appendix_placebo.R

datalab

2023-06-16

```
#############################################################
##############           Placebo Tests  ####################
###############     October 10, 2018     ##################
#######  Rerun December 17, 2022

rm(list=ls())
library(foreign)
library(plyr)
library(readstata13)
library(multiwayvcov)
library(sandwich)
library(lmtest)
library(stargazer)
library(ggplot2)


################################################
########### Multiplot Function #################
######   Source: http://www.cookbook-r.com/Graphs/Multiple_graphs_on_one_page_(ggplot2)/



multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL) {
  library(grid)

  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)

  numPlots = length(plots)

  # If layout is NULL, then use 'cols' to determine layout
  if (is.null(layout)) {
    # Make the panel
    # ncol: Number of columns of plots
    # nrow: Number of rows needed, calculated from # of cols
    layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
                     ncol = cols, nrow = ceiling(numPlots/cols))
  }

  if (numPlots==1) {
    print(plots[[1]])
```

```
  } else {
    # Set up the page
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))

    # Make each plot, in the correct location
    for (i in 1:numPlots) {
      # Get the i,j matrix positions of the regions that contain this subplot
      matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))

      print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
                                      layout.pos.col = matchidx$col))
    }
  }
}
#########################################################


data=read.csv("~/Dropbox/Personal Research 2017/replications/karn_nov16.csv")
names(data)
```

```
##   [1] "X.1"              "dist_name"        "vilname91"        "v1"               "dist_code"
##  [15] "phc_cntr"         "hc_cntr"          "fpc_cntr"         "tb_cntr"          "nh_cntr"
##  [29] "st_town"          "agri_land"        "near_town"        "circl_code"       "m_pop"
##  [43] "m_sc"             "f_sc"             "tot_st"           "m_st"             "f_st"
##  [57] "ngmf_char"        "ngmfprwmedd"      "ngmfprwothd"      "ngmfprwnod"       "ngmftrpr"
##  [71] "taptr"            "tapuntr"          "hp"               "covwell"          "uncovwell"
##  [85] "phs_cntr"         "prhsc"            "stname"           "stname1991"       "d_name"
##  [99] "all_hosp"         "area_na_cu"       "ayu_disp"         "ayu_hosp"         "canal_govt"
## [113] "ind_sch"          "lake"             "m_home"           "m_sch"            "nw_fac"
## [127] "power_oth"        "power_supl"       "p_sch"            "p_t_fac"          "rang_mcw"
## [141] "s_sch"            "s_s_sch"          "tot_exp"          "tot_inc"          "tr_sch"
## [155] "gov_ps_n"         "pr_ps_n"          "gov_ms_n"         "pr_ms_n"          "gov_secs_n"
## [169] "nviltmsna"        "nviltsecs"        "nviltsecsna"      "power"            "hplost"
## [183] "pnt_fac"          "power_ea"         "power_eag"        "power_edea"       "power_eo"
## [197] "medfac"           "rangmed"          "tot_hh"           "pucca_binary"     "kucha_binary"
## [211] "dist_fr_town"     "tbcl"             "tank"             "tap"              "X"
## [225] "M_POP"            "F_POP"            "TOT_L6"           "M_L6"             "F_L6"
## [239] "TOT_ILLT"         "M_ILLT"           "F_ILLT"           "TOT_W"            "M_W"
## [253] "F_AGLB"           "TOT_MFHH"         "M_MFHH"           "F_MFHH"           "TOT_OTH_W"
## [267] "M_MRG_AGLB"       "F_MRG_AGLB"       "T_MRG_HH"         "M_MRG_HH"         "F_MRG_HH"
## [281] "NEAR_DIST_border1" "NEAR_ANGLE"       "temp_av"          "wc2010mt_1"       "TerrainRug"
```

```
summary(data$NEAR_DIST_border1)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
##     45.1  17162.2  40186.1  45728.7  67400.2 133049.0     5146
```

```
