NOVA IMS Information Management School

MEGI

Master's Degree Program in
Statistics and Information Management

Spatial Modeling of COVID-19 Vaccine Hesitancy in Texas

Maria Teresa França Gouveia Dinis Correia

Master Thesis
presented as partial requirement for obtaining the Master Degree in Statistics and Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Spatial Modelling of COVID-19 Vaccine Hesitancy in Texas

by
Maria Teresa França Gouveia Dinis Correia

Master Thesis presented as partial requirement for obtaining the Master's degree in Statistics and Information Management, with a specialization in Information Analysis and Management

Supervised by

Ana Cristina Costa, PhD, NOVA IMS, Universidade NOVA de Lisboa

July, 2024

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisbon, July 7th, 2024,
Maria Teresa Correia

ABSTRACT

This thesis investigates the spatial dynamics and factors influencing COVID-19 vaccination rates across the state of Texas. Throughout the search of data in several databases, this research identifies several significant variables such as gender, education level, employment, income, insurance coverage, race and ethnicity, age, party choice, religion and flu vaccine intake, that affect vaccine acceptance. Using spatial analysis and Geographical Information Systems we reach to conclusions about the variables above described. The study reveals distinct spatial patterns of vaccine hesitancy, particularly highlighting the differences between the several areas within the state (north, south, west and east) and offers insights of the impact of socioeconomic and demographic disparities. By employing Local and Global Moran's I statistics, hot spot analysis, and Multiscale Geographically Weighted Regression (MGWR), (using ArcGIS Pro) the research provides a nuanced understanding of the geographical variability in the COVID-19 vaccine intake, in the state of Texas both for year 2021 and 2022. The findings aim to inform authority figures of public health and politics in order to create better interventions and policies. This measures would help to enhance equity vaccine distribution and improve health communication strategies. Ultimately contributing to better management of future public health crises.

KEYWORDS

COVID-19; Vaccine; Hesitancy; Texas

Sustainable Development Goals (SDG):

TABLE OF CONTENTS

Statement of Integrity 1
Abstract 2
List of Figures 5
List of Tables...6
List of Abbreviations and Acronyms 7
1. Introduction 8
  1.1 Background and study relevance 8
  1.2 Problem statement and research gap 9
  1.3 Research questions and objectives 10
  1.4 Report structure 11
2. Literature review 13
  2.1 Vaccine hesitancy in other diseases 13
  2.2 Factors affecting covid-19 vaccine hesitancy 14
3. Methodology..24
  3.1 Study Region and Data 25
  3.2 Exploratory spatial data analysis 29
  3.3 MGWR model and diagnostics 31
4. Results of the empirical study for 2021 34
  4.1 Visualise the data 35
  4.2 Explore the distribution of the variables. 36
  4.3 Spatial patterns in covid-19 vaccination 37
    4.3.1 Spatial Autocorrelation 37
    4.3.2 Spatial non-stationarity 38
  4.4 Bivariate relationships 39
    4.4.1 Global Relationships 39
    4.4.2 Local Relationships 41
  4.5 Global Multicollinearity 43
  4.6 MGWR model 44
    4.6.1 Local Multicollinearity 45
    4.6.2 Influential Observations 46
    4.6.3 Goodness-of-fit of the final MGWR model 46
    4.6.4 Residual analysis of the final MGWR model 48
    4.6.5 Local coefficients of the final MGWR model 48
5. Results of the empirical study for 2022 50
  5.1 Visualise the data 50
  5.2 Explore the distribution of the variables. 51

5.3 Investigate spatial patterns in covid-19 vaccination in 2022 ... 52
5.3.1 Spatial Autocorrelation ... 52
5.3.2 Spatial non-stationarity ... 53
5.4 Analyse bivariate relationships ... 54
5.4.1 Global Relationships ... 54
5.4.2 Local Relationships ... 56
5.5 Global Multicollinearity ... 58
5.6 MGWR Model ... 59
5.6.1 Local multicollinearity ... 60
5.6.2 Influential Observations ... 60
5.6.3 Goodness-of-fit of the final MGWR model ... 61
5.6.4 Residual analysis of the final MGWR model ... 62
5.6.5 Local coefficients of the final MGWR model ... 62
6. Final Discussion ... 65
7. Conclusions and future works ... 69
Bibliographical References ... 72
Appendix A ... 78
Appendix B ... 82
Appendix C ... 90
Appendix D ... 91
Appendix E ... 93

LIST OF FIGURES

Figure 1 – Flowchart of the methodology 24
Figure 2 - Texas state and its 254 counties 26
Figure 3 – Incidence rate of the COVID-19 vaccination, in Texas for 2021 35
Figure 4 – Histogram chart of the percentage population that took the COVID-19 vaccine, for 2021 36
Figure 5 - Box plot charts of the distribution of the potential explanatory variables, for 2021 37
Figure 6 – Local Moran's I map of the percentage of county population that took the COVID-19 vaccine, for 2021 38
Figure 7 – Hot Spot Analysis of the percentage of county population that took the COVID-19 vaccine, for 2021 39
Figure 8 - Scatter and Pearson's correlation matrix, for 2021 40
Figure 9 – Local Multicollinearity, 2021 46
Figure 10 – Influential Observations, 2021 46
Figure 11 – Local R2 Map, 2021 47
Figure 12 - Standardized residuals of the MGWR model, 2021 48
Figure 13 - Incidence rate of the COVID-19 vaccination, in Texas for 2022 50
Figure 14 - Histogram chart of the percentage population that took the COVID-19 vaccine, for 2022 51
Figure 15 - Box plot charts of the distribution of the potential explanatory variables, for 2022 52
Figure 16 - Local Moran's I map of the percentage of county population that took the COVID-19 vaccine, for 2021 53
Figure 17 - Hot Spot Analysis of the percentage of county population that took the COVID-19 vaccine, for 2022 53
Figure 18 - Scatter and Pearson's correlation matrix, for 2022 55
Figure 19 - Local Multicollinearity, 2022 60
Figure 20 - Influential Observations, 2022 61
Figure 21 - Local R2 Map, 2022 62
Figure 22 - Standardized residuals of the MGWR model, 2022 62

LIST OF TABLES

Table 1 – Summary of studies addressing COVID-19 vaccine hesitancy 18
Table 2 – Description of independent variables included in the analysis 29
Table 3 - Name of the independent variables included in the analysis 34
Table 4 - Number of counties with significant relationships, for 2021.................................. 43
Table 5 – Global multicollinearity Results for 2021 data 44
Table 6 - Individual bandwidth and summary statistics of the local coefficients of each explanatory variable for 2021 45
Table 7 – Goodness-of-fit of the final MGWR model, 2021 47
Table 8 - Number of counties with significant relationships 57
Table 9 - Global multicollinearity Results for 2022 data 58
Table 10 - Individual bandwidth and summary statistics of the local coefficients of each explanatory variable for 2022 59
Table 11 - Goodness-of-fit of the final MGWR model, 2022.. 61

LIST OF ABBREVIATIONS AND ACRONYMS

AIC	Akaike's Information Criterion
AICc	Corrected Akaike's Information Criterion
COVID-19	COrona VIrus Disease 2019
GIS	Geographical Information Systems
GWR	Geographically Weighted Regression
MGWR	Multiscale Geographically Weighted Regression
MMR	Measles mumps and rubella
Pct_Asian	Percent of people that are Asians
Pct_Black_NH	Percent of people that are black, non-Hispanic
Pct_Children	Percent of people under 18years old
Pct_Democrat2020	Percent of people that vote for the democrat party
Pct_Elderly	Percent of people over 65years old
Pct_Female	Percent of females
Pct_FluVac	Percent of people that took the flu vaccine
Pct_HighSchool	Percent of people with high school degree
Pct_Hispanic	Percent of people that are Hispanic
Pct_IncomeGT200000	Percent of people with annual income higher than 200 000
Pct_IncomeLT10000	Percent of people with annual income lower than 10 000
Pct_Republican2020	Percent of people that vote for the republican party
Pct_SomeCollege	Percent of people with some college degree
Pct_Unemployment	Percent of people that do not have a job
Pct_Uninsured	Percent of people that don't have health insurance
Pct_White_NH	Percent of people that are white
Pct_Young	Percent of people between 19 - 39 years old
Religion	Percent of people that are religious
USA	United States of America
VPD	Vaccine Preventable Diseases
WHO	World Health Organization

1. INTRODUCTION

1.1 BACKGROUND AND STUDY RELEVANCE

Coronavirus (COVID-19) is an infectious disease created by SARS-CoV-2 virus, discovered in Wuhan in China, in December 2019. On March 11th of 2020, World Health Organization (WHO) (*WHO, COVID-19,* 2023) declared a pandemic due to the virus fast spread worldwide (Marfe et al., 2021). This pandemic was unlike anything we ever expected since its impact will be noticed, over society and global health, for the next decade. This virus disrupted not just people but also economic and social matters. Globally, the total amount of COVID-19 cases reached over 771 million as of October 22nd of 2023 (*WHO, COVID-19 cases,* 2023).

As the geographic dimension of the pandemic kept increasing, there was a time-sensitive need for the creation of public health interventions and approaches to mitigate this issue (Faisal et al., 2022). Among these strategies, the fast and efficient manufacture and distribution of vaccines was the main hope (Sadarangani et al., 2021). Previous research suggested that achieving group immunization by vaccinating a significative percentage of the population, should be the most effective strategy to control the effects of the pandemic (Fattah et al., 2022; Mollalo et al., 2021). An efficient vaccine against COVID-19 would decrease both mortality and morbidity, enabling a meaningful ease of social distance measures (Sadarangani et al., 2021). In 2020, several vaccines like Pfizer- BioNTech, Moderna, and Johnson & Johnson were developed and rapidly administered all over USA (Faisal et al., 2022; Mollalo et al., 2021). However, the administration of vaccines to prevent infection or disease, was not as easy as expected and relied on diverse obstacles (Sadarangani et al., 2021).

Vaccine hesitancy, defined by the WHO as the reluctance or refusal to vaccinate despite the availability of vaccines, emerged and was a significant and critical obstacle to control the pandemic (Adhikari et al., 2021; Malik et al., 2020; Mollalo et al., 2021). While only 5 to 10% of the population strongly opposes to vaccination, a larger and increasing percentage is considered as hesitant in receiving the vaccine (Dubé et al., 2013). It is a complex problem that depends on factors such as location, time and specific vaccine involved. Contains elements like confidence, complacency and convenience and threatens to set back the improvements made so far in talking vaccine-preventable diseases (*Vaccine hesitancy, WHO*, 2015). The WHO identified this problem as part of the principal ten threats to global health (WHO, 2019).

Previous research indicated that the leading reasons to vaccine hesitancy include many factors such as, media communications and misinformation (Kricorian et al., 2022; Scheufele et al., 2019) religion, historical influences, culture, gender (Agrawal et al., 2021), socioeconomic and sociodemographic factors (Litaker et al., 2022), geographic barriers, political influences, previous vaccination experiences, the design of the vaccination program perception of risk (Soares et al., 2021) , and the lack of trust in government or public health systems (Lee et al., 2022; Xavier et al., 2021).

Several studies, mostly from USA, try to explain spatial variations of COVID-19 vaccine hesitancy (Lee et al., 2022; Pallathadka et al., 2023). Texas is a particularly relevant state of the U.S. to focus on, since it is the second most populated state in the United States and one of the most demographically diverse (*US Ranked by Population 2023*, 2023). It contains a wide variety of communities, including urban centers, suburban areas and rural regions, each with its unique socio-economic and cultural factors. Texas also struggles with health disparities and varying accesses to healthcare services, which can also significantly impact vaccination rates (Xavier et al., 2021). Additionally, the state had a huge COVID-19 impact, reporting over 8.5 million cases of COVID-19 and over 92 thousand fatalities on the end of July 2023. By the end of May 2023, Texas registered that more than 18 million individuals (64%) have completed their full vaccination program against the virus. However, 22 million (77%) have incomplete vaccination which means they took at least one dose of the vaccine (*Texas Vaccination progress COVID-19*, 2023). Despite these numbers, there is still a significant percentage individuals left to receive the necessary immunization since Texas has a population of over 30 million people. This study will be focused on spatial analytics and spatial modelling to find what factors have a bigger influence of vaccination choices, in Texas.

1.2 PROBLEM STATEMENT AND RESEARCH GAP

Over the last years, spatial analysis of COVID-19 vaccine hesitancy gained significant attention since it has a strong potential for improving several disease prevention measures (Lee et al., 2022; Pallathadka et al., 2023). Nevertheless, there is still limited research on this topic. Understanding the factors that influence vaccine hesitancy can improve vaccination rates and contribute to the goal of controlling the virus spread and achieve group immunization faster (Truong et al., 2022).

There are only a few studies that have comprehensively assessed factors linked to COVID-19 vaccine hesitancy, (Pallathadka et al., 2023) and even fewer have employed Geographical Information Systems (GIS) and spatial analytics to comprehend, analyze and model the spatial patterns of vaccine hesitancy (Mollalo et al., 2021).

Studying COVID-19 vaccine hesitancy in Texas is important for several reasons. Texas offers a captivating background for exploring the spatial dynamics of vaccine hesitancy as well as the factors influencing them (Litaker et al., 2022). Despite the end of COVID19 pandemic, this state demographic heterogeneity can offer valuable lessons and strategies that could be applied in other regions. Furthermore, there is a large amount of data from Texas available in open access databases (e.g., demographic, socioeconomic, political), which allows the investigation of local aspects that might influence vaccine hesitancy in different locations of the state.

Finally, in public health research, the utilization of GIS emerges as an extremely helpful tool, offering a transformative way to evaluate and address the complexities of COVID-19 vaccine hesitancy (Kamel Boulos et al., 2020). GIS not only eases spatial modeling but also allows researchers to recognize patterns and associations (Gao et al., 2008). The geographic context becomes a crucial dimension, that allows comprehensive understanding of the interaction between demographic, socioeconomic, cultural factors in vaccine hesitancy. Disease reports demonstrated significant spatial dimensions, including locations of disease cases and patterns of the disease spread (Gao et al., 2008). Mapping these spatial patterns of diseases would help comprehend some issues associated with disease outbreak (Gao et al., 2008).

This study will be focused on the spatial dynamics and identification of factors linked to COVID-19 vaccine hesitancy in Texas. By using spatial analytics in GIS, we seek to reveal the complexity of vaccine hesitancy in Texas, offering a useful perspective for the creation of public health interventions (Ulugtekin et al., 2006).

1.3 Research questions and objectives

The main research questions of this study are:

- What are the spatial patterns of COVID-19 vaccine hesitancy in Texas?
- Which demographic, socioeconomic, political, or healthcare factors are associated with vaccine hesitancy in each county of Texas?

- How has COVID-19 vaccine hesitancy evolved spatially over time in Texas? Are there identifiable trends or changes linked to key interventions?

The aim of this thesis is to analyze and model local factors that induce to different spatial patterns of vaccine incidence rate in Texas. The specific objectives of the study are:

- Identify factors associated with vaccine incidence rate reported in the literature.
- Understand what kind of factors are linked to COVID-19 vaccine incidence rate in Texas, and if the association changes in different geographic areas.
- Explore how socioeconomic factors such as income, education, and access to healthcare insurance influence vaccine hesitancy in different geographic areas of Texas.
- Analyse spatial patterns of socioeconomic disparities and their association with vaccine acceptance rates across counties.
- Offer spatially informed recommendations for public health interventions and communication strategies to address vaccine hesitancy.

A methodological approach based on spatial analytics and spatial regression was addressed. We adopted a multifaced methodological approach to investigate the COVID-19 incidence rate in Texas. First, Local and Global Moran's I statistics and hot spot analysis was used to analyse the distribution of the vaccination rates and spatial analysis of data. Then, a Multiscale Geographically Weighted Regression (MGWR) was performed to explore the spatial correlations between vaccine incidence rate in Texas and the various explanatory variables at different geographic scales.

This study expects to help on the evolution of public health interventions, improve health communications, as well as policy decisions designed to tackle vaccine hesitancy, like reducing health disparities, ensuring equitable vaccine distribution, and safeguarding vulnerable populations from different socioeconomic and educational backgrounds (*OCDE, trust in COVID-19 vaccination*, 2021) (Xavier et al., 2021). Health education, including vaccine literacy are central aspects for success.

1.4 REPORT STRUCTURE

The structure of this thesis begins with the Introduction, which outlines the background and motivation for the study, emphasizes the significance of the study and displays the main research questions and objectives. Following this, the Literature Review presents an overview of existing research, establishes the theoretical framework and identifies key findings and gaps in the literature, summarizing the main articles and variables relevant to the study. The Methodology chapter details the study region and data, including a summary table of all variables, describes the data collection process, exploratory spatial data analysis and the MGWR model and diagnosis. The presentation of Results is divided into two chapters, the first one contains the results for the year of 2021 and the second one for the year of 2022. The Final Discussion chapter compares the results obtained for the two years and the thesis ends with a Conclusions and Future Recommendations chapter, summarizing the key findings and suggesting recommendations and directions for future situations.

2. LITERATURE REVIEW

The literature review was divided in two main parts: first, it focuses on vaccine hesitancy in other diseases, in order to comprehend the common patterns between them. Secondly, it addresses a more specific analysis for COVID-19, starting by evaluating the methodologies of other studies, to understand what are the main variables influencing COVID-19 vaccine hesitancy.

2.1 VACCINE HESITANCY IN OTHER DISEASES

Vaccination stands as one of the most significant accomplishments of public health (Dubé et al., 2013). However, vaccination impact depends a lot on the coverage rate, in order to reduce incidence of vaccine preventable diseases (VPD). Many experts believe that apprehensions regarding effectiveness and safety of vaccines are a risk to the vaccination programs (Dubé et al., 2013). Despite appearing in small percentage, a growing proportion of the population is considered to be hesitant regarding vaccination (Dubé et al., 2013).

Vaccine hesitancy is an individual behaviour shaped by a large number of factors, that include previous experiences and knowledge, but also historical, political, and socio-cultural influences. Trust in the systems that deliver vaccines, as well as in professionals' recommendations is also critical. Misinformation by the social media has a huge impact on vaccine hesitancy (Dubé et al., 2013).

Several studies explored hesitancy related to diseases like the Influenza, Smallpox, Measles, HPV and Polio (Aguolu et al., 2022; Gomes et al., 2022; Hotez et al., 2020; Soares et al., 2021; Ughasoro et al., 2015). Widespread of vaccination began in 1796 followed by Jenner's presentation that cowpox vaccine had potential to provide protection against smallpox. Despite the dramatic reduction of smallpox, many disapproved the use of this vaccine, including the important natural selection investigator, Alfred Russel Wallace (Dubé et al., 2014). The pertussis vaccine, controversy, is known as the match that led to the fire of active anti-vaccination groups that we still see nowadays (Baker, 2008). It started in UK, in the 1970s after a report publication, that stated that 36 children experienced severe neurological conditions after receiving pertussis vaccine (Kulenkampff et al., 1974). The study gathered interest of the media and set off widespread of public concerns, resulting in a concerning decrease on vaccine coverage, from 77% to 33%. As a result, three severe pertussis epidemics came after (Dubé et al., 2014). After 25 years Andrew Wakefield, a previous British surgeon, published a paper in 1998 (Wakefield et al., 1998) suggesting that MMR vaccine could cause autism(Marshall, 2019). A new break of global anti-vaccination belief spread worldwide. Measles, previously eliminated from UK in 1994, was making its way back to stage zero since parents were hesitant to give their children the measles-mumps-rubella (MMR) vaccine out of fear of its possible side effects (Dubé et al., 2014). Newspapers, magazines and websites had warnings of the existent risks of receiving vaccines (Dubé et al., 2014).

Individuals exhibiting vaccine hesitancy towards other diseases also display a higher probability of hesitancy towards the COVID-19 vaccine. Individuals who took other vaccines like Influenza, Hepatitis- B, or Polio and experienced short-term side effects, are more likely to be unwilling to get vaccinated (Alley et al., 2021; Biswas et al., 2021). This way, studies also discovered that those who did not take the flu vaccine were more likely to refuse or delay the COVID-19 vaccine (Gomes et al., 2022; Soares et al., 2021).

2.2 FACTORS AFFECTING COVID-19 VACCINE HESITANCY

There still exists a remarkable research gap in the literature concerning the application of advanced spatial analysis techniques to the research of vaccine hesitancy. This gap highlights the need for new studies to use innovative approaches to explain the correlation between vaccine hesitancy and its diverse influencing factors.

The dynamics of COVID-19 vaccine hesitancy were addressed with techniques such as Global and local Moran's I statistics and hot spot analysis (Mollalo et al., 2021; Pallathadka et al., 2023) to investigate spatial autocorrelation and to discover what areas had higher concentration of the phenomenon, respectively. He et al. (2023) explains the application of MGWR to COVID-19 incidence rates in the state of Arkansas, USA. The MGWR is a supplement of Geographically Weighted Regression (GWR) that permits the study of relationships at changing spatial scales. This happens because the regression allows the neighbourhoods around each spatial unit to vary between each predictor variable. This offers valuable insights of the spatial correlations between vaccine hesitancy and various explanatory variables at different geographic scales (He et al., 2023; Mollalo et al., 2021).

Several studies attempted to find the most relevant variables to explain COVID-19 vaccine hesitancy. As presented on **Table 1 – Summary of studies addressing COVID-19 vaccine hesitancy** these studies were diverse, from different geographic regions and variables changed with local needs and availability data.

Article	Region	Study Objective(s)	Research Method	Dependent/ response Variable (y)	Independent/ Explanatory Variable (X)
Mollalo et al., 2021	USA	1) Investigate the socioeconomic determinants of COVID-19 vaccination rates 2) Contribute to county-level policies and making informed public health decisions	1) Ordinary Least Squares Model (OLS) 2) Geographically Weighted Regression (GWR) 3) Multiscale Geographically Weighted Regression (MGWR)	1) COVID-19 Vaccination rates	Below poverty %; Unemployment rate %; Per capita income; No high school diploma %; Age 65 and older %; Age 17 and younger %; non-institutionalized with a disability %; Single-parent households with children %; Minority (except white, non-Hispanic) %; Age 5+ who speak limited English %; Housing in structures with 10+ units %; Mobile homes %; Over-occupied housing units %; Households with no vehicle available %; Institutionalized group quarters %; Uninsured people % 17) Population density per square mile
Pallathadka et al., 2023	USA	1) Identify spatially varying social, ecological, and technological factors that are associated with COVID-19 vaccination rates	1) Hotspot analysis 2) Spatial regression model 3) Geographically weighted regression (GWR) 4) GIS analysis and mapping	1) COVID-19 Vaccination rate at county level	Voters who voted for Republican; Presidential Nominee in 2020 (−); % Population Aged 65 and Over (+); % Population Change (2010–2020) (+); % Population (Aged 18 or Over) with; Bachelor's Degree or Higher (+); % Minority by each racial group (−);% Female Population of Fertility Age (15–44) Education (−); Median Household Income (+); National Risk Index (+); % Farmland (−); % Population with Broadband Access (+); Health Facilities (+); % Impervious Surface (+); Factories (−)
Lee et al., 2022	Texas	**1)** Investigate the spatial dimension of socioeconomic and demographic factors behind COVID-19 vaccine hesitancy **2)** Contribute to a better understanding of the drivers of vaccine acceptance **3)** Identify the social vulnerability characteristics and political views that influence vaccine hesitancy. **4)** Provide insights that can help public health officials and policymakers	1) Ordinary least squares (OLS) regressions are initially used to analyze the census tract data. 2) Explore Spatial patterns in vaccinations using the spatial autoregressive (SAR) approach, specifically the spatial Durbin model.	1) Share of the unvaccinated population in each census tract	1) Unvaccinated Pop. (%) 2) Poverty (%) 3) Unemployment (%) 4) Per Capita Income ($) 5) No High School (%) 6) Age 65 and older (%) 7) Age 17 and younger (%) 8) Disabilities (%) 9) Single parents (%) 10) Minorities (%) 11) Limited English (%) 12) No vehicle (%) 13) Biden voters (%) 14) Pop. Density

Article	Region	Study Objective(s)	Research Method	Dependent/ response Variable (y)	Independent/ Explanatory Variable (X)
Litaker et al., 2022	Central Texas	1) Identify sociodemographic factors associated with vaccine hesitancy among a population of individuals with health insurance	1) Online survey with sociodemographic and vaccine-related questions 2) Logistic regression analysis	1) Vaccine hesitancy in central Texas	1) age 2) gender 3) race 4) ethnicity 5) level of educational attainment 5) income
Malik et al., 2020	USA	1) Predict COVID-19 vaccine acceptance 2) Identify the most vulnerable population 3) Provide information for public health officials and politicians	1) Statistical analysis Descriptive statistics 2) Frequency and percentage of COVID-19 vaccine acceptance 3) Chi-square analysis	1) Acceptance of COVID-19 vaccine	1) age 2) gender 3) race 4) education 5) ethnicity 6) employment status
Liu et al., 2021	USA	1) How COVID-19 vaccine hesitancy, varies across sociodemographic groups? 2) How do different sociodemographic characteristics intersect to affect COVID-19 vaccine hesitancy?	1) Descriptive analyses 2) multi-level mixed-effects logistic regression models with state-level random intercepts effects	1) overall vaccine hesitancy	Gender; Race (White; Asian; Black; Hispanic; Other/Mixed); Education: (Less than High School; High School; Associate or Some College; Bachelor; Graduate); Age, Married; Number of Household Adults; Number of Household Children ; Had COVID-19
Soares et al., 2021	Portugal	1) Assess and identify factors associated with COVID- 19 vaccine hesitancy in Portugal.	1) Data from a community-based survey 2) Multinomial regression to identify factors associated with intention to delay or refuse to take COVID- 19 vaccines. (All statistical analyses were performed using R)	1) Intention to vaccinate	Gender; Age; Education; Monthly household income; Partial or total income loss during the pandemic; Occupation; Intention to take the flu vaccine; Perception of the health status; Number of comorbidities; Self-reported diabetes; Self-reported respiratory disease; Self-reported autoimmune disease; Having school-age children; Confidence in the capacity of health services to respond to the pandemic; View on the information provided by health authorities; Perception of the adequacy of measures implemented by the government; Self-perceived risk to get COVID-19 infection; Self-perceived risk to develop severe disease following COVID-19 infection; Frequency of agitation, sadness, or anxiety due to the physical distancing measures; Confidence in the efficacy and safety of COVID-19 vaccines being developed; Period of the questionnaire

Article	Region	Study Objective(s)	Research Method	Dependent/ response Variable (y)	Independent/ Explanatory Variable (X)
Gomes et al., 2022	Portugal	1) Assess and identify factors associated with COVID-19 vaccine hesitancy in Portugal, nine months after the rollout of the country's vaccination program, in a context of high vaccination coverage.	1) Variables described using absolute and relative frequencies. 2) Logistic regression models were fitted. 3) Estimated a crude odds-ratio (OR) 4) Sensitivity analysis (All statistical analyses were performed using R)	1) Binary: vaccinated or hesitant	Gender; Age; Education; Monthly household income; Partial or total income loss during the pandemic; Occupation; Month of the questionnaire; Intention to take the flu vaccine; Perception of the health status; Number of comorbidities; Having school-age children; Frequency of agitation, sadness, or anxiety; Confidence in the health services response to COVID-19; Confidence in the health services response to non-COVID-19; Perception of the adequacy of measures implemented by the Government; Self-perceived risk of getting COVID-19 infection; Self-perceived risk of developing severe disease following COVID-19 infection; Confidence in the safety of the COVID-19 vaccines; Confidence in the efficacy of the COVID-19 vaccines
Biswas et al., 2021	Worldwide	1) Explore the scientific literature and find the determinants for worldwide COVID-19 vaccine hesitancy as reported in the literature.	1) PRISMA Extension for Scoping Reviews (PRISMA-ScR) guidelines were followed in conducting this review.	1) Overall vaccine hesitancy	age, gender, level of education, profession, ethnicity, population size, and monthly income; (i) demographic factors (ii) environmental factors and (iii) vaccine-specific factors: Vaccine safety and efficacy, Vaccine side effects, Individuals believe that they are at less risk to get infected by COVID-19, Religious beliefs, Price of vaccine and lack of insurance, Mistrust in healthcare; Mistrust in government; The rapid development of a vaccine; Widespread misinformation in the social media; Past vaccine experience; Demographic influence; Political instability; Racist and ethnic minority; Trust in the vaccine manufacturer, Lockdown periods decrease the number of cases; Trust in natural remedies; Lack of information about vaccine; Inconsistent risk message from public health organization; Anti-vaccination movement

Table 1 – Summary of studies addressing COVID-19 vaccine hesitancy

Age

Age is an important variable to consider (Troiano et al., 2021). Policies and decisions made could be specifically designed to tackle the most problematic age group.

Desire to get vaccinated is linked to the perception of the probability of being infected with the virus (Karlsson et al., 2021). Younger people are at a lower threat of critical COVID-19 infection (Guidry et al., 2021; Karlsson et al., 2021). On the contrary, older and immunosuppressed patients, had more concerns about COVID-19 vaccination safety and its side effects (Kourlaba et al., 2021; Palamenghi et al., 2020). Since children had milder disease, there were concerns about safety, such as myocarditis, (*Pfizer COVID-19*, 2021) which made several parents uncertain about vaccinating their children.

Despite these findings, most of those studies concluded that older people, nevertheless, have a higher probability to accept the vaccine as soon as they can due to age and health worries (Chakraborty et al., 2021; Soares et al., 2021). Consequently, younger individuals have higher vaccine delay or refusal (Mollalo et al., 2021; Soares et al., 2021). As for females of fertility age, they are also less likely to take the vaccine due reproductive and health concerns (Markert et al., 2021). Finally, Reich (2020) also concluded that parents that have a greater social privilege have a higher probability to refuse vaccination for their children.

Gender

Previous studies show significant gender differences in vaccine intake choices (Soares et al., 2021). Liu et al. (2021) found that vaccine hesitancy among women was mostly due to circumspection. Initially, women, are more probable be hesitant in vaccination when compared to men, but this percentage also declined faster, as more information was given over tine (Liu et al., 2021). Additionally, in the most recent data, Liu et al. (2021) found a greater proportion of men that are hesitant due to confidence and complacency when compared to women.

Overall, most studies showed that females have a higher likelihood to have greater vaccine hesitancy percentages than man (Malik et al., 2020; Soares et al., 2021).This was due to of safety concerns regarding pregnancy, the potential negative outcomes for infants born to mothers who received vaccination and childbirth (Litaker et al., 2022; Markert et al., 2021).

Education

Several studies found that populations or individuals with college degree were less prone to vaccine hesitancy compared to those with no degree or less than college degree (Gomes et al., 2022; Malik et al., 2020; Pallathadka et al., 2023). Controversially (Liu et al., 2021) found that individuals holding a high school diploma, or its equivalent, have a higher likelihood of vaccine hesitancy than to those without it (Liu et al., 2021). Overall, the studies indicated that those that had less education are more prone to refuse or delay the intake of COVID-19 vaccines (Lee et al., 2022; Litaker et al., 2022; Liu et al., 2021; Malik et al., 2020; Reiter et al., 2020). However, this relationship can vary geographically (Pallathadka et al., 2023).

These results support that a more educated population has higher probability to take the vaccines since they have higher perception of awareness (Ehde et al., 2021). Additionally**,** higher educated people are more preoccupied with the chances of actually getting COVID-19 (Ehde et al., 2021). Additionally, they also have access to several sources of information (such as health agencies or social media personal networks), which are fundamental to get informed and foresee vaccine acceptance (Wang et al., 2020). Educationalists often get involved in protecting and promoting optimistic awareness concerning vaccination (Biswas et al., 2021). Given the fact that many students are members of those groups, it is important that public health organizations increase their attempts to target these individuals to have higher COVID-19 vaccine coverage (Karlsson et al., 2021).

Income

In general, people with higher income had a lower probability of vaccine hesitancy (Gomes et al., 2022; Litaker et al., 2022; Mollalo et al., 2021; Pallathadka et al., 2023). Indeed, wealthier individuals had a more awareness of the negative consequences of getting the virus (Ehde et al., 2021), (Yang et al., 2021) and were also more familiar with the effectiveness of the vaccine (Williams et al., 2021). On the other hand, individuals with lower monthly incomes were less disposed to receive the vaccines because of the poor health insurance and the cost of the vaccines (Callaghan et al., 2021). According to (Lee et al., 2022) individuals living in economically disadvantage neighborhoods with lower income and lower education achievements were less likely to receive the vaccines. The same way, individuals who suffered income loss during the pandemic or were unemployed also had a higher probability to refuse or delay the vaccine intake (Malik et al., 2020; Soares et al., 2021).

Offering affordable vaccine prices and exempting health insurance increased the probability higher vaccination rates. These results suggest that communities with lower income, that are also disproportionately affected by COVID-19, might have higher vulnerability to frequent outbreaks, even there is a vaccine available (Malik et al., 2020).

Political choice

Regarding state partisanship, several studies indicate that U.S. states that have a higher proportion of Republican voters are more likely to refuse or delay the vaccine intake (Lee et al., 2022; Liu et al., 2021). A recent study revealed that almost 40% of Republicans were not vaccinated, in contrast to only 10% of Democrats (Poland et al., 2009). This happens due to misconceptions about vaccines (Albrecht, 2022). Numerous studies demonstrate that misinformation spread by politicians significantly contributes to the spatial patterns of vaccine hesitancy (Allcott et al., 2020). In fact, (Pink et al., 2021) discovered that Republicans that did not take the vaccine were more willing to accept it if they saw support from an important political figure. The higher likelihood vaccine hesitancy among republicans in general, is explained by this factor.

Race and ethnicity

The variable race holds significant importance in the study of vaccine hesitancy, as evidenced by multiple studies. (Litaker et al., 2022) found that individuals that are Black or African American are 65% less likely to accept the COVID-19 vaccine when compared to the other races. Other study, performed by (Barello et al., 2020) demonstrated that the vaccine intake rate between the White ethnic group was 84.5%, and 14.9% among the Black minority. Hence, in general, Black population is negatively correlated with vaccination rates (Pallathadka et al., 2023).

Vaccine hesitancy among ethnic minorities emerged form a combination of history of racial injustices, absence of healthcare availability and misinformation (Hildreth et al., 2021) (DiRago et al., 2022). It is crucial to acknowledge that this problem differs across various ethnic groups, considering their racialized and minoritized communities, each characterized by their unique cultures and social norms (Williams et al., 2021). Notably, hesitancy across the Asian individuals consistently remains the smallest percentages and decreased very little across time (Liu et al., 2021). On the other hand, blacks and Other/Mixed groups exhibit the highest levels of vaccine hesitancy due to confidence and circumspection and lower levels due to complacency (Liu et al., 2021). Furthermore, in the majority of the states, white individuals continue to exhibit a higher vaccination compared to Hispanic and Black people (Nambi Ndugga, 2022). Individuals that categorize as ethnic minorities often tend to have less confidence in government institutions due to racism such as discrimination in education, employment, housing and political injustices (Hotez, 2020). The results align with previous research indicating that African Americans exhibit lack of government trust due to historical instances of medical abuse and discrimination. (Jamison et al., 2019).

Furthermore, Liu et al. 2021, found that even in predominant Democrat states, Black individuals tended to display great levels of vaccine hesitancy compared to the other racial groups. Controversially, higher levels of hesitancy were observed among Whites and Other/ Mixed group on mainly Republican states (Liu et al., 2021). These findings underscore the need for aimed policies and strategies to enhance vaccination percentages based on community and local contexts (Liu et al., 2021).

