HKS SUP-135 Lab 9: The Long-Run Causal Effects of HOLC Redlining

Matt Khinda

4/21/2023

```r
# determine distance from threshold
holc$dist_from_cut <- holc$pop_1930 - 40000

# subset data
holc1930 <- subset(holc, year == 1930)
```

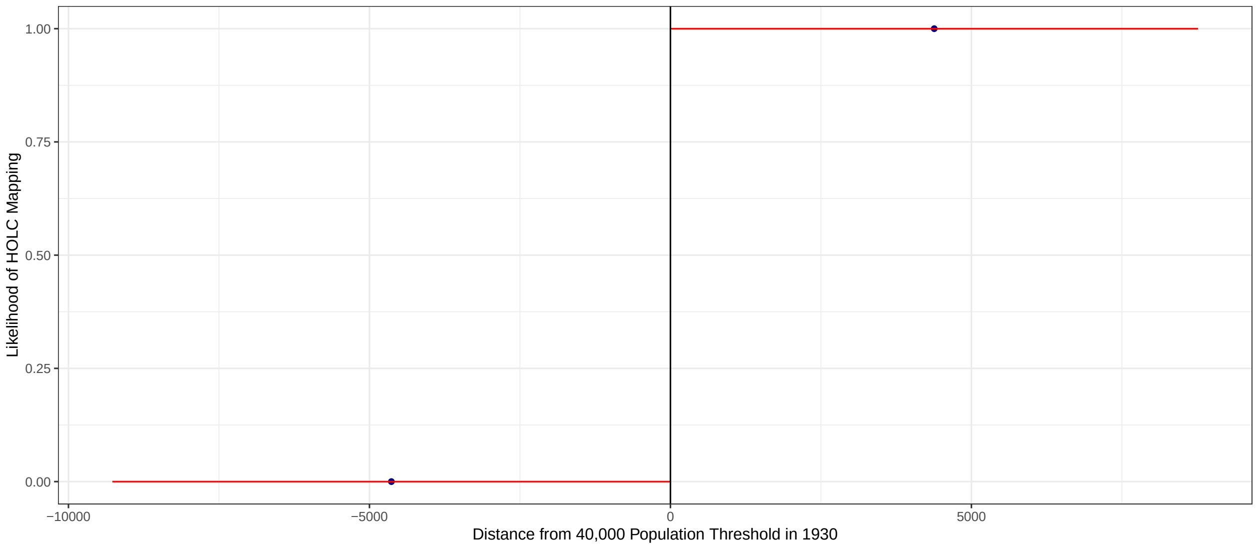
Question 1: Graphical Regression Discontinuity Analysis

```r
# Plot discontinuity of HOLC mapping
rdplot(y = holc1930$holc_map,
       x = holc1930$dist_from_cut,
       c = 0,
       p = 1,
       nbins = 20,
       binselect = "es",
       x.label = "Distance from 40,000 Population Threshold in 1930",
       y.label = "Likelihood of HOLC Mapping",
       title = "Discontinuity of HOLC Mapping and Population Size"
       )
```

1a: HOLC Maps and Population Size

```
## [1] "Warning: not enough variability in the outcome variable below the threshold"
## [1] "Warning: not enough variability in the outcome variable above the threshold"
```

