Effects of Demographic Factors on Heavy Drinking in College

Betsy Bersson Shrey Gupta Xiaojun Zheng

Case Study 3 - STA 723

February 17, 2020

Introduction

- Main Goal: Identify the importance of demographic factors and high school drinking habits in relation to heavy drinking in college.
- Also aim to determine how this may have changed from 1993 to 2001.

Data

Data Available:

- Survey data with about 15,000 observations taken in 4 different years.
- Focus on first and last years available (1993, 2001).

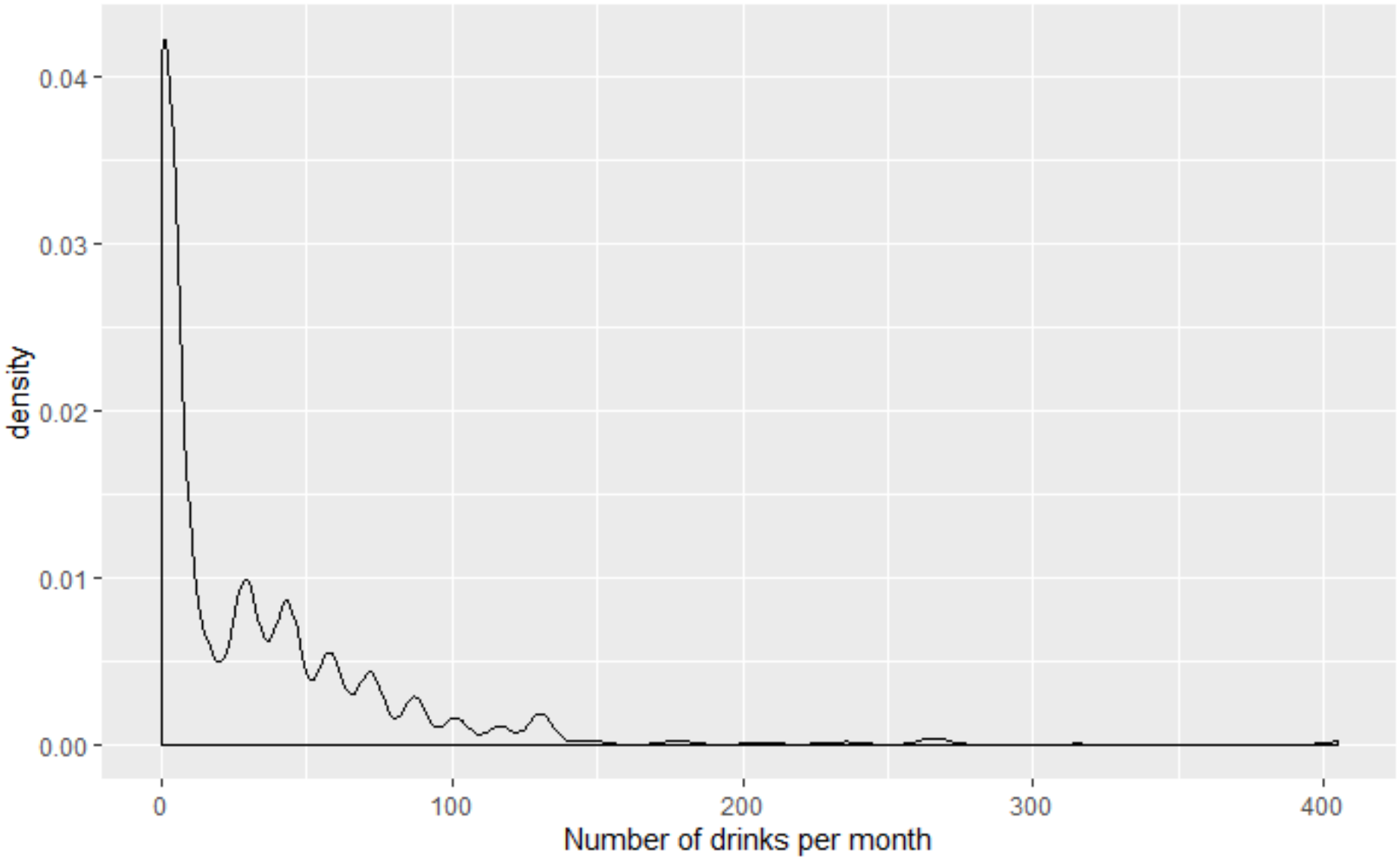
Missing Data:

- Impute some NA to 0 if instructed to not answer the survey question (e.g. C7:C9 based on C6).
- If NAs among response variables, drop these entries.
- MICE the rest NA or "don't know" entries (e.g. GPA).

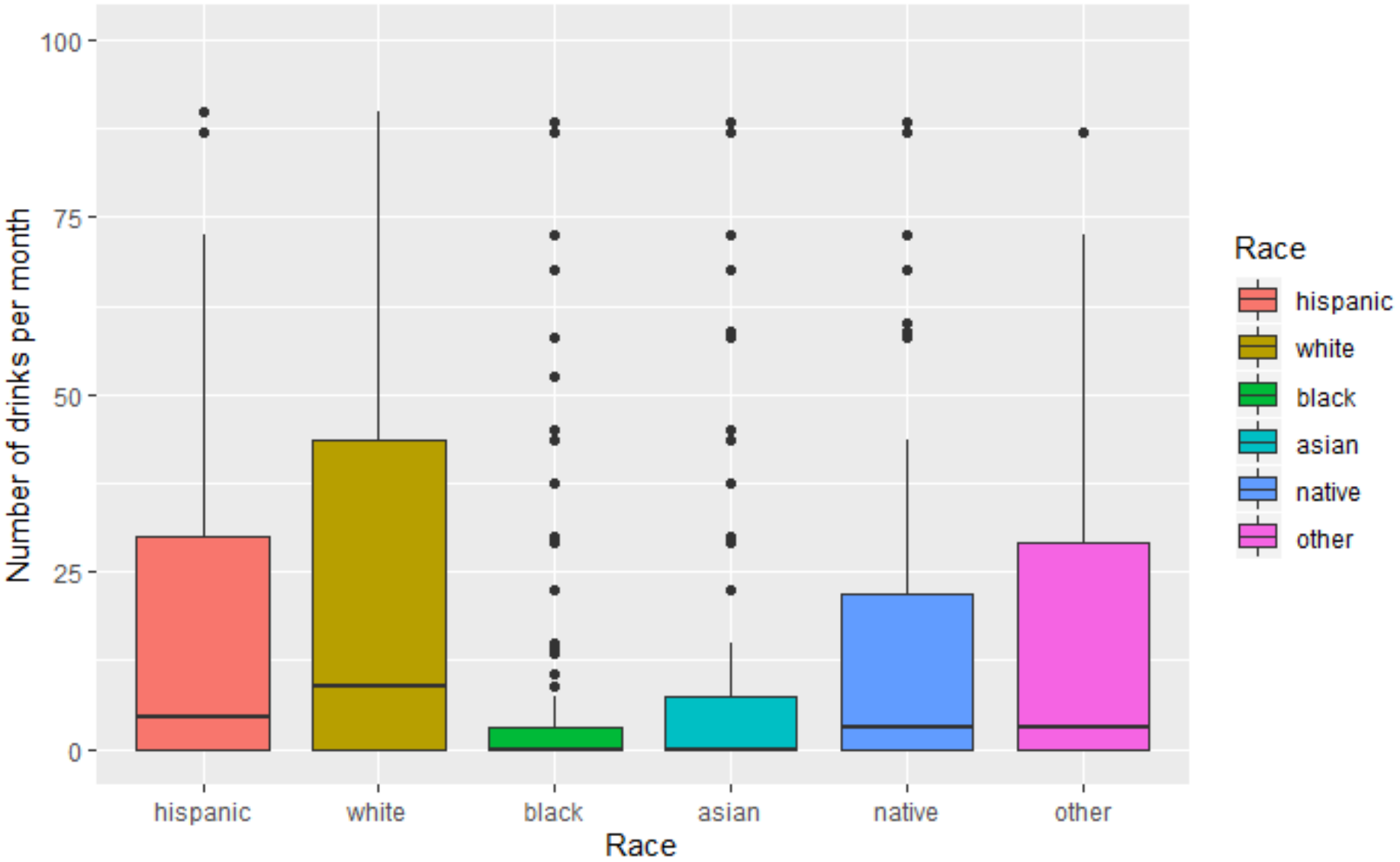
Data Organization

- Response: binary drink or not; binary heavy drink or not; number of drinks per month (mean number of drinks per occasion times number of occasions).
- Important demographic factors: high school drinking behavior, housing situation, race, age, gender, class year, relationship status, religion.
- Controls: greek life, importance of religion, importance of school work, school policy, other drug use, GPA, happiness, height, weight, parent drinking, highest parent education.

