Imperial College London

Health Care and Medical Analytics: Individual Assignment

The Relation between Marijuana Involvement and other Substance Involvement

–

A quantitive analysis among adolescent students in the US

Author: Cyrill Studer

May 24, 2019

Table of Contents

1 Introduction

In many Western countries, the legalisation of pharmaceutical marijuana or marijuana in general is either up for discussion or already a fast-growing industry. With the liberalisation, as shown by the example of Canada, the number of initiations rises steeply and consequently also the marijuana involvement of teens and young adults increases (Zuckermann et al. 2019, p.7). The goal of this paper is to evaluate the association of marijuana involvement with other substance involvement among teens and young adults. This is of particular interest as heavy involvement in drinking, smoking and other illegal drug involvement in the early stage of a life can have serious health consequences for the individual and be connected to a significant increase in health costs for a society (Gryczynski et al. 2016, p.16).

Most of the literature regarding adolescence marijuana involvement and its associated risk factors focus on particular relationships such as demographic factors, psychological condition or social factors while controlling only for some of the other relevant risk factors. A focused but comprehensive modeling of all significant risk factors is rarely the focus. Already in 2005, the work of Van den Bree and Pickworth identified own and peer substance involvement, delinquency and school problems as strong predictors of marijuana involvement for teens and young adults. Consequently, they recommended designing prevention and policy interventions around these risk areas. The work of Mahalik et al., 2015, found that social economical variables and a range of basic demographic variables are valuable indicators to determine marijuana involvement. In the same year, De La Haye et al., 2015, specifically found that the social environment and friends networks form highly relevant, collective risk clusters and are therefore strong determinators for marijuana involvement. This paper is willing to connect the above-introduced approaches and focus areas in order to comprehensively assess the significance of the marijuana involvement and other substance involvement relationship among teens and young adults.

The paper is being organised in three sections. First of all, the methodology and exploratory data analysis will be introduced in order to select and engineer meaningful variables for the quantitative analysis. Secondly, a range of different logistic regression models will be used in order to assess the relationship between marijuana involvement and other substance involvement. Lastly, a conclusion will address the consequences as well as the limitations of the model.

2 Method and Descriptive Statistics

This study assesses the relationship between *marijuana involvement* and *smoking, heavy drinking* and *other drug involvement* while controlling for the three groups of variables demographic factors, social norms and psychological condition. For each of the four variable groups, four to nine variables were pre-selected based on the considered variables from the three papers introduced in the introduction section and the data availability. The data of this paper's analysis is from the National Longitudinal Study of Adolescent Health, Wave 2, a comprehensive survey conducted in 1996 that contains a systematic random sample of 4,834 observations across the US. A detailed overview of the selected variables can be found in Appendix 1. This section will describe in more detail the selection of the variables, the two steps how this study handled missing data and subsequently how the relevant variables of this study's model were engineered and modified followed by the most important insights of the exploratory data analysis.

For the variable selection, attention was paid to select complementary variables rather than correlated variables. With regard to the variable group demographics, standard variables such as *gender* or *age* are used. Moreover, as household income data was not available, the social economic status of the participants was approximated with a combination of the variables *neighborhood safety* and *neighborhood happiness*. As school life and family/ friends constitute the social norms of most adolescent students, close attention was paid to cover all relevant dimensions of these two factors. In terms of school, the factors *teacher trouble, peer trouble, four different grades and suspension* were considered. For the family and friends dimension, *family love and marijuana involvement of friends* was accounted for. Lastly, previous research suggests that there is a significant association between an adolescent student's marijuana involvement and an individual's psychological condition. The study controls in accordance with previous studies for the four complementary emotions *depressive, fearful, sad* and *lonely*.

The challenge of modeling such a wide range of risk factors is to avoid a sample size selection bias due to a pattern in missing data with regard to the predicted variable. Therefore, as a first step, attention was paid to the number of missing values and refused answers with regard to the participants' most recent marijuana consumption, the question that constitutes the basis for the construction of the *marijuana involvement* variable. However, by having a closer look at the data, it becomes apparent that only 0.3% of the participants refused to answer or did not know what to answer. As this is a marginal proportion, the threat of a selection bias on the dependent variable can be ignored.

As a second step, the sample size was narrowed down to individuals who attended school and received a full grade report. This selection is in accordance with the procedure of previous research (De La Haye et al., p. 1916). However, by making this pre-selection, the validation of the study's results will be limited to teens and young adults attending school.

Moreover, due to accurate comparison reasons, observations with missing data or refused answers for the variables *gender, neighborhood happiness, neighborhood safety, best friends marijuana consumption, family love* and *lonely* were disregarded. This was assumed to be a minor bias as the missing data or the refused answers were less than one percent for all above-mentioned variables. For the variables *teacher trouble* and *peer trouble,* there were 385 participants who were allowed to legitimately skip these questions. However, no further details were provided why these skips were legitimate. With respect to the completeness approach, the still large enough sample size and the fact of no strong pattern behind the missing data, it was decided that disregarding these data points causes the least bias compared to for example replacing the missing values by the average. In the end, the data set used for this paper's study contained 1,961 observations.

With regard to engineering the dependent variable, unlike as for an example for alcohol consumption where the WHO published accurate research what is considered heavy drinking, no such universally valid definition for regular marijuana consumption exists (World Health Organisation, 2019). Therefore, marijuana involvement was defined as being something regular and consequently, the threshold was set for students who stated that they smoked/ experimented with marijuana in the last month. This threshold is in accordance with the value used in other studies (Johnson et al., 2016, p. 583). Naturally, this definition also includes students who are coincidently first-time consumers within the last month before the questionnaire took place and never consumed marijuana ever after. However, this proportion was regarded as negligible. As a result, in this study, 15.67% of students were regarded as involved in marijuana. Taking into account the increase in marijuana involvement over the past 20 years, this number is in accordance with recent literature (Johnson et al., 2016, p. 583). Similar procedures were used to engineer the independent binary variable*s GPA*, *smoker dummy* and *heavy drinker dummy.* Details can be found in Appendix 2.

A full overview of the variables used for modeling, their summary statistics as well as a correlation table can be found in Appendix 3-4. The most interesting observations are a strong correlation between different psychological conditions as well as a strong correlation between marijuana involvement and an individual's best friends marijuana involvement. Moreover, one can see that the peak of marijuana involvement is between age 16-18 and that marijuana involvement is somehow related to other substance involvement (Appendix 5).

3 Model and Interpretation

In order to comprehensively assess the association between adolescent student marijuana involvement and other substance involvement, a range of logistic regression models was constructed. As a first step, a model consisting of only the three other-substance-involvement variables *smoker, other drug involvement* and *heavy drinker* was made. This model explains approximately 20% of the variance in marijuana involvement ($Pseudo\ R^2 = 0.21$) while all three predictors are all highly statistically significant and positively associated with marijuana involvement (p-values<2e-10) as outlined under Appendix 6 and 12. According to this simple model which does not control for any other potential influencing factors and is therefore strongly biased, smoking increases the probability of marijuana involvement by 16.67%, heavy drinking by 13.65% and other drug involvement by 17.56%.

By separately adding the other three groups of control variables to the model, *smoking, heavy drinking* and *other drug involvement* all remain statistically significant predictors (at least p-values<1e-2) while their effect on marijuana involvement decrease as outlined in Appendix 7-9 and Appendix 12. This could be expected when controlling for other explanatory factors. However, the change when adding demographic and psychological factors is marginal (minus 0-1%) compared to the decrease when adding the group of social norm variables (minus 12-16%). These results suggest that the demographic and psychological factors in the model have little explanatory power compared to the other two groups of explanatory variables. Hence, social norms and other substance involvement appear to be the most powerful associations of marijuana involvement. As a next step, the goal is to confirm these assumptions with a complete model and to correctly quantify the effects with a correctly specified model of high explanatory power.

In the complete model, the other substance involvement variables all remain highly statistically significant (p-values<1e-3) and are the second to fourth most powerful explanatory factors. However, the probability association of the other substance involvement variables on marijuana involvement drops to 0.88-3.11% (Appendix 10 and 12). This is most likely due to the noise caused by a significant number of irrelevant predictors within the model that potentially distort the true impact of the coefficients. This can be solved by constructing a correctly specified model with high explanatory power. For this cause, the forward and backward variable selection algorithm was used. The final model consists of the three other substance involvement variables plus the variables *best friends marijuana involvement, teacher trouble, fear and suspension.* Interaction terms did not appear to have any relevance. In this final model *heavy drinking* is associated with a probability effect on marijuana involvement of 1.80%, *other substance involvement* of 1.40% and *smoking* of 2.17% (Appendix 11 and 12). These numbers are likely the most precise estimations of the other substance involvement effects. However, it is important to see that the direction of the effects remains unclear. More precisely, with these models, one cannot say whether, for example, other substance involvement or marijuana consumption of friends causes marijuana involvement or vice versa. The same applies to the other variables. However, there it is easier to find a logic behind, such as marijuana involvement leads more likely to suspension and teacher trouble than the other way around.

4 Conclusion

The result of this paper's analysis shows that there is a significant, positive association between marijuana involvement and smoking, heavy alcohol consumption and other illegal drug involvement when comprehensively controlling for other potential explanatory factors. However, the exact direction of the effects remains unclear and would be subject to further exploratory research. Nevertheless, for governments and health-related organisations, it is important to see that marijuana legalisation and the consequential increase in marijuana involvement of adolescent students, is somehow related to other substance involvement which subsequently can be connected to serious health consequences for an individual as well as to a significant increase in health costs for a society (Gryczynski et al. 2016, p.16). It is important to mention that the study does not come without limitations. First of all, the results only apply to enrolled adolescents in the US. Moreover, the sample size is randomly drawn and the sample size weight was specifically and exclusively controlled for the correct representation of marijuana involvement. However, except for gender, the weights of the sample size are highly accurate as one can see in Appendix 13. Secondly, social norms and socioeconomic status could potentially, with for example data from other waves, be modeled more accurately with regard to household income, family relationships, social network and love relationships. As a consequence, this could potentially have an effect on the variable selection of the final model and on the explanatory power of the effects.

References

De La Haye, K., Green, H.D., Pollard, M.S., Kennedy, D.P., and Trucker, J.S., 2015. Befriending risky peers: factors driving adolescents' selection with similar marijuana use. *Journal of youth and adolescence, 44(10),* pp. 1914-1928. doi: https://doi.org/10.1007/s10964-014-0210-z

Gryczynski J., Schwartz, R.P., O'Grady, K.E., Restivo, L., Mithcell, S.G. and Jaffe, J.H, 2016. Understanding patterns of high-cost health care use across different substance user groups. *Health affairs, 35*(1), pp. 12-19 doi: https://doi.org/10.1377/hlthaff.2015.0618

Johnson, R.M., Brooks-Russell, A., Ma, M., Fairman, B.J., Tolliver Jr., R.L. and Levinson, A.H., 2016. Usual models of marijuana consumption among high school students in Colorado. *Journal of studies,* 77(4), pp. 580–588.

Mahalik, J.R., Lombardi, C.M., Sims, J., Coley, R.L., Lynch, A.D., 2015. Gender, male-typicality, and social norms predicting adolescent alcohol intoxication and marijuana use. *Social science & medicine, 143*, pp. 71–80.

Van den Bree, M.B. and Pickworth, W.B., 2005. Risk factors predicting changes in marijuana involvement in teenagers. *Archives of general psychiatry, 62(3)*, pp. 311–319. doi: 10.1001/archpsyc.62.3.311.

World Health Organisation, 2019, *Heavy episodic drinking among drinkers*, viewed 18 May 2019,https://www.who.int/gho/alcohol/consumption_patterns/heavy_episodic_drinkers_text/en/

Zuckermann, A.M., Battista, K., de Groh, M., Jiang, Y., & Leatherdale, S.T., 2019. Prelegalisation patterns and trends of cannabis use among Canadian youth: results from the COMPASS prospective cohort study. *BMJ open, 9*(3), pp. 1-9. doi: https://dx.doi.org/10.1136/bmjopen-2018-026515

Appendix

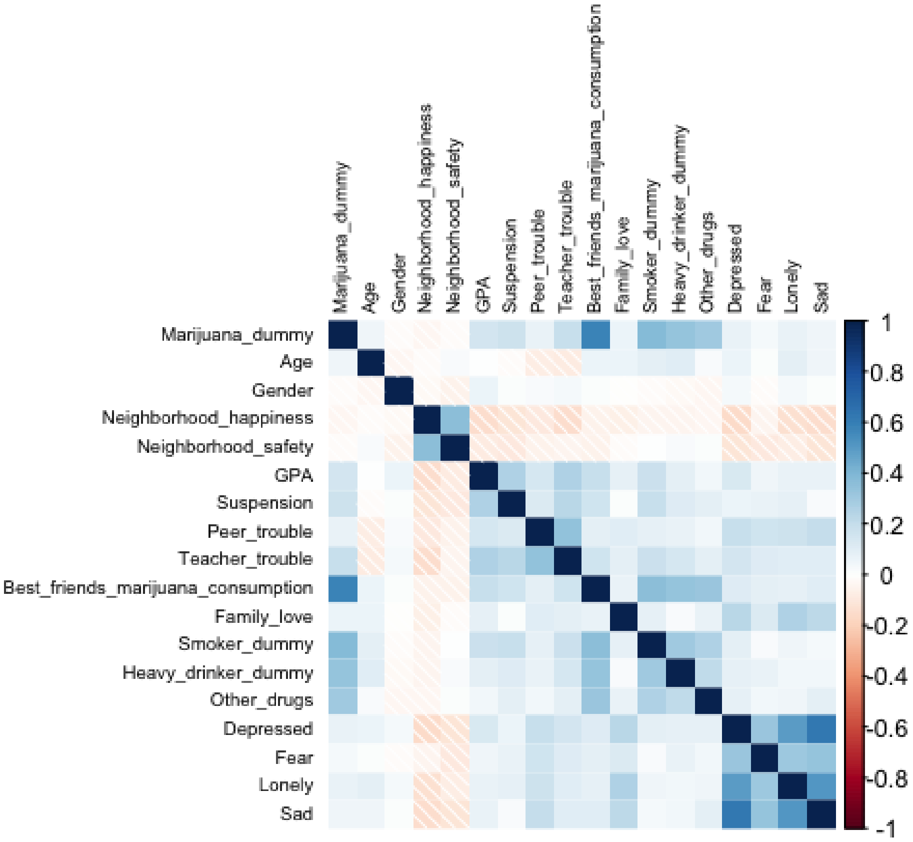
Appendix 1: Considered Variables

Name	Code	Information	Group
Marijuana	H2TO47	Most recent marijuana consumption (12 points in time)	Dep. variable
Age	CALCAGE2	Age of survey participant (11-21)	demographic
Gender	H2HR3A	Sex of the survey participant (binary)	demographic
Neighborhood happiness	H2NB6	Happiness about living in own neighborhood (1-5)	demographic
Neighborhood safety	H2NB5	Safety sense about living in own neighborhood (1-5)	demographic
Grade English	H2ED7	Grade in English class (1-4)	social
Grade Math	H2ED8	Grade in Math class (1-4)	social
Grade History	H2ED9	Grade in History class (1-4)	social
Grade Science	H2ED10	Grade in Science class (1-4)	social
Suspension	H2ED3	Suspension from School (binary)	social
Peer trouble	H2ED14	Trouble with other students (0-4)	social
Teacher trouble	H2ED11	Trouble with teachers (0-4)	social
Best friends marijuana Invovement	H2TO48	Marijuana involvement of the best three friends (1-3)	social
Family love	H2PF27	Feeling loved by your family (1-5)	social
Smoking	H2TO8	Last time smoked a cigarette (7 points in time)	substance
Alcohol	H2TO21	Frequency of 5 or more glasses per day in the last 12 months ()	substance
Other Drug Involvement	H2TO58	Tried other illegal drugs e.g. LSD, speed, ecatasy etc (binary)	substance
Depressed	H2FS6	Feeling depressed (0-3)	psycological
Fear	H2FS10	Feeling fear (0-3)	psycological
Lonely	H2FS13	Feeling lonely (0-3)	psycological
Sad	H2FS16	Feeling Sad (0-3)	psycological

Appendix 2: Feature Engineering

New Variable	Basis Variable	Information	Category
Marijuana dummy	Marijuana	marijuana consumption within the last month	binary
GPA	Grade History Grade Math Grade Science Grade English	Sum of all grades divided by four	Within range 1-4
Smoker dummy	Smoking	Regular Smoker, smoked cigarettes today or yesterday	binary
Heavy drinking dummy	Alcohol	At least 2-3 per month drinking more than 5 glasses a day[1]	binary

Appendix 3: Correlation Table

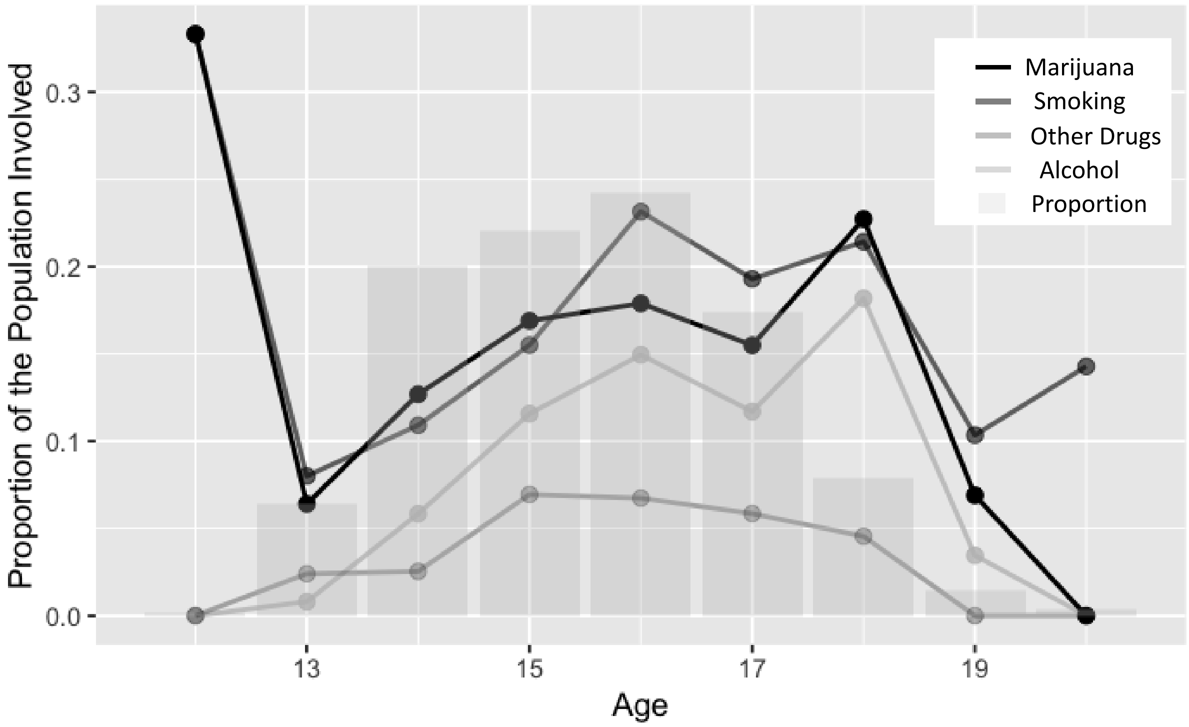


[1] Benchmark: National Center for Biotechnology Information, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6104966/

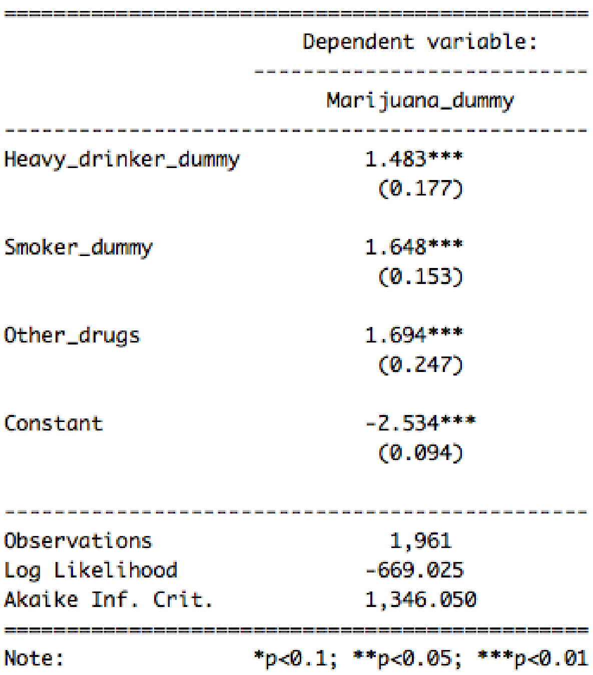
Appendix 4: Summary Statistics

Variable	Min	Max	Mean	Std Dev
Marijuana dummy	0	1	0.16	0.36
Age	12	20	15.57	1.46
Gender	0	1	0.65	0.48
Neighborhood happiness	0	3	2.37	0.82
Neighborhood safety	0	1	0.86	0.34
GPA	0	3	1.16	0.75
Suspension	0	1	0.11	0.32
Peer trouble	0	4	0.87	0.91
Teacher trouble	0	4	0.85	0.90
Best friends marijuana invovement	0	3	0.66	0.98
Family love	0	4	0.69	0.69
Smoker dummy	0	1	0.17	0.37
Heavy drinker dummy	0	1	0.11	0.31
Other drug involvement	0	1	0.05	0.22
Depressed	0	3	0.51	0.74
Fear	0	3	0.29	0.54
Lonely	0	3	0.45	0.69
Sad	0	3	0.59	0.68

