Advanced Impact Evaluation Term Paper

Natalia Esquenazi

2024-03-25

Import libraries and dataset

```
setwd("C:/Users/usuario/Desktop/Masters Degree CEU/Advanced Impact Evaluation/Term Paper")

library(tidyverse)
library(dplyr)
library(readr)
library(estimatr)
library(ggplot2)
library(hrbrthemes)
library(texreg)
library(MatchIt)
library(Matching)
library(cobalt)

df = read_csv("C:/Users/usuario/Desktop/Masters Degree CEU/Advanced Impact Evaluation/Term Paper/ENFR_ba
```

Data Transformation

```
# Marriage
df <- df %>%
  mutate(marital_status = case_when(marital_status == 1 ~ "single",
                                    marital_status == 2 ~ "married",
                                    marital_status == 3 ~ "single",
                                    marital_status == 4 ~ "divorced",
                                    marital_status == 5 ~ "widow",
                                    marital_status == 6 ~ "single"))

# employment
df <- df %>%
  mutate(employment = case_when(employment == 1 ~ "employed",
                                employment == 2 ~ "unemployed",
                                employment == 3 ~ "inactive"))

# education levels
df <- df %>%
  mutate(education = case_when(education == 1 ~ "no education",
                               education == 2 ~ "primary incomplete",
```

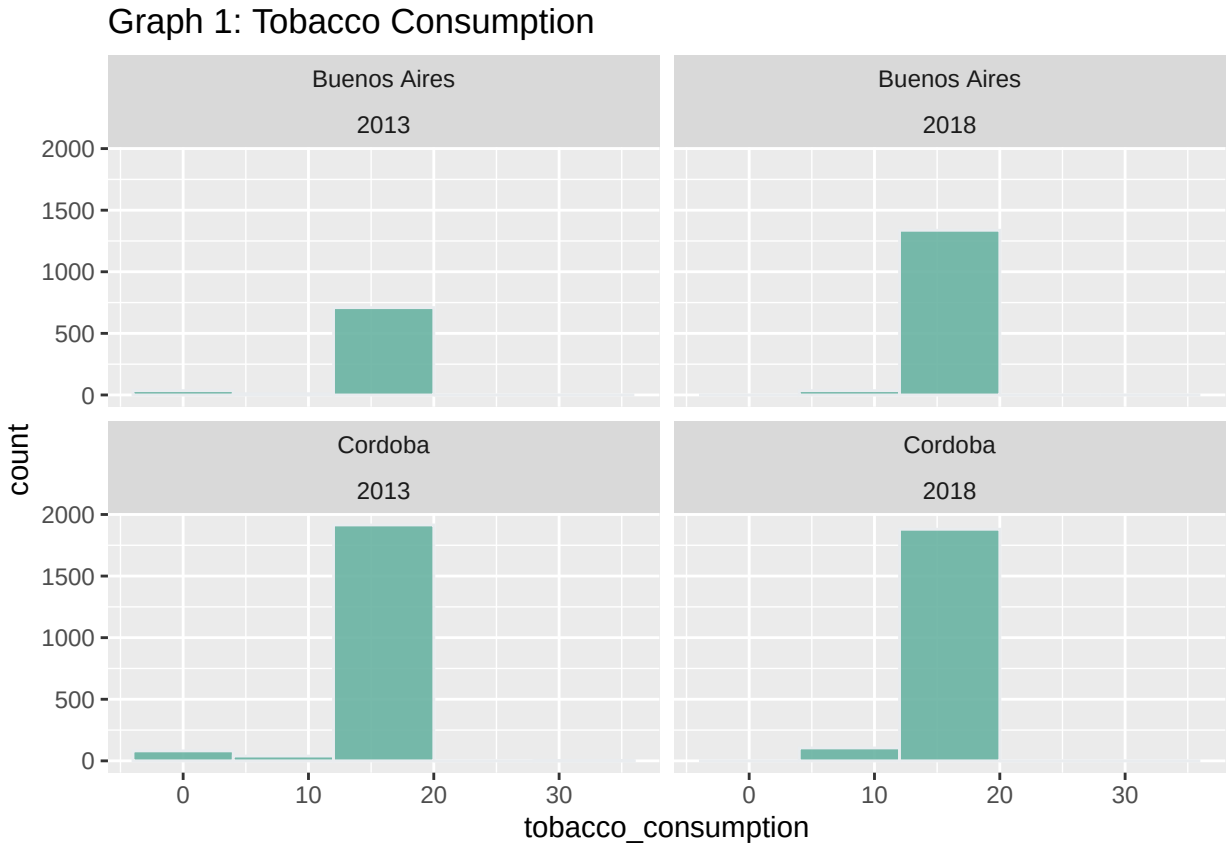
```
                                education == 3 ~ "primary complete",
                                education == 4 ~ "secondary incomplete",
                                education == 5 ~ "secondary complete",
                                education == 6 ~ "university incomplete",
                                education == 7 ~ "university complete",
                                education == 8 ~ "special education"))

# gender
df <- df %>%
  mutate(gender = case_when(gender == 1 ~ "male",
                            gender == 2 ~ "female"))

# Age range
df <- df %>%
  mutate(age_range = case_when(age >= 18 & age <= 24 ~ "18-24",
                               age >= 25 & age <= 34 ~ "25-34",
                               age >= 35 & age <= 49 ~ "35-49",
                               age >= 50 & age <= 64 ~ "50-64",
                               age >= 65 ~ "65+"))


# Transform variables class
df$employment <- as.factor(df$employment)
df$education <- as.factor(df$education)
df$tobacco_consumption <- as.numeric(df$tobacco_consumption)
df$income_quintile <- as.numeric(df$income_quintile)
df$year <- as.factor(df$year)
df$state <- as.factor(df$state)
df$region <- as.factor(df$region)
df$gender <- as.factor(df$gender)
df$age_range <- as.factor(df$age_range)

summary(df)
```

```
##       ...1            id                      state      region
##  Min.   :   1   Min.   :4.509e+04   Buenos Aires:2126   1:2126
##  1st Qu.:1538   1st Qu.:6.913e+08   Cordoba     :4021   2:4021
##  Median :3074   Median :1.375e+09
##  Mean   :3074   Mean   :1.555e+09
##  3rd Qu.:4610   3rd Qu.:2.077e+09
##  Max.   :6147   Max.   :4.293e+09
##
##  monthly_income    income_range   income_quintile    gender          age
##  Min.   :     0   Min.   : 0.00   Min.   :1.000   female:3445   Min.   :18.00
##  1st Qu.:  5000   1st Qu.: 6.00   1st Qu.:2.000   male  :2702   1st Qu.:31.00
##  Median : 12000   Median :10.00   Median :4.000                 Median :44.00
##  Mean   : 18979   Mean   :19.61   Mean   :3.389                 Mean   :47.21
##  3rd Qu.: 25000   3rd Qu.:13.00   3rd Qu.:5.000                 3rd Qu.:62.00
##  Max.   :300000   Max.   :99.00   Max.   :5.000                 Max.   :98.00
##                   NA's   :1549    NA's   :10
##  marital_status                    education         employment
##  Length:6147        university complete  :1657   employed  :4188
##  Class :character   secondary complete   :1308   inactive  :1767
```

```
##  Mode  :character   primary complete     : 998   unemployed: 192
##                     university incomplete: 860
##                     secondary incomplete : 822
##                     primary incomplete   : 429
##                     (Other)              :  73
##   min_con_age    tobacco_consumption     weight         year      age_range
##  Min.   : 5.00   Min.   : 0.0        Min.   : 39.00   2013:2790   18-24: 693
##  1st Qu.:15.00   1st Qu.:14.0        1st Qu.: 62.00   2018:3357   25-34:1274
##  Median :16.00   Median :14.0        Median : 72.00               35-49:1539
##  Mean   :17.68   Mean   :13.9        Mean   : 82.84               50-64:1279
##  3rd Qu.:18.00   3rd Qu.:14.0        3rd Qu.: 84.00               65+  :1362
##  Max.   :99.00   Max.   :30.0        Max.   :999.00
##  NA's   :2963                        NA's   :95
```

