Econometrics3_PS2_Q5

Yuhuan Huang

2025-04-18

```
# Q1

#a
#regression1
model1 <- lm_robust(lnmdvalhs0 ~ npl2000 + lnmeanhs8, data = all_sites_data)
summary(model1, se="hetero")
```

```
##
## Call:
## lm_robust(formula = lnmdvalhs0 ~ npl2000 + lnmeanhs8, data = all_sites_data)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper    DF
## (Intercept)  2.38976    0.13295  17.975 6.017e-72  2.12918  2.65034 38404
## npl2000      0.03451    0.01233   2.799 5.136e-03  0.01034  0.05868 38404
## lnmeanhs8    0.85707    0.01229  69.756 0.000e+00  0.83299  0.88115 38404
##
## Multiple R-squared:  0.5802 ,    Adjusted R-squared:  0.5802
## F-statistic:  2434 on 2 and 38404 DF,  p-value: < 2.2e-16
```

```
#coeftest(model1, vcov = vcovHC(model1, type = "HC1"))


#regression2
all_vars_sites_data <- names(all_sites_data)
basic_character_list <- c("firestoveheat80", "noaircond80", "nofullkitchen80", "zerofullbath80",
                          "detach80occ", "mobile80occ")
temp_vars <- grep("^(bedrms|blt)", all_vars_sites_data, value = TRUE)
housing_character_list <- c("npl2000", "lnmeanhs8", basic_character_list, as.list(temp_vars))
# note: need to drop one of the blt* to avoid perfect colinearity!!
housing_character_list <- housing_character_list[housing_character_list != "blt40_yrs80occ"]
housing_character_list <- housing_character_list[housing_character_list != "bedrms5_80occ"]
model_formula <- as.formula(paste("lnmdvalhs0", "~", paste(housing_character_list, collapse = " + ")))
model2 <- lm_robust(model_formula, data = all_sites_data)
summary(model2, se="hetero")
```

```
##
## Call:
## lm_robust(formula = model_formula, data = all_sites_data)
```

```
##
## Standard error type:  HC2
##
## Coefficients:
##                   Estimate Std. Error  t value  Pr(>|t|) CI Lower  CI Upper
## (Intercept)       3.339446   0.239476  13.9448 4.337e-44  2.87007  3.808826
## npl2000           0.037821   0.011937   3.1683 1.534e-03  0.01442  0.061218
## lnmeanhs8         0.864394   0.019237  44.9335 0.000e+00  0.82669  0.902100
## firestoveheat80  -0.031829   0.020991  -1.5163 1.295e-01 -0.07297  0.009315
## noaircond80       0.320326   0.007759  41.2825 0.000e+00  0.30512  0.335534
## nofullkitchen80  -1.610169   0.178137  -9.0390 1.654e-19 -1.95932 -1.261017
## zerofullbath80    0.670265   0.142780   4.6944 2.683e-06  0.39041  0.950119
## detach80occ      -0.265036   0.014075 -18.8305 9.632e-79 -0.29262 -0.237449
## mobile80occ      -0.264704   0.029169  -9.0748 1.192e-19 -0.32188 -0.207532
## bedrms0_80occ    -0.591451   0.231608  -2.5537 1.066e-02 -1.04541 -0.137493
## bedrms1_80occ    -0.335579   0.083276  -4.0297 5.594e-05 -0.49880 -0.172357
## bedrms2_80occ    -0.955865   0.069996 -13.6560 2.335e-42 -1.09306 -0.818672
## bedrms3_80occ    -1.006031   0.071486 -14.0732 7.182e-45 -1.14614 -0.865917
## bedrms4_80occ    -0.564611   0.077558  -7.2799 3.405e-13 -0.71663 -0.412596
## blt0_1yrs80occ   -0.140391   0.045749  -3.0687 2.151e-03 -0.23006 -0.050721
## blt2_5yrs80occ   -0.168493   0.031389  -5.3678 8.014e-08 -0.23002 -0.106969
## blt6_10yrs80occ  -0.126767   0.024089  -5.2624 1.429e-07 -0.17398 -0.079551
## blt10_20yrs80occ -0.064554   0.019543  -3.3032 9.568e-04 -0.10286 -0.026249
## blt20_30yrs80occ -0.005175   0.020196  -0.2562 7.978e-01 -0.04476  0.034410
## blt30_40yrs80occ -0.071157   0.026816  -2.6535 7.969e-03 -0.12372 -0.018597
##                     DF
## (Intercept)      38387
## npl2000          38387
## lnmeanhs8        38387
## firestoveheat80  38387
## noaircond80      38387
## nofullkitchen80  38387
## zerofullbath80   38387
## detach80occ      38387
## mobile80occ      38387
## bedrms0_80occ    38387
## bedrms1_80occ    38387
## bedrms2_80occ    38387
## bedrms3_80occ    38387
## bedrms4_80occ    38387
## blt0_1yrs80occ   38387
## blt2_5yrs80occ   38387
## blt6_10yrs80occ  38387
## blt10_20yrs80occ 38387
## blt20_30yrs80occ 38387
## blt30_40yrs80occ 38387
##
## Multiple R-squared:  0.6477 ,    Adjusted R-squared:  0.6475
## F-statistic:  1623 on 19 and 38387 DF,  p-value: < 2.2e-16
```

```
#coeftest(model2, vcov = vcovHC(model2, type = "HC1"))

#regression3
