

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

```
#| echo: false
```

```
score_data = read_csv("../inputs/data/scores_lm_demographics.csv")
```

New names:

Rows: 9823 Columns: 63

-- Column specification

----- Delimiter: "," chr
(9): state, DistrictName, subject, lea_name, fips, zip_location, urban... dbl

```
(51): ...1, county_code, covid_level, year, NCESDistrictID, lunch, miss... lgl
(2): spec_ed_students, english_language_learners date (1): ReportingDate
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.
* `` -> `...1`
```

```
clean_score_data_inperson <- score_data |>
  select(subject, change_2017_2018, change_2018_2019, change_2019_2021, change_2021_2022)
  rename_with(~ sub("^change_(\\d{4})_(\\d{4})$", "Spring_\\2", .), starts_with("change_"))
  mutate(share_inperson_grouped = cut(share_inperson * 100, breaks = seq(0, 100, 10)))

# Pivot the data to a long format
score_data_long_inperson <- clean_score_data_inperson |>
  pivot_longer(cols = starts_with("Spring"), names_to = "time_period", values_to = "change_score")

# Group by 'subject', 'share_inperson_grouped', and 'time_period', then summarize
score_data_summary_inperson <- score_data_long_inperson |>
  group_by(subject, share_inperson_grouped, time_period) |>
  summarise(
    mean_change = mean(change_score, na.rm = TRUE),
    .groups = 'drop'
  )

# Now prepare data for the 'urban-centric_locale' grouping
score_data_long_locale <- clean_score_data_inperson |>
  pivot_longer(cols = starts_with("Spring"), names_to = "time_period", values_to = "change_score")

# Group by 'subject', 'urban-centric_locale', and 'time_period', then summarize
score_data_summary_locale <- score_data_long_locale |>
  group_by(subject, urban_centric_locale, time_period) |>
  summarise(
    mean_change = mean(change_score, na.rm = TRUE),
    .groups = 'drop'
  )
```

```

)

ggplot(score_data_summary_inperson, aes(y = share_inperson_grouped, x = round(mean_
geom_point(position = position_dodge(width = 0.2)) +
scale_x_continuous(limits = c(-15, 5), breaks = seq(-15, 5, by = 5)) +
labs(
  title = "Average Grade Change by In Person Attendance",
  y = "In-Person Share Group (%)",
  x = "Average Change Score (%)",
  color = "Time Period"
) +
scale_color_brewer(palette = "Set1", labels = c("Spring 2018", "Spring 2019", "S
theme_minimal() +
theme(
  legend.position = "bottom",
  legend.background = element_rect(fill = "white", size = 0.3, linetype = "solid
  legend.text = element_text(size = 8),
  legend.title = element_text(size = 10, face = "bold"),
  legend.key.size = unit(0.2, "cm")
) +
facet_wrap(~subject)

```

Warning: The `size` argument of `element_rect()` is deprecated as of ggplot2 3.4.0.
 i Please use the `linewidth` argument instead.

Warning: Removed 1 rows containing missing values (`geom_point()`).

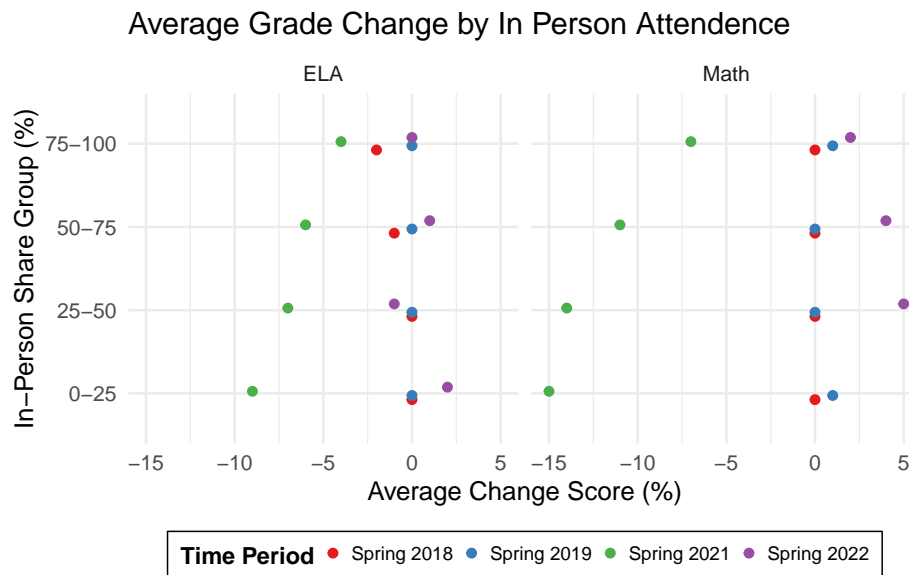


Figure 1: Average Grade Change by In Person Attendance

Warning: Removed 1 rows containing missing values (``geom_point()``).

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.

