Factors affecting GPA of

Langara College Students

PROJECT REPORT

Vivek Reddy 100328766

Contents

[Index of Figures 3](#_Toc36720465)

[Index of Tables 3](#_Toc36720466)

[1.0 Introduction 4](#_Toc36720467)

[2.0 Descriptive Analysis 7](#_Toc36720468)

[2.1 GPA 7](#_Toc36720469)

[2.2 Sleep time 8](#_Toc36720470)

[2.3 Working Hours 10](#_Toc36720471)

[2.4 Study Hours 11](#_Toc36720472)

[2.5 Study Time Preference 13](#_Toc36720473)

[3.0 Inferential Analysis 14](#_Toc36720474)

[3.1 Variable Selection 14](#_Toc36720475)

[3.2 Model building 20](#_Toc36720476)

[3.3 Residual Analysis**:** 23](#_Toc36720477)

[3.4 Estimation and Prediction 28](#_Toc36720478)

[4.0 Limitations 29](#_Toc36720479)

[5.0 Conclusion 29](#_Toc36720480)

[References 30](#_Toc36720481)

[Appendix I 30](#_Toc36720482)

[R outputs for model building 30](#_Toc36720483)

# Index of Figures

[Figure 1 Student Survey Form 7](#_Toc36720014)

[Figure 2 Confirmation message of form submission 8](#_Toc36720015)

[Figure 3 Histogram of GPA 9](#_Toc36720016)

[Figure 4 Boxplot of GPA 10](#_Toc36720017)

[Figure 5 Histogram of Sleep time 11](#_Toc36720018)

[Figure 6 Boxplot of Sleep time 11](#_Toc36720019)

[Figure 7 Histogram of Working hours 12](#_Toc36720020)

[Figure 8 Boxplot of Working hours 13](#_Toc36720021)

[Figure 9 Histogram of Study hours 14](#_Toc36720022)

[Figure 10 Boxplot of Study hours 14](#_Toc36720023)

[Figure 11 Distribution of preferred time period of study 15](#_Toc36720024)

[Figure 12 GPA vs Transit time graph 17](#_Toc36720025)

[Figure 13 GPA vs Social Media graph 17](#_Toc36720026)

[Figure 14 GPA vs Gaming Time graph 18](#_Toc36720027)

[Figure 15 GPA vs Sleep time graph 18](#_Toc36720028)

[Figure 16 GPA vs Working hours graph 19](#_Toc36720029)

[Figure 17 GPA vs Study hours graph 19](#_Toc36720030)

[Figure 18 GPA vs Work experience graph 20](#_Toc36720031)

[Figure 19 GPA vs Study time preference graph 20](https://langara-my.sharepoint.com/personal/nbatra01_mylangara_ca/Documents/DANA%20Project.docx#_Toc36720032)

[Figure 20 Residuals vs Sleep time graph 26](#_Toc36720033)

[Figure 21 Residuals vs Study time graph 26](#_Toc36720034)

[Figure 22 Standard Residuals vs Predicted GPA graph 27](#_Toc36720035)

[Figure 23 Normal Probability Plot (Q-Q Plot) 28](#_Toc36720036)

[Figure 24 Standardized residuals vs theoretical quantiles 28](https://langara-my.sharepoint.com/personal/nbatra01_mylangara_ca/Documents/DANA%20Project.docx#_Toc36720037)

[Figure 25 Residuals vs Normal scores (without outliers) 29](#_Toc36720038)

# Index of Tables

[Table 1 GPA statistics 9](#_Toc36719168)

[Table 2 Sleep time statistics 10](#_Toc36719169)

[Table 3 Working hours statistics 12](#_Toc36719170)

[Table 4 Study hours statistics 13](#_Toc36719171)

[Table 5 p-value, R-squared and Adjusted R-squared for different predictors 20](#_Toc36719172)

[Table 6 Table Summary of correlation test with each variable. 21](#_Toc36719173)

[Table 7 Correlation coefficients 24](#_Toc36719174)

[Table 8 VIF values table 24](#_Toc36719175)

[Table 9 Estimated values and 95% CI for GPA 29](#_Toc36719176)

[Table 10 Predicted values and 95% CI for GPA 30](#_Toc36719177)

# 1.0 Introduction

Students get equal opportunities, but why do some perform better than the others? How day-to-day activities can affect in an individual’s performance at school? In this project we are trying to analyze how leisure time activities such as using social media, travelling, relationships etcetera affects students’ GPA by taking a survey of Langara students. This will also help us to understand different strategies of students managing their time, either domestic or international. Usually, these students use social media such as Facebook, Instagram, WhatsApp, Twitter and many more applications in their leisure time and prefer to play either online or outdoor games, which are stress-buster activities for them. Through this survey, we will analyze how these factors affect their grades and gain insight on where students prefer to study, what time they prefer to study and how many hours they study per day. Besides this, we will also see how much their personal relationship(s) affect their study.

The performance of students in universities/colleges should be a concern not only to the administrators and educators, but also to corporations in the labor market. Academic achievement is one of the main factors considered by the employer in recruiting workers especially fresh graduates. The motivation of this project is that we want to analyze practically what we are learning theoretically during the course. With the help of a survey of few students, who are currently studying at Langara College, we want to understand the concept of R that how R deals with big data. Moreover, our team will use graphs such as boxplots and histograms to get an appropriate conclusion for each factor, which will help us to interpret the data. This study is interesting because during this project we learnt many new concepts of RStudio and got a chance to interact with different students and learnt their view on leisure time activities and studies.

We used simple random sampling to collect the samples. Each subject selected for sampling had an equal probability to participate in the survey. Simple Random Sampling (SRS) method is preferred because of its simplicity. We did face non-response bias during the process of sample collection. The sample size consists **101 students** with almost equal distribution between men and women.

**Questionnaire Design**: We used Google forms to create the web form and deployed the forms in Heroku which provides free cloud service for limited usage.

It was easy to create, manage and store data. Below is the attached screenshot of the web form.

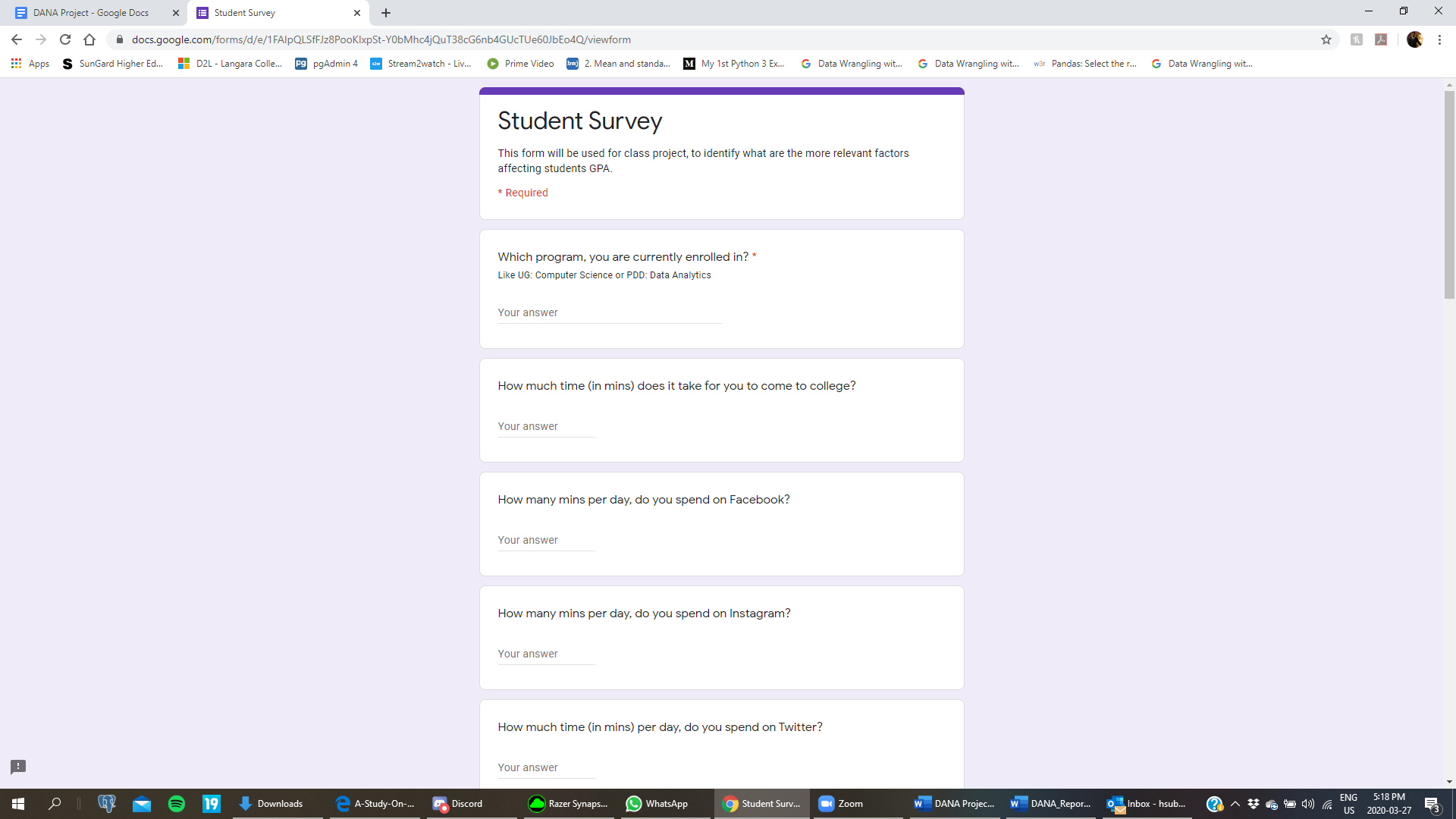


Figure 1 Student Survey Form

Questionnaire Link -<https://student-survey-form.herokuapp.com/>

1. Which course are you currently enrolled in?

2. How much time (in mins) does it take for you to come to college?

3. How many minutes per day, do you spend on social media?

4. How much average time (in hours) per week do you spend on online gaming?

5. How much average time (in hours) per week do you spend on outdoor sports?

6. How many hours of sleep do you get per day?

7. How many hours per week do you work?

8. What is your preferred time for study (Morning, Afternoon, Evening, Night, all are the same)?

9. Where do you prefer to study? (At Home, In Library, At Workplace, while traveling?)

10. How many hours do you study per day?

11. How much do you think your personal relationships affect your studies? (Affects very negatively, negatively, does not affect, positively, affects very positively)

12. What is your total professional work experience (in months)?

Once the survey is complete, a confirmation message is displayed to the respondent.

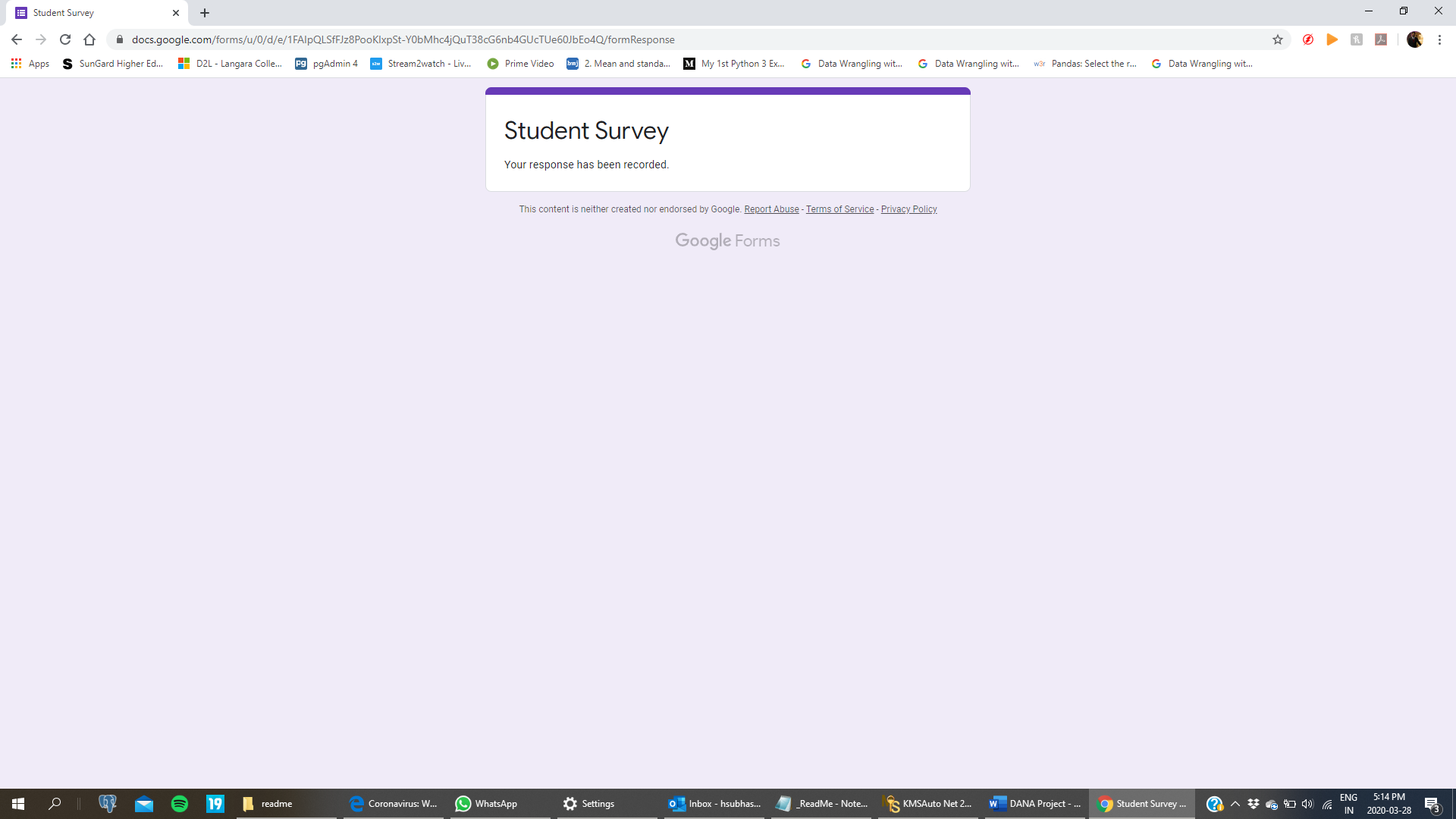


Figure 2 Confirmation message of form submission

# 2.0 Descriptive Analysis

The data used to conduct this project was taken from the survey done by Langara students in 2019. Though there are several factors which affect the ***GPA*** of a student, we consider ***Sleep Time, Work Hours and Study Hours*** for our analysis.

