Final Report

Project Members: Emma Gagnon, Jeimi Reyes, Isabella Demers, Karishma Christmas, Shannon Fuechslin

Professor Simone

Course: STA4164

Section: 0001

Motivation:

Prior research indicates that increased social media usage can be a response to feelings of poor familial connectedness (Winstone, 2021). Studies like this display the need for more research into why people use social media and what variables can effectively predict feelings of connectedness. Our objective in using this dataset collected by UCF researchers is to evaluate how three different types of motivation for using social media (SM), along with age and sex, impact social connection score. The three different types of motivation are information, connection, and friendship. Additionally, we would like to assess an interaction term between age and total social media use to determine the significance of their interaction in determining connection score. Our main question is: If we test age, sex, and type of motivation as predictors how would it affect social connection score? This data is useful for our main interest because we wanted to examine social media in our peers' lives. Our interest peaked upon reading this article because each of us was able to relate to using social media for connection. Gen Z spends about three hours per day using social media and we wanted to dive deeper to figure out what most peers our age use it for (Lindner, 2023).

Data Description:

This study was conducted by Allison E. Garibaldi, Grace E. Waldfogle, and James L. Szalma at the University of Central Florida in 2023. The study has 365 observations and 15 variables. This dataset examines a multitude of variables that focus on social media's impacts on relationships and time usage. The researchers also looked at people's social media use on each social media platform (Twitter, Snapchat, Instagram, and Facebook), affiliation importance, social connection, disconnection, and their motives for using social media. They identified three motives for using social media: friendship, connection, and information. Friendship explains

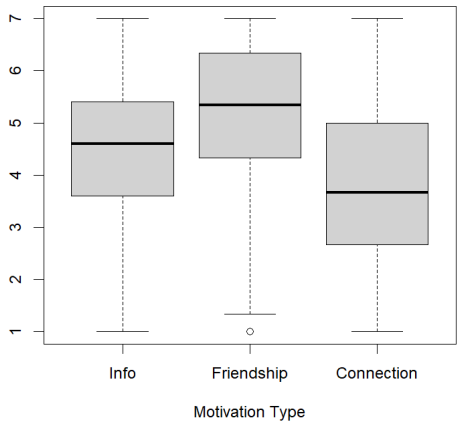
those who use social media to keep in touch with friends. Connection explains the use of social media to make new friends. Information refers to the use of social media for academic purposes or to share information.

There were 400 participants (356 after removing participants with missing data), 118 participants were male and 234 were female. Participants' ethnicity/race was also surveyed, with ages ranging from 18 to 49 years old. The study's data was gathered through a signup link from SONA and Qualtrics. Participants were told that they would be a part of a study that examines social media use. The researchers examined the participants' aspiration index, reasons for using SM, relatedness need satisfaction, and how much time they spent on SM. To evaluate their connection and disconnection (relatedness need satisfaction), the researchers used a six-item scale that included three positively worded items and three negatively worded items. They then asked participants to rate how much they agreed with the statement over the past week. To examine the reason for their SM use, they used an 11-item self-report scale that included use for friendship, connection, and information. Participants rated on a seven-point scale how much each statement resonates with them. To evaluate the quantity of social media use, participants were asked to report how many hours they spend on social media as a whole, and how many hours they spend on each social media platform. Considering all of the variables researched, this article was useful in investigating the relationship between information motivation, friendship motivation, connection motivation, age, sex, and total social media use, using social connection score as a response variable.