summary(data$border1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.000   0.000   1.000   0.599   1.000   1.000    5146
```

```
#####################################################
###Create Placebo Border to the North (-10 km to Princely State)
```

```
#calculate new distances

#border1
#mysore part
summary(data$NEAR_DIST_border1[data$border1==1])
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
##    119.9  20224.4  42551.1  47531.3  66465.8 133049.0     5146
```

```
data$distplacebo1[(data$border1==1)&!is.na(data$border1)]=
  data$NEAR_DIST_border1[(data$border1==1)&!is.na(data$border1)]+10000

summary(data$distplacebo1[data$border1==1])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   10120   30224   52551   57531   76466  143049    5146
```

```
#bombay part
data$distplacebo1[(data$border1==0)&!is.na(data$border1)
                  &(data$NEAR_DIST_border1<10000)]=
  10000-data$NEAR_DIST_border1[(data$border1==0)&!is.na(data$border1)
                               &(data$NEAR_DIST_border1<10000)]

data$distplacebo1[(data$border1==0)&!is.na(data$border1)
                  &(data$NEAR_DIST_border1>=10000)]=
  data$NEAR_DIST_border1[(data$border1==0)&!is.na(data$border1)
                         &(data$NEAR_DIST_border1>=10000)]-10000

summary(data$NEAR_DIST_border1[data$border1==0])
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
##     45.1  13709.4  34703.0  43034.2  68323.9 130614.8     5146
```

```
summary(data$distplacebo1[data$border1==0])
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.      NA's
##      3.57   7355.51  24702.97  34974.50  58323.93 120614.78      5146
```

```
#create a new treatment variable
#old mysore and chunk of the bombay are in the treatment,
#and the rest of bombay are in teh control group
data$borderplac1[data$border1==1]=1
data$borderplac1[(data$border1==0)&!is.na(data$border1)
                 &(data$NEAR_DIST_border1<10000)]=1
data$borderplac1[(data$border1==0)&!is.na(data$border1)
                 &(data$NEAR_DIST_border1>=10000)]=0
summary(data$borderplac1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.000   0.000   1.000   0.674   1.000   1.000    5146
```

```
##### Distances #####

#Distance to Mysore-Bombay Border
rd10.mb=data[which(data$distplacebo1<10000),] #20 km

table(rd10.mb$borderplac1)
```

```
##
##   0   1
## 455 559
```

```
#######Linear Polynomial #####

#baseline bandwidth
#Mysore-Bombay
mys.health1=lm(health_binary~borderplac1+TOT_POP+
        TOT_SC+TOT_ST+Slope+TerrainRug+Latitude+Longitude, data=rd10.mb)
summary(mys.health1)
```

```
##
## Call:
## lm(formula = health_binary ~ borderplac1 + TOT_POP + TOT_SC +
##     TOT_ST + Slope + TerrainRug + Latitude + Longitude, data = rd10.mb)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7468 -0.2738 -0.1436  0.3224  1.0358
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.571e+00  3.430e+00  -2.499  0.01261 *
## borderplac1 -5.106e-02  2.823e-02  -1.809  0.07075 .
## TOT_POP      9.626e-05  1.383e-05   6.962 6.04e-12 ***
## TOT_SC       1.272e-04  4.370e-05   2.911  0.00368 **
## TOT_ST       2.871e-04  5.617e-05   5.111 3.84e-07 ***
## Slope       -2.702e-04  1.159e-03  -0.233  0.81569
## TerrainRug  -2.827e-03  6.345e-03  -0.446  0.65598
## Latitude    -1.787e-01  7.933e-02  -2.253  0.02448 *
## Longitude    1.499e-01  5.146e-02   2.912  0.00367 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4083 on 1005 degrees of freedom
## Multiple R-squared:  0.1399, Adjusted R-squared:  0.133
## F-statistic: 20.43 on 8 and 1005 DF,  p-value: < 2.2e-16
```

```
mys.health1.cl=cluster.vcov(mys.health1, rd10.mb$dist_name)
mys.health1.se=sqrt(diag(mys.health1.cl)) #cluster standard errors

mys.pucca1=lm(pucca_binary~borderplac1+TOT_POP+
        TOT_SC+TOT_ST+Slope+TerrainRug+Latitude+Longitude, data=rd10.mb)
summary(mys.pucca1)
```

```
##
## Call:
## lm(formula = pucca_binary ~ borderplac1 + TOT_POP + TOT_SC +
##     TOT_ST + Slope + TerrainRug + Latitude + Longitude, data = rd10.mb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.00046  0.01794  0.08377  0.13515  0.23091
##
```

