# Problem Set 2

Andrew Simon

April 26, 2021

**Submission guide** This .pdf is meant to provide every answer necessary to complete this assignment. I did not include any code as I didn't think it was necessary, but it can be found in this github repository. All of the data is also included in the repository. Any questions about the assignment can be emailed to me at awsimon@uchicago.edu, or may be answered in the github readme.

## Question 1

Theorize on a variety of possible causal channels via which marijuana legalization might lead to more or fewer driving fatalities. For each causal channel, guesstimate a plausible sign and magnitude of the impact of the channel. Obviously, you are just speculating here using common sense and back of the envelope calculations.

**Response** There are several causal channels through which maraijuana legalization could shift driving outcomes.

**1) Legalization increases marijuana use and therefore impaired driving** Increased use of marijuana would increase driving fatalities through the mechanism of impairment. If marijuana does in fact impair driver and make them more likely to crash, increased use would lead to increased fatalities. Threats to this channel's viability are if legalization does not increase use and if weed does not impair driving ability.

I am fairly confident this is the most likely source of increased driving fatalities due to cannabis legalization. If we argue that the increase in driving deaths we have seen since 2012 is 100% due to legalization and impairment, we can take the difference of the 6 years post-2012 and the 2011 numbers and get a rough estimate of the imapct of impairment. In 2011 there were 10.42 driving deaths per 100,000 people in the US. From 2012-2018, the annual average number of driving deaths per 100,000 people was 10.9514286. So we estimate a positive impact of 0.5314 deaths per 100,000 people due to legalization. (Source)

**2) Legalization decreases drinking and driving** For some, weed and alcohol are subsitutes. Legalization could lead to increased substitution from alcohol to weed as the cost of smoking weed (both the price and risk associated with smoking) potentially decreases with legalization.

Data on the danger of driving high are mixed, with some studies suggesting that high drivers are actually better when they are high compared to when they are sober. Despite mixed data, it is generally accepted (both in the literature and anecdotally) that driving high is safer than driving drunk. I doubt this hypothesis because while I believe that smoking and drinking are substitutes, I don't believe people who are driving drunk will substitute at a significant rate. Clearly they are committed to drinking. I also have observed anecdotally that some people who are willing to drive high are unwilling to drive drunk.

28% of traffic deaths involved drunk driving in 2016. To get an upper bound on potential impact lets say 100% of drunk driving crashes could be avoided if the driver were high instead of drunk. With those assumptions, 10,497 deaths would've been avoided in 2016 through substitution (27.7654341%).

## Question 2

Describe a "natural experiment" research design based on law changes that you will implement using available data that you will collect. (Some things to think about: (a) there have been various versions of legalization/medical marijuana/decriminalization across states. Provide a compelling logic as to which state law changes have materially impacted the ability to obtain marijuana legally, (b) give some thought to timing… when do you think marijuana usage will increase relative to passage of relevant laws, and (c) justify your choice of a control group).

**Response** I will use state-level accident data from the National Highway Traffic Safety Administration (NHTSA) manually paired with legalization data from multiple sources (mainly this wikipedia page, which I confirmed is accurate). I coded this legalization data into a .csv file which can be viewed here. I used population data from the census (found in git repo) from 2010-2019. The 2010 start provides 3 years of data before treatment (because treatment is delayed until a year after the passage of legalization policy), and was the most consise and accurate data I could find. There is limited need for older data.

I will use these law changes to regress the annual number of car crash fatalities on the number of years a legalization law has been in place. The null hypothesis would correspond to 0 change with an alternative hypothesis of any impact.

After reviewing some data on the number of dispensaries open in each state after legalization, I decided the floor for treatment should be 1 year. The number of dispensaries increases significantly after the one year threshold. Additionally, legalization legislation often involves a delay, meaning that the time coded for this regression is likely before the full real-world increase in smoking that occurs as a result of legalization. For this reason, these estimates can be considered a conservative estimate of the impact of smoking on crashes.

