Project 1

William Buckey, Sidney Gehring, Micheal Purtle, 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, with higher levels of income disparity correlating with lower upward mobility.

Exploratory data analysis

Data cleaning and variable selection

```
##            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 `NA`s 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 `NA`s* | 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 `NA`s remained. These rows were dropped.

Simply dropping all rows with `NA`s 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

One of the principal research questions concerns whether better education can predict higher levels of economic mobility. The selected educational variables include School_spending (which is also discussed as a policy variable), Student-teacher_ratio, Test_scores, HS_dropout, Colleges, Tuition, and Graduation. These variables represent key aspects of the education system, including funding, resource allocation, academic performance, and attainment levels, and are critical in assessing educational impact. Since many of the features with the most missing values were related to education, a new data set was created of just the educational factors where all rows with NA values were dropped.

Educational predictors of mobility
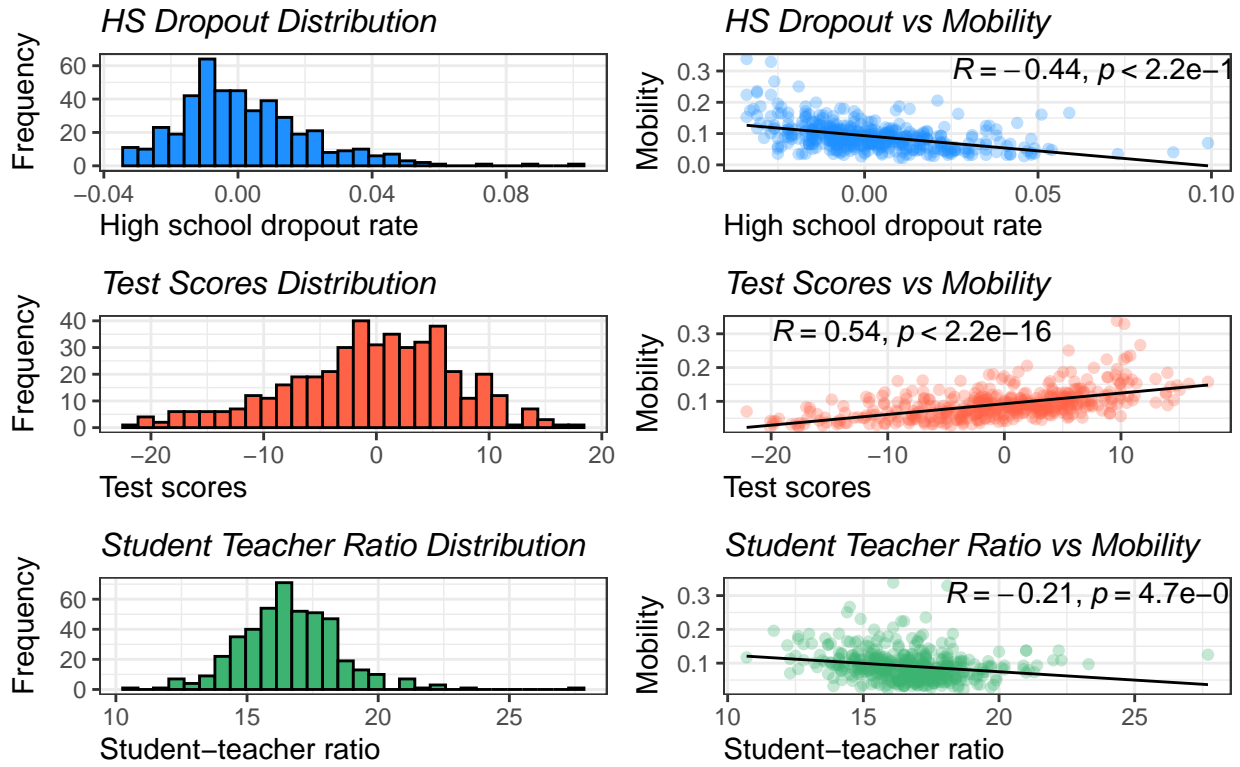
The initial step involved creating a correlation plot using pairwise relationships between social variables (see Appendix A.1). This analysis identified the top two predictive variables: test_scores, with a correlation of 0.54, and HS_dropout, with a correlation of -0.44. Student_teacher_ratio and colleges exhibited some correlation with mobility, whereas Tuition and Graduation showed little to no correlation.

To further analyze the most strongly correlated variable, the distribution of test_scores was examined. The data followed a bell curve with a slight left skew. A scatter plot of mobility and Test_scores revealed a positive linear relationship with some variation. To quantify this variation, the standard deviation was calculated as 7.43, aligning with the observed distribution. A similar analysis was conducted for HS_dropout. The distribution displayed a right skew, and the scatter plot indicated a negative linear relationship, consistent with the correlation value.

Policy analysis

By a similar procedure as above, three highly correlated variables related to government policy were identified:

1. Local tax rate | Fraction of all income going to local taxes

2. School expenditures | Average spending per pupil in public schools
3. Local government spending | Local government spending per capita

Further, while there were other variables such as Chinese_imports and Manufacturing that were considered inclusion in this category, both features were deemed to stem both from business practices and general economic state along with government policy, complicating analysis. On the other hand, variables like local tax rate and school spending are dependent on how the government chooses to raise and spend funds.

Government policy is first examined by identifying the relationship between the three variables that are selected as predictors and verifying that these variables have a positive correlation between themselves. To verify these relations, a correlation matrix was created using pairwise relations (see Appendix A.2). We highlight the policy-associated variables and display the top 5 variables that are correlated with each (also Appendix A.2).

School_spending, Local_tax_rate, and Local_gov_spending all have a similar steady correlation with each other, verifying the covariant reliance on government policy. The sample variables are then used to generate a relationship scatter plot between each specific predicting variable using scatterplots. A linear regression line and R & p values are displayed to examine any trend lines. Further analysis of the relationship between school spending, local tax rates, and local government spending and demographic factors like income inequality can be found in Appendix B.

Policy predictors of mobility
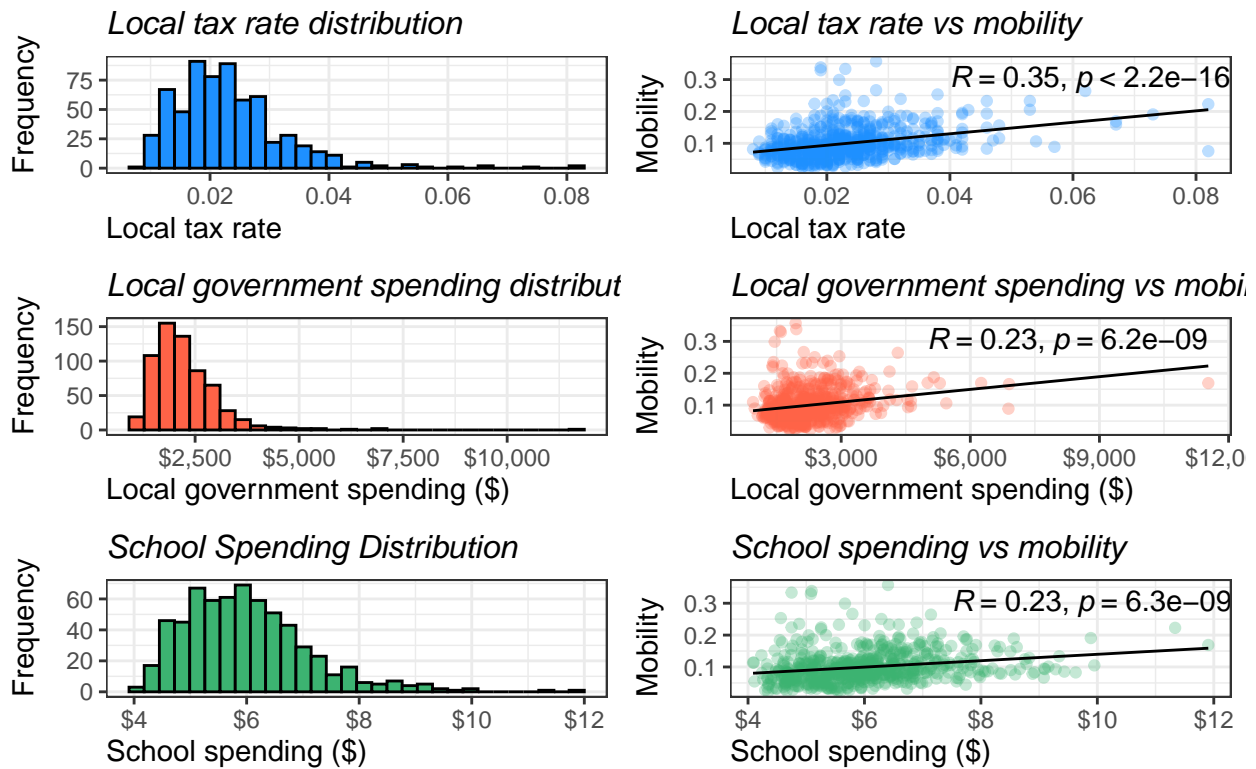
These results state to us that while there might be outlier data in all 3 relational graphs, there still shows to be somewhat of a linear relationship between Mobility and Tax rate. This is interesting because school_spending had a higher relationship with demographic factors. School_spending and local_gov_spending do not seem to show as strong of a linear trend, but since School_spending had a high relationship with other demographic factors, I still will use it in my model.

Segregation factors

Covariate analysis of segregation factors (`Seg_racial`, `Seg_poverty`, `Seg_affluence` and `Seg_income`- see **Appendix B.4**) indicated that, 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.

Additionally, other income related factors that are generally associated with integration were examined (see **Appendix B.3**). The `Middle_class` variable is colinear with `Gini` and `Gini_99`, while the `Share01` variable is co-linear 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
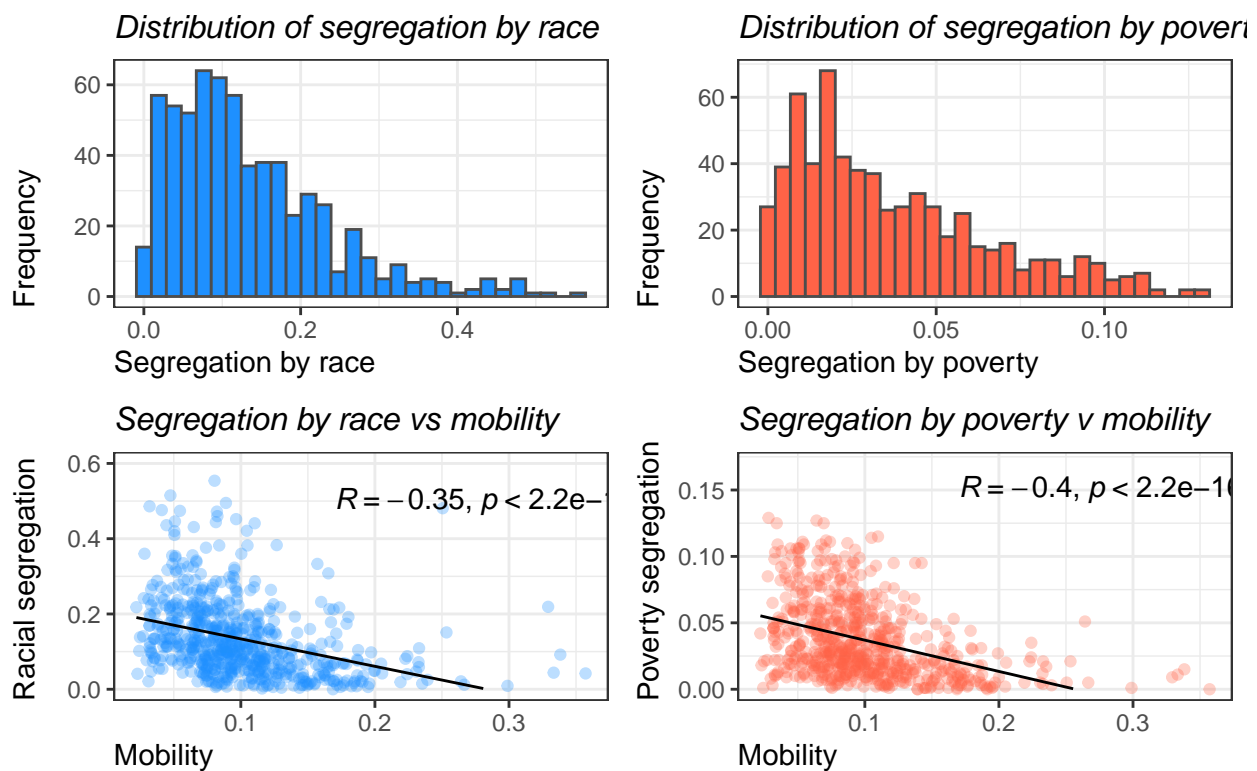
Modeling

In our analysis of the factors associated with mobility, we employed both exploratory data analysis (EDA) and a multiple linear regression model, and a beta regression model to determine which variables exhibit a strong relationship with our outcome measure. Based on the regression output and correlation values from our EDA, we identified a subset of variables that appear particularly influential in explaining mobility.

Linear model

