Project 7: Difference-in-Differences and Synthetic Control

Brenda Sciepura

April 23, 2024, 14:43

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:

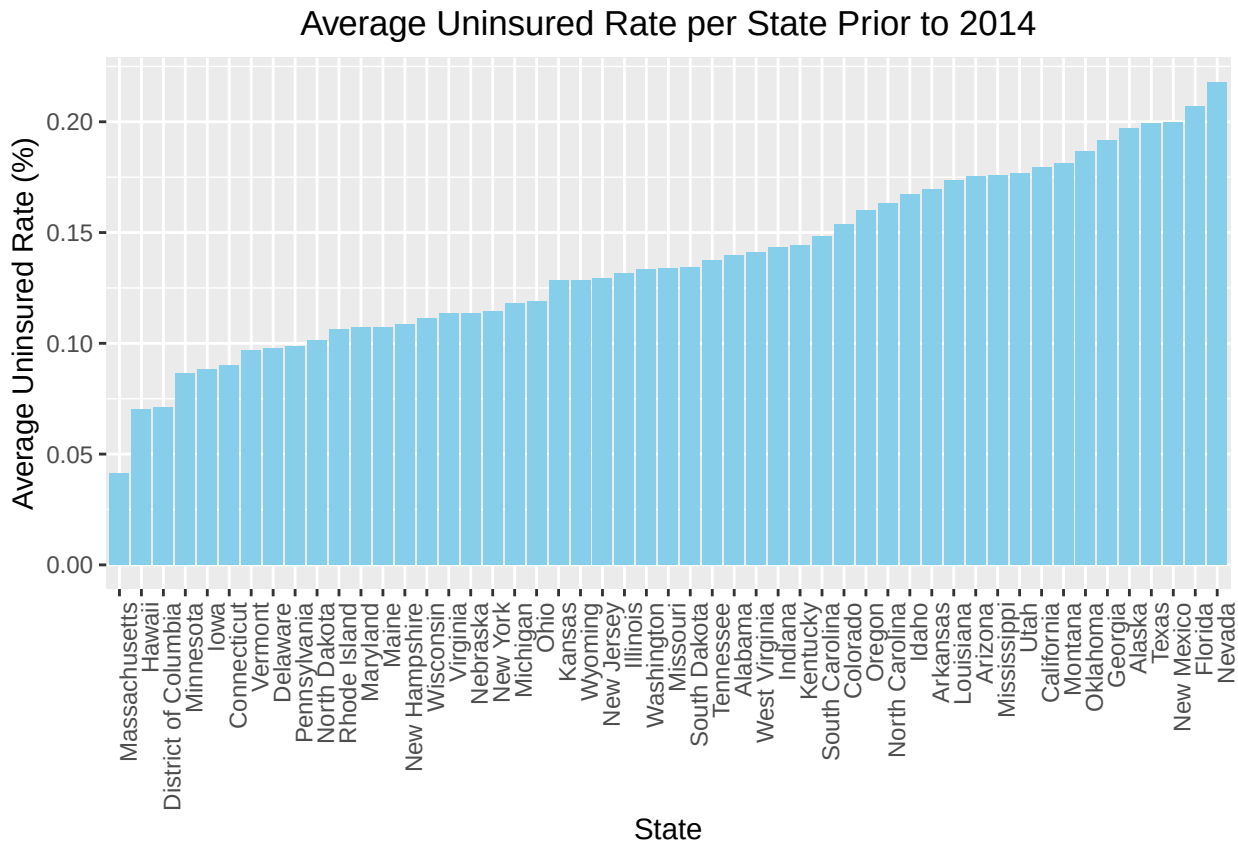
- **State**: Full name of state
- **Medicaid Expansion Adoption**: Date that the state adopted the Medicaid expansion, if it did so.
- **Year**: Year of observation.
- **Uninsured rate**: State uninsured rate in that year.

Exploratory Data Analysis

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

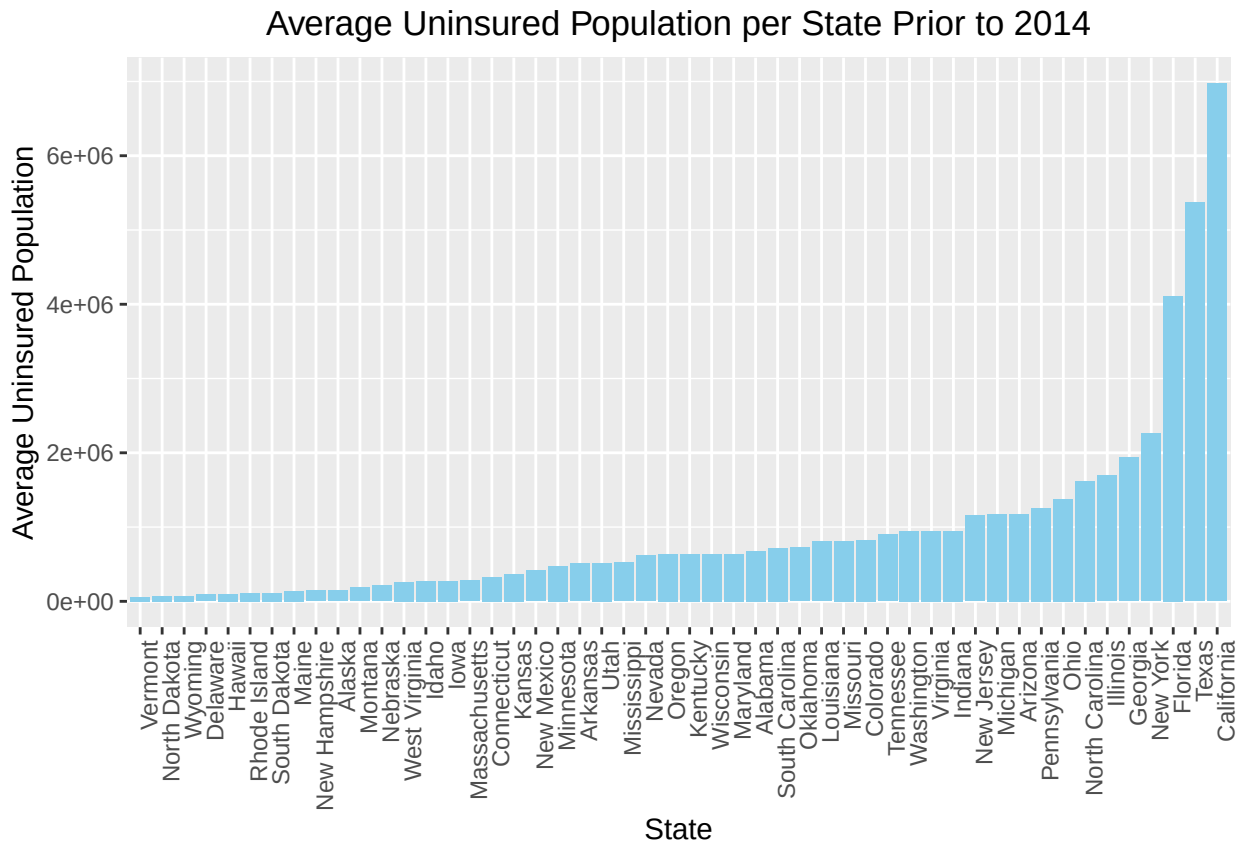
- Which states had the highest uninsured rates prior to 2014? The lowest?
- Which states were home to most uninsured Americans prior to 2014? How about in the last year in the data set? **Note**: 2010 state population is provided as a variable to answer this question. In an actual study you would likely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

```
# highest and lowest uninsured rates prior to 2014
average_rates_prior_2014 <- medicaid_expansion %>%
  filter(year < 2014) %>% ## subsetting dataset to the years prior to 2014
  group_by(State) %>%
  summarise(avg_uninsured_rate = mean(uninsured_rate))

ggplot(average_rates_prior_2014, aes(x = reorder(State, avg_uninsured_rate), y = avg_uninsured_rate)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(x = "State", y = "Average Uninsured Rate (%)", title = "Average Uninsured Rate per State Prior t
  theme(plot.title = element_text(hjust = 0.5),
      axis.text.x = element_text(angle = 90, hjust = 1))
```

