Project 7: Difference-in-Differences and Synthetic Control

```
# turn off scientific notation
options(scipen = 999)

# set chunk options
knitr::opts_chunk$set(echo = TRUE, message = FALSE, warning = FALSE)

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

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

```
# load other packages
pacman::p_load(ggthemes, tidyverse, magrittr, scales)

# load the augsynth functions from source files
path <- "./augsynth-master/R/"
file.sources <- paste0(path, list.files(path = path, pattern ="*.R"))
sapply(file.sources, source, .GlobalEnv)

# set seed
set.seed(44)

# load data
medicaid_expansion <- read.csv('./data/medicaid_expansion.csv')
```

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:

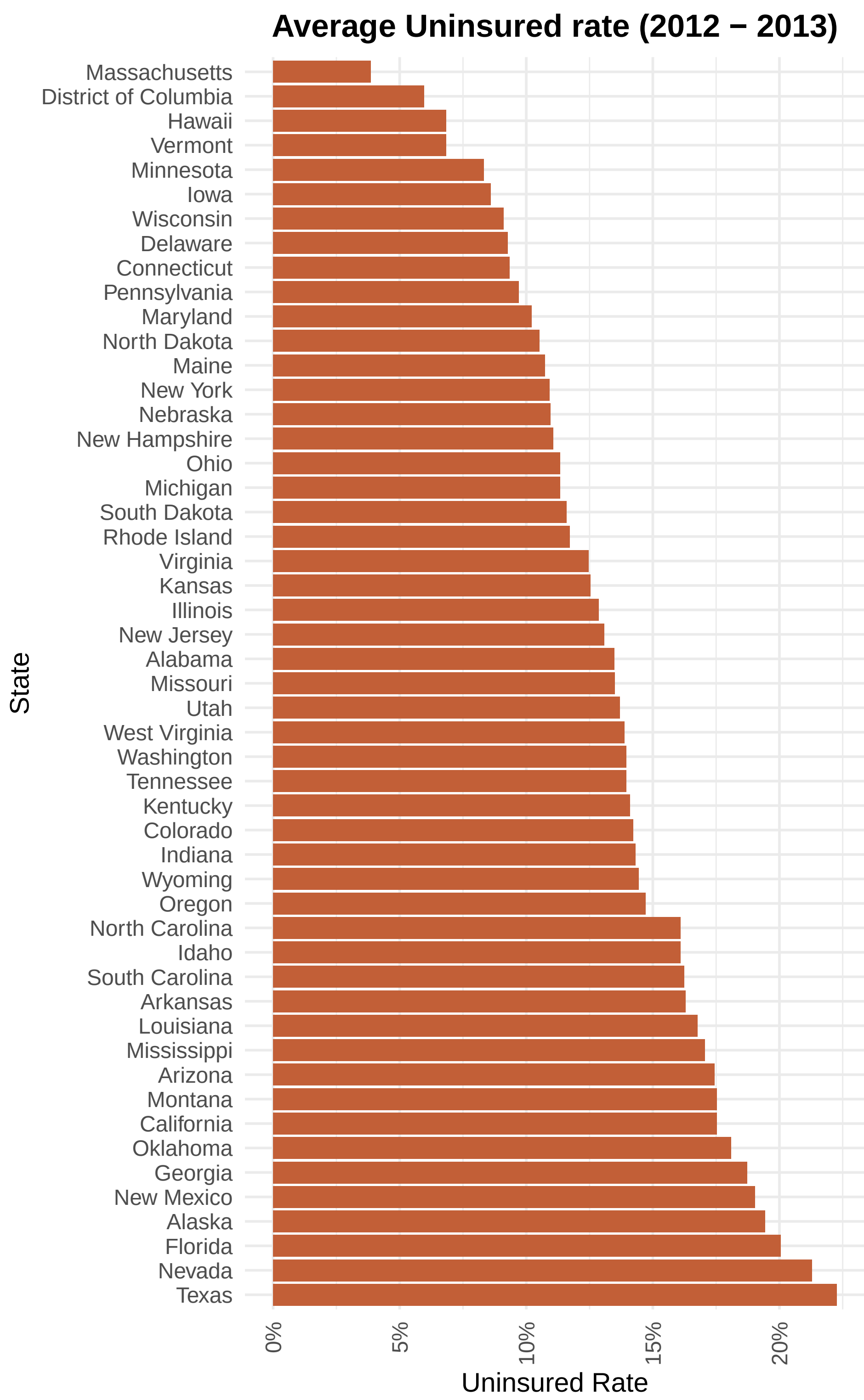
- **State**: Full name of state
- **Medicaid Expansion Adoption**: Date that the state adopted the Medicaid expansion, if it did so.
- **Year**: Year of observation.
- **Uninsured rate**: State uninsured rate in that year.

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?

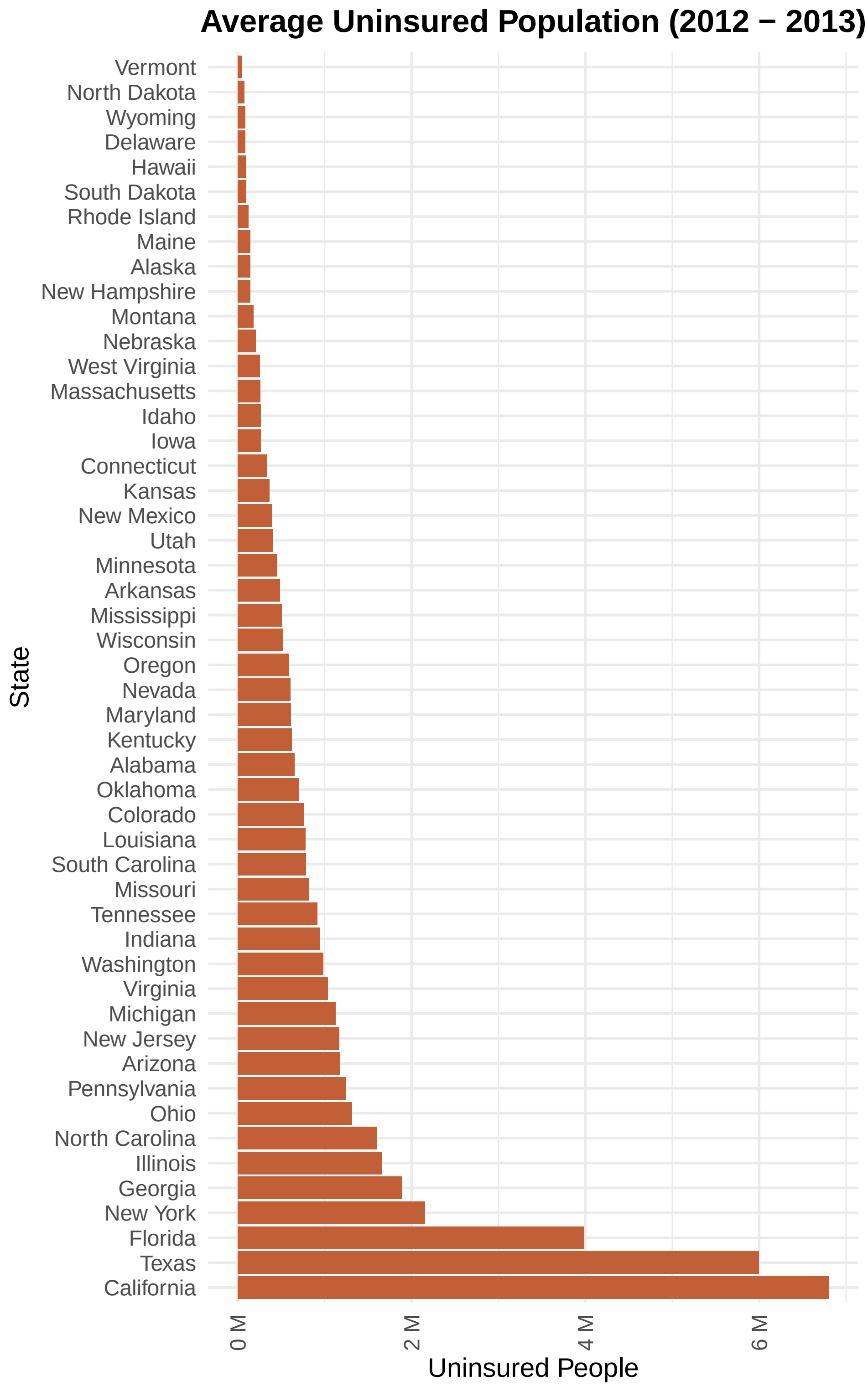
```r
# need to do something so this doesn't just add all the rates -
# averaging 2012 and 2013

