

# Descriptive Analysis & Visualization of individuals' time spent on reading, watching TV, and self-reported happiness

Zemelak Goraga

2024-01-27

```
head(df)
```

```
##   TimeReading TimeTV Happiness Gender
## 1           1     90      86.20      1
## 2           2     95      88.70      0
## 3           2     85      70.17      0
## 4           2     80      61.31      1
## 5           3     75      89.52      1
## 6           4     70      60.50      1
```

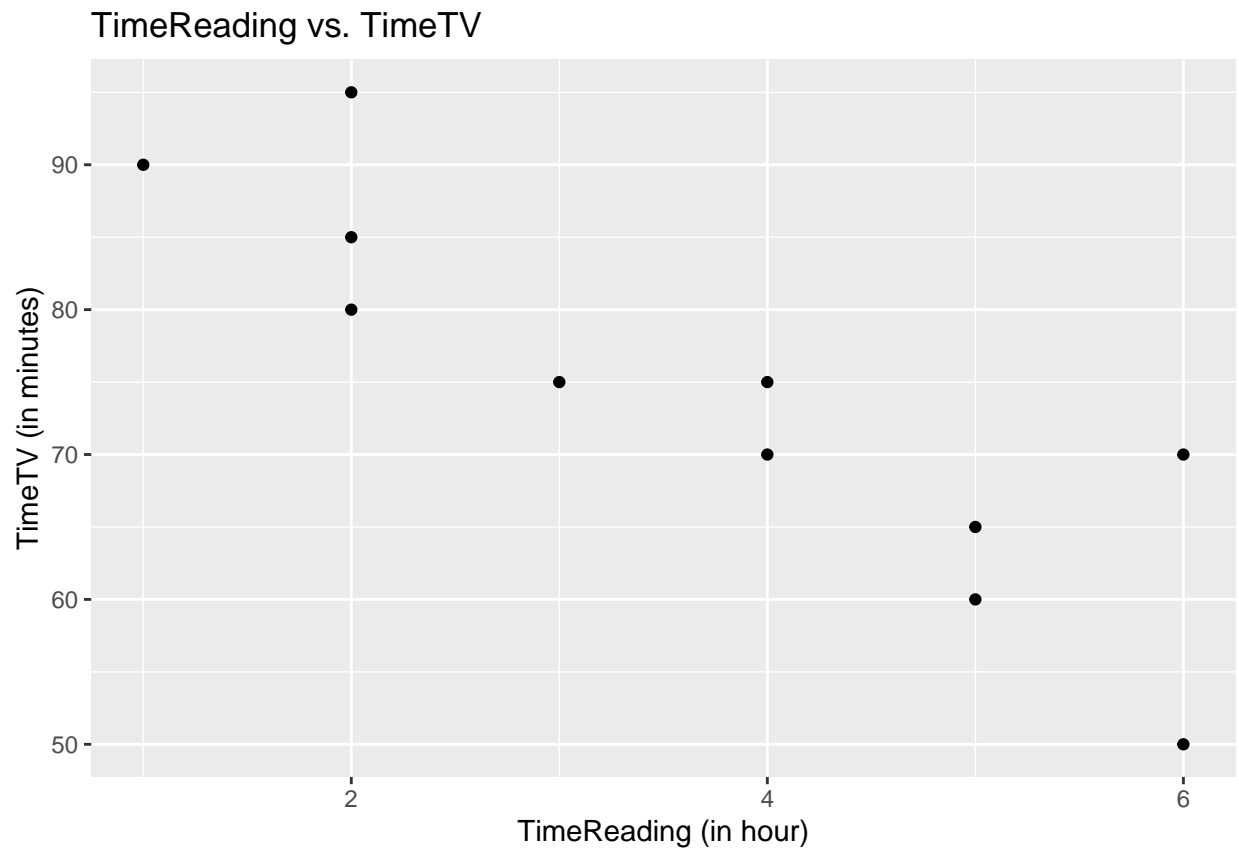
```
# Plots of Survey Variables
library(ggplot2)
```

