Project Preliminary Report

Matthew Blake and Donovan Roddy

Healthy sexual relationships and their determinants have been a nuanced and ongoing question among public health professionals. Variation in the number of sexual partners between genders, ages, and within different regions has been a topic of particular interest, as sociologists and sexual health professionals explore trends in sexual relationships over time. Exploring the role education plays in sexual health has also been a frequent point of discussion. This project explores whether there is an association between one's high school academic performance and both their self-reported sexual partners over time and their self-reported frequency of sexual encounters. The data used in this study comes from the National Longitudinal Study of Youth from 1997 (NLSY97). The survey looks at 9884 people, born between 1980-1984. The data was collected through yearly interviews, from 1997 until 2008. The interviews collected information about employment, demographics, sexual activity, dating and marriage, family, health trends, and various other attributes. This study focuses on information related to age when the respondent experiences puberty, age of first sex encounter, race, ethnicity, and sex identify. These measures have been reported to be associated with sexual partners (Kan et al. 2010). Our measure of academic performance is one's average GPA, and our dependent variables of interest are a respondent's reported number of sexual encounters since the last interview and a respondent's reported number of sexual partners since the last interview.

Kan, Marni L., Yen-hsin Alice Cheng, Nancy S. Landale, and Susan M. McHale. 2010. "Longitudinal Predictors of Change in Number of Sexual Partners across Adolescence and Early Adulthood." Journal of Adolescent Health 46 (1): 25–31. https://doi.org/10.1016/j.jadohealth.2009.05.002.

Variable Descriptions:

- year

  - Year that response was recorded (AD)
  - Level 1 variable and observational unit
  - Collinear with Age and year of birth, so not used in model

- PUBID

  - Unique ID for each individual (no units)
  - Level 2 variable and observation unit

- Avg_GPA

  - Average GPA among 7 different high schools surveyed (some attended multiple of the high schools)
  - No units
  - Level 2 predictor/independent variable

- Race

  - Race of respondent
  - AAPI, American Indian, Black, Other, or White

- Level 2 variable

- Ethnicity

  - Ethnicity of respondent
  - Reduced to "Hispanic" or "Not Hispanic" to reduce collinearity with "Race"
  - Level 2 variable

- Sex

  - Sex of respondent
  - Male or Female
  - Level 2 variable

- Puberty_Age

  - Age at which individual started puberty
  - Unit for age is years since birth
  - Level 2 variable

- na.prop

  - Proportion of responses left blank
  - Used to partially account for nonrepresentativity of more filled in responses
  - Level 2 variable

- Region

  - Region of the US in which respondent lives
  - US Census region: Northeast, North Central, South, or West
  - Level 1 variable

- Rural_Urban

  - Whether the respondent lives in a rural or urban area
  - Binary variable: Rural or Urban
  - Level 1 variable

- Marriage_Status

  - Whether the respondent is cohabitating, or has ever been married
  - Cohabitating, Never Married, Married, Separated, Divorced, or Widowed
  - Level 1 variable

- SEX_SINCE_DLI

  - Number of times respondent has had sex since Date of Last Interview
  - Level 1 response / dependent variable

- NUM_PARTNERS_EVER

  - Number of sexual partners respondent has ever had
  - Level 1 response / dependent variable

- NUM_PARTNERS_SINCE_DLI

  - Number of sexual partners since Date of Last Interview
 - Level 1 response / dependent variable

Data Cleaning Description

The data was originally collected in a wide format, with the observational unit being the respondent, and the columns indicated the question and the year in which it was asked. The original data listed the columns as numbers related to each question. The columns were renamed to better represent their measures, and the data was pivoted longer so that each row represented a respondent and their answers to the questions each year. This will allow us to do multilevel modeling. Only about one-third of the respondents had data recorded related to their sexual partners, so the other respondents were dropped. One of the challenges of this data set is that the number of sexual partners and number of sexual encounters did not have yearly responses for each respondent, listing NAs for years in which no response was recorded. To account for this, the data was interpolated for each respondent since in theory these numbers would not fluctuate between interview times as the question asked for the count since that last interview. We plan to do an offset for modeling, but this was what we did for the convenience of our exploratory analysis. In addition, for the level 2 variables, such as average GPA or race and ethnicity, there were also NA values for certain years. We also interpolated these values as they do not change over time. Finally, there were some significant outliers in reports of the number of sexual partners. A couple of respondents reported around 1000, while the majority were between 0 and 50. To account for this, we simply removed respondents with extreme values.

