# Alcohol Consumption

*Raju Adhikari & Mario Alonso Rodriguez*

*October 6th, 2016*

## Data

This is an assignment that utilizes fivethirtyeight.com's dataset on alcohol consumption in the world. This data is taken from the World Health Organization.

```
### Load the raw data from fivethirtyeight's Github
Drinks <- read.csv('https://raw.githubusercontent.com/fivethirtyeight/data/master/alcohol-consumption/di

### A simple summary of descriptive statistics of all the dataset
summary(Drinks)
```

```
##               country    beer_servings   spirit_servings  wine_servings
##  Afghanistan      :  1   Min.   :  0.0   Min.   :  0.00   Min.   :  0.00
##  Albania          :  1   1st Qu.: 20.0   1st Qu.:  4.00   1st Qu.:  1.00
##  Algeria          :  1   Median : 76.0   Median : 56.00   Median :  8.00
##  Andorra          :  1   Mean   :106.2   Mean   : 80.99   Mean   : 49.45
##  Angola           :  1   3rd Qu.:188.0   3rd Qu.:128.00   3rd Qu.: 59.00
##  Antigua & Barbuda:  1   Max.   :376.0   Max.   :438.00   Max.   :370.00
##  (Other)          :187
##  total_litres_of_pure_alcohol
##  Min.   : 0.000
##  1st Qu.: 1.300
##  Median : 4.200
##  Mean   : 4.717
##  3rd Qu.: 7.200
##  Max.   :14.400
##
```

```
### The means of all the variables show the average world consumption of each drink
for (i in 1:length(names(Drinks))) {
    Drinks[, i] %>%
    mean() %>%
    round(digits = 1) %>%
    paste(names(Drinks)[i], ., '\n') %>%  cat() }
```

```
## country NA
## beer_servings 106.2
## spirit_servings 81
## wine_servings 49.5
## total_litres_of_pure_alcohol 4.7
```

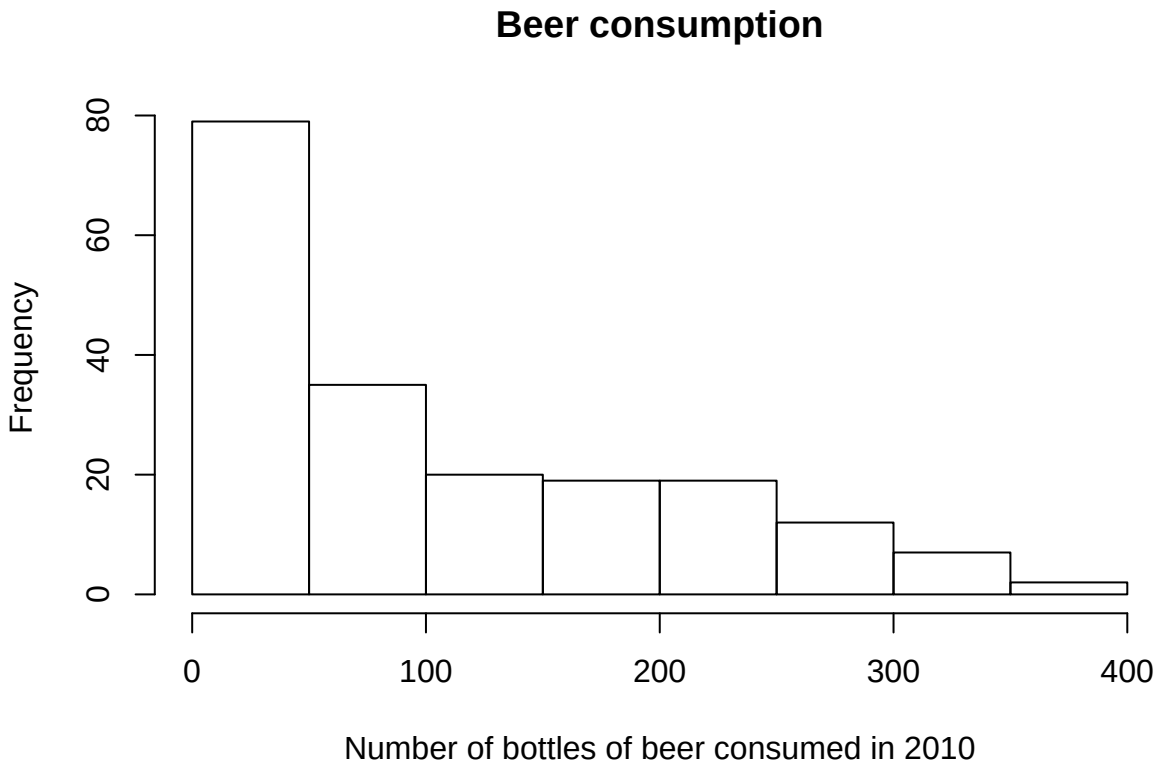
```
### The standard deviation of all the variables shows how dispersed is the data
for (i in 1:length(names(Drinks))) {
    Drinks[, i] %>%
    sd() %>%
    round(digits = 1) %>%
    paste(names(Drinks)[i], ., '\n') %>%  cat() }
```

```
## country 55.9
## beer_servings 101.1
## spirit_servings 88.3
## wine_servings 79.7
## total_litres_of_pure_alcohol 3.8
```

