The Impact of Recreational Marijuana Legalization on Binge Drinking in the U.S.: Evidence from a Difference-in-Differences Approach

2025-03-30

Setup

```
packages <- c("did", "ggplot2", "dplyr", "readr", "stargazer", "showtext")

#install packages if not installed
installed <- packages %in% rownames(installed.packages())
if (any(!installed)) {
  install.packages(packages[!installed])
}

#load packages
lapply(packages, library, character.only = TRUE)

theme_serif <- function(base_size = 14) {
  theme_minimal(base_family = "serif", base_size = base_size) +
    theme(
      text = element_text(family = "serif", size = base_size),
      legend.text = element_text(size = base_size * 0.9),
      axis.title = element_text(size = base_size),
      plot.title = element_text(size = base_size * 1.3, face = "bold"),
      plot.subtitle = element_text(size = base_size)
    )
}
```

Importing Data

```
url <- 'https://docs.google.com/spreadsheets/d/e/2PACX-1vS-Qiz6Osc9I39bvo_94Mo1GvYe_ghAVboqYvR5xu6ZSPDI-ESb6ECC
XgocUP3ALVQ3oxi_n3wDihQL/pub?gid=1277276287&single=true&output=csv'
data <- read.csv(url)
head(data)
```

```
##   Year State_Abbr      State Binge_Drinking_Prevalence Legalized Bachelors_Rate
## 1 2011         AL    Alabama                      13.7         0           22.3
## 2 2011         AK     Alaska                      20.8         1           26.4
## 3 2011         AZ    Arizona                      17.6         0           26.6
## 4 2011         AR   Arkansas                      14.1         0           20.3
## 5 2011         CA California                      18.6         1           30.3
## 6 2011         CO   Colorado                      20.1         1           36.7
##   Median_Age Urbanization_Rate Legalization_Year    G State_ID
## 1       38.1           0.58304                NA    0        1
## 2       33.9           0.59651              2014 2014        2
## 3       36.2           0.88484                NA    0        3
## 4       37.5           0.54736                NA    0        4
## 5       35.4           0.93649              2016 2016        5
## 6       36.2           0.83236              2012 2012        6
```

```
summary(data)
```

```
##       Year       State_Abbr           State           Binge_Drinking_Prevalence
##  Min.   :2011   Length:659         Length:659         Min.   : 9.60
##  1st Qu.:2014   Class :character   Class :character   1st Qu.:14.79
##  Median :2017   Mode  :character   Mode  :character   Median :16.50
##  Mean   :2017                                         Mean   :16.71
##  3rd Qu.:2020                                         3rd Qu.:18.30
##  Max.   :2023                                         Max.   :27.20
##
##    Legalized      Bachelors_Rate   Median_Age    Urbanization_Rate
##  Min.   :0.0000   Min.   :18.5   Min.   :29.60   Min.   :0.3117
##  1st Qu.:0.0000   1st Qu.:27.2   1st Qu.:37.05   1st Qu.:0.6161
##  Median :0.0000   Median :30.9   Median :38.40   Median :0.7209
##  Mean   :0.1973   Mean   :31.8   Mean   :38.42   Mean   :0.7224
##  3rd Qu.:0.0000   3rd Qu.:35.4   3rd Qu.:39.65   3rd Qu.:0.8551
##  Max.   :1.0000   Max.   :65.9   Max.   :45.10   Max.   :1.0000
##
##  Legalization_Year       G             State_ID
##  Min.   :2012      Min.   :   0.0   Min.   : 1.00
##  1st Qu.:2014      1st Qu.:   0.0   1st Qu.:13.00
##  Median :2015      Median :   0.0   Median :26.00
##  Mean   :2015      Mean   : 397.4   Mean   :26.01
##  3rd Qu.:2016      3rd Qu.:   0.0   3rd Qu.:39.00
##  Max.   :2016      Max.   :2016.0   Max.   :51.00
##  NA's   :529
```

