Estimation of Geographical Coordinates from Household Surveys using Machine Leaning: An application to Study Digital Gender Divide in Sierra Leone

Author: Bikalpa Baniya

Most data used in international development and in comparative education has some spatial aspect: children's distance to school, a community's access to the road network or the deployment of teachers in remote ares. These are also present at all levels from the largest country to the smallest villages. Incorporating these spatial components can be crucial for many important analysis conducted in international education and development. However, data without precise geolocations are much more common than those with a detailed level of information. Even statistics currently being disseminated are aggregated to a level that severely restrict information required for any spatial analysis.

For example, one could examine gender patterns in accessing digital tools like radio or internet in a country to better understand the opportunities for distance education during school closures due to COVID-19. Although "access to media" is collected in many household surveys, is often only available at a rather large administrative level. This information is therefore hardly compatible with very fine information of cellular towers, radio wave transmissions, or speed of internet data.

Machine Learning can become part of the solution. More information is being collected than ever before, as evidence is now central to every aspects of international development. With this increased profusion of data, many artificial intelligence algorithms have become viable tools that produce reliable predictions. This paper explores two of these tools - Random Forest and Affinity Propagation – to estimate geolocation of household survey data. Random Forest is used to

This paper briefly explains of both of the methods, the reasons for their selection and the context in which they should be used. Then it will demonstrate the ability of machine learning to enhance the analysis of important questions in the field. More specifically, it will investigate digital gender divide in Sierra Leone using UNICEF 2017 MICS data which has geolocations available only at the district level rather than specific geographic coordinates. We will use the Sierra Leone national 2004 language census, Open Street Maps and WorldPop data to estimate a finer geo-localisation of the households so that we can better understand the gendered differences in digital technologies used.

Machine Learning Algorithms

Machine Learning Algorithms usually use training datasets to make prediction on actual data. Depending on the type of prediction needed from the training dataset there are two types of machine learning algorithms: supervised and unsupervised. Supervised learning methods are ones that utilize classification already present in the training dataset. The goal is to train the algorithm with these classifications well enough to predict the classification of new data. For example, if we want to predict the geographical district of in a given dataset, we can utilize training data that has this information. Supervised learning model learns the properties of the district variable in the training data in relation with other variables and makes its prediction in the given dataset. Random forests, OLS regression Naïve Bayes are the more widely used of such tools.

Unsupervised learning methods are algorithms that creates its own classification in the training data. The goals here is to discover the underlining structure of data for some sort of clustering or association. For example, such algorithm can be used if the training datasets does not have district information, but we still need to predict the groups by clustering similar entries. Popular unsupervised learning algorithms are K-means, Deep Learning, Mean Shift and Affinity Propagation.

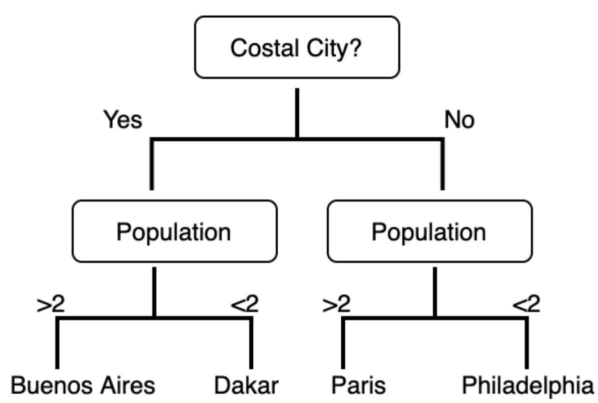
This paper uses both types of methods to estimate geographical coordinates. Random Forest is first used to predict chiefdom (the administrative level below district) because the training data has this information. In the second step, the paper predicts classification below the chiefdom level using Affinity Propagation in conjunction with Random Forest because further categorization is not present in the training data. Specifically, Affinity Propagation is used to create artificial categorization and Random Forest is used to match them to the survey data.

Random Forest

Random Forests are supervised machine learning algorithm developed by Breiman (2001). This algorithm is well-known for its ease use and its accuracy. The "forest" first uses training data to build on the predictions of individual decision trees, which are tree like models that classifies data into "branches" using the given parameters. A visual demonstration of decision trees is given below to show how the decision trees can be used to classify four cities, with two variables. This is a simple example but similar tree models can be used to classify data from larger datasets on a greater number of variables. Categorical variables are classified by type, whereas continuous variables are classified by a cut-off point (which, in the example below, is 2 millions). The variables used for decision points in a tree are ordered in such way that the final classification contain optimally distinct data. For example, if there were more cities in the example below, classifying first using population and then Costal City might produce better classification results faster.. After training is completed, the best order of variable is preserved to predict classification for non-training data.

