

# Investigating Yearly and Generational Attitudes towards Legalization of Marijuana in the United States

Martin Hsu, Statistics, California Polytechnic State University  
STAT 365, Rossman and Grant, Spring 2022

## Abstract

An investigation was conducted into trends regarding attitudes of US adults towards legalizing marijuana, using General Social Survey (GSS) data. Support for marijuana was used as a binary response variable, with predictors of age group and year. Significant differences between proportion of support in years 2008 and 2018 were found, as well as between age groups 18-34 and 50-64 in the year 2018. After fitting a logistic model to odds of a US adult supporting legal marijuana with year as a predictor, the model demonstrated a multiplicative increase in the aforementioned odds over time.

## 1. Introduction

Marijuana has long been a source of controversy in the United States. During the “War on Drugs,” marijuana was criminalized at the federal level, spawning a passionate campaign against its use [1]. However, a growing number of states have legalized either recreational or medical use of marijuana, indicating possible changing attitudes towards use of the drug.

Starting in 1973, the National Opinion Research Center (NORC) at the University of Chicago began collecting data on social and political opinion trends through large scale surveys. These surveys were collected for the express purpose of researching topics and trends in pertinent to the social sciences [2]. Regarding marijuana, this survey, called the General Social Survey (GSS) included the question “Should marijuana be made legal?” in addition to basic demographic data.

This study aims to investigate yearly and generational trends in attitudes regarding legal marijuana, using responses from the General Social Survey to the above question. Three main research questions were posed:

- 1. Was the proportion of the United States adult population supporting legal marijuana greater in 2018 than in 2008 (a decade before)?*
- 2. Did young adults (18-34 years) in 2018 tend to have a higher proportion of supporters of legal marijuana in the United States than older adults (50-64 years)?*
- 3. Overall (1973-2018), do the odds of an adult in the United States supporting legal marijuana tend to increase over time?*

## 2. Methods and Materials

The data analyzed originated from the GSS, collected by NORC either annually or biannually. According the GSS methodological primer for 2020, the GSS is a “nationally representative survey of the attitudes and behaviors of adults in the United States” and is a proportional sample of the demographics of the US population [3]. Around 700-2000 people are selected for the survey each year, with sampling methodologies varying year to year. GSS data were accessed at the following site:

<https://sda.berkeley.edu/sdaweb/analysis/?dataset=gss21>

Each observational unit was an adult resident of the United States, age 18 or older. The explanatory variables of interest to this analysis were the age of the participants, which for this analysis were aggregated into four categories: 18-34, 35-49, 50-64, and 65+ years-old respondents; and the year the study was conducted. The response variable of interest to this analysis was whether or not the respondent supported marijuana legalization.

Each respondent was asked to provide demographic data, among which was the explanatory variable of age. They were then asked to respond to (among other questions) the following question:

*“Should marijuana be made legal?”*

Participants in the survey responded either “should be legal” or “should not be legal.”

### 3. Descriptive Analysis

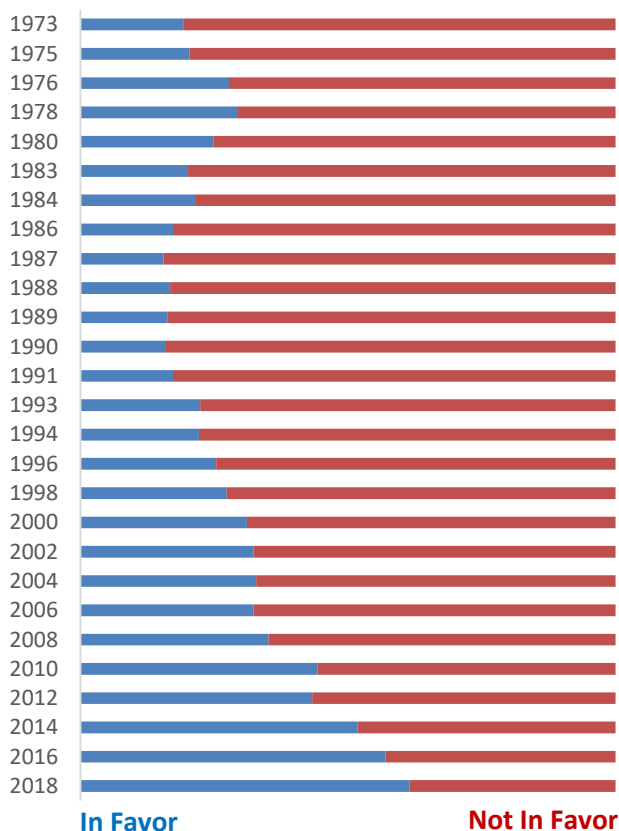
#### 3.1 Comparison and Analysis of Trends for Years