Outcome variable

```
# Outcome variable
df %>%
    ggplot( aes(x=tobacco_consumption)) +
    geom_histogram( binwidth=8, fill="#69b3a2", color="#e9ecef", alpha=0.9, ) +
    ggtitle("Graph 1: Tobacco Consumption") +
  facet_wrap( ~ state + year)
```



```r
df %>%
  group_by(state, year)%>%
  summarise(avg = mean(tobacco_consumption))
```

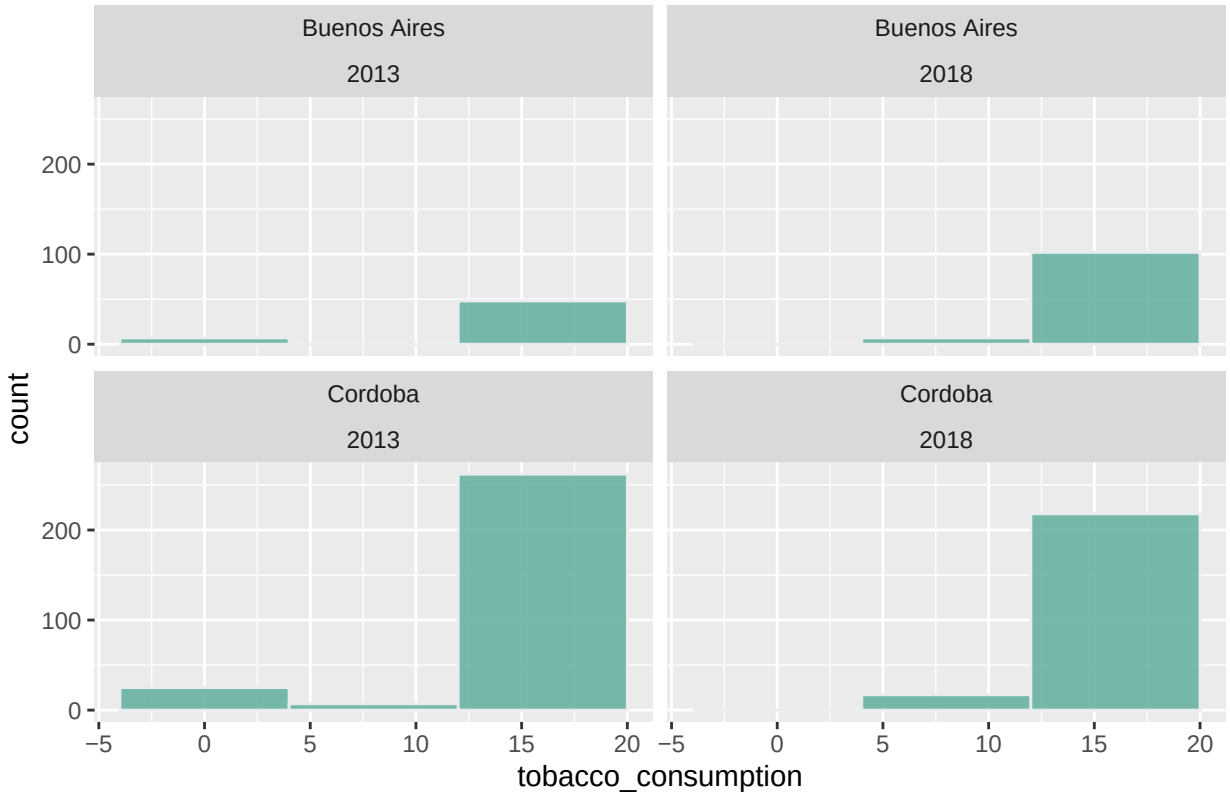
```
## # A tibble: 4 x 3
## # Groups:   state [2]
##   state        year    avg
##   <fct>        <fct> <dbl>
## 1 Buenos Aires 2013   13.3
## 2 Buenos Aires 2018   14.4
## 3 Cordoba      2013   13.4
## 4 Cordoba      2018   14.3
```

