Extension: In-Place Placebo for Pernambuco

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

```r
# Load Data
load("data/DATA_COMPLETE.RData")
load("data/abbr_code.RData")

# Subsetting data by grade and subject (for all states, including PE):
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"))

#-------------------------------------------------------------------------------------------#
# Function: prepare_p_ls_PE()
# Description: prepares data for the synth function - primary and lower secondary school
#-------------------------------------------------------------------------------------------#

prepare_p_ls_PE <- function(data){
  library(Synth)

  predictors <- c("homicides", "TWh", "ln_pop", "unemployment", "edu_invest_pc")

  DATA_PM <- dataprep(foo = data,
                      predictors = predictors,
                      dependent     = "score",
                      unit.variable = "code_state",
                      time.variable = "year",
                      unit.names.variable = "abbr_state",
                      treatment.identifier  = 26, # Pernambuco Code
                      controls.identifier   = c(11:17, 21, 22, 24:25, 27:29, 31:33, 35, 41:43, 50:53),
                      # Exclude PE (26) and CE (23) from controls
                      time.predictors.prior = seq(1995, 2007, 2),
                      time.optimize.ssr     = seq(1995, 2007, 2),
                      time.plot             = seq(1995, 2019, 2))

  return(DATA_PM)
};

#-------------------------------------------------------------------------------------------#
# Function: prepare_us_PE()
# Description: prepares data for the synth function - upper secondary school
#-------------------------------------------------------------------------------------------#

prepare_us_PE <- function(data){
  library(Synth)

  predictors <- c("homicides", "TWh", "unemployment", "ln_pop", "edu_invest_pc")
```

```
  DATA_PM <- dataprep(foo = data,
                      predictors = predictors,
                      dependent     = "score",
                      unit.variable = "code_state",
                      time.variable = "year",
                      unit.names.variable = "abbr_state",
                      treatment.identifier  = 26, # Pernambuco Code
                      controls.identifier   = c(11:17, 21, 22, 24:25, 27:29, 31:33, 35, 41:43, 50:53),
                      # Exclude PE (26) and CE (23) from controls
                      time.predictors.prior = seq(1995, 2009, 2),
                      time.optimize.ssr     = seq(1995, 2009, 2),
                      time.plot             = seq(1995, 2019, 2))

  return(DATA_PM)
};

#-----------------------------------------------------------------------------------------------#
# Preparing data for Synth for PE
#-----------------------------------------------------------------------------------------------#

# Prepare data for SCM (specifying PE as the treatment unit):
DATA_PE_PM <- prepare_p_ls_PE(PRIMARY_M)
DATA_PE_PP <- prepare_p_ls_PE(PRIMARY_P)
DATA_PE_LSM <- prepare_p_ls_PE(LOWERS_M)
DATA_PE_LSP <- prepare_p_ls_PE(LOWERS_P)
# DATA_PE_USM <- prepare_us_PE(UPPERS_M)
# DATA_PE_USP <- prepare_us_PE(UPPERS_P)

# Primary School Mathematics and Portuguese
DATA_PE_PM <- prepare_p_ls_PE(PRIMARY_M)
DATA_PE_PP <- prepare_p_ls_PE(PRIMARY_P)

# Lower Secondary School Mathematics and Portuguese
DATA_PE_LSM <- prepare_p_ls_PE(LOWERS_M)
DATA_PE_LSP <- prepare_p_ls_PE(LOWERS_P)

# Upper Secondary School Mathematics and Portuguese
# DATA_PE_USM <- prepare_us_PE(UPPERS_M)
# DATA_PE_USP <- prepare_us_PE(UPPERS_P)

#-----------------------------------------------------------------------------------------------#
# Function: plot_scm_PE()
# Description: prepares data from the synthetic control output to be plotted with ggplot
#-----------------------------------------------------------------------------------------------#

plot_scm_PE <- function(original_data, synth.tables){
