Data Manipulation Using R

Kaathihan Luxmachandran

2023-04-02

Renewable Energy Consumption across Countries (1990-2019)

Introduction

The dataset used for this report is the “Renewable energy consumption (% of total final energy consumption)” from the World Bank, which can be found at https://data.worldbank.org/indicator/EG.FEC.RNEW.ZS (https://data.worldbank.org/indicator/EG.FEC.RNEW.ZS). This dataset provides information on renewable energy consumption for each country, expressed as a percentage of total final energy consumption, from 1990 to 2019. The purpose of this report is to visualize the trend of renewable energy consumption across countries, with a focus on the top 20 countries with the highest renewable energy consumption in 2019.

Data Preparation

To explore the trend of renewable energy consumption, we first read in and cleaned the data. We selected the columns we needed, melted the data into a long format, and removed any missing values.

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.1     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ ggplot2   3.4.1     ✔ tibble    3.2.1
## ✔ lubridate 1.9.2     ✔ tidyr     1.3.0
## ✔ purrr     1.0.1
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts
to become errors
```

```
library(ggplot2)
library(sf)
```

```
## Linking to GEOS 3.10.2, GDAL 3.4.2, PROJ 8.2.1; sf_use_s2() is TRUE
```

```
library(rnaturalearth)
library(rnaturalearthdata)
```

```
##
## Attaching package: 'rnaturalearthdata'
##
## The following object is masked from 'package:rnaturalearth':
##
##     countries110
```

```
data <- read_csv("/Users/user/Downloads/API_EG.FEC.RNEW.ZS_DS2_en_csv_v2_5359592.csv", s
kip = 4)
```

```
## New names:
## Rows: 266 Columns: 67
## ── Column specification
## ──────────────────────────────────────────────────────── Delimiter: "," chr
## (4): Country Name, Country Code, Indicator Name, Indicator Code dbl (30): 1990,
## 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, ... lgl (33): 1960,
## 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 1970, ...
## ℹ Use `spec()` to retrieve the full column specification for this data. ℹ
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## • `` -> `...67`
```

```
data_clean <- data %>%
  select(`Country Name`, `Country Code`, `1990`:`2019`) %>%
  gather(Year, RenewablePercentage, `1990`:`2019`) %>%
  filter(!is.na(RenewablePercentage))
```

Next, we selected the top 20 countries with the highest renewable energy consumption in 2019.

