Introduction

The `Global_black_Money` dataset provides a comprehensive overview of illegal financial transactions taking place across various countries. It offers valuable insights into the patterns and trends of money movements associated with unlawful activities, such as tax evasion, money laundering, or corruption. This dataset includes detailed attributes like the transaction amounts, the types of transactions (e.g., cash withdrawals, cryptocurrency, or property purchases), the industries involved, and risk scores that highlight the likelihood of suspicious or high-risk activities.

By analyzing this data, we can gain a better understanding of how financial crimes are carried out, which countries or industries are most affected, and the scale of these activities. The dataset serves as a powerful resource for identifying irregularities and helping experts, policymakers, and law enforcement agencies tackle financial crimes more effectively.

Overview of the Dataset

There are 10000 rows and 15 columns in the data set. The Columns with their descriptions are as follows:

`Country` : Country where the transaction occurred.

`Transaction ID` : Unique identifier for each transaction

`Amount`: Transaction amount in US dollar

`Transaction Type`: Type of transaction

`Date of Transaction`: The date and time of the transaction

`Person Involved` : Name or identifier of the person/ entity involved

`Industry`:Industry associated with the transaction

`Destination Country` :Country where the money was sent

`Reported by Authority` : Whether the transaction was reported to authorities.

`Source of Money` : Origin of the money

`Money laundering Risk Score`: Risk score indicating the likelihood of money laundering

`Shell Companies Involved`: Number of shell companies used in the transaction

`Financial Institution`: Bank or Financial institution involved in the transaction

`Tax haven Country`:Country where the money was transferred to a tax haven.

The Countries where the transaction were occured

```
unique(Global_black_Money$Country)
```

```
##  [1] "Brazil"       "China"        "UK"           "UAE"          "South Africa"
##  [6] "Russia"       "Switzerland"  "India"        "USA"          "Singapore"
```

These are the unique Countries in the dataset where transaction were occured.

Type of Transactions

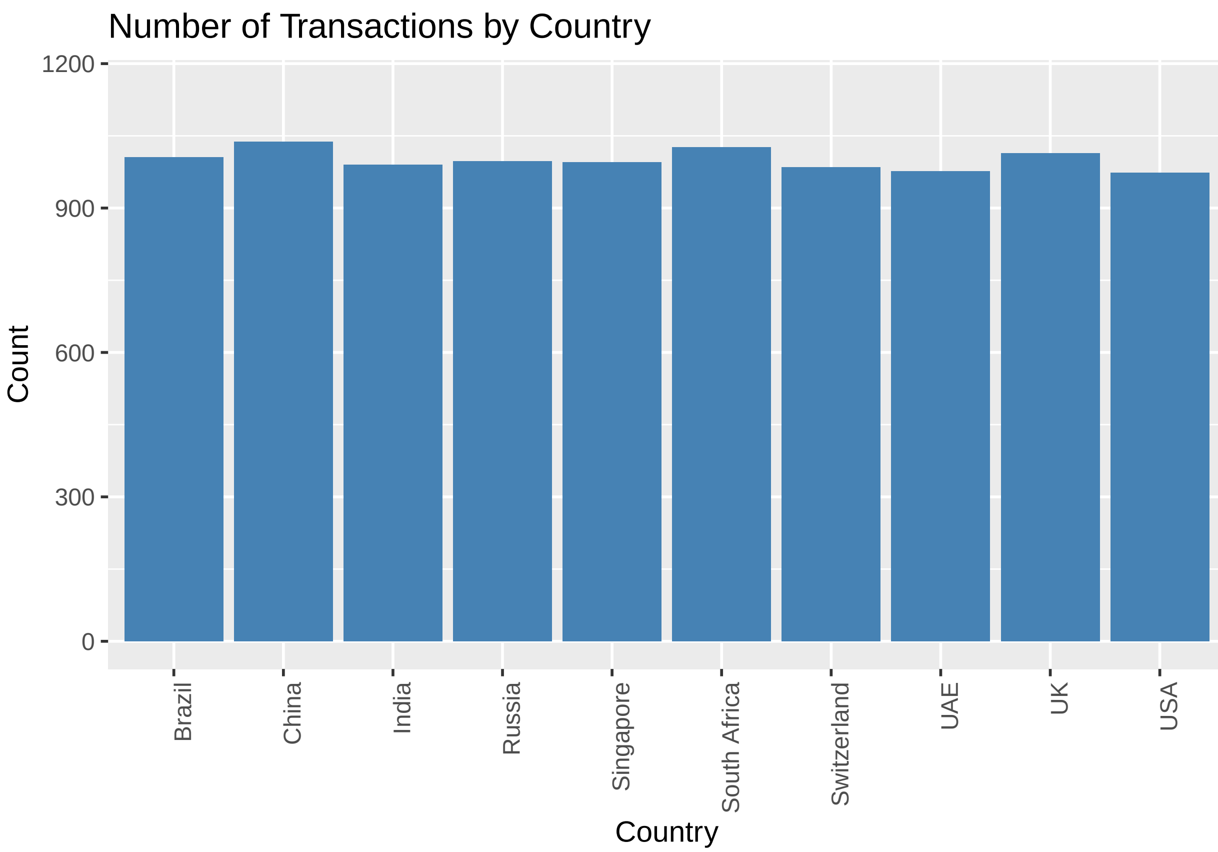
```
unique(Global_black_Money$`Transaction.Type`)
```

```
## [1] "Offshore Transfer" "Stocks Transfer"   "Cash Withdrawal"
## [4] "Cryptocurrency"    "Property Purchase"
```

