Assignment_Three Q2

Author - Eric Opoku

December 2, 2022

```
name<- Sys.info()
name[7]
   user
"erico"
```

```
library(rlang)
library(dplyr)
library(kableExtra)
library(ggdag)      # For plotting DAGs
library(dagitty)    # For working with DAG logic
library(modelsummary) # For making regression tables
library(AER) # this package has lots of applied metrics packages
library(foreign)
# Helpful for reading in data from Stata or other code languages
library(lubridate) # For figures
library(stargazer) # For tables
library(data.table)  #data manipulation and wrangling
library(lme4)
library(psych)
library(readxl) # Read in data
library(expss) #value labelling from spss style
library(readstata13)
library(marginaleffects) # To calculate marginal effects
library(knitr) # Alternative table package
library(ggplot2)





######
```

```r
### 0. Load the packages we will need for this file ####
library(tidyverse) # load the installed package for each new session of R
library(broom)
library(causaldata) # Useful toy data sets
library(here) # Helpful in working with directories and projects
library(zoo) # Helpful packages for organizing dates
library(tidysynth) # For synthetic controls
library(gsynth) # For synthetic controls
library(gghighlight) # For figures
library(binsreg) # For binscatters
library(nprobust) # Local linear regression

library(fixest)

set.seed(032620) # random number generators; same numbers across machines
```

```r
mydata<- read_excel("C:/Users/erico/Desktop/ACADEMICS/MY UofT COURSES/YEAR TWO/SEMESTER

#View(mydata)


#regdata$the_date<-mydata$date  # I will use this data later




#table(mydata$date)
#table(mydata$hits)
table(mydata$keyword)

   cereal  sandwich      soup sourdough
      214       214       214       214
table(mydata$time)

2020-01-01 2020-08-01
                  856
table(mydata$gprop)

web
856
table(mydata$category)

  0
856
table(mydata$geo)
```

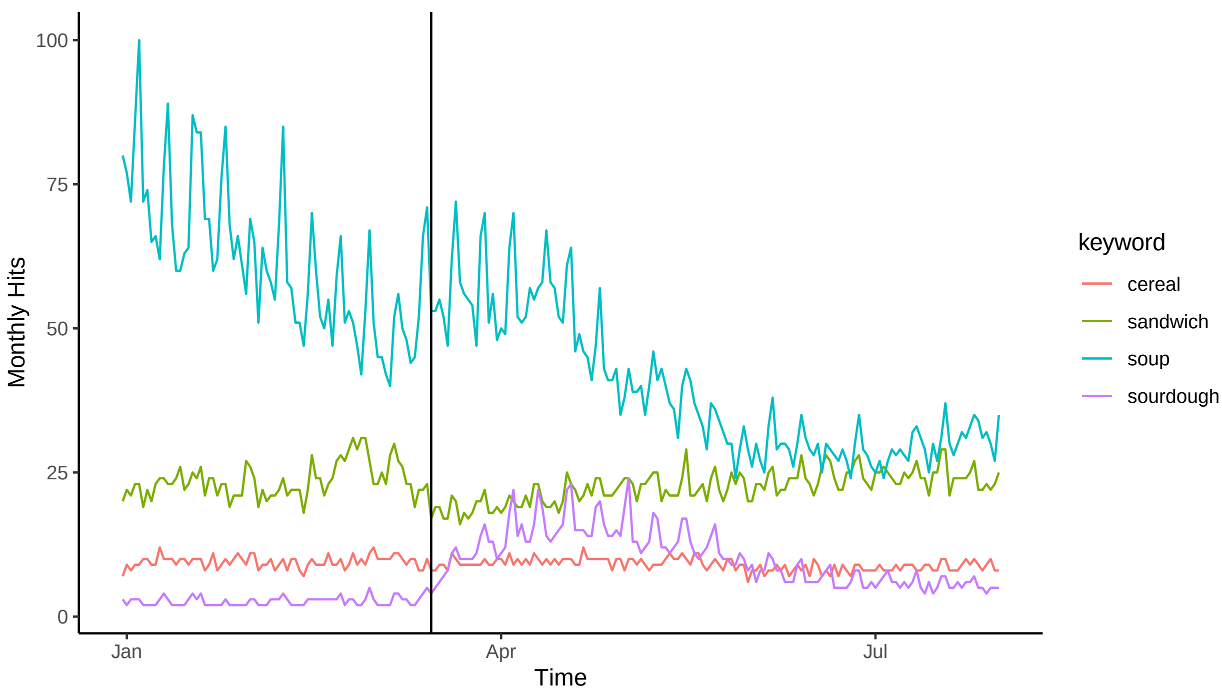
```
 US
856

mydata<-mydata%>%mutate(date=as.Date(date, format="YYYY-MM-DD"))
#View(mydata)
```

#QUESTION ONE A:

(i) Based on the line graph, it could be hypothesized that, after the pandemic, the popularity of the search term "sourdough" increased relative to the controls such as cereal which remained constant over time. Nonetheless, the potential effect of the pandemic on the popularity of the search term "sourdough" seems to be temporal as it declines over time.

(ii) Based on the line graph, soup and sandwich do not seem to be good controls as their search on google or hits were not constant over time.

```
# Group to month level
#mydata <- mydata %>% group_by(date, keyword) %>% summarize(hits = sum(hits,na.rm=T))


#Show hits over time for each Keyword
ggplot(mydata,aes(x=date,y=hits,group=keyword, col=keyword)) + geom_line() +
  theme_classic() + labs(x="Time",y="Monthly Hits")+
  geom_vline(xintercept = as.Date('2020-03-15'),color='black')
```



QUESTION ONE B:

Prior trends differ significantly between the treated and control groups. I will drop the soup and sandwich (in Model 2) as they are not good controls based on the graph.

```
table(mydata$keyword)

   cereal  sandwich      soup sourdough 
      214       214       214       214 

#View(mydata)
# Definition of key variables
mydata <- mydata %>% mutate(treated = ifelse(keyword == "sourdough", 1, 0),
                              interaction = ifelse(keyword == "sourdough",date,0))

# Statistical tests: are the two groups' trends different?
pretrend_t <- lm(hits ~ date + interaction, data=mydata)

#using better control variable and dropping others
# Regression data -- just cereal and sourdough
mydata <- mydata %>% filter(keyword %in% c("sourdough","cereal"))

pretrend_test <- lm(hits ~ date + interaction, data=mydata)
msummary(list(pretrend_t,pretrend_test),
         vcov=c(rep("robust",2)),
         stars=c('*' = .1, '**' = .05, '***' = .01)) # Interpret each coefficient here
```

	Model 1	Model 2
(Intercept)	953.138^{***}	-194.787^{***}
	(175.484)	(41.776)
date	-0.050^{***}	0.011^{***}
	(0.010)	(0.002)
interaction	-0.001^{***}	$-0.000\,08^{***}$
	(0.000 05)	(0.000 02)
Num.Obs.	856	428
R2	0.231	0.070
R2 Adj.	0.229	0.065
AIC	7170.9	2353.3
BIC	7189.9	2369.6
Log.Lik.	−3581.446	−1172.668
F		18.312
RMSE	15.88	3.75
Std.Errors	HC3	HC3

* p < 0.1, ** p < 0.05, *** p < 0.01

QUESTION ONE C:

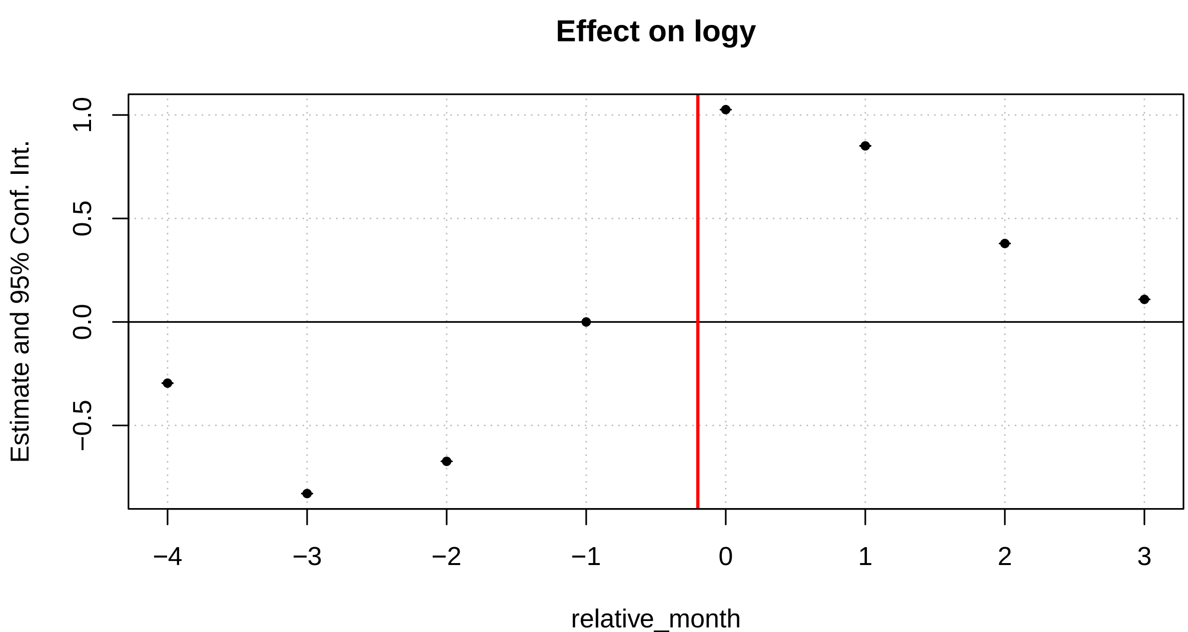
```
# Definition of key variables

mydata$date <- as.yearmon(mydata$date)


mydata <- mydata %>% mutate(the_keyword = ifelse(keyword == "sourdough",1,0))
mydata <- mydata %>% mutate(After = ifelse(date >= "Mar 2020",1,0))
#mydata <- mydata %>% mutate(inter = keyword * post)

table(mydata$hits)

 2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 22 23 24
35 30 12 22 21 15 57 80 75 26 10 10 10  5  6  3  2  3  1  3  1  1
table(mydata$After)

  0   1
122 306
table(mydata$the_keyword)

  0   1
214 214
```

```r
# transform the data #Run pre trends again?
mydata <- mydata %>% mutate(logy = log(hits))


# First, construct relative time variable
mydata <- mydata %>% mutate(relative_month = round((as.numeric(date) - 2020.211)*12))

#View(mydata)
```

