Matthew Fikes

DSC 680: Project 1

SUMMARY

The study looks at the connections between childhood poverty and education. This is not new research but confirming existing knowledge. This project looks at data from the Common Core of Data, the US Census Bureau's Small Area Income and Poverty Estimates, and the US Department of Education's EDFacts initiative. The results of a least squares model were not compelling enough that any set of variables could reliably predict graduation rates for economically-disadvantaged students, and more testing is required to predict this based on district-level data.

INTRODUCTION

BACKGROUND

The purpose of this project is to study the connection between childhood poverty and educational attainment. There exists previous research on the effects of childhood poverty and how it may affect children later in life. It has shown to lead to poorer physical health, behavior, and educational attainment as an adult (Magnuson & Votruba-Zrzal, 2008). While there was no conclusive evidence that outcomes will not improve when poverty is remedied, the immediate effect is well-documented.

The highest level of education completed has a strong link to earning potential. The difference between finishing high school or not can mean $410K less over a lifetime for men and $290K for women (Social Security Administration, 2015). If a better education can lead to a better financial situation then it may help to undo some of the other damage caused by poverty.

PROBLEM STATEMENT

Do the characteristics of a school district predict student success of economically-disadvantaged students?

SCOPE

This project will be looking at district finances, poverty levels, and graduation rates for 2017 in the United States. This will not predict individual student success, but only the likelihood of success in a district. The data does not include any row-record student information and so cannot predict that specifically.

METHODS

DATA SOURCES

The data was pulled from CSV files compiled by the Urban Institute Education Data Portal. The data itself comes from three sources. All datasets include a common id identified as 'leaid'. This number represents each individual school district and a 'fips' code that identifies the location of the district.

Common Core of Data (CCD) – the US Department of Education database for primary and secondary education.

- rev_total: The sum of revenue contributions emerging from local, state, and federal sources.
- exp_total: Total expenditures
- exp_current_elsec_total (exp_edu): Total current expenditures for elementary and secondary education
- enrollment_fall_responsible: Number of students for which the reporting local education agency is financially responsible

Small Area Income and Poverty Estimates (SAIPE) – a program under the US Census Bureau that provides poverty and income estimates for states and counties, in addition to estimates of the number of school-aged children in poverty.

- est_population_total: Estimated total population

- est_population_5_17_poverty ('youth_poverty'): Estimated total population ages 5–17

EDFacts – an initiative under the US Department of Education that analyzes assessment data and completion rates for student cohorts.

- Data from Edfacts includes graduation rates for students identified as economically-disadvantaged and for all students. Graduation rates are reported as low, midpoint and high ranges.

DATA PREPARATION

The final merged dataset comprised records from 12251 districts before districts with null values were removed, leaving 7409 for processing. Fips codes outside of the 50 states and the District of Columbia were removed due to many missing values.

MODELING

For testing I chose to work with ordinary least squares regression. The target variable I chose was the high graduation rate for economically-disadvantaged students. I ran several tests with the low, midrange, and high rates reported by districts and chose the high rate for disadvantaged students.

RESULTS

The results of the modeling were mixed. Some of the best results included the high value for overall graduation rate, with an R-squared of 0.693. I suspect that this is information leakage and using a variable not obtained until graduation makes it difficult to make predictions in time to do something about the results. Removing that variable from the model drops the R-squared to 0.031. Unfortunately it seems that while there is a connection between poverty and education in many articles and journals, this type of modeling at the district level is not a suitable method for predicting graduation rates for low-income students.