Appendix 5: Selected EDA insights

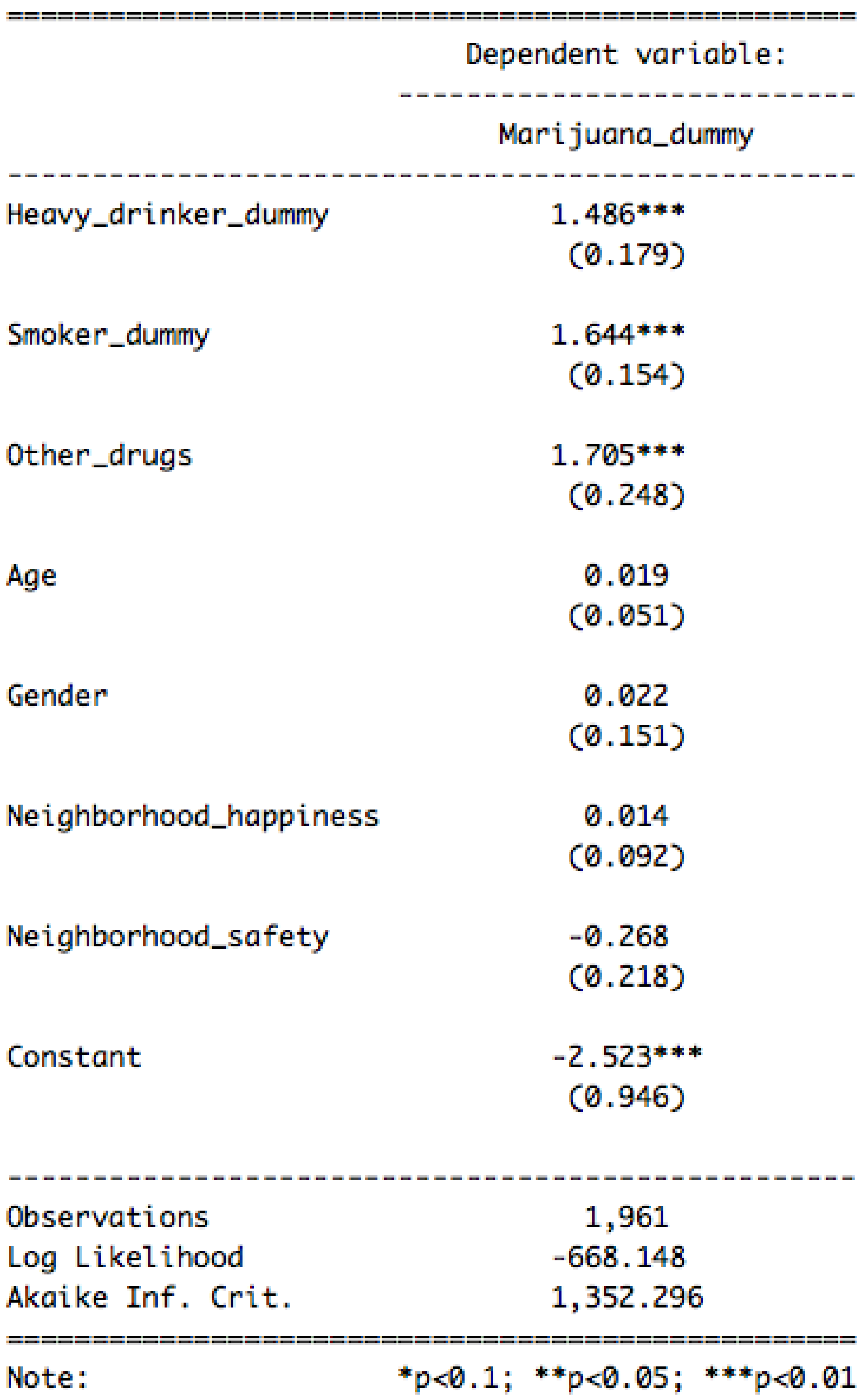


Appendix 6: Model 1 – Substance Involvement



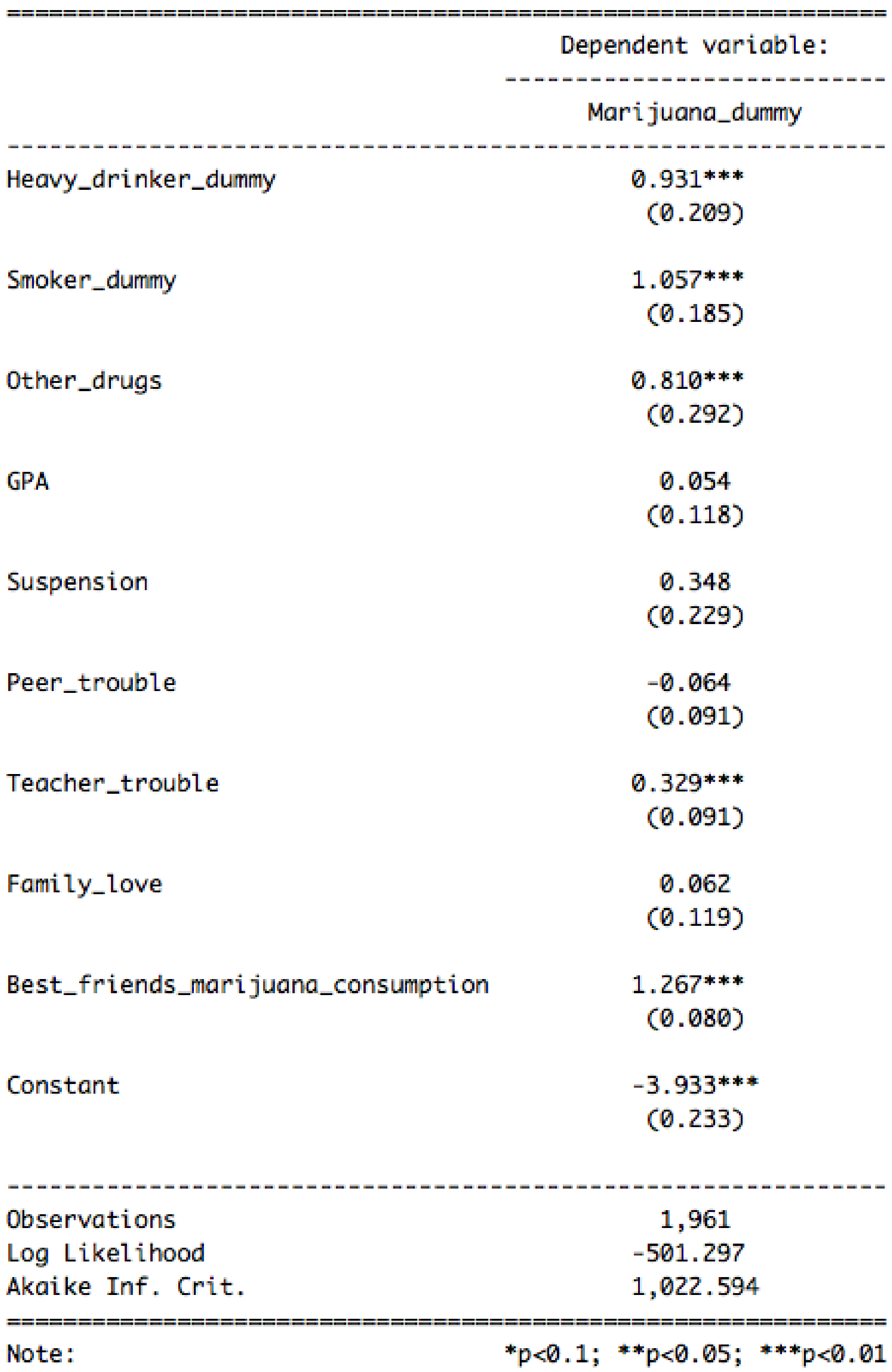
```
================================================
                        Dependent variable:
                    ----------------------------
                          Marijuana_dummy
------------------------------------------------
Heavy_drinker_dummy          1.483***
                              (0.177)

Smoker_dummy                 1.648***
                              (0.153)

Other_drugs                  1.694***
                              (0.247)

Constant                     -2.534***
                              (0.094)

------------------------------------------------
Observations                   1,961
Log Likelihood               -669.025
Akaike Inf. Crit.            1,346.050
================================================
Note:                 *p<0.1; **p<0.05; ***p<0.01
```

Appendix 7: Model 2 – Substance Involvement + First Control Group



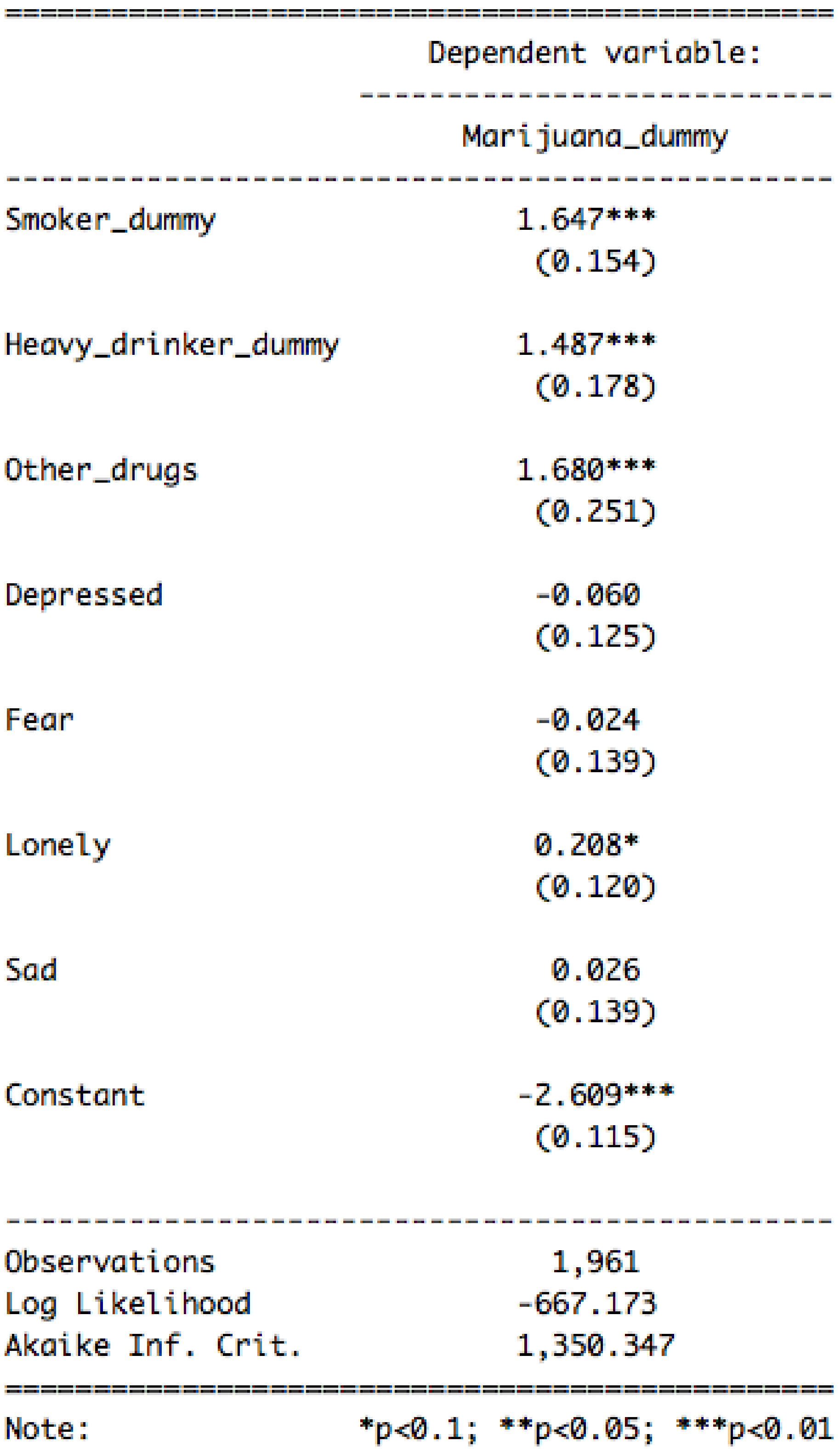
```
==================================================
                        Dependent variable:
                    ---------------------------
                          Marijuana_dummy
--------------------------------------------------
Heavy_drinker_dummy           1.486***
                              (0.179)

Smoker_dummy                  1.644***
                              (0.154)

Other_drugs                   1.705***
                              (0.248)

Age                            0.019
                              (0.051)

Gender                         0.022
                              (0.151)

Neighborhood_happiness         0.014
                              (0.092)

Neighborhood_safety           -0.268
                              (0.218)

Constant                      -2.523***
                              (0.946)

--------------------------------------------------
Observations                   1,961
Log Likelihood               -668.148
Akaike Inf. Crit.            1,352.296
==================================================
Note:                 *p<0.1; **p<0.05; ***p<0.01
```

Appendix 8: Model 3 – Substance Involvement + Second Control Group



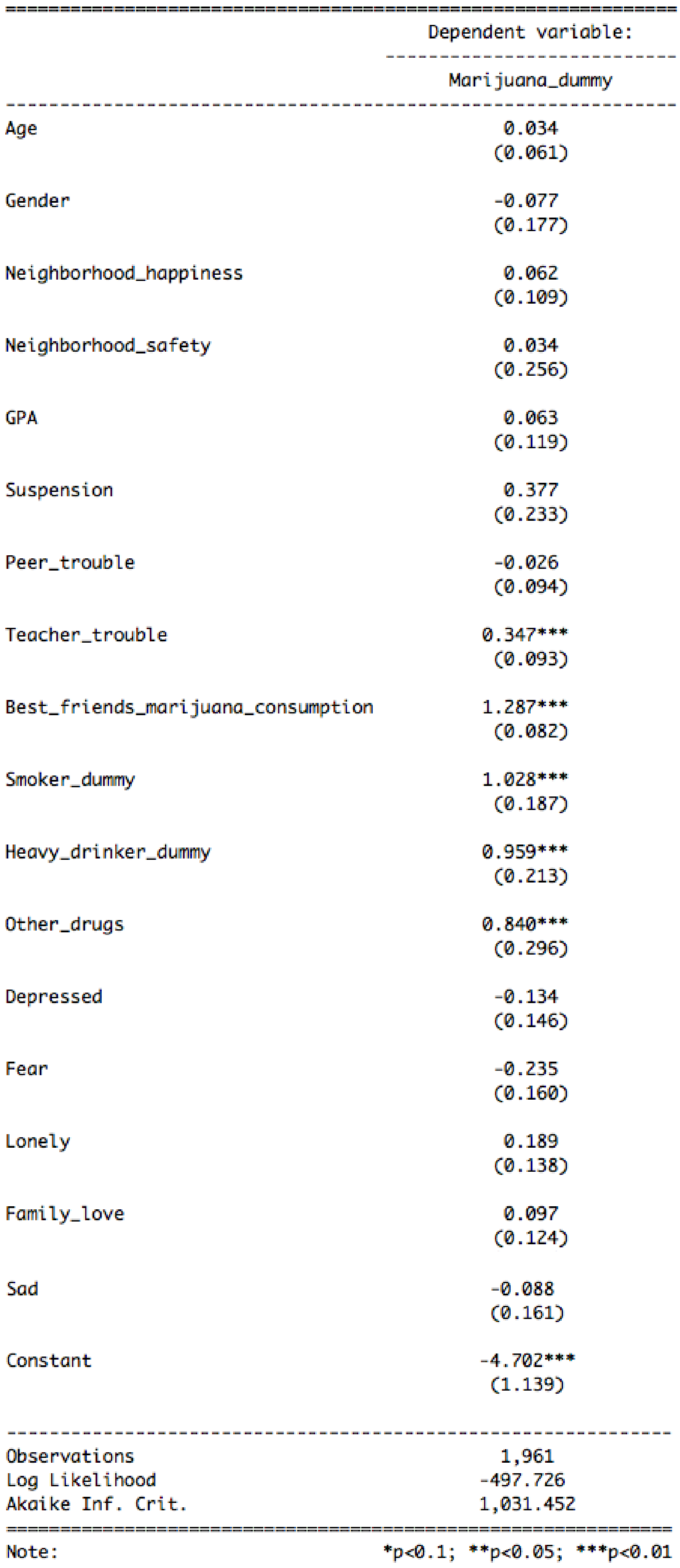
```
==============================================================
                                      Dependent variable:
                                  ---------------------------
                                        Marijuana_dummy
--------------------------------------------------------------
Heavy_drinker_dummy                         0.931***
                                            (0.209)

Smoker_dummy                                1.057***
                                            (0.185)

Other_drugs                                 0.810***
                                            (0.292)

GPA                                          0.054
                                            (0.118)

Suspension                                   0.348
                                            (0.229)

Peer_trouble                                -0.064
                                            (0.091)

Teacher_trouble                             0.329***
                                            (0.091)

Family_love                                  0.062
                                            (0.119)

Best_friends_marijuana_consumption          1.267***
                                            (0.080)

Constant                                   -3.933***
                                            (0.233)

--------------------------------------------------------------
Observations                                 1,961
Log Likelihood                             -501.297
Akaike Inf. Crit.                          1,022.594
==============================================================
Note:                              *p<0.1; **p<0.05; ***p<0.01
```

Appendix 9: Model 4 – Substance Involvement + Third Control Group



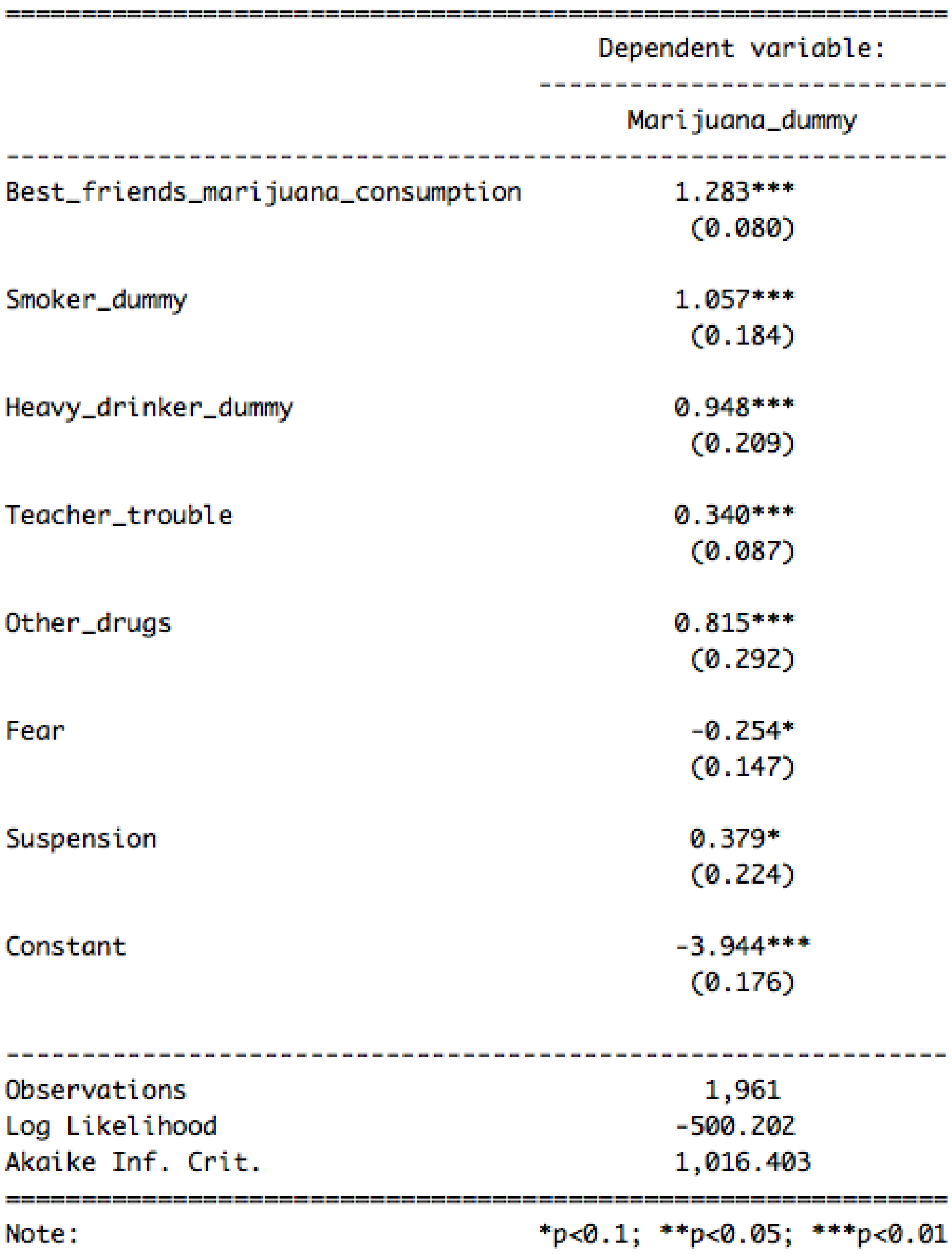
```
===============================================
                        Dependent variable:
                    ---------------------------
                          Marijuana_dummy
-----------------------------------------------
Smoker_dummy                 1.647***
                              (0.154)

Heavy_drinker_dummy          1.487***
                              (0.178)

Other_drugs                  1.680***
                              (0.251)

Depressed                     -0.060
                              (0.125)

Fear                          -0.024
                              (0.139)

Lonely                         0.208*
                              (0.120)

Sad                            0.026
                              (0.139)

Constant                     -2.609***
                              (0.115)

-----------------------------------------------
Observations                   1,961
Log Likelihood               -667.173
Akaike Inf. Crit.            1,350.347
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

Appendix 10: Model 5 - Complete Model



```
================================================================
                                        Dependent variable:
                                    ----------------------------
                                          Marijuana_dummy
----------------------------------------------------------------
Age                                           0.034
                                             (0.061)

Gender                                       -0.077
                                             (0.177)

Neighborhood_happiness                        0.062
                                             (0.109)

Neighborhood_safety                           0.034
                                             (0.256)

GPA                                           0.063
                                             (0.119)

Suspension                                    0.377
                                             (0.233)

Peer_trouble                                 -0.026
                                             (0.094)

Teacher_trouble                             0.347***
                                             (0.093)

Best_friends_marijuana_consumption          1.287***
                                             (0.082)

Smoker_dummy                                1.028***
                                             (0.187)

Heavy_drinker_dummy                         0.959***
                                             (0.213)

Other_drugs                                 0.840***
                                             (0.296)

Depressed                                    -0.134
                                             (0.146)

Fear                                         -0.235
                                             (0.160)

Lonely                                        0.189
                                             (0.138)