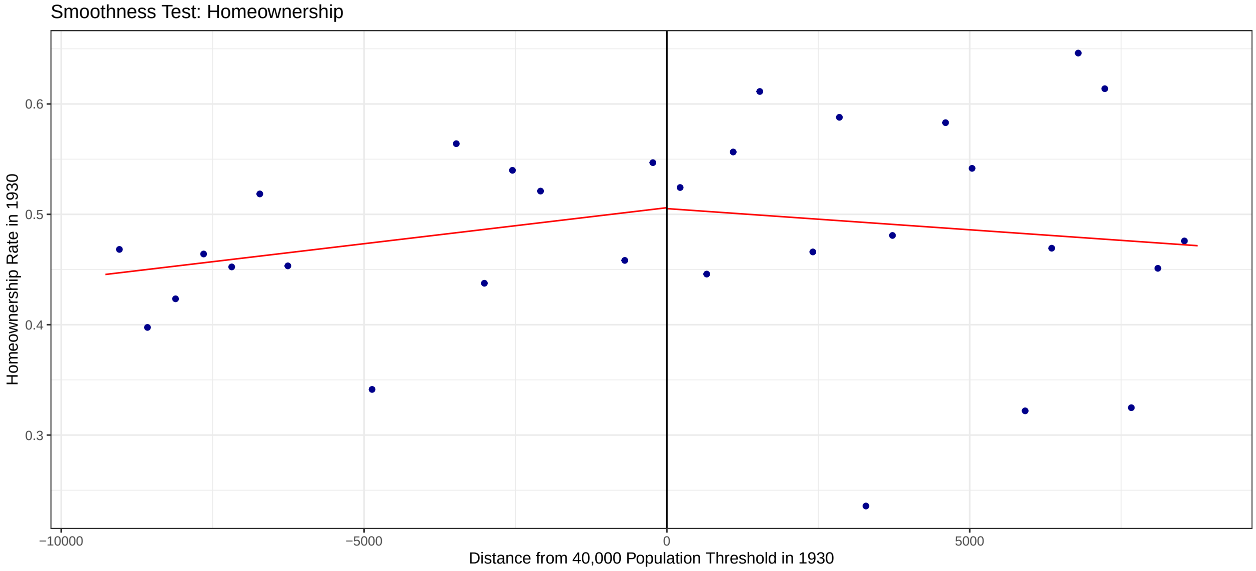
Discontinuity of HOLC Mapping and Population Size



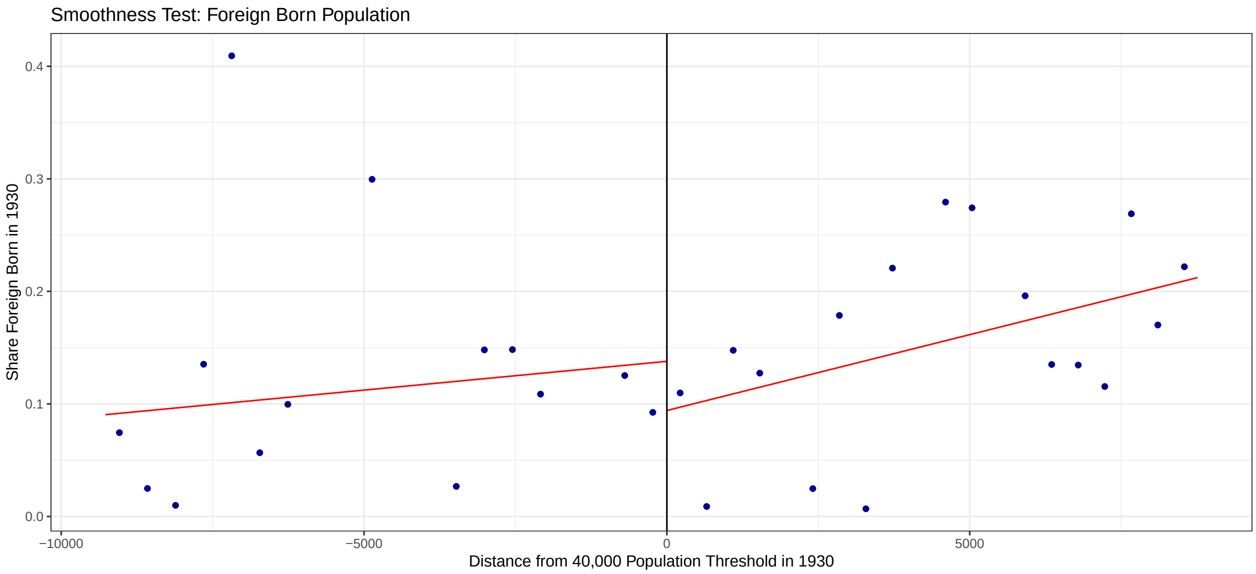

```r
# Plot discontinuity of ownership
rdplot(y = holc1930$ownhome,
       x = holc1930$dist_from_cut,
       c = 0,
       p = 1,
       nbins = 20,
       binselect = "es",
       x.label = "Distance from 40,000 Population Threshold in 1930",
       y.label = "Homeownership Rate in 1930",
       title = "Smoothness Test: Homeownership"
       )
```