LIMITATIONS/CHALLENGES

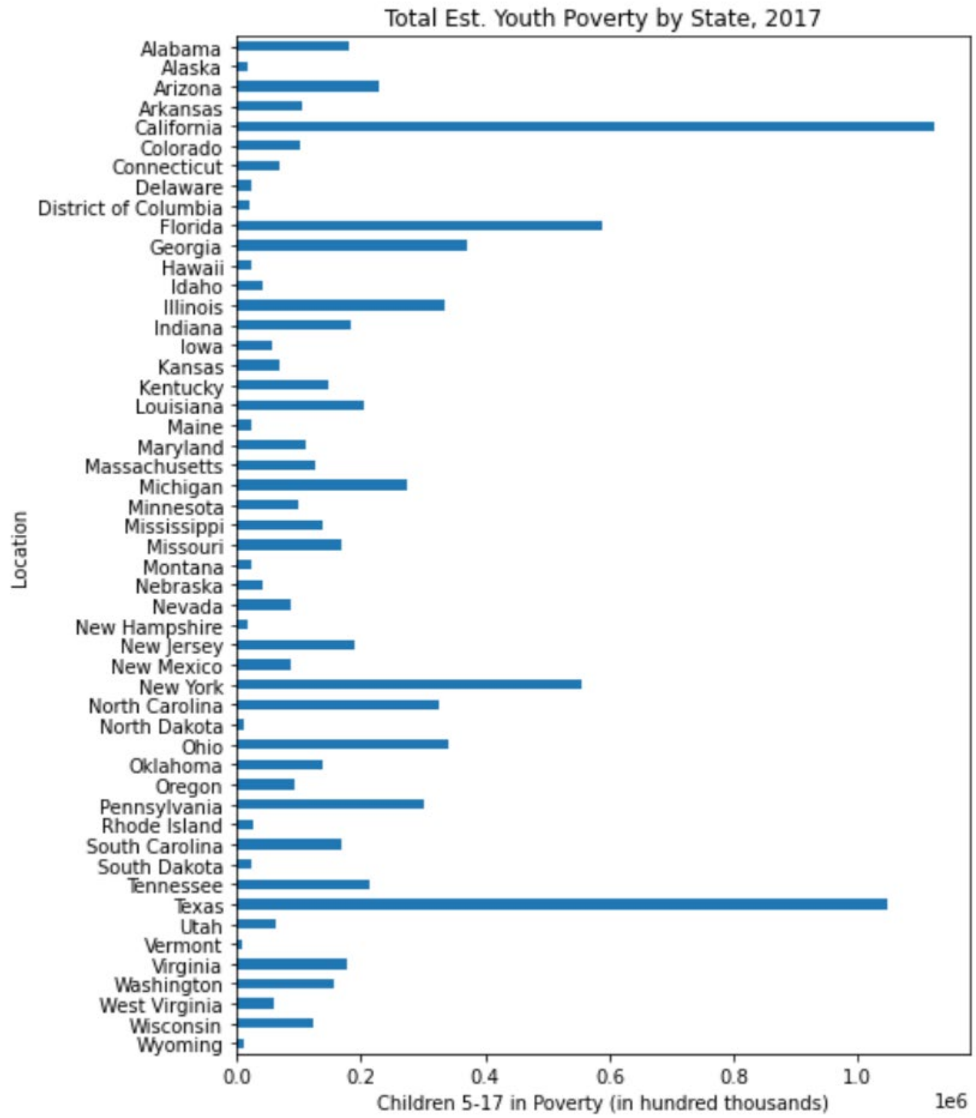
The data only goes to the level of school district so individual students or even some individual school data is unavailable. This means that success can only be predicted generally based on the district characteristics.

Graduation rates for each school district are divided into low, midpoint, and high rates which make setting a single target point difficult. Even just choosing the midpoint as a value can be misleading when the range is high. Some composite value representing the mean and range might work better.

