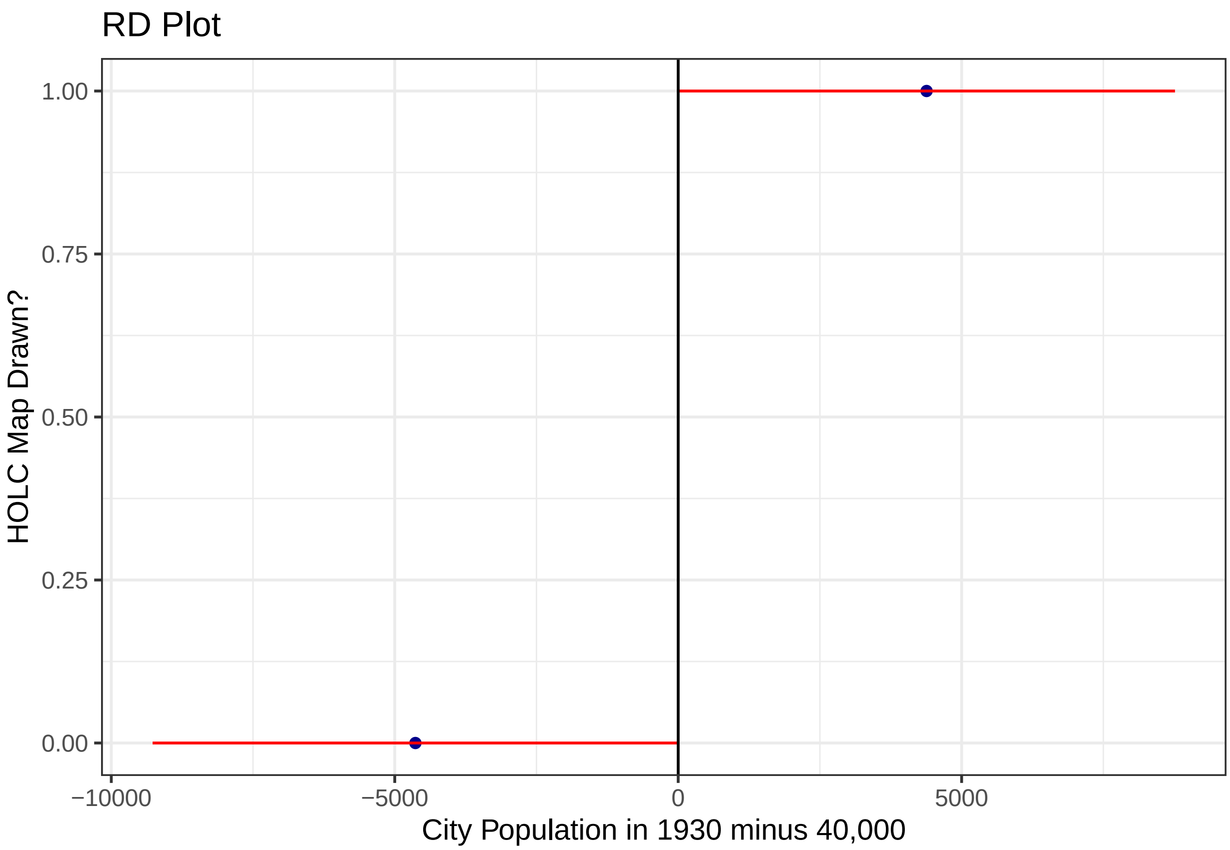
Lab9

2023-04-18

```
#Q1: Graphing an RDD analysis
#1A: Binned scatter plot of likelihood of having a HOLC map drawn conditional on 1930 population > 40,0
#creating indicator variable
dat$dist_from_cut <- dat$pop_1930 - 40000
dat1930 <- dat |> filter(year <= 1930)

#draw binned scatter plot with linear fit
redline_rdd <- rdplot(dat1930$holc_map,
 dat1930$dist_from_cut,
 p = 1,
 nbins = c(50, 50),
 binselect = "es",
 y.lim = c(0, 1),
 x.label = "City Population in 1930 minus 40,000",
 y.label = "HOLC Map Drawn?"
 )
```

```
## [1] "Mass points detected in the running variable."
## [1] "Warning: not enough variability in the outcome variable below the threshold"
## [1] "Warning: not enough variability in the outcome variable above the threshold"
```

