# final_project

Emily Phillips

8/1/2021

## Final Project Step 1

### Introduction

The idea and question of fertility has been a topic that I have been researching for some years now, especially being a young woman. It has become more of a prevalent issue, at least in the United States, for women to struggle with infertility and therefore not being able to have children naturally or maybe at all. In terms of birth rates, I would think that the issue of reproductive health in women (and maybe even men) has an impact on the number of kids that are born year-to-year in the country. Therefore, the problem that I want to research is the problem of infertility in America and how it has impacted birth rate. In order for any country's population to grow, people need to be born, and it could have a huge impact on the population if people experience issues with fertility and reproductive health. With the concept of infertility, I would also like to research how various factors influence one's fertility and whether we can predict the fertility of an individual based on their demographic and environmental variables.

### Research questions

1. What is the weight of women's reproductive health in influencing a couple's ability to have children?
2. What is the current difference in birth rates from one country to another?
3. What is the average age for women to try to start having children?
4. How have non-traditional methods of having children influenced birth rate, such as adoption/IVF/etc?
5. What resources are provided to people who are experiencing issues with infertility?
6. What role does proper sex education play in fertility and reproductive health?
7. Does the current calculation of birth rate account for non-traditional methods of child delivery?
8. What are the key factors that play a role in one's fertility, men and women?

### Approach

I plan to address this problem statement by first building regression models that will allow for me to determine what variables or factors allow for predicting one's fertility or at a more major level, which factors influence/predict the fertility rate of a country. From there, knowing the fertility rate of a region, which indicates the ability of someone to have a child, I want to visualize and compare the birth rates between people who experience fertility issues versus those who don't, and see how it has then impacted the birth rate in that region for the overall population growth of the area. I think both rates (fertility & birth) play into each other, and my approach will try to address how fertility influence both numerical values.

**How your approach addresses (fully or partially) the problem.**

In terms of partiality, I think it will be difficult to address this problem at a global level (at least within the context of this course) but my approach will partially address the problem by identifying how reproductive health plays into the overall birth rate of a certain area. It will identify how we can predict fertility and draw a comparison to understand how fertility or lack of it affects birth rate and therefore also population growth.

**Data (Minimum of 3 Datasets - but no requirement on number of fields or rows)**

1. Kaggle Fertility Dataset

- 100 volunteers provided a semen sample analyzed according to the WHO 2010 criteria
- Sperm concentration are related to socio-demographic data, environmental factors, health status and life habits
- 100 rows
- 10 attributes/columns
- Date the data was donated: 2013-01-17
- No missing values
- Original dataset can be found at UCI Machine Learning Repository

2. World Bank Data

- Collected data from 1960 to 2016
- Data pertaining to countries' population, fertility rate and life expectancy
- Three different CSV files:
  - country_population.csv (61 columns)
  - fertility_rate.csv (61 columns)
  - life_expectancy.csv (61 columns)
- Collected from over the years but downloaded in 2018
- 264 rows in each CSV file

3. Female Respondent Data File
4. Female Pregnancy Data File

- All three above datasets:
  - Original source: National Survey of Family Growth
  - Collected from 2017-2019
  - Collection methods:
    - In-person interviews by trained female interviewers, in respondents' homes
  - Number of columns:
    - Female Respondent (3087)
    - Female Pregnancy (244)
    - Male Respondent (3059)
  - Handling of missing data
    - Only included completed cases
    - If any survey participants answered "don't know", "refused" or "not ascertained", they were given a default value .. i.e., 9, 8, 7, etc.
    - A case was defined as being complete if the respondent answered the last applicable question before ACASI

    * The small number of respondents who did not complete the ACASI section, partially or completely, will have "not ascertained" values assigned to all variables after their break-off point
  - Survey Info

**Required Packages**

- readxl
- tidyverse
- car
- QuantPsyc
- boot
- ggplot2

**Plots and Table Needs**

**What types of plots and tables will help you to illustrate the findings to your research questions?**

- Scatterplots – understanding relationship between variables
- Residual plots – assessing outliers
- Histograms – assess normality of the data sets
- Boxplots for comparing distributions among groups of study participants
- Tables:
  - Population of each country
  - Current birth rate in each country
  - Historical birth rates in each country
  - Current/historical fertility rates
  - Assessments of reproductive health for sample participants
  - Reproductive health comparison between men and women

**Questions for future steps**

- Assessing categorical variables when used in regression
- Generalization of analysis on one region to the greater regions – EX) USA to the world
- Merging of datasets, especially when bringing in data on men & women

# Final Project Step 2

**How to import and clean my data**

```
fertility_df <- read.csv("fertility.csv")
head(fertility_df)
```

```
fertility__rate_df <- read.csv("fertility_rate.csv")
country_pop_df <- read.csv("country_population.csv")
```

```
preg <- read.csv("2015_2017_FemPregData.csv")
fem_resp <- read.csv("2015_2017_FemRespData.csv")

str(preg)
```

```
##                                 Season                                    Age
##                                      0                                      0
##                     Childish.diseases              Accident.or.serious.trauma
##                                      0                                      0
##                  Surgical.intervention           High.fevers.in.the.last.year
##                                      0                                      0
##      Frequency.of.alcohol.consumption                           Smoking.habit
##                                      0                                      0
## Number.of.hours.spent.sitting.per.day                               Diagnosis
##                                      0                                      0
```

No NA values in any of the columns in the 1st fertility dataset.

```
sapply(fertility__rate_df, function(x) sum(is.na(x)))
```

```
## ï..Country.Name    Country.Code  Indicator.Name  Indicator.Code           X1960
##               0               0               0               0              28
##           X1961           X1962           X1963           X1964           X1965
##              27              28              29              28              28
##           X1966           X1967           X1968           X1969           X1970
##              28              28              28              28              27
##           X1971           X1972           X1973           X1974           X1975
##              26              25              27              27              26
##           X1976           X1977           X1978           X1979           X1980
##              25              26              26              26              26
##           X1981           X1982           X1983           X1984           X1985
##              24              21              24              24              24
##           X1986           X1987           X1988           X1989           X1990
##              24              20              24              24              22
##           X1991           X1992           X1993           X1994           X1995
##              22              19              23              22              20
##           X1996           X1997           X1998           X1999           X2000
##              23              19              22              21              18
##           X2001           X2002           X2003           X2004           X2005
##              18              16              18              19              17
##           X2006           X2007           X2008           X2009           X2010
##              16              15              16              16              17
##           X2011           X2012           X2013           X2014           X2015
##              16              16              18              18              17
##           X2016
##              18
```

```
sapply(country_pop_df, function(x) sum(is.na(x)))
```

I am removing Indicator.Name & Indicator.Code from both fertility_rate_df & country_pop_df because these columns have the same values for each row and don't give any extra information around the datasets and their specifications.

For the columns in fertility_rate_df that represent the years from 1960-2016, there are ~18-30 NAs in each of the columns. I think this dataset could be cleaned up depending on the number of years that I really wanted to investigate and analyze. 56 years of data is nice to have, but I think it is a bit excessive if we could rather try to find a yearly trend from a subset of the dataset.

The country_pop_df dataset does not have as many NA values in the year columns as the fertility_rate_df. However, if I subset the fertility_rate_df dataset than I will subset the population one by the same columns to keep it consistent and better for analyzing the same years among the countries.

I am also going to replace the NAs in the rest of the year columns 1980-2016 with the median value for the year column. I chose median over mean, since I don't want the value to be affected by the extreme values and countries have highly varying population sizes so the fertility rates and population numbers will be quite different.

```
# exclude variables v1, v2, v3
cols <- names(fertility_rate_df) %in% c("Indicator.Name","Indicator.Code","X1960", "X1961", "X1962","X
cols2 <- names(country_pop_df) %in% c("Indicator.Name","Indicator.Code","X1960", "X1961", "X1962","X196

fertility_rate_df<- fertility_rate_df[!cols]

country_pop_df <- country_pop_df[!cols2]
```

```
fertility_rate_df[,5:41] <- impute(fertility_rate_df[,3:39], fun = median)

country_pop_df[,5:41] <- impute(country_pop_df[,3:39], fun = median)
```

```
colSums(is.na(preg))
preg[,colSums(is.na(preg)) >0]
```

148 of the 380 variables contained NA values. The NA counts range from a couple hundred to ~5500 which is basically the number of rows in the dataset as is row (numRows = 5554). Given that there are already a great amount of columns in this dataset, I decided to remove all of the columns with any NA values since I think the rest of the data is already representative of the females that were surveyed about their pregnancies.

```
preg <- preg[ , colSums(is.na(preg)) == 0]
```

```
colSums(is.na(fem_resp))
fem_resp[,colSums(is.na(fem_resp)) >0]
```

2,792 of the 3,024 total variables in the female resp dataset have NA values. The count of these NA values is mostly very high such as being around 5,554 which is the total number of rows in the dataset as is, which would mean that the entire column contains NAs. Given that this dataset already has many columns and many are not applicable to my problem/question, I am going to remove all columns with any NA values to make the dataset easier to consume, analyze and utilize.

```
fem_resp <- fem_resp[ , colSums(is.na(fem_resp)) == 0]
```

**Merging of similar datasets** I want to merge the fertility_rate_df & country_pop_df datasets on country code, given that both datasets provide data on the same countries over the 36 years from 1980 - 2016. I took out the year variables from 1960-1979 in order to subset the data and not have to handle as many NA values.

```
# merge fertility_rate_df & country_pop_df by country code
rate_pop_merged <- merge(fertility__rate_df,country_pop_df,by="Country.Code")
head(rate_pop_merged)
```

I also want to merge the preg & fem_resp dataframes on CASEID, since both datasets represent data for females surveyed on their pregnancies from 2015-2017. I chose these years, since the fertility rate & country population datasets go up until 2016, so the years from 2015-2017 will represent the more recent years for looking at women's fertility and pregnancy experiences. I want to get a more current idea of what is affecting women's ability to have children.

```
# merge preg & fem_resp dataframes on ï..CASEID
preg_resp_merged <- merge(preg,fem_resp,by="ï..CASEID")
```

