

Final Report

due November 16, 2021 by 11:59 PM

Lindsey Weyant, Ali Raich, Aden Clemente

11/16/21

```
#Load Data
```

```
## Your original .Renviron will be backed up and stored in your R HOME directory if needed.
```

```
## Your API key has been stored in your .Renviron and can be accessed by Sys.getenv("CENSUS_API_KEY").
```

```
## To use now, restart R or run `readRenviron("~/Renviron")`
```

```
## [1] "abc8289fa2ba274ced76d97c7f8ee31666a2c931"
```

```
#v18 <- load_variables(2018, "acs5", cache = TRUE)
```

```
#View(v18)
```

```
## Getting data from the 2014-2018 5-year ACS
```

```
## # A tibble: 6 x 5
```

	GEOID	NAME	variable	estimate	moe
	<chr>	<chr>	<chr>	<dbl>	<dbl>
## 1	01	Alabama	B07011_001	25375	132
## 2	02	Alaska	B07011_001	33413	428
## 3	04	Arizona	B07011_001	28815	147
## 4	05	Arkansas	B07011_001	24977	139
## 5	06	California	B07011_001	30797	65
## 6	08	Colorado	B07011_001	34109	231

```
## Getting data from the 2014-2018 5-year ACS
```

```
## # A tibble: 6 x 5
```

	GEOID	NAME	variable	estimate	moe
	<chr>	<chr>	<chr>	<dbl>	<dbl>
## 1	01	Alabama	B02001_002	3317453	3345
## 2	02	Alaska	B02001_002	478834	1368
## 3	04	Arizona	B02001_002	5364141	9871
## 4	05	Arkansas	B02001_002	2302874	2783
## 5	06	California	B02001_002	23529068	26419
## 6	08	Colorado	B02001_002	4655584	5852

```
## Getting data from the 2014-2018 5-year ACS
```

```
## # A tibble: 6 x 5
```

	GEOID	NAME	variable	estimate	moe
	<chr>	<chr>	<chr>	<dbl>	<dbl>
## 1	01	Alabama	B01003_001	4864680	NA
## 2	02	Alaska	B01003_001	738516	NA
## 3	04	Arizona	B01003_001	6946685	NA

```
## 4 05    Arkansas    B01003_001 2990671    NA
## 5 06    California B01003_001 39148760   NA
## 6 08    Colorado   B01003_001 5531141    NA
```

#Research Question

How do measles vaccination rates vary across the country and demographics in 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

```
#racerates <- left_join(race, population, by="GEOID") %>%
#pivot_wider(names_from = "variable", values_from = "estimate")
```

```
## # A tibble: 39,505 x 29
## # Groups:   state [26]
##   index state year name type city county district enroll mmr overall
##   <dbl> <chr> <chr> <chr> <chr> <chr> <chr> <lgl> <dbl> <dbl> <dbl>
## 1 1 Arizona 2018-19 A J M~ Publ~ Noga~ Santa~ NA 51 100 -1
## 2 2 Arizona 2018-19 Acade~ Char~ Tucs~ Pima NA 22 100 -1
## 3 3 Arizona 2018-19 Acade~ Char~ Tucs~ Pima NA 85 100 -1
## 4 4 Arizona 2018-19 Acade~ Char~ Phoe~ Maric~ NA 60 100 -1
## 5 5 Arizona 2018-19 Accla~ Char~ Phoe~ Maric~ NA 43 100 -1
## 6 6 Arizona 2018-19 Alfre~ Publ~ Phoe~ Maric~ NA 36 100 -1
## 7 7 Arizona 2018-19 All A~ Char~ Phoe~ Maric~ NA 24 100 -1
## 8 8 Arizona 2018-19 Ameri~ Char~ Yuma Yuma NA 22 100 -1
## 9 9 Arizona 2018-19 Arizo~ Priv~ Phoe~ Maric~ NA 26 100 -1
## 10 10 Arizona 2018-19 Arizo~ Publ~ San ~ Yuma NA 78 100 -1
## # ... with 39,495 more rows, and 18 more variables: xrel <lgl>, xmed <dbl>,
## # xper <dbl>, lat <dbl>, lng <dbl>, realrate <dbl>, numvaxx <dbl>,
## # unvaxx <dbl>, numxmed <dbl>, numxrel <dbl>, numxper <dbl>, numxother <dbl>,
## # statefac <fct>, statemean <dbl>, GEOID <chr>, variable <chr>,
## # estimate <dbl>, moe <dbl>
```