## 2.1 GPA

According to Kevin Potter at Study Portals, ”*Grade Point Average, is a number that indicates how well or how high you scored in your courses on average…….This number is then used to assess whether you meet the standards and expectations set by the degree program or university.”* In this project we will analyze the variables that affect GPA grades of Langara students with other factors. GPA is the response variable which is a quantitative variable with a ratio scale of measurement.

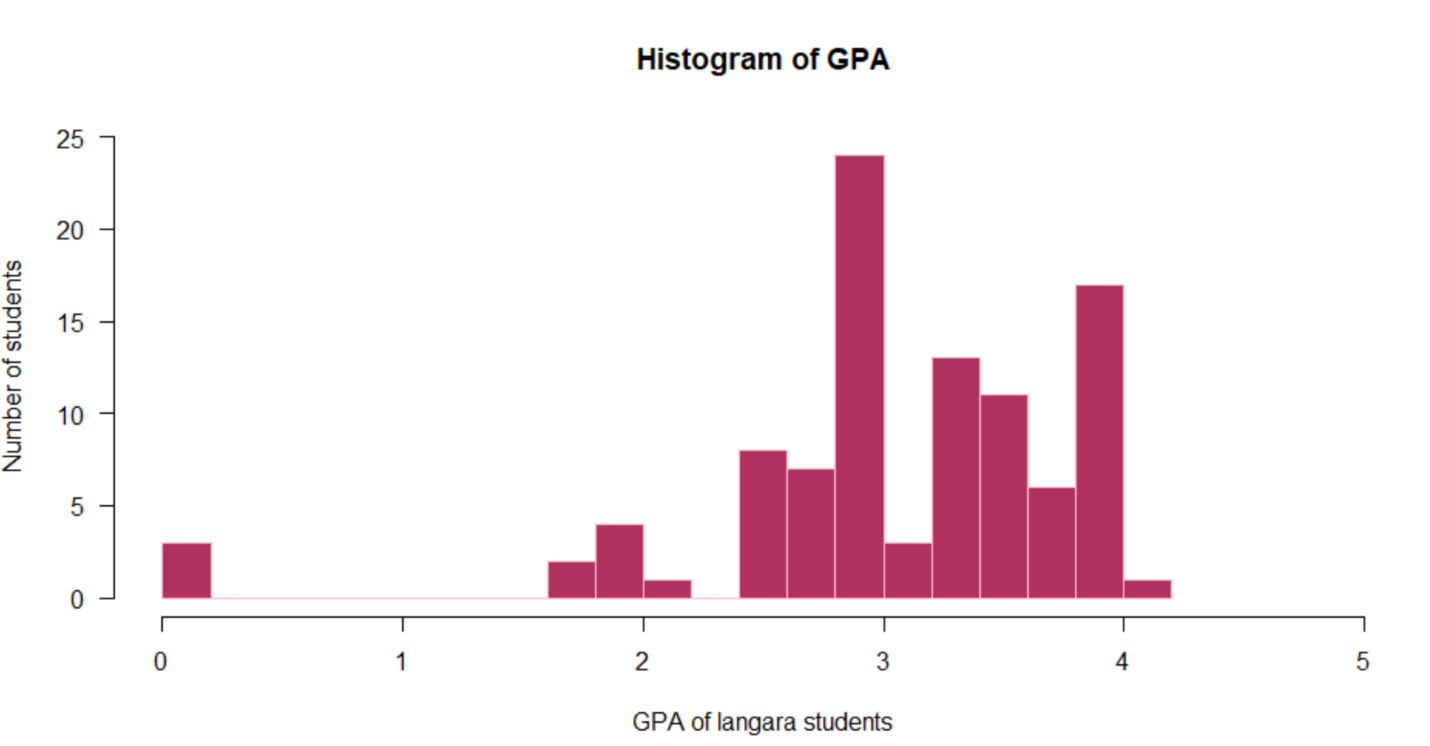


Figure 3 Histogram of GPA

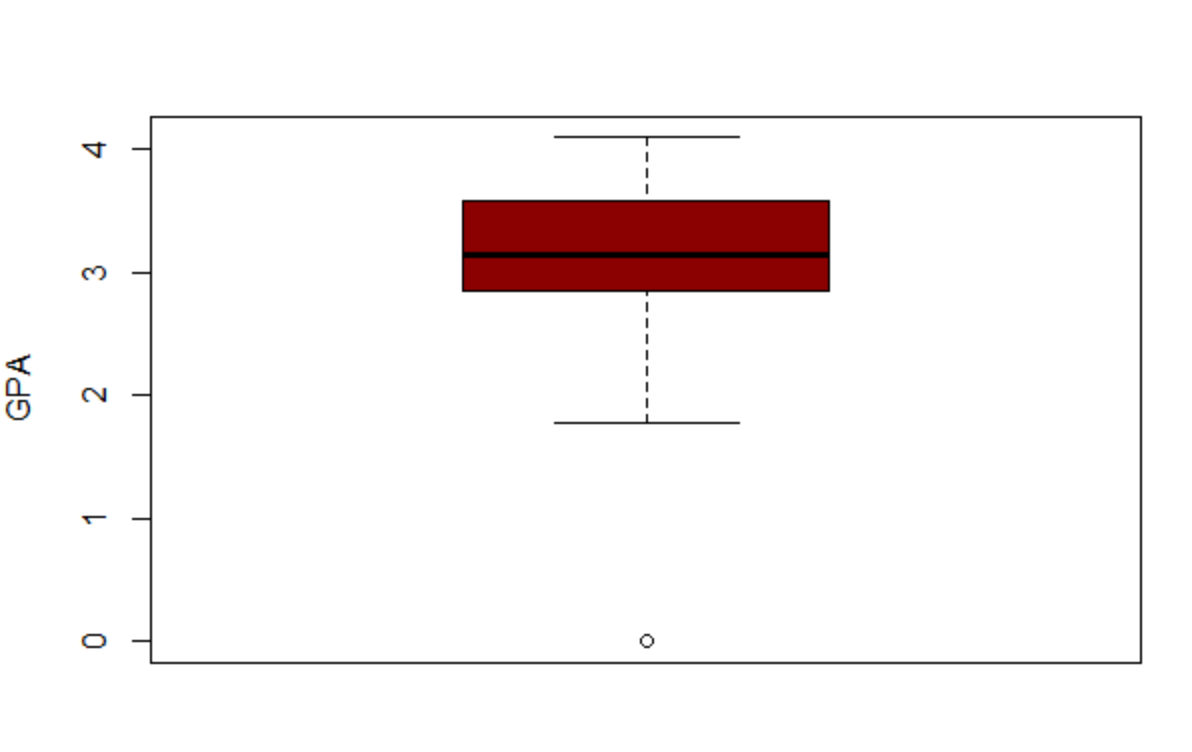


Figure 4 Boxplot of GPA

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Mean | Standard Deviation | Variance | Minimum | IQR | First Quartile | Median | Third Quartile | Maximum |
| 3.096 | 0.7849382 | 0.6161279 | 0.00 | 0.695 | 2.875 | 3.150 | 3.570 | 4.100 |

Table 1 GPA statistics

The above table illustrates the summarized statistics of a sample. There were no

outliers observed in this with a standard deviation of 0.7849382.

## 2.2 Sleep time

The relationship between GPA and Sleep time per day is being observed by this survey. Below boxplot illustrates that most students sleep for 7-8 hours per day and hours of sleep for all students lies between 5 to 10 hours. *“While previous studies show that many sleep variables share a relationship with academic performance, to date very few have examined all the sleep variables simultaneously. Further, many studies failed to control for significant predictors of academic performance such as standardized test scores, high school GPA, and SES. This makes it difficult to interpret the true strength of the relationship between sleep variables and academic performance.”*, stated by the US National Library of Medicine.

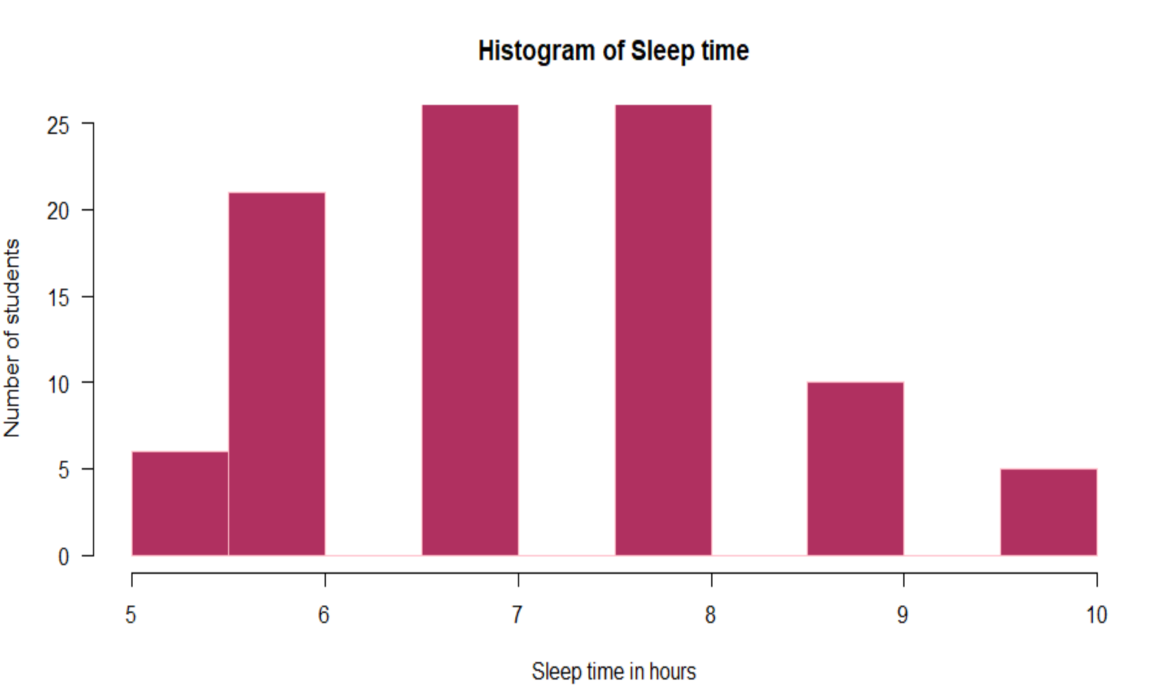


Figure 5 Histogram of Sleep time

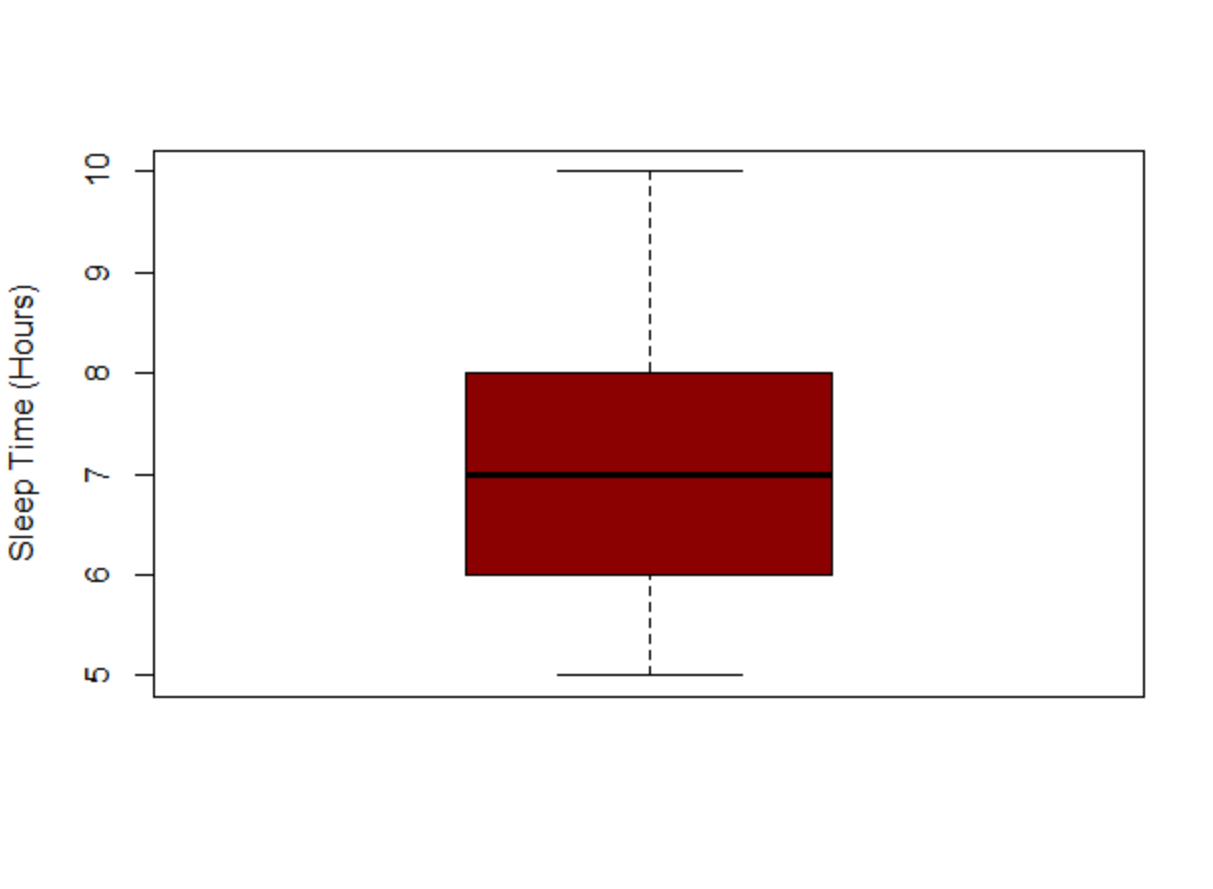


Figure 6 Boxplot of Sleep time

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Mean | Standard Deviation | Variance | Minimum | IQR | First Quartile | Median | Third Quartile | Maximum |
| 7.3 | 1.226805 | 1.505051 | 5.0 | 2.0 | 6.0 | 7.0 | 8.0 | 10.0 |

Table 2 Sleep time statistics

Above figure describes the statistics of the sleep time. The mean of the distribution is 7.3 and the median is 7 which is close to each other. It has a standard deviation of 1.226. From the above boxplot no outlier is visible.