**What does the final data set look like?**

```
dplyr::glimpse(fertility_df)
```

```
## Rows: 100
## Columns: 10
## $ Season                                <chr> "spring", "spring", "spring", "s~
## $ Age                                   <int> 30, 35, 27, 32, 30, 30, 30, 36, ~
## $ Childish.diseases                     <chr> "no", "yes", "yes", "no", "yes",~
## $ Accident.or.serious.trauma            <chr> "yes", "no", "no", "yes", "yes",~
## $ Surgical.intervention                 <chr> "yes", "yes", "no", "yes", "no",~
## $ High.fevers.in.the.last.year          <chr> "more than 3 months ago", "more ~
## $ Frequency.of.alcohol.consumption      <chr> "once a week", "once a week", "h~
## $ Smoking.habit                         <chr> "occasional", "daily", "never", ~
## $ Number.of.hours.spent.sitting.per.day <int> 16, 6, 9, 7, 9, 9, 8, 7, 5, 5, 6~
## $ Diagnosis                             <chr> "Normal", "Altered", "Normal", "~
```

```
str(fertility_df)
```

```
## 'data.frame':    100 obs. of  10 variables:
##  $ Season                                : chr  "spring" "spring" "spring" "spring" ...
##  $ Age                                   : int  30 35 27 32 30 30 30 36 30 29 ...
##  $ Childish.diseases                     : chr  "no" "yes" "yes" "no" ...
##  $ Accident.or.serious.trauma            : chr  "yes" "no" "no" "yes" ...
##  $ Surgical.intervention                 : chr  "yes" "yes" "no" "yes" ...
##  $ High.fevers.in.the.last.year          : chr  "more than 3 months ago" "more than 3 months ago" "mo
##  $ Frequency.of.alcohol.consumption      : chr  "once a week" "once a week" "hardly ever or never" "ha
##  $ Smoking.habit                         : chr  "occasional" "daily" "never" "never" ...
##  $ Number.of.hours.spent.sitting.per.day: int  16 6 9 7 9 9 8 7 5 5 ...
##  $ Diagnosis                             : chr  "Normal" "Altered" "Normal" "Normal" ...
```

```
dplyr::glimpse(rate_pop_merged)
```

```
## Rows: 264
## Columns: 81
## $ Country.Code      <chr> "ABW", "AFG", "AGO", "ALB", "AND", "ARB", "ARE", "AR~
```

```
## $ ï..Country.Name.x <chr> "Aruba", "Afghanistan", "Angola", "Albania", "Andorr~
## $ X1980.x           <dbl> 2.392000, 7.449000, 7.504000, 3.621000, NA, 6.335756~
## $ X1981.x           <dbl> 2.37700, 7.44900, 7.46900, 3.53000, NA, 6.26037, 5.4~
## $ X1982.x           <dbl> 2.392000, 7.449000, 7.504000, 3.621000, 2.914000, 6.~
## $ X1983.x           <dbl> 2.37700, 7.44900, 7.46900, 3.53000, 2.91400, 6.26037~
## $ X1984.x           <dbl> 2.364000, 7.450000, 7.438000, 3.452000, 2.914000, 6.~
## $ X1985.x           <dbl> 2.353000, 7.452000, 7.413000, 3.383000, 2.914000, 6.~
## $ X1986.x           <dbl> 2.34200, 7.45500, 7.39400, 3.32300, 2.91400, 5.99415~
## $ X1987.x           <dbl> 2.332000, 7.458000, 7.380000, 3.269000, 2.914000, 5.~
## $ X1988.x           <dbl> 2.320000, 7.460000, 7.366000, 3.217000, 2.914000, 5.~
## $ X1989.x           <dbl> 2.307000, 7.461000, 7.349000, 3.164000, 2.914000, 5.~
## $ X1990.x           <dbl> 2.291000, 7.461000, 7.324000, 3.108000, 2.914000, 5.~
## $ X1991.x           <dbl> 2.272000, 7.461000, 7.291000, 3.046000, 2.914000, 5.~
## $ X1992.x           <dbl> 2.249000, 7.466000, 7.247000, 2.978000, 2.914000, 5.~
## $ X1993.x           <dbl> 2.221000, 7.479000, 7.193000, 2.905000, 2.914000, 5.~
## $ X1994.x           <dbl> 2.187000, 7.502000, 7.130000, 2.829000, 2.914000, 4.~
## $ X1995.x           <dbl> 2.149000, 7.535000, 7.063000, 2.751000, 2.914000, 4.~
## $ X1996.x           <dbl> 2.108000, 7.572000, 6.992000, 2.672000, 2.914000, 4.~
## $ X1997.x           <dbl> 2.064000, 7.606000, 6.922000, 2.591000, 2.914000, 4.~
## $ X1998.x           <dbl> 2.021000, 7.630000, 6.854000, 2.507000, 2.914000, 4.~
## $ X1999.x           <dbl> 1.978000, 7.635000, 6.791000, 2.422000, 2.914000, 4.~
## $ X2000.x           <dbl> 1.939000, 7.616000, 6.734000, 2.334000, 2.914000, 4.~
## $ X2001.x           <dbl> 1.903000, 7.569000, 6.683000, 2.246000, 2.914000, 3.~
## $ X2002.x           <dbl> 1.872000, 7.494000, 6.639000, 2.157000, 2.914000, 3.~
## $ X2003.x           <dbl> 1.846000, 7.392000, 6.602000, 2.068000, 2.914000, 3.~
## $ X2004.x           <dbl> 1.823000, 7.271000, 6.568000, 1.981000, 2.914000, 3.~
## $ X2005.x           <dbl> 1.803000, 7.136000, 6.536000, 1.897000, 2.914000, 3.~
## $ X2006.x           <dbl> 1.78700, 6.98800, 6.50200, 1.82100, 2.91400, 3.55976~
## $ X2007.x           <dbl> 1.774000, 6.827000, 6.465000, 1.754000, 2.914000, 3.~
## $ X2008.x           <dbl> 1.766000, 6.651000, 6.420000, 1.703000, 1.240000, 3.~
## $ X2009.x           <dbl> 1.763000, 6.460000, 6.368000, 1.668000, 1.180000, 3.~
## $ X2010.x           <dbl> 1.764000, 6.254000, 6.307000, 1.650000, 1.250000, 3.~
## $ X2011.x           <dbl> 1.769000, 6.038000, 6.238000, 1.646000, 1.190000, 3.~
## $ X2012.x           <dbl> 1.776000, 5.816000, 6.162000, 1.653000, 1.270000, 3.~
## $ X2013.x           <dbl> 1.783000, 5.595000, 6.082000, 1.668000, 2.914000, 3.~
## $ X2014.x           <dbl> 1.791000, 5.380000, 6.000000, 1.685000, 2.914000, 3.~
## $ X2015.x           <dbl> 1.796000, 5.174000, 5.920000, 1.700000, 2.914000, 3.~
## $ X2016.x           <dbl> 1.800000, 4.981000, 5.841000, 1.710000, 2.914000, 3.~
## $ X2015.1.x         <dbl> 1.80100, 4.80200, 5.76600, 1.71400, 2.91400, 3.37384~
## $ X2016.1.x         <dbl> 1.800000, 4.635000, 5.694000, 1.713000, 2.914000, 3.~
## $ ï..Country.Name.y <chr> "Aruba", "Afghanistan", "Angola", "Albania", "Andorr~
## $ X1980.y           <dbl> 60096, 13248370, 8929900, 2671997, 36067, 165689490,~
## $ X1981.y           <dbl> 60567, 13053954, 9244507, 2726056, 37500, 171051950,~
## $ X1982.y           <dbl> 60096, 13248370, 8929900, 2671997, 36067, 165689490,~
## $ X1983.y           <dbl> 60567, 13053954, 9244507, 2726056, 37500, 171051950,~
## $ X1984.y           <dbl> 61345, 12749645, 9582156, 2784278, 39114, 176490084,~
## $ X1985.y           <dbl> 62201, 12389269, 9931562, 2843960, 40867, 182005827,~
## $ X1986.y           <dbl> 62836, 12047115, 10277321, 2904429, 42706, 187610756~
## $ X1987.y           <dbl> 63026, 11783050, 10609042, 2964762, 44600, 193310301~
## $ X1988.y           <dbl> 62644, 11601041, 10921037, 3022635, 46517, 199093767~
## $ X1989.y           <dbl> 61833, 11502761, 11218268, 3083605, 48455, 204942549~
## $ X1990.y           <dbl> 61079, 11540888, 11513968, 3142336, 50434, 210844771~
## $ X1991.y           <dbl> 61032, 11777609, 11827237, 3227943, 52448, 216787402~
## $ X1992.y           <dbl> 62149, 12249114, 12171441, 3286542, 54509, 224735446~
```

```
## $ X1993.y          <dbl> 64622, 12993657, 12553446, 3266790, 56671, 230829868~
## $ X1994.y          <dbl> 68235, 13981231, 12968345, 3247039, 58888, 235037179~
## $ X1995.y          <dbl> 72504, 15095099, 13403734, 3227287, 60971, 241286091~
## $ X1996.y          <dbl> 76700, 16172719, 13841301, 3207536, 62677, 247435930~
## $ X1997.y          <dbl> 80324, 17099541, 14268994, 3187784, 63850, 255029671~
## $ X1998.y          <dbl> 83200, 17822884, 14682284, 3168033, 64360, 260843462~
## $ X1999.y          <dbl> 85451, 18381605, 15088981, 3148281, 64327, 266575075~
## $ X2000.y          <dbl> 87277, 18863999, 15504318, 3128530, 64142, 272235146~
## $ X2001.y          <dbl> 89005, 19403676, 15949766, 3108778, 64370, 277962869~
## $ X2002.y          <dbl> 90853, 20093756, 16440924, 3089027, 65390, 283832016~
## $ X2003.y          <dbl> 92898, 20966463, 16983266, 3060173, 67341, 289850357~
## $ X2004.y          <dbl> 94992, 21979923, 17572649, 3051010, 70049, 296026575~
## $ X2005.y          <dbl> 97017, 23064851, 18203369, 3039616, 73182, 302434519~
## $ X2006.y          <dbl> 98737, 24118979, 18865716, 3026939, 76244, 309162029~
## $ X2007.y          <dbl> 100031, 25070798, 19552542, 3011487, 78867, 31626472~
## $ X2008.y          <dbl> 100832, 25893450, 20262399, 2992547, 80991, 32377326~
## $ X2009.y          <dbl> 101220, 26616792, 20997687, 2970017, 82683, 33165379~
## $ X2010.y          <dbl> 101353, 27294031, 21759420, 2947314, 83861, 33982548~
## $ X2011.y          <dbl> 101453, 28004331, 22549547, 2927519, 84462, 34814509~
## $ X2012.y          <dbl> 101669, 28803167, 23369131, 2913021, 84449, 35650890~
## $ X2013.y          <dbl> 102053, 29708599, 24218565, 2905195, 83751, 36489587~
## $ X2014.y          <dbl> 102577, 30696958, 25096150, 2900401, 82431, 37330699~
## $ X2015.y          <dbl> 103187, 31731688, 25998340, 2895092, 80788, 38170208~
## $ X2016.y          <dbl> 103795, 32758020, 26920466, 2889104, 79223, 39004302~
## $ X2015.1.y        <dbl> 104341, 33736494, 27859305, 2880703, 78014, 39830496~
## $ X2016.1.y        <dbl> 104822, 34656032, 28813463, 2876101, 77281, 40645269~
```