**Question 1:** Create the following 3 plots of survey variables (first variable is x-axis, second is y-axis)

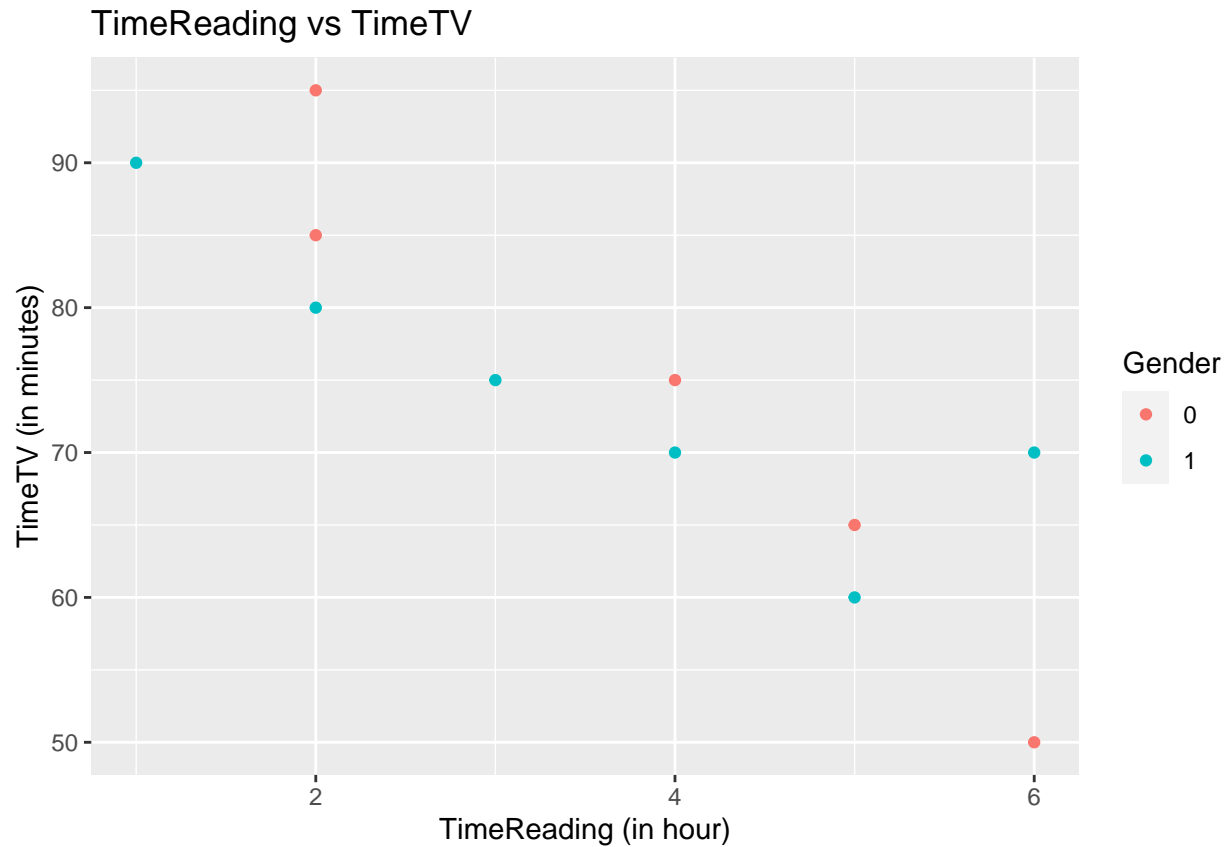
TimeReading, TimeTV TimeReading, Happiness TimeTV, Happiness

Based on the plots you created, provide a rough estimate of the relationship between the variables; do the slopes indicate a positive or negative relationship? Instead of the plots, use the descriptive statistics

```
# Plot TimeReading vs. TimeTV
plot1 <- ggplot(df, aes(x = TimeReading, y = TimeTV)) +
  geom_point() +
  labs(title = "TimeReading vs. TimeTV", x = "TimeReading (in hour)", y = "TimeTV (in minutes)")
print(plot1)
```