#QUESTION ONE D

There is a positive an significant effect of the pandemic on the popularity of the search term "sourdough".

```r
#difference-in-difference estimate of the effect of lockdown on sourdough popularity.
#mydata <- mydata %>% mutate(logy = log(hits))


dte <- feols(logy ~ i(relative_month, the_keyword, ref = -1) |
                date + keyword, data = mydata)

# And use coefplot() for a graph of effects
iplot(dte,ref.line=-0.2,
         ref.line.par=list(col="red",lty=1,lwd=2))
```




```r
msummary(list(dte),
          vcov=c(~keyword),
         stars=c('*' = .1, '**' = .05, '***' = .01)) # Interpret each coefficient here
```

	Model 1
relative_month = -4 × the_keyword	-0.296***
	(1×10^{-14})
relative_month = -3 × the_keyword	-0.829***
	(2×10^{-14})
relative_month = -2 × the_keyword	-0.673***
	(2×10^{-14})
relative_month = 0 × the_keyword	1.026***
	(2×10^{-14})
relative_month = 1 × the_keyword	0.851***
	(2×10^{-14})
relative_month = 2 × the_keyword	0.379***
	(1×10^{-14})
relative_month = 3 × the_keyword	0.109***
	(1×10^{-14})
Num.Obs.	428
R2	0.807
R2 Adj.	0.800
R2 Within	0.630
R2 Within Adj.	0.624
RMSE	0.25
Std.Errors	by: keyword
FE: date	X
FE: keyword	X

* p < 0.1, ** p < 0.05, *** p < 0.01

#QUESTION TWO

#A The rate of living organ donations for Louisiana and other states is stable over time.

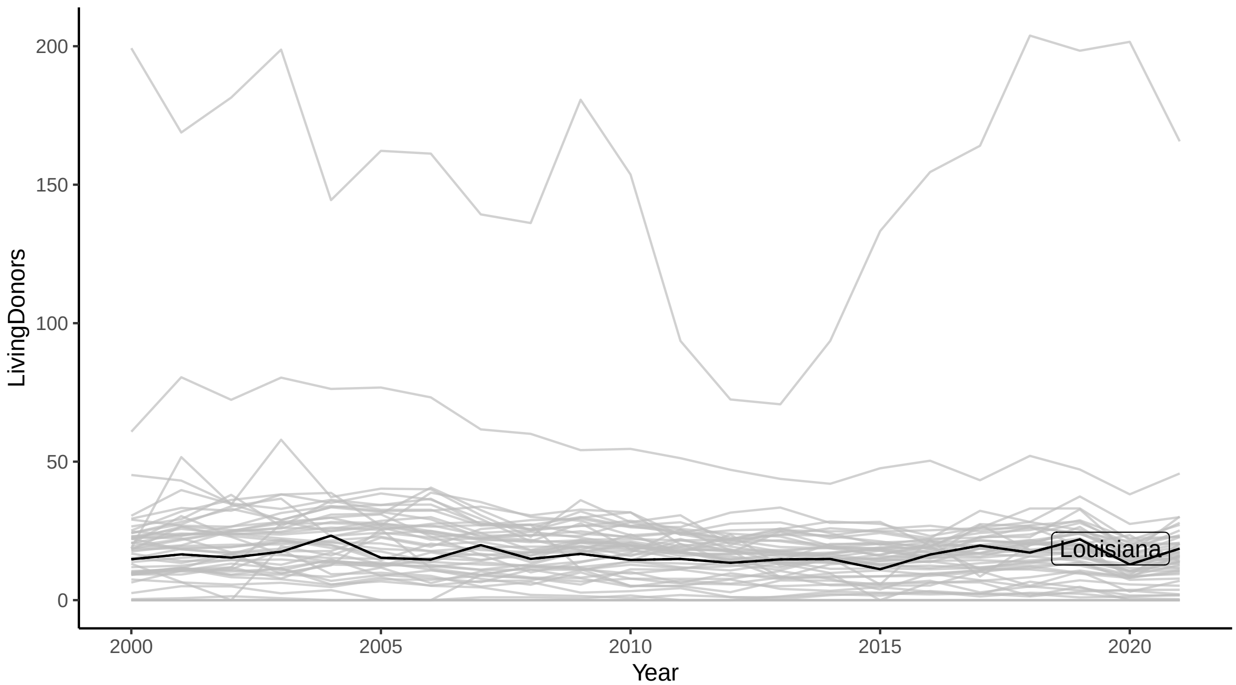
```
mydata2<- read_excel("C:/Users/erico/Desktop/ACADEMICS/MY UofT COURSES/YEAR TWO/SEMESTER

#View(mydata2)
```

```
summary(mydata2$LivingDonors)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0.0    18.0    72.5   118.1   162.0   743.0

summary(mydata2$AllDonors)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0.0    48.0   192.0   272.2   357.0  1822.0


summary(mydata2$WaitinglistAdditions)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0.0   115.0   551.5   993.0  1210.0  7878.0
```

```
#population-adjusted
mydata2$LivingDonors<-mydata2$LivingDonors/mydata2$Population*1000000
mydata2$AllDonors<-mydata2$AllDonors/mydata2$Population*1000000
mydata2$WaitinglistAdditions<-mydata2$WaitinglistAdditions/mydata2$Population*1000000

mydata2$GDP_percapita<-mydata2$GDP/mydata2$Population*1000000

mydata2$GDP_percapita_mean<-mean(mydata2$GDP_percapita)

#install.packages("table1")
library(table1)
table1(~GDP_percapita+LivingDonors+AllDonors+WaitinglistAdditions, data=mydata2)   #or V
```

	Overall
	(N=1122)
GDP_percapita	
Mean (SD)	52500 (19700)
Median [Min, Max]	48700 [30600, 188000]
LivingDonors	
Mean (SD)	19.4 (22.7)
Median [Min, Max]	16.8 [0, 204]
AllDonors	
Mean (SD)	42.8 (46.3)
Median [Min, Max]	38.4 [0, 400]
WaitinglistAdditions	
Mean (SD)	143 (151)
Median [Min, Max]	133 [0, 1710]

```
#rate of living organ donations
#table(mydata2$rate_LivingDonors)

#The mean for all variables

# Some descriptive figures
View(mydata2)

linegraph2<-ggplot(mydata2, aes(x = Year,
                y = LivingDonors,
                group = State)) +
  geom_line() +
  gghighlight(State=="Louisiana",
              label_params = list(fill = NA, alpha=1))
linegraph2+
  geom_line(aes(xintercept=2015),
            linetype=4, colour="black")+
  theme_classic()
```



LivingDonors
200
150
100
50
0
Louisiana
2000
2005
2010
2015
2020
Year

#B

A typical difference-in-difference will not suffice here because, based on the line graph, the rates of living organ donations seems to be mimicking each other for the various states. There is no clear cut effect of a policy. It could be that most states have the policy and hence they may not serve as good controls. Hence, it will be advisable to create a “synthetic control” variable to measure the effect of the policy.

