

Final Writeup

Phuc Nguyen, Ezinne Nwankwo, Frances Hung

2/19/2020

Abstract

In this paper, we study the factors that influence college drinking behavior and its effects on other harmful behaviors, in particular sexual assault. We answer three main questions of interest: understanding student-specific factors that influence college drinking behavior, determining if current school policies and alcohol education have any influence on college drinking, and understanding the relationship between college drinking behavior and sexual assault while accounting for other confounding factors (i.e. family drinking behaviour). Through the use of mediation analysis and structural equation models, we are able to isolate the direct effect of other factors on sexual assault and their indirect effect through college drinking behavior. We conclude that in 1997, high school drinking behavior had the largest effects, both directly and indirectly, on college drinking behavior and on incidences of sexual assault.

Introduction

Drinking in college is akin to a rite of passage; while policies around college campus seem to vary in terms of strictness, enforcement is often less stringent than at bars and clubs. Rather than focusing on drinking, many schools care more about the dangerous situations that drinking can lead to, like sexual assault and drunk driving. For this study, we focus on understanding sexual assault, as it is a pervasive and pertinent issue on college campuses. To ground our study, we offer up Duke University as an example of this issue. In a 2018 Student Experience Survey, it was reported that 48 percent of females had been victims of sexual assault while enrolled. We hope to offer guidance to universities on how to address this issue as it relates alcohol use.

We are interested in studying three questions in the hopes that we can help colleges take preventative measures: (i.) which factors influence college drinking behavior (ii.) how school policies influence drinking behavior and (iii.) how those behaviors can lead to sexual assault. Our data comes from the 1997 Harvard School of Public Health College Alcohol Survey (CAS) survey, which was a multi-round survey that interviewed students from over 100 colleges across the US. The survey collected information on various aspects of students' daily activities, living arrangements, social life, sexual activity, satisfaction with education, attitude on campus alcohol policies, their high school drinking behavior and college drinking patterns. Demographic variables such as age, race, sex, parent education level, and parent drinking habits were also collected.

Data Cleaning

The survey had a very complex design that induced a lot of missingness in the data. For example, there were many instances where students skip certain questions based on their answers to previous questions. As a result, there is a lot of missing values for certain questions. We impute zero values for NAs if the variables are ordinal and if NAs are naturally lowest in the ordinal structure. Otherwise, we create indicator variables for whether a person answered NA or not for each variable in question. Afterwards, we drop any remaining missing values as we deemed these to be truly missing as opposed to design induced missingness.

Exploratory Data Analysis

We look at drinking patterns by visualizing a subset of survey question responses from Section C, which asks about a student's drinking patterns in college. By looking at histograms, we can see that drinking at off-campus parties and bars results in the highest number of drinks taken (Fig. 1). The reasons students cited as most important for drinking were getting drunk and as a reward (Fig 2). This suggests that drinking location and social groups may be important predictors of drinking behavior.

For our dependent variable measuring drinking behavior, we choose to use `drinkcat`, a variable created by researchers which classifies drinking behavior into four ordinal risk categories. The sexual assault index variable draws from three questions in the survey that are related to occurrences of sexual assault where alcohol is involved: D5_H (unwanted sexual advance, intoxicated perpetrator), D5_I (date rape/sexual assault, intoxicated perpetrator), E15 (intoxicated, unable to consent). If any one of these questions is marked as occurring at least once by the respondent, then the sexual assault variable is equal to one and zero otherwise.

Methods

We use Structural Equation Modeling (SEM) for modeling relationships between information collected via the survey and our dependent variables `drinkcat` and `sexassault_ind`. SEM is a way of implementing dimension reduction (by using latent factors) while retaining coarse information on all variables of interest. This is different from other dimension reduction strategies which get rid of some variables all together. The SEM package we use in R, `lavaan`, uses partial least squares regression to find a linear model relating the independent and dependent variables through latent factors. We standardize all of the variables we use in order to make our coefficients interpretable and comparable in the final model.