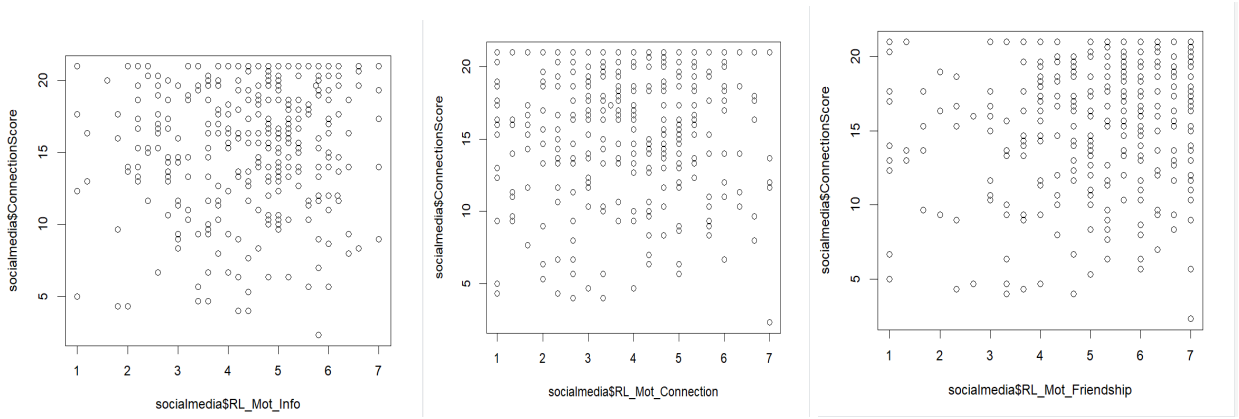
Model Diagnostics:

```
summary(socialmedia$RL_Mot_Info)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   1.00    3.60    4.60    4.41    5.40    7.00       2
summary(socialmedia$RL_Mot_Friendship)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  1.000   4.333   5.333   5.088   6.333   7.000       2

prop.table(table(socialmedia$Sex, exclude = NULL))

         1          2       <NA>
0.33146067 0.65730337 0.01123596
(1 is male and 2 is female)
```

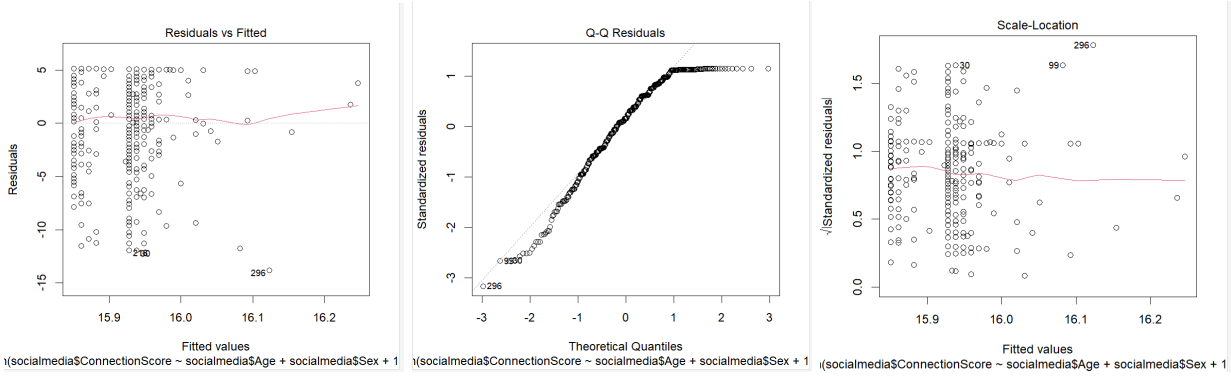
```
summary(socialmedia$Sex)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  1.000   1.000   2.000   1.665   2.000   2.000       4
summary(socialmedia$SNS_Total)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  0.000   3.000   4.000   4.974   7.000  18.000       8
summary(socialmedia$RL_Mot_Connection)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  1.000   2.667   3.667   3.790   5.000   7.000       2
summary(socialmedia$ConnectionScore)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  2.333  13.333  16.667  15.942  19.667  21.000       4
summary(socialmedia$Age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  18.00   18.00   19.00   20.05   20.00   49.00       9
```

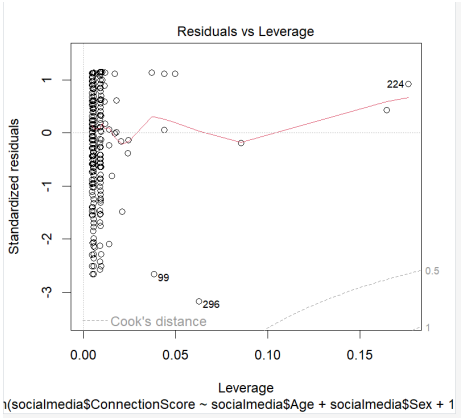


Boxplots for the three different motivation types.



