Milestone2

Julianne; [Julianne.Ammirati@emory.edu](mailto:Julianne.Ammirati@emory.edu)

March 27, 2017

#### The CDC has allowed open access to the NSFG survey results from 1973-2015:

<https://www.cdc.gov/nchs/nsfg/index.htm>.

#### The code used to generate this report is available from my GitHub Account:

<https://github.com/JulianneA/N741_Milestone1>

# Initial Exploration of Available Data

|  |  |  |  |
| --- | --- | --- | --- |
| Year of NFGS | Number of Respondents | Variables | Notes |
| 2013-2015 | 5699 | 3207 | This was the first file imported as a CSV. |
| 2011-2013 | 5601 | 3094 | Imported as CSV. |
| 2006-2010 | 1.227910^{4} | 3315 | Imported as CSV. |
| 2002 | 7643 | 3087 | Cycle 6, 3 SAS setup files. |
| 1995 | 1.084710^{4} | 5753 | Cycle 5, Imported as CSV. |
| 1988 | 8450 | 1437 | Cycle 4, Imported as CSV. |
| 1982 | 7969 | 742 | Cycle 3, Imported as CSV. |
| 1976 | 8611 | 470 | Cycle 2, Imported as CSV. |
| 1973 | 9797 | 451 | Cycle 1, Imported as CSV. |
| All Respondents | 7.689610^{4} |  |  |
| 1982-2015 | 5.848810^{4} |  |  |

Upon further inspection of code books, it became evident that the 1973 and 1976 files could not be used in analysis. Throughout these cycles, the CDC interviewed only married women or unmarried women who had borne children for the NSFG. From 1982 onward, unmarried and childless women were included in the NSFG, therefore, the initial two cycles of the NSFG were excluded here because the differences in sampling populations introduces too great a threat to data validity and generalizability.

## Data Cleaning Summary

Sixty-three variables were extracted from the NSFG datasets. These variables capture information regarding respondent demographics, contraceptive utilization, sexual education history, and reproductive history. These variables were selected with special consideration for variables present in the most recent three cycles of the NSFG (2006, 2011-2013, and 2013-2015), which are most similar in their data cleaning and accessibility of resources detailing data coding methodology. Two variables were added prior to datamerge: CYCLE and YEAR which both indicate which NSFG survey dataset that respondent had been included.

#### Below is a summary of the sixty-three variables chosen for extraction:

* MAINTENANCE
  + CASEID: Respondent's (R's) Study ID.
  + CMINTVW: Century Month of Interview.
  + INTVWYEAR: Year of Interview.
* DEMOGRAPHICS
  + AGE\_R: Age of R
  + MARSTAT: Marital Status of R
  + HIGRADE: R's current grade in school or highest grade/year attended
  + DEGREES: R's highest college or university degree received
  + PARMARR: Whether R's biological parents married at R's birth
  + ACASILANG: Language to be used in ACASI
  + AGER: Recode of R's Age
  + HISPANIC: Whether R identifies as Hispanic
  + RACE: R's Race
  + CURR\_INS: Current health insurance status
  + METRO: Metro versus Non-Metro as place of residence
  + RELIGION: R's current religions affiliation
  + LABORFOR: R's employment status
  + POVERTY: R's income as a percentage of poverty level
  + TOTINCR: R's total household income
  + PUBASSIS: Whether R currently receives public assistance
* CONTRACEPTIVE USE
  + PILL- Whether R has ever used birth control pills?
  + NORPLANT- Whether R has ever used hormonal implants?
  + CONDOM- Whether R has ever used condoms?
  + VASECTMY- Whether R has ever depended on partner's vasectomy?
  + DIAFRAGM- Whether R has ever used a diaphragm?
  + DEPOPROV- Whether R has ever used depo-provera or other injectable?
  + WIDRAWAL- Whether R has ever used withdrawal?
  + RHYTHM- Whether R has ever used rhythm method?
  + SDAYCBDS- Whether R has ever used Standard Days or CycleBeads?
  + TEMPSAFE- Whether R has ever used natural family planning (safe temperature)?
  + WOCONDOM- Whether R has ever used a women's condem?
  + FOAMALON- Whether R has ever used contraceptive foam?
  + JELCRMAL- Whether R has ever used spermicide jelly or cream?
  + CERVLCAP- Whether R has ever used a cervical cap?
  + SUPPOSIT- Whether R has ever used suppositories?
  + TODAYSPG- Whether R has ever used the Today Sponge?
  + IUD- Whether R has ever had an IUD?
  + LUNELLE- Whether R has ever used Lunelle injection?
  + PATCH- Whether R has ever used contraceptive patch?
  + RING- Whether R has ever used the contracpetive ring?
  + OTHRMETH01- Has R used another method?
  + ECTIMESX- Number of times R used emergency contraception.
  + FIRSMETH1- What was first method R used?
  + CURRMETH1- What is the current method R uses (1)?
  + CURRMETH2- What is the current method R uses (1)?
  + CURRMETH3- What is the current method R uses (1)?
  + CURRMETH4- What is the current method R uses (1)?
* REPRO HEALTH
  + MAYBPREG- Currently, does R believe she could be pregnant?
  + NPREGS\_S- Total number of pregnancies
  + PARITY- Total number of live births
  + ABORTION- Total number of induced abortions
  + HYST- Has R had a hysterectomy?
  + OVAREM- Has R had an ovarectomy?
  + OTHR- Is R surgically sterile due to another operation?
  + PREGNOWQ- Is R currently pregnant?
* SEXUAL HX
  + EVERSEX- Has R ever had hetereosexual intercourse?
  + RHADSEX- Has R ever had hetereosexual intercourse (recoded)?
  + YNOSEX- Main reason R has not had hetereosexual intercourse?
* SEX ED
  + SEDBC- Did R receive formal Sex Ed Before 18: Methods of Birth Control
  + SEDBCG- R's Grade When Received Instruction on Birth Control
  + SEDBCSX- Received Sex Ed on Birth Control Methods Before/After 1st Sex
  + SEDWHBC- Did R receive formal Sex Ed Before 18: Where to Get Birth Control
  + SEDWHBCG- R's Grade When Received Instruction on Where to Get Birth Control
  + SEDWBCSX- Received Sex Ed on Where to get Birth Control Before/After 1st Sex

