ANA 515 Assignment

Gyu Hyun Byun

2022-09-18

## 

## Section 1: Description of the data

## 

The dataset I am using is alochol consumption, the data behind the story, “Dear Mona Followup: Where Do People Drink The Most Beer, Wine And Spirits?” The population is countries with four variables - “beer\_servings”, “spirit\_servings”, “wine\_servings”, and “total\_litres\_of\_pure\_alcohol”. The data would likely live in a database as the data is well structured and organized with columns and rows. The dataset is governed and published by the World Health Organization (“WHO”). The WHO assumes the roles and responsibilities for the original dataset. I don’t see any security or ethics issues related to this dataset as it is publicly published data from the WHO, a credible United Nations agency, for benefit of the world’s public health. The data shows countries as population, not speicifc individuals, which mitigates any ethical issues.

Please find the dataset at <https://github.com/fivethirtyeight/data/blob/50ec9ccfe3aa20f328bfd5e5da584fa5e15efee6/alcohol-consumption/drinks.csv>

## 

## Section 2: Reading the data into R

## 

drinks <- read.csv("https://raw.githubusercontent.com/fivethirtyeight/data/50ec9ccfe3aa20f328bfd5e5da584fa5e15efee6/alcohol-consumption/drinks.csv")  
drinks.data <- data.frame(drinks)

## 

## Section 3: Clean the data

## 

The data are sorted by the varaible total\_litres\_of\_pure\_alcohol in reverse order.

drinks.data <- drinks.data[order(-drinks.data$total\_litres\_of\_pure\_alcohol),]  
drinks.data = drinks.data%>%select(beer\_servings, spirit\_servings, wine\_servings, total\_litres\_of\_pure\_alcohol)

## 

## Section 4: Characteristics of the data

## 

The data has 193 rows and 4 columns. The names of the columns and a brief description of each are in the table below:

Variable = colnames(drinks.data)  
Description = c("12-ounce cans of beer consumed per person per year",  
 "1.5-fluid ounces / shot-glasses of spirit consumed per person per year",  
 "Glasses of wine per person per year",  
 "Total liters of pure alcohol consumed per person per year")  
vars = data.frame(Variable, Description)  
kable(vars, caption = "Variables and Their Descriptions")

Variables and Their Descriptions

| Variable | Description |
| --- | --- |
| beer\_servings | 12-ounce cans of beer consumed per person per year |
| spirit\_servings | 1.5-fluid ounces / shot-glasses of spirit consumed per person per year |
| wine\_servings | Glasses of wine per person per year |
| total\_litres\_of\_pure\_alcohol | Total liters of pure alcohol consumed per person per year |

## 

## Section 5: Summary statistics

## 

sub.data = drinks.data%>%select(beer\_servings, spirit\_servings, wine\_servings, total\_litres\_of\_pure\_alcohol)  
sub.summary = summary(sub.data)  
sub.summary

## beer\_servings spirit\_servings wine\_servings total\_litres\_of\_pure\_alcohol  
## Min. : 0.0 Min. : 0.00 Min. : 0.00 Min. : 0.000   
## 1st Qu.: 20.0 1st Qu.: 4.00 1st Qu.: 1.00 1st Qu.: 1.300   
## Median : 76.0 Median : 56.00 Median : 8.00 Median : 4.200   
## Mean :106.2 Mean : 80.99 Mean : 49.45 Mean : 4.717   
## 3rd Qu.:188.0 3rd Qu.:128.00 3rd Qu.: 59.00 3rd Qu.: 7.200   
## Max. :376.0 Max. :438.00 Max. :370.00 Max. :14.400

beer\_servings\_0 <- sum(is.na(drinks.data$beer\_servings))  
spirit\_servings\_0 <- sum(is.na(drinks.data$spirit\_servings))  
wine\_servings\_0 <- sum(is.na(drinks.data$wine\_servings))  
total\_litres\_of\_pure\_alcohol\_0 <- sum(is.na(drinks.data$total\_litres\_of\_pure\_alcohol))

There are 0 countries with missing beer\_servings.

There are 0 countries with missing spirit\_servings.

There are 0 countries with missing wine\_servings.

There are 0 countries with missing total\_litres\_of\_pure\_alcohol.