Orangutan Spatial Pattern

*Liwen Zhang, Xuechun Liu, Kailun Huang**

4/11/2018

Data Input

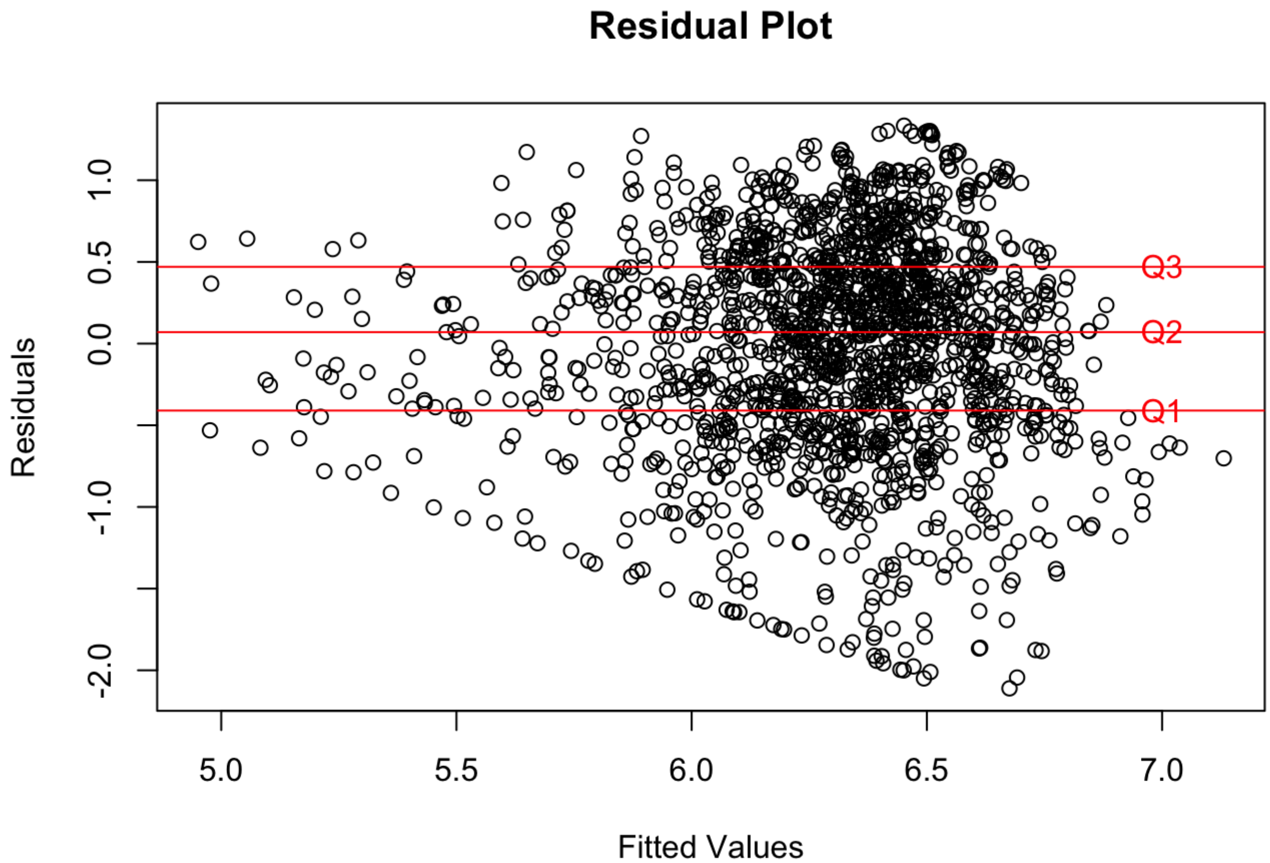
Introduction

We get this project from our client, Andrea Blackburn, an anthropology student at Boston University. This project studies how spatial parameters (altitude, slope, distance to river, and normalized difference vegetation index) predict an orangutan spatial point pattern. The orangutan point data was collected with handheld GPS units, and the spatial parameter data was extracted from landsat and DEM imagery using geographic information systems (GIS). In this project, we use the Kernel Density Value for each point as the dependent variable, and the geological information of the data points as independent variables, including Normalized Difference Vegetation Index (NDVI), the degree of the slope (0-90 degrees), elevation measured in meters, Euclidean distance from point to the river (meters).