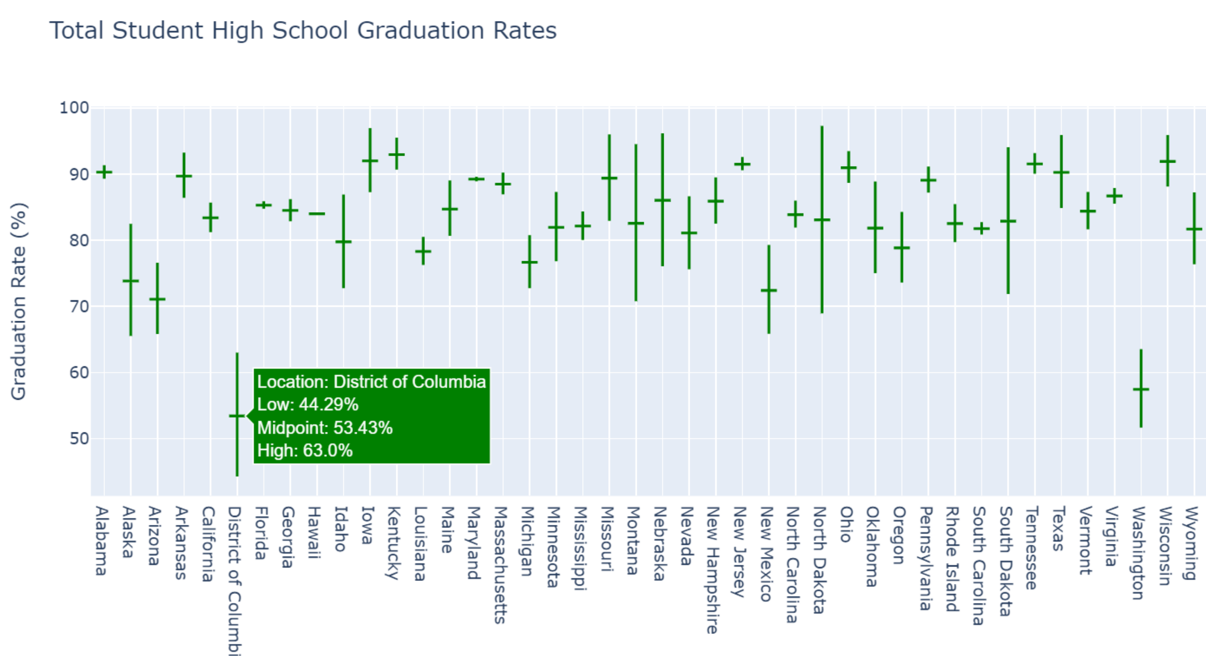
Next Steps

Other models may perform better and are worth further investigation. Ideally row-record data for students would be the best way of working on an issue like this but there are a number of obstacles involved in working with actual student data.

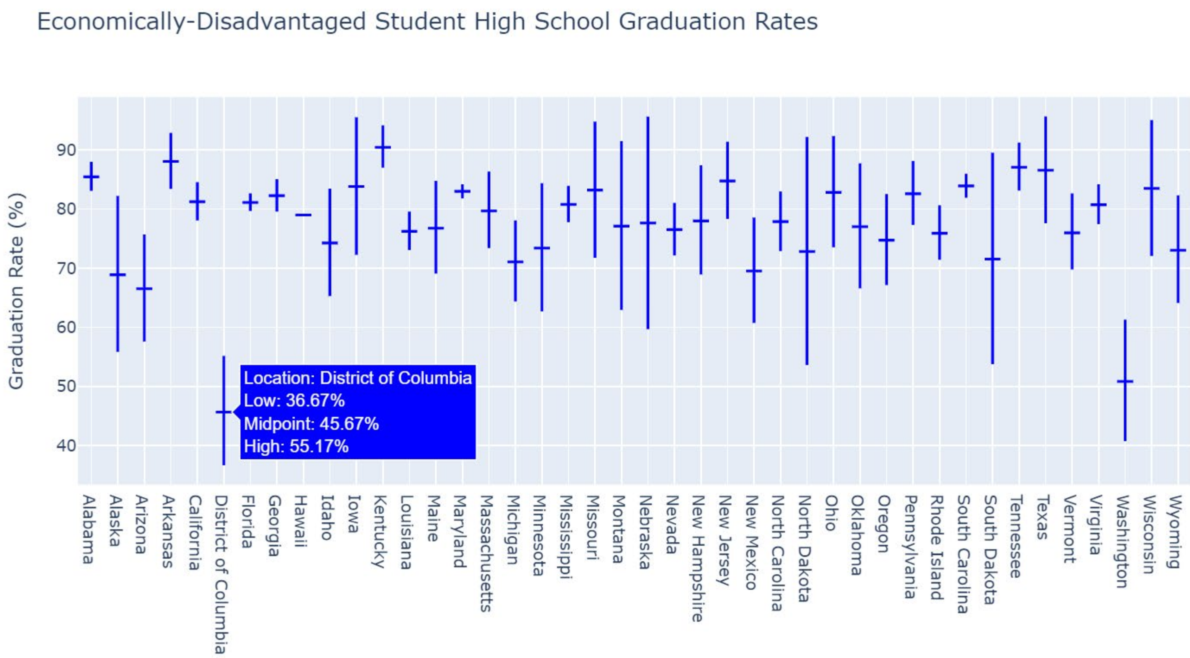
Appendix:



Total Est. Youth Poverty by State, 2017
Location
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District of Columbia
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia
Wisconsin
Wyoming
0.0
0.2
0.4
0.6
0.8
1.0
Children 5-17 in Poverty (in hundred thousands)
1e6



Total Student High School Graduation Rates
Graduation Rate (%)
100
90
80
70
60
50
Location: District of Columbia
Low: 44.29%
Midpoint: 53.43%
High: 63.0%
Alabama
Alaska
Arizona
Arkansas
California
District of Columbia
Florida
Georgia
Hawaii
Idaho
Iowa
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Vermont
Virginia
Washington
Wisconsin
Wyoming




```
                            OLS Regression Results
==============================================================================
Dep. Variable:             gr_econ_hi   R-squared:                       0.693
Model:                            OLS   Adj. R-squared:                  0.692
Method:                 Least Squares   F-statistic:                     2781.
Date:                Wed, 22 Sep 2021   Prob (F-statistic):               0.00
Time:                        23:42:26   Log-Likelihood:                 9978.1
No. Observations:                7409   AIC:                        -1.994e+04
Df Residuals:                    7402   BIC:                        -1.989e+04
Df Model:                           6
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
const                   -0.0493      0.008     -6.477      0.000      -0.064      -0.034
youth_poverty            1.0661      0.098     10.832      0.000       0.873       1.259
exp_edu                 -0.2673      0.328     -0.816      0.415      -0.910       0.375
rev_total                0.8690      0.532      1.632      0.103      -0.175       1.913
est_population_total    -1.0469      0.137     -7.634      0.000      -1.316      -0.778
exp_total               -0.6715      0.506     -1.328      0.184      -1.663       0.320
gr_tot_hi                1.0374      0.008    126.282      0.000       1.021       1.053
==============================================================================
Omnibus:                     1987.003   Durbin-Watson:                   1.882
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            97277.647
Skew:                          -0.485   Prob(JB):                         0.00
Kurtosis:                      20.725   Cond. No.                     1.29e+03
==============================================================================
```

Figure 1 - Regression Results (including total grad feature)

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             gr_econ_hi   R-squared:                       0.031
Model:                            OLS   Adj. R-squared:                  0.030
Method:                 Least Squares   F-statistic:                     47.02
Date:                Thu, 23 Sep 2021   Prob (F-statistic):           5.09e-48
Time:                        00:05:54   Log-Likelihood:                 5722.3
No. Observations:                7409   AIC:                        -1.143e+04
Df Residuals:                    7403   BIC:                        -1.139e+04
Df Model:                           5
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
const                    0.9072      0.001    662.637      0.000       0.904       0.910
youth_poverty           -0.1717      0.174     -0.987      0.324      -0.513       0.169
exp_edu                 -1.2331      0.582     -2.120      0.034      -2.373      -0.093
rev_total                2.4139      0.945      2.554      0.011       0.561       4.267
est_population_total    -0.9067      0.244     -3.723      0.000      -1.384      -0.429
exp_total               -0.8581      0.898     -0.955      0.339      -2.619       0.902
==============================================================================
Omnibus:                     2184.453   Durbin-Watson:                   1.639
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             7611.241
Skew:                          -1.465   Prob(JB):                         0.00
Kurtosis:                       7.008   Cond. No.                         952.
==============================================================================
```

Figure 2- Regression Results (excluding total grad feature)

References:

Magnuson, K. A., & Votruba-Drzal, E. (2008). *Enduring influences of childhood poverty*. Madison: University of Wisconsin-Madison, Institute for Research on Poverty.

Social security Administration. Research Summary: Education and Lifetime Earnings. (2015, November). Retrieved September 17, 2021, from https://www.ssa.gov/policy/docs/research-summaries/education-earnings.html.