## 2.3 Working Hours

“*Researches for the Bureau of Labor Statistics*[*found that students who worked less than 20 hours*](http://www.bls.gov/ore/abstract/ec/ec080020.htm)*per week had an average GPA of 3.13, while nonworking students had an average GPA of 3.04*”, Dr. Grey Pike, Indiana University.

According to statistical results there is a relationship between GPA and working hours per week of students. Most of the students work for 0 to 20 hours per week. However, 2% students work for more than 20 hours.

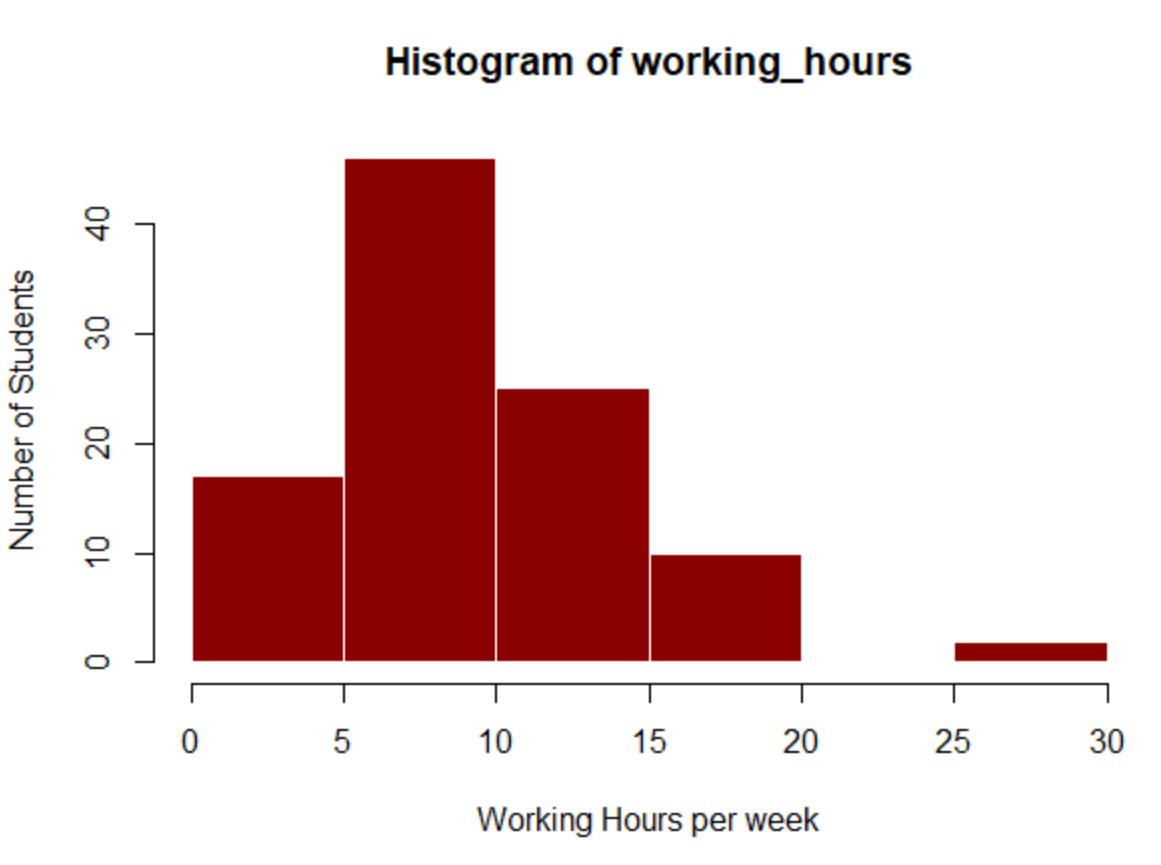


Figure 7 Histogram of Working hours

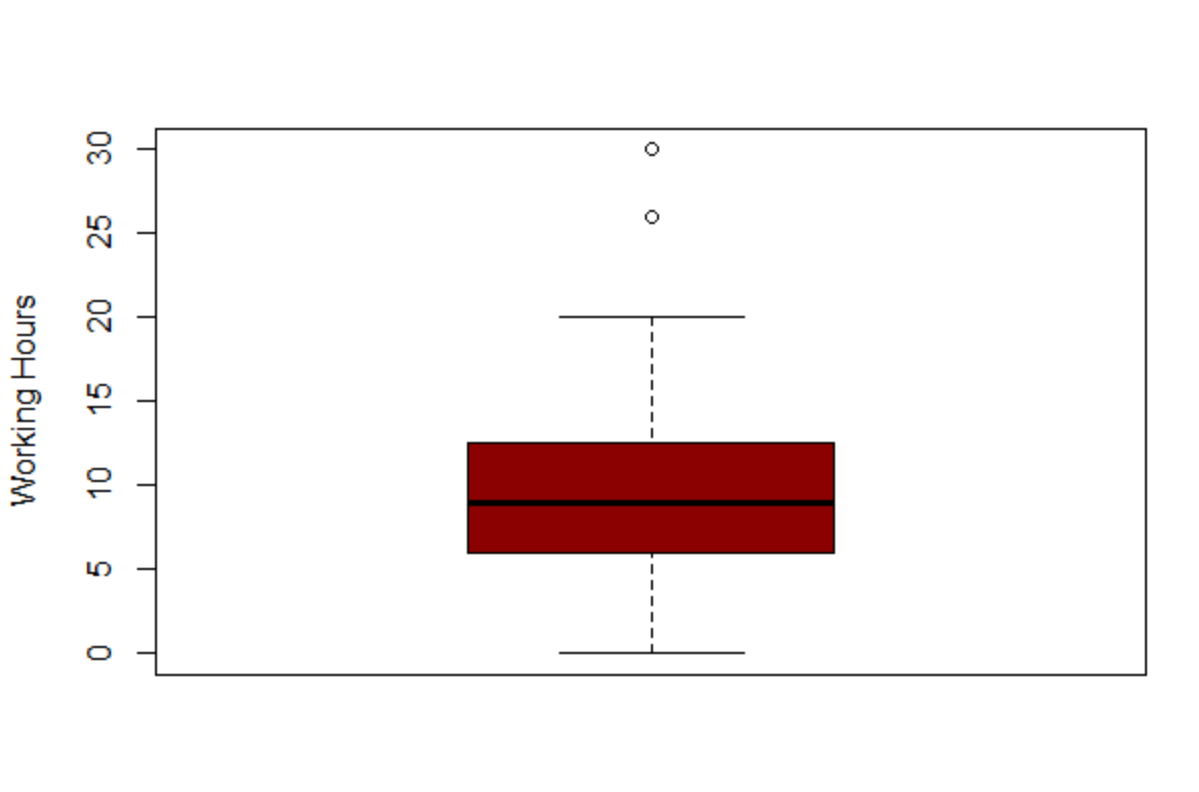


Figure 8 Boxplot of Working hours

From the above boxplot, we observe that there are two outliers in our dataset for the number of working hours. Before going ahead, we want to make sure that we take care of them and make our conclusions accordingly. The presence of outliers will also affect our statistical summary.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Mean | Standard Deviation | Variance | Minimum | IQR | First Quartile | Median | Third Quartile | Maximum |
| 9.86 | 5.310595 | 28.20242 | 0.00 | 3.86 | 6.00 | 9.00 | 9.86 | 30.00 |

Table 3 Working hours statistics

Above figure describes the statistics of the working hours. The mean of the distribution is 9.86 and the median is 9 which is close to each other. It has a standard deviation of 5.310. From the above boxplot no outlier is visible.

## 2.4 Study Hours

The more time a student spends studying, the higher you would expect the student's grade to be. This is something that we all believe is true, but let’s analyze how many hours do students’ need to dedicate in studies to score well. This includes studying outside of classroom. Textbook reading or notes, everything that adds value to academic performance is taken into consideration.

