A black and white photograph of a person sitting cross-legged on a polished floor, leaning against a wall. They are holding and reading a book. The scene is dimly lit from above, creating a dramatic shadow on the floor and wall. The background shows a closed door.

Depression & Drug Use

Vanessa Okosun
Capstone Project – BDA 650
Mercer University Stetson School of Business

AGENDA

01

Statistics

02

Objective

03

NHNES
Background

04

Insights

05

Results

06

Recomm-
endations

07

References

STATISTICS

- Depression is defined as a period of at least two weeks when a person experienced a depressed mood or loss of interest or pleasure in daily activities and had specified symptoms, such as problems with sleep, eating, energy, concentration, or self-worth.
- Major depression is one of the most common mental disorders in the United States and about 18.5% of adults had symptoms of depression that were either mild, moderate, or severe in the past 2 weeks.

- The number of overdose deaths increases annually at 4.0%.



50% of people from the **age of 12** have used illicit drugs at least once.



22% of males and **17% of females** used illegal drugs or misused prescription drugs within the last year.



OBJECTIVE

Does being depressed and other external factors increase the chances of a person being a drug user?

National Health and Nutrition Examination Survey

- 2017-2018
- program of studies designed to assess the health and nutritional status of adults and children in the United States
- combines interviews and physical examinations
- Findings from this survey will be used to determine the prevalence of major diseases and risk factors for diseases.
- The survey examines a nationally representative sample of about 5,000 persons each year. These persons are located in counties across the country, 15 of which are visited each year.

Example of Questions asked in scales

DPINTEREST	Have little interest in doing things
DPFEEL	how often have you been bothered by the following problems: feeling down, depressed, or hopeless?
DPSLEEP	Trouble sleeping or sleeping too much
DPENERGY	Feeling tired or having little energy
DPAPPETITE	Poor appetite or overeating
DPSELF	Feeling bad about yourself
DPCONCEN	Trouble concentrating on things
DPSPEAK	Moving or speaking slowly or too fast
DPTHOUGHT	Thought you would be better off dead
DPDifficulty	Difficulty these problems have caused

DUQ211	Used marijuana every month for a year?
DUQ215Q	Time since last used marijuana regularly
DUQ230	# days used marijuana or hashish/month
DUQ250	Ever use any form of cocaine
DUQ280	# of days used cocaine/month
DUQ290	Ever used heroin
DUQ320	# of days used heroin/month
DUQ330	Ever used methamphetamine
DUQ340	Age first used methamphetamine
DUQ350Q	Last time used methamphetamine
DUQ360	# days used methamphetamine/month
DUQ370	Ever use a needle to inject illegal drug
DUQ410	# times injected drugs/lifetime
DUQ430	Ever been in rehabilitation program

Data Insights

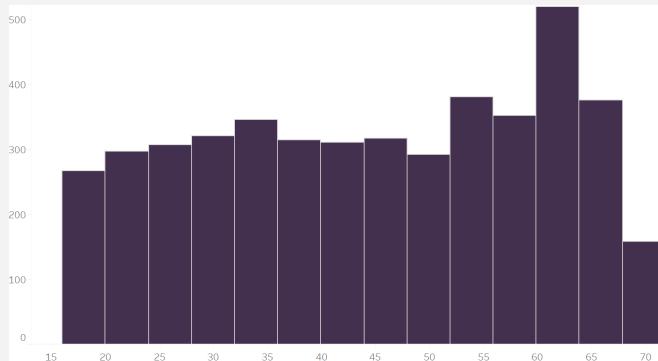
Data Insights

Gender

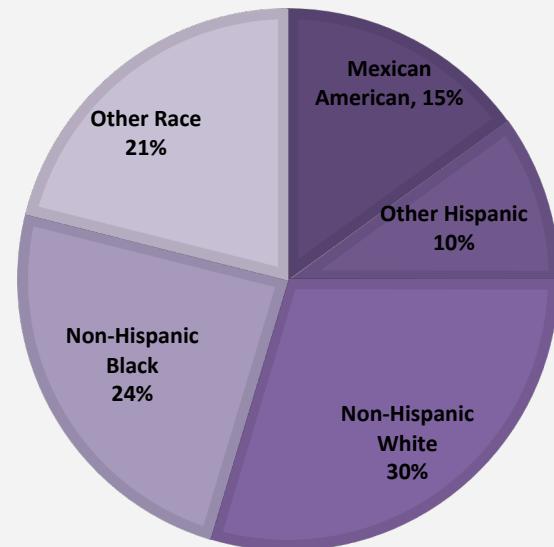


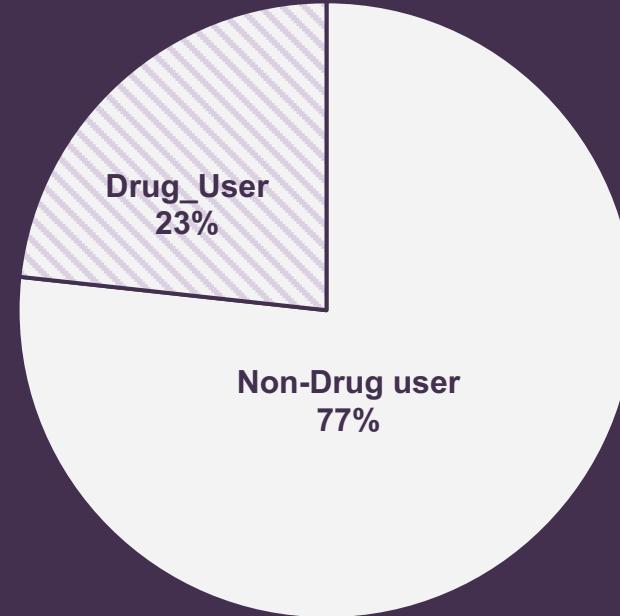
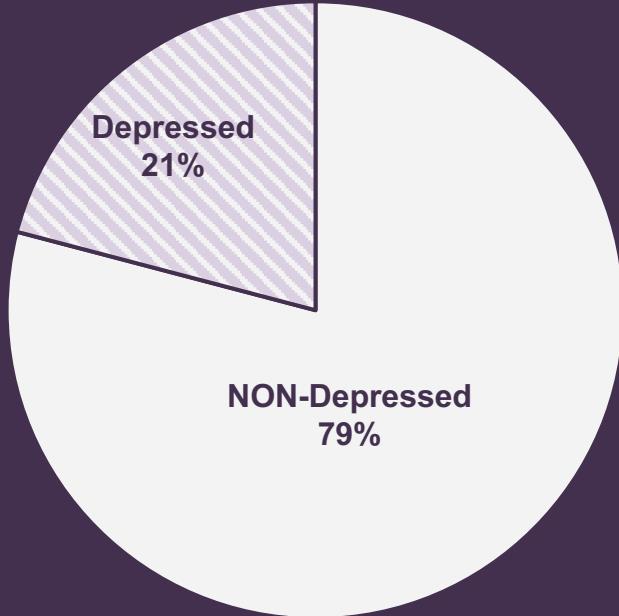
Male

Age



Race/Hispanic Origin



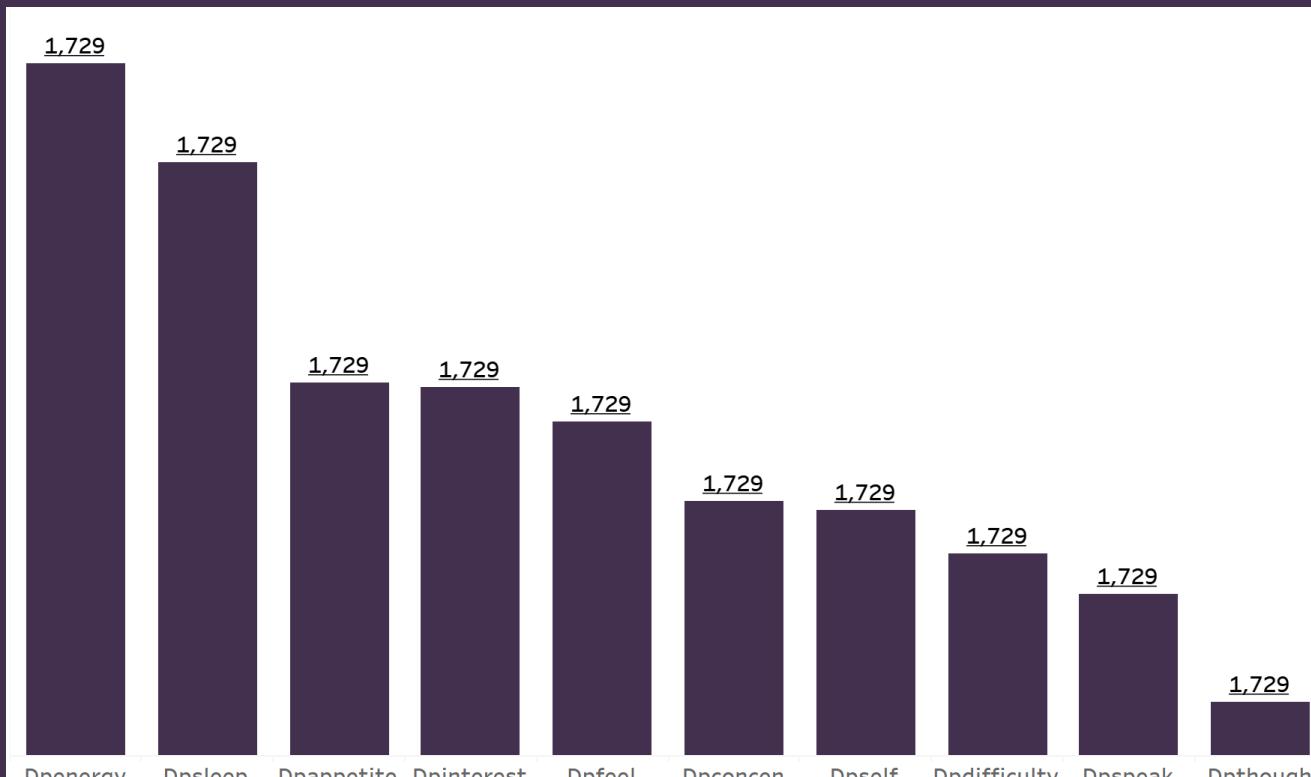


48 variables and **4572** observations in this dataset.
10 variables were used for **depression** classification.
6 variables were used for **drug user** classification

Depression Classification

Depression Scale
(categorized into 5 levels):

normal (score 0 – 4), minimal (score 5 – 9), mild (score 10 – 14), moderate (score 15 – 19), and moderately severe (score > 19).



