My title*

My subtitle if needed

First author Another author

February 13, 2024

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

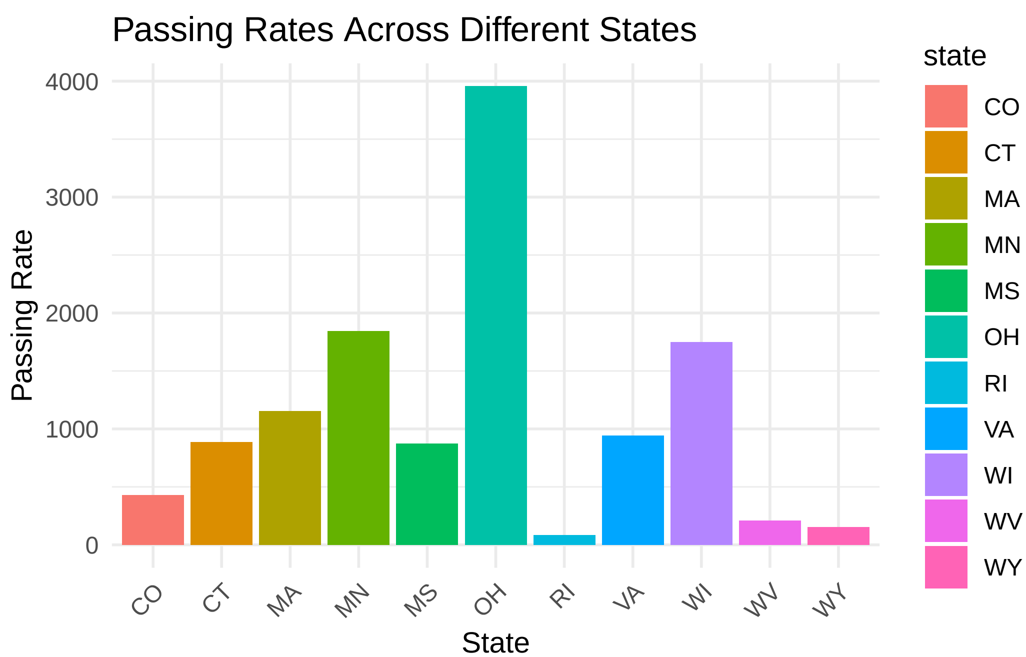

Figure 1: lets see

```
Warning: Removed 6882 rows containing missing values (`position_stack()`).
```

```
Warning: Removed 8192 rows containing missing values (`position_stack()`).
```

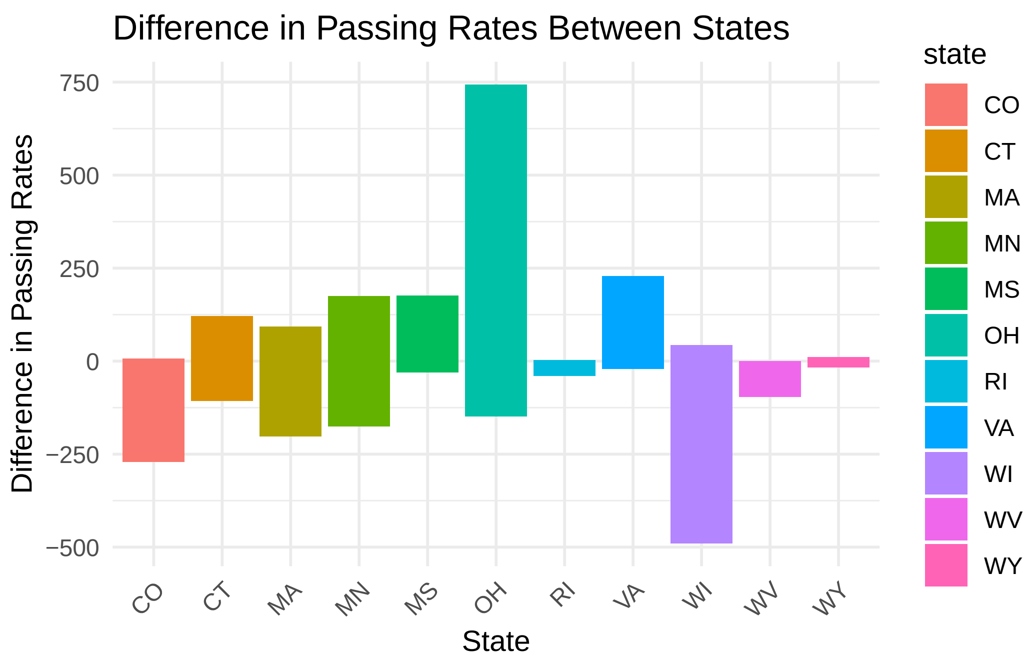

Figure 2: lets see

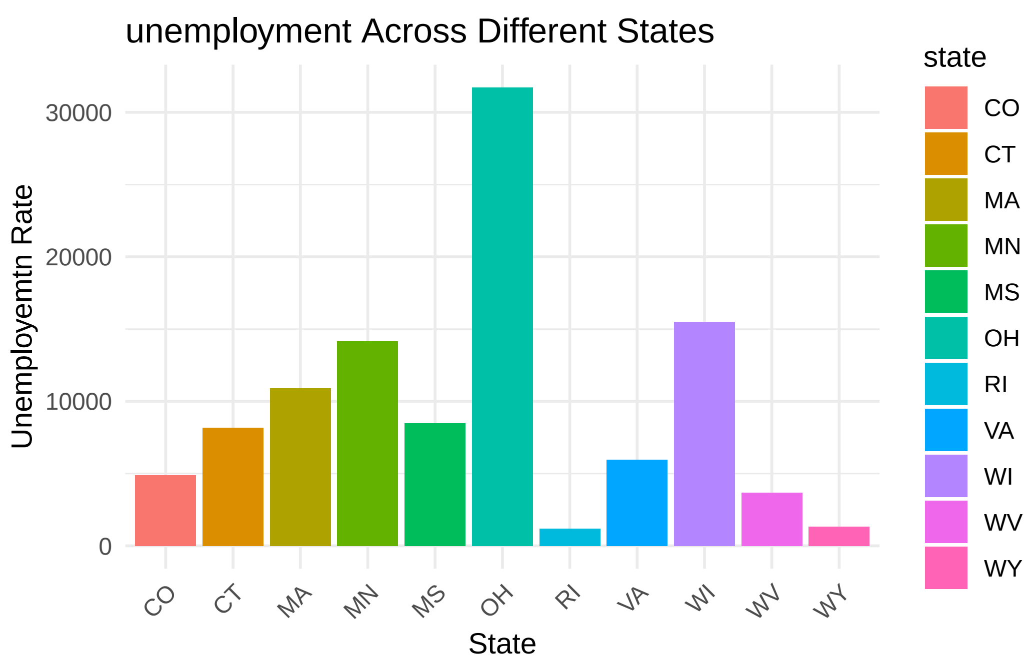

Figure 3: lets see

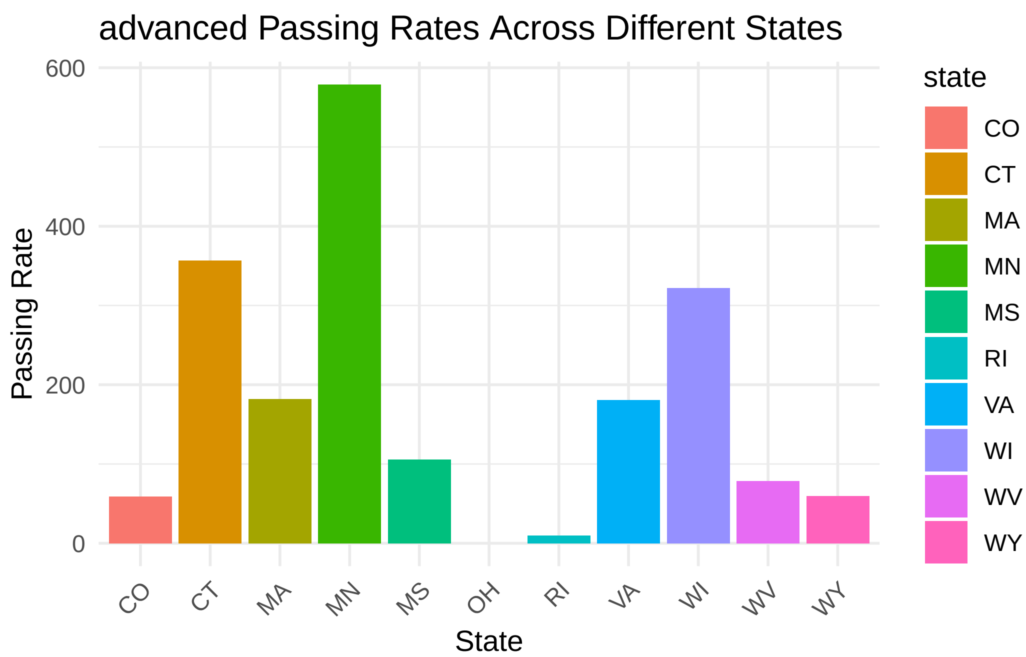

Figure 4: lets see

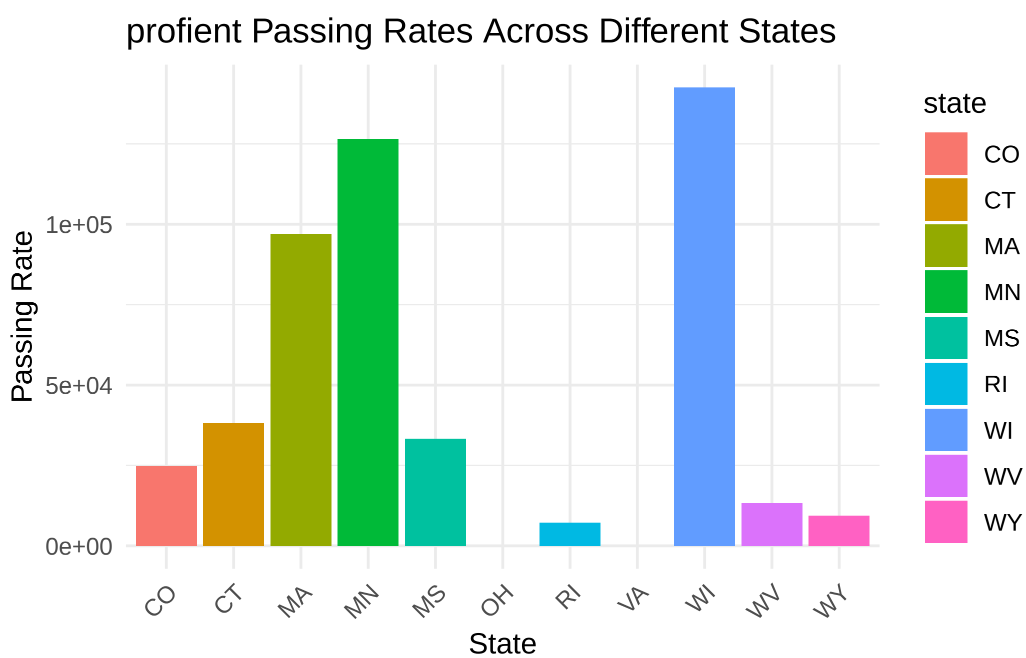

Figure 5: lets see

```
Warning: Removed 7979 rows containing missing values (`position_stack()`).
```

