ANA 515 Assignment 2

Jiawei Zhao

2024-06-16

knitr::opts\_chunk$set(echo = TRUE)  
# Section 1: Description of the data  
  
# The dataset used in this analysis measures alcohol consumption across different countries. The data includes the average servings of beer, spirits, and wine consumed per person per year, as well as the total liters of pure alcohol consumed. This data could be used to answer research questions about the patterns of alcohol consumption globally, such as which countries have the highest alcohol consumption, or the type of alcohol most consumed in different regions.  
  
# The dataset is saved in a CSV file, which is a type of flat file where each line of the file corresponds to a data record, and each record is delimited by a comma. CSV files are widely used for data exchange because they are simple and can be opened by many programs, including text editors and spreadsheet software like Microsoft Excel.

# Load required package for reading CSV files  
library(readr)  
# Read the data from the provided URL and store it in a dataframe  
url <- "https://raw.githubusercontent.com/fivethirtyeight/data/master/alcohol-consumption/drinks.csv"  
drinks\_data <- readr::read\_csv(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.

# In this code chunk, I used the readr package's read\_csv.

# Clean the data by renaming columns for better readability  
colnames(drinks\_data)[1] <- "Country"  
colnames(drinks\_data)[2] <- "Beer\_Servings"  
colnames(drinks\_data)[3] <- "Spirit\_Servings"  
colnames(drinks\_data)[4] <- "Wine\_Servings"  
colnames(drinks\_data)[5] <- "Total\_Litres\_of\_Pure\_Alcohol"

# Output the number of rows and columns in the dataframe  
n\_rows <- nrow(drinks\_data)  
n\_cols <- ncol(drinks\_data)  
  
# Create a table with column names and brief descriptions  
column\_names <- c("Country", "Beer\_Servings", "Spirit\_Servings", "Wine\_Servings", "Total\_Litres\_of\_Pure\_Alcohol")  
column\_descriptions <- c("Name of the country", "Servings of beer per person", "Servings of spirits per person", "Servings of wine per person", "Total liters of pure alcohol consumed per person")  
column\_info <- data.frame(Column\_Name = column\_names, Description = column\_descriptions)  
  
knitr::kable(column\_info, caption = "Column Descriptions")

Column Descriptions

| Column\_Name | Description |
| --- | --- |
| Country | Name of the country |
| Beer\_Servings | Servings of beer per person |
| Spirit\_Servings | Servings of spirits per person |
| Wine\_Servings | Servings of wine per person |
| Total\_Litres\_of\_Pure\_Alcohol | Total liters of pure alcohol consumed per person |

# Select three columns for summary statistics  
selected\_columns <- c("Beer\_Servings", "Spirit\_Servings", "Wine\_Servings")  
selected\_data <- drinks\_data[, selected\_columns]  
  
# Calculate summary statistics for the selected columns  
summary\_stats <- summary(selected\_data)  
summary\_stats

## Beer\_Servings Spirit\_Servings Wine\_Servings   
## Min. : 0.0 Min. : 0.00 Min. : 0.00   
## 1st Qu.: 20.0 1st Qu.: 4.00 1st Qu.: 1.00   
## Median : 76.0 Median : 56.00 Median : 8.00   
## Mean :106.2 Mean : 80.99 Mean : 49.45   
## 3rd Qu.:188.0 3rd Qu.:128.00 3rd Qu.: 59.00   
## Max. :376.0 Max. :438.00 Max. :370.00