DATA 606 Final Project

Leticia Salazar

12/09/2021

Contents

Abstract

Historically social drinking or moderate alcohol consumption holds an important role in social engagement for many. According to WHO, World Health Organization, the recorded alcohol consumption per capita (years 15+) in the United States is 47% beer, 35% spirits and 18% wine. The purpose of this project is to get more insight on how alcohol consumption varies between the countries presented in the data set below. Being that alcohol is consumed worldwide, there is no statistical evidence that there is a preferred type of alcohol countries are more inclined in drinking. I will be performing a Linear Regression model in hopes to find a correlation between the alcohol types: beer, spirits, and wine along with total liters of pure alcohol and be able to target what is the most popular alcohol type.

Introduction

Data was collected from FiveThirtyEight's article called "Dear Mona Followup: Where Do People Drink The Most Beer, Wine And Spirits?" This data was collected by World Health Organization, Global Information System on Alcohol and Health (GISAH), 2010. Each of the 193 total observations in this case represents a country around the world along with their beer, spirits and/or wine number of servings, as well as the total liters of pure alcohol.

Research Question In general, is there a significant difference in the preferred type of alcohol?

Type of study This is an observational study.

Dependent Variable The dependent variable is alcohol consumption and it is quantitative.

Independent Variable The independent variables are country and types of alcohol and they are qualitative.

Data

```
# load libraries
library("DATA606")
```

```
##
## Welcome to CUNY DATA606 Statistics and Probability for Data Analytics
## This package is designed to support this course. The text book used
## is OpenIntro Statistics, 4th Edition. You can read this by typing
## vignette('os4') or visit www.OpenIntro.org.
##
## The getLabs() function will return a list of the labs available.
##
## The demo(package='DATA606') will list the demos that are available.
```

```
library("tidyverse")
library("tidyr")
library("dplyr")
library("ggplot2")
library("infer")
library("openintro")
library("GGally")
```

```
# load data
drinks <- read.csv("https://raw.githubusercontent.com/letisalba/Data-606/main/Project/drinks.csv", head
```

Load Data

Exploratory Data Analysis

```
glimpse(drinks)
```

```
## Rows: 193
## Columns: 5
## $ country                       <chr> "Afghanistan", "Albania", "Algeria", "And~
## $ beer_servings                 <int> 0, 89, 25, 245, 217, 102, 193, 21, 261, 2~
## $ spirit_servings               <int> 0, 132, 0, 138, 57, 128, 25, 179, 72, 75,~
## $ wine_servings                 <int> 0, 54, 14, 312, 45, 45, 221, 11, 212, 191~
## $ total_litres_of_pure_alcohol <dbl> 0.0, 4.9, 0.7, 12.4, 5.9, 4.9, 8.3, 3.8, ~
```

```
# Check for any missing values
sum(is.na(drinks))
```

```
## [1] 0
```

```
summary(drinks)
```

```
##    country          beer_servings   spirit_servings  wine_servings
##  Length:193         Min.   :  0.0   Min.   :  0.00   Min.   :  0.00
##  Class :character   1st Qu.: 20.0   1st Qu.:  4.00   1st Qu.:  1.00
##  Mode  :character   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
##  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
```

```
# Get column names
names(drinks)
```

Data Wrangling

```
## [1] "country"                       "beer_servings"
## [3] "spirit_servings"               "wine_servings"
## [5] "total_litres_of_pure_alcohol"
```

```r
# Rename columns
colnames(drinks) <- c("Country", "Beer_Servings", "Spirit_Servings", "Wine_Servings", "Total_Litres_Pur
head(drinks, n = 5)
```

```
##       Country Beer_Servings Spirit_Servings Wine_Servings
## 1 Afghanistan             0               0             0
## 2     Albania            89             132            54
## 3     Algeria            25               0            14
## 4     Andorra           245             138           312
## 5      Angola           217              57            45
##   Total_Litres_Pure_Alcohol
## 1                       0.0
## 2                       4.9
## 3                       0.7
## 4                      12.4
## 5                       5.9
```

```r
# commenting out so that it doesn't replicate in my system

#write.csv(drinks, "/Users/letiix3/Desktop/Data-606/Project/csv file/drinks2.csv")
```

Write .CSV

Inference

Hypthesis Testing:

H_0**:** There is no preference between alcohol types consumed per country.

H_A**:** There is a preference between alcohol types consumed per country.

Number of countries with less than 1 and 0 `Total_Litres_Pure_Alcohol` *Out of 40 countries with less than 1 `Total_Litres_Pure_Alcohol` there are 13 counties that have 0 `Total_Litres_Pure_Alcohol`. This leaves 153 countries with an alcohol consumption greater than 1.*

```r
# Countries with less than 1 liter in  alcohol consumption
low_alcohol <- subset(drinks, Total_Litres_Pure_Alcohol < 1)
head(low_alcohol, n = 10)
```

```
##        Country Beer_Servings Spirit_Servings Wine_Servings
## 1  Afghanistan             0               0             0
## 3      Algeria            25               0            14
## 14  Bangladesh             0               0             0
## 20      Bhutan            23               0             0
## 25      Brunei            31               2             1
## 35        Chad            15               1             1
## 39     Comoros             1               3             1
## 47 North Korea             0               0             0
## 54       Egypt             6               4             1
## 57     Eritrea            18               0             0
##    Total_Litres_Pure_Alcohol
## 1                        0.0
## 3                        0.7
## 14                       0.0
## 20                       0.4
## 25                       0.6
## 35                       0.4
## 39                       0.1
## 47                       0.0
## 54                       0.2
## 57                       0.5
```

```
# Countries with No alcohol consumption
no_alcohol <- subset(drinks, Total_Litres_Pure_Alcohol == 0)
no_alcohol
```

```
##             Country Beer_Servings Spirit_Servings Wine_Servings
## 1       Afghanistan             0               0             0
## 14       Bangladesh             0               0             0
## 47      North Korea             0               0             0
## 80             Iran             0               0             0
## 91           Kuwait             0               0             0
## 98            Libya             0               0             0
## 104        Maldives             0               0             0
## 107 Marshall Islands             0               0             0
## 108      Mauritania             0               0             0
## 112          Monaco             0               0             0
## 129        Pakistan             0               0             0
## 148      San Marino             0               0             0
## 159         Somalia             0               0             0
##     Total_Litres_Pure_Alcohol
## 1                           0
## 14                          0
## 47                          0
## 80                          0
## 91                          0
## 98                          0
## 104                         0
## 107                         0
## 108                         0
## 112                         0
## 129                         0
## 148                         0
```

```
## 159                          0
```

Find Mean, Median, Standard Deviation, IQR, Standard Error, Margin of Error for `Beer_Servings`

```
# Mean
mean_beer <- mean(drinks$Beer_Servings)
  print (paste0("Mean for `Beer_Servings` = ", mean_beer))
```

```
## [1] "Mean for `Beer_Servings` = 106.160621761658"
```

```
# Median
median_beer <- median(drinks$Beer_Servings)
 print (paste0("Median for `Beer_Servings` = ", median_beer))
```

```
## [1] "Median for `Beer_Servings` = 76"
```

```
# Standard Deviation
sd_beer <- sd(drinks$Beer_Servings)
 print (paste0("Standard Deviation for `Beer_Servings` = ", sd_beer))
```

```
## [1] "Standard Deviation for `Beer_Servings` = 101.143102539313"
```

```
# IQR
iqr_beer <- IQR(drinks$Beer_Servings, na.rm = TRUE)
 print (paste0("IQR for `Beer_Servings` = ", iqr_beer))
```

```
## [1] "IQR for `Beer_Servings` = 168"
```

```
# Standard Error
se_beer <- sd_beer / sqrt(nrow(drinks))
 print (paste0("Standard Error for `Beer_Servings` = ", se_beer))
```

```
## [1] "Standard Error for `Beer_Servings` = 7.28043982873881"
```

```
# Margin of Error
z <- 1.96

me_beer <- z * se_beer
 print (paste0("Margin of Error for `Beer_Servings` = ", me_beer))
```

```
## [1] "Margin of Error for `Beer_Servings` = 14.2696620643281"
```

The probability that a randomly chosen country has a `Beer_Servings` more than 100 is 52.43%

```r
#Probability that a randomly chosen country has a total liters of pure alcohol more than 10
mean_beer <- 106.1606
sd_beer <- 101.1431

P_100_beer <- 1 - pnorm(100, mean_beer, sd_beer)
 print (paste0("The probability that a randomly chosen country has `Beer_Servings` more than 10 is = ",
```

```
## [1] "The probability that a randomly chosen country has `Beer_Servings` more than 10 is = 0.52428445
```

Find Mean, Median, Standard Deviation, IQR, Standard Error, Margin of Error for Spirit_Servings

```r
# Mean
mean_spirit <- mean(drinks$Spirit_Servings)
 print (paste0("Mean for `Spirit_Servings` = ", mean_spirit))
```

```
## [1] "Mean for `Spirit_Servings` = 80.9948186528497"
```

```r
# Median
median_spirit <- median(drinks$Spirit_Servings)
 print (paste0("Median for `Spirit_Servings` = ", median_spirit))
```

```
## [1] "Median for `Spirit_Servings` = 56"
```

```r
# Standard Deviation
sd_spirit <- sd(drinks$Spirit_Servings)
 print (paste0("Standard Deviation for `Spirit_Servings` = ", sd_spirit))
```

```
## [1] "Standard Deviation for `Spirit_Servings` = 88.2843121096862"
```

```r
# IQR
iqr_spirit <- IQR(drinks$Spirit_Servings, na.rm = TRUE)
 print (paste0("IQR for `Spirit_Servings` = ", iqr_spirit))
```

```
## [1] "IQR for `Spirit_Servings` = 124"
```

```r
# Standard Error
se_spirit <- sd_spirit / sqrt(nrow(drinks))
 print (paste0("Standard Error for `Spirit_Servings` = ", se_spirit))
```

```
## [1] "Standard Error for `Spirit_Servings` = 6.35484384005658"
```

```r
# Margin of Error
z <- 1.96

me_spirit <- z * se_spirit
 print (paste0("Margin of Error for `Spirit_Servings` = ", me_spirit))
```

```
## [1] "Margin of Error for ‘Spirit_Servings‘ = 12.4554939265109"
```

The probability that a randomly chosen country has a `Spirit_Servings` more than 100 is 41.5%

```
#Probability that a randomly chosen country has a `Spirit_Servings` more than 10
mean_spirit<- 80.99482
sd_spirit <- 88.28431

P_10_spirit <- 1 - pnorm(100, mean_spirit, sd_spirit)
 print (paste0("The probability that a randomly chosen country has a `Spirit_Servings` more than 100 is
```

```
## [1] "The probability that a randomly chosen country has a ‘Spirit_Servings‘ more than 100 is = 0.414
```

Find Mean, Median, Standard Deviation, IQR, Standard Error, Margin of Error for `Wine_Servings`

```
# Mean
mean_wine <- mean(drinks$Wine_Servings)
 print (paste0("Mean for `Wine_Servings` = ", mean_wine))
```

```
## [1] "Mean for ‘Wine_Servings‘ = 49.4507772020725"
```

```
# Median
median_wine <- median(drinks$Wine_Servings)
 print (paste0("Median for `Wine_Servings` = ", median_wine))
```

```
## [1] "Median for ‘Wine_Servings‘ = 8"
```

```
# Standard Deviation
sd_wine <- sd(drinks$Wine_Servings)
 print (paste0("Standard Deviation for `Wine_Servings` = ", sd_wine))
```

```
## [1] "Standard Deviation for ‘Wine_Servings‘ = 79.6975984576301"
```

```
# IQR
iqr_wine <- IQR(drinks$Wine_Servings, na.rm = TRUE)
 print (paste0("IQR for `Wine_Servings` = ", iqr_wine))
```

```
## [1] "IQR for ‘Wine_Servings‘ = 58"
```

```
# Standard Error
se_wine <- sd_wine / sqrt(nrow(drinks))
 print (paste0("Standard Error for `Wine_Servings` = ", se_wine))
```

```
## [1] "Standard Error for ‘Wine_Servings‘ = 5.7367586666647"
```

```r
# Margin of Error
z <- 1.96

me_wine <- z * se_wine
  print (paste0("Margin of Error for `Wine_Servings` = ", me_wine))
```

```
## [1] "Margin of Error for ‘Wine_Servings‘ = 11.2440469866628"
```

The probability that a randomly chosen country has a `Wine_Servings` more than 100 is 26.3%

```r
#Probability that a randomly chosen country has a `Wine_Servings` more than 10
mean_wine <- 49.45078
sd_wine <- 79.6976

P_100_wine <- 1 - pnorm(100, mean_wine, sd_wine)
 print (paste0("The probability that a randomly chosen country has a `Wine_Servings` more than 100 = ",
```

```
## [1] "The probability that a randomly chosen country has a ‘Wine_Servings‘ more than 100 = 0.26295467
```

