# Milestone 7

Maria Burzillo

3/28/2020

## Abstract

Trounstine (2016) suggests that high levels of residential segregation are associated with increased political polarization and decreased public spending. In this analysis, I was able to successfully replicate Trounstine (2016)'s main results. Adding onto her original analysis, I impute a large amount of missing data from the original dataset and re-run the analysis. Finally, I add in income segregation as a predictor in her regressions to assess whether this could be a confounding variable.

## Introduction

There is a large degree of variation in public goods spending across local governments. As a result, many scholars have worked to determine what factors may lead to the underprovision of public goods spending. Research in the past has associated racial diversity or changes in levels of diversity with the under-provision of public goods (Baqir, Easterly, and Alesina 1999; Hopkins 2009).However, Trounstine (2016) argues that it is racial segregation, not diversity in and of itself, that results in the under-provision of public goods. Trounstine's analysis consists of two main parts. First, she uses election and demographic data from 25 of America's largest cities between 1990 and 2010 to run a multilevel mixed-effects linear regression with fixed effects for region and year and with random effects for cities in order to show that polarization increases with segregation. Trounstine measures segregation with Theil's H index, which measures the degree to which the diversity of a neighborhood differs from the diversity of the entire city. The second main part of Trounstine's analysis looks at the ability of the Theil's H segregation index to explain a variety of types of public expenditures by city using a sample of 2,637 cities with 13,742 city-year observations. Using linear regressions with fixed effects for cities and robust standard errors clustered by city, Trounstine finds a significant, negative correlation between segregation and public goods spending that is robust to the inclusion of a variety of relevant controls and an alternative specification in which the number of waterways is used as an instrument for segregation.

In this analysis, I first work to replicate the main results of Trounstine (2016) using R statistical software (R Core Team 2019). The original data and Stata code made publically available by the author were downloaded via the Harvard Dataverse (Trounstine 2015). I also make all of my code and analysis available on Github.[1] I was successfully able to replicate the main results of Trounstine (2016) in R with the exception of some of the marginal analyses, which I was nevertheless able to replicate in Stata.

One concern with the original analysis in Trounstine (2016) is the large amount of missing data, which substantially constrains the sample size used in the regression analysis. For example, the regression analysis in the main specification using the the original racial polarization dataset exludes %0.55 of the observations, and the regression analysis in the main specification using the original financial segregation data excludes %0.95 of observations from the original dataset. As an extension of Trounstine (2016), I impute missing values in the original data using the mice package in R, which generates multivariate imputations using

[1] Link to my Github repository for this project.

chained equations (van Buuren and Groothuis-Oudshoorn 2011). Then, I use the multiply imputed datasets to re-estimate the original models, pooling the results to generate final pooled regression coefficients and parameters.

Comparing the results of the original regressions and those done with the imputed data yields similar big picture results in terms of the direction of the signs on the coefficients on the segregation indices. Like in Trounstine (2016), I find that segregation is positively associated with political polarization and negatively associated with spending on public goods. However, the magnitude of the effects in most specifications has diminished and most results become statistically insignificant. While these findings do not necessarily challenge the results of Trounstine (2016), they do call into question the relative importance of segregation in determining public goods spending and political polarization and suggest that the results of Trounstine (2016) may not be quite as robust as once thought.

## Literature Review

Despite some progress made towards racial equity in the U.S. on other fronts, residential racial segregation continues to be prevasive and deeply entrenched in society (Fischer et al. 2004; Oliver 2010; Massey 1993). Research suggests that this kind of segregation has political consequences, as political cleavages in segregated cities tend to have racial as well as spatial dimensions (Massey 1993). Neighborhoods are often important actors within local politics because local governments provide many functions that are allocational in nature and concern geographical space (Trounstine 2016). Thus, when neighborhoods are divided on racial lines as well as spacial lines, it is natural to expect higher degrees of racial polarization as a result.

Studying residential segregation is difficult because its effects tend to differ by geographic level. On the neighborhood level, the kind of geographic racial isolation brought on by residential segregation has been associated with racial intolerance, resentment, and competition between racial groups (Oliver 2010). Living within segregated neighborhoods has also been associated with holding negative stereotypes and perceptions about out groups (Eric Oliver and Wong 2003). As a result, homogeneous neighborhoods have been associated with increased racial tension and political polarization in comparison to integrated, diverse neighborhoods. However, at the city or metropolitan level, the opposite seems to be true: when considering larger geographic areas, diversity and integration are correlated with racial tension, competition, prejudice, lower levels of cooperation, and lower spending on public goods (Oliver 2010; Baqir, Easterly, and Alesina 1999; Hopkins 2009). While these differences in the expected effect of segregation on the geographic level may seem confusing at first, they make sense as they suggest that the most severely segregated areas are those that are diverse overall, but have many homogeneous neighborhoods. Thus, while people of different races co-exist within a highly segregated city, they live separately within their own neighborhoods, which creates an environment ripe for racial tension (Trounstine 2016). It is thus not simply the level of diversity or integration that matters for racial harmony and cooperation, but their patterns within a larger geographic framework (Trounstine 2016; Oliver 2010; Bharathi et al. 2018).

Political polarization along racial lines may lead to decreased public spending and goods provision because groups may have different preferences, which can make compromise hard, and because groups may preceive a disutility in out-groups receiving public goods expenditure Baqir, Easterly, and Alesina (1999). Einstein (2012) found evidence that racial segregation predicts large political divisions at the metropolitan level and that these divisions can create a lack of willingness to compromise and collaborate on local policy problems. Trounstine (2016) finds similar results at the city level: that residential racial segregation is associated with both increased political division and decreased public spending. Thus, these authors suggest that it is the combination of homogenous neighborhoods within a much larger, diverse geographic area that leads to increased political polarization and reduced public goods spending in local governments.

More recently, some scholars have called this hypothesis and its importance into question. For example, Lee (2018) finds evidence that larger inequalities within the political system favoring socially powerful groups, not local diversity patterns leading to decreased cooperation, may be a better explanation of failures in public goods provision in diverse areas. Others suggest that additional factors, such as income segregation, may be important confounding factors in public goods provision. An, Levy, and Hero (2018), for example, suggests that the more closely related income inequality is to racial inequality, the less investment is made in public goods, and that this interaciton was a better predictor of public goods spending patterns than measures of diversity. There is, in fact, a variety of evidence suggesting that it is meaningful to consider the effects of income inequaltiy and diversity and segregation jointly (An, Levy, and Hero 2018; Massey 1993). Given the active debate in the literature over the relaitonships between diversity, segregation, public spending, and other factors, it is increasingly important to re-examine previously reported findings as a means of robustness checks.

## Replication

I was able to successfully replicate all of the main results from Trounstine (2016). All regressions and tables were fully replicated in R. As an example, in table X, I replicate Table 1 from page 713 of Trounstine (2016), which is also included as Figure 1 for comparison. However, I was unable to successfully replicate Trounstine (2016)'s marginal effects analyses and margins plots using R. There does not yet appear to be a built-in R function to calculate marginal or predicted effects or to generate margins plots from the complicated multi-level models employed in the original paper, and creating such a function was outside of the scope of this analysis. Nevertheless, these results were successfully replicated in Stata.

**TABLE 1 Racial Polarization in Segregated Cities**

| | Racial Divide with Multigroup Segregation Index | | | Racial Divide with Two-Group Segregation Index | | | Racial Divide with Ideology Control | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\beta$ | SE | $P > \|t\|$ | $\beta$ | SE | $P > \|t\|$ | $\beta$ | SE | $P > \|t\|$ |
| Multigroup *H* Index | 0.932 | 0.39 | 0.02 | | | | | | |
| White/Nonwhite *H* Index | | | | 0.756 | 0.30 | 0.01 | 0.835 | 0.30 | 0.01 |
| Diversity | 0.385 | 0.36 | 0.29 | 0.518 | 0.32 | 0.11 | 0.584 | 0.32 | 0.07 |
| % Asian | –0.115 | 0.53 | 0.83 | 0.120 | 0.56 | 0.83 | –0.004 | 0.52 | 0.99 |
| % Black | –0.432 | 0.27 | 0.11 | –0.237 | 0.22 | 0.27 | –0.133 | 0.21 | 0.53 |
| % Latino | –0.191 | 0.26 | 0.46 | –0.059 | 0.25 | 0.82 | 0.095 | 0.28 | 0.73 |
| Median HH Income (1,000s) | –0.004 | 0.00 | 0.52 | –0.007 | 0.00 | 0.32 | –0.002 | 0.00 | 0.81 |
| % Renters | –0.580 | 0.42 | 0.17 | –0.806 | 0.43 | 0.06 | –0.419 | 0.45 | 0.36 |
| % College Degree | 0.328 | 0.71 | 0.65 | 0.723 | 0.73 | 0.32 | 0.123 | 0.87 | 0.89 |
| Biracial Contest | 0.210 | 0.04 | 0.00 | 0.208 | 0.04 | 0.00 | 0.192 | 0.04 | 0.00 |
| Nonpartisan Election | –0.090 | 0.07 | 0.18 | –0.089 | 0.07 | 0.18 | –0.034 | 0.06 | 0.60 |
| Primary Election | –0.092 | 0.03 | 0.00 | –0.09 | 0.03 | 0.01 | –0.071 | 0.03 | 0.02 |
| Population (logged) | 0.035 | 0.06 | 0.53 | 0.048 | 0.05 | 0.38 | –0.011 | 0.06 | 0.86 |
| White Ideology | | | | | | | –0.051 | 0.03 | 0.11 |
| Constant | –0.242 | 0.57 | 0.67 | –0.393 | 0.56 | 0.49 | 0.236 | 0.61 | 0.70 |
| Wald $\chi^2$ | 187.12 | | 0.00 | 189.68 | | 0.00 | 222.92 | | 0.00 |
| N | | 91 | | | 91 | | | 86 | |

*Note:* Multilevel mixed-effects linear regressions with fixed effects for region and year, and random effects for cities are presented.