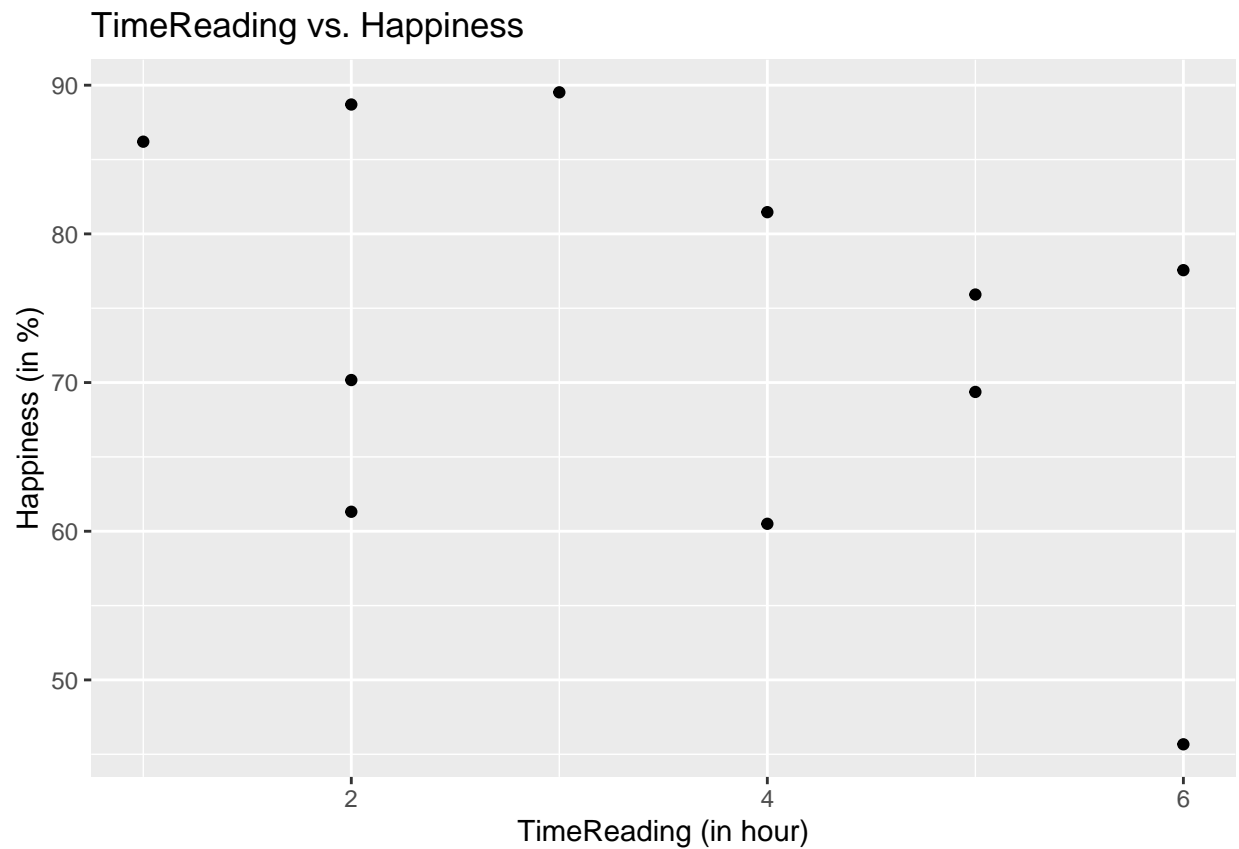
```
# Create plot1 using 'Gender' as a grouping variable
plot1 <- ggplot(df, aes(x = TimeReading, y = TimeTV, color = factor(Gender))) +
  geom_point() +
  labs(title = "TimeReading vs TimeTV",
        x = "TimeReading (in hour)",
        y = "TimeTV (in minutes)",
        color = "Gender")
print(plot1)
```



```
# Descriptive statistics for TimeReading and TimeTV based on Gender
summary(df[c("TimeReading", "TimeTV")])
```

```
##   TimeReading      TimeTV
##   Min.    :1.000   Min.    :50.00
##   1st Qu.:2.000   1st Qu.:67.50
##   Median :4.000   Median :75.00
##   Mean    :3.636   Mean    :74.09
##   3rd Qu.:5.000   3rd Qu.:82.50
##   Max.    :6.000   Max.    :95.00
```

