

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

##	Timestamp	Age	Primary.streaming.service	Hours.per.day
## 1	8/27/2022 19:29:02	18	Spotify	3.0
## 2	8/27/2022 19:57:31	63	Pandora	1.5
## 3	8/27/2022 21:28:18	18	Spotify	4.0
## 4	8/27/2022 21:40:40	61	YouTube Music	2.5
## 5	8/27/2022 21:54:47	18	Spotify	4.0
## 6	8/27/2022 21:56:50	18	Spotify	5.0
## 7	8/27/2022 22:00:29	18	YouTube Music	3.0
## 8	8/27/2022 22:18:59	21	Spotify	1.0
## 9	8/27/2022 22:33:05	19	Spotify	6.0
## 10	8/27/2022 22:44:03	18	I do not use a streaming service.	1.0

##	While.working	Instrumentalist	Composer	Fav.genre	Exploratory
## 1	Yes	Yes	Yes	Latin	Yes
## 2	Yes	No	No	Rock	Yes
## 3	No	No	No	Video game music	No
## 4	Yes	No	Yes	Jazz	Yes
## 5	Yes	No	No	R&B	Yes
## 6	Yes	Yes	Yes	Jazz	Yes
## 7	Yes	Yes	No	Video game music	Yes
## 8	Yes	No	No	K pop	Yes
## 9	Yes	No	No	Rock	No
## 10	Yes	No	No	R&B	Yes

##	Foreign.languages	BPM	Frequency..Classical.	Frequency..Country.
## 1	Yes	156	Rarely	Never
## 2	No	119	Sometimes	Never
## 3	Yes	132	Never	Never
## 4	Yes	84	Sometimes	Never
## 5	No	107	Never	Never
## 6	Yes	86	Rarely	Sometimes
## 7	Yes	66	Sometimes	Never
## 8	Yes	95	Never	Never

## 9	No	94	Never	Very frequently	
## 10	Yes	155	Rarely	Rarely	
##	Frequency..EDM.	Frequency..Folk.	Frequency..Gospel.	Frequency..Hip.hop.	
## 1	Rarely	Never	Never	Sometimes	
## 2	Never	Rarely	Sometimes	Rarely	
## 3	Very frequently	Never	Never	Rarely	
## 4	Never	Rarely	Sometimes	Never	
## 5	Rarely	Never	Rarely	Very frequently	
## 6	Never	Never	Never	Sometimes	
## 7	Rarely	Sometimes	Rarely	Rarely	
## 8	Rarely	Never	Never	Very frequently	
## 9	Never	Sometimes	Never	Never	
## 10	Rarely	Rarely	Sometimes	Rarely	
##	Frequency..Jazz.	Frequency..K.pop.	Frequency..Latin.	Frequency..Lofi.	
## 1	Never	Very frequently	Very frequently	Rarely	
## 2	Very frequently	Rarely	Sometimes	Rarely	
## 3	Rarely	Very frequently	Never	Sometimes	
## 4	Very frequently	Sometimes	Very frequently	Sometimes	
## 5	Never	Very frequently	Sometimes	Sometimes	
## 6	Very frequently	Very frequently	Rarely	Very frequently	
## 7	Sometimes	Never	Rarely	Rarely	
## 8	Rarely	Very frequently	Never	Sometimes	
## 9	Never	Never	Never	Never	
## 10	Rarely	Never	Rarely	Rarely	
##	Frequency..Metal.	Frequency..Pop.	Frequency..R.B.	Frequency..Rap.	
## 1	Never	Very frequently	Sometimes	Very frequently	
## 2	Never	Sometimes	Sometimes	Rarely	
## 3	Sometimes	Rarely	Never	Rarely	
## 4	Never	Sometimes	Sometimes	Never	
## 5	Never	Sometimes	Very frequently	Very frequently	
## 6	Rarely	Very frequently	Very frequently	Very frequently	
## 7	Rarely	Rarely	Rarely	Never	
## 8	Never	Sometimes	Sometimes	Rarely	
## 9	Very frequently	Never	Never	Never	
## 10	Never	Sometimes	Sometimes	Rarely	
##	Frequency..Rock.	Frequency..Video.game.music.	Anxiety	Depression	Insomnia
## 1	Never		Sometimes	3	0
## 2	Very frequently		Rarely	7	2
## 3	Rarely		Very frequently	7	7
## 4	Never		Never	9	7
## 5	Never		Rarely	7	2
## 6	Very frequently		Never	8	8
## 7	Never		Sometimes	4	8
## 8	Never		Rarely	5	3
## 9	Very frequently		Never	2	0
## 10	Sometimes		Sometimes	2	2
##	OCD Music.effects	Permissions			
## 1	0	I understand.			
## 2	1	I understand.			
## 3	2	No effect	I understand.		
## 4	3	Improve	I understand.		
## 5	9	Improve	I understand.		
## 6	7	Improve	I understand.		
## 7	0	Improve	I understand.		

```
## 8      3      Improve I understand.
## 9      0      Improve I understand.
## 10     1      Improve I understand.

## Loading required package: tidyr

## 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

## Loading required package: ggplot2
```

Making categories for Anxious, Depressed, and Insomniac

Since our data had people rate their anxiety, depression, insomnia, and OCD and a scale of 1 to 10 (only integers), we will make categories for **Anxiety**, **Depression**, **Insomnia**, and **OCD** so that the categories will be binary. For example, for **Anxiety**, we will have a category called **Anxious** that is 1 if **Anxiety** > 5 and 0 otherwise.

