Final Report

due November 16, 2021 by 11:59 PM

Lindsey Weyant, Ali Raich, Aden Clemente

11/16/21

Research Question

We are choosing to study a data set about Measles Vaccination rates in schools across the country. This data set pulls from 46,412 schools across 32 states in the years of 2017-2019. The data comes from a Wall Street Journal article published in October 2019 called "What's the Measles Vaccination Rate at Your Child's School?". The article discusses how increasing rates of unvaccinated people caused a high number of measles cases in the beginning of 2019. The Wall Street Journal compiled the data by reaching out to state health departments for kindergarten rates for individual schools across the country. It is important to note that there is not a universal method for collecting and keeping track of immunization rates so each state's data set is slightly different. The World Health Organization recommends a 95% vaccination rate among elementary schools. Our overarching research question is: How do vaccination rates vary across the country and different types of schools?

Data Wrangling

Our data had significant inconsistencies across different states and school types, which required that we consolidate certain variables. Also, we needed to create several categorical variables corresponding to continuous ones in order to be able to conduct logistic regression. The major changes to the dataset are outlined below.

For most schools, a value was only provided for either "overall" (overall vaccination rate) or "mmr" (measles, mumps, and rubella vaccination rate). Choosing to conduct our analyses on one of these variables would entail losing a massive number of observations. So, we created a new variable, "realrate," which took on the value of "overall" if present and the value of "mmr" otherwise. This way, we are able to retain most observations from the data set. However, this may have unfairly increased the vaccination rates of types of schools or states which favored reporting mmr rates over overall rates, since the mmr rate can only be equal to or greater than the overall rate.

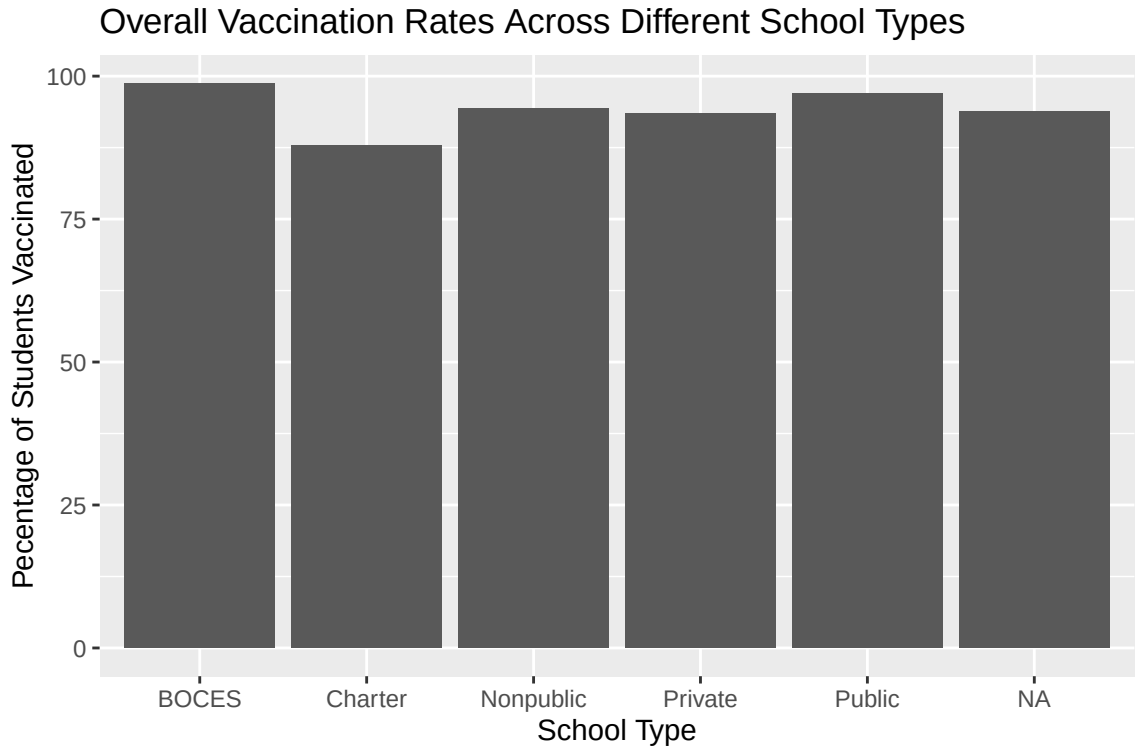
We eliminated California, Colorado, and Ohio from our analysis because these states had only 1, 2, and 2 observations, respectively. All other states had over 200 observations.