```
lm.1 <- lm(formula = Mobility ~ .,
    data = mobility)
```

```
c1 <- summary(lm.1)
```

From the first model built (see Appendix C.1 for output), iSchool_spending and Local_tax_rate do not have statistical evidence to prove an effect on economic mobility. However, another model was made to solely focus on the effects of government policy on mobility position. This model calls the 3 predicting government policy variables, as well as the sample demographic variables used earlier. These variables are grouped and fitted into a regression formula. From here, the formula is fitted, and the initial summary is complete. From the full model, it is already shown that the p value for Local_gov_spending was 0.225, which indicates that it not statistically significant for predicting mobility. The variables that proved to be statistically significant were then moved to the next step.

1. **Educational Spending**

- *School Spending* | Higher per-pupil expenditures can signal greater investment in educational resources, facilities, and student support programs, thereby fostering improved long-term mobility outcomes.
- *Test Scores* | These serve as a proxy for overall educational quality and student achievement, and they often correlate positively with economic and social mobility.
- *Colleges* | The presence and density of colleges in an area can enhance the availability of higher education and skill development opportunities, contributing to upward mobility.

2. **Social and Demographic Factors**

- `Black` & `Seg_racial` | The proportion of Black residents and the degree of racial segregation in a community are crucial indicators, reflecting underlying social structures and potential barriers or pathways to mobility. Areas with higher racial segregation often experience limited socioeconomic opportunities.
- `Commute` | Longer or more prevalent commutes could signal broader labor markets or suburban sprawl. Where commutes are feasible, individuals may have access to a wider range of job opportunities.
- `Gini` | Income inequality is often inversely related to mobility; higher inequality can concentrate disadvantage and limit pathways for advancement.
- `Middle_class` & `Progressivity` | Communities with a robust middle class or more progressive tax structures may provide support systems (e.g., social services, quality infrastructure) that encourage upward mobility.
- `Manufacturing` | A heavy reliance on manufacturing might limit economic diversification; in many cases, deindustrialization or technological shifts in manufacturing can hinder long-term mobility prospects.
- `Migration_in` | Areas experiencing inbound migration may be more economically dynamic, suggesting better job prospects and growth, which in turn can improve mobility opportunities.
- `Religious` | Higher religiosity can be associated with stronger community networks or social support, potentially facilitating resource sharing and stability that foster mobility.
- `Violent_crime` | High crime rates often correspond to reduced social cohesion and economic investment, negatively affecting people's prospects for advancement.
- `Single_mothers` & `Divorced` | Family structure measures can indicate economic vulnerability or instability, impacting children's outcomes and the intergenerational transmission of opportunity.

3. **Government Spending**

- `Local_tax_rate`: The percentage of local taxes levied on residents and businesses. While higher rates can reduce disposable income, they can also fund public services that may boost economic mobility.
- `Local_gov_spending`: Reflects how much local authorities invest in public services, infrastructure, and community programs. Effective spending can expand opportunities and resources, potentially promoting upward mobility.

From a methodological standpoint, these variables are significant predictors in our regression and/or showed string correlation in EDA phase. To refine this model, we plan to:

- Check for Multicollinearity using Variance Inflation Factors (VIFs) to ensure no subset of variables is overly redundant.

- Check for Heteroskedasticity see if any of the variable have varying variance
- How influential are outliers in the model and see if we need to remove them
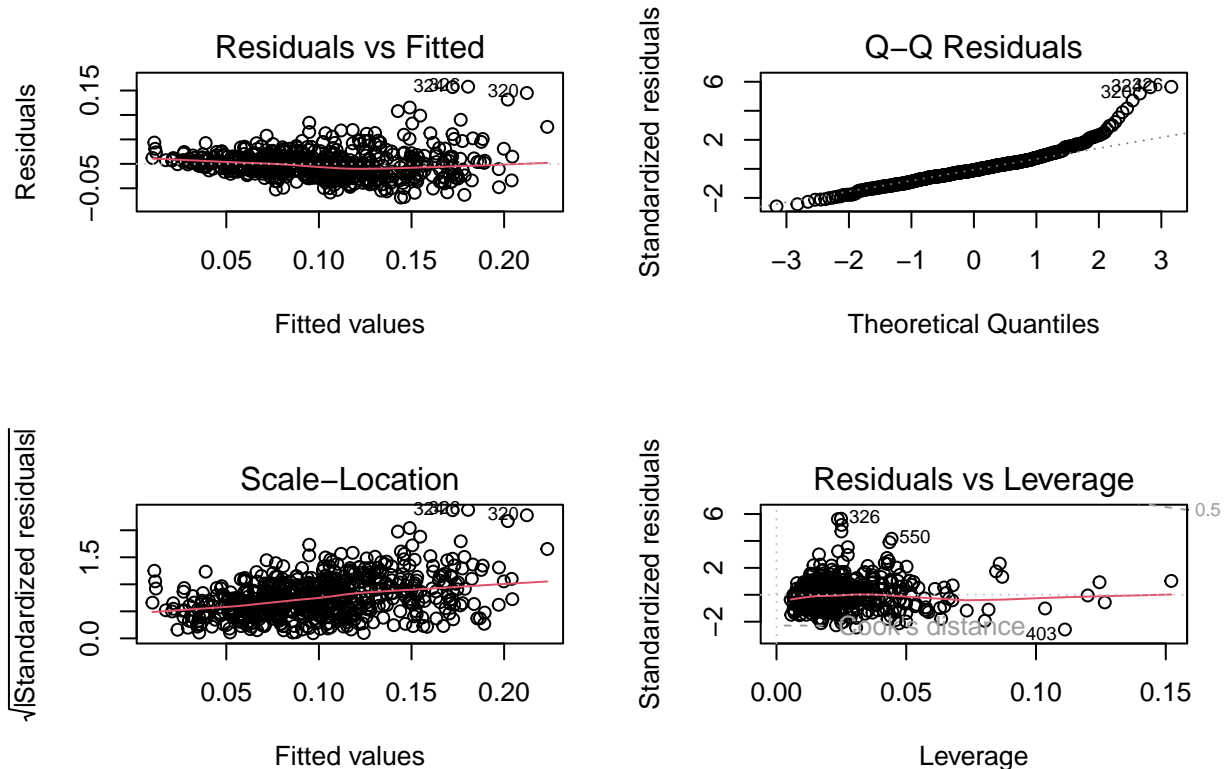
A numerical summary can be found in Appendix C.2.

We fitted the model, now let look for co-linearity between the variables then we can decide on how to deal with them.

