Final Report

due November 16, 2021 by 11:59 PM

Lindsey Weyant, Ali Raich, Aden Clemente

11/16/21

#Load Data

```
census_api_key("abc8289fa2ba274ced76d97c7f8ee31666a2c931", overwrite = TRUE, install = TRUE)
```

```
## 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"
```

```
if (FALSE) {
census_api_key("abc8289fa2ba274ced76d97c7f8ee31666a2c931", install = TRUE)
# First time, reload your environment so you can use the key without restarting R.
readRenviron("~/.Renviron")
# You can check it with:
Sys.getenv("CENSUS_API_KEY")
}
```

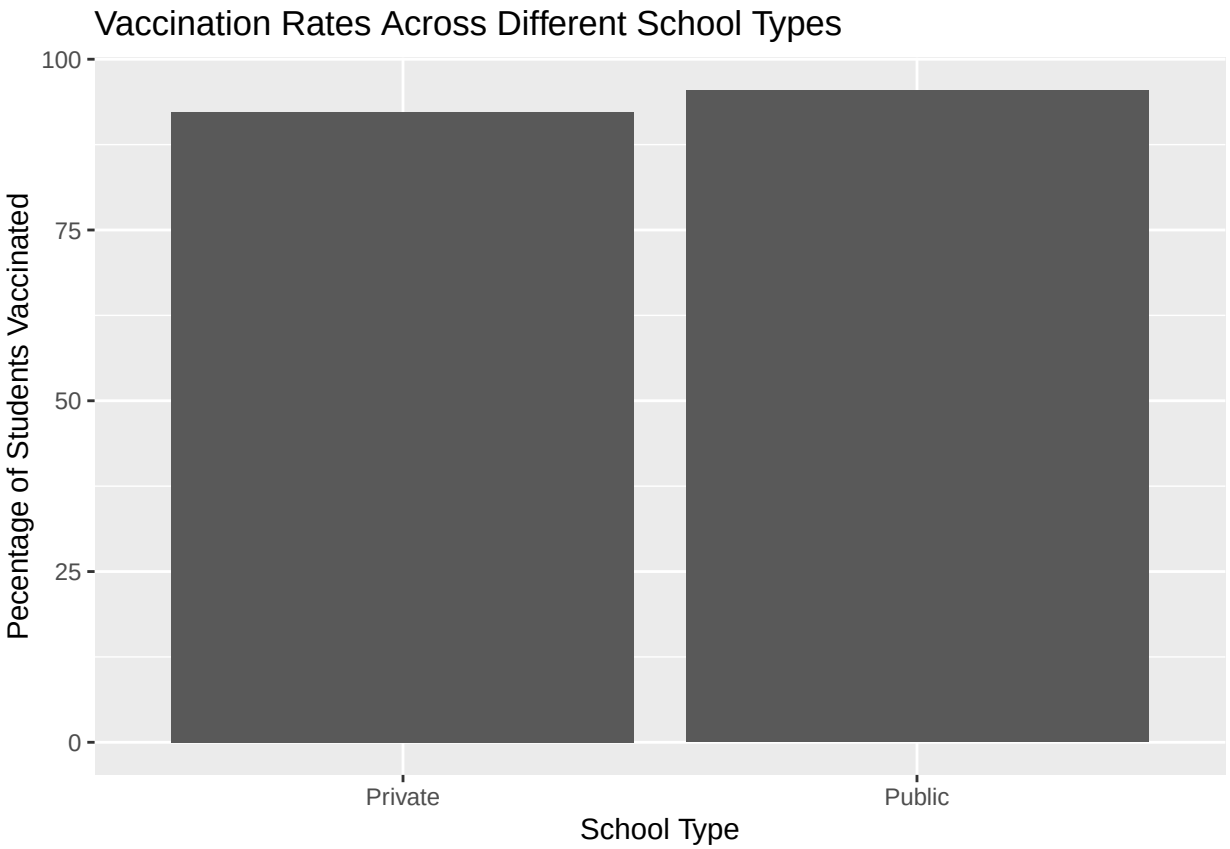
```
#v18 <- load_variables(2018, "acs5", cache = TRUE)

#View(v18)
```

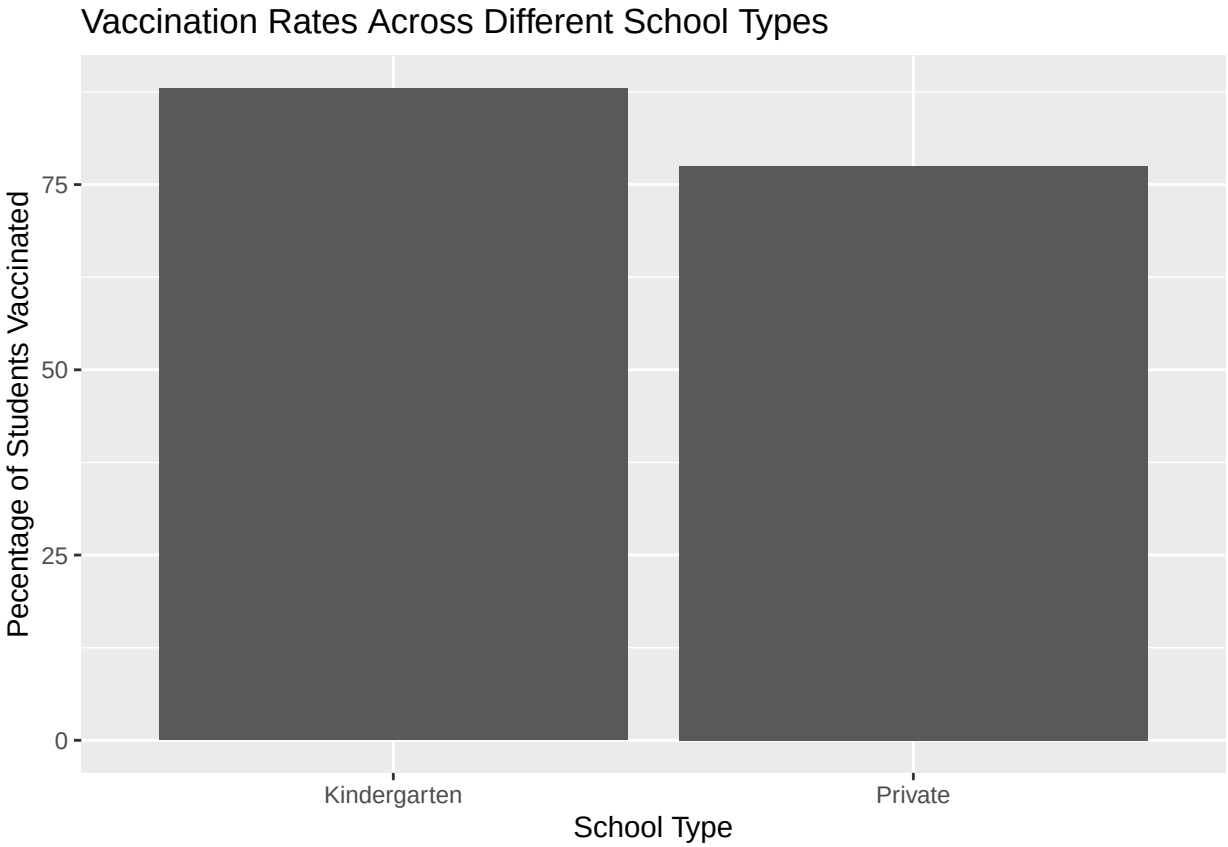
#Research Question:

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

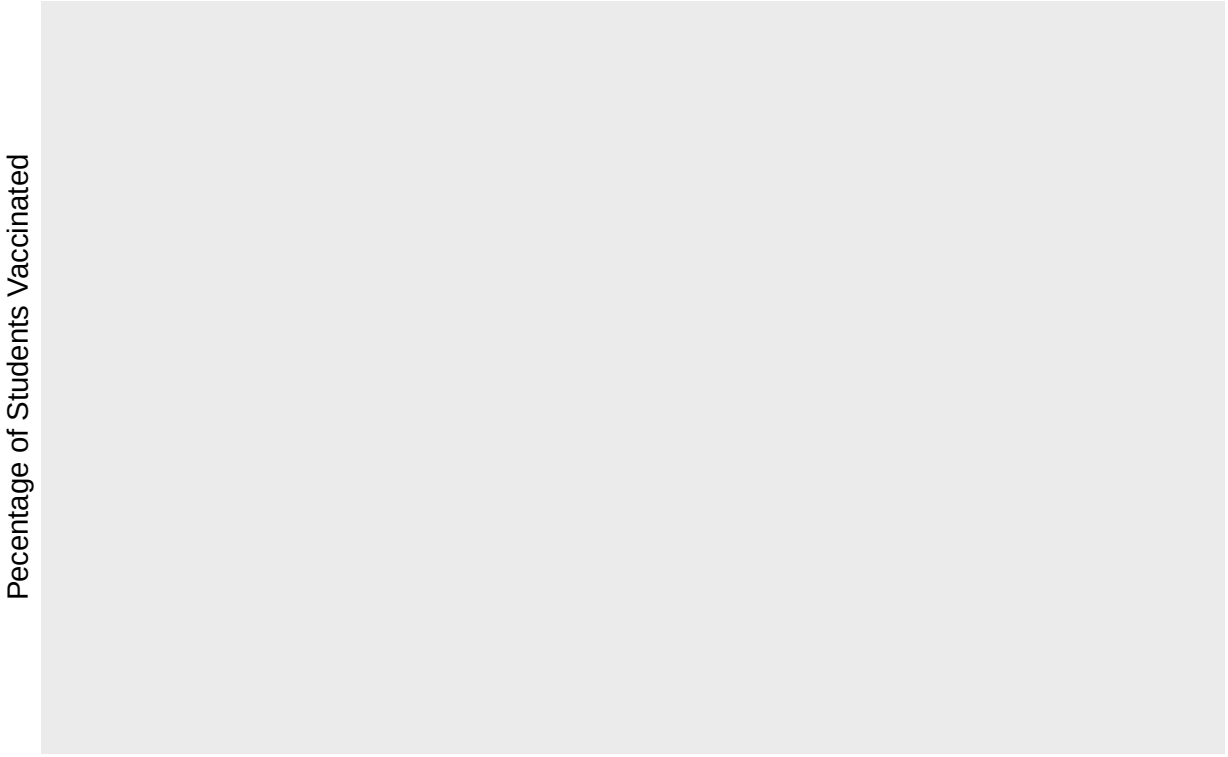
overall vaccination status vs. state, overall vaccination status vs. type of school, each type of exemption (personal, religious, and medical) vs. state exemption vs. type of school. To analyze vaccination and exemption rates by states, we will use spatial data to show the change in these rates across 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.

#Variable Manipulation


Vaccination Rates Across Different School Types
Pecentage of Students Vaccinated
100
75
50
25
0
Private
Public
School Type

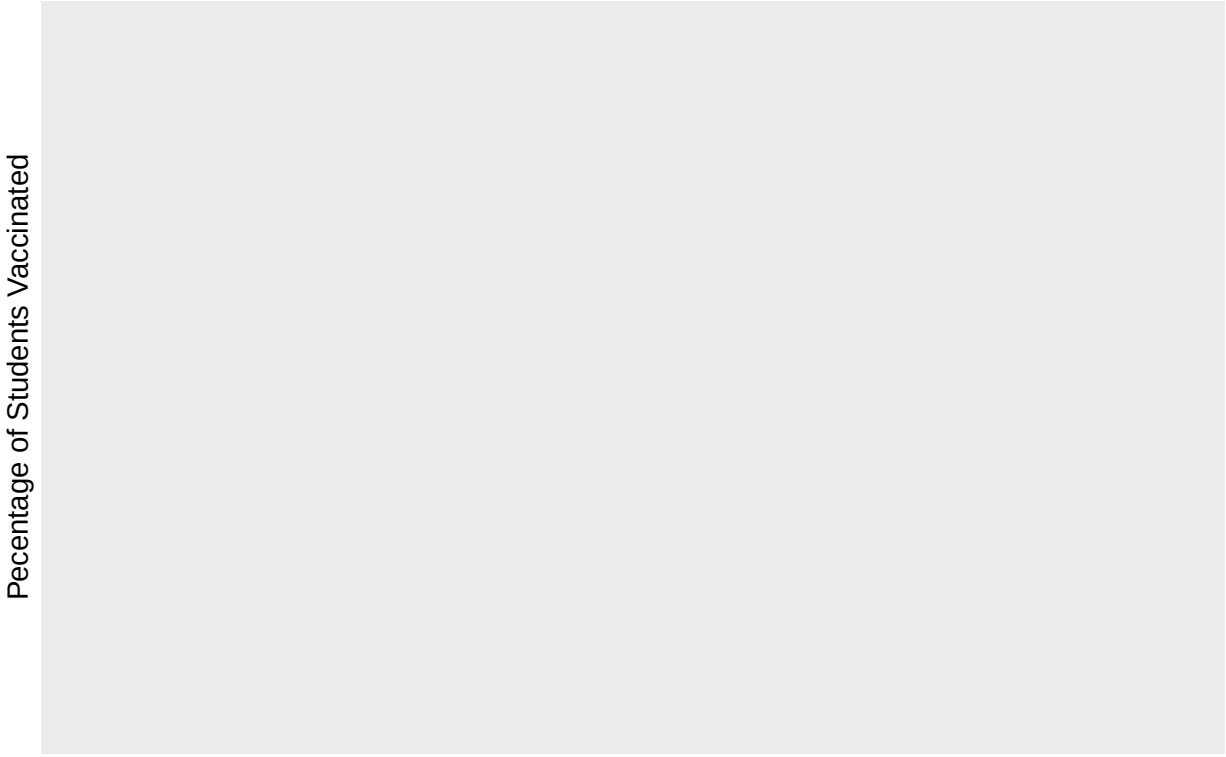

Vaccination Rates Across Different School Types
Pecentage of Students Vaccinated
75
50
25
0
Kindergarten
Private
School Type