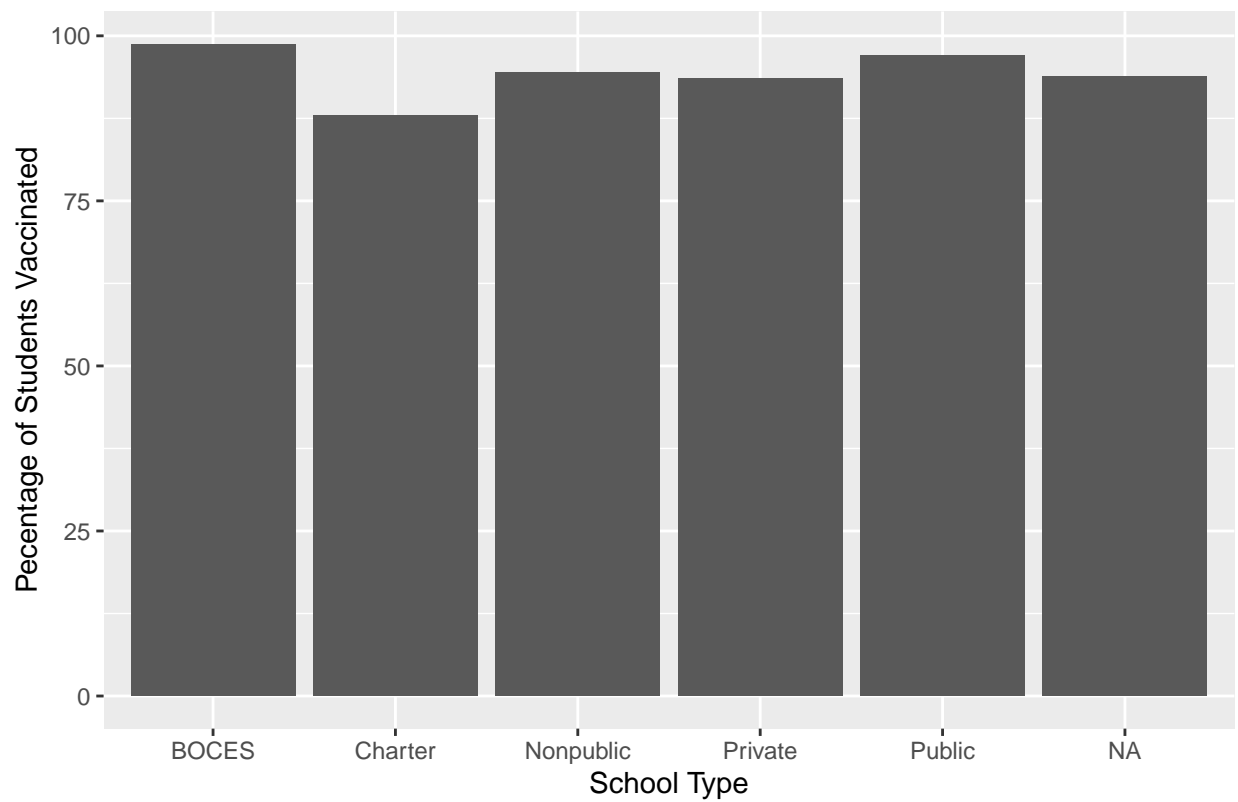
Exploratory Data Analysis

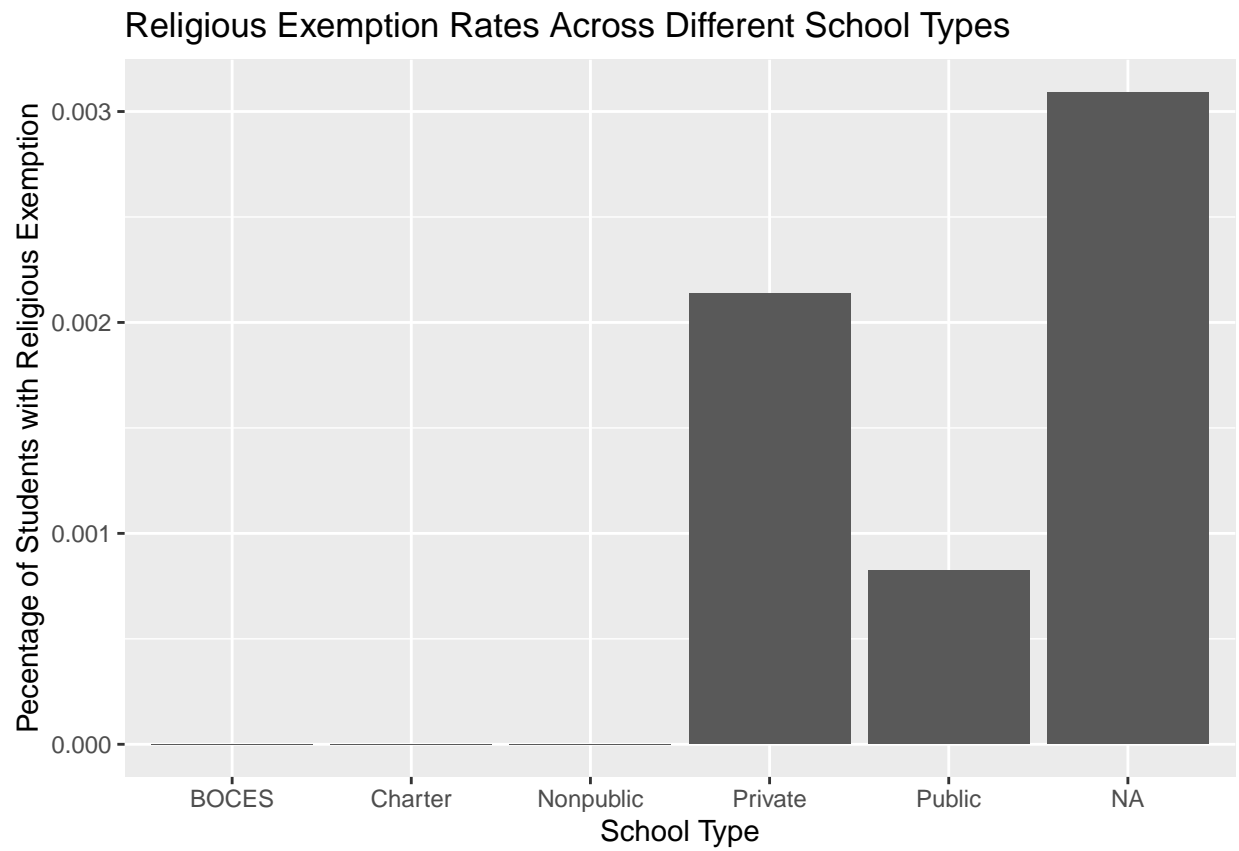
```
## # A tibble: 6 x 2
## # Groups:   type [6]
##   type n
##   <chr> <int>
## 1 BOCES 47
## 2 Charter 217
## 3 