

Volunteer Nurses in North Carolina during COVID-19

Feiya Suo
Client: Jill Forcina

Date: April 30th, 2022

Executive Summary

Health care workers were in massive demand during COVID-19, and a considerable number of volunteers participated to help. To collect information about available volunteer nurses in each county of North Carolina, an online questionnaire survey about the willingness of volunteer nursing during the pandemic was designed. Using quantitative data from the survey with 3,089 registered nurses in NC, we describe the geographic distribution of available and volunteered nurses across the state. We also use additional data on vaccination to analyze the correlation between volunteer nurses and the COVID-19 vaccination coverage rate at the county level. We found that urban states have more nurses willing to volunteer and who actually volunteered. The percentages of available and volunteered nurses over the county population are also high in those urban counties, but it is also the highest in some rural counties with small populations.

The correlation analysis indicates that the numbers of available and volunteered nurses are significantly associated with the vaccination rate. The proportion of volunteered nurses over the number of available nurses does not significantly correlate with the vaccination coverage rate. After aggregating the data at the region level, we found the same trends.

Introduction

During the COVID-19 pandemic, there was a shortage of health care workers because of the rapidly increasing demand. Considering the negative effect of the worker shortage, it is essential to determine the problems in supply, distribution, and retention of health care professionals. The North Carolina Area Health Education Centers (AHEC)

aims at providing and supporting educational activities and services relevant to health care inside the state, mainly focusing on primary care in rural communities and those with less access to resources to recruit, train, and retain the workforce.

AHEC worked with the NC Board of Nursing to design and distribute a questionnaire about nurses' willingness to volunteer work during the pandemic. By using the survey data, we better understand how many volunteer nurses are available in each county and how many of them actually volunteered. Combined with county-level data on vaccination, we study the variation in available nurses and its correlation with vaccination coverage rates.

Data

AHEC designed two rounds of questionnaire surveys to ask about the volunteer information in North Carolina during the COVID-19 pandemic. In the first survey, respondents were invited to participate in the voluntary survey when accessing the NC Board of Nursing website. If they chose the option that they are willing to be a volunteer nurse, they would be redirected to the survey, asking for their primary county, other counties of volunteer willingness, nurse license type, and employment status. Their personal contact information was recorded, and the county government or local health care providers may contact them for potential volunteer work. There were 3,089 responses in the first round.

In the second survey, the survey asked if the respondent volunteered since the first survey. If the answer is yes, the county where they volunteered was recorded. There were

562 responses in the second round. The two surveys are separate and cannot be connected to each other.

To conduct the analysis, we also collected information on the county-level population, the number of fully vaccinated people, and the vaccination coverage rates. We first transferred the survey data from the individual level to the county level and then merged the data to generate the final dataset for future analysis. The final data has 100 rows, representing the 100 counties inside North Carolina. There are many variables of interest in this project. Table 1 summarizes the descriptive statistics of the major variables.

Table 1: Descriptive Statistics of the Major Variables

Statistic	N	Mean	St. Dev.	Min	Max
Available Volunteers	100	92.36	121.85	9	853
% of Available Volunteers over Population	100	0.00121	0.00076	0.00025	0.00412
Volunteered Nurses	100	1.58	3.26	1.00	19
% of Volunteered Nurses over Available Volunteers	100	0.0141	0.0239	0	0.1389
Vaccinated Population	100	59,334	110421.7	1,926	784,394
Vaccination Rate	100	0.5326	0.0797	0.3987	0.7869
% of NP	100	0.1105	0.0455	0.0278	0.2500
% of RN	100	0.8555	0.2095	0.6740	0.9229
% of LPN	100	0.1210	0.0716	0.0328	0.3636
% of Other License	100	0.1201	0.0647	0.0212	0.3077
% of Retired Nurses	100	0.2500	0.0489	0.0280	0.0939
Ratio of Active Nurses	100	0.9615	0.0184	0.8936	0.9915

Note: NP, RN, LPN, and Other License all refer to the types of nursing license.

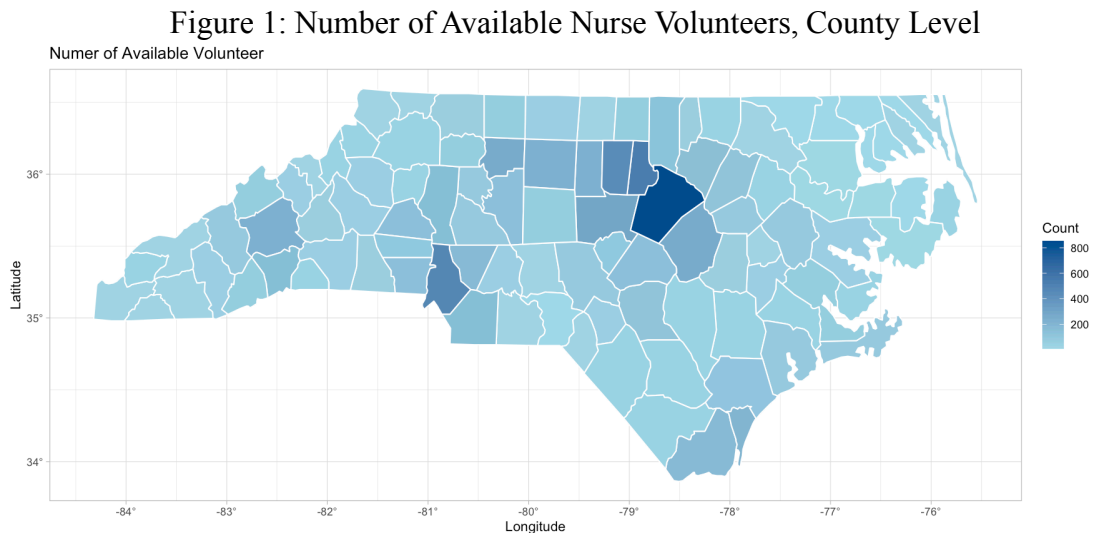
Analysis

1. Data visualization

We first use the Shiny app in R to visualize the survey data at the county level. We create an interactive map application that helps the client get the visualization easily. They can choose the metric of interest from the interactive menu and obtain the detailed county-level data by clicking the county on the map.

We will first introduce the number and percentage of available volunteers in each county. Figure 1 indicates the number of available nurses, and Figure 2 shows the percentage based on the county population. Darker blue represents higher counts/percentages of available nurses during the pandemic in the two figures.