medicaid_pre2014 <- medicaid_expansion %>%
  filter(year == 2012 | year == 2013) %>%
  group_by(State) %>%
  summarise(uninsured_rate = mean(uninsured_rate))

ggplot(medicaid_pre2014, aes(x = reorder(State,-uninsured_rate), uninsured_rate))+
  geom_bar(stat ="identity", fill = "#C25F37")+
  theme_minimal()+
  scale_y_continuous(labels = scales::percent)+
  coord_flip()+
  xlab("State")+
  ylab("Uninsured Rate")+
  labs(title = "Average Uninsured rate (2012 - 2013)")+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1),
        plot.title = element_text(hjust = 0.5, face = "bold"))
```



The above plot shows that Texas, Nevada, Florida, and Alaska had the highest rates of uninsured in their population prior to 2014 (using averaged 2012 - 2013 rate). Massachusetts, DC, Hawaii, and Vermont had the lowest rates of uninsured in their population.

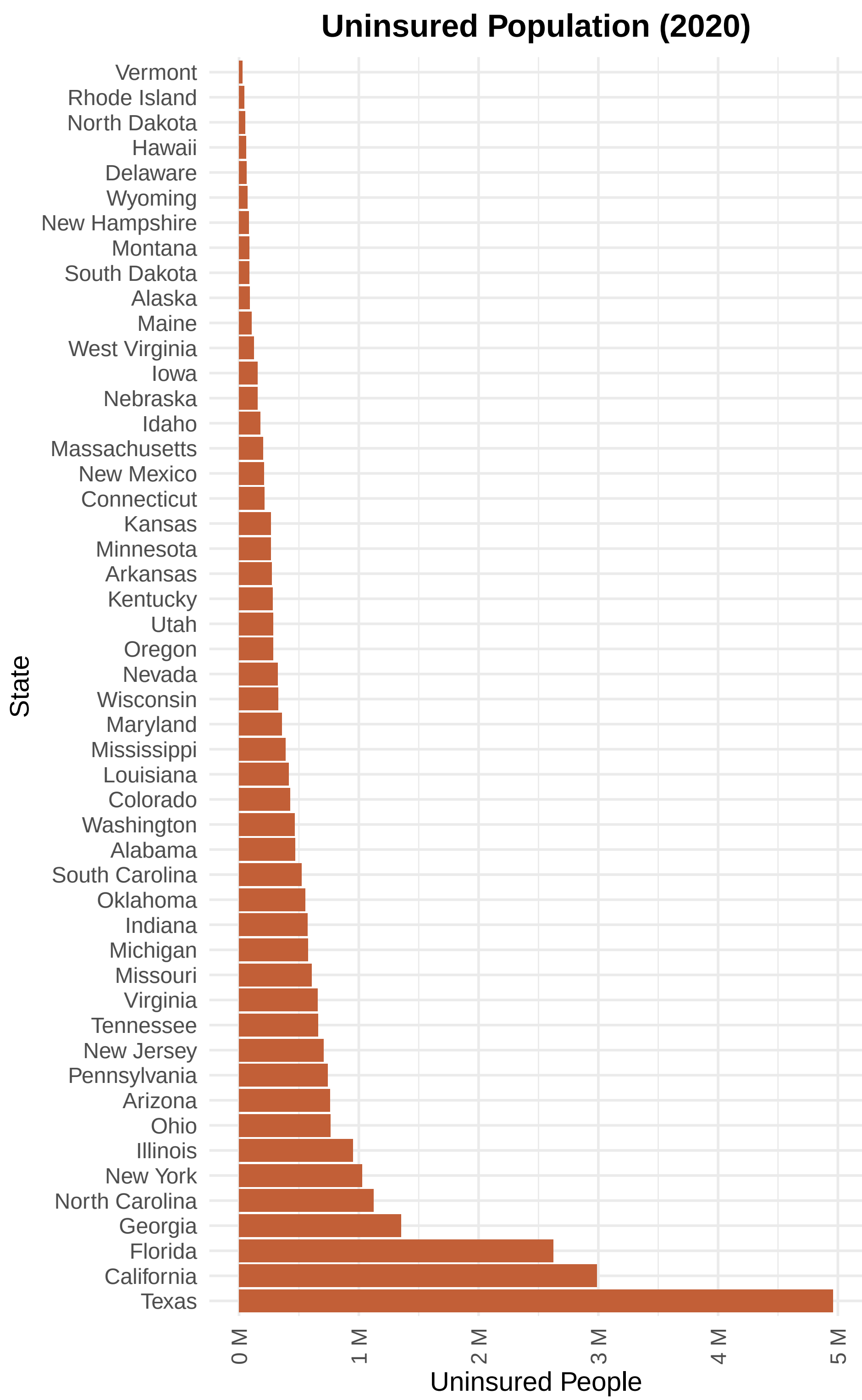
- Which states were home to most uninsured Americans prior to 2014? How about in the last year in the data set? **Note**: 2010 state population is provided as a variable to answer this question. In an actual study you would likely use population estimates over time, but to simplify you can assume these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

