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

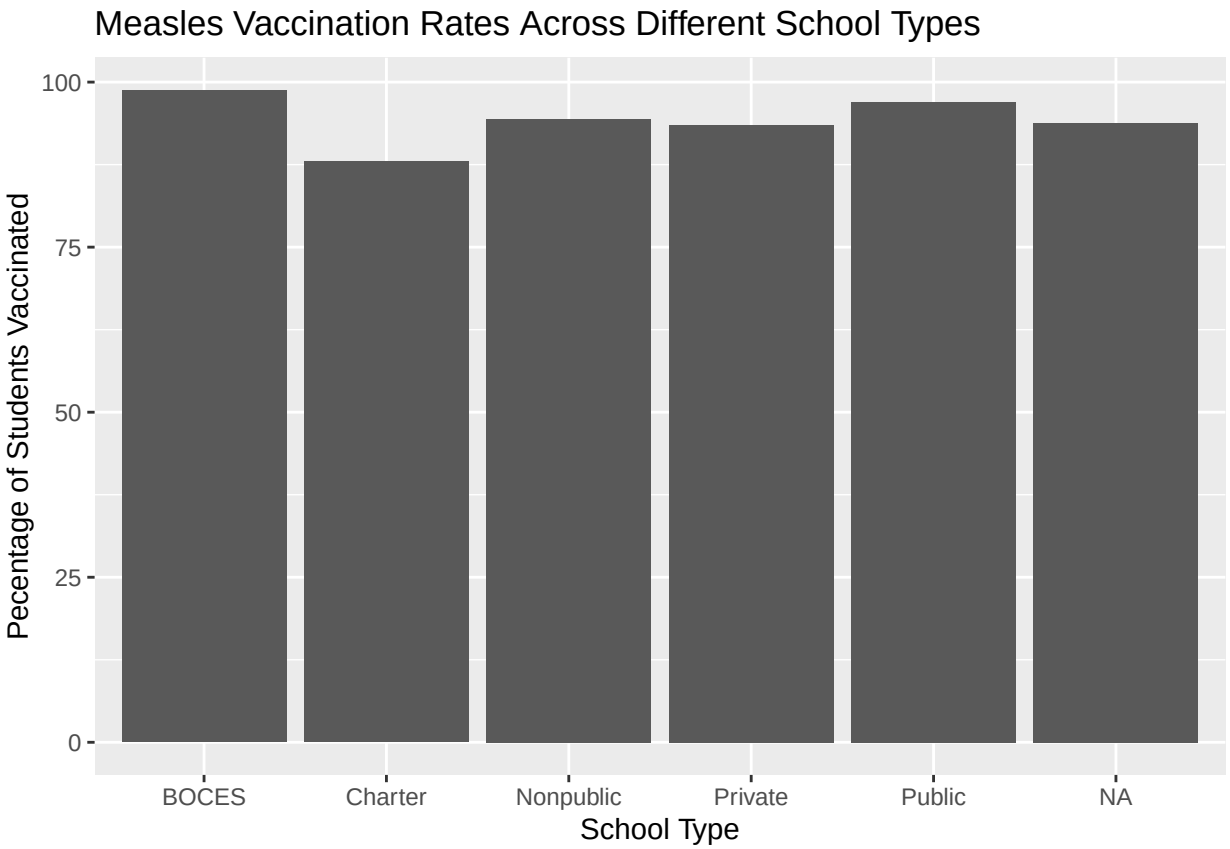
```
#racerates <- left_join(race, population, by="GEOID") %>%
#pivot_wider(names_from = "variable", values_from = "estimate")
```

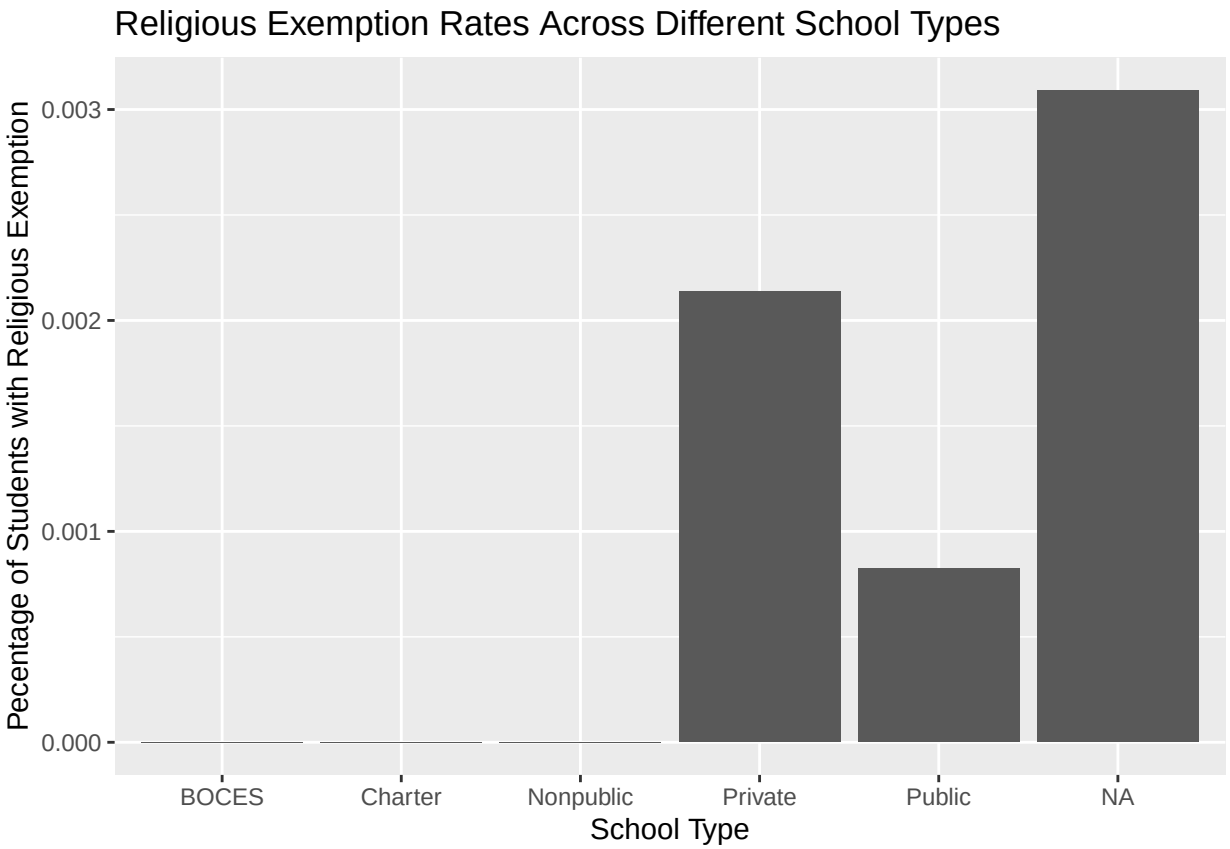
```
#incomerates <- left_join(income, measles, by = "statemean")
```

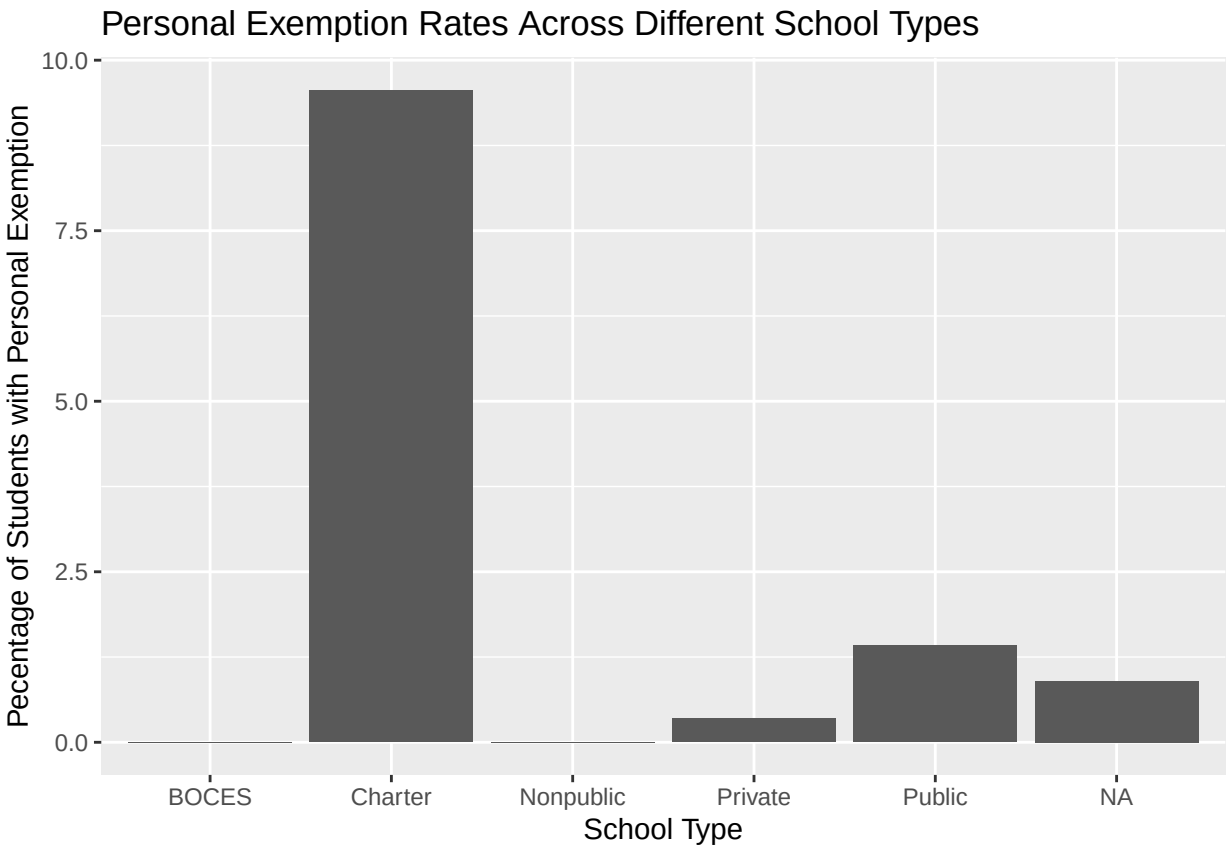
Exploratory Data Analysis

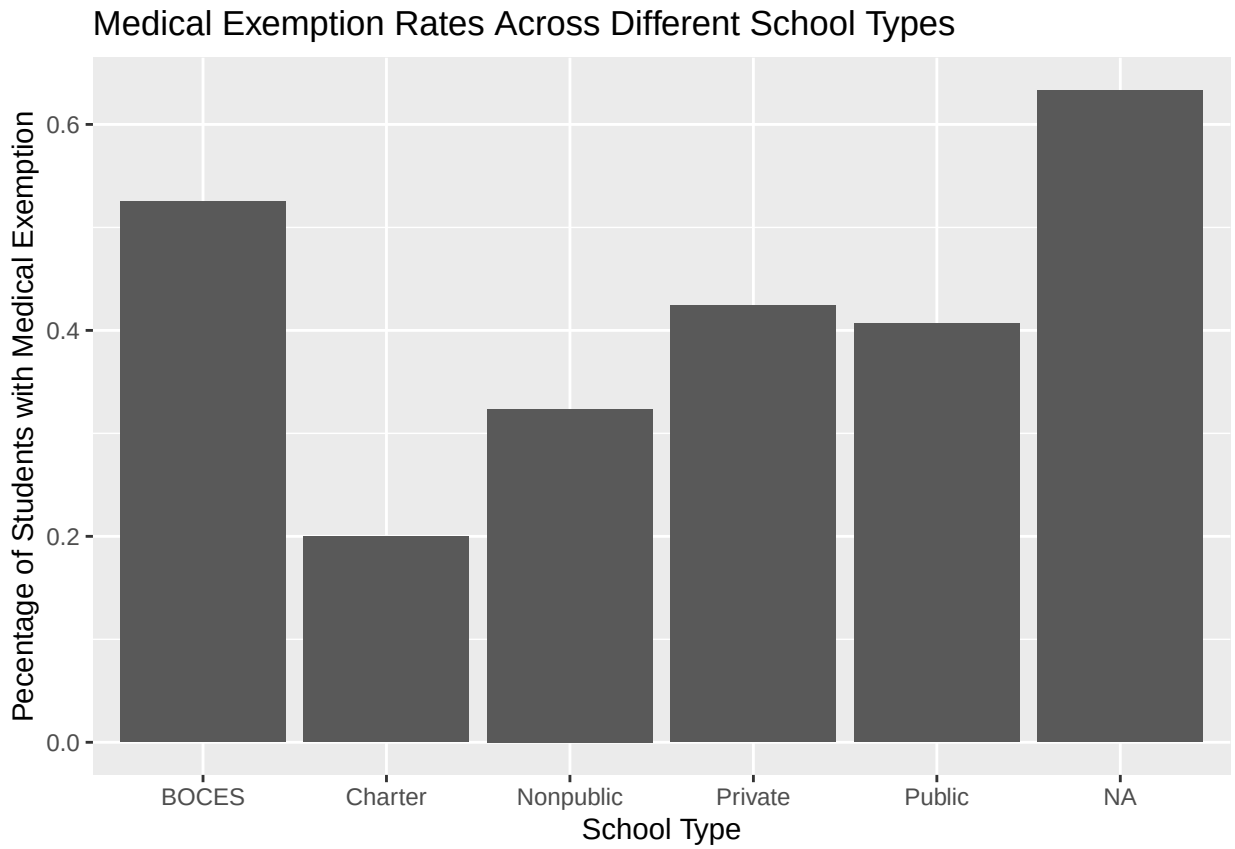
```
## # A tibble: 6 x 2
## # Groups:   type [6]
##   type          n
##   <chr>     <int>
## 1 BOCES        47
## 2 Charter     217
## 3 Nonpublic    18
## 4 Private    2339
## 5 Public     4855
## 6 <NA>      32029
```