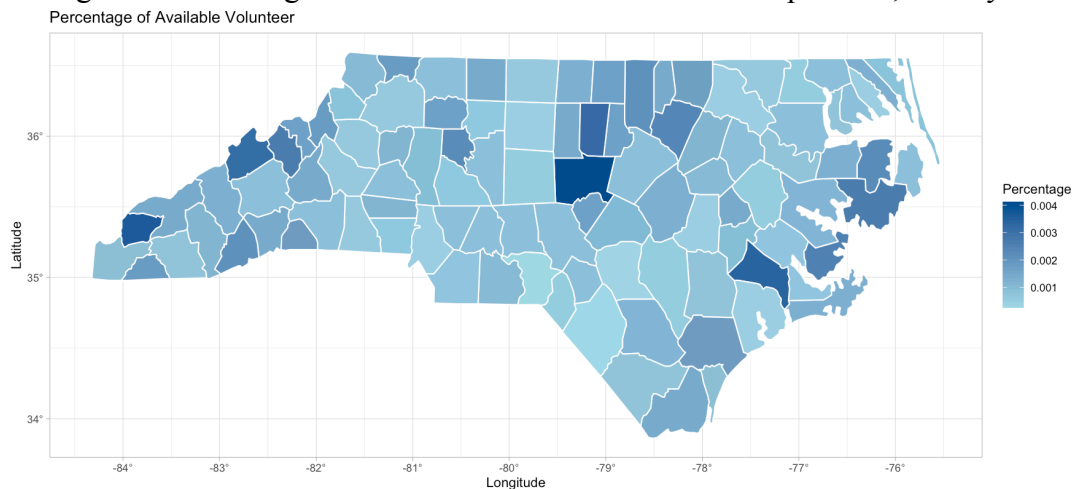
From Figure 1, the number of nurses willing to volunteer is the highest in Wake County. The nearby counties, Durham and Orange, also have a large number of available nurses. The other high counts are from more urban¹ and populous counties, including Mecklenburg (Charlotte) and Buncombe (Asheville). It is consistent with our hypothesis that more urban counties have more available nurses. However, urban counties also have larger populations; it is also important to understand the percentage of available nurses based on county population. If urban counties have a smaller percentage of nurses, it can potentially lead to negative impacts to health outcomes.



¹ The definition of rural/urban county comes from the NC Rural Center (www.ncruralcenter.org). There are 78 rural counties in NC, with an average population density of 250 people per square mile or less. Regional cities and suburban counties refer to those that have 250-750 people per square mile; there are 16 in NC. There are six urban counties in NC, and they have more than 750 people per square mile on average.

We visualize the percentage of available nurses in Figure 2. The counties with a higher percentage of volunteer nurses based on population are different. Although some urban counties, such as Durham and Orange, still have a high percentage of available volunteers, some rural counties also have a higher percentage. For example, Chatham, Graham, Madison, Yancey, and Jones have more available nurses based on county population. They all rank relatively low in GDP and have smaller population sizes. Although the absolute number of available nurses is low, it is still comparable to urban counties when adjusted by population.

Figure 2: Percentage of Available Nurse Volunteers on Population, County Level



We then visualize the number of nurses who actually volunteered in each county. Figure 3 and Figure 4 show the counts and percentages, respectively. In Figure 3, we have similar findings as Figure 1: urban counties, regional cities, and suburban counties, such as Wake, Durham, Orange, and Buncombe, have more nurses who volunteered during COVID-19. AHEC and the NC Board of Nursing collected the available volunteers' information from the first survey and shared it with the local governments.

Those counties might use the information more effectively, and nurses in those counties may be more willing to participate in relevant health care services.

In Figure 4, we can still find that the urban counties mentioned above (Wake, Durham, Orange) and some regional cities and suburban counties (Buncombe, Cabarrus, and Davidson) have higher percentages of volunteered nurses, even though their population sizes are big. Meanwhile, Polk County, which is a rural county, has the highest percentage of volunteered nurses based on the county population. Similarly, rural counties with smaller populations, including Scotland, Wilkes, Moore, and Watauga, also have a large percentage of volunteered nurses.

Figure 3: Number of Volunteered Nurse, County Level

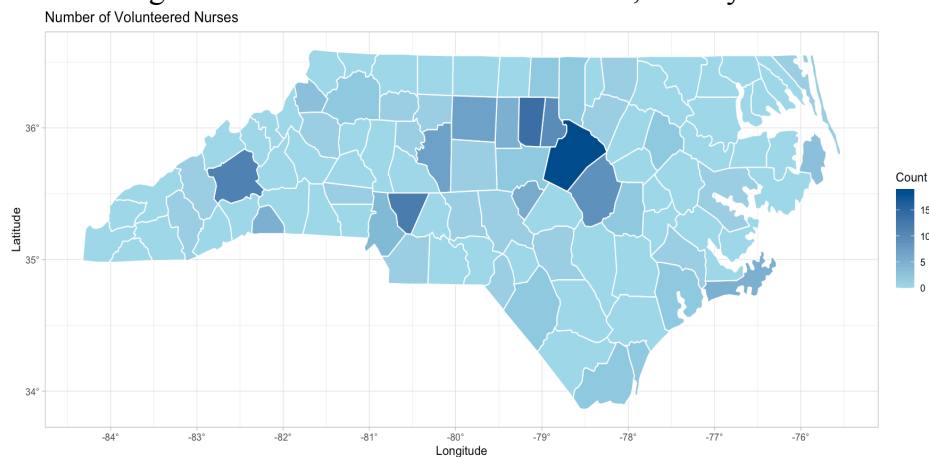
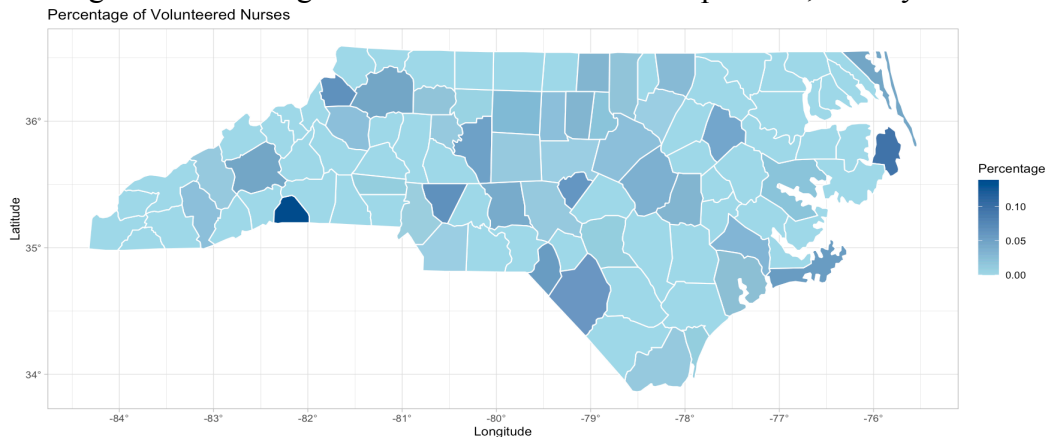


Figure 4: Percentage of Volunteered Nurses on Population, County Level



We are also interested in the correlation between volunteer nurses and county-level vaccination coverage. We first show the number of vaccinated population and the coverage rates in Figures 5 and 6. Again, darker blue represents higher counts/rates. In Figure 5, urban counties like Wake, Mecklenburg, Guilford, Forsyth, Durham, and Buncombe have the largest vaccinated population. It is consistent with the findings that urban counties have more available nurse volunteers. One potential explanation is that urban counties have more health care resources, and it is highly correlated with the vaccination rates and the number of available volunteer nurses.

