Sonia Kopel, Amber Widmer, and Christopher Mullinax, *College of Charleston*

A Mathematical Analysis of Unplanned Pregnancies in the United States

[[1]](#footnote-1)

*Abstract*—The issue of unplanned pregnancy is among the more divisive in the United States. The mission of this project is to further investigate existing data sets to gain a new insight into this problem. Specifically, we focused on targeting the causes of unplanned pregnancy and identify at risk groups in order to identify potential interventions that may result in fewer unplanned pregnancies nationwide. We also use survey data to study how these at risk groups feel about current public health policy. Analysis of twitter data showed that public sentiment regarding terminating unwanted pregnancies was strongly negative, which led us to seek interventions to lower unwanted conceptions rather than just unwanted births.

*Index Terms*—Unplanned pregnancy, public health, sexual education

# Introduction

U

NPLANNED pregnancy has the potential to ruin a woman’s financial stability, educational goals, and her personal relationships. Nearly 50% of all pregnancies carried to term in the United States are unintended [5]. Beyond the impact to the women, an unintended pregnancy could have health impacts on the baby and economic impacts on the country.

\*\*\*We may want to exclude this \*\*\* In order to gain a basic level of insight into the historic trends of unplanned pregnancy in the United States, we examined a data set that summarized the number of unmarried births in the United States from 1940 to 2015 [1]. The trends are summarized in Figure 1. However, the assumption that all or even most of these unmarried births are unintentional is likely to be faulty. However,

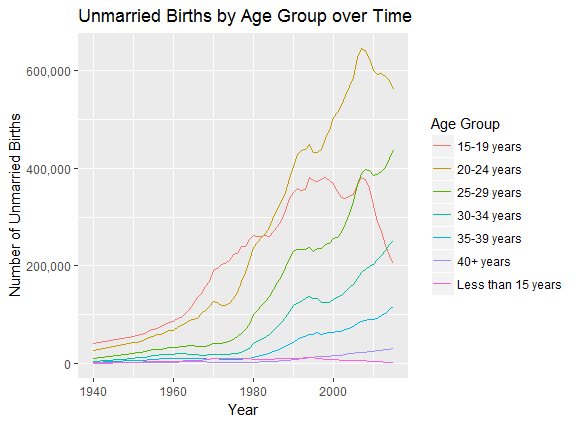


Figure 1: Trends in unmarried births by age group over time

Here is some background information. Here are some graphs that illustrate trends over time. We probably want a bunch of citations in this section. Heat map of where the unintended pregnancies are. Compare to other relative heatmaps. Look at rate of unintended over time.

# Data

This study analyzed public data gathered from online sources.

## Pregnancy Risk Assessment Monitoring System (PRAMS)

The PRAMS data is a surveillance system of the CDC which collects population-based based data from participating states reflecting maternal attitudes before, during, and shortly after pregnancy. The surveillance currently covers about 83% of all U.S. births. The surveillance includes a questionnaire sent to a random sample selected from birth certificates issued in participating states.

The results of this questionnaire, specifically the pregnancy questionnaire, were used for this study. The years 2009-2011 were included, which are the three years for Phase 6 of the survey. The nature of the data is population-based, as mentioned, and thus includes information about the number of respondents in a subgroup which answered a question in a particular way. There are questions pertaining to the intendedness of the pregnancy and well as many socioeconomic characteristics of the population, so we were able to use this data to identify key differences between the intended and unintended pregnancy populations.

As the sample is selected from birth certificates, this data does not contain information on unintended pregnancies which were terminated or resulted in fetal loss.

## Kaiser 1998

## Kaiser 2017

More political related polls regarding Medicaid

# Methods

. The analysis was for the most part exploratory and varied with the format of each analyzed data set. We used a number of mathematical techniques to model the

## Modelling Birth Control Efficacy with Markov Chains

We used a mathematical technique called a Markov Chain to model the probability of various birth control methods failing after a certain number of years. This was done to test the assumption that if one truly wishes to avoid having an unplanned pregnancy, the only truly “safe” bet is abstinence.

A Markov Chain shows how transition probabilities change over time and eventually converge. In order for the method to work, we must have a stochastic matrix. That is, the sum of all entries in any row must equal one.