Nonpublic 18
## 4 Private 2339
## 5 Public 4855
## 6 <NA> 32029

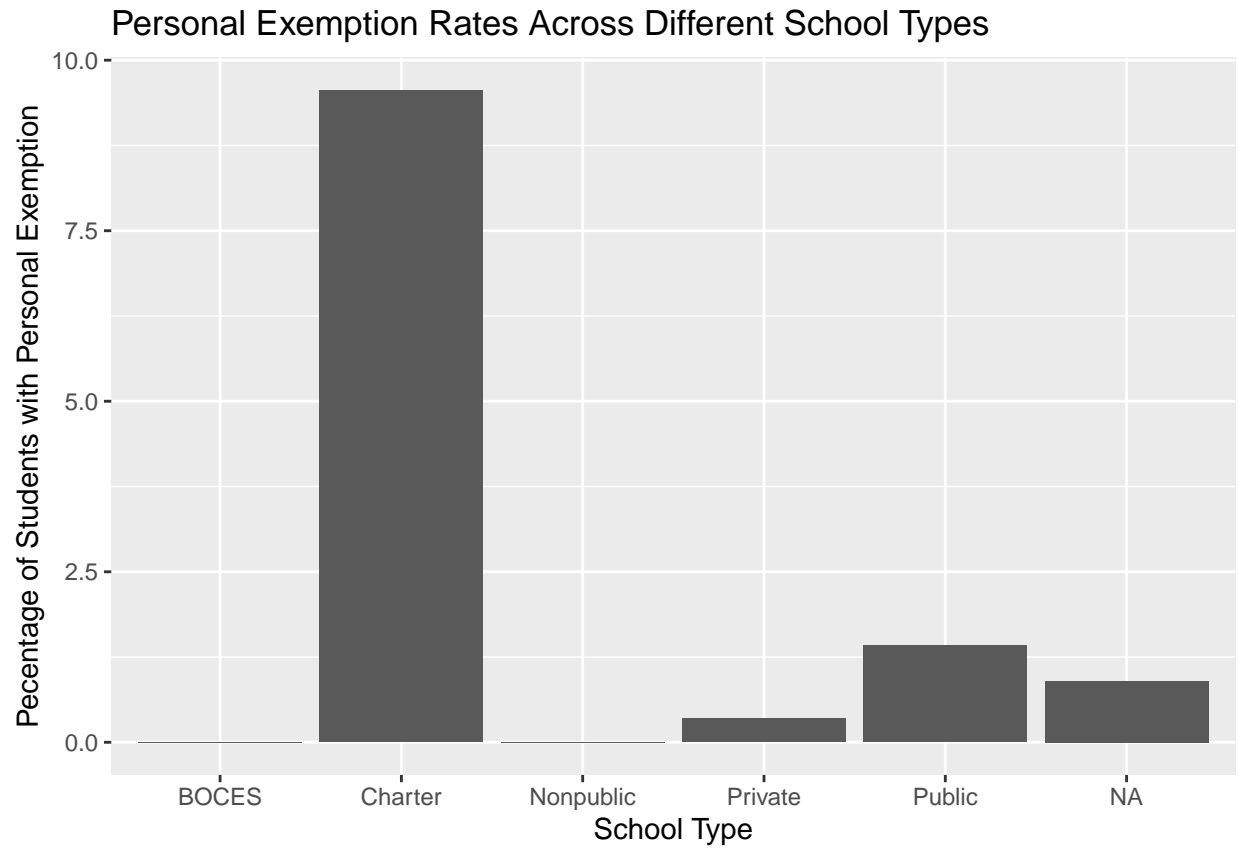
## # A tibble: 26 x 2
## # Groups:   state [26]
```

```
##      state      n
##      <chr>    <int>
## 1 Arizona    1171
## 2 Arkansas    567
## 3 Connecticut  589
## 4 Florida    2672
## 5 Idaho       467
## 6 Illinois    7686
## 7 Iowa       1163
## 8 Maine       357
## 9 Massachusetts 954
## 10 Michigan   2351
## # ... with 16 more rows
```

