Final Project: Music vs Mental Health Analysis

Calvin, Vy, Karen

December 2, 2023

Executive Summary

We aim to explore how music impact mental health by analyzing relevant data. This will help us determine how music is the most beneficial for mental health. We'll compare this to the amount of people suffering from certain mental conditions.

Importing the dataset

```
data <- read.csv('mxmh_survey_results.csv')
head(data)
```

```
##            Timestamp Age Primary.streaming.service Hours.per.day While.working
## 1 8/27/2022 19:29:02  18                   Spotify           3.0           Yes
## 2 8/27/2022 19:57:31  63                   Pandora           1.5           Yes
## 3 8/27/2022 21:28:18  18                   Spotify           4.0            No
## 4 8/27/2022 21:40:40  61             YouTube Music           2.5           Yes
## 5 8/27/2022 21:54:47  18                   Spotify           4.0           Yes
## 6 8/27/2022 21:56:50  18                   Spotify           5.0           Yes
##   Instrumentalist Composer        Fav.genre Exploratory Foreign.languages BPM
## 1             Yes      Yes            Latin         Yes               Yes 156
## 2              No       No             Rock         Yes                No 119
## 3              No       No Video game music          No               Yes 132
## 4              No      Yes             Jazz         Yes               Yes  84
## 5              No       No              R&B         Yes                No 107
## 6             Yes      Yes             Jazz         Yes               Yes  86
##   Frequency..Classical. Frequency..Country. Frequency..EDM. Frequency..Folk.
## 1                Rarely               Never          Rarely            Never
## 2             Sometimes               Never           Never           Rarely
## 3                 Never               Never Very frequently            Never
## 4             Sometimes               Never           Never           Rarely
## 5                 Never               Never          Rarely            Never
## 6                Rarely           Sometimes           Never            Never
##   Frequency..Gospel. Frequency..Hip.hop. Frequency..Jazz. Frequency..K.pop.
## 1              Never           Sometimes            Never   Very frequently
## 2          Sometimes              Rarely  Very frequently            Rarely
## 3              Never              Rarely           Rarely   Very frequently
## 4          Sometimes               Never  Very frequently         Sometimes
## 5             Rarely     Very frequently            Never   Very frequently
```

```
## 6             Never           Sometimes  Very frequently   Very frequently
##   Frequency..Latin. Frequency..Lofi. Frequency..Metal. Frequency..Pop.
## 1   Very frequently           Rarely             Never Very frequently
## 2         Sometimes           Rarely             Never       Sometimes
## 3             Never        Sometimes         Sometimes          Rarely
## 4   Very frequently        Sometimes             Never       Sometimes
## 5         Sometimes        Sometimes             Never       Sometimes
## 6            Rarely  Very frequently            Rarely Very frequently
##   Frequency..R.B. Frequency..Rap. Frequency..Rock. Frequency..Video.game.music.
## 1       Sometimes Very frequently            Never                   Sometimes
## 2       Sometimes          Rarely  Very frequently                      Rarely
## 3           Never          Rarely           Rarely             Very frequently
## 4       Sometimes           Never            Never                       Never
## 5 Very frequently Very frequently            Never                      Rarely
## 6 Very frequently Very frequently  Very frequently                       Never
##   Anxiety Depression Insomnia OCD Music.effects   Permissions
## 1       3          0        1   0               I understand.
## 2       7          2        2   1               I understand.
## 3       7          7       10   2     No effect I understand.
## 4       9          7        3   3       Improve I understand.
## 5       7          2        5   9       Improve I understand.
## 6       8          8        7   7       Improve I understand.
```

```r
require('tidyr')
```

```
## Loading required package: tidyr
```

```r
require('dplyr')
```

```
## Loading required package: dplyr

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
require('ggplot2')
```

```
## Loading required package: ggplot2
```

