

Community Impacts of College Covid Policies

Jordan Striebig and Candace Todd

May 5, 2022

Introduction

COVID-19 is a highly transmissible novel virus that has the potential to put even healthy college students on a ventilator in critical condition. In 2021, COVID-19 vaccines became approved for safe use among the public, and leaders throughout communities around the nation have been trying to spread information to encourage as many people to get vaccinated as possible. Leaders of many colleges and universities have done the same. For the return of students in the Fall 2021 semester, many universities aimed for a fully in-person learning experience. However, with COVID-19 cases still on the rise and many places and limited data on its long-term effects, learning institutions are still taking precautions. Some of these precautions include requiring proof of a COVID-19 vaccination to come back to campus. The student body of college or university can play a very active role in the surrounding communities. With this in mind, we are interested in examining whether a college or university's precautionary COVID-19 policies could have an impact on the COVID-19 cases rates in the school's surrounding county. Our research question is as follows: *In the Fall 2021 semester, did COVID-19 vaccine policies at American colleges and universities have a significant impact on COVID-19 case rates in the school's surrounding county?*

Study Design and Methods

Here we discuss some data processing decisions which impacted our study design. Detailed comments on the data cleaning process are included in the Data Processing section of our Code Appendix.

To gauge COVID-19 case rates at the county level, we found [historical CDC data](#) on county-level estimates of “new-cases per 100,000 persons in the past 7 days” ranging from January 1, 2021, to the present. From this we calculated the arithmetic mean new-cases per capita for each county in the time period from September 1, 2021, up to and excluding December 1, 2021. Throughout the report we will refer to this time period as “fall” and we will refer to the calculated COVID-19 case attribute as “average new-cases per capita”. Average new-cases per capita was the response variable in our study.

We took a random sample of 280 schools from post-processed survey data from the 2019 Integrated Postsecondary Education Data System's (IPEDS) [national survey](#) on the educational offerings of higher education institutions. We decided that our population of interest was 4-year undergraduate-serving U.S. institutions that the IPEDS survey placed in the “Colleges, Universities, and Professional Schools” category. This included schools from any of the 50 states, Puerto Rico, Washington D.C., Guam, the Northern Mariana Islands, or the Virgin Islands (although no schools from the latter four locations made it into the sample). If there were multiple schools in the same county, we only allowed the school with the largest student enrollment to enter the selection pool. We also opted to filter out online schools in light of our research interests.

We collected data on the COVID-19 Vaccine policies and the type of school (whether a school was public or private) with a combination of manual look-up and external data. The vaccine policy attribute was coded as a binary variable in reference to whether a school required all students to get the COVID-19 vaccine for the fall 2021 semester (barring religious or health exemptions). External data for these attributes came from

data on [college vaccine policies](#) from The Chronicle of Higher Education, filtered for schools that announced the vaccine policy by September 1, 2021. We expected that private schools would be more comfortable putting various COVID-19 precautions in place than a public school, which could impact local COVID-19 case rates. To account for this we decided to include school type as a block in our study. We did not stratify our selection pool of schools before randomly sampling, but we did end up with balance across the block in our sample.

