Community Impacts of College Covid Policies

Jordan Striebig and Candace Todd

May 5, 2022

Introduction

COVID-19 is a novel virus that is highly transmissible and for much of a county's general population, dangerous. COVID-19 has the potential to put even healthy college students on a ventilator in critical condition. In 2021, COVID vaccines became approved for safe use among the public. Leaders throughout the community have been trying to spread information to encourage as many people to get vaccinated as possible. In colleges and universities, many of their leaders have done the same. For the return of students in Fall 2021, most universities aimed for a fully in-person learning experience. However, with COVID cases still on the rise and data on its long-term effects limited, learning institutions still had to take precautions. These precautions include requiring proof of a COVID vaccination to come back to campus. Many colleges play a prevalent role in their communities. So, it would make sense that a college's decisions on covid precautionary policies could have an impact on the county's COVID cases. Our research question is as follows: How did Fall 2021 COVID-19 policies at American universities relate to the arithmetic mean of daily new COVID-19 cases/capita among the universities' home counties in the time-frame of September 1st to but not including December 1st?

We were going to also include mask policies in our analysis, however, when we finished gathering our data, the factors of mask requirements and vaccine requirements were extremely unbalanced. Out of our sample of 280 schools, we only found 4 schools that did not have an indoor mask mandate at any point during their fall semester. Vaccine requirements however were much more varied among different schools so we decided to focus on vaccines instead.

specify 50 states + PR + surrogate response

Study Design & Methods

In our study, we collected data on the COVID policies of different colleges and universities in the United States of America. In the "United States" we also included schools from Puerto Rico as it is a US territory. First, we decided to take our sample from undergraduate-serving 4-year colleges and universities as well as professional schools within a county in the United States. If our sample ended up having multiple schools from the same county, we chose the school that had the larger/largest student body population. In drawing a sample we opted to filter out schools that were primarily online and focused on schools with physical campuses. We drew a total of 280 schools in our sample. We got our covid data for our sample from a dataset provided by the CDC. This dataset included a list of 877 counties in the United States and its corresponding average daily new COVID cases per capita, averaged from September 1st to but not including December 1st. The capita was per 100k people. We then merged together the schools that we selected, its corresponding county, and its corresponding COVID information. To find out what a school's COVID policies were, we initially gathered a smaller list of schools that had vaccine mandates from the Chronicle of Education's dataset. From there, we selected our sample of colleges from a dataset that included all the schools in the US that got through our filter. Then, we gathered our final data on school COVID policies
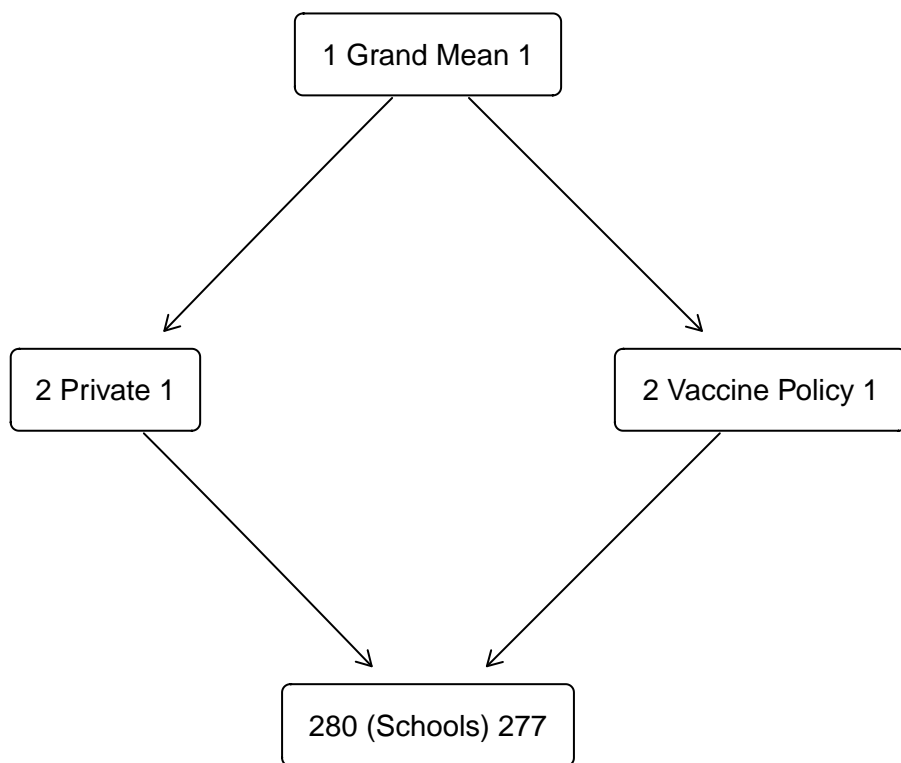
Figure 1: Hasse Diagram