1b: Predetermined Characteristics

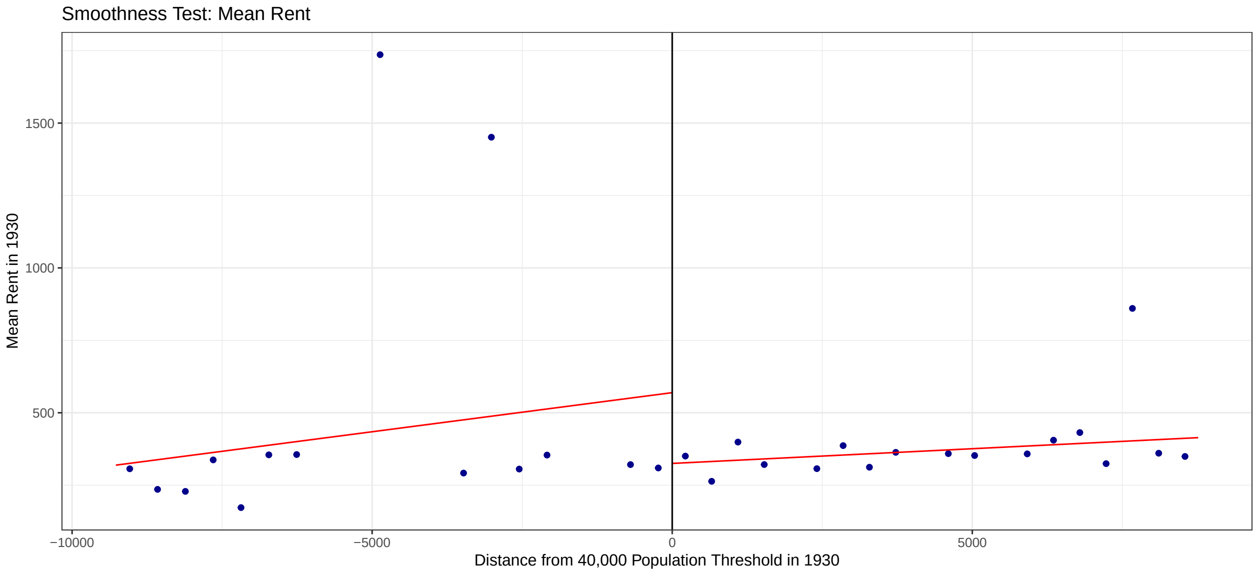
Smoothness Test: Homeownership



```r
# Plot discontinuity of foreign born pop
rdplot(y = holc1930$foreign_born,
       x = holc1930$dist_from_cut,
       c = 0,
       p = 1,
       nbins = 20,
       binselect = "es",
       x.label = "Distance from 40,000 Population Threshold in 1930",
       y.label = "Share Foreign Born in 1930",
       title = "Smoothness Test: Foreign Born Population"
       )
```




```r
# Plot discontinuity of rent
rdplot(y = holc1930$rent,
       x = holc1930$dist_from_cut,
       c = 0,
       p = 1,
       nbins = 20,
       binselect = "es",
       x.label = "Distance from 40,000 Population Threshold in 1930",
       y.label = "Mean Rent in 1930",
       title = "Smoothness Test: Mean Rent"
       )
```