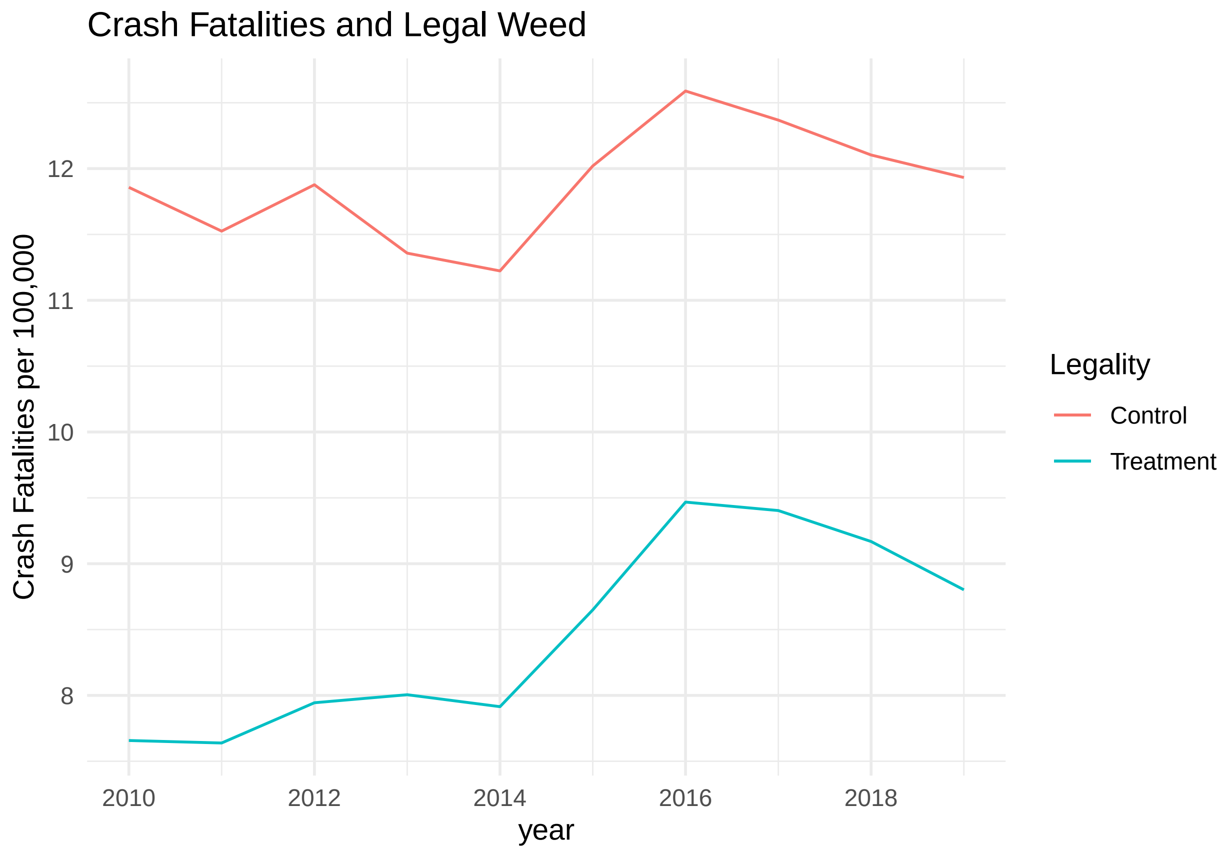
## Question 3

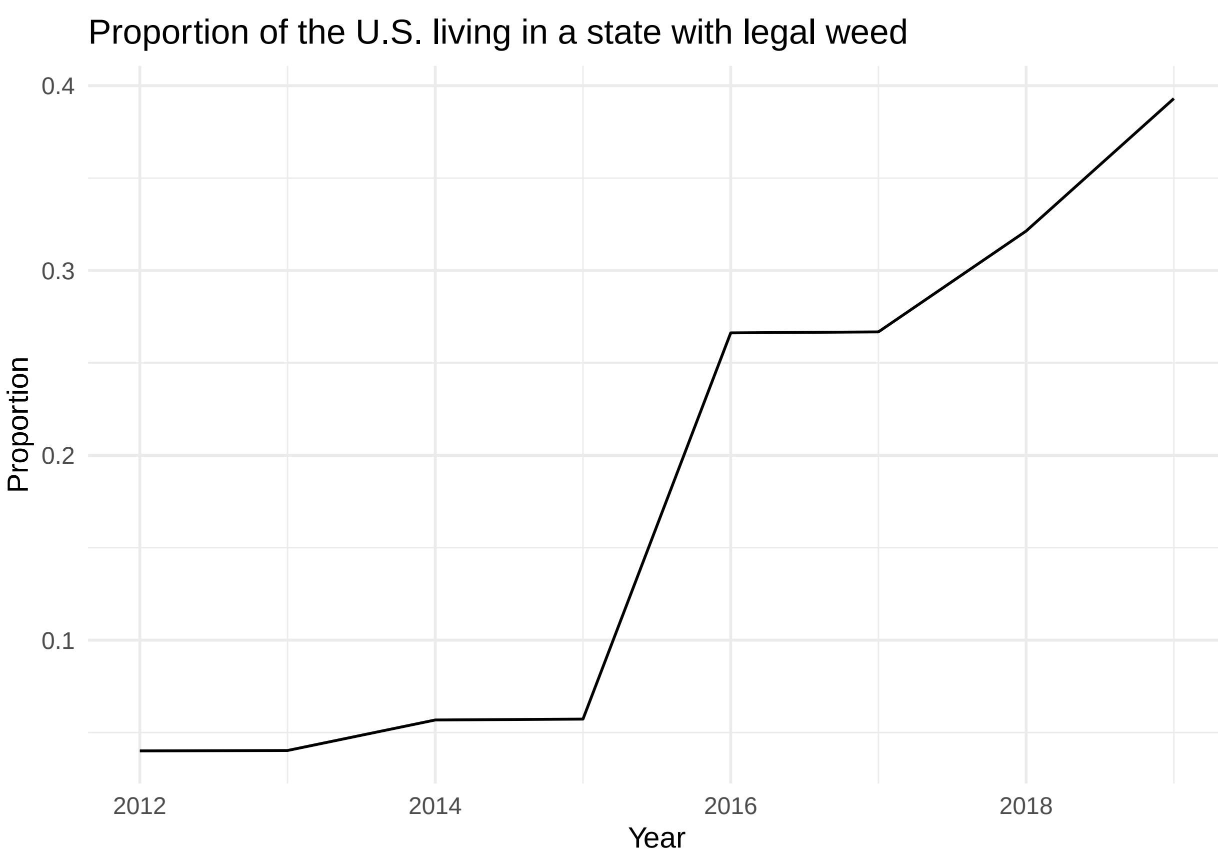
Collect data at the state-year level on the legal status of marijuana and on traffic fatality death rates. NO OTHER CONTROL VARIABLES ARE NECESSARY. Make a summary statistics table, and also a graph showing how traffic deaths and the legal status of marijuana have evolved over time.

**Response** Here I import the data from the NHTSA, the Census Bureau, and my own hard coded data about weed legalization. The code for these imports can be found in the .Rmd file in the git repo for this assignment.

From looking at these charts, it seems as though a small but potentially significant increase in fatalities could be linked to the increase in population living under legal weed from 2014-2016. The delay in actual changes in weed use after legalization complicates this intial hypothesis. It is also worth noting that states who have approved legal weed currently have much lower crash fatalities per 100,000 people. This significant difference will limit the external validity of the results of this natural experiment.

```
## <environment: R_GlobalEnv>
```


Crash Fatalities and Legal Weed
Crash Fatalities per 100,000
12
11
10
9
8
2010
2012
2014
2016
2018
year
Legality
Control
Treatment


Proportion of the U.S. living in a state with legal weed
Proportion
0.4
0.3
0.2
0.1
2012
2014
2016
2018
Year

## Question 4

Using the research design in question 2 and the data collected in question 3, come up with an estimate of the causal impact of marijuana legalization on traffic fatalities. Please provide both a point estimate and a standard error. Explain whether you think the magnitude of your estimate is large or small.

