

# DevCon 6/15/16

---

Michigan Longitudinal Study

Stephanie Zellers

# Recruitment

- Data collected from 3 generations
  - G2 targets are children of fathers with alcohol problems
  - Biological son between ages 3-5 designated the male target child (MTC)
  - One full female sibling of MTC designated the female target child (FTC), no info given on how chosen if more than one full female sibling
  - All full siblings within +/- 6 years of age from MTC eligible to participate
    - FTC and other full siblings not recruited from project start
    - Half siblings recruited from 2006 onwards
- Data Collection begins when MTC is 3-5 years old
- Subjects recruited in 3 groups: court alcoholic/high risk, community alcoholic/intermediate risk, and community control/low risk

# Reading the G2 ID numbers

- 4-7 digits depending on length of family ID (no leading zeros in this dataset)
- Four positions WXYZ
- W is Family ID
  - 1-599 are sample 1 families (1980s)
  - 600-899 are sample 2 families (1990s)
  - 9000 onwards are sample 3 families (early 2000s)
- X is member
  - 3 is Male target child
  - 4 is Female target child
  - 5-9 are siblings
- Y is Sex
  - 1 is Female
  - 2 is Male
- Z is step
  - 0 is biological child
  - 1 is half sibling on mother's side
  - 2 is half sibling on father's side

# Assessments

- Assessments every 3 years (T) for a total of 8 waves
- In home assessment at baseline
- Shorter battery annually from ages 11-26
  - Annual assessment is omitted during the years when T wave collection is due
  - Annuals sorted into adolescent and adult
    - Adolescent annuals labeled A1-A7 during ages 11-17
    - Adult annuals labeled A18-A26 during ages 18-26
- Retrospective reporting

# Descriptives

- G2 data only: MTC, FTC, and other siblings
- Annuals only, no T wave data
- N = 1250
- 544 families represented in this subset, all three samples represented
- Gender is M/F binary, 70.48% male
- Race
  - 74% Caucasian
  - 7.52% African American
  - .32% Native American
  - 4.08% Bi-racial
  - 5.28% Hispanic ethnicity, Caucasian race
  - 8.6% not specified

# Variables of Interest: Quantity and Frequency

- dv- Number of drinks in the year specified
  - not present in any documentation, neither as a question nor composite variable
- maxdrnk- maximum drinks in 24 hours in year specified
  - composite based on self reported quantity across beer, wine, liquor
  - not present in documentation as is, but there is a max drinks in 24 hours in last 3 years
- smoke- number of days of cigarette use in specified year
  - discrete categories, variable name/categories do not match data codebook
- mj- number of days of marijuana use in specified year
  - discrete categories, variable name/categories do not match data codebook

# Variables of Interest: Binge Drinking Behavior

- binge- presence of binge drinking (binary y/n) in past 6 months for specified year
  - composite based on responses to earlier reported quantities across beer, wine, and liquor
- bingedk- sum of beer, wine, and liquor drinking days in specified year
  - composite based on estimate for one year based on past 6 months across beer, wine, and liquor
- cbinge- ever had a binge drinking occasion until and including specified year
  - likely a composite of binge, computation not noted in documentation
- cbingedk- binge drinking days until and including specified year
  - composite of bingedk, computation not noted in documentation

# Variables examined: five number summaries, means, and histograms

- maxdrnk at ages 13, 17, 21, 25
- binge at all ages
- cbinge at all ages
- bingedk at ages 13, 17, 21, 25
- smoke at ages 13, 17, 21, 25
- mj at ages 12, 19, 21, 25
  - Excluded mj at ages 13-18 due to unresolved coding issue, multiple instances of frequency of 450 days per year, does not persist within subject
  - I intend to go back into documentation to check before just removing the outliers
- Representative ages chosen due of general trend of means and frequencies when looking across all ages
  
- Did not provide examples of cbingedk since I was unsure what utility this composite variable would have over the yearly reporting
- Did not provide examples of dv because it does not seem as informative as other variables like frequency of drinking per week or month, drinks in typical drinking session, or max drinks in single session



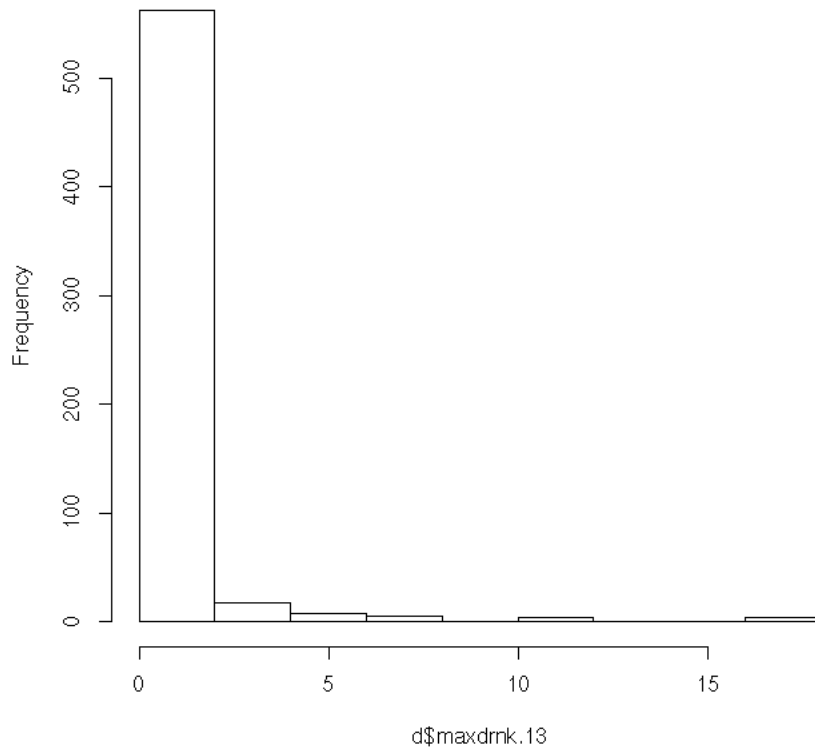
# maxdrnk: maximum drinks in 24 hours in specified year

