Evaluation of the Relationship Between Household Income and the Percentage of High School Students Who Decided to do Remote Learning in Dallas Independent School District

Geo 327G

GIS & GPS Applications in the Earth Sciences

Fall 2020

Dr. Mark Helper

Semester Project

Prepared By: Meredith Martin

December 7, 2020

A. Background

In March of 2020, the United States was hit by the spread of the COVID-19 virus. Many aspects of daily life were brought to a halt due to the fear of the rapid spread of this deadly virus. One of the aspects that was hit the hardest were schools as they were forced to rapidly shift all lessons and activities to virtual platforms. Students were thrown into taking classes through computers in their homes and teachers forced to teach entire classes from their homes using various virtual resources. As the 2019-2020 school year came to a close, a lot was learned, and it was now up to schools to prepare for the 2020-2021 school year if the precautionary conditions of COVID-19 were to continue. As August quickly approached, many students were given the option to either continue instruction through virtual platforms or return to in-person instruction with appropriate safety measures. Every parent and student had their own reasons for either going back to school in person or continuing classes from their own home. It could have been for safety reasons, for economic reasons, or numerous other personal reasons.

B. Problem Formulation

The purpose of this project is to see if there is a correlation between household income and the percentage of high school student in the Dallas Independent School District (DISD) that chose remote learning. Are students that are living in areas that reported incomes in the lower to lower-middle income class range more likely to opt out for remote learning instead of going to school in person?

To evaluate this relationship between household income and percentage of remote learning, many different actions were performed in ArcGIS and Excel to find the percentage of area within each high school attendance zone in DISD that had census block groups that reported a household income less than $42,000 in the 2010 Census. Then that percentage was compared to the percentage of students that decided to do remote learning for each high school for the 2020-2021 school year in order to see if there was a correlation.

C. Data Collection

First, an article by *US News*[1] provided an income value that is considered lower class and lower-middle class. According to this article, an income of $31,000 or less is considered in the lower income group and an income of $31,000 to $42,000 is considered in the lower-

middle income group. That is why a household income of $42,000 or less was considered the baseline in this evaluation.

In order to find the household income, Census.gov[2] provides shapefiles of the census block groups for all the states filled with different demographic data. When this is opened up in ArcMap, there were hundreds of fields of data in the attribute table that have different codes that corresponded to different types of demographic data. In order to more efficiently manipulate this data in ArcMap, the fields that was not needed had to be removed. Census.gov[3] also provides a link to the corresponding metadata which provides a key to which codes belong to what demographic data, so all the codes that did not correspond to household income values were removed.

The Dallas Independent School District sent out a survey in August to all of the parents of students in the district to see how many students were going to do remote learning or return to campus. The data received from this survey was published in a chart in an article published by *The Dallas Morning News*[4]. The chart included the name of the school, the total number of students who attend the school, the percentage of survey completion, and the normalized percentage of students who chose remote learning and the percentage of students who went in person based on the people who completed the survey.

The Dallas Independent School District provides shapefiles for the attendance zones for each level of school in the district[5]. They also provide point data for the location of each of their campuses within the district. The data used in this evaluation that was provided by DISD is represented in the map provided in figure 1.

