Final Report

due November 16, 2021 by 11:59 PM

Lindsey Weyant, Ali Raich, Aden Clemente

11/16/21

```
## [1] "abc8289fa2ba274ced76d97c7f8ee31666a2c931"
## Getting data from the 2014-2018 5-year ACS
## Getting data from the 2014-2018 5-year ACS
## Getting data from the 2014-2018 5-year ACS
```

#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. Thus, our overarching research question is: How do measles vaccination rates vary across the country and different types of schools?

realrate vaccination status vs. state, realrate vaccination status vs. type of school, each type of exemption (personal, religious, and medical) vs. state exemption vs. type of school. To analyze vaccination and exemption rates by states, we will use spatial data to show the change in these rates across the country. Then, we can use two-sample t-tests to test for significance of vaccination and exemption rates between different types of schools. If there are significantly lower vaccination rates in private schools vs. other types of schools, this will support our main hypothesis.

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.

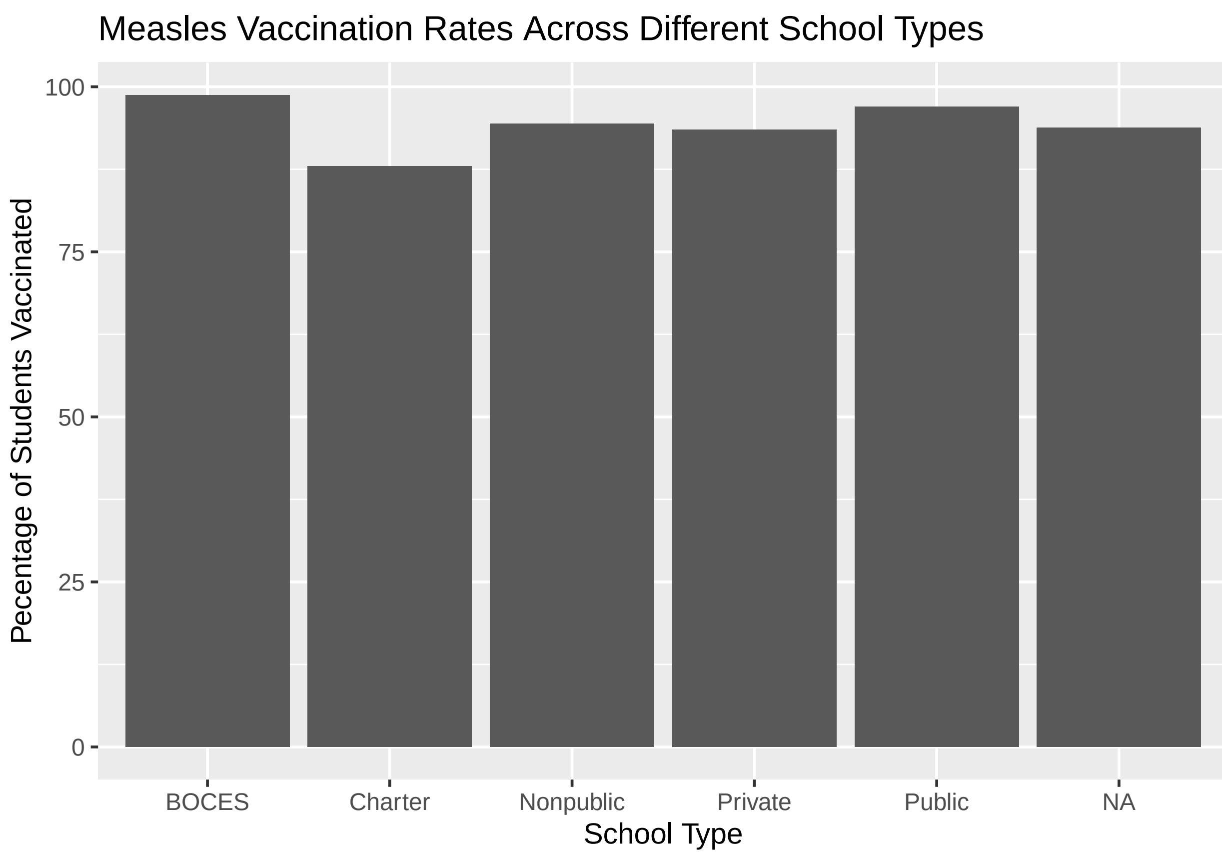
For most schools, a value was only provided for one of the "overall" (overall vaccination rate) and "mmr" (measles, mumps, and rubella vaccination rate) variables. Choosing to conduct our analyses on one of these variables would entail losing a massive number of observations. So, we created a new variable, "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.

