DUKE FUQUA SCHOOL OF BUSINESS

Empirical Economic Analysis —— Drug Use And Mental Health Issues

Sec. C Team 62 – Samir Kaddoura, Jiakai Liu, Xinyi Zhou, Maggie Guo, Tapan Agarwal

2) Project Motivation

- There's a clear indication that the drug use—including alcohol, cigarettes, and marijuana—has profound and lasting impacts on individuals' mental health and overall well-being. These impacts manifest across a spectrum of outcomes, from the heightened risk of developing substance use disorders (SUD) to significant impairments in social competence, academic performance, employment prospects, and even increased susceptibility to psychiatric disorders. The interplay between drug use and these adverse outcomes underscores not only a correlation but, a directional pathway that may lead from substance use to complex health and social challenges.

- **We are interested in re-investigate if drug use directly leads to mental health disorders or if both share common risk factors, thus aiming to offer insights for targeted prevention and intervention.** For instance, providing more effective prevention and treatment programs for those at highest risk, guiding policymakers, educators, and healthcare providers on the necessity of intervention and comprehensive strategies, and encouraging a well-informed and proactive community approach.

3) Literature Review

The Impact of Drug Use on Mental Health:

In response to the insights garnered from traditional analytical approaches, our research aims to **reevaluate and substantiate the connection between drug use and mental health through innovative econometric models.** These models promise to refine our understanding by accounting for a wider array of economic and social variables, thereby offering a more nuanced perspective on the interplay between substance use and mental health trajectories. Meanwhile, the **onset age of drug use** based on the previous studies will be consider as an important influencing explanatory variable in our study.

Supporting Literatures:

- The intricate relationship between drug use, including tobacco, alcohol, and cannabis, and its impact on mental health remains a pivotal area of investigation within psychiatric research.
- Historically, the examination of this relationship has been pursued through various methodological lenses, with substantial reliance on Logistic Regression Models, observational methods, and standardized psychiatric symptom evaluations to assert the influence of substance use on mental health outcomes(York et al,2004 & Gove, Geerken, Huges, 2024 & Sullivan et al, 2006).
- It becomes evident that the age of initial substance use is a critical factor. In the 1992 National Longitudinal Alcohol Epidemiologic Survey by Grant and Dawson (1997), it underscore the association between the age of first alcoholic drink and the probability of lifetime alcohol abuse or mental dependence. The initiation of drinking at an early age correlates with a spectrum of problematic outcomes including mental disorder in later adolescence and young adulthood.

4) Data Source

About our dataset NSDUH:

NSDUH is an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years or older. It is the primary source of statistical information on the use of tobacco, alcohol, prescription psychotherapeutic drugs (pain relievers, tranquilizers, stimulants, and sedatives), and other substances (e.g., marijuana, cocaine) by people aged 12 or older in that population. The survey also includes extensive information on substance use disorders (SUDs), substance use treatment, mental health issues, and mental health treatment.

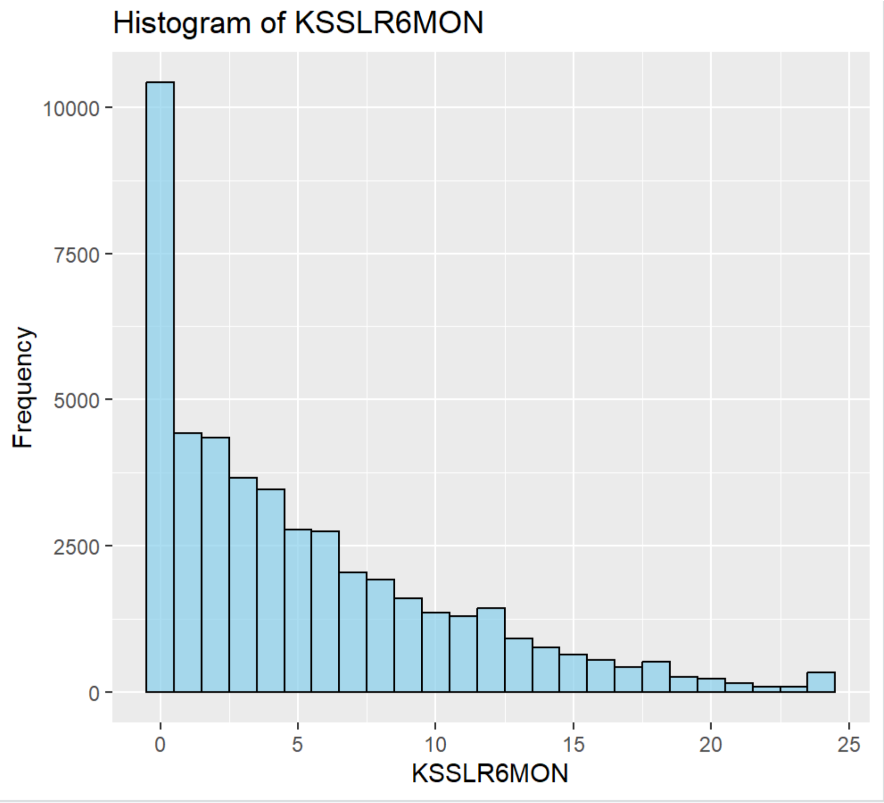
The whole dataset countians 2605 variables and 59069 observations. After data cleaning and feature processing, we selected and created 33 variables and 46585 observations from the original dataset for better explaining the impact of drug use on mental health. We have filtered out any individual under the age of 18, as there is no data on drug consumption for minors.