```
# need to do something so this doesn't just add all the rates -
# averaging 2012 and 2013

medicaid_pre2014 <- medicaid_expansion %>%
  filter(year == 2012 | year == 2013) %>%
  # No population data for DC
  filter(State != "District of Columbia") %>%
  mutate(uninsured_n = uninsured_rate*population) %>%
  group_by(State) %>%
  summarise(uninsured_n = mean(uninsured_n))

ggplot(medicaid_pre2014, aes(x = reorder(State,-uninsured_n), uninsured_n))+
  geom_bar(stat ="identity", fill = "#C25F37")+
  theme_minimal()+
  scale_y_continuous(labels = label_number(suffix = " M", scale = 1e-6))+
  coord_flip()+
  xlab("State")+
  ylab("Uninsured People")+
  labs(title = "Average Uninsured Population (2012 - 2013)")+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1),
        plot.title = element_text(hjust = 0.5, face = "bold"))
```



Prior to 2014, Califonia and Texas by far had the largest numbers of uninsured people, followed by Florida (using 2012 - 2013 average). Vermont, North Dakota, and Wyoming had the lowest. This makes sense due to large size difference of some of these states.

```r
# need to do something so this doesn't just add all the rates -
# averaging 2012 and 2013

medicaid_2020 <- medicaid_expansion %>%
  filter(year == 2020) %>%
  # No population data for DC
  filter(State != "District of Columbia") %>%
  mutate(uninsured_n = uninsured_rate*population) %>%
  group_by(State) %>%
  summarise(uninsured_n = mean(uninsured_n))

ggplot(medicaid_2020, aes(x = reorder(State,-uninsured_n), uninsured_n))+
  geom_bar(stat ="identity", fill = "#C25F37")+
  theme_minimal()+
  scale_y_continuous(labels = label_number(suffix = " M", scale = 1e-6))+
  coord_flip()+
  xlab("State")+
  ylab("Uninsured People")+
  labs(title = "Uninsured Population (2020)")+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1),
        plot.title = element_text(hjust = 0.5, face = "bold"))
```



In 2020, Texas now far surpassed California as having the highest number of uninsured people in the state. California has the second highest number, followed by Florida. Vermont still has the fewest uninsured people, however Rhode Island now has the second fewest.

Difference-in-Differences Estimation

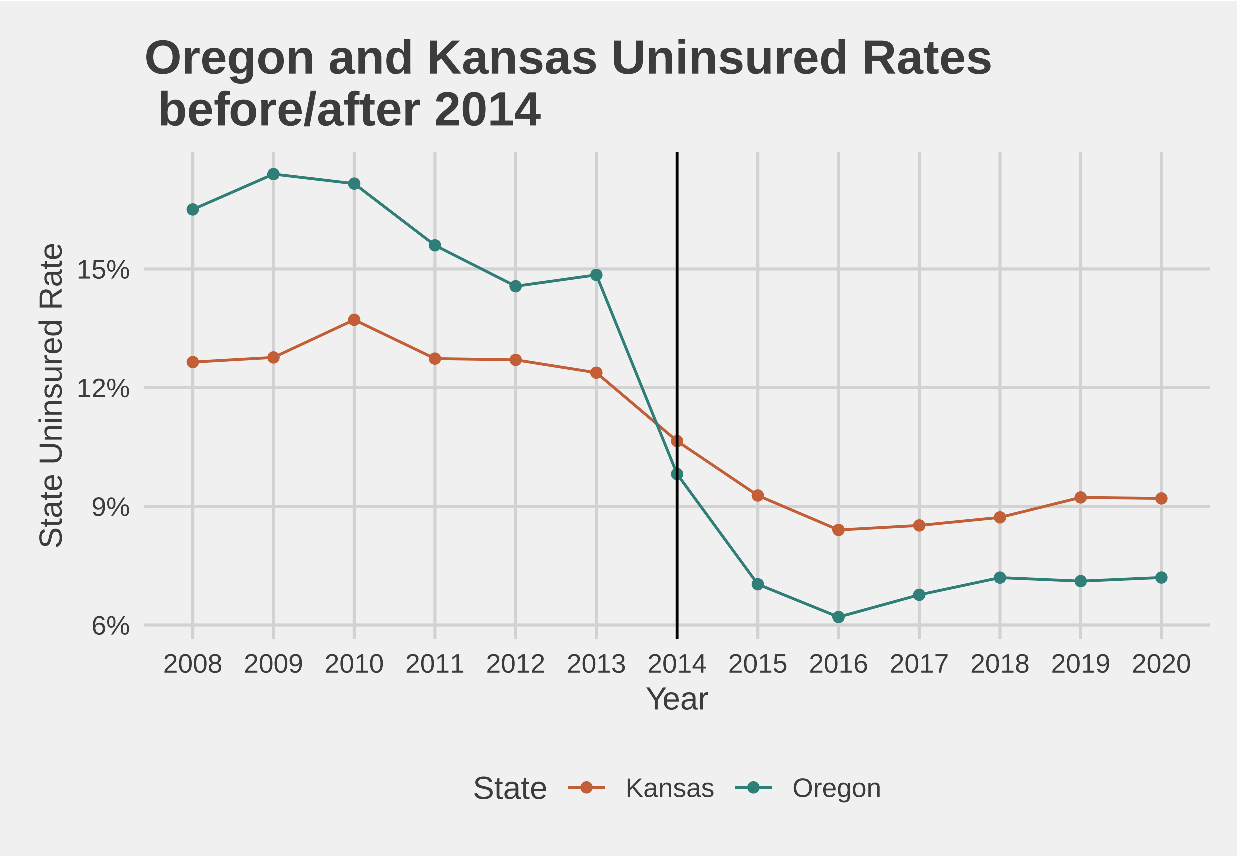
Estimate Model

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).

We chose Oregon as the state that adopted Medicaid expansion on January 1, 2014, and are comparing to Kansas.

```
years <- unique(medicaid_expansion$year)

ggplot(medicaid_expansion %>% filter(State == "Oregon" | State == "Kansas")) +
  geom_point(aes(x = year,
                 y = uninsured_rate,
                 color = State)) +
  geom_line(aes(x = year,
                y = uninsured_rate,
                color = State))  +
  scale_x_continuous(breaks = years)+
  scale_y_continuous(labels = scales::percent)+
  scale_color_manual(values = c("#C25F37", "#307E78"))+
  geom_vline(aes(xintercept = 2014)) +
  theme_fivethirtyeight() +
  theme(axis.title = element_text()) +
  ggtitle('Oregon and Kansas Uninsured Rates \n before/after 2014') +
  xlab('Year') +
  ylab('State Uninsured Rate')
```



Oregon and Kansas follow similar trends prior to 2014, so we can be satisfied the parallel trends assumption is met. Additionally, we are not concerned about the divergence that shows right at 2014, since the 2014 data is for the year and accounts for Medicaid expansion changes in Oregon (it is not data from the beginning of the year).

- 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

```
# Oregon adopted first day of 2014

# kansas-oregon
ko <- medicaid_expansion %>%
  filter(State == "Kansas" | State == "Oregon") %>%
  filter(year == 2013 | year == 2014)

# pre-treatment difference
pre_diff <- ko %>%
  filter(year == 2013) %>%
  select(State,
         uninsured_rate) %>%
  spread(State,
         uninsured_rate) %>%
  summarise(Oregon - Kansas)
```

```r
# post-treatment difference
post_diff <- ko %>%
  filter(year == 2014) %>%
  select(State,
         uninsured_rate) %>%
  spread(State,
         uninsured_rate) %>%
  summarise(Oregon - Kansas)

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

```
##   Oregon - Kansas
## 1        -0.03305
```

The difference in difference estimate is -0.03305, which means that the treatment effect for Medicaid expansion in Oregon is a 3.3% reduction in the uninsured rate.