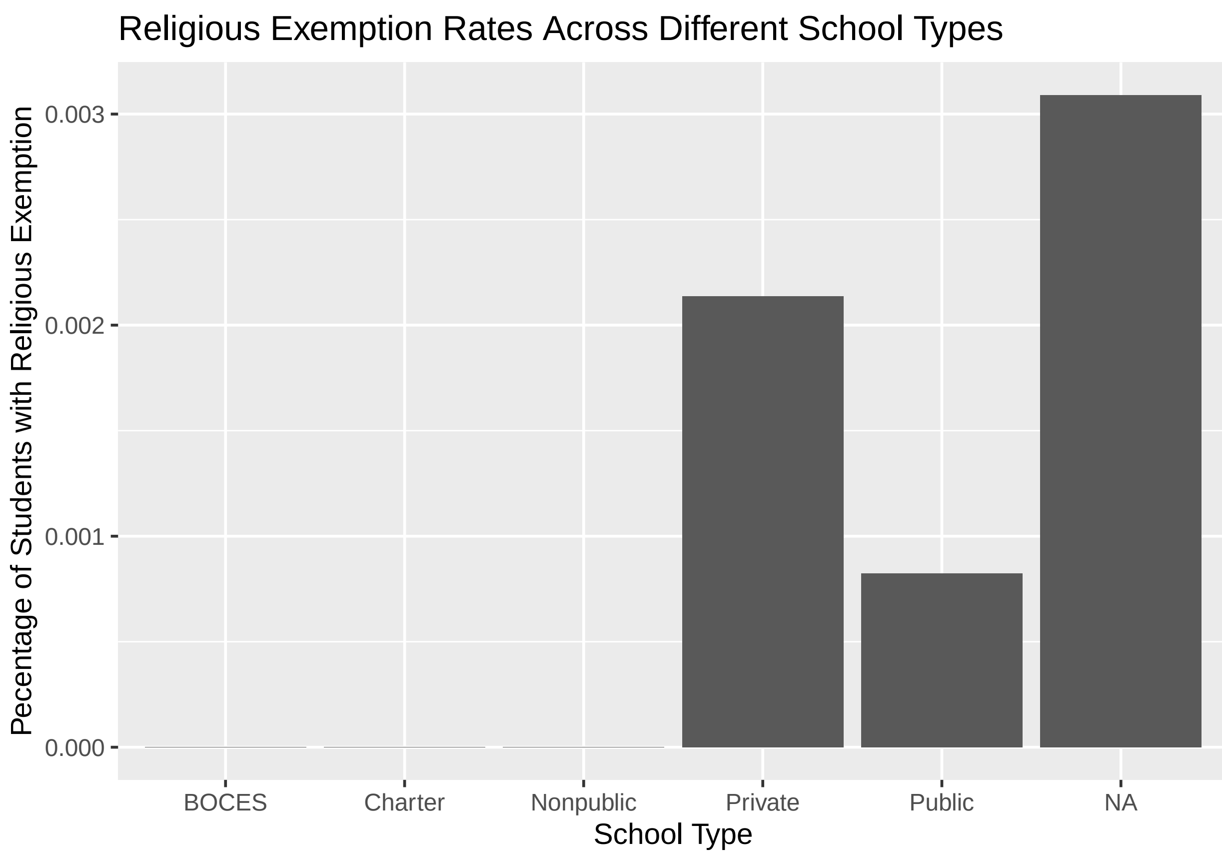
In regards to school type, we kept all types in the dataset since the lowest was "nonpublic" at a count of 18. However, due to the relatively low number of nonpublic and BOCES schools (which had a count of 47), the majority of our analysis by type of school was conducted between private, public, and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