Methods

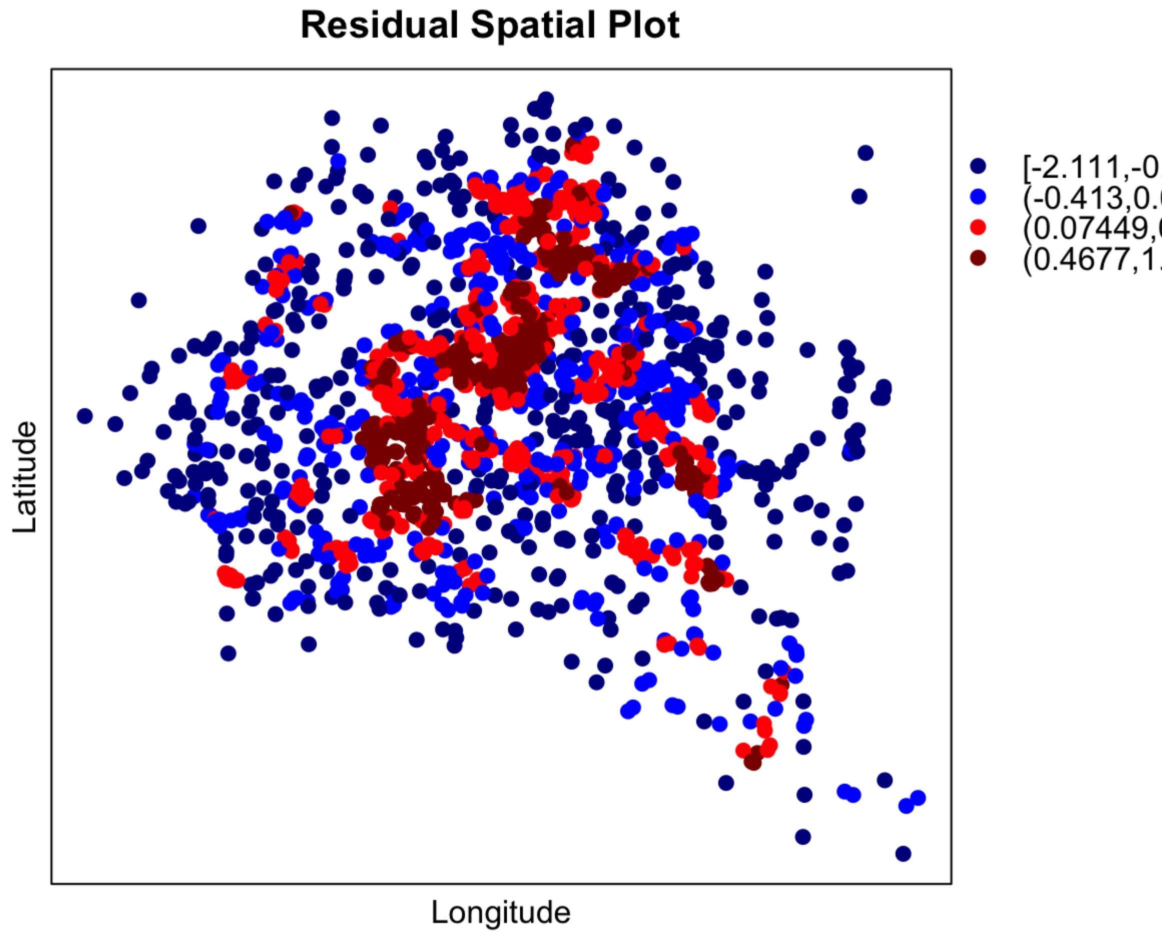
1. Multiple regression model

First, we run a multiple regression model lm(RASTERVALU ~ NDVI_v2 + slope1 + altitude + dist_riv_b) and look at its residual plot to check model assumptions.



The residual plot reveals a funnel shape, which indicates that there is a violation of heteroscedasticity assumptions. We tried to transform the response Y using a concave function (ex. logY, sqrt(Y)), but the shape wasn't improving much. Another option would be to fit a new model by weighted least squares, which geographically weighted regression(GWR) may be a good choice.

Another way to check is to plot the residuals by its quantile to see if there is any obvious spatial patterning.