Latex Summary Statistics Table

```
#subset to key variables only
summary_vars <- data[c("Binge_Drinking_Prevalence", "Legalized", "Bachelors_Rate", "Median_Age", "Urbanization_
Rate")]

#create summary table in LaTeX format (default)
stargazer(summary_vars, type = "latex", title = "Summary Statistics", digits = 2, summary.stat = c("min", "mea
n", "sd", "max"))
```

```
##
## % Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmai
l.com
## % Date and time: Wed, Apr 23, 2025 - 14:06:50
## \begin{table}[!htbp] \centering
##   \caption{Summary Statistics}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lcccc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
## Statistic & \multicolumn{1}{c}{Min} & \multicolumn{1}{c}{Mean} & \multicolumn{1}{c}{St. Dev.} & \multicolumn
{1}{c}{Max} \\
## \hline \\[-1.8ex]
## Binge\_Drinking\_Prevalence & 9.60 & 16.71 & 3.03 & 27.20 \\
## Legalized & 0 & 0.20 & 0.40 & 1 \\
## Bachelors\_Rate & 18.50 & 31.80 & 6.79 & 65.90 \\
## Median\_Age & 29.60 & 38.42 & 2.41 & 45.10 \\
## Urbanization\_Rate & 0.31 & 0.72 & 0.15 & 1.00 \\
## \hline \\[-1.8ex]
## \end{tabular}
## \end{table}
```

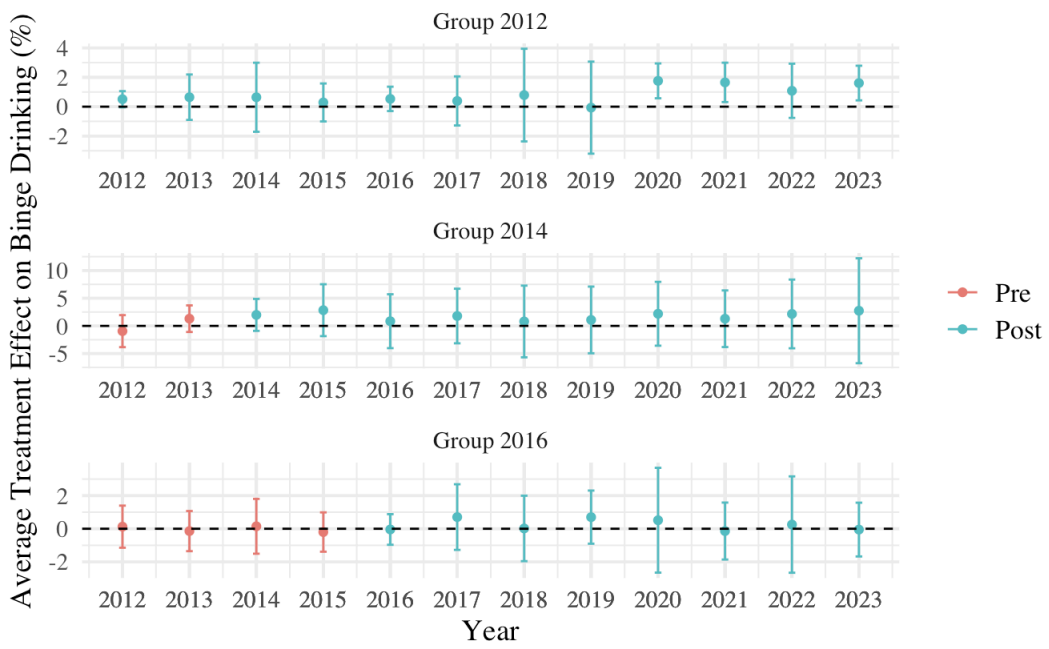
Analysis

Event Study Estimates of Binge Drinking Prevalence by Legalization Cohort

```
att_did <- att_gt(
  yname = "Binge_Drinking_Prevalence",
  tname = "Year",
  idname = "State_ID",
  gname = "G",
  xformla = ~ Bachelors_Rate + Median_Age + Urbanization_Rate,
  data = data,
  est_method = "dr"
)

summary(att_did)
```

