# Final Project

AUTHOR
Sarah Morrison

PUBLISHED
December 7, 2024

```
RendererRegistry.enable('html')
```

Section: 1

Github Username: smorrison21

For my final project, I explored the research question: *How Does State Spending Relate to Enrollment Trends and Quality Standards in State Pre-K?* Using preconfigured datasets from the National Institute for Early Education Research, I analyzed data spanning from 2002 to 2023. Each dataset included variables such as *Year*, *State*, *Program Name* (when applicable), and metrics related to spending, enrollment, and program quality.

## Part I: State Pre-K Metrics Data Cleaning and Storing

### Read in Data

### Cleaning Spending Data:

The spending data includes variables such as **Total State Pre-K Spending (2023 Dollars)**, **State Spending per Child (2023 Dollars)**, **Total All-Reported Spending (2023 Dollars)**, and **All-Reported Spending per Child (2023 Dollars)**. I pivoted the DataFrame to analyze each variable separately. For my first dynamic plot, I examined the relationship between spending, enrollment, and quality standards over time for individual states, using **Total State Pre-K Spending (2023 Dollars)** to represent investments over the years. For static and comparative dynamic plots, I used **State Spending per Child (2023 Dollars)** to account for state size differences. A challenge arose with the **Program Name** column, as rows with "NA" indicated total spending across all programs. Summing values initially led to double-counting. To address this, I used the `max` function, selecting the largest value for each state-year-variable combination, ensuring accurate aggregation.

## Cleaning Enrollment Data

The enrollment data includes variables such as **Number of 3- and 4-year-olds Enrolled**, **Percentage of 3- and 4-year-olds Enrolled**, and **Total State Pre-K Enrollment**. Cleaning involved pivoting the table and using the `max` function for accurate aggregation. For the first dynamic plot, I analyzed trends using **Total State Pre-K Enrollment**, while for static and comparative dynamic plots, I used **Percentage of 4-Year-Olds Enrolled** to account for population differences across states.

### Cleaning Quality Standards Data

Cleaning the quality standards data was more complex due to varying benchmark evaluations over time. Benchmarks measured from 2002–2015 included items like **Teacher Degree** and **Class Size**, while from 2016–2023, additional measures like **Curriculum Supports** were added. Unlike spending and enrollment, standards data used categorical values in a **Benchmark Met?** column. States with multiple programs often included NA rows, some labeled **NA - Program Level Only**, indicating mixed compliance. Initially, I assigned 1 for "yes" and 0 for all others, but totals sometimes exceeded the expected 10 due to averaging in the source data. To address this, I adjusted the code to include NA rows only when they matched the average and recalculated benchmarks met annually by averaging compliance and summing results for each state-year combination.

```
['Yes' 'No program' 'No' 'Not reported' 'NA – Program level only'
 'Not applicable' 'No ' 'yes']
```

### Merge Data and Save to .csv

I combined the three DataFrames for further analysis and to create my first dynamic plot.

## Part II: Creating GeoDataFrame

### Read in Data

I used a shapefile from a previous problem set to get state geometries.

### Merge GeoData with Main DataFrame

I merged the GeoDataFrame with the DataFrame I created in the previous section to make said Dataframe into a GeoDataFrame for plotting. I then removed irrelevant columns to clean the GeoDataFrame.

### Save Data to .csv

I saved the GeoDataFrame to a .csv to create my second dynamic plot. The file is too large, so I saved it to my Google Drive. The file, preschool_geometry.csv, can be downloaded through the link to run the shiny app. The shiny app calls the GeoDataFrame from the data folder.