Making categories for Anxious, Depressed, and Insomniac

```r
# mark people as anxious (1) if anxiety > 5, not anxious (0) otherwise
data$Anxious = ifelse(data$Anxiety > 5, 1, 0)
# mark people as anxious (1) if anxiety > 5, not anxious (0) otherwise
data$Depressed = ifelse(data$Depression > 5, 1, 0)
# mark people as anxious (1) if anxiety > 5, not anxious (0) otherwise
data$Insomniac = ifelse(data$Insomnia > 5, 1, 0)

head(data)
```

```
##           Timestamp Age Primary.streaming.service Hours.per.day While.working
## 1 8/27/2022 19:29:02  18                   Spotify           3.0           Yes
## 2 8/27/2022 19:57:31  63                   Pandora           1.5           Yes
## 3 8/27/2022 21:28:18  18                   Spotify           4.0            No
## 4 8/27/2022 21:40:40  61             YouTube Music           2.5           Yes
## 5 8/27/2022 21:54:47  18                   Spotify           4.0           Yes
## 6 8/27/2022 21:56:50  18                   Spotify           5.0           Yes
##   Instrumentalist Composer        Fav.genre Exploratory Foreign.languages BPM
## 1             Yes      Yes            Latin         Yes               Yes 156
## 2              No       No             Rock         Yes                No 119
## 3              No       No Video game music          No               Yes 132
## 4              No      Yes             Jazz         Yes               Yes  84
## 5              No       No              R&B         Yes                No 107
## 6             Yes      Yes             Jazz         Yes               Yes  86
##   Frequency..Classical. Frequency..Country. Frequency..EDM. Frequency..Folk.
## 1                Rarely               Never          Rarely            Never
## 2             Sometimes               Never           Never           Rarely
## 3                 Never               Never Very frequently            Never
## 4             Sometimes               Never           Never           Rarely
## 5                 Never               Never          Rarely            Never
## 6                Rarely           Sometimes           Never            Never
##   Frequency..Gospel. Frequency..Hip.hop. Frequency..Jazz. Frequency..K.pop.
## 1              Never           Sometimes            Never   Very frequently
## 2          Sometimes              Rarely  Very frequently            Rarely
## 3              Never              Rarely           Rarely   Very frequently
## 4          Sometimes               Never  Very frequently         Sometimes
## 5             Rarely     Very frequently            Never   Very frequently
## 6              Never           Sometimes  Very frequently   Very frequently
##   Frequency..Latin. Frequency..Lofi. Frequency..Metal. Frequency..Pop.
## 1   Very frequently           Rarely             Never Very frequently
## 2         Sometimes           Rarely             Never       Sometimes
## 3             Never        Sometimes         Sometimes          Rarely
## 4   Very frequently        Sometimes             Never       Sometimes
## 5         Sometimes        Sometimes             Never       Sometimes
## 6            Rarely  Very frequently            Rarely Very frequently
##   Frequency..R.B. Frequency..Rap. Frequency..Rock. Frequency..Video.game.music.
## 1       Sometimes Very frequently            Never                   Sometimes
## 2       Sometimes          Rarely  Very frequently                      Rarely
## 3           Never          Rarely           Rarely             Very frequently
## 4       Sometimes           Never            Never                       Never
## 5 Very frequently Very frequently            Never                      Rarely
## 6 Very frequently Very frequently  Very frequently                       Never
##   Anxiety Depression Insomnia OCD Music.effects   Permissions Anxious Depressed
## 1       3          0        1   0               I understand.       0         0
```

```
## 2        7           2          2    1             I understand.        1          0
## 3        7           7         10    2    No effect I understand.        1          1
## 4        9           7          3    3      Improve I understand.        1          1
## 5        7           2          5    9      Improve I understand.        1          0
## 6        8           8          7    7      Improve I understand.        1          1
##   Insomniac
## 1         0
## 2         0
## 3         1
## 4         0
## 5         0
## 6         1
```