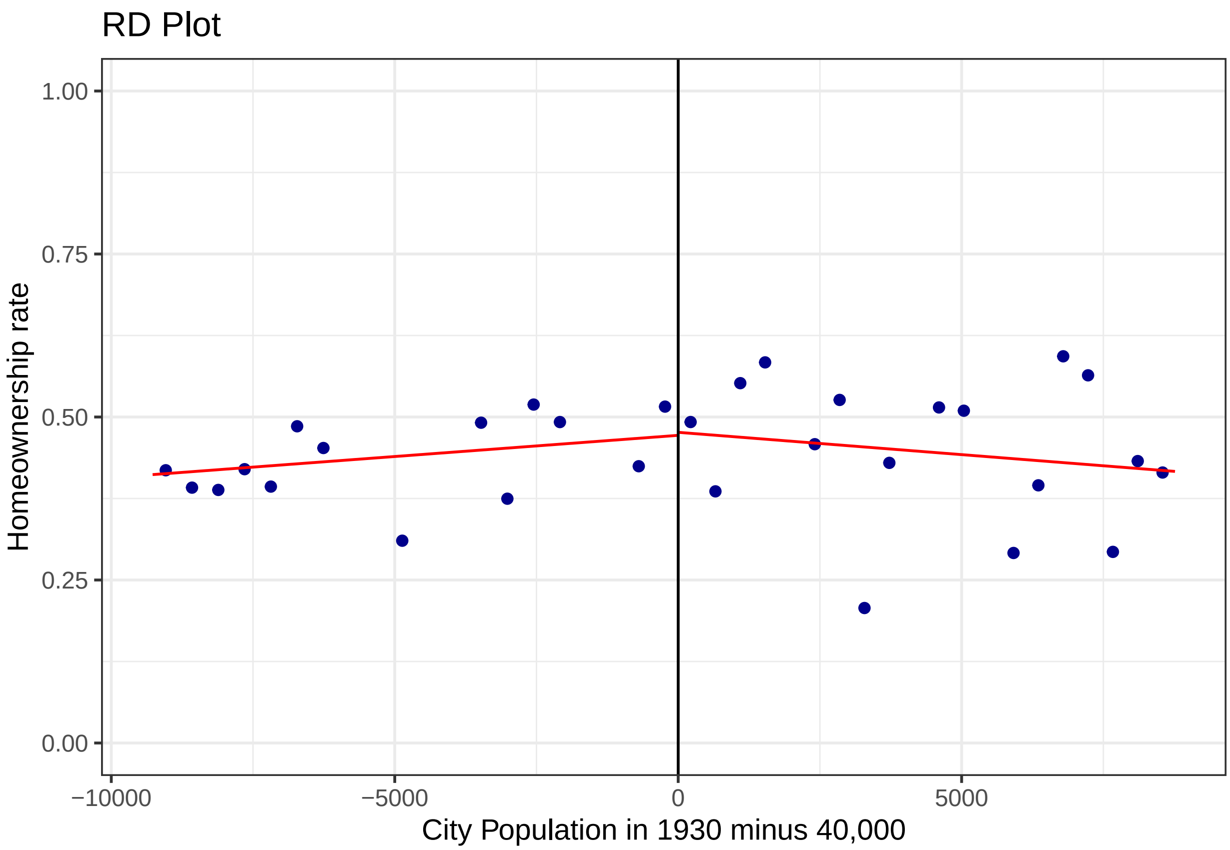



```
redline_rdd
```

```
## Call: rdplot
##
## Number of Obs.                 159
## Kernel                     Uniform
##
## Number of Obs.                  78               81
## Eff. Number of Obs.             78               81
## Order poly. fit (p)              1                1
## BW poly. fit (h)          9271.000         8764.000
## Number of bins scale         1.000            1.000
```

```
#1B: Binned scatter plots of other 1930 city characteristics to verify discontinuity of HOLC map
#Homeownership rate
#further subset to only 1930
homeown_rdd <- rdplot(dat1930$ownhome,
 dat1930$dist_from_cut,
 p = 1,
 nbins = c(20, 20),
 binselect = "es",
 y.lim = c(0, 1),
 x.label = "City Population in 1930 minus 40,000",
 y.label = "Homeownership rate"
 )
```

