Project 7: Difference-in-Differences and Synthetic Control

```
# Install and load packages
if (!require("pacman")) install.packages("pacman")
```

```
## Loading required package: pacman
```

```
devtools::install_github("ebenmichael/augsynth")
```

```
## Using GitHub PAT from the git credential store.

## Skipping install of 'augsynth' from a github remote, the SHA1 (0f4f1bcc) has not changed since last i
##   Use `force = TRUE` to force installation
```

```
pacman::p_load(# Tidyverse packages including dplyr and ggplot2
               tidyverse,
               ggthemes,
               augsynth,
               gsynth)


library(readr)
library(tidyverse)

# set seed
set.seed(44)

# load data
medicaid_expansion <- read.csv('/Users/marisatsai/Library/CloudStorage/Box-Box/Spring 2024/CSSTP/project
summary(medicaid_expansion)
```

```
##     State            Date_Adopted            year       uninsured_rate
##  Length:663         Length:663         Min.   :2008   Min.   :0.02495
##  Class :character   Class :character   1st Qu.:2011   1st Qu.:0.07702
##  Mode  :character   Mode  :character   Median :2014   Median :0.10475
##                                        Mean   :2014   Mean   :0.10978
##                                        3rd Qu.:2017   3rd Qu.:0.13888
##                                        Max.   :2020   Max.   :0.24082
##
##    population
##  Min.   :  584153
##  1st Qu.: 1850326
##  Median : 4531566
##  Mean   : 6364343
##  3rd Qu.: 7061530
##  Max.   :38802500
##  NA's   :13
```

```
data<-medicaid_expansion
```

Introduction

For this project, you will explore the question of whether the Affordable Care Act increased health insurance coverage (or conversely, decreased the number of people who are uninsured). The ACA was passed in March 2010, but several of its provisions were phased in over a few years. The ACA instituted the "individual mandate" which required that all Americans must carry health insurance, or else suffer a tax penalty. There are four mechanisms for how the ACA aims to reduce the uninsured population:

- Require companies with more than 50 employees to provide health insurance.
- Build state-run healthcare markets ("exchanges") for individuals to purchase health insurance.
- Provide subsidies to middle income individuals and families who do not qualify for employer based coverage.
- Expand Medicaid to require that states grant eligibility to all citizens and legal residents earning up to 138% of the federal poverty line. The federal government would initially pay 100% of the costs of this expansion, and over a period of 5 years the burden would shift so the federal government would pay 90% and the states would pay 10%.

In 2012, the Supreme Court heard the landmark case NFIB v. Sebelius, which principally challenged the constitutionality of the law under the theory that Congress could not institute an individual mandate. The Supreme Court ultimately upheld the individual mandate under Congress's taxation power, but struck down the requirement that states must expand Medicaid as impermissible subordination of the states to the federal government. Subsequently, several states refused to expand Medicaid when the program began on January 1, 2014. This refusal created the "Medicaid coverage gap" where there are indivudals who earn too much to qualify for Medicaid under the old standards, but too little to qualify for the ACA subsidies targeted at middle-income individuals.

States that refused to expand Medicaid principally cited the cost as the primary factor. Critics pointed out however, that the decision not to expand primarily broke down along partisan lines. In the years since the initial expansion, several states have opted into the program, either because of a change in the governing party, or because voters directly approved expansion via a ballot initiative.

You will explore the question of whether Medicaid expansion reduced the uninsured population in the U.S. in the 7 years since it went into effect. To address this question, you will use difference-in-differences estimation, and synthetic control.

Data

The dataset you will work with has been assembled from a few different sources about Medicaid. The key variables are:

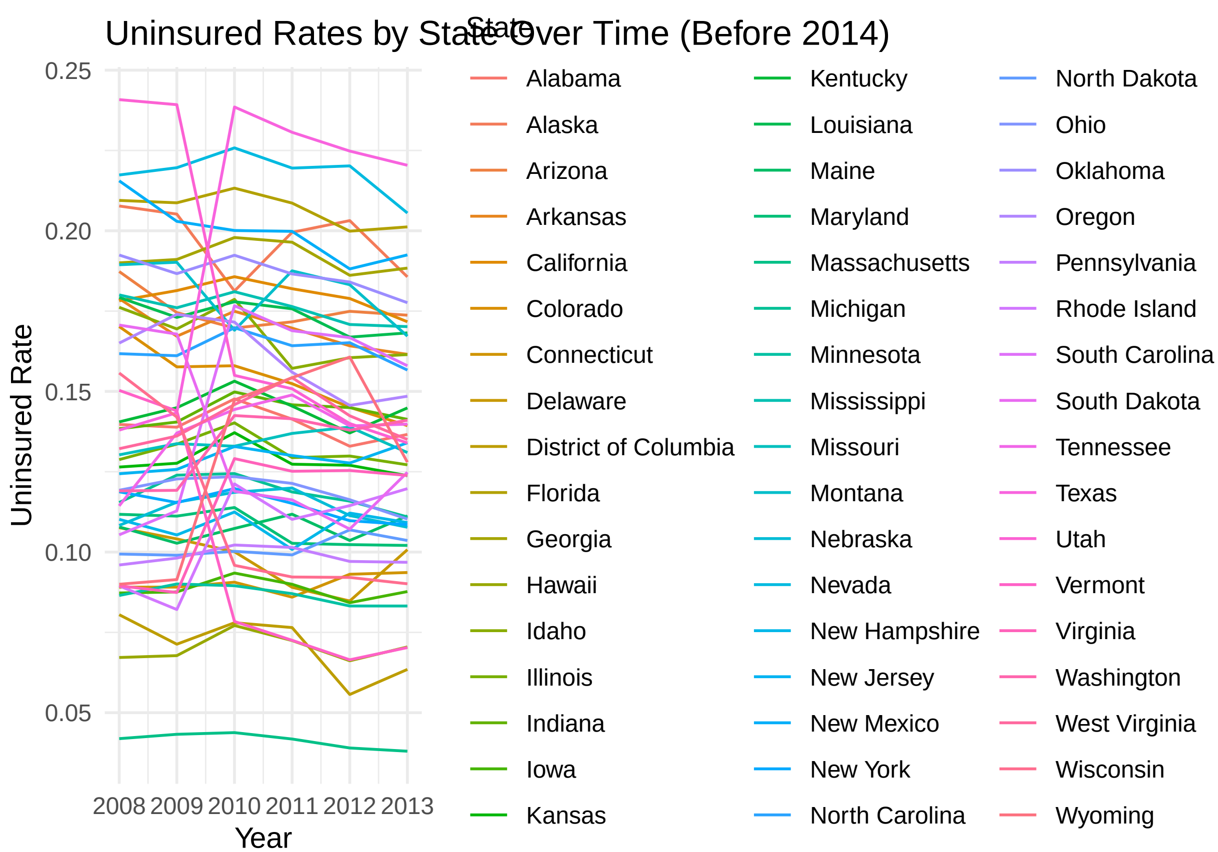
- **State**: Full name of state
- **Medicaid Expansion Adoption**: Date that the state adopted the Medicaid expansion, if it did so.
- **Year**: Year of observation.
- **Uninsured rate**: State uninsured rate in that year.

