YUSUF TAYO

Assignment 2

The Department of Health in Texas, decided to give out a relief fund to counties that are mostly affected by the covid-19 cases and areas with a high level of poverty. This necessitates the need to carry out a study to identify counties with a high percentage of cases and a high percentage of the population below the poverty line to receive the funding. To achieve this, I will be examining the pattern of spatial autocorrelation for the 2022 percentage of covid-19 cases and the percentage of population below the poverty line in the state of Texas. The hypothesis for the study would be categorized based on the two variables.

1. The Null hypothesis for the percentage of covid-19 cases is that there is no spatial autocorrelation in the percentage of covid-19 cases among counties in the state of Texas and are randomly distributed. And the First alternative hypothesis is that there is a positive correlation in the percentage of covid-19 cases among counties in Texas, while the second alternative hypothesis is that there is a negative correlation in the percentage of covid-19 cases among counties in the state of Texas.

2. The Null hypothesis for the percentage of population below the poverty line is that there is no spatial autocorrelation in the percentage of population below the poverty line among counties in the state of Texas and are randomly distributed. And the First alternative hypothesis is that there is a positive correlation in the percentage of population below the poverty line among counties in Texas, while the second alternative hypothesis is that there is a negative correlation in the percentage of population below the poverty line among counties in the state of Texas.

In my analysis, I will be calculating a Global Moran's I spatial weights matrix to understand the clustering of the two variables I will be examining, which are: the percentage of covid-19 cases (PER_CASE); and the percentage of the population below the poverty line (RAT_POV) and their permutations to know the significance that would assist me in making my decision. And to ascertain the robustness of my analysis, two types of this matrix will be calculated. The queen contiguity and the K-nearest neighbors' matrix.

The Local indicators of spatial association (LISA) is another spatial weight matrix I will be calculating to examine the percentage of covid-19 cases and percentage of population below the poverty line, showing

their significance and cluster map using only the queen contiguity matrix filtered at 5%. This will assist me in understanding the cluster and locations of the affected counties.

Comparison of Global Moran's I for Percentage of Cases

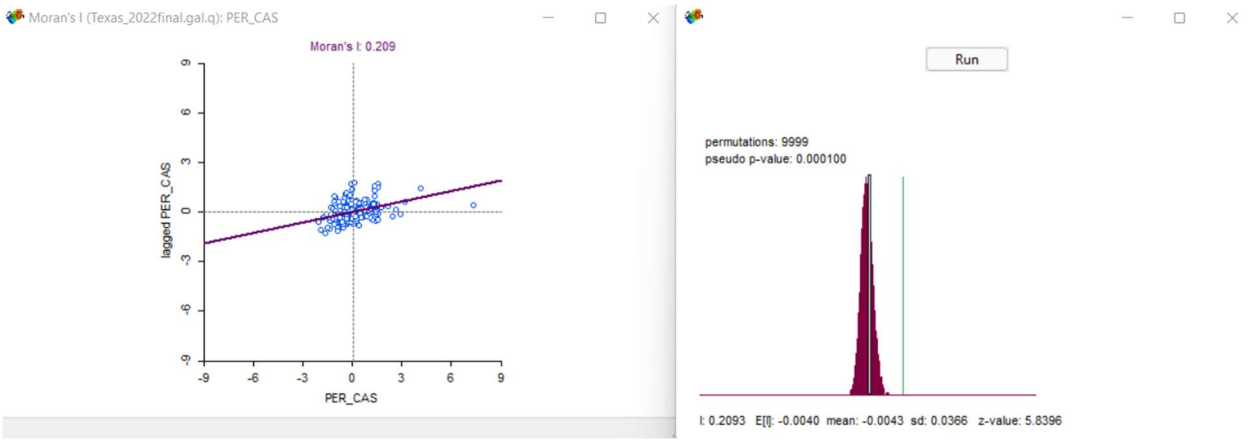

Figure 1: The scatter plot at the left and permutations at the right of Global Moran's I for percentage of cases using a queen contiguity matrix

