

Project 7: Difference-in-Differences and Synthetic Control

Saron Goitom

04/26/2024

```
## Loading required package: pacman

## Using GitHub PAT from the git credential store.

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

## Rows: 663 Columns: 5
## -- Column specification -----
## Delimiter: ","
## chr  (1): State
## dbl  (3): year, uninsured_rate, population
## date (1): Date_Adopted
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

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 individuals 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
check <- medicaid_expansion %>%
  filter(year<2014) %>%
  group_by(State) %>%
  summarise(minIns = min(uninsured_rate),
            maxIns = max(uninsured_rate)) %>%
  arrange(maxIns, minIns)

check
```

```
## # A tibble: 51 x 3
##   State                minIns maxIns
##   <chr>                <dbl> <dbl>
## 1 Massachusetts      0.0380 0.0438
## 2 Hawaii              0.0662 0.0771
## 3 District of Columbia 0.0557 0.0805
## 4 Minnesota           0.0832 0.0901
## 5 Iowa                0.0842 0.0935
## 6 Connecticut         0.0860 0.0936
```

```
## 7 Pennsylvania      0.0960 0.102
## 8 North Dakota      0.0990 0.107
## 9 Delaware          0.0848 0.108
## 10 Maine            0.103  0.112
## # i 41 more rows
```

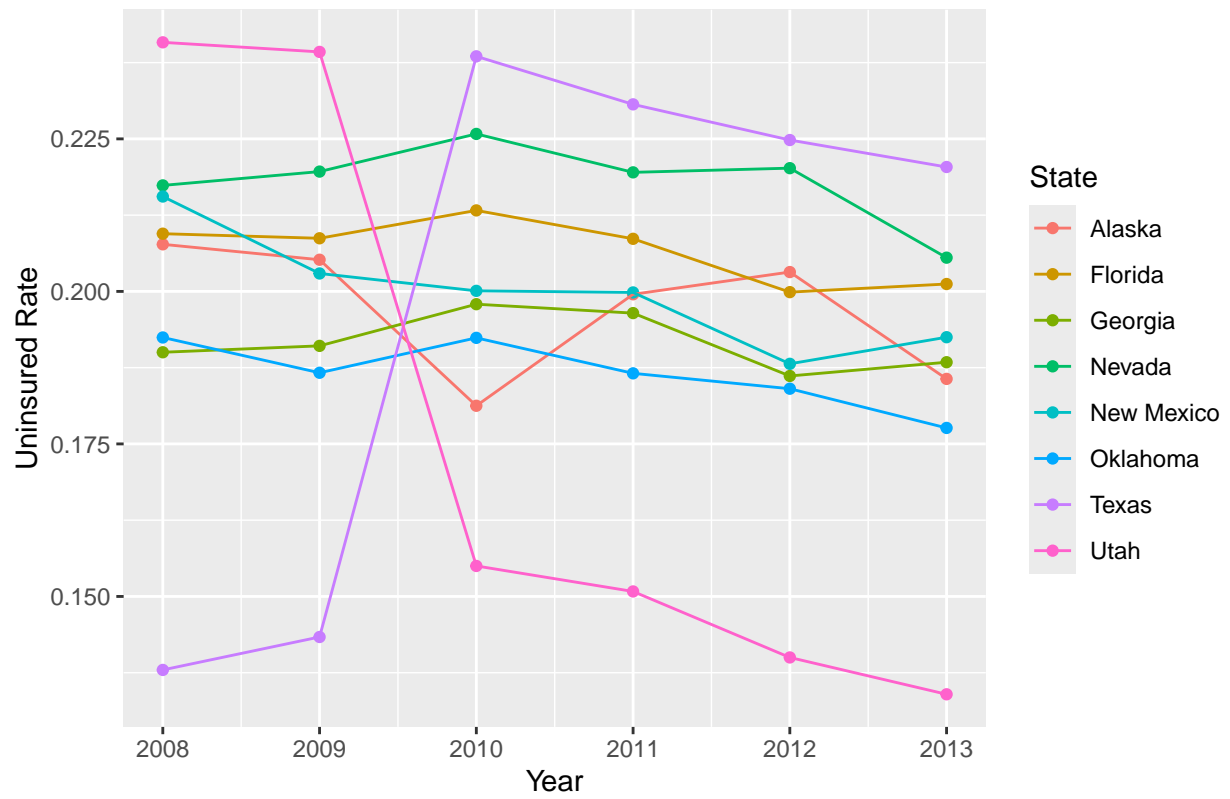
Prior to 2014, Utah, Texas, Nevada, New Mexico, Florida, Alaska, Georgia, Oklahoma are the 8 states with the highest uninsured rates. States with lowest uninsured rates pre 2014 are Massachusetts, Hawaii, DC, Minnesota, Iowa, Connecticut, Pennsylvania, and North Dakota

```
highplot <- medicaid_expansion %>%
  filter(year<2014 & State %in% c("Utah","Texas", "Nevada", "New Mexico", "Florida", "Alaska", "G
  ggplot() +
    geom_point(aes(x = year,
    y = uninsured_rate,
    color = State)) +
    geom_line(aes(x = year,
    y = uninsured_rate,
    color = State)) +
    theme(axis.title = element_text()) +
    ggtitle('States With Highest Unsured Rates (pre 2014)') +
    xlab('Year') +
    ylab('Uninsured Rate')

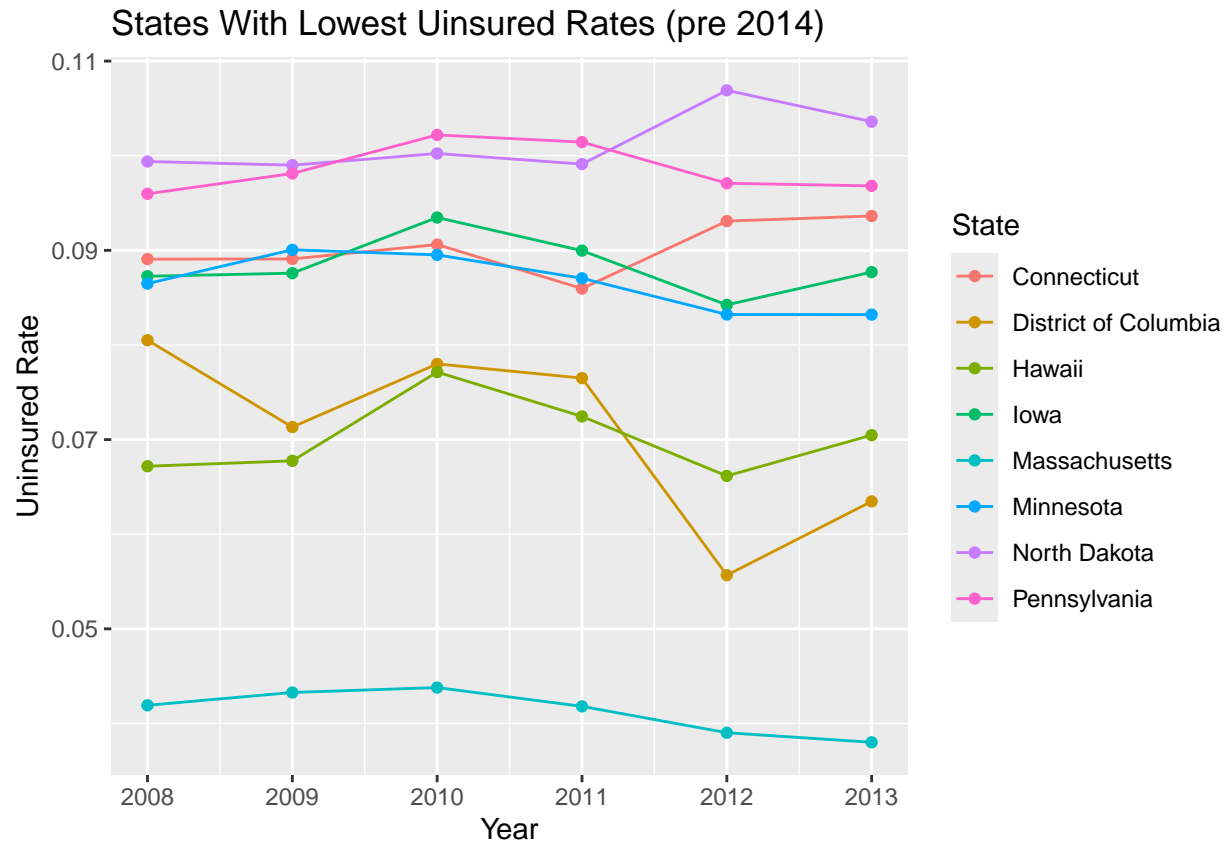
lowplot <- medicaid_expansion %>%
  filter(year<2014 & State %in% c("Massachusetts","Hawaii", "District of Columbia", "Minnesota",
  ggplot() +
    geom_point(aes(x = year,
    y = uninsured_rate,
    color = State)) +
    geom_line(aes(x = year,
    y = uninsured_rate,
    color = State)) +
    theme(axis.title = element_text()) +
    ggtitle('States With Lowest Unsured Rates (pre 2014)') +
    xlab('Year') +
    ylab('Uninsured Rate')

highplot
```