manually by looking up each school and verifying what state it was in, verifying its name, checking the date its vaccine policy was enforced, checking its mask policy, checking if the school was public or private, checking the date its vaccine policy was enforced, and what its vaccine policy actually was. We coded masks on a binary level for 0 being the school never had a mask policy where masks were required indoors for all students from September 1st to and not including December 1st and 1 being that it did. We coded the type of school as being either "private" or "public" and we decided to use this as a block in our study. We know that a school's status as public can put far more limits on it as for vaccination mades than a private school so we wanted to account for that. We then coded vaccine requirements on a binary level as well with 0 being the school never required vaccines for ALL (not including special health/religious exemptions) students at the school BY Sept. 1, 2021 and a 1 if they did.

Our primary response is the average daily new COVID cases per capita from September 1st to December 1st. Our only factor of interest is vaccine requirements which has two levels (0, Vaccines NOT required for students at the school BY Sept. 1, 2021 or 1, vaccines required for ALL [not including special health/religious exemptions] students at the school BY Sept. 1, 2021). We can make causal inference about undergraduate-serving 4-year colleges and universities as well as professional schools within a county in the United States because we took a random sample from this population so we have the required randomization. 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.

Let y_{ijk} represent the a new daily covid cases per capita in a US county, $\mu_{...}$ represent the Grand Sample Arithmetic Mean (GSAM) of the average new daily covid cases , β_j represent the effect of public vs private school block, α_i represent the effect of vaccine policy levels, and ϵ_{ijk} represent the residual error for individual school 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 at a college or university are NOT related to the daily new COVID-19 cases/capita among the universities' home counties.
- $H_A : y_{ijk} = \mu... + \beta_j + \alpha_i + \epsilon_{ijk}$ for some $\alpha_i \neq 0$. Vaccine requirements at a college or university ARE related to the daily new COVID-19 cases/capita among the universities' home counties

We will control our overall Type 1 error risk at 5% and an unusualness threshold of 5%.
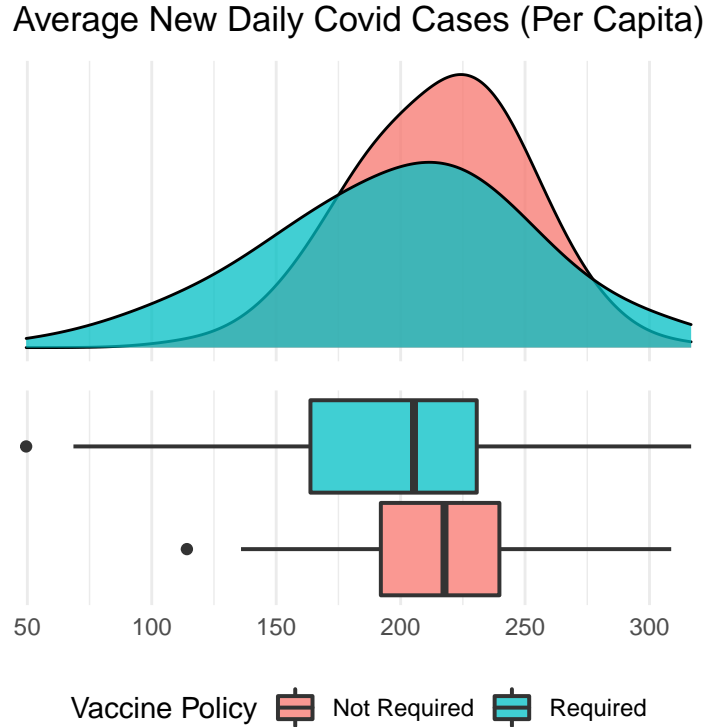
Exploration of the Data

Figure 2: Graph

This graph shows the distribution of average daily covid cases per capita by vaccine requirements. There is a good bit of overlap here and both distributions are unimodal. The group of no vaccine requirement seems to have had a high peak as the vaccine requirement group began to fall down. The peak for no vaccines required as a substantially larger peak than that of the vaccine required group.