DUKE FUQUA SCHOOL OF BUSINESS

4) EDA

Outcome Variable: KSSLR6MON : total score in past month

```
   KSSLR6MON
Min.    : 0.000
1st Qu.: 1.000
Median : 4.000
Mean    : 5.185
3rd Qu.: 8.000
Max.    :24.000
```

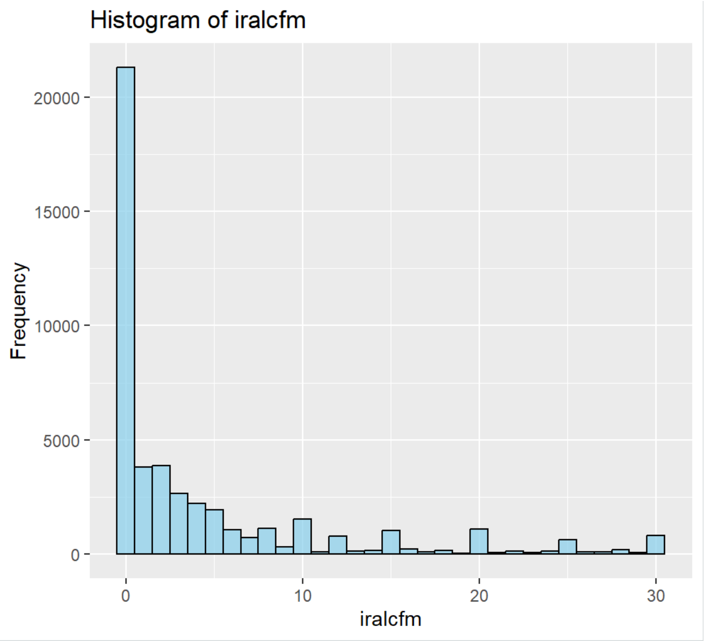


DUKE FUQUA SCHOOL OF BUSINESS

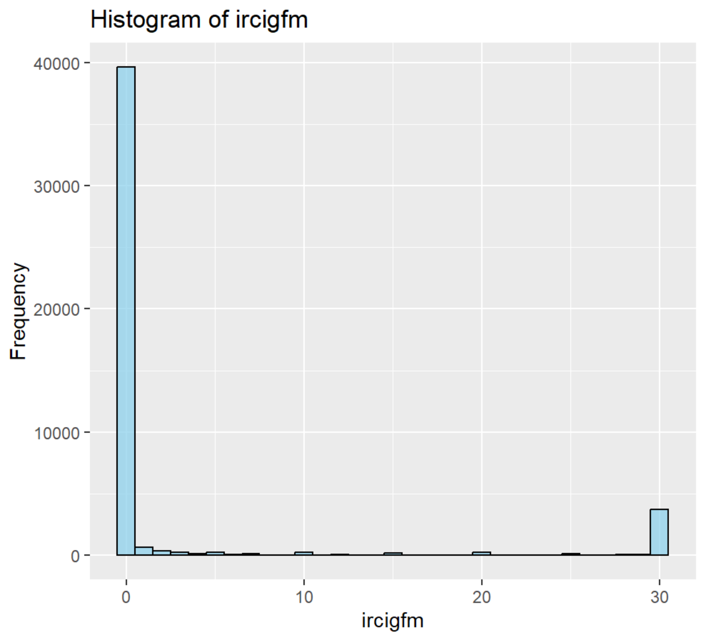
4) EDA

EDA: Explore the numerical features

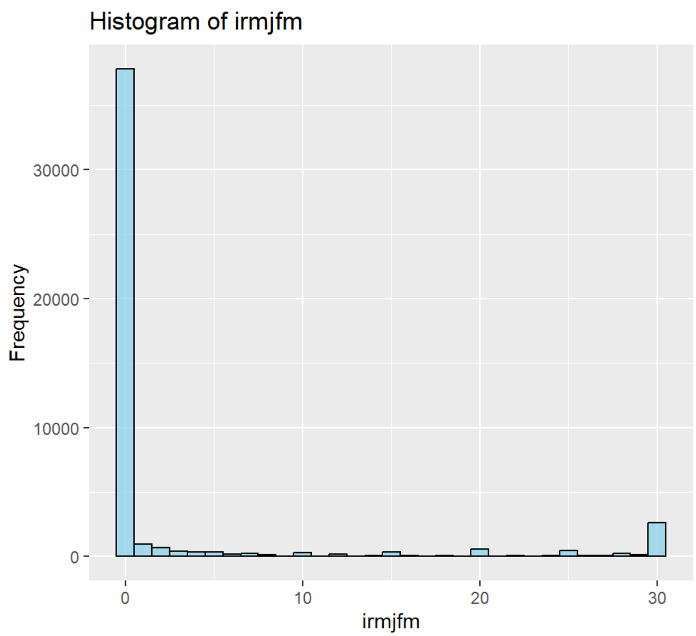
iralcfm: alcohol frequency past month



ircigfm: cig frequency past month



irmjfm: marijuana frequency past month



