Student Survey

Torres Gloria

2023-04-28

```
setwd('C:/Users/glori/OneDrive/Documents/Gloria GIT/Gloria_Torres_DSC_520')


student_survey <- read.csv("data/student-survey.csv")
head(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
```

```
## Use R to calculate the covariance of the Survey variables and provide an explanation of why you woul

cov(student_survey)
```

```
##              TimeReading       TimeTV  Happiness      Gender
## TimeReading   3.05454545 -20.36363636 -10.350091 -0.08181818
## TimeTV      -20.36363636 174.09090909 114.377273  0.04545455
## Happiness   -10.35009091 114.37727273 185.451422  1.11663636
## Gender       -0.08181818   0.04545455   1.116636  0.27272727
```

```
## The measure of covariance helps to identify the relation/ change  between two variables.


## Examine the Survey data variables. What measurement is being used for the variables? Explain what ef

#Data is quantitative and categorical (Gender). The level of data measurement of the student survey dat

class(student_survey$TimeReading)
```

```
## [1] "integer"
```

```
class(student_survey$TimeTV)
```

```
## [1] "integer"
```

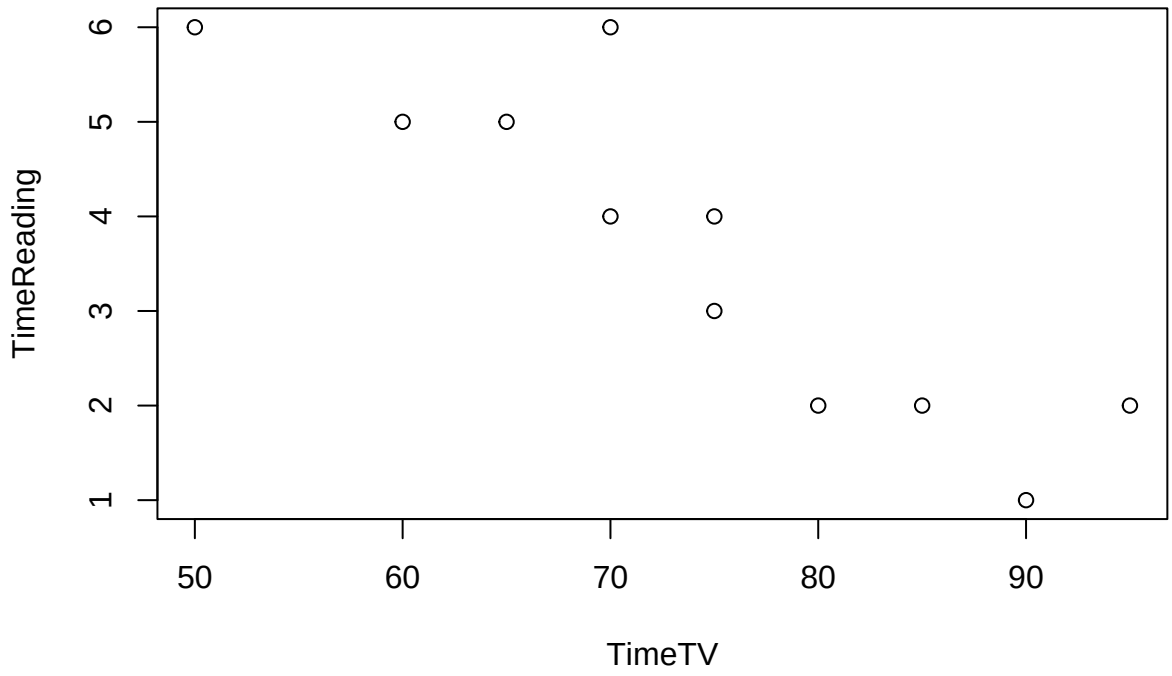
```r
class(student_survey$Happiness)
```

```
## [1] "numeric"
```