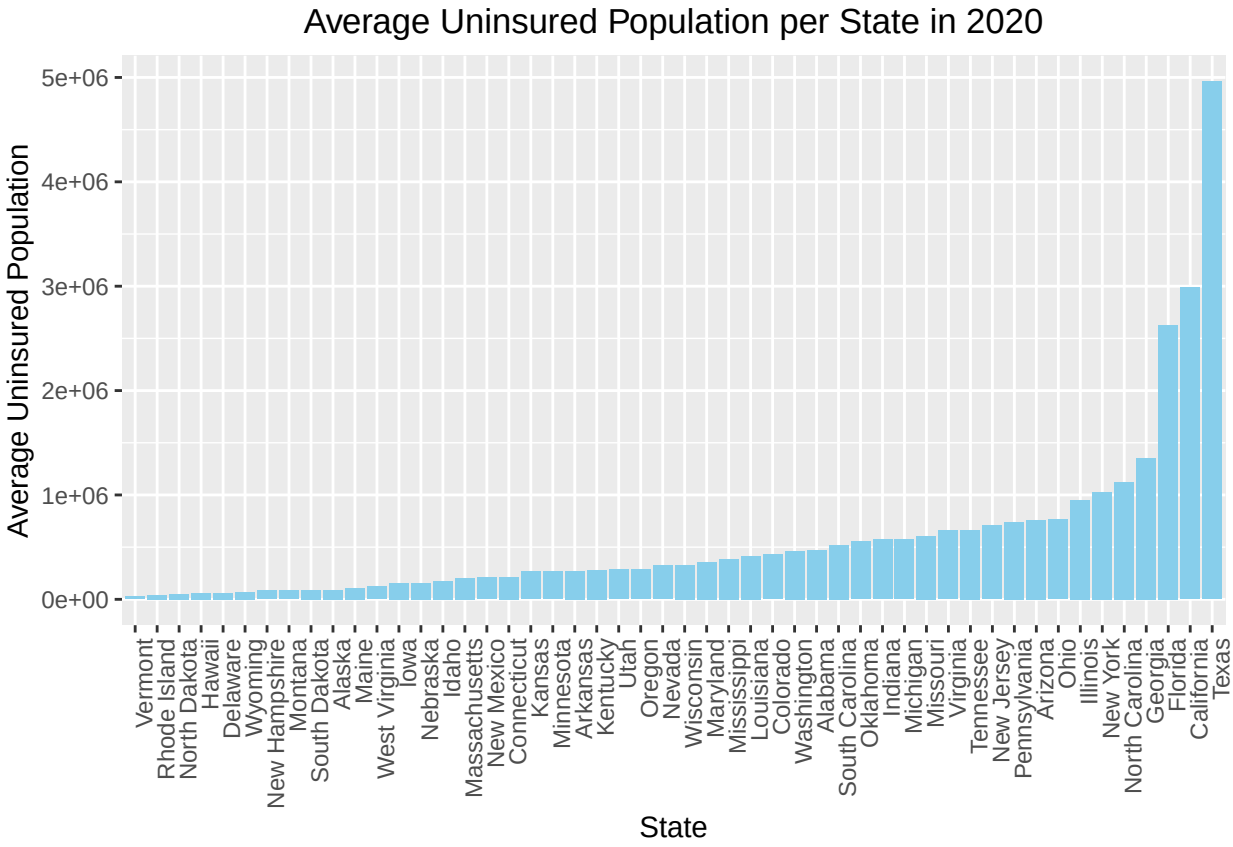


Answer: Massachusetts is the state with the lowest uninsured rate and Nevada the one with the highest prior to 2014.

```
# most uninsured Americans

## uninsured population prior to 2014
medicaid_expansion <- medicaid_expansion %>% mutate(uninsured_pop = uninsured_rate * population)

average_uninsured_population <- medicaid_expansion %>%
  filter(year < 2014) %>% ## subsetting dataset to the years prior to 2014
  group_by(State) %>%
  summarise(avg_uninsured_pop = mean(uninsured_pop)) %>%
  filter(avg_uninsured_pop != "NA")

ggplot(average_uninsured_population, aes(x = reorder(State, avg_uninsured_pop), y = avg_uninsured_pop))
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(x = "State", y = "Average Uninsured Population",
       title = "Average Uninsured Population per State Prior to 2014") +
  theme(plot.title = element_text(hjust = 0.5),
      axis.text.x = element_text(angle = 90, hjust = 1))
```




```r
## uninsured population in the last year of the dataset (2020)
average_uninsured_population_2020 <- medicaid_expansion %>%
  filter(year == 2020) %>% ## subsetting dataset to the years prior to 2014
  group_by(State) %>%
  summarise(avg_uninsured_pop_2020 = mean(uninsured_pop)) %>%
  filter(avg_uninsured_pop_2020 != "NA")

ggplot(average_uninsured_population_2020, aes(x = reorder(State, avg_uninsured_pop_2020),
                                              y = avg_uninsured_pop_2020)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(x = "State", y = "Average Uninsured Population",
       title = "Average Uninsured Population per State in 2020") +
  theme(plot.title = element_text(hjust = 0.5),
      axis.text.x = element_text(angle = 90, hjust = 1))
```



Answer: Prior to 2014, NYC, Florida, Texas and California were the 4 states with the highest proportion of uninsured population prior to 2014. In 2020 we don't see NYC and we see California among the 4 states with the highest uninsured population.

Difference-in-Differences Estimation

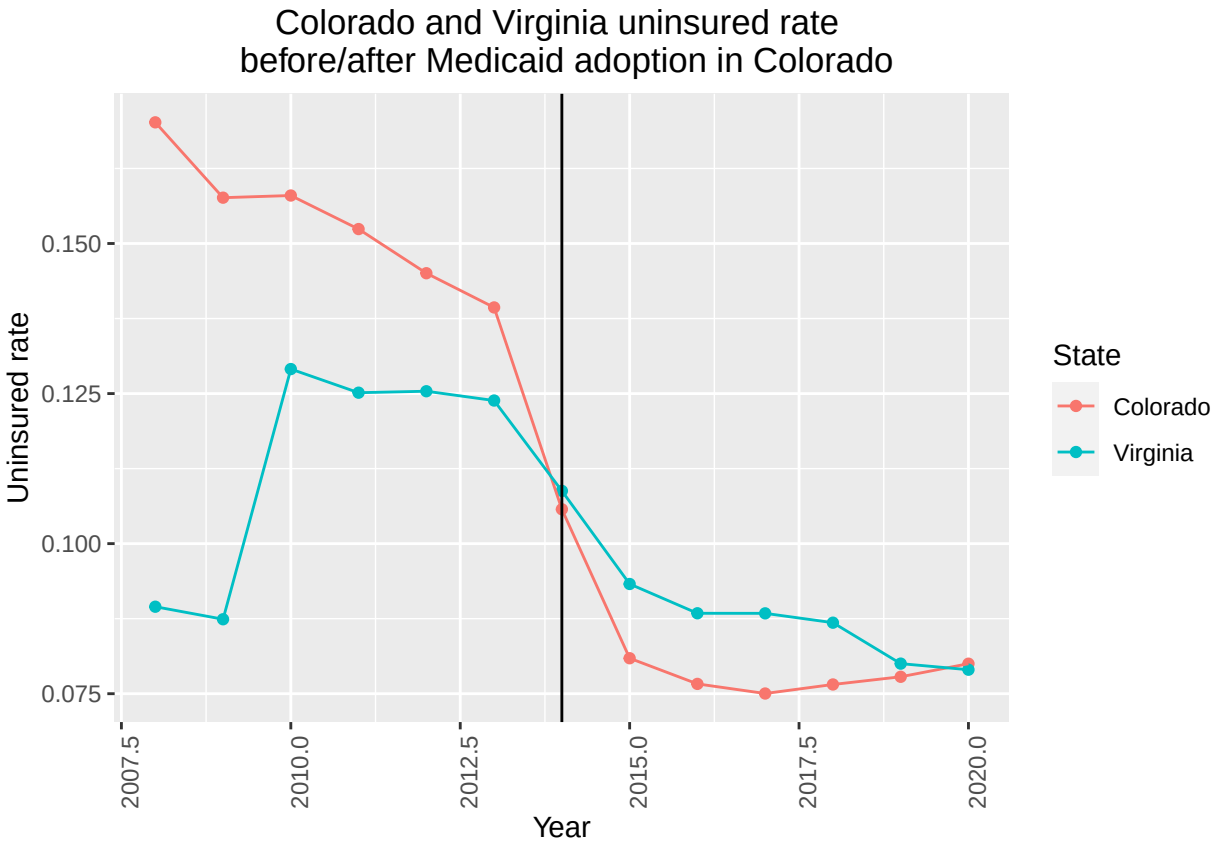
Estimate Model

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: Try 1

## Picks:
## State that adopted Medicaid expansion on Jan 1, 2014 --> Colorado
## State that did not: Virginia (adopted Jan 1, 2019)