Neighbourhoods with a higher concentration of Hispanic residents generally exhibit a greater vaccine intake than those with higher non-Hispanic White individual share (Lee et al., 2022). Problems concerning racial and spatial inequality in public health are still a concern (Pallathadka et al., 2022). This way, there is an urgent need for specific policies and strategies aimed to enhance vaccination rates according to community and local contexts (Liu et al., 2021).

Mistrust in healthcare services/government or Mistrust in COVID-19 vaccine

Reports have demonstrated that due to the level of mistrust in government/healthcare services, several individuals were worried with the evidence provided by its institutions, which resulted in increased vaccine refusal (Park et al., 2021; Pogue et al., 2020). This way, those who considered measures implemented by the government/health services as inadequate, inconsistent and contradictory had higher probability to reject vaccines (Soares et al., 2021). In addition, those who perceived COVID-19 vaccines as unsafe and ineffective also exhibited a larger probability of vaccine refusal (Gomes et al., 2022).

Misinformation

Social media bares a powerful influence on our lives. Many people rely on social media instead of traditional sources such as newspaper or TV news to find vaccine information (Biswas et al., 2021). Regrettably, some individuals firmly believe and share information the collect on social media within their social circles without verifying its accuracy. (Biswas et al., 2021). Groups opposed to vaccination actively use social media and spread misinformation, which ends up influencing individual's disposition to get vaccinated (Alley et al., 2021; Sallam, 2021).

Previous studies concluded that widespread misinformation communicated via social media contributed to the decline in vaccine acceptance rates (Barello et al., 2020). Also, when individuals are submitted to unclear communication of the risks or benefits the vaccines, they are more likely to refuse or delay its intake (Allcott et al., 2020; Gomes et al., 2022).

With the increase of misinformation about COVID-19 trough several media channels, it is becoming crucial for USA public health officials and politicians to strategize for effective communication and policy making (Malik et al., 2020). Health communication needs to extend to all communities, with a particular focus on the most vulnerable, to provide individuals with information about vaccine safety and prevent future infections and causalities (Malik et al., 2020).

Vaccine side effects

The COVID-19 vaccine is rather recent, was approved faster and most of the vaccine's short-term and long-term side effects were unknow (Wang et al., 2020). Indeed, some users refused to receive the vaccine due to the fear of the treatment possible side effects, as reported in (Guidry et al., 2021; Sherman et al., 2021; Williams et al., 2021; Yang et al., 2021). Thus, establishing trust among the general population and individuals with negative past experiences with medical figures and government agencies stands as an unprecedent challenge for public health authorities (Ruiz et al., 2021).

Religious beliefs

The struggle between vaccination and religious individuals arises from conspiracy theories centered on morality. As a result, certain religious individuals may choose to reject or postpone vaccination (Bertin et al., 2020).

3. METHODOLOGY

In order to address the research questions a methodological approach based on spatial analytics and spatial regression, was applied. The methodology involved mapping the study area, collecting and processing data, choosing and addressing the influence that specific variables have within that area. **Figure 1 – Flowchart of the methodology** shows a flowchart representing our overall methodology steps.

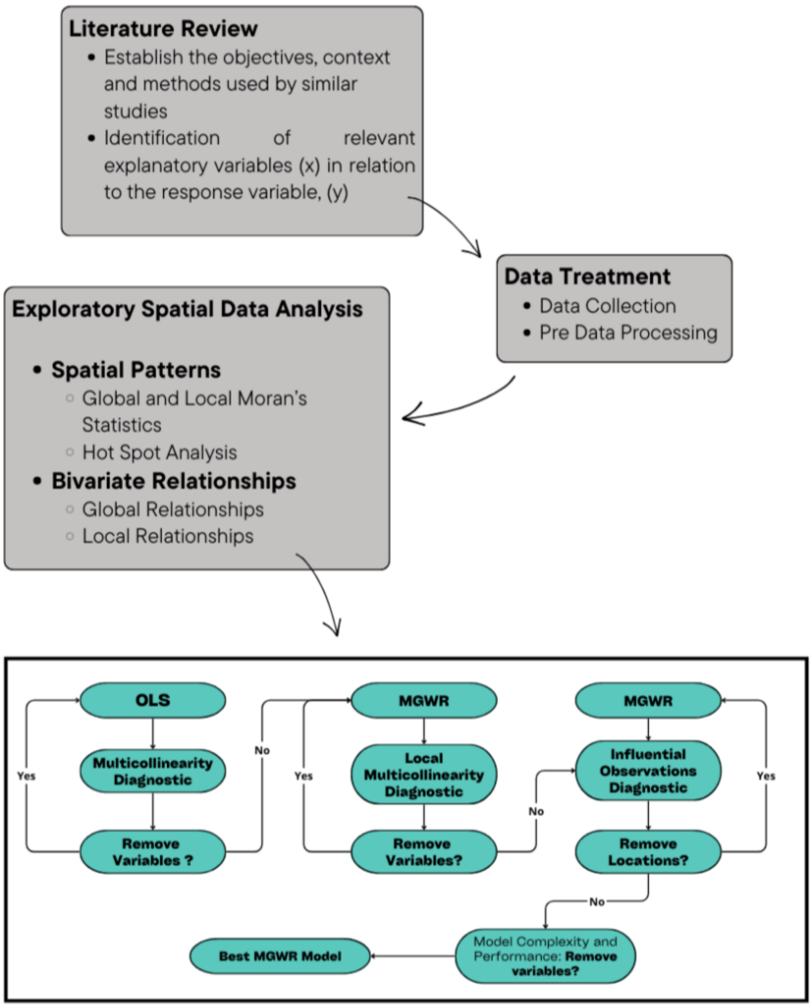


Figure 1 – Flowchart of the methodology

Initially we conducted a literature review, leading to the identification of relevant factors associated with the vaccine incidence rate (dependent variable). With the defined factors, data were collected, and then pre-processed to harmonize the data from different sources into a single table containing all the variables of analysis within the Texas counties. Additionally, a spatial exploratory analysis was conducted to characterize distributions, relationships, and potential spatial heterogeneity, helping data visualization, and allowing for easy interpretation of spatial patterns. To achieve this, the techniques employed include, Local and Global Moran's I Statistics (Anselin, 1995; Moran, 1950) and Hot Spot Analysis (Getis et al., 1992). Global relationships between vaccine hesitancy and potential explanatory variables were assessed with scatterplots and the Pearson's correlation coefficient. Local relationships were investigated with a local entropy statistic (Guo, 2010). Finally, using local spatial multiple regressions, specifically the Multiscale Geographically Weighted Regression (MGWR) (Fotheringham et al., 2017) it was possible to investigate how the relationship between various variables and vaccine incidence varied in space. This allowed us to identify the most relevant factors in locations where vaccination intake is higher. Subsequently, in areas where vaccine acceptance is lower, the explanatory factors contributing to the decrease efficiency in that region can be examined. Several diagnostic techniques were applied to MGWR models with different sets of independent variables as detailed in **Table 3 - Name of the independent variables included in the analysis**. The final MGWR model results were then compared with Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) (Brunsdon et al., 1996)(Brunsdon et al., 2002) models with the same final set of explanatory variables.

A 5% significance level was used in all statistical tests, otherwise stated.

We will start this chapter with an overview of our study region and data. Afterwards we proceed to a chapter of exploratory spatial data analysis, where we theoretically explain the tests that we will perform. Finally, the chapter ends with a chapter of the MGWR model and diagnostics.

3.1 Study Region and Data

Texas is the second largest, the second most populous state of the USA and tops the list of states with the most counties, with 254 counties. Geographically, Texas is in the South-Central United States of America. It is considered to be part of the South and Southwest of USA, as shown in **Figure 2 - Texas state and its 254 counties**.

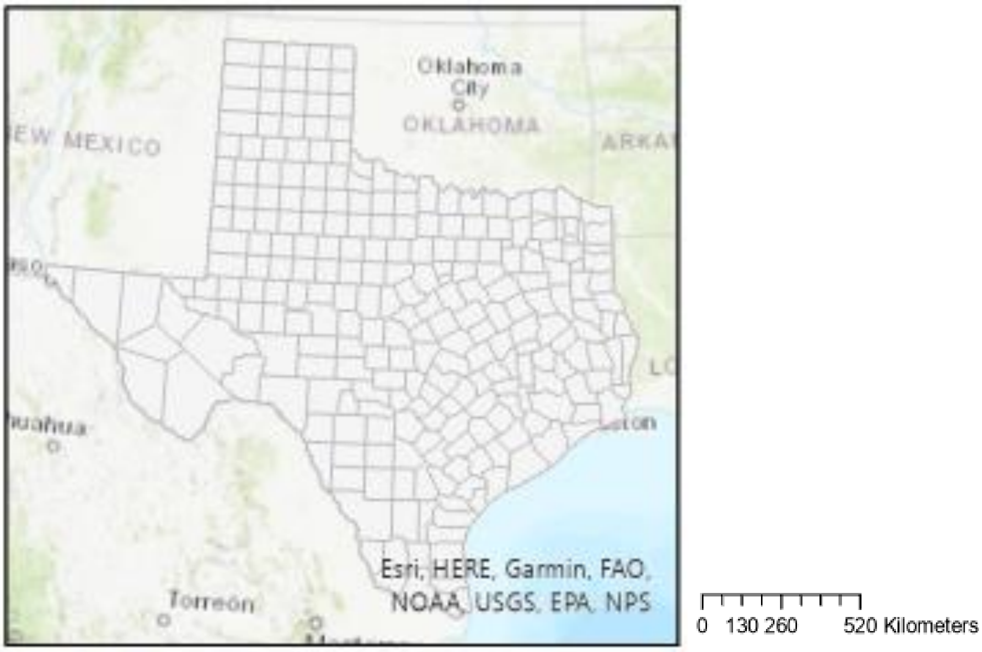


Figure 2 - Texas state and its 254 counties

The state's area covers a total of 268,581 square miles (695,622 km2) and its coastline has a total of 367 miles (591 km). The geographic center of Texas is about 15 miles (24 km) northeast of Brady in northern McCulloch County. It has five state forests and 120 state parks and 15 major river systems. The state was selected for this study since it allows for a comprehensive analysis of spatial patterns and relationships. The decision to concentrate on Texas as the study region is based on its size, diversity, and socio-economic dynamism, providing a strong backdrop for our research objectives. Lee et al. (2022) also considered this study region to investigate socioeconomic and demographic factors behind COVID-19 vaccine hesitancy, but they used a different approach based on the classical global regression model by OLS and spatial SAR models (**Table 1 – Summary of studies addressing COVID-19 vaccine hesitancy**).

Spatial data are data that have a geographical component. Every spatial coordinate from the study will be based on Texas counties' centroids. The input data used for this study was collected from different sources (**Table 2 – Description of independent variables included in the analysis**). The dependent (or response) variable (y) taken under consideration was the percentage of individuals that were vaccinated in the years of 2021 and 2022, and data collection from the Centres for disease control and prevention (CDC). The demographic variables were collected from the USA census for the years 2022 and 2021. This includes Age and Sex, Income, Employment Status, Education levels and lastly Race and Ethnicity. As for the variables such as religion beliefs and political choice the data was obtained from US Religion census and MIT Election Data, respectively. For the political choice, only data from 2020 was collected since it was not possible to obtain data for other years. We assumed that the percentage of voters of the republican and democrat party in 2021 were identical to those of 2020. Since this assumption could not be very realistic for the model of 2022, the variables of the category political choice were not considered.

It is important to notice that, since there are several different religions believes, we collected the 2020 observations from the U.S. Religion Census (Grammich et al., 2023), so we considered the following variable: **Adherents as % of Population**. The definition of adherents varies from one religion to another but can be generally described as the number of reported members of religious groups (for details see,(Grammich et al., 2023)). We only collected 2020 data because this type of data was not available for other periods. Hence, we assume that the percentage of adherents to a religion is similar in 2021 and 2022.

In this study we used the software ArcGIS Pro version 3.2.2, since it storages the data from several sources and evaluates them using algorithms and techniques of spatial data. The utilization of GIS software is crucial for investigating and understanding the spatial dynamics of COVID-19 vaccine hesitancy in Texas. As a powerful GIS software, ArcGIS Pro allows for the integration and analysis of several spatial data layers, being a useful platform for mapping, analyzing and modelling vaccine hesitancy patterns.

Category	Variable	Source	Justification	References
Age	**Young:** 18-39 years / Younger Individuals (-)	US Census Bureau	Younger individuals have higher vaccine delay or refusal	(Mollalo & Tatar, 2021; Soares et al., 2021)
	Elderly: Older Individuals (> 65) (+)	US Census Bureau	Older individuals are more likely to take the vaccine as soon as possible	(Chakraborty et al., 2021; Mollalo & Tatar, 2021; Soares et al., 2021)
	Children: < 18 Years / Children (-)	US Census Bureau	Parents are more likely to decline vaccines for their children	(Reich, 2020)
Gender	**Gender:** Female (-)	US Census Bureau	Females are more likely to have higher vaccine hesitancy rates when compared to man	(Malik et al., 2020)

Category	Variable	Source	Justification	References
Education	**Some College:** College or associate degree (+)	Texas County Health Rankings & Roadmaps	Populations or individuals with college degree were less likely to have vaccine hesitancy	(Gomes et al., 2022; Pallathadka et al., 2023)
	Hight school: High school diploma (-)	Texas County Health Rankings & Roadmaps	Those with a high school diploma or equivalent are more likely to have vaccine hesitancy	(Liu et al., 2021; Mollalo & Tatar, 2021)
Socioeconomic	**Income > 200 000:** Higher than 200 000 (+)	US Census Bureau	People with higher income had a lower probability of vaccine hesitancy	(Gomes et al., 2022; Litaker et al., 2022; Mollalo & Tatar, 2021; Pallathadka et al., 2023)
	Income < 10 000: lower than 10 000 (-)	US Census Bureau	Those with low monthly incomes were less willing to get vaccinated	(Callaghan et al., 2021; Mollalo & Tatar, 2021)
	Employment: Unemployed (-)	Texas County Health Rankings & Roadmaps	Those unemployed had a higher probability to refuse or delay the vaccine intake	Malik et al., 2020; Mollalo & Tatar, 2021; Soares et al., 2021)
Race and Ethnicity	**Black:** Black or African American (-)	Texas County Health Rankings & Roadmaps	Black or African American are 65% less likely to receive the COVID-19 vaccine	(Litaker et al., 2022; Pallathadka et al., 2023)
	White: White ethnic group (+)	Texas County Health Rankings & Roadmaps	Vaccine acceptance rate among the White ethnic group was 84.5%	(Barello et al., 2020)
	Asian: Asians (+)	Texas County Health Rankings & Roadmaps	Hesitancy among Asians consistently remains the lowest	(Liu et al., 2021)
	Hispanic: Hispanic or Latino (+)	Texas County Health Rankings & Roadmaps	Hispanic residents generally exhibit a greater vaccine acceptance	(Lee et al., 2022)
Religious beliefs	**Religious:** Adherents as % of Population (-)	US Religion Census	Religious people tend to reject or delay vaccination	(Bertin et al., 2020)
Political choice	**Democrat:** Democrat (+)	MIT Election Data and Science Lab	Almost 40% of Republicans were not vaccinated, in contrast to only 10% of Democrats	(Poland et al., 2009)

Category	Variable	Source	Justification	References
	Republican: Republican (-)	MIT Election Data and Science Lab	States that have a higher proportion of Republican voters are more likely to refuse or delay the vaccine intake	(Lee et al., 2022; Liu et al., 2021)
Health	**Flu Vaccination**: Flu Vaccinations (+)	Texas County Health Rankings & Roadmaps	Those who did not take the flu vaccine were more likely to refuse or delay the COVID-19 vaccine.	(Gomes et al., 2022; Soares et al., 2021).
	Uninsured: No Health Insurance Coverage (-)	Texas County Health Rankings & Roadmaps	Those with poor health insurance were less disposed to receive the vaccines	(Callaghan et al., 2021; Mollalo & Tatar, 2021)

Table 2 – Description of independent variables included in the analysis

3.2 EXPLORATORY SPATIAL DATA ANALYSIS

This section describes how to assess spatial autocorrelation and spatial heterogeneity in the percentage of individuals that were not vaccinated using the Global Moran's I, Anselin Local Moran's I, and Hot spot analysis (Getis-Ord Gi* statistic).

Spatial autocorrelation is a statistical concept that explores the spatial patterns of similarity or dissimilarity in the values of a variable across different locations. If there is spatial autocorrelation, the closest locations have a higher degree of similarity in terms of the measured value when compared to what would be anticipated on random chance. The spatial autocorrelation degree is measured by the Moran's Index (I). It falls between -1 and 1. Similar values tend to cluster together in space when there is high spatial autocorrelation, shown by a positive result of the Moran's I test. The expected index represents the value of Moran's I under the assumption of spatial randomness. Which means the expected value if there were no spatial autocorrelation. Variance measures the degree of spread or dispersion of the data around the mean. The z-score expresses in standard deviations the difference between the observed and expected Moran's I.

Positive spatial autocorrelation occurs when similar values cluster together in space. Conversely, negative spatial autocorrelation occurs when different values are found near each other (high values might be surrounded by low values and vice versa). Understanding spatial autocorrelation helps researchers and analysts make more informed decisions and predictions about the spatial distribution of variables.

The Local Moran's I statistic and hot spot analysis are commonly used to evaluate non-stationarity trough local analysis of spatial autocorrelation. The Local Moran's I statistic that was developed by (Anselin, 1995) uses the randomization distribution to test the null hypothesis of no local autocorrelation (Acharya et al., 2018). The main applications of these local spatial statistics are the detection of univariate spatial clusters (concentration of high or low values, for example) and the investigation of the geographical variability of such clusters. However, current univariate and bivariate local statistics rely on a measure that is intended to capture a particular kind of link (such as high-high and low-low associations), which in the meantime will overlook or fail to capture other kinds of correlations (Guo, 2010).

The Local Moran's I statistic evaluates the interdependence of a variable with itself at different geographic locations. It can be used to identify spatial outliers (i.e., atypical clusters in data) that correspond to negative spatial autocorrelation, and spatial clusters (i.e., clusters of similar values) corresponding to positive spatial autocorrelation. Hot spots and cold spots, or clusters of low and high values, indicate locations where the variable does not differ significantly from the surrounding areas. These spatial patterns are strong when significant clusters at different confidence levels (e.g., 90%, 95%, 99%) are present.

Examining the connection between pairs of variables throughout the entire dataset, as opposed to concentrating on local neighbourhoods, is known as a global bivariate relationship analysis. This link can be shown visually with a matrix scatterplot map, which shows scatterplots for pairs of variables in a matrix format. This can help inform further analysis and provide insights into factors influencing COVID-19 vaccine hesitancy in Texas. Moreover, global bivariate relationships were assessed using the Pearson's correlation coefficient.

The examination of local bivariate relationships dives deeper into the shades of the associations between the percentage of vaccinated population and potential explanatory variables. This involves analysing specific conditions within the dataset to identify variations in the direction of the relationships. If one variable depends on the other, then there is a "good" relationship between the two. This means, a positive correlation indicates that it is possible to estimate the value of one variable given the value of the other (Guo, 2010). These results of local statistical tests based on entropy measure allow to create maps with the significance of different types of local relationships. These maps are very important to choose variables for the model. One variable could not be relevant globally but relevant locally. If a linear relationship is not observed, we should not use that independent variable in the linear model. The relationships identified by this test can be classified as:

Not Significant—The relationship between the variables is not statistically significant.

Positive Linear—The dependent variable increases linearly as the explanatory variable increases.

Negative Linear—The dependent variable decreases linearly as the explanatory variable increases.

Concave—The dependent variable changes by a concave curve as the explanatory variable increases.

Convex—The dependent variable changes by a convex curve as the explanatory variable increases.

Undefined Complex—The variables are significantly related, but the type of relationship cannot be reliably described by any other category.

3.3 MGWR MODEL AND DIAGNOSTICS

The Multiscale Geographically Weighted Regression (MGWR) is an advanced spatial regression technique. It has evolved from the Geographically Weighted Regression (GWR) model, which utilizes explanatory and dependent variables' values in the neighbourhood of a target feature to create a local linear regression model for interpretation or prediction. MGWR captures variations in relationships at different scales whereas GWR operates at a single scale. This is because GWR requires specifying a single bandwidth that determines the extent of spatial influence of each predictor for each target feature (i.e., county), whereas MGWR uses a different bandwidth (i.e., neighbourhood) for each predictor.

The equation for the GWR model is close to the OLS model except that now it includes location, and a value for each coefficient is assigned to a location (u, v) in space (e.g., u is longitude and v is latitude) and we denote it as $\beta j(u, v)$. In MGWR, since each predictor can have its own bandwidth, we use the bw subscript in $\beta bwj(u, v)$ to denote it. The MGWR model was used to estimate the local relationships between COVID-19 vaccine incidence rate in 2021 and eight county-level characteristics:

$$Y_i = \beta_{bw0}(u_i, v_i) + \beta_{bw1}(u_i, v_i)X_{1i} + \beta_{bw2}(u_i, v_i)X_{2i} + \beta_{bw3}(u_i, v_i)X_{3i} + \beta_{bw4}(u_i, v_i)X_{4i} + \beta_{bw5}(u_i, v_i)X_{5i} + \beta_{bw6}(u_i, v_i)X_{6i} + \beta_{bw7}(u_i, v_i)X_{7i} + \beta_{bw8}(u_i, v_i)X_{8i} + \varepsilon_i,$$

$$i=1, ..., 254$$

where:

• Y accounts for the COVID-19 vaccine incidence rate (FullVac_pct)
• X1 accunts for the the percentage of Uninsured population (Pct_Uninsured),
• X2 accunts for the the percentage of population with some college degree (Pct_SomeCollege),
• X3 accunts for the the percentage of population with the Flu vaccine (Pct_FluVac),
• X4 accunts for the the percentage of population with annual income greater than 200000(Pct_IncomeGT200000),
• X5 accunts for the the percentage of Asian population (Pct_Asian),
• X6 accunts for the the percentage of Hispanic population (Pct_Hispanic),
• X7 accunts for the the percentage of White non Hispanic population (Pct_White_NH),
• X8 accunts for the the percentage of Democrat population (Pct_Democrat2020),
• (u_i, v_i) are the coordinates of the centroid of each county,
• $\{\beta_{bw0}(u, v), \ldots, \beta_{bw3}(u, v)\}$ are continuous functions of the location (u, v),
• bwj in β_{bwj} denotes a specific optimal bandwidth used in the calibration of the intercept and the jth conditional relationship (j = 0,1,2,3,4,5,6,7,8).

As explained before, the political variable X8 (Pct_Democrat2020) was not considered in the 2022 analyses. On the other hand, the local relationships analysis in 2022 allowed to include in the MGWR model the following variable:
• X8 accounts for the percentage of people over 65years old (Pct_Elderly).

The MGWR models were estimated using the Bi-square weighting function and the Golden Search Algorithm, which determines the optimal number of neighbors.
When redundant information is represented by the explanatory variables, this is recognized as collinearity. Global multicollinearity has the capacity to bias OLS and MGWR models results and produce inaccurate data. Local multicollinearity might also stop MGWR from operating. It is advisable to check for both global and local multicollinearity, even if (Fotheringham et al., 2016) showed that GWR models are robust to the effects of multicollinearity when the sample

size is large. Each explanatory variable's Variance Inflation Factor (VIF) can be used to evaluate global multicollinearity. Some authors like (O'Brien, 2007) use a more conservative approach of 4 to identify the presence of multicollinearity but the rule of thumb of 10 is also a commonly used limit. In our analysis we considered a threshold of 7.5, therefore predictors with VIF >7.5 were removed, one at a time, until the VIF values for all the rest of the explanatory variables are below that limit.

Local multicollinearity in (M)GWR models can be a truly complicated problem, since it reduces model reliability and precision. We computed and mapped local condition numbers for the final MGWR models to check local multicollinearity, since values above 30 may imply the occurrence of multicollinearity.

As for the influential observations, Cook's D values depend on both the residual and the leverage. Influential observations may be unusual in terms of the dependent variable (high residual, low leverage) or in terms of the independent variables (low residual, high leverage). Not every high leverage observation or outlier has a significant impact on the regression analysis. Cook's D values greater than 1 will be indicative of influential features (Fotheringham et al., 2002, p. 216).

According to certain writers, a linear regression model is generally regarded as significantly better than another if it differs from it by more than two Akaike's Information Criterion (AIC) units (e.g., Symonds & Moussalli, 2011).

Moreover, Local R2 values were mapped to assess the goodness of fit of the final MGWR models. In places where the MGWR model fits poorly, mapping the Local R2 values may reveal details about significant factors that could be absent from the model.

The residuals' statistically significant spatial autocorrelation indicates that the model's specification is incorrect. Consequently, it's essential to determine if the standardized residuals of the MGWR model exhibit a random distribution throughout space. Global Moran's *I* statistic was employed on the model's residuals to determine whether the residuals are spatially autocorrelated or spatial dependent (Mollalo, Rivera, et al., 2021).

After identifying the optimal models for 2021 and 2022, we joined the estimations obtained from the model to the corresponding county shapefile and mapped the coefficients to depict the local effects of each covariate on the COVID-19 vaccination rate in the United States.

4. RESULTS OF THE EMPIRICAL STUDY FOR 2021

In the regression models investigated here, the dependent variable is the percentage of people that accepted to take the COVID-19 vaccine in 2021, in Texas, per county. We are interested to see how it is related to age, gender, education, socioeconomic factors, race and ethnicity, religion beliefs, political choice and Health. To be more specific we want to see how it relates to the explanatory variables described in **Table 2 – Description of independent variables included in the analysis**. **Table 3 - Name of the independent variables included in the analysis** shows how these variables are named throughout this chapter.

Variable abbreviation	**Description:**
Pct_Female	Percent of females
Pct_Children	Percent of people under 18years old
Pct_Young	Percent of people between 19 - 39 years old
Pct_Elderly	Percent of people over 65years old
Pct_Uninsured	Percent of people that don't have health insurance
Pct_HighSchool	Percent of people with high school degree
Pct_SomeCollege	Percent of people with some college degree
Pct_FluVac	Percent of people that took the flu vaccine
Pct_Unemployment	Percent of people that do not have a job
Pct_IncomeLT10000	Percent of people with annual income lower than 10 000
Pct_IncomeGT200000	Percent of people with annual income higher than 200 000
Pct_Black_NH	Percent of people that are black non-Hispanic
Pct_White_NH	Percent of people that are white, non-Hispanic
Pct_Asian	Percent of people that are Asians
Pct_Hispanic	Percent of people that are Hispanic
Pct_Democrat2020:	Percent of people that vote for the democrat party
Pct_Republican2020	Percent of people that vote for the republican party
Religion	Percent of people that are religious

Table 3 - Name of the independent variables included in the analysis

After visualizing the data (section 4.1) and exploring the distribution of the variables (section 4.2), the spatial patterns of the dependent variable are explored to assess spatial autocorrelation and spatial non-stationarity (section 4.3). Then, we will analyse global and local relationships between the COVID-19 vaccine intake/acceptance and those potential explanatory variables (section 4.4). Global multicollinearity in the predictors is investigated in section 4.5. Finally, section 4.6 describes de MGWR model of 2021, its diagnostics analyses, and goodness-of-fit results.

4.1 VISUALISE THE DATA

The map represented below (**Figure 3 – Incidence rate of the COVID-19 vaccination, in Texas for 2021**) shows the state of Texas and all the countries within its border area. The different colors in the counties represented on the map stand for the incidence rate of the COVID-19 vaccination in 2021. Lighter shades imply lower vaccination rates, while darker shades indicate higher percentages of those who have had vaccinations. With this spatial representation of vaccination coverage throughout the state of Texas, this map sheds light on regional differences and the general acceptance of vaccines in the state.

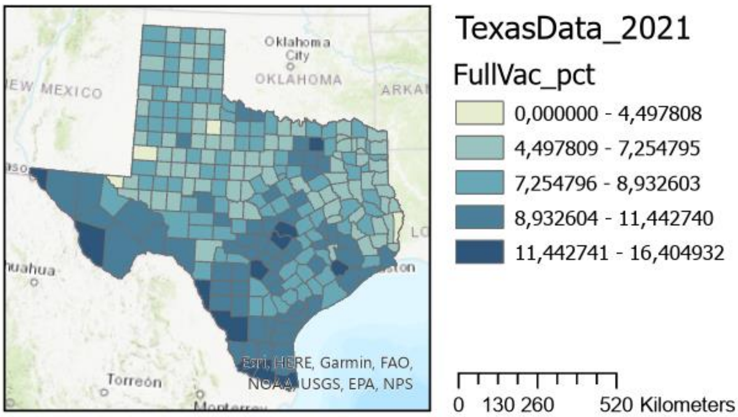


Figure 3 – Incidence rate of the COVID-19 vaccination, in Texas for 2021

The maps of the independent variables are disposed on the Appendix A. The map of the percentage of people under 18years old shows that there is a larger number of children in the northwest and southwest of the state when compared to the East side. The opposite is observed for the variable of elderly.

There is a larger number of uninsured people in the southwest, west and north, close to the border of the state. There is a larger number of people with a high school degree in the East of the state. The map in Appendix A shows that there is a larger number of people that took the flu vaccine in all the East side of the when compared to the west. There is a lower number of people that are unemployed in the north side of the state. There is a higher number of people that are black or African American in the east side of the state. There is a higher number of people that are Hispanic in the west and southwest side of the state. There is a higher number of people that are white in the northeast side of the state. There is a lower number of people that are democrats in the north of the state. On the other hand, for republican voters, there is a higher number of people in the north.

4.2 EXPLORE THE DISTRIBUTION OF THE VARIABLES.

In this section we created histogram charts in order to analyse the frequency distribution of the percentage of the county population that took the COVID-19 vaccine (FullVac_pct). The distribution is positively skewed (**Figure 4 – Histogram chart of the percentage population that took the COVID-19 vaccine, for 2021Figure 4**), the mean (8,1834%) is higher than the median (7,84945%). The variable varies between 0% and 16.4% and the typical deviation of the values to the mean is 1,897% (standard deviation). The histograms for the explanatory variables are on the Appendix B.

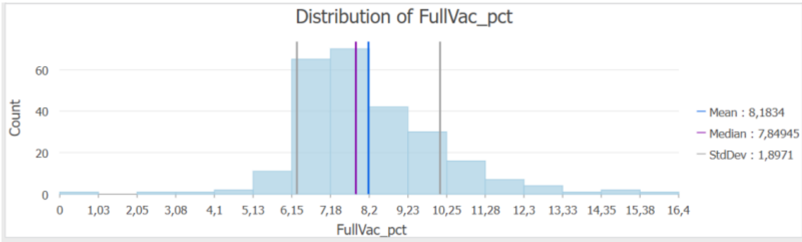


Figure 4 – Histogram chart of the percentage population that took the COVID-19 vaccine, for 2021

Afterwards we computed the Box plot charts in order to examine the distribution of the potential explanatory variables (**Figure 5 - Box plot charts of the distribution of the potential explanatory variables, for 2021**). The data's median is shown by the line inside the box. The variability or spread of the data within the interquartile range is shown by the length of the boxes in the graph. The whiskers extend from the box to the lower [upper] adjacent value of the 1st [3rd] quartile. The data points outside of this range are considered as potential outliers. They stand for the dataset's extreme values. Although they are not always mistakes, outliers might point to odd or unexpected observations that need more research.

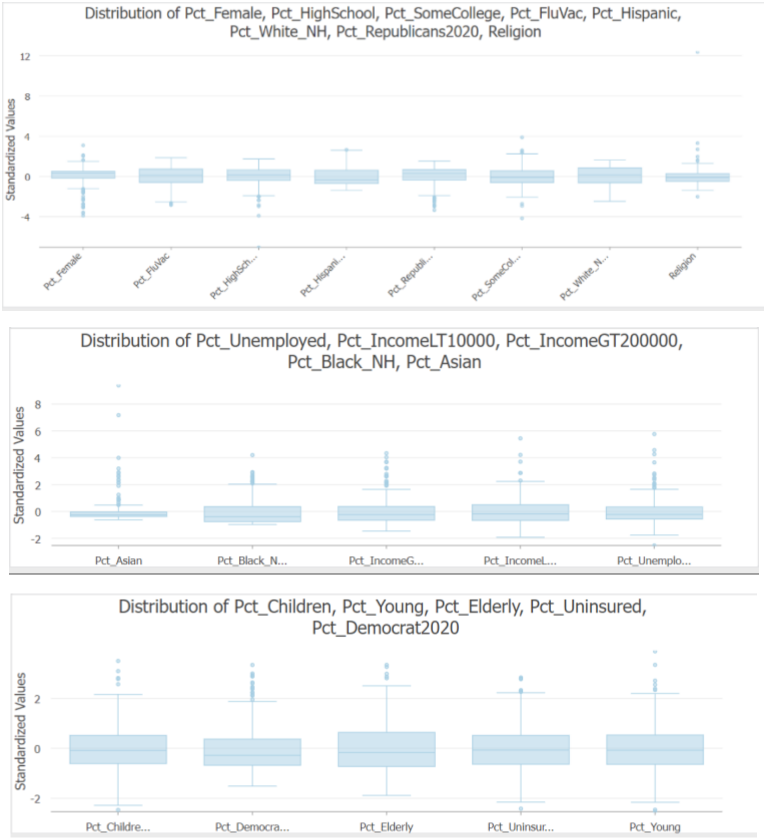


Figure 5 - Box plot charts of the distribution of the potential explanatory variables, for 2021

4.3 SPATIAL PATTERNS IN COVID-19 VACCINATION

This section explains how to discover spatial autocorrelation and spatial heterogeneity in COVID-19 vaccination using the Global Moran's I, Anselin Local Moran's I, and Getis-Ord Gi* statistics.

4.3.1 SPATIAL AUTOCORRELATION

We use the spatial autocorrelation (Global Moran’s I) test to access the spatial autocorrelation in the percentage of the county population that took the COVID-19 vaccine. In our case, the Global Moran’s value is equal to 0.258605 indicating a positive spatial autocorrelation. The p-value is 0.000000, which is much less than the usual significance level of 0.05. The null hypothesis of spatial randomness is strongly rejected by this, indicating a statistically significant spatial autocorrelation for the percentage of individuals that took the COVID-19 vaccine across the counties of Texas. This result implies that the dependent variable's values across the counties in our study area exhibit clustering or spatial dependency. This suggests that the values of the dependent variable tend to be similar to the nearby counties. This way, the classical global regression model by Ordinary Least Squares (OLS) is not appropriate and a spatial regression model should be used.

4.3.2 SPATIAL NON-STATIONARITY

Now, we will investigate spatial non-stationarity in the percentage of county population that took the COVID-19 vaccine with the Anselin Local Moran's I and the Getis-Ord Gi* statistics.

The Anselin Local Moran's I test identifies an appropriate scale of analysis using the distribution of the weighted features. The Local Moran's I map (**Figure 6 – Local Moran's I map of the percentage of county population that took the COVID-19 vaccine, for 2021Figure 6**) shows: 67 counties classified as clusters of low values, 44 counties are clusters of high values; 2 counties classified as spatial outliers because their low values correlate with neighbouring high values (Low-High outliers), and 10 county's high value correlates with neighbouring low values (High-Low outlier).