Figure 1: This is the original table displaying the results of the models for the effects of residential segregation on political polarization from page 713 of Trounstine (2016). My replication of these results can be found in Table 1 in this paper.

There was one interesting outcome from my attempt to replicate the original Stata code in R. Due to the differences in R and Stata in dealing with missing values, the results of my first replication of analyses using

Table 1: **Racial Polarization in Segregated Cities**

| | *Dependent variable:* | | |
|---|---|---|---|
| | biggestsplit | | |
| | (1) | (2) | (3) |
| Multigroup H Index | 0.932** (0.394) | | |
| White/Nonwhite H Index | | 0.756** (0.297) | 0.835*** (0.296) |
| Diversity | 0.385 (0.362) | 0.518 (0.323) | 0.584* (0.323) |
| Percent Asian | −0.115 (0.527) | 0.120 (0.558) | −0.004 (0.522) |
| Percent Black | −0.432 (0.269) | −0.237 (0.216) | −0.133 (0.212) |
| Percent Latino | −0.191 (0.257) | −0.059 (0.254) | 0.095 (0.278) |
| Medain HH Income (1000s) | −0.004 (0.007) | −0.007 (0.007) | −0.002 (0.006) |
| Percent Renters | −0.580 (0.422) | −0.806* (0.431) | −0.419 (0.454) |
| Percent College Degree | 0.328 (0.711) | 0.723 (0.729) | 0.123 (0.869) |
| Biracial Contest | 0.210*** (0.037) | 0.208*** (0.037) | 0.192*** (0.036) |
| Nonpartisan Election | −0.090 (0.066) | −0.089 (0.066) | −0.034 (0.065) |
| Primary Election | −0.092*** (0.032) | −0.090*** (0.032) | −0.071** (0.030) |
| Population (logged) | 0.035 (0.055) | 0.048 (0.055) | −0.011 (0.061) |
| White Ideology | | | −0.051 (0.032) |
| Constant | −0.242 (0.569) | −0.393 (0.563) | 0.236 (0.605) |
| Wald Chi Squarred | a | b | c |
| Observations | 91 | 91 | 86 |
| Akaike Inf. Crit. | −55.548 | −56.381 | −64.922 |
| Bayesian Inf. Crit. | 32.332 | 31.499 | 20.981 |
| *Note:* | | | *p<0.1; **p<0.05; ***p<0.01 |
| | | | This table displays... |

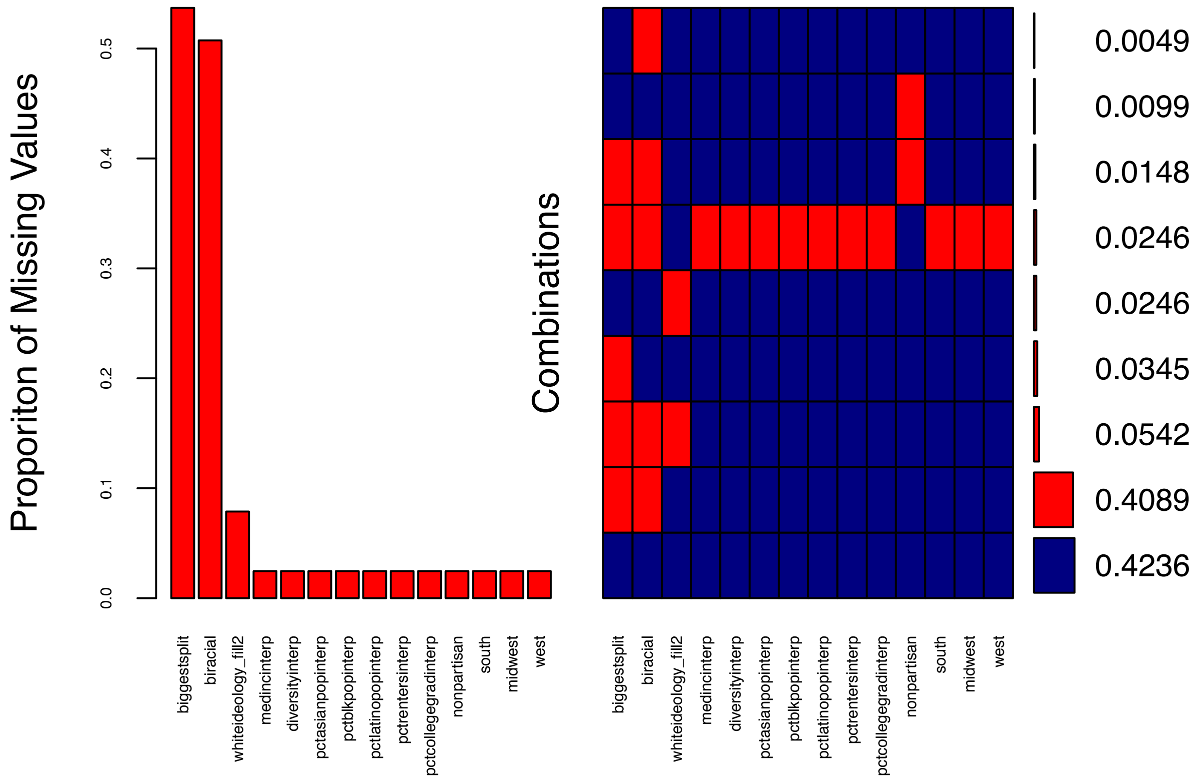
the financial segregation dataset were slightly different than the results in Stata from the original paper. In order to exclude cities from her analysis with only one census tract, Trounstine conditions her regressions in Stata such that the value for the number of census tracts is greater than one. In Stata, this does not remove missing values, whereas in R, it does. Since this variable is used only as a conditional filter and not as a regression variable (and thus, the observations with missing values for number of census tracts are not dropped), Trounstine's analysis includes 14 cities and 58 observations with missing values for census tracts in addition to cities with two or more census tracts. This is a potential oversight on the part of the author, and I would suggest also dropping observations with missing census tract data, or else imputing them given that the aim of her condition was to exclude cities without a sufficient number of census tracts. Nevertheless, dropping these values did not have much of an effect on the subsequent results, likely because they represent a small proportion of the overall sample.

# Extension

One concern with the original analysis in Trounstine (2016) is the large amount of missing data values in her original datasets. Because both R and Stata drop any observations with missing values for any of the variables used in a regression, this can exlcude a large portion of the data from the analysis and potentially bias the results if the data is not missing completely at random. A large amount of data is missing in both of the main datasets used in the analyses of Trounstine (2016). The main model specification using the racial polarization dataset exludes 112 of the original 203 observations, or %55.17 of the data. Additionally, the main model specification using the financial segregation data set excludes $2.88291 \times 10^5$ of the original 302033 observations, or %95.45 of observations from the original dataset.

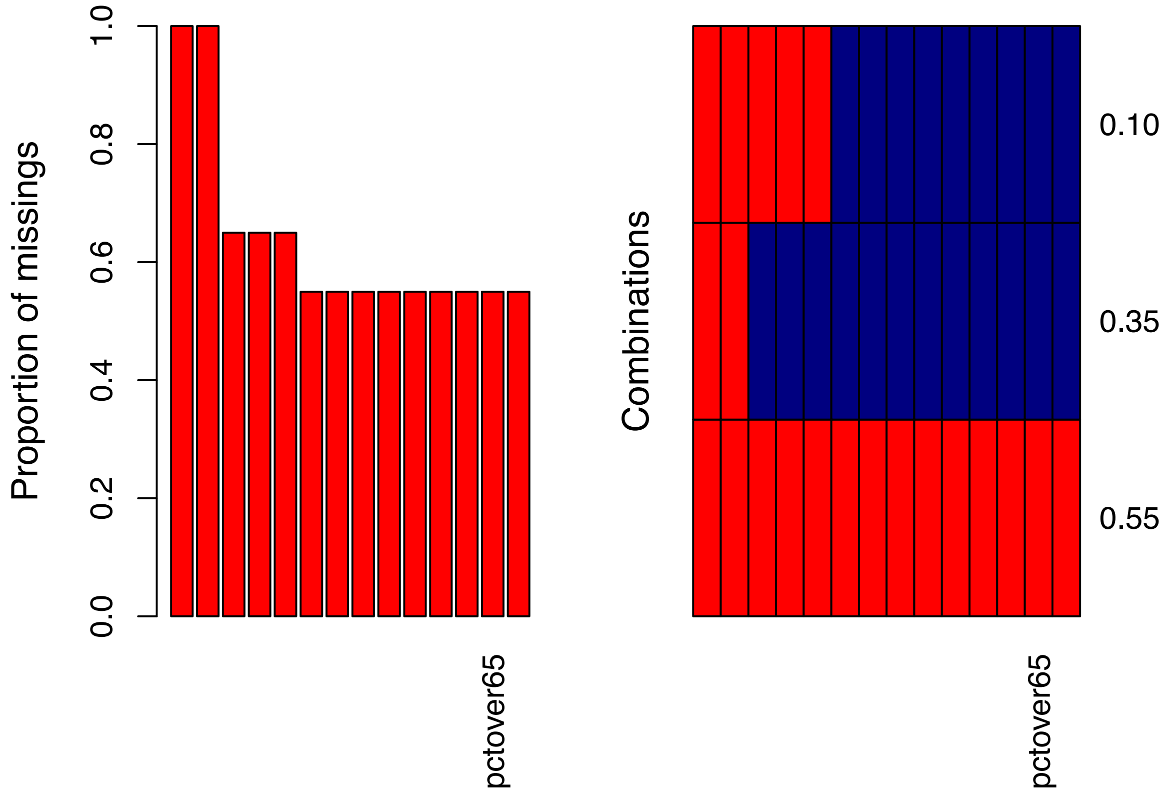
In an attempt to better deal with the problem of missing data in Trounstine (2016), I impute missing values in the original data using the mice package in R, which generates multivariate imputations using chained equations (van Buuren and Groothuis-Oudshoorn 2011). While there are a variety of different imputation methods that could have been employed, multiple imputation (such as the multiple multivariate imputations generated by mice) is desireable because instead of inputting a single value such as the mean for missing values, it instead uses the distribution of the available data to estimate multiple potential values for the missing data. As a result, multiple imputation helps to account for the uncertainty inherent in the imputation process and allows for the calculation of standard errors around estimators. As a result, multiple imputation allows the researcher to more accurately assess the of uncertainty in the analysis in general.

