

Does Substance Abuse Lead to Higher Sexual Activity During One's Lifetime?

Abstract

Background

Due to earlier and longer exposure to STD leading to higher likelihood of lifetime complications, previous public health studies¹ focusing on adolescents have found positive associations between substance use and sexual risk activity. This association is a result of complex intertwined factors that are not only limited to direct behavioral change from drug use, but also socioeconomic background and psychological factors that may encourage drug misuse and subsequent feedback loop of bidirectional causality. In this paper, we hope to expand upon this literature by seeing how this relationship may hold across all ages in general.

Method

The dataset comes from the R package NHANES data from the R package "NHANES" which are resampled 10,000 individuals from the original 2009-2010 and 2011-2012 NHANES cross-sectional survey data. Using preliminary analysis with two simple linear models, we initially investigate the relationship between substance use and first sexual activity, and substance use and the number of sexual partners. After this, we develop a more complex linear model to incorporate both first sexual activity and number of sexual partners in the outcome variable. The outcome variable is log transformed average frequency of sexual activity during the duration between the age of first sexual activity and current age. We sequentially adjust with blocks of socioeconomic factors, demographics, and potential confounding variables to see if any of the association can be explained away by other factors. We also removed observations which had missing values and possible outliers, and examined any systematic differences between the observations with missing values to see if there might be any bias introduced into the results.

Results

Among 4897 observations, those with self-reported substance use history have reported 1-2 years earlier sexual activity on average. On the other hand, among 1626 observations, on average, the average number of sexual activities during one's lifetime since first reported sexual activity is higher by a statistically significant percentage for those with substance use experience. Accounting for the interaction covariate, those with regular marijuana use controlling for hard drug use on average have had 46.23 percent more average sexual activity and those with hard drug use controlling for regular marijuana use have had on average have had 66.3 percent more average sexual activity.

Conclusion

While those with self-reported substance use history have earlier self-reported first sexual activity on average, this does not necessarily imply that they are likely to have engaged in more sexual activity during their lifetimes. Further analysis incorporating information of total number of sexual partners during one's lifetime, controlling for socioeconomic factors, demographics, and potential confounding variables shows that average number of sexual activities during one's life

¹ https://www.cdc.gov/healthyyouth/factsheets/substance_use_fact_sheet-detailed.htm

lifetime since first reported sexual activity is higher for those with reported history of either hard drug use or regular marijuana use. This supports the argument that substance misuse is likely to lead to higher sexual activity during one's lifetime.

Introduction

According to the Surgeon General's report published in 2016, the misuse of substances including alcohol and drugs is a growing problem and has contributed to the decrease in life expectancy in the United States. One of the ways in which this finding could occur is that those who misuse substances are more likely to have engaged in sexual risk behavior. To help develop effective intervention policies, previous studies have primarily focused on teenage and young adults whose substance misuse is associated with sexual risk behavior which leads to a higher risk of HIV, STDs, and early pregnancy.

Substance misuse itself is not the sole contributor to sexual risk behavior. Factors such as poverty, family history, parenting, association with peer groups associated with substance misuse, isolation in an education may cause predisposition to vulnerability to substance misuse and subsequent sexual risk behavior. This is, thus, a complex problem and it is important to approach this problem not only at an individual level but from the bigger picture of social context as well.

Given this, we wanted to expand upon this literature and see what the association between substance misuse and sexual activity would be across all age groups. We believed that while it might be important to identify the factors that influence at-risk groups to be exposed to substance use and sexual risk behavior to help develop early intervention for adolescents and young adults, it is also important to see the individual's behavior during their lifetime to see the risk of further transmission of HIV and STDs by analyzing the frequency of sexual activity and their contribution to public health risks.

Methods

Study Population

The NHANES dataset from the "NHANES" R package is complex survey data from a survey sample that aims to represent "the non-institutionalized civilian resident population of the United States." It was collected during the 2009-10 and 2011-2012 series of NHANES surveys. Approximately 20,293 individuals of all ages were interviewed in their homes and 10,000 individuals were resampled for this dataset to account for oversampling certain demographics. For the purposes of this study, we treat the dataset as if from a simple random sampling, or equal probability of selection for each observation.