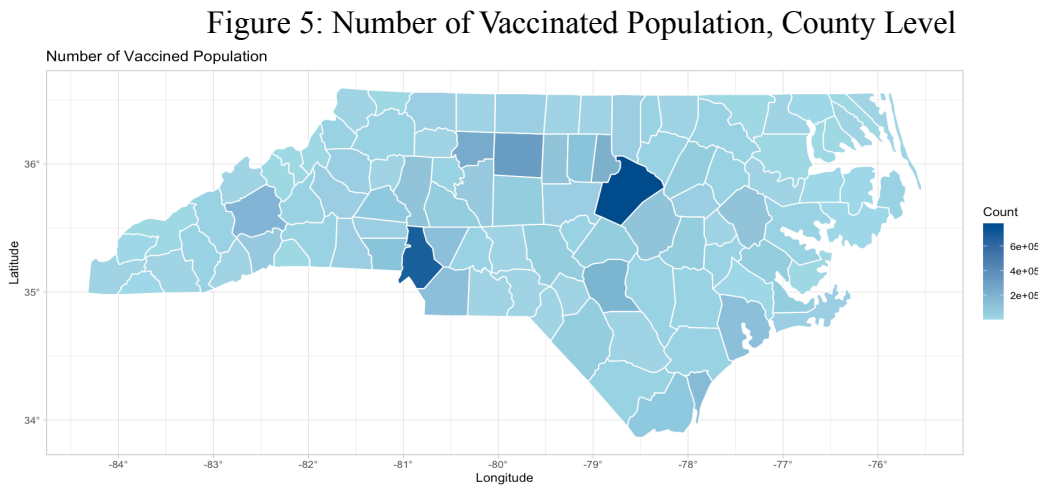
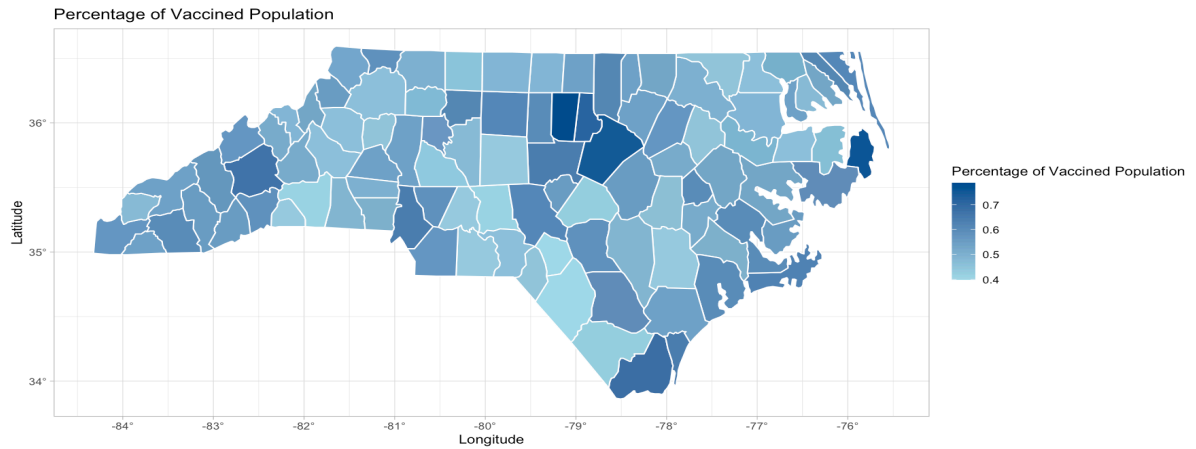


Figure 6 shows that most urban counties mentioned above have high vaccination coverage rates. However, the vaccination rates in rural counties are also high. Similar to the findings on volunteer percentage, the smaller population sizes in rural counties can be one potential explanation.

Figure 6: Vaccination Coverage Rates, County Level



We also use the same visualization method to show the percentage of available nurses based on their license types (NP, RN, LPN, NAIL, and other), license status (active or inactive), and retirement status (retired or not). These figures can be found in the appendix.

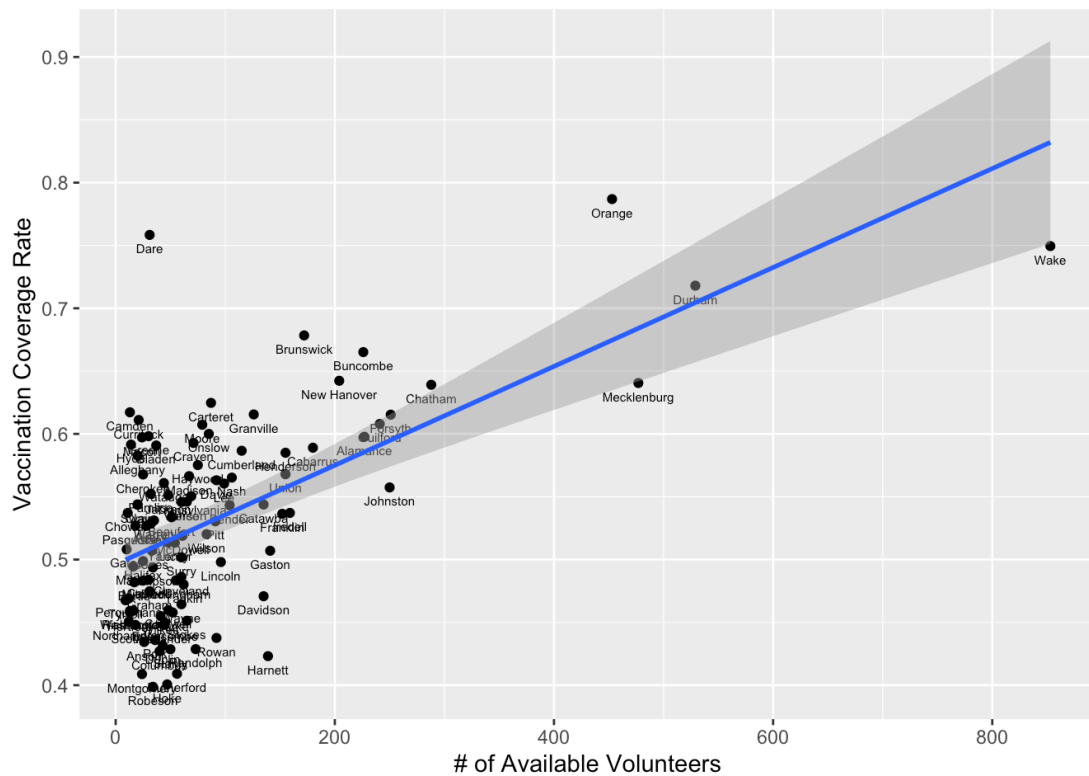
2. Correlation Analysis

We are interested in the correlations between volunteer nurses and COVID-19 vaccination rates. The hypothesis is that in counties with (1) more available volunteer nurses; (2) more nurses who actually volunteered; and (3) a higher proportion of nurses who volunteered over the available nurses, the COVID-19 vaccination coverage rate would also be higher.

Figure 7 shows the scatter plot between the vaccination coverage rate and the number of available volunteers. Each point represents a county. Many points are on the left side of the figure, indicating that many counties have zero or very low numbers of available volunteers. The counties on the right are the urban counties with more available

volunteers. The vertical axis shows the vaccination rates; the points on the top represent the counties with higher vaccination coverage rates. The blue line represents the estimated regression line, showing a positive correlation between the two variables. The grey area indicates the 95% confidence interval of the regression line, which shows the uncertainty of the estimation.

