Statistical Report

Gishor Thavakumar(48032875)

2023-11-01

Abstract

This study analyzes a dataset of 274 weightlifters, exploring variables such as gender, body weight, age, and maximum weight lifted. The research questions focus on the difference in average age between female and male weightlifters and the relationship between body weight and weight lifted, accounting for gender distinctions. The dataset, although simulated, represents a significant sample. Hypothesis testing is used to compare the mean ages of male and female weightlifters in order to answer the first question. In order to validate the relevant assumptions, the normality of the age distributions for both genders is evaluated. In order to answer the second question, the relationship between weight lifted and body weight is examined while accounting for gender influence. To find out if separate regression analyses for men and women are required, assumptions are verified. The answers to these analyses will definitively answer this query. Following established guidelines, this report presents the results, methods, and interpretations without using pronouns. The study highlights important factors influencing weightlifting performance and offers clear insights into age differences and the relationship between body weight and weight lifted through rigorous statistical analysis using R.

Introduction

Research Question

Question 1: Is there any difference in the average age of female and male weightlifters?
Question 2: What is the relation between the body weight of weightlifters and the weight lifted?

Methods

In this study, statistical methods in R were used to analyze the weightlifting data. Different groups were compared using tests, and graphs were created to visually understand the relationships between variables. The data were carefully examined to answer the research questions accurately, leading to meaningful conclusions about the weightlifters' characteristics and performance. Below are the statistical methods used:

- **Hypothesis Testing (t-test):** Utilized to compare the mean ages of female and male weightlifters, investigating differences between the two groups.
- **Exploratory Data Analysis:** Employed graphical representations such as histograms, bar charts, and boxplots to visually explore the distribution and relationships among variables, providing insights into the dataset.
- **Regression Analysis:** Conducted regression analyses to understand the relationships between body weight and weight lifted, accounting for gender differences. Separate regressions were performed for females and males when necessary to ensure accurate interpretations.

Methods

Experimental Design

The study utilized a cross-sectional design, examining a sample of 274 weightlifters. This design was appropriate as it allowed for the collection of data at a single point in time, providing a snapshot of weightlifting performance among diverse individuals.

Subjects

Subjects were selected from a random sample of weightlifters, ensuring a representative mix of both genders. The inclusion criteria encompassed individuals actively participating in weightlifting activities, leading to a well-rounded and diverse sample.

Variables

- **ID:** Subject ID served as a unique identifier for each participant.
- **Gender:** Categorized participants as either "female" or "male."
- **Bodyweight:** Recorded the weight of each subject.
- **Age:** Captured the age of the participants in years.
- **Weightlifted:** Measured the maximum weight lifted by each

Preliminary Data Exploration

Data Import and Exploration

```
# Read the data into R
weightlifters_data <- read.csv("48032875_data_StatReport.csv")

# Explore the dataset
head(weightlifters_data)  # Display the first few rows of the dataset
```

```
##      ID gender bodyweight   age weightlifted
## 1 subj1 female       96.4 22.49        116.0
## 2 subj2 female       98.4 31.22        120.9
## 3 subj3   male      111.7 40.11        181.6
## 4 subj4   male      123.3 34.52        183.1
## 5 subj5   male      114.9 23.63        173.3
## 6 subj6 female       84.8 31.00        114.3
```

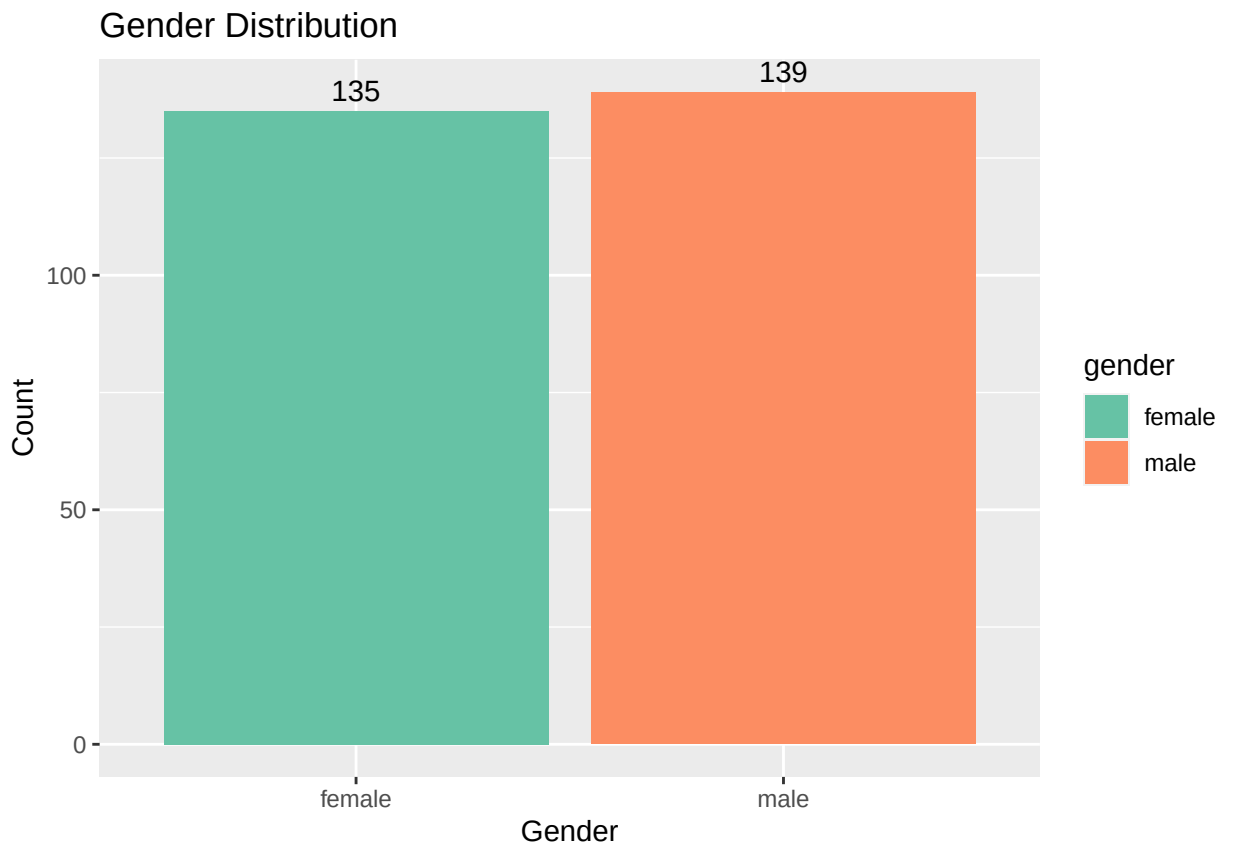
```
summary(weightlifters_data)  # Summary statistics for numerical variables
```

```
##       ID              gender           bodyweight          age
##  Length:274         Length:274         Min.   : 60.00   Min.   :14.15
##  Class :character   Class :character   1st Qu.: 86.42   1st Qu.:24.25
##  Mode  :character   Mode  :character   Median : 99.55   Median :28.24
##                                        Mean   : 98.15   Mean   :27.73
##                                        3rd Qu.:111.58   3rd Qu.:30.89
```

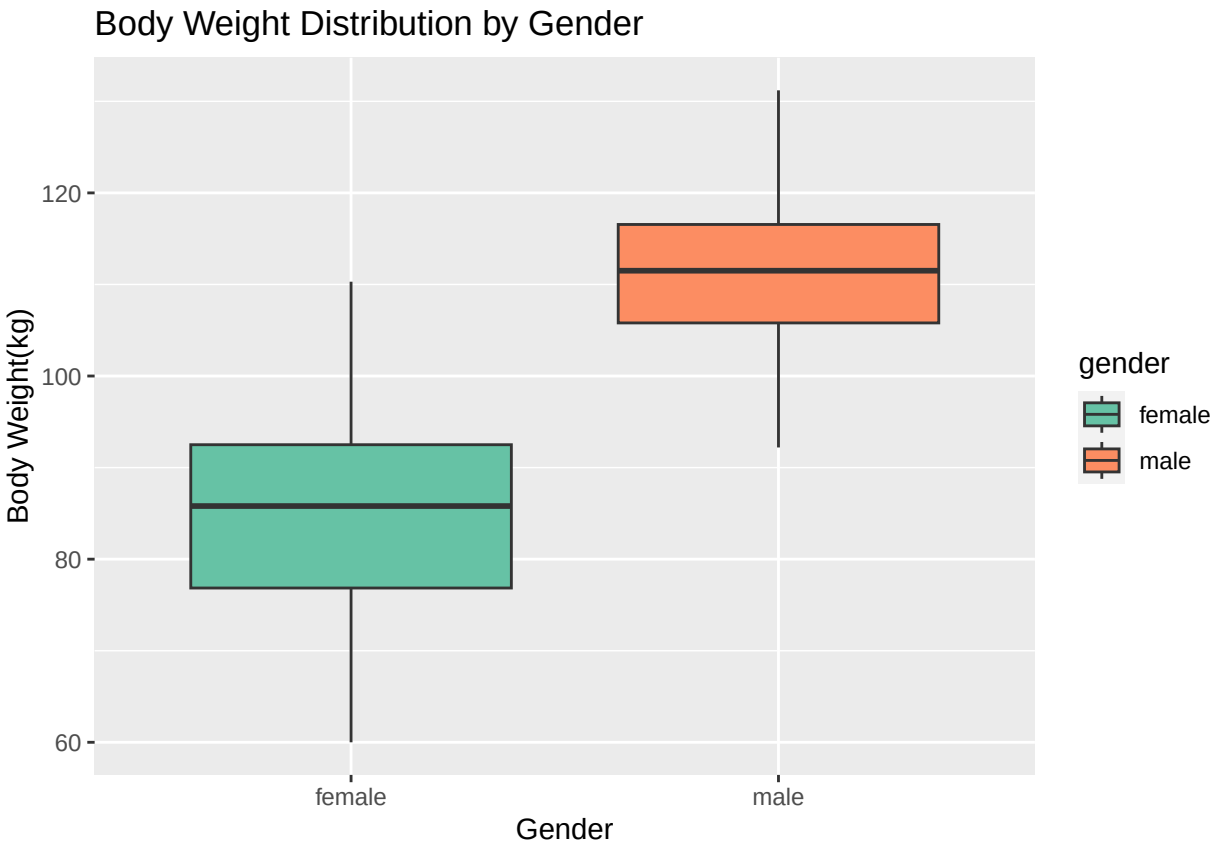
```
##                                     Max.   :131.20   Max.   :42.21
##   weightlifted
##  Min.   :102.9
##  1st Qu.:115.4
##  Median :162.8
##  Mean   :147.0
##  3rd Qu.:178.5
##  Max.   :195.6
```

Frequency Distributions

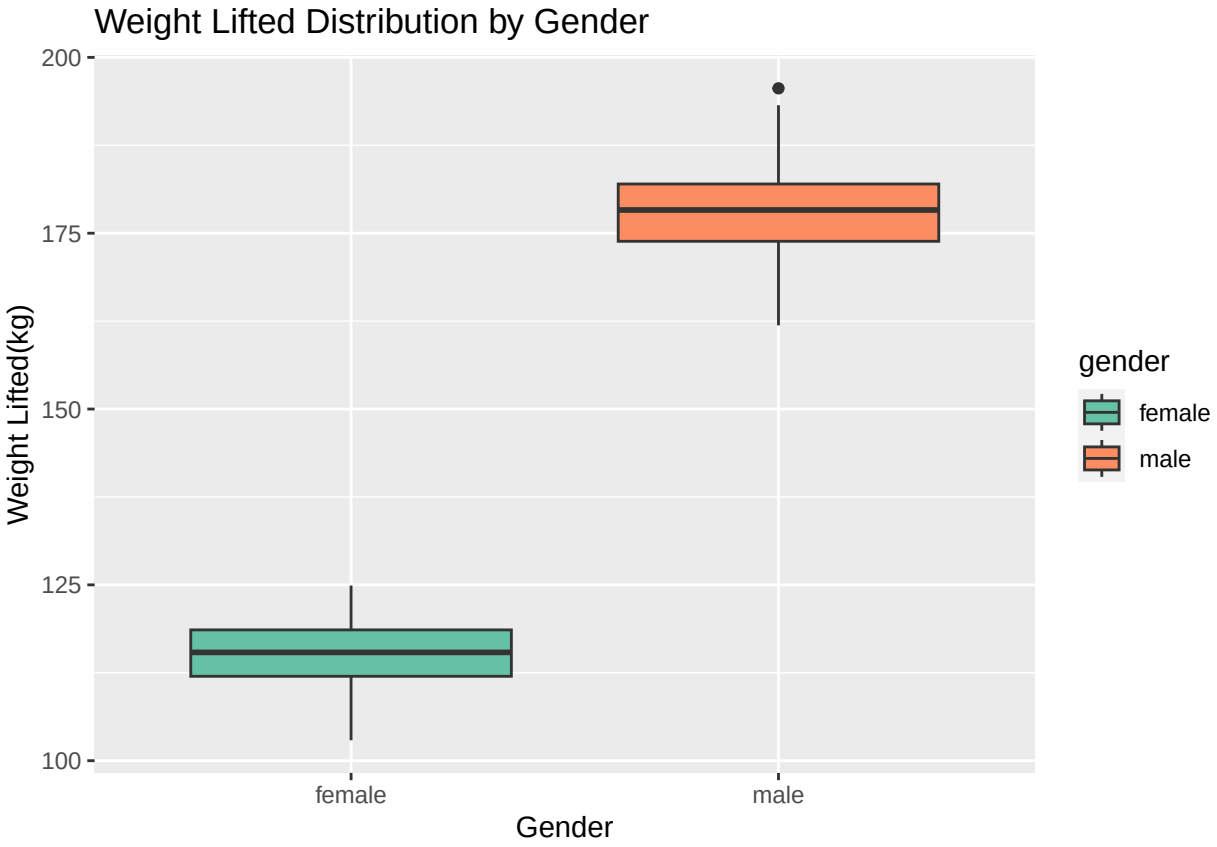
```r
# Load necessary libraries
library(ggplot2)

# Assuming your dataset is named 'data', replace it with your actual dataset name

# Bar chart for gender distribution
bar_gender <- ggplot(weightlifters_data, aes(x = gender, fill = gender)) +
  geom_bar() +
  geom_text(stat = "count", aes(label = after_stat(count)), vjust = -0.5) +  # Display count inside bar
  labs(title = "Gender Distribution", x = "Gender", y = "Count") +
  scale_fill_manual(values = c("#66c2a5", "#fc8d62"))  # Green for female, orange for male


boxplot_bodyweight <- ggplot(weightlifters_data, aes(x = gender, y = bodyweight, fill = gender)) +
  geom_boxplot() +
  labs(title = "Body Weight Distribution by Gender", x = "Gender", y = "Body Weight(kg)") +
  scale_fill_manual(values = c("#66c2a5", "#fc8d62"))

boxplot_weightlifted <- ggplot(weightlifters_data, aes(x = gender, y = weightlifted, fill = gender)) +
  geom_boxplot() +
  labs(title = "Weight Lifted Distribution by Gender", x = "Gender", y = "Weight Lifted(kg)") +
  scale_fill_manual(values = c("#66c2a5", "#fc8d62"))

# Save the plots as separate files
ggsave("bar_gender.png", plot = bar_gender, width = 6, height = 4)
ggsave("boxplot_bodyweight.png", plot = boxplot_bodyweight, width = 6, height = 4)
ggsave("boxplot_weightlifted.png", plot = boxplot_weightlifted, width = 6, height = 4)

# To display the plots in R Markdown, simply call the variables where the plots are stored
bar_gender
```




```
boxplot_bodyweight
```




```
boxplot_weightlifted
```



Summaries for the plots

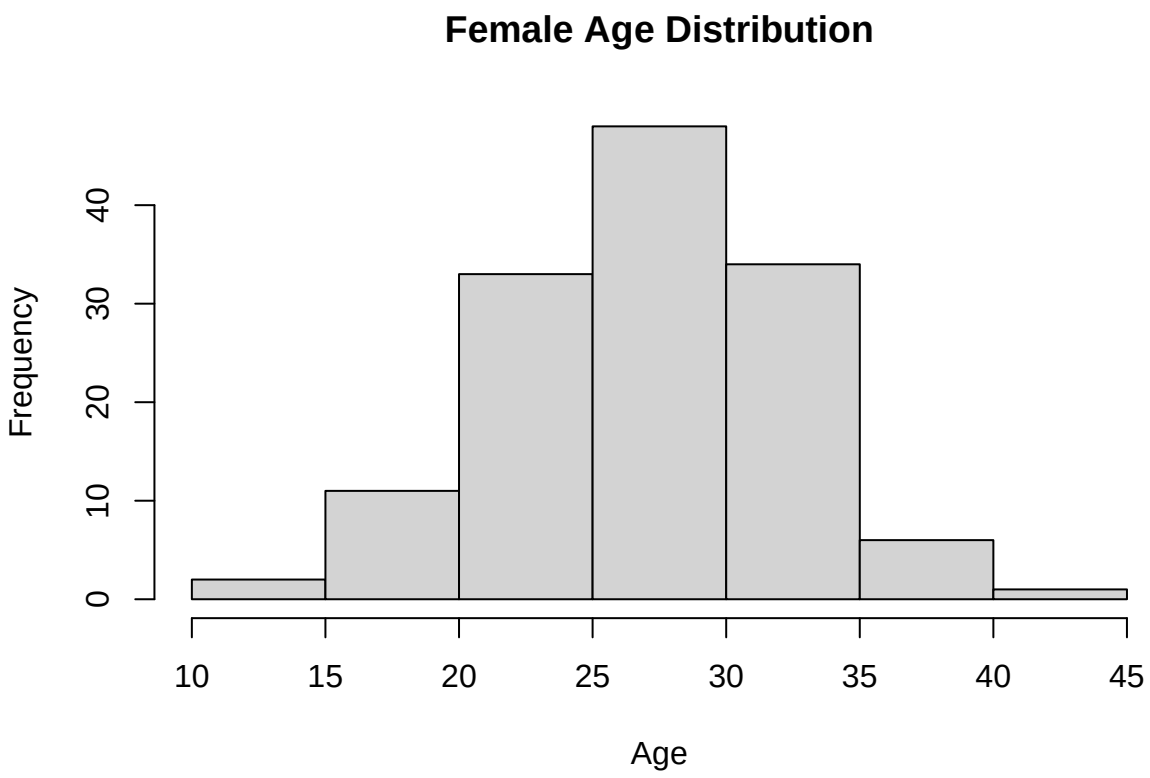
In gender comparison, an equal gender balance is observed with 135 females and 139 males. Females' median body weight slightly exceeds 85 kg, while males have a median above 110 kg. In terms of weight lifted, males dominate with a median of 179 kg, including an outlier, whereas females' median is 114 kg.

Analyses

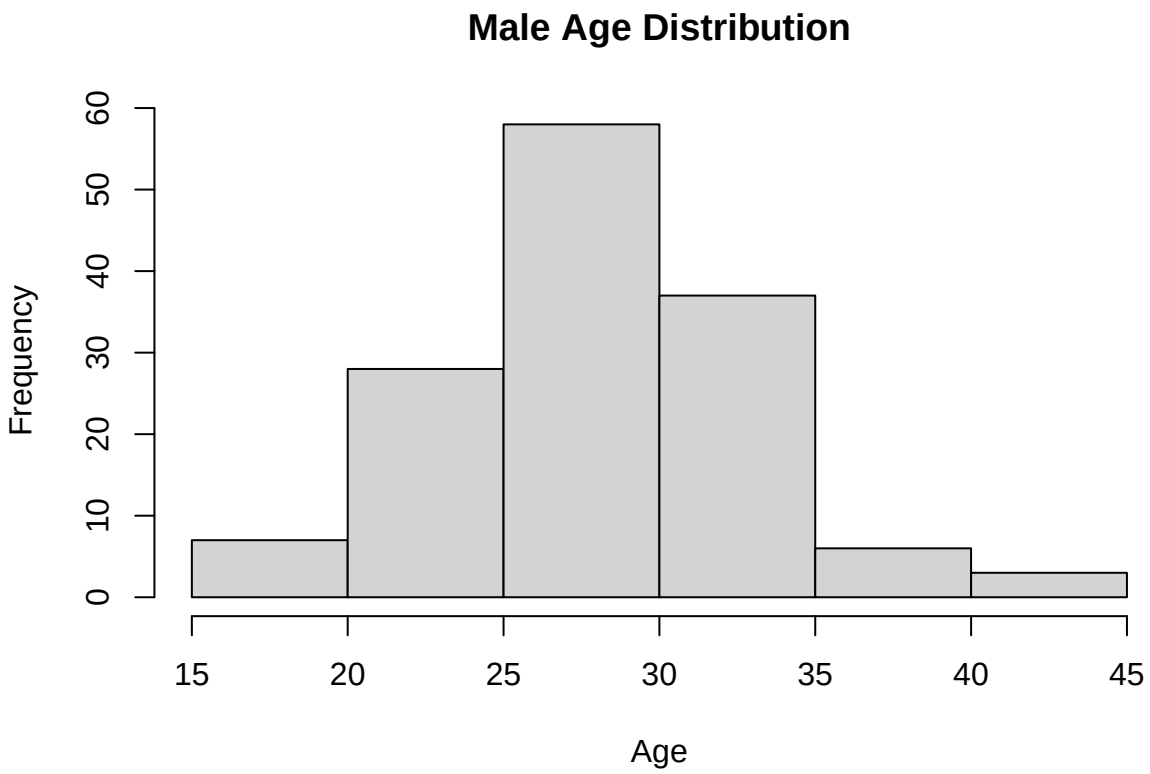
Address Research Questions:

Question (a): Is there any difference in the average age of female and male weightlifters?

```
# Check assumptions and conduct t-test
hist(weightlifters_data$age[weightlifters_data$gender == "female"], main = "Female Age Distribution", xl
```




```
hist(weightlifters_data$age[weightlifters_data$gender == "male"], main = "Male Age Distribution", xlab =
```




```
t_test_result <- t.test(age ~ gender, data = weightlifters_data)
t_test_result
```

```
##
##  Welch Two Sample t-test
##
## data:  age by gender
## t = -1.5229, df = 269.85, p-value = 0.1289
## alternative hypothesis: true difference in means between group female and group male is not equal to
## 95 percent confidence interval:
##  -2.1613822  0.2759913
## sample estimates:
## mean in group female   mean in group male
##             27.25133             28.19403
```

Based on the Sample t-test comparing the average age of female and male weightlifters, the results reveal a t-statistic of approximately -1.5229 with a corresponding p-value of 0.1289. With a degrees of freedom (df) of 269.85, the test does not provide strong evidence to support a significant difference in the means of age between female and male weightlifters at the conventional alpha level of 0.05.

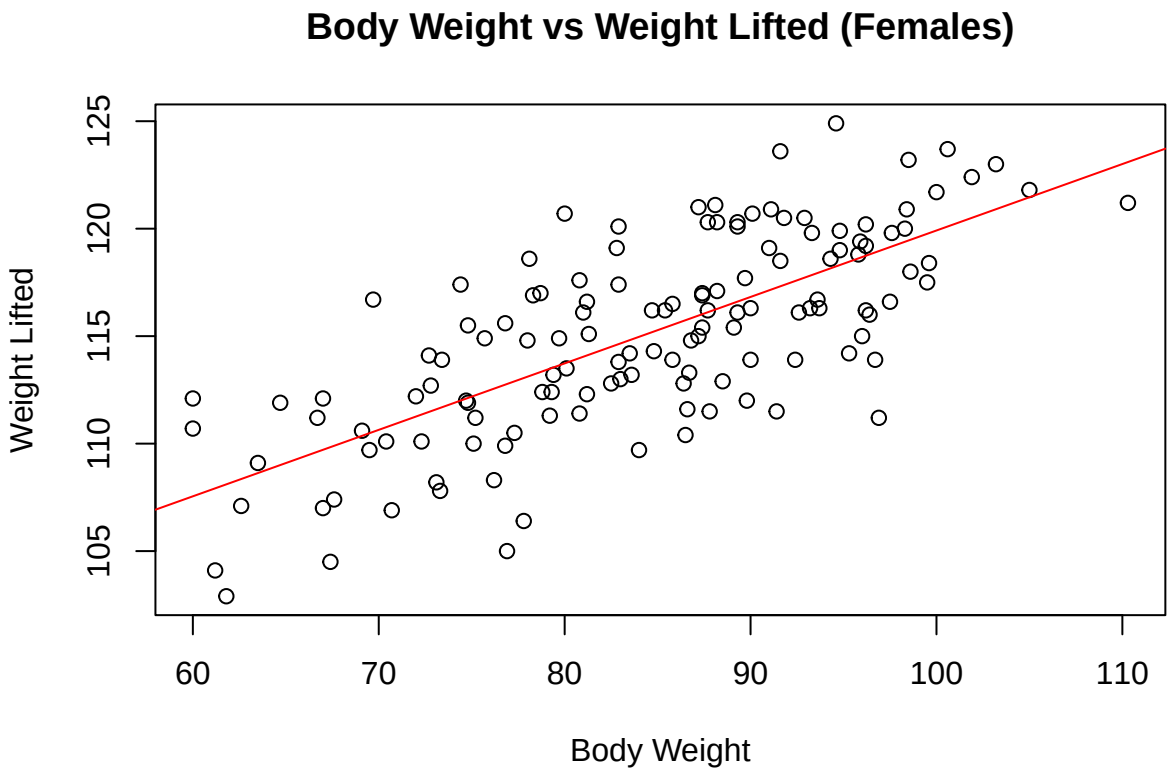
The 95% confidence interval for the difference in means spans from approximately -2.16 to 0.28, indicating that the true difference in average age between female and male weightlifters is likely to fall within this range.

While there is a slight difference in the mean ages between female (27.25 years) and male (28.19 years) weightlifters, the p-value of 0.1289 suggests that this difference is not statistically significant at the conventional significance level of 0.05. In other words, the observed difference in means could likely be due to

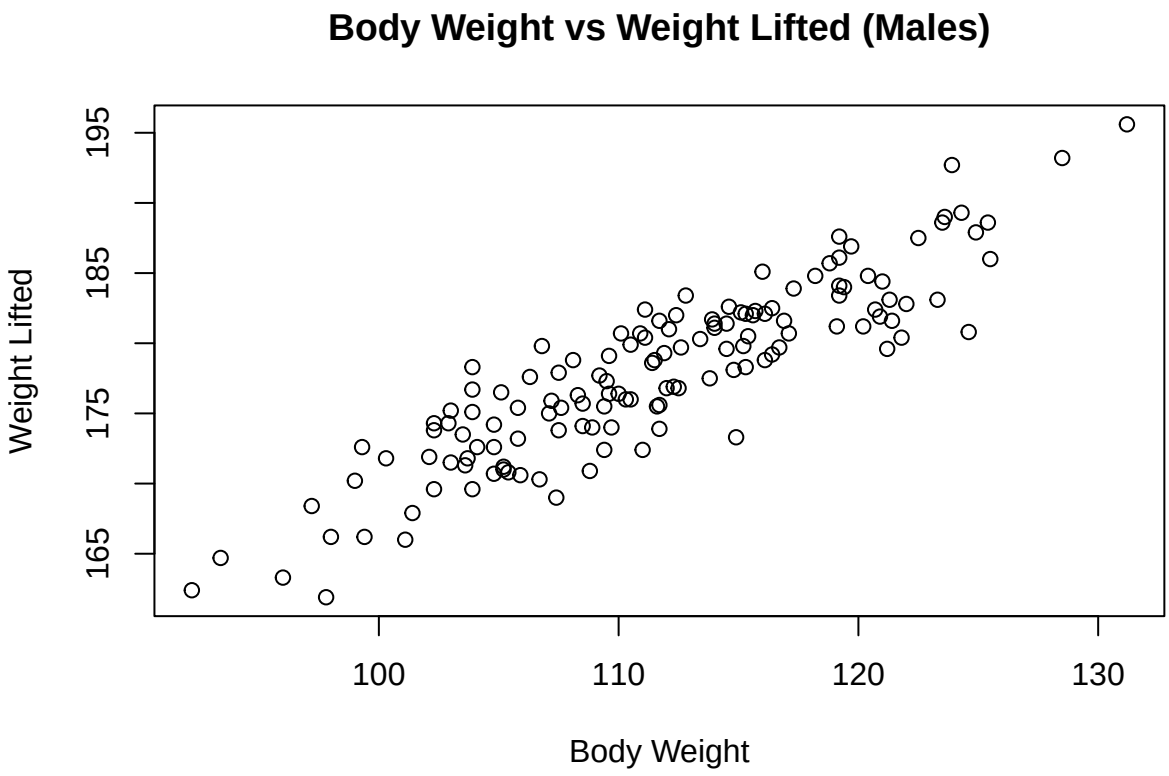
random chance, and there is no strong evidence to conclude that the average age differs significantly between female and male weightlifters based on the given data.

Question (b): What is the relation between the body weight of weightlifters and the weight lifted?

```
# Check assumptions and perform regression analysis

# Scatter plot for females
females_data <- weightlifters_data[weightlifters_data$gender == "female",]
plot(females_data$bodyweight, females_data$weightlifted,
     xlab = "Body Weight", ylab = "Weight Lifted",
     main = "Body Weight vs Weight Lifted (Females)")

# Add a linear trendline
abline(lm(weightlifted ~ bodyweight, data = females_data), col = "red")
```




```
# Scatter plot for males
males_data <- weightlifters_data[weightlifters_data$gender == "male",]
plot(males_data$bodyweight, males_data$weightlifted,
     xlab = "Body Weight", ylab = "Weight Lifted",
     main = "Body Weight vs Weight Lifted (Males)")
```




```r
# Check if separate regressions are needed based on gender
lm_female <- lm(weightlifted ~ bodyweight, data = weightlifters_data[weightlifters_data$gender == "fema
lm_male <- lm(weightlifted ~ bodyweight, data = weightlifters_data[weightlifters_data$gender == "male",
summary(lm_female)  # Summary of regression for females
```

```
##
## Call:
## lm(formula = weightlifted ~ bodyweight, data = weightlifters_data[weightlifters_data$gender ==
##     "female", ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.7738 -2.2011  0.0871  2.1516  6.9678
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 88.99778    2.16117   41.18   <2e-16 ***
## bodyweight   0.30918    0.02542   12.16   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.143 on 133 degrees of freedom
## Multiple R-squared:  0.5266, Adjusted R-squared:  0.523
## F-statistic: 147.9 on 1 and 133 DF,  p-value: < 2.2e-16
```

```r
summary(lm_male)    # Summary of regression for males
```

```
##
## Call:
## lm(formula = weightlifted ~ bodyweight, data = weightlifters_data[weightlifters_data$gender ==
##     "male", ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.1083 -2.1380  0.2952  2.0150  5.9497
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  96.2387     3.4101   28.22   <2e-16 ***
## bodyweight    0.7326     0.0305   24.02   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.752 on 137 degrees of freedom
## Multiple R-squared:  0.8081, Adjusted R-squared:  0.8067
## F-statistic: 576.9 on 1 and 137 DF,  p-value: < 2.2e-16
```

Based on the linear regression analysis conducted separately for females and males, there is a positive relationship between the body weight of weightlifters and the weight lifted. Specifically, as the body weight increases, the weight lifted also tends to increase for both females and males. This relationship is indicated by the positive coefficients of the bodyweight variable in the regression equations.

For males, the close proximity of scatter points to the rising trendline indicates a relatively strong and consistent positive linear relationship between body weight and weight lifted. This suggests that, on average, as the body weight of male weightlifters increases, the weight they can lift also tends to increase, and this relationship is consistent across individuals.

For females, although there is a rising trend, the larger spaces between scatter points and the trendline suggest more variability in the relationship. This means that while there is a general positive trend, there are greater differences in weightlifting performance among females with similar body weights.

Conclusion

In conclusion, this comprehensive analysis of a dataset comprising 274 weightlifters sheds light on the intricate relationships between age, gender, body weight, and weight lifted in the context of weightlifting performance. The study revealed that, on average, there is no significant difference in the average age between female and male weightlifters, emphasizing the equality in age distribution among the studied population.

Furthermore, the study confirmed a positive relationship between body weight and weight lifted for both genders. Male weightlifters displayed a strong and consistent positive correlation, indicating that as their body weight increased, so did the weight they could lift. In contrast, while females also exhibited a positive trend, there was greater variability in weightlifting performance among individuals with similar body weights. This variability underscores the influence of additional factors beyond body weight, possibly encompassing training techniques, body composition, or individual strengths and weaknesses, in shaping weightlifting abilities among female.