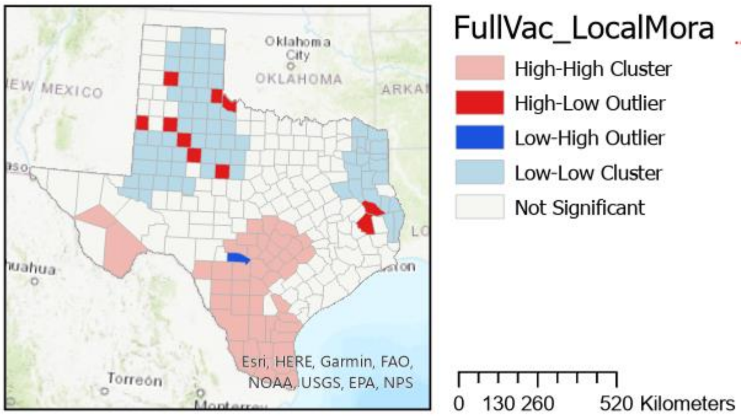


Figure 6 – Local Moran's I map of the percentage of county population that took the COVID-19 vaccine, for 2021

Afterwards, we computed the hot spot analysis (**Figure 7 – Hot Spot Analysis of the percentage of county population that took the COVID-19 vaccine, for 2021Figure 7**). This test identifies an appropriate scale of analysis and corrects for both multiple testing and spatial dependence. The map shows 55 clusters of high values (i.e., hot-spots), 6 with 90% confidence, 27 with 95% confidence and 22 with 99% confidence. Additionally, we can see that we also have 87 clusters of low values (i.e., cold-spots), 3 with 99% confidence, 58 with 95% confidence and 26 with 90% confidence. These clusters of cold-spots highlight the counties with higher vaccine hesitancy.

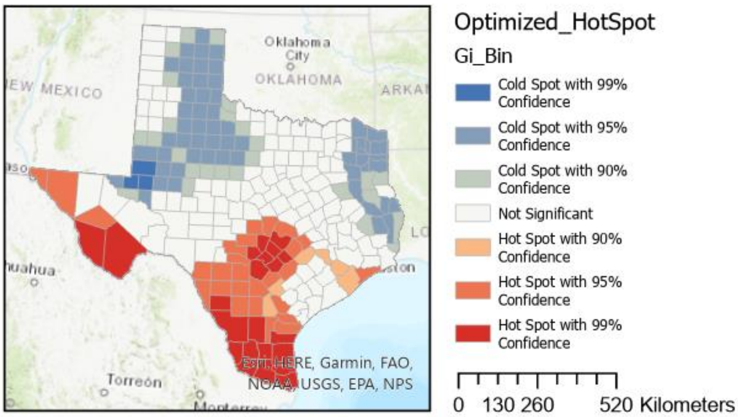


Figure 7 – Hot Spot Analysis of the percentage of county population that took the COVID-19 vaccine, for 2021

To sum up, the results show that the dependent variable has evidence spatial non-stationarity, as shown by the existence of significant clusters, spatial outliers, hot spots, and cold spots. This implies that there may be spatial variations in the relationship between the dependent variable and its determinants across different regions within the study area. This way, a global OLS regression is inappropriate.

4.4 Bivariate relationships

In these sections we analyse the bivariate relationships. Here we explore global and local relationships between the percentage of the county population with the COVID-19 vaccine (Y) and eighteen potential explanatory variables shown in **Table 2 – Description of independent variables included in the analysis**.

We will use scatterplots and the correlation coefficient to explore global relationships. Then, we will investigate if there are any geographical variations in the relationships between the incidence rate of COVID-19 vaccination and those variables.

4.4.1 Global Relationships

The Scatter Plot Matrix graph allows us to examine global relationships between the COVID-19 vaccination incidence rate and a variety of other variables (Appendix C). What we are looking for in the scatterplots with trend lines fitted are clouds of points close to the lines, showing that values in one variable change linearly with changes in another. The closer the values to the straight line, the stronger the linear relationship.

In the map, selecting the counties with highest percentage of people that took the COVID-19 vaccine, represents the relationship of this variable with the others in those counties. This allowed to confirm that the strength of the relationships vary across the state, and only a few explanatory variables seem to have a strong relationship in several counties.

	FullVac_pct	Pct_Female	Pct_Children	Pct_Young	Pct_Elderly	Pct_Uninsured	Pct_HighSchool	Pct_SomeCollege	Pct_FluVac	Pct_Unemployed	Pct_IncomeLT10000	Pct_IncomeGT200000	Pct_Black_NH	Pct_Asian	Pct_Hispanic	Pct_White_NH	Pct_Democrat2020	Pct_Republicans2020	Religion
FullVac_pct																			
Pct_Female	0.016 (.795)																		
Pct_Children	0.098 (.121)	0.197 (.002)																	
Pct_Young	0.258 (<.001)	-0.334 (<.001)	0.135 (.033)																
Pct_Elderly	-0.228 (<.001)	0.148 (.019)	-0.551 (<.001)	-0.759 (<.001)															
Pct_Uninsured	0.016 (.798)	-0.073 (.250)	0.139 (.028)	-0.094 (.136)	0.097 (.125)														
Pct_HighSchool	-0.231 (<.001)	0.329 (<.001)	-0.361 (<.001)	-0.171 (.006)	0.170 (.007)	-0.611 (<.001)													
Pct_SomeCollege	0.057 (.366)	0.424 (<.001)	0.007 (.916)	0.012 (.848)	-0.070 (.270)	-0.504 (<.001)	0.669 (<.001)												
Pct_FluVac	0.111 (.079)	0.143 (.024)	-0.090 (.155)	0.132 (.036)	-0.150 (.017)	-0.330 (<.001)	0.342 (<.001)	0.252 (<.001)											
Pct_Unemployed	0.214 (.001)	0.080 (.208)	0.060 (.341)	-0.025 (.693)	0.005 (.933)	0.133 (.035)	-0.257 (<.001)	-0.188 (.003)	-0.009 (.893)										
Pct_IncomeLT10000	0.208 (.001)	-0.005 (.931)	-0.064 (.315)	0.167 (.008)	-0.017 (.788)	0.225 (<.001)	-0.327 (<.001)	-0.241 (<.001)	-0.151 (.017)	0.428 (<.001)									
Pct_IncomeGT200000	0.200 (.001)	0.180 (.004)	0.018 (.781)	0.050 (.432)	-0.201 (.001)	-0.378 (<.001)	0.437 (<.001)	0.488 (<.001)	0.415 (<.001)	-0.240 (<.001)	-0.375 (<.001)								
Pct_Black_NH	-0.074 (.242)	-0.001 (.988)	-0.152 (.016)	0.223 (<.001)	-0.174 (.006)	-0.222 (<.001)	0.230 (<.001)	0.072 (.255)	0.326 (<.001)	0.190 (.003)	0.080 (.204)	0.087 (.171)							
Pct_Asian	0.404 (<.001)	0.062 (.329)	0.081 (.200)	0.236 (<.001)	-0.307 (<.001)	-0.271 (<.001)	0.163 (.010)	0.325 (<.001)	0.284 (<.001)	-0.056 (.378)	-0.121 (.055)	0.499 (<.001)	0.267 (<.001)						
Pct_Hispanic	0.508 (<.001)	-0.217 (.001)	0.465 (<.001)	0.360 (<.001)	-0.399 (<.001)	0.363 (<.001)	-0.734 (<.001)	-0.379 (<.001)	-0.331 (<.001)	0.192 (.002)	0.312 (<.001)	-0.275 (<.001)	-0.398 (<.001)	-0.033 (.604)					
Pct_White_NH	-0.563 (<.001)	0.224 (<.001)	-0.462 (<.001)	-0.478 (<.001)	0.512 (<.001)	-0.291 (<.001)	0.698 (<.001)	0.349 (<.001)	0.229 (<.001)	-0.258 (<.001)	-0.346 (<.001)	0.222 (<.001)	0.095 (.133)	-0.149 (.018)	-0.946 (<.001)				
Pct_Democrat2020	0.805 (<.001)	0.057 (.366)	0.194 (.002)	0.413 (<.001)	-0.375 (<.001)	-0.016 (.795)	-0.259 (<.001)	0.027 (.673)	0.145 (.022)	0.369 (<.001)	0.329 (<.001)	0.082 (.193)	0.185 (.003)	0.431 (<.001)	0.546 (<.001)	-0.687 (<.001)			
Pct_Republicans2020	-0.807 (<.001)	-0.060 (.343)	-0.194 (.002)	-0.418 (<.001)	0.381 (<.001)	0.022 (.733)	0.250 (<.001)	-0.034 (.588)	-0.149 (.018)	-0.360 (<.001)	-0.323 (<.001)	-0.090 (.154)	-0.186 (.003)	-0.437 (<.001)	-0.540 (<.001)	0.681 (<.001)	-1.000 (<.001)		
Religion	-0.121 (.055)	0.155 (.014)	0.134 (.034)	-0.186 (.003)	0.145 (.021)	0.173 (.006)	-0.138 (.028)	-0.069 (.275)	-0.304 (<.001)	0.032 (.614)	0.112 (.077)	-0.226 (<.001)	-0.112 (.077)	-0.160 (.011)	0.109 (.086)	-0.068 (.285)	-0.120 (.057)	0.124 (.049)	

Computed correlation used pearson-method with listwise-deletion.

Figure 8 - Scatter and Pearson's correlation matrix, for 2021

This way, population that took the COVID-19 vaccine (FullVac_pct) seems to have a global positive relationship with Democrats, Hispanics, and Asians. It seems to have a global negative relationship with Republicans and White. It does not seem to be correlated with female, children, uninsured, college, flu vaccine intake, religion and Black or African American.

We observe that the correlations of the dependent variable with the explanatory variables that were significant were: Pct_Republican2020, Pct_Democrat2020, Pct_White_NH, Pct_Hispanic, Pct_Asian, Pct_IncomeGT200000, Pct_IncomeLT10000, Pct_Unemployed, Pct_HighSchool, Pct_Elderly and Pct_Young.

4.4.2 LOCAL RELATIONSHIPS

The local bivariate relationships maps show statistically significant relationships using local entropy. These maps are very important to choose variables for the model since some of them may not be relevant globally but relevant locally. If a linear relationship is not observed, we should not use the variable. Appendix D shows all local bivariate relationships maps, except those where none of the counties had a significant relationship.

The local relationships between COVID-19 vaccination and female population are not significant in all counties. This fact explains the apparent inexistent relationship in the scatterplot graph of the FullVac and the Pct_Female variables. The same happens for Children, Young, High School, unemployment, Income > 10 000, Black or African American and Religion. Therefore, these variables were not included in the MGWR models.

For the variable of percentage of **elderly** population, the local relationship is negative linear in 6 counties, convex in 60 counties and concave in one county. This variable does not have significant relationship in the whole study area (r = -0.228), but it has a significant linear relationship in 6 counties. As for the percentage of **uninsured population**, the local relationship is positive linear in 20 counties and convex in 31 counties. This variable is not relevant in the whole study area (r = 0.016) but has a significant linear relationship in at least 20 counties. For the variable of percentage of population that has some **college degree**, the local relationship is positive linear in 48 counties and convex in 13 counties. This variable is not relevant in the whole study area (r = 0.057), but it has a significant linear relationship in at least 48 counties. For the percentage of population that took the **flu vaccine**, the local relationship is positive linear in 9 counties, concave in 3 and convex in 37 counties. This variable is not relevant in the whole study area (r = 0.111), but it has a significant linear relationship in at least 9 counties. For the variable of percentage of population that has an annual **income higher than 200 000** dollars, the local relationship is positive linear in 41 counties and convex in 4 counties. This variable is not relevant in the whole study area (r = 0.2) but it has a significant linear relationship in at least 41 counties.

As expected from the scatter and Pearson's correlation matrix displayed above (**Figure 8 - Scatter and Pearson's correlation matrix, for 2021Error! Reference source not found.**), population that took the COVID-19 vaccine shows a moderate positive relation (r=0.404) with the percent of **Asian** population in Texas. The local relationship is positive linear in 65 counties and concave in 49 counties. Population that took the COVID-19 vaccine shows a positive moderate relation (r = 0.508) with the percent of **Hispanic** population. The local relationship is positive linear in 17 counties. Additionally, population that took the COVID-19 vaccine shows a negative moderate relation (r = - 0.563) with the percent of **White** population in Texas. The local relationship is negative linear in 32 counties and convex in 8 counties.

Finally, the population that took the COVID-19 vaccine shows a global positive relation (r = 0.805) with the percent of **democrats** and a global negative relation (r = - 0.807) with the percent **republicans**. The local relationship is positive linear in 181 counties, concave in 2 and convex in 4 counties, for the democrats and negative linear in 181 counties and concave in 2 and convex in 4 counties, for the republicans.

Variable	Negative linear	Positive linear	Convex	Concave
Pct_Female	0	0	0	0
Pct_Children	0	0	0	0
Pct_Young	0	0	0	0
Pct_Unemployment	0	0	0	0
Pct_IncomeLT10000	0	0	0	0
Pct_Black_NH	0	0	0	0
Religion	0	0	0	0
Pct_HighSchool	0	0	0	0
Pct_Uninsured	0	20	31	0
Pct_Elderly	6	0	60	1
Pct_SomeCollege	0	48	13	0
Pct_FluVac	0	9	37	3
Pct_IncomeGT200000	0	41	4	0
Pct_Asian	0	65	0	49
Pct_Hispanic	0	17	0	0

Variable	Negative linear	Positive linear	Convex	Concave
Pct_White_NH	32	0	8	0
Pct_Democrat2020	0	181	2	4
Pct_Republican2020	181	0	4	2

Table 4 - Number of counties with significant relationships, for 2021

To conclude, according to the obtained results and as we can see in **Table 4 - Number of counties with significant relationships, for 2021**, the variables of Pct_Female, Pct_Children, Pct_Young, Pct_HighSchool, Pct_unemployment, Pct_IncomeLT10000, Pct_Black and Religion should not be included in linear regression models such as MGWR, because they do not exhibit significant linear relationships. Additionally, after carefully observing all the maps (presented on the Appendix D) we also concluded that would be best for our model if we don´t include the variables **Pct_Elderly and Pct_Republican** since only a few counties show significant linear relationships with the dependent variable. This way, we proceeded the analysis with the remaining eight variables since they show a reasonable number of counties with significant linear relationships.

4.5 GLOBAL MULTICOLLINEARITY

The global multicollinearity analysis regarding the eight potential explanatory variables to be included in the MGWR model showed that the variables Pct_Hispanic and Pct_White_NH had high Variance Inflation Factor (VIF) values, exceeding 7.5. In this case, these values suggested that they may be redundant or highly correlated with other variables in the model, potentially leading to unstable coefficient estimates. This way, the Pct_Hispanic variable was then eliminated, because it has a significant linear relationship only in 17 counties. The VIF values were then computed for models with the remaining seven potential explanatory variables.

Table 5 – Global multicollinearity Results for 2021 data, summarises the results of the global multicollinearity analysis regarding the seven independent variables to be included in the MGWR model.

Variable	VIF
Intercept	-----------------------
Pct_Uninsured	1.479356
Pct_SomeCollege	1.7729
Pct_FluVac	1.449129

Variable	VIF
Pct_IncomeGT200000	1.778825
Pct_Asian	1.716258
Pct_White_NH	3.108637
Pct_Democrat2020	2.884483

Table 5 – Global multicollinearity Results for 2021 data

The VIF values of each predictor are all lower than 4 so there is no relation of multicollinearity among them. These low VIF values suggest that their coefficient estimates are likely stable and reliable.

4.6 MGWR MODEL

As demonstrated above by the local bivariate analysis certain relations may function over greater scales than others. This behavior is typical of most spatially heterogeneous processes. In order to model this, each predictor can have a unique bandwidth identified by MGWR, allowing for different spatial scales in the relationships between each predictor and the dependent variable. The MGWR model was estimated with the seven independent variables listed in Table 5. When we did the analysis in ArcGIS, there is a warning message that says that it had problems reading 2 out of the 254 total records. Since it was not possible to correct this error, we considered that 100% of the counties corresponds to 252 counties in the following analysis and model results.

In this section we examined the individual bandwidth and summary statistics of the local coefficients of each explanatory variable (**Table 6 - Individual bandwidth and summary statistics of the local coefficients of each explanatory variable for 2021**). The local bandwidths of the MGWR showed that the relationship between Asian population (Pct_Asian) and population that has some college degree (Pct_SomeCollege) with COVID-19 incidence rate is global, because the bandwidth of that variable included all 252 counties. The relationship of Pct_Democrat2020, Pct_WhiteNH and Pct_IncomeGT200000 with the dependent variable operates at a regional scale (the bandwidth includes 58.33%, 53.57% and 53.57%, respectively, of the nearest counties). As for Pct_Uninsured and Pct_FluVac, they have a more local relationship (the bandwidth includes 24.6% of the neighboring counties, for both variables). As for the intercept, we can see that it also operates at a local scale since the bandwidth includes 32.54% of the nearest counties.

Variable (Scaled)	Bandwidth (nr. and % of counties)	Nr. (%) of counties with significant coefficients	Mean	Standard Deviation	Minimum	Median	Maximum
Intercept	82 (32.54%)	105 (41.67%)	-0.0829	0.2168	-0.4834	-0.1095	0.3118
Pct_Uninsured	62 (24.6%)	19 (7.54%)	-0.0458	0.1409	-0.2941	-0.0504	0.3118
Pct_SomeCollege	252 (100%)	0 (0%)	-0.0649	0.0082	-0.0786	-0.0641	-0.0517
Pct_FluVac	62 (24.6%)	0 (0%)	0.0398	0.0662	-0.1370	0.0331	0.1743
Pct_IncomeGT200000	135 (53.57%)	149 (59.13%)	0.1447	0.1067	-0.0686	0.1765	0.2646
Pct_Asian	252 (100%)	0 (0%)	0.0518	0.0025	0.0468	0.0528	0.0551
Pct_White_NH	135 (53.57%)	0 (0%)	-0.0061	0.0597	-0.1719	0.0046	0.1049
Pct_Democrat2020	147 (58.33%)	252 (100%)	0.5533	0.0499	0.4799	0.5436	0.7551

Table 6 - Individual bandwidth and summary statistics of the local coefficients of each explanatory variable for 2021

As we can see on the local coefficients variability alongside the bandwidths of
, high bandwidths result in less variability in the coefficients. This way, the local coefficients of Pct_SomeCollege have the smallest range of values (Maximum–Minimum=0.0269) and standard deviation (0.0082). The same happens for the other variable of high bandwidths, Pct_Asian (Maximum–Minimum=0.0083) and standard deviation (0.0025). Therefore, we can reach to the conclusion that the influence of college degree and being Asian on COVID-19 vaccine incidence rate is effectively stationary over space.

Since the dependent and independent variables were standardized, the values of the coefficients can be directly compared. On average, Pct_Democrat2020 is the most influential factor (mean= 0.5533) and Pct_SomeCollege is the least one (mean= –0.0649). We also observed that no county exhibited any significant association between Pct_SomeCollege, Pct_ FluVac, Pct_Asian and Pct_White_NH with COVID-19 vaccine incidence rate.

4.6.1 Local Multicollinearity

Local multicollinearity in the MGWR model was diagnosed using Local Condition Numbers. Since every value is below the rule of thumb of 30, none of the local condition numbers in this MGWR model indicate that local multicollinearity is a problem, as showed in **Figure 9 – Local Multicollinearity, 2021**.

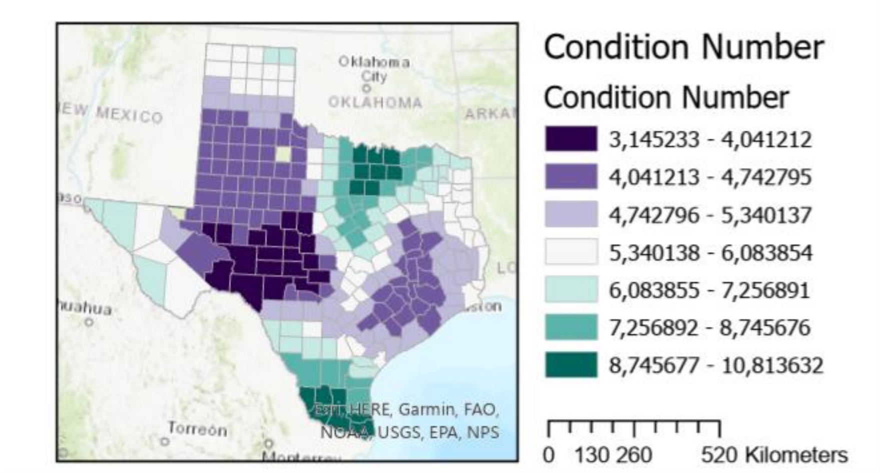


Figure 9 – Local Multicollinearity, 2021

4.6.2 INFLUENTIAL OBSERVATIONS

This section demonstrates the application of Cook's Distance, a metric that quantifies the impact of an observation on the calibration of the model. Since every value in the map of **Figure 10 – Influential Observations, 2021Figure 10** is below 1, there are no influential observations that could represent a problem for the MGWR model.

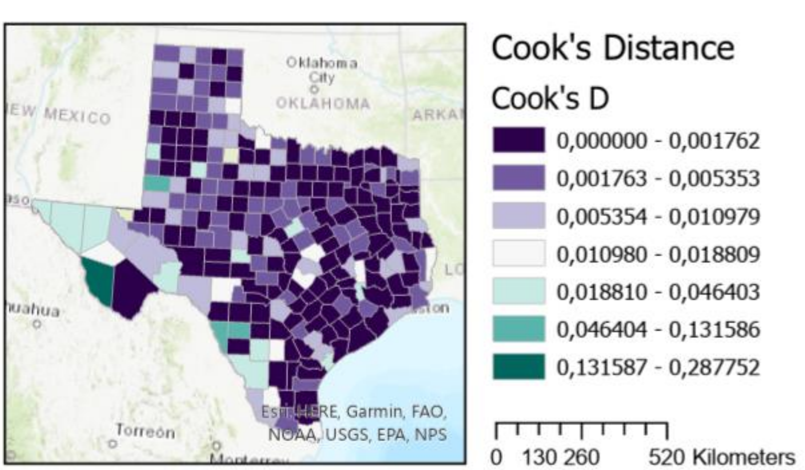


Figure 10 – Influential Observations, 2021

4.6.3 GOODNESS-OF-FIT OF THE FINAL MGWR MODEL

The final MGWR model only includes seven explanatory variables (Pct_Uninsured, pct_FluVac, Pct_SomeCollege, Pct_IncomeGT200000, Pct_White_NH, Pct_Asian and Pct_Democrat2020) and its performance (AICc = 355.7537) is better than the corresponding GWR model (AICc = 370.1123) as shown in **Table 7 – Goodness-of-fit of the final MGWR model, 2021** below. The single optimal bandwidth obtained in the GWR calibration is 118 counties. In MGWR, Pct_someCollege, and Pct_Asian operated at a global scale (all 252 counties), Pct_Democrat2020, Pct_White_NH and Pct_IncomeGT200000 at a regional scale with a bandwidth of 147 (58.33%) 135 (53.57%) and 135 (53.57%) counties, respectively and Pct_Uninsured and Pct_FluVac at a local scale with a bandwidth of 62 (24.6%) counties for both variables. As (Yu et al., 2020) point out, these results are informative because they demonstrate the dissimilarity between the GWR and MGWR results when the optimal bandwidths of the predictors are not identical to that of the single GWR bandwidth.

Model	R2	Adj – R2	AICc	Effective degrees of freedom
GWR with 7 predictors	0.8171	0.7732	378.9396	203.4393
MGWR with 7 predictors	**0.8343**	**0.8044**	**355.7537**	**213.7069**

Table 7 – Goodness-of-fit of the final MGWR model, 2021

The map of Local R2 values in **Figure 11 – Local R2 Map, 2021Figure 11 – Local R2 MapFigure 11** show that MGWR has good explanatory capability in most of the counties, particularly in the northeast and west region. However, the model has a poorer fit in the northern region, where clusters of higher vaccine hesitancy were identified (i.e., cold-spots of the vaccine incidence rate). This may indicate that the model is missing a relevant explanatory variable in that region.

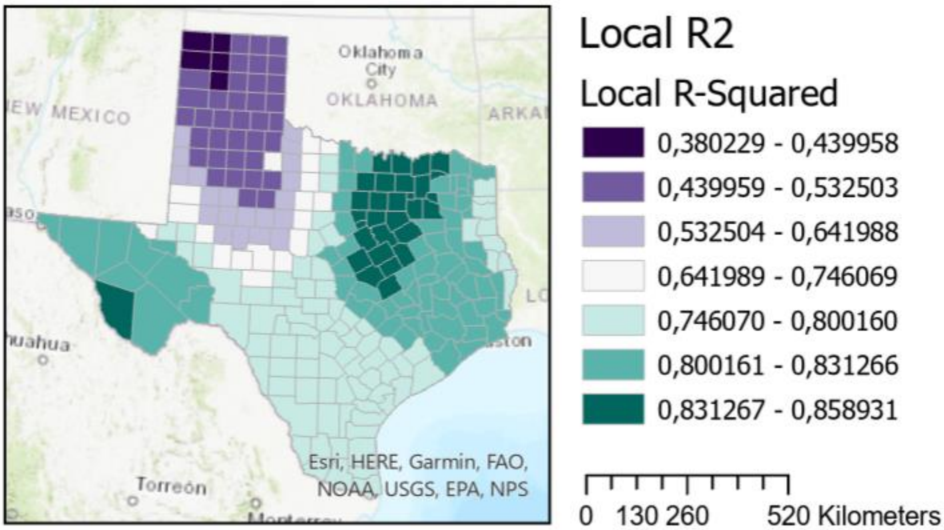


Figure 11 – Local R2 Map, 2021

4.6.4 RESIDUAL ANALYSIS OF THE FINAL MGWR MODEL

It is essential to determine if the standardized residuals of the MGWR model (**Figure 12 - Standardized residuals of the MGWR model, 2021**) exhibit a random distribution throughout space.

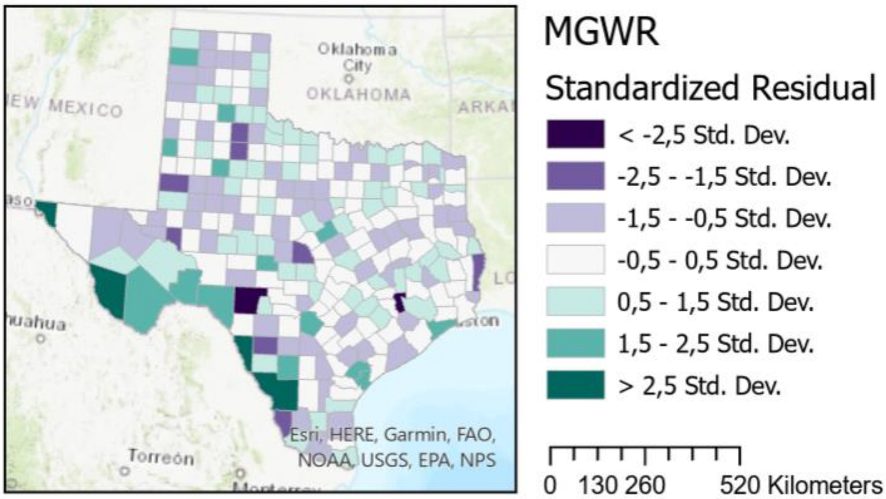


Figure 12 - Standardized residuals of the MGWR model, 2021

We use the spatial autocorrelation (Global Moran's I) test to access the spatial autocorrelation in the standardized residuals of the MGWR model. In our case, the Global Moran's value is approximately zero and the p-value is 0.912073, which is higher than the usual significance level of 0.05. The null hypothesis of spatial randomness is failed to be rejected, indicating a statistically non-significant spatial autocorrelation. In other words, the standardized residuals pattern does not appear to be significantly different than random.

4.6.5 LOCAL COEFFICIENTS OF THE FINAL MGWR MODEL

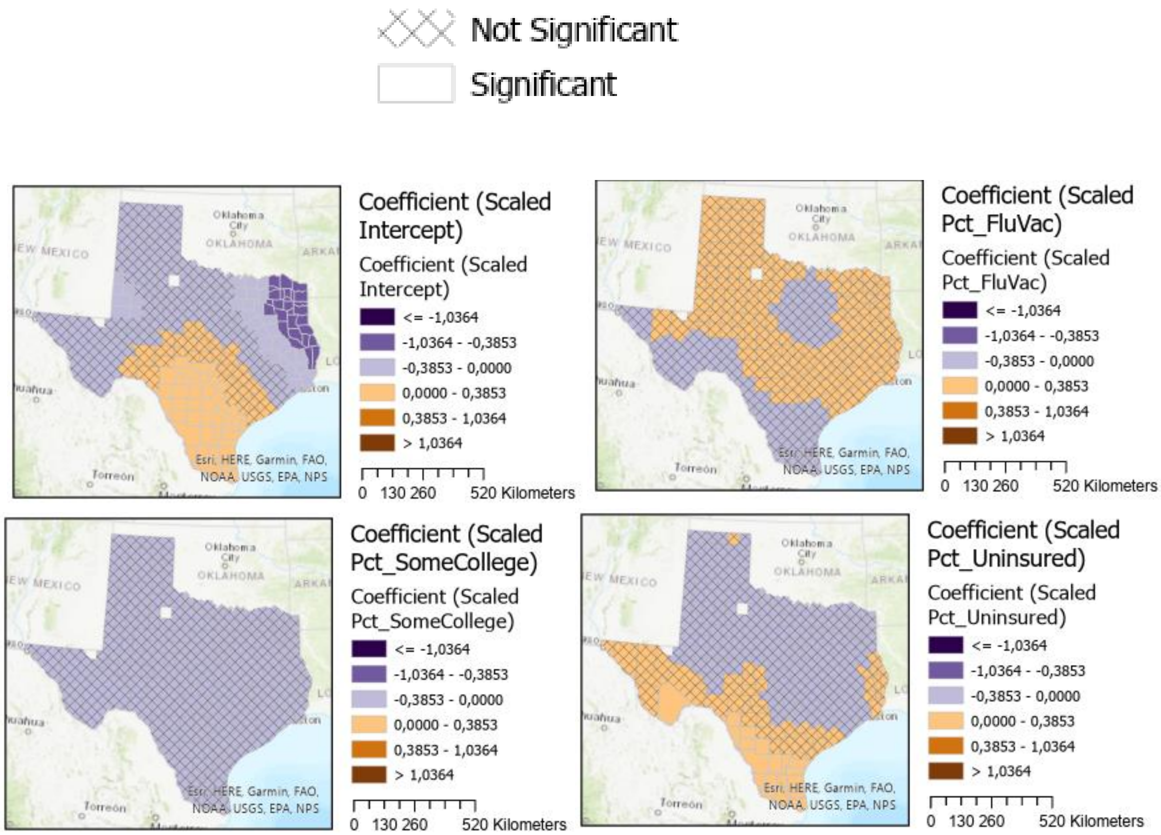
The maps of the local coefficients of the intercept and the seven predictors of the MGWR model are disclosed below. As expected from the results of the initial MGWR model (**Table 6 - Individual bandwidth and summary statistics of the local coefficients of each explanatory variable for 2021**

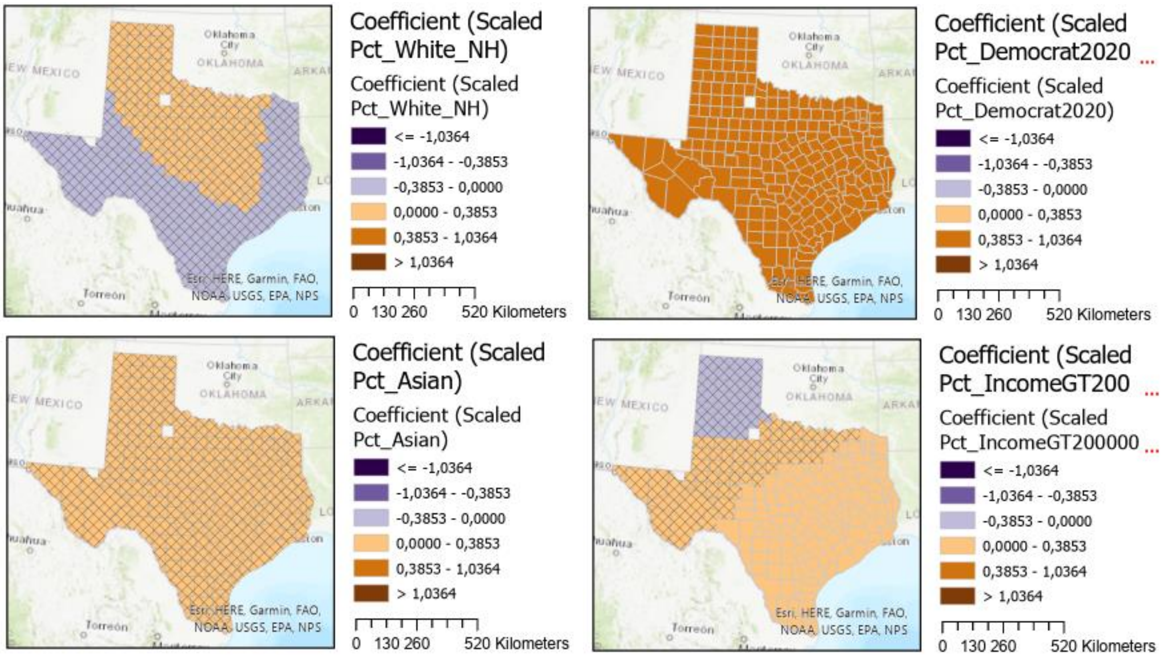
), the coefficients of Pct_SomeCollege are all negative but not significant. In northern counties, where the model has the poorest fit, the only significant coefficients are those of the Pct_Democrat2020.

The coefficients of Pct_IncomeGT200000 and Pct_Democrat2020 are all positive and significantly different from zero in 135 and 147 counties, respectively.

In the local maps presented below, we can observe that IncomeGT200000 is significantly positively associated with COVID-19 vaccine intake on the southeast of the state. As for Pct_Democrat2020, the variable is positively associated with vaccine intake in all the areas of Texas state. Lastly, for Pct_Uninsured, the variable is positively associated with COVID-19 vaccine intake but only in the south of the estate and the southwest close to the border.

2021: Local coefficients of the MGWR model




Coefficient (Scaled Pct_White_NH)
Coefficient (Scaled Pct_White_NH)
<= -1,0364
-1,0364 - -0,3853
-0,3853 - 0,0000
0,0000 - 0,3853
0,3853 - 1,0364
> 1,0364
0 130 260 520 Kilometers
Coefficient (Scaled Pct_Democrat2020 ...
Coefficient (Scaled Pct_Democrat2020)
<= -1,0364
-1,0364 - -0,3853
-0,3853 - 0,0000
0,0000 - 0,3853
0,3853 - 1,0364
> 1,0364
0 130 260 520 Kilometers
Coefficient (Scaled Pct_Asian)
Coefficient (Scaled Pct_Asian)
<= -1,0364
-1,0364 - -0,3853
-0,3853 - 0,0000
0,0000 - 0,3853
0,3853 - 1,0364
> 1,0364
0 130 260 520 Kilometers
Coefficient (Scaled Pct_IncomeGT200 ...
Coefficient (Scaled Pct_IncomeGT200000 ...
<= -1,0364
-1,0364 - -0,3853
-0,3853 - 0,0000
0,0000 - 0,3853
0,3853 - 1,0364
> 1,0364
0 130 260 520 Kilometers
Esri, HERE, Garmin, FAO, NOAA, USGS, EPA, NPS

5. RESULTS OF THE EMPIRICAL STUDY FOR 2022

In the regression models investigated here, the dependent variable is the COVID-19 vaccine incidence rate in 2022, in Texas, per county. The explanatory variables are the same as the ones showed in **Table 3 - Name of the independent variables included in the analysis**, but the political variables (Pct_Democrat2020 and Pct_Republican2020) were not considered as explained in section 3.1.

