

Project 1

Ruth Walters

2025-02-19

Introduction

Economic mobility, or the ability of an individual to raise their economic status throughout their lifetime, is a marker of a healthy society. As economic mobility declines and income inequality rises throughout the United States, it is of increasing interest to determine which factors contribute to immobility. In this paper, we will investigate the correlation between economic, educational, and policy factors that contribute to economic mobility. We hypothesize that economic factors such as income inequality, will be most predictive of economic mobility.

Exploratory data analysis

```
# View NAs
nas <- colSums(is.na(mobility))
print(nas[nas > 0])
```

```
##           Mobility           Share01           Gini_99
##           12           32           32
##           Middle_class           Local_tax_rate           Local_gov_spending
##           32           1           2
##           School_spending           Student_teacher_ratio           Test_scores
##           10           30           36
##           HS_dropout           Colleges           Tuition
##           148           157           161
##           Graduation           Chinese_imports           Teenage_labor
##           160           19           32
##           Migration_in           Migration_out           Social_capital
##           17           17           19
##           Violent_crime
##           27
```

This dataset contains several rows for which one or more than one value is **NA**. Three steps were taken to eliminate NAs from the dataset while preserving its integrity.

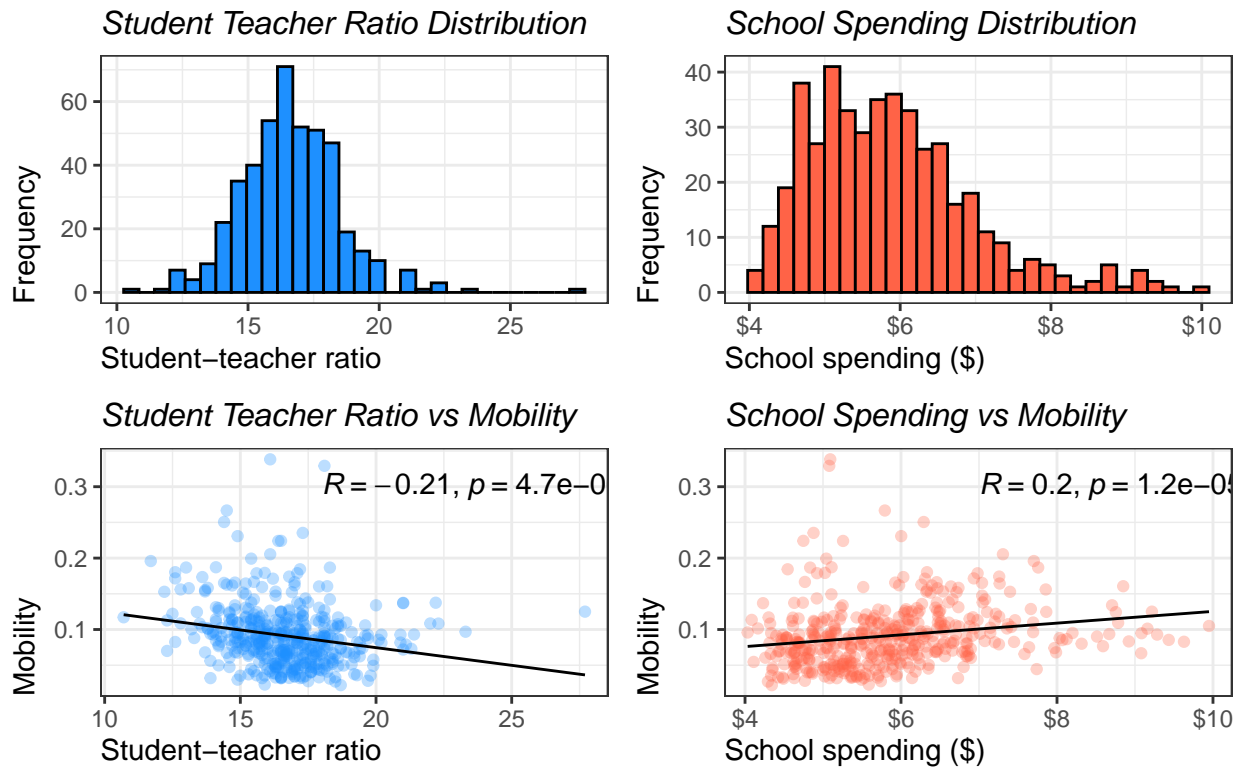
1. *Drop the 12 rows that do not contain a value for **Mobility*** | These rows are useless for linear analysis because they do not contain the variable we are attempting to predict.
2. *Drop the features that contain a high incidence of **NAs*** | Any features that contained more than 100 NAs were designated as too poor in quality to be useful for the linear model. While some of these features were used for exploratory data analysis, they were removed from the dataset prior to modeling.
3. *Drop all remaining **NA** values* | After removing the most **NA** values, a small amount of rows with NAs remained. These rows were dropped.

Simply dropping all rows with NAs would have resulted in a reduction of 39.4062078% of the data whereas our three-step procedure only resulted in a reduction of 14.5748988%.