Age	Min	1st Q	Median	3rd Q	Max	Mean (SD)	NA
13	0.00	0.00	0.00	0.00	17.00	.53 (2.01)	649
17	0.00	0.00	0.00	12.00	31.00	5.66 (7.78)	785
21	0.00	5.50	12.00	17.00	31.00	12.28 (8.54)	980
25	0.00	5.50	12.00	17.00	31.00	10.89 (8.36)	1052

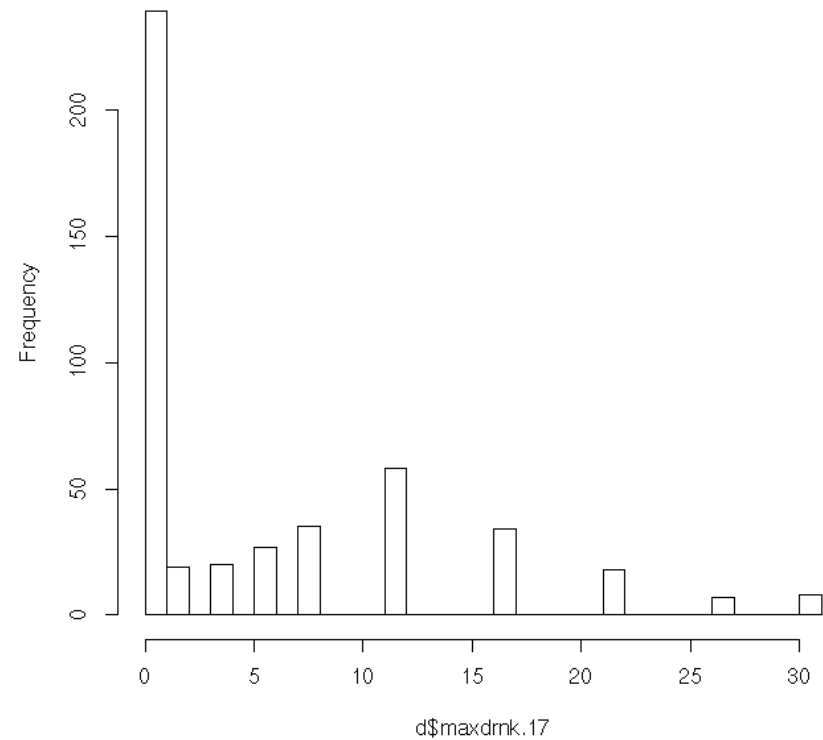
- codebook: max3y, recall time is last three years, as opposed to past year
  - 0=None
  - 2=1-2 drinks
  - 4=3-4 drinks
  - 6=5-6 drinks
  - 8=7-9 drinks
  - 12=10-14 drinks
  - 17=15-19 drinks
  - 22=20-24 drinks
  - 27=25-29 drinks
  - 30=30 or more drinks
- dataset
  - 0, 1.5, 3.5, 5.5, 8, 12, 17, 22, 27, 31, no labeling