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**: States are not necessarily similar in demographics, economics, or state policies that can have an influence on uninsured rates. Additionally, in this data we don't have variables for these various aspects of the states.
- What are the strengths and weaknesses of using the parallel trends assumption in difference-in-differences estimates?
- **Answer**: In this assumption we assume that the treatment and control groups (in this case, states) have followed a similar pattern in the outcome over time and that they would likely continue to do so in the absence of an intervention. One weakness is that we are assuming the treatment and control groups have similar traits/aspects, but in reality there might be different relevant state factors that can influence the outcome that aren't captured when looking at the parallel trends. Similarly are also assuming that there were no other interventions or large events that could influence these numbers (like natural disasters or pandemics, for example), which isn't always the case. Another weakness is that there is no clear statistical test to "prove" trends are similar enough to meet this assumption. However, a strength is that this approach can help with controlling for unobserved factors that may affect the outcome for both the treatment and control groups.

Synthetic Control

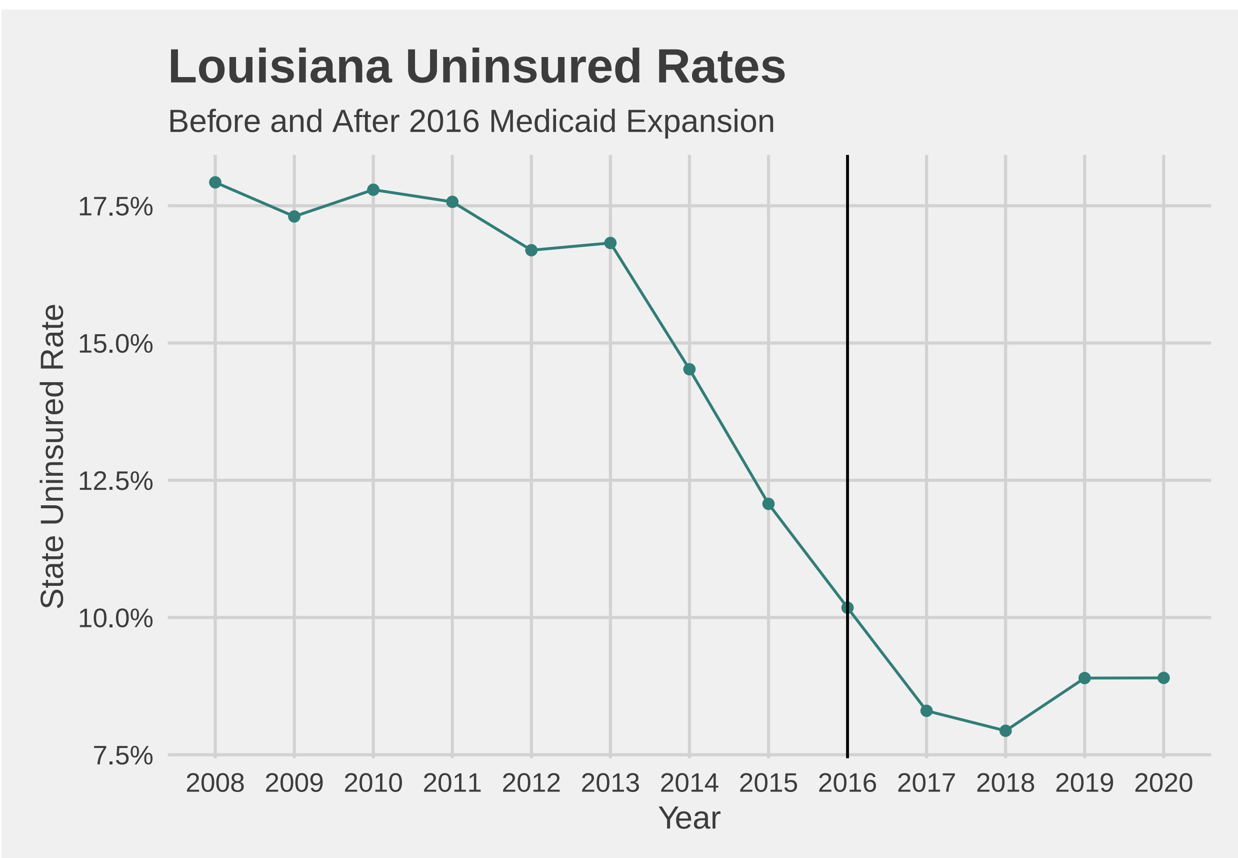
Estimate Synthetic Control

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.

```
###Choosing and plotting Louisiana. Chose Louisiana since they adopted 2016, which leaves adequate post-
### There was a dip before 2016, which may pose some analysis issues.
  ggplot(data = medicaid_expansion[which(medicaid_expansion$State == 'Louisiana'),]) +
  geom_point(aes(x = year,
                 y = uninsured_rate,
                 color = State), color = "#307E78") +
  geom_line(aes(x = year,
                y = uninsured_rate,
                ), color = "#307E78")  +
  geom_vline(aes(xintercept = 2016)) +
  scale_x_continuous(breaks = years)+
  scale_y_continuous(labels = scales::percent)+
  theme_fivethirtyeight() +
  theme(axis.title = element_text()) +
  labs(title = 'Louisiana Uninsured Rates',
       subtitle = "Before and After 2016 Medicaid Expansion") +
  xlab('Year') +
  ylab('State Uninsured Rate')
```