5.1 VISUALISE THE DATA

The map represented below (**Figure 13 - Incidence rate of the COVID-19 vaccination, in Texas for 2022**) shows the state of Texas and all the countries within its border area. The different colors in the counties represented on the map stand for the incidence rate of the COVID-19 vaccination in 2022. Lighter shades imply lower vaccination rates, while darker shades indicate higher percentages of those who have had vaccinations. Similarly to 2021, the northern counties show a lower vaccination coverage.

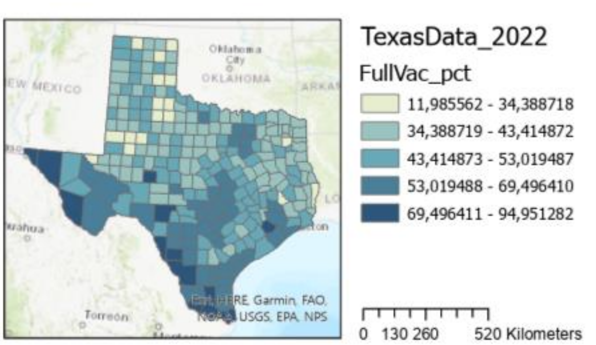


Figure 13 - Incidence rate of the COVID-19 vaccination, in Texas for 2022

The maps of the independent variables are disposed on the Appendix A. The map of the percentage of people under 18years old shows that there is a larger number of children in the northwest and southwest of the state when compared to the East side. The opposite is observed for the variable of High school. There is a larger number of uninsured people in the southwest, west and north, close to the border of the state. The map in Appendix A shows that there is a larger number of people that took the flu vaccine in all the East side of the when compared to the west. There is a lower number of people that are unemployed in the north side of the state. There is a higher number of people that are black or African American in the east side of the state. There is a higher number of people that are Hispanic in the west and southwest side of the state. There is a higher number of people that are white in the northeast side of the state.

5.2 Explore the distribution of the variables.

In this section we created histogram charts in order to analyse the frequency distribution of the percentage of the county population that took the COVID-19 vaccine (FullVac_pct) in 2022. The distribution is positively skewed (**Figure 14 - Histogram chart of the percentage population that took the COVID-19 vaccine, for 2022Figure 14**) the mean (47,25982%) is higher than the median (44,97872%). The variable varies between 0% and 16.4% and the typical deviation of the values to the mean is 12,01292% (standard deviation). The histograms for the explanatory variables are also on the Appendix B.

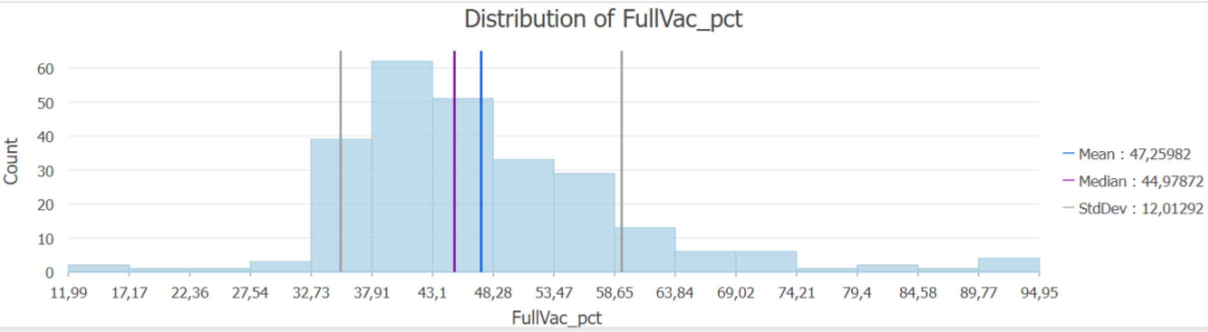


Figure 14 - Histogram chart of the percentage population that took the COVID-19 vaccine, for 2022

Afterwards we computed the Box plot charts in order to examine the distribution of the potential explanatory variables (**Figure 15 - Box plot charts of the distribution of the potential explanatory variables, for 2022**).

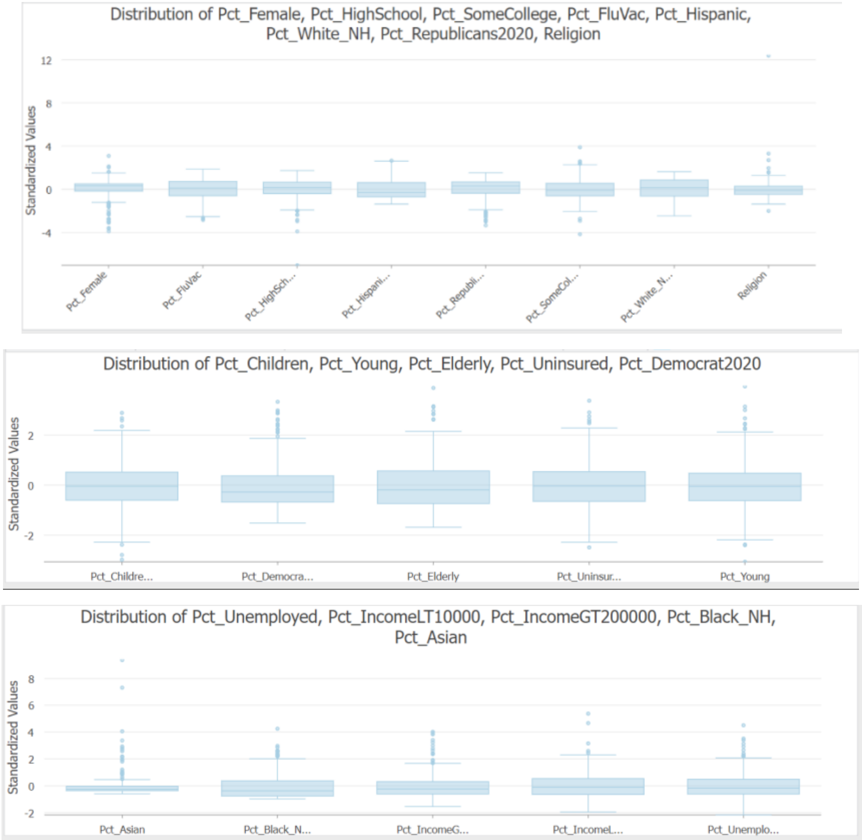


Figure 15 - Box plot charts of the distribution of the potential explanatory variables, for 2022

5.3 Investigate spatial patterns in COVID-19 vaccination in 2022

This section explains how to discover spatial autocorrelation and spatial heterogeneity in COVID-19 vaccination using the Global Moran's I, Anselin Local Moran's I, and Getis-Ord Gi* statistics.

5.3.1 Spatial Autocorrelation

We use the spatial autocorrelation (Global Moran’s I) test to access the spatial autocorrelation in the percentage of the county population that took the COVID-19 vaccine in 2022.

For this year data, the Global Moran’s value is equal to 0.257321 (p-value <0.0000) indicating a significative positive spatial autocorrelation for the percentage of individuals that took the COVID-19 vaccine across the counties of Texas. This way, the classical global regression model by Ordinary Least Squares (OLS) is not appropriate and a spatial regression model should be used, as concluded previously for 2021.

5.3.2 SPATIAL NON-STATIONARITY

Now, we will investigate spatial non-stationarity in the percentage of county population that took the COVID-19 vaccine with the Anselin Local Moran's I and the Getis-Ord Gi* statistics.

The Local Moran's I map for 2022 (**Figure 16 - Local Moran's I map of the percentage of county population that took the COVID-19 vaccine, for 2021**) shows 74 counties classified as clusters of low values, 37 counties are clusters of high values; 3 counties classified as spatial outliers because their low values correlate with neighbouring high values (Low-High outliers), and 5 county's high value correlates with neighbouring low values (High-Low outlier).

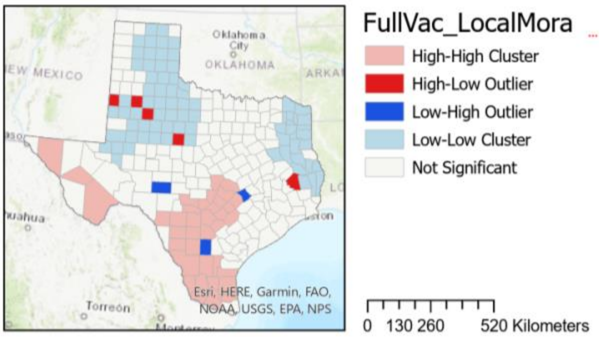


Figure 16 - Local Moran's I map of the percentage of county population that took the COVID-19 vaccine, for 2021

Afterwards we computed the hot spot analysis (**Figure 17 - Hot Spot Analysis of the percentage of county population that took the COVID-19 vaccine, for 2022**). The map shows 47 clusters of high values (i.e., hot-spots), 7 with 90% confidence, 23 with 95% confidence and 17 with 99% confidence. Additionally, we can see that we also have 80 clusters of low values (i.e., cold-spots), 1 with 99% confidence, 59 with 95% confidence and 20 with 90% confidence. These clusters of cold-spots highlight the counties with higher vaccine hesitancy.

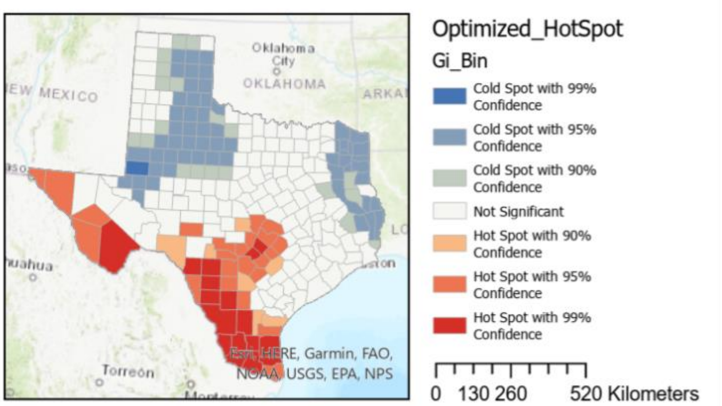


Figure 17 - Hot Spot Analysis of the percentage of county population that took the COVID-19 vaccine, for 2022

To sum up, the results show that the dependent variable has evidence spatial non-stationarity. As expected, there may be spatial variations in the relationship between the dependent variable and its determinants across different counties. This way, a global OLS regression is inappropriate.

5.4 ANALYSE BIVARIATE RELATIONSHIPS

In these sections we analyse the bivariate relationships. Here we explore global and local relationships between the percentage of the county population with the COVID-19 vaccine (Y) and sixteen potential explanatory variables. The variables are shown in **Table 2 – Description of independent variables included in the analysis** but we will not be using the political variables for this year analysis. We will use scatterplots and the correlation coefficient to explore global relationships. Then, we will investigate if there are any geographical variations in the relationships between educational attainment and those variables.

5.4.1 GLOBAL RELATIONSHIPS

The Scatter Plot Matrix graph allows us to examine global relationships between the COVID-19 vaccination incidence rate and a variety of other variables in 2022 (Appendix C).

In the map, selecting the counties with the highest percentage of people that took the COVID-19 vaccine, represents the relationship of this variables with the others in those counties. This allowed to confirm that the strength of the relationships vary across the state, and only a few explanatory variables seem to have a strong relationship in those counties.

	FullVac_pct	*Pct_Female*	*Pct_Children*	*Pct_Young*	*Pct_Elderly*	*Pct_Uninsured*	*Pct_HighSchool*	*Pct_SomeCollege*	*Pct_FluVac*	*Pct_Unemployed*	*Pct_IncomeLT10000*	*Pct_IncomeGT200000*	*Pct_Black_NH*	*Pct_Asian*	*Pct_Hispanic*	*Pct_White_NH*	*Religion*
FullVac_pct																	
Pct_Female	-0.036 (.568)																
Pct_Children	0.097 (.123)	0.283 (<.001)															
Pct_Young	0.255 (<.001)	-0.326 (<.001)	0.181 (.004)														
Pct_Elderly	-0.203 (.001)	0.130 (.039)	-0.604 (<.001)	-0.744 (<.001)													
Pct_Uninsured	0.027 (.665)	-0.082 (.192)	0.146 (.020)	-0.041 (.521)	0.056 (.375)												
Pct_HighSchool	-0.252 (<.001)	0.347 (<.001)	-0.274 (<.001)	-0.217 (.001)	0.156 (.013)	-0.618 (<.001)											
Pct_SomeCollege	0.113 (.072)	0.396 (<.001)	0.018 (.770)	-0.068 (.285)	-0.013 (.841)	-0.439 (<.001)	0.654 (<.001)										
Pct_FluVac	0.088 (.166)	0.217 (.001)	-0.002 (.975)	0.110 (.081)	-0.167 (.008)	-0.329 (<.001)	0.355 (<.001)	0.290 (<.001)									
Pct_Unemployed	0.299 (<.001)	0.076 (.230)	0.238 (<.001)	0.165 (.009)	-0.267 (<.001)	0.044 (.486)	-0.290 (<.001)	-0.163 (.009)	-0.052 (.407)								
Pct_IncomeLT10000	0.171 (.006)	0.031 (.623)	-0.055 (.384)	0.135 (.032)	0.042 (.507)	0.195 (.002)	-0.228 (<.001)	-0.138 (.029)	-0.069 (.279)	0.336 (<.001)							
Pct_IncomeGT200000	0.140 (.026)	0.132 (.036)	0.083 (.188)	0.013 (.833)	-0.202 (.001)	-0.391 (<.001)	0.426 (<.001)	0.498 (<.001)	0.374 (<.001)	-0.122 (.052)	-0.355 (<.001)						
Pct_Black_NH	-0.111 (.079)	0.025 (.693)	-0.121 (.056)	0.214 (.001)	-0.165 (.009)	-0.211 (.001)	0.218 (.001)	0.074 (.244)	0.311 (<.001)	0.162 (.010)	0.123 (.052)	0.058 (.358)					
Pct_Asian	0.328 (<.001)	0.052 (.413)	0.069 (.278)	0.236 (<.001)	-0.288 (<.001)	-0.265 (<.001)	0.143 (.024)	0.337 (<.001)	0.284 (<.001)	0.024 (.701)	-0.107 (.090)	0.480 (<.001)	0.278 (<.001)				
Pct_Hispanic	0.520 (<.001)	-0.224 (<.001)	0.427 (<.001)	0.384 (<.001)	-0.389 (<.001)	0.374 (<.001)	-0.718 (<.001)	-0.370 (<.001)	-0.351 (<.001)	0.384 (<.001)	0.215 (.001)	-0.252 (<.001)	-0.391 (<.001)	-0.036 (.574)			
Pct_White_NH	-0.557 (<.001)	0.227 (<.001)	-0.426 (<.001)	-0.500 (<.001)	0.496 (<.001)	-0.304 (<.001)	0.686 (<.001)	0.335 (<.001)	0.254 (<.001)	-0.461 (<.001)	-0.253 (<.001)	0.205 (.001)	0.086 (.173)	-0.151 (.016)	-0.945 (<.001)		
Religion	-0.110 (.083)	0.081 (.202)	0.114 (.072)	-0.178 (.004)	0.143 (.023)	0.162 (.010)	-0.106 (.092)	-0.116 (.065)	-0.302 (<.001)	-0.024 (.700)	0.050 (.432)	-0.235 (<.001)	-0.113 (.074)	-0.161 (.011)	0.109 (.085)	-0.067 (.288)	

Figure 18 - Scatter and Pearson's correlation matrix, for 2022

This way, population that took the COVID-19 vaccine (FullVac_pct) seems to have a global positive relationship with Hispanic population. It seems to have a global negative relationship with White population. It does not seem to be correlated with religion, female, children, Uninsured, college, flu vaccine intake, and Black or African American.

We observe that the correlations of the dependent variable with the explanatory variables that were significant were: Pct_White_NH, Pct_Hispanic, Pct_Asian, Pct_IncomeGT200000, Pct_IncomeLT10000, Pct_Unemployed, Pct_HighSchool, Pct_Elderly and Pct_Young.

5.4.2 Local Relationships

The local bivariate relationships maps show statistically significant relationships using local entropy. These maps are crucial for selecting the variables because some of them might only be relevant locally rather than globally. It would be inappropriate to use a variable if a linear relationship did not exist. Appendix DAppendix shows all local bivariate relationships maps, except those where none of the counties had a significant relationship.

The local relationships between COVID-19 vaccination and female population are not significant in all counties. This fact explains the apparent inexistent relationship in the scatterplot graph of the FullVac and the Pct_Female variables. The same happens for Children, Young, unemployment, Income < 10 000, Black or African American and Religion. Therefore, these variables were not included in the MGWR models.

For the variable of percentage of **elderly** population, the local relationship is negative linear in 28 counties and convex in 70 counties. As for the of percentage of **uninsured population**, the local relationship is positive linear in 1 county, negative linear in 17 counties and convex in 26 counties. This variable is not relevant in the whole study area (r = 0.027) but relevant locally. For the percentage of population with a **high school** degree, the local relationship is positive linear for 1 county, concave for 11 counties and convex for 14 counties. For the percentage of population that has some **college degree**, the local relationship is positive linear in 58 counties, negative linear in 17 counties, concave in 9 counties and convex in 11 counties. For the variable of percentage of population that took the **flu vaccine**, the local relationship is positive linear in

18 counties and convex in 31 counties. This variable is not relevant in the whole study area (r = 0.088), but relevant locally. For the percentage of population that has an annual **income higher than 200 000** dollars, the local relationship is positive linear in 61 counties, negative linear in 4 counties and convex in 6 counties.
All these variables described above can be excluded from OLS but not from MGWR because locally they are relevant.

As expected, form the scatter and Pearson's correlation matrix displayed above (**Error! Reference source not found.**), population that took the COVID-19 vaccine shows a moderate positive relation (r=0.328) with the percent of **Asian** population in most counties. The local relationship is positive linear in 68 counties and concave in 41 counties. Population that took the COVID-19 vaccine shows a moderate positive relation (r = 0.52) with the percent of **Hispanic** population in most counties. The local relationship is positive linear in 26 counties and concave in 2 counties. Finally, population that took the COVID-19 vaccine shows a negative moderate relation (r = - 0.557) with the percent of **White** population in most counties. The local relationship is Negative linear in 20 counties and convex in 11 counties.

Variable	Negative linear	Positive linear	Convex	Concave
Pct_Female	0	0	0	0
Pct_Children	0	0	0	0
Pct_Young	0	0	0	0
Pct_Unemployment	0	0	0	0
Pct_IncomeLT10000	0	0	0	0
Pct_Black_NH	0	0	0	0
Religion	0	0	0	0
Pct_HighSchool	1	0	14	11
Pct_Uninsured	1	17	26	0
Pct_Elderly	0	28	70	0
Pct_SomeCollege	17	58	11	9
Pct_FluVac	0	18	31	0
Pct_IncomeGT200000	4	61	6	0
Pct_Asian	0	68	0	41
Pct_Hispanic	0	26	0	0
Pct_White_NH	20	0	11	0

Table 8 - Number of counties with significant relationships

To conclude, as we can see in **Table 8 - Number of counties with significant relationships**, the variables of Pct_Female, Pct_Children, Pct_Young, Pct_unemployment, Pct_IncomeLT10000, Pct_Black and Religion should not be included in linear regression models such as MGWR, because they do not exhibit significant linear relationships. Additionally, after carefully observing all the maps (presented on the Appendix D) we also concluded that would be best for our model if we don´t include the variable of Pct_HighSchool since only a few counties show significant linear relationships with the dependent variable. This way, we proceeded the analysis with the remaining eight variables since they show a reasonable number of counties with significant linear relationships.

5.5 Global Multicollinearity

The global multicollinearity analysis regarding the eight potential explanatory variables we observed that the variables Pct_Hispanic and Pct_White_NH had high VIF values, exceeding 4 (15.709594 and 17.771529, respectively). In this case, these values suggested that they may be redundant or highly correlated with other variables in the model, potentially leading to unstable coefficient estimates. This way, the Pct_White_NH variable was then eliminated, because it has a significant linear relationship only in 20 counties. The VIF values were then computed for models with the remaining seven potential explanatory variables, solving this way the problem of multicollinearity and obtained good results as showed in **Table 9 - Global multicollinearity Results for 2022 data**.

Table 9 - Global multicollinearity Results for 2022 data summarises the results of the global multicollinearity analysis regarding the seven independent variables to be included in the MGWR model.

Table 9 - Global multicollinearity Results for 2022 data

Variable	VIF
Intercept	-----------------------
Pct_Elderly	1.490489
Pct_Uninsured	1.428597
Pct_SomeCollege	1.570909
Pct_FluVac	1.401662
Pct_IncomeGT200000	1.707543
Pct_Asian	1.426944
Pct_Hispanic	1.796068

The VIF values of each predictor are all lower than 4 so there is no relation of multicollinearity among them. These low VIF values suggest that their coefficient estimates are likely stable and reliable.

5.6 MGWR MODEL

As demonstrated above by the local bivariate analysis certain relations may function over greater scales than others. This behavior is typical of most spatially heterogeneous processes. In order to model this, each predictor can have a unique bandwidth identified by MGWR, allowing for different spatial scales in the relationships between each predictor and the dependent variable. The MGWR model was estimated with the seven independent variables listed in **Table 9 - Global multicollinearity Results for 2022 data**. When we did the analysis in ArcGIS, there is a warning message that says that it had problems reading 2 out of the 254 total records. Since it was not possible to correct this error, we considered that 100% of the counties corresponds to 252 counties in the following analysis and model results.

In this section we examined the individual bandwidth and summary statistics of the local coefficients of each explanatory variable (**Table 10 - Individual bandwidth and summary statistics of the local coefficients of each explanatory variable for 2022).** The local bandwidths of the MGWR showed that the relationship between Asian population (Pct_Asian), population with annual income higher than 200 000 (Pct_IncomeGT200000), population that took the flu vaccine (Pct_FluVac) and population that has more than 65 years old (Pct_Elderly) with COVID-19 incidence rate is global, because the bandwidth of that variable included all 252 counties. As for Pct_Uninsured, Pct_SomeCollege and Pct_Hispanic, they have a more local relationship (the bandwidth includes 11.9%, 19.44% and 19.84% of the neighboring counties, respectively). As for the intercept, we can see that it also operates at a local scale since the bandwidth includes 37.7% of the nearest counties.

Variable (Scaled)	Bandwidth (nr. and % of counties)	Nr. (%) of counties with significant coefficients	Mean	Standard Deviation	Minimum	Median	Maximum
Intercept	95 (37.7%)	91 (36.11%)	-0.0552	0.2305	-0.4588	0.0283	0.2272
Pct_Elderly	252 (100%)	252 (100%)	0.1541	0.0046	0.1435	0.1536	0.1684
Pct_Uninsured	30 (11.9%)	32 (12.7%)	-0.0985	0.2555	-0.8371	-0.1459	0.7542
Pct_SomeCollege	49 (19.44%)	10 (3.97%)	0.0488	0.1214	-0.1794	0.0392	0.4135
Pct_FluVac	252 (100%)	231 (91.67%)	0.1164	0.0132	0.0833	0.1179	0.137
Pct_IncomeGT200000	252 (100%)	0 (0.00%)	0.0788	0.0123	0.0439	0.0849	0.0916
Pct_Asian	252 (100%)	252 (100%)	0.2286	0.0062	0.2221	0.2256	0.2459
Pct_Hispanic	50 (19.84%)	154 (61.11%)	0.5490	0.2065	0.0785	0.5937	-0.8733

Table 10 - Individual bandwidth and summary statistics of the local coefficients of each explanatory variable for 2022

As we can see on the local coefficients variability alongside the bandwidths of **Table 10 - Individual bandwidth and summary statistics of the local coefficients of each explanatory variable for 2022**, high bandwidths result in less variability in the coefficients. This way, the local coefficients of Pct_Elderly have the smallest range of values (Maximum–Minimum=0.0249) and standard deviation (0.0046). The same happens for the other variables of high bandwidths: Pct_Asian (Maximum–Minimum=0.0238) and standard deviation (0.0062). Therefore, we can reach to the conclusion that the influence of being elderly and being Asian on COVID-19 vaccine incidence rate is effectively stationary over space.

Since the dependent and independent variables were standardized, the values of the coefficients can be directly compared. On average, Pct_Hispanic is the most influential factor (mean= 0.5490) and Pct_SomeCollege is the least one (mean= 0.0488).

5.6.1 LOCAL MULTICOLLINEARITY

Local multicollinearity in the MGWR model was diagnosed using Local Condition Numbers. Since every value is below the rule of thumb of 30, none of the local condition numbers in this MGWR model indicate that local multicollinearity is a problem, as showed in **Figure 19 - Local Multicollinearity, 2022**.

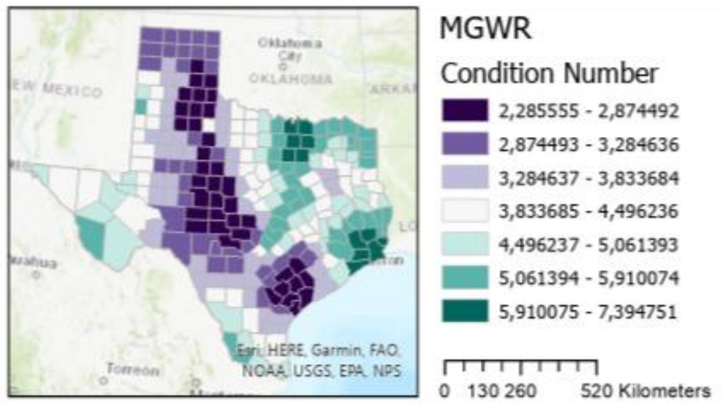


Figure 19 - Local Multicollinearity, 2022

5.6.2 INFLUENTIAL OBSERVATIONS

This section demonstrates the application of Cook's Distance, a metric that quantifies the impact of an observation on the calibration of the model. All the values in the map of **Figure 20 -**

Influential Observations, 2022Figure 20 are below 1. This way, there are no influential observations that could represent a problem for the MGWR model.

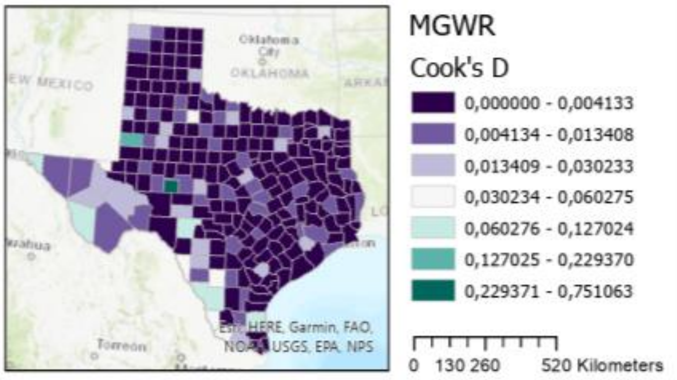


Figure 20 - Influential Observations, 2022

5.6.3 GOODNESS-OF-FIT OF THE FINAL MGWR MODEL

The final MGWR model only includes seven explanatory variables (Pct_Elderly, Pct_Uninsured, Pct_FluVac, Pct_SomeCollege, Pct_IncomeGT200000, Pct_Hispanic, and Pct_Asian) and its performance (AICc = 43573.58) is better than the corresponding GWR model (AICc = 477.3245 as shown in **Table 11 - Goodness-of-fit of the final MGWR model, 2022** below. The single optimal bandwidth obtained in the GWR calibration is 118 counties. In MGWR, Pct_Elderly, Pct_IncomeGT200000, Pct_FluVac and Pct_Asian operated at a global scale (all 252 counties), no variable at a regional scale and Pct_Uninsured, Pct_SomeCollege and Pct_White_NH at a local scale with a bandwidth of 30 (11.9%), 49 (19.44%) and 50 (19.84%) counties, respectively. As (Yu et al., 2020) point out, these results are informative because they demonstrate the dissimilarity between the GWR and MGWR results when the optimal bandwidths of the predictors are not identical to that of the single GWR bandwidth.

Model	R2	Adj – R2	AICc	Effective degrees of freedom
GWR with 7 predictors	0.7276	0.6636	477.3245	204.2724
MGWR with 7 predictors	**0.8071**	**0.7566**	**435.7358**	**199.8781**

Table 11 - Goodness-of-fit of the final MGWR model, 2022

The map of Local R2 values in **Figure 21 - Local R2 Map, 2022Error! Reference source not found.** show that MGWR has good explanatory capability in most of the counties, particularly in the east and west region. However, the model has a poorer fit in the northern region, where clusters of higher vaccine hesitancy were identified (i.e., cold-spots of the vaccine incidence rate). This may indicate that the model is missing a relevant explanatory variable in that region.

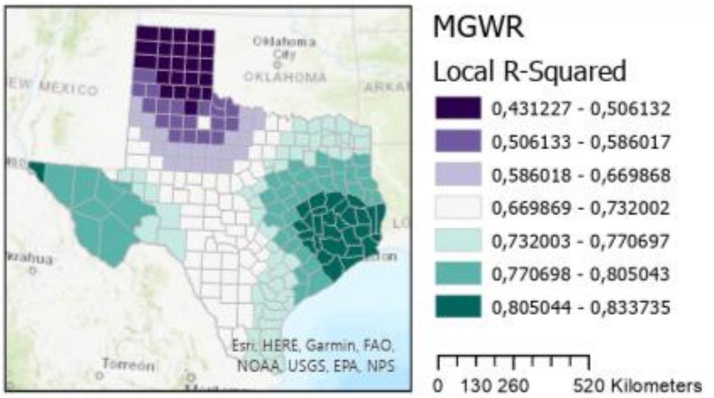


Figure 21 - Local R2 Map, 2022

5.6.4 Residual analysis of the final MGWR model

It is essential to determine if the standardized residuals of the MGWR model (**Figure 22 - Standardized residuals of the MGWR model, 2022**) exhibit a random distribution throughout space.

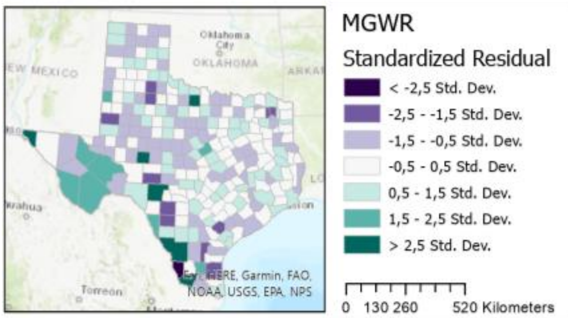


Figure 22 - Standardized residuals of the MGWR model, 2022

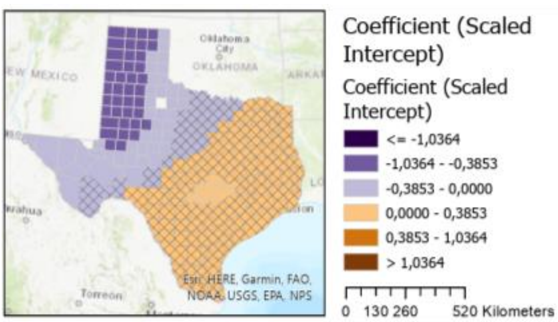
We use the spatial autocorrelation (Global Moran's I) test to access the spatial autocorrelation in the standardized residuals of the MGWR model. In our case, the Global Moran's value is equal to -0.065804. The p-value is 0.049531, which is lower than the usual significance level of 0.05. The null hypothesis of spatial randomness is rejected by this, indicating a statistically significant spatial autocorrelation for the standardized residuals across the counties of Texas. In other words, the standardized residuals pattern appears to be significantly different than random with 95% confidence. However, if the confidence level is increased to 96% (significance level of 0.04), then there is not enough evidence of spatial autocorrelation.

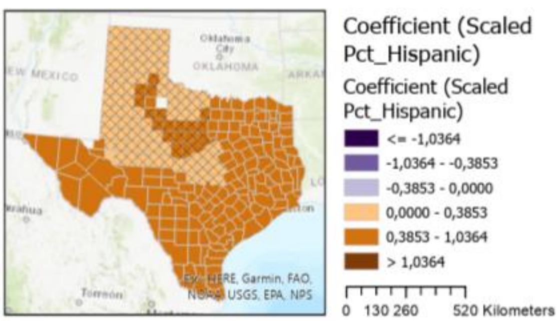
5.6.5 Local coefficients of the final MGWR model

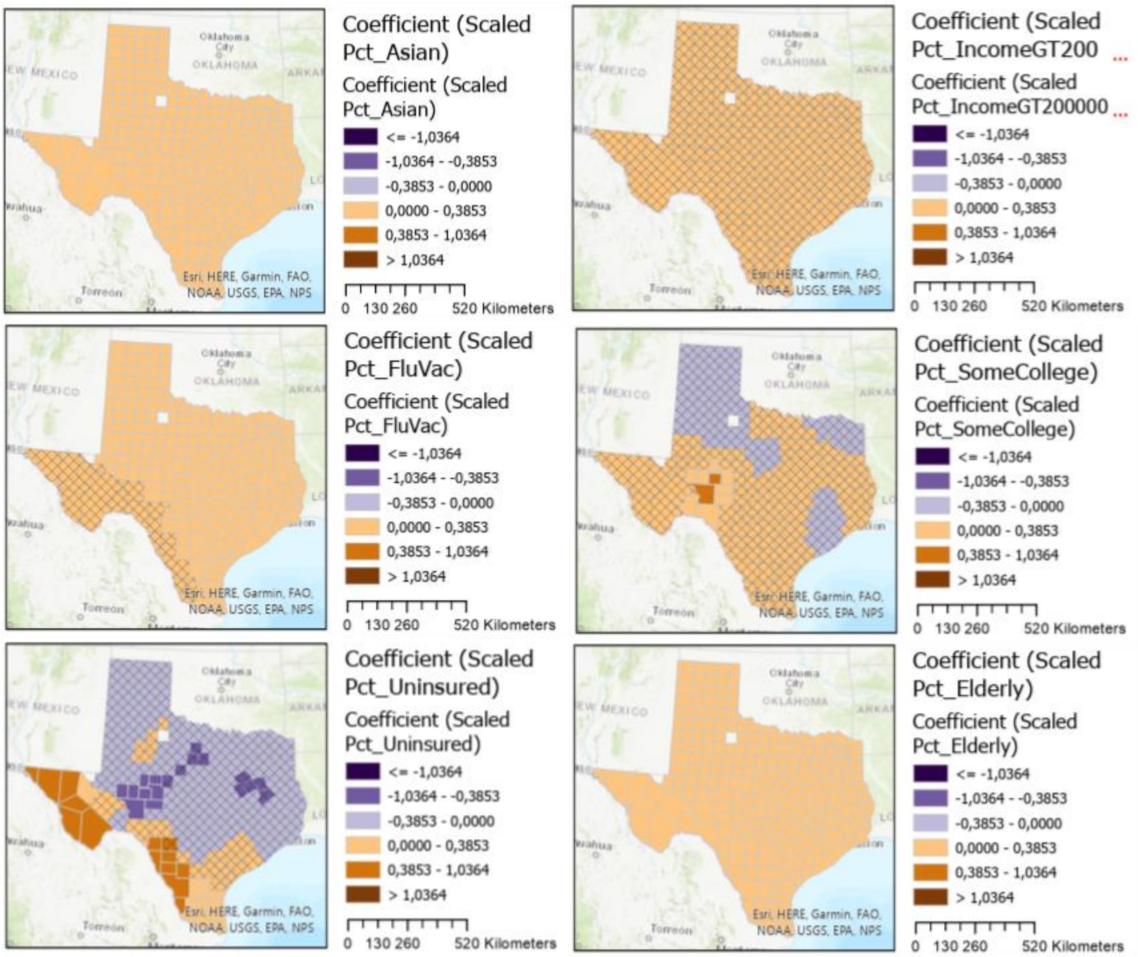
The maps of the local coefficients of the intercept and the seven predictors of the MGWR model are disclosed below. As expected from the results of the initial MGWR model (**Table 10 - Individual bandwidth and summary statistics of the local coefficients of each explanatory variable for 2022**), the coefficients of Pct_FluVac, Pct_Asian and Pct_Elderly are all positive and significantly different from zero in all 252 counties, except Pct_FluVac that is significant in 231 counties. These coefficients exhibit very little variability because the variable has a global effect on the COVID-19 vaccine incidence rate in Texas. In northern counties, where the model has the poorest fit, the only significant coefficients are the ones of Pct_Asian, Pct_Elderly and Pct_FluVac.