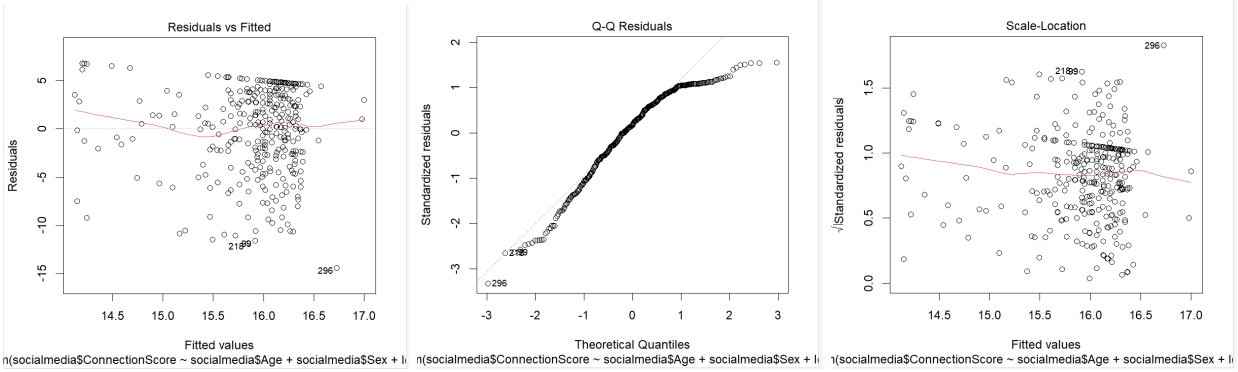
Scatterplots for each motivation type indicate linearity issues.

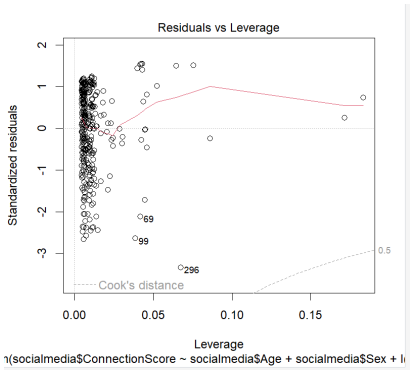




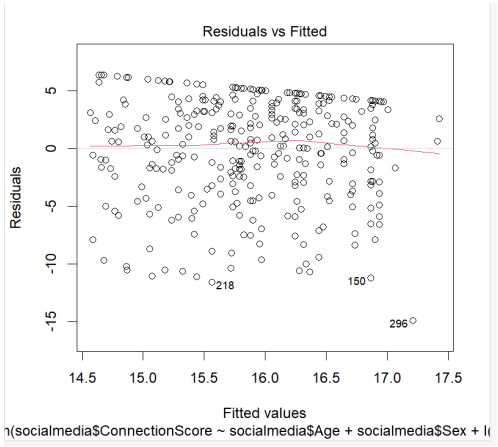
These are the plots for the model with no transformations.

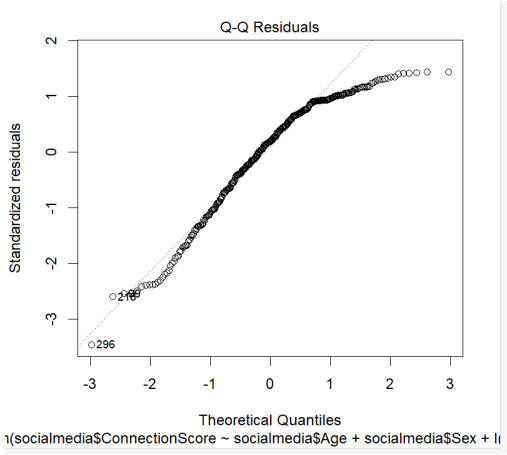
Using basic analytics for our model we can analyze our Q-Q plot and assume that normality is violated. Since the data points appear to veer downwards towards the end of the Q-Q we predict that we will have to apply a transformation (square root, log, or squared). Additionally, after looking at the graphs (Residuals vs. Fitted, Q-Q Residuals, Scale-Location, and Residuals vs. Leverage) there appear to be outliers, one of which being observation 296. Below are the plots of all the transformations we tried, but we ultimately used the log(x) transformation on friendship motivation. We did not try a Box-Cox transformation because we were advised that it probably would not fix the assumption issues for our dataset.