Before performing the multiple imputations on the datasets, I first examined the missing data for patterns. To better understand any potential patterns in missing data, I plotted the pattern of missingness and created a histogram showing the frequency of missing values for those variables with missing values in figure X1 for the racial polarization data and figure X2 for the financial segregation data. Looking at figure X1, the histogram shows that the variables for the largest vote split along racial lines and the variable indicating whether or not the election had candidates of more than one race had by far the largest percentages of missing data, missing %53.69 and %50.74, respectively. It is important to note that the variable for the largest vote split along racial lines is the dependent variable in the regresison analysis, and thus, we are missing a large percentage of this key variable. From the plot on the right, we can see that approximately 42% of observations are complete (have no missing values). There seem to be a correspondence between missing a value for the largest vote split and missing a value for the biracial election indicator. There also seem to be about 2% of values for wich most of the variables are missing. However, most observations are not missing more than 2-3 values.

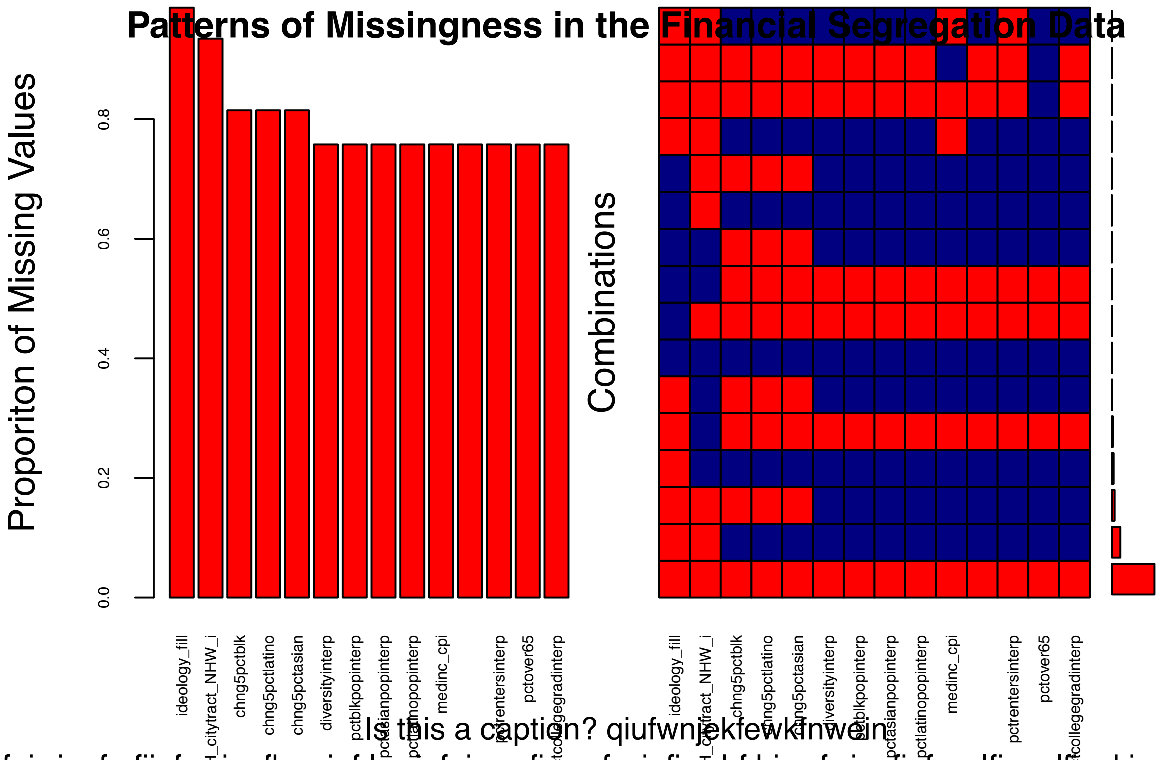



```
Variables sorted by number of missings:
             Variable      Count
         biggestsplit 0.53694581
             biracial 0.50738916
  whiteideology_fill2 0.07881773
         medincinterp 0.02463054
      diversityinterp 0.02463054
    pctasianpopinterp 0.02463054
      pctblkpopinterp 0.02463054
   pctlatinopopinterp 0.02463054
     pctrentersinterp 0.02463054
 pctcollegegradinterp 0.02463054
          nonpartisan 0.02463054
                south 0.02463054
              midwest 0.02463054
                 west 0.02463054
```

In figure X2, we can similarly observe the trends for the financial segregation dataset. In this dataset, there is a very high proportion of missing variables for a number of variables. For example, there are 282334 missing values for the segregation index, which is the main independent variable, or %93.48 of the data. In this dataset, there are also many more observations for which values are missing for multiple variables in comparison to the racial polarization dataset. In total, a mere %0.58 of observations are complete for all variables included in the main specification for the financial segregaion dataset.

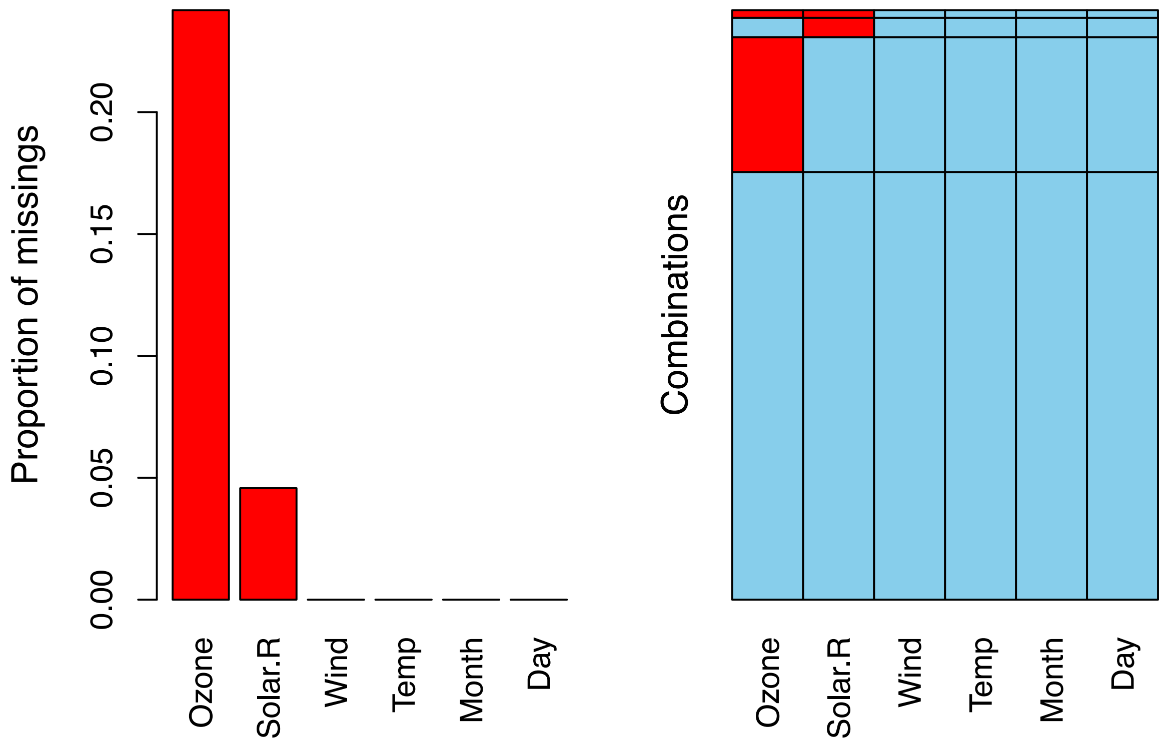



```
Variables sorted by number of missings:
              Variable Count
          ideology_fill  1.00
      H_citytract_NHW_i  1.00
            chng5pctblk  0.65
         chng5pctlatino  0.65
          chng5pctasian  0.65
        diversityinterp  0.55
        pctblkpopinterp  0.55
      pctasianpopinterp  0.55
     pctlatinopopinterp  0.55
             medinc_cpi  0.55
pctlocalgovworker_100  0.55
      pctrentersinterp  0.55
              pctover65  0.55
 pctcollegegradinterp  0.55
```




Is this a caption? qiufwnjekfewkfnwein
wfuiwjnef nfijefnwjenfkqwjnf lqwjefnjqwefjwnef winfjqwhf hjwefwiuefief welfiunqlfnqkjwe c