```
## # 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
```

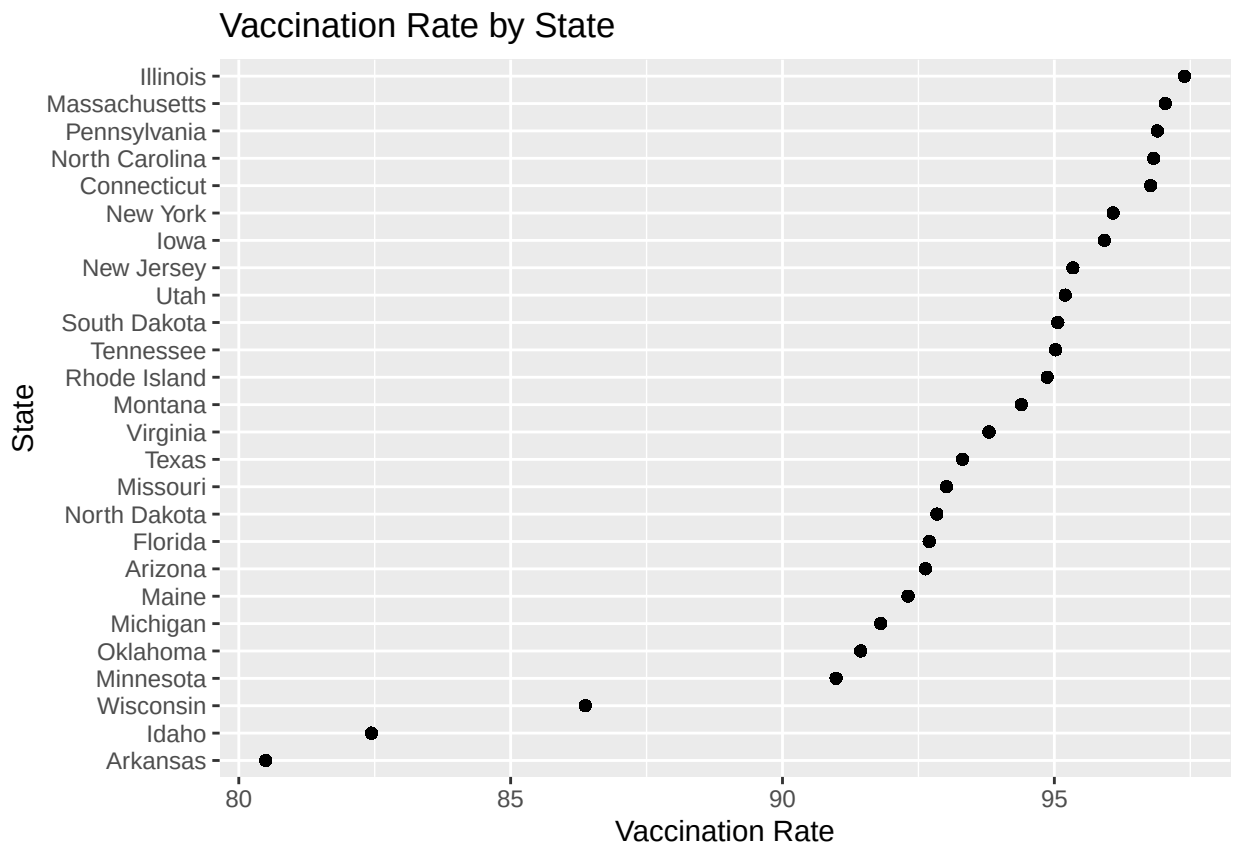

Measles Vaccination Rates Across Different School Types
Pecentage of Students Vaccinated
100
75
50
25
0
BOCES
Charter
Nonpublic
Private
Public
NA
School Type


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

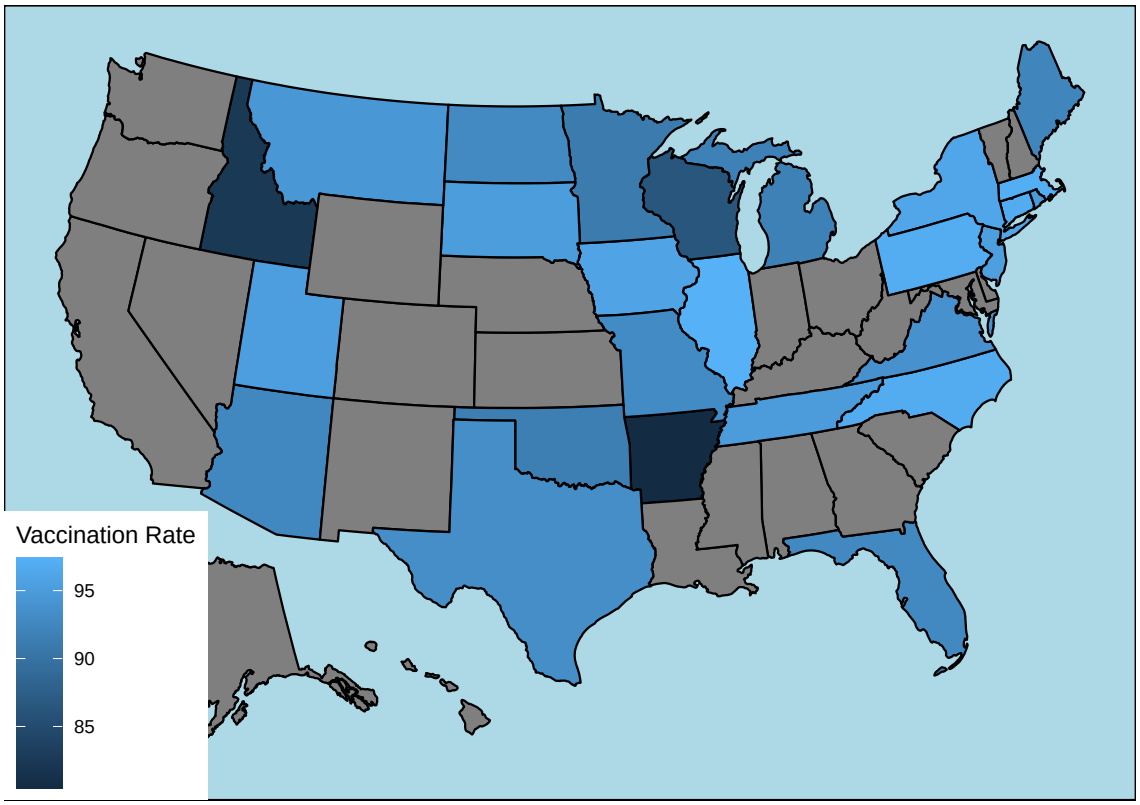


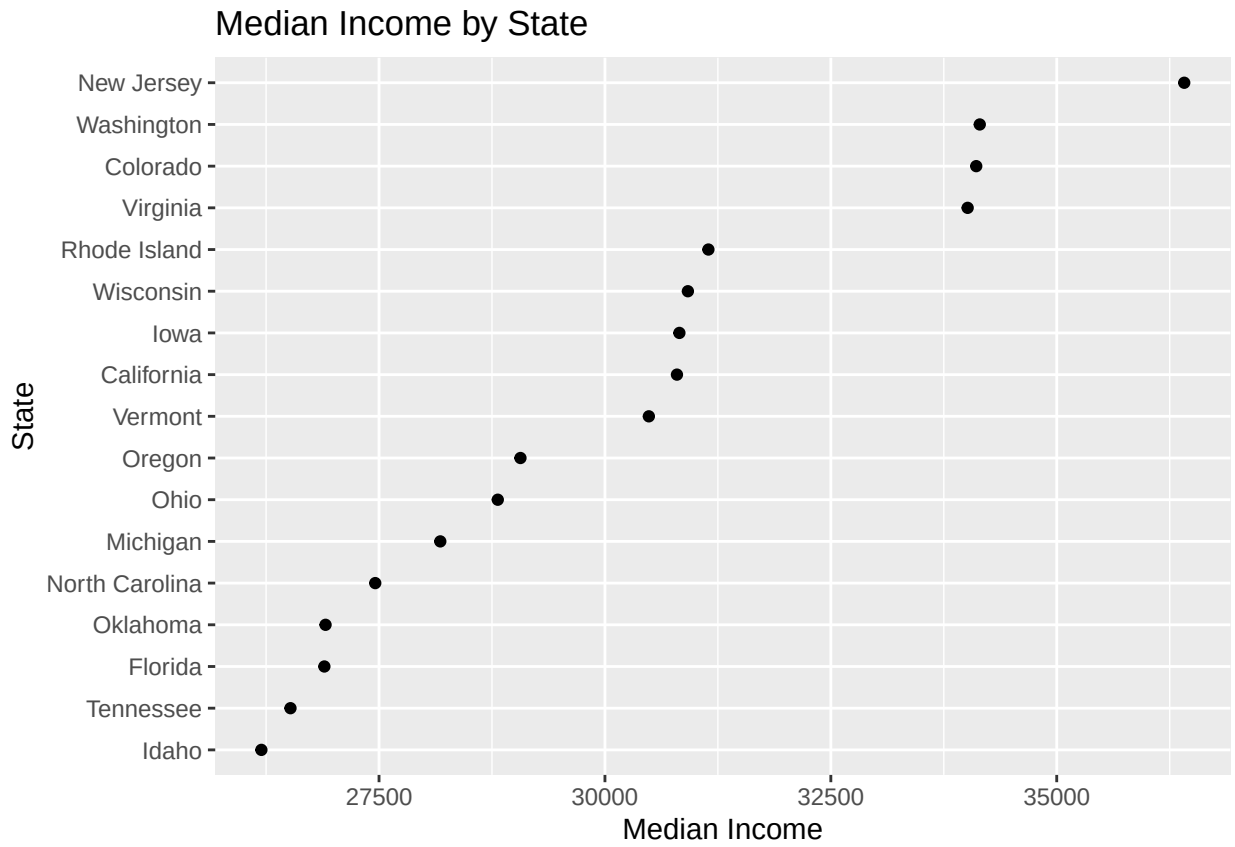

```
measles %>%
  ggplot(aes(x = statemean, y = reorder(state, statemean))) +
  geom_point() +
  labs(x = "Vaccination Rate", y = "State", title = "Vaccination Rate by State")
```




```