medicaid_expansion %>%
filter(State %in% c("Colorado", "Virginia")) %>%
```

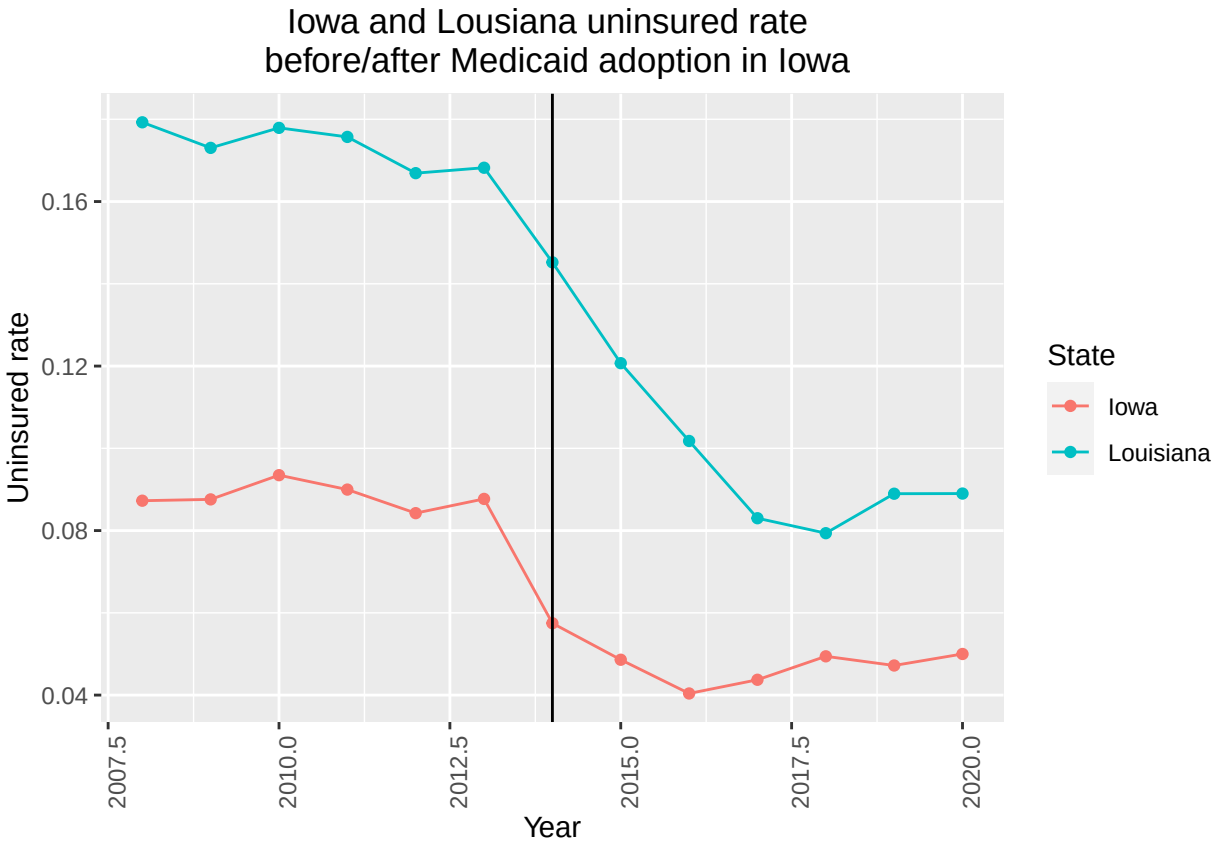
```r
# plotting all of the time periods -- not filtering out any of them

# plot
# ---------
ggplot() +
# add in point layer
geom_point(aes(x = year,
               y = uninsured_rate,
               color = State)) +
# add in line layer
geom_line(aes(x = year,
              y = uninsured_rate,
              color = State)) +
# add a horizontal line
geom_vline(aes(xintercept = 2014)) +

# themes
theme(plot.title = element_text(hjust = 0.5),
  axis.text.x = element_text(angle = 90, hjust = 1)) +

# labels
ggtitle('Colorado and Virginia uninsured rate \n before/after Medicaid adoption in Colorado') +
xlab('Year') +
ylab('Uninsured rate')
```



Answer: Does not satisfy the parallel trends assumption. The gap between Colorado and Virginia closes

quickly before the intervention period, but before that we can observe that at times when Colorado was reducing the uninsured rate, Virginia was growing it, so they don't move in the same direction in the pre-period even though they end up in a similar condition right before intervention period.

```
# Parallel Trends plot: Try 2

## Picks:
## State that adopted Medicaid expansion on Jan 1, 2014 --> Iowa
## State that did not: Louisiana (adopted Jan 1, 2019)

medicaid_expansion %>%
filter(State %in% c("Iowa", "Louisiana")) %>%
  # plotting all of the time periods -- not filtering out any of them

  # plot
  # ---------
  ggplot() +
  # add in point layer
  geom_point(aes(x = year,
                 y = uninsured_rate,
                 color = State)) +
  # add in line layer
  geom_line(aes(x = year,
                y = uninsured_rate,
                color = State)) +
  # add a horizontal line
  geom_vline(aes(xintercept = 2014)) +

  # themes
  theme(plot.title = element_text(hjust = 0.5),
      axis.text.x = element_text(angle = 90, hjust = 1)) +

  # labels
  ggtitle('Iowa and Lousiana uninsured rate \n before/after Medicaid adoption in Iowa') +
  xlab('Year') +
  ylab('Uninsured rate')
```



Answer: This performs a bit better in terms of parallel trends. Event though Iowa is always performing better (having lower uninsured rates compared to Louisiana), they move in a similar fashion in the pre-period. We see them sort of stagnating, and then decreasing the uninsured rate about one year prior to the intervention period.

- 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: Iowa-Louisiana
# ------------------------------------------------
# create a dataset for Iowa-Louisiana
il <-
  medicaid_expansion %>%
  filter(State %in% c("Iowa","Louisiana"))

# pre-treatment difference
# ----------
pre_diff <-
  il %>%
  # filter out only the quarter we want
  filter(year < 2014) %>%
  # subset to select only vars we want
```

```
  select(State,
         uninsured_rate) %>%
  group_by(State) %>%
  summarize(mean(uninsured_rate))

pre_diff <- pre_diff$`mean(uninsured_rate)`[1] - pre_diff$`mean(uninsured_rate)`[2]

# post-treatment difference
# ----------
post_diff <-
  il %>%
  # filter out only the quarter we want
  filter(year >= 2014) %>%
  # subset to select only vars we want
  select(State,
         uninsured_rate) %>%
  group_by(State) %>%
  summarize(mean(uninsured_rate))

post_diff <- post_diff$`mean(uninsured_rate)`[1] - post_diff$`mean(uninsured_rate)`[2]

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