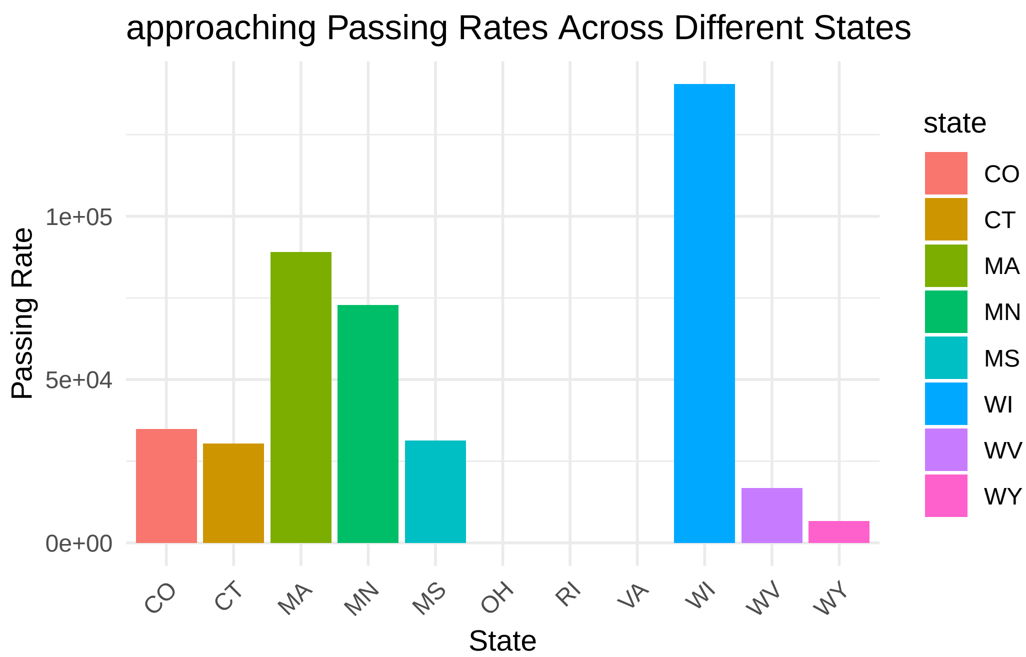

Figure 6: lets see

```
Warning: Removed 20981 rows containing missing values (`position_stack()`).

Warning: Removed 7873 rows containing missing values (`position_stack()`).

Attaching package: 'reshape2'

The following object is masked from 'package:tidyr':

    smiths

Warning: Removed 12 rows containing missing values (`geom_point()`).
```

