NUS National University of Singapore

BT4015: Geospatial Analytics

Project Report

Group Members:

1. Yong Zi Ping (A0183529W)
2. Zhang Jiawei (A0155667R)
3. Isaac Hong (A0184054H)
4. Ng Hong Quan (A0187489A)

Table of Content

Context of Data

Main Dataset:	Spatial Data Repository
Supplementary Datasets:	Data Science for Good: Kiva Crowdfunding, Additional Resources for Kiva Crowdfunding, ACLED African Conflicts, Global MPI Tables (table 7) , GADM
Area of Focus:	Africa Continent
Year:	2014 - 2017
Kiva's Background:	Online crowdfunding platform to extend financial services to poor and financially excluded people around the world.

Multidimensional Poverty Index (MPI)

MPI will be one of the main focus of analysis in this project to evaluate the poverty level in the different countries and sub-national regions. It was developed by the Oxford Poverty & Human Development Initiative (OPHI) and the United Nations Development Programme (UNDP) to measure poverty in developing countries.

The MPI assesses poverty at both individual and household levels, providing a more complete representation of the deprivations experienced by each person. The measure takes into account the 3 main areas, which includes health, education and standard of living, as shown in Figure 0 below.

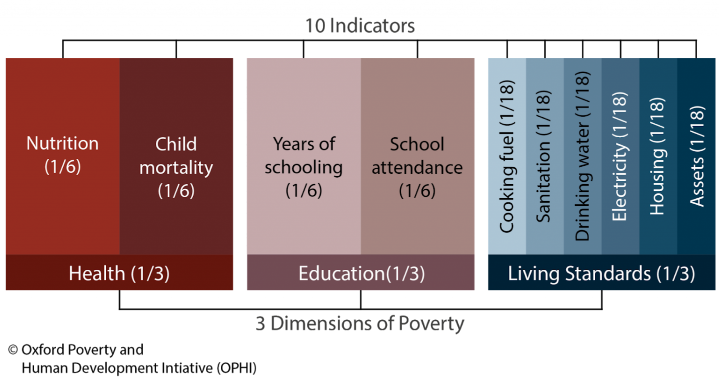


Figure 0: Indicators considered in the calculation of MPI

As the MPI is computed based on a standard set of indicators, it permits fair comparisons across countries and regions. Higher MPI figures will indicate a higher level of poverty experienced in a country or region. Hence, the MPI is especially valuable in our project to identify the more vulnerable regions which are experiencing higher levels of poverty and require more assistance.

Motivation

1) Help expand Kiva's initiative to alleviate global poverty in Africa Continent.
2) Evaluate Kiva's past distribution of loans and its effectiveness in Africa Continent.
3) Understanding poverty by gaining insights into contributing factors and thereafter provide recommendations of better ways to distribute and set investment priorities.
4) Propose alternative measurement of poverty in areas where MPI is absent.

Target Insights

1) Identify Kiva's area of focus for the past 4 years by performing spatial temporal analysis
2) Evaluate effectiveness of Kiva's loans on Africa Continent via spatial operations, in particular the use of boolean operators.
3) Propose future area of focus, both in global (countries) and local (regions) context based on presence of clusters with high MPI attained via spatial autocorrelation between neighbouring spatial data points.
4) Gain insights into contributing factors to poverty, with specific emphasis on environmental factors and accessibility via point pattern analysis.
5) Recommends new approach in distributing loans, specifically in the sector for funded activities to future areas of focus.
6) With the use of spatial operations and insights from poverty analysis, propose alternative measurement of poverty in areas where MPI is absent.

Data Description

File Name	**Data Type**	**Attributes**	**Number of Rows**	**Number of Columns**
MPIData_augmented.csv	SpatialPolygon DataFrame	Loan_amount, Temperature, Elevation, Precipitation, TimeToCity	994	31
kivaData_augmented.csv	SpatialPolygon DataFrame	MPI_Country	671,205	43
all_loan_theme_merged_with_geo_mpi_regions.csv	SpatialPointsDataFrame	geo	15,736	31
african_conflicts.csv	SpatialPointsDataFrame	LATITUDE, LONGITUDE	165,808	29
Tables_7_MPI_estimations_country_levels.xlsx *Did not directly use this dataset as it requires further preprocessing, Used MPI_old.csv instead*			268	24
MPI_old.csv *Cleaned dataset from 'Table-7_All-MPI.csv'*			262	4
Various sp objects in RDS format representing sub-national regions	SpatialPolygon DataFrame		14 sp objects, representing 7 countries	

Figure 1: Table of Raw Data Used

Descriptive Analysis

Prior to conducting spatial data analysis, descriptive analysis is conducted to summarize our historical data to identify patterns or trends within Africa Continent. This yields useful information from the various attributes and possibly prepares the data for further spatial analysis.

<u>Numerical Data</u>

<u>MPI across Countries</u>

The plot seeks to summarize and rank the countries in Africa on their MPI. Higher MPI figures indicate higher levels of poverty. More resources should be directed to these countries to alleviate their poverty level.

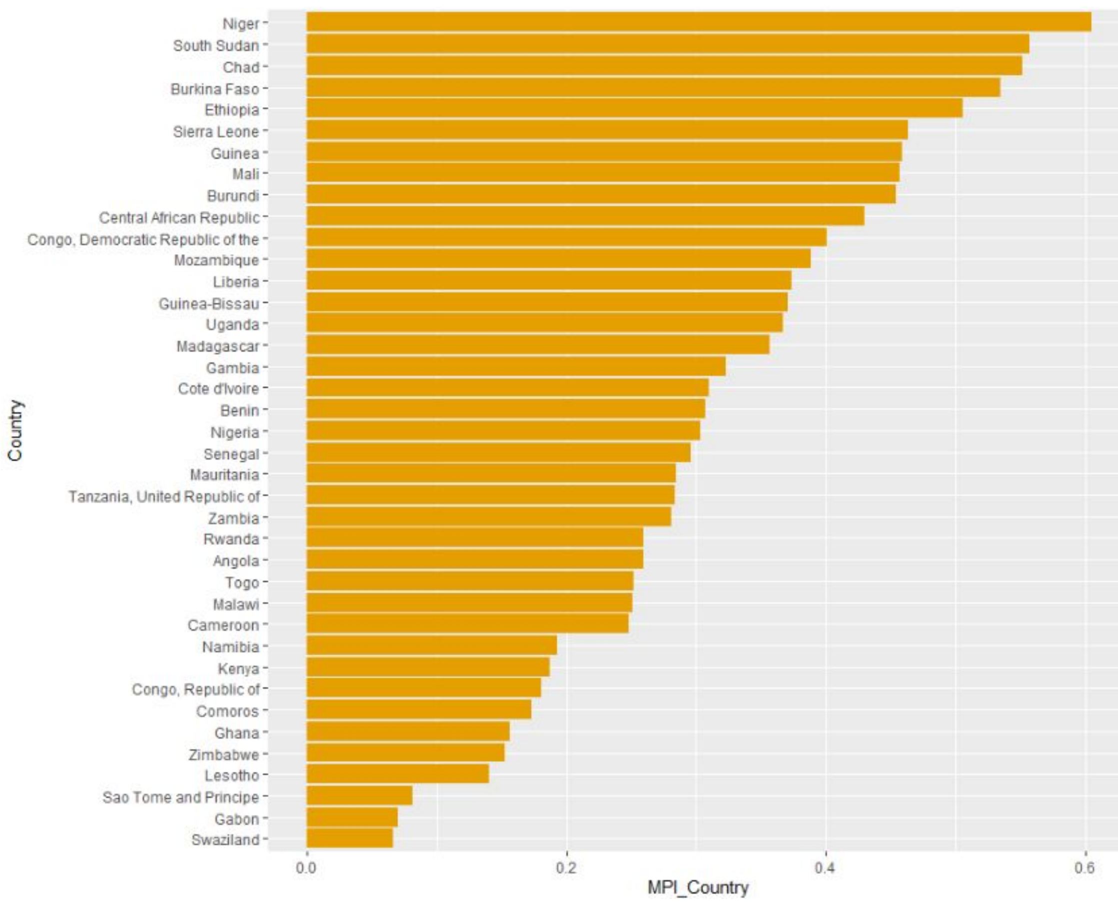


Figure 2: MPI across Countries

Number of Loans across Countries

The plot suggests the distribution of loans by Kiva for each country sorted in descending order. Here, countries like Kenya and Uganda are being targeted by Kiva which is contrary to the MPI ranking above. The distribution of loans also seemed to be uneven and disproportionate. This encourages further analysis.

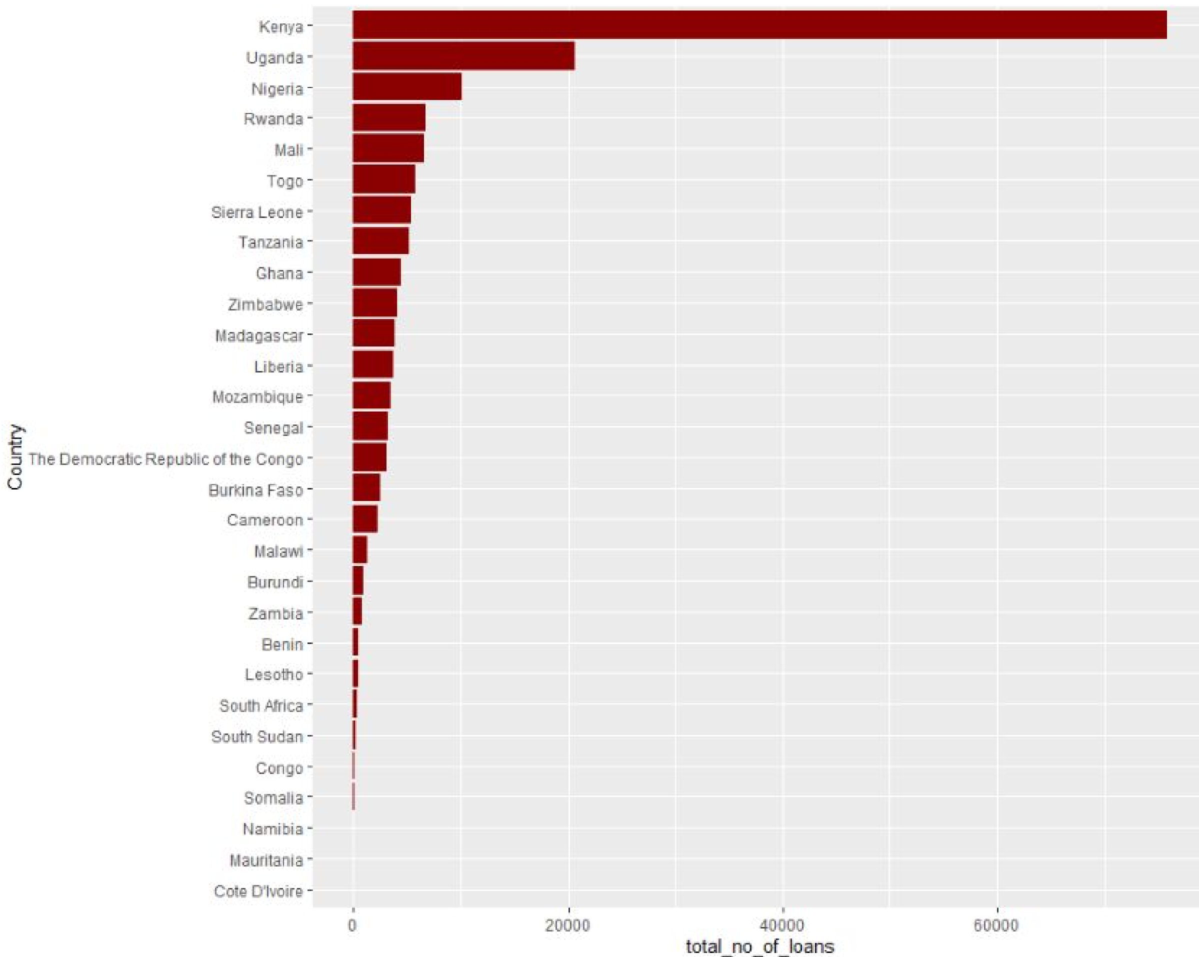


Figure 3: Total Number of Loans across Countries

Total Amount of Loans across Countries

The plot suggests the distribution of the total amount of loans by Kiva for each country sorted in descending order. Here, countries like Kenya and Rwanda are being provided with more funds by Kiva which is contrary to the MPI ranking above. Besides, the total amount of loans are not totally consistent with the number of loans. The distribution of loan amounts also seemed to be uneven and disproportionate. This encourages further analysis.

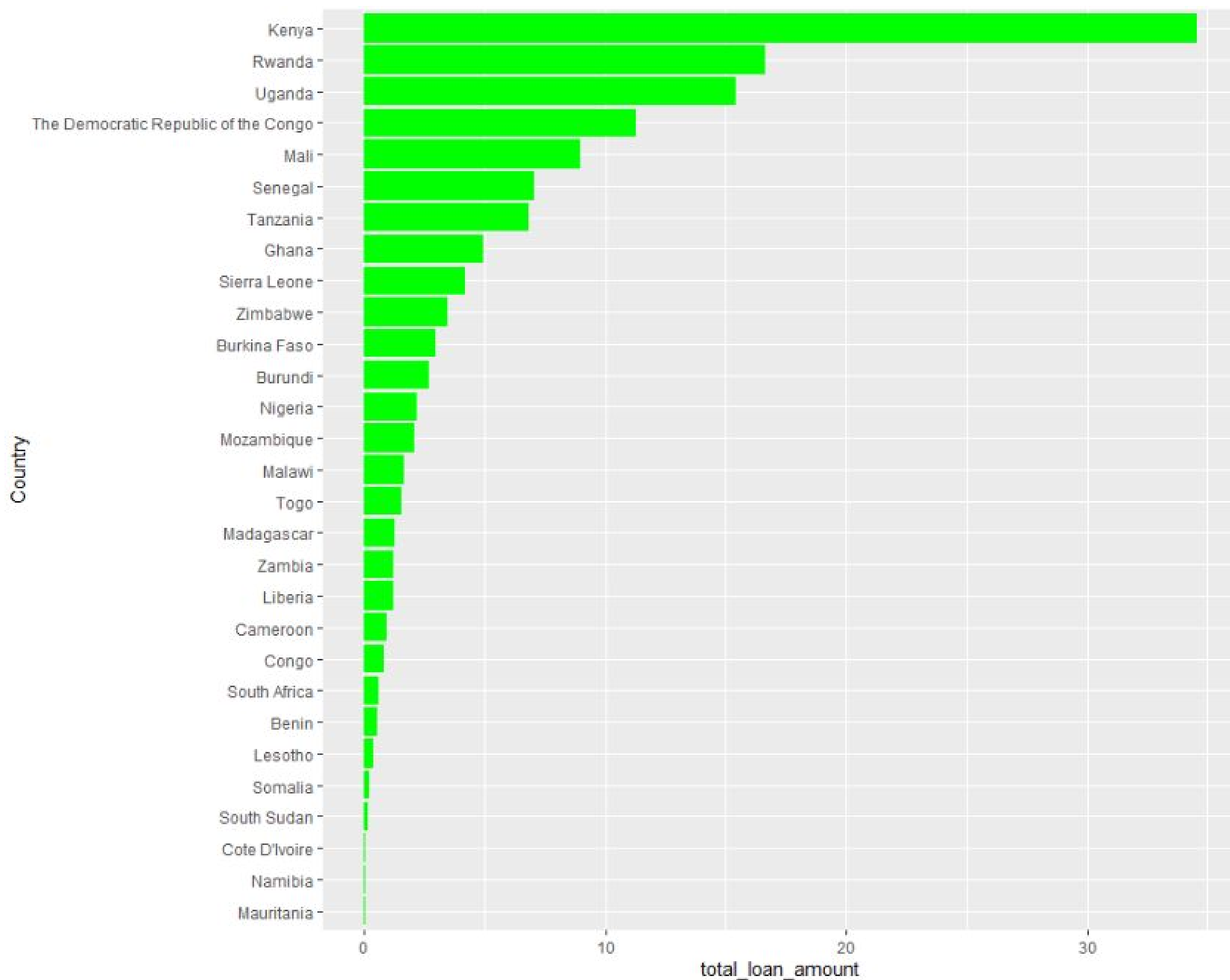


Figure 4: Total Amount of Loans across Countries

Gender of Loanees

The demographics and in specific, genders is plotted to analyse if there is a correlation between gender and loans by Kiva.