```
## [1] 0.03210726
```

Answer: Looks like our treatment effect is about 3pp, which doesn't seem trivial if it were a true effect. However, here we checked that Iowa and Louisiana are similar in terms of the uninsured rate pre-treatment, but we did not check other similarities. We are implicitly assuming that Iowa and Louisiana are similar in other regards as well (other factors that might be correlated with the outcome).

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: In this original study, because towns were neighboring you could establish that they were similar in a lot of regards that could affect the outcome and that they could be comparable groups. However, in a study spanning multiple states, we encounter substantial heterogeneity. Unlike adjacent towns, states exhibit significant variations in demographic composition, economic indicators, and healthcare policies, regulations, among other contextual factors. These disparities can introduce confounding variables, posing challenges in isolating the causal impact of a specific policy adoption by a certain state. Additionally, geographical features, population densities, socioeconomic conditions, and access to healthcare services also vary across states, which might cause uninsured rates to vary not solely on the grounds of a policy change. Additionally, spillover effects are plausible, wherein policy changes in one state may influence neighboring states, and these can make it hard to set the boundaries between treatment and control groups, complicating identification of causal effects.

- What are the strengths and weaknesses of using the parallel trends assumption in difference-in-differences estimates?

Answer: **Strength:** This assumption, if it holds, is easy to interpret. The parallel trends assumption suggests that any difference in outcomes observed after the treatment is likely attributable to the treatment itself, rather than pre-existing differences between the treatment and control groups. **Weaknesses:** (1) This assumption is sort of weak as any deviation from parallel trends can bias the estimated treatment effect, (2) It depends on the length and quality of the pre-treatment period, and with limited data you wouldn't be able to test it empirically with credible results, (3) even if it holds for the pre-treatment period, it may not be valid for all the post-treatment period, especially if there are time-varying confounders that affect the treatment and control groups differently.

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:

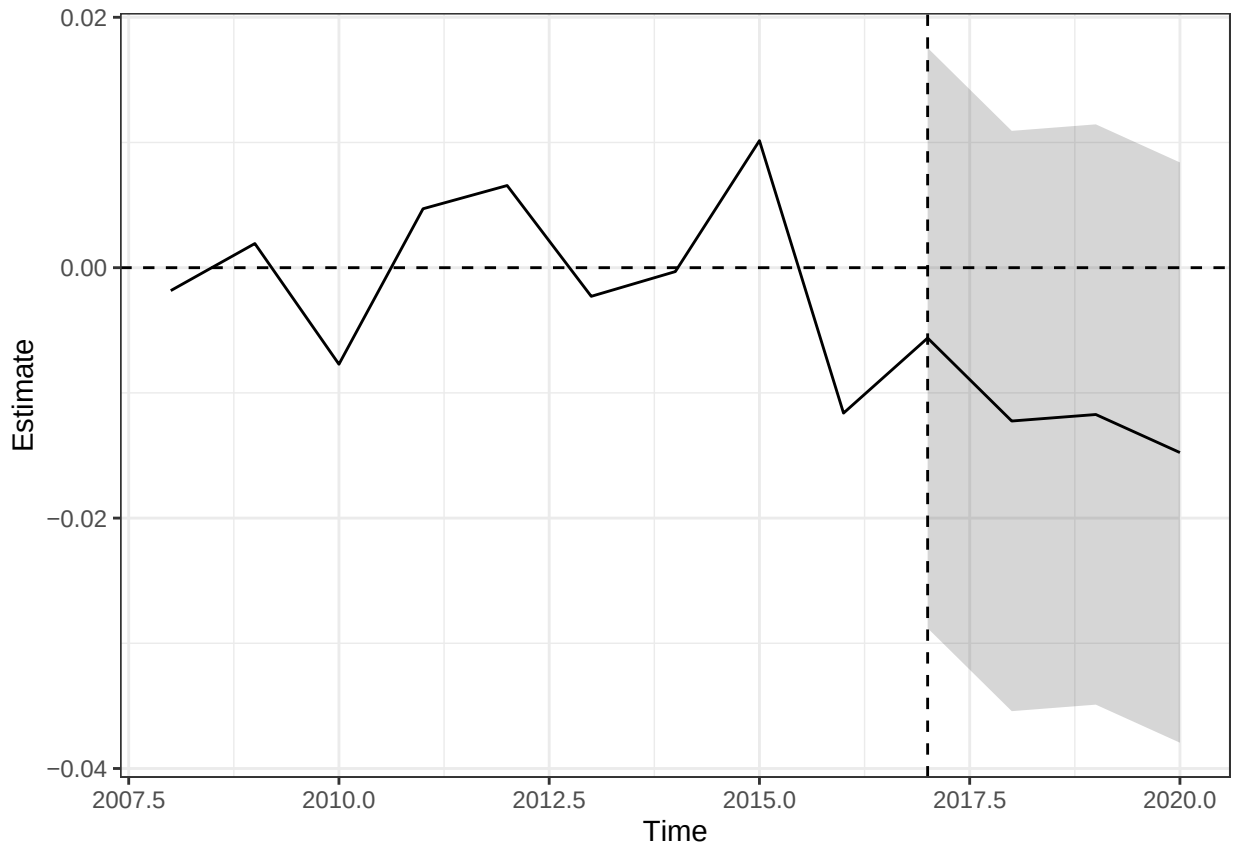
- Choose a state that adopted the Medicaid expansion after January 1, 2014. Construct a non-augmented synthetic control and plot the results (both pre-treatment fit and post-treatment differences). Also report the average ATT and L2 imbalance.

```
# non-augmented synthetic control

# create a treatment indicator
# ----------

medicaid_expansion$year_adopted <- substr(medicaid_expansion$Date_Adopted, 1, 4)

medicaid_expansion <-
  medicaid_expansion %>%
  mutate(treatment = if_else(State == "Montana" & year > year_adopted, 1, 0))

## State chosen: Montana

syn <-                                              # save object
  augsynth(uninsured_rate ~ treatment,              #
                            State,                  # unit
                            year,                   # time
                            medicaid_expansion,     # data
                            progfunc = "None",      # plain syn control
                            scm = T)                # synthetic control

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):  -0.0111   ( 0.73 )
## L2 Imbalance: 0.019
## Percent improvement from uniform weights: 83.8%
##
## Avg Estimated Bias: NA
##
## Inference type: Conformal inference
##
##  Time Estimate 95% CI Lower Bound 95% CI Upper Bound p Value
##  2017   -0.006             -0.029              0.018   0.718
##  2018   -0.012             -0.035              0.011   0.594
##  2019   -0.012             -0.035              0.011   0.516
##  2020   -0.015             -0.038              0.008   0.516
```

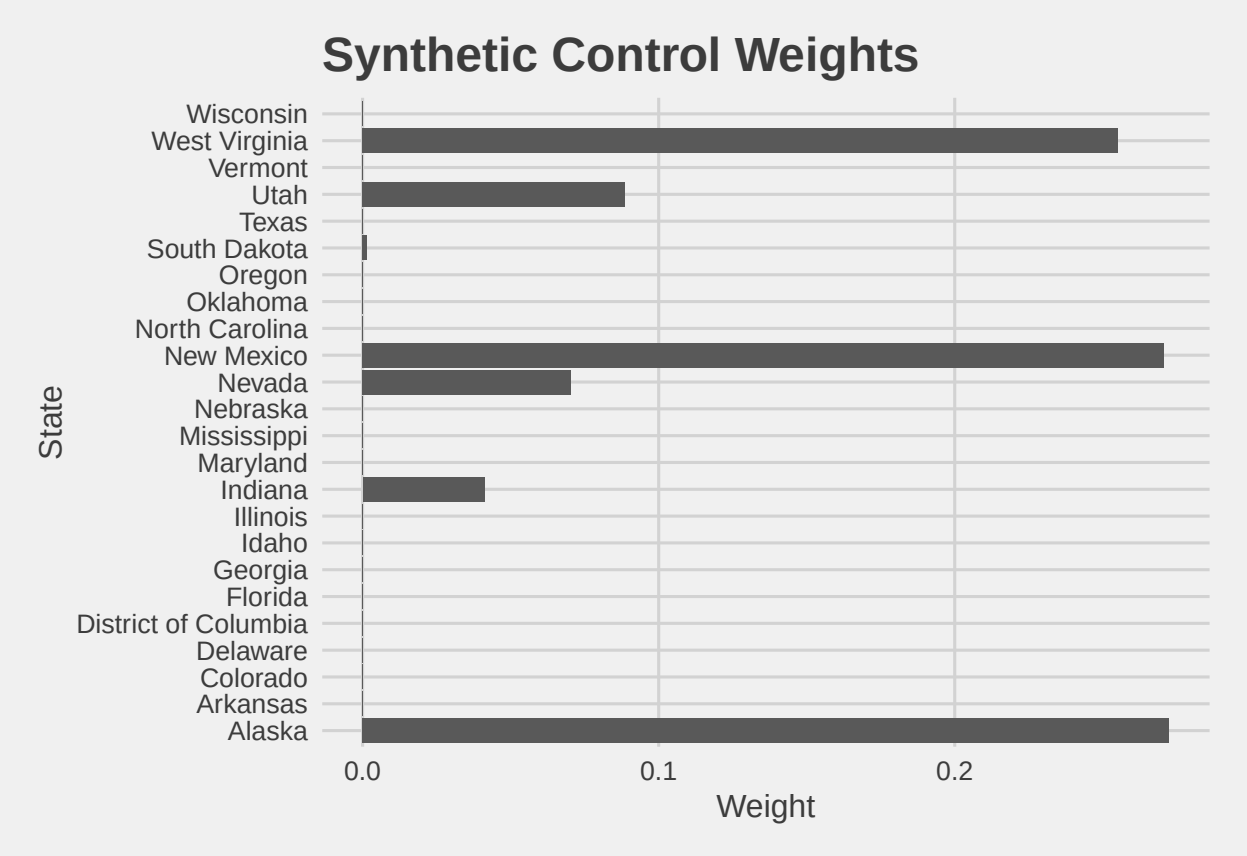
```
plot(syn)
```




```
## Average ATT: -0.0111
## L2 Imbalance: 0.019
```

```
# view each state's contribution

data.frame(syn$weights) %>% # coerce to data frame since it's in vector form
  # process
```