Exploratory Data Analysis

Create plots and provide 1-2 sentence analyses to answer the following questions:

- Which states had the highest uninsured rates prior to 2014? The lowest?
- Which states were home to most uninsured Americans prior to 2014? How about in the last year in the data set? **Note**: 2010 state population is provided as a variable to answer this question. In an actual study you would likely use population estimates over time, but to simplify you can assume these numbers stay about the same.

```
# highest and lowest uninsured rates


library(ggplot2)

# Filter data to include only years prior to 2014
filtered_data <- subset(data, year < 2014)

# Create the plot using ggplot2
ggplot(filtered_data, aes(x = year, y = uninsured_rate, group = State, color = State)) +
  geom_line() +
  labs(title = "Uninsured Rates by State Over Time (Before 2014)",
       x = "Year",
       y = "Uninsured Rate") +
  theme_minimal()
```




```
#getting list of states since it's hard to ID states by color

# Filter data to include only year 2013
```

```r
data_2013 <- subset(data, year == 2013)

# Calculate average uninsured rate for each state in 2013
avg_uninsured_rate_2013 <- aggregate(uninsured_rate ~ State, data_2013, mean)

# Sort states based on average uninsured rate in descending order
sorted_states <- avg_uninsured_rate_2013[order(-avg_uninsured_rate_2013$uninsured_rate), ]

# Print the sorted list of states
print(sorted_states)
```

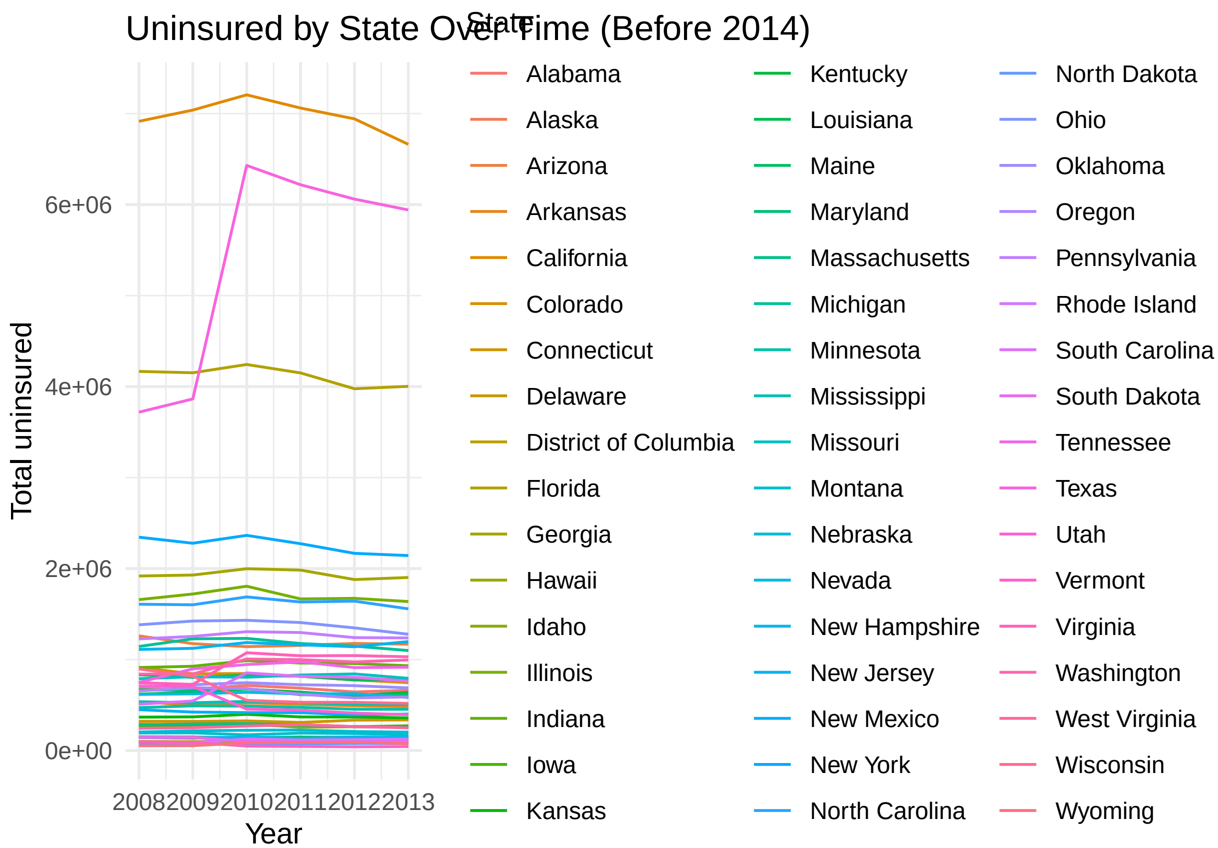
```
##                   State uninsured_rate
## 44                Texas        0.22039
## 29               Nevada        0.20553
## 10              Florida        0.20121
## 32           New Mexico        0.19249
## 11              Georgia        0.18839
## 2                Alaska        0.18566
## 37             Oklahoma        0.17762
## 3               Arizona        0.17373
## 5            California        0.17169
## 25          Mississippi        0.17017
## 19            Louisiana        0.16822
## 27              Montana        0.16721
## 4              Arkansas        0.16155
## 13                Idaho        0.16147
## 41       South Carolina        0.15798
## 34       North Carolina        0.15665
## 38               Oregon        0.14850
## 18             Kentucky        0.14486
## 15              Indiana        0.14131
## 48           Washington        0.14103
## 43            Tennessee        0.13981
## 6              Colorado        0.13935
## 1               Alabama        0.13661
## 49        West Virginia        0.13504
## 31           New Jersey        0.13400
## 45                 Utah        0.13396
## 26             Missouri        0.13097
## 51              Wyoming        0.12796
## 14             Illinois        0.12716
## 42         South Dakota        0.12483
## 47             Virginia        0.12384
## 17               Kansas        0.12378
## 40         Rhode Island        0.11973
## 20                Maine        0.11125
## 23             Michigan        0.11093
## 36                 Ohio        0.11026
## 30        New Hampshire        0.10918
## 33             New York        0.10854
## 28             Nebraska        0.10772
## 35         North Dakota        0.10360
## 21             Maryland        0.10206
```

```
## 8              Delaware        0.10074
## 39         Pennsylvania        0.09681
## 7           Connecticut        0.09363
## 50            Wisconsin        0.09012
## 16                 Iowa        0.08770
## 24            Minnesota        0.08321
## 12               Hawaii        0.07047
## 46              Vermont        0.07027
## 9  District of Columbia        0.06347
## 22        Massachusetts        0.03801
```