We simulated a number of environments with varying assumptions. In each environment we tested the same methods of birth control: withdrawal, family planning (also known as the calendar method), intrauterine devices (IUD), condoms, injections, the pill, and implants. For all of these methods, the probability of having the birth control method fail within a year was provided by Planned Parenthood [2]. Those probabilities account for user error, so they were selected over the theoretical efficacy of a given method to simulate a more realistic environment. However, Planned Parenthood only provided two decimals of accuracy, and rounded down the efficacies of the implant and the IUD to 0.99. Because of the 100% efficacy of abstinence, this method was not included in the Markov Chain. The time step for all Markov Chains was one year.

Once the transition matrices were created, the matrices were raised to higher and higher powers until they eventually converged. However, as the converge occurs far beyond the fertile period or even life span of most women, this study was more interested in the intermediate states of the matrices particularly after five and 30 time steps. Thus, if **T** is a transition matrix, **T**5 would show how the probabilities changed after five time steps.

## Twitter Sentiment Analysis

Do we even want to include this??

Could be good for establishing public sentiment toward abortion 🡪 want to focus more on prevention

## Chi Squared Testing

In all of the data sets that were used in the analysis, the majority of the variables were nominal (categorical) in nature with two or more possible values. We tested the variables for independence for questions that were considered to be interesting. The G-test of independence was used to test whether the proportions of one variable are different for different values of a second variable. The G-test of independence and the Chi-square test of independence provide almost identical results. The biggest difference is that the G-values are additive and can be used for more elaborate statistical designs.

Numerous variables were tested in all of the data sets. Those that produced a significant result were then investigated further. To do this, an odds ratio was calculated to describe the magnitude of the difference. For example, in the PRAMS data set we were looking for potential predictors of unplanned pregnancies. There was a noticeable difference between women who exercised three or more time per week versus those who did not exercise regularly. The odds ratio for the two groups was calculated to be 0.69. The interpretation of this value indicated that women who exercised three or more times per week were 31% less likely to have an unintended pregnancy. The results for the odds ratio comparisons could lead to potential paths for further research.

### Two Proportion Tests

For the PRAMS data, we performed a chi-square test to determine whether there were significant differences between the responses for each question in the Pregnancy survey for the group of respondents who said their pregnancy was unintended and those whose pregnancy was intended. This analysis covered only the questions with binary responses (180 of the 220 questions in the survey).

### Tests for Association

## Variable Correlations

Correlations were used to determine the strengths of linear relationships between two variables in multiple data sets. They were performed on entire data sets to gain a top level understanding of some of the trends within the data as well as on individual questions to gain a detailed understanding of what variables were correlated with the proportions of certain responses.

These investigations were all done in R using default and public libraries. Any given correlation value ranges from -1 to 1 with values close to 0 representing weaker correlations (little no linear relationship), values close to 1 representing strong positive correlations (as one variable increases, so does the other), and values close to -1 representing strong negative correlations (as one variable increases, the other decreases).

### Full Data Correlations

In order to find questions with correlated responses, numerical weights were assigned to the qualitative responses in the Kaiser surveys. For example, <Strongly Disagree, Somewhat Disagree, Don’t Know, Somewhat Agree, Strongly Agree> was translated to <-10, -8, 0, 8, 10>. For responses that covered ranges of values, such as for Age Groups and Income Brackets, the median was assigned. For Income, this was reduced down to the same order of magnitude. With a numerical interpretation of each response, correlations were calculated for each pair of questions/variables in the data set.

The visualizations that summarize these findings were created using the “corrplot” library in R.

### Answer Level Correlations

Correlations were also done to show the strength of the linear relationships between grouping variables and responses to particular questions. For example, in the PRAMS data one question asked if the respondent’s pregnancy was unplanned with binary responses “yes” and “no.” We analyzed how those responses were affected by the grouping variable “income”. The incomes of respondents were broken up into several categorical income brackets. All of the brackets were converted into numeric values that approximated their respective central values. For each answer to the question, the correlation between the numeric incomes and the proportion of responses that gave the answer was calculated and analyzed.