Figure 1: Demonstration of Decision Trees with Hypothetical Data

City	Costal City?	Population in millions
Paris	No	2.16
Dakar	Yes	1.05
Buenos Aires	Yes	3.05
Philadelphia	No	1.5



Source: Author's own

Individual decision trees are flexible and easy to train. However, they have low predictive accuracy for new data. Random Forest overcomes this deficit by creating a large number of trees, for which predicted results are compared with one another. Each tree is provided with a random subset of variables and training data. They make their classification, and every tree preserves their optimum order of variable. When new input data is supplied, each tree uses their order to classify and predict output value. Finally, the Random Forest pools the results from these trees and makes the final estimations to predict a result.

Another advantage of Random forest is the opportunity to calculate the accuracy of the prediction. The full dataset is separated in two: a first part, representing about 70% of the data is the 'bootstrap sample' and is used for training the model, the other part (~30%) is the 'out of bag' sample and is used for checking the accuracy of the predictions Whenever the forest randomly assigns data to individual trees it uses only the bootstrap sample. The remaining part of the training data is then utilized to measure the accuracy of the prediction using *out of bag* error. By comparing the prediction that the random forest can make on the out-of-bag data with the real value of the real value of the predicted variable, we can calculate the proportion of classification errors.

Affinity Propagation

Affinity Propagation (AP) is an unsupervised machine learning algorithm developed by Frey and Dueck (2007). It is a graph-based approach like the popular K-means clustering algorithm. However, unlike many other algorithms, users do not have to specify the number of clusters beforehand. AP instead considers the similarity among data entry and find a representative example, or exemplar, that can represent an entire cluster. In other words, every entry "votes" on another entry that best represents them, and suitable entries are considered to be the exemplar. Data point with the same exemplar are then assigned into the same cluster. This automatic calculation of the number of cluster presents great value, as described in the methodology.

Algorithmically, all AP identify clusters using four subsequent matrices: Similarity matrix, Responsibility matrix, Availability matrix and Criterion matrix. A demonstration is given below using a hypothetical data on a specific population that is dispersed over five districts. The objective of the AP in this case would be to identify which districts might be grouped into the same administrative division (assuming district in the same administrative division are more similar than the ones in the other administrative division).

	Population (in percent)				
	With high school diploma	Access to internet	Aged 15-20 years	Speaking language 1	Speaking language 2
District 1	30	40	30	20	10
District 2	40	30	50	10	10
District 3	30	50	30	30	30
District 4	20	10	30	30	20
District 5	10	10	30	20	30

1) Similarity Matrix

The first step to identify clusters using AP is the similarity matrix. This matrix is the negative square distance of each value on a multidimensional plane and signifies the similarity between two data points.

For example, we have five variables above so on a five-dimensional plane the distance between *District 1* and *District 2* is going to be the following. Values in the diagonal is zero because they compare the same districts.

$$Square\ of\ distance = (30 - 40)^2 + (40 - 30)^2 + (30 - 50)^2 + (20 - 10)^2 + (10 - 10)^2 = 700$$

$$Negative\ Square\ of\ distance = -700$$

Table 1.1: Similarity Matrix

	District 1	District 2	District 3	District 4	District 5
District 1	0	-700	-600	-1200	-1700
District 2	-700	0	-1700	-1700	-2200
District 3	-600	-1700	0	-1800	-2100
District 4	-1200	-1700	-1800	0	-300
District 5	-1700	-2200	-2100	-300	0

According to Frey and Dueck (2007), the size of the number in the diagonal proportionally controls the number of classes the algorithm is able to produce. They recommend setting it to the smallest number in the matrix, which in this case is -2200. Adjusting for this, we finally have the similarity matrix below.