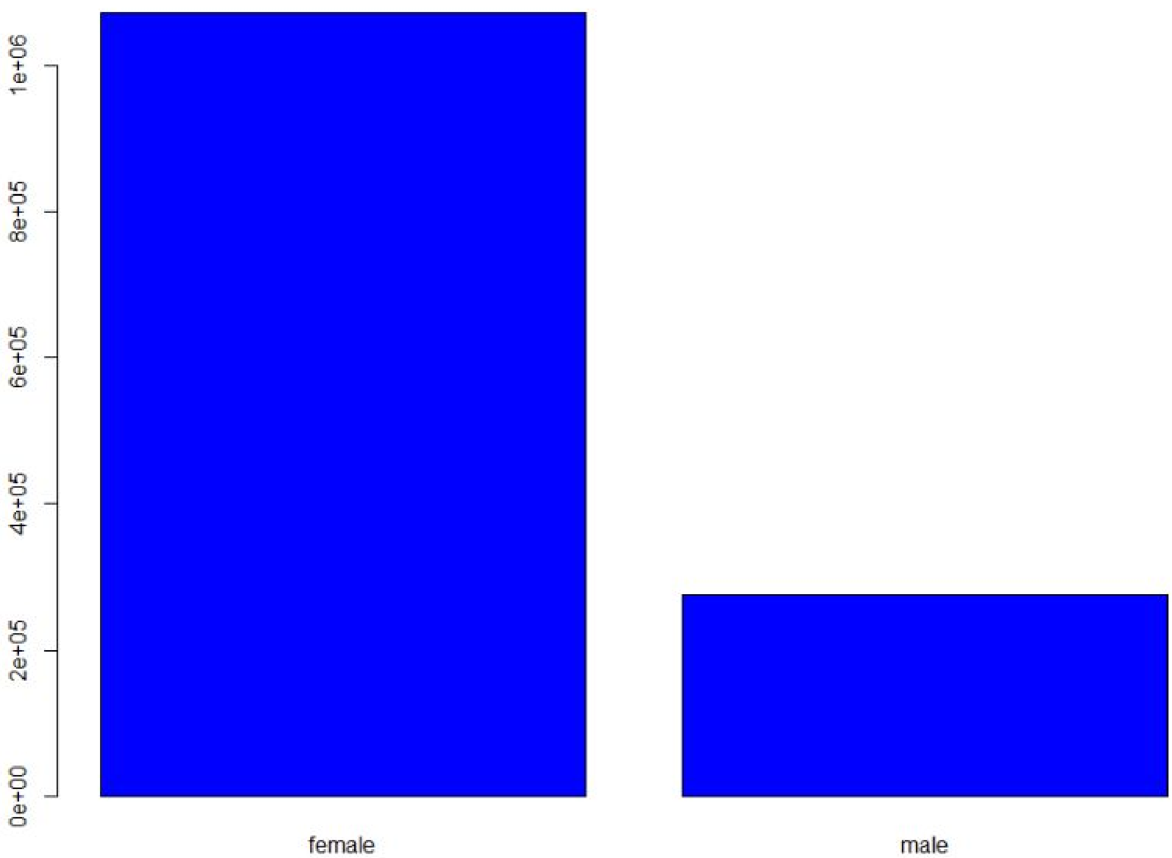


Figure 5: Gender of Loanees

Spatial Data

Points of Loans Application

Below is the point plot of individual loan applications from 2014 - 2017. These spatial points are important as they will be used as a proxy to measure the poverty levels at certain points in Africa continent. From the plot (Figure 6), some form of clustering does exist and to investigate further into that, Kernel Density Estimates (KDE) will be performed.

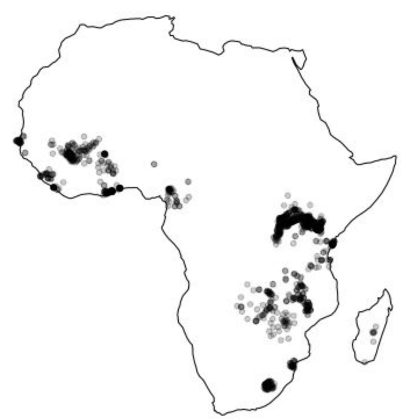


Figure 6: Points of Loans Applications in Africa

Density Based Analysis

To identify clusters, Kernel Density Estimates (KDE) is first employed to convert from discrete points to continuous density estimates allowing the detection of hotspots and local variations. To ensure that various radii are considered, 3 levels of kernel bandwidth is plotted (Figure 7). The plots do show signs of clustering with specific hotspots narrowed when kernel bandwidth decreases, implying that there are certain regions that appear to be submitting more loan applications.

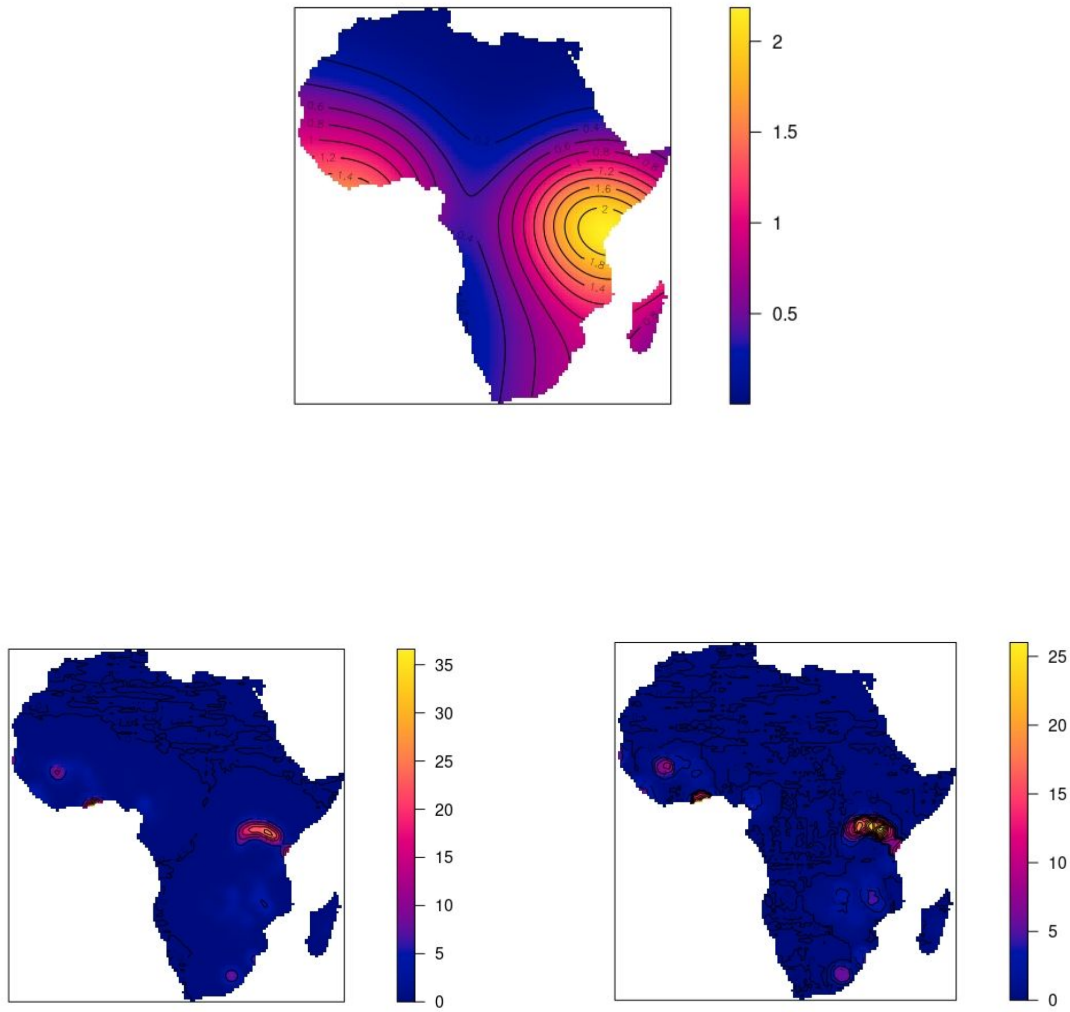


Figure 7: Kernel Density Estimates for First (top); Second (left); Third (right) Order

To further emphasize the point on hotspots, quadrat analysis is further employed to examine the distribution based on frequency of occurrence. Quadratic density of 10x10 grid is used (Figure 8). The bright grid indicates the existence of hotspots and is coherent with KDE analysis.

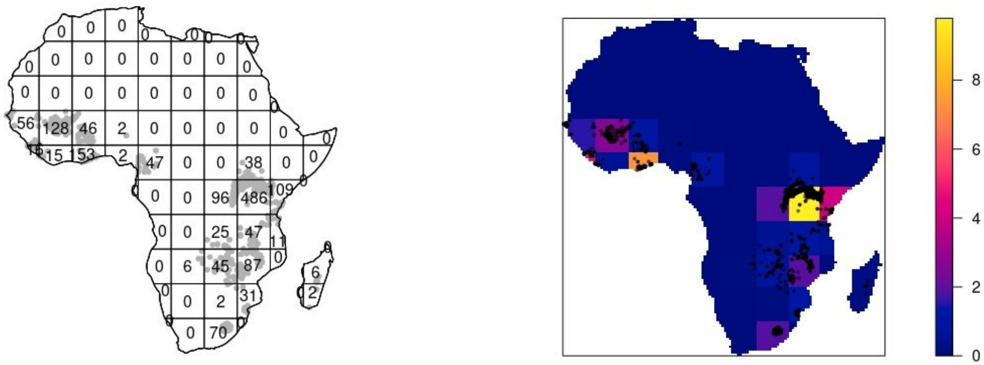


Figure 8: Quadrat Analysis for Loan Applications

Distance Based Approach

Nearest neighbor analysis is used to determine the spatial arrangement of the loans application. As from the ANN plot (Figure 9, Left), at 1000 neighbours, clusters can be seen as the line is no longer smooth and continuous. To further check for point patterns, K, L and g functions are subsequently conducted to ensure coherence in results.

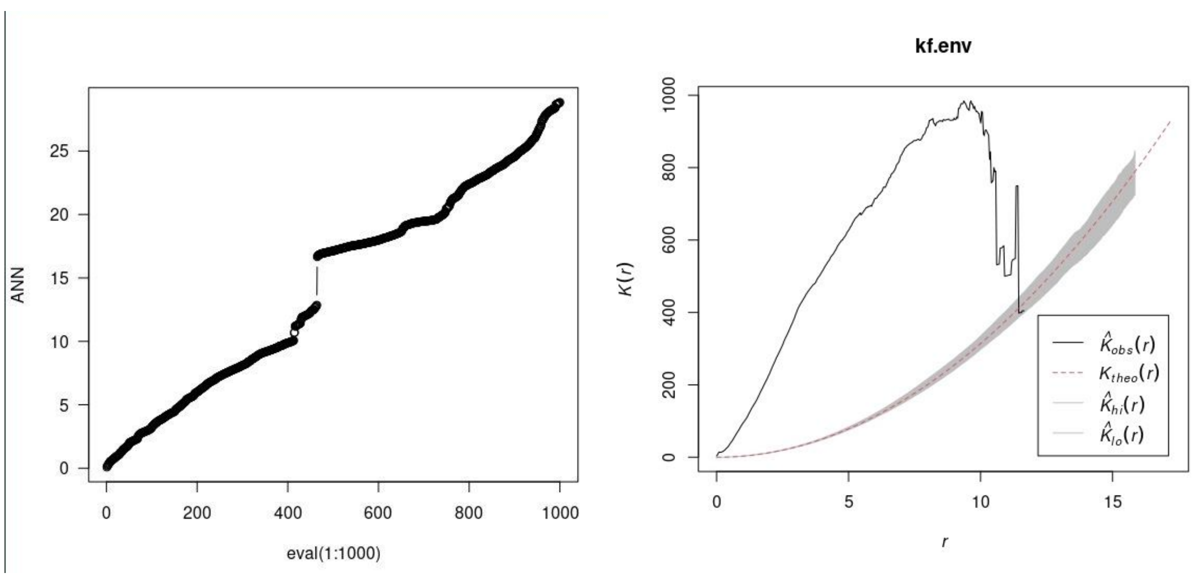


Figure 9: Average Nearest Neighbours Distance for 1000 Neighbours (left)
K function Plot (right)

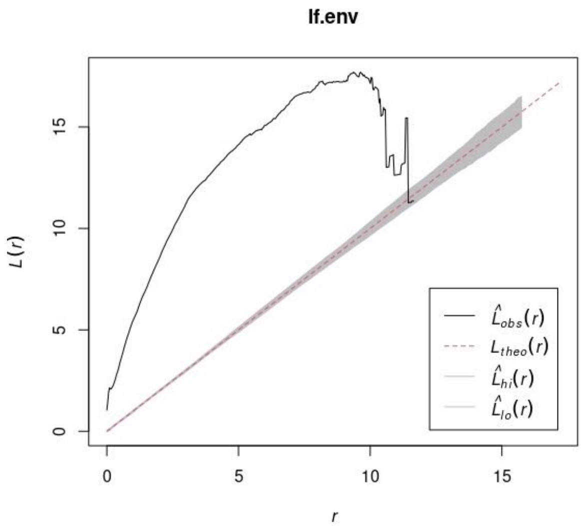


Figure 10: L function Plot

As seen from the K function (Figure 9, Right), with reference to the IRP/CSR process represented by the red dotted line (K_{theo}), there exist various distances/ radii where the observed K values (K_{obs}) are above the reference line (K_{theo}), implying the presence of clustering at a given distance band. This is further verified by both the L (Figure 10, left) and G function plots (Figure 10, right).