In regards to school type, we kept all types in the dataset since the lowest was "nonpublic" at a count of 18. However, due to the relatively low number of nonpublic and BOCES schools (which had a count of 47), the majority of our analysis by type of school was conducted between private, public, and charter schools.

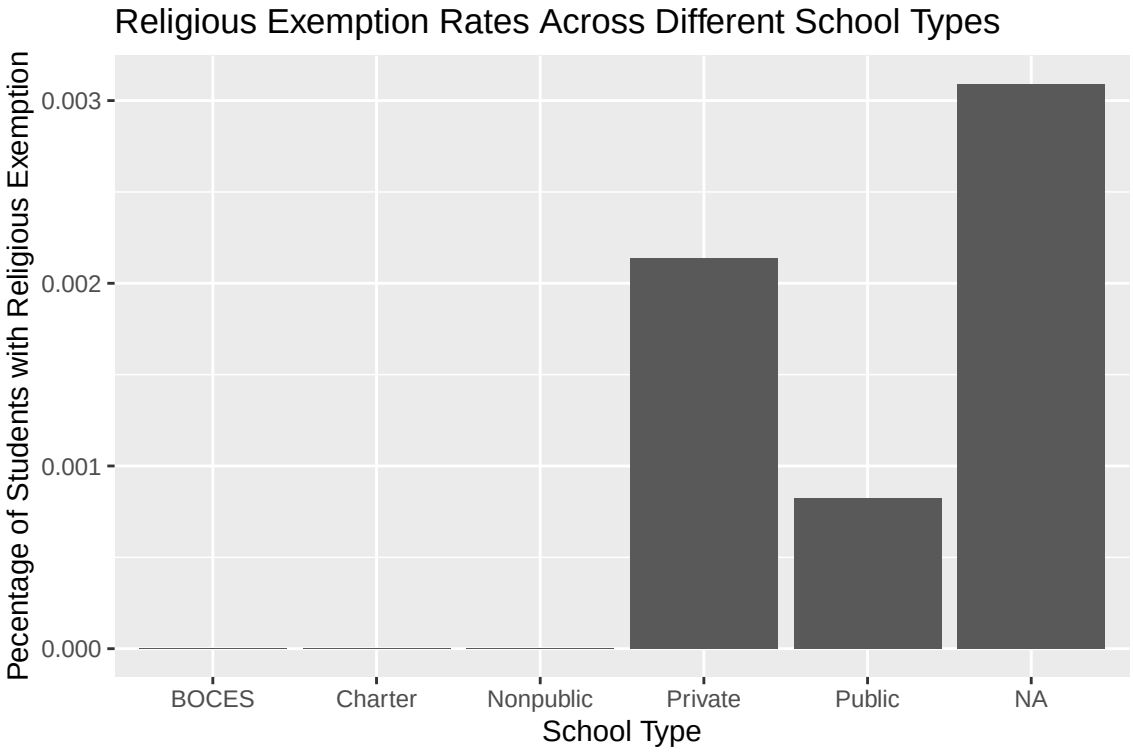
We created the "numvaxx" and "unvaxx" variables by using overall vaccination rates and enrollment rates of each school to be able to plot a logistic regression – this wouldn't have worked otherwise because both the predictor (state) and response (vax rate) have to be categorical, not continuous as the overall vaccination rate would have been.

