# R Notebook

```
library(readr)

sleep <- read_csv("https://raw.githubusercontent.com/JA-
McLean/STOR455/master/data/SleepStudy.csv")

## Rows: 253 Columns: 27

## -- Column specification ------------------------------------------------
------
## Delimiter: ","
## chr  (5): LarkOwl, DepressionStatus, AnxietyStatus, Stress, AlcoholUse
## dbl (22): Gender, ClassYear, NumEarlyClass, EarlyClass, GPA,
ClassesMissed, ...

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this
message.
```
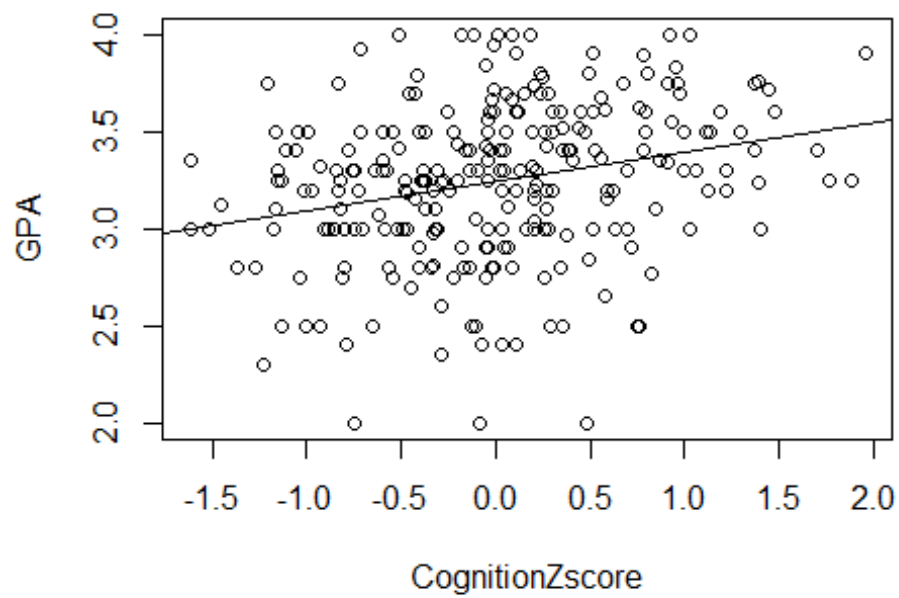
```
#Transformations
#Building a Model with a single predictors
```

```
head(sleep)