Find Mean, Median, Standard Deviation, IQR, Standard Error, Margin of Error for `Total_Litres_Pure_Alcohol`

```r
# Mean
mean_tlpa <- mean(drinks$Total_Litres_Pure_Alcohol)
 print (paste0("Mean for `Total_Litres_Pure_Alcohol` = ", mean_tlpa))
```

```
## [1] "Mean for ‘Total_Litres_Pure_Alcohol‘ = 4.71709844559586"
```

```r
# Median
median_tlpa <- median(drinks$Total_Litres_Pure_Alcohol)
 print (paste0("Median for `Total_Litres_Pure_Alcohol` = ", median_tlpa))
```

```
## [1] "Median for ‘Total_Litres_Pure_Alcohol‘ = 4.2"
```

```r
# Standard Deviation
sd_tlpa <- sd(drinks$Total_Litres_Pure_Alcohol)
 print (paste0("Standard Deviation for `Total_Litres_Pure_Alcohol` = ", sd_tlpa))
```

```
## [1] "Standard Deviation for ‘Total_Litres_Pure_Alcohol‘ = 3.77329816435608"
```

```r
# IQR
iqr_tlpa <- IQR(drinks$Total_Litres_Pure_Alcohol, na.rm = TRUE)
 print (paste0("IQR for `Total_Litres_Pure_Alcohol` = ", iqr_tlpa))
```

```
## [1] "IQR for ‘Total_Litres_Pure_Alcohol‘ = 5.9"
```

```r
# Standard Error
se_tlpa <- sd_tlpa / sqrt(nrow(drinks))
 print (paste0("Standard Error for `Total_Litres_Pure_Alcohol` = ", se_tlpa))
```

```
## [1] "Standard Error for ‘Total_Litres_Pure_Alcohol‘ = 0.271607945097465"
```

```r
# Margin of Error
z <- 1.96

me_tlpa <- z * se_tlpa
 print (paste0("Margin of Error for `Total_Litres_Pure_Alcohol` = ", me_tlpa))
```

```
## [1] "Margin of Error for ‘Total_Litres_Pure_Alcohol‘ = 0.53235157239103"
```

The probability that a randomly chosen country has a `Total_Litres_Pure_Alcohol` more than 10 is 8.07%

```r
#Probability that a randomly chosen country has a total litres of pure alcohol more than 10
mean_tlpa <- 4.717098
sd_tlpa <- 3.773298

P_10_tlpa <- 1 - pnorm(10, mean_tlpa, sd_tlpa)
 print (paste0("The probability that a randomly chosen country has a `Total_Litres_Pure_Alcohol` more t
```

```
## [1] "The probability that a randomly chosen country has a ‘Total_Litres_Pure_Alcohol‘ more than 100
```

Linear Regression Model

Linear models can be used for prediction or to evaluate whether there is a linear relationship between two numerical variables.

By using the Linear Regression model I will be able to evaluate whether or not there is a linear relationship for the alcohol types, `Beer_Servings`, `Spirit_Servings`, `Wine_Servings` and `Total_Litres_Pure_Alcohol`.

`Beer_Servings`

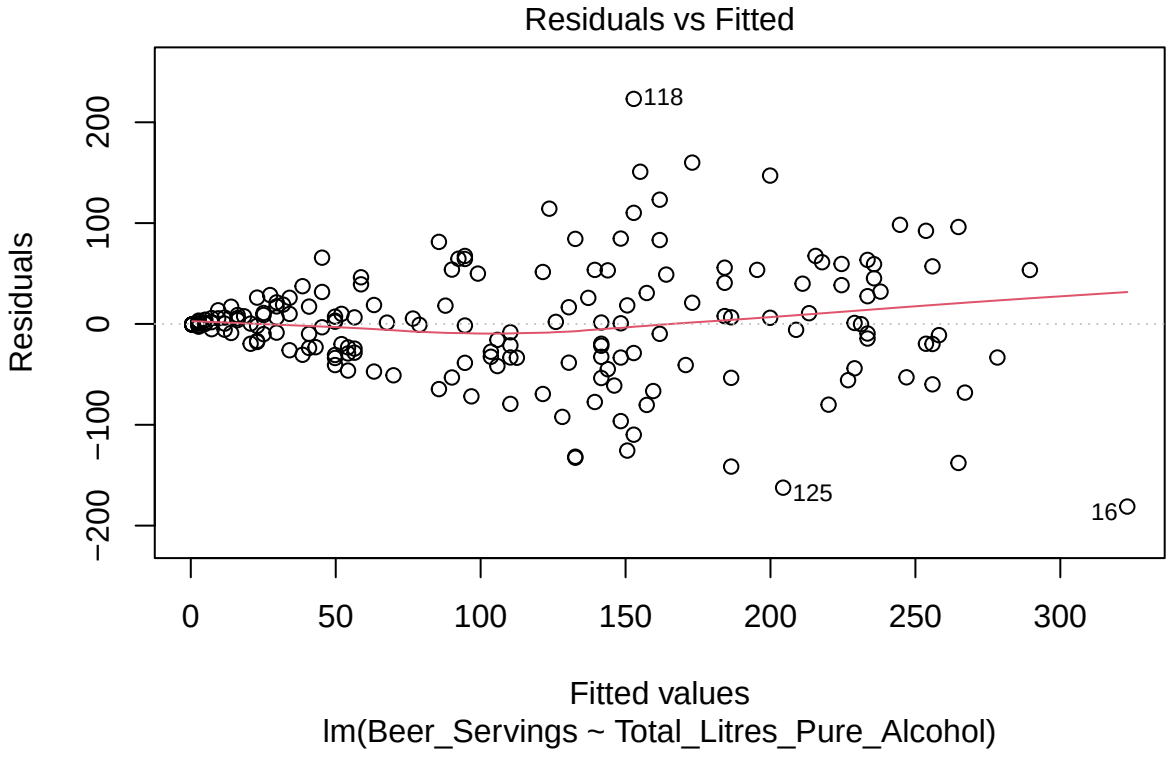
$$\hat{y} = 0.4758 + 22.4046 \times Total\_Litres\_Pure\_Alcohol$$

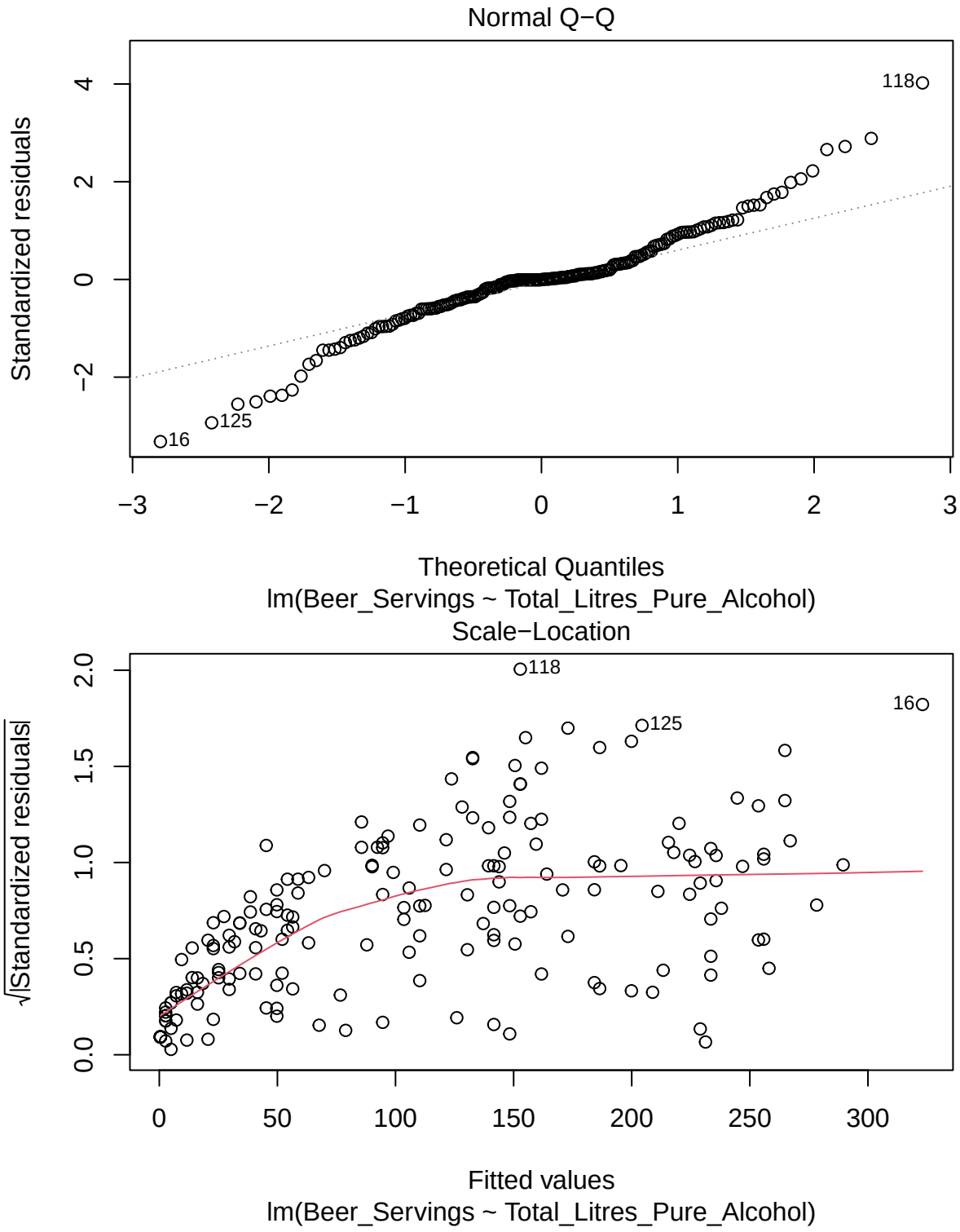
The adjusted R^2 is 69.75% and the P-Value is $<$ 2.2e-16, which is smaller than the typical threshold of 0.05.

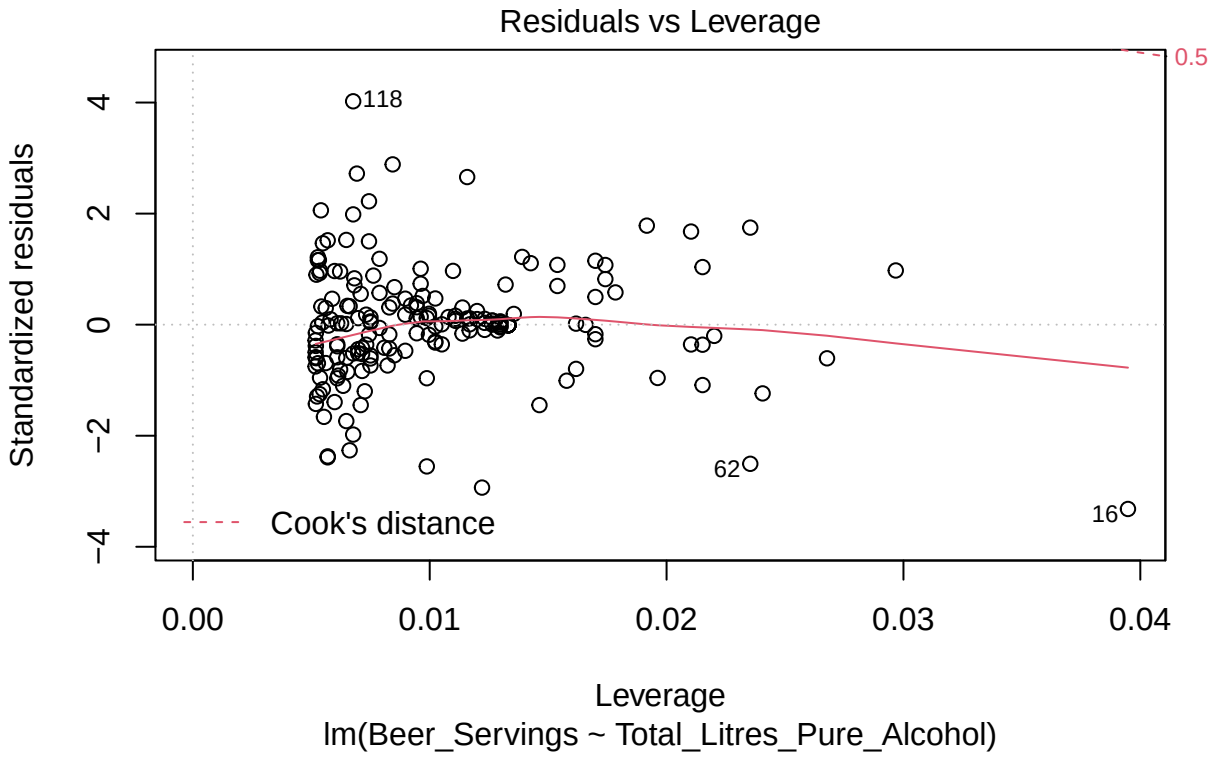
```r
# lm for `Beer_Servings` against `Total_Litres_Pure_Alcohol`
beer_model <- lm(Beer_Servings~ Total_Litres_Pure_Alcohol, data = drinks)
summary(beer_model)
```

```
##
## Call:
## lm(formula = Beer_Servings ~ Total_Litres_Pure_Alcohol, data = drinks)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -181.102  -27.537   -0.243   21.398  223.173
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  0.4758     6.4253   0.074    0.941
## Total_Litres_Pure_Alcohol   22.4046     1.0648  21.042   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 55.67 on 191 degrees of freedom
## Multiple R-squared:  0.6986, Adjusted R-squared:  0.697
## F-statistic: 442.8 on 1 and 191 DF,  p-value: < 2.2e-16
```

```
# Residual v. Fitted, Normal Probability, Scale-Location, and Residuals v. Leverage
plot(beer_model)
```




Normal Q−Q
Standardized residuals
118
125
16
−3
−2
−1
0
1
2
3
−2
0
2
4
Theoretical Quantiles
lm(Beer_Servings ~ Total_Litres_Pure_Alcohol)
Scale−Location
√|Standardized residuals|
118
125
16
0.0
0.5
1.0
1.5
2.0
0
50
100
150
200
250
300
Fitted values
lm(Beer_Servings ~ Total_Litres_Pure_Alcohol)



Spirit_Servings

$$\hat{y} = 8.708 + 15.324 \times Total\_Litres\_Pure\_Alcohol$$

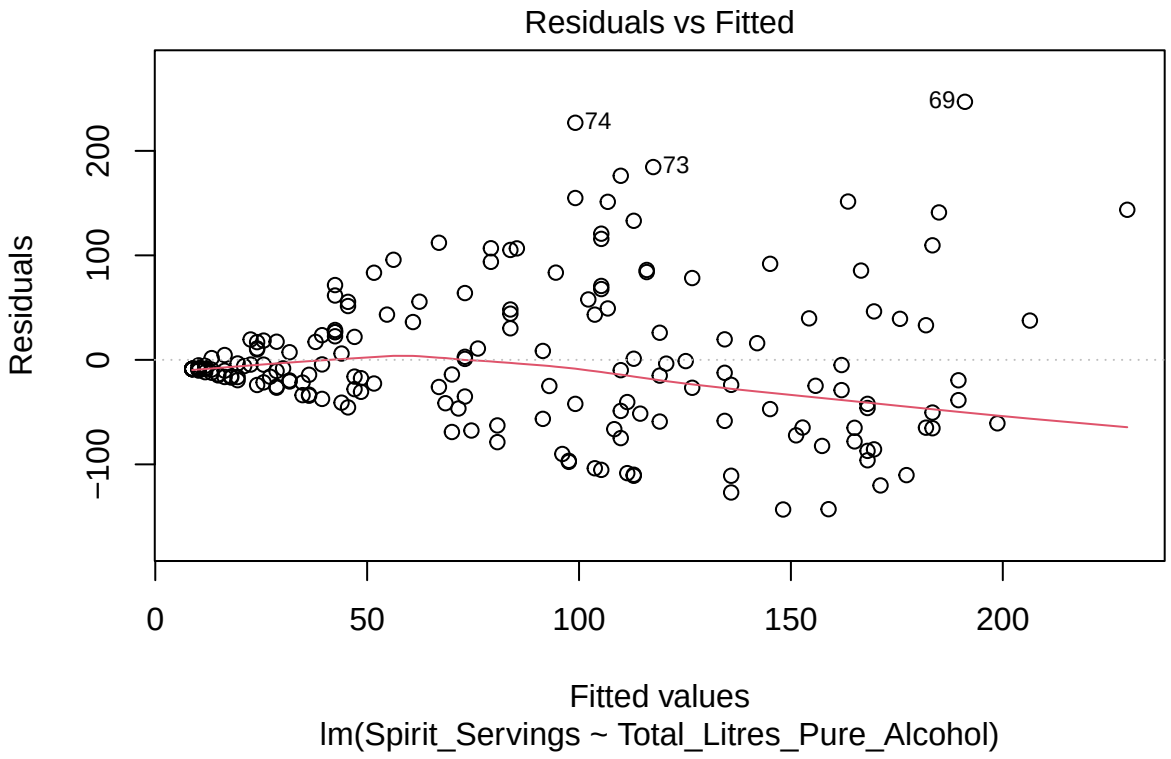
The adjusted R^2 is 42.6% and the P-Value is $<$ 2.2e-16, which is smaller than the typical threshold of 0.05.

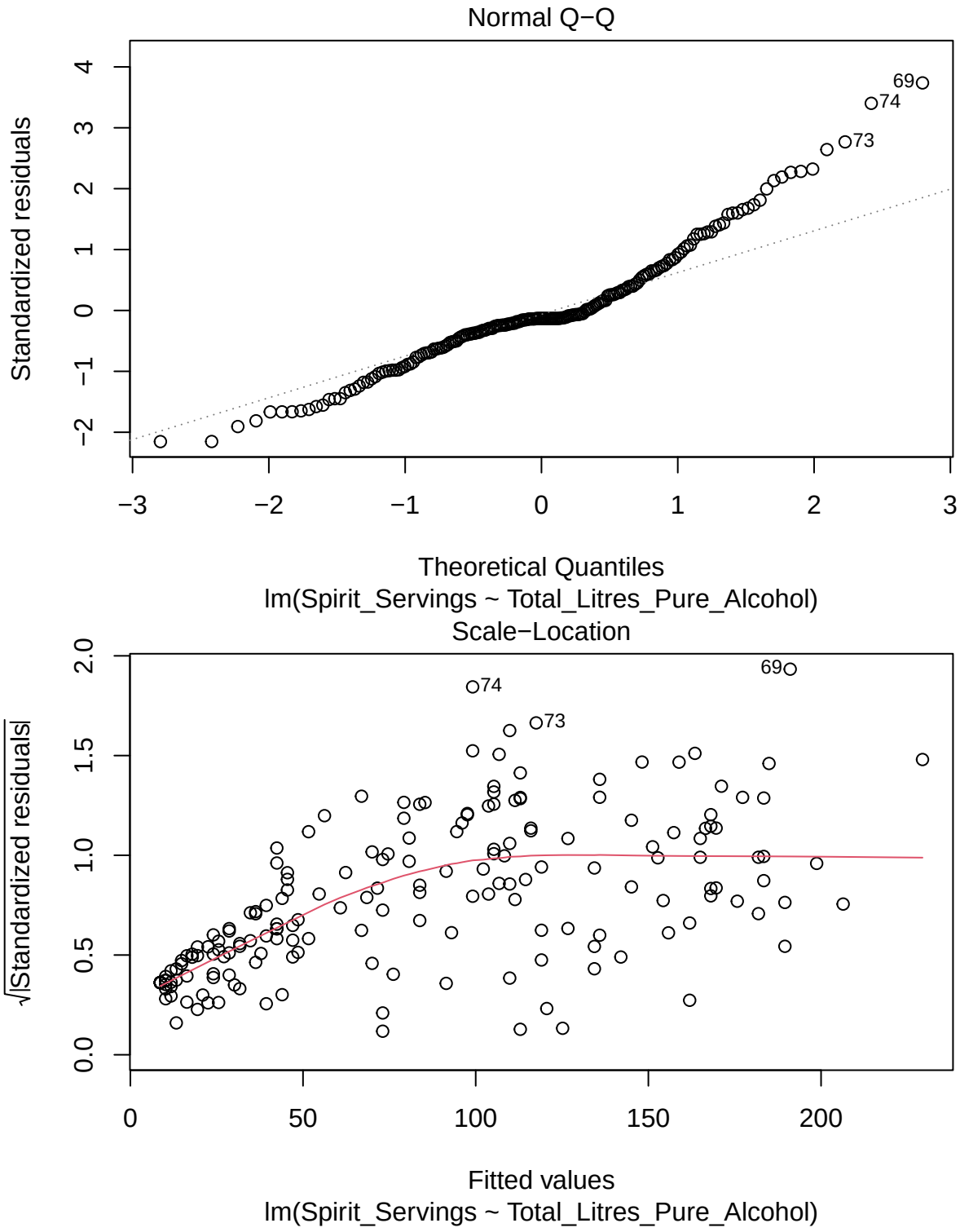
```
# lm for `Spirit_Servings` against `Total_Litres_Pure_Alcohol`
spirit_model <- lm(Spirit_Servings~Total_Litres_Pure_Alcohol, data = drinks)
summary(spirit_model)
```

```
##
## Call:
## lm(formula = Spirit_Servings ~ Total_Litres_Pure_Alcohol, data = drinks)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -143.160  -35.071   -8.708   26.578  246.932
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   8.708      7.720   1.128    0.261
## Total_Litres_Pure_Alcohol    15.324      1.279  11.979   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 66.89 on 191 degrees of freedom
```

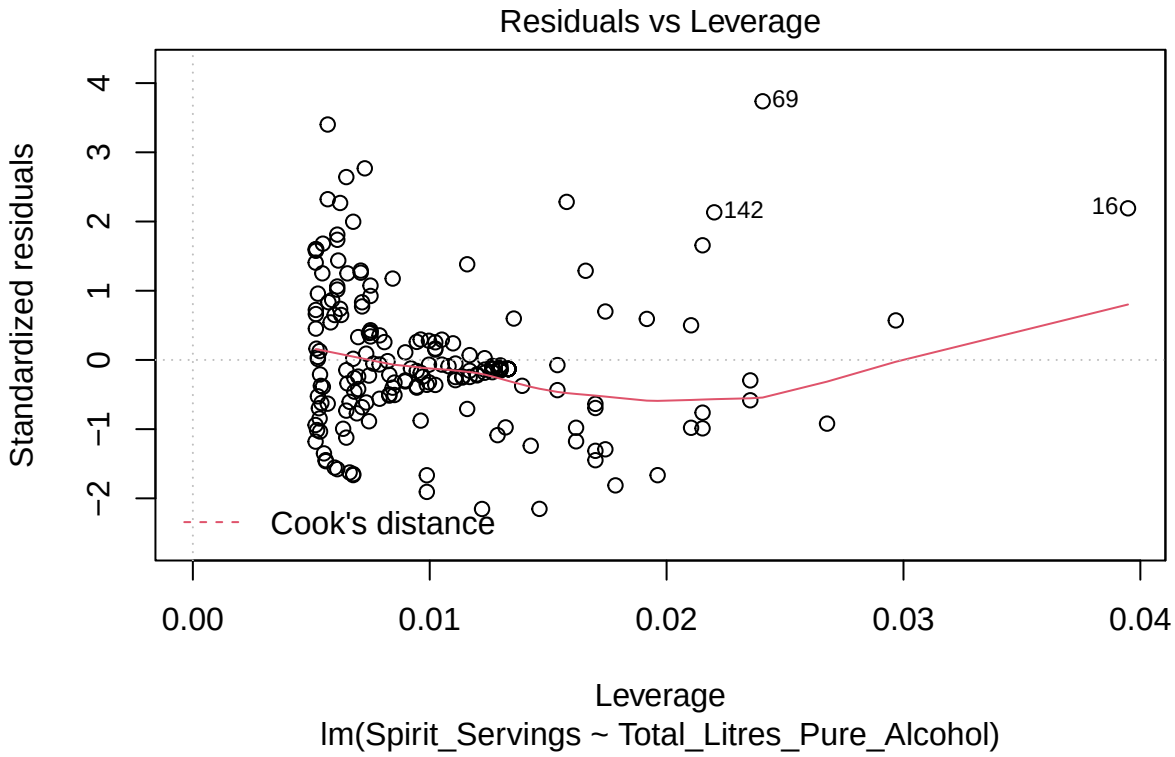
```
## Multiple R-squared:  0.429,  Adjusted R-squared:  0.426
## F-statistic: 143.5 on 1 and 191 DF,  p-value: < 2.2e-16
```

```
# Residual v. Fitted, Normal Probability, Scale-Location, and Residuals v. Leverage
plot(spirit_model)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(Spirit_Servings ~ Total_Litres_Pure_Alcohol)
69
74
73
−3
−2
−1
0
1
2
3
4

Scale−Location
√|Standardized residuals|
Fitted values
lm(Spirit_Servings ~ Total_Litres_Pure_Alcohol)
69
74
73
0.0
0.5
1.0
1.5
2.0
0
50
100
150
200



Wine_Servings

$$\hat{y} = -17.063 + 14.101 \times Total\_Litres\_Pure\_Alcohol$$

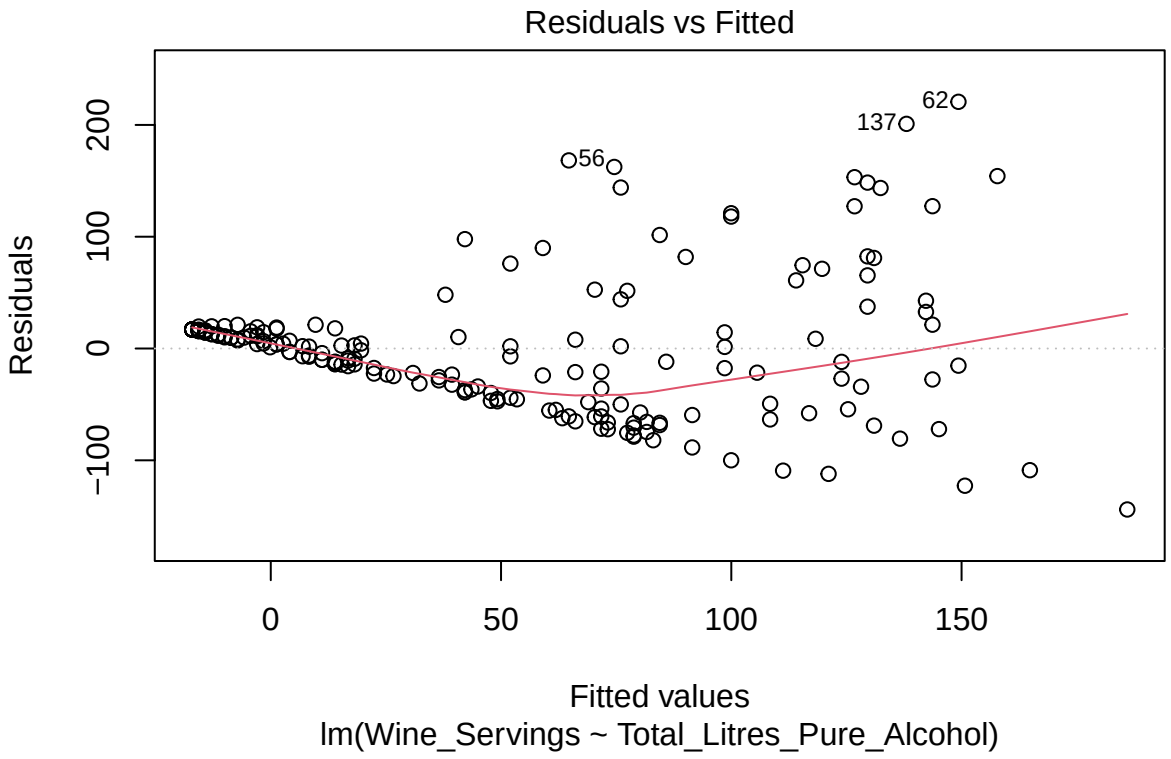
The adjusted R^2 is 44.3% and the P-Value is $<$ 2.2e-16, which is smaller than the typical threshold of 0.05.

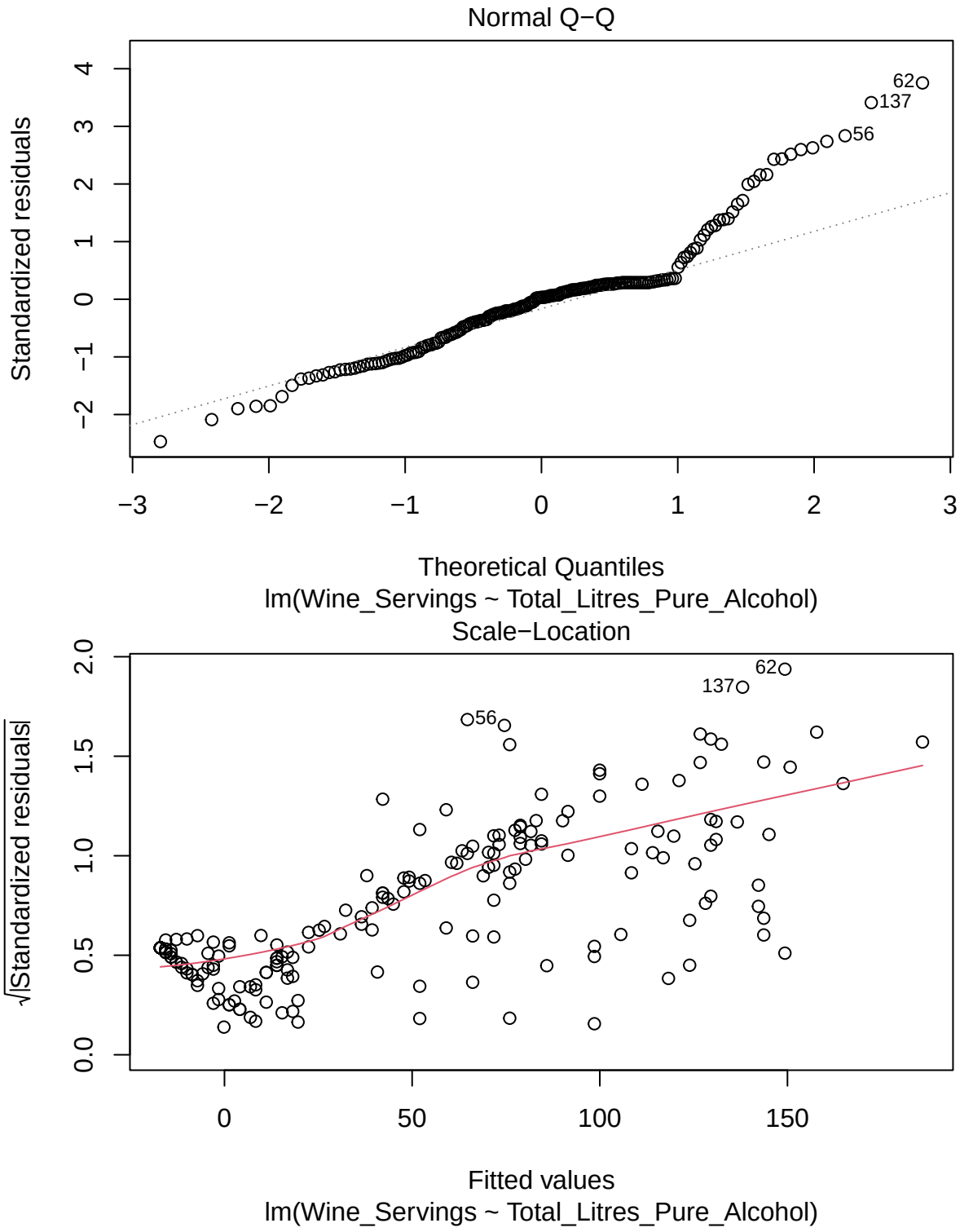
```
# lm for `Wine_Servings` against `Total_Litres_Pure_Alcohol`
wine_model <- lm(Wine_Servings~Total_Litres_Pure_Alcohol, data = drinks)
summary(wine_model)
```

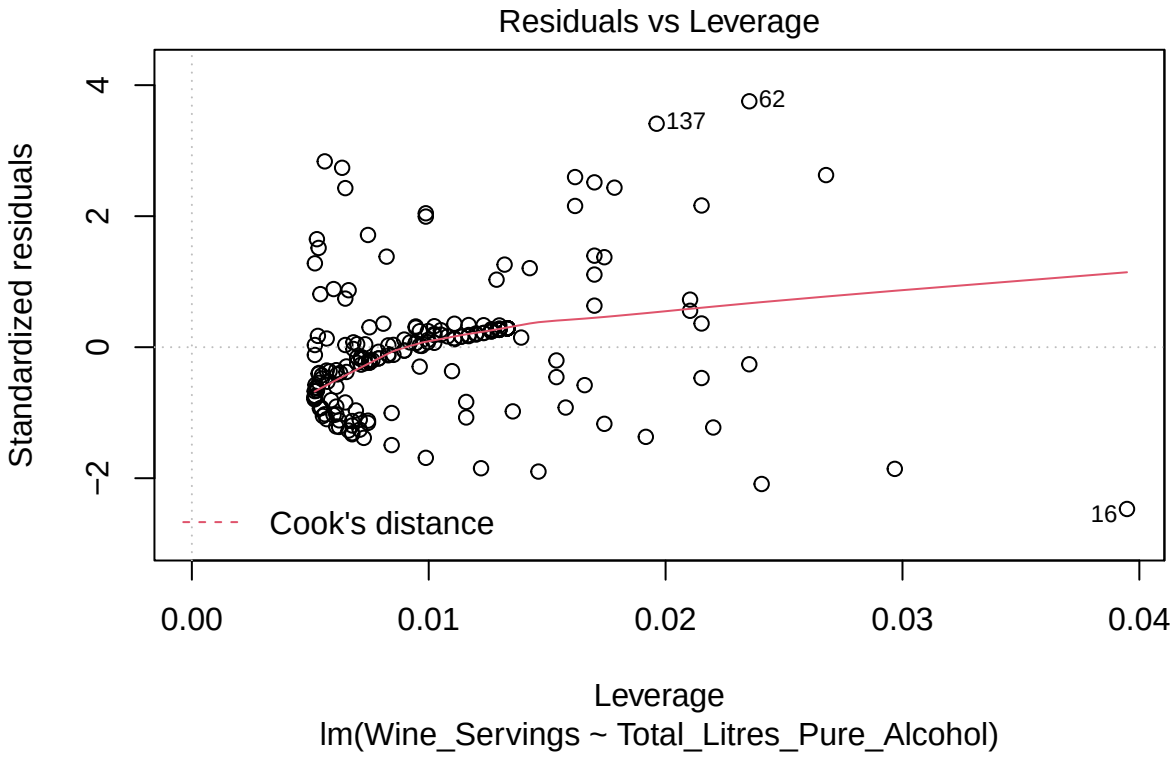
```
##
## Call:
## lm(formula = Wine_Servings ~ Total_Litres_Pure_Alcohol, data = drinks)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -143.99  -36.57    1.97   17.06  220.68
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -17.063      6.866  -2.485   0.0138 *
## Total_Litres_Pure_Alcohol    14.101      1.138  12.392   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 59.49 on 191 degrees of freedom
```

```
## Multiple R-squared:  0.4457, Adjusted R-squared:  0.4428
## F-statistic: 153.6 on 1 and 191 DF,  p-value: < 2.2e-16
```

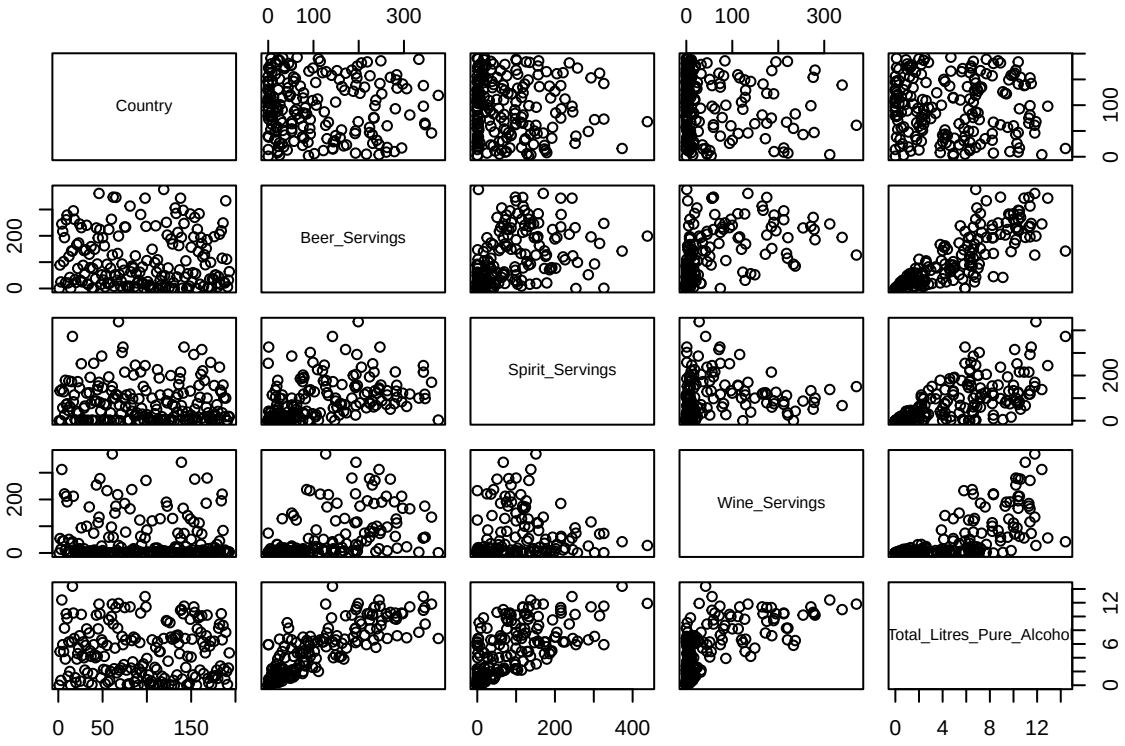
```
# Residual v. Fitted, Normal Probability, Scale-Location, and Residuals v. Leverage
plot(wine_model)
```




Normal Q−Q
Standardized residuals
62
137
56
Theoretical Quantiles
lm(Wine_Servings ~ Total_Litres_Pure_Alcohol)
Scale−Location
√|Standardized residuals|
62
137
56
Fitted values
lm(Wine_Servings ~ Total_Litres_Pure_Alcohol)

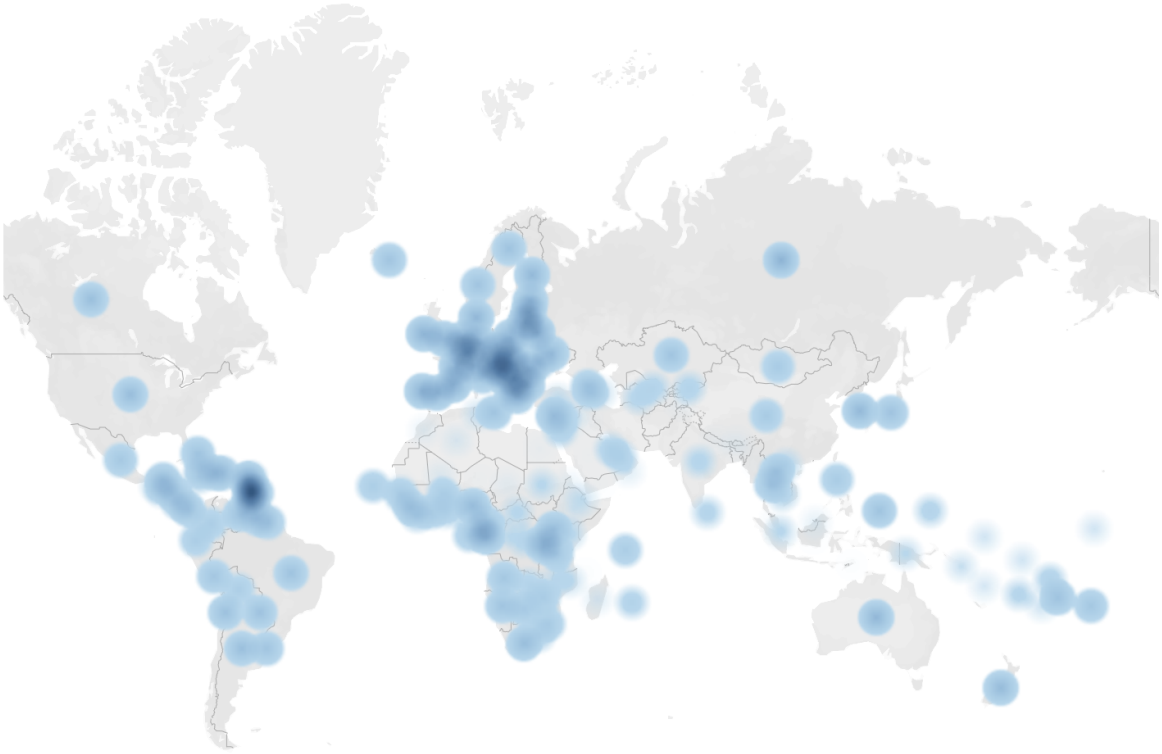



```
# Multiple plots from data set
par(mfrow = c(2, 2)) # creates a 2 x 2 plotting matrix
plot(drinks)
```



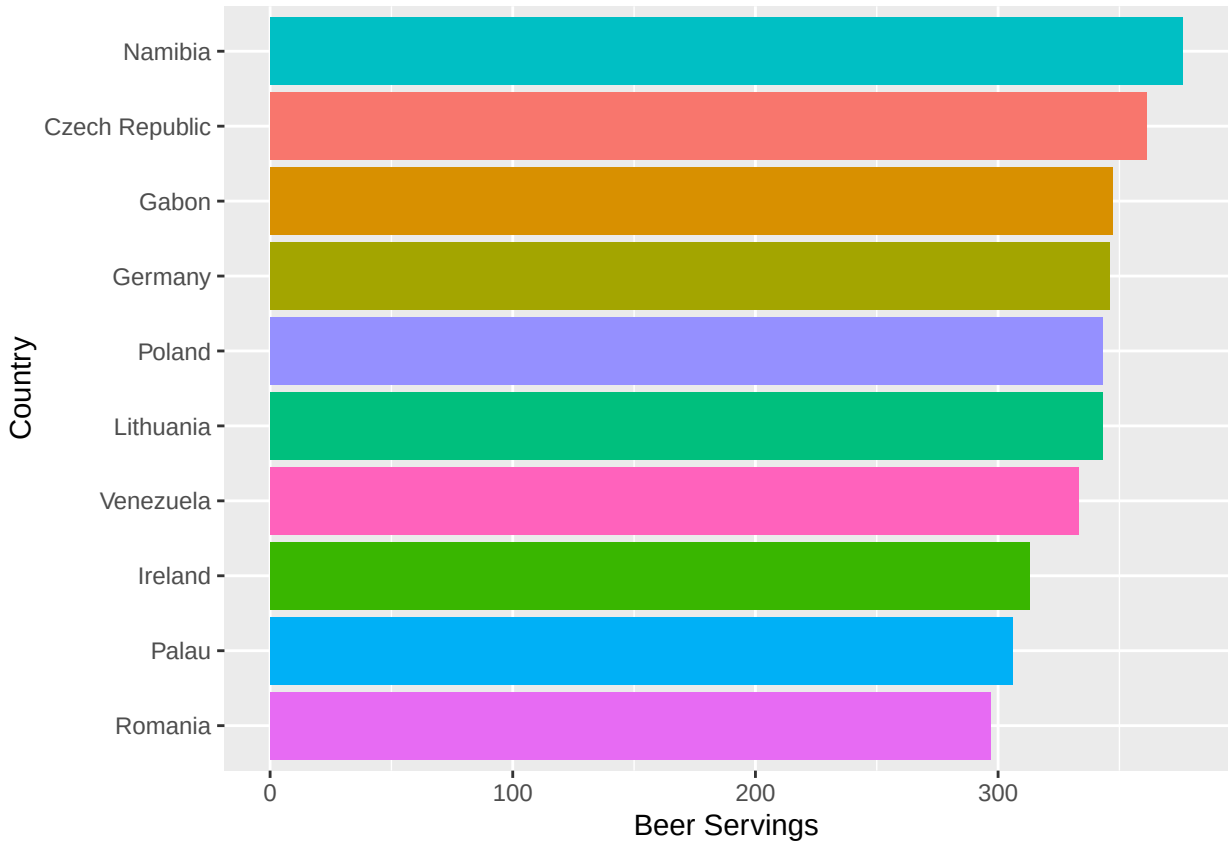
Plots

Tableau Public Visualization

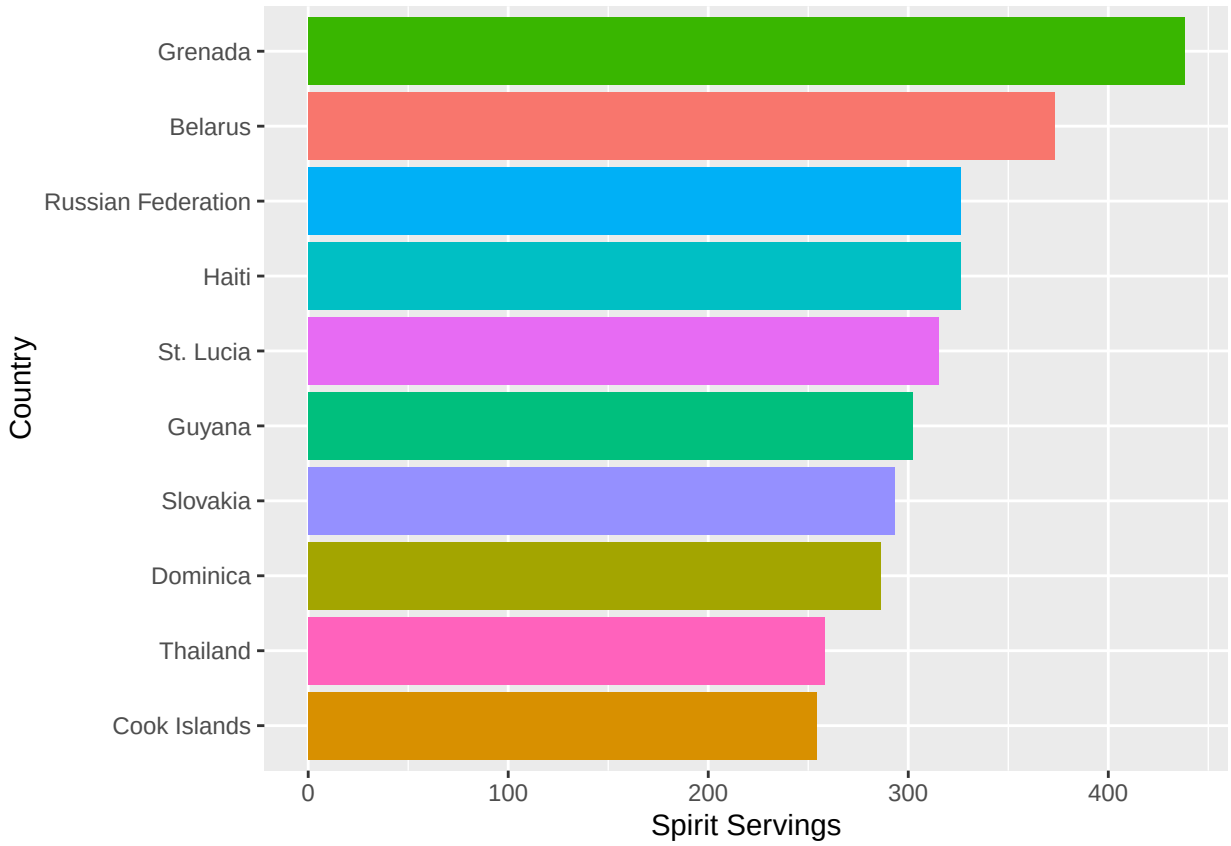


GGPPLOT

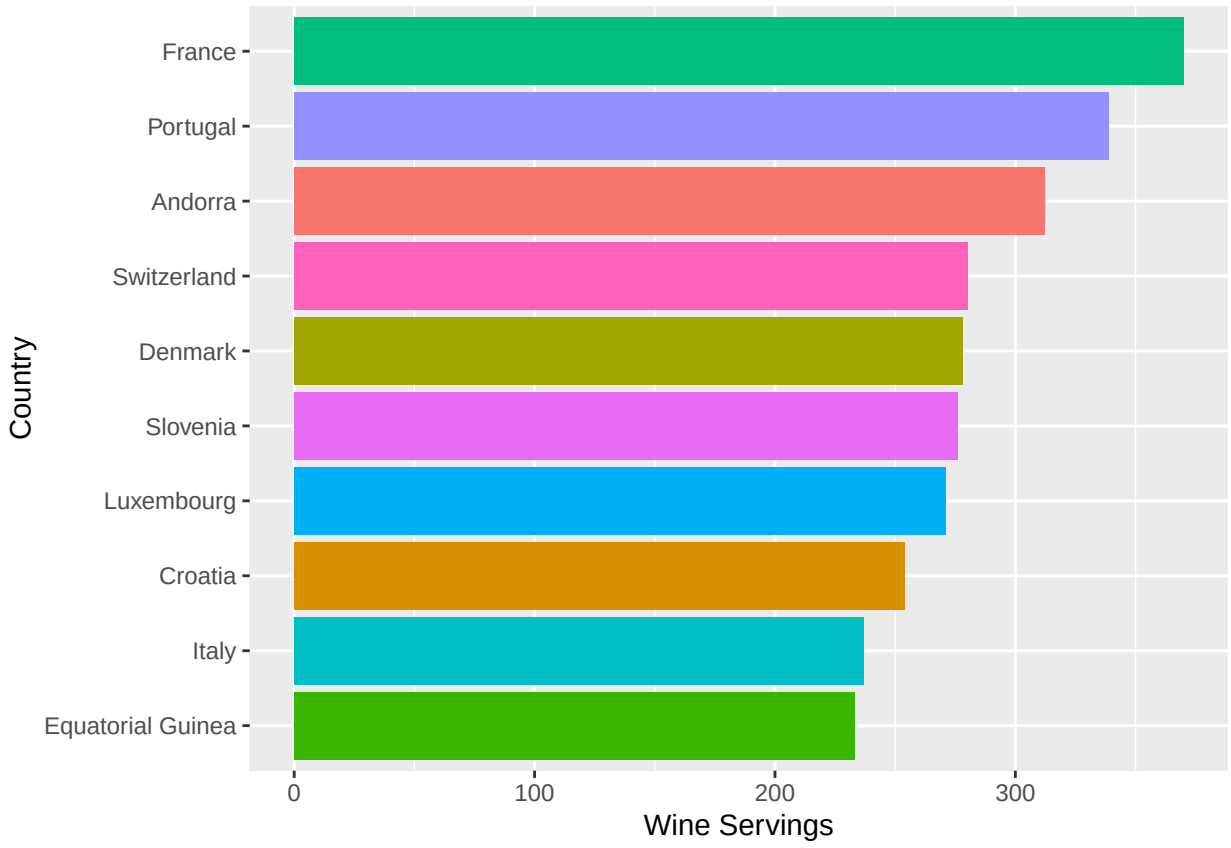
```
# Beer servings by Country; top 10
drinks %>%
  arrange(desc(Beer_Servings)) %>%
  head(10) %>%
  ggplot(aes(y = reorder(Country, Beer_Servings),
             x = Beer_Servings,
                  fill= Country))+
  geom_col()+
  xlab("Beer Servings") +
  ylab("Country") +
  theme(legend.position = "none")
```



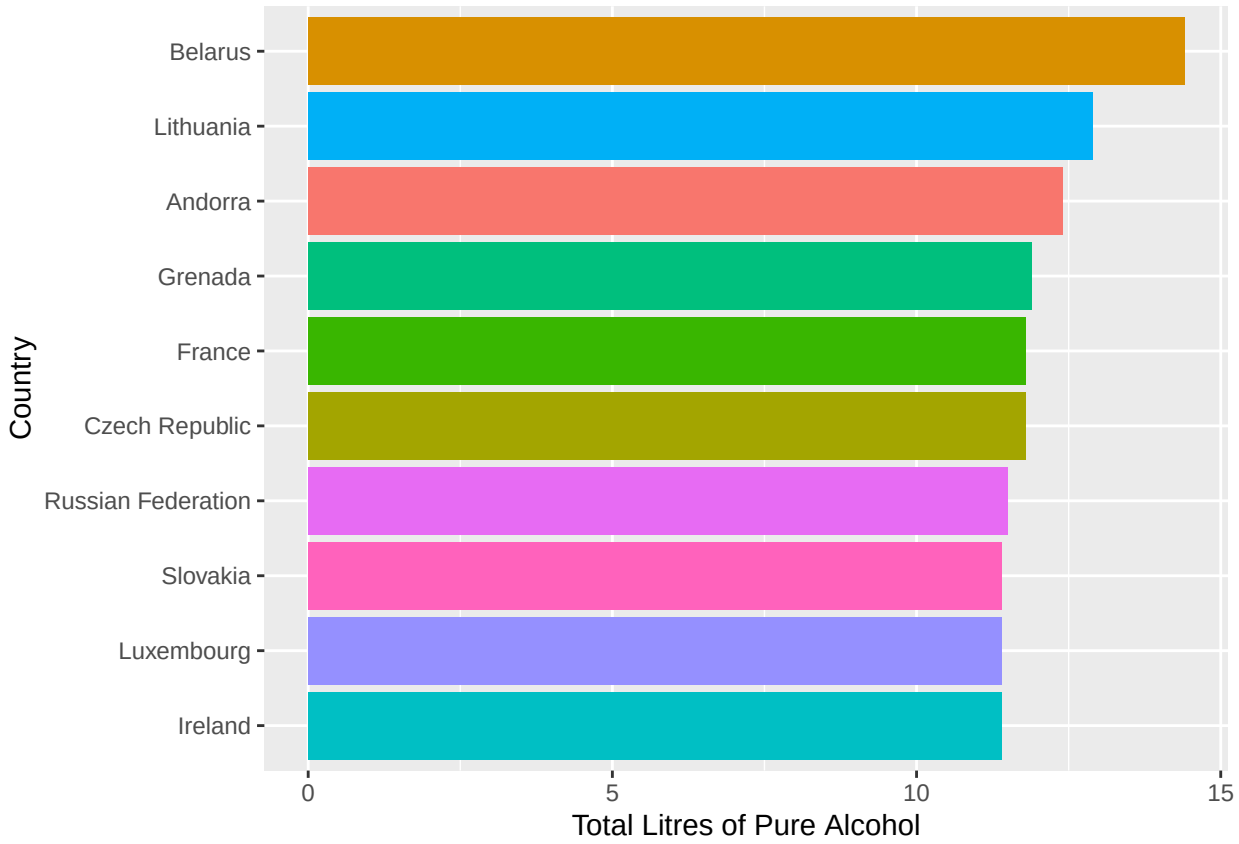

```
# Spirits serving by Country; top 10
drinks %>%
  arrange(desc(Spirit_Servings)) %>%
  head(10) %>%
  ggplot(aes(y = reorder(Country, Spirit_Servings),
             x = Spirit_Servings,
                  fill= Country))+
  geom_col()+
  xlab("Spirit Servings") +
  ylab("Country") +
  theme(legend.position = "none")
```




```
# Wine servings by Country; top 10
drinks %>%
  arrange(desc(Wine_Servings)) %>%
  head(10) %>%
  ggplot(aes(y = reorder(Country, Wine_Servings),
             x = Wine_Servings,
                  fill= Country))+
  geom_col()+
  xlab("Wine Servings") +
  ylab("Country") +
  theme(legend.position = "none")
```

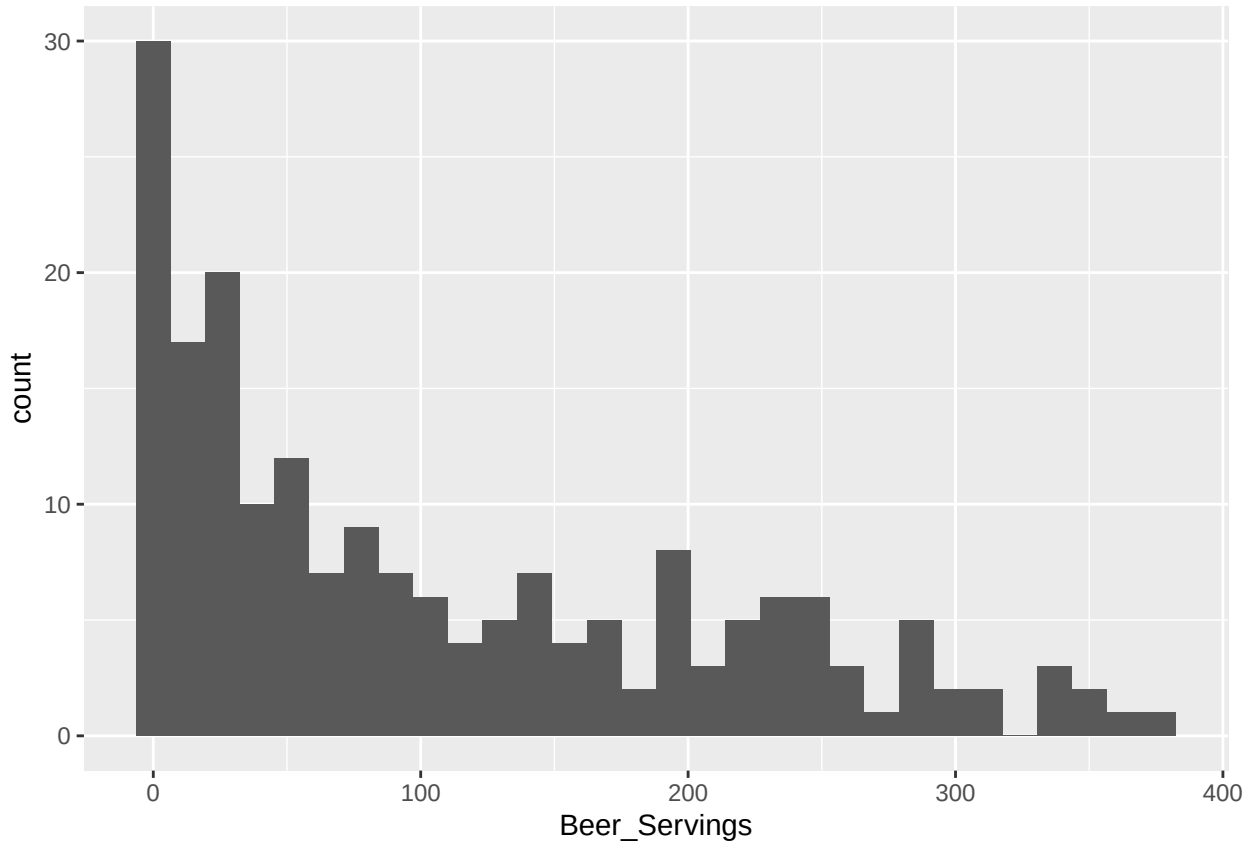



```
# Total alcohol in litres by Country; top 10
drinks %>%
  arrange(desc(Total_Litres_Pure_Alcohol)) %>%
  head(10) %>%
  ggplot(aes(y = reorder(Country, Total_Litres_Pure_Alcohol),
             x = Total_Litres_Pure_Alcohol,
                  fill= Country))+
  geom_col() +
  xlab("Total Litres of Pure Alcohol") +
  ylab("Country") +
  theme(legend.position = "none")
```



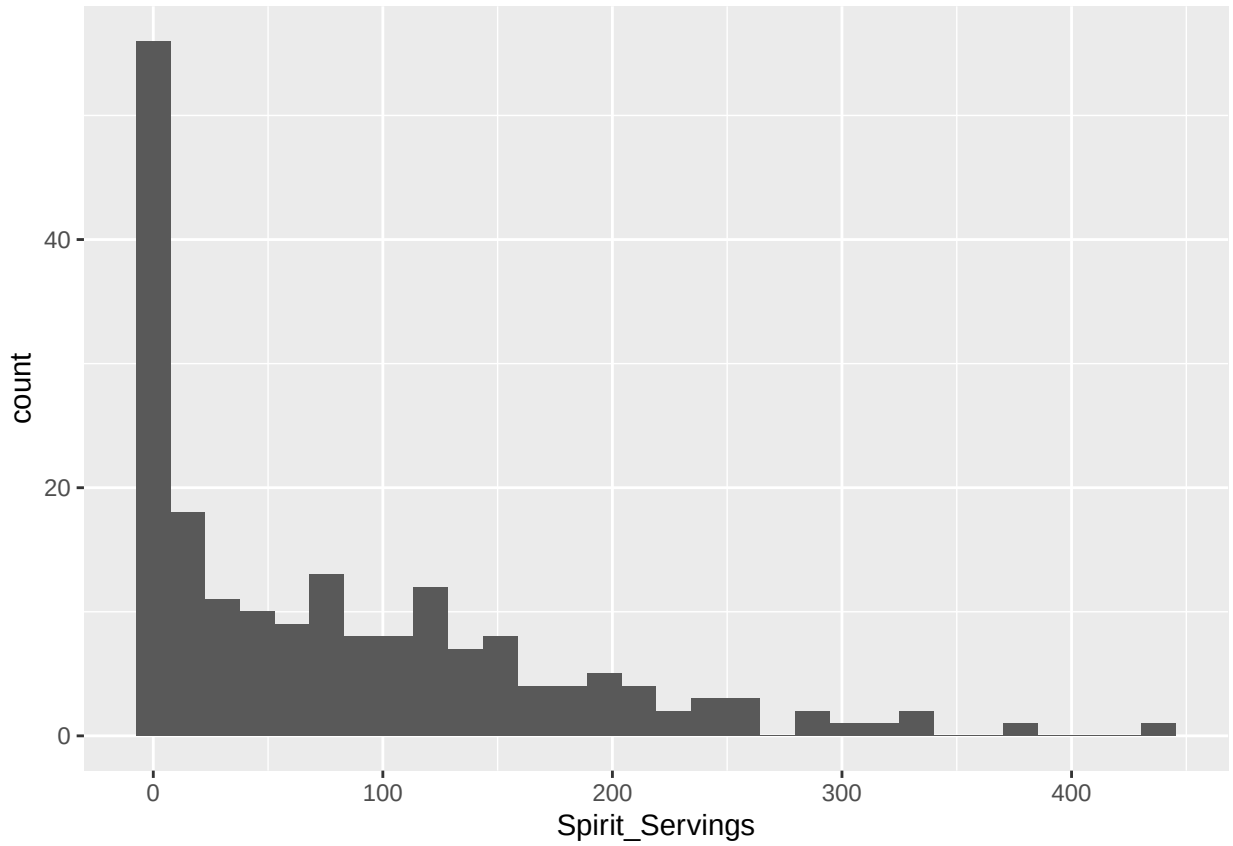

```
# Histogram for Beer_Servings
drinks %>%
ggplot(aes(x = Beer_Servings)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



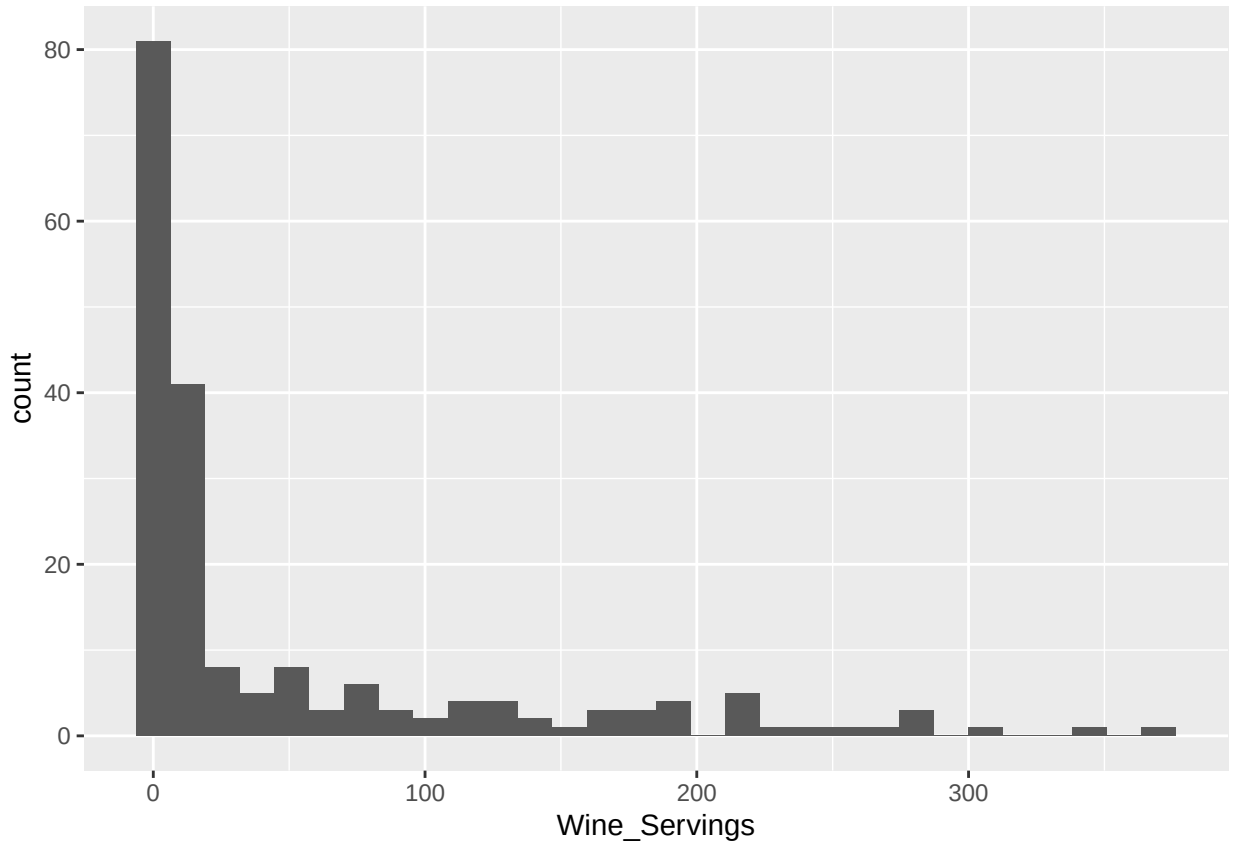

```
# Histogram for Spirit_Servings
drinks %>%
ggplot(aes(x = Spirit_Servings)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



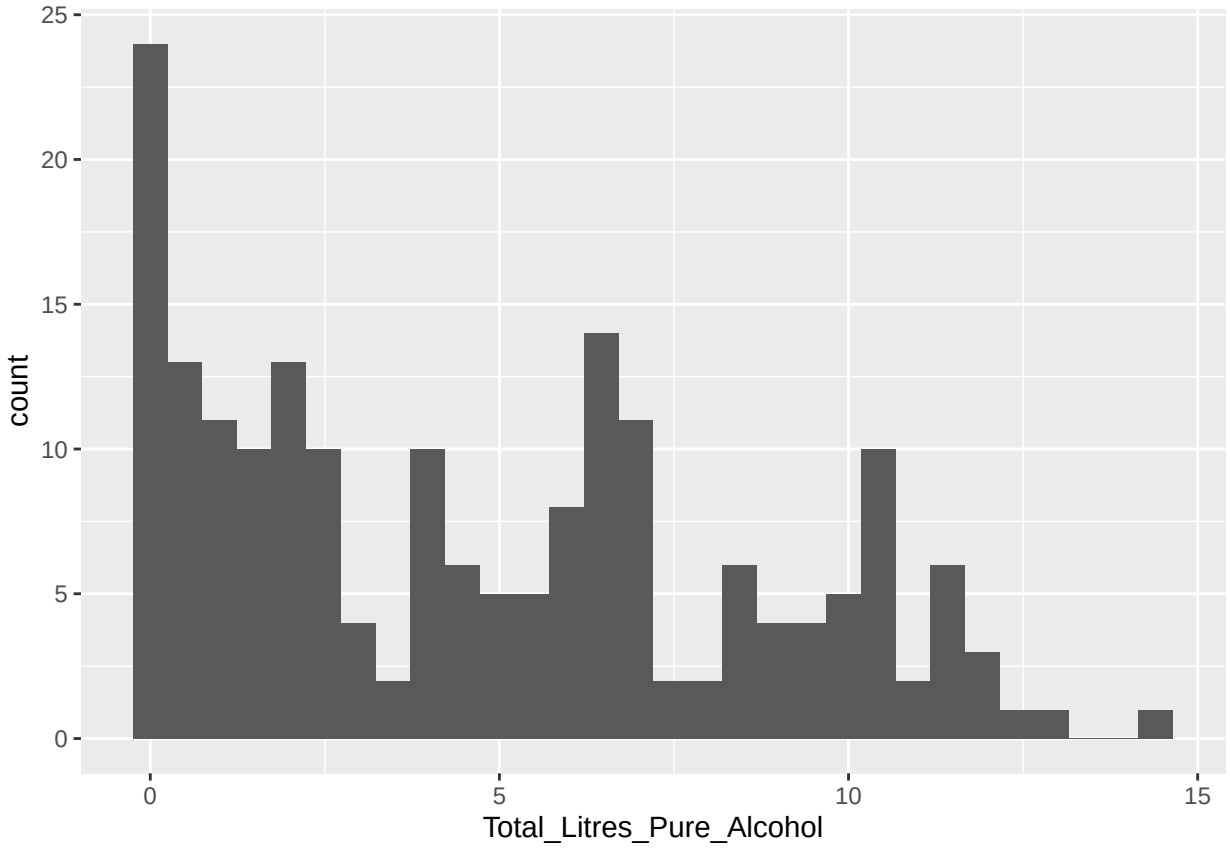

```
# Histogram for Wine_Servings
drinks %>%
ggplot(aes(x = Wine_Servings)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



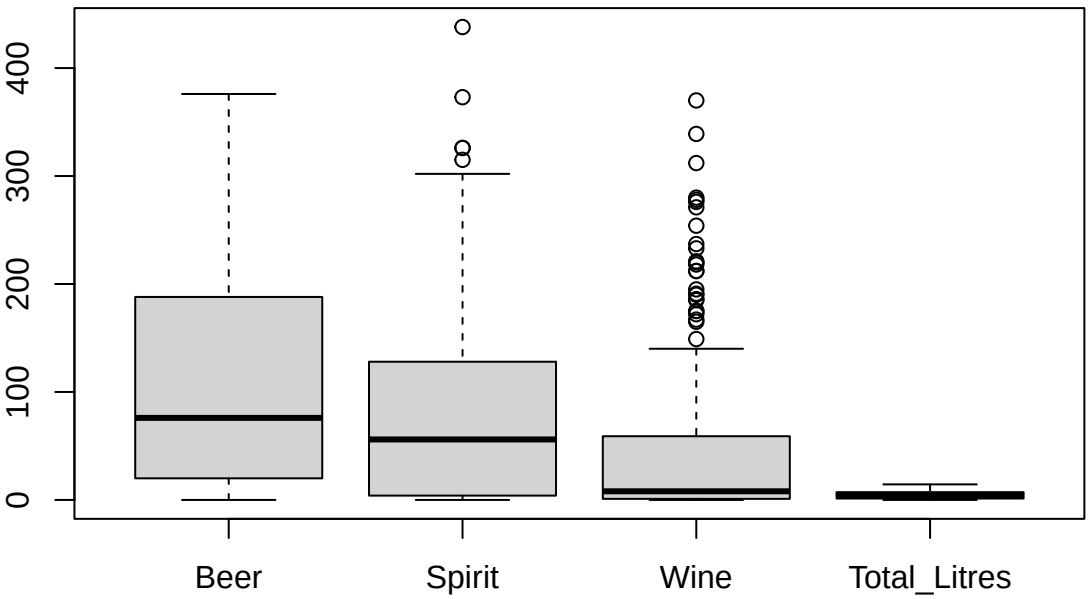

```
# Histogram for Total_Litres_Pure_Alcohol
drinks %>%
ggplot(aes(x = Total_Litres_Pure_Alcohol)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

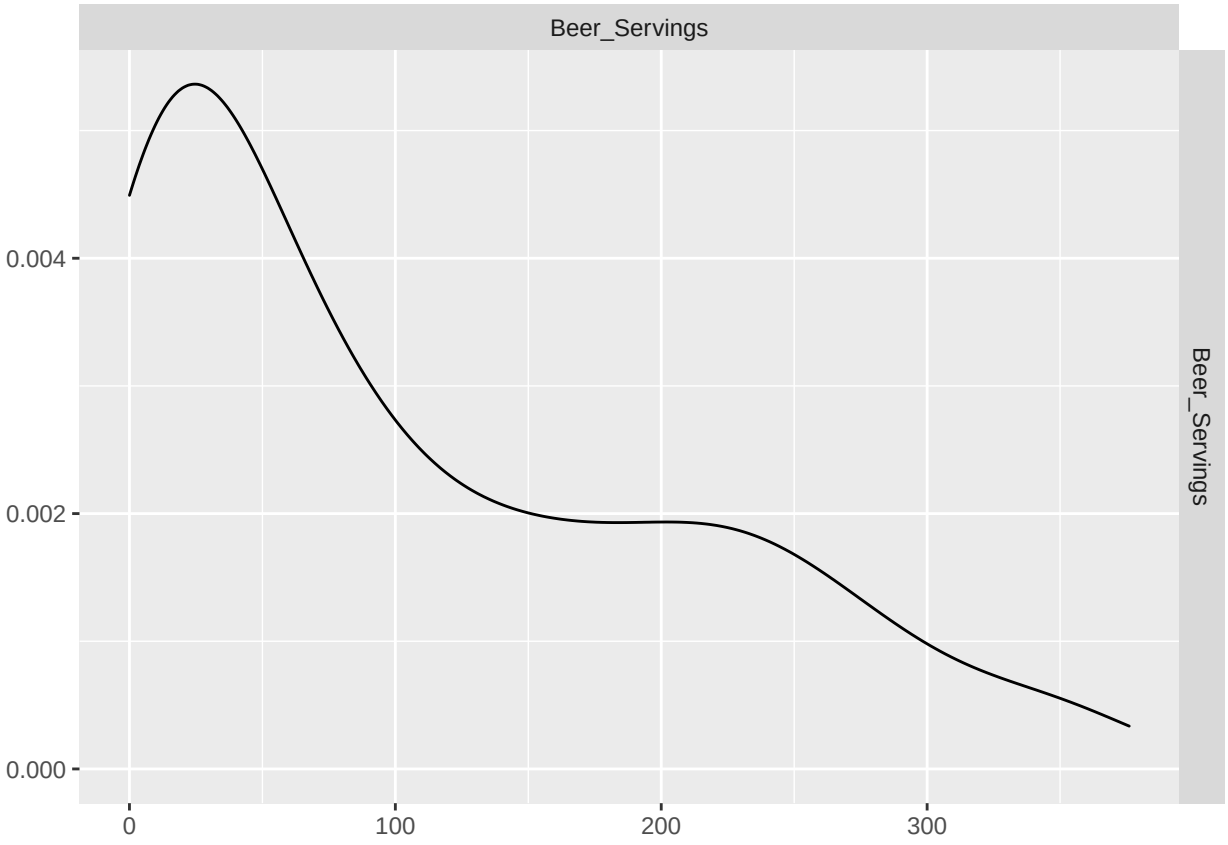


The outliers are more visible within the boxplots and you see how the mean changes per alcohol type.

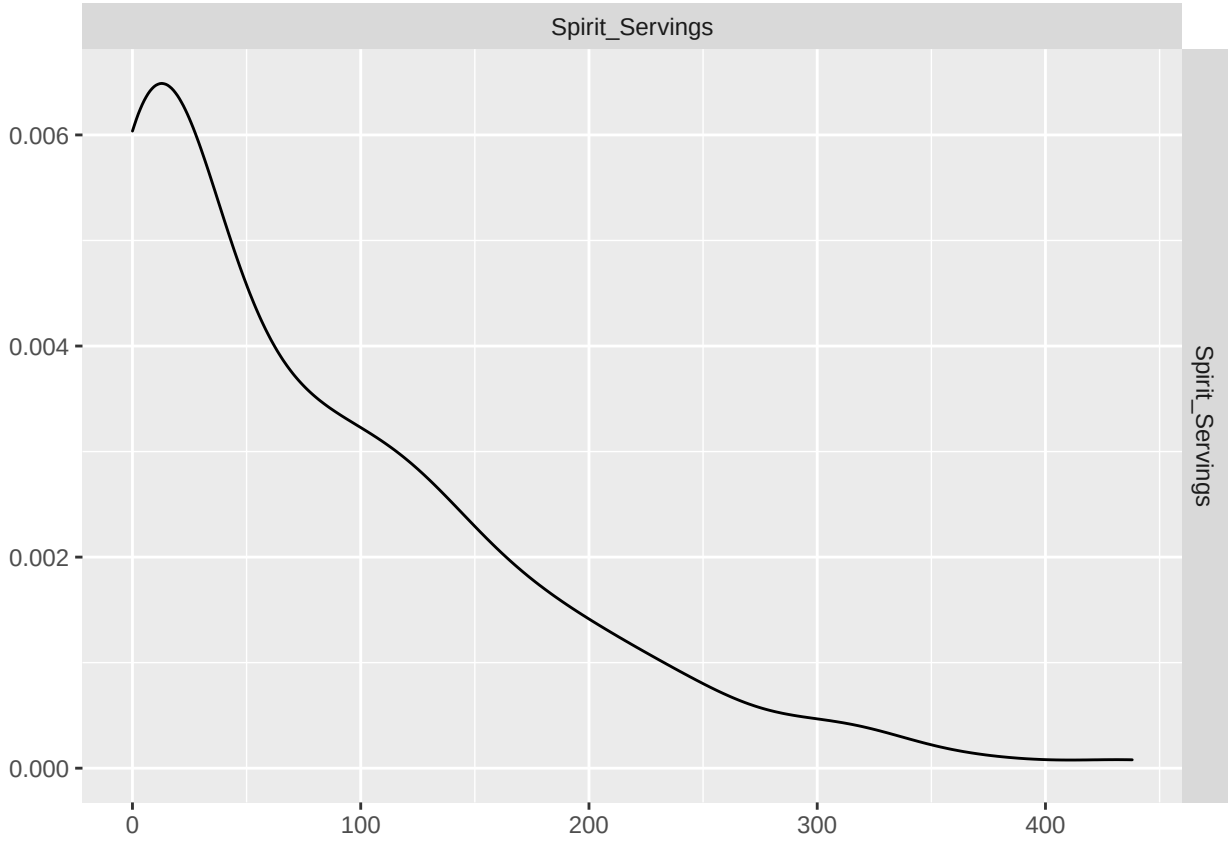
```
# Boxplots side by side alcohol types and total litres of pure alcohol
boxplot(drinks$Beer_Servings, drinks$Spirit_Servings, drinks$Wine_Servings, drinks$Total_Litres_Pure_Al
        names = c("Beer", "Spirit", "Wine", "Total_Litres"))
```



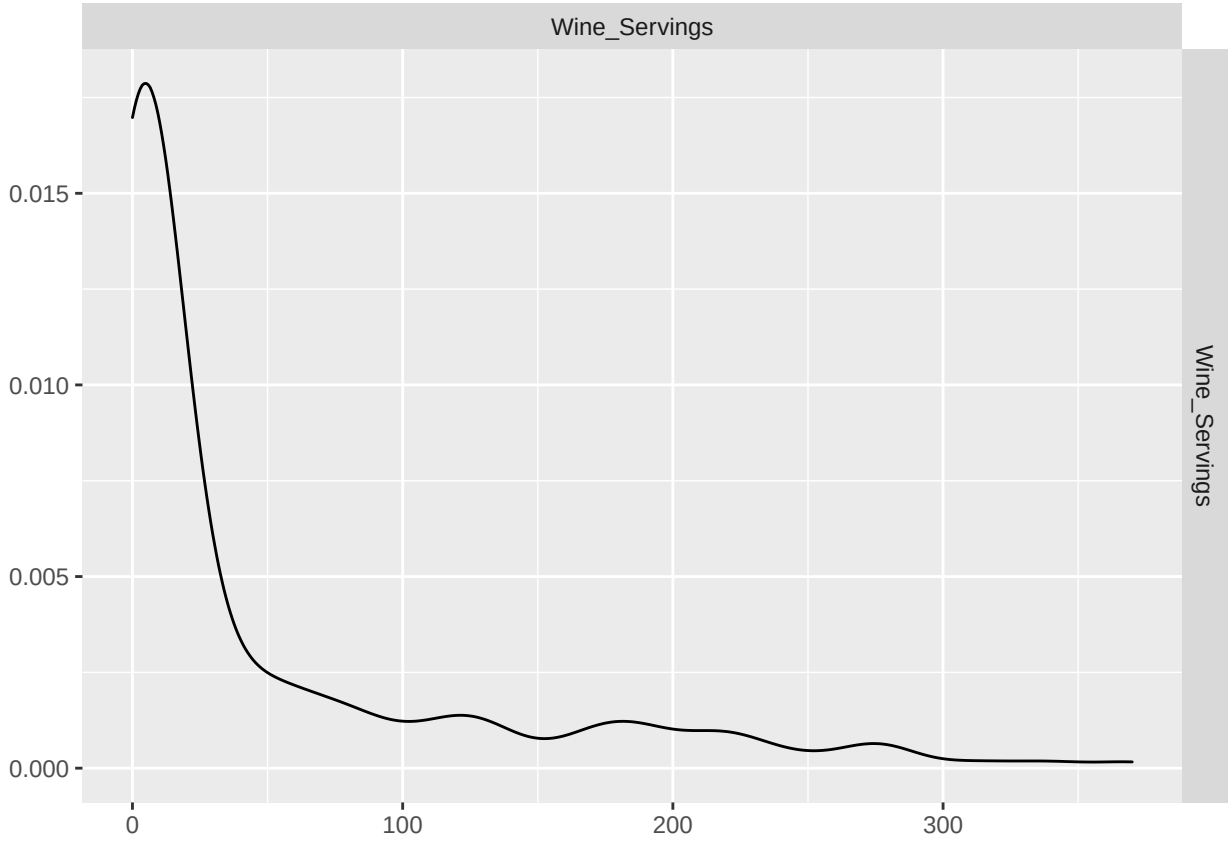
```r
# Beer Servings
g1 <- drinks %>%
  select(contains("Beer_Servings")) %>%
  ggpairs()
g1
```



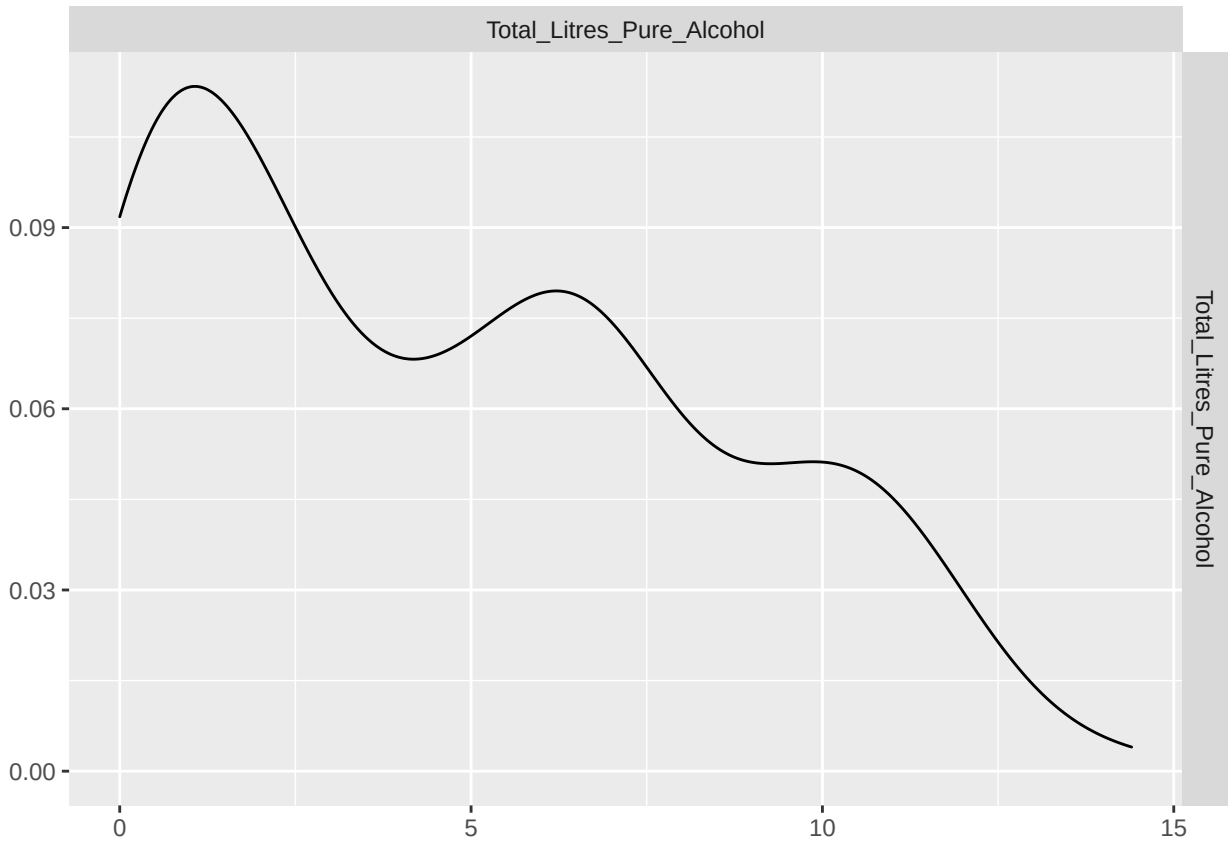

```r
# Spirit Servings
g2 <- drinks %>%
  select(contains("Spirit_Servings")) %>%
  ggpairs()
g2
```




```
# Wine Servings
g3 <- drinks %>%
  select(contains("Wine_Servings")) %>%
  ggpairs()
g3
```




```
# Total Litres Pure Alcohol
g4 <- drinks %>%
  select(contains("Total_Litres_Pure_Alcohol")) %>%
  ggpairs()
g4
```



Conclusion

H_0**:** There is no preference between alcohol types consumed per country.

H_A**:** There is a preference between alcohol types consumed per country.

After exploring this data we can safely reject the H_0. There is a preference between alcohol types consumed per country. The probability that a randomly chosen country has more than 100 servings of Beer is 52.43%, Spirit is 41.5% and Wine is 26.3% therefore, most countries in this data prefer Beer and Spirit over Wine. Yet, there are 13 countries that do no have alcohol consumption recorded due to strict alcohol consumption laws or religious believes. Some of these countries include Afghanistan, Bangladesh, North Korea, Iran and Kuwait, among others. The linear regression model showed us that there's a correlation between `Beer_Servings`, `Spirit_Servings`, and `Wine_Servings` with `Total_Litres_Pure_Alcohol`. I did notice that within each model the p-value was close to 0 and below the threshold of 0.05 and an indicator to reject the null hypothesis.

References

World Health Organization. (n.d.). Global information system on alcohol and health. World Health Organization. Retrieved October 19, 2021, from https://www.who.int/data/gho/data/themes/global-information-system-on-alcohol-and-health.

Ritchie, H., & Roser, M. (2018, April 16). Alcohol consumption. Our World in Data. Retrieved November 15, 2021, from https://ourworldindata.org/alcohol-consumption.

WHO. (2016). 19089 United States of america - world health organization. Retrieved December 5, 2021, from https://www.who.int/substance_abuse/publications/global_alcohol_report/profiles/usa.pdf.