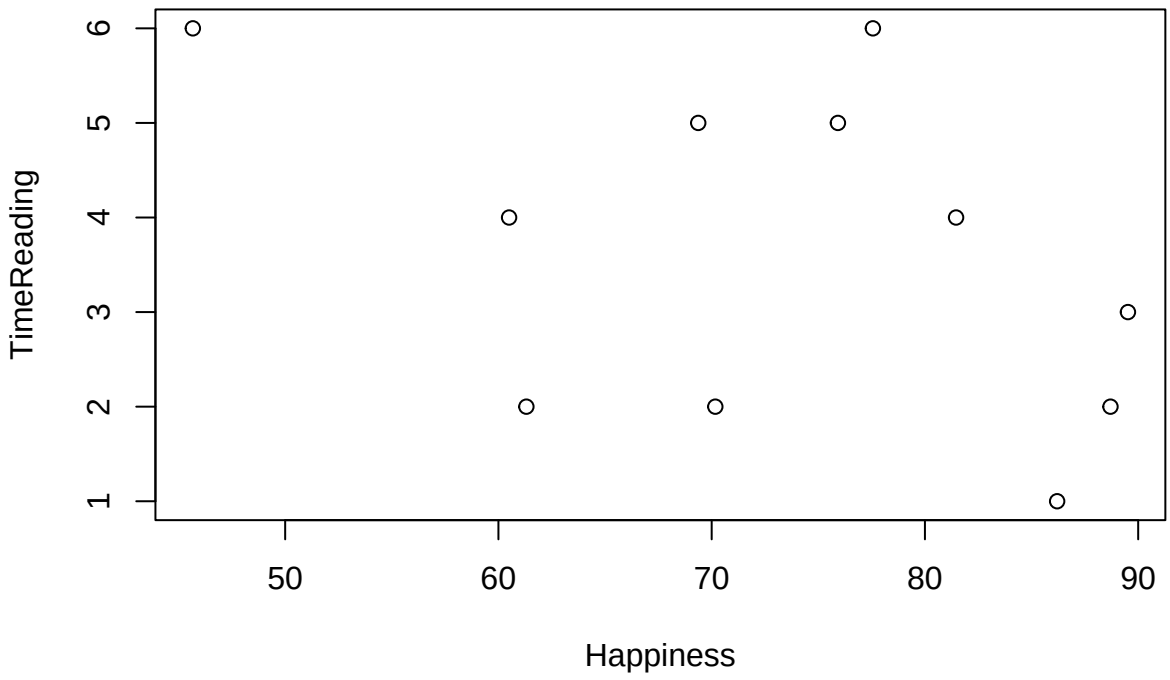
```r
class(student_survey$Gender)
```

```
## [1] "integer"
```