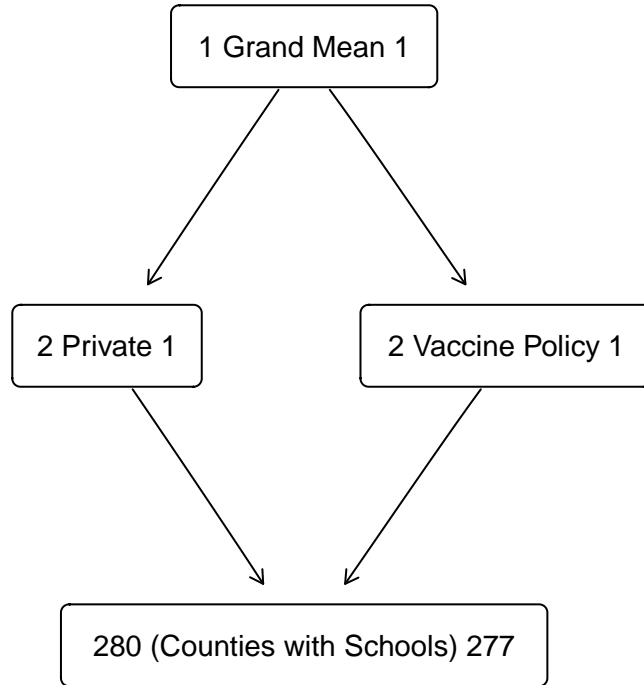


Figure 1: Hasse Diagram for Vaccine Policy Study

In summary, we have a quasi-experiment, thus we can make generalizable correlational statements on the relationship between a county's average new COVID-19 cases per capita in fall 2021 and the COVID-19 vaccine policy of the largest 4-year undergraduate college in the county. We have a continuous numeric response attribute of average new-cases per capita, a binary block of school type, and one binary factor for vaccine policy. ANOVA methods seem to be an appropriate way to address our research question with this data. Figure 1 shows the Hasse diagram for this study. We have positive degrees of freedom everywhere in our Hasse diagram; we have sufficient degrees of freedom to estimate effects and residuals, and we can use an additive model. We will control our Type 1 error risk at 5% and use an unusualness threshold of 5%.

Let y_{ijk} represent the average new-cases per capita for US county k with a college with vaccine policy i and school type j , let $\mu_{...}$ represent the Grand Sample Arithmetic Mean (GSAM) of the average new-cases per capita for US counties with open colleges in the fall of 2021, let β_j represent the effect of whether the largest college in the county is private, let α_i represent the effect of whether the largest college in the county requires vaccines, and let ϵ_{ijk} represent the residual error for for US county k with a college with vaccine policy i and school type j . Then, we will adopt the following hypotheses:

- $H_0 : y_{ijk} = \mu_{...} + \beta_j + \epsilon_{ijk}$. Vaccine requirements for the largest college or university in a county do *not* have a significant impact on average new COVID-19 cases per capita for the college's home counties.
- $H_A : y_{ijk} = \mu_{...} + \beta_j + \alpha_i + \epsilon_{ijk}$ for some $\alpha_i \neq 0$. Vaccine requirements for the largest college or university in a county *do* have a significant impact on average new COVID-19 cases per capita for the

college's home counties.

Exploration of the Data

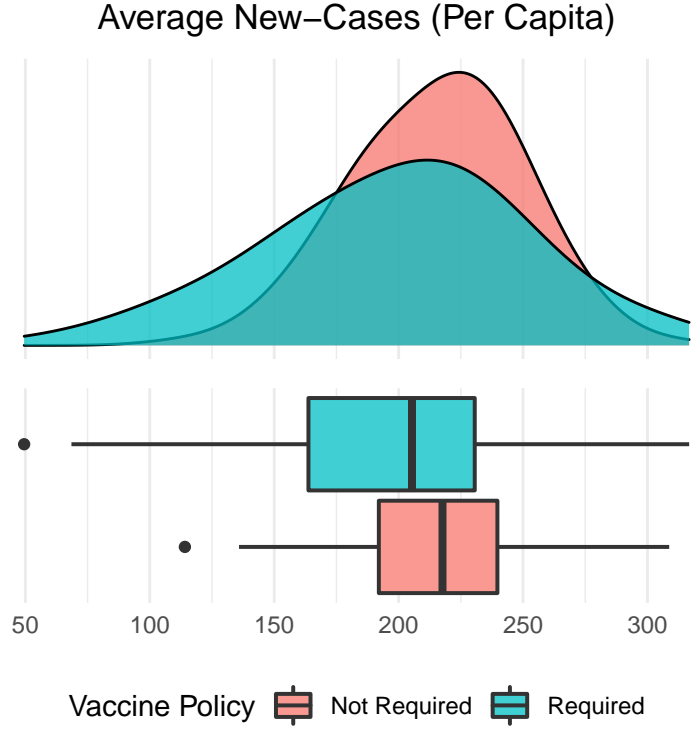


Figure 2: COVID-19 Cases Rates by Vaccine Policy

Figure 2 shows the distribution of average new COVID-19 cases per capita of the counties in our sample grouped by the vaccine policy for the largest college in that county. The distributions are similarly shaped, there is a good bit of overlap here and both distributions are unimodal. As a group, the counties whose largest college did not have a vaccine requirement seem to have had a higher mode (~ 220 average new-cases per capita) than the counties whose largest college did have a vaccine requirement (mode at ~ 205 average new-cases per capita). We ask the readers to keep in mind that these distributions are not COVID-19 case rates plotted over time, as one common visualization that has been used throughout the COVID-19 pandemic often shows. The counties whose largest college did have a vaccine requirement appear to have a wider spread of spread of average new-cases per capita than the other counties.