```
  # ---------
  tibble::rownames_to_column('State') %>%
  filter(syn$weights > 0) %>%
  # plot
  # ---------
  ggplot() +
  geom_bar(aes(x = State,
               y = syn.weights),
           stat = 'identity') +
  coord_flip() +
  theme_fivethirtyeight() +
  theme(axis.title = element_text()) +
  ggtitle('Synthetic Control Weights') +
  xlab('State') +
  ylab('Weight')
```



Answer: Only a few states are contributing to creating a fake counterfactual: West Virginia, Utah, New Mexico, Nevada, Indiana and Alaska.

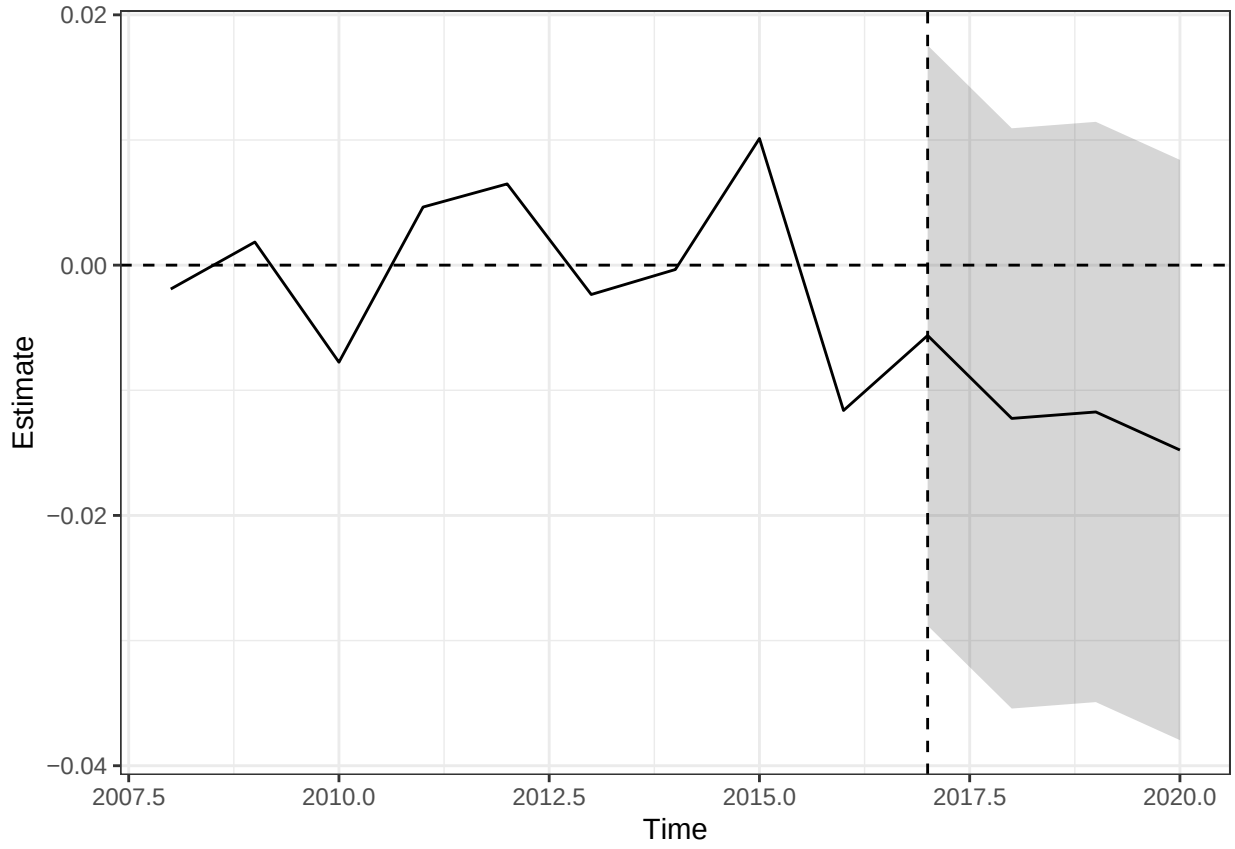
- Re-run the same analysis but this time use an augmentation (default choices are Ridge, Matrix Completion, and GSynth). Create the same plot and report the average ATT and L2 imbalance.

```
# augmented synthetic control
ridge_syn <-
  augsynth(uninsured_rate ~ treatment,
```

```
                            State,
                            year,
                            medicaid_expansion,
                            progfunc = "ridge",
                            scm = T)

summary(ridge_syn) # same L2 balance, no improvement
```

```
##
## 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):  -0.0111   ( 0.71 )
## L2 Imbalance: 0.019
## Percent improvement from uniform weights: 83.8%
##
## Avg Estimated Bias: 0.000
##
## Inference type: Conformal inference
##
##  Time Estimate 95% CI Lower Bound 95% CI Upper Bound p Value
##  2017   -0.006             -0.029              0.018   0.700
##  2018   -0.012             -0.035              0.011   0.567
##  2019   -0.012             -0.035              0.011   0.462
##  2020   -0.015             -0.038              0.008   0.596
```

```
plot(ridge_syn)
```




```
## Average ATT Estimate:  -0.0111
## L2 balance: 0.019

gsynth_syn <-
  augsynth(uninsured_rate ~ treatment,
                            State,
                            year,
                            medicaid_expansion,
                            progfunc = "GSYN",  # specify
                            scm = T)
```

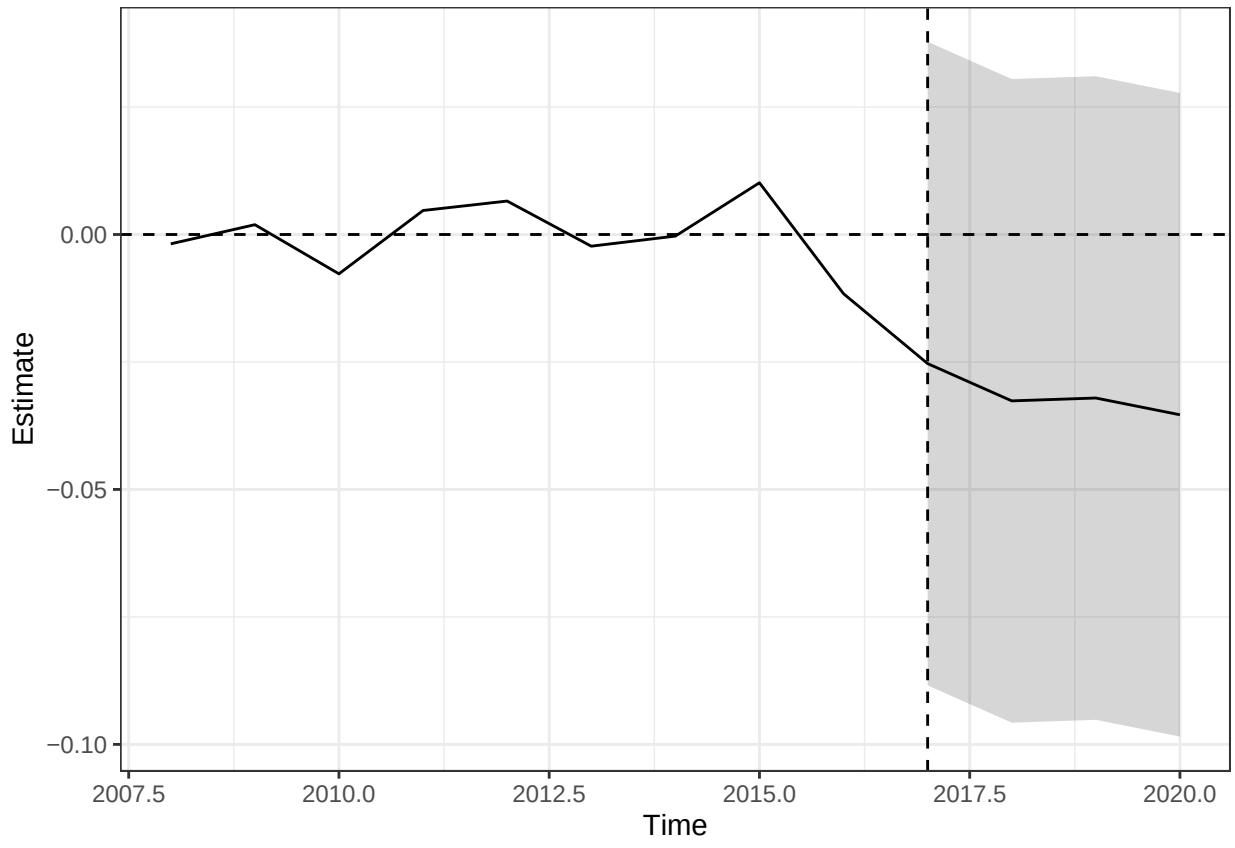
```
## Cross-validating ...
##  r = 0; sigma2 = 0.00024; IC = -7.71591; PC = 0.00022; MSPE = 0.00029
##  r = 1; sigma2 = 0.00010; IC = -8.00201; PC = 0.00021; MSPE = 0.00015*
##  r = 2; sigma2 = 0.00002; IC = -8.82752; PC = 0.00008; MSPE = 0.00016
##  r = 3; sigma2 = 0.00002; IC = -8.64913; PC = 0.00007; MSPE = 0.00015
##  r = 4; sigma2 = 0.00001; IC = -8.28725; PC = 0.00008; MSPE = 0.00020
##  r = 5; sigma2 = 0.00001; IC = -7.88880; PC = 0.00008; MSPE = 0.00050
##
##  r* = 1
```