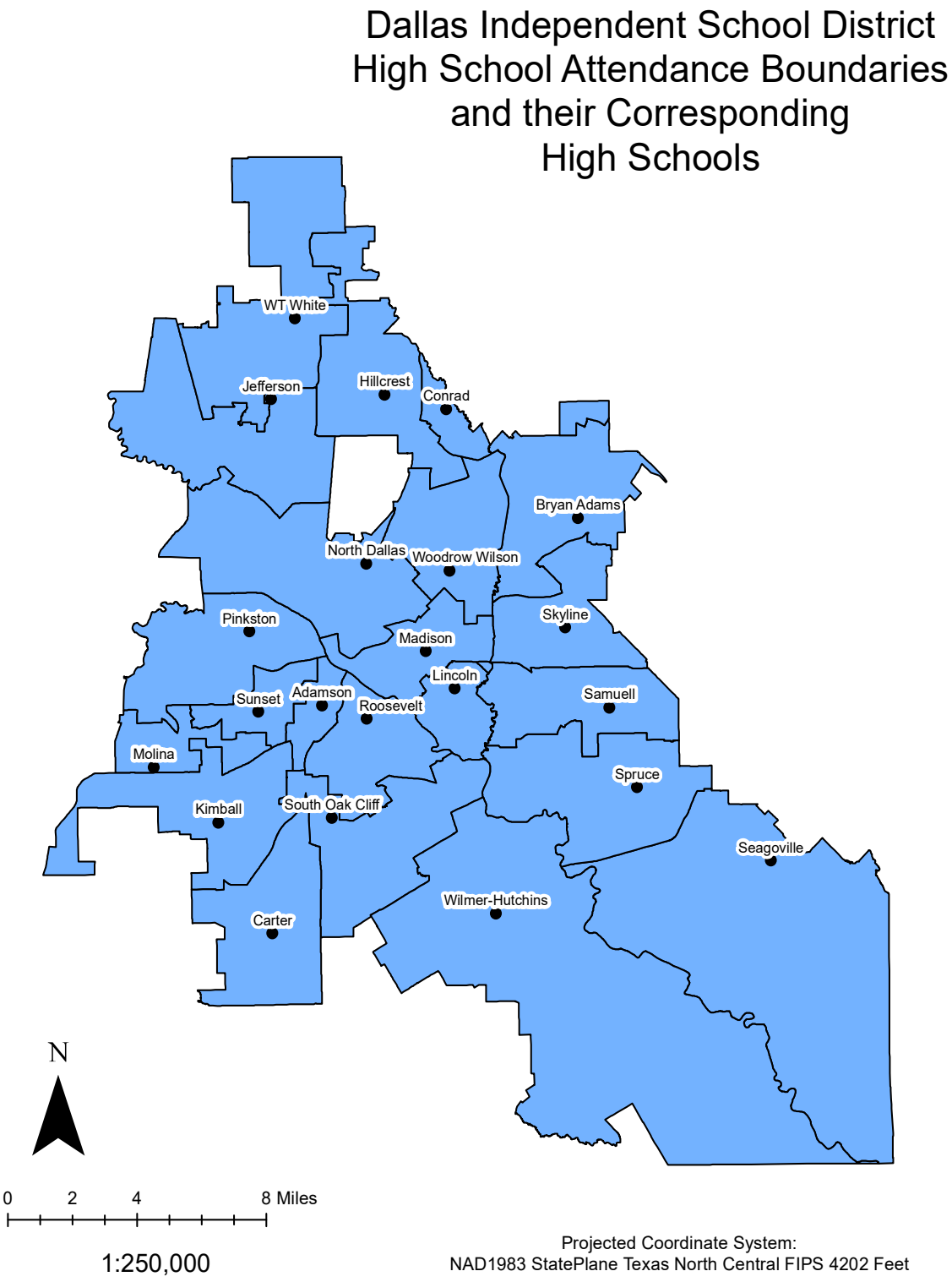

Figure 1. DISD high school attendance boundaries with the corresponding high schools

D. Data Preprocessing

Before any evaluation could be done, all of the GIS data needed to be put into the same projected coordinate system. The high school attendance boundaries and the campus points had the same projected coordinate system; however, the Census block group data was only defined by a geographic coordinate system. First, the block group data for Texas was pulled into ArcMap as well as the high school attendance boundaries. In order to give the census block group shapefile a projected coordinate system, the Project tool was used. (Arc Toolbox →Data Management →Projections and Transformations →Project). The block group layer represented as "ACS_10_5YR_BG_48_TEXAS" in Figure 2 was selected as the input dataset. The correct storage location was chosen, then the same coordinate system as the high school attendance boundary was chosen as the Output coordinate system, in this case it is NAD83 2011 StatePlane Texas North Central FIPS 4202 Feet.

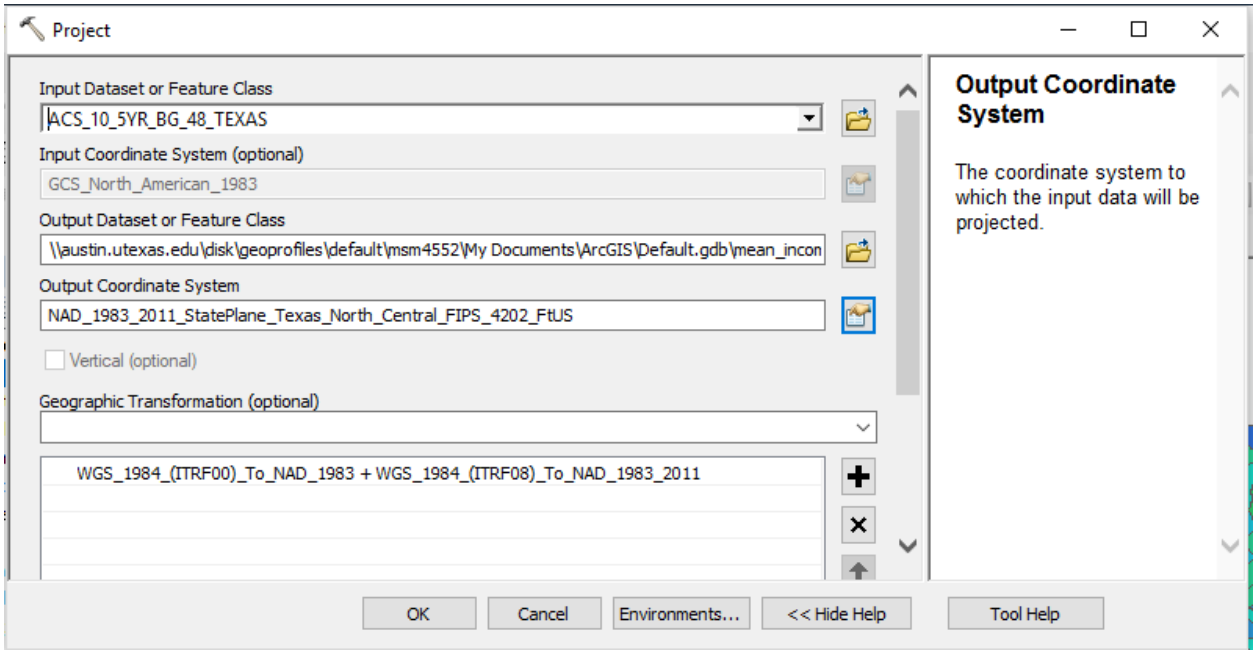

Figure 2: Projecting the Census Block Group data to the correct projected coordinate system using the Project tool

The block group data that was provided was for the entire state of Texas, so that data needed to be clipped to the outer boundary of the Dallas Independent School District. In order to do that, the Clip tool was used. The search function in ArcMap was used to access this tool. The input feature was the block group data, and the clip feature was the high school attendance boundary, also known as "DISD_High" then the appropriate storage location was

chosen. This process was not able to be performed when all of the provided fields for the census data were present. That is why only the appropriate household income fields were kept in the attribute table. Then a new census block group layer was created that fell within the boundaries of DISD. This operation is shown in figure 3 and the result of this operation is represented in figure 4 which also shows where in Texas this evaluation is focused.