The L function is a simple transformation of the K-function to allow the L_{theo} to have a linear slope of 1. Hence, the L function plot will allow us to visualize the difference between K_{obs} and K_{theo} more easily, whereas the G function will describe the clustering based on cumulative distribution of the nearest neighbour distance by summing only points that fall within the narrow distance band. In both L (Figure 10, left) and G function plots (Figure 10, right), the observed values (L_{obs} and G_{obs}) also lie above the red dotted lines which represent the IRP/CSR process (L_{theo} and G_{theo}). Therefore, all 3 plots are indicative of a strong presence of clusters.

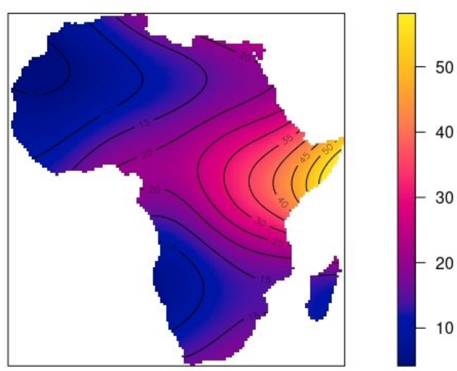
<u>Conflict Points</u>

The point plot for conflicts is plotted over the years (Figure 11). Before delving into deeper analysis, clustering of points will first be determined. From the plot, some form of clustering does exist and to investigate further into that, Kernel Density Estimates (KDE) will be performed.

Figure 11: Points of Conflict Events in Africa

Density Based Analysis

Similarly, to ensure that various radii are considered, 3 levels of kernel bandwidth is plotted (Figure 12). The plots do show signs of clustering with specific hotspots narrowed when kernel bandwidth decreases, implying that there are certain regions that appear to be experiencing more frequent conflicts.



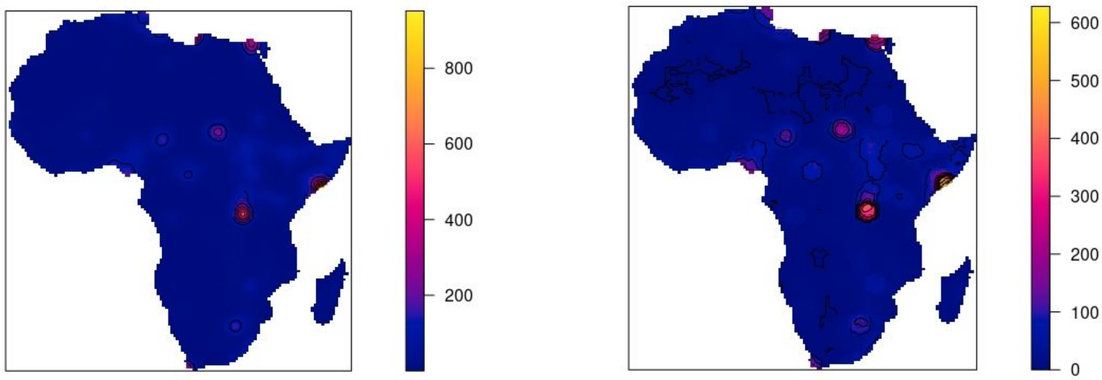


Figure 12: Kernel Density Estimates for First (top); Second (left); Third (right) Order

To further emphasize the point on hotspots, quadrat analysis is further employed to determine the distribution based on frequency of occurrence (Figure 13). 10x10 grids are used and the bright grid indicates the existence of hotspots and is coherent with KDE analysis.

Figure 13: Quadrat Analysis for Conflict Events

Distance Based Approach

Nearest neighbor analysis is used to determine the spatial arrangement of the loans application. K, L and g functions are subsequently conducted to ensure coherence in results.

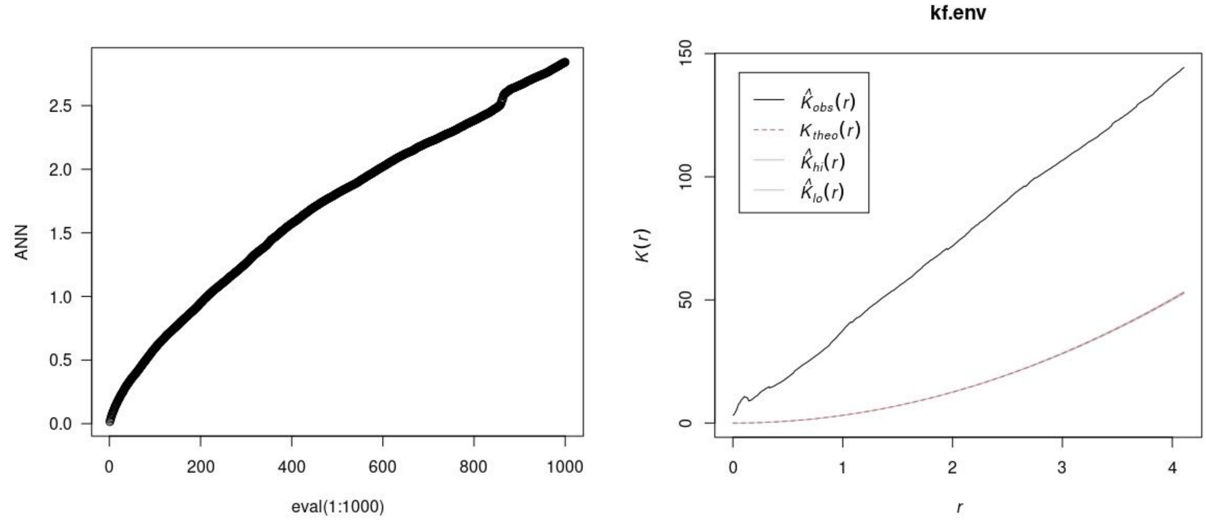


Figure 14: Average Nearest Neighbours Distance for 1000 Neighbours (left)
K function Plot (right)

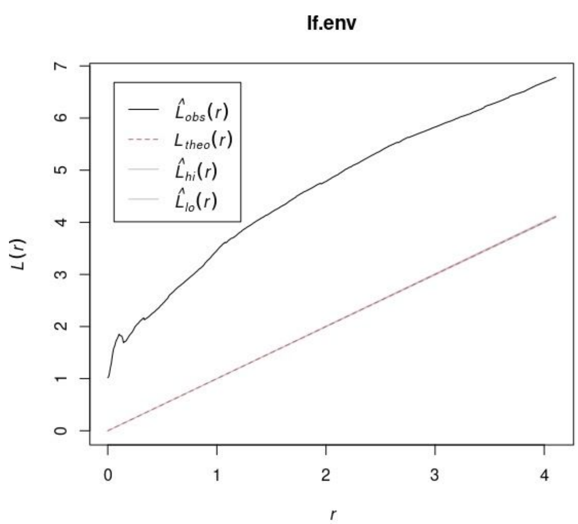


Figure 15: L function Plot

As seen from the K function (Figure 14, Right), with reference to the IRP/CSR process represented by the red dotted line (K_{theo}), there exist various distances/ radii where the observed K values (K_{obs}) are above the reference line (K_{theo}), implying the presence of clustering at a given distance band. This is further verified by both the L (Figure 15).

With the identification of the various hotspots using both density based analysis and distance based approach, it motivates us to layer different raster plots and test our hypothesis to identify the contributing factors to poverty in the following section of spatial data analysis.

Spatial Data Analysis

<u>Kiva's Evaluation</u>

Before generating insights, understanding and evaluating current approaches will be key to determine our focus of analysis. This section will study Kiva's approach in providing loans from 2014 - 2017 and in specifics, 3 main aspects will be discussed:

1) Identification Kiva's area of focus for the past 4 years
2) Evaluation on effectiveness of Kiva's loans on Africa Continent
3) Proposal on future area of focus in country and regional level

<u>Identification Kiva's Area of Focus for the Past 4 Years</u>

To identify Kiva's focus in the Africa continent, the aggregation of loan amounts across all 4 years for each country will be plotted. A choropleth plot is chosen, with the gradient of colours determining the proportion of aggregate loan amounts for each country (Figure 16). As from the plot below, Kiva appears to be investing heavily in the Eastern side of Africa.

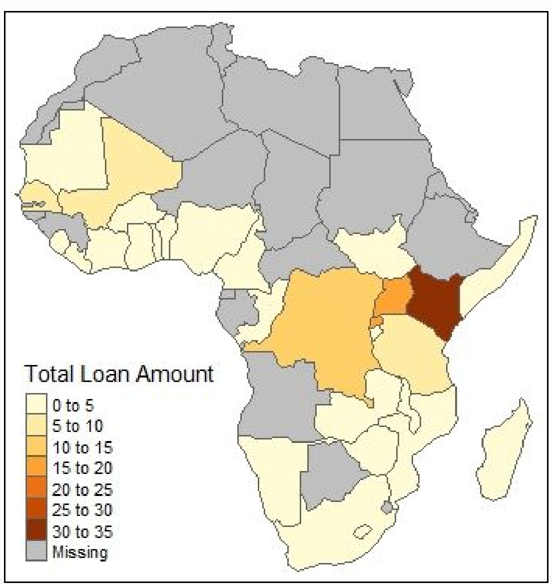


Figure 16: Aggregate Loan Amounts for Each Country

To analyse if a distribution pattern exists in Kiva's approach, a hypothesis is then generated to measure if there exists spatial autocorrelation in total loan amounts between countries for which Kiva distributes their loans to. In order to test the hypotheses, the Global Moran Method is employed (Figure 17). Here, the neighbouring polygons are defined using the Queen's Method. The simulation-based approach is used in this context as compared to the theoretical approach to eliminate the need for any restrictive assumptions which may not hold true for the dataset. Hence, Monte-Carlo simulation of Moran I is conducted instead of the Moran I test under the normality assumption.

```
        Monte-Carlo simulation of Moran I

data:  sub_africa.sp$MPI_Country
weights: sub_africa.sp.lw
number of simulations + 1: 10001

statistic = 0.15231, observed rank = 9244, p-value = 0.07569
alternative hypothesis: greater
```

Figure 17: Global Moran on Total Loan Amounts

From the results, by setting the level of significance at 0.05, the p-value from the test is 0.076 and hence the null hypothesis cannot be rejected, suggesting the absence of spatial autocorrelation in the distribution of loan amounts across countries.

However, as Global Moran only analyses if there is any presence of spatial autocorrelation in the study area at a global level, it has the tendency to omit any local clustering at a lower level. Hence, to further analyse the local effects between neighbouring countries, Local Moran's Method is employed. Again, the neighbouring polygons are defined using the Queen's Method and the computation of average neighbouring loan amounts are performed for each country. Moran's statistics are plotted on the choropleth map where the higher the statistics, implies correlation of loan amounts across the countries and the presence of a cluster (Figure 18, Left).

Monte Carlo test is then conducted to determine if the clusters are significant by testing against the hypothesis that the computed Moran's statistic is significantly different from 0. As from the plotted choropleth plot (Figure 18, Right), only one cluster is deemed to be significant and suggest a presence of a cluster of countries where Kiva has focused on over the years.

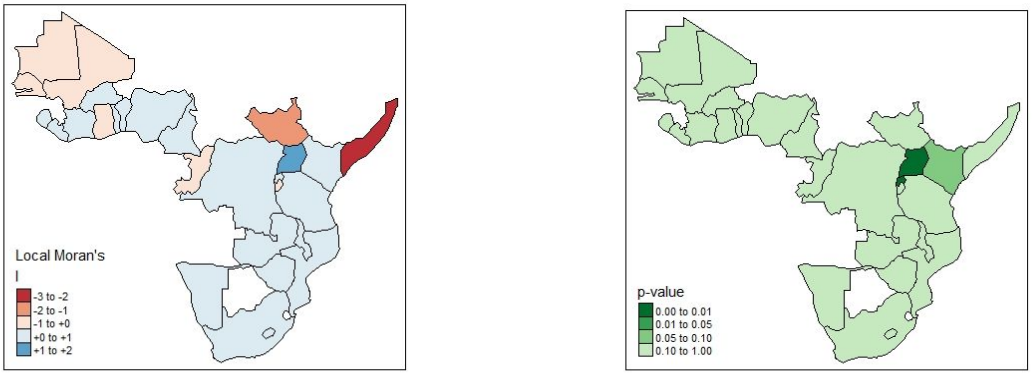


Figure 18: Local Moran Statistic on Total Loan Amounts (left)
P-value of Local Moran Statistic (right)

To further delve into understanding how Kiva changed their approach in handling the loans, temporal analysis via spatial operations will be performed. From the plot, the difference in the amount of loans Kiva provides between the year 2017 and 2014 can be inferred from the colours on a choropleth plot where green represents an increase, while red represents a decrease (Figure 19). This implies that Kiva has shifted their focus away from the Eastern side of Africa with the reduction in loan amounts over the years, as indicated by the red and orange regions. In addition, it appears that Kiva has shifted their focus from larger to smaller countries in general.

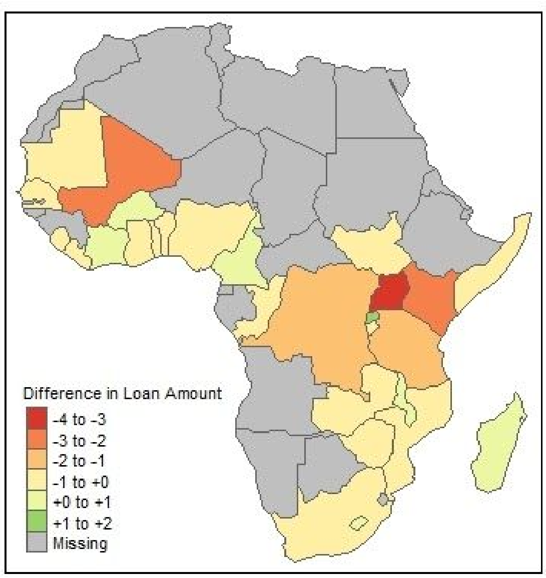


Figure 19: Temporal Analysis on Total Loan Amounts between 2014- 2017
(Green represents an increase, while Red represents a decrease)

Evaluation on Effectiveness of Kiva's Loans on Africa Continent

Effectiveness is defined as the improvement in MPI as compared to the amount of loans that Kiva provides across the years. Having visualized the amount of loans above, MPI improvement must be understood by conducting temporal analysis via spatial operations. Here, choropleth plot is also used, with the color representing the changes in MPI from 2014 to 2017 (Figure 20).