Economic_population_list <- c("pop_den8", "shrblk8", "shrhsp8", "child8", "old8",
```

```
                              "shrfor8", "ffh8", "smhse8", "hsdrop8", "unemprt8",
                              "povrat8", "welfare8", "avhhin8", "tothsun8", "ownocc8",
                              "occupied80")
temp_vars_2 <- grep("^(no_hs_dipl|ba_or_b)", all_vars_sites_data, value = TRUE)
combined_list <- c(housing_character_list,Economic_population_list,as.list(temp_vars_2))
model_formula2 <- as.formula(paste("lnmdvalhs0", "~", paste(combined_list, collapse = " + ")))
model3 <- lm_robust(model_formula2, data = all_sites_data)
summary(model3, se="hetero")
```

```
##
## Call:
## lm_robust(formula = model_formula2, data = all_sites_data)
##
## Standard error type:  HC2
##
## Coefficients:
##                     Estimate Std. Error  t value   Pr(>|t|)   CI Lower
## (Intercept)       5.460e+00  2.285e-01  23.8896 3.235e-125  5.012e+00
## npl2000           6.986e-02  1.002e-02   6.9710  3.197e-12  5.022e-02
## lnmeanhs8         5.592e-01  2.098e-02  26.6510 4.528e-155  5.181e-01
## firestoveheat80  -1.462e-02  2.229e-02  -0.6561  5.118e-01 -5.831e-02
## noaircond80       4.369e-01  9.287e-03  47.0483  0.000e+00  4.187e-01
## nofullkitchen80  -4.796e-01  1.608e-01  -2.9817  2.868e-03 -7.948e-01
## zerofullbath80    4.565e-01  1.250e-01   3.6520  2.606e-04  2.115e-01
## detach80occ       3.683e-03  1.394e-02   0.2642  7.917e-01 -2.365e-02
## mobile80occ       2.793e-01  2.748e-02  10.1657  3.027e-24  2.255e-01
## bedrms0_80occ    -6.643e-01  2.340e-01  -2.8384  4.537e-03 -1.123e+00
## bedrms1_80occ    -1.546e-01  8.227e-02  -1.8795  6.018e-02 -3.159e-01
## bedrms2_80occ    -3.775e-01  6.185e-02  -6.1044  1.042e-09 -4.988e-01
## bedrms3_80occ    -5.145e-01  5.988e-02  -8.5930  8.789e-18 -6.319e-01
## bedrms4_80occ    -4.606e-01  7.156e-02  -6.4368  1.235e-10 -6.009e-01
## blt0_1yrs80occ    1.255e-01  4.868e-02   2.5788  9.917e-03  3.012e-02
## blt2_5yrs80occ    1.559e-01  2.960e-02   5.2676  1.390e-07  9.790e-02
## blt6_10yrs80occ   1.233e-01  2.308e-02   5.3433  9.178e-08  7.809e-02
## blt10_20yrs80occ  6.930e-02  1.660e-02   4.1743  2.995e-05  3.676e-02
## blt20_30yrs80occ  1.559e-02  1.528e-02   1.0199  3.078e-01 -1.437e-02
## blt30_40yrs80occ  2.852e-02  2.583e-02   1.1043  2.695e-01 -2.210e-02
## pop_den8          7.708e-06  4.077e-07  18.9085  2.232e-79  6.909e-06
## shrblk8          -1.111e-01  1.381e-02  -8.0409  9.170e-16 -1.381e-01
## shrhsp8          -2.886e-01  2.087e-02 -13.8296  2.149e-43 -3.295e-01
## child8           -5.503e-01  5.801e-02  -9.4858  2.539e-21 -6.640e-01
## old8             -3.463e-01  4.786e-02  -7.2349  4.744e-13 -4.401e-01
## shrfor8           1.224e+00  4.564e-02  26.8177 5.662e-157  1.135e+00
## ffh8             -3.821e-02  3.578e-02  -1.0681  2.855e-01 -1.083e-01
## smhse8            4.271e-01  2.402e-02  17.7808  1.910e-70  3.800e-01
## hsdrop8           5.810e-03  2.570e-02   0.2260  8.212e-01 -4.457e-02
## unemprt8         -1.235e+00  7.559e-02 -16.3364  8.661e-60 -1.383e+00
## povrat8          -3.926e-01  5.162e-02  -7.6051  2.911e-14 -4.937e-01
## welfare8          8.426e-01  6.984e-02  12.0642  1.877e-33  7.057e-01
## avhhin8           1.295e-05  7.375e-07  17.5614  9.046e-69  1.151e-05
## tothsun8          1.096e-05  6.924e-06   1.5834  1.133e-01 -2.608e-06
## ownocc8          -1.344e-04  1.104e-05 -12.1717  5.081e-34 -1.561e-04
## occupied80        1.171e-01  4.715e-02   2.4829  1.303e-02  2.466e-02
```

```
## no_hs_diploma8   -3.011e-01  3.618e-02  -8.3208  9.025e-17 -3.720e-01
## ba_or_better8     4.910e-01  3.557e-02  13.8037  3.074e-43  4.213e-01
##                    CI Upper    DF
## (Intercept)       5.908e+00 38369
## npl2000           8.951e-02 38369
## lnmeanhs8         6.004e-01 38369
## firestoveheat80   2.906e-02 38369
## noaircond80       4.552e-01 38369
## nofullkitchen80  -1.643e-01 38369
## zerofullbath80    7.015e-01 38369
## detach80occ       3.101e-02 38369
## mobile80occ       3.332e-01 38369
## bedrms0_80occ    -2.056e-01 38369
## bedrms1_80occ     6.624e-03 38369
## bedrms2_80occ    -2.563e-01 38369
## bedrms3_80occ    -3.972e-01 38369
## bedrms4_80occ    -3.204e-01 38369