Prior to 2014, DC and Massachusetts had the lowest uninsured rates; Texas, Nevada, and Florida had the highest.

```
# most uninsured Americans

library(ggplot2)

# Filter data to include only years prior to 2014
filtered_data <- subset(data, year < 2014)

# Calculate number of uninsured people for each state
filtered_data$total_uninsured <- filtered_data$uninsured_rate * filtered_data$population

ggplot(filtered_data, aes(x = year, y = total_uninsured, group = State, color = State)) +
  geom_line() +
  labs(title = "Uninsured by State Over Time (Before 2014)",
       x = "Year",
       y = "Total uninsured") +
  theme_minimal()
```

```
## Warning: Removed 6 rows containing missing values or values outside the scale range
## (‘geom_line()‘).
```




```r
#getting list of states since it's hard to ID states by color

# Filter data to include only year 2013
data_2013 <- subset(data, year == 2013)

# Calculate number of uninsured people for each state in 2013
data_2013$total_uninsured <- data_2013$uninsured_rate * data_2013$population

# Aggregate total uninsured by state
total_uninsured_by_state <- aggregate(total_uninsured ~ State, data_2013, sum)

# Sort states based on total uninsured people in descending order
sorted_states_uninsured <- total_uninsured_by_state[order(-total_uninsured_by_state$total_uninsured), ]

# Print the sorted list of states
print(sorted_states_uninsured)
```

```
##             State total_uninsured
## 5      California      6662001.23
## 43          Texas      5941043.97
## 9         Florida      4002730.29
## 32       New York      2143255.48
## 10        Georgia      1902238.45
## 13       Illinois      1637894.55
## 33 North Carolina      1557721.96
```

```
## 35           Ohio      1278372.41
## 38   Pennsylvania      1237929.70
## 30     New Jersey      1197715.45
## 3         Arizona      1169460.72
## 22       Michigan      1099302.66
## 46       Virginia      1031127.63
## 47     Washington       995887.58
## 14        Indiana       932201.58
## 42      Tennessee       915664.90
## 25       Missouri       794148.25
## 18      Louisiana       782168.50
## 40 South Carolina       763435.51
## 6        Colorado       746338.53
## 36       Oklahoma       688819.42
## 1         Alabama       662473.39
## 17       Kentucky       639333.38
## 20       Maryland       609952.10
## 37         Oregon       589580.49
## 28         Nevada       583520.02
## 49      Wisconsin       518871.67
## 24    Mississippi       507944.86
## 4        Arkansas       483693.46
## 23      Minnesota       454091.37
## 31     New Mexico       401451.75
## 44           Utah       394231.15
## 16         Kansas       359459.72
## 7     Connecticut       336756.87
## 15           Iowa       272494.95
## 12          Idaho       263916.90
## 21  Massachusetts       256392.96
## 48  West Virginia       249868.02
## 27       Nebraska       202675.50
## 26        Montana       171152.64
## 19          Maine       147972.40
## 29  New Hampshire       144861.44
## 2          Alaska       136967.32
## 39   Rhode Island       126335.86
## 41   South Dakota       106501.84
## 11         Hawaii       100036.46
## 8        Delaware        94253.75
## 34   North Dakota        76610.34
## 50        Wyoming        74748.22
## 45        Vermont        43989.79
```

Prior to 2014, Vermont had the lowest number of uninsured; California had the highest.

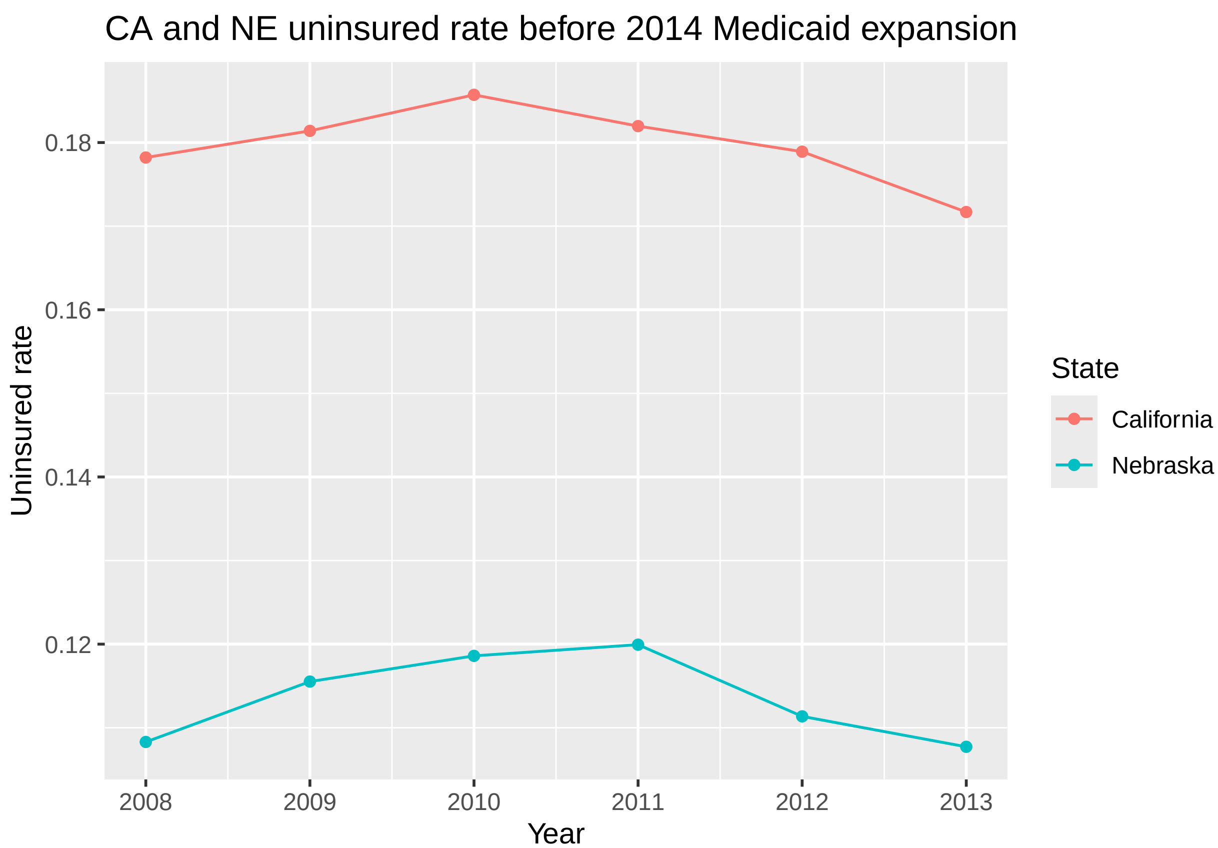
Difference-in-Differences Estimation

Estimate Model

Do the following:

- Choose a state that adopted the Medicaid expansion on January 1, 2014 and a state that did not. **Hint**: Do not pick Massachusetts as it passed a universal healthcare law in 2006, and also avoid picking a state that adopted the Medicaid expansion between 2014 and 2015.
- Assess the parallel trends assumption for your choices using a plot. If you are not satisfied that the assumption has been met, pick another state and try again (but detail the states you tried).

```
# Parallel Trends plot
library(dplyr)
# visualize intervention
# ----------
data %>%
  # processing
  # ----------
  filter(State %in% c("California","Nebraska")) %>%  # use "%in% to filter values in a vector
  filter(year >= 2007 & year< 2014) %>%
  #filter(between(year_qtr, 2012.5, 2012.75)) %>%   # same filtering but using between() instead which

  # plot
  # ----------
  ggplot() +
  # add in point layer
  geom_point(aes(x = year,
                 y = uninsured_rate,
                 color = State)) + # color by state
  # add in line
  geom_line(aes(x = year,
                y = uninsured_rate,
                color = State)) +

  # themes
 # theme_fivethirtyeight() +
  #theme(axis.title = element_text()) +

  # labels - PREFER TO USE labs() SO THAT IT IS ALL IN ONE ARGUMENT
  ggtitle('CA and NE uninsured rate before 2014 Medicaid expansion') +
  xlab('Year') +
  ylab('Uninsured rate')
```



I tried California and Utah, but they were extremely different - Utah experienced a big drop in 2010. I then tried NE, which was more similar to each other, in that both increased in 2010/2011 and decreased afterwards.

- Estimate a difference-in-differences estimate of the effect of the Medicaid expansion on the uninsured share of the population. You may follow the lab example where we estimate the differences in one pre-treatment and one post-treatment period, or take an average of the pre-treatment and post-treatment outcomes

```
# Difference-in-Differences estimation

#
# DiD for: CA-NE
# -----------------------------------------------------------------
# create a dataset for CA and NE
kc <-
  data %>%
  filter(State %in% c("California","Nebraska")) %>%
  filter(year >= 2013 & year<= 2014)

# pre-treatment difference
# ----------

library(tidyr)
pre_diff <-
  kc %>%
```

```r
  # filter out only the quarter we want
  filter(year == 2013) %>%
  # subset to select only vars we want
  select(State,
         uninsured_rate) %>%
  # make the data wide
  pivot_wider(names_from = State,
              values_from = uninsured_rate) %>%
  #subtract to make calculation
  summarise(diff = Nebraska - California)

# post-treatment difference
# ----------
post_diff <-
  kc %>%
  # filter out only the quarter we want
  filter(year == 2014) %>%
  # subset to select only vars we want
  select(State,
         uninsured_rate) %>%
  # make the data wide
  pivot_wider(names_from = State,
              values_from = uninsured_rate) %>%
  # subtract to make calculation
  summarise(Nebraska - California)

# diff-in-diffs
# ----------
diff_in_diffs <- post_diff - pre_diff
diff_in_diffs
```

```
##   Nebraska - California
## 1               0.03257
```

Discussion Questions

- Card/Krueger's original piece utilized the fact that towns on either side of the Delaware river are likely to be quite similar to one another in terms of demographics, economics, etc. Why is that intuition harder to replicate with this data?
- **Answer**: It is harder to replicate because our data represents state, which has larger variation in geography, demographic make-up, and surrounding policies than towns across a river.
- What are the strengths and weaknesses of using the parallel trends assumption in difference-in-differences estimates?
- **Answer**: Having parallel trends makes it more conceiveable that a deviation from the trend at the timepoint of an intervention might be due to the intervention; it suggests that factors (outside of the intervention) affecting trends between the two groups might be consistent. A weakness is that people may overinterpret differences since there could still be factors that cause deviations from trends that are not due to an interbentions.

Synthetic Control

Estimate Synthetic Control

Although several states did not expand Medicaid on January 1, 2014, many did later on. In some cases, a Democratic governor was elected and pushed for a state budget that included the Medicaid expansion, whereas in others voters approved expansion via a ballot initiative. The 2018 election was a watershed moment where several Republican-leaning states elected Democratic governors and approved Medicaid expansion. In cases with a ballot initiative, the state legislature and governor still must implement the results via legislation. For instance, Idaho voters approved a Medicaid expansion in the 2018 election, but it was not implemented in the state budget until late 2019, with enrollment beginning in 2020.

Do the following:

- Choose a state that adopted the Medicaid expansion after January 1, 2014. Construct a non-augmented synthetic control and plot the results (both pre-treatment fit and post-treatment differences). Also report the average ATT and L2 imbalance.

```
# create a treatment indicator
# ----------
data <-
  data %>%
  # select subset of variables
  select(year, State, Date_Adopted, population, uninsured_rate) %>%
  # create new treatment flag just to see
  mutate(treatment = case_when(State == "Nebraska" & year >= 2013 ~ 1,
                               TRUE ~ 0))
# view head
head(data)
```

```
##   year      State Date_Adopted population uninsured_rate treatment
## 1 2008    Alabama         <NA>    4849377       0.139716         0
## 2 2008     Alaska   2015-09-01     737732       0.207716         0
## 3 2008    Arizona   2014-01-01    6731484       0.187312         0
## 4 2008   Arkansas   2014-01-01    2994079       0.178883         0
## 5 2008 California   2014-01-01   38802500       0.178212         0
## 6 2008   Colorado   2014-01-01    5355856       0.170183         0
```

```
# install libraries - install "augsynth" here since it is now on CRAN
pacman::p_load(# Tidyverse packages including dplyr and ggplot2
               tidyverse,
               ggthemes,
               augsynth)
#install.packages('BiocManager')
#library(augsynth)
#install.packages("augsynth")
#install.packages("devtools")
#devtools::install_github("ebenmichael/augsynth")


# non-augmented synthetic control

  syn <-                                # save object
  augsynth(uninsured_rate ~ treatment, # treatment
```

