Replication: Synthetic Ceará (Figure 06)

```
install.packages("dplyr")
install.packages("ggplot2")
install.packages("Synth")
install.packages("xtable")
install.packages("ggpubr")
```

```
source("functions/plot_scm.R")
source("functions/prepare_p_ls.R")
source("functions/prepare_us.R")
source("functions/prepare_time_placebo.R")

load("data/DATA_COMPLETE.RData")
load("data/abbr_code.RData")
```

```
# Subsetting data:

PRIMARY_M <- as.data.frame(filter(DATA_COMPLETE, grade == "P", subject == "math"))
PRIMARY_P <- as.data.frame(filter(DATA_COMPLETE, grade == "P", subject == "port"))
LOWERS_M <- as.data.frame(filter(DATA_COMPLETE, grade == "LS", subject == "math"))
LOWERS_P <- as.data.frame(filter(DATA_COMPLETE, grade == "LS", subject == "port"))
UPPERS_M <- as.data.frame(filter(DATA_COMPLETE, grade == "US", subject == "math"))
UPPERS_P <- as.data.frame(filter(DATA_COMPLETE, grade == "US", subject == "port"))


# Preparing data for Synth:
DATA_PM <- prepare_p_ls(PRIMARY_M)
DATA_PP <- prepare_p_ls(PRIMARY_P)

DATA_LSM <- prepare_p_ls(LOWERS_M)
DATA_LSP <- prepare_p_ls(LOWERS_P)

DATA_USM <- prepare_us(UPPERS_M)
DATA_USP <- prepare_us(UPPERS_P)


################################################
# SCM for primary and lower secondary school #
################################################

# Run synth
SCM_PM    <- synth(DATA_PM)
```

```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
## ****************
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
```

```
##
## MSPE (LOSS V): 19.73758
##
## solution.v:
##  0.3449389 0.1771021 0.1243577 0.3535089 9.23628e-05
##
## solution.w:
##  4.7371e-06 3.1603e-06 3.136e-06 2.7913e-06 3.4114e-06 2.1615e-06 3.8e-06 4.3647e-06 0.4793483 5.112
```

```
TABLES_PM <- synth.tab(dataprep.res = DATA_PM, synth.res = SCM_PM)

SCM_PP    <- synth(DATA_PP)
```

```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
## ****************
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
##
## MSPE (LOSS V): 9.588537
##
## solution.v:
##  0.1659437 0.388797 0.3242967 0.1038635 0.0170991
##
## solution.w:
##  3.1327e-06 3.4293e-06 9.0371e-06 2.7722e-06 1.3516e-06 3.016e-06 3.9832e-06 6.276e-07 0.4531438 1.2
```

```
TABLES_PP <- synth.tab(dataprep.res = DATA_PP, synth.res = SCM_PP)

SCM_LSM    <- synth(DATA_LSM)
```

```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
## ****************
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
##
## MSPE (LOSS V): 12.12755
##
## solution.v:
##  0.1141571 0.3845373 0.2922767 0.2090279 1.0478e-06
##
## solution.w:
```

```
##  4.1573e-06 3.3566e-06 3.7345e-06 2.8728e-06 3.1331e-06 2.4117e-06 4.0757e-06 3.0407e-06 0.4657458 5
```

```
TABLES_LSM <- synth.tab(dataprep.res = DATA_LSM, synth.res = SCM_LSM)

SCM_LSP    <- synth(DATA_LSP)
```

```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
## ****************
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
##
## MSPE (LOSS V): 6.792952
##
## solution.v:
##  0.01212731 0.5569337 0.09954846 0.3197918 0.01159867
##
## solution.w:
##  0.0002073634 0.0001753144 0.0002039686 0.0001401451 0.000157976 0.0001195055 0.0002135721 0.0002698
```

```
TABLES_LSP <- synth.tab(dataprep.res = DATA_LSP, synth.res = SCM_LSP)


# Graphs in ggplot
PM <- plot_scm(PRIMARY_M, TABLES_PM)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v stringr   1.5.1