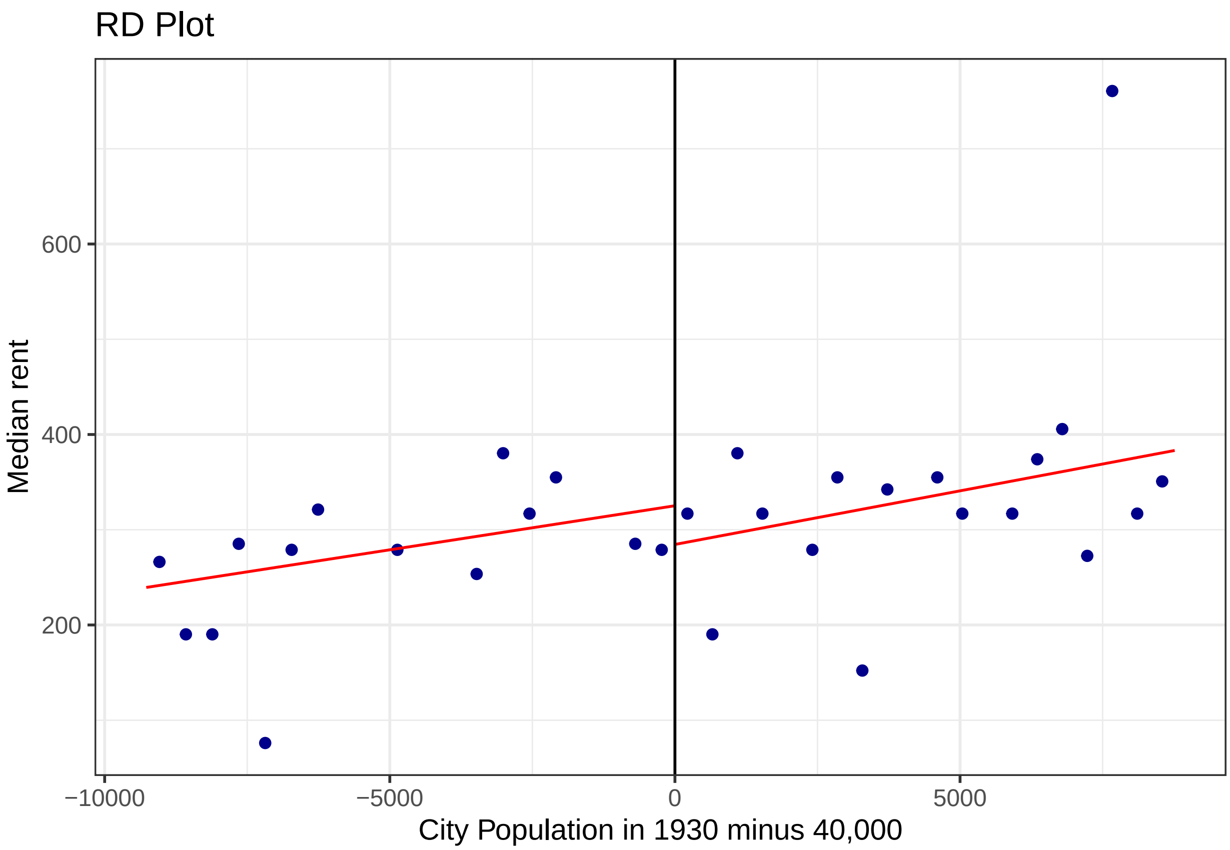
```
## [1] "Mass points detected in the running variable."
```




```
homeown_rdd
```

```
## Call: rdplot
##
## Number of Obs.                  159
## Kernel                      Uniform
##
## Number of Obs.                   78               81
## Eff. Number of Obs.              78               81
## Order poly. fit (p)               1                1
## BW poly. fit (h)           9271.000         8764.000
## Number of bins scale          1.000            1.000
```

```
#Median rent
rent_rdd <- rdplot(dat1930$median_gross_rent,
 dat1930$dist_from_cut,
 p = 1,
 nbins = c(20, 20),
 binselect = "es",
 x.label = "City Population in 1930 minus 40,000",
 y.label = "Median rent"
 )
```




```
rent_rdd
```

```
## Call: rdplot
##
## Number of Obs.                  53
## Kernel                     Uniform
##
## Number of Obs.                  26              27
## Eff. Number of Obs.             26              27
## Order poly. fit (p)              1               1
## BW poly. fit (h)          9271.000        8764.000
## Number of bins scale         1.000           1.000
```

