

ASS#3.doc.Rmd

Alecia Clarke

2023-07-02

Import Nat2021US - January 2021 - Heading Only Excel file

```
library(readxl)
Nat2021US_January_2021_Headings_Only <- read_excel("~/SGU MPH Program/Summer 2023/PUBH 835 Practical Da
```

```
## Warning: Expecting numeric in W1057 / R1057C23: got '1 1'
## Warning: Expecting numeric in W10021 / R10021C23: got '1 1'
## Warning: Expecting numeric in W10558 / R10558C23: got '1 1'
## Warning: Expecting numeric in W12418 / R12418C23: got '1 1'
## Warning: Expecting numeric in W12574 / R12574C23: got '1 1'
## Warning: Expecting numeric in W13010 / R13010C23: got '1 1'
## Warning: Expecting numeric in W13832 / R13832C23: got '1 1'
## Warning: Expecting numeric in W14101 / R14101C23: got '1 1'
## Warning: Expecting numeric in W14124 / R14124C23: got '1 1'
## Warning: Expecting numeric in H20118 / R20118C8: got '1 31'
## Warning: Expecting numeric in H28585 / R28585C8: got '1 31'
## Warning: Expecting numeric in H38570 / R38570C8: got '1 28'
## Warning: Expecting numeric in FI38570 / R38570C165: got '1 9'
## Warning: Expecting numeric in H40008 / R40008C8: got '1 29'
## Warning: Expecting numeric in H41062 / R41062C8: got '1 30'
## Warning: Expecting numeric in H45100 / R45100C8: got '1 26'
```

Warning: Expecting numeric in H45232 / R45232C8: got '1 31'

Warning: Expecting numeric in H45240 / R45240C8: got '1 26'

Warning: Expecting numeric in H45992 / R45992C8: got '1 28'

Warning: Expecting numeric in H46023 / R46023C8: got '1 28'

Warning: Expecting numeric in FI46023 / R46023C165: got '1 9'

Warning: Expecting numeric in H46060 / R46060C8: got '1 28'

Warning: Expecting numeric in FI46060 / R46060C165: got '1 9'

Warning: Expecting numeric in H46076 / R46076C8: got '1 28'

Warning: Expecting numeric in H46085 / R46085C8: got '1 28'

Warning: Expecting numeric in FI46085 / R46085C165: got '1 9'

Warning: Expecting numeric in H46090 / R46090C8: got '1 28'

Warning: Expecting numeric in H46098 / R46098C8: got '1 28'

Warning: Expecting numeric in W46645 / R46645C23: got '1 1'

Warning: Expecting numeric in W46966 / R46966C23: got '1 1'

Warning: Expecting numeric in W47296 / R47296C23: got '1 1'

Warning: Expecting numeric in W47530 / R47530C23: got '1 1'

Warning: Expecting numeric in W47566 / R47566C23: got '1 1'

Warning: Expecting numeric in W47603 / R47603C23: got '1 1'

Warning: Expecting numeric in W47788 / R47788C23: got '1 1'

Warning: Expecting numeric in W49454 / R49454C23: got '1 1'

Warning: Expecting numeric in W49623 / R49623C23: got '1 1'

Warning: Expecting numeric in W50009 / R50009C23: got '1 1'

Warning: Expecting numeric in W50825 / R50825C23: got '1 1'

Warning: Expecting numeric in W53463 / R53463C23: got '1 1'

```
## Warning: Expecting numeric in W53511 / R53511C23: got '1 1'
## Warning: Expecting numeric in W53774 / R53774C23: got '1 1'
## Warning: Expecting numeric in W53911 / R53911C23: got '1 1'
## Warning: Expecting numeric in W54704 / R54704C23: got '1 1'
## Warning: Expecting numeric in W54914 / R54914C23: got '1 1'
## Warning: Expecting numeric in H58468 / R58468C8: got '1 26'
## Warning: Expecting numeric in W58962 / R58962C23: got '1 1'
## Warning: Expecting numeric in W68574 / R68574C23: got '1 1'
## Warning: Expecting numeric in W70731 / R70731C23: got '1 1'
## Warning: Expecting numeric in W71449 / R71449C23: got '1 1'
## Warning: Expecting numeric in W71452 / R71452C23: got '1 1'
## Warning: Expecting numeric in FI71635 / R71635C165: got '1 9'
## Warning: Expecting numeric in FI71641 / R71641C165: got '1 9'
## Warning: Expecting numeric in W73818 / R73818C23: got '1 1'
## Warning: Expecting numeric in W77596 / R77596C23: got '1 1'
## Warning: Expecting numeric in W77767 / R77767C23: got '1 1'
## Warning: Expecting numeric in W78442 / R78442C23: got '1 1'
## Warning: Expecting numeric in W79837 / R79837C23: got '1 1'
## Warning: Expecting numeric in W80562 / R80562C23: got '1 1'
## Warning: Expecting numeric in W81289 / R81289C23: got '1 1'
```

```
View(Nat2021US_January_2021_Headings_Only)
```