Drug User Classification

Drug Use was classified if they have tried the following drugs at least once. The more drugs they try the higher the drug index

- Marijuana
- Cocaine
- Heroine
- Methamphetamine
- Illegal Injected Drugs

Also being Rehab Patient was considered

Drug Users

Cocaine Users	639
Heroin Users	107
Marijuana Users	1,738
Methamphetamine Users	285
Rehab Patient	186

Logic Regression

MODEL 1:

Drug User

$$\begin{aligned} &= B_0 + B_1(\text{Depression}) + B_2(\text{Gender}) \\ &+ B_3(\text{Race}) + B_4(\text{Educ}) + B_5(\text{Martl stat}) \\ &+ B_6(\text{Poverty Rat}) \end{aligned}$$

Variable	P-value
Depressed	Significant
Gender	Significant
Race	Significant
EDUC	Not significant
MARTL	Significant
POVRAT	Not Significant

MODEL 2:

DU(Drug User)

$$\begin{aligned} &= B_0 + B_1(\text{Depression}) + B_2(\text{Gender}) \\ &+ B_3(\text{Race}) + B_5(\text{Martl stat}) \end{aligned}$$

Variable	P-Value
Depressed	Significant
Gender	Significant
MARTL	Significant
Race	Significant

- Those who are depressed are 49% more likely to be drug users
- Females are 48% less likely to be drug users.
- African Americans are 29% more likely to be drug users
- Those who are married are 57% less likely to be drug users

Model Comparison

Model selection based on AICc:

	K	AICc	Delta_AICc	AICcWt	Cum.Wt	LL
dep.gen.mart1	5	3291.60	0.00	0.82	0.82	-1640.79
dep.gen.rac.educ.mart1.pov	7	3294.65	3.05	0.18	1.00	-1640.31

AICc

The AIC value of the model. The model that explains the most variation in the data has a lower value

AICcWt

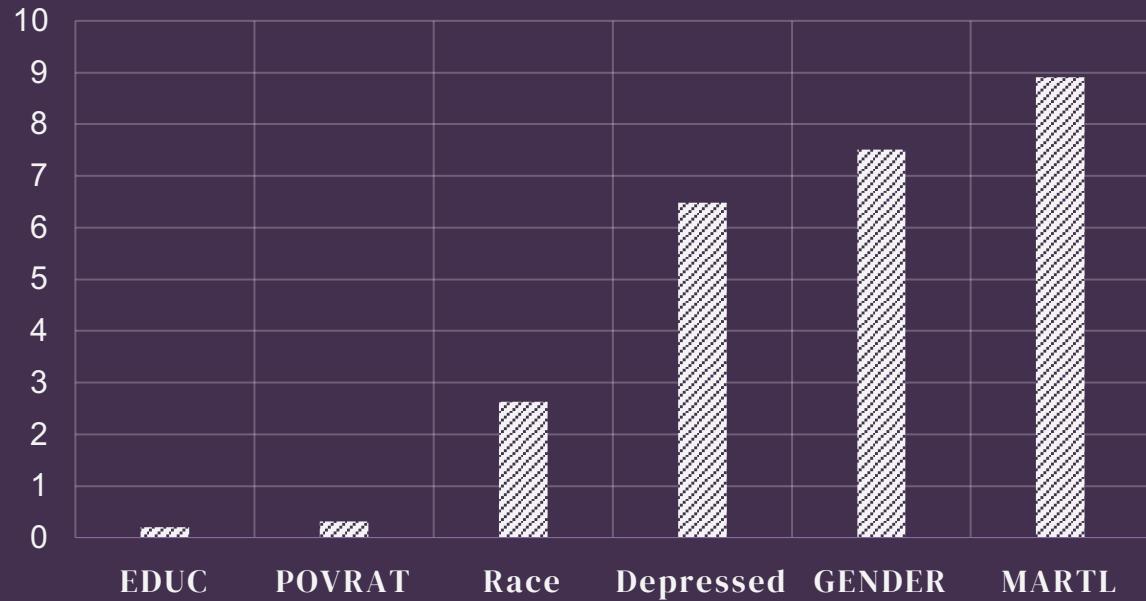
The proportion of the total predictive power that can be found in the model.

Variable Importance

Shows the importance of each predictor variable in the model.

The higher the score, the more important the variable is

Variable Importance



Prediction

PERSON A

Gender	Female
Race	African American
Education	Highschool or below
Marital Status	Not Married
Poverty Ratio	2.17
Depressed	Yes

Being Female, African American, Up to high school education, not being married, a poverty ratio of 2.17, and Depressed status of "Yes" has a predicted probability of being a drug user of .3796.

PERSON B

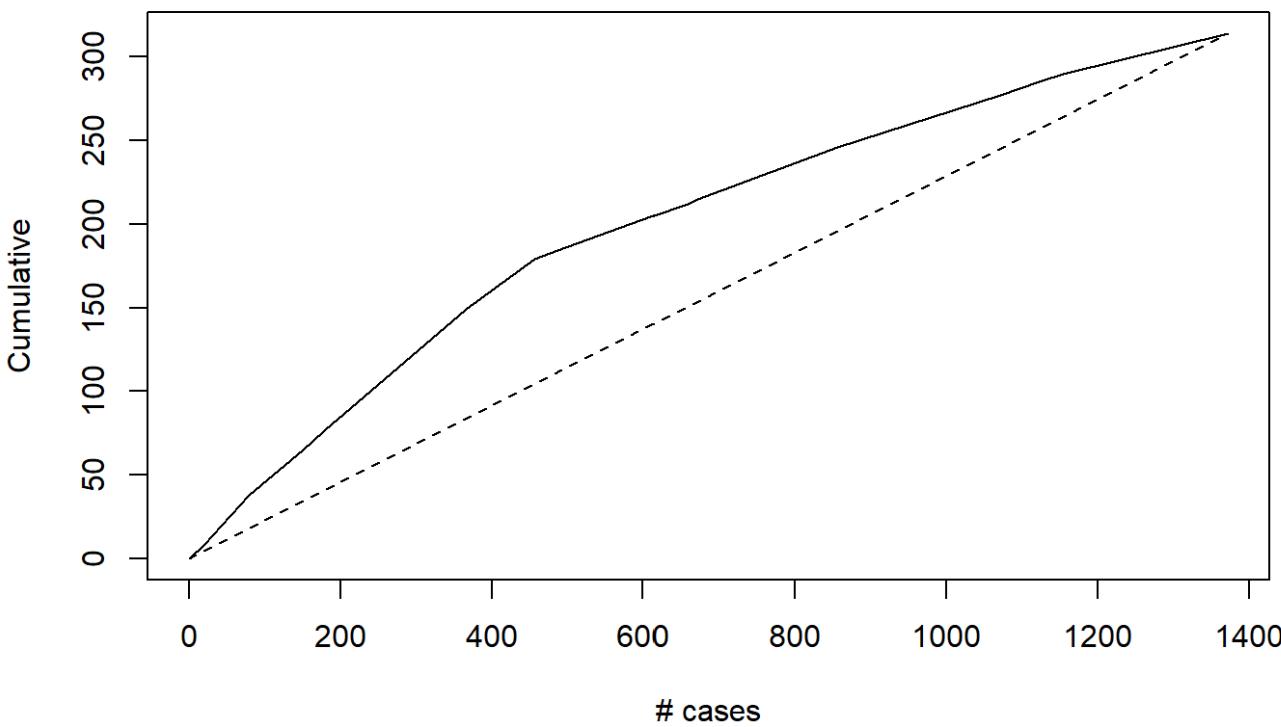
Gender	Female
Race	African American
Education	Highschool or below
Marital Status	Not Married
Poverty Ratio	2.17
Depressed	No

Conversely, an individual with the same gender, race, marital status, educ, and poverty rate but with a depression status of "No" has a predicted probability of being a drug user of 0.2414.

Lift Chart

The charts are based on the validation data of 1400 people.

The model's predictive performance in terms of lift is better than the baseline model since its lift curve is higher than that of the baseline model.



Simultaneous Equations Model

DRUG USE

Depression
Marij_price - \$7.38 price/g
Cocaine_price - \$84 price/g
Heroin_price - \$69 price/g
Meth_price - \$40 price/g

Variable	Coefficient	P- value
Depressed	0.1498	0.0082
Marij_price	0.0610	2.22E-16
Cocaine_price	0.0043	2.22E-16
Heroin_price	0.0034	9.99E-14
Meth_price	0.0022	9.11E-05

DEPRESSION

Drug User
Marital Status
Gender
Income

Variable	Coefficient	P- value
Drug_user	0.215	2.22E-16
Gender	0.076	2.17873E-10
MARTL	-0.056	7.9546E-06
Income	-0.121	2.22E-16

Recommendations



Mental health checkups should be a part of yearly physicals for prevention purposes.