Table 1: Summary Statistics for Avg. Daily New Covid Cases 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 daily new covid cases by vaccine policy broken down by no vaccine required and vaccine required. This table shows that the median average daily new covid cases for schools without a vaccine requirement are higher at ~217.63 while schools without a vaccine requirement had a lower median at ~205.37. Our sample arithmetic mean of the average daily new covid cases for schools without a vaccine requirement are significantly higher at ~49.94 while schools with a vaccine had a sample arithmetic mean of ~198.89. The minimum average daily new covid cases for schools without a vaccine requirement are much higher at ~114.12 compared to the average daily new covid cases for schools with a vaccine requirement. Schools with a vaccine requirement had a minimum average daily new covid cases of ~49.52. Surprisingly, the maximum average daily new covid cases for schools without a vaccine requirement is lower at ~308.72 while for schools with a vaccine requirement, the maximum is 316.74.
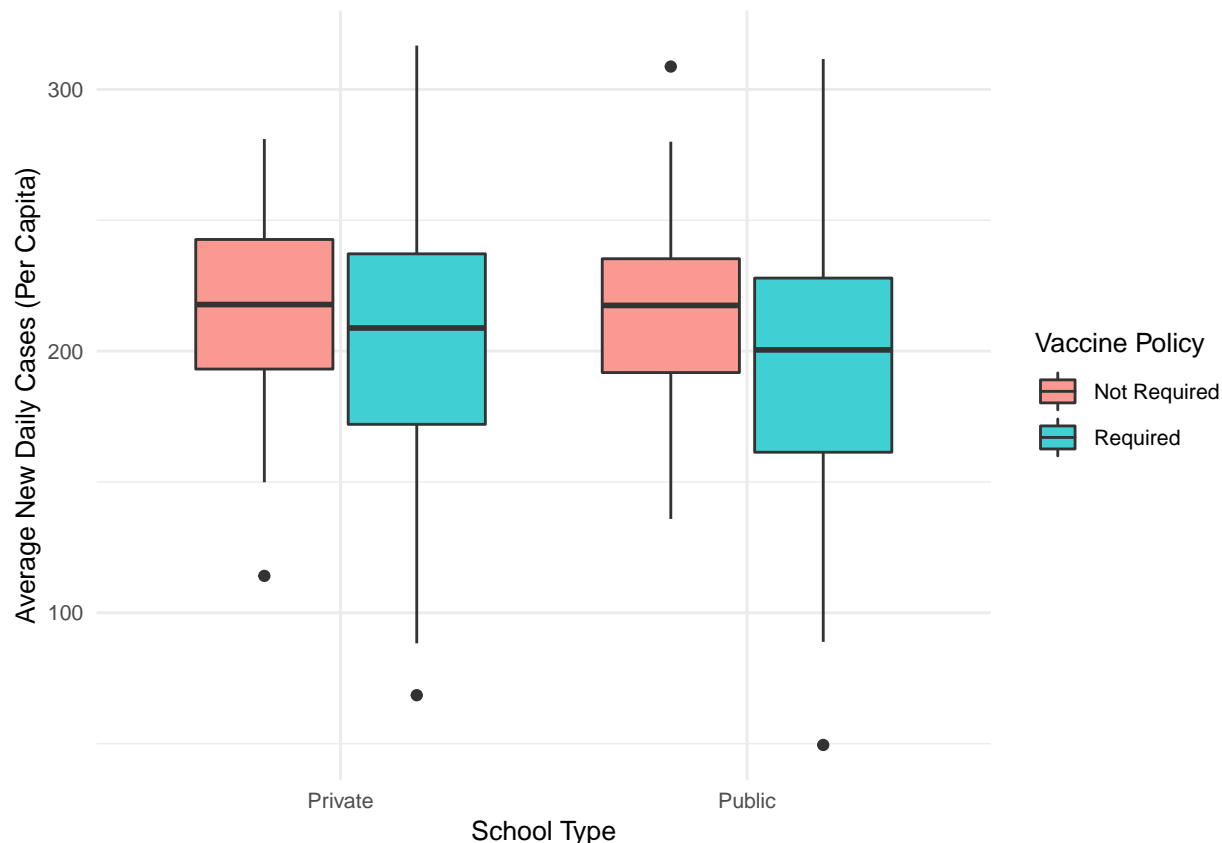
Looking at the boxplots for private schools in Figure **??**, we see again that the median average daily new covid cases for schools without a vaccine requirement is higher than that of schools with a vaccine requirement. We can also see that there is a larger spread for vaccine-required schools. Both vaccine-required and non-vaccine-required schools have low outliers. Looking at boxplots for public schools, we see again that the median average daily new covid cases for schools without a vaccine requirement is higher than that of schools with a vaccine requirement- however this time, there is a greater difference between the two. The schools that require vaccines have a larger spread and a low outlier and the schools that do not require vaccines have a high outlier.