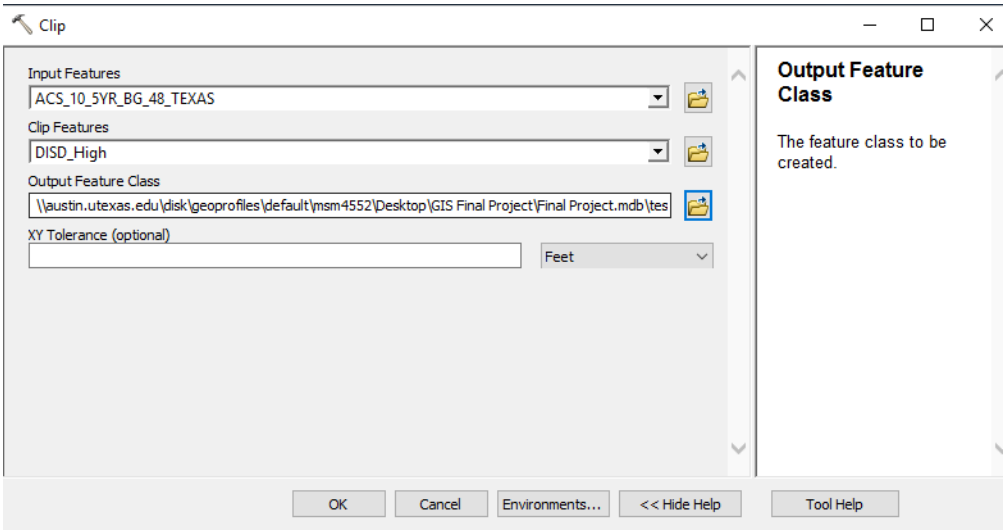

Figure 3. Clipping the block group data to the outer boundary of DISD

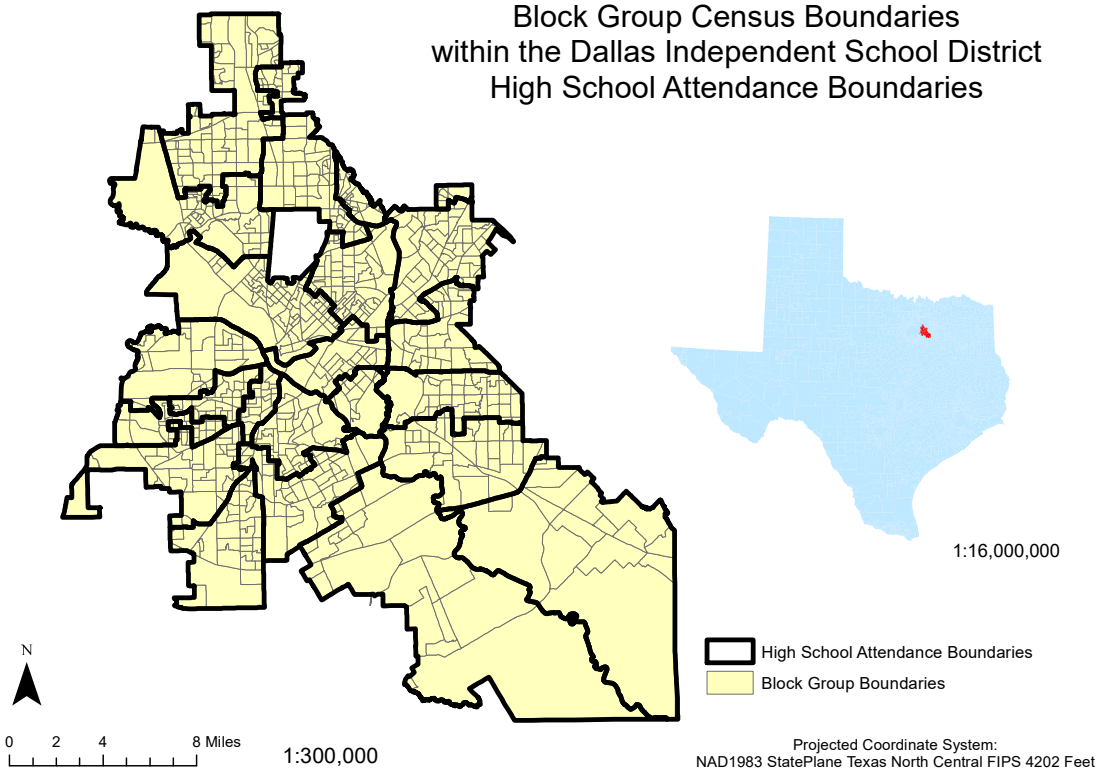

Figure 4. Block Group Boundaries Clipped to the Dallas Independent School District Outer boundary; includes an inset of where in Texas this is focuses

Next, the percentage of remote learning data per high school was added to the corresponding polygon for that high school's attendance zone. In order to do this, a new field was added to the "DISD_High" layer called "Remo_Learn". Then an editing session was started. The DISD_High layer's attribute table was opened, then the percentage of students that chose remote learning that came from the survey published by *The Dallas Morning News* was added under the "Remo_learn" field. Once all of the data has been added to this new field, the edits were saved, and the editing session was ended.

Now all of the data needed in for this evaluation has been imported into ArcGIS and is ready for analysis.

E. ArcGIS Processing

The main goal in the ArcGIS processing was to determine the area within each high school attendance boundary that had block groups that reported a household income less than $42,000. Then that area could be turned into a percentage by area within each high school attendance boundary that had a household income that fell in the low to lower-middle income group.

The first step was to determine which block groups fell within each high school attendance boundaries. In order to do that, the Spatial Join tool was used (Arc Toolbox → Analysis Tools → Overlay → Spatial Join). The target feature was the newly clipped block group layer also known as "mean_income". The join feature was the DISD_High layer and then the appropriate storage location was selected. Attributes that were not needed to be spatial joined were deleted and then for "match option", "have their center in" was selected. This meant that when the layers were spatial joined and there happened to be a block group that fell between two attendance zone boundaries, whatever attendance zone that block group had its center in was the attendance zone assigned to that block group. A new layer was then created that contained these newly spatially joined layers This operation can be seen in figure 5.