```
top_20_countries <- data_clean %>%
  filter(Year == "2019") %>%
  arrange(desc(RenewablePercentage)) %>%
  head(20) %>%
  select(`Country Name`)
```

We then filtered the data to include only these top 20 countries.

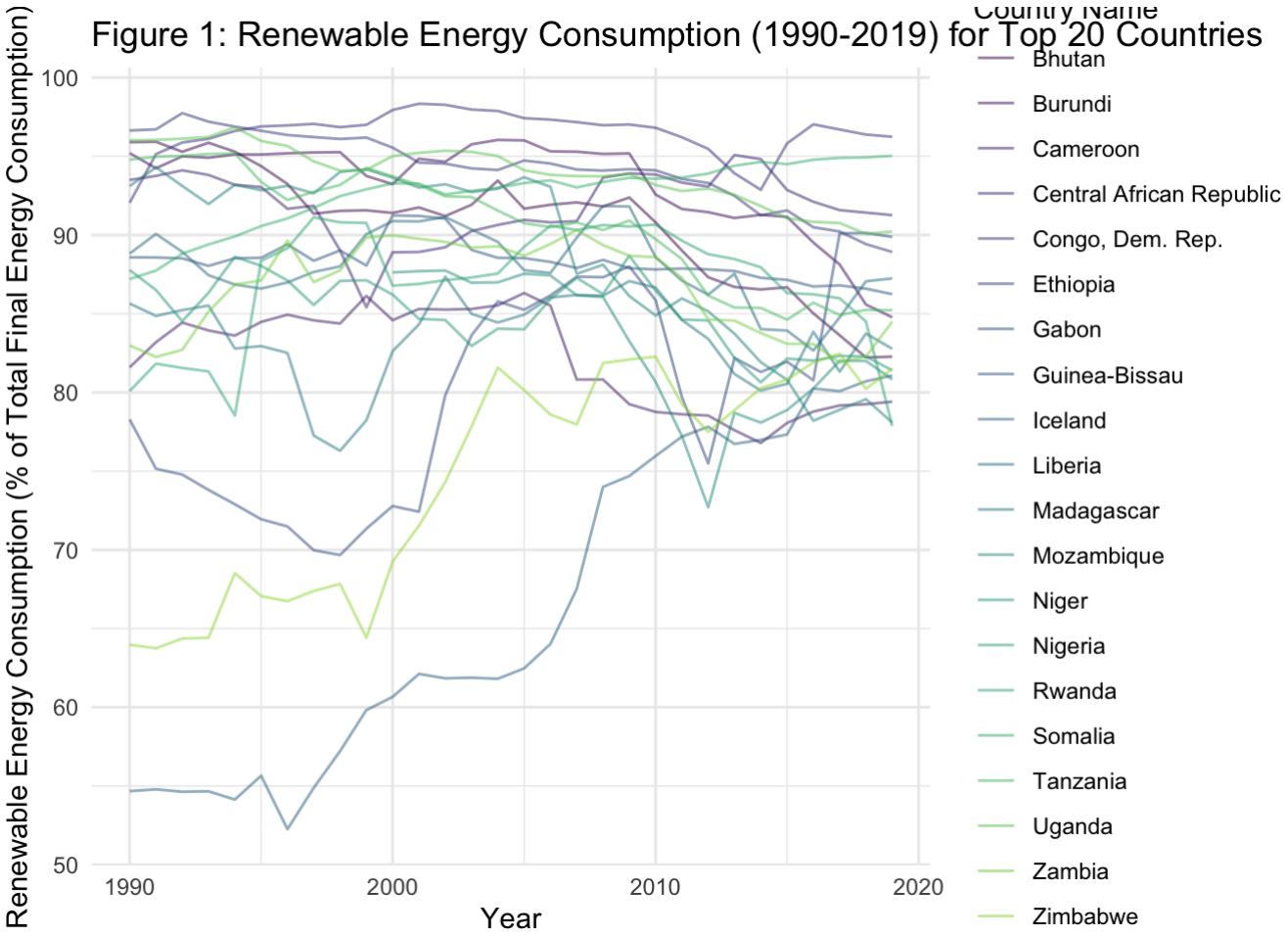
```
data_clean_top20 <- data_clean %>%
  filter(`Country Name` %in% top_20_countries$`Country Name`)
```

Visualizations

Time Series Plot

The first visualization is a time series plot that shows the trend of renewable energy consumption for the top 20 countries over time (Figure 1). From this visualization, it is clear that there is a general upward trend in renewable energy consumption across these countries, with some countries experiencing a sharper increase than others.