## blt0_1yrs80occ    2.209e-01 38369
## blt2_5yrs80occ    2.139e-01 38369
## blt6_10yrs80occ   1.686e-01 38369
## blt10_20yrs80occ  1.018e-01 38369
## blt20_30yrs80occ  4.554e-02 38369
## blt30_40yrs80occ  7.914e-02 38369
## pop_den8          8.507e-06 38369
## shrblk8          -8.399e-02 38369
## shrhsp8          -2.477e-01 38369
## child8           -4.366e-01 38369
## old8             -2.525e-01 38369
## shrfor8           1.314e+00 38369
## ffh8              3.191e-02 38369
## smhse8            4.742e-01 38369
## hsdrop8           5.619e-02 38369
## unemprt8         -1.087e+00 38369
## povrat8          -2.914e-01 38369
## welfare8          9.795e-01 38369
## avhhin8           1.440e-05 38369
## tothsun8          2.453e-05 38369
## ownocc8          -1.128e-04 38369
## occupied80        2.095e-01 38369
## no_hs_diploma8   -2.301e-01 38369
## ba_or_better8     5.607e-01 38369
##
## Multiple R-squared:  0.732 , Adjusted R-squared:  0.7317
## F-statistic:  2283 on 37 and 38369 DF,  p-value: < 2.2e-16
```

```
#coeftest(model3, vcov = vcovHC(model3, type = "HC1"))


#regression4
model_formula3 <- as.formula(paste("lnmdvalhs0 ~", paste(c(combined_list, "statefips"), collapse = " + "
model4 <- lm(model_formula3, data = all_sites_data)
summary(model4, se="hetero")
```

```
##
```

```
## Call:
## lm(formula = model_formula3, data = all_sites_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8687 -0.1777  0.0005  0.1806  2.3335
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       6.006e+00  8.190e-02  73.340  < 2e-16 ***
## npl2000           7.883e-02  1.091e-02   7.226 5.08e-13 ***
## lnmeanhs8         5.259e-01  5.131e-03 102.499  < 2e-16 ***
## firestoveheat80  -8.593e-03  2.127e-02  -0.404 0.686150
## noaircond80       4.355e-01  7.342e-03  59.321  < 2e-16 ***
## nofullkitchen80  -5.459e-01  1.062e-01  -5.141 2.75e-07 ***
## zerofullbath80    5.891e-01  8.908e-02   6.613 3.81e-11 ***
## detach80occ      -7.908e-03  1.098e-02  -0.720 0.471318
## mobile80occ       2.638e-01  2.361e-02  11.173  < 2e-16 ***
## bedrms0_80occ    -7.964e-01  1.420e-01  -5.609 2.05e-08 ***
## bedrms1_80occ    -2.485e-01  5.519e-02  -4.504 6.70e-06 ***
## bedrms2_80occ    -5.067e-01  4.400e-02 -11.516  < 2e-16 ***
## bedrms3_80occ    -5.792e-01  4.296e-02 -13.481  < 2e-16 ***
## bedrms4_80occ    -5.262e-01  5.201e-02 -10.117  < 2e-16 ***
## blt0_1yrs80occ    1.156e-01  3.908e-02   2.957 0.003106 **
## blt2_5yrs80occ    1.729e-01  2.415e-02   7.158 8.36e-13 ***
## blt6_10yrs80occ   8.013e-02  1.953e-02   4.102 4.10e-05 ***
## blt10_20yrs80occ  5.052e-02  1.384e-02   3.650 0.000262 ***
## blt20_30yrs80occ -4.989e-03  1.351e-02  -0.369 0.711828
## blt30_40yrs80occ  4.981e-02  2.129e-02   2.340 0.019282 *
## pop_den8          8.240e-06  2.679e-07  30.755  < 2e-16 ***
## shrblk8          -1.268e-01  1.207e-02 -10.504  < 2e-16 ***
## shrhsp8          -2.977e-01  1.787e-02 -16.663  < 2e-16 ***
## child8           -4.961e-01  4.442e-02 -11.170  < 2e-16 ***
## old8             -3.737e-01  3.835e-02  -9.743  < 2e-16 ***
## shrfor8           1.107e+00  3.028e-02  36.556  < 2e-16 ***
## ffh8             -4.527e-02  2.680e-02  -1.689 0.091241 .
## smhse8            4.589e-01  1.989e-02  23.070  < 2e-16 ***
## hsdrop8          -2.107e-02  1.978e-02  -1.066 0.286575
## unemprt8         -1.326e+00  5.773e-02 -22.963  < 2e-16 ***
## povrat8          -4.021e-01  3.887e-02 -10.345  < 2e-16 ***
## welfare8          6.515e-01  5.084e-02  12.814  < 2e-16 ***
## avhhin8           1.246e-05  3.410e-07  36.541  < 2e-16 ***
## tothsun8          1.744e-05  4.831e-06   3.610 0.000307 ***
## ownocc8          -1.457e-04  7.735e-06 -18.831  < 2e-16 ***
## occupied80        1.090e-01  3.453e-02   3.158 0.001591 **
## no_hs_diploma8   -2.214e-01  2.669e-02  -8.297  < 2e-16 ***
## ba_or_better8     5.776e-01  2.892e-02  19.973  < 2e-16 ***
## statefips        -3.704e-03  1.176e-04 -31.505  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.32 on 38368 degrees of freedom
## Multiple R-squared:  0.7387, Adjusted R-squared:  0.7385
## F-statistic:  2855 on 38 and 38368 DF,  p-value: < 2.2e-16
```

```
coeftest(model4, vcov = vcovHC(model4, type = "HC1"))
```