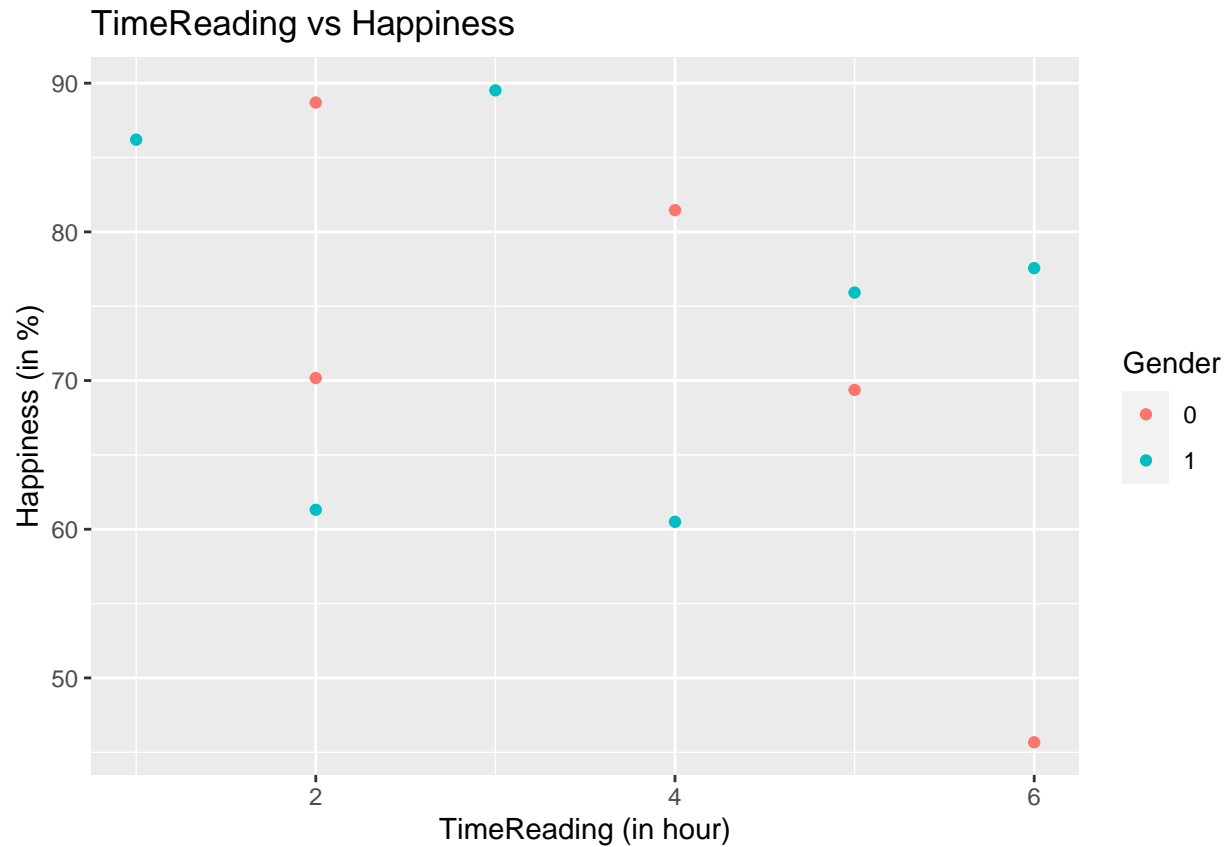
```
# Plot TimeReading vs. Happiness
plot2 <- ggplot(df, aes(x = TimeReading, y = Happiness)) +
  geom_point() +
  labs(title = "TimeReading vs. Happiness", x = "TimeReading (in hour)", y = "Happiness (in %)")
print(plot2)
```



```
# Create plot2 using 'Gender' as a grouping variable

plot2 <- ggplot(df, aes(x = TimeReading, y = Happiness, color = factor(Gender))) +