From this plot it is apparent that there is some spatial patterning of the residuals (i.e. the red, dark red and blue points are not randomly distributed, but there appear to be small clusters of red, dark red and blue points in certain parts of the map). As there appears to be some spatial patterning in these residuals, we will now run a geographically weighted regression (GWR) model to see how the coefficients of the model might vary across space.

2. GWR Model

First we will calibrate the bandwidth of the kernel that will be used to capture the points for each regression and then run the GWR model.

```
## Adaptive q: 0.381966 CV score: 261775996
## Adaptive q: 0.618034 CV score: 279645632
## Adaptive q: 0.236068 CV score: 239575934
## Adaptive q: 0.145898 CV score: 207796245
## Adaptive q: 0.09016994 CV score: 179683484
## Adaptive q: 0.05572809 CV score: 151941088
## Adaptive q: 0.03444185 CV score: 125035078
## Adaptive q: 0.02128624 CV score: 102131946
## Adaptive q: 0.01315562 CV score: 83196736
## Adaptive q: 0.008130619 CV score: 66859624
## Adaptive q: 0.005024999 CV score: 55087813
## Adaptive q: 0.00310562 CV score: 46120938
## Adaptive q: 0.001919379 CV score: 42256632
## Adaptive q: 0.001186241 CV score: NA
## Adaptive q: 0.002372483 CV score: 43219380
## Adaptive q: 0.001458373 CV score: 105694743
## Adaptive q: 0.00174329 CV score: 45221355
## Adaptive q: 0.002110677 CV score: 42280944
## Adaptive q: 0.0020067 CV score: 42111101
## Adaptive q: 0.00204739 CV score: 42147487
## Adaptive q: 0.00196601 CV score: 42131076
## Adaptive q: 0.0020067 CV score: 42111101
```

```
## Call:
## gwr(formula = RASTERVALU ~ NDVI_v2 + slope1 + altitude + dist_riv_b,
##     data = OHD1, coords = cbind(OHD1$Latitude, OHD1$Longitude),
##     adapt = GWRbandwidth, hatmatrix = TRUE, se.fit = TRUE)
## Kernel function: gwr.Gauss
## Adaptive quantile: 0.0020067 (about 3 of 1721 data points)
## Summary of GWR coefficient estimates at data points:
##                     Min.     1st Qu.      Median     3rd Qu.        Max.
## X.Intercept. -1.2535e+04 -4.4579e+02  6.3131e+02  1.7081e+03  1.1860e+04
## NDVI_v2      -2.0257e+04 -1.9828e+03 -4.9991e+01  2.3047e+03  2.4758e+04
## slope1       -1.2838e+02 -5.3532e+00  1.9257e+00  9.8206e+00  1.7378e+02
## altitude     -1.9426e+01 -1.8616e+00 -5.4563e-02  1.5639e+00  1.9537e+01
## dist_riv_b   -1.1770e+01 -1.0940e+00  1.8199e-01  1.7501e+00  1.6116e+01
##                   Global
## X.Intercept. -2113.3680
## NDVI_v2       5689.9372
## slope1           7.8514
## altitude        -0.2179
## dist_riv_b      -0.3207
## Number of data points: 1721
## Effective number of parameters (residual: 2traceS - traceS'S): 948.8707
## Effective degrees of freedom (residual: 2traceS - traceS'S): 772.1293
## Sigma (residual: 2traceS - traceS'S): 144.2538
## Effective number of parameters (model: traceS): 757.0835
## Effective degrees of freedom (model: traceS): 963.9165
## Sigma (model: traceS): 129.1078
## Sigma (ML): 96.62329
## AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 23329.38
## AIC (GWR p. 96, eq. 4.22): 21373.83
## Residual sum of squares: 16067358
## Quasi-global R2: 0.954228
```

The output from the GWR model reveals how the coefficients vary across the Study region. You will see how the global coefficients are exactly the same as the coefficients in the earlier lm model. In this particular model, if we take altitude, we can see that the coefficients range from a minimum value of -19.43 (1 unit change in altitude resulting in a decrease in kernel density value of 19.43) to +19.54 (1 unit change in altitude resulting in an increase in kernel density value of +19.54). For half of the points in the dataset, as altitude rises by 1 meter, kernel density scores will increase between -1.86 and 1.56 points (the inter-quartile range between the 1st Quartile and the 3rd Quartile).