Table 1.2: Similarity Matrix

	District 1	**District 2**	**District 3**	**District 4**	**District 5**
District 1	-2200	-700	-600	-1200	-1700
District 2	-700	-2200	-1700	-1700	-2200
District 3	-600	-1700	-2200	-1800	-2100
District 4	-1200	-1700	-1800	-2200	-300
District 5	-1700	-2200	-2100	-300	-2200

2) Responsibility matrix

The second step is the calculation of the responsibility matrix, which signals how well-suited any data point x is to be an exemplar of data point *y,* relative to other candidate exemplars. It is calculated iteratively (where a different point are used in each iteration) using the following formula.

$$r(x, y) \leftarrow s(x, y) - \{a(x, y') + s(x, y')\}$$

where *r(x,y)* is the responsibility value and *s(x,y)* is the value from the similarity matrix. *a(x,y)* is the "availability" defined below and, for the first iteration, it is set to zero.

Table 2: Affinity Propagation Responsibility Matrix

	District 1	**District 2**	**District 3**	**District 4**	**District 5**
District 1	-1600	-100	100	-600	-1100
District 2	1000	-1500	-1000	-1000	-1500
District 3	1100	-1100	-1600	-1200	-1500
District 4	-900	-1400	-1500	-1900	900

District 5	-1400	-1900	-1800	1400	-1900

3) Availability matrix

to the third step is the availability matrix. It calculates how "available" data point x is for data point y to serve as its exemplar, after accounting for other point's preference for x to be an exemplar. It is calculated iteratively using the following formula.

$$a(x, y) \leftarrow \{(\ 0, \ r(y, y) + \sum\left(0, r\left(x^{'}, y\right)\right) \ ) \quad for\ x' \neq y \ \ \sum\left(0, r\left(x^{'}, y\right)\right) \quad for\ x' = y$$

Table 3: Affinity Propagation Availability Matrix

	District 1	**District 2**	**District 3**	**District 4**	**District 5**
District 1	2100	-1500	-1600	-500	-1000
District 2	-500	0	-1500	-500	-1000
District 3	-600	-1500	100	-1500	-1000
District 4	0	-1500	-1500	1400	-1900
District 5	0	-1500	-1500	-1900	900

4) Criterion Matrix

Finally, the Criterion matrix is the sum of the Responsibility matrix and the Availability matrix. The row with the highest criterion value is designated as the exemplar and rows that share the same exemplar are allocated into the same cluster. From Table 4 below, we can see that the highest value on the rows for Districts 1, 2, and 3 are 500, so this means that these three districts are assigned into one cluster, while Districts 4 and 5 have a common highest value of -500, which means that they are composing all together another cluster.

Table 4: Affinity Propagation Criterion Matrix

	District 1	**District 2**	**District 3**	**District 4**	**District 5**
District 1	**500**	-1600	-1500	-1100	-2100
District 2	**500**	-1500	-2500	-1500	-2500
District 3	**500**	-2600	-1500	-1700	-2500
District 4	-900	-2900	-3000	**-500**	-1000
District 5	-1400	-3400	-3300	**-500**	-1000

So in conclusion, the Affinity Propagation method, once applied to the above matrix reveals that the five districts presented can be grouped in two clusters.

Data

This paper uses UNICEF 2017 Multiple Indicator Cluster Survey Data to study the digital gender divide in Sierra Leone. This data was selected because it contains multiple variables relevant to the literature on digital divide and gender divide like access to internet and age of marriage. Furthermore, the country has three levels of administrative divisions: province (including one area), district and chiefdom.

There are currently 4 Provinces, one area, 16 districts and 186 chiefdoms. Similar to many household surveys, the geospatial information is aggregated at a district level. So, performing detailed geospatial analysis would be difficult as the variables of interest are very sensitive to location.

Nonetheless, we can take advantage of the structure of the data to estimate the geolocation. Specifically, we can use cluster number, household number, age, native language and district information. Considering these variables, this paper constructed a training set with similar variables from World Pop, Open Street Maps and the country's 2004 language census performed by UNICEF. These datasets were used to create the training dataset because they either have precise coordinates or administrative division below the lowest level in the household survey (i.e., at a chiefdom level). These external data were combined to form the training dataset. However,

Table 5.1: Relevant Variables for geographical Coordinate Estimation in Household Survey (25,289 Observations, 53 Variables)