EDA Graphs

Sex Frequency versus High School GPA
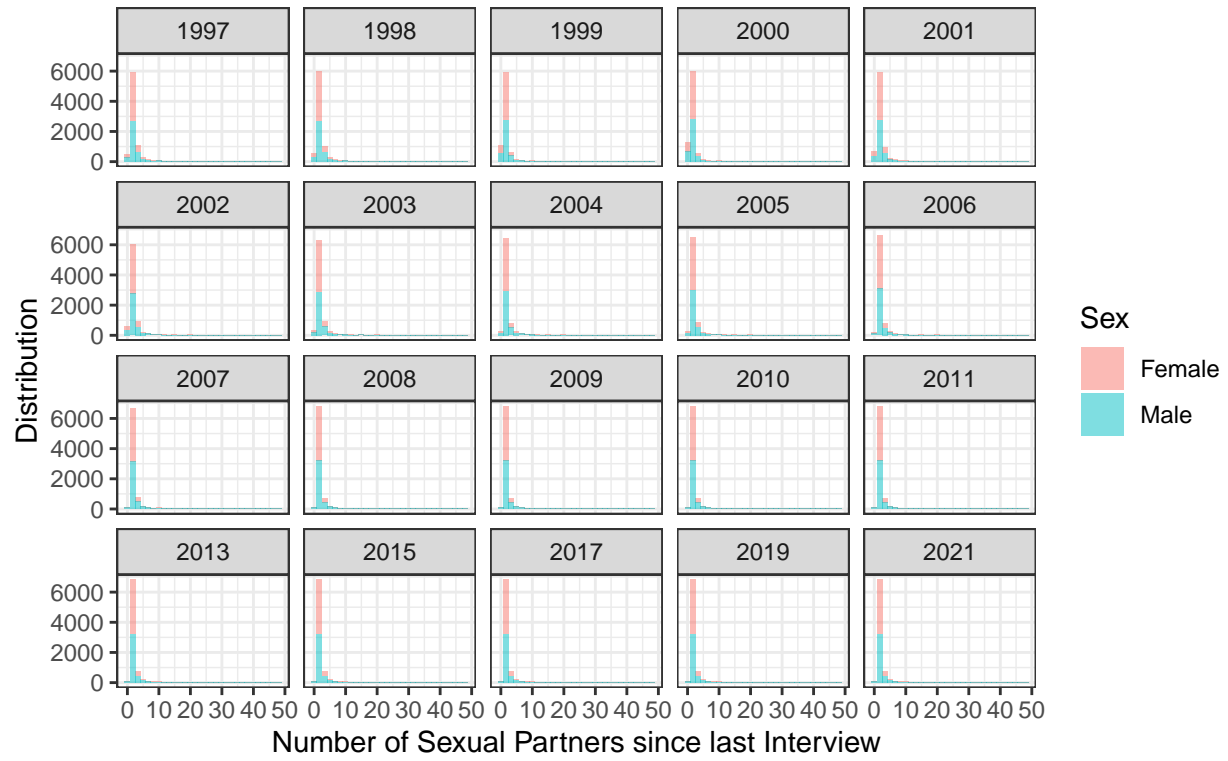
Amount of Sex since Last