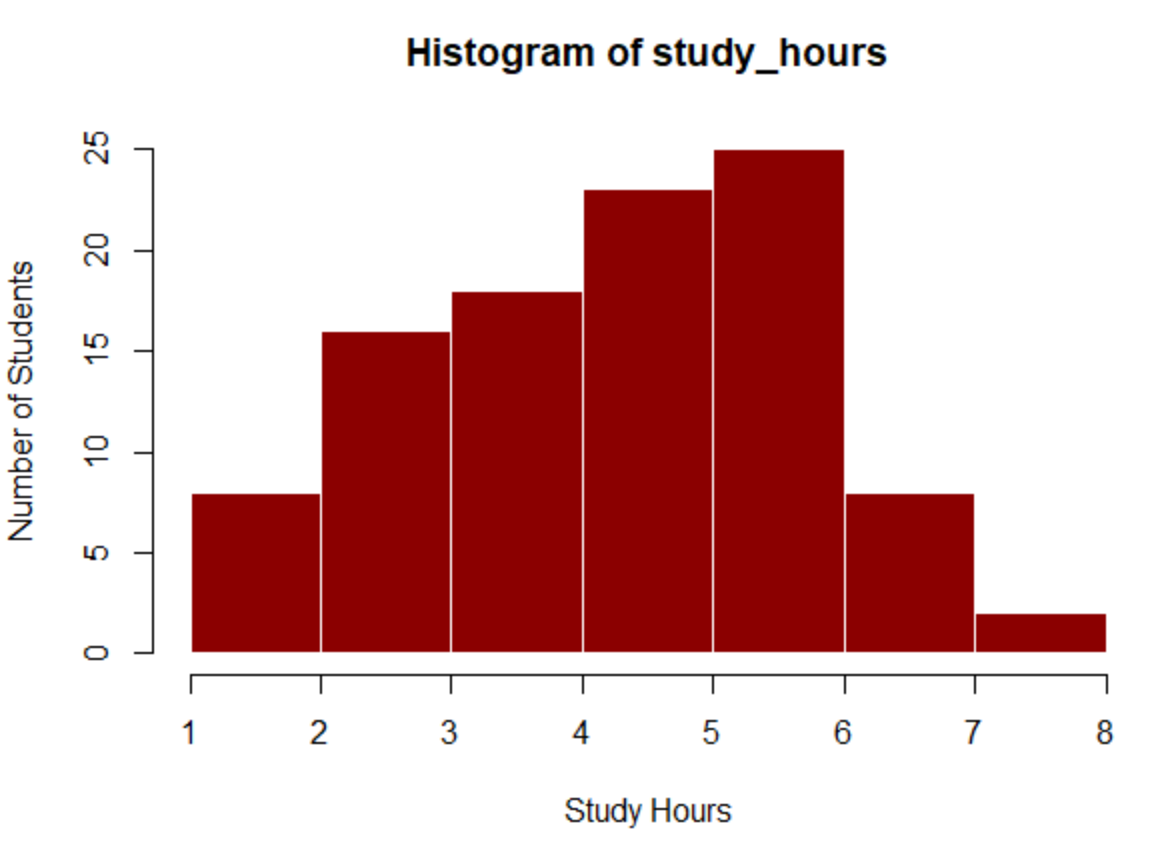


Figure 9 Histogram of Study hours

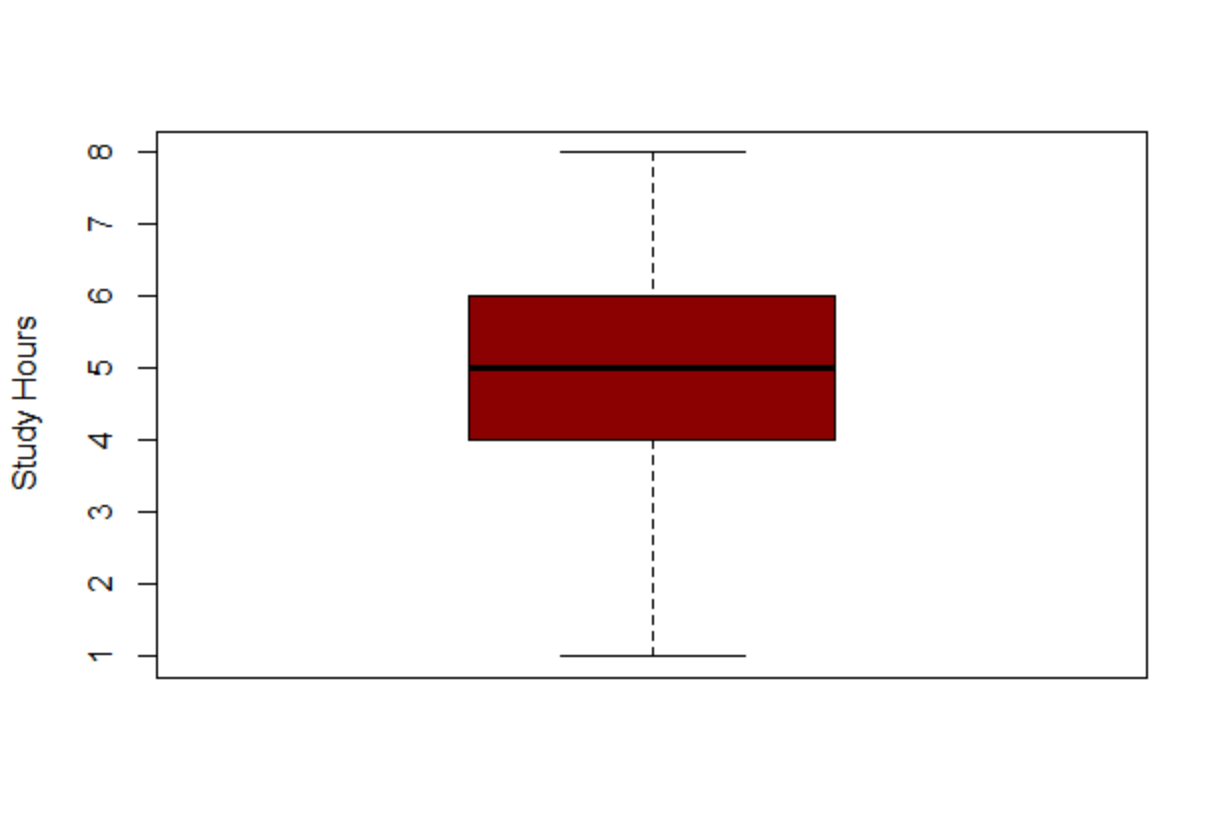


Figure 10 Boxplot of Study hours

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Mean | Standard Deviation | Variance | Minimum | IQR | First Quartile | Median | Third Quartile | Maximum |
| 4.71 | 1.539349 | 2.369596 | 1.00 | 2.0 | 4.00 | 5.00 | 6.00 | 8.00 |

Table 4 Study hours statistics

Above figure describes the statistics of the study hours. The mean of the distribution is 4.71 and the median is 5 which is close to each other. It has a standard deviation of 1.539. From the above boxplot no outlier is visible.

## 2.5 Study Time Preference

Figure 11 Distribution of preferred time period of study

Above figure describes the statistics of the study time preference. It appears that most students prefer to study during morning hours and few students prefer to study during the afternoon time. However, 16 students prefer studying during evening hours while 17 people agree on studying at any time of the day.

# 3**.0 Inferential Analysis**

## 3.1 Variable Selection

Variables of interest:

1. How much time (in mins) does it take for you to come to college? - **transit\_time**
2. How many mins per day, do you spend on Facebook+ Instagram+ Twitter? - **social\_media**
3. How much time (in hours) per week do you spend on online gaming + outdoor sports? - **game**
4. How many hours of sleep do you get per day? - **sleep\_time**
5. How many hours per week do you work? - **working\_hours**
6. What is your preferred time to study? - **study\_time\_prefer**
7. Where do you prefer to study? - **study\_place\_prefer**
8. How many hours do you study per day? - **study\_hours**
9. How much you think your personal Relationships affect your studies? - **relation**
10. What is your total professional work experience (in months)? - **experience**

Relationship between GPA and selected predictors:

In building our model, 10 variables were of interest in understanding their relationship with GPA. These variables were transit time, social media usage time, playing time, sleep time, working hours, study time preference, study place preference, study hours, personal relationships and previous work experience.

Relationship between each variable of interest and GPA were plotted to gain further insights into the sample. They are shown in Figures 12-19.