Vaccination Rates Across Different School Types



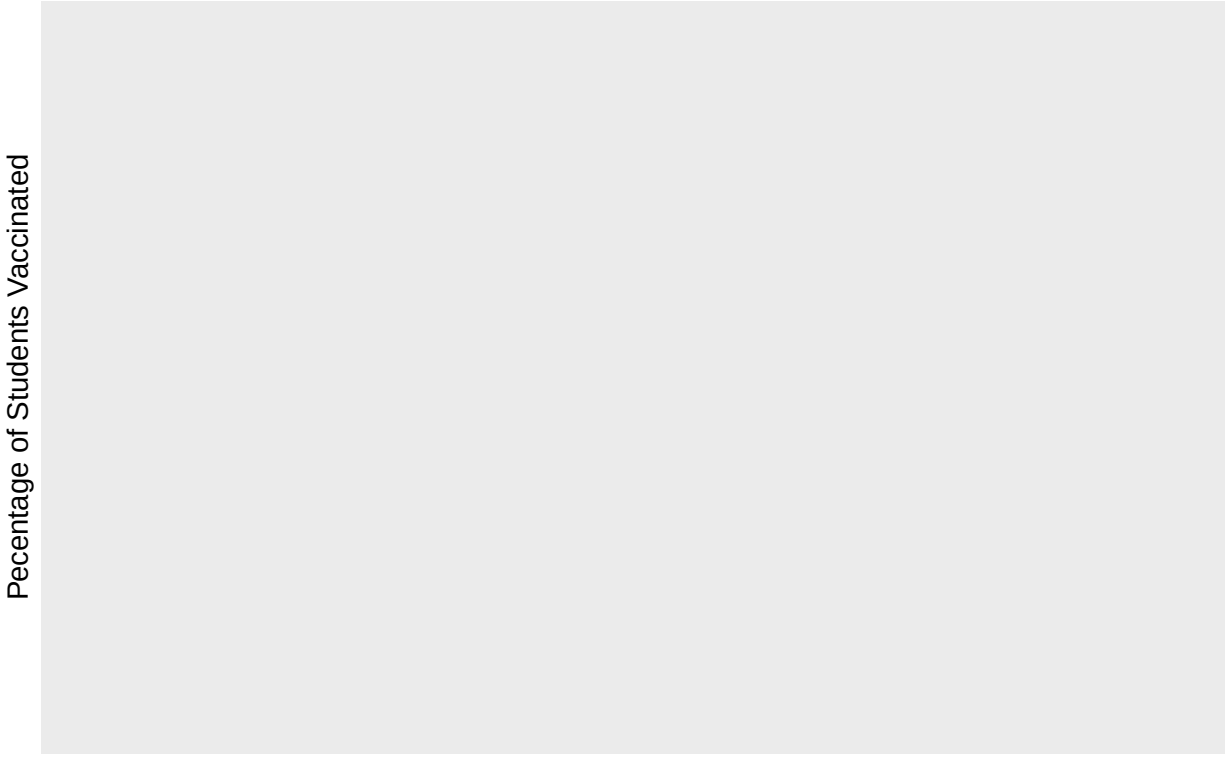

Vaccination Rates Across Different School Types
Pecentage of Students Vaccinated
School Type

Vaccination Rates Across Different School Types

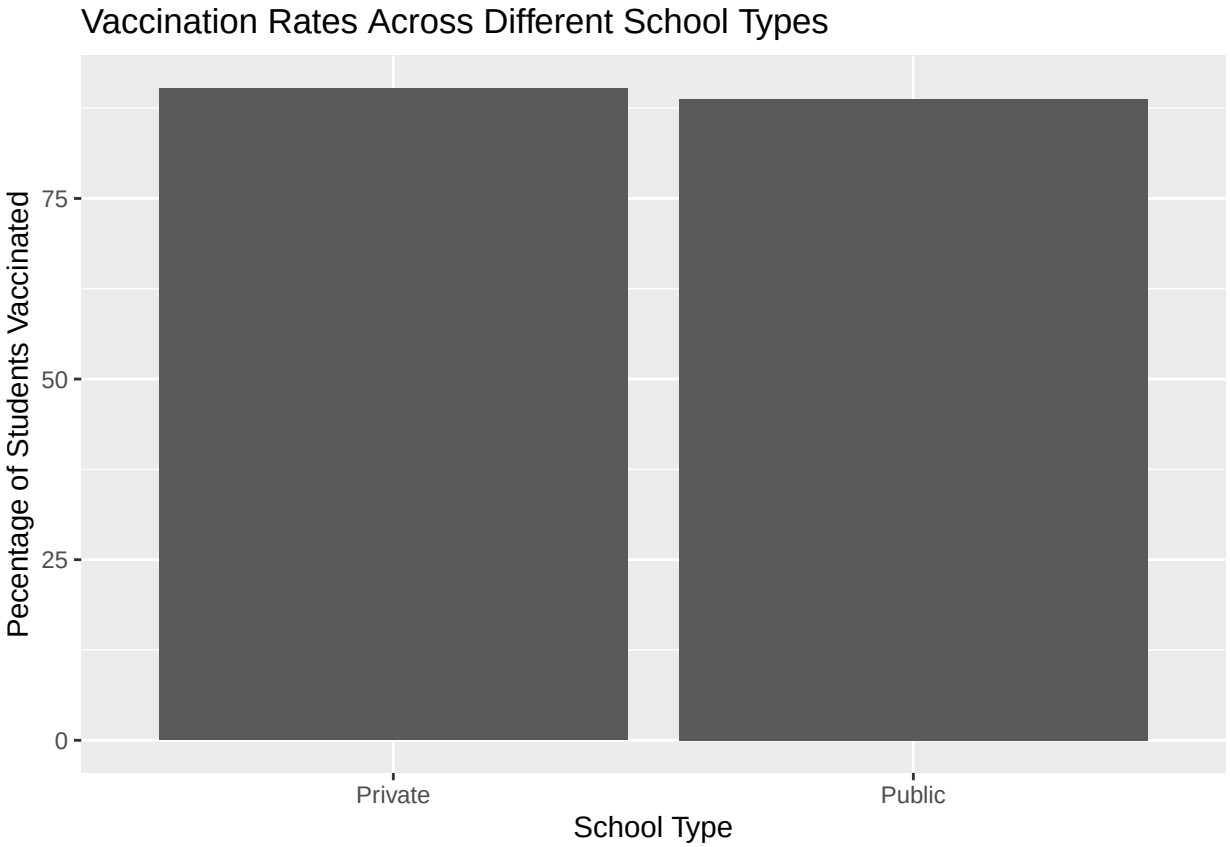



Vaccination Rates Across Different School Types
Pecentage of Students Vaccinated
School Type