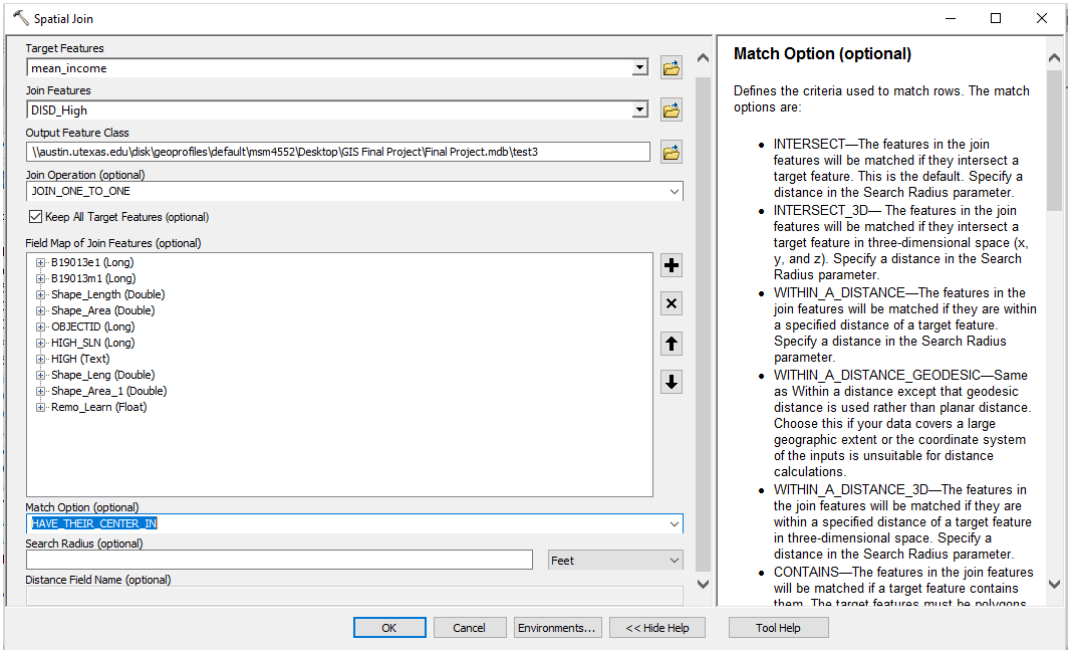

Figure 5. Using the Spatial Join tool to combine the DISD_High layer to the newly clipped block group data

After this step however, an issue began to appear. The whole point of the analysis was to determine the area of block groups within each attendance boundary zone that reported an income of less than $42,000. Although the block group data now had which attendance zone they fell in, which was useful for the project, because the match option chosen was “Have their center in”, there was block group data that did not correctly correspond to the right high school zone. So, when the area of block groups that fell below $42,000 for household income were compared to the area of the attendance zones, there were some percentages that were above 100%. In order to fix this discrepancy, the Intersect tool was used.

According to ArcGIS.com[6], the intersect tool “Computes a geometric intersection of the input features. Features or portions of features which overlap in all layers and/or feature classes will be written to the output feature class.” To use this tool, geoprocessing was selected at the top of the ArcMap window, then “intersect” was selected. In the intersect window, the two layers that needed to be intersected were the DISD_High layer and the clipped block group data. This function combined all of the attributes in those two layers and produced a new layer that had the maximum number of polygons that were formed by the intersection of the two layer’s polygons. This process can be seen in Figure 6.

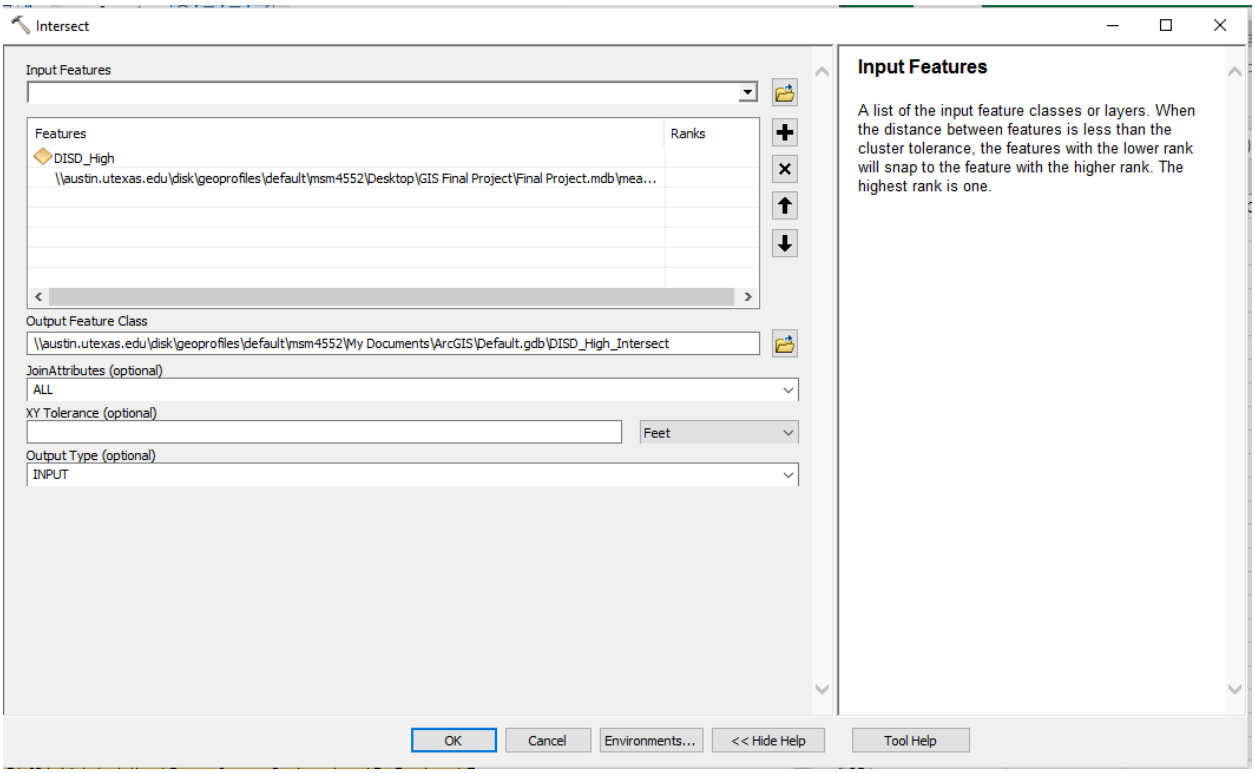

Figure 6. Using the Intersect tool to combine the DISD_High layer to the newly clipped block group data more effectively than the Spatial Join tool

Now that the two layers had been combined more effectively, it was now time to select the block groups that reported a household income less than or equal to $42,000 in order to determine the ones that are considered in the lower to lower-middle income group. In order to do this, the polygons were selected by attribute. At the top of the ArcMap window, under Selection, "Select by attribute" was used. In the select by attribute window, the newly intersected layer that was created was selected (the layer chosen in figure 7 was before intersect was chosen as the right tool for analysis, so disregard the spatial join layer). For method, "create a new selection" was chosen , then the attribute that contains the household income (B19013e1) was selected by double clicking, which put it in the box at the bottom of the window. In order to select the block groups that are below $42,000, "[B19013e1] <= 42000" was typed into the bottom box. This process is shown in figure 7.