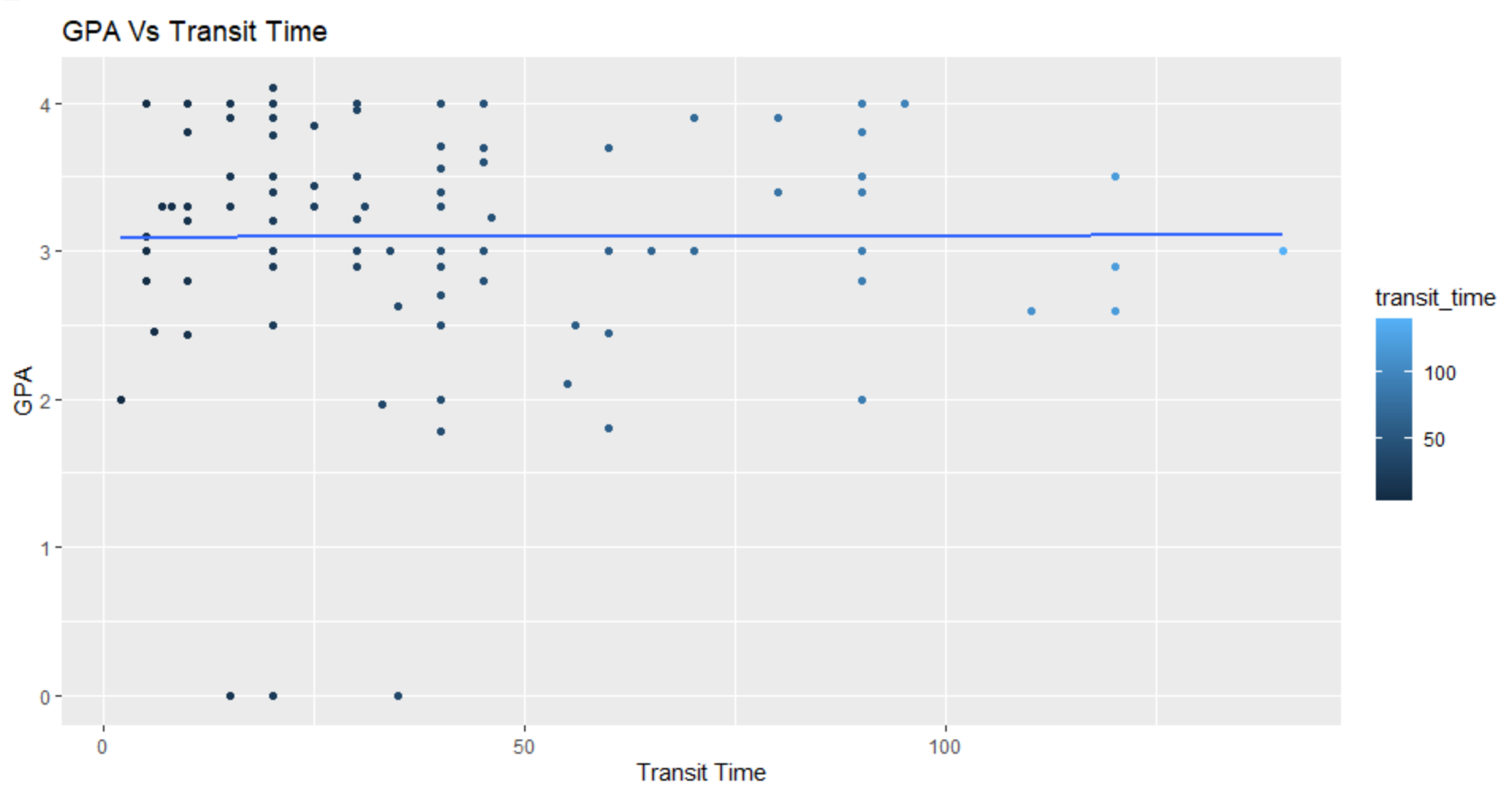


Figure 12 GPA vs Transit time graph

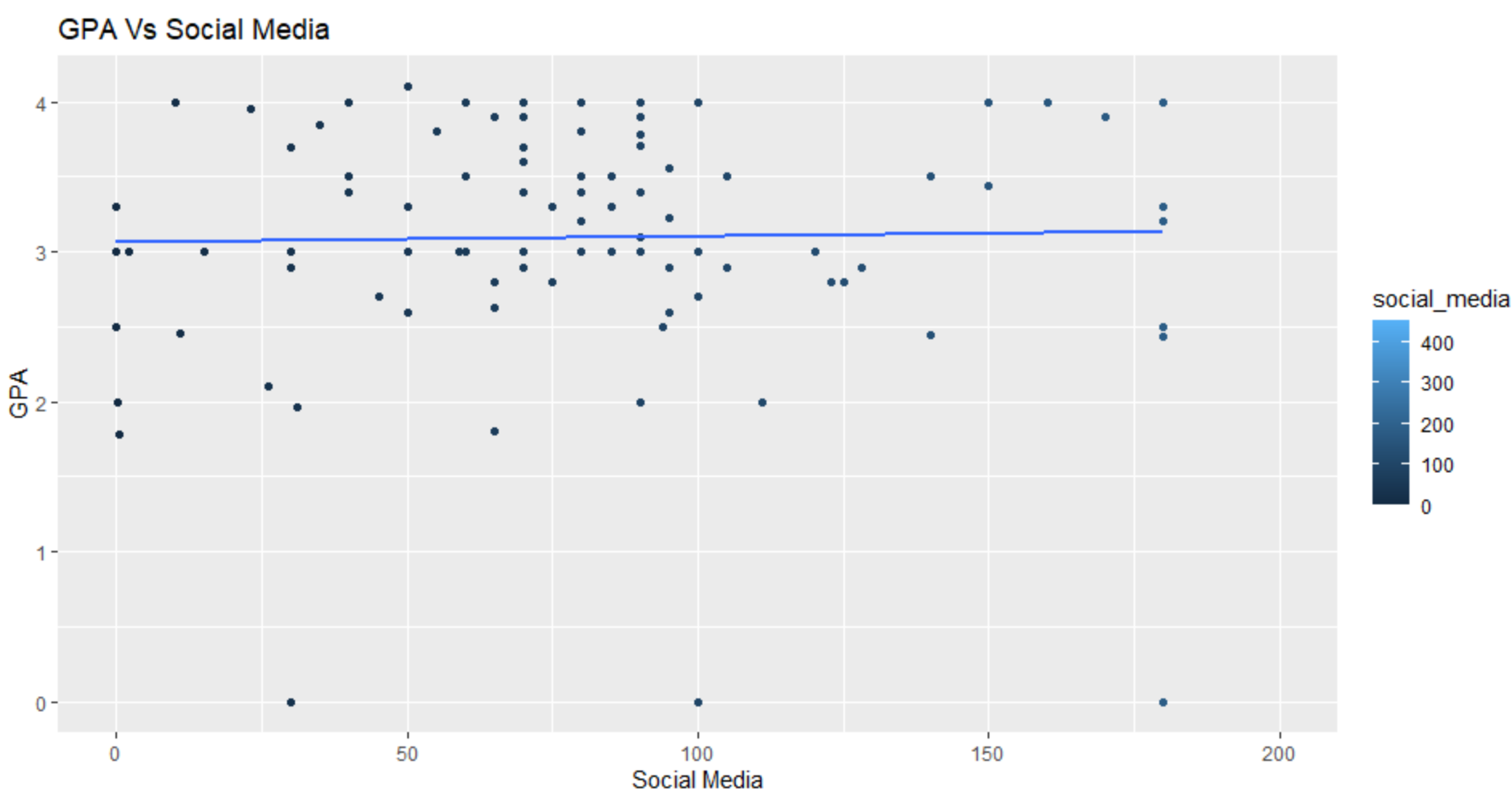


Figure 13 GPA vs Social Media graph

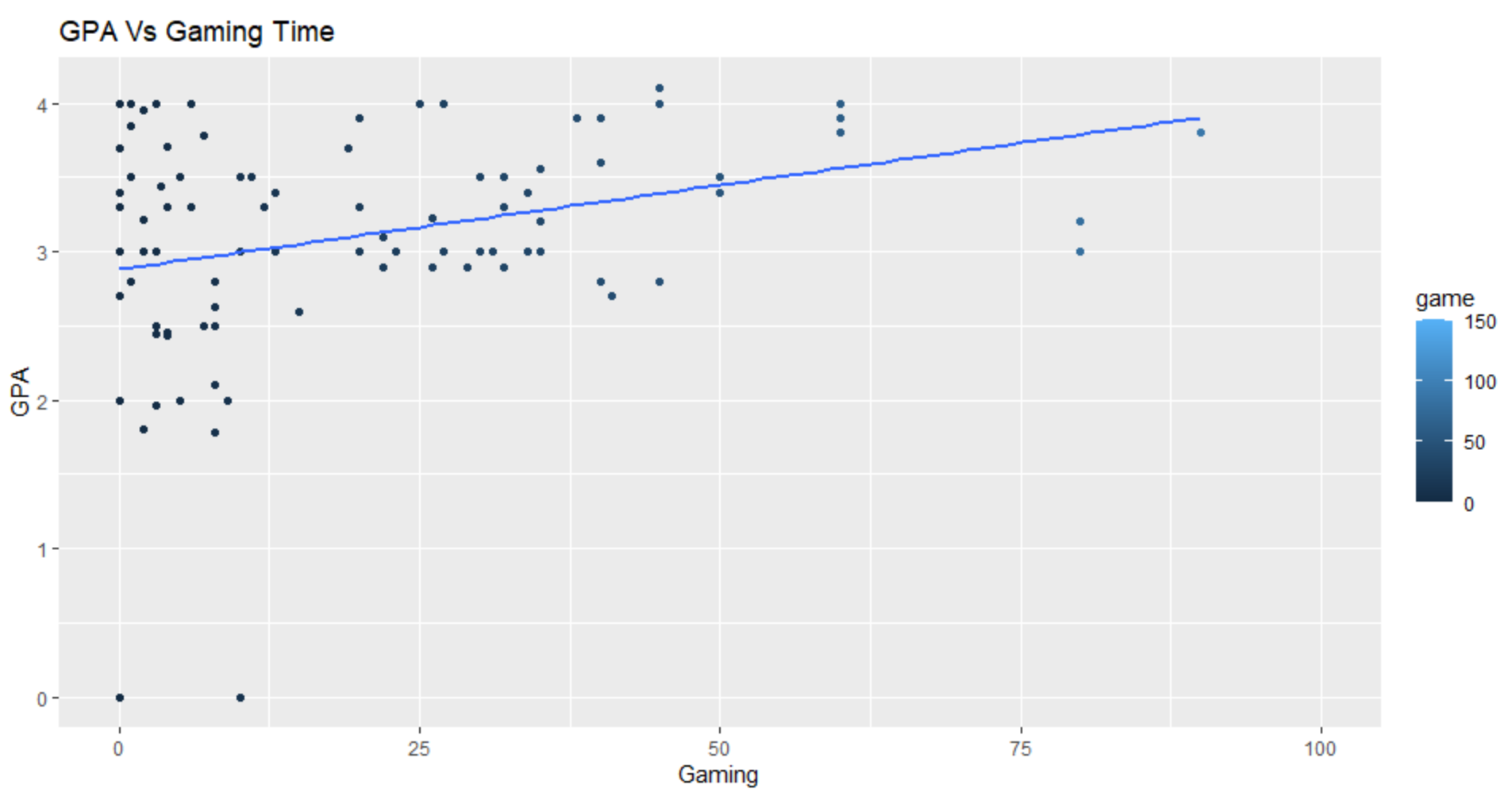


Figure 14 GPA vs Gaming Time graph

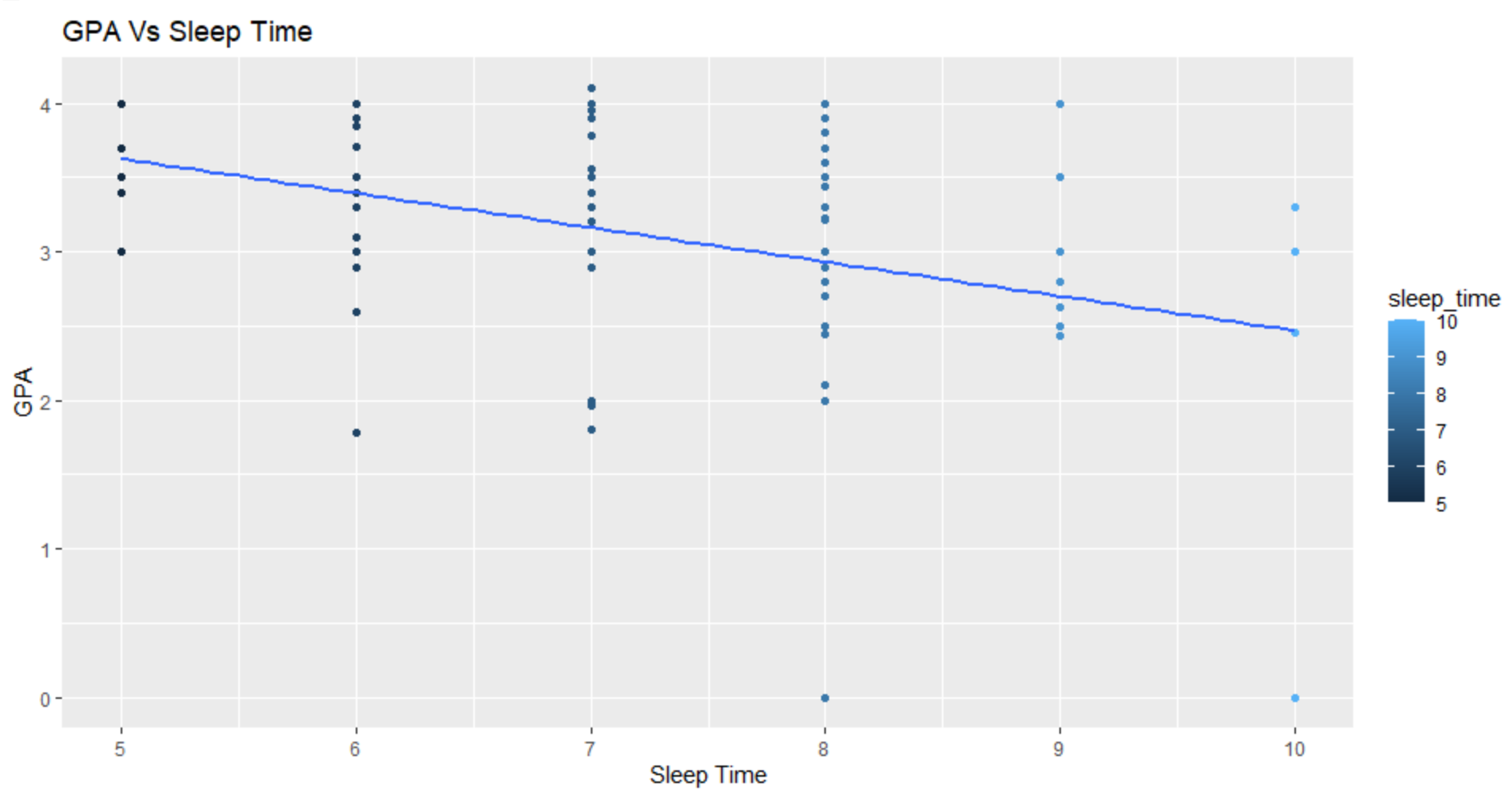


Figure 15 GPA vs Sleep time graph

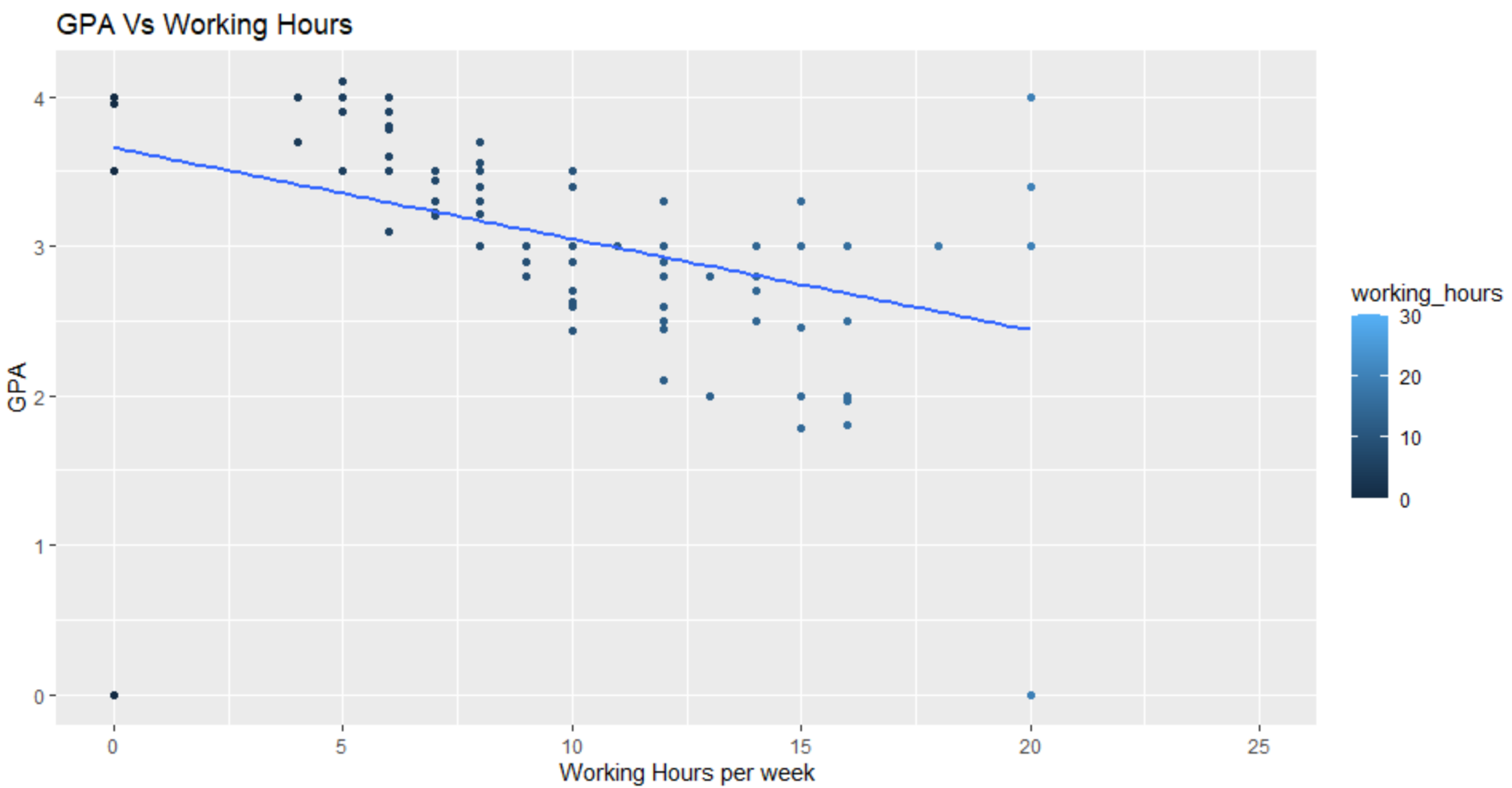


Figure 16 GPA vs Working hours graph

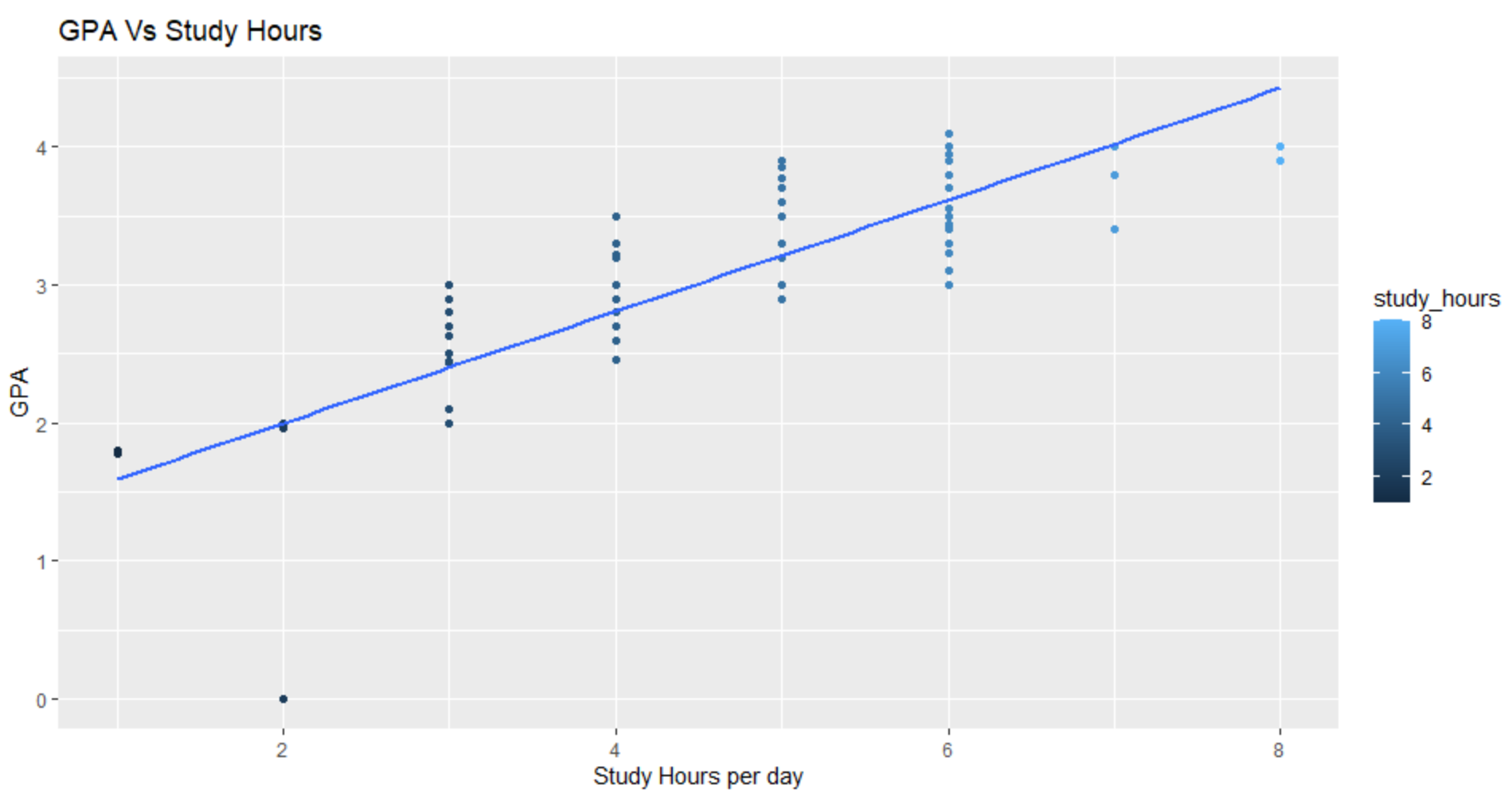


Figure 17 GPA vs Study hours graph

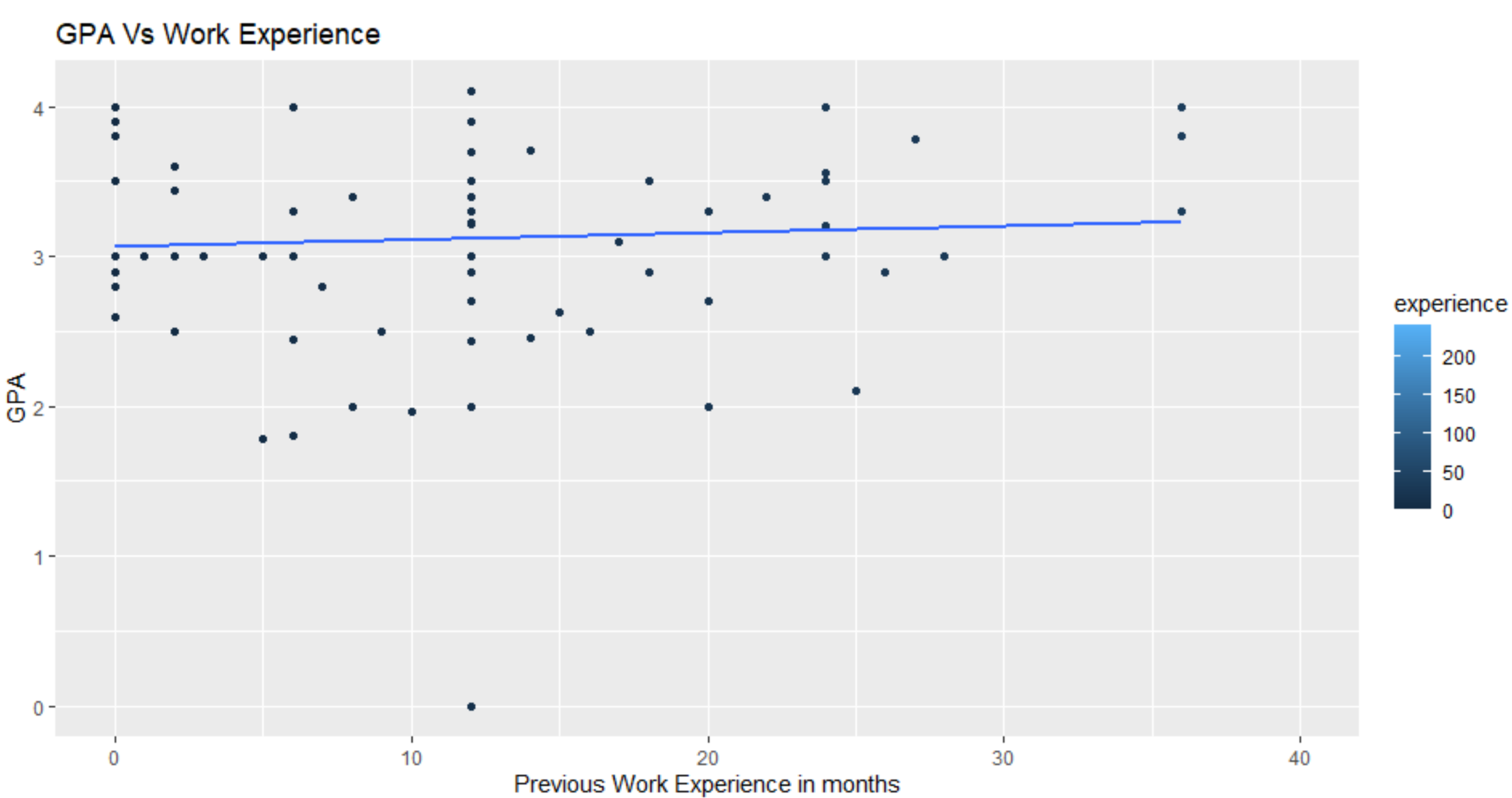


Figure 18 GPA vs Work experience graph

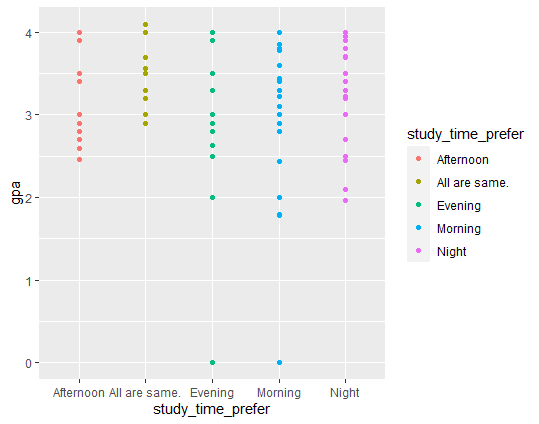


Figure 19 GPA vs Study time preference graph

GPA of Langara students depends upon all above mentioned factors. After observing figures, when we analyze all factors individually, we came to know that transit time, social media doesn’t affect GPA at large extent because for these factors we got a constant plot.

However, study hours showed a very positive linear relationship and game time also showed slightly positive relationship. Also, we can observe working hours and sleep time shows a negative linear relationship.

During the analysis of study place preference, personal relationships and previous work experience, we found GPA doesn’t depend upon these factors.

|  |  |  |  |
| --- | --- | --- | --- |
| Predictors | P-Value | R-Squared | Adjusted R-Squared |
| Transit Time | 0.98 | 1.67e-05 | -0.01019 |
| Social Media Usage Time | 0.77 | 0.0009125 | -0.009282 |
| Playing Games Time | 0.07 | 0.03333 | 0.02337 |
| Sleep Time | 0.0002 | 0.1306 | 0.1217 |
| Working Hours | 0.012 | 0.06301 | 0.05345 |
| Study Time Preference | 0.203 | 0.06 | 0.02043 |
| Study Place Preference | 0.994 | 0.03859 | -0.1067 |
| Study Hours | 2.2e-16 | 0.6325 | 0.6287 |
| Personal Relationships | 0.0008 | 0.1987 | 0.156 |
| Previous Work Experiences | 0.95 | 3.802e-05 | -0.01017 |

Table 5 p-value, R-squared and Adjusted R-squared for different predictors

Here, we used stepwise variable selection method in which we found p-values, R-squared and adjusted R-squared for all predictors. From the table, we can see p-values of sleep time, study hours, working hours, gaming time, personal relationships and study time preference are less. Though gaming time and personal relationships show less p-value, the graphs showed inverse

relations than expected in a conventional scenario. Hence, we are excluding these two predicters in our final model.

Therefore, we have finalized with sleep time, study time preference and study hours as predictors to make our final model.

## 3.2 Model building

*“In Regression Analysis, Model Building is the process of developing a probabilistic model that best describes the relation between the dependent and independent variables.”*

From above graphs, we can predict that relationship between response variable and predictors is not curvilinear, so we cannot include quadratic terms in final model.

The first model we had is:

E(y) = β0 + β1x1 + β2x2. + β3x3 + β4x4 + β5x5 + β6x6+ (main terms)

β 7x1x2 + β8x1x3 + β9x1x4 + β10x1x5 + β11x1x6 + (interaction terms for sleep time)

β12x2x3 + β13x2x4 + β14x2x5 + β15x2x6 (interaction terms for study hours)

Where y = GPA (response variable);

x1 = sleep time;

x2 = study hours;

x3 = 1 if all are same, else 0;

x4 = 1 if Evening, else 0;

x5 = 1 if Morning, else 0;

x6 = 1 if Night, else 0;

where x3, x4, x5 and x6 are variables for study time preference.

Here, Afternoon is considered as baseline.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Correlation | GPA | Sleep Time | Study Hours | Working Hours | Experience |
| GPA | 1.00 | -0.36 | 0.80 | -0.25 | -0.01 |
| Sleep Time | -0.36 | 1.00 | -0.29 | 0.05 | 0.02 |
| Study Hours | 0.80 | -0.29 | 1.00 | -0.38 | 0.09 |
| Working Hours | -0.25 | 0.05 | -0.38 | 1.00 | -0.14 |
| Experience | -0.01 | 0.02 | 0.09 | -0.14 | 1.00 |

Table 6 Table Summary of correlation test with each variable.