Analysis

Here is the visualization of Number of Transactions by Country

```
ggplot(Global_black_Money, aes(x = Country)) +
  geom_bar(fill = "steelblue") + ylim(0,1150)+
  labs(title = "Number of Transactions by Country", x = "Country", y = "Count") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



This bar chart illustrates the number of black money transactions across various countries. Each bar corresponds to a specific country, and the height of the bar reflects the transaction count. The countries represented include Brazil, China, India, Russia, Singapore, South Africa, Switzerland, UAE, the UK, and the USA.

From the chart, it's evident that the number of transactions is relatively similar across all the countries, ranging from about 900 to just over 1,000 transactions. China appears to have the highest number of transactions, slightly exceeding 1,000, while UAE & USA has the lowest, with just under 950 transactions.

This uniformity in transaction counts highlights that black money activities are prevalent and consistently distributed across these countries, suggesting widespread financial irregularities on a global scale. Such data can provide valuable insights for identifying patterns and understanding the scale of illegal financial activities worldwide.

The maximum & Minimum transaction amount in the data set

```
max(Global_black_Money$Amount..USD.)
```

```
## [1] 4999812
```

```r
min(Global_black_Money$Amount..USD.)
```

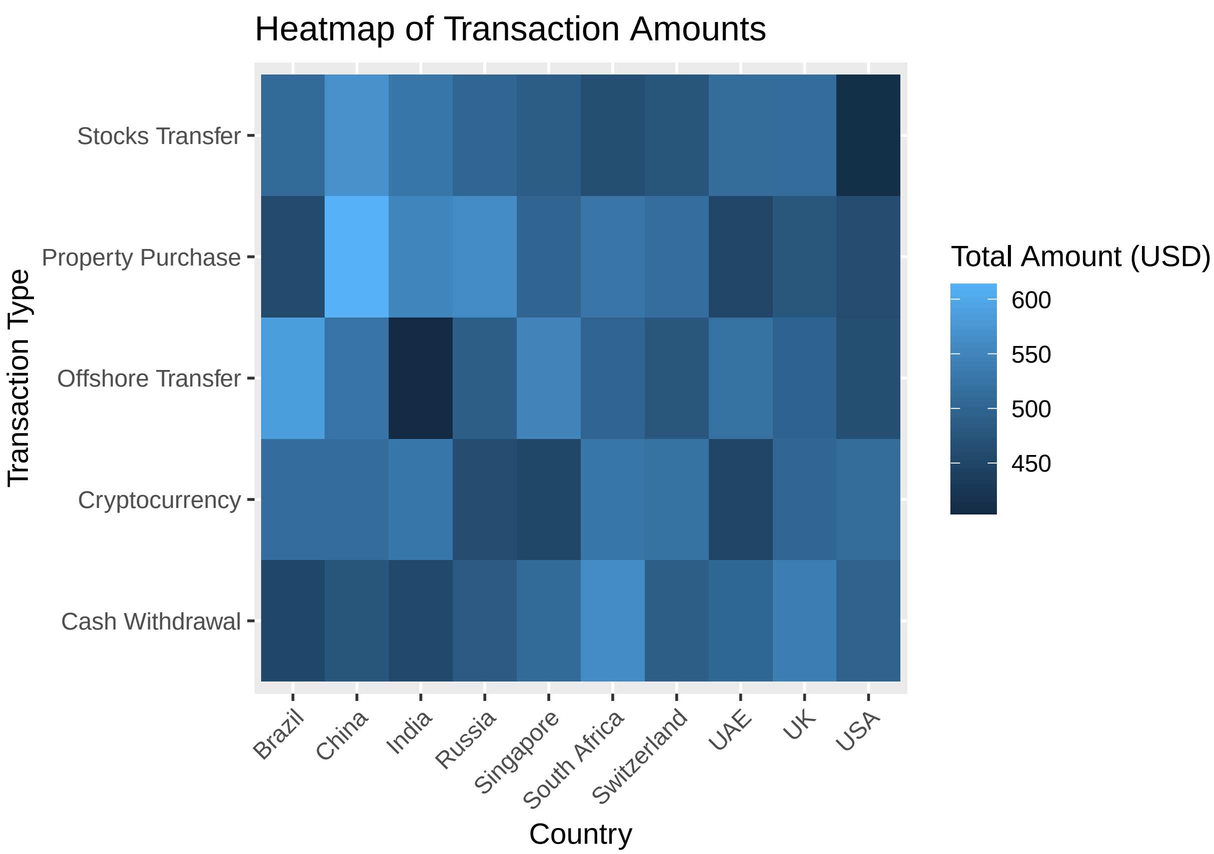
```
## [1] 10031.8
```

The total amount of Transaction(in million)