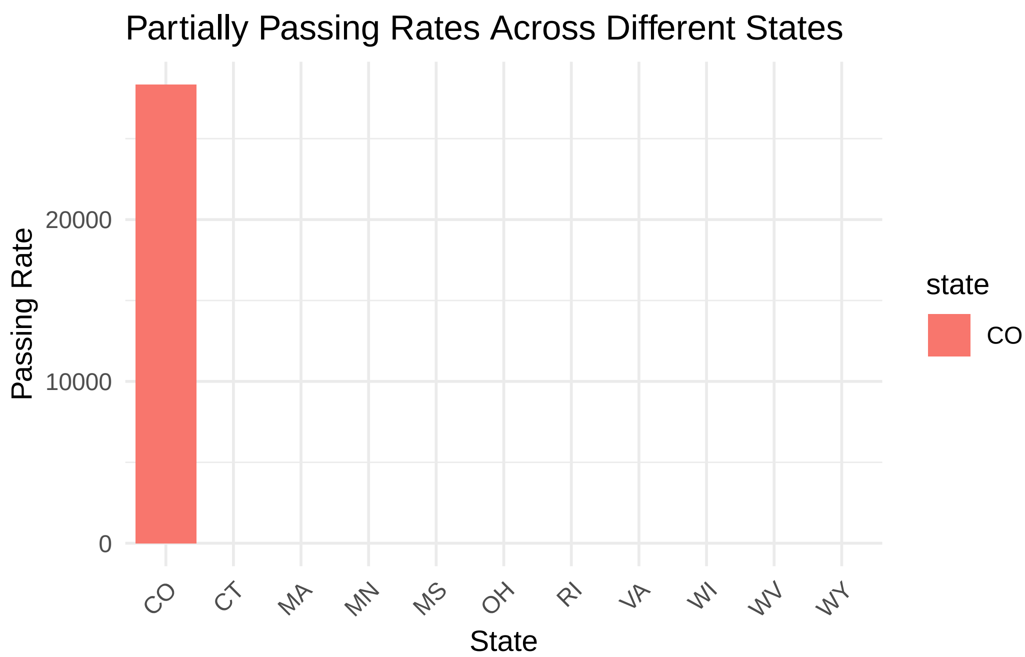

Figure 7: lets see

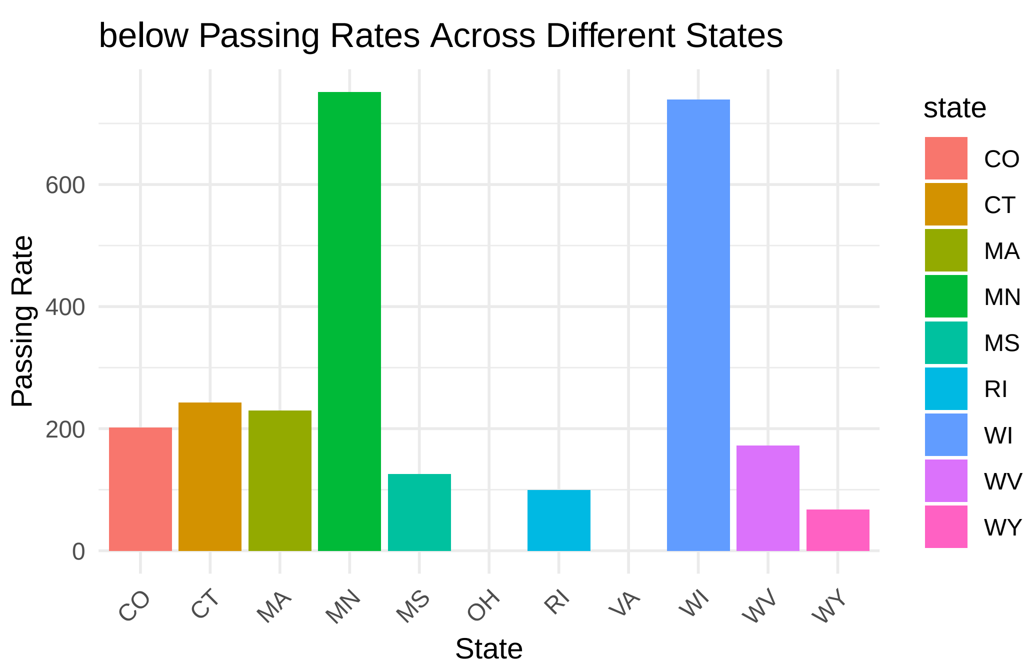

Figure 8: lets see

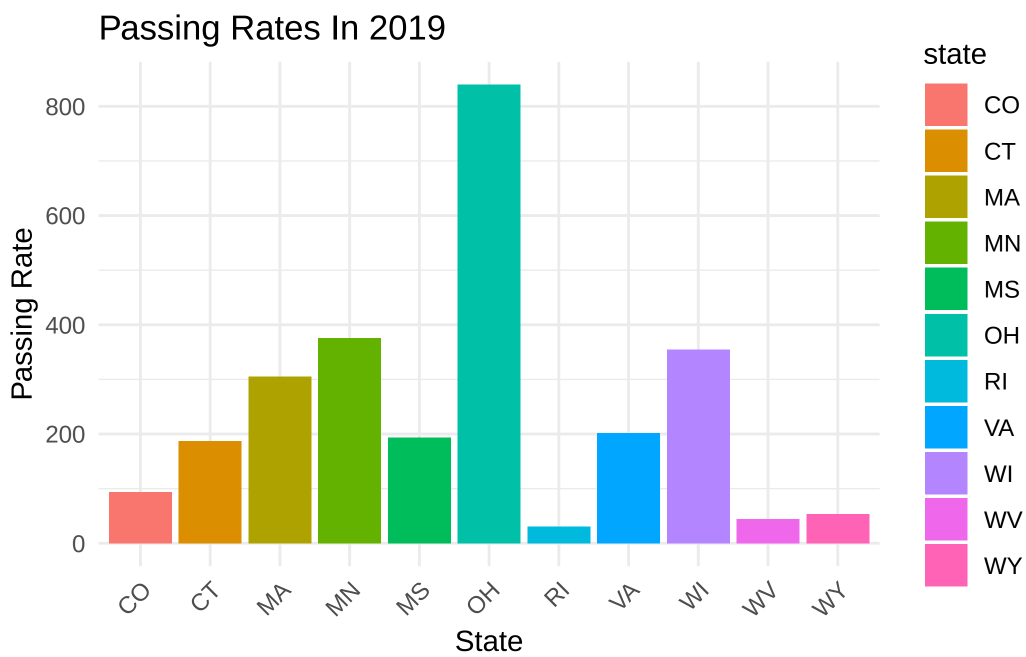

Figure 9: lets see

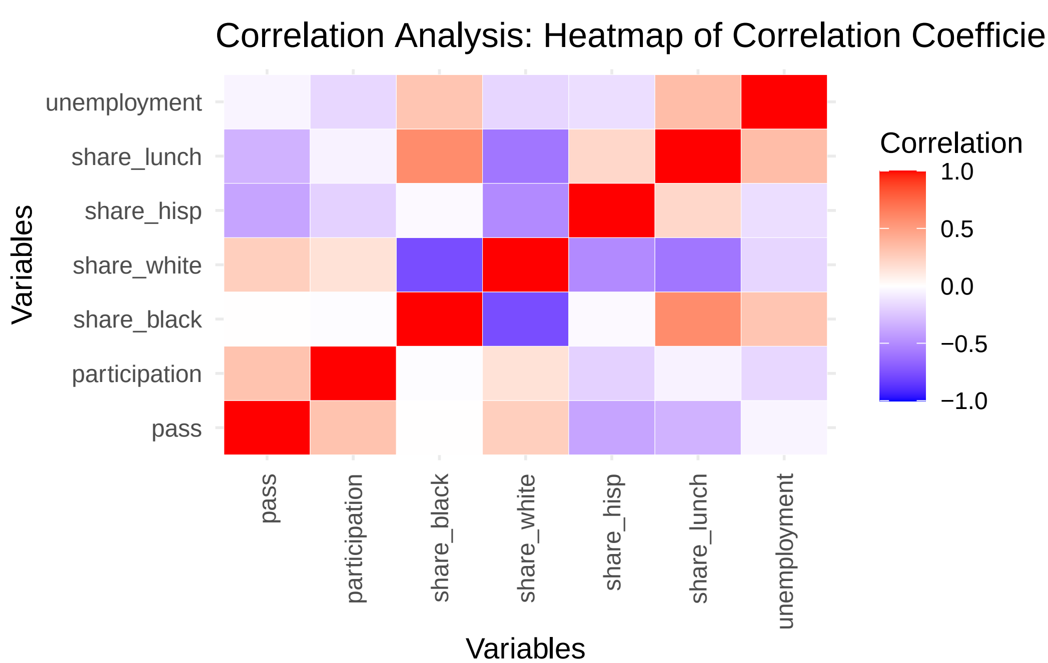

Figure 10: lets see

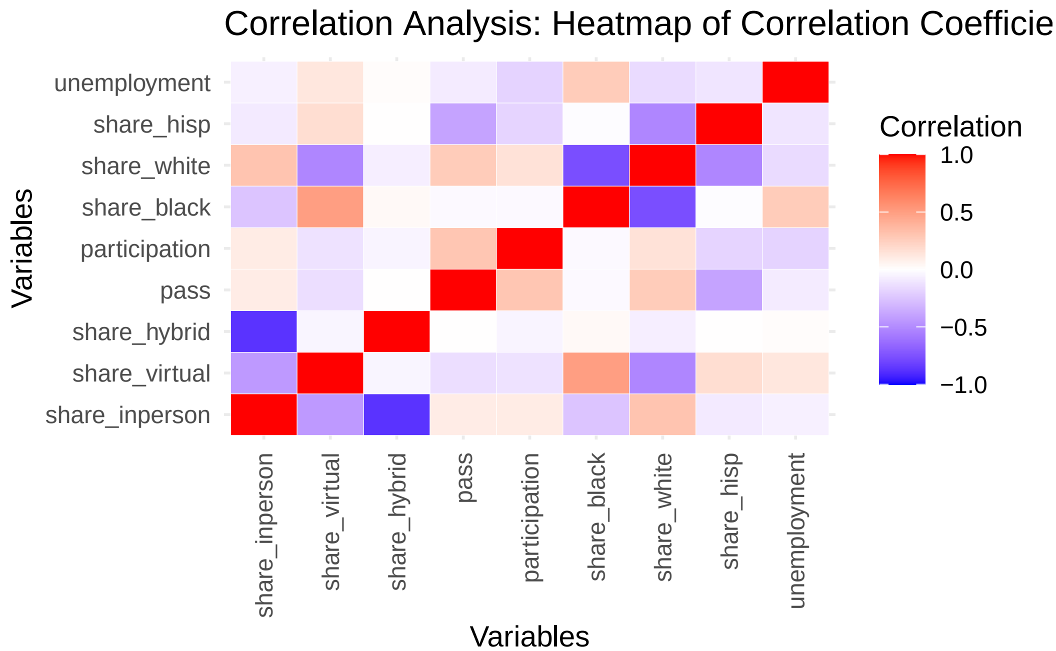

Figure 11: lets see

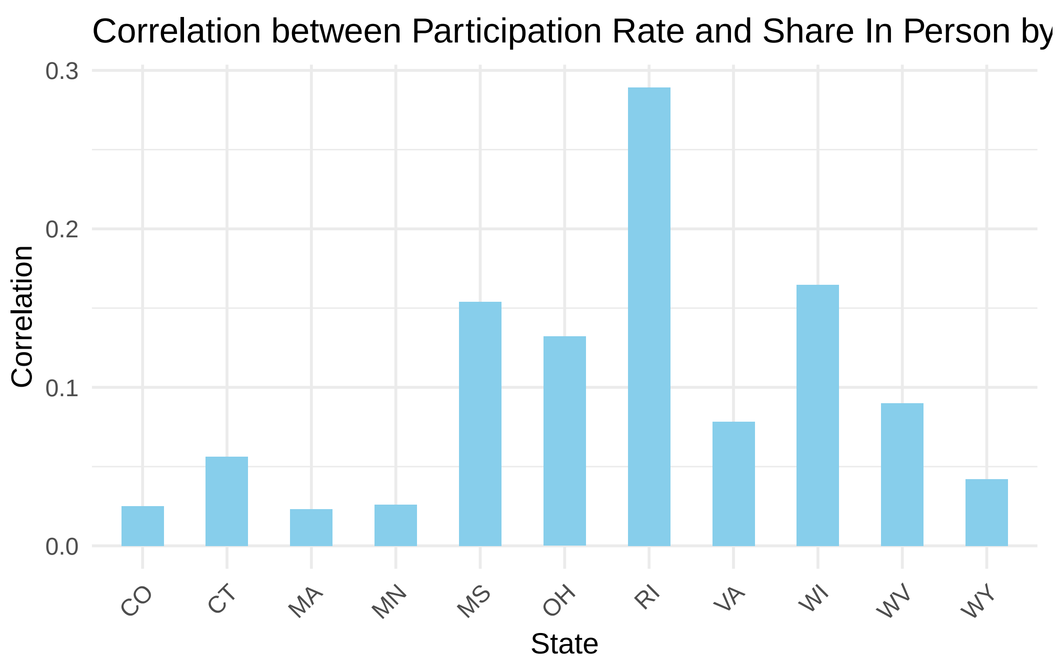

Figure 12: lets see

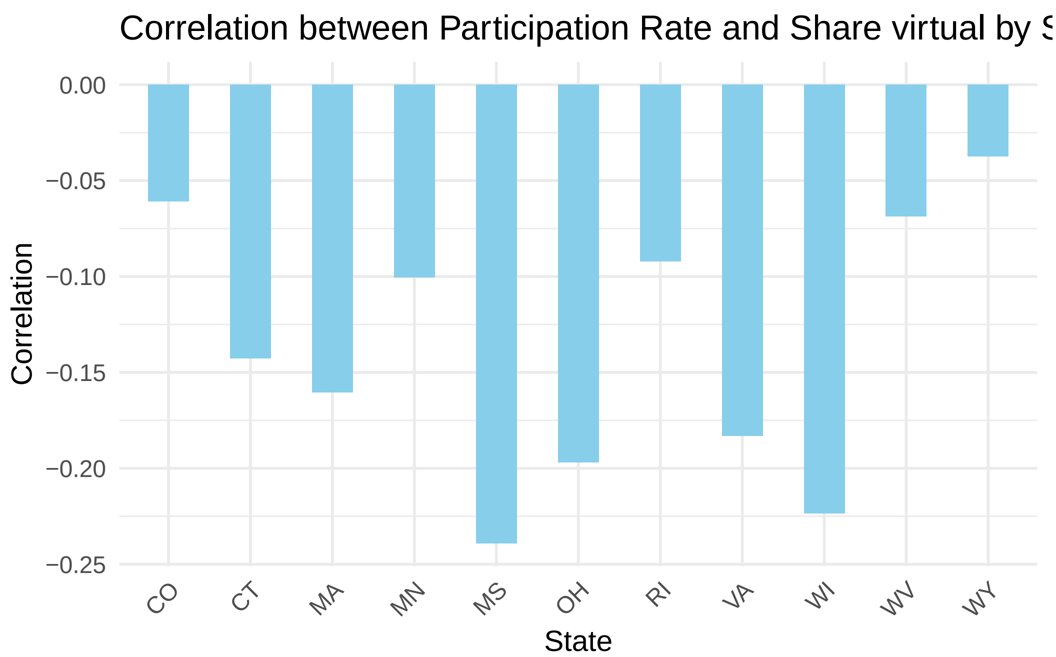

Figure 13: lets see

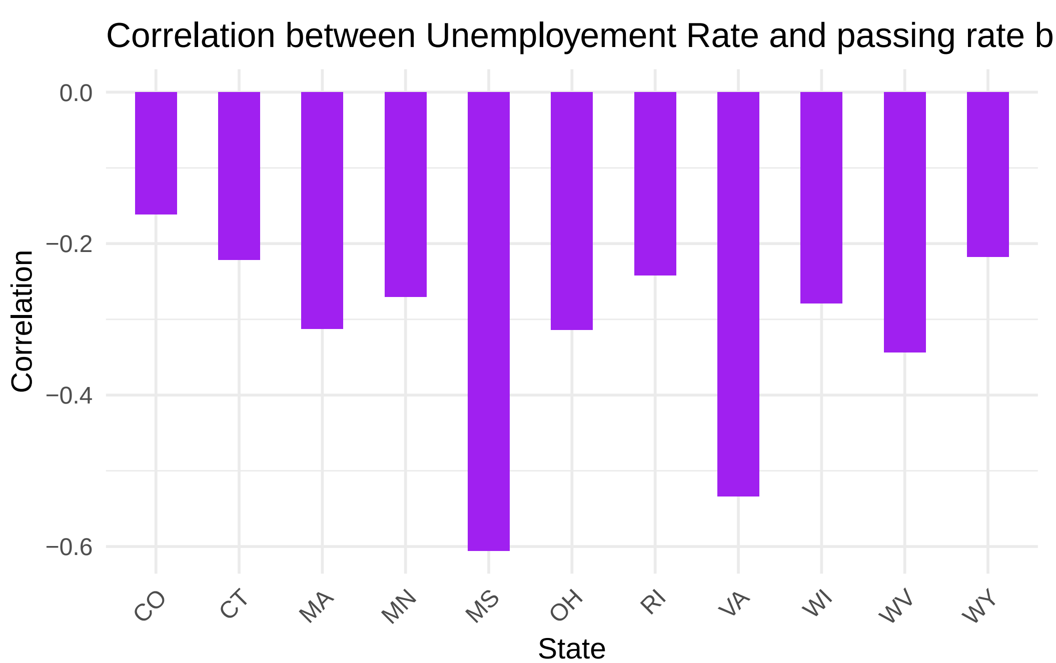

Figure 14: lets see

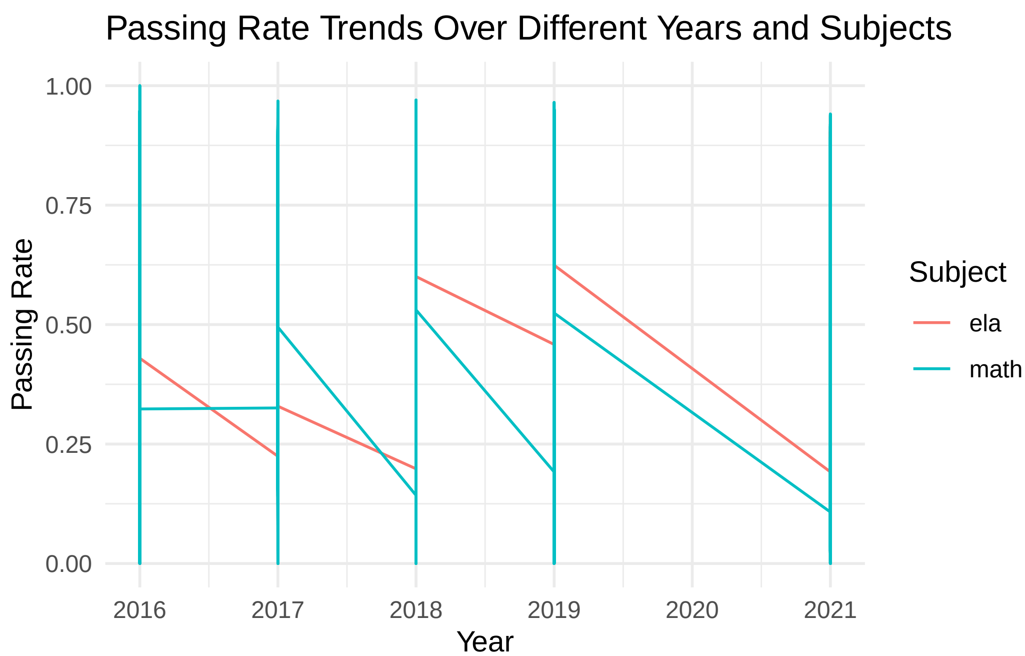

Figure 15: lets see

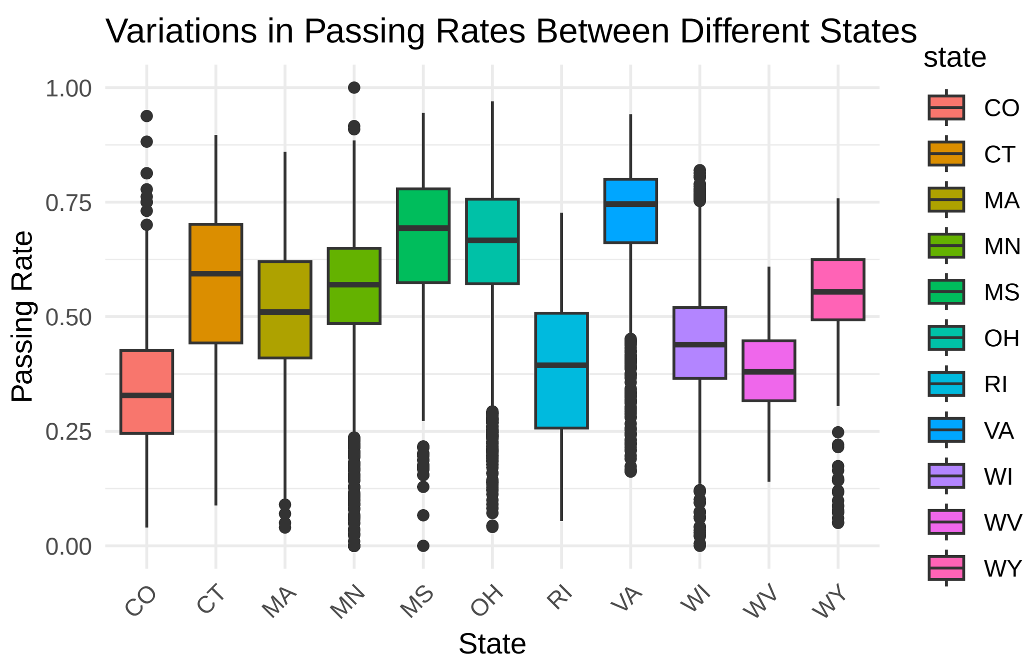

Figure 16: lets see

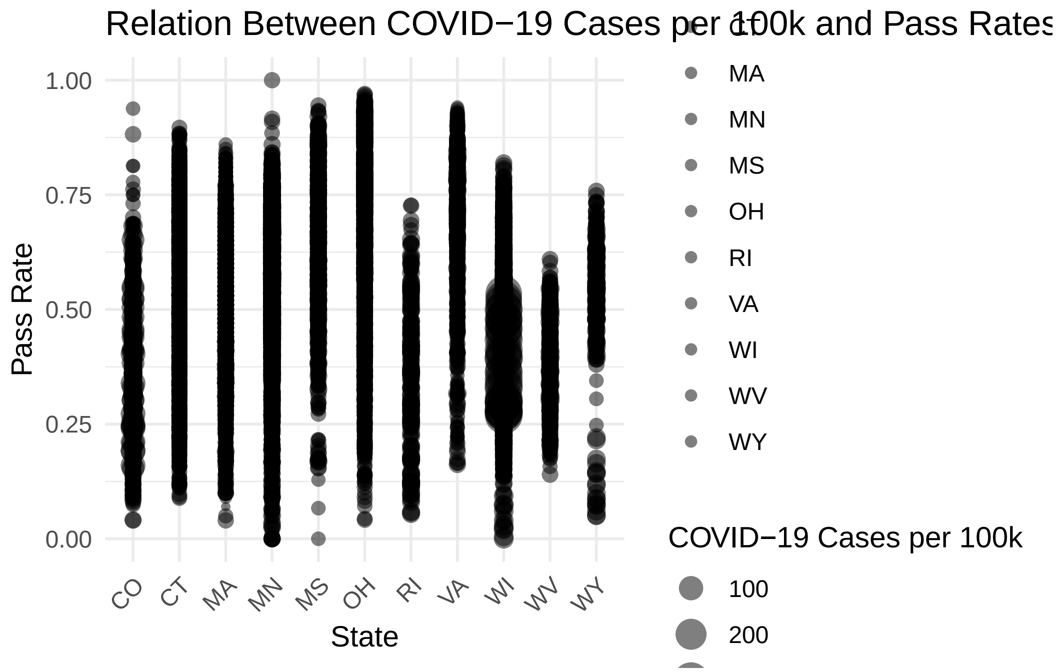

Figure 17: lets see

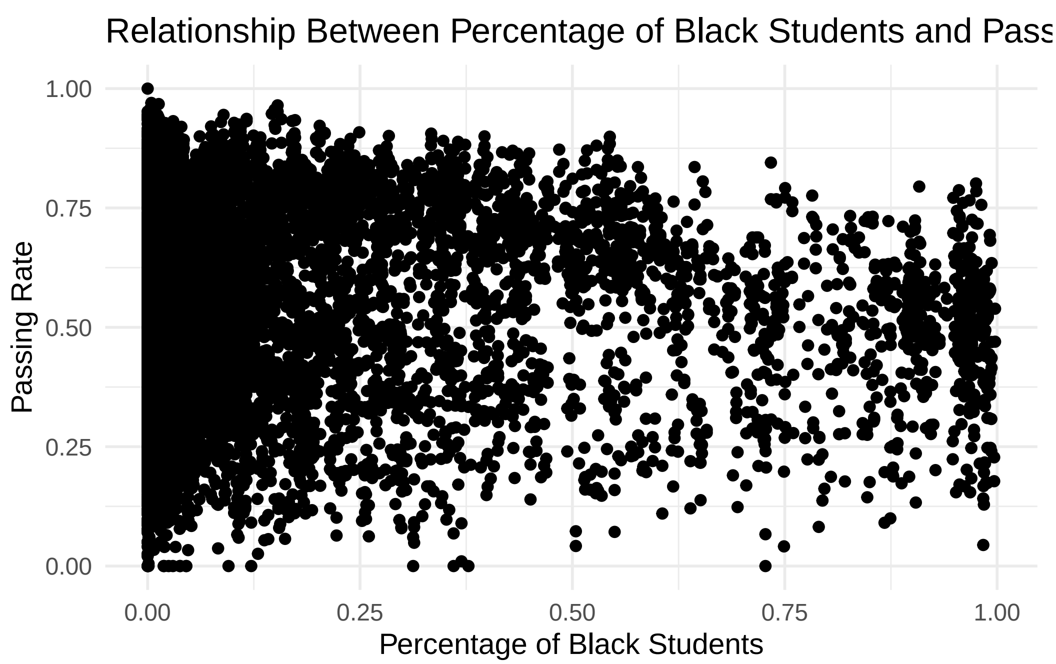

Figure 18: lets see

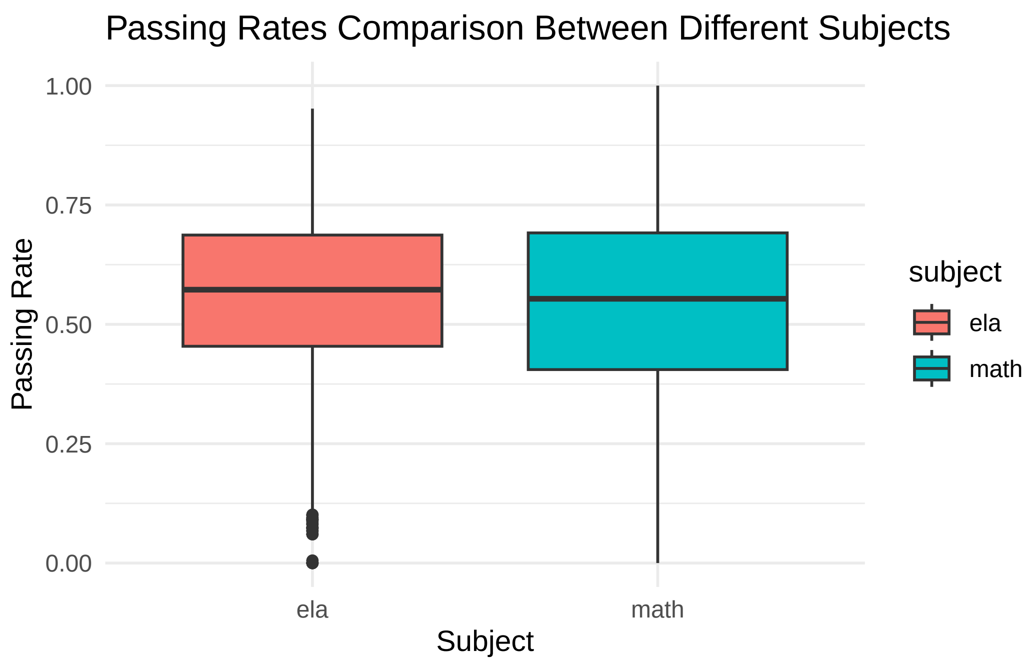

Figure 19: lets see

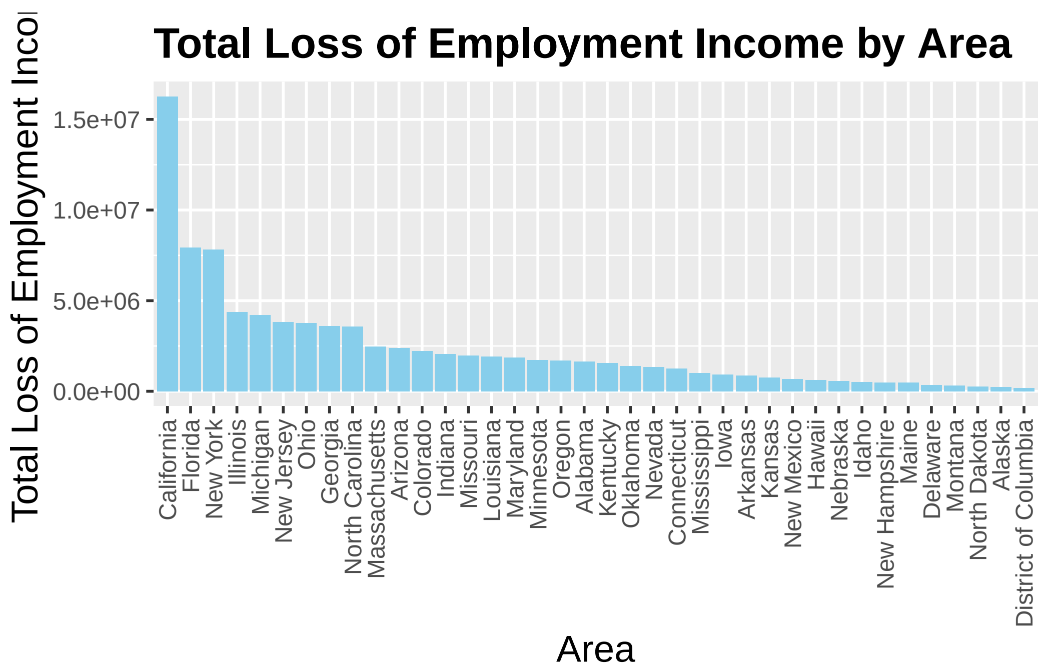

Figure 20: lets see

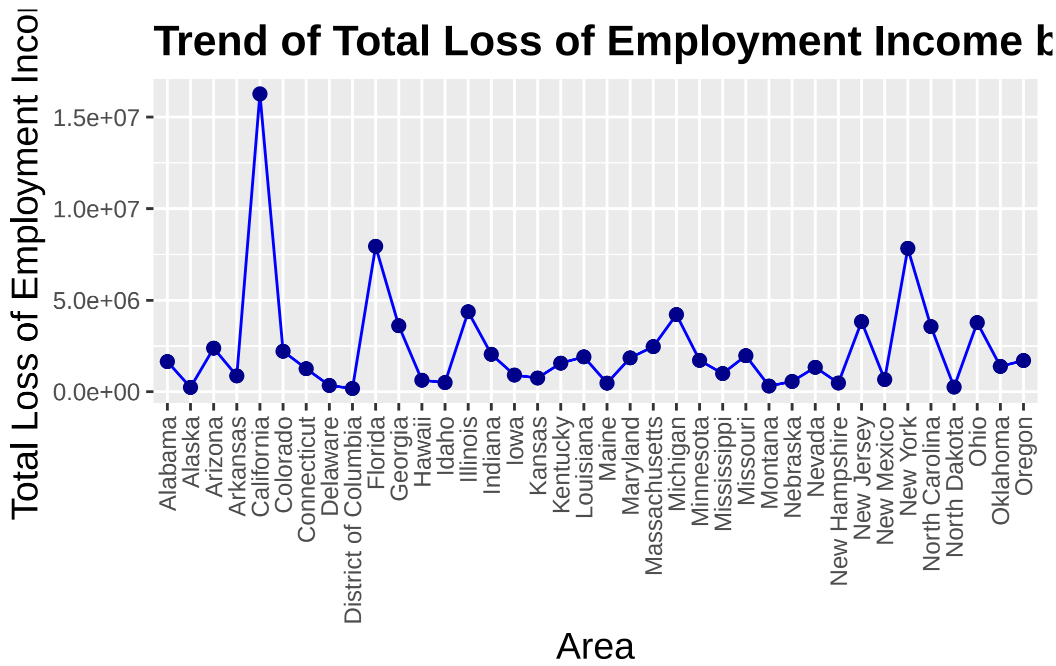

Figure 21: lets see

```
Warning: Removed 2 rows containing missing values (`geom_point()`).
```

```
`geom_smooth()` using formula = 'y ~ x'
```

1 Introduction

