How well is minority’s life in NC

First of all, we did some test on the authenticity of the dataset. Compared to the real poverty percentage of NC in 2014 which is 17.2%, the computed poverty rate(16.8%) is pretty close, proving the truth value of the dataset.

Surprised by the great poverty rate of North Carolina, our group decided to start our project with the focus on Socioeconomic Status. We plotted the percent of people live in poverty using QGIS. According to the graph, the suburban of the 4 most populous metropolitan statistical areas in North Carolina, Charlotte, Raleigh, and Greensboro, and Fayetteville, tend to have lower poverty rate compared to the rest of the state. This shows that the suburban area of a large city tends to have lower poverty rate compared to urban and rural areas.

Furthermore, as international students from China, we are particularly interested in the wellbeing of minority groups in the US, such as Asians. Thus, we decide to set our focus in the relationship between the set Socioeconomic Status and Minority Status&Language. Because of this, we also plotted the percent of people who speak English less than well, From the 2 graphs, we can see that there is a relatively strong relationship geographically between the Percent of people who speak English less than well and the Percent of people who are minority.

Another reason for our choice of theme 1 and theme 3 is that # According to the correlation plot, we found an interesting point that

# there exists a very high correlation between the Sum of series for Theme1 and the

# overall SVI for the four themes.

# Based on this finding and our primary interest in the theme of Minority, we made some

# analysis on the relationship between the two variables in Theme 3 and their correlation with

# the four variables in Theme 1.

Apart We tried many different combinations of models to find out some interesting relationships that are hidden behind the huge amount of data. For example, we tried to look at the correlation between the percentage of people who speak English less than well and the percentage of unemployed within one area. Most of the correlation test is significant to reject Ho. Particularly, we find a fairly strong positive relationship between the percentage of people who are under poverty and the percentage of people without high school diploma. However, in many cases the r-squared is very small, indicating that even though there is an existing relationship between the two variables that we want to analyze, the correlation is too small to give us any interesting information.

Moreover, there are also several cases in which we do not expect any relationship between the two variables but surprisingly find a fairly significant r-squared. Because of the problem of small r-squared, our group starts to use full variables of theme 1 or even all variables of the dataset to get the best model through model-selection method. In this way, we believe we can get some more promising correlated factors between the minority theme and the socioeconomic theme.

After using the all subset model-selection method, we find it surprising that the best single-variable prediction model for EP\_LIMENG is EP\_CROWD, which stands for Percentage of occupied housing units with more people than rooms estimate. We are interested in this unexpected but significant correlation between the two variables. We tried to explain this phenomenon : people who do speak English less than well probably do not live alone, they definitely have someone who is proficient in English with them so that they don’t find it a problem to live in US where English is in everyone’s daily life.

Seeing the negative impact of speaking English less than well, one question pops to our mind: who are those people? What are they like? As shown in the statistic, there is a strong positive relationship between the minority group and the percentage of people who speak English less than well. Taking that into circumstance, we can imagine that the minority group, especially those who just immigrated to the U.S. have a really hard time learning English, adjusting to American lifestyle, and getting their life onto the right track. From this idea, we further analyzed the relationship between minority groups and all those variables in the group of socioeconomics. For the analysis of the minority variable, According to the linear model regression and anova analysis,

# poverty and unemployment have significant correlation with the percentage minority.

# Comparatively, the personal income and percentage of people who own high school diploma

# are not that significant in the explanation of minority.

And above are all our analysis for the group project. Thanks for listening.