Vaccination Rates Across Different School Types



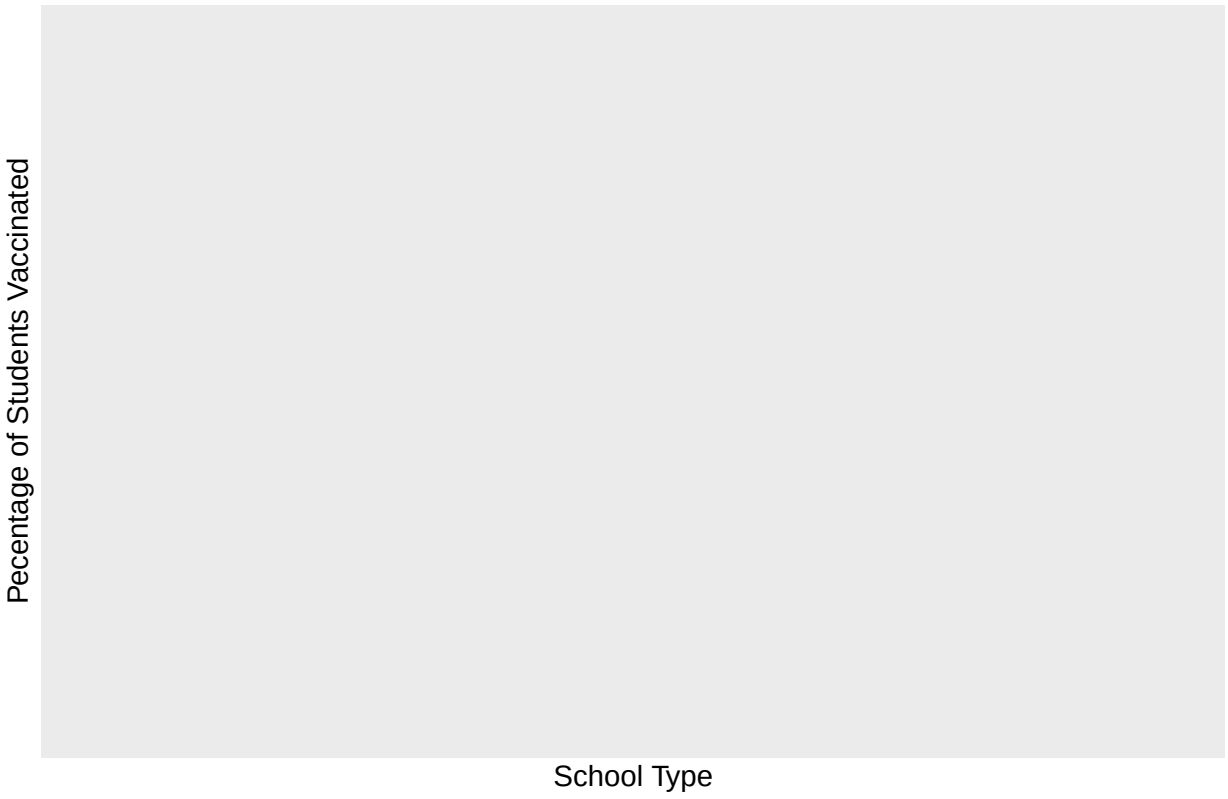

Vaccination Rates Across Different School Types
Pecentage of Students Vaccinated
School Type


Vaccination Rates Across Different School Types
Pecentage of Students Vaccinated
75
50
25
0
Private
Public
School Type


Vaccination Rates Across Different School Types
Pecentage of Students Vaccinated
School Type


Vaccination Rates Across Different School Types
Pecentage of Students Vaccinated
School Type

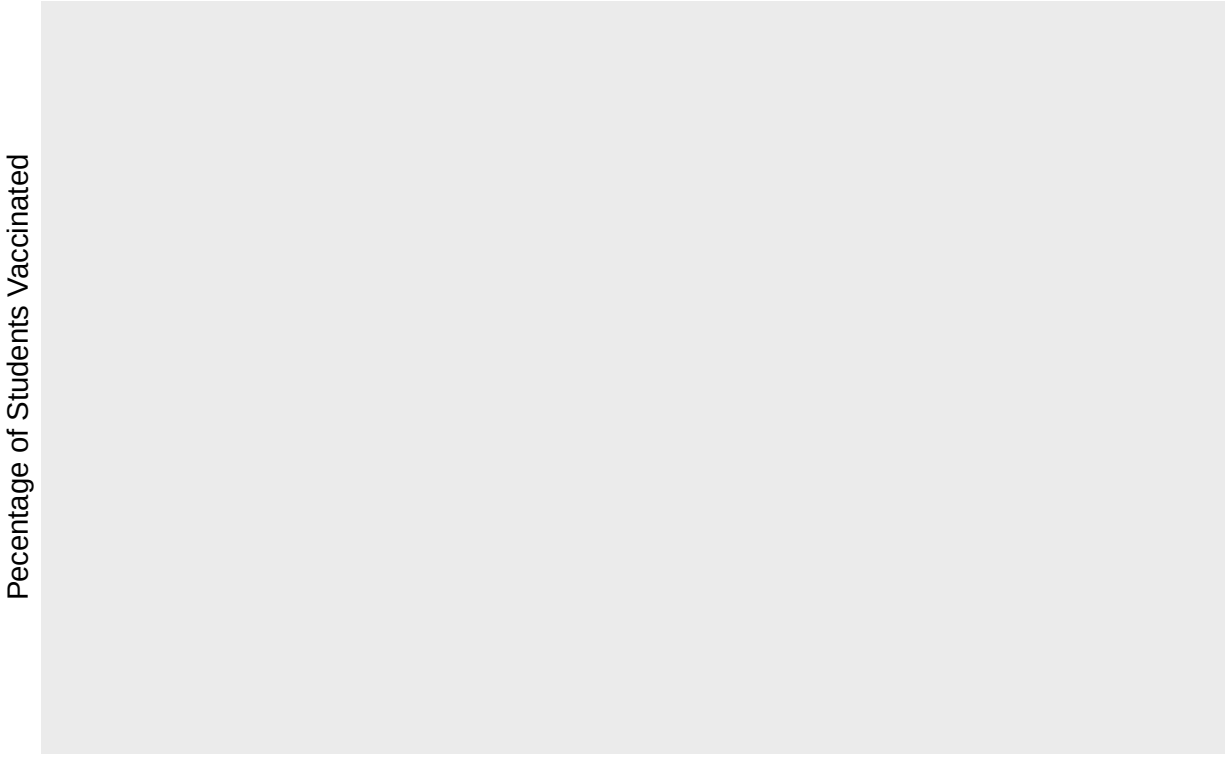

Vaccination Rates Across Different School Types
Pecentage of Students Vaccinated
School Type


Vaccination Rates Across Different School Types
Pecentage of Students Vaccinated
School Type


Vaccination Rates Across Different School Types
Pecentage of Students Vaccinated
School Type

