Czech University of Life Sciences Prague

Faculty of Economics and Management

Department of Information Engineering

Bachelor Thesis

ANALYSIS OF LITERACY RATE WITH THE HELP OF GEOGRAPHIC INFORMATION SYSTEMS (GIS)

ATAKAN AYDOĞAR

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Economics and Management

BACHELOR THESIS ASSIGNMENT

Atakan Aydoğar

Informatics

Thesis title

ANALYSIS OF LITERACY RATE WITH THE HELP OF GEOGRAPHIC INFORMATION SYSTEMS (GIS)

Objectives of thesis

Although many studies have been carried out to increase the literacy rate and positive results have been obtained, unfortunately, the results have not reached sufficient levels. In this sense, it is necessary to better understand the reasons for the literacy rate, which is among the sustainable development goals. This study aims to determine the possible causes of the literacy rate in the environment of geographic information systems with the help of general and local analyzes, the levels of problems that may occur in the models created, whether it varies according to the regions, and to produce solutions or suggestions for increasing the literacy rate.

Methodology

Literacy rate and GIS will be researched and introduced.

Subsequently, The Least Squares method and Geographically Weighted Regression analysis in ArcGIS environment will be mentioned.

The Least Squares method and Geographical Weighted Regression analysis will be performed in ArcGIS environment using the data of the year 2020 obtained from the World Bank.

According to the results of the analysis, the increase or decrease in the literacy rate will be examined. According to the results analyzed, the factors in the regions where the literacy rate is low will be evaluated and solutions will be produced.

The proposed extent of the thesis

40

Keywords

GIS, Literacy, Geographically Weighted Regression (GWR), Spatial analysis, Statistical analysis,Least Squares Method

Recommended information sources

ALLEN, D W. *Getting to know ArcGIS : modelbuilder.* Redlands: ESRI, 2011. ISBN 978-1-58948-255-5.

GORR, W L. – KURLAND, K S. *GIS tutorial for ArcGIS Pro 2.8.* Redlands, California: Esri Press, 2021. ISBN 978-1-58948-680-5.

https://data.worldbank.org/indicator/SE.ADT.1524.LT.ZS

https://ourworldindata.org/literacy

LAW, M. – COLLINS, A. *Getting to know arcGIS for desktop.* Redlands: Esri Press, 2013. ISBN 1589483081.

TOMS, S. – O'BEIRNE, D. *ArcPy and ArcGIS : automating ArcGIS for Desktop and ArcGIS online with Python.* Birmingham: Packt, 2017. ISBN 978-1-78728-251-3.

Expected date of thesis defence

2022/23 WS – FEM

The Bachelor Thesis Supervisor

Ing. Jakub Konopásek, Ph.D.

Supervising department

Department of Information Engineering

Electronic approval: 30. 11. 2022

Ing. Martin Pelikán, Ph.D.

Head of department

Electronic approval: 30. 11. 2022

doc. Ing. Tomáš Šubrt, Ph.D.

Dean

Prague on 18. 01. 2023

Declaration

I declare that I have worked on my bachelor thesis titled " ANALYSIS OF LITERACY RATEWITH THE HELP OF GEOGRAPHIC INFORMATION SYSTEMS (GIS)" by myself and I have used only the sources mentioned at the end of the thesis. As the author of the bachelor thesis, I declare that the thesis does not break any copyrights.

In Prague on the 13.03.2023 _____

Acknowledgment

I would like to thank Jakub Konopásek And all other persons, for their advice and support during my work on this Thesis.

ANALYSIS OF LITERACY RATE WITH THE HELP OF GEOGRAPHIC INFORMATION SYSTEMS (GIS)

Abstract

The aim of this study is to determine the possible causes of the literacy rate in the world and the levels of problems that may be caused in the regions where the literacy rate is low, with the help of general and local analyzes in the geographical information systems environment. The models created are to produce solutions or suggestions to increase the literacy rate and whether they differ according to the regions. Based on the human development index, comments were made by applying the least squares method and geographically weighted regression analysis separately for the countries examined in 4 groups low, medium, high, and very high levels. No meaningful models and results were obtained for the high-level country groups. Overall and local analyzes yielded significant results for very high, medium, and lower-level groups. The lowest literacy rate is seen in Africa and partially in the Asian region. Accordingly, it has been observed that GNI per capita is effective in countries with the lowest literacy rate in Africa. In the models obtained by the OLS method, the $AdjR2$ value for the low-level group was 0.024, while this value was found to be 0.35 according to the GWR analysis. The $AdjR2$ value for the medium-level group was determined as 0.049 as a result of the OLS method and as 0.062 as a result of the GWR analysis. For the very high-level group, the $AdjR2$ value was found to be 0.1 as a result of the OLS method and 0.39 as a result of the GWR analysis.

Keywords: GIS, Literacy, Geographically Weighted Regression (GWR), Spatial analyse, Statistical analyse, Ordinary Least Squares Method(OLS)

ANALÝZA MÍRY GRAMOTNOSTI S POMOCÍ GEOGRAFICKÝCH INFORMAČNÍCH SYSTÉMŮ (GIS)

Abstract

Cílem této studie je pomocí obecných a lokálních analýz v geografických informačních systémech určit možné příčiny míry gramotnosti ve světě a míru problémů, které mohou být způsobeny v regionech s nízkou mírou gramotnosti. životní prostředí. Vytvořené modely mají přinášet řešení nebo návrhy na zvýšení míry gramotnosti a zda se liší podle regionů. Na základě indexu lidského rozvoje byly provedeny komentáře použitím metody nejmenších čtverců a geograficky vážené regresní analýzy zvlášť pro zkoumané země ve 4 skupinách jako nízká, střední, vysoká a velmi vysoká úroveň. Pro skupiny zemí na vysoké úrovni nebyly získány žádné smysluplné modely a výsledky. Celkové a místní analýzy přinesly významné výsledky pro skupiny velmi vysoké, střední a nižší úrovně. Nejnižší míra gramotnosti je patrná v Africe a částečně v asijském regionu. V souladu s tím bylo pozorováno, že HDP na hlavu je efektivní v zemích s nejnižší mírou gramotnosti v Africe. V modelech získaných metodou OLS byla hodnota $AdjR2$ pro nízko úrovňovou skupinu 0,024, zatímco tato hodnota byla podle analýzy GWR zjištěna jako 0,35. Hodnota $AdjR2$ pro skupinu střední úrovně byla stanovena jako 0,049 jako výsledek OLS metody a jako 0,062 jako výsledek GWR analýzy. Pro skupinu na velmi vysoké úrovni byla zjištěna hodnota $AdjR2$ jako výsledek metody OLS 0,1 a jako výsledek analýzy GWR 0,39.

Keywords: GIS, gramotnost, geograficky vážená regrese (GWR), prostorová analýza, statistická analýza, metoda obyčejných nejmenších čtverců (OLS)

Table of Content

1.Introduction

Literacy is an essential skill that enables individuals to effectively communicate, comprehend, and use written language for a variety of purposes. It involves the ability to read, write, and interpret information in various forms, including books, newspapers, digital media, and other written materials. Literacy is crucial for personal development, socialization, and success in both academic and professional settings. It is a fundamental human right that allows individuals to access knowledge, participate in their communities, and exercise their rights and responsibilities as citizens. While literacy rates have improved in recent years, there are still millions of people worldwide who struggle with basic reading and writing skills. As such, promoting literacy remains a critical priority for governments, educators, and advocates around the world.

Poverty and literacy rates are lower in underdeveloped countries, especially in the African continent, compared to other countries. Literacy rates are one of the important factors that show the level of development of a country. For this reason, it is among the data that is closely followed by various international organizations. Development studies on the collection, processing, and reporting of the said data are also ongoing. The correct presentation of data on literacy as well as many data such as poverty, access to clean water, education, and schooling rates, will guide the struggles to be made and policies to be developed in this area. The actual situation should be conveyed to the decision-making and implementing units in the most accurate way.

There are studies to increase the literacy rate. The analysis methods used in these studies are based on general models. In other words, such approaches create a common model for the dataset regardless of location. However, the change in the spatial distribution of the variables used and their interactions with each other depending on the location can provide a clearer presentation of the relationship between them. While looking for solutions to the problems to be explained, it would be more appropriate to examine the data spatially together with geographical conditions in statistical analysis. This perspective brought by the developing technology makes it possible to analyze the data in question in spatial dimensions. As a result, Geographic Information Systems (GIS) that can process and analyze location-dependent relationships have been developed.