```
#str(rate_pop_merged)
```

```
dplyr::glimpse(preg_resp_merged)
```

```
## Rows: 5,554
## Columns: 463
## $ ï..CASEID    <dbl> 7.157213e+07, 7.404652e+07, 7.454052e+07, 7.486513e+07, 7~
## $ PREGORDR     <dbl> 323232112, 2323235, 2626265, 3535345, 4242425, 2323235, 3~
## $ HOWPREG_N.x  <dbl> 3.220001e+06, 1.000000e+00, 1.000000e+00, 1.000000e+00, 1~
## $ HOWPREG_P.x  <dbl> 1.120000e+02, 2.000000e+12, 2.000006e+06, 4.121100e+12, 3~
## $ MOSCURRP.x   <dbl> 5, 15512010, 50, 1352, 112, 50, 915, 50, 1, 50, 11115, 1,~
## $ NOWPRGDK.x   <dbl> 915, 12, 5, 111119985, 5, 1, 91994, 5, 16512011, 5, 18111~
## $ PREGEND1     <dbl> 51994, 5, 13512007, 116256000000, 16111992, 15512011, 112~
## $ PREGEND2     <dbl> 5.000000e+00, 5.000000e+00, 1.200000e+01, 3.350000e+02, 1~
## $ HOWENDDK     <dbl> 2.111110e+05, 2.555121e+10, 5.000000e+00, 1.350000e+02, 1~
## $ NBRNALIV     <dbl> 1.100000e+01, 2.000000e+00, 5.000000e+00, 2.150000e+02, 2~
## $ MULTBRTH     <dbl> 1, 135, 255111, 21, 11, 2, 145, 31, 3, 2, 3, 135, 4, 125,~
## $ BORNALIV     <dbl> 5, 125, 12, 2, 4, 235, 415, 1, 5, 135, 5, 125, 5, 55, 135~
## $ DATPRGEN_Y   <dbl> 1.25000e+02, 5.50000e+01, 3.00000e+00, 2.20052e+21, 5.000~
## $ AGEATEND     <dbl> 215, 5, 5, 11235, 125, 55, 2, 105, 55, 411201, 215, 5, 55~
## $ HPAGEEND     <dbl> 2.10000e+01, 0.00000e+00, 1.25000e+02, 5.00000e+00, 3.150~
## $ GESTASUN_M   <dbl> 2.00000e+00, 0.00000e+00, 5.50000e+01, 5.50000e+01, 3.100~
## $ GESTASUN_W   <dbl> 2.20132e+21, 0.00000e+00, 5.00000e+00, 1.50000e+01, 2.000~
## $ WKSGEST      <dbl> 5.500000e+01, 5.000000e+00, 0.000000e+00, 1.100000e+01, 2~
## $ MOSGEST      <dbl> 5.00000e+00, 5.50000e+01, 0.00000e+00, 2.00500e+03, 5.500~
## $ DK1GEST      <dbl> 5.100000e+01, 5.000000e+00, 0.000000e+00, 2.519790e+16, 5~
```

```
## $ DK2GEST      <dbl> 2002, 5, 5, 112, 11, 5, 199751, 55, 15, 1, 2719725, 22155~
## $ DK3GEST      <dbl> 2.519765e+06, 2.015000e+03, 5.500000e+01, 2.012000e+07, 1~
## $ BABYSEX1     <dbl> 2.115100e+04, 2.719880e+12, 5.000000e+00, 2.005120e+08, 2~
## $ BIRTHWGT_LB1 <dbl> 1.550000e+02, 3.500000e+01, 1.000000e+00, 3.519800e+12, 1~
## $ BIRTHWGT_OZ1 <dbl> 2.002000e+04, 3.200000e+01, 0.000000e+00, 8.000000e+00, 1~
## $ LOBTHWGT1    <dbl> 2.00255e+05, 1.00000e+00, 5.50000e+01, 2.01511e+05, 1.993~
## $ BABYSEX2     <dbl> 0, 201551, 0, 1, 0, 142014000000, 4, 42111, 21, 411, 7199~
## $ BIRTHWGT_LB2 <dbl> 1.000000e+00, 0.000000e+00, 4.245510e+05, 1.200200e+04, 1~
## $ BIRTHWGT_OZ2 <dbl> 3.200212e+10, 2.200000e+01, 0.000000e+00, 2.219800e+16, 1~
## $ LOBTHWGT2    <dbl> 1.00000e+00, 1.00000e+00, 0.00000e+00, 1.10000e+01, 1.000~
## $ BABYSEX3     <dbl> 7.000000e+00, 1.200913e+10, 5.555500e+04, 1.000000e+00, 5~
## $ BIRTHWGT_LB3 <dbl> 1.00000e+00, 1.00000e+01, 5.55550e+04, 1.61995e+11, 4.000~
## $ BIRTHWGT_OZ3 <dbl> 1.020150e+13, 1.234568e+07, 5.000000e+00, 9.000000e+00, 1~
## $ LOBTHWGT3    <dbl> 1.00000e+00, 8.00000e+00, 5.55000e+02, 5.00000e+00, 4.000~
## $ BABYDOB_Y    <dbl> 0.00000e+00, 1.10000e+01, 1.00000e+00, 4.00000e+00, 6.000~
## $ KIDAGE       <dbl> 1.000000e+00, 1.100000e+01, 5.000000e+00, 5.199710e+12, 6~
## $ HPAGELB      <dbl> 1, 111, 5, 8, 1, 6, 552, 1, 8, 9112, 71392122015, 0, 6, 0~
## $ BIRTHPLC     <dbl> 1, 111, 0, 7, 1, 5, 55555, 6, 5, 915, 1, 55555, 5, 55555,~
## $ PAYBIRTH1    <dbl> 1.00000e+01, 1.00000e+00, 1.00000e+00, 5.00000e+00, 6.100~
## $ PAYBIRTH2    <dbl> 55555, 8, 5, 1, 71386, 41, 1, 6, 6, 4, 4, 5, 1, 5, 31, 4,~
## $ PAYBIRTH3    <dbl> 5.55555e+05, 1.00000e+00, 5.50000e+01, 1.00000e+00, 6.201~
## $ CSECPRIM     <dbl> 552, 8, 95, 71, 1, 3134, 55, 6, 6, 1, 10, 1, 5, 1, 22016,~
## $ CSECMED1     <dbl> 55555, 11, 15, 81391112015, 1920, 1, 11555555155, 11, 11,~
## $ CSECMED2     <dbl> 1.000000e+00, 8.000000e+00, 3.000000e+00, 1.000000e+00, 6~
## $ CSECMED3     <dbl> 1.00000e+00, 5.00000e+00, 1.00000e+00, 2.22300e+03, 9.199~
## $ CSECMED4     <dbl> 5.000000e+00, 4.000000e+00, 1.000000e+00, 6.000000e+00, 1~
## $ CSECMED5     <dbl> 5.110000e+02, 5.200910e+12, 1.364000e+03, 7.200310e+12, 1~
## $ CSECMED6     <dbl> 51555555555, 6, 820131364, 10, 555, 131387, 6, 3, 4, 2119~
## $ CSECPLAN     <dbl> 9.500000e+01, 3.000000e+00, 2.400000e+01, 1.000000e+00, 1~
## $ KNEWPREG     <dbl> 1.500000e+01, 1.000000e+00, 1.345500e+234, 5.550000e+02, ~
## $ TRIMESTR     <dbl> 4.000000e+00, 1.000000e+00, 1.364139e+07, 1.555150e+05, 5~
## $ LTRIMEST     <dbl> 1.000000e+00, 3.100000e+01, 2.700000e+01, 5.522005e+06, 1~
## $ PRIORSMK     <dbl> 1.000000e+00, 8.139111e+10, 3.000000e+00, 5.000000e+00, 1~
## $ POSTSMKS     <dbl> 6.000000e+00, 1.000000e+00, 3.000000e+00, 1.515111e+06, 1~
## $ NPOSTSMK     <dbl> 5.136900e+04, 2.227000e+03, 1.555556e+08, 1.011510e+12, 1~
## $ GETPRENA     <dbl> 1.201414e+08, 4.000000e+00, 3.000000e+00, 9.500000e+01, 1~
## $ BGNPRENA     <dbl> 3.000000e+01, 2.201510e+12, 3.000000e+00, 1.100000e+01, 1~
## $ PNCTRIM      <dbl> 5.134500e+230, 1.000000e+01, 3.000000e+00, 3.000000e+00, ~
## $ LPNCTRI      <dbl> 13691390, 55555, 3, 5, 1, 11551555555, 44995, 55555, 4, 0~
## $ LIVEHERE1    <dbl> 2.100000e+01, 5.555550e+05, 3.000000e+00, 9.000000e+00, 1~
## $ ALIVENOW1    <dbl> 4.000000e+00, 5.520000e+02, 3.000000e+00, 4.000000e+00, 2~
## $ WHENDIED_Y1  <dbl> 4.000000e+00, 5.555500e+04, 3.000000e+00, 1.000000e+00, 1~
## $ WHENLEFT_Y1  <dbl> 5.555556e+08, 1.000000e+00, 3.000000e+00, 1.000000e+00, 4~
## $ LASTAGE1     <dbl> 4.00000e+00, 1.00000e+00, 3.00000e+00, 2.00000e+00, 1.134~
## $ WHERENOW1    <dbl> 4, 5, 3, 11150, 13451389, 1, 4, 0, 55555, 515, 6, 3, 1116~
## $ LEGAGREE1    <dbl> 4.000000e+00, 5.500000e+01, 3.000000e+00, 4.000000e+00, 4~
## $ PARENEND1    <dbl> 4.000000e+00, 5.151116e+10, 3.000000e+00, 1.134600e+235, ~
## $ ANYNURSE1    <dbl> 4, 95, 3, 13451391, 6, 3200913111, 4, 15, 55555, 1, 6, 55~
## $ FEDSOLID1    <dbl> 4, 15, 3, 46, 6, 17, 4, 4, 1, 1, 6, 5, 37, 0, 3, 13201313~
## $ FRSTEATD_N1  <dbl> 4, 8, 3, 84, 555555555, 3, 4, 7, 1, 6, 6, 51323355, 24, 8~
## $ FRSTEATD_P1  <dbl> 4.0000e+00, 1.0000e+00, 3.0000e+00, 1.2610e+03, 6.0000e+0~
## $ FRSTEATD1    <dbl> 4.000000e+00, 1.000000e+00, 3.000000e+00, 6.000000e+00, 6~
## $ QUITNURS1    <dbl> 4.000000e+00, 1.000000e+00, 3.000000e+00, 6.000000e+00, 6~
## $ AGEQTNUR_N1  <dbl> 4, 11287, 3, 555555555, 6, 13451389, 4, 11331, 95, 3, 6, ~
```

```