```r
sum(Global_black_Money$Amount)
```

```
## [1] 25018.18
```

```r
Global_black_Money %>%
  group_by(Country, Transaction.Type) %>%
  summarise(Total_Amount = sum(Amount, na.rm = TRUE)) %>%
  ggplot(aes(x = Country, y = Transaction.Type, fill = Total_Amount)) +
  geom_tile()+
  labs(title = "Heatmap of Transaction Amounts", x = "Country", y = "Transaction Type", fill = "Total Ar
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



This heatmap shows the total amounts of money spent on different types of transactions in various countries. The x-axis lists the countries, including Brazil, China, India, Russia, Singapore, South Africa, Switzerland, UAE, UK, and the USA. The y-axis shows the types of transactions, such as Stocks Transfers, Property Purchases, Offshore Transfers, Cryptocurrency, and Cash Withdrawals. The colors represent how much money was spent, with lighter blue meaning more money and darker blue meaning less money.

From the heatmap, we can see that Property Purchases have the highest amounts in some countries, especially in the China, which is shown by the lighter shade of blue. Offshore Transfers also show high amounts in many countries, but India has a darker patch, meaning less activity in this type of transaction. Cryptocurrency transactions look similar across all countries, without big differences. Stocks Transfers and Cash Withdrawals also show fairly equal amounts in all countries.

This heatmap gives us some important insights. Property Purchases seem to involve a lot of money, especially in certain countries, likely because of real estate or investment. Offshore Transfers are also important in many countries, possibly for financial planning or tax reasons. Cryptocurrency transactions are steady and similar worldwide, showing its global usage. Some darker areas, like Offshore Transfers in India, suggest lower activity in those areas. Overall, this chart helps us compare money activity by country and transaction type in an easy-to-see way.

The most commonly used Tax Havens in these illegal financial activities

```r
unique(Global_black_Money$Tax.Haven.Country)
```

```
## [1] "Singapore"      "Bahamas"        "Switzerland"    "Panama"
## [5] "Luxembourg"     "Cayman Islands"
```

```r
#  Group the data by 'Tax.Haven.Country'
grouped_data <- group_by(Global_black_Money, Tax.Haven.Country)

# Summarise the grouped data to count the occurrences of each country
summarised_data <- summarise(grouped_data, count = n())

# Finding the maximum count from the summarised data
max_count <- max(summarised_data$count)
max_count
```

```
## [1] 1743
```