```
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.057e+00  2.355e+00  -3.422 0.000647 ***
## borderplac1  9.152e-04  1.938e-02   0.047 0.962341
## TOT_POP      1.550e-05  9.492e-06   1.633 0.102829
## TOT_SC       2.281e-05  3.000e-05   0.760 0.447169
## TOT_ST       5.406e-05  3.856e-05   1.402 0.161287
## Slope        1.551e-04  7.955e-04   0.195 0.845441
## TerrainRug  -4.616e-03  4.356e-03  -1.060 0.289529
## Latitude    -4.469e-02  5.447e-02  -0.820 0.412142
## Longitude    1.275e-01  3.533e-02   3.610 0.000322 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2803 on 1005 degrees of freedom
## Multiple R-squared:  0.04659,    Adjusted R-squared:  0.039
## F-statistic: 6.139 on 8 and 1005 DF,  p-value: 9.476e-08
```

```
mys.pucca1.cl=cluster.vcov(mys.pucca1, rd10.mb$dist_name)
mys.pucca1.se=sqrt(diag(mys.pucca1.cl)) #cluster standard errors


##############################
#### Create Placebo Border to the South (+10 km to the princely state)

#border1
#mysore part
summary(data$NEAR_DIST_border1[data$border1==0])
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
##     45.1  13709.4  34703.0  43034.2  68323.9 130614.8     5146
```

```
data$distplacebo1[(data$border1==0)&!is.na(data$border1)]=
  data$NEAR_DIST_border1[(data$border1==0)&!is.na(data$border1)]+10000

summary(data$distplacebo1[data$border1==0])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   10045   23709   44703   53034   78324  140615    5146
```

```
#bombay part
data$distplacebo1[(data$border1==1)&!is.na(data$border1)
                  &(data$NEAR_DIST_border1<10000)]=
  10000-data$NEAR_DIST_border1[(data$border1==1)&!is.na(data$border1)
                               &(data$NEAR_DIST_border1<10000)]

data$distplacebo1[(data$border1==1)&!is.na(data$border1)
                  &(data$NEAR_DIST_border1>=10000)]=
  data$NEAR_DIST_border1[(data$border1==1)&!is.na(data$border1)
                         &(data$NEAR_DIST_border1>=10000)]-10000

summary(data$NEAR_DIST_border1[data$border1==1])
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
##    119.9  20224.4  42551.1  47531.3  66465.8 133049.0     5146
```

```r
summary(data$distplacebo1[data$border1==1])
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.      NA's
##     12.96  10224.44  32551.09  38913.18  56465.82 123049.02      5146
```

```r
#create a new treatment variable
#old mysore and chunk of the bombay are in the treatment,
#and the rest of bombay are in teh control group

data$borderplac1[data$border1==0]=1
data$borderplac1[(data$border1==1)&!is.na(data$border1)
                 &(data$NEAR_DIST_border1<10000)]=1
data$borderplac1[(data$border1==1)&!is.na(data$border1)
                 &(data$NEAR_DIST_border1>=10000)]=0
summary(data$borderplac1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.000   0.000   0.000   0.481   1.000   1.000    5146
```

```r
##### Distances #####

#Distance to Mysore-Bombay Border
rd10.mb=data[which(data$distplacebo1<10000),] #20 km

table(rd10.mb$borderplac1)
```

```
##
##   0   1
## 511 599
```

```r
##### Simple OLS with controls ######
## and linear polynomial of lat and long  ###


#baseline bandwidth
#Mysore-Bombay
mys.health2=lm(health_binary~borderplac1+TOT_POP+
        TOT_SC+TOT_ST+Slope+TerrainRug+Latitude+Longitude, data=rd10.mb)
summary(mys.health2)
```

```
##
## Call:
## lm(formula = health_binary ~ borderplac1 + TOT_POP + TOT_SC +
##     TOT_ST + Slope + TerrainRug + Latitude + Longitude, data = rd10.mb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.52151 -0.23538 -0.13628 -0.05932  1.00326
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -9.155e+00  3.085e+00  -2.968  0.00307 **
## borderplac1  6.455e-03  2.722e-02   0.237  0.81262
## TOT_POP     -2.105e-06  2.080e-05  -0.101  0.91943
## TOT_SC       1.824e-04  4.725e-05   3.861  0.00012 ***
## TOT_ST       3.298e-04  6.071e-05   5.432 6.84e-08 ***
```