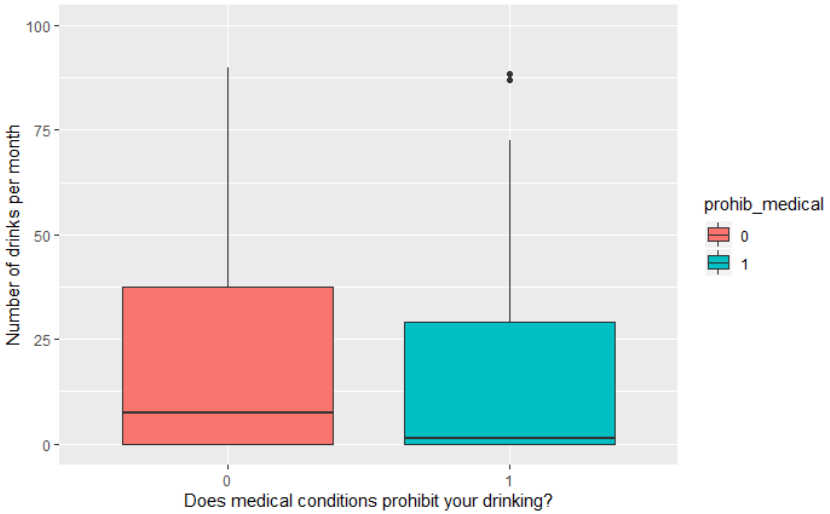
EDA

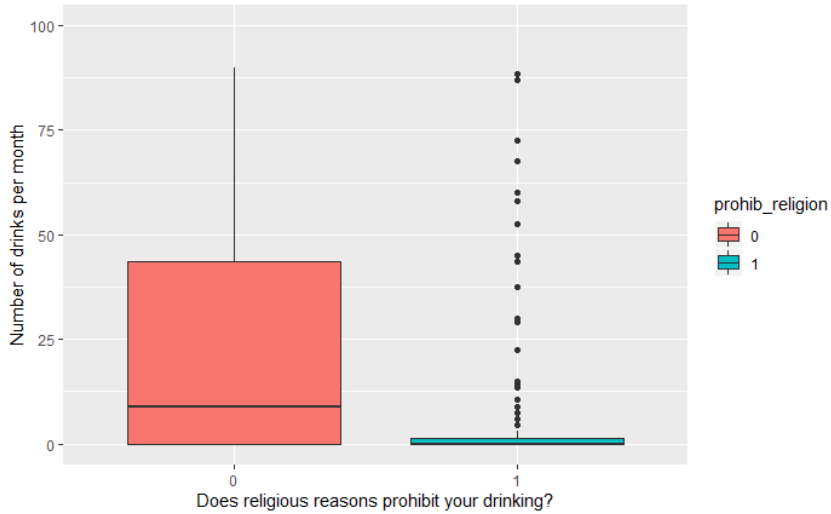

0.04
0.03
0.02
0.01
0.00
density
0
100
200
300
400
Number of drinks per month

EDA