Youth population and tobacco consumption analysis

Treatment column

```r
df_young <- df %>%
  filter(age_range == "18-24") %>%
  mutate(treat = case_when(state == "Cordoba" ~ "treat",
                           state == "Buenos Aires" ~ "control"))

# treatment variable
df_young %>%
    ggplot( aes(x=tobacco_consumption)) +
    geom_histogram( binwidth=8, fill="#69b3a2", color="#e9ecef", alpha=0.9, ) +
    ggtitle("Graph 2: Youth population tobacco consumption") +
    facet_wrap( ~ state + year)
```

Graph 2: Youth population tobacco consumption




```
df_young %>%
  group_by(state, year)%>%
  summarise(avg = mean(tobacco_consumption))
```

```
## # A tibble: 4 x 3
## # Groups:   state [2]
##   state        year    avg
##   <fct>        <fct> <dbl>
## 1 Buenos Aires 2013   12.5
## 2 Buenos Aires 2018   14.0
## 3 Cordoba      2013   12.8
## 4 Cordoba      2018   14.2
```

Matching

```
# coarsened exact matching:
df_young <- df_young %>%
  filter(!is.na(weight)) %>%
  filter(!is.na(income_quintile))

match1 <- matchit(treat ~ income_quintile + gender + education + employment + weight,
                  data = df_young, method = "cem",
                  replace = FALSE)
summary(match1)
```

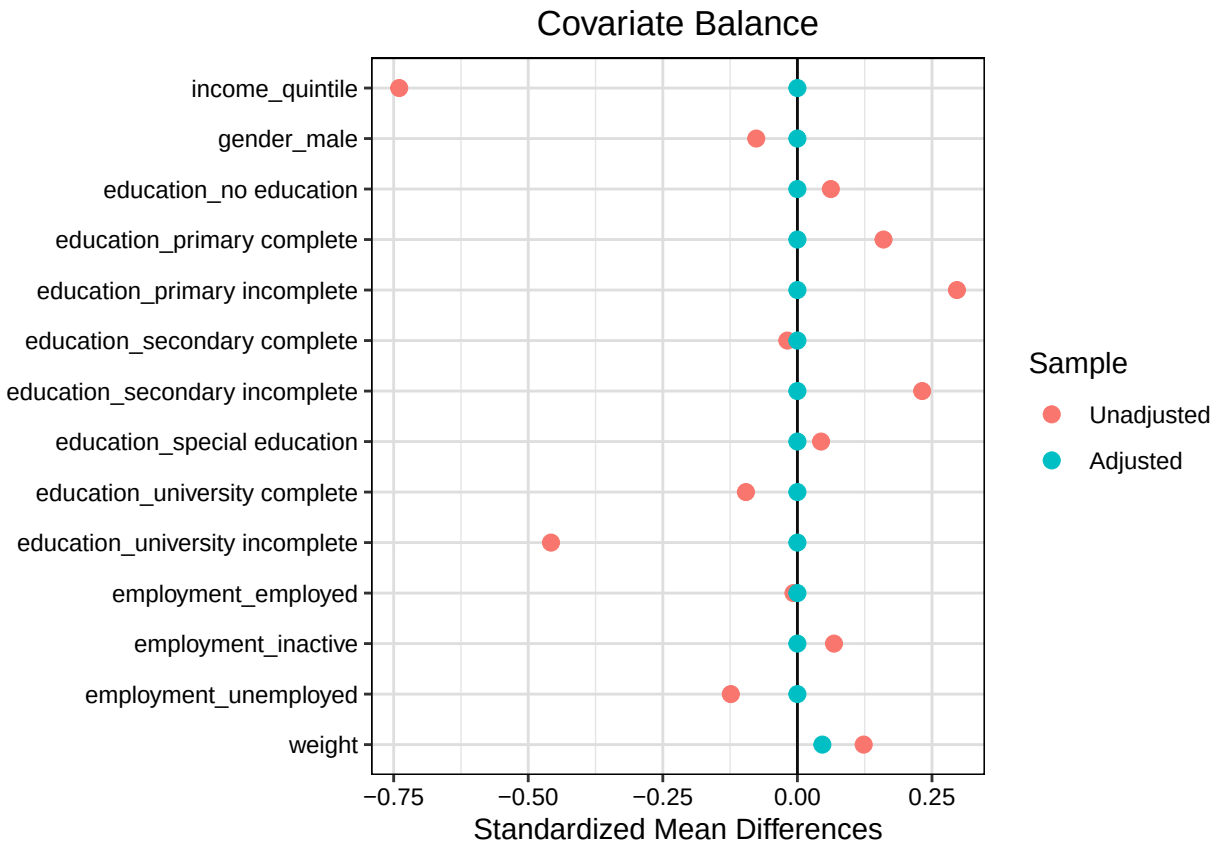
```
##
## Call:
## matchit(formula = treat ~ income_quintile + gender + education +
##     employment + weight, data = df_young, method = "cem", replace = FALSE)
##
## Summary of Balance for All Data:
##                               Means Treated Means Control Std. Mean Diff.
## income_quintile                      2.7135        3.7081         -0.7397
## genderfemale                         0.5288        0.4907          0.0765
## gendermale                           0.4712        0.5093         -0.0765
## educationno education                0.0038        0.0000          0.0621
## educationprimary complete            0.1577        0.0994          0.1600
## educationprimary incomplete          0.0808        0.0000          0.2964
## educationsecondary complete          0.1731        0.1801         -0.0186
## educationsecondary incomplete        0.1904        0.0994          0.2318
## educationspecial education           0.0019        0.0000          0.0439
## educationuniversity complete         0.1577        0.1925         -0.0956
## educationuniversity incomplete       0.2346        0.4286         -0.4577
## employmentemployed                   0.6923        0.6957         -0.0072
## employmentinactive                   0.2596        0.2298          0.0680
## employmentunemployed                 0.0481        0.0745         -0.1237
## weight                              73.8115       64.9255          0.1231
##                               Var. Ratio eCDF Mean eCDF Max
## income_quintile                   0.9730    0.1989   0.3019
## genderfemale                           .    0.0382   0.0382
## gendermale                             .    0.0382   0.0382
## educationno education                  .    0.0038   0.0038
## educationprimary complete              .    0.0583   0.0583
## educationprimary incomplete            .    0.0808   0.0808
## educationsecondary complete            .    0.0070   0.0070
## educationsecondary incomplete          .    0.0910   0.0910
## educationspecial education             .    0.0019   0.0019
## educationuniversity complete           .    0.0349   0.0349
## educationuniversity incomplete         .    