```
##          Black      Seg_racial         Commute           Gini    Middle_class
##       2.394938        1.320021        2.932599       2.684005        4.097972
## Local_tax_rate   Progressivity  School_spending    Test_scores   Manufacturing
##       1.712259        1.229092        1.655802       2.264826        1.723279
##    Migration_in       Religious   Violent_crime       Divorced
##       1.389331        1.787569        1.613089       1.765527
```

Since the above factors are around 1-3, which suggest that colinearity isn't a huge concern. The highest VIF score is the "Middle_class" (about 4.097), which is borderline but still not extremely high. It's something to keep an eye on, but not a immediate concern.

```
##
## t test of coefficients:
##
##                    Estimate  Std. Error t value  Pr(>|t|)
## (Intercept)      0.05698998  0.03222521  1.7685 0.0774722 .
## Black           -0.04578197  0.01188349 -3.8526 0.0001291 ***
## Seg_racial      -0.08977309  0.01695497 -5.2948 1.658e-07 ***
## Commute          0.01104064  0.01722973  0.6408 0.5218967
## Gini            -0.01518672  0.02609392 -0.5820 0.5607773
## Middle_class     0.18999556  0.02912158  6.5242 1.424e-10 ***
## Local_tax_rate   0.56297971  0.24900925  2.2609 0.0241137 *
## Progressivity    0.00463411  0.00109617  4.2276 2.720e-05 ***
## School_spending -0.00063444  0.00158336 -0.4007 0.6887862
```

6

```
## Test_scores     -0.00019584  0.00022102 -0.8861 0.3759231
## Manufacturing    -0.16876108  0.02165467 -7.7933 2.777e-14 ***
## Migration_in     -0.34146670  0.12459017 -2.7407 0.0063077 **
## Religious         0.04806408  0.00958220  5.0160 6.905e-07 ***
## Violent_crime    -2.50481983  1.39186687 -1.7996 0.0724096 .
## Divorced         -0.52885532  0.08737476 -6.0527 2.471e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The result is a more refined model of each predictors estimated coefficient, its standard error, the t-statistic (coefficient / standard error), and the associated p-value (See Appendix C.3 for output). In short, there are variables like Black, Seg_Racial, Middle_class, Progressivity, Manufacturing, Religious, Divorced, and Local_tax_rate show significant relationships while some of the others do not reach conventional significance thresholds. And we in fact use the more significant variables when choosing our next model below.

Having refined our model to include only statistically significant variables, we can now delve deeper into the data by identifying influential observations. The Q-Q plot below helps us pinpoint these observations, highlighting points that deviate from the expected distribution. These deviations may indicate outliers or data points that have a disproportionate impact on the model's performance, warranting further investigation.
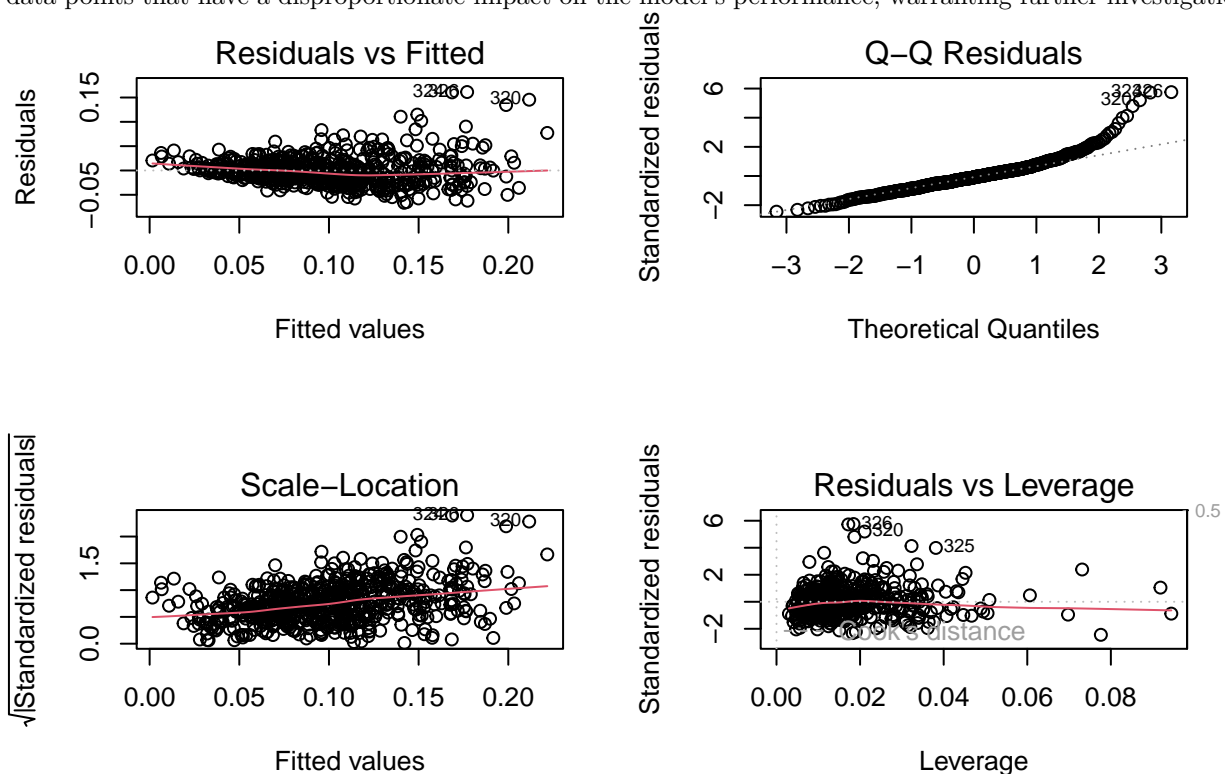

After examining the diagnostic plots (Residuals vs. Fitted, Q-Q Plot, Scale-Location, and Residuals vs. Leverage), we identified observations 320, 325, 326, and others as potential outliers or influential points. These points appeared to deviate substantially from the overall trend, suggesting they might exert an undue influence on the regression results.

Why Remove Outliers?

1. *Influence on Parameter Estimates* | Outliers can disproportionately affect the estimated coefficients, leading to skewed interpretations.
2. *Violation of Model Assumptions* | If extreme points violate assumptions of normality or homoscedasticity, they can compromise the validity of the model's inferences.
3. *Model Fit* | Removing influential points may improve the model fit and reveal a clearer relationship

among the remaining data.

