Extension: In-Place Placebo for Piauí

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

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


# Subsetting data by grade and subject (for all states, including PI):
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_PI()
# Description: prepares data for the synth function - primary and lower secondary school
#-------------------------------------------------------------------------------------------------#

prepare_p_ls_PI <- 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  = 22, # Piauí Code
                      controls.identifier   = c(11:17, 21, 24:29, 31:33, 35, 41:43, 50:53),
                      # Exclude PI (22) 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_PI()
# Description: prepares data for the synth function - upper secondary school
#-------------------------------------------------------------------------------------------------#

prepare_us_PI <- 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  = 22, # Piauí Code
                      controls.identifier   = c(11:17, 21, 24:26, 28:29, 31:33, 35, 41:43, 50:53),
                      # Exclude PI (22) 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 PI
#-----------------------------------------------------------------------------------------------#

# Prepare data for SCM (specifying PI as the treatment unit):
DATA_PI_PM <- prepare_p_ls_PI(PRIMARY_M)
DATA_PI_PP <- prepare_p_ls_PI(PRIMARY_P)
DATA_PI_LSM <- prepare_p_ls_PI(LOWERS_M)
DATA_PI_LSP <- prepare_p_ls_PI(LOWERS_P)
DATA_PI_USM <- prepare_us_PI(UPPERS_M)
DATA_PI_USP <- prepare_us_PI(UPPERS_P)

# Primary School Mathematics and Portuguese
DATA_PI_PM <- prepare_p_ls_PI(PRIMARY_M)
DATA_PI_PP <- prepare_p_ls_PI(PRIMARY_P)

# Lower Secondary School Mathematics and Portuguese
DATA_PI_LSM <- prepare_p_ls_PI(LOWERS_M)
DATA_PI_LSP <- prepare_p_ls_PI(LOWERS_P)

# Upper Secondary School Mathematics and Portuguese
DATA_PI_USM <- prepare_us_PI(UPPERS_M)
DATA_PI_USP <- prepare_us_PI(UPPERS_P)

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

plot_scm_PI <- 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))

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

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

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

  GG_DATA$unit[GG_DATA$unit == "score"] <- "Piauí"
  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 (PI)
SCM_PI_PM <- synth(DATA_PI_PM)
```

```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
## ****************
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
##
## MSPE (LOSS V): 3.721776
##
## solution.v:
##  0.2217879 1.9274e-06 0.05259182 1.4035e-06 0.7256169
##
## solution.w:
```

```
##  2.826e-07 4.5732e-06 1.2587e-06 8.076e-07 8.55e-08 6.889e-07 0.2006367 0.7843464 8.30438e-05 2.789e
```

```
TABLES_PI_PM <- synth.tab(dataprep.res = DATA_PI_PM, synth.res = SCM_PI_PM)

# Run SCM for Primary School Portuguese (PI)
SCM_PI_PP <- synth(DATA_PI_PP)
```

```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
## ****************
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
##
## MSPE (LOSS V): 2.328019
##
## solution.v:
##  0.2679382 0.2692525 0.3424016 0.01790257 0.1025051
##
## solution.w:
##  3.906e-07 5.719e-07 3.559e-07 3.107e-07 2.231e-07 2.397e-07 0.3178909 0.3319879 0.22772 4.4972e-06
```

```
TABLES_PI_PP <- synth.tab(dataprep.res = DATA_PI_PP, synth.res = SCM_PI_PP)

# Run SCM for Lower Secondary School Mathematics (PI)
SCM_PI_LSM <- synth(DATA_PI_LSM)
```

```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
## ****************
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
##
## MSPE (LOSS V): 3.474292
##
## solution.v:
##  0.1356106 0.2475876 0.5476153 0.00824473 0.06094185
##
## solution.w:
##  2.2678e-06 2.1181e-06 1.2029e-06 1.5922e-06 1.0822e-06 1.2055e-06 0.2051833 0.1434177 0.4222922 0.1
```

```
TABLES_PI_LSM <- synth.tab(dataprep.res = DATA_PI_LSM, synth.res = SCM_PI_LSM)

# Run SCM for Lower Secondary School Portuguese (PI)
SCM_PI_LSP <- synth(DATA_PI_LSP)
```

```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
## ****************
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
##
## MSPE (LOSS V): 13.90849
##
## solution.v:
##  0.09991008 0.8858695 0.005624223 0.002671255 0.005924897
##
## solution.w:
##  3.276e-07 2.258e-07 2.016e-07 1.964e-07 2.108e-06 1.409e-07 0.38369 4.78991e-05 0.6160384 4.045e-06
```

```
TABLES_PI_LSP <- synth.tab(dataprep.res = DATA_PI_LSP, synth.res = SCM_PI_LSP)