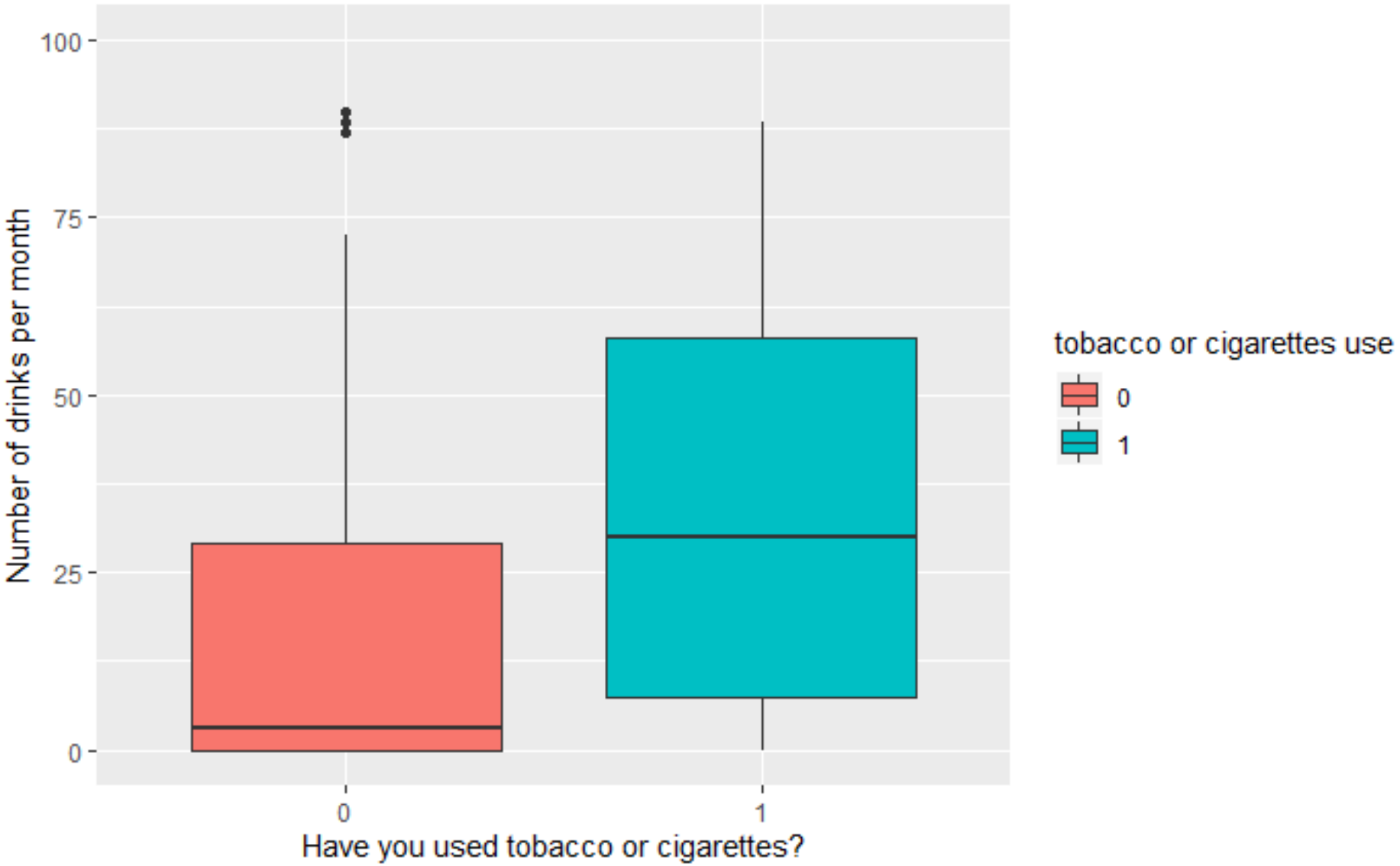

100
75
50
25
0
Number of drinks per month
hispanic
white
black
asian
native
other
Race
Race
hispanic
white
black
asian
native
other

