Week 7 Exercise

Miles Peña

2024-01-27

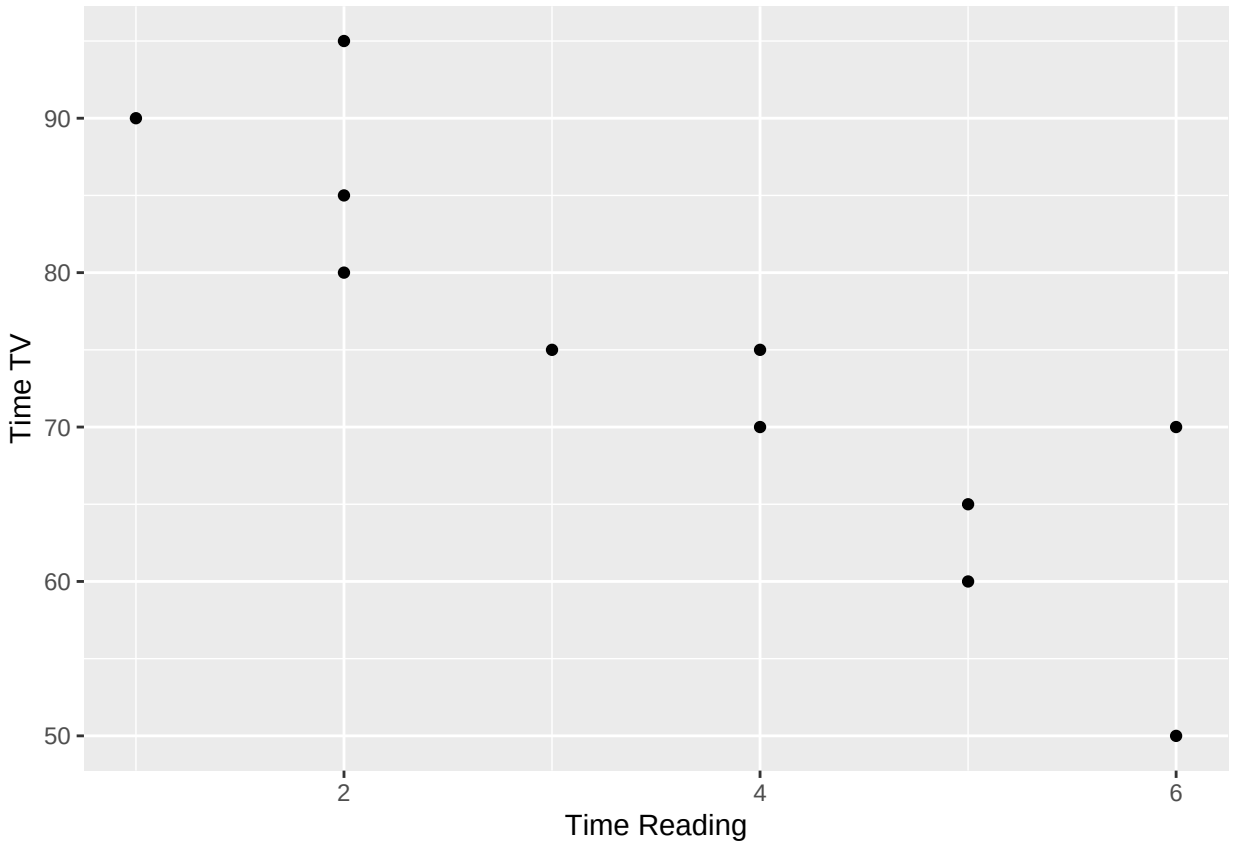
Student Survey

```
# Load Student Survey
setwd("/Users/milespena/Documents/R")
studentSurvey <- read.csv("student_survey.csv")
```