GIS is a system where geographical, numerical, and verbal data are combined with different disciplines and processed together, information is obtained, and solutions are produced. Helping decision-making processes based on spatial data and spatial-place relationships focused on solving complex social, economic, environmental, and similar problems in the world; It is the whole of hardware, software, personnel, geographical data, and method that performs the collection, storage, processing, management, spatial analysis, querying and presentation. GIS is important in the mapping of diseases, the study of plant health, geospatial and sociological analysis based on location, the spatial study of basic sciences such as physics and chemistry, the description of populations, spatial statistics, and modelling.

In this study, literacy rate data according to the human development index data were used. The data were collected in 4 groups: 0-0.55 group low human development level, 0.55-0.7 group medium level, 0.7-0.8 group high level, and 0.8 and above group very high. Then, in order to create a preview image showing how the data is distributed over the world, thematic maps created by highlighting the information to be highlighted on the topographic base were drawn. Looking at the thematic maps, the lowest literacy rate is in Africa and partially in Asia. Significant results have been obtained in the regression models established in these regions where the gross national income per capita is low.

2.Objective and Methodology

2.1. Objective

Although many studies have been carried out to increase the literacy rate and positive results have been obtained, unfortunately, the results have not reached sufficient levels. In this sense, it is necessary to better understand the reasons for the literacy rate, which is among the sustainable development goals.

This study aims to determine the possible causes of the literacy rate in the environment of geographic information systems with the help of general and local analyzes, the levels of problems that may occur in the models created, whether it varies according to the regions, and to produce solutions or suggestions for increasing the literacy rate.

2.2.Methodology

General regression models are widely used in statistical analyses used in literacy rate studies. These models are called general because a single model emerges in the analyses made for the area of study. In regional models, different models and different results for these models can be obtained for data points in the same area. Considering regional differences in understanding and solving a problem can provide a better result from the study. In this study, geographic weighted regression analysis was applied together with the ordinary least squares method applied in the ArcGIS environment. The reason for this is not only the regional perspective that the geographically weighted regression analysis adds but also that it is applicable if the general model is first created, and its assumptions are met for the application of this analysis. In other words, this analysis can be carried out on models that have not undergone any manipulation and are correctly established. In this sense, it has been concluded that this regional analysis in the GIS environment will give more accurate and higher-quality results than the general analysis methods within the framework of logical and scientific ethics.

3.Literature Review

3.1. Literacy

Literacy is the ability to read and write. It is a fundamental skill for individuals to participate fully in society. Without literacy, individuals are unable to access information, communicate effectively, and make informed decisions.

Literacy is important for personal development as well. Being able to read and write opens up a world of knowledge and self-expression. It allows individuals to learn about new subjects, gain new skills, and share their thoughts and ideas with others.

"Once you learn to read, you will be forever free." (Douglass)

However, not everyone has the same level of literacy. According to the United Nations Educational, Scientific and Cultural Organization (UNESCO), there are over 774 million adults and 126 million youth who are illiterate. This disparity in literacy can lead to several problems, such as poverty, unemployment, and poor health.

Many organizations and governments have implemented programs to promote literacy to combat this issue. These programs often focus on providing access to education, resources, and support for individuals to improve their reading and writing skills.

In addition, technology has also played a major role in promoting literacy. The widespread availability of smartphones and the internet have made it easier for individuals to access information and learn new skills. Online resources such as e-books, educational apps, and online tutorials have made it possible for people to learn at their own pace and in their own time.

Overall, literacy is a crucial skill that enables individuals to participate fully in society and to better their lives. It is important that we continue to support and promote literacy to ensure that everyone has the opportunity to read and write. (Ortiz-Ospina, 2018)

3.2. Literacy Rate

The literacy rate is a measure of the percentage of a population that can read and write. It is a commonly used indicator of a country's level of education and is closely linked to overall societal development. The higher the literacy rate, the more developed a society is considered to be.

One of the most important factors that influence literacy rate is access to education. Countries with high literacy rates typically have well-funded and well-organized education systems that provide access to schooling for all children. This allows for a higher percentage of the population to learn how to read and write, leading to an increased literacy rate.

Economic development also plays a role in the literacy rate. Countries with strong economies tend to have higher literacy rates because they can invest more in education. Additionally, individuals in these countries are more likely to have access to resources such as books and computers, which can also contribute to an increased literacy rate.

Gender also plays a role in the literacy rate. In many countries, girls and women have less access to education than boys and men. This leads to a lower literacy rate among females. Efforts to promote gender equality in education can help to close this gap and increase overall literacy rates (Unesco Institute for Statistics).

Despite the many factors that contribute to the literacy rate, there are still many countries where the literacy rate is low. This is particularly true in developing countries, where access to education is limited. Governments, organizations, and individuals need to work together to improve literacy rates through initiatives such as adult literacy programs and community-based education.

In conclusion, the literacy rate is a measure of the percentage of a population that can read and write. It is closely linked to overall societal development and is influenced by factors such as access to education, economic development, and gender equality. Efforts to improve literacy rates are crucial for promoting growth and providing individuals with the skills they need to participate fully in society.

3.3. Geographic Information Systems (GIS)

Geographic Information Systems (GIS) is a technology that allows users to capture, store, analyse, and visualize geographical data. This data can include information such as maps, satellite imagery, and demographics, and is used in a wide range of applications, including urban planning, natural resource management, and emergency response.

One of the key advantages of GIS is its ability to integrate and analyse large amounts of data from different sources. This allows users to gain a more complete understanding of a particular area or problem, and to make more informed decisions. For example, GIS can be used to analyse the impact of a proposed development on the surrounding environment by overlaying information about land use, soil types, and endangered species.

Another advantage of GIS is its ability to visualize data in a way that is easy to understand. This can be done through the use of maps, 3D models, and other graphical representations. This is particularly useful for communicating complex information to non-experts, such as policymakers or the public.

GIS is also a very flexible technology, which can be used in a wide range of applications. Some examples of GIS applications include:

- Urban planning and management, where GIS can be used to analyse population density, land use, and transportation networks.
- Natural resource management, where GIS can be used to track the distribution of wildlife, monitor changes in land cover, and manage water resources.
- Emergency response, where GIS can be used to track the spread of a disaster, identify evacuation routes, and allocate resources. (Fagerholm, 2021)

Despite the many benefits of GIS, there are also some challenges associated with this technology. One of the main challenges is ensuring that the data used in GIS is accurate and up to date. This can be difficult to achieve, particularly in developing countries where data collection infrastructure may be limited. Another challenge is that GIS can be complex and difficult to use, particularly for non-experts.

In conclusion, Geographic Information Systems (GIS) is a powerful technology that allows users to capture, store, analyze, and visualize geographical data. It is useful in a wide range

of applications, including urban planning, natural resource management, and emergency response. GIS has many advantages, such as integrating and analyzing a large amount of data, easily visualizing the data, and flexibility. However, there are also challenges, such as ensuring data accuracy and the complexity of the technology.

3.4. ArcGIS

ArcGIS is a software developed by Esri, one of the leading companies in the field of Geographic Information Systems (GIS). ArcGIS is a comprehensive platform that allows users to create, manage, analyze, and share geographical data. It is widely used in various industries such as government, transportation, health, and education.

One of the key features of ArcGIS is its ability to create and edit maps. ArcGIS provides a wide range of tools that allow users to create maps with a high level of detail and accuracy. These tools include the ability to add layers of information, such as satellite imagery, demographic data, and street maps. Users can also customize the appearance of their maps by changing the colors, labels, and symbols used.

ArcGIS also offers a wide range of analytical tools that allow users to extract insights from their data. These tools include spatial analysis, network analysis, and 3D analysis. For example, spatial analysis tools can be used to identify patterns and relationships in data, such as the correlation between land use and crime rates. Network analysis tools can be used to model transportation networks, while 3D analysis tools can be used to create 3D models of buildings and other structures.

ArcGIS also offers a range of tools for sharing and collaborating with others. The software allows users to share their maps and data through a variety of platforms, such as web maps and mobile apps. This allows others to view and interact with the data and to contribute their data and insights. This feature enables different agencies or departments within an organization to work together and share information seamlessly, resulting in more efficient workflows and better decision-making. (Meg Miller, 2021)

ArcGIS is also a very flexible platform, which can be used in a wide range of applications. Some examples of ArcGIS applications include:

- Urban planning and management, where ArcGIS can be used to analyse population density, land use, and transportation networks.
- Natural resource management, where ArcGIS can be used to track the distribution of wildlife, monitor changes in land cover, and manage water resources.
- Emergency response, where ArcGIS can be used to track the spread of a disaster, identify evacuation routes, and allocate resources.

