ARE 213 Problem Set 3

Nick Depsky, Will Gorman, Peter Worley

December 3, 2018

Import Data

```
#setwd("~/Dropbox/Berkeley_tings/Fall 2018/ARE213/Problem Sets/SharedFiles/are213/PS2b")
setwd("C:\\Users\\will-\\Desktop\\are213\\PS3")
#setwd("C:\\Users\\Will\\Desktop\\are213\\PS3")

site_covariates <- read.dta("sitecovariates.dta")
all_sites <- read.dta13("allsites.dta")
all_covariates <- read.dta("allcovariates.dta")
mile_data <- read.dta("2miledata.dta")
```

1a - tract housing analysis

Outcome Variable: Median housing price in 2000 Control Variables: NPL site in tract in 2000, mean housing price in 1980

```
ols1 <- felm(lnmdvalhs0 ~ npl2000 + lnmeanhs8, data=all_sites)
summary(ols1, robust=T)
```

```
##
## Call:
##    felm(formula = lnmdvalhs0 ~ npl2000 + lnmeanhs8, data = all_sites)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.0581 -0.2173 -0.0241  0.1921  2.8403
##
## Coefficients:
##             Estimate Robust s.e t value Pr(>|t|)
## (Intercept)  2.40427    0.12410  19.373  < 2e-16 ***
## npl2000      0.04004    0.01195   3.352 0.000804 ***
## lnmeanhs8    0.85575    0.01147  74.622  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4057 on 42971 degrees of freedom
## Multiple R-squared(full model): 0.5792   Adjusted R-squared: 0.5791
## Multiple R-squared(proj model): 0.5792   Adjusted R-squared: 0.5791
## F-statistic(full model, *iid*):2.957e+04 on 2 and 42971 DF, p-value: < 2.2e-16
## F-statistic(proj model):  2784 on 2 and 42971 DF, p-value: < 2.2e-16
```

Outcome Variable: Median housing price in 2000 Control Variables: NPL site in tract in 2000, mean housing price in 1980 and various housing characteristics

```
ols2 <- felm(lnmdvalhs0 ~ npl2000 + lnmeanhs8 + firestoveheat80 + noaircond80 +
               nofullkitchen80 + zerofullbath80 + bedrms0_80occ + bedrms1_80occ +
               bedrms2_80occ + bedrms3_80occ +bedrms4_80occ  + blt0_1yrs80occ +
```

```
              blt2_5yrs80occ +blt6_10yrs80occ +blt10_20yrs80occ +blt20_30yrs80occ +
              blt30_40yrs80occ +detach80occ+attach80occ, data=all_sites)
summary(ols2, robust=T)
```

```
##
## Call:
##    felm(formula = lnmdvalhs0 ~ npl2000 + lnmeanhs8 + firestoveheat80 +      noaircond80 + nofullkitch
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.9335 -0.1848 -0.0048  0.1896  2.8751
##
## Coefficients:
##                    Estimate Robust s.e t value Pr(>|t|)
## (Intercept)       3.170e+00  2.142e-01  14.796  < 2e-16 ***
## npl2000           4.395e-02  1.156e-02   3.801 0.000145 ***
## lnmeanhs8         8.605e-01  1.791e-02  48.054  < 2e-16 ***
## firestoveheat80  -2.917e-02  1.964e-02  -1.485 0.137459
## noaircond80       3.201e-01  7.340e-03  43.618  < 2e-16 ***
## nofullkitchen80  -1.683e+00  1.680e-01 -10.015  < 2e-16 ***
## zerofullbath80    6.795e-01  1.350e-01   5.033 4.85e-07 ***
## bedrms0_80occ    -6.831e-01  2.224e-01  -3.071 0.002133 **
## bedrms1_80occ    -3.823e-01  7.846e-02  -4.873 1.11e-06 ***
## bedrms2_80occ    -1.008e+00  6.532e-02 -15.427  < 2e-16 ***
## bedrms3_80occ    -1.065e+00  6.664e-02 -15.989  < 2e-16 ***
## bedrms4_80occ    -6.188e-01  7.197e-02  -8.599  < 2e-16 ***
## blt0_1yrs80occ   -1.365e-01  4.249e-02  -3.213 0.001315 **
## blt2_5yrs80occ   -1.633e-01  2.985e-02  -5.472 4.46e-08 ***
## blt6_10yrs80occ  -1.240e-01  2.245e-02  -5.522 3.37e-08 ***
## blt10_20yrs80occ -6.766e-02  1.822e-02  -3.713 0.000205 ***
## blt20_30yrs80occ  7.271e-04  1.903e-02   0.038 0.969522
## blt30_40yrs80occ -6.922e-02  2.547e-02  -2.718 0.006578 **
## detach80occ      -8.333e-05  2.463e-02  -0.003 0.997301
## attach80occ       2.686e-01  2.749e-02   9.771  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3711 on 42954 degrees of freedom
## Multiple R-squared(full model): 0.648   Adjusted R-squared: 0.6478
## Multiple R-squared(proj model): 0.648   Adjusted R-squared: 0.6478
## F-statistic(full model, *iid*): 4162 on 19 and 42954 DF, p-value: < 2.2e-16
## F-statistic(proj model):  1871 on 19 and 42954 DF, p-value: < 2.2e-16
```

Outcome Variable: Median housing price in 2000 Control Variables: NPL site in tract in 2000, mean housing price in 1980 and various housing and demographic characteristics

```
ols3 <- felm(lnmdvalhs0 ~ npl2000 + lnmeanhs8 + firestoveheat80 + noaircond80 +
              nofullkitchen80 + zerofullbath80 + bedrms0_80occ + bedrms1_80occ +
              bedrms2_80occ + bedrms3_80occ +bedrms4_80occ  + blt0_1yrs80occ +
              blt2_5yrs80occ +blt6_10yrs80occ +blt10_20yrs80occ +blt20_30yrs80occ +
              blt30_40yrs80occ +detach80occ+attach80occ+pop_den8+shrblk8+
              shrhsp8+child8+old8+shrfor8+ffh8+smhse8+hsdrop8+no_hs_diploma8+
              ba_or_better8+unemprt8+povrat8+welfare8+avhhin8+tothsun8+ownocc8, data=all_sites)
summary(ols3, robust=T)
```

```
##
## Call:
##    felm(formula = lnmdvalhs0 ~ npl2000 + lnmeanhs8 + firestoveheat80 +      noaircond80 + nofullkitc
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8489 -0.1768  0.0039  0.1822  2.2764
##
## Coefficients:
##                     Estimate Robust s.e t value Pr(>|t|)
## (Intercept)       5.892e+00  2.200e-01  26.783  < 2e-16 ***
## npl2000           7.392e-02  9.693e-03   7.626 2.47e-14 ***
## lnmeanhs8         5.584e-01  1.945e-02  28.708  < 2e-16 ***
## firestoveheat80  -2.632e-02  2.053e-02  -1.282 0.199693
## noaircond80       4.374e-01  8.811e-03  49.635  < 2e-16 ***
## nofullkitchen80  -6.013e-01  1.535e-01  -3.917 8.96e-05 ***
## zerofullbath80    5.089e-01  1.206e-01   4.221 2.44e-05 ***
## bedrms0_80occ    -7.461e-01  2.248e-01  -3.319 0.000906 ***
## bedrms1_80occ    -2.029e-01  7.749e-02  -2.619 0.008830 **
## bedrms2_80occ    -4.166e-01  5.806e-02  -7.176 7.31e-13 ***
## bedrms3_80occ    -5.544e-01  5.583e-02  -9.930  < 2e-16 ***
## bedrms4_80occ    -4.961e-01  6.673e-02  -7.435 1.07e-13 ***
## blt0_1yrs80occ    9.559e-02  4.351e-02   2.197 0.028019 *
## blt2_5yrs80occ    1.633e-01  2.814e-02   5.802 6.62e-09 ***
## blt6_10yrs80occ   1.168e-01  2.153e-02   5.426 5.80e-08 ***
## blt10_20yrs80occ  6.760e-02  1.556e-02   4.343 1.41e-05 ***
## blt20_30yrs80occ  1.914e-02  1.446e-02   1.323 0.185760
## blt30_40yrs80occ  3.502e-02  2.458e-02   1.425 0.154253
## detach80occ      -2.706e-01  2.282e-02 -11.860  < 2e-16 ***
## attach80occ      -2.643e-01  2.630e-02 -10.049  < 2e-16 ***
## pop_den8          7.436e-06  3.903e-07  19.054  < 2e-16 ***
## shrblk8          -1.129e-01  1.311e-02  -8.609  < 2e-16 ***
## shrhsp8          -2.854e-01  1.975e-02 -14.448  < 2e-16 ***
## child8           -5.629e-01  5.529e-02 -10.180  < 2e-16 ***
## old8             -3.520e-01  4.442e-02  -7.925 2.34e-15 ***
## shrfor8           1.230e+00  4.333e-02  28.384  < 2e-16 ***
## ffh8             -4.794e-02  3.423e-02  -1.400 0.161391
## smhse8            4.376e-01  2.260e-02  19.369  < 2e-16 ***
## hsdrop8           1.450e-02  2.438e-02   0.595 0.552085
## no_hs_diploma8   -3.103e-01  3.401e-02  -9.124  < 2e-16 ***
## ba_or_better8     4.784e-01  3.335e-02  14.345  < 2e-16 ***
## unemprt8         -1.236e+00  7.165e-02 -17.243  < 2e-16 ***
## povrat8          -3.838e-01  4.938e-02  -7.773 7.84e-15 ***
## welfare8          8.570e-01  6.608e-02  12.970  < 2e-16 ***
## avhhin8           1.306e-05  6.865e-07  19.030  < 2e-16 ***
## tothsun8          1.059e-05  6.411e-06   1.651 0.098732 .
## ownocc8          -1.325e-04  1.017e-05 -13.024  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3244 on 42937 degrees of freedom
## Multiple R-squared(full model): 0.7311   Adjusted R-squared: 0.7309
## Multiple R-squared(proj model): 0.7311   Adjusted R-squared: 0.7309
## F-statistic(full model, *iid*): 3242 on 36 and 42937 DF, p-value: < 2.2e-16
```

```
## F-statistic(proj model):  2641 on 36 and 42937 DF, p-value: < 2.2e-16
```