0.1940   0.1940
## employmentemployed                     .    0.0033   0.0033
## employmentinactive                     .    0.0298   0.0298
## employmentunemployed                   .    0.0265   0.0265
## weight                           34.5928    0.0529   0.1638
##
## Summary of Balance for Matched Data:
##                               Means Treated Means Control Std. Mean Diff.
## income_quintile                      3.1698        3.1698          0.0000
## genderfemale                         0.5123        0.5123          0.0000
## gendermale                           0.4877        0.4877         -0.0000
## educationno education                0.0000        0.0000          0.0000
## educationprimary complete            0.1235        0.1235         -0.0000
## educationprimary incomplete          0.0000        0.0000          0.0000
## educationsecondary complete          0.1481        0.1481          0.0000
## educationsecondary incomplete        0.1883        0.1883          0.0000
## educationspecial education           0.0000        0.0000          0.0000
## educationuniversity complete         0.1698        0.1698          0.0000
## educationuniversity incomplete       0.3704        0.3704          0.0000
## employmentemployed                   0.7593        0.7593          0.0000
## employmentinactive                   0.2315        0.2315          0.0000
```

```
## employmentunemployed                  0.0093        0.0093          0.0000
## weight                               68.5278       65.1822          0.0464
##                                Var. Ratio eCDF Mean eCDF Max Std. Pair Dist.
## income_quintile                   0.9903     0.000    0.000          0.0000
## genderfemale                           .     0.000    0.000          0.0000
## gendermale                             .     0.000    0.000          0.0000
## educationno education                  .     0.000    0.000          0.0000
## educationprimary complete              .     0.000    0.000          0.0000
## educationprimary incomplete            .     0.000    0.000          0.0000
## educationsecondary complete            .     0.000    0.000          0.0000
## educationsecondary incomplete          .     0.000    0.000          0.0000
## educationspecial education             .     0.000    0.000          0.0000
## educationuniversity complete           .     0.000    0.000          0.0000
## educationuniversity incomplete         .     0.000    0.000          0.0000
## employmentemployed                     .     0.000    0.000          0.0000
## employmentinactive                     .     0.000    0.000          0.0000
## employmentunemployed                   .     0.000    0.000          0.0000
## weight                            1.8078     0.057    0.181          0.1713
##
## Sample Sizes:
##                Control Treated
## All             161.       520
## Matched (ESS)    78.42     324
## Matched         147.       324
## Unmatched        14.       196
## Discarded         0.         0
```

```
match1
```

```
## A matchit object
##  - method: Coarsened exact matching
##  - number of obs.: 681 (original), 471 (matched)
##  - target estimand: ATT
##  - covariates: income_quintile, gender, education, employment, weight
```

```
love1 <- love.plot(match1, stars ="std",
                   stats = c("mean.diffs"),
                   binary = "std", abs = FALSE, grid = TRUE)
love1
```