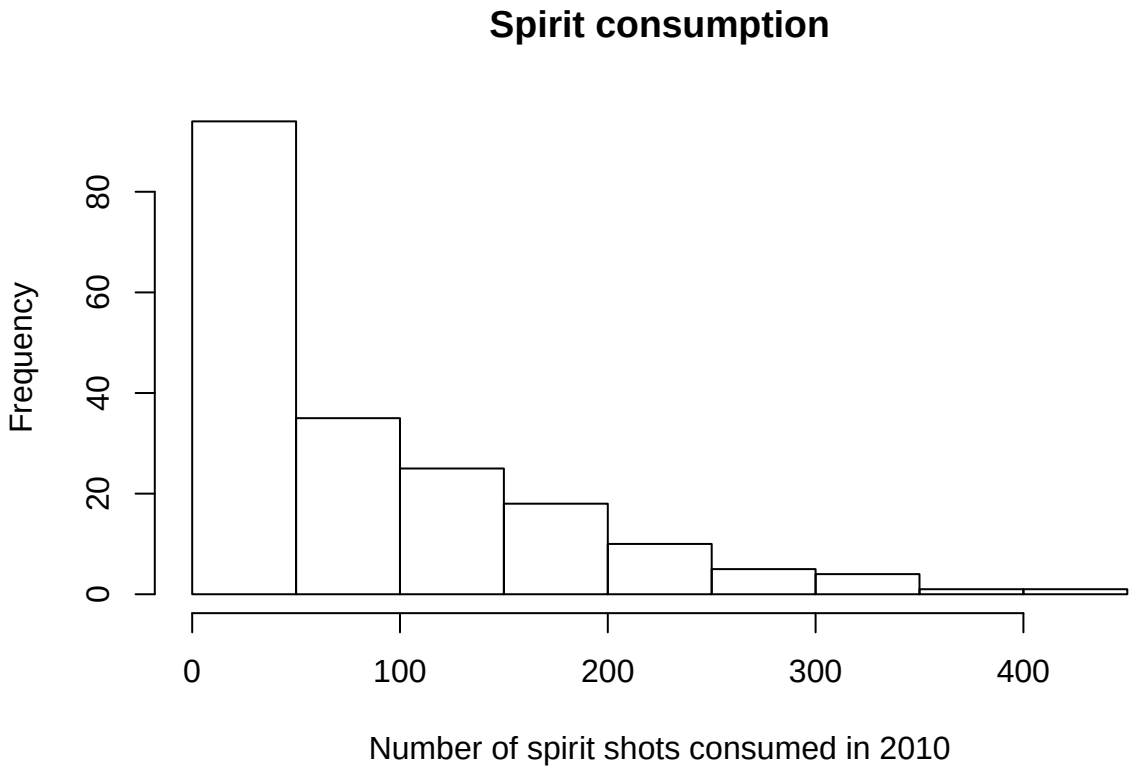
### Histograms

Histograms are really usefull to look at the distribution of the data.