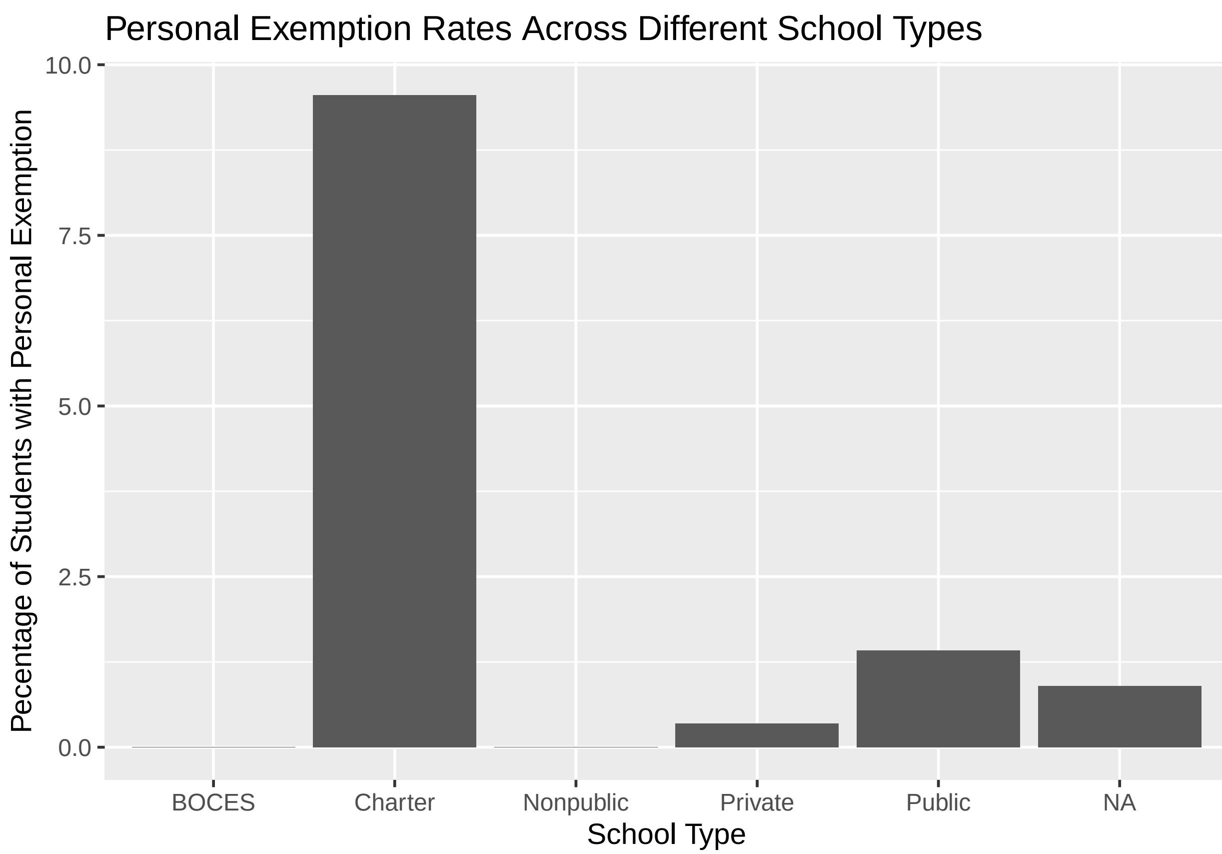
We created the numvaxx and unvaxx variables by using overall vax rates and enrollment rates of each school to be able to plot a logistic regression – this wouldn't have worked otherwise because both the predictor (state) and response (vax rate) have to be categorical, not continuous as the overall vax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