## $ AGEQTNUR_P1  <dbl> 4.000000e+00, 8.000000e+00, 3.000000e+00, 6.000000e+00, 6~
## $ AGEQTNUR1    <dbl> 4, 3200712871, 3, 6, 6, 1, 4, 7, 3, 13451389, 6, 1, 3, 55~
## $ LIVEHERE2    <dbl> 4.0000e+00, 1.4000e+01, 3.0000e+00, 6.0000e+00, 6.0000e+0~
## $ ALIVENOW2    <dbl> 4, 3, 3, 6, 6, 555555555, 4, 12, 1, 12, 6, 1, 3, 21, 1, 6~
## $ WHENDIED_Y2  <dbl> 4.0000e+00, 1.0000e+00, 3.0000e+00, 6.0000e+00, 6.0000e+0~
## $ WHENLEFT_Y2  <dbl> 4.000000e+00, 8.000000e+00, 3.000000e+00, 6.000000e+00, 6~
## $ LASTAGE2     <dbl> 4.00000e+00, 1.13460e+235, 3.00000e+00, 6.00000e+00, 6.00~
## $ WHERENOW2    <dbl> 4, 13451391, 3, 6, 6, 1, 4, 3, 7, 555555555, 6, 725, 3, 3~
## $ LEGAGREE2    <dbl> 4, 46, 3, 6, 6, 1, 4, 3, 4200913121, 4, 6, 27, 3, 6, 1, 6~
## $ PARENEND2    <dbl> 4.0000e+00, 3.6000e+01, 1.3451e+187, 6.0000e+00, 6.0000e+~
## $ ANYNURSE2    <dbl> 4.000000e+00, 1.309000e+03, 1.111110e+27, 6.000000e+00, 6~
## $ FEDSOLID2    <dbl> 1.345100e+183, 8.000000e+00, 0.000000e+00, 6.000000e+00, ~
## $ FRSTEATD_N2  <dbl> 1.111110e+21, 8.000000e+00, 1.000010e+26, 6.000000e+00, 6~
## $ FRSTEATD_P2  <dbl> 0, 555515555, 2, 6, 6, 1, 4, 1, 7, 4, 6, 5, 3, 11155, 1, ~
## $ FRSTEATD2    <dbl> 1.00000e+20, 8.00000e+00, 3.00000e+00, 6.00000e+00, 6.000~
## $ QUITNURS2    <dbl> 1, 8, 3, 6, 6, 1, 4, 1, 13451391, 1, 6, 0, 3, 5, 1, 6, 5,~
## $ AGEQTNUR_N2  <dbl> 4, 8, 93, 6, 6, 1, 4, 1, 46, 1, 6, 0, 3, 5555555555, 1, 6~
## $ AGEQTNUR_P2  <dbl> 5, 8, 5, 6, 6, 1, 4, 1, 1995, 1, 6, 5555, 3, 5, 19, 6, 5,~
## $ AGEQTNUR2    <dbl> 8, 8, 15, 6, 6, 1, 4, 1, 420091312, 1, 6, 5, 3, 55223455,~
## $ LIVEHERE3    <dbl> 8, 8, 11155551, 6, 6, 1, 4, 4, 3, 1, 6, 21, 3, 3, 19, 6, ~
## $ ALIVENOW3    <dbl> 1.0000e+00, 8.0000e+00, 5.5000e+01, 6.0000e+00, 6.0000e+0~
## $ WHENDIED_Y3  <dbl> 0, 8, 5, 6, 6, 1, 4, 1, 3, 1, 6, 1111115, 3, 1, 19, 6, 5,~
## $ WHENLEFT_Y3  <dbl> 1, 8, 31, 4, 6, 1, 4, 1, 7, 1, 6, 55, 3, 2, 19, 6, 5, 6, ~
## $ LASTAGE3     <dbl> 555556000000, 8, 3, 6, 6, 1, 4, 4, 155555555, 1, 6, 5, 3,~
## $ WHERENOW3    <dbl> 5, 8, 1, 4, 6, 1, 4, 4, 3, 1, 6, 5, 3, 6, 19, 6, 5, 6, 55~
## $ LEGAGREE3    <dbl> 9, 8, 53, 6, 6, 1, 4, 1, 7, 1, 6, 55, 3, 8, 19, 6, 5, 6, ~
## $ PARENEND3    <dbl> 0.00000e+00, 8.00000e+00, 1.00000e+00, 4.00000e+00, 6.000~
## $ ANYNURSE3    <dbl> 2125, 8, 6, 6, 6, 1, 4, 1, 7, 1, 6, 1, 3, 3, 19, 6, 5, 6,~
## $ FEDSOLID3    <dbl> 5.000000e+00, 8.000000e+00, 1.000000e+00, 4.000000e+00, 6~
## $ FRSTEATD_N3  <dbl> 1.000000e+00, 8.000000e+00, 6.000000e+00, 6.000000e+00, 6~
## $ FRSTEATD_P3  <dbl> 2125, 8, 15, 4, 6, 1, 4, 1, 3, 1, 6, 1, 1, 1, 19, 6, 5, 6~
## $ FRSTEATD3    <dbl> 11, 8, 991555, 6, 6, 4, 4, 7, 7, 1, 6, 3, 1, 711, 19, 6, ~
## $ QUITNURS3    <dbl> 5125, 8, 5, 4, 6, 7, 4, 7, 3, 1, 6, 8, 4, 3, 19, 6, 5, 6,~
## $ AGEQTNUR_N3  <dbl> 8.5555e+04, 8.0000e+00, 5.0000e+00, 6.0000e+00, 6.0000e+0~
## $ AGEQTNUR_P3  <dbl> 2.52000e+02, 8.00000e+00, 1.50000e+01, 4.00000e+00, 6.000~
## $ AGEQTNUR3    <dbl> 5.000000e+00, 8.000000e+00, 5.000000e+00, 6.000000e+00, 6~
## $ PRGOUTCOME   <dbl> 5.000000e+00, 8.000000e+00, 5.515556e+09, 4.000000e+00, 6~
## $ OUTCOM_S     <dbl> 4.5e+01, 8.0e+00, 5.0e+00, 6.0e+00, 6.0e+00, 1.0e+00, 1.0~
## $ DATEND       <chr> "1", "1", "3E+23", "5", "1", "5", "12001", "1", "3", "5",~
## $ FMARITAL     <chr> "2530", "1", "42420000173", "245", "5", "0", "1", "19", "~
## $ RMARITAL     <chr> "20072013", "1", "820002847", "355121144", "1151", "0", "~
## $ HIEDUC       <chr> "11", "2", "", "111144", "11", "5", "55555555", "2", "1.1~
## $ METRO        <chr> "0", "32", "", "5", "2", "1", "4", "1", "1", "3", "3E+23"~
## $ DATEND_I     <chr> "1.12889E+18", "1", "", "1", "4E+23", "23", "19202325", "~
## $ AGEPREG_I    <chr> "1", "1.21323E+17", "", "1", "3", "1", "1.9972E+15", "8",~
## $ DATECON_I    <chr> "2", "1", "", "1", "3", "2E+23", "18202224", "910", "25",~
## $ FMARCON5_I   <chr> "11", "45", "", "5.55556E+12", "202227", "22", "1.99819E+~
## $ RMARCON6_I   <chr> "11", "2", "", "313", "1.9932E+11", "0", "1", "5", "1", "~
## $ LEARNPRG_I   <chr> "11", "1", "", "313", "202126", "0", "121998", "515", "0"~
## $ LBW1_I       <chr> "2", "5", "", "9.22222E+42", "551", "1.11222E+18", "1995"~
## $ LIVCHILD_I   <chr> "222", "15", "", "2", "551", "1.51139E+11", "1211", "735"~
## $ OLDWANTR_I   <chr> "41270000", "11", "", "0", "55", "4", "0", "11", "995", "~
## $ OLDWANTP_I   <chr> "93", "2.11656E+11", "", "2", "11", "1E+66", "1.19972E+42~
## $ WANTRESP_I   <chr> "610005968.1", "3", "", "20211", "121995", "1.11114E+14",~
```

```
## $ WANTPART_I   <chr> "75.64", "3", "", "20022005", "22", "2", "1E+69", "1", "1~
## $ TOOSOON_I    <chr> "", "1", "", "2224", "5", "1", "1.12889E+18", "0", "3", "~
## $ NEWWANTR_I   <chr> "", "1", "", "20022004", "1995", "2", "1", "0", "5", "2E+~
## $ AGER_I       <chr> "", "1", "", "2123", "2211", "5", "2", "5551", "5", "4", ~
## $ FMARITAL_I   <chr> "", "0", "", "51", "1", "1", "1", "65", "15", "3", "4", "~
## $ RMARITAL_I   <chr> "", "55", "", "51", "0", "2", "2", "45", "5", "0", "8E+69~
## $ EDUCAT_I     <chr> "", "55555555", "", "55", "1.19932E+26", "5", "1", "1", "~
## $ HIEDUC_I     <chr> "", "1", "", "56", "3.31139E+14", "1", "11732222", "1", "~
## $ RACE_I       <chr> "", "2", "", "1120022234", "6", "1311", "3222", "0", "551~
## $ HISPANIC_I   <chr> "", "1.11112E+12", "", "4411", "6E+72", "0", "3222", "211~
## $ HISPRACE_I   <chr> "", "1", "", "102005", "188888811", "1", "2222", "115135"~
## $ HISPRACE2_I  <chr> "", "3", "", "2011", "2", "2000222", "3222", "11", "6", "~
## $ RCURPREG_I   <chr> "", "1", "", "1", "1", "20", "22", "3120", "5", "555", "4~
## $ PREGNUM_I    <chr> "", "5", "", "2.43031E+13", "4", "4.216E+20", "4", "5", "~
## $ PARITY_I     <chr> "", "5", "", "4111", "2", "66.84", "4", "3", "1", "5555",~
## $ CURR_INS_I   <chr> "", "5.55556E+30", "", "1", "1078661", "", "4", "3", "111~
## $ PUBASSIS_I   <chr> "", "2", "", "2", "552", "", "4", "1", "15", "3.1201E+15"~
## $ POVERTY_I    <chr> "", "2E+23", "", "3.20022E+26", "552", "", "4", "1", "1",~
## $ LABORFOR_I   <chr> "", "0", "", "1.91139E+14", "661", "", "1", "1", "111", "~
## $ RELIGION_I   <chr> "", "0", "", "6", "662", "", "1000222", "1", "15", "4111"~
## $ METRO_I      <chr> "", "0", "", "7E+69", "0", "", "22", "51", "1595", "0", "~
## $ WGT2015_2017 <chr> "", "0", "", "1", "8", "", "4E+26", "2", "1", "1", "2", "~
## $ SECU         <chr> "", "11", "", "188888811", "8", "", "820005755.5811421221~
## $ SEST         <chr> "", "0", "", "2", "2", "", "", "21", "17", "61138811123",~
## $ CMINTVW      <chr> "", "1", "", "1", "222", "", "", "5.51511E+12", "16", "29~
## $ CMLSTYR      <chr> "", "1.20152E+26", "", "4", "22", "", "", "1", "7512314",~
## $ CMJAN3YR     <chr> "", "41139121123", "", "1", "3.231E+21", "", "", "5", "11~
## $ CMJAN4YR     <chr> "", "9", "", "4", "86.84", "", "", "3995", "11115", "21",~
## $ CMJAN5YR     <chr> "", "4", "", "2", "", "", "", "5", "1", "1.30889E+18", ""~
## $ QUARTER      <chr> "", "3E+69", "", "115022", "", "", "", "5", "1.31112E+17"~
## $ PHASE        <chr> "", "21", "", "32", "", "", "", "5.15556E+25", "1", "2", ~
## $ INTVWYEAR    <chr> "", "588888821", "", "32", "", "", "", "11", "0", "6", ""~
## $ X            <chr> "74", "", "4", "0", "", "", "1", "9999", "", "", "4111", ~
## $ X.1          <chr> "4", "", "4E+69", "0", "", "", "2", "1.11556E+32", "", ""~
## $ X.2          <chr> "1E+67", "", "1", "0", "", "", "1", "1", "", "", "1", "",~
## $ X.3          <chr> "11", "", "1", "1", "", "", "1", "3E+23", "", "", "1.2012~
## $ X.4          <chr> "3.21089E+11", "", "2888811", "1.21172E+18", "", "", "1",~
## $ X.5          <chr> "2", "", "2", "1.3214E+11", "", "", "1", "2", "", "", "12~
## $ X.6          <chr> "117", "", "1", "0", "", "", "1", "2", "", "", "53", "", ~
## $ X.7          <chr> "2", "", "3", "10", "", "", "10843232225", "0", "", "", "~
## $ X.8          <chr> "119532", "", "4", "54", "", "", "5555235", "222", "", ""~
## $ X.9          <chr> "32", "", "1", "7E+66", "", "", "5555235", "20072014", ""~
## $ X.10         <chr> "32", "", "3", "11", "", "", "3232225", "2027", "", "", "~
## $ X.11         <chr> "32", "", "4", "1.11221E+11", "", "", "6666236", "2007201~
## $ X.12         <chr> "32", "", "2", "2", "", "", "0", "2027", "", "", "2", "",~
## $ X.13         <chr> "1E+132", "", "11822222", "1", "", "", "8", "55", "", "",~
## $ X.14         <chr> "8", "", "3222", "2", "", "", "8", "66", "", "", "2", "",~
## $ X.15         <chr> "8", "", "3222", "1", "", "", "8", "55", "", "", "130053"~
## $ X.16         <chr> "1", "", "2222", "2", "", "", "8", "66", "", "", "33", ""~
## $ X.17         <chr> "1000222", "", "4222", "1", "", "", "8", "2", "", "", "33~
## $ X.18         <chr> "22", "", "1E+132", "13233", "", "", "8", "6", "", "", "5~
## $ X.19         <chr> "11250000170", "", "2", "3", "", "", "8", "0", "", "", "4~
## $ X.20         <chr> "820002907.1", "", "222", "3", "", "", "2", "996", "", ""~
## $ X.21         <chr> "71.85", "", "20", "3", "", "", "222", "4", "", "", "8", ~
```