Multiple regression

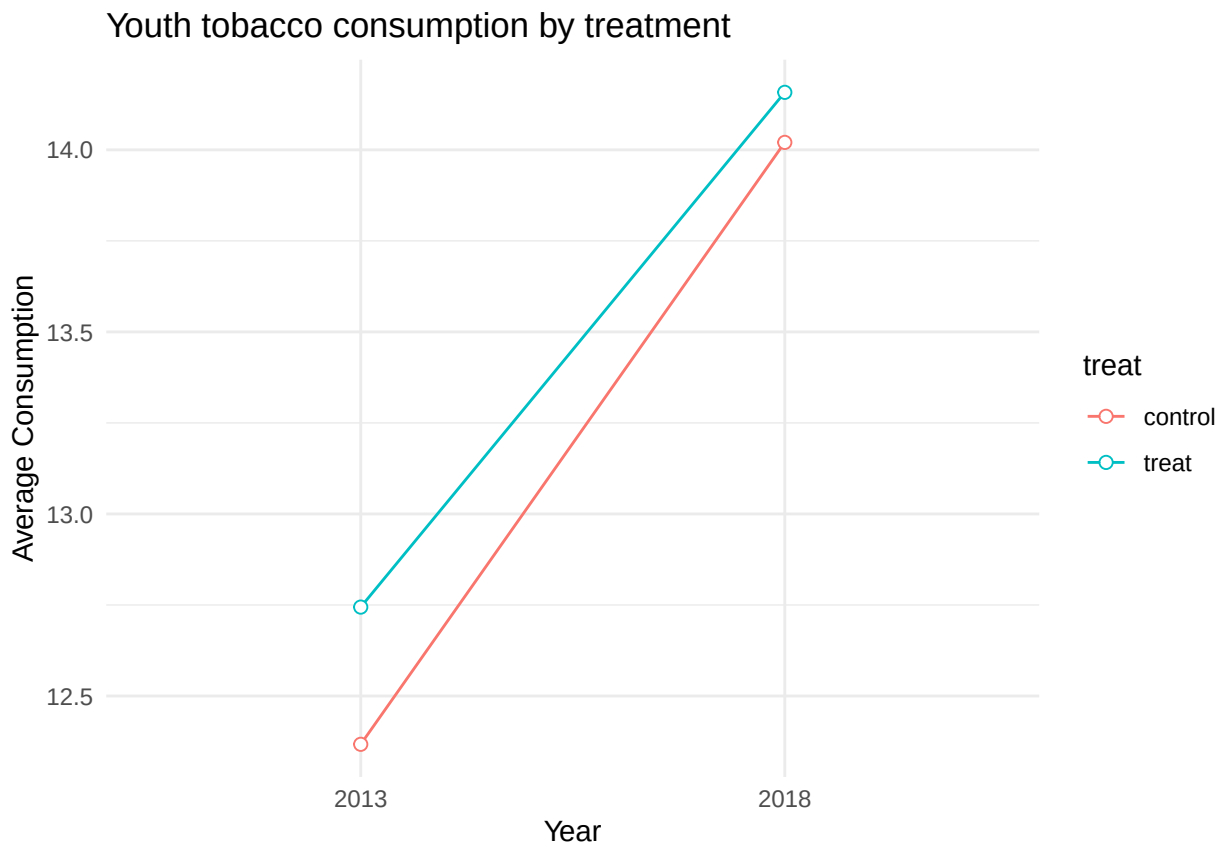
```
# create dataset
match_dat1 <- match.data(match1)

mod1 <- lm_robust(tobacco_consumption ~ treat, data = match_dat1)

mod2 <- lm_robust(tobacco_consumption ~ treat + year + income_quintile + employment, data = match_dat1)

mod3 <- lm_robust(tobacco_consumption ~ treat + year + income_quintile + employment + gender + educatio

screenreg(list(mod1, mod2, mod3))
```

```
##
## ============================================================================
##                                 Model 1         Model 2         Model 3
## ----------------------------------------------------------------------------
## (Intercept)                      13.47 *         12.09 *         12.19 *
##                                 [12.98; 13.96]  [10.89; 13.29]  [10.44; 13.93]
## treattreat                       -0.06           0.29            0.19
##                                 [-0.66;  0.53]  [-0.34;  0.93]  [-0.43;  0.81]
## year2018                                         1.49 *          1.53 *
##                                                 [ 0.90;  2.08]  [ 0.93;  2.13]
## income_quintile                                  0.13            0.16
```

```
##                                                [-0.09;  0.35]  [-0.07;  0.40]
## employmentinactive                              -0.30           0.10
##                                                [-0.98;  0.37]  [-0.65;  0.85]
## employmentunemployed                            -1.49          -1.05
##                                                [-5.56;  2.57]  [-5.01;  2.92]
## gendermale                                                      -0.61 *
##                                                                [-1.18; -0.03]
## educationsecondary complete                                     -0.23
##                                                                [-1.13;  0.67]
## educationsecondary incomplete                                    0.26
##                                                                [-0.60;  1.13]
## educationuniversity complete                                    -0.16
##                                                                [-1.19;  0.86]
## educationuniversity