

COVIDSchoolClosures*

Christina, Yan, Sakura

February 12, 2024

Abstract

COVID-19 was an outbreak of virus that forced many institutions to shut down for 2-3 years. Schools were no different and this paper aims to look at the effects of the said closures in school and how it affected the population. This paper finds that with more inperson schooling provided the less the enrollment rates drop(more in depth analysis in later sections).

Introduction

Data

Data Source and Collection //Methodology

We use R Core Team (2022) to make this paper as well as the graphs and topic were taken inspiration from Jack and Oster (2023). Various helpful packages were used in order to clean, sort and graph this paper in a way

*Code and some data from this paper are available at:
<https://github.com/NotSakura/COVIDSchoolClosures.git>

such that the reader will not have difficulty understanding neither the topic nor the data sets of this paper. The packages are, Wickham (2016), Wickham et al. (2019), Wickham et al. (2023), Wickham, Hester, and Bryan (2023), Xie (2014), Firke (2023), Zhu (2021), Wickham, Vaughan, and Girlich (2024), Wickham and Miller (2021), Hyndman and O'Hara-Wild (2021).

Data Cleaning

What data set did we clean and why. Explain the variable here too if The data provided originally was called nces

Results

The graphs we made and describe the trends. Only talk about results not what they mean

First Graph

##Second Graph

Third Graph

Rows: 14967 Columns: 7

-- Column specification -----

Delimiter: ","

chr (3): StateAbbrev, DistrictName, StateAssignedDistrictID

dbl (4): NCESDistrictID, share_inperson, share_hybrid, share_virtual

i Use `spec()` to retrieve the full column specification for this data.

i Specify the column types or set ``show_col_types = FALSE`` to quiet this message.

```
data <- raw_d |>
  filter(grade == 0 | grade == 99) |>
  filter(race == 99)

data <- subset(data, select = -race)

# Drop observations where enrollment is less than 0
data <- data[data$enrollment >= 0, ]

# Drop observations where fips is greater than 56
data <- data[data$fips <= 56, ]

data <- data %>%
  filter(!is.na(sex))

# Reshape 'enrollment' wide by 'leaid', 'year', and 'grade'
data_wide <- pivot_wider(data,
  names_from = sex,
  values_from = enrollment) |>
  rename(
    enrollment1 = `1`,
    enrollment2 = `2`,
    enrollment9 = `9`,
    enrollment99 = `99`)

# Drop the 'enrollment9' variable
data_wide$enrollment9 <- NULL
```

```

# Drop observations where 'enrollment99' is missing or less than or equal to 0
data_wide <- data_wide[!is.na(data_wide$enrollment99) & data_wide$enrollment99 > 0]

# Reshape 'enrollment1', 'enrollment2', and 'enrollment99' wide by 'leaid' and 'year'
data_final <- pivot_wider(data_wide,
                          names_from = grade,
                          values_from = c(enrollment1, enrollment2, enrollment99))

# Can get rid of fips

data_final <- subset(data_final, select = -fips)

# Rename columns
colnames(data_final) <- c("year", "leaid",
                          "enroll_male_kinder", "enroll_female_kinder",
                          "enroll_male_total", "enroll_female_total",
                          "enroll_all_kinder", "enroll_all_total")

# Convert 'leaid' to numeric (assuming it's a string)
data$leaid <- as.numeric(data$leaid)
other_raw$leaid <- as.numeric(other_raw$leaid)

write_csv(x = data_final,
          file = "../data/cleaned_raw_data.csv")

# Link data with another dataset (replace 'other_data' with your actual dataset)
# Assuming 'leaid' is the common identifier
merged_data <- merge(data_final, other_raw, by = "leaid")

write_csv(x = merged_data,
          file = "../data/merged_data.csv")