```
## $ X.22         <chr> "", "", "00000000000000032270000157111000320 23.7129952865~
## $ X.23         <chr> "", "", "", "1E+132", "", "", "1.213E+21", "0", "", "", "~
## $ X.24         <chr> "", "", "", "8", "", "", "143.2", "1", "", "", "8", "", "~
## $ X.25         <chr> "", "", "", "8", "", "", "", "1.20152E+26", "", "", "8", ~
## $ X.26         <chr> "", "", "", "8", "", "", "", "5.41141E+11", "", "", "8", ~
## $ X.27         <chr> "", "", "", "1", "", "", "", "50", "", "", "1", "", "", "~
## $ X.28         <chr> "", "", "", "1000222", "", "", "", "7", "", "", "1000222"~
## $ X.29         <chr> "", "", "", "2.213E+26", "", "", "", "1E+70", "", "", "22~
## $ X.30         <chr> "", "", "", "61", "", "", "", "1", "", "", "22310000", ""~
## $ X.31         <chr> "", "", "", "620004617.8", "", "", "", "6.88889E+16", "",~
## $ X.32         <chr> "", "", "", "71.85", "", "", "", "1", "", "", "610007011.~
## $ X.33         <chr> "", "", "", "", "", "", "", "1", "", "", "103.8", "", "",~
## $ X.34         <chr> "", "", "", "", "", "", "", "1", "", "", "", "", "", "", ~
## $ X.35         <chr> "", "", "", "", "", "", "", "1", "", "", "", "", "", "", ~
## $ X.36         <chr> "", "", "", "", "", "", "", "2", "", "", "", "", "", "", ~
## $ X.37         <chr> "", "", "", "", "", "", "", "125052", "", "", "", "", "",~
## $ X.38         <chr> "", "", "", "", "", "", "", "33", "", "", "", "", "", "",~
## $ X.39         <chr> "", "", "", "", "", "", "", "33", "", "", "", "", "", "",~
## $ X.40         <chr> "", "", "", "", "", "", "", "52", "", "", "", "", "", "",~
## $ X.41         <chr> "", "", "", "", "", "", "", "44", "", "", "", "", "", "",~
## $ X.42         <chr> "", "", "", "", "", "", "", "1E+134", "", "", "", "", "",~
## $ X.43         <chr> "", "", "", "", "", "", "", "4", "", "", "", "", "", "", ~
## $ X.44         <chr> "", "", "", "", "", "", "", "4", "", "", "", "", "", "", ~
## $ X.45         <chr> "", "", "", "", "", "", "", "4", "", "", "", "", "", "", ~
## $ X.46         <chr> "", "", "", "", "", "", "", "4", "", "", "", "", "", "", ~
## $ X.47         <chr> "", "", "", "", "", "", "", "4", "", "", "", "", "", "", ~
## $ X.48         <chr> "", "", "", "", "", "", "", "1", "", "", "", "", "", "", ~
## $ X.49         <chr> "", "", "", "", "", "", "", "2000222", "", "", "", "", ""~
## $ X.50         <chr> "", "", "", "", "", "", "", "23", "", "", "", "", "", "",~
## $ X.51         <chr> "", "", "", "", "", "", "", "21420000", "", "", "", "", "~
## $ X.52         <chr> "", "", "", "", "", "", "", "39", "", "", "", "", "", "",~
## $ X.53         <chr> "", "", "", "", "", "", "", "211106049.189770496554351141~
## $ X.54         <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ X.55         <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ X.56         <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ X.57         <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ X.58         <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ X.59         <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ X.60         <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ X.61         <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ X.64         <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ X.68         <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ X.74         <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ X.75         <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ X.77         <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ X.83         <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ X.86         <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ X.89         <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ X.90         <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ X.119        <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ RSCRNINF     <dbl> 323232112, 2323235, 2626265, 3535345, 4242425, 2323235, 3~
## $ RSCRAGE      <dbl> 3.220001e+06, 1.000000e+00, 1.000000e+00, 1.000000e+00, 1~
## $ RSCRHISP     <dbl> 1.120000e+02, 2.000000e+12, 2.000006e+06, 4.121100e+12, 3~
## $ RSCRRACE     <dbl> 5, 15512010, 50, 1352, 112, 50, 915, 50, 1, 50, 11115, 1,~
```

```
## $ AGE_A        <dbl> 915, 12, 5, 111119985, 5, 1, 91994, 5, 16512011, 5, 18111~
## $ AGE_R        <dbl> 51994, 5, 13512007, 116256000000, 16111992, 15512011, 112~
## $ AGESCRN      <dbl> 5.000000e+00, 5.000000e+00, 1.200000e+01, 3.350000e+02, 1~
## $ HISP         <dbl> 2.111110e+05, 2.555121e+10, 5.000000e+00, 1.350000e+02, 1~
## $ HISPGRP      <dbl> 1.100000e+01, 2.000000e+00, 5.000000e+00, 2.150000e+02, 2~
## $ PRIMLANG1    <dbl> 1, 135, 255111, 21, 11, 2, 145, 31, 3, 2, 3, 135, 4, 125,~
## $ PRIMLANG2    <dbl> 5, 125, 12, 2, 4, 235, 415, 1, 5, 135, 5, 125, 5, 55, 135~
## $ PRIMLANG3    <dbl> 1.25000e+02, 5.50000e+01, 3.00000e+00, 2.20052e+21, 5.000~
## $ ROSCNT       <dbl> 215, 5, 5, 11235, 125, 55, 2, 105, 55, 411201, 215, 5, 55~
## $ NUMCHILD     <dbl> 2.10000e+01, 0.00000e+00, 1.25000e+02, 5.00000e+00, 3.150~
## $ HHKIDS18     <dbl> 2.00000e+00, 0.00000e+00, 5.50000e+01, 5.50000e+01, 3.100~
## $ DAUGHT918    <dbl> 2.20132e+21, 0.00000e+00, 5.00000e+00, 1.50000e+01, 2.000~
## $ SON918       <dbl> 5.500000e+01, 5.000000e+00, 0.000000e+00, 1.100000e+01, 2~
## $ NONBIOKIDS   <dbl> 5.00000e+00, 5.50000e+01, 0.00000e+00, 2.00500e+03, 5.500~
## $ MARSTAT      <dbl> 5.100000e+01, 5.000000e+00, 0.000000e+00, 2.519790e+16, 5~
## $ FMARSTAT     <dbl> 2002, 5, 5, 112, 11, 5, 199751, 55, 15, 1, 2719725, 22155~
## $ FMARIT       <dbl> 2.519765e+06, 2.015000e+03, 5.500000e+01, 2.012000e+07, 1~
## $ EVRMARRY     <dbl> 2.115100e+04, 2.719880e+12, 5.000000e+00, 2.005120e+08, 2~
## $ HPLOCALE     <dbl> 1.550000e+02, 3.500000e+01, 1.000000e+00, 3.519800e+12, 1~
## $ MANREL       <dbl> 2.002000e+04, 3.200000e+01, 0.000000e+00, 8.000000e+00, 1~
## $ GOSCHOL      <dbl> 2.00255e+05, 1.00000e+00, 5.50000e+01, 2.01511e+05, 1.993~
## $ VACA         <dbl> 0, 201551, 0, 1, 0, 142014000000, 4, 42111, 21, 411, 7199~
## $ HIGRADE      <dbl> 1.000000e+00, 0.000000e+00, 4.245510e+05, 1.200200e+04, 1~
## $ COMPGRD      <dbl> 3.200212e+10, 2.200000e+01, 0.000000e+00, 2.219800e+16, 1~
## $ DIPGED       <dbl> 1.00000e+00, 1.00000e+00, 0.00000e+00, 1.10000e+01, 1.000~
## $ EARNHS_Y     <dbl> 7.000000e+00, 1.200913e+10, 5.555500e+04, 1.000000e+00, 5~
## $ HISCHGRD     <dbl> 1.00000e+00, 1.00000e+01, 5.55550e+04, 1.61995e+11, 4.000~
## $ LSTGRADE     <dbl> 1.020150e+13, 1.234568e+07, 5.000000e+00, 9.000000e+00, 1~
## $ MYSCHOL_Y    <dbl> 1.00000e+00, 8.00000e+00, 5.55000e+02, 5.00000e+00, 4.000~
## $ HAVEDEG      <dbl> 0.00000e+00, 1.10000e+01, 1.00000e+00, 4.00000e+00, 6.000~
## $ DEGREES      <dbl> 1.000000e+00, 1.100000e+01, 5.000000e+00, 5.199710e+12, 6~
## $ EARNBA_Y     <dbl> 1, 111, 5, 8, 1, 6, 552, 1, 8, 9112, 71392122015, 0, 6, 0~
## $ EXPSCHL      <dbl> 1, 111, 0, 7, 1, 5, 55555, 6, 5, 915, 1, 55555, 5, 55555,~
## $ EXPGRADE     <dbl> 1.00000e+01, 1.00000e+00, 1.00000e+00, 5.00000e+00, 6.100~
## $ WTHPARNW     <dbl> 55555, 8, 5, 1, 71386, 41, 1, 6, 6, 4, 4, 5, 1, 5, 31, 4,~
## $ ONOWN        <dbl> 5.55555e+05, 1.00000e+00, 5.50000e+01, 1.00000e+00, 6.201~
## $ ONOWN18      <dbl> 552, 8, 95, 71, 1, 3134, 55, 6, 6, 1, 10, 1, 5, 1, 22016,~
## $ INTACT       <dbl> 55555, 11, 15, 81391112015, 1920, 1, 