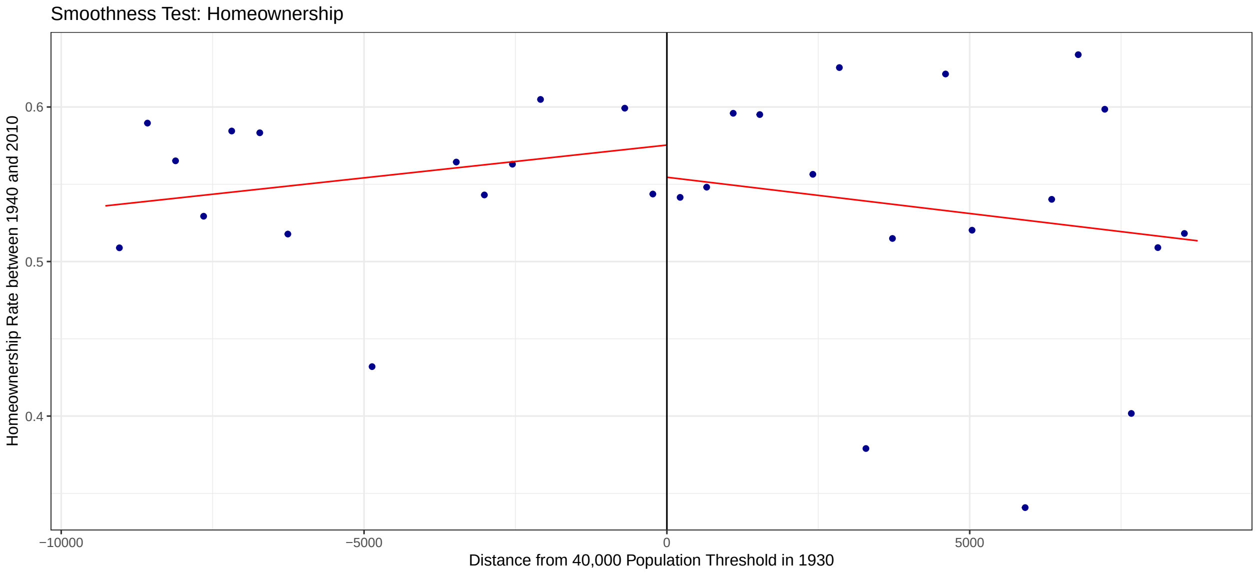
1c: Regression Discontinuity Design Validity As with any regression discontinuity design, we can evaluate the validity of the identification assumption by looking at the continuity of predetermined characteristics just above and just below the threshold. In the three charts above, we can see that homeownership in 1930 is essentially identical across the cut-off, and that the trend lines for share foreign born and mean rents in 1930 were only slightly discontinuous which is primarly driven by outliers. From this we can say that an RDD is a plausible approach to take for analyzing these data.

```r
# subset data for specific date range
holc_1940_2010 <- subset(holc, year >= 1940 & year <= 2010)

# Plot discontinuity of ownership
rdplot(y = holc_1940_2010$ownhome,
       x = holc_1940_2010$dist_from_cut,
       c = 0,
       p = 1,
       nbins = 20,
       binselect = "es",
       x.label = "Distance from 40,000 Population Threshold in 1930",
       y.label = "Homeownership Rate between 1940 and 2010",
       title = "Smoothness Test: Homeownership"
       )
```

1d: Homeownership and Population Size

```
## [1] "Mass points detected in the running variable."
```



