

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?

Study Design & Methods

The Hasse Diagram has to be changed.

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 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

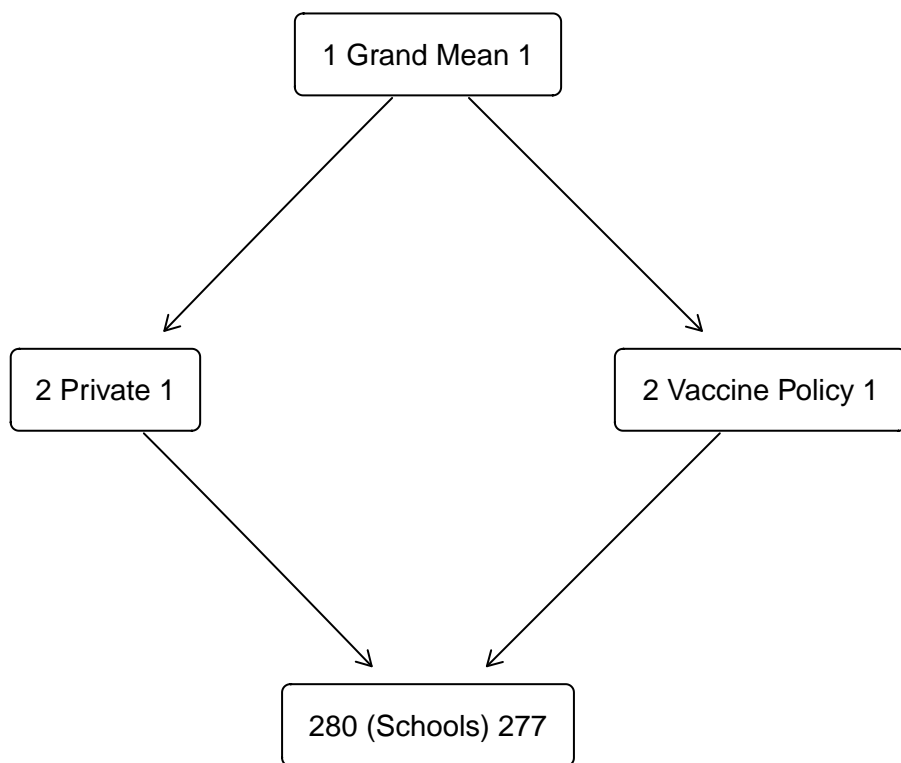


Figure 1: Hasse Diagram

that a school's status as public can put far more limits on it as for vaccination makes 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.

We will adopt the following hypotheses:

- H_0 : 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 : 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

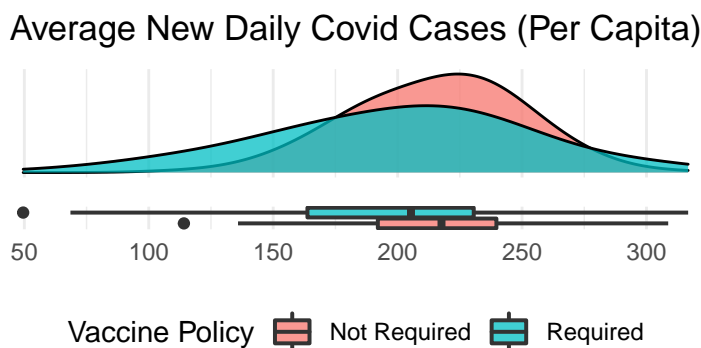


Figure 2: Graph

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

Comment on vaccine policy stats.

Results

ANOVA Assumptions

To answer the question of whether a college or university'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.