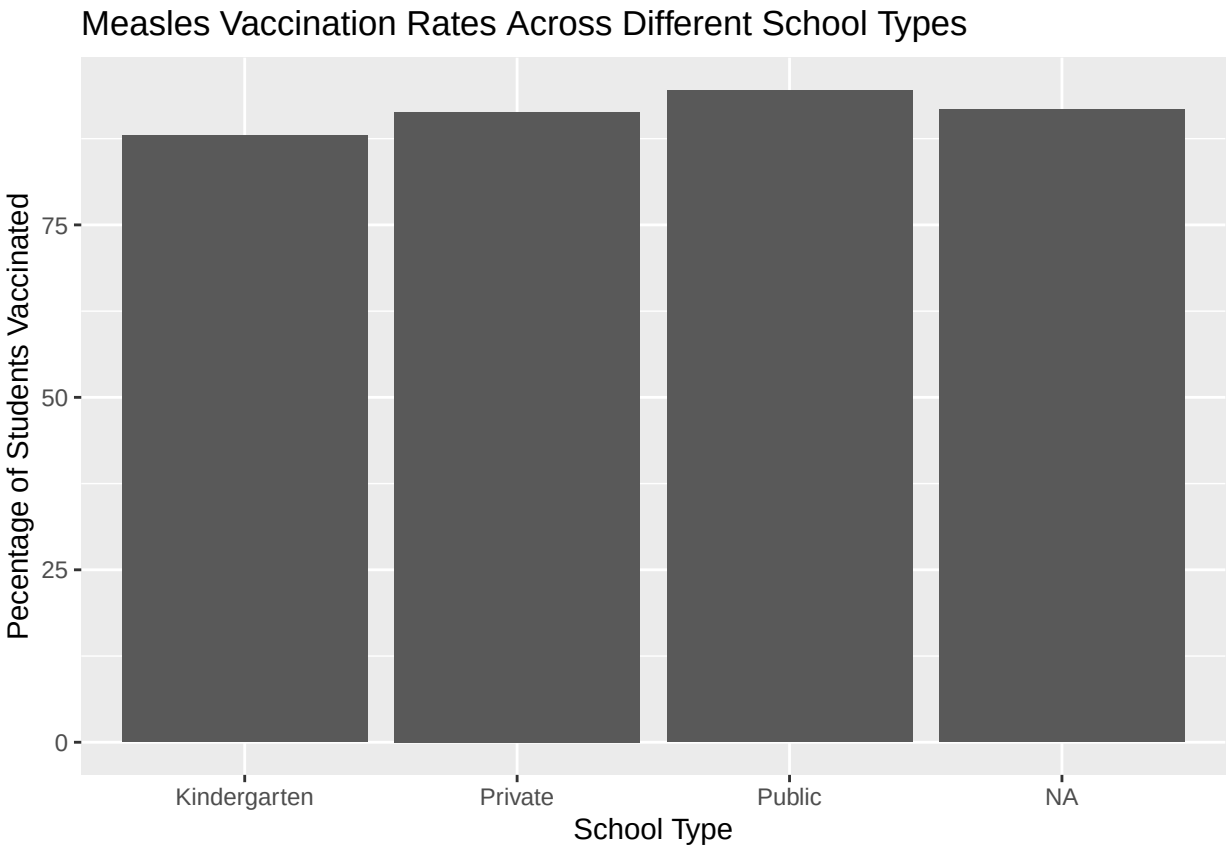

Vaccination Rates Across Different School Types
Pecentage of Students Vaccinated
School Type

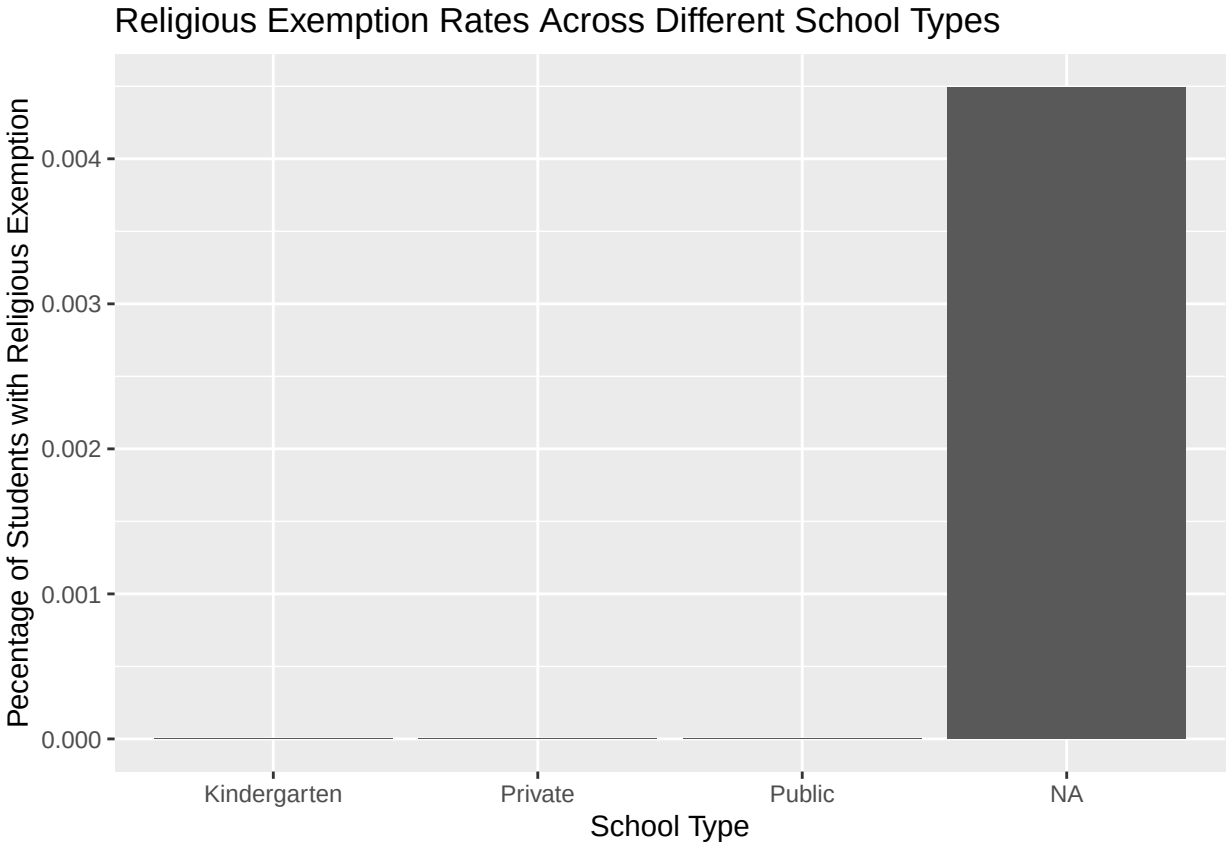
Vaccination Rates Across Different School Types