#C

Based on poking around the data, I believe that Alaska, Idaho, Montana, Vermont, and Wyoming should be dropped as the rate of living organ donations were zero for each of them. Our focus is to examine whether a change from tax deductions to tax credit affects living organs donations hence such states where there are no donations may not be helpful in our analysis.

```
#drop those with zeros for living donors

`%!in%` <- Negate(`%in%`)
mydata2 <- mydata2 %>% filter(State  %!in% c("Alaska", "Idaho", "Montana",
                                             "Vermont", "Wyoming"))
```

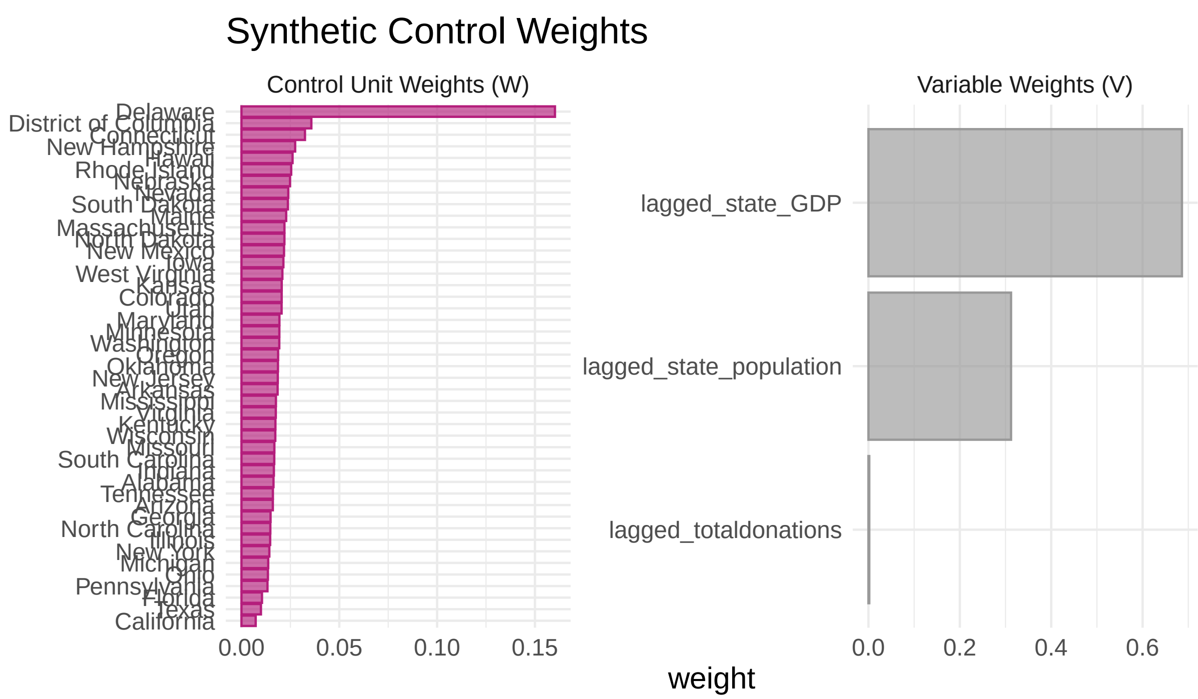
#D

The synthetic control weights for lagged GDP and Delaware were relatively higher.

```r
mydata2$centered_year<-mydata2$Year-2015
# Construct Synthetic Controls
donation_out <-
  mydata2  %>%
  # initial the synthetic control object
  synthetic_control(outcome = LivingDonors, # outcome
                    unit = State, # unit index in the panel data
                    time = Year, # time index in the panel data
                    i_unit = "Louisiana", # unit where the intervention occurred
                    i_time = 2015, # time period when the intervention occurred
                    generate_placebos=T # generate placebo synthetic controls (for infe
  ) %>%
    generate_predictor(time_window = 2000:2014,
    lagged_state_population=mean(Population, na.rm=T),
    lagged_state_GDP=mean(GDP, na.rm=T),
    lagged_totaldonations = mean(AllDonors, na.rm=T)) %>%
    # Generate the fitted weights for the synthetic control
  generate_weights(optimization_window = 2000:2014, # time to use in the optimization t
                   margin_ipop = .02,sigf_ipop = 7,bound_ipop = 6 # optimizer options
  ) %>%
  # Generate the synthetic control
  generate_control()


#relative weights

# Which states are we using, and what weights are they given?
donation_out %>%
  grab_unit_weights() %>%
  mutate(weights = round(weight, digits = 4)) %>%
  select(unit, weights) %>%
  filter(weights>0.0001) %>%
  as.data.frame() %>%
  stargazer(summary = FALSE, rownames = FALSE,type="text")

============================
unit                 weights
----------------------------
Alabama               0.016
```

```
Arizona              0.016
Arkansas             0.018
California           0.007
Colorado             0.020
Connecticut          0.032
Delaware             0.160
District of Columbia 0.036
Florida              0.011
Georgia              0.015
Hawaii               0.026
Illinois             0.015
Indiana              0.017
Iowa                 0.021
Kansas               0.021
Kentucky             0.017
Maine                0.023
Maryland             0.019
Massachusetts        0.022
Michigan             0.014
Minnesota            0.019
Mississippi          0.018
Missouri             0.017
Nebraska             0.025
Nevada               0.024
New Hampshire        0.027
New Jersey           0.019
New Mexico           0.022
New York             0.014
North Carolina       0.015
North Dakota         0.022
Ohio                 0.014
Oklahoma             0.019
Oregon               0.019
Pennsylvania         0.013
Rhode Island         0.025
South Carolina       0.017
South Dakota         0.024
Tennessee            0.016
Texas                0.010
Utah                 0.020
Virginia             0.018
Washington           0.019
West Virginia        0.021
Wisconsin            0.017
```

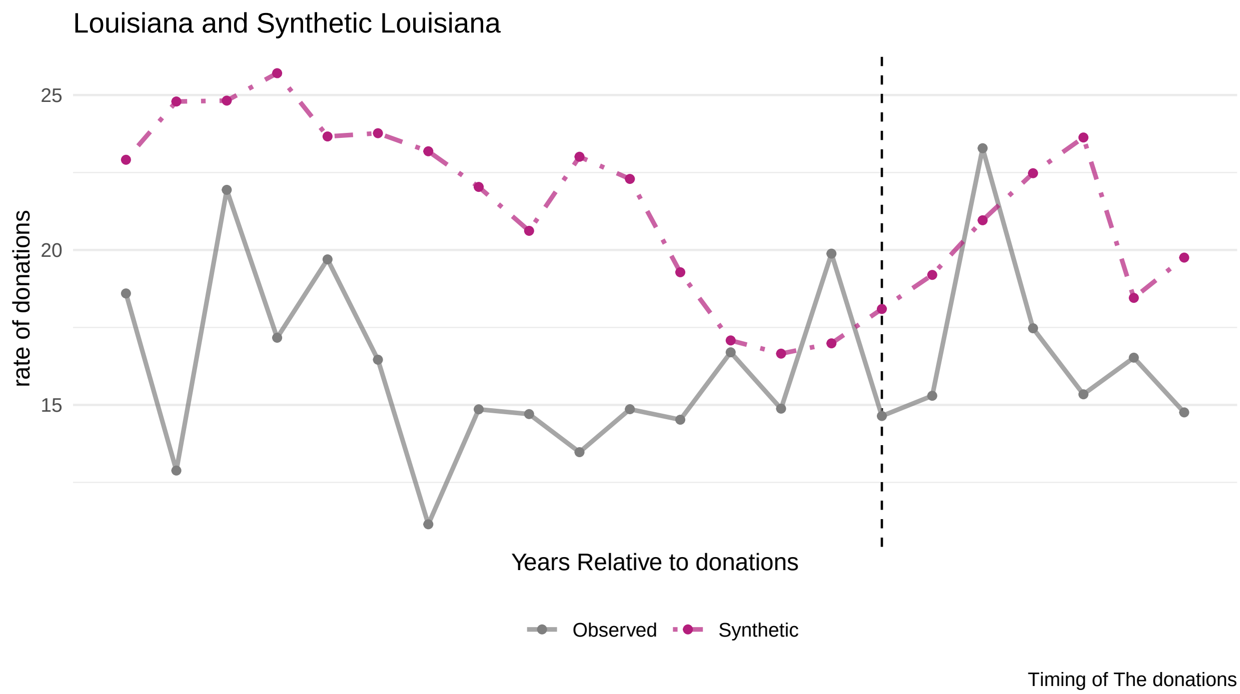
```
----------------------------

# What about the independent variables?
donation_out %>%
  plot_weights() +
  labs(title="Synthetic Control Weights")
```



#E

The synthetic control construction was not successful because, graphically, the "before the policy" trend were not close to each other. The reason could be that the real or observed Louisiana trend of living organs donations may be affected by time varying factors which do not affect the synthetic trend.