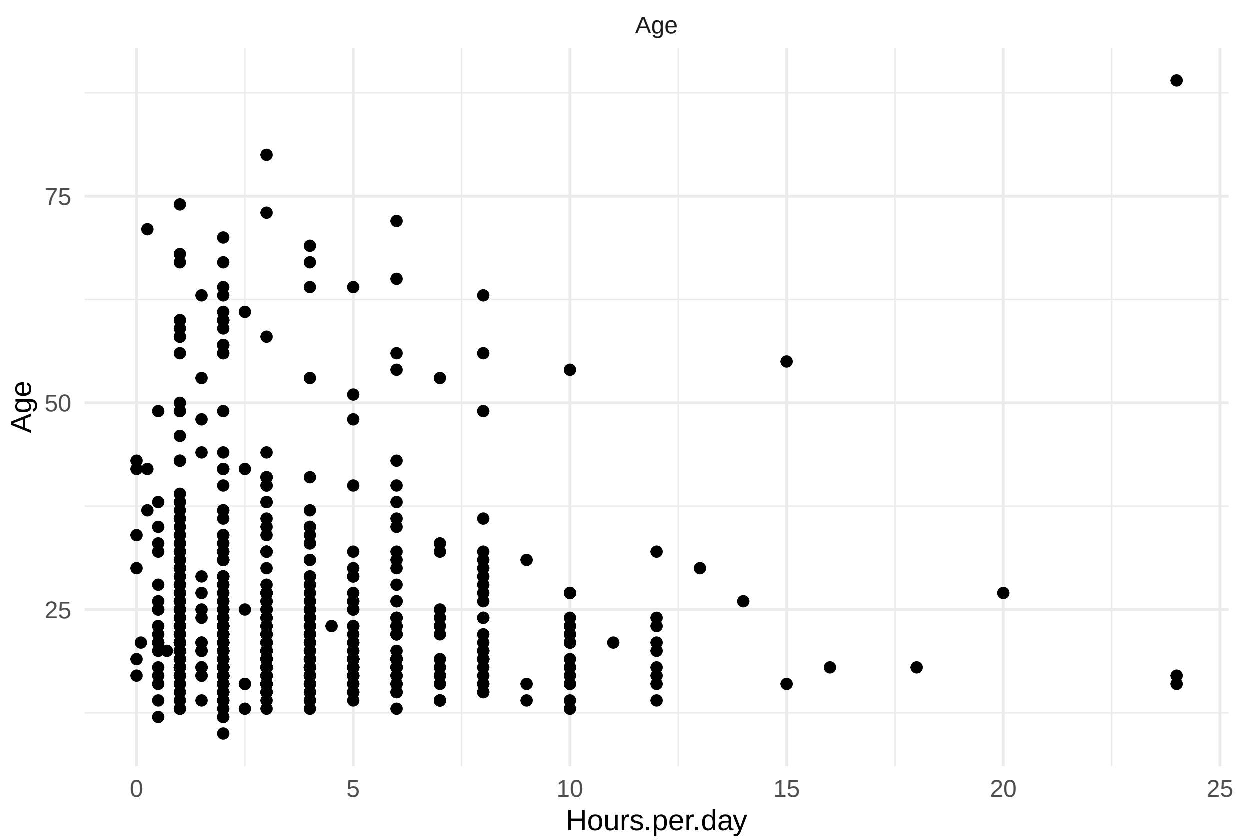
```r
attach(data)
```

Testing Non-linearity

We don't need to test for non-linearity for categorical variables so we will only test for non-linearity on the continuous variables, namely age.

```r
ggplot(pivot_longer(data = data, 2), aes(Hours.per.day, Age)) +
    theme_minimal() + geom_point() + facet_wrap('name', scales = 'free')
```

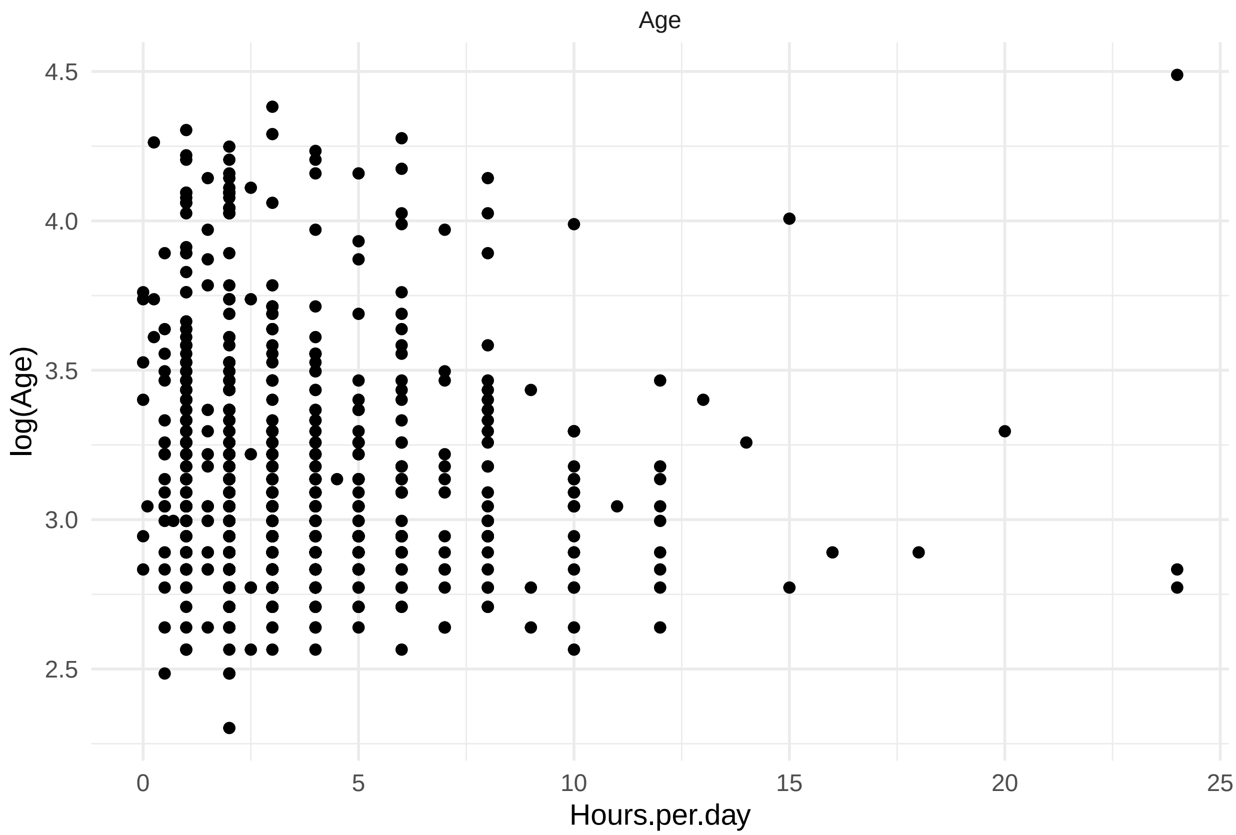
```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```



Plotting the covariate age against hours per day, it doesn't seem that there is a linear relationship between the 2. To try to fix this, let's take the log of age. With log age:

```
ggplot(pivot_longer(data = data, 2), aes(Hours.per.day, log(Age))) +
    theme_minimal() + geom_point() + facet_wrap('name', scales = 'free')
```

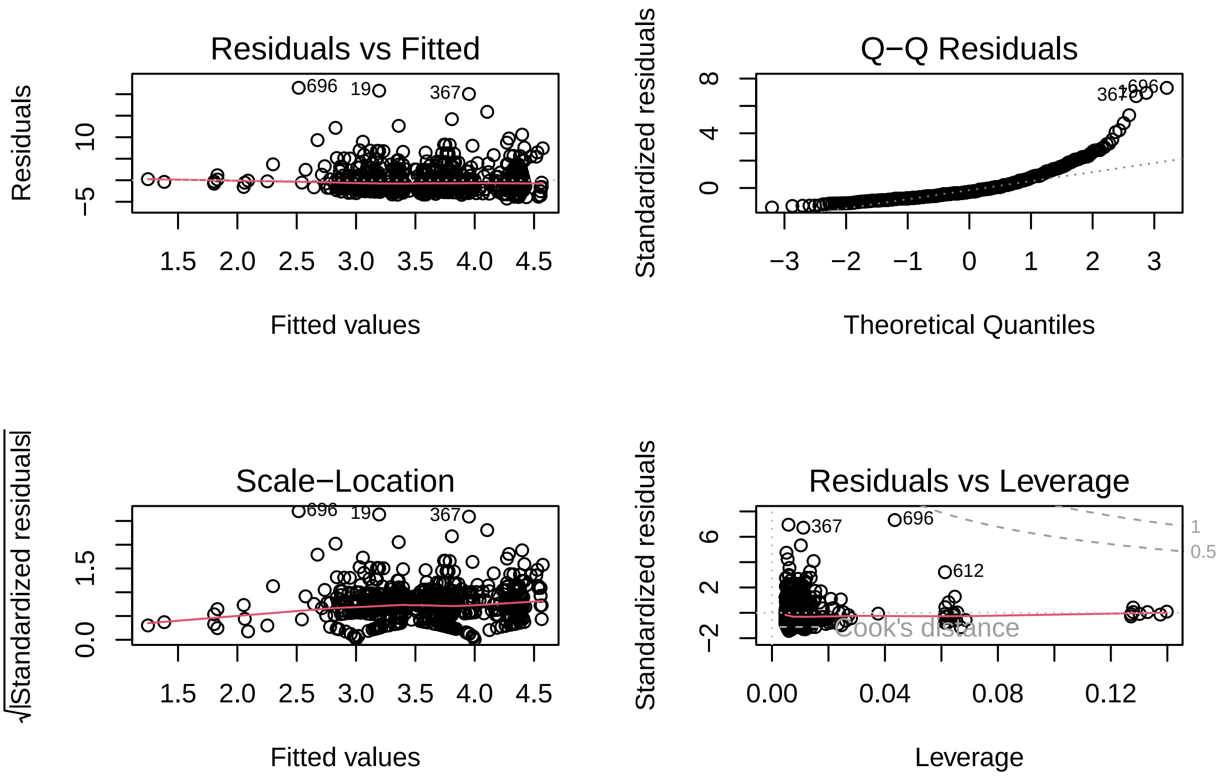
```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```



Fitted model

```
##
## Call:
## lm(formula = data$Hours.per.day ~ data$Age + data$Anxious + data$Depressed +
##     data$Insomniac + data$Music.effects)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2717 -1.8169 -0.8077  0.8563 21.4835
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    1.996420   1.107135   1.803   0.0718 .
## data$Age                      -0.009144   0.009348  -0.978   0.3283
## data$Anxious                  -0.173814   0.247245  -0.703   0.4823
## data$Depressed                 0.622019   0.244447   2.545   0.0111 *
## data$Insomniac                 0.573701   0.248678   2.307   0.0213 *
```