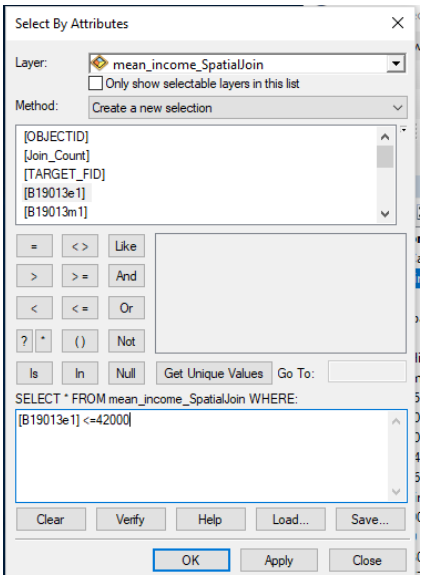

Figure 7: Using the select by attribute tool to select the block groups that have a household income less than $42,000

In order to make the analysis of the area within each attendance zone that had a household income of less than $42,000 more efficient, the currently selected data was turned into its own layer. The layer in the table of contents that currently contains all the selections was double-clicked on and the Data option was chosen then the “export data” option was selected. A pop-up window will appear and it will ask if the currently selected features are what is to be exported and then the appropriate storage location was chosen. This new layer was then saved with the name “low_intersect”. This process is shown in figure 8. After hitting save, another pop-up window will come up that allows for the newly exported data to be imported into the current data frame. The outcome of this process can be seen in the yellow areas in the map in Figure 9.