Early intervention strategies can reduce the impact of substance use and mental disorders in America's communities



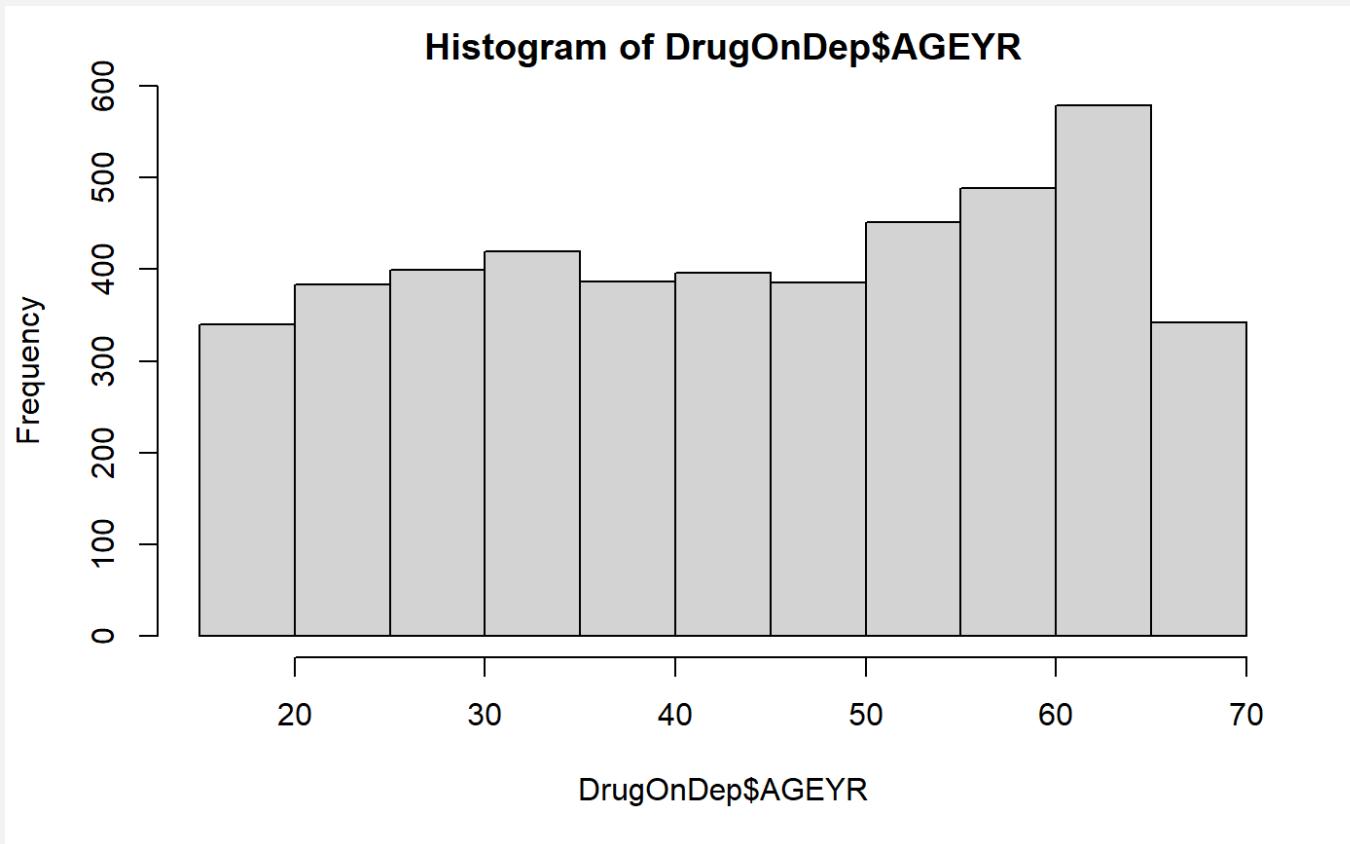
More research to boost protective factors and eliminate or reduce risk factors for drug use



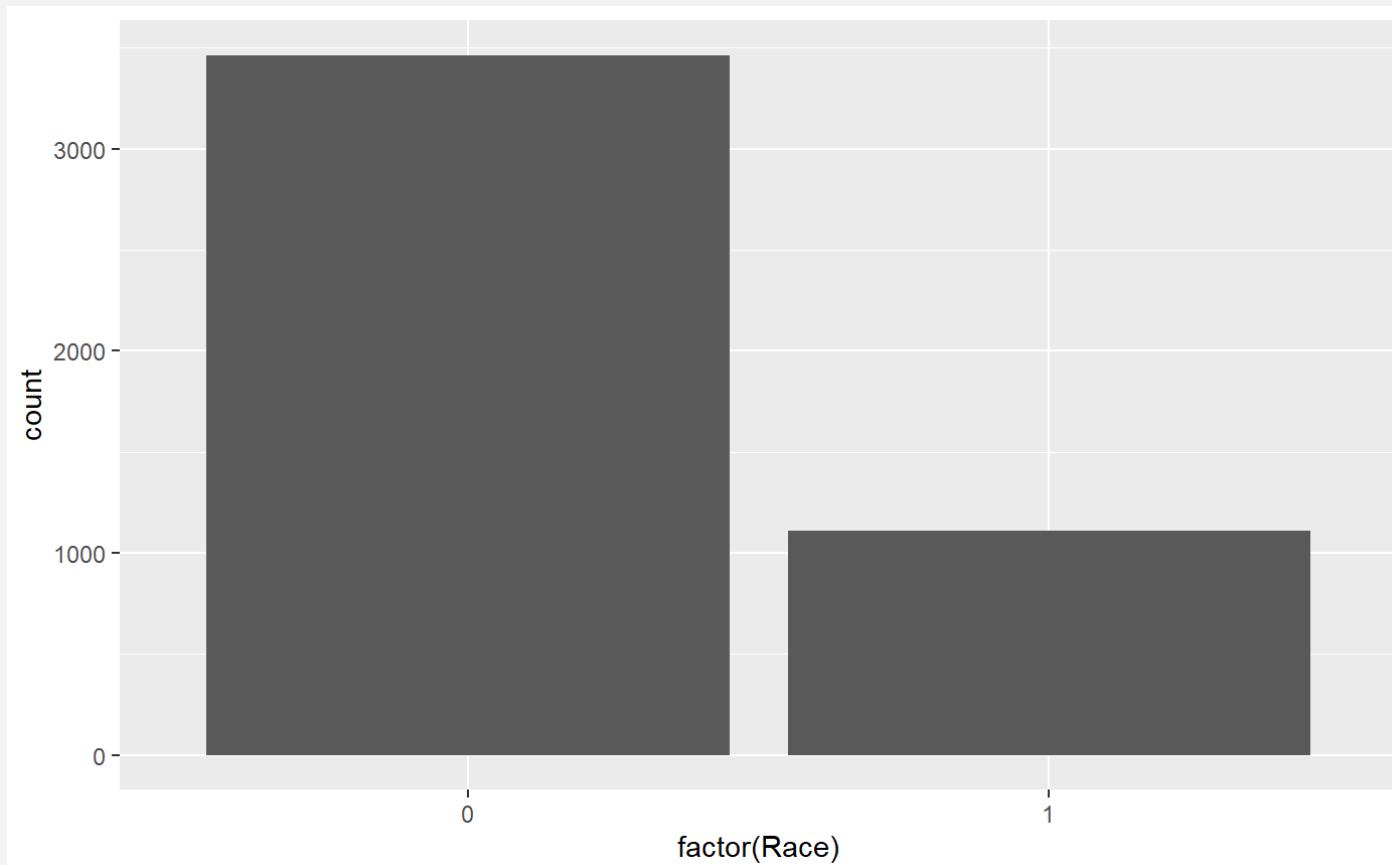
Questions?

