SIADS 591 (Winter 2021)

Exploring the Relationship between Education and Prison Expenditures

GitHub Repository: https://github.com/albemlee/siads_592_school_prison_exploration
Sean Cafferty (caffsean@umich.edu)
Albert Lee (albemlee@umich.edu)
Damaris Rodgers (damarisr@umich.edu)

Motivation

Our group has decided to explore the relationship between incarceration and education spending. All our team members have a strong interest in the topic based on our backgrounds: Sean has a social science background, Albert teaches students in Detroit, and Damaris is curious about the topic as a parent.

The United States has a fraught history of using incarceration to shape its civil society. As such, we're hoping to gain a deeper understanding of contemporary social issues in the United States. Our group would thus like to explore these dynamics by trying to tease out correlations among various data sets.

- Are there historical trends related to education and incarceration?
- Are the geographic trends related to education and incarceration?
- What correlations exist between education and incarceration?

Because impactful social science research needs to be grounded in appropriate theory in order to establish claims of causality, we will situate any notions of causal relationships within existing social science frameworks that account the history and economics of both education and carceral systems.

Data Sources

The data sources that we used are:

- Education Data Portal (CCD, CRDC)
- Incarceration Trends Dataset
- State and Local Government Expenditures on Police Protection in the U.S., 2000-2017
- State-by-State Spending on Kids
- Property Tax Rates
- County Geography

Education Data Portal

Source	https://educationdata.urban.org/api/v1
Format	JSON (API)
Important Variables	Detailed breakdown of revenue and expenditures of every school district (Common Core of Data Finance dataset). Record of disciplinary actions (e.g. suspensions) of every school grouped by race (Civil Rights Data Collection Discipline dataset).
Size	The 2015 Common Core of Data Finance dataset had over 18,000 rows and 130 columns. The 2015 Civil Rights Data Collection Discipline dataset had 5,963,076 rows and 19 columns.
Time Periods	The Common Core of Data Finance dataset included years 1994 - 2016. The Civil Rights Data Collection Discipline dataset included years 2011, 2013, and 2015.

Incarceration Trends by the Vera Institute

Source	https://github.com/vera-institute/incarceration-trends/blob/master/incarceration_trends.csv
Format	CSV
Important Variables	Jail and prison populations per race/ethnicity
Size	153,811 rows, 121 columns
Time Periods	1970 - 2018

State and Local Government Expenditures on Police Protection in the U.S., 2000-2017

Source	https://www.bjs.gov/content/pub/sheets/slgeppus0017.zip
Format	CSV
Important Variables	Total expenditures for all police protection in the U.S.
Size	The combined data (across all years) had 18 rows and 14 columns
Time Periods	2000 - 2017

State-by-State Spending on Kids

Source	https://datacatalog.urban.org/dataset/state-state-spending-kids-dataset
Format	Excel
Important Variables	The State-by-State Spending on Kids dataset provides the expenditure data on elementary and secondary education, subsidies, special services, and the Head Start program.
Size	The combined and reshaped data (across all years) had 867 rows and 12 columns
Time Periods	1997 - 2016

Property Tax Rates

Source	https://www.lincolninst.edu/research-data/data-toolkits/significant-features-property-tax/access-property-tax-database/property-tax-rates https://www.nahbclassic.org/generic.aspx?genericContentID=250239&fromGSA=1
Format	CSV
Important Variables	Average house values and property taxes of all US Counties
Size	3,109 rows, 8 columns
Time Periods	2010-2014

County Geography

Source	https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html
Format	ShapeFile
Important Variables	Geometry of all US counties (for mapping)
Size	N/A
Time Periods	2018

Data Manipulation

There were three phases of data manipulation for this project:

- Extract Data
- Reshape Data
- Merge Data
- Additional Processing Steps

Extract Data

The following data were available for direct download from webpages:

- Incarceration Trends
- State-by-State Spending on Kids
- State and Local Government Expenditures on Police Protection in the U.S., 2000-2017
- Property Tax Rates
- County Geography

The Common Core of Data and Civil Rights Data Collection datasets were acquired through the Education Data Portal API and converted from JSON to a tabular format (Pandas dataframe and CSV). One specific challenge we encountered during this process was that the API only returned one page of entries (100 records) at a time. We used a loop to retrieve results from all pages for the datasets, and the entire loop took over 24 hours to complete for the discipline datasets within Civil Rights Data Collection. The code used to complete this can be found in "01_gather_education_data.ipynb."

Reshape Data

For our analysis, we need all analysis variables in columns and counties or states as the rows. The State-by-State Spending on Kids came in a unique format as an Excel document with separate worksheets for each variable, rows as states, and columns as years, and we needed to reshape the data. The code used to complete this can be found in "02_statebystate_wrangling.ipynb."