According to the National Survey of Family Growth survey for 2015-2019, across both genders, the median sexual partners in a lifetime for adults aged 25-49 is 5.3.² The median for the number of sexual partners for our dataset without missing data was approximately similar despite the wider age range and we initially thought of removing any observations with self-reported sexual partners above the third quartile of approximately 23. According to the UN 2022 World Population Prospects, however, the US birth per woman was approximately 0.24 percentage points higher in 2009 compared to 2015-2019³ and our study population included older groups who were likely to

² https://www.cdc.gov/nchs/nsfg/key_statistics/n-keystat.htm

³ <https://www.bls.gov/opub/mlr/2023/article/labor-force-and-macroeconomic-projections.htm>

have had more sexual partners during their lifetime; we also wanted to account for the possibility of genuine datapoints. Therefore, we doubled the cutoff to a more generous 40 sexual partners. 318 observations deemed outliers from either disingenuous responses and possible typos were removed. Once the final covariates were selected after a sequential extra sum of squares test, and observations with missing data for the selected covariates were removed, we were left with 1639 observations.

Some observations were deemed possible outliers based on factors such as deffits, Cook's distance, and leverage. Based on case-by-case observation analysis, however, we noticed only one observation being a possible outlier. Fitting new model without the possible outlier did not noticeably the change coefficients or their statistical significance, however, and we did not remove the observation.

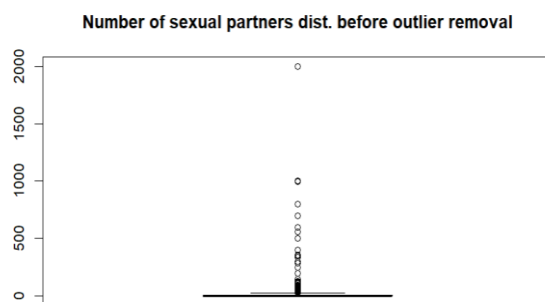


Figure 1

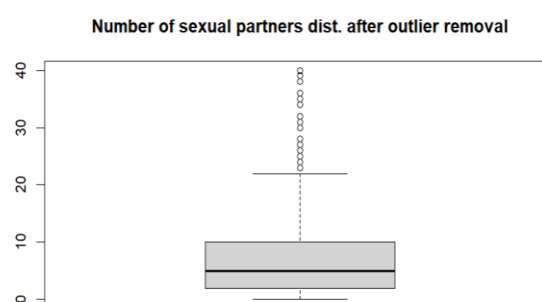


Figure 2

Variables

Predictors of Interest

Based on literature search, the initial predictors of interest were divided into four categories: substance use, socioeconomic status and demographics, lifestyle, and same sexual partner experience. Our main focus was in the block substance use which includes self-reported current regular smoking (SmokeNow, Yes/No), estimated number of days of alcohol drinking over the last year (AlcoholYear), history of regular marijuana of at least once in a year (RegularMarij, Yes/No), experience of using cocaine, crack cocaine, heroin, or methamphetamine (HardDrugs, Yes/No). Among these, our interest primarily lies in regular marijuana use and hard drug use as during preliminary analysis using simple regression and controlling for potential confounding variables. Marijuana is a frequent polysubstance use candidate with hard drugs, which have serious health consequences due to high likelihood strong physical dependence. Regular marijuana use and hard drug use experience were also the predictors that held the largest magnitude in effect on the outcome variable and maintained extremely strong statistical significance. We also noticed strong interaction between the two covariates. Given the literature with moderate evidence of association, albeit with differing views on the cause such as marijuana serving as gateway drugs, or potentially enhancing the hard drug use experience,⁴⁵ we decided to include this sole interaction in the model.