```
## Slope        1.701e-03  1.059e-03   1.607  0.10843
## TerrainRug   8.472e-04  7.005e-03   0.121  0.90376
## Latitude     1.525e-01  8.493e-02   1.795  0.07286 .
## Longitude    9.180e-02  4.225e-02   2.173  0.03001 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3929 on 1101 degrees of freedom
## Multiple R-squared:  0.07367,    Adjusted R-squared:  0.06694
## F-statistic: 10.95 on 8 and 1101 DF,  p-value: 6.085e-15
```

```
mys.health2.cl=cluster.vcov(mys.health2, rd10.mb$dist_name)
mys.health2.se=sqrt(diag(mys.health2.cl)) #cluster standard errors

mys.pucca2=lm(pucca_binary~borderplac1+TOT_POP+
       TOT_SC+TOT_ST+Slope+TerrainRug+Latitude+Longitude, data=rd10.mb)
summary(mys.pucca2)
```

```
##
## Call:
## lm(formula = pucca_binary ~ borderplac1 + TOT_POP + TOT_SC +
##     TOT_ST + Slope + TerrainRug + Latitude + Longitude, data = rd10.mb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.03623 -0.01632  0.16068  0.25919  0.52007
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.479e-01  3.157e+00  -0.047 0.962656
## borderplac1  3.281e-02  2.786e-02   1.178 0.239244
## TOT_POP      1.322e-04  2.129e-05   6.212 7.41e-10 ***
## TOT_SC       2.330e-04  4.836e-05   4.817 1.66e-06 ***
## TOT_ST       2.322e-04  6.213e-05   3.738 0.000195 ***
## Slope        3.137e-03  1.084e-03   2.894 0.003880 **
## TerrainRug  -7.225e-03  7.170e-03  -1.008 0.313837
## Latitude     1.606e-02  8.692e-02   0.185 0.853428
## Longitude    2.421e-03  4.324e-02   0.056 0.955369
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4022 on 1101 degrees of freedom
## Multiple R-squared:  0.09589,    Adjusted R-squared:  0.08932
## F-statistic:  14.6 on 8 and 1101 DF,  p-value: < 2.2e-16
```

```
mys.pucca2.cl=cluster.vcov(mys.pucca2, rd10.mb$dist_name)
mys.pucca2.se=sqrt(diag(mys.pucca2.cl)) #cluster standard errors



########################################
#border2

#####Calculate new distances (-10 km from the princely state)
```

```
#hyd part
summary(data$NEAR_DIST_border2[data$border2==1])
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
##    307.7  21747.8  46276.5  49566.0  73561.2 139213.6     6425
```

```
data$distplacebo2[(data$border2==1)&!is.na(data$border2)]=
  data$NEAR_DIST_border2[(data$border2==1)&!is.na(data$border2)]+10000

summary(data$distplacebo2[data$border2==1])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   10308   31748   56276   59566   83561  149214    6425
```

```
#bombay part
data$distplacebo2[(data$border2==0)&!is.na(data$border2)
                  &(data$NEAR_DIST_border2<10000)]=
  10000-data$NEAR_DIST_border2[(data$border2==0)&!is.na(data$border2)
                               &(data$NEAR_DIST_border2<10000)]

data$distplacebo2[(data$border2==0)&!is.na(data$border2)
                  &(data$NEAR_DIST_border2>=10000)]=
  data$NEAR_DIST_border2[(data$border2==0)&!is.na(data$border2)
                         &(data$NEAR_DIST_border2>=10000)]-10000

summary(data$NEAR_DIST_border2[data$border2==0])
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
##    235.7  17397.5  40491.5  45437.5  74356.1 111267.2     6425
```

```
summary(data$distplacebo2[data$border2==0])
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.      NA's
##     19.45   8982.04  30491.49  37324.40  64356.14 101267.22      6425
```

```
#create a new treatment variable
#old mysore and chunk of the bombay are in the treatment,
#and the rest of bombay are in teh control group
data$borderplac2[data$border2==1]=1
data$borderplac2[(data$border2==0)&!is.na(data$border2)
                 &(data$NEAR_DIST_border2<10000)]=1