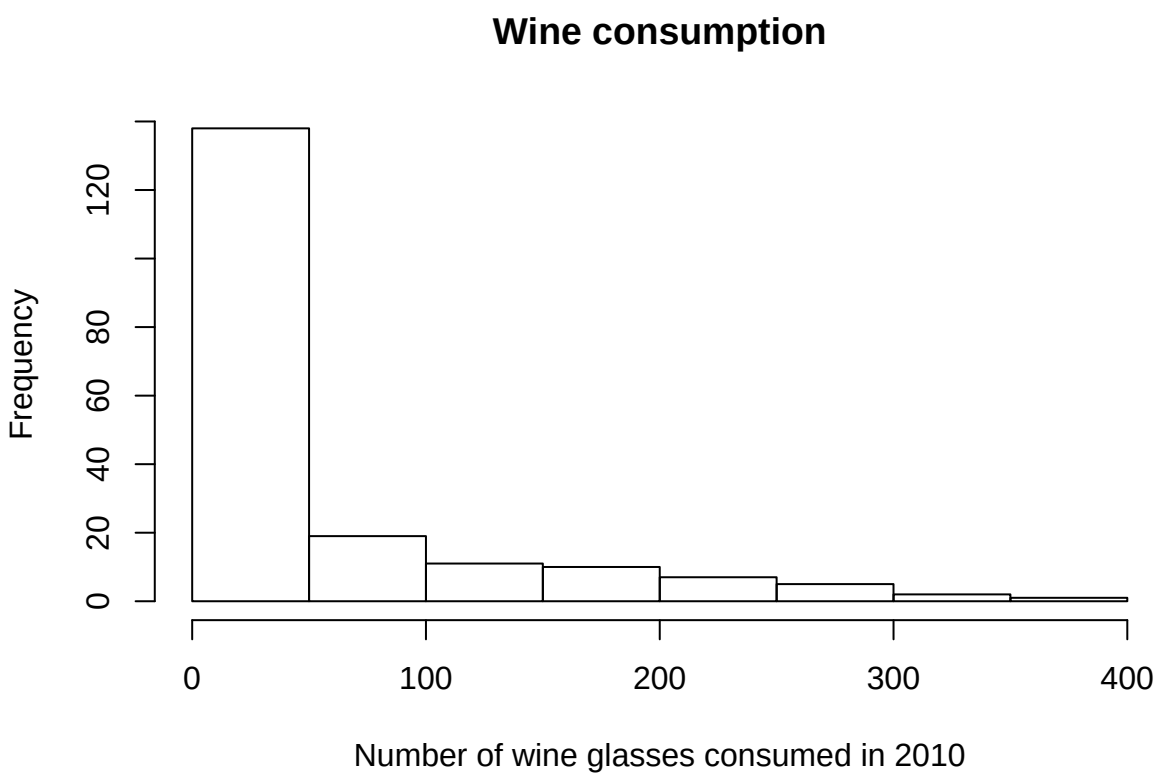
```
### Distribution of beer consumption
hist(Drinks$beer_servings,
     main = 'Beer consumption',
     xlab = 'Number of bottles of beer consumed in 2010')
```




```
### Distribution of spirit consumption
hist(Drinks$spirit_servings,
     main = 'Spirit consumption',
     xlab = 'Number of spirit shots consumed in 2010')
```




```
### Distribution of wine consumption
hist(Drinks$wine_servings,
     main = 'Wine consumption',
     xlab = 'Number of wine glasses consumed in 2010')
```