Merge Data

To conduct correlation analysis between education and incarceration trends, we needed to merge all data at the county level (FIPS code). The figure below shows the merge pipeline.
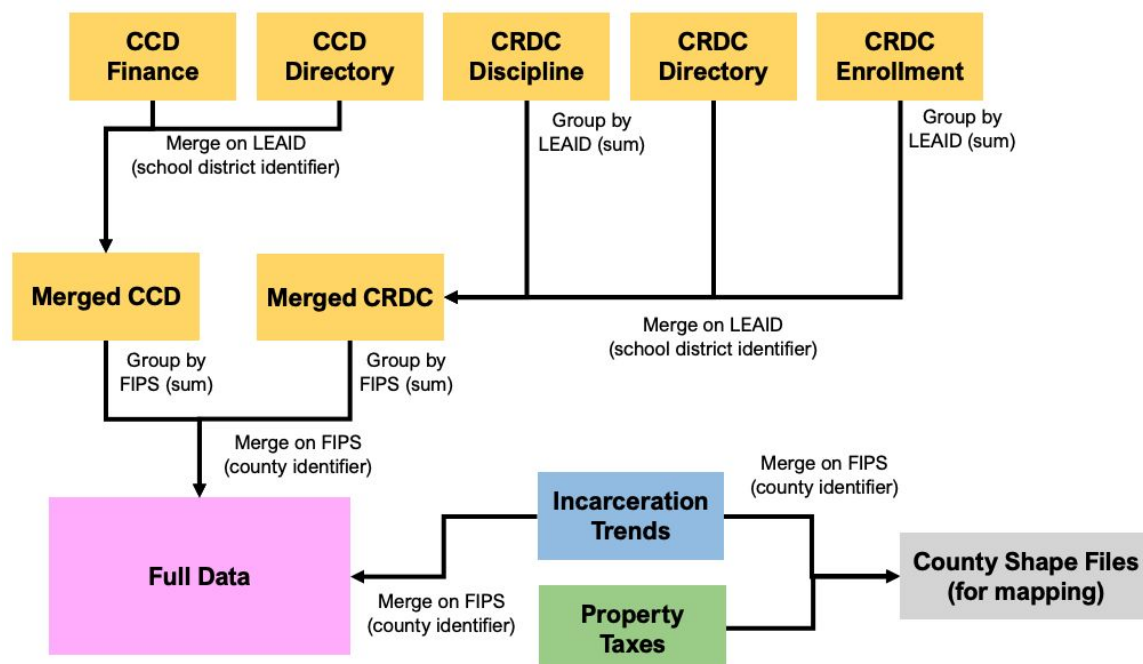
To obtain the FIPS codes for the educational datasets, they needed to be merged with their directory files using the LEAID (school district identifier). Since CRDC Discipline and CRDC Enrollment were both at the school level, they needed to be grouped by the county level (FIPS) before they could be merged.

Once the data was merged by school districts, they needed to be aggregated by counties (FIPS code) before being merged again. Since Incarceration Trends and Property taxes were already at the county level and both contained FIPS codes, they were ready to merge with the rest of the data.

Some counties were lost because all merges were inner joins, but most of the data (> 90 percent) remained after the merging process.

The code used to complete this can be found in "03_education_wrangling.ipynb."

Additional Processing Steps

Not all of our datasets were used in the correlation analysis, and the State and Local Government Expenditures on Police Protection dataset and the State-by-State Spending on Kids dataset were used in a separate analysis. To prepare these datasets for analysis, the State-by-State Spending on Kids dataset was totaled at the national level and then merged with the State and Local Government Expenditures on Police Protection dataset by year.

Data Analysis and Visualization

Main Findings of Map Visualizations

The main findings from the map visualizations were:

- Trends towards mass incarceration began in the Deep South–in Georgia in particular.
- By the 1980s mass incarceration began to proliferate in areas in which Black Americans represented a high proportion of the population. In other words, the higher the proportion of Black residents in a given county at the time, the higher that county's incarceration rate was likely to be.
- The relatively high incarceration for Black prisoners relative to the Black population in some counties appears to reflect the extent to which incarcerated individuals are moved from high-density counties to low-density counties. Google searches reveal that the counties with a high proportion of their Black population in prison are often home to large corrections facilities, thereby confirming the suspicion that this variable serves as a proxy for the mass displacement of Black prisoners from their communities of origin. This pattern is not visible with White prisoners indicating the disproportionate impact of incarceration-related displacement on Black Americans.
- This analysis can also be found in the '08_map_generator.ipynb' notebook

Social Science and Map Visualizations

Data collection in the social sciences always involves a spatial element insofar as humans are geographically bound objects of analysis. As such, the use of spatial visualizations such as maps serve as a useful tool for exploring the implications of a given social science data set. Maps are effective tools in the social science domain due to their ability to display historical narratives that might otherwise go unnoticed. The ubiquity of maps as a tool for data communication ensures that users of differing backgrounds are readily able to extract useful information from geographic representations of data. Although maps effectively communicate some variables, some features of data (such as uncertainty) are difficult to effectively encode in map form. Even so, maps should be considered an essential tool for any data exploration in the domain of social sciences.