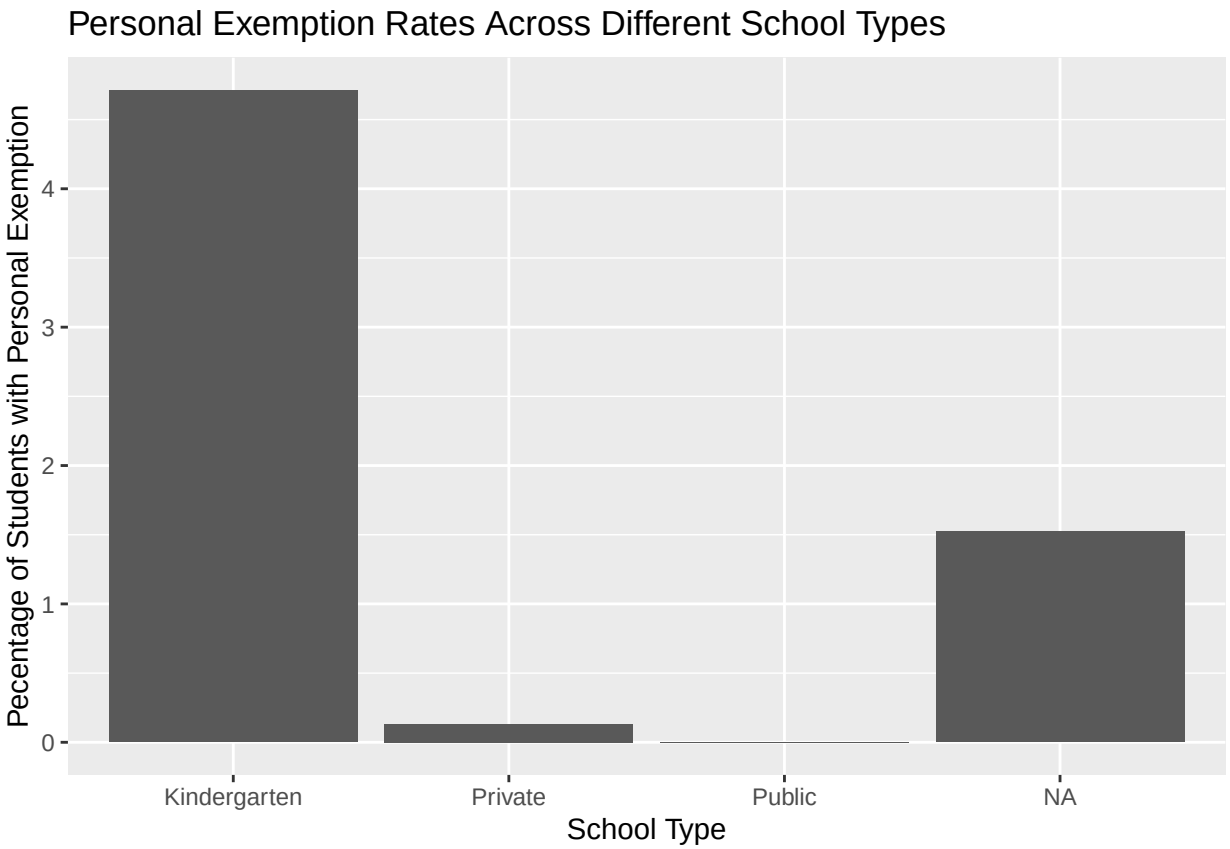


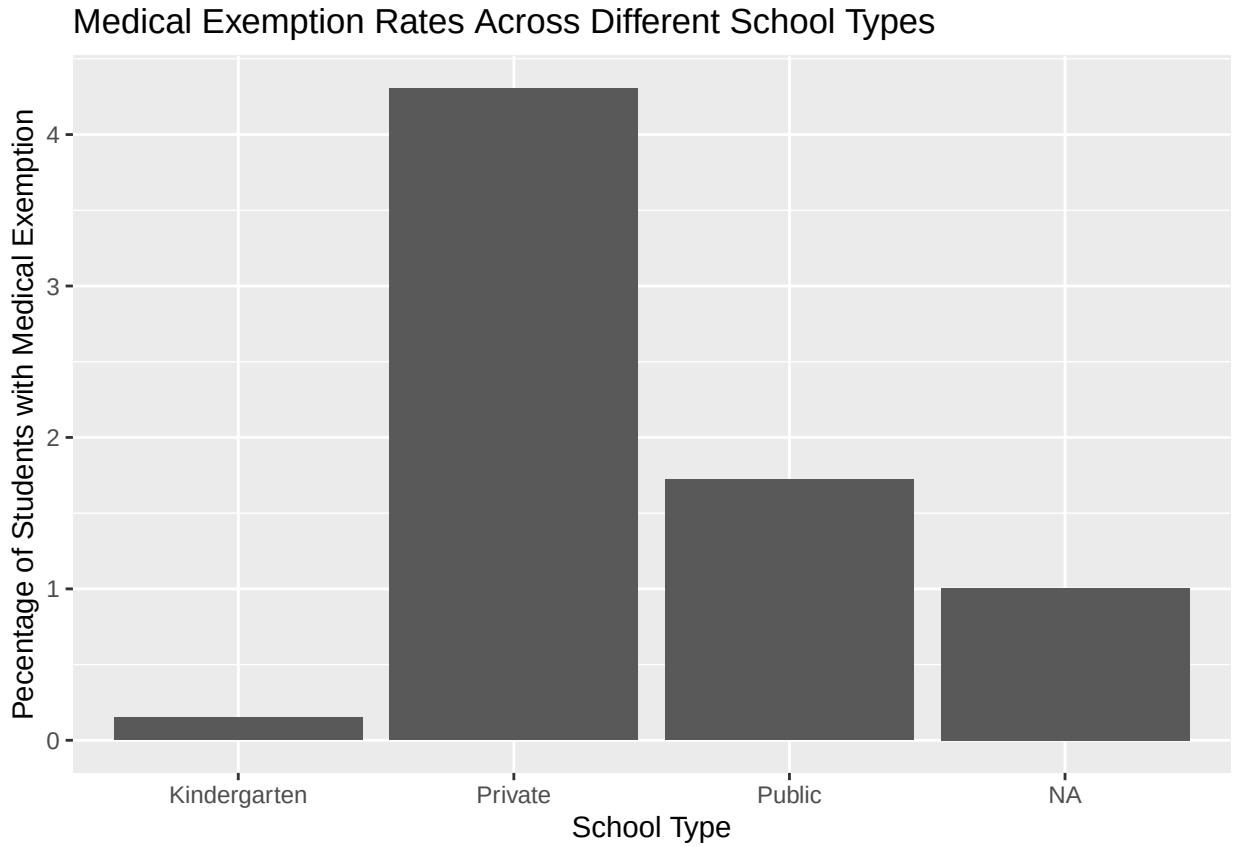

Vaccination Rates Across Different School Types
Pecentage of Students Vaccinated
School Type

Preliminary Bar Graphs






Personal Exemption Rates Across Different School Types
Pecentage of Students with Personal Exemption
0
1
2
3
4
Kindergarten
Private
Public
NA
School Type