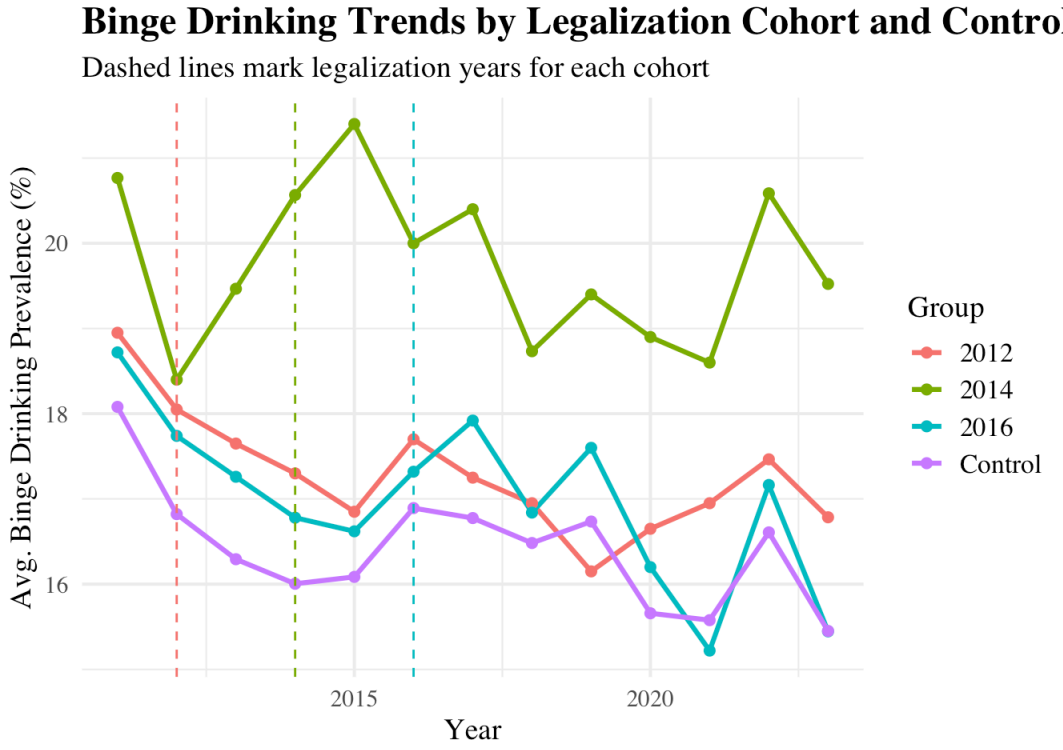
```
##
## Call:
## att_gt(yname = "Binge_Drinking_Prevalence", tname = "Year", idname = "State_ID",
##     gname = "G", xformla = ~Bachelors_Rate + Median_Age + Urbanization_Rate,
##     data = data, est_method = "dr")
##
## Reference: Callaway, Brantly and Pedro H.C. Sant'Anna.  "Difference-in-Differences with Multiple Time Period
s." Journal of Econometrics, Vol. 225, No. 2, pp. 200-230, 2021. <https://doi.org/10.1016/j.jeconom.2020.12.001
>, <https://arxiv.org/abs/1803.09015>
##
## Group-Time Average Treatment Effects:
##  Group Time ATT(g,t) Std. Error [95% Simult.  Conf. Band]
##   2012 2012   0.5111     0.2300       -0.0385      1.0608
##   2012 2013   0.6463     0.6469       -0.8999      2.1925
##   2012 2014   0.6429     0.9835       -1.7076      2.9935
##   2012 2015   0.2889     0.5407       -1.0033      1.5812
##   2012 2016   0.5320     0.3465       -0.2960      1.3601
##   2012 2017   0.3912     0.6985       -1.2783      2.0607
##   2012 2018   0.7913     1.3201       -2.3639      3.9466
##   2012 2019  -0.0632     1.3115       -3.1977      3.0713
##   2012 2020   1.7594     0.4965        0.5727      2.9462 *
##   2012 2021   1.6551     0.5606        0.3153      2.9949 *
##   2012 2022   1.0835     0.7724       -0.7626      2.9296
##   2012 2023   1.6094     0.4941        0.4285      2.7903 *
##   2014 2012  -0.9431     1.2111       -3.8376      1.9515
##   2014 2013   1.3045     1.0055       -1.0987      3.7077
##   2014 2014   1.9879     1.2093       -0.9025      4.8784
##   2014 2015   2.8387     1.9624       -1.8515      7.5289
##   2014 2016   0.8468     2.0372       -4.0223      5.7159
##   2014 2017   1.7910     2.0644       -3.1430      6.7250
##   2014 2018   0.8031     2.7129       -5.6810      7.2873
##   2014 2019   1.0778     2.5220       -4.9499      7.1056
##   2014 2020   2.1930     2.4137       -3.5760      7.9620
##   2014 2021   1.2927     2.1414       -3.8254      6.4108
##   2014 2022   2.1638     2.5995       -4.0492      8.3767
##   2014 2023   2.7419     3.9673       -6.7402     12.2240
##   2016 2012   0.1248     0.5319       -1.1465      1.3962
##   2016 2013  -0.1422     0.5067       -1.3532      1.0688
##   2016 2014   0.1464     0.6920       -1.5075      1.8003
##   2016 2015  -0.2007     0.4966       -1.3875      0.9862
##   2016 2016  -0.0449     0.3862       -0.9680      0.8783
##   2016 2017   0.7029     0.8287       -1.2778      2.6836
##   2016 2018   0.0182     0.8266       -1.9574      1.9939
##   2016 2019   0.6997     0.6709       -0.9037      2.3031
##   2016 2020   0.5131     1.3237       -2.6508      3.6769
##   2016 2021  -0.1407     0.7190       -1.8592      1.5778
##   2016 2022   0.2516     1.2165       -2.6559      3.1591
##   2016 2023  -0.0524     0.6791       -1.6754      1.5706
## ---
## Signif. codes: `*' confidence band does not cover 0
##
## P-value for pre-test of parallel trends assumption:  0.75395
## Control Group:  Never Treated,  Anticipation Periods:  0
## Estimation Method:  Doubly Robust
```

```
ggdid(att_did) + theme_serif() +
  labs(
    title = "Event Study Estimates by Legalization Year",
    subtitle = "ATT by calendar year for 2012, 2014, and 2016 legalization groups",
    x = "Year",
    y = "Average Treatment Effect on Binge Drinking (%)"
  )
```