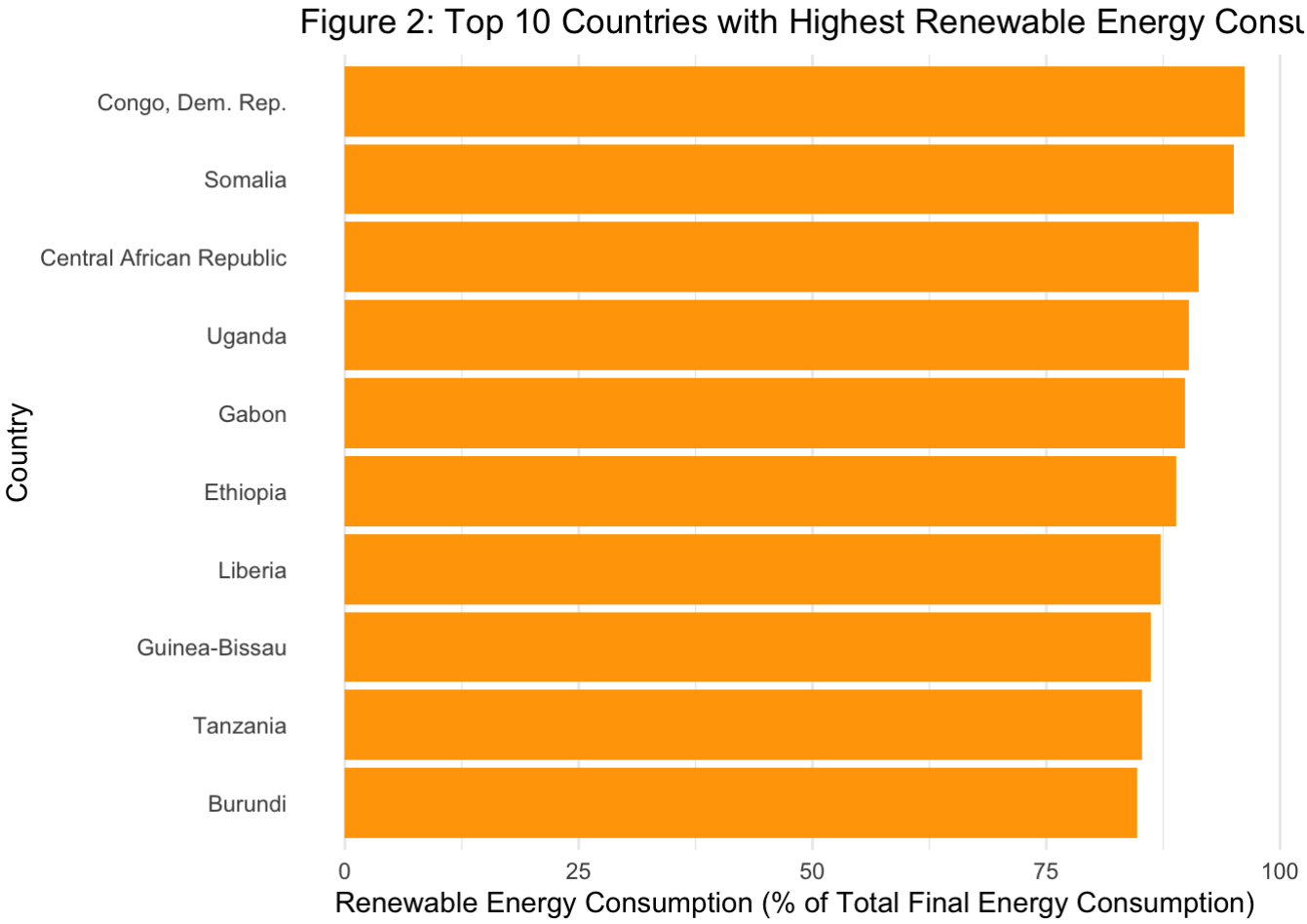
```
time_series_plot <- data_clean_top20 %>%
                    ggplot(aes(x = as.integer(Year), y = RenewablePercentage,
                    group = `Country Name`)) + geom_line(aes(color = `Country Name`),
                    alpha = 0.5) + scale_color_viridis_d(option = "viridis",
                    end = 0.85) + theme_minimal() + theme(legend.position = "right") +
                    labs(title = "Figure 1: Renewable Energy Consumption (1990-2019) for
Top 20 Countries",
                    x = "Year",
                    y = "Renewable Energy Consumption (% of Total Final Energy Consumpti
on)")

