Exercise for Week 1: Introduction to R and RStudio

Objective:
The goal of this exercise is to familiarize yourself with the R environment, learn basic R commands, and create your first simple data visualization using `ggplot2`.

Duration: 1-2 hours

Part 1: Getting Started with R and RStudio

1. **Install R and RStudio**:
   - If you haven't already, install R from CRAN.
   - Install RStudio from RStudio.
2. **Explore the RStudio Interface**:
   - **Console**: Type `2 + 2` and press Enter to see the result in the Console.
   - **Script Editor**: Create a new script (File > New File > R Script) and save it as `week1_exercise.R`.
   - **Environment Pane**: Notice how variables and data are listed here as you create them.
   - **Plots Pane**: This is where your visualizations will appear.

Part 2: Basic R Commands

1. **Creating Variables**:
   - Open your `week1_exercise.R` script and enter the following commands:

     ```
     # Creating a numeric vector
     data_vector <- c(1, 2, 3, 4, 5)

     # Calculate basic statistics
     mean_value <- mean(data_vector)
     sum_value <- sum(data_vector)
     ```

   - **Task**: Run the script and observe the output in the Console. What is the mean and sum of `data_vector`?

2. **Working with Data Frames**:
   - Data frames are a key data structure in R for storing datasets.
   - Enter the following code in your script:

     ```
     # Creating a simple data frame
     df <- data.frame(
       Name = c("Species A", "Species B", "Species C"),
       Count = c(23, 15, 19)
     )
     ```

```
# View the data frame
print(df)
```

- **Task**: Run the code and explore the `df` data frame. What are the species names and their counts?

Part 3: Introduction to `ggplot2`

1. **Installing and Loading `ggplot2`**:
   - Install `ggplot2` if you haven't already:

     ```
     install.packages("ggplot2")
     ```

   - Load the package:

     ```
     library(ggplot2)
     ```

2. **Creating Your First Plot**:
   - Use `ggplot2` to create a simple bar plot:

     ```
     # Basic bar plot
     ggplot(data = df, aes(x = Name, y = Count)) +
       geom_bar(stat = "identity") +
       labs(
         title = "Species Count",
         x = "Species",
         y = "Count"
       ) +
       theme_minimal()
     ```

   - **Task**: Run the script to generate the plot. This plot shows the count of each species in the data frame.

3. **Customizing the Plot**:
   - Modify the code to change the color of the bars and the title of the plot:

     ```
     # Customizing the bar plot
     ggplot(data = df, aes(x = Name, y = Count, fill = Name)) +
       geom_bar(stat = "identity") +
       labs(
         title = "Species Count by Name",
         x = "Species",
         y = "Count"
       ) +
       theme_minimal() +
       theme(legend.position = "none")
     ```

   - **Task**: Run the modified code. How does changing the `fill` aesthetic affect the plot? How does the plot look without the legend?

Part 4: Reflect and Document

1. **Reflect**:
    - Take a moment to reflect on what you've learned. What was challenging? What was straightforward?
2. **Document Your Work**:
    - Add comments to your script explaining each section of code.
    - Save your script and be prepared to discuss your experience in the next class.

Submission Instructions

- **Save your script**: Ensure your script is well-commented and saved as `week1_exercise_name.R`.
- **Upload**: Submit your script on TEAMS.
- **Discussion**: Be ready to discuss your experiences and any questions you have during the next class.

Expected Outcome

By the end of this exercise, you should be comfortable navigating RStudio, writing basic R scripts, and creating simple visualizations with `ggplot2`. This foundational knowledge will be built upon in subsequent weeks.