ANA Assignment 2\_2

Laykumar Lad

2024-06-11

#The dataset in use, sourced from FiveThirtyEight, details alcohol consumption by country. It captures the average annual per capita consumption of different types of alcohol across various nations. This dataset aims to provide insights into global alcohol consumption patterns. Research questions it can address include identifying countries with the highest and lowest alcohol consumption and determining the most and least consumed types of alcohol.  
  
#The dataset is stored in a CSV (Comma Separated Values) file, a flat file format. Delimited by commas, CSV files are easily readable and writable using numerous software programs, including R.

library(tidyverse)

## ── 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() ──  
## ✖ 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

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

## 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.

#The read\_csv function from the readr package is used to read the CSV file.

# Inline code to print number of rows and columns

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

# Create a table of column names and descriptions  
library(knitr)  
column\_descriptions <- data.frame(  
 column\_name = colnames(data),  
 description = c("Country name",   
 "Beer servings per person per year",   
 "Spirit servings per person per year",   
 "Wine servings per person per year",   
 "Total litres of pure alcohol per person per year")  
)  
kable(column\_descriptions, col.names = c("Column Name", "Description"))

| Column Name | Description |
| --- | --- |
| country | Country name |
| beer\_servings | Beer servings per person per year |
| spirit\_servings | Spirit servings per person per year |
| wine\_servings | Wine servings per person per year |
| total\_litres\_of\_pure\_alcohol | Total litres of pure alcohol per person per year |

# Select three columns for summary  
selected\_columns <- data[, c("beer\_servings", "spirit\_servings", "wine\_servings")]  
  
# Get summary statistics  
summary\_stats <- data.frame(  
 column = colnames(selected\_columns),  
 min\_value = sapply(selected\_columns, min, na.rm = TRUE),  
 max\_value = sapply(selected\_columns, max, na.rm = TRUE),  
 mean\_value = sapply(selected\_columns, mean, na.rm = TRUE),  
 missing\_values = sapply(selected\_columns, function(x) sum(is.na(x)))  
)  
  
# Print summary statistics  
kable(summary\_stats, col.names = c("Column", "Min Value", "Max Value", "Mean Value", "Missing Values"))

|  | Column | Min Value | Max Value | Mean Value | Missing Values |
| --- | --- | --- | --- | --- | --- |
| beer\_servings | beer\_servings | 0 | 376 | 106.16062 | 0 |
| spirit\_servings | spirit\_servings | 0 | 438 | 80.99482 | 0 |
| wine\_servings | wine\_servings | 0 | 370 | 49.45078 | 0 |