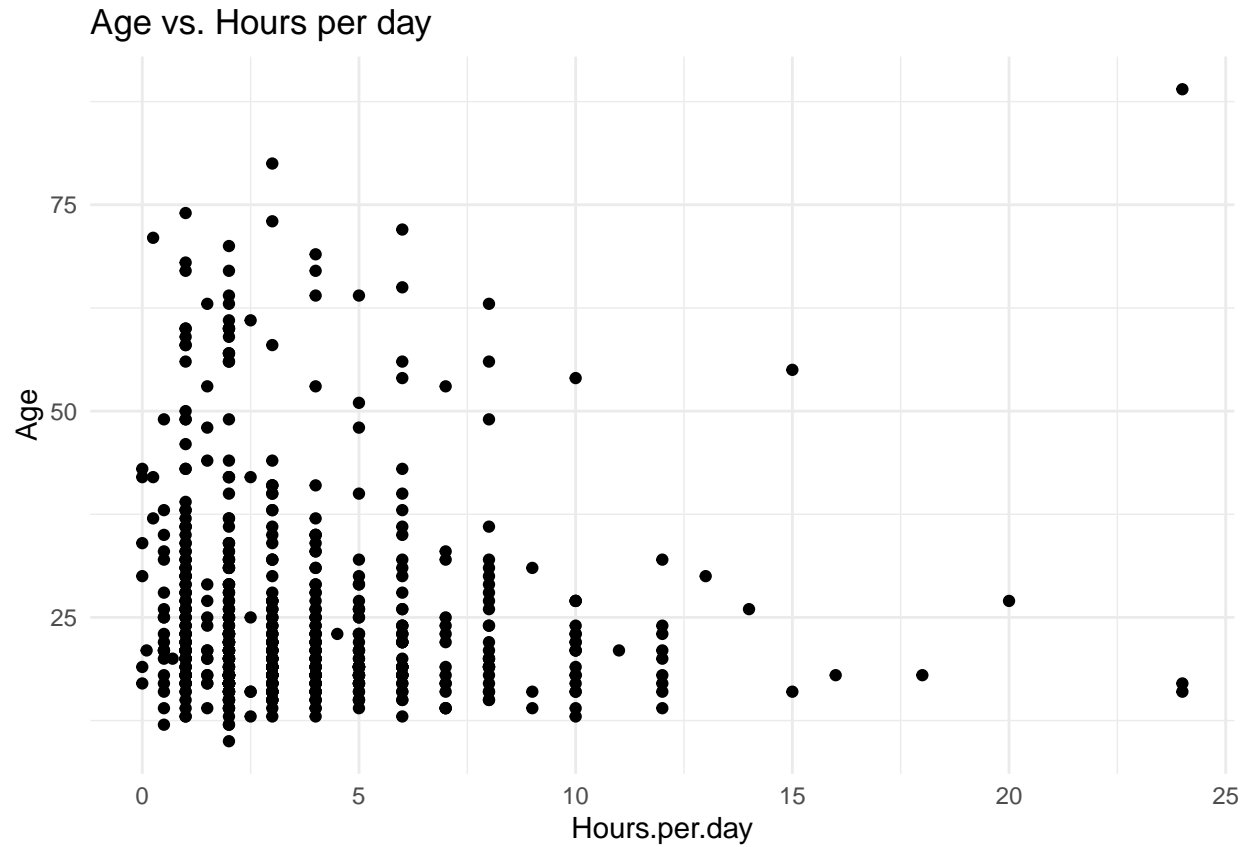
```
##      Timestamp Age      Primary.streaming.service Hours.per.day
## 1  8/27/2022 19:29:02 18      Spotify              3.0
## 2  8/27/2022 19:57:31 63      Pandora              1.5
## 3  8/27/2022 21:28:18 18      Spotify              4.0
## 4  8/27/2022 21:40:40 61      YouTube Music        2.5
## 5  8/27/2022 21:54:47 18      Spotify              4.0
## 6  8/27/2022 21:56:50 18      Spotify              5.0
## 7  8/27/2022 22:00:29 18      YouTube Music        3.0
## 8  8/27/2022 22:18:59 21      Spotify              1.0
## 9  8/27/2022 22:33:05 19      Spotify              6.0
## 10 8/27/2022 22:44:03 18 I do not use a streaming service. 1.0
##      While.working Instrumentalist Composer      Fav.genre Exploratory
## 1      Yes              Yes      Yes      Latin      Yes
## 2      Yes              No      No      Rock      Yes
## 3      No              No      No      Video game music  No
## 4      Yes              No      Yes      Jazz      Yes
## 5      Yes              No      No      R&B      Yes
## 6      Yes              Yes      Yes      Jazz      Yes
## 7      Yes              Yes      No      Video game music  Yes
## 8      Yes              No      No      K pop      Yes
## 9      Yes              No      No      Rock      No
## 10     Yes              No      No      R&B      Yes
##      Foreign.languages BPM Frequency..Classical. Frequency..Country.
```