```

```

# Assuming 'year' is numeric (e.g., 2019, 2020)
merged_data$date <- as.Date(paste0(merged_data$year, "-01-01"))

my_tsibble <- as_tsibble(merged_data, key = leaid, index = date)
my_tsibble <- my_tsibble %>%
  filter(year >= 2020)

# Assuming you have a tsibble called my_tsibble with appropriate columns

# Calculate change_kinder
my_tsibble <- my_tsibble %>%
  mutate(change_kinder = enroll_all_kinder - lag(enroll_all_kinder),
         perc_change_kinder = change_kinder / lag(enroll_all_kinder))

# Calculate change_all
my_tsibble <- my_tsibble %>%
  mutate(change_all = enroll_all_total - lag(enroll_all_total),
         perc_change_all = change_all / lag(enroll_all_total))

# Calculate change_male
my_tsibble <- my_tsibble %>%
  mutate(change_male = enroll_male_kinder - lag(enroll_male_kinder),
         perc_change_male = change_male / lag(enroll_male_kinder))

# Calculate change_female
my_tsibble <- my_tsibble %>%
  mutate(change_female = enroll_female_kinder - lag(enroll_female_kinder),
         perc_change_female = change_female / lag(enroll_female_kinder))

# Drop rows where absolute change_all is greater than enroll_all_total

```

```

my_tsibble <- my_tsibble %>%
  filter(abs(change_all) <= enroll_all_total)

# Drop rows where absolute change_kinder is greater than enroll_all_total
my_tsibble <- my_tsibble %>%
  filter(abs(change_kinder) <= enroll_all_total)

my_tsibble <- my_tsibble %>%
  mutate(inperson_bin = cut(share_inperson, breaks = 11))

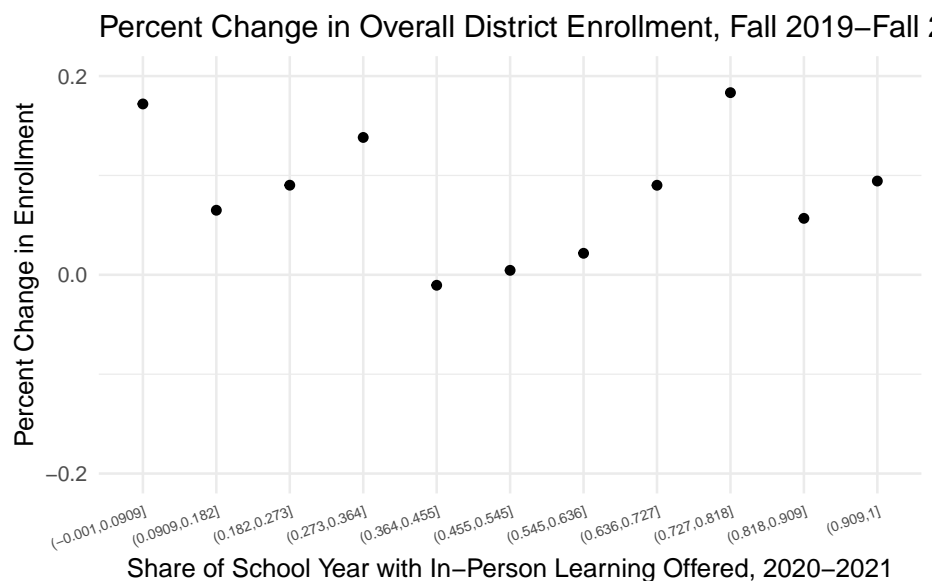
# Compute weighted mean change in enrollment for each bin
weighted_means_all <- my_tsibble %>%
  group_by(inperson_bin) %>%
  summarise(
    weighted_mean_change = weighted.mean(perc_change_all, enroll_all_total)
  )

weighted_means_all <- subset(weighted_means_all, select = -date)
weighted_means_all <- weighted_means_all %>%
  group_by(inperson_bin) %>%
  summarise_at(vars(weighted_mean_change), list(weighted_mean_change = mean)
  )

weighted_means_kinder <- my_tsibble %>%
  group_by(inperson_bin) %>%
  summarise(
    weighted_mean_change = weighted.mean(perc_change_kinder, enroll_all_kinder)
  )
weighted_means_kinder <- subset(weighted_means_kinder, select = -date)

```

```
##Graph
ggplot(weighted_means_all, aes(x = inperson_bin , y = weighted_mean_change)) +
  geom_point() +
  scale_y_continuous(breaks = seq(-0.2, 1.0, by = 0.2), limits = c(-0.2, 0.2))+
  labs(
    title = "Percent Change in Overall District Enrollment, Fall 2019-Fall 2020",
    x = "Share of School Year with In-Person Learning Offered, 2020-2021",
    y = "Percent Change in Enrollment"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 20, vjust = 1, hjust = 1, size = 6))
```