# Run SCM for Upper Secondary School Mathematics (PI)
SCM_PI_USM <- synth(DATA_PI_USM)
```

```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
## ****************
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
##
## MSPE (LOSS V): 14.74899
##
## solution.v:
##  0.0876074 0.1164132 0.01563386 0.6786735 0.1016721
##
## solution.w:
##  1.9562e-06 5.323e-07 1.2371e-06 5.026e-07 3.827e-07 4.595e-07 0.4235325 0.5760261 2.4744e-05 0.0003
```

```
TABLES_PI_USM <- synth.tab(dataprep.res = DATA_PI_USM, synth.res = SCM_PI_USM)

# Run SCM for Upper Secondary School Portuguese (PI)
SCM_PI_USP <- synth(DATA_PI_USP)
```

```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
## ****************
```

```
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
##
## MSPE (LOSS V): 20.3554
##
## solution.v:
##  0.0492984 0.1164686 0.04759728 0.5881315 0.1985042
##
## solution.w:
##  6.427e-07 1.385e-07 8.41e-08 1.073e-07 6.55e-08 6.06e-08 0.3966485 0.6033413 1.387e-07 1.6328e-06 1
```

```
TABLES_PI_USP <- synth.tab(dataprep.res = DATA_PI_USP, synth.res = SCM_PI_USP)

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

# Graphs in ggplot for Piauí
PM_PI <- plot_scm_PI(PRIMARY_M, TABLES_PI_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 error
```

```
## 'data.frame':    25 obs. of  3 variables:
##  $ w.weights   : num  0 0 0 0 0 0 0.201 0.784 0 0 ...
##  $ unit.names  : chr  "RO" "AC" "AM" "RR" ...
##  $ unit.numbers: num  11 12 13 14 15 16 17 21 24 25 ...
## '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':    3 obs. of  3 variables:
##  $ w.weights   : num  0.201 0.784 0.015
##  $ abbr_state  : chr  "TO" "MA" "SC"
##  $ unit.numbers: num  17 21 42
```

```r
PM_PI_SC <- PM_PI[[1]]
PM_PI_GAP <- PM_PI[[2]]

PP_PI <- plot_scm_PI(PRIMARY_P, TABLES_PI_PP)
```

```
## 'data.frame':    25 obs. of  3 variables:
##  $ w.weights   : num  0 0 0 0 0 0 0.318 0.332 0.228 0 ...
##  $ unit.names  : chr  "RO" "AC" "AM" "RR" ...
##  $ unit.numbers: num  11 12 13 14 15 16 17 21 24 25 ...
## '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':    4 obs. of  3 variables:
##  $ w.weights   : num  0.318 0.332 0.228 0.122
##  $ abbr_state  : chr  "TO" "MA" "RN" "SC"
##  $ unit.numbers: num  17 21 24 42
```

```r
PP_PI_SC <- PP_PI[[1]]
PP_PI_GAP <- PP_PI[[2]]

LSM_PI <- plot_scm_PI(LOWERS_M, TABLES_PI_LSM)
```

```
## 'data.frame':    25 obs. of  3 variables:
##  $ w.weights   : num  0 0 0 0 0 0 0.205 0.143 0.422 0.105 ...
##  $ unit.names  : chr  "RO" "AC" "AM" "RR" ...
##  $ unit.numbers: num  11 12 13 14 15 16 17 21 24 25 ...
## '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':    5 obs. of  3 variables:
```

```
##  $ w.weights   : num  0.205 0.143 0.422 0.105 0.124
##  $ abbr_state  : chr  "TO" "MA" "RN" "PB" ...
##  $ unit.numbers: num  17 21 24 25 42
```

```
LSM_PI_SC <- LSM_PI[[1]]
LSM_PI_GAP <- LSM_PI[[2]]

