Week 7 Assignment

Ruben Brionez Jr

January 28th 2024

Scatter Plots of Variables from Student Survey Data

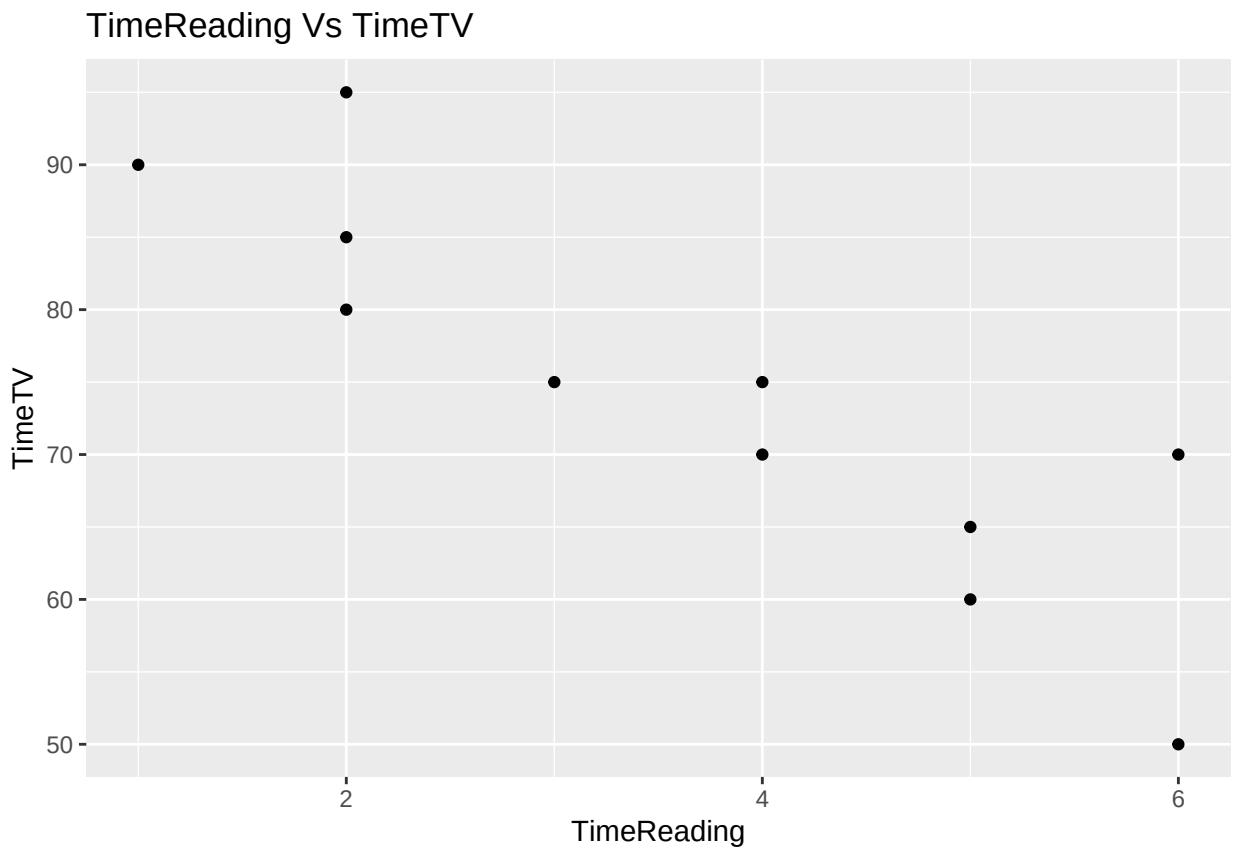
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.2
```

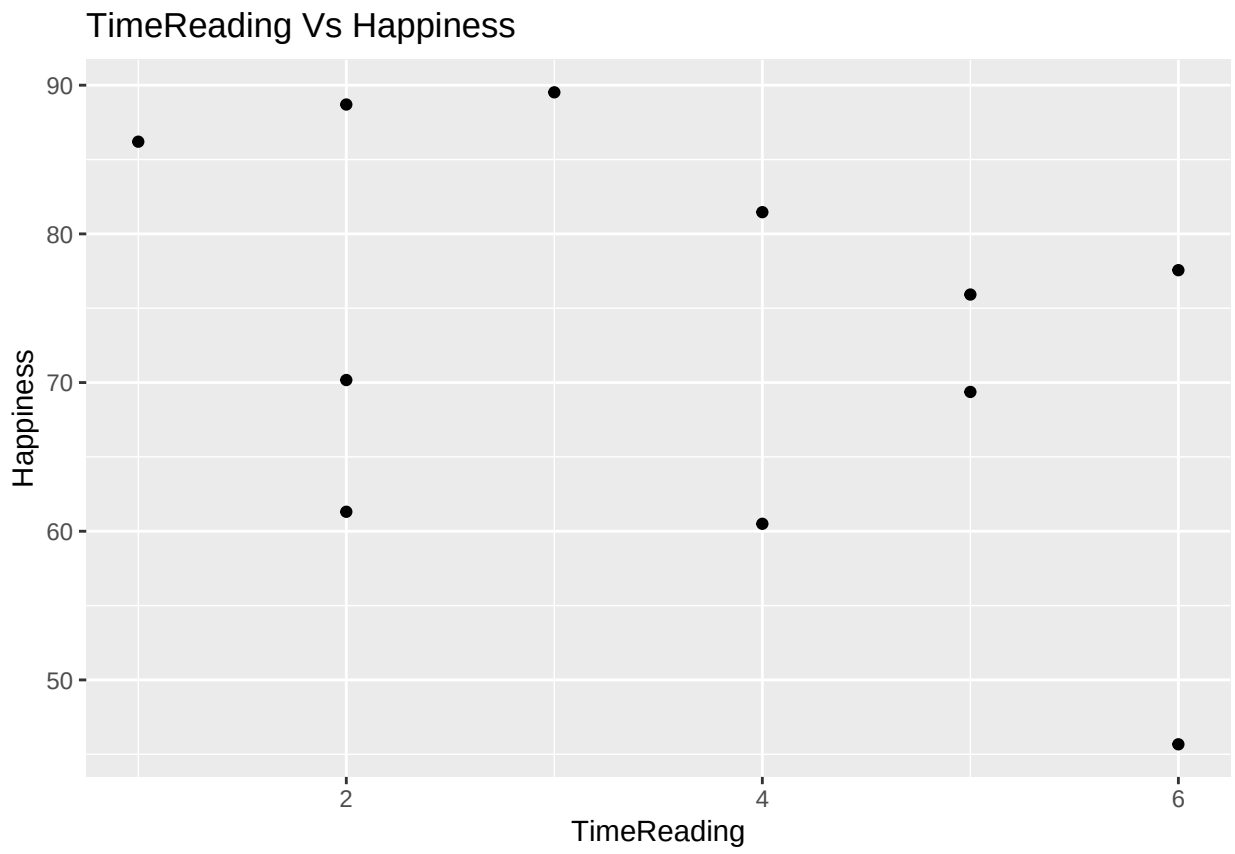
```
#Reading csv and saving to a variable
student_survey <- read.csv("student-survey.csv.crdownload")
student_survey
```