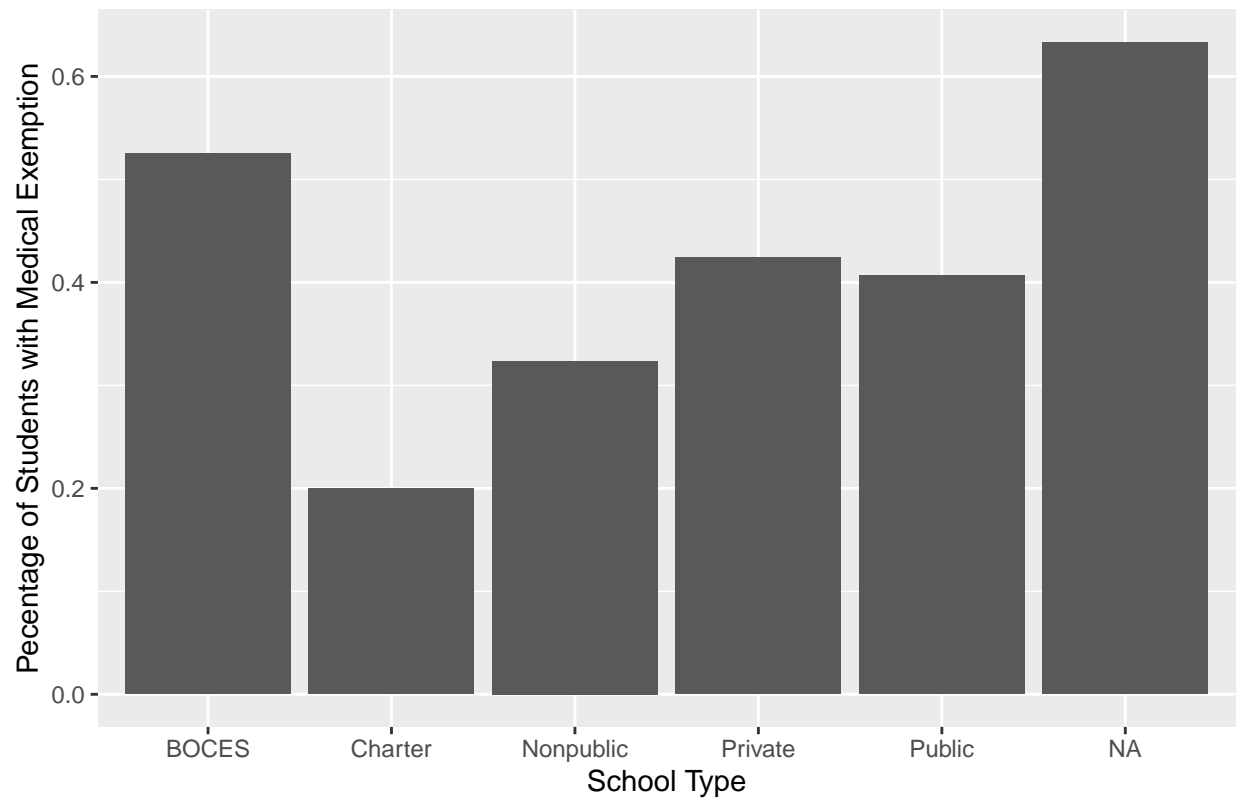
Measles Vaccination Rates Across Different School Types