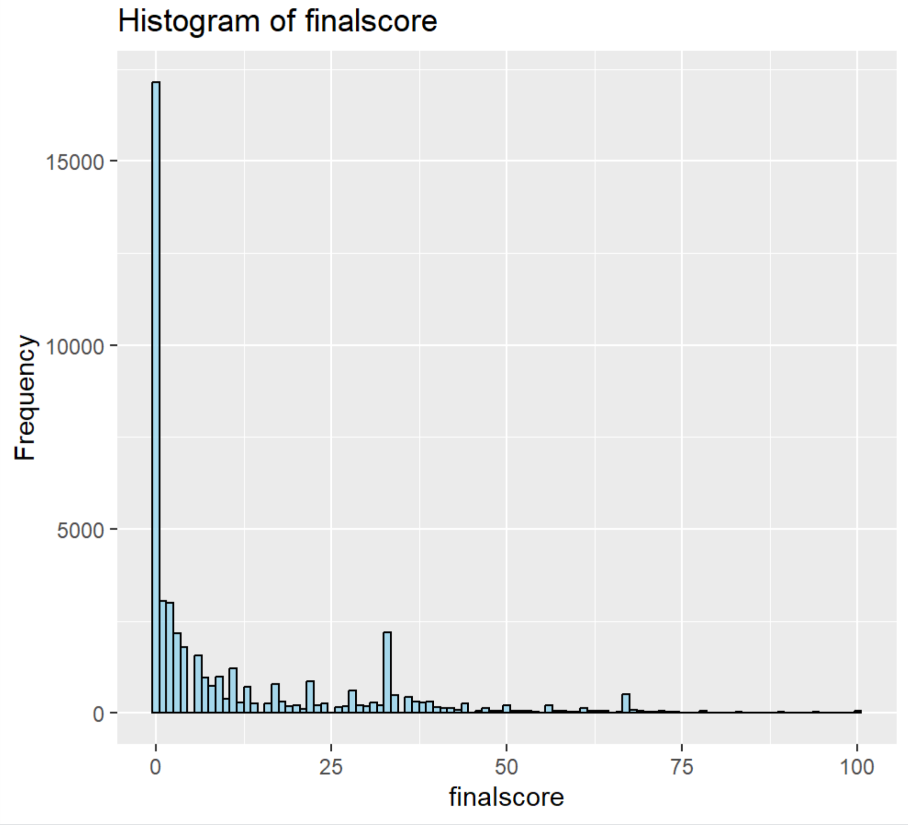
DUKE FUQUA SCHOOL OF BUSINESS

4) EDA

EDA: Explore the numerical features

finalscore: (iralcfm+ircigfm+irmjfm)/90 * 100

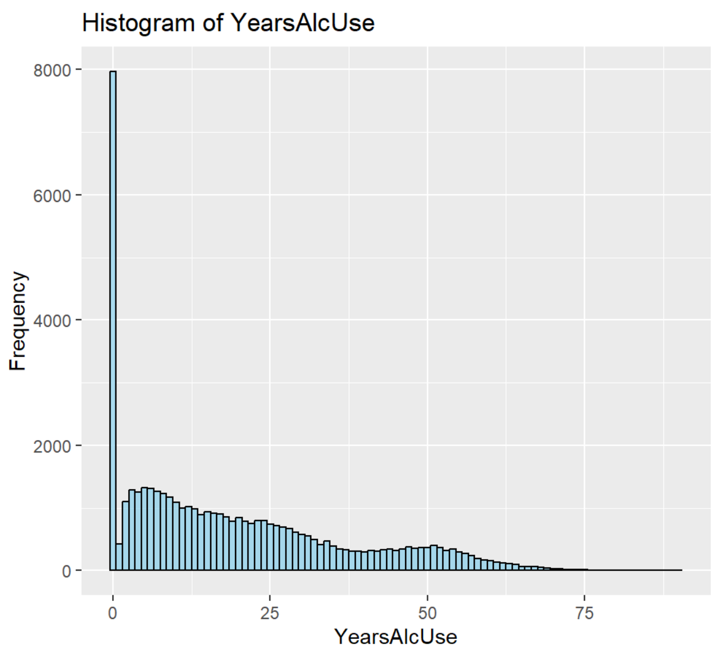
```
   finalscore
Min.   :  0.000
1st Qu.:  0.000
Median :  3.333
Mean   : 11.554
3rd Qu.: 16.667
Max.   :100.000
```



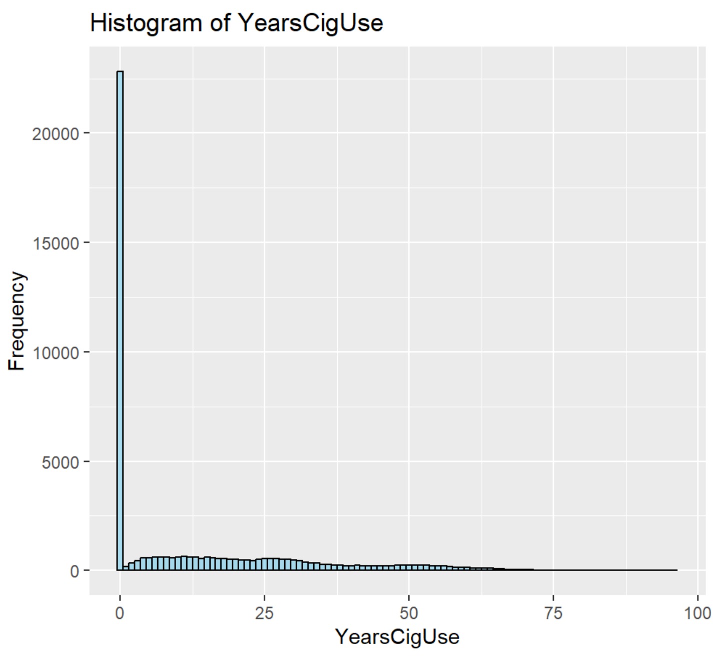
4) EDA

EDA: Explore the numerical features

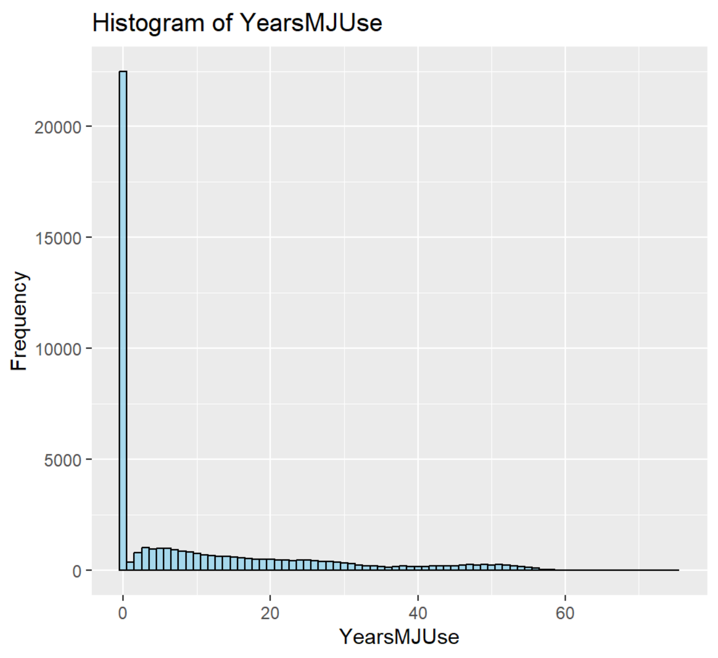
YearsAlcUse: How many years of use in Alcohole