Results

ANOVA Assumptions

To answer the question of whether a college or univerity's COVID-19 vaccine policy had a significant impact on the amount of COVID-19 cases in the county where the school's campus is located, we will perform a parametric one-way ANOVA F-test (omnibus). There are three assumptions that our data must satisfy to use this approach: Gaussian residuals, homoscedastic residuals, and independence of observations.

To get a better view of exactly how many of our residuals had non-Guassian behavior, we show a worm plot (a de-trended QQ plot) in Figure 3 with a 90% confidence envelope. The residuals appear approximately Gaussian ; only 9 of the points (~3% of the residuals) fall outside of the 90% confidence envelope for following a Guassian 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 Guassian distribution. So, we will act as if the assumption of Gaussian residuals is satisfied.
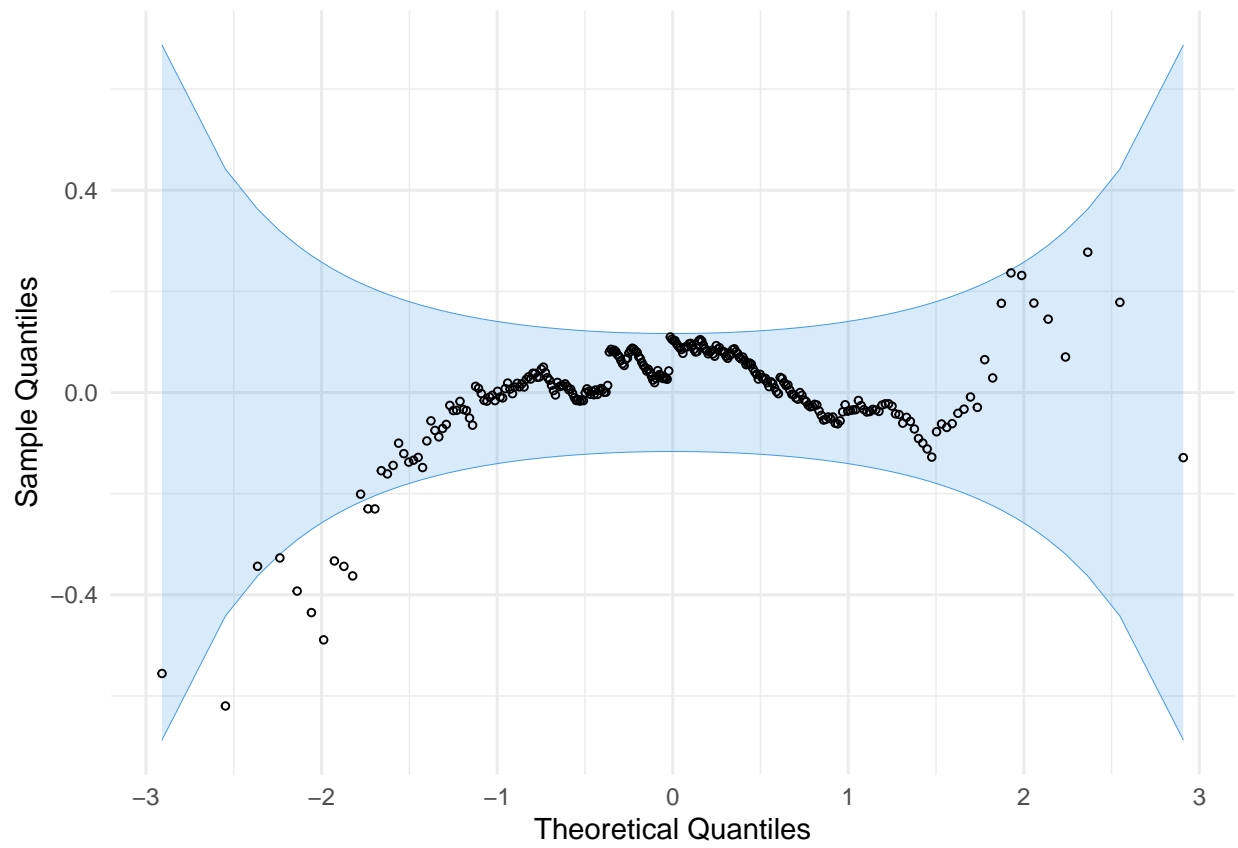
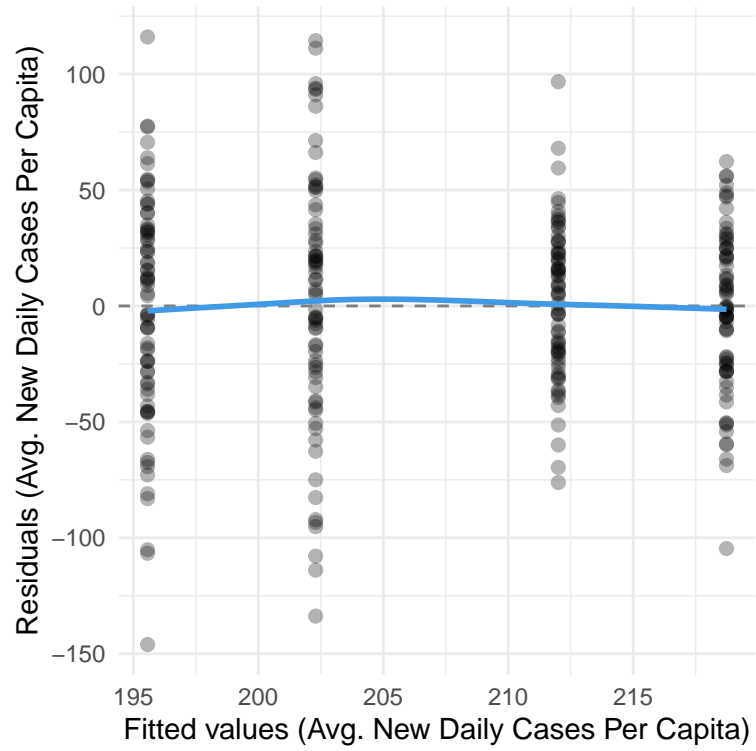Figure 3: Worm Plot for Vaccine Policy Study