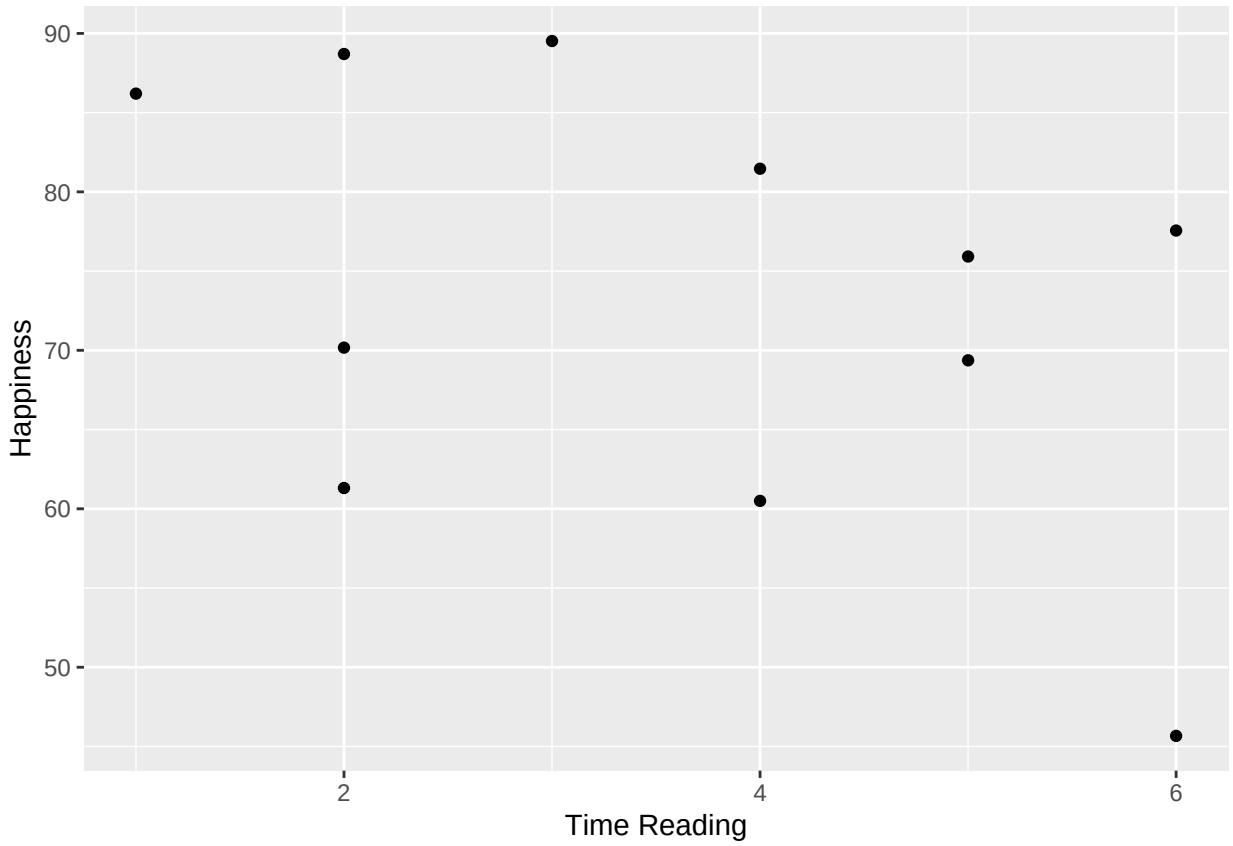
```
# Load ggplot

library(ggplot2)
```

Create plots for survey variables

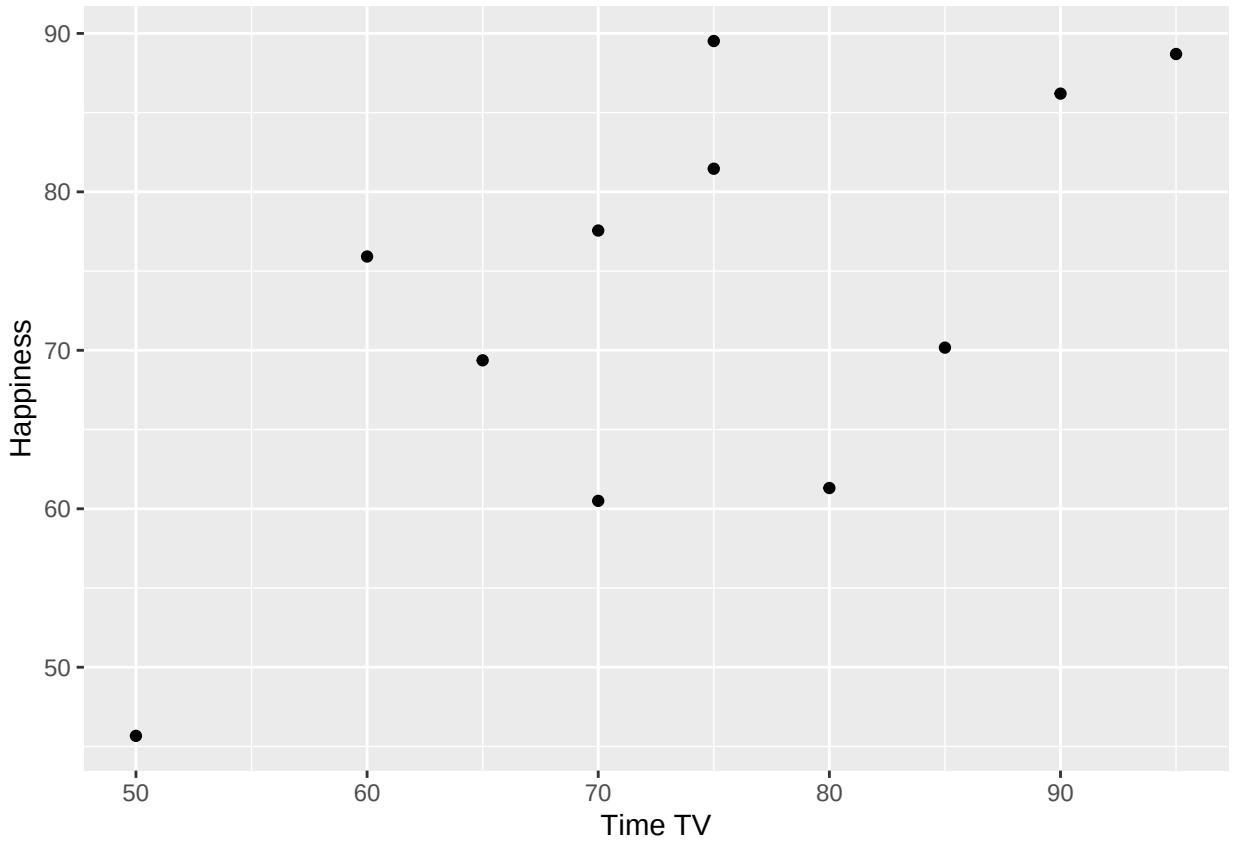
```
# Plot TimeReading vs. TimeTv

readingVsTv <- ggplot(data.frame(studentSurvey$TimeReading,
                                 studentSurvey$TimeTV)) +
  aes(x = studentSurvey$TimeReading, y = studentSurvey$TimeTV) +
  geom_point() + labs(x = "Time Reading", y = "Time TV")

print(readingVsTv)
```




```
# Plot TimeReading vs. Happiness

readingVsHappy <- ggplot(data.frame(studentSurvey$TimeReading,
                                    studentSurvey$Happiness)) +
  aes(x = studentSurvey$TimeReading, y = studentSurvey$Happiness) +
  geom_point() + labs(x = "Time Reading", y = "Happiness")

print(readingVsHappy)
```



