527_Project2_Analysis

Weixin Xiao

2025-09-16

```r
# Load data and calculate Delta
data <- read.csv("EdemaData - Copy.csv") %>%
  mutate(Delta = Edema_Post - Edema_Pre)

# Preview of first 10 rows (full analysis uses all 180 patients)
kable(head(data, 10), digits = 3)
```

PatientID	Edema_Pre	Edema_Post	Group	Delta
1	9.901	9.556	Placebo	-0.345
2	9.147	7.708	Placebo	-1.440
3	10.638	10.336	Placebo	-0.302
4	10.483	10.150	Placebo	-0.333
5	6.810	7.423	Placebo	0.614
6	9.320	9.142	Placebo	-0.177
7	8.765	8.524	Placebo	-0.242
8	7.764	8.606	Placebo	0.842
9	8.795	8.106	Placebo	-0.689
10	8.805	9.213	Placebo	0.408

```r
# Summary statistics by group
summary_stats <- data %>%
  group_by(Group) %>%
  summarise(
    N = n(),
    Pre_Mea = mean(Edema_Pre),
    Pre_SD = sd(Edema_Pre),
    Pre_Med = median(Edema_Pre),
    Post_Mea = mean(Edema_Post),
    Post_SD = sd(Edema_Post),
    Post_Med = median(Edema_Post),
    Delta_Mea = mean(Delta),
    Delta_SD = sd(Delta),
    Delta_Med = median(Delta)
  )

kable(summary_stats, digits = 3)
```

Group	N	Pre_Mea	Pre_SD	Pre_Med	Post_Mea	Post_SD	Post_Med	Delta_Mea	Delta_SD	Delta_Med
Placebo	90	8.997	1.456	8.988	8.717	1.515	8.671	-0.280	0.652	-0.294
Treatment	90	8.924	1.315	8.713	8.408	1.302	8.302	-0.516	0.651	-0.510

```r
# Correlation between Pre and Post by group
correlations <- data %>%
  group_by(Group) %>%
  summarise(Correlation_Pre_Post = cor(Edema_Pre, Edema_Post))

kable(correlations, digits = 3)
```

Group	Correlation_Pre_Post
Placebo	0.904
Treatment	0.876

```r
# Function to detect outliers
detect_outliers <- function(x) {
  Q1 <- quantile(x, 0.25)
  Q3 <- quantile(x, 0.75)
  IQR <- Q3 - Q1
  outliers <- x < (Q1 - 1.5 * IQR) | x > (Q3 + 1.5 * IQR)
  return(outliers)
}

# Add outlier flag and summarize
data <- data %>%
  group_by(Group) %>%
  mutate(Is_Outlier = detect_outliers(Delta)) %>%
  ungroup()

kable(table(data$Group, data$Is_Outlier), col.names = c("Non-Outlier", "Outlier"))
```

	Non-Outlier	Outlier
Placebo	87	3
Treatment	90	0

```r
# Long format for pre/post boxplot
data_long <- data %>%
  pivot_longer(cols = c(Edema_Pre, Edema_Post), names_to = "Time", values_to = "Edema")

# Boxplot for edema levels
ggplot(data_long, aes(x = Time, y = Edema, fill = Group)) +
  geom_boxplot(outlier.shape = 19, outlier.alpha = 0.5) +
  theme_minimal() +
  labs(title = "Edema Levels by Time and Group", x = "Time Point", y = "Edema Level") +
  scale_fill_manual(values = c("Placebo" = "#ADD8E6", "Treatment" = "#90EE90")) +
  scale_x_discrete(labels = c("Edema_Pre" = "Pre", "Edema_Post" = "Post"))
```