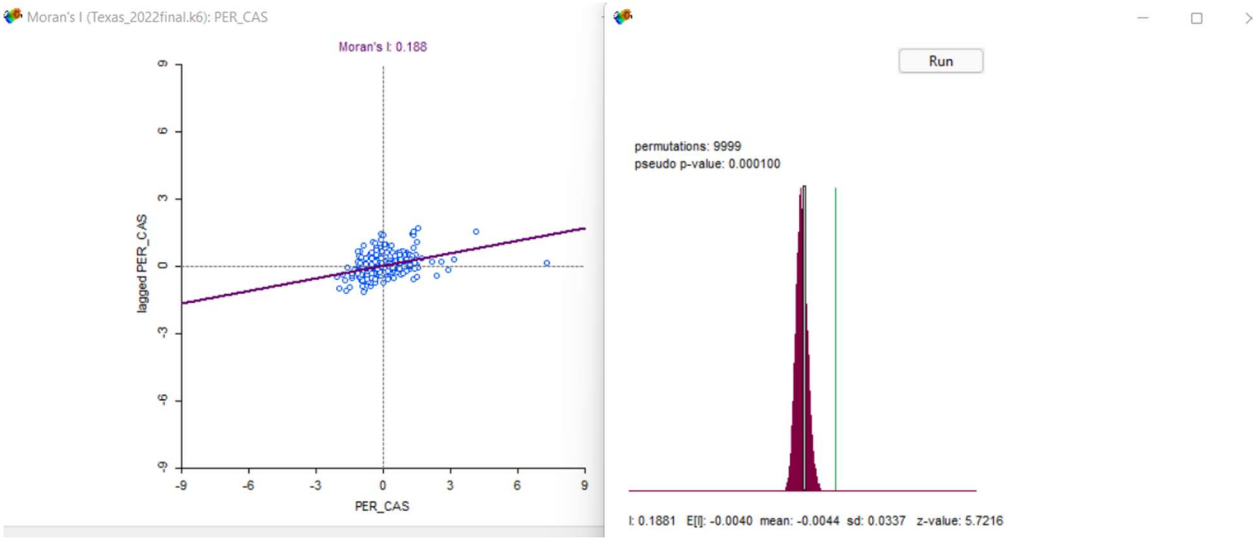

Figure 2: The scatter plot at the left and permutations at the right of Global Moran's I for percentage of cases using a K-nearest neighbors' matrix

As shown in figure 1, the scatter plot of Global Morans I for the percentage of cases using a queen contiguity matrix shows that there is a clustering of covid-19 cases at the center of the scatter plot and

the permutation with a P-value of 0.0001 allows us to reject the null hypothesis for the percentage of covid-19 cases as the P-value is below 0.05 and accept the first alternative hypothesis since the P-value is positive. Similarly, figure 2 displayed the scatter plot of global morans I for the percentage of cases using K-nearness neighbors matric where K is six (6) and its permutation has a P-value of 0.0001 which also allow us to reject the null hypothesis. The two matrixes rejected the null hypothesis and have the same p-value of positive 0.0001 that allowed us to accept the first alternative hypothesis wish states that there is positive correlation in the percentage of covid-19 cases among the counties in the state of Texas. This shows how robust the data and the analysis are.

Also not that in understanding the K value, I plotted an histogram, click on the highest bar to see it value and compared it with the mean score which gives me six (6)