We create the following latent factors from existing survey question variables (Fig. 3). Since these factors must be correlated, interpretable, and continuous (due to our use of the `lavaan` package), we ensure that the survey question variables included in each factor are uniformly ordinal in an interpretable way. Our two factors related to students' pre-college lives are Family Education/Drinking (G14-17) and High School Drinking Behavior (G9-11). The external-policy related factors are College Alcohol Policy (B1, B2, B9), College Alcohol Education (B7, B8), and City Alcohol Policy (B11-13). The rest of the latent factors have to do with students' college experiences and opinions: Academic Wellbeing (F1, F4), Personal Wellbeing (F6), Support for Stricter Policies (D3), Opinions on # Appropriate Drinks (D1), Time Spent on Activities (F5), and General Interests for College Experience (split into activities involving and not involving partying/Greek life, A10). All of the latent factors which draw from one survey question actually consist of several sub-categories within each question (i.e. D3 consists of student rankings of agreement with various school policies).

In order to understand the relationships between drinking behaviors and sexual assault, we fit another latent factor model with mediation on the indicator of being a victim of sexual assault given family education/drinking background, high school drinking, `drinkcat`, and participation in parties. Mediation analysis allows us to estimate how much of the association between a risk factor, such as participation in parties, and the risk of sexual assault is mediated by the victim's drinking behavior.

Results

Drinking Behavior

Table ?? shows the estimates and 95% confidence intervals for the coefficients of latent variables in the drinking behavior model. We accept the model as its RMSEA 95% CI of [0.053, 0.054] is around the conventional threshold of 0.05 for a good fit (CITE). Since the survey answers generally did not have physical units, we standardized the variables to compare the size of the estimated coefficients. The dependent variable, i.e. the rating of levels of drinking, also does not have physical units. Thus, we will interpret these coefficients in terms of the relative magnitude and direction of their association with heavier drinking. High school drinking history is most correlated with more drinking in college. The latent factor Parties summarizing the importance of fraternity/sorority, parties and athletics to students, with a lower value corresponding to more importance, has the second largest association. In other words, students who participate more in the mentioned activities tend to drink more. The latent factor Communities measures the importance of other activities such as volunteering, religion, arts, activism and academic, has the third largest association. Students who care less about these activities tend to drink more. Fourthly, students who perceive a larger number of drink appropriate tend to drink more. The loadings of these four variables are shown in Table 1. Other latent variables with significant but smaller associations with heavier drinking in college include

Personal Wellbeing, Family Education/Drinking and Support for More Lenient Policy. Specifically, students who are generally more happy, whose families approve of drinking, or who support more lenient drinking policies on campus tend to drink more. There is not enough evidence to establish a relationship between stricter college drinking policy and drinking behaviors. We find a small association between more exposure to alcohol/drinking education and heavier drinkers. We hypothesize that students who already have drinking problems or schools with more prevalent drinking culture might consequently have more educational programs on the issue.

Risk of Sexual Assault

We accept the mediation models for when the victim was drinking and when the assaulter was drinking which have RMSEA 95% CI's of [0.029, 0.035] and [0.029, 0.034] respectively. More acceptance of alcohol use from family, drinking history in high school, more participation in parties/Greek life/athletics and heavier drinking are all significantly associated with higher risk of sexual assault in both models, as seen in Table 2 and Table 3. Table and show that the indirect effects of family, high school drinking history and parties mediated through a student's drinking behavior are statistically significant at 5% level. Interestingly, the indirect effects are larger when the victim was drinking compared to when the assaulter was drinking at the time of the incident.

Sensitivity Analysis

We can compare our SEM model for college drinking severity to an elastic net model created using the `glmnet` package. Both of these models carry out dimension reduction using different methods, so it is interesting and reassuring to note that the results of both seem to corroborate one another. From the elastic net model, the most important predictors for heavy drinkers are A10_F(parties, -), B6_D (times they were part of drinking group which was asked to be quieter/less disruptive, +), D1_B (opinion on appropriate amount for off-campus bar drinking, +), D9_B (percentage of friends who are binge drinkers, +), and G11 (HS number of binges, +). This supports the results of our latent factor model, which finds strong correlations of college drinking with drinking attitudes, emphasis on partying, and high school drinking.