EDA

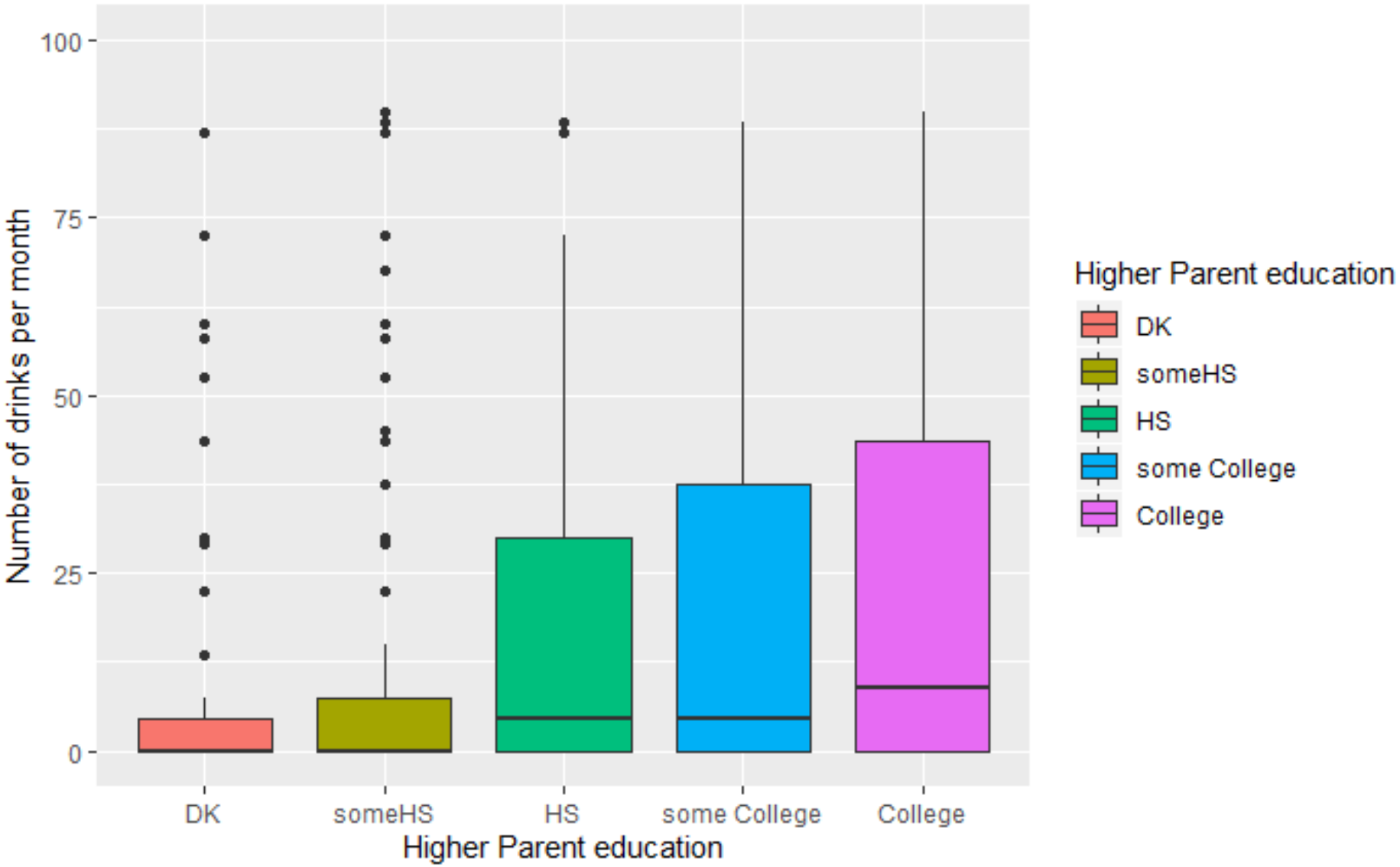

100
75
50
25
0
Number of drinks per month
0
1
Does medical conditions prohibit your drinking?
prohib_medical
0
1


100
75
50
25
0
Number of drinks per month
0
1
Does religious reasons prohibit your drinking?
prohib_religion
0
1

EDA


100
75
50
25
0
Number of drinks per month
0
1
Have you used tobacco or cigarettes?
tobacco or cigarettes use
0
1

EDA


100
75
50
25
0
Number of drinks per month
DK
someHS
HS
some College
College
Higher Parent education
Higher Parent education
DK
someHS
HS
some College
College

Methods

Implement three models over 2 survey years (1993, 2001) and compare output.