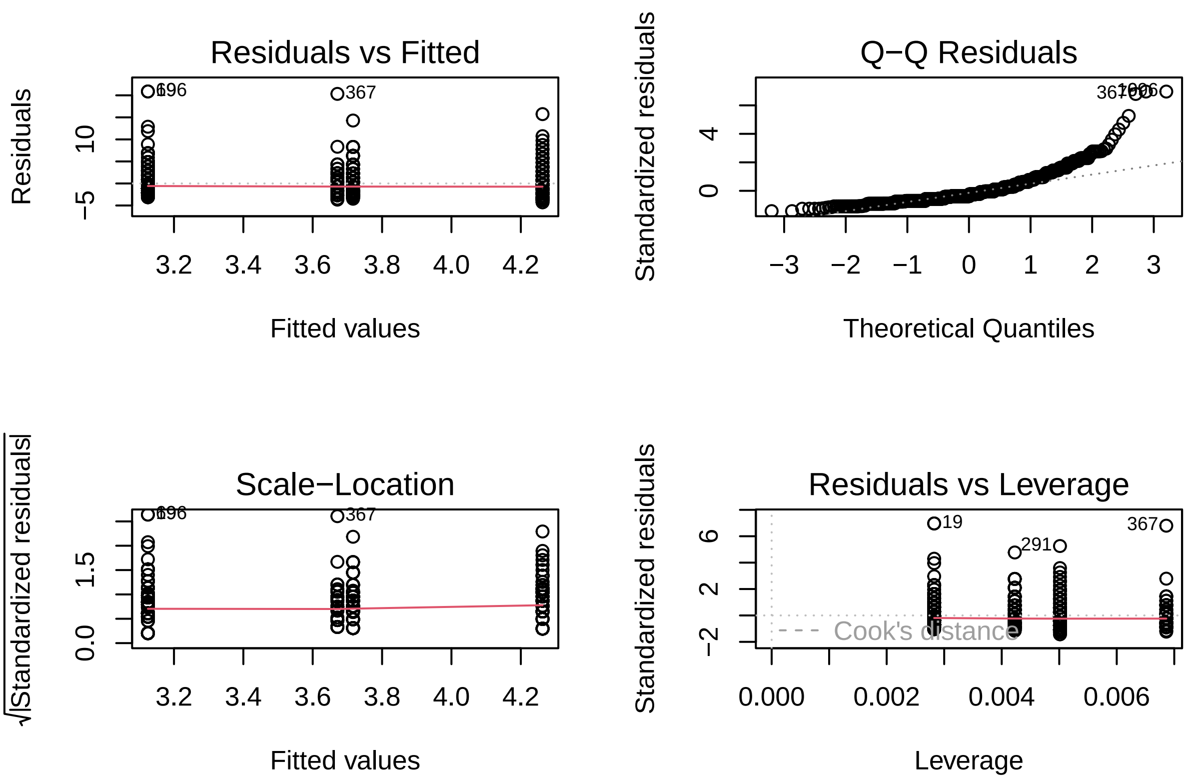
```
## data$Music.effectsImprove    1.527684   1.072628   1.424   0.1548
## data$Music.effectsNo effect  1.333873   1.087927   1.226   0.2206
## data$Music.effectsWorsen     0.422756   1.293733   0.327   0.7439
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.001 on 727 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.02839,    Adjusted R-squared:  0.01903
## F-statistic: 3.034 on 7 and 727 DF,  p-value: 0.003758
```



FitA is not normal as shown in the Q-Q Plot where the error points are rising off the normal line which show a right skew. Additionally, FitA does have a constant variance which was evident in Residual VS Fitted plot where there is no clear pattern however, there appear to be many outliers.

```
##
## Call:
## lm(formula = data$Hours.per.day ~ data$Depressed + data$Insomniac)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2627 -1.7163 -1.1248  0.8752 20.8752
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.1248     0.1595  19.591   <2e-16 ***
```

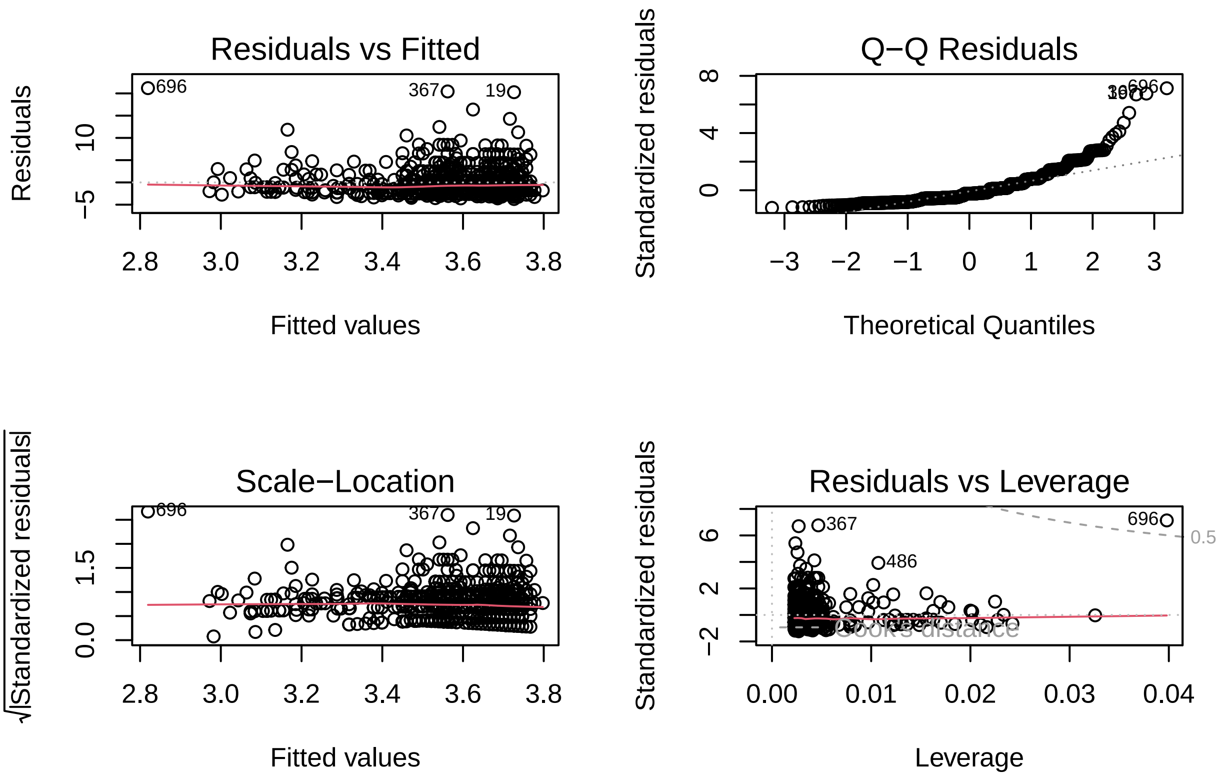
```
## data$Depressed   0.5914     0.2295   2.577   0.0102 *
## data$Insomniac   0.5464     0.2467   2.215   0.0271 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.001 on 733 degrees of freedom
## Multiple R-squared:  0.02075,    Adjusted R-squared:  0.01808
## F-statistic: 7.767 on 2 and 733 DF,  p-value: 0.0004594
```