Discussion and Future Work

Influential Factors: * High school drinking habits are positively associated with an increase college drinking behavior * Worse academic well-being is associated with an increase in college drinking, while better personal wellbeing is associated with a decrease in college drinking Policies: * No evidence of a relationship between strictness of college or city drinking policy and drinking behaviors * Less education material on drinking is associated with less drinking Sexual Assault: * High school drinking habits alone and through mediator contributes the most to incidences of sexual assault We assume linearity and normality in our models, so in the future we want to extend to non-parametric models with SEM

References

- [1] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2819368/>
- [2] <http://lavaan.ugent.be/index.html>
- [3] An Introduction to Structural Equation Modeling. J.J. Hox. University of Amsterdam/Utrecht University
- [4] <http://archive.sph.harvard.edu/cas/What-We-Learned-08.pdf>

Table 1: Loadings of survey questions for latent factors.

latent_var	survey_question	est	ci_lower	ci_upper	p_value
parties	A10_A	1.00	1.00	1.00	NA
parties	A10_D	1.15	0.99	1.32	0
parties	A10_F	3.46	3.06	3.86	0
communities	A10_B	1.00	1.00	1.00	NA
communities	A10_C	1.29	1.08	1.50	0
communities	A10_E	1.16	0.98	1.34	0
communities	A10_G	2.50	2.16	2.84	0
communities	A10_H	2.86	2.43	3.29	0
attitude_drink	D1_A	1.00	1.00	1.00	NA
attitude_drink	D1_B	0.93	0.91	0.95	0
attitude_drink	D1_C	0.58	0.55	0.60	0
attitude_drink	D1_D	0.75	0.72	0.77	0
attitude_drink	D1_E	0.30	0.27	0.32	0
highschool_drink	G8	1.00	1.00	1.00	NA
highschool_drink	G9	7.21	5.43	8.98	0
highschool_drink	G10	8.63	6.50	10.77	0

Appendix

Figures

Tables

\begin{table}[ht]

	latent_var	est	ci_lower	ci_upper	p_value
1	attitude_policy	0.10	0.05	0.14	0.00
2	communities	0.37	0.18	0.56	0.00
3	parties	-1.10	-1.28	-0.92	0.00
4	attitude_drink	0.37	0.34	0.41	0.00
5	college_policy	0.03	-0.03	0.09	0.29
6	college_edu	-0.09	-0.12	-0.05	0.00
7	city_policy	-0.17	-0.37	0.02	0.08
8	family_drink	0.09	0.04	0.13	0.00
9	highschool_drink	2.21	1.67	2.75	0.00
10	personal_wellbeing	-0.05	-0.09	-0.00	0.04
11	academic_wellbeing	-0.04	-0.13	0.06	0.46
12	social_groups	0.19	-0.14	0.53	0.26

\caption{Estimates and 95% confident intervals of latent factors in the SEM for levels of drinking. High school drinking, support of more lenient drinking policy, importance of parties (parties), less exposure to alcohol education, family drinks, and personal wellbeing are associated with more drinking in college.}

\end{table}

\begin{table}[t]

\caption{Coefficient estimates of three models: SEXASSAULT \sim communities + family_drink + highschool_drink, DRINKCAT \sim communities + family_drink + highschool_drink, SEXASSAULT \sim DRINKCAT. The response is indicator of sexual assault when the assaulter was drinking.}

