

NHANES Manual Reproduce

HEIR Analyst Claudia Solis-Roman

November 11, 2019

Manual for Visualizing Risk for Heart Attack by Race for Women Ages 50 and Above

Objective

This Health Equity Institute of Research (HEIR) manual is a reference for reproducible NHANES analyses relating to Heart Health Equity. It includes training to help analysts perform some basic data cleaning, analysis, and visualization methods in R with an eye toward reproducibility and basic instructions on frequently used packages at HEIR. In this analysis manual, we will be producing statistics from NHANES on female heart health, looking closely at women above age 50 and examining how race interacts with risk for heart attack.

Step 0. Environment set up

0. We load necessary packages and run source code from R script to set up packrat, a reproducibility system that will allow adequate archival reference of this code and analysis.

```
if(!requireNamespace("here"))  
  install.packages("here", repos = "https://cloud.r-project.org")
```

```
## Loading required namespace: here
```

```
library("here")
```

```
## Warning: package 'here' was built under R version 3.5.3
```

```
## here() starts at C:/Users/Cloud/Box/A3SR Social Impact/repro_NHANES_manual
```

```
here::here()  
  
if(!requireNamespace("table1"))  
  install.packages("table1", repos = "https://cloud.r-project.org")
```

```
## Loading required namespace: table1
```

```
library("table1")
```

```
## Warning: package 'table1' was built under R version 3.5.3
```

```
##  
## Attaching package: 'table1'
```

```
## The following objects are masked from 'package:base':  
##  
##      units, units<-
```

```
if(!requireNamespace("readstata13"))  
  install.packages("readstata13", repos = "https://cloud.r-project.org")
```

```
## Loading required namespace: readstata13
```

```
library("readstata13")
```

```
## Warning: package 'readstata13' was built under R version 3.5.3
```

```
if(!requireNamespace("dplyr"))  
  install.packages("dplyr", repos = "https://cloud.r-project.org")
```

```
## Loading required namespace: dplyr
```

```
library("dplyr")
```

```
## Warning: package 'dplyr' was built under R version 3.5.3
```

```
##  
## 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
```

```
if(!requireNamespace("tidyr"))  
  install.packages("tidyr", repos = "https://cloud.r-project.org")
```

```
## Loading required namespace: tidyr
```

```
library("tidyr")
```

```
## Warning: package 'tidyr' was built under R version 3.5.3
```

```
if(!requireNamespace("ggplot2"))  
  install.packages("ggplot2", repos = "https://cloud.r-project.org")
```

```
## Loading required namespace: ggplot2
```

```
library("ggplot2")
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
source(here("src", "script.R"))
```

```
## Initializing packrat project in directory:  
## - "C:/Users/Claud/Box/A3SR Social Impact/repro_NHANES_manual"
```

```
## Warning in Sys.junction(from, to): cannot set reparse point 'C:/Users/  
## Claud/Box/A3SR Social Impact/repro_NHANES_manual/packrat/lib-R/x86_64-w64-  
## mingw32/3.5.2/base', reason 'Incorrect function'
```

```
## Initialization complete!
```

```
## Warning in Sys.junction(from, to): cannot set reparse point 'C:/Users/  
## Claud/Box/A3SR Social Impact/repro_NHANES_manual/packrat/lib-R/x86_64-w64-  
## mingw32/3.5.2/base', reason 'Incorrect function'
```

Step 1. Open, read in, and subset data for population

1. We will use data from the Center for Disease Control's National Health and Nutrition Examination Survey II (NHANES II) for this analysis. Up to date data can be found at <http://www.stata-press.com/data/r14/nhanes2.dta> (<http://www.stata-press.com/data/r14/nhanes2.dta>) (here). This is an unweighted analysis. We use `readdata13` to read in Stata data from the website and save the attributes so we can quickly view the variable labels in R in conjunction with online NHANES documentation. We then subselect from the data as a whole using base R to arrive at our population of interest, women over age 40 or above age 50.

```
nhanes <- read.dta13("http://www.stata-press.com/data/r14/nhanes2.dta")  
var.labels <- attr(data, "var.labels")  
nhanes <- nhanes[nhanes$female==1 & nhanes$age>=50,]
```

Step 2. Create summary NHANES race and heart health statistics table

2. Create a summary table or figure with descriptive statistics for five variables that relate to heart disease stratified by race. In particular, we generate mean, standard deviation, median, min, max, and cross tabulated frequencies for heart attack, high blood pressure, age, diabetes, and body mass index.

```
table1::label(nhanes$age) <- "Age"
table1::label(nhanes$race) <- "Race"
table1::label(nhanes$highbp) <- "High Blood Pressure"
table1::label(nhanes$diabetes) <- "Diabetes"
table1::label(nhanes$bmi) <- "Body Mass Index"

table1::label(nhanes$heartatk) <- "Heart Attack"

table1(~heartatk + highbp + age + diabetes + bmi | race, data=nhanes)
```