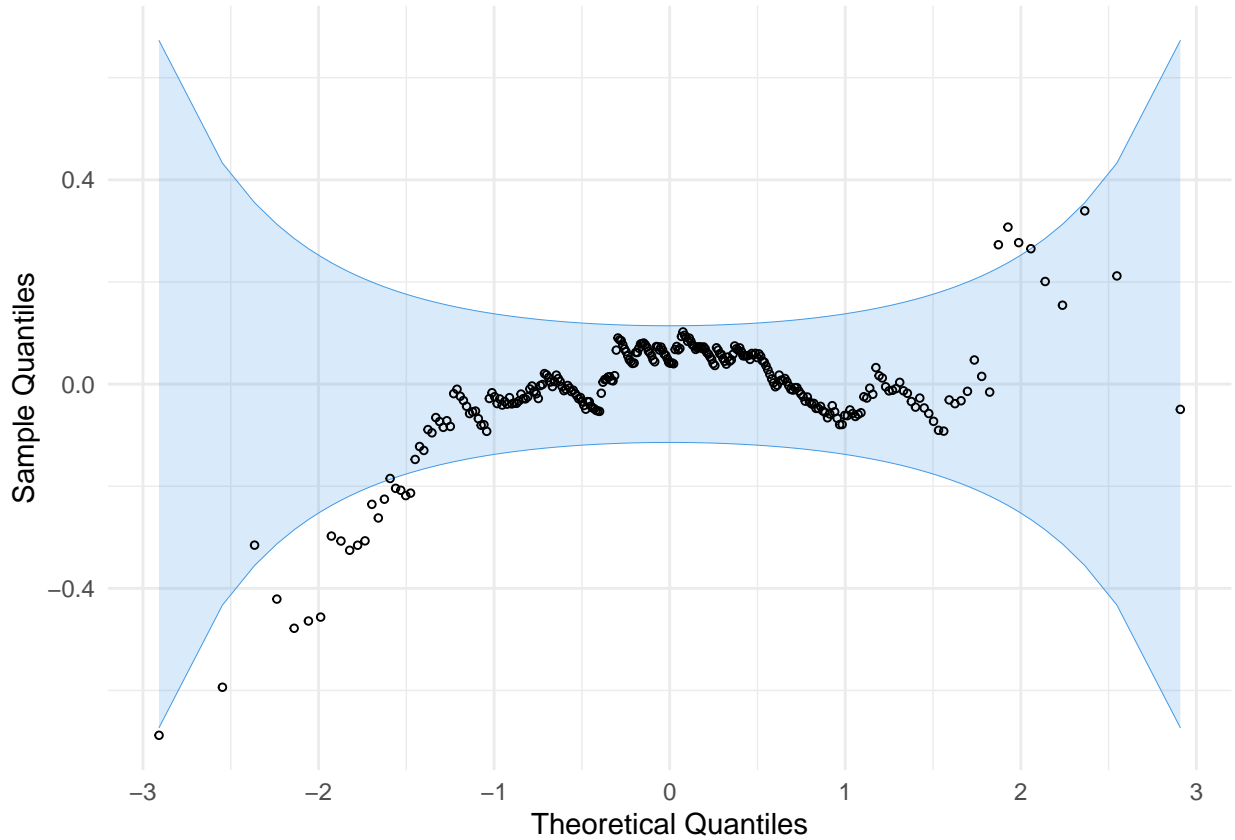


Figure 3: Worm Plot for Vaccine Policy Study

To get a better view of exactly how many of our residuals had non-Gaussian 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 22 of the points (~7% of the residuals) fall outside of the 90% confidence envelope. Additionally, the residuals have a sample skewness of ~ -0.5 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 Gaussian assumption is satisfied.

In Figure 4 we show a strip chart for our residuals. There seems to be no significant difference in the spread of the residuals across different fitted values, so we will act as if the assumption of homoscedastic residuals is satisfied.

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.