States With Highest Uninsured Rates (pre 2014)



lowplot



```
# states with most uninsured Americans (create new variable to calculate # uninsured)
uninsured <- medicaid_expansion %>%
  filter(year<2014) %>%
  mutate(num_unIns = population*uninsured_rate) %>%
  group_by(State) %>%
    summarise(minNum = min(num_unIns),
              maxNum = max(num_unIns)) %>%
  arrange(maxNum, minNum)
uninsured
```

```
## # A tibble: 51 x 3
##   State      minNum  maxNum
##   <chr>      <dbl>   <dbl>
## 1 North Dakota 73210.  79058.
## 2 Wyoming     52562.  93868.
## 3 Vermont     41611.  94105.
## 4 Delaware     79312. 100659.
## 5 Hawaii       93918. 109501.
## 6 Rhode Island 86640. 127901.
## 7 South Dakota 91426. 145599.
## 8 Maine       136673. 148675.
## 9 New Hampshire 133767. 149220.
## 10 Alaska      133716. 153239.
## # i 41 more rows
```

```

uninsured20 <- medicaid_expansion %>%
  filter(year==2020) %>%
  mutate(num_unIns = population*uninsured_rate) %>%
  group_by(State) %>%
    summarise(minNum = min(num_unIns),
              maxNum = max(num_unIns)) %>%
  arrange(maxNum, minNum)
uninsured20

```

```

## # A tibble: 51 x 3
##   State      minNum maxNum
##   <chr>      <dbl> <dbl>
## 1 Vermont    28170. 28170.
## 2 Rhode Island 43262. 43262.
## 3 North Dakota 51024. 51024.
## 4 Hawaii     59622. 59622.
## 5 Delaware   61751. 61751.
## 6 Wyoming    71851. 71851.
## 7 New Hampshire 83589. 83589.
## 8 Montana    84957. 84957.
## 9 South Dakota 87024. 87024.
## 10 Alaska    90003. 90003.
## # i 41 more rows