As mentioned in previous studies, drug use and sexual risk behavior is relevant to social context for adolescents and young adults. After including older people in the dataset, we wished to if any of the explanatory power of regular marijuana use and hard drug use on the outcome variable

⁴ [Cannabis Use and the Abuse of Other Substances - The Health Effects of Cannabis and Cannabinoids - NCBI Bookshelf](#)

⁵ https://link.springer.com/chapter/10.1007/978-1-4939-2294-9_16

may be nullified once social economic status and demographics (Age, Gender[M/F], HHIncome, Education), lifestyle and health (BMI, Depressed[None/Several/Majority/AlmostAll], LittleInterest[None/Several/Majority/AlmostAll], PhysActive[Y/N]) were controlled for. Same gender sexual activity experience (SameSex[Y/N]) was also taken into account considering the literature on sexual minorities being more vulnerable to polysubstance misuse.⁶

Table 1. Initial exploratory linear regression between sexual activity and substance misuse

SexAge MLR				SexNumPartnLife MLR			
Characteristic	Beta	95% CI [†]	p-value	Characteristic	Beta	95% CI [†]	p-value
(Intercept)	18	18, 18	<0.001	(Intercept)	8.4	6.3, 10	<0.001
RegularMarij	-2.2	-2.5, -1.9	<0.001	RegularMarij	15	9.8, 20	<0.001
HardDrugs	-1.7	-2.1, -1.3	<0.001	HardDrugs	14	6.5, 21	<0.001
RegularMarij * HardDrugs				RegularMarij * HardDrugs			
Yes * Yes	1.4	0.90, 2.0	<0.001	Yes * Yes	0.82	-8.7, 10	0.9
[†] CI = Confidence Interval				[†] CI = Confidence Interval			

Outcome Variable

To see a more comprehensive picture of sexual activity during one's lifetime given drug use experience controlling for other variables, we initially selected SexAge, the self-reported first age at which sexual activity occurred, and SexNumPartnLife, the total number of sexual partners during the respondent's lifetime. We realized, however, that SexNumPartnLife is discrete count data which does not conform to the assumption of a normally distributed outcome variable for the linear regression model used in this course. This would potentially lead to undesirable outcomes such as biased estimates. Therefore, we created a new continuous variable where we took SexNumPartnLife and divided by the difference between Age and SexAge, where Age >= SexAge, to indicate duration from which first sexual activity occurred and divided it by. The new variable would give the average number of sexual activity frequency since first sexual activity occurred. Since the new variable was heavily skewed to the right, we applied a log transformation. Since some people had 0 sexual partners in life, the log transformation produced some observations with negative infinity. We corrected them as 0. Some people also reported the same Age and SexAge, resulting in 0 for the denominator. We corrected the denominator as 1, essentially giving them an annual average number of sexual partners.

$$AvgSexFreq = \log \left(\frac{SexNumPartnLife}{Age - SexAge} \right)$$

⁶ [Trends in Binge Drinking, Marijuana, Illicit Drug, and Polysubstance Use by Sexual Identity in the United States \(2006–2017\) - PMC](#)