Despite the many benefits of ArcGIS, there are also some challenges associated with this software. One of the main challenges is that ArcGIS can be complex and difficult to use, particularly for non-experts. This can be mitigated by providing adequate training and support to users. Another challenge is the cost of the software, which can be expensive for some organizations. (A. A. Emelyianov, 2023)

In conclusion, ArcGIS is a powerful software developed by Esri that allows users to create, manage, analyze, and share geographical data. It offers a wide range of tools for creating and editing maps, analyzing data, and collaborating with others. ArcGIS is widely used in various industries and is flexible enough to be used in a wide range of applications. However, there are also some challenges associated with ArcGIS, such as complexity and cost. (Michael Law, 2021)

4. Practical Part

In this study, which was carried out around the world with the data of 2020 obtained from the World Bank, the factors that could explain the increase or decrease in the literacy rate were searched with the Least Squares method and Geographically Weighted Regression analysis in the ArcGIS environment, and it was investigated whether it changed according to the regions with the maps created. In addition, according to the models obtained, whether the possible causes of the literacy rate vary regionally and the explanatory nature of the established local models was examined.

4.1. Material

Literacy is recognized as a worldwide problem by international communities. The fact that it is not a problem of any society but a common problem of all countries of the world has revealed the necessity of worldwide struggle efforts. The fact that the study is carried out worldwide means the search for a global solution, while at the same time, it means that the data to be used in the analysis are complete or close to completion based on countries. For this reason, data for 2020 was used as it was aimed to use the closest, most complete, and most accessible data.

4.2. Method

After collecting the data to be used in the study, first, the world base map was loaded in the ArcGIS program, and then the data groups were processed into the attribute tables. Thematic maps were created, previewed, and interpreted by showing the symbolization of each data group on the map on the processed base created. Subsequently, "Multiple Linear Regression Analysis" was conducted to question the existence, direction, and strength of the relationship between the literacy rate and the variables thought to be related to the literacy rate. To establish a linear model to explain the cause-effect relationship between

two or more independent variables affecting an event and to analyze. The strengths of the effects of these variables on the event.

The method used to determine the data is called "Multiple Linear Regression Analysis." Regression analyses are available in many package programs, and although it seems easy to implement, it is an error-prone analysis because it has assumptions, and these assumptions need to be carefully controlled. The Multiple Linear regression model is as follows.

Figure 1: The Multiple Linear regression Model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta 2 X_{i2} + ... + \beta_p Xi_p + \varepsilon_i$$

Source: *(Chowdary, 2020)*

"Y" indicates the variable we want to be explained (the dependent variable), and "X" represents the variable or variables (independent variable) that are thought to explain the dependent variable." β" represents the regression coefficients in the model created. Finally, "ε" represents the error term. (Dastan Hussen Maulud, 2020)

In order to create an accurate regression model, there are certain assumptions that the data used in the analysis and the model must provide. These are the normality assumption, the independence assumption of the residues, the homoscedasticity assumption, and the collinearity assumption. The assumption of a normal distribution is very important for the application of parametric tests. Regression analysis with a data set that does not show normal distribution gives misleading results. Normality can be tested with many tests, such as Kolmogorov-Smirnov and Shapiro-Wilks, to control normality. In addition, the error terms, "ε" values, must also be normally distributed. Under the assumption of the independence of the residuals, the mean of the error terms is expected to be zero, the variance is constant, and it is expected to show an uncorrelated relationship between them. In addition, the explanatory variables should not be linearly related to each other. This linear relationship is a multicollinearity causing erroneous results. (Daoud, 2017)

It is seen that the literacy rate data is not normally distributed when the controls of the data are provided. The logarithmic transformation, which is a method frequently used in the field of econometrics, was applied to the dependent variable.

Afterward, the assumption of normality was realized by dividing them into four groups based on the human development index. (Benoit, 2011)

In the next step, in order to test the significance of the models for each group created, variance analysis is performed, and the hypothesis that the model is invalid or valid is tested. Then, the hypothesis that the second hypothesis, the regression coefficient (β), is significant or unimportant is tested. Since the regression coefficient is insignificant as a result of the hypothesis, it is expected that this coefficient will be significant since it will result in the conclusion that 1-unit changes in the explanatory variable or variables will not cause a change in the dependent variable. Finally, by looking at the coefficient of determination R^2It is interpreted how much of the change in the dependent variable can be explained with the help of independent variables. As variables are added to the regression model, the disclosure rate may increase, but at the same rate, an error arising from the randomly explained change is also included. To avoid this error, the model is based on $AdjR$^2 (Adjusted Determination Coefficient) instead of R^2. **(J. G Liao, 2012)**

The analysis of spatial characteristics, spatial distribution, spatial interaction, and relationship of the information obtained today has been much improved by the incorporation of statistical methods into geographic information systems. Because the distribution and interaction of the variables according to their positions in the space also have an important place in the explanation of the relationship between the variables.

Regional differences or interactions can change the effect of each independent variable on the dependent variable. It is not possible to analyze these differences by including them in general regression models such as the ordinary least squares (OLS) method. Global statistics are usually single-valued. Examples of mean value,

standard deviation, and spatial autocorrelation measure in the dataset. Therefore, in addition to the OLS method, a Geographically Weighted Regression (GWR) analysis, which can make separate calculations for each location, taking into account regional differences, was performed.

Local statistics are invaluable. Different values of statistics may occur at different locations within the study area. Each local statistic is a measure of the attribute or relationship studied near a location within the study. Local statistics may take different values as this location changes. In this sense, the geographically weighted regression model has turned the general regression phenomenon into geographical regression by highlighting the concept of spatial non-stationary. In the working principle of the GWR model, the regression coefficient "β" has been developed with the spatial weight matrix.

Finally, since separate regression results can be obtained for each country with the GWR model, the homogeneity and heterogeneity of the spatial distribution were determined, and the results and local R^2 values were mapped. (Majid Nazeer, 2018)

5. Results and Discussion

The attribute table was created by processing the data in the world base. Afterward, preview maps, mapping, and interpretation of results obtained from OLS and GWR models are mentioned.

5.1. Results

Thematic maps were created and interpreted for the three indicators. In addition, general and local maps were produced according to the results of OLS and GWR analyses.