plotdata <- measles %>%
filter(realrate != (-1)) %>%
group_by(state) %>%
summarise(statemean = mean(realrate))
plot_usmap(data=plotdata, values = "statemean") +
  labs(title = "Vaccination Rate by State", fill = "Vaccination Rate") +
  theme(panel.background = element_rect(color = "black", fill = "lightblue"))
```

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

```
measlereg <- glm(cbind(numvaxx, unvaxx) ~ statefac, data=measles, family = binomial)
measlereg
```

```
##
## Call:  glm(formula = cbind(numvaxx, unvaxx) ~ statefac, family = binomial,
##     data = measles)
##
## Coefficients:
##            (Intercept)        statefacArkansas         statefacFlorida
##               2.619172              -1.214998                0.002825
##       statefacIllinois            statefacIowa           statefacMaine
##               1.122926               0.587924                0.091579
##       statefacMichigan       statefacMinnesota         statefacMontana
##              -0.047001              -0.002653               -0.259731
##     statefacNew Jersey  statefacNorth Carolina    statefacNorth Dakota
##               0.670881               0.716962                0.062234
##   statefacPennsylvania    statefacRhode Island    statefacSouth Dakota
##               0.900324               0.511840                0.790630
##      statefacTennessee            statefacUtah        statefacVirginia
##               0.304427               0.476347                0.024633
##
## Degrees of Freedom: 28126 Total (i.e. Null);  28109 Residual
##   (11347 observations deleted due to missingness)
## Null Deviance:       316100
## Residual Deviance: 177000    AIC: 255100
```

```
summary(measlereg)
```

```
##
## 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)              2.619172   0.013599 192.594  < 2e-16 ***
## statefacArkansas        -1.214998   0.014364 -84.584  < 2e-16 ***