Telling the Story of Mass Incarceration with Maps

Georgia and the Beginnings of Mass Incarceration in the Deep South
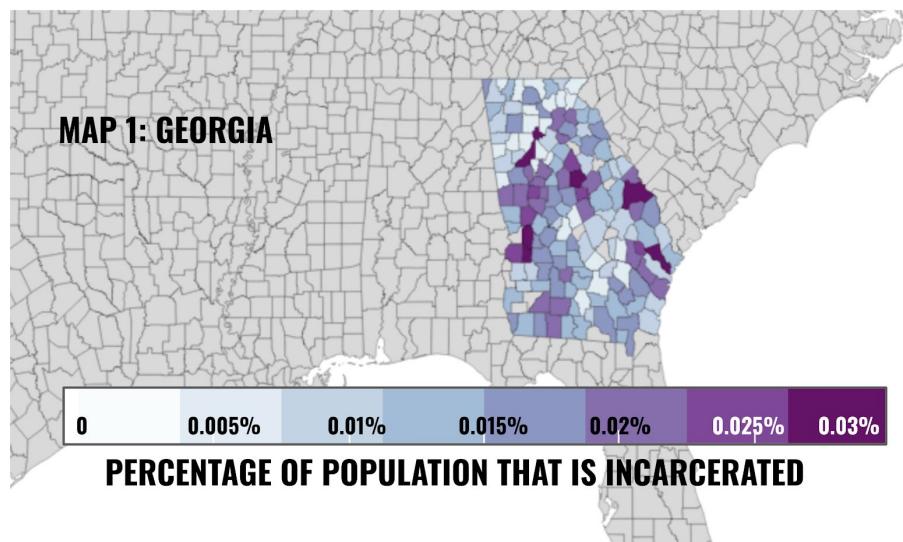
In exploring the Vera Institute's incarceration data, it made sense to start at the beginning. Using the map generator to look at incarceration rates in 1970 (the first year of available data), we see that only Georgia has any data **(MAP 1)**. What does this mean? One way to interpret it is that Georgia was home to the vast majority of incarcerated Americans in 1970. Alternatively, one might note that perhaps Georgia started collecting incarceration data relatively early. The correct interpretation might be a mix of both. Perhaps Georgia was at the vanguard of data collection regarding prisoners in part due to its extensive incarceration apparatus. A review of the history of the Jim Crow South indicates that Georgia began the trend of 'convict leasing'–a widespread practice used to re-enslave Black Americans after the 13th Amendment allowed for a labor-as-punishment loophole.[1] Given the early beginnings of Georgia's carceral state, it is reasonable that Georgia would already be engaging in robust data collection by 1970. Prisoners were assets after all.[2]
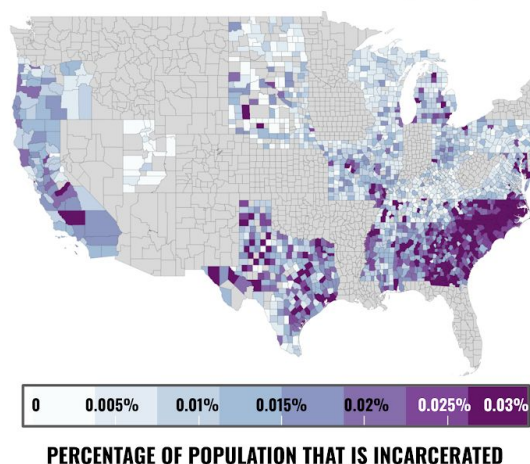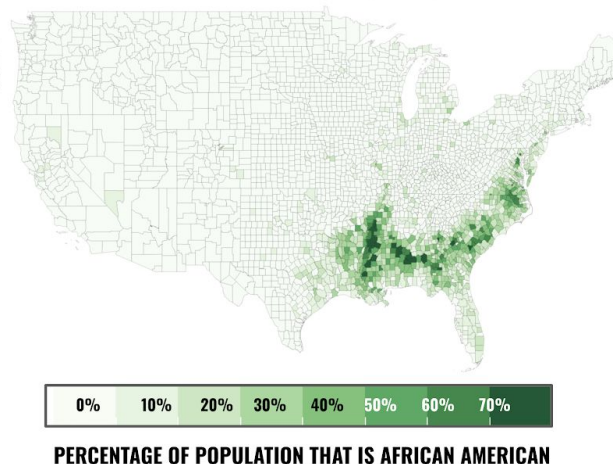
The Beginning of Modern Mass Incarceration

By 1983 data collection regarding incarceration begins to take off across the country **(MAP 2)**. Given the racialized origins of mass incarceration, it is clear in the data that the areas experiencing increasing incarceration rates in the 1980s are home to more concentrated Black communities. **(MAP 3)** Indeed, in the 1970s, Nixon began the 'War on Drugs' with the 'tough on crime' campaign that has come to be seen as 'thinly veiled racial rhetoric' by modern analysts.[3] Considering that the modern prison system emerged from the 'convict leasing' system that began in Georgia, it is perhaps unsurprising that the trends toward mass incarceration began in the Deep South before spreading throughout the country.