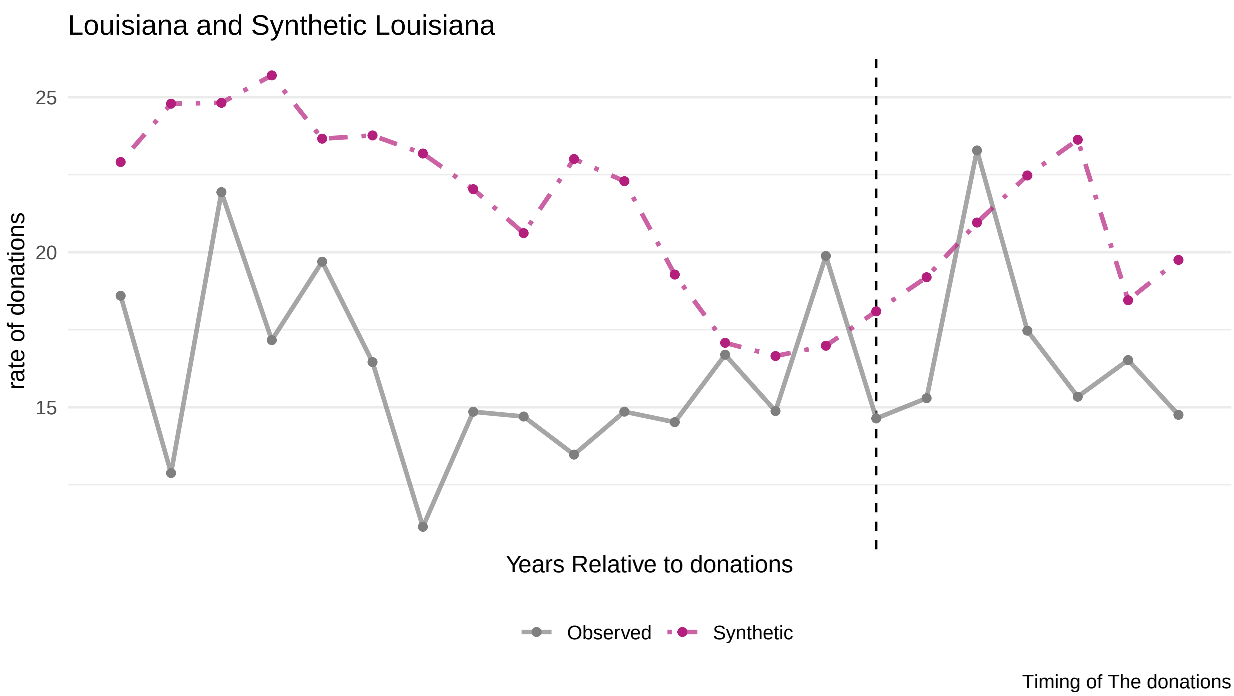
```
# Balance Table
donation_out %>%
  grab_balance_table() %>%
  mutate(difference = Louisiana - synthetic_Louisiana) %>%
  select(variable, Louisiana, synthetic_Louisiana, difference, donor_sample) %>%
  as.data.frame() %>%
  stargazer(summary = FALSE, rownames = FALSE,
            caption = "Balance Table",
            label = "balancetable",type="text") #

===========================================================================
lagged_state_GDP         230,341.000   230,340.900   0.089   325,360.800
lagged_state_population 4,512,806.000 4,512,810.000 -4.484  6,489,964.000
lagged_totaldonations      43.279        43.279     -0.0001    45.120
-------------------------------------------------------------------------

=============
Balance Table
-------------


donation_out %>% plot_trends() +
  scale_x_continuous(breaks = c(-15,-10,-5,0,5)) +
  labs(
    title = "Louisiana and Synthetic Louisiana",
    caption = "Timing of The donations",
    x="Years Relative to donations",
    y="rate of donations"
  )
```


Louisiana and Synthetic Louisiana
25
20
15
rate of donations
Years Relative to donations
Observed
Synthetic
Timing of The donations

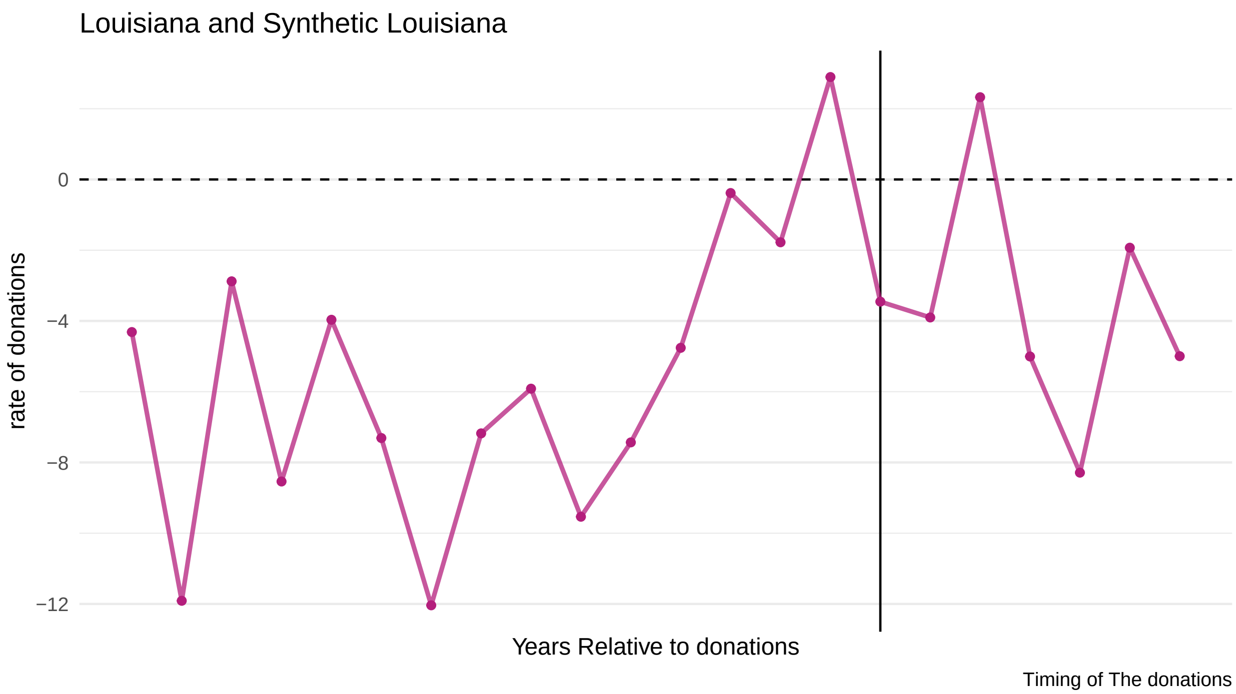
#F Checking effectiveness

It can be realised that the policy only had a temporal positive effect on the rate of living organs donations but it declines over time.

```r
#So what's the effect? To find out, let's compare Louisiana and Synthetic Louisiana ov

donation_out %>% plot_trends() +
  scale_x_continuous(breaks = c(-15,-10,-5,0,5)) +
  labs(
    title = "Louisiana and Synthetic Louisiana",
    caption = "Timing of The donations",
    x="Years Relative to donations",
    y="rate of donations"
  )
```




```r
#So what's the effect? To find out, let's compare Louisiana and Synthetic Louisiana ov
# Plot Model Differences
donation_out %>% plot_differences() +
  scale_x_continuous(breaks = c(-15,-10,-5,0,5)) +
  labs(
   title = "Louisiana and Synthetic Louisiana",
    caption = "Timing of The donations",
    x="Years Relative to donations",
    y="rate of donations"
  )
```

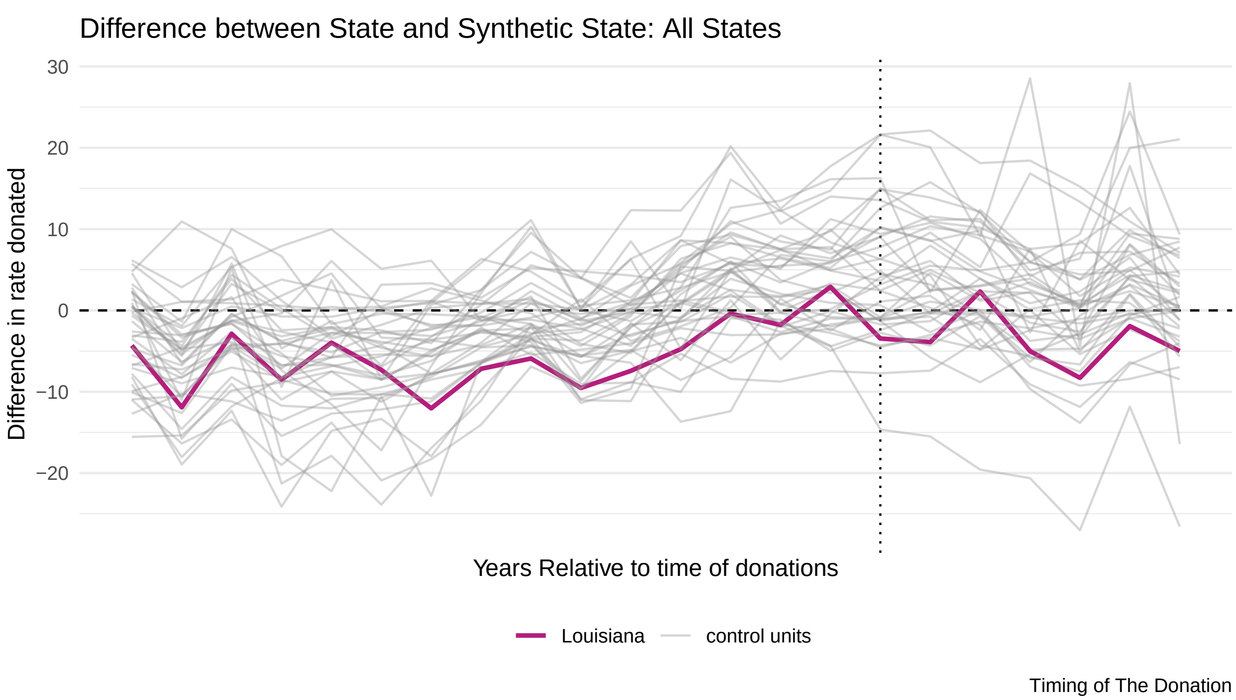


Louisiana and Synthetic Louisiana
0
−4
−8
−12
rate of donations
Years Relative to donations
Timing of The donations

#G

The change of date of implementation do not have any impact on the effectiveness of the policy. It is still not increasing the rate of living organ donations.

```
#i:
# Plot placebos of different states' assignments
### Inference
#How do we get a sense of whether the effect was significant? To do this, let's plot t


donation_out %>% plot_placebos() +
  scale_x_continuous(breaks = c(-15,-10,-5,0,5)) +
  labs(
    title = "Difference between State and Synthetic State: All States",
    caption = "Timing of The Donation",
    x="Years Relative to time of donations",
    y="Difference in rate donated"
  )
```