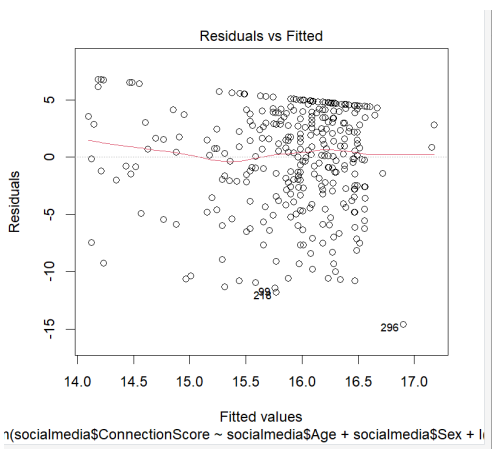


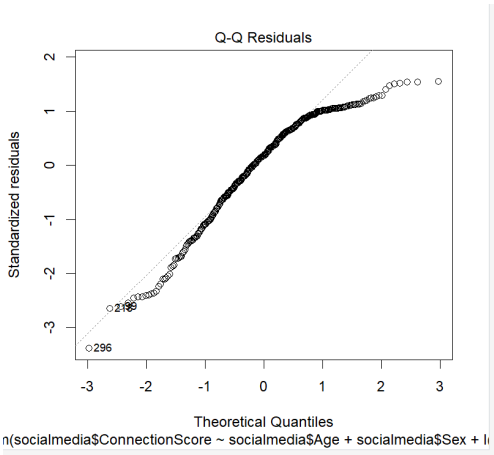
These are the plots for the model with the log(x) transformation of the friendship motivation variable.



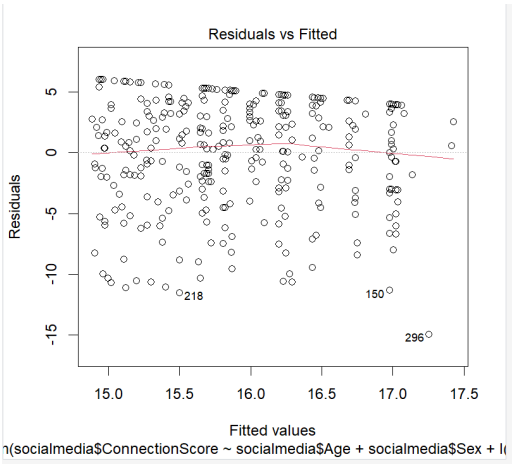


Plots for the model with the squared transformation of the friendship motivation variable.

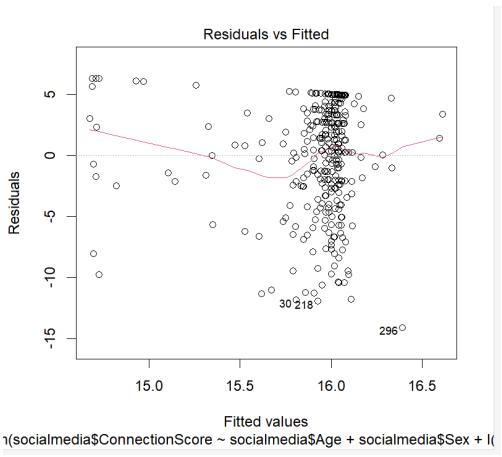




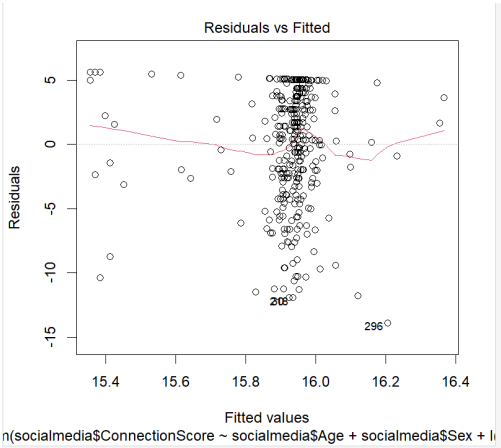
Plots for the model with the square root transformation of the friendship motivation variable.