# maxdrnk: maximum drinks in 24 hours in specified year

Histogram of d\$maxdrnk.13

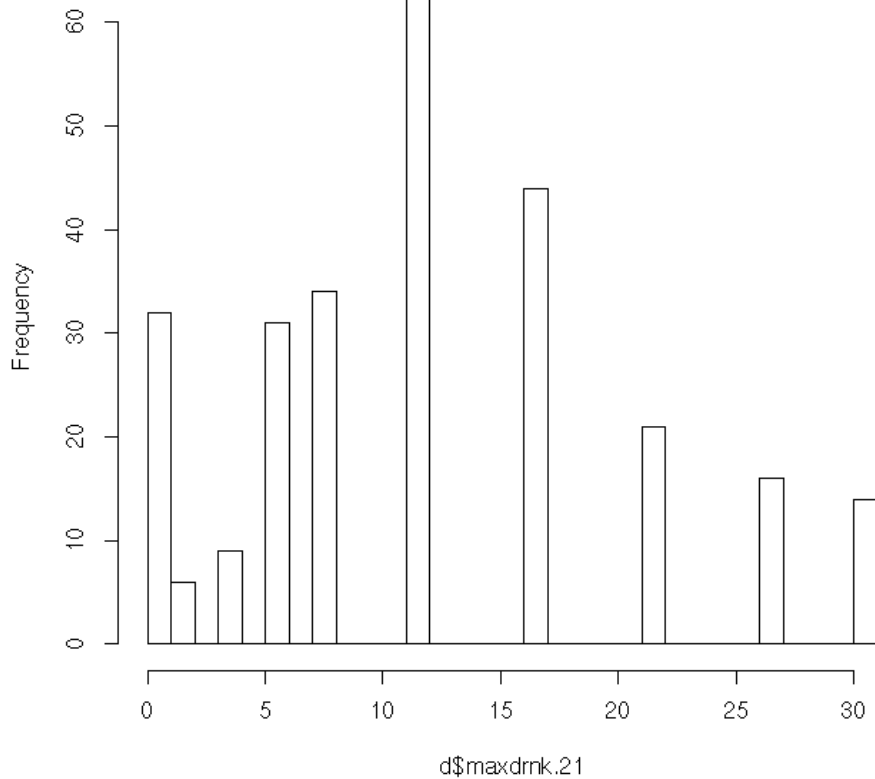


Histogram of d\$maxdrnk.17

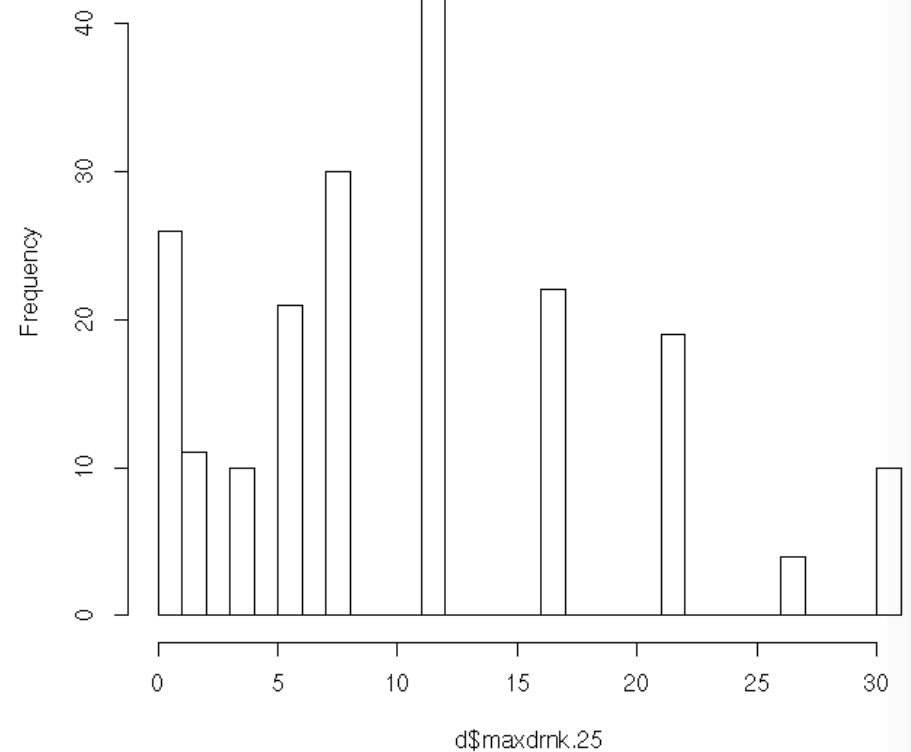


# maxdrnk: maximum drinks in 24 hours in specified year

Histogram of d\$maxdrnk.21



Histogram of d\$maxdrnk.25



# Binge Drinking Behavior

Age	Binge %Y	Binge N	Cbinge % Y	Cbinge N
11	0	505	0	1051
12	0.36	556	0.19	1031
13	2.84	598	1.79	1004
14	5.31	622	4.53	971
15	14.68	613	12.27	929
16	27.2	592	23.68	874
17	39.46	555	36.36	814
18	49.67	304	44.74	751
19	59.66	290	54.27	715
20	62.5	305	64.35	676
21	74.28	311	72.53	637
22	72.93	314	75.71	597
23	67.96	309	78.64	557
24	68.09	282	80.72	498
25	71.54	253	80.99	426
26	62.5	208	80.31	320

- Note that there is a large difference in number of responses for binge and cbinge at all ages
- Also note that binge is described as presence of binge drinking in previous 6 months, whereas cbinge is described as presence of binge drinking up through the specified age (lifetime binge presence)

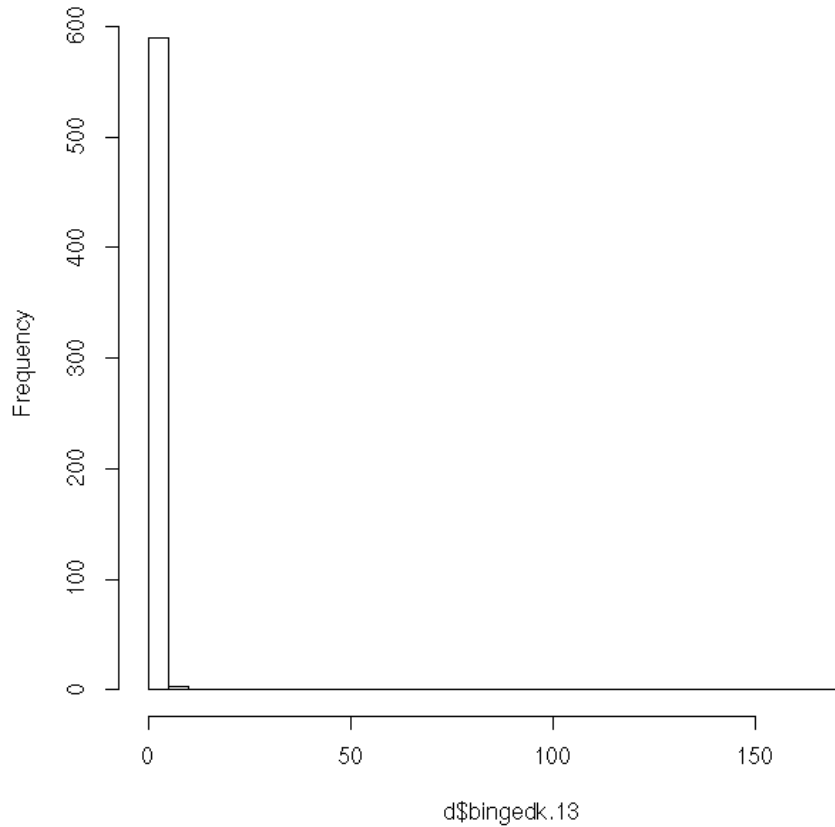
# bingedk: number of binge drinking days in a given year

Age	Min	1st Q	Median	3rd Q	Max	Mean (SD)	NA
13	0.00	0.00	0.00	0.00	178.20	.84 (9.87)	652
17	0.00	0.00	0.00	11.40	365.00	17.98 (47.60)	695
21	0.00	0.00	26.40	84.00	365.00	62.18 (85.27)	939
25	0.00	0.00	25.20	84.00	365.00	57.29 (81.64)	997