### Response

Based on these law changes, I ran a regression of crash fatalities per 100,000 citizens on the number of years post-legalization, with a one year floor. Dummy variables were included for every year and every state to control for state and time effects. Additionally, decriminalization and legal medical use were controlled for. This multiple linear regression returned a coefficient of 0.25496 more crash fatalities per year of legalization per 100,000 citizens. This coefficient is associated with a standard error of 0.08817 and a p value of .00402, making it significant at the 0.01 level. A full summary of the model is included below this response. This regression involved all states.

In 2019, the highest crash fatalities rate per 100,000 was 25.5 in Wyoming, while the lowest was 3.28 in Washington, DC. Given the positive impact of ~0.26 per year, this is fairly significant today. What makes this finding even more significant, however, is the potential impact for this number of deaths to grow. There is not enough data currently to fully understand the impact of legal weed on smoking and driving habits, but it seems safe to assume that fatal crashes will increase with weed smoking as legality increases nation wide.

Additionally, while this 0.26 per 100,000 number may not sound large immediately, it is worth considering in context. In California, a state with legal weed, there were 9.1 crash deaths per 100,000 people or around 3,595 people. If this 0.26 per 100,000 increase were to happen next year (which is a much stronger outcome of this analysis than I support but a valuable exercise), 3,698 people will die traffic deaths, in increase of around 100 people. This is an immensely high cost of legalization.

```
weed_accidents <- inner_join(accidents , weeddata , by = c('STATE' , 'year')) %>%
  inner_join(pop_data , by = c('STATE' , 'year')) %>%
  mutate(years2legal = legal_year - year) %>%
  mutate(fatals_pc = (fatals/pop)*100000) %>%
  dummy_cols(select_columns = c('STATE' , 'year')) %>%
  mutate(treated_legal = ifelse(is.na(years2legal) == F , 1 , 0)) %>%
  mutate(yearsafter = case_when(year-legal_year > 0 ~ year-legal_year ,
                                year-legal_year <= 0 ~ 0 ,
                                is.na(legal_year) ~ 0))

reg1 <- lm(fatals_pc ~ yearsafter +  DCRIM + MED + STATE_1 + STATE_2 + STATE_4
           +  STATE_5    +  STATE_6  +  STATE_8 +  STATE_9 + STATE_10 +
            STATE_11 + STATE_12 +  STATE_13    +  STATE_15    +   STATE_16    +
            STATE_17    +   STATE_18 + STATE_19  +    STATE_20   +  STATE_21   +
            STATE_22   +    STATE_23   +    STATE_24   +    STATE_25   +  STATE_26+
            STATE_27   +    STATE_28   +    STATE_29   +    STATE_30   +
            STATE_31   +    STATE_32   +    STATE_33   +    STATE_34   +
            STATE_35   +    STATE_36   +    STATE_37   +    STATE_38   +
            STATE_39   +  STATE_40   +    STATE_41     +  STATE_42   +
            STATE_44     +  STATE_45   +    STATE_46   +    STATE_47+
             STATE_48   +    STATE_49   +    STATE_50     +  STATE_51   +
            STATE_53+    STATE_54   +    STATE_55+    STATE_56   +
            year_2010    +  year_2011    +  year_2012 +    year_2013 +
            year_2014 + year_2015 +year_2016 + year_2017 + year_2018 +
            year_2019 , data = weed_accidents)
summary(reg1)
```

```
##
## Call:
## lm(formula = fatals_pc ~ yearsafter + DCRIM + MED + STATE_1 +
##     STATE_2 + STATE_4 + STATE_5 + STATE_6 + STATE_8 + STATE_9 +
##     STATE_10 + STATE_11 + STATE_12 + STATE_13 + STATE_15 + STATE_16 +
##     STATE_17 + STATE_18 + STATE_19 + STATE_20 + STATE_21 + STATE_22 +
##     STATE_23 + STATE_24 + STATE_25 + STATE_26 + STATE_27 + STATE_28 +
##     STATE_29 + STATE_30 + STATE_31 + STATE_32 + STATE_33 + STATE_34 +
##     STATE_35 + STATE_36 + STATE_37 + STATE_38 + STATE_39 + STATE_40 +
##     STATE_41 + STATE_42 + STATE_44 + STATE_45 + STATE_46 + STATE_47 +
##     STATE_48 + STATE_49 + STATE_50 + STATE_51 + STATE_53 + STATE_54 +
##     STATE_55 + STATE_56 + year_2010 + year_2011 + year_2012 +
##     year_2013 + year_2014 + year_2015 + year_2016 + year_2017 +
##     year_2018 + year_2019, data = weed_accidents)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.8705 -0.5798 -0.0027  0.5899  6.6052
##
## Coefficients: (2 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  22.09311    0.42679  51.766  < 2e-16 ***
## yearsafter    0.25496    0.08817   2.892  0.00402 **
## DCRIM         0.01002    0.30414   0.033  0.97373
## MED          -0.17839    0.25388  -0.703  0.48265
## STATE_1      -3.61587    0.54870  -6.590 1.24e-10 ***
## STATE_2     -13.16341    0.62597 -21.029  < 2e-16 ***
## STATE_4      -9.10906    0.60459 -15.067  < 2e-16 ***
## STATE_5      -4.44439    0.54870  -8.100 5.29e-15 ***
## STATE_6     -13.71109    0.66745 -20.542  < 2e-16 ***
## STATE_8     -13.12002    0.68574 -19.133  < 2e-16 ***
## STATE_9     -14.54397    0.63344 -22.960  < 2e-16 ***
## STATE_10     -9.89076    0.60531 -16.340  < 2e-16 ***
## STATE_11    -18.64068    0.54870 -33.973  < 2e-16 ***
## STATE_12     -8.55661    0.54870 -15.594  < 2e-16 ***
## STATE_13     -8.98888    0.54870 -16.382  < 2e-16 ***
## STATE_15    -14.46129    0.60351 -23.962  < 2e-16 ***
## STATE_16     -9.45229    0.54870 -17.227  < 2e-16 ***
## STATE_17    -14.45342    0.58415 -24.743  < 2e-16 ***
## STATE_18    -10.12729    0.54870 -18.457  < 2e-16 ***
## STATE_19    -11.17459    0.54870 -20.366  < 2e-16 ***
## STATE_20     -8.43795    0.54870 -15.378  < 2e-16 ***
## STATE_21     -5.62813    0.54870 -10.257  < 2e-16 ***
## STATE_22     -6.37325    0.56319 -11.316  < 2e-16 ***
## STATE_23    -10.88711    0.60980 -17.854  < 2e-16 ***
## STATE_24    -13.75435    0.59125 -23.263  < 2e-16 ***
## STATE_25    -17.03867    0.65271 -26.104  < 2e-16 ***
## STATE_26    -12.45240    0.60514 -20.578  < 2e-16 ***
## STATE_27    -15.21124    0.63517 -23.948  < 2e-16 ***
## STATE_28     -0.78188    0.62735  -1.246  0.21330
## STATE_29     -8.26124    0.57817 -14.289  < 2e-16 ***
## STATE_30     -2.74390    0.60459  -4.538 7.29e-06 ***
```

```
## STATE_31    -10.71264    0.62735 -17.076  < 2e-16 ***
## STATE_32    -12.08554    0.60980 -19.819  < 2e-16 ***
## STATE_33    -13.53768    0.57992 -23.344  < 2e-16 ***
## STATE_34    -15.61211    0.60459 -25.823  < 2e-16 ***
## STATE_35     -4.59811    0.60351  -7.619 1.54e-13 ***
## STATE_36    -16.66008    0.63517 -26.229  < 2e-16 ***
## STATE_37     -8.85257    0.62735 -14.111  < 2e-16 ***
## STATE_38     -4.81794    0.54954  -8.767  < 2e-16 ***
## STATE_39    -12.90191    0.62848 -20.529  < 2e-16 ***
## STATE_40     -4.96579    0.55104  -9.012  < 2e-16 ***
## STATE_41    -12.51648    0.62597 -19.995  < 2e-16 ***
## STATE_42    -12.77027    0.55802 -22.885  < 2e-16 ***
## STATE_44    -16.37495    0.63792 -25.669  < 2e-16 ***
## STATE_45     -3.61610    0.54870  -6.590 1.24e-10 ***
## STATE_46     -7.41939    0.54870 -13.522  < 2e-16 ***
## STATE_47     -6.89063    0.54870 -12.558  < 2e-16 ***
## STATE_48     -9.50757    0.54870 -17.327  < 2e-16 ***
## STATE_49    -13.76440    0.55104 -24.979  < 2e-16 ***
## STATE_50    -12.26808    0.62956 -19.487  < 2e-16 ***
## STATE_51    -13.01910    0.54870 -23.727  < 2e-16 ***
## STATE_53    -15.91753    0.66554 -23.917  < 2e-16 ***
## STATE_54     -5.98599    0.55396 -10.806  < 2e-16 ***
## STATE_55    -12.30555    0.54870 -22.427  < 2e-16 ***
## STATE_56           NA         NA      NA       NA
## year_2010     0.36809    0.27205   1.353  0.17673
## year_2011    -0.02809    0.26942  -0.104  0.91701
## year_2012     0.36461    0.26579   1.372  0.17081
## year_2013    -0.27726    0.26156  -1.060  0.28970
## year_2014    -0.32718    0.25613  -1.277  0.20213
## year_2015     0.33744    0.25294   1.334  0.18286
## year_2016     0.80542    0.24968   3.226  0.00135 **
## year_2017     0.65153    0.24639   2.644  0.00847 **
## year_2018     0.31276    0.24437   1.280  0.20127
## year_2019          NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.227 on 447 degrees of freedom
## Multiple R-squared:  0.9369, Adjusted R-squared:  0.9281
## F-statistic:   107 on 62 and 447 DF,  p-value: < 2.2e-16
```