APPENDIX

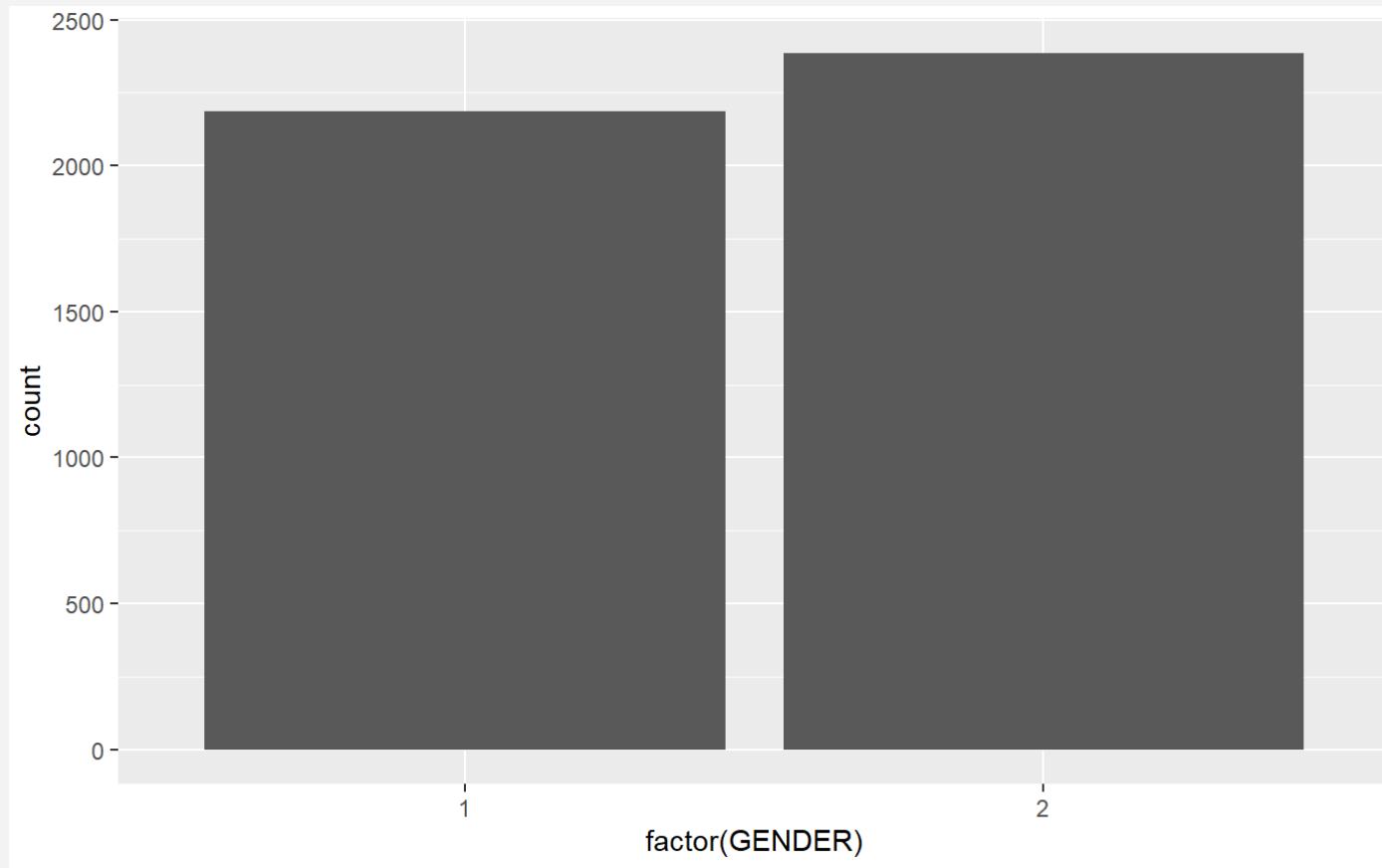
Appendix 1 - Visualizations



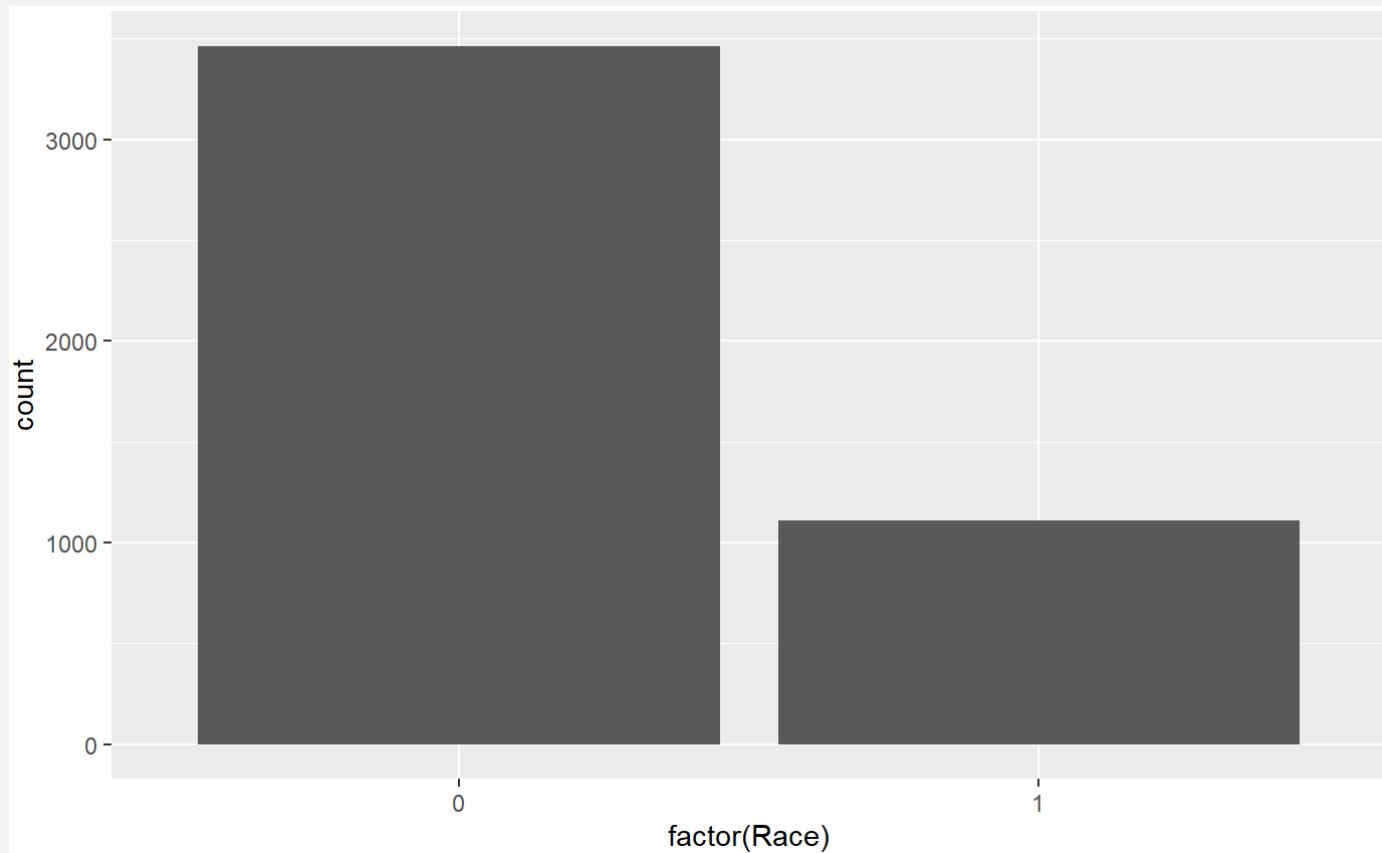
Appendix 2 - Visualizations



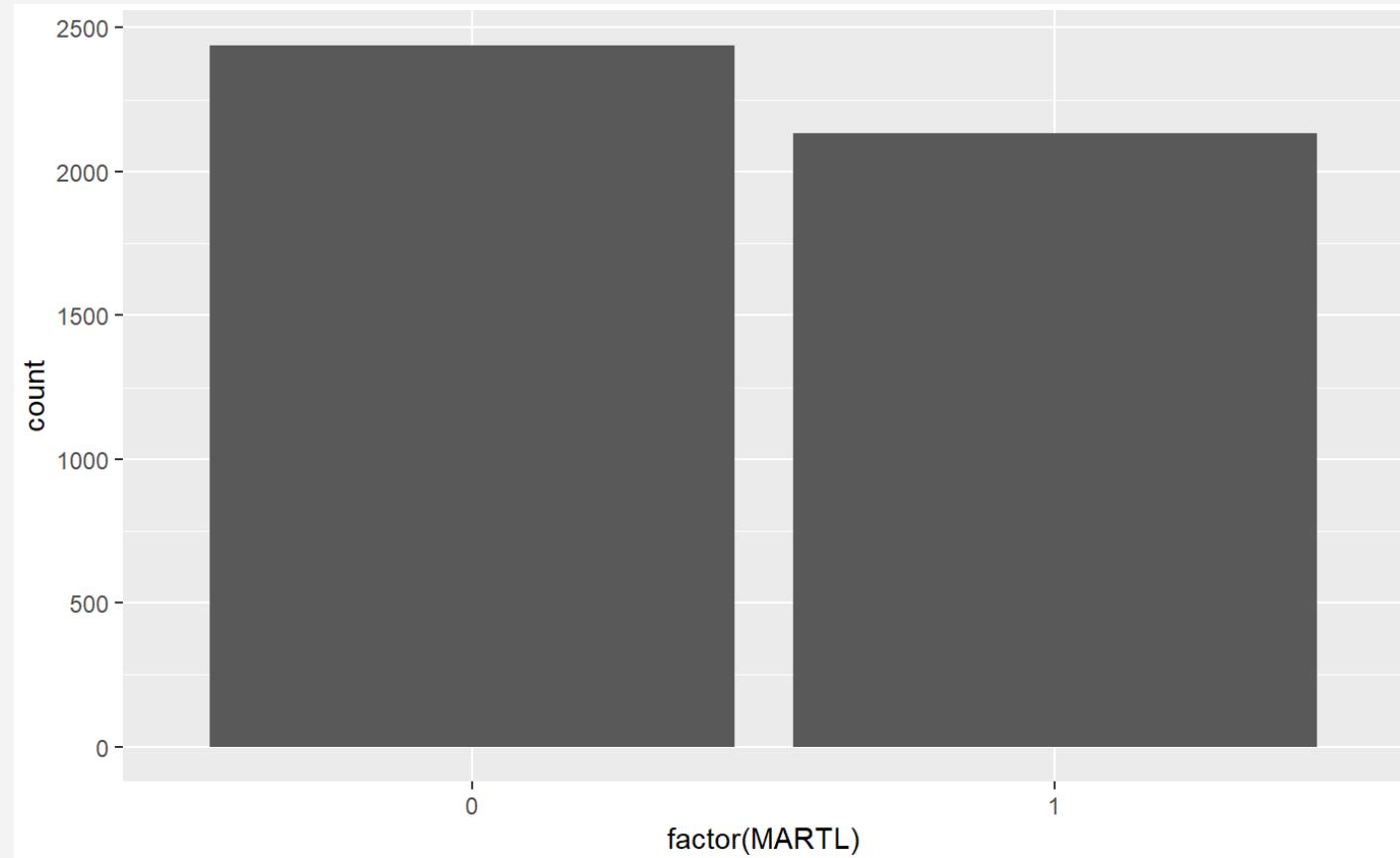
Appendix 3 - Visualizations



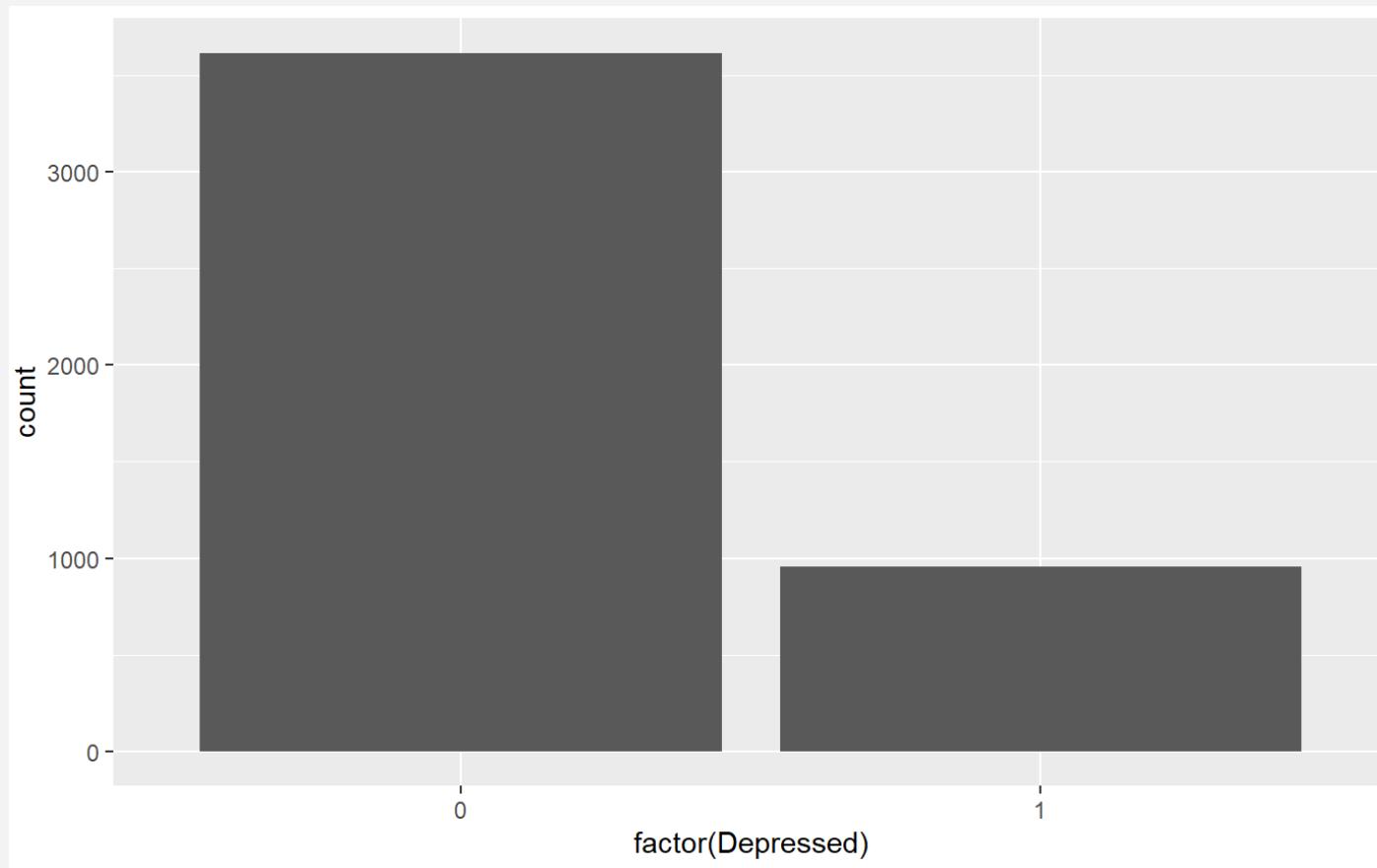
Appendix 4 - Visualizations



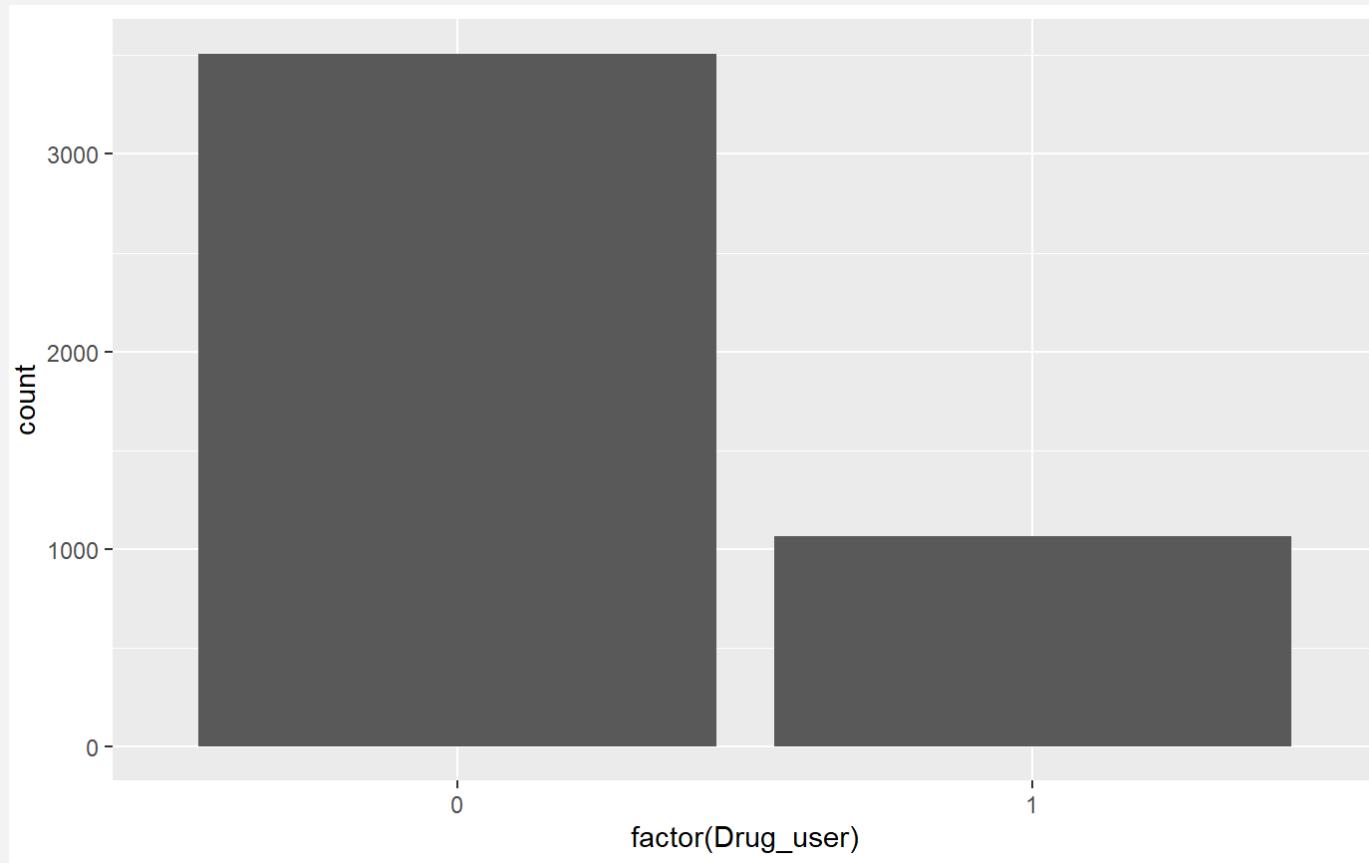
Appendix 5 - Visualizations



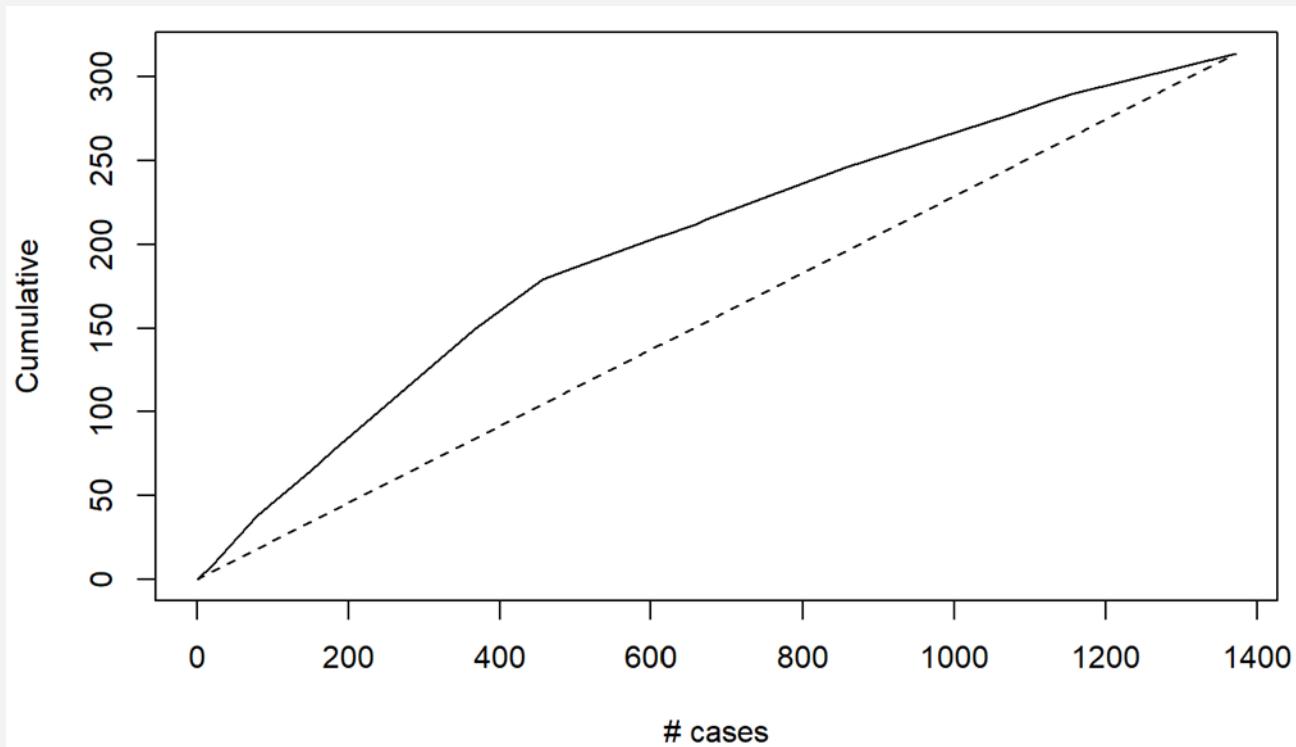
Appendix 6 - Visualizations



Appendix 7 - Visualizations



Appendix 8 - Visualizations



Appendix 9 – Variable Counts

Row Labels	# of Depressed/Non-Depressed
0	3616
1	956
Grand Total	4572

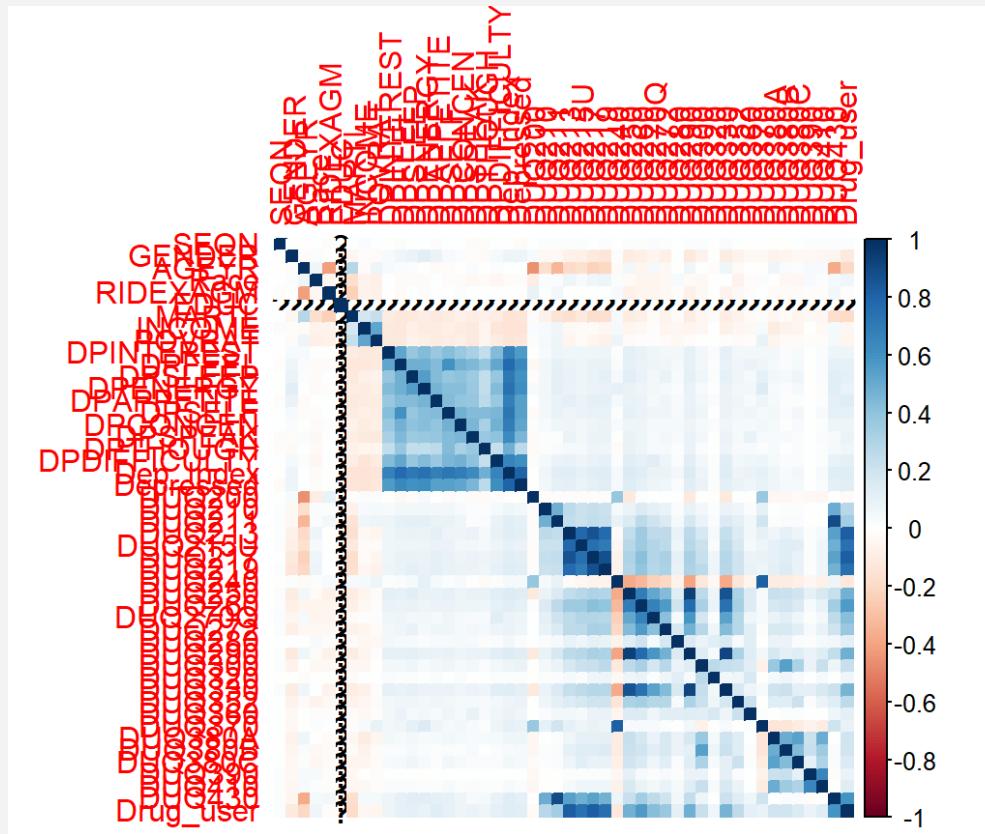
Row Labels	Count of Drug_user
0	3508
1	1064
Grand Total	4572

Row Labels	Count of GENDER
1	2186
2	2386
Grand Total	4572

Row Labels	Cocaine Users
0	3901
1	639
2	30
7	2
Grand Total	4572

Row Labels	Count of Race
1	687
2	458
3	1354
4	1109
5	964
Grand Total	4572

Appendix 10 - Visualizations



Appendix 11 – Creating Binary Variables

```
CLEANING DATA
```{r}
#remove unimportant columns
DrugOnDep <- subset(DrugOnDep, select = -c(RIDEXAGM))

#reduce values in race origin level 0 - other races ;1 - black non-hispanic
DrugOnDep$Race <- ifelse(DrugOnDep$Race==4,1,0)

#reduce values in Education level 0 - highschool equiv or less than; 1 - college and above
DrugOnDep$EDUC <- ifelse(DrugOnDep$EDUC==4 | DrugOnDep$EDUC==5, 1,0)

#reduce values in Marital level 0 - not married; 1 - married
DrugOnDep$MARTL <- ifelse(DrugOnDep$MARTL==1, 1,0)

#reduce values in Income level; 0- under $20,000, 1 - over $20,000
DrugOnDep$INCOME <- ifelse(DrugOnDep$INCOME<= 4 | DrugOnDep$INCOME == 13, 0, 1)
```

## Appendix 12 – Data Structure

	skim_variable <chr>	n_missing <int>	complete_rate <dbl>	mean <dbl>	sd <dbl>	p0 <dbl>	p25 <dbl>	▶
