

# Assignment 2 – Alcohol Consumption

This dataset shows the average serving sizes of different kinds of alcohol per person for each country in the world. The data was collected using government records, statistics from the alcohol industry and the United Nation's Food and Agriculture Organization database. By analyzing this dataset, we can observe the trend of alcohol consumption across the world.

The data is saved as a CSV file, in which the data are separated by commas.

Read data:

```
#read data
library(readr)
drinks <- read_csv("C:\\Users\\lyy03\\OneDrive\\Documents\\Mcdaniel\\ANA 515\\Week 4\\
```

```
## 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.
```

## Data Cleaning

```
#data cleaning, some of the countries in this dataset are prohibited to drink legal w
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

```
allowed_countries <- c("Monaco", "San Marino", "North Korea") ## source:https://en.wik
drinks_legal <- drinks %>% filter (!(rowSums(select(., beer_servings, spirit_servings
summary(drinks_legal)
```

```
##      country      beer_servings spirit_servings wine_servings
## Length:183      Min.   : 0      Min.   : 0.00      Min.   : 0.00
## Class :character 1st Qu.: 25      1st Qu.: 8.00      1st Qu.: 1.50
## Mode  :character Median : 78      Median : 63.00     Median : 9.00
##                  Mean   :112      Mean   : 85.42     Mean   : 52.15
##                  3rd Qu.:193      3rd Qu.:132.50     3rd Qu.: 72.00
##                  Max.   :376      Max.   :438.00     Max.   :370.00
## total_litres_of_pure_alcohol
## Min.   : 0.000
## 1st Qu.: 1.600
## Median : 4.700
## Mean   : 4.975
## 3rd Qu.: 7.450
## Max.   :14.400
```

```
rowcount <- nrow(drinks_legal)
```

```
colcount <- ncol(drinks_legal)
```

This dataframe has 183 rows and 5 columns. The names of the columns and a brief description of each are in the table below:

```
library(knitr)
column_info <- data.frame(
  Column_Name = c("country", "beer_servings", "spirit_servings", "wine_servings", "total
  Description = c("Name of the country",
                  "Average size of beer servings per person per year",
                  "Average size of spirit servings per person per year",
                  "Average size of wine servings per person per year",
                  "Total litres of pure alcohol consumed per person per year")
```

```
)
kable(column_info, col.names = c("Column Name", "Description"), caption = "Descriptio
```

Column Name	Description
country	Name of the country
beer_servings	Average size of beer servings per person per year
spirit_servings	Average size of spirit servings per person per year
wine_servings	Average size of wine servings per person per year
total_litres_of_pure_alcohol	Total litres of pure alcohol consumed per person per year

### Description of dataset columns

```
##Summary stats
subset_drinks <- drinks_legal %>%
  select(country, beer_servings, wine_servings)
summary_stat <- summary(subset_drinks)
print(summary_stat)
```

```
##   country      beer_servings wine_servings
## Length:183      Min.   : 0      Min.   : 0.00
## Class :character 1st Qu.: 25     1st Qu.: 1.50
## Mode  :character Median : 78     Median : 9.00
##              Mean   :112     Mean   : 52.15
##              3rd Qu.:193     3rd Qu.: 72.00
##              Max.   :376     Max.   :370.00
```