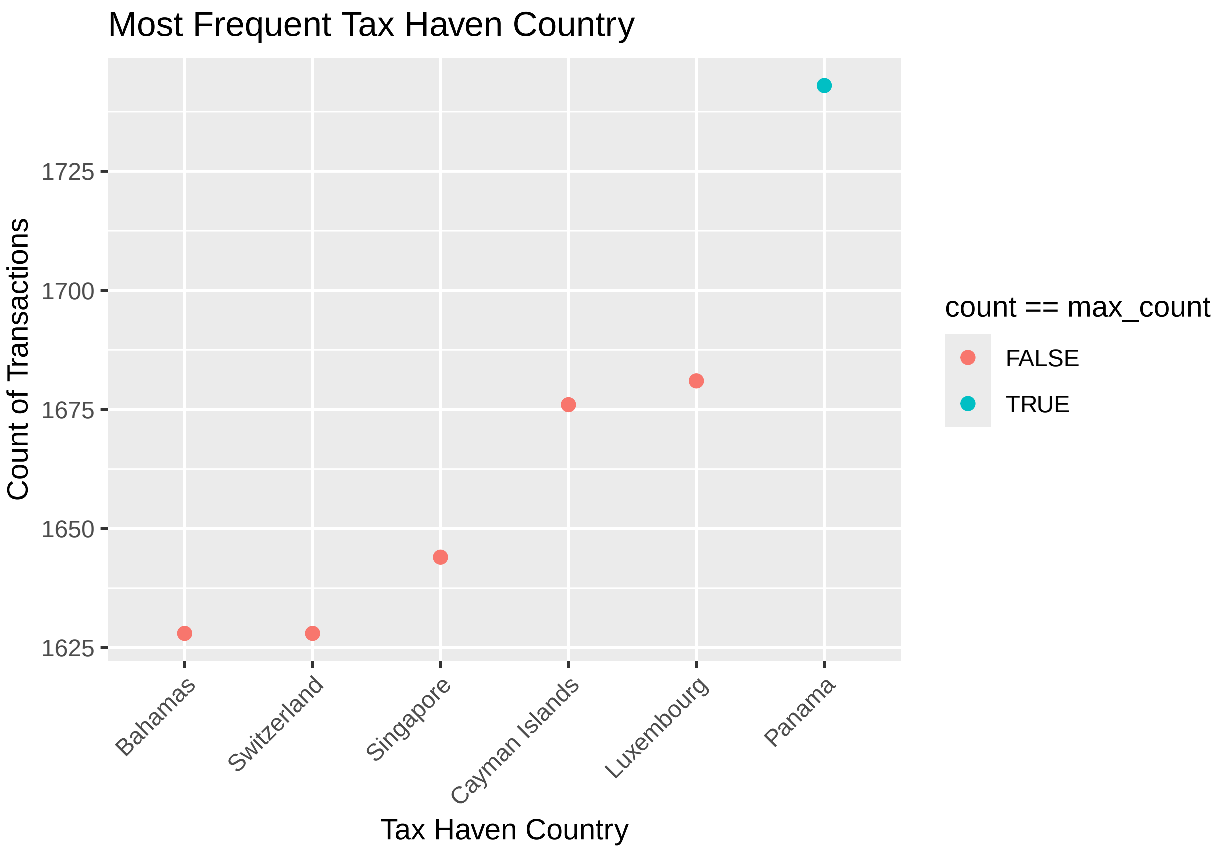
```r
# Filter the rows where the count is equal to the maximum count
most_frequent_data <- filter(summarised_data, count == max_count)

# Extracting the most frequent tax haven country
most_frequent_tax_haven_country <- pull(most_frequent_data, Tax.Haven.Country)

# Printing the result
print(most_frequent_tax_haven_country)
```

```
## [1] "Panama"
```

```r
ggplot(summarised_data, aes(x = reorder(Tax.Haven.Country, count), y = count, color = count == max_coun
  geom_point(size = 2) +
  labs(
    title = "Most Frequent Tax Haven Country",
    x = "Tax Haven Country",
    y = "Count of Transactions"
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



This is a dot plot that displays the frequency of transactions in various tax haven countries. The x-axis represents the countries, which include the