1	SEQN	0	1.0000000	98278.96653543	2690.9602843	93705	95949.50	
2	GENDER	0	1.0000000	1.52187227	0.4995760	1	1.00	
3	AGEYR	0	1.0000000	44.24693788	15.4712188	18	31.00	
4	Race	0	1.0000000	0.24256343	0.4286801	0	0.00	
5	RIDEXAGM	0	1.0000000	13.05249344	52.9700332	0	0.00	
6	EDUC	1	0.9997813	0.53905054	0.4985273	0	0.00	
7	MARTL	0	1.0000000	0.46653543	0.4989334	0	0.00	
8	INCOME	0	1.0000000	0.78193351	0.4129780	0	1.00	
9	POVRAT	0	1.0000000	2.17251094	1.7451292	0	0.74	
10	DPINTEREST	0	1.0000000	0.37292213	0.7963997	0	0.00	

1-10 of 48 rows | 1-8 of 11 columns

Previous 1 2 3 4 5 Next

## Appendix 13 – Logistic Regression

LOGISTIC REGRESSION - MODEL 1

```
```{r}
library(AICcmodavg)

#partition data
set.seed(2)
train.index <- sample(c(1:dim(DrugOnDepFinal)[1]), dim(DrugOnDepFinal)[1]*0.7)
train.df <- DrugOnDepFinal[train.index, ]
valid.df <- DrugOnDepFinal[-train.index, ]

#run logistic regression
logit.reg <- glm(Drug_user~Depressed+ GENDER+ Race+ EDUC+ MARTL+ POVRAT, data = train.df, family = 'binomial')
options(scipen=999)
summary(logit.reg)
```

MODEL 2

```
```{r}
logit.reg2 <- glm(Drug_user~Depressed+ GENDER+ MARTL+ Race, data = train.df, family = 'binomial')
options(scipen=999)
summary(logit.reg2)
```
```

Appendix 14 – Logistic Regression

```
call:  
glm(formula = Drug_user ~ Depressed + GENDER + Race + EDUC +  
MARTL + POVRAT, family = "binomial", data = train.df)
```

Deviance Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|---------|---------|--------|
| -1.2667 | -0.7439 | -0.6150 | -0.4467 | 2.1726 |

Coefficients:

| | Estimate | Std. Error | z value | Pr(> z) |