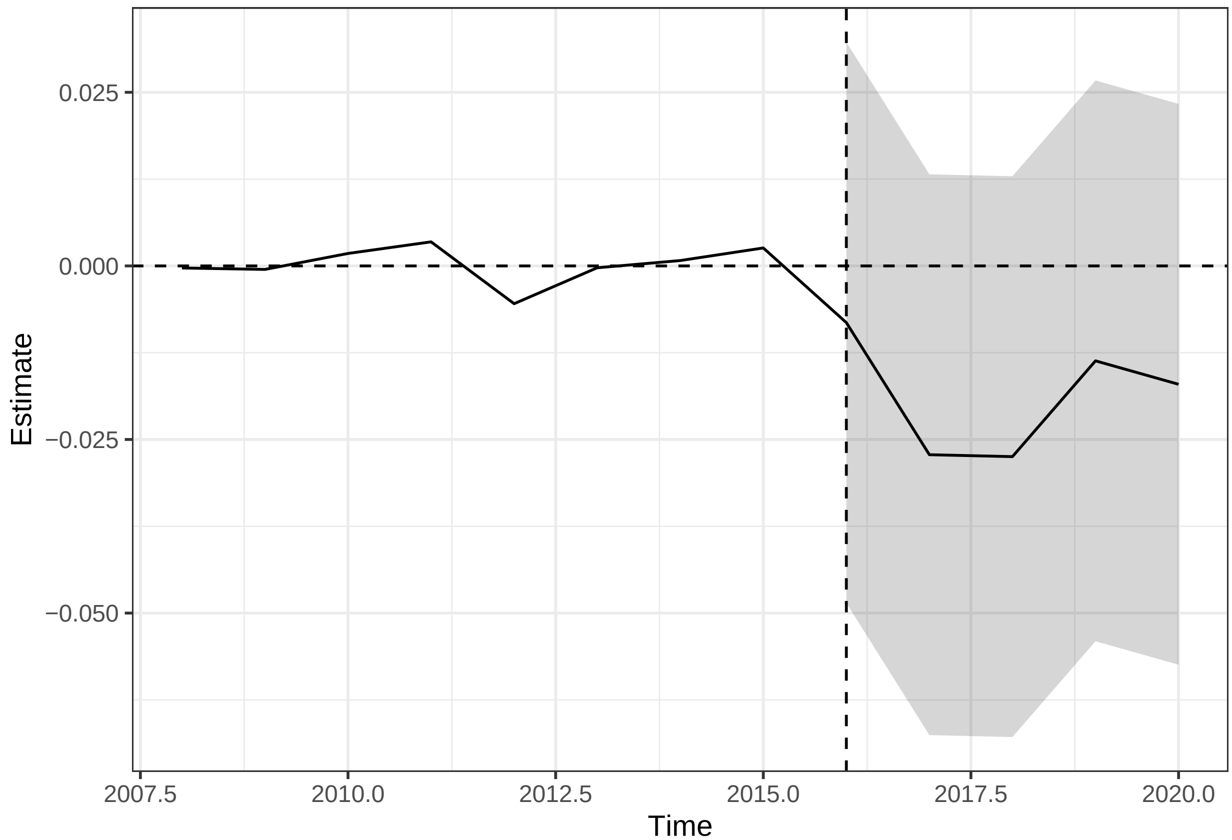


```r
# remove DC since it is a unique case and we don't have population data
medicaid_expansion <-  medicaid_expansion %>%
  filter(State != "District of Columbia")

medicaid_expansion$Date_Adopted[is.na(medicaid_expansion$Date_Adopted)] <- 0

###Dropping rows with missing values for date adopted. Since this is a synthetic control, we don't nece

medicaid_expansion_syn <- medicaid_expansion %>%
  filter(Date_Adopted != 0) %>%
  mutate(la_treated = ifelse(State == "Louisiana" & year > 2015, 1, 0))
 # Since focus state is Louisiana, treatment is Louisiana after adoption in 2016.

# In case needed, adding another treatment variable for all years in a state after its adoption date
  medicaid_expansion_syn$allstate_treated <- 0
  for (i in 1:nrow( medicaid_expansion_syn)) {
    if ( medicaid_expansion_syn[i,]$year >  medicaid_expansion_syn[i,]$Date_Adopted) {
      medicaid_expansion_syn[i,]$allstate_treated <- 1
    }
    else {  medicaid_expansion_syn[i,]$allstate_treated <- 0
    }
  }

# It's unclear to me from these instructions if I should be using the treatment for a specific state, o

  syn_la <- augsynth(uninsured_rate ~ la_treated, State, year, medicaid_expansion_syn,
                  progfunc = "None", scm = T)
  summary(syn_la)
```

```
##
## 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):  -0.019   ( 0.186 )
## L2 Imbalance: 0.007
## Percent improvement from uniform weights: 94%
##
## Avg Estimated Bias: NA
##
## Inference type: Conformal inference
##
##  Time Estimate 95% CI Lower Bound 95% CI Upper Bound p Value
##  2016   -0.008             -0.049              0.032   0.333
##  2017   -0.027             -0.068              0.013   0.222
##  2018   -0.027             -0.068              0.013   0.111
##  2019   -0.014             -0.054              0.027   0.111
##  2020   -0.017             -0.057              0.023   0.111
```

- Average ATT Estimate (p Value for Joint Null): -0.019 (0.186)
- L2 Imbalance: 0.007245917
- Percent improvement from uniform weights: 94%
- Avg Estimated Bias: NA

```r
#Ran with all state treatment variable, but realized this is "multisynth" b/c multiple time conditions.

#We can use the built in plot function to see how Louisina did relative to synthetic Lousiana:

plot(syn_la)
```



- 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.