```

States with most uninsured people prior to 2014: California, Texas, Florida, New York, Georgia, Illinois, North Carolina, Ohio
 States with the most uninsured people in 2020: Texas, California, Florida, Georgia, North Carolina, New York, Illinois, Ohio

```

highplot_num <- medicaid_expansion %>%
  mutate(num_unIns = population*uninsured_rate) %>%
  filter(year<2014 &
         State %in% c("California", "Texas", "Florida", "New York", "Georgia", "Illinois", "North Carolina", "Ohio"))
ggplot() +
  geom_point(aes(x = year,
                 y = num_unIns,
                 color = State)) +
  geom_line(aes(x = year,
                y = num_unIns,
                color = State)) +
  theme(axis.title = element_text()) +
  ggtitle('States With Highest # of Uninsured People (pre 2014)') +
  xlab('Year') +
  ylab('# of People Uninsured')

highplot_num20 <- medicaid_expansion %>%
  mutate(num_unIns = population*uninsured_rate) %>%
  filter(year==2020 &
         State %in% c("Texas", "California", "Florida", "Georgia", "North Carolina", "New York", "Illinois", "Ohio"))
ggplot() +
  geom_bar(aes(x = State,
               y = num_unIns),
           stat = 'identity') +
  theme(axis.title = element_text()) +

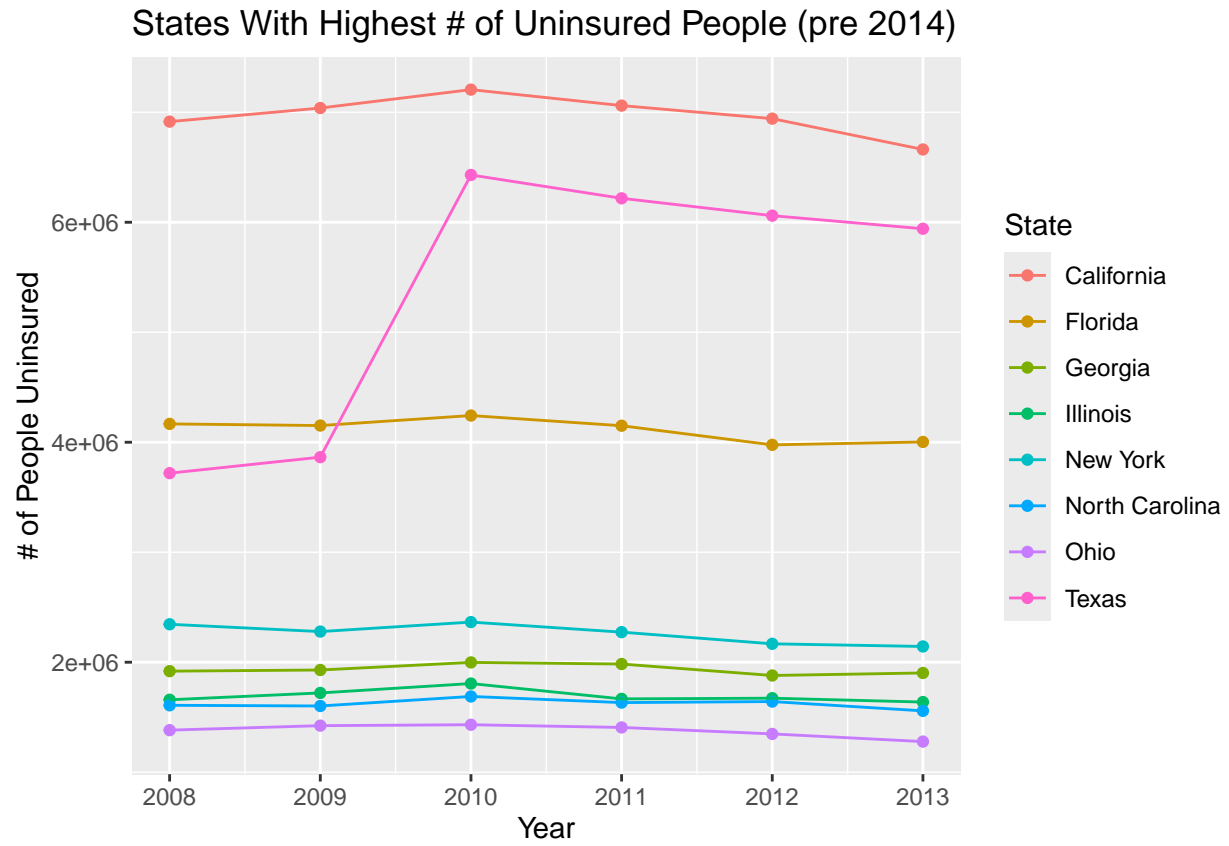
```

```

ggtitle('States With Highest # of Uninsured People (2020)' ) +
xlab('State') +
ylab('# of People Uninsured')

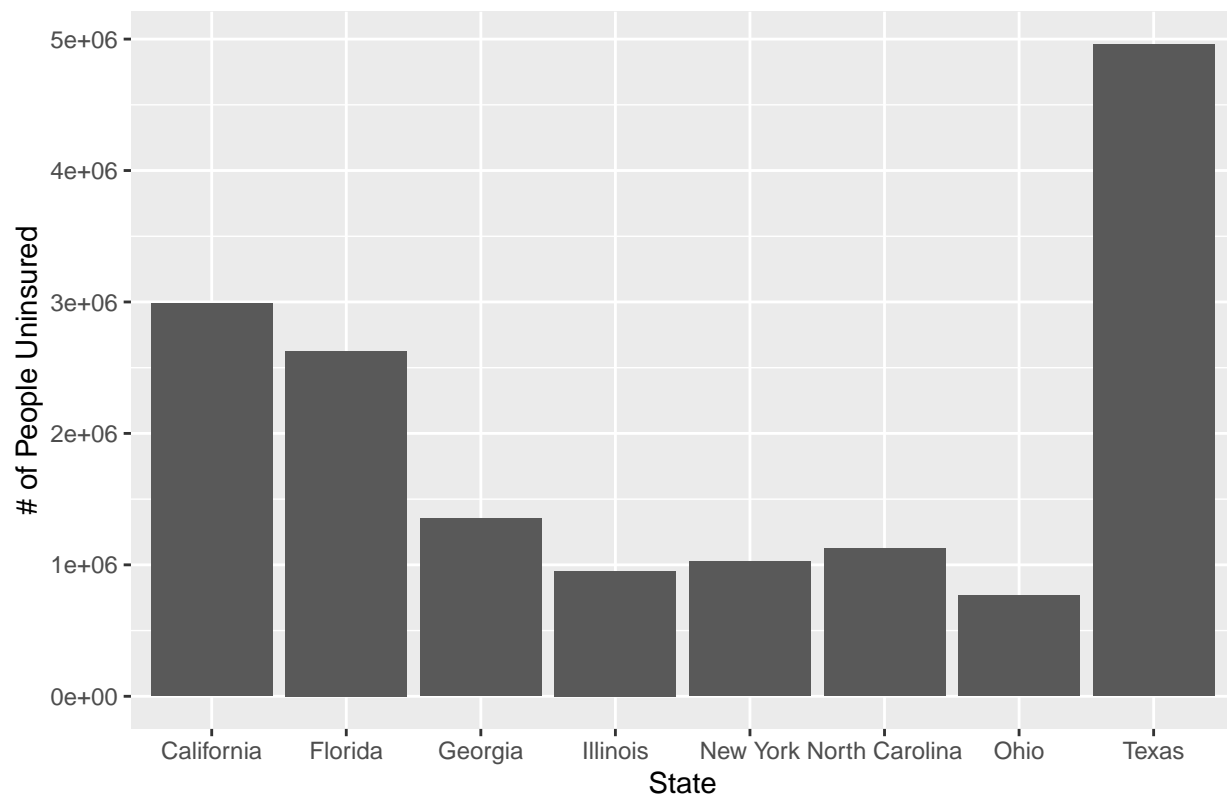
```

highplot_num



highplot_num20

States With Highest # of Uninsured People (2020)



Difference-in-Differences Estimation

Estimate Model

Do the following:

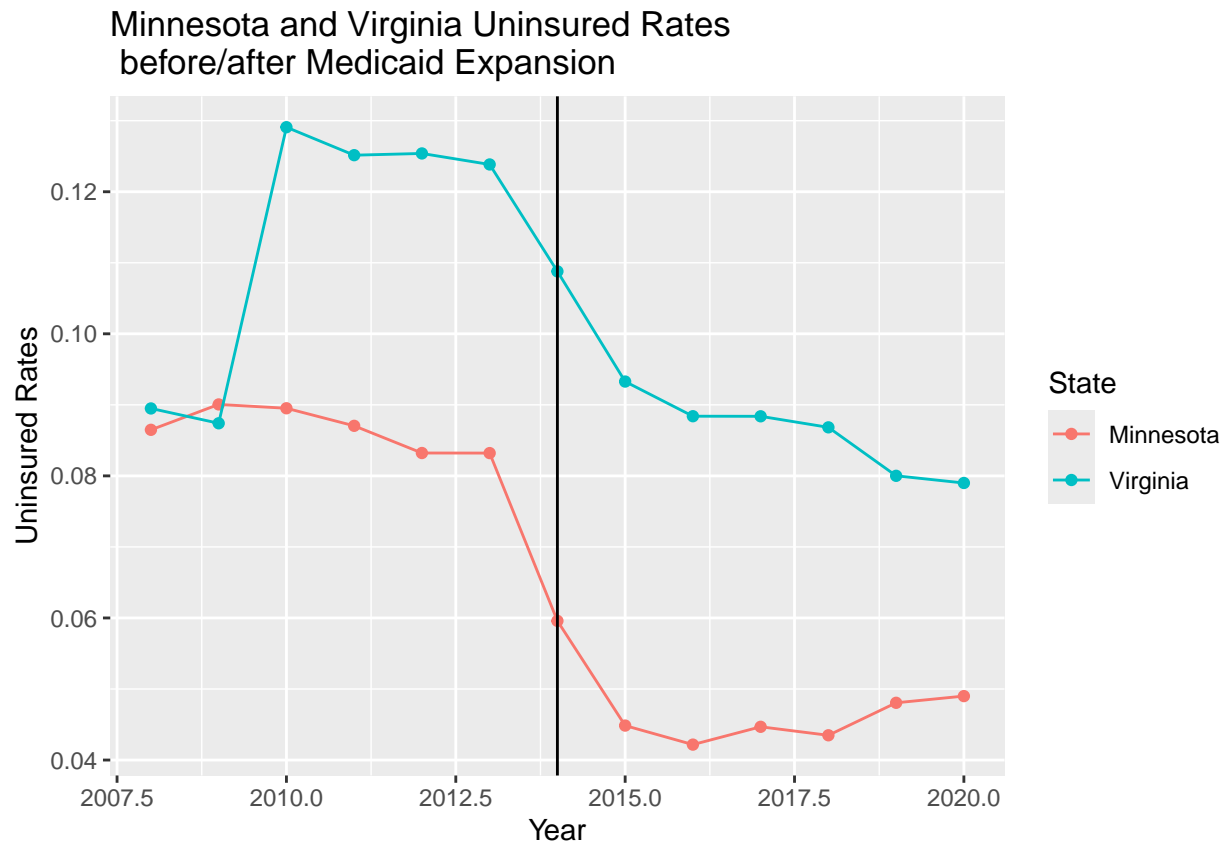
- Choose a state that adopted the Medicaid expansion on January 1, 2014 and a state that did not. **Hint:** Do not pick Massachusetts as it passed a universal healthcare law in 2006, and also avoid picking a state that adopted the Medicaid expansion between 2014 and 2015.
- Assess the parallel trends 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

# Parallel trends plot #1: comparing Minnesota and Virginia (adopted 07/01/2019)
medicaid_expansion %>%
  filter(State %in% c("Minnesota", "Virginia")) %>%
  ggplot() +
    geom_point(aes(x = year,
                    y = uninsured_rate,
                    color = State)) +
    geom_line(aes(x = year,
                  y = uninsured_rate,
                  color = State)) +
```



```
geom_vline(aes(xintercept = 2014)) +
# theme_fivethirtyeight() +
theme(axis.title = element_text()) +
ggtitle('Minnesota and Virginia Uninsured Rates \n before/after Medicaid Expansion') +
xlab('Year') +
ylab('Uninsured Rates')
```



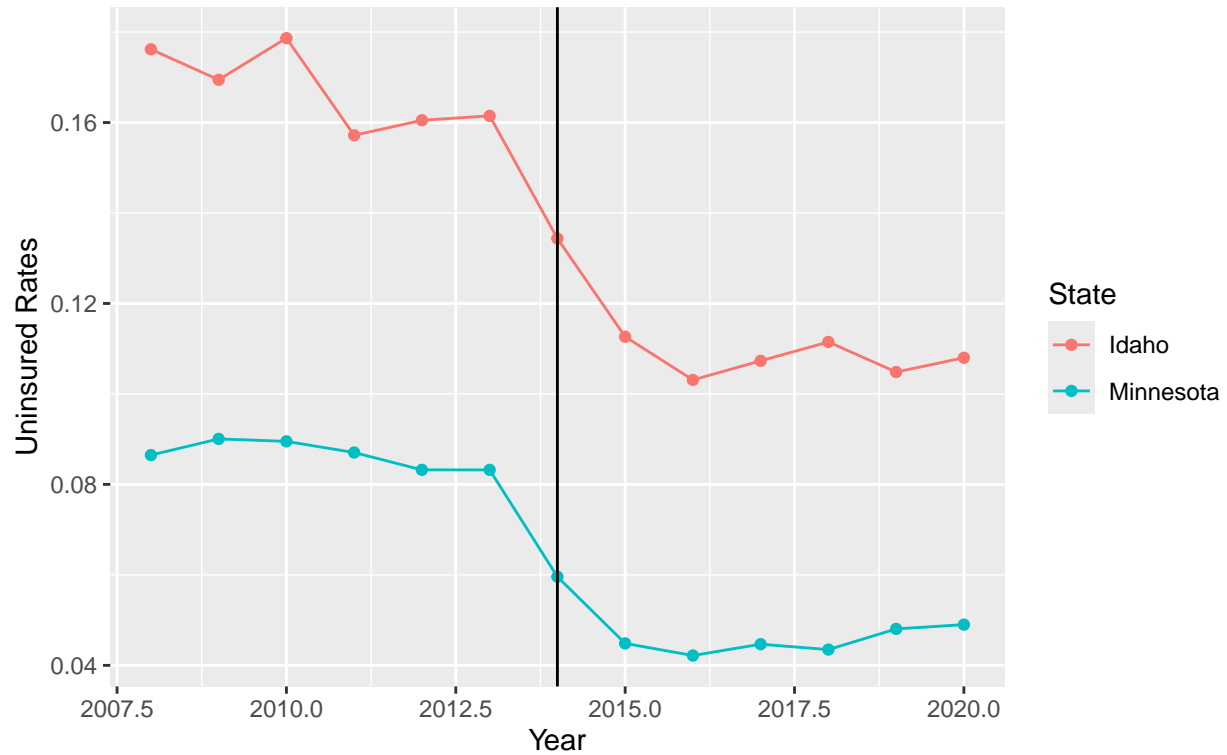
We see a small gap between the two states up until 2010 where the gap becomes drastically larger. Will choose a different state.