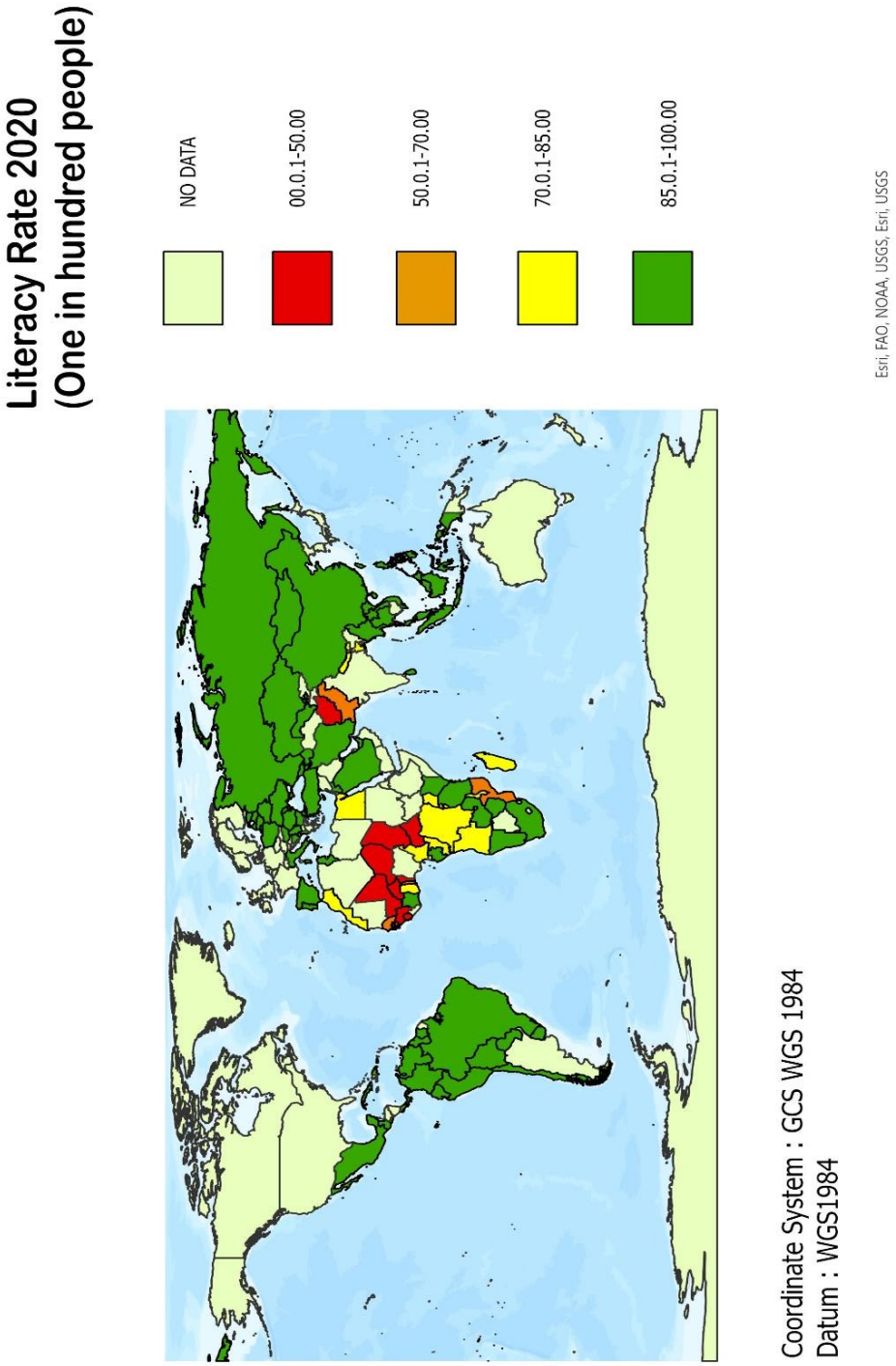
5.1.1. Thematic Maps of Parameters

Thematic maps obtained under this title are given. The best and worst-off countries and critical-level countries are indicated on the maps obtained for each indicator.

5.1.1.1. Literacy Rate

In the “Literacy” map of 2020, the country with the lowest literacy rate, Chad, ranks first with a literacy rate of 26.76. Afghanistan is the second country with the lowest literacy rate at 37.27. Niger is third with 37.34, the Central African Republic is fourth with 37.49, and Burkina Faso is fifth with 46.04. The countries with the highest literacy rate are Ukraine, first with 100. Latvia is second with 99.89, and Belarus and Estonia are third with 99.87. Lithuania is fourth with 99.83, and Kazakhstan is fifth with 99.80. (Figure 1.)

Figure 2. Literacy Rate Map for 2020

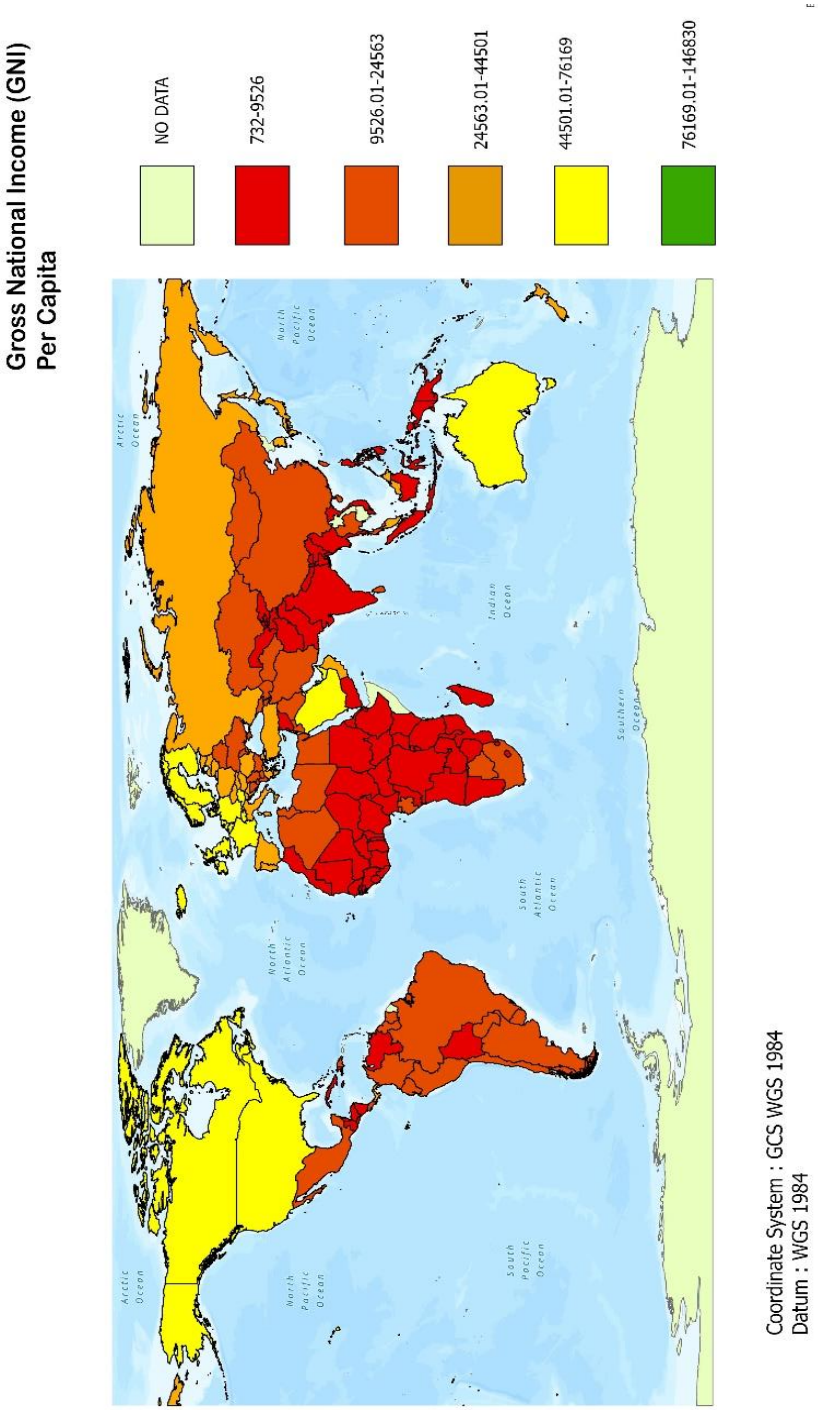


Source: Own Processing

5.1.1.2 Gross National Income (GNI) Per Capita

In the “Human Development Index and Its Components” map of 2021, the country with the Gross National Income (GNI) per capita, Burundi, ranks first with a Gross national income of 732$. South Sudan is the second country with the lowest income at 768$. Central African Republic is third with 966$. The Congo Democratic Republic is fourth with 1076$, and Mozambique is fifth with 1198$. The countries with the highest gross national income per capita are Liechtenstein is first with 146830$. Singapore is second with 90919$, and Qatar is third with 87134$. Luxembourg is fourth with 84649$, and Ireland is fifth with 76169$. (Figure 3.)