incomplete                                  -0.74
##                                                                [-1.64;  0.16]
## weight                                                           0.01
##                                                                [-0.01;  0.02]
## ---------------------------------------------------------------------------------
## R^2                                 0.00            0.06            0.08
## Adj. R^2                           -0.00            0.05            0.06
## Num. obs.                         471             471             471
## RMSE                                3.07            2.98            2.97
## =================================================================================
## * Null hypothesis value outside the confidence interval.
```

Difference in differences

```
# parallel trends
tab1 <- match_dat1 %>%
  group_by(year, treat) %>%
  summarise(avg_consumption = mean(tobacco_consumption))

ggplot(tab1, aes(x=year, y=avg_consumption, group = treat, colour = treat)) +
  geom_line()  +
  geom_point( size=2, shape=21, fill="white") +
  theme_minimal() +
  labs(title = "Youth tobacco consumption by treatment",
       y = "Average Consumption",
       x = "Year")
```




```
match_dat1 <- match_dat1 %>%
  mutate(treat = case_when(treat == "control" ~ 0,
                           treat == "treat" ~ 1),
         time = case_when(year == "2013"~ 0,
                          year == "2018"~ 1))

match_dat1$treat <- as.numeric(match_dat1$treat)
match_dat1$year <- as.numeric(match_dat1$year)

mod4 <- lm_robust(tobacco_consumption ~ treat + time + I(treat*time), data = match_dat1)
screenreg(mod4)
```

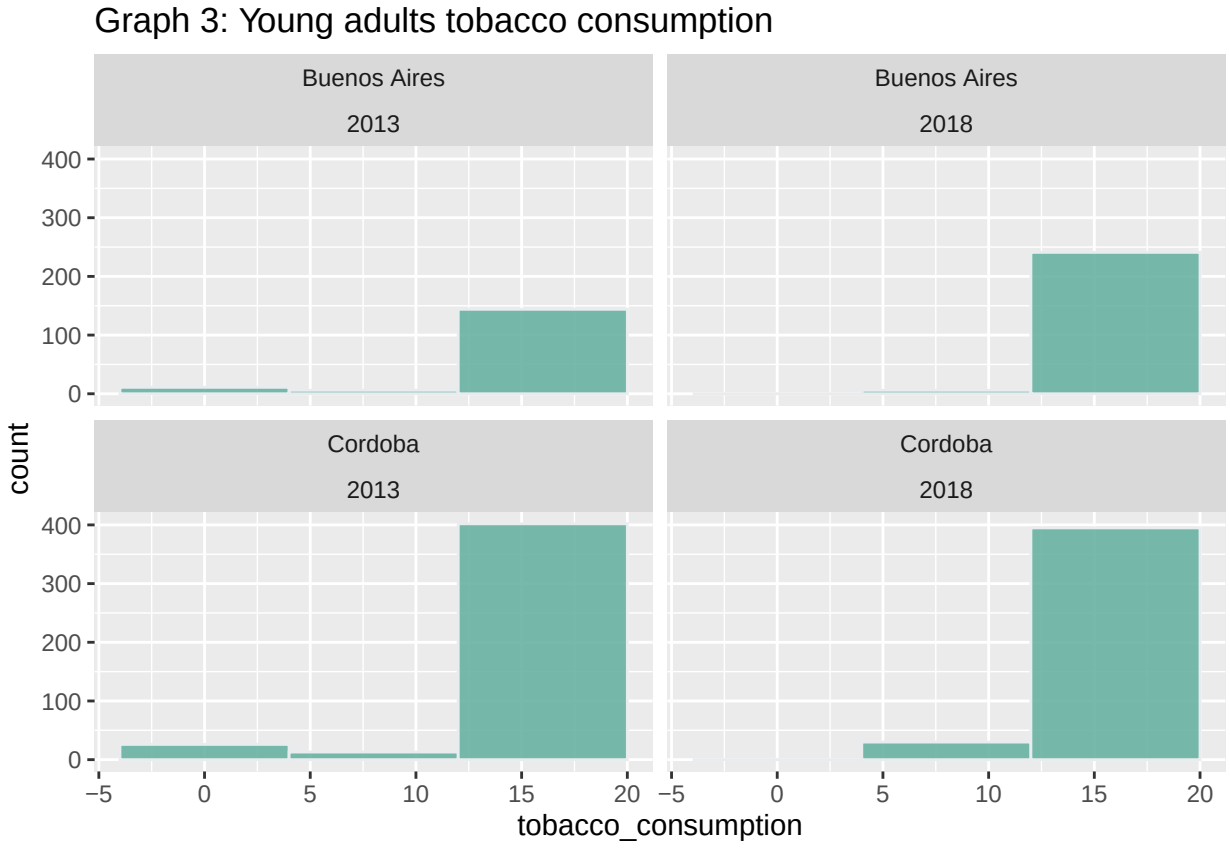
```
##
## ============================
##               Model 1
## ----------------------------
## (Intercept)    12.37 *
##               [11.11; 13.63]
## treat           0.38
##               [-0.99;  1.74]
## time            1.65 *
##               [ 0.35;  2.96]
## treat * time   -0.24
##               [-1.69;  1.21]
## ----------------------------
```

```
## R^2              0.06
## Adj. R^2         0.05
## Num. obs.      471
## RMSE             2.99
## ============================
## * 0 outside the confidence interval.
```