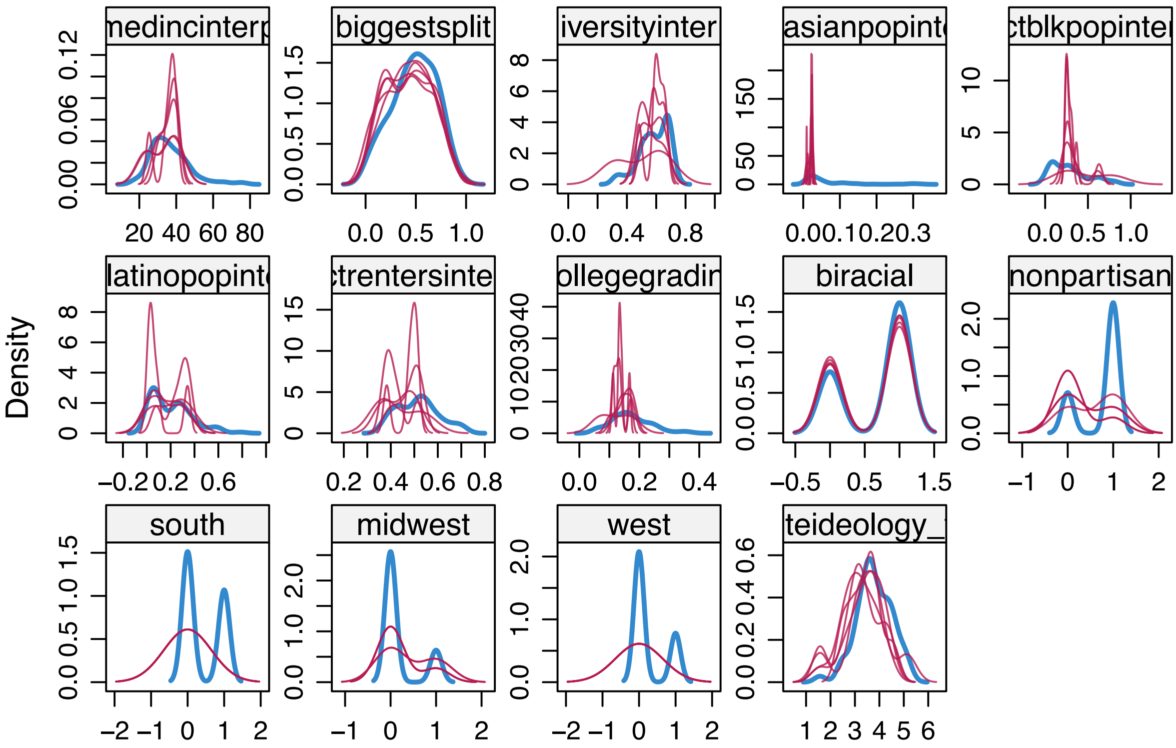
```r
aggr(airquality, bars = TRUE, numbers = TRUE, varheight = TRUE)
```



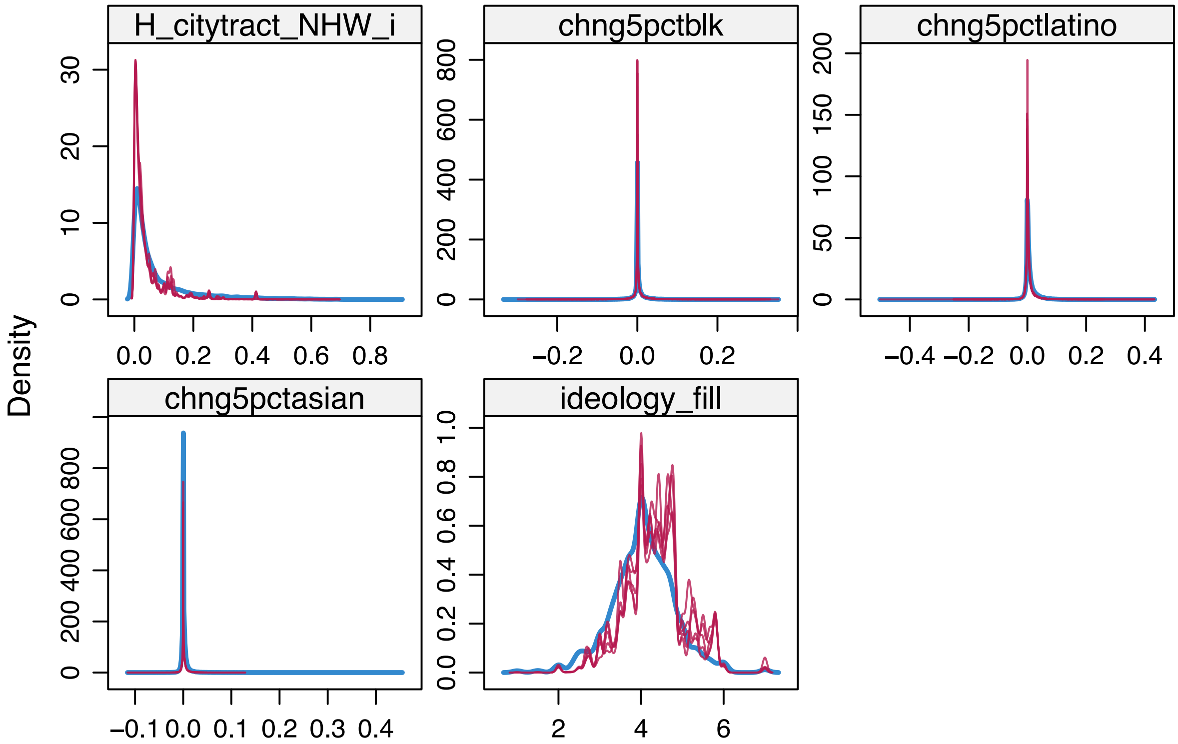

Because neither dataset seems to exhibit any clear patterns in the missing, data, it makes sense to proceed with the imputation. For the racial polarization dataset, I performed multiple imputations with 20 iterations using mice, while for the financial segregation dataset, I performed only 5 iterations due to the large size of the dataset and computing limitaitons. A non-stochastic imputation method, Classification and Regression Trees (CART), was used instead of the default for the imputation for both datasets because of an error with matrix inversion caused by the data that prevented the use of the default method, predictive mean matching. For the racial polarization dataset, I included all of the variables used in the analysis associated with the dataset and was able to impute all missing values. For the financial segregation dataset, however, I only imputed values for the main independent variable, the segregation index, and a few other variables, although all variables used in the analysis were included in the data subset input into the mice function. More of the data could not be imputed from the financial segregation dataset due to computing and time constraints for this project given the dataset's large size and high proportion of missing values.

### make sure to talk about the fact that still a large number of values missing for the imputed variables here

Before examining the results of Trounstine's model using the multiply imputed data, I first ran some diagnostic tests on the imputation results to make sure that everything ran as expected. First, I checked the convergence of the algorithm used within mice() by plotting the trace lines as a function of the number of iterations for each of the variables. Then, I visually inspected the distributions of the imputed data in comparison to the original data with density and strip plots. All of these checks suggested that the imputed values were within a plausible range of the data and that their distribution fit the underlying distribution of the data relatively well. The only cases in which there was some cause for concern were for some of the variables with very few missing values in the racial polarization dataset (such as the indicators for region). However, this is more or less to be expected given the small number of imputations performed in these cases. Thus, and especially because there are so few of these values missing in the actual datasets, this was not a major concern. See figures X and X1 to see the density plots of the imputed values overlayed on the density plots of the original variables. The rest of the plots and results of the diagnostic tests discussed here are presented in the Appendix.




```
Error in density.default(x = c(NA_real_, NA_real_, NA_real_, NA_real_, : need at least 2 points to sele
```




Given the promising results of the diagnostic checks, I next proceeded to re-estimate the original model using the multiply imputed datasets, pooling the results to produce final pooled regression coefficients and parameters. The results for the analyses using the racial polarization dataset are presented in tables 1-3.

The results of the model with the new dataset are slightly different in comparison to the original results from Trounstine (2016). With the imputed data, we now have a total of 203 observations in our model as compared to the original model, which had only 91 observations. Interestingly, while the sign of the coefficient on the main variable of interest, the Theil's H segregation index, is the same and the standard error has decreased slightly, the result has become statisically insignificant and the 95% confidence interval contains zero. Thus while these results still suggest that segregation may be associated with increased political polarization, they confer a lesser degree of certainty than Trounstine's original analysis. The coefficients for % Black population, % Latino population, and median household income have also switched signs, and all coefficients except the indicator for a biracial election are also statistically insignificant, as was the case in the original analysis. The coefficient on the indicator for a primary election, has also changed from being significant in the original analysis to insignificant here. In general, the standard errors on the coefficients have decreased slightly.

| | term | estimate | std.error | p.value |
|---|---|---|---|---|
| 1 | (Intercept) | 0.102790142 | 0.493934326 | 0.837614939 |
| 2 | H_citytract_multi_i | 0.453326497 | 0.348694304 | 0.205951039 |
| 3 | diversityinterp | 0.123967102 | 0.312377696 | 0.693743276 |
| 4 | pctasianpopinterp | -0.074241964 | 0.444583120 | 0.868590049 |
| 5 | pctblkpopinterp | 0.207899537 | 0.193167426 | 0.287790076 |
| 6 | pctlatinopopinterp | 0.163002438 | 0.204286528 | 0.432303470 |
| 7 | medincinterp | 0.001562794 | 0.004287399 | 0.717720249 |
| 8 | pctrentersinterp | -0.419230064 | 0.297441281 | 0.166170001 |
| 9 | pctcollegegradinterp | 0.307747279 | 0.611607314 | 0.622253286 |
| 10 | biracial | 0.188377352 | 0.038608047 | 0.000590573 |
| 11 | nonpartisan | -0.064507988 | 0.056677862 | 0.265434489 |
| 12 | primary | -0.048685964 | 0.027900379 | 0.093728871 |
| 13 | logpop | 0.009863395 | 0.042472137 | 0.819901297 |

```
\begin{table}[!htbp] \centering
  \caption{\textbf{Racial Polarization in Segregated Cities: Racial Divide with Multigroup Segregation I
  \label{}
\begin{tabular}{@{\extracolsep{5pt}}lccccccc}
\\[-1.8ex]\hline
\hline \\[-1.8ex]
Statistic & \multicolumn{1}{c}{N} & \multicolumn{1}{c}{Mean} & \multicolumn{1}{c}{St. Dev.} & \multicolu
\hline \\[-1.8ex]
estimate & 13 & 0.073 & 0.214 & $-$0.419 & $-$0.049 & 0.188 & 0.453 \\
std.error & 13 & 0.237 & 0.200 & 0.004 & 0.042 & 0.349 & 0.612 \\
p.value & 13 & 0.462 & 0.310 & 0.001 & 0.206 & 0.718 & 0.869 \\
\hline \\[-1.8ex]
\end{tabular}
\end{table}
```