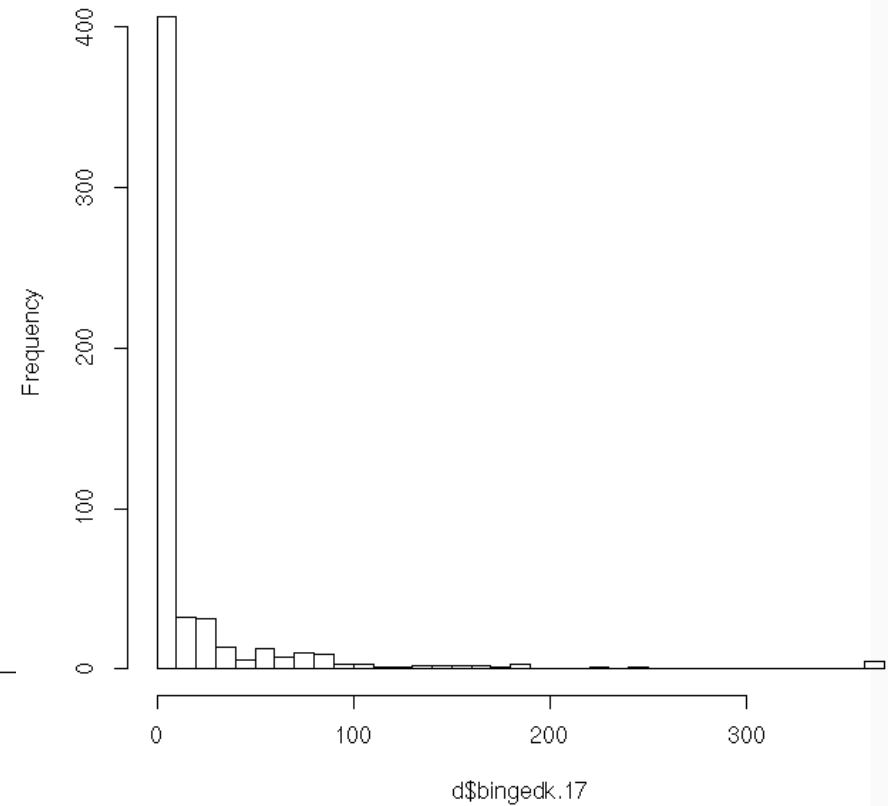
- Frequency tables for dataset and coding manual both suggest that this is continuous variable for exact estimate of number of days, as opposed to discrete values like for smoke and mj
- Intermediate variables computed based on quantity and frequency in past 6 months for each wine, beer, and liquor (winedk, beerdk, and liqdk variables not present in this data set), to get an estimate for amount in one year
- bingedk is a sum of beerdk, winek, and liqdk
- maximum set to 365 for bingedk