## 1	Yes	156	Rarely	Never		
## 2	No	119	Sometimes	Never		
## 3	Yes	132	Never	Never		
## 4	Yes	84	Sometimes	Never		
## 5	No	107	Never	Never		
## 6	Yes	86	Rarely	Sometimes		
## 7	Yes	66	Sometimes	Never		
## 8	Yes	95	Never	Never		
## 9	No	94	Never	Very frequently		
## 10	Yes	155	Rarely	Rarely		
##	Frequency..EDM.	Frequency..Folk.	Frequency..Gospel.	Frequency..Hip.hop.		
## 1	Rarely	Never	Never	Sometimes		
## 2	Never	Rarely	Sometimes	Rarely		
## 3	Very frequently	Never	Never	Rarely		
## 4	Never	Rarely	Sometimes	Never		
## 5	Rarely	Never	Rarely	Very frequently		
## 6	Never	Never	Never	Sometimes		
## 7	Rarely	Sometimes	Rarely	Rarely		
## 8	Rarely	Never	Never	Very frequently		
## 9	Never	Sometimes	Never	Never		
## 10	Rarely	Rarely	Sometimes	Rarely		
##	Frequency..Jazz.	Frequency..K.pop.	Frequency..Latin.	Frequency..Lofi.		
## 1	Never	Very frequently	Very frequently	Rarely		
## 2	Very frequently	Rarely	Sometimes	Rarely		
## 3	Rarely	Very frequently	Never	Sometimes		
## 4	Very frequently	Sometimes	Very frequently	Sometimes		
## 5	Never	Very frequently	Sometimes	Sometimes		
## 6	Very frequently	Very frequently	Rarely	Very frequently		
## 7	Sometimes	Never	Rarely	Rarely		
## 8	Rarely	Very frequently	Never	Sometimes		
## 9	Never	Never	Never	Never		
## 10	Rarely	Never	Rarely	Rarely		
##	Frequency..Metal.	Frequency..Pop.	Frequency..R.B.	Frequency..Rap.		
## 1	Never	Very frequently	Sometimes	Very frequently		
## 2	Never	Sometimes	Sometimes	Rarely		
## 3	Sometimes	Rarely	Never	Rarely		
## 4	Never	Sometimes	Sometimes	Never		
## 5	Never	Sometimes	Very frequently	Very frequently		
## 6	Rarely	Very frequently	Very frequently	Very frequently		
## 7	Rarely	Rarely	Rarely	Never		
## 8	Never	Sometimes	Sometimes	Rarely		
## 9	Very frequently	Never	Never	Never		
## 10	Never	Sometimes	Sometimes	Rarely		
##	Frequency..Rock.	Frequency..Video.game.music.	Anxiety	Depression	Insomnia	
## 1	Never		Sometimes	3	0	1
## 2	Very frequently		Rarely	7	2	2
## 3	Rarely	Very frequently		7	7	10
## 4	Never		Never	9	7	3
## 5	Never		Rarely	7	2	5
## 6	Very frequently		Never	8	8	7
## 7	Never		Sometimes	4	8	6
## 8	Never		Rarely	5	3	5
## 9	Very frequently		Never	2	0	0
## 10	Sometimes		Sometimes	2	2	5

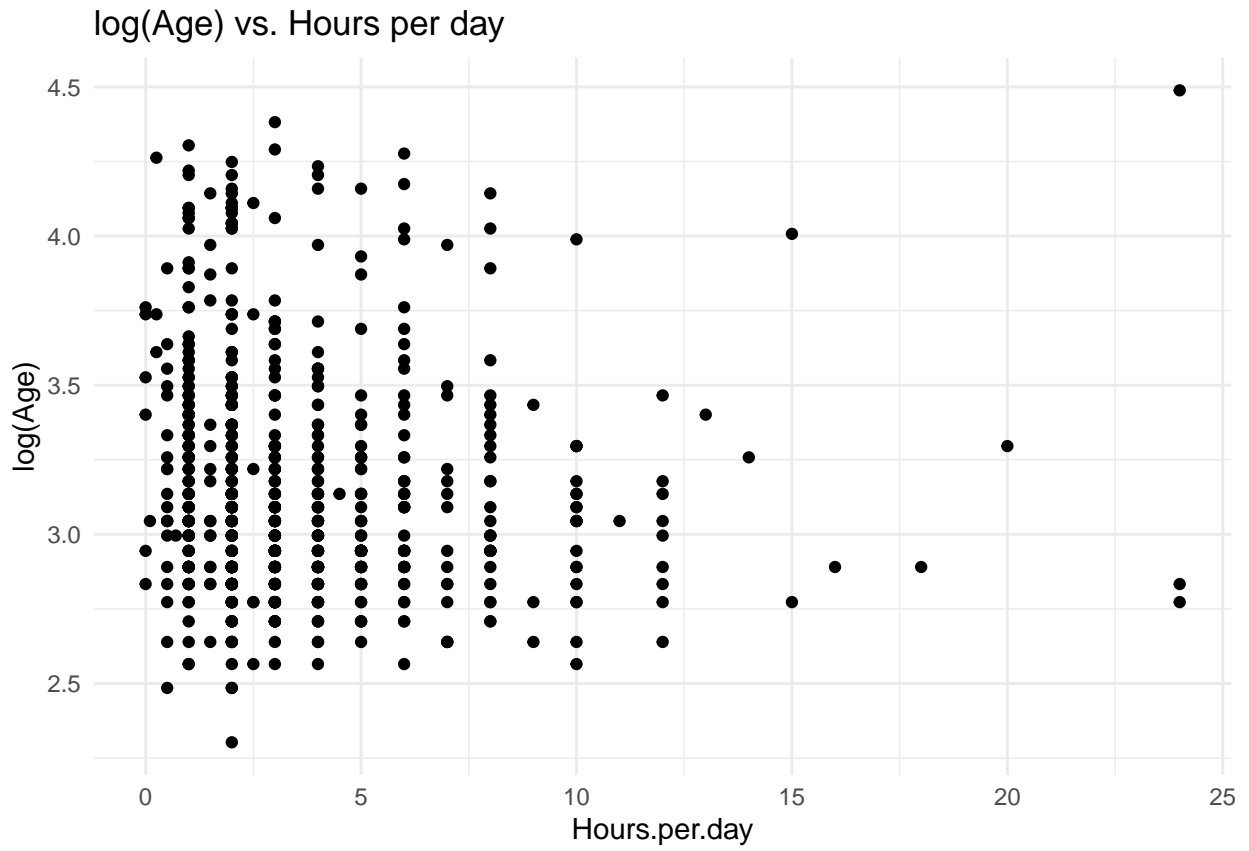
##	OCD	Music.effects	Permissions	Anxious	Depressed	Insomniac
## 1	0		I understand.	0	0	0
## 2	1		I understand.	1	0	0
## 3	2	No effect	I understand.	1	1	1
## 4	3	Improve	I understand.	1	1	0
## 5	9	Improve	I understand.	1	0	0
## 6	7	Improve	I understand.	1	1	1
## 7	0	Improve	I understand.	0	1	1
## 8	3	Improve	I understand.	0	0	0
## 9	0	Improve	I understand.	0	0	0
## 10	1	Improve	I understand.	0	0	0