Event Study Estimates by Legalization Year
ATT by calendar year for 2012, 2014, and 2016 legalization groups
Group 2012
Group 2014
Group 2016
Average Treatment Effect on Binge Drinking (%)
Year
Pre
Post
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023

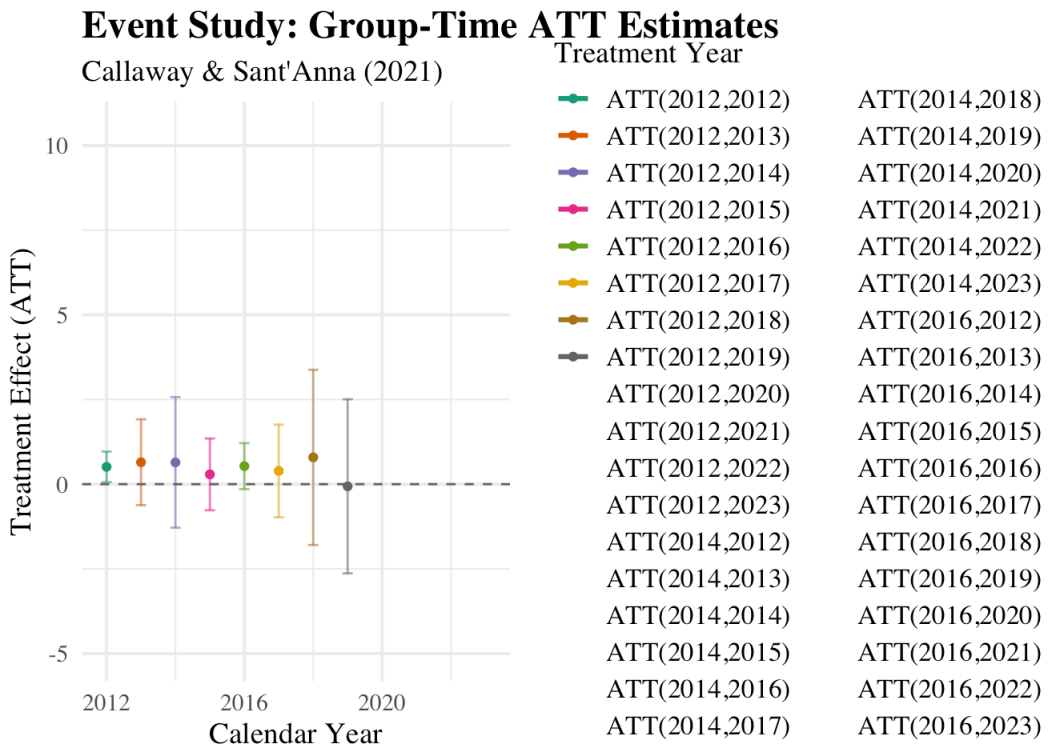
Trends in Binge Drinking Prevalence by Legalization Timing

```
#treated states by year (2012, 2014, 2016)
treated_data <- data %>%
  filter(!is.na(Legalization_Year)) %>%
  mutate(Cohort = as.factor(Legalization_Year))

#control group
control_data <- data %>%
  filter(is.na(Legalization_Year)) %>%
  mutate(Cohort = "Control")

#combine groups
combined_data <- bind_rows(treated_data, control_data)

#average binge drinking by group and year
avg_trends <- combined_data %>%
  group_by(Year, Cohort) %>%
  summarise(Avg_Binge = mean(Binge_Drinking_Prevalence, na.rm = TRUE), .groups = "drop")

#get legalization years per cohort (excluding control)
cohort_lines <- treated_data %>%
  distinct(Cohort, Legalization_Year)

#plot
ggplot(avg_trends, aes(x = Year, y = Avg_Binge, color = Cohort)) +
  geom_line(size = 1.1) +
  geom_point(size = 2) +
  #add vertical line for each treated cohort
  geom_vline(data = cohort_lines, aes(xintercept = Legalization_Year, color = Cohort),
             linetype = "dashed", show.legend = FALSE) +
  labs(
    title = "Binge Drinking Trends by Legalization Cohort and Control Group",
    subtitle = "Dashed lines mark legalization years for each cohort",
    x = "Year",
    y = "Avg. Binge Drinking Prevalence (%)",
    color = "Group"
  ) +
  theme_minimal(base_size = 13) +
  theme(legend.position = "bottom") + theme_serif()
```