# bingedk: number of binge drinking days in a given year

Histogram of d\$bingedk.13

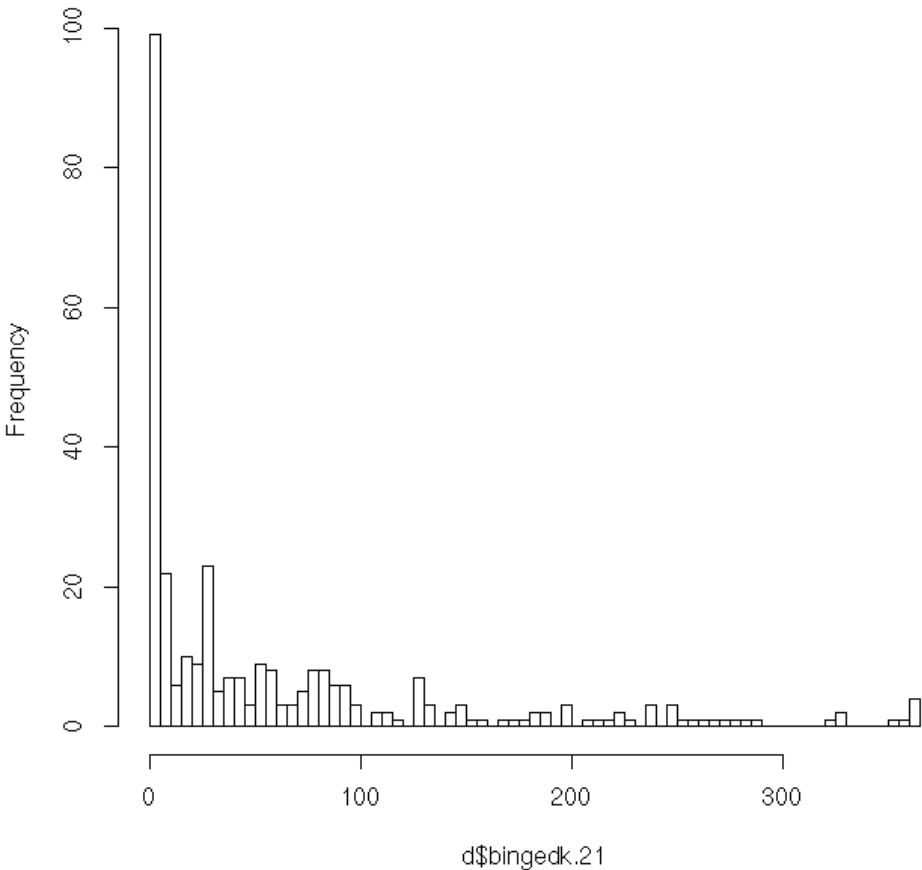


Histogram of d\$bingedk.17

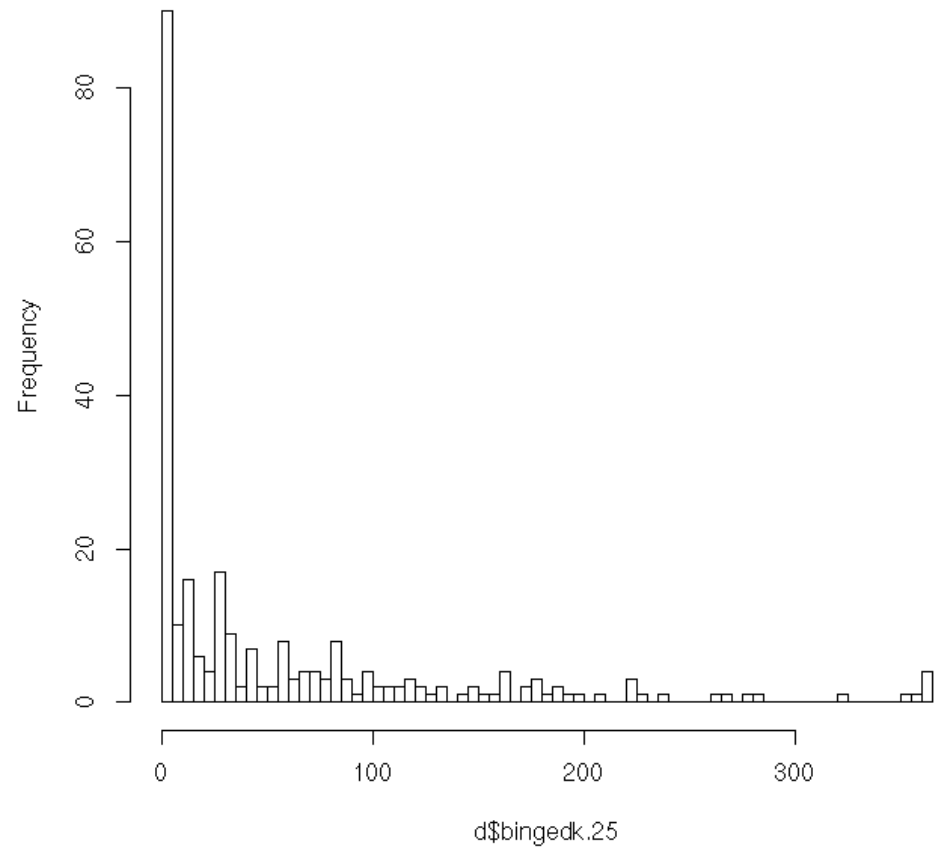


# bingedk: number of binge drinking days in a given year

Histogram of d\$bingedk.21



Histogram of d\$bingedk.25



# smoke: number of days in a given year smoking cigarettes

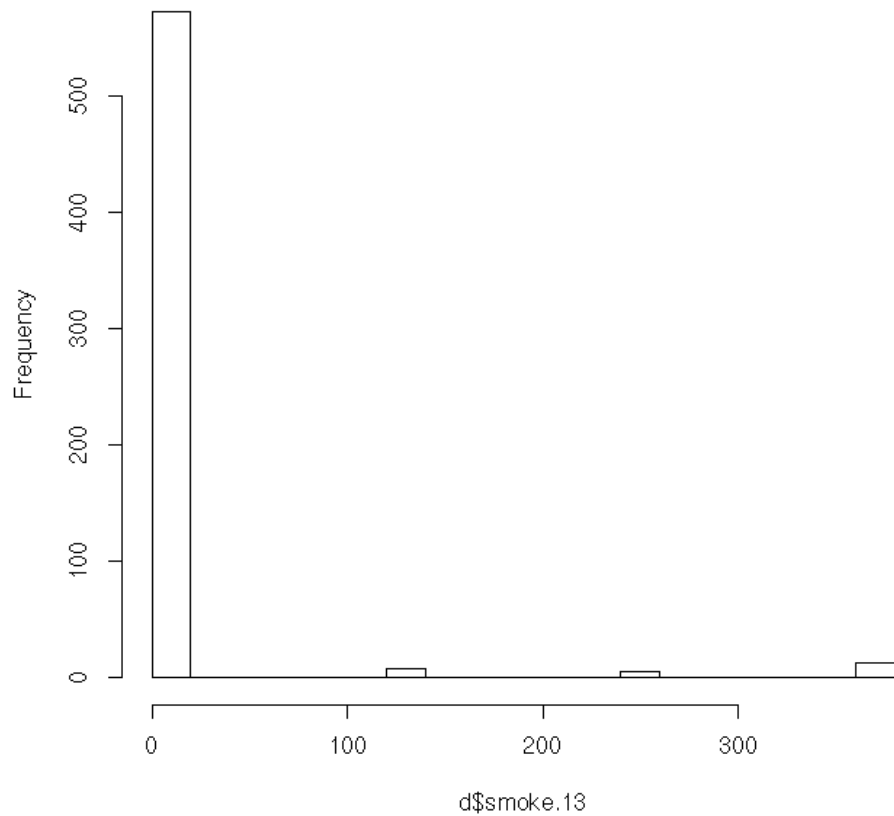
Age	Min	1st Q	Median	3rd Q	Max	Mean (SD)	NA
13	0.00	0.00	0.00	0.00	365.00	11.77 (59.12)	652
17	0.00	0.00	0.00	1.50	365.00	67.61 (128.57)	701
21	0.00	0.00	1.50	250.00	365.00	116.05 (149.80)	946
25	0.00	0.00	1.50	365.00	365.00	127.04 (1159.31)	1009

- Coding manual
  - 0=never
  - 1=once or twice
  - 2=occasionally but not regularly
  - 3=regularly for a while, but not now
  - 4=regularly now
  - 8=N/A
  - 9=missing
- Data
  - 1.5, 125, 250, 365, unlabeled