---

[1] https://en.wikipedia.org/wiki/Convict_leasing
[2] Further reading on Georgia's history of mass incarceration:
https://theconversation.com/prison-records-from-1800s-georgia-show-mass-incarcerations-racially-charged-beginnings-96612
[3] https://www.brennancenter.org/our-work/analysis-opinion/history-mass-incarceration

MAP 2: INCARCERATION RATES (1983)

| 0 | 0.005% | 0.01% | 0.015% | 0.02% | 0.025% | 0.03% |

PERCENTAGE OF POPULATION THAT IS INCARCERATED

MAP 3: BLACK POPULATION (2015)

| 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% |

PERCENTAGE OF POPULATION THAT IS AFRICAN AMERICAN

The Expansion of Modern Mass Incarceration

As we watch incarceration rates skyrocket across America from the 1980s onwards, it becomes clear that the numbers reflect changes in federal policy as the Nixon administration began a tough-on-crime political culture. The prison population truly began to expand under Reagan, however, as the crack epidemic (a public health crisis) was used as a pretext to terrorize communities of color–Black communities in particular.[4] By 1995, incarceration rates were exploding across the country (**MAP 4**) and by 2015 (**MAP 6**) it's clear that mass incarceration is a national phenomena, with the darkest color representing 1 in 100 people being incarcerated.

MAP 4: INCARCERATION RATES (1995)   **MAP 5: INCARCERATION RATES (2005)**   **MAP 6: INCARCERATION RATES (2015)**

| 0 | 0.2% | 0.4% | 0.6% | 0.8% | 1% |

PERCENTAGE OF POPULATION THAT IS INCARCERATED
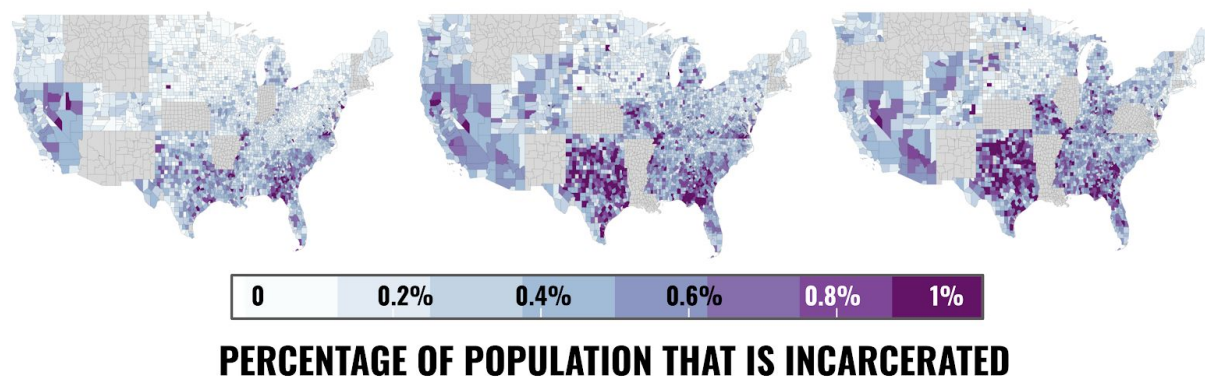
The Mass Displacement of Black Americans

Despite the usefulness of map visualizations, mapping populations alone can distort deeper patterns. Indeed, what becomes lost in this geospatial data is the extent to which people move. With respect to prisons, it is well-documented that prisoners are frequently moved from high-density counties to low-density counties. This pattern isn't immediately visible in the data. In order to examine this trend via mapping techniques, one must be creative with the data. As such, a new column was created in the data frame ('black_prison_ratio') which calculated the ratio of the incarcerated Black population to the overall Black population (aged 15-64). (**MAP 8**)

---

[4] https://en.wikipedia.org/wiki/Crack_epidemic_in_the_United_States

Though analytically imperfect, this new column reveals a pattern whereby Black residents as a percentage of the population appear to be incarcerated at extremely high rates (often more than 10%) in some counties. This pattern suggests that Black prisoners are being moved from high-density counties to low-density counties. After using several iterations of the map to identify counties with high rates of Black prisoners, it becomes evident that this variable can also be used to reliably identify counties with large correctional facilities. In order to test for the statistical relationship between a high Black incarceration rate and the presence of large correctional facilities, however, we would need to create a new categorical variable that indicates the presence of such facilities–a gargantuan task that unfortunately goes beyond the scope of this project. Nonetheless, high incarceration rates among a county's Black population seem to represent counties that are likely to have large correctional facilities. For instance, Madison County, North Carolina (with a 'black_prison_ratio' score of over 30%) appears to be home to several large correctional facilities. Furthermore, one of the counties identified with the tooltip, Dearborn County, Indiana, was even the subject of an scathing journalistic account entitled "Dearborn County, Indiana: A Microcosm for America's Prison Problem"[5]. The fact that this measure of Black prison populations appears to allow for the identification of large prison facilities in rural counties reflects the extent to which Black populations are displaced from their communities via America's mass incarceration apparatus. Moreover, this process is not replicable for White prisoners. **(MAP 9)**