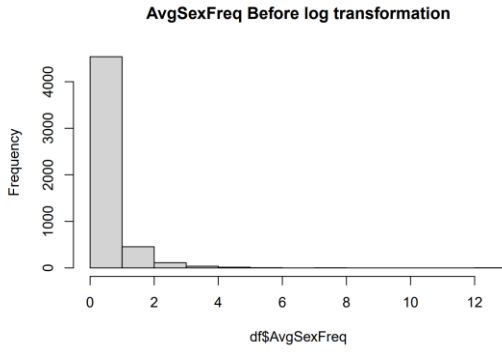


Figure 3

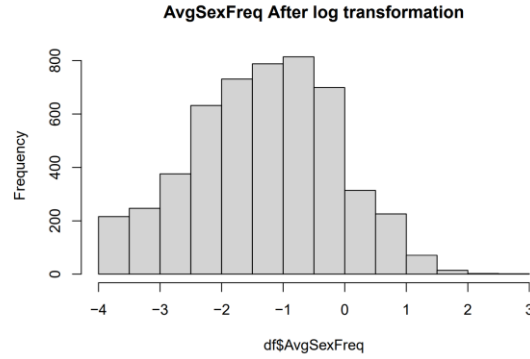


Figure 4

Covariates

Dividing the predictors by blocks, we ran a sequential extra sums of squares test in the order we deemed most relevant to our analysis to see if any of the socioeconomic status and demographics, lifestyle, and same sexual partner experience blocks of covariates added statistically significant explanatory power to the model. We found that the lifestyle block (β_{health}) failed to reject the null hypothesis at a significance level of $\alpha = 0.001$. We used a much stricter statistical significance level since the number of missing values from choosing so many covariates removed too many observations and left only 30, which would potentially result in higher variance in the estimated coefficients of the model. We also noted their individual statistical significance in the initial model and the small magnitude of their effect on the outcome variable.

- (i) $\beta_{\text{substance}} = (\beta_{\text{SmokeNow}}, \beta_{\text{AlcoholYear}}, \beta_{\text{RegularMarij}}, \beta_{\text{HardDrugs}}, \beta_{\text{RegularMarij*HardDrugs}})^T$
- (ii) $\beta_{\text{Demo}} = (\beta_{\text{Age}}, \beta_{\text{Gender}}, \beta_{\text{HHIncome}}, \beta_{\text{Education}})^T$
- (iii) $\beta_{\text{Health}} = (\beta_{\text{BMI}}, \beta_{\text{DiabetesAges}}, \beta_{\text{Depressed}}, \beta_{\text{LittleInterest}}, \beta_{\text{PhysActive}})^T$
- (iv) $\beta_{\text{SameSex}} = (\beta_{\text{SameSex}})^T$

Table 2. Sequential Extra Sums of Squares Test

Step	Tested Var.	SS(Num.)	SS(Denom.)	Test Stat.	Dist.	p-value	Decision	Stopping Rule	Decision
I	$\beta_{\text{Substance}}$	13.88444	26.9329	5.155204576	$F_{5,14}$	0.001262146	Reject	Do not stop	Collect
II	β_{Demo}	55.61473	26.9329	25.81174	$F_{4,14}$	6.872507e-10	Reject	Do not stop	Collect
III	β_{Health}	5.687399	26.9329	2.11169493	$F_{5,14}$	0.08788892	Fail to Reject	Stop	Not Collect
IV	β_{SameSex}	0.001708498	26.9329	10.55847467	$F_{1,14}$	0.00260712	NA	NA	NA

The final model is thus,

$$\text{AvgSexFreq} = X_{\text{Substance}}\beta_{\text{Substance}} + X_{\text{Demo}}\beta_{\text{Demo}} + \epsilon, \epsilon \sim N(0, \sigma^2 I)$$

After a multiple linear regression, we noticed that for the categorical covariate household income, the "5000-9999" and "10000-14999" income ranges had p-values 0.08 and 0.04, respectively. Therefore, with the null hypothesis that the estimated coefficients for the two categories were the same and the estimated coefficients for the rest of the income categories were the same, we performed a general linear hypothesis testing and failed to reject the null hypothesis at a significance level 0.01. Thus, we collapsed the two income ranges and the rest into two categories.