Outcome Variable: Median housing price in 2000 Control Variables: NPL site in tract in 2000, mean housing price in 1980 and various housing and demographic characteristics and state fixed effects (exlcuded variables that returned errors in fixed effects treatment bedrms5_80occ, blt40_yrs80occ, percent mobile home single family housing 1980)

```
ols4 <- felm(lnmdvalhs0 ~ npl2000 + lnmeanhs8 + firestoveheat80 + noaircond80 +
               nofullkitchen80 + zerofullbath80 + bedrms0_80occ + bedrms1_80occ +
               bedrms2_80occ + bedrms3_80occ +bedrms4_80occ  + blt0_1yrs80occ +
               blt2_5yrs80occ +blt6_10yrs80occ +blt10_20yrs80occ +blt20_30yrs80occ +
               blt30_40yrs80occ +detach80occ+attach80occ+pop_den8+shrblk8+
               shrhsp8+child8+old8+shrfor8+ffh8+smhse8+hsdrop8+no_hs_diploma8+
               ba_or_better8+unemprt8+povrat8+welfare8+avhhin8+tothsun8+
               ownocc8 | statefips, data=all_sites)
summary(ols4, robust=T)
```

```
##
## Call:
##    felm(formula = lnmdvalhs0 ~ npl2000 + lnmeanhs8 + firestoveheat80 +      noaircond80 + nofullkitch
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6913 -0.1541  0.0003  0.1584  2.5369
##
## Coefficients:
##                    Estimate Robust s.e t value Pr(>|t|)
## npl2000           6.635e-02  8.797e-03   7.542 4.71e-14 ***
## lnmeanhs8         4.964e-01  2.067e-02  24.010  < 2e-16 ***
## firestoveheat80   7.965e-02  3.192e-02   2.495 0.012599 *
## noaircond80       3.252e-01  9.485e-03  34.282  < 2e-16 ***
## nofullkitchen80  -5.408e-01  1.463e-01  -3.697 0.000218 ***
## zerofullbath80    5.940e-01  1.142e-01   5.202 1.98e-07 ***
## bedrms0_80occ    -4.740e-01  2.239e-01  -2.117 0.034296 *
## bedrms1_80occ    -1.241e-01  7.453e-02  -1.666 0.095801 .
## bedrms2_80occ    -3.759e-01  5.548e-02  -6.775 1.26e-11 ***
## bedrms3_80occ    -4.438e-01  5.368e-02  -8.267  < 2e-16 ***
## bedrms4_80occ    -4.767e-01  6.305e-02  -7.561 4.06e-14 ***
## blt0_1yrs80occ    1.017e-01  4.365e-02   2.330 0.019821 *
## blt2_5yrs80occ    1.129e-01  2.625e-02   4.301 1.70e-05 ***
## blt6_10yrs80occ   4.059e-02  2.081e-02   1.951 0.051089 .
## blt10_20yrs80occ -2.183e-02  1.409e-02  -1.549 0.121327
## blt20_30yrs80occ -2.801e-02  1.308e-02  -2.142 0.032204 *
## blt30_40yrs80occ  1.059e-02  2.250e-02   0.470 0.638055
## detach80occ      -2.530e-01  2.210e-02 -11.448  < 2e-16 ***
## attach80occ      -1.886e-01  2.541e-02  -7.421 1.19e-13 ***
## pop_den8          6.770e-06  3.880e-07  17.446  < 2e-16 ***
## shrblk8          -9.578e-02  1.265e-02  -7.571 3.78e-14 ***
## shrhsp8          -8.938e-02  2.124e-02  -4.208 2.58e-05 ***
## child8           -3.631e-01  5.189e-02  -6.997 2.65e-12 ***
## old8             -1.940e-01  4.325e-02  -4.486 7.26e-06 ***
## shrfor8           5.863e-01  3.925e-02  14.937  < 2e-16 ***
## ffh8             -1.125e-01  3.324e-02  -3.384 0.000716 ***
## smhse8            3.747e-01  2.290e-02  16.364  < 2e-16 ***
## hsdrop8           2.231e-02  2.323e-02   0.961 0.336699
```

```
## no_hs_diploma8   -3.568e-01  3.294e-02 -10.829  < 2e-16 ***
## ba_or_better8     5.054e-01  3.464e-02  14.590  < 2e-16 ***
## unemprt8         -1.377e+00  7.431e-02 -18.527  < 2e-16 ***
## povrat8          -9.174e-02  4.770e-02  -1.923 0.054441 .
## welfare8          2.959e-01  6.490e-02   4.559 5.14e-06 ***
## avhhin8           1.383e-05  6.198e-07  22.314  < 2e-16 ***
## tothsun8          1.429e-05  6.148e-06   2.324 0.020119 *
## ownocc8          -1.323e-04  9.683e-06 -13.667  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2941 on 42887 degrees of freedom
## Multiple R-squared(full model): 0.7793   Adjusted R-squared: 0.7789
## Multiple R-squared(proj model): 0.6876   Adjusted R-squared: 0.687
## F-statistic(full model, *iid*): 1761 on 86 and 42887 DF, p-value: < 2.2e-16
## F-statistic(proj model):  2032 on 36 and 42887 DF, p-value: < 2.2e-16
```

The coefficient for NPL status is significant and positive in all of the regressions. Since all of the coefficients for NPL are positive, it suggests this designation has an unambiguously positive correlation with housing value. The magnitude of the coefficient on NPL term doesn't change notably as the additional demographic and economic variables are added .

In the first three regressions, we have to make the assumptions associated with a selection on observables research design.

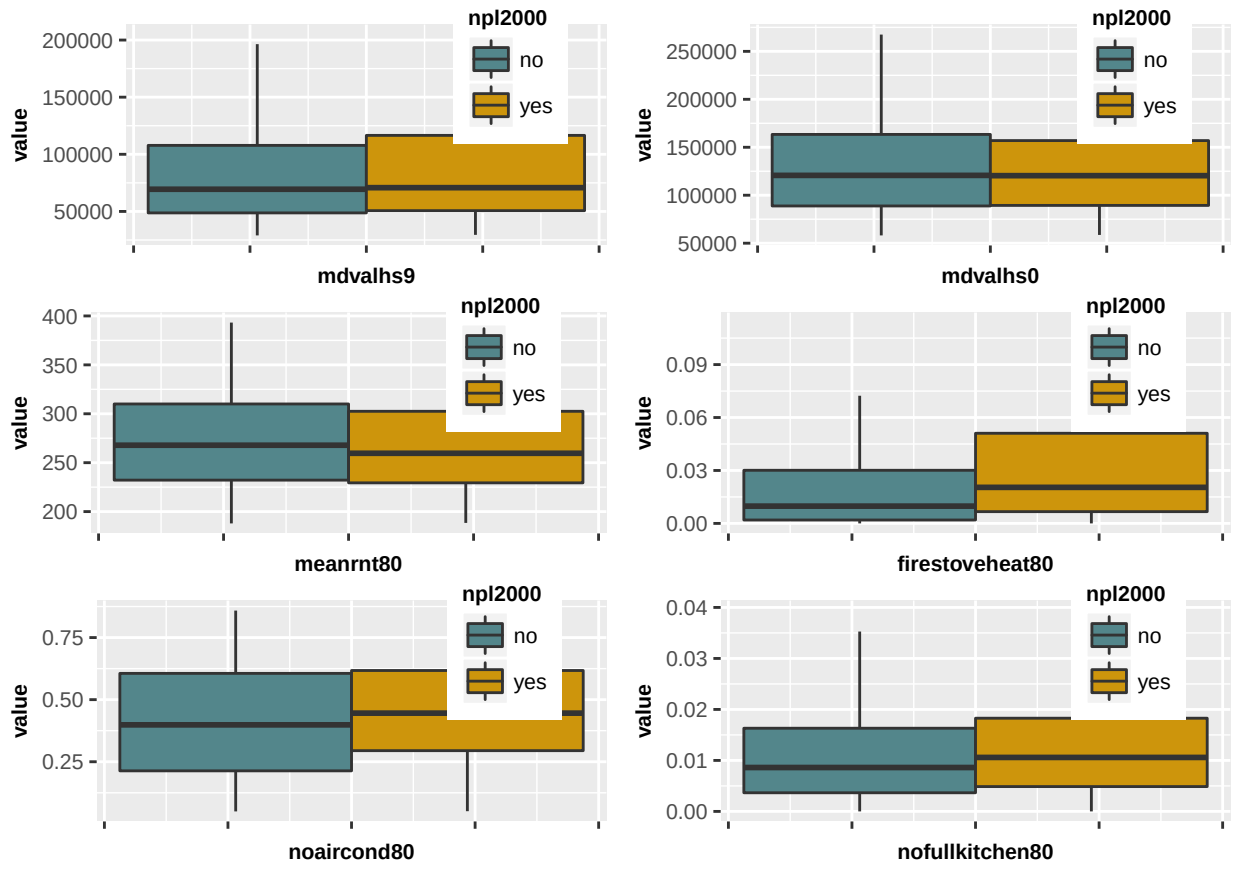
The coefficient will be unbiased if the following assumptions are met: - residuals are mean zero - residuals are homoskedastic with finite variance - residuals are uncorrelated with NPL2000

The fixed effect design, on the other hand, does not have to rely on the selection of observables assumption but rather that the the observables individual specific effects are not correlated with the error term and that the observeables are not correlated with the individual specific effects.

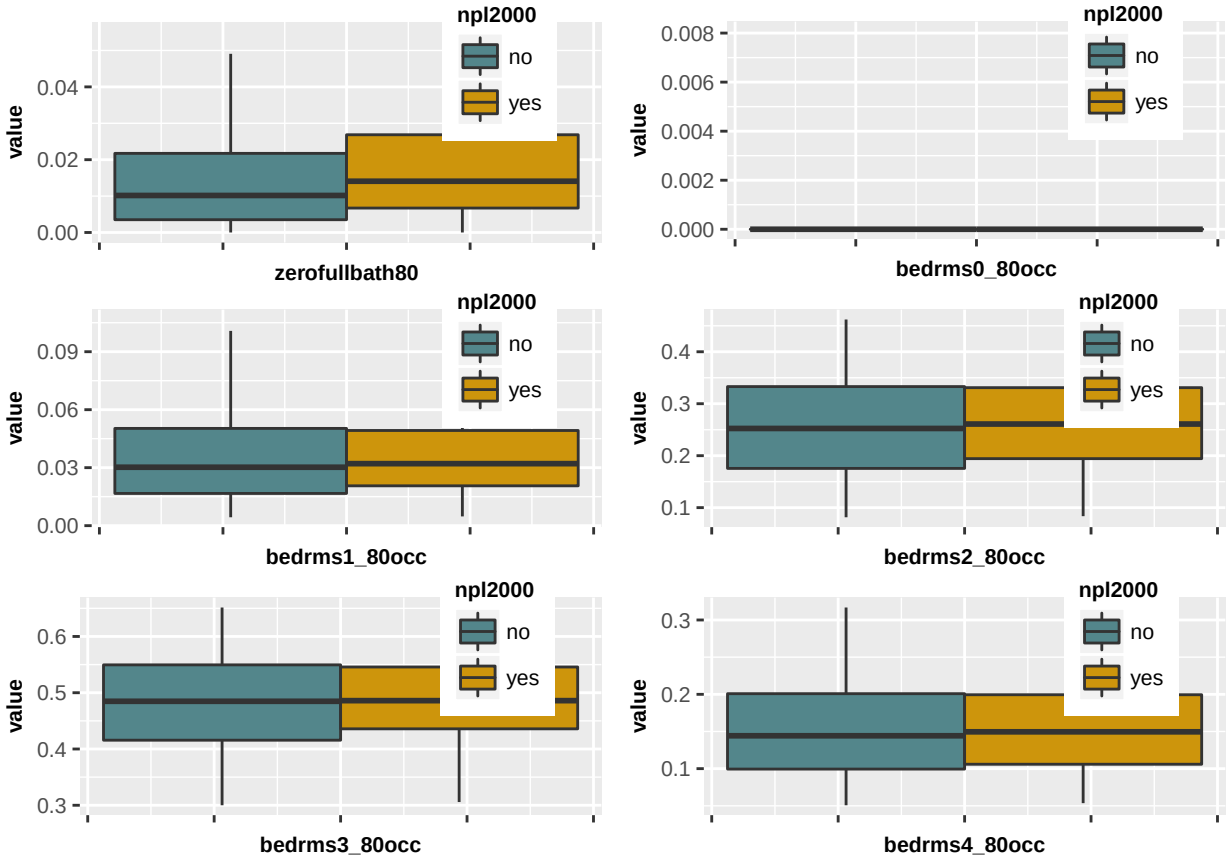
1b - treatment vs. comparison

```
all_covariates$npl2000 <- as.character(all_covariates$npl2000)
covariates <- c("mdvalhs9","mdvalhs0","meanrnt80","firestoveheat80","noaircond80","nofullkitchen80",
                "zerofullbath80",
                "bedrms0_80occ","bedrms1_80occ","bedrms2_80occ","bedrms3_80occ","bedrms4_80occ",
                "blt0_1yrs80occ","blt2_5yrs80occ","blt6_10yrs80occ","blt10_20yrs80occ",
                "blt20_30yrs80occ","blt30_40yrs80occ","detach80occ","attach80occ","pop_den8",
                "shrblk8","shrhsp8","child8","shrfor8","ffh8","smhse8","hsdrop8",
                "no_hs_diploma8","ba_or_better8","unemprt8","povrat8","welfare8","avhhin8",
                "tothsun8","ownocc8")
pl <- list()

for (i in c(1:6)) {

  pl[[i]] <- ggplot(all_covariates, aes_string(x = i,y=covariates[i], fill="npl2000")) +
  geom_boxplot(outlier.shape=NA)  + ylab(label = "value") + xlab(label = covariates[i]) +
  theme(axis.text.x=element_blank(),axis.text=element_text(size=8),
        axis.title=element_text(size=8,face="bold"),
        legend.text=element_text(size=8),legend.title=element_text(size=8,face="bold"),
        legend.position = c(0.8,.8)) +
  scale_fill_manual(values = c("cadetblue4", "darkgoldenrod3"),labels = c("no", "yes")) +
```