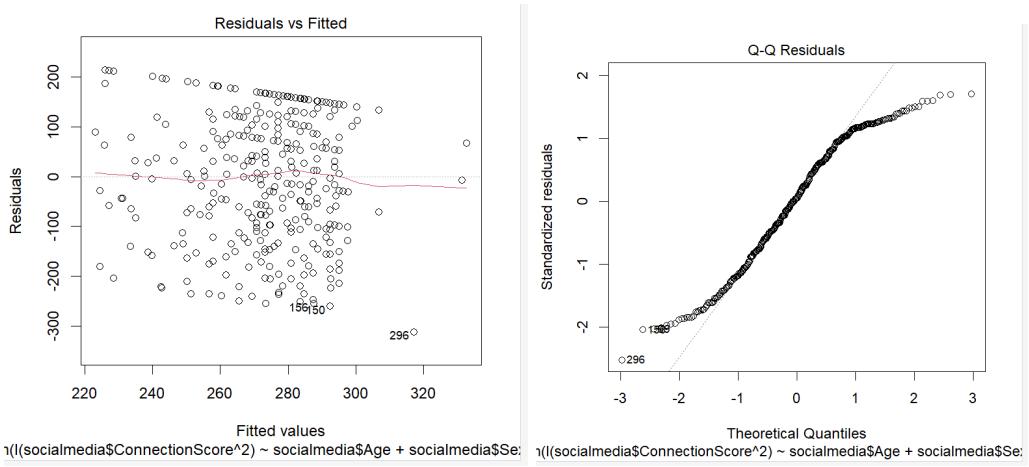
Plots for the model with the cubed transformation of the friendship motivation variable.



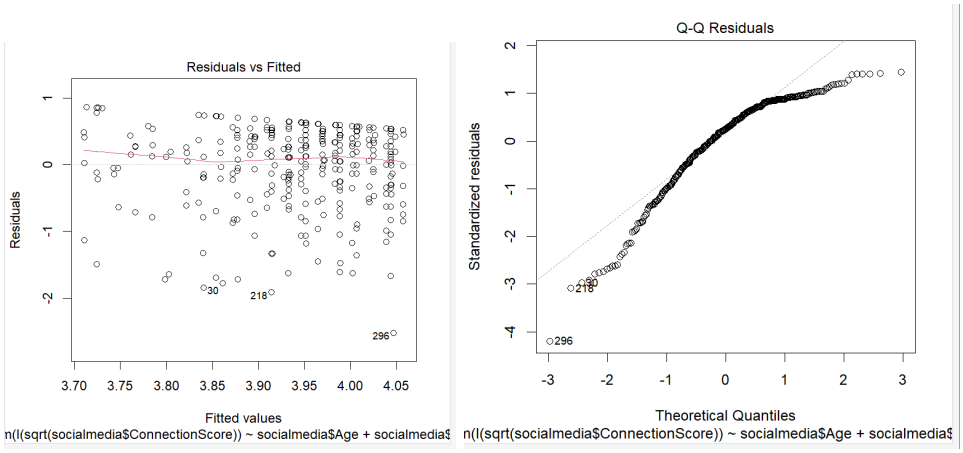
Plots for the model with the 1/x transformation of the friendship motivation variable.



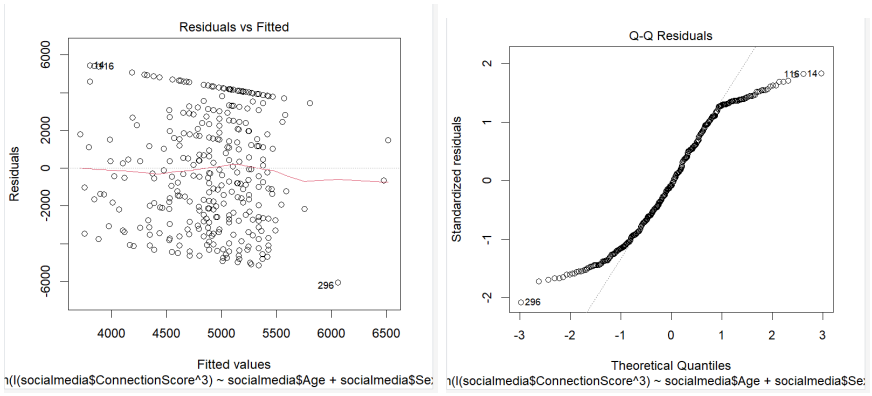
Plots for the model with the 1/x^2 transformation of the friendship motivation variable.