YearsCigUse: How many years of use in Cigarette



YearsMJUse: How many years of use in Marijuana

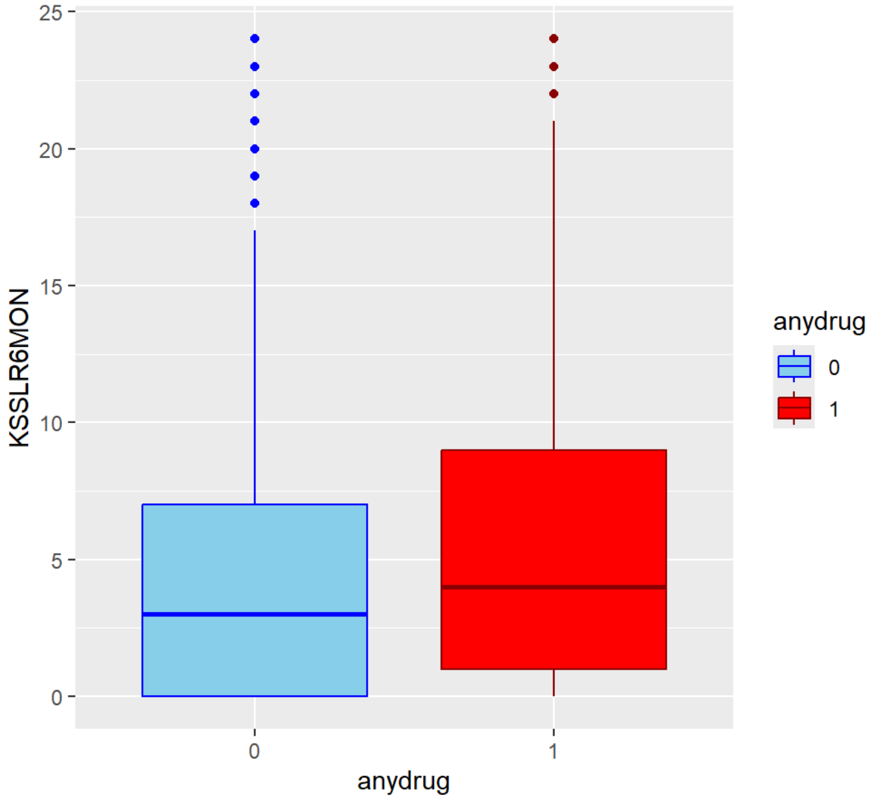


FUQUA
SCHOOL OF BUSINESS

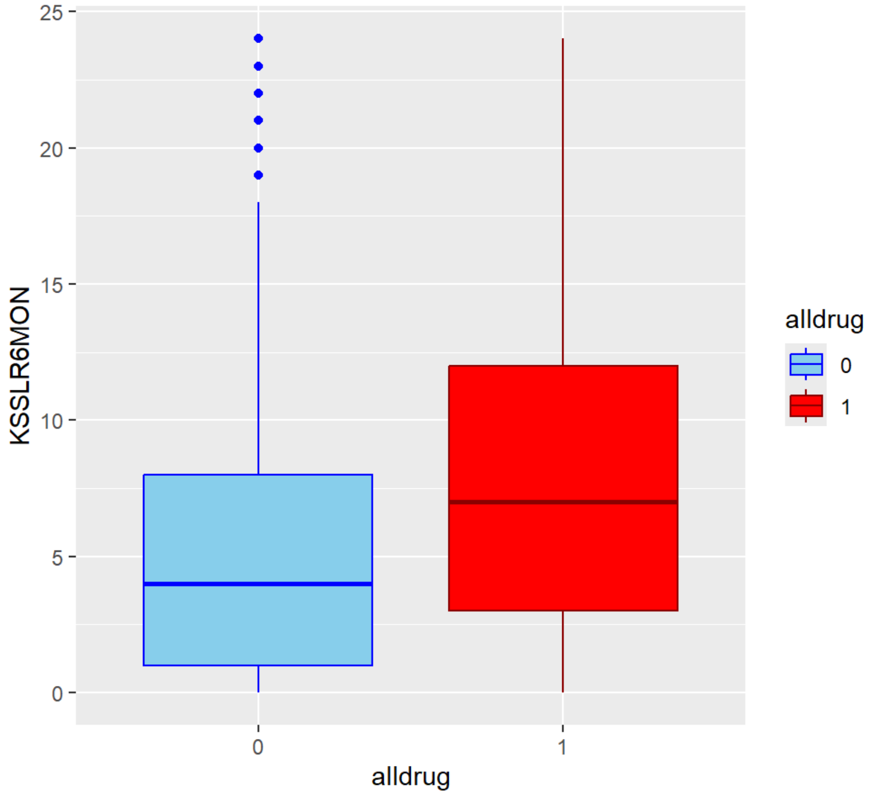
4) EDA

EDA: Explore the categorical features

anydrug: if the person had either alcohol, cigarette, or marijuana in the past 30 days (Yes = 1, No = 0)



alldrug: If the person had alcohol, cigarette and marijuana in the past 30 days (Yes = 1, No = 0)