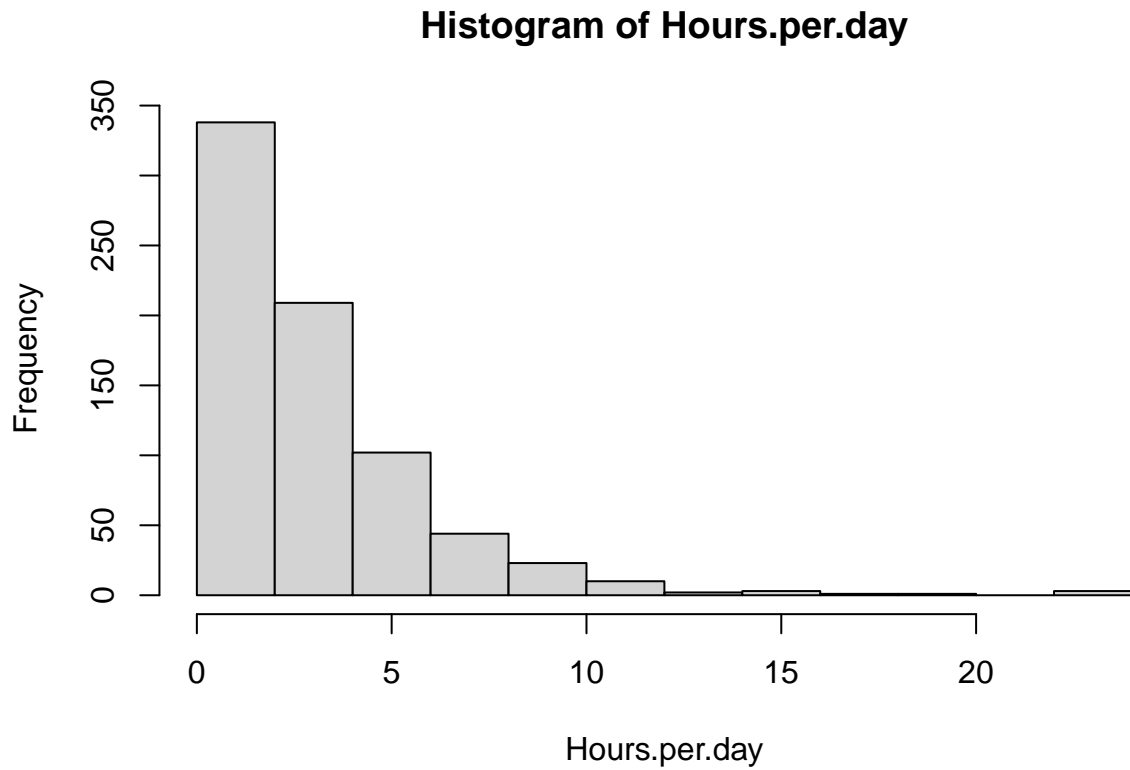
Testing Non-linearity

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



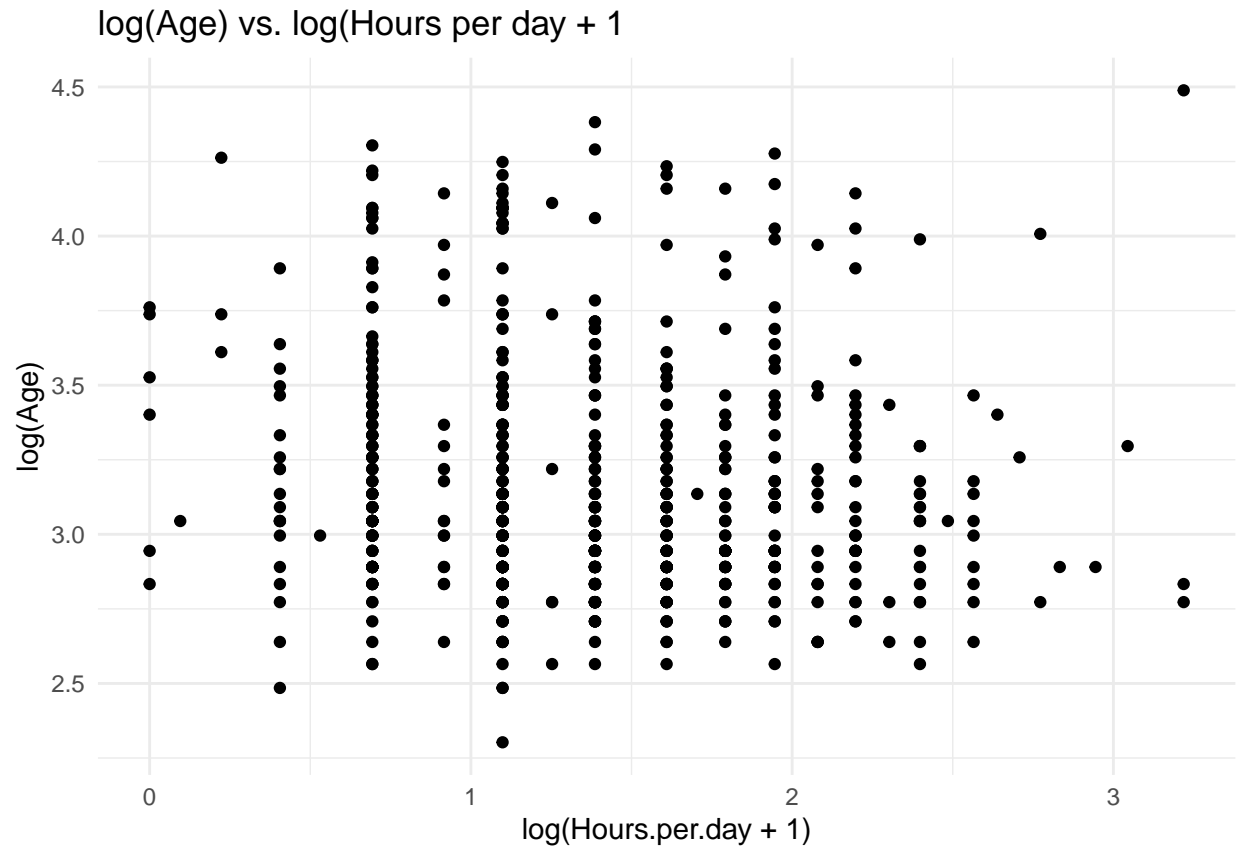
There doesn't seem to be a linear relationship between age and hours per day. We can transform age by logging it so that the relationship looks less non-linear. With log age:





Since the hours per day that people listen to music is not normally distributed, for each of the