- Logistic model on binary response if students drink or not.
- Logistic model on binary response if students are heavy drinkers or not.
  - Male: Heavy drinker if count of monthly drinks ≥ 65.
  - Female: Heavy drinker if count of monthly drinks ≥ 35.
- Negative Binomial model on count of monthly drinks among students who do drink.
  - The negative binomial model is a poisson model where λ is also random.
  - If $Y|\lambda \sim \text{poisson}(\lambda)$, and $\lambda \sim \text{gamma}(r, \frac{1-p}{p})$.
  - Then $Y \sim NegBinom(r, p)$ where $\text{E(Y)} = \mu$ and $\text{Var(Y)} = \mu + \frac{\mu^2}{r}$.
  - The overdispersion is given by $\frac{\mu^2}{r}$, which approaches 0 as r increases.

Results: Negative Binomial Model, 1993

	Lower bound	Mean	Upper bound
(Intercept)	20.74	45.06	97.91
age_18_20	0.40	0.63	0.98
age_21_22	0.39	0.62	0.96
age_23up	0.36	0.57	0.90
gender	0.77	0.81	0.86
roommate	1.07	1.13	1.20
family_roomate	0.85	0.91	0.97
greek	1.18	1.23	1.29
academic_important	0.77	0.85	0.93
religion_important	0.86	0.90	0.93
school_disc_once	1.27	1.33	1.39
school_disc_mult	1.37	1.43	1.50
prohib_medical	0.82	0.89	0.96
prohib_religion	0.69	0.76	0.83
tobac	1.33	1.38	1.44
marijuana	1.25	1.32	1.39
hard_drugs	1.07	1.16	1.25
gpa	0.86	0.88	0.91
married	0.70	0.77	0.84
div_sep_wid	0.73	0.83	0.94
white	1.07	1.16	1.26
black	0.69	0.78	0.88
asian	0.76	0.85	0.96
religion_catholic	1.09	1.16	1.23
religion_jewish	0.78	0.86	0.96
religion_prot	1.02	1.09	1.16
height	1.01	1.01	1.02
HS_num_of_drinks	1.00	1.00	1.00
mom_heavy_drinker	1.05	1.17	1.30
dad_heavy_drinker	1.00	1.06	1.12

Results: Negative Binomial Model, 2001

	Lower bound	Mean	Upper bound
(Intercept)	8.13	20.59	52.13
gender	0.74	0.79	0.84
on_housing	1.03	1.11	1.19
off_housing	1.05	1.13	1.22
roommate	1.06	1.13	1.21
family_roommate	0.78	0.84	0.91
greek	1.22	1.30	1.38
academic_important	0.74	0.82	0.91
religion_important	0.89	0.94	0.98
prohib_medical	0.83	0.90	0.98
prohib_religion	0.78	0.81	0.85
tobac	1.32	1.39	1.46
marijuana	1.17	1.24	1.32
hard_drugs	1.14	1.22	1.31
aa_attendance	1.19	1.34	1.51
married	0.74	0.83	0.93
white	1.03	1.12	1.21
black	0.77	0.88	1.00
asian	0.76	0.87	0.98
religion_catholic	1.10	1.18	1.26
height	1.00	1.01	1.02
HS_num_of_drinks	1.00	1.00	1.00
school_prohibs	1.02	1.08	1.15
school_disc_once	1.03	1.18	1.36
school_disc_mult	1.33	1.48	1.64

Results: Logistic Model for Heavy Drinkers, 1993 (Females)