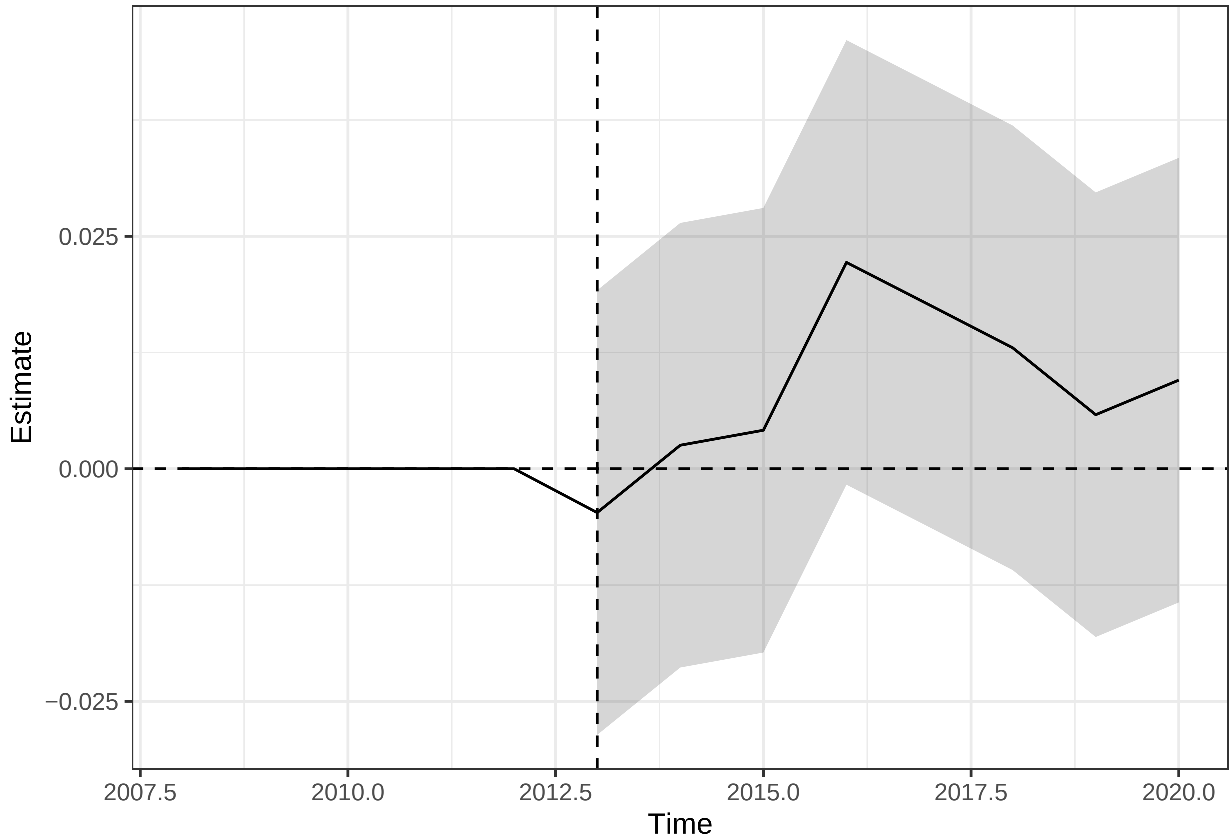
```r
                         State,     # unit
                         year,  # time
                         data,    # data
           progfunc = "None",       # plain syn control
           scm = T)                 # synthetic control
```

```
## One outcome and one treatment time found. Running single_augsynth.
```

```r
# summary
summary(syn)
```

```
##
## Call:
## single_augsynth(form = form, unit = !!enquo(unit), time = !!enquo(time),
##     t_int = t_int, data = data, progfunc = "None", scm = ..2)
##
## Average ATT Estimate (p Value for Joint Null):  0.00877   ( 0.24 )
## L2 Imbalance: 0.000
## Percent improvement from uniform weights: 100%
##
## Avg Estimated Bias: NA
##
## Inference type: Conformal inference
##
##  Time Estimate 95% CI Lower Bound 95% CI Upper Bound p Value
##  2013   -0.005             -0.029              0.019   0.154
##  2014    0.003             -0.021              0.026   1.000
##  2015    0.004             -0.020              0.028   1.000
##  2016    0.022             -0.002              0.046   0.175
##  2017    0.018             -0.006              0.042   0.505
##  2018    0.013             -0.011              0.037   0.485
##  2019    0.006             -0.018              0.030   1.000
##  2020    0.010             -0.014              0.033   0.156
```

```r
# plot
plot(syn)
```




```
- Re-run the same analysis but this time use an augmentation (default choices are Ridge, Matrix Complet
```

```
```r
# augmented synthetic control
ridge_syn <-
augsynth(uninsured_rate ~ treatment,
State,
year,
data,
progfunc = "ridge",
scm = T)
```

```
## One outcome and one treatment time found. Running single_augsynth.
```

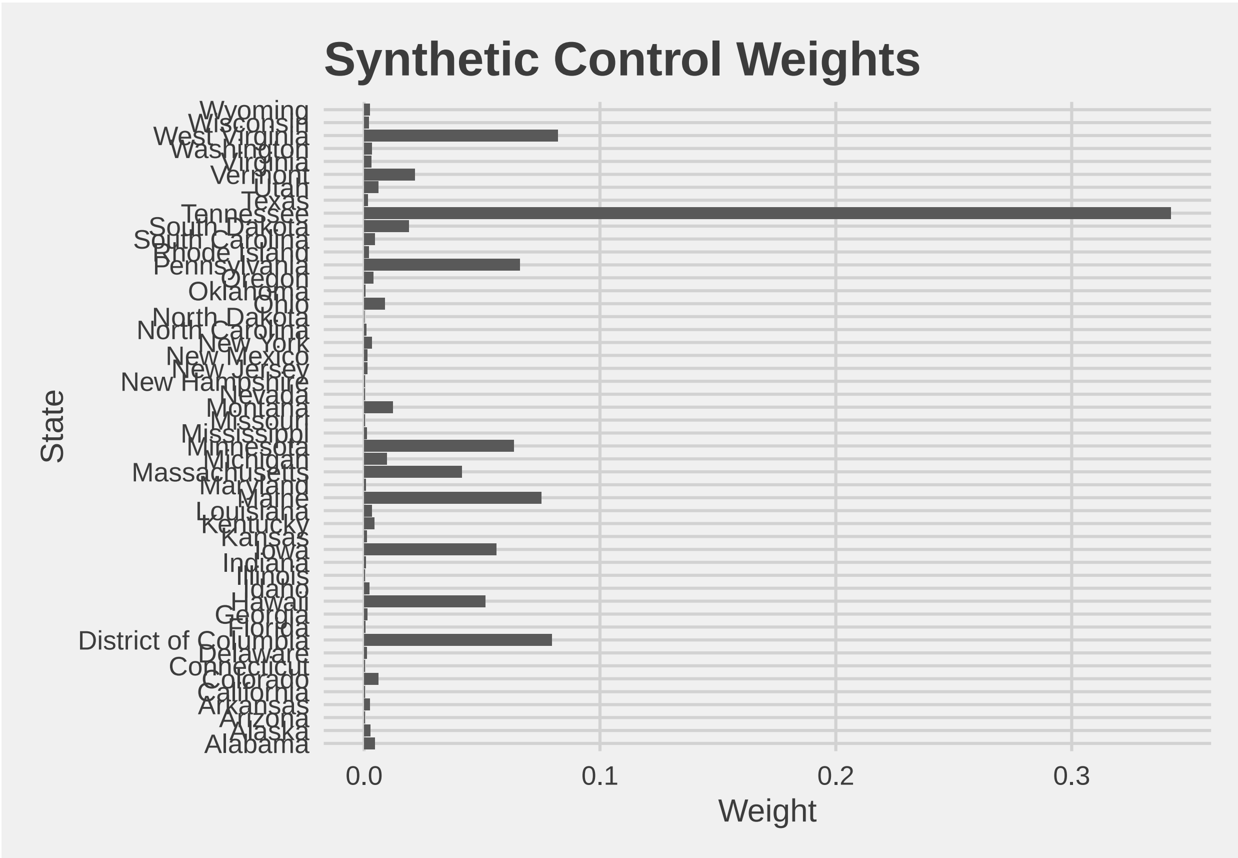
```r
# summary
summary(ridge_syn)
```