Interview vs Average GPA
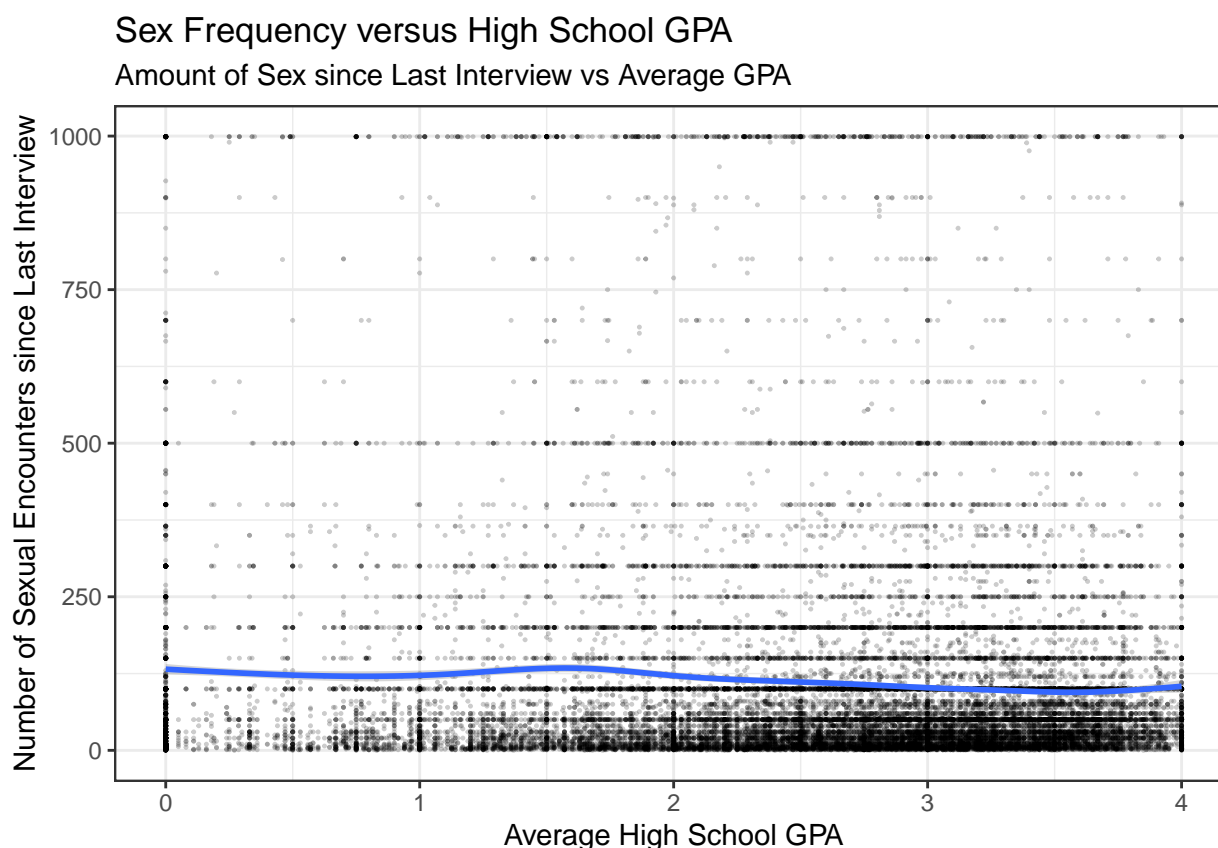
Number of Sexual Partners by GPA Achievement