Family_love                                   0.097
                                             (0.124)

Sad                                          -0.088
                                             (0.161)

Constant                                    -4.702***
                                             (1.139)

----------------------------------------------------------------
Observations                                  1,961
Log Likelihood                               -497.726
Akaike Inf. Crit.                            1,031.452
================================================================
Note:                                 *p<0.1; **p<0.05; ***p<0.01
```

Appendix 11: Model 6- Final Model



```
==============================================================
                                        Dependent variable:
                                   ---------------------------
                                          Marijuana_dummy
--------------------------------------------------------------
Best_friends_marijuana_consumption          1.283***
                                             (0.080)

Smoker_dummy                                1.057***
                                             (0.184)

Heavy_drinker_dummy                         0.948***
                                             (0.209)

Teacher_trouble                             0.340***
                                             (0.087)

Other_drugs                                 0.815***
                                             (0.292)

Fear                                         -0.254*
                                             (0.147)

Suspension                                   0.379*
                                             (0.224)

Constant                                    -3.944***
                                             (0.176)

--------------------------------------------------------------
Observations                                  1,961
Log Likelihood                              -500.202
Akaike Inf. Crit.                           1,016.403
==============================================================
Note:                              *p<0.1; **p<0.05; ***p<0.01
```

Appendix 12: Marginal Probability Effect (%) associated with marijuana invovment and Pseudo-R^2 [2]

Model	Drinking	Smoking	Other Drugs	Pseudo-R^2 [3]
Model 1 – Other Substance	13.65	16.67	17.56	0.21
Model 2 – Control Group 1	13.81	16.71	17.92	0.21
Model 3 - Control Group 2	1.76	2.09	1.40	0.41
Model 4 – Control Group 3	12.98	15.79	16.41	0.22
Model 5 - Complete	0.88	3.11	2.50	0.42
Model 6 - Final	1.80	2.17	1.4	0.41

Appendix 13: Population Representation vs Sample Representation

Variable	Model 1997 (%)	Real World (%)
Marijuana involvement	17	~22[4]
Heavy drinking	11	~11[5]
Other illegal drugs	5	~4[6]
Smoking	16	~16[7]
Male	65	~ 50

[2] Calculation Reference: https://sebastiansauer.github.io/convert_logit2prob/

[3] McFadden Psuedo-R-Squared

[4] National Center for Biotechnology Information, 2016, accurate when accounting for increase in marijuana involvement, https://www.ncbi.nlm.nih.gov/pubmed/27340962

[5] National Center for Biotechnology Information, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6104966/

[6] US Department of Health & Human Services, 2016, https://www.hhs.gov/ash/oah/adolescent-development/substance-use/drugs/opioids/index.html

[7] US Department of Health & Human Services, 1996 Data, https://www.hhs.gov/ash/oah/adolescent-development/substance-use/drugs/tobacco/trends/index.html

Code

R Notebook

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

Hide

```
library(ggplot2)
library(dplyr)
library(tidyverse)
library("Hmisc")
library(car)
library(BaylorEdPsych)
library(aod)
```

Hide

```
setwd("~/Desktop")
```

```
The working directory was changed to /Users/Cyrill/Desktop inside a notebook chunk. The working directory wi
ll be reset when the chunk is finished running. Use the knitr root.dir option in the setup chunk to change t
he working directory for notebook chunks.
```

Hide

```
df <- read.csv('/Users/Cyrill/Desktop/ds8.csv')
df1 <- read.csv('/Users/Cyrill/Desktop/df1.csv')
df2 <- read.csv('/Users/Cyrill/Desktop/data_w2.csv')
```

COUNT NUMBER OF OBSEVATIONS

Hide

```
nrow(df2)
```

```
[1] 4834
```

RENAME COLUMNS

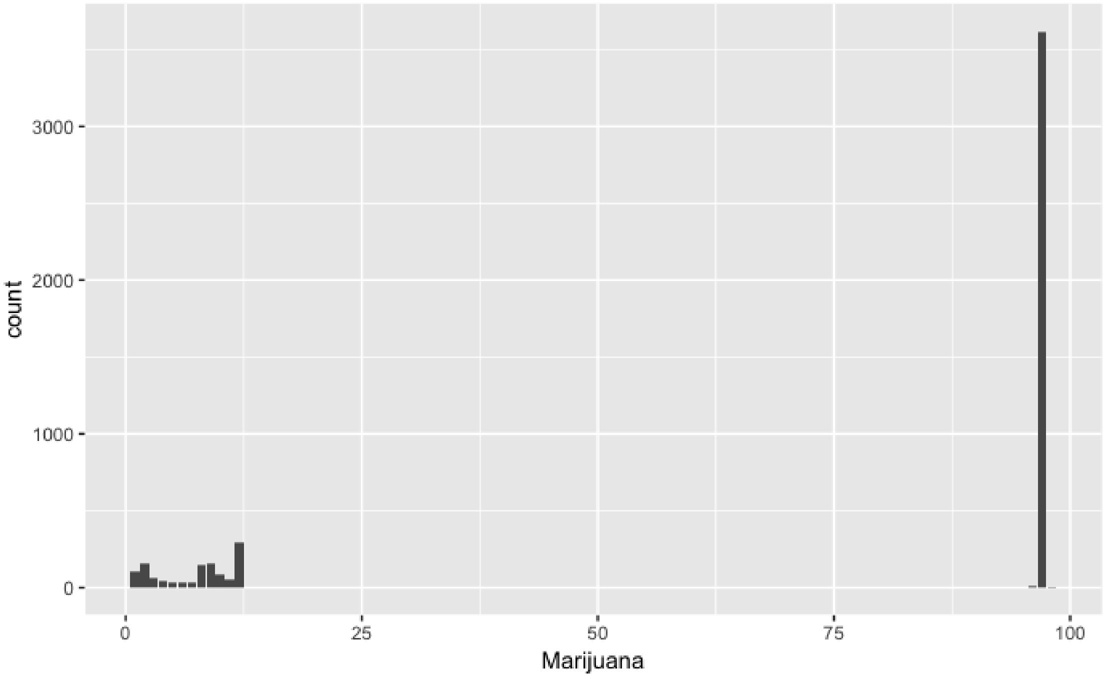
Hide

```
df2 <- rename(df2, Marijuana = H2TO47 )
df2 <- rename(df2, Age = CALCAGE2)
df2 <- rename(df2, Gender= H2HR3A )
df2 <- rename(df2, Neighborhood_happiness = H2NB6)
df2 <- rename(df2, Neighborhood_safety = H2NB5 )
df2 <- rename(df2, English = H2ED7)
df2 <- rename(df2, Science = H2ED10)
df2 <- rename(df2, Math = H2ED8)
df2 <- rename(df2, History = H2ED9)
df2 <- rename(df2, Suspension= H2ED3 )
df2 <- rename(df2, Peer_trouble = H2ED14 )
df2 <- rename(df2, Teacher_trouble = H2ED11 )
df2 <- rename(df2, Best_friends_marijuana_consumption = H2TO48 )
df2 <- rename(df2, Smoking = H2TO8 )
df2 <- rename(df2, Drinking = H2TO21 )
df2 <- rename(df2, Other_drugs = H2TO58 )
df2 <- rename(df2, Depressed = H2FS6 )
df2 <- rename(df2, Fear = H2FS10 )
df2 <- rename(df2, Lonely= H2FS13 )
df2 <- rename(df2, Sad= H2FS16  )
df2 <- rename(df2, Family_love= H2PF27 )
```

EDA OF CRITICAL VARIABLES

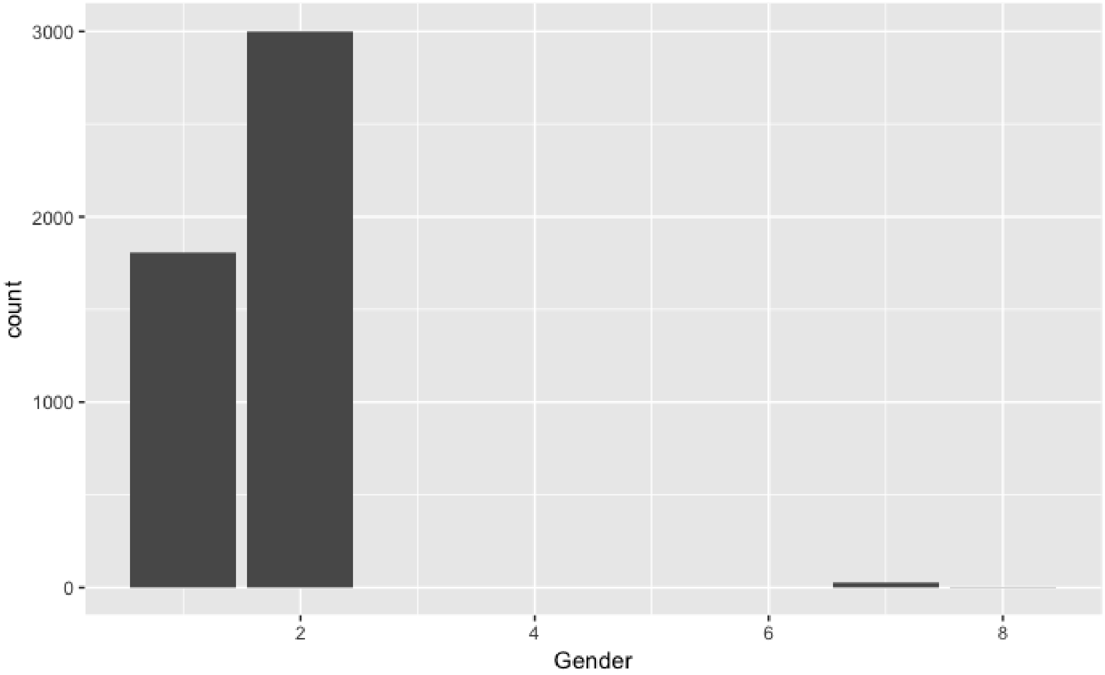
Hide

```
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Marijuana))
```



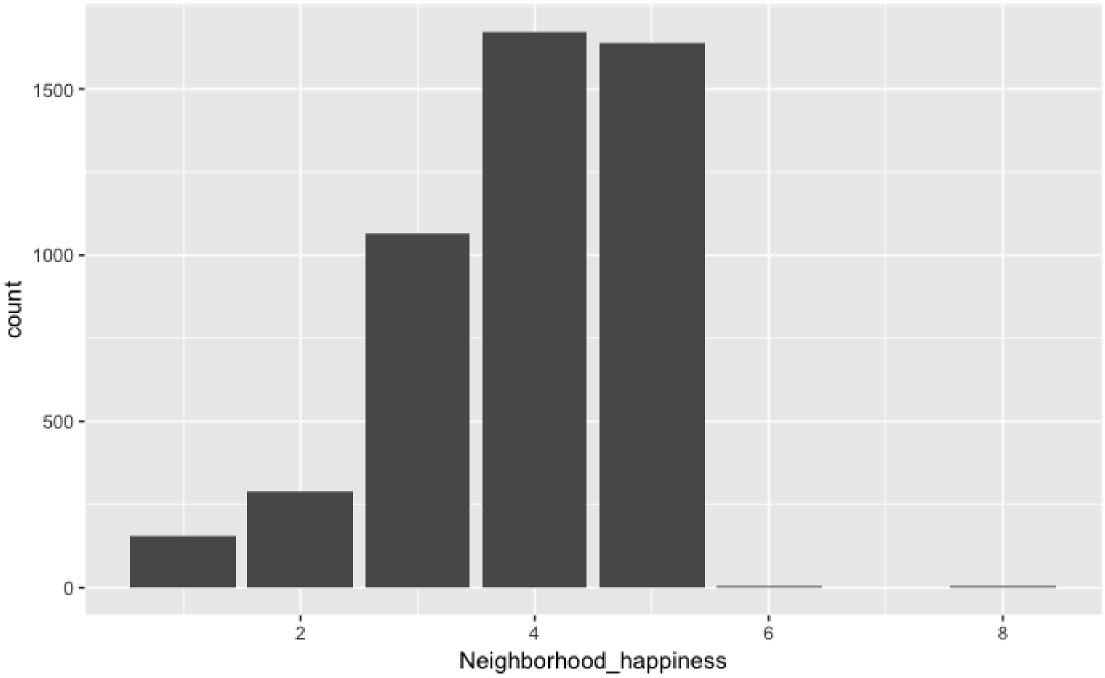
Hide

```
#gender wave 2 | indendent variable 2 unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Gender))
```



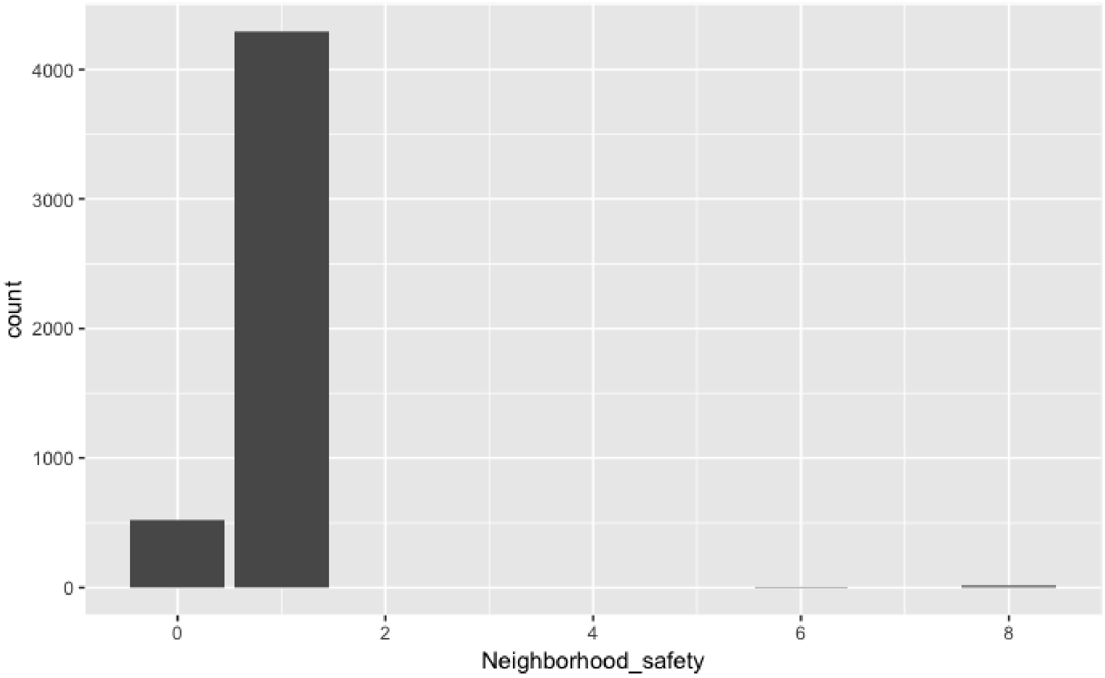
Hide

```
#neighborhood (=happy to live in this neighborhood) wave 2| indendent variable 4a unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Neighborhood_happiness ))
```



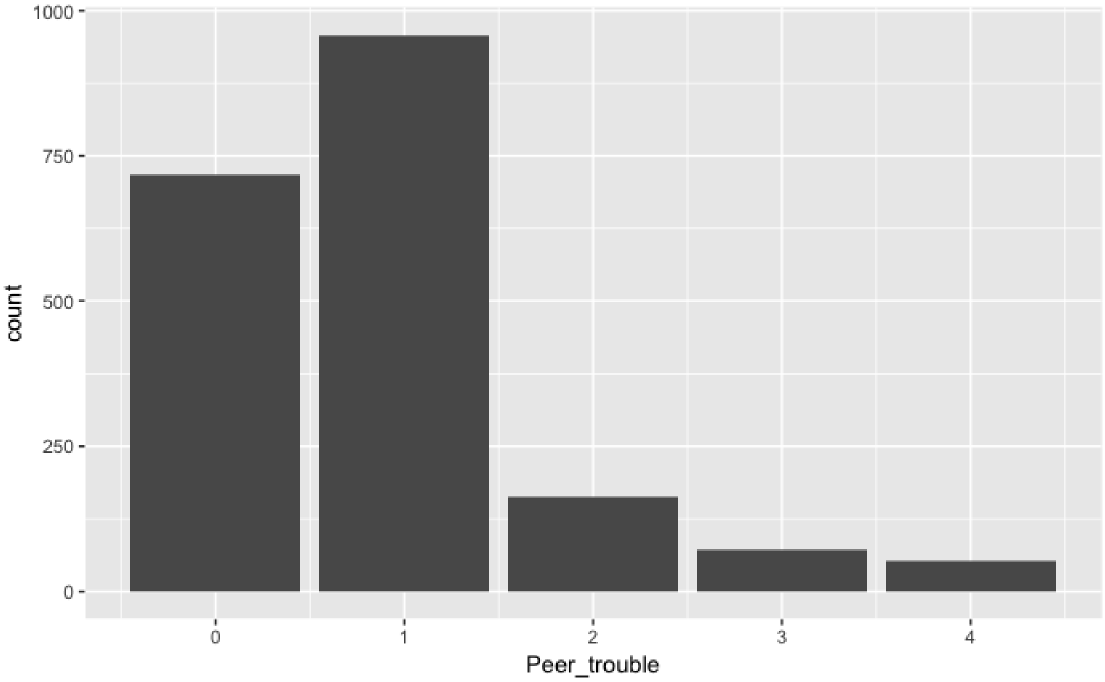
Hide

```
#neighborhood (=feel save in this neighborhood) wave 2 | indendent variable 4b unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Neighborhood_safety ))
```



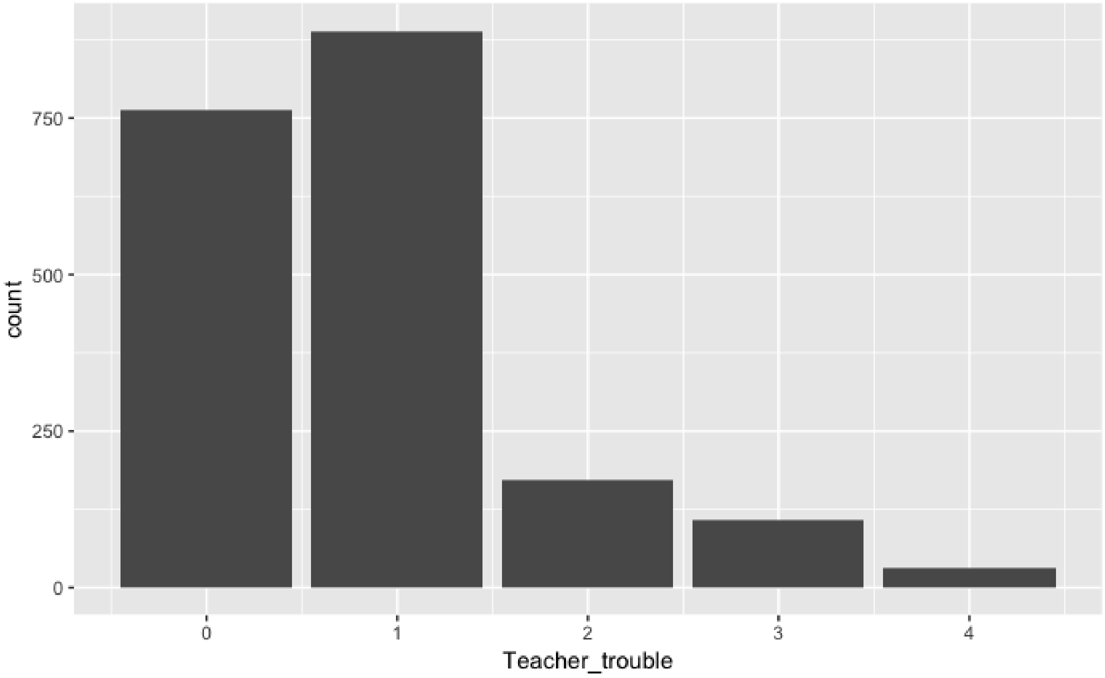
Hide

```
#trouble with peers wave 2 | indendent variable 9 unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Peer_trouble ))
```



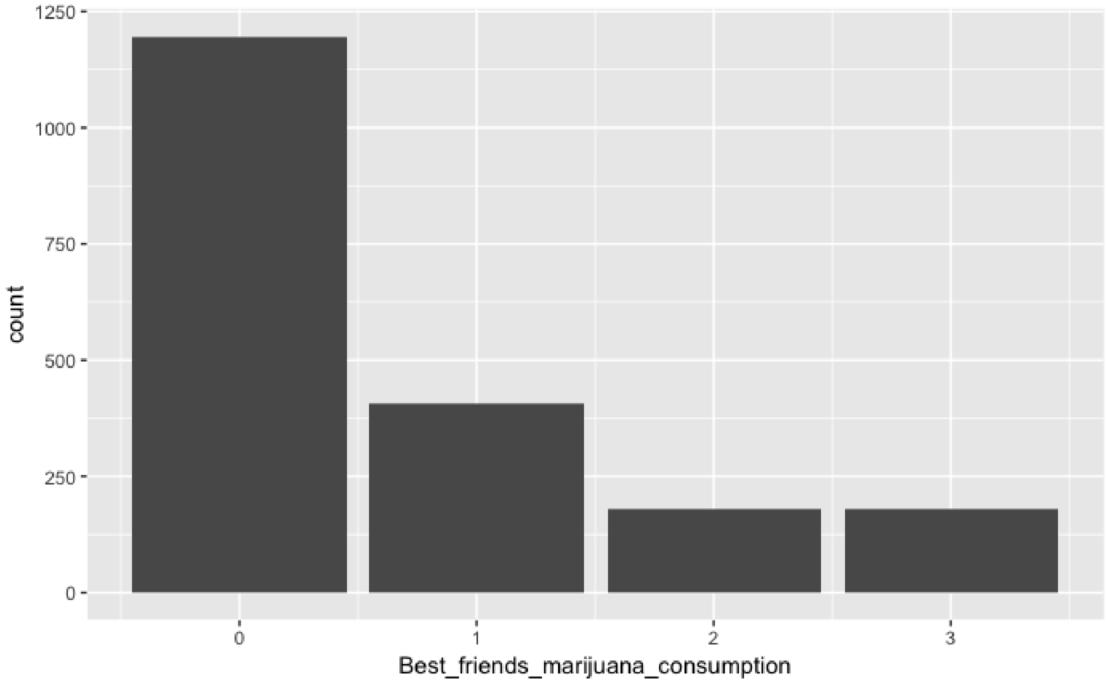
Hide