Edema Levels by Time and Group
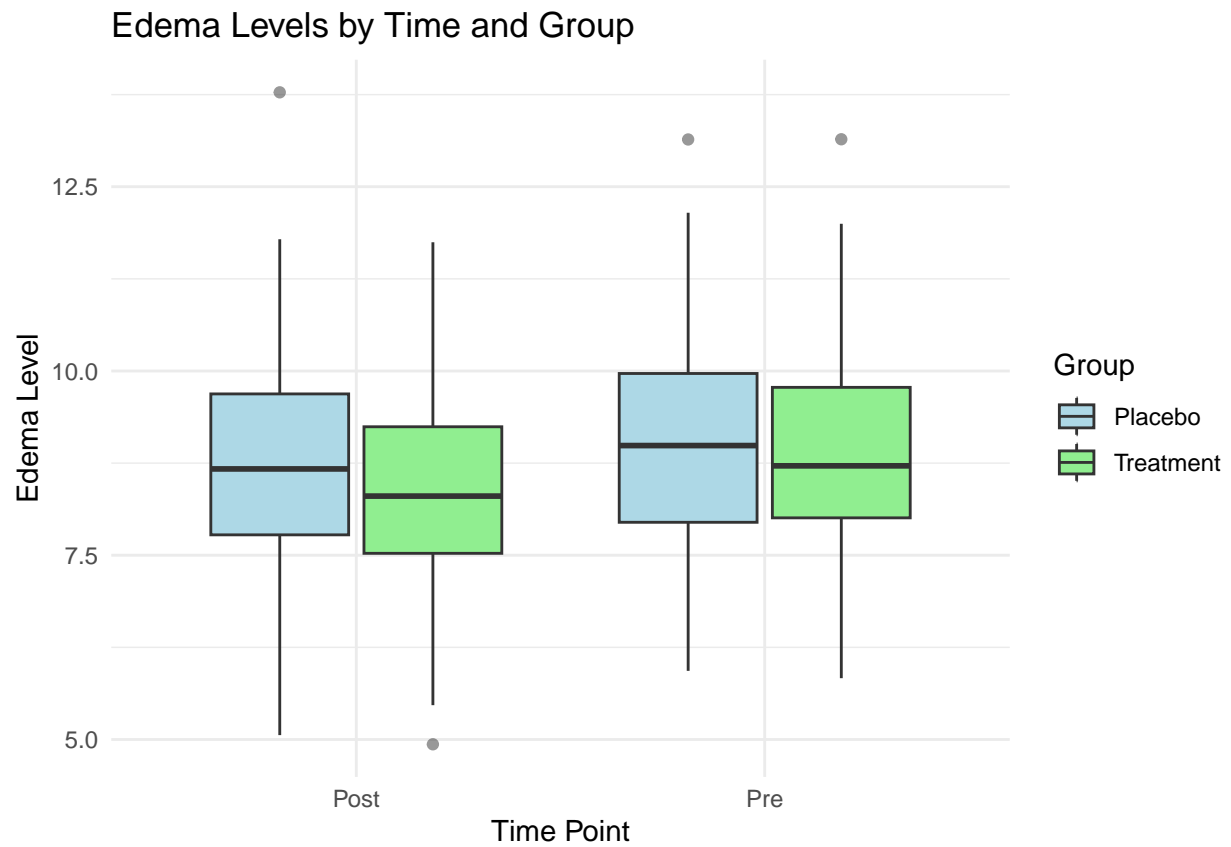

```
# Boxplot for Delta
ggplot(data, aes(x = Group, y = Delta, fill = Group)) +
  geom_boxplot(outlier.shape = 19, outlier.alpha = 0.5) +
  theme_minimal() +
  labs(title = "Change in Edema (Post - Pre)", x = "Group", y = "Delta") +
  scale_fill_manual(values = c("Placebo" = "#ADD8E6", "Treatment" = "#90EE90"))
```