```
##
## t test of coefficients:
##
##                     Estimate  Std. Error  t value  Pr(>|t|)
## (Intercept)       6.0065e+00  2.2858e-01  26.2771 < 2.2e-16 ***
## npl2000           7.8827e-02  9.9233e-03   7.9436 2.017e-15 ***
## lnmeanhs8         5.2592e-01  2.0490e-02  25.6676 < 2.2e-16 ***
## firestoveheat80  -8.5934e-03  2.2232e-02  -0.3865 0.6991067
## noaircond80       4.3554e-01  9.0197e-03  48.2875 < 2.2e-16 ***
## nofullkitchen80  -5.4587e-01  1.5986e-01  -3.4147 0.0006391 ***
## zerofullbath80    5.8910e-01  1.2429e-01   4.7397 2.148e-06 ***
## detach80occ      -7.9077e-03  1.3985e-02  -0.5654 0.5717857
## mobile80occ       2.6382e-01  2.7225e-02   9.6906 < 2.2e-16 ***
## bedrms0_80occ    -7.9638e-01  2.3874e-01  -3.3358 0.0008514 ***
## bedrms1_80occ    -2.4854e-01  8.2399e-02  -3.0164 0.0025599 **
## bedrms2_80occ    -5.0672e-01  6.1975e-02  -8.1762 3.018e-16 ***
## bedrms3_80occ    -5.7919e-01  5.9853e-02  -9.6769 < 2.2e-16 ***
## bedrms4_80occ    -5.2619e-01  7.1358e-02  -7.3740 1.689e-13 ***
## blt0_1yrs80occ    1.1556e-01  4.8192e-02   2.3979 0.0164931 *
## blt2_5yrs80occ    1.7287e-01  2.9258e-02   5.9085 3.482e-09 ***
## blt6_10yrs80occ   8.0129e-02  2.2324e-02   3.5894 0.0003319 ***
## blt10_20yrs80occ  5.0520e-02  1.6057e-02   3.1463 0.0016545 **
## blt20_30yrs80occ -4.9892e-03  1.4805e-02  -0.3370 0.7361162
## blt30_40yrs80occ  4.9815e-02  2.5737e-02   1.9355 0.0529379 .
## pop_den8          8.2398e-06  4.0697e-07  20.2464 < 2.2e-16 ***
## shrblk8          -1.2684e-01  1.3755e-02  -9.2213 < 2.2e-16 ***
## shrhsp8          -2.9769e-01  2.0348e-02 -14.6303 < 2.2e-16 ***
## child8           -4.9614e-01  5.7131e-02  -8.6843 < 2.2e-16 ***
## old8             -3.7368e-01  4.7395e-02  -7.8844 3.245e-15 ***
## shrfor8           1.1069e+00  4.3422e-02  25.4908 < 2.2e-16 ***
## ffh8             -4.5267e-02  3.5644e-02  -1.2700 0.2040997
## smhse8            4.5894e-01  2.3733e-02  19.3375 < 2.2e-16 ***
## hsdrop8          -2.1074e-02  2.5557e-02  -0.8246 0.4096055
## unemprt8         -1.3258e+00  7.5126e-02 -17.6476 < 2.2e-16 ***
## povrat8          -4.0211e-01  5.0913e-02  -7.8979 2.911e-15 ***
## welfare8          6.5150e-01  6.8708e-02   9.4820 < 2.2e-16 ***
## avhhin8           1.2462e-05  6.9441e-07  17.9456 < 2.2e-16 ***
## tothsun8          1.7438e-05  6.8717e-06   2.5377 0.0111636 *
## ownocc8          -1.4567e-04  1.0907e-05 -13.3554 < 2.2e-16 ***
## occupied80        1.0904e-01  4.6705e-02   2.3347 0.0195639 *
## no_hs_diploma8   -2.2140e-01  3.4508e-02  -6.4159 1.416e-10 ***
## ba_or_better8     5.7756e-01  3.5525e-02  16.2577 < 2.2e-16 ***
## statefips        -3.7035e-03  1.4475e-04 -25.5863 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#b: compare covariates
covariate_list <- c(basic_character_list,Economic_population_list)
covariate_list <- covariate_list[covariate_list!="old8"]

compare_table <- all_covariates_data %>%
```

```
  group_by(npl2000) %>%
  summarise(across(all_of(covariate_list), \(x) mean(x,na.rm=TRUE)))
print(compare_table)
```

```
## # A tibble: 2 x 22
##   npl2000 firestoveheat80 noaircond80 nofullkitchen80 zerofullbath80 detach80occ
##     <dbl>           <dbl>       <dbl>           <dbl>          <dbl>       <dbl>
## 1       0          0.0437       0.427          0.0174         0.0230       0.879
## 2       1          0.0535       0.489          0.0183         0.0246       0.879
## # i 16 more variables: mobile80occ <dbl>, pop_den8 <dbl>, shrblk8 <dbl>,
## #   shrhsp8 <dbl>, child8 <dbl>, shrfor8 <dbl>, ffh8 <dbl>, smhse8 <dbl>,
## #   hsdrop8 <dbl>, unemprt8 <dbl>, povrat8 <dbl>, welfare8 <dbl>,
## #   avhhin8 <dbl>, tothsun8 <dbl>, ownocc8 <dbl>, occupied80 <dbl>
```

```
diff_by_npl2000 <- compare_table[2,] - compare_table[1,]
print(diff_by_npl2000)
```

```