#### *Challenges:*

This survey has been collected in large cohorts of women over the past 34 years. While providing the user with enormous amounts of data, the CDC has been consistently inconsistent in their variable collection and availability of specific variables. Since its inception in 1973, there has been a seven-fold increase in the number of variables maintained in the dataset, while many of these are recodes and imputed variables, some are useful for further examining reproductive health history or sexual education experiences. Additionally, many contraceptive methods have been developed or fallen out of use in the last 35 years, so the contraceptive use variables have also changed greatly overtime. Many coding schemes have remained consistent across NSFG datasets, allowing the collapse across sets to be more easily handled.

# Preliminary Analyses

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year of NFGS | Age | Highest Grade of Education | % Hispanic | % Black | % White |
| 1982 | 27.26 +/- 8.18 | 12.71 +/- 4.48 | 6.24 | 40.17 | 57.44 |
| 1988 | 29.54 +/- 8.13 | 12.99 +/- 4.14 | 7.59 | 32.79 | 63.36 |
| 1995 | 30.59 +/- 8.31 | 13.03 +/- 2.82 | 14.32 | 23.18 | 71.41 |
| 2002 | 29.5 +/- 8.43 | 13.31 +/- 3.13 | 20.79 | 21.85 | 69.24 |
| 2006-2010 | 28.64 +/- 8.45 | 13.3 +/- 3.25 | 22.18 | 22.17 | 65.82 |
| 2011-2013 | 28.61 +/- 8.38 | 13.46 +/- 3.67 | 26.05 | 23.76 | 63.36 |
| 2013-2015 | 28.88 +/- 8.4 | 13.59 +/- 3.33 | 24.46 | 22.57 | 66.66 |
| All Respondents | 29.08 +/- 8.39 | 13.17 +/- 3.55 | 16.85 | 26.49 | 65.65 |

#### Additional Demographics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Year of NFGS | % Currently Married | % Never Married | % Residing in Metro Areas | % Perfering English/Spanish |
| 1982 | 44.56 | 39.5 | 39.65 | Data Unavailable |
| 1988 | 47.7 | 34.52 | 29.34 | Data Unavailable |
| 1995 | 48.78 | 36.9 | Data Unavailable | 97.09; 2.91 |
| 2002 | 40.3 | 38.55 | 47.23 | 92.59; 7.41 |
| 2006-2010 | 32.34 | 45.58 | Data Unavailable | 92.33; 7.58 |
| 2011-2013 | 30.55 | 45.21 | 41.9 | 94.25; 5.73 |
| 2013-2015 | 29.81 | 46.89 | 41.05 | 93.58; 6.37 |
| All Surveys | 39.9 | 40.72 | 39.06 | 94.03; 5.94 |

#### Participant Financial Indicators

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year of NFGS | % Private Insured | % Medicaid Insured | % Uninsured | % Receiving Public Assistance | % Employeed Full Time |
| 1982 | Data Unavailable | Data Unavailable | Data Unavailable | Data Unavailable | 35.1 |
| 1988 | Data Unavailable | Data Unavailable | Data Unavailable | Data Unavailable | 46.47 |
| 1995 | Data Unavailable | Data Unavailable | Data Unavailable | Data Unavailable | 42.09 |
| 2002 | 63.27 | 12.84 | 6.8 | 10.24 | 43.06 |
| 2006-2010 | 55.41 | 20.07 | 3.66 | 0.35 | 39.67 |
| 2011-2013 | 48.49 | 23.62 | 6.09 | 0.67 | 37.14 |
| 2013-2015 | 52.48 | 26.32 | 4 | 87.99 | 38.78 |
| All Surveys | 55.56 | 20.08 | 4.93 | 1.11 | 40.59 |

***Goal 3: Visualization*** I plan on investing a great deal of time into the visualization step, since I believe that when working with so many variables, the interesting, effective dissemination of results depends on good visualization. This step will be the final week of "processing type" work before I begin writing my manuscript. Therefore, I will have functional visualizations completed before March 1st.

***MILESTONE 2: Working Prototype*** The next milestone is due on March 22nd, and I will spend from March 1st-15th working on this manuscript.

***Manuscript Rewrites*** After discussing my project with a professor to get feedback, I plan to rewrite my manuscript and edit any analysis steps and implement any suggestions to improve the visualization of the data. The analysis changes will occur from March 15th-April1st. Visualization changes will take the next week, and from April 7th to the 15th, I will conduct the final manuscript edits.

***MILESTONE 3: Final Manuscript*** Due Wednesday, April 19th.