11555555155, 11, 11,~
## $ PARMARR      <dbl> 1.000000e+00, 8.000000e+00, 3.000000e+00, 1.000000e+00, 6~
## $ INTACT18     <dbl> 1.00000e+00, 5.00000e+00, 1.00000e+00, 2.22300e+03, 9.199~
## $ LVSIT14F     <dbl> 5.000000e+00, 4.000000e+00, 1.000000e+00, 6.000000e+00, 1~
## $ LVSIT14M     <dbl> 5.110000e+02, 5.200910e+12, 1.364000e+03, 7.200310e+12, 1~
## $ WOMRASDU     <dbl> 51555555555, 6, 820131364, 10, 555, 131387, 6, 3, 4, 2119~
## $ MOMDEGRE     <dbl> 9.500000e+01, 3.000000e+00, 2.400000e+01, 1.000000e+00, 1~
## $ MOMWORKD     <dbl> 1.500000e+01, 1.000000e+00, 1.345500e+234, 5.550000e+02, ~
## $ MOMFSTCH     <dbl> 4.000000e+00, 1.000000e+00, 1.364139e+07, 1.555150e+05, 5~
## $ MOM18        <dbl> 1.000000e+00, 3.100000e+01, 2.700000e+01, 5.522005e+06, 1~
## $ MANRASDU     <dbl> 1.000000e+00, 8.139111e+10, 3.000000e+00, 5.000000e+00, 1~
## $ R_FOSTER     <dbl> 6.000000e+00, 1.000000e+00, 3.000000e+00, 1.515111e+06, 1~
## $ EVRFSTER     <dbl> 5.136900e+04, 2.227000e+03, 1.555556e+08, 1.011510e+12, 1~
## $ MNYFSTER     <dbl> 1.201414e+08, 4.000000e+00, 3.000000e+00, 9.500000e+01, 1~
## $ DURFSTER     <dbl> 3.000000e+01, 2.201510e+12, 3.000000e+00, 1.100000e+01, 1~
## $ MENARCHE     <dbl> 5.134500e+230, 1.000000e+01, 3.000000e+00, 3.000000e+00, ~
## $ PREGNOWQ     <dbl> 13691390, 55555, 3, 5, 1, 11551555555, 44995, 55555, 4, 0~
```

```
## $ MAYBPREG     <dbl> 2.100000e+01, 5.555550e+05, 3.000000e+00, 9.000000e+00, 1~
## $ NUMPREGS     <dbl> 4.000000e+00, 5.520000e+02, 3.000000e+00, 4.000000e+00, 2~
## $ EVERPREG     <dbl> 4.000000e+00, 5.555500e+04, 3.000000e+00, 1.000000e+00, 1~
## $ CURRPREG     <dbl> 5.555556e+08, 1.000000e+00, 3.000000e+00, 1.000000e+00, 4~
## $ HOWPREG_N.y  <dbl> 4.00000e+00, 1.00000e+00, 3.00000e+00, 2.00000e+00, 1.134~
## $ HOWPREG_P.y  <dbl> 4, 5, 3, 11150, 13451389, 1, 4, 0, 55555, 515, 6, 3, 1116~
## $ NOWPRGDK.y   <dbl> 4.000000e+00, 5.500000e+01, 3.000000e+00, 4.000000e+00, 4~
## $ MOSCURRP.y   <dbl> 4.000000e+00, 5.151116e+10, 3.000000e+00, 1.134600e+235, ~
## $ NPREGS_S     <dbl> 4, 95, 3, 13451391, 6, 3200913111, 4, 15, 55555, 1, 6, 55~
## $ HASBABES     <dbl> 4, 15, 3, 46, 6, 17, 4, 4, 1, 1, 6, 5, 37, 0, 3, 13201313~
## $ NUMBABES     <dbl> 4, 8, 3, 84, 555555555, 3, 4, 7, 1, 6, 6, 51323355, 24, 8~
## $ NBABES_S     <dbl> 4.0000e+00, 1.0000e+00, 3.0000e+00, 1.2610e+03, 6.0000e+0~
## $ CMLASTLB     <dbl> 4.000000e+00, 1.000000e+00, 3.000000e+00, 6.000000e+00, 6~
## $ CMLSTPRG     <dbl> 4.000000e+00, 1.000000e+00, 3.000000e+00, 6.000000e+00, 6~
## $ CMFSTPRG     <dbl> 4, 11287, 3, 555555555, 6, 13451389, 4, 11331, 95, 3, 6, ~
## $ CMPG1BEG     <dbl> 4.000000e+00, 8.000000e+00, 3.000000e+00, 6.000000e+00, 6~
## $ NPLACED      <dbl> 4, 3200712871, 3, 6, 6, 1, 4, 7, 3, 13451389, 6, 1, 3, 55~
## $ NDIED        <dbl> 4.0000e+00, 1.4000e+01, 3.0000e+00, 6.0000e+00, 6.0000e+0~
## $ NADOPTV      <dbl> 4, 3, 3, 6, 6, 555555555, 4, 12, 1, 12, 6, 1, 3, 21, 1, 6~
## $ TOTPLACD     <dbl> 4.0000e+00, 1.0000e+00, 3.0000e+00, 6.0000e+00, 6.0000e+0~
## $ OTHERKID     <dbl> 4.000000e+00, 8.000000e+00, 3.000000e+00, 6.000000e+00, 6~
## $ NOTHRKID     <dbl> 4.00000e+00, 1.13460e+235, 3.00000e+00, 6.00000e+00, 6.00~
## $ SEXOTHKD     <dbl> 4, 13451391, 3, 6, 6, 1, 4, 3, 7, 555555555, 6, 725, 3, 3~
## $ RELOTHKD     <dbl> 4, 46, 3, 6, 6, 1, 4, 3, 4200913121, 4, 6, 27, 3, 6, 1, 6~
## $ ADPTOTKD     <dbl> 4.0000e+00, 3.6000e+01, 1.3451e+187, 6.0000e+00, 6.0000e+~
## $ TRYADOPT     <dbl> 4.000000e+00, 1.309000e+03, 1.111110e+27, 6.000000e+00, 6~
## $ TRYEITHR     <dbl> 1.345100e+183, 8.000000e+00, 0.000000e+00, 6.000000e+00, ~
## $ STILHERE     <dbl> 1.111110e+21, 8.000000e+00, 1.000010e+26, 6.000000e+00, 6~
## $ DATKDCAM_Y   <dbl> 0, 555515555, 2, 6, 6, 1, 4, 1, 7, 4, 6, 5, 3, 11155, 1, ~
## $ OTHKDFOS     <dbl> 1.00000e+20, 8.00000e+00, 3.00000e+00, 6.00000e+00, 6.000~
## $ OKDDOB_Y     <dbl> 1, 8, 3, 6, 6, 1, 4, 1, 13451391, 1, 6, 0, 3, 5, 1, 6, 5,~
## $ OKBORNUS     <dbl> 4, 8, 93, 6, 6, 1, 4, 1, 46, 1, 6, 0, 3, 5555555555, 1, 6~
## $ OKDISABL1    <dbl> 5, 8, 5, 6, 6, 1, 4, 1, 1995, 1, 6, 5555, 3, 5, 19, 6, 5,~
## $ OKDISABL2    <dbl> 8, 8, 15, 6, 6, 1, 4, 1, 420091312, 1, 6, 5, 3, 55223455,~
## $ SEXOTHKD2    <dbl> 8, 8, 11155551, 6, 6, 1, 4, 4, 3, 1, 6, 21, 3, 3, 19, 6, ~
## $ RELOTHKD2    <dbl> 1.0000e+00, 8.0000e+00, 5.5000e+01, 6.0000e+00, 6.0000e+0~
## $ ADPTOTKD2    <dbl> 0, 8, 5, 6, 6, 1, 4, 1, 3, 1, 6, 1111115, 3, 1, 19, 6, 5,~
## $ TRYADOPT2    <dbl> 1, 8, 31, 4, 6, 1, 4, 1, 7, 1, 6, 55, 3, 2, 19, 6, 5, 6, ~
## $ TRYEITHR2    <dbl> 555556000000, 8, 3, 6, 6, 1, 4, 4, 155555555, 1, 6, 5, 3,~
## $ STILHERE2    <dbl> 5, 8, 1, 4, 6, 1, 4, 4, 3, 1, 6, 5, 3, 6, 19, 6, 5, 6, 55~
## $ DATKDCAM_Y2  <dbl> 9, 8, 53, 6, 6, 1, 4, 1, 7, 1, 6, 55, 3, 8, 19, 6, 5, 6, ~
## $ OTHKDFOS2    <dbl> 0.00000e+00, 8.00000e+00, 1.00000e+00, 4.00000e+00, 6.000~
## $ OKDDOB_Y2    <dbl> 2125, 8, 6, 6, 6, 1, 4, 1, 7, 1, 6, 1, 3, 3, 19, 6, 5, 6,~
## $ OKBORNUS2    <dbl> 5.000000e+00, 8.000000e+00, 1.000000e+00, 4.000000e+00, 6~
## $ OKDISABL5    <dbl> 1.000000e+00, 8.000000e+00, 6.000000e+00, 6.000000e+00, 6~
## $ OKDISABL6    <dbl> 2125, 8, 15, 4, 6, 1, 4, 1, 3, 1, 6, 1, 1, 1, 19, 6, 5, 6~
## $ SEXOTHKD3    <dbl> 11, 8, 991555, 6, 6, 4, 4, 7, 7, 1, 6, 3, 1, 711, 19, 6, ~
## $ RELOTHKD3    <dbl> 5125, 8, 5, 4, 6, 7, 4, 7, 3, 1, 6, 8, 4, 3, 19, 6, 5, 6,~
## $ ADPTOTKD3    <dbl> 8.5555e+04, 8.0000e+00, 5.0000e+00, 6.0000e+00, 6.0000e+0~
## $ TRYADOPT3    <dbl> 2.52000e+02, 8.00000e+00, 1.50000e+01, 4.00000e+00, 6.000~
## $ TRYEITHR3    <dbl> 5.000000e+00, 8.000000e+00, 5.000000e+00, 6.000000e+00, 6~
## $ STILHERE3    <dbl> 5.000000e+00, 8.000000e+00, 5.515556e+09, 4.000000e+00, 6~
## $ DATKDCAM_Y3  <dbl> 4.5e+01, 8.0e+00, 5.0e+00, 6.0e+00, 6.0e+00, 1.0e+00, 1.0~
## $ SEXOTHKD7    <chr> "1", "1", "3E+23", "5", "1", "5", "12001", "1", "3", "5",~
```

```
## $ OKDISABL30   <chr> "2530", "1", "42420000173", "245", "5", "0", "1", "19", "~
## $ SEXOTHKD9    <chr> "20072013", "1", "820002847", "355121144", "1151", "0", "~
## $ ADPTOTKD9    <chr> "11", "2", "", "111144", "11", "5", "55555555", "2", "1.1~
## $ TRYADOPT10   <chr> "0", "32", "", "5", "2", "1", "4", "1", "1", "3", "3E+23"~
## $ OKBORNUS10   <chr> "1.12889E+18", "1", "", "1", "4E+23", "23", "19202325", "~
## $ OKDISABL37   <chr> "1", "1.21323E+17", "", "1", "3", "1", "1.9972E+15", "8",~
## $ OKDISABL38   <chr> "2", "1", "", "1", "3", "2E+23", "18202224", "910", "25",~
## $ TRYEITHR11   <chr> "11", "45", "", "5.55556E+12", "202227", "22", "1.99819E+~
## $ STILHERE11   <chr> "11", "2", "", "313", "1.9932E+11", "0", "1", "5", "1", "~
## $ DATKDCAM_Y11 <chr> "11", "1", "", "313", "202126", "0", "121998", "515", "0"~
## $ OKBORNUS11   <chr> "2", "5", "", "9.22222E+42", "551", "1.11222E+18", "1995"~
## $ OKDISABL41   <chr> "222", "15", "", "2", "551", "1.51139E+11", "1211", "735"~
## $ SEXOTHKD12   <chr> "41270000", "11", "", "0", "55", "4", "0", "11", "995", "~
## $ RELOTHKD12   <chr> "93", "2.11656E+11", "", "2", "11", "1E+66", "1.19972E+42~
## $ ADPTOTKD12   <chr> "610005968.1", "3", "", "20211", "121995", "1.11114E+14",~
## $ TRYADOPT12   <chr> "75.64", "3", "", "20022005", "22", "2", "1E+69", "1", "1~
## $ TRYEITHR12   <chr> "", "1", "", "2224", "5", "1", "1.12889E+18", "0", "3", "~
## $ STILHERE12   <chr> "", "1", "", "20022004", "1995", "2", "1", "0", "5", "2E+~
## $ DATKDCAM_Y12 <chr> "", "1", "", "2123", "2211", "5", "2", "5551", "5", "4", ~
## $ OTHKDFOS12   <chr> "", "0", "", "51", "1", "1", "1", "65", "15", "3", "4", "~
## $ OKDDOB_Y12   <chr> "", "55", "", "51", "0", "2", "2", "45", "5", "0", "8E+69~
## $ OKBORNUS12   <chr> "", "55555555", "", "55", "1.19932E+26", "5", "1", "1", "~
## $ OKDISABL45   <chr> "", "1", "", "56", "3.31139E+14", "1", "11732222", "1", "~
## $ OKDISABL46   <chr> "", "2", "", "1120022234", "6", "1311", "3222", "0", "551~