Figure 7: Scatter Plot between Vaccination Coverage Rate & the Number of Available Volunteers, County Level



We then use hypothesis testing to examine the correlation further. We use the vaccination coverage rate as the outcome variable and the number of available volunteers as the explanatory variable in a linear regression model. The coefficient is $3.937\text{e-}04$, and it is statistically significant at the 5% significance level ($p\text{-value} = 3.56\text{e-}11$). We have strong evidence to reject the null hypothesis that the correlation between the two

variables is zero. It indicates a significantly positive correlation between the two variables.

From the regression model, we can also use the R-squared value to explain to what extent the variance of one variable explains the variance of the second variable. R-squared value is a statistical measure representing the proportion of the variance for a dependent variable explained by an independent variable or variables in a regression model. It ranges from 0 to 1 and is commonly stated as percentages from 0% to 100%. An R-squared of 100% means that all movements of the dependent variable are completely explained by movements in the independent variable(s)². The R-squared value in this model is 0.3621, which means the explanatory variable can interpret 36.21% of the variations inside the outcome variable.

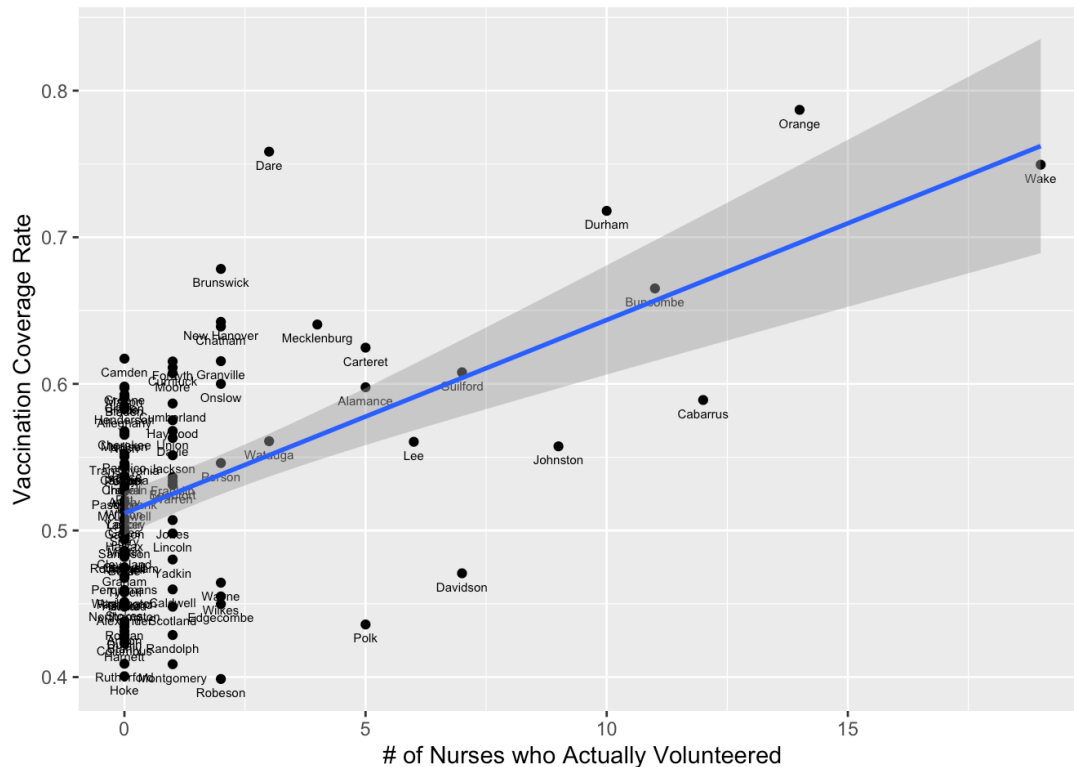
Similarly, Figure 8 shows the correlation between vaccination coverage rate and the number of nurses who actually volunteered. We can still observe that many urban counties are on the right of the plot, indicating a higher volunteer percentage. Many rural counties are also on the right, which is consistent with the findings from the map. The blue regression line shows a positive correlation between the two variables.

From the regression model, the coefficient is 0.0132, and the p-value is 6.94e-09. We have strong evidence that the two variables are significantly correlated. Thus, a higher number of nurses who volunteered is significantly associated with a higher vaccination rate. The R-squared value is 0.2911, which means the explanatory variable can interpret

² We usually want to have high R-squared values because it indicates the model explains more variation inside the dependent variable. However, R-squared doesn't always tell whether the model is good or bad, nor will it tell whether the data and predictions are biased. A high or low R-square isn't necessarily good or bad, as it doesn't convey the reliability of the model, nor whether you've chosen the right regression. You can get a low R-squared for a good model, or a high R-squared for a poorly fitted model, and vice versa. Generally, when you add more variables into the regression, the R-squared value increases even if those variables are not highly correlated with the outcome variable.

29.11% of the variations inside the outcome variable. It is lower than the R-squared value from the previous model, but it is still very high considering that we only have one explanatory variable in the regression model.

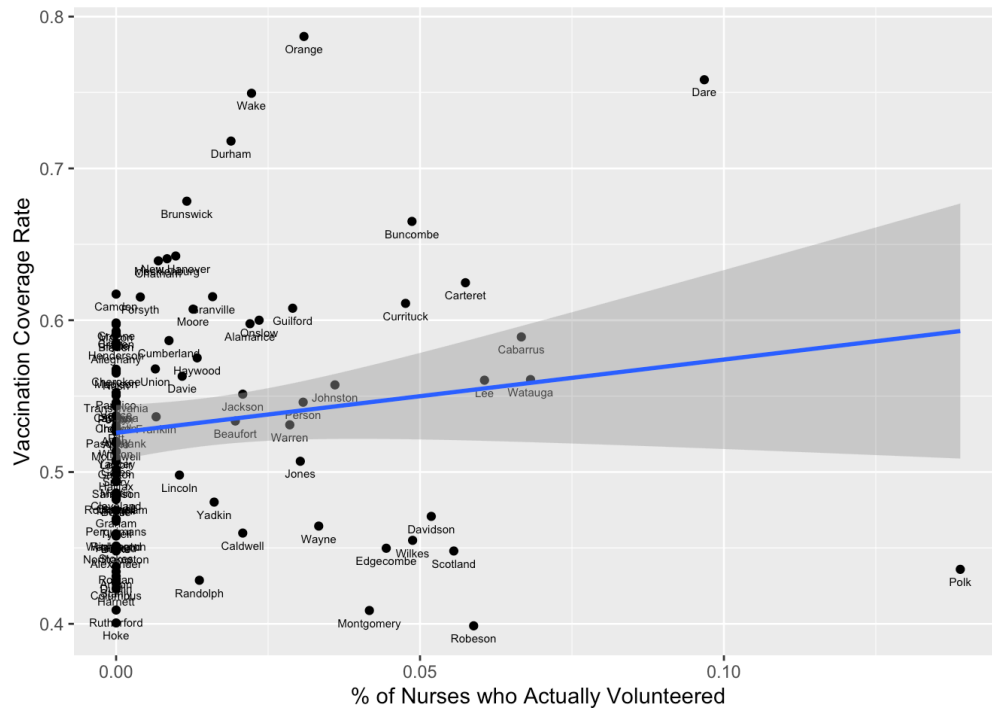
Figure 8: Scatter Plot between Vaccination Coverage Rate & the Number of Nurses who Actually Volunteered, County Level



For the two analyses above, we also use the percentage of available nurses and the percentage of volunteered nurses as the explanatory variable in the regression model. The scatter plot with the estimated regression line can be found in the appendix. The results are similar: vaccination coverage rate is positively associated with the two percentages based on the county population. The p-values of the coefficients are smaller than 0.05. Thus, we have strong evidence that the correlation is statistically significant.