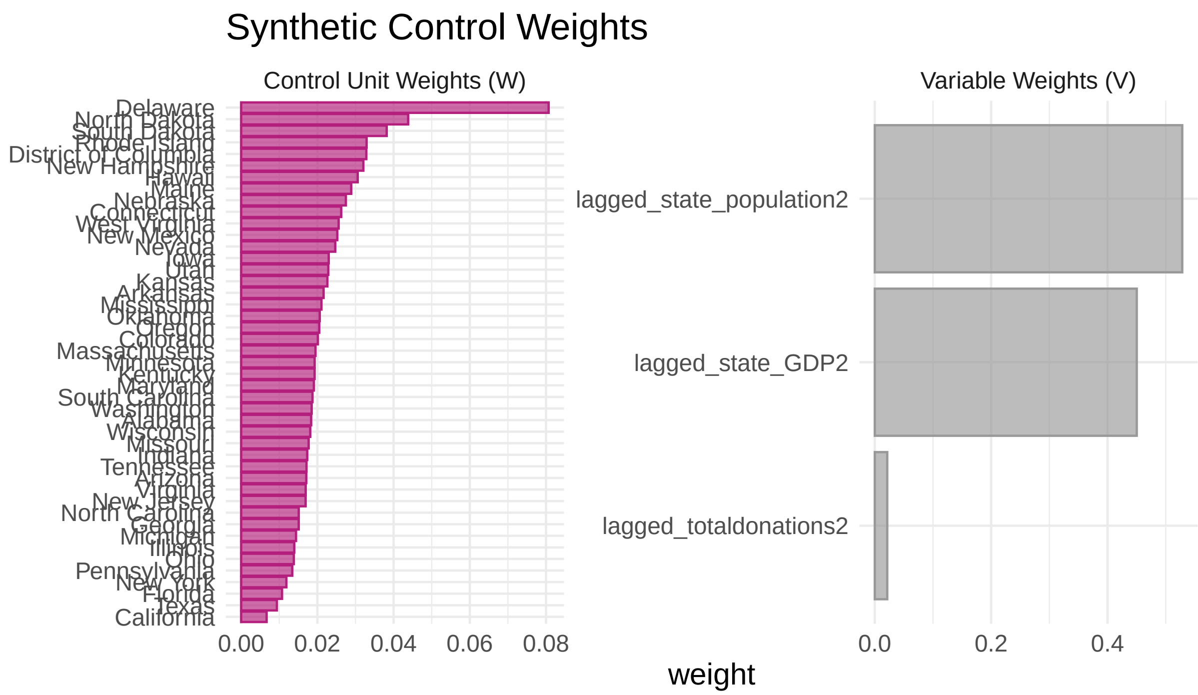



```
#ii

# This test shifts the pre-treatment window back five years
```

```
placebo_out <-
  mydata2 %>%
  filter(Year <= 2018) %>%
  # initial the synthetic control object
  synthetic_control(outcome = LivingDonors, # outcome
                    unit = State, # unit index in the panel data
                    time = Year, # time index in the panel data
                    i_unit = "Louisiana", # unit where the intervention occurred
                    i_time = 2018, # time period when the intervention occurred
                    generate_placebos=T # generate placebo synthetic controls (for infe
  ) %>%
    generate_predictor(time_window = 2000:2017,
    lagged_state_population2=mean(Population, na.rm=T),
    lagged_state_GDP2=mean(GDP, na.rm=T),
    lagged_totaldonations2 = mean(AllDonors, na.rm=T)) %>%
    # Generate the fitted weights for the synthetic control
  generate_weights(optimization_window = 2000:2017, # time to use in the optimization t
                   margin_ipop = .02,sigf_ipop = 7,bound_ipop = 6 # optimizer options
  ) %>%
  # Generate the synthetic control
  generate_control()

  ###########################

#relative weights

# Which states are we using, and what weights are they given?
placebo_out %>%
  grab_unit_weights() %>%
  mutate(weights = round(weight, digits = 4)) %>%
  select(unit, weights) %>%
  filter(weights>0.0001) %>%
  as.data.frame() %>%
  stargazer(summary = FALSE, rownames = FALSE,type="text")

============================
unit                 weights
----------------------------
Alabama                0.018
Arizona                0.017
Arkansas               0.022
California             0.007
Colorado               0.020
Connecticut            0.026
```

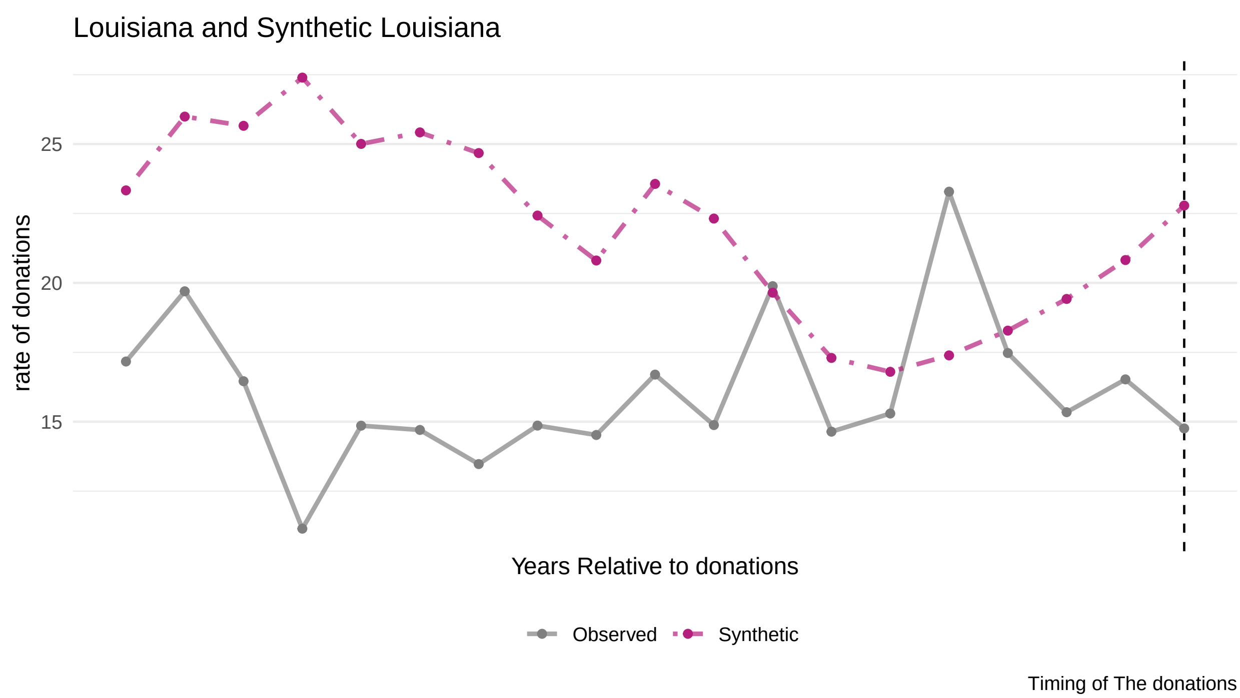
```
Delaware                  0.081
District of Columbia      0.033
Florida                   0.011
Georgia                   0.015
Hawaii                    0.031
Illinois                  0.014
Indiana                   0.017
Iowa                      0.023
Kansas                    0.023
Kentucky                  0.019
Maine                     0.029
Maryland                  0.019
Massachusetts             0.020
Michigan                  0.014
Minnesota                 0.019
Mississippi               0.021
Missouri                  0.018
Nebraska                  0.028
Nevada                    0.025
New Hampshire             0.032
New Jersey                0.017
New Mexico                0.025
New York                  0.012
North Carolina            0.015
North Dakota              0.044
Ohio                      0.014
Oklahoma                  0.021
Oregon                    0.020
Pennsylvania              0.013
Rhode Island              0.033
South Carolina            0.019
South Dakota              0.038
Tennessee                 0.017
Texas                     0.009
Utah                      0.023
Virginia                  0.017
Washington                0.018
West Virginia             0.026
Wisconsin                 0.018
----------------------------

# What about the independent variables?
placebo_out %>%
  plot_weights() +
```

```
  labs(title="Synthetic Control Weights")
```