  geom_point() +
  labs(title = "TimeReading vs Happiness",
       x = "TimeReading (in hour)",
       y = "Happiness (in %)",
       color = "Gender")

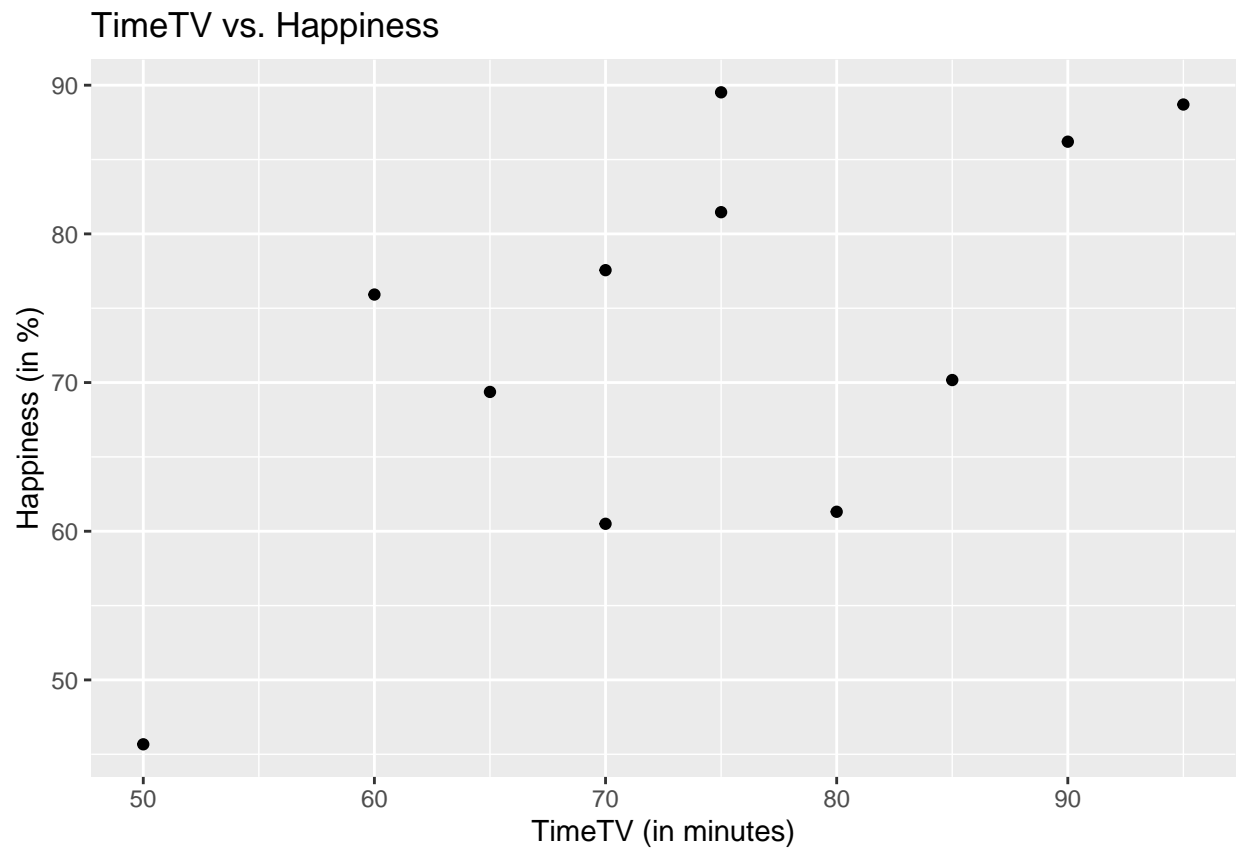
print(plot2)
```



```
# Descriptive statistics for TimeReading and Happiness based on Gender
summary(df[c("TimeReading", "Happiness")])
```