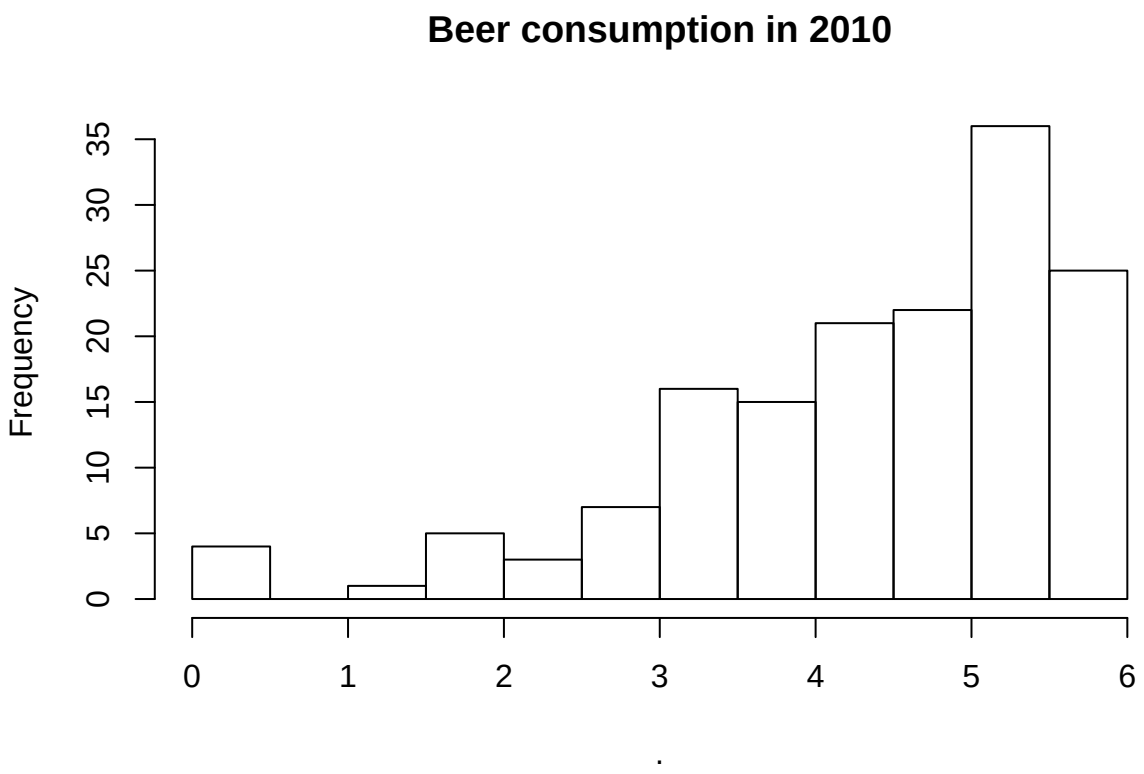


### Transformation of the data

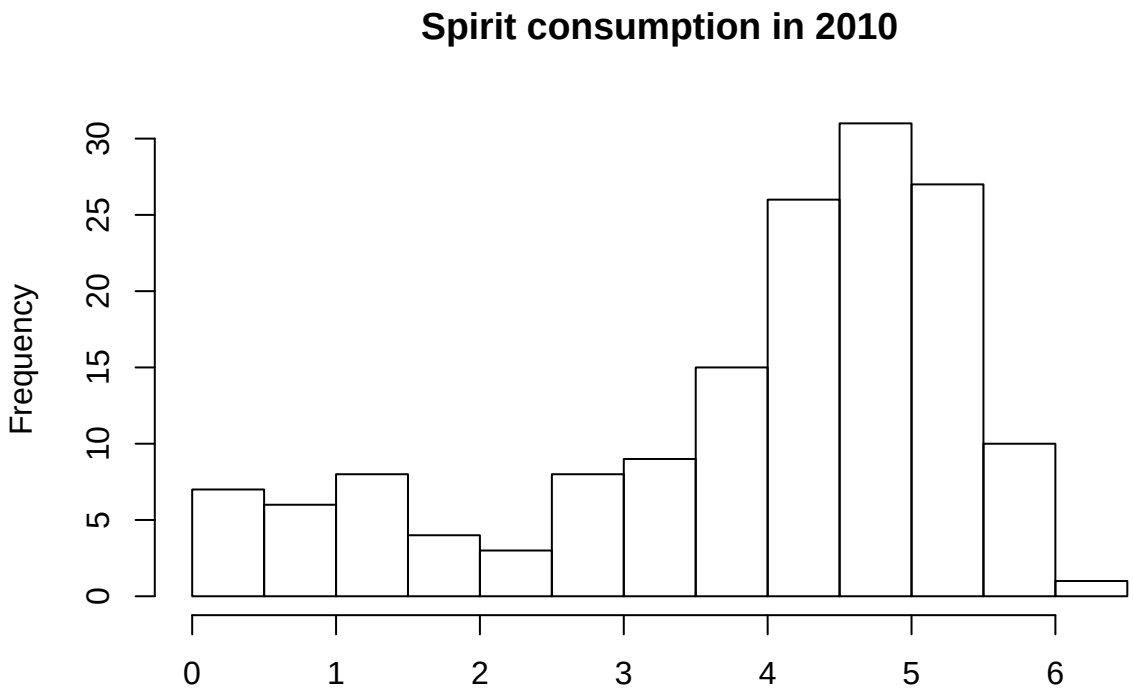
Seems like most of the countries consumed less than 50 of each in 2010. Let's unskew the distributions.

```
## Remove cases with zeros
Drinks[Drinks==0] <- NA
Drinks <- na.omit(Drinks)

## Transformation of beer consumption
log(Drinks$beer_servings) %>% hist(main = "Beer consumption in 2010")
```