Question 2: Regression Discontinuity Calculations

```
# Indicator for threshold
holc_1940_2010$above_threshold <- ifelse(holc_1940_2010$dist_from_cut >= 0, 1, 0)

# Create interaction term
holc_1940_2010$interaction_term <- (holc_1940_2010$dist_from_cut * holc_1940_2010$above_threshold)

# calculate rdd
rdd <- lm(ownhome ~ above_threshold +
            dist_from_cut +
            interaction_term,
          data = holc_1940_2010)

coeftest(rdd, vcovCL(rdd, cluster = holc_1940_2010$city_id))
```

```
##
## t test of coefficients:
##
##                     Estimate  Std. Error t value Pr(>|t|)
## (Intercept)       5.7538e-01  1.6876e-02 34.0941   <2e-16 ***
## above_threshold  -2.0823e-02  3.5684e-02 -0.5835   0.5599
## dist_from_cut     4.2424e-06  3.0817e-06  1.3767   0.1694
## interaction_term -8.9344e-06  6.9890e-06 -1.2783   0.2019
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

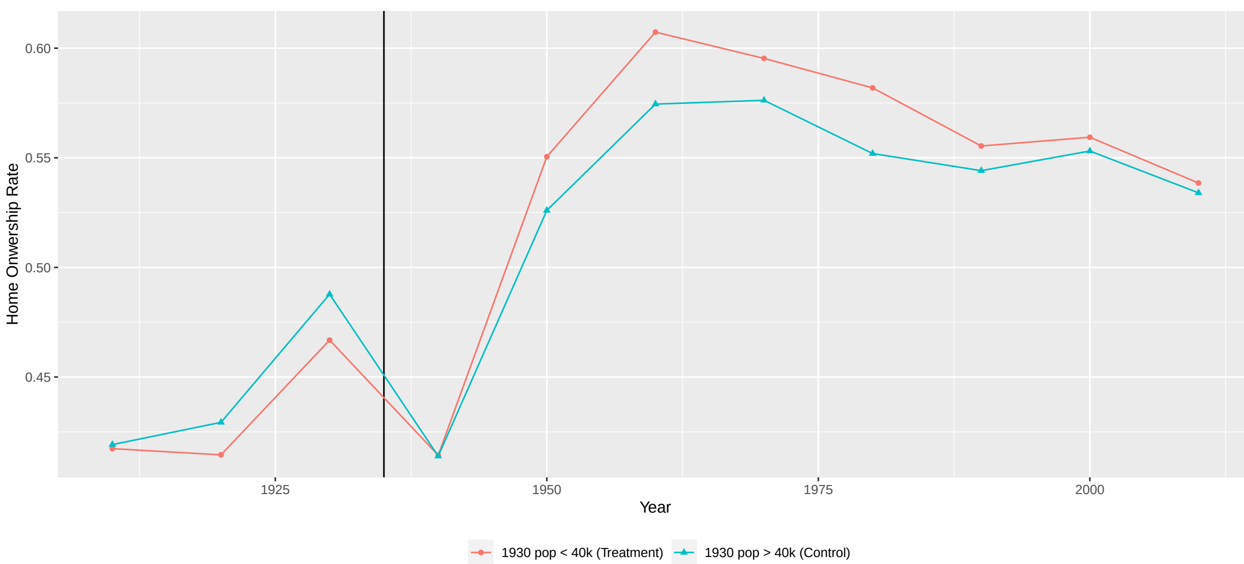
From the regression discontinuity calculations it appears as if the effect of being above the threshold on homeownership rates is not statistically significant. This is reflected in the plot in question 1d which shows only slight discontinuity at the threshold of a population size of 40,000. However, rather than concluding that there is no effect of HOLC mapping on homeownership rates, it is possible that a regression discontinuity design is not best suited for this type of analysis. To better understand the effects of the policy on trends over time, a difference-in-differences approach may be more appropriate.

Question 3: Graphical Difference-in-Differences Analysis

