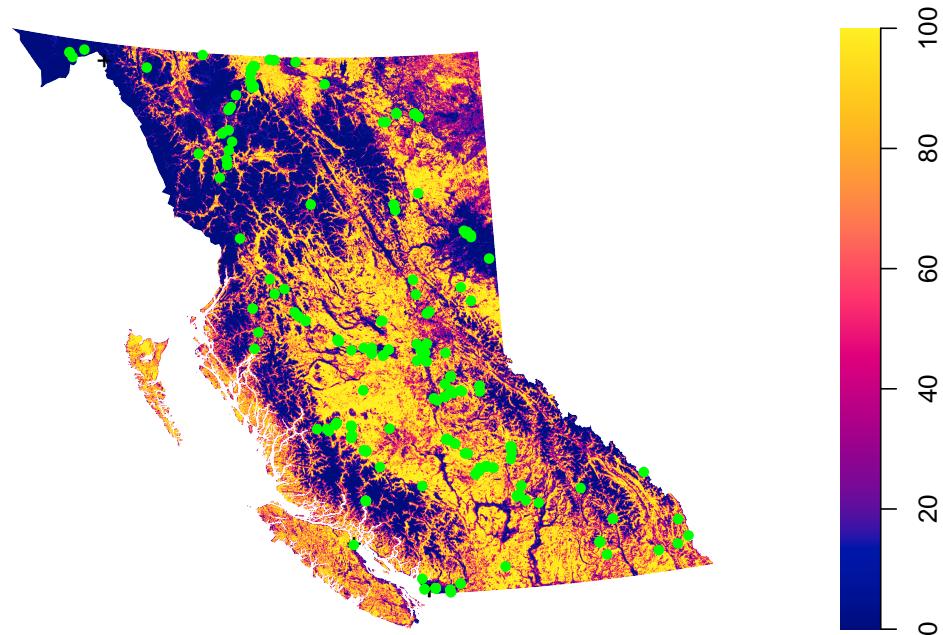
## [1] 10420.78

```

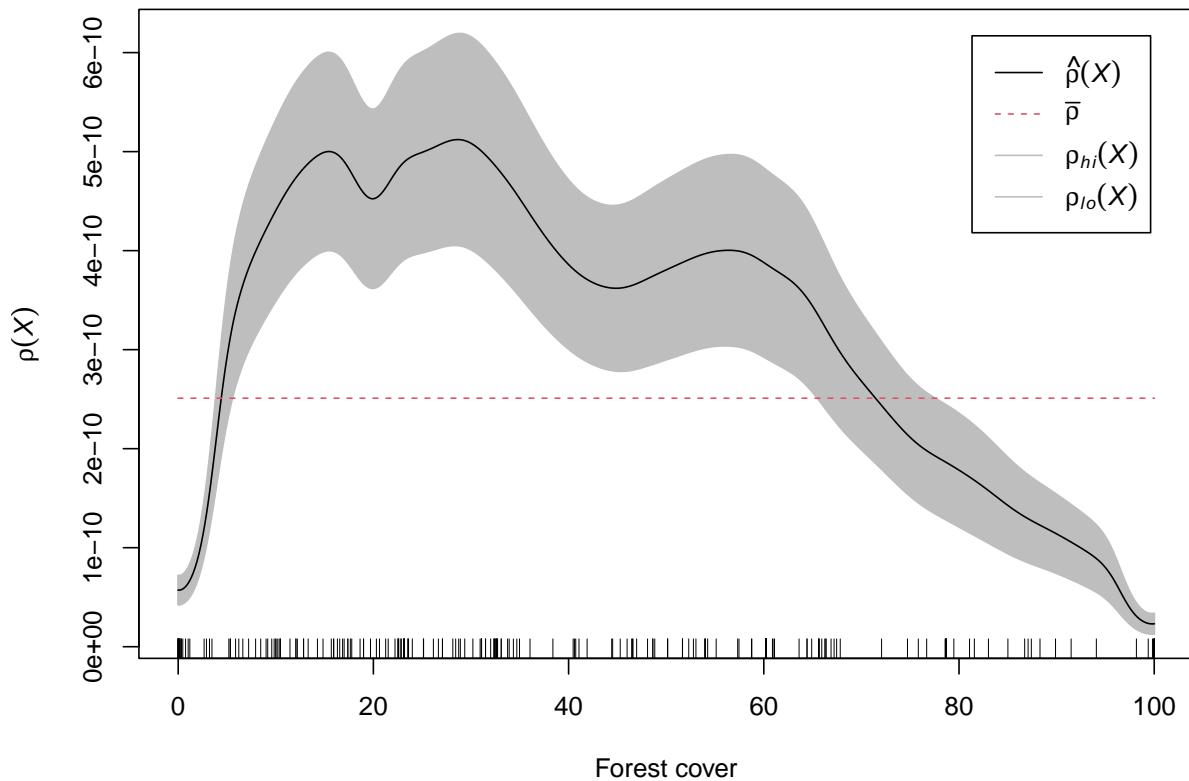


Forest Cover

Forest



1. To add : Add other covariates together - side by side

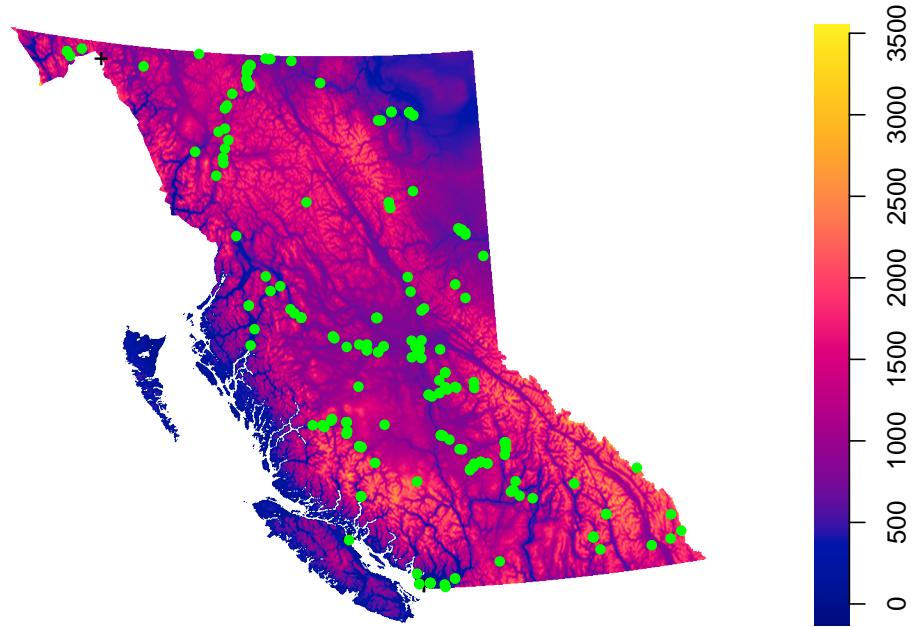


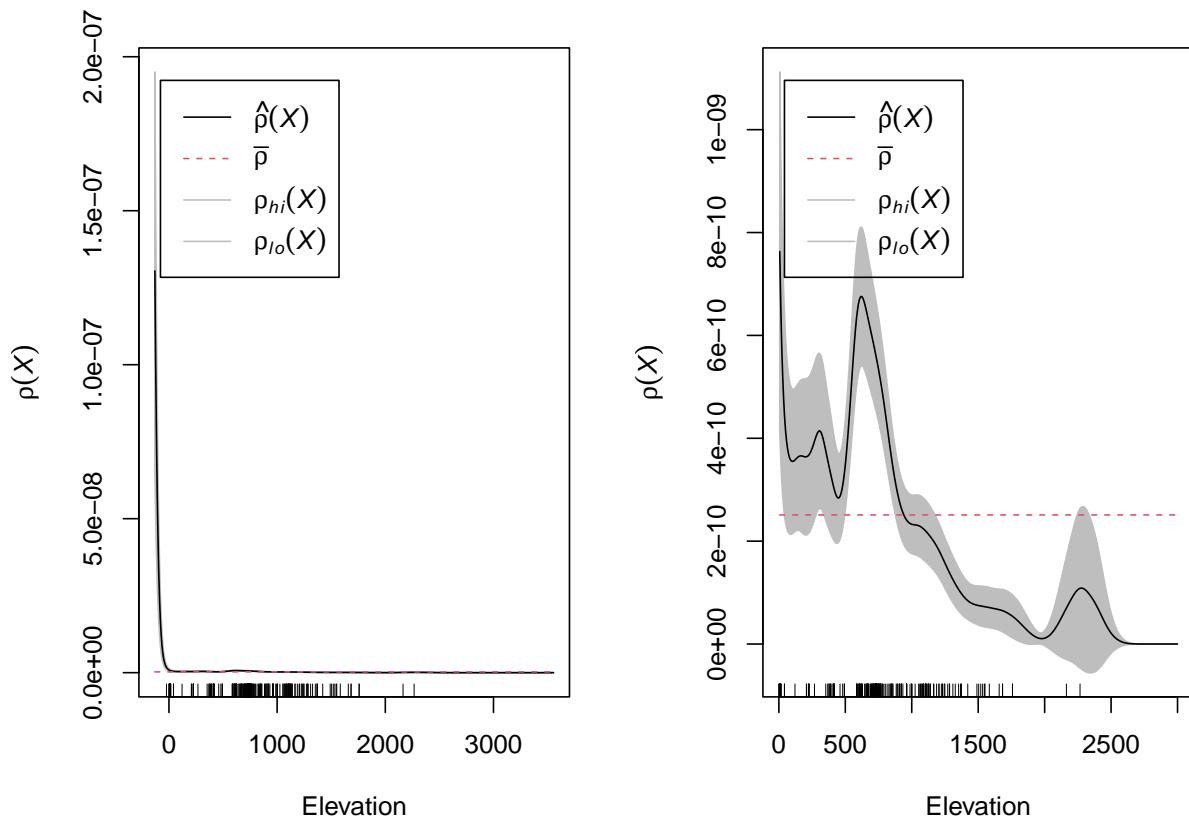
