Final Project

DATASET REFERENCE:

"https://data.europa.eu/data/datasets/e50ca3e5-787e-4911-bc9e-d8012fa5af7a?locale=en"

```
# Load necessary library
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
```

```
# Read the CSV file
data <- read.csv("cattle birth 2016.csv")

# Create a mapping from month abbreviations to numbers
month_to_num <- c(JAN = 1, FEB = 2, MAR = 3, APR = 4, MAY = 5, JUN = 6,
                  JUL = 7, AUG = 8, SEP = 9, OCT = 10, NOV = 11, DEC = 12)

# Convert BIRTH_MONTH column to uppercase (if not already) and then to numeric
data$BIRTH_MONTH <- sapply(data$BIRTH_MONTH, function(x) month_to_num[toupper(x)])

# Display the updated data
head(data)
```

```
  BIRTH_YEAR COUNTY_ORIGIN BIRTH_MONTH CALF_BREED_TYPE TOTAL_BIRTH
1       2016        CARLOW           4            BEEF        3670
2       2016        CARLOW           4           DAIRY         680
3       2016        CARLOW           8            BEEF         778
4       2016        CARLOW           8           DAIRY          80
5       2016        CARLOW          12            BEEF         835
6       2016        CARLOW          12           DAIRY         123
```

PART:1

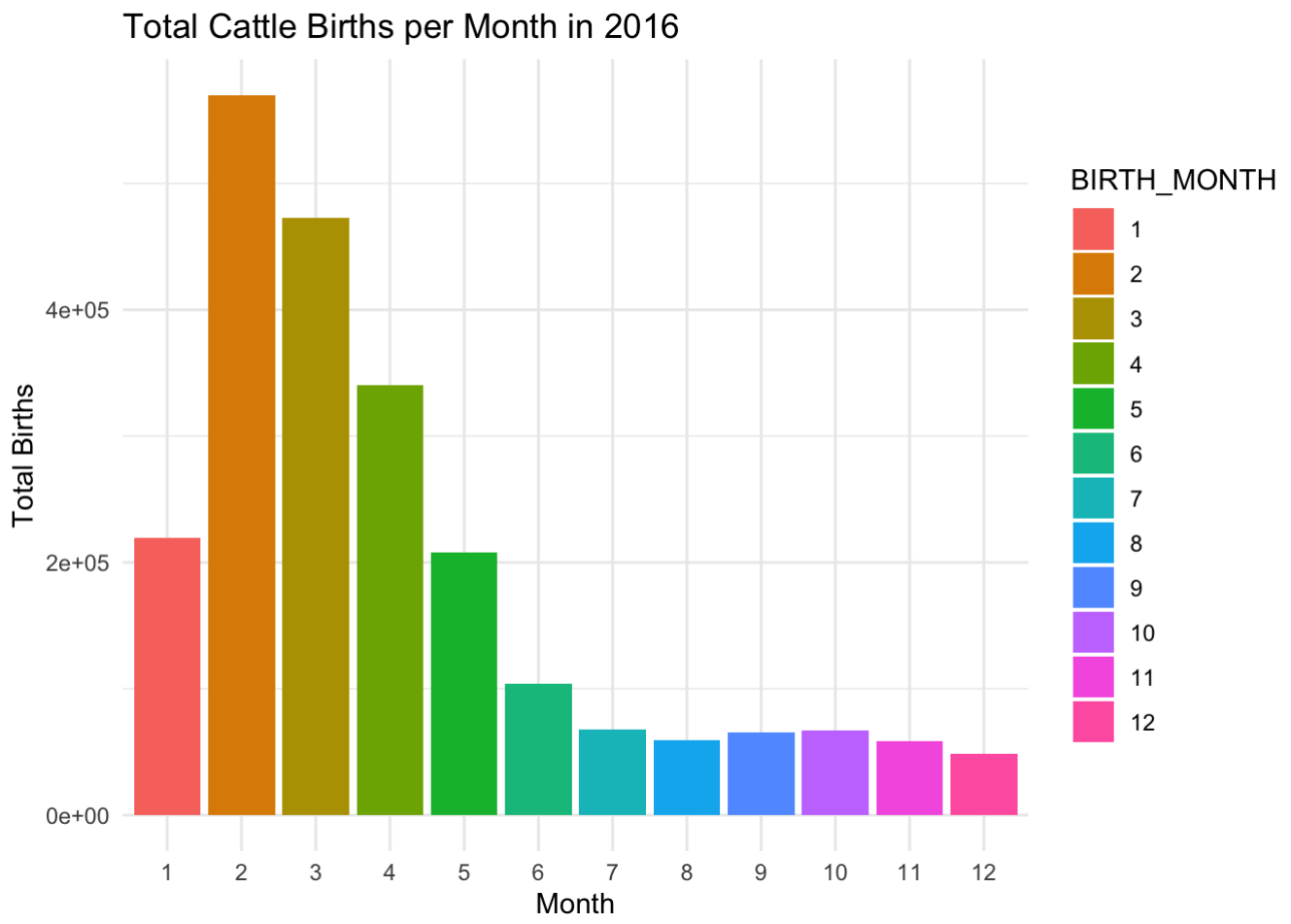
```
summary(data)
```

```
   BIRTH_YEAR   COUNTY_ORIGIN       BIRTH_MONTH    CALF_BREED_TYPE
 Min.   :2016   Length:624         Min.   : 1.00   Length:624
 1st Qu.:2016   Class :character   1st Qu.: 3.75   Class :character
 Median :2016   Mode  :character   Median : 6.50   Mode  :character
 Mean   :2016                      Mean   : 6.50
 3rd Qu.:2016                      3rd Qu.: 9.25
 Max.   :2016                      Max.   :12.00
```

```
 TOTAL_BIRTH
Min.   :    6.0
1st Qu.:  454.5
Median : 1561.5
Mean   : 3655.0
3rd Qu.: 3825.8
Max.   :95720.0
```