## # A tibble: 6 x 27
##    Gender ClassYear LarkOwl NumEarlyClass EarlyClass   GPA ClassesMissed
##     <dbl>     <dbl> <chr>           <dbl>      <dbl> <dbl>         <dbl>
## 1      0         4 Neither             0          0  3.6             0
## 2      0         4 Neither             2          1  3.24            0
## 3      0         4 Owl                 0          0  2.97           12
## 4      0         1 Lark                5          1  3.76            0
## 5      0         4 Owl                 0          0  3.2             4
## 6      1         4 Neither             0          0  3.5             0
## # ... with 20 more variables: CognitionZscore <dbl>, PoorSleepQuality
<dbl>,
## #   DepressionScore <dbl>, AnxietyScore <dbl>, StressScore <dbl>,
## #   DepressionStatus <chr>, AnxietyStatus <chr>, Stress <chr>, DASScore
<dbl>,
## #   Happiness <dbl>, AlcoholUse <chr>, Drinks <dbl>, WeekdayBed <dbl>,
## #   WeekdayRise <dbl>, WeekdaySleep <dbl>, WeekendBed <dbl>, WeekendRise
<dbl>,
## #   WeekendSleep <dbl>, AverageSleep <dbl>, AllNighter <dbl>
```
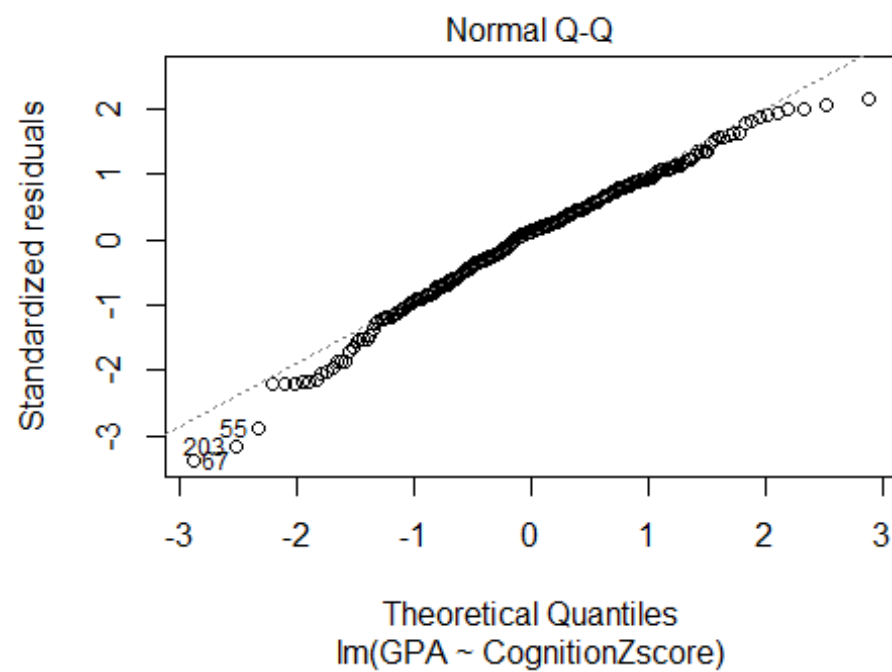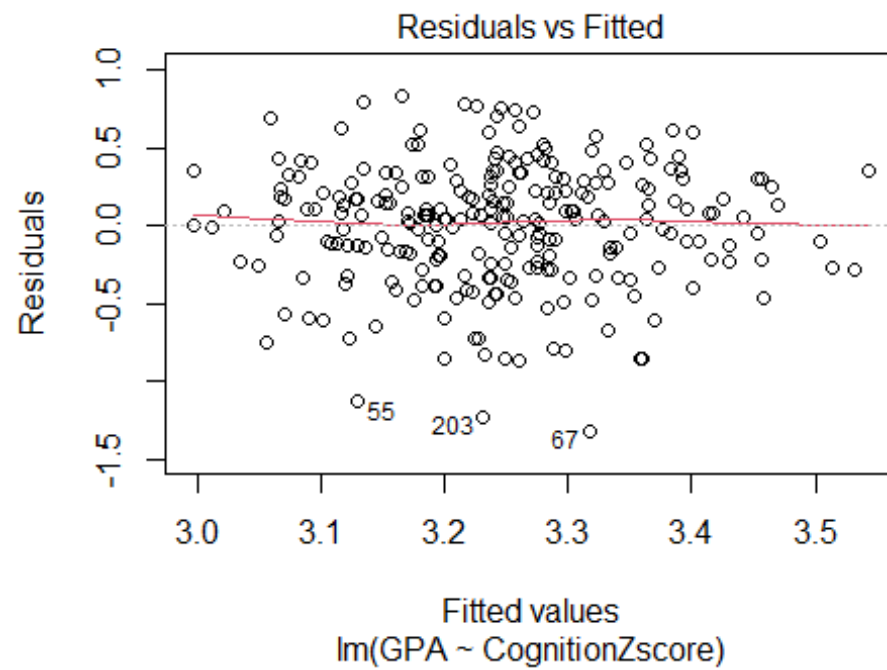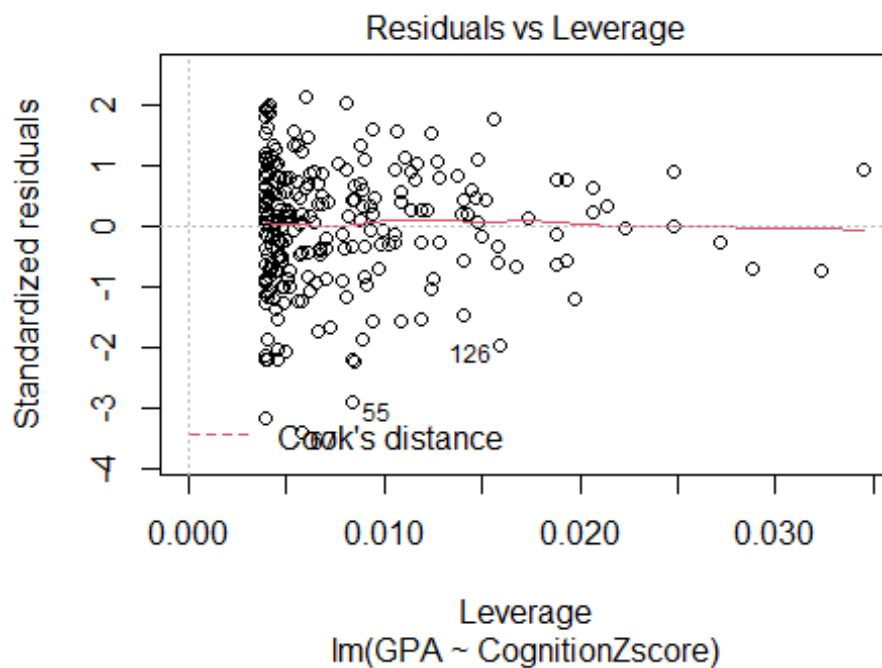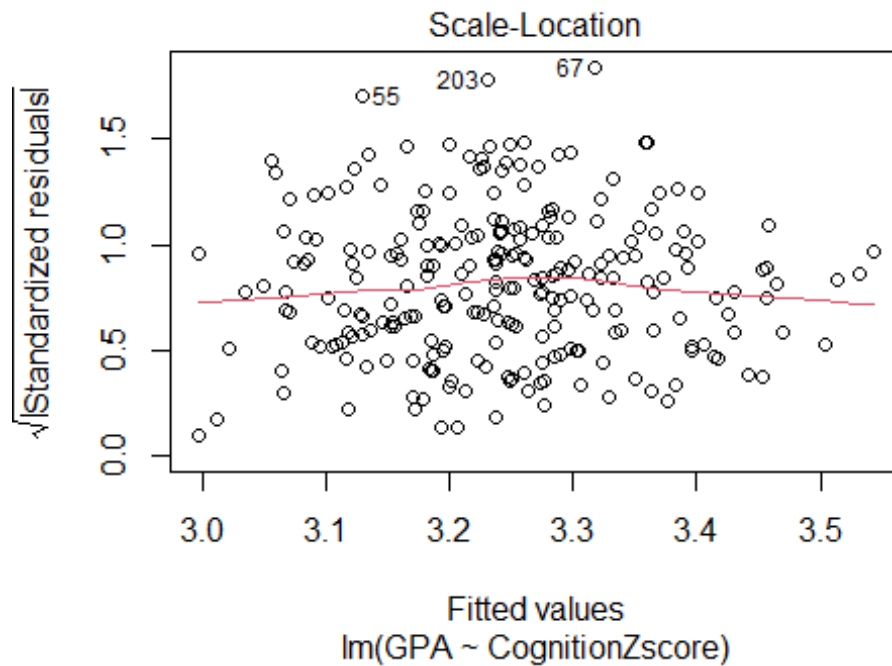
1. Using CognitionZscore as a predictor for GPA, calculate the least squares regression line that best fits your data and plot this.

```
mod1 = lm(GPA ~ CognitionZscore, data = sleep)
plot(GPA ~ CognitionZscore, data = sleep)
abline(mod1)
```



```
plot(mod1)
```

