Assingment 7

Surenther

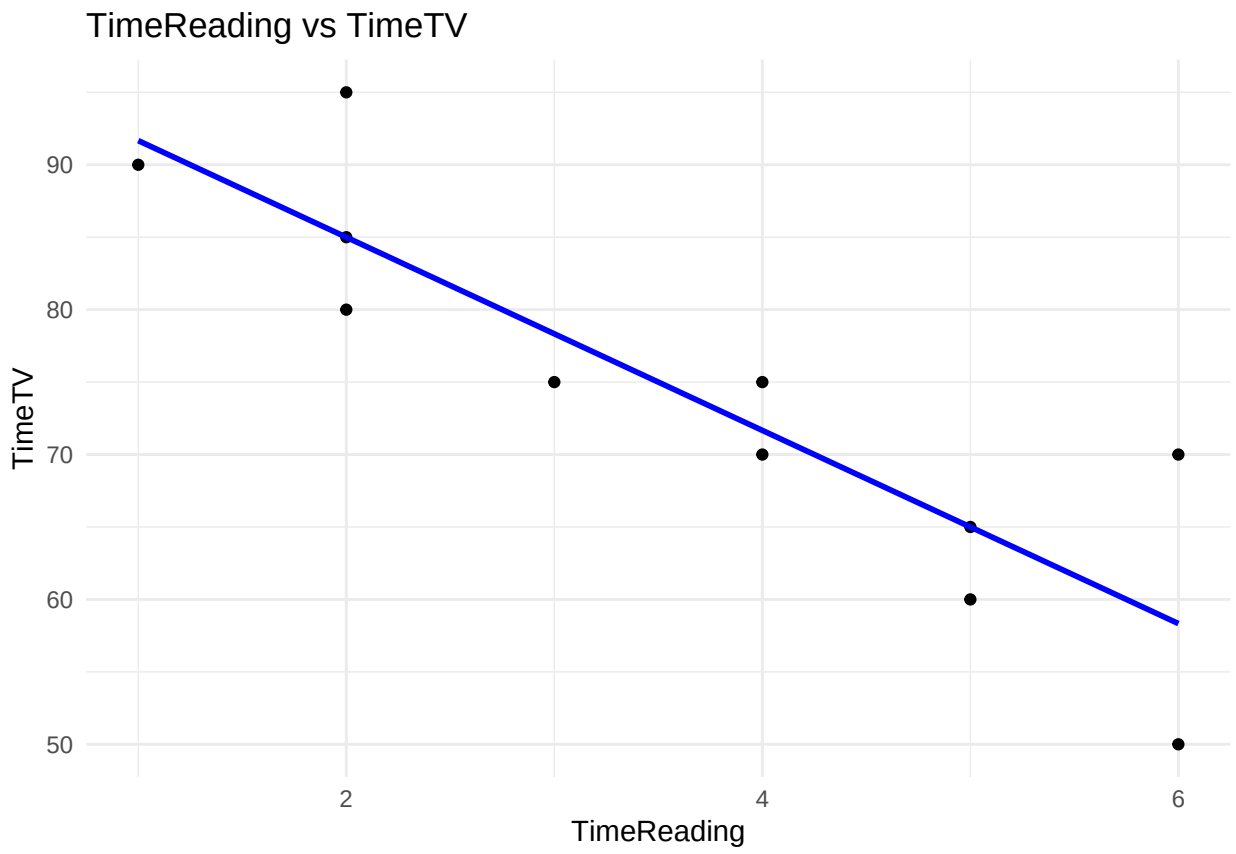
2024-10-10

Import CSV

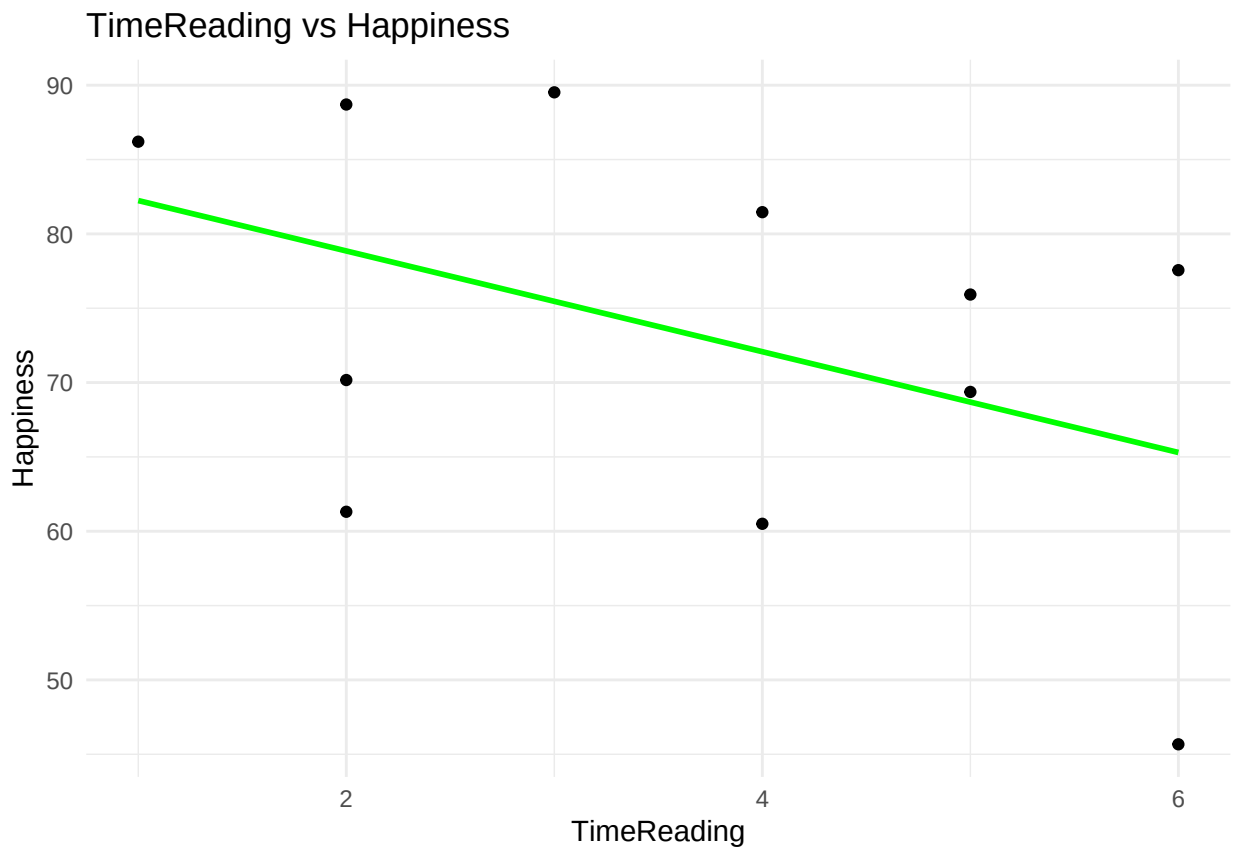
```
# Import CSV
c_data <- read.table(file = "student-survey.csv", header = TRUE, sep = ",")
```