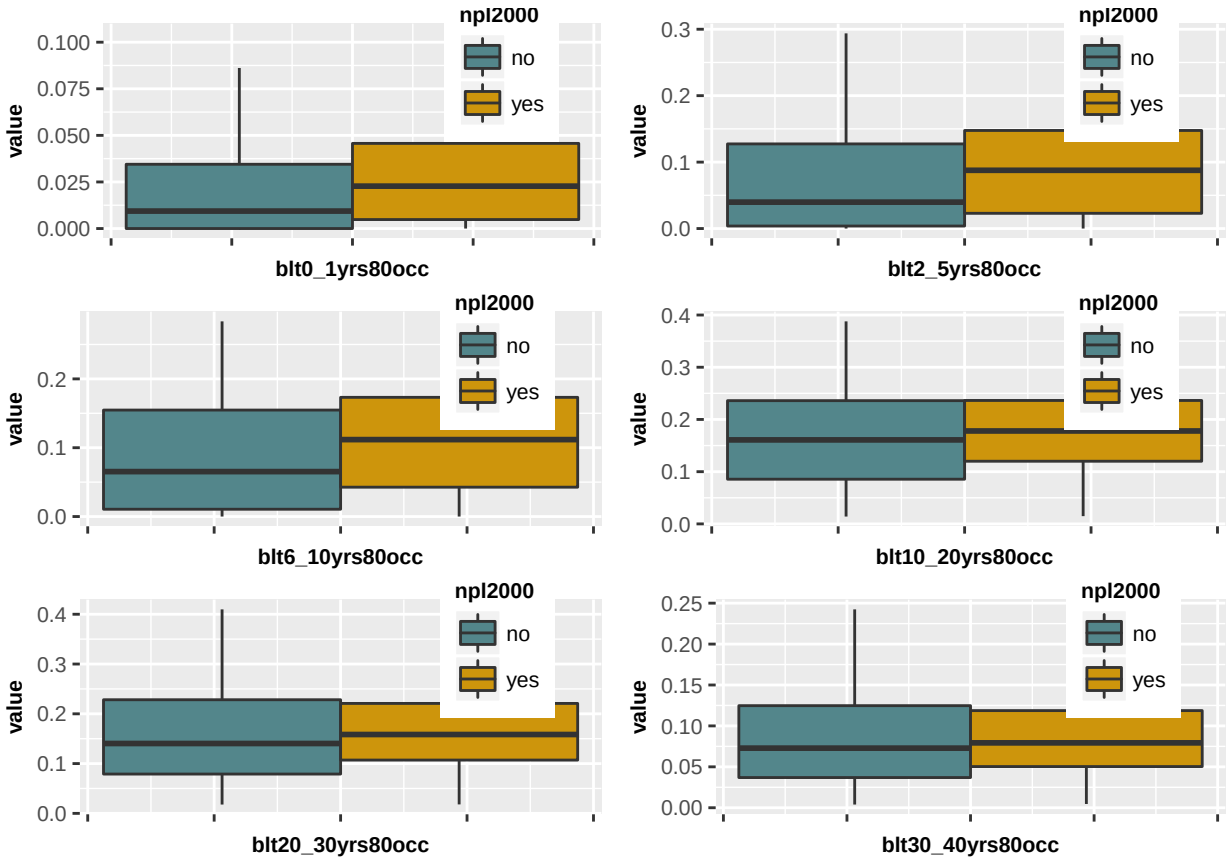
```
  guides(colour = guide_legend(override.aes = list(size=10))) +
    scale_y_continuous(limits = quantile(all_covariates[,covariates[i]], c(0.1, 0.9),na.rm=T))

}

suppressWarnings(grid.arrange(grobs = pl))
```



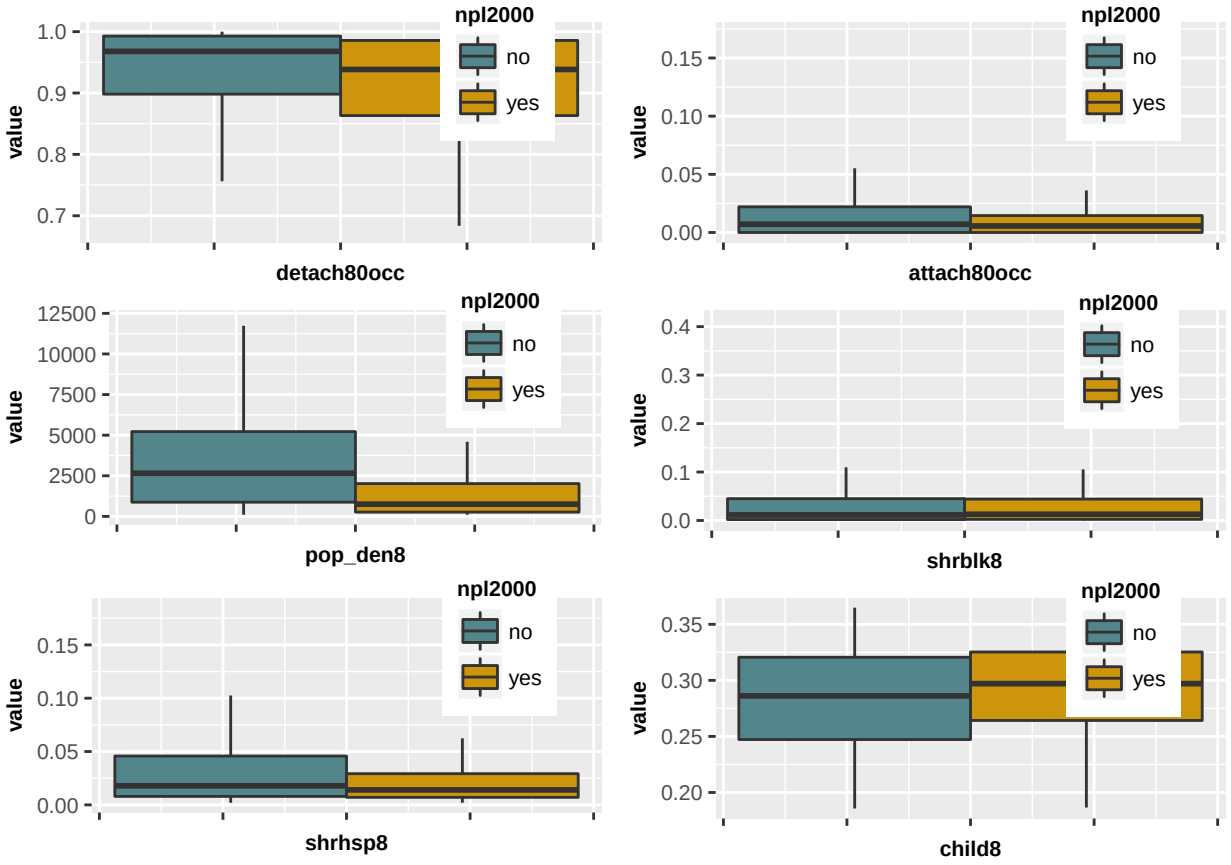

```
pl <- list()

for (i in c(7:12)) {

  pl[[i-6]] <- ggplot(all_covariates, aes_string(x = i,y=covariates[i], fill="npl2000")) +
  geom_boxplot(outlier.shape=NA)  + ylab(label = "value") + xlab(label = covariates[i]) +
  theme(axis.text.x=element_blank(),axis.text=element_text(size=8),
        axis.title=element_text(size=8,face="bold"),
        legend.text=element_text(size=8),legend.title=element_text(size=8,face="bold"),
        legend.position = c(0.8,.8)) +
  scale_fill_manual(values = c("cadetblue4", "darkgoldenrod3"),labels = c("no", "yes")) +
  guides(colour = guide_legend(override.aes = list(size=10))) +
    scale_y_continuous(limits = quantile(all_covariates[,covariates[i]], c(0.1, 0.9)))

}

suppressWarnings(grid.arrange(grobs = pl))
```



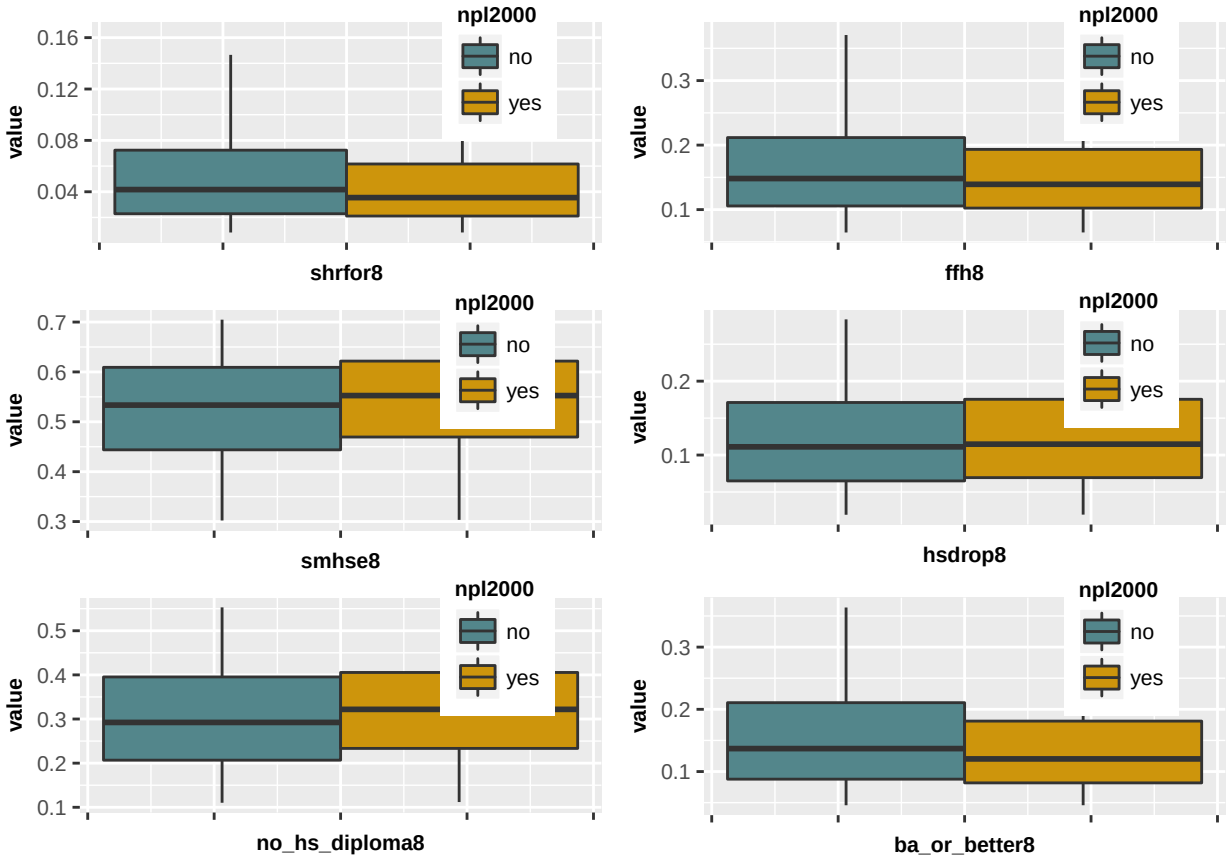

```
pl <- list()

for (i in c(13:18)) {

  pl[[i-12]] <- ggplot(all_covariates, aes_string(x = i,y=covariates[i], fill="npl2000")) +
  geom_boxplot(outlier.shape=NA)  + ylab(label = "value") + xlab(label = covariates[i]) +
  theme(axis.text.x=element_blank(),axis.text=element_text(size=8),
        axis.title=element_text(size=8,face="bold"),
        legend.text=element_text(size=8),legend.title=element_text(size=8,face="bold"),
        legend.position = c(0.8,.8)) +
  scale_fill_manual(values = c("cadetblue4", "darkgoldenrod3"),labels = c("no", "yes")) +
  guides(colour = guide_legend(override.aes = list(size=10))) +
    scale_y_continuous(limits = quantile(all_covariates[,covariates[i]], c(0.1, 0.9)))

}

suppressWarnings(grid.arrange(grobs = pl))
```