In the PRAMS data set, the <Less than $10,000, $10,000 to $24,999, $25,000 to $49,999, $50,000 and above> were coerced to numeric values <5000, 12500 , 37500, 99000>. In the 2017 Kaiser survey, the income brackets <Less than $20,000, $20,000 to less than $30,000, $30,000 to less than $40,000, $40,000 to less than $50,000, $50,000 to less than $75,000, $75,000 to less than $90,000, $90,000 to less than $100,000, $100,000 or more> were coerced to <10000, 25000, 35000, 45000, 62500, 100000, 100000>. The bracket “$90,000 to less than $100,000” was grouped with the “$100,000 or more” category to reduce sparsity and offset the effect of a heavy right skew.

## Association Learning

In order to gain more insight into the person-level data, we ran a number of association learners from the RWeka package. The rules output from the association learner provided insight into how individuals were distributed across multiple groups. The rules also confirmed some of the findings from the correlation plots.

These rules were of the following form:

Σ (Questioni = Answerj) 🡪 Questionx = Answery

Which roughly translates to the following: if an individual responded to certain questions with particular answers, then she would likely answer Questionx with Answery.

# Results

Table 1: Transition matrix for Markov Chain with absorbing pregnancy state and random switching after each step

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Withdraw** | **FAM** | **IUD** | **Condom** | **Injection** | **Pill** | **Implant** | **Pregnant** |
| **Withdraw** | 0.062 | 0.016 | 0.125 | 0.187 | 0.055 | 0.320 | 0.016 | 0.220 |
| **FAM** | 0.061 | 0.015 | 0.122 | 0.182 | 0.053 | 0.312 | 0.015 | 0.240 |
| **IUD** | 0.079 | 0.02 | 0.158 | 0.238 | 0.069 | 0.406 | 0.02 | 0.010 |
| **Condom** | 0.068 | 0.017 | 0.136 | 0.204 | 0.060 | 0.349 | 0.017 | 0.150 |
| **Injection** | 0.075 | 0.019 | 0.15 | 0.226 | 0.066 | 0.385 | 0.019 | 0.060 |
| **Pill** | 0.073 | 0.018 | 0.146 | 0.218 | 0.064 | 0.373 | 0.018 | 0.090 |
| **Implant** | 0.079 | 0.020 | 0.158 | 0.238 | 0.069 | 0.406 | 0.020 | 0.010 |
| **Pregnant** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 |

## Modelling Birth Control Efficacy with Markov Chains

The first simulation (Sim. 1) assumed pregnancy to be an absorbing state (meaning that once an individual becomes pregnant she is removed from the population). It was also assumed that after each step in the Markov Chain a woman would randomly switch her birth control. The probability of switching to any given birth control method (including the method that the woman was using at the current time step) was proportional to the percentage of Americans who used that particular method of birth control. Those probabilities were calculated from statistics provided by the Guttmacher Institute [3]. The probabilities were scaled to account for the probability of being put in the absorbing pregnant state. The transition matrix is given by Table 1.

Under these assumptions, after five time steps (five years), the probability of having an unplanned pregnancy ranged from 0.354 to 0.504. Within 30 years, approximately the average length of fertility for a woman, the probability of unplanned pregnancy was over 0.95 in all rows. Thus, under these assumptions, most sexually active women would experience an unplanned pregnancy in their lifetimes.

However, to more closely simulate reality, we ran an additional trial (Sim. 2) that biased the random switching greatly to keeping the birth control method from the previous time step. For all methods, it was assumed that if the woman did not become pregnant, the probability to keeping the same birth control method was 0.99. The remaining probabilities were maintained to reflect their relative popularity, but additionally scaled to keep each row sum equal to one.

In this simulation, after five time steps, the pregnancy probabilities varied greatly with the lowest being at 0.057 for the implant (with the IUD at 0.058) and the highest being at 0.741 for the family planning method. After 30 time steps, the probabilities of pregnancy ranged from 0.374 (implant) to 0.994 (family planning).

To account for the rounding of birth control failure, we did a final simulation (Sim. 3) in which all assumptions were the same as the previous simulation, but the efficacies of the pill and implant were both assumed to be 0.999. The probability of having an unplanned pregnancy after 30 years is 0.206 and 0.184 respectively.

Finally, to demonstrate the effects of proper use of birth control on unplanned pregnancies, we ran a fourth simulation (Sim. 4) that assumed perfect use for all of the birth control methods. Methods with efficacies that were reported as greater than 99% were assumed to be 99.9%. The probabilities of switching methods after each time step were calculated in the same fashion as in simulations 1 and 2. After five time steps, probability of pregnancy ranged between 0.006 (IUD and implant) and 0.465 (family planning). After 30 time steps, the probabilities ranged between 0.058 (implant) and 0.928 (family planning).

