

When considering macroeconomic measures of long-term growth, prioritization of growing capital is widely regarded as one of the most important ways to continue increasing laborer productivity. On the side of firms and businesses, this comes through in terms of investments in physical capital and expansion that improves employment and increases the total output of the firm, then aggregated over many firms to create a largely felt effect on the economy. However, on the other side of the coin, the government's investments and usage of fiscal policy in fostering long-term, sustainable growth generally has two focuses. Assuming rule of law and strongly defined property rights are already defined and not largely changed, one way spending in government expenditures can promote long-term economic growth through spending on physical infrastructure like roads, public transportation, bridges, or sewage, hopefully outweighing the "crowding out" effect of running deficits on government budgets on national saving. The other way, and the method I am investigating further with this project, is with investment into human capital through funding of educational institutions, programs, services, and access. The return on human capital cannot be understated: when we examine how countries have recovered from great devastation to become international leaders in manufacturing and production, like Japan and Germany after World War II or China after the Cultural Revolution, one key trait we find in common is high investment into human capital by the government. From an extensive apprenticeship system to an on-the-job training emphasis, human capital has not only been a key driver in innovation and thus productivity spikes, but also in ensuring overall economic prosperity in the long-term rather than short-term relief.

Reviews, literature, and studies on the topic of the effects of fiscal policy as well as general policies around education and its workers on education quality and its subsequent link on economic growth are indeed extensive; however, there exists a lack of recent data collection and

analysis on the differences between states. There exists a unique opportunity with the United States of America, having fifty largely varying territories with differing state laws, natural resources, populations, and demographics, while still having many variables controlled through federal laws and regulations, language, and general culture surrounding work and labor. My project is to empirically determine a link between variables on education spending by the state, policies and culture promoting or discouraging union presence, membership, and collective bargaining strength, as well as their interaction through interaction terms in regression, on GDP per capita long term growth, if there indeed is one. Such variables would include public education spending as a percent of the states' total GDP, per-pupil expenditure, instructional teacher union participation rates, and teacher union strength. I began with a two model approach, testing two questions: first, can we establish a link between expenditure alongside union variables with education quality? Next, if there is a link, I will include education quality and outcomes, as measured by state standardized testing data in the National Assessment of Educational Progress, as an independent variable alongside the original independent variables from the first model to test a link between education quality and economic growth. Due to a lack of extensive time-series data for all variables, though I would like to run panel regression to compare each state to itself as its output, spending, and other educational variables changed, I instead will be comparing states versus each other. Thus, I will attempt to control for some differences in each state that can affect growth, like poverty rate, population density, and initial GDP per capita from 10 years prior in the case of current GDP per capita, by including them in the multiple linear regression. In my research, I also attempted to measure any statistically significant correlative factors between the variables measured and 10-year GDP per capita growth by state. I gathered all of this data from various public sources, including NAEP and

BEA, and compiled them into an Excel spreadsheet before importing into Python and performing exploratory data analysis to look for distributions and outliers, cleaning, and adding interaction terms.

For visualizations, I used heatmaps to determine correlation between dependent variables and independent variables to gain a first good understanding for which ones are relevant to be included within the next machine learning models. For variables with high correlation, scatterplots would be used to gain a better understanding and explore the data's ranges, alongside adding trendlines to better visualize relationships. Additionally, to see how certain variables have changed, I include trend lines of historical data on independent variables. After creating models, I included coefficient plots to more succinctly show variable effects' significance alongside residual diagnostics to determine if model assumptions were met. In all of these, I used pandas, scikitlearn, seaborn, matplotlib, numpy, math, and statsmodels libraries. I additionally used bootstrapping in place of testing and training data splits due to small sample size and the importance of each state, and measured error in the models, built off the mean of bootstrapping coefficients using Root Mean Squared Error.

Here are links to the main datasets I will be pulling from and combining:

[BEA Interactive Data Application](#) (GDP)
[Union Density Estimates by State, 1964-2021](#)
[States with Teacher Unions 2025](#)
[Teacher Union Strength](#)
[Nation's Report Card](#)
[Poverty Rate by State 2025](#)
[Real per capita GDP by state U.S. 2024 | Statista](#)
[COE - Public School Expenditures](#)
[Useful Stats: Per Capita Gross State Product, 1998-2018 | SSTI](#)
[BEA Interactive Data Application](#) (Population)
[Fiscal Effort Index.](#)
[Population Density](#)