data$borderplac2[(data$border2==0)&!is.na(data$border2)
                 &(data$NEAR_DIST_border2>=10000)]=0
summary(data$borderplac2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.000   0.000   1.000   0.641   1.000   1.000    6425
```

```
#Distance to Hyderabad-Bombay Border
rd10.hb=data[which(data$distplacebo2<10000),] #20 km
table(rd10.hb$borderplac2)
```

```
##
##   0   1
## 293 447
```

```
#hyderabad
```

```r
hyd.health1=lm(health_binary~borderplac2+TOT_POP+
        TOT_SC+TOT_ST+Slope+TerrainRug+Latitude+Longitude, data=rd10.hb)
summary(hyd.health1)
```

```
##
## Call:
## lm(formula = health_binary ~ borderplac2 + TOT_POP + TOT_SC +
##     TOT_ST + Slope + TerrainRug + Latitude + Longitude, data = rd10.hb)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7378 -0.3469 -0.2292  0.4792  0.9156
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.010e+00  7.881e+00  -0.255   0.7988
## borderplac2 -9.133e-03  3.814e-02  -0.239   0.8108
## TOT_POP      1.109e-04  1.715e-05   6.468 1.82e-10 ***
## TOT_SC       1.088e-04  5.169e-05   2.104   0.0357 *
## TOT_ST       1.838e-04  7.360e-05   2.498   0.0127 *
## Slope       -4.688e-04  1.006e-03  -0.466   0.6413
## TerrainRug  -1.031e-02  2.011e-02  -0.513   0.6083
## Latitude    -2.376e-02  4.253e-02  -0.559   0.5765
## Longitude    3.331e-02  1.095e-01   0.304   0.7611
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4579 on 731 degrees of freedom
## Multiple R-squared:  0.08649,    Adjusted R-squared:  0.07649
## F-statistic: 8.651 on 8 and 731 DF,  p-value: 2.558e-11
```

```r
hyd.health1.cl=cluster.vcov(hyd.health1, rd10.hb$dist_name)
hyd.health1.se=sqrt(diag(hyd.health1.cl)) #cluster standard errors

hyd.pucca1=lm(pucca_binary~borderplac2+TOT_POP+
        TOT_SC+TOT_ST+Slope+TerrainRug+Latitude+Longitude, data=rd10.hb)
summary(hyd.pucca1)
```

```
##
## Call:
## lm(formula = pucca_binary ~ borderplac2 + TOT_POP + TOT_SC +
##     TOT_ST + Slope + TerrainRug + Latitude + Longitude, data = rd10.hb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.01086  0.02143  0.08926  0.13991  0.25204
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.087e+00  5.076e+00   1.593 0.111541
## borderplac2 -1.297e-02  2.456e-02  -0.528 0.597714
## TOT_POP      4.268e-05  1.105e-05   3.864 0.000122 ***
## TOT_SC       4.674e-05  3.329e-05   1.404 0.160797
## TOT_ST       8.877e-05  4.740e-05   1.873 0.061501 .
```

```
## Slope       -4.650e-04  6.479e-04  -0.718 0.473208
## TerrainRug   2.770e-02  1.295e-02   2.139 0.032764 *
## Latitude    -1.608e-02  2.739e-02  -0.587 0.557391
## Longitude   -9.228e-02  7.053e-02  -1.308 0.191149
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2949 on 731 degrees of freedom
## Multiple R-squared:  0.05689,    Adjusted R-squared:  0.04656
## F-statistic: 5.511 on 8 and 731 DF,  p-value: 8.832e-07
```

```
hyd.pucca1.cl=cluster.vcov(hyd.pucca1, rd10.hb$dist_name)
hyd.pucca1.se=sqrt(diag(hyd.pucca1.cl)) #cluster standard errors


####Calculate new distances (+10 km to the princely state)

#hyd part
summary(data$NEAR_DIST_border2[data$border2==0])
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
##    235.7  17397.5  40491.5  45437.5  74356.1 111267.2     6425
```

```
data$distplacebo2[(data$border2==0)&!is.na(data$border2)]=
  data$NEAR_DIST_border2[(data$border2==0)&!is.na(data$border2)]+10000

summary(data$distplacebo2[data$border2==0])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   10236   27397   50491   55437   84356  121267    6425
```

```
#bombay part
data$distplacebo2[(data$border2==1)&!is.na(data$border2)
                  &(data$NEAR_DIST_border2<10000)]=