```
#trouble with teacher wave 2 | indendent variable 10 unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Teacher_trouble ))
```



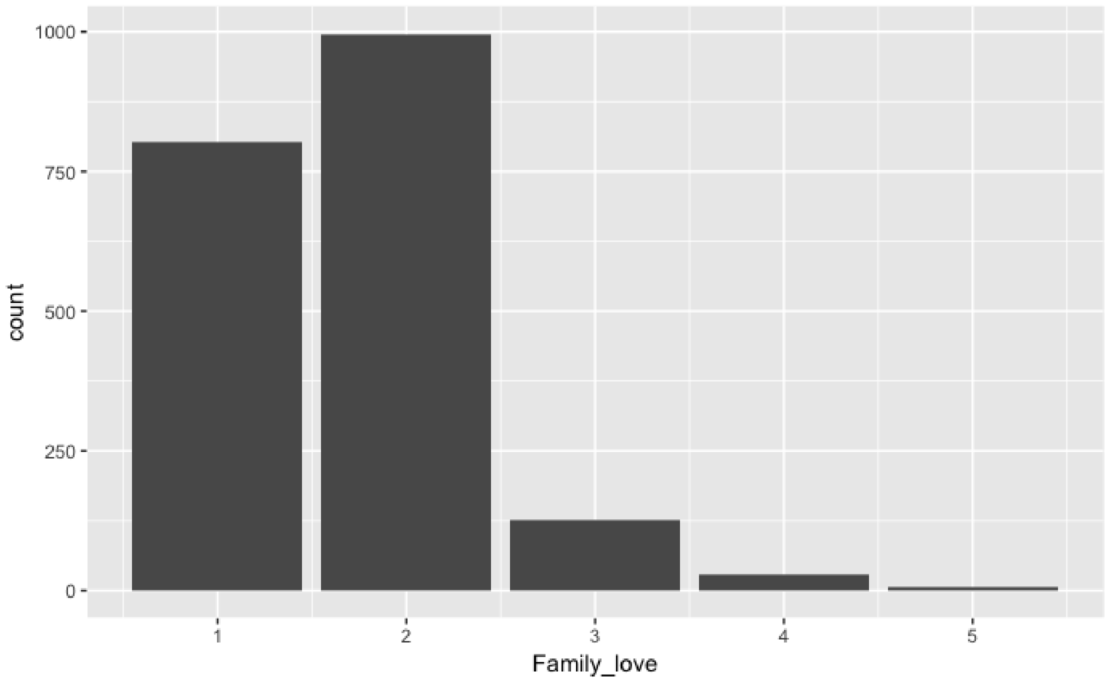
Hide

```
#Friends marijuana involvement wave 2 | indendent variable 11 unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Best_friends_marijuana_consumption ))
```



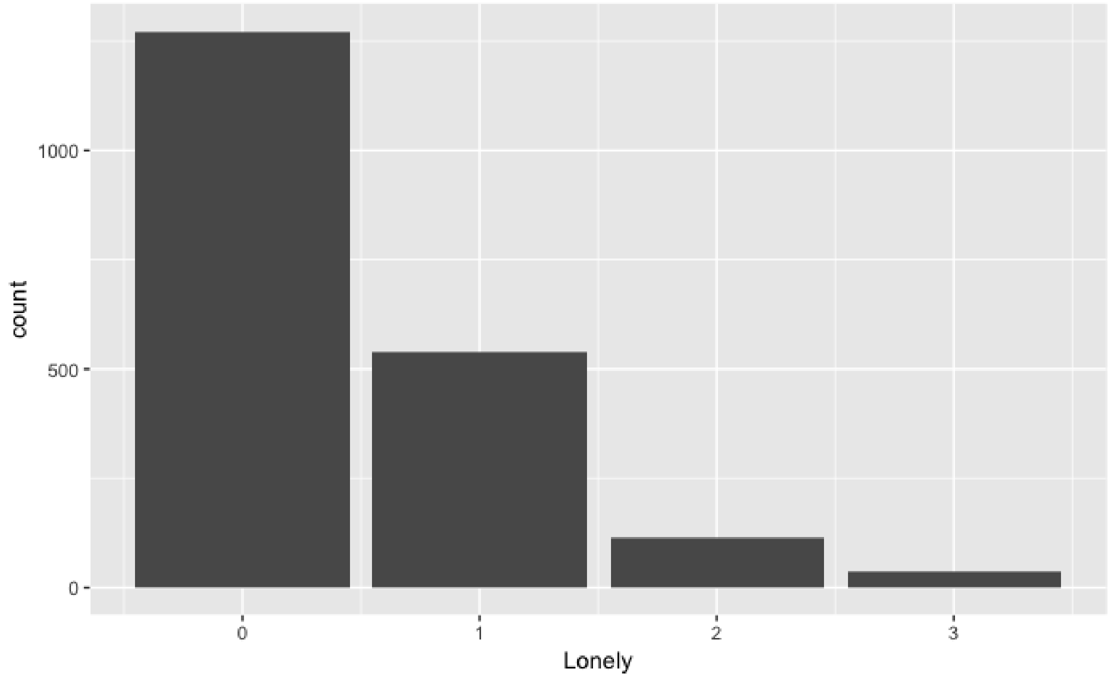
Hide

```
#feeling loved by family (5 stages) | indendent variable 20 unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Family_love))
```



Hide

```
#feeling loney (4 stages) | indendent variable 18 unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Lonely ))
```



DECREASE IN SAMPLE SIZE DUE TO SCHOOL ENROLLMENT AND FULL GRADE REPORT AND FILTER NA

Hide

```
#Filter wether school is signifacnt
df2<-df2 %>%
  filter(English<=4 & Science<=4 & Math<=4 & History<=4 )
```

Hide

```
nrow(df2)
```

```
[1] 1961
```

NUMBER OF OBSEVATIONS

Hide

```
df2<-df2 %>%
  filter(English<=4 & Science<=4 & Math<=4 & History<=4 &
           Gender<=2 &
           Neighborhood_happiness<=4 &
           Neighborhood_safety<=2 &
           Peer_trouble <=4&
           Best_friends_marijuana_consumption <=3 &
           Other_drugs<=2 &
           Lonely<=5&
           Family_love<=5)
```

DEPENDENT VARIABLE

Hide

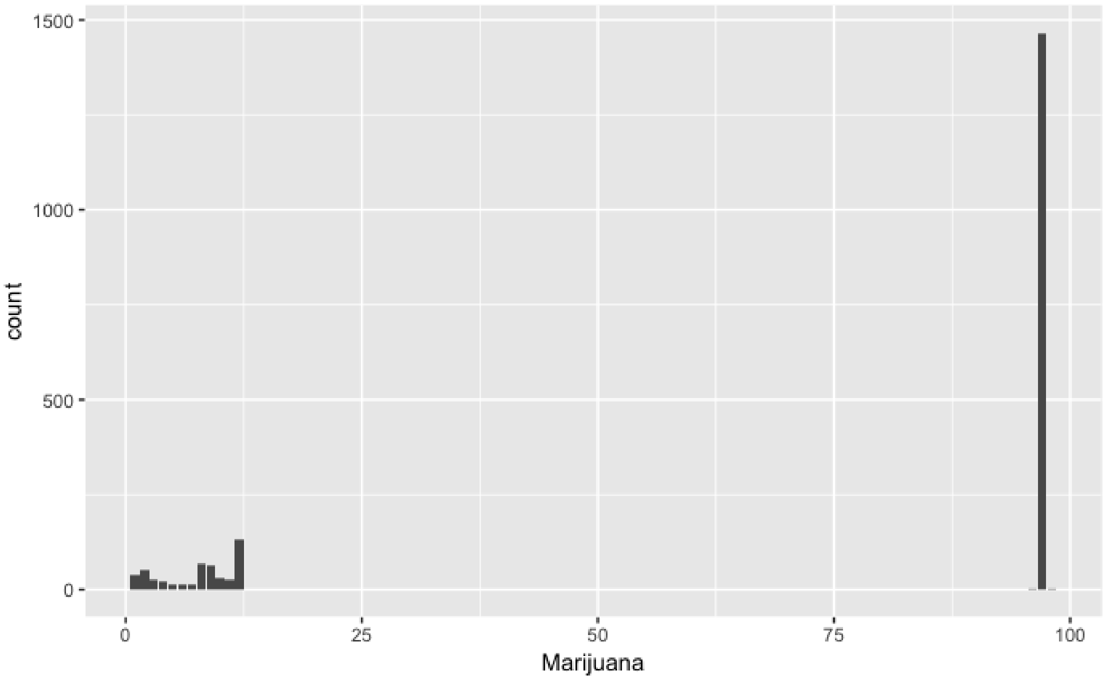
```
nrow(df2)
```

```
[1] 1961
```

CONTROL GROUP 1 (BASICS DEMOGRAPHIC)

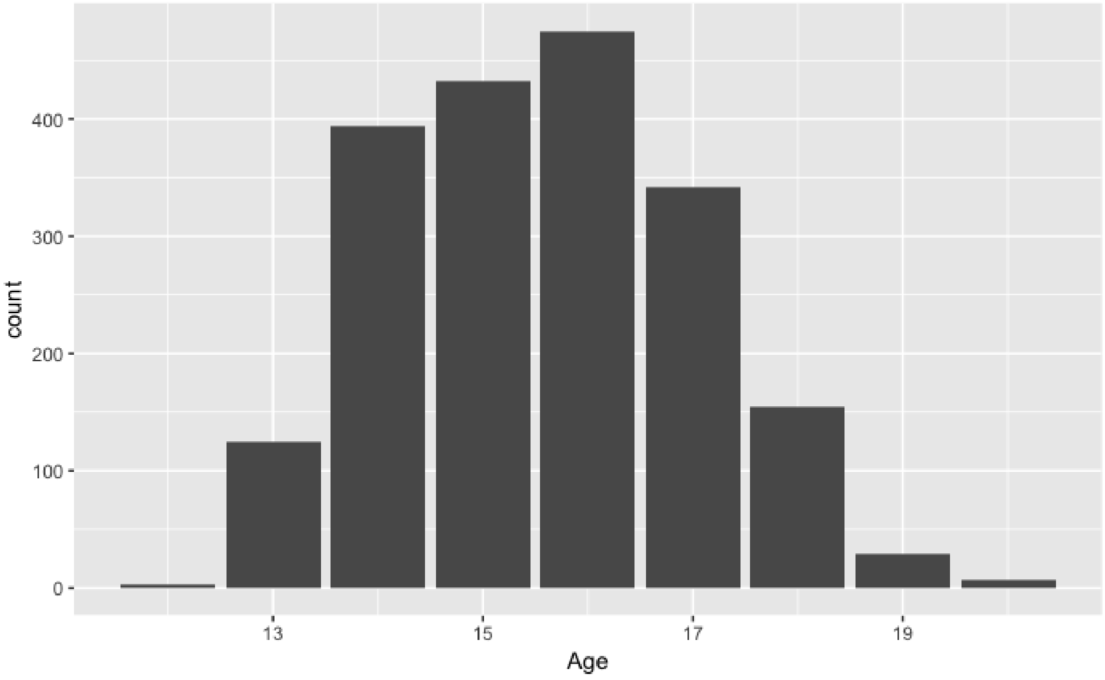
Hide

```
#marijuana consumption wave 2 | Dependent Variable unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Marijuana))
```



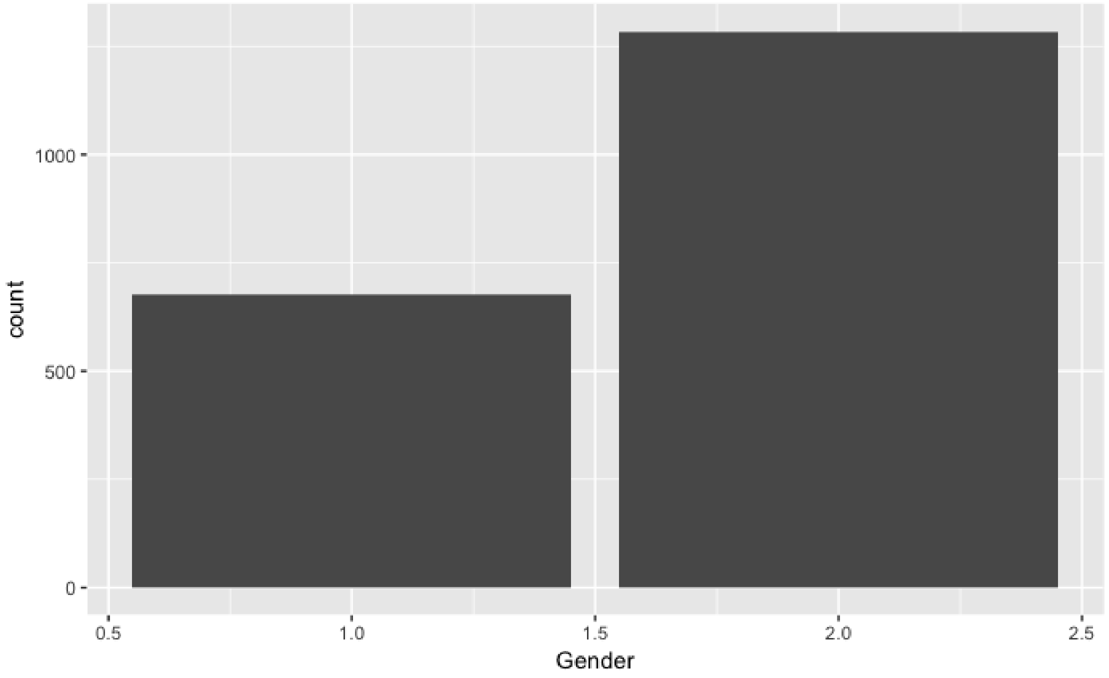
Hide

```
#age wave 2 | indendent variable 1a unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Age))
```



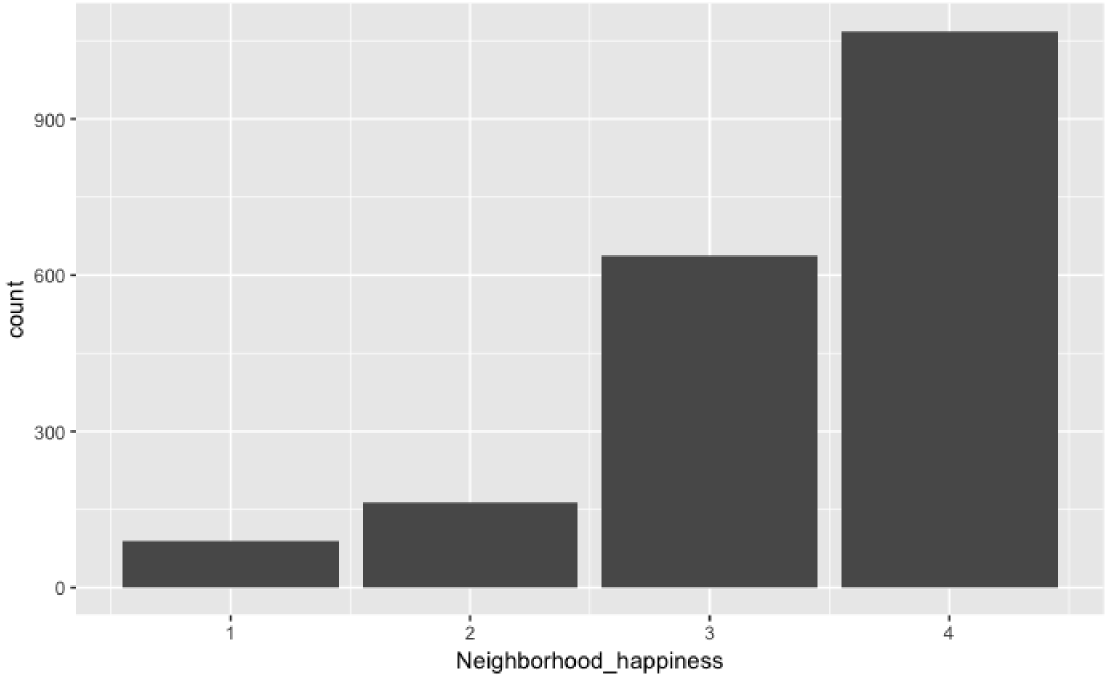
Hide

```
#gender wave 2 | indendent variable 2 unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Gender))
```



Hide

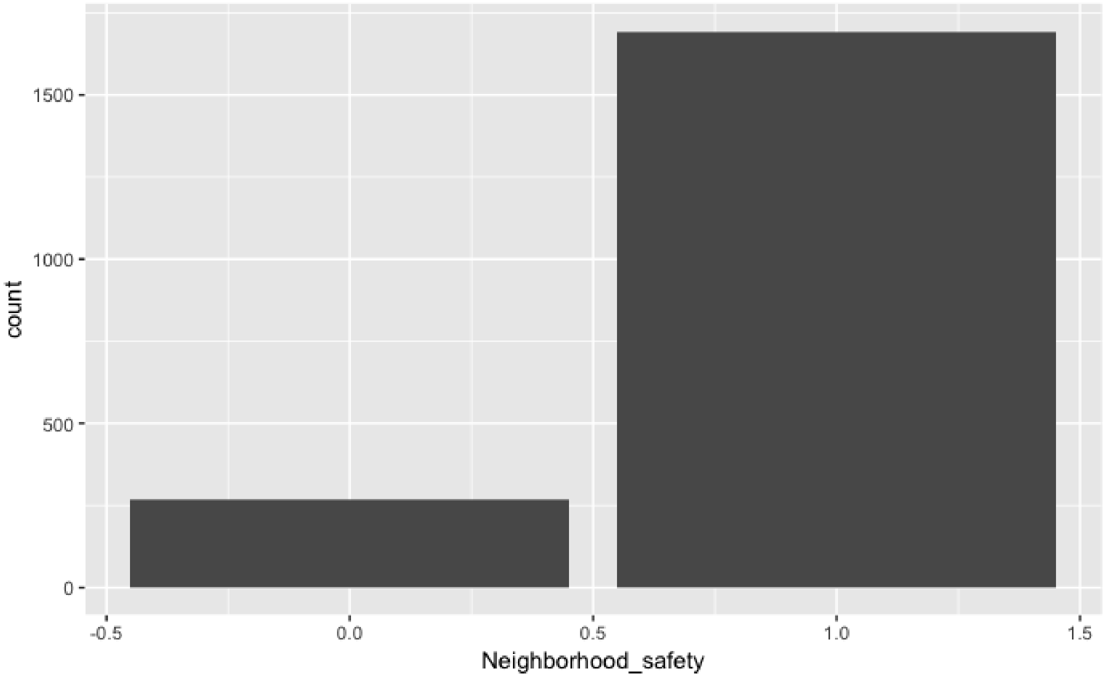
```
#neighborhood (=happy to live in this neighborhood) wave 2| indendent variable 4a unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Neighborhood_happiness ))
```



CONTROL GROUP 2 (SOCIAL NORMS)

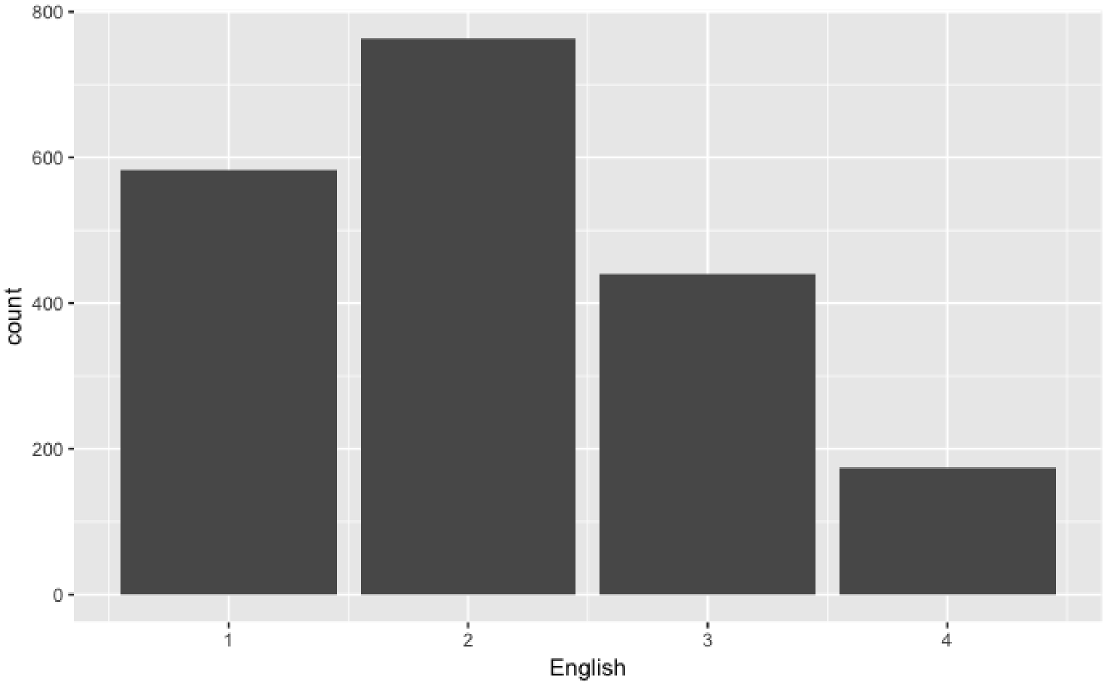
Hide