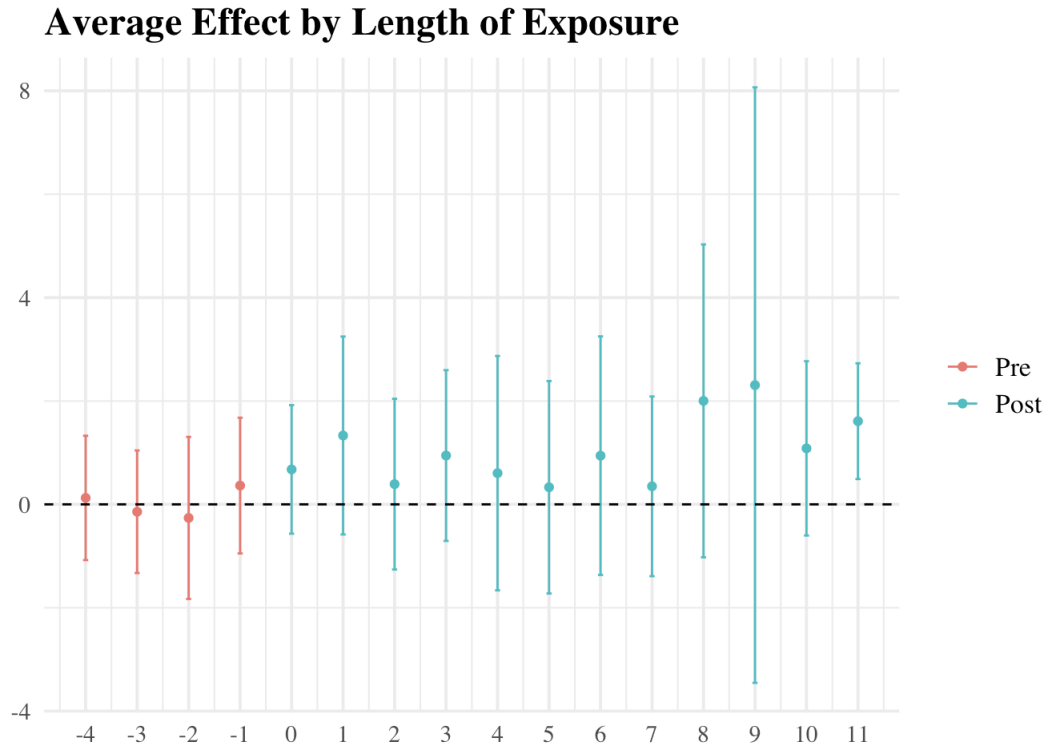
Group-Time ATT Estimates of Legalization Effects

```
library(broom)
#convert att_gt object into a tidy dataframe
gt_df <- tidy(att_did)

ggplot(gt_df, aes(x = time, y = estimate, color = as.factor(term), group = term)) +
  geom_line(linewidth = 1.1) +
  geom_point(linewidth = 2) +
  geom_errorbar(aes(ymin = estimate - 1.96 * std.error, ymax = estimate + 1.96 * std.error),
                width = 0.3, alpha = 0.6) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "gray40") +
  scale_color_brewer(palette = "Dark2", name = "Treatment Year") +
  labs(
    title = "Event Study: Group-Time ATT Estimates",
    subtitle = "Callaway & Sant'Anna (2021)",
    x = "Calendar Year",
    y = "Treatment Effect (ATT)"
  ) +
  theme_minimal(base_size = 14) + theme_serif()
```



Effect of Length of Exposure to Legalization

```
#aggregate plotting code
agg_event <- aggte(att_did, type = "dynamic")
ggdid(agg_event) + theme_serif()
```