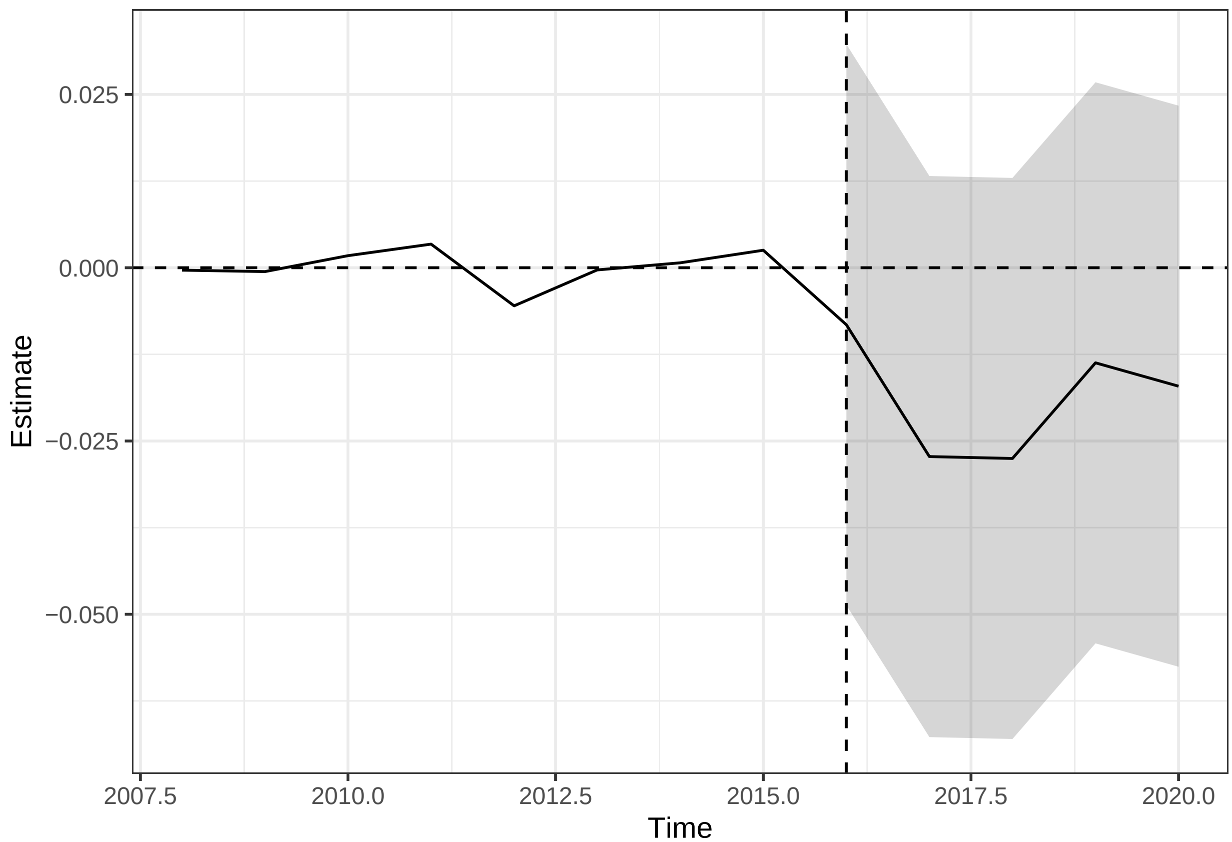
```r
ridge_syn_la <- augsynth(uninsured_rate ~ la_treated, State, year, medicaid_expansion_syn,
                      progfunc = "ridge", scm = T)
summary(ridge_syn_la)
```

```
##
## 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):  -0.019   ( 0.172 )
## L2 Imbalance: 0.007
## Percent improvement from uniform weights: 94.1%
##
## Avg Estimated Bias: 0.000
##
## Inference type: Conformal inference
##
```

```
##   Time Estimate 95% CI Lower Bound 95% CI Upper Bound p Value
##   2016   -0.008             -0.049              0.032   0.333
##   2017   -0.027             -0.068              0.013   0.222
##   2018   -0.028             -0.068              0.013   0.111
##   2019   -0.014             -0.054              0.027   0.111
##   2020   -0.017             -0.058              0.023   0.111
```

- Average ATT Estimate (p Value for Joint Null): -0.019 (0.172)
- L2 Imbalance: 0.007228185
- Percent improvement from uniform weights: 94.1%
- Avg Estimated Bias: 0.000

```
plot(ridge_syn_la)
```

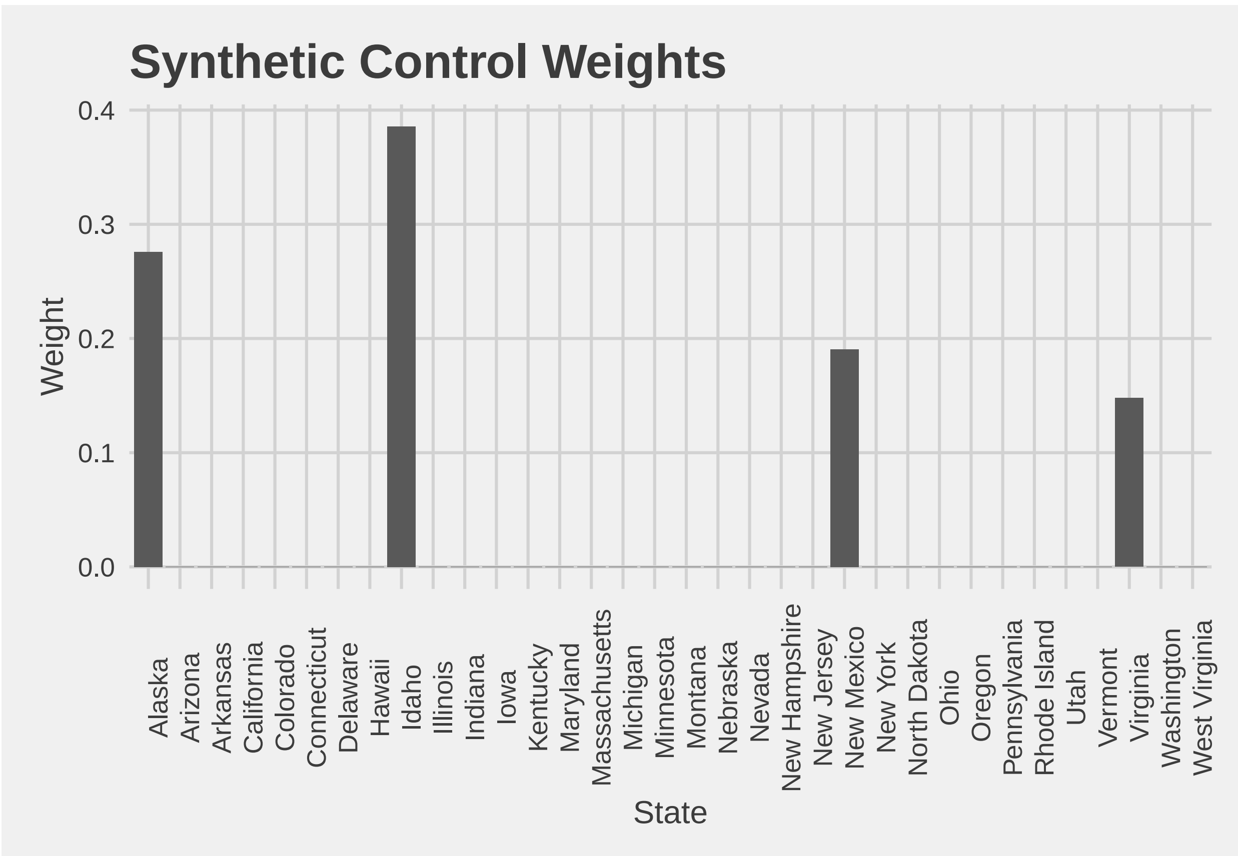


Overall, the augmentation provided little improvement over the analysis with the non-augmented synthetic control.