```
##    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
## 7            4     75     81.46      0
## 8            5     60     75.92      1
## 9            5     65     69.37      0
## 10           6     50     45.67      0
## 11           6     70     77.56      1
```

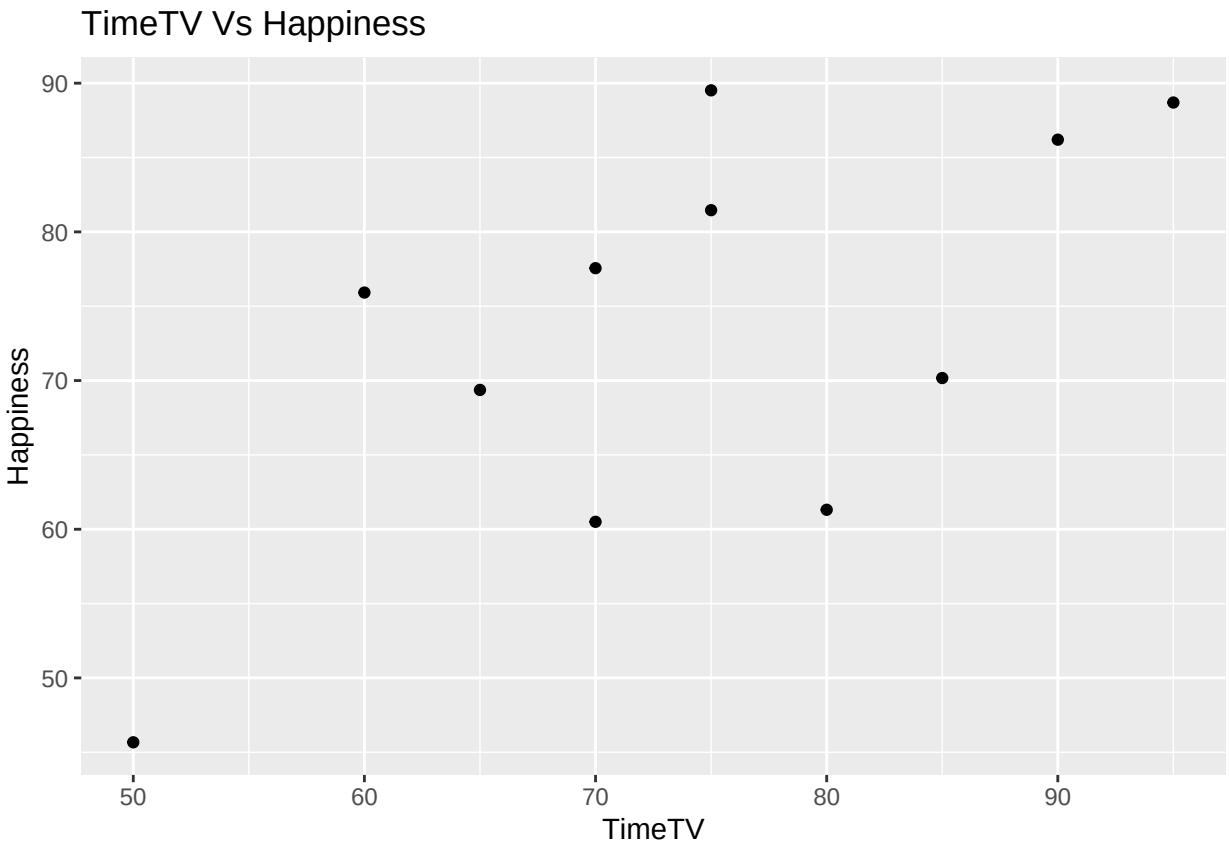
```
# Scatter plot of TimeReading VS TimeTV
ggplot(student_survey, aes(x = TimeReading, y = TimeTV)) +
  geom_point() + ggtitle("TimeReading Vs TimeTV")
```