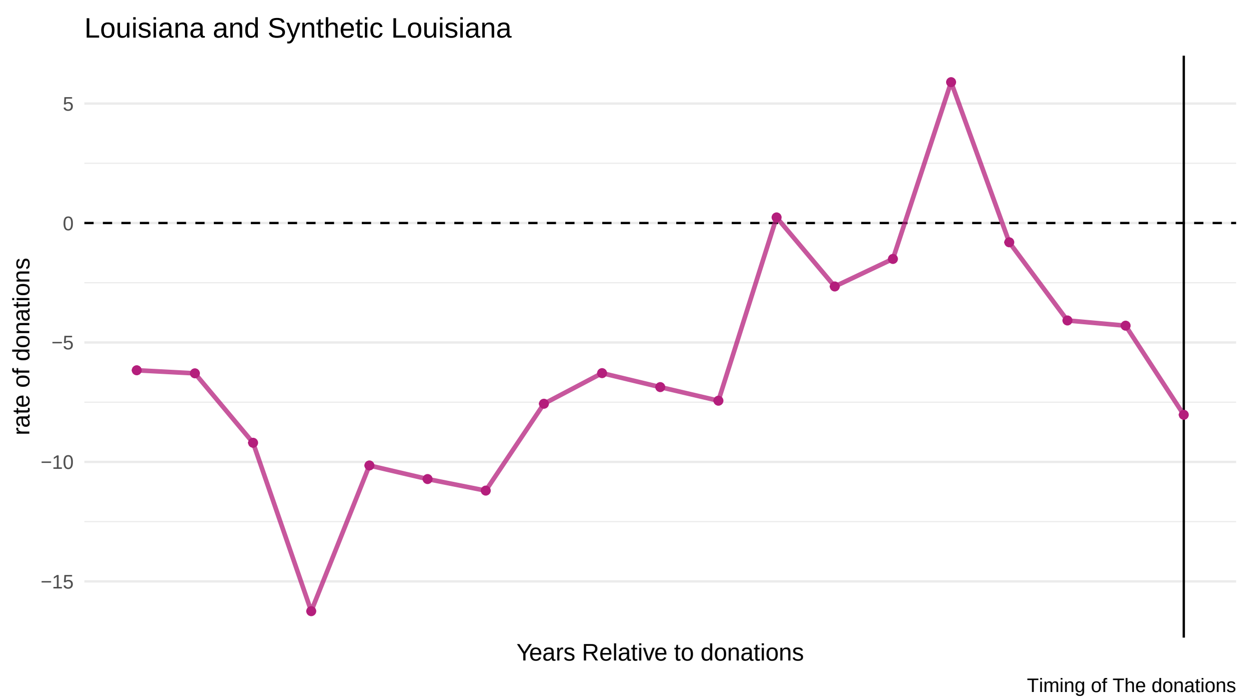



```
# Balance Table
placebo_out %>%
  grab_balance_table() %>%
  mutate(difference = Louisiana - synthetic_Louisiana) %>%
  select(variable, Louisiana, synthetic_Louisiana, difference, donor_sample) %>%
  as.data.frame() %>%
  stargazer(summary = FALSE, rownames = FALSE,
            caption = "Balance Table",
            label = "balancetable",type="text") #

===========================================================================
lagged_state_GDP2          230,574.100   230,574.000   0.047     334,483.200
lagged_state_population2 4,539,189.000 4,539,190.000  -1.056   6,569,455.000
lagged_totaldonations2       44.403        44.403     -0.00000     45.428
---------------------------------------------------------------------------

=============
Balance Table
-------------
```

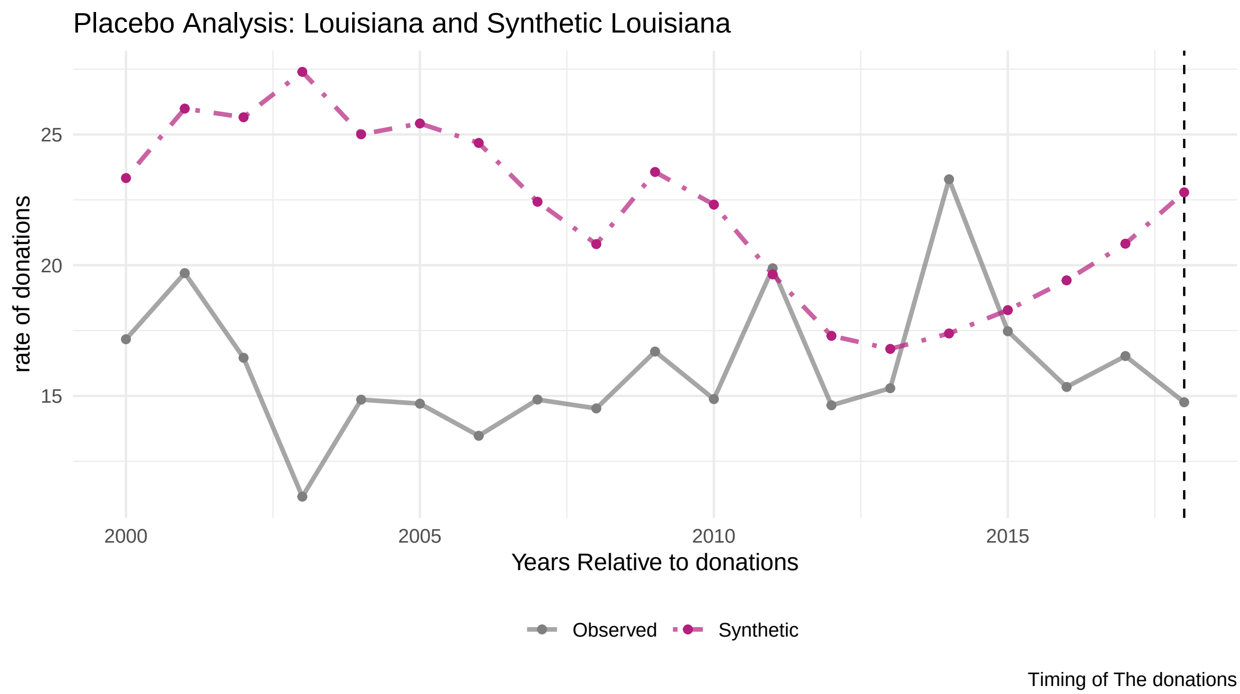
```
placebo_out %>% plot_trends() +
  scale_x_continuous(breaks = c(-15,-10,-5,0,5)) +
  labs(
    title = "Louisiana and Synthetic Louisiana",
    caption = "Timing of The donations",
    x="Years Relative to donations",
    y="rate of donations"
  )
```



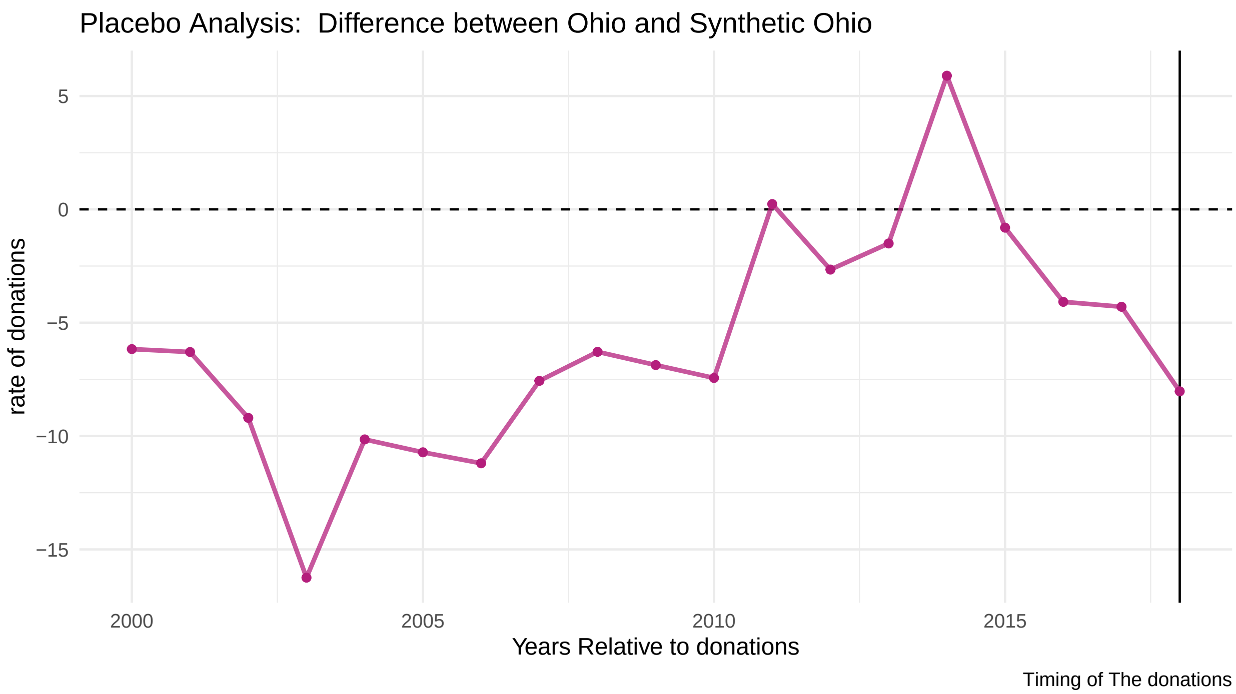

```
#So what's the effect? To find out, let's compare Louisiana and Synthetic Louisiana ov
# Plot Model Differences
placebo_out %>% plot_differences() +
  scale_x_continuous(breaks = c(-15,-10,-5,0,5)) +
  labs(
   title = "Louisiana and Synthetic Louisiana",
    caption = "Timing of The donations",
    x="Years Relative to donations",
    y="rate of donations"
  )
```




```
####################





  ####################

placebo_out %>% plot_trends()  +
  labs(
    title = "Placebo Analysis: Louisiana and Synthetic Louisiana",
  caption = "Timing of The donations",
    x="Years Relative to donations",
    y="rate of donations"
  )
```




```
placebo_out %>% plot_differences()  +
  labs(
    title = "Placebo Analysis:  Difference between Ohio and Synthetic Ohio",
   caption = "Timing of The donations",
    x="Years Relative to donations",
    y="rate of donations"
  )
```




```
#placebo_out %>% grab_signficance() %>% filter(unit_name=="Louisiana")
#placebo_out %>% grab_unit_weights() %>% arrange(desc(weight))

#placebo_out %>% plot_mspe_ratio()
```

#H

Moving from tax deductions to tax credits may have a long-term detrimental effect on the rate of living organ donations.

```r
packages <- knitr::write_bib(file = 'packages.bib')
packages
```