```
summary(gsynth_syn) # same L2 balance, higher ATT
```

```
##
## Call:
```

```
## single_augsynth(form = form, unit = !!enquo(unit), time = !!enquo(time),
##     t_int = t_int, data = data, progfunc = "GSYN", scm = ..2)
##
## Average ATT Estimate (p Value for Joint Null):  -0.0313   ( 0.72 )
## L2 Imbalance: 0.019
## Percent improvement from uniform weights: 83.8%
##
## Avg Estimated Bias: 0.006
##
## Inference type: Conformal inference
##
##  Time Estimate 95% CI Lower Bound 95% CI Upper Bound p Value
##  2017   -0.025             -0.088              0.038   0.668
##  2018   -0.033             -0.096              0.030   0.600
##  2019   -0.032             -0.095              0.031   0.488
##  2020   -0.035             -0.098              0.028   0.506
```

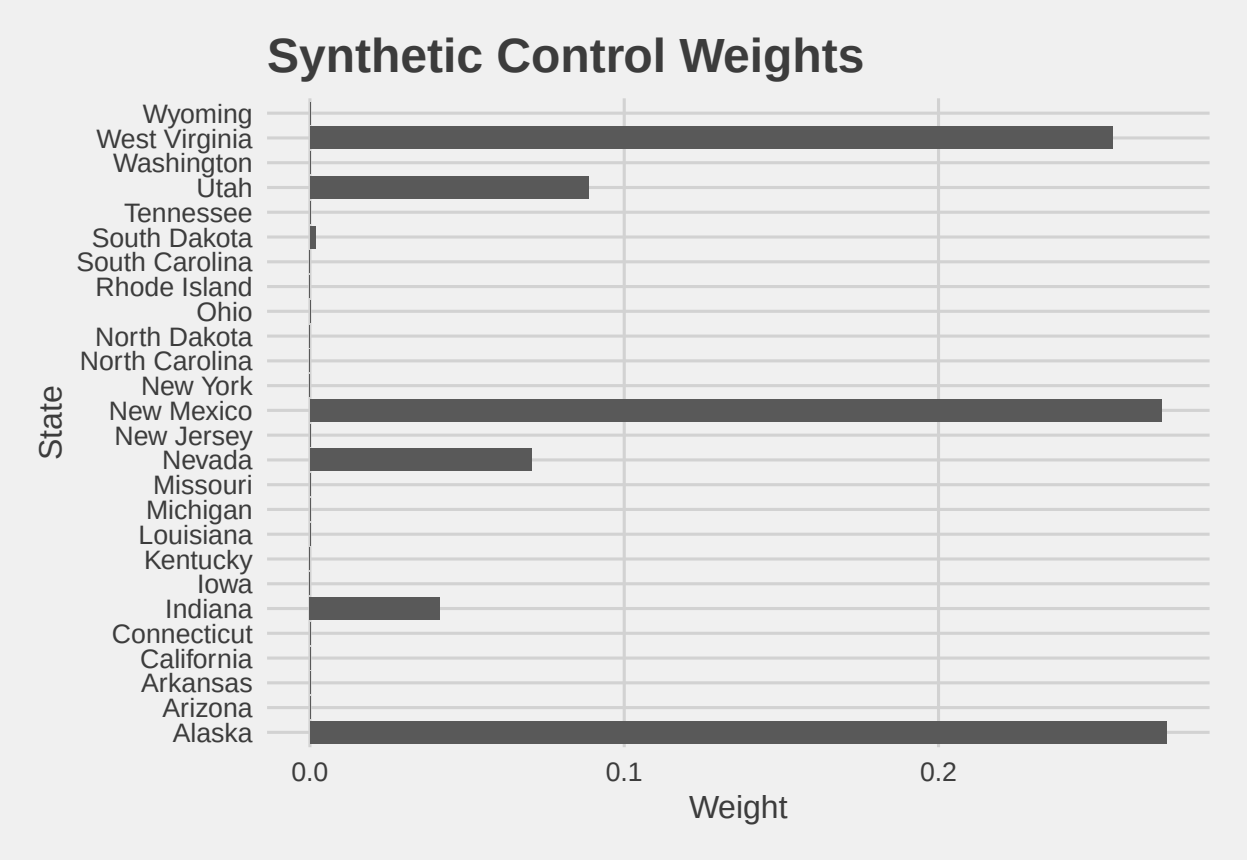
```
plot(gsynth_syn)
```




```
## Average ATT Estimate: -0.0313
## L2 balance: 0.019
```

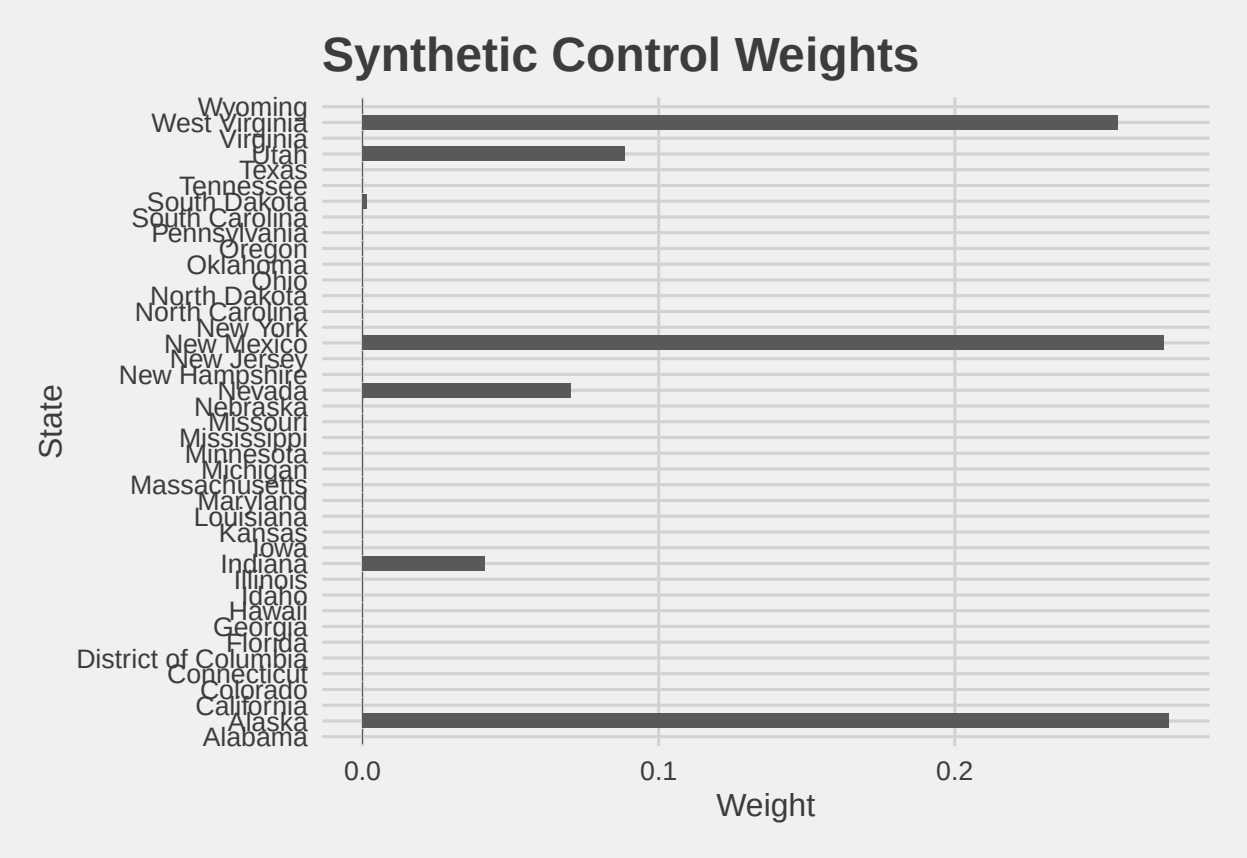
- Plot barplots to visualize the weights of the donors.

```r
# barplots of weights

## plotting weights of donors for ridge
data.frame(ridge_syn$weights) %>%
  # process
  # ---------
  tibble::rownames_to_column('State') %>%
  filter(ridge_syn.weights >0) %>%
  # plot
  # ---------
  ggplot() +
  geom_bar(aes(x = State,
               y = ridge_syn.weights),
           stat = 'identity') +
  coord_flip() +
  theme_fivethirtyeight() +
  theme(axis.title = element_text()) +
  ggtitle('Synthetic Control Weights') +
  xlab('State') +
  ylab('Weight')
```




```r
## plotting weights of donors for GSYN
data.frame(gsynth_syn$weights) %>%
  # process
  # ---------
```

```r
# change index to a column
tibble::rownames_to_column('State') %>%
filter(gsynth_syn.weights >0) %>%
# plot
# ---------
ggplot() +
geom_bar(aes(x = State,
             y = gsynth_syn.weights),
         stat = 'identity') +
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: (1) SC provides flexibility in selecting the control group (that is a weighted combination of multiple donor units) and allows for a more customized comparison to the treated unit,

(2) It can be useful in the context of small sample sizes because it uses information from multiple donors to create a counterfactual (you don't rely on a single state to do this). Disadvantages: (1) Same as DID, it relies on the assumption of parallel pre-trends. (2) Model choice can influence results (I remember a problem set where depending on the covariates I chose the created counterfactual changed significantly, and the professor told that if results are so sensitive to the model choice then SC might not be the best model to use).