DUKE FUQUA SCHOOL OF BUSINESS

4) EDA

EDA: Explore the categorical features

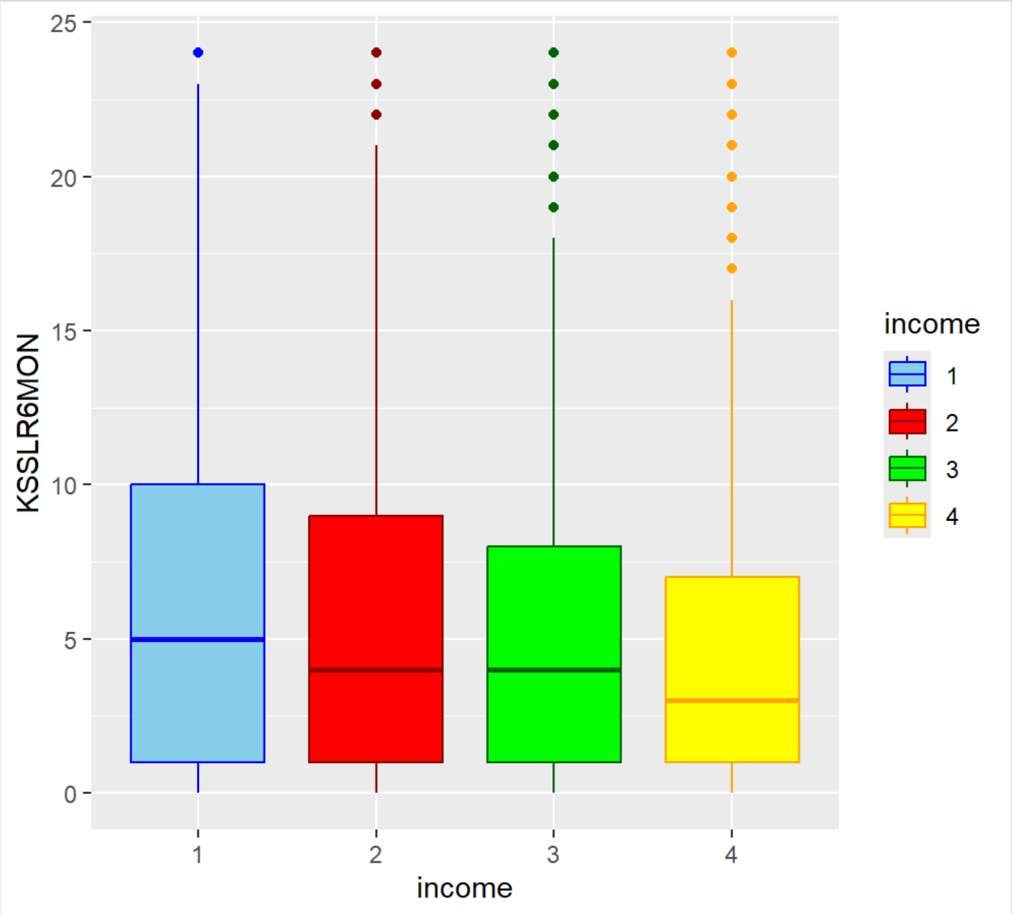
income: total family income recode

1 = Less than $20,000

2 = $20,000 - $49,999

3 = $50,000 - $74,999

4 = $75,000 or More



4) EDA

EDA: Explore the categorical features

CATAG6: age category recode (6 levels)

1 = 12-17 Years Old

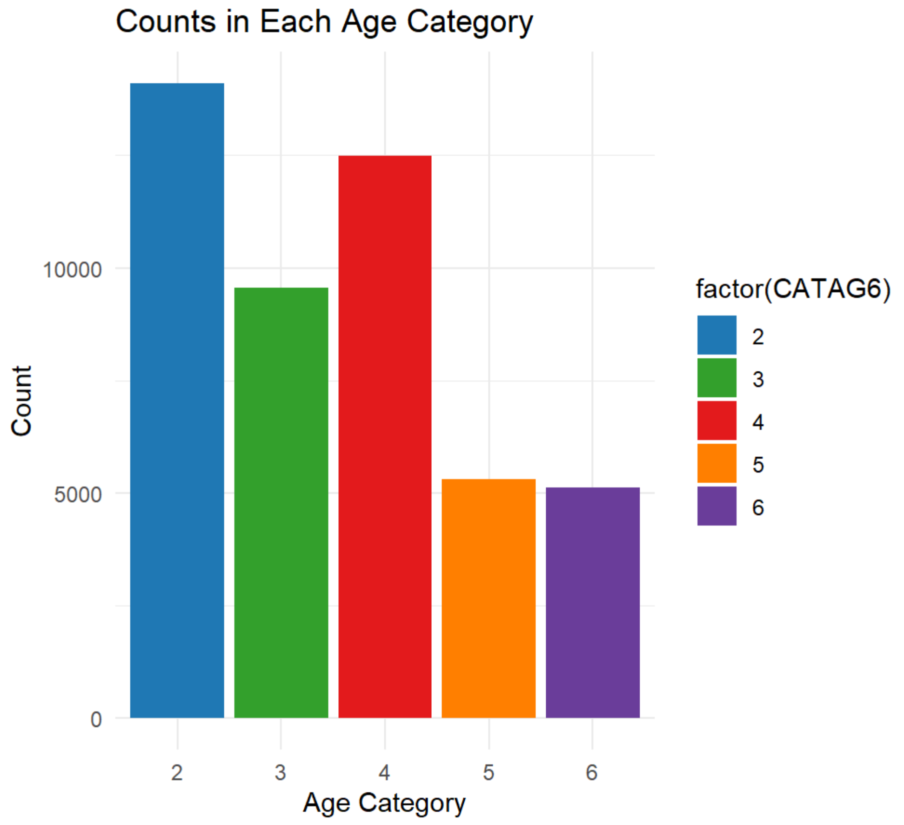
2 = 18-25 Years Old

3 = 26-34 Years Old

4 = 35-49 Years Old

5 = 50-64 Years Old

6 = 65 or Older



4) EDA

EDA: Explore the categorical features

eduhighcat: education categories

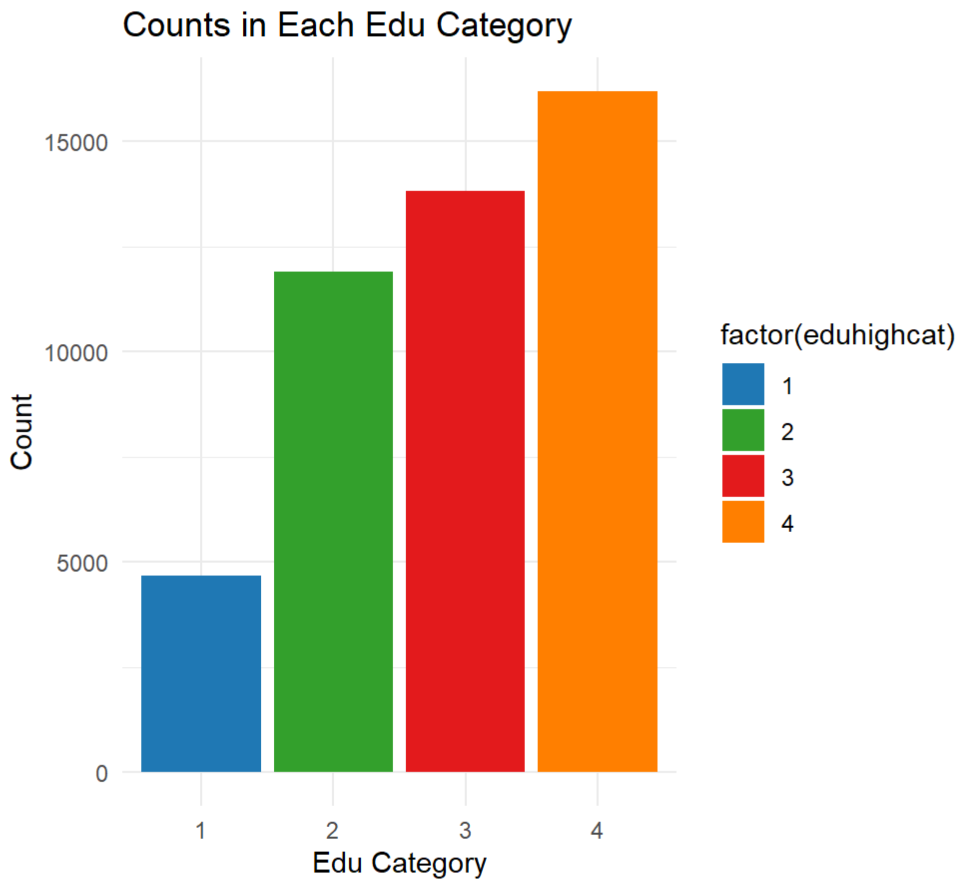
1 = less high school (ireduhighst2=1-7 & age3=4-11)

