My title*

My subtitle if needed

First author Another author

February 13, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

The COVID-19 pandemic has profoundly disrupted education systems worldwide, compelling schools and districts across the United States to rapidly adapt their learning models. In 2020 and 2021, educational institutions faced unprecedented challenges, transitioning between in-person, hybrid, and virtual learning models in response to evolving public health guidelines and infection rates. This shift has brought into sharp focus the need to understand the implications of these various schooling models on educational outcomes.

This paper delves into a comprehensive analysis across 11 U.S. states, examining the relationship between different schooling models and their impact on student enrollment and staffing patterns. The unique dataset compiled for this study encompasses a variety of metrics, including total enrollment, in-person and virtual attendance, as well as staff count across different learning models. By analyzing data from kindergarten to 12th grade, this study provides a granular view of how the pandemic has reshaped the educational landscape across these states.

Our findings reveal significant variations in how states have navigated the challenges posed by the pandemic, with notable differences in the adoption of in-person, hybrid, and virtual learning models. We observe that these variations are not only reflective of public health policies but also indicate broader socio-economic and demographic influences. For instance, preliminary analysis suggests that shifts to virtual learning correlate with changes in student enrollment, raising questions about equity and access in education during these challenging times.

*Code and data are available at: LINK.

The remainder of the paper is structured as follows: Section 2 provides a detailed overview of the data and methodology employed in this study. Section 3 presents an in-depth analysis of the schooling models across the 11 states, offering insights into the trends and patterns observed in the data. Section 4 discusses the key findings, exploring the implications of these schooling models on educational outcomes. Finally, Section 5 concludes with a reflection on the study's findings, limitations, and potential directions for future research in this critical area of educational policy.

The remainder of this paper is structured as follows. Section 2....

2 Data

2.1 Data Source and Collection

The data utilized in this study is obtained from the compilation of district-level schooling models and state standardized assessment data, originally featured in the paper "Pandemic Schooling Mode and Student Test Scores: Evidence from US School Districts" (Jack 2023) published in the American Economic Review: Insights in June 2023. (AEA 2022). The primary dataset encompasses schooling mode data from the 2020–2021 academic year across 11 U.S. states, integrating various educational approaches during the pandemic, including in-person, hybrid, and virtual learning environments. These datasets were sourced from COVID-19 School Data Hub (USA 2022), district-level state standardized assessment data from spring 2016–2019 and 2021, and additional data demographic statistics from the National Center for Education Statistics (NCES) to establish a comprehensive analytical framework. We also used additional year 2020 household data by Household Pulse Survey (Bureau 2020) from U.S. Census Bureau spanning weeks 1 to 12.

Our analyses incorporate three categories of data: district-level schooling modes, state standardized assessment results from the academic years 2016–2019 and 2021, and auxiliary that encompass district demographics and county-level variables. The following subsections detail the sources, collection methodologies, and data cleaning procedures undertaken to ensure the accuracy and reliability of the datasets for our analysis.

State Score Data: This dataset encompasses state-level academic performance statistics from 2016 to 2020. Variables include the proportion of in-person, virtual, and hybrid learning, participation rates, standardized test pass rates, COVID-19 case rates per 100,000 in school zip codes, peak monthly cases, total enrollment figures, and political voting shares by district.

Learning Model Data: This dataset details the learning models adopted by school districts in Colorado, Connecticut, Ohio, Virginia, West Virginia, Wyoming, Mississippi, Rhode Island, Minnesota, Massachusetts, and Wisconsin. Each state's data outlines the specific learning model (in-person, virtual, hybrid) employed over time, enrollment figures, and staffing counts.

2.2 Data Cleaning and Processing

We used R (R Core Team 2022) for data cleaning and processing, utilizing packages like tidyverse (Wickham et al. 2019) for data manipulation and janitor (Firke 2023) for cleaning column names. The cleaning process involved standardizing learning model categories, selecting columns of interest, and simplifying data for analysis. We integrated data from different states, and made sure that variations in reporting and measurement were addressed while combining them into one. For instance, some states categorized their schooling mode as "In-person," while others used "In-Person." To ensure they are equivalent, we standardized these terms. We also conducted a check for missing data values, and imputed or excluded them as appropriate based on the context. The variables were also renamed for clarity and consistency across different states' datasets. A sample of cleaned state score data can be seen in (**tbl_cleaned_data?**).

