Part 1: Creating and Analyzing Histograms

1. **Create a Sample Dataset**:
   o   Start by generating a random dataset:

   ```
   # Sample data
   set.seed(42)
   data <- rnorm(200, mean = 5, sd = 2)
   ```

2. **Basic Histogram**:
   o   Use ggplot2 to create a basic histogram:

   ```
   # library(ggplot2)

   # Basic histogram
   ggplot(data.frame(data), aes(x = data)) +
     geom_histogram(binwidth = 0.5, fill = "skyblue", color =
   "white") +
     labs(title = "Histogram of Data", x = "Value", y =
   "Frequency") +
     theme_minimal()
   ```

   o   **Task**: Run the code to create the histogram. Experiment with different bin
       sizes by adjusting the binwidth parameter. How does changing the bin width
       affect the interpretation of the data distribution?

Part 2: Creating and Analyzing Density Plots

1. **Basic Density Plot**:

   ```
   # Density plot
   ggplot(data.frame(data), aes(x = data)) +
     geom_density(fill = "blue", alpha = 0.5) +
     labs(title = "Density Plot of Data", x = "Value", y =
   "Density") +
     theme_minimal()
   ```

   o   **Task**: Run the code to create the density plot. Compare it to the histogram.
       How does the density plot enhance your understanding of the data distribution?

2. **Overlaying Histogram and Density Plot:**:

   ```
   # Overlaying histogram and density plot
   ggplot(data.frame(data), aes(x = data)) +
     geom_histogram(aes(y = ..density..), binwidth = 0.5, fill =
   "skyblue", color = "white") +
     geom_density(color = "red", size = 1) +
     labs(title = "Overlayed Histogram and Density Plot", x =
   "Value", y = "Density") +
     theme_minimal())
   ```

   o   **Task:** Run the code to overlay the histogram and density plot. What new
       insights does this combined visualization prsovide?

Part 3: Creating and Analyzing Boxplots

1. **Basic Boxplot:**

```
# Basic boxplot
ggplot(data.frame(data), aes(y = data)) +
  geom_boxplot(fill = "lightgreen") +
  labs(title = "Boxplot of Data", y = "Value") +
  theme_minimal()
```

   o **Task:** Run the code to create the boxplot. Identify any outliers in the data. How does the boxplot help you understand the spread and skewness of the data?

2. **Boxplot with Jittered Points:**

```
# Boxplot with jittered points
ggplot(data.frame(data), aes(y = data)) +
  geom_boxplot(fill = "lightgreen", outlier.color = "red") +
  geom_jitter(width = 0.2, alpha = 0.4, color = "blue") +
  labs(title = "Boxplot with Jittered Points", y = "Value") +
  theme_minimal()
```

   o **Task:** Run the code to create a boxplot with jittered points. How do the jittered points enhance the interpretation of the boxplot?

Part 4: Reflect and Document

1. **Reflect**:
   o Reflect on the process of creating and analyzing visualizations. Which visualization (histogram, density plot, or boxplot) provided the most insights into the data distribution? Why?
2. **Document Your Work**:
   o Add comments to your script explaining each step.
   o Save your script as week4_exercise.R and be prepared to discuss your findings in the next class.

Submission Instructions

- **Save your script**: Ensure your script is well-commented and saved as `week4_exercise_name.R`.
- **Upload**: Submit your script via 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 using ggplot2 to create histograms, density plots, and boxplots. You will also gain experience identifying outliers and understanding the underlying distribution of your data through different visualizations.