```
# Parallel Trends plot

# Parallel trends plot #2: comparing Minnesota and Idaho (adopted 01/01/2020)
medicaid_expansion %>%
  filter(State %in% c("Minnesota", "Idaho")) %>%
  ggplot() +
    geom_point(aes(x = year,
                    y = uninsured_rate,
                    color = State)) +
    geom_line(aes(x = year,
                  y = uninsured_rate,
                  color = State)) +
    geom_vline(aes(xintercept = 2014)) +
# theme_fivethirtyeight() +
theme(axis.title = element_text()) +
ggtitle('Minnesota and Idaho Uninsured Rates \n before/after Medicaid Expansion') +
```

```
xlab('Year') +
ylab('Uninsured Rates')
```

Minnesota and Idaho Uninsured Rates before/after Medicaid Expansion

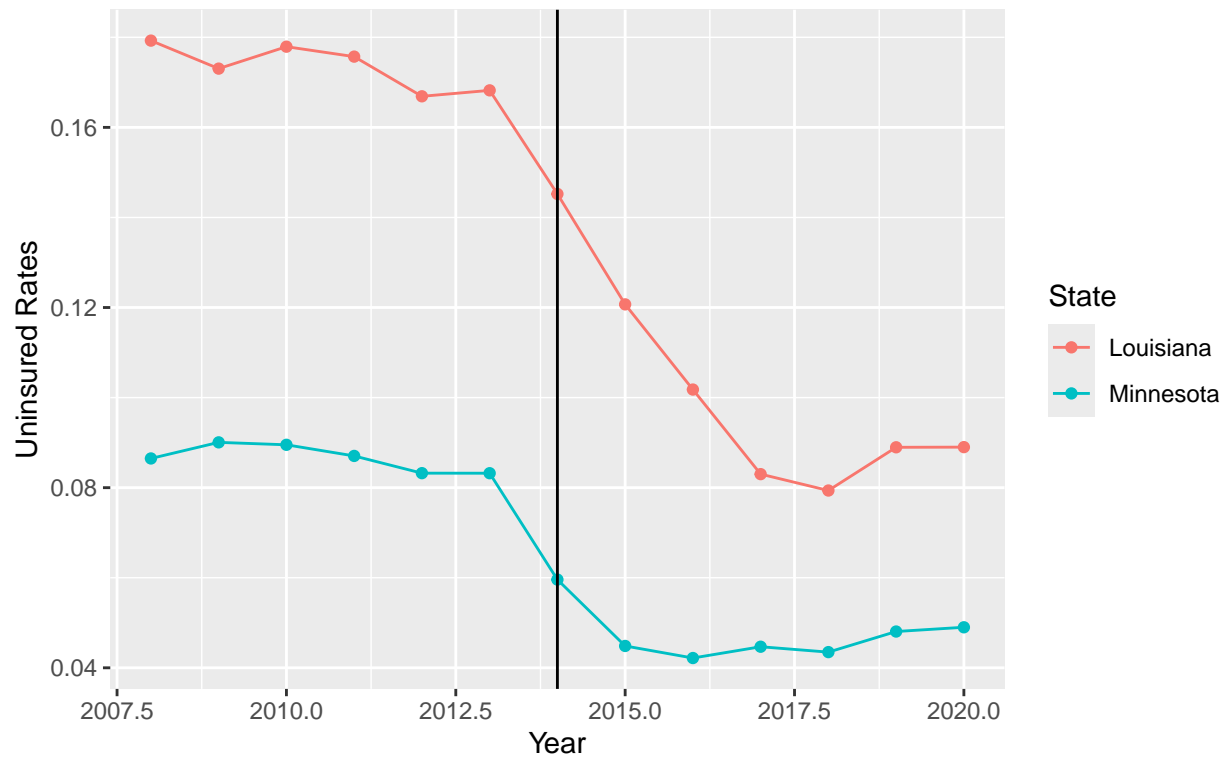


Almost parallel but some divergence in 2010.

```
# Parallel Trends plot

# Parallel trends plot #3: comparing Minnesota and Louisiana (adopted 07/01/2016)
medicaid_expansion %>%
  filter(State %in% c("Minnesota", "Louisiana")) %>%
  ggplot() +
    geom_point(aes(x = year,
                   y = uninsured_rate,
                   color = State)) +
    geom_line(aes(x = year,
                  y = uninsured_rate,
                  color = State)) +
    geom_vline(aes(xintercept = 2014)) +
    # theme_fivethirtyeight() +
    theme(axis.title = element_text()) +
    ggtitle('Minnesota and Louisiana Uninsured Rates \n before/after Medicaid Expansion') +
    xlab('Year') +
    ylab('Uninsured Rates')
```

Minnesota and Louisiana Uninsured Rates before/after Medicaid Expansion



Closest one so far!!!! (Minnesota vs Louisiana)

- Estimates 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

```
# Diff in Diff Estimate Comparing Uninsured Rates from Minnesota vs Louisiana
kc <- medicaid_expansion %>%
  filter(State %in% c("Minnesota", "Louisiana"))

# pre-treatment difference (average of pretreatment outcomes)
pre_diff <- kc %>%
  filter(year < 2014) %>%
  select(State,
    uninsured_rate) %>%
  group_by(State) %>%
  mutate(pre_avg = mean(uninsured_rate)) %>%
  distinct(State, pre_avg) %>%
  spread(State, pre_avg) %>%
  summarise(Louisiana - Minnesota)

# post-treatment difference
post_diff <- kc %>%
  filter(year >= 2014) %>%
  select(State,
```

```

uninsured_rate) %>%
group_by(State) %>%
mutate(post_avg = mean(uninsured_rate)) %>%
distinct(State, post_avg) %>%
spread(State, post_avg) %>%
summarise(Louisiana - Minnesota)

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

## Louisiana - Minnesota
## 1 -0.03317561

```

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:** This is because our dataset covers all states across the US, meaning there are drastic differences in demographics, economics and ultimately politics/policy decisions.
- What are the strengths and weaknesses of using the parallel trends assumption in difference-in-differences estimates?
- **Answer:** The parallel trends assumption is fundamental requirement of the Diff in Diff analysis. Strength: If this assumption holds, the control group in the Diff in Diff analysis approximates the treatment group across all time points (pre-treatment) and thus accounts for all observed and unobserved confounding. Limitation: Difficult to test for parallel trends if the treatment is endogenous.

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.