The COVID-19 pandemic has had a profound impact on education systems worldwide, leading to rapid adaptations in learning models across schools and districts in the United States. From 2020 to 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 in learning models in unknown circumstances led to severe decline in learning rate and knowledge of the students. As reported by NAEP, from 2019 to 2022, the average performance on the National Assessment of Educational Progress (NAEP) declined by eight points in eighth-grade math, representing approximately three-quarters of a grade level, and by three points in eighth-grade reading, equivalent to roughly one-quarter of a grade level. This impacted every student in the US and their learning but all the factors

*Code and data are available at: LINK.

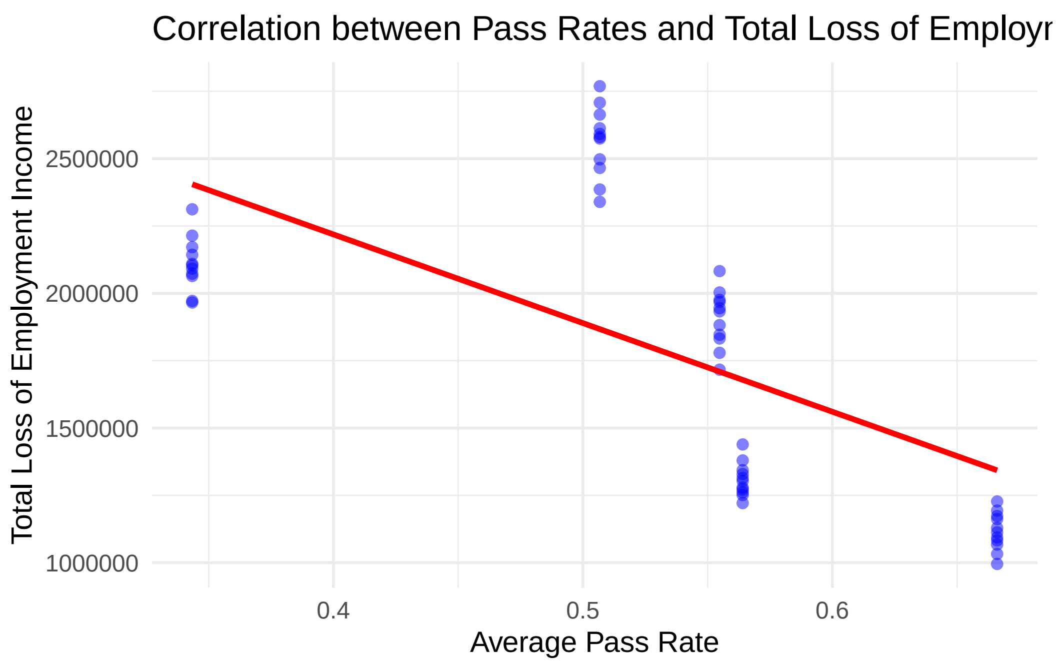

Figure 22: lets see

that led to this large impact is still not well understood. In this paper, we embark on a comprehensive analysis spanning 11 U.S. states, with a keen focus on explaining the connection between various learning models