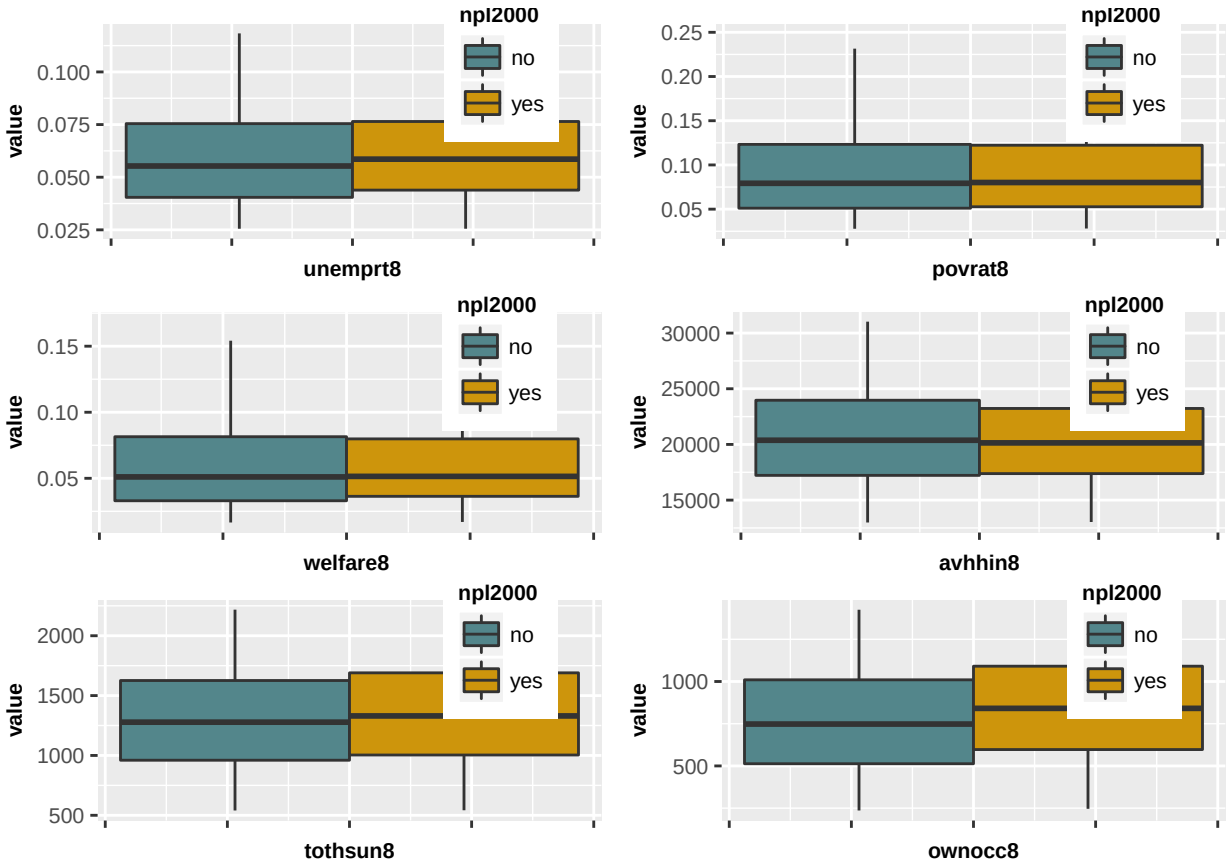

```
pl <- list()

for (i in c(19:24)) {

  pl[[i-18]] <- ggplot(all_covariates, aes_string(x = i,y=covariates[i], fill="npl2000")) +
  geom_boxplot(outlier.shape=NA)  + ylab(label = "value") + xlab(label = covariates[i]) +
  theme(axis.text.x=element_blank(),axis.text=element_text(size=8),
        axis.title=element_text(size=8,face="bold"),
        legend.text=element_text(size=8),legend.title=element_text(size=8,face="bold"),
        legend.position = c(0.8,.8)) +
  scale_fill_manual(values = c("cadetblue4", "darkgoldenrod3"),labels = c("no", "yes")) +
  guides(colour = guide_legend(override.aes = list(size=10))) +
    scale_y_continuous(limits = quantile(all_covariates[,covariates[i]], c(0.1, 0.9)))

}

suppressWarnings(grid.arrange(grobs = pl))
```




```
pl <- list()

for (i in c(25:30)) {

  pl[[i-24]] <- ggplot(all_covariates, aes_string(x = i,y=covariates[i], fill="npl2000")) +
  geom_boxplot(outlier.shape=NA)  + ylab(label = "value") + xlab(label = covariates[i]) +
  theme(axis.text.x=element_blank(),axis.text=element_text(size=8),
        axis.title=element_text(size=8,face="bold"),
        legend.text=element_text(size=8),legend.title=element_text(size=8,face="bold"),
        legend.position = c(0.8,.8)) +
  scale_fill_manual(values = c("cadetblue4", "darkgoldenrod3"),labels = c("no", "yes")) +
  guides(colour = guide_legend(override.aes = list(size=10))) +
    scale_y_continuous(limits = quantile(all_covariates[,covariates[i]], c(0.1, 0.9)))

}

suppressWarnings(grid.arrange(grobs = pl))
```




```
pl <- list()

for (i in c(31:36)) {

  pl[[i-30]] <- ggplot(all_covariates, aes_string(x = i,y=covariates[i], fill="npl2000")) +
  geom_boxplot(outlier.shape=NA)  + ylab(label = "value") + xlab(label = covariates[i]) +
  theme(axis.text.x=element_blank(),axis.text=element_text(size=8),
        axis.title=element_text(size=8,face="bold"),
        legend.text=element_text(size=8),legend.title=element_text(size=8,face="bold"),
        legend.position = c(0.8,.8)) +
  scale_fill_manual(values = c("cadetblue4", "darkgoldenrod3"),labels = c("no", "yes")) +
  guides(colour = guide_legend(override.aes = list(size=10))) +
    scale_y_continuous(limits = quantile(all_covariates[,covariates[i]], c(0.1, 0.9)))

}

suppressWarnings(grid.arrange(grobs = pl))
```




```
site_covariates$above_285 <- ifelse(site_covariates$hrs_82>28.5,1,0)

sitec.vars <- attributes(site_covariates)$var.labels

range_285 <- site_covariates %>% group_by(above_285) %>% summarise_all(mean, na.rm=T) %>% t() %>%
  round(3)
```

```
## Warning: package 'bindrcpp' was built under R version 3.4.4
```

```
range_285 <- cbind(c("HRS test result of 28.5 or greater before 1982",sitec.vars),range_285) %>% as.data

kable(range_285, caption = "Mean Values of Covariates for tracts that had an HRS test in 1982 with a val
```

Table 1: Mean Values of Covariates for tracts that had an HRS test in 1982 with a value greater than 28.5 (treated) and those with a value less than or equal to 28.5 (control)

	Variable	Control	Treated
above_285	HRS test result of 28.5 or greater before 1982	0	1
hrs_82		15.538	44.468
npl1990	NPL Before 1990	0.127	0.99
npl2000	NPL Before 2000	0.16	0.99
pop_den8	Tract Pop Density 1980	1670.241	1157.464
shrblk8	Share of Pop Black 1980	0.113	0.071
shrhsp8	Share of Pop Hispanic 1980	0.044	0.042
child8	Share of Pop < 18 1980	0.293	0.294
shrfor8	Share of Pop Forgn Born 1980	0.053	0.051

	Variable	Control	Treated
ffh8	Share of Female Headed HHs 80	0.188	0.158
smhse8	% Pop in Same House 5 Yrs Ago 80	0.603	0.562
hsdrop8	% Pop 16 <= <= 19 NO HS Diploma 80	0.145	0.134
no_hs_diploma8	% Pop > 25 NO HS Diploma 80	0.405	0.343
ba_or_better8	% Pop > 25 BA or Better 80	0.1	0.138
unemprt8	% Pop > 16 Unemployed 80	0.086	0.071
povrat8	% Pop < Poverty Line 80	0.114	0.1
welfare8	% HHolds Pub Ass Recipients 80	0.089	0.074
favinc8	average family income 80	21657.675	23041.604
avhhin8	average HHold Income 80	19635.32	20868.854
meanrnt80	Mean House Rent 1980	247.313	268.205
mdvalhs9	Median Housing Price 1990	80185.072	96752.082
meanrnt9	Mean House Rent 1990	470.387	527.323
mdvalhs0	Median Housing Price 2000	115479.006	135435.621
meanrnt0	Mean House Rent 2000	609.843	690.646
tothsun8	Total Housing Units in Tract 80	1356.74	1352.81
ownocc8	number of owner occupied housing units	906.37	901.853
owner_occupied80	% Housing Units Owner Occ 80	0.679	0.68
bltlast5yrs80	% Housing Units Built < 5 Yrs 80	0.119	0.14
bltlast10yrs80	% Housing Units Built < 10 Yrs 80	0.237	0.281
firestoveheat80	% Units Fire Stove Port No Heat 80	0.053	0.052
noaircond80	% Housing Units No AC 80	0.506	0.48
nofullkitchen80	% Housing Units No Full Kitchen 80	0.021	0.019
zerofullbath80	% Housing Units No Full Bath 80	0.032	0.026
northeast	=1 if Tract in Northeast Region	0.331	0.477
midwest	=1 if Tract in Midwest Region	0.348	0.225
south	=1 if Tract in South Region	0.215	0.193
west	=1 if Tract in West Region	0.105	0.105
meanhs8	Mean Housing Prices 1980	45027.338	52136.801
bedrms02_80	% Housing Units 0-2 Bedrooms 80	0.469	0.444
bedrms34_80	% Housing Units 3-4 Bedrooms 80	0.51	0.529
detach80	percent of detached housing units 1980	0.658	0.69
bedrms0_80occ	percent owner-occupied housing units 0 bedrooms 1980	0.002	0.002
bedrms1_80occ	percent owner-occupied housing units 1 bedrooms 1980	0.046	0.044
bedrms2_80occ		0.308	0.269
bedrms3_80occ		0.483	0.488
bedrms4_80occ		0.134	0.163
bedrms5_80occ		0.027	0.033
blt0_1yrs80occ	percent owner-occupied housing units built in last year 1980	0.027	0.033
blt2_5yrs80occ	percent owner-occupied housing units 2-5 years old in 1980	0.089	0.104
blt6_10yrs80occ		0.109	0.13
blt10_20yrs80occ		0.178	0.203
blt20_30yrs80occ		0.185	0.193
blt30_40yrs80occ		0.113	0.098
blt40_yrs80occ		0.299	0.238
detach80occ	percent detached single family housing 1980	0.858	0.891
attach80occ	percent attached single family housing 1980	0.06	0.031
mobile80occ	percent mobile home single family housing 1980	0.081	0.078
occupied80	% housing units that are occupied 1980	0.941	0.941
bltmore30_80		0.409	0.333
og82list	=1 if 1 of the sites was tested by EPA by 1982	1	1
nbr_dummy	=1 if census tract nbrs a census tract with a hazardous waste site	0.204	0.337

```
site_covariates$above_btwn <- ifelse((site_covariates$hrs_82>28.5) &
                                     (site_covariates$hrs_82<=40.5),1,
                                     ifelse((site_covariates$hrs_82<=28.5) &
                                              (site_covariates$hrs_82>=16.5),2,0))

sitec.vars <- attributes(site_covariates)$var.labels

range_285 <- site_covariates %>% group_by(above_btwn) %>% summarise_all(mean, na.rm=T) %>% t() %>%
  round(3)

range_285 <- cbind(c("HRS test result of 28.5 or greater before 1982",sitec.vars),range_285)%>%
  as.data.frame() %>% set_colnames(c("Variable","Control","Treated","not included"))
```

```
## Warning in cbind(c("HRS test result of 28.5 or greater before 1982",
## sitec.vars), : number of rows of result is not a multiple of vector length
## (arg 1)
```

```
kable(range_285, caption = "Mean Values of Covariates for tracts that had an HRS test in 1982 with a va
```

