ANA 515 Assignment 2

Clarentia Kurnia

4/8/2022

## Libraries

#install.packages("tidyverse")  
#install.packages("readr")  
#install.packages("dplyr")  
#install.packages("stringr")  
#install.packages("knitr")  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✓ ggplot2 3.3.5 ✓ purrr 0.3.4  
## ✓ tibble 3.1.3 ✓ dplyr 1.0.7  
## ✓ tidyr 1.2.0 ✓ stringr 1.4.0  
## ✓ readr 2.1.2 ✓ forcats 0.5.1

## Warning: package 'tidyr' was built under R version 4.1.2

## Warning: package 'readr' was built under R version 4.1.2

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(readr)  
library(dplyr)  
library(stringr)  
library(knitr)

## Warning: package 'knitr' was built under R version 4.1.2

## Description of the data

The dataset Alcohol Consumption measures the beer, spirit, and wine servings of 193 countries in 2010. It also measures the total litres of pure alcohol each country consumes. The data was collected by World Health Organization (WHO) and Global Information System on Alcohol and Health (GISAH) in 2010. WHO utilized government records and statistics from the alcohol industry and the United Nation’s Food and Agriculture Organization database. The data is saved in a CSV file that is 4.28 KB in size. It is delimited with comma as its delimiter.

## Reading the data into R

url <- "https://raw.githubusercontent.com/fivethirtyeight/data/master/alcohol-consumption/drinks.csv"  
drinks <- read\_csv(url(url))

## Rows: 193 Columns: 5

## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): country  
## dbl (4): beer\_servings, spirit\_servings, wine\_servings, total\_litres\_of\_pure...

##   
## 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.

To read the data into R and assign it to a dataframe object, I installed the package ‘tidyverse’ and used the function ‘read\_csv’. I assigned the data to a dataframe called ‘drinks’.

## Cleaning the data

drinks$country <- str\_to\_title(drinks$country)  
drinks$beer\_servings <- str\_to\_title(drinks$beer\_servings)  
drinks$spirit\_servings <- str\_to\_title(drinks$spirit\_servings)  
drinks$wine\_servings <- str\_to\_title(drinks$wine\_servings)  
drinks$total\_litres\_of\_pure\_alcohol <- str\_to\_title(drinks$total\_litres\_of\_pure\_alcohol)

I installed the package ‘stringr’ to rename the columns. The function str\_to\_title replaces the column names into a title format. For example, “country” becomes “Country”.

## Characteristics of the data

a <- nrow(drinks)  
b <- ncol(drinks)

This dataframe has 193 rows and 5 columns. The names of the columns and a brief description of each are in the table below:

|  |  |
| --- | --- |
| Column Name | Description |
| Country | Name of country being measured |
| Beer servings | Number of 12-ounce cans of beer consumed |
| Spirit servings | Number of measures of spirit consumed |
| Wine servings | Number of glasses of wine consumed |
| Total litres of pure alcohol | Number of litres of hard liquor consumed |

## Summary Statistics

drinks <- subset(drinks, select=-c(country,total\_litres\_of\_pure\_alcohol))  
drinks\_summary <- summary(drinks)  
print(drinks\_summary)

## beer\_servings spirit\_servings wine\_servings   
## Length:193 Length:193 Length:193   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character

Here I created a summary of my data, where it provides the length, class, and mode of each column.