Now we just have to remove the outliers and refit the model.
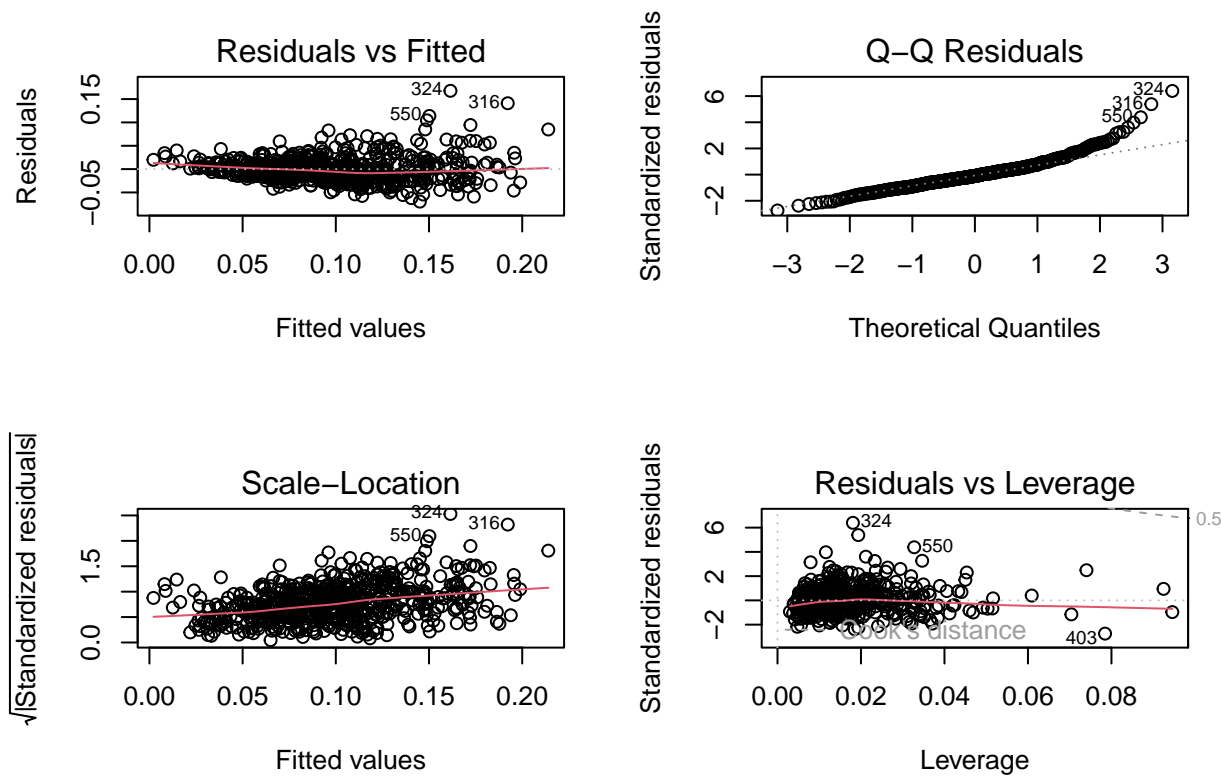
This is likely the best we can achieve using a simple linear regression model, given that we're attempting to predict a probability (mobility). Because probabilities are bounded between 0 and 1, linear regression can struggle to provide accurate predictions or valid inferences in this context. As a result, we may need a more advanced modeling technique—such as logistic regression or another specialized method—to capture the probabilistic nature of mobility. This modeling mismatch is probably why we've encountered multiple issues when trying to predict mobility with a simple linear approach.

Beta regression

For the next model, I plan to use beta regression because it is specifically designed for response variables that lie between 0 and 1. However, before proceeding with beta regression, we need to verify that all values of our response variable lie strictly within the (0, 1) interval—a key assumption for beta regression.

We can confirm this in R with the following code:

```r
all(high_cor$Mobility > 0 & high_cor$Mobility < 1)
```

```
## [1] TRUE
```

If this returns TRUE, it confirms that all values meet the (0, 1) requirement, and we can confidently move forward with the beta regression analysis.

While the histogram does not provide sufficient evidence that mobility follows a beta distribution, we can use the Kolmogorov-Smirnov (K-S), which essentially asks: 'How likely is it that we would observe these two samples if they were drawn from the same probability distribution?'

Fit of Beta Distribution
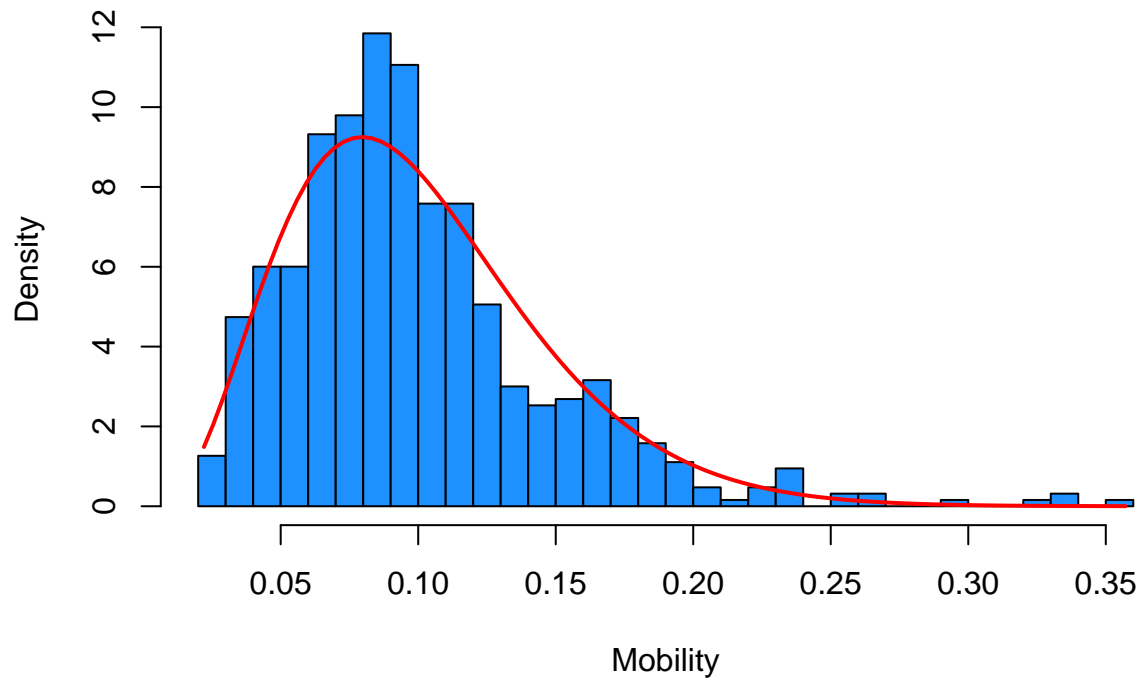

```
## 
##   Asymptotic one-sample Kolmogorov-Smirnov test
## 
## data:  high_cor$Mobility
## D = 0.052811, p-value = 0.05856
## alternative hypothesis: two-sided
```

The p-value from the one-sample Kolmogorov-Smirnov test is 0.05856, which is slightly above the conventional 0.05 threshold. This suggests that we do not have strong evidence to reject the null hypothesis that our data follow a beta distribution.
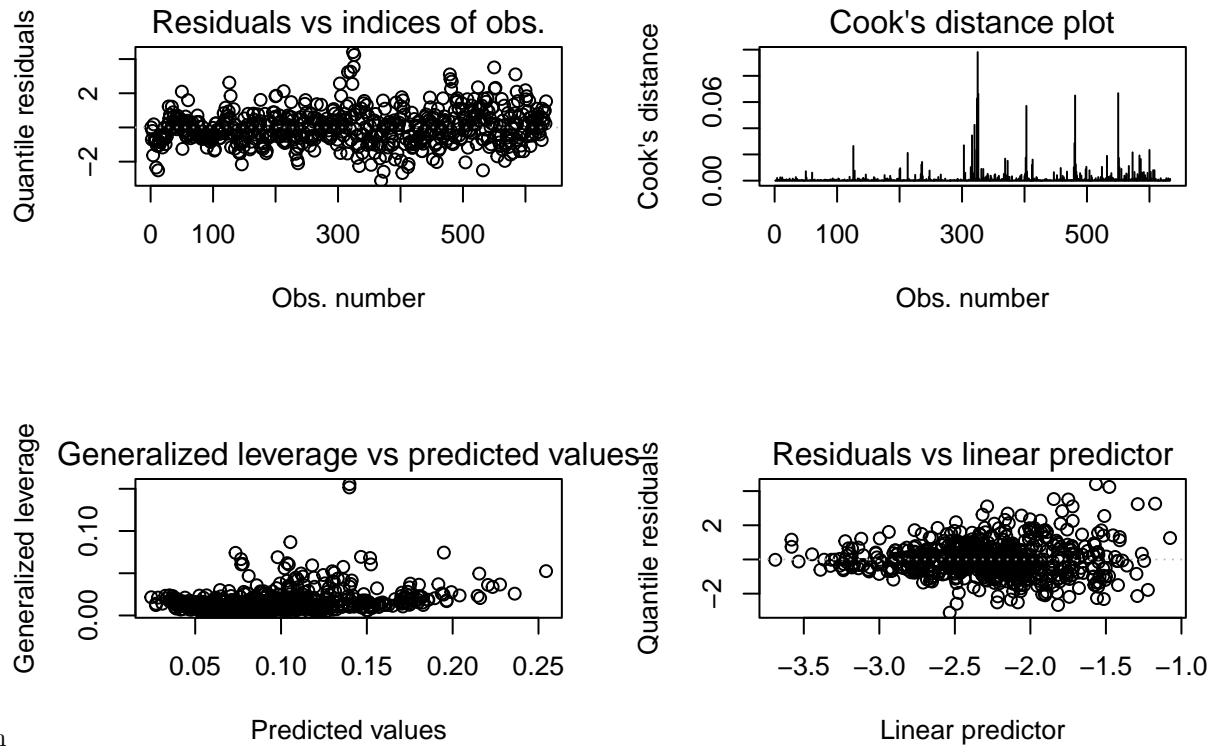
Beta regression

Overall, the beta regression model appears to be performing well based on several diagnostic indicators. First, the significance of the predictors in the mean model is high, with all coefficients (except Migration_in, which is still below the 0.05 threshold) showing very low p-values. This suggests that each variable contributes meaningfully to explaining the variability in Mobility. Furthermore, the pseudo R-squared value of approximately 0.75 indicates that the model accounts for a substantial portion of the variation in the data—quite high for a beta regression.