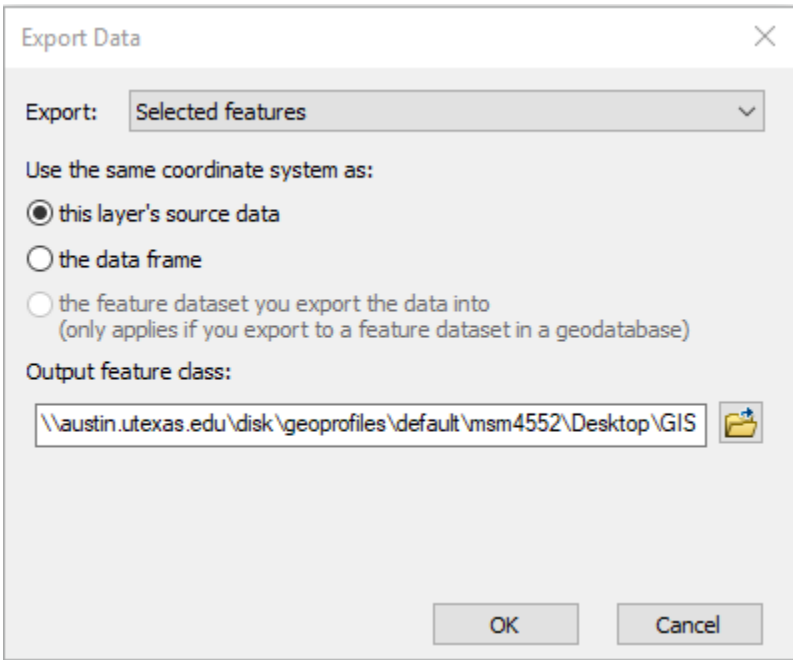

Figure 8. Exporting the selected features and making it its own layer

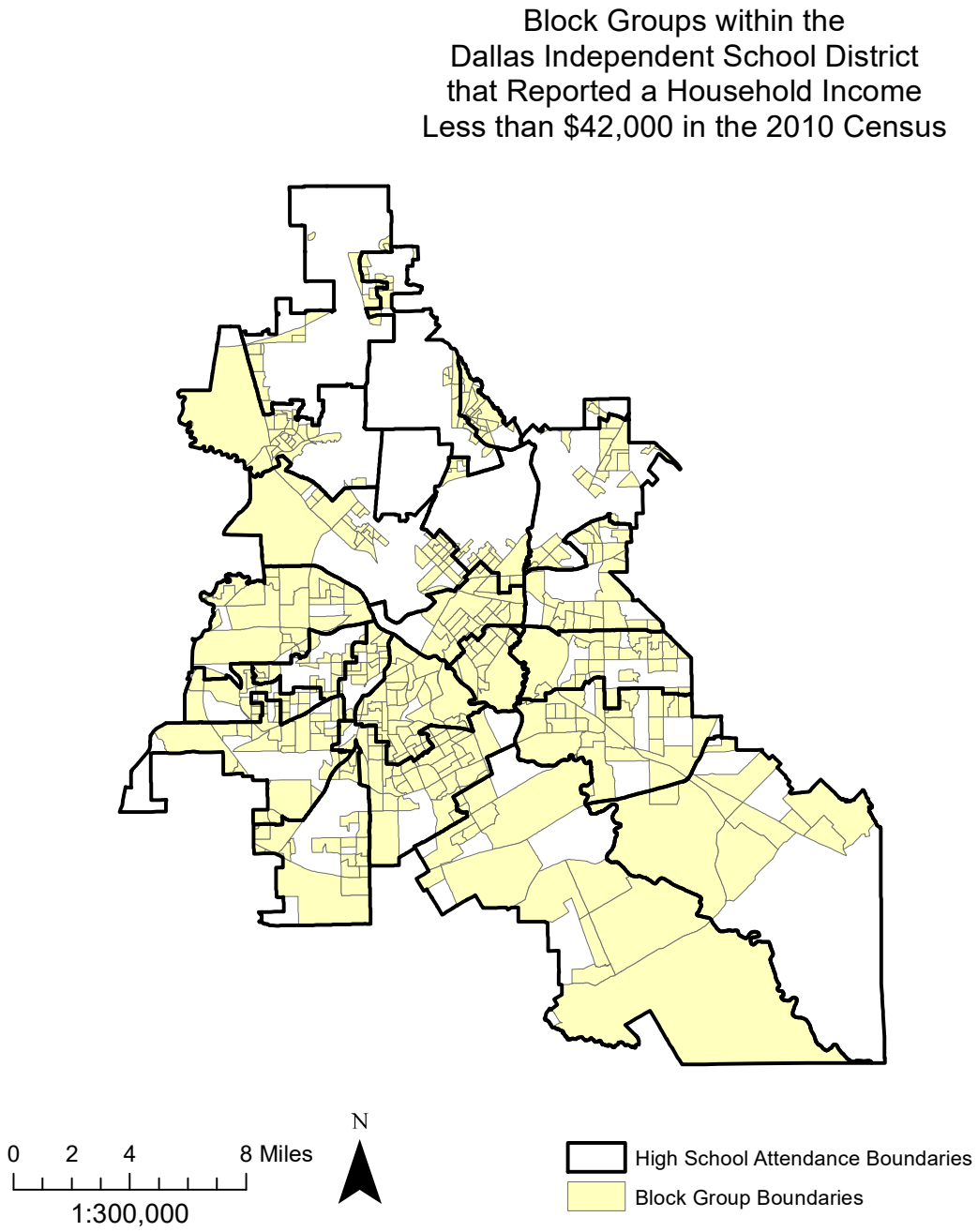

Figure 9. Map that shows the block groups within each attendance zone boundary that reported a household income of less than $42,000 in the 2010 Census. This is the result after intersecting the two data layers, selecting the desired block groups, and exporting them into their own layer.

Next, in order to determine the percentage by area that reported a household income less than $42,000 per high school attendance boundaries, the total area of each high school attendance zone needed to be determined. This data is under the “Shape_Area” field in the “DISD_High” layer attribute table. This data is in square feet based on the projected coordinate system that this project was given. Then, the total area for each attendance boundary was copied over to an excel sheet where it will be compared to the area within each zone that reported a household income less than $42,000.

To determine the area within each attendance zone that had block groups that reported a household income in the low to lower-middle income zone, the attribute table for the last made layer, named “low_intersect” was used. The field named “Shape_Area” provides the area in square feet of each polygon that has a reported income of less than $42,000. The attribute table was then sorted by high school name which means the polygons were organized by which high school attendance boundary they fell in. Using the Shift button on the keyboard, all of the polygons within in one attendance boundary were selected. Then at the bottom of the attribute window, the “show selected records” was clicked which would only show the polygons of interest. In order to get the total area of the combination of the polygons, the Statistic tool was used by left clicking on the “Shape_area” field name, then clicking on “statistics” then a “Selection Statistics of low_intersect” will pop up, shown in figure 10.

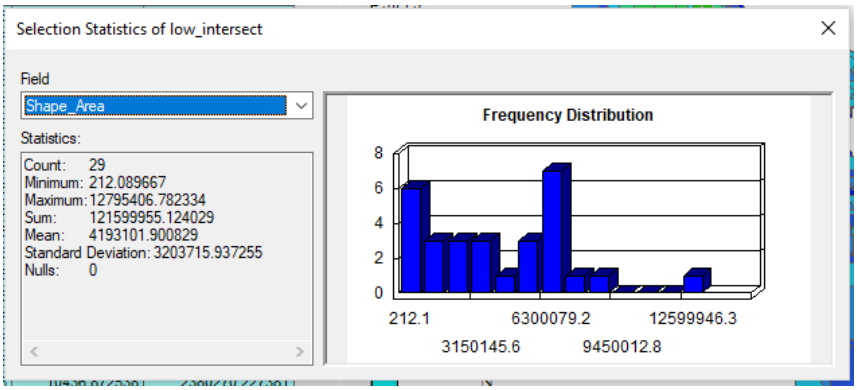

Figure 10. Statistics of the “shape_area” field for the selected polygons that will give the total area of the block groups that have a household income of less than $42,000 per high school attendance boundary

Within this window, the number in the row named “sum” is the total area of interest. This number was then copied over to the Excel sheet where it will be compared to the total area of the high school attendance boundaries. This step was done for each high school attendance boundary.

Finally to get the percentage by area of people who reported a household income less than $42,000 per high school attendance boundary, in Excel, the total block group area was divided by the total area of the high school attendance boundary then was multiplied by 100 to get a percentage. The resulting percentage and the areas used are shown in the chart in figure 11.

High School	Attendance Boundary Area (square feet)	Area of Block Groups Under $42,000 Within Attendance Boundary (square feet)	Percentage by Area that Reported less than $42,000 Within Specific High School Attendance Boundary
Bryan Adams	437960505.1	121599955.1	27.77%
Adamson	90423364.4	62785978.25	69.44%
Carter	430774234.2	271132243.5	62.94%
Conrad	113332173.2	66198780.87	58.41%
Hillcrest	443465105.1	62598219.75	14.12%
Jefferson	538812247.3	306342251.7	56.86%
Kimball	451291928.4	286714408.9	63.53%
Lincoln	145633667.4	145598203.8	99.98%
Madison	230989888.4	184506477.7	79.88%
Molina	163168554.7	113629246	69.64%
North Dallas	624385430.8	340636090.8	54.56%
Pinkston	421286165.9	379725694.1	90.13%
Roosevelt	280766560.9	272315486.5	96.99%
Samuell	337380858.5	246868093	73.17%
Seagoville	1599998707.0	712898213.3	44.56%
Skyline	310266014.5	198316340.1	63.92%
South Oak Cliff	454076743.5	353447991.4	77.84%
Spruce	535075235.3	420803460.8	78.64%
Sunset	136427587.5	61769909.09	45.28%
WT White	656847267.2	82819430.4	12.61%
Wilmer-Hutchins	1918294031.0	1610936000	83.98%
Wilson	349262249.4	76791012.91	21.99%

Figure 11. The area of each high school attendance zone and the corresponding total area within each zone that had a household income less than $42,000. Then there is the percentage by area resulting from the comparison between the two areas.