##   npl2000 firestoveheat80 noaircond80 nofullkitchen80 zerofullbath80
## 1       1     0.009760101  0.06245345    0.0009031986    0.001564953
##     detach80occ mobile80occ  pop_den8     shrblk8     shrhsp8     child8
## 1 -0.0004402019  0.03607002 -4023.342 -0.02277776 -0.01835318 0.01459556
##       shrfor8        ffh8     smhse8     hsdrop8    unemprt8      povrat8
## 1 -0.01543198 -0.02584867 0.02940218 0.001789729 0.003777791 -0.006202148
##       welfare8   avhhin8 tothsun8 ownocc8  occupied80
## 1 -0.001655657 -1196.139 50.40042 111.576 0.007297695
```

```
compare_data <- site_covariates_data %>%
  mutate(
    hrs_group = case_when(
      hrs_82 < 16.5 ~ 0,
      hrs_82 >= 16.5 & hrs_82 <= 28.5 ~ 1,
      hrs_82 >= 28.5 & hrs_82 <= 40.5 ~ 2,
      hrs_82 > 40.5 ~ 3
    )
  )

compare_table2 <- compare_data %>%
  group_by(hrs_group) %>%
  summarise(across(all_of(covariate_list), \(x) mean(x,na.rm=TRUE)))
print(compare_table2)
```

```
## # A tibble: 4 x 22
##   hrs_group firestoveheat80 noaircond80 nofullkitchen80 zerofullbath80
##       <dbl>           <dbl>       <dbl>           <dbl>          <dbl>
## 1         0          0.0444       0.506          0.0183         0.0292
## 2         1          0.0637       0.498          0.0253         0.0362
## 3         2          0.0603       0.535          0.0229         0.0311
## 4         3          0.0518       0.459          0.0159         0.0226
## # i 17 more variables: detach80occ <dbl>, mobile80occ <dbl>, pop_den8 <dbl>,
## #   shrblk8 <dbl>, shrhsp8 <dbl>, child8 <dbl>, shrfor8 <dbl>, ffh8 <dbl>,
## #   smhse8 <dbl>, hsdrop8 <dbl>, unemprt8 <dbl>, povrat8 <dbl>, welfare8 <dbl>,
## #   avhhin8 <dbl>, tothsun8 <dbl>, ownocc8 <dbl>, occupied80 <dbl>
```

```r
diff_by_npl2000_0_1 <- compare_table2[2,] - compare_table2[1,]
diff_by_npl2000_1_2 <- compare_table2[3,] - compare_table2[2,]
diff_by_npl2000_2_3 <- compare_table2[4,] - compare_table2[3,]
print(diff_by_npl2000_0_1)
```

```
##   hrs_group firestoveheat80  noaircond80 nofullkitchen80 zerofullbath80
## 1         1      0.01926776 -0.008023427     0.006930967    0.007012545
##     detach80occ mobile80occ  pop_den8     shrblk8     shrhsp8      child8
## 1 -0.0003659126  0.02500993 -535.7442 -0.05990577 -0.03466658 -0.01048222
##       shrfor8        ffh8      smhse8    hsdrop8    unemprt8    povrat8
## 1 -0.02887849 -0.05515569 -0.02801687 0.01542357 -0.02225875 -0.0109622
##      welfare8  avhhin8 tothsun8  ownocc8   occupied80
## 1 -0.01677323 343.0021 24.58212 92.52349 -0.001652427
```

```r
print(diff_by_npl2000_1_2)
```

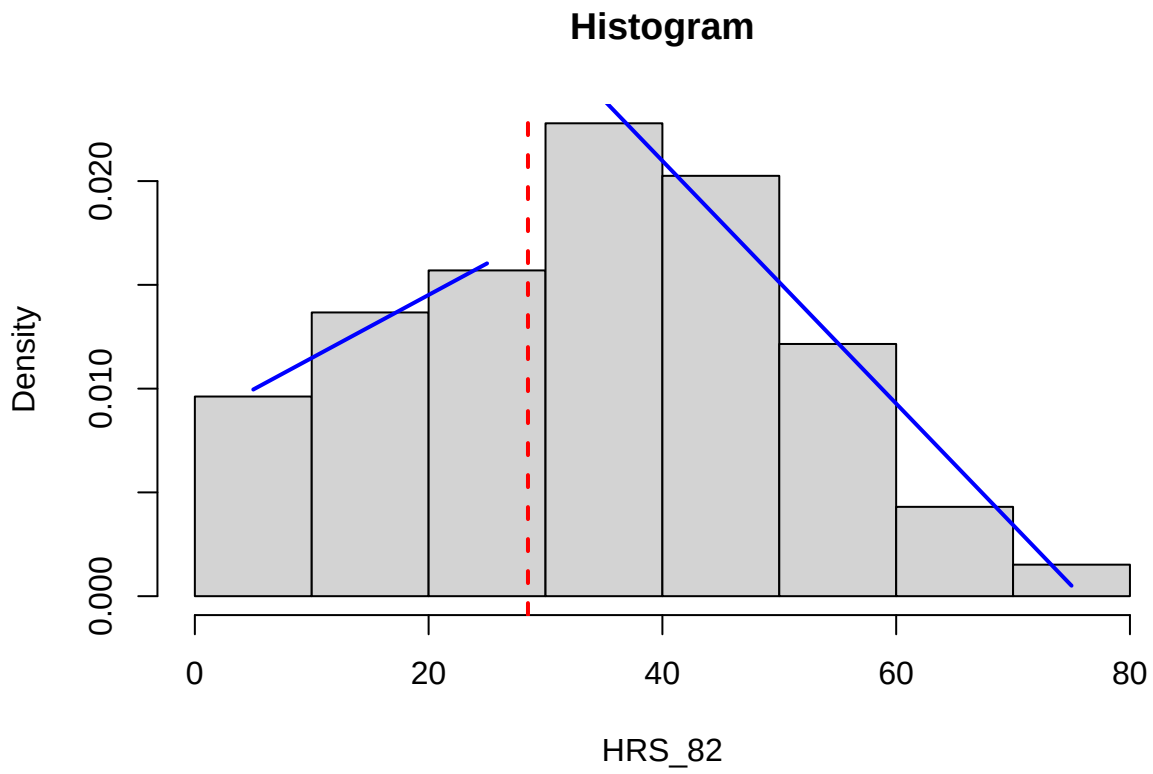
```
##   hrs_group firestoveheat80 noaircond80 nofullkitchen80 zerofullbath80
## 1         1    -0.003330449  0.03740652     -0.00232992   -0.005159094
##   detach80occ mobile80occ  pop_den8     shrblk8     shrhsp8      child8
## 1  0.04239224 -0.01372709 -254.6245 -0.00514858 -0.00213929 0.002317222
##        shrfor8        ffh8      smhse8     hsdrop8     unemprt8     povrat8
## 1 -0.003660411 0.007222568 -0.01853137 -0.01563834 0.0006116025 0.003377021
##       welfare8   avhhin8  tothsun8   ownocc8  occupied80
## 1 -0.005945316 -132.8292 -8.232644 -52.73702 0.003410623
```

```r
print(diff_by_npl2000_2_3)
```

```
##   hrs_group firestoveheat80 noaircond80 nofullkitchen80 zerofullbath80
## 1         1    -0.008516104  -0.0761523    -0.007071843   -0.008450063
##    detach80occ  mobile80occ pop_den8     shrblk8    shrhsp8     child8
## 1 -0.002661128 -0.001812645 -146.174 -0.03039714 0.02789394 0.00202846
##      shrfor8        ffh8      smhse8      hsdrop8     unemprt8    povrat8
## 1 0.02203435 -0.01757916 -0.02272595 0.0002385719 -0.007649737 -0.0233712
##      welfare8  avhhin8  tothsun8  ownocc8  occupied80
## 1 -0.00176901 1656.868 -11.99188 6.235567 -0.00171391
```

```r
# Q2

#b
hrs_bin <- pretty(mile_data$hrs_82, n = 10)
hist_obj <- hist(mile_data$hrs_82, breaks = hrs_bin, freq = FALSE,