and students' "learning loss," as measured by their passing and participation rates, emblematic of missed educational opportunities. Our objective is to examine and closely study the many reasons behind students' learning problems. Initially, we harness a bespoke dataset tailored for this study, encompassing metrics such as total enrollment, in-person and virtual attendance, unemployment rate, races, voting data, and staff count across diverse learning modalities from kindergarten to 12th grade. Through examination, we show the transformative impact of the pandemic on the educational landscape across these states. Our analytical framework understands patterns of various factors that contributed to the decline in test scores between 2019 and 2021, shedding light on the nuanced associations between students' success on state exams and various factors such as learning modalities, staff count, unemployment rates, and COVID-19 caseloads. Subsequently, we delve deeper into the origin of this decline by leveraging an extensive dataset from the Household Pulse Survey, probing into myriad household dynamics that may have precipitated students' lower academic performance and states' diminished passing percentages. Our findings reveal that the learning losses incurred during the pandemic were both substantial and highly heterogeneous across communities. We contend that these losses cannot be attributed to a singular cause but rather stem from a confluence of diverse factors. Notably, the depletion of household income during the pandemic exhibited a direct correlation with learning loss, with mathematics suffering a more pronounced decline compared to ELA courses. Additionally, we elucidate the correlation between passing rates in each state and the type of learning modality adopted, demonstrating the significant impact of in-person, hybrid, or virtual models on children's educational trajectories. Furthermore, we underscore the pronounced effect of varying COVID-19 caseloads on learning outcomes, with states experiencing higher infection rates manifesting greater learning losses compared to their counterparts with lower case burdens.