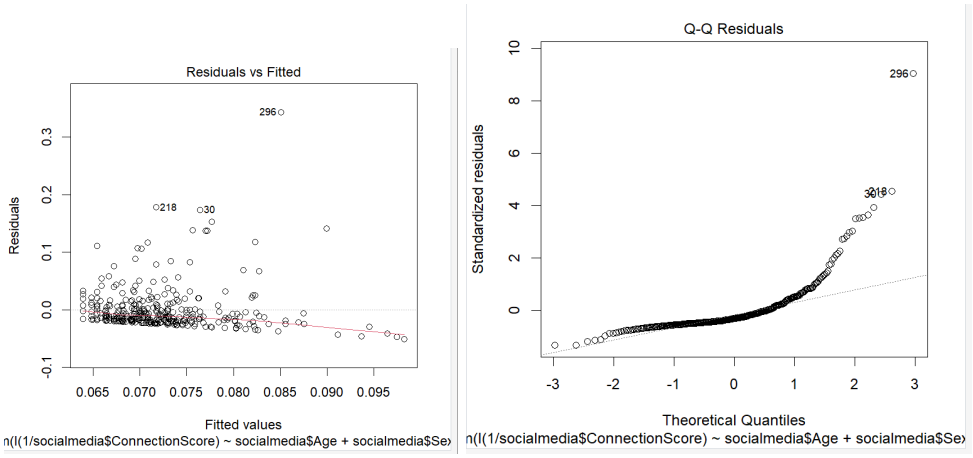
Plots for the model with the squared transformation of the connection score variable.



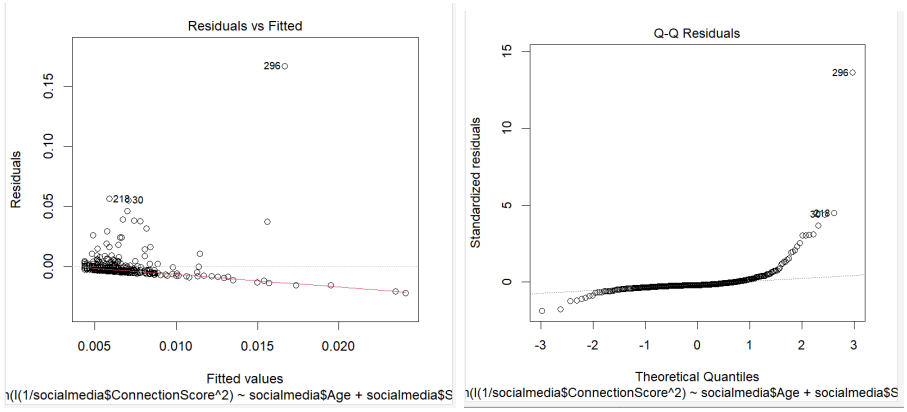
Plots for the model with the square root transformation of the connection score variable.



Plots for the model with the cubed transformation of the connection score variable.



Plots for the model with the 1/y transformation of the connection score variable.



Plots for the model with the 1/y^2 transformation of the connection score variable.

To check for any outliers in our data we tested for cook's distance, leverage and jackknife residuals. From the cook's distance we compare the points to 1 from the leverage plot. After conducting this test we didn't find any observations that exceeded 1 so it gave us no outliers. Testing for leverage we compare the points to the calculated leverage value 0.0898. From this we find that the two points, 255 and 224, both have higher leverages causing them both to be considered outliers. Finally from conducting jackknife residuals, we compare the points to the calculated value 1.96, finding 19 observations to have a greater value than 1.96 causing them to be considered outliers.

```
cor(socialmediadf, use = "complete.obs")
```

None of the correlations in the correlation matrix were higher than 0.8, so indicates no collinearity issue.

Model Selection:

```
                                   Pr(>|t|)
(Intercept)                        5.72e-12 ***
socialmedia$Age                       0.791
socialmedia$SNS_Total                 0.942
socialmedia$Age:socialmedia$SNS_Total    0.894
---
```

Testing for interaction between age and total social media use, the p-value is greater than alpha at 0.05, therefore we fail to reject HO & conclude there is no interaction between age and total social media use.

"model9" Connection Motivation AIC = 2000.036
"model8" Friendship Motivation AIC = 1993.609
"model7" Information Motivation AIC = 1998.552
Since the lowest AIC indicates the best model, we would select "model8" as our final model, which in our code had the variables for friendship motivation, age, and sex.

For our full model, we ultimately considered our variables age, sex, friendship motivation, and connection score as our response variable. After transforming the data using log, squared, square root, cubed, 1/x, and 1/x^2 we tried to find the best full model. We decided that the log(x) transformation was by far the best one because the residual plots looked better than any other transformation. The residual plot improved, however, the normality plot got worse with this transformation. With the assumptions in mind, the log model violated independence, linearity, homoscedasticity, and normality, but was still far better than other transformations.

Results Summary and Interpretations:

Our new model transformed with log(x) was our best one by far, although it did violate normality. Our final model used age, sex, and the log of friendship motivation to predict the connection scores. From the overall significance test, we determined that age, sex, and the log of friendship motivation do not predict connection score.