Figure 4: Tukey-Anscombe Plot for Vaccine Policy Study

In Figure 4 we show a Tukey-Anscombe plot for our residuals. The fitted line for this plot is alomst 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.

Average New Daily Cases Per Capita
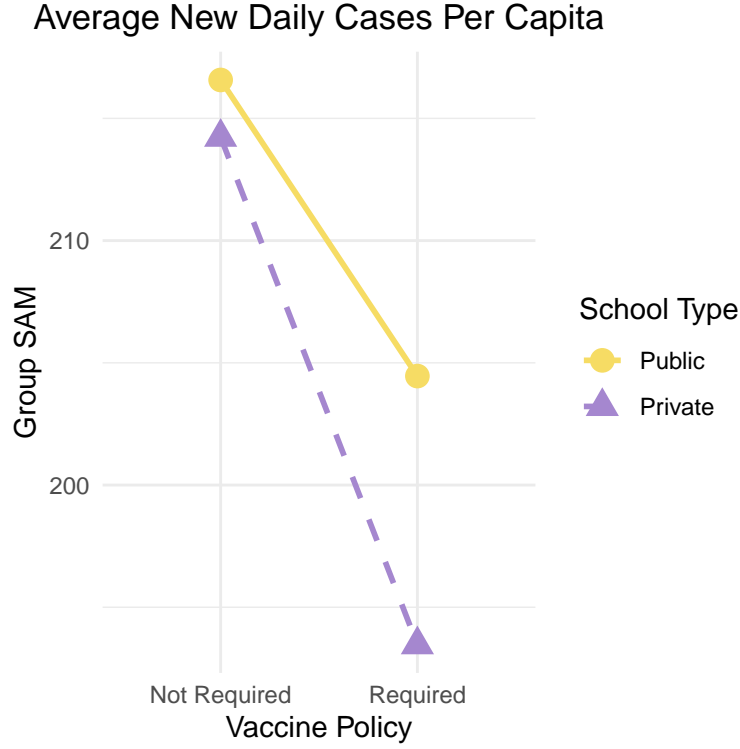
The interaction plot in Figure **??** shows the sample arithmetic mean (SAM) for each combinations of School Type and Vaccine Policy present in our study. Although the lines aren't parallel, we see the same trend in Group SAM Average New Daily Cases within both blocks. We will act as if there is no interaction 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.