Figure 3 Gross National Income (GNI) Per Capita Map

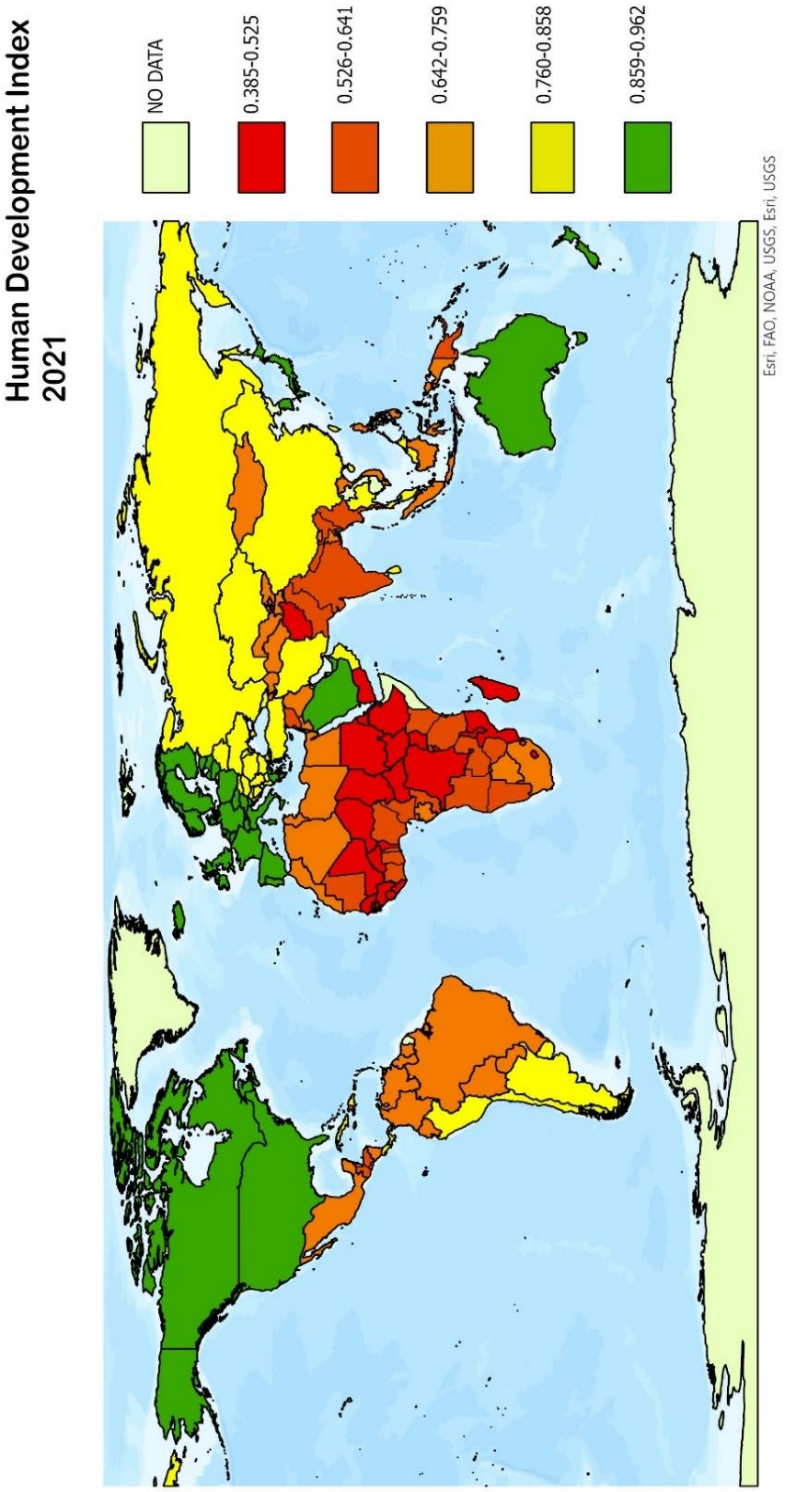


Source: Own Processing

5.1.1.3. Human Development Index

In the “Human Development Index and its components” map of 2021, the country with the Human Development Index (HDI), South Sudan, ranks first with the lowest human development index of 0.385. Chad is the second country with the lowest human development index at 0.394. Niger is third with 0.400, the Central African Republic is fourth with 0.404, and Burundi is fifth with 0.426. The countries with the highest human development index are Switzerland is first, with 0.962. Norway is second with 0.961, and Iceland is third with 0.959. Hong Kong is fourth with 0.952, and Australia is fifth with 0.951. (Figure 4.)

Figure 4 Human Development Index Map for 2021



Source: Own Processing

5.1.2. Regression (OLS) Analysis of Parameters

OLS analysis, which produces a general regression model, was applied to the data set divided into four groups according to human development levels and interpreted for each group.

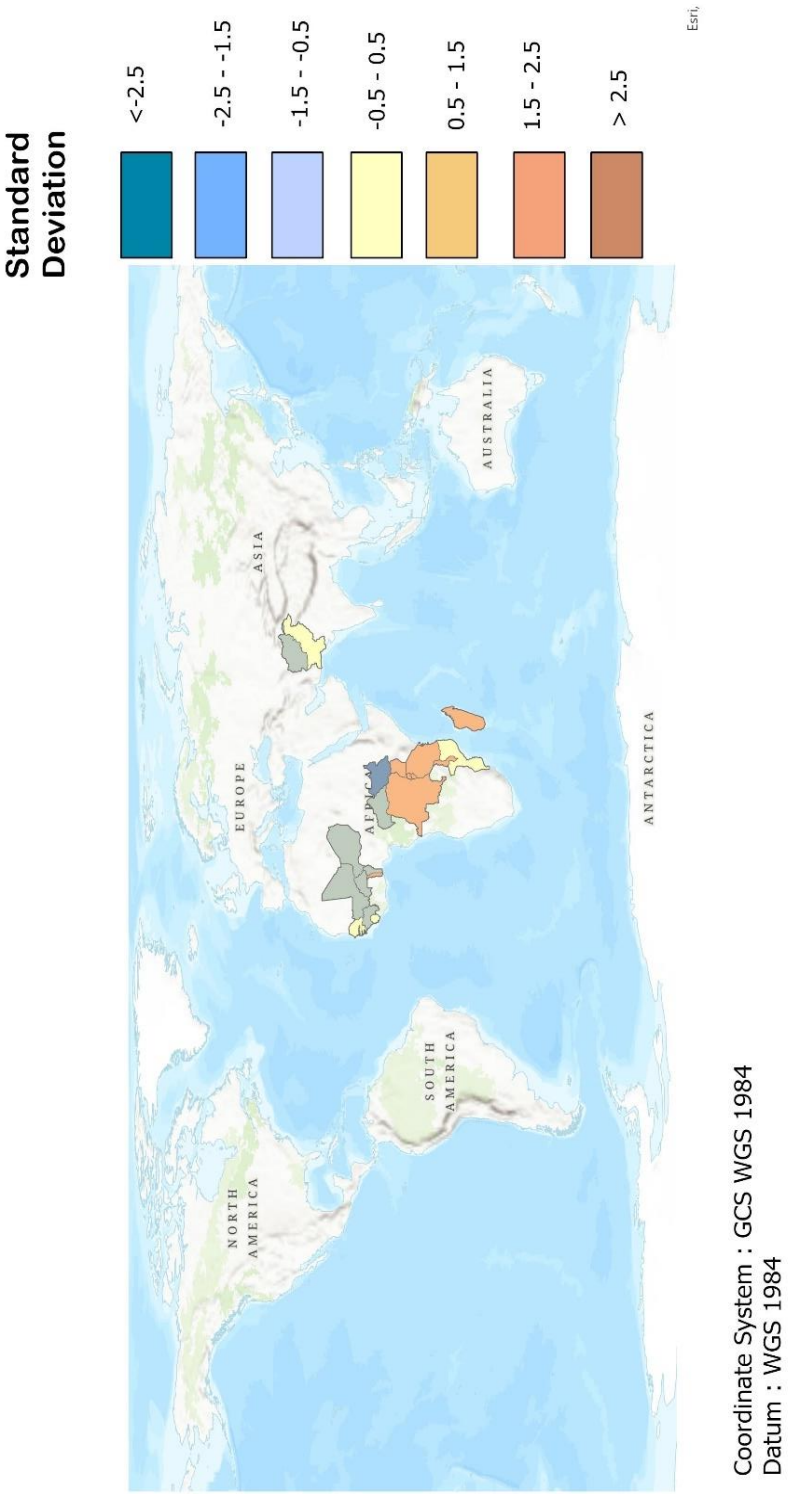
5.1.2.1. Countries with a Low Level of Human Development Index

Variables that are thought to explain the literacy rate were included in the model, and all variations were tested. According to the results obtained in the OLS (Ordinary Least Square) method performed in the ArcGIS environment, a significant regression model could be established between the literacy rate and gross national income per capita in countries with low human development levels. In this case, the change in literacy rate can be explained by 0.24% gross national income per capita.

It has been determined that when the gross national income per capita increases by one unit, the literacy rate increases by 0.09%.

According to the OLS Map, the deviation of the literacy values observed in the model created for countries with a low level of human development from the expected literacy rate values is shown on the map from blue to pink. Negative values indicate that the observed values are smaller than the expected values. Positive values indicate that the observed values are greater than the expected values. Based on these results, it can be said that the literacy rate is lower than expected in South Sudan, Mali, Guinea, Burkina Fatso, Niger, Central African Republic, and Afghanistan. On the other hand, it can be said that the literacy rate in the Democratic Republic of Congo, Tanzania, Uganda, and Madagascar is higher than expected. (Figure 5.)