Table 1: Summary Statistics for Avg. New-Cases Per Capita by Vaccine Policy

| | n | Min | Q1 | Median | Q3 | Max | MAD | SAM | SASD | Sample Skew | Sample Ex. Kurtosis |
|--------------|-----|---------|---------|---------|---------|---------|--------|---------|--------|-------------|---------------------|
| Not Required | 138 | 114.116 | 192.041 | 217.629 | 239.684 | 308.722 | 35.802 | 215.417 | 33.679 | -0.201 | -0.070 |
| Required | 142 | 49.524 | 163.760 | 205.365 | 230.580 | 316.738 | 49.938 | 198.885 | 52.602 | -0.239 | -0.097 |

Table 1 shows the summary statistics for the average new-cases per capita with counties grouped by the vaccine policy of their largest college. This table shows that half of the counties whose largest college had no vaccine requirement had average new-cases per capita greater than ~ 217.63 while half of counties whose largest college did have a vaccine requirement had new-cases per capita lower than ~ 205.37 (per sample

medians). Counties whose largest college did require students to get vaccinated accumulated an average new-case rate of ~ 198.9 per county, while counties whose largest college did not require students to get vaccinated accumulated an average new-case rate of ~ 215.4 per county. The minimum average new-case rate for counties whose largest college did not require students to get vaccinated was much higher at ~ 114.12 compared to ~ 49.52 , the minimum average new-case rate for counties whose largest college did not require students to get vaccinated. On the other hand, the maximum average new-case rate for counties whose largest college did not require students to get vaccinated, ~ 308.72 , is lower than the maximum average new-case rate for counties whose largest college did require students to get vaccinated, ~ 316.74 .

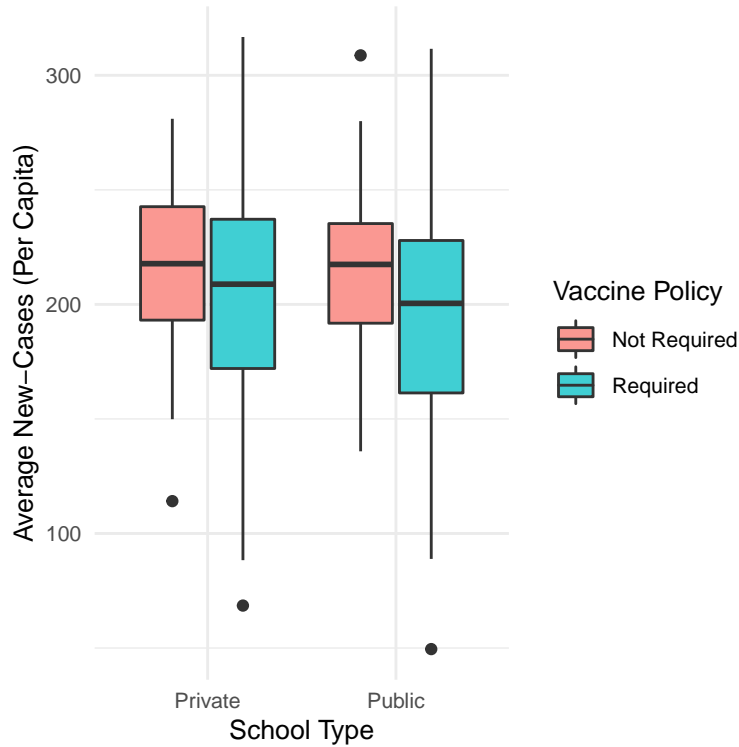


Figure 3: COVID-19 Case Rates by School Type

In the boxplots for private schools in Figure 3, we see again that even within blocks the median average new-cases per capita for counties whose largest college did not require students to get vaccinated is higher than that of counties whose largest college did require students to get vaccinated. We can also see that within both blocks there is a larger spread in average new-cases per capita for counties whose largest college did require students to get vaccinated.

ANOVA

Assumptions

To answer our research question, we will perform a parametric F-test for a 1-factor ANOVA with 1 block. There are four assumptions that our data must satisfy to use this approach: Gaussian residuals, homoscedastic residuals, lack of interaction between the factor and the block, and independence of observations.