Looking at the diagnostic plots, the residuals versus observation indices and the residuals versus the linear predictor do not exhibit strong patterns, suggesting no obvious violations of model assumptions. The Cook's distance plot reveals that no single observation exerts undue influence on the model, and generalized leverage values are mostly low, indicating limited leverage points. Lastly, the estimated precision $\phi \approx 150$ is relatively large, pointing to a tightly clustered distribution around the fitted values. Taken together, these diagnostics imply that the beta regression model is a good fit, although you may want to confirm its predictive performance through additional fit metrics or cross-validation.

```
## AIC: -2972.293
```

```
## BIC: -2923.338
```

```
## Mean RMSE (5-fold CV): 0.02762679
```

The model's fit is strong, as indicated by the AIC (-2972.293) and BIC (-2923.338), which suggest excellent in-sample fit, with lower values signifying a better model. The mean RMSE from the 5-fold cross-validation is 0.0276, meaning that, on average, the model's predictions are off by just 2.8 percentage points—indicating good predictive accuracy. Overall, these results suggest that your beta regression model is well-calibrated, performing well both in terms of fit and prediction. However, comparing these metrics across different models can provide additional insights into the relative performance.

The negative AIC and BIC values indicate that the model has a high likelihood relative to its complexity, though these measures are most informative when comparing across multiple models. Meanwhile, the mean RMSE of approximately 0.028 suggests that, on average, the model's predictions deviate from the actual values by about 2.8 percentage points on the 0–1 scale, which is fairly small. Taken together, these results suggest that the beta regression model fits the data well, although further comparisons with alternative

10

models or additional diagnostics would provide a more comprehensive assessment.