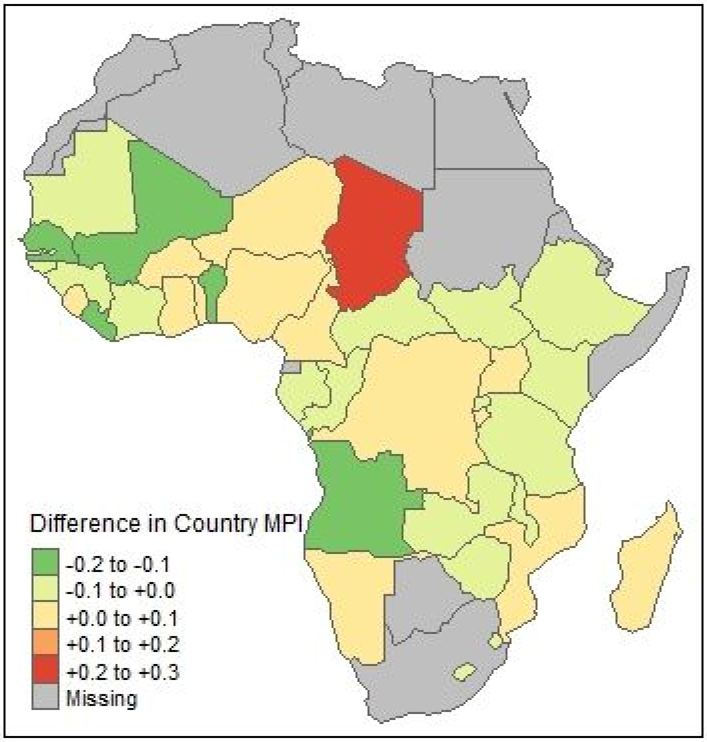


Figure 20: Temporal Analysis on MPI between 2014- 2017 across 2014- 2017
(Green represents an improvement, while Red represents deterioration)

The presence of orange and red polygons implies that there are countries having worsened poverty levels over the years. To visualize the effectiveness of Kiva loans, 4 different scenarios are crafted and our focus on analysis will be directed based on the results. Below are the possible scenarios and interpretations on Kiva's approach respective to their colors (Figure 21).

MPI Increase Loan Amount Increase	MPI Decrease Loan Amount Increase
MPI Increase Loan Amount Decrease	MPI Decrease Loan Amount Decrease

Figure 21: Table of Possible Scenarios

The green scenarios are the optimal scenarios where the MPI in the countries decreases which implies an improvement, be it due to Kiva's loans or not. Red implies the situation where the poverty index rises/ worsened despite the injection of a larger number of loans. While there are many factors that contribute to the poverty levels, there is a high probability and a reasonable suspicion on the poor distribution of the loans by Kiva. Blue, on the other hand, implies that poverty worsened in countries and Kiva has reduced the amount of loans. In such a situation, our proposal would be for them to focus on such countries which will be further elaborated later.

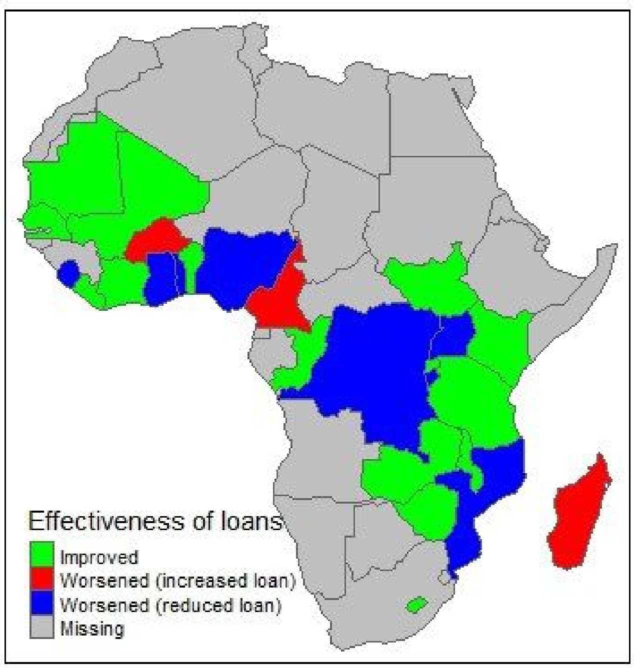


Figure 22: Scenarios that Countries Fall Under

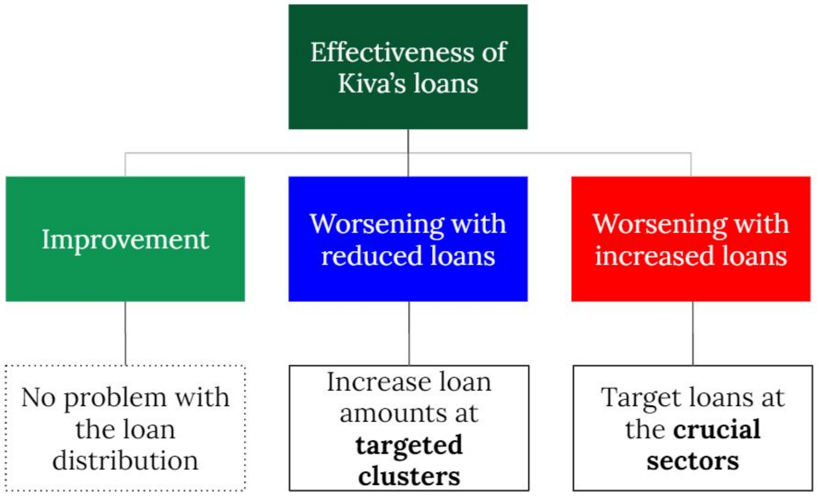


Figure 23: Chart of Recommendations for Possible Scenarios

Proposal on Future Area of Focus in Country and Regional Level

The presence of many polygons falling within blue and red scenarios calls for a real concern (Figure 22). This motivates the focus of analysis to generate 2 main insights:

1) Clusters of countries and regions with high MPI index
2) Sectors/ Industry to allocate loans to for individual countries

Identification of Clusters with High MPI

Country Level

To identify the spatially autocorrelated clusters that require urgent attention, Local Moran is conducted on the MPI of countries. The purpose for the analysis is to:

1) Identify clusters in Africa Continent that require assistance/ loans (High MPI)
2) Monitor if MPI across countries are spatially autocorrelated hence the assistance to a carefully selected country will benefit neighbouring countries

In the Local Moran method, the neighbouring polygons are defined using the Queen's Method and the computation of average neighbouring MPI is performed for each country. Moran's statistics are plotted on the choropleth map where the higher the statistics, implies spatial autocorrelation of MPI across the countries and the presence of a cluster (Figure 24, Left).

Monte Carlo test is then conducted to determine if the clusters are significant by testing against the hypothesis that the computed Moran's statistic is significantly different from 0. As from the plotted choropleth plot (Figure 24, Right), the Northern side of Africa is deemed to be significant which suggests the presence of clustering of countries and correlation in MPI among countries.

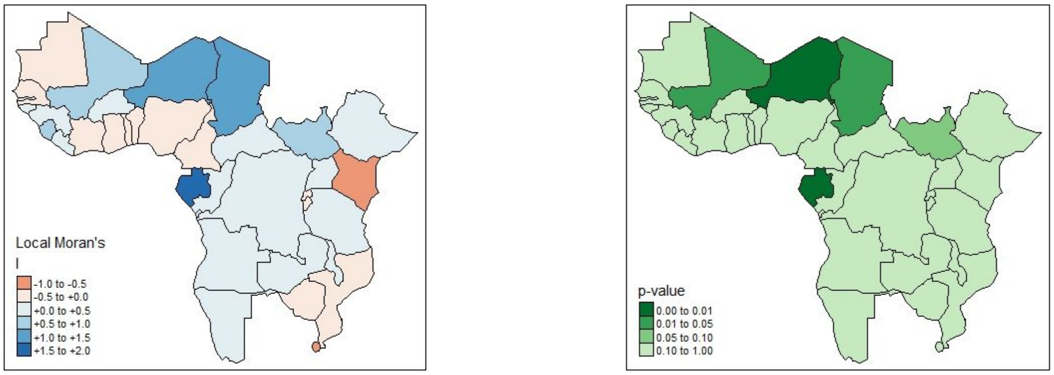


Figure 24: Local Moran Statistic on MPI (left)
P-value of Local Moran Statistic (right)

Having identified clusters and the presence of correlation in some of the neighbouring polygons, the next focus is to identify the clusters with high MPI index to ensure the target countries satisfy 2 criterion mentioned above (Figure 25); in urgent need of assistance and have a higher probability of benefiting neighbouring countries due to the presence of high correlation in MPI.

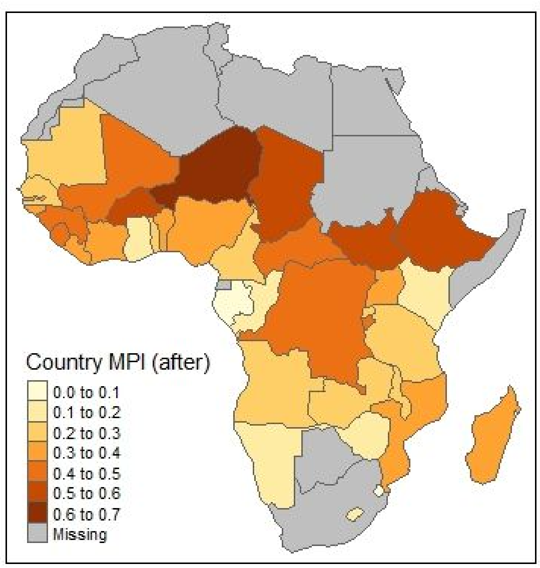


Figure 25: MPI of Countries

The plot of MPI across countries coincides with the significant clusters as seen in Figure 24. The clipped out countries (Figure 26) will hence be the area of focus in the future as these countries satisfy the above mentioned conditions.

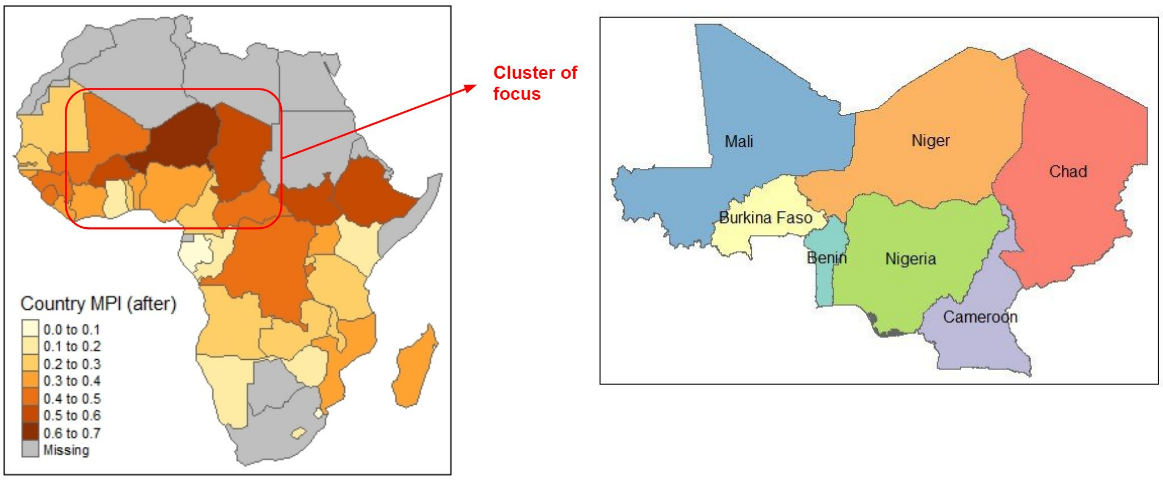


Figure 26: Countries of Focus (Clipped)

Regional Level

While the countries of focus have been sieved out, it would be naive to set such a huge area of land to target the loans on. Coupled with the reasonable assumption that not all parts of a country are suffering from the same level of poverty, a microscopic and more local analysis is conducted at the regional level, to identify hotspots in countries to employ a more targeted injection of funds.

The presence of many regions that are not too spatially far apart from each other motivates the evaluation of both the first and second order effects via the use of spatial autoregression model, Conditional Autoregressive (Figure 27) and Simultaneous Autoregressive (Figure 28) respectively.

```
Call:
spautolm(formula = MPI_Region ~ 1, data = sub_cluster_mpi, listw = sub_cluster_mpi.lw,
    family = "CAR")

Residuals:
      Min        1Q    Median        3Q       Max
-0.393077 -0.051829  0.012845  0.070084  0.427007

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.232309   0.099304 -2.3394  0.01932

Lambda: 0.9882 LR test value: 100.3 p-value: < 2.22e-16
Numerical Hessian standard error of lambda: NaN

Log likelihood: 73.00583
ML residual variance (sigma squared): 0.013958, (sigma: 0.11814)
Number of observations: 120
Number of parameters estimated: 3
AIC: -140.01
```

Figure 27: Conditional Autoregressive Summary of MPI Regional

```
Call:
spautolm(formula = MPI_Region ~ 1, data = sub_cluster_mpi, listw = sub_cluster_mpi.lw,
    family = "SAR")

Residuals:
      Min        1Q    Median        3Q       Max
-0.347938 -0.084045  0.015544  0.079034  0.401247

Coefficients:
            Estimate Std. Error z value  Pr(>|z|)
(Intercept) 0.364658   0.050894   7.165 7.778e-13

Lambda: 0.77465 LR test value: 91.365 p-value: < 2.22e-16
Numerical Hessian standard error of lambda: 0.055159

Log likelihood: 68.5367
ML residual variance (sigma squared): 0.015784, (sigma: 0.12564)
Number of observations: 120
Number of parameters estimated: 3
AIC: -131.07
```

Figure 28: Simultaneous Autoregressive Summary of MPI Regional

From the 2 analyses, both lambdas at 2.22e-16 and are significantly different from the null hypothesis of 0, hence the hypothesis can be rejected and imply the presence of possible spill-overs effects between different regions. These will be imperative in setting the future area of focus as it will fulfil the aforementioned criterion; to identify regions that are spatially autocorrelated and high in MPI such that assistance to a carefully selected region will also benefit neighbouring regions.

Next, to identify clusters of regions to focus on, Local Moran Method is again employed with Queen's Method selected to determine the neighbouring regions (Figure 29). The choropleth plot shows the colour gradient where the darker the color, implies a stronger correlation.

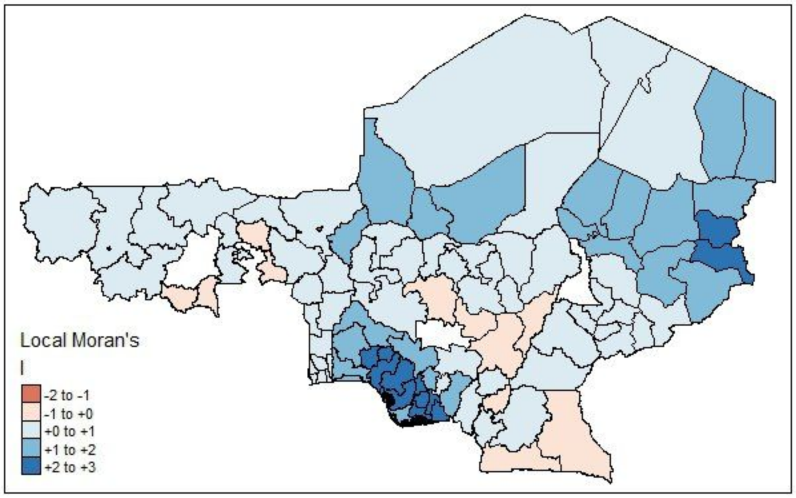


Figure 29: Local Moran Statistic on MPI Regional

Similarly, to test for significance from 0, a Monte Carlo test is conducted. As from the plotted choropleth plot (Figure 30), There are 3 clusters present that exhibit strong correlation.

