31_question_6

Sudarshan Budhathoki

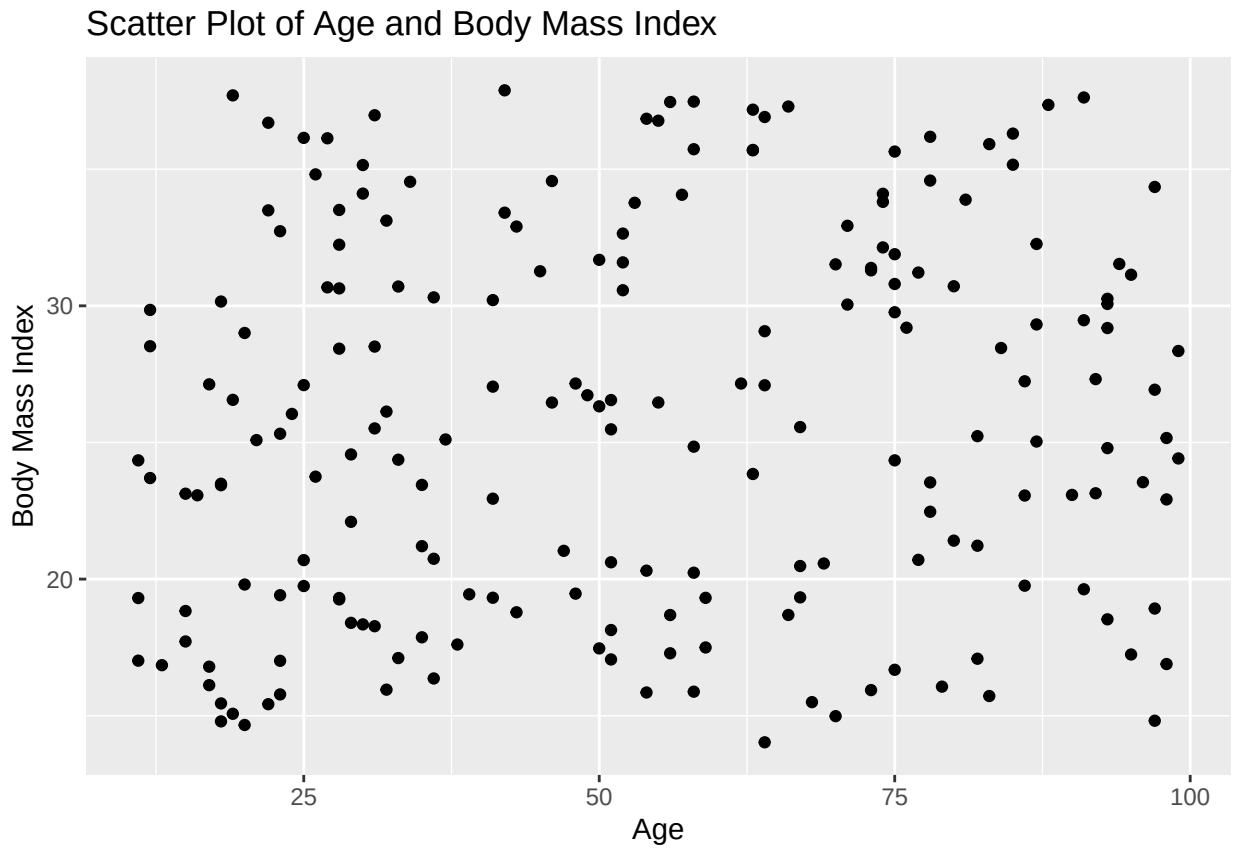
2024-05-31

```r
# Set random seed
set.seed(31)

# Create a dataset with random 200 cases for each variable
age <- sample(10:99, 200, replace = TRUE)
sex <- sample(c("male", "female"), 200, replace = TRUE)
edu_levels <- sample(c("No education", "Primary", "Secondary", "Beyond secondary"),
                     200, replace = TRUE)
soc_econ_status <- sample(c("Low", "Middle", "High"), 200, replace = TRUE)
body_mass_index <- runif(200, 14, 38)

data <- data.frame(age, sex, edu_levels, soc_econ_status, body_mass_index)
```

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

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

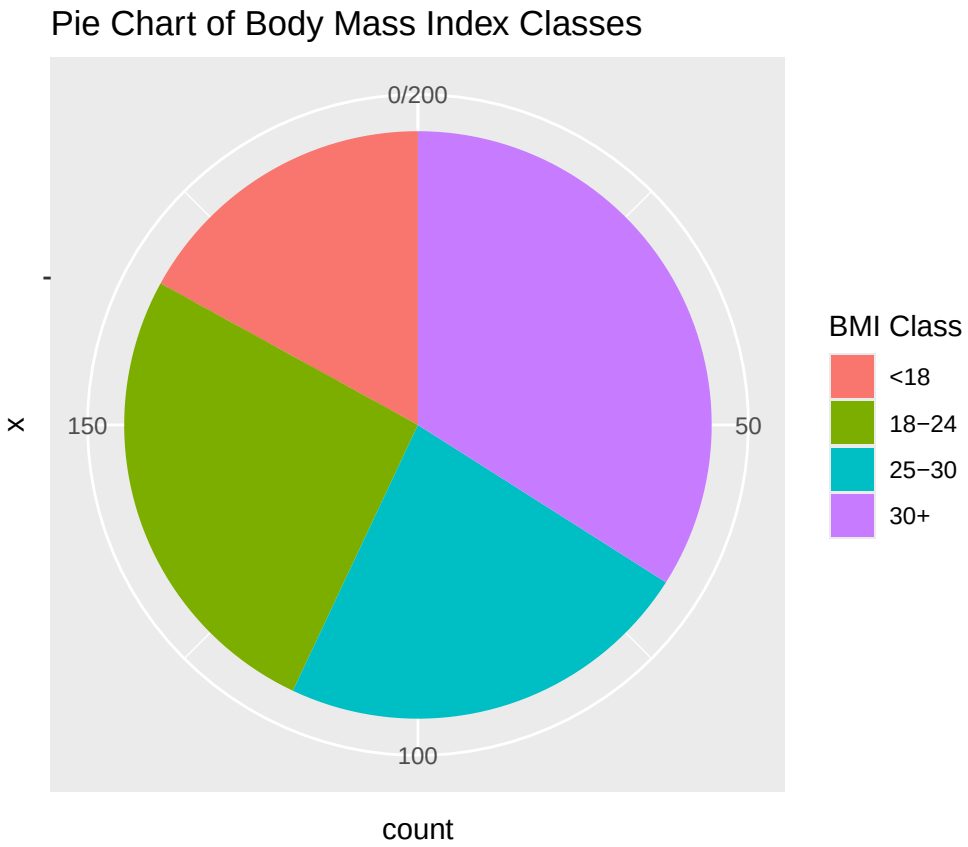
```r
# Create scatter plot of age and body mass index
ggplot(data, aes(x = age, y = body_mass_index)) +
  geom_point() +
  labs(title = "Scatter Plot of Age and Body Mass Index", x = "Age", y = "Body Mass Index")
```



*Interpretation: The scatter plot shows the relationship between age and body mass index. It can help identify any patterns or trends between the two variables.

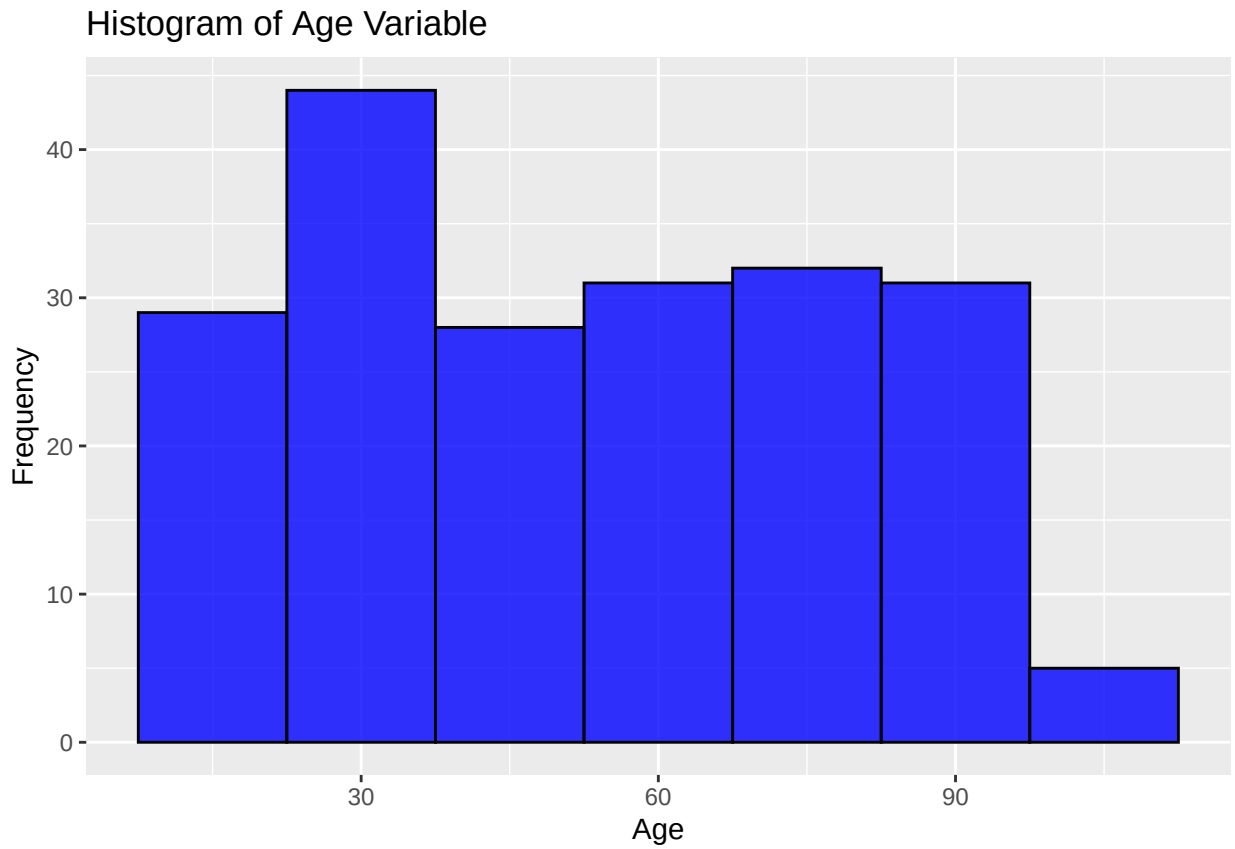
```r
# Create classes for body mass index
data$BMI_class <- cut(data$body_mass_index, breaks = c(0, 18, 24, 30, 100),
                      labels = c("<18", "18-24", "25-30", "30+"))
```

```r
# Create a pie chart of body mass index classes
ggplot(data, aes(x = "", fill = BMI_class)) +
  geom_bar(width = 1) +
  coord_polar("y") +
  labs(title = "Pie Chart of Body Mass Index Classes", fill = "BMI Class")
```



*Interpretation: The pie chart shows the distribution of body mass index classes among the 200 cases. It gives a visual representation of how the BMIs are distributed across the specified classes.

```
# Create histogram of age variable with bin size of 15
ggplot(data, aes(x = age)) +
  geom_histogram(binwidth = 15, fill = "blue", color = "black", alpha = 0.8) +
  labs(title = "Histogram of Age Variable", x = "Age", y = "Frequency")
```



*Interpretation: The histogram displays the distribution of ages in the dataset with bins of size 15.It allows us to visualize the frequency of ages in the specified bins.