	White (n=2414)	Black (n=264)	Other (n=40)	Overall (n=2718)
Heart Attack				
Mean (SD)	0.0514 (0.221)	0.0758 (0.265)	0.0750 (0.267)	0.0541 (0.226)
Median [Min, Max]	0.00 [0.00, 1.00]	0.00 [0.00, 1.00]	0.00 [0.00, 1.00]	0.00 [0.00, 1.00]
High Blood Pressure				
Mean (SD)	0.565 (0.496)	0.720 (0.450)	0.625 (0.490)	0.581 (0.494)
Median [Min, Max]	1.00 [0.00, 1.00]	1.00 [0.00, 1.00]	1.00 [0.00, 1.00]	1.00 [0.00, 1.00]
Age				
Mean (SD)	63.3 (6.47)	62.8 (6.60)	61.9 (6.78)	63.2 (6.48)
Median [Min, Max]	64.0 [50.0, 74.0]	63.0 [50.0, 74.0]	62.5 [50.0, 74.0]	64.0 [50.0, 74.0]
Diabetes				
Mean (SD)	0.0783 (0.269)	0.133 (0.340)	0.0750 (0.267)	0.0835 (0.277)
Median [Min, Max]	0.00 [0.00, 1.00]	0.00 [0.00, 1.00]	0.00 [0.00, 1.00]	0.00 [0.00, 1.00]
Body Mass Index				
Mean (SD)	26.4 (5.40)	29.2 (6.15)	24.5 (4.46)	26.6 (5.53)
Median [Min, Max]	25.5 [14.1, 54.1]	28.2 [15.4, 53.3]	23.5 [17.4, 38.2]	25.7 [14.1, 54.1]

Step 3. Describe NHANES heart health variables table

3. Provide a brief description of the table you have generated. We recommend an intuitive check to see if your results make sense using your health equity content and knowledge and common sense. One way to practice this is to guess at what your output could look like, developing detached hypotheses, and then interpreting actual results. Describe your results using general trends such as compare and contrast and pattern-finding, and use in-text figures where possible (e.g. Suzy ran a faster mile time than Joe [6 versus 7 minute miles, respectively]). While we include minimum and maximum for binary variables in this table as part of our data quality checking process, note that this part of the table is not exported into the published scorecard.

Description: In the results contained in Table 1, we see that even though black women, white women, and women of other race/ethnicity groups at or above age 50 have similar average ages and age distributions (ranging from about 62 to about 63, with standard deviations exceeding the difference between them), white women have the lowest rates of heart attacks (.05 vs. .08). High blood pressure rates are 18 percent higher among black women compared to white women in this sample, and 8 percent higher among other race women compared to white women. Diabetes rates are highest among black women at 13.3%, nearly double those of white women and other women, and while all three race groups have BMI averages within one standard deviation of the “healthy” range,

women in other race groups fall within the maximum of the healthy range at 24.5, compared to black and white women (29.2 and 26.4, respectively). We note that the sample size 40 for other race groups is quite small and we will exclude this group for the subsequent graphical analysis.