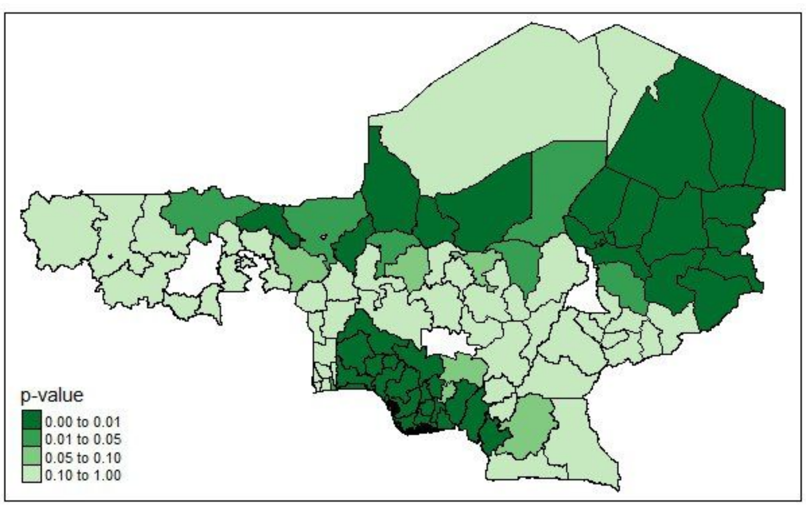


Figure 30: P-value of Local Moran Statistic (right)

Lastly, to narrow down our focus even further, as per the above criterion to sieve out only regions with urgent need for assistance, chosen focus regions must exhibit high MPI in that region. The map of MPI region with respect to their cluster is plotted (Figure 31) and is evident that only 2 out of the 3 clusters are both high in MPI and strong in correlation, namely the cluster to the Central and Eastern side of the clipped map.

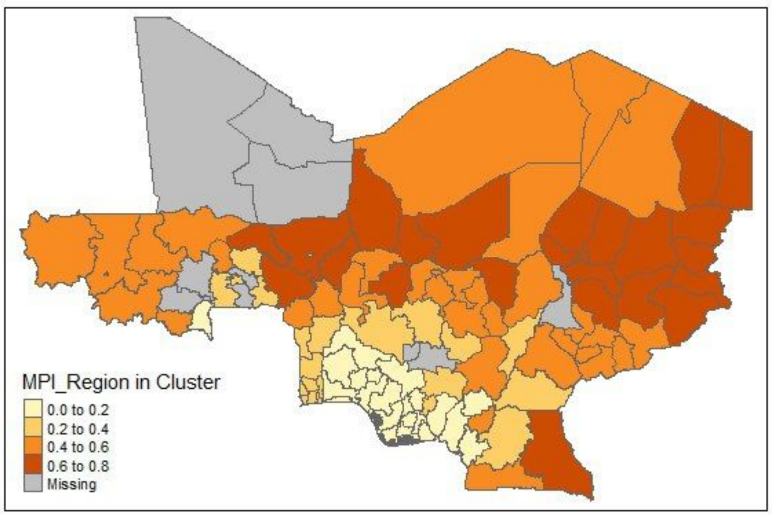


Figure 31: MPI Region across Regions in Clipped Countries

Upon further map operations by clipping, the regions of focus and the countries they belonged to are presented in the clipped polygon (Figure 33A, 33B), as the clusters we have identified in Figure 32.

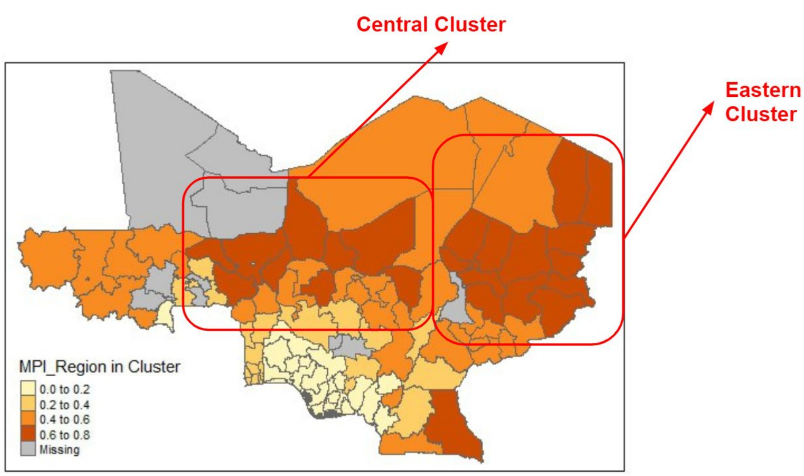


Figure 32: Identification of 2 clusters on a regional level

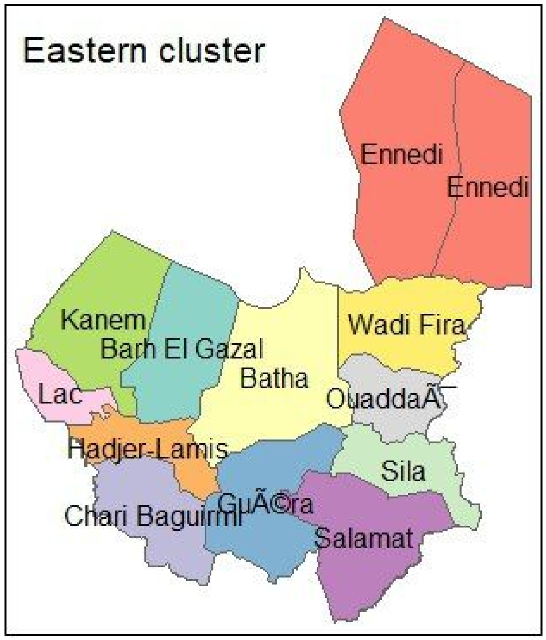


Figure 33A: Eastern Regions of Focus in Clipped Countries

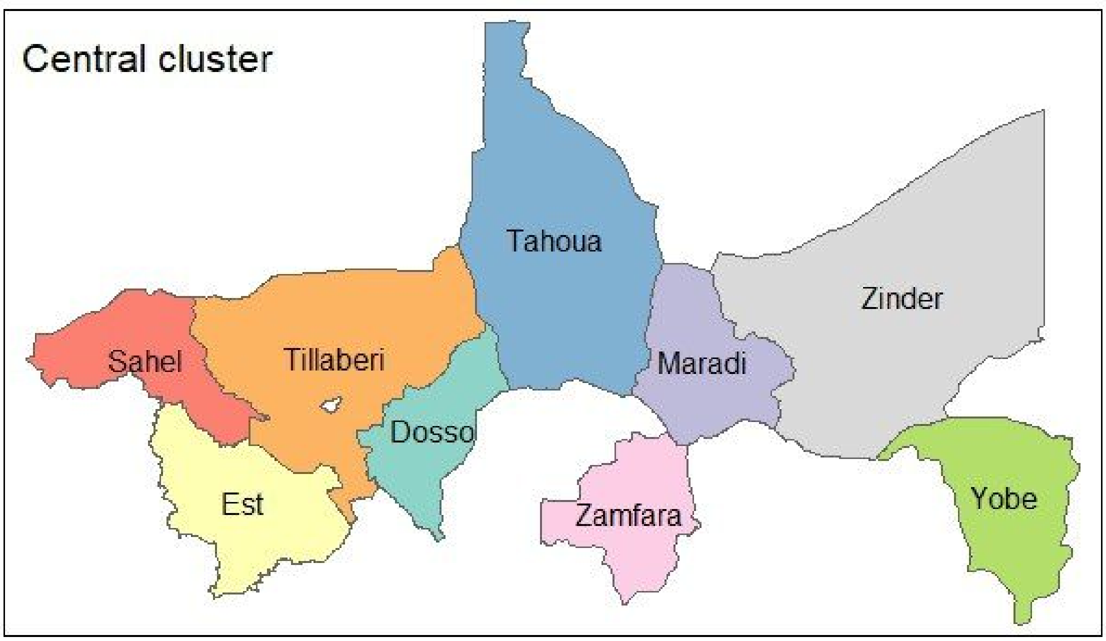


Figure 33B: Central Regions of Focus in Clipped Countries

Identification of Sectors to Allocate Loans

In the above section, evaluation was done on the effectiveness of Kiva's approach for the past 4 years and showed room for improvement as many countries, despite receiving larger amounts of loans, rose in poverty levels over the years. The results spur the analysis into identifying and thereafter suggesting sectors/ industry to allocate loans to for individual countries to ensure targeted distribution of loans to uplift effectiveness. In order to achieve this, understanding poverty is imperative.

<u>Understanding Poverty</u>

According to reliefweb, climate change and access to livelihoods are the main causes of poverty. While there are still many contributing factors, they are often intangible or have no open source data to access. Therefore, our focus will be on the above 2 factors, to uncover their correlation and effects on poverty specifically in Africa. Doing so, targeted approach and better use of the loans can be recommended.

<u>Accessibility</u>

Accessibility here is defined as the elevation of the location. To identify correlation of accessibility on poverty, the required elevation data must first be preprocessed. As elevation can be seen as a continuous data and measurements are only done on specific discrete points (Figure 34), spatial interpolation is required. As such, a raster of elevation is obtained after applying Inverse Distance Weighted (IDW) Method (Figure 35).

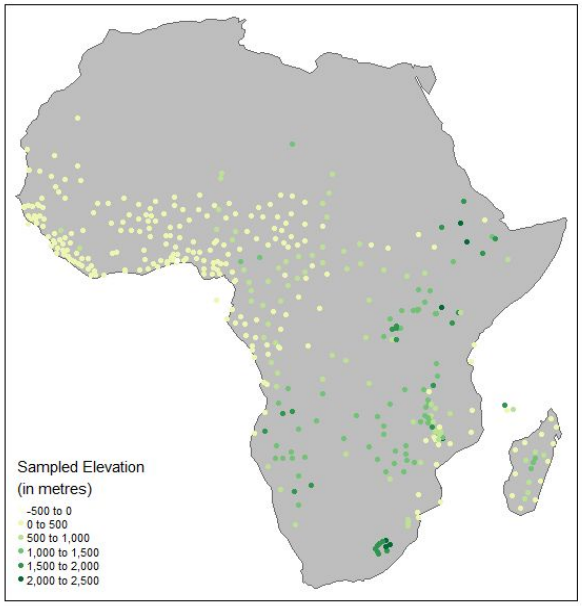


Figure 34: Points where Elevation is Measured

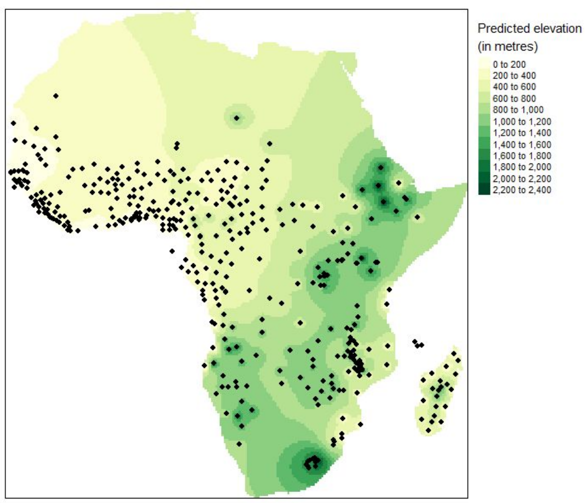


Figure 35: Raster Plot of Elevation After Interpolation using IDW

Trend surface modeling is also used to identify trends in the elevation values and performed in both 1st and 2nd order surface trends (Figure 36). Both orders are slightly coherent and with the raster plot above in highlighting the prominent East-West trend where the elevation tends to increase when moving Eastwards.

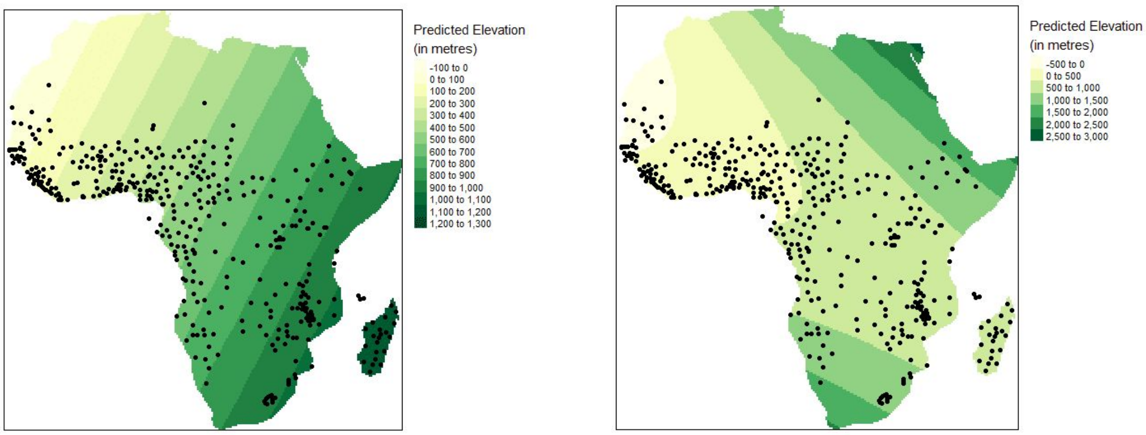


Figure 36: First (left) and Second (right) Order Surface Trends

In order to perform point pattern and quadratic density analysis to identify the number of loan applications following the elevation raster covariate, the plot is discretized and tesselized (Figure 39) into non-uniform quadrats. In order to achieve that, histograms (Figure 37) are first plotted to understand the distribution of predicted elevation before breaking down into 4 regions for tessellation. Log transformed elevation values are used to divide the tessellation quadrats with equal interval classification scheme (Figure 38).