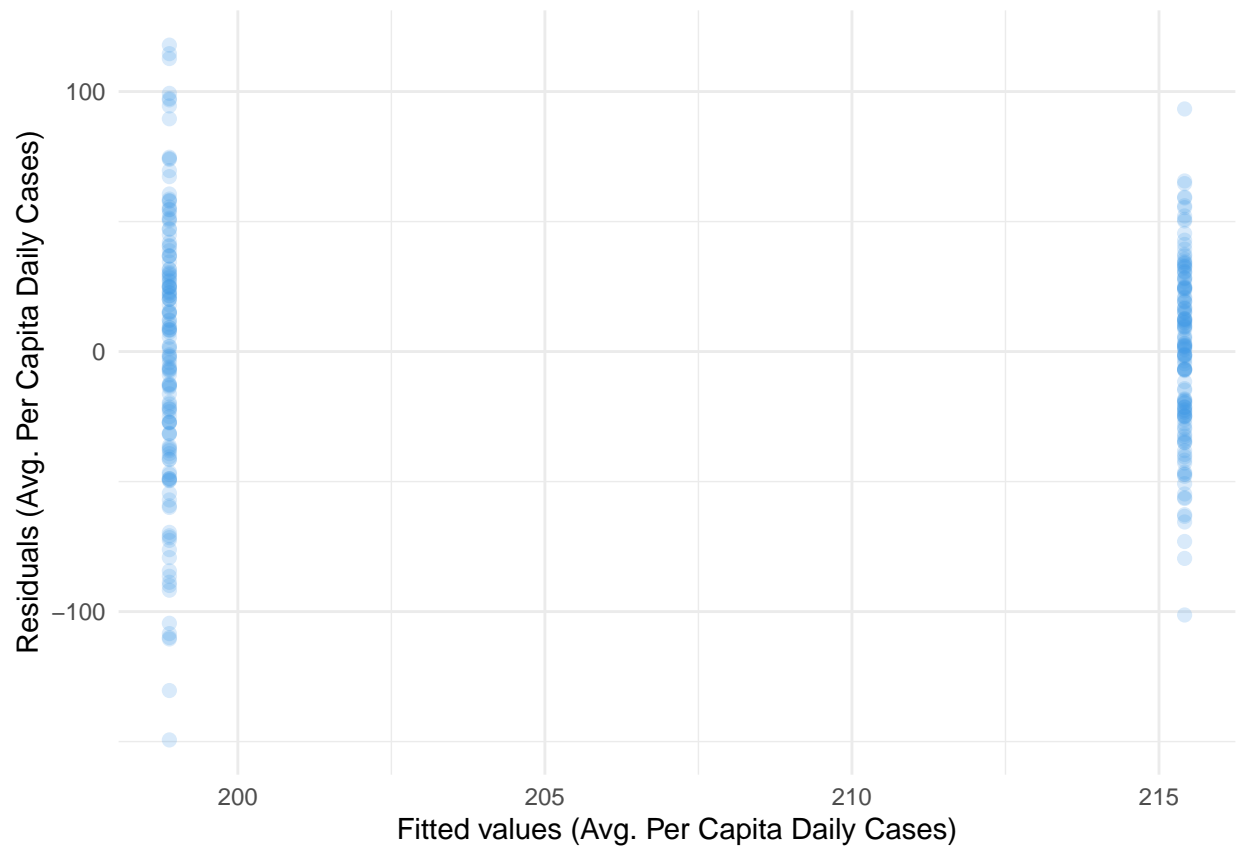


Figure 4: Strip Chart for Vaccine Policy Study

Omnibus

Table 2: ANOVA Table for Vaccine Policy Study

Source	SS	df	MS	F	p-value	Omega Sq.	Eta Sq.	Epsilon Sq.
Policy	19127.45	1	19127.452	9.747	0.002	0.0303	0.0339	0.0304
Residuals	545543.51	278	1962.387					

We can see in Table ?? that a college or univeristy’s vaccine policy accounted for ~ 9.747 times as much variation in the average new daily COVID-19 cases (per capita) of the county that the school is located in as the residuals in this model in the fall of 2021. Since our p-value is less than our unusualness threshold of 5% ($0.002 < 0.05$), we will reject the null hypothesis; We decide to act as if a college or university’s vaccine policy for the fall 2021 semester did impact the average new daily COVID-19 cases per capita of the county that the school is located in. In particular, a school’s vaccine policy accounted for a $\sim 30.3\%$ of the variation in a county’s average new daily COVID-19 cases per capita ($\omega^2 = 0.0303$, $\epsilon^2 = 0.0304$, $\eta^2 = 0.0339$).

Table 3: Point Estimates from the Vaccine Policy Study

	Estimate
GSAM	207.15
No Vaccine Requirement	8.27
Vaccine Required	-8.27

As a whole, the sample of undergraduate-serving 4-year colleges, universities, and professional schools accumulated 207.15 average daily new COVID-19 cases per capita per school. We see that schools that did not start the Fall 2021 semester requiring students to get the COVID-19 vaccine performed worse than the GSAM baseline; such schools accumulated 8.27 *more* average daily new COVID-19 cases per capita per school (keep in mind that *more* cases is the *worse* outcome in this context). Schools that started the Fall 2021 semester requiring students to get the COVID-19 vaccine performed better than the GSAM baseline; such schools accumulated 8.27 *less* average daily new COVID-19 cases per capita per school.

Discussion

Here we can put some thing about our conclusion with vaccines and what that means in a broader real-world context.

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. TOther 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 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.

The final reduction of scope resulted in the block of school type being dropped from the study as a result of the low relative efficiency of the ANOVA model including the block, the lack of apparent variation being screened by the block, and our own faulty blocking technique. The latter of these was the final deciding factor; it wasn't until after we had begun exploring the data (post manual data collection) that we realized we had not stratified our population of schools before making our sample selection. Without the time to manually collect data from a new stratified sample, blocks were dropped from the study.

Better planning and organization could have prevented much of the scope reduction we've discussed here. We believe we have learned from this process and hope others observing our work can learn from us as well.

Author Contributions

The authors of this report would like to acknowledge their individual contributions to the report.

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 (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"))
```

```

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 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),

```

```

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"))
covidAOV <-
aov(formula= avgCasePerCapita ~ Policy,
     data = covid,
     na.action = "na.omit")

covidANOVA <- anova(covidAOV)
anovaTable <-
parameters::model_parameters(model = covidAOV,
                              omega_squared = "raw",
                              eta_squared = "raw",
                              epsilon_squared = "raw") %>%
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", "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()

# QQ Plot for Residuals
car::qqPlot(x = residuals(covidAOV),
            distribution = "norm",
            envelope = 0.9,
            id = FALSE,
            pch = 20,
            ylab = "Residuals (Cases Per Capita)")
ggplot(data = data.frame(residuals = residuals(covidAOV),
                        fitted = fitted.values(covidAOV)),
       mapping = aes(x = fitted, y = residuals)) +
  geom_point(size = 2, alpha=0.2, col="#419AE6") +
  theme_minimal() +
  xlab("Fitted values (Avg. Per Capita Daily Cases)") +
  ylab("Residuals (Avg. Per Capita Daily Cases)")
anovaTable
pointEstimatesTable
covidOldAOV <-
  aov(formula= avgCasePerCapita ~ Type + Policy,
      data = covid,
      na.action = "na.omit")

```