|-------------|----------|------------|--------------------------------|-----------------------|
| (Intercept) | -0.10954 | 0.15378 | -0.712 | 0.47627 |
| Depressed | 0.65794 | 0.10144 | 6.486 | 0.000000000088358 *** |
| GENDER | -0.65949 | 0.08789 | -7.504 | 0.000000000000062 *** |
| Race | 0.25803 | 0.09800 | 2.633 | 0.00847 ** |
| EDUC | 0.01821 | 0.09231 | 0.197 | 0.84364 |
| MARTL | -0.83238 | 0.09348 | -8.905 < 0.000000000000002 *** | |
| POVRAT | 0.00848 | 0.02751 | 0.308 | 0.75791 |

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 3484.3 on 3198 degrees of freedom

Residual deviance: 3280.6 on 3192 degrees of freedom

(1 observation deleted due to missingness)

AIC: 3294.6

Number of Fisher Scoring iterations: 4

```
call:  
glm(formula = Drug_user ~ Depressed + GENDER + MARTL + Race,  
family = "binomial", data = train.df)
```

Deviance Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|---------|---------|--------|
| -1.2489 | -0.7435 | -0.6150 | -0.4527 | 2.1583 |

Coefficients:

| | Estimate | Std. Error | z value | Pr(> z) |
|-------------|----------|------------|--------------------------------|------------------------|
| (Intercept) | -0.08898 | 0.14369 | -0.619 | 0.53574 |
| Depressed | 0.65365 | 0.10066 | 6.493 | 0.0000000000839039 *** |
| GENDER | -0.65729 | 0.08767 | -7.497 | 0.0000000000000652 *** |
| MARTL | -0.82318 | 0.09163 | -8.984 < 0.000000000000002 *** | |
| Race | 0.25905 | 0.09755 | 2.655 | 0.00792 ** |

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 3484.9 on 3199 degrees of freedom

Residual deviance: 3281.6 on 3195 degrees of freedom

AIC: 3291.6

Number of Fisher Scoring iterations: 4

Appendix 15 – Model Comparison

MODEL Comparison

K: The number of parameters in the model.