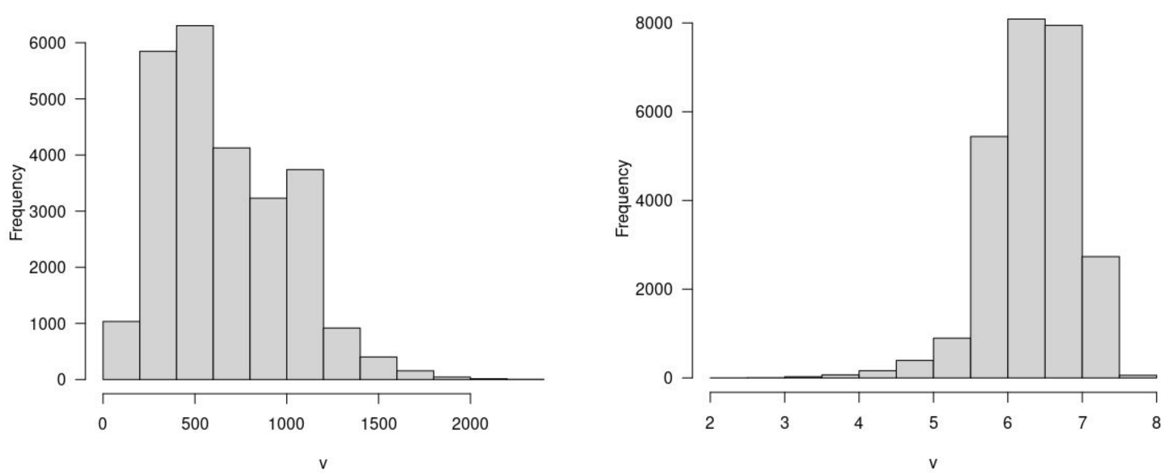


Figure 37: Histograms of Distribution of Elevation (Left) and Logged Elevation (Right)

Break	Logged elevation value
1	[-Inf; 5]
2	[5; 6]
3	[6; 7]
4	[7; Inf]

Figure 38: Table of Values Used for Tessellation

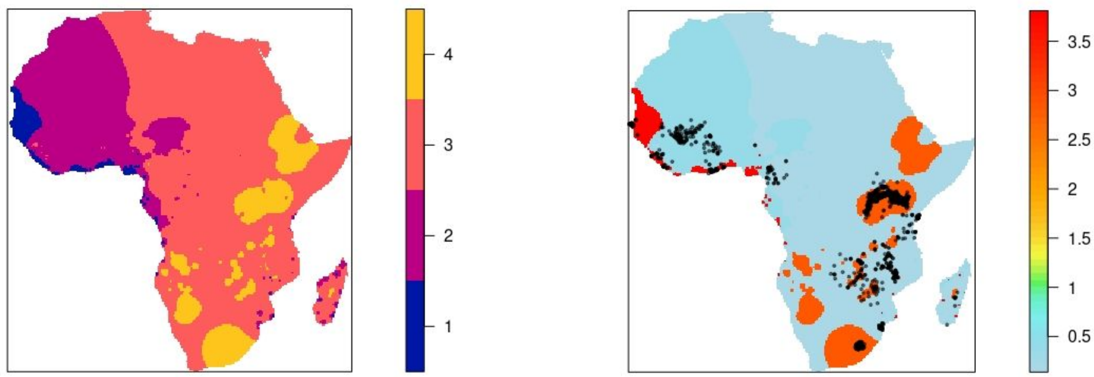


Figure 39: Tessellated Plot for Elevation

Kernel density adjusted for covariate is then plotted to identify density of loan applications points. A non-parametric curve is fit to the data, which describes the shape of the relationship between point density and covariate. As seen from the plot (Figure 40, Left), in the area where the confidence interval is high and uncertainty is lower, there is an exponential increase in intensity of loan applications as logged elevation values increases, implying that elevation is a contributing factor to poverty levels. This is further confirmed by plot (Figure 40, Right), where intensity of loan applications are higher in regions with high elevation.

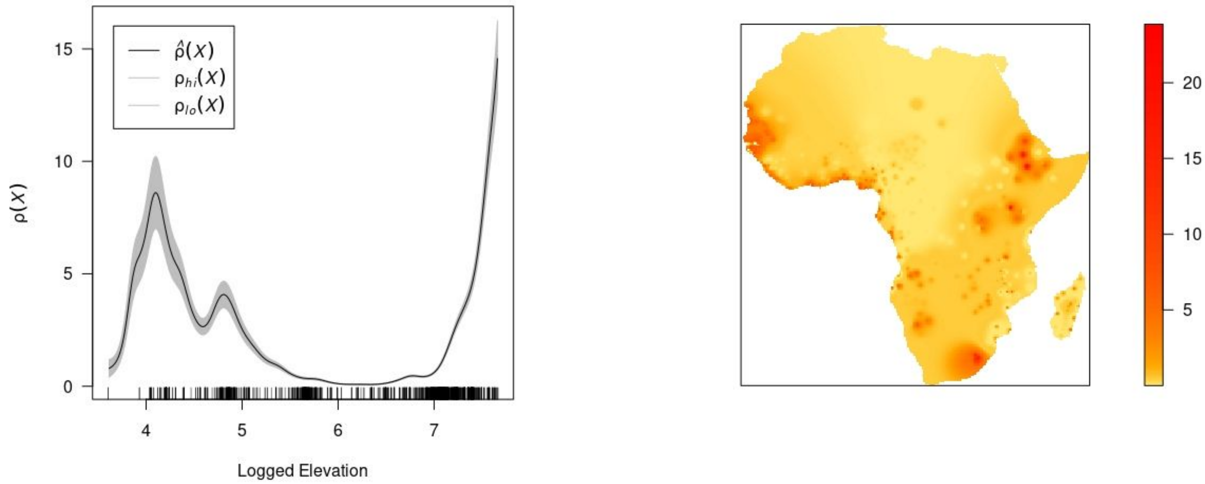


Figure 40: Kernel Density Adjusted for Covariates

The plot (Figure 41) of predicted intensity (above) against observed intensity (K1) also emphasizes this observation, where if the modelled intensity was comparable to observed intensity, clusters forming around a one-to-one diagonal line (red line) will be observed. In this case, it can be concluded that there is a strong relationship between elevation and the number of loan applications.

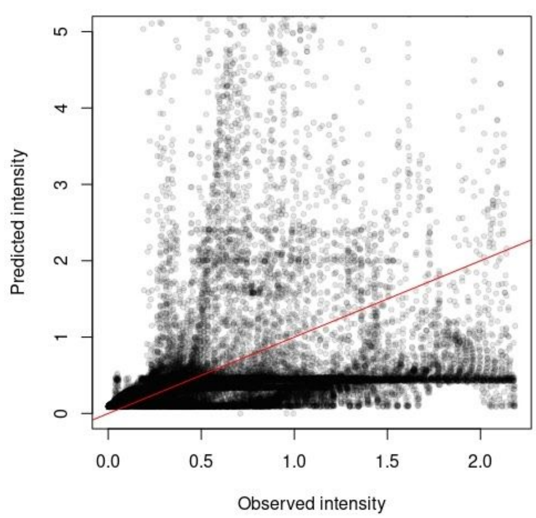


Figure 41: Predicted VS Observed Intensity (K1)

To ensure the validation of the model fit in explaining the observed point pattern, hypothesis testing is conducted with elevation as covariate. From the results in Figure 42C, the p-value is 0.01176, hence there is sufficient evidence to reject the null hypothesis and conclude that the model can explain the observed point pattern.

```
Nonstationary Poisson process

Log intensity:  ~elev.lg.im

Fitted trend coefficients:
(Intercept)  elev.lg.im
 -1.2097577   0.1090025

              Estimate       S.E.     CI95.lo    CI95.hi Ztest      Zval
(Intercept) -1.2097577 0.27993805 -1.75842624 -0.6610892   *** -4.321519
elev.lg.im   0.1090025 0.04370911  0.02333419  0.1946707     *  2.493816
```

Figure 42A: Model Assumes that Point Process' Intensity is a Function of Logged Elevation

```
Stationary Poisson process
Intensity: 0.5955582
              Estimate       S.E.    CI95.lo    CI95.hi Ztest      Zval
log(lambda) -0.5182561 0.02559898 -0.5684292 -0.4680831   *** -20.24518
```

Figure 42B: Model Assumes that Process' Intensity is not a Function of Elevation

```
Analysis of Deviance Table

Model 1: ~1      Poisson
Model 2: ~elev.lg.im     Poisson
  Npar Df Deviance Pr(>Chi)
1    1
2    2  1   6.3463  0.01176 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 42C: Likelihood Ratio Test (Anova)

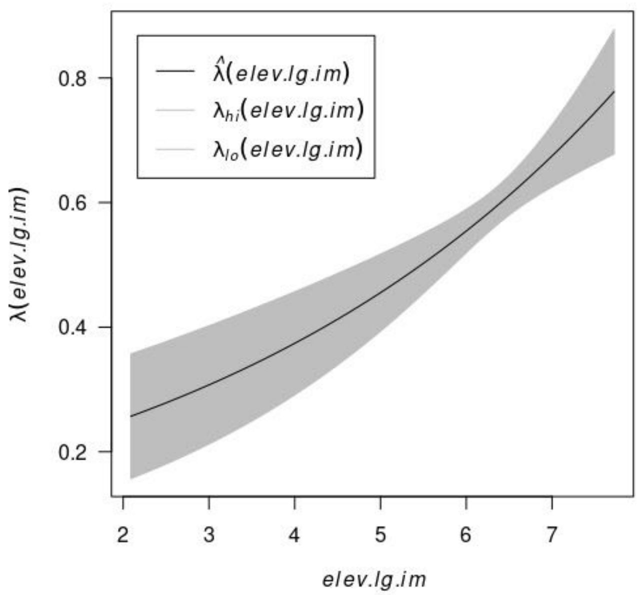


Figure 43: Poisson Point Process Model with Elevation as Covariate

<u>*Climate/ Environment*</u>

Temperature

To identify correlation of temperature on poverty, temperature data must first be preprocessed. As temperature is a continuous data and measurements are only done on specific discrete points (Figure 44), spatial interpolation is required. As such, a raster of temperature is obtained after applying Inverse Distance Weighted Method (Figure 45).

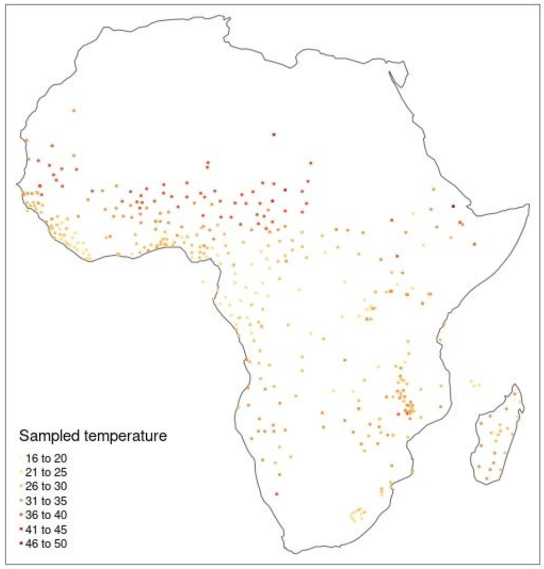


Figure 44: Points where Temperature is Measured

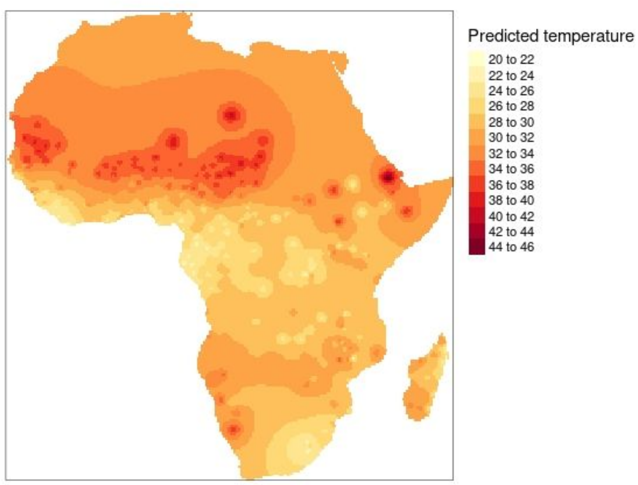


Figure 45: Raster Plot of Temperature After Interpolation

Trend surface modeling is then used to identify trends in the temperature values and performed in both 1st and 2nd order surface trends (Figure 46). Both orders are coherent and with the raster plot above in highlighting the prominent North-South trend where the temperature tends to increase when moving Northwards.

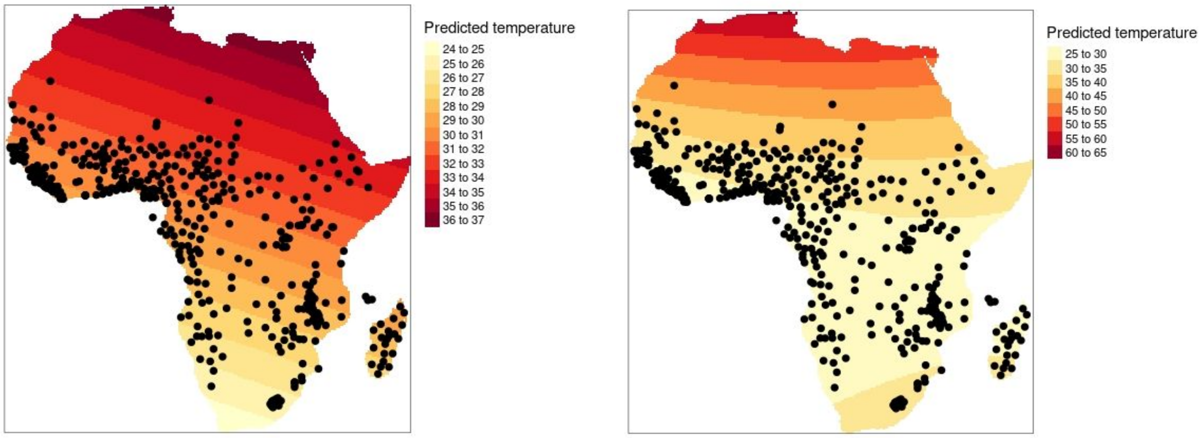


Figure 46: First (left) and Second (right) Order Surface Trends

In order to perform point pattern and quadratic density analysis to identify the number of loan applications following the temperature raster covariate, the plot is discretized and tesselized (Figure 49) into non-uniform quadrats. In order to achieve that, histograms (Figure 47) are first plotted to understand the distribution of predicted temperature before breaking down into 4 regions for tessellation. Log transformed temperature values are used to divide the tessellation quadrats with equal interval classification scheme. (See Figure 48 below)

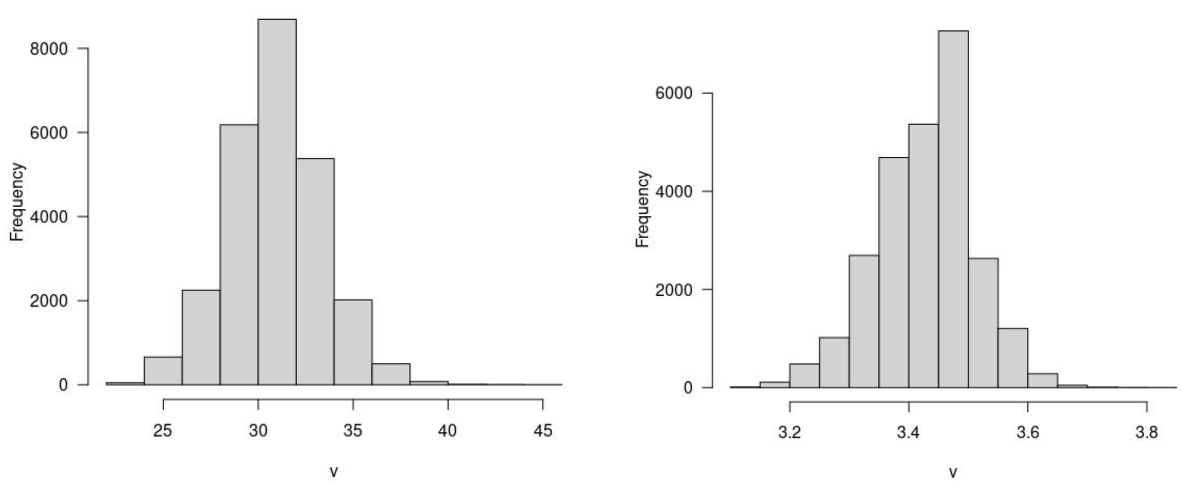


Figure 47: Histograms of Distribution of Temperature (Left) and Logged Temperature (Right)

Break	Logged temperature value
1	[-Inf; 3.2]
2	[3.2; 3.4]
3	[3.4; 3.6]
4	[3.6; Inf]

Figure 48: Table of Values Used for Tessellation

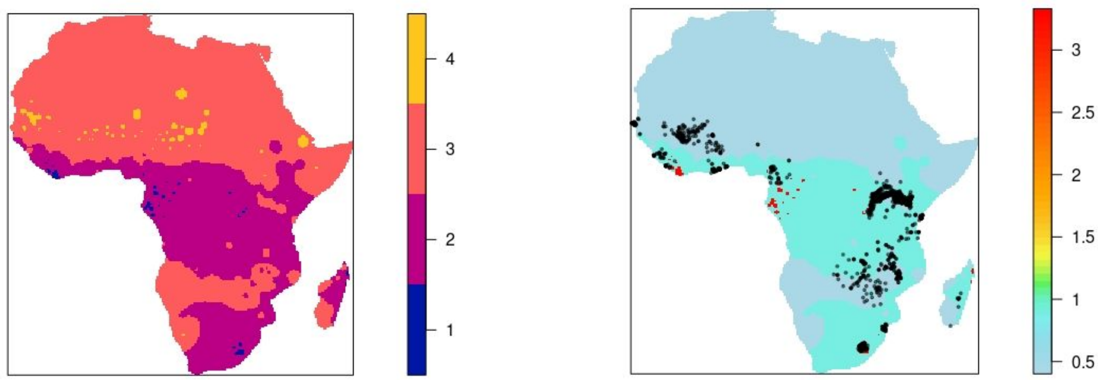


Figure 49: Tessellated Plot for Temperature

Kernel density adjusted for covariate is then plotted to identify density of loan applications points. A non-parametric curve is fit to the data, which describes the shape of the relationship between point density and covariate. As seen from the plot (Figure 50, Left), in the area where the confidence interval is high and uncertainty is higher, there is an exponentially decreasing intensity of loan applications as logged temperature values increase, implying that temperature might be a contributing factor to poverty levels. This is further confirmed by plot (Figure 50, Right), where intensity of loan applications are higher in regions with lower temperatures.

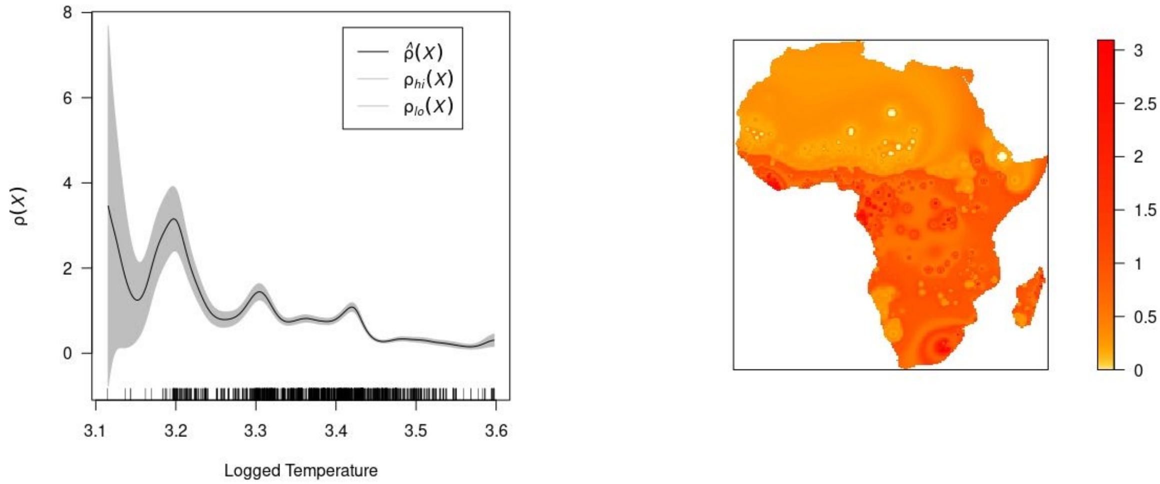


Figure 50: Kernel Density Adjusted for Covariates

The plot (Figure 51) of predicted intensity (above) against observed intensity (K1) also emphasizes this observation, where if the modelled intensity was comparable to observed intensity, clusters forming around a one-to-one diagonal line (red line) will be observed. In this case, it can be concluded that there indeed exists some form of relationship between temperature and the number of loan applications.

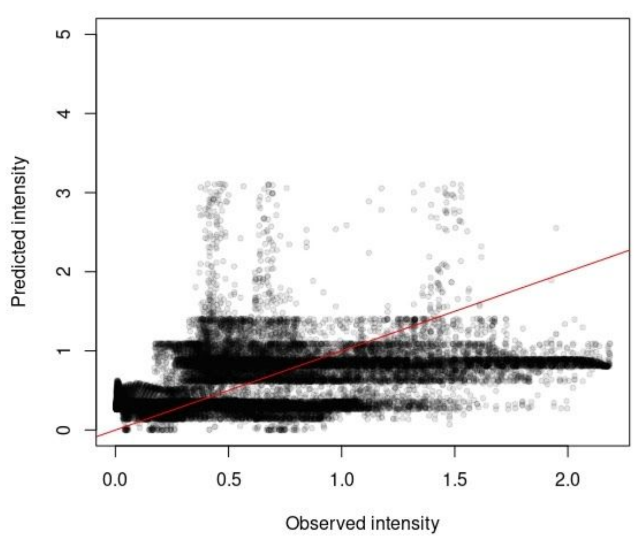


Figure 51: Predicted VS Observed Intensity (K1)

To ensure the validation of the model fit in explaining the observed point pattern, hypothesis testing is conducted with temperature as covariate. From the results in Figure 52C, the p-value is 2.2e-16, hence there is sufficient evidence to reject the null hypothesis and conclude that the model can explain the observed point pattern.