In the local maps presented below, we can observe that Pct_IncomeGT200000 is positively associated with COVID-19 vaccine intake in all the area of the state, but not significantly. Asian and Elderly they are positively associates with COVID-19 incidence rate in all areas of the state, FluVac as well except in the southwest region of the state, close to the border. Pct_Hispanic is significant and positively associated with the COVID-19 vaccine intake in all areas except for the northern region. Finally, Pct_Uninsured is significant and positively associated with the vaccine intake in the western region close to the border of the state and very negative in some areas in the centre.

2022: Local coefficients of the MGWR model







Coefficient (Scaled Pct_Asian)
Coefficient (Scaled Pct_Asian)
<= -1,0364
-1,0364 - -0,3853
-0,3853 - 0,0000
0,0000 - 0,3853
0,3853 - 1,0364
> 1,0364
0 130 260 520 Kilometers
Coefficient (Scaled Pct_IncomeGT200 ...
Coefficient (Scaled Pct_IncomeGT200000 ...
<= -1,0364
-1,0364 - -0,3853
-0,3853 - 0,0000
0,0000 - 0,3853
0,3853 - 1,0364
> 1,0364
0 130 260 520 Kilometers
Coefficient (Scaled Pct_FluVac)
Coefficient (Scaled Pct_FluVac)
<= -1,0364
-1,0364 - -0,3853
-0,3853 - 0,0000
0,0000 - 0,3853
0,3853 - 1,0364
> 1,0364
0 130 260 520 Kilometers
Coefficient (Scaled Pct_SomeCollege)
Coefficient (Scaled Pct_SomeCollege)
<= -1,0364
-1,0364 - -0,3853
-0,3853 - 0,0000
0,0000 - 0,3853
0,3853 - 1,0364
> 1,0364
0 130 260 520 Kilometers
Coefficient (Scaled Pct_Uninsured)
Coefficient (Scaled Pct_Uninsured)
<= -1,0364
-1,0364 - -0,3853
-0,3853 - 0,0000
0,0000 - 0,3853
0,3853 - 1,0364
> 1,0364
0 130 260 520 Kilometers
Coefficient (Scaled Pct_Elderly)
Coefficient (Scaled Pct_Elderly)
<= -1,0364
-1,0364 - -0,3853
-0,3853 - 0,0000
0,0000 - 0,3853
0,3853 - 1,0364
> 1,0364
0 130 260 520 Kilometers
Esri, HERE, Garmin, FAO, NOAA, USGS, EPA, NPS

6. FINAL DISCUSSION

In this section we will discuss our overall results for both years of analysis (2022 and 2021) and make a comparative discussion with other previous made studies.

The purpose of this GIS-based study was to examine the spatial variability of COVID-19 complete vaccination rates throughout all counties in the state of Texas.

Critical steps in reducing the transmission of SARS-CoV2 include vaccination, face mask use, sanitization, and social distancing. Numerous pharmaceutical companies created COVID-19 vaccines, and immunization campaigns started in countries. However, partisanship, vaccine knowledge, vulnerability to COVID-19, politics and COVID-19 risk factors all influence vaccination intention (Ruiz et al., 2021). To discover what specific factors influence vaccine intention is the main reason we performed this study. According (Lazarus et al., 2021), vaccination intentions differ from county to county as well. Overall, our study showed that, on average, for all counties of Texas, 47.26% of the individuals took the complete vaccination for 2022 and 8,18% for 2021.

From what we can see in the local coefficient's maps of the final MGWR model for 2021, we can take spatially significant conclusions on three variables: Uninsured, annual income greater than 200 000 and Democrat. All of those variables have a local relation except for the variable of democrats that has a global relation. Additionally, we can see that for 2022 there are seven explanatory variables with significant coefficients: Uninsured, College degree, Hispanic, Asian, Flu Vaccine and elderly. The results of the MGWR model, which was the best-fitting model in this study, indicated that the model fitted well in most areas of south and east for 2021 and most areas of east, south and a small part of west for 2022. But in most counties of the north of the state, all of which had reported lower vaccination rates, the model was primarily under-fitted, for both years.

Our findings showed no significance between county-level vaccination rates and the proportion of people with a high school diploma, for both years. These results are consistent with (Mollalo et al., 2021).

For the variable of **Pct_Uninsured**, our findings indicated low positive coefficients in the southwest close to the border of the state for 2021. As for 2022, the results were even more positive around the southwest border. This might be due to the fact that the higher the number of uninsured individuals, the lower the access to healthcare centers in case of illness. In this case the vaccine was free of charge (U.S Department of Health and Human Services, n.d.), so it makes sense that they took the vaccine if uninsured to decrease the risk of other health problems. Despite this, still in 2022, we also observed that a small proportion of the center had high negative values.

Mollalo et al., 2021, found that the Uninsured rate was negatively associated with vaccination rates, which does not go in accordance with what we found. Furthermore, (Lindemer et al., 2021) indicates that the vaccination coverage has been noticeably slower in counties with lower insurance coverage. Even though efficient preventive health services may reduce concerns, uninsured adults receive significantly lower preventive services than insured people (Holden et al., 2015).

For the variable of **annual income greater than 200 000**, our findings indicated that it was positively associated with COVID-19 vaccine incidence rate in Texas on the southeast region of the state in 2021. These results were in accordance with the ones found by (Biswas et al., 2021; Mollalo et al., 2021). However, in our case, this is only true for the year of 2021 since for 2022 the variable was not significant for any county within the state area. Mollalo et al (2021) found that per capita income was positively associated with vaccination rates. However, the strengths of associations significantly varied by geographic location. Additionally, Biswas et al (2021) also concluded that higher income individuals demonstrated greater acceptance of vaccination due to their increased sense of the risk of contracting the virus. These individuals also more informed about the effectiveness of the vaccine intake.

Regarding ethnicity for the year of 2022, our data identified a statistically significant association between **Asian** and an increased likelihood to obtain the COVID-19 vaccine in the entire state area. As for Hispanic ethnicity, we found a high positive likelihood to obtain the COVID-19 vaccine in all areas of the state expect for the north.

For 2021, local statistical significance was not retained in the final model for both ethnicity variables, White and Asian. This finding was surprising to us, we had hypothesized that there would be an association between COVID-19 vaccine hesitancy and ethnicity for both years based on previous research (Liu et al., 2021).

This way, our findings are in accordance with (Liu et al., 2021) in a way that we found significant racial differences in hesitancy towards COVID-19 vaccines. However, we were not in accordance with the fact that that Black Americans were less likely to accept a potential

COVID-19 vaccine (Liu et al., 2021; Malik et al., 2020; Pallathadka et al., 2023), since we found no significance in this variable in the local relationships analyses. The reasons for this to happen are beyond the scope of this study but we recon we might be due to the fact that the first year of our data collection is 2021 and by then racial disparities among black or African American individuals are less when compared to studies with data from earlier years.

For the variable of **Pct_Democrat2020**, our findings indicated that it was positively associated with COVID-19 vaccine incidence rate in Texas on the entire state equally, for 2021. We did not include political choice in the analysis of 2022 nor other potential explanatory variables related to politics. Those findings agree with the ones of (Lee et al., 2022). These authors found that vaccine hesitancy was lower where the relative number of voters for Biden (Democratic party) in the 2020 presidential election is higher, even after accounting for a wide range of sociodemographic characteristics.
(Liu et al., 2021), also agreed in the fact that state partisanship and the percentage of people who are vaccine hesitant were shown to be strongly correlated. This supports our hypothesis that the differences in vaccination rates and attitudes are a direct reflection of partisan differences. Furthermore, unlike some other differences, the partisan divide did not close over time. However, in our case, this could not be verified due to one of the limitations of our study (i.e., lack of political data for 2022).

For the variable of **Pct_FluVac** we found that it was significant and positively associated with the COVID-19 vaccine incidence rate in all the state except on the southwest close to the border, for 2022. For the year of 2021 there was no significance found in any county of the state. These results were in accordance with the ones found by (Gomes et al., 2022).

Finally, for **Pct_Elderly** we also only found significance for the year of 2022. Within the whole area of the state, it was positively associated with the COVID-19 vaccine intake. Our findings for 2021 are consistent with some studies that found no association (Biasio et al., 2021; Harapan et al., 2020). The results of 2022 were in accordance with (Gomes et al., 2022; Litaker et al., 2022) who stated that, people older than 65 years had the greatest percentage of full-dose vaccine coverage rate. Moreover, we found that the Elderly variable only had a significant effect in COVID-19 vaccination incidence in 2022. (Soares et al., 2021) found that younger individuals had higher odds of vaccine refusal. This author also found that older individuals were more likely to take the vaccine as soon as possible. Additionally, some authors pointed out that younger people are reluctant to receive the COVID-19 vaccine because they do not believe they would be affected by it, while elderly people are reluctant because they do not feel safe about what might happen after immunization (Kourlaba et al., 2021).

There are several limitations within this study. The disease is highly dynamic due to factors like the highly contagious delta variant and ongoing fluctuations in vaccination rates. This way, more recent data from follow-up studies are needed to give policymakers the latest information possible when combating the disease. In addition, it is advised that in future research, environmental, demographic, and health-related variables be included together with higher-resolution geographical analysis carried out at several scales. Further research is necessary on the local healthcare facilities and social resources that may have an impact on immunization rates in various counties. We also only considered three databases for this study (Google scholar Scopus and PubMed). There are other databases that we did not explore. Also, for the variables of Democrat and Republican voters we were only able to retrieve data from the 2020 presidential election (for the year of 2021; we did not include the politics variables in the 2022 analysis).

7. CONCLUSIONS AND FUTURE WORKS

This thesis had the objective to investigate the spatial dimension of socioeconomic and demographic factors behind COVID-19 vaccine hesitancy, in every Texas county, as measured by the incidence rate of COVID-19 vaccine intake.

High monthly income individuals were found to be more open to receive vaccinations, in our case for the year of 2021, which agrees with previous studies (Yang et al., 2021). According to the research, those with higher incomes are more conscious of the harmful effects of COVID-19 and desire to receive the immunization in order to protect their health (Williams et al., 2021). Therefore, a conclusion/recommendation we reached for future works was that future public health campaigns should focus more intently on lower-income populations. These campaigns should concentrate on raising vaccination awareness and solving obstacles that these people may encounter. Future studies should also look into the underlying causes of the vaccine hesitancy of those with lower incomes. Qualitative research may be used to investigate misinformation's impact, healthcare access, and socioeconomic factors. Having a better understanding of these factors can support the creation of interventions that are more fair in their approach based on the needs of various income groups. Additionally, longitudinal research should be done to determine whether these patterns remain true over time and in a pandemic context or health emergency. This would offer a more thorough comprehension of the relationships between health habits and income levels, guiding the development of future policies and health communication plans.

As for uninsured individuals, we found positive relationships close to the southwest border, for both years. Health insurance plans have the potential to enhance health outcomes and assist members in managing their health. Future efforts should focus on targeted outreach and improving healthcare access in these central areas to address the disparity and ensure more equitable vaccination coverage across the state (Mollalo et al., 2021).

As for the observed racial disparities there are also some strategies that could be done to decrease this problem. The main base is to increase the transparency of the processes involved in developing and approving vaccines. If accessible information is provided about the efficacy and side effects of vaccines, offer free vaccinations regardless of insurance status, expand vaccination sites, develop more accessible scheduling systems, and carry out other community-specific engagement initiatives, racial disparities could be effectively reduced. Finally, the examination of ethnic variables showed that, for the entire state in 2022, there was statistical significance between the Asian population and COVID-19 vaccine intake percentage. Similarly, throughout Texas, with the exception of the northern areas, Hispanic people demonstrated a high positive likelihood of immunization. These results demonstrate the efficacy of immunization efforts aimed at these ethnic groups, however more research is necessary given the lower Hispanic uptake in the northern regions. Nevertheless, in the 2021 model, the variables of White and Asian ethnicities lost their local statistical significance, indicating that the influence of ethnicity on vaccination rates might have changed over time. Future studies should investigate the fundamental causes of these differences and create customized plans to tackle the distinct requirements and obstacles encountered by many ethnic groups, with a special emphasis on areas with reduced vaccination rates (Liu et al., 2021).

As for Elderly, we concluded that it only had influence in 2022. This variable showed a positive statistical correlation with COVID-19 vaccine uptake for the entire state. Looking for future suggestions, we found that they should focus on preserving and increasing vaccination campaigns for this population. It is specifically advised to keep using campaigns that focus on the special requirements of the elderly, like making immunization venues easily accessible, providing transportation, and guaranteeing that the advantages and safety of vaccines are communicated clearly. Furthermore, enhancing vaccination uptake among older individuals can guarantee their continued protection against COVID-19 and other illnesses.

Additionally, future initiatives should concentrate on more successfully marketing the advantages of vaccination in underserved areas, given the notable positive correlation found in 2022 between flu vaccine and COVID-19 vaccine incidence rates across most of the state, with the exception of the southwest border region. Public health initiatives that are specifically designed to highlight the mutually beneficial protection provided by the COVID-19 and flu vaccines could potentially boost vaccination rates in areas where rates are now lower. Moreover, community involvement and targeted interventions will be essential to removing obstacles to vaccination acceptability and access in the southwest border region in order to guarantee more equal health results throughout the entire state.

Finally, the findings for the democratic party voter variable highlight the ways in which political decisions affect the public's confidence in government policies and information. Increasing media coverage and combating misinformation would increase public knowledge of the scope and effects of the COVID-19 pandemic. These results also demand more focused policies and tactics to raise immunization rates in accordance with regional and social settings.

In conclusion, unlike other articles, which mainly use analysis based on survey responses, our analysis used official vaccination records and official data on demographic, socioeconomic, political factors. Therefore, even the largest surveys ever undertaken for COVID-19 vaccination have been found to be affected by survey sampling bias, which should not apply to our findings. Since there isn't much research on COVID-19 vaccine spatial modeling in Texas, as far as the authors are aware, this current work can act as a geospatial reference to help public health decision-makers create region-specific vaccinations monitoring programs.

BIBLIOGRAPHICAL REFERENCES

Acharya, B. K., Cao, C. X., Lakes, T., Chen, W., Naeem, S., & Pandit, S. (2018). Modeling the spatially varying risk factors of dengue fever in Jhapa district, Nepal, using the semi-parametric geographically weighted regression model. *International Journal of Biometeorology*, *62*(11), 1973–1986. doi: 10.1007/s00484-018-1601-8

Adhikari, B., & Cheah, P. Y. (2021). Vaccine hesitancy in the COVID-19 era. *The Lancet Infectious Diseases*, *21*(8), 1086. doi: 10.1016/S1473-3099(21)00390-X

Agrawal, H., Das, N., Nathani, S., Saha, S., Saini, S., Kakar, S. S., & Roy, P. (2021). An Assessment on Impact of COVID-19 Infection in a Gender Specific Manner. In Stem Cell Reviews and Reports (Vol. 17, Issue 1, pp. 94–112). Springer. doi: 10.1007/s12015-020-10048-z

Aguolu, O. G., Malik, A. A., Ahmed, N., & Omer, S. B. (2022). Overcoming Vaccine Hesitancy for Future COVID-19 and HIV Vaccines: Lessons from Measles and HPV Vaccines. *Current HIV/AIDS Reports*, *19*(5), 328–343. doi: 10.1007/s11904-022-00622-0

Albrecht, D. (2022). Vaccination, politics and COVID-19 impacts. *BMC Public Health*, *22*(1). doi: 10.1186/s12889-021-12432-x

Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., & Yang, D. (2020). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. *Journal of Public Economics*, *191*. doi: 10.1016/j.jpubeco.2020.104254

Alley, S. J., Stanton, R., Browne, M., To, Q. G., Khalesi, S., Williams, S. L., Thwaite, T. L., Fenning, A. S., & Vandelanotte, C. (2021). As the pandemic progresses, how does willingness to vaccinate against covid-19 evolve? *International Journal of Environmental Research and Public Health*, *18*(2), 1–14. doi: 10.3390/ijerph18020797

Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, *27*(2), 93–115. doi: 10.1111/j.1538-4632.1995.tb00338.x

Baker, J. P. (2008). Mercury, vaccines, and autism: One controversy, three histories. In American Journal of Public Health (Vol. 98, Issue 2, pp. 244–253). doi: 10.2105/AJPH.2007.113159

Barello, S., Nania, T., Dellafiore, F., Graffigna, G., & Caruso, R. (2020). 'Vaccine hesitancy' among university students in Italy during the COVID-19 pandemic. *European Journal of Epidemiology*, *35*(8), 781–783. doi: 10.1007/s10654-020-00670-z

Bertin, P., Nera, K., & Delouvée, S. (2020). Conspiracy Beliefs, Rejection of Vaccination, and Support for hydroxychloroquine: A Conceptual Replication-Extension in the COVID-19 Pandemic Context. *Frontiers in Psychology*, *11*. doi: 10.3389/fpsyg.2020.565128

Biasio, L. R., Bonaccorsi, G., Lorini, C., & Pecorelli, S. (2021). Assessing COVID-19 vaccine literacy: a preliminary online survey. *Human Vaccines and Immunotherapeutics*, *17*(5), 1304–1312. doi: 10.1080/21645515.2020.1829315

Biswas, M. R., Alzubaidi, M. S., Shah, U., Abd-Alrazaq, A. A., & Shah, Z. (2021). A scoping review to find out worldwide covid-19 vaccine hesitancy and its underlying determinants. In Vaccines (Vol. 9, Issue 11). MDPI. doi: 10.3390/vaccines9111243

Brunsdon, C., Fotheringham, A. S., & Charlton, M. (2002). *Geographically weighted summary statistics-a framework for localised exploratory data analysis*. Retrieved from www.elsevier.com/locate/compenvurbsys

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, *28*(4), 281–298. doi: 10.1111/j.1538-4632.1996.tb00936.x

Callaghan, T., Moghtaderi, A., Lueck, J. A., Hotez, P., Strych, U., Dor, A., Fowler, E. F., & Motta, M. (2021). Correlates and disparities of intention to vaccinate against COVID-19. *Social Science and Medicine*, *272*. doi: 10.1016/j.socscimed.2020.113638

Chakraborty, C., Sharma, A. R., Bhattacharya, M., Agoramoorthy, G., & Lee, S. S. (2021). All nations must prioritize the COVID-19 vaccination program for elderly adults urgently. In Aging and Disease (Vol. 12, Issue 3, pp. 688–690). International Society on Aging and Disease. doi: 10.14336/AD.2021.0426

DiRago, N. V., Li, M., Tom, T., Schupmann, W., Carrillo, Y., Carey, C. M., & Gaddis, S. M. (2022). COVID-19 Vaccine Rollouts and the Reproduction of Urban Spatial Inequality: Disparities Within Large US Cities in March and April 2021 by Racial/Ethnic and Socioeconomic Composition. *Journal of Urban Health*, *99*(2), 191–207. doi: 10.1007/s11524-021-00589-0

Dubé, E., Laberge, C., Guay, M., Bramadat, P., Roy, R., & Bettinger, J. (2013). Vaccine hesitancy: An overview. In Human Vaccines and Immunotherapeutics (Vol. 9, Issue 8, pp. 1763–1773). doi: 10.4161/hv.24657

Dubé, E., Vivion, M., & MacDonald, N. E. (2014). Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: Influence, impact and implications. In Expert Review of Vaccines (Vol. 14, Issue 1, pp. 99–117). Expert Reviews Ltd. doi: 10.1586/14760584.2015.964212

Ehde, D. M., Roberts, M. K., Herring, T. E., & Alschuler, K. N. (2021). Willingness to obtain COVID-19 vaccination in adults with multiple sclerosis in the United States. *Multiple Sclerosis and Related Disorders*, *49*. doi: 10.1016/j.msard.2021.102788

Faisal, K., Alshammari, S., Alotaibi, R., Alhothali, A., Bamasag, O., Alghanmi, N., & Bin Yamin, M. (2022). Spatial Analysis of COVID-19 Vaccine Centers Distribution: A Case Study of the City of Jeddah, Saudi Arabia. *International Journal of Environmental Research and Public Health*, *19*(6). doi: 10.3390/ijerph19063526

Fattah, A., Mohammadtaghizadeh, M., & Azadi, H. (2022). Factors Associated with COVID-19 Vaccine Acceptance Worldwide: A Rapid Review. *Med Edu Bull*, *3*(7). doi: 10.22034/MEB.2021.318247.1040

Fotheringham, A. S., & Oshan, T. M. (2016). Geographically weighted regression and multicollinearity: dispelling the myth. *Journal of Geographical Systems*, *18*(4), 303–329. doi: 10.1007/s10109-016-0239-5

Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale Geographically Weighted Regression (MGWR). *Annals of the American Association of Geographers*, *107*(6), 1247–1265. doi: 10.1080/24694452.2017.1352480

Gao, S., Mioc, D., Anton, F., Yi, X., & Coleman, D. J. (2008). Online GIS services for mapping and sharing disease information. *International Journal of Health Geographics*, *7*. doi: 10.1186/1476-072X-7-8

Getis, A., & Ord, J. K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, *24*(3), 189–206. doi: 10.1111/j.1538-4632.1992.tb00261.x

Gomes, I. A., Soares, P., Rocha, J. V., Gama, A., Laires, P. A., Moniz, M., Pedro, A. R., Dias, S., Goes, A. R., Leite, A., & Nunes, C. (2022). Factors Associated with COVID-19 Vaccine Hesitancy after Implementation of a Mass Vaccination Campaign. *Vaccines*, *10*(2). doi: 10.3390/vaccines10020281

Grammich, C., Dollhopf, E. J., Gautier, M. L., Houseal, R., Jones, D. E., Krindatch, A., Stanley, R., Thumma, S., Dollhopf, G., Houseal, G., Krindatch, J., & Thumma, S. (2023). *2020 U.S. Religion Census*. Retrieved from https://www.thearda.com

Guidry, J. P. D., Laestadius, L. I., Vraga, E. K., Miller, C. A., Perrin, P. B., Burton, C. W., Ryan, M., Fuemmeler, B. F., & Carlyle, K. E. (2021). Willingness to get the COVID-19

vaccine with and without emergency use authorization. *American Journal of Infection Control*, *49*(2), 137–142. doi: 10.1016/j.ajic.2020.11.018
Guo, D. (2010). Local entropy map: A nonparametric approach to detecting spatially varying multivariate relationships. *International Journal of Geographical Information Science*, *24*(9), 1367–1389. doi: 10.1080/13658811003619143
Harapan, H., Wagner, A. L., Yufika, A., Winardi, W., Anwar, S., Gan, A. K., Setiawan, A. M., Rajamoorthy, Y., Sofyan, H., & Mudatsir, M. (2020). Acceptance of a COVID-19 Vaccine in Southeast Asia: A Cross-Sectional Study in Indonesia. *Frontiers in Public Health*, *8*. doi: 10.3389/fpubh.2020.00381
He, Y., Seminara, P. J., Huang, X., Yang, D., Fang, F., & Song, C. (2023). Geospatial Modeling of Health, Socioeconomic, Demographic, and Environmental Factors with COVID-19 Incidence Rate in Arkansas, US. *ISPRS International Journal of Geo-Information*, *12*(2). doi: 10.3390/ijgi12020045
Hildreth, J. E. K., & Alcendor, D. J. (2021). Targeting covid-19 vaccine hesitancy in minority populations in the us: Implications for herd immunity. In Vaccines (Vol. 9, Issue 5). MDPI AG. doi: 10.3390/vaccines9050489
Holden, C. D., Chen, J., & Dagher, R. K. (2015). Preventive care utilization among the uninsured by race/ethnicity and income. *American Journal of Preventive Medicine*, *48*(1), 13–21. doi: 10.1016/j.amepre.2014.08.029
Hotez, P. J. (2020). Anti-science extremism in America: escalating and globalizing. In Microbes and Infection (Vol. 22, Issue 10, pp. 505–507). Elsevier Masson s.r.l. doi: 10.1016/j.micinf.2020.09.005
Hotez, P. J., Nuzhath, T., & Colwell, B. (2020). Combating vaccine hesitancy and other 21st century social determinants in the global fight against measles. In Current Opinion in Virology (Vol. 41, pp. 1–7). Elsevier B.V. doi: 10.1016/j.coviro.2020.01.001
Jamison, A. M., Quinn, S. C., & Freimuth, V. S. (2019). “You don’t trust a government vaccine”: Narratives of institutional trust and influenza vaccination among African American and white adults. *Social Science and Medicine*, *221*, 87–94. doi: 10.1016/j.socscimed.2018.12.020
Kamel Boulos, M. N., & Geraghty, E. M. (2020). Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: How 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. *International Journal of Health Geographics*, *19*(1). doi: 10.1186/s12942-020-00202-8
Karlsson, L. C., Soveri, A., Lewandowsky, S., Karlsson, L., Karlsson, H., Nolvi, S., Karukivi, M., Lindfelt, M., & Antfolk, J. (2021). Fearing the disease or the vaccine: The case of COVID-19. *Personality and Individual Differences*, *172*. doi: 10.1016/j.paid.2020.110590
Kourlaba, G., Kourkouni, E., Maistreli, S., Tsopela, C. G., Molocha, N. M., Triantafyllou, C., Koniordou, M., Kopsidas, I., Chorianopoulou, E., Maroudi-Manta, S., Filippou, D., & Zaoutis, T. E. (2021). Willingness of Greek general population to get a COVID-19 vaccine. *Global Health Research and Policy*, *6*(1). doi: 10.1186/s41256-021-00188-1
Kricorian, K., Civen, R., & Equils, O. (2022). COVID-19 vaccine hesitancy: misinformation and perceptions of vaccine safety. *Human Vaccines and Immunotherapeutics*, *18*(1). doi: 10.1080/21645515.2021.1950504
Kulenkampff, M., Schwartzman, J. S., & Wilson, J. (1974). Neurological complications of pertussis inoculation. In Archives of Disease in Childhood (Vol. 49).
Lazarus, J. V., Ratzan, S. C., Palayew, A., Gostin, L. O., Larson, H. J., Rabin, K., Kimball, S., & El-Mohandes, A. (2021). A global survey of potential acceptance of a COVID-19 vaccine. *Nature Medicine*, *27*(2), 225–228. doi: 10.1038/s41591-020-1124-9

Lee, J., & Huang, Y. (2022). COVID-19 Vaccine Hesitancy: The Role of Socioeconomic Factors and Spatial Effects. *Vaccines*, *10*(3). doi: 10.3390/vaccines10030352

Lindemer, E., Choudhary, M., Donadio, G., Pawlowski, C., & Soundararajan, V. (2021). *Counties with lower insurance coverage are associated with both slower vaccine rollout and higher COVID-19 incidence across the United States*. doi: 10.1101/2021.03.24.21254270

Litaker, J. R., Tamez, N., Bray, C. L., Durkalski, W., & Taylor, R. (2022). Sociodemographic factors associated with vaccine hesitancy in central texas immediately prior to covid-19 vaccine availability. *International Journal of Environmental Research and Public Health*, *19*(1). doi: 10.3390/ijerph19010368

Liu, R., & Li, G. M. (2021). Hesitancy in the time of coronavirus: Temporal, spatial, and sociodemographic variations in COVID-19 vaccine hesitancy. *SSM - Population Health*, *15*. doi: 10.1016/j.ssmph.2021.100896

Malik, A. A., McFadden, S. A. M., Elharake, J., & Omer, S. B. (2020). Determinants of COVID-19 vaccine acceptance in the US. *EClinicalMedicine*, *26*. doi: 10.1016/j.eclinm.2020.100495

Marfe, G., Perna, S., & Shukla, A. (2021). Effectiveness of COVID-19 vaccines and their challenges (Review). *Experimental and Therapeutic Medicine*, *22*(6). doi: 10.3892/etm.2021.10843

Markert, U. R., Szekeres-Bartho, J., & Schleußner, E. (2021). Adverse effects on female fertility from vaccination against COVID-19 unlikely. *Journal of Reproductive Immunology*, *148*. doi: 10.1016/j.jri.2021.103428

Marshall, G. S. (2019). Vaccine hesitancy, history, and human nature: The 2018 Stanley A. Plotkin lecture. *Journal of the Pediatric Infectious Diseases Society*, *8*(1), 1–8. doi: 10.1093/jpids/piy082

Mollalo, A., Rivera, K. M., & Vahabi, N. (2021). Spatial statistical analysis of pre-existing mortalities of 20 diseases with COVID-19 mortalities in the continental United States. *Sustainable Cities and Society*, *67*. doi: 10.1016/j.scs.2021.102738

Mollalo, A., & Tatar, M. (2021). Spatial modeling of covid-19 vaccine hesitancy in the united states. *International Journal of Environmental Research and Public Health*, *18*(18). doi: 10.3390/ijerph18189488

Moran, P. A. P. (1950). *Notes on Continuous Stochastic Phenomena* (Vol. 37, Issue 1). Retrieved from https://www.jstor.org/stable/2332142

Nambi Ndugga, L. H. S. A. and S. H. (2022, July 14). *Latest Data on COVID-19 Vaccinations by Race/Ethnicity*.

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, *41*(5), 673–690. doi: 10.1007/s11135-006-9018-6

OCDE, trust in COVID-19 vaccination. (2021).

Palamenghi, L., Barello, S., Boccia, S., & Graffigna, G. (2020). Mistrust in biomedical research and vaccine hesitancy: the forefront challenge in the battle against COVID-19 in Italy. *European Journal of Epidemiology*, *35*(8), 785–788. doi: 10.1007/s10654-020-00675-8

Pallathadka, A., Chang, H., & Han, D. (2023). What explains spatial variations of COVID-19 vaccine hesitancy?: a social-ecological-technological systems approach. *Environmental Research: Health*, *1*(1), 011001. doi: 10.1088/2752-5309/ac8ac2

Pallathadka, A., Pallathadka, L., Rao, S., Chang, H., & Van Dommelen, D. (2022). Using GIS-based spatial analysis to determine urban greenspace accessibility for different racial groups in the backdrop of COVID-19: a case study of four US cities. *GeoJournal*, *87*(6), 4879–4899. doi: 10.1007/s10708-021-10538-8

Park, S., Massey, P. M., & Stimpson, J. P. (2021). Primary Source of Information About COVID-19 as a Determinant of Perception of COVID-19 Severity and Vaccine Uptake: Source of Information and COVID-19. *Journal of General Internal Medicine*, *36*(10), 3088–3095. doi: 10.1007/s11606-021-07080-1

Pfizer COVID-19. (2021, May 11).

Pink, S. L., Chu, J., Druckman, J. N., Rand, D. G., Willer, R., & Levi, M. (2021). *Elite party cues increase vaccination intentions among Republicans*. *118*. doi: 10.1073/pnas.2106559118/-/DCSupplemental

Pogue, K., Jensen, J. L., Stancil, C. K., Ferguson, D. G., Hughes, S. J., Mello, E. J., Burgess, R., Berges, B. K., Quaye, A., & Poole, B. D. (2020). Influences on attitudes regarding potential covid-19 vaccination in the united states. *Vaccines*, *8*(4), 1–14. doi: 10.3390/vaccines8040582

Poland, G. A., Jacobson, R. M., Tilburt, J., & Nichol, K. (2009). The social, political, ethical, and economic aspects of biodefense vaccines. *Vaccine*, *27*(SUPPL. 4). doi: 10.1016/j.vaccine.2009.08.054

Reich, J. A. (2020). "We are fierce, independent thinkers and intelligent": Social capital and stigma management among mothers who refuse vaccines. *Social Science and Medicine*, *257*. doi: 10.1016/j.socscimed.2018.10.027

Reiter, P. L., Pennell, M. L., & Katz, M. L. (2020). Acceptability of a COVID-19 vaccine among adults in the United States: How many people would get vaccinated? *Vaccine*, *38*(42), 6500–6507. doi: 10.1016/j.vaccine.2020.08.043

Ruiz, J. B., & Bell, R. A. (2021). Predictors of intention to vaccinate against COVID-19: Results of a nationwide survey. *Vaccine*, *39*(7), 1080–1086. doi: 10.1016/j.vaccine.2021.01.010

Sadarangani, M., Abu Raya, B., Conway, J. M., Iyaniwura, S. A., Falcao, R. C., Colijn, C., Coombs, D., & Gantt, S. (2021). Importance of COVID-19 vaccine efficacy in older age groups. *Vaccine*, *39*(15), 2020–2023. doi: 10.1016/j.vaccine.2021.03.020

Sallam, M. (2021). Covid-19 vaccine hesitancy worldwide: A concise systematic review of vaccine acceptance rates. In Vaccines (Vol. 9, Issue 2, pp. 1–15). MDPI AG. doi: 10.3390/vaccines9020160

Scheufele, D. A., & Krause, N. M. (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(16), 7662–7669. doi: 10.1073/pnas.1805871115

Sherman, S. M., Smith, L. E., Sim, J., Amlôt, R., Cutts, M., Dasch, H., Rubin, G. J., & Sevdalis, N. (2021). COVID-19 vaccination intention in the UK: results from the COVID-19 vaccination acceptability study (CoVAccS), a nationally representative cross-sectional survey. *Human Vaccines and Immunotherapeutics*, *17*(6), 1612–1621. doi: 10.1080/21645515.2020.1846397

Soares, P., Rocha, J. V., Moniz, M., Gama, A., Laires, P. A., Pedro, A. R., Dias, S., Leite, A., & Nunes, C. (2021). Factors associated with COVID-19 vaccine hesitancy. *Vaccines*, *9*(3). doi: 10.3390/vaccines9030300

Texas Vaccination progress COVID-19. (2023).

Troiano, G., & Nardi, A. (2021). Vaccine hesitancy in the era of COVID-19. In Public Health (Vol. 194, pp. 245–251). Elsevier B.V. doi: 10.1016/j.puhe.2021.02.025

Truong, J., Bakshi, S., Wasim, A., Ahmad, M., & Majid, U. (2022). What factors promote vaccine hesitancy or acceptance during pandemics? A systematic review and thematic analysis. *Health Promotion International*, *37*(1). doi: 10.1093/heapro/daab105

Ughasoro, M. D., Tagbo, B. N., & Esangbedo, D. O. (2015). Introduction of Inactivated Polio Vaccine and Specific Determinants of Vaccine Hesitancy. *World Journal of Vaccines*, *05*(01), 8–18. doi: 10.4236/wjv.2015.51002

Ulugtekin, N., Alkoy, S., Seker, D., & Goksel, C. (2006). Use of GIS in epidemiology: A case study in Istanbul. *Journal of Environmental Science and Health - Part A Toxic/Hazardous Substances and Environmental Engineering*, *41*(9), 2013–2026. doi: 10.1080/10934520600780636

U.S Department of Health and Human Services. (n.d.). *COVID-19 Vaccines*. Retrieved from https://www.hhs.gov/coronavirus/covid-19-vaccines/index.html

US Ranked by Population 2023. (2023).

Vaccine hesitancy, WHO. (2015).