```
#neighborhood (=feel save in this neighborhood) wave 2 | indendent variable 4b unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Neighborhood_safety ))
```



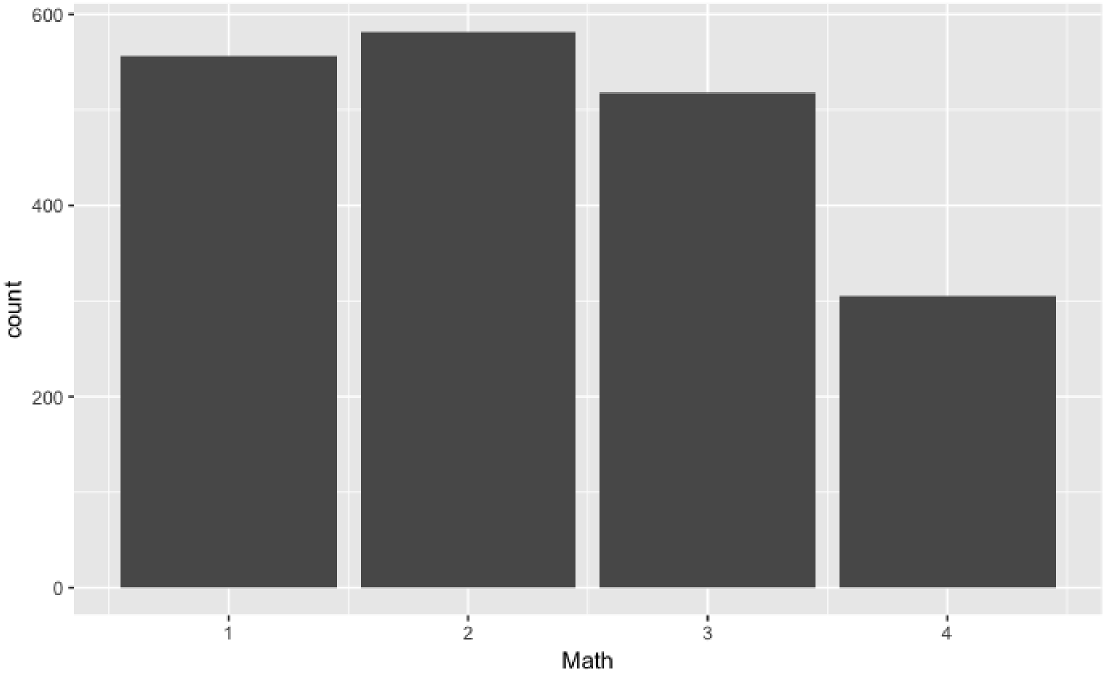
Hide

```
#grade English wave 2 | indendent variable 5a unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   English))
```



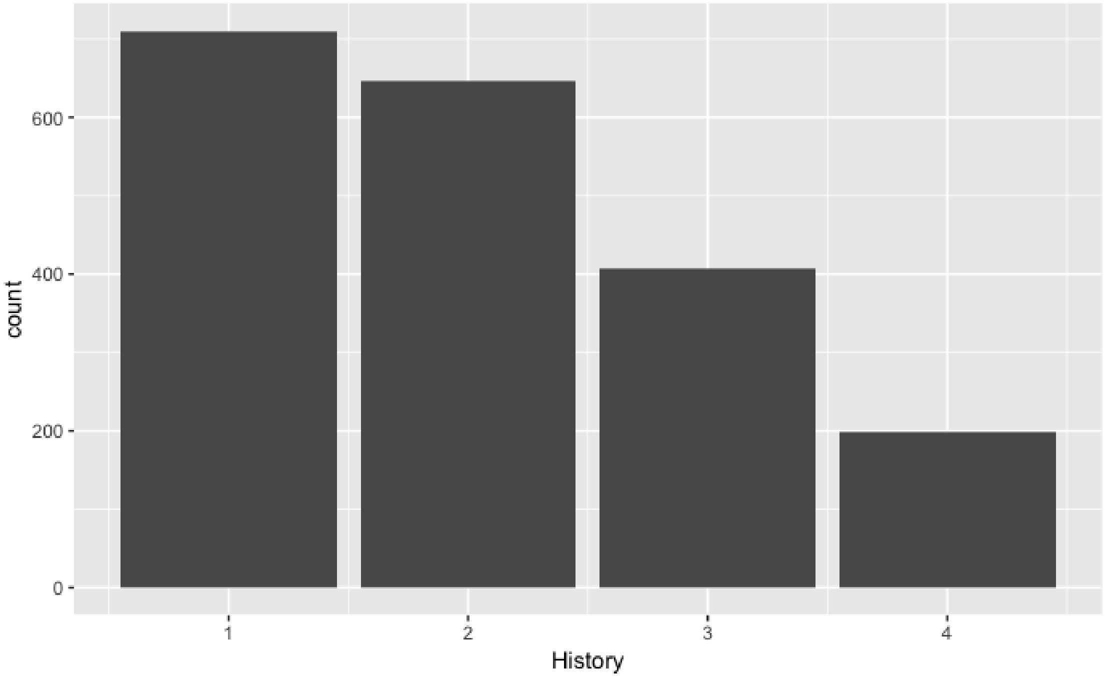
Hide

```
#grade Math wave 2 | indendent variable 5b unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Math))
```



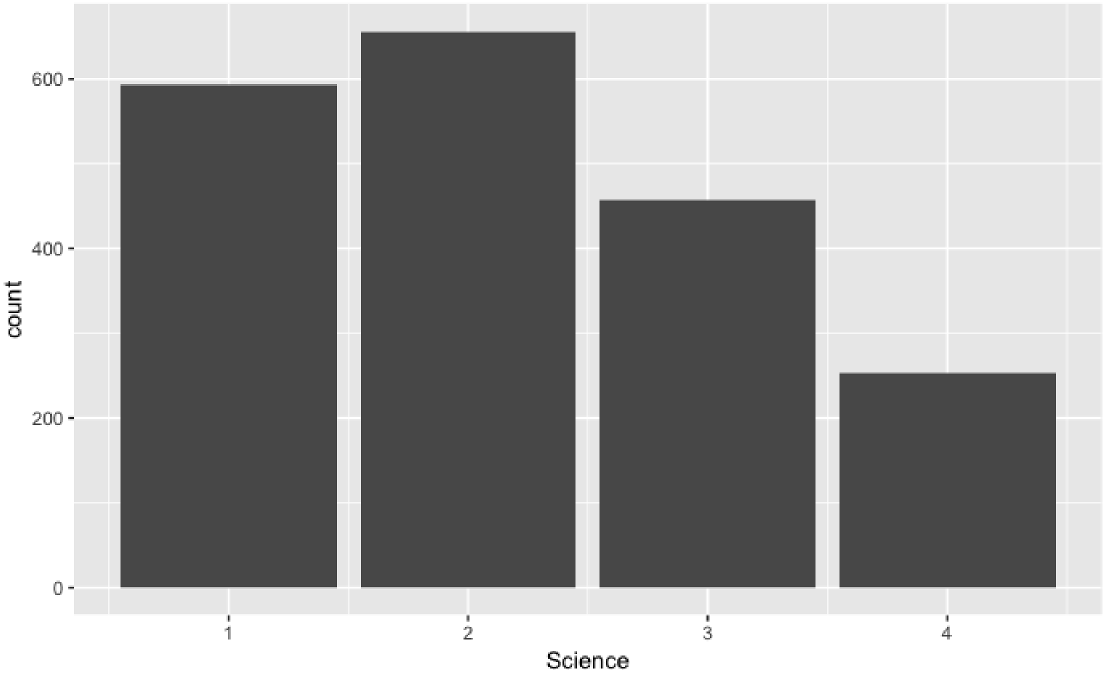
Hide

```
#grade History wave 2| indendent variable 5c unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   History))
```



Hide

```
#grade Science wave 2 | indendent variable 5d unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Science))
```

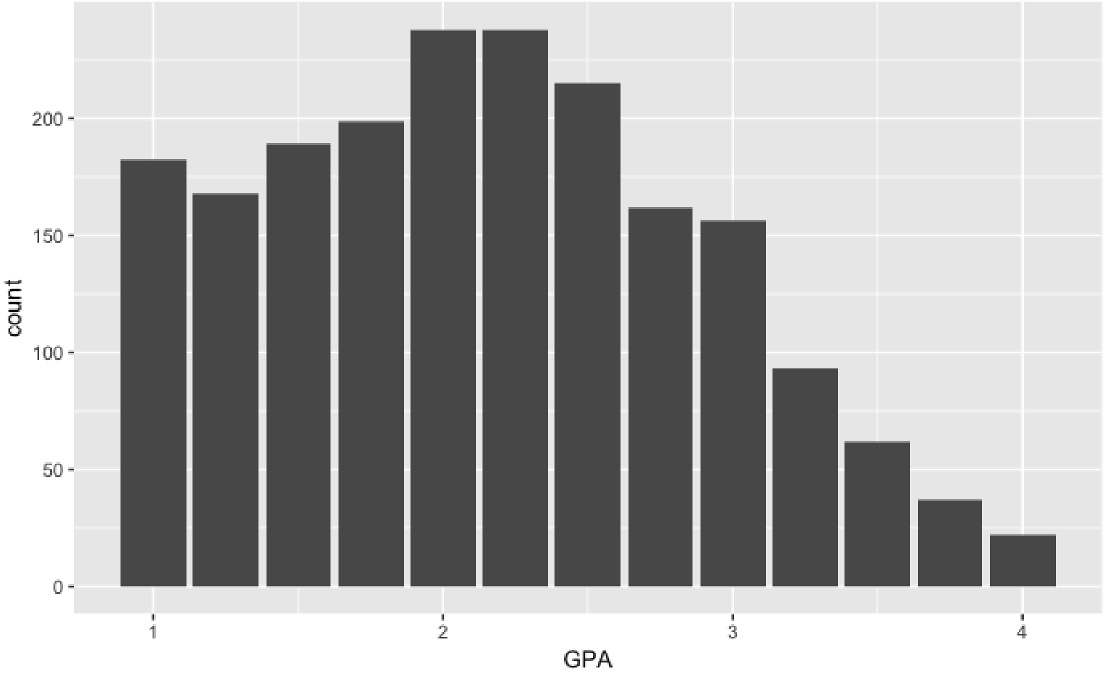


Hide

```
#Create GPA variable
df2<-mutate(df2, GPA= (History+English+Math+Science)/4)
```

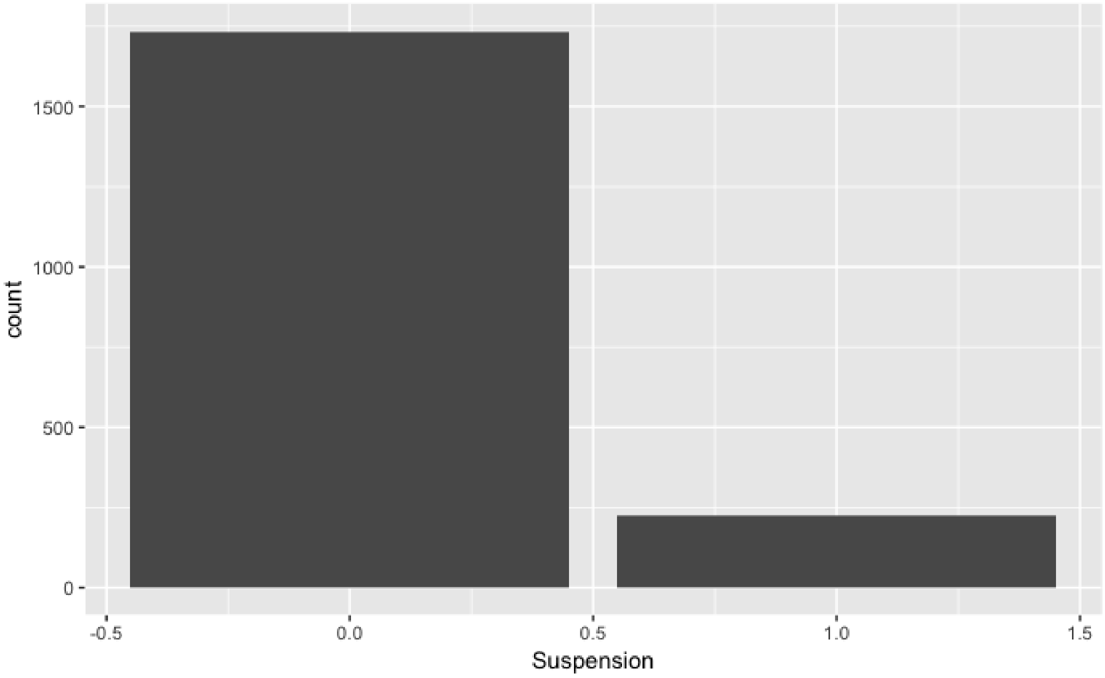
Hide

```
#GPA wave 2 | indendent variable 5 unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =GPA))
```



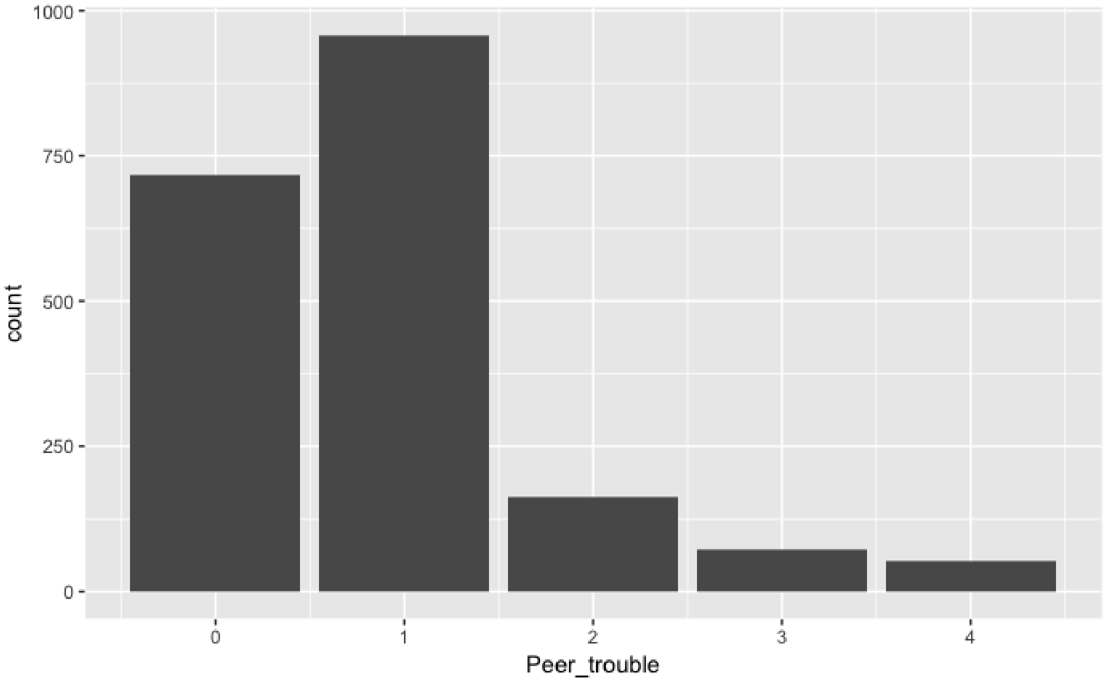
Hide

```
#school suspension wave 2| indendent variable 8 unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Suspension))
```



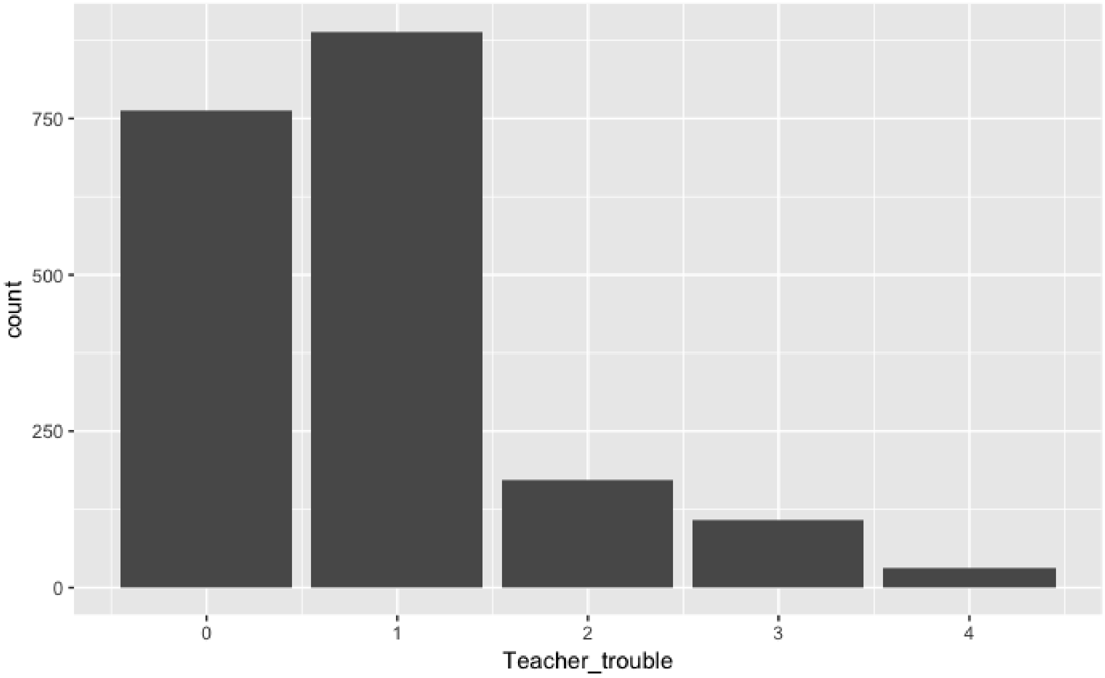
Hide

```
#trouble with peers wave 2 | indendent variable 9 unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Peer_trouble ))
```



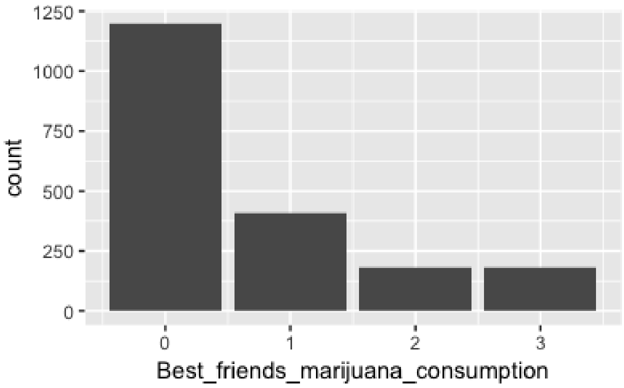
Hide

```
#trouble with teacher wave 2 | indendent variable 10 unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Teacher_trouble ))
```



Hide

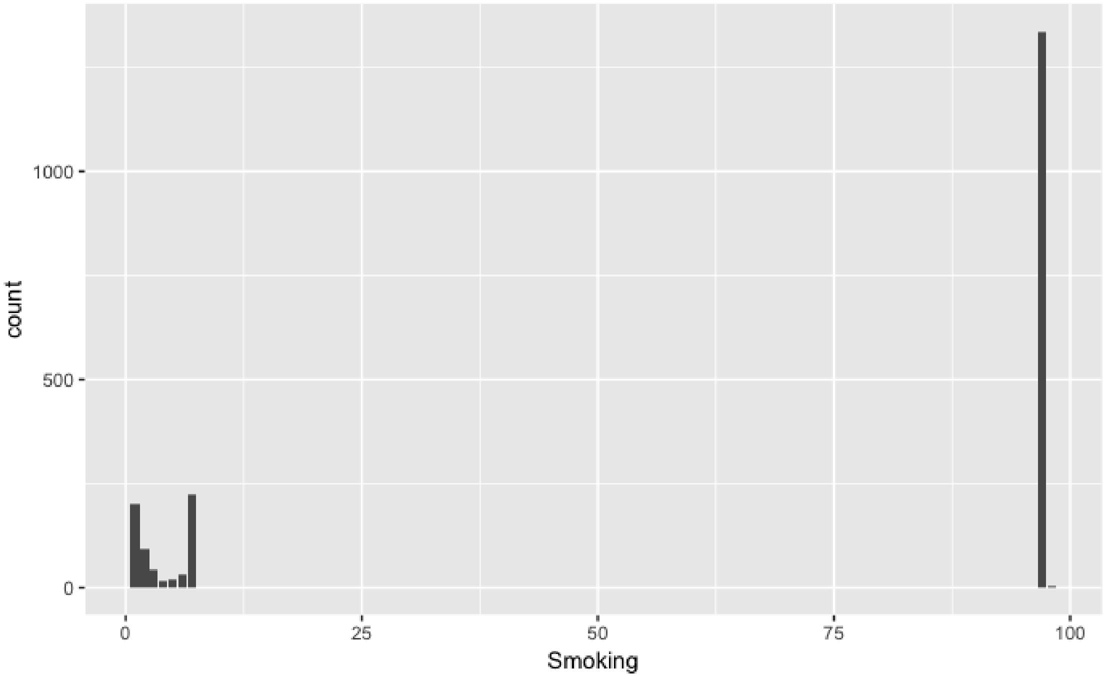
```
#Friends marijuana involvement wave 2 | indendent variable 11 unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Best_friends_marijuana_consumption ))
```



CONTROL GROUP 3 (OTHER SUBSTANCE INVOLVEMENT)

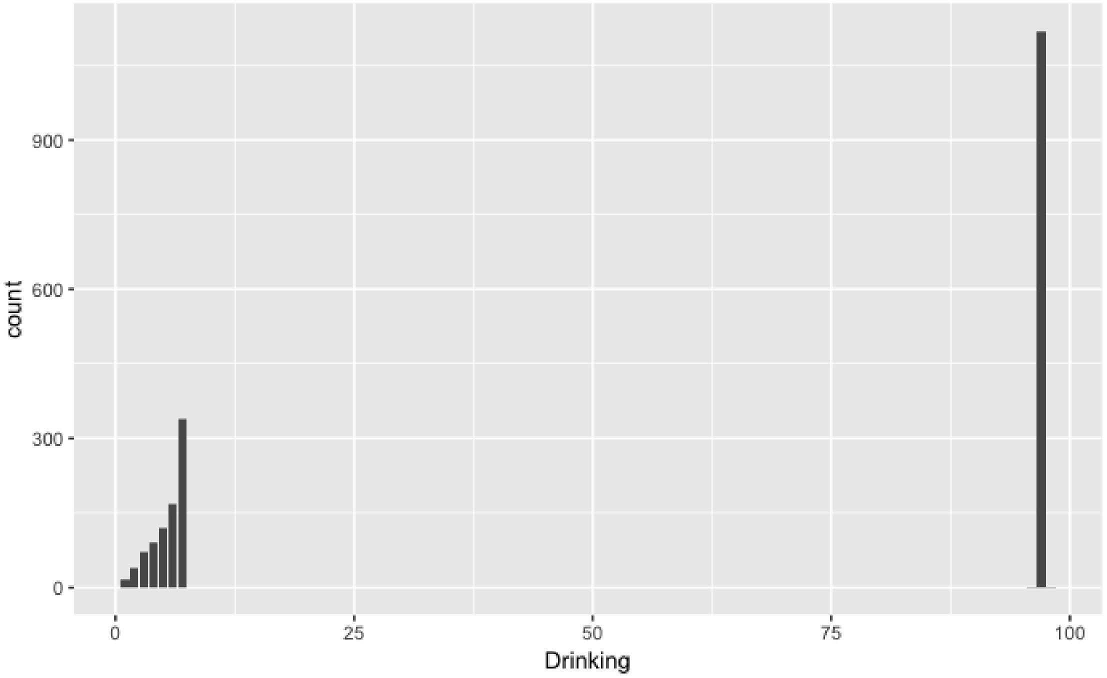
Hide