## Question 5

Discuss the public policy implication of your finding, and what your results say about the theories in question 1.

### Response

There are several potential policy responses to the positive causal impact legal smoking has on crash fatalities. With the significant funding that can be provided through taxes on legal marijuana, many of these responses are feasible and could have a significant impact in decreasing crash fatalities that result due to high driving and those that do not.

First, expanding and increasing the requirements of driver's education programs in states where weed is legal could help prevent these crashes. By explaining to new drivers the risks of high driving (as we explain the risk of drunk driving), policy makers could begin to dispel the myth that smoking does not inhibit driving ability. This policy change is easily implementable and cost effective.

Second, policy makers should consider expanding funding for highway patrols. There is a strong correlation between driving speed and fatality risk. By directing funding toward highway patrols, municipalities and states can help mitigate the riskiest high driving. Technologies like speed cameras and radar guns could assist significantly in preventing fatalities that result from high driving. The CDC reports that high driving increases at night time. Through an expansion of night-time highway patrol, the likelihood of being caught and punished for risky driving increases and this will help deter bad behavior.

Third, the CDC and U.S. government should invest significantly in more effective methods of measuring weed inhibition. Currently, there are ineffective and subjective ways of measuring intoxication from marijuana. This not only decreases the likelihood an individual is punished for driving high, it also increases individual police officers discretion when acting in high driving situations. A widely-publicized and highly-effective 'weed breathalyzer' would deter drivers from smoking before getting behind the wheel.

Like any policy, legalizing marijuana has trade offs. This increase in driving deaths, while negative, must be considered in context with the benefits of legalizing weed. Thinking strictly in terms of finances, which is to ignore the injustices of criminalizing marijuana, tax revenue from weed likely offsets the social cost of driving deaths completely. For an additional ~100 road deaths, California earned 1.8 billion dollars in tax revenue. This is a high number of deaths that has to be seriously considered seriously when proposing legalization policies.