Week 5 Visualizations Activity

This dataset shows the average serving sizes of different kinds of alcohol per person for each country in the world.

Read data:

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

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
```

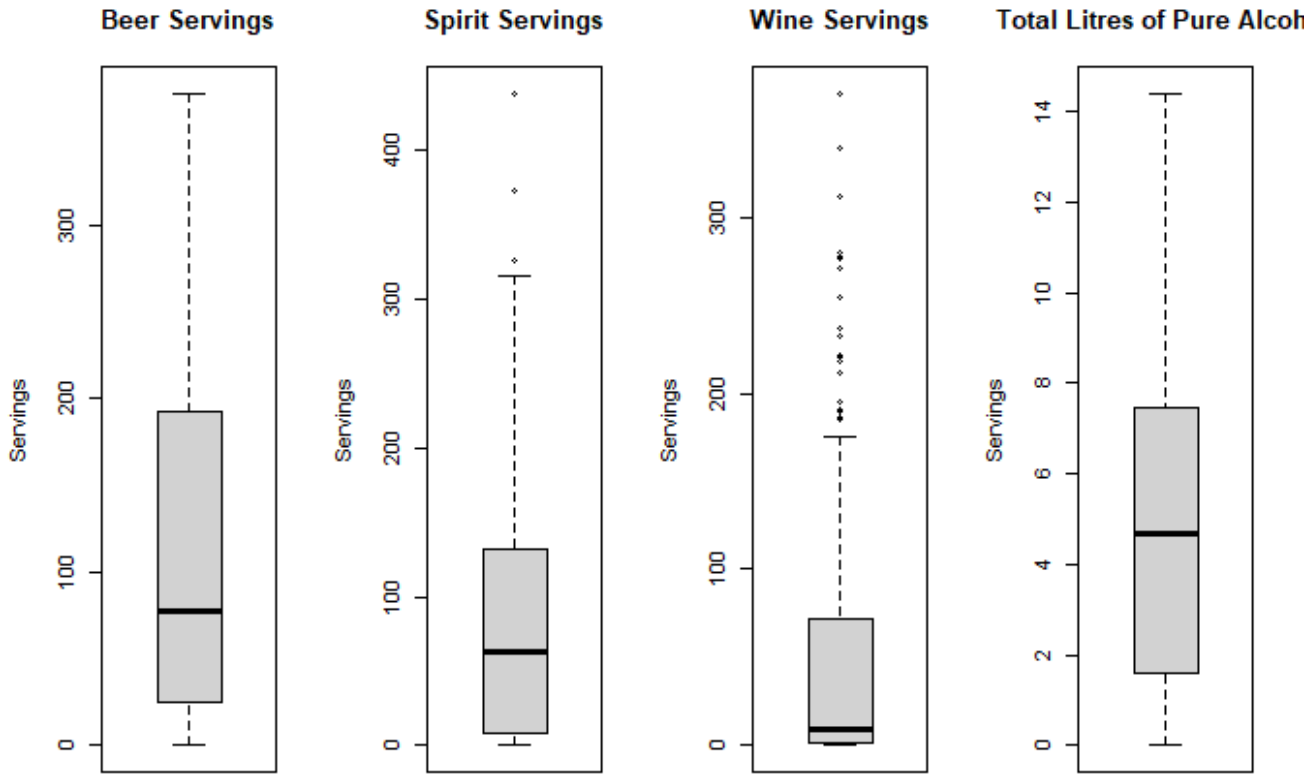
```
##  Class :character   1st Qu.: 25   1st Qu.:  8.00   1st Qu.:  1.50
##  Mode  :character   Median : 78   Median : 63.00   Median :  9.00
##                     Mean   :112   Mean   : 85.42   Mean   : 52.15
##                     3rd Qu.:193   3rd Qu.:132.50   3rd Qu.: 72.00
##                     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
```

Data Visualization

———Box Plot———-

```
library(ggplot2)
par(mfrow = c(1,4))

boxplot(drinks_legal$beer_servings, main = "Beer Servings", ylab = "Servings")
boxplot(drinks_legal$spirit_servings, main = "Spirit Servings", ylab = "Servings")
boxplot(drinks_legal$wine_servings, main = "Wine Servings", ylab = "Servings")
boxplot(drinks_legal$total_litres_of_pure_alcohol, main = "Total Litres of Pure Alcoh
```



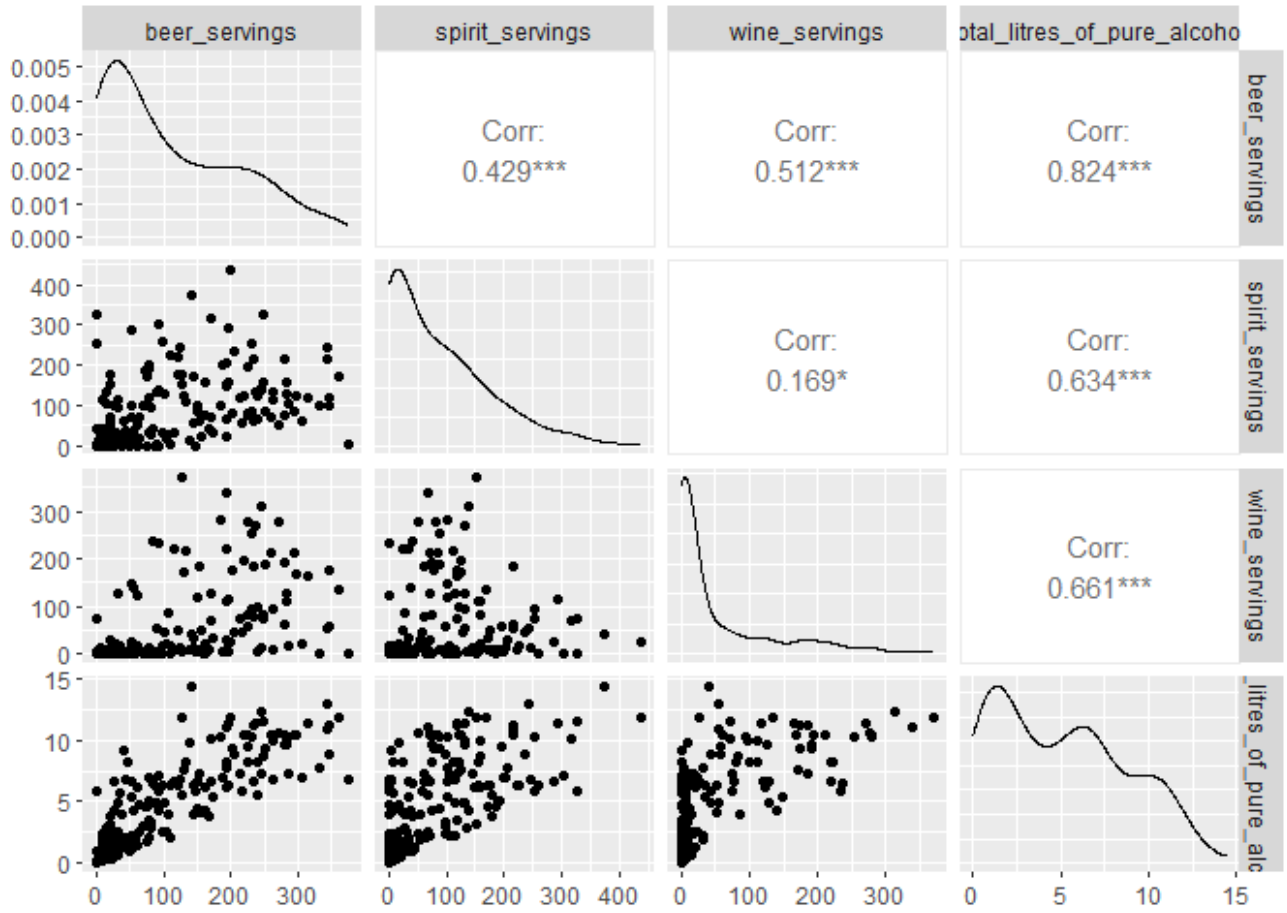
From the Box Plot, we can see that beer servings, spirit servings and wine servings all show right-skewed distribution, and wine servings has most outliers.

—Scatter Plot—

```
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
ggpairs(drinks_legal[, c("beer_servings", "spirit_servings", "wine_servings", "total_
```



From the scatter plot we can tell that, all fields are positive correlated.