## statefacFlorida          0.002825   0.015823   0.179  0.85828
## statefacIllinois         1.122926   0.014195  79.107  < 2e-16 ***
## statefacIowa             0.587924   0.015962  36.833  < 2e-16 ***
## statefacMaine            0.091579   0.029166   3.140  0.00169 **
## statefacMichigan        -0.047001   0.017099  -2.749  0.00598 **
## statefacMinnesota       -0.002653   0.018816  -0.141  0.88786
## statefacMontana         -0.259731   0.016949 -15.324  < 2e-16 ***
## statefacNew Jersey       0.670881   0.021321  31.466  < 2e-16 ***
## statefacNorth Carolina   0.716962   0.020218  35.461  < 2e-16 ***
## statefacNorth Dakota     0.062234   0.036011   1.728  0.08395 .
## statefacPennsylvania     0.900324   0.021047  42.776  < 2e-16 ***
## statefacRhode Island     0.511840   0.049686  10.302  < 2e-16 ***
## statefacSouth Dakota     0.790630   0.052304  15.116  < 2e-16 ***
## statefacTennessee        0.304427   0.020575  14.796  < 2e-16 ***
## statefacUtah             0.476347   0.016331  29.169  < 2e-16 ***
## statefacVirginia         0.024633   0.018849   1.307  0.19125
## ---
## 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
```

```
# linear_reg() %>%
# set_engine("lm") %>%
# fit(statemean ~ estimate, data = measleincome) %>%
# tidy()
```