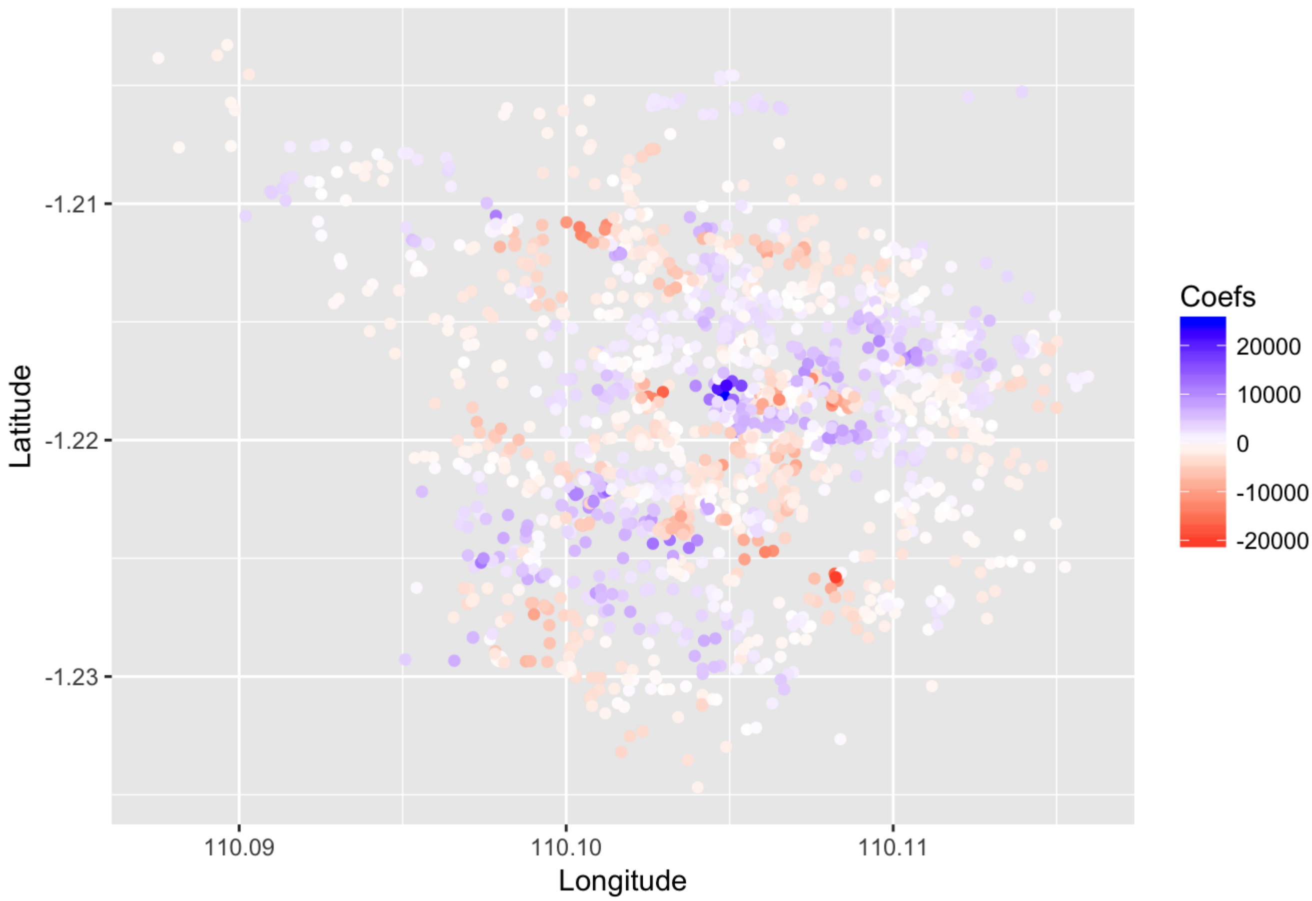
Coefficient ranges can also be seen for the other variables and they suggest some interesting spatial patterning. To explore this we can plot the GWR coefficients for different variables.