```
#smoking (=number of daily cigaretes) wave 2 | indendent variable 12a unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Smoking))
```



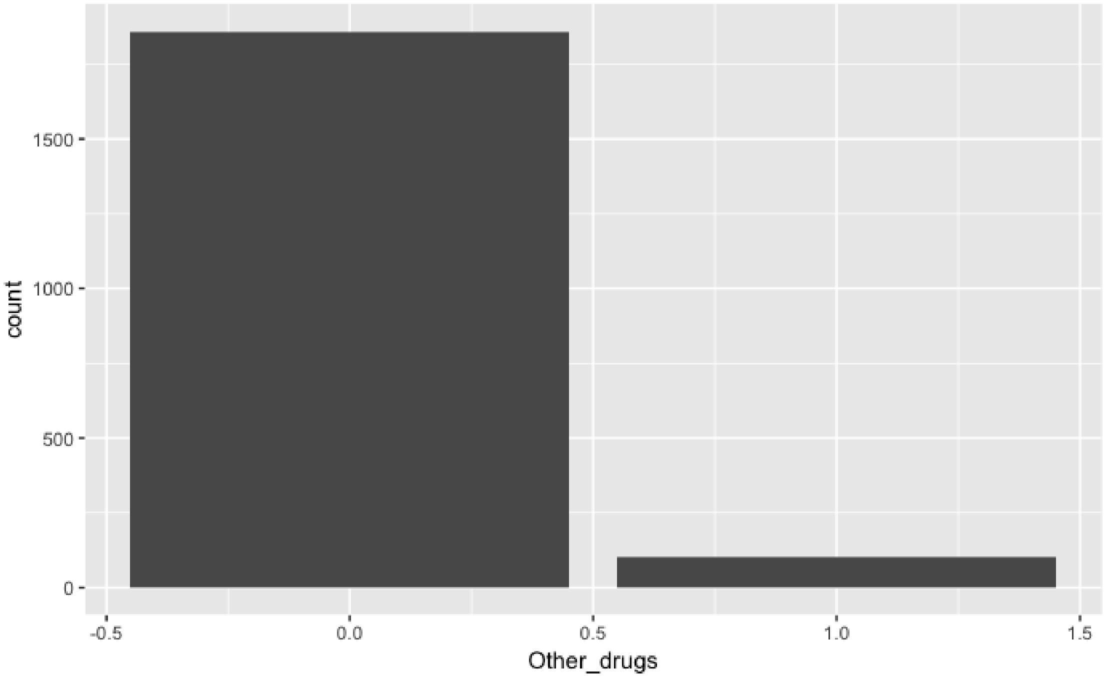
Hide

```
#drinking(=five plus drinks in the last 12 months) wave 2 | indendent variable 13 unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Drinking))
```



Hide

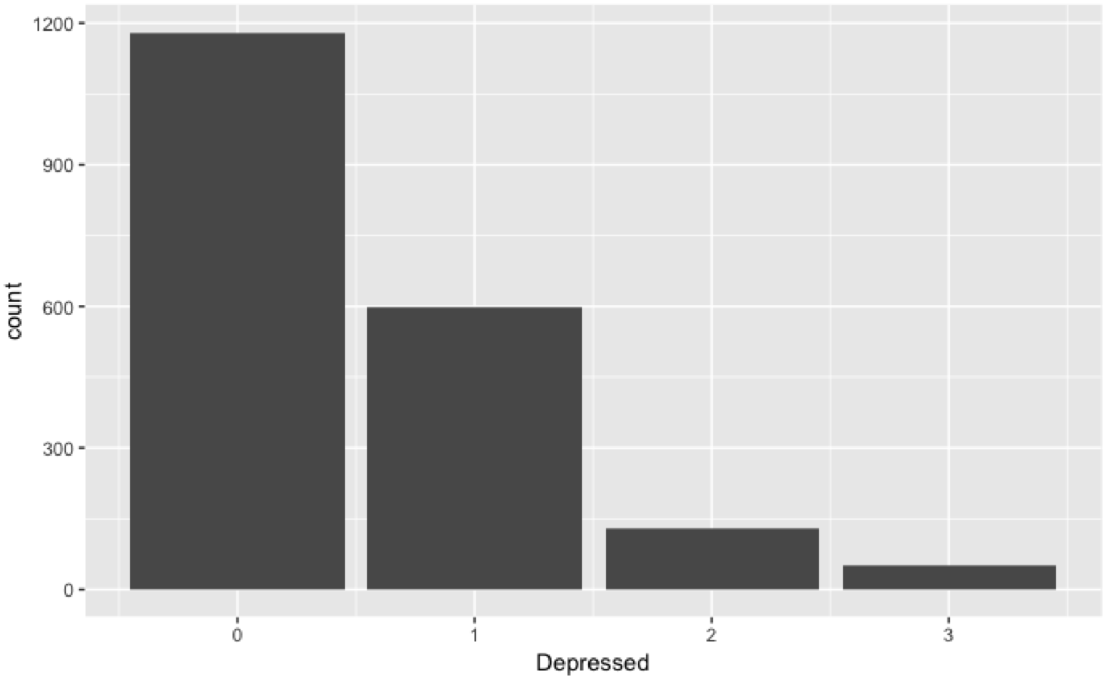
```
#Tried other illegal drugs(e.g LSD) wave 2 (=binary) | indendent variable 14 unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Other_drugs))
```



CONTROL GROUP 4 (PSYCHOLOGICAL CONDITION)

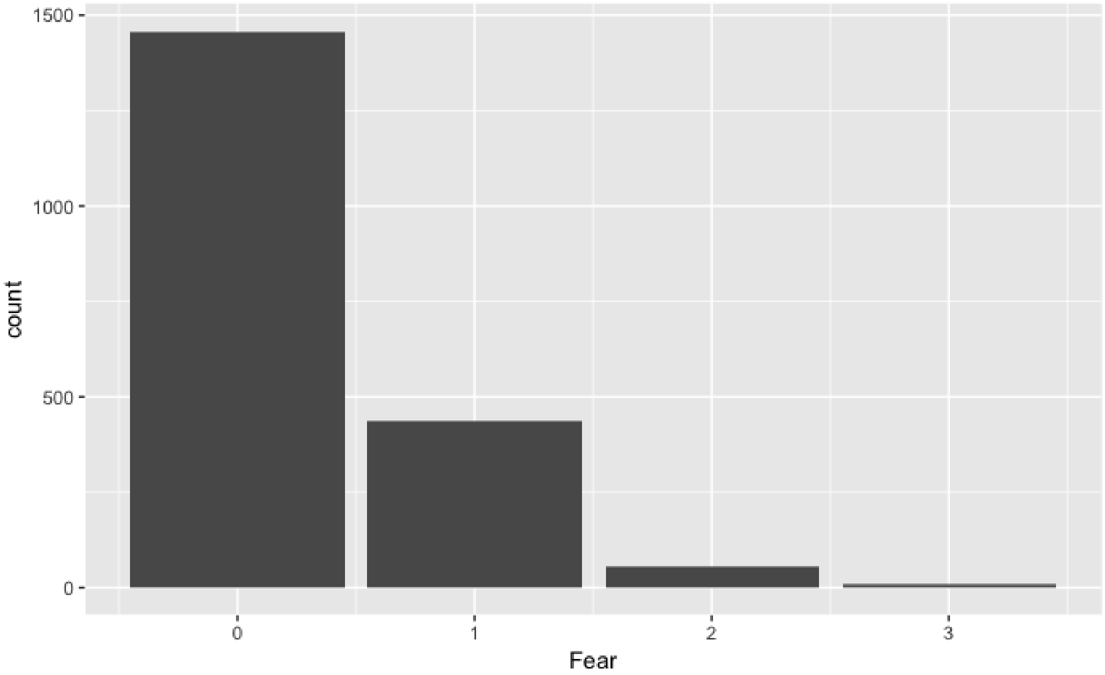
Hide

```
#feeling depressed wave 2 (=binary) | indendent variable 15 unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =    Depressed ))
```



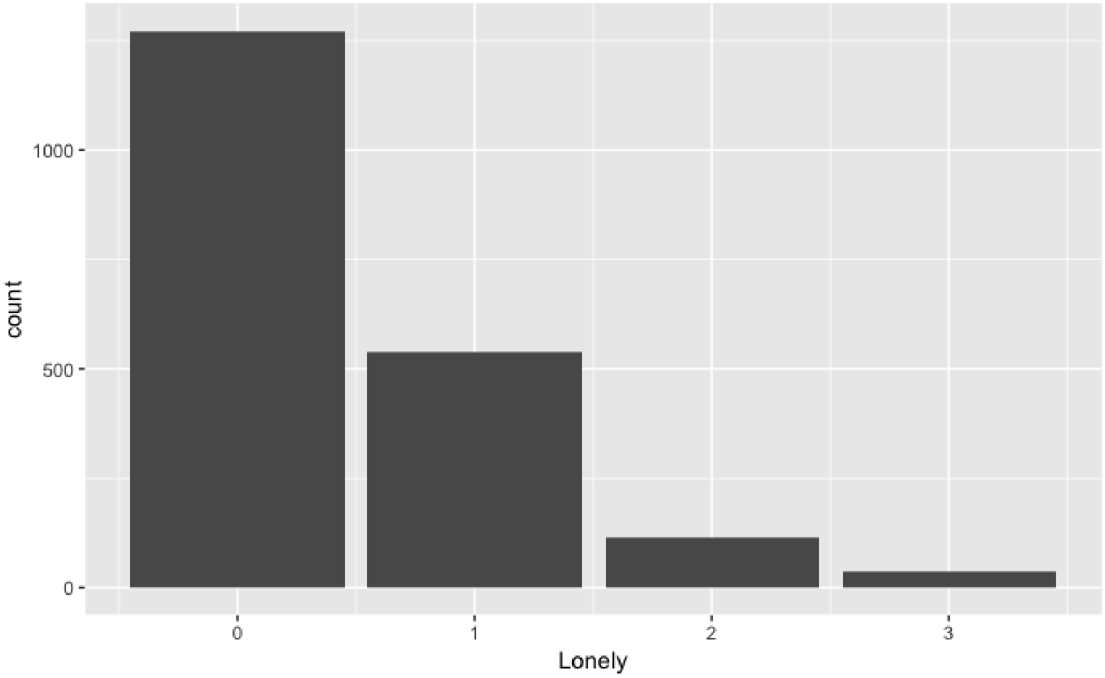
Hide

```
#feeling fear (4 stages) | indendent variable 17 unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Fear))
```



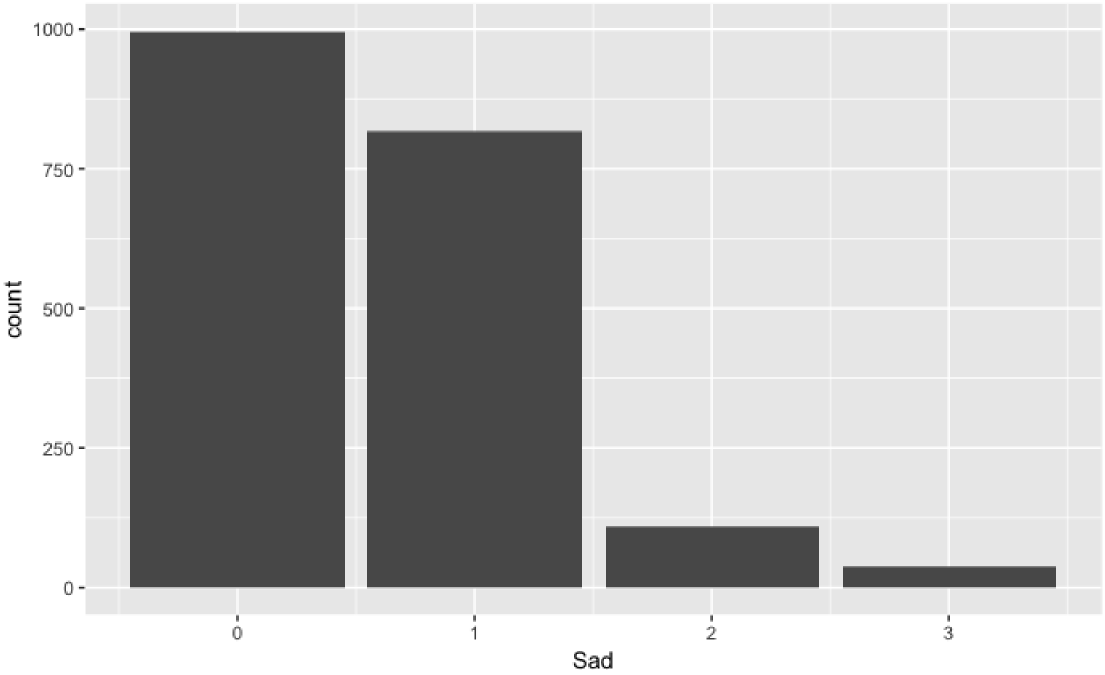
Hide

```
#feeling loney (4 stages) | indendent variable 18 unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Lonely ))
```



Hide

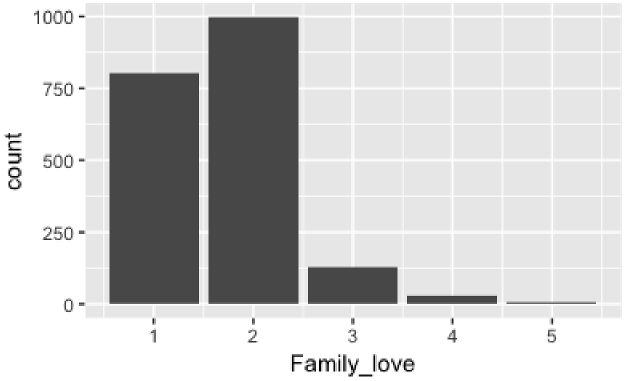
```
#feeling sad (4 stages) | indendent variable 19 unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Sad ))
```



MANIPULATE DATA

Hide

```
#feeling loved by family (5 stages) | indendent variable 20 unmodified
df2 %>%
  ggplot()  +
    geom_bar(mapping    = aes(x =   Family_love))
```



Hide

```
df2<- mutate(df2, Marijuana_dummy = ifelse( Marijuana<=9 , 1, 0))
df2<- mutate(df2, Smoker_dummy = ifelse( Smoking<=3 , 1, 0))
df2<- mutate(df2, Heavy_drinker_dummy = ifelse( Drinking<=4 , 1, 0))
```

SELECT RELEVANT VARIABLES

Hide

```
df2 %>% count (Marijuana_dummy)  %>%
        mutate(freq = n / sum(n))
```

Hide

```
df_model <-df2 %>%
  select( 'Marijuana_dummy',

          'Age',
          'Gender',
          'Neighborhood_happiness',
          'Neighborhood_safety',

          'GPA',
          'Suspension',
          'Peer_trouble',
          'Teacher_trouble',
          'Best_friends_marijuana_consumption',
          'Family_love',

          'Smoker_dummy',
          'Heavy_drinker_dummy',
          'Other_drugs',

          'Depressed',
          'Fear',
          'Lonely',
          'Sad'

          )
```

Hide

```
#variable standartisation
df_model<- mutate(df_model, Gender = Gender-1)
df_model<- mutate(df_model, Neighborhood_happiness = Neighborhood_happiness-1)
df_model<- mutate(df_model, GPA = GPA-1)
df_model<- mutate(df_model, Family_love = Family_love-1)
```

Hide

```
summary(df_model)
```

```
 Marijuana_dummy       Age            Gender       Neighborhood_happiness Neighborhood_safety      GPA       
Suspension    
 Min.   :0.0000   Min.   :12.00   Min.   :-4.000   Min.   :-3.0000        Min.   :0.0000      Min.   :-2.00
00   Min.   :0.0000  
 1st Qu.:0.0000   1st Qu.:14.00   1st Qu.:-4.000   1st Qu.:-1.0000        1st Qu.:1.0000      1st Qu.:-1.500
0   1st Qu.:0.0000  
 Median :0.0000   Median :16.00   Median :-3.000   Median : 0.0000        Median :1.0000      Median :-0.75
00   Median :0.0000  
 Mean   :0.1566   Mean   :15.57   Mean   :-3.345   Mean   :-0.6303        Mean   :0.8633      Mean   :-0.84
18   Mean   :0.1158  
 3rd Qu.:0.0000   3rd Qu.:17.00   3rd Qu.:-3.000   3rd Qu.: 0.0000        3rd Qu.:1.0000      3rd Qu.:-0.250
0   3rd Qu.:0.0000  
 Max.   :1.0000   Max.   :20.00   Max.   :-3.000   Max.   : 0.0000        Max.   :1.0000      Max.   : 1.00
00   Max.   :1.0000  
  Peer_trouble    Teacher_trouble  Best_friends_marijuana_consumption  Family_love      Smoker_dummy    Heav
y_drinker_dummy
 Min.   :0.0000   Min.   :0.0000   Min.   :0.0000                     Min.   :0.0000   Min.   :0.0000   Min
.   :0.0000   
 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000                     1st Qu.:0.0000   1st Qu.:0.0000   1st 
Qu.:0.0000   
 Median :1.0000   Median :1.0000   Median :0.0000                     Median :1.0000   Median :0.0000   Med
ian :0.0000   
 Mean   :0.8715   Mean   :0.8542   Mean   :0.6645                     Mean   :0.6945   Mean   :0.1703   Mea
n   :0.1091   
 3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000                     3rd Qu.:1.0000   3rd Qu.:0.0000   3rd 
Qu.:0.0000   
 Max.   :4.0000   Max.   :4.0000   Max.   :3.0000                     Max.   :4.0000   Max.   :1.0000   Max
.   :1.0000   
  Other_drugs        Depressed          Fear            Lonely           Sad        
 Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000  
 1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000  
 Median :0.00000   Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000  
 Mean   :0.05201   Mean   :0.5181   Mean   :0.2953   Mean   :0.4462   Mean   :0.5859  
 3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000  
 Max.   :1.00000   Max.   :3.0000   Max.   :3.0000   Max.   :3.0000   Max.   :3.0000  
```

Hide

```
#std
sd(df_model$Marijuana_dummy)
```

```
[1] 0.3634713
```

Hide

```
#std  Age
sd(df_model$Age)
```

```
[1] 1.455193
```

Hide

```
#std  Gender
sd(df_model$Gender)
```

```
[1] 0.4755651
```

Hide

```
#std  Neighborhood_happiness
sd(df_model$Neighborhood_happiness)
```

```
[1] 0.8221924
```

Hide

```
#std  Neighborhood_safety
sd(df_model$Neighborhood_safety)
```

```
[1] 0.3435809
```

Hide

```
#std  GPA
sd(df_model$GPA)
```

```
[1] 0.7453864
```

Hide

```
#std  Suspension
sd(df_model$Suspension)
```

```
[1] 0.3200152
```

Hide

```
#std  peer_trouble
sd(df_model$Peer_trouble)
```

```
[1] 0.9050923
```

Hide

```
#std  Teacher_trouble
sd(df_model$Teacher_trouble)
```

```
[1] 0.9018939
```

Hide

```
#std  Best_friends_marijuana_consumption
sd(df_model$Best_friends_marijuana_consumption)
```

```
[1] 0.9770877
```

Hide

```
#std  Family_love
sd(df_model$Family_love)
```

```
[1] 0.6851306
```

Hide

```
#std Smoker_dummy
sd(df_model$Smoker_dummy)
```

```
[1] 0.3760107
```

Hide

```
#std  Heavy_drinker_dummy
sd(df_model$Heavy_drinker_dummy)
```

```
[1] 0.3118793
```

Hide

```
#std  Other_drugs
sd(df_model$Other_drugs )
```

```
[1] 0.2221125
```

Hide

```
#std  Depressed
sd(df_model$Depressed)
```

```
[1] 0.7366524
```

Hide

```
#std  Fear
sd(df_model$Fear)
```

```
[1] 0.5421263
```

Hide

```
#std  Lonely
sd(df_model$Lonely)
```

```
[1] 0.688304
```

CORRELATION TABLE

Hide

```
#std  Sad
sd(df_model$Sad)
```

```
[1] 0.6842885
```

Hide

```
# correlation table with p-value
rcorr(as.matrix(df_model))
```

```
                                   Marijuana_dummy   Age Gender Neighborhood_happiness Neighborhood_safety
GPA Suspension
Marijuana_dummy                               1.00  0.06  -0.01                  -0.04               -0.02
0.15        0.17
Age                                           0.06  1.00  -0.04                  -0.02                0.02
0.01       -0.02
Gender                                       -0.01 -0.04   1.00                  -0.02               -0.05
0.06        0.01
Neighborhood_happiness                       -0.04 -0.02  -0.02                   1.00                0.36
-0.14      -0.12
Neighborhood_safety                          -0.02  0.02  -0.05                   0.36                1.00
-0.09      -0.09
GPA                                           0.15  0.01   0.06                  -0.14               -0.09
1.00        0.27
Suspension                                    0.17 -0.02   0.01                  -0.12               -0.09
0.27        1.00
Peer_trouble                                  0.08 -0.08   0.02                  -0.10               -0.05
0.14        0.13
Teacher_trouble                               0.20 -0.08   0.04                  -0.14               -0.04
0.27        0.24
Best_friends_marijuana_consumption            0.59  0.07   0.01                  -0.05               -0.06
0.20        0.16
Family_love                                   0.07  0.06  -0.01                  -0.07               -0.02
0.08        0.02
Smoker_dummy                                  0.39  0.09  -0.01                  -0.05                0.00
0.19        0.20
Heavy_drinker_dummy                           0.34  0.10  -0.02                  -0.05                0.02
0.09        0.11
Other_drugs                                   0.31  0.03  -0.04                  -0.03                0.02
0.05        0.09
Depressed                                     0.07  0.07   0.03                  -0.16               -0.12
0.13        0.07
Fear                                          0.04  0.01  -0.01                  -0.04               -0.09
```