There seems to be non-linear relationship between forest cover and number of red foxs observed. The observance increase with increase in forest cover at intermediate coverage and then it decreases.

Next is to observe if there is any correlation between the covariates (collinearity in the covariates dataset). This is necessary to avoid any identifiability issues in modeling the data.

Elevation

Elevation

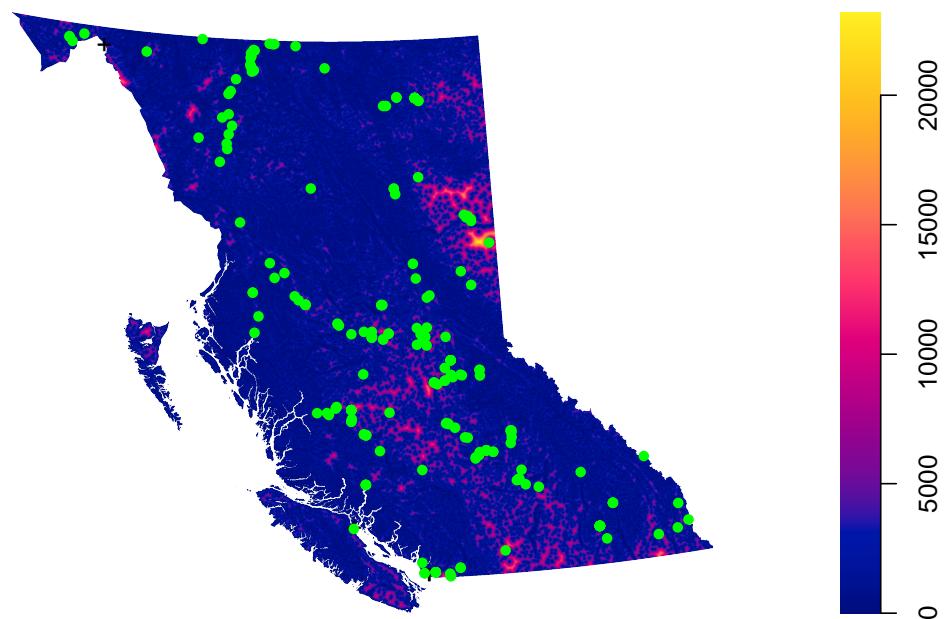


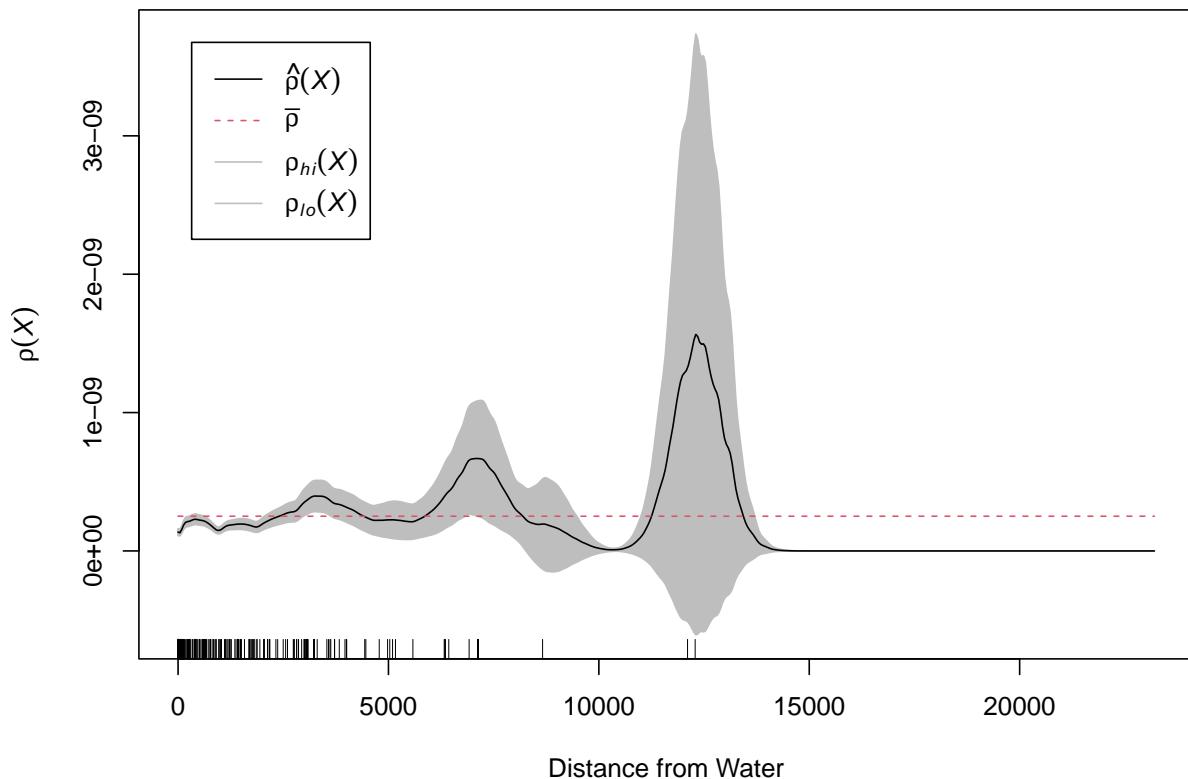


In the first figure, there seems to be no relationship but after taking a closer look we see that there is non-linear relationship. The relationship appears to be non-linear as the graph is showing different results for different elevation and we cannot see any type of specific pattern from the same.

Distance to Water

Distance from Water





In the above figure, we see that there is non linear relationship.