*Key Question: Was the proportion of the United States adult population supporting legal marijuana greater in 2018 than in 2008 (a decade before)?*

*Key Question: Overall (1973-2018), do the odds of an adult in the United States supporting legal marijuana tend to increase over time?*

Firstly, we conducted a visual analysis of the data to detect the possibility of trends of interest to the research questions posed in this analysis.

**Figure 3.1: Yearly Proportion of GSS Respondents In Favor/Not in Favor of Legalization**

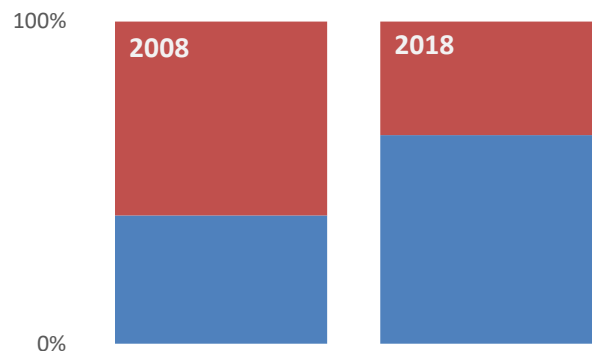


As seen above in Figure 3.1, every year up until about 2014 appears to have a majority of respondents not in favor of legal marijuana. From 2014 onwards, most respondents appear to be in favor of legal marijuana. Overall, there appears to be a positive trend for those in favor. In other words, as time goes on, more people appear to be in favor of legalizing marijuana. It would also be plausible to state that as year increases, the odds of a respondent being in favor of legalization increases.

An inspection of the proportions for 2018 and 2008 appear to show that the 2018 survey has a larger proportion of respondents in favor of legalization than

that of 2008. Looking at Figure 3.2 and Table 3.1, this observation appears to be supported by the survey statistics.

**Figure 3.2: Proportion of GSS Respondents in Favor/Not in Favor of Legalization in 2008 and 2018**



**Table 3.1: Yearly Proportion of GSS Respondents In Favor of Legalization**

Year	Count	No. In Favor	Pct. in Favor
2008	1247	496	39.8
2018	1447	938	64.8

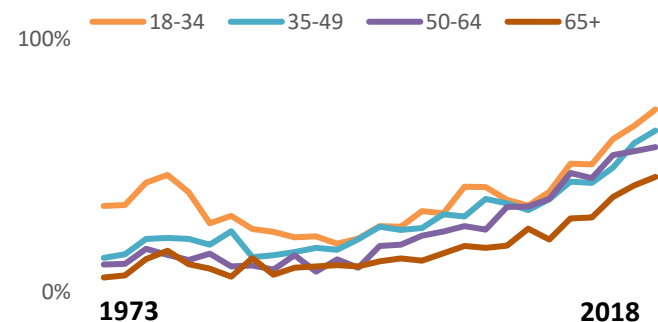
According to Table 3.1, the proportion of valid respondents in favor of legalization in 2018 was 25.0 percentage points greater than that of 2008. This results in a z-score of 12.7, which is extremely high.