Wakefield, A. J., Murch, S. H., Anthony, A., Linnell, J., Casson, D. M., Malik, M., Berelowitz, M., Dhillon, A. P., Thomson, M. A., Harvey, P., Valentine, A., Davies, S. E., & Walker-Smith, J. A. (1998). Retracted: Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet*, *351*(9103), 637–641. doi: 10.1016/S0140-6736(97)11096-0

Wang, J., Jing, R., Lai, X., Zhang, H., Lyu, Y., Knoll, M. D., & Fang, H. (2020). Acceptance of covid-19 vaccination during the covid-19 pandemic in china. *Vaccines*, *8*(3), 1–14. doi: 10.3390/vaccines8030482

WHO. (2019). *Ten threats to global health in 2019*. Retrieved from https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019

WHO, COVID-19. (2023).

WHO, COVID-19 cases. (2023).

Williams, L., Flowers, P., McLeod, J., Young, D., & Rollins, L. (2021). Social patterning and stability of intention to accept a COVID-19 vaccine in scotland: Will those most at risk accept a vaccine? *Vaccines*, *9*(1), 1–9. doi: 10.3390/vaccines9010017

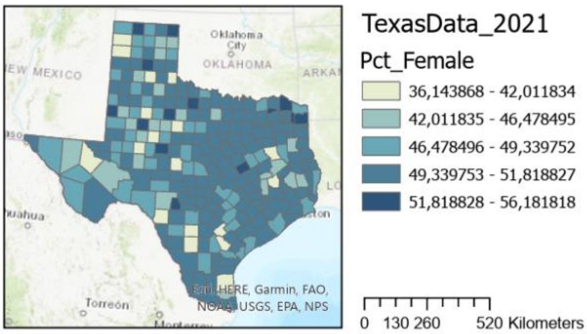
Xavier, C., & Rasu, R. S. (2021). Health Disparities of Coronavirus Disease 2019 in Texas, March-July 2020. *Southern Medical Journal*, *114*(10), 649–656. doi: 10.14423/SMJ.0000000000001308

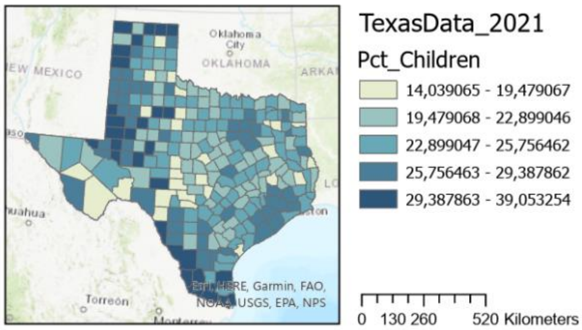
Yang, F., Li, X., Su, X., Xiao, T., Wang, Y., Hu, P., Li, H., Guan, J., Tian, H., Wang, P., & Wang, W. (2021). A study on willingness and influencing factors to receive COVID-19 vaccination among Qingdao residents. *Human Vaccines and Immunotherapeutics*, *17*(2), 408–413. doi: 10.1080/21645515.2020.1817712

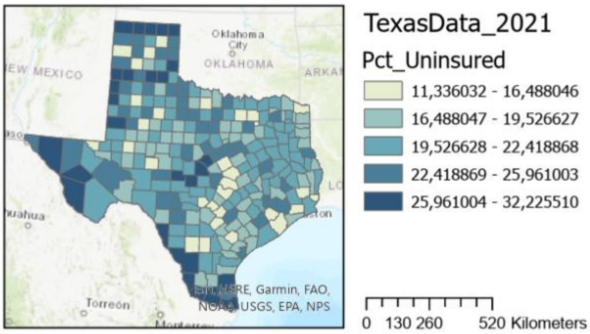
Yu, H., Fotheringham, A. S., Li, Z., Oshan, T., & Wolf, L. J. (2020). On the measurement of bias in geographically weighted regression models. *Spatial Statistics*, *38*. doi: 10.1016/j.spasta.2020.100453

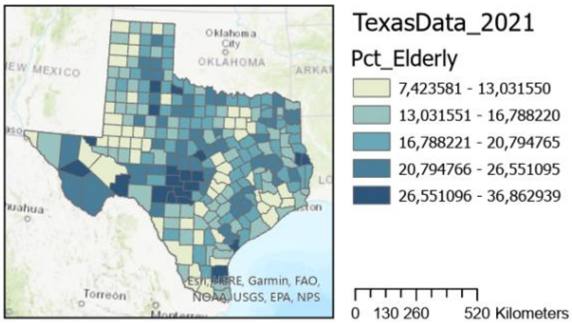
APPENDIX A

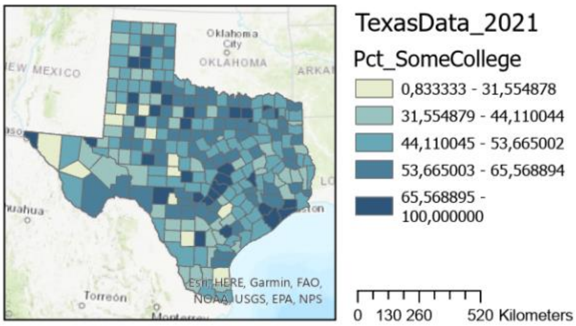
Visualize the data: These images represent the incidence rate for each explanatory variable within the state of Texas, in **2021.**

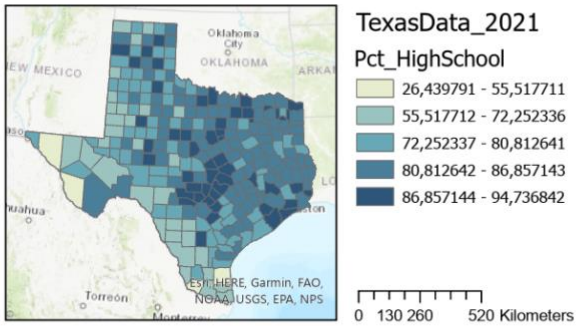


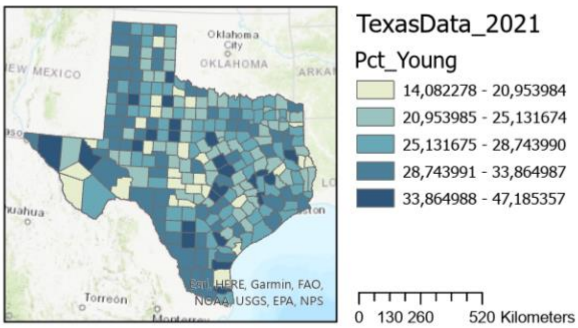


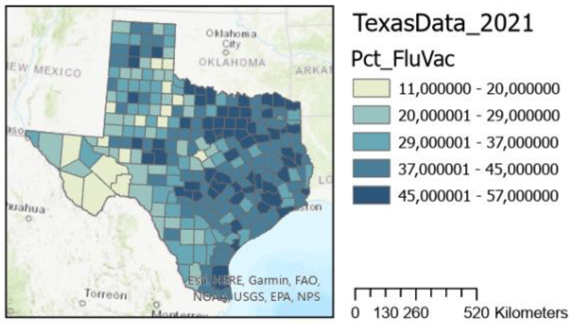


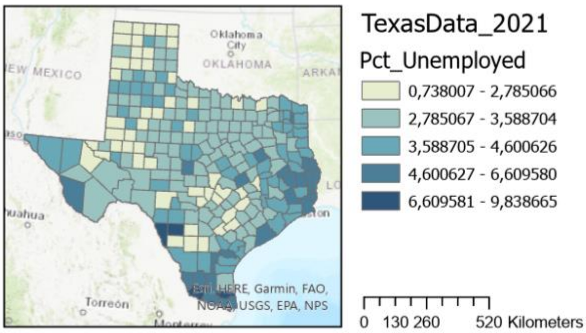


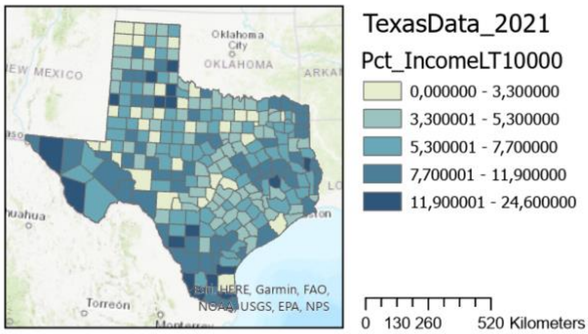


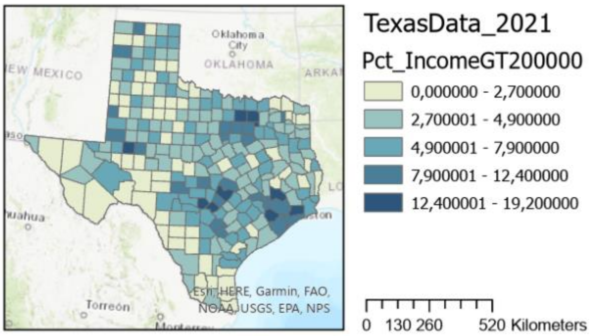


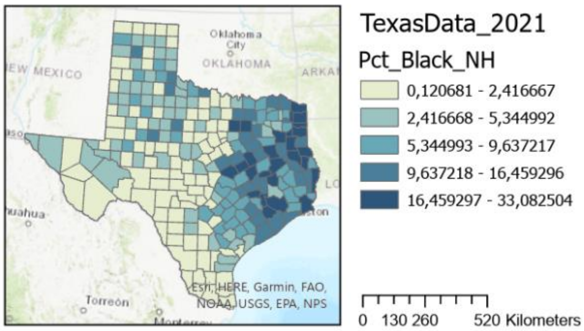


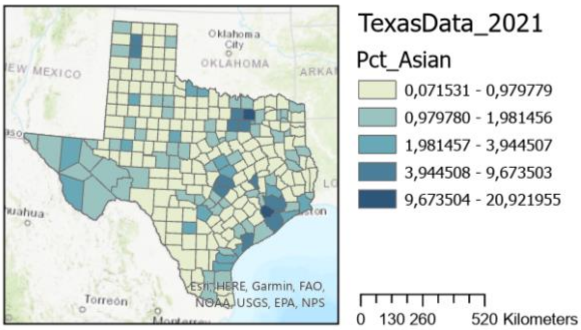


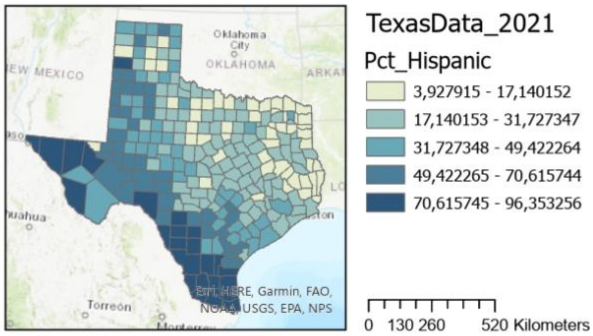

TexasData_2021
Pct_Unemployed
0,738007 - 2,785066
2,785067 - 3,588704
3,588705 - 4,600626
4,600627 - 6,609580
6,609581 - 9,838665
0 130 260 520 Kilometers
Oklahoma City
OKLAHOMA
Esri, HERE, Garmin, FAO, NOAA, USGS, EPA, NPS

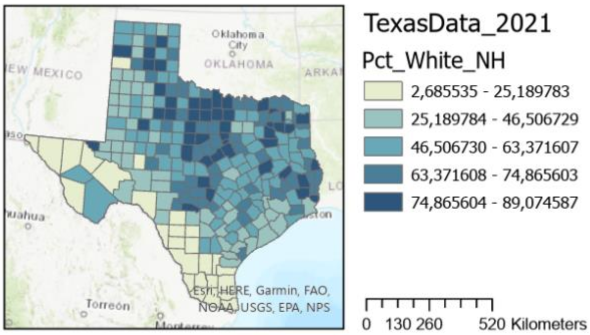

TexasData_2021
Pct_IncomeLT10000
0,000000 - 3,300000
3,300001 - 5,300000
5,300001 - 7,700000
7,700001 - 11,900000
11,900001 - 24,600000
0 130 260 520 Kilometers
Oklahoma City
OKLAHOMA
Esri, HERE, Garmin, FAO, NOAA, USGS, EPA, NPS

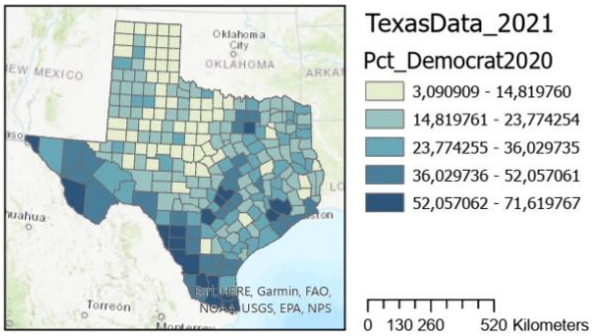

TexasData_2021
Pct_IncomeGT200000
0,000000 - 2,700000
2,700001 - 4,900000
4,900001 - 7,900000
7,900001 - 12,400000
12,400001 - 19,200000
0 130 260 520 Kilometers
Oklahoma City
OKLAHOMA
Esri, HERE, Garmin, FAO, NOAA, USGS, EPA, NPS

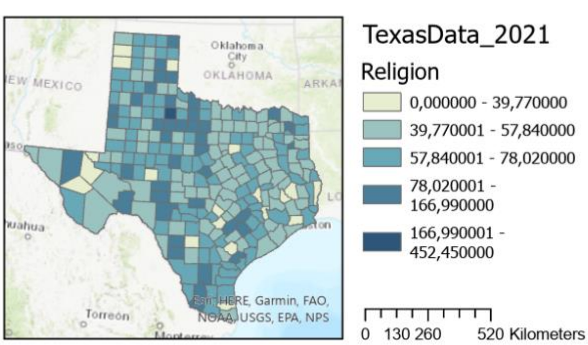

TexasData_2021
Pct_Black_NH
0,120681 - 2,416667
2,416668 - 5,344992
5,344993 - 9,637217
9,637218 - 16,459296
16,459297 - 33,082504
0 130 260 520 Kilometers
Oklahoma City
OKLAHOMA
Esri, HERE, Garmin, FAO, NOAA, USGS, EPA, NPS

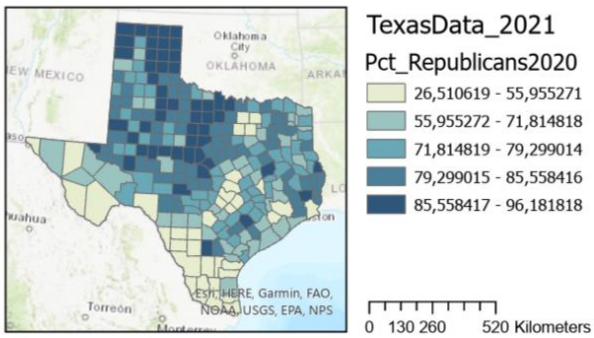

TexasData_2021
Pct_Asian
0,071531 - 0,979779
0,979780 - 1,981456
1,981457 - 3,944507
3,944508 - 9,673503
9,673504 - 20,921955
0 130 260 520 Kilometers
Oklahoma City
OKLAHOMA
Esri, HERE, Garmin, FAO, NOAA, USGS, EPA, NPS


TexasData_2021
Pct_Hispanic
3,927915 - 17,140152
17,140153 - 31,727347
31,727348 - 49,422264
49,422265 - 70,615744
70,615745 - 96,353256
0 130 260 520 Kilometers
Oklahoma City
OKLAHOMA
Esri, HERE, Garmin, FAO, NOAA, USGS, EPA, NPS

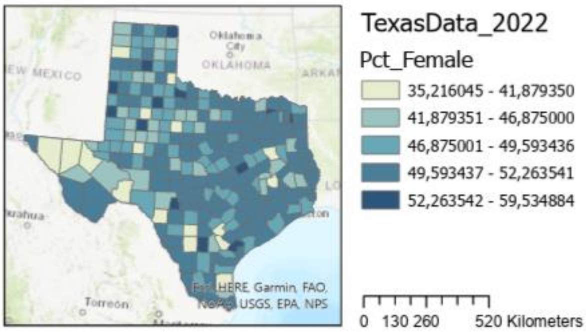

TexasData_2021
Pct_White_NH
2,685535 - 25,189783
25,189784 - 46,506729
46,506730 - 63,371607
63,371608 - 74,865603
74,865604 - 89,074587
0 130 260 520 Kilometers
Oklahoma City
OKLAHOMA
Esri, HERE, Garmin, FAO, NOAA, USGS, EPA, NPS

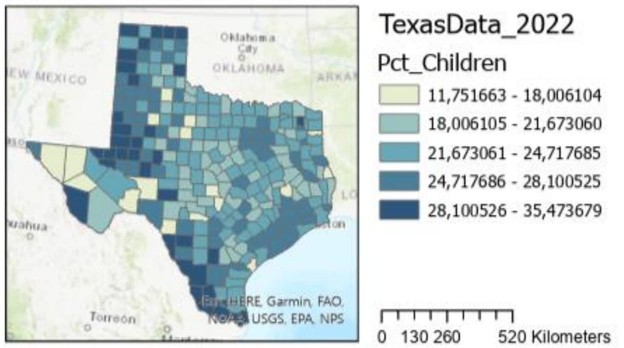

TexasData_2021
Pct_Democrat2020
3,090909 - 14,819760
14,819761 - 23,774254
23,774255 - 36,029735
36,029736 - 52,057061
52,057062 - 71,619767
0 130 260 520 Kilometers
Oklahoma City
OKLAHOMA
Esri, HERE, Garmin, FAO, NOAA, USGS, EPA, NPS

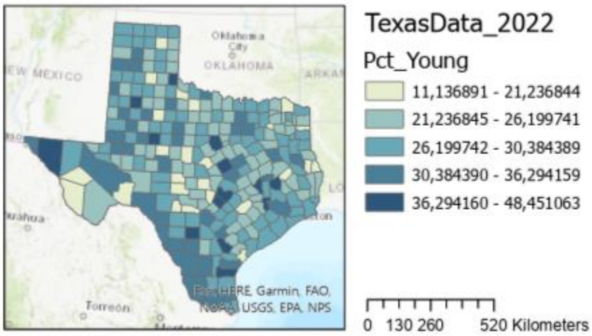

TexasData_2021
Religion
0,000000 - 39,770000
39,770001 - 57,840000
57,840001 - 78,020000
78,020001 - 166,990000
166,990001 - 452,450000
0 130 260 520 Kilometers
Oklahoma City
OKLAHOMA
Esri, HERE, Garmin, FAO, NOAA, USGS, EPA, NPS

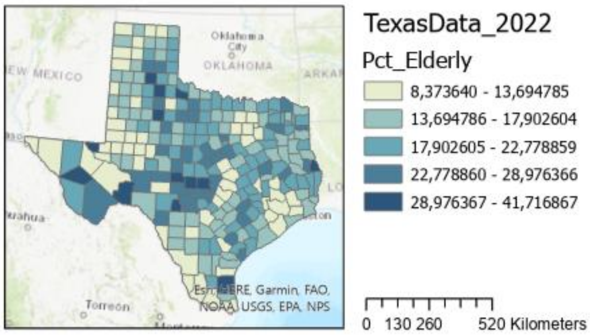

TexasData_2021
Pct_Republicans2020
26,510619 - 55,955271
55,955272 - 71,814818
71,814819 - 79,299014
79,299015 - 85,558416
85,558417 - 96,181818
0 130 260 520 Kilometers
Oklahoma City
OKLAHOMA
Esri, HERE, Garmin, FAO, NOAA, USGS, EPA, NPS

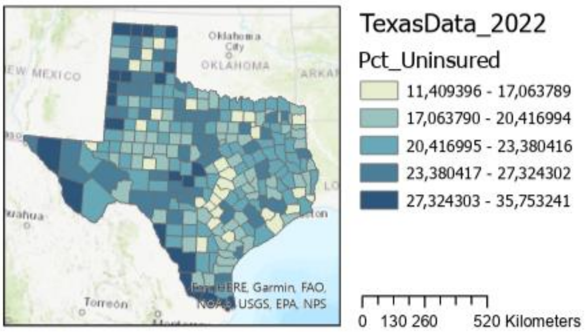
Visualize the data: These images represent the incidence rate for each explanatory variable within the state of Texas, in **2022.**

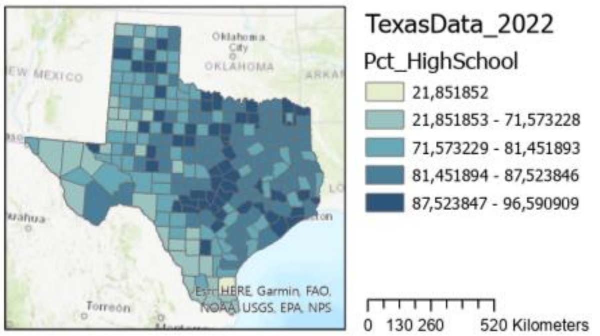


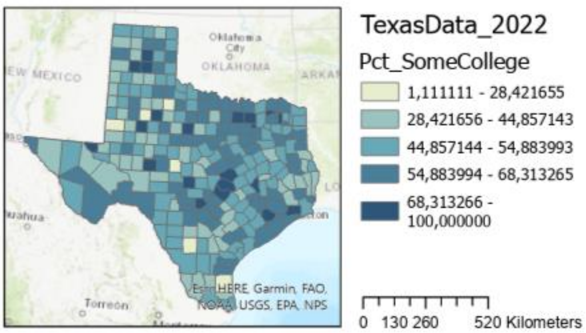


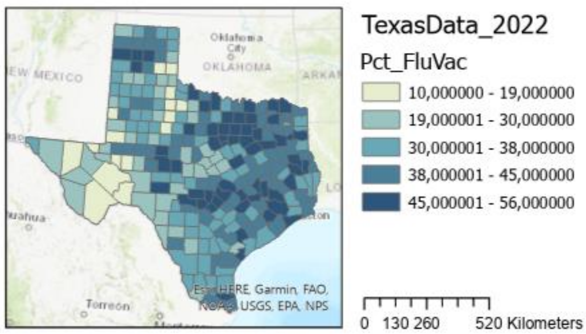


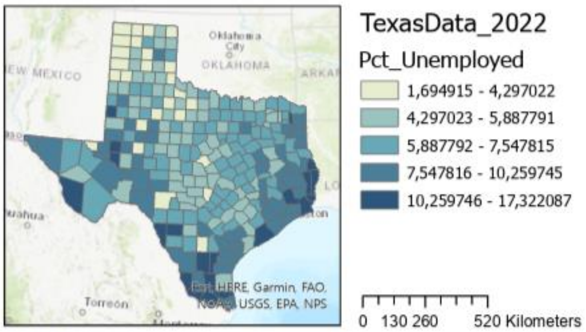


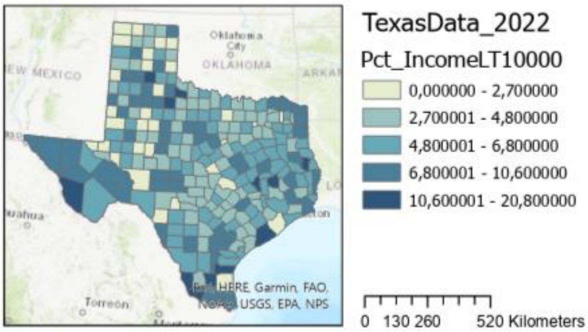


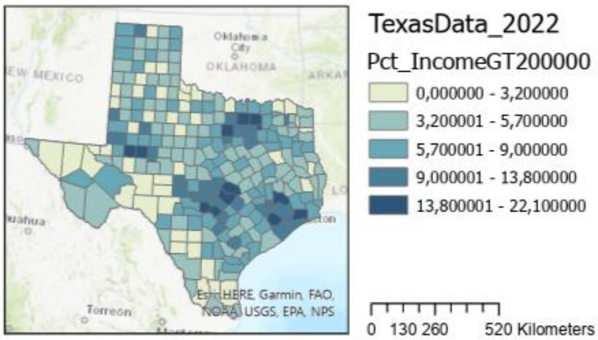


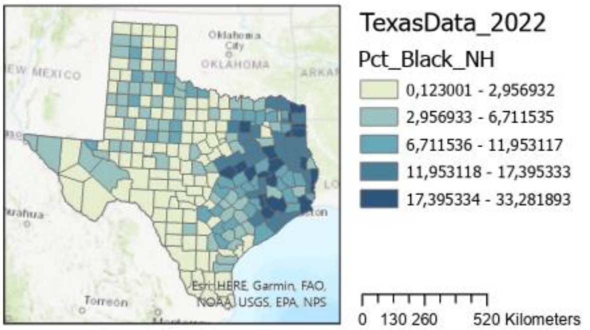


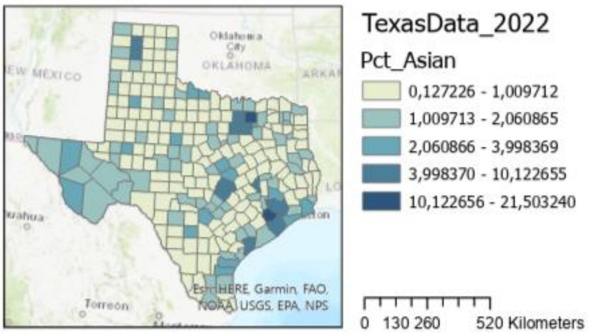


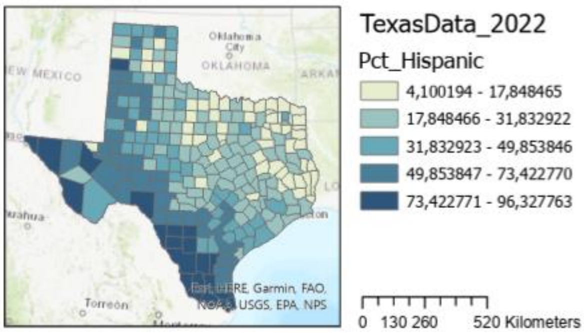

TexasData_2022
Pct_Unemployed
1,694915 - 4,297022
4,297023 - 5,887791
5,887792 - 7,547815
7,547816 - 10,259745
10,259746 - 17,322087
0 130 260 520 Kilometers

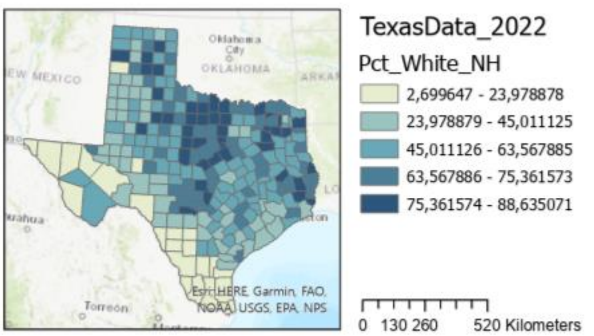

TexasData_2022
Pct_IncomeLT10000
0,000000 - 2,700000
2,700001 - 4,800000
4,800001 - 6,800000
6,800001 - 10,600000
10,600001 - 20,800000
0 130 260 520 Kilometers

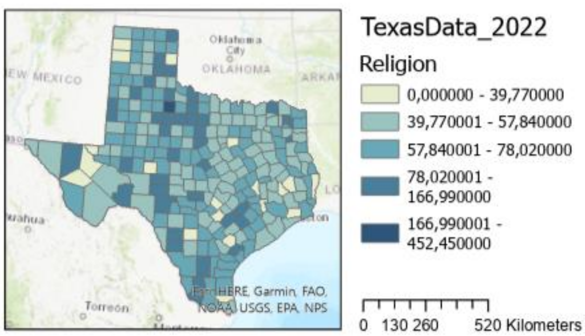

TexasData_2022
Pct_IncomeGT200000
0,000000 - 3,200000
3,200001 - 5,700000
5,700001 - 9,000000
9,000001 - 13,800000
13,800001 - 22,100000
0 130 260 520 Kilometers


TexasData_2022
Pct_Black_NH
0,123001 - 2,956932
2,956933 - 6,711535
6,711536 - 11,953117
11,953118 - 17,395333
17,395334 - 33,281893
0 130 260 520 Kilometers


TexasData_2022
Pct_Asian
0,127226 - 1,009712
1,009713 - 2,060865
2,060866 - 3,998369
3,998370 - 10,122655
10,122656 - 21,503240
0 130 260 520 Kilometers

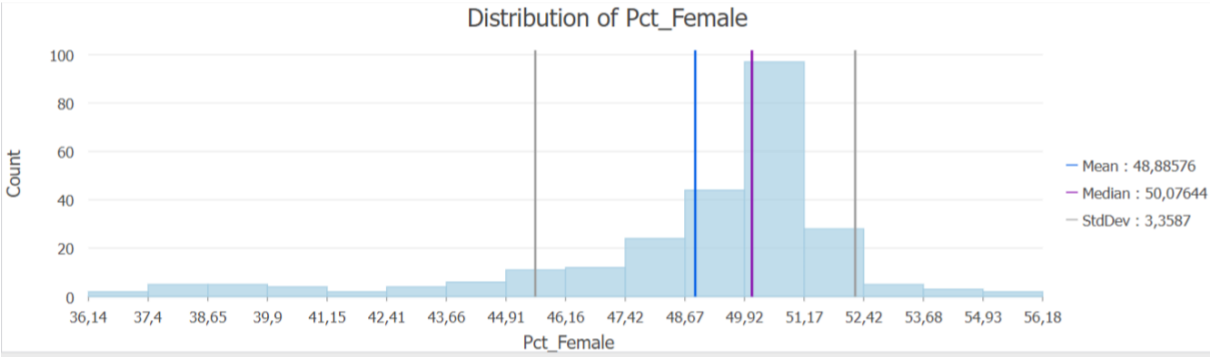

TexasData_2022
Pct_Hispanic
4,100194 - 17,848465
17,848466 - 31,832922
31,832923 - 49,853846
49,853847 - 73,422770
73,422771 - 96,327763
0 130 260 520 Kilometers

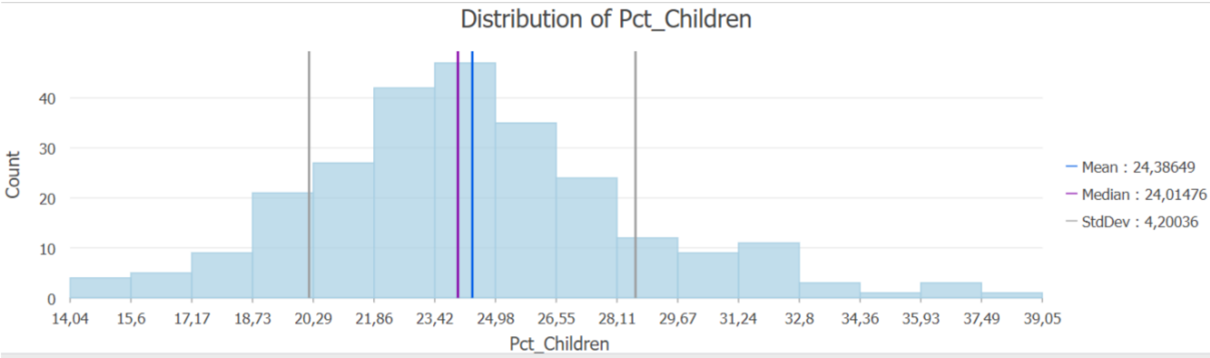

TexasData_2022
Pct_White_NH
2,699647 - 23,978878
23,978879 - 45,011125
45,011126 - 63,567885
63,567886 - 75,361573
75,361574 - 88,635071
0 130 260 520 Kilometers

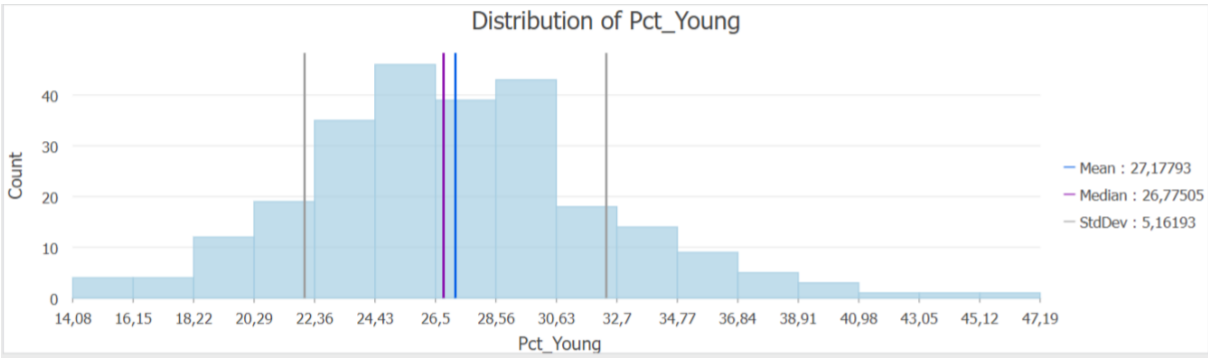

TexasData_2022
Religion
0,000000 - 39,770000
39,770001 - 57,840000
57,840001 - 78,020000
78,020001 - 166,990000
166,990001 - 452,450000
0 130 260 520 Kilometers

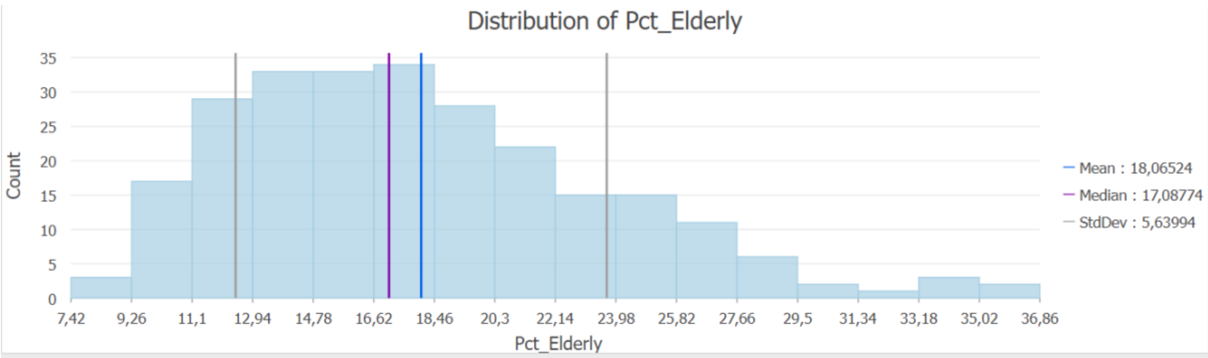
APPENDIX B

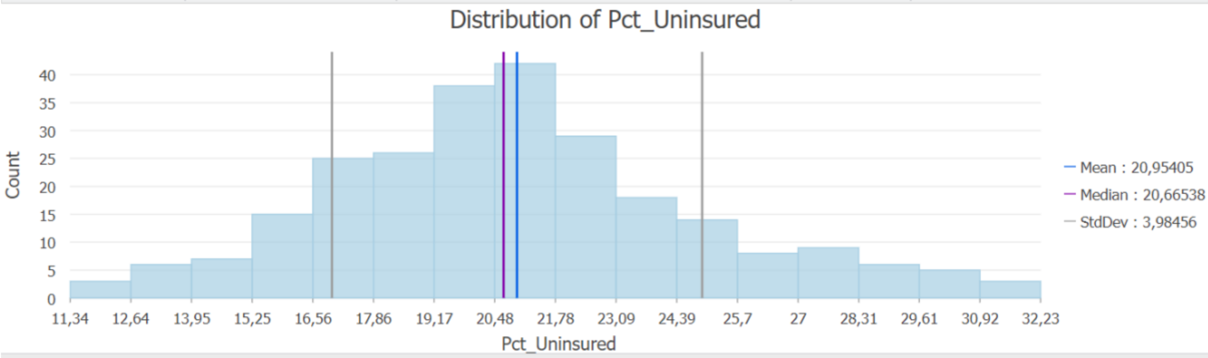
Histograms for each of the explanatory variables that have the maximum, minimum, the mean, the median and the standard deviation for each one of them (for the year of **2021**)