Figure 9 is the scatter plot between the vaccination coverage rate and the proportion of nurses who volunteered over the total number of available volunteers. We can find a slightly positive regression line in this figure. Many counties have zeros (and also in Figure 8) because the many counties do not have nurses who actually volunteered; it might bias the regression results. Another interesting finding is that there is a more obvious positive correlation in urban counties on the top right corner of this plot (including Orange, Wake, Durham, and Brunswick).

Figure 9: Scatter Plot between Vaccination Coverage Rate & Proportion of Nurses who Actually Volunteered over the Number of Available Volunteers, County Level



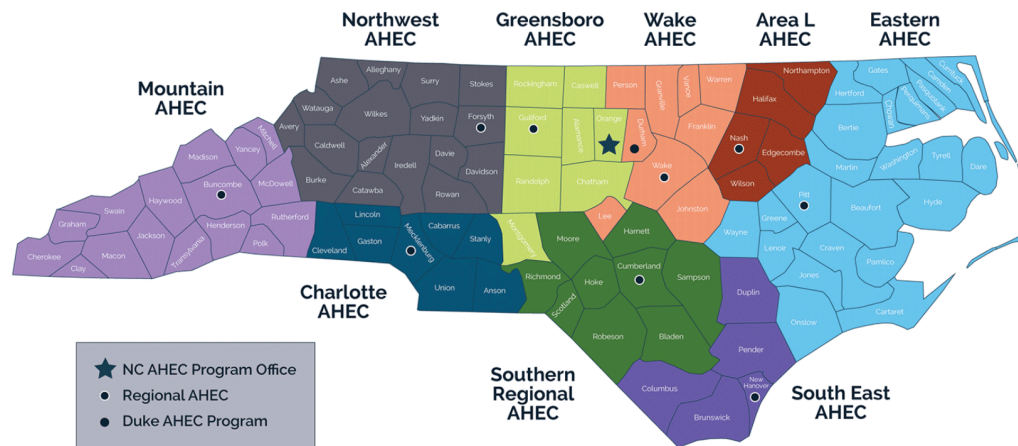
The regression model's coefficient is 0.4826, but the p-value is 0.15. We cannot reject the null hypothesis that the correlation between the two variables is zero. Thus, the vaccination rate and the volunteered nurse proportion are not significantly correlated. The R-squared value is also low, which is 0.021. It indicates that the explanatory variable can

only interpret 2.1% of the variations inside the outcome variable. It shows that there is no strong connection between the two variables.

As we can observe, there are many zeros in the three figures above. The number of volunteer nurses, especially those who actually volunteered, is low in many counties. It may influence the correlation analysis results. Thus, we use the NC AHEC regions to aggregate the data further and then analyze the correlations. The regions are defined by AHEC's regional offices and programs. Figure 10 shows the AHEC regions.

The 100 counties in NC are divided into nine AHEC regions. We first aggregate the raw data at the regional level and then calculate the metrics of interest, such as the proportions of volunteered nurses over available nurses and the vaccination coverage rates. We use the aggregated data to analyze the correlation by the same method.

Figure 10: AHEC Regions

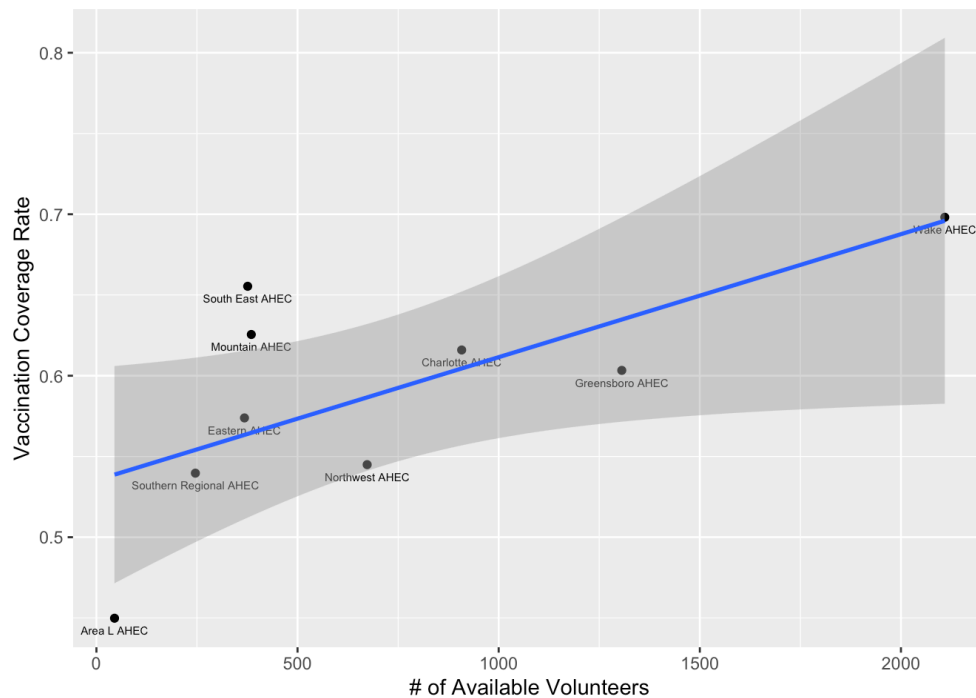


We first analyze the correlation between vaccination rates and the number of available volunteers. From Figure 11, we can still observe the positive correlation between the two variables, consistent with the county-level findings. Area L, South East, and Mountain are farther away from the regression line and its 95% confidence interval,

and they are on the left side of the scatter plot. They are all regions including more rural counties, and the number of available volunteers is low.

We still use the vaccination rate as the outcome variable in the hypothesis testing. When analyzing its correlation with the number of available volunteers, the correlation coefficient is 7.616×10^{-5} , with a p-value of 0.046. It shows that we can reject the null hypothesis that the correlation between the two variables is zero. The R-squared value is 0.4559, which indicates that the explanatory variable can interpret 45.59% of the variations inside the outcome variable.