Statistical analysis

Table 3

AvgSexFreq MLR				Education			
Characteristic	Beta	95% CI [†]	p-value				
(Intercept)	0.60	0.36, 0.85	<0.001	8th Grade	—	—	
SmokeNow	0.15	0.08, 0.23	<0.001	9 - 11th Grade	0.01	-0.19, 0.22	0.9
AlcoholYear	0.00	0.00, 0.00	<0.001	High School	-0.07	-0.27, 0.13	0.5
RegularMarij	0.38	0.29, 0.48	<0.001	Some College	-0.08	-0.28, 0.12	0.4
HardDrugs	0.51	0.38, 0.64	<0.001	College Grad	-0.01	-0.22, 0.19	>0.9
Age	-0.05	-0.05, -0.05	<0.001	RegularMarij * HardDrugs			
Gender	0.19	0.11, 0.26	<0.001	Yes * Yes	-0.32	-0.49, -0.15	<0.001
segmentincome	0.24	0.09, 0.40	0.002	[†] CI = Confidence Interval			

Comparison of basic demographics and social economic status using a binomial GLM with missing values of the chosen covariates as an outcome variable yielded statistically significant differences for Age, Gender, Education below high school, and marital status at a statistical significance level 0.05. With limited information regarding the survey, we know that there was no data collected for our covariates for those aged below 18, but it is not clear why there are missing values for those above 60. Overall, those that were older, were male, had education level at least above 8th grade, were divorced or living with a partner had a higher response rate for our covariates. Thus, the missingness does not seem to be at completely at random. We believe there might be a positive bias for these observations as it seems to be that there is less pressure to conform to social desirability for these respondents to report their use of cigarettes, alcohol, marijuana use, hard drug use.

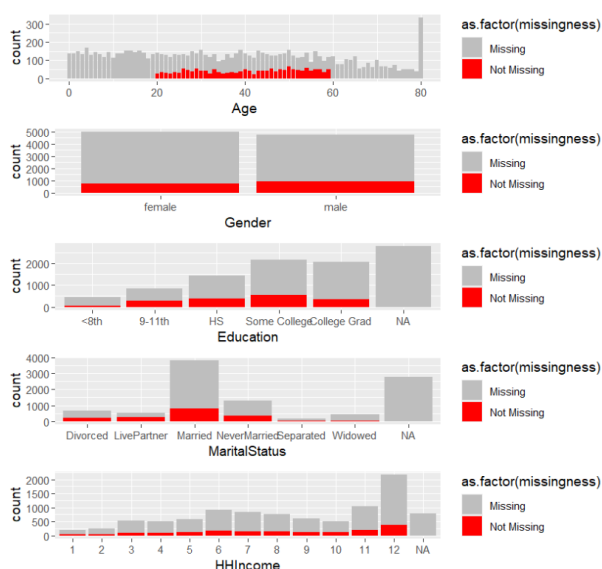


Figure 5. Histogram for basic demographics differences

A QQ-Plot with residuals for the model suggests that normality of residuals assumption holds. Residual plots showed no signs of violation of homoscedasticity. Partial regression plots with

the covariates displayed no evidence of linearity assumption violation. Unfortunately, a Durbin Watson test statistic of 1.13 suggested the presence of autocorrelation which would lead to biased coefficient estimates and incorrect variance. For the purposes of this course, we assumed independence of errors. A VIF test showed no signs of multicollinearity.