To get a better view of exactly how many of our residuals had non-Gaussian behavior, we show a detrended QQ plot rather than a typical QQ plot in Figure 4, equipped with a 90% confidence envelope. The residuals appear approximately Gaussian ; only 10 of the points ($\sim 3.6\%$ of the residuals) fall outside of the

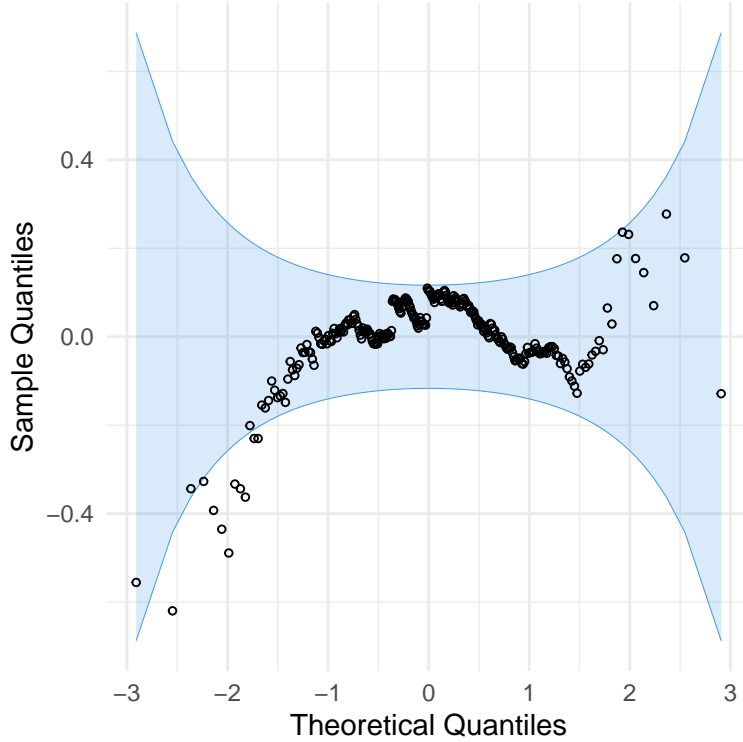


Figure 4: Worm Plot for Vaccine Policy Study

90% confidence envelope for following a Gaussian distribution, which we consider an acceptable amount. Additionally, the residuals have a sample skewness of ~ -0.27 and a sample kurtosis of ~ 0.43 , both of which are close to 0 as they would be for a Gaussian distribution. So, we will act as if the assumption of Gaussian residuals is satisfied.

In Figure 5 we show a Tukey-Anscombe plot for our residuals. The fitted line for this plot is almost perfectly horizontal, and there seems to be no significant difference in the spread of the residuals across different fitted values. We also see no clear patterns in the plot that would make us concerned about a lack of equal variance across the groups. We will act as if the assumption of homoscedastic residuals is satisfied.

The interaction plot in Figure 6 shows the sample arithmetic mean (SAM) for each combinations of School Type and Vaccine Policy present in our study. Although the lines aren't parallel, we see the same trend in Group SAM Average new-cases within both blocks. We will act as if there is no interaction between the Vaccine Policy factor and the School Type block.

To assess whether the independence of observations assumption is met, we will consider the context of the data. In our selection process, we verified that no two schools in our sample came from the same county in order to ensure independence of observations. However, there are instances of multiple counties in the same state. We will keep the potential for spatial dependence in mind but will assume that the independence assumption is satisfied and continue with our study.