The differences in the resultant probabilities of unplanned pregnancy in Sim. 4 highlight the importance of not only proper use of birth control, but also the importance of selecting a method. Thus, if a woman wishes to prevent pregnancy, it is imperative for her to select a more effective method as, over time, the probability of having an unplanned pregnancy will vary greatly from method to method. Furthermore, the woman must use the selected method properly as this too can greatly impact efficacy as summarized in Table 2. For example, if condoms are used perfectly, the probability of unplanned pregnancy after 30 years is 41.9% which, although high, is significantly less than average use which, after 30 years, is 97.9%.

It is worth noting that the more effective methods of birth control (less than 1% chance of pregnancy after 30 years) such as IUD, injection, pill, and implant, are more difficult to obtain, cost relatively more, and require an appointment with a doctor to obtain which could potentially make them more difficult to access.

Table 2: Summary of Markov Chain intermediate phases for n = 5 and 30 for all simulated environments

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Withdraw** | **FAM** | **IUD** | **Condom** | **Injection** | **Pill** | **Implant** |
| **Sim. 1, n = 5** | 0.491 | 0.504 | 0.354 | 0.446 | 0.387 | 0.406 | 0.354 |
| **Sim. 1, n = 30** | 0.965 | 0.966 | 0.955 | 0.962 | 0.958 | 0.959 | 0.955 |
| **Sim. 2, n = 5** | 0.705 | 0.741 | 0.058 | 0.553 | 0.269 | 0.377 | 0.057 |
| **Sim. 2, n = 30** | 0.993 | 0.994 | 0.391 | 0.981 | 0.851 | 0.930 | 0.374 |
| **Sim. 3, n = 5** | 0.705 | 0.741 | 0.015 | 0.552 | 0.269 | 0.377 | 0.014 |
| **Sim. 3, n = 30** | 0.992 | 0.993 | 0.206 | 0.979 | 0.848 | 0.926 | 0.184 |
| **Sim 4, n = 5** | 0.182 | 0.465 | 0.006 | 0.095 | 0.006 | 0.007 | 0.006 |
| **Sim 4, n = 30** | 0.649 | 0.928 | .0625 | 0.419 | 0.060 | 0.074 | 0.058 |

The consequences of these results is that certain methods of birth control if used consistently are virtually guaranteed to result in many unplanned births given a reasonable amount of time (≤ 30 years). Even for more effective methods, improper use can dramatically increase the probability of having an unwanted conception. Thus, for a woman who wishes to avoid pregnancy, it is very important for her to be using an effective method of birth control, but also to be using it correctly.

## Chi Squared Testing

### Two Proportion Tests

Comparing the differences in responses to the PRAMS questionnaire between the unintended and intended pregnancy populations, there were significant differences in the answers to nearly every question. As the sample size is large, there are several statistically significant results that are not practically significant. As such, the questions which had the largest absolute differences in the proportions of each population which responded a particular way were examined more closely.

It was discovered that there are four main categories in which the unintended and intended populations differ: contraception use, economic, health/care, and personal relationships.

#### Contraception Use

Whereas only 6% of those intending to get pregnant were using birth control, 49% of those who labeled their pregnancy as unintended were using some form of birth control. Among unintended pregnancies, condoms, withdrawal, and Calendar rhythm methods are cited at much higher rates than the general population.

Of those who were not using birth control but still had an intended pregnancy, there may not be enough drive to actively prevent pregnancy as 31% did not necessarily mind getting pregnant. Another common reason given for lack of contraception use that occurs more prominently in the unintended group (P-value ≈ 0) is the belief that the respondent cannot become pregnant. 32% of the respondents with unintended pregnancies believed this to be the case compare to only 20% of intended pregnancies.

#### Economic

In the respondents who termed their pregnancy as unintended, the rates of usage of Medicaid for delivery and being enrolled in the WIC program when surveyed were significantly higher. The usage of the programs are indicators of low income. As these are federally funded programs, this serves as evidence that the economic burden on the country for an unintended pregnancy is greater than for an intended pregnancy.