Plots

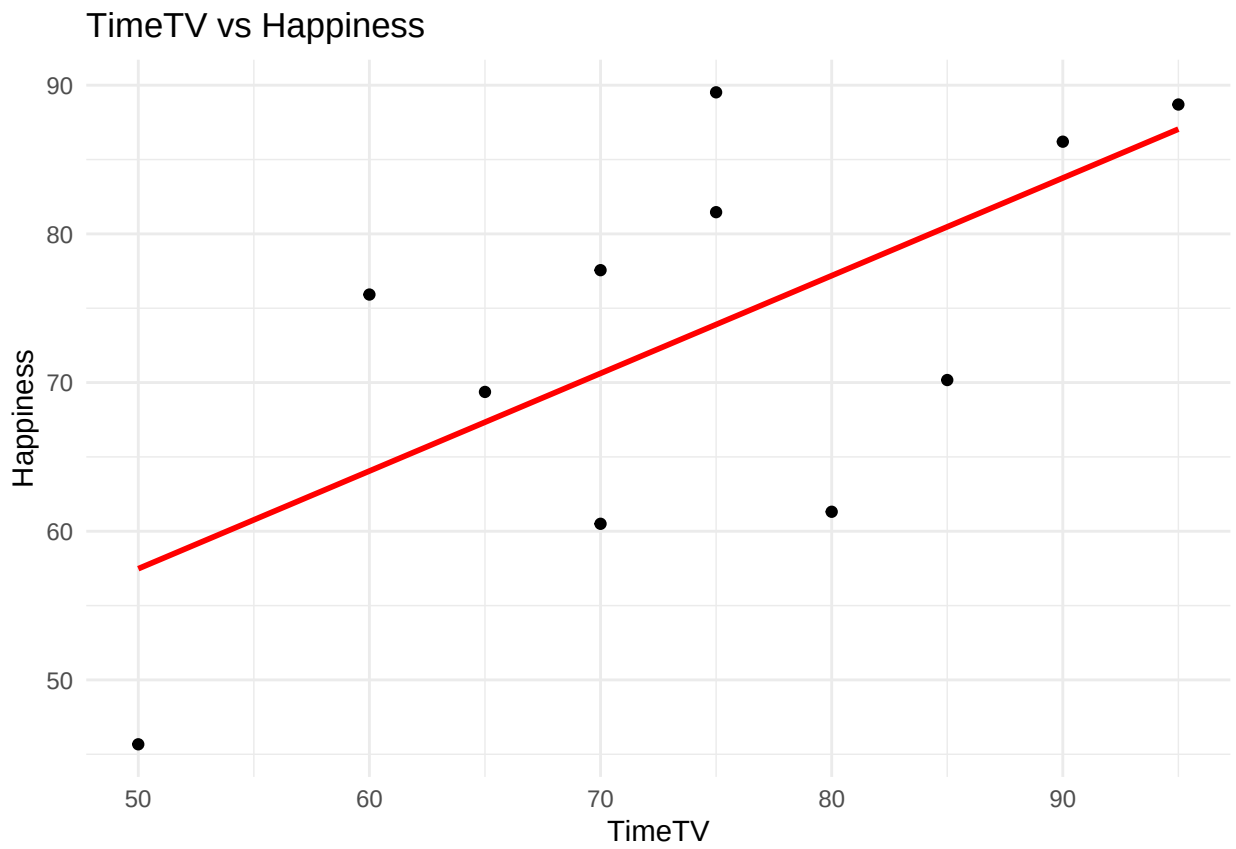
```
# Load ggplot2
library(ggplot2)

# Plot 1: TimeReading vs TimeTV
ggplot(c_data,aes(x = TimeReading, y = TimeTV)) + geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "blue") +
  labs(title = "TimeReading vs TimeTV", x = "TimeReading", y = "TimeTV") +
  theme_minimal()
```




```
# Plot 2: TimeReading vs Happiness
ggplot(c_data,aes(x = TimeReading, y = Happiness)) + geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "green") +
  labs(title = "TimeReading vs Happiness", x = "TimeReading", y = "Happiness") +
  theme_minimal()
```




```
# Plot 3: TimeTV vs Happiness
ggplot(c_data,aes(x = TimeTV, y = Happiness)) + geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title = "TimeTV vs Happiness", x = "TimeTV", y = "Happiness") +
  theme_minimal()
```



Findings
1. **TimeReading vs TimeTV:** Indicates a Negative relationship, as the more time spent reading decreases the time watching TV.
2. **TimeReading vs Happiness:** The trend seems mixed, but there might be a weak negative relationship overall.
3. **TimeTV vs Happiness:** There's a Positive correlation; as time watching TV increases, happiness tends to increase.

Covariance Matrix

```
#Create the correlation matrix
cov_matrix <- cov(data)
print(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
```