Question 1C The binned scatterplot showing the likelihood of a HOLC map drawn clearly shows a discontinuity at the 40,000 1930 population mark. In contrast, two other related variables, homeownership rate and median gross rent, do not show discontinuities around any population threshold. This suggests that the RDD research design is valid for this question.

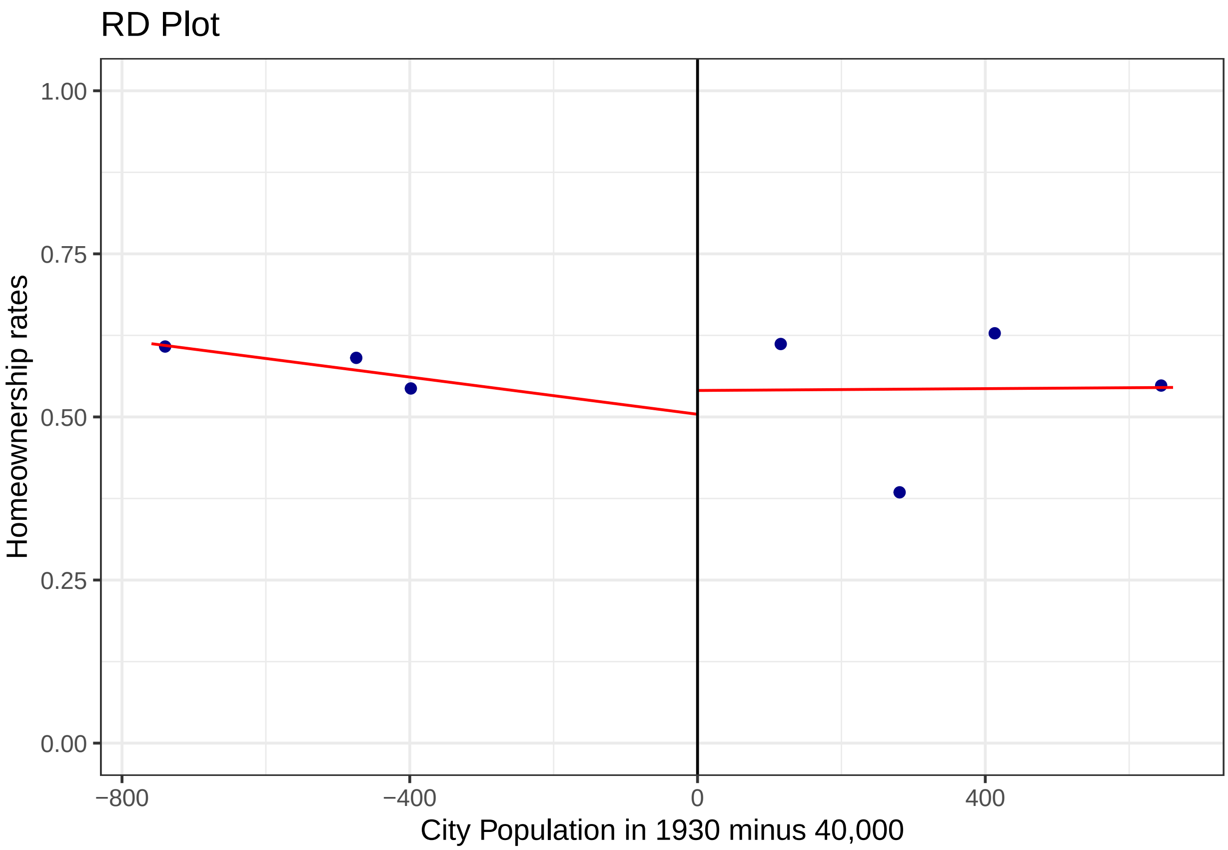
```
#Question 1D: Binned scatter plot of homeownership rates 1940-2010 discontinuity conditional on 1930 po

dfthrupresent <- dat %>% filter(year >= 1940 & year <= 2010)
narrow <- subset(dfthrupresent, dist_from_cut <= 1000 & dist_from_cut >= -1000)

present_rdd <- rdplot(narrow$ownhome,
 narrow$dist_from_cut,
 p = 1,
 nbins = c(20, 20),
 binselect = "es",
```

```
y.lim = c(0, 1),
x.label = "City Population in 1930 minus 40,000",
y.label = "Homeownership rates"
)
```