Trends in Sexual Partners for those with High versus Low GPAs
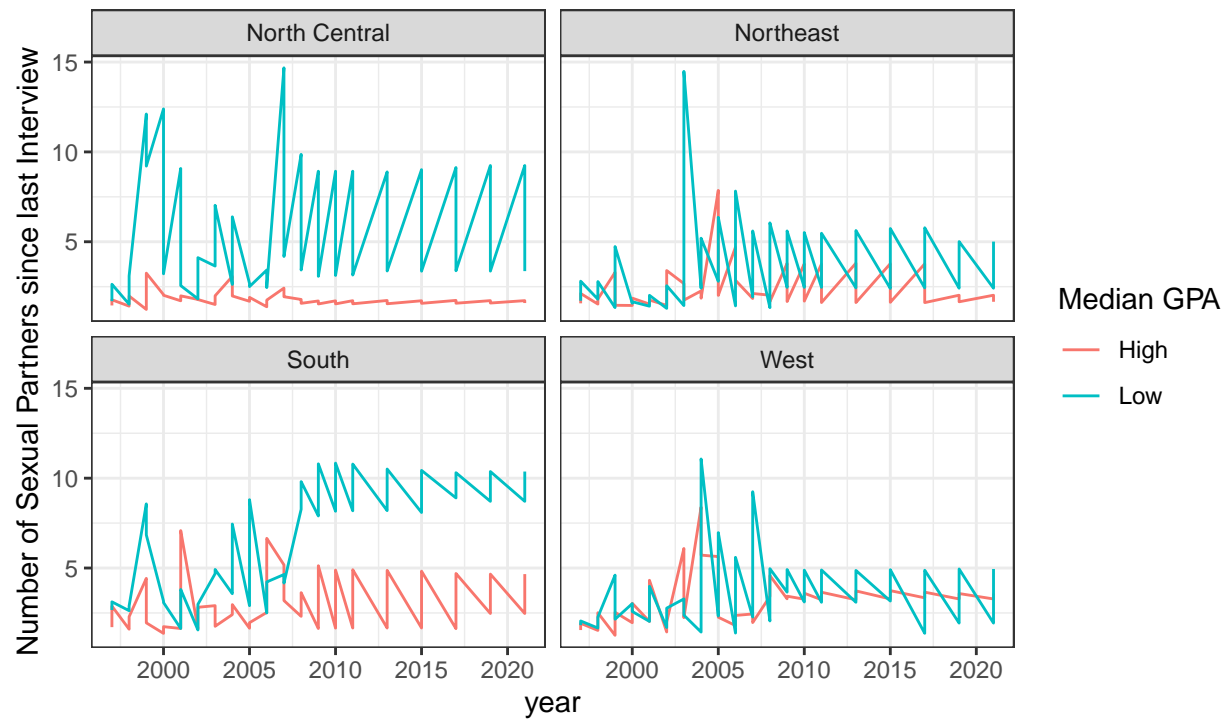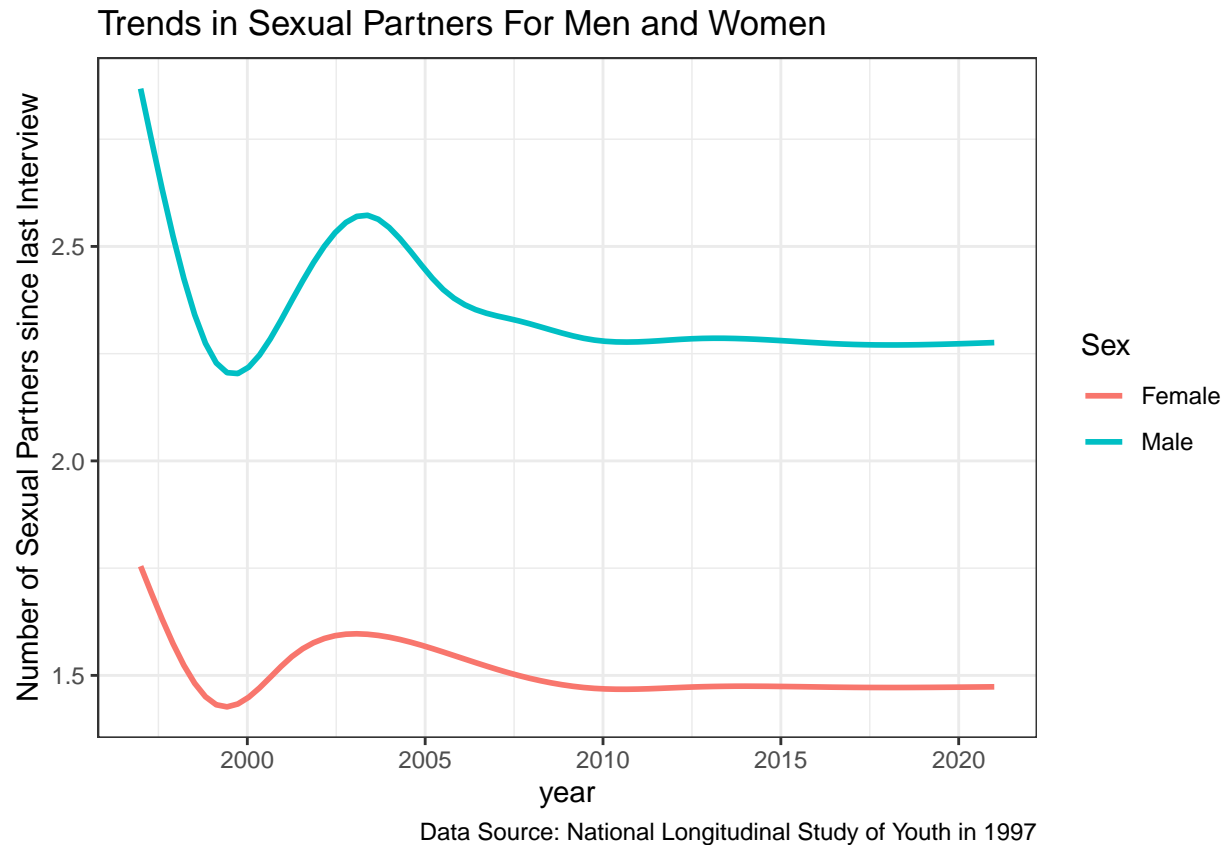
Data Source: National Longitudinal Study of Youth in 1999