Table 2: Mean Values of Covariates for tracts that had an HRS test in 1982 with a value greater than 28.5 (treated) and those with a value less than or equal to 28.5 (control)

	Variable	Control	Treated	not incl
above_btwn	HRS test result of 28.5 or greater before 1982	0	1	2
hrs_82		36.975	34.89	22.513
npl1990	NPL Before 1990	0.658	0.985	0.222
npl2000	NPL Before 2000	0.665	0.985	0.267
pop_den8	Tract Pop Density 1980	1447.394	1151.05	1360.90
shrblk8	Share of Pop Black 1980	0.09	0.084	0.082
shrhsp8	Share of Pop Hispanic 1980	0.054	0.03	0.031
child8	Share of Pop < 18 1980	0.295	0.293	0.288
shrfor8	Share of Pop Forgn Born 1980	0.062	0.039	0.04
ffh8	Share of Female Headed HHs 80	0.172	0.166	0.164
smhse8	% Pop in Same House 5 Yrs Ago 80	0.581	0.565	0.585
hsdrop8	% Pop 16 <= <= 19 NO HS Diploma 80	0.136	0.134	0.148
no_hs_diploma8	% Pop > 25 NO HS Diploma 80	0.365	0.353	0.388
ba_or_better8	% Pop > 25 BA or Better 80	0.123	0.134	0.109
unemprt8	% Pop > 16 Unemployed 80	0.079	0.073	0.075
povrat8	% Pop < Poverty Line 80	0.102	0.112	0.107
welfare8	% HHolds Pub Ass Recipients 80	0.082	0.075	0.081
favinc8	average family income 80	22740.116	22587.848	21820.05
avhhin8	average HHold Income 80	20675.211	20300.79	19812.21
meanrnt80	Mean House Rent 1980	265.214	256.727	252.898
mdvalhs9	Median Housing Price 1990	92181.954	91611.453	84461.75
meanrnt9	Mean House Rent 1990	529.153	481.591	477.147
mdvalhs0	Median Housing Price 2000	134029.231	123502.92	117527.7
meanrnt0	Mean House Rent 2000	689.631	621.767	634.392
tothsun8	Total Housing Units in Tract 80	1368.3	1319.328	1366.93
ownocc8	number of owner occupied housing units	905.538	871.533	946.444
owner_occupied80	% Housing Units Owner Occ 80	0.678	0.673	0.694
bltlast5yrs80	% Housing Units Built < 5 Yrs 80	0.127	0.14	0.137
bltlast10yrs80	% Housing Units Built < 10 Yrs 80	0.258	0.276	0.267
firestoveheat80	% Units Fire Stove Port No Heat 80	0.048	0.055	0.061

	Variable	Control	Treated	not incl
noaircond80	% Housing Units No AC 80	0.47	0.51	0.516
nofullkitchen80	% Housing Units No Full Kitchen 80	0.017	0.022	0.023
zerofullbath80	% Housing Units No Full Bath 80	0.026	0.029	0.034
northeast	=1 if Tract in Northeast Region	0.435	0.423	0.389
midwest	=1 if Tract in Midwest Region	0.246	0.285	0.322
south	=1 if Tract in South Region	0.204	0.204	0.189
west	=1 if Tract in West Region	0.115	0.088	0.1
meanhs8	Mean Housing Prices 1980	50049.516	50647.975	46135.1
bedrms02_80	% Housing Units 0-2 Bedrooms 80	0.45	0.45	0.467
bedrms34_80	% Housing Units 3-4 Bedrooms 80	0.527	0.521	0.509
detach80	percent of detached housing units 1980	0.67	0.685	0.693
bedrms0_80occ	percent owner-occupied housing units 0 bedrooms 1980	0.002	0.003	0.002
bedrms1_80occ	percent owner-occupied housing units 1 bedrooms 1980	0.043	0.046	0.046
bedrms2_80occ		0.28	0.275	0.308
bedrms3_80occ		0.496	0.48	0.469
bedrms4_80occ		0.151	0.162	0.144
bedrms5_80occ		0.029	0.035	0.03
blt0_1yrs80occ	percent owner-occupied housing units built in last year 1980	0.031	0.031	0.031
blt2_5yrs80occ	percent owner-occupied housing units 2-5 years old in 1980	0.095	0.102	0.103
blt6_10yrs80occ		0.121	0.126	0.12
blt10_20yrs80occ		0.194	0.196	0.19
blt20_30yrs80occ		0.188	0.193	0.192
blt30_40yrs80occ		0.105	0.1	0.108
blt40_yrs80occ		0.267	0.251	0.255
detach80occ	percent detached single family housing 1980	0.881	0.89	0.855
attach80occ	percent attached single family housing 1980	0.044	0.032	0.051
mobile80occ	percent mobile home single family housing 1980	0.075	0.079	0.094
occupied80	% housing units that are occupied 1980	0.941	0.941	0.941
bltmore30_80		0.364	0.354	0.366
og82list	=1 if 1 of the sites was tested by EPA by 1982	1	1	1
nbr_dummy	=1 if census tract nbrs a census tract with a hazardous waste site	0.3	0.292	0.244
above_285	HRS test result of 28.5 or greater before 1982	0.65	1	0

We can see from these above comparisons that with the NPL 2000 treatment vs. control comparison the hosuing values in 2000 tend to be 8-10% higher in NPL tracts than those with non-NPLs, as expected. Most of the covariates are significantly different between treatment vs. control, which is not ideal. However, the differences are quite small. NPL tracts tend to be less population dense and had families making slightly lower incomes in 1980. The 1982 HRS test result comparisons are less clear, with incomes and housing prices actually looking to be somewhat higher in tracts that had a high 1982 HRS score compared to those that didn't. Therefore, this suggests that perhaps something occurred to cause this change (hence the need for regression discontinuity).

2a - HRS and regression discontinuity

Regression discontinuity design requires that treatment assignment is "as good as random" at the threshold for treatment. If this holds, then it guarantees that those who just barely received treatment are comparable to those who just barely did not receive treatment, as treatment status is effectively random. Treatment assignment at the threshold can be "as good as random" if there is randomness in the assignment variable and the agents considered individuals cannot perfectly manipulate their treatment status.

Regarding the facts:

i. This fact limits the randomness of the of the assignment variable since the EPA is selecting the number of sites that it can manage to treat. This could suggest that the treated units are not balanced in some way. EPA might have selected the few sites that had the largest potential for benefit. On the other hand, if the threshold was determined for some reason like all HRS scores above 28.5 produce visible or noxious signs of pollution, then there may be other reasons that housing prices might respond to this HRS value other than due to the NPL listing.

ii. This fact should alleviate concerns that there was gaming around the threshold. The fact that none of the officials involved in site identification, testing and HRS scoring knew about the cutoff threshold at the time means that they were not subject to any bias to score sites near the threshold one way or the other. One can imagine a scenario in which someone may want to score a site with an HRS score of say 29 as being 28.5 so that it might be listed if they have a vested interest in that community being aware of an NPL site. Conversely it is reasonable to think that one might not want to subject a community to an NPL listing if it has an HRS score just below the cutoff so as to prevent housing prices from dropping in response to NPL listing. We could also check this issue by evaluating bunching on our own.

iii. This fact could potentially muddy the treatment effect on the home's prices since scores may be off in a way that leads some homes that have really bad pollution not to be treated and homes that don't have that bad of pollution to be treated (when these homes are close to the threshold). The is a reason for fuzzy RD. The fact that the EPA emphasized that the HRS score is an imperfect scoring measure again means that there may be some slight element of randomness to the assignment and that the HRS score does not reflect detectable characteristics of the tracts other than causing them to be NPL listed or not.

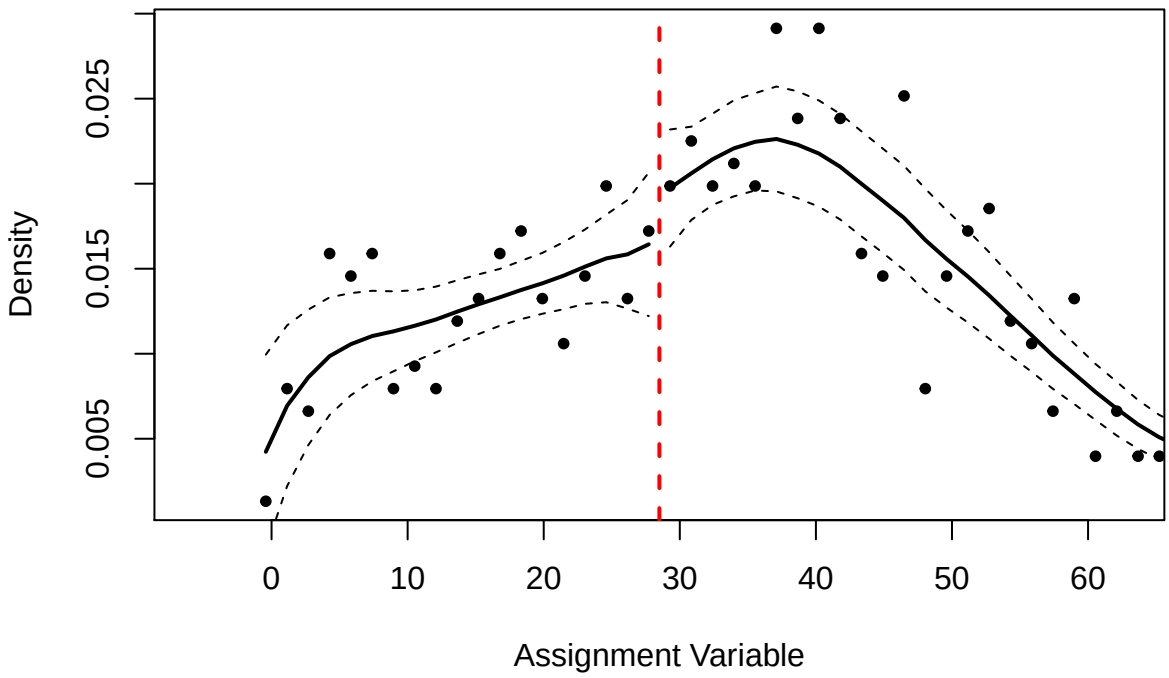
2b - histogram

```
DCdensity(
    runvar      = mile_data$hrs_82,
    cutpoint    = 28.5,
    plot        = TRUE,
    ext.out     = TRUE)
```

```
## $theta
## [1] 0.1463121
##
## $se
## [1] 0.3003333
##
## $z
## [1] 0.4871657
##
## $p
## [1] 0.6261409
##
## $binsize
## [1] 1.563345
##
## $bw
## [1] 12.21568
##
## $cutpoint
## [1] 28.5
##
## $data
```

```
##        cellmp     cellval
## 1  -0.4218786 0.001324336
## 2   1.1414662 0.007946014
## 3   2.7048110 0.006621679
## 4   4.2681558 0.015892029
## 5   5.8315005 0.014567693
## 6   7.3948453 0.015892029
## 7   8.9581901 0.007946014
## 8  10.5215349 0.009270350
## 9  12.0848797 0.007946014
## 10 13.6482245 0.011919022
## 11 15.2115693 0.013243357
## 12 16.7749141 0.015892029
## 13 18.3382589 0.017216365
## 14 19.9016037 0.013243357
## 15 21.4649484 0.010594686
## 16 23.0282932 0.014567693
## 17 24.5916380 0.019865036
## 18 26.1549828 0.013243357
## 19 27.7183276 0.017216365
## 20 29.2816724 0.019865036
## 21 30.8450172 0.022513708
## 22 32.4083620 0.019865036
## 23 33.9717068 0.021189372
## 24 35.5350516 0.019865036
## 25 37.0983963 0.029135386
## 26 38.6617411 0.023838043
## 27 40.2250859 0.029135386
## 28 41.7884307 0.023838043
## 29 43.3517755 0.015892029
## 30 44.9151203 0.014567693
## 31 46.4784651 0.025162379
## 32 48.0418099 0.007946014
## 33 49.6051547 0.014567693
## 34 51.1684995 0.017216365
## 35 52.7318442 0.018540700
## 36 54.2951890 0.011919022
## 37 55.8585338 0.010594686
## 38 57.4218786 0.006621679
## 39 58.9852234 0.013243357
## 40 60.5485682 0.003973007
## 41 62.1119130 0.006621679
## 42 63.6752578 0.003973007
## 43 65.2386026 0.003973007
## 44 66.8019474 0.003973007
## 45 68.3652921 0.002648671
## 46 69.9286369 0.005297343
## 47 71.4919817 0.003973007
## 48 73.0553265 0.003973007
## 49 74.6186713 0.001324336
```

```
title(xlab = 'Assignment Variable', ylab = 'Density')
abline(v = 28.5, col = "red", lty = "dashed", lwd = 2)
```



As we can see in the histogram and in the local linear regression lines, there is a limited discontinuity around the threshold density which suggests we should not be concerned about bunching. Furthermore, the p-value is 0.62 which is further evidence of the insignificance of the discontinuity gap.