                 main = "Histogram", xlab = "HRS_82", ylab = "Density")

abline(v = 28.5, col = "red", lwd = 2, lty = 2)

#hrs_mid <- (head(hrs_bin,-1)+tail(hrs_bin,-1)) / 2
bin_data <- data.frame(
  mid = hist_obj$mids,
  density = hist_obj$density
)
```

```r
left_bin <- bin_data %>% filter(mid<=28.5)
right_bin <- bin_data %>% filter(mid>28.5)

model_left <- lm(density ~ mid, data=left_bin)
model_right <- lm(density ~ mid, data=right_bin)

lines(left_bin$mid, predict(model_left), col = "blue", lwd = 2)
lines(right_bin$mid, predict(model_right), col = "blue", lwd = 2)
```




```r
# Q3

#a
new_covariates <- c("firestoveheat80_nbr", "noaircond80_nbr", "nofullkitchen80_nbr", "zerofullbath80_nbr",
                    "detach80occ_nbr", "mobile80occ_nbr", "pop_den8_nbr","shrblk8_nbr",
                    "shrhsp8_nbr", "child8_nbr", "shrfor8_nbr", "ffh8_nbr", "smhse8_nbr", "hsdrop8_nbr",
                    "unemprt8_nbr", "povrat8_nbr", "welfare8_nbr", "avhhin8_nbr", "tothsun8_nbr",
                    "ownocc8_nbr", "occupied80_nbr")
temp_vars3 <- grep("^(bedrms|blt|no_hs_dipl|no_hs_dipl)", mile_data, value = TRUE)
new_cov_list <- c(new_covariates,temp_vars3)
data_RD <- mile_data
data_RD$z <- ifelse(data_RD$hrs_82 > 28.5, 1, 0)

formula_rd_fs <- as.formula(
  paste("npl2000 ~ z +", paste(new_cov_list, collapse = " + "))
)
```

```r
model_first_stage <- lm(formula_rd_fs, data = data_RD)
summary(model_first_stage, se="hetero")
```

```
##
## Call:
## lm(formula = formula_rd_fs, data = data_RD)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.92306 -0.11785 -0.01807  0.03601  0.89073
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           4.195e-01  5.219e-01   0.804  0.42200
## z                     7.799e-01  2.708e-02  28.802  < 2e-16 ***
## firestoveheat80_nbr -4.882e-02  2.833e-01  -0.172  0.86331
## noaircond80_nbr       2.040e-01  6.767e-02   3.015  0.00275 **
## nofullkitchen80_nbr   4.706e-01  1.589e+00   0.296  0.76734
## zerofullbath80_nbr  -2.189e+00  1.417e+00  -1.545  0.12331
## detach80occ_nbr       1.417e-01  1.424e-01   0.995  0.32049
## mobile80occ_nbr       5.494e-01  2.760e-01   1.991  0.04726 *
## pop_den8_nbr         -2.075e-06  9.926e-06  -0.209  0.83450
## shrblk8_nbr           3.703e-02  1.852e-01   0.200  0.84168
## shrhsp8_nbr           1.053e-01  2.213e-01   0.476  0.63429
## child8_nbr           -4.871e-01  4.247e-01  -1.147  0.25223
## shrfor8_nbr          -3.516e-02  3.676e-01  -0.096  0.92385
## ffh8_nbr              8.344e-02  4.369e-01   0.191  0.84863
## smhse8_nbr            3.687e-01  1.863e-01   1.979  0.04855 *
## hsdrop8_nbr          -2.228e-01  2.270e-01  -0.981  0.32705
## unemprt8_nbr         -1.269e+00  5.772e-01  -2.198  0.02854 *
## povrat8_nbr           3.322e-01  5.111e-01   0.650  0.51606
## welfare8_nbr          2.685e-01  6.804e-01   0.395  0.69331
## avhhin8_nbr           3.196e-06  4.347e-06   0.735  0.46261
## tothsun8_nbr          1.414e-06  6.361e-06   0.222  0.82417
## ownocc8_nbr          -3.608e-06  8.459e-06  -0.427  0.66996
## occupied80_nbr       -5.338e-01  5.249e-01  -1.017  0.30984
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2447 on 372 degrees of freedom
## Multiple R-squared:  0.7308, Adjusted R-squared:  0.7149
## F-statistic:  45.9 on 22 and 372 DF,  p-value: < 2.2e-16
```

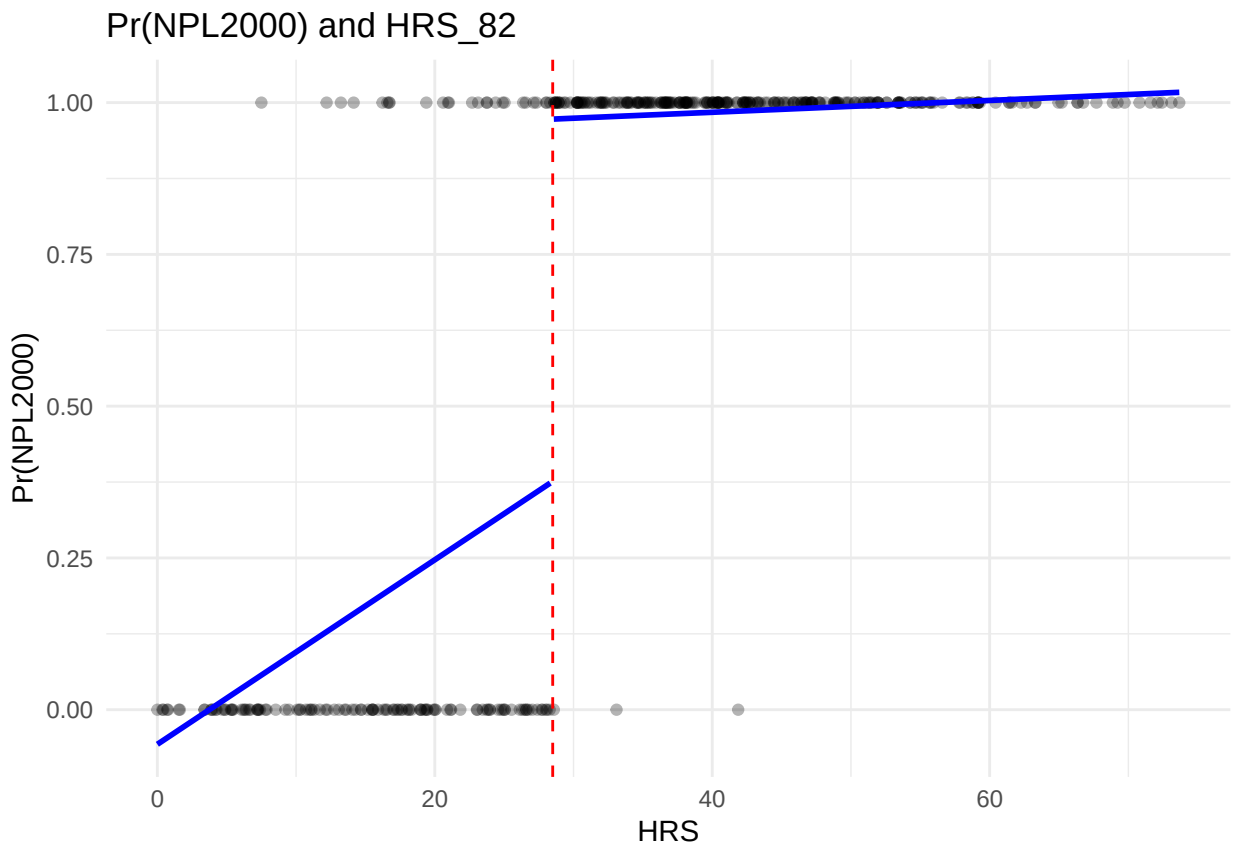
```r
data_RD_between <- data_RD %>% filter(hrs_82 >= 16.5 & hrs_82 <= 40.5)

model_first_stage_with_bandwidth <- lm(formula_rd_fs, data = data_RD_between)
summary(model_first_stage_with_bandwidth, se="hetero")
```

```
##
## Call:
## lm(formula = formula_rd_fs, data = data_RD_between)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.84433 -0.17186 -0.03057  0.08744  0.74733
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           1.081e+00  1.074e+00   1.007   0.3154