Rename the dataset so that it will be easier to work with

```
NatJan2021 <- Nat2021US_January_2021_Headings_Only
NatJan2021
```

```
## # A tibble: 81,393 x 219
##   'Birth Year' 'Birth Month' 'Time of Birth' 'Day of Birth' 'Birth Place'
##   <chr>          <dbl>          <dbl>          <dbl>          <dbl>
## 1 2021              1              636              7              1
## 2 2021              1              259              7              1
## 3 2021              1              223              1              1
## 4 2021              1              241              1              1
## 5 2021              1              503              1              1
## 6 2021              1             2341              7              1
## 7 2021              1             1800              7              1
## 8 2021              1              652              1              1
## 9 2021              1              227              6              1
## 10 2021             1             2151              6              1
## # i 81,383 more rows
## # i 214 more variables: 'Reporting Birth Place' <dbl>,
## #   'Birth Facility Recode' <dbl>, 'Mother's Single Year of Age' <dbl>,
## #   'Mother's Age Recode 14' <dbl>, 'Mother's Age Recode 9' <dbl>,
## #   'Mother's Nativity' <dbl>, 'Residence Status' <dbl>,
## #   'Mother's Race Recode 31' <dbl>, 'Mother's Race Recode 6' <dbl>,
## #   'Mother's Race Recode 15' <dbl>, 'Mother Race Imputed Flag' <dbl>, ...
```

Load the dplyr package to remove multiple columns

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

Use the dplyr package to remove multiple columns for the desired data to remain

```
NatJan2021.1 <- NatJan2021 %>% select(-c(`Birth Year`, `Birth Month`, `Time of Birth`, `Day of Birth`,
NatJan2021.1
```

```
## # A tibble: 81,393 x 211
##   'Mother's Single Year of Age' 'Mother's Age Recode 9' 'Mother's Nativity'
##   <dbl>          <dbl>          <dbl>
## 1              22              3              1
## 2              31              5              1
## 3              29              4              1
## 4              39              6              1
## 5              20              3              2
## 6              29              4              1
## 7              23              3              1
## 8              34              5              1
```

```
## 9 19 2 1
## 10 25 4 1
## # i 81,383 more rows
## # i 208 more variables: 'Residence Status' <dbl>,
## # 'Mother's Race Recode 31' <dbl>, 'Mother's Race Recode 6' <dbl>,
## # 'Mother's Race Recode 15' <dbl>, 'Mother Race Imputed Flag' <dbl>,
## # 'Mother's Hispanic Origin' <dbl>, 'Mother's Hispanic Origin Recode' <dbl>,
## # 'Reporting Mother's Origin' <dbl>, 'Mother's Race/Hispanic Origin' <dbl>,
## # 'Paternity Acknowledged' <chr>, 'Marital Status' <dbl>, ...
```

```
NatJan2021.2 <- NatJan2021.1 %>% select(-c('Mother's Age Recode 9', 'Mother's Nativity', 'Residence Sta
NatJan2021.2
```

```
## # A tibble: 81,393 x 153
##   'Mother's Single Year of Age' Mother's Race/Hispanic~1 Month Prenatal Care ~2
##   <dbl> <dbl> <dbl>
## 1 22 7 3
## 2 31 6 99
## 3 29 1 3
## 4 39 7 2
## 5 20 5 3
## 6 29 7 3
## 7 23 3 3
## 8 34 1 3
## 9 19 3 5
## 10 25 3 3
## # i 81,383 more rows
## # i abbreviated names: 1: 'Mother's Race/Hispanic Origin',
## # 2: 'Month Prenatal Care Began'
## # i 150 more variables: 'Number of Prenatal Visits' <dbl>, 'WIC Program' <chr>,
## # 'Cigarette Recode' <chr>, 'Body Mass Index' <dbl>,
## # 'Body Mass Index Recode' <dbl>,
## # 'Pre-pregnancy Weight Recode (in Pounds)' <dbl>, ...
```

```
NatJan2021.3 <- NatJan2021.2 %>% select(-c('Body Mass Index Recode', 'Pre-pregnancy Weight Recode (in P
NatJan2021.3
```

```
## # A tibble: 81,393 x 57
##   'Mother's Single Year of Age' Mother's Race/Hispanic~1 Month Prenatal Care ~2
##   <dbl> <dbl> <dbl>
## 1 22 7 3
## 2 31 6 99
## 3 29 1 3
## 4 39 7 2
## 5 20 5 3
## 6 29 7 3
## 7 23 3 3
## 8 34 1 3
## 9 19 3 5
## 10 25 3 3
## # i 81,383 more rows
## # i abbreviated names: 1: 'Mother's Race/Hispanic Origin',
## # 2: 'Month Prenatal Care Began'
```

```
## # i 54 more variables: 'Number of Prenatal Visits' <dbl>, 'WIC Program' <chr>,
## #   'Cigarette Recode' <chr>, 'Body Mass Index' <dbl>, 'Weight Gain' <dbl>,
## #   'Pre-pregnancy Diabetes' <chr>, 'Gestational Diabetes' <chr>,
## #   'Pre-pregnancy Hypertension' <chr>, 'Gestational Hypertension' <chr>, ...
```

```
NatJan2021.FC <- NatJan2021.3 %>% select(-c(`Assisted Ventilation (immediately)`, `Assisted Ventilation
NatJan2021.FC
```

```
## # A tibble: 81,393 x 16
##   'Mother's Single Year of Age' Mother's Race/Hispanic~1 Month Prenatal Care ~2
##                                     <dbl>                                     <dbl>                                     <dbl>
## 1                                     22                                     7                                     3
## 2                                     31                                     6                                    99
## 3                                     29                                     1                                     3
## 4                                     39                                     7                                     2
## 5                                     20                                     5                                     3
## 6                                     29                                     7                                     3
## 7                                     23                                     3                                     3
## 8                                     34                                     1                                     3
## 9                                     19                                     3                                     5
## 10                                    25                                     3                                     3
## # i 81,383 more rows
## # i abbreviated names: 1: 'Mother's Race/Hispanic Origin',
## #   2: 'Month Prenatal Care Began'
## # i 13 more variables: 'Number of Prenatal Visits' <dbl>, 'WIC Program' <chr>,
## #   'Cigarette Recode' <chr>, 'Body Mass Index' <dbl>, 'Weight Gain' <dbl>,
## #   'Pre-pregnancy Diabetes' <chr>, 'Gestational Diabetes' <chr>,
## #   'Pre-pregnancy Hypertension' <chr>, 'Gestational Hypertension' <chr>, ...
```