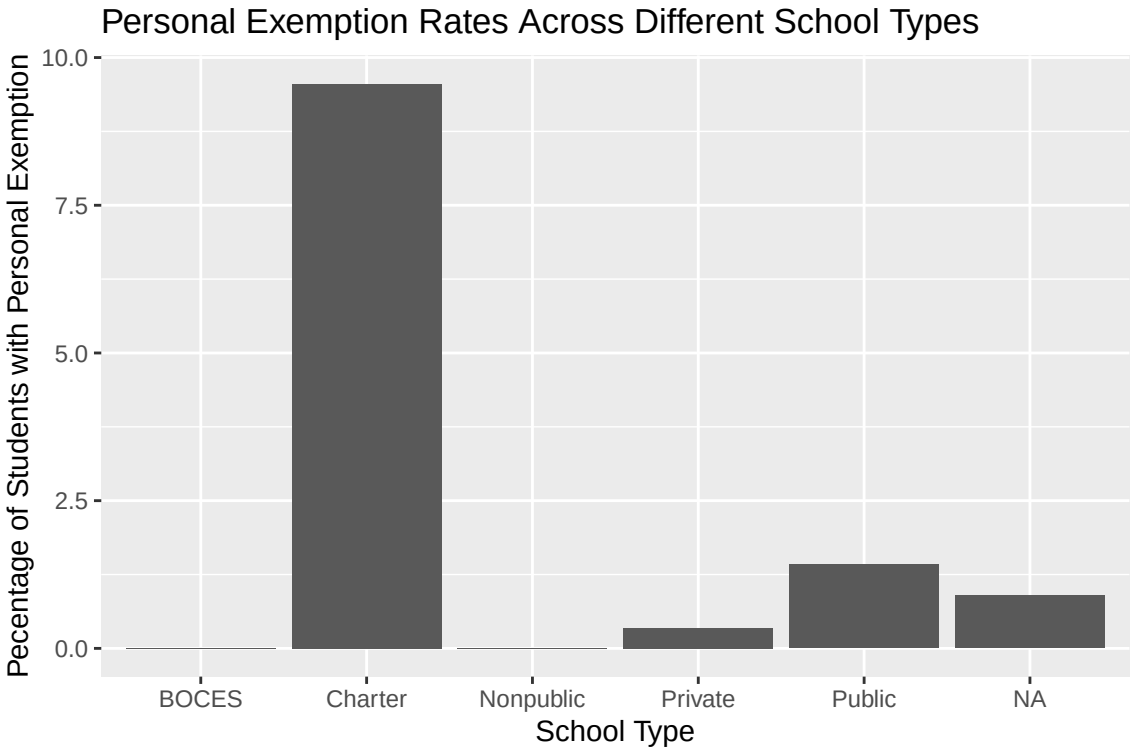
Exploratory Data Analysis

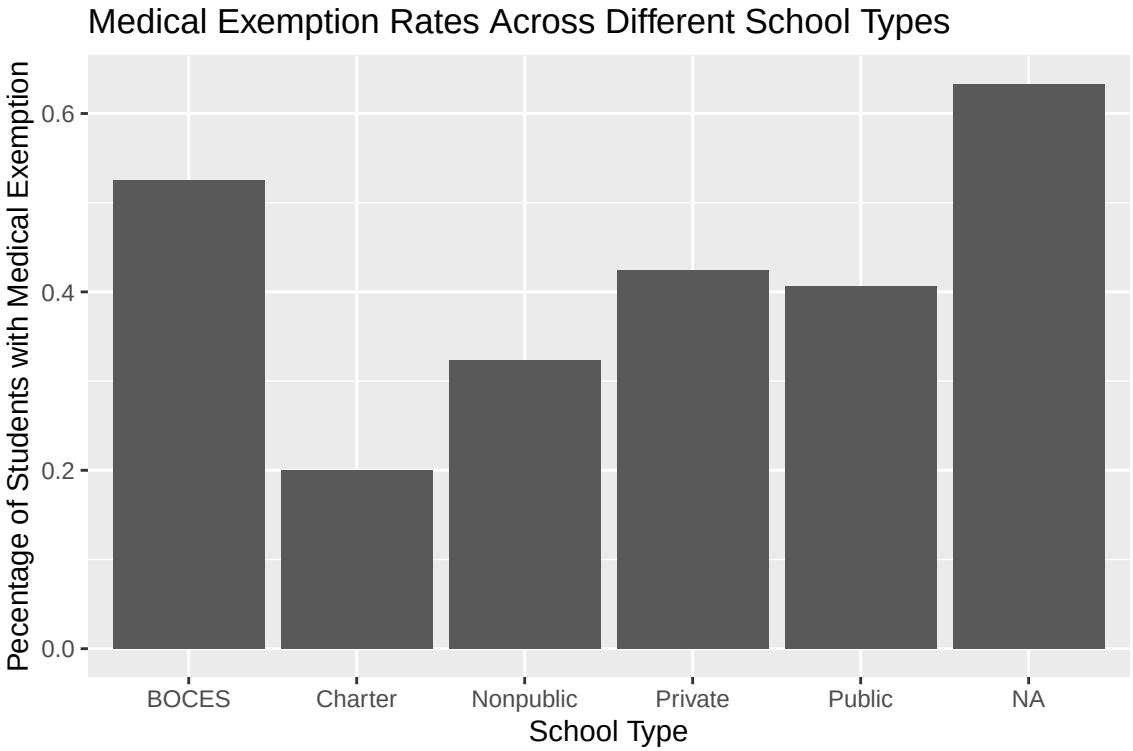
In terms of overall vaccination rate, we can see that charter schools have the lowest rates, whereas public and BOCES schools have the highest. It is important to note, however, that because the number of observations for certain types of schools is so small, not all of these differences may be significant.