models, we will have a model where we log Hours.per.day and a model where we fit it to a log-link Gamma distribution.

Also, since half our models will have log Hours.per.day, we should also look to see if log(Age) and log(Hours.per.day + 1).

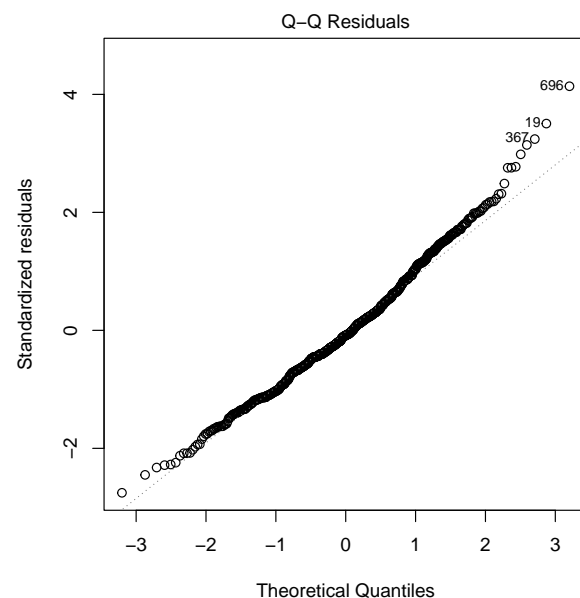
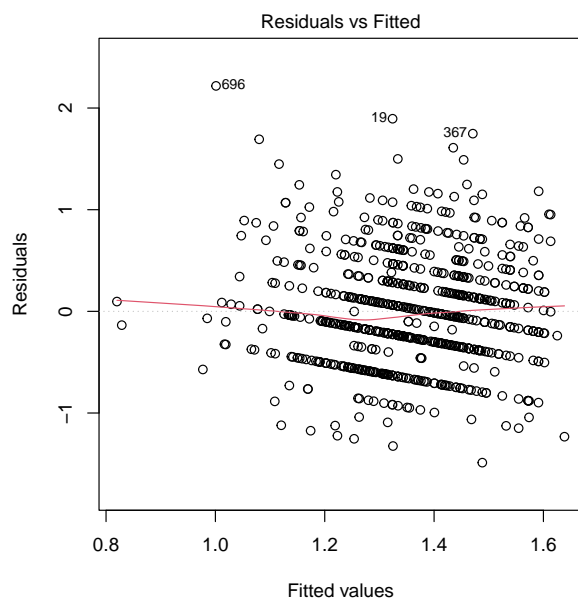


The relationship between $\log(\text{Age})$ and $\log(\text{Hours.per.day} + 1)$ is not obviously non-linear.