3a - 2SLS

First Stage Equation: $d_i = \gamma_0 + \gamma_1 * z_i + \gamma_2 * z_i * (X_i - c) + \gamma_3(X_i - c) + u_i$

Where z_i is the IV, in this case the binary 0/1 variable for whether a tract has an HRS 1982 score above 28.5 (1) or not (0), and d_i is the variable we want to instrument for, in this case the treatment variable of the NPL 2000 listing.

```
mile_data$hrs_82_indicator <- ifelse(mile_data$hrs_82>=28.5,1,0)
mile_data$running <- mile_data$hrs_82 - 28.5
covariates <- colnames(mile_data)
remove <- c("mdvalhs9_nbr","mdvalhs9","mdvalhs0","meanhs8",
            "lnmdvalhs0_nbr","fips","statefips","npl1990","hrs_82", "blt40_yrs80occ_nbr")
covariates <- covariates[! covariates %in% remove]
mile_data_reg <- mile_data[,covariates]
mile_data_reg$interaction <- mile_data_reg$hrs_82_indicator * mile_data_reg$running


reg1 <- lm(npl2000 ~ ., data = mile_data_reg)
summary(reg1)
```

```
##
```

```
## Call:
## lm(formula = npl2000 ~ ., data = mile_data_reg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.96611 -0.07643 -0.00946  0.03896  0.93677
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             5.754e+05  8.341e+05   0.690  0.49069
## tothsun8_nbr           -1.438e-05  1.645e-05  -0.874  0.38239
## occupied80_nbr         -8.407e-01  4.681e-01  -1.796  0.07321 .
## pop_den8_nbr           -1.832e-07  9.261e-06  -0.020  0.98422
## no_hs_diploma8_nbr     -5.041e-01  2.547e-01  -1.979  0.04840 *
## ba_or_better8_nbr      -1.893e-01  3.416e-01  -0.554  0.57978
## shrblk8_nbr             1.359e-01  1.523e-01   0.892  0.37294
## shrhsp8_nbr             2.168e-01  2.012e-01   1.077  0.28185
## child8_nbr             -8.106e-01  5.476e-01  -1.480  0.13953
## old8_nbr               -9.230e-01  4.433e-01  -2.082  0.03791 *
## shrfor8_nbr             2.385e-01  3.484e-01   0.684  0.49410
## ffh8_nbr                1.343e-02  2.839e-01   0.047  0.96230
## smhse8_nbr              6.311e-01  2.197e-01   2.872  0.00428 **
## hsdrop8_nbr            -3.799e-02  2.238e-01  -0.170  0.86528
## unemprt8_nbr           -1.299e-01  5.275e-01  -0.246  0.80555
## povrat8_nbr             5.852e-01  4.891e-01   1.196  0.23219
## welfare8_nbr           -4.721e-02  5.951e-01  -0.079  0.93681
## avhhin8_nbr             1.462e-06  5.270e-06   0.277  0.78154
## zerofullbath80_nbr     -1.448e+00  1.248e+00  -1.160  0.24661
## firestoveheat80_nbr    -2.089e-01  2.400e-01  -0.871  0.38441
## nofullkitchen80_nbr    -2.796e-01  1.408e+00  -0.199  0.84264
## noaircond80_nbr         1.510e-01  6.266e-02   2.409  0.01641 *
## ownocc8_nbr             6.743e-06  8.177e-06   0.825  0.41006
## bedrms0_80occ_nbr       7.880e+04  5.716e+05   0.138  0.89041
## bedrms1_80occ_nbr       7.880e+04  5.716e+05   0.138  0.89041
## bedrms2_80occ_nbr       7.880e+04  5.716e+05   0.138  0.89041
## bedrms3_80occ_nbr       7.880e+04  5.716e+05   0.138  0.89041
## bedrms4_80occ_nbr       7.880e+04  5.716e+05   0.138  0.89041
## bedrms5_80occ_nbr       7.880e+04  5.716e+05   0.138  0.89041
## detach80occ_nbr        -6.542e+05  5.619e+05  -1.164  0.24500
## attach80occ_nbr        -6.542e+05  5.619e+05  -1.164  0.24500
## mobile80occ_nbr        -6.542e+05  5.619e+05  -1.164  0.24500
## blt0_1yrs80occ_nbr     -3.542e-01  7.018e-01  -0.505  0.61400
## blt2_5yrs80occ_nbr      4.773e-01  3.164e-01   1.509  0.13210
## blt6_10yrs80occ_nbr    -1.657e-01  2.620e-01  -0.633  0.52737
## blt10_20yrs80occ_nbr    9.367e-02  1.831e-01   0.512  0.60917
## blt20_30yrs80occ_nbr   -1.373e-01  1.660e-01  -0.827  0.40851
## blt30_40yrs80occ_nbr   -4.876e-01  2.941e-01  -1.658  0.09802 .
## tothsun9_nbr            2.913e-05  3.203e-05   0.909  0.36366
## tothsun0_nbr           -1.798e-05  2.312e-05  -0.778  0.43718
## lnmeanhs8_nbr          -4.400e-02  4.915e-02  -0.895  0.37123
## hrs_82_indicator        6.417e-01  4.110e-02  15.612  < 2e-16 ***
## running                 1.427e-02  2.304e-03   6.194 1.35e-09 ***
## interaction            -1.455e-02  2.585e-03  -5.631 3.20e-08 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.226 on 439 degrees of freedom
## Multiple R-squared:  0.784,  Adjusted R-squared:  0.7628
## F-statistic: 37.05 on 43 and 439 DF,  p-value: < 2.2e-16
```

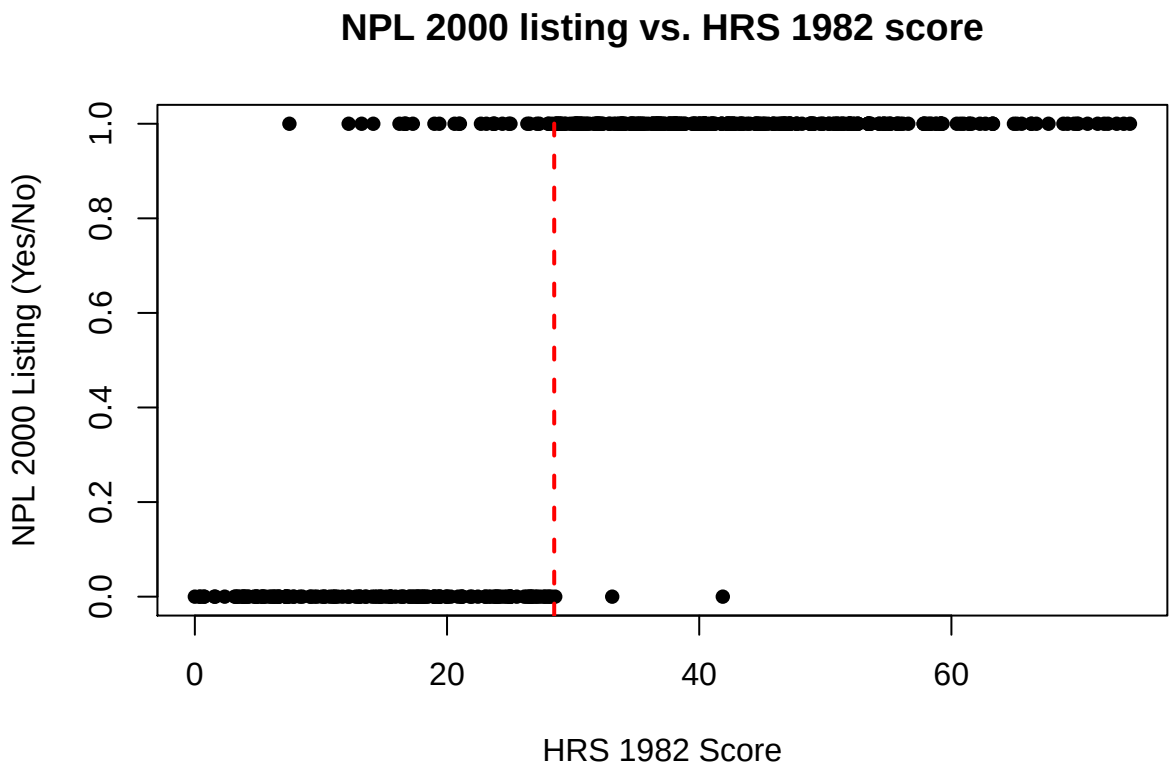
```
mile_data_reg <- mile_data_reg[which(mile_data_reg$running >= -12 & mile_data_reg$running <= 12),]

reg2_1st <- lm(npl2000 ~ ., data = mile_data_reg)
summary(reg2_1st)
```

```
##
## Call:
## lm(formula = npl2000 ~ ., data = mile_data_reg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.88833 -0.16683 -0.02606  0.07462  0.80964
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             3.395e+05  1.627e+06   0.209   0.8350
## tothsun8_nbr           -3.463e-05  4.249e-05  -0.815   0.4161
## occupied80_nbr         -1.307e+00  1.023e+00  -1.277   0.2031
## pop_den8_nbr            5.627e-06  2.100e-05   0.268   0.7891
## no_hs_diploma8_nbr     -4.542e-01  5.194e-01  -0.875   0.3829
## ba_or_better8_nbr      -1.901e-01  6.608e-01  -0.288   0.7740
## shrblk8_nbr            -4.753e-02  3.336e-01  -0.142   0.8869
## shrhsp8_nbr             2.895e-01  4.334e-01   0.668   0.5050
## child8_nbr             -1.378e+00  9.513e-01  -1.448   0.1492
## old8_nbr               -1.128e+00  8.317e-01  -1.356   0.1766
## shrfor8_nbr            -1.868e-01  9.667e-01  -0.193   0.8470
## ffh8_nbr                9.577e-01  8.025e-01   1.193   0.2343
## smhse8_nbr              1.124e+00  4.499e-01   2.499   0.0133 *
## hsdrop8_nbr            -3.023e-01  4.795e-01  -0.630   0.5293
## unemprt8_nbr            4.959e-01  9.659e-01   0.513   0.6083
## povrat8_nbr             9.404e-01  9.958e-01   0.944   0.3462
## welfare8_nbr           -2.973e-02  1.282e+00  -0.023   0.9815
## avhhin8_nbr             1.160e-05  1.091e-05   1.064   0.2889
## zerofullbath80_nbr     -2.835e-01  2.689e+00  -0.105   0.9162
## firestoveheat80_nbr     9.736e-02  5.444e-01   0.179   0.8583
## nofullkitchen80_nbr    -3.536e+00  2.945e+00  -1.201   0.2314
## noaircond80_nbr         2.223e-01  1.210e-01   1.837   0.0679 .
## ownocc8_nbr             1.278e-05  1.655e-05   0.772   0.4409
## bedrms0_80occ_nbr      -4.094e+05  1.084e+06  -0.378   0.7062
## bedrms1_80occ_nbr      -4.094e+05  1.084e+06  -0.378   0.7062
## bedrms2_80occ_nbr      -4.094e+05  1.084e+06  -0.378   0.7062
## bedrms3_80occ_nbr      -4.094e+05  1.084e+06  -0.378   0.7062
## bedrms4_80occ_nbr      -4.094e+05  1.084e+06  -0.378   0.7062
## bedrms5_80occ_nbr      -4.094e+05  1.084e+06  -0.378   0.7062
## detach80occ_nbr         6.993e+04  1.137e+06   0.062   0.9510
## attach80occ_nbr         6.993e+04  1.137e+06   0.062   0.9510
## mobile80occ_nbr         6.993e+04  1.137e+06   0.062   0.9510
## blt0_1yrs80occ_nbr     -1.148e+00  1.560e+00  -0.736   0.4627
## blt2_5yrs80occ_nbr      1.547e+00  7.266e-01   2.130   0.0345 *
```

```
## blt6_10yrs80occ_nbr   2.532e-01  5.691e-01   0.445   0.6569
## blt10_20yrs80occ_nbr  3.191e-01  4.200e-01   0.760   0.4484
## blt20_30yrs80occ_nbr -2.631e-01  3.360e-01  -0.783   0.4347
## blt30_40yrs80occ_nbr -3.060e-01  6.279e-01  -0.487   0.6267
## tothsun9_nbr          6.202e-05  7.720e-05   0.803   0.4228
## tothsun0_nbr         -3.348e-05  5.122e-05  -0.654   0.5142
## lnmeanhs8_nbr        -4.095e-02  9.000e-02  -0.455   0.6496
## hrs_82_indicator      6.287e-01  8.549e-02   7.354 6.33e-12 ***
## running               1.705e-02  9.321e-03   1.830   0.0690 .
## interaction          -1.702e-02  1.191e-02  -1.429   0.1547
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2955 on 182 degrees of freedom
## Multiple R-squared:  0.6629, Adjusted R-squared:  0.5832
## F-statistic: 8.322 on 43 and 182 DF,  p-value: < 2.2e-16
```