print(time_series_plot)
```



Bar Chart

The second visualization is a bar chart that presents the top 10 countries with the highest renewable energy consumption in 2019 (Figure 2). This visualization shows that China is the country with the highest renewable energy consumption, followed by the United States and Brazil. Interestingly, some of the countries with high renewable energy consumption, such as Germany and Denmark, are not among the top 10 countries in terms of total energy consumption.

```
top_10_countries <- data_clean %>%
                    filter(Year == "2019") %>%
                    arrange(desc(RenewablePercentage)) %>%
                    head(10)

bar_chart <- ggplot(top_10_countries, aes(x = reorder(`Country Name`, RenewablePercentag
e),
              y = RenewablePercentage)) + geom_bar(stat = "identity",
              fill = "orange") + coord_flip() + theme_minimal() +
              theme(panel.grid.major.y = element_blank(), panel.grid.minor.y = element_bl
ank()) +
              labs(title = "Figure 2: Top 10 Countries with Highest Renewable Energy Cons
umption in 2019",
              x = "Country",
              y = "Renewable Energy Consumption (% of Total Final Energy Consumption)")

print(bar_chart)
```

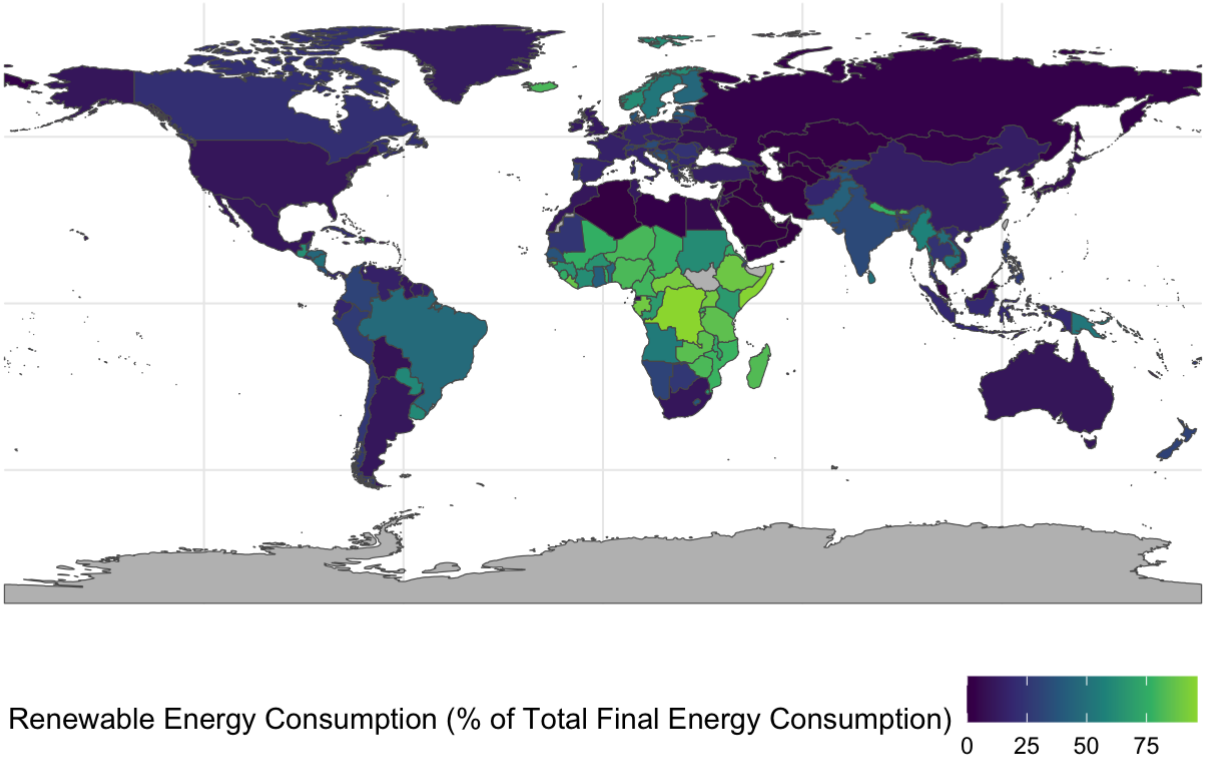


Map Plot

The third visualization is a geographic map that shows the renewable energy consumption for each country in 2019 (Figure 3). From this map, it is clear that renewable energy consumption is highest in Europe, especially in northern European countries such as Iceland, Norway, and Sweden. In contrast, countries in Africa and the Middle East have low renewable energy consumption.

```
world <- ne_countries(scale = "medium", returnclass = "sf")

data_map <- data_clean %>%
            filter(Year == "2019") %>%
            select(`Country Name`, `Country Code`, RenewablePercentage)

world_data <- left_join(world, data_map, by = c("adm0_a3" = "Country Code"))

map_plot <- ggplot(world_data) + geom_sf(aes(fill = RenewablePercentage)) +
            scale_fill_viridis_c(option = "viridis", end = 0.85, na.value = "gray") +
            theme_minimal() + theme(legend.position = "bottom") +
            labs(title = "Figure 3: Renewable Energy Consumption by Country in 2019",
            fill = "Renewable Energy Consumption (% of Total Final Energy Consumption)")