	Lower bound	Mean	Upper bound
(Intercept)	0.02	0.13	0.71
roommate	1.54	1.75	1.99
greek	1.55	1.79	2.06
religion_important	0.67	0.76	0.87
school_disc_once	1.83	2.10	2.42
school_disc_mult	2.07	2.38	2.74
prohib_medical	0.59	0.74	0.93
prohib_religion	0.28	0.38	0.52
tobac	2.12	2.40	2.73
marijuana	1.57	1.86	2.21
hard_drugs	1.31	1.73	2.29
gpa	0.66	0.73	0.81
married	0.30	0.41	0.55
div_sep_wid	0.32	0.48	0.73
white	1.20	1.51	1.90
black	0.26	0.40	0.60
asian	0.40	0.60	0.89
native	0.14	0.40	1.15
religion_catholic	1.20	1.44	1.73
religion_jewish	0.55	0.76	1.07
religion_prot	1.17	1.40	1.69
height	1.00	1.02	1.04
HS_num_of_drinks	1.01	1.01	1.01

Results: Logistic Model for Heavy Drinkers, 2001 (Females)

	Lower bound	Mean	Upper bound
(Intercept)	0.04	0.26	1.76
age_23up	0.67	0.82	1.00
roommate	1.05	1.27	1.52
family_roommate	0.53	0.66	0.82
greek	1.45	1.72	2.05
religion_important	0.77	0.89	1.02
prohib_medical	0.51	0.65	0.82
prohib_religion	0.54	0.61	0.70
tobac	2.30	2.63	3.02
marijuana	1.63	1.92	2.27
hard_drugs	1.10	1.35	1.66
married	0.35	0.52	0.78
hispanic	0.51	0.66	0.86
black	0.48	0.66	0.90
asian	0.36	0.50	0.67
native	0.10	0.30	0.92
religion_catholic	1.18	1.34	1.53
weight	0.99	1.00	1.00
height	0.99	1.02	1.05
HS_num_of_drinks	1.00	1.00	1.01
dad_heavy_drinker	0.69	0.85	1.03
max_parent_edu_someHS	0.13	0.37	1.05
max_parent_edu_HS	0.15	0.40	1.05
max_parent_edu_someCol	0.13	0.33	0.84
max_parent_edu_col	0.14	0.35	0.91
school_prohibs	1.11	1.31	1.55
school_disc_once	1.15	1.73	2.59
school_disc_mult	1.83	2.60	3.69

Results: Logistic Model for Heavy Drinkers, 1993 (Males)

	Lower bound	Mean	Upper bound
(Intercept)	0.01	0.04	0.10
age_18_20	1.52	1.89	2.33
age_21_22	1.25	1.53	1.89
off_housing	1.07	1.26	1.49
roommate	1.03	1.29	1.63
family_roommate	0.51	0.67	0.89
greek	1.54	1.82	2.14
academic_important	0.54	0.74	1.00
religion_important	0.66	0.78	0.93
school_disc_once	1.68	2.00	2.39
school_disc_mult	1.99	2.36	2.80
prohib_medical	0.38	0.53	0.73
prohib_religion	0.32	0.47	0.70
tobac	2.01	2.31	2.67
marijuana	1.57	1.88	2.25
hard_drugs	1.05	1.38	1.81
aa_attendance	0.22	0.45	0.92
happy	1.14	1.50	1.97
not_happy	1.19	2.01	3.42
gpa	0.72	0.81	0.91
married	0.41	0.64	1.00
white	1.33	1.63	2.01
native	0.82	2.11	5.44
religion_catholic	1.21	1.40	1.61
religion_jewish	0.49	0.71	1.03
weight	1.00	1.01	1.01
HS_num_of_drinks	1.01	1.01	1.01
dad_heavy_drinker	1.08	1.34	1.67

Results: Logistic Model for Heavy Drinkers, 2001 (Males)