3b - graph results

```
plot(mile_data$hrs_82, mile_data$npl2000, main = "NPL 2000 listing vs. HRS 1982 score",
     pch = 16, ylab = "NPL 2000 Listing (Yes/No)", xlab = "HRS 1982 Score")
abline(v = 28.5, lty = 'dashed', col = 2, lwd = 2)
```



The graph shows a clear trend of houses with scores less than 28.5 in 1982 being classified as non-priority

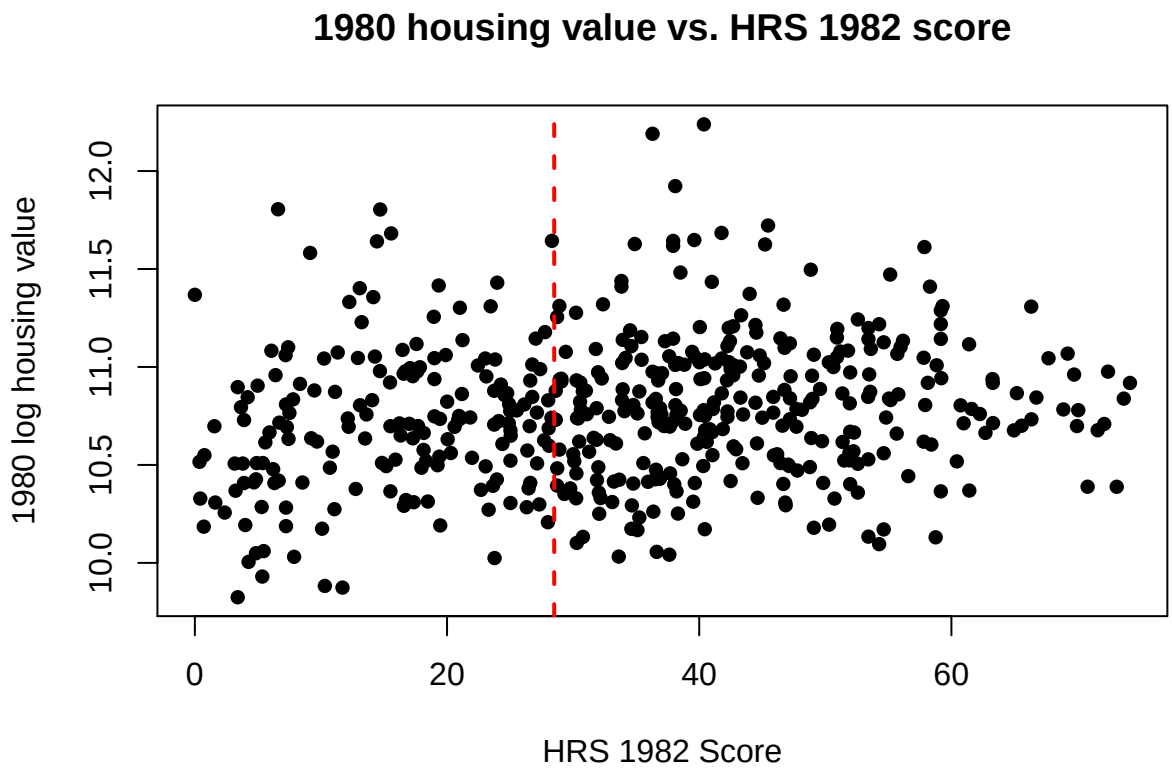
(NPL value of 0) in 2000, while houses with scores greater than 28.5 in 1982 are on the National Priorities List (NPL value of 1) for clean-up. There are also a few houses with scores lower than 28.5 that are on the NPL in 2000 showing that there is an increase in the number of hazardous waste sites that are legally obligated to be cleaned up since 1982, when their initial scores were assigned under the Superfund program.

This shows that there is a fuzzy discontinuity in NPL classification at an HRS score of 28.5 which can be used in a regression discontinuity design study to study the differential effects of hazardous waste material above a certain threshold on housing prices.

The fact that there is a correlation suggests that the first assumption needed to use HRS 82 scores as an IV is valid because clearly the covariance between HRS 82 scores and NPL 2000 listing is not equal to zero, or $Cov(z_i, d_i) \neq 0$.

3c - graph results #2

```
plot(mile_data$hrs_82, mile_data$lnmeanhs8_nbr, main = "1980 housing value vs. HRS 1982 score",
     pch = 16, ylab = "1980 log housing value", xlab = "HRS 1982 Score")
abline(v = 28.5, lty = 'dashed', col = 2, lwd = 2)
```



There is no clear trend between the HRS score in 1982 and mean housing prices in 1980. This shows that ostensibly there may not be a relationship between the two variables prior to the Superfund program, which is desirable because we would not want a covariate to have a regression discontinuity.

4 - More 2SLS

reduced form Equation: $y_i = \pi_0 + \pi_1 z_i + \pi_2(x_i - c)z_i + \pi_2(x_i - c) + \varepsilon_i$

Where y_i is the outcome variable, in this case the difference in log housing value between 1980 and 2000, and z_i is the instrument, and x_i are the rest of the exogenous covariates.

In the second stage, we regress the outcome variable on predicted values from the first stage (x_star) and all the exogenous variables in the model.

The 2 standard assumptions for a valid IV are as follows:

1. It is uncorrelated with the disturbance - Cov(instrument, error) = 0
2. In the linear projection of the endogenous variable on all exogenous variables and the instrument, the coefficient on the instrument is not equal to 0. This implies that the only way the instrument affects the outcome variable is through the endogenous variable.