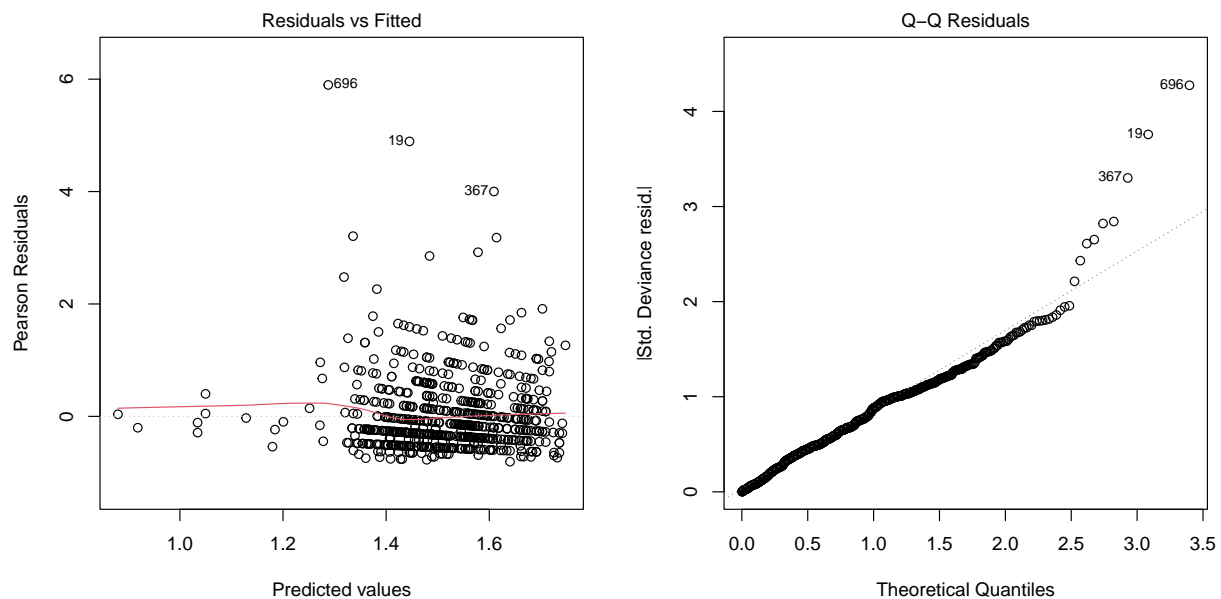
Fitted model

Since the hours per day that people listen to music is not normally distributed, we created two fits, one where we log Hours.per.day and the other where we fit it to a log-link Gamma distribution.

```
##
## Call:
## lm(formula = log(Hours.per.day + 1) ~ log(Age) + Anxious + Depressed +
##      Insomniac + Music.effects)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.48822 -0.35553 -0.04605  0.33077  2.21770
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.51836    0.26176   5.801 9.85e-09 ***
## log(Age)         -0.16400    0.05365  -3.057  0.00232 **
## Anxious          -0.01895    0.04460  -0.425  0.67113
## Depressed         0.13938    0.04408   3.162  0.00163 **
## Insomniac        0.11785    0.04492   2.624  0.00888 **
## Music.effectsImprove 0.28938    0.19371   1.494  0.13565
## Music.effectsNo effect 0.21897    0.19647   1.115  0.26543
## Music.effectsWorsen -0.02302    0.23368  -0.099  0.92156
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5422 on 727 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.05533,    Adjusted R-squared:  0.04623
## F-statistic: 6.083 on 7 and 727 DF,  p-value: 6.435e-07
```



```
##
## Call:
## glm(formula = Hours.per.day + 1 ~ log(Age) + Anxious + Depressed +
##       Insomniac + Music.effects, family = Gamma(link = "log"))
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.33875    0.32175   4.161 3.55e-05 ***
## log(Age)         -0.09998    0.06595  -1.516   0.1300
## Anxious          -0.04465    0.05483  -0.814   0.4157
## Depressed         0.13867    0.05419   2.559   0.0107 *
## Insomniac         0.11312    0.05522   2.049   0.0409 *
## Music.effectsImprove 0.43442    0.23811   1.824   0.0685 .
## Music.effectsNo effect 0.39773    0.24150   1.647   0.1000
## Music.effectsWorsen  0.18991    0.28723   0.661   0.5087
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 0.444237)
##
## Null deviance: 241.13  on 734  degrees of freedom
## Residual deviance: 231.11  on 727  degrees of freedom
## (1 observation deleted due to missingness)
## AIC: 3285.5
##
## Number of Fisher Scoring iterations: 6
```