Findings
Diagonal Values : These are the variances of the individual variables

- TimeReading : 3.05

- TimeTV : 174.09
- Happiness : 185.45

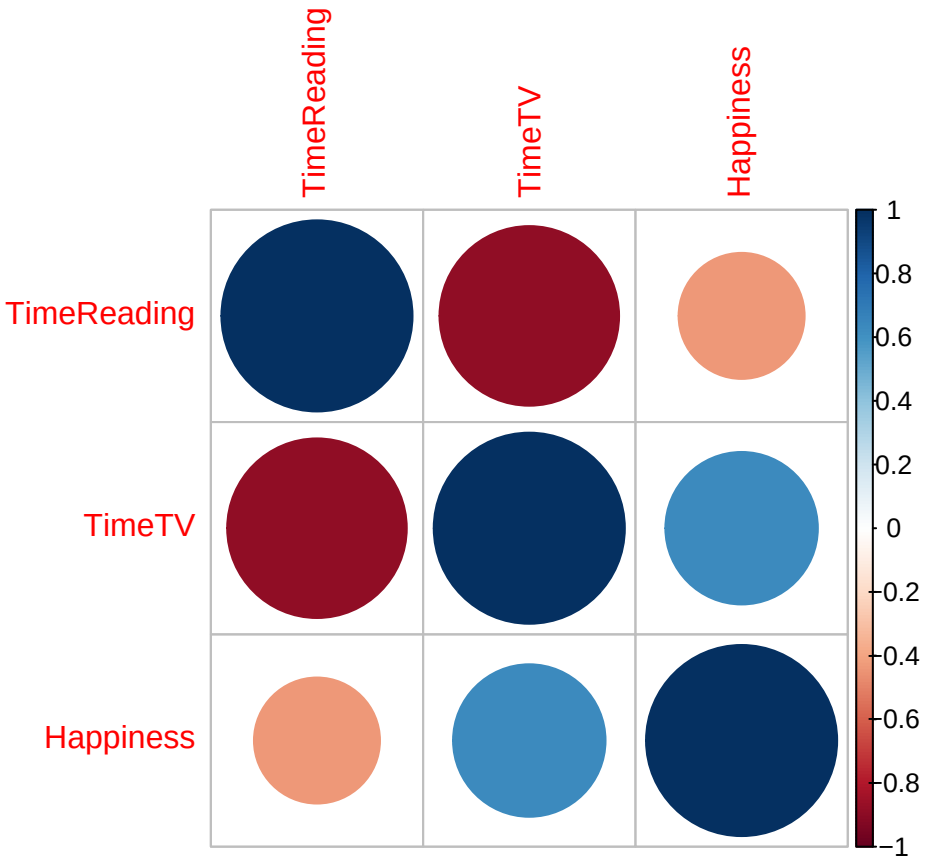
Off-diagonal values: These represent the covariances between pairs of variables
1. **Covariance between TimeReading and TimeTV** is -20.36, which indicates negative relationship (as reading time increases, TV time decreases)
2. **Covariance between TimeReading and Happiness** is -10.35, which indicates negative relationship (as reading time increases, happiness decreases)
3. **Covariance between TimeTV and Happiness** is 114.377, which indicated positive relationship (as TV time increase, Happiness increases as well)

Correlation Matrix

```
# Create the correlation matrix
cor_matrix <- cor(data)
print(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
```

```
#Visualizing Correlation Matrix
corrplot(cor_matrix)
```



Findings
1. **TimeReading vs TimeTV**: The correlation between time spent reading and watching TV is -0.883, indicating a strong negative relationship. As reading time increases, TV time tends to decrease significantly.
2. **TimeReading vs Happiness:** The correlation between reading time and happiness is -0.43, suggesting a slight negative relationship. While there is some trend, it is not strongly negative, indicating mixed results.
3. **TimeTV vs Happiness:** The correlation between TV time and happiness is 0.63, reflecting a moderately positive relationship. More time spent watching TV is associated with an increase in happiness.

Covarience vs Correlation Matrix

Correlation Matrix is easier to interpret because it gives standardized values between -1 and 1, making it easy to understand the strength and direction of relationships between variables. Covariance Matrix provides raw co-variation, the varying scales make it harder to interpret. Thus, correlation is generally more user-friendly for understanding relationships.

Correlation Test

```r
# Perform correlation test
cor_test_result <- cor.test(data$TimeReading, data$TimeTV,method="pearson")
print(cor_test_result)
```

```
##
##  Pearson's product-moment correlation
##
## data:  data$TimeReading and data$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
```

Findings
1. **Correlation Coefficient (cor = -0.883):** This value indicates a strong negative correlation between TimeReading and TimeTV. As the time spent on reading increases, the time spent watching TV tends to decrease significantly.
2. **p-value = 0.0003153:** Since the p-value is much smaller than 0.05, the correlation is statistically significant. This means there's strong evidence that the relationship between TimeReading and TimeTV is not due to random chance.
3. **Confidence Interval (-0.969 to -0.602):** The 95% confidence interval for the correlation is quite narrow and does not include 0, further confirming a strong negative relationship.

Based on the correlation test, while TimeReading and TimeTV have a strong negative correlation (meaning as reading time increases, TV time decreases), this does not imply causality. Correlation alone shows a relationship but cannot confirm an effect or direction of influence between the variables.