MAP 8: BLACK INCARCERATION RATE   MAP 9: WHITE INCARCERATION RATE

| 0 | 2% | 4% | 6% | 8% | 10% |
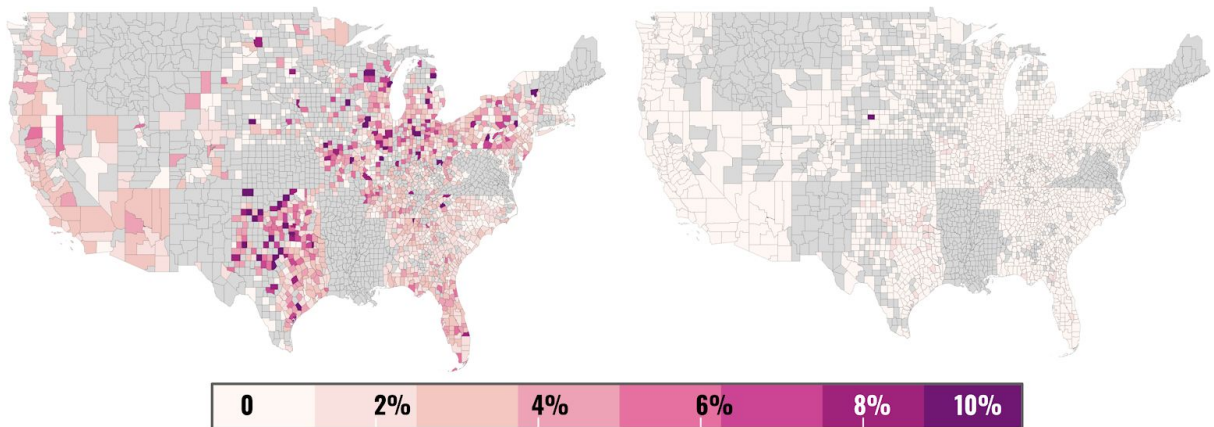
PERCENTAGE OF DEMOGRAPHIC GROUP THAT IS INCARCERATED

Concluding Remarks on Mapping Visualizations

The use of map data allowed us to visualize the extent to which Black Americans are displaced from high-density counties to low-density counties–a pattern that was not replicated for White Americans. This pattern further indicates the extent to which Black Americans have been targeted by mass incarceration, as well as the extent to which geographically bound data analysis has its shortcomings. Given that the relationship between social inputs (e.g. education spending) and social outcomes (e.g. incarceration status) is not necessarily bound by geography, it may not be justifiable to dwell on statistical relationships that hinge on geographic location (e.g. county).

Moreover, the usage of mapping techniques allowed for a rapid evaluation of the quality of the Vera Institute Data Set. Although substantial in size, this data set has significant gaps which would undermine any statistical relationships that might be gleaned from a large-N analysis. For instance, data from Illinois are often missing from the data set. Given the state's complex history of race and mass incarceration, its exclusion suggests that any statistical correlations derived from the data set overall should be taken with a grain of salt. Nonetheless, the data largely bear out the documented history of an incarceration

---

[5] https://www.wnycstudios.org/podcasts/takeaway/segments/small-indiana-countys-rising-prison-population