In the second and third regressions, the results are similar. The main coefficient on segregation index has decreased in magnitude, although in these specifications, it remains statistically significant. The other coefficients except for the biracial indicator are insignificant, and the coefficients on percent Asian, Black, and Latino have all switched signs. Standard errors have also reduced slightly in these specifications compared to the original analysis in Trounstine (2016).

```
                   term    estimate   std.error      p.value
1           (Intercept)  0.06359131 0.485904355 0.8973585131
2      H_citytract_NHW_i  0.58781597 0.258004304 0.0270850500
3        diversityinterp  0.07233701 0.287492968 0.8027800741
4      pctasianpopinterp  0.13801742 0.456035952 0.7643173911
5        pctblkpopinterp  0.21298251 0.168036405 0.2117744468
6    pctlatinopopinterp  0.20511675 0.213275459 0.3485866272
7           medincinterp  0.00050681 0.004346694 0.9079867879
8      pctrentersinterp -0.51495197 0.300841418 0.0947111090
9  pctcollegegradinterp  0.46544162 0.625563151 0.4695679555
10              biracial  0.18726958 0.038984753 0.0007818527
11           nonpartisan -0.04836187 0.056358296 0.3986015846
12               primary -0.04750543 0.027579664 0.0977656454
13                logpop  0.01227514 0.041592434 0.7722869246
```

```
\begin{table}[!htbp] \centering
  \caption{\textbf{Racial Polarization in Segregated Cities: Racial Divide with Two-Group Segregation In
  \label{}
\begin{tabular}{@{\extracolsep{5pt}}lccccccc}
\\[-1.8ex]\hline
\hline \\[-1.8ex]
Statistic & \multicolumn{1}{c}{N} & \multicolumn{1}{c}{Mean} & \multicolumn{1}{c}{St. Dev.} & \multicolu
\hline \\[-1.8ex]
estimate & 13 & 0.103 & 0.266 & $-$0.515 & 0.001 & 0.205 & 0.588 \\
std.error & 13 & 0.228 & 0.200 & 0.004 & 0.042 & 0.301 & 0.626 \\
p.value & 13 & 0.446 & 0.346 & 0.001 & 0.098 & 0.772 & 0.908 \\
\hline \\[-1.8ex]
\end{tabular}
\end{table}
```

```
                   term    estimate   std.error     p.value
1           (Intercept)  0.03161226 0.494480802 0.949809318
2      H_citytract_NHW_i  0.56382756 0.257506957 0.033361645
```

```
3       diversityinterp  0.11339515 0.278801961 0.686073897
4     pctasianpopinterp  0.17314648 0.461168461 0.710306652
5       pctblkpopinterp  0.23765154 0.171764553 0.174915825
6    pctlatinopopinterp  0.19193817 0.209104301 0.369550944
7          medincinterp  0.00129454 0.004451916 0.773565336
8      pctrentersinterp -0.57173192 0.315168406 0.079913167
9  pctcollegegradinterp  0.30239740 0.625335770 0.635705666
10             biracial  0.18115919 0.040312506 0.001464491
11          nonpartisan -0.04181535 0.054897786 0.451931475
12              primary -0.04646529 0.027383436 0.102509685
13               logpop  0.02233954 0.044099579 0.621581568
```

```
\begin{table}[!htbp] \centering
  \caption{\textbf{Racial Polarization in Segregated Cities: Racial Divide with Ideology Control}}
  \label{}
\begin{tabular}{@{\extracolsep{5pt}}lccccccc}
\\[-1.8ex]\hline
\hline \\[-1.8ex]
Statistic & \multicolumn{1}{c}{N} & \multicolumn{1}{c}{Mean} & \multicolumn{1}{c}{St. Dev.} & \multicol
\hline \\[-1.8ex]
estimate & 13 & 0.089 & 0.259 & $-$0.572 & 0.001 & 0.192 & 0.564 \\
std.error & 13 & 0.230 & 0.201 & 0.004 & 0.044 & 0.315 & 0.625 \\
p.value & 13 & 0.430 & 0.323 & 0.001 & 0.103 & 0.686 & 0.950 \\
\hline \\[-1.8ex]
\end{tabular}
\end{table}
```

For the financial segregation dataset, we similarly fit each of our 5 imputed datasets to the original models from Trounstine (2016) using this data and then pool the results for each. The pooled results are presented in tables 4-6.

## FS data results

```r
# Main Analysis 2: Imputations


## regression 1 Table 2

# fit multiple imputed datasets

fit_imp_felm1 <- with(imp_fs, felm(dgepercap_cpi ~ H_citytract_NHW_i + diversityinterp + pctblkpopinter

# pool the analyses

pool_imp_felm1 <- pool(fit_imp_felm1)
imp_felm1_sum <- summary(pool_imp_felm1)


## regression 2 Table 2

# fit multiple imputed datasets
```

```r
fit_imp_felm2 <- with(imp_fs, felm(dgepercap_cpi ~ H_citytract_NHW_i + pctblkpopinterp + pctasianpopint

# pool the analyses

pool_imp_felm2 <- pool(fit_imp_felm2)
imp_felm2_sum <- summary(pool_imp_felm2)



## Regression 3 Table 2

fit_imp_felm3 <- with(imp_fs, felm(dgepercap_cpi ~ H_citytract_NHW_i + diversityinterp + pctblkpopinter

# pool the analyses

pool_imp_felm3 <- pool(fit_imp_felm3)
imp_felm3_sum <- summary(pool_imp_felm3)
```

```r
# how many observations were included in those 3 analyses?

# basic filter for table for all regressions

# xtreg dgepercap_cpi H_citytract_NHW_i diversityinterp pctblkpopinterp
# pctasianpopinterp pctlatinopopinterp medinc_cpi pctlocalgovworker_100
# pctrentersinterp pctover65 pctcollegegradinterp logpop if totaltracts>1 &
# dgepercap_cpi~=0,fe vce(cluster geo_id2)

fin <- complete(imp_fs) %>%
  filter(!(is.na(dgepercap_cpi)), !(is.na(H_citytract_NHW_i)), !(is.na(diversityinterp)),
         !(is.na(pctblkpopinterp)), !(is.na(pctasianpopinterp)), !(is.na(pctlatinopopinterp)),
         !(is.na(medinc_cpi)), !(is.na(pctlocalgovworker_100)), !(is.na(pctrentersinterp)),
         !(is.na(pctover65)), !(is.na(pctcollegegradinterp)), !(is.na(logpop)))

# create dge variable used in regression

fin_dge <-fin %>% filter(dgepercap_cpi != 0)

fin_dge_tab <- tibble(
  Variable = "Direct General Expenditure per Capita",
  Obs = nrow(fin_dge),
  Mean = mean(fin_dge$dgepercap_cpi, na.rm = T),
  SD = sd(fin_dge$dgepercap_cpi, na.rm = T),
  Min = min(fin_dge$dgepercap_cpi, na.rm = T),
  Max = max(fin_dge$dgepercap_cpi, na.rm = T)
)

# now we have 73,119 observations compared to 13,742. Calculation:

nrow(fin_dge)
```

```
[1] 73119
```

```r
# number of missing seg indexes before imputation: 282334
```

```r
# number missing now: 222957 -> Calculation:

new_missing <- sum(is.na(complete(imp_fs)$H_citytract_NHW_i))

# calculate number of observations we added to the sample: 59,377

num_added <- 282334 - new_missing

# INTERPRETATION BEFORE AFTER CHANGED The results of the first regression present a stark difference to
```

The results of the first regression present a stark difference to the results of Trounstine (2016). Crucially, the effect size of the segregation index on public spending has essentially gone to zero and has become statistically insignificant. Interestingly, the coefficient on the diversity variable has increased substantially from .106 to .759 and has become statistically significant. The sign of the coefficient on percent Asian popultion has switched signs from negative to positive; however, it remains statistically insignificant.