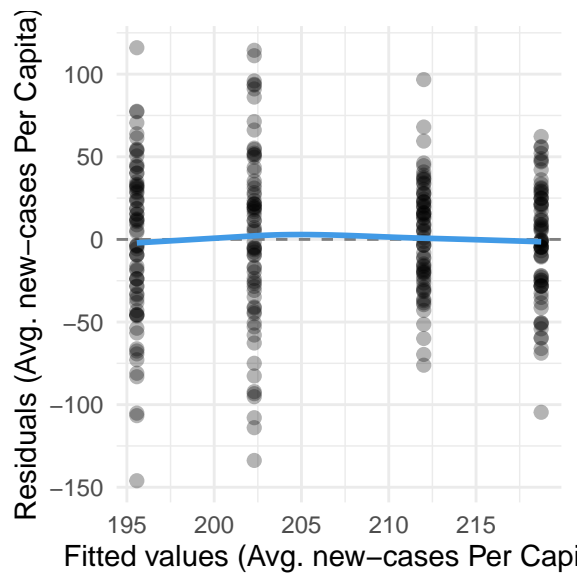


Figure 5: Tukey-Anscombe Plot for Vaccine Policy Study

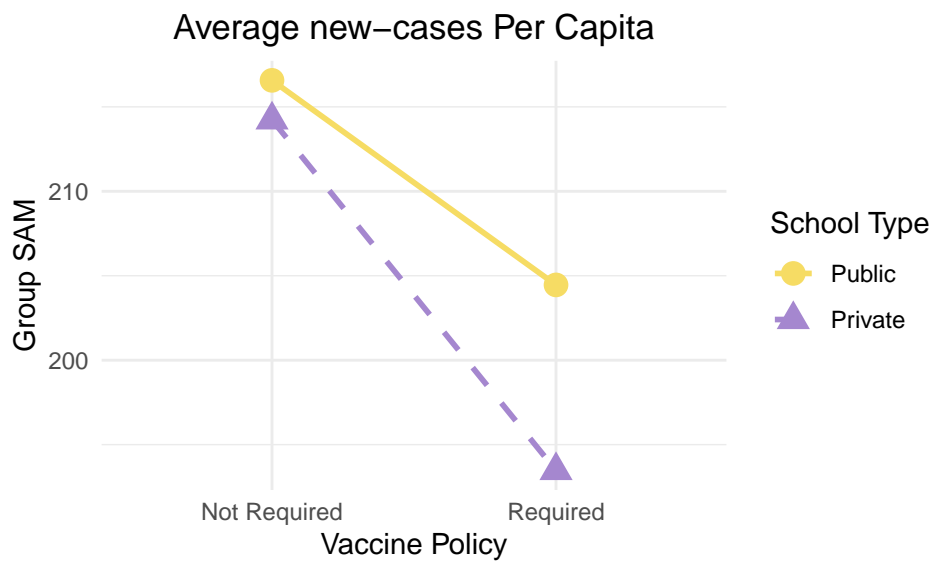


Figure 6: Interaction Plot for School Type and Vaccine Policy

Results

Table 2: ANOVA Table for Vaccine Policy Study

| Source | SS | df | MS | F | p-value | Omega Sq. | Eta Sq. | Epsilon Sq. |
|-----------|------------|-----|-----------|--------|---------|-----------|---------|-------------|
| Type | 3392.672 | 1 | 3392.672 | 1.7327 | 0.1892 | 0.0026 | 0.0062 | 0.0026 |
| Policy | 18901.819 | 1 | 18901.819 | 9.6534 | 0.0021 | 0.0300 | 0.0337 | 0.0302 |
| Residuals | 542376.470 | 277 | 1958.038 | | | | | |

In Table 2 can see here in our ANOVA table that the status of a county’s largest school as either requiring vaccines or not requiring vaccines (“Policy”) accounts for ~ 9.65 times as much variation in average new COVID-19 cases per capita as the residuals in this model. Since our p-value is less than our unusualness threshold of 5% ($\sim 0.0021 < 0.05$), we will reject the null hypothesis and decide to act as if vaccine requirements for the largest college in a county *do* have a significant impact on average new COVID-19 cases per capita for the college’s home county. In particular, vaccine policy for the largest college in a county accounts for a $\sim 3\%$ of the variation average " new covid cases ($\omega^2 = 0.0300$, $\epsilon^2 = 0.0337$, $\eta^2 = 0.0302$).

The relative efficiency of the School Type block is 1.037; we would need about 3.7% more observations per group if we were to conduct this study without the School Type block. Although blocking did not have a big impact on the sample size for our study, introducing the School Type block was helpful for reducing some of the unexplained variation in our model and addressing the Non-Gaussian behavior of some of the residuals.

Table 3: Point Estimates from the Vaccine Policy Study

| | Estimate |
|------------------------|----------|
| GSAM | 207.15 |
| Private School | 3.36 |
| Public School | -3.36 |
| No Vaccine Requirement | 8.22 |
| Vaccine Required | -8.22 |

The sample of counties as a whole accumulated an average new-case rate of about 207.15 per county (GSAM). Counties whose largest school did not have a vaccine requirement accumulated a average new-case rate that was 8.22 higher than the GSAM baseline. Counties whose largest school did not have a vaccine requirement accumulated a average new-case rate that was 8.22 lower than the GSAM baseline. Counties whose largest school did not have a vaccine requirement performed worse than counties whose largest school did have a vaccine requirement (lower average new-case rate is a better performance).