Comparison of Global Moran's I for Percentage of people below poverty line

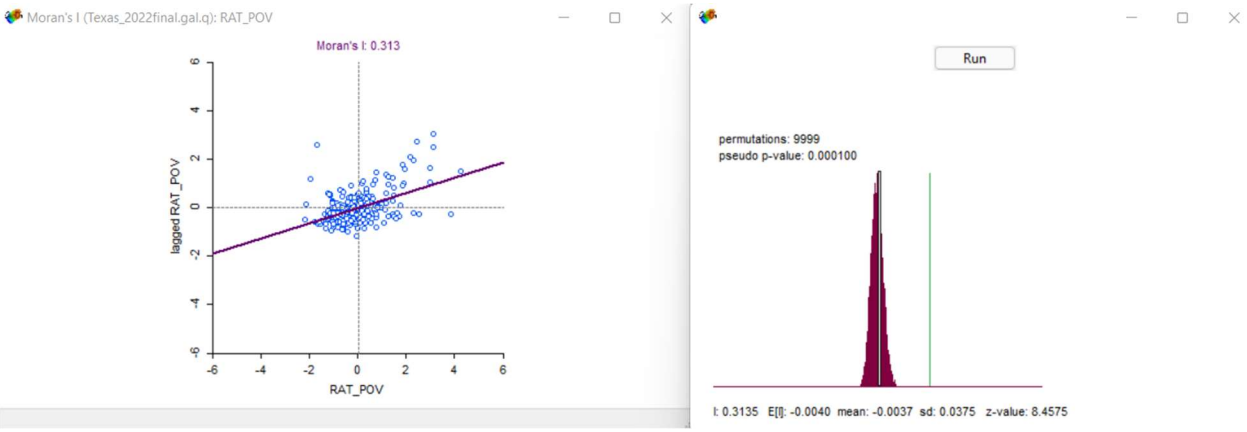

Figure 3: The scatter plot at the left and permutations at the right of Global Moran's I for the percentage of population below poverty line using a queen contiguity matrix

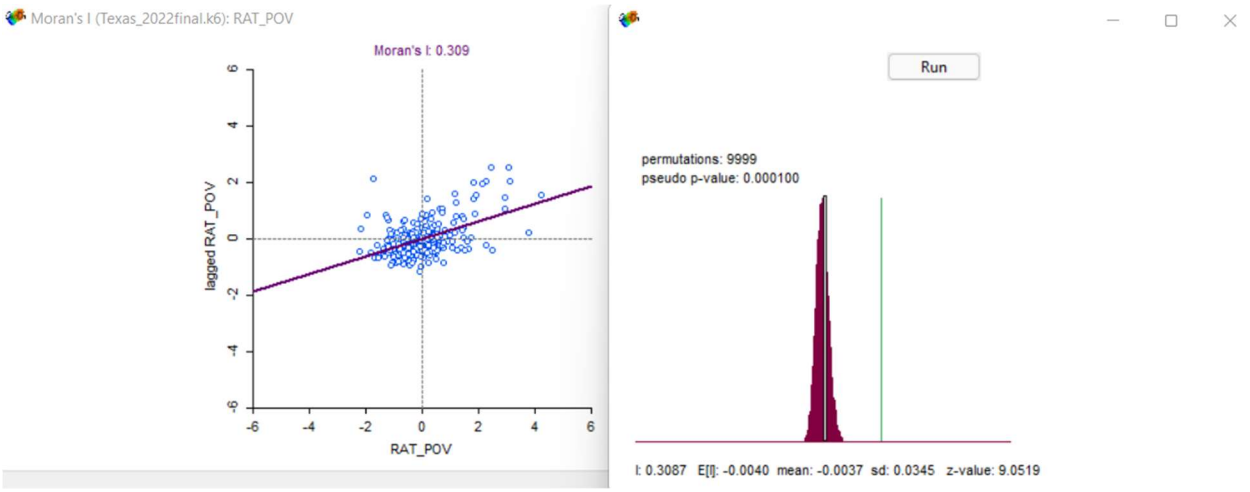

Figure 4: The scatter plot at the left and permutations at the right of Global Moran's I for the percentage of population below poverty line using a K-nearest neighbors' matrix

It was revealed in figure 3 that the scatter plot, which shows the global morans I for the percentage of population below the poverty line using the queen contiguity matrix, has the majority of its population clustering around the center, and its permutation with a p-value of 0.0001 allows us to reject the null hypothesis and accept the first alternative hypothesis as the value is positive. Similarly, figure 4, which is the scatter plot of the global morans I for the percentage of population below the poverty line using K-nearness neighbors matrix where K is six (6), display a similar value with its permutation P-value of 0.0001 also allow us to reject the null hypothesis and accept the first alternative hypothesis which state that there is a positive correlation in the percentage of people below poverty line among the counties in the state of Texas.

The Local Indicators of Spatial Association (LISA)