Young adults and tobacco consumption analysis

Treatment column

```
df_young_adults <- df %>%
  filter(age_range == "25-34") %>%
  mutate(treat = case_when(state == "Cordoba" ~ "treat",
                           state == "Buenos Aires" ~ "control"))

# treatment variable
df_young_adults %>%
    ggplot( aes(x=tobacco_consumption)) +
    geom_histogram( binwidth=8, fill="#69b3a2", color="#e9ecef", alpha=0.9, ) +
    ggtitle("Graph 3: Young adults tobacco consumption") +
    facet_wrap( ~ state + year)
```



```r
df_young_adults %>%
  group_by(state, year)%>%
  summarise(avg = mean(tobacco_consumption))
```

```
## # A tibble: 4 x 3
## # Groups:   state [2]
##   state        year    avg
##   <fct>        <fct> <dbl>
## 1 Buenos Aires 2013   12.9
## 2 Buenos Aires 2018   14.7
## 3 Cordoba      2013   13.1
## 4 Cordoba      2018   14.4
```

Matching

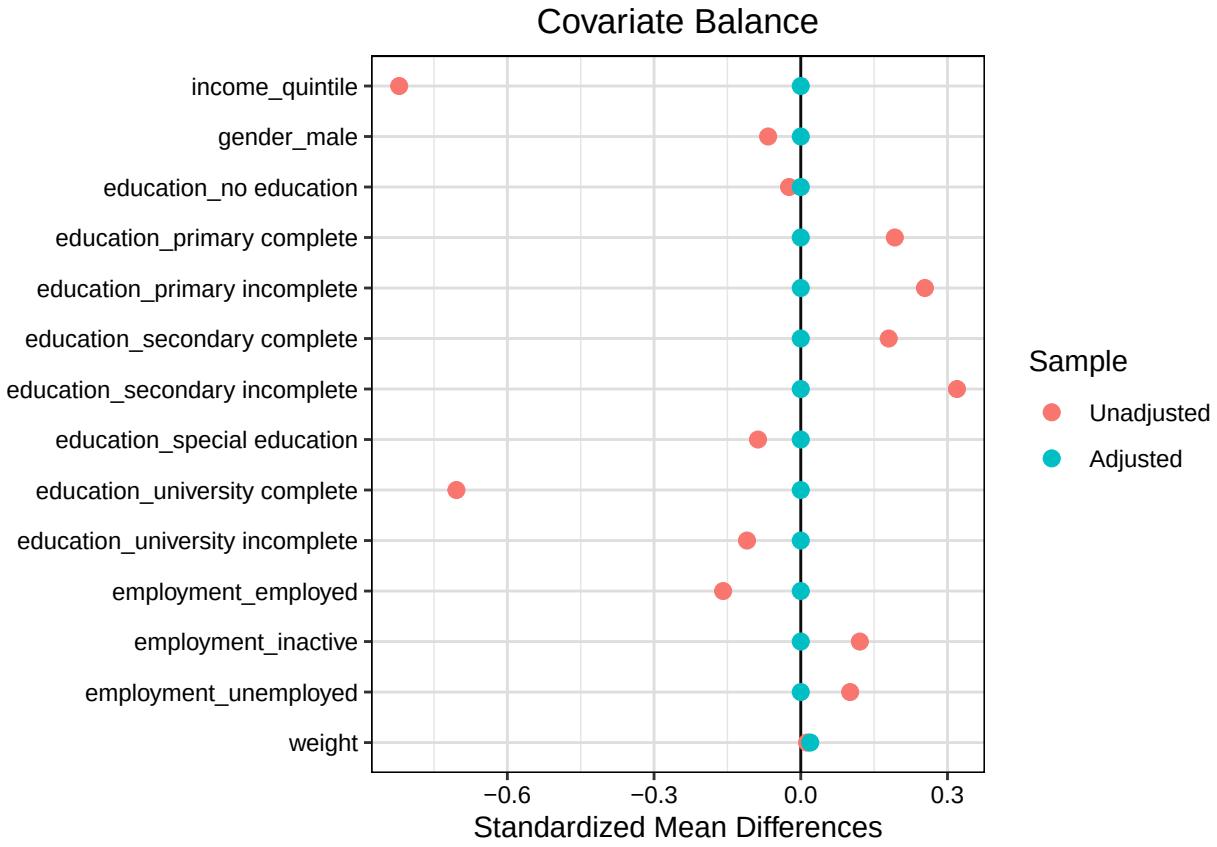
```r
# coarsened exact matching:
df_young_adults <- df_young_adults %>%
  filter(!is.na(weight))%>%
  filter(!is.na(income_quintile))

match2 <- matchit(treat ~ income_quintile + gender + education + employment + weight,
                  data = df_young_adults, method = "cem",
                  replace = FALSE)
summary(match2)
```

```
##
## Call:
## matchit(formula = treat ~ income_quintile + gender + education +
##     employment + weight, data = df_young_adults, method = "cem",
##     replace = FALSE)
##
## Summary of Balance for All Data:
##                                Means Treated Means Control Std. Mean Diff.
## income_quintile                       3.0515        4.1872         -0.8211
## genderfemale                          0.5602        0.5271          0.0668
## gendermale                            0.4398        0.4729         -0.0668
## educationno education                 0.0035        0.0049         -0.0240
## educationprimary complete             0.0994        0.0419          0.1923
## educationprimary incomplete           0.0737        0.0074          0.2538
## educationsecondary complete           0.2749        0.1946          0.1798
## educationsecondary incomplete         0.1825        0.0591          0.3194
## educationspecial education            0.0000        0.0025         -0.0876
## educationuniversity complete          0.2023        0.4852         -0.7041
## educationuniversity incomplete        0.1637        0.2044         -0.1100
## employmentemployed                    0.8409        0.8990         -0.1588
## employmentinactive                    0.1263        0.0862          0.1207
## employmentunemployed                  0.0327        0.0148          0.1010
## weight                               85.1450       83.7906          0.0128
##                                Var. Ratio eCDF Mean eCDF Max
## income_quintile                    1.3137    0.2271   0.3969
## genderfemale                            .    0.0331   0.0331
```

```
## gendermale                               .    0.0331   0.0331
## educationno education                    .    0.0014   0.0014
## educationprimary complete                .    0.0575   0.0575
## educationprimary incomplete              .    0.0663   0.0663
## educationsecondary complete              .    0.0803   0.0803
## educationsecondary incomplete            .    0.1233   0.1233
## educationspecial education               .    0.0025   0.0025
## educationuniversity complete             .    0.2829   0.2829
## educationuniversity incomplete           .    0.0407   0.0407
## employmentemployed                       .    0.0581   0.0581
## employmentinactive                       .    0.0401   0.0401
## employmentunemployed                     .    0.0180   0.0180
## weight                              0.8707    0.0321   0.1122
##
## Summary of Balance for Matched Data:
##                                 Means Treated Means Control Std. Mean Diff.
## income_quintile                        3.2303        3.2303         -0.0000
## genderfemale                           0.5758        0.5758          0.0000
## gendermale                             0.4242        0.4242         -0.0000
## educationno education                  0.0015        0.0015          0.0000
## educationprimary complete              0.0714        0.0714          0.0000
## educationprimary incomplete            0.0175        0.0175          0.0000
## educationsecondary complete            0.3163        0.3163          0.0000
## educationsecondary incomplete          0.1778        0.1778          0.0000
## educationspecial education             0.0000        0.0000          0.0000
## educationuniversity complete           0.2362        0.2362          0.0000
## educationuniversity incomplete         0.1793        0.1793          0.0000
## employmentemployed                     0.9315        0.9315          0.0000
## employmentinactive                     0.0641        0.0641          0.0000
## employmentunemployed                   0.0044        0.0044          0.0000
## weight                                72.0466       69.9955          0.0194
##                                 Var. Ratio eCDF Mean eCDF Max Std. Pair Dist.
## income_quintile                     0.9941    0.0000   0.0000          0.0000
## genderfemale                             .    0.0000   0.0000          0.0000
## gendermale                               .    0.0000   0.0000          0.0000
## educationno education                    .    0.0000   0.0000          0.0000
## educationprimary complete                .    0.0000   0.0000          0.0000
## educationprimary incomplete              .    0.0000   0.0000          0.0000
## educationsecondary complete              .    0.0000   0.0000          0.0000
## educationsecondary incomplete            .    0.0000   0.0000          0.0000
## educationspecial education               .    0.0000   0.0000          0.0000
## educationuniversity complete             .    0.0000   0.0000          0.0000
## educationuniversity incomplete           .    0.0000   0.0000          0.0000
## employmentemployed                       .    0.0000   0.0000          0.0000
## employmentinactive                       .    0.0000   0.0000          0.0000
## employmentunemployed                     .    0.0000   0.0000          0.0000
## weight                              0.9350    0.0241   0.1203          0.1179
##
## Sample Sizes:
##               Control Treated
## All            406.       855
## Matched (ESS)  136.24     686
## Matched        385.       686
## Unmatched       21.       169
```

```
## Discarded        0.         0
```

```
love2 <- love.plot(match2, stars ="std",
                   stats = c("mean.diffs"),
                   binary = "std", abs = FALSE, grid = TRUE)
love2
```