Discussion and Limitations

Since the early days of the COVID-19 pandemic, many colleges and universities have regularly disseminated information on the impact of the COVID-19 virus on the school community and the surrounding municipalities. In an understandably non-uniform manner, each school responded the the state of emergency in a unique manner. Many schools have archives of policies concerning COVID-19 and dashboards of COVID-19 infection and hospitalization rates for the student body and faculty, but there is no consistent formatting across schools nor is there a standard protocol for archiving past data. Individual schools typically have information regarding COVID-19 as easily accessible, but for a study this size (i.e. sample size in the hundreds) it did not seem feasible to manually search for the policies for every school in our study. Thus, we

opted to use a surrogate response from county-level data, which are more consistently formatted because they are provided by the government, instead of direct observations of the attributes we were interested in.

Other studies like this likely suffer from the same issues with finding consistently formatted information, and it is likely that many studies have a hard time getting past an exploratory phase because of the difficulty in finding data. This is unfortunate because there is much value to be found in the research of the impact of the COVID-19 virus in education.

One main limitation for this study was time; with a research team of only two individuals and a large amount of time dedicated to finding and cleaning the data, there was not much room for flexibility in our design and the scope of our study. This is further discussed in the next section.

Scope Reduction

Originally, we wanted to investigate the impact of a school’s vaccine policy and masking policy with respect to the COVID-19 pandemic, with nuanced levels of a vaccine requirement factor and mask requirement factor. However, finding a comprehensive list of the vaccine and masking policies for US schools that were members of our desired population proved difficult. This resulted in a dramatic reduction in the scope of our study in many ways:

The first reduction was in the COVID-19 data we wanted to use as a response. The lack of a consistent standard on COVID-19 dashboards and the processes around archiving past dashboard data made it impossible to automate any selection process; thus, we opted to use county-level data on COVID-19 case rates.

The next reduction in scope impacted the nuance of our factor levels. Comprehensive higher-education data on COVID-19 policies was hard to find at all, and the few lists we did find typically listed binary responses (“Mask Requirement” or “No Mask Requirement”). At this point we reduced the study to the inspection of two binary factors (vaccine requirement and mask requirement) and one binary block (public school versus private school).

The scope of our study was finally reduced to a single-factor study because of our data selection process. We ended up randomly selecting our sample of schools from comprehensive government data courtesy of The Integrated Postsecondary Education Data System’s (IPEDS) surveys. We intended to matching some of the schools to a list of vaccine and mask policies obtained from The Chronicle of Higher Education and manually research the rest. Unfortunately, this resulted in 276 out of the 280 sampled schools having a masking policy (specifically, a policy that required masks indoors for all students). This heavy imbalance led us to decide to drop the mask factor from our study and focus only on schools’ vaccine policies, accounting for the block of school type.

Author Contributions

- Candace cleaned the pre-processed data and contributed to the EDA, Results, and Discussion sections. The code for this document and the analysis contained in it is available in the appendix below, but Candace's project directory for data cleaning can be seen on [GitHub](#) (Currently still being formatted).
- Jordan cleaned the post-processed data and contributed to the Introduction, Study Design, EDA, Results, and Discussion sections.