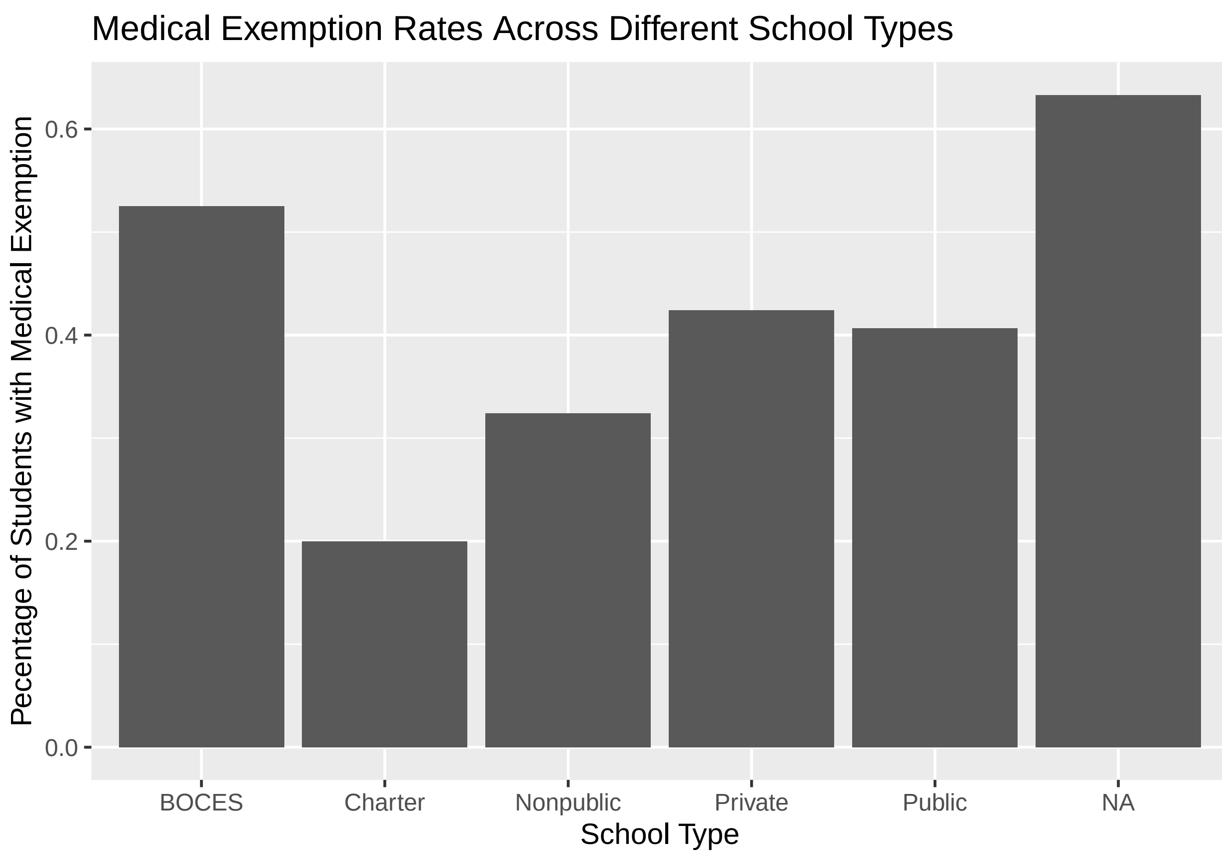
Exploratory Data Analysis



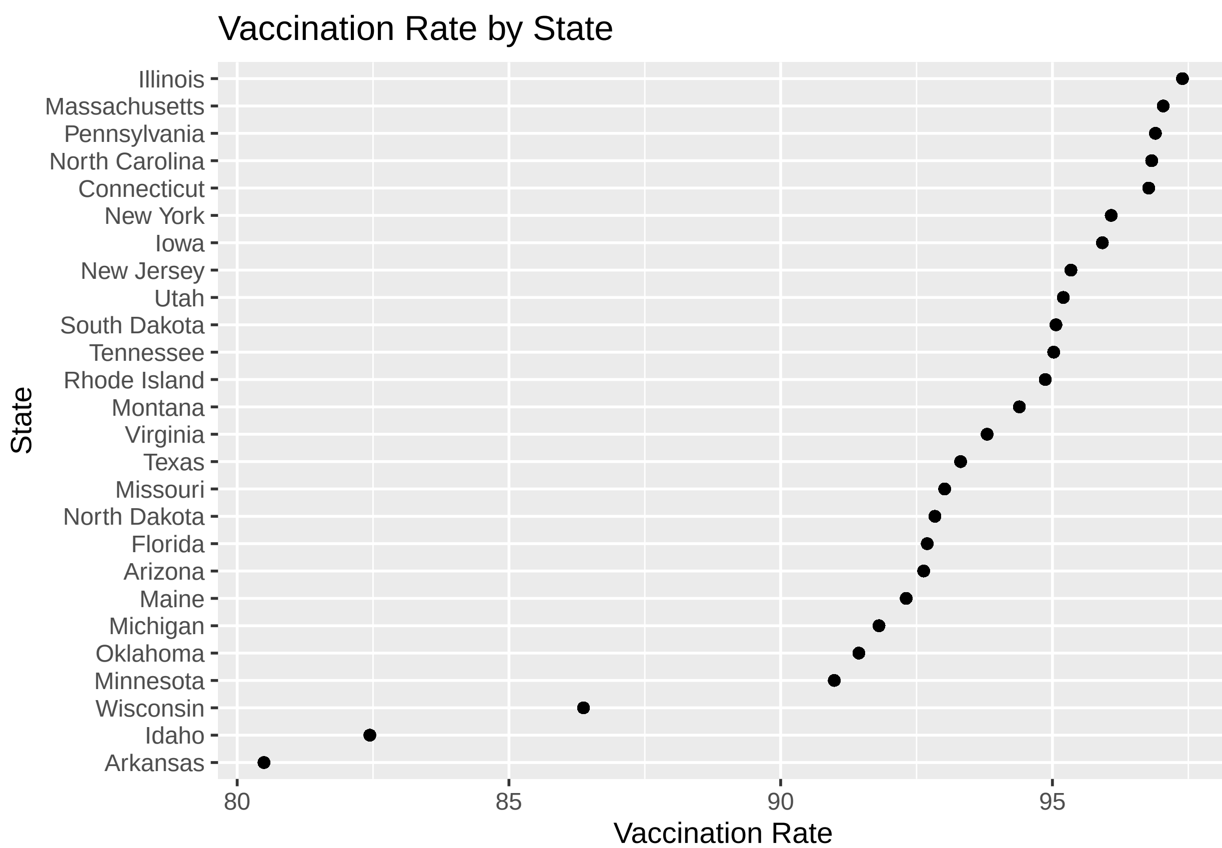
Looking at the y axis for these 4 graphs, it is evident that personal exemption rates are the most common type of exemption among all of the school types. Personal exemptions have the highest rate in charter schools. Additionally, private schools have higher religious and medical exemption rates than public schools. Overall, we can see that charter schools have the lowest measles vaccination rates. These graphs demonstrate that one is most likely to be unvaccinated and have an exemption at a charter school. However it is important to note that because the percentage differences are so small we do not know for certain that these differences are not due to random variation until we do more tests on the difference in vaccination rates in different types of schools.