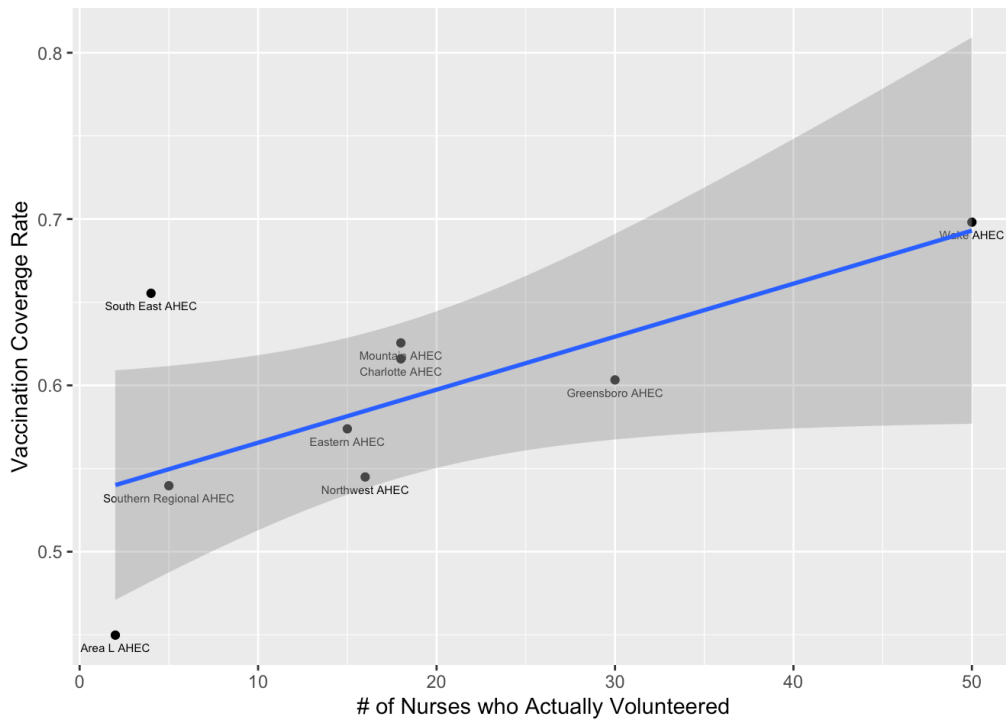
Figure 11: Scatter Plot between Vaccination Coverage Rate & the Number of Available Volunteers, Region Level



In Figure 12, there is a positive correlation between the vaccination rate and the number of volunteered nurses. The 95% confidence interval of the regression line is

wider compared to the county-level analysis. It is because there are fewer data points inside the aggregated data, and the uncertainty increases accordingly. The regression model's coefficient is 0.0032, but the p-value is 0.0552. Thus, we do not have strong evidence to reject the null hypothesis of zero correlation at the 5% significance level. However, the p-value is very close to the 0.05 threshold, so we are very close to having the conclusion that the two variables are correlated. It still shows evidence to reject the null hypothesis, but it is not as strong as the two county-level models above. The value of R-squared is still high, which is 0.4299. It shows that the number of nurses who volunteered can explain about 43% of the variation in vaccination rates in this specific model.

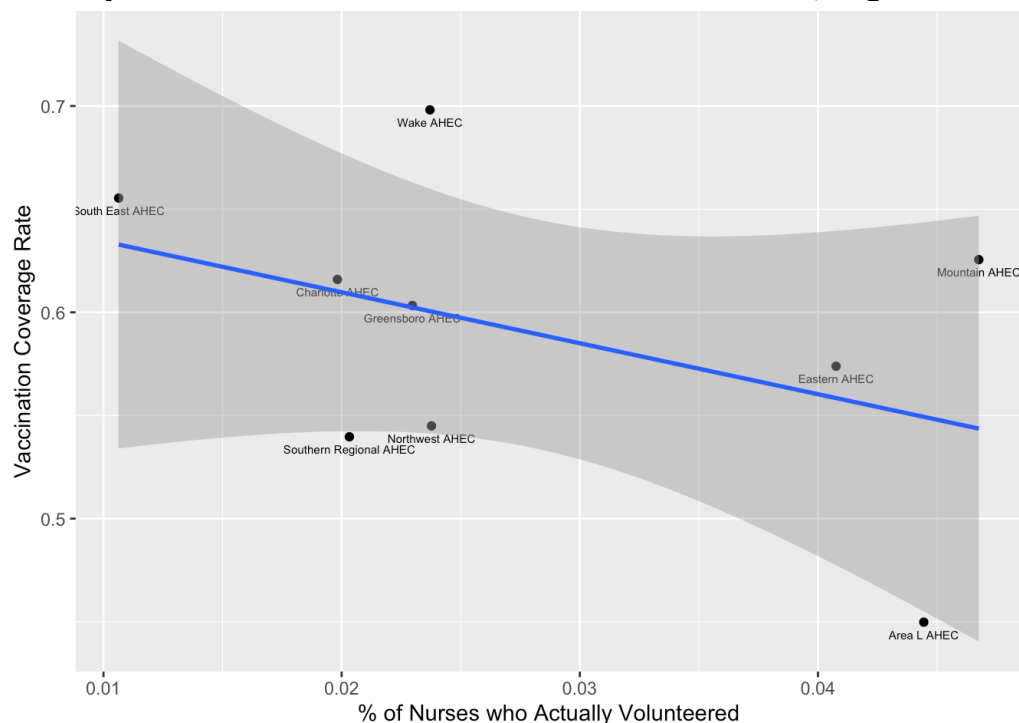
Figure 12: Scatter Plot between Vaccination Coverage Rate & the Number of Nurses who Actually Volunteered, Region Level



We also use the percentage of available nurses and volunteered nurses over the county population as the dependent variable. The scatter plots can be found in the appendix. They show similar trends that the vaccination rate is positively correlated with the percentage of available nurses and volunteered nurses. However, the p-values are no longer smaller than 0.05. We do not have evidence to reject the null hypothesis of zero correlation. By using the region-level data, there are fewer data points, so the conclusion is not as conclusive as the county-level analyses.

Figure 13 shows a different trend. The correlation between the vaccination coverage rate and the proportion of the volunteered nurses is negative. This is different from Figures 11 and 12 at the region level, and it is also inconsistent with the regression line at the county level in Figure 9.

Figure 13: Scatter Plot between Vaccination Coverage Rate & Proportion of Nurses who Actually Volunteered over the Number of Available Volunteers, Region Level



It seems that a lower proportion of volunteered nurses over available nurses is associated with a lower vaccination rate. However, the regression coefficient is not statistically significant. The estimated coefficient is -2.470, and the p-value is 0.251. Thus, we cannot reject the null hypothesis of zero correlation. The R-squared is also lower (0.1826) than the two models above.

Conclusion

In this project, we first visualize the data to explore the county-level difference in volunteer nurses. From the county maps, we find that urban counties, regional cities and suburban counties have more available nurses and nurses who actually volunteered. However, urban counties also have more population. The percentage of nurses over the county population is the highest in rural counties and some urban counties. The number of vaccinated people is also higher in urban counties than in rural counties, but the vaccination coverage rate in NC is high in both urban and rural counties.