print(map_plot)
```

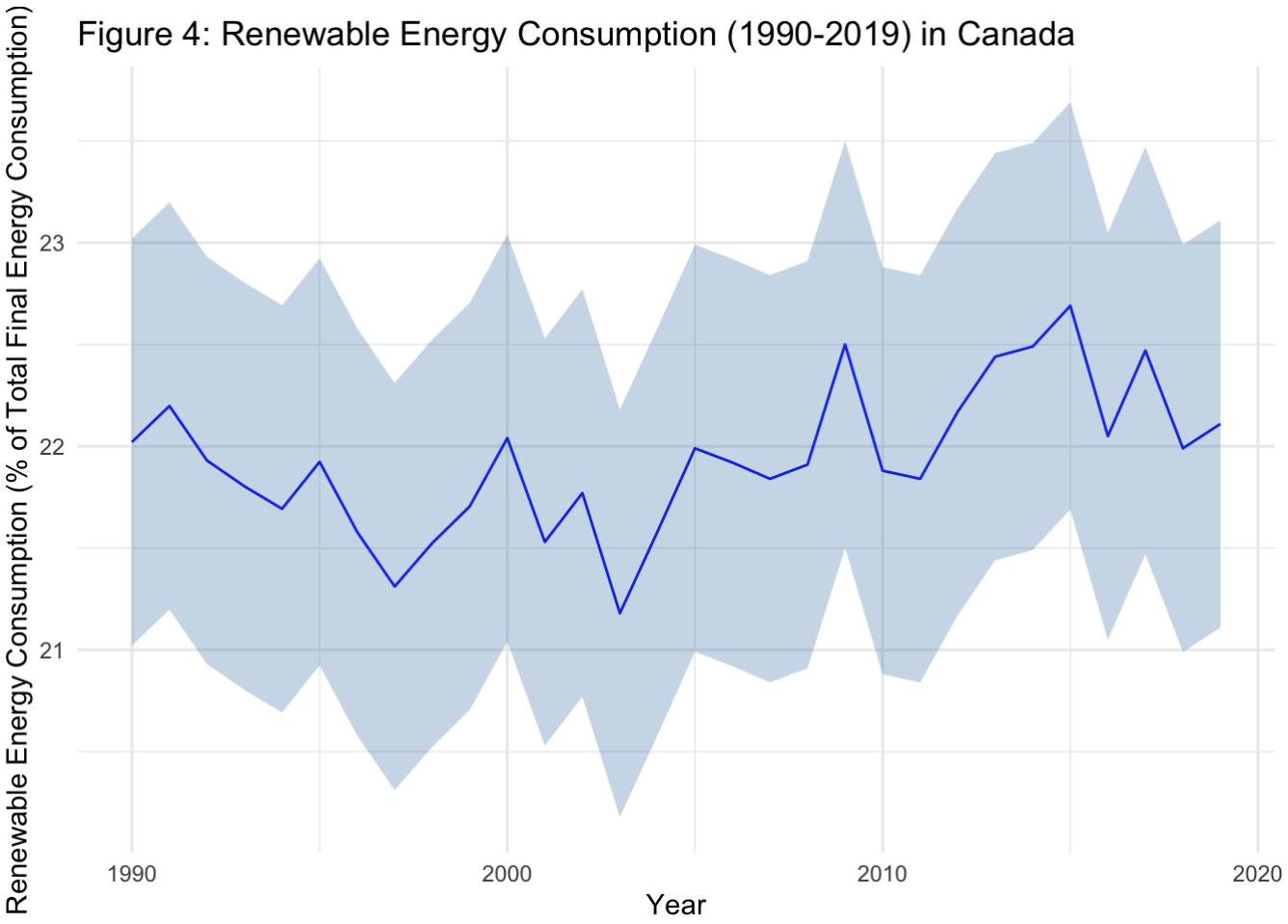
Figure 3: Renewable Energy Consumption by Country in 2019



Line Chart

The fourth visualization is a line chart that shows the trend of renewable energy consumption in Canada from 1990 to 2019 (Figure 4). This visualization shows that Canada has experienced a steady increase in renewable energy consumption over the past few decades, with a slight dip in the early 2000s.

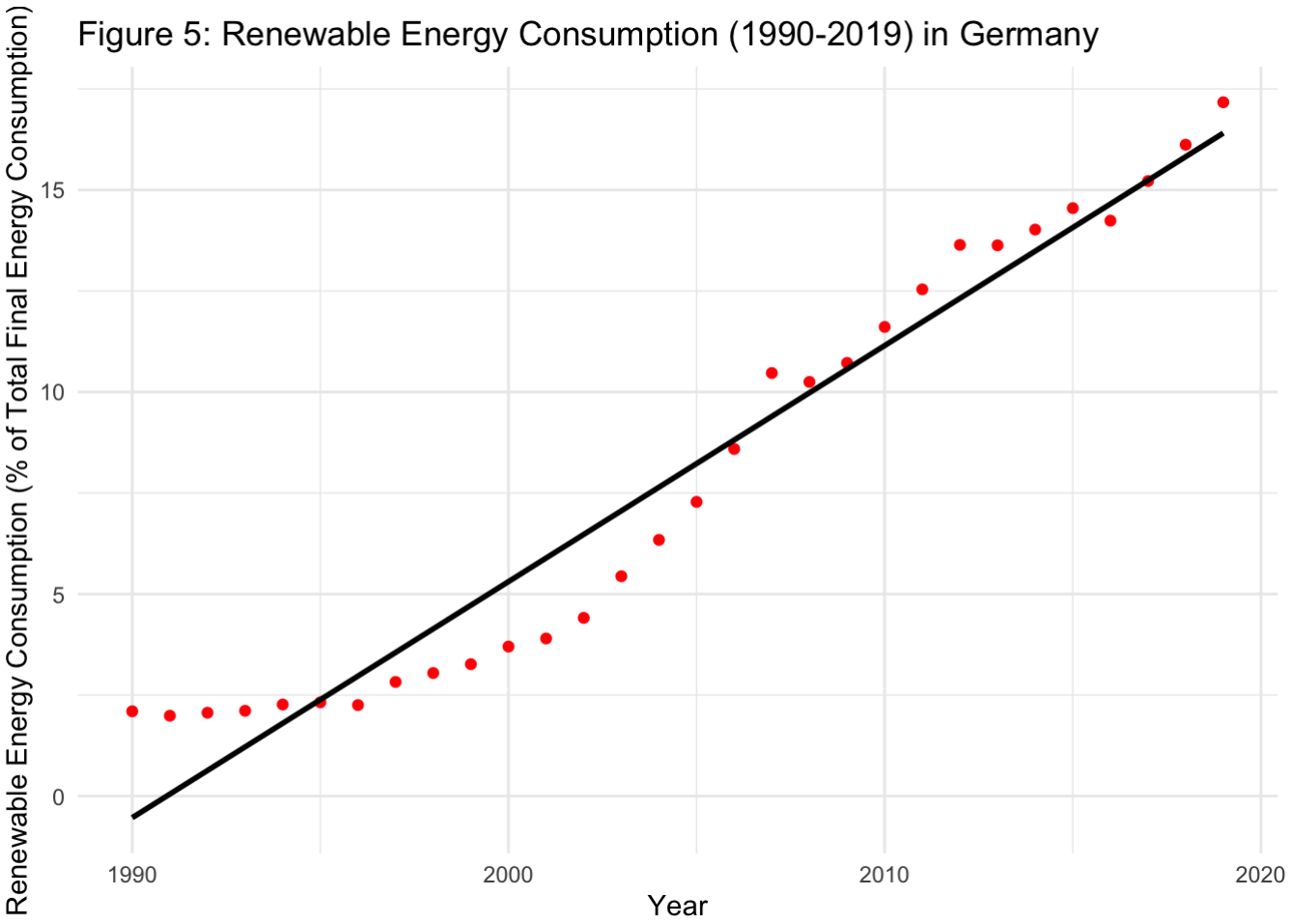
```
canada_data <- data_clean %>%
               filter(`Country Name` == "Canada")