```

# non-augmented synthetic control
#install.packages("Matrix", dependencies = TRUE)
#install.packages("augsynth")
#install.packages("dplyr")
library(Matrix)

```

```
##
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

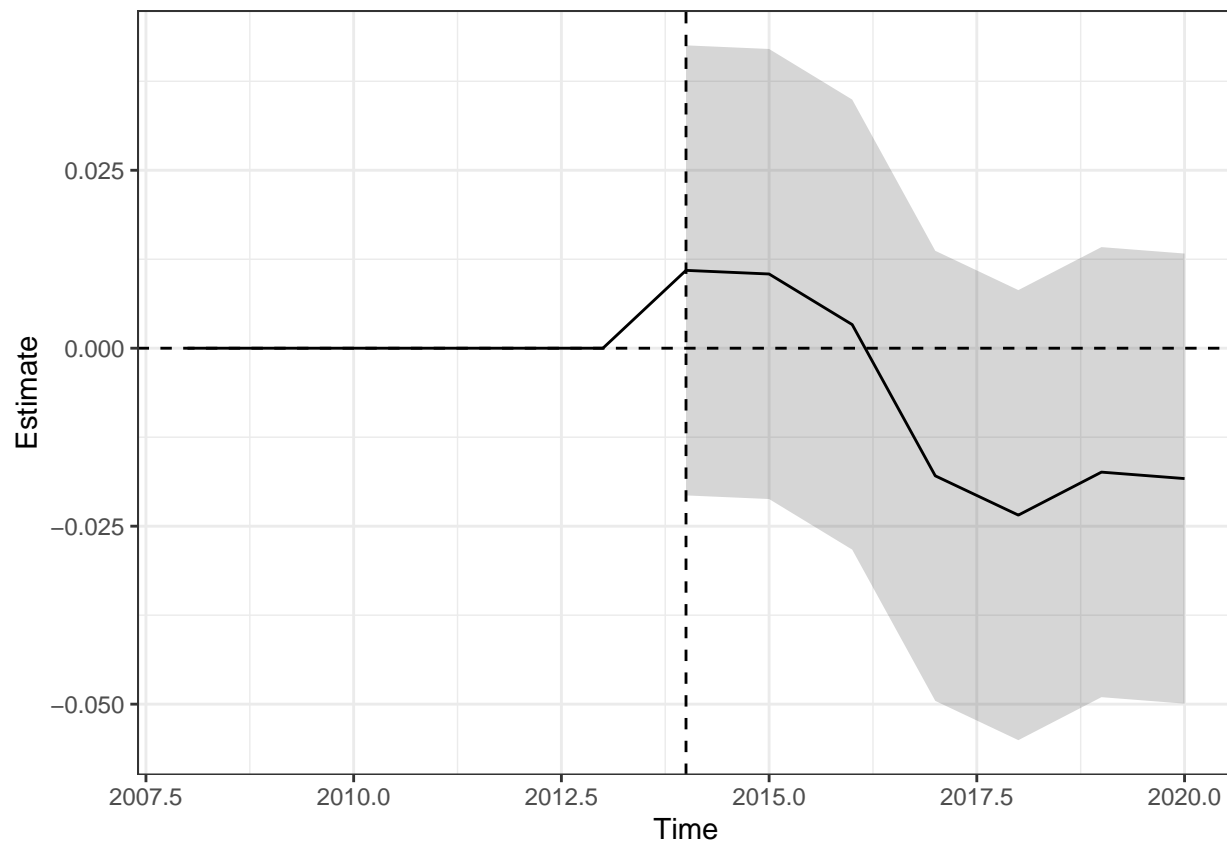
```
library(tidyverse)
library(augsynth)
data(kansas)
dat <- medicaid_expansion %>%
# filter(State %in% c("Louisiana")) %>%
  mutate(treated = case_when(State == "Louisiana" & year >= 2014 ~ 1,
                             TRUE ~ 0))
syn <- augsynth(uninsured_rate ~ treated, State, year, dat,
progfunc = "None", scm = T)
```

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

```
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): -0.00749 ( 0.02 )
## 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
## 2014 0.011 -0.021 0.043 0.140
## 2015 0.010 -0.021 0.042 0.741
## 2016 0.003 -0.028 0.035 0.870
## 2017 -0.018 -0.050 0.014 1.000
## 2018 -0.023 -0.055 0.008 0.880
## 2019 -0.017 -0.049 0.014 1.000
## 2020 -0.018 -0.050 0.013 1.000
```

```
plot(syn)
```



- Re-run the same analysis but this time use an augmentation (default choices are Ridge, Matrix Completion, and GSynth). Create the same plot and report the average ATT and L2 imbalance.

```
# augmented synthetic control
ridge_syn <- augsynth(unsured_rate ~ treated, State, year, dat,
  progfunc = "ridge", scm = T)
```

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

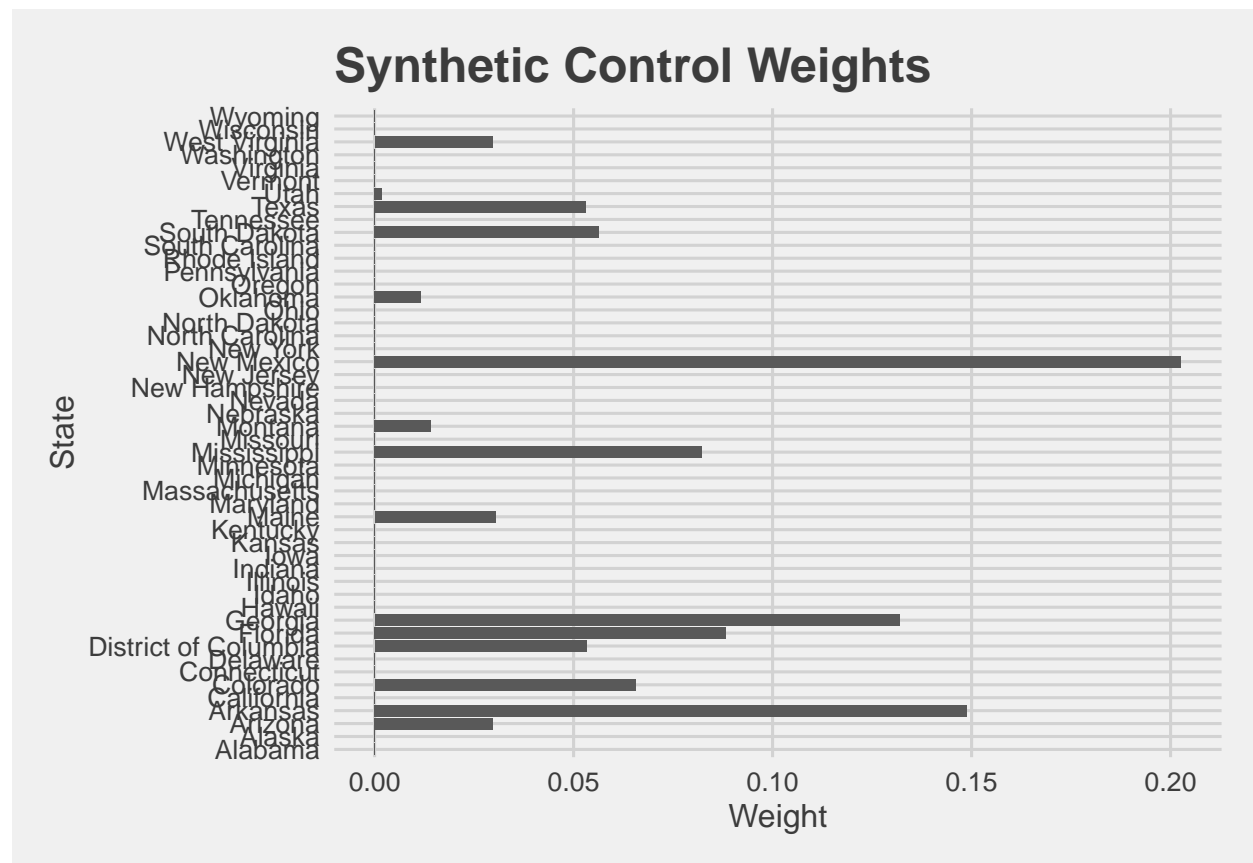
```
summary(ridge_syn)
```

```
##
## 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):  -0.00749   ( 0.019 )
## 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
##	2014	0.011	-0.021	0.043	0.133
##	2015	0.010	-0.021	0.042	0.726
##	2016	0.003	-0.028	0.035	0.852
##	2017	-0.018	-0.050	0.014	1.000
##	2018	-0.023	-0.055	0.008	0.849
##	2019	-0.017	-0.049	0.014	1.000
##	2020	-0.018	-0.050	0.013	1.000

- Plot barplots to visualize the weights of the donors.