The results of the first regression indicate that Trounstine's results are robust to the inclusion of the imputed data, which led to the inclusion of an additional 59,377 observations in the analysis. The coefficients, standard errors, and significance levels are essentially unchanged. The same is largely true for the second and third specifications, but the magnitude of the coefficient on the segregation decreases slightly.

```
                    term     estimate   std.error      p.value
1      H_citytract_NHW_i -0.010918705 0.426554875 9.803087e-01
2        diversityinterp  0.759651272 0.294331931 9.861985e-03
3        pctblkpopinterp  0.472621835 0.349769823 1.766383e-01
4      pctasianpopinterp  0.968895301 1.263478353 4.431825e-01
5     pctlatinopopinterp  1.574170460 0.439849234 3.460155e-04
6             medinc_cpi  0.005639282 0.006681596 3.986797e-01
7  pctlocalgovworker_100  0.031531454 0.035515698 3.746512e-01
8      pctrentersinterp  0.094340566 0.336314649 7.790880e-01
9              pctover65  0.614973630 0.774697667 4.273087e-01
10  pctcollegegradinterp  3.128726475 0.786162431 6.926933e-05
11                logpop -0.588480024 0.146933940 6.225702e-05


\begin{table}[!htbp] \centering
  \caption{\textbf{Effect of Segregation on Overall per Capita City Expenditures: Direct General Expendi
  \label{}
\begin{tabular}{@{\extracolsep{5pt}}lccccccc}
\\[-1.8ex]\hline
\hline \\[-1.8ex]
Statistic & \multicolumn{1}{c}{N} & \multicolumn{1}{c}{Mean} & \multicolumn{1}{c}{St. Dev.} & \multicolu
\hline \\[-1.8ex]
estimate & 11 & 0.641 & 1.011 & $-$0.588 & 0.019 & 0.864 & 3.129 \\
std.error & 11 & 0.442 & 0.372 & 0.007 & 0.221 & 0.607 & 1.263 \\
p.value & 11 & 0.326 & 0.333 & 0.0001 & 0.005 & 0.435 & 0.980 \\
\hline \\[-1.8ex]
\end{tabular}
\end{table}

                    term     estimate   std.error      p.value
1      H_citytract_NHW_i  0.012718453 0.427763524 9.771293e-01
2        pctblkpopinterp  1.104995604 0.360504058 2.317801e-03
3      pctasianpopinterp  1.306800498 0.851137418 1.248055e-01
4     pctlatinopopinterp  2.338971573 0.480813839 1.220193e-06
5            chng5pctblk -0.713920215 0.769426147 3.563561e-01
```

```
6         chng5pctlatino -0.397370293 0.885289696 6.545084e-01
7          chng5pctasian  1.905116396 2.173524919 3.812072e-01
8             medinc_cpi  0.005878473 0.006694755 3.799170e-01
9  pctlocalgovworker_100  0.030699929 0.035453313 3.865429e-01
10      pctrentersinterp  0.135631862 0.333205088 6.839751e-01
11             pctover65  0.544248802 0.773949667 4.819346e-01
12  pctcollegegradinterp  3.422216653 0.822109404 3.159807e-05
13                logpop -0.580005381 0.146467758 7.526216e-05

\begin{table}[!htbp] \centering
  \caption{\textbf{Effect of Segregation on Overall per Capita City Expenditures: Direct General Exp. wi
  \label{}
\begin{tabular}{@{\extracolsep{5pt}}lccccccc}
\\[-1.8ex]\hline
\hline \\[-1.8ex]
Statistic & \multicolumn{1}{c}{N} & \multicolumn{1}{c}{Mean} & \multicolumn{1}{c}{St. Dev.} & \multicolu
\hline \\[-1.8ex]
estimate & 11 & 0.641 & 1.011 & $-$0.588 & 0.019 & 0.864 & 3.129 \\
std.error & 11 & 0.442 & 0.372 & 0.007 & 0.221 & 0.607 & 1.263 \\
p.value & 11 & 0.326 & 0.333 & 0.0001 & 0.005 & 0.435 & 0.980 \\
\hline \\[-1.8ex]
\end{tabular}
\end{table}
                    term     estimate   std.error      p.value
1      H_citytract_NHW_i -0.004676988 0.422650975 9.914818e-01
2        diversityinterp  0.762327416 0.294473572 9.640194e-03
3        pctblkpopinterp  0.482537267 0.349539756 1.674529e-01
4      pctasianpopinterp  0.944675103 1.259673829 4.533032e-01
5     pctlatinopopinterp  1.572055130 0.439073048 3.440240e-04
6             medinc_cpi  0.005683481 0.006684777 3.952186e-01
7  pctlocalgovworker_100  0.031133753 0.035577460 3.815328e-01
8       pctrentersinterp  0.071059876 0.340584575 8.347342e-01
9              pctover65  0.615130696 0.775647776 4.277572e-01
10  pctcollegegradinterp  3.130303577 0.786094452 6.858870e-05
11                logpop -0.589199901 0.147150404 6.252054e-05
12         ideology_fill -0.024847163 0.023399481 3.035981e-01

\begin{table}[!htbp] \centering
  \caption{\textbf{Effect of Segregation on Overall per Capita City Expenditures: Direct General Exp. wi
  \label{}
\begin{tabular}{@{\extracolsep{5pt}}lccccccc}
\\[-1.8ex]\hline
\hline \\[-1.8ex]
Statistic & \multicolumn{1}{c}{N} & \multicolumn{1}{c}{Mean} & \multicolumn{1}{c}{St. Dev.} & \multicolu
\hline \\[-1.8ex]
estimate & 11 & 0.641 & 1.011 & $-$0.588 & 0.019 & 0.864 & 3.129 \\
std.error & 11 & 0.442 & 0.372 & 0.007 & 0.221 & 0.607 & 1.263 \\
p.value & 11 & 0.326 & 0.333 & 0.0001 & 0.005 & 0.435 & 0.980 \\
\hline \\[-1.8ex]
\end{tabular}
\end{table}
```

# Conclusion

Why cities differ in their amount of spending on public goods is an interesting questions that is frequently debated in the literature. Trounstine (2016) posits that segregation may be an important underlying cause of the underprovsion of public goods. Finding evidence that residential segregation by race increases political partisanship, Trounstine (2016) argues that residential segregation can increase division and make cooperation amongst competing groups difficult, which results in decreased spending on public goods. The findings of Trounstine (2016) are particularly important because they suggest that segregation, as defined by many homogeneous neighborhoods within a larger diverse geographical area, is the key factor in disinvestment in public goods. This contradicts previous research which has suggested that levels of diversity were instead most important (see for example Baqir, Easterly, and Alesina (1999) or Hopkins (2009)). While Trounstine (2016) still represents relatively new reasearch as of 2020, it has yet to be fully vetted by additional research. Furthermore, some research that has been published on the topic since its publication suggests that other factors such as larger inequities in the political system favoring socially powerful groups (Lee 2018) or the interaction of income and racial inequality (An, Levy, and Hero 2018) may be more important in explaining public goods provisions.

I first worked to provide a check on the analysis presented in Trounstine (2016) by attempting to replicate the work in R. I was able to successfully replicate all of the main results from Trounstine (2016), with the exception of a few of a few marginal analyses of the complex, multi-level models which, to my knowledge, are currently not supported by available R packages. I was, however, able to successfully replicate these results in Stata.

As an additional robustness check I also tested the models in Trounstine (2016) by re-running them with additional data that I imputed from missing values in the original datasets. The large number of missing values in both of the datasets used in Trounstine (2016)'s main analysis created a concern for potential bias in the original results and a lack of representativeness due to the large amount of data excluded from the regression analysis due to missing values. Thus, I used the mice package in R to create multiply imputed datasets, upon which I re-ran the models and pooled the results to generate the final model coefficients and parameters.

The results of the data imputation exercise differed between the two datasets. For the analyses using the smaller, racial polarization dataset, the results from the imputed data resulted in a smaller magnitude of the coefficient on the main independent variable, and also made it statistically insignificant in the main model specification. This suggests that the segregation index may not have as strong of a positive association with political polarization as Trounstine (2016) suggests. However, for the analyses using the much larger financial segregation dataset, the results of the models using the multiply imputed data largely mirrored the results of Trounstine (2016), and the coefficients and significance levels were essentially the same between the two, with the standard errors on the model estimates being generally slightly lower for the analyses using the imputed data. This provides strong support for Trounstine (2016)'s finding that residential racial segregation is associated with diminshed spending on public goods.

The replication exercise and robustness check using multiply imputed data here provide strong evidence that residential segregation along racial lines does in fact correlate negatively with spending on a variety of public goods. What my analysis calls more into question, is by what mechanism this happens. Using the multiply imputed data, I still found some evidence that segregation may be associated with increased political polarization which Trounstine (2016) suggests may lead to diminished public goods spending; however, the effect size is much smaller than that found in Trounstine (2016), and some of the results have become insignificant. It is possible that other mechanisms may be more important in explaining why segregation might be associated with decreased spending on public goods. For example, it could be that there is an important interaction effect between income and segregation, as others have previously suggested, or there could be some other intervening mechanism or a confounding variable that is leading to the negative association between segregation and lower spending on public goods. Trounstine (2016) has identified an important relationship between segregation and the provision of public goods. However, my analysis suggests that further research is needed to assess the potential mechanisms by which residential segregation by race or even a confounding variable may result in dimished spending. If this mechanism can be better understood, than it will be possible to more effectively understand and address the underprovision of public goods, which may inhibit social mobility and unfairly disadvantage certain groups.

# Appendix

## Table 2

## Table 3

## Table 3

```r
# once I figure out how to make this sideways, I will combine a and b into this
# table:

table3c <- stargazer(felm2_1, felm2_2, felm2_3, felm2_4, felm2_5, felm2_6,
          omit = c("year", "south", "midwest", "west"),
          header = FALSE,
          title = "\\textbf{Effect of Segregation on Public Goods}",
          float.env = "sidewaystable")
```

## Main Analysis 4

## Main Analysis 5

## Table 5

# Appendix

**TABLE A2 Cities Included in Racial Polarization Data**

Replication

| | **Segregation: Mean *H* Index** | | **Largest Racial Divide, Number of Elections** | | |
|---|---|---|---|---|---|
| City Name | Multigroup | Two-Group | Black/White | Latino/White | Black/Latino |
| Austin, TX | 0.204 | 0.208 | 1 | 0 | 0 |
| Baltimore, MD | 0.510 | 0.516 | 3 | 1 | 0 |
| Charlotte, NC | 0.269 | 0.287 | 2 | 0 | 0 |
| Chicago, IL | 0.572 | 0.460 | 7 | 0 | 1 |
| Cleveland, OH | 0.558 | 0.531 | 2 | 0 | 0 |
| Columbus, OH | 0.316 | 0.284 | 3 | 0 | 1 |
| Dallas, TX | 0.359 | 0.339 | 4 | 0 | 1 |
| Denver, CO | 0.289 | 0.254 | 1 | 2 | 0 |
| Detroit, MI | 0.398 | 0.255 | 1 | 0 | 1 |
| Houston, TX | 0.339 | 0.308 | 7 | 0 | 2 |
| Indianapolis, IN | 0.292 | 0.293 | 0 | 0 | 1 |
| Jacksonville, FL | 0.233 | 0.222 | 2 | 0 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| Los Angeles, CA | 0.351 | 0.366 | 3 | 0 | 5 |
| Memphis, TN | 0.470 | 0.474 | 2 | 0 | 0 |
| Milwaukee, WI | 0.423 | 0.360 | 3 | 0 | 0 |
| New York, NY | 0.468 | 0.474 | 5 | 3 | 1 |
| Oklahoma, OK | 0.231 | 0.165 | 1 | 0 | 0 |
| Philadelphia, PA | 0.492 | 0.487 | 5 | 0 | 0 |
| Phoenix, AZ | 0.255 | 0.270 | 0 | 1 | 0 |
| San Antonio, TX | 0.237 | 0.225 | 0 | 4 | 0 |
| San Diego, CA | 0.255 | 0.266 | 3 | 0 | 1 |
| San Francisco, CA | 0.223 | 0.161 | 3 | 0 | 1 |
| San Jose, CA | 0.186 | 0.198 | 0 | 2 | 1 |
| Tucson, AZ | 0.185 | 0.192 | 1 | 0 | 0 |
| Washington, DC | 0.464 | 0.491 | 3 | 0 | 0 |