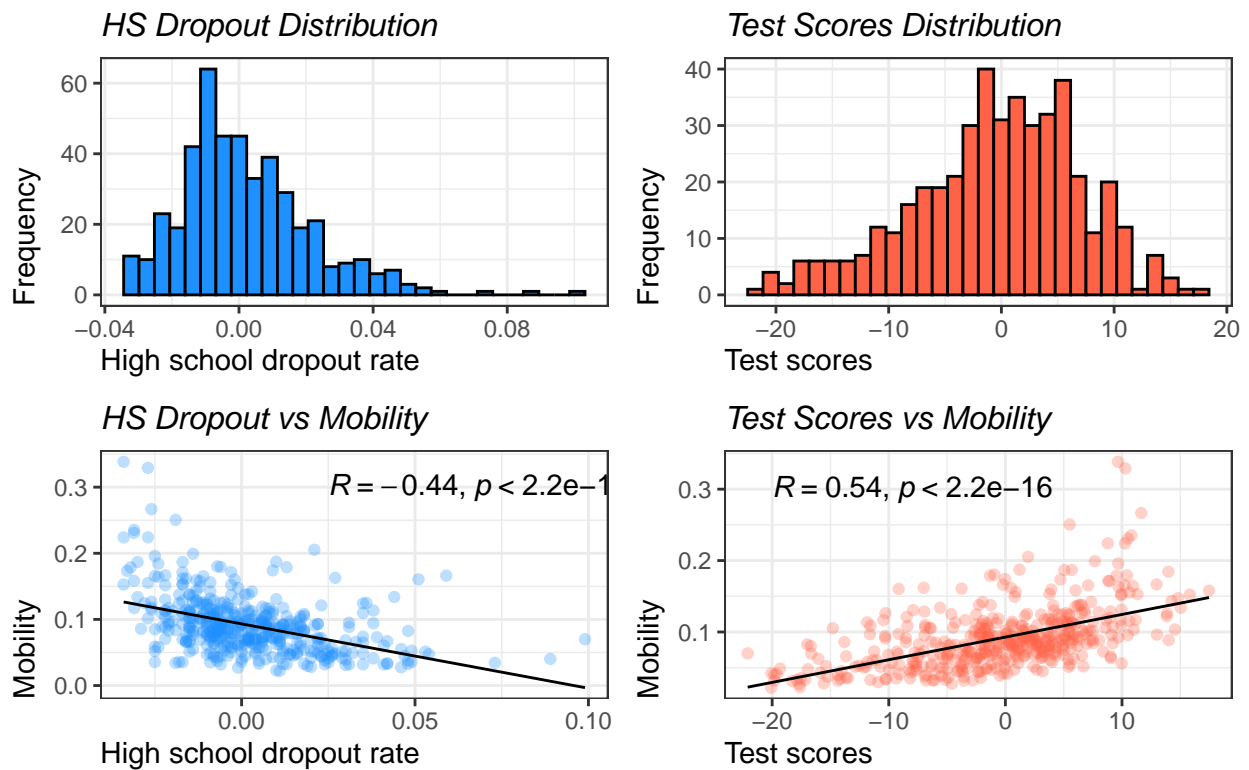
Additionally, qualitative (non-numeric and non-quantitative) variables such as those representing latitude and longitude, state/region names, and the ID tag, were removed.

Education analysis

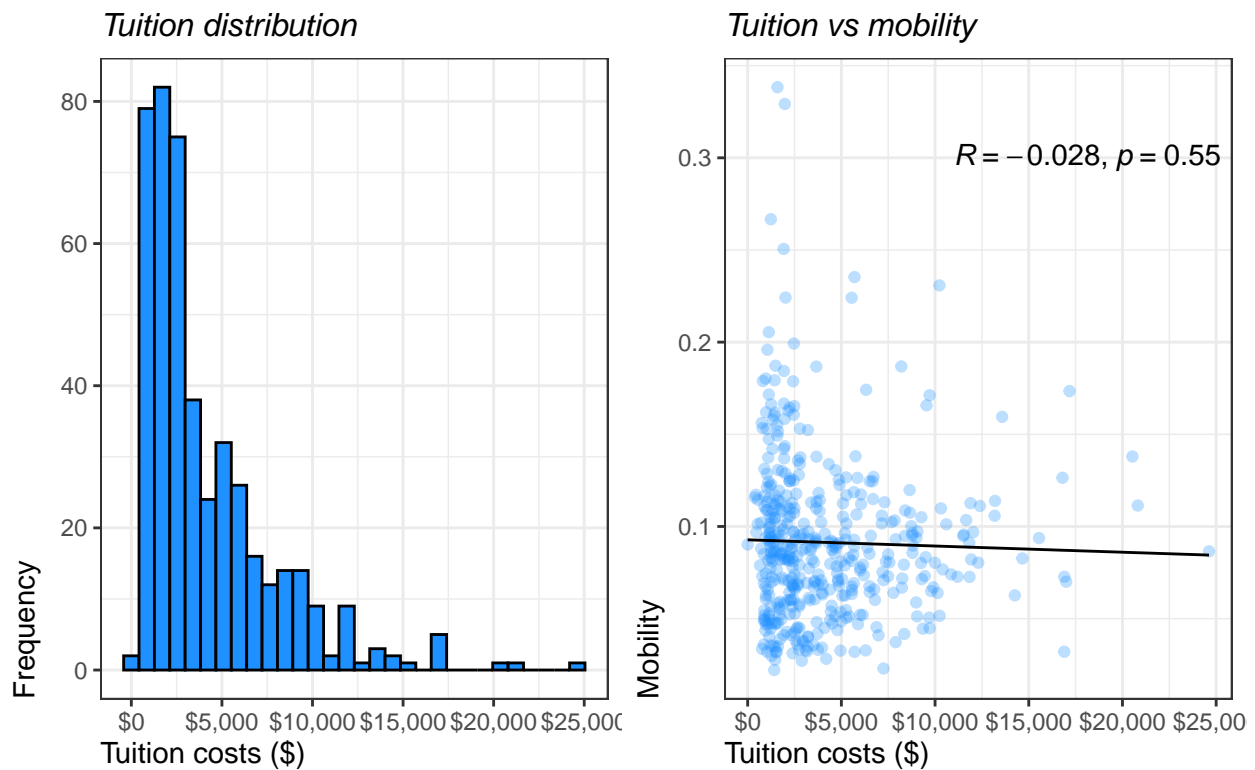
Educational investment as a predictor of economic mobility