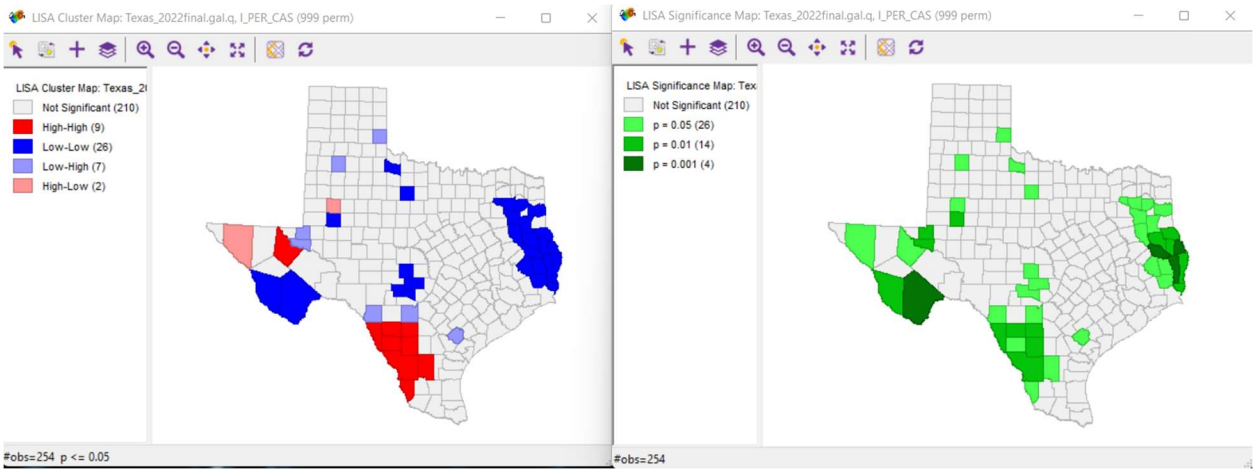

Figure 5: The LISA for the Percentage of Covid-19 Cases Cluster Map at the left and the Significant Map at the right filtered at 0.05.

Figure 5 displays the cluster and the significance map of the percentage of covid-19 cases among the counties in the Texas state. From the cluster map it was revealed that 9 counties have high percentage of covid-19 cases with high lagged percentage of cases, 26 counties have Low-Low, 7 counties have Low-High, and 2 counties have High-low. However only counties where we have high-high, and High-low will be given consideration. And on the right side, the significance map indicates that when the percentage of cases was filtered at 0.05, we have 26 counties and when filtered at 0.01 we have 14 counties.

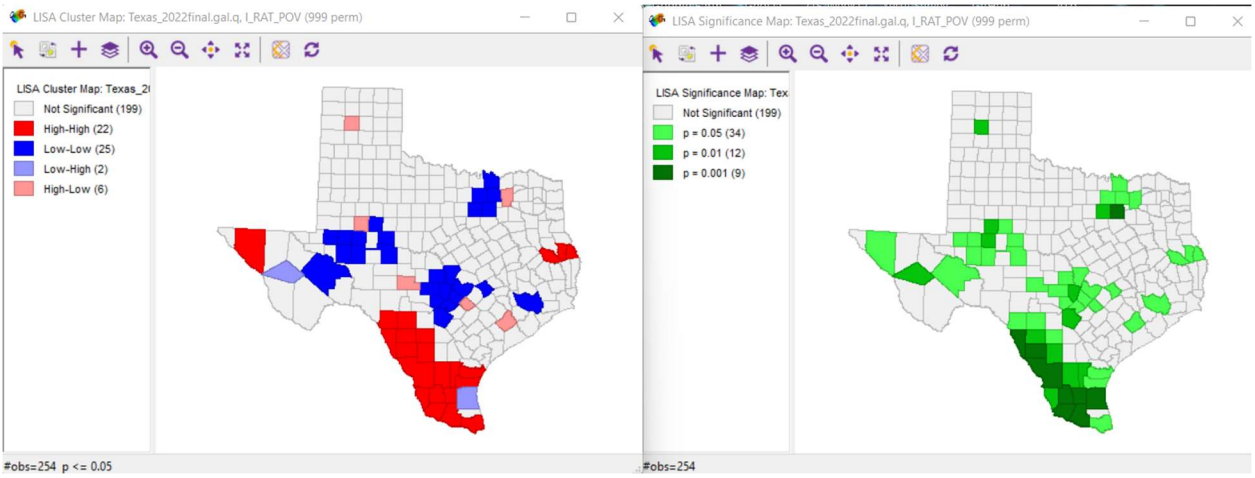

Figure 6: The LISA for the Percentage of people below poverty line Cluster Map at the left and the Significant Map at the right filtered at 0.05.

As displayed in figure 6, the cluster map shows that twenty-two (22) counties have high percentage of people below poverty line and high levels of lagged percentage of people below poverty line, while

twenty-five (25) counties have low-low, two (2) counties have low-high, and six (6) counties have high-low. However, counties with high-high, and high-low will be given consideration.

COMPARISON OF THE MORANS I VALUE FOR THE SPATIAL WEIGHT MATRCES.