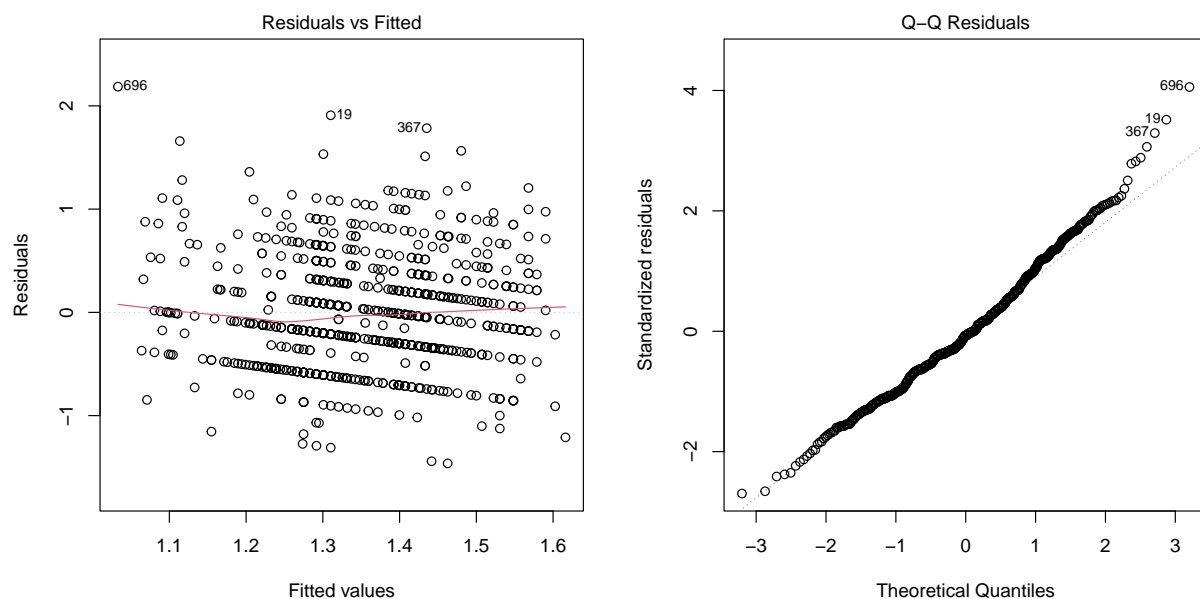
Analysis:

- For Log: From looking at the Residual Vs Fitted plot for fitA_log, it seems that the normality assumption is not violated because the shape of the fit is cloud-like without any noticeable pattern.

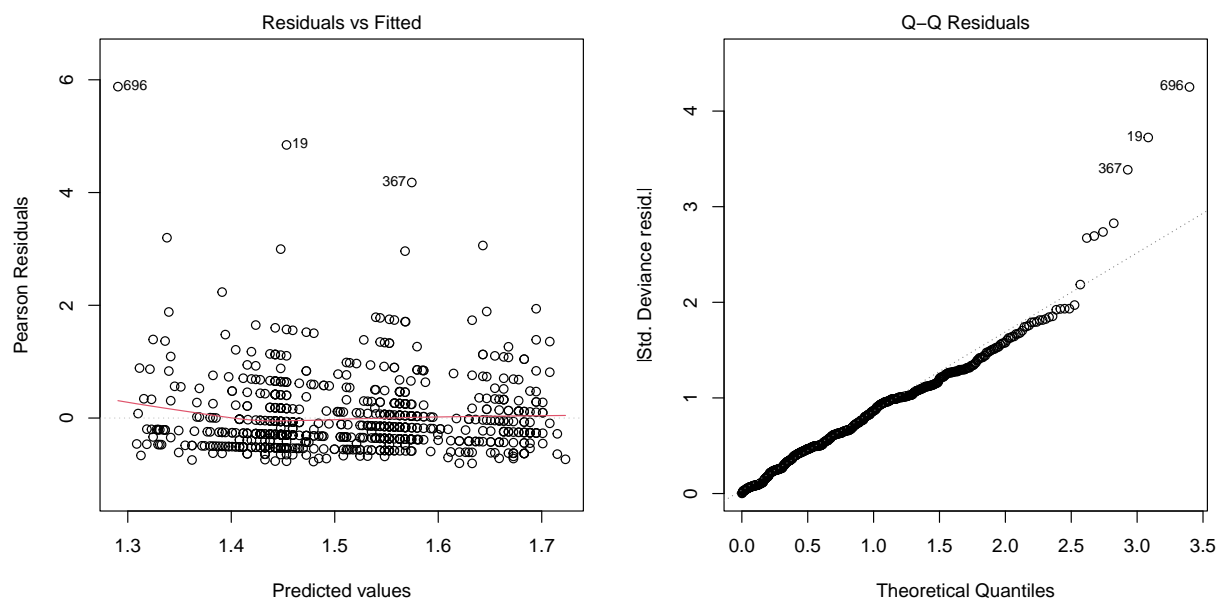
This means that `fitA_log` does have constant variance. However, there are possible outliers such as point 696 or 19. For the Normal Q-Q plot, normality doesn't seem to be violated because most error points remain on the normality line. Nevertheless, there are still evidences of outliers.