## v lubridate 1.9.3     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.0
## v readr     2.1.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
## 'data.frame':    3 obs. of  3 variables:
##  $ w.weights   : num  0.479 0.164 0.356
##  $ abbr_state  : chr  "PI" "PE" "BA"
##  $ unit.numbers: num  22 26 29
## 'data.frame':    351 obs. of  14 variables:
##  $ abbr_state   : chr  "AC" "AC" "AC" "AC" ...
##  $ code_state   : num  12 12 12 12 12 27 27 27 27 27 ...
##  $ year         : num  1995 1997 1999 2001 2003 ...
##  $ grade        : chr  "P" "P" "P" "P" ...
##  $ subject      : chr  "math" "math" "math" "math" ...
##  $ score        : num  168 166 162 152 159 ...
##  $ TWh          : num  0.0139 0.0132 0.0137 0.0198 0.0227 ...
##  $ homicides    : num  22.62 19.99 9.66 21.24 22.04 ...
##  $ pop          : num  455242 500185 527937 574355 600595 ...
```

```
##  $ ln_pop       : num  13 13.1 13.2 13.3 13.3 ...
##  $ unemployment : num  9.8 8.89 12.2 8.56 7.45 ...
##  $ gini         : num  0.582 0.574 0.621 0.623 0.578 ...
##  $ edu_invest   : num  3.69e+08 4.34e+08 4.54e+08 6.19e+08 6.26e+08 ...
##  $ edu_invest_pc: num  811 868 860 1077 1042 ...
```

```
PM_SC <- PM[[1]]
PM_GAP <- PM[[2]]

PP <- plot_scm(PRIMARY_P, TABLES_PP)
```

```
## 'data.frame':    3 obs. of  3 variables:
##  $ w.weights   : num  0.453 0.17 0.375
##  $ abbr_state  : chr  "PI" "PE" "BA"
##  $ unit.numbers: num  22 26 29
## 'data.frame':    351 obs. of  14 variables:
##  $ abbr_state   : chr  "AC" "AC" "AC" "AC" ...
##  $ code_state   : num  12 12 12 12 12 27 27 27 27 27 ...
##  $ year         : num  1995 1997 1999 2001 2003 ...
##  $ grade        : chr  "P" "P" "P" "P" ...
##  $ subject      : chr  "port" "port" "port" "port" ...
##  $ score        : num  166 160 154 148 158 ...
##  $ TWh          : num  0.0139 0.0132 0.0137 0.0198 0.0227 ...
##  $ homicides    : num  22.62 19.99 9.66 21.24 22.04 ...
##  $ pop          : num  455242 500185 527937 574355 600595 ...
##  $ ln_pop       : num  13 13.1 13.2 13.3 13.3 ...
##  $ unemployment : num  9.8 8.89 12.2 8.56 7.45 ...
##  $ gini         : num  0.582 0.574 0.621 0.623 0.578 ...
##  $ edu_invest   : num  3.69e+08 4.34e+08 4.54e+08 6.19e+08 6.26e+08 ...
##  $ edu_invest_pc: num  811 868 860 1077 1042 ...
```

```
PP_SC <- PP[[1]]
PP_GAP <- PP[[2]]

LSM <- plot_scm(LOWERS_M, TABLES_LSM)
```