- 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: By performing augmentation and allowing for negative weights, it becomes less intuitive to understand how each donor contributes to the creation of the counterfactual, which makes the interpretation of the causal effect more challenging. However, by allowing augmentation we contribute to achieving more accurate estimates. It is, in the end, a trade-off between interpretability and accuracy. We could potentially do sensitivity analysis to check the robustness of the treatment effect in different model specifications. It's important to be transparent as a researcher regarding the decisions we make: our model choice and weights, and how that might impact results, and also provide information on how to interpret results.

Staggered Adoption Synthetic Control

Estimate Multisynth

Do the following:

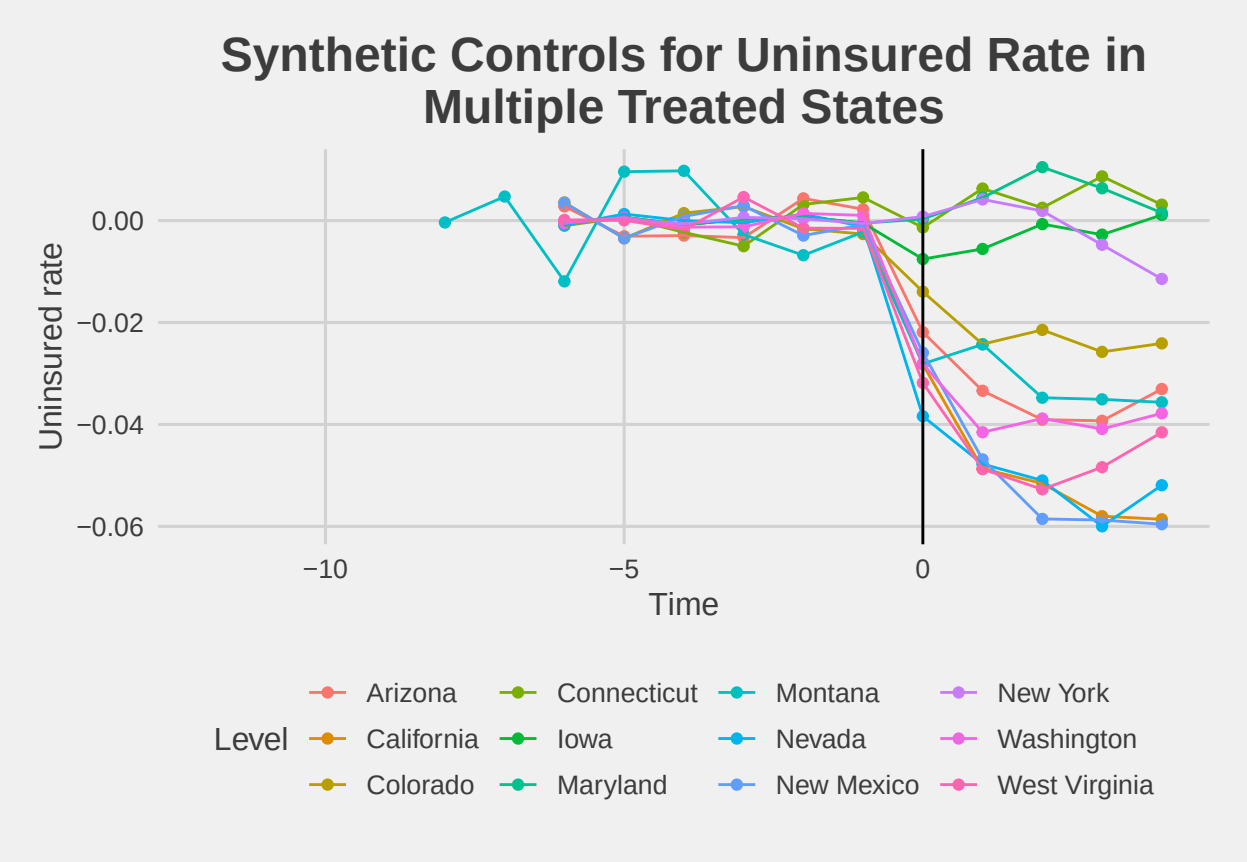
- Estimate a multisynth model that treats each state individually. Choose a fraction of states that you can fit on a plot and examine their treatment effects.

```
medicaid_expansion <-
  medicaid_expansion %>%
  filter(!State %in% c("Massachusetts")) %>%
  mutate(treatment_staggered = if_else(year >= year_adopted, 1, 0))

# multisynth model states

# setting nu to 0.5
# ---------
multi_syn <- multisynth(uninsured_rate ~ treatment_staggered,
                        State,                      # unit
                        year,                       # time
                        medicaid_expansion,         # data
                        n_leads = 5)                # post-treatment periods to estimate

summary(multi_syn)
```

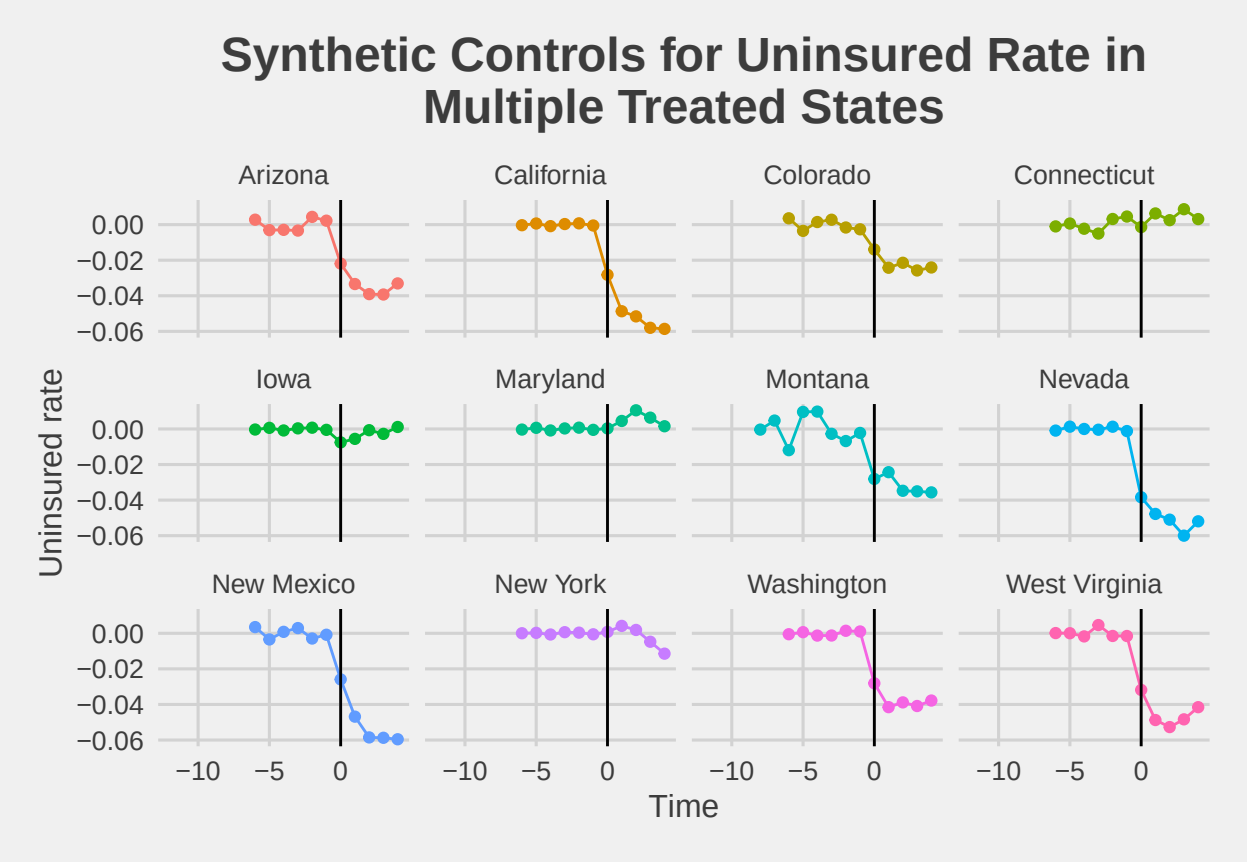
```
##
## Call:
## multisynth(form = uninsured_rate ~ treatment_staggered, unit = State,
##     time = year, data = medicaid_expansion, n_leads = 5)
##
## Average ATT Estimate (Std. Error): -0.015  (0.005)
```

```
##
## Global L2 Imbalance: 0.000
## Scaled Global L2 Imbalance: 0.017
## Percent improvement from uniform global weights: 98.3
##
## Individual L2 Imbalance: 0.004
## Scaled Individual L2 Imbalance: 0.101
## Percent improvement from uniform individual weights: 89.9
##
##  Time Since Treatment   Level    Estimate   Std.Error lower_bound  upper_bound
##                     0 Average -0.01103586 0.004224219 -0.01919225 -0.002965219
##                     1 Average -0.01700362 0.005780981 -0.02895244 -0.005718429
##                     2 Average -0.01583007 0.006565701 -0.02901812 -0.003125473
##                     3 Average -0.01890269 0.006620784 -0.03210388 -0.006105403
##                     4 Average -0.02006543 0.006415738 -0.03273069 -0.008324886
```

```r
multi_syn_summ <- summary(multi_syn)

# List of states
selected_states <- c("Arizona", "California", "Colorado", "Connecticut",
                     "Florida", "Georgia", "Iowa", "Kansas", "Maryland",
                     "Montana", "Nevada", "New Mexico", "New York",
                     "Texas", "Washington", "West Virginia")

# Filter data for selected states
multi_syn_summ$att <- multi_syn_summ$att %>%
  filter(Level %in% selected_states)

multi_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 Uninsured Rate in\nMultiple Treated States') +
  theme(plot.title = element_text(hjust = 0.5)) +
  xlab('Time') +
  ylab('Uninsured rate')
```




```
multi_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 Uninsured Rate in\nMultiple Treated States') +
  theme(plot.title = element_text(hjust = 0.5)) +
  xlab('Time') +
  ylab('Uninsured rate') +
  facet_wrap(~Level)
```