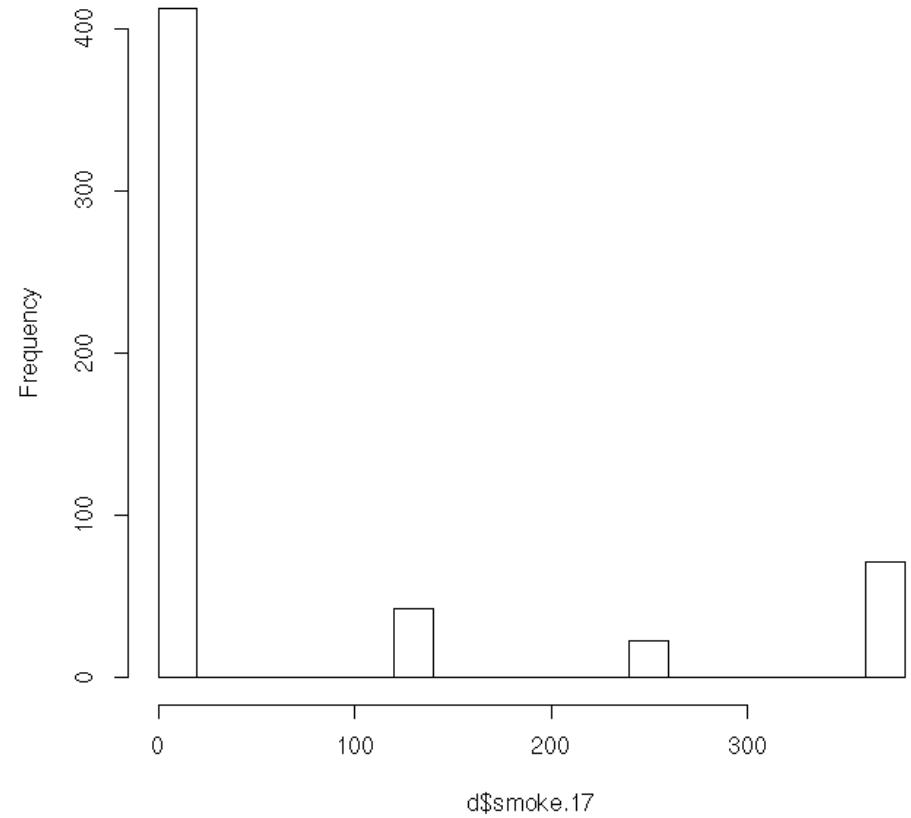


# smoke: number of days in a given year smoking cigarettes

Histogram of d\$smoke.13

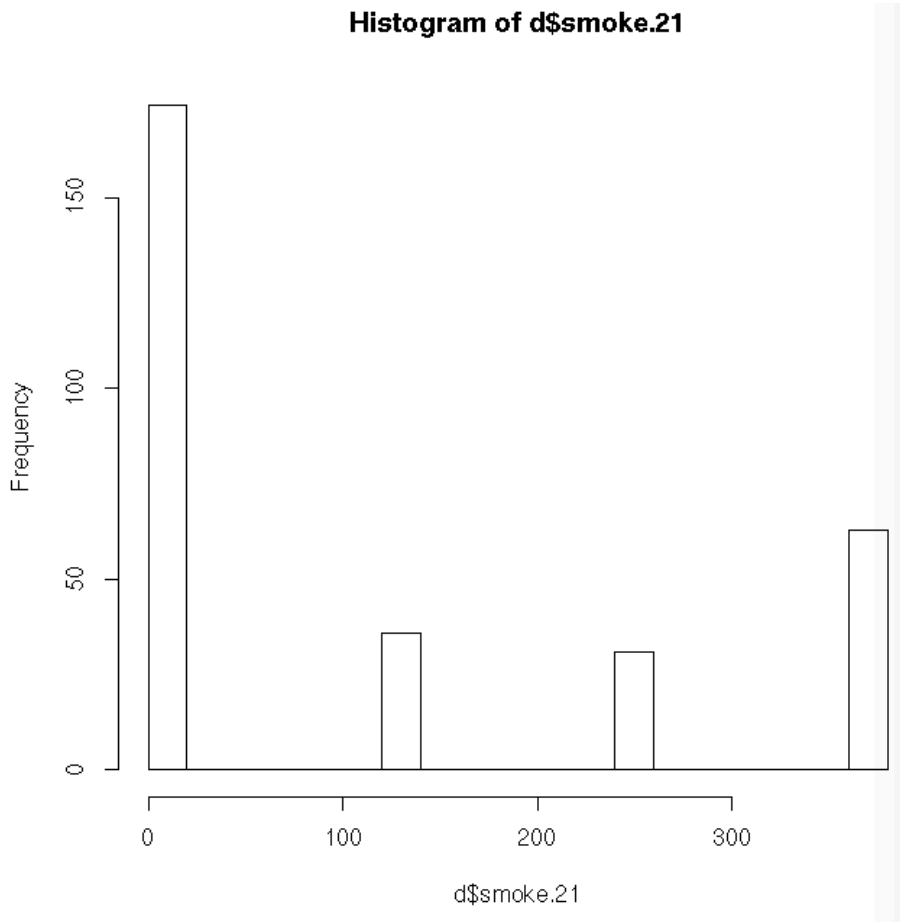


Histogram of d\$smoke.17

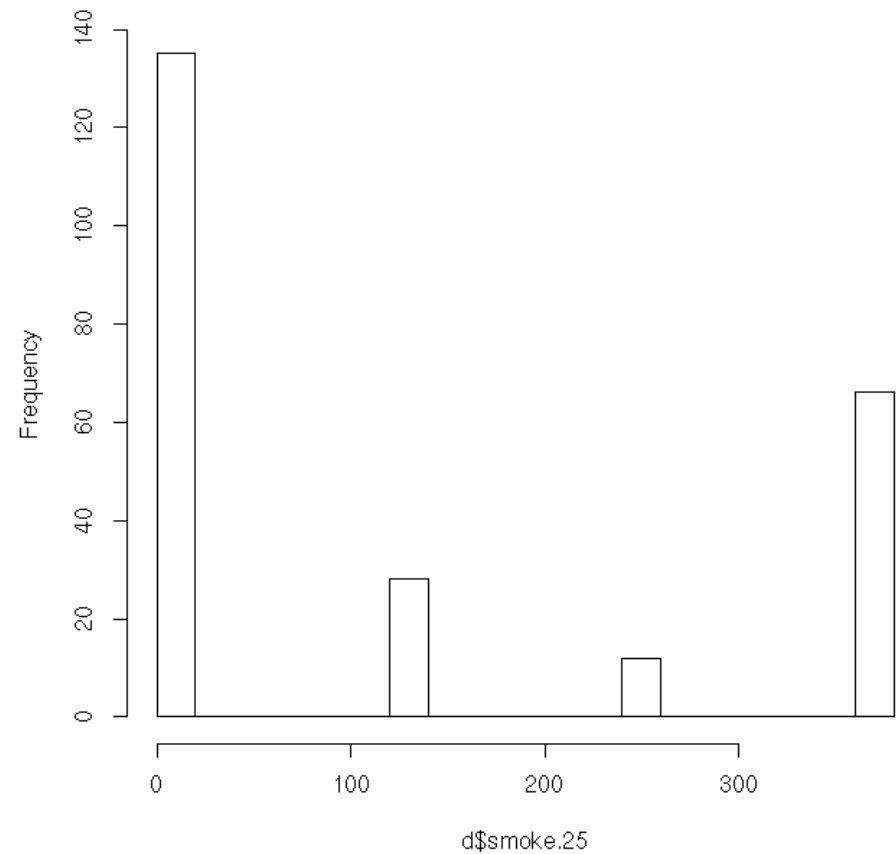


# smoke: number of days in a given year smoking cigarettes

Histogram of d\$smoke.21



Histogram of d\$smoke.25



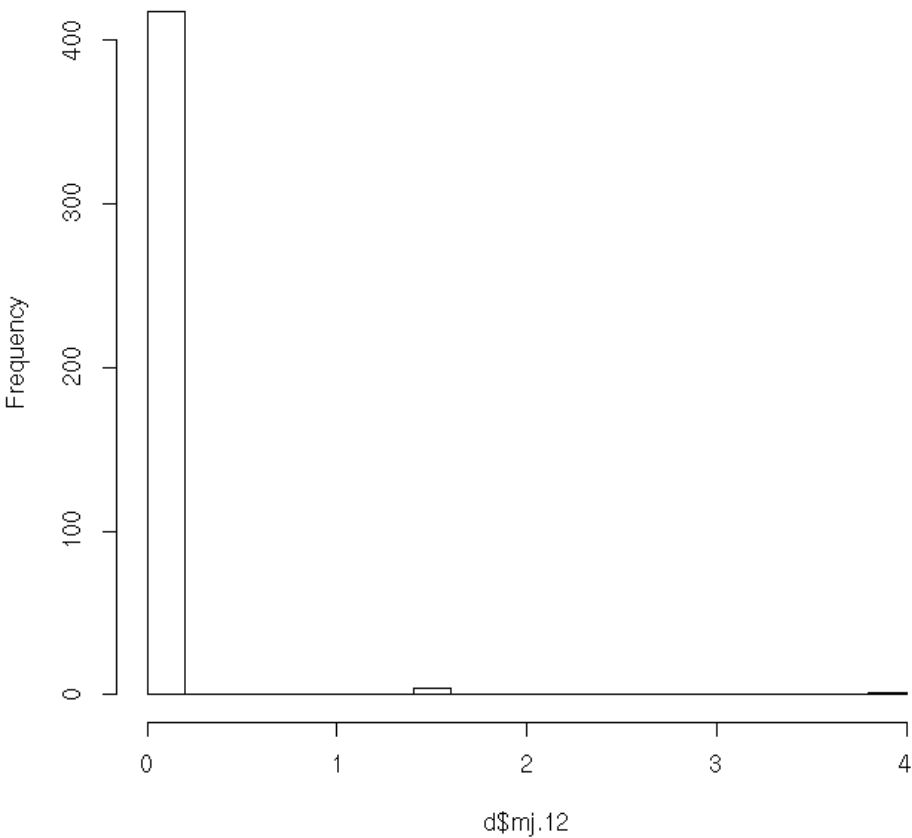
# mj: number of days in a year using marijuana

Age	Min	1st Q	Median	3rd Q	Max	Mean (SD)	NA
12	0.00	0.00	0.00	0.00	4.00	.02 (.24)	828