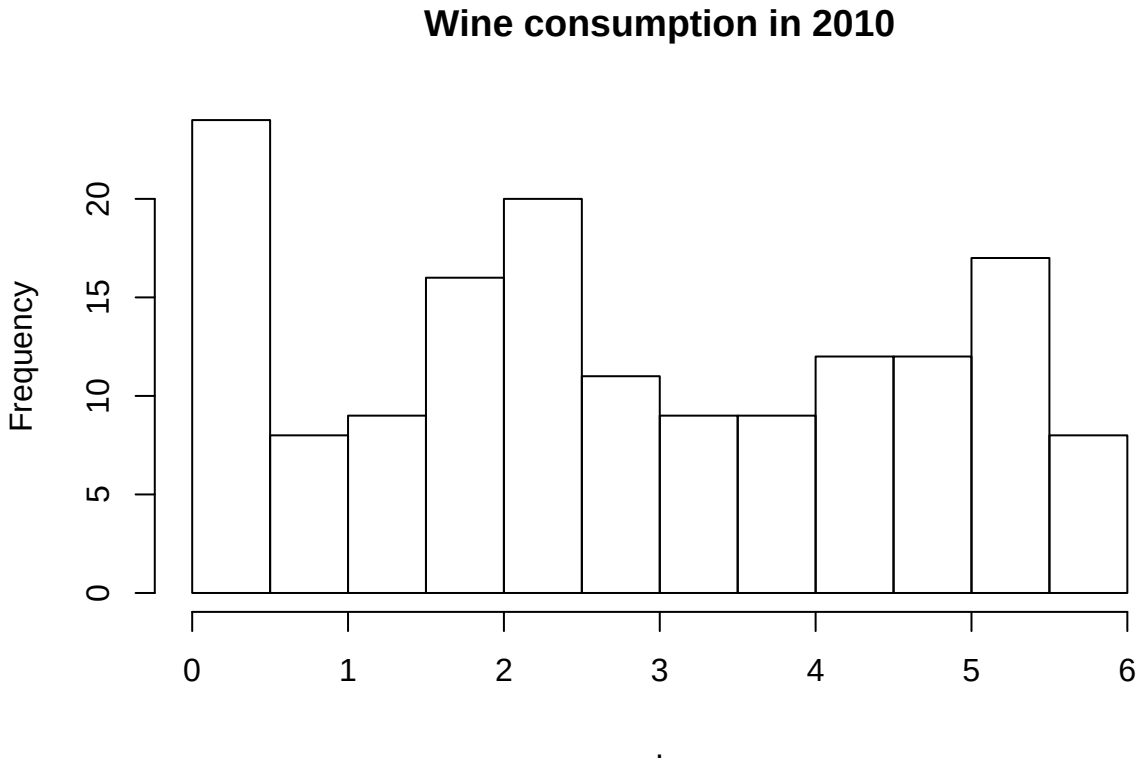



```
## Transformation of spirits consumption
log(Drinks$spirit_servings) %>% hist(main = "Spirit consumption in 2010")
```




.

```
## Transformation of spirits consumption
log(Drinks$wine_servings) %>% hist(main = "Wine consumption in 2010")
```




.

Thanks for the review. You can check the other dataset for this assignment in this link