```
# 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:** Advantages: synthetic controls can provide a more accurate counterfactual by providing a better fit between the treated and synthetic control in the period before treatment. Synthetic control methods are also more flexible because it relaxes the need for the parallel trends assumption. Disadvantages: using synthetic controls are typically computationally intensive. Additionally, depending on how predictors/weights are chosen, there may be concerns of internal validity if done incorrectly.
- 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:** Yes, having negative weights may introduce some problems with interpretability. To assess the tradeoff between using the augmentation vs non-augmentation, it's important to consider whether we are more interested in having a better pre-treatment fit vs having interpretable estimates.

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
library(lubridate)
dat_mul <- medicaid_expansion %>%
  mutate(yearAdop = year(Date_Adopted),
         DA = ifelse(year>yearAdop, 1, 0))

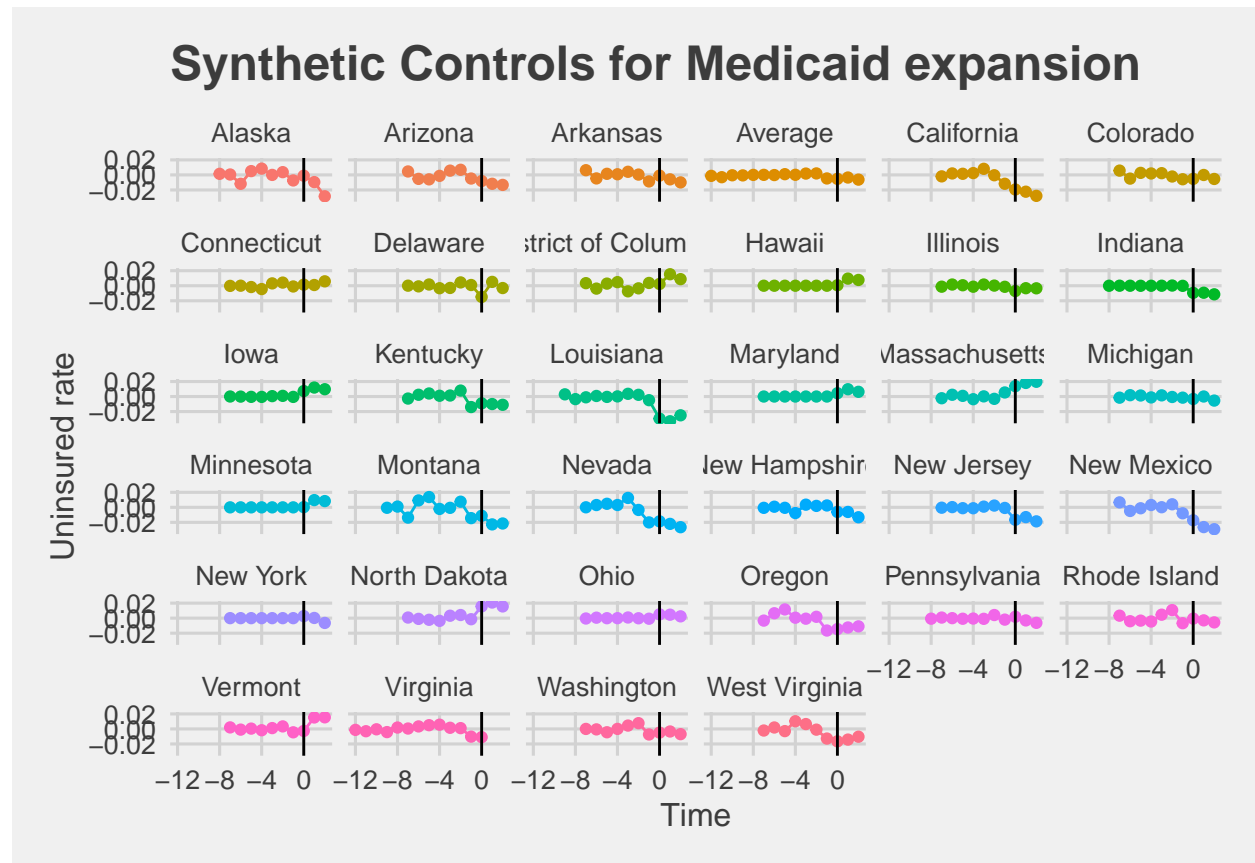
ppool_syn <- multisynth(uninsured_rate ~ DA,
                       State,                # unit
                       year,                 # time
                       nu = 0,               # varying degree of pooling
                       dat_mul, # data
                       n_leads = 3)    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

```
##
## Call:
## multisynth(form = uninsured_rate ~ DA, unit = State, time = year,
##   data = dat_mul, n_leads = 3, nu = 0)
##
## Average ATT Estimate (Std. Error): -0.005 (0.011)
##
## Global L2 Imbalance: 0.002
```



```
## Scaled Global L2 Imbalance: 0.088
## Percent improvement from uniform global weights: 91.2
##
## Individual L2 Imbalance: 0.004
## Scaled Individual L2 Imbalance: 0.094
## Percent improvement from uniform individual weights: 90.6
##
## Time Since Treatment   Level      Estimate  Std.Error lower_bound upper_bound
##                      0 Average -0.005220697  0.01129281 -0.02826507  0.009526926
##                      1 Average -0.003433478  0.01188007 -0.02777941  0.012100290
##                      2 Average -0.006239754  0.01205297 -0.03135487  0.009263188
```

```
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
```



