

NHANES_Manual_Reproduce

HEIR Analyst Claudia Solis-Roman

November 11, 2019

NHANES Analysis Reproduce Manual

Manual for Visualizing how the Relationship between Blood Pressure and Risk for Heart Attack Varies by Race for Women Ages 50 and Above”

Objective

This Health Equity Institute of Research (HEIR) manual is intended to serve as a reference for junior analysts on how to perform routine data analyses for the website to help inform health policy for Health Equity Org. It includes training to help analysts like yourself perform some basic data cleaning, analysis, and visualization methods in R. In this manual, we will be producing statistics from NHANES for our female heart health scorecard, looking closely at women above age 50 and examining how race interacts with high blood pressure and risk for heart attack.

Step 0. Environment set up

```
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
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'

packrat::init
```

Step 1. Open and Clean Data

We will use data from the Center for Disease Control's National Health and Nutrition Examination Survey II (NHANES II) for this assignment. A Stata version of this dataset can be found at <http://www.stata-press.com/data/r14/nhanes2.dta> (here). For the purposes of this analysis, we are not required to use the survey weights. We use read data 13 to read in Stata 13 data from the online location, save the attributes so we can quickly view the variable labels here in conjunction with NHANES documentation.

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

We then subselect from the data as a whole using base R to arrive at our population of interest, women over age 50 or above age 50.

```
nhanes <- nhanes[nhanes$female==1 & nhanes$age>=50,]
```

Step 2. Tables for 5 variables

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)
```

```
## [1] "<table class='Rtable1'>\n<thead>\n<tr>\n<th class='rowlabel firstrow lastrow'></th>\n<th class='>
```

Step 3. Describe results

3. Provide a brief description of what the table/figure shows you. We recommend an intuitive check to see if your results make sense given your health equity content and knowledge, by practicing guessing

at what your output could look like and then investigating what results actually show. Describe your results specifically and use in text figures where possible.

In the results contained in the table above, 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 (ranging from about 62 to about 63) with standard distributions (6.5 to 7) exceeding their range, white women have the lowest rates of heart attacks.

Step 4.

4. Perform an analysis or create a visualization that will be useful to your organization. [Note to student: you can use any command that you want as long as (a) it incorporates at least two variables and (b) you have never used it before.] Describe why you chose that command (how it uses data to inform the mission of the organization) and how you implemented it. Examples of types of commands: visualizations (graphs, charts, plots), data sorting, missing data summaries, predictive models. Pie charts of any sort are not allowed!

Step 5.

5. Bundle this using packrat as demonstrated by Vicky Steeves.

```
packrat::bundle
```

6. Create a title for your manual that is appropriate given what you did, for instance “Manual for visualizing how the relationship between blood pressure and heart attacks varies by race”