19	0.00	0.00	0.00	14.50	365.00	40.22 (93.83)	967
21	0.00	0.00	0.00	11.00	365.00	43.28 (102.56)	947
25	0.00	0.00	1.50	14.50	365.00	48.16 (109.06)	1009

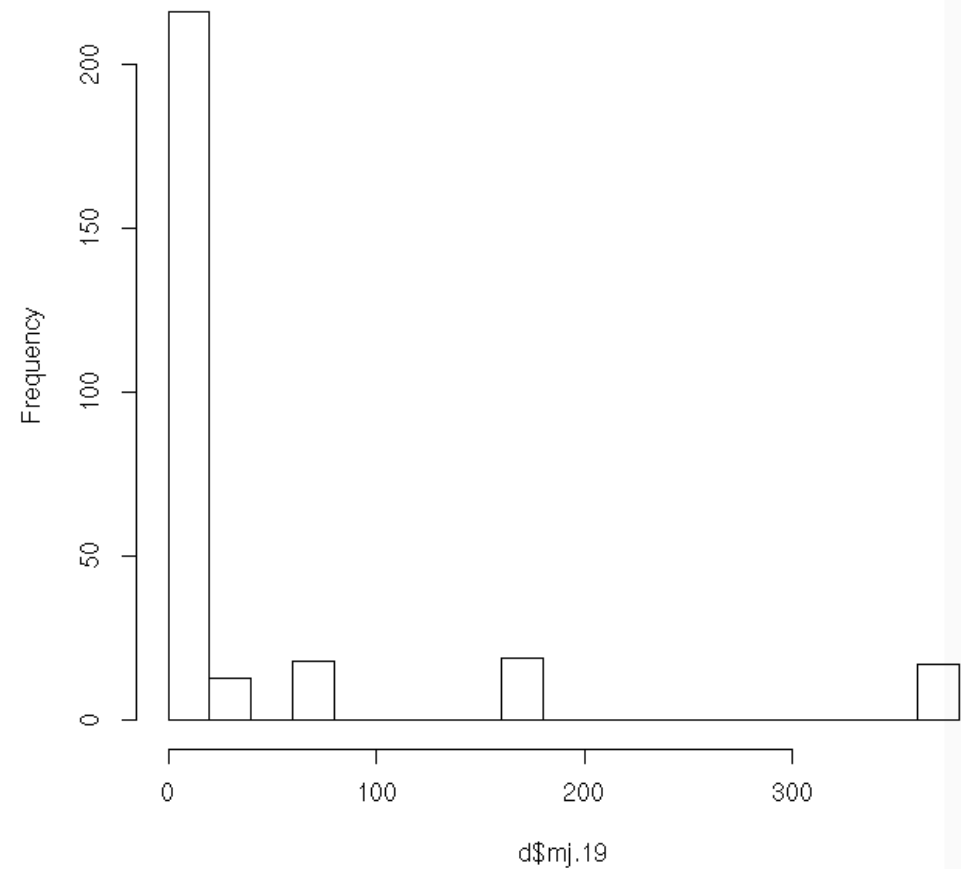
- Coding manual
  - 0=never (0 occasions)
  - 1=1-2 occasions
  - 2=3-5 occasions
  - 3=6-9 occasions
  - 4=10-19 occasions
  - 5=20-39 occasions
  - 6=40-99 occasions
  - 7=100-249 occasions
  - 8=250-499 occasions
  - 9=500 & above
- Data
  - 0, 1.5, 4, 7.5, 14.5, 24.5, 69.5, 174.5, 365, unlabeled