Code Appendix

```
# Clear workspace
rm(list=ls())

# Setting Document Options
knitr::opts_chunk$set(
  echo = FALSE,
  warning = FALSE,
  message = FALSE,
  fig.align = "center"
)

# Load Packages
packages <- c(
  "tidyverse", "knitr", "kableExtra", "hasseDiagram",
  "psych", "car", "parameters", "ggpubr", "data.table",
  "qqplotr")

lapply(X = packages, FUN = library, character.only = TRUE)

# Loading Helper Files
source("https://raw.githubusercontent.com/neilhatfield/STAT461/master/rScripts/ANOVATools.R")
source("https://raw.githubusercontent.com/neilhatfield/STAT461/master/rScripts/shadowgram.R")

# Setting Global Options
options("contrasts" = c("contr.sum", "contr.poly"))
options(knitr.kable.NA = "")
modelLabels <- c("1 Grand Mean 1", "2 Private 1", "2 Vaccine Policy 1", "280 (Counties with Schools) 27")
modelMatrix <- matrix(
  data = c(FALSE, FALSE, FALSE, FALSE, TRUE, FALSE, FALSE, FALSE, TRUE, FALSE, FALSE, FALSE, TRUE, TRUE),
  nrow = 4,
  ncol = 4,
  byrow = FALSE
)
hasseDiagram::hasse(
  data = modelMatrix,
  labels = modelLabels
)

# Load in the data
covidRaw <-
  readRDS("../Data/finalSampleVerified")

covidRaw$mask <- factor(covidRaw$mask,
  levels=c(0,1),
  labels=c("No Mask Req","Mask Req"))

covidRaw$vaccine <- factor(covidRaw$vaccine,
  levels=c(0,1),
  labels=c("Not Required","Required"))
```

```

covidRaw$private <- as.factor(covidRaw$private)

# This drastic imbalance is what led us to decide to
# reduce the scope of our study
table(covidRaw$vaccine, covidRaw$mask)

# Renaming Variables
setnames(covidRaw,
         old = c("vaccine", "private"),
         new = c("Policy", "Type"))

# Select only the variables we need for One-Way Blocked ANOVA
covid <- covidRaw[,c("avgCasePerCapita", "Policy", "Type")]
# Create Density/Boxplot
densityByVax <-
ggplot(covid, aes(x=avgCasePerCapita, fill=Policy)) +
  geom_density(adjust=1.5, alpha=.75, color="black") +
  ggtitle("Average New-Cases (Per Capita)") +
  ylab("Density") +
  scale_x_continuous(position = "top",
                    breaks=seq(50,320, by=50)) +
  theme_minimal() +
  theme(axis.text.y = element_blank(),
        axis.ticks.y = element_blank(),
        axis.title.y = element_blank(),
        axis.title.x = element_blank(),
        axis.ticks.x = element_blank(),
        axis.text.x = element_blank(),
        panel.grid.major.y = element_blank(),
        panel.grid.minor.y = element_blank(),
        legend.position="none",
        plot.title = element_text(hjust = 0.5))

boxByVax <-
ggplot(covid, mapping=aes(fill=Policy)) +
  geom_boxplot(size=0.7, aes(x=avgCasePerCapita,
                           group=Policy)) +
  theme_minimal() +
  theme(axis.text.y = element_blank(),
        axis.ticks.y = element_blank(),
        panel.grid.major.y = element_blank(),
        panel.grid.minor.y = element_blank(),
        axis.title.x = element_blank(),
        legend.position="bottom") +
  scale_x_continuous(position = "bottom",
                    breaks=seq(50,300, by=50)) +
  scale_fill_manual(name = "Vaccine Policy",
                   labels = c("Not Required", "Required"),
                   values = c("#FA9892", "#40CFD3"))
# Calculate summary statistics by vaccine policy
covidStatsByVax <-
psych::describeBy(
  x = covid$avgCasePerCapita,

```

```

group = c(covid$Policy),
na.rm = TRUE,
skew = TRUE,
ranges = TRUE,
quant = c(0.25, 0.75),
IQR = FALSE,
mat = TRUE,
digits = 4) %>%
tibble::remove_rownames() %>%
tibble::column_to_rownames(var = "group1") %>%
dplyr::select(n,
              min,
              Q0.25,
              median,
              Q0.75,
              max,
              mad,
              mean,
              sd,
              skew,
              kurtosis)
# Show stacked density/boxplot
ggarrange(densityByVax, boxByVax, ncol=1)
# Show summary statistics by Vaccine policy
covidStatsByVax %>%
knitr::kable(
  caption = "Summary Statistics for Avg. New-Cases Per Capita by Vaccine Policy",
  digits = 3,
  format.args = list(big.mark = ","),
  align = rep('c', 11),
  col.names = c("n", "Min", "Q1",
                "Median", "Q3", "Max",
                "MAD", "SAM", "SASD",
                "Sample Skew", "Sample Ex. Kurtosis"),
  booktabs = TRUE) %>%
kableExtra::kable_styling(
  font_size = 12,
  latex_options = c("scale_down", "HOLD_position"))
# Make and show blocked boxplots
ggplot(data = covid,
       mapping = aes(
         x = Type,
         y = avgCasePerCapita,
         fill = Policy)) +
geom_boxplot() +
theme_minimal() +
xlab("School Type") +
ylab("Average New-Cases (Per Capita)") +
labs(fill = "Vaccine Policy") +
theme(legend.position = "right",
      text = element_text(size = 10)) +
scale_fill_manual(name = "Vaccine Policy",
                  labels = c("Not Required", "Required"),