## $ SEXOTHKD13   <chr> "", "1.11112E+12", "", "4411", "6E+72", "0", "3222", "211~
## $ RELOTHKD13   <chr> "", "1", "", "102005", "188888811", "1", "2222", "115135"~
## $ ADPTOTKD13   <chr> "", "3", "", "2011", "2", "2000222", "3222", "11", "6", "~
## $ TRYADOPT13   <chr> "", "1", "", "1", "1", "20", "22", "3120", "5", "555", "4~
## $ TRYEITHR13   <chr> "", "5", "", "2.43031E+13", "4", "4.216E+20", "4", "5", "~
## $ STILHERE13   <chr> "", "5", "", "4111", "2", "66.84", "4", "3", "1", "5555",~
## $ DATKDCAM_Y13 <chr> "", "5.55556E+30", "", "1", "1078661", "", "4", "3", "111~
## $ OTHKDFOS13   <chr> "", "2", "", "2", "552", "", "4", "1", "15", "3.1201E+15"~
## $ OKDDOB_Y13   <chr> "", "2E+23", "", "3.20022E+26", "552", "", "4", "1", "1",~
## $ OKBORNUS13   <chr> "", "0", "", "1.91139E+14", "661", "", "1", "1", "111", "~
## $ OKDISABL49   <chr> "", "0", "", "6", "662", "", "1000222", "1", "15", "4111"~
## $ OKDISABL50   <chr> "", "0", "", "7E+69", "0", "", "22", "51", "1595", "0", "~
## $ SEXOTHKD14   <chr> "", "0", "", "1", "8", "", "4E+26", "2", "1", "1", "2", "~
## $ RELOTHKD14   <chr> "", "11", "", "188888811", "8", "", "820005755.5811421221~
## $ ADPTOTKD14   <chr> "", "0", "", "2", "2", "", "", "21", "17", "61138811123",~
## $ TRYADOPT14   <chr> "", "1", "", "1", "222", "", "", "5.51511E+12", "16", "29~
## $ TRYEITHR14   <chr> "", "1.20152E+26", "", "4", "22", "", "", "1", "7512314",~
## $ STILHERE14   <chr> "", "41139121123", "", "1", "3.231E+21", "", "", "5", "11~
## $ DATKDCAM_Y14 <chr> "", "9", "", "4", "86.84", "", "", "3995", "11115", "21",~
## $ OTHKDFOS14   <chr> "", "4", "", "2", "", "", "", "5", "1", "1.30889E+18", ""~
## $ OKDDOB_Y14   <chr> "", "3E+69", "", "115022", "", "", "", "5", "1.31112E+17"~
## $ OKBORNUS14   <chr> "", "21", "", "32", "", "", "", "5.15556E+25", "1", "2", ~
## $ OKDISABL53   <chr> "", "588888821", "", "32", "", "", "", "11", "0", "6", ""~
## $ OKDISABL54   <chr> "", "6", "", "22", "", "", "", "3", "0", "13095522", "", ~
## $ SEXOTHKD15   <chr> "", "1", "", "42", "", "", "", "1E+23", "5555", "3322", "~
## $ RELOTHKD15   <chr> "", "6", "", "0", "", "", "", "0", "5", "3322", "", "", "~
## $ ADPTOTKD15   <chr> "", "1", "", "8", "", "", "", "0", "55", "5522", "", "", ~
## $ TRYADOPT15   <chr> "", "6", "", "8", "", "", "", "0", "6", "4422", "", "", "~
## $ TRYEITHR15   <chr> "", "6", "", "2", "", "", "", "0", "2", "3E+132", "", "",~
## $ STILHERE15   <chr> "", "1287", "", "222", "", "", "", "22", "1", "1", "", ""~
```

```
## $ DATKDCAM_Y15 <chr> "", "11", "", "21", "", "", "", "0", "5311111", "1", "", ~
## $ OTHKDFOS15   <chr> "", "1", "", "1.221E+21", "", "", "", "0", "5", "8", "", ~
## $ OKDDOB_Y15   <chr> "", "1", "", "57.56", "", "", "", "0", "115111511", "8", ~
## $ OKBORNUS15   <chr> "", "1", "", "", "", "", "", "1", "3116", "2E+19", "", ""~
## $ OKDISABL57   <chr> "", "1", "", "", "", "", "", "1.21202E+18", "1195511511",~
## $ OKDISABL58   <chr> "", "1", "", "", "", "", "", "1.91139E+11", "6", "22", ""~
## $ SEXOTHKD16   <chr> "", "1.11E+21", "", "", "", "", "", "24", "6", "8E+23", "~
## $ RELOTHKD16   <chr> "", "1000222", "", "", "", "", "", "4", "2", "93", "", ""~
## $ ADPTOTKD16   <chr> "", "23", "", "", "", "", "", "3E+66", "2", "710003170.84~
## $ TRYADOPT16   <chr> "", "43320000", "", "", "", "", "", "11", "2", "", "", ""~
## $ TRYEITHR16   <chr> "", "86", "", "", "", "", "", "41", "2", "", "", "", "", ~
## $ STILHERE16   <chr> "", "520003520.4", "", "", "", "", "", "688888811", "1511~
## $ DATKDCAM_Y16 <chr> "", "56.2", "", "", "", "", "", "2", "555511", "", "", ""~
## $ OTHKDFOS16   <chr> "", "", "", "", "", "", "", "5", "1", "", "", "", "", "",~
## $ OKDDOB_Y16   <chr> "", "", "", "", "", "", "", "1", "2", "", "", "", "", "",~
## $ OKBORNUS16   <chr> "", "", "", "", "", "", "", "1", "2", "", "", "", "", "",~
## $ OKDISABL61   <chr> "", "", "", "", "", "", "", "5", "0", "", "", "", "", "",~
## $ OKDISABL62   <chr> "", "", "", "", "", "", "", "1", "18", "", "", "", "", ""~
## $ SEXOTHKD17   <chr> "", "", "", "", "", "", "", "1", "5", "", "", "", "", "",~
## $ RELOTHKD17   <chr> "", "", "", "", "", "", "", "5", "33", "", "", "", "", ""~
## $ ADPTOTKD17   <chr> "", "", "", "", "", "", "", "2", "5", "", "", "", "", "",~
## $ TRYADOPT17   <chr> "", "", "", "", "", "", "", "5", "5.55552E+12", "", "", "~
## $ TRYEITHR17   <chr> "", "", "", "", "", "", "", "1331", "1", "", "", "", "", ~
## $ STILHERE17   <chr> "", "", "", "", "", "", "", "0", "2", "", "", "", "", "",~
## $ DATKDCAM_Y17 <chr> "", "", "", "", "", "", "", "8", "315", "", "", "", "", "~
## $ OTHKDFOS17   <chr> "", "", "", "", "", "", "", "8", "5", "", "", "", "", "",~
## $ OKDDOB_Y17   <chr> "", "", "", "", "", "", "", "8", "5.55556E+30", "", "", "~
## $ OKBORNUS17   <chr> "", "", "", "", "", "", "", "1", "4", "", "", "", "", "",~
## $ OKDISABL65   <chr> "", "", "", "", "", "", "", "2000222", "3E+23", "", "", "~
## $ OKDISABL66   <chr> "", "", "", "", "", "", "", "22", "0", "", "", "", "", ""~
## $ SEXOTHKD18   <chr> "", "", "", "", "", "", "", "8E+36", "0", "", "", "", "",~
## $ RELOTHKD18   <chr> "", "", "", "", "", "", "", "59.18", "0", "", "", "", "",~
## $ ADPTOTKD18   <chr> "", "", "", "", "", "", "", "", "0", "", "", "", "", "", ~
## $ TRYADOPT18   <chr> "", "", "", "", "", "", "", "", "11", "", "", "", "", "",~
## $ TRYEITHR18   <chr> "", "", "", "", "", "", "", "", "0", "", "", "", "", "", ~
## $ STILHERE18   <chr> "", "", "", "", "", "", "", "", "1", "", "", "", "", "", ~
## $ DATKDCAM_Y18 <chr> "", "", "", "", "", "", "", "", "1.20152E+26", "", "", ""~
## $ OTHKDFOS18   <chr> "", "", "", "", "", "", "", "", "3.92139E+11", "", "", ""~
## $ OKDDOB_Y18   <chr> "", "", "", "", "", "", "", "", "7", "", "", "", "", "", ~
## $ OKBORNUS18   <chr> "", "", "", "", "", "", "", "", "0", "", "", "", "", "", ~
## $ OKDISABL69   <chr> "", "", "", "", "", "", "", "", "47", "", "", "", "", "",~
## $ OKDISABL70   <chr> "", "", "", "", "", "", "", "", "6E+69", "", "", "", "", ~
## $ SEXOTHKD19   <chr> "", "", "", "", "", "", "", "", "31", "", "", "", "", "",~
## $ RELOTHKD19   <chr> "", "", "", "", "", "", "", "", "621888811", "", "", "", ~
## $ ADPTOTKD19   <chr> "", "", "", "", "", "", "", "", "1", "", "", "", "", "", ~
## $ TRYADOPT19   <chr> "", "", "", "", "", "", "", "", "5", "", "", "", "", "", ~
## $ TRYEITHR19   <chr> "", "", "", "", "", "", "", "", "1", "", "", "", "", "", ~
## $ STILHERE19   <chr> "", "", "", "", "", "", "", "", "1", "", "", "", "", "", ~
## $ DATKDCAM_Y19 <chr> "", "", "", "", "", "", "", "", "5", "", "", "", "", "", ~
## $ OTHKDFOS19   <chr> "", "", "", "", "", "", "", "", "1", "", "", "", "", "", ~
## $ OKDDOB_Y19   <chr> "", "", "", "", "", "", "", "", "1", "", "", "", "", "", ~
## $ OKBORNUS19   <chr> "", "", "", "", "", "", "", "", "5", "", "", "", "", "", ~
## $ OKDISABL73   <chr> "", "", "", "", "", "", "", "", "1", "", "", "", "", "", ~
## $ OKDISABL74   <chr> "", "", "", "", "", "", "", "", "1312", "", "", "", "", "~
```

```
## $ SEXOTHKD20   <chr> "", "", "", "", "", "", "", "", "0", "", "", "", "", "", ~
## $ TRYADOPT20   <chr> "", "", "", "", "", "", "", "", "1000222", "", "", "", ""~
## $ OTHKDFOS20   <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ SEEKADPT     <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ CONTAGEM     <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ KNOWADPT     <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ APROCESS2    <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ TIMESMAR     <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ AGEMARHX     <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ HXAGEMAR     <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
## $ ENGAGHX2     <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "~
```

```
#str(preg_resp_merged)
```