  library(tidyverse)
  W <- as.data.frame(synth.tables[["tab.w"]])
  str(W)
  W <- W %>%
    filter(w.weights > 0.01) %>%
    mutate(w.weights = round(w.weights, digits = 3)) %>%
    rename(abbr_state = unit.names)
```

```
  str(original_data)
  str(W)
  SC <- left_join(original_data, select(W, -unit.numbers), by = "abbr_state") %>%
    na.omit() %>%
    group_by(year) %>%
    summarise(sc = weighted.mean(score, w.weights))

  PE <- original_data %>%
    filter(abbr_state == "PE") %>%
    select(year, score)

  GAP <- left_join(PE, SC, by = "year") %>%
    mutate(gap = score - sc)
  GAP$grade <- unique(original_data$grade)
  GAP$subject <- unique(original_data$subject)

  GG_DATA <- left_join(PE, SC, by = "year") %>%
    pivot_longer(!year, names_to = "unit", values_to = "score")

  GG_DATA$unit[GG_DATA$unit == "score"] <- "Pernambuco"
  GG_DATA$unit[GG_DATA$unit == "sc"] <- "Synthetic Control"
  GG_DATA$grade <- unique(original_data$grade)
  GG_DATA$subject <- unique(original_data$subject)

  return(list(GG_DATA, GAP))
}

#-----------------------------------------------------------------------------------------#
# Data Preparation for Plotting
#-----------------------------------------------------------------------------------------#

# Run SCM for Primary School Mathematics (PE)
SCM_PE_PM <- synth(DATA_PE_PM)
```

```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
## ****************
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
##
## MSPE (LOSS V): 21.63649
##
## solution.v:
##  0.008754374 0.3649107 0.08927948 0.01204785 0.5250076
##
## solution.w:
##  2.5e-08 1.45e-08 3.46e-08 1.14e-08 0 1.3e-08 1.75e-08 0 5.1e-09 2.14e-08 1.46e-08 0.7530414 1.95e-08
```

```r
TABLES_PE_PM <- synth.tab(dataprep.res = DATA_PE_PM, synth.res = SCM_PE_PM)

# Run SCM for Primary School Portuguese (PE)
SCM_PE_PP <- synth(DATA_PE_PP)
```

```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
## ****************
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
##
## MSPE (LOSS V): 12.77149
##
## solution.v:
##  0.00265162 0.302572 0.08881434 0.08620696 0.5197551
##
## solution.w:
##  2.75e-08 2.27e-08 3.7971e-06 1.91e-08 0 3.37e-08 2.16e-08 1e-10 5.2e-09 5.47e-08 1.34e-08 0.6920162
```

```r
TABLES_PE_PP <- synth.tab(dataprep.res = DATA_PE_PP, synth.res = SCM_PE_PP)

# Run SCM for Lower Secondary School Mathematics (PE)
SCM_PE_LSM <- synth(DATA_PE_LSM)
```

```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
## ****************
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
##
## MSPE (LOSS V): 9.068007
##
## solution.v:
##  2.04918e-05 0.5305243 0.01069398 0.129317 0.3294442
##
## solution.w:
##  3.11e-08 1.96e-08 0.0922407 1.61e-08 0.06827577 1.96e-08 2.57e-08 8.42e-08 3.18e-08 5.68e-08 4.95e-
```

```r
TABLES_PE_LSM <- synth.tab(dataprep.res = DATA_PE_LSM, synth.res = SCM_PE_LSM)

# Run SCM for Lower Secondary School Portuguese (PE)
SCM_PE_LSP <- synth(DATA_PE_LSP)
```

```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
## ****************
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
##
## MSPE (LOSS V): 4.909805
##
## solution.v:
##  0.00366472 0.6049595 1.22621e-05 0.1490142 0.2423493
##
## solution.w:
##  2.11e-08 2.09e-08 0.1161376 2.14e-08 0.02076024 3.19e-08 1.84e-08 1.486e-07 9e-09 2.12e-08 1.39e-08
```

```
TABLES_PE_LSP <- synth.tab(dataprep.res = DATA_PE_LSP, synth.res = SCM_PE_LSP)

# Run SCM for Upper Secondary School Mathematics (PE)
# SCM_PE_USM <- synth(DATA_PE_USM)
# TABLES_PE_USM <- synth.tab(dataprep.res = DATA_PE_USM, synth.res = SCM_PE_USM)

# Run SCM for Upper Secondary School Portuguese (PE)
# SCM_PE_USP <- synth(DATA_PE_USP)
# TABLES_PE_USP <- synth.tab(dataprep.res = DATA_PE_USP, synth.res = SCM_PE_USP)

#-------------------------------------------------------------------------------------------#

# Graphs in ggplot for Pernambuco
PM_PE <- plot_scm_PE(PRIMARY_M, TABLES_PE_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':    25 obs. of  3 variables:
##  $ w.weights   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ unit.names  : chr  "RO" "AC" "AM" "RR" ...
##  $ unit.numbers: num  11 12 13 14 15 16 17 21 22 24 ...
## '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 ...
## 'data.frame':    2 obs. of  3 variables:
##  $ w.weights   : num  0.753 0.247
##  $ abbr_state  : chr  "AL" "BA"
##  $ unit.numbers: num  27 29
```

```
PM_PE_SC <- PM_PE[[1]]
PM_PE_GAP <- PM_PE[[2]]

PP_PE <- plot_scm_PE(PRIMARY_P, TABLES_PE_PP)
```

```
## 'data.frame':    25 obs. of  3 variables:
##  $ w.weights   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ unit.names  : chr  "RO" "AC" "AM" "RR" ...
##  $ unit.numbers: num  11 12 13 14 15 16 17 21 22 24 ...
## '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 ...
## 'data.frame':    2 obs. of  3 variables:
##  $ w.weights   : num  0.692 0.308
##  $ abbr_state  : chr  "AL" "BA"
##  $ unit.numbers: num  27 29
```

```
PP_PE_SC <- PP_PE[[1]]
PP_PE_GAP <- PP_PE[[2]]

LSM_PE <- plot_scm_PE(LOWERS_M, TABLES_PE_LSM)
```

```
## 'data.frame':    25 obs. of  3 variables:
##  $ w.weights   : num  0 0 0.092 0 0.068 0 0 0 0 0 ...
##  $ unit.names  : chr  "RO" "AC" "AM" "RR" ...
##  $ unit.numbers: num  11 12 13 14 15 16 17 21 22 24 ...
## '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 ...
## 'data.frame':    3 obs. of  3 variables:
##  $ w.weights   : num  0.092 0.068 0.839
##  $ abbr_state  : chr  "AM" "PA" "AL"
##  $ unit.numbers: num  13 15 27
```

```
LSM_PE_SC <- LSM_PE[[1]]
LSM_PE_GAP <- LSM_PE[[2]]

LSP_PE <- plot_scm_PE(LOWERS_P, TABLES_PE_LSP)
```

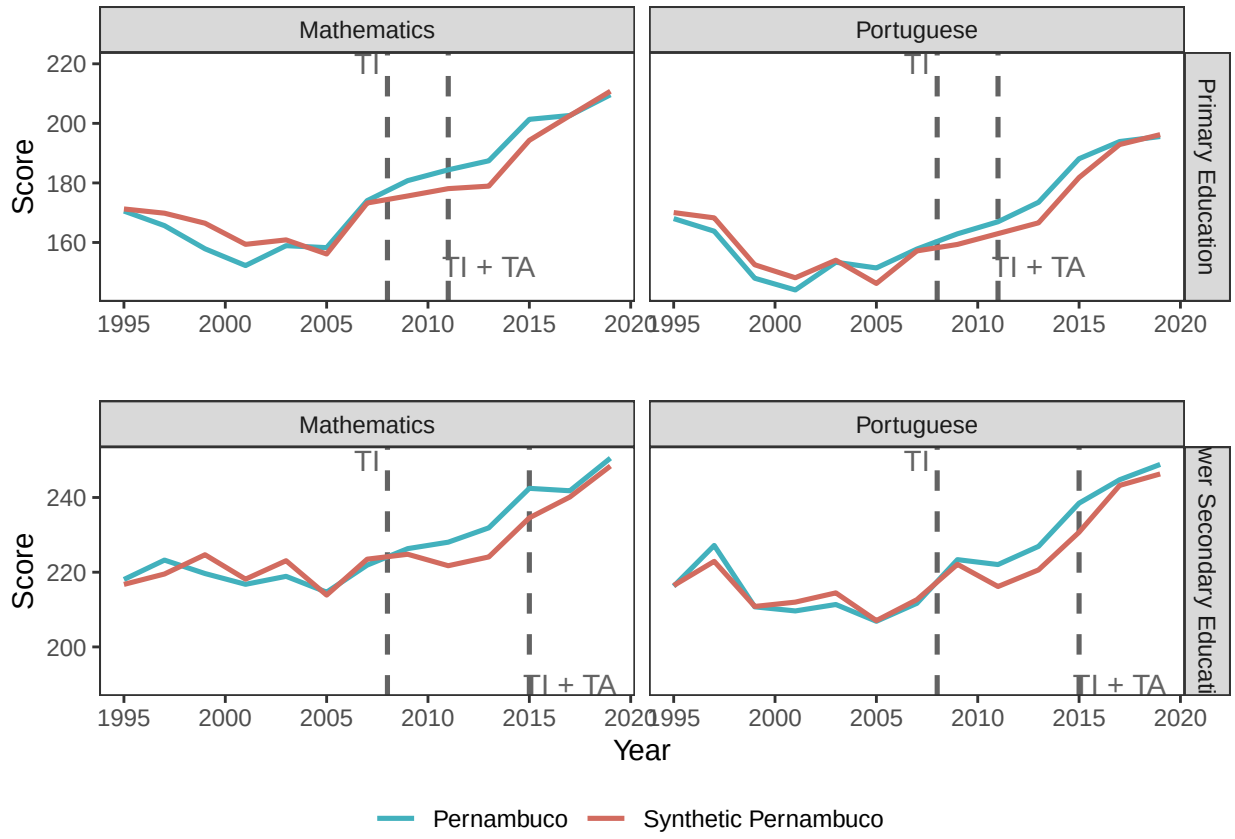
```
## 'data.frame':    25 obs. of  3 variables:
##  $ w.weights   : num  0 0 0.116 0 0.021 0 0 0 0 0 ...
##  $ unit.names  : chr  "RO" "AC" "AM" "RR" ...
##  $ unit.numbers: num  11 12 13 14 15 16 17 21 22 24 ...
## 'data.frame':    351 obs. of  14 variables:
##  $ abbr_state   : chr  "AC" "AC" "AC" "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 ...
## 'data.frame':    3 obs. of  3 variables:
##  $ w.weights   : num  0.116 0.021 0.863
##  $ abbr_state  : chr  "AM" "PA" "AL"
##  $ unit.numbers: num  13 15 27
```