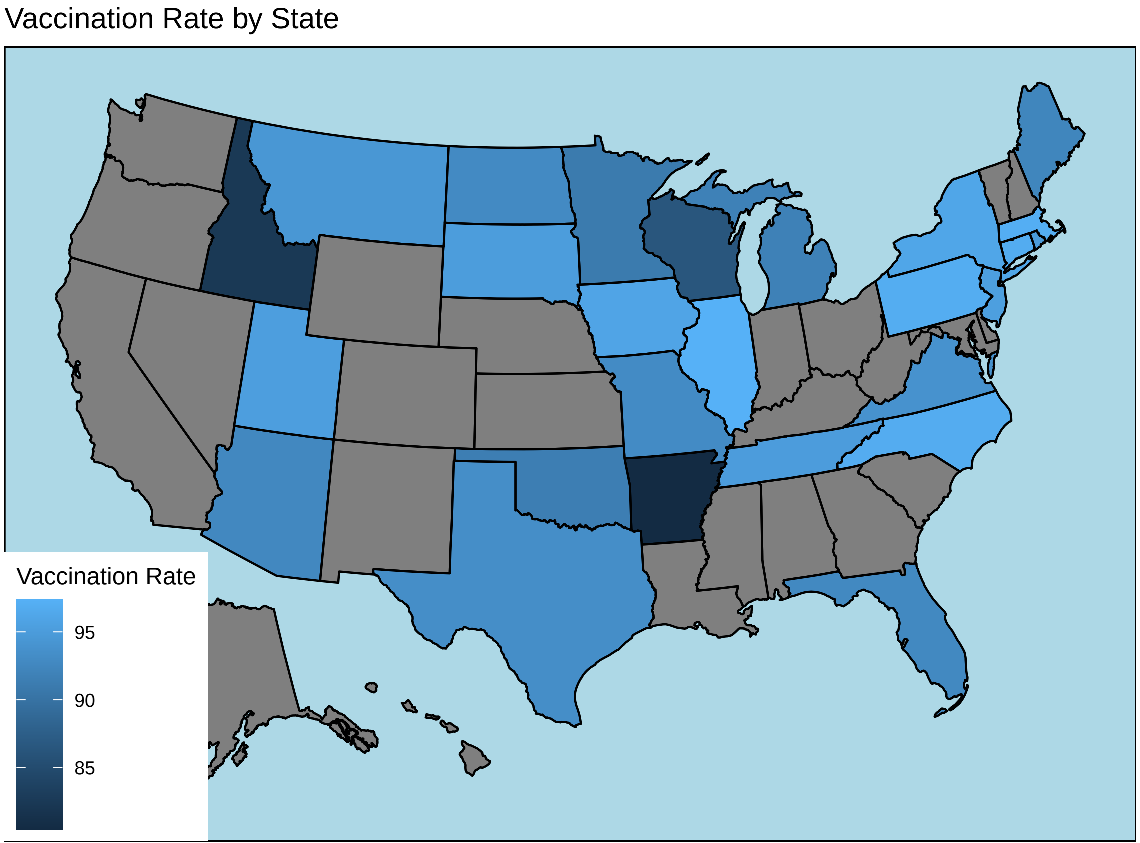

Religious Exemption Rates Across Different School Types
Pecentage of Students with Religious Exemption
0.003
0.002
0.001
0.000
BOCES
Charter
Nonpublic
Private
Public
NA
School Type


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

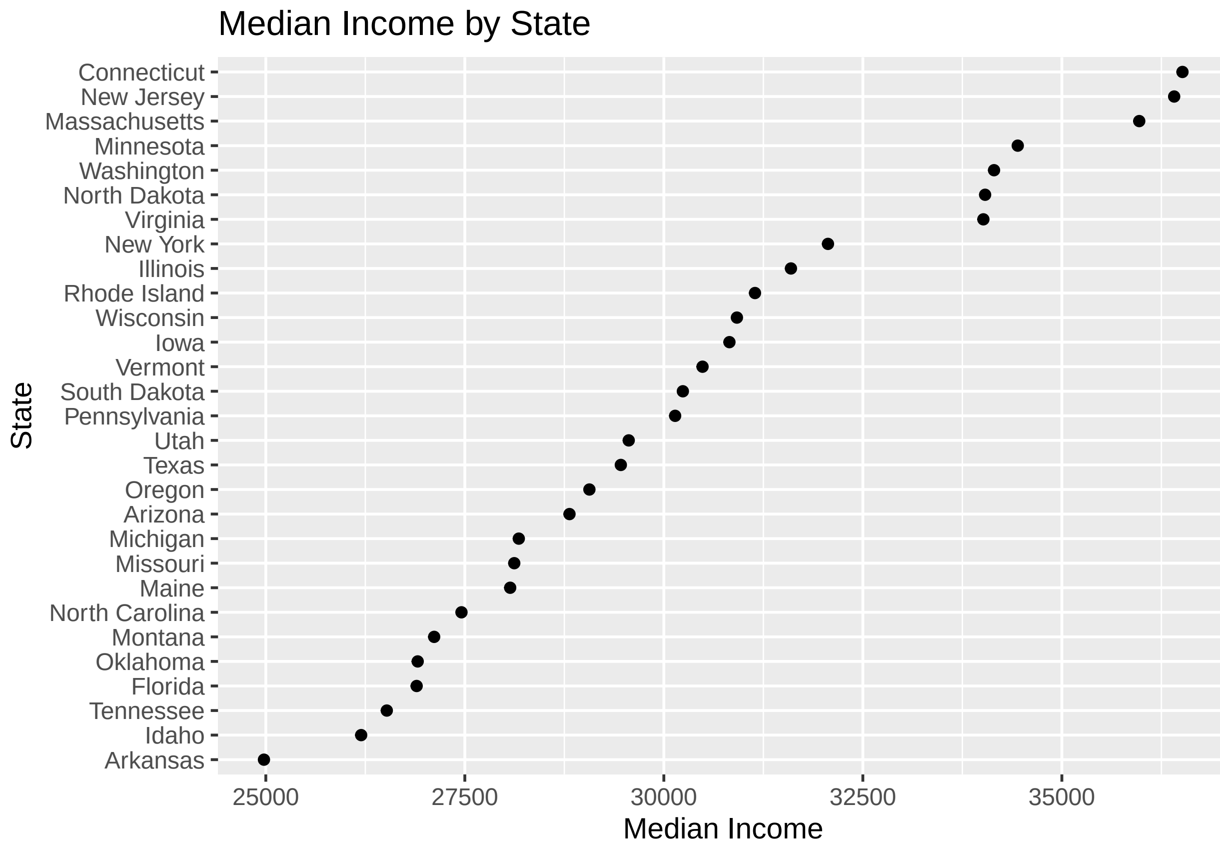


From the overall vaccination rate by state graph we can see that Illinois has the highest vaccination rate and Arkansas has the lowest. Comparatively, Idaho and Wisconsin have low vaccination rates at roughly 82.5% and 86% respectively. The rest of the states all have vaccination rates ranging from roughly 91% to 97.5%. Although it is expected that the overall vaccination rate to be lower than the measles, mumps and rubella rate, it is alarming that only 7 about 7 states are clearly above the 95% vaccination rate recommended by the World Health Organization (WHO).