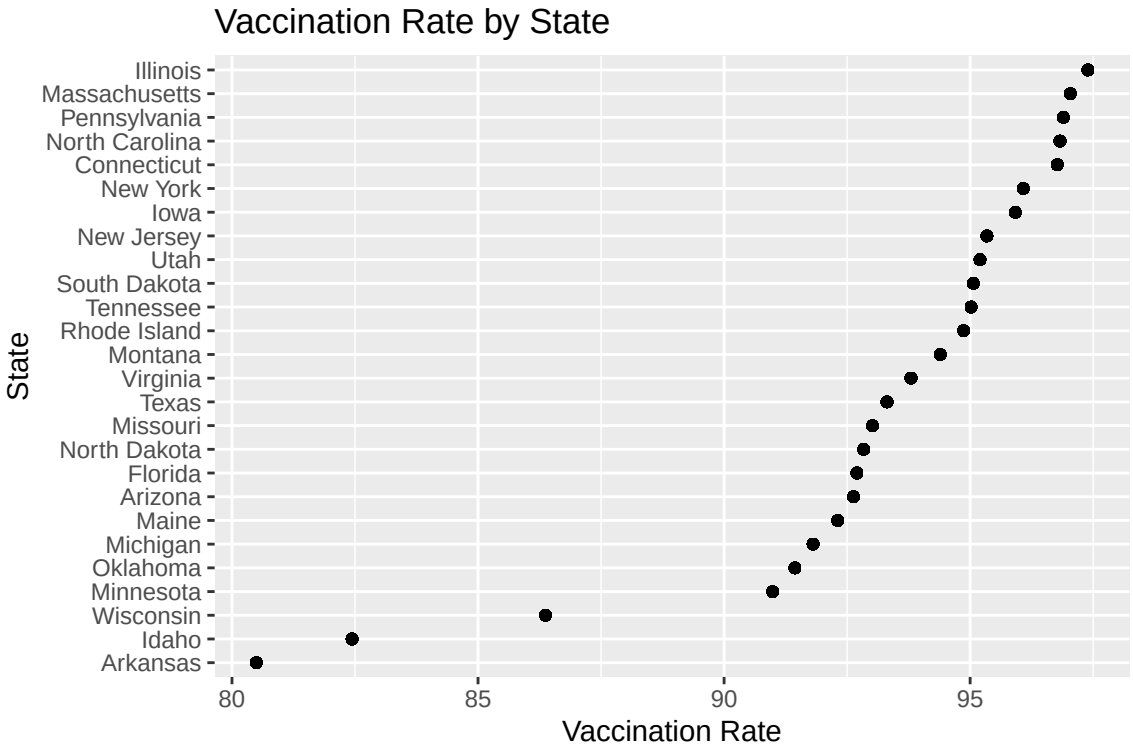
Looking at the y axis for the 3 exemption graphs, it is evident that personal exemption rates are the most common type of exemption among all of the school types. Personal exemptions have the highest rate in charter schools. Additionally, private schools have higher religious and medical exemption rates than public schools. However, since the number of religious exemptions is so small (e.g. 0.002% for private schools), it is difficult to tell whether differences in religious exemption rates would be statistically significant.