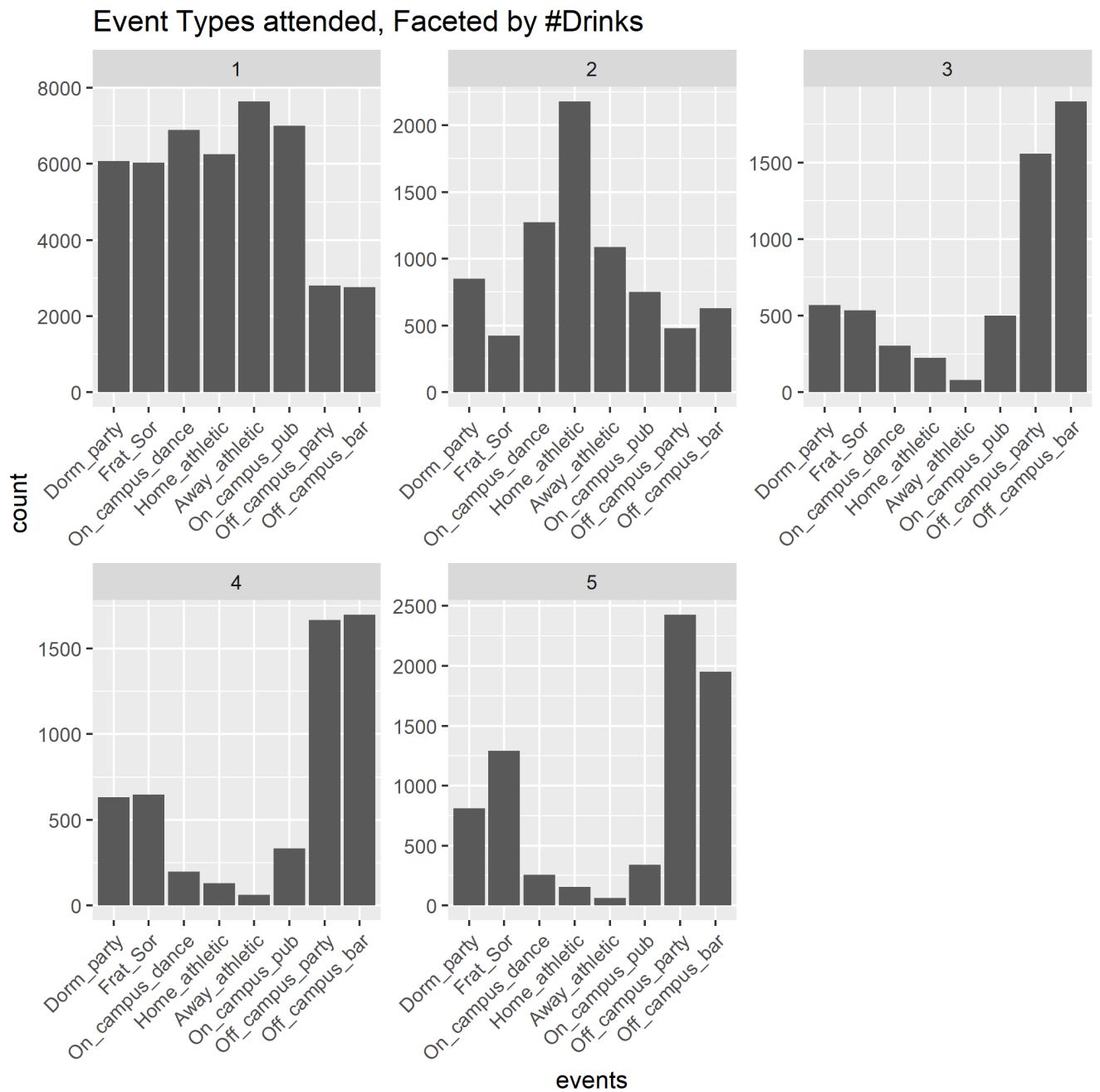


Figure 1: The figure shows the number of students attending certain events, faceted by number of drinks had at an event.