  10000-data$NEAR_DIST_border2[(data$border2==1)&!is.na(data$border2)
                               &(data$NEAR_DIST_border2<10000)]

data$distplacebo2[(data$border2==1)&!is.na(data$border2)
                  &(data$NEAR_DIST_border2>=10000)]=
  data$NEAR_DIST_border2[(data$border2==1)&!is.na(data$border2)
                         &(data$NEAR_DIST_border2>=10000)]-10000

summary(data$NEAR_DIST_border2[data$border2==1])
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
##    307.7  21747.8  46276.5  49566.0  73561.2 139213.6     6425
```

```
summary(data$distplacebo2[data$border2==1])
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.      NA's
##      0.82  11747.78  36276.47  41199.57  63561.22 129213.56      6425
```

```
#create a new treatment variable
#old mysore and chunk of the bombay are in the treatment,
#and the rest of bombay are in teh control group
data$borderplac2[data$border2==0]=1
data$borderplac2[(data$border2==1)&!is.na(data$border2)
```

```
                &(data$NEAR_DIST_border2<10000)]=1
data$borderplac2[(data$border2==1)&!is.na(data$border2)
                &(data$NEAR_DIST_border2>=10000)]=0
summary(data$borderplac2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.000   0.000   1.000   0.511   1.000   1.000    6425
```

```
#Distance to Hyderabad-Bombay Border
rd10.hb=data[which(data$distplacebo2<10000),] #20 km
table(rd10.hb$borderplac2)
```

```
##
##   0   1
## 326 493
```

```
#hyderabad
hyd.health2=lm(health_binary~borderplac2+TOT_POP+
        TOT_SC+TOT_ST+Slope+TerrainRug+Latitude+Longitude, data=rd10.hb)
summary(hyd.health2)
```

```
##
## Call:
## lm(formula = health_binary ~ borderplac2 + TOT_POP + TOT_SC +
##     TOT_ST + Slope + TerrainRug + Latitude + Longitude, data = rd10.hb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.60432 -0.26117 -0.15722  0.03745  0.98974
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.306e+01  7.321e+00   3.150 0.001691 **
## borderplac2 -5.162e-02  3.272e-02  -1.577 0.115120
## TOT_POP      5.248e-05  1.727e-05   3.038 0.002457 **
## TOT_SC       3.319e-04  4.763e-05   6.970 6.59e-12 ***
## TOT_ST       1.950e-04  5.476e-05   3.561 0.000391 ***
## Slope        1.236e-04  1.096e-03   0.113 0.910253
## TerrainRug   1.594e-02  1.574e-02   1.013 0.311505
## Latitude     1.281e-01  3.805e-02   3.368 0.000793 ***
## Longitude   -3.296e-01  1.017e-01  -3.240 0.001244 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3985 on 810 degrees of freedom
## Multiple R-squared:  0.1119, Adjusted R-squared:  0.1031
## F-statistic: 12.76 on 8 and 810 DF,  p-value: < 2.2e-16
```

```
hyd.health2.cl=cluster.vcov(hyd.health2, rd10.hb$dist_name)
hyd.health2.se=sqrt(diag(hyd.health2.cl)) #cluster standard errors

hyd.pucca2=lm(pucca_binary~borderplac2+TOT_POP+
        TOT_SC+TOT_ST+Slope+TerrainRug+Latitude+Longitude, data=rd10.hb)
summary(hyd.pucca2)
```

```
##
```

```
## Call:
## lm(formula = pucca_binary ~ borderplac2 + TOT_POP + TOT_SC +
##     TOT_ST + Slope + TerrainRug + Latitude + Longitude, data = rd10.hb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.02943  0.03998  0.10971  0.16980  0.39534
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.711e+00  6.080e+00   0.281  0.77850
## borderplac2  1.015e-02  2.718e-02   0.374  0.70886
## TOT_POP      6.068e-05  1.435e-05   4.229 2.61e-05 ***
## TOT_SC       1.167e-04  3.956e-05   2.949  0.00328 **
## TOT_ST       7.076e-05  4.548e-05   1.556  0.12012
## Slope       -1.364e-03  9.106e-04  -1.498  0.13452
## TerrainRug  -1.404e-02  1.307e-02  -1.074  0.28310
## Latitude     6.979e-02  3.160e-02   2.209  0.02749 *
## Longitude   -2.608e-02  8.450e-02  -0.309  0.75768
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.331 on 810 degrees of freedom