Cluster number	**Househol d number**	**Age**	**Native Language**	**District**	**Area**	**…**	**Gender Dummy**
1	1	17	Krio	Bombali	Urban		1
1	2	36	Mende	Bombali	Rural		1
1	2	37	Mende	Bombali	Rural		0
1	3	45	Mende	Bombali	Rural		1
…	…	…	…	…	…	…	…
600	1	31	Temne	Moyamba	Urban		1
600	1	29	Temne	Moyamba	Urban		0
600	2	21	Kono	Moyamba	Rural		0

Table 5.2: Compiled Training Dataset to create the estimates (9,976 observations, 55 variable)

District Name (*16 District*)	**Chiefdom name *(149 Chiefdom)*[1]**	**X coordinate**	**Y coordinate**	**Population (Proportion)**					
				Kiro speakers	Mende speakers	**…**	**Female Age 15-20**	**Male Age 15- 20**	**…**
Bombali	Biriwa	-11.9363	9.05948	0.29	0.82	…	0.122	0.101	…
Bombali	Biriwa	-11.8576	9.33099	0.28	0.86	…	0.123	0.101	…
Bombali	Biriwa	-11.8576	9.33099	0.31	0.35	…	0.123	0.122	…
Bombali	Tambakha	-12.2601	9.83141	0.1	0.05	…	0.134	0.155	…
…	…	…	…	…	…	…	…	…	…
Moyamba	Ribbi	-12.8178	8.19570	0	0.27	…	0.101	0.113	…
Moyamba	Ribbi	-12.7869	8.20922	0	0.27	…	0.101	0.113	…
Moyamba	Dasse	-12.7760	8.21858	0	0.23	…	0.148	0.131	…
Source:									
2004 Language Census		Open Street Maps		2004 Language Census			World Pop		

Estimating geographical coordinates for Sierra Leone's MICS households

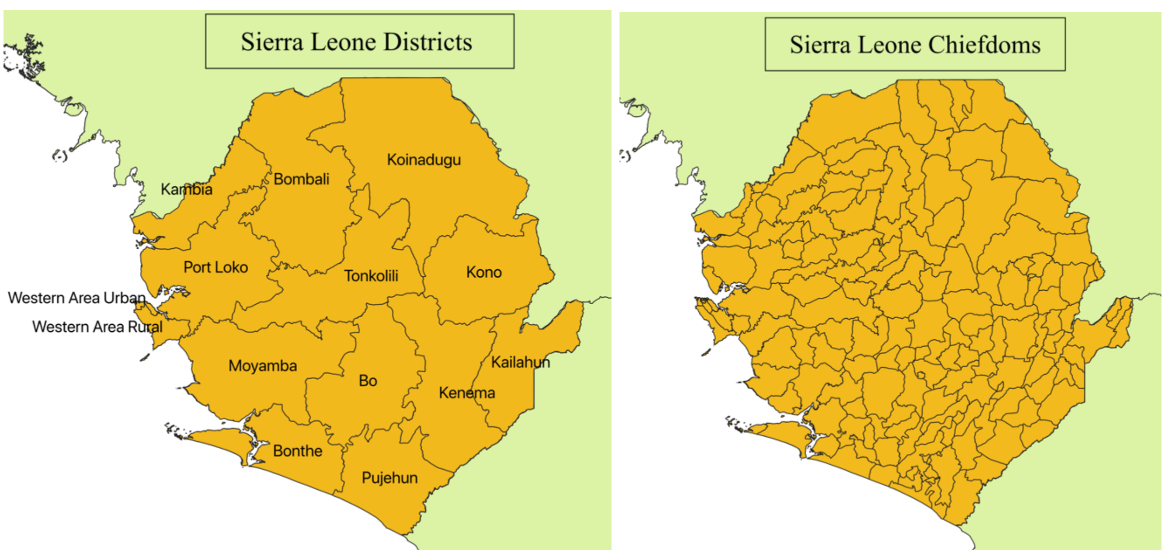
[1] Since this is 2004 data, the training dataset only has 149 Chiefdom. So, this classification will be used to create the estimates not the current classification with 186 chiefdom.

There are four main steps to estimating the geographical coordinates (both chiefdom information and the specific coordinates) for the survey data using the training dataset. Random Forest will be used to predict the chiefdom information and a combination of Affinity Propagation and Random Forest will be used to predict the specific coordinates.

First, all of the variables in the survey data are condensed using their cluster number so that entries with the same cluster number are represent in one entry. This is done because for any given survey being in the same cluster signifies some geographical proximity. The aggregation is also done ensuring that the survey data precisely matches the training data, which is an amalgamation of World Pop, Open Street Maps and UNICEF 2004 Sierra Leone national language census. Eventually, every cluster will be a chiefdome and geo-coordiantes.