```
0.05       0.07
Lonely                                        0.07  0.09   0.04                  -0.14               -0.08
0.07       0.09
Sad                                           0.05  0.05   0.02                  -0.14               -0.11
0.08       0.03
                                   Peer_trouble Teacher_trouble Best_friends_marijuana_consumption Family_l
ove Smoker_dummy
Marijuana_dummy                            0.08            0.20                               0.59
0.07         0.39
Age                                       -0.08           -0.08                               0.07
0.06         0.09
Gender                                     0.02            0.04                               0.01        -
0.01        -0.01
Neighborhood_happiness                    -0.10           -0.14                              -0.05        -
0.07        -0.05
Neighborhood_safety                       -0.05           -0.04                              -0.06        -
0.02         0.00
GPA                                        0.14            0.27                               0.20
0.08         0.19
Suspension                                 0.13            0.24                               0.16
0.02         0.20
Peer_trouble                               1.00            0.34                               0.09
0.10         0.08
Teacher_trouble                            0.34            1.00                               0.17
0.10         0.18
Best_friends_marijuana_consumption         0.09            0.17                               1.00        0
.07         0.36
Family_love                                0.10            0.10                               0.07
1.00         0.09
Smoker_dummy                               0.08            0.18                               0.36
0.09         1.00
Heavy_drinker_dummy                        0.08            0.15                               0.33
0.02         0.31
Other_drugs                                0.05            0.09                               0.32
0.08         0.27
Depressed                                  0.20            0.15                               0.11
0.24         0.09
Fear                                       0.16            0.11                               0.10
0.12         0.03
Lonely                                     0.18            0.11                               0.08
0.26         0.05
Sad                                        0.20            0.11                               0.11
0.23         0.03
                                   Heavy_drinker_dummy Other_drugs Depressed  Fear Lonely   Sad
Marijuana_dummy                                   0.34        0.31      0.07  0.04   0.07  0.05
Age                                               0.10        0.03      0.07  0.01   0.09  0.05
Gender                                           -0.02       -0.04      0.03 -0.01   0.04  0.02
Neighborhood_happiness                           -0.05       -0.03     -0.16 -0.04  -0.14 -0.14
Neighborhood_safety                               0.02        0.02     -0.12 -0.09  -0.08 -0.11
GPA                                               0.09        0.05      0.13  0.05   0.07  0.08
Suspension                                        0.11        0.09      0.07  0.07   0.09  0.03
Peer_trouble                                      0.08        0.05      0.20  0.16   0.18  0.20
Teacher_trouble                                   0.15        0.09      0.15  0.11   0.11  0.11
Best_friends_marijuana_consumption                0.33        0.32      0.11  0.10   0.08  0.11
Family_love                                       0.02        0.08      0.24  0.12   0.26  0.23
Smoker_dummy                                      0.31        0.27      0.09  0.03   0.05  0.03
Heavy_drinker_dummy                               1.00        0.21      0.09  0.07   0.04  0.05
Other_drugs                                       0.21        1.00      0.09  0.05   0.06  0.10
Depressed                                         0.09        0.09      1.00  0.32   0.49  0.63
Fear                                              0.07        0.05      0.32  1.00   0.31  0.33
Lonely                                            0.04        0.06      0.49  0.31   1.00  0.51
Sad                                               0.05        0.10      0.63  0.33   0.51  1.00

n= 1961


P
                                   Marijuana_dummy Age    Gender Neighborhood_happiness Neighborhood_safety
GPA    Suspension
Marijuana_dummy                                    0.0106 0.6001 0.0890                 0.4647
0.0000 0.0000
Age                                0.0106                 0.1121 0.4767                 0.3445
0.7085 0.4228
```

```
0.7085 0.4228
Gender                             0.6001          0.1121        0.3644                0.0223
0.0076 0.5170
Neighborhood_happiness             0.0890          0.4767 0.3644                       0.0000
0.0000 0.0000
Neighborhood_safety                0.4647          0.3445 0.0223 0.0000
0.0000 0.0000
GPA                                0.0000          0.7085 0.0076 0.0000                0.0000
0.0000
Suspension                         0.0000          0.4228 0.5170 0.0000                0.0000
0.0000
Peer_trouble                       0.0004          0.0005 0.2940 0.0000                0.0224
0.0000 0.0000
Teacher_trouble                    0.0000          0.0003 0.1110 0.0000                0.0674
0.0000 0.0000
Best_friends_marijuana_consumption 0.0000          0.0039 0.5651 0.0172                0.0088
0.0000 0.0000
Family_love                        0.0029          0.0058 0.6880 0.0021                0.4507
0.0002 0.3904
Smoker_dummy                       0.0000          0.0000 0.6410 0.0312                0.9101
0.0000 0.0000
Heavy_drinker_dummy                0.0000          0.0000 0.3514 0.0336                0.2689
0.0000 0.0000
Other_drugs                        0.0000          0.1891 0.0959 0.1475                0.3844
0.0361 0.0000
Depressed                          0.0018          0.0036 0.1256 0.0000                0.0000
0.0000 0.0036
Fear                               0.0999          0.6143 0.5334 0.0477                0.0000
0.0178 0.0011
Lonely                             0.0016          0.0000 0.0835 0.0000                0.0005
0.0017 0.0002
Sad                                0.0176          0.0190 0.4180 0.0000                0.0000
0.0004 0.2161
                                   Peer_trouble Teacher_trouble Best_friends_marijuana_consumption Family_l
ove Smoker_dummy
Marijuana_dummy                    0.0004       0.0000          0.0000                             0.0029
0.0000
Age                                0.0005       0.0003          0.0039                             0.0058
0.0000
Gender                             0.2940       0.1110          0.5651                             0.6880
0.6410
Neighborhood_happiness             0.0000       0.0000          0.0172                             0.0021
0.0312
Neighborhood_safety                0.0224       0.0674          0.0088                             0.4507
0.9101
GPA                                0.0000       0.0000          0.0000                             0.0002
0.0000
Suspension                         0.0000       0.0000          0.0000                             0.3904
0.0000
Peer_trouble                                    0.0000          0.0000                             0.0000
0.0003
Teacher_trouble                    0.0000                       0.0000                             0.0000
0.0000
Best_friends_marijuana_consumption 0.0000       0.0000                                             0.0012
0.0000
Family_love                        0.0000       0.0000          0.0012
0.0002
Smoker_dummy                       0.0003       0.0000          0.0000                             0.0002

Heavy_drinker_dummy                0.0007       0.0000          0.0000                             0.3222
0.0000
Other_drugs                        0.0419       0.0000          0.0000                             0.0006
0.0000
Depressed                          0.0000       0.0000          0.0000                             0.0000
0.0000
Fear                               0.0000       0.0000          0.0000                             0.0000
0.2072
Lonely                             0.0000       0.0000          0.0002                             0.0000
0.0185
Sad                                0.0000       0.0000          0.0000                             0.0000
0.1525
                                   Heavy_drinker_dummy Other_drugs Depressed Fear  Lonely Sad
Marijuana_dummy                    0.0000              0.0000      0.0018    0.0999 0.0016 0.0176
```

```
Age                                0.0000              0.1891      0.0036    0.6143 0.0000 0.0190
Gender                             0.3514              0.0959      0.1256    0.5334 0.0835 0.4180
Neighborhood_happiness             0.0336              0.1475      0.0000    0.0477 0.0000 0.0000
Neighborhood_safety                0.2689              0.3844      0.0000    0.0000 0.0005 0.0000
GPA                                0.0000              0.0361      0.0000    0.0178 0.0017 0.0004
Suspension                         0.0000              0.0000      0.0036    0.0011 0.0002 0.2161
Peer_trouble                       0.0007              0.0419      0.0000    0.0000 0.0000 0.0000
Teacher_trouble                    0.0000              0.0000      0.0000    0.0000 0.0000 0.0000
Best_friends_marijuana_consumption 0.0000              0.0000      0.0000    0.0000 0.0002 0.0000
Family_love                        0.3222              0.0006      0.0000    0.0000 0.0000 0.0000
Smoker_dummy                       0.0000              0.0000      0.0000    0.2072 0.0185 0.1525
Heavy_drinker_dummy                                    0.0000      0.0000    0.0009 0.0514 0.0379
Other_drugs                        0.0000                          0.0000    0.0412 0.0148 0.0000
Depressed                          0.0000              0.0000                0.0000 0.0000 0.0000
Fear                               0.0009              0.0412      0.0000           0.0000 0.0000
Lonely                             0.0514              0.0148      0.0000    0.0000        0.0000
Sad                                0.0379              0.0000      0.0000    0.0000 0.0000
```

Hide

```
install.packages("corrplot")
```

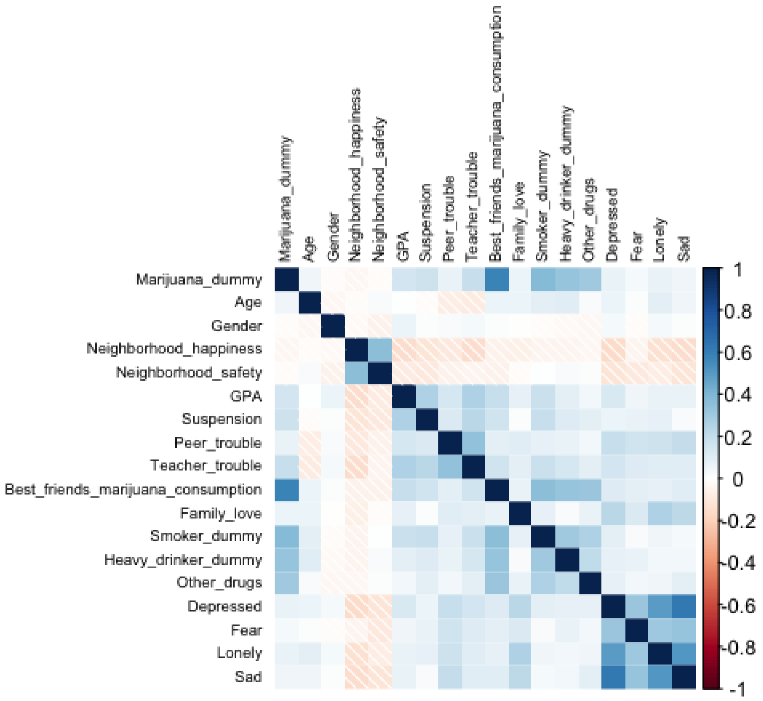
```
Error in install.packages : Updating loaded packages
```

Hide

```
library(corrplot)
```

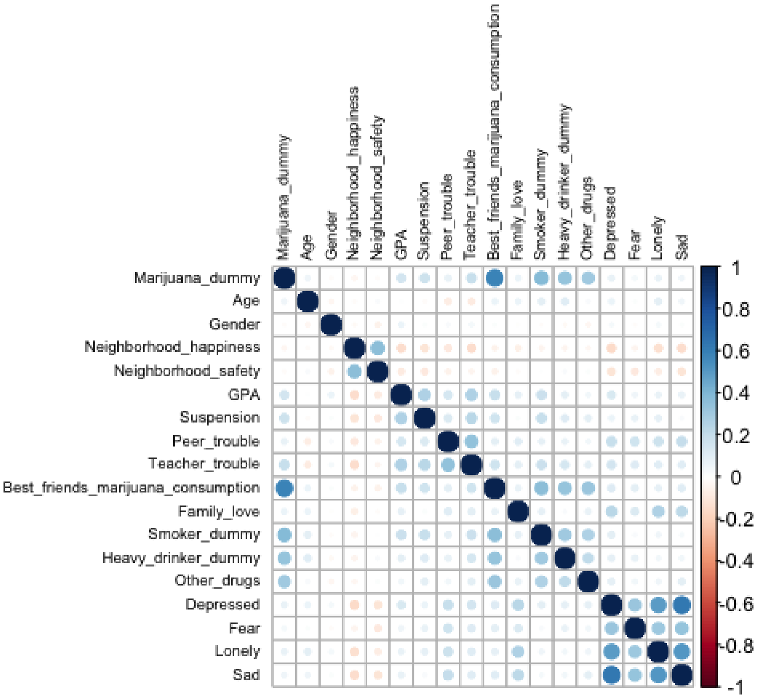
Hide

```
corrplot(cor(df_model), method="shade", tl.cex =0.6,tl.col="black")
```



Hide

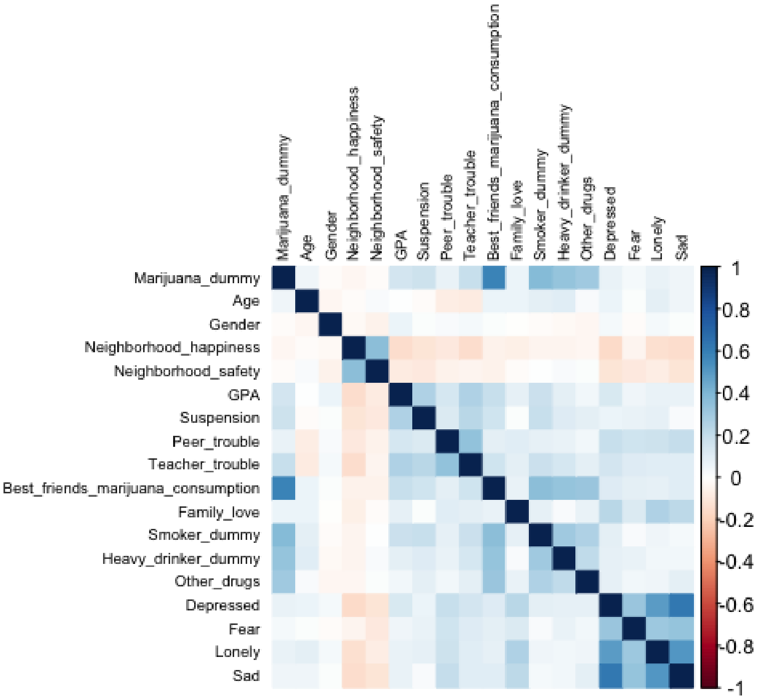
```
corrplot(cor(df_model), method="circle",tl.cex =0.6,tl.col="black")
```



Basic EDA

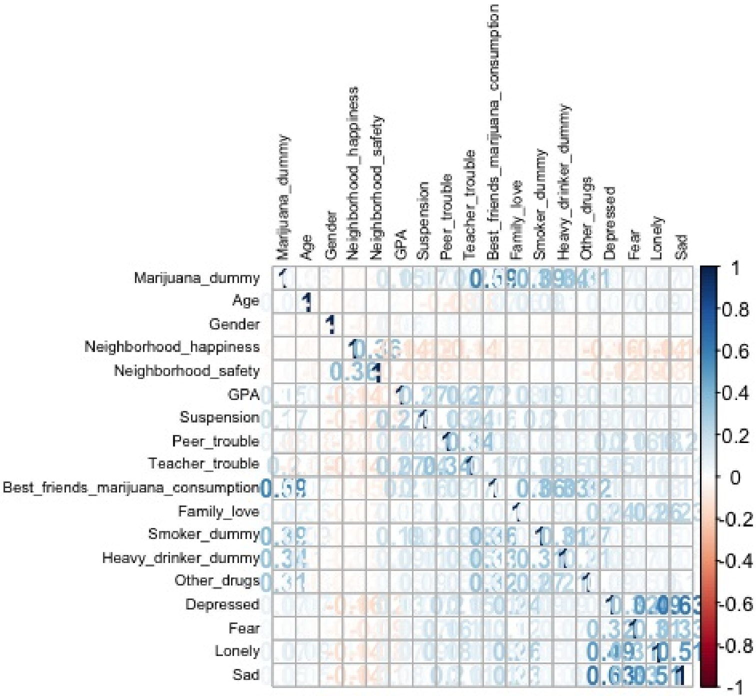
Hide

```
corrplot(cor(df_model), method="color",tl.cex =0.6,tl.col="black")
```



Hide

```
corrplot(cor(df_model), method="number",tl.cex =0.6,tl.col="black")
```



Hide

```
i<-df_model %>%
  group_by(Age) %>%
  summarise(smoke_age = mean(Smoker_dummy) )
i
```

Hide

```
j<-df_model %>%
  group_by(Age) %>%
  summarise(drink_age = mean(Heavy_drinker_dummy))
```

Hide

```
k<-df_model %>%
  group_by(Age) %>%
  summarise(Marijuana_age = mean(Marijuana_dummy))
```

Hide

```
l<-df_model %>%
  group_by(Age) %>%
  summarise(Drugs_age = mean(Other_drugs))
```

Hide

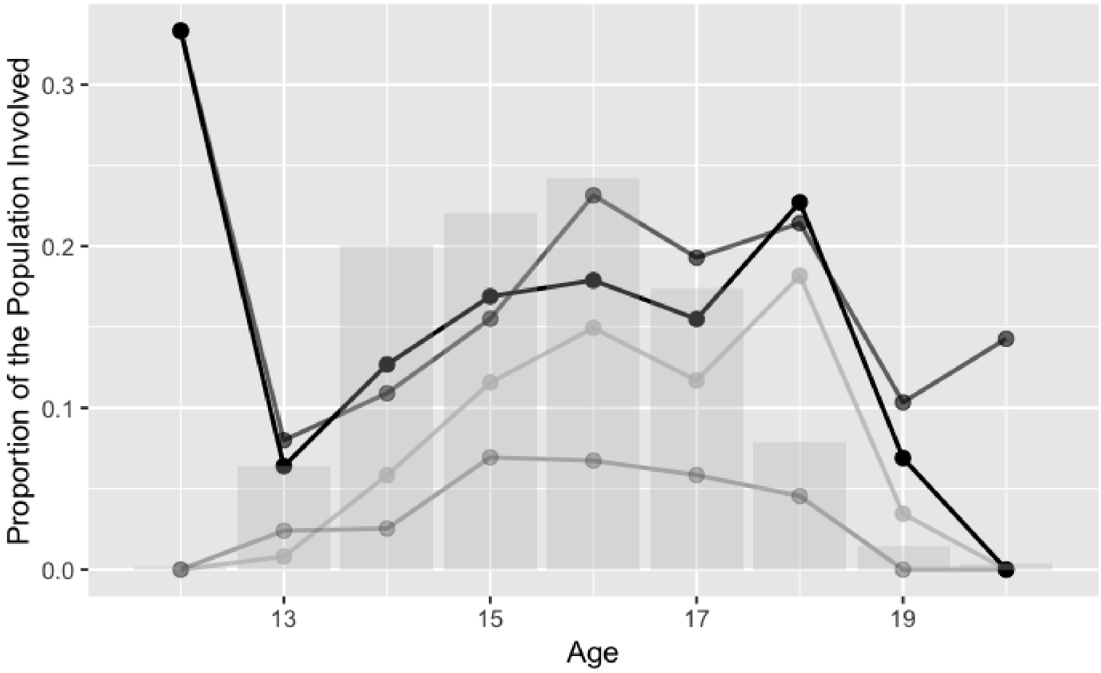
```
m<-df_model %>%
  group_by(Age) %>%
  count()
m
```

Logistic Regression, step-wise extension

model 1

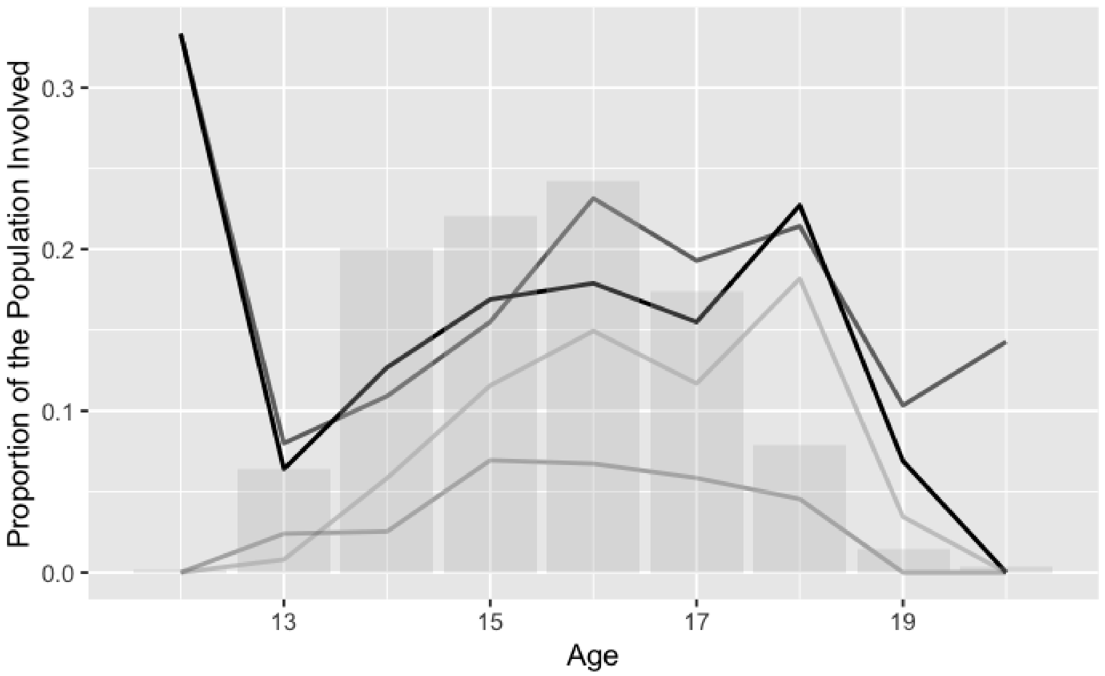
Hide