apparatus that targets Black Americans. This is a mass tragedy that will force American society to contemplate the meaning of 'justice' for generations to come.
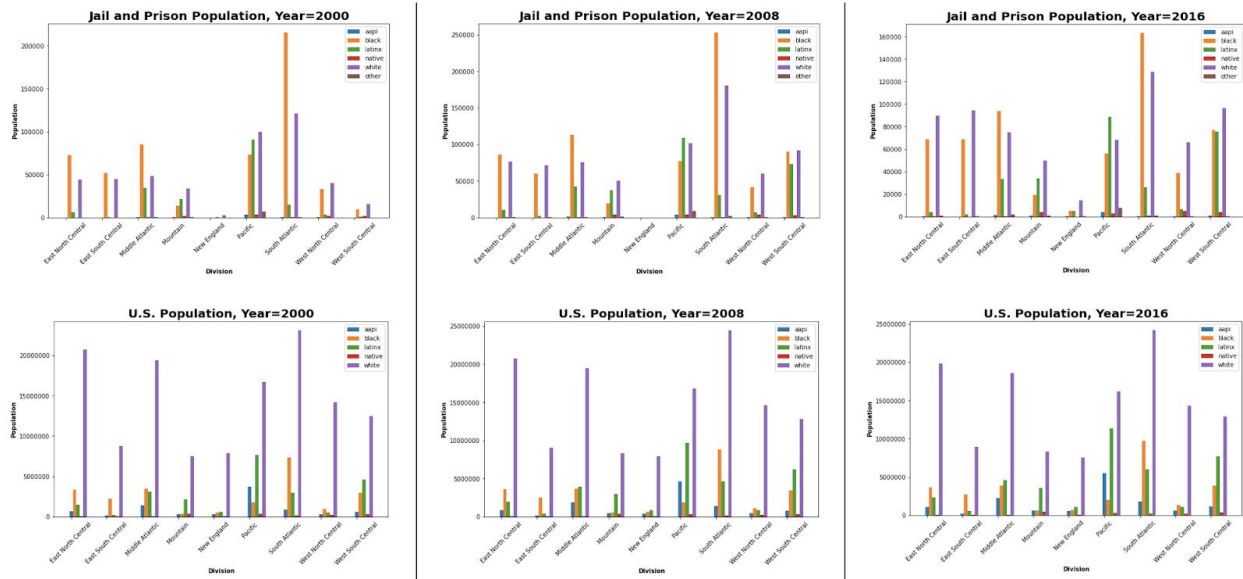
Incarceration Findings
- Finding 1: Jail and Prison Populations per U.S. Geographical Division
- Finding 2: Accurate Representation of Jail and Prison Populations

Finding 1: Jail and Prison Populations per U.S. Geographical Division

For the years 2000 - 2016, we combined the jail and prison populations per race/ethnicity, then grouped them by the U.S. geographic divisions (below is a breakdown of what the geographic divisions are as well as the states in each division as determined by the U.S. census). We then compared the jail and prison populations to the U.S. populations for the race/ethnicity sub-populations and geographic divisions. From this you can see how the Black and Latinx jail and prison populations (in particular the Black population) are more likely to be incarcerated as often as (if not more than) the White population. However, there is a much higher White population than the other races and ethnicities in the U.S. Even with the missing data from some states this goes to show that the Black and Latinx populations are being incarcerated at a much higher rate than the White population. This is illustrated by the sets of bar plots for the years 2000, 2008, and 2016, below. This analysis can be found in the '04_incarceration_trends.ipynb' notebook. Additional visualizations can be found in this project's GitHub repository:

- East North Central Division: Illinois, Indiana, Michigan, Ohio, Wisconsin
- East South Central Division: Alabama, Kentucky, Mississippi, and Tennessee
- Middle Atlantic Division: New Jersey, New York, and Pennsylvania
- Mountain Division: Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming
- New England Division: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont
- Pacific Division: Alaska, California, Hawaii, Oregon, and Washington
- South Atlantic Division: Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, and West Virginia
- West North Central Division: Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota
- West South Central: Arkansas, Louisiana, Oklahoma, and Texas

POPULATIONS PER U.S. CENSUS DIVISIONS

([GitHub link](#))

Finding 2: Accurate Representation of Jail and Prison Populations

As was mentioned in the incarceration trends documentation, the researchers were not able to gather the data consistently for all the years or for all of the states. Because of this we do not have an accurate representation of the jail and prison populations. The states that do not show any jail and prison population data include: Alaska, Arizona, Connecticut, District of Columbia, Delaware, Hawaii, Idaho, Kansas, Louisiana, Montana, New Mexico Rhode Island, and Vermont. States that are missing jail and prison population data for one or more years include: Iowa, Illinois, Massachusetts, Maryland, Maine, Mississippi, New Hampshire, Nevada, Ohio, Oregon, Texas, Virginia, West Virginia, and Wyoming. This means that 27 of the 50 states and DC contain incomplete or no data at all. This analysis can be found in the '04_incarceration_trends.ipynb' notebook. Additional visualizations can be found in this project's [GitHub repository](#):
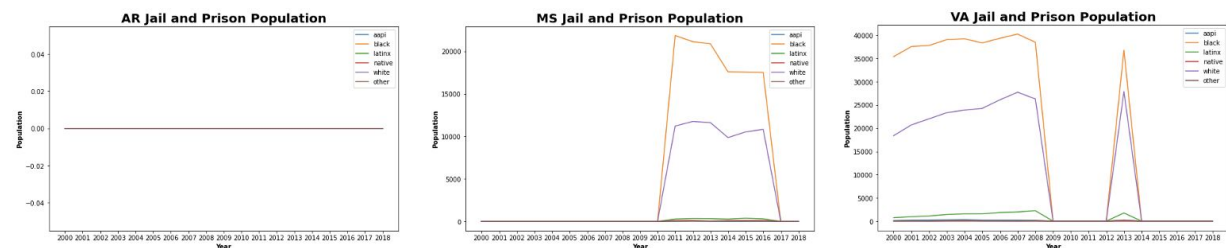