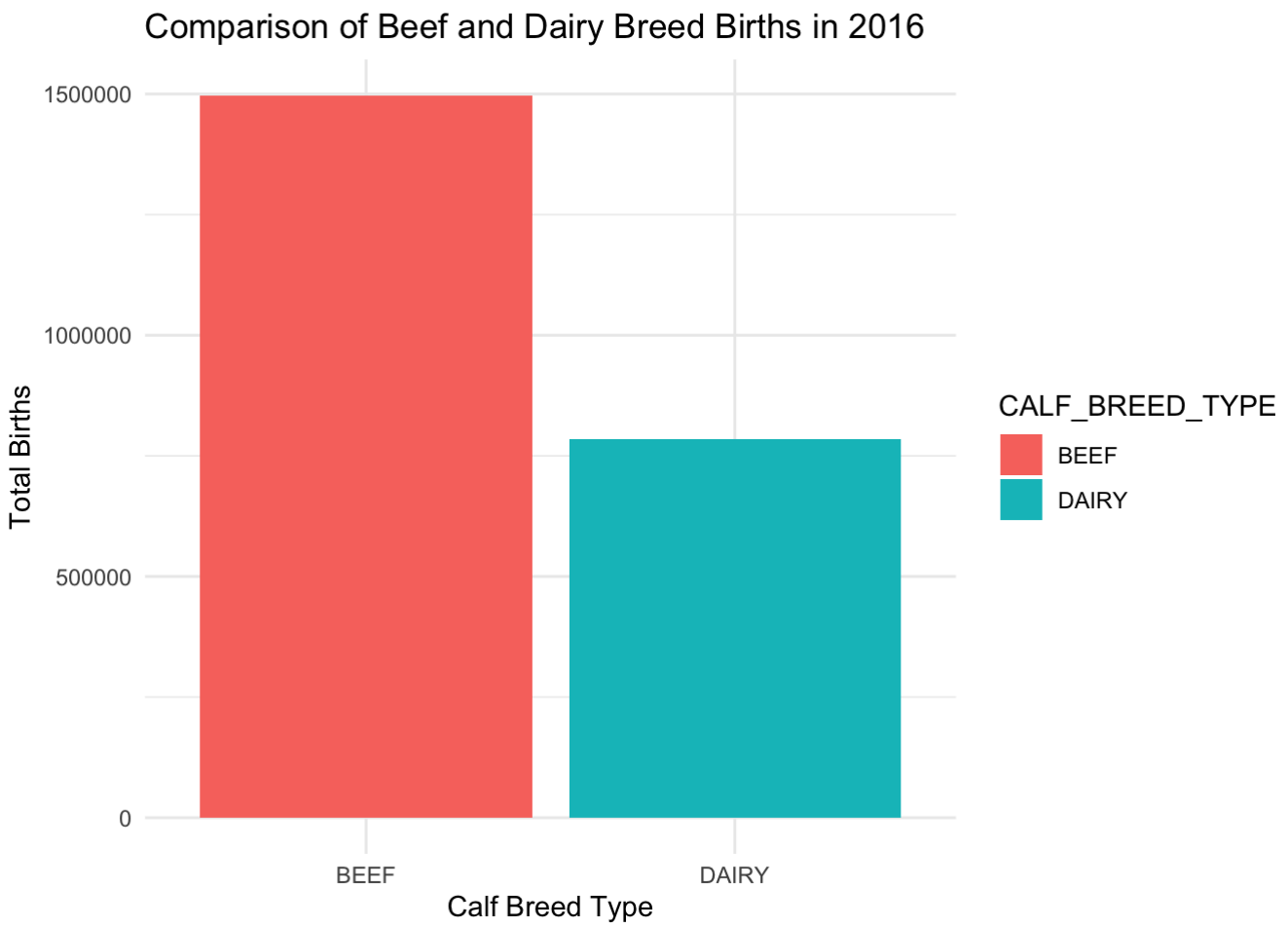
```r
library(ggplot2)
library(dplyr)

# Plot 1: Total Cattle Births per Month in 2016
data %>%
  group_by(BIRTH_MONTH) %>%
  summarise(TOTAL_BIRTH = sum(TOTAL_BIRTH)) %>%
  mutate(BIRTH_MONTH = factor(BIRTH_MONTH, levels=c(1,2,3,4,5,6,7,8,9,10,11,12))) %>%
  ggplot(aes(x = BIRTH_MONTH, y = TOTAL_BIRTH, fill = BIRTH_MONTH)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Total Cattle Births per Month in 2016", x = "Month", y = "Total Births
```




```r
# Plot 2: Comparison of Beef and Dairy Breed Births in 2016
data %>%
  group_by(CALF_BREED_TYPE) %>%
  summarise(TOTAL_BIRTH = sum(TOTAL_BIRTH)) %>%
  ggplot(aes(x = CALF_BREED_TYPE, y = TOTAL_BIRTH, fill = CALF_BREED_TYPE)) +
  geom_bar(stat = "identity") +
```

```
  theme_minimal() +
  labs(title = "Comparison of Beef and Dairy Breed Births in 2016", x = "Calf Breed Ty
```