Figure 5 OLS Map of Low Level Human Development Index Countries



Source: Own Processing

5.1.2.2. Countries with a Medium Level of Human Development Index

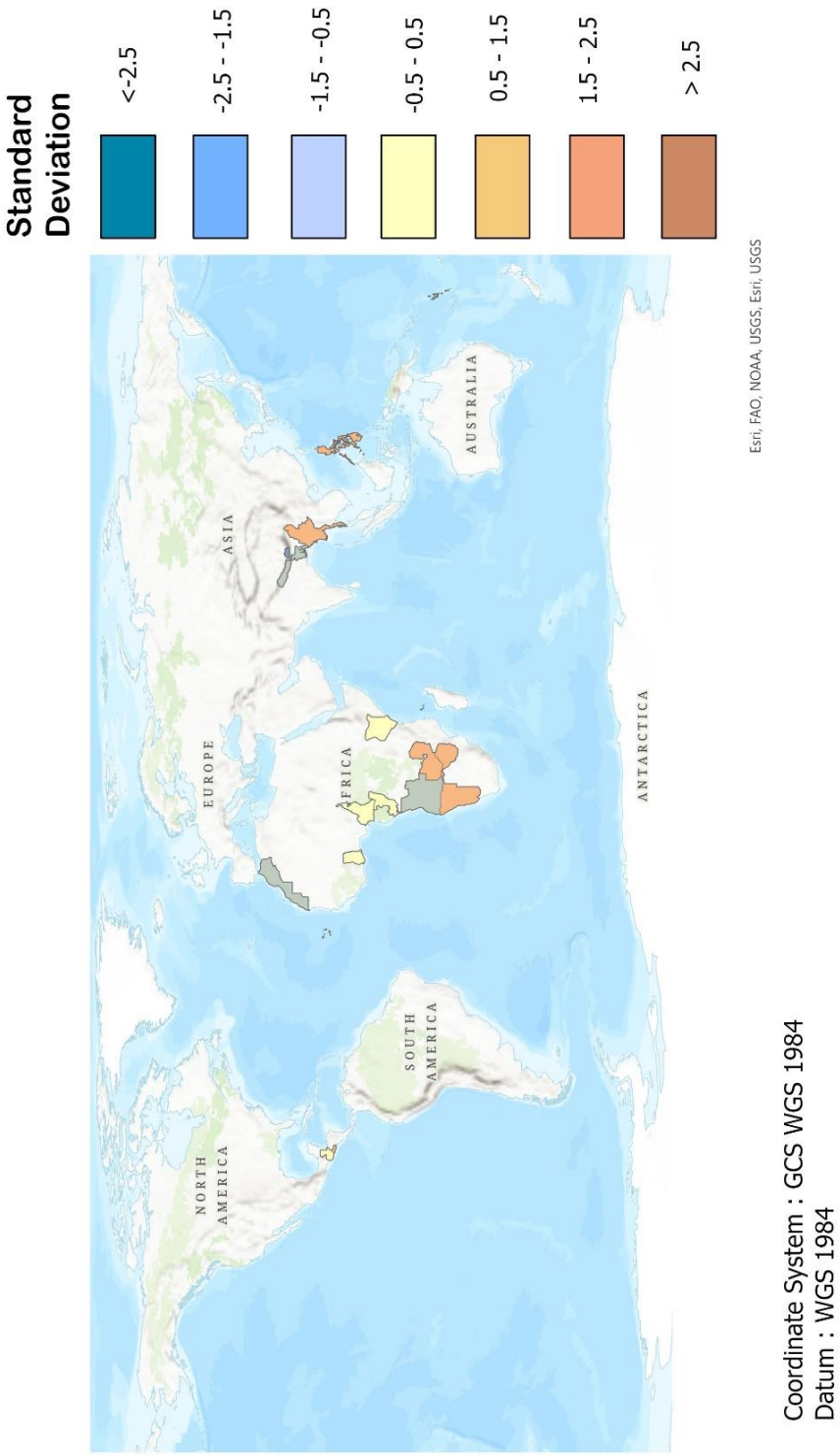
As a result of the analyses made for countries with a medium level of human development, a model was established in which the initial assumptions of the regression analysis were provided. It has been commented that the per capita income variable can explain the literacy rate adjR2=0.0493, that is, 4.93%. It has been determined that when the gross national income per capita increases by one unit, the literacy rate increases by 0.09%

According to these results, it can be said that the literacy rate in Angola, Morocco, Bhutan, and Bangladesh is lower than expected. On the other hand, it can be said that the literacy rate is higher than expected in El Salvador, Namibia, Zimbabwe, Zambia, Myanmar, and the Philippines. (Figure 6.)

5.1.2.3. Countries with a High Level of Human Development Index

As a result of the analyses made for countries with a high level of human development, a model with the initial assumptions of the regression analysis could not be established. It was observed that the residual deviations in the model were not normally distributed. Since the normality assumption could not be provided, a suitable model could not be established for countries with a high level of human development. Therefore, geographic weighted regression analysis will not be performed for this group.

Figure 6 OLS Map of Medium Level Human Development Index Countries



.

Source: Own Processing

5.1.2.4. Countries with a Very High Level of Human Development Index

As a result of the analyses made for countries with a very high level of human development, a model was established in which the initial assumptions of the regression analysis were provided. It has been commented that the per capita income variable can explain the literacy rate of 0.1%. It has been determined that when the gross national income per capita increases by one unit, the literacy rate increases by 0.0036%

According to these results, it can be said that the literacy rate in Turkey, Spain, Portugal, and Russia is equal to what is expected. On the other hand, it can be said that the literacy rate in Kazakhstan, Belarus, and Georgia is higher than expected. (Figure 7.)

Figure 7 OLS Map of Very High Level Human Development Index Countries

Source: Own Processing

5.1.3. Geographically Weighted Regression (GWR) Analyses of Parameters

At the end of the OLS analyses, meaningful models were created in the low, medium and very high level groups. Afterwards, local models were created for these groups and GWR analyses were performed to monitor and interpret regional changes.

5.1.3.1. Countries with a Low Level Human Development Index

After the OLS method was applied to countries with a low level of human development, the GWR analysis was inserted into the model with the same variables for the same countries, and it was investigated whether the location and neighborhood of the countries had a positive effect on the model.

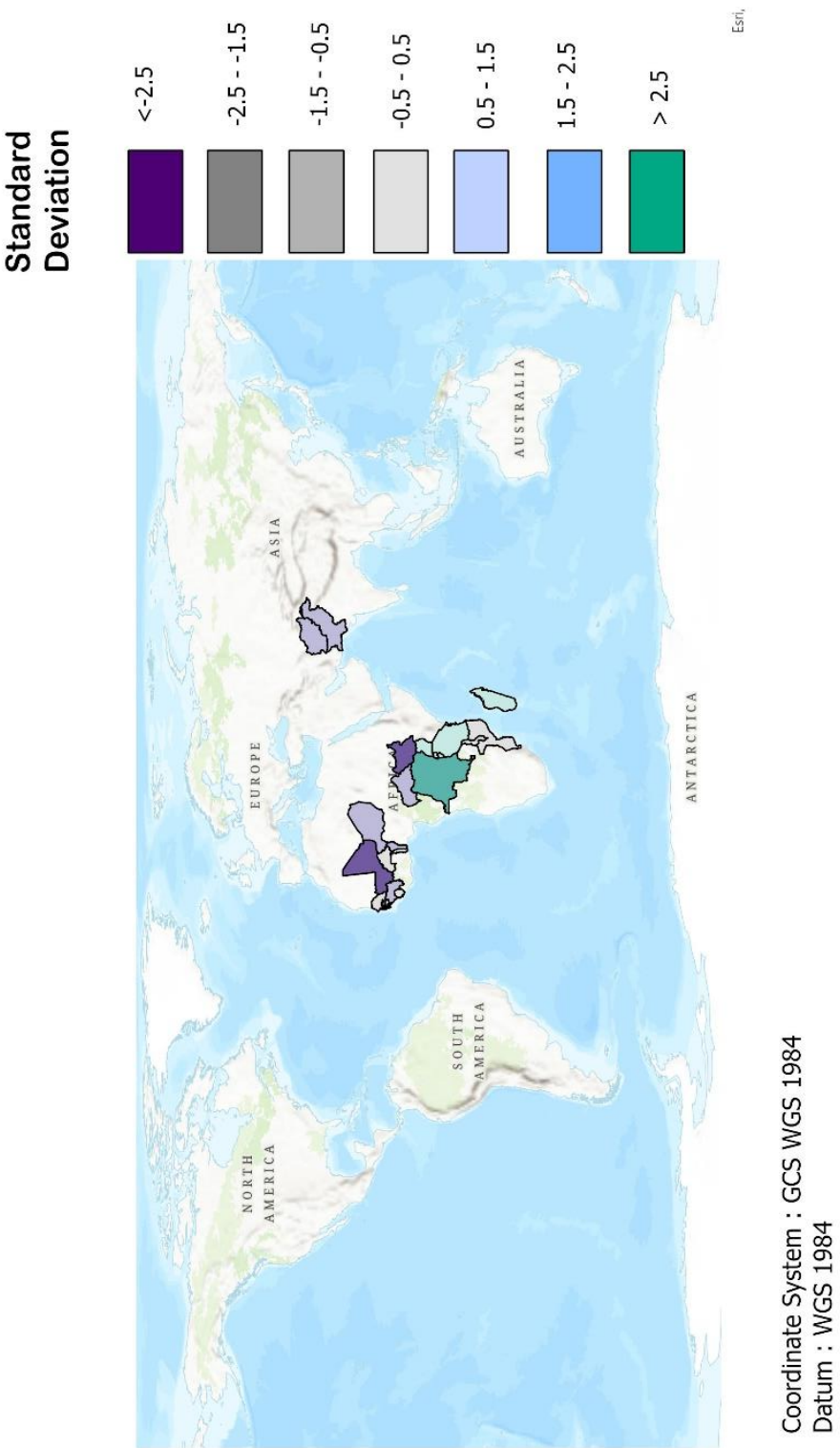
Compared to the OLS model with an AdjR2 value of 0.0024, the GWR model gave better results, and the AdjR2 value increased to 0.35.