AICc: The AIC value of the model. model that explains the most variation in the data has lower value

Delta_AICc: The difference between the AIC of the best model compared to the current model being compared.

AICcWt: The proportion of the total predictive power that can be found in the model.

Cum.Wt: The cumulative sum of the AIC weights.

LL: The log-likelihood of the model. This tells us how likely the model is, given the data we used.

MODEL 2 offers a better fit

```
```{r}
#List of models
models <- list(logit.reg, logit.reg2)

#model names
mod.names <- c('dep.gen.rac.educ.martl.pov','dep.gen.martl')

#Calculate AIC
aictab(cand.set = models, modnames = mod.names)
```
```

Model selection based on AICc:

| | K | AICc | Delta_AICc | AICcWt | Cum.Wt | LL |
|----------------------------|---|---------|------------|--------|--------|----------|
| dep.gen.martl | 5 | 3291.60 | 0.00 | 0.82 | 0.82 | -1640.79 |
| dep.gen.rac.educ.martl.pov | 7 | 3294.65 | 3.05 | 0.18 | 1.00 | -1640.31 |

Appendix 16 – Accessing the model

ACCESSING THE MODEL

McFadden's R², which ranges from 0 to just under 1. Values close to 0 indicate that the model has no predictive power. In practice, values over 0.40 indicate that a model fits the data very well.

```
```{r}
library(caret)

#PREDICTIVE POWER
pscl::pR2(logit.reg2)["McFadden"]
pscl::pR2(logit.reg)["McFadden"]
```

```
#COMPUTE IMPORTANCE OF EACH VARIABLE
caret::varImp(logit.reg)
```

```
#VIF - VALUE OVER 5 IS SEVERE multicollinearity
car::vif(logit.reg2)
````
```

R Console

```
Fitting null model for pseudo-R2
McFadden
0.05833019
Fitting null model for pseudo-R2
McFadden
0.05846401
0.05833019 0.05846401 0.05833019
0.05833019 0.05846401 0.05833019
0.05833019 0.05846401 0.05833019
0.05833019 0.05846401 0.05833019
0.05833019 0.05846401 0.05833019
0.05833019 0.05846401 0.05833019
```

data.frame
6 x 1

```
Variables: 6 (3 x 2)
          Overall
Dependent: F 4.029752
Df1,Df2: 7 3.013259
Model: 0.05833019
Residual: 0.05846401
AIC: 0.05833019
BIC: 0.05846401
Log Likelihood: 0.05833019
Deviance: 0.05846401
Nagelkerke: 0.05833019
```

```
fitting null model for pseudo-r2
  McFadden
0.05833019
fitting null model for pseudo-r2
  McFadden
0.05846401
```

Appendix 17 - Accessing the model

| | Overall
<dbl> |
|-----------|------------------|
| Depressed | 6.4856358 |
| GENDER | 7.5038355 |
| Race | 2.6329248 |
| EDUC | 0.1972357 |
| MARTL | 8.9045640 |
| POVRAT | 0.3082307 |

Depressed GENDER MARTL Race
1.028362 1.018499 1.037879 1.022086

Appendix 18 - Predictions

MAKE Predictions

a Depressed status of "Yes" has a probability of being a drug user of .3796. Conversely, an individual with the same gender, race, martl, educ, and poverty rate but with a depression status of "No" has a probability of being a drug user of 0.2414

```{r}



```
#define two individuals
new <- data.frame(GENDER = 2, Race = 1, MARTL = 0, EDUC = 0, POVRAT = 2.17, Depressed = c(1, 0))

#Predicted probability
predict(logit.reg2, new, type = "response")
```

...

```
1 2
0.3796904 0.2414937
```



## Appendix 19 - MODEL DIAGNOSTIC

```
MODEL DIAGNOSTIC
``{r}
Predicted <- predict(logit.reg2, valid.df, type = 'response')

u <- union(Predicted, valid.df$Drug_user)
t <- table(factor(Predicted, u), factor(valid.df$Drug_user, u))

confusionMatrix(t)

class
```
```

Appendix 20 - MODEL DIAGNOSTIC

0.320840782899427	0.211811687602561	0.212420092754185	0.341469189634343	0	1
0.241493745803033	0	0	0	0	98 19
0.0973754343535728	0	0	0	0	194 24
0.172295216192635	0	0	0	0	184 31
0.476860285596239	0	0	0	0	30 30
0.380548234855239	0	0	0	0	44 27
0.32163456556838	0	0	0	0	105 66
0.379690435951193	0	0	0	0	24 16
0.197252483647745	0	0	0	0	154 31
0.285814022293889	0	0	0	0	34 7
0.171776669370003	0	0	0	0	43 8
0.122637809764446	0	0	0	0	27 5
0.541510035726082	0	0	0	0	11 8
0.320840782899427	0	0	0	0	60 30
0.211811687602561	0	0	0	0	10 3
0.212420092754185	0	0	0	0	37 7
0.341469189634343	0	0	0	0	3 2
0	0	0	0	0	0 0 0
1	0	0	0	0	0 0 0

Appendix 21 - Simultaneous Equations Model

#Simultaneous Equations Model#

the endogenous variables are \$Depression\$ and \$Druguser\$.

the exogenous variables are the indicator variables for demographics of the user.

the identification variable for the drug user equation are gender, martial status and income , which will only show up in the depression equation but not in the drug user equation;

the identification variables for the depression equation will be drug prices, as none of them show up in the drug user equation.

```
```{r}
#Creating Drug Prices variables
DrugOnDepFinal$Marij_price <- ifelse(DrugOnDepFinal$DUQ200=="1",7.38, 0)
DrugOnDepFinal$Cocaine_price <- ifelse(DrugOnDepFinal$DUQ250=="1",84.00, 0)
DrugOnDepFinal$Heroin_price <- ifelse(DrugOnDepFinal$DUQ290=="1",69.00, 0)
DrugOnDepFinal$Meth_price <- ifelse(DrugOnDepFinal$DUQ330=="1",40.00, 0)

view(DrugOnDepFinal)
```

## Appendix 22 - Simultaneous Equations Model (2SLS)

```
```{r}
#install.packages('systemfit')
library(systemfit)

DrugOnDep.DU <- Drug_user~Depressed+Marij_price+Cocaine_price+Heroin_price+Meth_price
DrugOnDep.Dep <- Depressed ~ Drug_user+GENDER+MARTL+INCOME

DrugOnDep.eqs <- list(DrugOnDep.DU,DrugOnDep.Dep)
DrugOnDep.ivs <- ~Marij_price+Cocaine_price+Heroin_price+Meth_price+GENDER+MARTL+INCOME
DrugOnDep2SLS <- systemfit(DrugOnDep.eqs, method= "2SLS",
                           inst=DrugOnDep.ivs, data= DrugOnDepFinal)

summary(DrugOnDep2SLS)