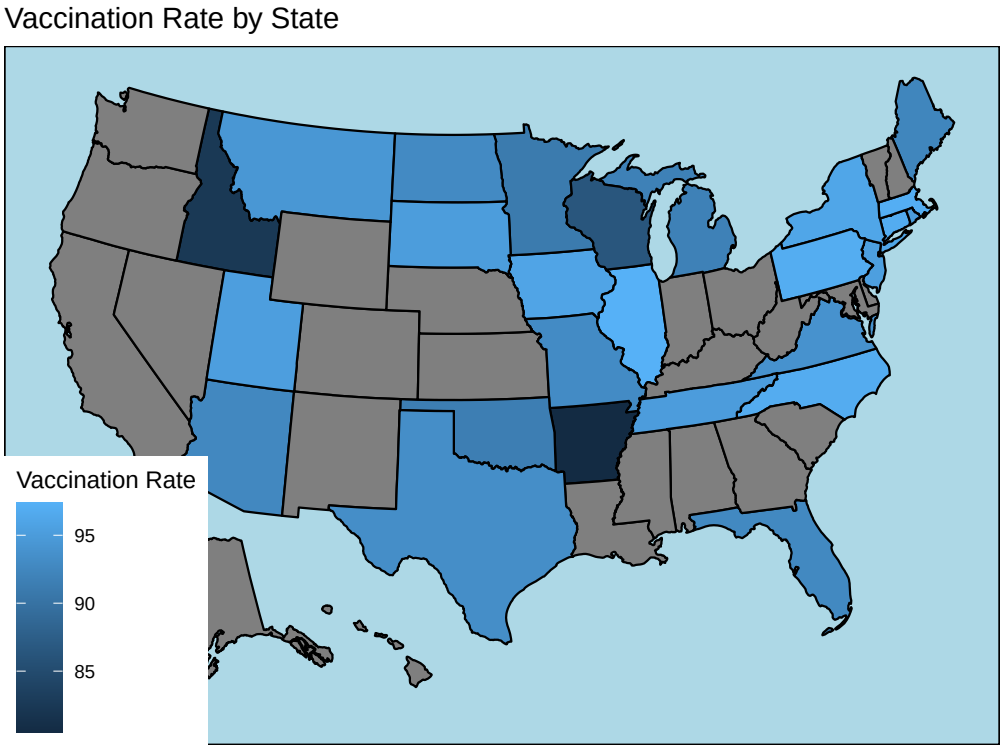

Religious Exemption Rates Across Different School Types
Pecentage of Students with Religious Exemption
0.003
0.002
0.001
0.000
BOCES
Charter
Nonpublic
Private
Public
NA
School Type


Personal Exemption Rates Across Different School Types
Pecentage of Students with Personal Exemption
10.0
7.5
5.0
2.5
0.0
BOCES
Charter
Nonpublic
Private
Public
NA
School Type



