

# Project Report

## Outline

- Project Report
  - Summarizing reading and research
  - Lit review/background
  - Dataset descriptions
    - \* Academic performance data
      - Explain their grade calculation
    - \* ACS Data
  - Methodology
    - \* Data prep + Data cleaning
    - \* Consideration for district coverage
  - Results
    - \* Description of the app
    - \* Takeaways + Conclusions
  - Discussion/Conclusion
    - \* Overall takeways answering the research question
      - The state of education post-pandemic
    - \* LImitations of my analysis/research
    - \* Future research steps
- Reflection on the project and experience

- Citing code/libraries
  - Only big ones like tidyverse and shiny

## References Test

I used the (Chang et al. 2023) package to create the interactive dashboard and tidyverse (Wickham et al. 2019) package for plotting, modeling, and displaying. I referenced several articles and used the academic performance data.

## Reading/Research/Background

The Covid-19 pandemic has had a profound effect on schools in the US. For many students, the transformation from live personal engagement to remote learning or worse, schools shutting down, has been difficult to adjust to. This struggle is evident when comparing academic performance metrics pre-pandemic to post-pandemic.

The primary research questions I wanted to answer were:

- What are the factors that influence the academic performance of school districts?
- What factors cause discrepancies between math and reading academic performance per school district?

I first wanted to search for current research and publicly available data. Throughout many articles and reports, I found many references to the Educational Opportunity Project by Stanford University for academic performance data. Additionally, I found a Census supplement dataset that provided information about Internet/Computer usage. I also found many datasets from the American Community Survey that covered topics such as population demographics, socioeconomic variables, and child population statistics.

From the Educational Opportunity project, I was able to get extensive academic performance data across thousands of school districts. From the Census supplement, I was able to get data on internet and device usage behavior. And from the American Community Survey, I was able to examine a lot of demographic and socioeconomic factors.

The Educational Opportunity project calculates their academic performance scale based off a 2019 national standard that is standardized across all school districts. The grade year standard (GYS) dataset specifically indicates each unit represents a full grade level of proficiency.

The Census supplement was of particular interest because I had a hypothesis that internet/technology usage could be one of the primary factors that would differentiate performance in math and reading. As such, I was hoping to compare academic performance and technology usage across matching geographical areas.

The American Community Survey data provides aggregate statistics and margin of errors across numerous variables [per school district](#). This is the dataset that I ended up relying on to draw comparisons across school districts and academic performance. This dataset was challenging to work with and required a lot of data preparation and cleaning to be machine readable.

## Methodology

My initial goal was to create a map where school districts across the US could be represented as points and be interacted with to glean more information. I experimented with various R packages like the maps (Becker, Minka, and Deckmyn. 2023), usmap (Di Lorenzo 2024), and sf (Pebesma and Bivand 2023) packages to determine which had the capabilities, customization, and functionality I was looking for. The deliberation between these options was very difficult as I would learn and discover new things about each one as I experimented more. However, after much deliberation I concluded that a map should not be my ultimate goal for the project due to restrictions on the geographic spread of my datasets. The academic performance dataset for instance, only represented \_\_\_\_ school districts across \_\_\_\_ states which would have not translated well on a map. Additionally, the computer/internet use dataset was distributed by county instead of school district, which made it much more difficult to justify trying to use a map.

I drew heavy inspiration by the plots on the website of the [Educational Opportunity website](#) where they compared math/reading scores with different socioeconomic factors in a nice comprehensible scatter plot. I noted how effective this method was at mitigating the geographic spread issue so I adapted and extended this idea with my Shiny dashboard. Their dashboard offers filters and divisions by subgroup, subject, and year, but I wanted to extensively look into variables that could be compared to academic performance. The American Community Survey offers a diverse range of population characteristics on a school district level basis and as a result became my primary data source for interesting factors. This is also the source of the socioeconomic data on the Educational Opportunity's plots.