Multiple regression

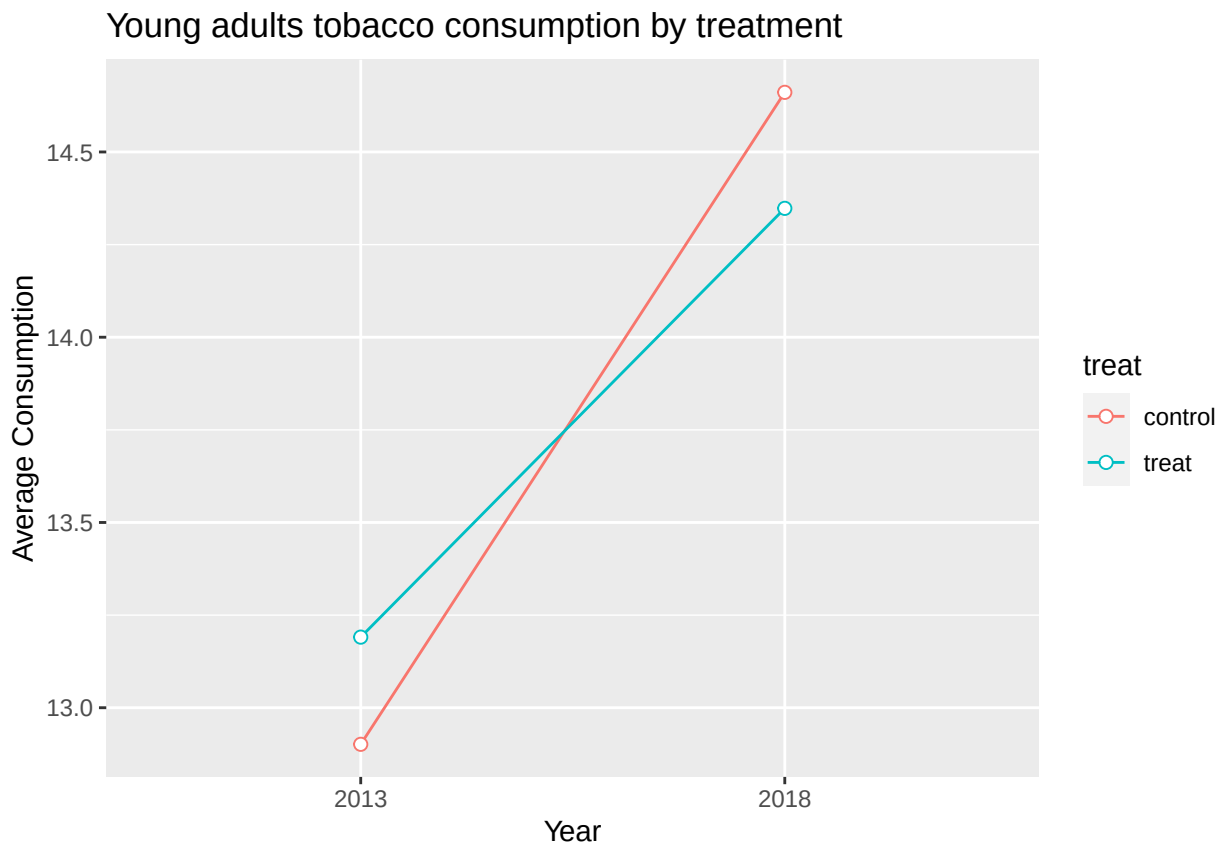
```
# create dataset
match_dat2 <- match.data(match2)

mod5 <- lm_robust(tobacco_consumption ~ treat, data = match_dat2)

mod6 <- lm_robust(tobacco_consumption ~ treat + year + income_quintile + employment, data = match_dat2)

mod7 <- lm_robust(tobacco_consumption ~ treat + year + income_quintile + employment + gender + educatio

screenreg(list(mod5, mod6, mod7))
```

```
##
## ============================================================================
##                                Model 1         Model 2         Model 3
```

```
## ------------------------------------------------------------------------------
## (Intercept)                      13.97 *         12.41 *          7.80
##                                [13.69; 14.25]  [11.65; 13.18]  [-2.94; 18.54]
## treattreat                       -0.19            0.12            0.12
##                                [-0.54;  0.15]  [-0.25;  0.49]  [-0.26;  0.50]
## year2018                                         1.31 *          1.27 *
##                                                [ 1.00;  1.62]  [ 0.96;  1.58]
## income_quintile                                  0.19 *          0.16
##                                                [ 0.05;  0.33]  [-0.00;  0.32]
## employmentinactive                              -0.56           -0.48
##                                                [-1.37;  0.24]  [-1.32;  0.36]
## employmentunemployed                             0.68            0.61
##                                                [-2.53;  3.88]  [-2.65;  3.86]
## gendermale                                                       0.03
##                                                                [-0.32;  0.39]
## educationprimary complete                                        4.16
##                                                                [-6.59; 14.90]
## educationprimary incomplete                                      3.94
##                                                                [-6.88; 14.77]
## educationsecondary complete                                      4.63
##                                                                [-6.07; 15.33]
## educationsecondary incomplete                                    4.48
##                                                                [-6.23; 15.19]
## educationuniversity complete                                     4.66
##                                                                [-6.05; 15.36]
## educationuniversity incomplete                                   4.52
##                                                                [-6.18; 15.23]
## weight                                                           0.00
##                                                                [-0.01;  0.01]
## ------------------------------------------------------------------------------
## R^2                                0.00            0.08            0.08
## Adj. R^2                           0.00            0.07            0.07
## Num. obs.                       1071            1071            1071
## RMSE                               2.70            2.61            2.60
## ==============================================================================
## * Null hypothesis value outside the confidence interval.
```

Difference in differences

```
# parallel trends
tab2 <- match_dat2 %>%
  group_by(year, treat) %>%
  summarise(avg_consumption = mean(tobacco_consumption))

ggplot(tab2, aes(x=year, y=avg_consumption, group = treat, colour = treat)) +
  geom_line()  +
  geom_point( size=2, shape=21, fill="white") +
  labs(title = "Young adults tobacco consumption by treatment",
       y = "Average Consumption",
       x = "Year")
```




```
match_dat2 <- match_dat2 %>%
  mutate(treat = case_when(treat == "control" ~ 0,
                           treat == "treat" ~ 1),
         time = case_when(year == "2013"~ 0,
                          year == "2018"~ 1))

match_dat2$treat <- as.numeric(match_dat2$treat)
match_dat2$year <- as.numeric(match_dat2$year)

mod8 <- lm_robust(tobacco_consumption ~ treat + time + I(treat*time), data = match_dat2)
screenreg(mod8)
```

```
##
## ============================
##               Model 1
## ----------------------------
## (Intercept)     12.90 *
##               [12.37; 13.43]
## treat            0.29
##               [-0.33;  0.90]
## time             1.76 *
##               [ 1.16;  2.36]
## treat * time    -0.60
##               [-1.32;  0.11]
## ----------------------------
```

```
## R^2              0.07
## Adj. R^2         0.06
## Num. obs.     1071
## RMSE             2.62
## ============================
## * 0 outside the confidence interval.
```