The bar chart illustrates the total number of cattle births for each month in the year 2016, revealing a seasonal pattern in cattle breeding. The most significant number of births occur in February, with nearly 400,000, followed closely by March. After the spring peak, there's a steady decline in births from April through to the end of the year. The chart indicates that the cattle industry likely plans breeding schedules to coincide with favorable conditions, leading to a concentration of births in the early part of the year. The consistency in birth numbers from June to December suggests a deliberate management of breeding cycles to suit operational or environmental conditions.

The bar chart "Comparison of Beef and Dairy Breed Births in 2016" presents a side-by-side comparison of the total number of births for beef and dairy cattle breeds within the year. The red bar represents beef breeds and towers significantly, indicating over 1,000,000 births. In contrast, the blue bar represents dairy breeds, showing more than 500,000 births but considerably fewer than beef breeds. The disparity suggests a higher number of beef breed cattle were born in 2016 compared to dairy breeds, which could be attributed to various factors such as market demand or specific breeding programs within the cattle industry during that year. The chart clearly demonstrates the predominance of beef breed births over dairy in the recorded period.

PART:2

The `janitor` package in R significantly streamlines data cleaning and preparation, making it a valuable asset for data analysts and researchers. Key features include `clean_names()`, which standardizes column names for consistency, and `tabyl()`, which simplifies the creation of frequency tables for data examination. It also offers robust tools for managing duplicates, reordering and