Table 1: Sample of Cleaned State Score Data

State	Year	Share In-person	Share Virtual	Share Hybrid	Participation	Pass Rate	Covid cases rate per 100k	Total Enrollment
CO	2016	0.8586572	0.0600707	0.0812721	86.28333	0.3296667	29.04677	1132
CO	2016	0.9399293	0.0600707	0.0000000	91.75000	0.2460000	31.04686	8723
CO	2016	0.0000000	0.2756184	0.7243816	95.86667	0.2526667	31.53898	18657
CO	2016	0.0000000	0.5759717	0.4240283	98.58334	0.0950000	31.54341	5725
CO	2016	0.2155477	0.0600707	0.7243816	96.35000	0.3298333	29.81067	1001
CO	2016	0.2155477	0.4664311	0.3180212	97.26667	0.3730000	31.54496	35817

2.3 Data Modifications

In this study, we created a new dataset by carefully selecting and combining data from the Household Pulse Survey, available from the U.S. Census Bureau for weeks 1 to 12. This process involved choosing key variables that were crucial for our research. We aimed to analyze the connection between economic changes, particularly the loss of employment income, and their impact on educational outcomes, focusing specifically on school passing rates. This helped us to understand the relationship between these economic factors and educational performance. To ensure consistency, we standardized terminologies and measurements, and aligned variables such as timeframes and demographic categories across the datasets. For example, we combined the state score data with the weekly household data, aligned variables relevant to our study. The merged data file was thereafter saved as a CSV file to study the trends later on.

3 Results

In the Results section of our study, our analysis discusses two key geographical representations:

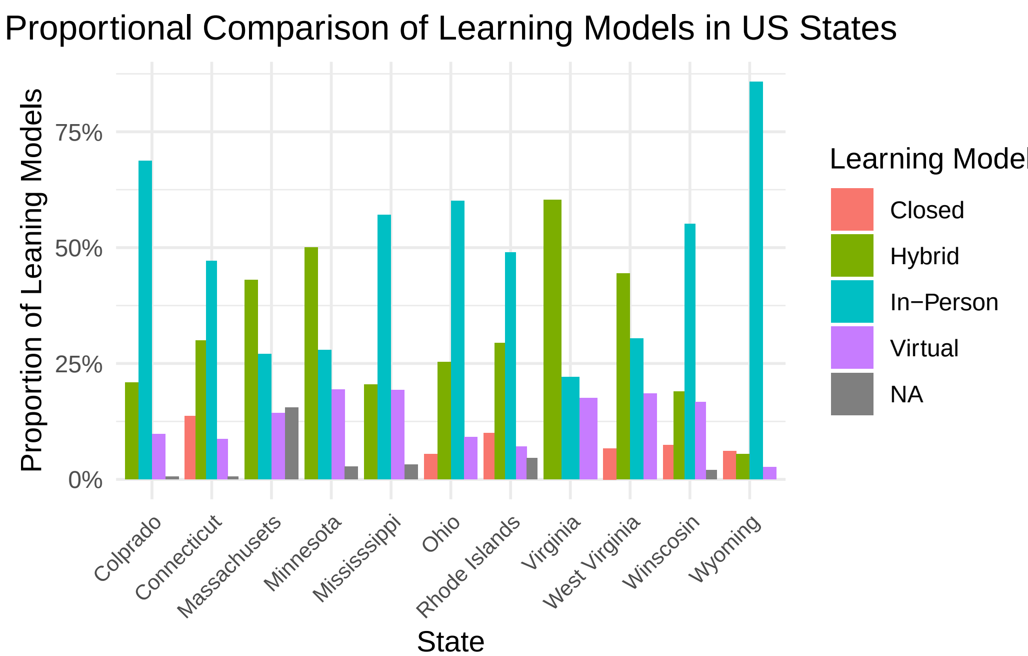

Figure 1: Bills of penguins

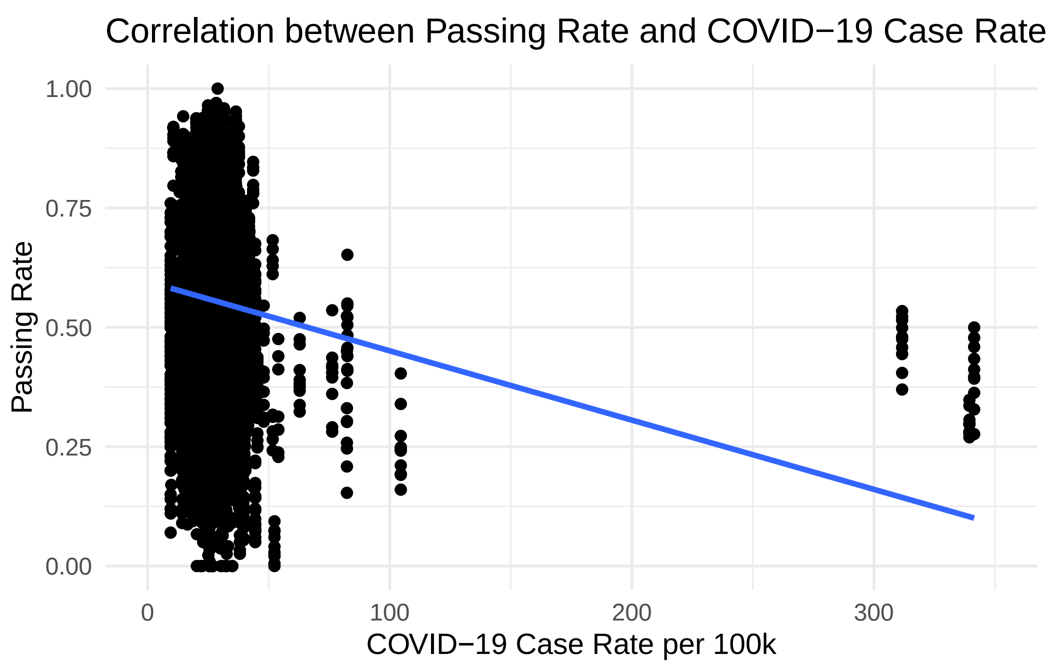

Figure 2: Bills of penguins

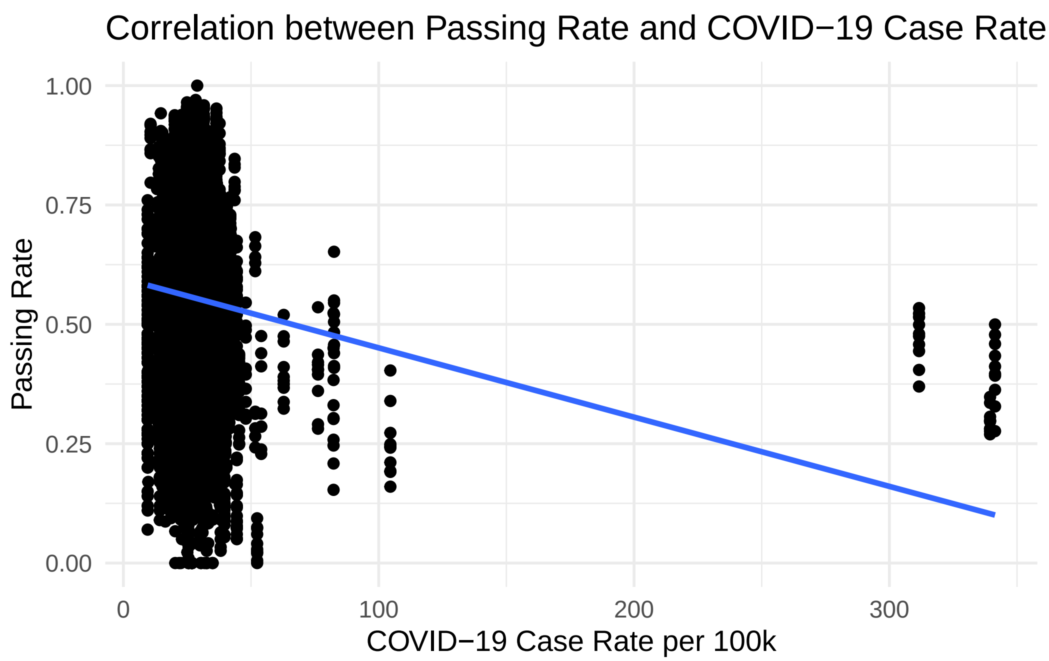

Figure 3: Bills of penguins

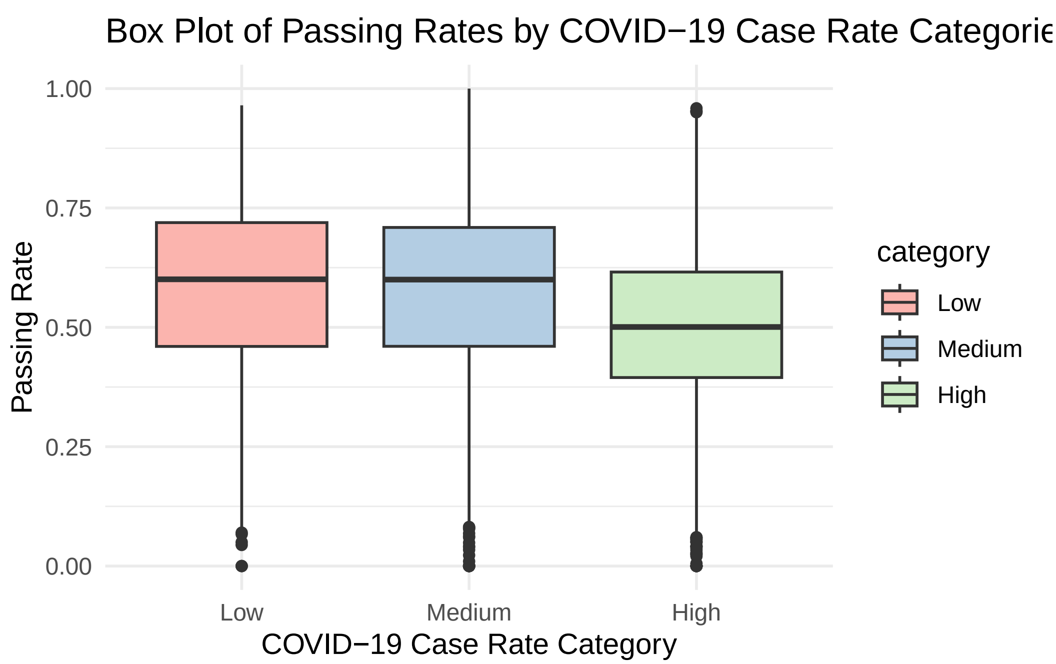

Figure 4: Bills of penguins

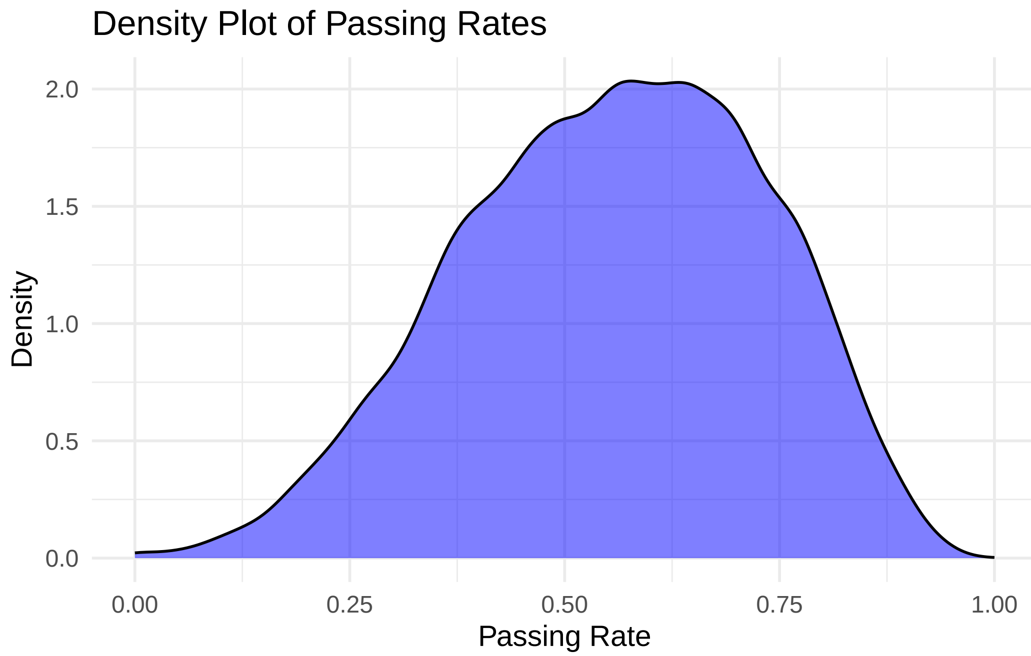

Figure 5: Bills of penguins

1. Passing Rate Trends Over Different Years and Subjects: Figue (**fig?**) shows the decline in passing rates in English Language Arts (ELA) and Mathematics from 2016 to 2021, depicting a consistent downward trend across both subjects especially during pandemic period from 2019 to 2021.

2. Variations in Passing Rates Between Different States: The box plot comparison (**fig2?**) across states showcases the wide range of passing rates, with some states showing a wide spread, which indicates the variability in performance within those states.

Building on these overviews, we visualized the distribution of learning models across states using box plots. The box plot for in-person learning (**box_plot_inperson?**) revealed a range of engagement across states, with Virginia showing a median of 0 percentage in-person learning during 2019 to 2021, whereas Wyoming having a median of approximately 80 percent in-person learning. The virtual learning model box plot (**box_plot_virtual?**) showed a significant number of outliers, suggesting that certain states had districts with a notably high share of virtual learning.The states Wyoming and Rhode Island had a median of 0 percent virtual learning whereas Virginia had highest median of virtual learning. Lastly, the hybrid learning model (**box_plot_hybrid?**) box plot displayed variability, with some states having a wide interquartile range, showing a diverse adoption of this learning model. However, Colorado had a median of 0 percent adoption of hybrid learning model relative to other states. These visual insights set the foundation for further statistical examination in subsequent analyses.

Following the initial box plots, we continued our investigation with a scatter plot (**scatterplot?**) and examined the relationship between the share of virtual learning and student passing rates. The scatter plot indicates a downward trend, with a line suggesting that an increase in the share of virtual learning correlates with a decrease in passing rates. The data points suggest that higher virtual learning is associated with lower academic performance of students.

After the initial analysis of learning model, we further studied the impact of Covid-19 on student's learning. We again plotted box plots (**covidboxplot?**) to observe the relationship between COVID-19 case rates per 100,000 and passing rates. COVID-19 case rates per 100,000 were categorized into 'Low,' 'Medium,' and 'High' based on the data's quartile distribution. It is revealed from the plot that districts with 'Low' COVID-19 case rates showed a median passing rates near the 50th percentile, while those categorized as 'High' showed a decline in median passing rates.