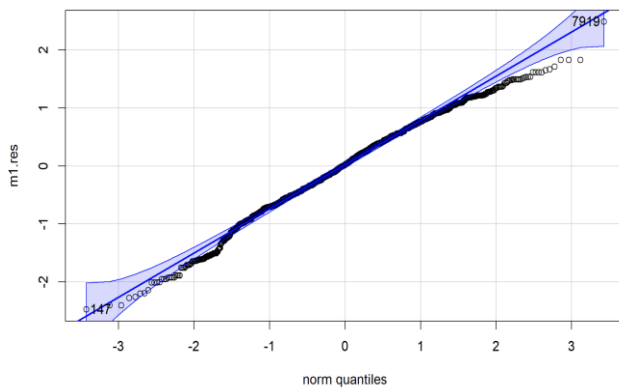


Figure 6. QQ Plot

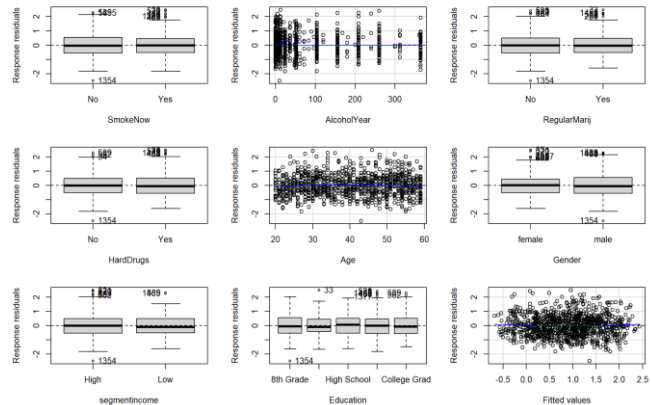


Figure 7. Residuals vs. Covariates

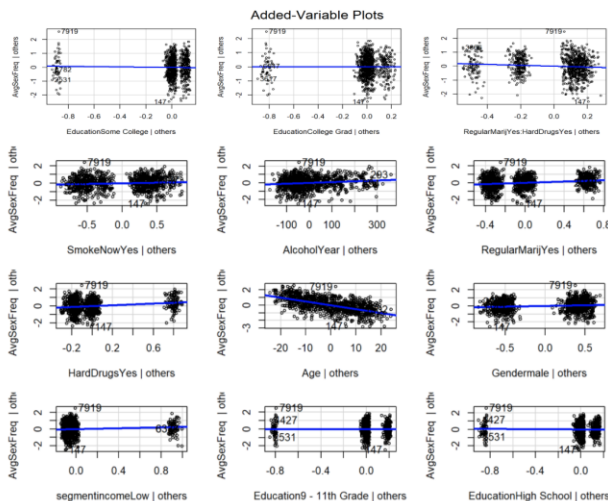


Figure 8. Partial Regression Plots

	GVIF <dbl>	Df <dbl>	GVIF^(1/(2*Df)) <dbl>
SmokeNow	1.129942	1	1.062987
AlcoholYear	1.103500	1	1.050476
RegularMarij	1.024763	1	1.012306
Age	1.079130	1	1.038812
Gender	1.035586	1	1.017638
segmentincome	1.025706	1	1.012772
Education	1.195547	4	1.022577

Results

As our model is a log transformed model, the coefficients reflect the percentage change in the outcome variable for a unit change in the covariate. For some generalized estimated coefficient β_1 , the interpretation for the outcome variable change for one unit change in x_1 is as such where the outcome variable changes in percentages:

$$\frac{f(x+1)}{f(x)} - 1 = (e^{\beta_1} - 1) * 100$$

Our main covariates of interest are regular marijuana use and hard drug use experience, and we must take into account interaction effects when interpreting the coefficients.

Accounting for the interaction covariate, those with regular marijuana use controlling for hard drug use have had 46.23 percent more average sexual activity on average, and those with hard

drug use controlling for regular marijuana use have had on average have had 66.3 percent more average sexual activity on average. For those that had regular marijuana use and hard use experience, they have had 37.24 percent less average sexual activity on average. These effects were statistically significant at a significance level 0.01.

Conclusions and Discussion

While those with self-reported substance use history have earlier self-reported first sexual activity on average, a head start does not necessarily imply that they are likely to have engaged in more sexual activity during their lifetimes. Upon further in-depth analysis, controlling for socioeconomic factors, demographics, and potential confounding variables, our analysis shows that the average number of sexual activities during one's life lifetime since first reported sexual activity is higher for those with reported history of either hard drug use or regular marijuana use. On the other hand, however, there is a negative interaction effect from having a history of hard drug use and regular marijuana use. Perhaps this reflects the fact that those that engage in polysubstance misuse are negatively afflicted to the point that sexual activity is hindered.