#### 3.2 Comparison of Age Groups

*Key Question: Did young adults (18-34 years) in 2018 tend to have a higher proportion of supporters of legal marijuana in the United States than older adults (50-64 years)?*

Looking at Figure 3.3, we can see that generally, US adults aged 18-34 tend to have the highest proportion of supporters of legalization. In 2018, it appears that the age group with the most supporters of legalization are ages 18-34, followed by 35-49, 50-64, and finally 65+.

**Figure 3.3: Yearly Proportion of GSS Respondents In Favor of Legalization, by Age Group**



**Table 3.2: Proportion of 2018 GSS Respondents In Favor of Legalization, by Age Group**

Age Group	Count	No. In Favor	Pct. in Favor
18-34	401	310	77.3
34-49	349	242	69.2
50-64	379	238	62.6
65+	318	148	50.1

The data in Table 3.2 supports this observation. For all age groups, more than 50 percent of valid survey participants in that age group responded in favor of legalization. In particular, the proportion of 18- to 34-year-olds in favor of legalization was 14.7 percentage points greater than 40- to 64-year-olds. This results in a z-score of 4.5, which is relatively high.

## 4. Inferential Analysis

### 4.1 Comparison of Years

*Key Question: Was the proportion of the United States adult population supporting legal marijuana greater in 2018 than in 2008 (a decade before)?*

To see if there was a significant difference between the proportion of support for legalization in 2018 versus 2008, we performed a two-proportion z-test. Based on the two-proportion z-test, we found that there was sufficient evidence ( $p < 0.001$ ) to conclude that the proportion of legalization supporters in 2018 was greater than that of 2008.

As a result of finding a significant difference between proportion of supporters of legalization in 2018 and 2008, we constructed a 95% confidence interval for the population difference in proportion. Based on the confidence interval, we are 95% confident that the true proportion of supporters of legalization among adults in the US is between 21.3 and 28.7 percentage points higher in 2018 than in 2008.

### 4.2 Comparison of Age Groups

*Key Question: Did young adults (18-34 years) in 2018 tend to have a higher proportion of supporters of legal marijuana in the United States than older adults (50-64 years)?*

To assess whether the proportion of 18- to 34-year-olds supporting legalization was higher than that of 50- to 64-year-olds, we performed yet another two-proportion z-test. Based on the two-proportion z-test, we found sufficient evidence ( $p < 0.001$ ) to conclude that the proportion of legalization supporters for 18- to 34-year-olds was greater than that of 50- to 64-year-olds.

As a result of finding a significant difference between proportion of supporters of legalization between the aforementioned age groups, we constructed a 95% confidence interval for the population difference in proportion. Based on the confidence interval, we are 95% confident that the true proportion of US supporters of legalization in 2018 is between 8.1 and 20.8 percentage points higher for 18- to 34-year-olds than for 50- to 64-year-olds.

### 4.3 Analysis of Yearly Trends

*Key Question: Overall (1973-2018), do the odds of an adult in the United States supporting legal marijuana tend to increase over time?*

Finally, we would like to see if the odds of a US adult being in support of legalization increases over time. Because the response variable is a proportion, we decided to generate a logistic model of the odds. Performing a logistic regression, we generate the following model:

$$\log(\widehat{\text{Odds for Legalize}}) = -84.51 + 0.042(\text{Year})$$

This model has a positive slope, indicating that as the year increases (the first instance of the survey where the legalization question appears) increases, the log-odds of a US adult being in favor of legalization increases as well. We can rewrite the equation as follows:

$$\widehat{\text{Odds for Legalize}} = \exp(-84.51 + 0.042(\text{Year}))$$

As an interpretation, based on the GSS data, for every 1 year that passes, the predicted odds that a US adult is in favor of legal marijuana increases by a factor of  $\exp(0.042)$ , which equals 1.043.