Once I created my first few interactive properties for my plots, fleshing out the functionality was not too challenging and I was able to make more additions to my dashboard to make it stand out more compared its inspiration. I added features to better help identify trends like fixed plot scaling, lines of best fit, and a more distinct color gradient. I also wanted to explore more variables and factors that could explain academic performance and specifically discrepancies between math and reading scores.

I also wanted to include a differentiation by year to indicate the effect of the pandemic.

## Results

The dashboard I created offers filters and divisions similar to the existing dashboard, but I analyzed a greater number of socioeconomic and demographic factors that could affect academic performance. From plotting socioeconomic factors, it was clear that indicators of areas being lower-income or lower-earning tended to be associated with lower academic performance in both math and reading.

I also made a plot of math vs reading scores and it demonstrated that performance in one subject was indicative of performance in the other across school districts.

## Discussion/Conclusion

My original research questions cannot be conclusively answered from the work of this project. I was able to discover and confirm the findings of many other researchers that lower-income/poorer school districts generally had worse academic performance than high-income/wealthier school districts. However, I was not able to identify any socioeconomic variables that resulted in noticeable discrepancies between reading and math scores. Similarly, the plot of reading scores vs math scores showed that how school districts were performing in one subject reflected how they performed in the other. As such, it remains difficult to answer the second question.

Comparing pre-pandemic to post-pandemic . . .

To take this project further, I would like to continue exploring different factors and variables per school district to draw more insights into what seems to affect academic performance. I would also like to include more interactivity with the scatterplot itself to aid in engagement. Additionally, I would also want to look into more COVID-related school district statistics to better demonstrate the effect of the pandemic on academic performance.

## Reflection (not part of report)

I am genuinely really happy with the work I have put in and with the final product that I have developed and will continue to maintain. I am very grateful to Professor Tackett for talking with me throughout the semester and being an incredible mentor for my project.

I think that I spent too much time working with the computer/internet use dataset before realizing its incompatibility with the primary dataset I was working with and in particular not recognizing the limitations of geospatial mapping sooner. I wish that I had started working with the American Community Survey dataset much earlier in the project as I could have been developing the scatter plots much earlier on which would have greatly assisted in app development.

Overall, I recognize that my process and evaluation could be improved greatly but I am nonetheless very proud of my work.

#### Questions

- How much should I write about the research I did before coding?
  - Reading various articles about academic recovery post-pandemic
  - Reading articles about factors that relate to better/worse recovery (COVID statistics, socioeconomic, etc.)
  - Discovering the primary sources of information and data (ie: Educational Opportunity Project & American Community Survey)
  - Looking into other sources for academic performance data (NCES)
  - The visit with the librarian for statistical science
- How in depth should I write about the data-prep/cleaning?
  - Ex: For the American Community Survey data, I had to transpose the dataset and shift a column down such that it aligned with the estimated values . . .
  - Ex: In order to add functionality of dividing by year/subject, I had to pivot wider the academic performance dataset
- Should I mention the extensive work I did with the computer/internet use Census Supplement Data? (even though it isn't relevant or included in the final product)

Becker, Original S code by Richard A., Allan R. Wilks. R version by Ray Brownrigg. Enhancements by Thomas P Minka, and Alex Deckmyn. 2023. "Maps: Draw Geographical Maps." <https://CRAN.R-project.org/package=maps>.

Chang, Winston, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert, and Barbara Borges. 2023. "Shiny: Web Application Framework for r." <https://CRAN.R-project.org/package=shiny>.

Di Lorenzo, Paolo. 2024. "Usmap: US Maps Including Alaska and Hawaii." <https://CRAN.R-project.org/package=usmap>.

Pebesma, Edzer, and Roger Bivand. 2023. "{Spatial Data Science: With Applications in r}." <https://doi.org/10.1201/9780429459016>.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the {Tidyverse}" 4: 1686. <https://doi.org/10.21105/joss.01686>.