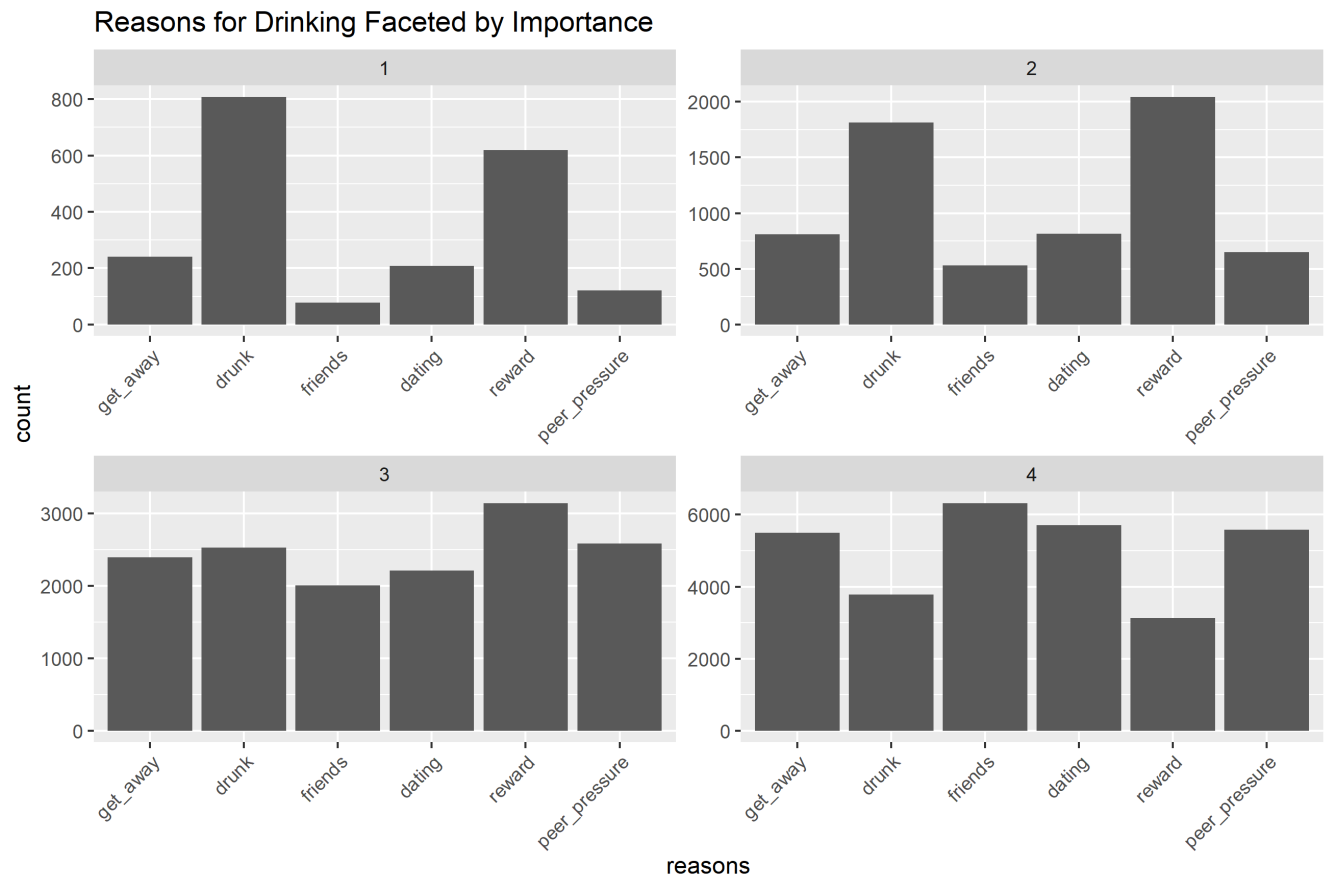


Figure 2: The figure shows survey.

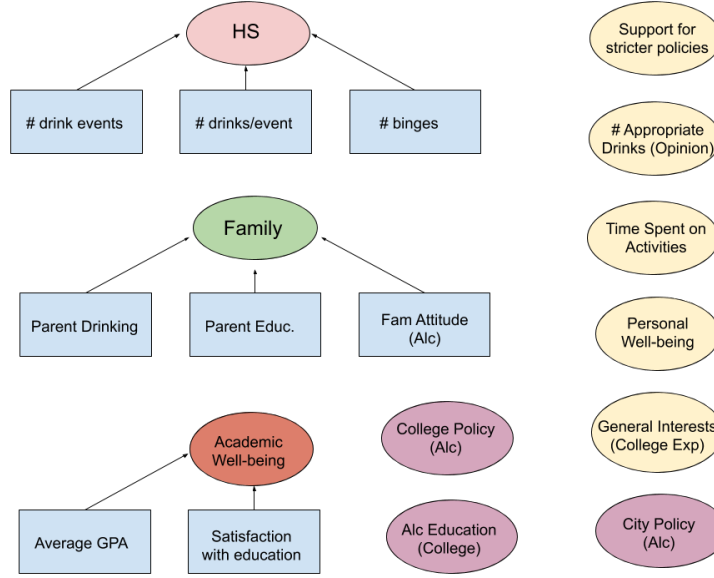


Figure 3: Figure shows latent factors (ovals) in the SEM for drinking behaviors. Squares show summary of actual questions in the survey.