SAMPLE STATES WITH MISSING OR NO DATA

([GitHub link](#))

Education and Police Expenditures Findings
- Finding 1: Education Expenditures by State
- Finding 2: U.S. Police Protection Expenditures vs. Education Expenditures

Finding 1: Education Expenditures by State

We also looked at how much each state was spending on education for the years 2000 - 2016 (inflation adjusted). The state by state spending did not contain data about the race/ethnicity of the student populations. However, for each state you can see how the majority of the education funds are spent on the public spending of elementary and secondary education, while little to none is spent on special education services, subsidies, and head start programs. This analysis can be found in the '07_state_spending' notebook.
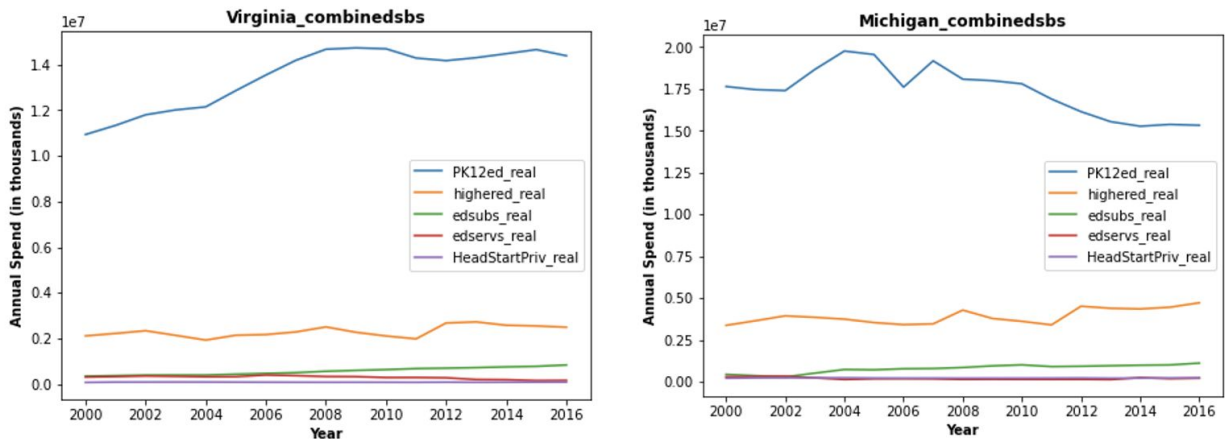

SAMPLE STATE SPENDING ON EDUCATION

(GitHub link)

Finding 2: U.S. Police Protection Expenditures vs. Education Expenditures

The police protection expenditures (inflation adjusted) were for the entire U.S. and not broken down by states. This includes expenditures for police protection, all judicial and legal functions, and corrections. It does not say if it includes police spending in schools. So while we can't glean much from this dataset we did compare it to the total U.S. spending on education and as you may have expected more money was spent on education than police protection. However, on a per capita/child basis, in 2000 approximately only $40 more was spent on education than police protection. In 2009, approximately $85 more was spent on education than police protection and in 2017, approximately $95 more was spent on education than police protection. The education spending is increasing more rapidly than the police protection spending. This analysis can be found in the '05_police_vs_education_spending.ipynb' notebook.

U.S. Education (Per Child) and Police Protection Expenditures (Per Capita)
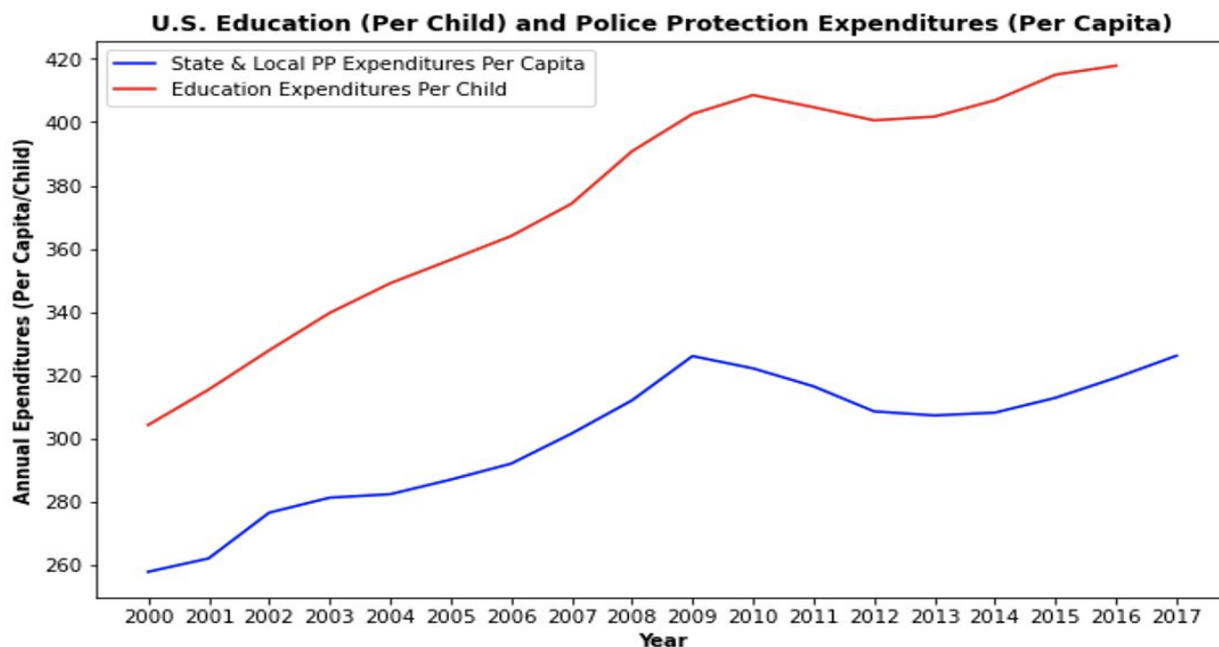
PER CAPITA / CHILD EXPENDITURES