The second step is to estimate the chiefdoms using a Random Forest. This administrative division is a level below district, the current administrative division of the household survey. The algorithm learns the characteristic of the classification of the chiefdom from the training data and classifies the clusters in the survey accordingly. The process is also outlined in figure 3 and the out of bag error for each district is listed in table 6.

Figure 2: Sierra Leone districts vs Sierra Leone chiefdoms



Source: Open Street Maps Shapefiles

Next, the paper estimates divisions below the chiefdom level using a combination of Affinity Clustering and Random Forest. The training dataset does not have classifications below the chiefdom level. However, we can utilize the structure of the data to create clusters (groups with similar characteristics) that act as the required classification. More specifically, Affinity Propagation is used because, unlike most an unsupervised learning method, the number of clusters do not need to be predefined. (This serves an important advantage because predefining the number for every chiefdom will be inefficient).

Now similar entries are assigned into the same cluster within each chiefdom (notified as ML cluster in figure 4). This assignment is used to train another Random Forest model and match household survey clusters to these new Machine Learning (ML) clusters. Note that each entry in the ML clusters also has geographical coordinates and every household survey cluster in now matched with a ML cluster. So, for the final step, the paper randomly picks an entry from the ML cluster to assign to the survey clusters that were matched by the Random Forest Model. After this final step, the survey data now has estimated chiefdom and specific geo-coordinates.

Figure 3: Methodology to estimate geographical coordinates for the household survey entries