cleaning factor levels, and performing advanced cross-tabulations. Additionally, the package provides functions for decorative formatting, like adding totals and subtotals to tables, and facilitates flexible data type conversions. Renowned for its user-friendly and intuitive functions, `janitor` enhances the efficiency and effectiveness of data cleaning processes, making it a popular choice in the R community.

```
#Package
library(janitor)
```

```
Attaching package: 'janitor'

The following objects are masked from 'package:stats':

    chisq.test, fisher.test
```

```
#Function 1
cattle_birth <- data %>% clean_names()
head(cattle_birth)
```

```
  birth_year county_origin birth_month calf_breed_type total_birth
1       2016        CARLOW           4            BEEF        3670
2       2016        CARLOW           4           DAIRY         680
3       2016        CARLOW           8            BEEF         778
4       2016        CARLOW           8           DAIRY          80
5       2016        CARLOW          12            BEEF         835
6       2016        CARLOW          12           DAIRY         123
```

```
#Function 2
breed_frequency <- cattle_birth %>% tabyl(calf_breed_type)
print(breed_frequency)
```

```
 calf_breed_type   n percent
            BEEF 312     0.5
           DAIRY 312     0.5
```

```
#Function 3
library(dplyr)
library(janitor)
monthly_totals <- cattle_birth %>%
                  group_by(birth_month) %>%
                  summarise(Total_Births = sum(total_birth))
monthly_totals_with_total <- monthly_totals %>%
                             adorn_totals("row")
print(monthly_totals_with_total)
```

```
 birth_month Total_Births
           1       219716
           2       569903
           3       472714
           4       340127
```

```
            5       207919
            6       103965
            7        67717
            8        59592
            9        65185
           10        66817
           11        58691
           12        48352
        Total      2280698
```

PART:3

```
library(dplyr)
# Define the S3 class
setClass("CattleStats", slots = c(stats = "data.frame"))

# The main function
analyzeCattleBirths <- function(data, group_by) {
  require(dplyr)

  # Calculating statistics
  stats <- data %>%
    group_by(!!sym(group_by)) %>%
    summarise(Mean = mean(TOTAL_BIRTH),
              Median = median(TOTAL_BIRTH),
              Std_Dev = sd(TOTAL_BIRTH)) %>%
    as.data.frame()

  # Creating S3 object
  stats_obj <- new("CattleStats", stats = stats)
  class(stats_obj) <- "CattleStats"
  return(stats_obj)
}

# Print method for CattleStats
print.CattleStats <- function(object) {
  print(object@stats)
}

# Summary method for CattleStats
summary.CattleStats <- function(object) {
  summary(object@stats)
}

# Plot method for CattleStats
plot.CattleStats <- function(object, group_by, ...) {
  barplot(as.matrix(object@stats[, -1]), beside = TRUE,
          main = "Cattle Birth Statistics",
          ylab = "Value", xlab = group_by,
          col = rainbow(nrow(object@stats)),
          legend.text = rownames(object@stats), ...)
}

# Now execute the function to create the result object
result <- analyzeCattleBirths(data, "BIRTH_MONTH")
```