In regards of how to interpret the results, let’s take the first point as an example. The model we got for the first point after fitting GWR model is

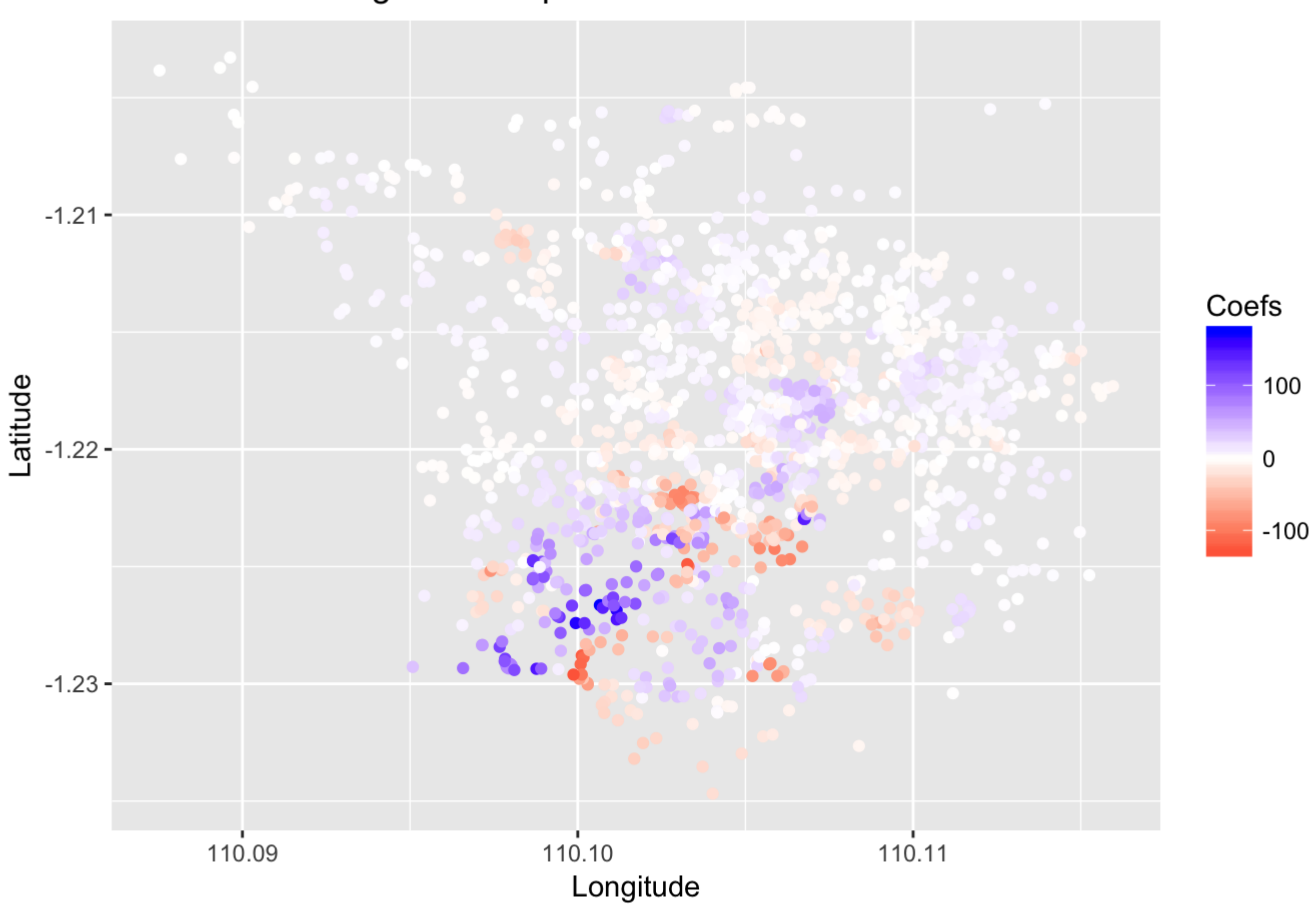
$$RASTERVALU = 880.98 + 858.16 * NDVI_v2 - 4.78 * slope1 + 1.17 * altitude - 4.27 * dist_riv_b$$

. It means that 1 unit change in Normalized Difference Vegetation Index (NDVI) results in an increase in kernel density value of 858.16, when holding other three variables the same and for 1 unit change in slope results in a decrease in kernel density value of 4.78 Also, for 1 unit change in altitude results in an increase in kernel density value of 1.17, while 1 unite change in Euclidean Distance from point to the river in meters results in a decrase of 4.27.