# mj: number of days in a year using marijuana

Histogram of d\$mj.12

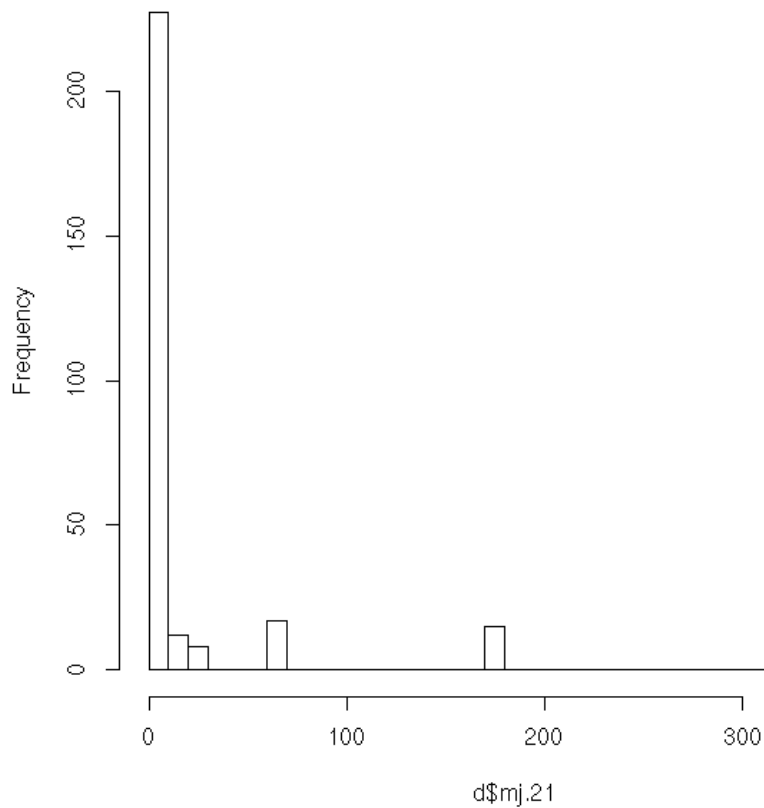


Histogram of d\$mj.19

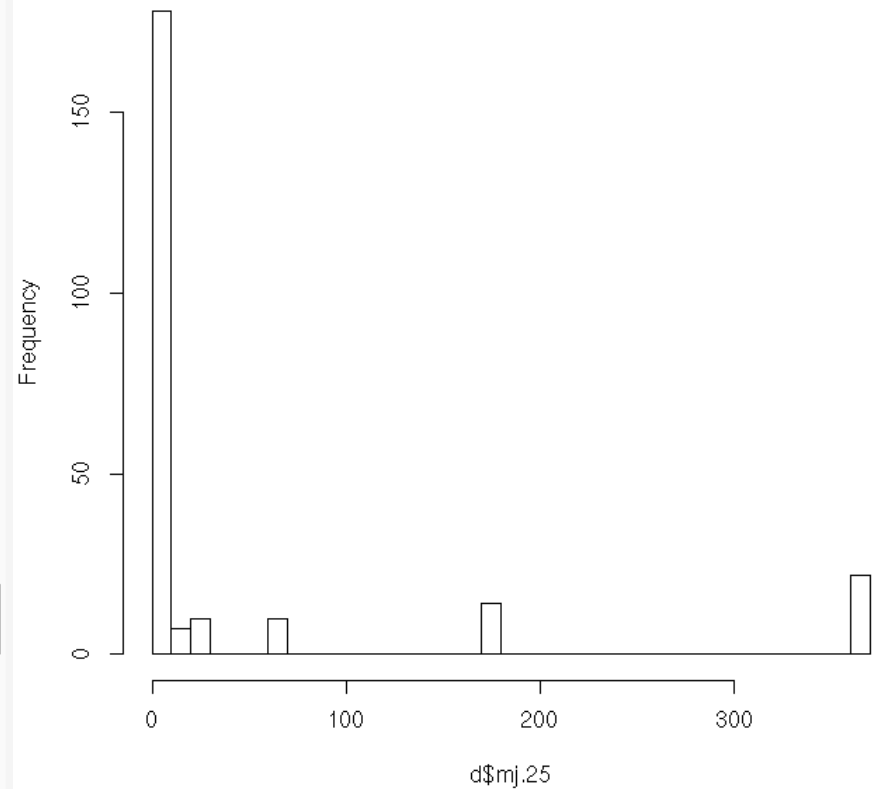


# mj: number of days in a year using marijuana

Histogram of d\$mj.21



Histogram of d\$mj.25



# To do list

- Tuesday I received more documentation that I still need to fully go through, I went through focusing on variables currently available
  - Orientation to MLS data
  - Pedigree File
  - Annual substance use measures, both adolescent and adult
  - Codebooks for data files
- Is there information on which families came from which group?
- How does this mesh with other data sets we have?
  - What other variables should I seek out? Ideas up next
- Reshape data into long format
  - coming after I acquire more variables of interest
- Modeling trajectories of use, within individuals or globally

# Potential variables of interest sourced from codebooks

- drinking days per month
- drinks per day on typical drinking occasion
- alcohol quantity/frequency broken down by type (wine, beer, liquor)
- cigarettes per day
- nicotine dependence