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




```
present_rdd
```

```
## Call: rdplot
##
## Number of Obs.                     56
## Kernel                        Uniform
##
## Number of Obs.                     24              32
## Eff. Number of Obs.                24              32
## Order poly. fit (p)                 1               1
## BW poly. fit (h)              759.000         661.000
## Number of bins scale            1.000           1.000
```

```
#Question 2: RDD regression model

#Create indicator for being above probation threshold
dat$above <- 0
dat$above[which(dat$dist_from_cut >= 0)] <- 1
view(dat)
#Interact dist_from_cut with non-probation
dat$interaction <- dat$dist_from_cut*dat$above
```

```r
##Subset data to [1940,2010] with new variables added
dat_narrow <- subset(dat, year<=2010 & year>=1940)
#Estimate regression
linear <- lm(ownhome ~ above + dist_from_cut + interaction , data =
dat_narrow)
#Report coefficients and standard errors
coeftest(linear, vcovCL(linear, cluster = dat_narrow$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         -2.0823e-02  3.5684e-02 -0.5835   0.5599
## dist_from_cut  4.2424e-06  3.0817e-06  1.3767   0.1694
## interaction   -8.9344e-06  6.9890e-06 -1.2783   0.2019
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 2 interpretation The regression discontinuity estimate is -0.020823, indicating that a city's 1930 population being above the HOLC redline threshold contributed to a decrease in the home ownership rate pooled from 1940 to 2010 by 2 percent, however, the result is not statistically significant.

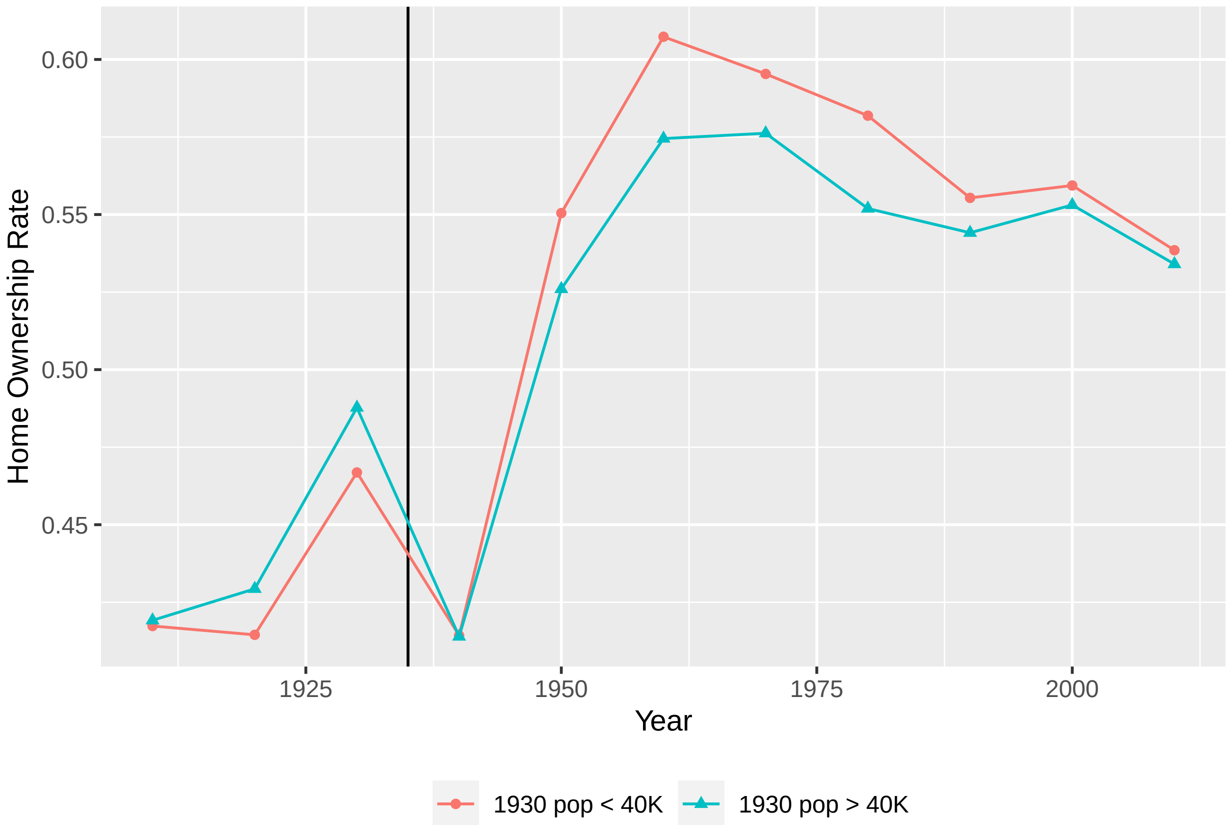
```r
#Question 3: Graphing DiD analysis
#creating indicator variable
dat$treat <- ifelse(dat$pop_1930>40000, 1, 0)
mean(dat$treat)
```