FitB is not normal as shown in the Q-Q Plot where the error points are rising off the normal line which show a right skew. Additionally, FitB does have a constant variance which was evident in Residual VS Fitted plot where there is no clear pattern however, there appear to be many outliers.

```
##
## Call:
## lm(formula = data$Hours.per.day ~ data$Age + data$Anxious)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7264 -1.7264 -0.7162  1.2634 21.1805
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.724573   0.312308  11.926   <2e-16 ***
## data$Age     -0.010169   0.009381  -1.084    0.279
## data$Anxious  0.174711   0.230953   0.756    0.450
```

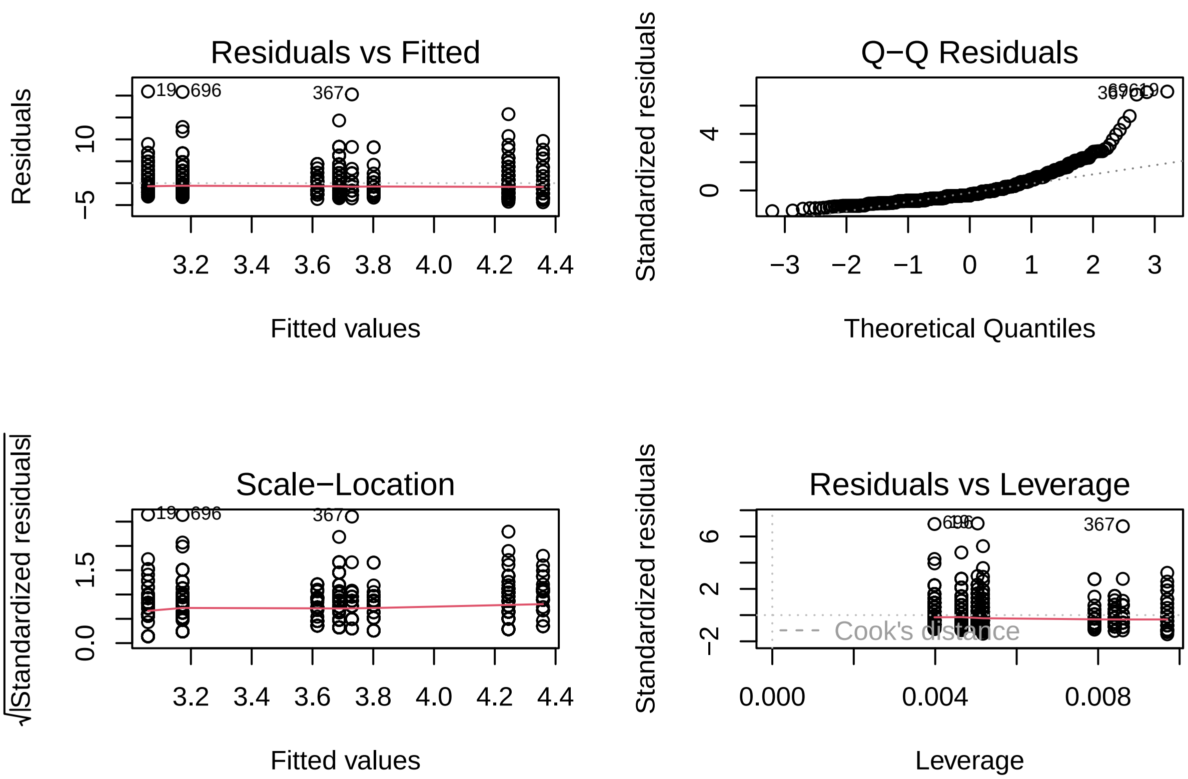
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.03 on 732 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.002772,   Adjusted R-squared:  4.744e-05
## F-statistic: 1.017 on 2 and 732 DF,  p-value: 0.362
```



FitC is not normal as shown in the Q-Q Plot where the error points are rising off the normal line which show a right skew. Additionally, FitC does have a constant variance which was evident in Residual VS Fitted plot where there is no clear pattern however, there appear to be many outliers.The outliers can be clearly seen in Residuals vs Leverage plot where point 696 is outside of the Cookline.