```
## 'data.frame':    3 obs. of  3 variables:
##  $ w.weights   : num  0.466 0.174 0.36
##  $ abbr_state  : chr  "PI" "PE" "BA"
##  $ unit.numbers: num  22 26 29
## 'data.frame':    351 obs. of  14 variables:
##  $ abbr_state   : chr  "AC" "AC" "AC" "AC" ...
##  $ code_state   : num  12 12 12 12 12 27 27 27 27 27 ...
##  $ year         : num  1995 1997 1999 2001 2003 ...
##  $ grade        : chr  "LS" "LS" "LS" "LS" ...
##  $ subject      : chr  "math" "math" "math" "math" ...
##  $ score        : num  223 220 223 219 225 ...
##  $ TWh          : num  0.0139 0.0132 0.0137 0.0198 0.0227 ...
##  $ homicides    : num  22.62 19.99 9.66 21.24 22.04 ...
##  $ pop          : num  455242 500185 527937 574355 600595 ...
##  $ ln_pop       : num  13 13.1 13.2 13.3 13.3 ...
##  $ unemployment : num  9.8 8.89 12.2 8.56 7.45 ...
##  $ gini         : num  0.582 0.574 0.621 0.623 0.578 ...
##  $ edu_invest   : num  3.69e+08 4.34e+08 4.54e+08 6.19e+08 6.26e+08 ...
##  $ edu_invest_pc: num  811 868 860 1077 1042 ...
```

```
LSM_SC <- LSM[[1]]
LSM_GAP <- LSM[[2]]

LSP <- plot_scm(LOWERS_P, TABLES_LSP)
```

```
## 'data.frame':    4 obs. of  3 variables:
##  $ w.weights   : num  0.491 0.262 0.206 0.034
##  $ abbr_state  : chr  "PI" "PE" "BA" "GO"
##  $ unit.numbers: num  22 26 29 52
## 'data.frame':    351 obs. of  14 variables:
##  $ abbr_state   : chr  "AC" "AC" "AC" "AC" ...
##  $ code_state   : num  12 12 12 12 12 27 27 27 27 27 ...
##  $ year         : num  1995 1997 1999 2001 2003 ...
##  $ grade        : chr  "LS" "LS" "LS" "LS" ...
##  $ subject      : chr  "port" "port" "port" "port" ...
##  $ score        : num  227 223 217 218 223 ...
##  $ TWh          : num  0.0139 0.0132 0.0137 0.0198 0.0227 ...
##  $ homicides    : num  22.62 19.99 9.66 21.24 22.04 ...
##  $ pop          : num  455242 500185 527937 574355 600595 ...
##  $ ln_pop       : num  13 13.1 13.2 13.3 13.3 ...
##  $ unemployment : num  9.8 8.89 12.2 8.56 7.45 ...
##  $ gini         : num  0.582 0.574 0.621 0.623 0.578 ...
##  $ edu_invest   : num  3.69e+08 4.34e+08 4.54e+08 6.19e+08 6.26e+08 ...
##  $ edu_invest_pc: num  811 868 860 1077 1042 ...
```

```
LSP_SC <- LSP[[1]]
LSP_GAP <- LSP[[2]]

DATA_GRAPH <- rbind(PM_SC, PP_SC, LSM_SC, LSP_SC)

DATA_GRAPH$grade[DATA_GRAPH$grade=="P"] <- "Primary Education"
DATA_GRAPH$grade[DATA_GRAPH$grade=="LS"] <- "Lower Secondary Education"
DATA_GRAPH$grade <- factor(DATA_GRAPH$grade, levels = c("Primary Education", "Lower Secondary Education"

DATA_GRAPH$subject[DATA_GRAPH$subject=="math"] <- "Mathematics"
DATA_GRAPH$subject[DATA_GRAPH$subject=="port"] <- "Portuguese"