$AER @Manual{R-AER, title = {AER: Applied Econometrics with R}, author = {Christian Kleiber and Achim Zeileis}, year = {2022}, note = {R package version 1.2-10}, url = {https://CRAN.R-project.org/package=AER}, }

$base @Manual{R-base, title = {R: A Language and Environment for Statistical Computing}, author = {{R Core Team}}, organization = {R Foundation for Statistical Computing}, address = {Vienna, Austria}, year = {2022}, url = {https://www.R-project.org/}, }

$binsreg @Manual{R-binsreg, title = {binsreg: Binscatter Estimation and Inference}, author = {Matias D. Cattaneo and Richard K. Crump and Max H. Farrell and Yingjie Feng}, year = {2021}, note = {R package version 0.7}, url = {https://CRAN.R-project.org/package=binsreg}, }

$broom @Manual{R-broom, title = {broom: Convert Statistical Objects into Tidy Tibbles}, author = {David Robinson and Alex Hayes and Simon Couch}, year = {2022}, note = {R package version 1.0.1}, url = {https://CRAN.R-project.org/package=broom}, }

$car @Manual{R-car, title = {car: Companion to Applied Regression}, author = {John Fox and Sanford Weisberg and Brad Price}, year = {2022}, note = {R package version 3.1-1}, url = {https://CRAN.R-project.org/package=car}, }

$carData @Manual{R-carData, title = {carData: Companion to Applied Regression Data Sets}, author = {John Fox and Sanford Weisberg and Brad Price}, year = {2022}, note = {R package version 3.0-5}, url = {https://CRAN.R-project.org/package=carData}, }

$causaldata @Manual{R-causaldata, title = {causaldata: Example Data Sets for Causal Inference Textbooks}, author = {Nick Huntington-Klein and Malcolm Barrett}, year = {2021}, note = {R package version 0.1.3}, url = {https://github.com/NickCH-K/causaldata}, }

$dagitty @Manual{R-dagitty, title = {dagitty: Graphical Analysis of Structural Causal Models}, author = {Johannes Textor and Benito {van der Zander} and Ankur Ankan}, year = {2021}, note = {R package version 0.3-1}, url = {https://CRAN.R-project.org/package=dagitty}, }

$data.table @Manual{R-data.table, title = {data.table: Extension of `data.frame`}, author = {Matt Dowle and Arun Srinivasan}, year = {2022}, note = {R package version 1.14.4}, url = {https://CRAN.R-project.org/package=data.table}, }

$dplyr @Manual{R-dplyr, title = {dplyr: A Grammar of Data Manipulation}, author = {Hadley Wickham and Romain François and Lionel Henry and Kirill Müller}, year = {2022}, note = {R package version 1.0.10}, url = {https://CRAN.R-project.org/package=dplyr}, }

$expss @Manual{R-expss, title = {expss: Tables, Labels and Some Useful Functions from Spreadsheets and SPSS Statistics}, author = {Gregory Demin}, year = {2022}, note = {R package version 0.11.2}, url = {https://gdemin.github.io/expss/}, }

$fixest @Manual{R-fixest, title = {fixest: Fast Fixed-Effects Estimations}, author = {Laurent Berge}, year = {2022}, note = {R package version 0.11.0}, url = {https://CRAN.R-project.org/package=fixest}, }

$forcats @Manual{R-forcats, title = {forcats: Tools for Working with Categorical Variables (Factors)}, author = {Hadley Wickham}, year = {2022}, note = {R package version 0.5.2}, url = {https://CRAN.R-project.org/package=forcats}, }

$foreign @Manual{R-foreign, title = {foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, ...}, author = {{R Core Team}}, year = {2022}, note = {R package version 0.8-83}, url = {https://svn.r-project.org/R-packages/trunk/foreign/}, }

$ggdag @Manual{R-ggdag, title = {ggdag: Analyze and Create Elegant Directed Acyclic Graphs}, author = {Malcolm Barrett}, year = {2022}, note = {R package version 0.2.7}, url = {https://github.com/malcolmbarrett/ggdag}, }

$gghighlight @Manual{R-gghighlight, title = {gghighlight: Highlight Lines and Points in ggplot2}, author = {Hiroaki Yutani}, year = {2022}, note = {R package version 0.4.0}, url = {https://CRAN.R-project.org/package=gghighlight}, }

$ggplot2 @Manual{R-ggplot2, title = {ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics}, author = {Hadley Wickham and Winston Chang and Lionel Henry and Thomas Lin Pedersen and Kohske Takahashi and Claus Wilke and Kara Woo and Hiroaki Yutani and Dewey Dunnington}, year = {2022}, note = {R package version 3.3.6}, url = {https://CRAN.R-project.org/package=ggplot2}, }

$gsynth @Manual{R-gsynth, title = {gsynth: Generalized Synthetic Control Method}, author = {Yiqing Xu and Licheng Liu}, year = {2021}, note = {R package version 1.2.1}, url = {https://yiqingxu.org/packages/gsynth/gsynth_examples.html}, }

$here @Manual{R-here, title = {here: A Simpler Way to Find Your Files}, author = {Kirill Müller}, year = {2020}, note = {R package version 1.0.1}, url = {https://CRAN.R-project.org/package=here}, }

$kableExtra @Manual{R-kableExtra, title = {kableExtra: Construct Complex Table with kable and Pipe Syntax}, author = {Hao Zhu}, year = {2021}, note = {R package version 1.3.4}, url = {https://CRAN.R-project.org/package=kableExtra}, }

$knitr @Manual{R-knitr, title = {knitr: A General-Purpose Package for Dynamic Report Generation in R}, author = {Yihui Xie}, year = {2022}, note = {R package version 1.40}, url = {https://yihui.org/knitr/}, }

$lme4 @Manual{R-lme4, title = {lme4: Linear Mixed-Effects Models using Eigen and S4}, author = {Douglas Bates and Martin Maechler and Ben Bolker and Steven Walker}, year = {2022}, note = {R package version 1.1-31}, url = {https://github.com/lme4/lme4/}, }

$lmtest @Manual{R-lmtest, title = {lmtest: Testing Linear Regression Models}, author = {Torsten Hothorn and Achim Zeileis and Richard W. Farebrother and Clint Cummins},

year = {2022}, note = {R package version 0.9-40}, url = {https://CRAN.R-project.org/package=lmtest}, }

$lubridate @Manual{R-lubridate, title = {lubridate: Make Dealing with Dates a Little Easier}, author = {Vitalie Spinu and Garrett Grolemund and Hadley Wickham}, year = {2021}, note = {R package version 1.8.0}, url = {https://CRAN.R-project.org/package=lubridate}, }

$maditr @Manual{R-maditr, title = {maditr: Fast Data Aggregation, Modification, and Filtering with Pipes and data.table}, author = {Gregory Demin}, year = {2022}, note = {R package version 0.8.3}, url = {https://github.com/gdemin/maditr}, }

$marginaleffects @Manual{R-marginaleffects, title = {marginaleffects: Marginal Effects, Marginal Means, Predictions, and Contrasts}, author = {Vincent Arel-Bundock}, year = {2022}, note = {R package version 0.8.0}, url = {https://vincentarelbundock.github.io/marginaleffects/}, }

$Matrix @Manual{R-Matrix, title = {Matrix: Sparse and Dense Matrix Classes and Methods}, author = {Douglas Bates and Martin Maechler and Mikael Jagan}, year = {2022}, note = {R package version 1.5-1}, url = {https://CRAN.R-project.org/package=Matrix}, }

$modelsummary @Manual{R-modelsummary, title = {modelsummary: Summary Tables and Plots for Statistical Models and Data: Beautiful, Customizable, and Publication-Ready}, author = {Vincent Arel-Bundock}, year = {2022}, note = {R package version 1.1.0}, url = {https://vincentarelbundock.github.io/modelsummary/}, }

$nprobust @Manual{R-nprobust, title = {nprobust: Nonparametric Robust Estimation and Inference Methods using Local Polynomial Regression and Kernel Density Estimation}, author = {Sebastian Calonico and Matias D. Cattaneo and Max H. Farrell}, year = {2020}, note = {R package version 0.4.0}, url = {https://CRAN.R-project.org/package=nprobust}, }

$psych @Manual{R-psych, title = {psych: Procedures for Psychological, Psychometric, and Personality Research}, author = {William Revelle}, year = {2022}, note = {R package version 2.2.9}, url = {https://personality-project.org/r/psych/ https://personality-project.org/r/psych-manual.pdf}, }

$purrr @Manual{R-purrr, title = {purrr: Functional Programming Tools}, author = {Lionel Henry and Hadley Wickham}, year = {2022}, note = {R package version 0.3.5}, url = {https://CRAN.R-project.org/package=purrr}, }

$readr @Manual{R-readr, title = {readr: Read Rectangular Text Data}, author = {Hadley Wickham and Jim Hester and Jennifer Bryan}, year = {2022}, note = {R package version 2.1.3}, url = {https://CRAN.R-project.org/package=readr}, }

$readstata13 @Manual{R-readstata13, title = {readstata13: Import Stata Data Files}, author = {Jan Marvin Garbuszus and Sebastian Jeworutzki}, year = {2021}, note = {R package version 0.10.0}, url = {https://github.com/sjewo/readstata13}, }

$readxl @Manual{R-readxl, title = {readxl: Read Excel Files}, author = {Hadley Wickham and Jennifer Bryan}, year = {2022}, note = {R package version 1.4.1}, url = {https://CRAN.R-project.org/package=readxl}, }