To see if the slope is significantly greater than zero, we performed a slope z-test on the coefficient for year. Based on the z-test we found significant evidence ( $p < 0.001$ ) to conclude that the slope is greater than 0. Therefore, we can conclude there is strong evidence that as year increases, the odds of a US adult being in favor of marijuana increases by a greater factor than one as well.

Because of this significant result, we constructed a 95% confidence interval for the factor by which the odds increase each year. Based on the confidence interval, we are 95% confident that for every one year that passes, the expected odds that a US adult is in favor of legal marijuana increases by a factor of between 1.041 and 1.045.

## 5. Conclusion

From an analysis of the GSS data, we can conclude that the expected proportion of adult supporters of legal marijuana in the US was likely greater in 2018 than it was in 2008. Additionally, in 2018, the expected proportion of 18- to 34-year-olds in the US that supported legal marijuana was likely larger than the proportion of 50- to 64-year-olds that supported marijuana. Finally, for every year that passes, the odds of a US adult supporting legal marijuana tends to increase as well.

## 6. Study Limitations

As mentioned before, the data were collected as part of the General Social Survey (GSS). Because of the observational nature of the study, it is ill advised to draw cause and effect conclusions from these findings. There is a high likelihood that there exist confounding factors that significantly affect support for legal marijuana. Further investigation, therefore, is needed to ascertain whether or not age or year has a conclusive effect on proportion of support for legal marijuana.

Additionally, it is not advised that the logistic regression used to predict the odds of support for legal marijuana using the year be seen as a robust predictive model, or generalized beyond the years included in the data. Steps that are important to the development of a regression model, such as the checking of model assumptions and data validation, have not been performed. Because year is the predictor, the independence of each observation is suspect. Additionally, the model implies that the odds that a US adult supports legal marijuana increases monotonically over time, which is unlikely.

## References

- [1] History.com Editors. "Marijuana." *History.com*, A&E Television Networks, 31 May 2017, <https://www.history.com/topics/crime/history-of-marijuana>.
- [2] "GSS Data Explorer: NORC at the University of Chicago." *GSS Data Explorer | NORC at the University of Chicago*, University of Chicago, <https://gssdataexplorer.norc.org/about>.
- [3] "2020 Panel Methodological Primer." *GSS Data Explorer | NORC at the University of Chicago*, University of Chicago, <https://gss.norc.org/Documents/other/2020%20Panel%20Methodological%20Primer.pdf>.

## Appendix

### Appendix 4.1A – R Code for 2-Sample z-Test for Difference in Proportion in Favor of Legalization, Year 2008 against Year 2018

```
prop.test(x = c(938, 496), n = c(1447, 1247),  
          alternative = "greater", correct = FALSE)
```

```
2-sample test for equality of proportions without continuity  
correction
```

```
data: c(938, 496) out of c(1447, 1247)  
X-squared = 168.8, df = 1, p-value < 2.2e-16  
alternative hypothesis: greater  
95 percent confidence interval:  
 0.2197247 1.0000000  
sample estimates:  
 prop 1 prop 2  
0.6482377 0.3977546
```

### Appendix 4.1B – R Code for 95% Confidence Interval for Difference in Proportion in Favor of Legalization, Year 2008 against Year 2018

```
diffscoreci(938, 1447, 496, 1247, conf.level = 0.95)
```

```
data:
```

```
95 percent confidence interval:  
 0.2135005 0.2867657
```

### Appendix 4.2A – R Code for 2-Sample z-Test for Difference in Proportion in Favor of Legalization, Age Group 18-34 against Age Group 50-64, Year 2018

```
prop.test(x = c(310, 238), n = c(401, 379),  
          alternative = "greater", correct = FALSE)
```

```
2-sample test for equality of proportions without continuity  
correction
```

```
data: c(310, 238) out of c(401, 379)  
X-squared = 19.631, df = 1, p-value = 4.697e-06  
alternative hypothesis: greater  
95 percent confidence interval:  
 0.09170039 1.0000000  
sample estimates:  
 prop 1 prop 2  
0.7730673 0.6279683
```