Vaccination Rate by State
State
Illinois
Massachusetts
Pennsylvania
North Carolina
Connecticut
New York
Iowa
New Jersey
Utah
South Dakota
Tennessee
Rhode Island
Montana
Virginia
Texas
Missouri
North Dakota
Florida
Arizona
Maine
Michigan
Oklahoma
Minnesota
Wisconsin
Idaho
Arkansas
80
85
90
95
Vaccination Rate



We loaded data from tidycensus to find the median income by state, as a way to compare income rates with vaccination rates later on. After graphing, it is shown that Connecticut, New Jersey, and Massachusetts have the highest median income, with Connecticut having an estimated value of $36,515. Idaho and Arkansas have the lowest, with Arkansas having a median income of $24,977. This is good to note, as these two states also have the lowest vaccination rate by state. So, it is possible that there is a connection between state vaccination rate and median state income, which we will explore later with a 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, so you are more likely to be vaccinated if you go to a public school over a private or charter school. The difference between the three school types is quite large, especially when combined with the earlier information that WHO recommends a vaccination rate of 95%. Only public schools reach this percentage 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

```
##                Df  Sum Sq Mean Sq F value Pr(>F)
## state          25  516286   20651   320.7 <2e-16 ***
## Residuals   39479 2542092      64
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Regression Analysis

```
##
## Call:  glm(formula = cbind(numvaxx, unvaxx) ~ statefac, family = binomial,
##     data = measles)
##
## Coefficients:
##           (Intercept)         statefacArizona         statefacFlorida
##                1.4042                  1.2150                  1.2178
##      statefacIllinois            statefacIowa           statefacMaine
##                2.3379                  1.8029                  1.3066
##      statefacMichigan       statefacMinnesota         statefacMontana
##                1.1680                  1.2123                  0.9553
##    statefacNew Jersey  statefacNorth Carolina    statefacNorth Dakota
##                1.8859                  1.9320                  1.2772
##  statefacPennsylvania    statefacRhode Island    statefacSouth Dakota
##                2.1153                  1.7268                  2.0056
##     statefacTennessee            statefacUtah        statefacVirginia
##                1.5194                  1.6913                  1.2396
##
## Degrees of Freedom: 28126 Total (i.e. Null);  28109 Residual
##   (11347 observations deleted due to missingness)
## Null Deviance:       316100
## Residual Deviance: 177000    AIC: 255100
```

```
##
## Call:
## glm(formula = cbind(numvaxx, unvaxx) ~ statefac, family = binomial,
##     data = measles)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -154.356    -0.709     0.554     1.628    12.531
```