**Questions for future steps.**

**What do you not know how to do right now that you need to learn to import and cleanup your dataset?** I believe I know everything I need to know right know in order to import and cleanup my dataset. I don't know how to merge all 5 of my datasets since they represent different forms of information pertaining to women's fertility, but I'm not sure if that's needed since it might be nice and more beneficial to deeper diver into each set of data depending on my problem questions.

**What information is not self-evident?**

**Discuss how you plan to uncover new information in the data that is not self-evident.** I think my next steps for each dataset (1 solo & 2 merged) is to analyze each of their variables and uncover how I can recode them and/or generate new columns based on existing ones to find new information. There are already many variables to investigate, but there is so much more we can learn by generating new variables that will build on already existing details & info.

I also want to look into the normality of the dataset variables, and also investigate the relationships between any of the variables to ensure there is no multicollinearity.

**What are different ways you could look at this data?**

**What are different ways you could look at this data to answer the questions you want to answer?**

**How could you summarize your data to answer key questions?** One way I want to look at the data is by building aggregations out of it, especially for the fertility rate and country population merged dataset. I want to look into it country-wise and year-wise. It will allow me to visualize any trends (or lack there of) over the 36 years of data, which spans from the 1980's to the 2010's. By looking at the data year-wise, I want to understand how fertility rate has changed with the massive population growth in the world. With more people existing in the world, there are going to be more people assessing their reproductive abilities and depending on the outcome, it can have an impact on the fertility rate of a country/year.

The fertility_df only has 100 rows of data so it is quite smaller than the other 2 datasets, but it includes some great information on a participant and their given symptoms/life habits in relation to a 'Normal' or 'Altered' diagnosis of fertility. I want to build logistic regression models on this data to uncover the variables which have the greatest effect on the diagnosis of a patient/study participant. I am trying to uncover the factors that play into one's fertility, and I think this dataset will be really useful for that information.

I have a few questions regarding non-traditional methods of conception, i.e. adoption, IVF, etc. The merged preg & resp dataset provides information regarding a participant's birth control & conception methods even if they are not pregnant, which could show that they are having trouble conceiving. Therefore, this dataset will be really great for looking into those questions in how non-traditional methods are included in fertility data and information. I want to look at the distributions of these variables and understand how the sample can be generalized to the population of women trying to get pregnant. I also want to subset the data by women using traditional vs. non-traditional methods and do data comparisons to dive into how their fertility cases differ or are similar.

**How do you plan to slice and dice the data?**

**Do you plan to slice and dice the data in different ways, create new variables, or join separate data frames to create new summary information? Explain.** Already mentioned above :)

I created 2 merged datasets: 1. Combined fertility_rate_df & country_pop_df 2. Combined preg & fem_resp dataframes

**What types of plots and tables will help you to illustrate the findings to your questions?**

1. What is the weight of women's reproductive health in influencing a couple's ability to have children?

- Frequency tables
- Pie charts

2. What is the current difference in birth rates from one country to another?

- Bar charts with country code on the x-axis
- Histogram of birth rates for each year represented in the merged dataset

3. What is the average age for women to try to start having children?

- Aggregation tables
- Summary statistics

4. How have non-traditional methods of having children influenced birth rate, such as adoption/IVF/etc?

- Regression models, residual plots
- Correlation plots

5. What resources are provided to people who are experiencing issues with infertility?

- Subset table focused on resources mentioned in the preg & resp merged dataset
- Count tables for number of people actually accessing and utilizing those resources
- Bar charts for showing ranking of resources in terms of actual usage and popularity

6. What role does proper sex education play in fertility and reproductive health?

- Regression models, residual plots
- Correlation plots

7. Does the current calculation of birth rate account for non-traditional methods of child delivery?

- Summary stastistics
- Aggregation of birth rate by method of conception – querying

8. What are the key factors that play a role in one's fertility, men and women?

- Regression, residual plots
- Correlation plots

**Do you plan on incorporating any machine learning techniques to answer your research questions? Explain.**

K-Means Clustering would be interesting to use to cluster the various countries in the rate_pop_merged dataset by their fertility rates to understand which are more similar and also different from eachother. It will give a global perspective and allow for more understanding on how the similar countries' characteristics play into/affect their fertility rates. I have never given much thought to how a country itself can affect its citizens' fertility, and by visualizing/grouping countries based on their fertility rates, I would hopefully be able to understand this fact in more detail.

I could also potentially use the machine learning technique of K-Nearest Neighbors to classify new records into the groupings of either being fertile or infertile, in terms of ease of conception. I would have to deliberate on which variables to include for the groupings, but I think this would be very interesting for seeing how fertility can be precited for an individual baed on the values of the given prediction variables.

**Questions for future steps.**

1. How are machine learning techniques applied using R?
2. How do you create aggregation/summary tables effectively in R?
3. What is the best way to rearrange data? What ideas/thinking should go into arranging data in an usable and valuable manner?

# Final Project Step 3

**Introduction.**

Infertility is a problem that affects every individual, male or female, in every part of the world; it is a global issue. Unfortunately, even with being such a major problem, there is still minimal information and evidence around what contributes to infertility and how people can possibly improve their fertility and/or prevent infertility.

It is a topic that desires more visibility and research, and I think statistical analysis and investigation can make great strides in the progress towards increasing one's fertility awareness and condition. Demographic factors, health factors and environmental factors seem to play a role in fertility and possibly infertility, and with this effort of building a model, I aimed to answer the question of: "what factors play a significant role in predicting an individual's possibility of being infertile?". I wanted to use fertility diagnosis as a categorical outcome variable for my model and apply logistic regression to understand which variables can help predict this binary outcome (fertile or infertile).

**The problem statement you addressed.**

The ambiguity around people's fertility, until they are at the time in their lives to start having children, is a concern that should be addressed as early as possible by identifying the factors that play into fertility and developing detection methods for infertility to be able to provide proper aid and support.

**How you addressed this problem statement**

I started out with five datasets for addressing my problem statement: country_pop_df, fem_resp, fertility_rate_df, fertility_df & preg. 1. country_pop_df: Countries' population numbers from 1960-2016 2. fertility_rate_df: Countries' fertility rates from 1960-2016 3. fem_resp: Female respondents' information from a survey of females around their health, life events, possible previous pregnancies, etc. 4. preg: NSFG survey data for pregnant females 5. fertility_df: - 100 volunteers provided a semen sample that was analyzed according to the WHO 2010 criteria deciding whether the individual was fertile

I merged two sets of datasets: 1. country_pop_df & fertility_rate_df (countries' population numbers & fertility rates in one dataset) 2. fem_resp & preg (the females' survey responses and pregnancy information in one dataset)

By merging the datasets, it allowed for me to have all applicable information in one place. It made it easier to conduct summary statistics and analyze the variables in the dataset to draw insights such as relationships between variables, distributions, and finally predictive modeling.

**Analysis**

There are a few models that I would want to build for answering the questions around my problem statement.

One of those models would be a logistic model, using the fertility_df & merged 'preg_resp_merged' datasets. I would utilize an outcome variable which depicts fertility as a binary categorical variable with values of 'Fertile' & 'Infertile'. There are many variables in both of those datsets, and so, I would attempt to uncover which variables hold the greatest predictive measure in identifying whether one is infertile. Fertility would be my baseline for the model, as I want to uncover more information around people's infertility since that is the main problem at hand. From building these logistic models with various explanatory variables and combinations, I would run comparisons between them to understand their varying strengths and precisions of fit. I want to ensure the models are providing accurate predictions, and it would take some trial and effort to uncover the variables which have the greatest influence on uncovering a solution to my problem statement.

Also, with the merged 'rate_pop_merged' dataset, I would suggest utilizing a time series analysis model which would allow us to draw meaningful insights regarding fertility rate in terms of population growth over time for the countries in the world. This dataset holds information around the years from 1960-2016 (I subsetted to be from 1980-2016). From having such a large range of dates to apply our model on, we can use previous data to predict future results for countries and their fertility rates, which could also help us understand how the people in their countries will be with their fertility and how resources can be allocated to plan accordingly.

A big focus of mine is on early detection and prevention of infertility, and I believe both of those models will help with those principles at both the individual characteristic level and the global resource level.

**Implications**

The insights from my analysis can help the target audience of post-puberty humans by giving them understanding and knowledge around their fertility health and how they can hopefully take proper action if infertility is a concern that is identified for them.

I read an article about how machine learning has been used on patients who think they may have dementia by comparing their brain scans with a database of dementia patients' scans to diagnose and/or decide whether the patients have a possibility of developing the neurological disorder. Therefore, from the results of this technique, the patients can be provided with proper care to ensure that they are prepared for an onset of dementia and can be as ready as possible for what's to come.

The target audience of this analysis will hopefully be provided with a similar outcome as these dementia patients. They will be able to understand if there is a possibility of being infertile and from there, they can be guided on how to prepare and possibly increase their chances/odds of having children. If the degree of their fertility is not ideal for natural conception,they can also start looking into other conception methods such as adoption and IVF, earlier than normal, since they are options that require time, money, and energy.

**Limitations**

Time is obviously a limitation of this analysis :) with more time, I would hope to take this analysis from data exploration to model building to analysis of results and possibly repetition of any steps after and in-between. There is a lot of data to explore in relation to fertility, and I would love to have the time to spend on diving into this subject and how statistics can be utilized to increase awareness, detection and prevention.

I would like to gain more education on incorporating categorical variables into modeling and exactly how to interpret them when put into models' output. I believe they can hold huge weight in their factors/coding for the problem or question at hand, and I would hate to leave them out just because I don't have a ton of experience with using and understanding their impact on the outcome variable at each level of the category.

Fertility is a major issue and a global issue as mentioned earlier. I worry that any one data set or even two merged together would not be enough to generalize the sample model results to the population. In doing this analysis, I would want to ensure that I can address the problem at a higher degree, and this could only come from having proper data that is representative of the world's population and the factors that affect both men and women.

Given the health nature of this issue with infertility, I would love to see and explore how artificial intelligence could be incorporated into the building out of the models and their predictions. It is an incredibly powerful too that is making waves in how computers can be trained to think and behave like humans. Doctors and scientists have been attempting to dive into infertility for many years, and I think with the help of smart technology and behavioral modeling, there could be major insights gained around one's fertility and possible infertility.

**Concluding Remarks**

Detection and prevention are the key words in this analysis of fertility and the modeling recommendations. There is major predictive power that can be yielded from building models on theses datasets regarding females' health & wellness and countries' fertility numbers. In examining the data and diving into the key factors that influence fertility and how countries can better prepare for their populations, we can be prepared as a nation for issues around people being able to have children when they are ultimately ready to.

There is major stress, pain and exhaustion that is caused from the battle of infertility and trying to find ways to conceive children when it is not in the body's best favor. With modeling and predictive insights, we can build a better global system that is educated around the issue of infertility and in how people can be prepared for their futures as early and as soon as possible. We all deserve to know what is happening in our body, and the earlier we can detect any problems against healthy living, the better.

Here's to a better reproductive future for our world and its people!