```
#Illustration 1
ggplot() +
  geom_line(data=i, aes( x  =   Age, y=(smoke_age)),size=1, alpha=0.6, color='black') +
  geom_line(data=x,aes(x    =   Age, y=(drink_age)),size=1,alpha=0.8, color='grey') +
  geom_point(data=x,aes(x   =   Age, y=(drink_age)),size=3,alpha=0.8, color='grey') +
  geom_point(data=i, aes( x =   Age, y=(smoke_age)),size=3, alpha=0.6, color='black') +
  geom_line(data=k, aes( x  =   Age, y=(Marijuana_age)),size=1, alpha=1, color='black') +
  geom_line(data=l,aes(x    =   Age, y=(Drugs_age)),size=1,alpha=0.3, color='black') +
  geom_point(data=l,aes(x   =   Age, y=(Drugs_age)),size=3,alpha=0.3, color='black') +
  geom_point(data=k, aes( x =   Age, y=(Marijuana_age)),size=3, alpha=1, color='black') +
  geom_bar(data=m,aes(x    =   Age, y=n/1961),stat = "identity", alpha=0.3, fill='grey') +
  ylab('Proportion of the Population Involved') +
  theme_grey(base_size = 14)
```



Hide

```
#Illustration 1
ggplot() +
  geom_line(data=i, aes( x  =   Age, y=(smoke_age)),size=1, alpha=0.6, color='black') +
  geom_line(data=x,aes(x    =   Age, y=(drink_age)),size=1,alpha=0.8, color='grey') +
  geom_line(data=k, aes( x  =   Age, y=(Marijuana_age)),size=1, alpha=1, color='black') +
  geom_line(data=l,aes(x    =   Age, y=(Drugs_age)),size=1,alpha=0.3, color='black') +
  geom_bar(data=m,aes(x    =   Age, y=n/1961),stat = "identity", alpha=0.3, fill='grey') +
  ylab('Proportion of the Population Involved') +
  theme_grey(base_size = 14)
```



Hide

```
#take only the focus group of variables as predictors
Other_involvement_logit <- glm(Marijuana_dummy  ~ Heavy_drinker_dummy+ Smoker_dummy+ Other_drugs, data = df_
model, family = "binomial")
summary(Other_involvement_logit)
```

Call: glm(formula = Marijuana_dummy ~ Heavy_drinker_dummy + Smoker_dummy + Other_drugs, family = “binomial”, data = df_model)

Deviance Residuals: Min 1Q Median 3Q Max
-2.1852 -0.3908 -0.3908 -0.3908 2.2849

Coefficients: Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.53391 0.09419 -26.903 < 2e-16 ***Heavy_drinker_dummy 1.48315 0.17725 8.368 < 2e-16*** Smoker_dummy 1.64823 0.15313 10.764 < 2e-16 ***Other_drugs 1.69373 0.24715 6.853 7.22e-12*** — Signif. codes: 0 ‘***’ 0.001 ‘**’ *0.01* ‘*’ 0.05 ‘.’ 0.1 ‘’ 1

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 1701.8  on 1960  degrees of freedom
```

Residual deviance: 1338.0 on 1957 degrees of freedom AIC: 1346

Number of Fisher Scoring iterations: 5

Hide

```
install.packages("stargazer", repos = "http://cran.us.r-project.org")
```

```
trying URL 'http://cran.us.r-project.org/src/contrib/stargazer_5.2.2.tar.gz'
Content type 'application/x-gzip' length 315967 bytes (308 KB)
==================================================
downloaded 308 KB

* installing *source* package ‘stargazer’ ...
** package ‘stargazer’ successfully unpacked and MD5 sums checked
** R
** inst
** preparing package for lazy loading
** help
*** installing help indices
** building package indices
** installing vignettes
** testing if installed package can be loaded
* DONE (stargazer)

The downloaded source packages are in
    ‘/private/var/folders/g4/9x186yqx14b1_jz1vkfp913r0000gn/T/RtmpS4042i/downloaded_packages’
Updating HTML index of packages in '.Library'
Making 'packages.html' ... done
```

Hide

```
library(stargazer)
```

```


Please cite as:

 Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
 R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

Hide

```
stargazer(Other_involvement_logit,type = "text")
```

```

===============================================
                        Dependent variable:
                    ---------------------------
                          Marijuana_dummy
-----------------------------------------------
Heavy_drinker_dummy          1.483***
                              (0.177)

Smoker_dummy                 1.648***
                              (0.153)

Other_drugs                  1.694***
                              (0.247)

Constant                     -2.534***
                              (0.094)

-----------------------------------------------
Observations                   1,961
Log Likelihood                -669.025
Akaike Inf. Crit.             1,346.050
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

Hide

```
#psudo R_SSquared
PseudoR2(Other_involvement_logit)
```

```
        McFadden     Adj.McFadden        Cox.Snell       Nagelkerke McKelvey.Zavoina           Effron 
Count        Adj.Count 
       0.2137407        0.2078645        0.1693012        0.2918316        0.2528374        0.2262361 
0.8628251        0.1237785 
             AIC    Corrected.AIC 
    1346.0495582     1346.0700081 
```

Hide

```
logit2prob <- function(logit){
  odds <- exp(logit)
  prob <- odds / (1 + odds)
  return(prob)
}
```

Hide

```
prob_model1<-logit2prob(coef(Other_involvement_logit))
```

Hide

```
intercept1<-coef(Other_involvement_logit)[1]
```

model 2

Hide

```
#probability increase marijuana involvment intercept + heavy drinkin-model 1
(logit2prob(intercept1+coef(Other_involvement_logit)[2]*0.81493648))-0.07360378
```

```
(Intercept) 
  0.1363443 
```

Hide

```
#probability increase marijuana involvment intercept + smoker -model 1
(logit2prob(intercept1+coef(Other_involvement_logit)[3]*0.83851618))-0.07360378
```

```
(Intercept) 
  0.1665486 
```

Hide

```
#probability increase marijuana involvment intercept + other drug involvement -model 1
(logit2prob(intercept1+coef(Other_involvement_logit)[4]*0.84463878 ))-0.07360378
```

```
(Intercept) 
  0.1755151 
```

Hide

```
#take only one predictor and control group variables 1
Control_group1_logit <- glm(Marijuana_dummy ~ Heavy_drinker_dummy + Smoker_dummy+ Other_drugs+ Age+ Gender + 
Neighborhood_happiness +Neighborhood_safety  , data = df_model, family = "binomial")
summary(Control_group1_logit)
```

```

Call:
glm(formula = Marijuana_dummy ~ Heavy_drinker_dummy + Smoker_dummy + 
    Other_drugs + Age + Gender + Neighborhood_happiness + Neighborhood_safety, 
    family = "binomial", data = df_model)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.1912  -0.3952  -0.3844  -0.3773   2.3288  

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)    
(Intercept)            -2.52325    0.94557  -2.668  0.00762 ** 
Heavy_drinker_dummy     1.48644    0.17885   8.311  < 2e-16 ***
Smoker_dummy            1.64367    0.15397  10.675  < 2e-16 ***
Other_drugs             1.70536    0.24828   6.869 6.48e-12 ***
Age                     0.01923    0.05090   0.378  0.70563    
Gender                  0.02179    0.15056   0.145  0.88494    
Neighborhood_happiness  0.01423    0.09185   0.155  0.87686    
Neighborhood_safety    -0.26816    0.21834  -1.228  0.21938    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1701.8  on 1960  degrees of freedom
Residual deviance: 1336.3  on 1953  degrees of freedom
AIC: 1352.3

Number of Fisher Scoring iterations: 5
```

Hide

```
#psudo R_SSquared
PseudoR2(Control_group1_logit )
```

```
        McFadden     Adj.McFadden        Cox.Snell       Nagelkerke McKelvey.Zavoina           Effron 
Count        Adj.Count 
       0.2147709        0.2041938        0.1700436        0.2931112        0.2547284        0.2269665 
0.8628251        0.1237785 
             AIC    Corrected.AIC 
    1352.2962399     1352.3700104 
```

Hide

```
#CI for standard errors
confint.default(Control_group1_logit)
```

```
                             2.5 %     97.5 %
(Intercept)            -4.37653791 -0.6699624
Heavy_drinker_dummy     1.13590395  1.8369768
Smoker_dummy            1.34189825  1.9454400
Other_drugs             1.21874337  2.1919781
Age                    -0.08054051  0.1189951
Gender                 -0.27330679  0.3168832
Neighborhood_happiness -0.16579971  0.1942647
Neighborhood_safety    -0.69608440  0.1597734
```

Hide

```
#CI for log-liklihood
confint(Control_group1_logit)
```

```
Waiting for profiling to be done...
```

```
                            2.5 %     97.5 %
(Intercept)            -4.38249564 -0.6731315
Heavy_drinker_dummy     1.13471887  1.8364407
Smoker_dummy            1.34114729  1.9451546
Other_drugs             1.22201467  2.1972966
Age                    -0.08092875  0.1187623
Gender                 -0.27106742  0.3197425
Neighborhood_happiness -0.16380537  0.1965582
Neighborhood_safety    -0.68899196  0.1682253
```

Hide

```
# wald test for joint significance e.g neighborhood
library(aod)
wald.test(b = coef(Control_group1_logit), Sigma = vcov(Control_group1_logit), Terms = 6:7)
```

```
Wald test:
----------

Chi-squared test:
X2 = 0.046, df = 2, P(> X2) = 0.98
```

Hide

```
#take only one predictor and control group variables 2
Control_group2_logit <- glm(Marijuana_dummy  ~ Heavy_drinker_dummy + Smoker_dummy+ Other_drugs+ GPA + Suspen
sion +Peer_trouble +Teacher_trouble + Family_love+ Best_friends_marijuana_consumption   , data = df_model, f
amily = "binomial")
summary(Control_group2_logit)
```

```

Call:
glm(formula = Marijuana_dummy ~ Heavy_drinker_dummy + Smoker_dummy + 
    Other_drugs + GPA + Suspension + Peer_trouble + Teacher_trouble + 
    Family_love + Best_friends_marijuana_consumption, family = "binomial", 
    data = df_model)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.5580  -0.3684  -0.2231  -0.1893   2.8613  

Coefficients:
                                   Estimate Std. Error z value Pr(>|z|)    
(Intercept)                        -3.93261    0.23347 -16.844  < 2e-16 ***
Heavy_drinker_dummy                 0.93113    0.20899   4.455 8.37e-06 ***
Smoker_dummy                        1.05730    0.18489   5.719 1.07e-08 ***
Other_drugs                         0.80952    0.29247   2.768 0.005642 ** 
GPA                                 0.05376    0.11802   0.455 0.648777    
Suspension                          0.34810    0.22932   1.518 0.129030    
Peer_trouble                       -0.06403    0.09060  -0.707 0.479766    
Teacher_trouble                     0.32922    0.09109   3.614 0.000301 ***
Family_love                         0.06153    0.11906   0.517 0.605272    
Best_friends_marijuana_consumption  1.26673    0.07978  15.879  < 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: 1701.8  on 1960  degrees of freedom
Residual deviance: 1002.6  on 1951  degrees of freedom
AIC: 1022.6

Number of Fisher Scoring iterations: 6
```

Hide

```
prob_model2<-logit2prob(coef(Control_group2_logit))
prob_model2
```

```
                       (Intercept)                Heavy_drinker_dummy                       Smoker_dummy
                        0.01921601                         0.71730388                         0.74217440
                       Other_drugs                                GPA                         Suspension
                        0.69200620                         0.51343568                         0.58615631
                      Peer_trouble                    Teacher_trouble                        Family_love
                        0.48399846                         0.58156910                         0.51537809
Best_friends_marijuana_consumption
                        0.78018236
```

Hide

```
intercept2<-coef(Control_group2_logit)[1]
```

model 3

Hide

```
#probability increase marijuana involvment intercept + heavy drinkin-model 2
(logit2prob(intercept2+coef(Other_involvement_logit)[2]*0.71316770))- 0.01202153
```

```
(Intercept)
 0.04138816
```

Hide

```
#probability increase marijuana involvment intercept + smoker -model 2
(logit2prob(intercept2+coef(Other_involvement_logit)[3]*0.74019310 ))-0.01202153
```

(Intercept) 0.05021237

Hide

```
#probability increase marijuana involvment intercept + other drug involvement -model 1
(logit2prob(intercept2+coef(Other_involvement_logit)[4]*0.69035431 ))-0.01202153
```

```
(Intercept)
 0.04731661
```

Hide

```
#take only one predictor and control group variables 3
Control_group3_logit <- glm(Marijuana_dummy  ~ Smoker_dummy + Heavy_drinker_dummy + Other_drugs+ Depressed +
Fear + Lonely + Sad  , data = df_model, family = "binomial")
summary(Control_group3_logit)
```

```

Call:
glm(formula = Marijuana_dummy ~ Smoker_dummy + Heavy_drinker_dummy + 
    Other_drugs + Depressed + Fear + Lonely + Sad, family = "binomial", 
    data = df_model)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.2875  -0.4117  -0.3768  -0.3707   2.3626  

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)    
(Intercept)         -2.60928    0.11513 -22.664  < 2e-16 ***
Smoker_dummy         1.64699    0.15413  10.686  < 2e-16 ***
Heavy_drinker_dummy  1.48662    0.17823   8.341  < 2e-16 ***
Other_drugs          1.67989    0.25063   6.703 2.05e-11 ***
Depressed           -0.06027    0.12520  -0.481   0.6302    
Fear                -0.02396    0.13851  -0.173   0.8627    
Lonely               0.20785    0.12026   1.728   0.0839 .  
Sad                  0.02629    0.13882   0.189   0.8498    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1701.8  on 1960  degrees of freedom
Residual deviance: 1334.3  on 1953  degrees of freedom
AIC: 1350.3

Number of Fisher Scoring iterations: 5
```

Hide

```
stargazer(Control_group1_logit,type = "text")
```

```
==================================================
                           Dependent variable:    
                       ---------------------------
                             Marijuana_dummy      
--------------------------------------------------
Heavy_drinker_dummy             1.486***          
                                 (0.179)          
                                                  
Smoker_dummy                    1.644***          
                                 (0.154)          
                                                  
Other_drugs                     1.705***          
                                 (0.248)          
                                                  
Age                               0.019           
                                 (0.051)          
                                                  
Gender                            0.022           
                                 (0.151)          
                                                  
Neighborhood_happiness            0.014           
                                 (0.092)          
                                                  
Neighborhood_safety              -0.268           
                                 (0.218)          
                                                  
Constant                        -2.523***         
                                 (0.946)          
                                                  
--------------------------------------------------
Observations                      1,961           
Log Likelihood                  -668.148          
Akaike Inf. Crit.               1,352.296         
==================================================
Note:                  *p<0.1; **p<0.05; ***p<0.01
```

Hide

```
stargazer(Control_group2_logit,type = "text")
```

```
==============================================================
                                       Dependent variable:    
                                   ---------------------------
                                         Marijuana_dummy      
--------------------------------------------------------------
Heavy_drinker_dummy                         0.931***          
                                             (0.209)          
                                                              
Smoker_dummy                                1.057***          
                                             (0.185)          
                                                              
Other_drugs                                 0.810***          
                                             (0.292)          
                                                              
GPA                                           0.054           
                                             (0.118)          
                                                              
Suspension                                    0.348           
                                             (0.229)          
                                                              
Peer_trouble                                 -0.064           
                                             (0.091)          
                                                              
Teacher_trouble                             0.329***          
                                             (0.091)          
                                                              
Family_love                                   0.062           
                                             (0.119)          
                                                              
Best_friends_marijuana_consumption          1.267***          
                                             (0.080)          
                                                              
Constant                                    -3.933***         
                                             (0.233)          
                                                              
--------------------------------------------------------------
Observations                                  1,961           
Log Likelihood                              -501.297          
Akaike Inf. Crit.                           1,022.594         
==============================================================
Note:                              *p<0.1; **p<0.05; ***p<0.01
```

Hide

```
stargazer(Control_group3_logit,type = "text")
```

```
===============================================
                        Dependent variable:    
                    ---------------------------
                          Marijuana_dummy      
-----------------------------------------------
Smoker_dummy                 1.647***          
                              (0.154)          
                                               
Heavy_drinker_dummy          1.487***          
                              (0.178)          
                                               
Other_drugs                  1.680***          
                              (0.251)          
                                               
Depressed                     -0.060           
                              (0.125)          
                                               
Fear                          -0.024           
                              (0.139)          
                                               
Lonely                        0.208*           
                              (0.120)          
                                               
Sad                            0.026           
                              (0.139)          
                                               
Constant                     -2.609***         
                              (0.115)          
                                               
-----------------------------------------------
Observations                   1,961           
Log Likelihood               -667.173          
Akaike Inf. Crit.            1,350.347         
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

Hide

```
#psudo R_SSquared - model 3
PseudoR2(Control_group2_logit )
```

model 4

Hide

```
#take only one predictor and control group variables 3 - model 4
Control_group3_logit <- glm(Marijuana_dummy  ~ Smoker_dummy + Heavy_drinker_dummy + Other_drugs+ Depressed +
Fear + Lonely + Sad  , data = df_model, family = "binomial")
summary(Control_group3_logit)
```

Hide

```
stargazer(Control_group3_logit,type = "text")
```

Hide

```
prob_model4<-logit2prob(coef(Control_group3_logit))
prob_model4
```

Hide

```
intercept4<-coef(Control_group3_logit)[1]
```

Hide

```
#probability increase marijuana involvment intercept + heavy drinkin-model 4
(logit2prob(intercept4+coef(Control_group3_logit)[3]* 0.81557067))- 0.06854373
```

Hide

```
#probability increase marijuana involvment intercept + smoker -model 4
(logit2prob(intercept4+coef(Control_group3_logit)[2]* 0.83848325 ))-0.06854373
```

Hide

```
#probability increase marijuana involvment intercept + other drug involvement -model 4
(logit2prob(intercept4+coef(Control_group3_logit)[4]*0.84288949  ))-0.06854373
```

Hide

```
#psudo R_SSquared - model 4
PseudoR2(Control_group3_logit )
```

MDOEL 5

Hide

```
#complete model - model 5
Complete_logit <- glm(Marijuana_dummy  ~ Age + Gender  + Neighborhood_happiness +Neighborhood_safety + GPA +
Suspension +Peer_trouble +Teacher_trouble + Best_friends_marijuana_consumption + Smoker_dummy + Heavy_drinke
r_dummy + Other_drugs +Depressed  + Fear + Lonely + Family_love + Sad  , data = df_model, family = "binomia
l")
summary(Complete_logit)
```

Hide

```
stargazer(Complete_logit,type = "text")
```

Hide

```
#psudo R_SSquared - model 4
PseudoR2(Complete_logit )
```

Hide

```
prob_model5<-logit2prob(coef(Complete_logit))
prob_model5
```

Hide

```
intercept5<-coef(Complete_logit)[1]
```

Hide

```
#probability increase marijuana involvment intercept + heavy drinkin-model 5
(logit2prob(intercept5+coef(Complete_logit)[12]* 0.722927663 ))- 0.008993329
```

Hide

```
#probability increase marijuana involvment intercept + smoke model 5
(logit2prob(intercept3+coef(Complete_logit)[11]*0.736436995))- 0.008993329
```

Hide

```
#probability increase marijuana involvment intercept +other drugs-model 5
(logit2prob(intercept3+coef(Complete_logit)[13]*0.698453269  ))- 0.008993329
```

COLLINEARITY

Hide

```
#correlation problem
vif(Complete_logit)
```

STEPWISE SELECTION

Hide

```
#forward selection
null<- glm(Marijuana_dummy  ~ 1 , data = df_model, family = "binomial")
step(null, scope = list(lower = null, upper = Complete_logit), direction = "forward")
```

MODEL 6 - Final model

Hide

```
#forward model- model 6
Forward_logit <- glm(Marijuana_dummy  ~ Best_friends_marijuana_consumption +
    Smoker_dummy + Heavy_drinker_dummy + Teacher_trouble + Other_drugs +
    Fear + Suspension , data = df_model, family = "binomial")
summary(Forward_logit)
```

Hide

```
#psudo R_SSquared forward model - model 6
PseudoR2(Forward_logit)
```

Hide

```
#backward selection
step(Complete_logit, null, direction = "backward")
```

Hide

```
step(null, scope = list(lower = null, upper = Complete_logit), direction = "both")
```

Hide

```
#plot residuals
ggplot(Forward_logit, aes(.fitted, .resid)) + geom_point() + geom_hline(yintercept = 0) + geom_smooth(aes(.f
itted, .resid), model= "loess")
```

Hide

```
#forward model
model_final<- glm(Marijuana_dummy  ~  Best_friends_marijuana_consumption +
    Smoker_dummy + Heavy_drinker_dummy + Teacher_trouble + Other_drugs +
    Fear + Suspension , data = df_model, family = "binomial")
summary(Forward_logit)
```

Hide

```
#psudo R_SSquared -  model 6
PseudoR2(model_final )
```

Hide

```
prob_model_f<-logit2prob(coef(model_final))
prob_model_f
```

Hide

```
stargazer(model_final,type = "text")
```

Hide

```
intercept_f<-coef(Forward_logit)[1]
```

Hide

```
#probability increase marijuana involvment intercept + heavy drinkin-model final
(logit2prob(intercept_f+coef(Forward_logit)[4]*0.72004991 ))- 0.01894008
```

Hide

```
#probability increase marijuana involvment intercept + smoke model final
(logit2prob(intercept_f+coef(Forward_logit)[3]*0.74133675))- 0.01894008
```

Hide

Hide

```
#probability increase marijuana involvment intercept +other drugs-model final
(logit2prob(intercept_f+coef(Forward_logit)[6]*0.69252969  ))- 0.01894008
```

Hide

```
Sys.getenv("RSTUDIO_PANDOC")
```

Hide

```
install.packages("knitr")
Sys.setenv(RSTUDIO_PANDOC='/anaconda3/bin/pandoc')
rmarkdown::render('Health_marijuana_v13.Rmd', 'all')
```

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Cmd+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.