Correlation matrix analysis is very useful to study dependences or associations between variables. From table we can observe that multicollinearity does not exist between the predictors we considered for the model.

So, the **Fitted Equation** will be:

E(y)= 6.13 - 0.63x1 - 0.55x2 + 0.75x3 - 1.71x4 + 0.50x5 + 0.76x6 +

0.124x1x2 - 0.115x1x3 + 0.163x1x4 + 0.003x1x5 – 0.08x1x6 +

0.021x2x3 + 0.107x2x4 + 0.116x2x5 - 0.02x1x6

**Check for model Utility**:

**Hypothesis testing:**

H0:

At least one of the values is non-zero.

**F-test:**

F=17.42

F-value for α=0.05, df1=15, df2=101-15=84

* F0.05,15,60=1.84
* F0.05,15,94=1.77
* 1.77<F0.05,15,84<1.84

As, F > Fα Hence, we can reject null Hypothesis and conclude that at least one of β is non-zero. So, this model is adequate.

If we consider sleep time and study hours don’t depend upon study time preference, then we can remove interaction terms between sleep time-study time preference and study hours-study time preference.

Hence the **Reduced model** is:

E(y)= β0 +β1x1+β2x2 + β3x3+β4x4+ β5x5+ β6x6+β7x1x2

E(y)= 5.77 - 0.62x1 - 0.44x2 + 0.04x3 +0.04x4 + 0.05x5 + 0.12x6 +0.115x1x2

**Comparing complete model and reduced model:**

Now, Check for model Utility:

**Hypothesis Testing**

H0:

At least one of the values is non-zero.

**Test statistic:**

F = [SSE(r)-SSE(f)/k-g]/[SSE(f)/n-(k+1)]

= 0.8911

F-value for α=0.05, df1=k-g=8, df2=101-16=85

* F0.05,8,80=2.06
* F0.05,8,90=2.04
* 2.04<F0.05,8,85<2.06

As, F < Fα, Hence, we fail to reject null Hypothesis and conclude that Interaction terms for sleep time-study time preference and study hours-study time preference doesn’t play important role for prediction or estimation of GPA. So, we can neglect these terms.

Now, check for interaction between sleep time and study hours.

**Hypothesis Testing:**

H0:

Ha:

Test statistic = 5.108

* tα/2,df=t0.025,60=2
* t0.025,120=1.98
* 1.98<t0.025,92<2

As t > |tα/2|which includes that we can reject null hypothesis and include alternative hypothesis. This conclude that interaction term for sleep time-study hours plays important role for finding GPA so we can’t reject that term.

So, our **final model** is:

E(y)= β0 +β1x1+β2x2 + β3x3+β4x4+ β5x5+ β6x6+β7x1x2

Now, check for **adequacy of final model by F-test**:

**Hypothesis Testing:**

H0:

Ha: At least one of the values is non-zero.

**F-test:**

F=36.44

F-value for α=0.05, df1=7, df2=101-7=92

* F0.05,90=2.11
* F0.05,7,100=2.10
* 2.10<F0.05,7,92<2.11

As F > Fα, Hence, we can reject null Hypothesis and conclude that at least one of β is non-zero. So, this model is adequate.

Multicollinearity Test:

Co-relation coefficient for predictors:

|  |  |  |  |
| --- | --- | --- | --- |
| Predictors | Sleep Time | Study Hours | GPA |
| Sleep Time | 1 | -0.28 | -0.36 |
| Study hours | -0.285 | 1 | 0.79 |
| GPA | -0.36 | 0.79 | 1 |

Table 7 Correlation coefficients

VIF values:

|  |  |  |  |
| --- | --- | --- | --- |
| Variable(s) | Sleep Time | Study Hours | Study Time Preference |
| VIF values | 1.0552 | 1.1152 | 1.0633 |

Table 8 VIF values table

From above observations, we come to know that multicollinearity is not present between predictors it means variables are not corelated. Also, for variable selection method since sample size n > number of predictors, we choose backward regression here. So that the full model can be fit. Since multicollinearity does not exist and there is colinear relation, variable transformation is not required here.

## 3.3 Residual Analysis**:**

We can use our model for prediction and estimation, if these three assumptions have been met. The three assumptions to be addressed are:

1. Lack of fit (Linearity)
2. Homoscedasticity
3. Normality

**Lack of fit:**

We can check for this assumption by plotting residual plots. If we get any particular pattern in plot, it shows this assumption of E(€)=0 hasn’t been met.

The residual plots showed no patterns or trends here. We can observe four outliers here but since 95% of the data are within two standard deviations, our model does not exhibit any lack of fit.