Step 4. Women's heart health by race equity plot

4. Perform an equity analysis using a visualization that will be useful to HEIR's scorecard reporting on racial disparities in heart disease for women ages 50 and over. Note we are returning to the full data set and subselecting our sample again to show you another way to manipulate data using dplyr and tidyr. We will be using dplyr and ggplot, as a means to introduce you to two versatile and beloved packages. Welcome to ggplot R visualization heaven and the magical tidyverse. Time permitting, explore the dataset with additional health equity questions. Can you redesign this step to examine relationships between systolic blood pressure or high blood pressure rates, or across other populations, such as rural and urban?

```
nhanes <- read.dta13("http://www.stata-press.com/data/r14/nhanes2.dta")

# Filtering and subsetting NHANES data
subdata <- nhanes %>%
  filter ((race == "White" | race == "Black") & (sex == "Female")) %>%
  subset(select=c(heartatk, race, age))

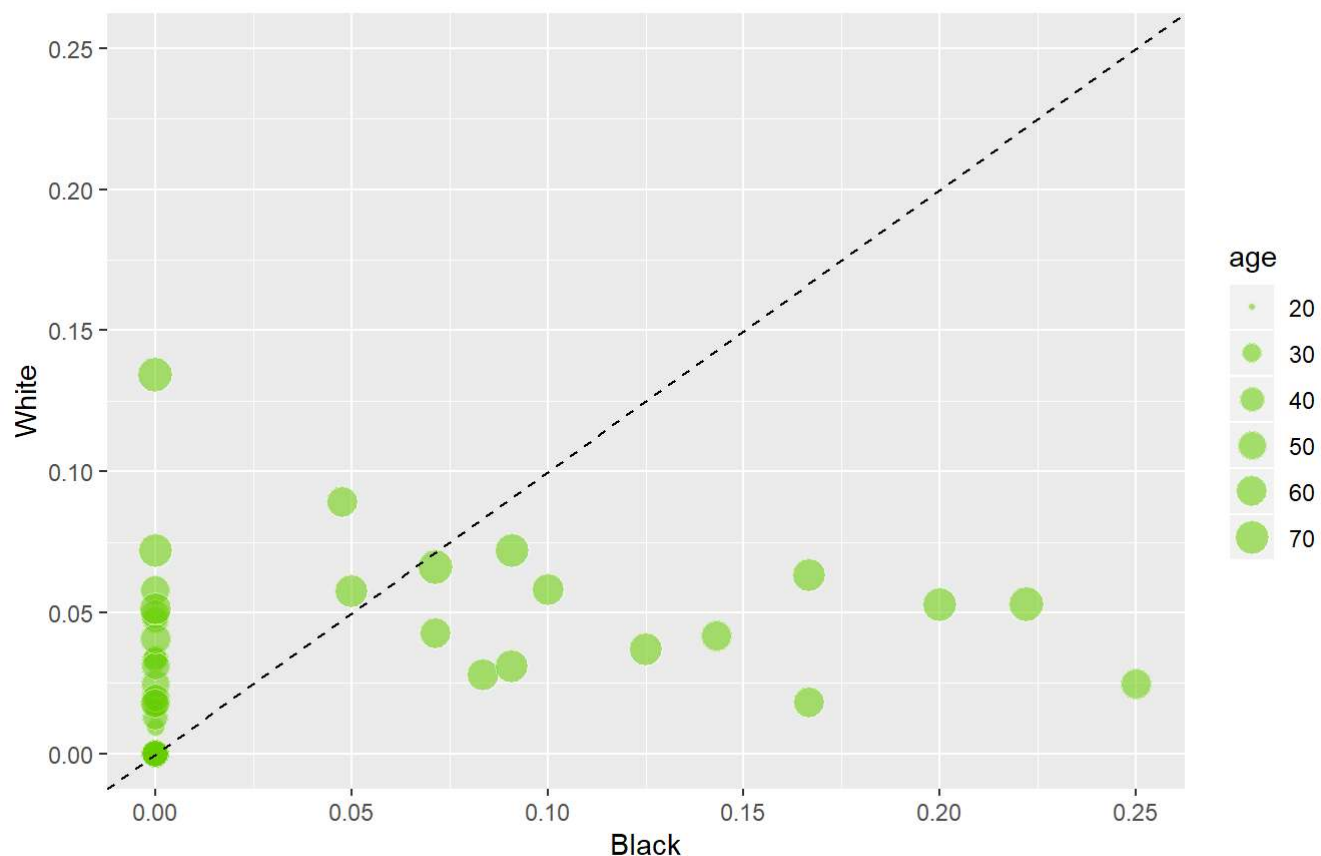
# Summarize heart rate risk by age
metadata <- subdata %>%
  group_by(age, race) %>%
  summarize(heartatk_rate = mean(heartatk, na.rm = TRUE))

# Reshaping the NHANES by age data
spreaddata <- metadata %>%
  spread(key = c(race), value=heartatk_rate)

# Plotting female heart attack rate for black and white with age sized dots
# Add an abline and change the scale of axes to be consistent and show line of equal outcomes

ggplot(spreaddata, aes(x=Black, y=White))+
  geom_point(colour="white", fill="chartreuse3", shape=21, alpha=.55, aes(size=age))+
  geom_abline(intercept = 0, slope = 1, linetype=2)+
  scale_x_continuous(limits=c(0,.25))+
  scale_y_continuous(limits=c(0,.25))+
  labs(title="Female Heart Attack Rate by Age and Race",
        Subtitle="Ages: 50 to 75, Average Rates",
        caption="Source: NHANES",
        x="Black",
        y="White")
```

Female Heart Attack Rate by Age and Race



Source: NHANES

Step 5. Bundle for easy reproducibility

5. Bundle this using packrat using `::bundle`, just as you did `::init` at start.

```
packrat::bundle
source(here("src", "script2.R"))
```

```
## The packrat project has been bundled at:
## - "C:\Users\Claud\Box\A3SR Social Impact\repro_NHANES_manual\packrat\bundles\repro_NHANES_manual-2019-11-11.tar.gz"
```

Step 6. Review, check for errors, and share

6. Create a title for your results that adequately describes your analysis. Review your output - are results sensible? Check knitted output, writing, and formatting for legibility. Bonus points if you can implement a shareable version control system such as github alongside packrat. Share your results.

Thank you for reading. Please feel free to make changes and provide feedback on this manual.