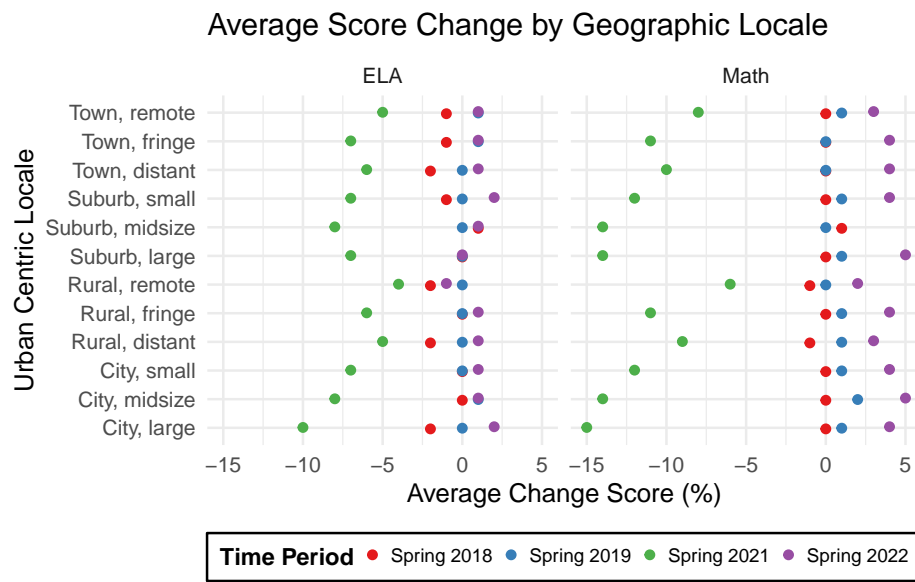
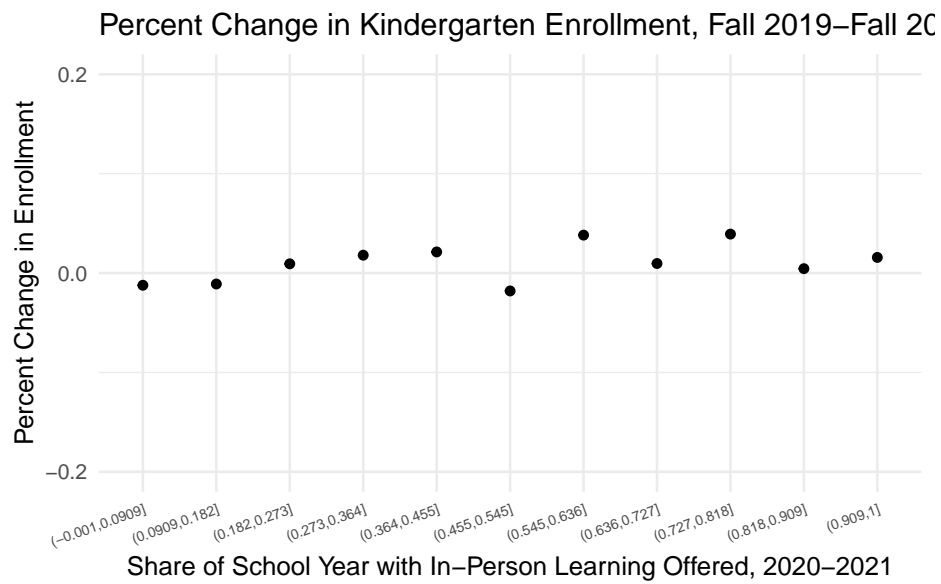
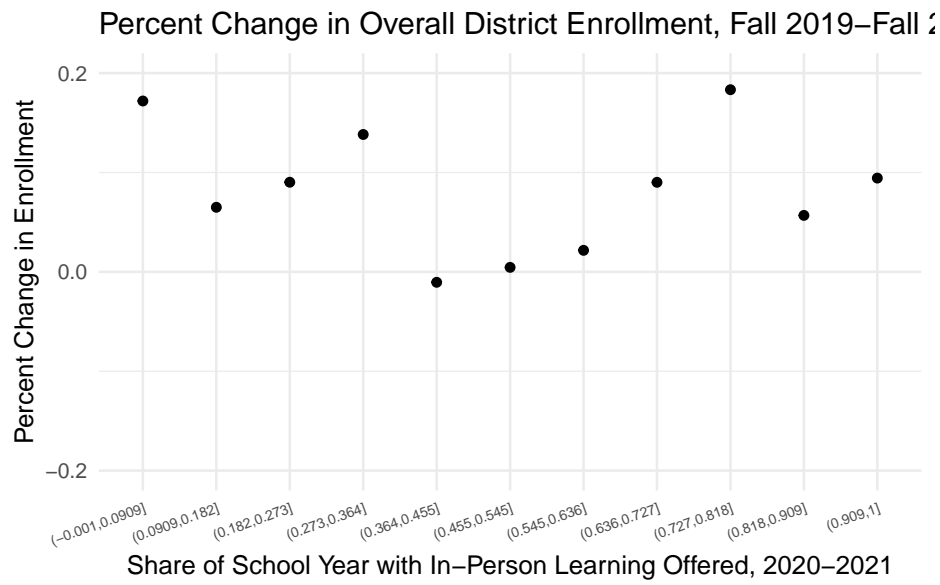


Figure 2: Average Score Change by Geographic Locale



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/>.