## z                     6.867e-01  4.654e-02  14.757   <2e-16 ***
## firestoveheat80_nbr -1.404e-01  4.719e-01  -0.298   0.7664
## noaircond80_nbr       2.887e-01  1.156e-01   2.498   0.0135 *
## nofullkitchen80_nbr -2.351e+00  2.861e+00  -0.822   0.4124
## zerofullbath80_nbr  -3.170e-01  2.687e+00  -0.118   0.9063
## detach80occ_nbr       3.348e-01  2.744e-01   1.220   0.2242
## mobile80occ_nbr       9.421e-01  5.307e-01   1.775   0.0778 .
## pop_den8_nbr         -1.050e-06  2.068e-05  -0.051   0.9596
## shrblk8_nbr           2.141e-01  3.550e-01   0.603   0.5472
## shrhsp8_nbr           3.667e-03  4.169e-01   0.009   0.9930
## child8_nbr           -1.004e+00  7.342e-01  -1.367   0.1735
## shrfor8_nbr          -8.012e-01  8.331e-01  -0.962   0.3376
## ffh8_nbr              3.905e-01  7.921e-01   0.493   0.6227
## smhse8_nbr            6.136e-01  3.336e-01   1.839   0.0677 .
## hsdrop8_nbr          -1.630e-01  3.700e-01  -0.441   0.6600
## unemprt8_nbr         -6.631e-01  1.008e+00  -0.658   0.5117
## povrat8_nbr          -2.314e-01  8.893e-01  -0.260   0.7950
## welfare8_nbr          8.691e-01  1.278e+00   0.680   0.4976
## avhhin8_nbr           8.834e-06  8.681e-06   1.018   0.3103
## tothsun8_nbr          2.709e-06  1.281e-05   0.212   0.8327
## ownocc8_nbr          -4.545e-06  1.476e-05  -0.308   0.7585
## occupied80_nbr       -1.575e+00  1.103e+00  -1.427   0.1554
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2978 on 164 degrees of freedom
## Multiple R-squared:  0.617,  Adjusted R-squared:  0.5656
## F-statistic: 12.01 on 22 and 164 DF,  p-value: < 2.2e-16
```

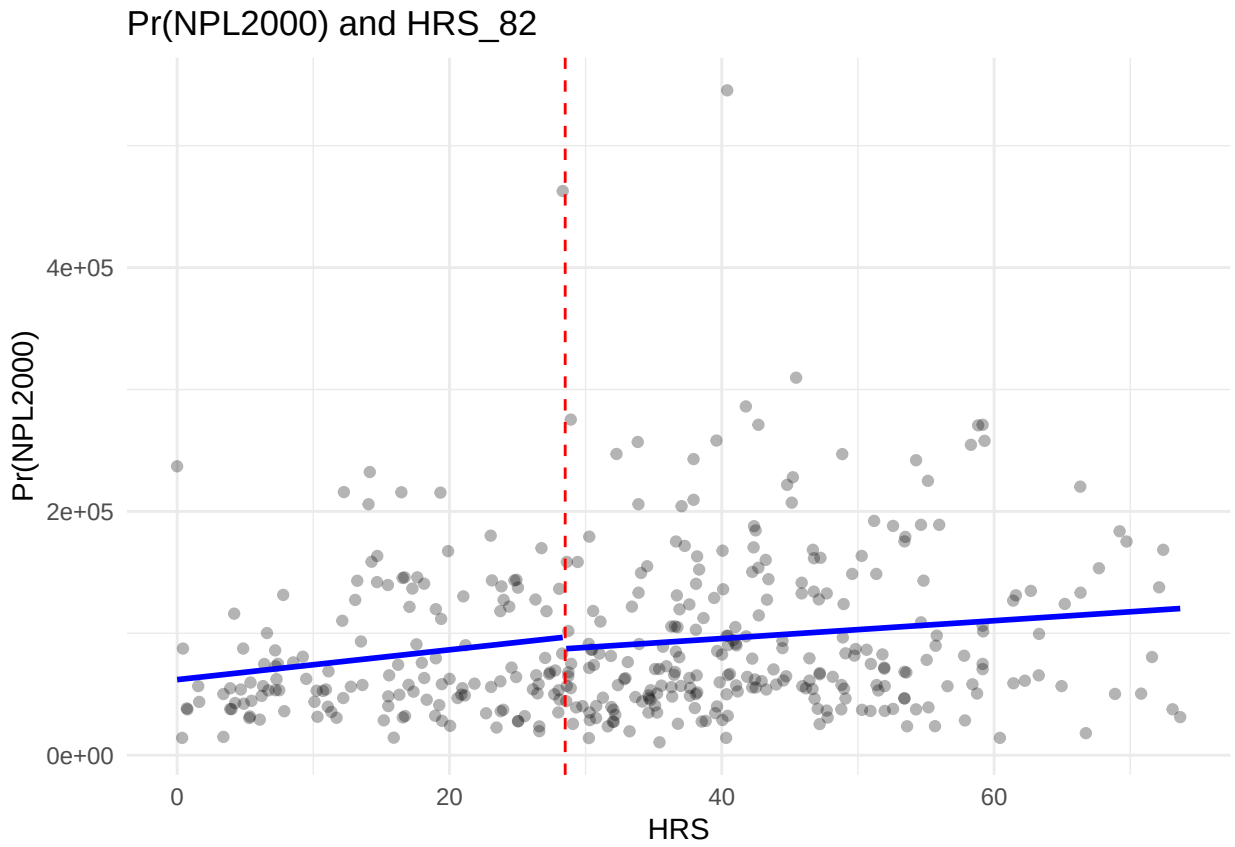
```
#b
data_RD_plot <- data_RD %>%
  mutate(group = ifelse(hrs_82 < 28.5, "l", "r"))

ggplot(data_RD_plot, aes(x = hrs_82, y = npl2000)) +
  geom_point(alpha = 0.3, size = 1.5) +
  geom_smooth(data = subset(data_RD_plot, group == "l"), method = "lm", se = FALSE, color = "blue") +
  geom_smooth(data = subset(data_RD_plot, group == "r"), method = "lm", se = FALSE, color = "blue") +
  geom_vline(xintercept = 28.5, linetype = "dashed", color = "red") +
  labs( title = "Pr(NPL2000) and HRS_82", x = "HRS", y = "Pr(NPL2000)") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```




```
#c
ggplot(data_RD_plot, aes(x = hrs_82, y = mdvalhs9)) +
  geom_point(alpha = 0.3, size = 1.5) +
  geom_smooth(data = subset(data_RD_plot, group == "l"), method = "lm", se = FALSE, color = "blue") +
  geom_smooth(data = subset(data_RD_plot, group == "r"), method = "lm", se = FALSE, color = "blue") +
  geom_vline(xintercept = 28.5, linetype = "dashed", color = "red") +
  labs( title = "Pr(NPL2000) and HRS_82", x = "HRS", y = "Pr(NPL2000)") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```




```
# Q4
data_RD$pred_npl <- predict(model_first_stage)
formula_rd_ss <- as.formula(paste("mdvalhs0 ~ pred_npl +", paste(new_cov_list, collapse = " + ")))
model_second_stage <- lm(formula_rd_ss, data = data_RD)
summary(model_second_stage, se="hetero")
```

```
##
## Call:
## lm(formula = formula_rd_ss, data = data_RD)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -89097 -28017  -6621  19523 196034
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -2.726e+05  9.021e+04  -3.022  0.00268 **