The third condition is a bit redundant, but we will state it anyway. The instrument is not one of the other variables in the original estimation equation.

```
covariates <- colnames(mile_data)
remove <- c("mdvalhs9_nbr","mdvalhs9","mdvalhs0","meanhs8",
            "fips","statefips","npl1990","hrs_82","npl2000", "blt40_yrs80occ_nbr")
covariates <- covariates[! covariates %in% remove]
mile_data_2sls <- mile_data[,covariates]
mile_data_2sls$interaction <- mile_data_2sls$hrs_82_indicator * mile_data_2sls$running

reg1 <- lm(lnmdvalhs0_nbr ~ ., data = mile_data_2sls)
summary(reg1)
```

```
##
## Call:
## lm(formula = lnmdvalhs0_nbr ~ ., data = mile_data_2sls)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.71261 -0.11227  0.01351  0.11088  0.75039
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -8.793e+05  7.337e+05  -1.199 0.231347
## tothsun8_nbr          3.489e-05  1.446e-05   2.412 0.016285 *
## occupied80_nbr        1.419e+00  4.117e-01   3.447 0.000621 ***
## pop_den8_nbr         -2.024e-06  8.146e-06  -0.248 0.803882
## no_hs_diploma8_nbr   -1.611e-01  2.240e-01  -0.719 0.472538
## ba_or_better8_nbr     3.391e-01  3.005e-01   1.128 0.259743
## shrblk8_nbr          -3.247e-01  1.340e-01  -2.423 0.015803 *
## shrhsp8_nbr          -3.875e-01  1.770e-01  -2.190 0.029071 *
## child8_nbr           -1.077e+00  4.817e-01  -2.236 0.025879 *
## old8_nbr             -8.285e-01  3.899e-01  -2.125 0.034159 *
## shrfor8_nbr           6.906e-01  3.065e-01   2.253 0.024730 *
## ffh8_nbr              6.379e-01  2.497e-01   2.555 0.010965 *
## smhse8_nbr            5.400e-01  1.933e-01   2.794 0.005433 **
## hsdrop8_nbr           4.968e-01  1.968e-01   2.524 0.011954 *
## unemprt8_nbr         -1.962e+00  4.640e-01  -4.230 2.85e-05 ***
## povrat8_nbr          -1.137e-02  4.302e-01  -0.026 0.978917
## welfare8_nbr          2.074e+00  5.234e-01   3.962 8.67e-05 ***
```

```
## avhhin8_nbr           4.220e-05  4.635e-06   9.106  < 2e-16 ***
## zerofullbath80_nbr    3.911e-01  1.098e+00   0.356 0.721871
## firestoveheat80_nbr   4.655e-01  2.111e-01   2.205 0.027957 *
## nofullkitchen80_nbr  -1.104e+00  1.238e+00  -0.892 0.373033
## noaircond80_nbr       2.249e-01  5.512e-02   4.081 5.32e-05 ***
## ownocc8_nbr          -2.552e-05  7.192e-06  -3.548 0.000430 ***
## bedrms0_80occ_nbr     7.098e+05  5.027e+05   1.412 0.158690
## bedrms1_80occ_nbr     7.098e+05  5.027e+05   1.412 0.158685
## bedrms2_80occ_nbr     7.098e+05  5.027e+05   1.412 0.158685
## bedrms3_80occ_nbr     7.098e+05  5.027e+05   1.412 0.158685
## bedrms4_80occ_nbr     7.098e+05  5.027e+05   1.412 0.158686
## bedrms5_80occ_nbr     7.098e+05  5.027e+05   1.412 0.158685
## detach80occ_nbr       1.695e+05  4.943e+05   0.343 0.731742
## attach80occ_nbr       1.695e+05  4.943e+05   0.343 0.731743
## mobile80occ_nbr       1.695e+05  4.943e+05   0.343 0.731742
## blt0_1yrs80occ_nbr    2.536e-01  6.173e-01   0.411 0.681438
## blt2_5yrs80occ_nbr    2.957e-02  2.783e-01   0.106 0.915414
## blt6_10yrs80occ_nbr   3.502e-01  2.305e-01   1.520 0.129321
## blt10_20yrs80occ_nbr -2.514e-01  1.610e-01  -1.561 0.119231
## blt20_30yrs80occ_nbr -8.850e-02  1.460e-01  -0.606 0.544665
## blt30_40yrs80occ_nbr -2.597e-01  2.587e-01  -1.004 0.315906
## tothsun9_nbr         -1.087e-04  2.817e-05  -3.858 0.000132 ***
## tothsun0_nbr          8.701e-05  2.033e-05   4.280 2.30e-05 ***
## lnmeanhs8_nbr         5.299e-01  4.323e-02  12.257  < 2e-16 ***
## hrs_82_indicator      1.889e-03  3.615e-02   0.052 0.958346
## running              -3.082e-04  2.027e-03  -0.152 0.879211
## interaction           8.251e-04  2.273e-03   0.363 0.716847
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1988 on 439 degrees of freedom
## Multiple R-squared:  0.8156, Adjusted R-squared:  0.7975
## F-statistic: 45.15 on 43 and 439 DF,  p-value: < 2.2e-16
```

```
mile_data_2sls <- mile_data_2sls[which(mile_data_2sls$running >= -12 & mile_data_2sls$running <= 12),]

reg2 <- lm(lnmdvalhs0_nbr ~ ., data = mile_data_2sls)
summary(reg2)
```

```
##
## Call:
## lm(formula = lnmdvalhs0_nbr ~ ., data = mile_data_2sls)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.42902 -0.12521 -0.00043  0.12517  0.49872
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -6.407e+05  1.144e+06  -0.560  0.57618
## tothsun8_nbr         -1.958e-06  2.987e-05  -0.066  0.94782
## occupied80_nbr        2.242e+00  7.193e-01   3.117  0.00212 **
## pop_den8_nbr         -2.904e-06  1.477e-05  -0.197  0.84429
## no_hs_diploma8_nbr   -3.544e-01  3.651e-01  -0.971  0.33297
## ba_or_better8_nbr    -3.221e-01  4.645e-01  -0.693  0.48896
```

```
## shrblk8_nbr          -1.665e-01  2.345e-01  -0.710  0.47858
## shrhsp8_nbr           3.148e-01  3.047e-01   1.033  0.30286
## child8_nbr           -1.305e+00  6.688e-01  -1.951  0.05263 .
## old8_nbr             -9.120e-01  5.847e-01  -1.560  0.12056
## shrfor8_nbr           1.145e+00  6.796e-01   1.685  0.09379 .
## ffh8_nbr              5.876e-03  5.641e-01   0.010  0.99170
## smhse8_nbr            3.145e-01  3.163e-01   0.994  0.32145
## hsdrop8_nbr           1.499e-01  3.371e-01   0.445  0.65712
## unemprt8_nbr         -1.766e+00  6.790e-01  -2.601  0.01007 *
## povrat8_nbr           6.695e-01  7.000e-01   0.956  0.34017
## welfare8_nbr          1.604e+00  9.013e-01   1.780  0.07679 .
## avhhin8_nbr           4.610e-05  7.667e-06   6.012 9.81e-09 ***
## zerofullbath80_nbr   -6.193e-01  1.891e+00  -0.328  0.74359
## firestoveheat80_nbr   8.181e-01  3.827e-01   2.138  0.03389 *
## nofullkitchen80_nbr   3.204e-01  2.071e+00   0.155  0.87721
## noaircond80_nbr       2.116e-01  8.509e-02   2.487  0.01379 *
## ownocc8_nbr          -5.347e-06  1.163e-05  -0.460  0.64632
## bedrms0_80occ_nbr     7.874e+05  7.623e+05   1.033  0.30302
## bedrms1_80occ_nbr     7.874e+05  7.623e+05   1.033  0.30301
## bedrms2_80occ_nbr     7.874e+05  7.623e+05   1.033  0.30301
## bedrms3_80occ_nbr     7.874e+05  7.623e+05   1.033  0.30301
## bedrms4_80occ_nbr     7.874e+05  7.623e+05   1.033  0.30301
## bedrms5_80occ_nbr     7.874e+05  7.623e+05   1.033  0.30301
## detach80occ_nbr      -1.467e+05  7.993e+05  -0.184  0.85458
## attach80occ_nbr      -1.467e+05  7.993e+05  -0.184  0.85458
## mobile80occ_nbr      -1.467e+05  7.993e+05  -0.184  0.85458
## blt0_1yrs80occ_nbr   -1.016e-01  1.097e+00  -0.093  0.92631
## blt2_5yrs80occ_nbr    1.360e-01  5.108e-01   0.266  0.79029
## blt6_10yrs80occ_nbr   1.671e-01  4.001e-01   0.418  0.67666
## blt10_20yrs80occ_nbr -3.040e-01  2.953e-01  -1.030  0.30458
## blt20_30yrs80occ_nbr -1.713e-01  2.362e-01  -0.725  0.46914
## blt30_40yrs80occ_nbr -8.920e-01  4.414e-01  -2.021  0.04478 *
## tothsun9_nbr         -5.527e-05  5.427e-05  -1.018  0.30981
## tothsun0_nbr          6.258e-05  3.601e-05   1.738  0.08392 .
## lnmeanhs8_nbr         3.812e-01  6.327e-02   6.025 9.18e-09 ***
## hrs_82_indicator      5.924e-02  6.010e-02   0.986  0.32562
## running              -1.886e-03  6.553e-03  -0.288  0.77387
## interaction          -6.125e-03  8.375e-03  -0.731  0.46554
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2077 on 182 degrees of freedom
## Multiple R-squared:  0.8051, Adjusted R-squared:  0.759
## F-statistic: 17.48 on 43 and 182 DF,  p-value: < 2.2e-16
```

```
reg2$coefficients[[42]] / reg2_1st$coefficients[[42]]
```

```
## [1] 0.09422333
```

```
mile_data_2sls$fitted <- reg2_1st$fitted.values
covariates <- colnames(mile_data_2sls)
remove <- c("mdvalhs9_nbr","mdvalhs9","mdvalhs0","meanhs8",
            "fips","statefips","npl1990","hrs_82","npl2000", "blt40_yrs80occ_nbr",
            "hrs_82_indicator")
covariates_2sls <- covariates[! covariates %in% remove]
```

```r
mile_data_2sls <- mile_data_2sls[,covariates_2sls]
mile_data_2sls$interaction <- mile_data_2sls$fitted * mile_data_2sls$running

reg_2sls <- lm(lnmdvalhs0_nbr ~ ., data = mile_data_2sls)
summary(reg_2sls)
```

```
##
## Call:
## lm(formula = lnmdvalhs0_nbr ~ ., data = mile_data_2sls)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.43290 -0.12584  0.00115  0.12550  0.50203
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -6.403e+05  1.143e+06  -0.560  0.57601
## tothsun8_nbr          1.896e-06  2.996e-05   0.063  0.94960
## occupied80_nbr        2.377e+00  7.297e-01   3.257  0.00134 **
## pop_den8_nbr         -3.669e-06  1.484e-05  -0.247  0.80509
## no_hs_diploma8_nbr   -3.102e-01  3.681e-01  -0.843  0.40044
## ba_or_better8_nbr    -3.013e-01  4.683e-01  -0.643  0.52076
## shrblk8_nbr          -1.573e-01  2.352e-01  -0.669  0.50433
## shrhsp8_nbr           2.868e-01  3.061e-01   0.937  0.35014
## child8_nbr           -1.162e+00  6.996e-01  -1.660  0.09855 .
## old8_nbr             -7.796e-01  5.995e-01  -1.300  0.19512
## shrfor8_nbr           1.158e+00  6.801e-01   1.702  0.09038 .
## ffh8_nbr             -1.039e-01  5.774e-01  -0.180  0.85741
## smhse8_nbr            1.822e-01  3.249e-01   0.561  0.57570
## hsdrop8_nbr           1.824e-01  3.409e-01   0.535  0.59334
## unemprt8_nbr         -1.841e+00  6.805e-01  -2.705  0.00748 **
## povrat8_nbr           5.459e-01  7.131e-01   0.766  0.44491
## welfare8_nbr          1.610e+00  9.027e-01   1.783  0.07624 .
## avhhin8_nbr           4.480e-05  7.882e-06   5.684 5.16e-08 ***
## zerofullbath80_nbr   -5.486e-01  1.898e+00  -0.289  0.77284
## firestoveheat80_nbr   7.987e-01  3.840e-01   2.080  0.03891 *
## nofullkitchen80_nbr   7.186e-01  2.114e+00   0.340  0.73430
## noaircond80_nbr       1.874e-01  8.673e-02   2.160  0.03205 *
## ownocc8_nbr          -7.144e-06  1.162e-05  -0.615  0.53954
## bedrms0_80occ_nbr     8.217e+05  7.658e+05   1.073  0.28472
## bedrms1_80occ_nbr     8.217e+05  7.658e+05   1.073  0.28471
## bedrms2_80occ_nbr     8.217e+05  7.658e+05   1.073  0.28471
## bedrms3_80occ_nbr     8.217e+05  7.658e+05   1.073  0.28471
## bedrms4_80occ_nbr     8.217e+05  7.658e+05   1.073  0.28471
## bedrms5_80occ_nbr     8.217e+05  7.658e+05   1.073  0.28471
## detach80occ_nbr      -1.814e+05  8.016e+05  -0.226  0.82118
## attach80occ_nbr      -1.814e+05  8.016e+05  -0.226  0.82118
## mobile80occ_nbr      -1.814e+05  8.016e+05  -0.226  0.82119
## blt0_1yrs80occ_nbr   -4.033e-03  1.108e+00  -0.004  0.99710
## blt2_5yrs80occ_nbr   -3.396e-02  5.307e-01  -0.064  0.94905
## blt6_10yrs80occ_nbr   1.320e-01  4.048e-01   0.326  0.74482
## blt10_20yrs80occ_nbr -3.352e-01  2.984e-01  -1.124  0.26269
## blt20_30yrs80occ_nbr -1.476e-01  2.375e-01  -0.622  0.53499
## blt30_40yrs80occ_nbr -8.513e-01  4.419e-01  -1.926  0.05560 .
```

```
## tothsun9_nbr         -6.311e-05  5.451e-05  -1.158  0.24848
## tothsun0_nbr          6.738e-05  3.615e-05   1.864  0.06394 .
## lnmeanhs8_nbr         3.826e-01  6.354e-02   6.021 9.36e-09 ***
## running              -5.467e-03  8.819e-03  -0.620  0.53610
## interaction          -1.591e-03  9.652e-03  -0.165  0.86926
## fitted                1.068e-01  9.374e-02   1.139  0.25613
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2079 on 182 degrees of freedom
## Multiple R-squared:  0.8048, Adjusted R-squared:  0.7587
## F-statistic: 17.45 on 43 and 182 DF,  p-value: < 2.2e-16
```

When running the two regressions separately, we calculate a treatment effect of 0.09. While the number is slightly positive, the results are insignificant. This implies that housing prices in 2000 are insignificantly affected by the NPL 2000 status.

When running the 2sls regression completely, the point estimate was very close, at 0.1, but also insignificant.

5 - Summary

We find that a simple OLS regression of housing prices on NPL 2000 status reveals a significant positive relationship when none of the control variables are included in the model. The relationship becomes less significant on adding the housing and demographic characteristics to the equation. We try to resolve ommitted variable bias using an IV model as suggested, using the HRS score in 1982 as an instrument for NPL 2000 status. This reveals an insignificant relationship between pollution levels indicated by inclusion on the NPL 2000 list and housing prices. The goal of the paper is to test the Market Willingness to Pay (WTP) for hazardous waste site cleanup using housing prices, and the results of the IV approach reveal that home-owners do not yet value the property on the basis of pollution levels or hazardous waste cleanup status.