```
##
## Call:
## lm(formula = data$Hours.per.day ~ data$Anxious + data$Depressed +
##     data$Insomniac)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3586 -1.8013 -1.0589  0.8275 20.9411
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.1725     0.1895  16.744   <2e-16 ***
## data$Anxious    -0.1137     0.2434  -0.467   0.6407
```

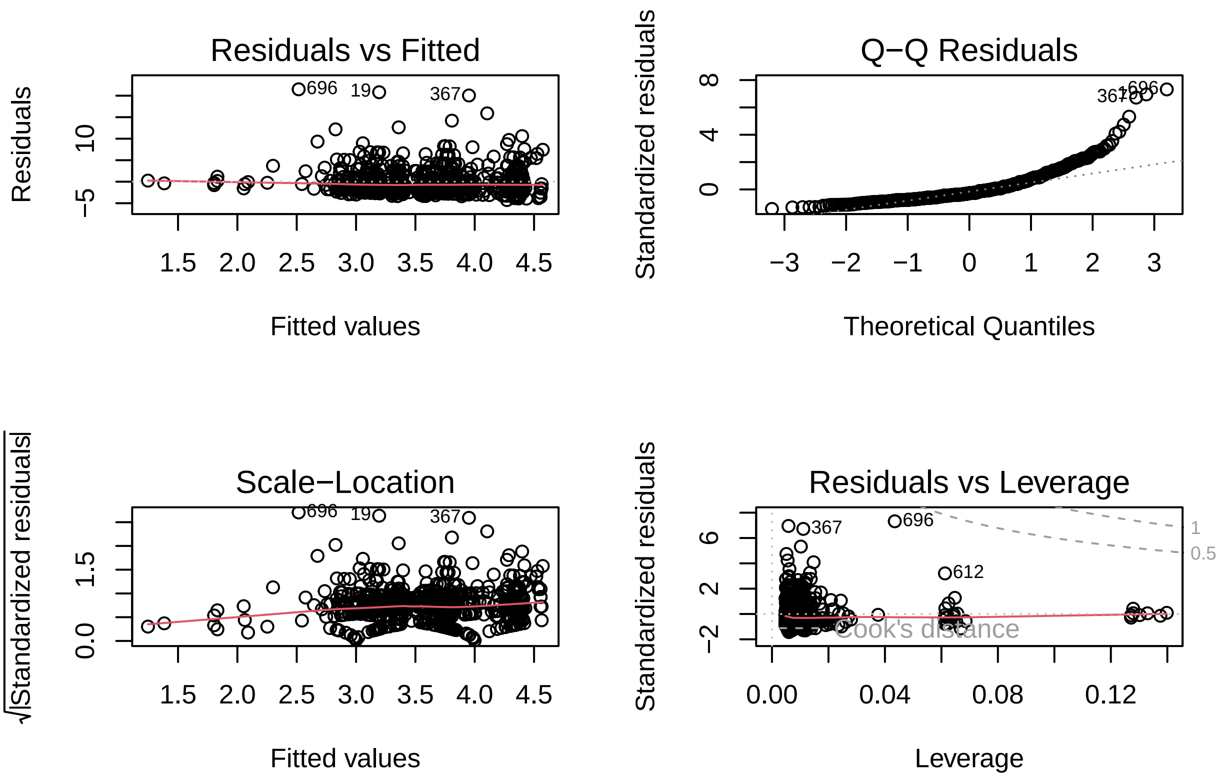
```
## data$Depressed   0.6288     0.2432   2.586   0.0099 **
## data$Insomniac   0.5573     0.2480   2.247   0.0249 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.002 on 732 degrees of freedom
## Multiple R-squared:  0.02104,    Adjusted R-squared:  0.01703
## F-statistic: 5.245 on 3 and 732 DF,  p-value: 0.001383
```



FitC is not normal as shown in the Q-Q Plot where the error points are rising off the normal line which show a right skew. Additionally, FitC does have a constant variance which was evident in Residual VS Fitted plot where there is no clear pattern however, there appear to be possible outliers.