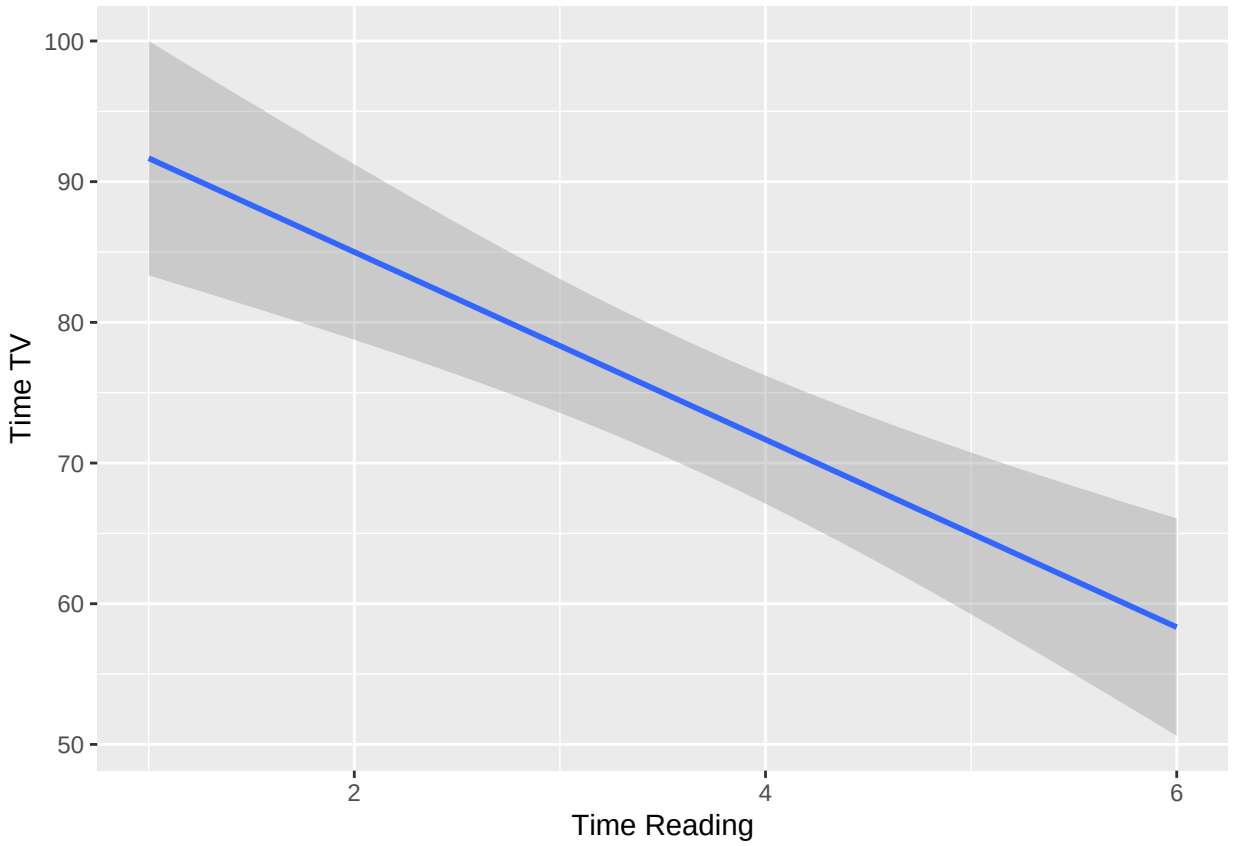

```
# Plot TimeTv vs. Happiness

tvVsHappy <- ggplot(data.frame(studentSurvey$TimeTV, studentSurvey$Happiness)) +
  aes(x = studentSurvey$TimeTV, y = studentSurvey$Happiness) +
  geom_point() + labs(x = "Time TV", y = "Happiness")

print(tvVsHappy)
```



Do the slopes indicate a positive or negative relationship?