Parameter estimates:

$\widehat{\beta_0}$ = Not interpretable

$\widehat{\beta}_1$ = We expect the connection score to increase by 0.02286 for every one year increase in age, holding sex and friendship motivation constant.

$\widehat{\beta}_2$ = The predicted connection score of a female is 0.07245 units lower than the predicted connection score for a male, holding age and friendship motivation constant.

$\widehat{\beta}_3$ = We expect the connection score to increase by 1.11491 for every one unit increase in friendship motivation, holding age and sex constant.

```
Call:
lm(formula = socialmedia$ConnectionScore ~ socialmedia$Age +
    socialmedia$Sex + I(log(socialmedia$RL_Mot_Friendship)),
    data = socialmedia)

Residuals:
     Min       1Q   Median       3Q      Max
-14.3940  -2.6133   0.7352   3.8098   6.8007

Coefficients:
                                   Estimate Std. Error t value
(Intercept)                        13.86360    1.72208   8.051
socialmedia$Age                     0.02268    0.06498   0.349
socialmedia$Sex                    -0.07245    0.52265  -0.139
I(log(socialmedia$RL_Mot_Friendship))  1.11491    0.55607   2.005
                                   Pr(>|t|)
(Intercept)                        1.43e-14 ***
socialmedia$Age                      0.7273
socialmedia$Sex                      0.8898
I(log(socialmedia$RL_Mot_Friendship))   0.0458 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.478 on 337 degrees of freedom
  (15 observations deleted due to missingness)
Multiple R-squared:  0.01194,   Adjusted R-squared:  0.003147
F-statistic: 1.358 on 3 and 337 DF,  p-value: 0.2556
```

After conducting the overall f test we conclude that the variables aren't significant in predicting connectedness, with the p-value 0.2556 > alpha at 0.05. In simpler terms, this means our hypothesis was incorrect and the variables we thought would be good at predicting connection scores are not actually.

Adjusted R^2 = 0.003147

About 0.3147% of the variation in Connection Score is explained by Age, Sex, and Friendship Motivation. The coefficient of determination, R^2, helps to determine the influence the independent variables have on the dependent variable. A low R^2 indicates that the influence of the independent variables is low, which means it is not good.

Conclusion and limitations:

For our study, we were able to find that age, sex, and friendship motivation as predictors do not influence social connection score. This conclusion means that our data was useful in solving our main question from the journal article. For limitations, our project was difficult due to the issues with the final log model violating independence, linearity, homoscedasticity, and normality. Also, our response variable seemed discrete the way it presented itself in the scatterplots, but was a continuous variable. As for improvements, we could try a nonparametric alternative for the multiple linear regression. We learned how to apply our knowledge on producing a multiple linear regression in R to a real life scenario. We got first hand experience on trying to analyze a journal article's data and how to transform the data while picking out specific variables. We also got hands-on experience with working as a team to come up with solutions from difficulties in R and the full model selection. Our full model transformations weren't perfect and without violations of assumptions, so it made it difficult to make a final selection. However, working with a team allowed each of us to express our knowledge of statistics and R to make a decision on which full model transformation to choose.

Works Cited

Auxier, B., & Anderson, M. (2021). Social Media Use in 2021. In Policy File. Pew Research Center. https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2021/04/PI_2021.04.07_Social-Media-Use_FINAL.pdf

Garibaldi, A. E., Waldfogle, G. E., & Szalma, J. L. (2023). Motivational Underpinnings of Social Media Use. Psychology of Popular Media, 12(4), 434–449. https://doi.org/10.1037/ppm0000431

Lindner, J. (2023, December 16). The most Surprising Gen Z Social Media Statistics and Trends in 2024. Gitnux. GITNUX. https://gitnux.org/gen-z-social-media-statistics/

Winstone, L., Mars, B., Haworth, C. M. A., & Kidger, J. (2021). Social media use and social connectedness among adolescents in the United Kingdom: a qualitative exploration of displacement and stimulation. BMC Public Health, 21(1), 1–1736. https://doi.org/10.1186/s12889-021-11802-9

We, the project team members, certify that below is an accurate account of the percentage of effort contributed by each team member in the project and report

Project Team Member	Percentage of Total Effort
Jeimi	20%
Emma	20%
Isabella	20%
Karishma	20%
Shannon	20%