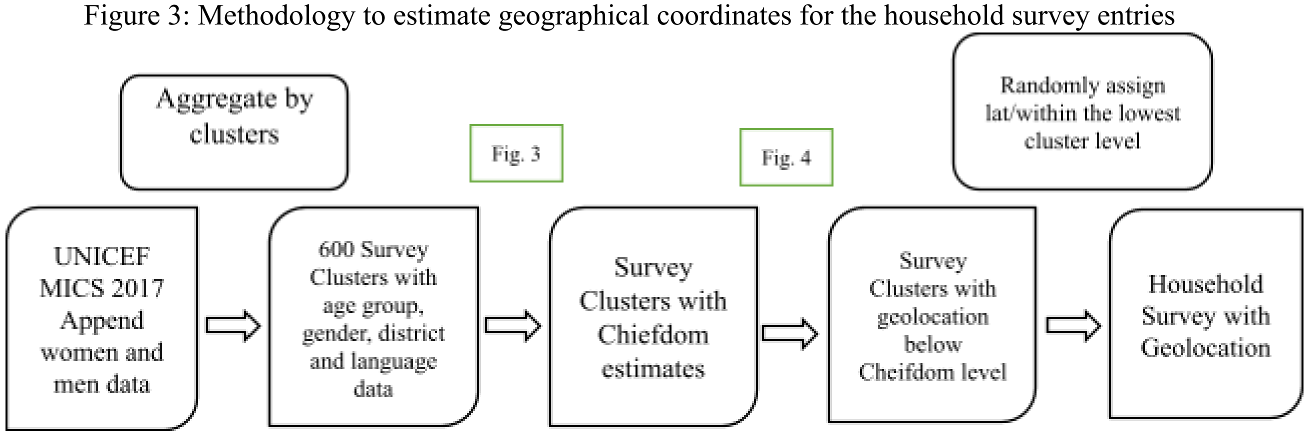


Figure 4: Creating Training Dataset and Using Random Forest to estimate Chiefdom

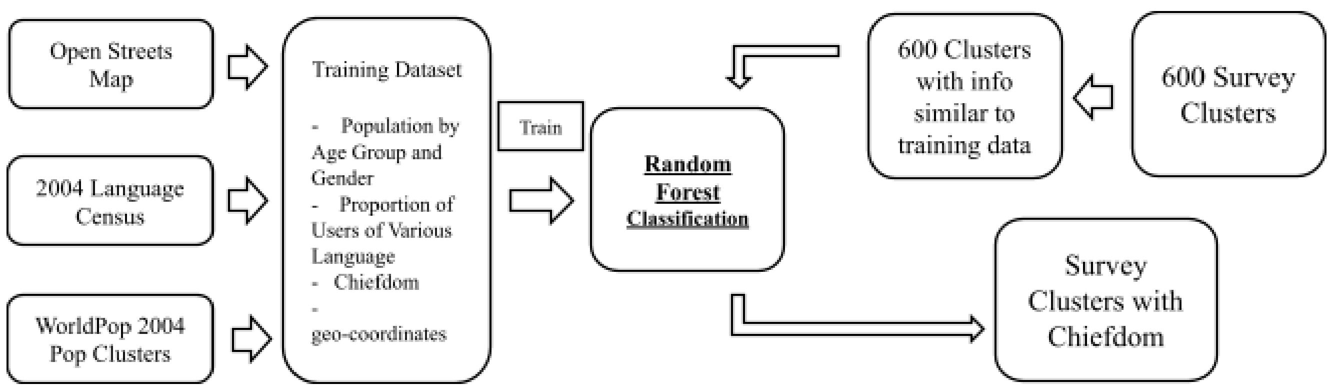


Figure 5: Using Affinity Proposition and Random Forest to produce cluster within chiefdom

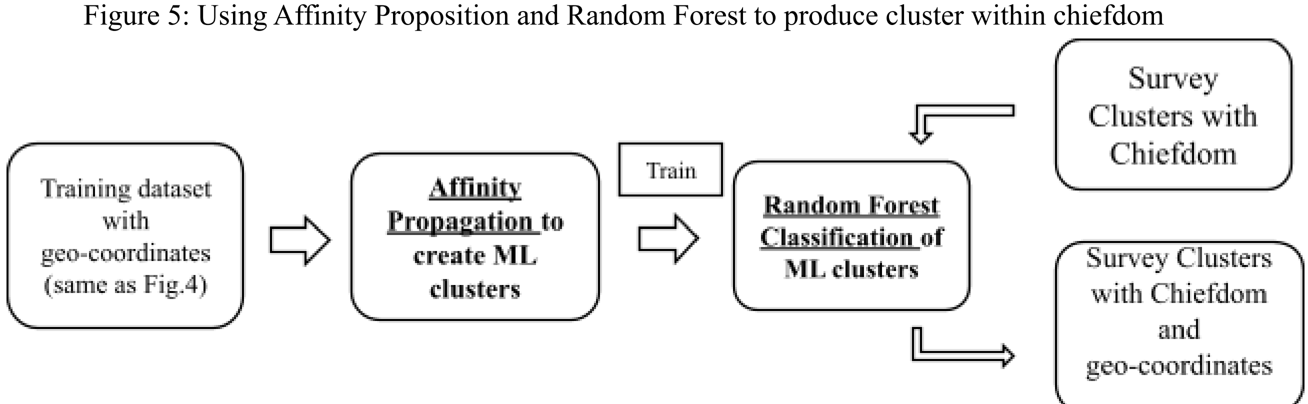


Table 6: Random Forest Out-of-Bag error for each district

District	Number of Observation in training dataset	Out-of-bag error	District	Number of Observation in training dataset	Out-of-bag error
Kailahun	782	19.18%	Kenema	918	22.55%
Kono	756	37.17%	Bombali	742	23.99%
Kambia	459	8.71%	Koinadugu	726	12.81%
Port Loko	886	6.66%	Tonkolili	1033	14.81%
Bo	1049	14.59%	Bonthe	534	41.76%
Moyamba	1048	19.66%	Pujehun	519	21.39%
Western Area Rural	381	6.3%	Western Area Urban	134	1.49%

The household survey now has estimated geographical coordinates for each of its entries, similar to the table 7 below. This new estimation adds a spatial variation to the data which can be used to enrich the analysis of questions like the Digital Gender Divide in Sierra Leone.