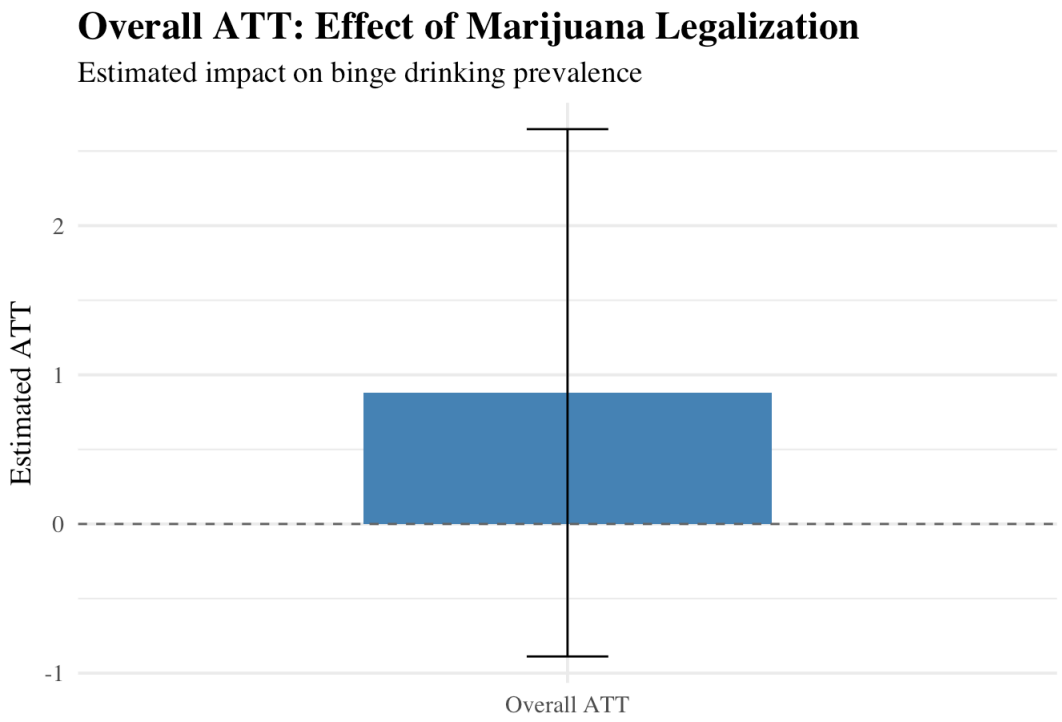


Overall Estimated ATT

```
#overall ATT

agg_overall <- aggte(att_did, type = "simple")
summary(agg_overall)
```

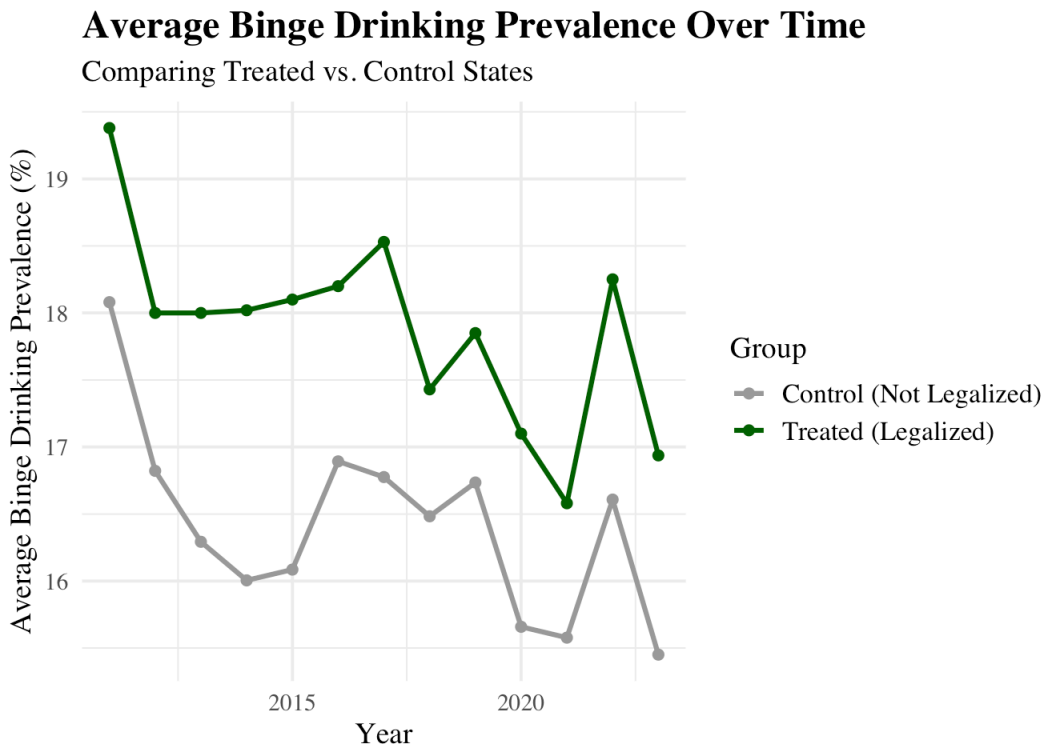
```
##
## Call:
## aggte(MP = att_did, type = "simple")
##
## Reference: Callaway, Brantly and Pedro H.C. Sant'Anna.  "Difference-in-Differences with Multiple Time Period
s." Journal of Econometrics, Vol. 225, No. 2, pp. 200-230, 2021. <https://doi.org/10.1016/j.jeconom.2020.12.001
>, <https://arxiv.org/abs/1803.09015>
##
##
##     ATT    Std. Error     [ 95%  Conf. Int.]
##  0.8792        0.9018    -0.8883      2.6466
##
##
## ---
## Signif. codes: `*' confidence band does not cover 0
##
## Control Group:  Never Treated,  Anticipation Periods:  0
## Estimation Method:  Doubly Robust
```

```
overall_df <- data.frame(
  Label = "Overall ATT",
  Estimate = agg_overall$overall.att,
  Lower = agg_overall$overall.att - 1.96 * agg_overall$overall.se,
  Upper = agg_overall$overall.att + 1.96 * agg_overall$overall.se
)

ggplot(overall_df, aes(x = Label, y = Estimate)) +
  geom_col(fill = "steelblue", width = 0.5) +
  geom_errorbar(aes(ymin = Lower, ymax = Upper), width = 0.1) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "gray40") +
  labs(
    title = "Overall ATT: Effect of Marijuana Legalization",
    subtitle = "Estimated impact on binge drinking prevalence",
    y = "Estimated ATT",
    x = ""
  ) +
  theme_minimal(base_size = 14) + theme_serif()
```