## Appendix 4.2B – R Code for 95% Confidence Interval for Difference in Proportion in Favor of Legalization, Year 2008 against Year 2018

```
diffscoreci(310, 401, 238, 379, conf.level = 0.95)
```

data:

95 percent confidence interval:

0.08114115 0.20838679

## Appendix 4.3A – R Code to Generate Dummy Variable Dataset for Logistic Regression, Based off GSS Data

```
legal <- c(rep(1, 275), rep(0, 1196),
          rep(1, 301), rep(0, 1113),
          rep(1, 416), rep(0, 1031),
          rep(1, 447), rep(0, 1017),
          rep(1, 364), rep(0, 1056),
          rep(1, 324), rep(0, 1222),
          rep(1, 332), rep(0, 1074),
          rep(1, 263), rep(0, 1174),
          rep(1, 300), rep(0, 1461),
          rep(1, 171), rep(0, 784),
          rep(1, 169), rep(0, 829),
          rep(1, 146), rep(0, 721),
          rep(1, 182), rep(0, 790),
          rep(1, 234), rep(0, 770),
          rep(1, 457), rep(0, 1450),
          rep(1, 492), rep(0, 1332),
          rep(1, 525), rep(0, 1263),
          rep(1, 597), rep(0, 1183),
          rep(1, 306), rep(0, 545),
          rep(1, 292), rep(0, 510),
          rep(1, 672), rep(0, 1156),
          rep(1, 496), rep(0, 751),
          rep(1, 603), rep(0, 656),
          rep(1, 586), rep(0, 648),
          rep(1, 870), rep(0, 704),
          rep(1, 1126), rep(0, 717),
          rep(1, 938), rep(0, 509)
        )
years <- c(rep(1973, 1471),
          rep(1975, 1414),
          rep(1976, 1447),
          rep(1978, 1464),
          rep(1980, 1420),
          rep(1983, 1546),
          rep(1984, 1406),
          rep(1986, 1437),
          rep(1987, 1761),
          rep(1988, 955),
          rep(1989, 998),
          rep(1990, 867),
          rep(1991, 972),
          rep(1993, 1004),
          rep(1994, 1907),
          rep(1996, 1824),
          rep(1998, 1788),
          rep(2000, 1780),
          rep(2002, 851),
          rep(2004, 802),
          rep(2006, 1828),
          rep(2008, 1247),
          rep(2010, 1259),
          rep(2012, 1234),
          rep(2014, 1574),
          rep(2016, 1843),
          rep(2018, 1447)
        )
grass_log <- tibble("legal" = factor(legal,
                                   levels = c(0, 1),
                                   labels = c('no', 'yes')),
                  "year" = years)
```

#### Appendix 4.3B – R Code to Generate Logistic Regression Model and z-Tests of In Favor/Not In Favor of Legalization Versus Year

```
grass_fit <- glm(legal ~ year, family = binomial, data = grass_log)
summary(grass_fit)
```

```
Call:
glm(formula = legal ~ year, family = binomial, data = grass_log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max 
-1.2436 -0.8729 -0.7030  1.1831  1.9474 

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -8.451e+01  1.761e+00  -47.98   <2e-16 ***
year          4.195e-02  8.819e-04   47.57   <2e-16 ***
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 46874  on 37545  degrees of freedom
Residual deviance: 44465  on 37544  degrees of freedom
AIC: 44469

Number of Fisher Scoring iterations: 4
```

#### Appendix 4.3C: R Code for 95% Confidence Interval for Logistic Regression Coefficient (Year)

```
exp(confint(grass_fit))
```

```
2.5 %      97.5 %
(Intercept) 6.270370e-39 6.248839e-36
year        1.041048e+00 1.044653e+00
```