Bahamas, Switzerland, Singapore, the Cayman Islands, Luxembourg, and Panama. The y-axis indicates the count of transactions for each country, allowing for a clear comparison of transaction volumes across these locations. This visualization emphasizes the most frequent tax haven country by marking it distinctly in blue. The remaining countries, shown in red, have relatively close transaction counts but fall below the maximum. The use of a dot plot makes it easy to compare and identify the country with the highest frequency while providing an overall sense of the distribution of transaction counts among the other tax havens.

Comparison between Finance & Real Estate Industry in Black Money Transaction

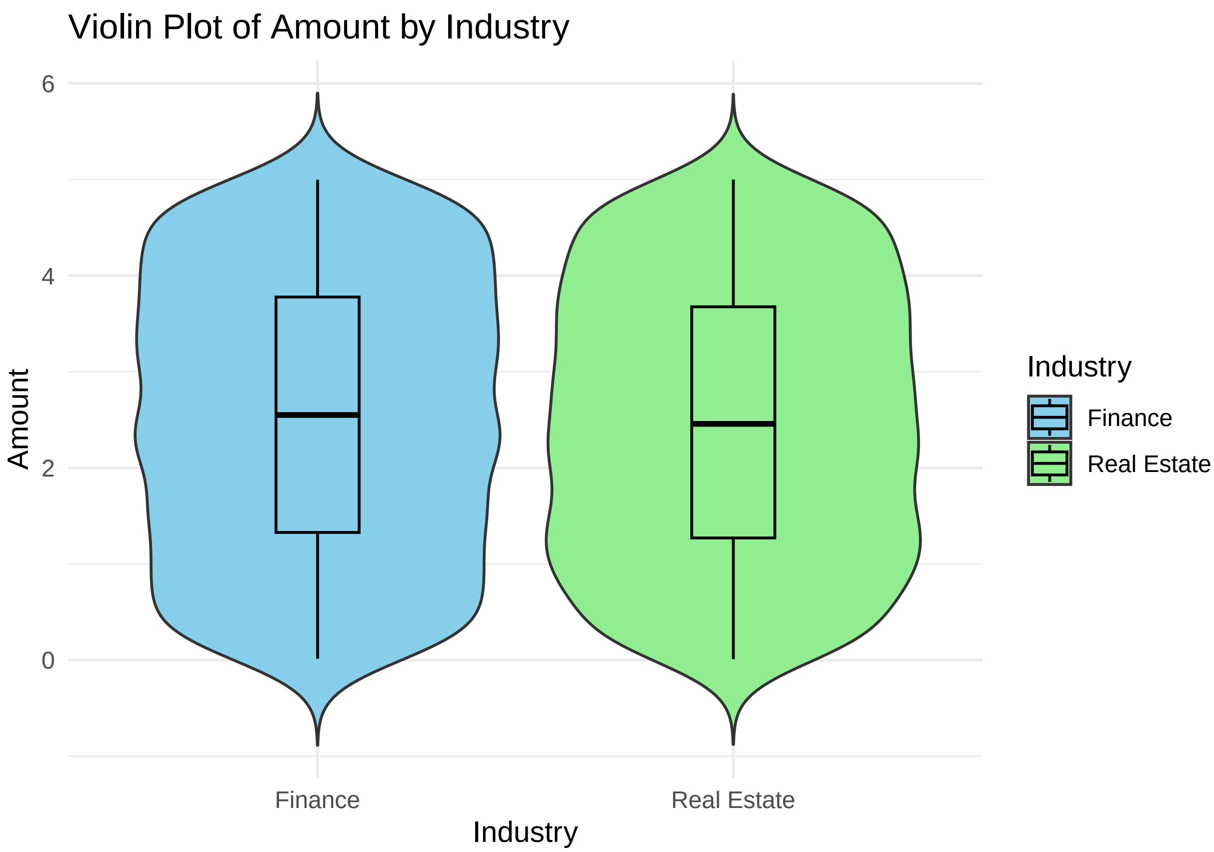
```
finance_real_estate <- Global_black_Money %>%
  filter(Industry== "Finance"|Industry=="Real Estate")