- Plot barplots to visualize the weights of the donors.

```
# barplots of weights

data.frame(syn_la$weights) %>%
  # change index to a column
  tibble::rownames_to_column('State') %>%
  ggplot() +
```

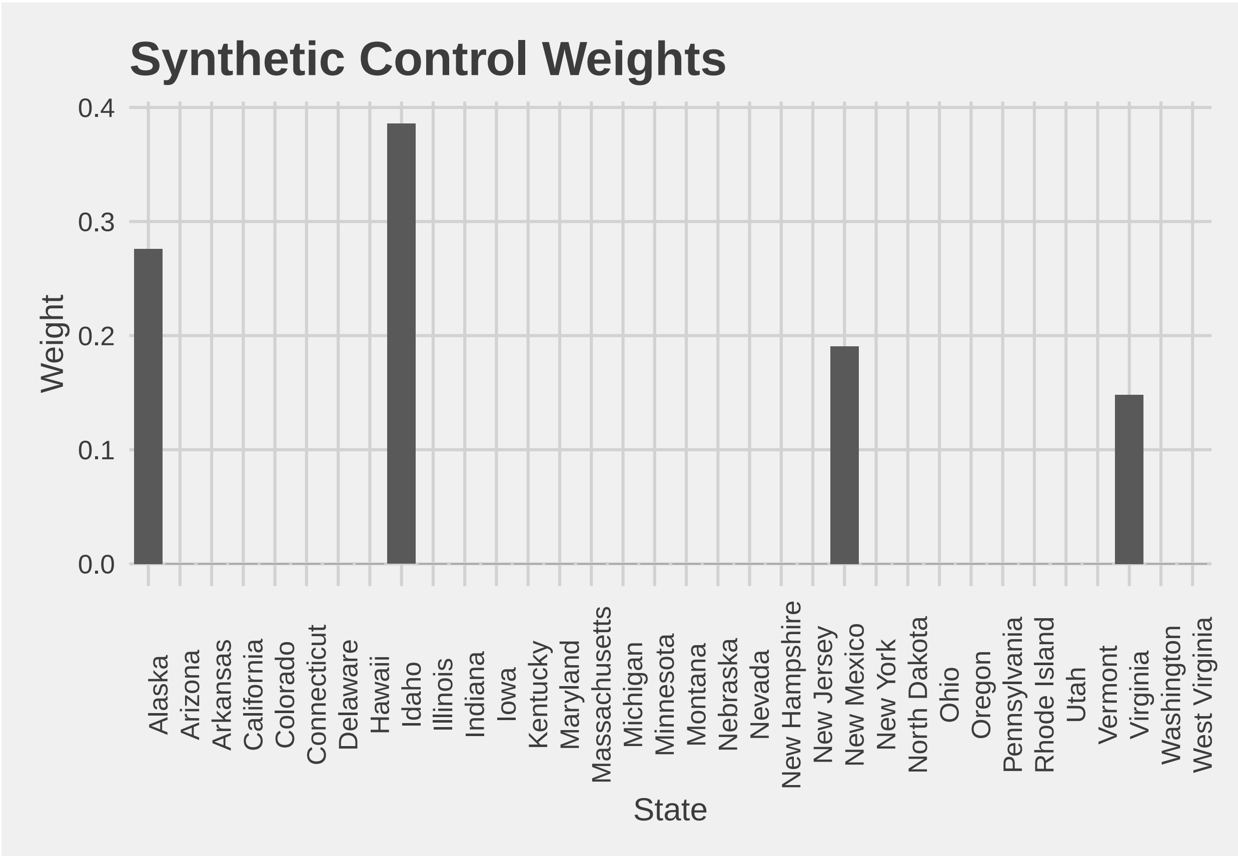
```r
  # stat = identity to take the literal value instead of a count for geom_bar()
  geom_bar(aes(x = State,
               y = syn_la.weights),
           stat = 'identity') +
  theme_fivethirtyeight() +
  theme(axis.title = element_text(),
        axis.text.x = element_text(angle = 90)) +
  ggtitle('Synthetic Control Weights') +
  xlab('State') +
  ylab('Weight')
```



- For the non augmented analysis, the weights come from: Alaska, Idaho, New Mexico, Virginia

```r
data.frame(ridge_syn_la$weights) %>%
  # change index to a column
  tibble::rownames_to_column('State') %>%
  ggplot() +
  # stat = identity to take the literal value instead of a count for geom_bar()
  geom_bar(aes(x = State,
               y = ridge_syn_la.weights),
           stat = 'identity') +
  theme_fivethirtyeight() +
  theme(axis.title = element_text(),
        axis.text.x = element_text(angle = 90)) +
  ggtitle('Synthetic Control Weights') +
  xlab('State') +
```

```
  ylab('Weight')
```



Discussion Questions

- What are the advantages and disadvantages of synthetic control compared to difference-in-differences estimators?
- **Answer**: Synthetic controls can help to create a better comparison group that would otherwise be hard to come by. This approach can help account for confounders that change over time, and is a more systematic way to select a control. However, synthetic controls also have disadvantes. The results are harder to interpret than a traditional difference in difference analysis, and can be sensitive to model specification (such as covariate selection).
- 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**: Using negative weights makes it harder to interpret. In the case of this analysis, the augmented version did not provide much benefit, as there was only a tiny improvement seen. However, there are some situations where this approach may provide more of a benefit and is worth considering, such as when there is significant imbalance in the pre-treatment period. The severity of the imbalance and the difficulty of interpretation should be caregfully considered for an analysis that is choosing between augmentation or no augmentation.

Staggered Adoption Synthetic Control

Estimate Multisynth

Do the following:

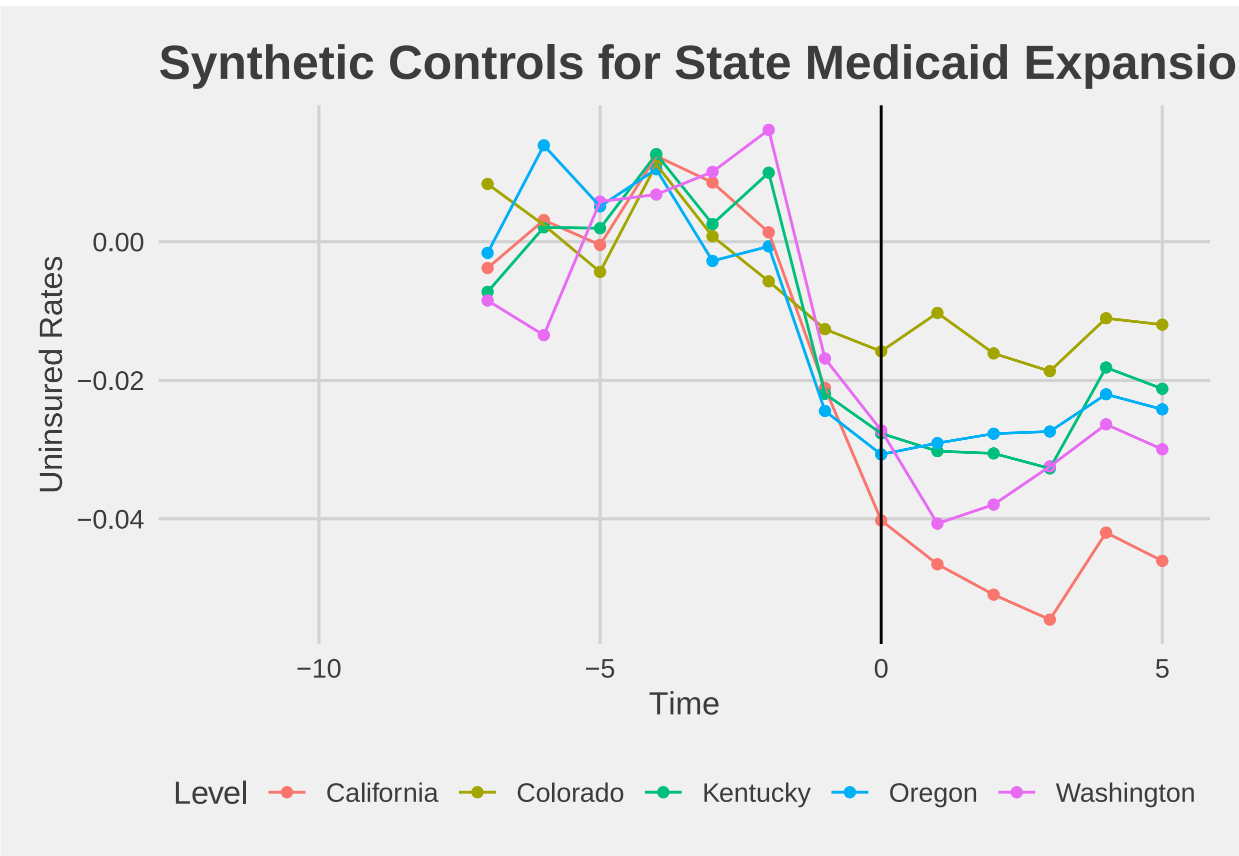
- Estimate a multisynth model that treats each state individually. Choose a fraction of states that you can fit on a plot and examine their treatment effects.

```
# multisynth model states

ppool_syn <- multisynth( uninsured_rate ~ allstate_treated, State, year, medicaid_expansion_syn, n_leads
ppool_syn_summ <- summary(ppool_syn)

ppool_syn_summ
```

```
##
## Call:
## multisynth(form = uninsured_rate ~ allstate_treated, unit = State,
##     time = year, data = medicaid_expansion_syn, n_leads = 10)
##
## Average ATT Estimate (Std. Error): -0.013  (0.011)
##
## Global L2 Imbalance: 0.003
## Scaled Global L2 Imbalance: 0.149
## Percent improvement from uniform global weights: 85.1
##
## Individual L2 Imbalance: 0.008
## Scaled Individual L2 Imbalance: 0.175
## Percent improvement from uniform individual weights: 82.5
##
##  Time Since Treatment   Level     Estimate  Std.Error lower_bound upper_bound
##                     0 Average -0.009559790 0.00781703 -0.02123382 0.006790910
##                     1 Average -0.016079986 0.01345043 -0.03397399 0.011010472
##                     2 Average -0.016265890 0.01163574 -0.03184703 0.007367055
##                     3 Average -0.016289735 0.01057833 -0.03069480 0.005497133
##                     4 Average -0.007341691 0.01041568 -0.02215642 0.014038164
##                     5 Average -0.010362574 0.01117834 -0.02604183 0.012179039
```

- Average ATT Estimate (Std. Error): -0.013 (0.011)
- Global L2 Imbalance: 0.003
- Scaled Global L2 Imbalance: 0.149
- Percent improvement from uniform global weights: 85.1
- Individual L2 Imbalance: 0.008
- Scaled Individual L2 Imbalance: 0.175
- Percent improvement from uniform individual weights: 82.5

```r
# select some states for the plot
states <- c("Washington", "Colorado", "California", "Oregon", "Kentucky")

ppool_syn_summ$att %>%
  filter(Level %in% states) %>%
  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 State Medicaid Expansion') +
  xlab('Time') +
  ylab('Uninsured Rates')
```



- 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.

```r
# multisynth model time cohorts

ppool_syn_time <- multisynth(uninsured_rate ~ allstate_treated, State, year, medicaid_expansion_syn, n_
```

```r
ppool_syn_time_summ <- summary(ppool_syn_time)
ppool_syn_time_summ
```

```
##
## Call:
## multisynth(form = uninsured_rate ~ allstate_treated, unit = State,
##     time = year, data = medicaid_expansion_syn, n_leads = 10,
##     time_cohort = TRUE)
##
## Average ATT Estimate (Std. Error): -0.012  (0.009)
##
## Global L2 Imbalance: 0.020
## Scaled Global L2 Imbalance: 0.138
## Percent improvement from uniform global weights: 86.2
##
## Individual L2 Imbalance: 0.053
## Scaled Individual L2 Imbalance: 0.136
## Percent improvement from uniform individual weights: 86.4
##
##  Time Since Treatment   Level     Estimate   Std.Error lower_bound upper_bound
##                     0 Average -0.008232467 0.006662747 -0.01908919 0.005902492
##                     1 Average -0.014413689 0.011164777 -0.03072973 0.009063390
##                     2 Average -0.014991845 0.009595911 -0.02900001 0.004978931
##                     3 Average -0.015346953 0.008807149 -0.02807785 0.003190512
##                     4 Average -0.005666535 0.008650598 -0.01890672 0.012994456
##                     5 Average -0.009498406 0.009788497 -0.02407709 0.011509913
```

- Average ATT Estimate (Std. Error): -0.012 (0.009)
- Global L2 Imbalance: 0.020
- Scaled Global L2 Imbalance: 0.138
- Percent improvement from uniform global weights: 86.2
- Individual L2 Imbalance: 0.053
- Scaled Individual L2 Imbalance: 0.136
- Percent improvement from uniform individual weights: 86.4

Overall, the time cohort provided some improvement over initial multisynth model.

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 plot above that shows 5 states (California, Colorado, Kentucky, Oregon, and Washington) shows a clear difference in effect size for different states. In the pre-treatment period, the states uninsured rates are much closer, however in the treatment period we see a much wider variation in uninsured rates even though all states saw improvements.

- Do you see evidence for the idea that early adopters of Medicaid expansion enjoyed a larger decrease in the uninsured population?
- **Answer**: This seems possible, however the analysis we did does not directly answer this question. Effects seem to vary substantially by state, regardless of whether the state was an early or late adopter.

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 help to account for unobserved differences and can be somewhat robust to confounding factors that are present at the aggregate level where it's hard to account for this heterogeneity.
- What role does selection into treatment play in DiD/synthetic control versus regression discontinuity? When would we want to use either method?
- **Answer**: Difference in difference and synthetic controls are used for observational studies where "selection" into treatment is frequently based on a policy or some outside large scale factor. We would want to use this approch to assess outcomes in situations such as this. Alternately, regression discontinuity is better suited for when selection into treatment is based on some continuous variable where we can compare effects for those right above and below some cutoff (for example, test scores or income cutoffs).