Distribtuions of Sexual Partners Each Year

Data Source: National Longitudinal Study of Youth in 1997

Trends in Sexual Partners For Men and Women

Data Source: National Longitudinal Study of Youth in 1997

This chart shows the SEX_SINCE_DLI versus the Average GPA. We see that there appears to be a slight negative correlation, from the linear fit line in red. However, the geom_smooth line reveals that for low SSD numbers, where we have the most data, there is a notable *positive* correlation between SSD and Avg_GPA, which levels out around 25.

Sex since DLI by sex (M/F)
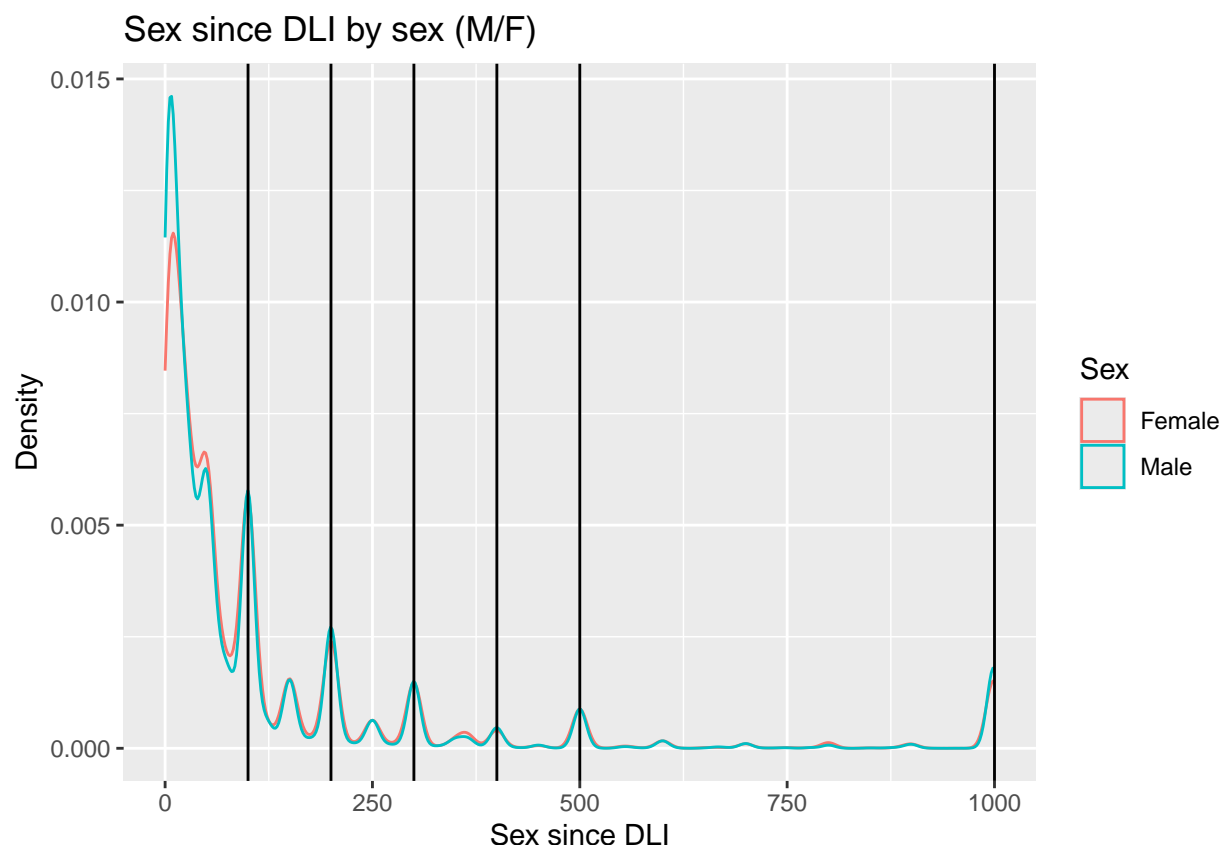
This is a density chart of Sex since DLI, broken up by sex (male and female). Vertical lines are included at multiples of 100, to show that the higher values are disproportionately around these round numbers.