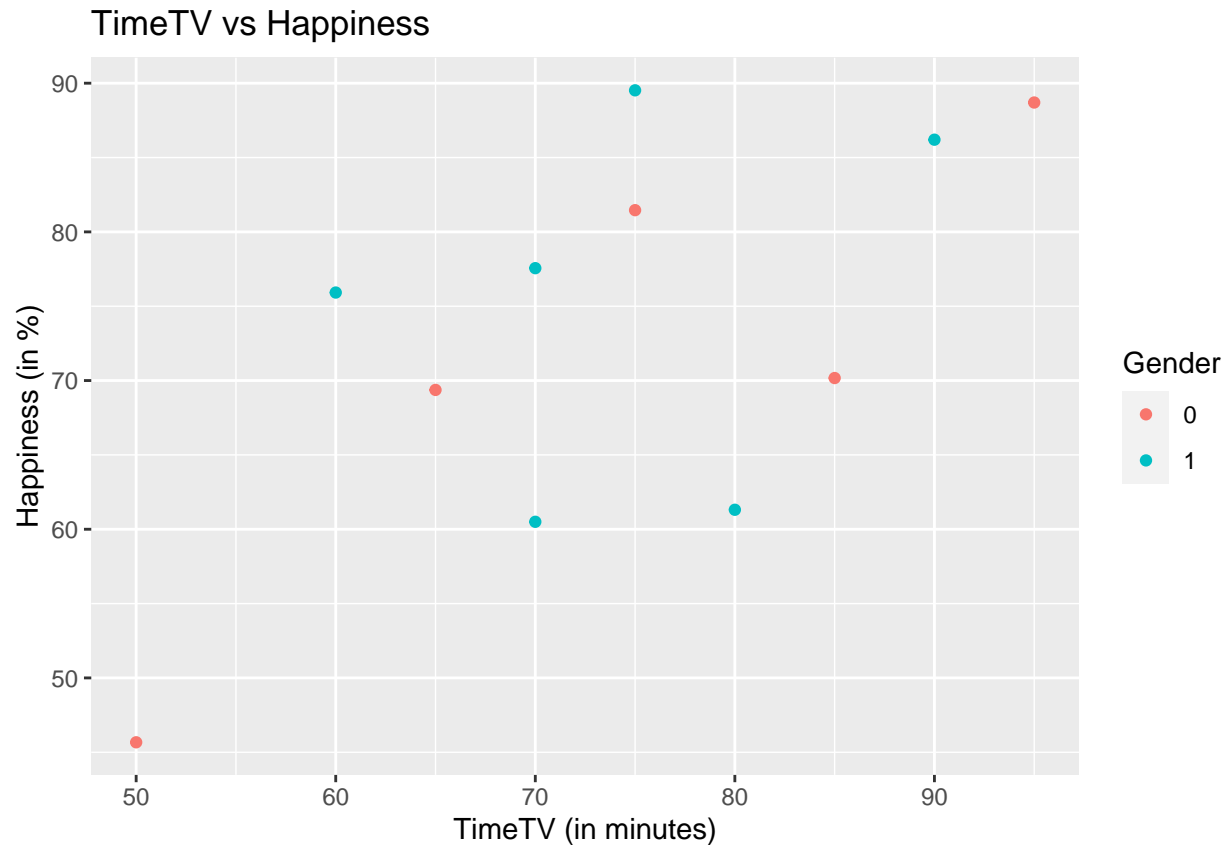
```
##   TimeReading      Happiness
##   Min.   :1.000   Min.   :45.67
##   1st Qu.:2.000   1st Qu.:65.34
##   Median :4.000   Median :75.92
##   Mean   :3.636   Mean   :73.31
##   3rd Qu.:5.000   3rd Qu.:83.83
##   Max.   :6.000   Max.   :89.52
```

```
# Plot TimeTV vs. Happiness
plot3 <- ggplot(df, aes(x = TimeTV, y = Happiness)) +
  geom_point() +
  labs(title = "TimeTV vs. Happiness", x = "TimeTV (in minutes)", y = "Happiness (in %)")
print(plot3)
```