```
Nonstationary Poisson process

Log intensity:  ~temp.lg.im

Fitted trend coefficients:
(Intercept)  temp.lg.im
  19.483291   -5.867553

             Estimate      S.E.   CI95.lo   CI95.hi Ztest      Zval
(Intercept) 19.483291 1.0041246 17.515243 21.451339   ***  19.40326
temp.lg.im  -5.867553 0.2963192 -6.448328 -5.286778   *** -19.80146
```

Figure 42A: Model Assumes that Point Process' Intensity is a Function of Logged Temperature

```
Stationary Poisson process
Intensity: 0.5955582
              Estimate       S.E.    CI95.lo    CI95.hi Ztest      Zval
log(lambda) -0.5182561 0.02559898 -0.5684292 -0.4680831   *** -20.24518
```

Figure 42B: Model Assumes that Process' Intensity is not a Function of Temperature

```
Analysis of Deviance Table

Model 1: ~1      Poisson
Model 2: ~temp.lg.im     Poisson
  Npar Df Deviance  Pr(>Chi)
1    1
2    2  1   376.21 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 52C: Likelihood Ratio Test (Anova)

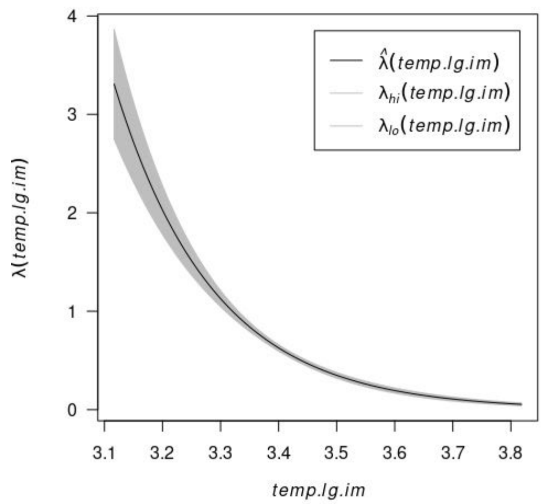


Figure 53: Poisson Point Process Model with Temperature as Covariate

Precipitation Levels

Similarly, to identify correlation between precipitation and poverty, precipitation data must first be preprocessed. As precipitation is a continuous data and measurements are only done on specific discrete points (Figure 54), spatial interpolation is required. As such, a raster of Precipitation is obtained after applying Inverse Distance Weighted Method (Figure 55).

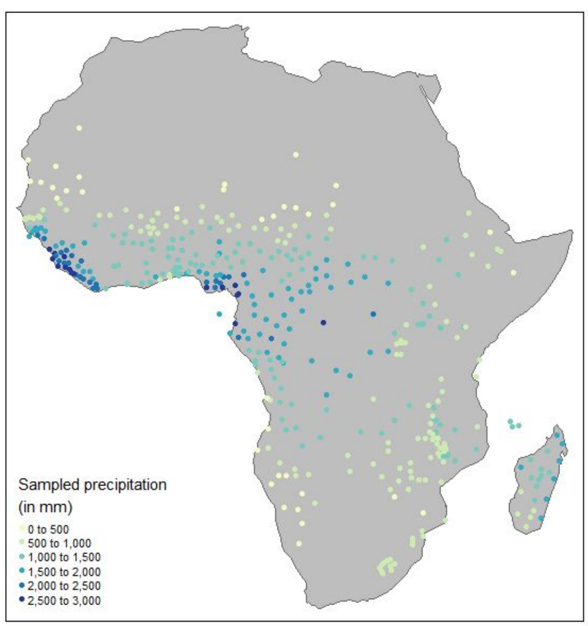


Figure 54: Points where Precipitation is Measured

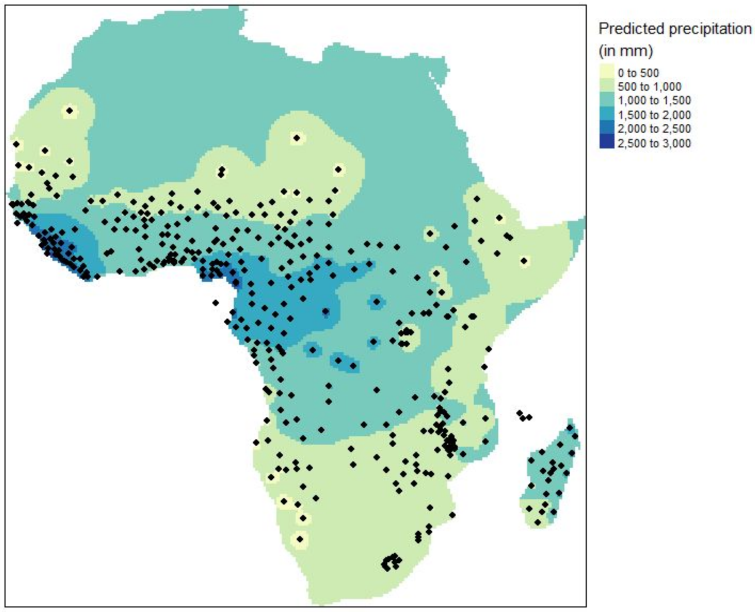


Figure 55: Raster Plot of Precipitation After Interpolation

Trend surface modeling is also used to identify trends in the precipitation values and performed in both 1st and 2nd order surface trends (Figure 56). Both orders are seemingly coherent and with the raster plot above in highlighting the East-West trend where the precipitation tends to increase when moving Westwards.

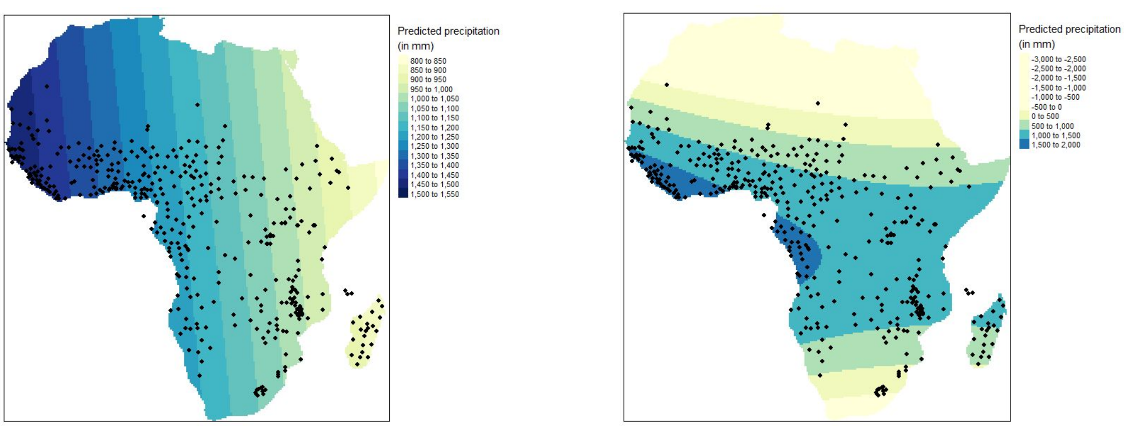


Figure 56: First (left) and Second (right) Order Surface Trends

In order to perform point pattern and quadratic density analysis to identify the number of loan applications following the precipitation raster covariate, the plot is discretized and tesselized

(Figure 59) into non-uniform quadrats. In order to achieve that, histograms (Figure 57) are first plotted to understand the distribution of predicted precipitation before breaking down into 4 regions for tessellation. Log transformed precipitation values are used to divide the tessellation quadrats with equal interval classification scheme (Figure 58).