## Static Plots

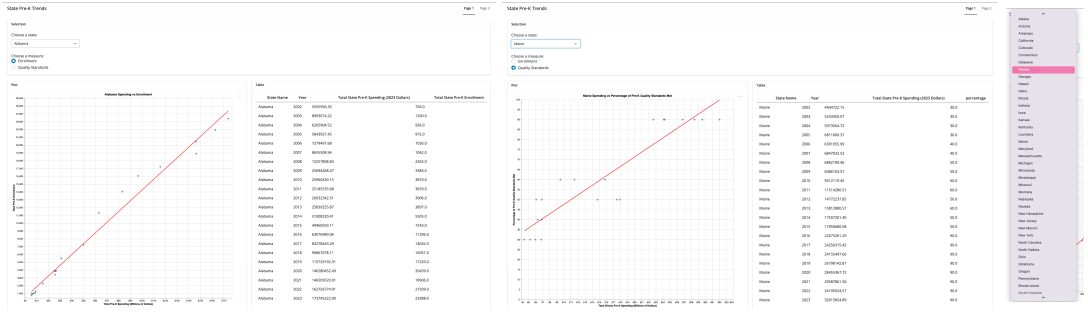
*Source: National Institute for Early Education Research (2023)*

Duplicate State Names in Charts: DC, New Jersey, Hawaii, Michigan, Mississippi, Florida, Utah, South Dakota, New Hampshire, Montana, Idaho, Indiana, Wyoming, Oklahoma, Texas, Arizona, Wisconsin

I created three static bar charts comparing the top 10 and bottom 10 states for spending per child in state pre-K, the percentage of 4-year-olds enrolled, and quality standards met. Each chart used color-coding to highlight disparities and identify states that consistently ranked high or low. The charts revealed that higher spending per child doesn't always align with higher enrollment or meeting more quality standards, highlighting areas for further research and policy considerations.

# Dynamic Plots

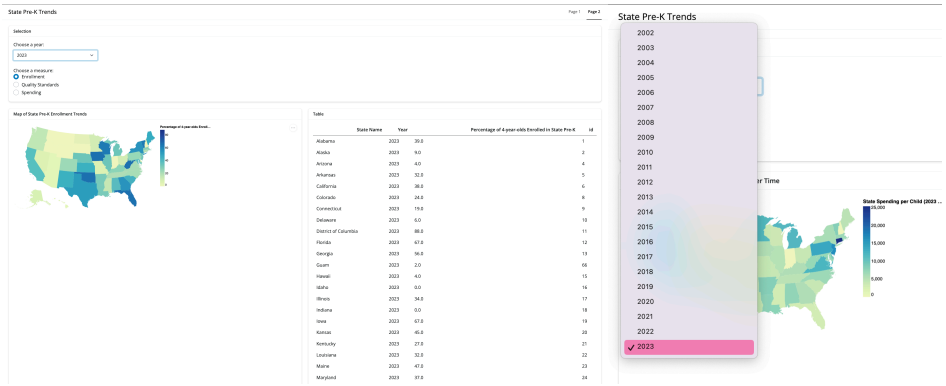
**Dynamic Plot 1**



Enrollment vs Spending, Quality Standards vs. Spending, Selection Options: States

The first dynamic plot explores the relationship between Total State Pre-K Spending (2023 Dollars) and either Total State Pre-K Enrollment or the Percentage of Quality Standards Met. Users can toggle between these options with radio buttons and select a state via a dropdown menu. An alternative visualization could adjust circle sizes on the enrollment plot to reflect the percentage of quality standards met. The charts show a strong positive correlation between spending and enrollment but only a weak correlation between spending and benchmarks met.

**Dynamic Plot 2**




The second dynamic plot is an interactive U.S. map where states are color-coded by one of three metrics: Spending Per Child (2023 Dollars), Percentage of 4-Year-Olds Enrolled, or Percentage of Quality Standards Met. Users can toggle metrics and select a year (2002–2023) via dropdown menus, revealing geographic patterns and trends in spending, enrollment, and quality standards over time.

## Policy Implications and Future Work

This analysis underscores key policy implications and research needs in state pre-K programs. States should optimize fund allocation to maximize impact, with research identifying areas for more effective spending. Policymakers must balance quality and access, exploring measures of quality and strategies to enhance program effectiveness while addressing the needs of unenrolled children. Additionally, studying long-term outcomes, such as the effects of low-quality pre-K on child development and equity, is crucial for guiding strategic investments and improving early education policies.