Appendices

Appendix A

A.1 Education covariate analysis

	Mobility	School_spending	Student_teacher_ratio	Test_scores	HS_dropout	Colleges	Tuition	Graduation
Mobility	1.00	0.20	−0.21	0.54	−0.44	0.27	−0.03	0.03
School_spending		1.00	−0.17	0.24	−0.12	0.11	0.28	0.08
Student_teacher_ratio			1.00	−0.25	0.19	−0.36	0.04	−0.05
Test_scores				1.00	−0.44	0.08	0.10	0.28
HS_dropout					1.00	−0.01	−0.02	−0.20
Colleges						1.00	0.04	−0.02
Tuition							1.00	0.36
Graduation								1.00
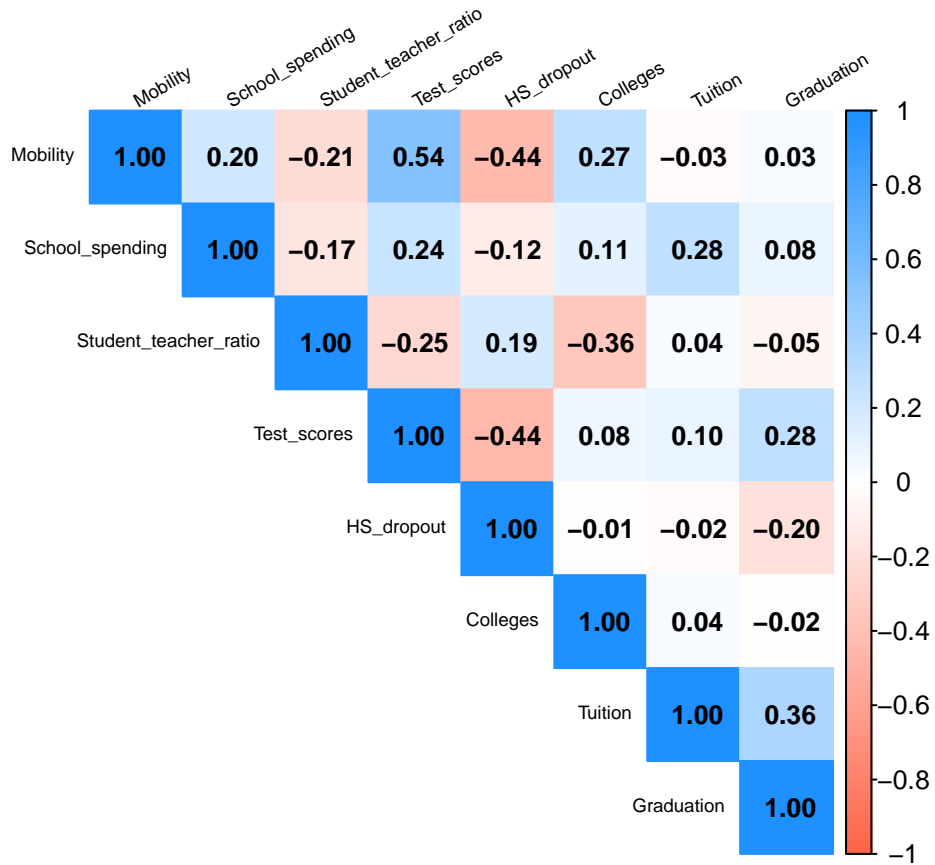
A.2 Policy covariate analysis
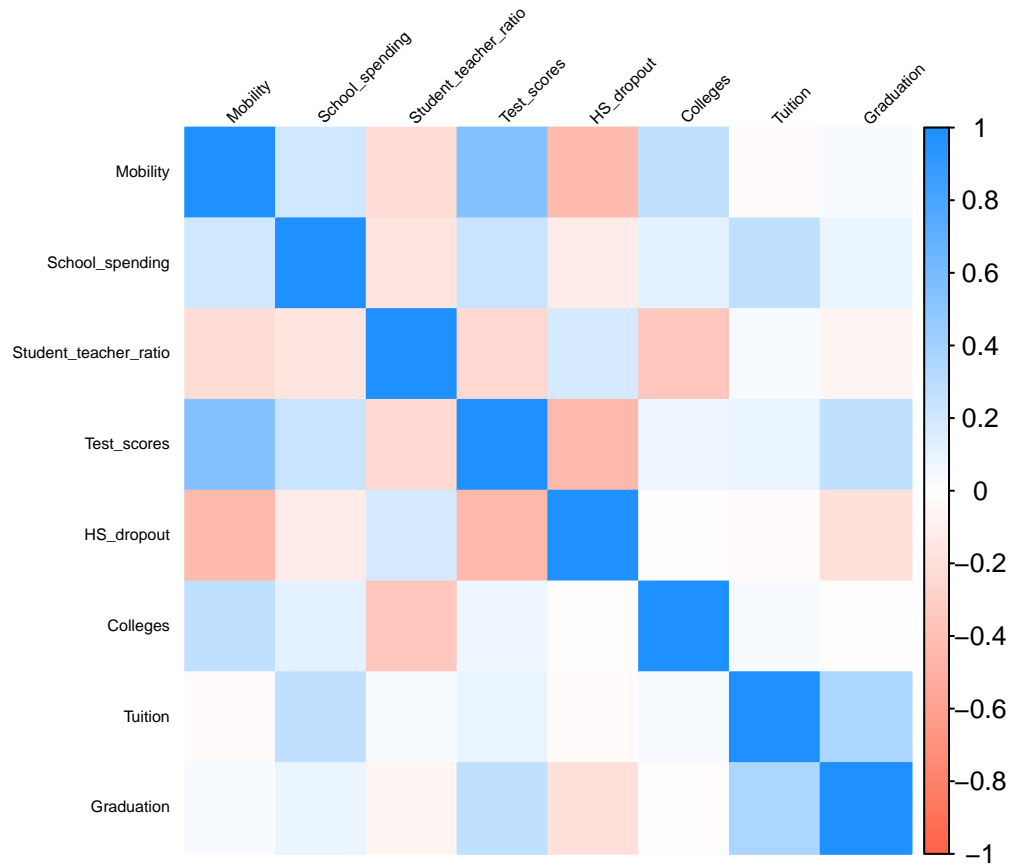


```
## [1] "School spending: "

## # A tibble: 5 x 3
## # Rowwise:
##   Var1            Var2              Freq
##   <fct>           <fct>             <dbl>
## 1 School_spending Local_tax_rate    0.524
## 2 School_spending EITC              0.382
## 3 School_spending Local_gov_spending 0.367
## 4 Teenage_labor   School_spending   0.344
## 5 School_spending Gini_99          -0.325

## [1] "Local tax rate: "

## # A tibble: 5 x 3
## # Rowwise:
##   Var1              Var2             Freq
##   <fct>             <fct>            <dbl>
## 1 School_spending   Local_tax_rate   0.524
## 2 Local_gov_spending Local_tax_rate  0.450
## 3 Manufacturing     Local_tax_rate  -0.373
## 4 Teenage_labor     Local_tax_rate   0.372
## 5 Local_tax_rate    Mobility         0.346

## [1] "Local government spending: "

## # A tibble: 5 x 3
```

```
## # Rowwise:
##   Var1               Var2               Freq
##   <fct>              <fct>              <dbl>
## 1 Local_gov_spending Local_tax_rate     0.450
## 2 School_spending    Local_gov_spending 0.367
## 3 Local_gov_spending Income             0.301
## 4 Foreign_born       Local_gov_spending 0.295
## 5 Teenage_labor      Local_gov_spending 0.293
```
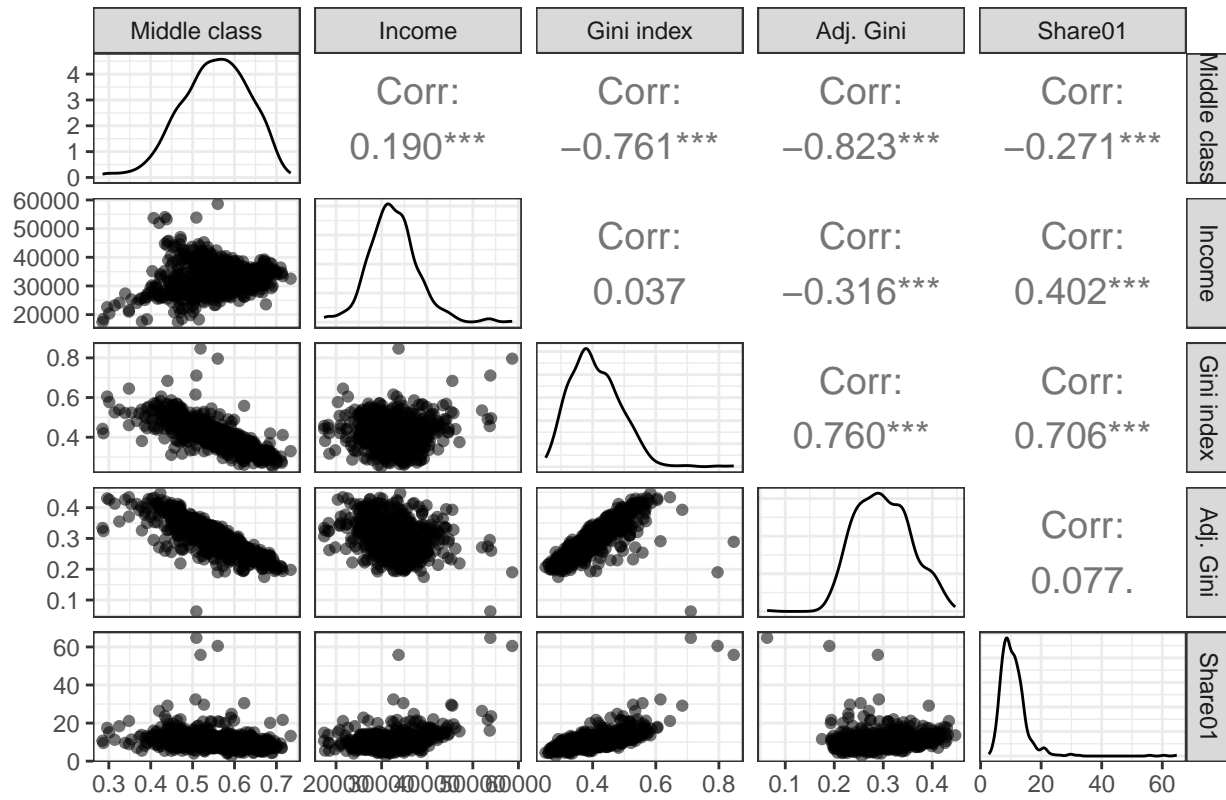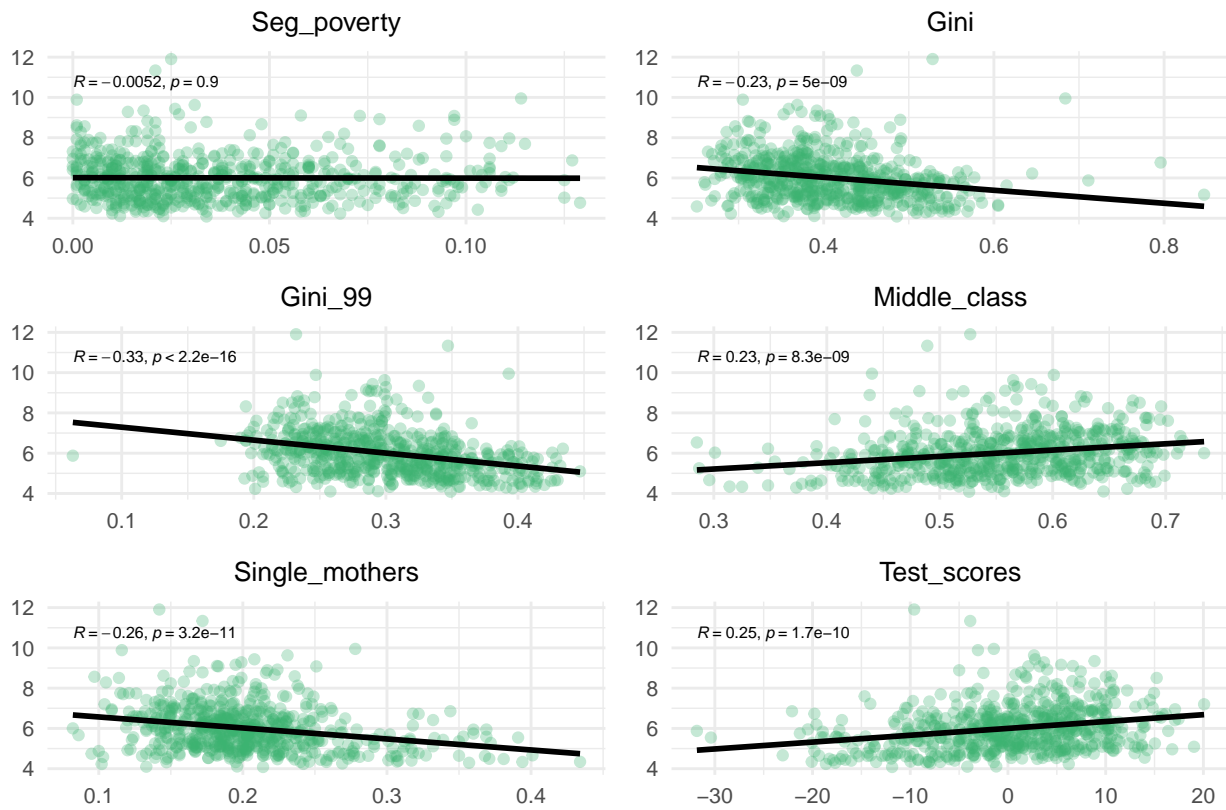
Colinearity analysis of segregation
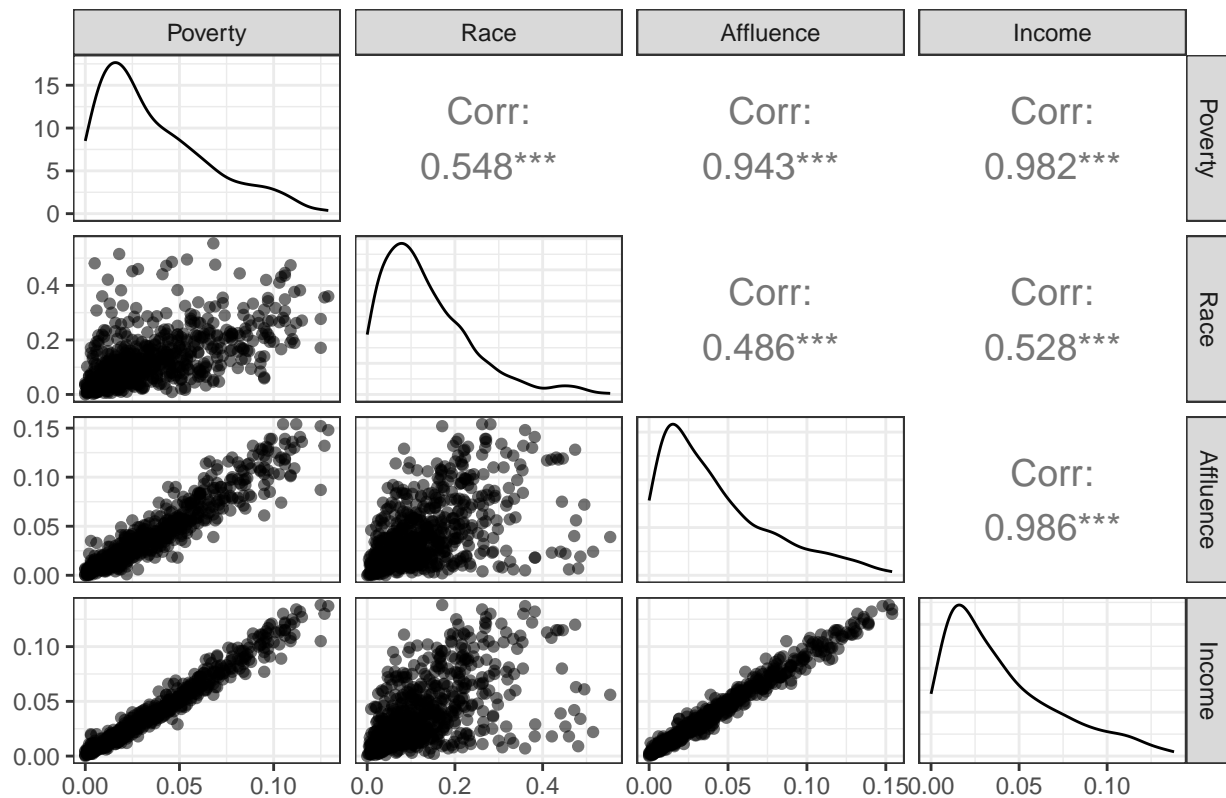
A.3 Segregation covariate analysis