	Lower bound	Mean	Upper bound
(Intercept)	0.03	0.14	0.58
age_23up	0.57	0.74	0.95
subfree_on_housing	0.57	0.78	1.06
off_housing	1.00	1.26	1.59
roommate	1.01	1.31	1.71
family_roommate	0.43	0.60	0.83
greek	1.32	1.68	2.13
academic_important	0.37	0.53	0.75
prohib_medical	0.40	0.59	0.85
prohib_religion	0.47	0.58	0.72
tobac	1.53	1.85	2.23
marijuana	1.27	1.59	1.97
hard_drugs	1.54	1.98	2.55
aa_attendance	1.14	1.68	2.48
happy	1.14	1.57	2.17
gpa	0.70	0.82	0.96
married	0.36	0.64	1.13
hispanic	0.47	0.67	0.97
black	0.30	0.50	0.84
asian	0.34	0.54	0.84
native	0.74	3.05	12.60
religion_catholic	1.11	1.34	1.61
weight	1.01	1.01	1.01
HS_num_of_drinks	1.00	1.01	1.01
max_parent_edu_someHS	0.12	0.37	1.16
max_parent_edu_HS	0.13	0.36	1.02
max_parent_edu_someCol	0.13	0.37	1.01
max_parent_edu_col	0.12	0.32	0.88
school_prohibs	0.94	1.22	1.57
school_disc_mult	2.52	3.53	4.93

Results: Logistic Model for Drinkers, 1993

	Lower bound	Mean	Upper bound
(Intercept)	0.07	0.24	0.81
age_21_22	1.25	2.92	6.84
age_23up	1.02	2.42	5.76
gender	1.23	1.45	1.70
year_in_school	1.14	1.23	1.33
on_housing	1.08	1.32	1.60
off_housing	1.73	2.05	2.43
family_roommate	0.44	0.52	0.62
greek	1.51	1.98	2.59
religion_important	0.54	0.63	0.74
school_prohibs	0.71	0.82	0.95
school_disc_once	1.76	2.17	2.68
school_disc_mult	1.30	1.58	1.92
prohib_religion	0.23	0.27	0.32
tobac	6.96	12.37	22.00
marijuana	2.48	17.94	129.75
gpa	0.75	0.84	0.95
married	1.39	1.87	2.50
div_sep_wid	1.59	2.85	5.09
white	1.13	1.36	1.63
asian	0.53	0.68	0.88
religion_catholic	1.21	1.43	1.69
religion_jewish	1.03	1.76	2.99
religion_moslem	0.34	0.56	0.92
HS_num_of_drinks	1.43	1.51	1.59
max_parent_edu_HS	1.45	2.85	5.60
max_parent_edu_someCol	1.25	2.42	4.71
max_parent_edu_col	1.24	2.40	4.65

Results: Logistic Model for Drinkers, 2001

	Lower bound	Mean	Upper bound
(Intercept)	0.06	0.20	0.73
age_18_20	1.23	3.43	9.58
age_21_22	1.67	4.79	13.71
age_23up	1.35	3.92	11.36
gender	1.33	1.57	1.84
year_in_school	1.10	1.20	1.31
off_housing	1.12	1.42	1.79
family_roommate	0.45	0.55	0.68
greek	1.19	1.59	2.12
religion_important	0.55	0.65	0.78
prohib_medical	0.27	0.33	0.39
prohib_religion	0.34	0.40	0.48
tobac	6.21	10.42	17.47
marijuana	4.33	11.83	32.31
hard_drugs	1.41	2.89	5.95
aa_attendance	2.34	15.26	99.64
white	1.18	1.43	1.73
black	1.30	1.72	2.26
religion_catholic	1.50	1.82	2.20
religion_jewish	1.39	2.80	5.64
religion_other	0.93	1.22	1.60
HS_num_of_drinks	1.32	1.37	1.43
max_parent_edu_someHS	1.01	2.19	4.75
max_parent_edu_HS	1.26	2.59	5.32
max_parent_edu_someCol	1.21	2.44	4.94
max_parent_edu_col	1.04	2.09	4.21
school_disc_once	0.12	0.17	0.24
school_disc_mult	0.05	0.07	0.10