```
##
## Call:
## lm(formula = data$Hours.per.day ~ -1 + data$Age + data$Anxious +
##     data$Depressed + data$Insomniac + data$Music.effects)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2717 -1.8169 -0.8077  0.8563 21.4835
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## data$Age                     -0.009144   0.009348  -0.978  0.32835
## data$Anxious                 -0.173814   0.247245  -0.703  0.48228
```

```
## data$Depressed              0.622019   0.244447   2.545  0.01115 *
## data$Insomniac              0.573701   0.248678   2.307  0.02133 *
## data$Music.effects          1.996420   1.107135   1.803  0.07177 .
## data$Music.effectsImprove   3.524104   0.326790  10.784  < 2e-16 ***
## data$Music.effectsNo effect 3.330293   0.376545   8.844  < 2e-16 ***
## data$Music.effectsWorsen    2.419176   0.791330   3.057  0.00232 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.001 on 727 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.5939, Adjusted R-squared:  0.5895
## F-statistic: 132.9 on 8 and 727 DF,  p-value: < 2.2e-16
```




```
anova_result <- anova(fitA, fitE, test = "LRT")
anova_result
```

```
## Analysis of Variance Table
##
## Model 1: data$Hours.per.day ~ data$Age + data$Anxious + data$Depressed +
##     data$Insomniac + data$Music.effects
## Model 2: data$Hours.per.day ~ -1 + data$Age + data$Anxious + data$Depressed +
##     data$Insomniac + data$Music.effects
##   Res.Df    RSS Df  Sum of Sq Pr(>Chi)
## 1    727 6548.3
## 2    727 6548.3  0 9.0949e-13
```

ANOVA test

```r
anova(fitA, fitE)
```

```
## Analysis of Variance Table
##
## Model 1: data$Hours.per.day ~ data$Age + data$Anxious + data$Depressed +
##     data$Insomniac + data$Music.effects
## Model 2: data$Hours.per.day ~ -1 + data$Age + data$Anxious + data$Depressed +
##     data$Insomniac + data$Music.effects
##   Res.Df    RSS Df  Sum of Sq F Pr(>F)
## 1    727 6548.3
## 2    727 6548.3  0 9.0949e-13
```

AIC

```r
AIC(fitA)
```

```
## [1] 3711.35
```

```r
AIC(fitB)
```

```
## [1] 3711.172
```

```r
AIC(fitC)
```

```
## [1] 3720.476
```

```r
AIC(fitD)
```

```
## [1] 3712.953
```

```r
AIC(fitE)
```

```
## [1] 3711.35
```

BIC

```r
BIC(fitA)
```

```
## [1] 3752.749
```

```r
BIC(fitB)
```

```
## [1] 3729.577
```

```r
BIC(fitC)
```

```
## [1] 3738.876
```

```r
BIC(fitD)
```

```
## [1] 3735.959
```

```r
BIC(fitE)
```

```
## [1] 3752.749
```

- fitB has the best fit according to AIC and BIC because it has the lowest AIC and BIC.
- For the AIC, fitA, fitB, and fitE have very similar AICs but, as noted, fitB has the lowest.

Analysis and Conclusion

Code Appendix:

```
knitr::opts_chunk$set(echo = TRUE)
data <- read.csv('mxmh_survey_results.csv')
head(data)
require('tidyr')
require('dplyr')
require('ggplot2')
# mark people as anxious (1) if anxiety > 5, not anxious (0) otherwise
data$Anxious = ifelse(data$Anxiety > 5, 1, 0)
# mark people as anxious (1) if anxiety > 5, not anxious (0) otherwise
data$Depressed = ifelse(data$Depression > 5, 1, 0)
# mark people as anxious (1) if anxiety > 5, not anxious (0) otherwise
data$Insomniac = ifelse(data$Insomnia > 5, 1, 0)

head(data)
attach(data)
ggplot(pivot_longer(data = data, 2), aes(Hours.per.day, Age)) +
    theme_minimal() + geom_point() + facet_wrap('name', scales = 'free')
ggplot(pivot_longer(data = data, 2), aes(Hours.per.day, log(Age))) +
    theme_minimal() + geom_point() + facet_wrap('name', scales = 'free')
fitA <- lm(data$Hours.per.day ~ data$Age + data$Anxious + data$Depressed + data$Insomniac + data$Music.
summary(fitA)


par(mfrow = c(2,2))
plot(fitA)


fitB <- lm(data$Hours.per.day ~ data$Depressed + data$Insomniac )
summary(fitB)

par(mfrow = c(2,2))
plot(fitB)


fitC <- lm(data$Hours.per.day ~ data$Age + data$Anxious )
summary(fitC)


par(mfrow = c(2,2))
plot(fitC)

fitD <- lm(data$Hours.per.day ~ data$Anxious + data$Depressed + data$Insomniac)
summary(fitD)

par(mfrow = c(2,2))
plot(fitD)
fitE <- lm(data$Hours.per.day ~ -1 + data$Age + data$Anxious + data$Depressed + data$Insomniac + data$Mu
summary(fitE)

par(mfrow = c(2,2))
plot(fitE)
```

```r
anova_result <- anova(fitA, fitE, test = "LRT")
anova_result
anova(fitA, fitE)
AIC(fitA)
AIC(fitB)
AIC(fitC)
AIC(fitD)
AIC(fitE)
BIC(fitA)
BIC(fitB)
BIC(fitC)
BIC(fitD)
BIC(fitE)
```