ANA 515 Assignment 2

Catherine Liao

2023

#Section 1: Description of the data #The dataset I’m using contains sleeping data about people who are in different relationship statuses - or more specifically, data about whether these people are sleeping in the same beds with their partners (if not, why). The data was collected from questionnaire responses. I’m hoping to use it answer questions about the correlations between relationship status and sleeping behaviors. This dataset is saved in a csv file that’s fixed width.

#Section 2: Reading the data into R #I used read\_csv, which comes from the “readr” package.

library(readr)  
url <- "https://github.com/fivethirtyeight/data/blob/master/sleeping-alone-data/sleeping-alone-data.csv"   
newdoc <- read\_csv(url)

## Warning: One or more parsing issues, call `problems()` on your data frame for details,  
## e.g.:  
## dat <- vroom(...)  
## problems(dat)

## Rows: 38526 Columns: 1  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): <!DOCTYPE html>  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

#Section 3: Clean the data

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.2 ✔ purrr 1.0.1  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.2 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(bslib)

##   
## Attaching package: 'bslib'  
##   
## The following object is masked from 'package:utils':  
##   
## page

colnames(newdoc) <- c("nd\_column1", "nd\_column2","nd\_column3", "nd\_column4", "nd\_column5")

#Section 4: Characteristics of the data #Inline Code This dataframe has 38526 rows and 38526 columns. The names of the columns and a brief description of each are in the table below:

library(knitr)  
column\_info <- data.frame(  
 Column\_Name = c("nd\_column1", "nd\_column2", "nd\_column3", "nd\_column4", "nd\_column5"),  
 Description = c("TheStartDate", "TheEndDate", "CurrentRelationshipStatus", "LengthofRelationship", "SleepingStatus")  
)  
kable(column\_info, format = "markdown", col.names = c("Column\_Name", "Description"))

| Column\_Name | Description |
| --- | --- |
| nd\_column1 | TheStartDate |
| nd\_column2 | TheEndDate |
| nd\_column3 | CurrentRelationshipStatus |
| nd\_column4 | LengthofRelationship |
| nd\_column5 | SleepingStatus |

#Section 5: Summary Statistics

data\_pick3 <- select(newdoc, "nd\_column1")  
summary <- summary(data\_pick3)  
missing\_values <- colSums(is.na(data\_pick3))  
summary <- cbind(summary, missing\_values)  
print(summary)

## nd\_column1 missing\_values  
## "Length:38526 " "2"   
## "Class :character " "2"   
## "Mode :character " "2"