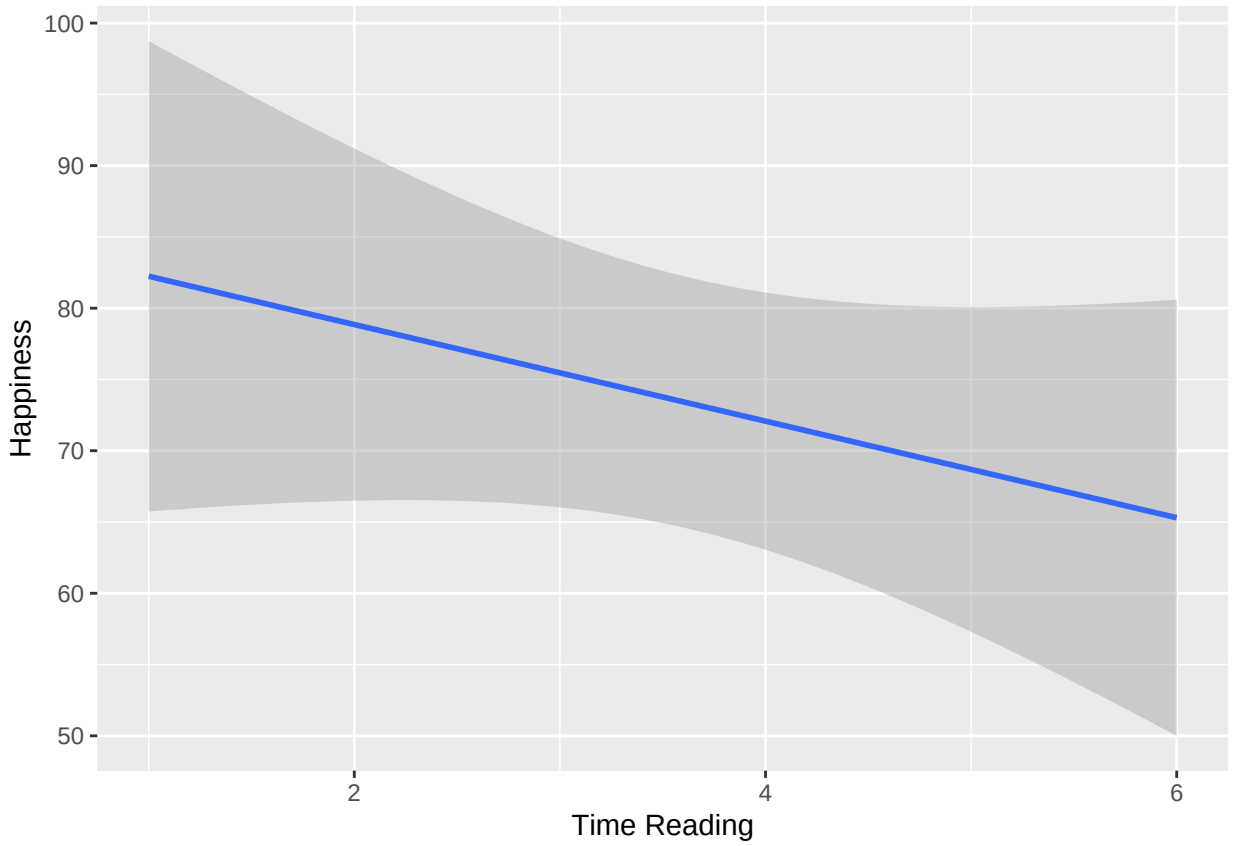
```
# TimeReading vs. TimeTV Slope using geom_smooth

readingVsTvSlope <- ggplot(data.frame(studentSurvey$TimeReading,
                                      studentSurvey$TimeTV)) +
  aes(x = studentSurvey$TimeReading, y = studentSurvey$TimeTV) +
  geom_smooth(method = lm) + labs(x = "Time Reading", y = "Time TV")

print(readingVsTvSlope)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



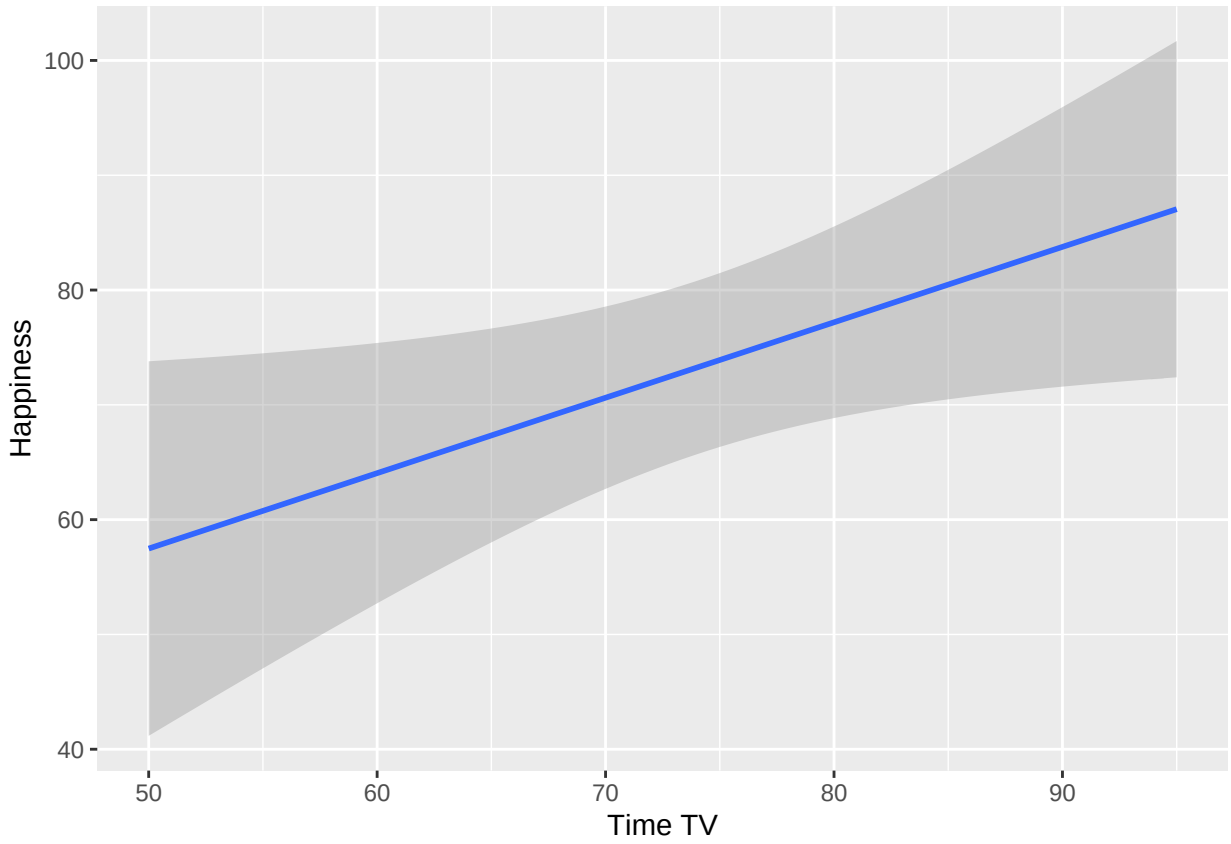