```
##
## Call:
## single_augsynth(form = form, unit = !!enquo(unit), time = !!enquo(time),
##     t_int = t_int, data = data, progfunc = "ridge", scm = ..2)
##
## Average ATT Estimate (p Value for Joint Null):  0.00877   ( 0.23 )
```
```

```
## L2 Imbalance: 0.000
## Percent improvement from uniform weights: 100%
##
## Avg Estimated Bias: 0.000
##
## Inference type: Conformal inference
##
##  Time Estimate 95% CI Lower Bound 95% CI Upper Bound p Value
##  2013   -0.005             -0.029              0.019   0.170
##  2014    0.003             -0.021              0.026   1.000
##  2015    0.004             -0.020              0.028   1.000
##  2016    0.022             -0.002              0.046   0.159
##  2017    0.018             -0.006              0.042   0.491
##  2018    0.013             -0.011              0.037   0.535
##  2019    0.006             -0.018              0.030   1.000
##  2020    0.010             -0.014              0.033   0.176
```

- Plot barplots to visualize the weights of the donors.

```r
# barplots of weights
data.frame(ridge_syn$weights) %>%
  tibble::rownames_to_column('State') %>%
  ggplot() +
  geom_bar(aes(x = State, y = ridge_syn.weights),
           stat = 'identity') +
  coord_flip() + # coord flip
  theme_fivethirtyeight() +
  theme(axis.title = element_text()) +
  ggtitle('Synthetic Control Weights') +
  xlab('State') +
  ylab('Weight')
```



HINT: Is there any preprocessing you need to do before you allow the program to automatically find weights for donor states?

Discussion Questions

- What are the advantages and disadvantages of synthetic control compared to difference-in-differences estimators?
- **Answer**:
- One of the benefits of synthetic control is that the weights are bounded between [0,1] and the weights must sum to 1. Augmentation might relax this assumption by allowing for negative weights. Does this create an interpretation problem, and how should we balance this consideration against the improvements augmentation offers in terms of imbalance in the pre-treatment period?
- **Answer**: I don't think it creates an interpretation issue– in weighting, it seems natural to have both up- and down-weighted influences. I don't see a problem if it improves balance issues.

Staggered Adoption Synthetic Control

Estimate Multisynth

Do the following:

- Estimate a multisynth model that treats each state individually. Choose a fraction of states that you can fit on a plot and examine their treatment effects.

```
# multisynth model states
summary(data)
```

```
##       year          State            Date_Adopted         population
##  Min.   :2008   Length:663         Length:663         Min.   :  584153
##  1st Qu.:2011   Class :character   Class :character   1st Qu.: 1850326
##  Median :2014   Mode  :character   Mode  :character   Median : 4531566
##  Mean   :2014                                         Mean   : 6364343
##  3rd Qu.:2017                                         3rd Qu.: 7061530
##  Max.   :2020                                         Max.   :38802500
##                                                       NA's   :13
##  uninsured_rate      treatment
##  Min.   :0.02495   Min.   :0.00000
##  1st Qu.:0.07702   1st Qu.:0.00000
##  Median :0.10475   Median :0.00000
##  Mean   :0.10978   Mean   :0.01207
##  3rd Qu.:0.13888   3rd Qu.:0.00000
##  Max.   :0.24082   Max.   :1.00000
##
```

```
# create dataset
# ---------
data_clean <-
  data %>%
#create "treatment" - year Medicaid expansion was adopted
    mutate(Date_Adopted = ifelse(is.na(Date_Adopted),
                                 Inf, Date_Adopted),
           DA = 1 * (year >= Date_Adopted))
```

```
# setting nu to 0.5
# ---------
ppool_syn <- multisynth(uninsured_rate ~ DA,
                        State,                       # unit
                        year,                        # time
                        nu = 0,                    # varying degree of pooling
                        data_clean, # data
                        n_leads = 3)                  # post-treatment periods to estimate


# view results
print(ppool_syn$nu)
```

```
## [1] 0
```

```
ppool_syn
```

```
##
## Call:
## multisynth(form = uninsured_rate ~ DA, unit = State, time = year,
```

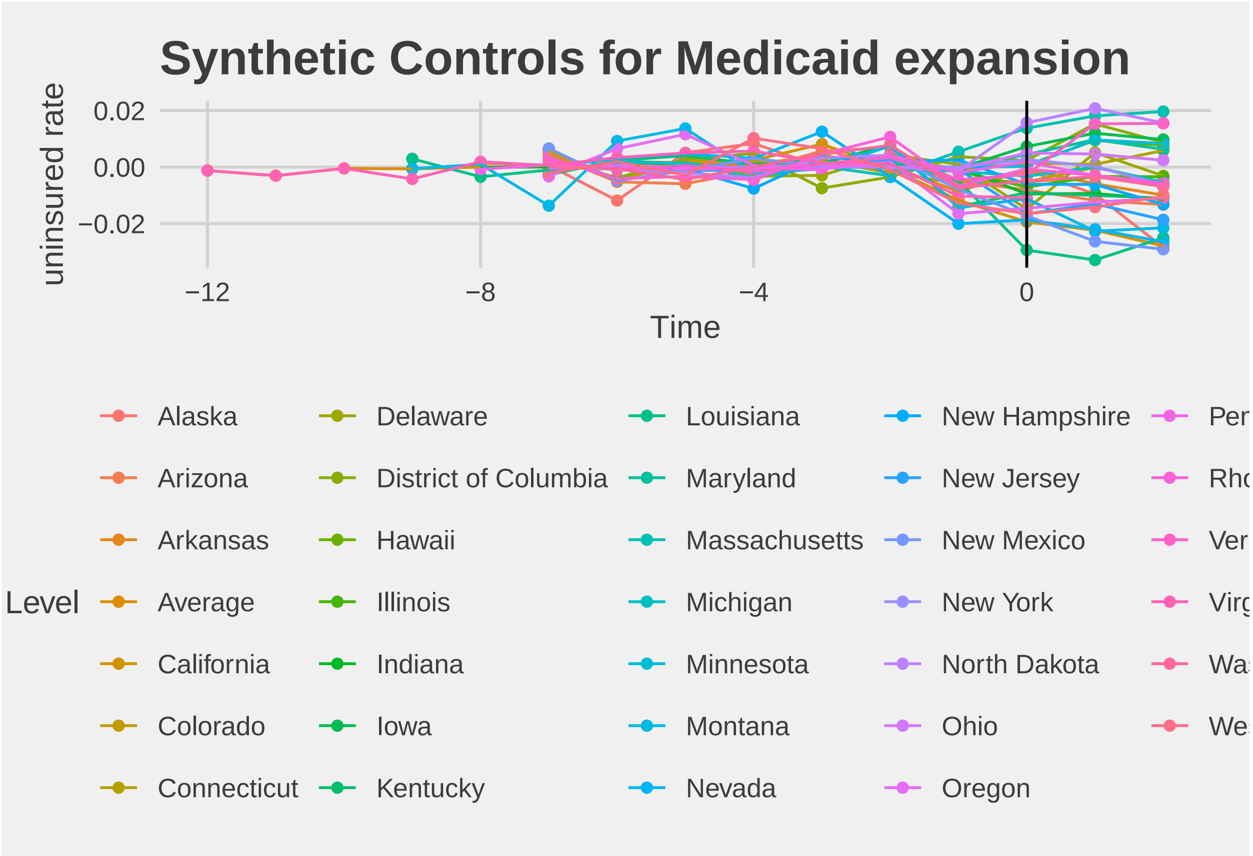
```
##     data = data_clean, n_leads = 3, nu = 0)
##
## Average ATT Estimate: -0.005
```

```r
# save ATT and balance stats
# ---------
ppool_syn_summ <- summary(ppool_syn)
```

```r
# plot actual estimates not values of synthetic controls
# ---------
ppool_syn_summ$att %>%
  ggplot(aes(x = Time, y = Estimate, color = Level)) +
  geom_point() +
  geom_line() +
  geom_vline(xintercept = 0) +
  theme_fivethirtyeight() +
  theme(axis.title = element_text(),
        legend.position = "bottom") +
  ggtitle('Synthetic Controls for Medicaid expansion') +
  xlab('Time') +
  ylab('uninsured rate')
```