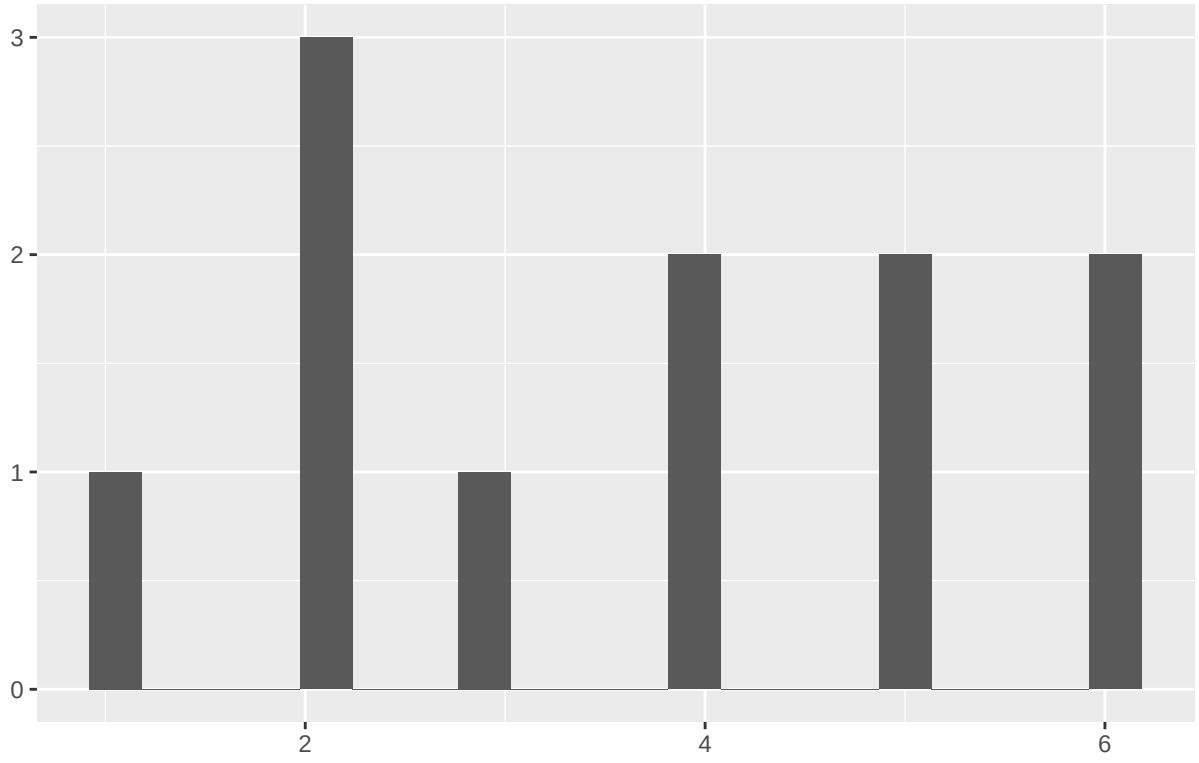
```r
## Choose the type of correlation test to perform, explain why you chose this test, and make a predicti

# I selected the Spearman method; for what I understand with the following plots, the data is not norma

library(ggplot2)

plot(TimeReading~TimeTV, data = student_survey)
```




```r
plot(TimeReading~Happiness, data = student_survey)
```




```
ggplot(student_survey, aes(TimeReading)) + geom_histogram(bins = 20)+ ggtitle('') + xlab('') + ylab('')
```




```
# Due to the high variance value between a couple variables (Happiness, Time TV and Gender), My predict
```

```
## Perform a correlation analysis of:

## All variables

all_test <- cor(student_survey, method = 'spearman')
all_test
```

```
##             TimeReading      TimeTV  Happiness      Gender
## TimeReading  1.00000000 -0.90725363 -0.4065196 -0.08801408
## TimeTV      -0.90725363  1.00000000  0.5662159 -0.02899963
## Happiness   -0.40651964  0.56621595  1.0000000  0.11547005
## Gender      -0.08801408 -0.02899963  0.1154701  1.00000000
```

```
## A single correlation between two a pair of the variables

single_test<- cor.test(student_survey$Happiness, student_survey$TimeReading, method = 'spearman')
```

```
## Warning in cor.test.default(student_survey$Happiness,
## student_survey$TimeReading, : Cannot compute exact p-value with ties
```

```
single_test
```

```
##
##  Spearman's rank correlation rho
##
## data:  student_survey$Happiness and student_survey$TimeReading
## S = 309.43, p-value = 0.2147
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## -0.4065196
```

```
## Repeat your correlation test in step 2 but set the confidence interval at 99%
c_score<- cor.test(student_survey$Happiness, student_survey$TimeReading, method = 'spearman', conf.level
```

```
## Warning in cor.test.default(student_survey$Happiness,
## student_survey$TimeReading, : Cannot compute exact p-value with ties
```

```
c_score
```

```
##
##  Spearman's rank correlation rho
##
## data:  student_survey$Happiness and student_survey$TimeReading
## S = 309.43, p-value = 0.2147
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## -0.4065196
```

```
## Describe what the calculations in the correlation matrix suggest about the relationship between the

# My analysis displays a moderate correlation between Time TV and Happiness (.56), While Happiness vs.

# Calculate the correlation coefficient and the coefficient of determination, describe what you conclud
library(ggplot2)
corr<- cor.test(student_survey$Happiness, student_survey$TimeTV)
corr
```

```
##
##  Pearson's product-moment correlation
##
## data:  student_survey$Happiness and student_survey$TimeTV
## t = 2.4761, df = 9, p-value = 0.03521
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.05934031 0.89476238
## sample estimates:
##      cor
## 0.636556
```

```r
corre1<- cor.test(student_survey$Happiness, student_survey$Gender, method = 'spearman')
```

```
## Warning in cor.test.default(student_survey$Happiness, student_survey$Gender, :
## Cannot compute exact p-value with ties
```

```r
corre1
```

```
##
##  Spearman's rank correlation rho
##
## data:  student_survey$Happiness and student_survey$Gender
## S = 194.6, p-value = 0.7353
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.1154701
```

```r
corre1<- cor.test(student_survey$TimeTV, student_survey$TimeReading, method = 'spearman')
```

```
## Warning in cor.test.default(student_survey$TimeTV, student_survey$TimeReading,
## : Cannot compute exact p-value with ties
```

```r
corre1
```

```
##
##  Spearman's rank correlation rho
##
## data:  student_survey$TimeTV and student_survey$TimeReading
## S = 419.6, p-value = 0.0001152
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.9072536
```

```r
library(datarium)

tryco <-lm(TimeTV~TimeReading, data = student_survey)
summary(tryco)
```

```
##
## Call:
## lm(formula = TimeTV ~ TimeReading, data = student_survey)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.333 -4.167 -1.667  1.667 11.667
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   98.333      4.723  20.818 6.37e-09 ***
```

```
## TimeReading   -6.667      1.181  -5.646 0.000315 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.526 on 9 degrees of freedom
## Multiple R-squared:  0.7798, Adjusted R-squared:  0.7553
## F-statistic: 31.87 on 1 and 9 DF,  p-value: 0.0003153
```

```
## Based on your analysis can you say that watching more TV caused students to read less? Explain.

# No, my analysis displays no positive correlation between TV time and Time Reading. Also, the coeffici
```

```
## Pick three variables and perform a partial correlation, documenting which variable you are "controll
library(ppcor)
```

```
## Loading required package: MASS
```

```
part_corr <- pcor.test(student_survey$Happiness, student_survey$TimeReading,student_survey$TimeTV)
part_corr
```

```
##    estimate  p.value statistic  n gp  Method
## 1 0.3516355 0.319059  1.062425 11  1 pearson
```

```
# Time TV  is the controlled variable; it was not the main focus of this analysis; however, due to the
```