```
> # Assignment: ASSIGNMENT Week 7
> # Name: Sharma, Dipika
> # Date: 2021-05-02
>
> getwd()
[1] "/Users/dipikasharma/R_Projects/DSC520"
> ## Set the working directory to the root of your DSC 520 directory
> setwd("/Users/dipikasharma/R_Projects/DSC520")
>
> ## Load the `data/student-survey.csv` to
> ssurvey_df <- read.csv("data/student-survey.csv")
> ssurvey_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
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
>
> ## Use R to calculate the covariance of the Survey variables
>
> cor(ssurvey_df$TimeTV, ssurvey_df$Happiness)
[1] 0.636556
> cor(ssurvey_df$TimeReading, ssurvey_df$Happiness)
[1] -0.4348663
>
> ## and provide an explanation of why you would use this calculation and what the results
indicate.
> ## Ans.
> ## I used  above calculation to see the relationship between the Time reading with Happiness
> ## and Time TV with Happiness. Using Cor function I found that Time TV and Happiness is
giving us positive
> ## correlation where as Time reading and Happiness is negative correlation which mean more
time students spent in
> ## watching TV their happiness increases but if student spent more time in reading their
happiness decreases.
>
> ## Examine the Survey data variables. What measurement is being used for the variables?
```

```
> ## Ans.
> ## We have four variables in student survey data.
> str(ssurvey_df)
'data.frame':   11 obs. of  4 variables:
 $ TimeReading: int  1 2 2 2 3 4 4 5 5 6 ...
 $ TimeTV     : int  90 95 85 80 75 70 75 60 65 50 ...
 $ Happiness  : num  86.2 88.7 70.2 61.3 89.5 ...
 $ Gender     : int  1 0 0 1 1 1 0 1 0 0 ...
> sapply(ssurvey_df, class)
TimeReading      TimeTV   Happiness      Gender
  "integer"   "integer"   "numeric"   "integer"
>
> ## Using above function we can clearly see TimeReading, TimeTV and Gender variables are
integer
> ## where as Happiness is numeric variable.
> ## TimeReading, TimeTV are interval variables.
> ## Gender is nominal variable with value 0 or 1
> ## Happiness is Ratio variable.
>
> ## Explain what effect changing the measurement being used for the variables would have on
the covariance calculation.
> ## Depending on the changes it can have significant effect on the covariance or it might have
no changes at all.
> ## Covariance is used to find out the relationship of of 2 variables
> ## or we can say finding out the dependency of one variable on other.
> ## if we will make any change to any varibale it will have some effect on covariance.
>
> ## Would this be a problem? Explain and provide a better alternative if needed.
> ## Ans.
> ## I would use nominal variables for Gender - Male and Female. currently we using the
numeric value 0 and 1
> ## which is not very clear in this way we can see different relationship between other
variables for specific Gender
> ## It would be interesting to see if that will change the overall relationship between the
variables.
>
> ## Choose the type of correlation test to perform,
> cor(ssurvey_df$TimeTV, ssurvey_df$Happiness)
[1] 0.636556
> ## explain why you chose this test,
> ## Ans.
> ## I would like to see relationship between time spent on TV and happiness.
>
> ## and make a prediction if the test yields a positive or negative correlation?
```

```
> ## It is positive correlation which indicate that if student spent more time watching TV their
happiness also increase.
>
> ## Perform a correlation analysis of:
> ##  All variables
> cor(ssurvey_df)
            TimeReading       TimeTV  Happiness       Gender
TimeReading  1.00000000 -0.883067681 -0.4348663 -0.089642146
TimeTV      -0.88306768  1.000000000  0.6365560  0.006596673
Happiness   -0.43486633  0.636555986  1.0000000  0.157011838
Gender      -0.08964215  0.006596673  0.1570118  1.000000000
> ##  A single correlation between two a pair of the variables
> cor.test(ssurvey_df$TimeTV, ssurvey_df$TimeReading, method = "pearson")

        Pearson's product-moment correlation

data:  ssurvey_df$TimeTV and ssurvey_df$TimeReading
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

>
> ##  Repeat your correlation test in step 2 but set the confidence interval at 99%
> cor.test(ssurvey_df$TimeTV, ssurvey_df$TimeReading, method = "pearson", conf.level = 0.99)

        Pearson's product-moment correlation

data:  ssurvey_df$TimeTV and ssurvey_df$TimeReading
t = -5.6457, df = 9, p-value = 0.0003153
alternative hypothesis: true correlation is not equal to 0
99 percent confidence interval:
 -0.9801052 -0.4453124
sample estimates:
      cor
-0.8830677

>
> ##  Describe what the calculations in the correlation matrix suggest about the relationship
between the variables.
> ## Be specific with your explanation.
```

```
> ## Ans.
> ## After looking at the correlation matrix we can clearly see that TimeReading and TimeTV
show negative correlation
> ## which mean with increase of one variable the second variable will decrease so if student
spent more time in watching
> ## TV they will spent less time in reading. Hence variables are opposite.
> ## if we look at TimeTV and Happiness we see positive correlation, we will see student are
more happy
> ## if they spent more time watching TV so if one variable
> ## increase the other variable will also increase
> ## Time reading and Happiness is negative correlation which mean if student spent more time
in reading their
> ## happiness decrease.
> ## and lastly all the gender are showing negative correlation with Time reading where as
gender are showing
> ## positive correlation wit TimeTV and Happiness.
>
> ## Calculate the correlation coefficient and the coefficient of determination,
> ## Correlation coefficient
> cor(ssurvey_df$TimeTV, ssurvey_df$TimeReading, method = "pearson")
[1] -0.8830677
> ## Correlation coefficient
> r = cor(ssurvey_df$TimeTV, ssurvey_df$TimeReading, method = "pearson")
> ## Square of corr. coef.
> lm.rel = lm(ssurvey_df$TimeTV~ssurvey_df$TimeReading)
> ## Coefficient of determination
> summary(lm.rel)$r.squared
[1] 0.7798085
>
> ## describe what you conclude about the results.
> ## Correlation define the strength of the relationship between an independent and
dependent variable
> ## and coefficient of determination tell us to what extent the variance of one variable
explains
> ## the variance of the second variable. In our case coefficient of determination is .77 then
approximately 70%
> ## of the observed variation can be explained by the inputs.
>
> ## Based on your analysis can you say that watching more TV caused students to read less?
Explain.
> ## Ans
> cor(ssurvey_df$TimeTV, ssurvey_df$TimeReading, method = "pearson")
[1] -0.8830677
```

```
> ## Using the Cor function on TimeTV and Time Reading we can cleary see bot variable have negative correlation
> ## which stat that if student spent more time in TV then they will spent less time reading.
>
> ## Pick three variables and perform a partial correlation,
> #install.packages("ppcor")
> library(ppcor)
> pcor.test(ssurvey_df$TimeTV, ssurvey_df$TimeReading, ssurvey_df$Happiness, method = "pearson")
   estimate      p.value statistic  n gp  Method
1 -0.872945 0.0009753126 -5.061434 11  1 pearson
> ## documenting which variable you are “controlling”.
> ## Ans: Happiness is the variable we are controlling.
> ## Explain how this changes your interpretation and explanation of the results.
> ## We already know that TimeTV and TimeReading are negative correlation that is if one variable increases
> ## second variable will decrease. After calculating partial correlation between the two where happiness is the controlling variable
> ## p value is 0.0009 which indicate if we can control happiness the relationship between two can show significant
> ## changes and it might improve.
```