- For Gamma: Since the distribution is Gamma, it is possible to observe clustering of negatively valued residuals in Residuals and Fitted. This means that `fitA_gamma`, doesn't seem to violate normality assumption. In another word, `fitA_gamma` have a constant Variance. However, there are possible outliers such as points 696 or 19. For the Normal Q-Q plot, Gamma violated normality assumption, as errors points are going off the normal line.

```
##
## Call:
## lm(formula = log(Hours.per.day + 1) ~ log(Age) + Depressed +
##      Insomniac)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.46264 -0.34592 -0.03795  0.32323  2.18588
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.78490    0.17085  10.447 < 2e-16 ***
## log(Age)     -0.16751    0.05327  -3.145  0.00173 **
## Depressed     0.13257    0.04175   3.175  0.00156 **
## Insomniac     0.11492    0.04479   2.566  0.01050 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5441 on 731 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.04363,    Adjusted R-squared:  0.0397
## F-statistic: 11.12 on 3 and 731 DF,  p-value: 3.851e-07
```



```
##
## Call:
## glm(formula = Hours.per.day + 1 ~ log(Age) + Depressed + Insomniac,
##      family = Gamma(link = "log"))
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.73224    0.20966   8.262 6.71e-16 ***
## log(Age)     -0.09842    0.06537  -1.506  0.1326
## Depressed     0.12017    0.05124   2.345  0.0193 *
## Insomniac     0.11508    0.05497   2.094  0.0366 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 0.4458175)
##
## Null deviance: 241.13  on 734  degrees of freedom
## Residual deviance: 233.57  on 731  degrees of freedom
## (1 observation deleted due to missingness)
## AIC: 3285.7
##
## Number of Fisher Scoring iterations: 6
```



Analysis:

- For Log: From looking at the Residual Vs Fitted plot for fitB_log, it seems that the normality assumption is not violated because the shape of the fit is cloud-like without any noticeable pattern. This means that fitB_log does have constant variance. However, there are possible outliers such as point 696 or 19. For the Normal Q-Q plot, normality doesn't seem to be violated because most error points remains on the normality line. Nevertheless, there are still evidences of outliers.
- For Gamma: Since the distribution is Gamma, it is possible to observe clustering of negatively valued residuals in Residuals and Fitted. This means that fitB_gamma, doesn't seem to violate normality assumption. In another word, fitA_gamma have a constant Variance. However, there are possible outliers such as points 696 or 19. For the Normal Q-Q plot, Gamma violated normality assumption, as errors points are going off the normal line.

AIC

```
## [1] 1196.068
```

```
## [1] 1197.113
```

```
## [1] 3285.542
```

```
## [1] 3285.735
```

BIC

```
## [1] 1237.467
```

```
## [1] 1220.112
```

```
## [1] 3326.941
```

```
## [1] 3308.735
```

- fitAlog is a better fit than fitBlog according to AIC because it has a lower AIC.
- fitBlog is a better fit than fitAlog according to BIC because it has a lower BIC.
- fitAgamma is a better fit than fitBgamma according to AIC because it has a lower AIC.
- fitBgamma is a better fit than fitAgamma according to BIC because it has a lower BIC.

Analysis and Conclusion

Code Appendix:

```
knitr::opts_chunk$set(echo = TRUE, warning = FALSE, fig.align='center')
#knitr::opts_chunk$set(fig.width = 12, fig.height = 6)
data <- read.csv('mxmh_survey_results.csv')
head(data, 10)
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, 10)
attach(data)
ggplot(pivot_longer(data = data, 2), aes(Hours.per.day, Age)) +
  labs(title = 'Age vs. Hours per day') +
  theme_minimal() + geom_point()
# log transform because there are a few high values and many low values
ggplot(pivot_longer(data = data, 2), aes(Hours.per.day, log(Age))) + labs(title = 'log(Age) vs. Hours p
  theme_minimal() + geom_point()
hist(Hours.per.day)
# Hours per day is not normal so we can log transform it or use gamma glm
# log transform because there are a few high values and many low values
ggplot(pivot_longer(data = data, 2), aes(log(Hours.per.day + 1), log(Age))) + labs(title = 'log(Age) vs
  theme_minimal() + geom_point()
# fitA where we log Hours.per.day + 1 since Hours.per.day is not normally distributed.
fitAlog <- lm(log(Hours.per.day + 1) ~ log(Age) + Anxious + Depressed + Insomniac + Music.effects)
summary(fitAlog)

# Residuals vs. Fitted Values Plot and QQ Plot
par(mfrow = c(1,2))
plot(fitAlog, which = c(1,2))
# fitA where it's fitted to a gamma distribution
# For both fitAlog and fitAgamma, we add 1 to Hours.per.day because we can't log zero.
fitAgamma <- glm(Hours.per.day + 1 ~ log(Age) + Anxious + Depressed + Insomniac + Music.effects,
  family = Gamma (link = 'log'))
summary(fitAgamma)

# Residuals vs. Fitted Values Plot and QQ Plot
par(mfrow = c(1,2))
plot(fitAgamma, which = c(1,2))
# log Hours.per.day +1
fitBlog <- lm(log(Hours.per.day + 1) ~ log(Age) + Depressed + Insomniac)
summary(fitBlog)

# Residuals vs. Fitted Values Plot and QQ Plot
par(mfrow = c(1,2))
plot(fitBlog, which = c(1,2))
# gamma glm
```



```

fitBgamma <- glm(Hours.per.day + 1 ~ log(Age) + Depressed + Insomniac,
                 family = Gamma (link = 'log'))
summary(fitBgamma)

# Residuals vs. Fitted Values Plot and QQ Plot
par(mfrow = c(1,2))
plot(fitBgamma, which = c(1,2))
AIC(fitAlog)
AIC(fitBlog)

AIC(fitAgamma)
AIC(fitBgamma)
BIC(fitAlog)
BIC(fitBlog)

BIC(fitAgamma)
BIC(fitBgamma)

```