Due to the limited information given by survey questions asking respondents whether they have had experience using hard drugs, it is impossible to tell whether they only tried hard drugs once out of curiosity or are hard drug addicts. Subsequently, it is also hard to tell in further detail how having hard drugs may positively affect average sexual activity. The initial model with health indicators relating to weight, physical activity, and psychological state did not add much explanatory power to the model, thus it does not seem to be the case that hard drugs may have noticeably changed the respondent's physical or psychological state based on the given information. Further information with distribution of sexual activity during one's lifetime will be helpful, but this kind of information may be hard to recall and potentially a breach of privacy that would be hard to conduct on a national scale. Another useful piece of information might be asking the respondent's overall mood regarding positive state of mind in addition to depression and lack of interest in things.

There are other limitations to this study. Since this is complex survey data, we would need to use survey weights to properly reflect the probability of selection for each observation and draw correct inference for the target population with weighted regression. We have also removed observations with missing data and outliers. This has resulted in 8,631 observations being deleted. While we were able to see some statistically significant differences between the observations with missing data and observations, and were able to analyze to make some inferences as to whether the missingness was completely at random, we could not salvage as much data using imputation as we lacked the means to properly implement it. From the point of a survey statistician, throwing away more than half of your sample would be simply disastrous not only because the studies are quite costly to carry out, but also because you would be throwing away observations with the weight of at least thousands of people. Any systematic differences between missing and non-missing data due to factors such as social desirability would lead to response bias and may drastically change the results of this analysis once accounted for.

Further study would ideally include self-reported sexual risk behavior and STD diagnosis from survey questions or actual diagnosis of HIV or STD presence in the respondent to get directly relevant information for public health intervention policy. The former is likely to be subject to interviewer effects which reduce the report of having STDs and the latter is likely to be a costly effort. Additionally, it would be better to have more specific data such as the number of days of substance misuse during the past year for both hard drug use and marijuana use. This is also likely to be subject to underreporting due to interviewer effects.

Bibliography

Centers for Disease Control and Prevention. (2018, July 11). *Substance Use and Sexual Risk Behaviors Among Youth*. Centers for Disease Control and Prevention.

https://www.cdc.gov/healthyyouth/factsheets/substance_use_fact_sheet-detailed.htm

Centers for Disease Control and Prevention. (2021, November 8). *NSFG - listing N - key statistics from the National Survey of Family Growth*. Centers for Disease Control and Prevention.

https://www.cdc.gov/nchs/nsfg/key_statistics/n-keystat.htm

Dubina, K. (2023). Labor force and macroeconomic projections overview and highlights, 2022–32. *Monthly Labor Review*. <https://doi.org/10.21916/mlr.2023.21>

Keane, H. (2018). *Facing addiction in America: The surgeon general's report on alcohol, Drugs, and Health*. U.S. Department of Health and Human Services, office of the surgeon general Washington, DC, USA: U.S. Department of Health and Human Services, 2016 382 pp. online (grey literature):

<https://addiction.surgeongeneral.gov/>. Drug and Alcohol Review, 37(2), 282–283.

<https://doi.org/10.1111/dar.12578>

McCabe, S. E., Engstrom, C. W., Kcomt, L., Evans-Polce, R., & West, B. T. (2022). Trends in binge drinking, marijuana use, illicit drug use, and polysubstance use by sexual identity in the United States (2006–2017). *Substance Abuse*, 43(1), 194–203.

<https://doi.org/10.1080/08897077.2021.1913696>

National Academies of Sciences, Engineering, and Medicine. 2017. *The health effects of cannabis and cannabinoids: The current state of evidence and recommendations for research*. Washington, DC: The National Academies Press. doi: 10.17226/24625.

Tambaro, S., & Bortolato, M. (2015). Interactions of cannabis and amphetamine-type stimulants. *Cannabinoid Modulation of Emotion, Memory, and Motivation*, 409–442.

https://doi.org/10.1007/978-1-4939-2294-9_16

Group Member Contributions

Jaehoon Kim is the de facto sole contributor to this project. Besides presentation and some meetings, the group members' contribution is solely within a separate final report for which I was only responsible for the abstract part. The instructors have been notified with documentation regarding work and communication done since October 29th.