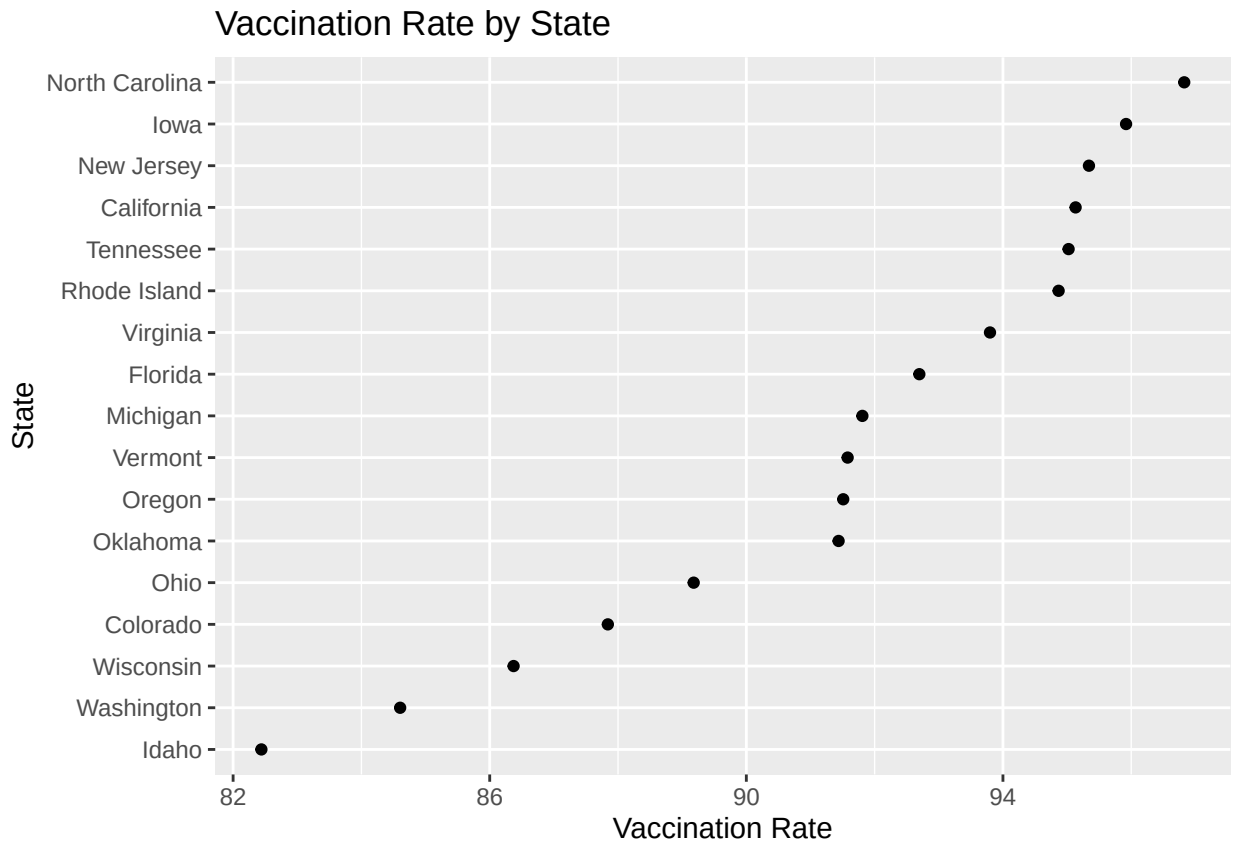
vaccination by state graph

```
measles %>%
  count(state) %>%
  group_by(state)
```

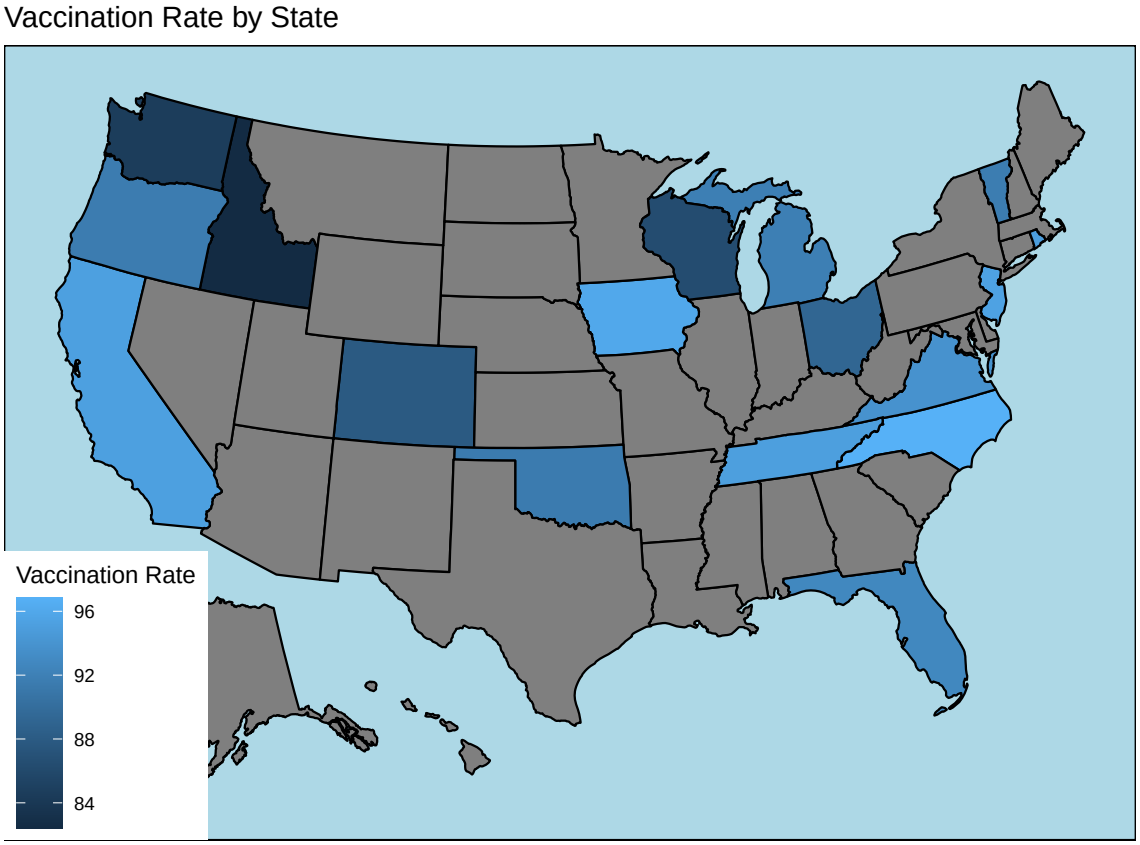
```
## # A tibble: 17 x 2
## # Groups:   state [17]
##    state              n
##    <chr>          <int>
##  1 California     14224
##  2 Colorado        1505
##  3 Florida         2672
##  4 Idaho            467
##  5 Iowa            1163
##  6 Michigan        2351
##  7 New Jersey      2044
##  8 North Carolina  2084
##  9 Ohio            2917
## 10 Oklahoma        1052
## 11 Oregon           806
## 12 Rhode Island     215
## 13 Tennessee       1152
## 14 Vermont          338
## 15 Virginia        1413
```

```
## 16 Washington      1978
## 17 Wisconsin       2508
```

```
measles %>%
  filter(overall != (-1)) %>%
  group_by(state) %>%
  summarise(statemean = mean(overall)) %>%
  ggplot(aes(x = statemean, y = reorder(state, statemean))) +
  geom_point() +
  labs(x = "Vaccination Rate", y = "State", title = "Vaccination Rate by State")
```