In addition, when a pregnancy was unintended, 31% of the mothers reported they had difficulty paying bills in the 12 months before delivery compared to 17% of mothers with intended pregnancies.

#### Health/Care

When trying to become pregnant, many women will begin taking vitamins to ensure a healthy system in which to cultivate a baby. For unintended births, this habit of taking vitamins is not established, potentially impacting the health of the child. (maybe add more scientific reasons to take vitamins here) Without having the intent to become pregnant, these women would not have had conversations with their doctors to learn fundamental facts that could help ensure a health pregnancy. For instance, only 67% of women with unintended pregnancies had exposure to information that folic acid could prevent birth defects. 83% of women with intended pregnancies had heard this fact.

In the unintended pregnancies the women entered into prenatal care later than desired, or in some cases not at all, and commonly were not able to get into prenatal care as soon as they desired (56%). If Medicaid is needed to pay for health care during pregnancy, which was found to often be the case for unintended pregnancies, there is a two to four-week period to gain approval on qualification after submitting the required documentation.

#### Personal Relationships

Partner related stressors were twice as likely to be reported in the women with unintended pregnancies (42%) versus intended pregnancies (21%). These women were more likely to argue with their partner more than usual in the 12 months before delivery (34%). This is an added source of stress which of course has many health impacts that could inversely affect the baby. (BE MORE SPECIFIC. FIND SOURCES)

## Impact of Income Bracket on Unplanned Pregnancy

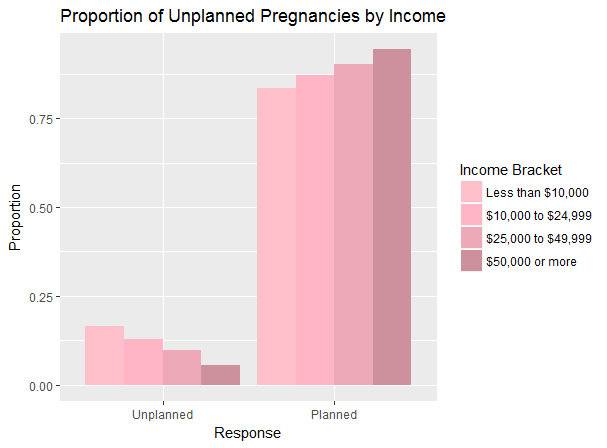
In order to unify disjoint data sets, an initial insight, however obvious, had to be made into unplanned pregnancies. Using the PRAMS data, we were able to identify a strong, negative linear correlation of -0.940 between income bracket and proportion of women who indicated they had an unplanned pregnancy. This means that the proportion of women reporting unplanned pregnancies decreases with increasing income levels. Because there were only two responses and no missing values, the correlation for planned pregnancies was the same magnitude, but with the opposite sign (0.940). To clarify, “proportion” refers to the percentage of women from a given income bracket to answer the question with the given answer. Figure 2 visualizes the data from which these correlations were calculated. 

Figure 2: Bar plot of the proportion of planned and unplanned pregnancies by income bracket from the PRAMS data set

With this linear relationship established, we then looked at the March 2017 Kaiser Health Tracking Poll data to extract which questions had answers that were correlated with income. The data set was subset to only look at female respondents. This survey had more income brackets than the PRAMS data, but because we were looking at linear relationships between income and answers, these additional brackets were kept. A total of four questions were identified. Of those four, three were knowledge based.

The first question was opinion based and only looked at a subset of the data. The data was subset only to individuals who supported a law requiring all private healthcare plans to cover the cost of maternity care. That subset of individuals was then asked whether or not they still supported the requirement for private health plans to cover maternity care even if it meant that some people had to pay for benefits they did not use. The correlation between income and the proportion of women from a given income bracket who now opposed this requirement was -0.841. This correlation is strongly negative and indicates that as income increases, the proportion of women opposing such a requirement decreases. This finding is counterintuitive as lower income women are at greater risk of having an unplanned pregnancy and may need those benefits which they now oppose. Because of missing values, the correlation of income and proportion of respondents is not the same magnitude for those who still support the requirement, but is still similar (0.866). As expected, the correlation is in the opposite direction which indicates a very similar trend. A log-likelihood ratio test confirmed that there were statistically significant differences between the income brackets and the proportions of people indicating they were in support of such a law.