Change in Edema (Post − Pre)
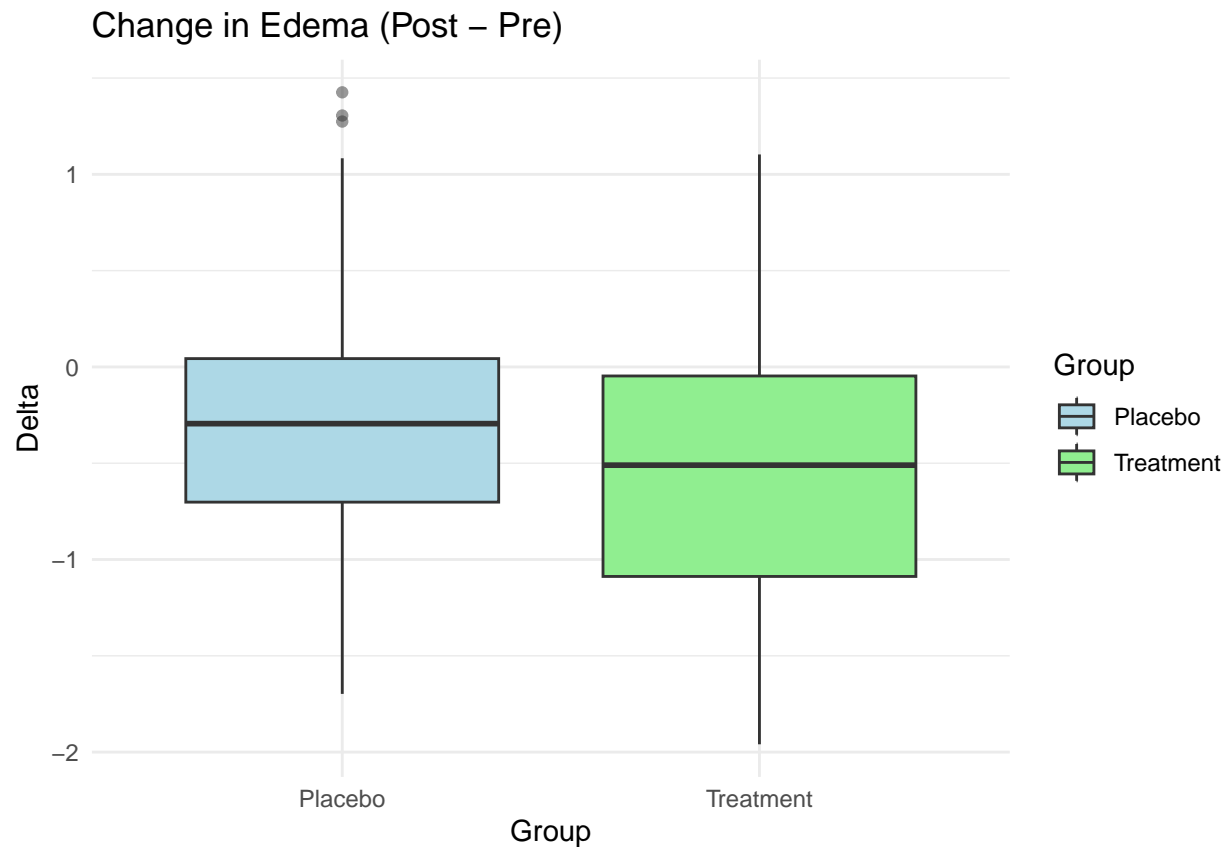

```r
# Reusable histogram function
plot_histogram <- function(df, var, group_name, fill_color, title_suffix) {
  ggplot(df, aes(x = !!sym(var))) +
    geom_histogram(bins = 20, fill = fill_color, color = "black") +
    theme_minimal() +
    labs(title = paste0(title_suffix, " (", group_name, ")"), x = var, y = "Count")
}
```