Figure Summaries.

The first figure indicates the relationship between the average GPA and the number of sexual encounters. The data suggests that initial increases in the number of sexual encounters seem to be had by those with higher GPAs, followed by a relatively steady decline. It generally suggests that those with high numbers of self-reported sexual encounters had lower GPAs. Figure two looks at the number of sexual partners. It suggests that generally, those with lower GPAs tend to have more sexual partners, but the trends in time seem to vary by region. In addition, the trend in time for those with high GPAs seems to be more consistent. Figures 3 look at the distribution of sexual partners each year separated by sex. They seem to suggest that men seem to have more sexual partners on average than women. They also seems to suggest that of the indicating having one sexual partner seems to be indicated significantly more than all all other numbers. In addition, Figure 4 suggests that men seem to have more partners on average than women, and this trend appears consistent over time. The final graph indicates the distribution of sexual encounters for both men and women. The plot suggests that men and women seem to round when expressing the number of sexual encounters they have had, as there are peaks at even numbers ending in zero. They tend to select numbers like 250, 500, and 1000.

Demographic Summaries

Looking at the number of sexual encounters over time, we find slight differences in most of the demographic characteristics. Women have more moderate amounts of sex, whereas men are more represented at the two extremes. Black Americans are most highly represented among people with few or no sexual encounters,

while white Americans are overrepresented among those with many. Hispanic Americans, however, have slightly fewer sexual encounters on average. The West has the most sexual encounters, while the South has the fewest, on average; urban people have slightly more sex than rural people, though not significantly. Those who are not married and not cohabitating have by far the least sex, while those who are unmarried and cohabitating have the most. Regarding number of sexual partners, this also seems to vary significantly depending on marriage type. Separated and divorced people seem to have higher number of sexual partners, while those who live with others seem to have less. In addition, those who went through puberty later seemed to have more sexual partners.

For more information on the code of our analysis, the Read Me found in our project Folder describes what each of the different Rmd files are intended for. It goes over our data cleaning code an exploratory analysis code.

Next Steps:

- Determine what variables, we want to include in our models.

- Determine what types of multilevel models we want to use.

Additional Questions

- We want to know if, specifically for number of sexual partners, our dependent variable should be made Binary. Where 0 indicates if a respondent has 1 or fewer sexual partners and 1 indicates if a respondent has more than 1 sexual partner.

- How do we account for the rounding challenges in the self-reported sexual encounters?

Sources

Kan, Marni L., Yen-hsin Alice Cheng, Nancy S. Landale, and Susan M. McHale. 2010. "Longitudinal Predictors of Change in Number of Sexual Partners across Adolescence and Early Adulthood." Journal of Adolescent Health 46 (1): 25–31. https://doi.org/10.1016/j.jadohealth.2009.05.002.

Our data was collected from the National Longitudinal Survey of Youth– 1997. This survey tracked individuals born from 1980 to 1984. https://www.nlsinfo.org/content/cohorts/nlsy97