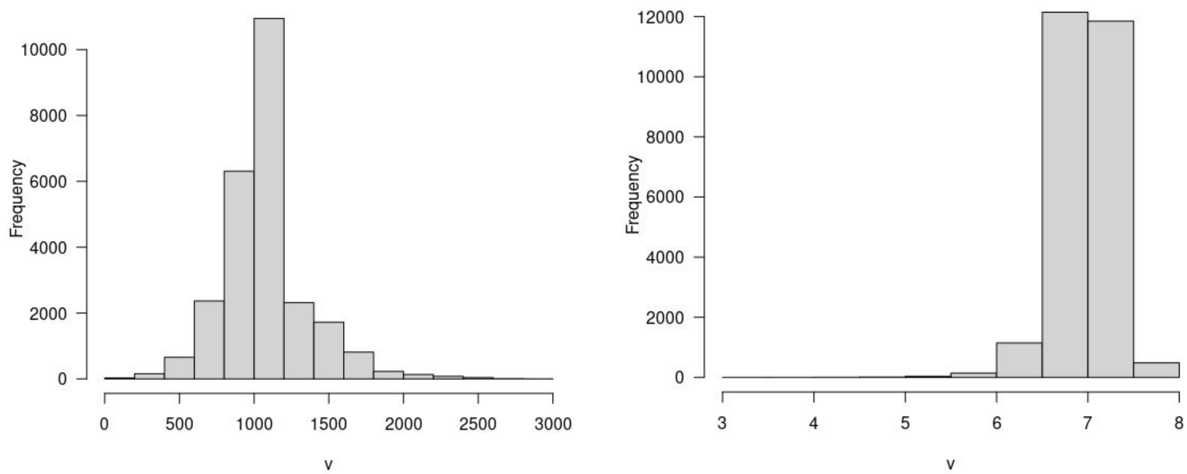


Figure 57: Histograms of Distribution of Precipitation (Left) and Logged Precipitation (Right)

Break	Logged precipitation value
1	[-Inf; 6.5]
2	[6.5; 7]
3	[7; 7.5]
4	[7.5; Inf]

Figure 58: Table of Values Used for Tessellation

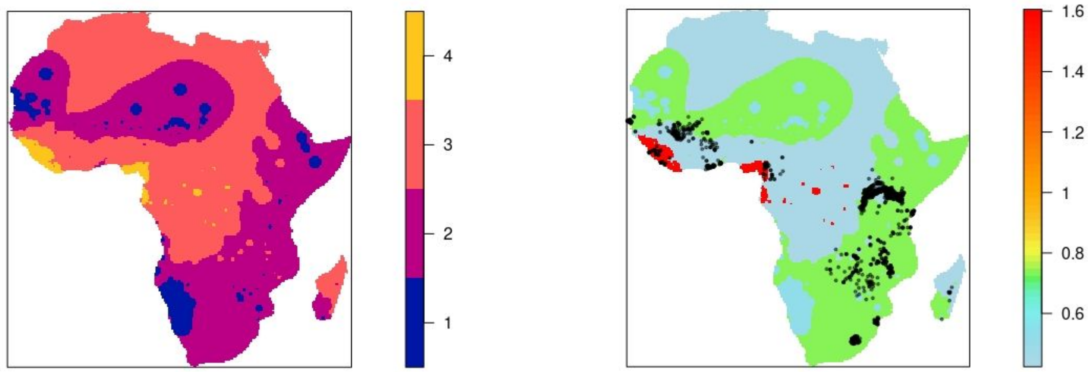


Figure 59: Tessellated Plot for Precipitation

Kernel density adjusted for covariate is then plotted to identify density of loan applications points. A non-parametric curve is fit to the data, which describes the shape of the relationship between point density and covariate. As seen from the plot (Figure 60, Left), in the area where the confidence interval is high and uncertainty is lower, there is an exponentially increasing intensity of loan applications as logged precipitation values increase, implying that precipitation might be a contributing factor to poverty levels. This is further confirmed by plot (Figure 60, right), where intensity of loan applications are higher in regions with higher precipitation.

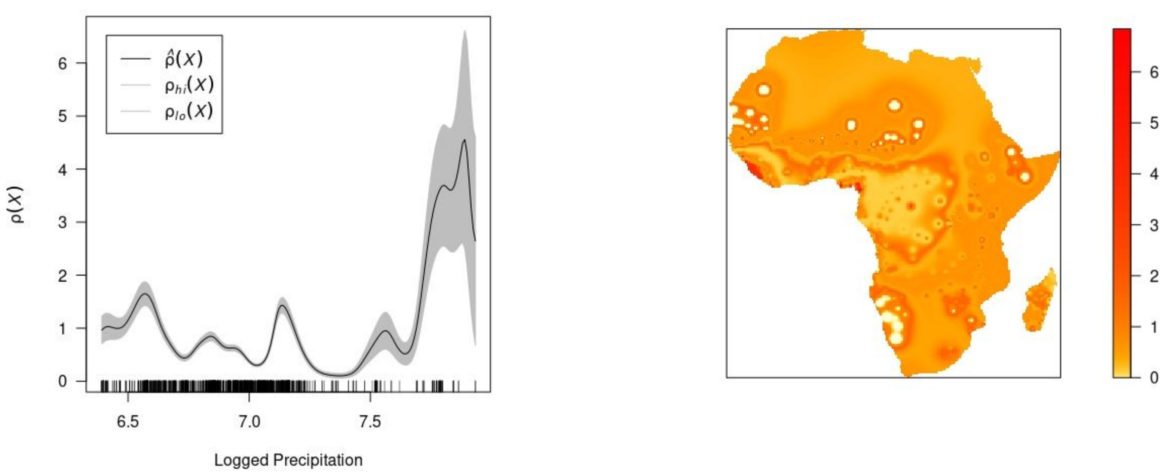


Figure 60: Kernel Density Adjusted for Covariates

The plot (Figure 61) of predicted intensity (above) against observed intensity (K1) also emphasizes this observation, where if the modelled intensity was comparable to observed intensity, clusters forming around a one-to-one diagonal line (red line) will be observed. In this case, it can be concluded that there indeed exists some form of relationship between precipitation and the number of loan applications.

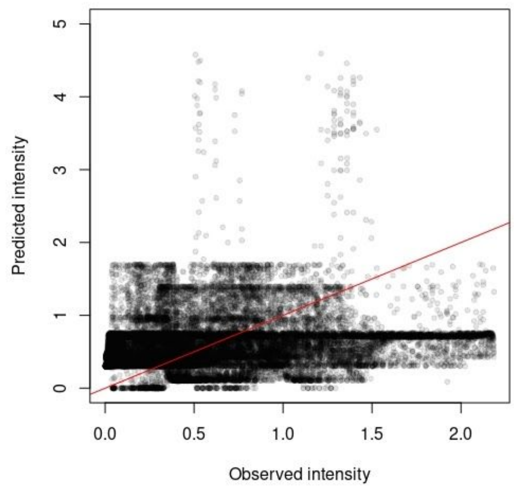


Figure 61: Predicted VS Observed Intensity (K1)

To ensure the validation of the model fit in explaining the observed point pattern, hypothesis testing is conducted with precipitation as covariate. From the results in Figure 62C, the p-value is 0.02908, hence there is sufficient evidence to reject the null hypothesis and conclude that the model can explain the observed point pattern.

```
Nonstationary Poisson process

Log intensity:  ~precip.lg.im

Fitted trend coefficients:
 (Intercept) precip.lg.im
   0.8467291   -0.1964182

               Estimate      S.E.    CI95.lo     CI95.hi Ztest      Zval
(Intercept)   0.8467291 0.6182606 -0.3650394  2.05849756        1.369534
precip.lg.im -0.1964182 0.0890968 -0.3710448 -0.02179171     * -2.204549
```

Figure 62A: Model Assumes that Point Process' Intensity is a Function of Logged Precipitation

```
Stationary Poisson process
Intensity: 0.5955582
              Estimate       S.E.    CI95.lo    CI95.hi Ztest      Zval
log(lambda) -0.5182561 0.02559898 -0.5684292 -0.4680831   *** -20.24518
```

Figure 62B: Model Assumes that Process' Intensity is not a Function of Precipitation

```
Analysis of Deviance Table

Model 1: ~1      Poisson
Model 2: ~precip.lg.im   Poisson
  Npar Df Deviance Pr(>Chi)
1    1
2    2  1   4.7628  0.02908 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 62C: Likelihood Ratio Test (Anova)

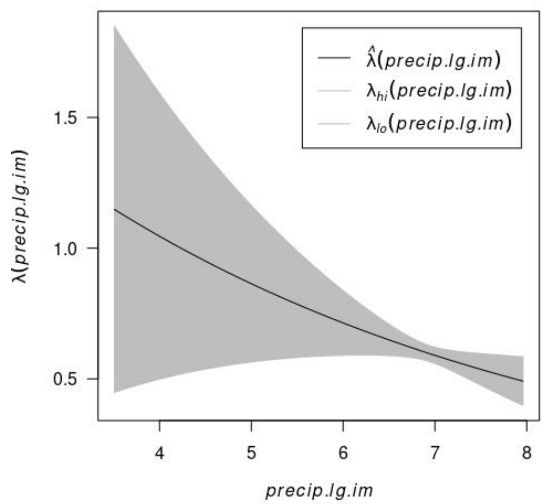


Figure 63: Poisson Point Process Model with Precipitation as Covariate

<u>Proposal on Alternative Estimations of Poverty Level</u>

Considering the importance of MPI in coming up with targeted strategies, it is imperative that MPI of all regions including those less accessible can be obtained. Coupled with the fact that the MPI assesses poverty at both individual and household levels making it a poverty measure difficult to obtain, accurate estimations of these values must be proposed. With the insights generated, including the contributing factors, layering of different rasters of these factors, scaled appropriately can provide a reasonable estimate to measure poverty.

Figure 64 shows the layered raster plot that is generated with the layering of raster plots from all the factors deemed correlated to poverty levels earlier on (namely temperature, precipitation and elevation). For positively correlated factors, addition is used while factors that are negatively correlated with poverty are subtracted. Scaling of the factors are also done to ensure proportionality during spatial operations. The red raster shows pixels with higher poverty levels while green/blue represent lower poverty levels. In this case, the gradient that the raster provides will provide an estimate on poverty levels for unknown countries and regions.

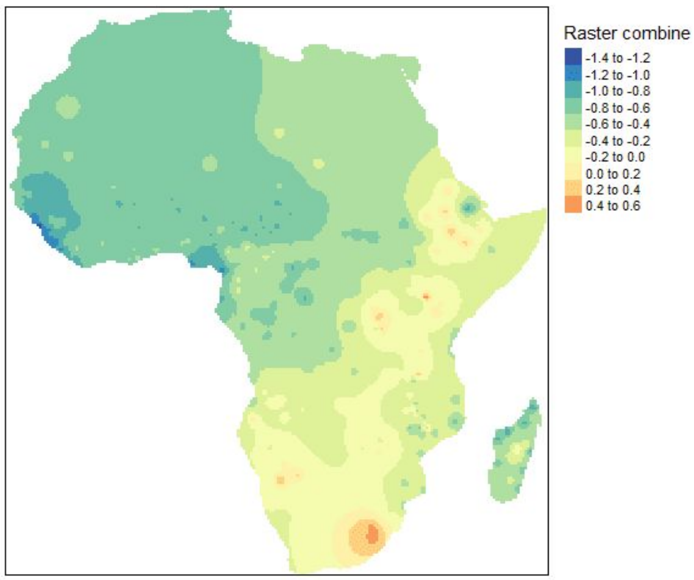


Figure 64: Layered Raster Plot of all Correlated Factors

Discussion

All in all, 3 main insights were discussed in the research:

1) Proposed future area of focus, both in global (countries) and local (regions) context
2) Recommends new approach in distributing loans, specifically looking at infrastructure
3) Proposed alternative measurement of poverty in areas where MPI is absent

Proposed Future Area of Focus

Upon evaluating Kiva's approach and loans distribution patterns, the worrying trend of countries facing an increase in MPI levels coupled with the gradual decrease in Kiva's assistance to these countries is a cause for concern. Kiva's shift in focus has thus motivated the study and analysis of areas to focus on, both in the global and local context. This ensures that Kiva does not unintentionally neglect countries and regions which require more assistance. With respect to significant spatial autocorrelation in MPI among countries and regions, 2 criterion are drawn up:

1) Area of focus must have urgent need for assistance reflected upon high MPI
2) Chosen area should be clusters of countries/ regions such that assistance to a carefully selected area will benefit neighbouring areas to maximize cost-effectiveness of loans

Upon detailed spatial autocorrelation analysis in the sections above, 1 cluster of countries (Figure 26) was chosen as our point of focus, with detailed analysis performed on regional level (Figure 32A, 32B), to identify hotspots in these countries to employ a more targeted injection of funds. In summary, the table of countries and regions to focus on is as follows (Figure 65):

Regions to Focus On

Region	Country	Cluster
Ennedi	Chad	Eastern
Wadi Fira	Chad	Eastern
OuaddaA	Chad	Eastern
Sila	Chad	Eastern
Salamat	Chad	Eastern
Kanem	Chad	Eastern

Barh El Gazal	Chad	Eastern
Batha	Chad	Eastern
Lac	Chad	Eastern
Hadjer-Lamis	Chad	Eastern
Chari Baguirml	Chad	Eastern
Guacra	Chad	Eastern
Dosso	Niger	Central
Maradi	Niger	Central
Tahoua	Niger	Central
Tillaberi	Niger	Central
Zinder	Niger	Central
Yobe	Nigeria	Central
Zamfara	Nigeria	Central
Est	Burkina Faso	Central
Sahel	Burkina Faso	Central

Figure 65: Table of Regions to Focus when Distributing Loans

Recommends New Approach in Distribution

Over the years, while Kiva increased their loans to certain regions, they often do not translate to tangible decrease in MPI levels as proven earlier and this essentially indicates an inefficient allocation of resources. Hence, a reasonable deduction is to question the effectiveness of Kiva's distribution approach and emphasis must be placed on the proposal of sectors to mitigate the losses.

With respect to this, and coupled with secondary research on poverty causes, accessibility (which is defined as elevation here) was chosen as our point of focus. Upon spatial interpolation and subsequent point pattern analysis on the loan applications, our results showed high positive correlation between elevation and poverty levels. This serves as strong evidence for us to propel injections to the infrastructure industry as compared to the past focus on individual loans and agriculture.

Estimation of Poverty Measurement

Considering the importance of MPI in coming up with targeted strategies, it is imperative that the MPI of all regions including those less accessible can be obtained. Coupled with the fact that the MPI assesses poverty at both individual and household levels, it is a poverty measure that may be difficult to obtain. Hence, accurate estimations of the poverty level must be proposed.

This requires Kiva to consider a larger set of factors, other than MPI, to estimate the poverty level. By using insights such as correlations drawn from environmental factors via spatial interpolation and point pattern analysis, a reasonable estimate of poverty levels which includes the use of spatial operations and scaling of poverty inducing factors is proposed.

Future Directions

Some possible future directions would be to do analysis on more environmental data to see if these variables will correlate to poverty to find more trends and insights and improve our strategies to recommend to KIVA, these variables include, Soil type, Modis Leaf Average Index, Modis Enhanced vegetation index, Average Night light time and Evaporation. While the socio-economic factors can also be another source of information to look at, MPI has been a reliable measure for these factors.

Data on the loans provided by KIVA can also be collected year after year to provide a more accurate estimate correlation. The presence of more sources of data will add validity to the proposed actions.