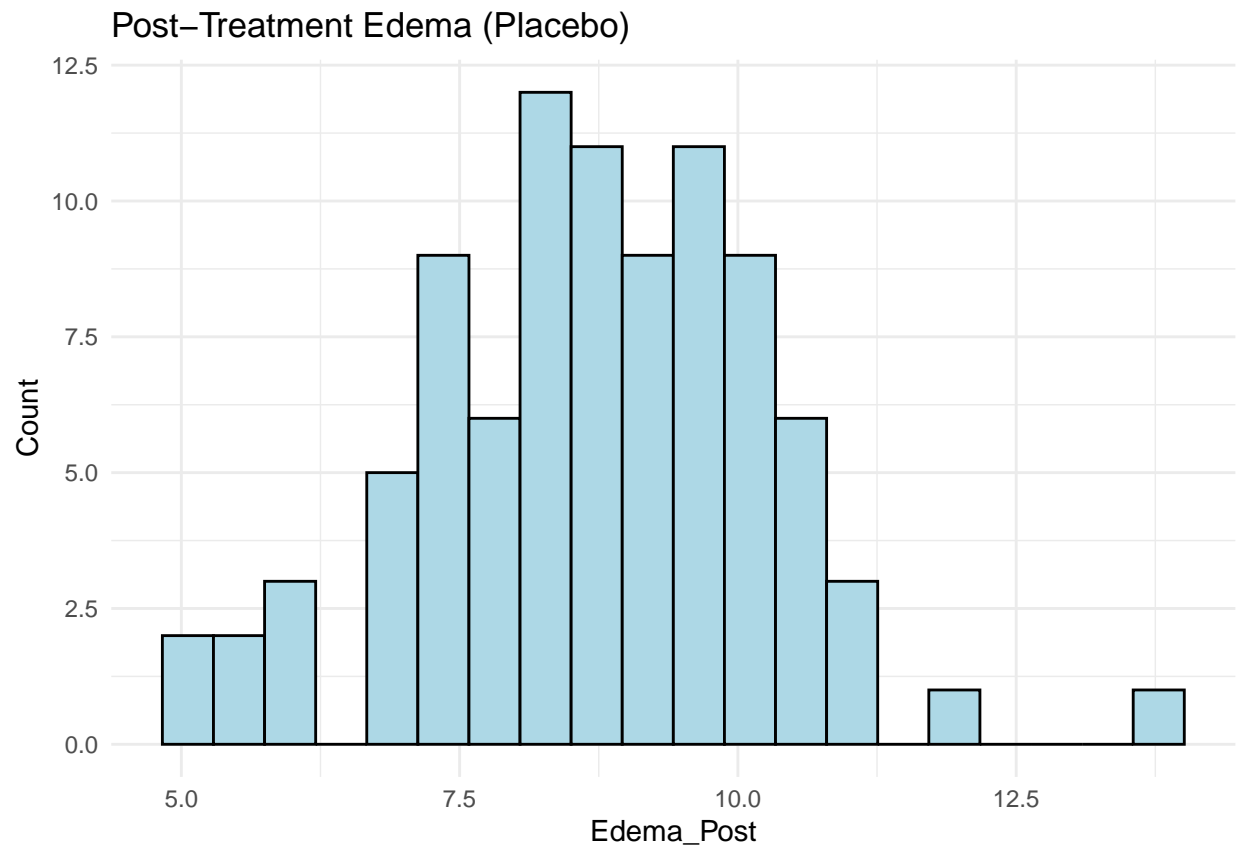
```r
# Pre - Placebo
plot_histogram(data[data$Group == "Placebo", ], "Edema_Pre", "Placebo", "#ADD8E6", "Pre-Treatment Edema
```
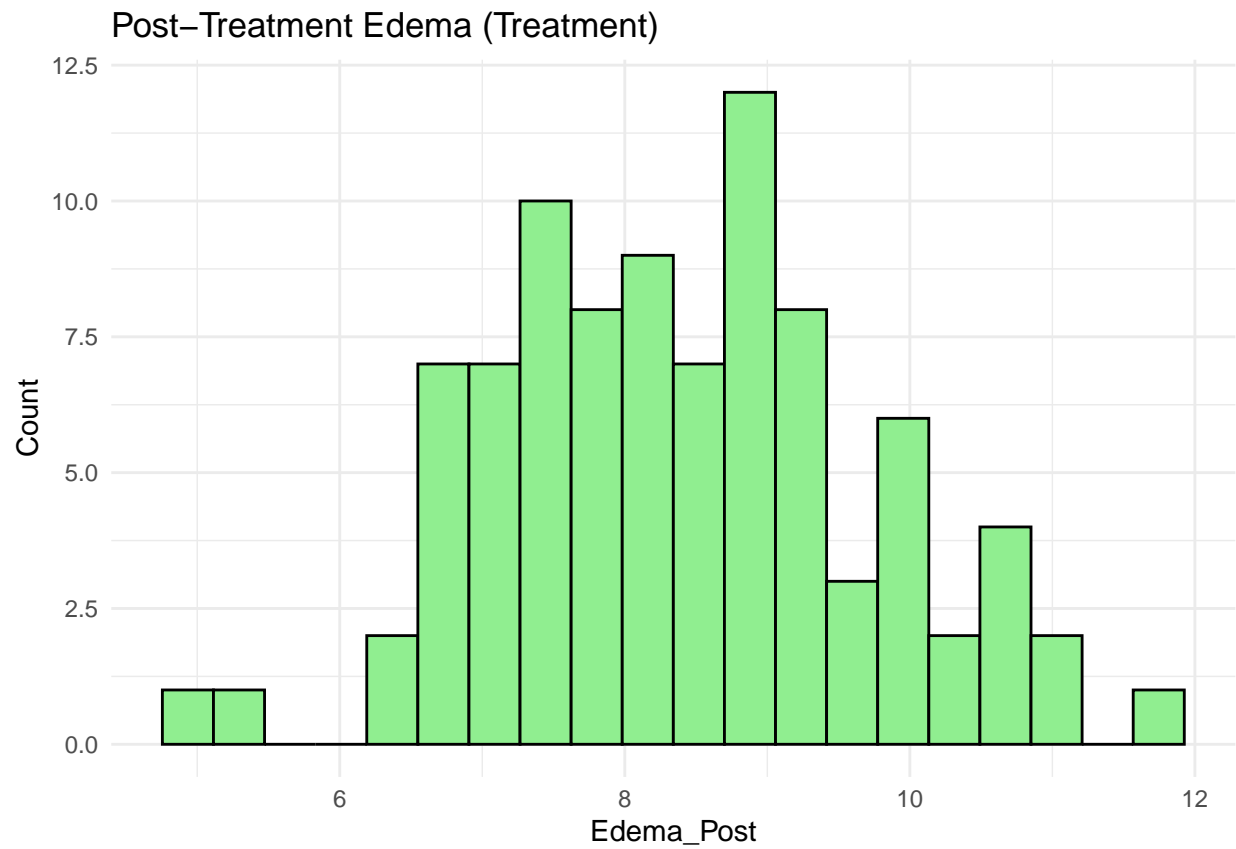
Pre–Treatment Edema (Placebo)
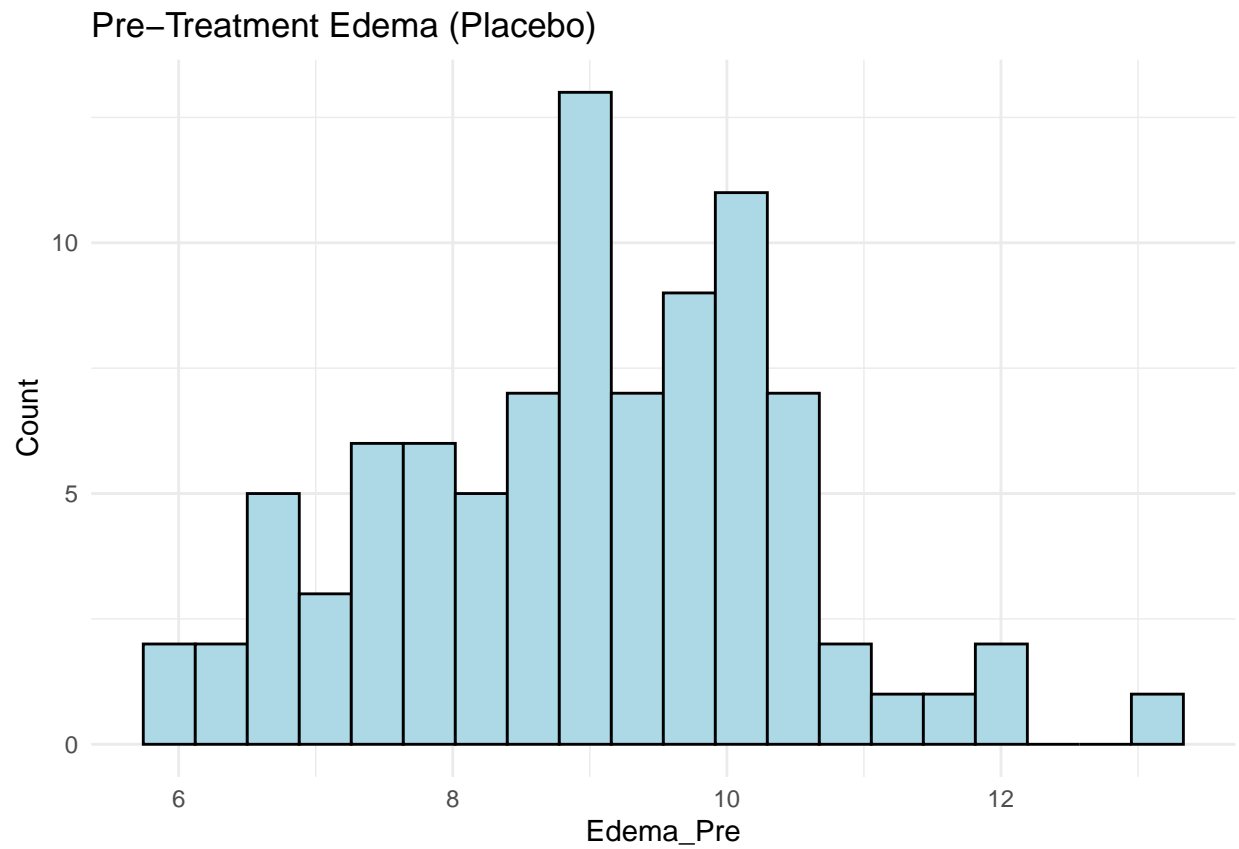

```
# Pre - Treatment
plot_histogram(data[data$Group == "Treatment", ], "Edema_Pre", "Treatment", "#90EE90", "Pre-Treatment E
```