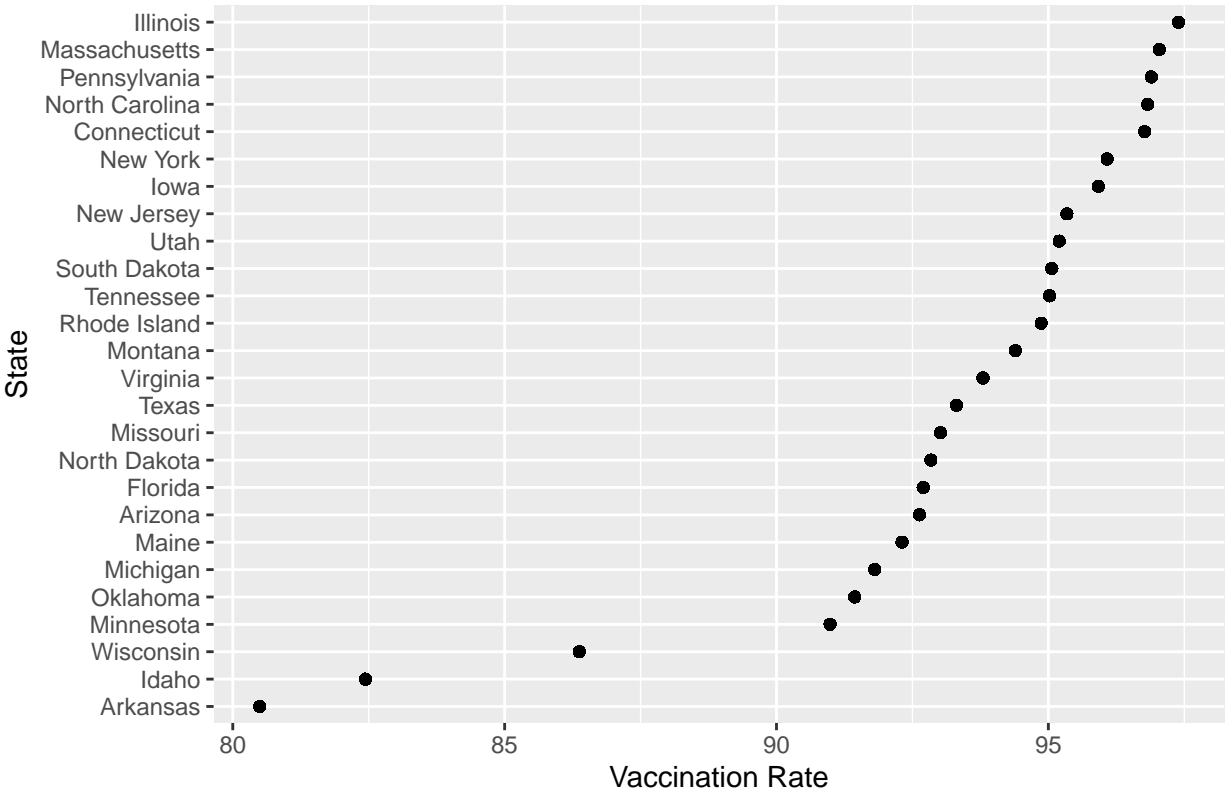




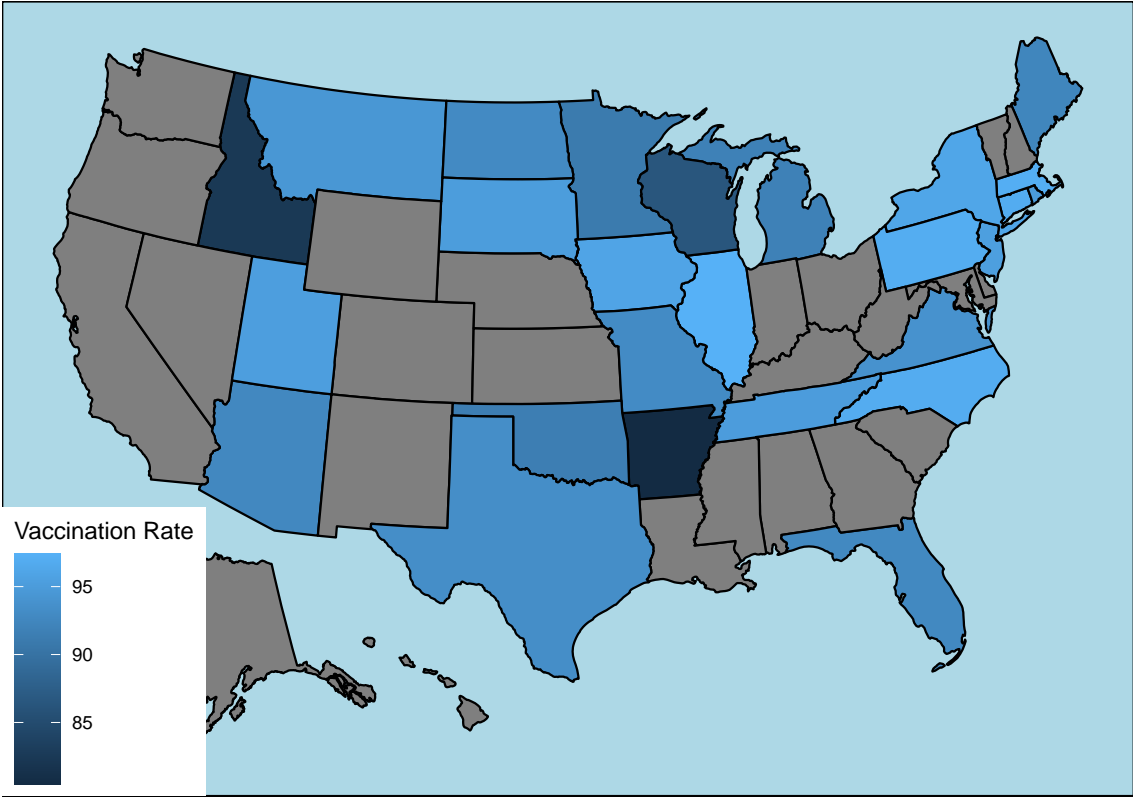
Medical Exemption Rates Across Different School Types