*# Create plot3 using 'Gender' as a grouping variable*

```
plot3 <- ggplot(df, aes(x = TimeTV, y = Happiness, color = factor(Gender))) +  
  geom_point() +  
  labs(title = "TimeTV vs Happiness",  
        x = "TimeTV (in minutes)",  
        y = "Happiness (in %)",  
        color = "Gender")  
  
print(plot3)
```



```
# Descriptive statistics for TimeTV and Happiness based on Gender
summary(df[c("TimeTV", "Happiness")])
```

```
##      TimeTV      Happiness
##  Min.   :50.00  Min.   :45.67
##  1st Qu.:67.50  1st Qu.:65.34
##  Median :75.00  Median :75.92
##  Mean   :74.09  Mean   :73.31
##  3rd Qu.:82.50  3rd Qu.:83.83
##  Max.   :95.00  Max.   :89.52
```

```
# Save plots to a PDF file as survey_plots.pdf
pdf("survey_plots.pdf", width = 10, height = 8)
print(plot1)
print(plot2)
print(plot3)
dev.off()
```

```
## pdf
## 2
```

**Answer for Question 1: Both the ‘plots’ and the ‘descriptive statistics’ for TimeReading, TimeTV, and Happiness provided insights into the relationships between these variables:**

TimeReading and TimeTV:

Descriptive Statistics: Mean TimeReading: 3.636 Mean TimeTV: 74.09 Interpretation: On average, as TimeReading increases, there is a slight decrease in TimeTV, suggesting a weak negative relationship.

TimeReading and Happiness:

Descriptive Statistics: Mean TimeReading: 3.636 Mean Happiness: 73.31 Interpretation: On average, as TimeReading increases, there is a moderate decrease in Happiness, indicating a moderate negative relationship.

TimeTV and Happiness:

Descriptive Statistics: Mean TimeTV: 74.09 Mean Happiness: 73.31 Interpretation: On average, as TimeTV increases, there is a moderate increase in Happiness, suggesting a moderate positive relationship. Overall Interpretation: The plots and the descriptive statistics provided a rough estimate that supports the relationships observed in the correlation matrix.

**Question 2: Explain the relationship between the variables using the covariance matrix.**

```
# Relationship between Variables  
# Create a covariance matrix  
cov_matrix <- cov(df[, c("TimeReading", "TimeTV", "Happiness")])  
cov_matrix
```

```
##           TimeReading    TimeTV Happiness  
## TimeReading    3.054545 -20.36364 -10.35009  
## TimeTV        -20.363636 174.09091 114.37727  
## Happiness     -10.350091 114.37727 185.45142
```

**Answer for Question 2: The covariance matrix indicates the covariances between pairs of variables:**

Covariance(TimeReading, TimeTV) = -20.36364:

Interpretation: A negative covariance implies an inverse relationship; as TimeReading increases, TimeTV tends to decrease.

Covariance(TimeReading, Happiness) = -10.35009:

Interpretation: A negative covariance suggests that as TimeReading increases, Happiness tends to decrease.

Covariance(TimeTV, Happiness) = 114.37727:

Interpretation: A positive covariance implies that as TimeTV increases, Happiness tends to increase. Overall Interpretation: Covariance provides a directional indication but lacks the standardized nature of correlation.