Colinearity analysis of income and income inequality

Appendix B

Demographic Variables vs School Spending

Demographic Variables vs Local Tax Rate

Seg_poverty

$R = -0.02, p = 0.61$

Gini

$R = -0.15, p = 0.00019$

Gini_99

$R = -0.21, p = 1.6e{-}07$

Middle_class

$R = 0.15, p = 0.00014$

Single_mothers

$R = -0.28, p = 5.7e{-}13$

Test_scores

$R = 0.23, p = 3.4e{-}09$
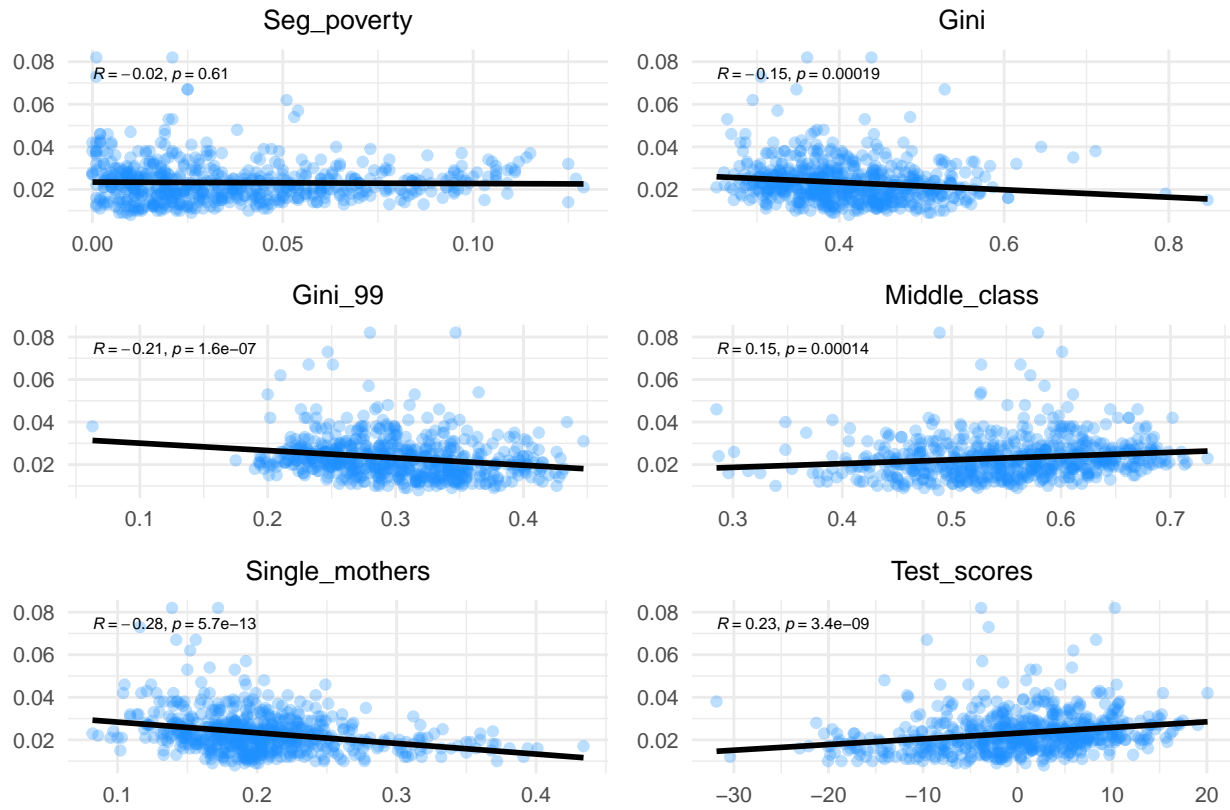
Demographic Variables vs Local Government Spending