Table 7: Estimation of Chiefdom and Geographical Coordinates in Household Survey (25,289 Observations, 53 + 3 Variables)

Cluster number	**House-hold number**	**Age**	**Native Language**	**District**	**…**	**Gender Dummy**	**Estimated**		
							Chiefdom	**xcoord**	**ycoord**
1	1	17	Krio	Bombali	…	1	Biriwa	-11.936	9.0594
1	2	36	Mende	Bombali	…	1	Biriwa	-11.857	9.3309
1	2	37	Mende	Bombali	…	0	Biriwa	-11.857	9.3309
2	1	45	Mende	Bombali	…	1	Tambakha	-12.260	9.8314
…	…	…	…	…	…	…	…	…	…
600	1	31	Temne	Moyamba	…	1	Ribbi	-12.817	8.1957
600	1	29	Temne	Moyamba	…	0	Ribbi	-12.786	8.2092
600	2	21	Kono	Moyamba	…	0	Dasse	-12.776	8.2185

Application to Digital Gender Divide in Sierra Leone

UNESCO defines digital literacy as "the ability to access, manage, understand, integrate, communicate, evaluate and create information safely and appropriately through digital technologies for employment, decent jobs and entrepreneurship" (UNESCO UIS 2018). According to a cross-national study, women are 25% less likely than men to leverage ICT skills for basic purposes ("Accountability …", n.d.). Here ICT includes not only the internet but also radio and cell phones. Another study conducted across 25 countries found that adolescent boys, in general, use mobile phones for a wider range of activities -from playing games to accessing financial service- than adolescent girls, who mostly utilize basic functions such as phone calls (Girl Effect & Vodafone Foundation, 2018). Given the evolving importance of ICT's role in developing socio-emotional skills and their prevalence in the labour market, bridging this digital gender divide is essential to meet the goals for gender equity, especially now that we know how useful alternative modes of teaching and learning can be in case of national school closures.

Issues of such disparity are now very pertinent for education in countries like Sierra Leone. Afrobarometer is a research network that measures public attitudes in Africa. They find that the proportion of women that use ICT in Sierra Leone, especially internet, has increased two-folds since from 2014 to 2019. However, the gender divide in regular internet usage has increased by 3% (Afrobarometer, 2019). This gap is expected to be further exacerbated during the COVID crisis as girls may disproportionately drop out of school and have less access to learning through ICT like the radio and internet (Ledoux et. al., 2020). Such trends in Sierra Leone are evident from the 2014-2016 Ebola epidemic, which the country saw an increase in gender disparity (Risso-Gill et.al., n.d.). As leaning moved outside of the classroom due to national school closures, it is essential to analyze how different digital medium can support students in ensuring continuous learning. Radio and mobile phones should especially be focused because, according to the survey used by this paper, 94% of women and 86% of men aged 15-49 have never used Internet.

Sierra Leone has shown great promise when tackled learning loss during the 2020 school closures through the innovative the use of radio (Mutseyekwa, 2020). When students could not go to school lessons were deliver using this ICT. Such steps were also taken during the previous closures during the Ebola outbreak and produced impressive results: coverage over 81.6% of vulnerable households with high listenership among school children of secondary age (Secretariat, 2020). However, little has been understood about the effects across gender. Such analysis is especially important to ensure equitable access because in 2018 the country was ranked 153rd out of 162 on the Gender Inequality Index ("Human Development Reports", n.d.).

This paper aims to fill the knowledge gap by investigating digital access by gender. The paper will specially look at senior secondary school aged and tertiary aged students because the Gender Parity Index in enrolment starts becoming unfavourable of girls starting from these levels (Government of Sierra Leone, n.d.). In addition, according other 2018 Integrated Household survey (SLIHS), more girls are enrolled at the primary and junior secondary level. Yet, the trend reverses at the upper secondary school level ("Sierra Leone..", n.d.). The widening of digital skills gaps is also noticeable at these level in many other regions around the world ("I'd Blush If I Could", 2020). Hence understanding this age group is very important.

The analysis will be performed on UNICEF's 2017 MICS for Sierra Leone, proceeded above, using Geographically Weighted Regression (GWR) as it contains household surveys across numerous indicators including ICT for both men and women aged 15-49. Similarly, GWR is a "spatial analysis technique that takes non-stationary variables (in this case demographic factors) into consideration and models the local relationships between these predictors and an outcome of interest."("Geographically Weighted Regression", n.d.) It is a useful exploratory technique as it allows for the visualization of relationship across space. More specifically, GWR creates regression for every location in the dataset with some associated bandwidth. The GWR analysis will use the following logistic regression.