```
## [1] 0.5111876
```

```r
#Bin scatter plot - connected dots
ownhomedid <- ggplot(dat,
 aes(x=year,y=ownhome,
 colour = factor(treat, labels = c("1930 pop < 40K", "1930 pop > 40K")),
 shape = factor(treat, labels = c("1930 pop < 40K", "1930 pop > 40K")))) +
 geom_vline(xintercept=1935) +
 stat_summary(fun = "mean",geom="point") +
 stat_summary(fun = "mean",geom="line") +
 labs(x = "Year", y = "Home Ownership Rate", shape = "", colour = "") +
 theme(legend.position="bottom")

ownhomedid
```

```
## Warning: Removed 9 rows containing non-finite values (`stat_summary()`).
## Removed 9 rows containing non-finite values (`stat_summary()`).
```



Question 3b The parallel trends assumption is plausibly satisfied in the data. Cities above and below 40,000 in population were following roughly the same trajectory with regards to home ownership rates, with cities with a 1930 population above 40,000 having higher home ownership rates until roughly 1935-1940, where the trends flip. From there on out, cities with a 1930 population below 40,000 start to outpace their counterparts, peaking around 1960.

```r
#Q4: Reporting conditional means
y_post_treat = dat %>% filter(pop_1930 > 40000 & year >= 1940 & year <= 1960) %>%
  summarise(mean(ownhome, na.rm=TRUE))

y_pre_treat = dat %>% filter(pop_1930 > 40000 & year <= 1930) %>%
  summarise(mean(ownhome, na.rm=TRUE))

y_post_control = dat %>% filter(pop_1930 <= 40000 & year >= 1940 & year <= 1960) %>%
  summarise(mean(ownhome, na.rm=TRUE))

y_pre_control = dat %>% filter(pop_1930 <= 40000 & year <= 1930) %>%
  summarise(mean(ownhome, na.rm=TRUE))

diff = (y_post_treat - y_pre_treat) - (y_post_control - y_pre_control)
diff
```

```
##   mean(ownhome, na.rm = TRUE)
## 1                 -0.03145226
```

Question 4 The difference in difference in the 1940-1960 homeownership rate and the 1910-1930 homeownership rate between cities with a 1930 population above and below 40,000 is approximately a decrease of 3.1 percentage points.