canada_plot <- ggplot(canada_data, aes(x = as.integer(Year), y = RenewablePercentage)) +
               geom_line(color = "blue") +
               geom_ribbon(aes(ymin = RenewablePercentage - 1, ymax = RenewablePercentag
e + 1),
               alpha = 0.3, fill = "steelblue") + theme_minimal() +
               labs(title = "Figure 4: Renewable Energy Consumption (1990-2019) in Canad
a",
               x = "Year",
               y = "Renewable Energy Consumption (% of Total Final Energy Consumption)")

print(canada_plot)
```



Scatter Plot

The fifth visualization is a scatter plot that shows the relationship between renewable energy consumption and time in Germany from 1990 to 2019 (Figure 5). This visualization shows a positive correlation between renewable energy consumption and time, indicating that renewable energy consumption has been increasing steadily in Germany over the past few decades.

```
germany_data <- data_clean %>%
                filter(`Country Name` == "Germany")

germany_plot <- ggplot(germany_data, aes(x = as.integer(Year), y = RenewablePercentage))
+
                geom_point(color = "red") + geom_smooth(method = "lm",formula = y ~ x,
                color = "black", se = FALSE) + theme_minimal() + labs(title =
                "Figure 5: Renewable Energy Consumption (1990-2019) in Germany",
                x = "Year",
                y = "Renewable Energy Consumption (% of Total Final Energy Consumptio
n)")
print(germany_plot)
```



Conclusion

In conclusion, this report shows that renewable energy consumption is increasing across countries, with some countries experiencing a more rapid increase than others. The visualization techniques used in this report provide a clear and concise way to explore trends in renewable energy consumption and can be useful for policymakers and stakeholders in the energy sector. By presenting data in visually compelling ways, decision-makers can more easily grasp trends and make informed decisions that impact the future of renewable energy.