Pre−Treatment Edema (Treatment)
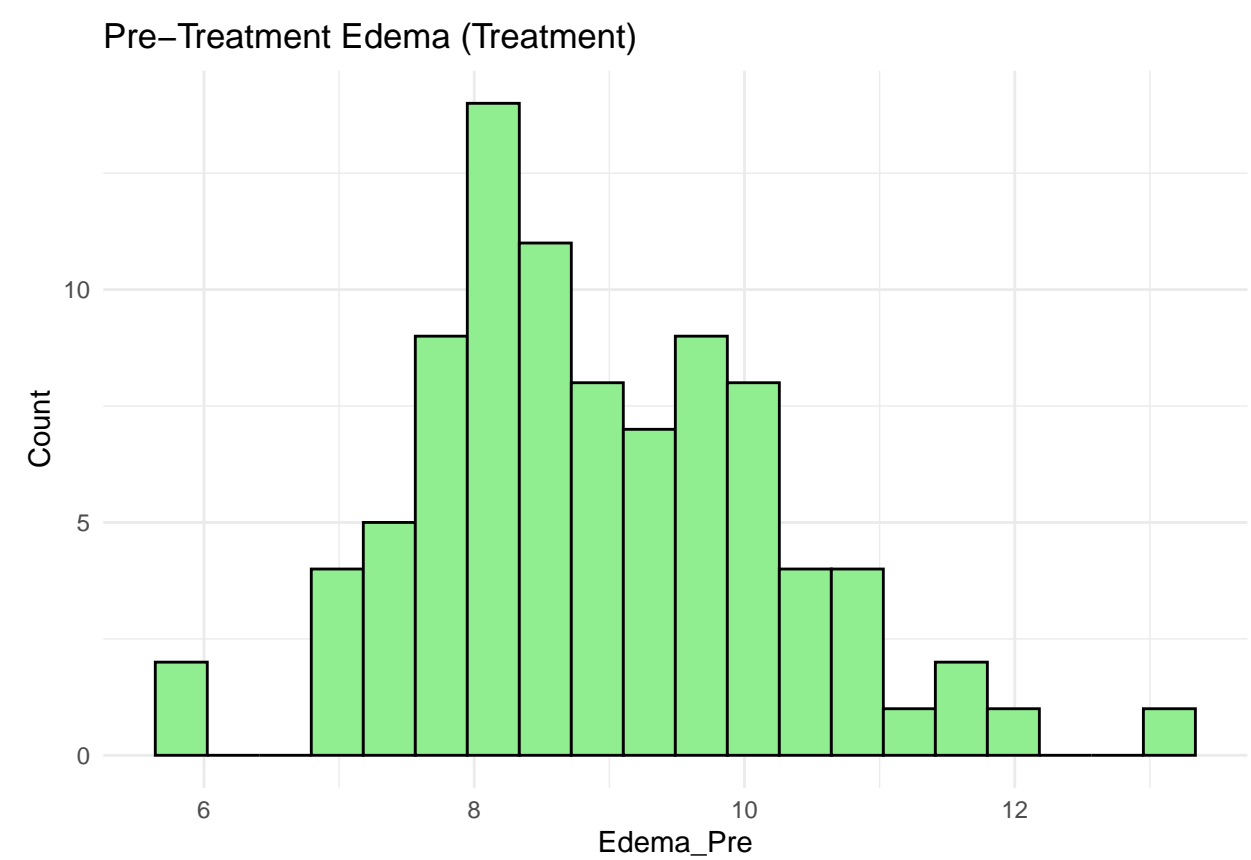


```
# Post - Placebo
plot_histogram(data[data$Group == "Placebo", ], "Edema_Post", "Placebo", "#ADD8E6", "Post-Treatment Ede
```