In other words, with this model, 35% of the change in literacy rate can be explained by gross national income per capita. Accordingly, the countries where the literacy rate is lower than expected are Afghanistan, Mali, Niger, Guinea, South Sudan, the Republic of Central Africa, and Pakistan.

Countries with higher literacy rates than expected are Congo Democratic Republic, Madagascar, Mozambique, Uganda, and Tanzania.

According to local $R2$ values, the country that can best be explained with 42% in South Sudan. Uganda is in second place with 41%. Rwanda is in 3rd place with 40%. (Figure 8.)

Figure 8 GWR Map of Low Level Human Development Index Countries



Source: Own Processing

5.1.3.2. Countries with a Medium Level Human Development Index

After the OLS method was applied to countries with a medium of human development, the GWR analysis was inserted into the model with the same variables for the same countries, and it was investigated whether the location and neighborhood of the countries had a positive effect on the model. Compared to the OLS model with an AdjR2 value of 0.049, the GWR model gave better results, and the AdjR2 value increased to 0.062.

In other words, with this model, 6.2% of the change in literacy rate can be explained by gross national income per capita. Accordingly, countries with lower-than-expected literacy rates are Guatemala, Angola, Morocco, Cameroon, Congo, Bangladesh, Nepal, and Bhutan.

Countries with higher-than-expected literacy rates are the Philippines, Myanmar, Namibia, Zimbabwe, and Zambia.

According to local $R2$ values, the country that can best be explained with 15.9% in Guatemala. El Salvador is in second place with 15.8%. Coba Verde is in 3rd place with 13%. (Figure 9.)

Figure 9 GWR Map of Medium Level Human Development Index Countries

Source: Own Processing

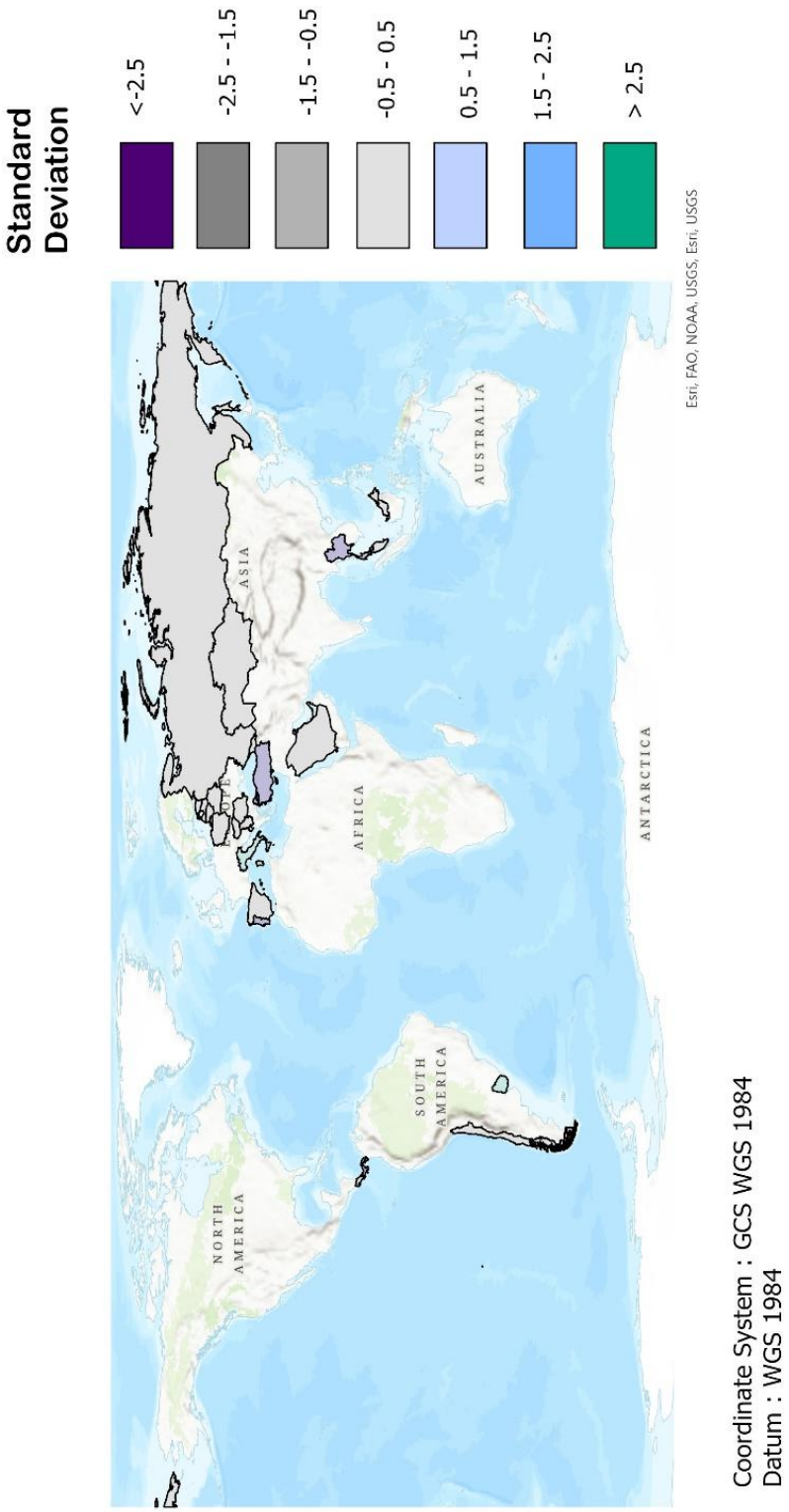
5.1.3.3. Countries with a Very High Human Development Index

After the OLS method was applied to countries with very high human development, the GWR analysis was inserted into the model with the same variables for the same countries, and it was investigated whether the location and neighborhood of the countries had a positive effect on the model. Compared to the OLS model with an AdjR2 value of 0.01, the GWR model gave better results, and the AdjR2 value increased to 0.39. In other words, with this model, 39% of the change in literacy rate can be explained by gross national income per capita.

Accordingly, the countries with lower-than-expected literacy rates are Turkey, Panama, Portugal, and Thailand.

Countries with higher-than-expected literacy rates are Italy, Uruguay, and Brunei Darussalam. Countries with equal expected literacy rates are Cyprus, Russia, Spain, Belarus, Hungary, Kazakhstan, Saudi Arabia, Estonia, Latvia, and Malaysia. According to local $R2$ values, the country that can best be explained with 92% is Brunei Darussalam. Chile ranks second with 89%. Uruguay ranks 3rd with 79%. (Figure 10.)

Figure 10 GWR Map of Very High Level Human Development Index Countries



Source: Own Processing

5.2. Discussion

For the future of societies, people need to know how to read and write in the best way. Literacy is the ability to read the literature of a language and to perceive and comprehend the items read. Many thinkers argue that "being literate is a human right." Because quality basic education equips students with the literacy skills necessary for further learning experiences, literate parents are more likely to raise healthy children and send their children to school. Literate people have higher access to other educational and job opportunities. When we look at society in general, literate societies cope better with the challenges they face. As a matter of fact, with the emergence of information societies, being literate has gained more vital importance.

Every year, thousands of children cannot go to school due to inadequate economic conditions. There are some children who do not have access to information while in the area they live in. Today, there is still a section that does not find it appropriate for girls to go to school. It is possible that we can overcome these problems with education.