# Figure 06a (SCM - Primary Education)

a_06 <- ggplot(data = filter(DATA_GRAPH, grade == "Primary Education"), aes(x=year, y= score, color = u
  geom_vline(xintercept = 2008, color = "#636363", linetype = "dashed", size = 0.9)+
  geom_vline(xintercept = 2011, color = "#636363", linetype = "dashed", size = 0.9)+
  geom_line(size=0.9)+
  scale_color_manual(values= c("#01665e","#d8b365"),
                     labels= c( "Ceará", "Synthetic Ceará"),
                     name = "")+
  ylab("Score")+
  xlab("")+
  annotate("text", x = 2007, y = 220, label = "TI", color = "#636363", size = 4)+
  annotate("text", x = 2013, y = 152, label = "TI + TA", color = "#636363", size = 4)+
  theme_bw()+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        strip.text = element_text(colour = "#636363"),
```

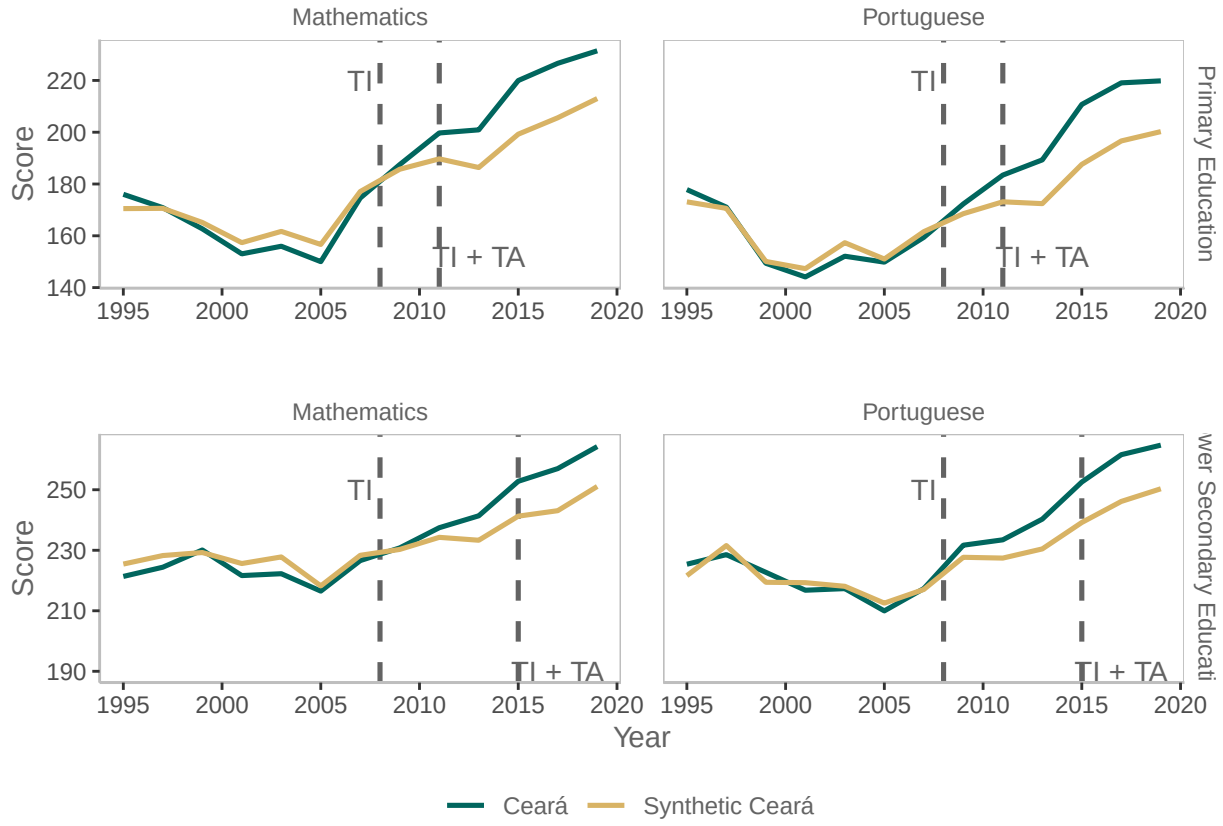
```
        axis.line = element_line(colour = "gray"),
        panel.border = element_rect(colour = "gray"),
        legend.position = "bottom",
        panel.spacing = unit(1.1, "lines"),
        strip.background = element_rect(fill="white", linetype = "blank"),
        text = element_text(family="Helvetica", color ="#636363"))+
  facet_grid(vars(grade),vars(subject))
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
# Figure 06b (SCM - Lower Secondary Education)

b_06 <- ggplot(data = filter(DATA_GRAPH, grade == "Lower Secondary Education"), aes(x=year, y= score, c
  geom_vline(xintercept = 2008, color = "#636363", linetype = "dashed", size = 0.9)+
  geom_vline(xintercept = 2015, color = "#636363", linetype = "dashed", size = 0.9)+
  geom_line(size=0.9)+
  scale_color_manual(values= c("#01665e","#d8b365"),
                     labels= c( "Ceará", "Synthetic Ceará"),
                     name = "")+
  ylab("Score")+
  xlab("Year")+
  annotate("text", x = 2007, y = 250, label = "TI", color = "#636363", size = 4)+
  annotate("text", x = 2017, y = 190, label = "TI + TA", color = "#636363", size = 4)+
  theme_bw()+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        strip.text = element_text(colour = "#636363"),
        axis.line = element_line(colour = "gray"),
        panel.border = element_rect(colour = "gray"),
        legend.position = "bottom",
        panel.spacing = unit(1.1, "lines"),
        strip.background = element_rect(fill="white", linetype = "blank"),
        text = element_text(family="Helvetica", color ="#636363"))+
  facet_grid(vars(grade),vars(subject))


ggarrange(a_06, b_06, ncol = 1, nrow = 2, common.legend = TRUE, legend = "bottom")
```




```
ggsave(filename = "figure06.png", path = "plots",   width = 21, height = 15, , units = "cm")
```