Post−Treatment Edema (Placebo)


```r
# Post - Treatment
plot_histogram(data[data$Group == "Treatment", ], "Edema_Post", "Treatment", "#90EE90", "Post-Treatment
```

Post−Treatment Edema (Treatment)


```r
# Delta - Placebo
plot_histogram(data[data$Group == "Placebo", ], "Delta", "Placebo", "#ADD8E6", "Delta")
```

Delta (Placebo)
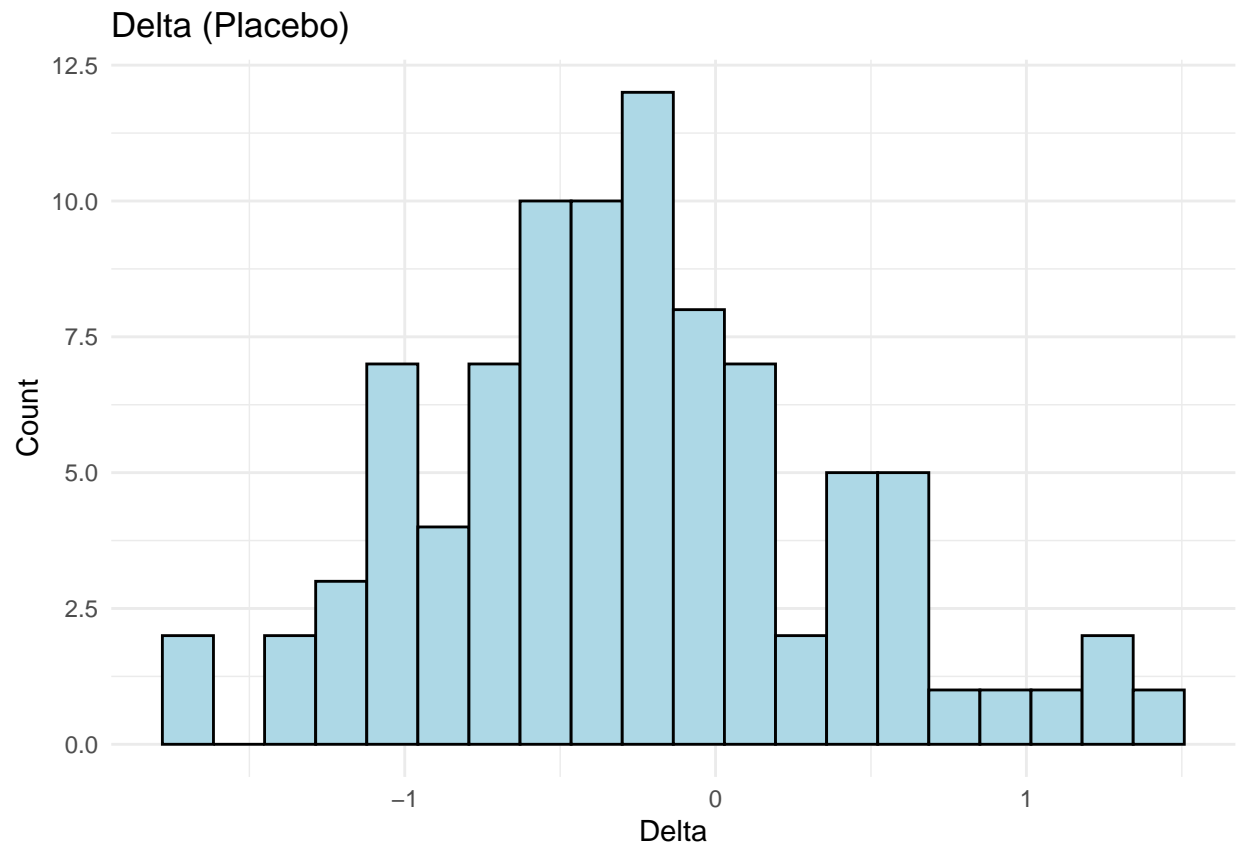


```r
# Delta - Treatment
plot_histogram(data[data$Group == "Treatment", ], "Delta", "Treatment", "#90EE90", "Delta")
```

Delta (Treatment)
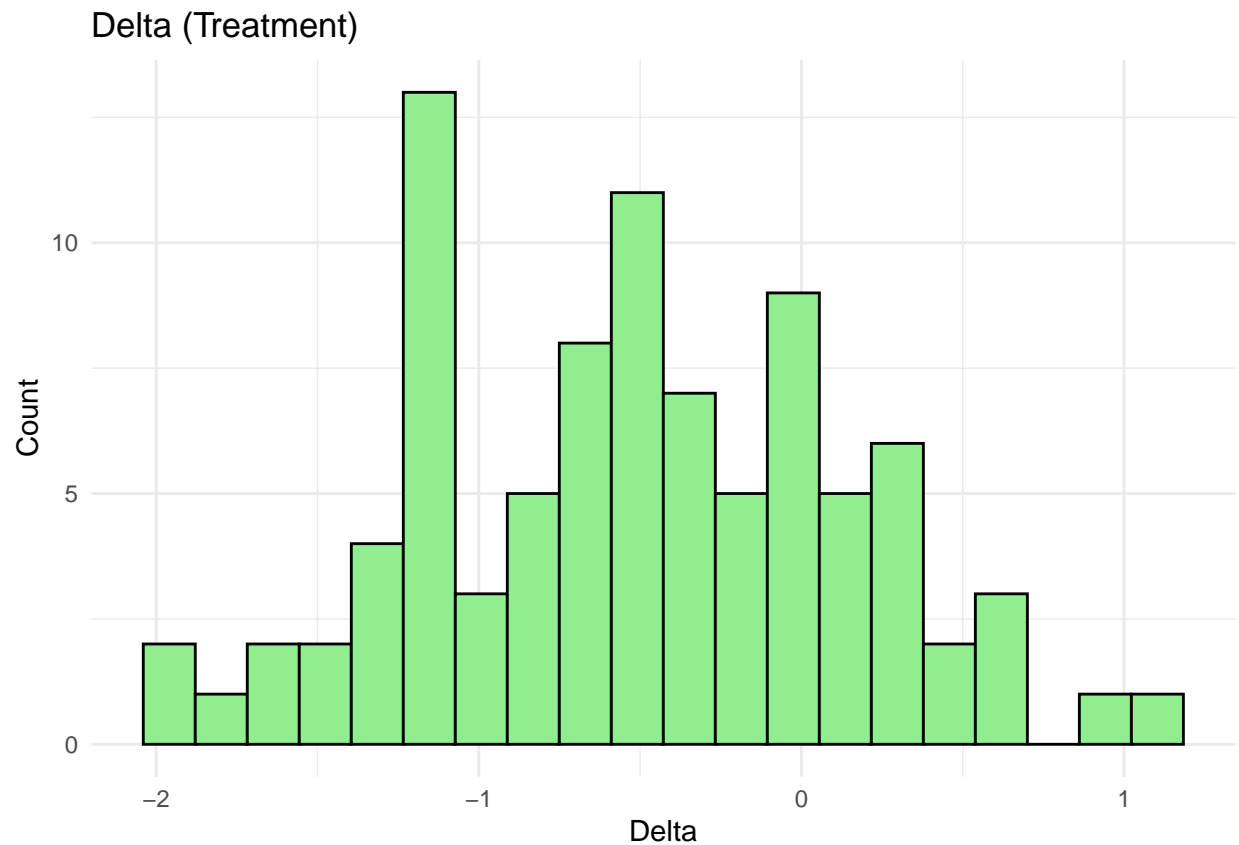

```
# Scatterplot of Post vs Pre
ggplot(data, aes(x = Edema_Pre, y = Edema_Post, color = Group)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE) +
  theme_minimal() +
  labs(title = "Scatterplot of Post vs Pre Edema by Group", x = "Edema_Pre", y = "Edema_Post") +
  scale_color_manual(values = c("Placebo" = "#ADD8E6", "Treatment" = "#90EE90"))
```
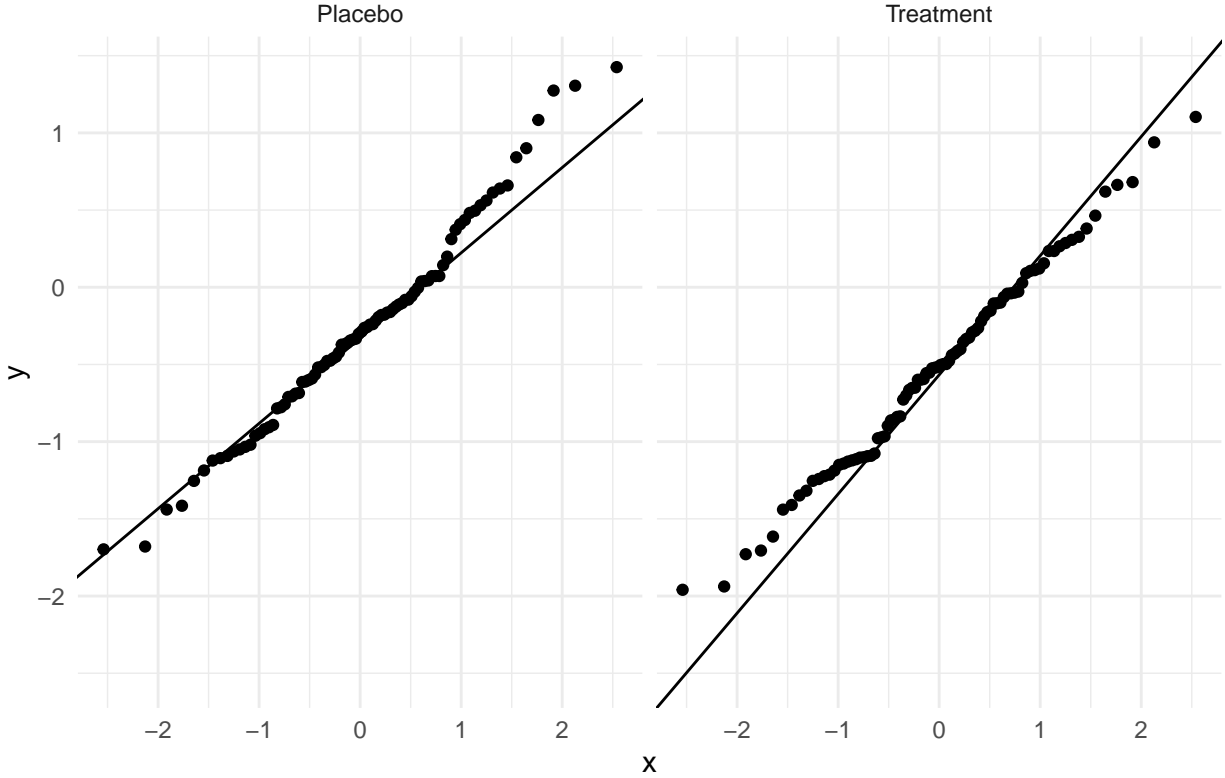
```
## 'geom_smooth()' using formula = 'y ~ x'
```

Scatterplot of Post vs Pre Edema by Group
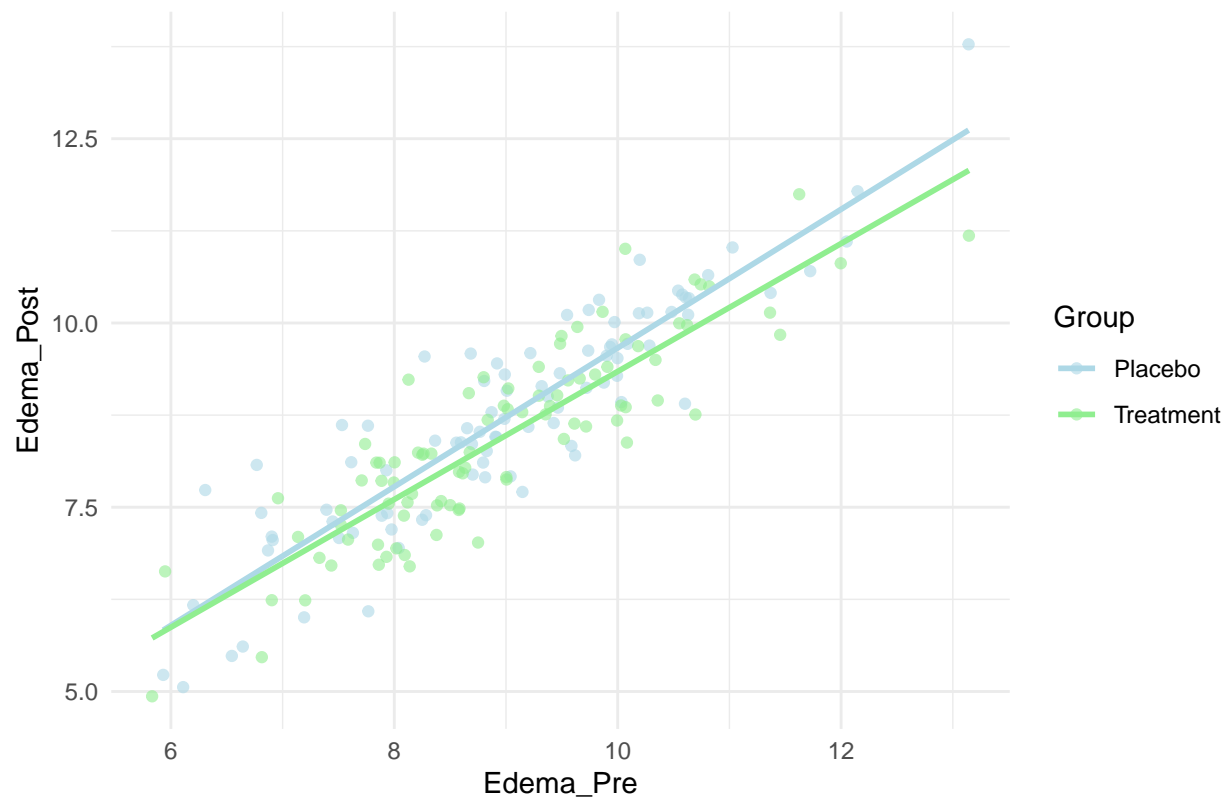

```r
# QQ plots for Delta
ggplot(data, aes(sample = Delta)) +
  stat_qq() +
  stat_qq_line() +
  facet_wrap(~ Group) +
  theme_minimal() +
  labs(title = "QQ Plots for Delta by Group")
```

QQ Plots for Delta by Group

Placebo

Treatment

```r
# Shapiro-Wilk tests
shapiro_placebo <- shapiro.test(data$Delta[data$Group == "Placebo"])
shapiro_treatment <- shapiro.test(data$Delta[data$Group == "Treatment"])

normality_results <- data.frame(
  Group = c("Placebo", "Treatment"),
  W = c(shapiro_placebo$statistic, shapiro_treatment$statistic),
  p_value = c(shapiro_placebo$p.value, shapiro_treatment$p.value)
)

kable(normality_results, digits = 3, caption = "Shapiro-Wilk Normality Tests for Delta")
```

Table 5: Shapiro-Wilk Normality Tests for Delta

Group	W	p_value
Placebo	0.984	0.351
Treatment	0.992	0.859