$rlang @Manual{R-rlang, title = {rlang: Functions for Base Types and Core R and Tidyverse Features}, author = {Lionel Henry and Hadley Wickham}, year = {2022}, note = {R package version 1.0.6}, url = {https://CRAN.R-project.org/package=rlang}, }

$sandwich @Manual{R-sandwich, title = {sandwich: Robust Covariance Matrix Estimators}, author = {Achim Zeileis and Thomas Lumley}, year = {2022}, note = {R package version 3.0-2}, url = {https://sandwich.R-Forge.R-project.org/}, }

$stargazer @Manual{R-stargazer, title = {stargazer: Well-Formatted Regression and Summary Statistics Tables}, author = {Marek Hlavac}, year = {2022}, note = {R package version 5.2.3}, url = {https://CRAN.R-project.org/package=stargazer}, }

$stringr @Manual{R-stringr, title = {stringr: Simple, Consistent Wrappers for Common String Operations}, author = {Hadley Wickham}, year = {2022}, note = {R package version 1.4.1}, url = {https://CRAN.R-project.org/package=stringr}, }

$survival @Manual{R-survival, title = {survival: Survival Analysis}, author = {Terry M Therneau}, year = {2022}, note = {R package version 3.4-0}, url = {https://github.com/therneau/survival}, }

$table1 @Manual{R-table1, title = {table1: Tables of Descriptive Statistics in HTML}, author = {Benjamin Rich}, year = {2021}, note = {R package version 1.4.2}, url = {https://github.com/benjaminrich/table1}, }

$tibble @Manual{R-tibble, title = {tibble: Simple Data Frames}, author = {Kirill Müller and Hadley Wickham}, year = {2022}, note = {R package version 3.1.8}, url = {https://CRAN.R-project.org/package=tibble}, }

$tidyr @Manual{R-tidyr, title = {tidyr: Tidy Messy Data}, author = {Hadley Wickham and Maximilian Girlich}, year = {2022}, note = {R package version 1.2.1}, url = {https://CRAN.R-project.org/package=tidyr}, }

$tidysynth @Manual{R-tidysynth, title = {tidysynth: A Tidy Implementation of the Synthetic Control Method}, author = {Eric Dunford}, year = {2021}, note = {R package version 0.1.0}, url = {https://CRAN.R-project.org/package=tidysynth}, }

$tidyverse @Manual{R-tidyverse, title = {tidyverse: Easily Install and Load the Tidyverse}, author = {Hadley Wickham}, year = {2022}, note = {R package version 1.3.2}, url = {https://CRAN.R-project.org/package=tidyverse}, }

$zoo @Manual{R-zoo, title = {zoo: S3 Infrastructure for Regular and Irregular Time Series (Z's Ordered Observations)}, author = {Achim Zeileis and Gabor Grothendieck and Jeffrey A. Ryan}, year = {2022}, note = {R package version 1.8-11}, url = {https://zoo.R-Forge.R-project.org/}, }

[[46]] @Book{AER2008, title = {Applied Econometrics with {R}}, author = {Christian Kleiber and Achim Zeileis}, year = {2008}, publisher = {Springer-Verlag}, address = {New York}, note = {{ISBN} 978-0-387-77316-2}, url = {https://CRAN.R-project.org/package=AER}, } [[47]] @Book{car2019, title = {An {R} Companion to Applied Regression}, edition = {Third}, author = {John Fox and Sanford Weisberg}, year = {2019}, publisher = {Sage}, address = {Thousand Oaks {CA}}, url = {https://socialsciences.mcmaster.ca/jfox/Books/Companion/}, }

[[48]] @Article{dagitty2016, title = {Robust causal inference using directed acyclic graphs: the R package 'dagitty'}, journal = {International Journal of Epidemiology}, author = {Johannes Textor and Benito {van der Zander} and Mark S Gilthorpe and Maciej Liśkiewicz and George TH Ellison}, volume = {45}, number = {6}, pages = {1887–1894}, year = {2016}, doi = {10.1093/ije/dyw341}, }

[[49]] @Article{fixest2018, title = {Efficient estimation of maximum likelihood models with multiple fixed-effects: the {R} package {FENmlm}}, author = {Laurent Berg'e}, year = {2018}, journal = {CREA Discussion Papers}, number = {13}, }

[[50]] @Book{ggplot22016, author = {Hadley Wickham}, title = {ggplot2: Elegant Graphics for Data Analysis}, publisher = {Springer-Verlag New York}, year = {2016}, isbn = {978-3-319-24277-4}, url = {https://ggplot2.tidyverse.org}, }

[[51]] @Book{knitr2015, title = {Dynamic Documents with {R} and knitr}, author = {Yihui Xie}, publisher = {Chapman and Hall/CRC}, address = {Boca Raton, Florida}, year = {2015}, edition = {2nd}, note = {ISBN 978-1498716963}, url = {https://yihui.org/knitr/}, }

[[52]] @InCollection{knitr2014, booktitle = {Implementing Reproducible Computational Research}, editor = {Victoria Stodden and Friedrich Leisch and Roger D. Peng}, title = {knitr: A Comprehensive Tool for Reproducible Research in {R}}, author = {Yihui Xie}, publisher = {Chapman and Hall/CRC}, year = {2014}, note = {ISBN 978-1466561595}, url = {http://www.crcpress.com/product/isbn/9781466561595}, }

[[53]] @Article{lme42015, title = {Fitting Linear Mixed-Effects Models Using {lme4}}, author = {Douglas Bates and Martin M{"a}chler and Ben Bolker and Steve Walker}, journal = {Journal of Statistical Software}, year = {2015}, volume = {67}, number = {1}, pages = {1–48}, doi = {10.18637/jss.v067.i01}, }

[[54]] @Article{lmtest2002, title = {Diagnostic Checking in Regression Relationships}, author = {Achim Zeileis and Torsten Hothorn}, journal = {R News}, year = {2002}, volume = {2}, number = {3}, pages = {7–10}, url = {https://CRAN.R-project.org/doc/Rnews/}, }

[[55]] @Article{lubridate2011, title = {Dates and Times Made Easy with {lubridate}}, author = {Garrett Grolemund and Hadley Wickham}, journal = {Journal of Statistical Software}, year = {2011}, volume = {40}, number = {3}, pages = {1–25}, url = {https://www.jstatsoft.org/v40/i03/}, }

[[56]] @Article{modelsummary2022, title = {{modelsummary}: Data and Model Summaries in {R}}, author = {Vincent Arel-Bundock}, journal = {Journal of Statistical Software}, year = {2022}, volume = {103}, number = {1}, pages = {1–23}, doi = {10.18637/jss.v103.i01}, } [[57]] @Article{nprobust2019, title = {{nprobust}: Nonparametric Kernel-Based Estimation and Robust Bias-Corrected Inference}, author = {Sebastian Calonico and Matias D. Cattaneo and Max H. Farrell}, journal = {Journal of Statistical Software}, year = {2019}, volume = {91}, number = {8}, pages = {1–33}, doi = {10.18637/jss.v091.i08}, }

[[58]] @Article{sandwich2020, title = {Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in {R}}, author = {Achim Zeileis and Susanne K"oll and Nathaniel Graham}, journal = {Journal of Statistical Software}, year = {2020}, volume = {95}, number = {1}, pages = {1–36}, doi = {10.18637/jss.v095.i01}, }

[[59]] @Article{sandwich2004, title = {Econometric Computing with {HC} and {HAC} Covariance Matrix Estimators}, author = {Achim Zeileis}, journal = {Journal of Statistical Software}, year = {2004}, volume = {11}, number = {10}, pages = {1–17}, doi = {10.18637/jss.v011.i10}, }

[[60]] @Article{sandwich2006, title = {Object-Oriented Computation of Sandwich Estimators}, author = {Achim Zeileis}, journal = {Journal of Statistical Software}, year = {2006}, volume = {16}, number = {9}, pages = {1–16}, doi = {10.18637/jss.v016.i09}, }

[[61]] @Book{survival-book, title = {Modeling Survival Data: Extending the {C}ox Model}, author = {{Terry M. Therneau} and {Patricia M. Grambsch}}, year = {2000}, publisher = {Springer}, address = {New York}, isbn = {0-387-98784-3}, }

[[62]] @Article{tidyverse2019, title = {Welcome to the {tidyverse}}, author = {Hadley Wickham and Mara Averick and Jennifer Bryan and Winston Chang and Lucy D'Agostino McGowan and Romain François and Garrett Grolemund and Alex Hayes and Lionel Henry and Jim Hester and Max Kuhn and Thomas Lin Pedersen and Evan Miller and Stephan Milton Bache and Kirill Müller and Jeroen Ooms and David Robinson and Dana Paige Seidel and Vitalie Spinu and Kohske Takahashi and Davis Vaughan and Claus Wilke and Kara Woo and Hiroaki Yutani}, year = {2019}, journal = {Journal of Open Source Software}, volume = {4}, number = {43}, pages = {1686}, doi = {10.21105/joss.01686}, }

[[63]] @Article{zoo2005, title = {zoo: S3 Infrastructure for Regular and Irregular Time Series}, author = {Achim Zeileis and Gabor Grothendieck}, journal = {Journal of Statistical Software}, year = {2005}, volume = {14}, number = {6}, pages = {1–27}, doi = {10.18637/jss.v014.i06}, }