F. Data Presentation

In order to see if there is a correlation between household income and the percentage of remote learners in Dallas Independent School District, the percent by area within each attendance zone that reported an income of less than $42,000 was compared to the percentage of remote learning provided by the survey published by *The Dallas Morning News*. One way to compare this data is represented in the Excel chart in figure 12.

High School	Precentage of Remote Learners per High School	Percentage by Area that Reported less than $42,000 Within Specific High School Attendance Boundary
Bryan Adams	56.55%	27.77%
Adamson	56.13%	69.44%
Carter	41.89%	62.94%
Conrad	46.45%	58.41%
Hillcrest	48.86%	14.12%
Jefferson	53.46%	56.86%
Kimball	62.11%	63.53%
Lincoln	44.00%	99.98%
Madison	47.73%	79.88%
Molina	61.12%	69.64%
North Dallas	43.22%	54.56%
Pinkston	52.74%	90.13%
Roosevelt	54.14%	96.99%
Samuell	50.73%	73.17%
Seagoville	50.61%	44.56%
Skyline	45.75%	63.92%
South Oak Cliff	39.13%	77.84%
Spruce	47.09%	78.64%
Sunset	53.92%	45.28%
WT White	52.10%	12.61%
Wilmer-Hutchins	52.71%	83.98%
Wilson	37.55%	21.99%

Figure 12. The percentage of remote learning per high school compared to the percentage by area that reported a household income of less than $42,00 per high school attendance boundary

In order to create a visual way to compare the percentage of remote learning to household income, two separate data frames were created. One had the block group data visible and the other one had the high school attendance boundaries visible. They were then each symbolized in a way that correlated high percentage of remote learning to lower household income. For the percentage of remote learning, the high school attendance boundary layer (DISD_High) property window was opened. Then under the symbology tab, “quantities” was chosen as the way to represent this data. Then under the “fields” area, “Remo_Learn” was selected for the value. Then under classification, “Natural Breaks (Jenks)” was used and then for this map representation, 5 classes were decided the best way to break up the percentages.

Then the higher percentage of remote learning correlated to a dark blue color and the lower percentage of remote learning was represented with a darker green. This procedure is shown in Figure 13.

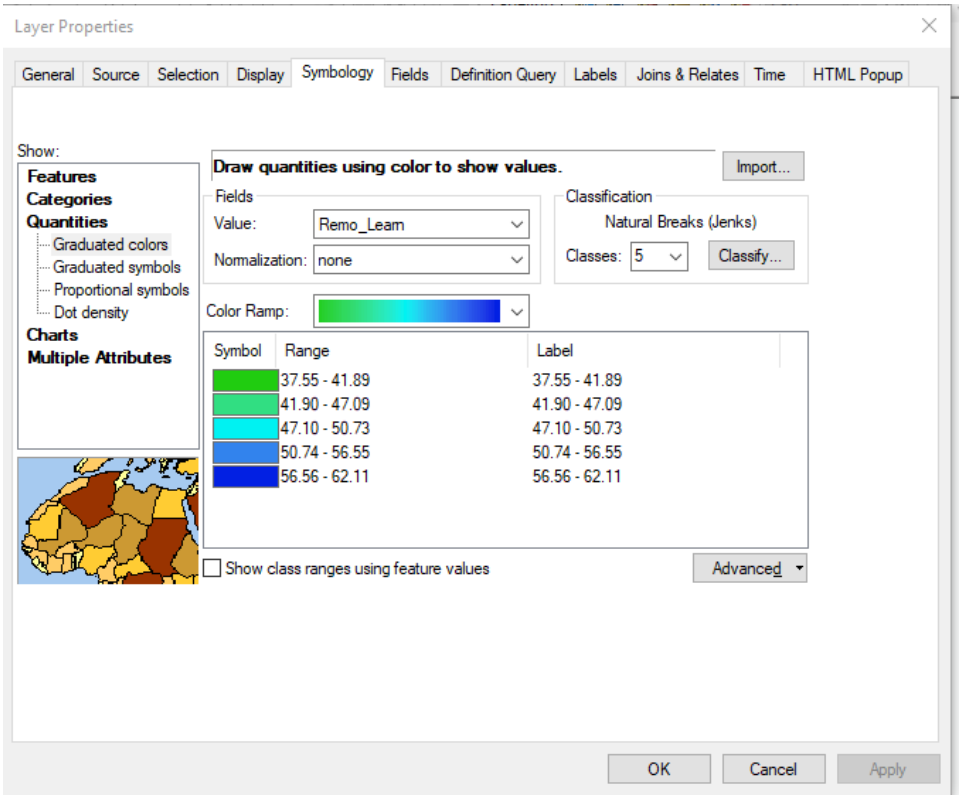

Figure 13. Choosing the symbology in order to represent the percentage of remote learning per high school attendance boundary.

The exact same steps were taken to represent household income per block group. In the other data frame, the layer that just contains the block group census data clipped to the outer boundary of the Dallas Independent School District was used. Then the same process as above was used but this time with the field that contains that household income selected as the value in the "Field" area and 8 classes were used instead of 5. Then the 8 classes were symbolized so that the lower income was represented by darker blues and high income was represented by darker greens.

This symbology was chosen in order to see if in areas that had darker blues for the percentage of remote learning also had darker blues for household income. If so, this would mean that areas that have higher percentages of remote learning also have block groups that reported a household income of less than $42,000. The resulting comparison map is shown in figure 14.