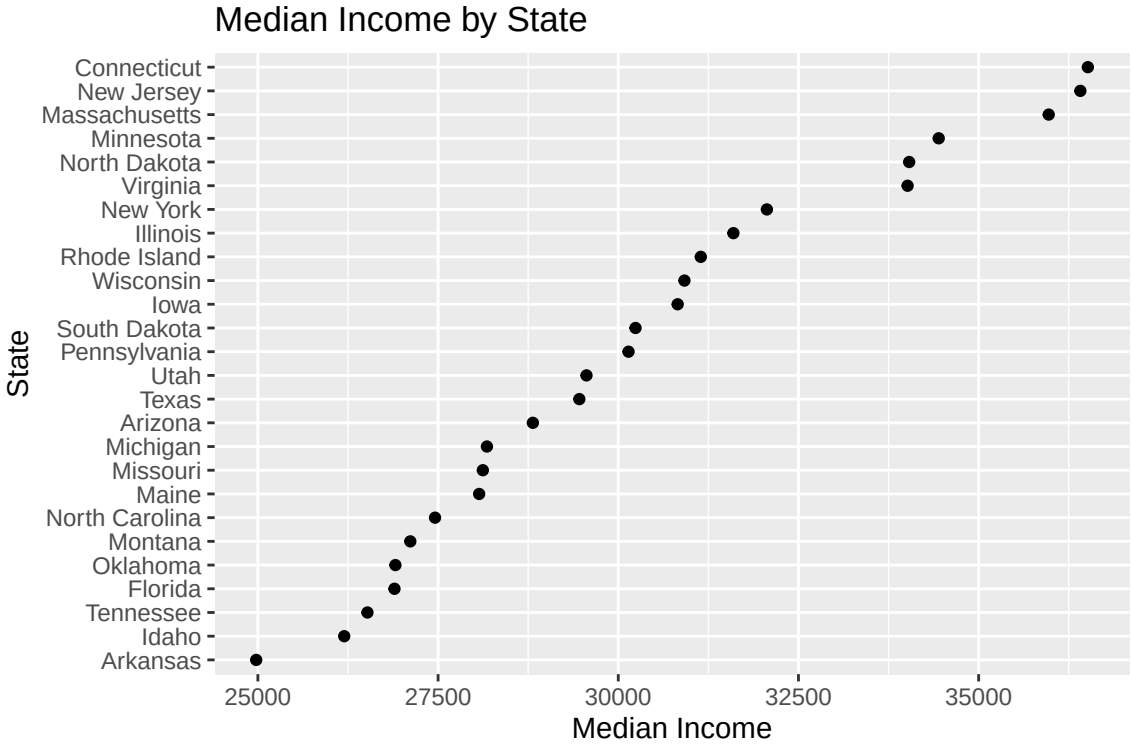
From the overall vaccination rate by state graph we can see that Illinois has the highest vaccination rate and Arkansas has the lowest. Comparatively, Idaho and Wisconsin have low vaccination rates at roughly 82.5% and 86% respectively. The rest of the states all have vaccination rates ranging from roughly 91% to 97.5%. As mentioned in our data wrangling, these figures represent a mix of overall and measles vaccination rates. Although it might be expected for the overall vaccination rate to be lower than the measles, mumps and rubella rate, it is still alarming that only about 7 out of 26 states are clearly above the 95% MMR vaccination rate recommended by the World Health Organization and the Center for Disease Control.





We loaded data from tidycensus to find the median income by state, as a way to compare income rates with vaccination rates. After graphing, we can see that Connecticut, New Jersey, and Massachusetts have the highest median income, with Connecticut having an estimated median income of $36,515. Idaho and Arkansas have the lowest, with Arkansas having a median income of $24,977. Interestingly, these two states

also have the lowest vaccination rate by state. We want to investigate whether there is a connection between state vaccination rate and median state income, which we will explore later by conducting linear regression.



T-Tests and ANOVA

We conducted three separate t-tests to evaluate the difference in vaccination rate between three different types of schools – public, private, and charter. In the first t-test, between private and public, the p-value is less than 0.05 so we can reject the null hypothesis that the two have the same means. The private school mean overall vaccination rate is 93.48%, and the public school mean is 97.01%. Between charter and public, the p-value is also below 0.05, so we can reject the null hypothesis that the two have the same mean overall vaccination rate. The charter mean is 87.96%. The last t-test was conducted between charter and private, with the p-value being less than 0.05, so we can reject the null hypothesis that the two are equal. From these tests we can see that out of the three, public schools have the highest mean vaccination rates, followed by private and charter. Notably, only public schools reach the WHO-recommended 95% threshold for vaccination rate on average.