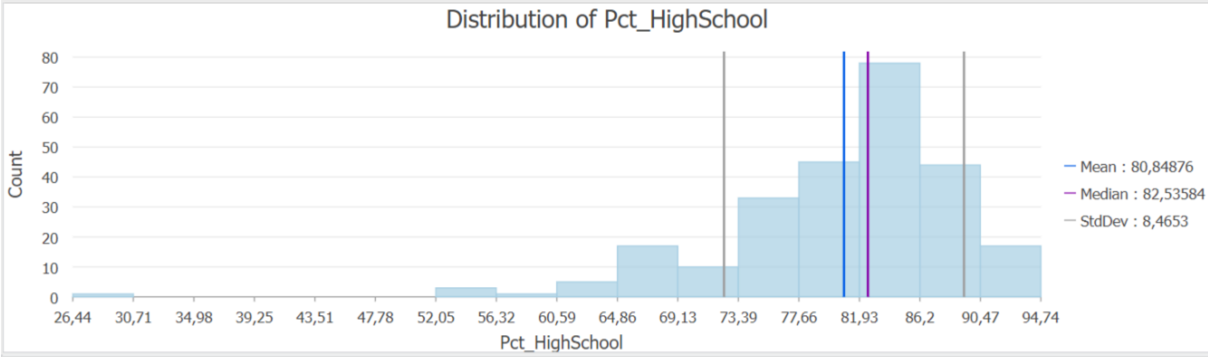


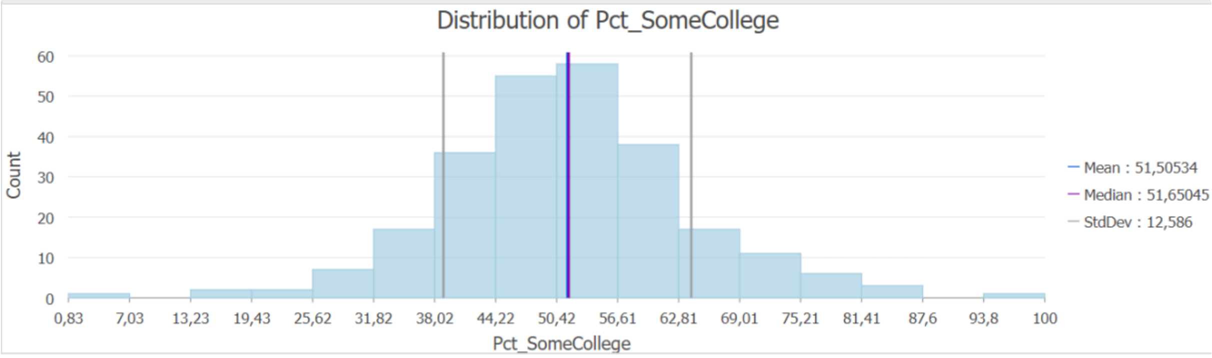


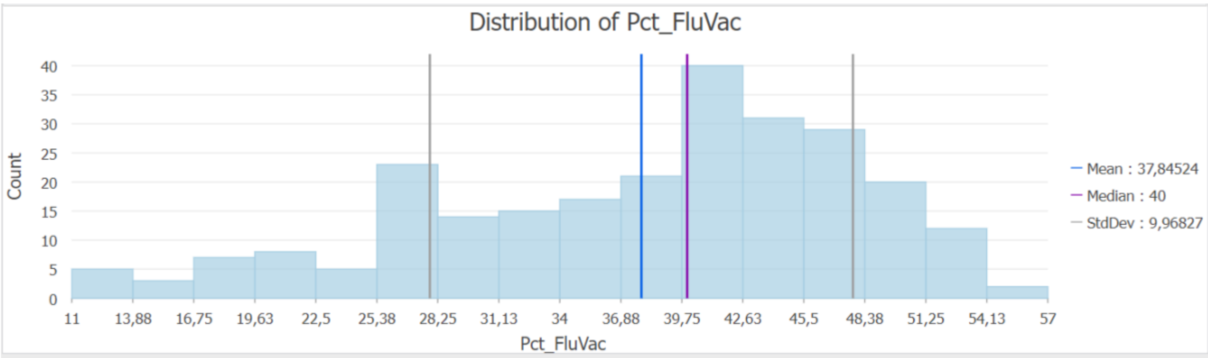


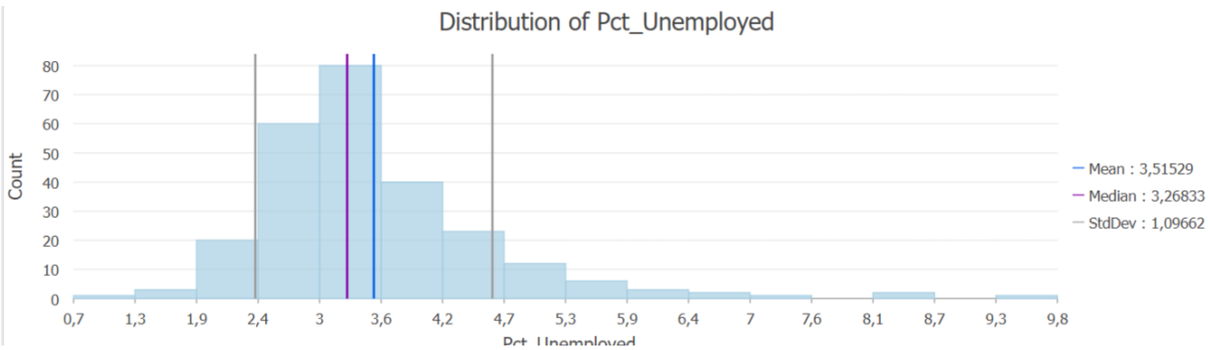


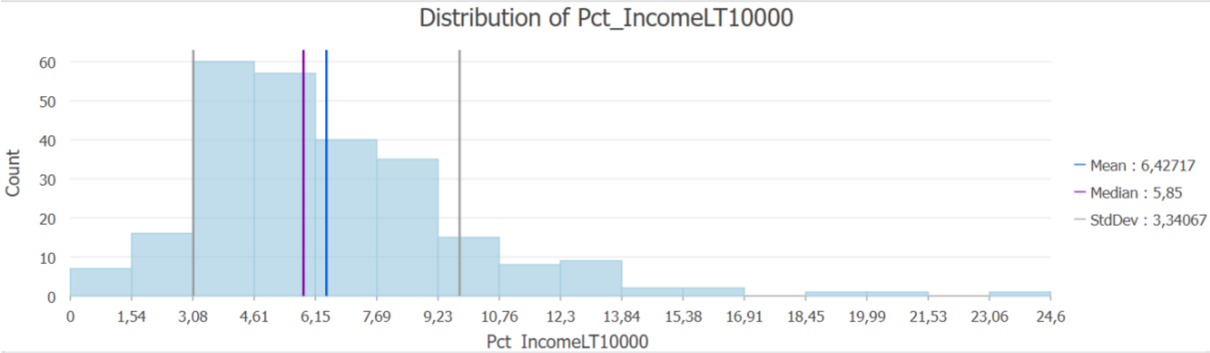

Distribution of Pct_Uninsured
Count
Pct_Uninsured
11,34 12,64 13,95 15,25 16,56 17,86 19,17 20,48 21,78 23,09 24,39 25,7 27 28,31 29,61 30,92 32,23
Mean : 20,95405
Median : 20,66538
StdDev : 3,98456

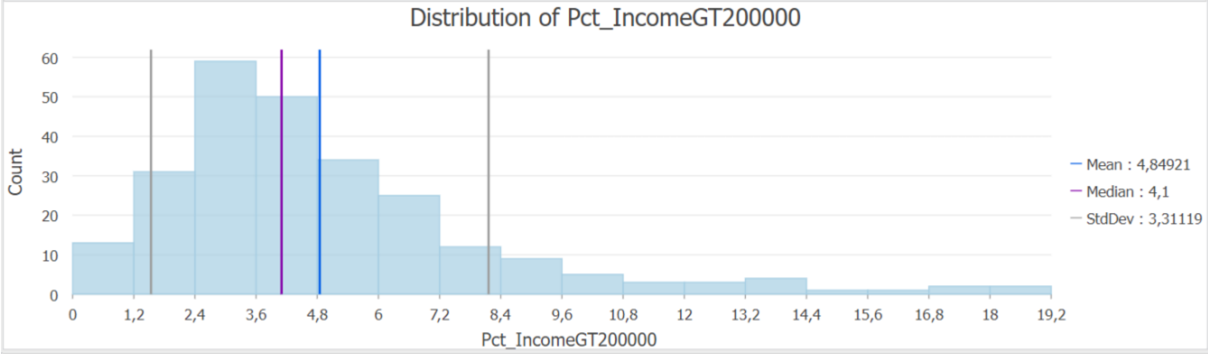

Distribution of Pct_HighSchool
Count
Pct_HighSchool
26,44 30,71 34,98 39,25 43,51 47,78 52,05 56,32 60,59 64,86 69,13 73,39 77,66 81,93 86,2 90,47 94,74
Mean : 80,84876
Median : 82,53584
StdDev : 8,4653

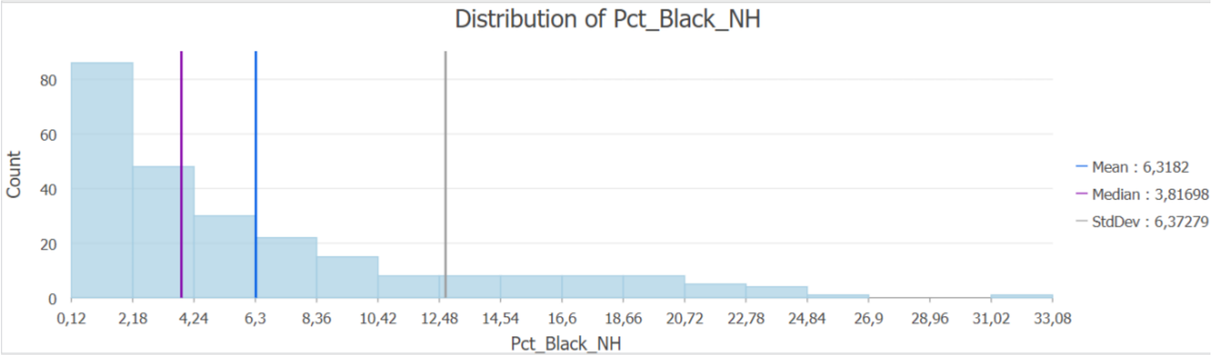

Distribution of Pct_SomeCollege
Count
Pct_SomeCollege
0,83 7,03 13,23 19,43 25,62 31,82 38,02 44,22 50,42 56,61 62,81 69,01 75,21 81,41 87,6 93,8 100
Mean : 51,50534
Median : 51,65045
StdDev : 12,586

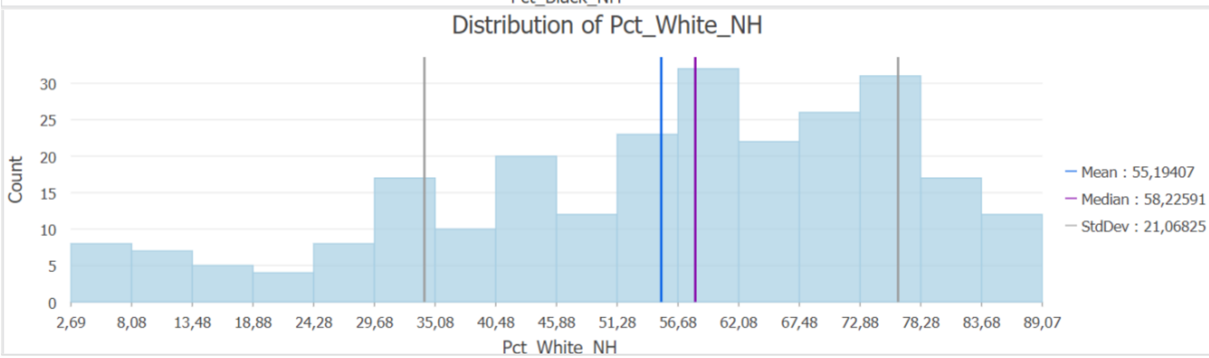

Distribution of Pct_FluVac
Count
Pct_FluVac
11 13,88 16,75 19,63 22,5 25,38 28,25 31,13 34 36,88 39,75 42,63 45,5 48,38 51,25 54,13 57
Mean : 37,84524
Median : 40
StdDev : 9,96827

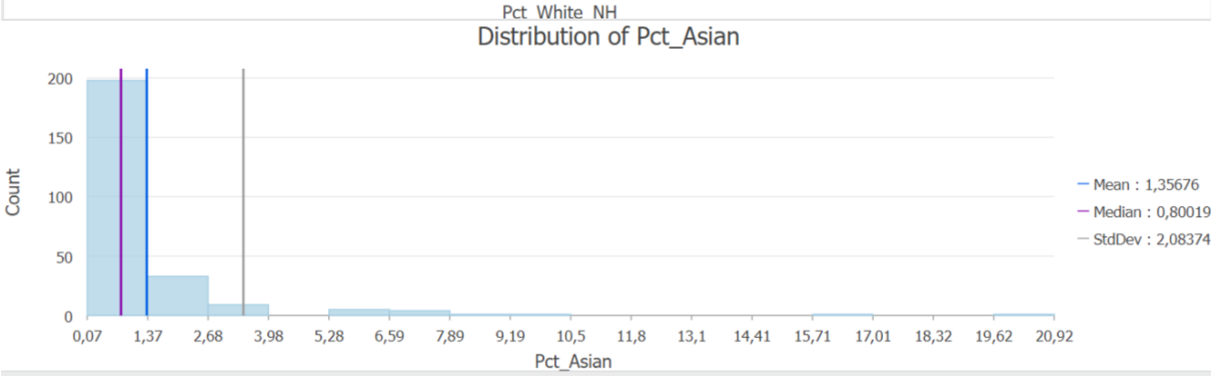

Distribution of Pct_Unemployed
Count
0,7 1,3 1,9 2,4 3 3,6 4,2 4,7 5,3 5,9 6,4 7 7,6 8,1 8,7 9,3 9,8
Mean : 3,51529
Median : 3,26833
StdDev : 1,09662

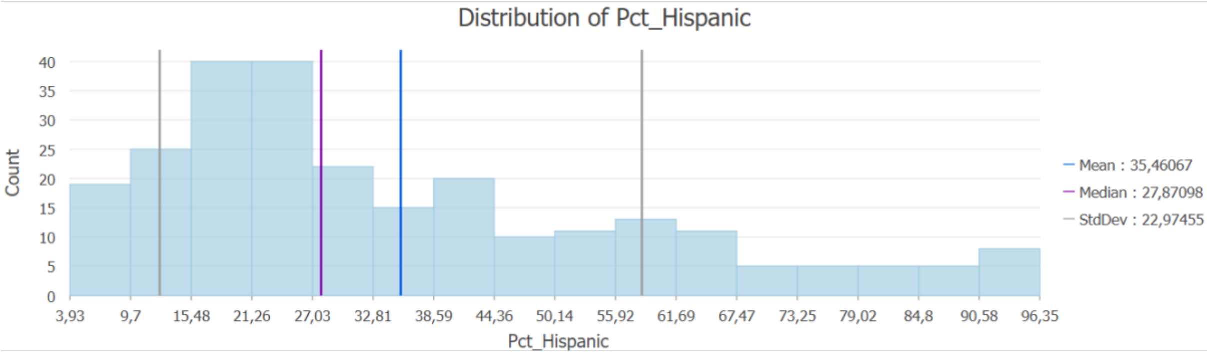

Distribution of Pct_IncomeLT10000
Count
60
50
40
30
20
10
0
0 1,54 3,08 4,61 6,15 7,69 9,23 10,76 12,3 13,84 15,38 16,91 18,45 19,99 21,53 23,06 24,6
Pct_IncomeLT10000
Mean : 6,42717
Median : 5,85
StdDev : 3,34067

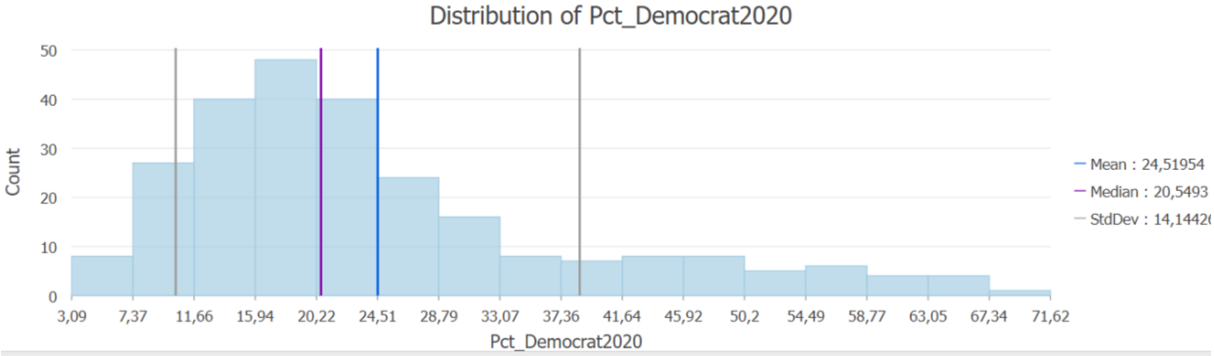

Distribution of Pct_IncomeGT200000
Count
60
50
40
30
20
10
0
0 1,2 2,4 3,6 4,8 6 7,2 8,4 9,6 10,8 12 13,2 14,4 15,6 16,8 18 19,2
Pct_IncomeGT200000
Mean : 4,84921
Median : 4,1
StdDev : 3,31119

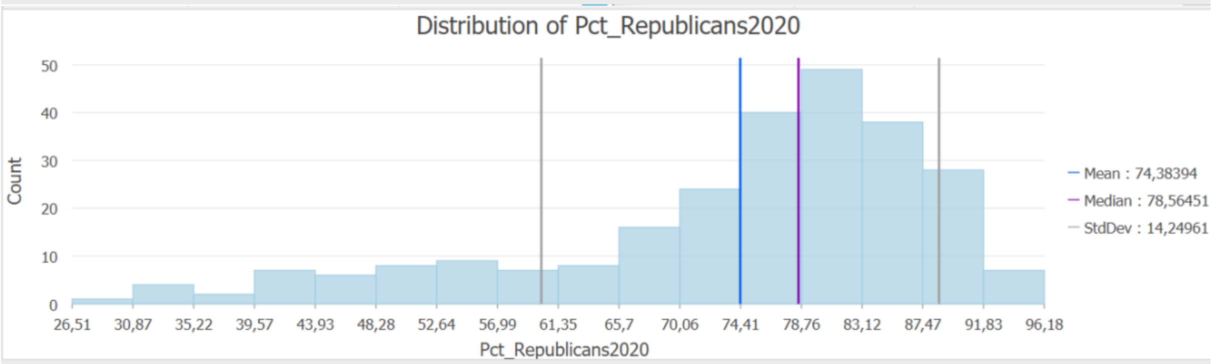

Distribution of Pct_Black_NH
Count
80
60
40
20
0
0,12 2,18 4,24 6,3 8,36 10,42 12,48 14,54 16,6 18,66 20,72 22,78 24,84 26,9 28,96 31,02 33,08
Pct_Black_NH
Mean : 6,3182
Median : 3,81698
StdDev : 6,37279

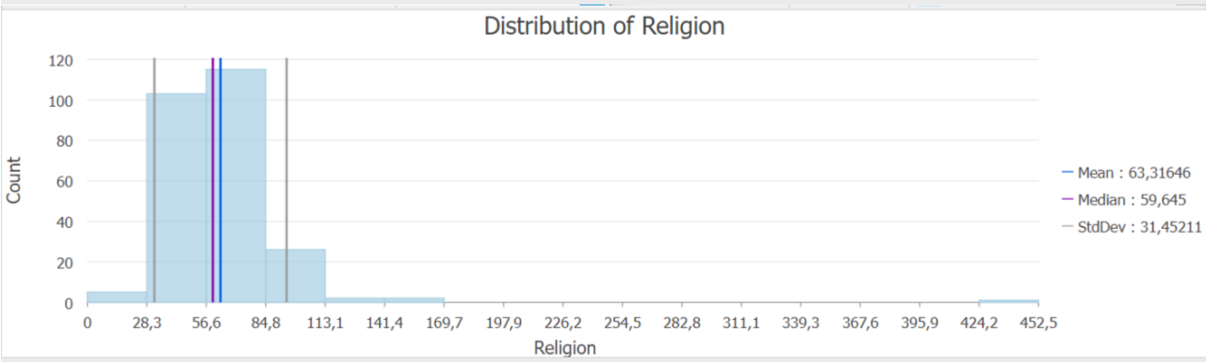

Distribution of Pct_White_NH
Count
30
25
20
15
10
5
0
2,69 8,08 13,48 18,88 24,28 29,68 35,08 40,48 45,88 51,28 56,68 62,08 67,48 72,88 78,28 83,68 89,07
Pct_White_NH
Mean : 55,19407
Median : 58,22591
StdDev : 21,06825


Distribution of Pct_Asian
Count
200
150
100
50
0
0,07 1,37 2,68 3,98 5,28 6,59 7,89 9,19 10,5 11,8 13,1 14,41 15,71 17,01 18,32 19,62 20,92
Pct_Asian
Mean : 1,35676
Median : 0,80019
StdDev : 2,08374

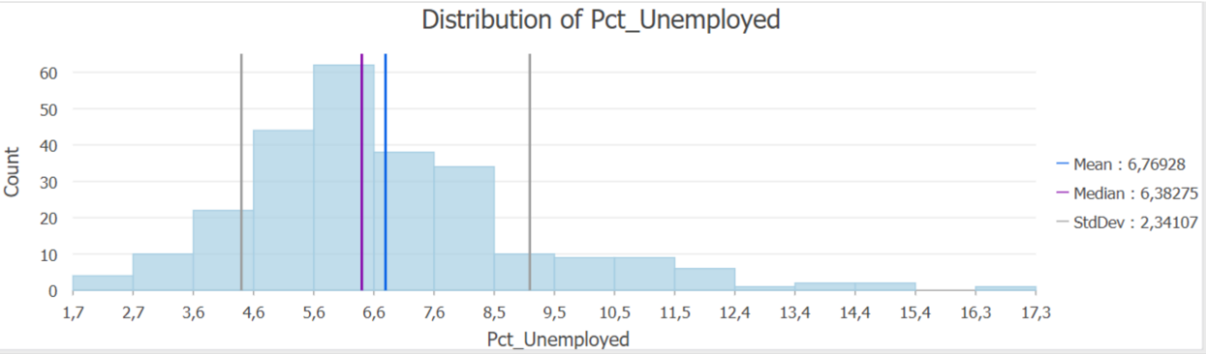

Distribution of Pct_Hispanic
Count
40
35
30
25
20
15
10
5
0
3,93
9,7
15,48
21,26
27,03
32,81
38,59
44,36
50,14
55,92
61,69
67,47
73,25
79,02
84,8
90,58
96,35
Pct_Hispanic
Mean : 35,46067
Median : 27,87098
StdDev : 22,97455

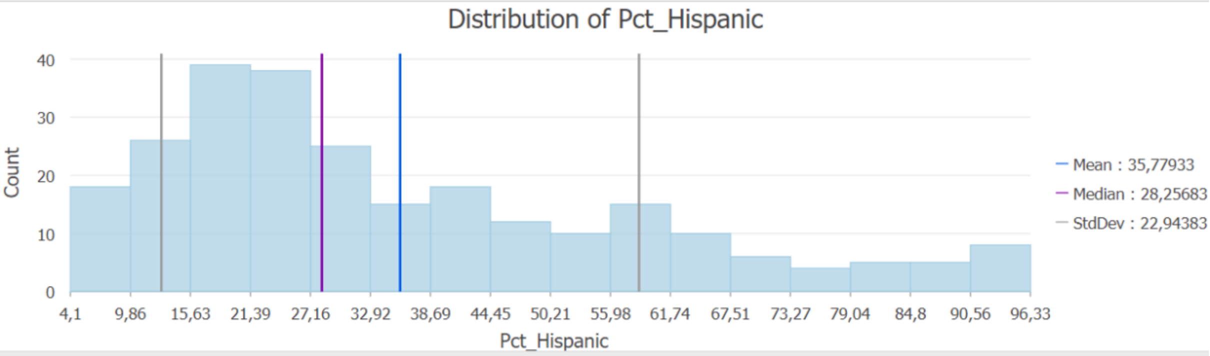

Distribution of Pct_Democrat2020
Count
50
40
30
20
10
0
3,09
7,37
11,66
15,94
20,22
24,51
28,79
33,07
37,36
41,64
45,92
50,2
54,49
58,77
63,05
67,34
71,62
Pct_Democrat2020
Mean : 24,51954
Median : 20,5493
StdDev : 14,1442

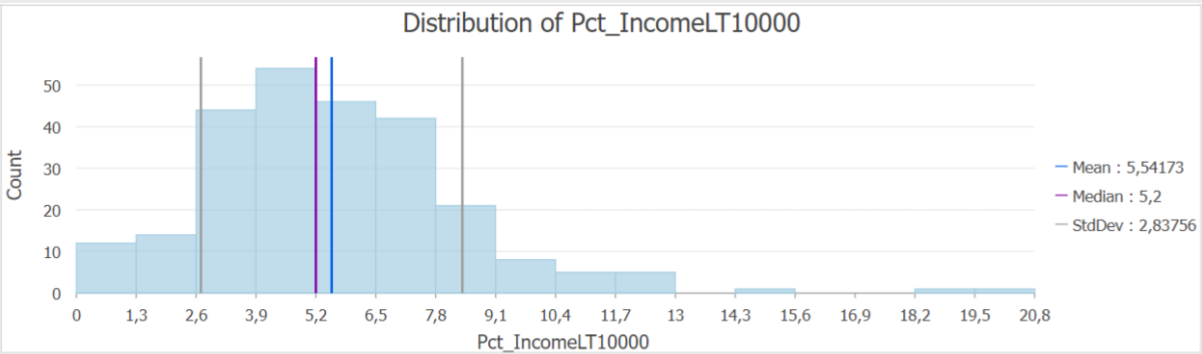

Distribution of Pct_Republicans2020
Count
50
40
30
20
10
0
26,51
30,87
35,22
39,57
43,93
48,28
52,64
56,99
61,35
65,7
70,06
74,41
78,76
83,12
87,47
91,83
96,18
Pct_Republicans2020
Mean : 74,38394
Median : 78,56451
StdDev : 14,24961

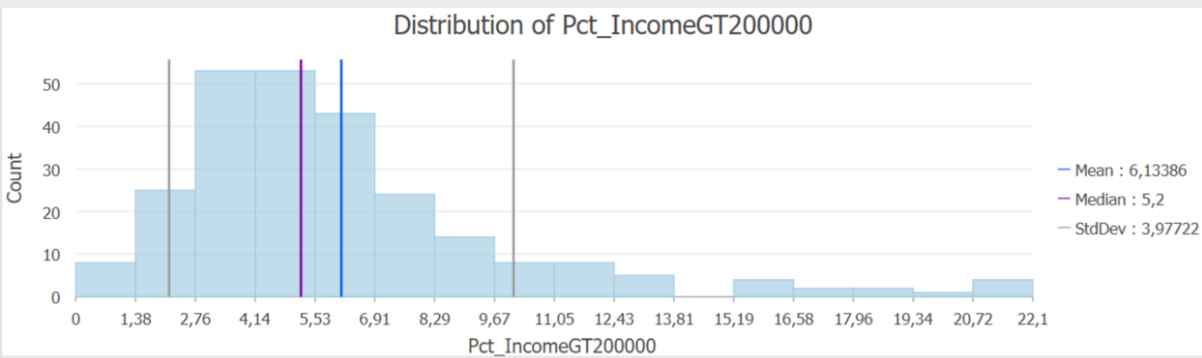

Distribution of Religion
Count
120
100
80
60
40
20
0
0
28,3
56,6
84,8
113,1
141,4
169,7
197,9
226,2
254,5
282,8
311,1
339,3
367,6
395,9
424,2
452,5
Religion
Mean : 63,31646
Median : 59,645
StdDev : 31,45211

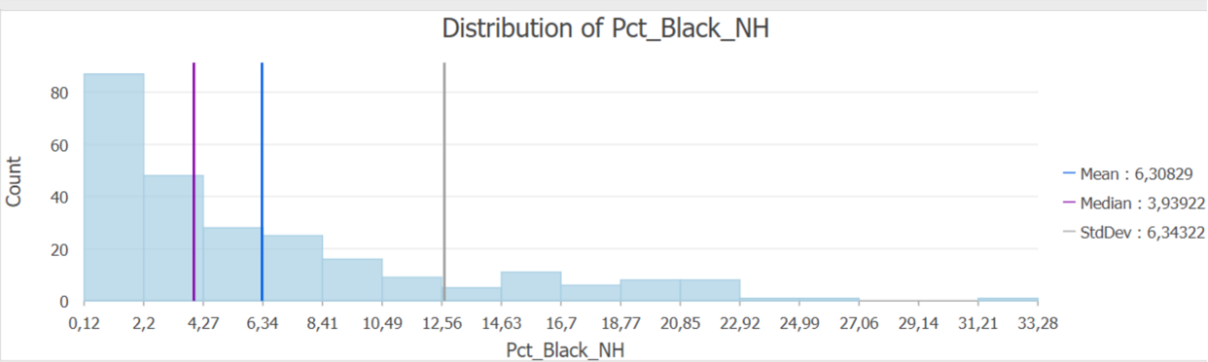
Now we present the same graphs as above but for the year of **2022**

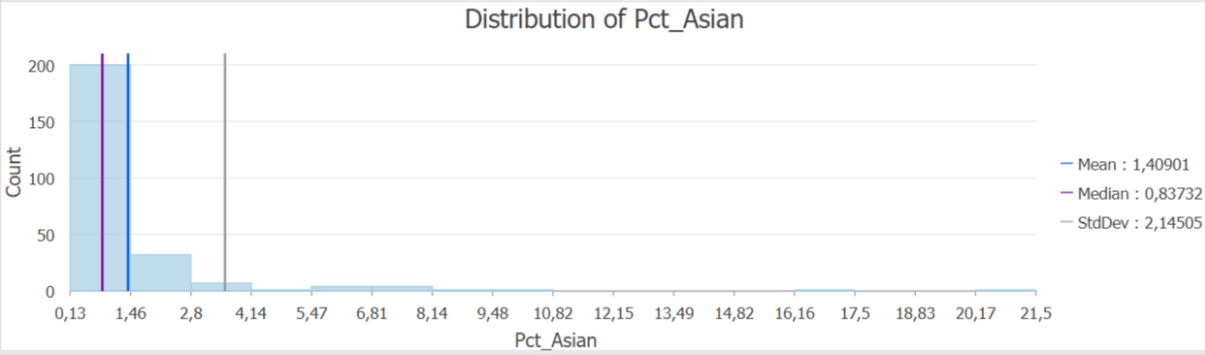


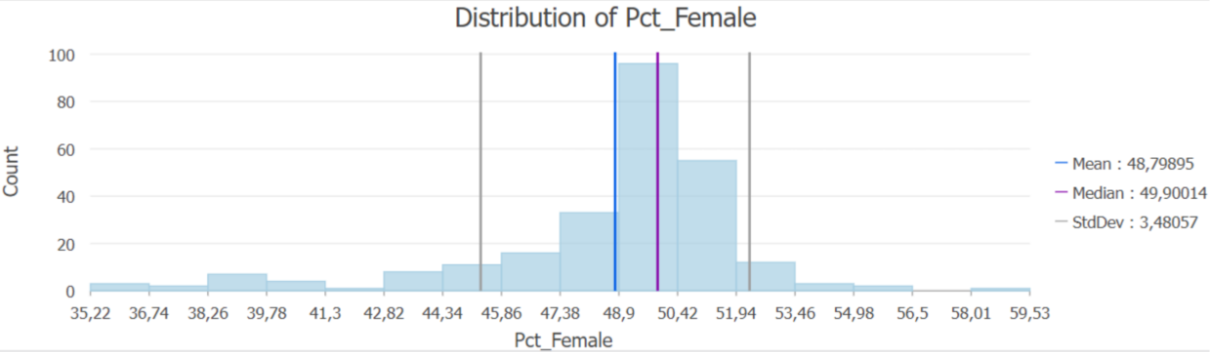


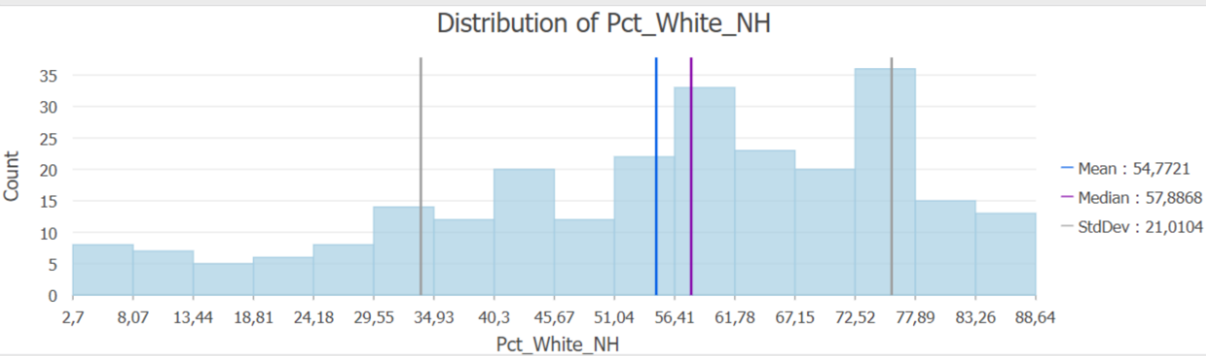


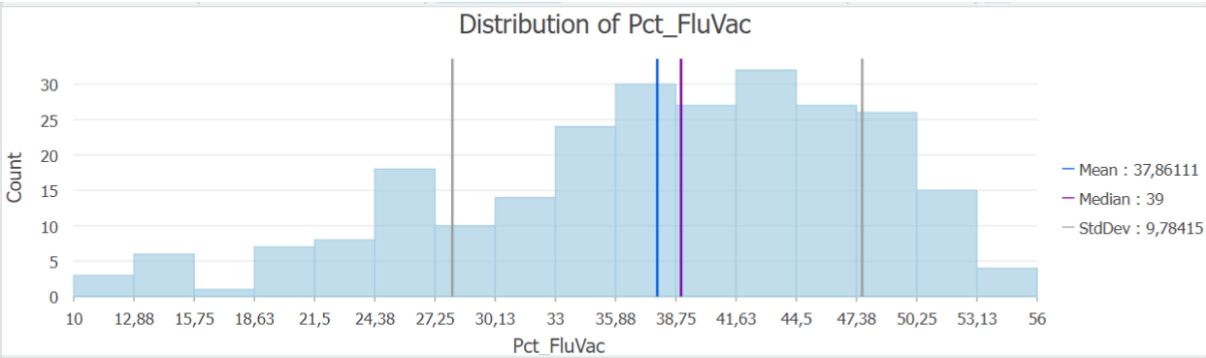


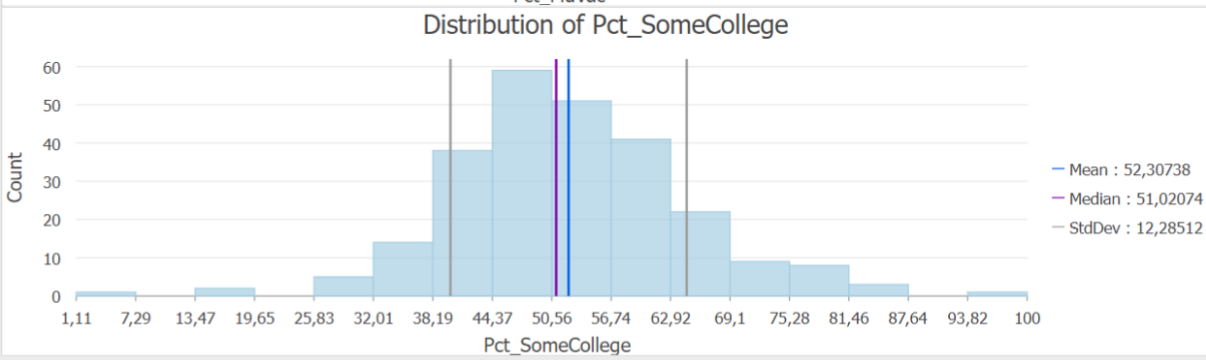


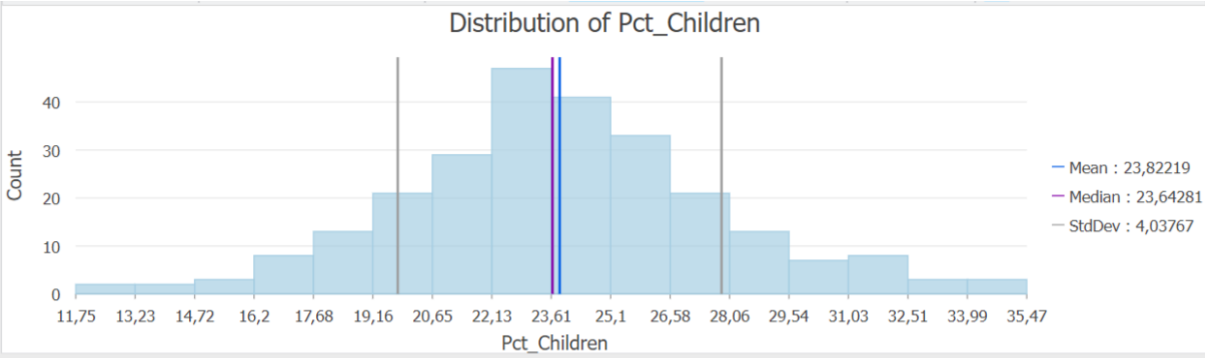

Distribution of Pct_Asian
Count
Pct_Asian
Mean : 1,40901
Median : 0,83732
StdDev : 2,14505