```
holc$treated <- ifelse(holc$dist_from_cut >= 0, 1, 0)

ggplot(holc, aes(x = year,
                 y = ownhome,
                 color = factor(treated,
                                labels = c("1930 pop < 40k (Treatment)",
                                           "1930 pop > 40k (Control)")),
                 shape = factor(treated,
                                labels = c("1930 pop < 40k (Treatment)",
                                           "1930 pop > 40k (Control)")))) +
      geom_vline(xintercept = 1935) +
      stat_summary(fun = "mean", geom = "point") +
      stat_summary(fun = "mean", geom = "line") +
      labs(x = "Year", y = "Home Onwership Rate", shape = "", color = "") +
      theme(legend.position = "bottom")
```

3a: Homeownership Rates Over Time (1910-2010)



3b: Parallel Trends Identification Assumption In order to establish a causal link between the effect of a policy and an outcome over time using the difference-in-differences method, the parallel trends assumption must be satisfied to preclude possible confounding effects. The parallel trends assumption requires that the trends in both the treatment and control groups are closely correlated before the intervention. Here, while there are only 3 data points before the 1935 threshold, we can stil observe that homeownership rates in cities with populations greater and less than 40,000 were moving parallel enough before the HOLC maps were drawn that a diff-in-diff analysis could be justified.

Question 4: Conditional Means

```r
# Calculate difference for treatment group pre and post
Y_treat_pre <- with(subset(holc, treated ==1 & year >= 1910 & year <= 1930),
                    mean(ownhome, na.rm = T))
Y_treat_post <- with(subset(holc, treated ==1 & year >= 1940 & year <= 1960),
                     mean(ownhome, na.rm = T))
Y_treat_diff <- Y_treat_post - Y_treat_pre

# Calculate difference for control group pre and post
Y_control_pre <- with(subset(holc, treated ==0 & year >= 1910 & year <= 1930),
                    mean(ownhome, na.rm = T))
Y_control_post <- with(subset(holc, treated ==0 & year >= 1940 & year <= 1960),
                     mean(ownhome, na.rm = T))
Y_control_diff <- Y_control_post - Y_control_pre

# Calculate difference in differences
Y_diffdiff <- Y_treat_diff - Y_control_diff
```

Using a conditional means approach, where difference in homeownership rates is calculated for both the control and treatment groups individually, then subtracted from one another to arrive at the difference in differences, the measured effect of HOLC maps is -0.0314523.

Question 5: Difference-in-Differences Calculations

```r
# Subset data
holc_1910_1960 <- subset(holc, year >= 1910 & year <= 1960)

# Assign indicator variable for post 1940
holc_1910_1960$post <- ifelse(holc_1910_1960$year >= 1940, 1, 0)
holc_1910_1960$above_threshold <- holc_1910_1960$treated

# Define interaction term
holc_1910_1960$dd_interaction_term <-
  (holc_1910_1960$above_threshold*holc_1910_1960$post)

# Create diff in diff regression
dd_lm <- lm(ownhome ~ dd_interaction_term +
              post + above_threshold, data = holc_1910_1960)
coeftest(dd_lm, vcovCL(dd_lm, cluster = holc_1910_1960$city_id))
```

```
##
## t test of coefficients:
##
##                       Estimate Std. Error t value  Pr(>|t|)
## (Intercept)          0.4328921  0.0150262 28.8091 < 2.2e-16 ***
## dd_interaction_term -0.0314523  0.0118392 -2.6566  0.008302 **
## post                 0.0896978  0.0088118 10.1793 < 2.2e-16 ***
## above_threshold      0.0125256  0.0269414  0.4649  0.642316
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Using a multivariable regression, the effect of the HOLC maps on homeownership rates is -0.0314523, which is equivalent to the value calculated using conditional means.