We further expanded our analysis to examine the relationship between the average number of students per staff and the average pass rate for virtual learners across the 11 states (**fig?**). Due to insufficient data, Rhode Island (RI) and Mississippi (MS) were not graphed. The graph plots both these variables and shows an overall positive correlation between the number of students per staff on average vs the average pass rate. The trend indicates that the states with the higher pass rates also had more students assigned to a teacher. We note that Virginia (VA) had the highest student-to-staff ratio, while Wyoming (WY) had the lowest. While this may be not expected of a general learning environment, the virtual learning model may be a factor contributing to this positive relationship between the student-to-staff ratio and the average pass rates across the states.

State	Avg. Share In-person	Avg. Share Virtual	Avg. Share Hybrid	Avg. Participation	Avg. Pass Rate
CO	65.68918	10.285258	24.02556	95.11431	36.03400
CT	55.21059	6.625613	31.45075	98.68581	59.33747
MA	33.46283	9.333824	57.20335	100.00000	53.69718
MN	35.23909	8.302583	56.45833	96.34993	56.35686
MS	56.80734	17.653758	25.53891	100.00000	72.84675
OH	63.31093	9.349329	26.61283	99.66106	69.27725
RI	54.77593	5.261545	34.39462	98.90301	41.80093
VA	16.28958	27.904606	55.80582	99.35298	76.24432
WI	68.34940	5.536344	17.69383	99.04141	44.81849
WV	31.84275	16.216216	48.20639	94.10710	40.05973
WY	85.46875	2.395833	5.62500	99.68056	55.27684

Relationship between wing length and width

State	Avg. Share In-person	Avg. Share Virtual	Avg. Share Hybrid	Avg. Participation	Avg. Pass Rate
CO	63.66436	10.832368	25.50328	81.01494	32.31432
CT	55.20204	6.625613	31.45931	93.31494	50.27606

MA	33.46283	9.333824	57.20335	95.58099	43.86268
MN	35.40717	8.277188	56.31564	87.20716	46.03220
MS	57.07332	17.624205	25.30248	97.04833	60.17785
OH	63.30790	9.319841	26.64837	97.11603	58.88783
RI	54.77593	5.261545	34.39462	90.82130	34.79347
VA	16.28958	27.904606	55.80582	84.78617	54.15240
WI	68.34940	5.536344	17.69383	95.60902	38.69319
WV	31.84275	16.216216	48.20639	87.97907	32.07227
WY	85.46875	2.395833	5.62500	97.16042	51.90012