# Two-sample t-test: Compare means between Finance and Real Estate industries
t_test_two_sample <- t.test(Amount ~ Industry, data = finance_real_estate)
print(t_test_two_sample)
```

```
##
##  Welch Two Sample t-test
##
## data:  Amount by Industry
## t = 1.04, df = 2916, p-value = 0.2984
## alternative hypothesis: true difference in means between group Finance and group Real Estate is not e
## 95 percent confidence interval:
##  -0.04842925  0.15783496
## sample estimates:
```

```
##     mean in group Finance mean in group Real Estate
##                  2.532829                  2.478126
```

```r
# Create a violin plot
ggplot(finance_real_estate, aes(x = Industry, y = Amount, fill = Industry)) +
  geom_violin(trim = FALSE) +
  geom_boxplot(width = 0.2, color = "black", alpha = 0.7) +
  labs(title = "Violin Plot of Amount by Industry",
       x = "Industry",
       y = "Amount") +
  theme_minimal() +
  scale_fill_manual(values = c("skyblue", "lightgreen"))
```



The goal was to compare the mean Amount between two industries: Finance and Real Estate, to determine if there is a significant difference in their averages. A Welch Two Sample t-test was conducted to compare the means of Amount.A violin plot was used to visualize the distribution of Amount for both industries.Both industries showed similar distributions with overlapping densities.This supports the conclusion that there is no major difference between the groups.The medians for Finance and Real Estate are nearly identical.

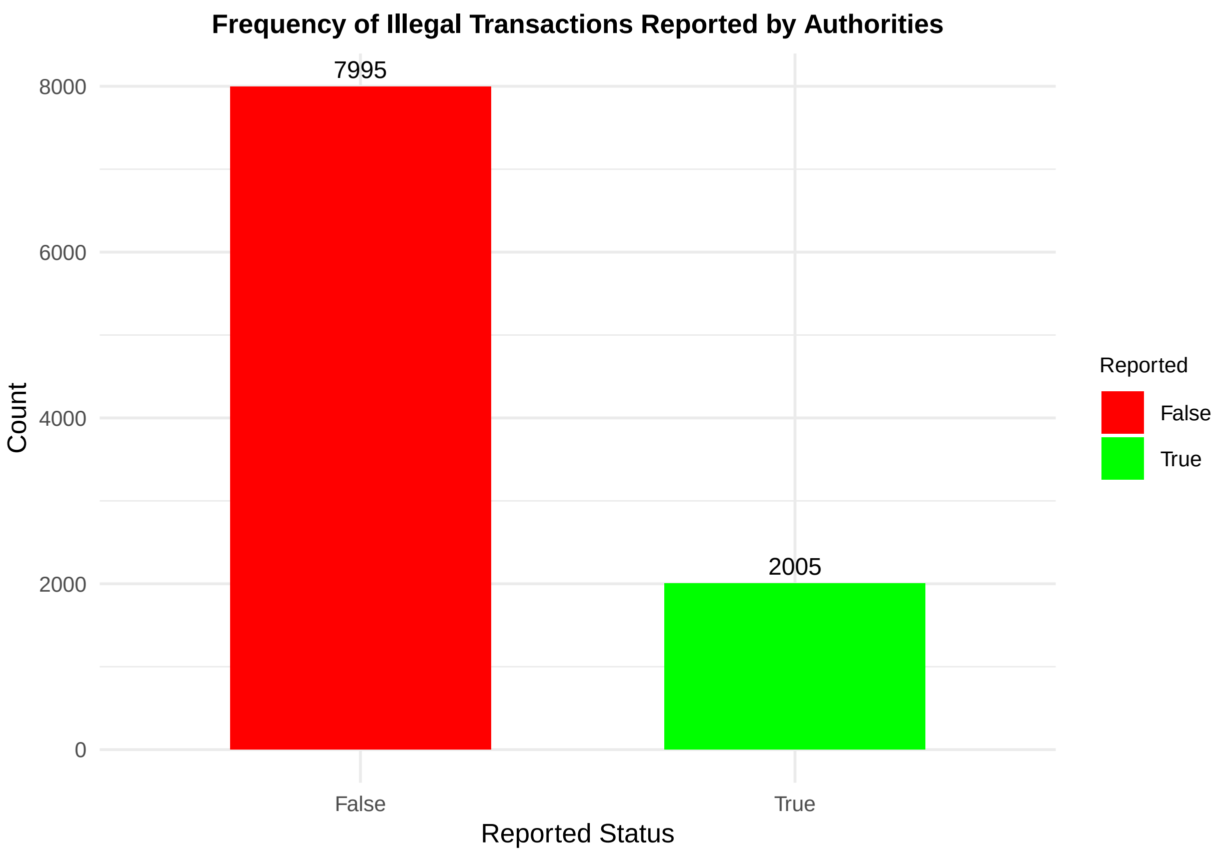
Reporting Summary

```r
reported_counts <- table(Global_black_Money$`Reported.by.Authority`)
print(reported_counts)
```

```
##
## False  True
##  7995  2005
```

```r
# Converting the table to a data frame for plotting
reported_df <- as.data.frame(reported_counts)
colnames(reported_df) <- c("Reported", "Count")