2. Checking collinearity

```
##           .1      .2      .3      .4
## .1  1.00000000  0.06616335 -0.26217406  0.04822162
## .2  0.06616335  1.00000000 -0.26625709  0.13249159
## .3 -0.26217406 -0.26625709  1.00000000 -0.03497584
## .4  0.04822162  0.13249159 -0.03497584  1.00000000
```

We see that no covariate is strongly collinear with another, and so we can move on without taking so into account and treating the covariates as independent.

3. Finalize and summarize about other two variables - elevation and water.

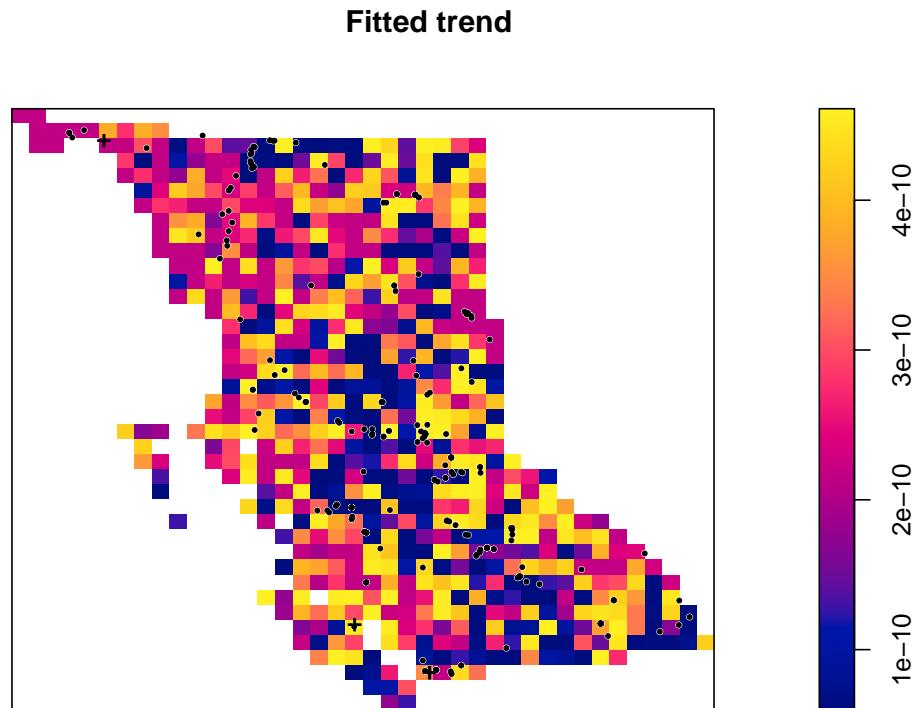
Model Fitting

Try different models - linear and quadratic, do necessary anova and other tests for model selection and validation

Model with Forest

```
## Nonstationary Poisson process
## Fitted to point pattern dataset 'parks_ppp'
##
## Log intensity: ~Forest + I(Forest^2)
##
## Fitted trend coefficients:
##   (Intercept)      Forest    I(Forest^2)
## -2.225373e+01  3.996522e-02 -5.280288e-04
##
##           Estimate       S.E.     CI95.lo     CI95.hi Ztest
## (Intercept) -2.225373e+01 1.330469e-01 -2.251450e+01 -2.199297e+01 *** 
## Forest       3.996522e-02 7.091557e-03  2.606603e-02  5.386442e-02 *** 
## I(Forest^2) -5.280288e-04 7.699618e-05 -6.789385e-04 -3.771191e-04 *** 
##
##          Zval
## (Intercept) -167.262353
## Forest       5.635606
## I(Forest^2) -6.857857
```

All variables look significant.