Table 2: Direct and indirect effects of latent factors on risk of sexual assault when victim was drinking.

latent_var	effect_type	est	ci_lower	ci_upper	p_value
communities	indir	-0.27	-0.43	-0.11	0.00
family	indir	0.04	0.02	0.07	0.00
highschool	indir	0.88	0.30	1.47	0.00
communities	dir	-0.48	-0.74	-0.21	0.00
family	dir	0.00	-0.15	0.14	0.99
highschool	dir	0.66	-0.39	1.71	0.22

response	latent_var	est	ci_lower	ci_upper	p_value
SEXASSAULT	communities	-0.83	-0.98	-0.67	0.00
SEXASSAULT	family_drink	0.07	-0.01	0.16	0.09
SEXASSAULT	highschool_drink	1.08	0.47	1.70	0.00
DRINKCAT	communities	-1.51	-1.63	-1.39	0.00
DRINKCAT	family_drink	0.25	0.20	0.29	0.00
DRINKCAT	highschool_drink	4.96	3.55	6.36	0.00
SEXASSAULT	DRINKCAT	0.12	0.06	0.18	0.00

\end{table}

\begin{table}[t]

\caption{Coefficient estimates of three models: SEXASSAULT \sim communities + family_drink + highschool_drink, DRINKCAT \sim communities + family_drink + highschool_drink, SEXASSAULT \sim DRINKCAT. The response is indicator of sexual assault when the victim was drinking.}

Table 3: Direct and indirect effects of latent factors on risk of sexual assault when assaulter was drinking.

latent_var	effect_type	est	ci_lower	ci_upper	p_value
communities	indir	-0.18	-0.27	-0.09	0.00
family	indir	0.03	0.01	0.04	0.00
highschool	indir	0.59	0.24	0.94	0.00
communities	dir	-0.83	-0.98	-0.67	0.00
family	dir	0.07	-0.01	0.16	0.09
highschool	dir	1.08	0.47	1.70	0.00

response	latent_var	est	ci_lower	ci_upper	p_value
SEXASSAULT	communities	-0.48	-0.74	-0.21	0.00
SEXASSAULT	family_drink	0.00	-0.15	0.14	0.99
SEXASSAULT	highschool_drink	0.66	-0.39	1.71	0.22
DRINKCAT	communities	-1.51	-1.63	-1.39	0.00
DRINKCAT	family_drink	0.25	0.20	0.29	0.00
DRINKCAT	highschool_drink	4.91	3.54	6.28	0.00
SEXASSAULT	DRINKCAT	0.18	0.07	0.29	0.00

\end{table}

Appendix

Data Cleaning

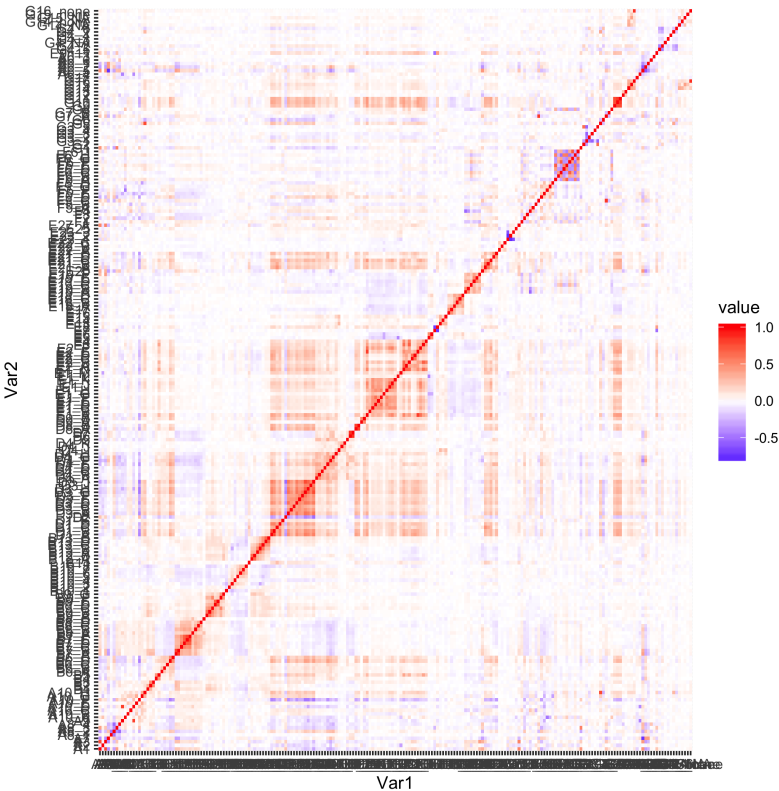


Figure 4: Correlations between Survey Questions.

Elastic Net Model