**TABLE A3 Summary Statistics: Census of Government Finance and Population**

| Variable | **Obs** | **Mean** | **SD** | **Min** | **Max** |
|---|---|---|---|---|---|
| Direct General Expenditure per Capita | 13742 | 1.186 | 1.220 | 0.019 | 70.457 |
| Highways per Capita | 13603 | 0.081 | 0.053 | 0.000 | 1.106 |
| Parks per Capita | 12905 | 0.061 | 0.061 | 0.000 | 1.111 |
| Police per Capita | 13626 | 0.181 | 0.094 | 0.000 | 1.546 |
| Sewers per Capita | 11223 | 0.092 | 0.077 | 0.000 | 1.591 |
| Welfare, Health, and Housing per Capita | 10871 | 0.057 | 0.131 | 0.000 | 4.984 |
| Own Source Revenue per Capita | 13741 | 0.942 | 1.118 | 0.021 | 76.123 |
| Two-Group H Index | 13742 | 0.076 | 0.099 | 0.000 | 0.767 |
| Diversity | 13742 | 0.309 | 0.188 | 0.007 | 0.772 |
| % Black | 13742 | 0.097 | 0.151 | 0.000 | 0.980 |
| % Asian | 13742 | 0.032 | 0.054 | 0.000 | 0.674 |
| % Latino | 13742 | 0.104 | 0.161 | 0.000 | 0.987 |
| 5Y Change, % Black | 11194 | 0.007 | 0.019 | −0.101 | 0.229 |
| 5Y Change, % Latino | 11194 | 0.016 | 0.020 | −0.171 | 0.207 |
| 5Y Change, % Asian | 11194 | 0.005 | 0.011 | −0.056 | 0.128 |
| Median Income | 13742 | 54,520.132 | 22,081.359 | 15,642.802 | 240,938.047 |
| % Local Gov. Employees | 13742 | 3.359 | 0.951 | 0.677 | 8.365 |
| % Renters | 13742 | 0.360 | 0.140 | 0.014 | 0.871 |
| % Over 65 | 13742 | 0.125 | 0.050 | 0.012 | 0.771 |
| % College Degree | 13742 | 0.160 | 0.099 | 0.003 | 0.587 |
| Population (logged) | 13742 | 10.132 | 1.016 | 6.071 | 15.921 |
| City Ideology | 2130 | 4.023 | 0.780 | 1.000 | 7.000 |
| Population | 13742 | 53,723.022 | 208,143.791 | 433.000 | 8,214,426.000 |

# Bibliography

An, Brian, Morris Levy, and Rodney Hero. 2018. "It's Not Just Welfare: Racial Inequality and the Local Provision of Public Goods in the United States." *Urban Affairs Review* 54 (5): 833–65. https://doi.org/10.1177/1078087417752476.

Baqir, Reza, William Easterly, and Alberto Alesina. 1999. "Public Goods and Ethnic Divisions" 114 (4): 1243–84. https://doi.org/10.1162/003355399556269.

Bharathi, N. Manjula, Deepak V. Malghan, Sumit Mishra, and Andaleeb Rahman. 2018. "Public Goods, and

Table 2: **Effect of Segregation on Overall per Capita City Expenditures**

| | *Dependent variable:* | | |
|---|---|---|---|
| | dgepercap_cpi | | |
| | (1) | (2) | (3) |
| H_citytract_NHW_i | −1.153*** | −1.011*** | −1.733*** |
| | (0.221) | (0.254) | (0.437) |
| diversityinterp | 0.106 | | −0.063 |
| | (0.134) | | (0.246) |
| pctblkpopinterp | 0.681*** | 0.741*** | 0.164 |
| | (0.167) | (0.161) | (0.523) |
| pctasianpopinterp | −0.385 | −0.852** | 0.197 |
| | (0.302) | (0.348) | (0.706) |
| pctlatinopopinterp | 1.543*** | 1.577*** | 1.622*** |
| | (0.186) | (0.205) | (0.390) |
| chng5pctblk | | −1.778*** | |
| | | (0.644) | |
| chng5pctlatino | | −2.055** | |
| | | (0.823) | |
| chng5pctasian | | −0.800 | |
| | | (1.093) | |
| medinc_cpi | 0.002* | 0.001 | 0.004 |
| | (0.001) | (0.002) | (0.003) |
| pctlocalgovworker_100 | 0.014 | 0.006 | −0.030 |
| | (0.016) | (0.018) | (0.046) |
| pctrentersinterp | 0.527 | 0.547 | 0.336 |
| | (0.333) | (0.385) | (0.656) |
| pctover65 | 0.093 | 0.487 | −0.865 |
| | (0.643) | (0.451) | (0.816) |
| pctcollegegradinterp | 5.395*** | 6.260*** | 6.527*** |
| | (0.403) | (0.419) | (1.029) |
| logpop | −0.243*** | −0.290*** | −0.447*** |
| | (0.044) | (0.068) | (0.088) |
| ideology_fill | | | −0.012 |
| | | | (0.034) |
| Observations | 13,742 | 11,194 | 2,130 |
| $R^2$ | 0.863 | 0.897 | 0.882 |
| Adjusted $R^2$ | 0.830 | 0.865 | 0.855 |
| Residual Std. Error | 0.503 (df = 11094) | 0.465 (df = 8544) | 0.405 (df = 1741) |
| *Note:* | | | *p<0.1; **p<0.05; ***p<0.01 |

Table 3: **Effect of Segregation on Public Goods A**

| | *Dependent variable:* | | |
|---|---|---|---|
| | highwayspercapNC_cpi | policepercapNC_cpi | parkspercapNC_cpi |
| | (1) | (2) | (3) |
| H_citytract_NHW_i | −0.039** | −0.215*** | −0.046*** |
| | (0.016) | (0.023) | (0.018) |
| diversityinterp | 0.005 | 0.059*** | 0.001 |
| | (0.010) | (0.013) | (0.013) |
| pctblkpopinterp | 0.052*** | 0.142*** | 0.031* |
| | (0.014) | (0.018) | (0.018) |
| pctasianpopinterp | −0.036 | −0.055 | −0.067*** |
| | (0.026) | (0.035) | (0.023) |
| pctlatinopopinterp | 0.025* | 0.335*** | 0.049*** |
| | (0.014) | (0.019) | (0.014) |
| medinc_cpi | 0.0003** | 0.00004 | −0.00002 |
| | (0.0001) | (0.0001) | (0.0001) |
| pctlocalgovworker_100 | −0.0003 | −0.001 | 0.001 |
| | (0.001) | (0.002) | (0.001) |
| pctrentersinterp | 0.011 | 0.075*** | 0.018 |
| | (0.023) | (0.028) | (0.021) |
| pctover65 | 0.140*** | 0.147*** | 0.127*** |
| | (0.032) | (0.045) | (0.040) |
| pctcollegegradinterp | 0.218*** | 0.793*** | 0.444*** |
| | (0.026) | (0.038) | (0.038) |
| logpop | −0.015*** | −0.054*** | −0.005* |
| | (0.004) | (0.004) | (0.003) |
| Observations | 13,603 | 13,626 | 12,905 |
| $R^2$ | 0.571 | 0.837 | 0.750 |
| Adjusted $R^2$ | 0.467 | 0.798 | 0.688 |
| Residual Std. Error | 0.039 (df = 10958) | 0.042 (df = 10991) | 0.034 (df = 10321) |
| *Note:* | | | *p<0.1; **p<0.05; ***p<0.01 |

Table 4: **Effect of Segregation on Public Goods B**

| | *Dependent variable:* | | |
|---|---|---|---|
| | sewerspercapNC_cpi | welfhoushealthNC_cpi | genrevownpercap_cpi |
| | (1) | (2) | (3) |
| H_citytract_NHW_i | −0.148*** | −0.138*** | −0.768*** |
| | (0.022) | (0.049) | (0.155) |
| diversityinterp | 0.039*** | −0.033 | 0.091 |
| | (0.015) | (0.025) | (0.085) |
| pctblkpopinterp | 0.012 | 0.016 | 0.272** |
| | (0.017) | (0.056) | (0.120) |
| pctasianpopinterp | −0.124*** | 0.130 | −0.147 |
| | (0.044) | (0.090) | (0.233) |
| pctlatinopopinterp | 0.091*** | 0.140*** | 1.202*** |
| | (0.019) | (0.028) | (0.120) |
| medinc_cpi | 0.001*** | −0.0003 | 0.004*** |
| | (0.0002) | (0.0003) | (0.001) |
| pctlocalgovworker_100 | −0.004* | −0.007** | 0.002 |
| | (0.002) | (0.003) | (0.013) |
| pctrentersinterp | 0.174*** | 0.079* | 0.569** |
| | (0.034) | (0.046) | (0.263) |
| pctover65 | 0.104* | −0.058 | 0.443 |
| | (0.053) | (0.070) | (0.471) |
| pctcollegegradinterp | 0.286*** | 0.421*** | 4.331*** |
| | (0.043) | (0.080) | (0.349) |
| logpop | −0.023*** | −0.012* | −0.126*** |
| | (0.003) | (0.007) | (0.032) |
| Observations | 11,223 | 10,871 | 13,741 |
| $R^2$ | 0.675 | 0.828 | 0.886 |
| Adjusted $R^2$ | 0.586 | 0.777 | 0.859 |
| Residual Std. Error | 0.049 (df = 8805) | 0.062 (df = 8380) | 0.420 (df = 11093) |
| *Note:* | | | *p<0.1; **p<0.05; ***p<0.01 |