The first of the knowledge based questions asked the respondent if she believed that there was currently a ban on federal Medicaid funds being used for abortions. The correct answer to this question is “Yes, there is currently such a ban.” Among people answering correctly, there was a 0.707 correlation between proportion of respondents from income category and income, meaning that, proportionally, more people from higher income brackets are getting this question wrong. However, the correlation is not as strong among people who answered “No, there is currently no such ban,” only -0.290. This large difference is due to higher numbers of people responding that they did not know the answer to the question. The weaker correlation here suggests there is a weaker linear relationship between believing in the wrong answer and income; however the relationship is still present. Because there were different ways people could fall into the “unknown” categories and because they generally had very few people in them, these “unknown” categories were not analyzed.

The final two questions dealt with Planned Parenthood services. The first asked respondents whether or not Planned Parenthood provided contraception/birth control. The proportions of people incorrectly answering “No, Planned Parenthood does not provide contraception/birth control” were generally lower than for all the other questions (ranging between 0% and 12.1% among the different income levels). However, despite higher proportions of people answering this question correctly, the responses were still strongly correlated with income. The correlation between the proportion of women answering “Yes, Planned Parenthood does provide contraception/birth control” and income bracket was 0.835, meaning greater proportions of women from higher income brackets were answering the question correctly. Similarly the correlation was -0.757 among women answering the question incorrectly. Interestingly, when the entire survey was analyzed (not just subset to women), the correlations were a little stronger for both answers, 0.896 and -0.902 respectively. A log likelihood ratio test confirmed statistically significant differences between the groups.

The last question asked respondents whether Planned Parenthood provided abortions. The responses to this question, summarized by Figure 3, varied greatly. We cannot be sure of the cause of these differences, but the strength of the correlations indicate that there is a strong linear relationship between income and the proportion of women to answer this question correctly as well as incorrectly. The correlation between proportion of women responding correctly to the question and income 0.927 which indicated a very strong linear relationship between income and knowing that Planned Parenthood provides abortions with higher proportions of women from higher income brackets answering correctly. Similarly, the correlation between women answering incorrectly was -0.893 which indicates a very similar trend. Again, a log likelihood ratio test confirmed statistically significant differences between the groups.

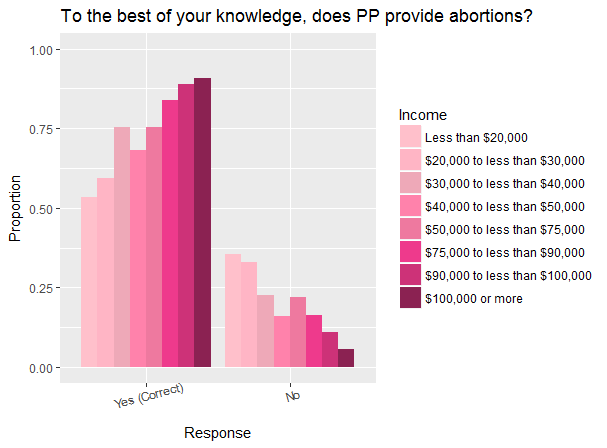
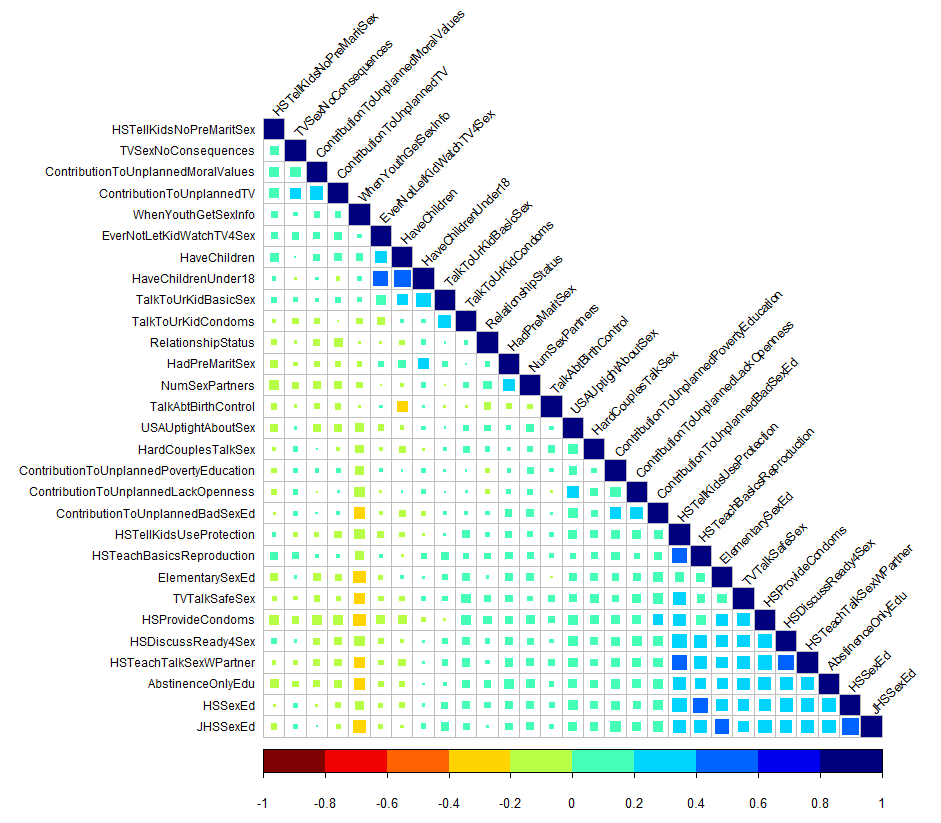