2 = high school grad (ireduhighst2=8 & age3=4-11)

3 = some coll/assoc dg (ireduhighst2=9-10 & age3=4-11)

4 = college graduate (ireduhighst2=11 & age3=4-11)

5 = 12 to 17 year olds (age3=1-3)



4) EDA

EDA: Explore the categorical features

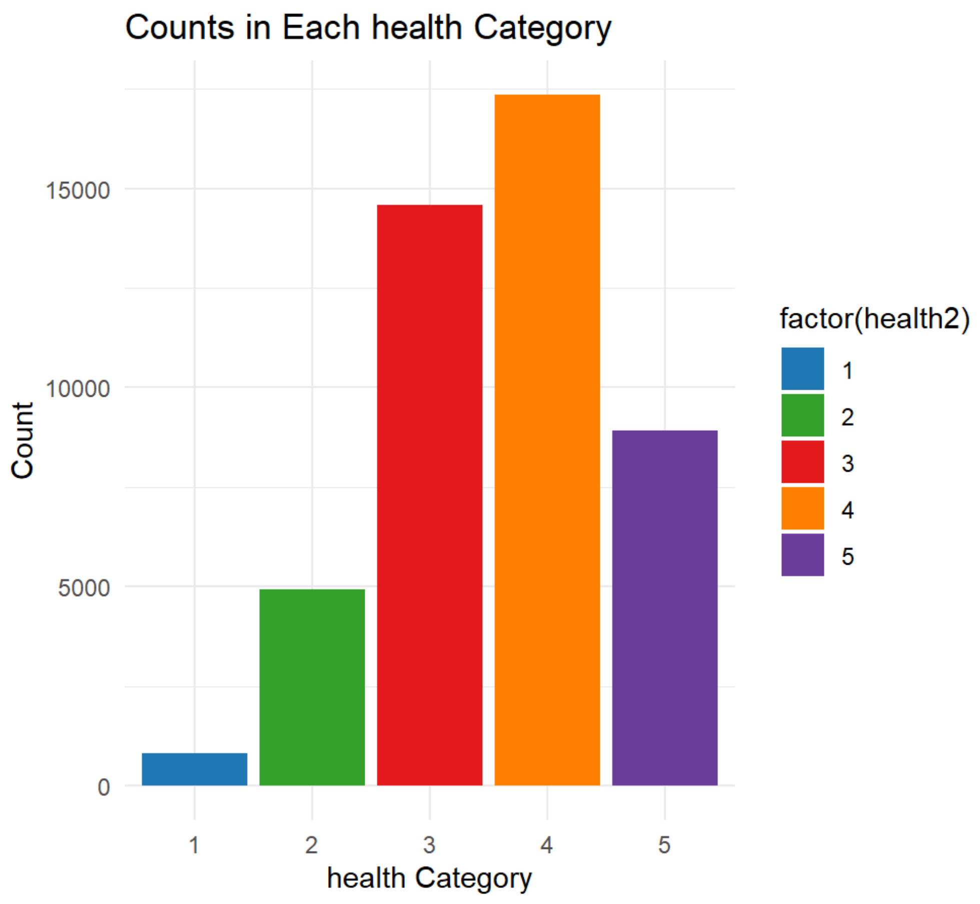
health2: overall health recode

1 = Excellent (HEALTH=1)

2 = Very Good (HEALTH=2)

3 = Good (HEALTH=3)

4 = Fair/Poor (HEALTH=4,5)



4) EDA

EDA: Explore the categorical features

race: race/hispanicity recode (7 levels)