```
#Question 5: DiD regression for 1910-1960
#Create indicators
dat$treat <- ifelse(dat$pop_1930>40000, 1, 0)
dat$post <- ifelse(dat$year>=1940, 1, 0)
#Interact treat and post
dat$dd <- dat$treat*dat$post

#Data frame with subset of years and new variables generated
dat_narrow <- subset(dat, year>=1910 & year <= 1960)
#Estimate regression
reg1 <- lm(ownhome ~ treat + post + dd, data=dat_narrow)
#Report coefficients and standard errors
coeftest(reg1, vcovCL(reg1, cluster = dat_narrow$city_id))
```

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

Question 5 The coefficient for the interaction variable of post and treat indicator variables is the same as the manually calculated difference in difference for treatment and control cities in the time periods specified.

```
#Question 6: Fixed effects regression

#Estimate regression
reg2 <- lm(ownhome ~ dd + factor(year) + factor(city_id) + dd,
 data=dat_narrow)
#Report coefficients and standard errors
coeftest(reg2, vcovCL(reg2, cluster = dat_narrow$city_id))
```

```
##
## t test of coefficients:
##
##                      Estimate  Std. Error     t value  Pr(>|t|)
## (Intercept)        5.5342e-01  6.1637e-03  8.9786e+01 < 2.2e-16 ***
## dd                -3.2502e-02  1.2899e-02 -2.5196e+00   0.01236 *
## 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   1.2407e-02  1.0683e-02  1.1614e+00   0.24655
## factor(year)1950   1.3168e-01  1.2201e-02  1.0792e+01 < 2.2e-16 ***
## factor(year)1960   1.8427e-01  1.3508e-02  1.3641e+01 < 2.2e-16 ***
## factor(city_id)2  -2.0838e-01  6.4497e-03 -3.2309e+01 < 2.2e-16 ***
## factor(city_id)3  -3.1166e-02  3.2891e-16 -9.4757e+13 < 2.2e-16 ***
## factor(city_id)4   7.0799e-02  3.3808e-16  2.0942e+14 < 2.2e-16 ***
## factor(city_id)5   1.1466e-02  6.4497e-03  1.7778e+00   0.07663 .
## factor(city_id)6  -2.6848e-01  2.7914e-03 -9.6181e+01 < 2.2e-16 ***
## factor(city_id)7  -1.9610e-01  6.4497e-03 -3.0404e+01 < 2.2e-16 ***
## factor(city_id)8  -3.4902e-01  3.1306e-16 -1.1149e+15 < 2.2e-16 ***
## factor(city_id)9  -1.8653e-01  3.3509e-16 -5.5665e+14 < 2.2e-16 ***
```