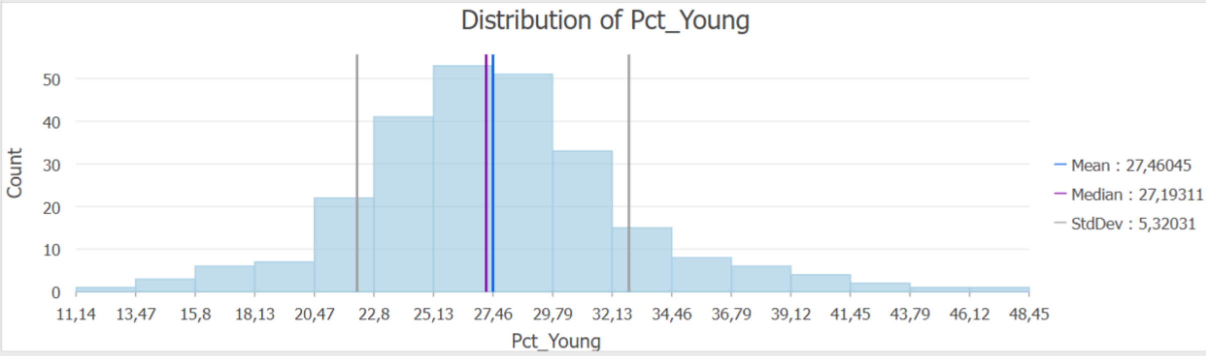

Distribution of Pct_Female
Count
Pct_Female
Mean : 48,79895
Median : 49,90014
StdDev : 3,48057

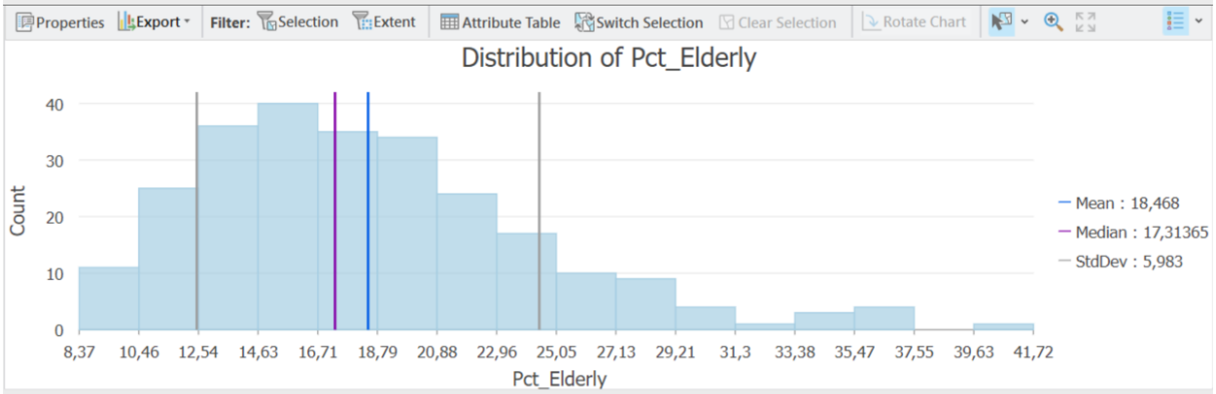

Distribution of Pct_White_NH
Count
Pct_White_NH
Mean : 54,7721
Median : 57,8868
StdDev : 21,0104

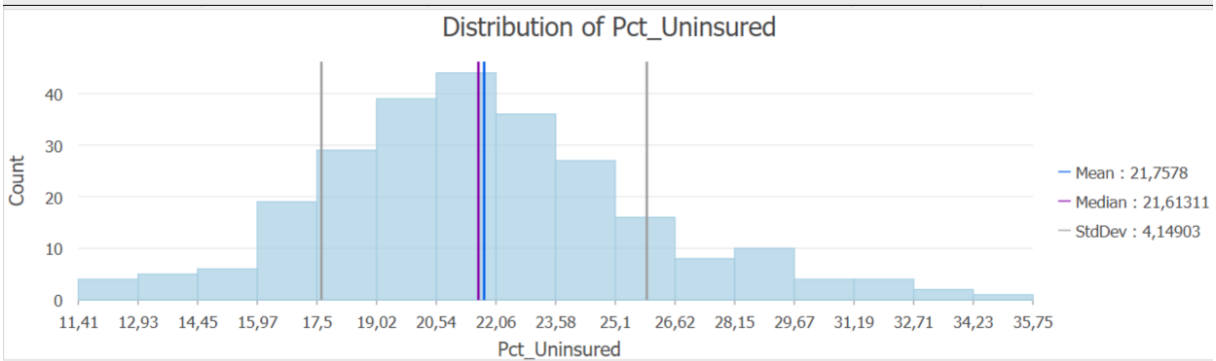

Distribution of Pct_FluVac
Count
Pct_FluVac
Mean : 37,86111
Median : 39
StdDev : 9,78415

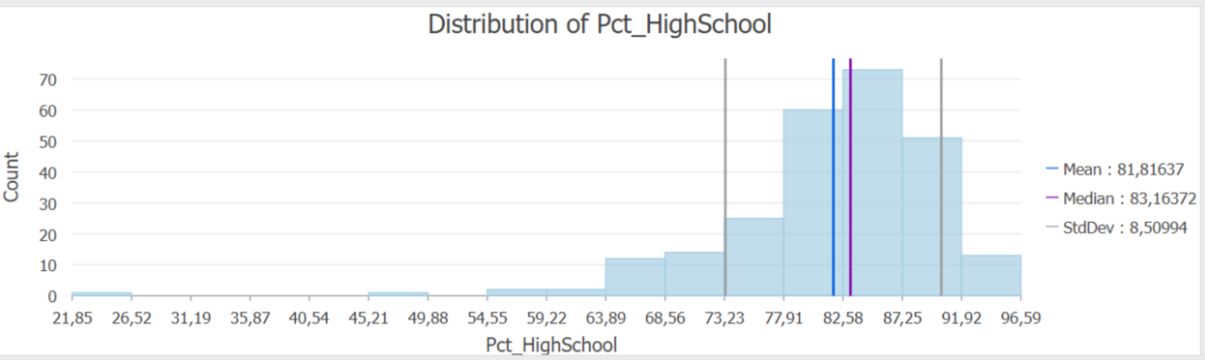

Distribution of Pct_SomeCollege
Count
Pct_SomeCollege
Mean : 52,30738
Median : 51,02074
StdDev : 12,28512

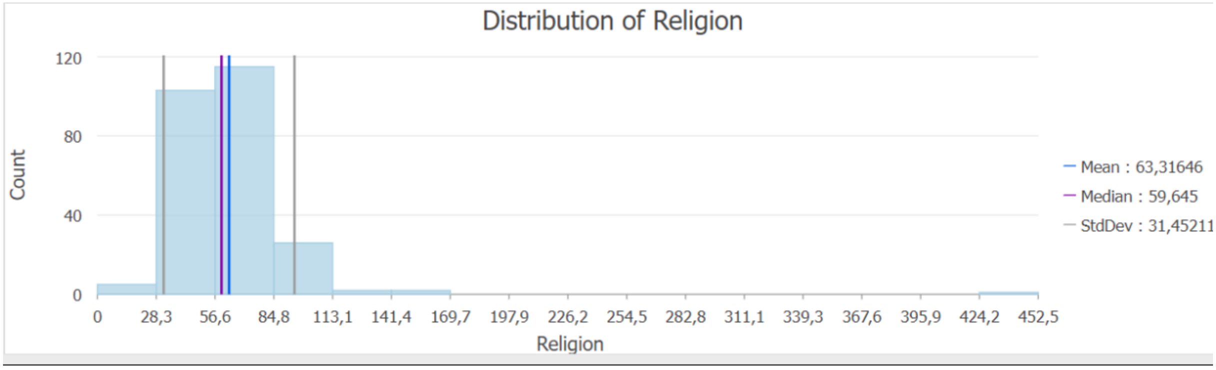

Distribution of Pct_Children
Count
0 10 20 30 40
11,75 13,23 14,72 16,2 17,68 19,16 20,65 22,13 23,61 25,1 26,58 28,06 29,54 31,03 32,51 33,99 35,47
Pct_Children
Mean : 23,82219
Median : 23,64281
StdDev : 4,03767


Distribution of Pct_Young
Count
0 10 20 30 40 50
11,14 13,47 15,8 18,13 20,47 22,8 25,13 27,46 29,79 32,13 34,46 36,79 39,12 41,45 43,79 46,12 48,45
Pct_Young
Mean : 27,46045
Median : 27,19311
StdDev : 5,32031


Properties
Export
Filter:
Selection
Extent
Attribute Table
Switch Selection
Clear Selection
Rotate Chart
Distribution of Pct_Elderly
Count
0 10 20 30 40
8,37 10,46 12,54 14,63 16,71 18,79 20,88 22,96 25,05 27,13 29,21 31,3 33,38 35,47 37,55 39,63 41,72
Pct_Elderly
Mean : 18,468
Median : 17,31365
StdDev : 5,983


Distribution of Pct_Uninsured
Count
0 10 20 30 40
11,41 12,93 14,45 15,97 17,5 19,02 20,54 22,06 23,58 25,1 26,62 28,15 29,67 31,19 32,71 34,23 35,75
Pct_Uninsured
Mean : 21,7578
Median : 21,61311
StdDev : 4,14903

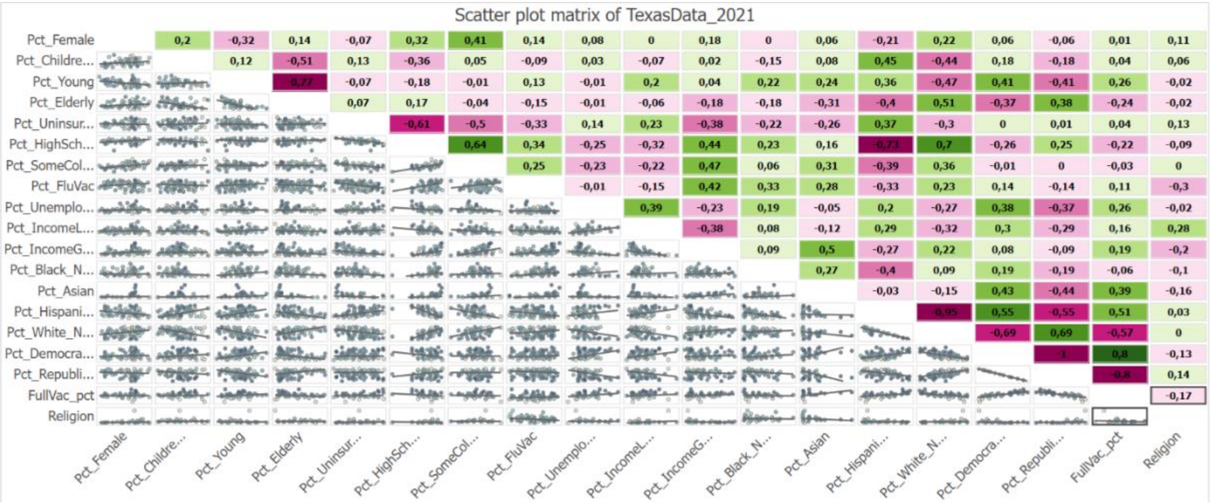

Distribution of Pct_HighSchool
Count
0 10 20 30 40 50 60 70
21,85 26,52 31,19 35,87 40,54 45,21 49,88 54,55 59,22 63,89 68,56 73,23 77,91 82,58 87,25 91,92 96,59
Pct_HighSchool
Mean : 81,81637
Median : 83,16372
StdDev : 8,50994


Distribution of Religion
120
80
40
0
Count
0
28,3
56,6
84,8
113,1
141,4
169,7
197,9
226,2
254,5
282,8
311,1
339,3
367,6
395,9
424,2
452,5
Religion
Mean : 63,31646
Median : 59,645
StdDev : 31,45211

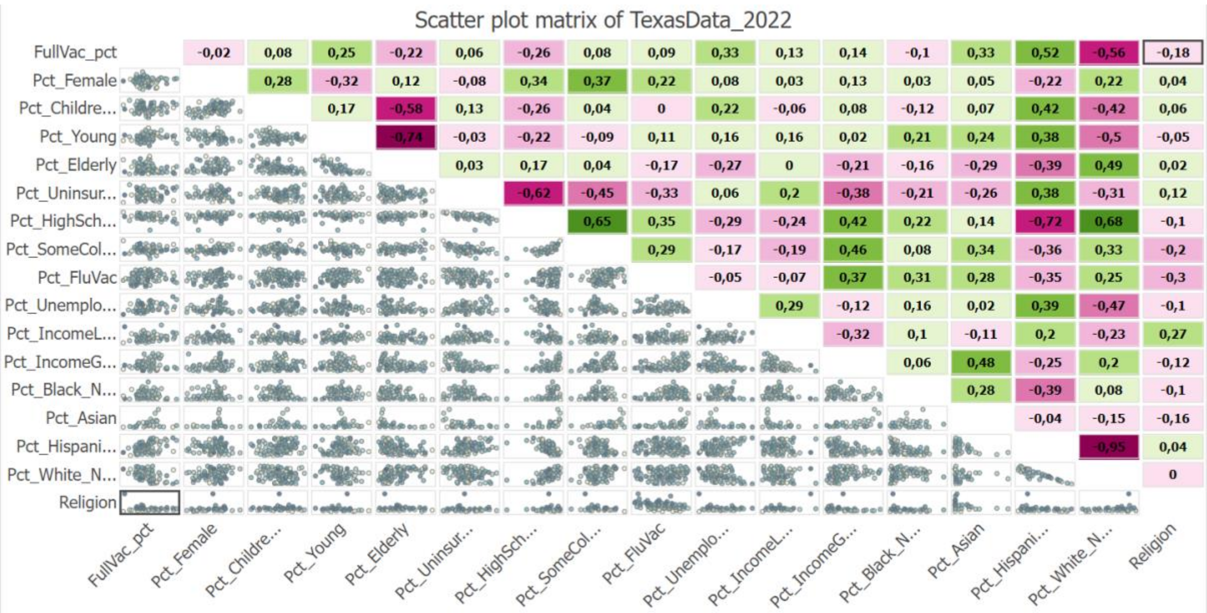
APPENDIX C

The R2 that we see on the graph is not the correlation coefficient but the coefficient of determination of the OLS regression line that is being fitted here.

Global relations for the variables in 2021

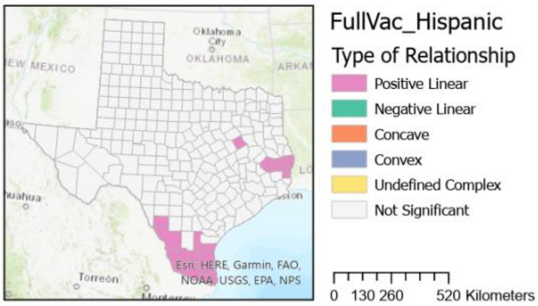


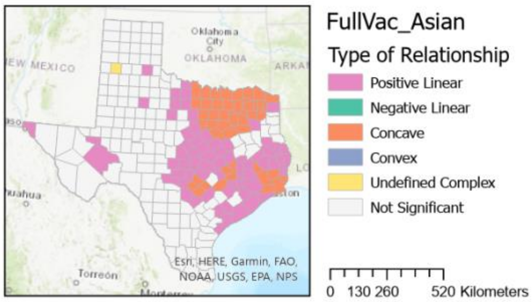
Global relations for the variables in 2022

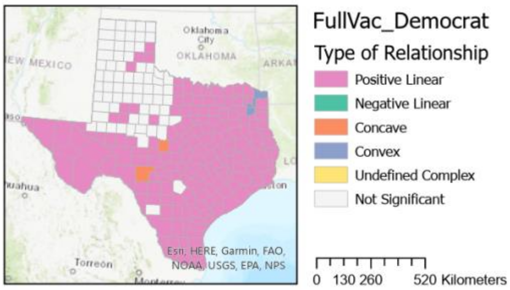


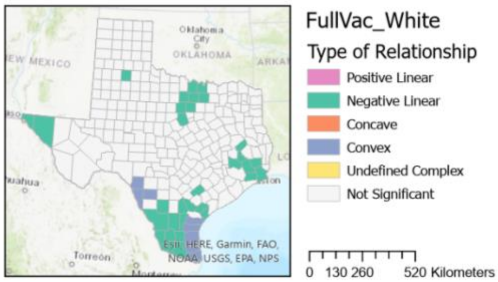
APPENDIX D

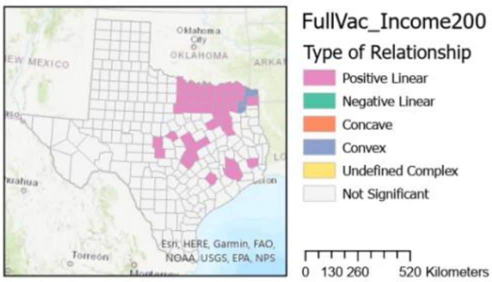
Local Relationships 2021 (Only included those variables that show significant results)

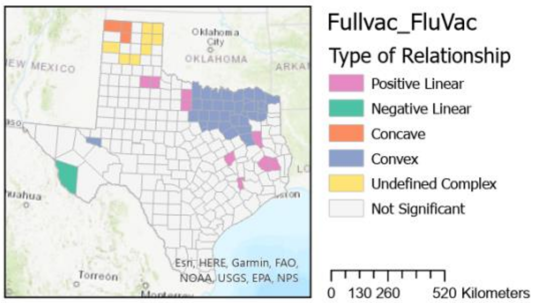


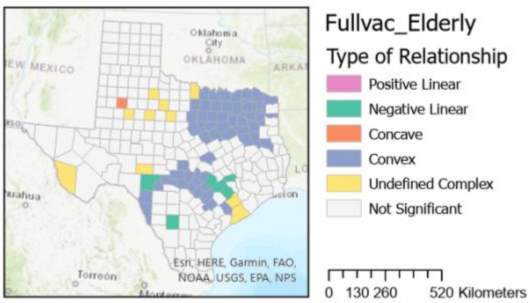


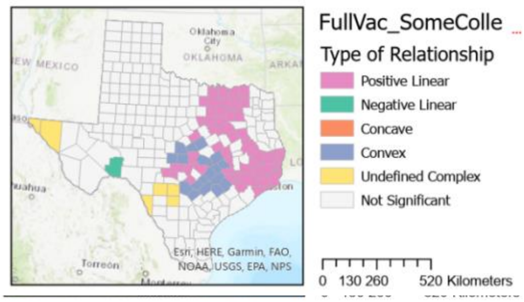


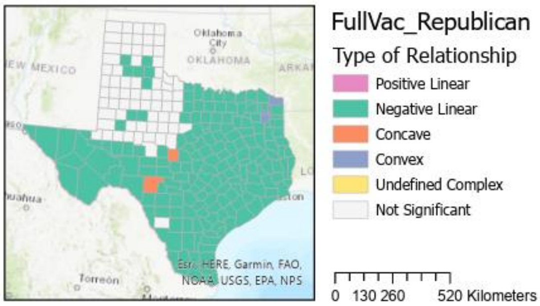




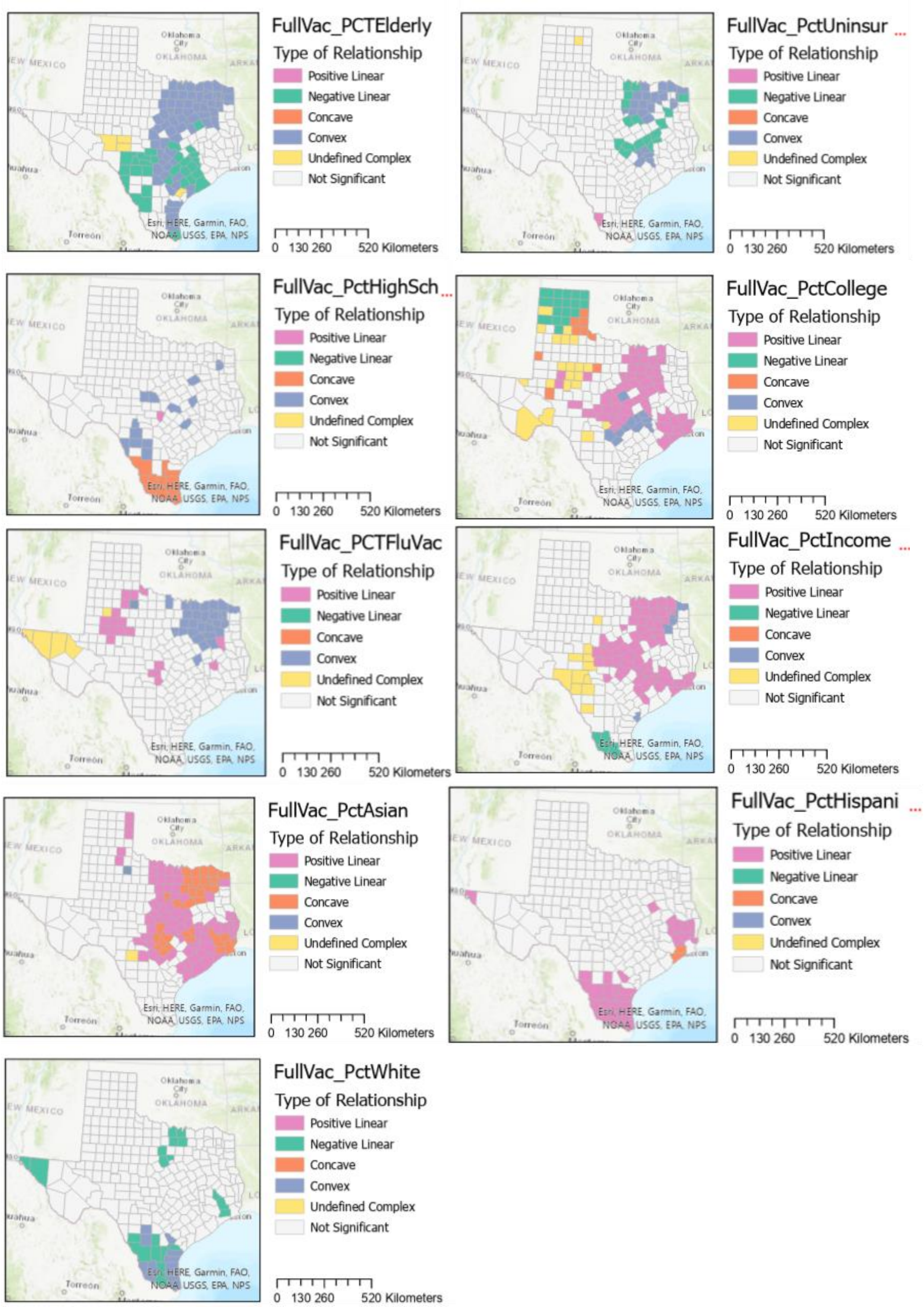






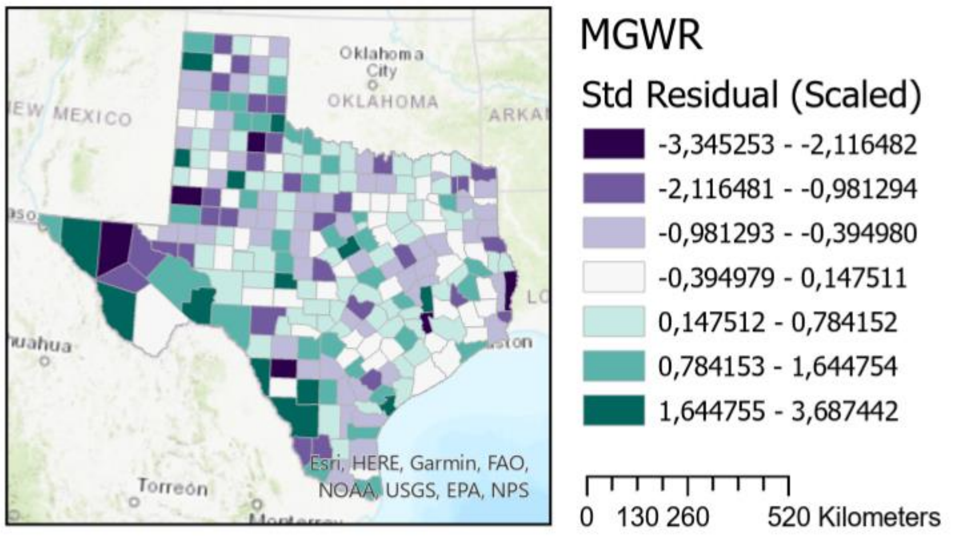


Local Relationships 2022 (Only included those variables that show significant results)



APPENDIX E

2021: MGWR (7 explanatory variables)



Bandwidth Statistics Summary

Explanatory Variables	Optimal Number of Neighbors	Effective Number of Parameters	Adjusted Value of Alpha	Adjusted Critical Value of Pseudo-t Statistics
Intercept (Scaled)	82	5,92	0,0085	2,6581
Pct_Uninsured (Scaled)	62	10,05	0,0050	2,8380
Pct_SomeCollege (Scaled)	252	1,45	0,0344	2,1293
Pct_FluVac (Scaled)	62	9,27	0,0054	2,8111
Pct_IncomeGT200000 (Scaled)	135	3,58	0,0140	2,4786
Pct_Asian (Scaled)	252	1,23	0,0407	2,0590
Pct_White_NH (Scaled)	135	3,63	0,0138	2,4841
Pct_Democrat2020 (Scaled)	147	3,17	0,0158	2,4332

Succeeded at 10 de maio de 2024 16:04:55 (Elapsed Time: 1 minutes 27 seconds)

Optimal Bandwidths Search History

Iterations	Intercept (Scaled)	Pct_Uninsured (Scaled)	Pct_SomeCollege (Scaled)	Pct_FluVac (Scaled)	Pct_IncomeGT200000 (Scaled)	Pct_Asian (Scaled)	Pct_White_NH (Scaled)	Pct_Democrat2020 (Scaled)	AICc
0	118	118	118	118	118	118	118	118	378,9396
1	82	62	252	82	115	232	117	122	363,9961
2	82	62	252	82	135	252	167	135	357,7390
3	82	62	252	82	135	252	147	135	356,0995
4	82	62	252	62	135	252	135	147	356,5481
5	82	62	252	62	135	252	135	147	356,1270
6	82	62	252	62	135	252	135	147	356,0185
7	82	62	252	62	135	252	135	147	355,9211
8	82	62	252	62	135	252	135	147	355,8290
9	82	62	252	62	135	252	135	147	355,7537

Summary of Explanatory Variables and Neighborhoods

Explanatory Variables	Neighbors (% of Features)[a]	Significance (% of Features)[b]
Intercept (Scaled)	82 (32,54)	105 (41,67)
Pct_Uninsured (Scaled)	62 (24,60)	19 (7,54)
Pct_SomeCollege (Scaled)	252 (100,00)	0 (0,00)
Pct_FluVac (Scaled)	62 (24,60)	0 (0,00)
Pct_IncomeGT200000 (Scaled)	135 (53,57)	149 (59,13)
Pct_Asian (Scaled)	252 (100,00)	0 (0,00)
Pct_White_NH (Scaled)	135 (53,57)	0 (0,00)
Pct_Democrat2020 (Scaled)	147 (58,33)	252 (100,00)

a: This number in the parenthesis ranges from 0 to 100%, and can be interpreted as a local, regional, global scale based on the geographical context from low to high.
b: In the parentheses, the percentage of features that have significant coefficients of an explanatory variable.

Model Diagnostics

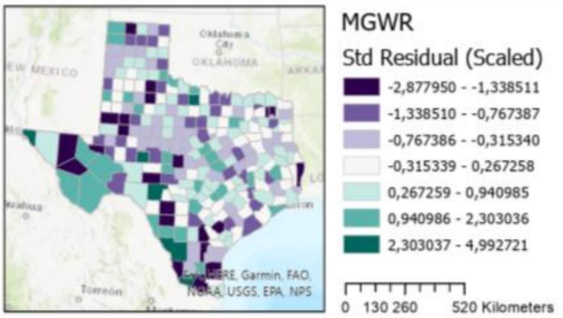
Statistic	GWR	MGWR
R-Squared	0,8171	0,8343
Adjusted R-Squared	0,7732	0,8044
AICc	378,9396	355,7537
Sigma-Squared	0,2266	0,1954
Sigma-Squared MLE	0,1829	0,1657
Effective Degrees of Freedom	203,4393	213,7069

Optimal GWR Bandwidth: 118 (K nearest neighbors).

Summary Statistics for Coefficients Estimates

Explanatory Variables	Mean	Standard Deviation	Minimum	Median	Maximum
Intercept (Scaled)	-0,0829	0,2168	-0,4834	-0,1095	0,3118
Pct_Uninsured (Scaled)	-0,0458	0,1409	-0,2941	-0,0504	0,3118
Pct_SomeCollege (Scaled)	-0,0649	0,0082	-0,0786	-0,0641	-0,0517
Pct_FluVac (Scaled)	0,0398	0,0662	-0,1370	0,0331	0,1743
Pct_IncomeGT200000 (Scaled)	0,1447	0,1067	-0,0686	0,1765	0,2646
Pct_Asian (Scaled)	0,0518	0,0025	0,0468	0,0528	0,0551
Pct_White_NH (Scaled)	-0,0061	0,0597	-0,1719	0,0046	0,1049
Pct_Democrat2020 (Scaled)	0,5533	0,0499	0,4799	0,5436	0,7551

2022: MGWR (7 explanatory variables)



Bandwidth Statistics Summary

Explanatory Variables	Optimal Number of Neighbors	Effective Number of Parameters	Adjusted Value of Alpha	Adjusted Critical Value of Pseudo-t Statistics
Intercept (Scaled)	95	4,26	0,0117	2,5433
Pct_Elderly (Scaled)	252	1,42	0,0353	2,1192
Pct_Uninsured (Scaled)	30	20,74	0,0024	3,0735
Pct_SomeCollege (Scaled)	49	10,67	0,0047	2,8601
Pct_FluVac (Scaled)	252	1,35	0,0371	2,0989
Pct_IncomeGT200000 (Scaled)	252	1,29	0,0387	2,0812
Pct_Asian (Scaled)	252	1,21	0,0414	2,0530
Pct_Hispanic (Scaled)	50	11,19	0,0045	2,8756

Summary Statistics for Coefficients Estimates

Explanatory Variables	Mean	Standard Deviation	Minimum	Median	Maximum
Intercept (Scaled)	-0,0552	0,2305	-0,4588	0,0283	0,2272
Pct_Elderly (Scaled)	0,1541	0,0046	0,1435	0,1536	0,1684
Pct_Uninsured (Scaled)	-0,0985	0,2555	-0,8371	-0,1459	0,7542
Pct_SomeCollege (Scaled)	0,0488	0,1214	-0,1794	0,0392	0,4135
Pct_FluVac (Scaled)	0,1164	0,0132	0,0833	0,1179	0,1370
Pct_IncomeGT200000 (Scaled)	0,0788	0,0123	0,0439	0,0849	0,0916
Pct_Asian (Scaled)	0,2286	0,0062	0,2221	0,2256	0,2459
Pct_Hispanic (Scaled)	0,5490	0,2065	0,0785	0,5937	0,8733

Model Diagnostics

Statistic	GWR	MGWR
R-Squared	0,7276	0,8071
Adjusted R-Squared	0,6636	0,7566
AICc	477,3245	435,7358
Sigma-Squared	0,3361	0,2432
Sigma-Squared MLE	0,2724	0,1929
Effective Degrees of Freedom	204,2724	199,8781

Optimal GWR Bandwidth: 118 (K nearest neighbors).

Summary of Explanatory Variables and Neighborhoods

Explanatory Variables	Neighbors (% of Features)[a]	Significance (% of Features)[b]
Intercept (Scaled)	95 (37,70)	91 (36,11)
Pct_Elderly (Scaled)	252 (100,00)	252 (100,00)
Pct_Uninsured (Scaled)	30 (11,90)	32 (12,70)
Pct_SomeCollege (Scaled)	49 (19,44)	10 (3,97)
Pct_FluVac (Scaled)	252 (100,00)	231 (91,67)
Pct_IncomeGT200000 (Scaled)	252 (100,00)	0 (0,00)
Pct_Asian (Scaled)	252 (100,00)	252 (100,00)
Pct_Hispanic (Scaled)	50 (19,84)	154 (61,11)

Optimal Bandwidths Search History

Iterations	Intercept (Scaled)	Pct_Elderly (Scaled)	Pct_Uninsured (Scaled)	Pct_SomeCollege (Scaled)	Pct_FluVac (Scaled)	Pct_IncomeGT200000 (Scaled)	Pct_Asian (Scaled)	Pct_Hispanic (Scaled)	AICc
0	118	118	118	118	118	118	118	118	477,3245
1	82	252	30	52	115	167	200	62	447,1022
2	95	252	30	50	200	252	252	50	438,4865
3	95	252	30	50	252	252	252	47	436,2205
4	95	252	30	49	252	252	252	50	435,9369
5	95	252	30	49	252	252	252	50	435,7485
6	95	252	30	49	252	252	252	50	435,7358

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa