**Introduction**

This report presents an analysis of alcohol consumption across various countries, utilizing data from the Gapminder\_Alcohol dataset. The analysis aims to explore the relationships between alcohol consumption and several socio-economic factors, including income per person, suicide rates, employment rates, and urbanization rates. The findings are visualized through interactive plots in a Shiny dashboard.

Data Exploration.

#Loaded necessary packages

library(tidyverse)

library(readxl)

library(dplyr)

library(tidyr)

library(ggplot2)

library(shiny) # For building dashboards

library(plotly)

corrplot: For visualizing correlation matrices.

Structure of the dataset

glimpse(gapminder\_alcohol\_data)

Rows: 213

Columns: 6

$ country *<chr>* "Afghanistan", "Albania", "Algeria", "An…

$ alcconsumption *<dbl>* 0.03, 7.29, 0.69, 10.17, 5.57, 8.17, 9.3…

$ incomeperperson *<dbl>* NA, 1914.9966, 2231.9933, 21943.3399, 13…

$ suicideper100th *<dbl>* 6.684385, 7.699330, 4.848770, 5.362179, …

$ employrate *<dbl>* 55.7, 51.4, 50.5, NA, 75.7, NA, 58.4, 40…

$ urbanrate *<dbl>* 24.04, 46.72, 65.22, 88.92, 56.70, 30.46…

Structure: Rows: 213 Columns: 6

variables:

country: The name of the country.

alcconsumption: Alcohol consumption per capita.

incomeperperson: Income per person.

suicideper100th: Suicide rate per 100,000 people.

employrate: Employment rate.

urbanrate: Urbanization rate.

**Socio-Economic Factors Influencing Alcohol Consumption**:

Independent Variables: incomeperperson, employrate, and urbanrate.

Dependent Variable: alcconsumption

Rationale: exploring how socio-economic factors affect alcohol consumption, which is a relevant public health issue.(Citation)

Data Exploratory Questions:

1. **What is the relationship between alcohol consumption and income per person?**/ **What is the average alcohol consumption across different income categories?**

This question aims to explore whether higher income correlates with higher alcohol consumption across different countries

1. **What is the distribution of alcohol consumption in urban vs. rural areas?**
   * This question explores whether urbanization impacts drinking habits.
2. **Is there a correlation between urbanization rate and alcohol consumption?**
   * whether countries with higher urbanization rates consume more or less alcohol compared to rural areas.
3. **What are the outliers in alcohol consumption, and what socio-economic characteristics do they have?**
   * Identify countries with exceptionally high or low alcohol consumption and examine their income, employment, and urbanization rates.
4. **Are there significant differences in alcohol consumption across different income categories?**
   * Compare average alcohol consumption among low, medium, and high-income countries to see how socio-economic status affects drinking behavior.

Additional insights, **Impact of Alcohol Consumption on Suicide Rates:**

Independent Variable: alcconsumption.

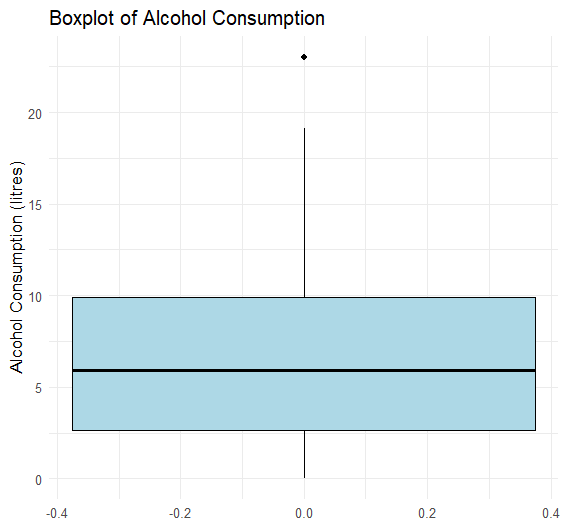
Dependent Variable: suicideper100th.

Rationale: This analysis can provide insights into the relationship between alcohol consumption and mental health outcomes, which is also crucial for public health discussions.(Citation)

Data Exploration questions

1. **What is the relationship between alcohol consumption and suicide rates?**
   * whether higher alcohol consumption correlates with higher suicide rates.
2. **Do countries with high alcohol consumption have significantly different suicide rates compared to those with low consumption?**
3. **How do average suicide rates differ between high and low alcohol-consuming countries?**
   * Compare suicide rates in countries that fall into high and low categories of alcohol consumption.

Identified ouliers



A tibble: 1 × 6

country alcconsumption incomeperperson suicideper100th

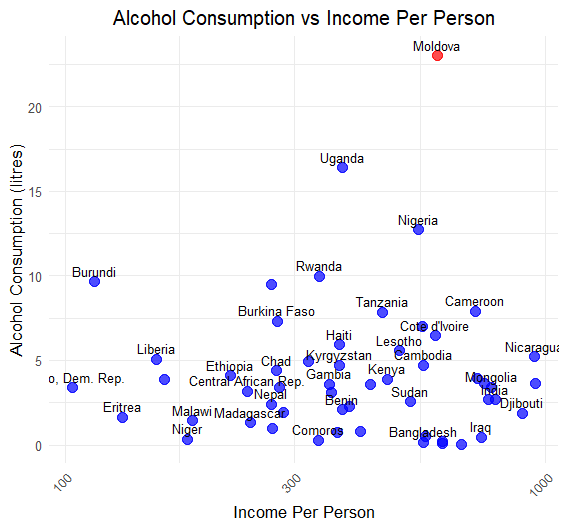
*<chr>* *<dbl>* *<dbl>* *<dbl>*

1 Moldova 23.0 596. 15.5

# ℹ 2 more variables: employrate <dbl>, urbanrate <dbl>

1. **Country**:
   * The output shows that the country is **Moldova**.
2. **Alcohol Consumption (alcconsumption)**:
   * Moldova has an alcohol consumption rate of **23.0 litres** per capita, which is significantly higher than the typical range observed in other countries. This suggests that Moldova is consuming a lot more alcohol compared to most other countries in your dataset.
3. **Income Per Person (incomeperperson)**:
   * The income per person in Moldova is **596**, which is relatively low compared to many other countries. This might indicate a potential relationship between lower income levels and higher alcohol consumption.
4. **Suicide Rate (suicideper100th)**:
   * The suicide rate in Moldova is **15.5 per 100,000 people**, which can be relevant when analyzing the impact of alcohol consumption on mental health outcomes.

Visualizing the outlier



Check for missing values:

print(missing\_values\_gapminder)

country alcconsumption incomeperperson suicideper100th

0 26 23 22

employrate urbanrate

35 10

How missing values were handled:

Removed rows with excessive missing values (4/5) columns

Imputation with median and mean

Result:

print(remaining\_missing\_values)

country alcconsumption incomeperperson suicideper100th

0 0 0 0

employrate urbanrate

0 0

Data Transformation:

A new column called IncomeCategory to our dataset. This categorization helps in analysing trends in alcohol consumption relative to income levels effectively. The following thresholds were used for categorization:

* **Low Income**: Countries with an income per person less than $1,000.
* **Medium Income**: Countries with an income per person between $1,000 and $10,000.
* **High Income**: Countries with an income per person greater than $10,000.

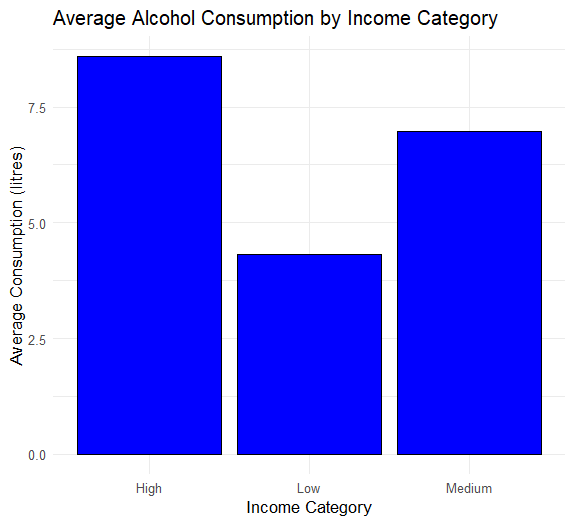
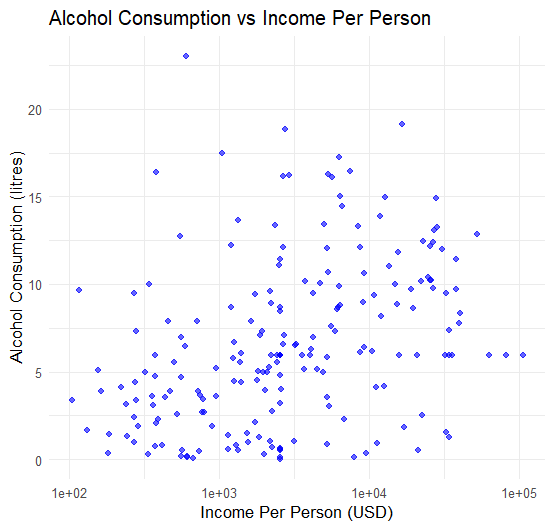
Here are a few entries from the transformed dataset:

|  |  |  |  |
| --- | --- | --- | --- |
| Afghanistan | 1381.004268 | High | 0 |
| **2** | Albania | 1381.004268 | High | 0 |
| **3** | Algeria | 1381.004268 | High | 0 |
| **4** | Andorra | 1381.004268 | High | 0 |
| **5** | Angola | 1381.004268 | High | 0 |
| **6** | Antigua and Barbuda | 1381.004268 | High | 0 |
| **7** | Afghanistan | 1914.996551 | High | 0 |
| **8** | Albania | 1914.996551 | High | 0 |
| **9** | Algeria | 1914.996551 | High | 0 |
| **10** | Andorra | 1914.996551 | High | 0 |

Part 2: Data visualizations

Line Plot Uing ggplo2 package, (There is no time series data so Not Applicable)

Bar Charts

High income, results in high alcohol consumption and low income in low alcohol consumption. There is a positive correlation between people’s level of income and alcohol consumption.

A graph of a graph with purple rectangles

Description automatically generated A graph of alcohol consumption

Description automatically generated

The urbanization rate of counties is high when income levels are high and significantly low when people’s incomes are low. There is a positive correlation between Alcohol consumption and urbanization rate.

A graph of suicide rate by income category

Description automatically generatedA graph with red dots

Description automatically generated

Low income, can possible result in high suicide rates as compared to high and medium income.

Histogram (showing the distribution of Alcohol consumption)

A graph of a distribution of alcohol consumption

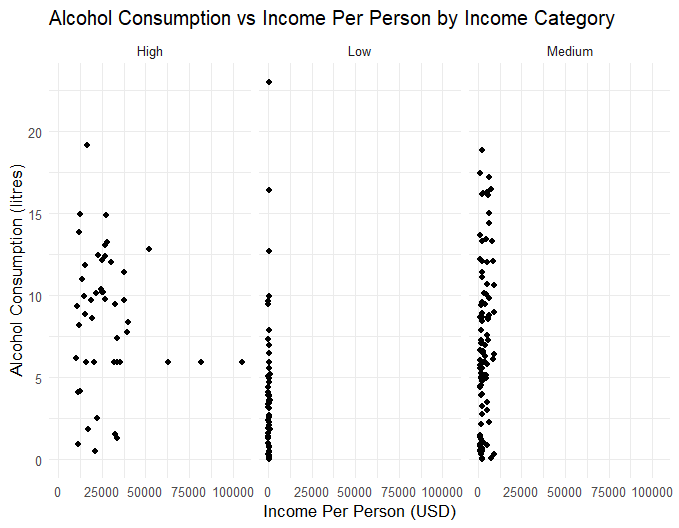
Description automatically generated

The data is skewed to the right.

Demonstrating Use of facet.

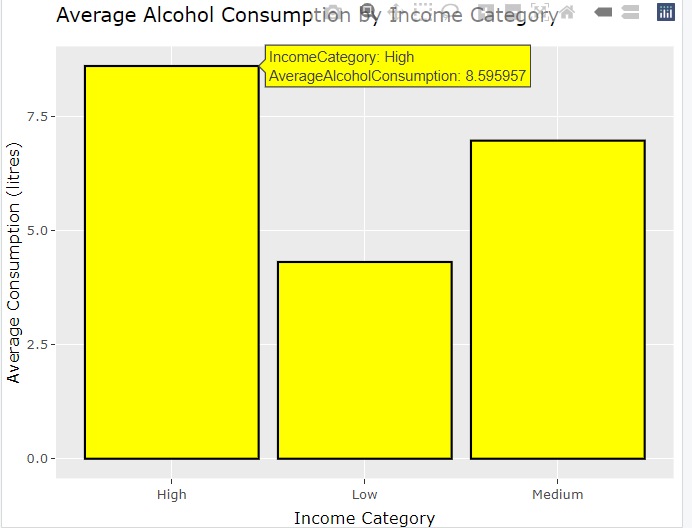
|  |
| --- |
| ggplot(data=gapminder\_cleaned, aes(x=incomeperperson, y=alcconsumption)) +  + geom\_point() +  + facet\_wrap(~ IncomeCategory) +  + labs(title="Alcohol Consumption vs Income Per Person by Income Category",  + x="Income Per Person (USD)", y="Alcohol Consumption (litres)") +  + theme\_minimal() |
|  |
| |  | | --- | |  | |

Output:

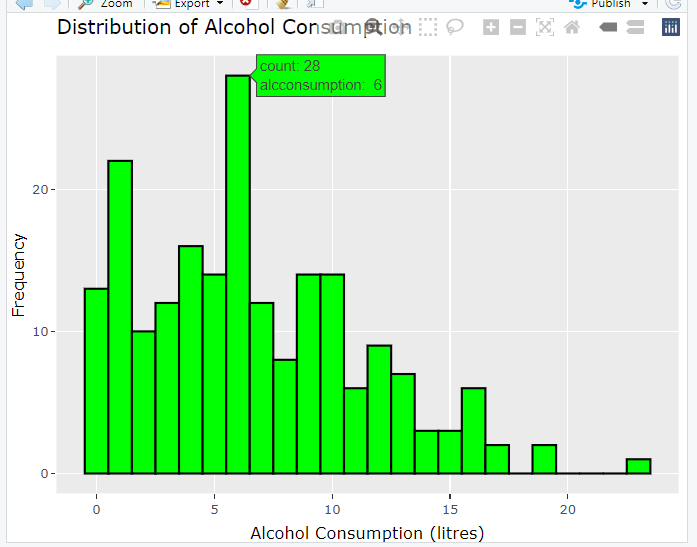


Interactive visualizations..

On hover,



Histogram with hover



It shows that there are 28 observations (countries) with alcohol consumption of 6 approximately litres

To verify:

countries\_with\_6\_litres <- gapminder\_cleaned %>%

+ filter(alcconsumption >= 5.5 & alcconsumption <= 6.5)

>

> print(countries\_with\_6\_litres)

Output:

A tibble: 28 × 7

country alcconsumption incomeperperson suicideper100th

*<chr>* *<dbl>* *<dbl>* *<dbl>*

1 Angola 5.57 1381. 14.6

2 Barbados 6.42 9244. 3.11

3 Belize 5.92 3546. 9.39

4 Bermuda 5.92 62682. 9.64

5 Bolivia 5.78 1233. 2.03

6 China 5.56 2425. 16.9

7 Costa Rica 5.81 5189. 6.60

8 Cote d'Iv… 6.47 591. 20.3

9 Dominican… 6.28 4049. 6.52

10 Equatoria… 6.12 8655. 10.1

# ℹ 18 more rows

Interactive facet (with hover)