Table 1: Output for T-Test between Public and Private Schools

estimate	estimate1	estimate2	statistic	p.value	parameter	conf.low	conf.high	method	alternative
-3.534196	93.47576	97.00995	-11.70155	0	2610.628	-4.126435	-2.941957	Welch Two Sample t-test	two.sided

Table 2: Output for T-Test between Public and Charter Schools

estimate	estimate1	estimate2	statistic	p.value	parameter	conf.low	conf.high	method	alternative
-9.054747	87.95521	97.00995	-11.42272	0	219.4524	-10.61702	-7.492476	Welch Two Sample t-test	two.sided

Table 3: Output for T-Test between Private and Charter Schools

estimate	estimate1	estimate2	statistic	p.value	parameter	conf.low	conf.high	method	alternative
-5.520551	87.95521	93.47576	-6.553249	0	279.4151	-7.178836	-3.862267	Welch Two Sample t-test	two.sided

We used an ANOVA to determine if all of the states have the same vaccination rates. The ANOVA test gave us a very high f-value of 320.7 and a p-value of $2 * 10^{-16}$ which is much smaller than 0.05. Because of the high f-value and small p-value, we can reject our null hypothesis that all of the state means are the same. Thus, vaccination rate is dependent on the state you are from for at least one state. We will now look at logistic regression to determine which states have a significant difference in vaccination rates.

Table 4: ANOVA

term	df	sumsq	meansq	statistic	p.value
state	25	516286.4	20651.45612	320.7196	0
Residuals	39479	2542092.1	64.39099	NA	NA

Regression Analysis

Arkansas is the reference, because it has the lowest vaccination rate per state. Our null hypothesis (that there is no relationship between state and vaccination rate) was rejected for every state, as every state has a p-value less than 0.05. All else held constant, the probability that you are vaccinated in any state is the "estimate" times the odds relative to Arkansas. However, the logistic regression may not be accurate because the coefficient estimates seem implausible, since Illinois being around 10 indicates that people in Illinois would be 10 times more likely to be vaccinated than people in Arkansas. This is far too high to be reasonable. The ordering of the states' coefficients does match their relative vaccination rates, however.

Table 5: Output for Logistic Regression

term	estimate	std.error	statistic	p.value
(Intercept)	4.072161	0.0046249	303.60851	0
statefacArizona	3.370289	0.0143644	84.58427	0
statefacFlorida	3.379824	0.0093171	130.70867	0
statefacIllinois	10.359710	0.0061598	379.54404	0
statefacIowa	6.067351	0.0095512	188.76406	0
statefacMaine	3.693512	0.0262128	49.84498	0
statefacMichigan	3.215546	0.0113500	102.90705	0
statefacMinnesota	3.361359	0.0138015	87.84142	0
statefacMontana	2.599367	0.0111232	85.88081	0
statefacNew Jersey	6.592150	0.0170592	110.54917	0
statefacNorth Carolina	6.903029	0.0156597	123.37174	0
statefacNorth Dakota	3.586700	0.0336633	37.94139	0