## pred_npl             5.209e+03  5.997e+03   0.869  0.38559
## firestoveheat80_nbr  1.724e+04  4.894e+04   0.352  0.72485
## noaircond80_nbr      3.800e+04  1.172e+04   3.244  0.00129 **
## nofullkitchen80_nbr -3.185e+05  2.748e+05  -1.159  0.24721
## zerofullbath80_nbr   2.930e+05  2.469e+05   1.187  0.23613
## detach80occ_nbr      1.717e+04  2.461e+04   0.697  0.48595
## mobile80occ_nbr      1.011e+05  4.792e+04   2.109  0.03558 *
## pop_den8_nbr         3.465e+00  1.715e+00   2.021  0.04405 *
## shrblk8_nbr          3.644e+03  3.198e+04   0.114  0.90936
```

```
## shrhsp8_nbr         -8.199e+04  3.821e+04  -2.146  0.03256 *
## child8_nbr          -2.205e+05  7.345e+04  -3.002  0.00287 **
## shrfor8_nbr          1.557e+05  6.348e+04   2.453  0.01462 *
## ffh8_nbr            -1.229e+04  7.547e+04  -0.163  0.87075
## smhse8_nbr          -8.014e+04  3.230e+04  -2.482  0.01352 *
## hsdrop8_nbr          5.091e+04  3.935e+04   1.294  0.19652
## unemprt8_nbr        -5.728e+05  1.009e+05  -5.678 2.75e-08 ***
## povrat8_nbr          6.587e+03  8.844e+04   0.074  0.94067
## welfare8_nbr         4.710e+05  1.176e+05   4.007 7.43e-05 ***
## avhhin8_nbr          1.106e+01  7.514e-01  14.719  < 2e-16 ***
## tothsun8_nbr        -2.954e-01  1.099e+00  -0.269  0.78820
## ownocc8_nbr         -3.655e+00  1.462e+00  -2.499  0.01287 *
## occupied80_nbr       2.564e+05  9.063e+04   2.829  0.00493 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 42270 on 372 degrees of freedom
## Multiple R-squared:  0.6192, Adjusted R-squared:  0.5967
## F-statistic: 27.49 on 22 and 372 DF,  p-value: < 2.2e-16
```

```
data_RD_between$pred_npl <- predict(model_first_stage_with_bandwidth)
model_second_stage_with_bandwidth <- lm(formula_rd_ss, data = data_RD_between)
summary(model_second_stage_with_bandwidth, se="hetero")
```

```
##
## Call:
## lm(formula = formula_rd_ss, data = data_RD_between)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -83315 -26044  -4355  19197 152890
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -2.843e+05  1.409e+05  -2.018  0.04526 *
## pred_npl            -4.285e+02  8.855e+03  -0.048  0.96147
## firestoveheat80_nbr  2.299e+04  6.166e+04   0.373  0.70981
## noaircond80_nbr      4.599e+04  1.525e+04   3.015  0.00298 **
## nofullkitchen80_nbr -2.031e+05  3.721e+05  -0.546  0.58594
## zerofullbath80_nbr   1.532e+05  3.515e+05   0.436  0.66356
## detach80occ_nbr      5.055e+04  3.607e+04   1.401  0.16297
## mobile80occ_nbr      1.537e+05  7.007e+04   2.193  0.02973 *
## pop_den8_nbr         4.149e+00  2.701e+00   1.536  0.12650
## shrblk8_nbr          3.140e+04  4.630e+04   0.678  0.49859
## shrhsp8_nbr         -8.187e+02  5.448e+04  -0.015  0.98803
## child8_nbr          -1.746e+05  9.609e+04  -1.818  0.07095 .
## shrfor8_nbr          3.046e+05  1.090e+05   2.795  0.00581 **
## ffh8_nbr            -1.037e+05  1.039e+05  -0.998  0.31964
## smhse8_nbr          -9.531e+04  4.379e+04  -2.176  0.03096 *
## hsdrop8_nbr          9.834e+03  4.846e+04   0.203  0.83946
## unemprt8_nbr        -4.863e+05  1.323e+05  -3.677  0.00032 ***
## povrat8_nbr          3.090e+04  1.161e+05   0.266  0.79052
## welfare8_nbr         5.178e+05  1.672e+05   3.097  0.00230 **
## avhhin8_nbr          9.882e+00  1.135e+00   8.704 3.27e-15 ***
```

```
## tothsun8_nbr        -1.569e+00  1.670e+00  -0.940  0.34883
## ownocc8_nbr         -1.306e+00  1.928e+00  -0.677  0.49918
## occupied80_nbr       2.453e+05  1.449e+05   1.693  0.09233 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 38910 on 164 degrees of freedom
## Multiple R-squared:  0.6134, Adjusted R-squared:  0.5616
## F-statistic: 11.83 on 22 and 164 DF,  p-value: < 2.2e-16
```