In closing, it's evident that understanding the complexities of the pandemic's impact on education is crucial for shaping future strategies. By understanding the complex web of factors influencing student learning during this time, we can better equip schools and districts to address these challenges and ensure that all students have the support they need to thrive academically and beyond.We can empower the education system to mitigate the effects of learning loss and ensure a more successful future for students. It's imperative that we learn from these challenges to prevent the continuation of the learning loss trend for years to come, fostering a more resilient and equitable educational environment for all.

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" (**citepaper?**) published in the American Economic Review: Insights in June 2023. (**citewebsite?**). 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 (**citedata?**), 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 (**citehousehold?**) 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 (**tidyverse?**) for data manipulation and janitor (**citejan?**) 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 developed a unique 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

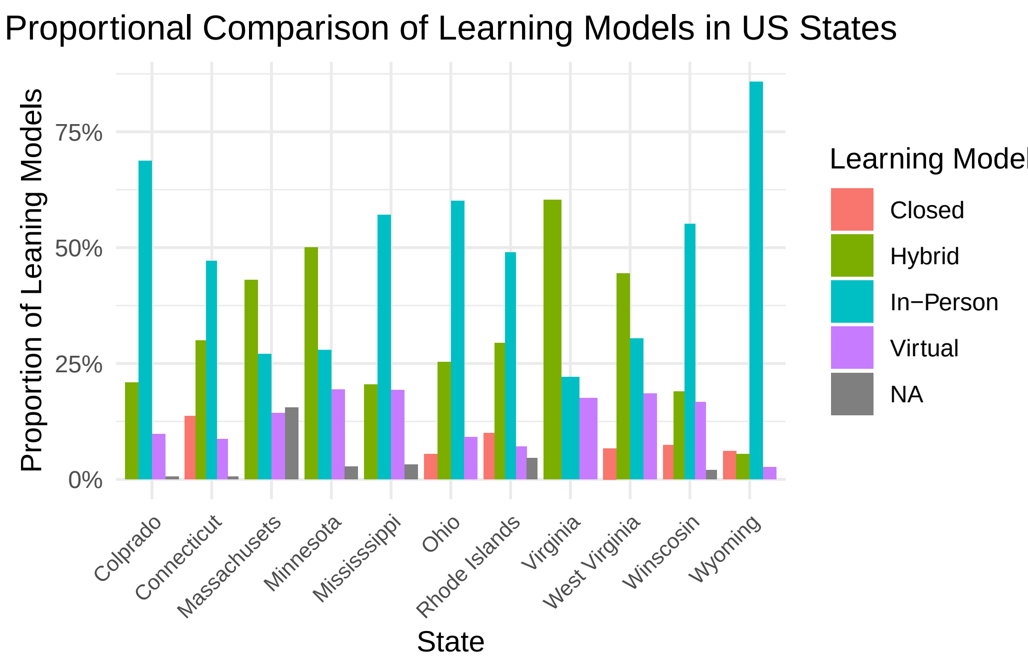

Figure 23: Bills of penguins

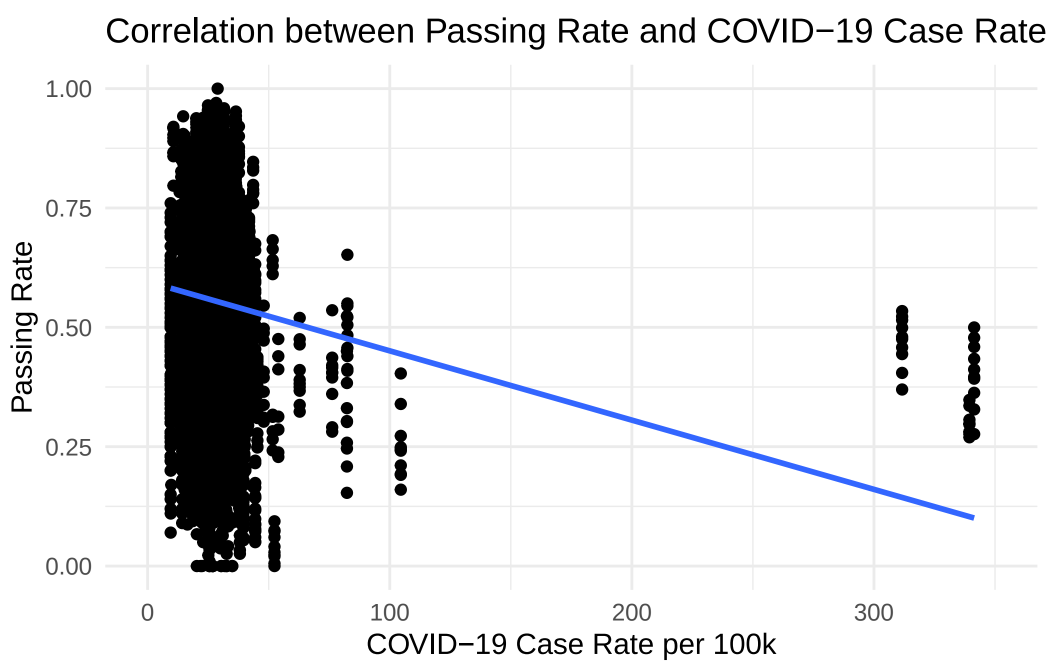

Figure 24: Bills of penguins

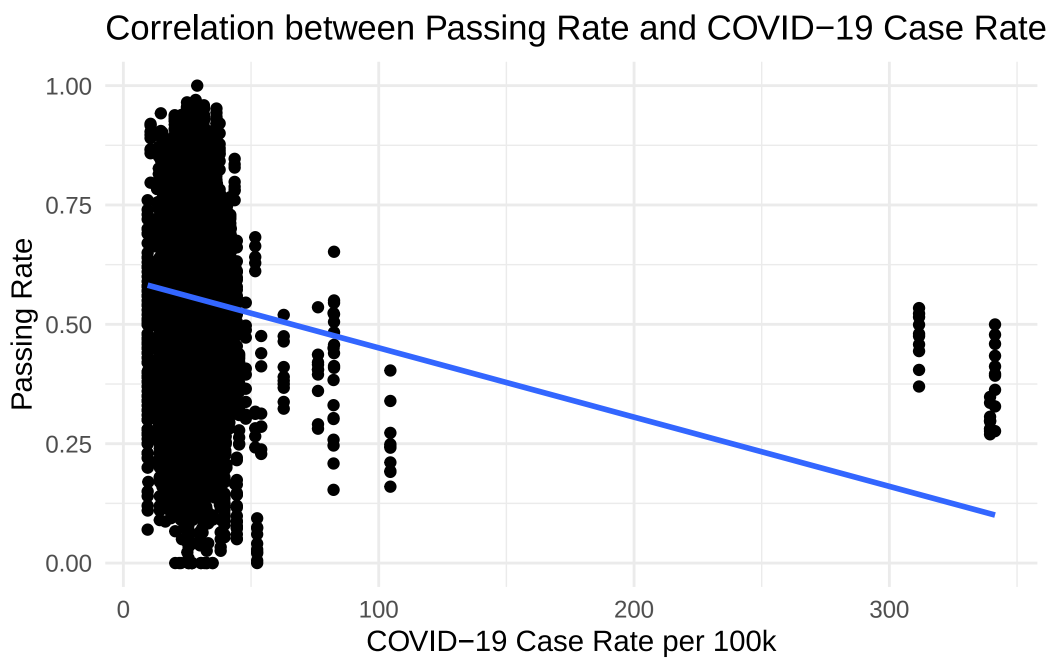

Figure 25: Bills of penguins

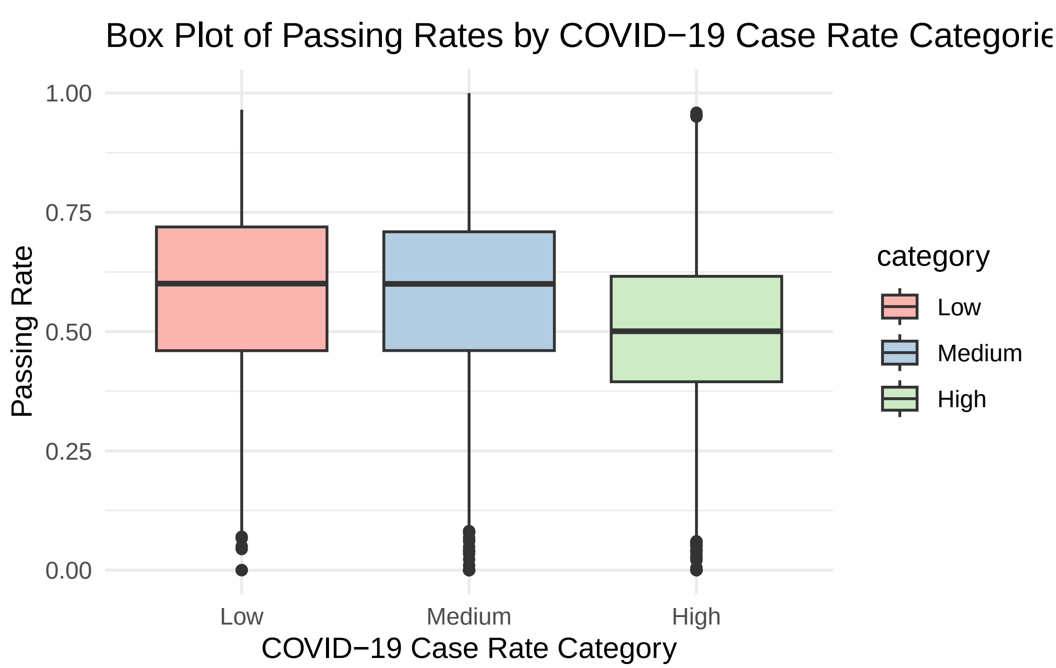

Figure 26: Bills of penguins

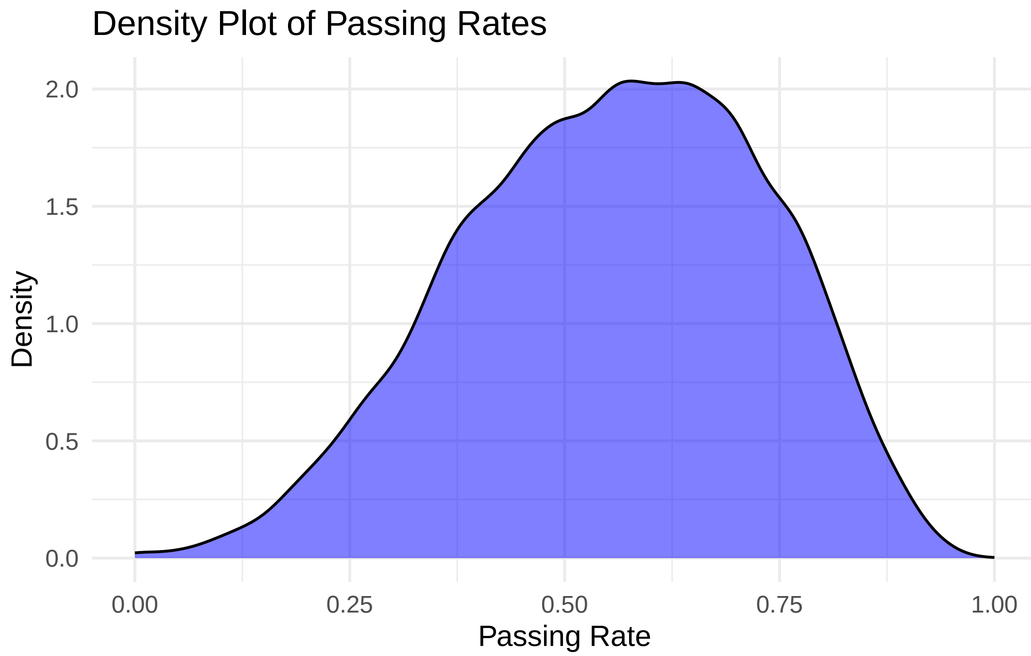

Figure 27: Bills of penguins

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.

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

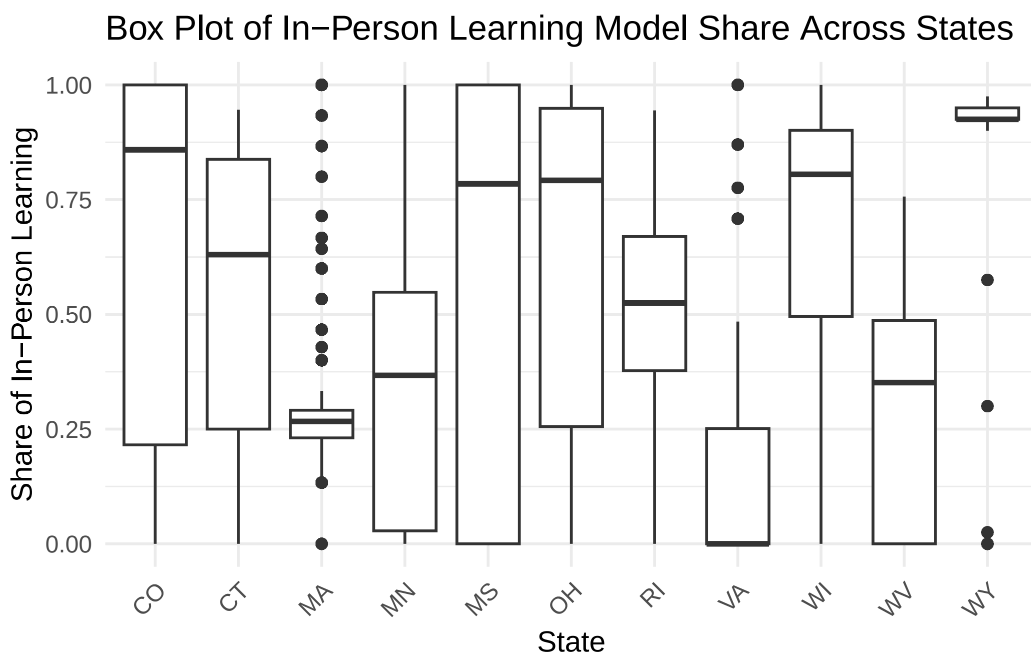

Figure 28: Relationship between wing length and width

Talk way more about it. Some of our data is of penguins (**?@fig-bills**), from Horst, Hill, and Gorman (2020).

Talk more about it.

And also planes (**?@fig-planes**). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

Talk way more about it.