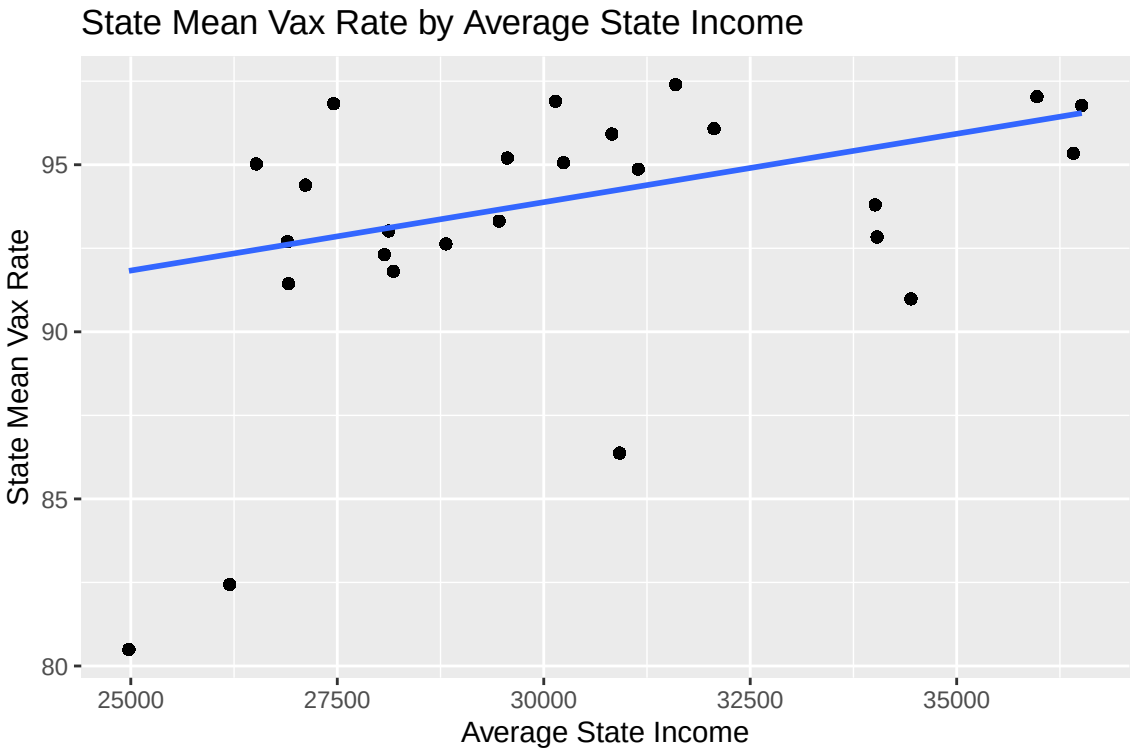
term	estimate	std.error	statistic	p.value
statefacPennsylvania	8.292260	0.0167164	126.54160	0
statefacRhode Island	5.622851	0.0480118	35.96695	0
statefacSouth Dakota	7.430762	0.0507160	39.54630	0
statefacTennessee	4.569600	0.0161181	94.26834	0
statefacUtah	5.426775	0.0101558	166.54017	0
statefacVirginia	3.454342	0.0138466	89.52595	0

Fitting a linear regression model with median income as the explanatory variable shows us that for every $1000 increase in the state's median income, we expect the state's mean vaccination rate to increase by 0.4092%. The line of best fit shows a positive relationship between average state income and state mean vaccination rate. There seems to be 3 outliers in regard to vaccination rate at roughly $25,000, $26,000, $31,000. We should also be cautious fitting a linear regression model on a variable with so few observations (26).

Equation for predicting state mean vax rate with income: $\hat{y} = 81.60 + 0.0004092 * x_i$

Table 6: Output for Linear Regression with Respect to Income

term	estimate	std.error	statistic	p.value
(Intercept)	81.6017595	0.1836077	444.43548	0
estimate	0.0004092	0.0000060	68.64739	0



Summary

In the data collected, there is a relationship between overall vaccination rate and type of school, as well as vaccination rate and state, and vaccination rate and median state income. Charter schools were the school type with the lowest vaccination rate, while public schools were the highest. The data is not very clear on vaccination rates in general, as discussed in the data wrangling section, because many schools don't report vaccination rates the same way – some report it with just MMR rates, whereas some report with overall rates. Also, we had to remove states which only provided only a few schools as case studies and applied these few observations to the overall state rates. These factors make it hard to be completely confident in the relationships we modeled in the project.

Hopefully, COVID will increase the demand to collect clearer, more standardized, and more comprehensive vaccination data from schools. Furthermore, since we've found that vaccination rate tends to be lower in states with a lower median state income, when working to increase COVID vaccinations, we should target low income states and areas. Also, since COVID vaccine eligibility has just extended to kids ages 5-11, it is important that we target places with children that have been less likely to be receive other vaccinations. For instance, in our analysis, we've shown that charter schools have had significantly lower rates of vaccination. Taking all of these factors into account will make for a more robust COVID vaccination campaign.