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



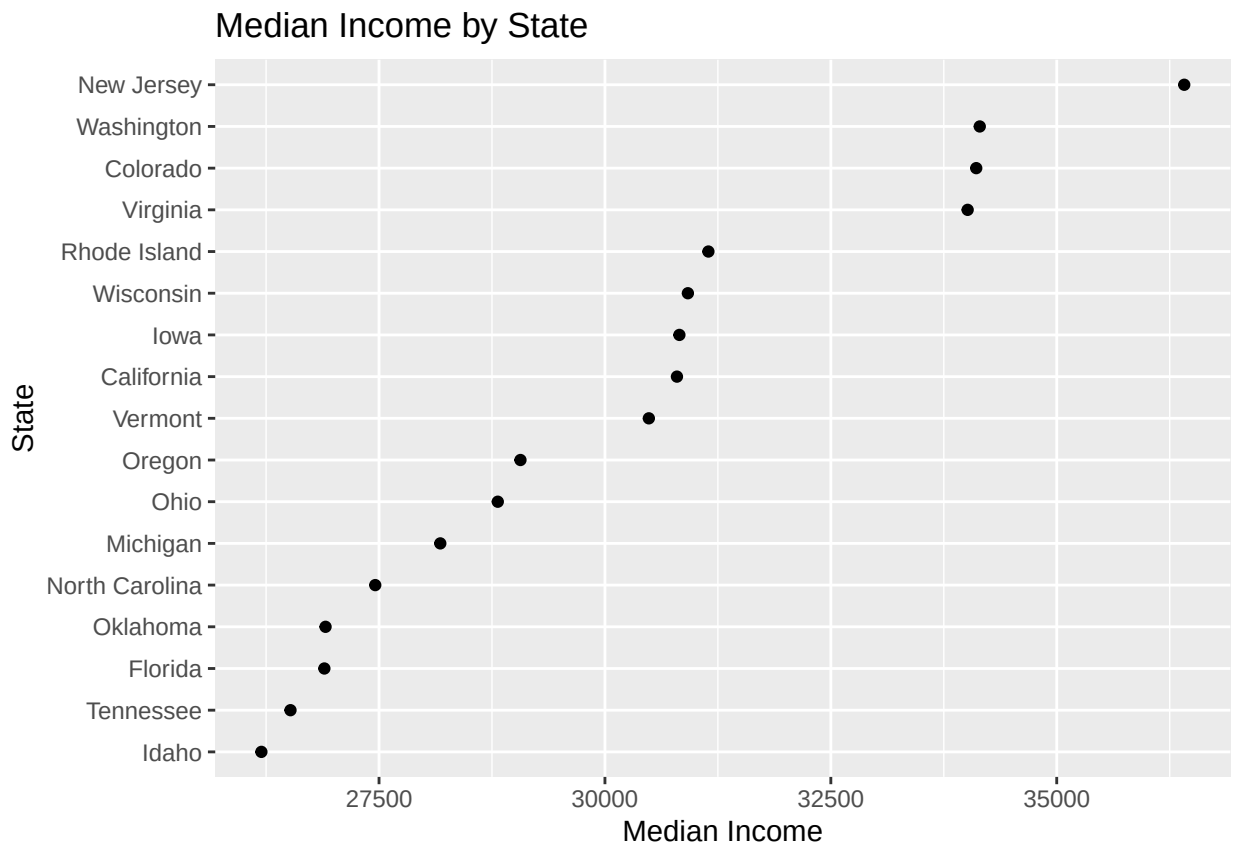

```
income <- get_acs(geography = "state",
                        variables = "B07011_001",
                        year = 2018)
```

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

```
head(income)
```

```
## # 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
```

```
income %>%
  filter(NAME %in% c("California", "Colorado", "Florida", "Idaho", "Iowa", "Michigan", "New Jersey", "Ne
  ggplot(aes(x = estimate, y = reorder(NAME, estimate))) +
  geom_point() +
  scale_y_discrete() +
  labs(x = "Median Income", y = "State", title = "Median Income by State")
```




```
race <- get_acs(geography = "state",
                        variables = "B02001_002",
                        year = 2018)
```

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

```
head(race)
```

```
## # 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
```

```
population <- get_acs(geography = "state",
                        variables = "B01003_001",
                        year = 2018)
```

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

```
head(population)
```

```
## # 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
```

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

Logistic Regression

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

```
##
## Call:  glm(formula = cbind(numvaxx, unvaxx) ~ statefac, family = binomial,
##     data = measles)
##
## Coefficients:
##            (Intercept)        statefacColorado         statefacFlorida
##                3.07058                -1.03239                -0.44858
##           statefacIowa        statefacMichigan      statefacNew Jersey
##                0.13652                -0.49841                 0.21947
## statefacNorth Carolina            statefacOhio          statefacOregon
##                0.26555                -0.86456                -0.51586
##   statefacRhode Island       statefacTennessee         statefacVermont
##                0.06043                -0.14698                -0.32595
##       statefacVirginia
##               -0.42677
##
## Degrees of Freedom: 32745 Total (i.e. Null);  32733 Residual
##   (6112 observations deleted due to missingness)
## Null Deviance:       178900
## Residual Deviance: 160100    AIC: 248300
```

```
summary(measlereg)
```

```
##
## Call:
## glm(formula = cbind(numvaxx, unvaxx) ~ statefac, family = binomial,
##     data = measles)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -59.779   -0.441    0.642    1.395   12.019
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)              3.070581   0.004376 701.752   <2e-16 ***
## statefacColorado        -1.032394   0.011724 -88.059   <2e-16 ***
## statefacFlorida         -0.448583   0.009196 -48.781   <2e-16 ***
## statefacIowa             0.136515   0.009433  14.472   <2e-16 ***
## statefacMichigan        -0.498410   0.011251 -44.300   <2e-16 ***
## statefacNew Jersey       0.219473   0.016993  12.915   <2e-16 ***
## statefacNorth Carolina   0.265553   0.015588  17.036   <2e-16 ***
```

```
## statefacOhio           -0.864555   0.009076 -95.262   <2e-16 ***
## statefacOregon         -0.515862   0.018428 -27.993   <2e-16 ***
## statefacRhode Island    0.060432   0.047988   1.259    0.208
## statefacTennessee      -0.146981   0.016048  -9.159   <2e-16 ***
## statefacVermont        -0.325945   0.018439 -17.677   <2e-16 ***
## statefacVirginia       -0.426775   0.013765 -31.004   <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: 178928  on 32745  degrees of freedom
## Residual deviance: 160130  on 32733  degrees of freedom
##   (6112 observations deleted due to missingness)
## AIC: 248304
##
## Number of Fisher Scoring iterations: 5
```