Developing technology, growing economies, ever-increasing social and international awareness, and despite their efforts to raise the literacy rate to the highest, the increase in the literacy rate is one of the development goals of a society.

In this study, the literacy rate, economy (per capita national income), human development index, 2020 literacy data, and 2021 human development index data are used to compare these variables. The relationship was examined in the GIS environment with the help of the OLS method and regression analysis with the help of GWR analysis. According to the human development index, the group with the low level of human development is 0-0.55, the group with the medium level is 0.55-0.7, the group with the high level is 0.7-0.8, and the group with the highest level is 0.8 and above. The data were collected in 4 groups. When we look at the thematic maps, the lowest literacy rate is seen in Africa and the Asian region. In South America, North Africa, and Asia, meaningful models could be established for gross national income per capita and literacy rates.

According to the results obtained in the OLS method applied in the ArcGIS environment, a significant regression model could be established between the child literacy rate and the gross national income per capita in the analyses made for the group with low human development levels. In this case, 0.24% of the change in literacy rate can be explained by gross national income per capita. In addition to the model, the GWR model, which gave more specific answers, gave better results and increased the disclosure value from 0.24% to 35%. In other words, according to the GWR model, gross national income per capita can explain 35% of the variation in literacy rates. In addition, local $R2$ values were obtained for the regional variation in explaining the literacy rate, which is one of the main objectives of the study. Accordingly, South Sudan is the country that can be best explained with 42% according to local $R2$ values. Uganda is in second place with 41%. Rwanda comes in 3rd place with 40%. According to the results obtained in the OLS method applied in the ArcGIS environment, a significant regression model could be established between the child literacy rate and the gross national income per capita in the analyses made for the group with medium human development level. In this case, 4.93% of the change in literacy rate can be explained by gross national income per capita. In addition to the model, the GWR model, which gave more specific answers, gave better results and increased the disclosure value from 4.93% to 6.2%. In other words, according to the GWR model, gross national income per capita can explain 6.2% of the variation in literacy rates. In addition, local $R2$ values were obtained for the regional variation in explaining the literacy rate, which is one of the main objectives of the study. Accordingly, Guatemala is the country that can be best explained with 15.9% according to local $R2$ values. El Salvador is in second place with 15.8%. Coba Verde comes in 3rd place with 13%. According to the results obtained in the OLS method applied in the ArcGIS environment, a significant regression model could be established between the child literacy rate and the gross national income per capita in the analyses made for the group with a very high human development level. In this case, 1% of the change in literacy rate can be explained by gross national income per capita. In addition to the model, the GWR model, which gave more specific answers, gave better results and increased the disclosure value from 1% to 39%. In other words, according to the GWR model, gross national income per capita can explain 39% of the variation in literacy rates.

In addition local $R2$ values were obtained for the regional variation in explaining the literacy rate, which is one of the main objectives of the study. Accordingly, Brunei Darussalam is the country that can be best explained with 92% according to local $R2$ values. Chile is in second place with 89%. Uruguay comes in 3rd place with 79%.

As a result, countries with low literacy rates experience economic difficulties. Economically prosperous, freer, more educated

The positive effects of more knowledgeable individuals on society and welfare are also inevitable. The impositions imposed on women by traditional patriarchal social structures, religious and traditional narrow-minded mentality, and many similar negative factors are not only applied to women but also to all children, including children, who constitute the most vulnerable group. Looking at the local $R2$ map created for the low-level group, it is seen how effective the per capita gross national income literacy rate is in African countries with the lowest literacy rate. Women who are not allowed or have no access to school in their region, who are regularly exposed to gender inequality in their social lives, and who cannot receive adequate health care due to poverty. While women who cannot reach school are not self-sufficient and constantly struggling for survival, it seems that societies' longing for education for children who are born healthy and grow up in healthy environments will continue for a long time. According to the GWR analysis, the concepts of children and women should be considered together in these regions mentioned in the studies aimed at increasing the literacy rate. Considering the importance of going to school and its positive effect on the development of economic conditions in the social policies and improvement studies to be created. It will be a good struggle tool in these countries where the literacy problem is the most common in the world.

6.Conclusion

In this study, literacy rate data according and human development index data were used. The data were collected in 4 groups: 0-0.55 group low human development level, 0.55-0.7 group medium level, 0.7-0.8 group high level, and 0.8 and above group very high. Then, in order to create a preview image showing how the data is distributed over the world, thematic maps created by highlighting the information to be highlighted on the topographic base were drawn. Looking at the thematic maps, the lowest literacy rate is in Africa and partially in Asia. Important results were obtained in the regression models established in these regions where the gross national income per capita is low. Analyses were made between the literacy rate and the gross national income per capita variables. First, the ordinary least squares method was applied to the data collected in 4 different groups. A meaningful model could not be established in the high-level group.

Significant models were established in the Low, Medium, and High-level groups, and OLS maps were created and interpreted by analyzing the Argis environment.

A geographically weighted regression analysis was applied to the collected data in the ArcGIS environment, and meaningful models were established. These models were mapped and interpreted.

7.References

A. A. Emelyianov, M.V. Ereshko. 2023. A Review of Advanced Cloud Platforms for ERS Data Processing and Analytics. *link.springer.* [Online] 7 February 2023. https://link.springer.com/article/10.1134/s0001433822090079.

Benoit, Kenneth. 2011. Linear Regression Models with Logarithmic Transformations. *links.sharezomics.* [Online] 17 March 2011. https://links.sharezomics.com/assets/uploads/files/1600247928973-from_slack_logmodels2.pdf.

Chowdary, Davuluri Hemanth. 2020. Multiple Linear Regression Explained. *medium.* [Online] 12 May 2020. https://medium.com/analytics-vidhya/multiple-linear-regression-explained-215f2683cd5a.

Daoud, Jamal I. 2017. Multicollinearity and Regression Analysis. *iopscience.iop.* [Online] 8 August 2017. https://iopscience.iop.org/article/10.1088/1742-6596/949/1/012009/pdf.

Dastan Hussen Maulud, , Adnan Mohsin Abdulazeez. 2020. A Review on Linear Regression Comprehensive in Machine. *semanticscholar.* [Online] 29 December 2020. https://pdfs.semanticscholar.org/5d7c/b4891c73ad8ec642e4d8ab2220bb52a64086.pdf.

Douglass, Frederick.

Fagerholm, Nora. 2021. A methodological framework for analysis of participatory mapping data in research, planning, and management. *tandfonline.* [Online] 14 Jan 2021. https://www.tandfonline.com/doi/full/10.1080/13658816.2020.1869747.

J. G Liao, Dan McGee. 2012. Adjusted Coefficients of Determination for Logistic Regression. *tandfonline.* [Online] 01 January 2012. https://www.tandfonline.com/doi/epdf/10.1198/0003130031964?needAccess=true&role=button.

Majid Nazeer, Muhammed Bilal. 2018. Evaluation of Ordinary Least Square (OLS) and Geographically Weighted Regression (GWR) for Water Quality Monitoring: A Case Study for the Estimation of Salinity. *link.springer.* [Online] 15 March 2018. https://link.springer.com/article/10.1007/s11802-018-3380-6.

Meg Miller, Mullai Manickavalli. 2021. ESRI Site License Integration using ArcGIS Online, Hub and Enterprise. *openjournals.uwaterloo.* [Online] 3 August 2021. https://openjournals.uwaterloo.ca/index.php/acmla/article/view/4253/5133.

Michael Law, Amy Collins. 2021. *Getting to Know ArcGIS Pro 2.8.* s.l. : Esri, 2021. p. 432.

Ortiz-Ospina, Max Roser and Esteban. 2018. Our World in Data. *Our World in Data.* [Online] September 20, 2018. https://ourworldindata.org/literacy?fbclid=IwAR06_PeRfWHkM0PcZopFd62LfbQj9COjAz-xqtF7ShAy-U8Jzfkjy85dSIM.

Unesco Institute for Statistics. Adult Education and Development 55/2000. *dvv-international.* [Online] https://www.dvv-international.de/en/adult-education-and-development/editions/aed-552000/dakar-education-for-all/statistical-document-education-for-all-ndash-2000-assessment.

8. Table Of The Figures and Abbreviations

8.1. Table Of The Figures

8.2. Table of Abbreviations

GIS	Geographical Information System
OLS	Ordinary Least Square Method
GWR	Geographically Weighted Regression
GNI	Gross National Income
HDI	Human Development Index
X	Independent Variable
Y	Dependent Variable
β	Regression Coefficients
ε	Error Term