```
# TimeReading vs. Happiness Slope using geom_smooth

readingVsHappySlope <- ggplot(data.frame(studentSurvey$TimeReading,
                                         studentSurvey$Happiness)) +
  aes(x = studentSurvey$TimeReading, y = studentSurvey$Happiness) +
  geom_smooth(method = lm) + labs(x = "Time Reading", y = "Happiness")

print(readingVsHappySlope)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
# TimeTv vs. Happiness Slope using geom_smooth

tvVsHappySlope <- ggplot(data.frame(studentSurvey$TimeTV,
                                    studentSurvey$Happiness)) +
  aes(x = studentSurvey$TimeTV, y = studentSurvey$Happiness) +
  geom_smooth(method = lm) + labs(x = "Time TV", y = "Happiness")

print(tvVsHappySlope)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
# I used geom_smooth in order to create a best line for the plots since there
# was no clear line to follow (except for on the TimeReading vs TimeTV plot).
# Based on these graphs, the slopes for TimeReading vs. TimeTv and for
# TimeReading vs. Happiness have a negative relationship and the slope for
# TimeTv vs. Happiness has a positive relationship. However, when looking at the
# original plots, TimeReading vs. TimeTV still has a negative relationship,
# TimeReading vs. Happiness has no clear pattern and thus their relationship is
# weak/non-existent, and the TimeTV vs. Happiness is still, for the most part,
# a positive relationship.
```

Create a covariance matrix

```
cov(studentSurvey[, c(1:3)])
```

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

```
# Looking at the covariance matrix, we can see that the relationship between
# TimeReading and TimeTV is represented by a negative value meaning that as one
# variable increases, the other value tends to decrease (the same thing the plot
# showed us). The relationship between TimeReading and Happiness is also
# represented by a negative value. The relationship between TimeTv and Happiness
# is represented by a positive value meaning that the two variables tend to
# increase or decrease sequentially.
```

Create a correlation matrix

```
cor(studentSurvey[, c(1:3)])
```

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

```
# The correlation matrix can be expressed as a range of values expressed within
# the interval [-1, 1]. The value -1 indicates a perfect non-linear (negative)
# relationship, 1 is a perfect positive linear relationship and 0 is an
# intermediate between neither positive nor negative linear interdependency.
# Based on the numbers we get, the relationships all remain the same.
# TimeReading vs. TimeTV is a negative relationship. TimeReading vs. Happiness
# is also a negative relationship. TimeTV vs. Happiness is a positive
# relationship. In my opinion, it is easier to interpret the relationships
# between the variables by using the correlation matrix as the numbers are
# smaller and can be interpreted more simply. For example, while TimeReading vs.
# TimeTV have a negative relationship, it is not a perfect non-linear (negative)
# relationship since it is not a -1 but we can determine how close to -1 it is.
# The same goes for the other relationships.
```

Correlation Test

```
cor.test(studentSurvey$TimeReading, studentSurvey$TimeTV)
```

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

```
# The p-value of the test is 0.0003153, which is less than the significance
# level alpha = 0.05. We can conclude that TimeReading and TimeTV are
# significantly correlated with a correlation coefficient of -0.88 and p-value
# of 0.0003153. The more time that one spends reading the less time that they
# will spend watching TV.
```