LSP_PI <- plot_scm_PI(LOWERS_P, TABLES_PI_LSP)
```

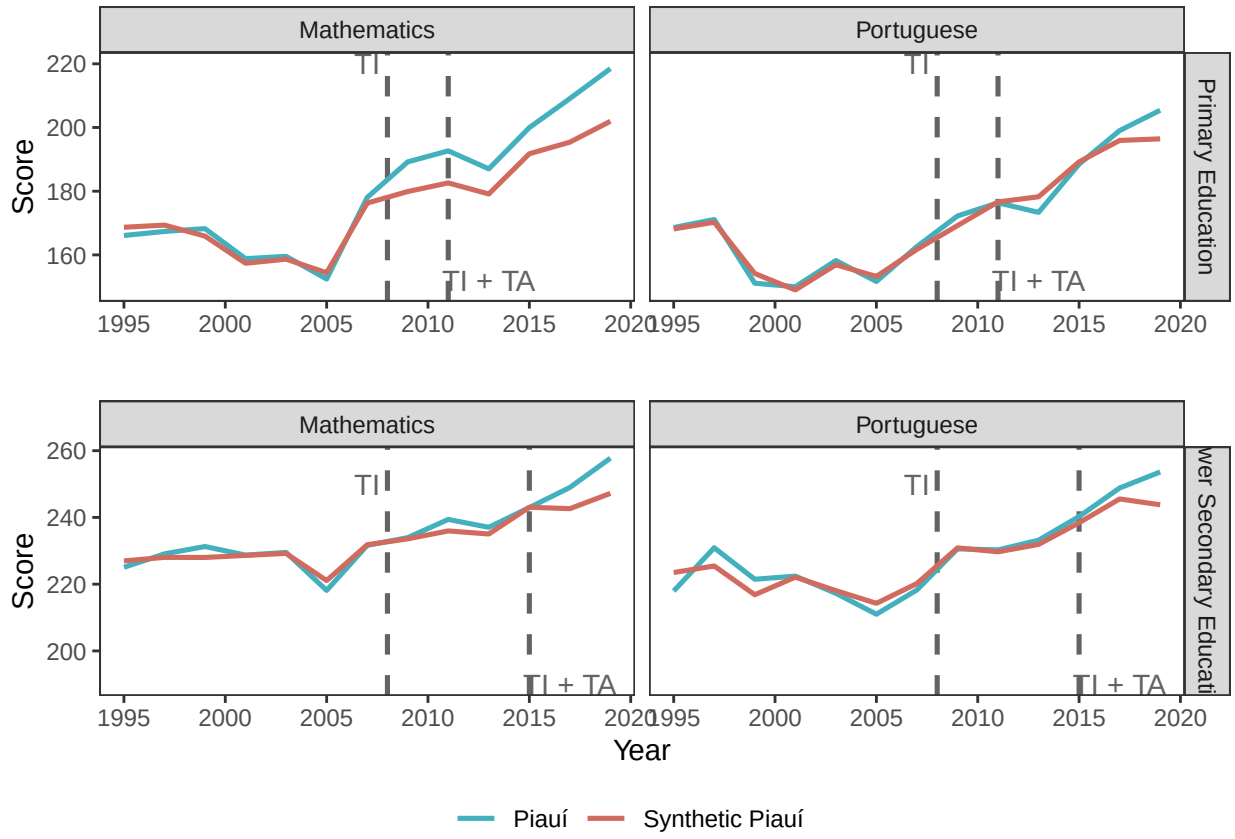
```
## 'data.frame':    25 obs. of  3 variables:
##  $ w.weights   : num  0 0 0 0 0 0 0.384 0 0.616 0 ...
##  $ unit.names  : chr  "RO" "AC" "AM" "RR" ...
##  $ unit.numbers: num  11 12 13 14 15 16 17 21 24 25 ...
## 'data.frame':    351 obs. of  14 variables:
##  $ abbr_state   : chr  "AC" "AC" "AC" "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':    2 obs. of  3 variables:
##  $ w.weights   : num  0.384 0.616
##  $ abbr_state  : chr  "TO" "RN"
##  $ unit.numbers: num  17 24
```

```
LSP_PI_SC <- LSP_PI[[1]]
LSP_PI_GAP <- LSP_PI[[2]]

# Combining Graph Data for Piauí
DATA_GRAPH_PI <- rbind(PM_PI_SC, PP_PI_SC, LSM_PI_SC, LSP_PI_SC)

# Adjusting Labels for Piauí data
DATA_GRAPH_PI$grade[DATA_GRAPH_PI$grade=="P"] <- "Primary Education"
DATA_GRAPH_PI$grade[DATA_GRAPH_PI$grade=="LS"] <- "Lower Secondary Education"
DATA_GRAPH_PI$grade <- factor(DATA_GRAPH_PI$grade, levels = c("Primary Education", "Lower Secondary Edu

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

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

# Figure for Primary Education in PI
a_06_PI <- ggplot(data = filter(DATA_GRAPH_PI, 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("Piauí", "Synthetic Piauí"), 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(), 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.
```

```
# Figure for Lower Secondary Education in PI
b_06_PI <- ggplot(data = filter(DATA_GRAPH_PI, 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("Piauí", "Synthetic Piauí"), 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(), legend.position = "botto
  facet_grid(vars(grade), vars(subject))

# Arrange and Save the Plots
ggarrange(a_06_PI, b_06_PI, ncol = 1, nrow = 2, common.legend = TRUE, legend = "bottom")
```




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