Now to remove the pre-pregnancy hypertensive mothers (Rows with Y or U)

```
NatJan2021.NoPreHTN <- NatJan2021.FC[!(NatJan2021.FC$`Pre-pregnancy Hypertension`=="Y"|NatJan2021.FC$`P
NatJan2021.NoPreHTN
```

```
## # A tibble: 79,407 x 16
##   'Mother's Single Year of Age' Mother's Race/Hispanic~1 Month Prenatal Care ~2
##                                     <dbl>                                     <dbl>                                     <dbl>
## 1                                     22                                     7                                     3
## 2                                     31                                     6                                    99
## 3                                     29                                     1                                     3
## 4                                     39                                     7                                     2
## 5                                     20                                     5                                     3
## 6                                     29                                     7                                     3
## 7                                     23                                     3                                     3
## 8                                     34                                     1                                     3
## 9                                     19                                     3                                     5
## 10                                    25                                     3                                     3
## # i 79,397 more rows
## # i abbreviated names: 1: 'Mother's Race/Hispanic Origin',
## #   2: 'Month Prenatal Care Began'
## # i 13 more variables: 'Number of Prenatal Visits' <dbl>, 'WIC Program' <chr>,
## #   'Cigarette Recode' <chr>, 'Body Mass Index' <dbl>, 'Weight Gain' <dbl>,
## #   'Pre-pregnancy Diabetes' <chr>, 'Gestational Diabetes' <chr>,
## #   'Pre-pregnancy Hypertension' <chr>, 'Gestational Hypertension' <chr>, ...
```

Then to remove the mother's with unknown statuses throughout file

```
NatJan2021.Final <- NatJan2021.NoPreHTN[!(NatJan2021.NoPreHTN$`Body Mass Index`=="99.9"|NatJan2021.NoPreHTN$`Pre-pregnancy Hypertension`=="99.9"),]  
NatJan2021.Final
```

```
## # A tibble: 73,584 x 16  
##   'Mother's Single Year of Age' Mother's Race/Hispanic~1 Month Prenatal Care ~2  
##           <dbl>           <dbl>           <dbl>  
## 1           29           1           3  
## 2           39           7           2  
## 3           20           5           3  
## 4           29           7           3  
## 5           34           1           3  
## 6           19           3           5  
## 7           25           3           3  
## 8           24           3           2  
## 9           28           3           1  
## 10          32           1           2  
## # i 73,574 more rows  
## # i abbreviated names: 1: 'Mother's Race/Hispanic Origin',  
## # 2: 'Month Prenatal Care Began'  
## # i 13 more variables: 'Number of Prenatal Visits' <dbl>, 'WIC Program' <chr>,  
## # 'Cigarette Recode' <chr>, 'Body Mass Index' <dbl>, 'Weight Gain' <dbl>,  
## # 'Pre-pregnancy Diabetes' <chr>, 'Gestational Diabetes' <chr>,  
## # 'Pre-pregnancy Hypertension' <chr>, 'Gestational Hypertension' <chr>, ...
```

The dataset is finally ready for analysis