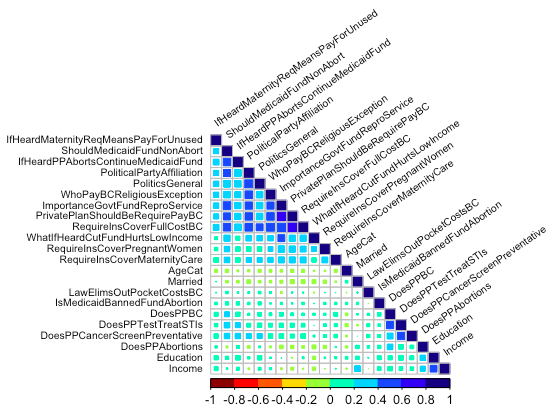
Figure 3: Bar plot of the proportion of women responding either "Yes" or "No" when asked if Planned Parenthood provides abortions grouped by income bracket

To summarize, there exist apparent linear relationships between income bracket and unplanned pregnancy and responses to certain questions to the March 2017 Kaiser Health Tracking Poll. The proportion of women answering knowledge based questions that relate to pregnancy and contraceptive services is shown to be lower when compared to women from higher income brackets. Seeing as women from lower income brackets are more greatly affected by unplanned pregnancy, this sort of information would be of greater value to them, but for one reason or another, they do not seem to have access to it.

## Variable Correlations

PARE DOWN PLOT TO VARIABLES WHICH SHOW HIGHEST CORRELATIONS





# Conclusions

In conclusion, by this time next year, unplanned pregnancies will be a thing of the past.

# Future Work

Check that research journal for future work ideas. I think we had a list of them in there.

# Acknowledgments

This work would not have been possible without the help and support of Dr. Amy Langville, Abhishek Mehta, Brittany Box, and the entire Tresata team. We would also like to acknowledge the 2018 spring Operations Research Class at the College of Charleston for their thoughtful input and advice throughout the course of this project.

Data Sources and References

1. Centers for Disease Control and Prevention. (2018). Births to Unmarried Women by Age Group: United States [Data File]. Retrieved from https://www.healthdata.gov/dataset/nchs-births-unmarried-women-age-group-united-states
2. Planned Parenthood. *All About Birth Control Methods*. Retrieved from https://www.plannedparenthood.org/learn/birth-control
3. Guttmacher Institute. Contraceptive use in the United States. Retrieved from https://www.guttmacher.org/fact-sheet/contraceptive-use-united-states

[4] What is PRAMS? (2017, August 30). Retrieved April 04, 2018, from <https://www.cdc.gov/prams/index.htm>

[5] Centers for Disease Control and Prevention. *Unintended Pregnancy Protection*. Retrieved from https://www.cdc.gov/reproductivehealth/unintendedpregnancy/index.htm

1. [↑](#footnote-ref-1)