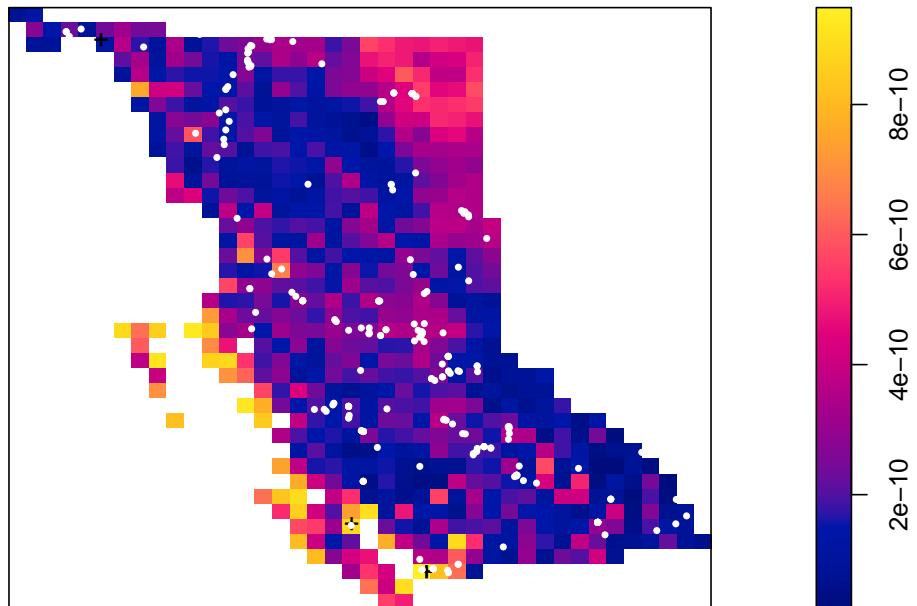


Does not look like a good fit for all points and also not easy to interpret.

Model with Elevation

```
## Nonstationary Poisson process
## Fitted to point pattern dataset 'parks_ppp'
##
## Log intensity: ~Elevation
##
## Fitted trend coefficients:
##   (Intercept)   Elevation
## -20.775522387 -0.001401046
##
##           Estimate      S.E.    CI95.lo    CI95.hi Ztest
## (Intercept) -20.775522387 0.1298745141 -21.030071757 -20.520973016 *** 
## Elevation    -0.001401046 0.0001421902  -0.001679734  -0.001122358 *** 
##
##          Zval
## (Intercept) -159.966122
## Elevation    -9.853321
```

Fitted trend



Model with Distance to Water

```
## Nonstationary Poisson process
## Fitted to point pattern dataset 'parks_ppp'
##
```

```

## Log intensity: ~Dist_Water
##
## Fitted trend coefficients:
##   (Intercept)  Dist_Water
## -2.211986e+01  1.071969e-05
##
##             Estimate      S.E.    CI95.lo    CI95.hi Ztest
## (Intercept) -2.211986e+01 8.745074e-02 -2.229126e+01 -2.194846e+01 *** 
## Dist_Water   1.071969e-05 3.419673e-05 -5.630466e-05  7.774404e-05
##
##            Zval
## (Intercept) -252.9408319
## Dist_Water   0.3134712

```



Here Elevation is significant but Dist to Water is not significant.

```
## [1] 10468.83
```

```
## [1] 10931.26
```

```
## [1] 10896.99
```

```
## [1] 11000.3
```

Analyze - linear+quad, AIC

Validate if we need GAM - 5. linear vs gam - table: evaluate, AIC, visualize

Results: Length: Describe your statistical findings. Tables and figures should be used throughout. Length: As long as necessary.

Findings from EDA

model: Tablulate observations -4. table to summarize models - linear+quad, AIC

model: GAM - Table if necessary

Discussion: Provide a brief summary of your findings. Length: ca. 1 page.

References: Include references to all necessary literature.

1. Data: GBIF.org (09 April 2023) GBIF Occurrence Download <https://doi.org/10.15468/dl.p6tsaa>
2. Research topics: <https://cwf-fcf.org/en/resources/encyclopedias/fauna/mammals/red-fox.html>