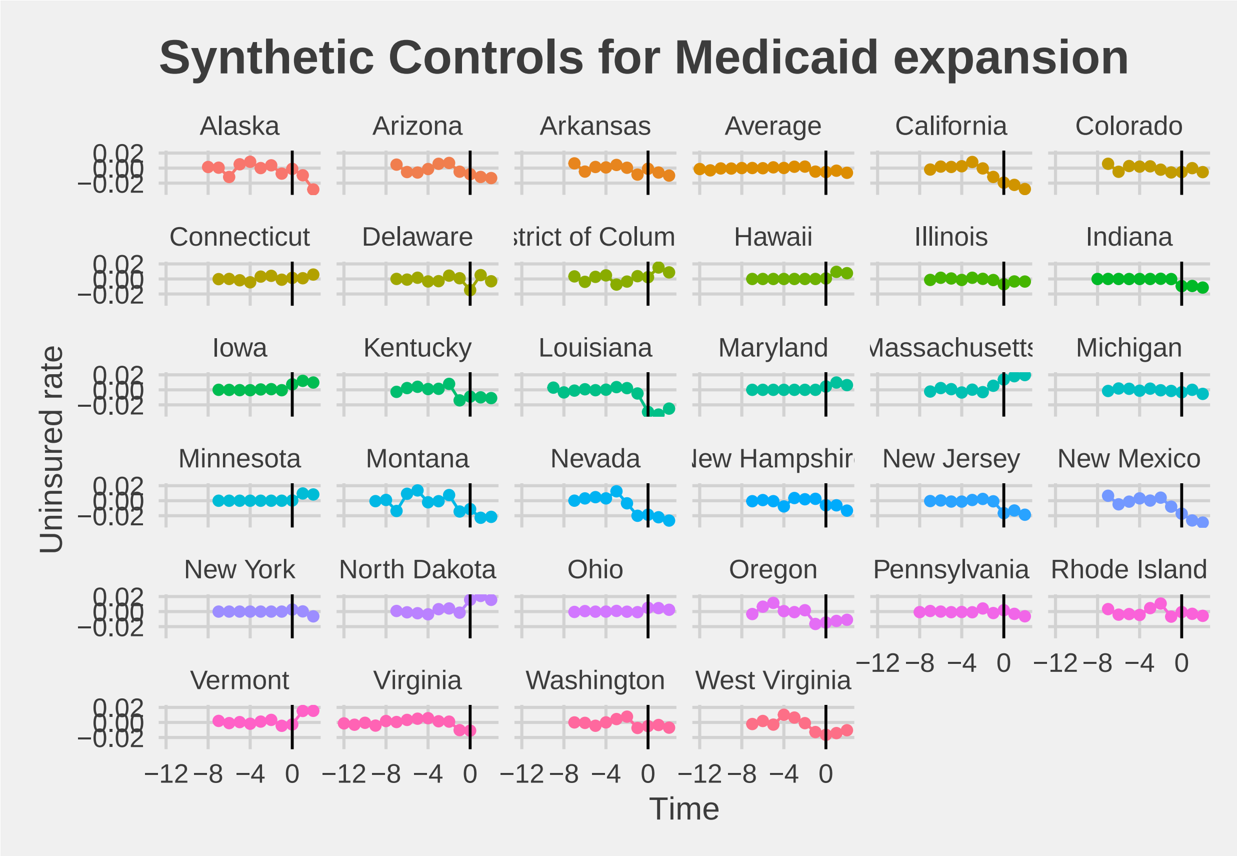
```
## Warning: Removed 189 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

```
## Warning: Removed 189 rows containing missing values or values outside the scale range
## (`geom_line()`).
```




```
# plot actual estimates not values of synthetic controls
# ---------

ppool_syn_summ$att %>%
  ggplot(aes(x = Time, y = Estimate, color = Level)) +
  geom_point() +
  geom_line() +
  geom_vline(xintercept = 0) +
  theme_fivethirtyeight() +
  theme(axis.title = element_text(),
        legend.position = 'None') +
  ggtitle('Synthetic Controls for Medicaid expansion') +
  xlab('Time') +
  ylab('Uninsured rate') +
  facet_wrap(~Level)    # facet-wrap by level (state in this case) for clearer presentation
```

```
## Warning: Removed 189 rows containing missing values or values outside the scale range
## (`geom_point()`).

## Warning: Removed 189 rows containing missing values or values outside the scale range
## (`geom_line()`).
```



- Estimate a multisynth model using time cohorts. For the purpose of this exercise, you can simplify the treatment time so that states that adopted Medicaid expansion within the same year (i.e. all states that adopted epxansion in 2016) count for the same cohort. Plot the treatment effects for these time cohorts.