Table 1: The comparison of the Moran's I value for the percentage of cases and the percentage of people below poverty line for both spatial weight matrices

Variable	Queen Contiguity Matrix		K6-Nearest neighbor Matrix	
	Standardized I	P-value	Standardized I	P-value
PER_CASE	5.8396	0.0001	5.7216	0.0001
RAT_POV	8.4575	0.0001	9.0519	0.0001

Table 1 summarize the results of the calculations of the Global Moran's I value, the P-value and the standardized (Z) value to assist in making decision on the best spatial weight matrix to select. Since the standardized value is used to measure the most appropriate spatial weight matrix, I will be selecting the queen contiguity matrix for the percentage of covid-19 cases and the k6-nearest neighbor matrix for the percentage of population below poverty line. This is to further analyze and identify the most affected

counties in the state of Texas. Counties where these two variables overlapped will be our final selected counties to be considered for the funding.

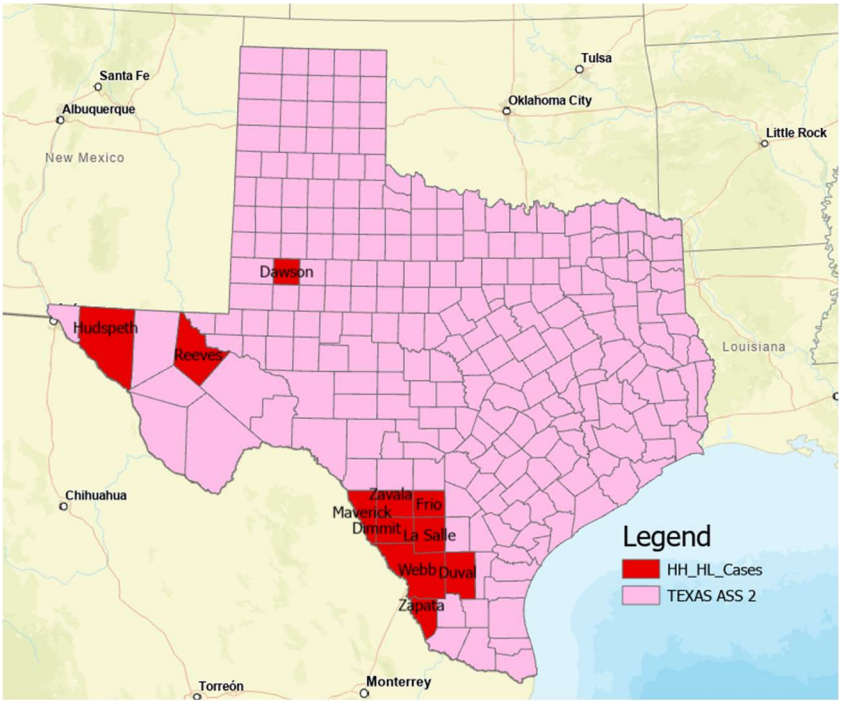

Figure 7: Map showing counties with High-High or High-Low percentage of covid--19 cases

As shown in figure 7, eleven counties namely: Dawson, Reeves, Hudspeth, Maverick, Zavala, Dimmit, Frio, La Salle, webb, Duval, and Zapata were the counties with either High-High or High-Low of percentage of covid-19 cases in the state of Texas.