1 = NonHisp White

2 = NonHisp Black/Afr Am

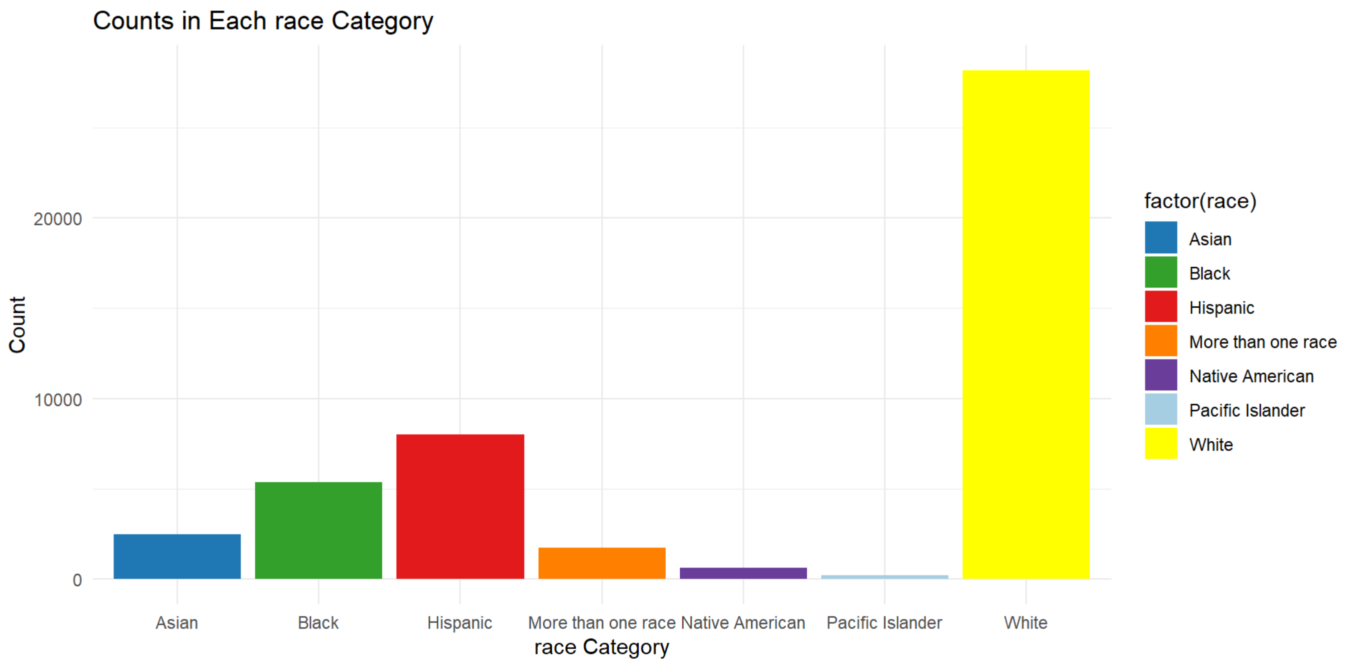
3 = NonHisp Native Am/AK Native

4 = NonHisp Native HI/Other Pac Isl

5 = NonHisp Asian

6 = NonHisp more than one race

7 = Hispanic



5) Methodology

- We used the following methods and models to estimate our average treatment effects: Regression adjustment, regression adjustment with matching, instrumental variables, and causal forests.

- Our outcome variable is a mental health score computed by summing the results of 6 different survey questions, each accepting an answer between 0 and 4. The score is therefore on a scale of 0 to 24 with 0 being the healthiest mentally, and 24 the unhealthiest. The following questions were used in the computation of that score:
  - How often have you felt so sad or depressed that nothing could cheered up in the last 30 days?
  - How often have you felt nervous in the last 30 days?
  - How often have you felt hopeless in the last 30 days?
  - How often have you felt restless or fidgety in the last 30 days?
  - How often have you felt that everything was an effort?
  - How often have you felt down on yourself, no good, or worthless in the last 30 days?

- We've used different treatments depending on the method. The treatments include the following:
  - The use or not of any of alcohol, cigarettes or marijuana in the last 30 days
  - The use of all 3 or not in the last 30 days
  - The number of days in the last 30 days where an individual has had one or more drinks
  - The number of days in the last 30 days where an individual has had one or more cigarettes
  - The number of days in the last 30 days where an individual has consumed Marijuana
  - A heuristic "drug score" computed as a percentage, representing the usage of drugs in the last 30 days

DUKE FUQUA SCHOOL OF BUSINESS

Regression Adjustments Methods

- We run different regression adjustments using the following controls: sex, age, income, highest education level, physical health, race, the interactions between education level and income, education level and age, age and income, and an interaction between education level, age and income all together.
- We apply a log-transformation to the outcome variable to obtain ATE's in terms of percentages.
- For the regressions with continuous treatments (# of days with one drink in last 30 days, # of days with at least one cigarette in last 30 days, # of days consumed marijuana in last 30 days, drug score), we extract the ATE by identifying the coefficient corresponding to the treatment variable.
- For the regressions with binary treatments (Had all three drugs in the last 30 days, had at least one of the three in the last 30 days), we use the regressions to generate predictions of mental health scores for the treatment and for the control groups separately. We take the mean of each group of predictions and subtract the control predictions from the treated predictions to generate the ATEs.

DUKE FUQUA SCHOOL OF BUSINESS

Matching Methods

- Due to the size of the dataset being too large, we take a random sample of 10 000 observations of the dataset to create matching results.
- We use full matching based on Mahalanobis distance to estimate ATE for two models:
    - Model with any drug in the last 30 days as treatment.
    - Model with all 3 drugs in the last 30 days as treatment.
- Full matching obtains propensity score weights for each observation to use in the OLS predictions
- Using the matched datasets, we run OLS with the same controls as before, still using the log of mental health as our outcome variable, with the weights generated from the propensity score weighing.
- We then compare the average contrast for treated and controls on the matched regressions to extract the ATEs

DUKE FUQUA SCHOOL OF BUSINESS

Instrumental Variables Methods

- We use years of alcohol use, tobacco use and marijuana use as IV's for number of days an individual had at least one drink, one cigarette, or consumed marijuana in the last 30 days respectively.
- The justifications of each of these IVs are symmetric. We use the alcohol example as follows (similar to cigarettes and marijuana):
    - The number of years as a drinker is related to your current drinking pattern. This is supported by the covariance of 28 (39 for cigarettes, 21 for marijuana), largely different from 0.
    - We believe there is exclusion as years of alcohol use is likely correlated to mental health only through current drinking patterns. We also believe we've included enough controls to justify independence. This is supported by the covariance of 0.46 (0.63 for cigarettes, 0.94 for marijuana) with the residuals, being close to 0, which reflects variable exogeneity.
- For all three regressions, we now check the weak instrument tests and the Wu-Hausman tests to see if the instruments are weak/strong, and if we do have an endogeneity problem to fix. We reject the null for both tests in all three regressions, both justifying our use of IV's and confirming the belief that the IV's are strong
- We look at the coefficients corresponding to each of the treatments to extract the ATE's for alcohol use in the last 30 days, cigarette use in the last 30 days, and marijuana use in the last 30 days respectively.

DUKE FUQUA SCHOOL OF BUSINESS

Causal Forests Methods

- We run the following methodology for the binary treatments mentioned previously.
- We create an 80/20 training/testing split of the dataset.
- We run a causal forest on the training dataset with the following controls: income, age, sex, physical health, highest education level, years of alcohol use, years of cigarette use, years of marijuana use, interactions previously mentioned, as well as all possible interactions from combinations of years of drugs used. We use 5000 trees
- We predict the individual treatment effects through the causal forest and observe the range and the mean of the conditional average treatment effect.
- We extract the importance of each variable use, and run a causal forest on the testing dataset with variables with above median importance, and obtain the ATE from it.
- We generate a best linear projection of the controls to see along which variables does the conditional average treatment effect vary by identifying the significant variables. The CATE for the any drug treatment varies along the interaction between years of cigarette and marijuana use, and the interaction for all three years of drug usage. The CATE for the all drugs treatment does not vary along any of the variables in our model.