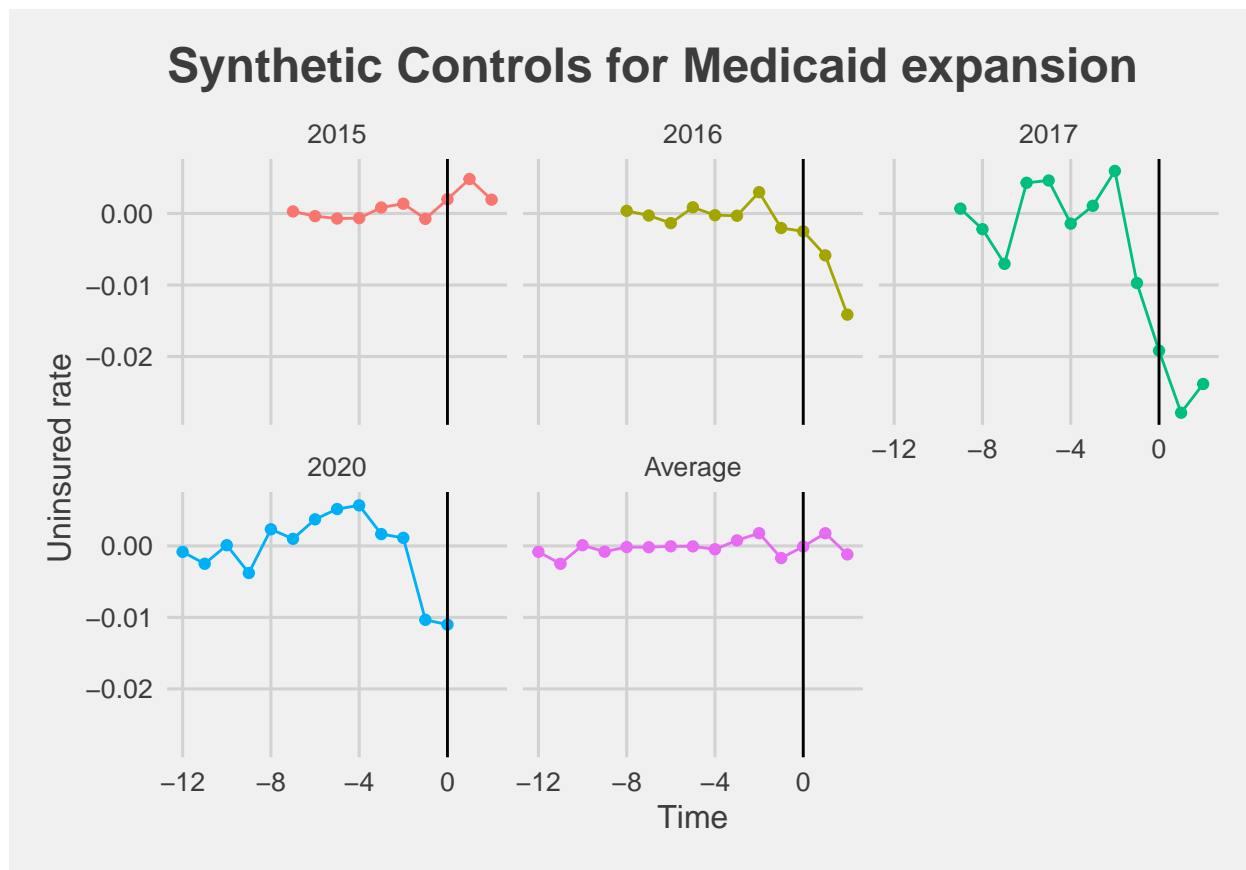
- Estimate a multisynth model using time cohorts. For the purpose of this exercise, you can simplify the

treatment time so that states that adopted Medicaid expansion within the same year (i.e. all states that adopted expansion in 2016) count for the same cohort. Plot the treatment effects for these time cohorts.

```
# multisynth model time cohorts
ppool_syn_coh <- multisynth(uninsured_rate ~ DA,
                           State,                # unit
                           year,                 # time
                           nu = 0,              # varying degree of pooling
                           dat_mul, # data
                           n_leads = 3,
                           time_cohort = TRUE)  # post-treatment periods to estimate
ppool_syn_coh_summ <- summary(ppool_syn_coh)
ppool_syn_coh_summ
```

```
##
## Call:
## multisynth(form = uninsured_rate ~ DA, unit = State, time = year,
##   data = dat_mul, n_leads = 3, nu = 0, time_cohort = TRUE)
##
## Average ATT Estimate (Std. Error): 0.000 (0.015)
##
## Global L2 Imbalance: 0.006
## Scaled Global L2 Imbalance: 0.042
## Percent improvement from uniform global weights: 95.8
##
## Individual L2 Imbalance: 0.012
## Scaled Individual L2 Imbalance: 0.031
## Percent improvement from uniform individual weights: 96.9
##
## Time Since Treatment   Level      Estimate Std.Error lower_bound upper_bound
##                      0 Average -9.497229e-05 0.01444801 -0.02882711 0.01721080
##                      1 Average  1.773862e-03 0.01511125 -0.02851186 0.02036881
##                      2 Average -1.188325e-03 0.01534298 -0.03164501 0.01748040
```

```
ppool_syn_coh_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)
```



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. Many of the states had different treatment effect sizes.
- Do you see evidence for the idea that early adopters of Medicaid expansion enjoyed a larger decrease in the uninsured population?
- **Answer:** No. States with early adoption of Medicaid did not show larger decreases in their uninsured population.

General Discussion Questions

- Why are DiD and synthetic control estimates well suited to studies of aggregated units like cities, states, countries, etc?
- **Answer:** Units like cities, states and countries have very clear exposure to a treatment (i.e. medicaid expansion) Further, for DID, for example, if the parallel-trends assumption is met, we are assuming

that all observed and unobserved confounders over time are accounted for (which means strong internal validity).

- 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/synthetic control, selection into treatment is not randomly assigned, however, there are certain assumptions that we make to handle observed/unobserved confounders (i.e. parallel trends assumption for DID and creating a strong model which predicts a synthetic control that most closely resembles the treated units in the pre-treatment period). It is best to use DID/synthetic control if you can reasonably make these assumptions. In regression discontinuity, selection into treatment is determined based on a cutoff of some variable.