Relationship between wing length and width

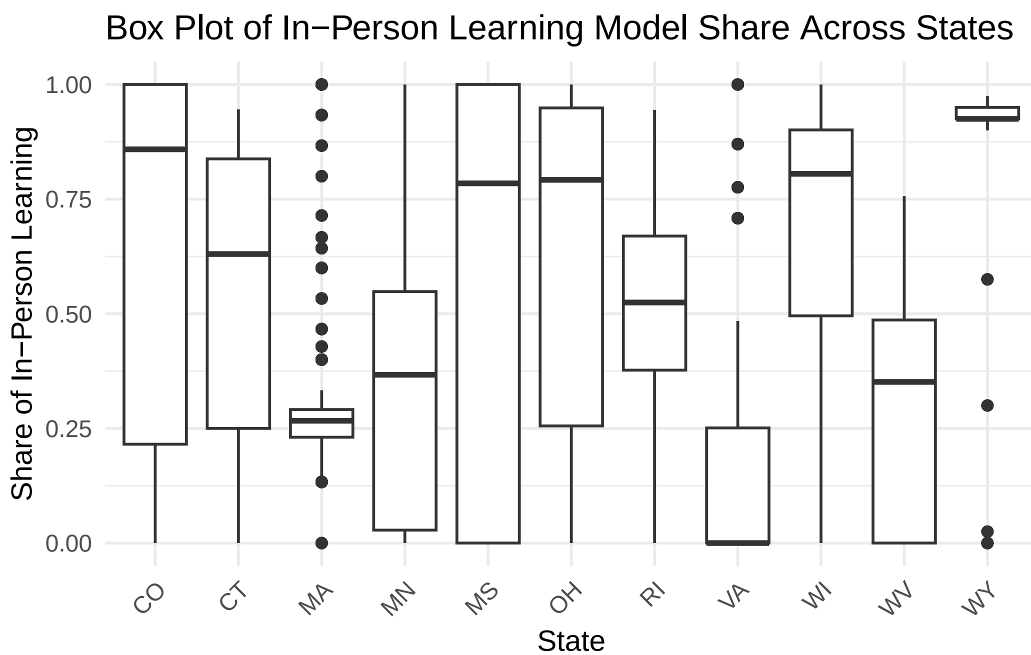

Figure 6: Relationship between wing length and width

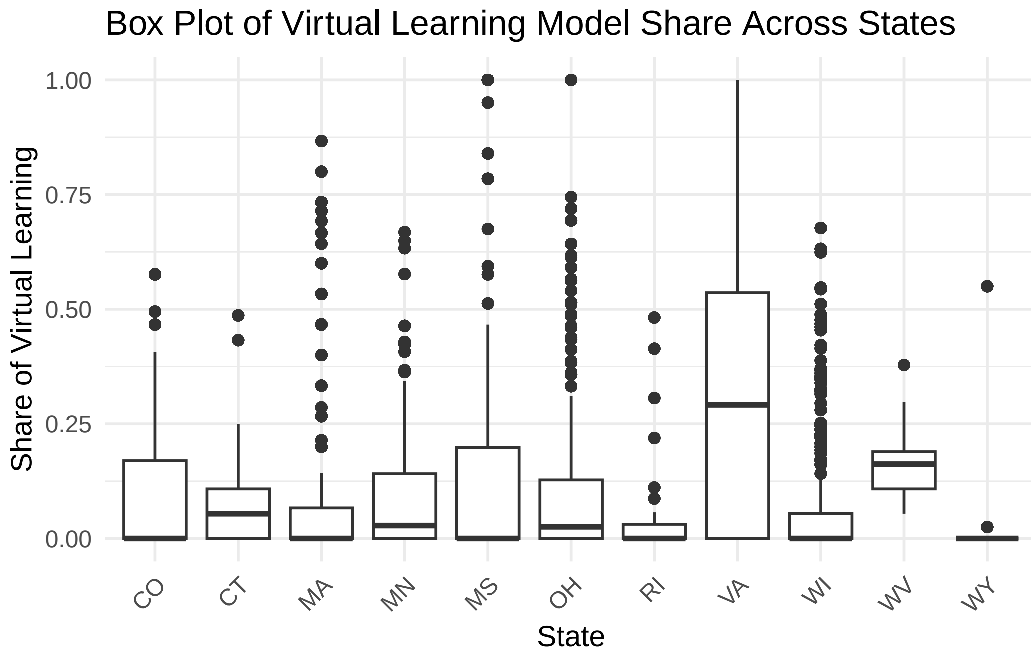

Figure 7: Relationship between wing length and width

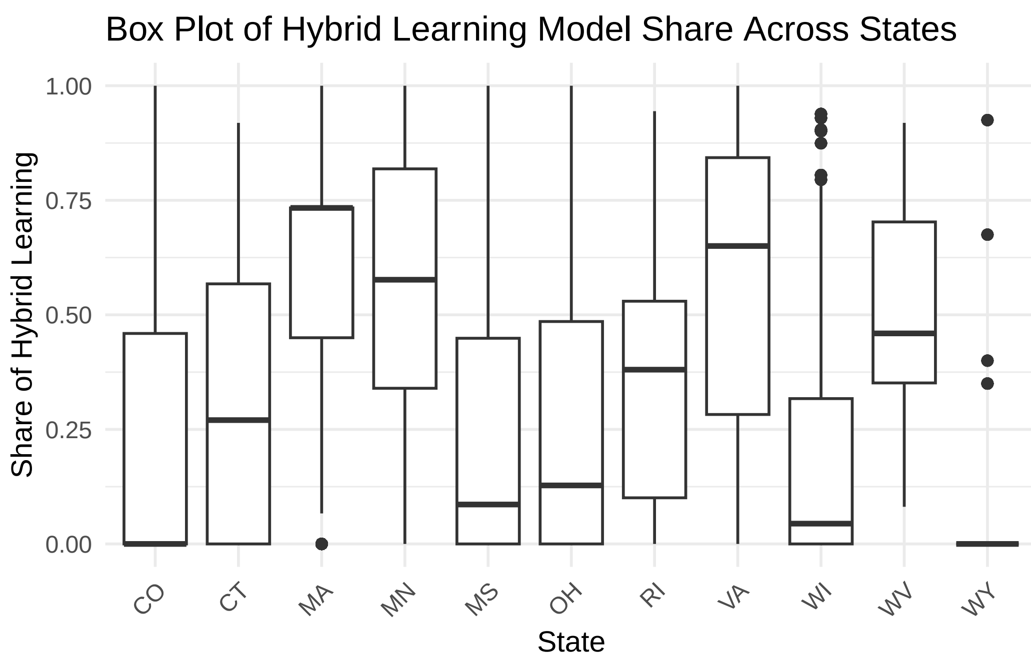

Figure 8: Relationship between wing length and width

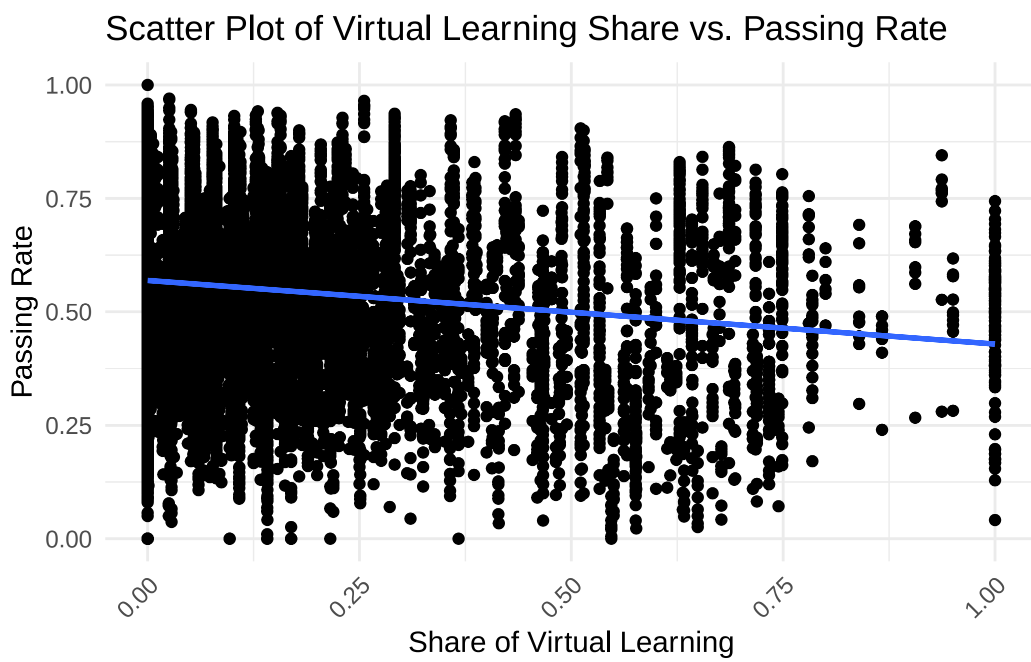

Figure 9: Relationship between wing length and width

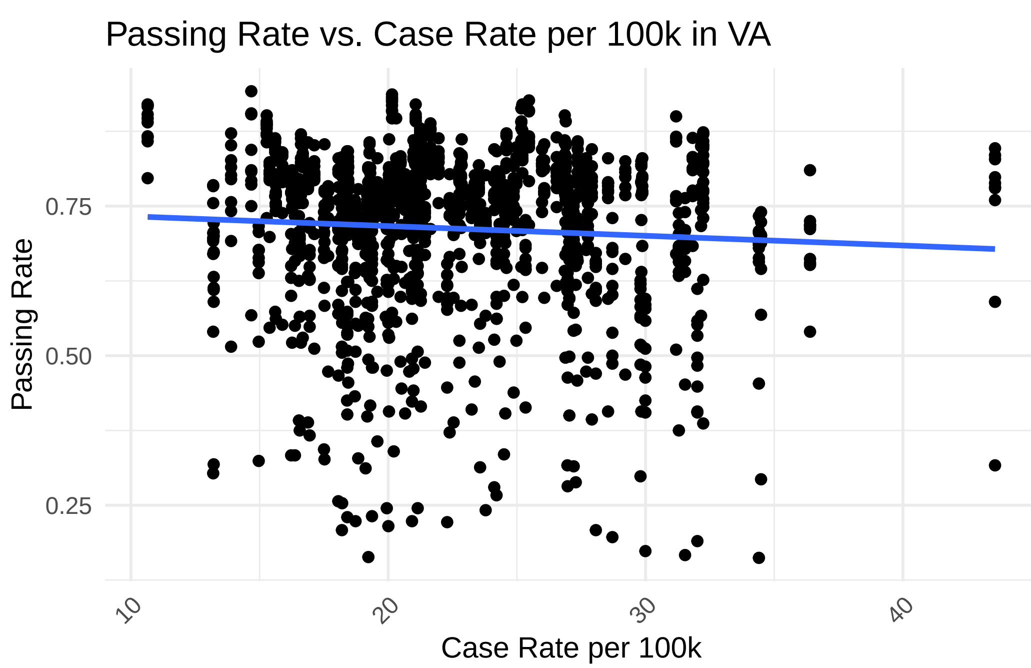

Figure 10: Relationship between wing length and width

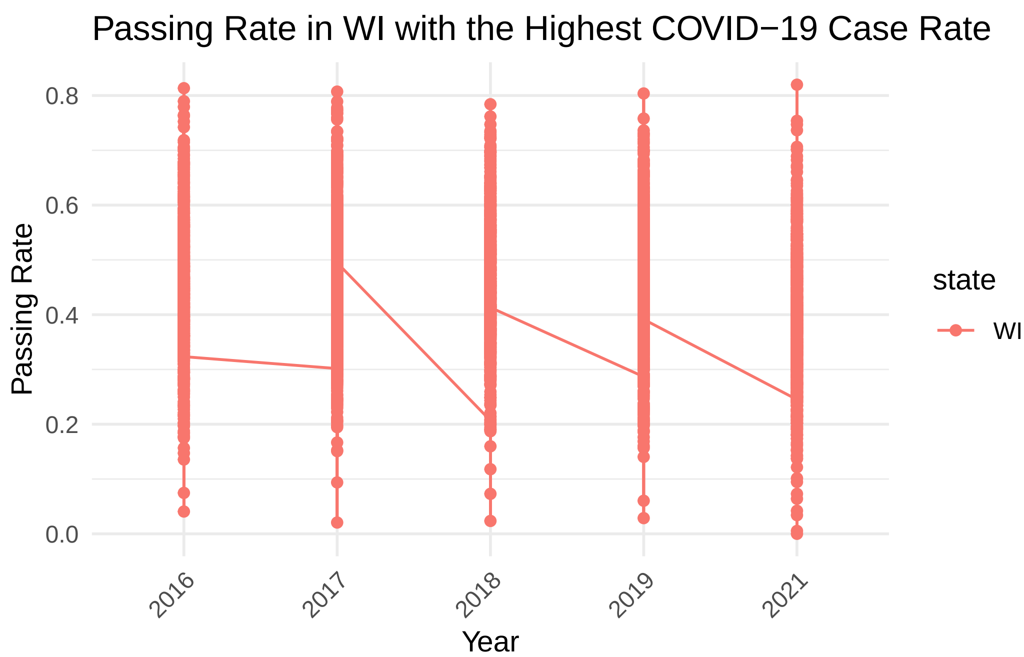

Figure 11: Relationship between wing length and width

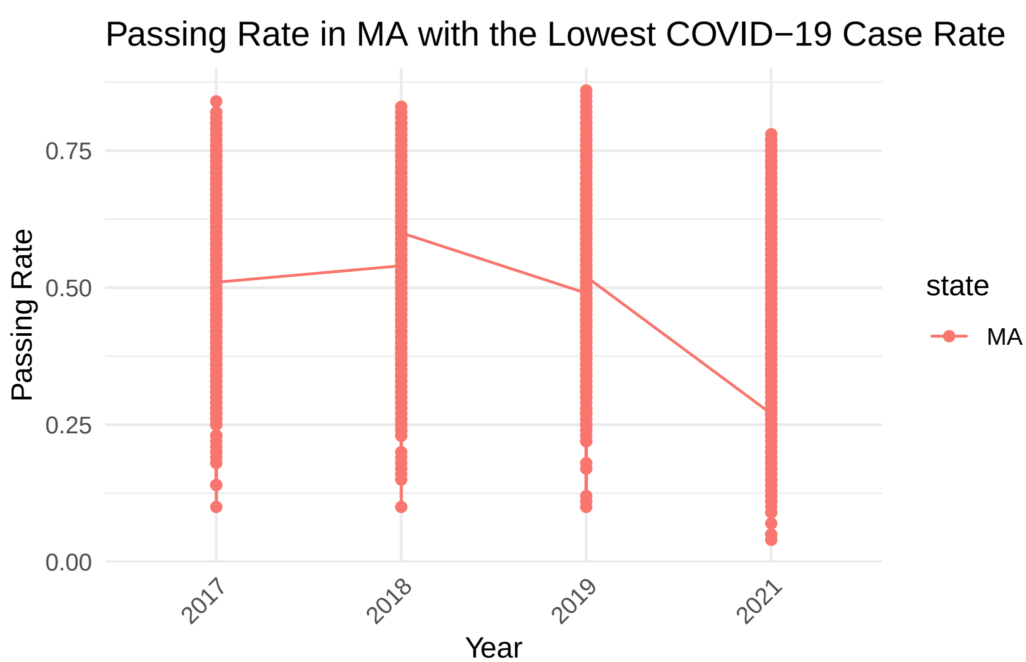

Figure 12: Relationship between wing length and width

4 Model

The goal of our modelling strategy is twofold. Firstly,...

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in Appendix B.