Table 5: **Effect of Segregation on City Expenditures, IV Approach A**

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | dgepercap_cpi | highwayspercapNC_cpi | policepercapNC_cpi | parkspercapNC_cp |
| | (1) | (2) | (3) | (4) |
| H_citytract_NHW_i | −2.676*** | −0.363*** | −0.350*** | −0.034* |
| | (0.935) | (0.056) | (0.109) | (0.019) |
| dgepercap_cpilag | 1.472*** | | | |
| | (0.007) | | | |
| highwayspercapNC_cpilag | | 0.477*** | | |
| | | (0.003) | | |
| policepercapNC_cpilag | | | 0.955*** | |
| | | | (0.004) | |
| parkspercapNC_cpilag | | | | 0.869*** |
| | | | | (0.006) |
| diversityinterp | 0.264 | −0.032 | −0.020 | 0.004 |
| | (0.355) | (0.022) | (0.042) | (0.007) |
| pctblkpopinterp | 0.376 | 0.085*** | 0.096** | 0.003 |
| | (0.325) | (0.020) | (0.038) | (0.007) |
| pctasianpopinterp | 0.143 | −0.111** | −0.064 | −0.022 |
| | (0.940) | (0.056) | (0.110) | (0.019) |
| pctlatinopopinterp | 0.087 | 0.088*** | 0.088*** | 0.007 |
| | (0.284) | (0.017) | (0.033) | (0.006) |
| medincinterp | −0.004 | 0.001*** | 0.001*** | 0.0002*** |
| | (0.003) | (0.0002) | (0.0004) | (0.0001) |
| pctlocalgovworker_100 | −0.104*** | 0.021*** | 0.026*** | 0.003*** |
| | (0.032) | (0.002) | (0.004) | (0.001) |
| pctrentersinterp | −0.553 | 0.165*** | 0.187*** | 0.035*** |
| | (0.350) | (0.021) | (0.041) | (0.007) |
| pctover65 | 0.301 | 0.267*** | 0.238*** | 0.072*** |
| | (0.730) | (0.044) | (0.086) | (0.015) |
| pctcollegegradinterp | 0.248 | −0.044 | −0.101 | 0.028** |
| | (0.567) | (0.034) | (0.067) | (0.012) |
| Constant | 0.328 | −0.125*** | −0.183*** | −0.029*** |
| | (0.278) | (0.017) | (0.033) | (0.006) |
| Observations | 21,145 | 20,704 | 20,627 | 19,056 |
| $R^2$ | 0.685 | 0.615 | 0.789 | 0.540 |
| Adjusted $R^2$ | 0.685 | 0.615 | 0.789 | 0.539 |
| Residual Std. Error | 4.877 (df = 21125) | 0.290 (df = 20684) | 0.566 (df = 20607) | 0.093 (df = 19036 |
| *Note:* | | | | *p<0.1; **p<0.05; ***p<0.0 |

Table 6: **Effect of Segregation on City Expenditures, IV Approach B**

| | *Dependent variable:* | | |
|---|---|---|---|
| | sewerspercapNC_cpi | genrevownpercap_cpi | welfhoushealthNC_cpi |
| | (1) | (2) | (3) |
| H_citytract_NHW_i | $-0.363^{***}$ | $-1.873^{**}$ | $-0.115^{**}$ |
| | (0.060) | (0.789) | (0.054) |
| sewerspercapNC_cpilag | $0.064^{***}$ | | |
| | (0.008) | | |
| genrevownpercap_cpilag | | $1.235^{***}$ | |
| | | (0.006) | |
| welfhoushealthNC_cpilag | | | $0.893^{***}$ |
| | | | (0.005) |
| diversityinterp | $0.080^{***}$ | 0.047 | $-0.047^{**}$ |
| | (0.024) | (0.300) | (0.022) |
| pctblkpopinterp | $0.058^{**}$ | 0.360 | $0.076^{***}$ |
| | (0.025) | (0.274) | (0.023) |
| pctasianpopinterp | $-0.223^{***}$ | −0.029 | 0.009 |
| | (0.068) | (0.793) | (0.053) |
| pctlatinopopinterp | $-0.050^{***}$ | 0.206 | $0.078^{***}$ |
| | (0.019) | (0.240) | (0.017) |
| medincinterp | 0.0002 | 0.0003 | $0.001^{**}$ |
| | (0.0003) | (0.003) | (0.0002) |
| pctlocalgovworker_100 | 0.001 | −0.003 | $0.016^{***}$ |
| | (0.002) | (0.027) | (0.002) |
| pctrentersinterp | $0.073^{***}$ | 0.263 | $0.098^{***}$ |
| | (0.024) | (0.295) | (0.023) |
| pctover65 | $0.287^{***}$ | 0.782 | $0.127^{**}$ |
| | (0.051) | (0.616) | (0.050) |
| pctcollegegradinterp | 0.029 | −0.035 | −0.038 |
| | (0.040) | (0.478) | (0.038) |
| Constant | 0.004 | −0.174 | $-0.093^{***}$ |
| | (0.019) | (0.234) | (0.018) |
| Observations | 16,616 | 21,148 | 14,711 |
| $R^2$ | 0.006 | 0.681 | 0.699 |
| Adjusted $R^2$ | 0.005 | 0.681 | 0.698 |
| Residual Std. Error | 0.284 (df = 16596) | 4.115 (df = 21128) | 0.252 (df = 14691) |
| *Note:* | | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

Table 7: **TABLE A1 Summary Statistics: Racial Polarization Data**

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| Largest Racial Divide | 91 | 0.481 | 0.213 | 0.016 | 0.934 |
| *H* Index: Multigroup | 91 | 0.376 | 0.119 | 0.183 | 0.635 |
| *H* Index: Two-Group | 91 | 0.353 | 0.114 | 0.156 | 0.614 |
| Diversity | 91 | 0.623 | 0.088 | 0.323 | 0.736 |
| % Asian | 91 | 0.067 | 0.074 | 0.008 | 0.318 |
| % Black | 91 | 0.275 | 0.181 | 0.030 | 0.815 |
| % Latino | 91 | 0.229 | 0.155 | 0.009 | 0.605 |
| Median HH Income | 91 | 36.725 | 10.114 | 17.267 | 75.982 |
| % Renters | 91 | 0.535 | 0.092 | 0.368 | 0.718 |
| % College Degree | 91 | 0.167 | 0.056 | 0.049 | 0.359 |
| Biracial Contest | 91 | 0.725 | 0.449 | 0 | 1 |
| Nonpartisan Election | 91 | 0.714 | 0.454 | 0 | 1 |
| Primary Election | 91 | 0.352 | 0.480 | 0 | 1 |
| Population (logged) | 91 | 14.166 | 0.826 | 13.065 | 15.921 |
| White Ideology | 86 | 3.835 | 0.648 | 2.667 | 5.250 |


Nested Subnational Units: Diversity, Segregation, or Hierarchy?" In. https://doi.org/10.2139/ssrn.3254095.

Einstein, Katherine Levine (Ph D. 2012. "Divided Regions: Race, Political Segregation, and the Fragmentation of American Metropolitan Policy." PhD thesis.

Eric Oliver, J., and Janelle Wong. 2003. "Intergroup Prejudice in Multiethnic Settings." *American Journal of Political Science* 47 (4): 567–82. https://doi.org/10.1111/1540-5907.00040.

Fischer, Claude, Gretchen Stockmayer, Jon Stiles, and Michael Hout. 2004. "Distinguishing the Geographic Levels and Social Dimensions of U.S. Metropolitan Segregation, 1960–2000." *Demography* 41 (1): 37–59. https://doi.org/10.1353/dem.2004.0002.

Hopkins, Daniel J. 2009. "The Diversity Discount: When Increasing Ethnic and Racial Diversity Prevents Tax Increases." *The Journal of Politics* 71 (1): 160–77. https://doi.org/10.1017/S0022381608090105.

Lee, Alexander. 2018. "Ethnic Diversity and Ethnic Discrimination: Explaining Local Public Goods Provision." *Comparative Political Studies* 51 (10): 1351–83. https://doi.org/10.1177/0010414017740604.

Massey, Douglas S. 1993. *American Apartheid: Segregation and the Making of the Underclass*. Democracy and Urban Landscapes. Cambridge, Mass.: Harvard University Press. http://nrs.harvard.edu/urn-3:hul.ebookbatch.ACLS_batch:20170426MIU01100000000000000000544.

Oliver, J. Eric. 2010. *The Paradoxes of Integration:Race, Neighborhood, and Civic Life in Multiethnic America*. Chicago ; London: University of Chicago Press.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Trounstine, Jessica. 2015. "Replication Data for: Segregation and Inequality in Public Goods." Harvard Dataverse. https://doi.org/10.7910/DVN/4LZXTY.

———. 2016. "Segregation and Inequality in Public Goods: SEGREGATION AND INEQUALITY IN PUBLIC GOODS." *American Journal of Political Science* 60 (3): 709–25. https://doi.org/10.1111/ajps.12227.

van Buuren, Stef, and Karin Groothuis-Oudshoorn. 2011. "mice: Multivariate Imputation by Chained Equations in R." *Journal of Statistical Software* 45 (3): 1–67. https://www.jstatsoft.org/v45/i03/.