Omnibus

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 **??** can see here in our ANOVA table that the status of a school as either requiring vaccines or not requiring vaccines ("Policy") accounts for ~9.65 times as much variation in average daily new covid cases as the residuals in this model. Since our p-value is less than our unusualness threshold of 5% (~0.0021 < 0.05), we will reject the null hypothesis and decide to act as if a school being public or private does have an impact on average daily new covid cases/capita in a school's home county. In particular, vaccine policy accounts for a ~3% of the variation average daily 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-Guassian 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 schools as a whole accumulated a rate of 207.15 average daily new covid cases (per capita)/school (GSAM). We can also see the factor level (vaccine policies) effects estimates. Schools without a vaccine requirement accumulated 8.22 more average daily new covid cases (per capita)/school than the GSAM baseline. Schools with a vaccine requirement accumulated 8.22 less average daily new covid cases (per capita)/school than the GSAM baseline. This suggests that the schools without vaccine requirements performed worse than the baseline, white the schools with vaccine requirements performed better than the baseline.

Discussion

After our analysis of the data and exploring our research question of whether or not Fall 2021 COVID-19 policies at American universities relate to the arithmetic mean of daily new COVID-19 cases/capita among the universities' home counties in the time-frame of September 1st to but not including December, 1st we found that a school's vaccination policy did in fact seem to have an impact on their county's average daily new COVID-19 cases/capita. According to our anova model, the vaccine policies of colleges and universities account for ~9.65 times as much variation in average daily new covid cases/capita in the school's home county as the residuals in this model.We found that this impact is statistically significant at the level of .02% and this impact had practical significance at the 3% level. We noticed that the data spread for schools that had a vaccination requirement was much larger than for those without a vaccination requirement. This meant that schools with a vaccine requirement had a much lower minimum value of average daily new COVID-19 cases/capita than schools without a vaccination requirement. Though this also meant that schools with a vaccination requirement had a larger maximum value for average daily new COVID-19 cases/capita, the difference was not as large as it was in comparing the two minimum values. In addition, when we compare our baseline GSAM of average daily new covid cases (per capita)/school, the schools that had a vaccine requirement accumulated 8.22 average daily new cases less than the baseline while the schools without a vaccine requirement did the inverse, accumulating 8.22 average daily new cases more than the baseline.

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 munici-palities. 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 reasearch 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 aroudn 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 problematic 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. Her code is available in the appendix below, but her entire project directory can be seen on https://github.com/CLTodd/Stat461Final (Currently still being formatted).
- Jordan cleaned the post-processed data and contributed to the Introduction, Study Design, EDA, Results, and Discussion sections.

Code Appendix

```r
# 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.github.com/neilhatfield/STAT461/master/rScripts/ANOVATools.R")
source("https://raw.github.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 (Schools) 277")
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"))
```

```r
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")]
densityByVax <-
ggplot(covid, aes(x=avgCasePerCapita, fill=Policy)) +
  geom_density(adjust=1.5, alpha=.75, color="black")  +
  ggtitle("Average New Daily Covid 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"))
covidStatsByVax <-
  psych::describeBy(
  x = covid$avgCasePerCapita,
  group = c(covid$Policy),
  na.rm = TRUE,
  skew = TRUE,
  ranges = TRUE,
  quant = c(0.25, 0.75),
```

```r
    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)
ggarrange(densityByVax, boxByVax, ncol=1)
covidStatsByVax %>%
  knitr::kable(
    caption = "Summary Statistics for Avg. Daily New Covid Cases 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"))
ggplot(data = covid,
       mapping = aes(
         x = Type,
         y = avgCasePerCapita,
         fill = Policy)) +
  geom_boxplot() +
  theme_minimal() +
  xlab("School Type") +
  ylab("Average New Daily 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"))
covidAOV <-
  aov(formula= avgCasePerCapita ~ Type + Policy,
      data = covid,
      na.action = "na.omit")

covidANOVA <- anova(covidAOV)
anovaTable <-
  parameters::model_parameters(model = covidAOV,
                               omega_squared = "partial",
```

```r
                                    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"))
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"))
# Standardized wormplot, used this because
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) +
```

```r
  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 Daily Cases Per Capita)") +
ylab("Residuals (Avg. New Daily Cases Per Capita)")

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

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) +
  # 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 Daily Cases Per Capita") +
  labs(shape = "School Type",
       color="School Type",
       linetype="School Type") +
  theme(legend.position = "right")

anovaTable
# Calculate Relative efficiency
block.RelEff(aov.obj = covidAOV,
```

```
            blockName = "Type",
            trtName = "Policy")
pointEstimatesTable
covidOldAOV <-
  aov(formula= avgCasePerCapita ~ Type + Policy,
      data = covid,
      na.action = "na.omit")
```