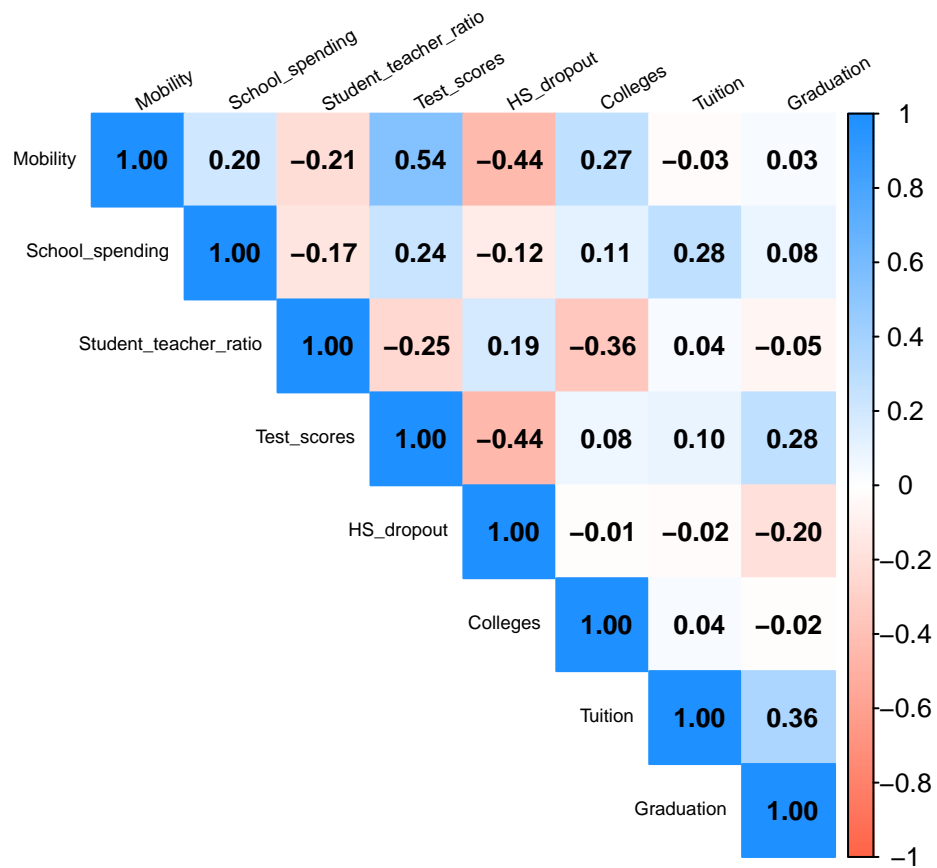


Educational outcomes as a predictor of economic mobility

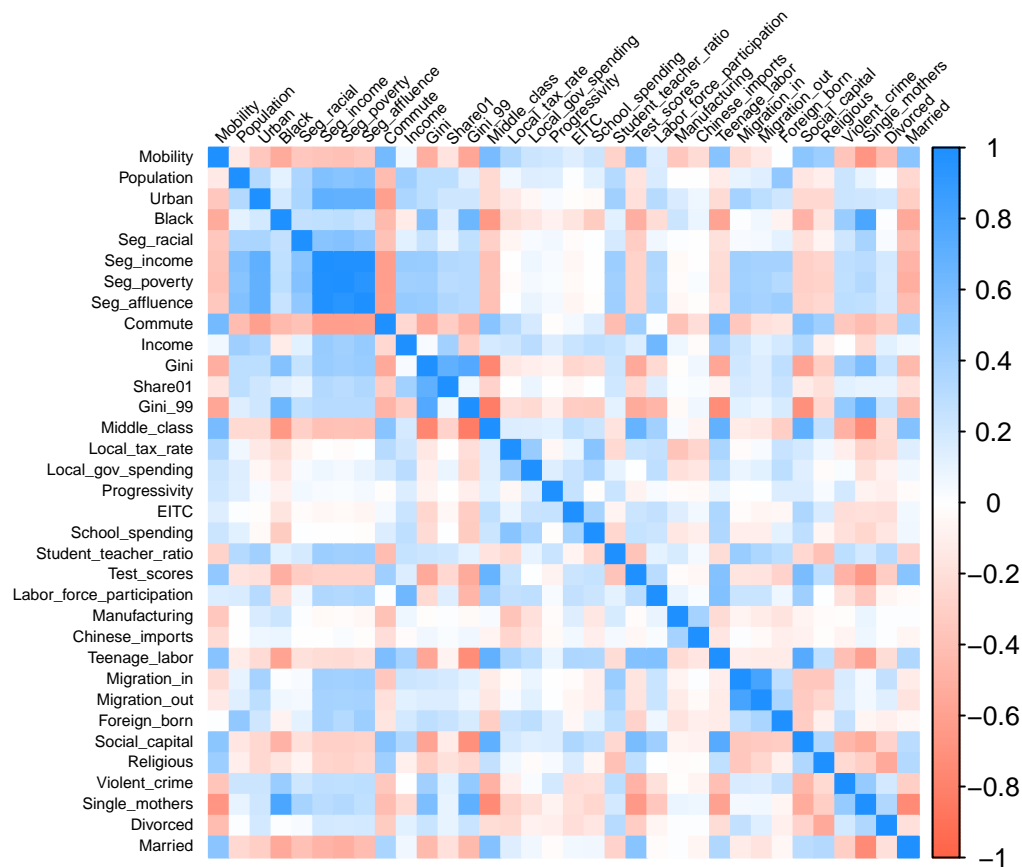


Tuition costs as a predictor of economic mobility



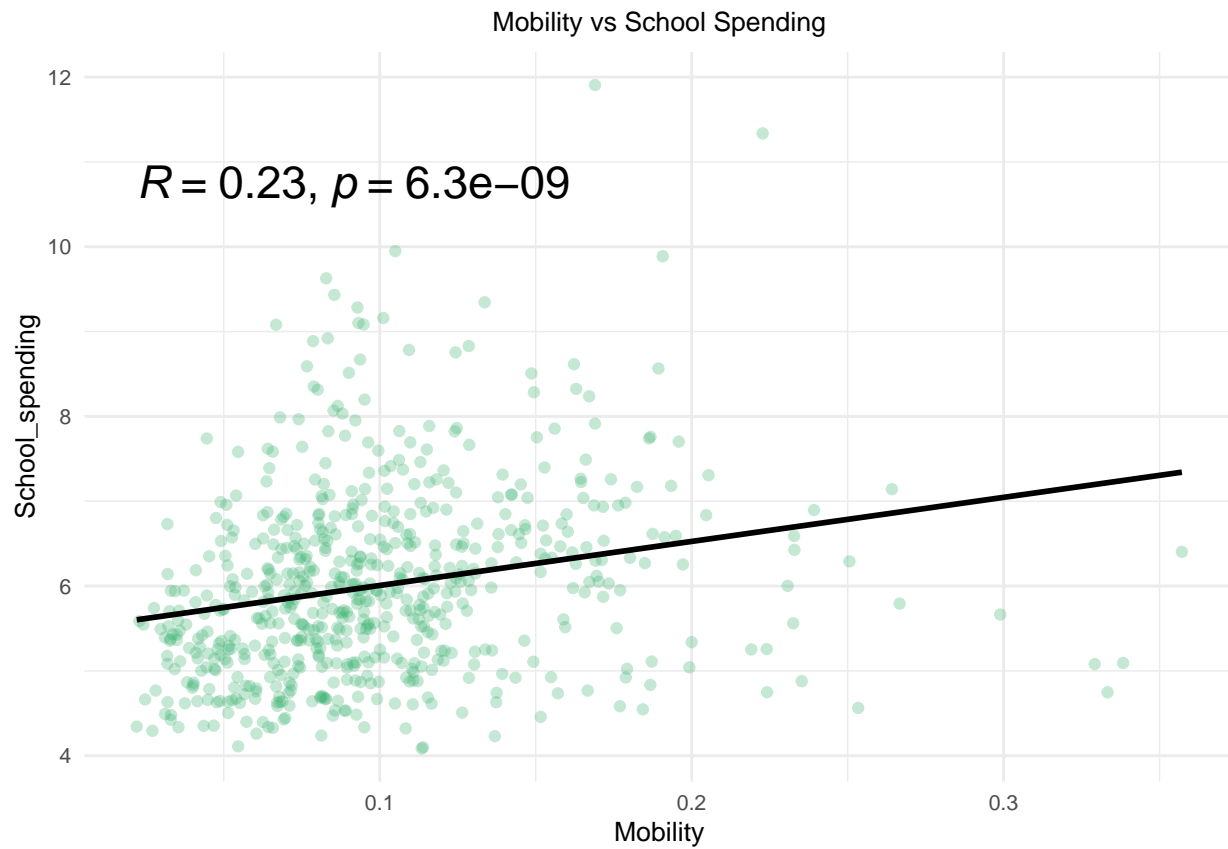


Policy variables



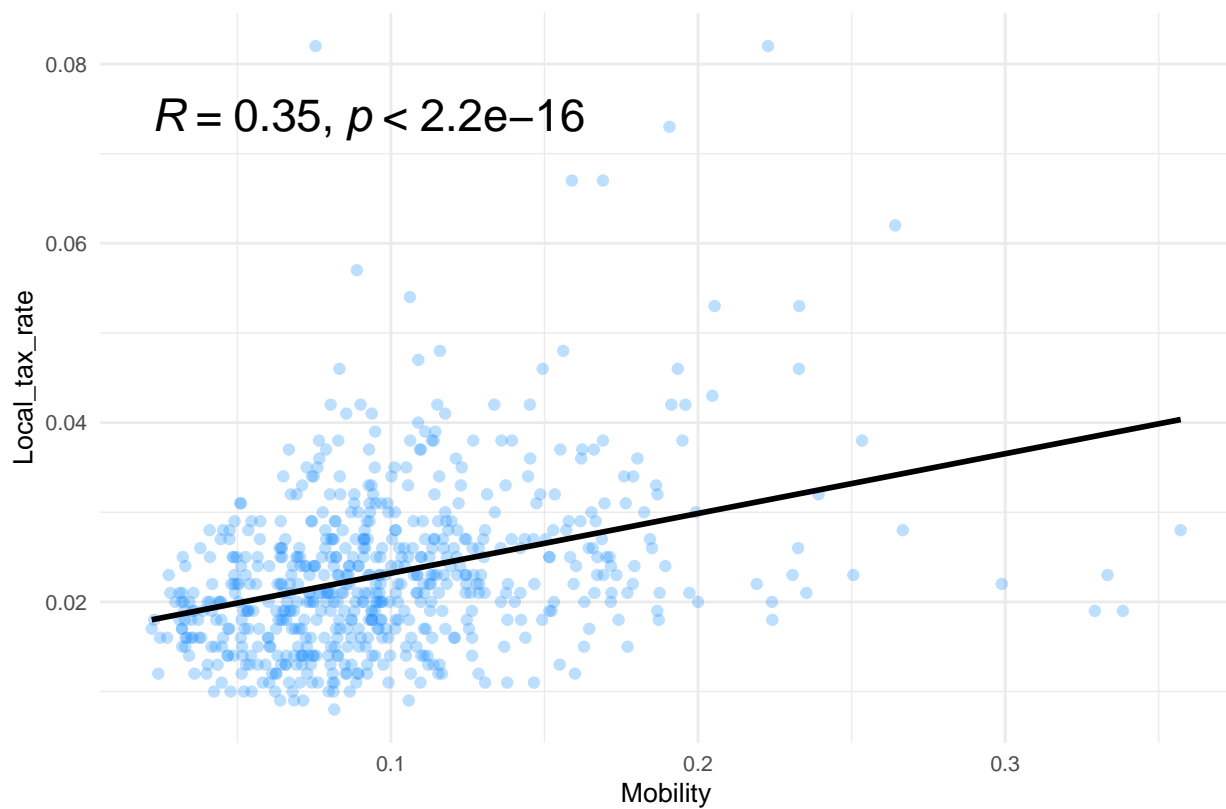
```
## # A tibble: 6 x 3
## # Rowwise:
##   Var1          Var2      Freq
##   <fct>        <fct>    <dbl>
## 1 Seg_affluence Seg_income 0.986
## 2 Seg_poverty   Seg_income 0.982
## 3 Seg_affluence Seg_poverty 0.943
## 4 Middle_class  Gini_99    -0.823
## 5 Migration_out Migration_in 0.817
## 6 Single_mothers Black       0.793

## `geom_smooth()` using formula = 'y ~ x'
```