```
LSP_PE_SC <- LSP_PE[[1]]
LSP_PE_GAP <- LSP_PE[[2]]

# Combining Graph Data for Pernambuco
DATA_GRAPH_PE <- rbind(PM_PE_SC, PP_PE_SC, LSM_PE_SC, LSP_PE_SC)

# Adjusting Labels for Pernambuco data
DATA_GRAPH_PE$grade[DATA_GRAPH_PE$grade=="P"] <- "Primary Education"
DATA_GRAPH_PE$grade[DATA_GRAPH_PE$grade=="LS"] <- "Lower Secondary Education"
```

```r
DATA_GRAPH_PE$grade <- factor(DATA_GRAPH_PE$grade, levels = c("Primary Education", "Lower Secondary Edu

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

#-------------------------------------------------------------------------------------------#
# Plotting
#-------------------------------------------------------------------------------------------#

# Figure for Primary Education in PE
a_06_PE <- ggplot(data = filter(DATA_GRAPH_PE, grade == "Primary Education"), aes(x = year, y = score,
  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("#42B1BD", "#D26B5F"), labels = c("Pernambuco", "Synthetic Pernambuco")
  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(), legend.position = "botto
  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.
```

```r
# Figure for Lower Secondary Education in PE
b_06_PE <- ggplot(data = filter(DATA_GRAPH_PE, grade == "Lower Secondary Education"), aes(x = year, y =
  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("#42B1BD", "#D26B5F"), labels = c("Pernambuco", "Synthetic Pernambuco")
  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(), legend.position = "botto
  facet_grid(vars(grade), vars(subject))

# Arrange and Save the Plots
ggarrange(a_06_PE, b_06_PE, ncol = 1, nrow = 2, common.legend = TRUE, legend = "bottom")
```




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