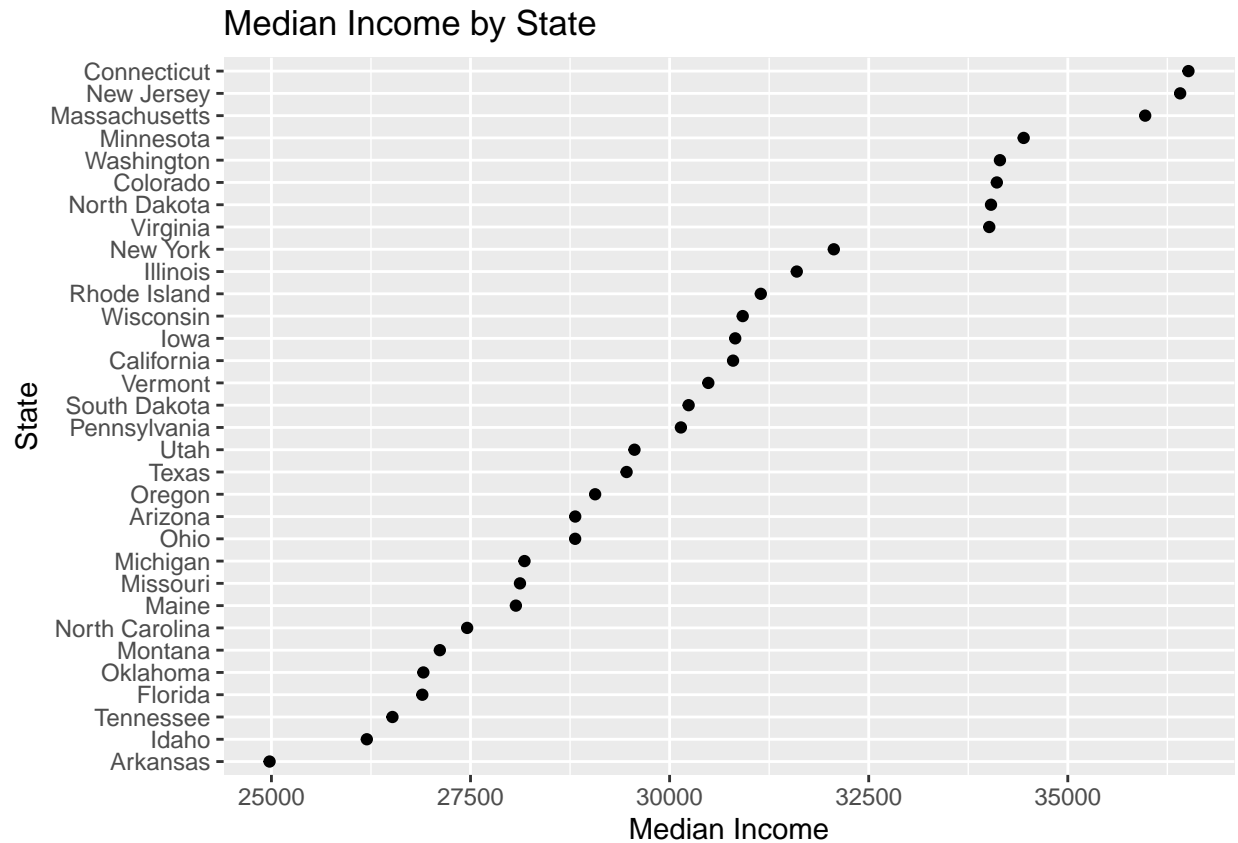


Vaccination Rate by State



Vaccination Rate by State





T-Tests and ANOVA

```
##           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

##
## Welch Two Sample t-test
##
## data: measles3$realrate by measles3$type
## t = -11.702, df = 2610.6, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group Private and group Public is not equal
## 95 percent confidence interval:
##  -4.126435 -2.941957
## sample estimates:
## mean in group Private mean in group Public
##           93.47576           97.00995

##
## Welch Two Sample t-test
##
## data: measles4$realrate by measles4$type
## t = -11.423, df = 219.45, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group Charter and group Public is not equal
```

```
## 95 percent confidence interval:
## -10.617019 -7.492476
## sample estimates:
## mean in group Charter mean in group Public
## 87.95521 97.00995

##
## Welch Two Sample t-test
##
## data: measles5$realrate by measles5$type
## t = -6.5532, df = 279.42, p-value = 2.704e-10
## alternative hypothesis: true difference in means between group Charter and group Private is not equal
## 95 percent confidence interval:
## -7.178836 -3.862267
## sample estimates:
## mean in group Charter mean in group Private
## 87.95521 93.47576
```

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
##
## parsnip model object
##
## Fit time: 10ms
##
## Call:
## stats::lm(formula = statemean ~ estimate, data = data)
##
## Coefficients:
## (Intercept)          estimate
##    8.160e+01     4.092e-04
##
## `geom_smooth()` using formula 'y ~ x'

```



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