Question 6: Fixed-Effects Regression

```r
# Create fixed-effects diff in diff regression
dd_lm_mv <- lm(ownhome ~ dd_interaction_term +
              post + above_threshold +
              factor(city_id) +
              factor(year), data = holc_1910_1960)
coeftest(dd_lm_mv, vcovCL(dd_lm_mv, cluster = holc_1910_1960$city_id))
```

```
##
## t test of coefficients:
##
##                       Estimate  Std. Error     t value  Pr(>|t|)
## (Intercept)          4.6597e-01  7.1548e-03  6.5127e+01 < 2.2e-16 ***
## dd_interaction_term -3.2502e-02  1.2899e-02 -2.5196e+00   0.01236 *
## post                 1.8427e-01  1.3508e-02  1.3641e+01 < 2.2e-16 ***
## above_threshold      8.7443e-02  6.4497e-03  1.3558e+01 < 2.2e-16 ***
## factor(city_id)2    -1.2094e-01  6.2384e-16 -1.9387e+14 < 2.2e-16 ***
## factor(city_id)3    -3.1166e-02  3.5041e-16 -8.8943e+13 < 2.2e-16 ***
## factor(city_id)4     7.0799e-02  3.7874e-16  1.8693e+14 < 2.2e-16 ***
## factor(city_id)5     9.8910e-02  6.2760e-16  1.5760e+14 < 2.2e-16 ***
## factor(city_id)6    -2.6848e-01  2.7914e-03 -9.6181e+01 < 2.2e-16 ***
## factor(city_id)7    -1.0865e-01  6.2908e-16 -1.7272e+14 < 2.2e-16 ***
## factor(city_id)8    -3.4902e-01  5.2263e-16 -6.6782e+14 < 2.2e-16 ***
## factor(city_id)9    -1.8653e-01  4.5277e-16 -4.1198e+14 < 2.2e-16 ***
## factor(city_id)10   -1.4583e-02  6.2660e-16 -2.3273e+13 < 2.2e-16 ***
## factor(city_id)11   -3.3300e-01  3.2808e-16 -1.0150e+15 < 2.2e-16 ***
## factor(city_id)12   -1.1713e-01  3.2631e-16 -3.5895e+14 < 2.2e-16 ***
## factor(city_id)13   -2.2186e-01  3.2814e-16 -6.7610e+14 < 2.2e-16 ***
## factor(city_id)14   -1.7126e-02  6.3271e-16 -2.7068e+13 < 2.2e-16 ***
## factor(city_id)15   -1.5023e-01  3.2756e-03 -4.5864e+01 < 2.2e-16 ***
## factor(city_id)16   -1.7569e-01  3.2719e-16 -5.3695e+14 < 2.2e-16 ***
## factor(city_id)17   -1.0664e-01  6.5625e-16 -1.6249e+14 < 2.2e-16 ***
## factor(city_id)18   -9.7141e-02  3.2709e-16 -2.9698e+14 < 2.2e-16 ***
## factor(city_id)19   -5.5162e-02  3.6308e-16 -1.5193e+14 < 2.2e-16 ***
## factor(city_id)20    2.9924e-03  6.2293e-16  4.8038e+12 < 2.2e-16 ***
## factor(city_id)21    6.0836e-03  6.2378e-16  9.7528e+12 < 2.2e-16 ***
## factor(city_id)22   -6.5000e-02  6.2278e-16 -1.0437e+14 < 2.2e-16 ***
## factor(city_id)23   -1.9413e-01  6.2290e-16 -3.1165e+14 < 2.2e-16 ***
## factor(city_id)24   -2.5625e-01  3.2790e-16 -7.8148e+14 < 2.2e-16 ***
## factor(city_id)25   -8.4757e-02  3.2773e-16 -2.5862e+14 < 2.2e-16 ***
## factor(city_id)26   -3.7331e-02  3.3661e-16 -1.1090e+14 < 2.2e-16 ***
## factor(city_id)27   -1.6372e-01  3.2650e-16 -5.0145e+14 < 2.2e-16 ***
## factor(city_id)28    2.7169e-02  6.2155e-16  4.3711e+13 < 2.2e-16 ***
## factor(city_id)29   -1.2684e-01  6.2171e-16 -2.0402e+14 < 2.2e-16 ***
## factor(city_id)30    1.9024e-02  6.2294e-16  3.0540e+13 < 2.2e-16 ***
## factor(city_id)31   -1.0701e-01  3.3278e-16 -3.2155e+14 < 2.2e-16 ***
## factor(city_id)32   -1.3127e-03  3.2754e-16 -4.0077e+12 < 2.2e-16 ***
## factor(city_id)33   -4.4282e-02  6.2829e-16 -7.0480e+13 < 2.2e-16 ***
## factor(city_id)34   -6.7513e-02  3.3079e-16 -2.0410e+14 < 2.2e-16 ***
## factor(city_id)35    4.2714e-02  3.2748e-16  1.3043e+14 < 2.2e-16 ***
## factor(city_id)36   -1.2012e-01  6.2142e-16 -1.9330e+14 < 2.2e-16 ***
## factor(city_id)37   -1.0362e-01  6.2112e-16 -1.6682e+14 < 2.2e-16 ***
```

```
## factor(city_id)38   -2.3401e-01  3.2659e-16 -7.1652e+14 < 2.2e-16 ***
## factor(city_id)39   -1.3629e-01  3.2816e-16 -4.1533e+14 < 2.2e-16 ***
## factor(city_id)40   -2.3948e-01  3.2627e-16 -7.3400e+14 < 2.2e-16 ***
## factor(city_id)41   -3.1911e-02  6.2061e-16 -5.1418e+13 < 2.2e-16 ***
## factor(city_id)42   -8.9931e-03  6.3058e-16 -1.4262e+13 < 2.2e-16 ***
## factor(city_id)43    2.3623e-02  6.2181e-16  3.7991e+13 < 2.2e-16 ***
## factor(city_id)44   -5.0395e-02  6.2144e-16 -8.1094e+13 < 2.2e-16 ***
## factor(city_id)45   -2.2873e-02  6.2129e-16 -3.6816e+13 < 2.2e-16 ***
## factor(city_id)46   -6.2460e-02  3.2591e-16 -1.9165e+14 < 2.2e-16 ***
## factor(city_id)47   -2.0210e-01  3.2680e-16 -6.1841e+14 < 2.2e-16 ***
## factor(city_id)48   -1.3746e-01  3.2552e-16 -4.2228e+14 < 2.2e-16 ***
## factor(city_id)49   -8.8358e-02  6.2003e-16 -1.4251e+14 < 2.2e-16 ***
## factor(city_id)50   -3.1738e-02  3.2496e-16 -9.7668e+13 < 2.2e-16 ***
## factor(city_id)51   -6.8233e-02  6.1935e-16 -1.1017e+14 < 2.2e-16 ***
## factor(city_id)52   -1.4111e-01  6.2017e-16 -2.2753e+14 < 2.2e-16 ***
## factor(year)1920     3.7876e-03  7.1031e-03  5.3320e-01   0.59434
## factor(year)1930     5.9229e-02  7.5596e-03  7.8349e+00  1.27e-13 ***
## factor(year)1940    -1.7186e-01  8.6895e-03 -1.9778e+01 < 2.2e-16 ***
## factor(year)1950    -5.2588e-02  5.0456e-03 -1.0423e+01 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

When controlling for differences between cities and years using a fixed-effects multivariable regression, the magnitude of the effect of HOLC maps on increasing homeownership rates grows slightly from -0.0314523 to -0.0325016.

Question 7: Causal Effect of HOLC Redlining on Homeownership

Due to the parallel trends assumption being satisfied in homeownership rates between cities with a population of less than 40,000 (control) and those with more than 40,000 (treatment), we can make a causal claim about the impact of HOLC mapping—which was contingent on population size—on the rates of homeownership in those cities. Specifically, we observe that in cities that were redlined, homeownership grew -3.2501617% less than in areas that were not.