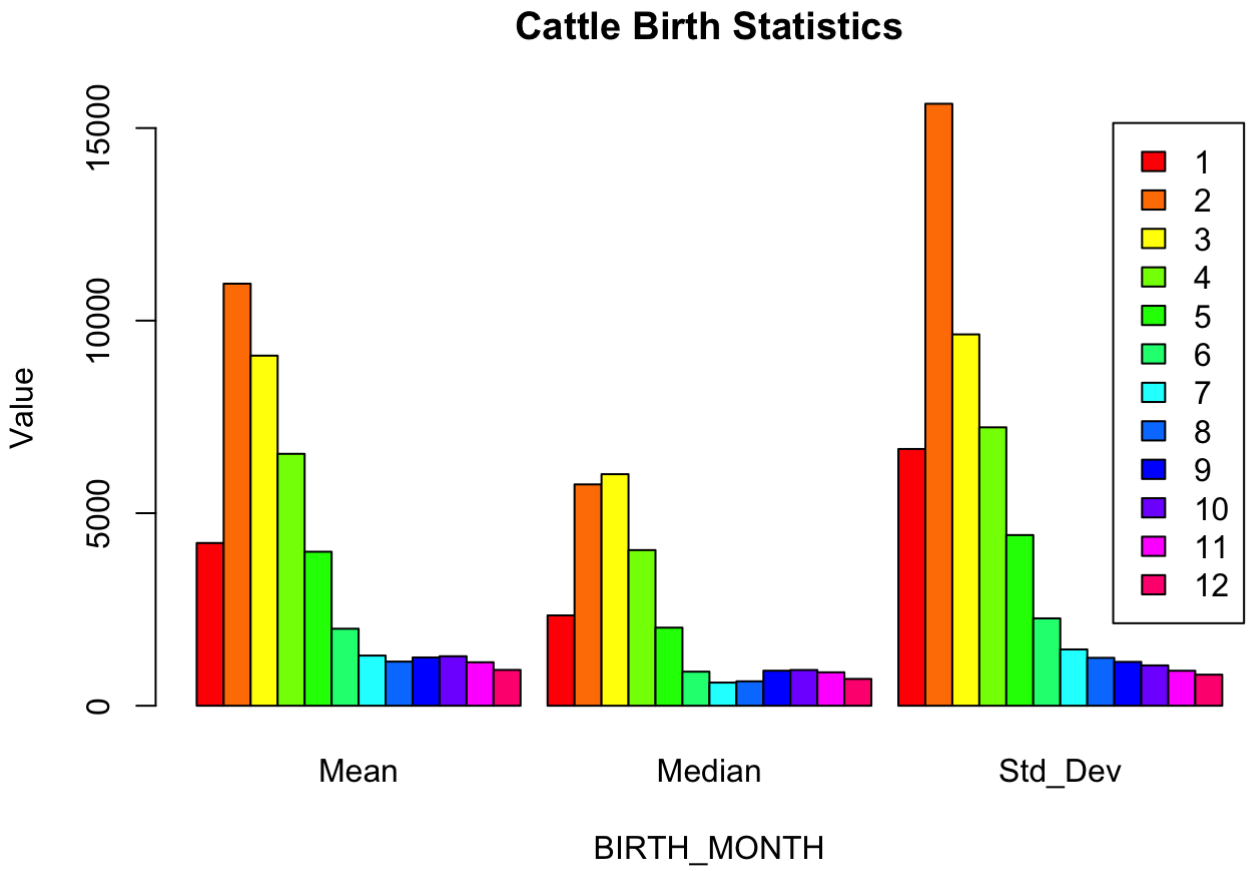
```
# Now you can print, summarize, and plot the result
print(result)
```

```
   BIRTH_MONTH       Mean Median    Std_Dev
1            1  4225.3077 2345.5  6666.5186
2            2 10959.6731 5747.0 15630.9017
3            3  9090.6538 6012.5  9642.6629
4            4  6540.9038 4041.0  7228.8914
5            5  3998.4423 2027.5  4431.4446
6            6  1999.3269  884.5  2268.3274
7            7  1302.2500  601.0  1461.9106
8            8  1146.0000  634.0  1243.7477
9            9  1253.5577  909.5  1138.9503
10          10  1284.9423  928.0  1047.1170
11          11  1128.6731  868.0   908.3569
12          12   929.8462  696.5   807.0202
```

```
summary(result)
```

```
  BIRTH_MONTH         Mean             Median          Std_Dev     
 Min.   : 1.00   Min.   :  929.8   Min.   : 601.0   Min.   :  807  
 1st Qu.: 3.75   1st Qu.: 1226.7   1st Qu.: 825.1   1st Qu.: 1116  
 Median : 6.50   Median : 1650.8   Median : 918.8   Median : 1865  
 Mean   : 6.50   Mean   : 3655.0   Mean   :2141.2   Mean   : 4373  
 3rd Qu.: 9.25   3rd Qu.: 4804.2   3rd Qu.:2769.4   3rd Qu.: 6807  
 Max.   :12.00   Max.   :10959.7   Max.   :6012.5   Max.   :15631  
```

```
plot(result, group_by = "BIRTH_MONTH") # The corrected call to your plot method
```



The “Cattle Birth Statistics” bar chart compares the mean, median, and standard deviation of cattle births by month for an unspecified year. The mean and median values peak in February, suggesting that this month typically has the highest average number of births. The standard deviation is greatest in January, indicating a larger variability in the number of births during this month compared to others. This could reflect a seasonal breeding pattern, with a more consistent number of births around the peak month and greater fluctuations at the beginning of the year. Overall, the chart provides a clear visual representation of cattle birth trends and variability throughout the year.

Citation

```
citation("janitor")
```

```
To cite package 'janitor' in publications use:

  Firke S (2023). _janitor: Simple Tools for Examining and Cleaning
  Dirty Data_. R package version 2.2.0,
  <https://CRAN.R-project.org/package=janitor>.

A BibTeX entry for LaTeX users is

  @Manual{,
    title = {janitor: Simple Tools for Examining and Cleaning Dirty Data},
    author = {Sam Firke},
    year = {2023},
    note = {R package version 2.2.0},
    url = {https://CRAN.R-project.org/package=janitor},
  }
```