ANA 515 Assignment 2

Saumya Rajamohan

2024-06-16

#echo = TRUE allows code to show up in final result  
library(magrittr)  
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.3.3

## Warning: package 'ggplot2' was built under R version 4.3.3

## Warning: package 'tidyr' was built under R version 4.3.3

## Warning: package 'dplyr' was built under R version 4.3.3

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ tidyr::extract() masks magrittr::extract()  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ✖ purrr::set\_names() masks magrittr::set\_names()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(dplyr)

#Section 1: Data Description

The dataset includes responses to a survey monkey questionnaire from around 1615 adult men, and explores their opinions behind all that entails masculinity. Most questions on the survey were multi-select multiple choice questions, while some were open ended that is not reflected in this dataset. The survey answers questions around the general perception of masculinity among men and their attitude towards being in vulnerable situations like asking for help, seeking therapy etc. It further explores their experience as a male at the work space, especially during the #MeToo movement. It also tries to assess actions taken if any to incidents of sexual harassment if witnessed at the workplace.It would like to use this survey to understand if there is an association between men’s ideas about masculinity and their knowledge and willingness to support the #MeToo Movement. The file is saved as a CSV file. It is a delimited flat file, where commas are used to separate every column (i.e. survey question) for each survey respondent.

# Section 2: Reading data into R  
  
#The read.csv function was used to read the data from a url. This R function can read delimited flat files with comma separated values.  
#read.csv is found within the Tidyverse package  
  
raw <- "https://raw.githubusercontent.com/fivethirtyeight/data/master/masculinity-survey/raw-responses.csv"  
survey <- read.csv(raw)

# Section 3: Clean the Data  
  
#Subset the dataset by selecting some columns  
#Rename columns  
subsurvey<- survey %>%   
 select(q0001, q0002, q0005, q0009, q0014, q0015, weight) %>%   
 rename("SelfPerception"="q0001", "ExternalPerception"="q0002", "SocietalPressure"="q0005", "Employment"="q0009", "MeTooAwareness"="q0014", "MeTooResponse"="q0015", "Weight"="weight")

# Section 4a: Characteristics of data  
#Inline code  
library(knitr)

## Warning: package 'knitr' was built under R version 4.3.3

This dataframe has 1615 rows and 7 columns. The names of the columns and a brief description of each are in the table below:

# Section 4b: Characteristics of data  
#Description table  
library(knitr)  
columns\_summary <- data.frame(  
Column\_Name = c(colnames(subsurvey)),  
Column\_Description = c(  
"In general, how masculine or manly do you feel?", "How important is it to you that others see you as masculine?", "Do you think that society puts pressure on men in a way that is unhealthy or bad for them?", "Which of the following categories best describes your employment status?","AMONG EMPLOYED: How much have you heard about the #MeToo movement?", "AMONG EMPLOYED WHO'VE HEARD OF METOO: As a man, would you say you think about your behavior at work differently in the wake of #MeToo?", "Weightage")  
)  
  
kable(columns\_summary, caption = "Masculinity Survey Dataframe Description Table")

Masculinity Survey Dataframe Description Table

| Column\_Name | Column\_Description |
| --- | --- |
| SelfPerception | In general, how masculine or manly do you feel? |
| ExternalPerception | How important is it to you that others see you as masculine? |
| SocietalPressure | Do you think that society puts pressure on men in a way that is unhealthy or bad for them? |
| Employment | Which of the following categories best describes your employment status? |
| MeTooAwareness | AMONG EMPLOYED: How much have you heard about the #MeToo movement? |
| MeTooResponse | AMONG EMPLOYED WHO’VE HEARD OF METOO: As a man, would you say you think about your behavior at work differently in the wake of #MeToo? |
| Weight | Weightage |

# Section 5a: Summary Statistics, including missing values  
  
#Subset dataframe to 3 columns  
survey3<-select(subsurvey,SelfPerception, MeTooAwareness, Weight)  
  
#Summarize columns  
surveysummary<-summary(survey3)  
print(surveysummary)

## SelfPerception MeTooAwareness Weight   
## Length:1615 Length:1615 Min. :0.01974   
## Class :character Class :character 1st Qu.:0.10258   
## Mode :character Mode :character Median :0.59689   
## Mean :1.00000   
## 3rd Qu.:1.01005   
## Max. :8.67661

#As the summary function does not show missing values for character variables by default, rewriting a code to summarize the 3 selected columns by factorizing the character variables  
  
surveysummary2<-survey3 %>%   
 mutate(across(where(is.character), factor))  
surveysummarymissing <-summary(surveysummary2)  
print(surveysummarymissing)

## SelfPerception MeTooAwareness Weight   
## No answer : 14 A lot :432 Min. :0.01974   
## Not at all masculine: 32 No answer : 1 1st Qu.:0.10258   
## Not very masculine :131 Nothing at all:137 Median :0.59689   
## Somewhat masculine :826 Only a little : 85 Mean :1.00000   
## Very masculine :612 Some :225 3rd Qu.:1.01005   
## NA's :735 Max. :8.67661

# Section 5b: Summary Statistics, excluding missing values  
  
#Filter out rows with missing values  
surveysummary3 <- na.omit(survey3)  
miss<- summary(surveysummary3)  
print(miss)

## SelfPerception MeTooAwareness Weight   
## Length:880 Length:880 Min. :0.01974   
## Class :character Class :character 1st Qu.:0.14255   
## Mode :character Mode :character Median :0.62053   
## Mean :1.06008   
## 3rd Qu.:1.15773   
## Max. :8.64367

#Factorize character variables and summarize the 3 selected variables  
surveysummarycomplete<-surveysummary3 %>%   
 mutate(across(where(is.character), factor))  
surveysummarynomissing <-summary(surveysummarycomplete)  
print(surveysummarynomissing)

## SelfPerception MeTooAwareness Weight   
## No answer : 7 A lot :432 Min. :0.01974   
## Not at all masculine: 15 No answer : 1 1st Qu.:0.14255   
## Not very masculine : 73 Nothing at all:137 Median :0.62053   
## Somewhat masculine :447 Only a little : 85 Mean :1.06008   
## Very masculine :338 Some :225 3rd Qu.:1.15773   
## Max. :8.64367