DUKE FUQUA SCHOOL OF BUSINESS

6) Results and findings (RA)

- Every increase in number of days with at least one alcoholic beverage in the last 30 days causes a 0.9% increase in mental distress.
- Every increase in number of days with at least one cigarette in the last 30 days causes a 0.9% increase in mental distress.
- Every increase in number of days with marijuana consumed in the last 30 days causes a1.2% increase in mental distress.
- People who've used any of the three drugs in the last 30 days have 16.7% higher mental distress than those who didn't.
- People who've used all 3 drugs in the last 30 days have 32.7% higher mental distress than those who didn't.
- Every unit increase in the drug score causes a 0.6% increase in mental distress.

DUKE FUQUA SCHOOL OF BUSINESS

6) Results and findings (Matching)

- Taking alcohol, cigarettes and marijuana in the last 30 days causes a 31.4% increase in mental distress
- Taking at least one of the three drugs in the last thirty days causes a 17.5% increase in mental distress.
- One key finding is that the 97.5th higher estimate for the Anydrug treatment is almost the same as the 2.5th lower estimate of the Alldrug treatment (24% and 20% percent respectively).
- This shows that the effect of taking all drugs together is significantly higher than any of the drugs individually.

DUKE FUQUA SCHOOL OF BUSINESS

6) Results and Findings (IV)

- Weak instrument and Wu-Hausman tests reject null hypotheses for all three IV regressions.
- New ATE estimate for alcohol use in the last 30 days: Every increase in number of days with at least one alcoholic beverage in the last 30 days causes a 4% increase in mental distress.
- New ATE estimate for cigarette use in the last 30 days: Every increase in number of days with at least one cigarette in the last 30 days causes a 2.85% increase in mental distress.
- New ATE estimate for marijuana use in the last 30 days: Every increase in days with marijuana consumption in the last 30 days causes a 5.1% increase in mental distress.

DUKE FUQUA SCHOOL OF BUSINESS

6) Results and Findings (Causal Forest)

- Outcome is mental health score (scale 0 to 24), not log of mental health. TE's correspond to change along that scale
- Treatment: At least one drug in the last 30 days:
  - CATE fluctuates between –1.59 and 2.16, with a 0.269 average.
  - ATE estimated at 0.24, statistically significant at the 95% confidence level.
  - CATE conditional on the following variables: Interaction between years of cigarette use and years of marijuana use, interaction between all three years of drug use variables.
- Treatment: All three drugs in the last 30 days:
  - CATE fluctuates between –0.35 and 2.65, with a 1.167 average.
  - ATE estimated at 0.59, statistically significant at the 10% confidence level.
  - CATE does not fluctuate along any of the variables tested.

DUKE FUQUA SCHOOL OF BUSINESS

6) Findings Summary

- Overall, all the methods suggest, as expected that the effect of all 3 drugs on mental health is much worse than that of taking any individual drug.
- The RA and IV regressions both suggest that Marijuana has the strongest effect on mental health, compared to alcohol and cigarettes, which goes against our expectations, as cigarettes were the likely biggest suspect.
- The causal forests have allowed us to identify Conditional Average Treatment Effects. One interesting finding is the fluctuation of the any drug treatment along the variables previously mentioned. Although not necessarily expected, this is a result that makes sense, as long-term cigarette and marijuana use can potentially negatively affect mental health.
- We believe therefore that the joint effect of drugs is in fact much worse on mental health than the use of individual drugs, and that, of the individual drugs tested, marijuana has the worst effect on mental health.

DUKE FUQUA SCHOOL OF BUSINESS

7) Conclusions

- In conclusion, we were able to identify the causal effect of taking alcohol in the last 30 days, cigarettes in the last 30 days, marijuana in the last 30 days, any or all three in the last 30 days, and have found how these effects compare to each other.

- Caveats and Limitations:
  - Complex problems: there are likely more controls to add to fine tune the models, models could suffer from OVB.
  - Improving Causal Forests: With more time, we would likely tune the causal forests to obtain the ideal hyperparameters for most accurate treatment effect predictions.
  - Simultaneity bias: We expect that a worse mental health state increases drug consumption. There is likely a simultaneous system that, although we were able to identify, we would need more time to properly model, as well as identify an IV solution to solve the simultaneity bias.

DUKE FUQUA SCHOOL OF BUSINESS

8) Reference

- https://www.datafiles.samhsa.gov/dataset/national-survey-drug-use-and-health-2022-nsduh-2022-ds0001
- Poudel A, Gautam S. Age of onset of substance use and psychosocial problems among individuals with substance use disorders. BMC Psychiatry. 2017 Jan 11;17(1):10. doi: 10.1186/s12888-016-1191-0. PMID: 28077106; PMCID: PMC5225546.

- Sullivan MD, Edlund MJ, Zhang L, Unützer J, Wells KB. Association Between Mental Health Disorders, Problem Drug Use, and Regular Prescription Opioid Use. Arch Intern Med. 2006;166(19):2087–2093. doi:10.1001/archinte.166.19.2087

- York JL, Welte J, Hirsch J, Hoffman JH, Barnes G. Association of age at first drink with current alcohol drinking variables in a national general population sample. Alcohol Clin Exp Res. 2004 Sep;28(9):1379-87. doi: 10.1097/01.alc.0000139812.98173.a4. PMID: 15365309.

- Gove, W. R., Geerken, M., & Hughes, M. (1979). Drug Use and Mental Health among a Representative National Sample of Young Adults. Social Forces, 58(2), 572–590. https://doi.org/10.2307/2577607

- Newcomb, M. D., & Bentler, P. M. (1979). Drug Use and Mental Health Among a Representative National Sample of Young Adults. Social Forces, 58(2), 572-590. https://doi.org/10.1093/sf/58.2.572

DUKE FUQUA SCHOOL OF BUSINESS