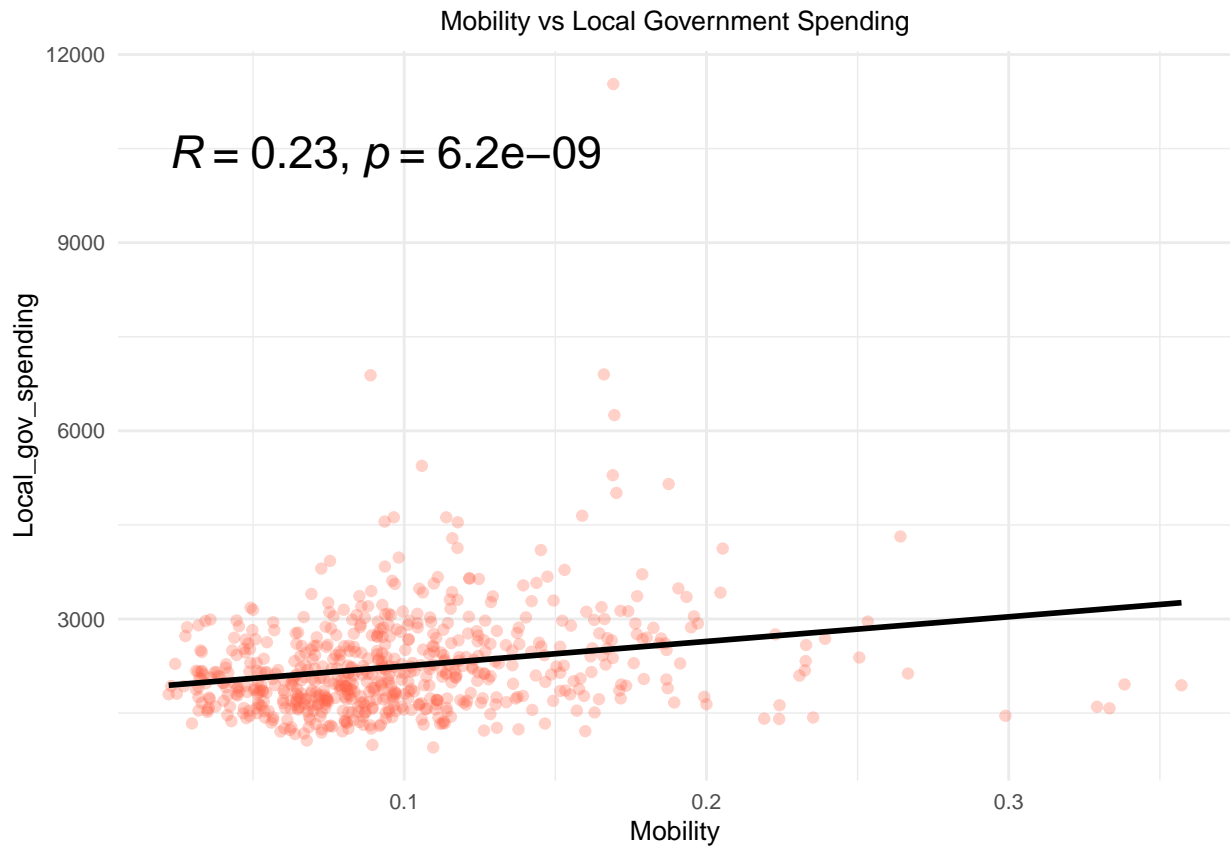
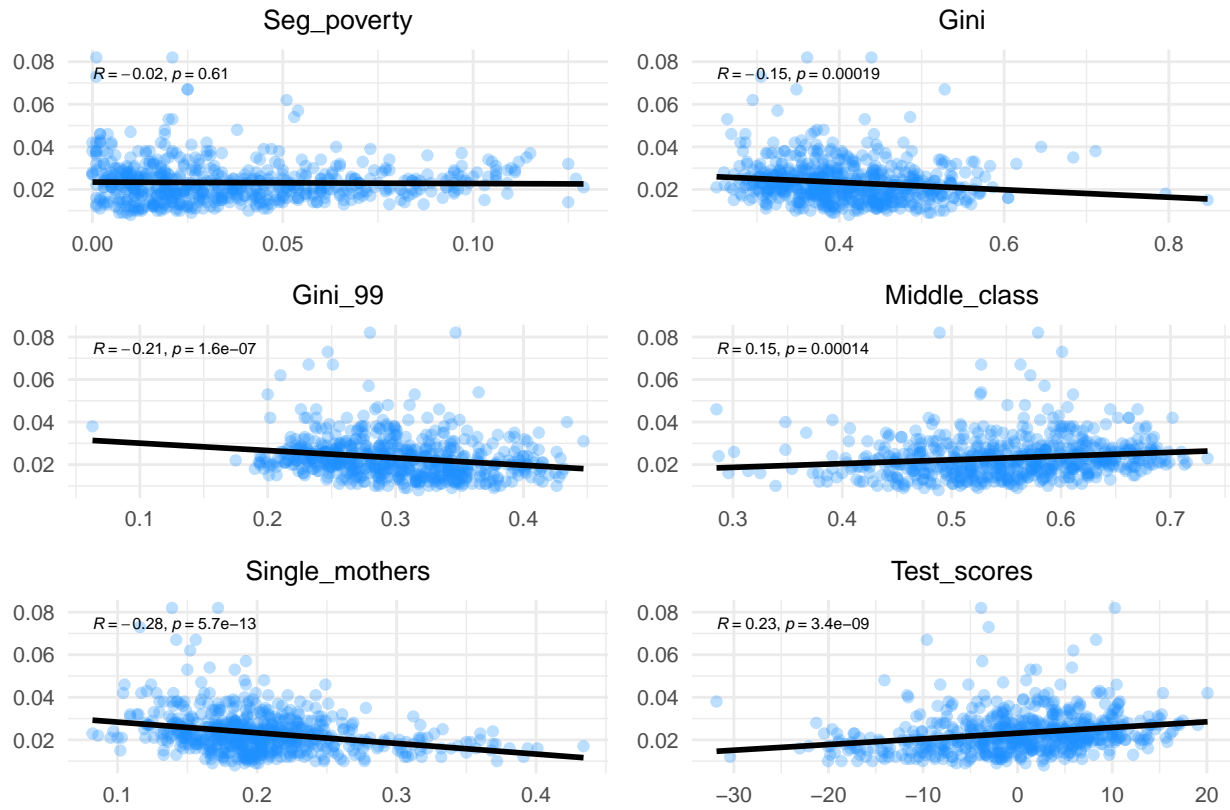


```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```

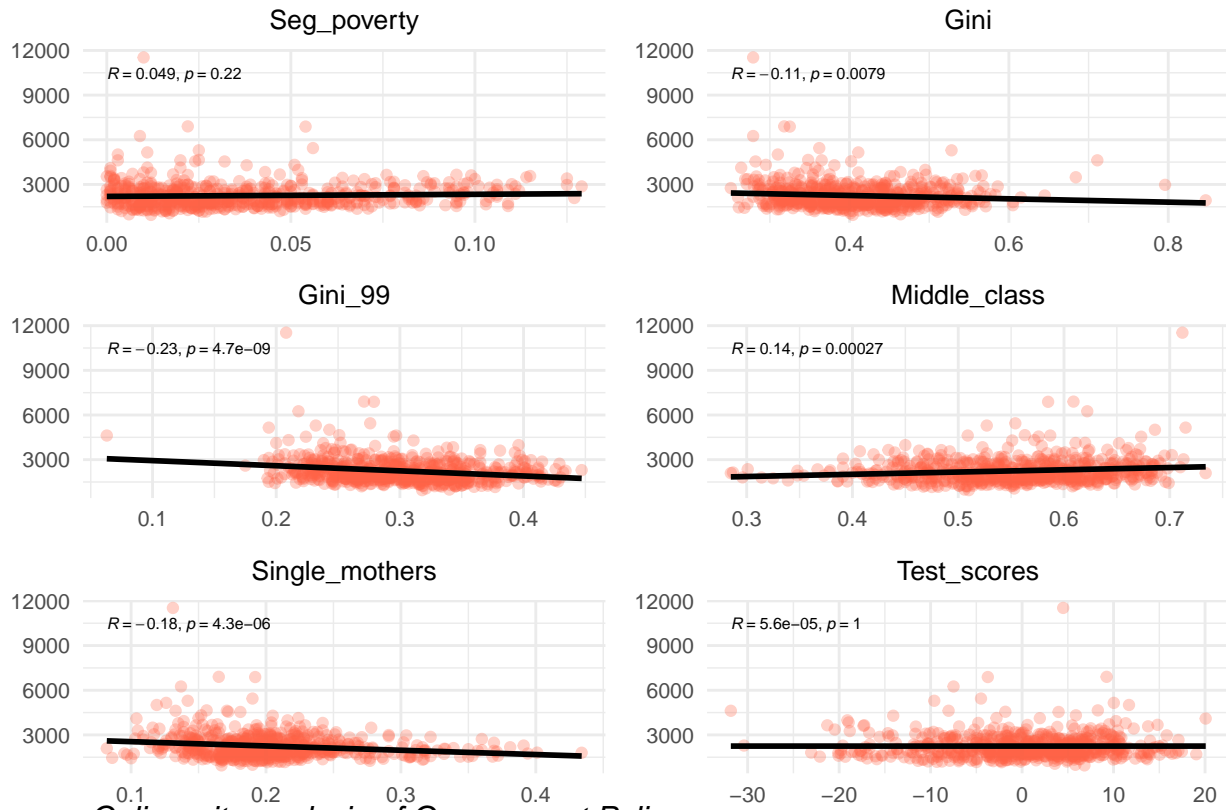
Demographic Variables vs School Spending



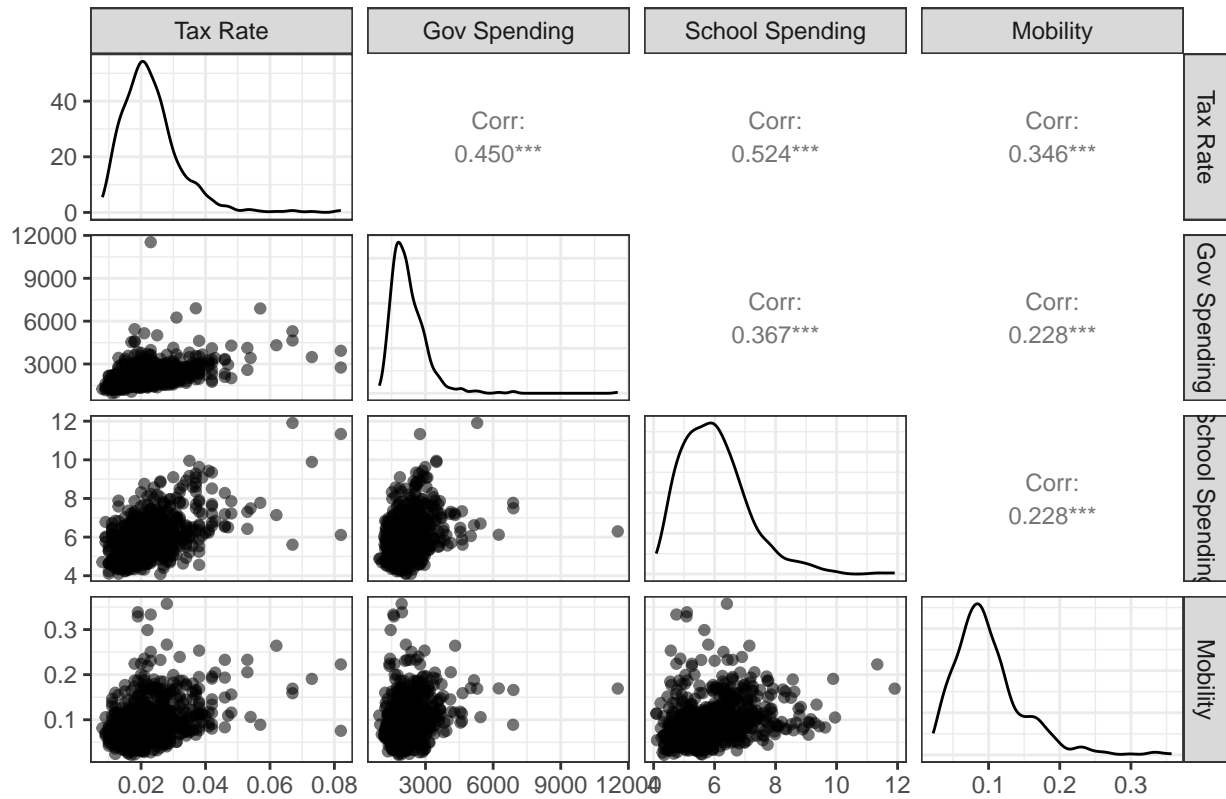
Demographic Variables vs Local Tax Rate



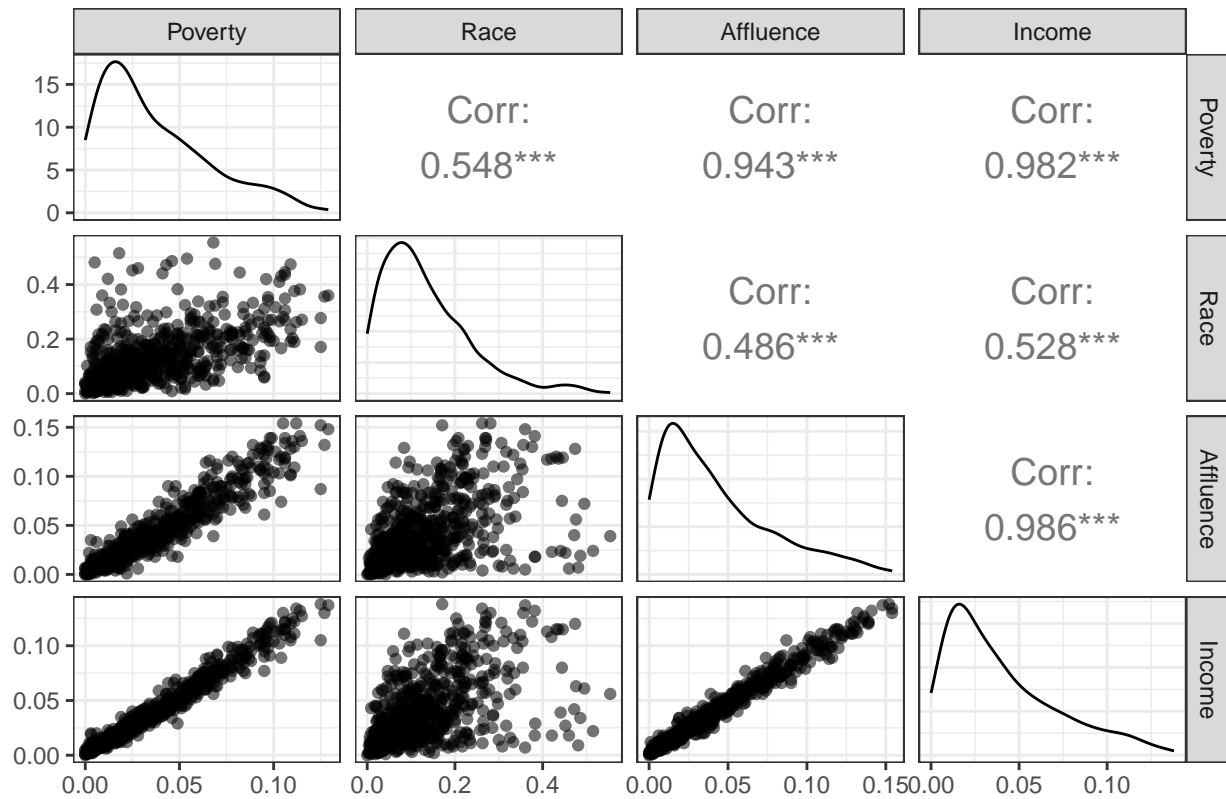