**Question 3: Explain the relationship between the variables using the correlation matrix. Additionally, in your opinion, are the relationships between variables easier to interpret using the covariance or correlation matrix? Explain your answer.**

```
# Create a correlation matrix
cor_matrix <- cor(df[, c("TimeReading", "TimeTV", "Happiness")])
cor_matrix
```

```
##           TimeReading    TimeTV  Happiness
## TimeReading    1.0000000 -0.8830677 -0.4348663
## TimeTV         -0.8830677  1.0000000  0.6365560
## Happiness      -0.4348663  0.6365560  1.0000000
```

**Answer for Question 3: The correlation matrix provides correlation coefficients, which are standardized measures of the strength and direction of linear relationships:**

Correlation(TimeReading, TimeTV) = -0.8830677:

Interpretation: A strong negative correlation indicates that as TimeReading increases, TimeTV tends to decrease.

Correlation(TimeReading, Happiness) = -0.4348663:

Interpretation: A moderate negative correlation suggests that as TimeReading increases, Happiness tends to decrease.

Correlation(TimeTV, Happiness) = 0.6365560:

Interpretation: A moderate positive correlation implies that as TimeTV increases, Happiness tends to increase. Opinion: Correlation matrices are easier to interpret as they provide standardized measures, making comparisons between different variable pairs straightforward.

**Question 4: Perform a correlation test on TimeReading and TimeTV. What does the correlation value tell you about the relationship between them? Can you say that TimeReading has an effect on TimeTV?**

```
# Correlation Test
# Correlation test between TimeReading and TimeTV
cor_test_result <- cor.test(df$TimeReading, df$TimeTV)
cor_test_result
```

```
##
## Pearson's product-moment correlation
##
## data: df$TimeReading and df$TimeTV
## t = -5.6457, df = 9, p-value = 0.0003153
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.9694145 -0.6021920
```

```
## sample estimates:  
##      cor  
## -0.8830677
```

### **Answer for Question 4: correlation test on TimeReading and TimeTV:**

Correlation coefficient: -0.8830677 p-value: 0.0003153 (statistically significant) Interpretation: The strong negative correlation suggests that there is a significant inverse relationship between TimeReading and TimeTV. As the time spent on reading increases, the time spent on TV tends to decrease.

Conclusion: Based on the correlation value and statistical significance, it can be concluded that TimeReading has a significant negative effect on TimeTV.

### **Overall Summary of Results:**

The exercise did focus on unraveling the intricate connection between TimeReading, TimeTV, and Happiness through an in-depth analysis of sample survey data. By employing various statistical techniques, such as plotting, descriptive statistics, and exploration of covariance and correlation matrices, the exercise aimed to draw meaningful conclusions from the collected sample data.

The dataset utilized encapsulated survey responses featuring variables such as TimeReading (measured in hours), TimeTV (measured in minutes), Happiness (expressed in percentage), and Gender as a categorical variable. The analytical approach used in this exercise involved the calculation of descriptive statistics for TimeReading, TimeTV, and Happiness. Moreover, the analysis delved into covariance and correlation matrices to unveil the intricate relationships between these variables, with a specific focus on gender-based differences.

Analysis of the relationships between TimeReading, TimeTV, and Happiness has yielded intriguing findings. Notably, a robust negative correlation (-0.88) emerged between TimeReading and TimeTV, indicating a tendency for increased reading time to coincide with decreased TV time. Moreover, a moderate negative correlation (-0.43) between TimeReading and Happiness suggests a potential connection between elevated reading hours and lower reported happiness. Conversely, a moderate positive correlation (0.64) between TimeTV and Happiness implies that increased TV time aligns with higher reported happiness levels. Covariance and correlation matrices further illustrated these associations, with standardized correlation coefficients offering a clearer picture of the strength and direction of these relationships. The correlation test underscored the significance of the negative correlation between TimeReading and TimeTV, providing statistical support for the observed pattern. As we look ahead, gender-based variations in these relationships warrant exploration, along with qualitative inquiries to elucidate the nuanced factors influencing reported happiness and time allocation. In essence, our project has unveiled intricate dynamics, paving the way for deeper understanding and future investigations.