```
# Scatter plot of TimeReading VS Happiness
ggplot(student_survey, aes(x = TimeReading, y = Happiness)) +
  geom_point() + ggtitle("TimeReading Vs Happiness")
```




```
#Scatter plot of TimeTV VS Happiness
ggplot(student_survey, aes(x = TimeTV, y = Happiness)) +
  geom_point() + ggtitle("TimeTV Vs Happiness")
```



Slope Analysis of Scatter Plots

In reviewing the slopes of the scatter plots, a rough estimate of the relationships can be attained. In the "TimeReading Vs TimeTV" plot, we can see that the slope appears to have a strong negative correlation. In the TimeReading Vs Happiness" plot it appears there is no correlation. In the "TimeReading Vs Happiness" plot there is a weak positive correlation.

Covariance Matrix

```
# Creating a Covariance Matrix of "student_survey" and dropping "Gender"
new_df <- subset(student_survey, select = -c(Gender))
cov(new_df)
```

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

When reviewing the covariance matrix, there is a negative value for "TimeReading" and "TimeTV. This shows that as"TimeReading" increases "TimeTV" decreases. The same is true with "TimeReading" and "Happiness", however the value being closer to zero indicates a weaker linear relationship.

Comparing "TimeTV" with "Happiness", there is a large positive value. This indicates that both variables increase and decrease in tandem. With a large value there appears to be a stronger linear correlation.

Creating Correlation Matrix

```
cor(new_df)
```

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

Reviewing the correlation matrix shows values for the strength and direction of the correlation between two variables. It's apparent that when comparing "TimeReading" and "TimeTV" there is a strong negative correlation based on the negative value being close to -1. The same is true when comparing "TimeReading" and "Happiness", but the negative correlation is slightly weaker. When comparing "TimeTV" with "Happiness" we find a positive correlation that is strong. One thing that stands out to me is that the visual correlation between "TimeTV" and "Happiness" appeared to be stronger than the correlation matrix value reflects.

Correltation Test

```
# Running a Correlation Test on "TimeReading" and "TimeTV"
cor.test(new_df$TimeReading, new_df$TimeTV, method="pearson")
```

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

The correlation test shows a very strong negative correlation. The value being close to negative one indicates this. I can say that TimeReading does have an effect on TimeTV. As TimeReading increases TimeTV would be decreasing. This makes sense in life in general, the more time spent reading the less there would be available for TV.