```
# multisynth model time cohorts

summary(data)
```

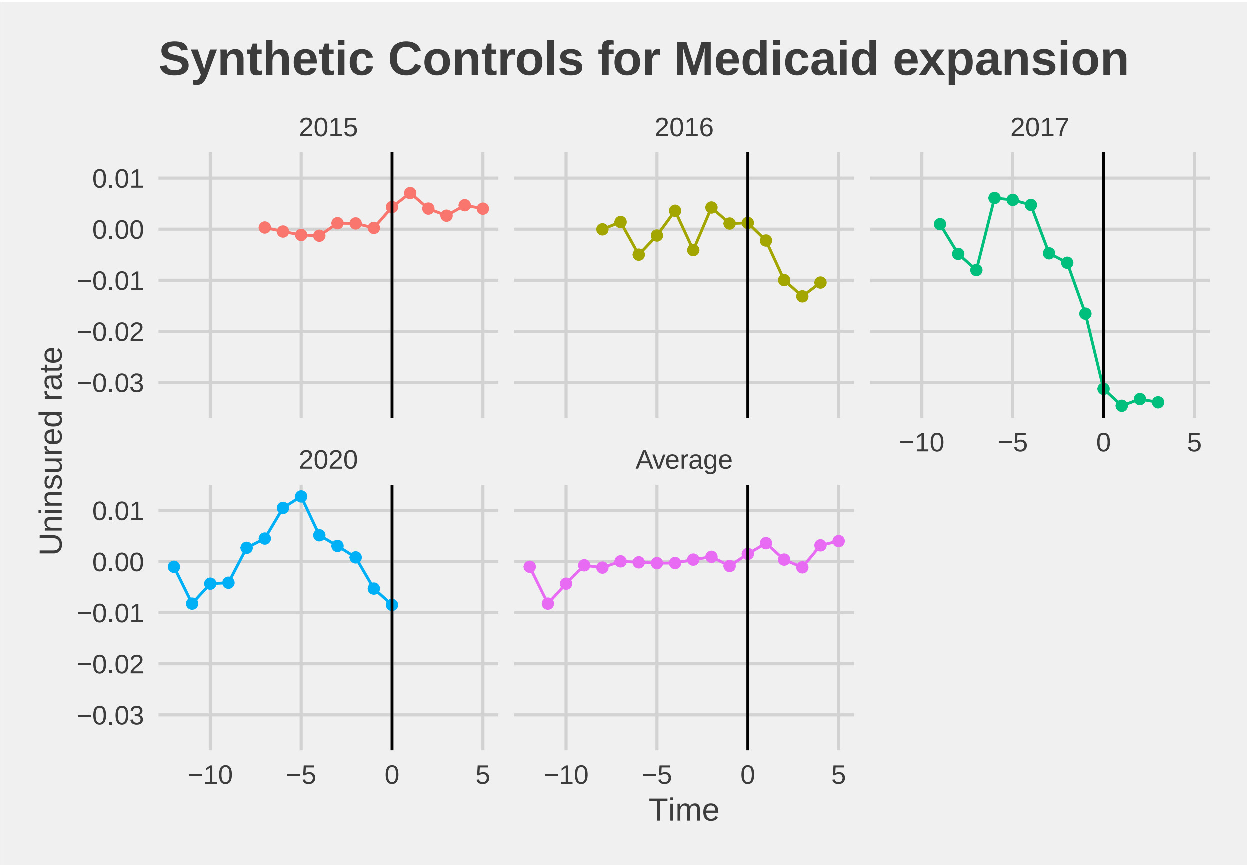
```
##       year          State            Date_Adopted         population      
##  Min.   :2008   Length:663         Length:663         Min.   :  584153  
##  1st Qu.:2011   Class :character   Class :character   1st Qu.: 1850326  
##  Median :2014   Mode  :character   Mode  :character   Median : 4531566  
##  Mean   :2014                                         Mean   : 6364343  
##  3rd Qu.:2017                                         3rd Qu.: 7061530  
##  Max.   :2020                                         Max.   :38802500  
##                                                       NA's   :13        
##  uninsured_rate       treatment      
##  Min.   :0.02495   Min.   :0.00000  
##  1st Qu.:0.07702   1st Qu.:0.00000  
##  Median :0.10475   Median :0.00000  
##  Mean   :0.10978   Mean   :0.01207  
##  3rd Qu.:0.13888   3rd Qu.:0.00000  
##  Max.   :0.24082   Max.   :1.00000  
## 
```

```r
# break observations into time cohorts
# -----------------------------------------------------------------
ppool_syn_time <- multisynth(uninsured_rate ~ DA,
                             State,
                             year,
                             data_clean,
                             n_leads = 6,
                             time_cohort = TRUE)           # time cohort set to TRUE

# save summary
ppool_syn_time_summ <- summary(ppool_syn_time)

# view
ppool_syn_time_summ
```

```
##
## Call:
## multisynth(form = uninsured_rate ~ DA, unit = State, time = year,
##     data = data_clean, n_leads = 6, time_cohort = TRUE)
##
## Average ATT Estimate (Std. Error): 0.001  (0.016)
##
## Global L2 Imbalance: 0.003
## Scaled Global L2 Imbalance: 0.022
## Percent improvement from uniform global weights: 97.8
##
## Individual L2 Imbalance: 0.016
## Scaled Individual L2 Imbalance: 0.038
## Percent improvement from uniform individual weights: 96.2
##
##  Time Since Treatment   Level      Estimate  Std.Error lower_bound upper_bound
##                     0 Average  0.0015026595 0.01586670 -0.02914111  0.02017969
##                     1 Average  0.0036030002 0.01671184 -0.02906730  0.02302021
##                     2 Average  0.0003869874 0.01679934 -0.03257295  0.02016377
##                     3 Average -0.0011230183 0.01607688 -0.03261802  0.01793574
##                     4 Average  0.0031793378 0.01894188 -0.03307690  0.02374738
##                     5 Average  0.0040117832 0.01964203 -0.03404167  0.02609780
```

```r
# plot effect for each time period (local treatment effects)
# ---------
ppool_syn_time_summ$att %>%
  ggplot(aes(x = Time, y = Estimate, color = Level)) +
  geom_point() +
  geom_line() +
  geom_vline(xintercept = 0) +
  theme_fivethirtyeight() +
  theme(axis.title = element_text(),
        legend.position = 'None') +
  ggtitle('Synthetic Controls for Medicaid expansion') +
  xlab('Time') +
  ylab('Uninsured rate') +
  facet_wrap(~Level)
```

```
## Warning: Removed 25 rows containing missing values or values outside the scale range
## (`geom_point()`).

## Warning: Removed 25 rows containing missing values or values outside the scale range
## (`geom_line()`).
```



Discussion Questions

- One feature of Medicaid is that it is jointly administered by the federal government and the states, and states have some flexibility in how they implement Medicaid. For example, during the Trump administration, several states applied for waivers where they could add work requirements to the eligibility standards (i.e. an individual needed to work for 80 hours/month to qualify for Medicaid). Given these differences, do you see evidence for the idea that different states had different treatment effect sizes?
- **Answer**: Yes, the trend does not look entirely consistent across states.
- Do you see evidence for the idea that early adopters of Medicaid expansion enjoyed a larger decrease in the uninsured population?
- **Answer**: No, that wasn't clear in the plots.

General Discussion Questions

- Why are DiD and synthetic control estimates well suited to studies of aggregated units like cities, states, countries, etc?
- **Answer**: DiD is often used to analyze policy-level changes which occur in aggregated unit. Policy-level changes are often focused on changing some aggregate level outcome, over time so having an method that can incorporate the unit-level and longituinal nature is helpful. Aggregated units may not have a siple
- What role does selection into treatment play in DiD/synthetic control versus regression discontinuity? When would we want to use either method?
- **Answer**: In DiD, selection into treatment is on an aggregate scale; in regression discontinuity it is individual, based on a cutoff of some metric.