```

```

        values = c("#FA9892", "#40CFD3"))
# Create the ANOVA model
covidAOV <-
  aov(formula= avgCasePerCapita ~ Type + Policy,
      data = covid,
      na.action = "na.omit")
# Create the ANOVA table
anovaTable <-
  parameters::model_parameters(model = covidAOV,
                              omega_squared = "partial",
                              eta_squared = "partial",
                              epsilon_squared = "partial") %>%
  knitr::kable(digits = 4,
              col.names = c("Source", "SS", "df",
                          "MS", "F", "p-value",
                          "Omega Sq.",
                          "Eta Sq.",
                          "Epsilon Sq."),
              caption = "ANOVA Table for Vaccine Policy Study",
              align = c('l', rep('c', 8)),
              booktab = TRUE) %>%
  kableExtra::kable_styling(bootstrap_options = c("striped",
                                                  "condensed"),
                          font_size = 10,
                          latex_options = c("scale_down",
                                              "HOLD_position"))

# Create Point estimates table
pointEst <- dummy.coef(covidAOV)
pointEst <- unlist(pointEst)
names(pointEst) <- c("GSAM", "Private School",
                  "Public School", "No Vaccine Requirement",
                  "Vaccine Required")

pointEstimatesTable <-
  data.frame("Estimate" = pointEst) %>%
  knitr::kable(
    digits = 2,
    caption = "Point Estimates from the Vaccine Policy Study",
    booktabs = TRUE,
    align = "c"
  ) %>%
  kableExtra::kable_styling(
    font_size = 12,
    latex_options = c("HOLD_position"))
# Create and show wormplot
ggplot(mapping = aes(sample=scale(residuals(covidAOV)))) +
  stat_qq_band(distribution="norm",
              col="#419AE6",
              fill="#419AE6",
              alpha=0.2,
              detrend=TRUE,
              conf=0.9) +
  xlab("Theoretical Quantiles") +

```

```

ylab("Sample Quantiles") +
stat_qq_point(size=1,
              shape=1,
              detrend=TRUE) +
theme_minimal()

ggplot(data =
       data.frame(residuals = residuals(covidAOV),
                  fitted = fitted.values(covidAOV)),
       mapping = aes(x = fitted, y = residuals)) +
geom_point(size = 2, alpha=0.3) +
geom_hline(yintercept = 0,
           linetype = "dashed",
           color = "grey50") +
geom_smooth(formula = y ~ x,
            method = stats::loess,
            method.args = list(degree = 1),
            se = FALSE,
            size = 1,
            color="#419AE6") +
theme_minimal() +
xlab("Fitted values (Avg. new-cases Per Capita)") +
ylab("Residuals (Avg. new-cases Per Capita)")

interactionDf <-
  covid %>%
  group_by(Policy, Type) %>%
  summarise(groupMean = mean(avgCasePerCapita))

interactionplot <-
ggplot2::ggplot(data = interactionDf,
               mapping = aes(
                 x = Policy,
                 y = groupMean,
                 shape = Type,
                 color=Type,
                 linetype=Type))+
geom_point(size=4) +
geom_line(mapping=aes(x=Policy,
                     y=groupMean,
                     group=Type),
          size=1) +
theme(plot.title = element_text(hjust = 0.5)) +
# Manally set the legend
scale_color_manual(name = "School Type",
                  labels = c("Public", "Private"),
                  values = c("#F6DC64", "#A587CF")) +
scale_shape_manual(name = "School Type",
                  labels = c("Public", "Private"),
                  values = c(16, 17)) +
scale_linetype_manual(name = "School Type",
                    labels = c("Public", "Private"),
                    values = c(1, 2)) +

```

```

theme_minimal() +
xlab("Vaccine Policy") +
ylab("Group SAM") +
theme(plot.title = element_text(hjust = 0.5)) +
ggtitle("Average new-cases Per Capita") +
labs(shape = "School Type",
      color="School Type",
      linetype="School Type") +
theme(legend.position = "right")

interactionplot
anovaTable
# Calculate Relative efficiency
block.RelEff(aov.obj = covidAOV,
             blockName = "Type",
             trtName = "Policy")
pointEstimatesTable

```