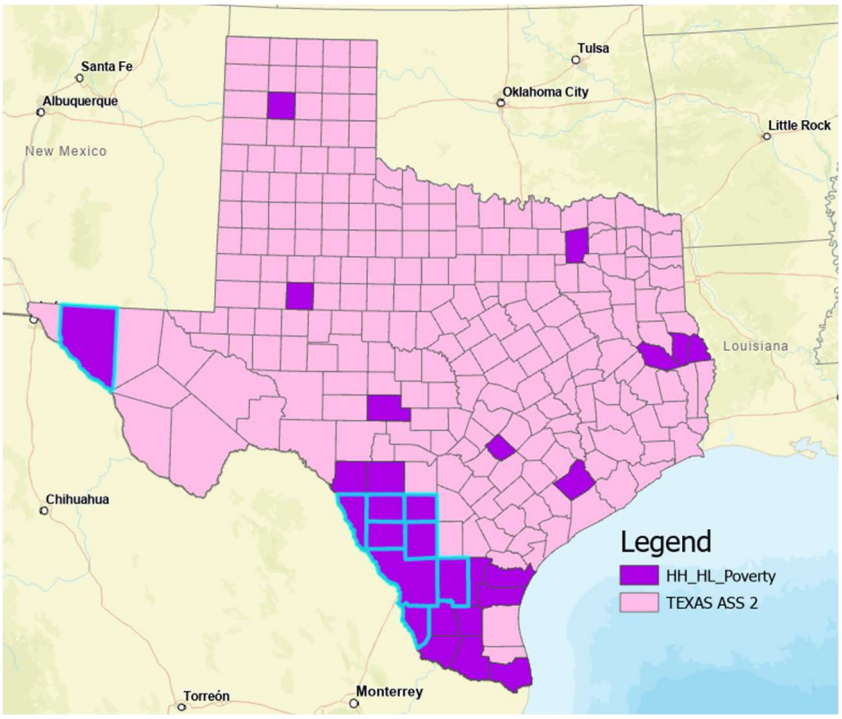

Figure 8: Map showing counties with either High-High or High-Low percentage of population below poverty line

Figure 8 identified twenty-eight (28) counties with either High-High or High-Low percentage of population below poverty line. These counties will be further examined to identified those that overlapped with the high percentage of cases.

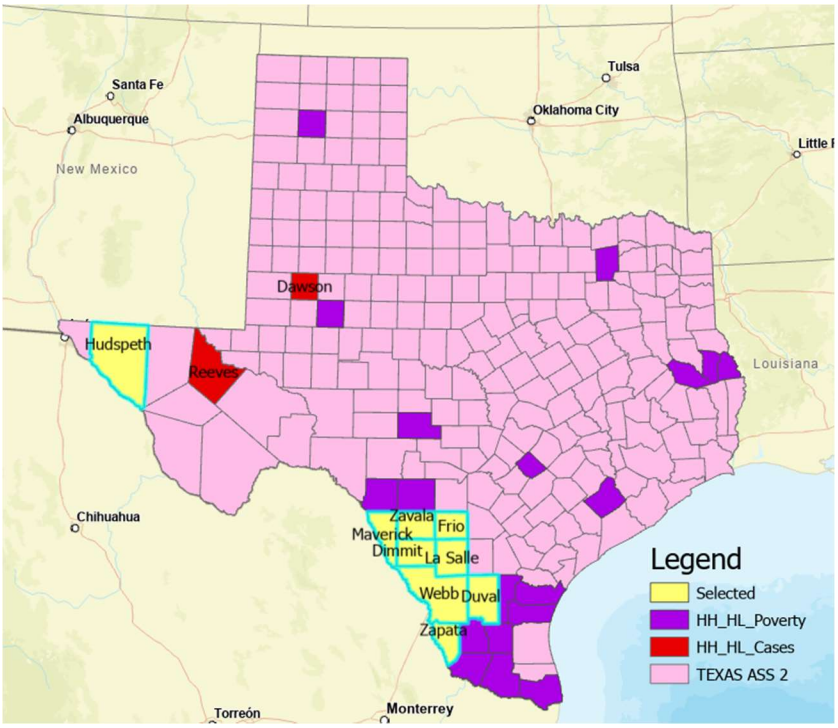

Figure 9: Map showing the counties with HH or HL percentage of covid-19 cases and percentage of population below poverty line

As shown in figure 9, the selected counties are those who have either High-High or High-Low percentage of covid-19 cases and percentage of population below poverty level. These counties are nine (9) namely: Hudspeth, Maverick, Zavala, Dimmit, Frio, La Salle, Webb, Duval, and Zapata. They would be the counties that would be considered for the funding by the department of Health, of the state of Texas.

Conclusion

After running series of calculations, plots, and maps in order to identified counties that possess the Highest level of financial needs based on the set parameters, nine counties were identified and found worthy of the state of Texas Department of Health funding assistance programs. They are: Hudspeth, Maverick, Zavala, Dimmit, Frio, La Salle, Webb, Duval, and Zapata.

It is important here to note the geographical location of these counties. These counties are located close to the united state border with Mexico at the south-eastern of the state of Texas. This explains the effect location can have on the demography and characteristics of certain locations.

It can be said that the high poverty rate at these counties is related to the country they share border with. And it would be necessary to further probe the sources to inform policies that can drastically reduce the rate of poverty.