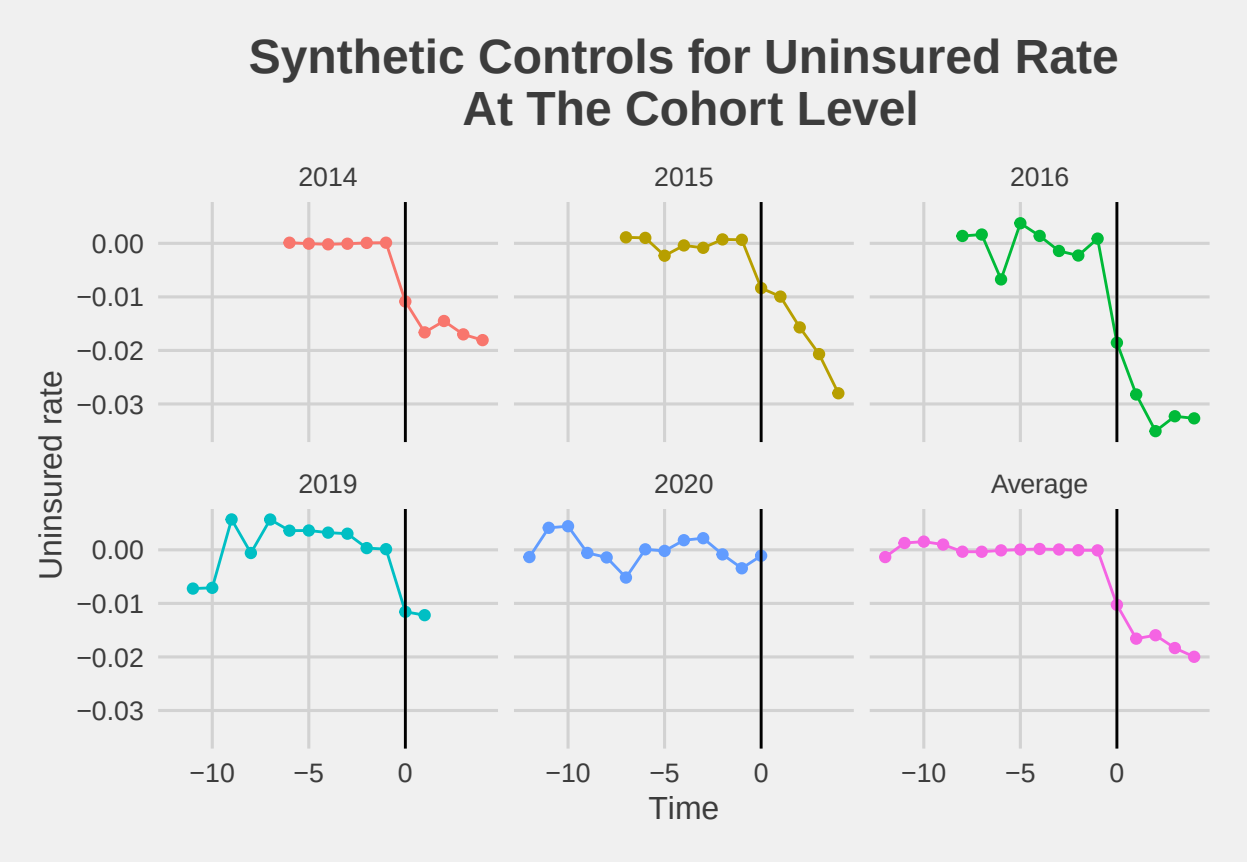


Answer:" In general, the rate of the uninsured population decreased post Medicaid adoption year at the state level. However, the rate of the decrease differs by state. Also, we observe states where the uninsured rate went up post adoption of the policy. This happened, for instance, in Connecticut, Iowa and Maryland. In other states, such as Washington and West Virginia, the uninsured rate went down right after the policy but then it started to go up again. All in all, we see heterogeneity across states.

- 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

multi_time_cohort <- multisynth(uninsured_rate ~ treatment_staggered,
                                State,
                                year,
                                medicaid_expansion,
                                n_leads = 5,
                                time_cohort = TRUE)

# save summary
multi_time_cohort_summ <- summary(multi_time_cohort)


## plot
multi_time_cohort_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 Uninsured Rate\n At The Cohort Level') +
theme(plot.title = element_text(hjust = 0.5)) +
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, we see similar pre-trends but we do see heterogeneity at the time of treatment, were some states saw a decrease in the uninsured rate (i.e. Arizona, California, Colorado),and others an increase (i.e. Connecticut, Iowa and Maryland).

- Do you see evidence for the idea that early adopters of Medicaid expansion enjoyed a larger decrease in the uninsured population?
  Answer: Well, not necessarily. We see a more steep decrease for those who adopted the policy in 2015 and 2016. We cannot see enough of a trend for those who adopted the policy in 2019 and 2020 because of lack of data in post-periods.

General Discussion Questions

- Why are DiD and synthetic control estimates well suited to studies of aggregated units like cities, states, countries, etc?
  Answer: Both of these methods rely on creating a sort of natural control group, they're quasi-experimental methods in nature. Then, it's important to rely on aggregated units like cities, states or countries because they exhibit greater homogeneity with themselves relative to individual-level data, and this mitigates the problem of unobserved heterogeneity within units. This works positively for identifying pre-trends. DID and SC are good methods to identify the effect of policy adoption, which is policy-relevant and many times hard (and sometimes even impossible) to detect with purely experimental methods.

- What role does selection into treatment play in DiD/synthetic control versus regression discontinuity? When would we want to use either method?
  Answer: For RDD selection into treatment has to do with a discontinuity at the cutoff. RDD is sensitive to manipulation around this threshold. If individuals or units have an incentive to manipulate their assignment variable scores to either side of the threshold, it could invalidate the assumption of comparability near the threshold and bias the treatment effect estimates. For DID/SC treatment assignment occurs when a policy change, intervention or natural event happens. For instance, a policy reform can be implemented in different states at different times. In both DID and SC assessing selection into treatment involves studying factors that influence treatment assignment and making sure that they are not systematically relate to the outcome of interest. I don't know if I fully understand the second question, but I think the scenarios to use either method are pretty different. For RDD you need to have a threshold that cannot be manipulated by the individual and you need a lot of observations around that threshold, so I think that if you have those conditions, you'd most likely use RDD. In other words, RDD is suitable when you have a clear threshold that separates units into treatment and control groups, and you want to estimate the causal effect of the treatment near that threshold. DID/SC is used when you have longitudinal data observed over multiple time periods (it can be just two periods, with one control and one treatment unit, or multiple periods when unit become treated at a certain point in time), and you want to estimate the causal effect of an intervention (usually policy adoption/reform, etc.) that occurs at a specific time point.