```
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)             1.404174   0.004625  303.61   <2e-16 ***
## statefacArizona         1.214998   0.014364   84.58   <2e-16 ***
## statefacFlorida         1.217824   0.009317  130.71   <2e-16 ***
## statefacIllinois        2.337924   0.006160  379.54   <2e-16 ***
## statefacIowa            1.802922   0.009551  188.76   <2e-16 ***
## statefacMaine           1.306578   0.026213   49.84   <2e-16 ***
## statefacMichigan        1.167997   0.011350  102.91   <2e-16 ***
## statefacMinnesota       1.212345   0.013802   87.84   <2e-16 ***
## statefacMontana         0.955268   0.011123   85.88   <2e-16 ***
## statefacNew Jersey      1.885880   0.017059  110.55   <2e-16 ***
## statefacNorth Carolina  1.931960   0.015660  123.37   <2e-16 ***
## statefacNorth Dakota    1.277233   0.033663   37.94   <2e-16 ***
## statefacPennsylvania    2.115323   0.016716  126.54   <2e-16 ***
## statefacRhode Island    1.726839   0.048012   35.97   <2e-16 ***
## statefacSouth Dakota    2.005628   0.050716   39.55   <2e-16 ***
## statefacTennessee       1.519426   0.016118   94.27   <2e-16 ***
## statefacUtah            1.691345   0.010156  166.54   <2e-16 ***
## statefacVirginia        1.239632   0.013847   89.53   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 316104  on 28126  degrees of freedom
## Residual deviance: 177040  on 28109  degrees of freedom
##   (11347 observations deleted due to missingness)
## AIC: 255099
##
## Number of Fisher Scoring iterations: 5
```

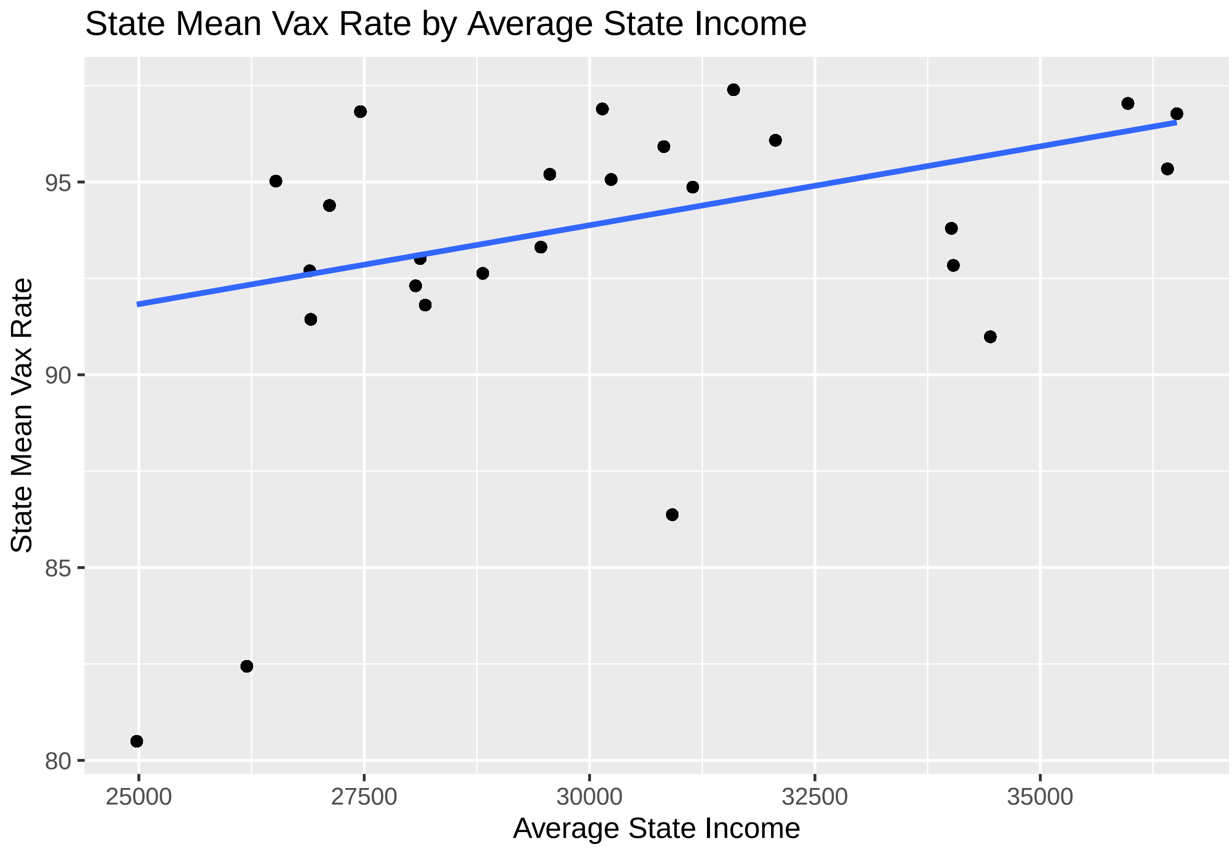
For every \$1000 increase in the state's median income we expect the state mean vaccination rate to increase by 0.4092%. The line shows a positive relationship between average state income and state mean vaccination rate. There seems to be 3 outliers at the average state income of roughly \$25,000, \$26,000, \$31,000.

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

Table 4: 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

```
## `geom_smooth()` using formula 'y ~ x'
```


State Mean Vax Rate by Average State Income
State Mean Vax Rate
95
90
85
80
25000
27500
30000
32500
35000
Average State Income