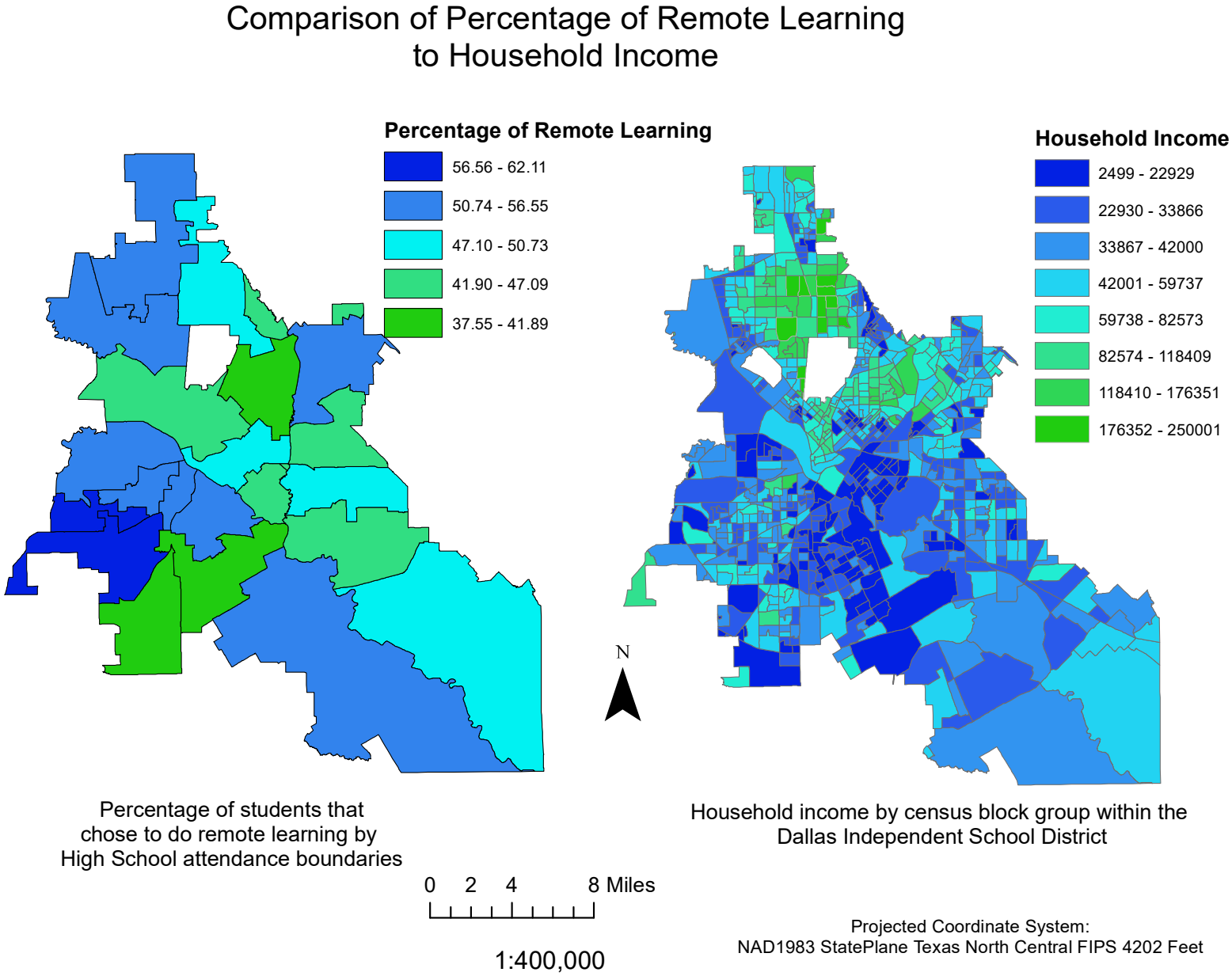

Figure 14. Map comparing the percentage of remote learning by high school attendance zone to the household income reported by each census block group within the boundary of Dallas Independent School District.

G. Conclusion

Comparing the percentage of remote learning per high school to the percentage by area of census block groups that reported a household income lower than $42,000 was used to see if there was a correlation to high percentages of remote learning and lower household income. A comparison map was also used to compare the percentage of remote learning per high school attendance boundaries to the reported household income per census block group in order to see if there was a correlation. The analysis indicated there was not a strong relationship between high percentages of remote learning and lower reported household incomes.

Several factors may have contributed to these results. One factor may have to do with the completion percentages for the survey sent out to DISD parents. For many of the high schools used in this project, the completion percentage was very low while others had very high completion percentages. Through this analysis however, some of the high schools that had the lowest completion percentage had a higher percentage by area of people that reported a household income of less than $42,000. There could be a correlation to income and access to a way to complete this survey if it was an online survey. This could be an interesting relationship to evaluate in another project to see if there is a relationship between income and access to technology and if that contributes to the percentage of remote learning. If there was 100% completion for each high school used in this analysis, there could have been a more obvious correlation to the percentage of remote learning to lower household incomes.

Another factor that could have contributed to this lack of relationship is the use of 2010 census data. The use of block group data, even though it is divided up into relatively small areas, it did not take into account the number of people in each block group. Also, the reported household incomes per block group were only estimates. If more recent data was used, the number of people in each block group that had incomes below that lower income range was considered, and the mean income in each block group was used, it is possible that there could have been a more obvious correlation between remote learning and income.

It is possible that there is no correlation between remote learning and income at all. Parents and students could have different reasons for choosing remote learning over going back to campus. During this time of continuous spread of the COVID-19 virus, people are

continuing to take precaution to try and stop the spread as well as to keep themselves safe from getting sick. This could be an even bigger reason for why people chose remote learning.

H. Sources

1. https://money.usnews.com/money/personal-finance/family-finance/articles/where-do-i-fall-in-the-american-economic-class-system#:~:text=A%20family%20earning%20between%20%2430%2C000,more%20than%20%24350%2C000%20are%20rich
2. https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-data.2010.html
3. https://www.census.gov/programs-surveys/geography/technical-documentation/records-layout/tiger-line-demo-record-layouts.html
4. https://www.dallasnews.com/news/education/2020/09/23/see-how-many-dallas-isd-parents-chose-remote-or-in-person-learning-for-their-kids-school-by-school/
5. https://www.dallasisd.org/Page/29132
6. https://pro.arcgis.com/en/pro-app/tool-reference/analysis/intersect.htm
7. http://www.geo.utexas.edu/courses/371c/project/2019F/Bushlow_Megan_Semester_Project.pdf
    a. Inspiration for format of project