Demographic Variables vs Local Government Spending



Colinearity analysis of Government Policy

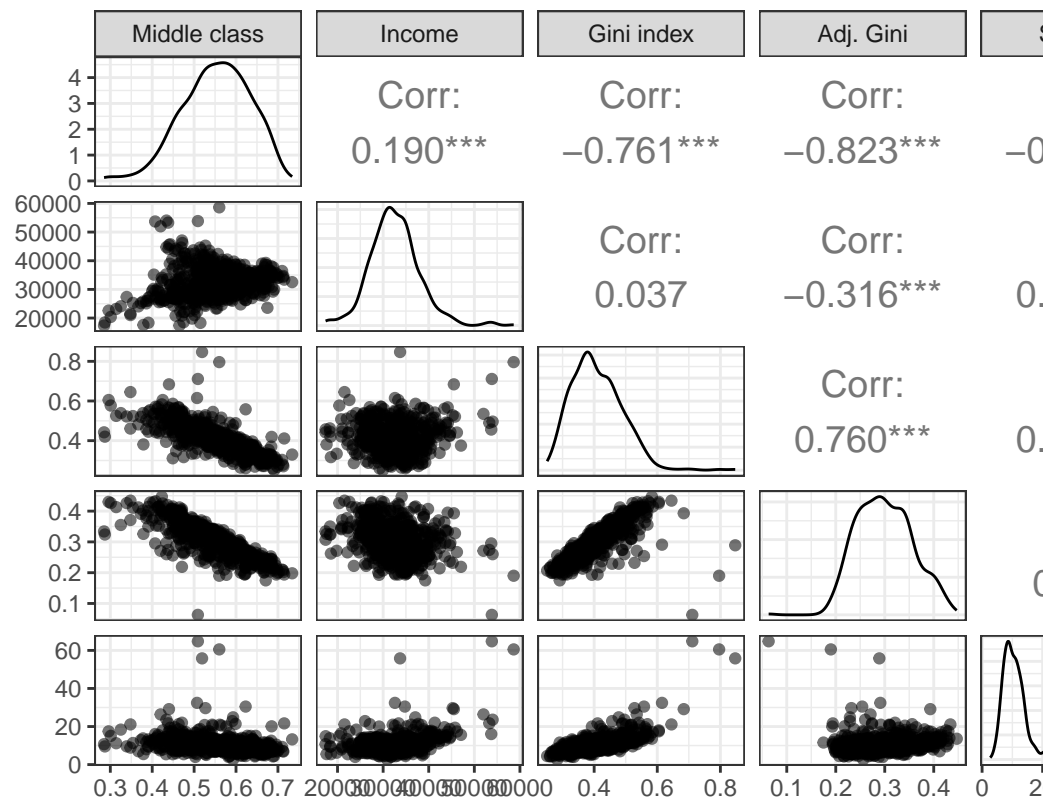


Colinearity analysis of segregation



While segregation on poverty lines is not particularly well correlated with segregation on racial lines, it is highly associated with segregation by affluence and segregation by income, which are also highly associated with each other. Since `Seg_poverty`, `Seg_affluence` and `Seg_income` are so strongly co-linear, `Seg_affluence` and `Seg_income` will be removed from the model.

Colinearity analysis of income and income inequality



Income and income inequality

The `Middle_class` variable is colinear with `Gini` and `Gini_99`, while the `Share01` variable is colinear with `Gini`. Additionally, `Gini` seems to be highly predictive of `Gini_99`. `Income` is not strongly associated with any of the other variables examined.

Social determinants of mobility

Segregation as a predictor of mobility