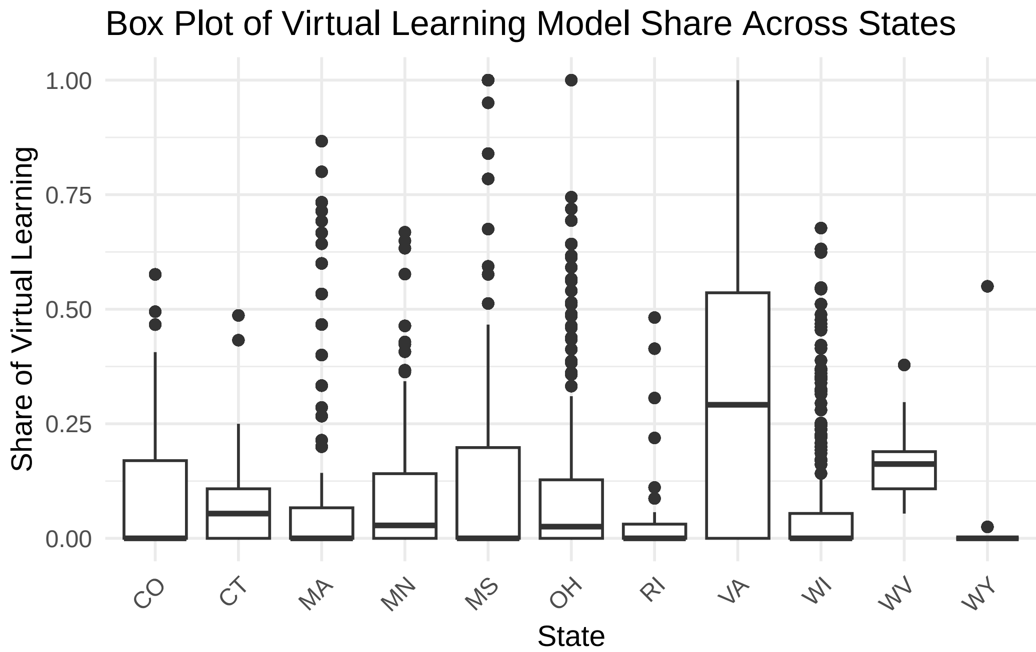

Figure 29: Relationship between wing length and width

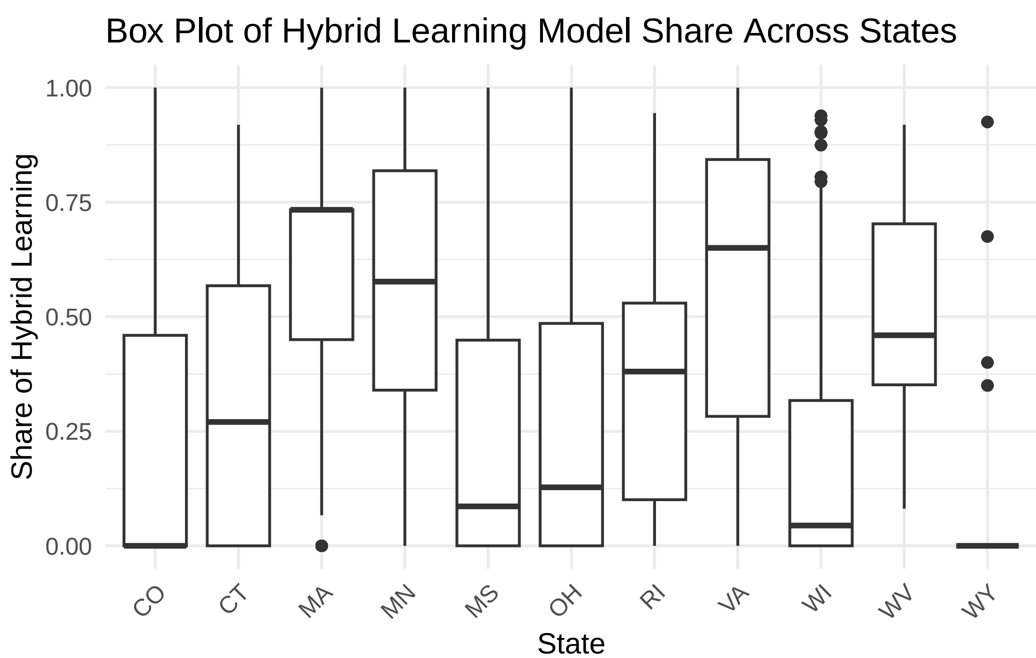

Figure 30: Relationship between wing length and width

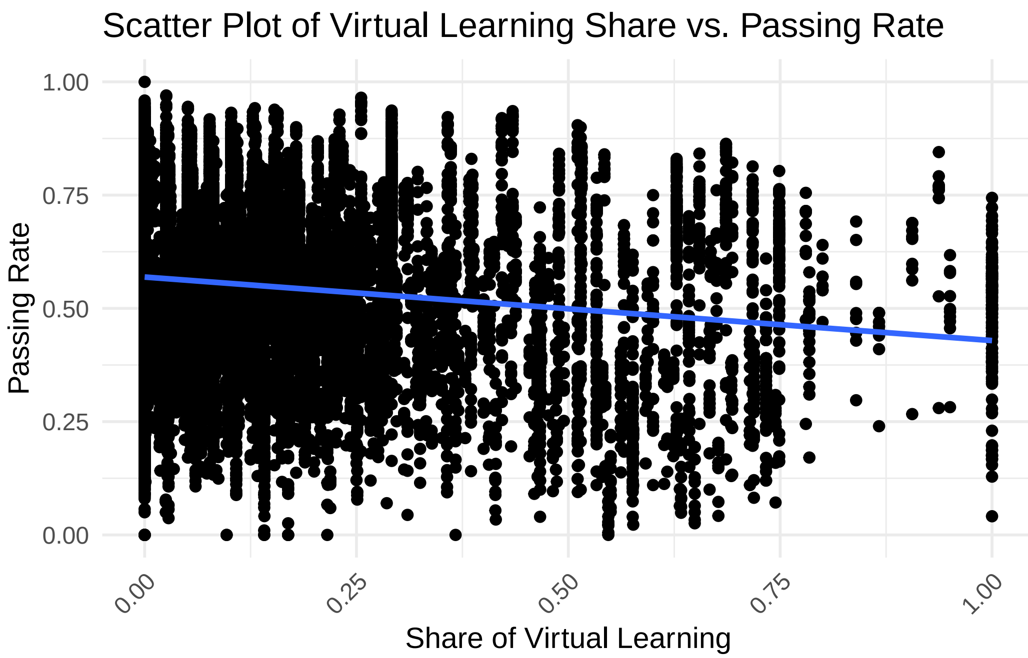

Figure 31: Relationship between wing length and width

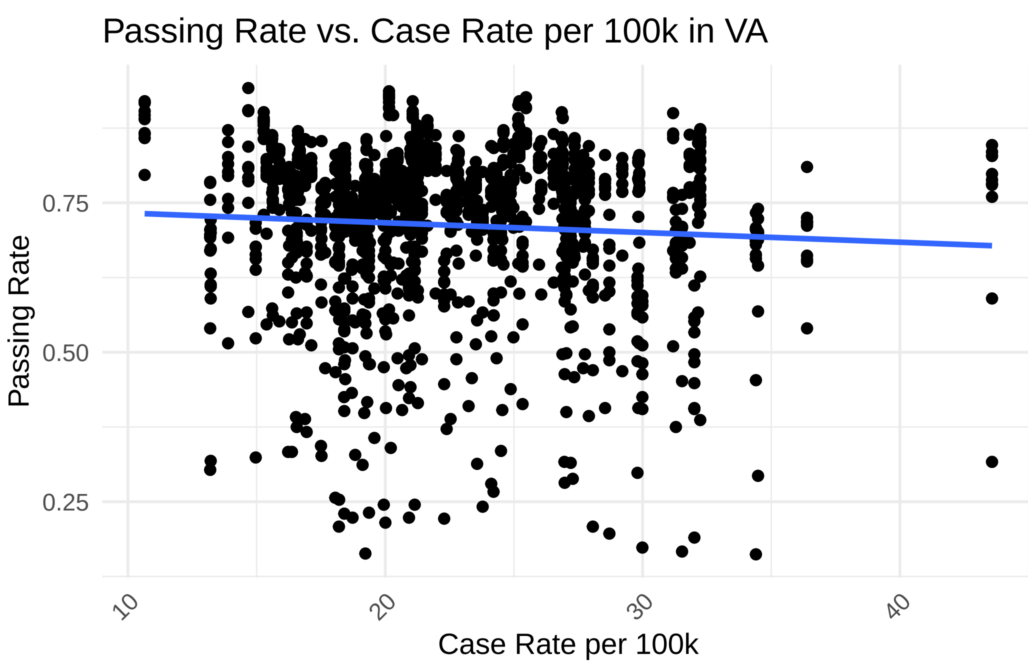

Figure 32: Relationship between wing length and width

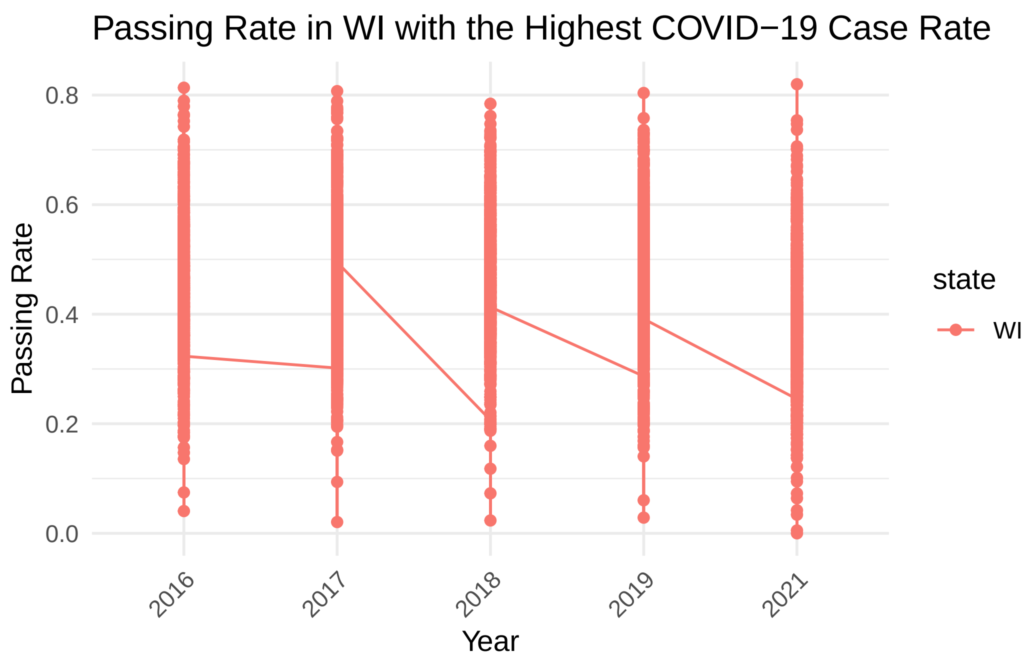

Figure 33: Relationship between wing length and width

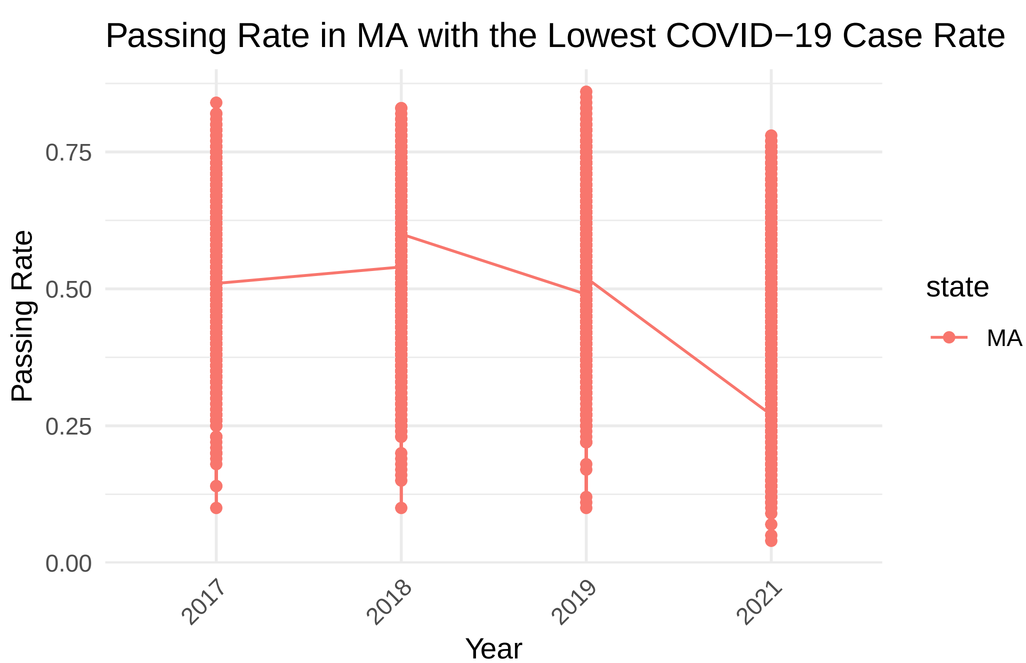

Figure 34: Relationship between wing length and width

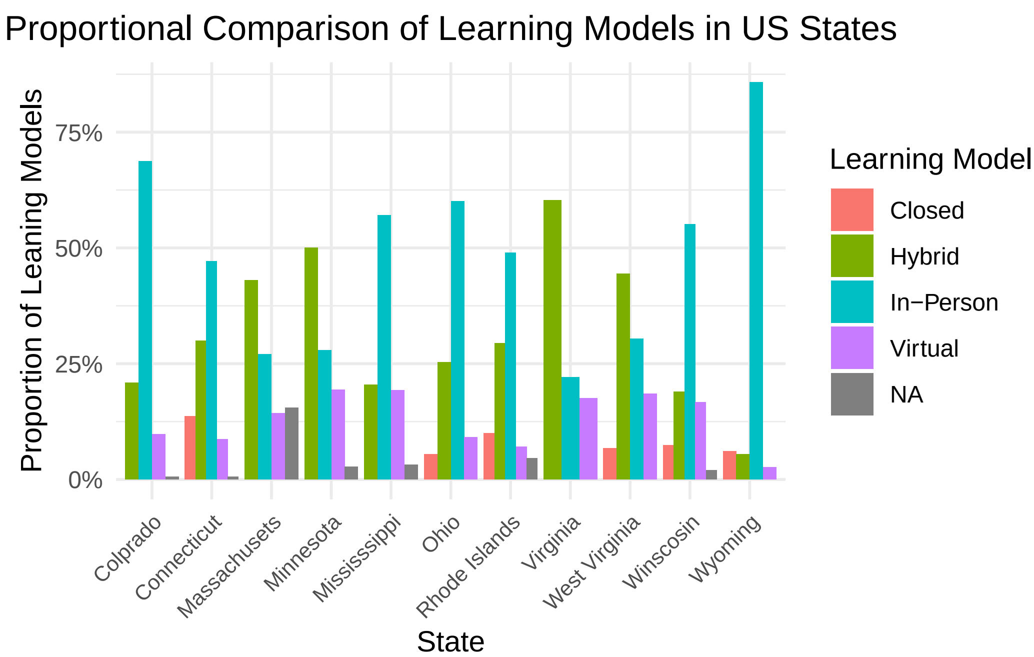

Figure 35: Bills of penguins

3 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.

3.1 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`.

3.1.1 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, for instance θ.

4 Results

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

5 Discussion

5.1 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.

5.2 Second discussion point

5.3 Third discussion point

5.4 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

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

Horst, Allison Marie, Alison Presmanes Hill, and Kristen B Gorman. 2020. *Palmerpenguins: Palmer Archipelago (Antarctica) Penguin Data*. https://doi.org/10.5281/zenodo.3960218.

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