Correlation Analysis Findings

We completed Spearman correlation tests to determine which educational variables had the highest correlation with incarceration rates at the county level. The table below displays the variables with the highest correlation based on the tests.

Variable	Description	Correlation Coefficient	P
benefits_supp_sch_admin	Employee benefits for school administration	-0.393143	5.899278e-116
benefits_employee_total	Total employee benefits	-0.391186	1.006437e-114
benefits_supp_operation_plant	Employee benefit for plant operations	-0.371717	6.403374e-103
benefits_supp_bco	Employee benefits for business, central, and other support services	-0.366352	8.303788e-100
exp_current_supp_serve_total	Expenditures for support services	-0.346700	6.671642e-89
exp_current_operation_plant	Expenditures for plant operations	-0.339861	2.755892e-85
exp_current_elsec_total	Total expenditures for elementary and secondary education	-0.338422	1.547400e-84
exp_current_sch_admin	Expenditures for administration and support services	-0.337931	2.780745e-84
exp_current_instruction_total	Total Instructional Expenditures	-0.335460	5.239860e-83
exp_total	Total Expenditures	-0.326933	1.071313e-78

rev_local_total	Total Revenue from Local Government	-0.326930	1.074940e-78
students_corporal_punish	Number of students who received corporal punishment	0.325807	3.882463e-78
rev_total	Total Revenue	-0.325578	5.037476e-78
salaries_total	Total Salary Expenditures	-0.321372	5.856625e-76
salaries_instruction	Total Salary for InstructionExpenditures	-0.318819	1.012355e-74

Many of the variables related to expenditures on employee benefits were negatively correlated with incarceration rates. Another interesting finding was that the number of students receiving corporal punishment was highly positive correlated with incarceration rates.

The code used for the analysis can be found in the "10_correlations.ipynb."

Concluding Remarks

Our analysis allowed us to identify some enlightening patterns in the selected data sets as well as the shortcomings that accompany social science analysis. For instance, the tendency of the incarceration apparatus to move prisoners throughout the country limits the potential for strong statistical relationships based on geography. Even so, the type of data set that would allow for an analysis of such prisoner mobility would likely necessitate the inclusion of enough data to allow for the identification of particular prisoners, thereby raising privacy concerns. As such, a potential next step might be the generation of a labeled data set. As our analysis of the Vera Institute data indicated, the Black incarceration rate of a given county seems to indicate the presence of large correctional facilities. Accordingly, a labeled data set that included the particular type of correctional facility (e.g. federal prison, immigration detention, etc.) might allow for the use of more sophisticated machine learning techniques to explore causal relationships between incarceration and the economic realities on the ground.

As American society grapples with the history and implications of its mass incarceration system, increasingly sophisticated data collection and analysis techniques will hopefully shed light on potential solutions. Nonetheless, as a bipartisan consensus grows about the inefficacy of incarcerating American citizens en masse, it is important to take into account the durability of a system bolstered by so many government agencies and publicly traded companies. In other words, the mass incarceration apparatus appears to be adapting itself so as to target other populations—namely, non-citizens.[6]

Team Contributions

Sean focused on creating maps to visualize incarceration trends and property taxes.

Albert focused on extracting and merging the Common Core of Data, Civil Rights Data Collection, and Incarceration Trends datasets and performing correlation tests. He also maintained the GitHub repository.

Damaris focused on analyzing state by state and incarceration trends across time.

---

[6] *"Significantly increasing the number of immigrants in detention means record-high profits for private prisons"*(https://www.americanprogress.org/issues/democracy/reports/2019/08/30/473966/private-prisons-profiting-trump-administration/)