```
## factor(city_id)10 -1.0203e-01  6.4497e-03 -1.5819e+01 < 2.2e-16 ***
## factor(city_id)11 -3.3300e-01  3.1564e-16 -1.0550e+15 < 2.2e-16 ***
## factor(city_id)12 -1.1713e-01  3.2941e-16 -3.5557e+14 < 2.2e-16 ***
## factor(city_id)13 -2.2186e-01  3.3193e-16 -6.6837e+14 < 2.2e-16 ***
## factor(city_id)14 -1.0457e-01  6.4497e-03 -1.6213e+01 < 2.2e-16 ***
## factor(city_id)15 -2.3767e-01  5.0649e-03 -4.6925e+01 < 2.2e-16 ***
## factor(city_id)16 -1.7569e-01  3.3783e-16 -5.2004e+14 < 2.2e-16 ***
## factor(city_id)17 -1.9408e-01  6.4497e-03 -3.0091e+01 < 2.2e-16 ***
## factor(city_id)18 -9.7141e-02  3.3490e-16 -2.9006e+14 < 2.2e-16 ***
## factor(city_id)19 -5.5162e-02  3.2894e-16 -1.6770e+14 < 2.2e-16 ***
## factor(city_id)20 -8.4451e-02  6.4497e-03 -1.3094e+01 < 2.2e-16 ***
## factor(city_id)21 -8.1360e-02  6.4497e-03 -1.2615e+01 < 2.2e-16 ***
## factor(city_id)22 -1.5244e-01  6.4497e-03 -2.3636e+01 < 2.2e-16 ***
## factor(city_id)23 -2.8157e-01  6.4497e-03 -4.3656e+01 < 2.2e-16 ***
## factor(city_id)24 -2.5625e-01  3.2458e-16 -7.8949e+14 < 2.2e-16 ***
## factor(city_id)25 -8.4757e-02  3.2663e-16 -2.5949e+14 < 2.2e-16 ***
## factor(city_id)26 -3.7331e-02  3.2178e-16 -1.1601e+14 < 2.2e-16 ***
## factor(city_id)27 -1.6372e-01  3.4505e-16 -4.7449e+14 < 2.2e-16 ***
## factor(city_id)28 -6.0275e-02  6.4497e-03 -9.3454e+00 < 2.2e-16 ***
## factor(city_id)29 -2.1428e-01  6.4497e-03 -3.3224e+01 < 2.2e-16 ***
## factor(city_id)30 -6.8419e-02  6.4497e-03 -1.0608e+01 < 2.2e-16 ***
## factor(city_id)31 -1.0701e-01  3.2582e-16 -3.2843e+14 < 2.2e-16 ***
## factor(city_id)32 -1.3127e-03  3.2296e-16 -4.0645e+12 < 2.2e-16 ***
## factor(city_id)33 -1.3173e-01  6.4497e-03 -2.0424e+01 < 2.2e-16 ***
## factor(city_id)34 -6.7513e-02  3.2518e-16 -2.0762e+14 < 2.2e-16 ***
## factor(city_id)35  4.2714e-02  3.2767e-16  1.3036e+14 < 2.2e-16 ***
## factor(city_id)36 -2.0756e-01  6.4497e-03 -3.2182e+01 < 2.2e-16 ***
## factor(city_id)37 -1.9106e-01  6.4497e-03 -2.9623e+01 < 2.2e-16 ***
## factor(city_id)38 -2.3401e-01  3.2396e-16 -7.2233e+14 < 2.2e-16 ***
## factor(city_id)39 -1.3629e-01  3.2303e-16 -4.2192e+14 < 2.2e-16 ***
## factor(city_id)40 -2.3948e-01  3.3739e-16 -7.0980e+14 < 2.2e-16 ***
## factor(city_id)41 -1.1935e-01  6.4497e-03 -1.8506e+01 < 2.2e-16 ***
## factor(city_id)42 -9.6436e-02  6.4497e-03 -1.4952e+01 < 2.2e-16 ***
## factor(city_id)43 -6.3821e-02  6.4497e-03 -9.8952e+00 < 2.2e-16 ***
## factor(city_id)44 -1.3784e-01  6.4497e-03 -2.1371e+01 < 2.2e-16 ***
## factor(city_id)45 -1.1032e-01  6.4497e-03 -1.7104e+01 < 2.2e-16 ***
## factor(city_id)46 -6.2460e-02  3.2742e-16 -1.9076e+14 < 2.2e-16 ***
## factor(city_id)47 -2.0210e-01  3.1920e-16 -6.3314e+14 < 2.2e-16 ***
## factor(city_id)48 -1.3746e-01  3.1702e-16 -4.3361e+14 < 2.2e-16 ***
## factor(city_id)49 -1.7580e-01  6.4497e-03 -2.7257e+01 < 2.2e-16 ***
## factor(city_id)50 -3.1738e-02  3.1737e-16 -1.0000e+14 < 2.2e-16 ***
## factor(city_id)51 -1.5568e-01  6.4497e-03 -2.4137e+01 < 2.2e-16 ***
## factor(city_id)52 -2.2855e-01  6.4497e-03 -3.5436e+01 < 2.2e-16 ***
## factor(city_id)53 -8.7443e-02  6.4497e-03 -1.3558e+01 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 6 The fixed effects regression's interaction term variable's coefficient is a similar decrease of 3.25 percentage points in the homeownership rate for treated cities.

Question 7 It is evident that HOLC redlining practices had a negative impact on howe ownership rates in the decades after 1930. The RDD and DiD graphical reviews, the RDD regression, the normal DiD regression, and the fixed effects DiD regression suggest, to varying degrees of statistical significance, that being 'treated' - that is, being a city with a 1930 population above 40,000 - is associated with an decrease in the homeownership rate between 1940 and 1960 compared to 'control' cities.