## Multiple R-squared:  0.0609, Adjusted R-squared:  0.05162
## F-statistic: 6.566 on 8 and 810 DF,  p-value: 2.534e-08
```

```
hyd.pucca2.cl=cluster.vcov(hyd.pucca2, rd10.hb$dist_name)
hyd.pucca2.se=sqrt(diag(hyd.pucca2.cl)) #cluster standard errors




stargazer(mys.health1, mys.pucca1, mys.health2, mys.pucca2,
          se=list(mys.health1.se, mys.pucca1.se, mys.health2.se, mys.pucca2.se), digits=3,
          omit=c("TOT_POP", "TOT_SC", "TOT_ST", "Slope", "TerrainRug", "Latitude", "Longitude"),
          dep.var.labels=c("Health Centers", "Paved Roads", "Health Centers", "Paved Roads"),
          column.labels =c("-10 km", "+10 km"), column.separate = c(2,2),
          covariate.labels = c("Placebo Indirect Rule", "Constant"),
          add.lines = list(c("Controls","\\checkmark","\\checkmark","\\checkmark","\\checkmark")),
          omit.stat = c("rsq", "f", "adj.rsq", "ser"))
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Fri, Jun 16, 2023 - 15:10:45
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lcccc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{4}{c}{\textit{Dependent variable:}} \\
## \cline{2-5}
## \\[-1.8ex] & Health Centers & Paved Roads & Health Centers & Paved Roads \\
##  & \multicolumn{2}{c}{-10 km} & \multicolumn{2}{c}{+10 km} \\
## \\[-1.8ex] & (1) & (2) & (3) & (4)\\
```

```
## \hline \\[-1.8ex]
##  Placebo Indirect Rule & $-$0.051 & 0.001 & 0.006 & 0.033 \\
##   & (0.032) & (0.027) & (0.007) & (0.021) \\
##   & & & & \\
##  Constant & $-$8.571$^{***}$ & $-$8.057$^{***}$ & $-$9.155 & $-$0.148 \\
##   & (0.474) & (2.923) & (5.599) & (0.970) \\
##   & & & & \\
## \hline \\[-1.8ex]
## Controls & \checkmark & \checkmark & \checkmark & \checkmark \\
## Observations & 1,014 & 1,014 & 1,110 & 1,110 \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{4}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
```

```
stargazer(hyd.health1, hyd.pucca1, hyd.health2, hyd.pucca2,
          se=list(hyd.health1.se, hyd.pucca1.se, hyd.health2.se, hyd.pucca2.se), digits=3,
          omit=c("TOT_POP", "TOT_SC", "TOT_ST", "Slope", "TerrainRug", "Latitude", "Longitude"),
          dep.var.labels=c("Health Centers", "Paved Roads", "Health Centers", "Paved Roads"),
          column.labels =c("-10 km", "+10 km"), column.separate = c(2,2),
          covariate.labels = c("Placebo Indirect Rule", "Constant"),
          add.lines = list(c("Controls","\\checkmark","\\checkmark","\\checkmark","\\checkmark")),
          omit.stat = c("rsq", "f", "adj.rsq", "ser"))
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Fri, Jun 16, 2023 - 15:10:45
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lcccc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{4}{c}{\textit{Dependent variable:}} \\
## \cline{2-5}
## \\[-1.8ex] & Health Centers & Paved Roads & Health Centers & Paved Roads \\
##  & \multicolumn{2}{c}{-10 km} & \multicolumn{2}{c}{+10 km} \\
## \\[-1.8ex] & (1) & (2) & (3) & (4)\\
## \hline \\[-1.8ex]
##  Placebo Indirect Rule & $-$0.009 & $-$0.013 & $-$0.052$^{***}$ & 0.010 \\
##   & (0.041) & (0.028) & (0.011) & (0.020) \\
##   & & & & \\
##  Constant & $-$2.010 & 8.087$^{***}$ & 23.063$^{**}$ & 1.711 \\
##   & (3.585) & (1.776) & (10.229) & (8.803) \\
##   & & & & \\
## \hline \\[-1.8ex]
## Controls & \checkmark & \checkmark & \checkmark & \checkmark \\
## Observations & 740 & 740 & 819 & 819 \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{4}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
```