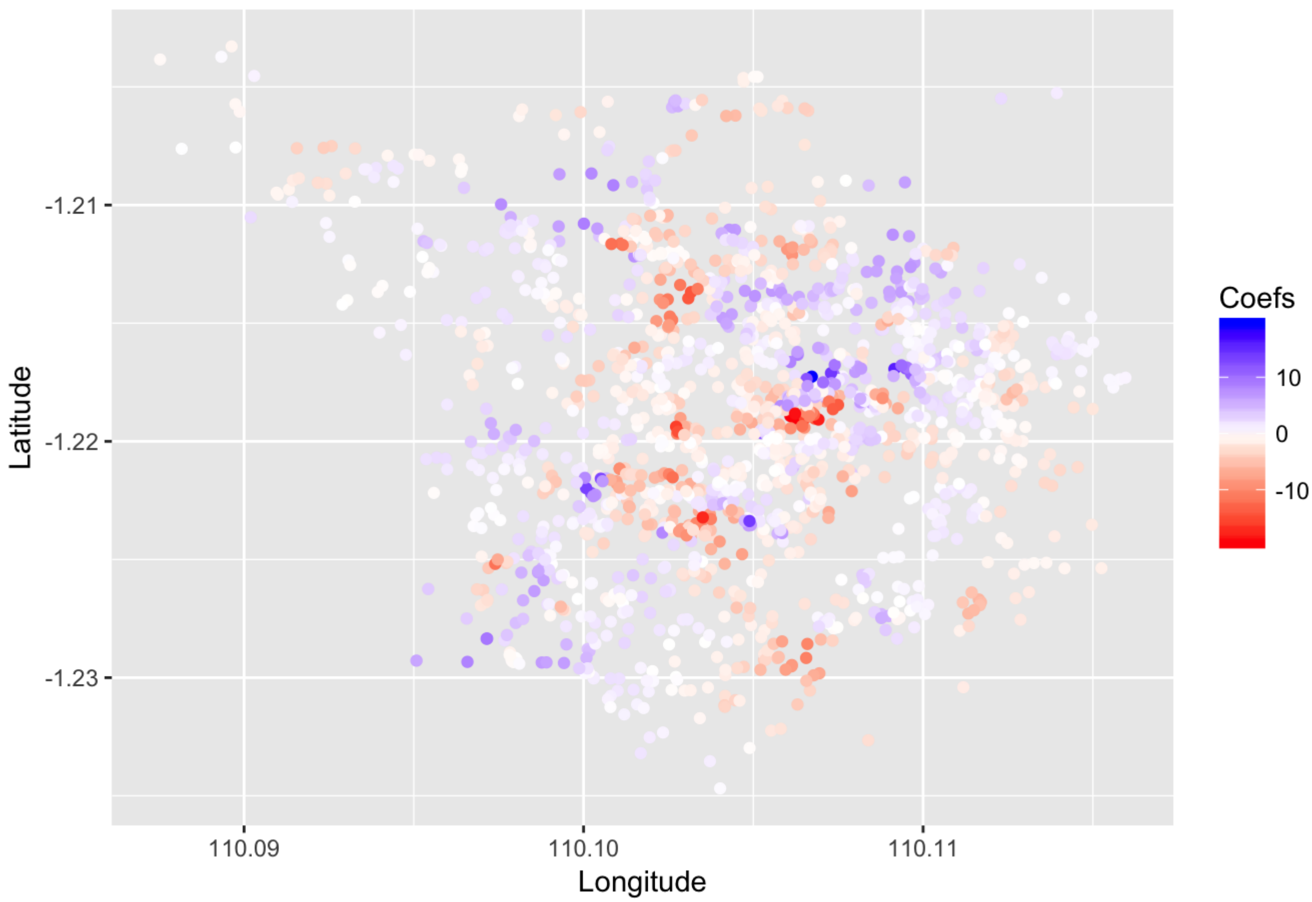
Coefficient of NDVI



Coefficient of Degree of Slope



Coefficient of Elevation



Coefficient of Euclidean Distance to River