

Project Proposal

due October 11, 2021 by 11:59 PM

GGPlot Gods – Ali Raich, Aden Clemente, Lindsey Weyant

10/11/21

Load Packages

```
library(tidyverse)
library(sf)
library(viridis)
library(ggspatial)
devtools::install_github("thebioengineer/tidytuesdayR")

tuesdata <- tidytuesdayR::tt_load('2020-02-25')

##
## Downloading file 1 of 1: `measles.csv`

tuesdata <- tidytuesdayR::tt_load(2020, week = 9)

##
## Downloading file 1 of 1: `measles.csv`

measles <- tuesdata$measles
```

Load Data

```
measles <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/02/2020-02-25/measles.csv')
```

Introduction and Data, including Research Questions

We are choosing to study a data set about Measles Vaccination rates in schools across the country. That data set pulls from about 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?” which discusses the increasing rates of unvaccinated people that 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. Thus, our overarching research question is: How does measles vaccination rates vary across the country and demographics in schools? To do this we will look at the reasons why students are exempted from being vaccinated such as religious, medical or personal reasons and examine if there are any trends between the reason and geographic location. Additionally we will examine trends between the type of school, public, private, or charter, and schools overall vaccination rate. For our analysis we will mostly be using the variables state, type of school, vaccination rate, percentage of students with religious exemptions, percentage with personal exemptions, and percentage with medical exemptions. Since this data is a pre-COVID demonstration of vaccination rates, we can use it to help

us tackle the current problem of COVID-19 vaccination. We can look at the vaccination trends of measles to target the geographic areas and types of schools which are less likely to be vaccinated. Furthermore, the rates of different measles vaccine exemptions can shed light on the reasoning behind lagging Covid-19 vaccination rates. We hypothesize that there will be a lower vaccination rate in private and religious schools rather than other types of schools. We also hypothesize that geographically, there will be a higher vaccination rate in states that have historically more progressive or liberal political standings, as well as less of a connection to religion.

Glimpse

(Please use `glimpse` for your data, uploaded into the data folder, [here](#).)

```
measles %>%
  glimpse()

## Rows: 66,113
## Columns: 16
## $ index      <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 10, 11, 12, 13, 14, 15, 15, 16~
## $ state      <chr> "Arizona", "Arizona", "Arizona", "Arizona", "Arizona", "Arizo~
## $ year       <chr> "2018-19", "2018-19", "2018-19", "2018-19", "2018-19", "2018--
## $ name       <chr> "A J Mitchell Elementary", "Academy Del Sol", "Academy Del So~
## $ type       <chr> "Public", "Charter", "Charter", "Charter", "Charter", "Public~
## $ city       <chr> "Nogales", "Tucson", "Tucson", "Phoenix", "Phoenix", "Phoenix~
## $ county     <chr> "Santa Cruz", "Pima", "Pima", "Maricopa", "Maricopa", "Marico~
## $ district   <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ enroll     <dbl> 51, 22, 85, 60, 43, 36, 24, 22, 26, 78, 78, 35, 54, 54, 34, 5~
## $ mmr        <dbl> 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 1~
## $ overall    <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, --
## $ xrel       <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ xmed       <dbl> NA, NA, NA, NA, 2.33, NA, NA, NA, NA, NA, NA, NA, 2.86, NA, 7.41,~
## $ xper       <dbl> NA, NA, NA, NA, 2.33, NA, 4.17, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ lat        <dbl> 31.34782, 32.22192, 32.13049, 33.48545, 33.49562, 33.43532, 3~
## $ lng        <dbl> -110.9380, -110.8961, -111.1170, -112.1306, -112.2247, -112.1~
```

Data Analysis Plan

Overall, our data analysis will seek to answer how vaccination rates and exemptions vary across states and the type of school. To accomplish this, we will examine the relationship between several variable pairs: overall vaccination status vs. state, overall vaccination status vs. type of school, each type of exemption (personal, religious, and medical) vs. state, and each type of 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 t-tests to test for significance of vaccination and exemption rates between different types of schools. We will rely on these methods in the place of linear regression since all of the explanatory variables are categorical. If there are significantly lower vaccination rates in private schools vs. other types of schools, this will support our hypothesis.

```
measles%>%
  mutate(averagevax = mean(mmr)) %>%
  ggplot(aes(x = type, y = averagevax)) +
  geom_col()
```