$$\begin{aligned}(Frequency\ of\ listening\ to\ the\ radio)_i \ = \ Const \ + (Gender\ Dummy)_i \ + \ Age_i \ + \\ + \ (Wealth\ Index)_i \ + (Education\ level\ dummy[s])_{id} \\ + \ (Married?)_i \ + (Ever\ given\ birth)_i \\ + \ (Domestic\ violence)_i \ + (Disability)_i \ + \epsilon_i.\end{aligned}$$

$$(Mobile\ phone\ usage\ in\ the\ last\ 3\ months\ )_i \ = \ Const\ + (Gender\ Dummy)_i \ + \ Age_i + (Wealth\ Index)_i$$
$$+ \ (Wealth\ Index)_i \ + (Education\ level\ dummy[s])_{id}.$$
$$+ (Married?)_i \ + (Ever\ given\ birth)_i$$
$$+ (Domestic\ violence)_i \ + (Disability)_i \ + \epsilon_i.$$

This paper will examine the effects of the depended variables on the usage of radio and mobile phones. Internet is excluded from the analysis because more than 90% of the population does not have access to computers in the dataset. Similarly, increased likelihood of forced marriage and pregnancy at the senior secondary and tertiary age are probably important factors affecting access and learning. Domestic violence is also an important consideration because it has increased in many parts of the world during the current pandemic and affects the learning capacity of children ("Infographic: The Shadow Pandemic", 2020). Finally, girls with disability face additional social and structural barriers to education. According to the 2018 SLIHS data, 33% of disabled girls between ages 3-24 were not able to access school compared to 20% among the total non-disabled population in the same age group.

Work Cited

UIS. (2018, June). A Global Framework of Reference on Digital Literacy Skills for Indicator 4.4.2. Retrieved March 31, 2021, from http://uis.unesco.org/sites/default/files/documents/ip51-global-framework-reference-digital-literacy-skills-2018-en.pdf

UNESCO Accountability in education: Meeting our commitments. (n.d.). Retrieved March 31, 2021, from https://en.unesco.org/gem-report/report/2017/accountability-education

Girl Effect & Vodafone Foundation. (2018, October 11). Real girls, real lives, connected. Retrieved March 31, 2021, from https://global.girleffect.org/stories/real-girls-real-lives-connected/

Afrobarometer. (2019, November 4). Africa's digital gender divide may be widening, Afrobarometer survey finds. Retrieved March 31, 2021, from https://afrobarometer.org/sites/default/files/press-release//ab_r7_pr3_africas_digital_gender_divide_may_be_widening.pdf

Ledoux, B., & Segniagbeto, K. (2020, April 15). What price will education pay for COVID-19? Retrieved March 31, 2021, from http://www.iiep.unesco.org/en/what-price-will-education-pay-covid-19-13366

Risso-Gill, I., & Finnegan, L. (n.d.). Children's Ebola Recovery Assessment: Sierra Leone. Retrieved March 31, 2021, from https://www.savethechildren.org/content/dam/global/reports/emergency-humanitarian-response/ebola-rec-sierraleone.pdf

Mutseyekwa, T. L. (2020, April 28). Covid-19 in Sierra Leone: Bridging the learning gap through radio. Retrieved March 31, 2021, from https://www.unicef.org/wca/stories/covid-19-sierra-leone-education

Secretariat, G. (2020, April 29). How did Sierra Leone implement radio instruction during the Ebola crisis? Retrieved March 31, 2021, from https://www.globalpartnership.org/blog/how-did-sierra-leone-implement-radio-instruction-during-ebola-crisis

Human Development Reports. (n.d.). Retrieved March 31, 2021, from http://hdr.undp.org/en/indicators/68606

Government of Sierra Leone. (n.d.). Education Sector Plan 2018-2020. Retrieved March 31, 2021, from https://www.globalpartnership.org/sites/default/files/2018-10-sierra-leone-esp.pdf

Sierra Leone Integrated Household Survey 2018 (SLIHS). (n.d.). Retrieved March 31, 2021, from https://www.statistics.sl/index.php/sierra-leone-integrated-household-survey-slihs.html

UNESCO, I'd Blush If I Could. (2020, January 08). Retrieved from https://en.unesco.org/Id-blush-if-I-could

Geographically Weighted Regression. (n.d.). Retrieved March 31, 2021, from https://www.publichealth.columbia.edu/research/population-health-methods/geographically-weighted-regression

Infographic: The Shadow Pandemic - Violence Against Women and Girls and COVID-19. (2020, April 6). Retrieved March 31, 2021, from https://www.unwomen.org/en/digital-library/multimedia/2020/4/infographic-covid19-violence-against-women-and-girls