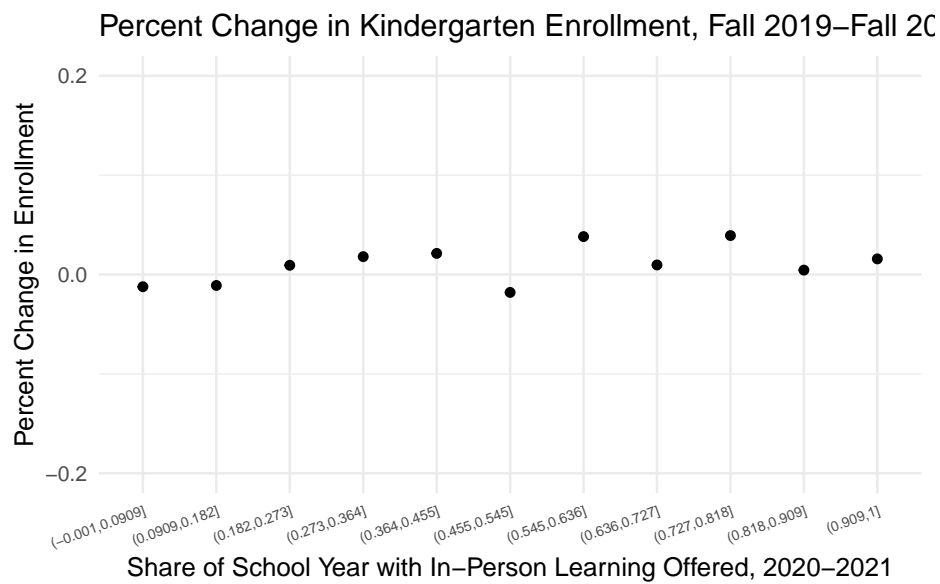


```
ggplot(weighted_means_kinder, aes(x = inperson_bin, y = weighted_mean_change)) +
  geom_point() +
```

```

scale_y_continuous(breaks = seq(-0.2, 1.0, by = 0.2), limits = c(-0.2, 0.2))+
labs(
  title = "Percent Change in Kindergarten Enrollment, Fall 2019–Fall 2020",
  x = "Share of School Year with In-Person Learning Offered, 2020–2021",
  y = "Percent Change in Enrollment"
) +
theme_minimal()+
theme(axis.text.x = element_text(angle = 20, vjust = 1, hjust = 1, size = 6))

```



Discussion

Interesting point 1

Intresting point 2

interesting point 3

##Ethics and Bias could talk about mental health maybe but it might apply to other “interesting point”

weakness and limitations

how to solve the limitations

Furthur questions?

Appendix

```
weighted_means_all |> kable(format = "pipe", padding = 2, col.names = c("Share of
```

Share of inperson	average enrollment rate
(-0.001,0.0909]	0.1719376
(0.0909,0.182]	0.0649727
(0.182,0.273]	0.0901854
(0.273,0.364]	0.1382330
(0.364,0.455]	-0.0105187
(0.455,0.545]	0.0045212
(0.545,0.636]	0.0216597
(0.636,0.727]	0.0901187

Share of inperson	average enrollment rate
(0.727,0.818]	0.1833064
(0.818,0.909]	0.0568220
(0.909,1]	0.0943735

Reference

- Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Hyndman, Rob, and Mitchell O’Hara-Wild. 2021. *Tsibble: Tidy Temporal Data Frames and Tools*. <https://CRAN.R-project.org/package=tsibble>.
- Jack, Rebecca, and Emily Oster. 2023. “COVID-19, School Closures, and Outcomes.” *Journal of Economic Perspectives* 37 (4): 51–70. <https://doi.org/10.1257/jep.37.4.51>.
- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemond, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. <https://CRAN.R-project.org/package=readr>.
- Wickham, Hadley, and Evan Miller. 2021. *Haven: Import and Export ‘SPSS’, ‘Stata’ and ‘SAS’ Files*. <https://CRAN.R-project.org/>

[package=haven](#).

Wickham, Hadley, Davis Vaughan, and Maximilian Girlich. 2024. *Tidyr: Tidy Messy Data*. <https://tidyr.tidyverse.org>.

Xie, Yihui. 2014. *Knitr: A Comprehensive Tool for Reproducible Research in R*. Edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. <http://www.crcpress.com/product/isbn/9781466561595>.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. <http://haozhu233.github.io/kableExtra/>.