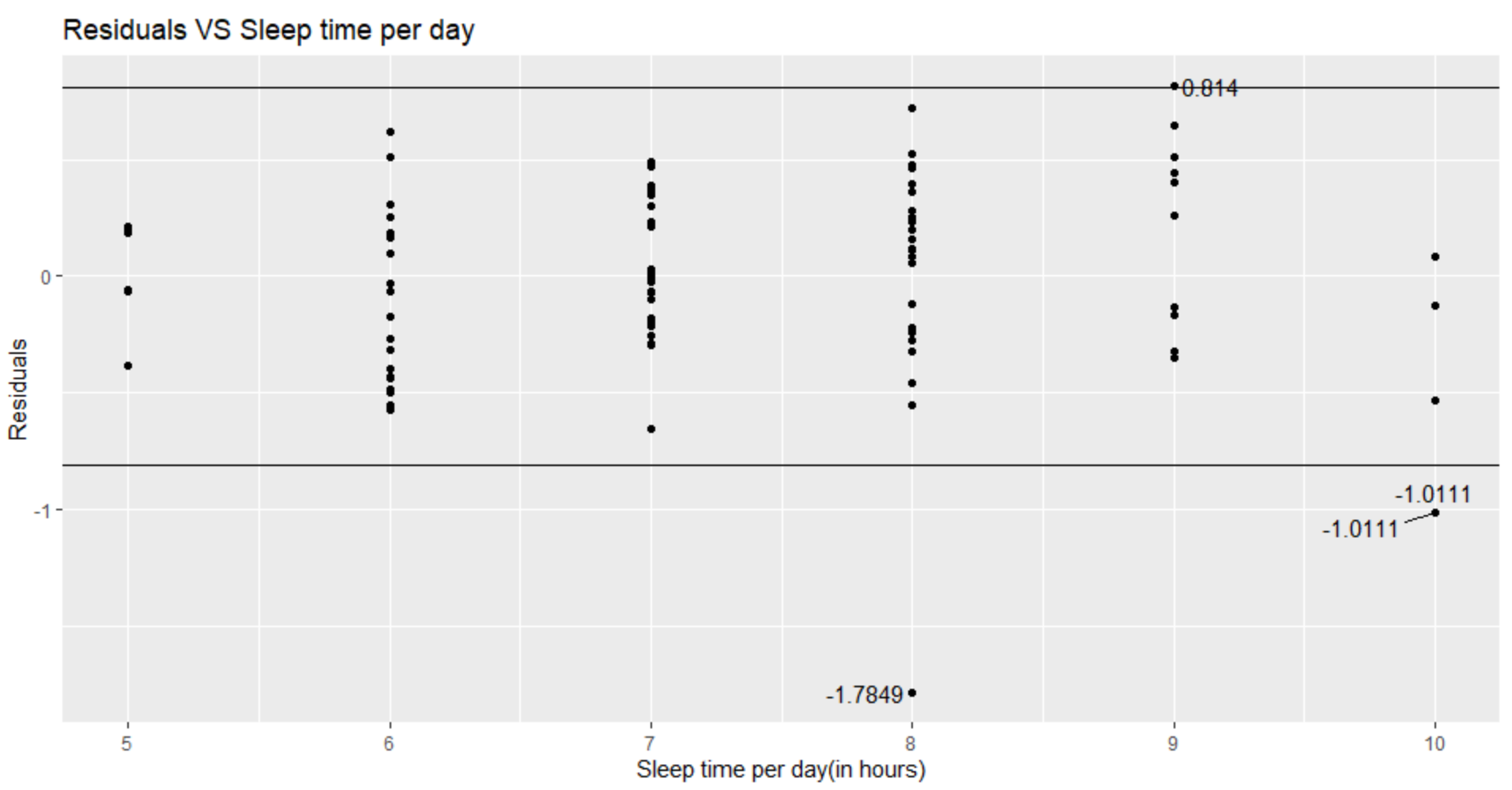


Figure 20 Residuals vs Sleep time graph

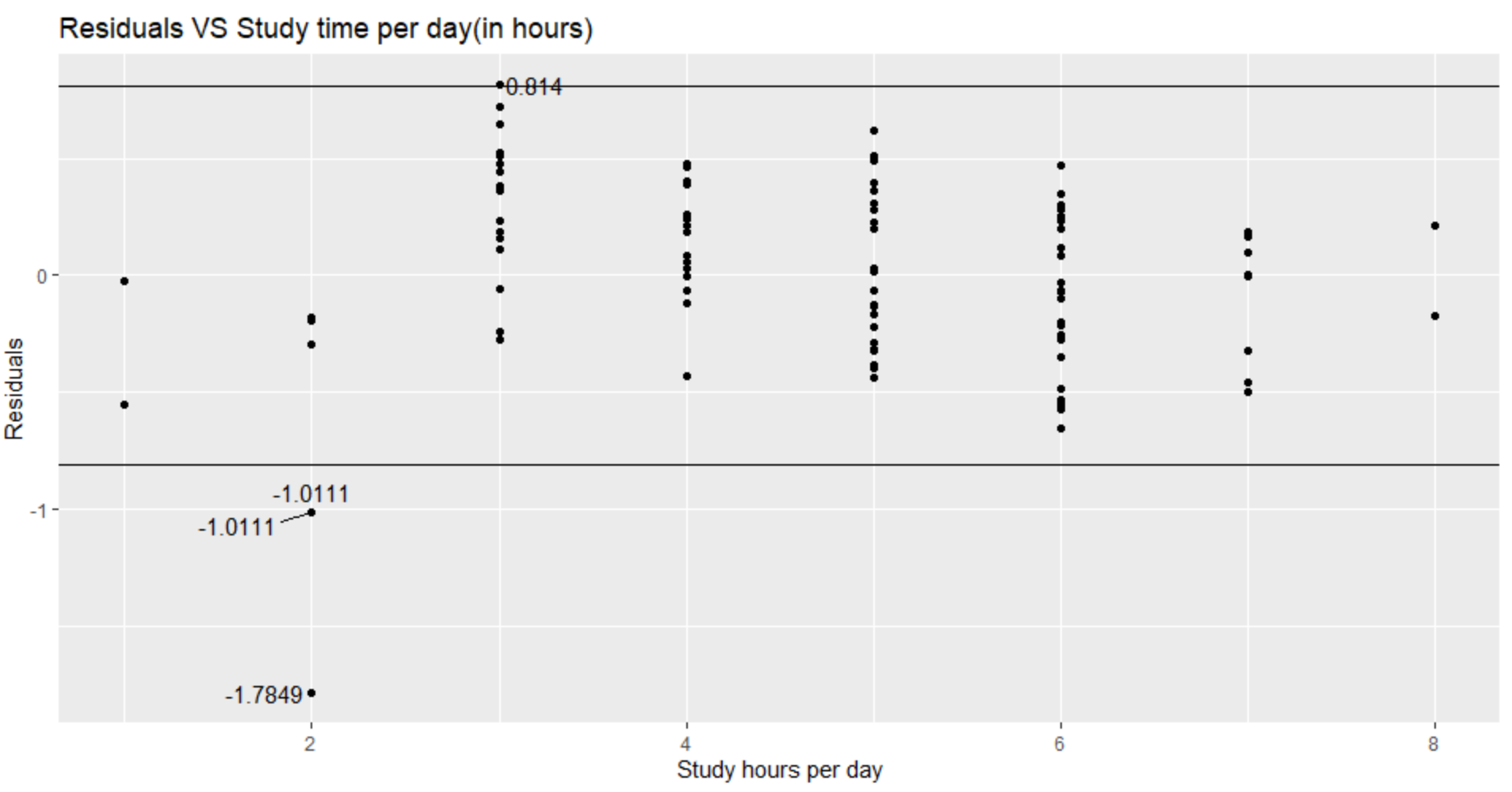


Figure 21 Residuals vs Study time graph

These residual plots don’t show any pattern, so we can conclude that this assumption has been met.

**Homoscedasticity:**

We can check for this assumption by plotting residuals with predicted values. If we get equal variance in graph, it concludes that his assumption of equal variance has been met.

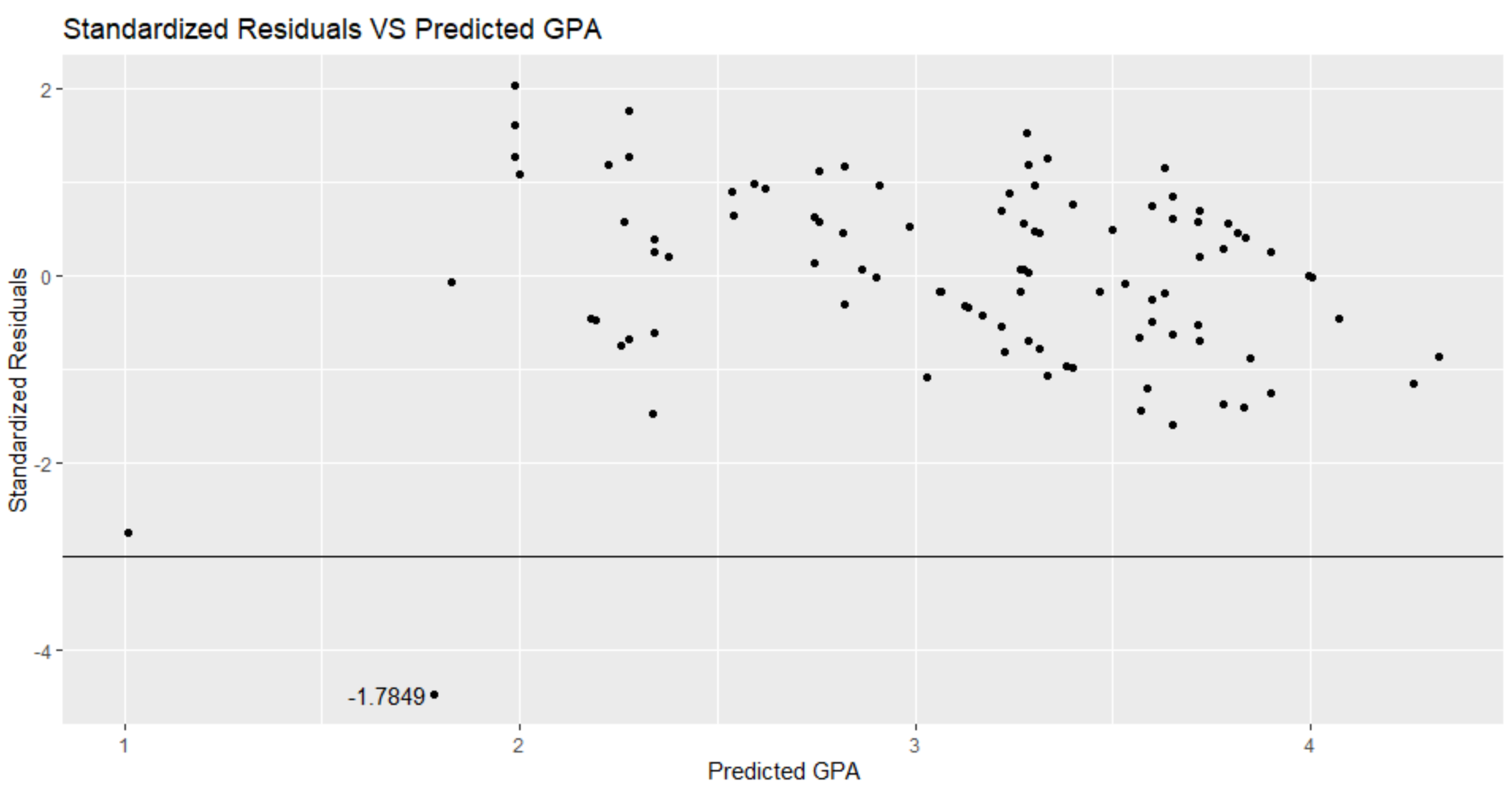


Figure 22 Standard Residuals vs Predicted GPA graph

This graph shows equal variance between -2s to +2s (s-standard deviation). Hence, we can say that this model has been met with Homoscedasticity assumption.

**Normality:**

We can check this assumption by plotting normal probability plot, if we get normal plot, it concludes that this assumption has been met.

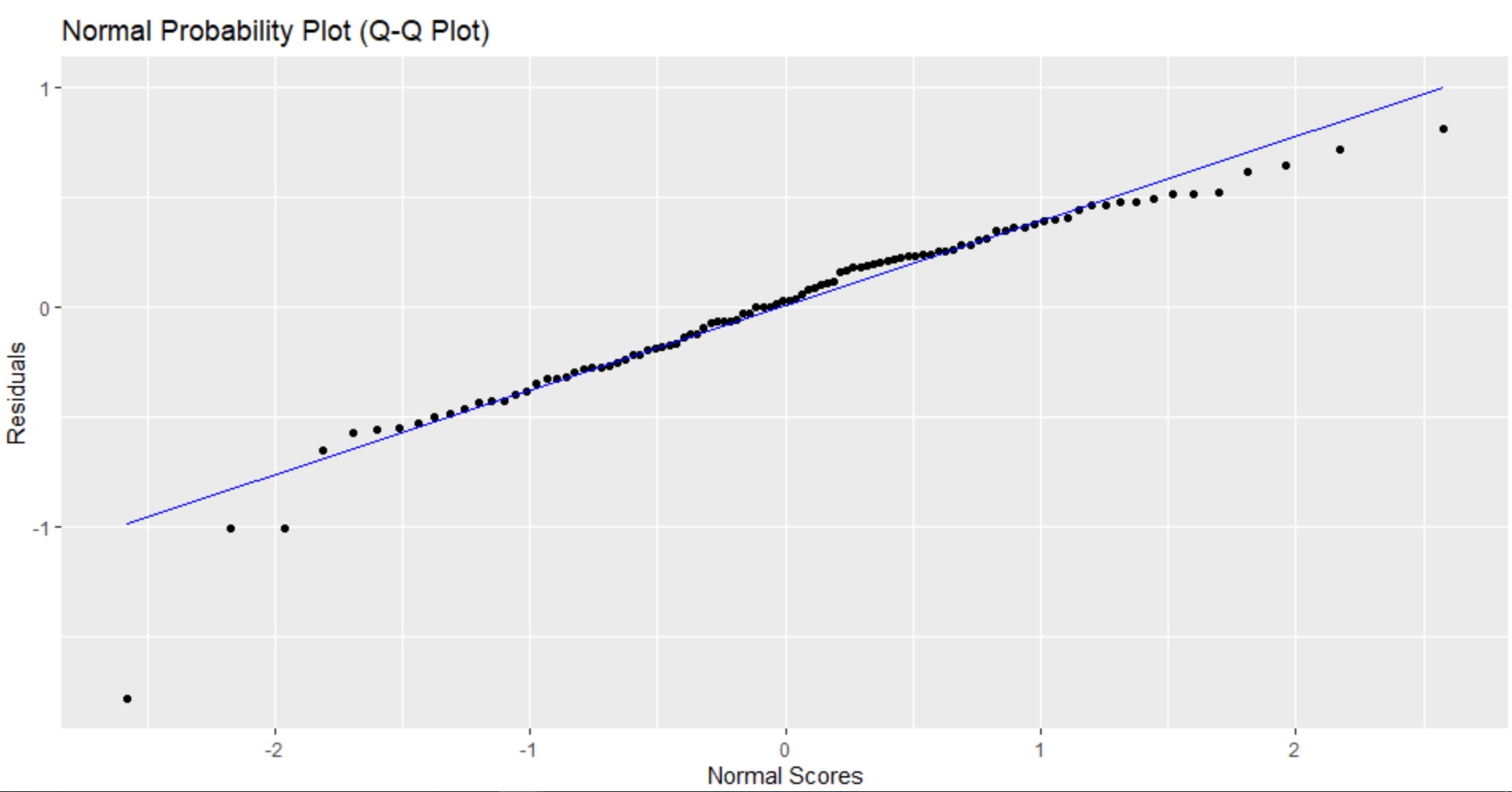


Figure 23 Normal Probability Plot (Q-Q Plot)