```
```

## Appendix 23 - Simultaneous Equations Model (2SLS)

systemfit results

method: 2SLS

| N      | DF   | SSR  | detRCov | OLS-R2   | McElroy-R2 |
|--------|------|------|---------|----------|------------|
| system | 9144 | 9133 | 1112.25 | 0.011778 | 0.292681   |
|        |      |      |         |          | 0.467351   |

| N   | DF   | SSR  | MSE     | RMSE     | R2       | Adj R2   |
|-----|------|------|---------|----------|----------|----------|
| eq1 | 4572 | 4566 | 386.241 | 0.084591 | 0.290845 | 0.526889 |
|     |      |      |         |          |          | 0.526371 |
| eq2 | 4572 | 4567 | 726.009 | 0.158969 | 0.398709 | 0.039799 |
|     |      |      |         |          |          | 0.038958 |

The covariance matrix of the residuals

|     | eq1        | eq2        |
|-----|------------|------------|
| eq1 | 0.0845907  | -0.0408551 |
| eq2 | -0.0408551 | 0.1589685  |

2SLS estimates for 'eq1' (equation 1)

Model Formula: Drug\_user ~ Depressed + Marij\_price + Cocaine\_price + Heroin\_price +  
Meth\_price  
Instruments: ~Marij\_price + Cocaine\_price + Heroin\_price + Meth\_price + GENDER +  
MARTL + INCOME

The correlations of the residuals

|     | eq1       | eq2       |
|-----|-----------|-----------|
| eq1 | 1.000000  | -0.352314 |
| eq2 | -0.352314 | 1.000000  |

|                | Estimate     | Std. Error  | t value                             | Pr(> t )                |
|----------------|--------------|-------------|-------------------------------------|-------------------------|
| (Intercept)    | -0.031498545 | 0.010587155 | -2.97517                            | 0.0029436 **            |
| Depressed      | 0.149813717  | 0.056620856 | 2.64591                             | 0.0081751 **            |
| Marij_price    | 0.061048606  | 0.001416135 | 43.10933 < 0.000000000000000222 *** |                         |
| Cocaine_price  | 0.004308407  | 0.000195478 | 22.04036 < 0.000000000000000222 *** |                         |
| Heroin_price   | 0.003396203  | 0.000455006 | 7.46409                             | 0.00000000000009992 *** |
| Meth_price     | 0.002229905  | 0.000569330 | 3.91672                             | 0.00009107414204301 *** |
| ---            |              |             |                                     |                         |
| Signif. codes: | 0 '***'      | 0.001 '**'  | 0.01 '*'                            | 0.05 '.'                |
|                | 1            |             |                                     |                         |

Residual standard error: 0.290845 on 4566 degrees of freedom

Number of observations: 4572 Degrees of Freedom: 4566

SSR: 386.240995 MSE: 0.084591 Root MSE: 0.290845

Multiple R-Squared: 0.526889 Adjusted R-Squared: 0.526371

## Appendix 24 - Simultaneous Equations Model (2SLS)

2SLS estimates for 'eq2' (equation 2)

Model Formula: Depressed ~ Drug\_user + GENDER + MARTL + INCOME

Instruments: ~Marij\_price + Cocaine\_price + Heroin\_price + Meth\_price + GENDER + MARTL + INCOME

|             | Estimate   | Std. Error | t value  | Pr(> t )              |     |
|-------------|------------|------------|----------|-----------------------|-----|
| (Intercept) | 0.1635764  | 0.0244160  | 6.69955  | 0.00000000023441      | *** |
| Drug_user   | 0.2154265  | 0.0197290  | 10.91930 | < 0.00000000000000222 | *** |
| GENDER      | 0.0763894  | 0.0120061  | 6.36256  | 0.000000000217873     | *** |
| MARTL       | -0.0555825 | 0.0124303  | -4.47153 | 0.000007954601208     | *** |
| INCOME      | -0.1214112 | 0.0145234  | -8.35970 | < 0.00000000000000222 | *** |

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.398709 on 4567 degrees of freedom

Number of observations: 4572 Degrees of Freedom: 4567

SSR: 726.009211 MSE: 0.158969 Root MSE: 0.398709

Multiple R-Squared: 0.039799 Adjusted R-Squared: 0.038958

# Appendix 25 - Simultaneous Equations Model (3SLS)

```
DrugOnDep3SLS <- systemfit(DrugOnDep.eqs, method="3SLS",
 inst=DrugOnDep.ivs, data=DrugOnDepFinal)
summary(DrugOnDep3SLS)
```

systemfit results  
method: 3SLS

|        | N    | DF   | SSR     | detRCov  | OLS-R2   | McElroy-R2 |
|--------|------|------|---------|----------|----------|------------|
| system | 9144 | 9133 | 1111.94 | 0.011786 | 0.292877 | 0.466763   |

|     | N    | DF   | SSR     | MSE      | RMSE     | R2       | Adj R2   |
|-----|------|------|---------|----------|----------|----------|----------|
| eq1 | 4572 | 4566 | 386.510 | 0.084650 | 0.290946 | 0.526559 | 0.526040 |
| eq2 | 4572 | 4567 | 725.431 | 0.158842 | 0.398550 | 0.040564 | 0.039724 |

The covariance matrix of the residuals used for estimation

|     | eq1        | eq2        |
|-----|------------|------------|
| eq1 | 0.0845907  | -0.0408551 |
| eq2 | -0.0408551 | 0.1589685  |

The covariance matrix of the residuals

|     | eq1        | eq2        |
|-----|------------|------------|
| eq1 | 0.0846496  | -0.0407382 |
| eq2 | -0.0407382 | 0.1588419  |

The correlations of the residuals

|     | eq1       | eq2       |
|-----|-----------|-----------|
| eq1 | 1.000000  | -0.351323 |
| eq2 | -0.351323 | 1.000000  |

3SLS estimates for 'eq1' (equation 1)  
Model Formula: Drug\_user ~ Depressed + Marij\_price + Cocaine\_price + Heroin\_price +  
Meth\_price  
Instruments: ~Marij\_price + Cocaine\_price + Heroin\_price + Meth\_price + GENDER +  
MARTL + INCOME

|                | Estimate                                       | Std. Error  | t value                         | Pr(> t )     |
|----------------|------------------------------------------------|-------------|---------------------------------|--------------|
| (Intercept)    | -0.030873815                                   | 0.010580021 | -2.91812                        | 0.0035386 ** |
| Depressed      | 0.150601549                                    | 0.056618351 | 2.65994                         | 0.0078427 ** |
| Marij_price    | 0.060552560                                    | 0.001380933 | 43.84901 < 0.000000000000000222 | ***          |
| Cocaine_price  | 0.004232901                                    | 0.000184733 | 22.91362 < 0.000000000000000222 | ***          |
| Heroin_price   | 0.003511245                                    | 0.000426965 | 8.22374 0.000000000000022204    | ***          |
| Meth_price     | 0.002752412                                    | 0.000535565 | 5.13927 0.0000028732584533842   | ***          |
| ---            |                                                |             |                                 |              |
| Signif. codes: | 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 |             |                                 |              |

Residual standard error: 0.290946 on 4566 degrees of freedom  
Number of observations: 4572 Degrees of Freedom: 4566  
SSR: 386.510264 MSE: 0.08465 Root MSE: 0.290946  
Multiple R-Squared: 0.526559 Adjusted R-Squared: 0.52604

## Appendix 26 - Simultaneous Equations Model (3SLS)

3SLS estimates for 'eq2' (equation 2)

Model Formula: Depressed ~ Drug\_user + GENDER + MARTL + INCOME

Instruments: ~Marij\_price + Cocaine\_price + Heroin\_price + Meth\_price + GENDER +  
MARTL + INCOME

|             | Estimate   | Std. Error | t value                          | Pr(> t )           |
|-------------|------------|------------|----------------------------------|--------------------|
| (Intercept) | 0.2055933  | 0.0230634  | 8.91428 < 0.0000000000000000222  | ***                |
| Drug_user   | 0.2080646  | 0.0196644  | 10.58079 < 0.0000000000000000222 | ***                |
| GENDER      | 0.0564730  | 0.0114106  | 4.94919                          | 0.000000772053 *** |
| MARTL       | -0.0672812 | 0.0117685  | -5.71705                         | 0.000000011529 *** |
| INCOME      | -0.1272117 | 0.0140355  | -9.06356 < 0.0000000000000000222 | ***                |

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.39855 on 4567 degrees of freedom

Number of observations: 4572 Degrees of Freedom: 4567

SSR: 725.430863 MSE: 0.158842 Root MSE: 0.39855

Multiple R-Squared: 0.040564 Adjusted R-Squared: 0.039724

## Appendix 27 - Hausman specification test

Hausman specification test

```
{r}
h <- hausman.systemfit(DrugOnDep2SLS, DrugOnDep3SLS)
print(h)
```

Hausman specification test for consistency of the 3SLS estimation

```
data: DrugOnDepFinal
Hausman = 36.715, df = 11, p-value = 0.0001286
```

---

X and Y (Depression and Drug user) are dependent of each other. Each variable can influence the other. This model helps correct for simultaneity

Based on the 2SLS model, the price of the 4 different drugs(marijuana, heroin, cocaine and meth) are very significant in the prediction of being a drug user.

In comparison, gender, marital status and income are also significant factor in the influence of depression.

Hausman specification test tells me to reject the 3SLS because the 2SLS is better.

## Appendix 28 – ODDS RATIO

```
exp(coef(logit.reg))
```

|        | (Intercept) | Depressed | GENDER    | Race      | EDUC      | MARTL     |
|--------|-------------|-----------|-----------|-----------|-----------|-----------|
|        | 0.8962454   | 1.9308043 | 0.5171146 | 1.2943755 | 1.0183733 | 0.4350136 |
| POVRAT |             |           |           |           |           |           |
|        | 1.0085160   |           |           |           |           |           |

```
exp(coef(logit.reg2))
```

|  | (Intercept) | Depressed | GENDER    | MARTL     | Race      |
|--|-------------|-----------|-----------|-----------|-----------|
|  | 0.9148597   | 1.9225359 | 0.5182562 | 0.4390346 | 1.2956963 |

## References

Depression. (n.d.). National Institute of Mental Health (NIMH).  
<https://www.nimh.nih.gov/health/topics/depression>

Major Depression. (n.d.). National Institute of Mental Health (NIMH).  
<https://www.nimh.nih.gov/health/statistics/major-depression>

Villarroel, M., & Terlizzi, E. P. (2020). Symptoms of Depression Among Adults: United States, 2019. NCHS Data Brief, 379, 1–8. <https://pubmed.ncbi.nlm.nih.gov/33054920/>

Bustamante, J. (2023, January 1). Drug Abuse Statistics. NCDAS. <https://drugabusestatistics.org/>

NHANES - About the National Health and Nutrition Examination Survey. (n.d.).  
[https://www.cdc.gov/nchs/nhanes/about\\_nhanes.htm](https://www.cdc.gov/nchs/nhanes/about_nhanes.htm)

Income inequality and depression: a systematic review and meta-analysis of the association and a scoping review of mechanisms. (n.d.).  
<https://onlinelibrary.wiley.com/doi/full/10.1002/wps.20492>