4.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$
\begin{aligned}
y_i|\mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) && (1)\\
\mu_i &= \alpha + \beta_i + \gamma_i && (2)\\
\alpha &\sim \text{Normal}(0, 2.5) && (3)\\
\beta &\sim \text{Normal}(0, 2.5) && (4)\\
\gamma &\sim \text{Normal}(0, 2.5) && (5)\\
\sigma &\sim \text{Exponential}(1) && (6)
\end{aligned}
$$

We run the model in R (R Core Team 2022) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

4.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, instance θ.

5 Results

Our results are summarized in **?@tbl-modelresults**.

6 Discussion

6.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

6.2 Second discussion point

6.3 Third discussion point

6.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows...

In **?@fig-ppcheckandposteriorvsprior-2** we compare the posterior with the prior. This shows...

References

AEA. 2022. *The American Economic Association Portal*. *The American Economic Association*. https://www.aeaweb.org/.

Bureau, United States Census. 2020. *Household Pulse Survey*. *The American Economic Association*. https://www.census.gov/data-tools/demo/hhp/#/?periodSelector=12&measures=JOBLOSS.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Jack, Rebecca. 2023. "Pandemic Schooling Mode and Student Test Scores: Evidence from US School Districts." *American Economic Review* 5 (2): 173–90. https://doi.org/10.1257/aeri.20210748.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

USA. 2022. *COVID-19 School Data Hub*. https://www.covidschooldatahub.com/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.