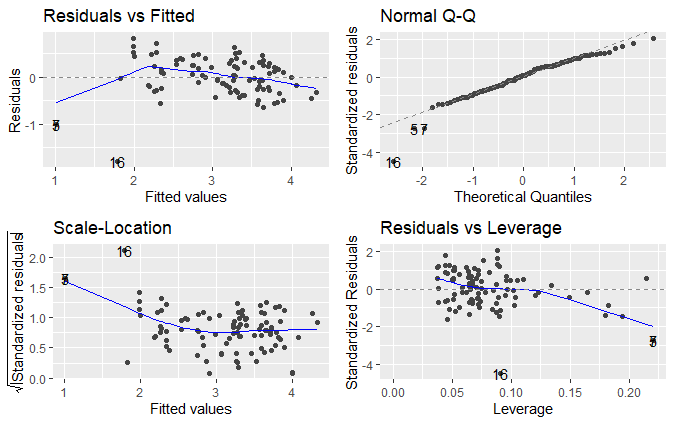
These graphs are almost normal, so we can conclude that our model has been met normality assumption.

Figure 24 Standardized residuals vs theoretical quantiles

From all plots, we can conclude that our model is valid because our model satisfies all 3 assumptions. However, from all plots we can see very little outliers. Although, these few outliers cannot affect validity of our model, even then we can remove these outliers to get a perfect model as shown.

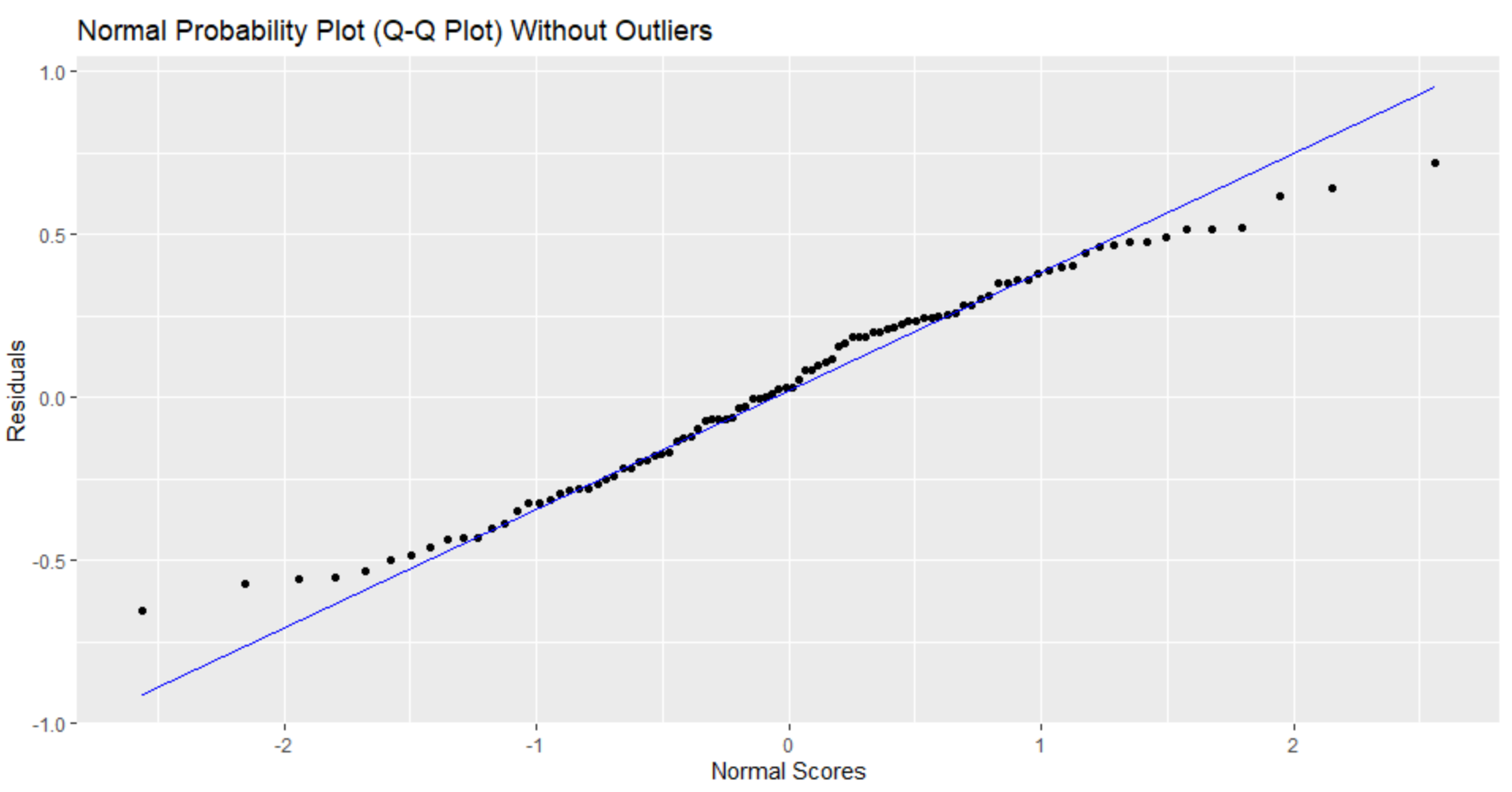


Figure 25 Residuals vs Normal scores (without outliers)

After removing all outliers, we get this normal probability plot to ensure that all outliers were removed successfully. Hence, from this plot, we can see that there is no outlier, and this is perfectly normal plot.

## 3.4 Estimation and Prediction

Our Reduced Model:

**E(y)= β0 +β1x1+β2x2 + β3x3+β4x4+ β5x5+ β6x6+β7x1x2**

**E(y)= 5.77 -0.62x1-0.44x2 + 0.04x3+0.04x4+ 0.05x5+ 0.12x6+0.115x1x2**

We use the above model and using different values for x1, x2, x3, x4, x5 and x6.

where y is GPA (response variable), x1 is sleep time, x2 is study hours, (x3, x4, x5, x6) are variables for study time preference.

x3= {1-All are Same,0-Otherwise

x4= {1-Evening,0-Otherwise

x5= {1-Morning,0-Otherwise

x6= {1-Night,0-Otherwise

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sleep time (Hours) | Study Hours | Study Time Preference | Estimated GPA | 95% Confidence Interval | GPA values as in the Student Survey |
| 6 | 1 | Morning | 2.34 | 1.97-2.70 | 1.78 |
| 7 | 2 | Evening | 2.18 | 1.91-2.45 | 2.00 |
| 8 | 3 | Night | 2.34 | 2.13-2.56 | 2.50 |
| 7 | 4 | Night | 2.98 | 2.80-3.17 | 3.20 |
| 6 | 5 | Morning | 3.34 | 3.15-3.52 | 3.38 |
| 8 | 6 | Morning | 3.72 | 3.51-3.93 | 3.74 |
| 6 | 7 | Night | 4.26 | 4.01-4.51 | 3.80 |

Table 9 Estimated values and 95% CI for GPA

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sleep time (Hours) | Study Hours | Study Time Preference | Predicted GPA | 95% Predicted Interval | GPA values as in the Student Survey |
| 6 | 1 | Morning | 2.34 | 1.43-3.24 | 1.78 |
| 7 | 2 | Evening | 2.18 | 1.30-3.05 | 2.00 |
| 8 | 3 | Night | 2.34 | 1.48-3.20 | 2.50 |
| 7 | 4 | Night | 2.98 | 2.13-3.84 | 3.20 |
| 6 | 5 | Morning | 3.34 | 2.48-4.19 | 3.38 |
| 8 | 6 | Morning | 3.72 | 2.86-4.58 | 3.74 |
| 6 | 7 | Night | 4.26 | 3.03-4.76 | 3.80 |

Table 10 Predicted values and 95% CI for GPA

The above two tables give the estimated and predicted values for GPA for certain values of the three variables**: sleep\_time, study\_hours** and **study\_time\_prefer** and compare with the actual values we have in the survey. The confidence interval for prediction is wider than the estimation confidence interval because the former includes the uncertainty in estimation of the value as well as the random variation of individual values. The individual value of GPA will lie in the prediction interval 95% of the times.

# 4.0 Limitations

We used simple random sampling method because it was easy to collect the required data with this method. Although Stratified sampling is mode accurate, even then we could not follow that as this could be complicated and more time-consuming process. On the other hand, we also had option of cluster sampling method to collect data, but cluster sampling is more expensive than other methods and we don’t have enough budget for that.

It was found that we would face Response Bias in this project as the response given by some students were biased and vague. We also expect the respondents to not answer a few questions. We were fortunate to have multiple variables to conduct the analysis.

We got some unexpected results because every student has different opinion and different way of studying. We were expecting to get inverse relation between GPA and gaming time. However, during analysis we come to know that if students play more games, their GPA get increased. We found some outliers in collected data because of wrong information given by some students.

# 5.0 Conclusion

The student’s GPA depends upon sleep time, study hours and study time preference. Other factors such as transit time, social media usage time, playing time, working hours, study place preference, personal relationships and previous work experience which were initially thought to be important predictors of GPA turned out to be insignificant in our model. Considering all these factors, students can increase their grades by properly adjusting their study hours, sleep time and study time.

# References

"What Is A GPA And Why Is It So Important? - Mastersportal.Com". 2020. *Mastersportal.Com*. https://www.mastersportal.com/articles/2126/what-is-a-gpa-and-why-is-it-so-important.html.

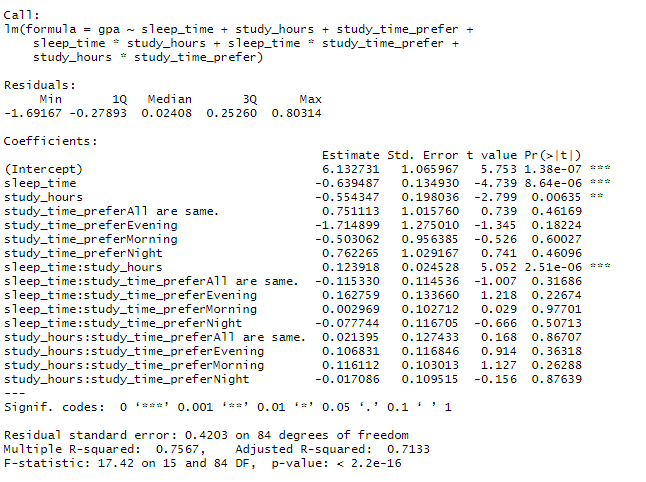
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3959895/>

Hess, Abigail. 2020. "Students Who Work Actually Get Better Grades—But There's A Catch". *CNBC*. <https://www.cnbc.com/2017/10/04/students-who-work-actually-get-better-grades-but-theres-a-catch.html>.

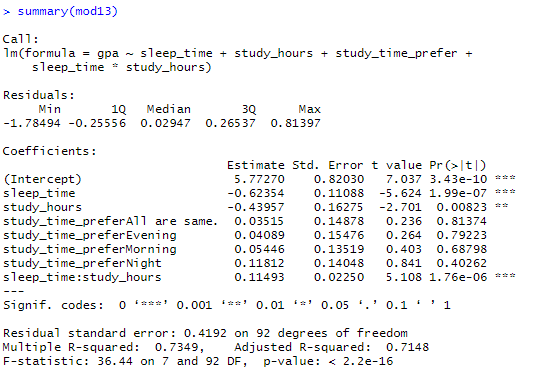
# Appendix I

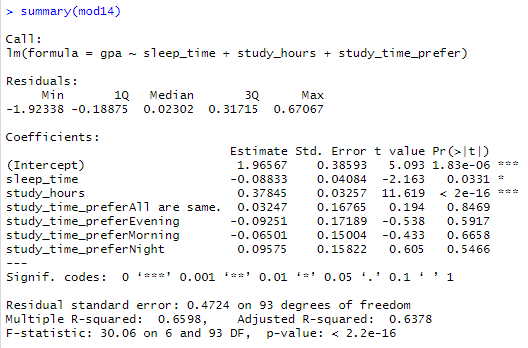
## R outputs for model building

>summary(mod12)



Output for model without interaction terms of study time preference.



Output for model with no interaction terms; just main terms with sleep time, study hours and study time preference. To test the significance of interaction terms.



Output of partial F-test to check whether interaction of study time preference is significant.