Seg_poverty

$R = 0.049, p = 0.22$

Gini

$R = -0.11, p = 0.0079$

Gini_99

$R = -0.23, p = 4.7e{-}09$

Middle_class

$R = 0.14, p = 0.00027$

Single_mothers

$R = -0.18, p = 4.3e{-}06$

Test_scores

$R = 5.6e{-}05, p = 1$
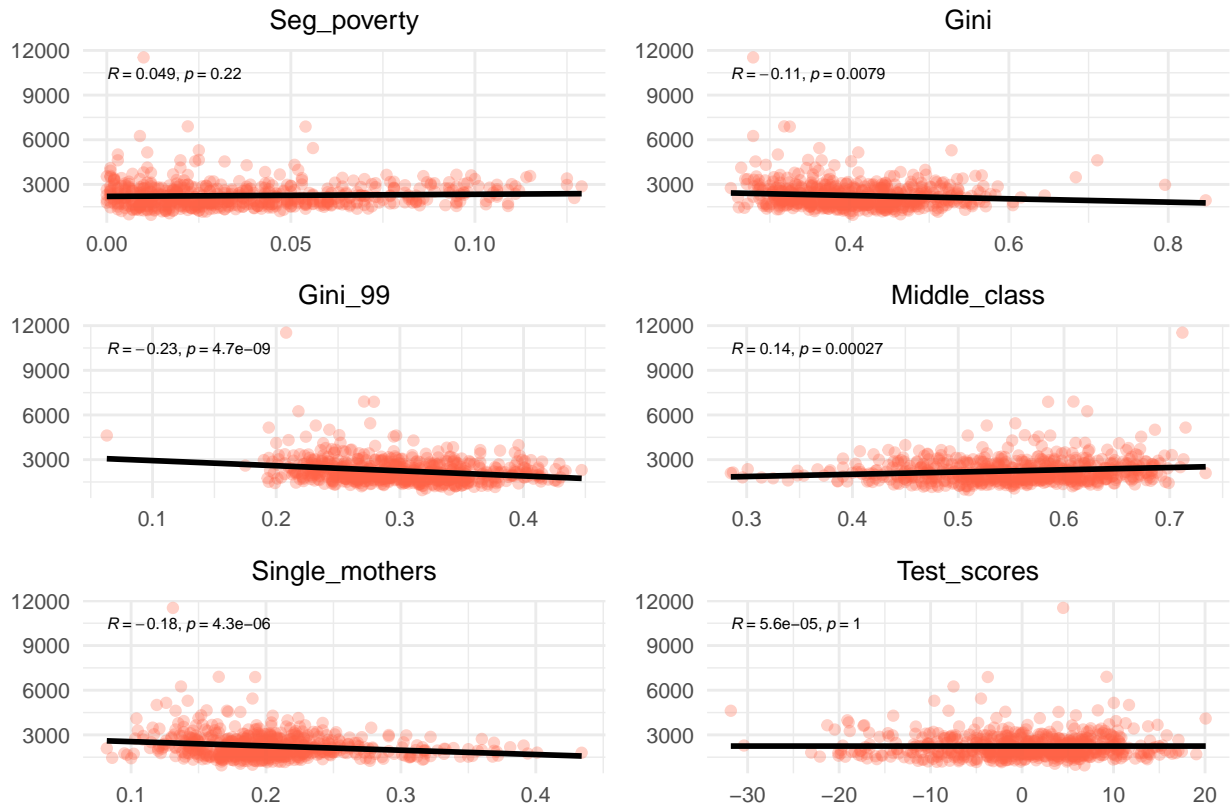
Appendix C

C.1 Initial linear modeling output

```
##
## Call:
## lm(formula = Mobility ~ ., data = mobility)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.072790 -0.014110 -0.001634  0.011117  0.155600
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1.408e-01  5.642e-02   2.496  0.01284 *
## Population                  1.265e-09  1.485e-09   0.852  0.39463
## Urban                       1.523e-03  3.434e-03   0.444  0.65753
## Black                       6.422e-02  2.302e-02   2.790  0.00544 **
## Seg_racial                 -4.778e-02  1.475e-02  -3.238  0.00127 **
## Seg_income                 -2.523e-01  7.164e-01  -0.352  0.72481
## Seg_poverty                -6.192e-02  3.826e-01  -0.162  0.87148
## Seg_affluence               2.447e-01  3.649e-01   0.671  0.50269
## Commute                     5.881e-02  2.005e-02   2.933  0.00348 **
## Income                      4.698e-07  4.603e-07   1.021  0.30787
## Gini                        2.227e+00  2.734e+00   0.814  0.41574
## Share01                    -2.247e-02  2.735e-02  -0.822  0.41148
## Gini_99                    -2.341e+00  2.731e+00  -0.857  0.39158
## Middle_class                1.117e-01  3.488e-02   3.204  0.00143 **
## Local_tax_rate              2.019e-01  1.670e-01   1.209  0.22714
## Local_gov_spending          8.749e-07  1.665e-06   0.525  0.59945
## Progressivity               5.321e-03  8.745e-04   6.085 2.08e-09 ***
## EITC                       -3.454e-04  3.526e-04  -0.980  0.32773
## School_spending             1.299e-03  1.509e-03   0.861  0.38941
## Student_teacher_ratio       4.133e-04  7.256e-04   0.570  0.56918
## Test_scores                -8.006e-05  2.294e-04  -0.349  0.72726
## Labor_force_participation  -3.130e-02  3.705e-02  -0.845  0.39854
## Manufacturing              -1.725e-01  2.094e-02  -8.237 1.11e-15 ***
## Chinese_imports            -8.960e-04  7.309e-04  -1.226  0.22072
## Teenage_labor              -1.819e+00  1.747e+00  -1.041  0.29819
## Migration_in               -5.001e-01  2.464e-01  -2.030  0.04281 *
## Migration_out               4.755e-02  2.990e-01   0.159  0.87369
## Foreign_born                1.930e-02  4.011e-02   0.481  0.63061
## Social_capital             -3.523e-03  1.985e-03  -1.775  0.07645 .
## Religious                   4.655e-02  9.886e-03   4.709 3.10e-06 ***
## Violent_crime              -4.199e+00  1.321e+00  -3.179  0.00156 **
## Single_mothers             -3.575e-01  6.966e-02  -5.132 3.89e-07 ***
## Divorced                   -1.735e-01  1.183e-01  -1.467  0.14282
## Married                     5.582e-04  5.074e-02   0.011  0.99123
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02658 on 599 degrees of freedom
## Multiple R-squared:  0.7201, Adjusted R-squared:  0.7047
## F-statistic:  46.7 on 33 and 599 DF,  p-value: < 2.2e-16
```

C.2 "High correlation" linear modeling output

```
##
## Call:
## lm(formula = Mobility ~ ., data = high_cor)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.068824 -0.015726 -0.002458  0.012251  0.157758
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.0569900  0.0285160   1.999 0.046097 *
## Black          -0.0457820  0.0137647  -3.326 0.000933 ***
## Seg_racial     -0.0897731  0.0128149  -7.005 6.46e-12 ***
## Commute         0.0110406  0.0148023   0.746 0.456027
## Gini           -0.0151867  0.0232378  -0.654 0.513655
## Middle_class    0.1899956  0.0283106   6.711 4.38e-11 ***
## Local_tax_rate  0.5629797  0.1562955   3.602 0.000341 ***
## Progressivity   0.0046341  0.0008336   5.559 4.03e-08 ***
## School_spending -0.0006344  0.0013005  -0.488 0.625834
## Test_scores    -0.0001958  0.0002067  -0.948 0.343753
## Manufacturing  -0.1687611  0.0184928  -9.126  < 2e-16 ***
## Migration_in   -0.3414667  0.1376914  -2.480 0.013406 *
## Religious       0.0480641  0.0093395   5.146 3.57e-07 ***
## Violent_crime  -2.5048198  1.3355098  -1.876 0.061188 .
## Divorced       -0.5288553  0.0870310  -6.077 2.15e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02827 on 618 degrees of freedom
## Multiple R-squared:  0.6732, Adjusted R-squared:  0.6658
## F-statistic: 90.94 on 14 and 618 DF,  p-value: < 2.2e-16
```

C.3 Refined linear modeling output

```
##
## Call:
## lm(formula = Mobility ~ Black + Seg_racial + Middle_class + Progressivity +
##     Manufacturing + Migration_in + Religious + Divorced + Local_tax_rate,
##     data = high_cor)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.066467 -0.016442 -0.002381  0.012039  0.161265
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.0417906  0.0183809   2.274 0.023330 *
## Black          -0.0496264  0.0128633  -3.858 0.000126 ***
## Seg_racial     -0.0932902  0.0118153  -7.896 1.31e-14 ***
## Middle_class    0.2054495  0.0203783  10.082  < 2e-16 ***
## Progressivity   0.0042545  0.0007718   5.512 5.19e-08 ***
## Manufacturing  -0.1745322  0.0160459 -10.877  < 2e-16 ***
## Migration_in   -0.3946780  0.1261807  -3.128 0.001843 **
```

```
## Religious        0.0515144  0.0088718   5.807 1.02e-08 ***
## Divorced        -0.5326058  0.0834021  -6.386 3.33e-10 ***
## Local_tax_rate  0.5195973  0.1362840   3.813 0.000151 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02829 on 623 degrees of freedom
## Multiple R-squared:  0.6702, Adjusted R-squared:  0.6654
## F-statistic: 140.6 on 9 and 623 DF,  p-value: < 2.2e-16
```

C.4 Final linear modeling output

```
##
## Call:
## lm(formula = Mobility ~ Black + Seg_racial + Middle_class + Progressivity +
##     Manufacturing + Migration_in + Religious + Divorced + Local_tax_rate,
##     data = high_cor_clean)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.069629 -0.015856 -0.002656  0.012126  0.167709
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.0480667  0.0172375   2.789  0.00546 **
## Black          -0.0549853  0.0120468  -4.564 6.05e-06 ***
## Seg_racial     -0.0985409  0.0111992  -8.799  < 2e-16 ***
## Middle_class    0.1890656  0.0191673   9.864  < 2e-16 ***
## Progressivity   0.0036112  0.0007258   4.975 8.45e-07 ***
## Manufacturing  -0.1581657  0.0151367 -10.449  < 2e-16 ***
## Migration_in   -0.3455706  0.1183115  -2.921  0.00362 **
## Religious       0.0479549  0.0083284   5.758 1.34e-08 ***
## Divorced       -0.5178954  0.0782790  -6.616 8.01e-11 ***
## Local_tax_rate  0.5675728  0.1278860   4.438 1.07e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02646 on 617 degrees of freedom
## Multiple R-squared:   0.68,  Adjusted R-squared:  0.6753
## F-statistic: 145.7 on 9 and 617 DF,  p-value: < 2.2e-16
```

C.5 Beta regression results

```
##
## Call:
## betareg(formula = Mobility ~ Black + Seg_racial + Middle_class + Progressivity +
##     Manufacturing + Migration_in + Religious + Divorced + Local_tax_rate,
##     data = high_cor)
##
## Quantile residuals:
##     Min      1Q  Median      3Q     Max
## -3.1124 -0.6053 -0.0501  0.5301  4.4049
##
## Coefficients (mean model with logit link):
##                  Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)    -2.80529     0.17635 -15.908  < 2e-16 ***
## Black          -1.26750     0.14661  -8.645  < 2e-16 ***
## Seg_racial     -0.90970     0.11543  -7.881 3.25e-15 ***
## Middle_class    1.93784     0.19057  10.169  < 2e-16 ***
## Progressivity   0.04289     0.00703   6.101 1.05e-09 ***
## Manufacturing  -1.66994     0.15308 -10.909  < 2e-16 ***
## Migration_in   -2.21954     1.20252  -1.846   0.0649 .
## Religious       0.50536     0.08251   6.125 9.09e-10 ***
## Divorced       -4.80191     0.78567  -6.112 9.85e-10 ***
## Local_tax_rate  4.68461     1.19788   3.911 9.20e-05 ***
##
## Phi coefficients (precision model with identity link):
##       Estimate Std. Error z value Pr(>|z|)
## (phi)  153.809      8.661   17.76   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Type of estimator: ML (maximum likelihood)
## Log-likelihood:  1497 on 11 Df
## Pseudo R-squared: 0.7531
## Number of iterations: 21 (BFGS) + 3 (Fisher scoring)
```