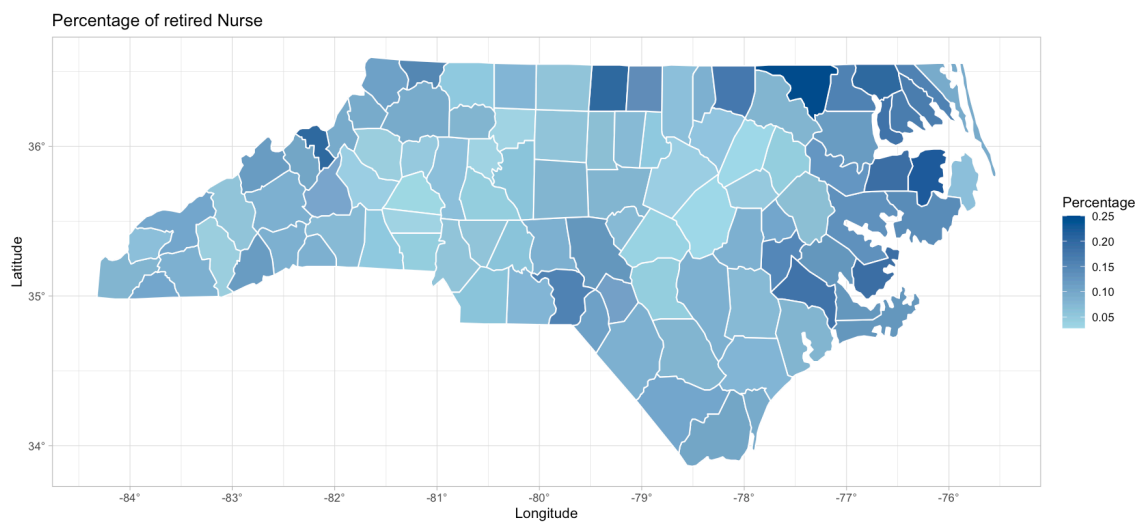
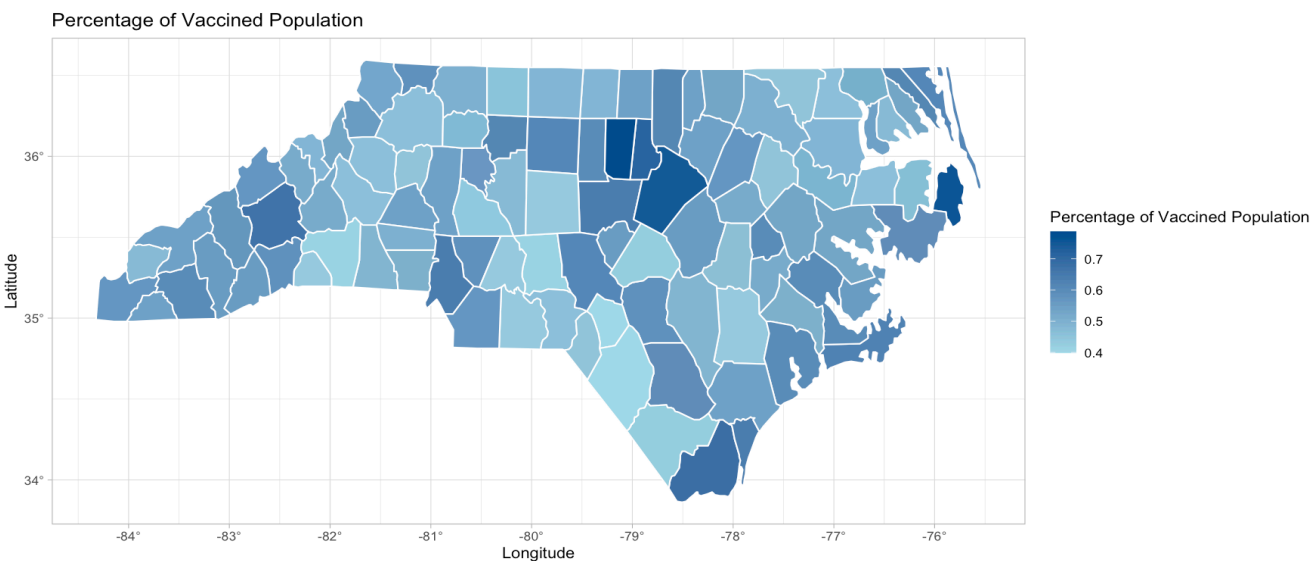
We also analyzed the correlation between volunteer nurses and vaccination rates. We hypothesized that a higher number or proportion of volunteer nurses is associated with a higher vaccination rate. We find a statistically significant correlation when using the number of available and volunteered nurses as the explanatory variables. We have strong evidence to reject the null hypothesis of zero correlation. However, the correlation coefficient is not statistically significant when using the proportion of volunteered nurses over the available nurses as the independent variable. We find the same trends when aggregating the data at the AHEC region level.

In counties with more available volunteer nurses, the healthcare system might be more developed and have more general healthcare workers. It is also associated with higher vaccination rates. However, compared to the proportion of volunteered nurses, the absolute number of nurses is the most important if we want to improve the vaccination coverage rate. Thus, it is essential for AHEC to participate in educational programs that help rural counties to train and recruit more health care workers.

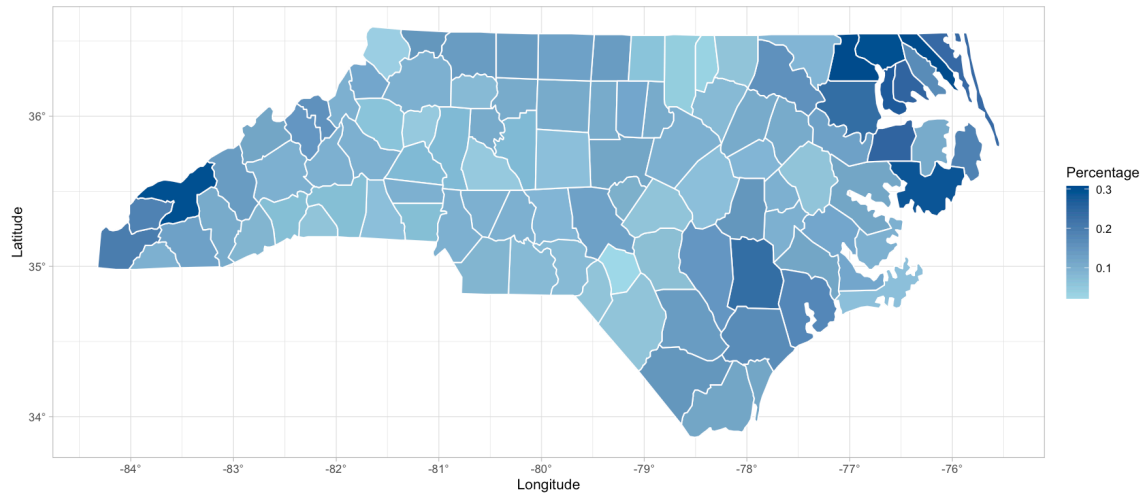
In addition, since we already collected the contact information of available volunteer nurses in the first survey, local governments also need to use the information more effectively. Some survey respondents mentioned that they were never contacted after providing their personal information. If local governments and health providers can make better use of the survey, it can help to facilitate the health care system and relevant programs to recruit more volunteer nurses.