# Create a bar plot
ggplot(reported_df, aes(x = Reported, y = Count, fill = Reported)) +
  geom_bar(stat = "identity", width = 0.6) +
  geom_text(aes(label = Count), vjust = -0.5, size = 3) +
  scale_fill_manual(values = c("False" = "red", "True" = "green")) +
  labs(
    title = "Frequency of Illegal Transactions Reported by Authorities",
    x = "Reported Status",
    y = "Count",
    fill = "Reported"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 10, face = "bold", hjust = 0.5),
    axis.text = element_text(size = 8),
    axis.title = element_text(size = 10),
    legend.title = element_text(size = 8),
    legend.text = element_text(size = 8)
  )
```



The bar plot illustrates the frequency of illegal transactions reported by authorities in the dataset. It represents two categories:

"False": Transactions not reported by authorities.

"True": Transactions reported by authorities

The dataset shows a noticeable discrepancy between the reported and unreported illegal transactions. Approximately 80% of transactions were not reported, which could suggest potential inefficiencies in detection or reporting mechanisms.

Correlation between Transaction Amount and Risk Score

```r
# Correlation coefficient
correlation_Amount_risk<-cor(Global_black_Money$`Money.Laundering.Risk.Score`,Global_black_Money$Amount
print(paste("Correlation between Transaction Amount and Risk Score: ", correlation_Amount_risk))
```

```
## [1] "Correlation between Transaction Amount and Risk Score:  0.0165285517397385"
```

The Correlation between Transaction Amount and Risk Score is 0.0165 which indicates a very weak positive relationship between Transaction Amount and Risk Score. In practical implication,Transaction amount alone is not a significant determinant of the money laundering risk score. It's likely that other variables or a combination of factors contribute more substantially.

Risk Score by Country

```r
risk_score<-Global_black_Money %>%
group_by(Country)%>%
summarise(Median_Risk_Score = median(Money.Laundering.Risk.Score, na.rm = TRUE))
print(risk_score)
```

```
## # A tibble: 10 x 2
##    Country      Median_Risk_Score
##    <chr>                    <dbl>
##  1 Brazil                       6
##  2 China                        6
##  3 India                        6
##  4 Russia                       5
##  5 Singapore                    6
##  6 South Africa                 6
##  7 Switzerland                  5
##  8 UAE                          5
##  9 UK                           5
## 10 USA                          5
```

The median score represents the "typical" money laundering risk for transactions in each country, unaffected by extreme outliers. Countries like Brazil, China, India, Singapore, and South Africa have a median risk score of 6, indicating a relatively higher typical risk. Countries like Russia, Switzerland, UAE, UK, and USA have a median risk score of 5, reflecting a slightly lower typical risk.Countries with a higher median risk score may need stricter regulatory frameworks to mitigate money laundering risks.Brazil, China, India, Singapore, and South Africa could be prioritized for interventions.

Conclusion

The dataset, Global_black_Money, provides a comprehensive overview of illegal financial activities occurring across various countries. It contains 10,000 rows and 15 columns with critical details about black money transactions, including Country, Transaction Amount, Transaction Type, Risk Score.Countries represented in the dataset include Brazil, China, the UK, UAE, South Africa, Russia, Switzerland, India, USA, and Singapore.The number of black money transactions is consistently distributed across countries. While

China has the highest transaction count, countries like UAE and USA have fewer transactions. This suggests a global prevalence of illicit financial activities, with no country being completely immune.The most frequent transaction types include Offshore Transfers, Cryptocurrency, and Property Purchases, with Property Purchases showing higher amounts in countries like China.The dataset reveals that Property Purchases and Offshore Transfers tend to involve higher amounts, particularly in countries like China and Brazil. This could indicate that these sectors are more heavily used for laundering money.

The Money Laundering Risk Score correlates weakly with the Transaction Amount, suggesting that other factors (e.g., transaction types or industries) might play a more significant role in determining the risk.Panama, Switzerland, and Singapore are the most frequent tax haven countries involved in these transactions. Panama stands out with the highest number of transactions linked to tax havens. A notable 80% of transactions were not reported by authorities, suggesting that many illicit activities go undetected, which could hinder efforts to mitigate financial crimes globally.A comparison of the Finance and Real Estate industries in terms of transaction amounts revealed no significant difference, with similar risk scores and transaction behaviors between the two sectors. The widespread nature of black money activities across countries calls for greater international collaboration to improve detection and reporting mechanisms.This dataset serves as a valuable tool in identifying financial irregularities and developing strategies to combat illicit financial flows.