Treated vs. Control Trends in Binge Drinking Prevalence

```
#create average trends by treatment status
avg_trends <- data %>%
  group_by(Year, Legalized) %>%
  summarize(
    avg_binge = mean(Binge_Drinking_Prevalence, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  mutate(
    Legalized = ifelse(Legalized == 1, "Treated (Legalized)", "Control (Not Legalized)")
  )

#plot
ggplot(avg_trends, aes(x = Year, y = avg_binge, color = Legalized)) +
  geom_line(size = 1.1) +
  geom_point(size = 2) +
  scale_color_manual(
    values = c("Treated (Legalized)" = "darkgreen", "Control (Not Legalized)" = "gray60")
  ) +
  labs(
    title = "Average Binge Drinking Prevalence Over Time",
    subtitle = "Comparing Treated vs. Control States",
    x = "Year",
    y = "Average Binge Drinking Prevalence (%)",
    color = "Group"
  ) +
  theme_minimal(base_size = 13) +
  theme(legend.position = "bottom") + theme_serif()
```



State-Level Trends in Binge Drinking Prevalence by Legalization Status

```
ggplot(data, aes(x = Year, y = Binge_Drinking_Prevalence, group = State, color = as.factor(Legalized))) +
  geom_line(alpha = 0.5, size = 0.4) +
  scale_color_manual(
    values = c("0" = "darkgray", "1" = "darkgreen"),
    labels = c("0" = "Not Legalized", "1" = "Legalized"),
    name = "Legal Status"
  ) +
  labs(
    title = "Binge Drinking Prevalence Over Time by State",
    subtitle = "Lines represent individual states; color indicates legalization status",
    x = "Year",
    y = "Binge Drinking Prevalence (%)"
  ) +
  theme_minimal(base_size = 12) +
  theme(legend.position = "bottom") + theme_serif()
```

Binge Drinking Prevalence Over Time by State

Lines represent individual states; color indicates legalization status