Appendix

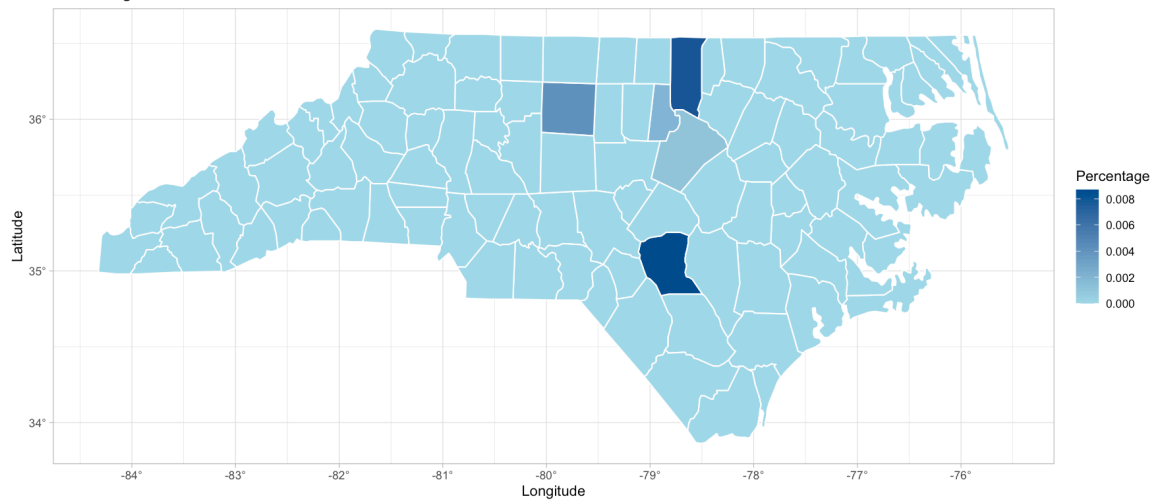
1. Map visualization



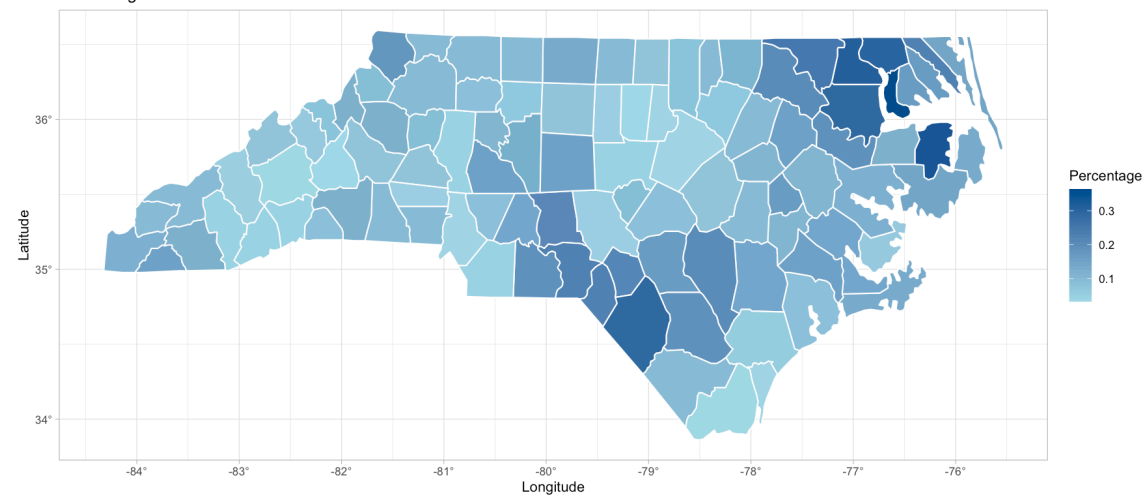
Percentage of Other License

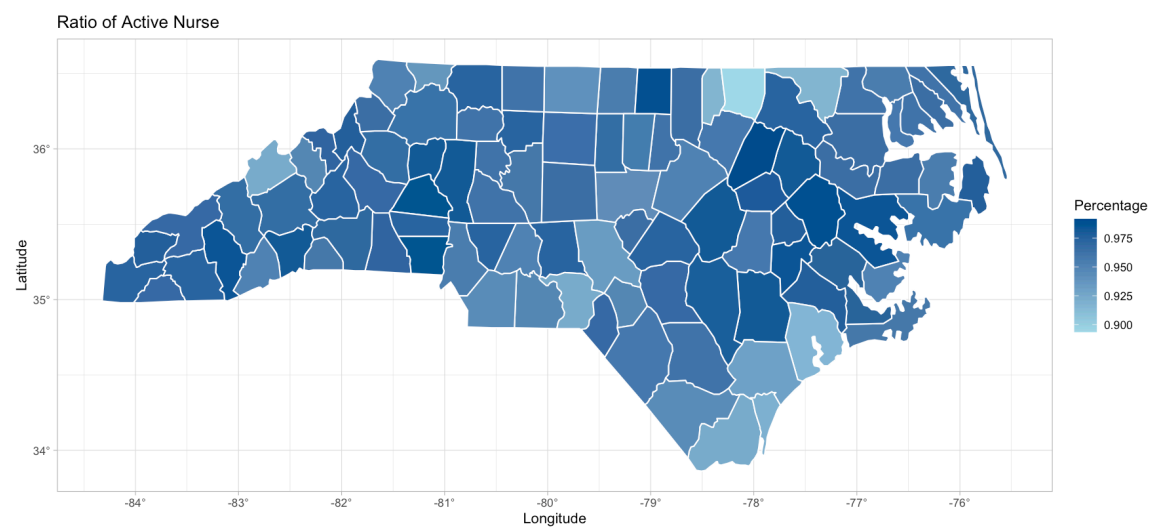
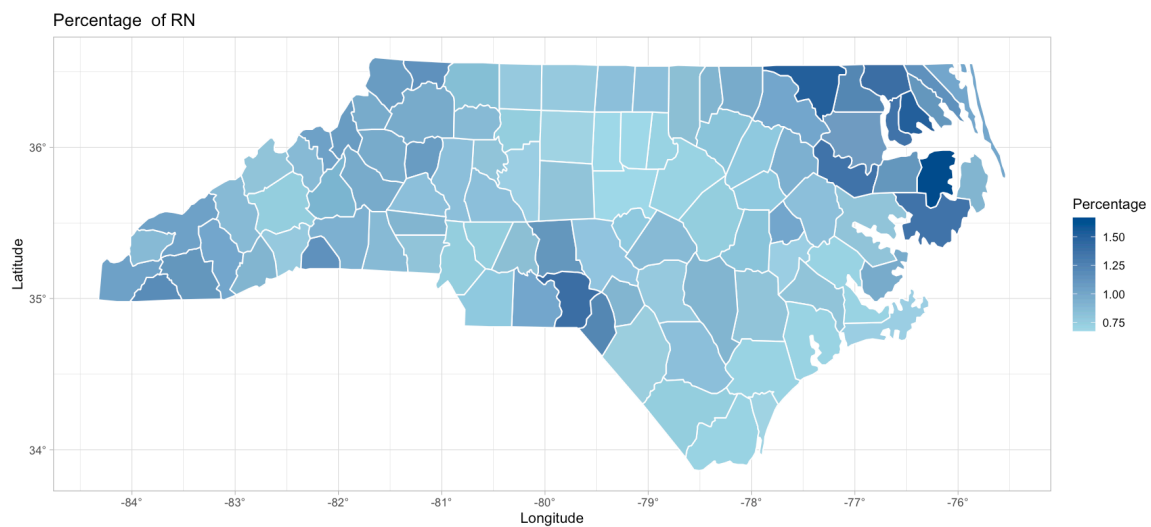


Percentage of NAI



Percentage of LPN





2. Correlation analysis using percentages