# Residuals vs Fitted



Fitted values
lm(GPA ~ CognitionZscore)

# Normal Q-Q



Theoretical Quantiles
lm(GPA ~ CognitionZscore)

Scale-Location
√|Standardized residuals|
Fitted values
lm(GPA ~ CognitionZscore)



Residuals vs Leverage
Standardized residuals
Leverage
lm(GPA ~ CognitionZscore)

2. Experiment with some transformation(s) to attempt to find one that seems to do a better job of satisfying the linear model conditions. Plot this as well.

```
mod2 = lm(log(GPA) ~ log(CognitionZscore), data = sleep)
```

```
## Warning in log(CognitionZscore): NaNs produced

plot(GPA ~ CognitionZscore, data = sleep)

b0 = summary(mod2)$coefficients[1,1]
b1 = summary(mod2)$coefficients[2,1]

curve(exp(b0+b1*(x)), add = TRUE)
```



```
plot(mod2)
```

Residuals vs Fitted

lm(log(GPA) ~ log(CognitionZscore))



Normal Q-Q

lm(log(GPA) ~ log(CognitionZscore))

Scale-Location
lm(log(GPA) ~ log(CognitionZscore))



Residuals vs Leverage
lm(log(GPA) ~ log(CognitionZscore))
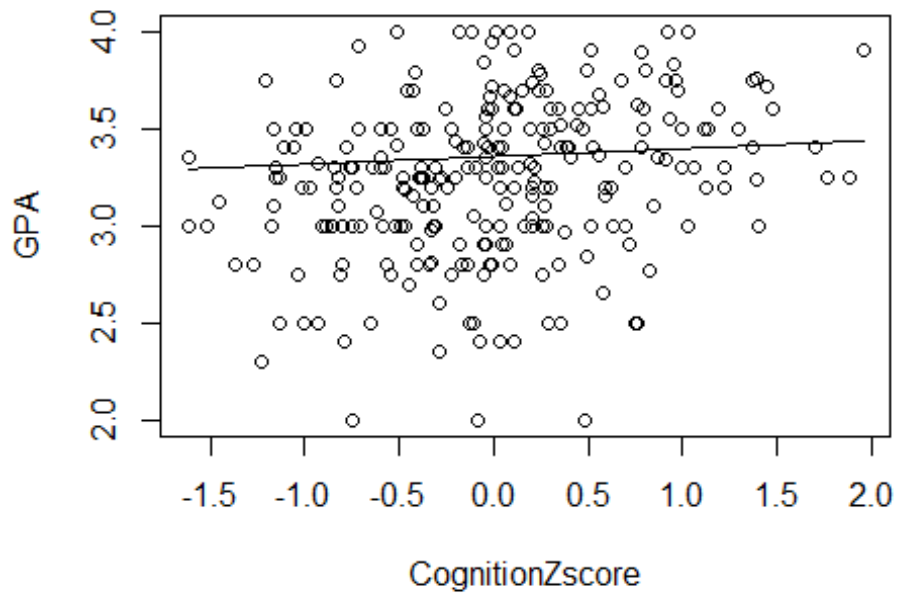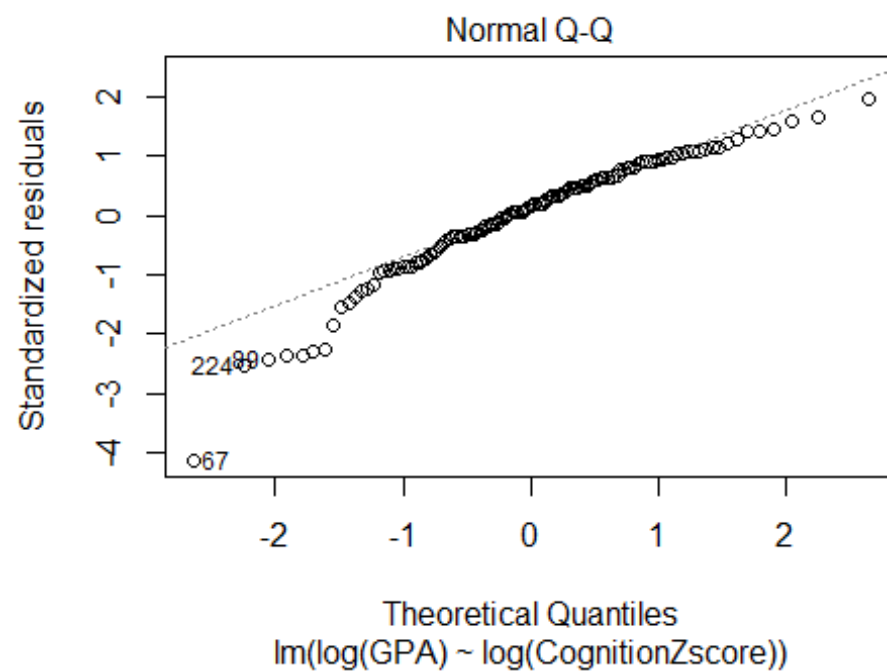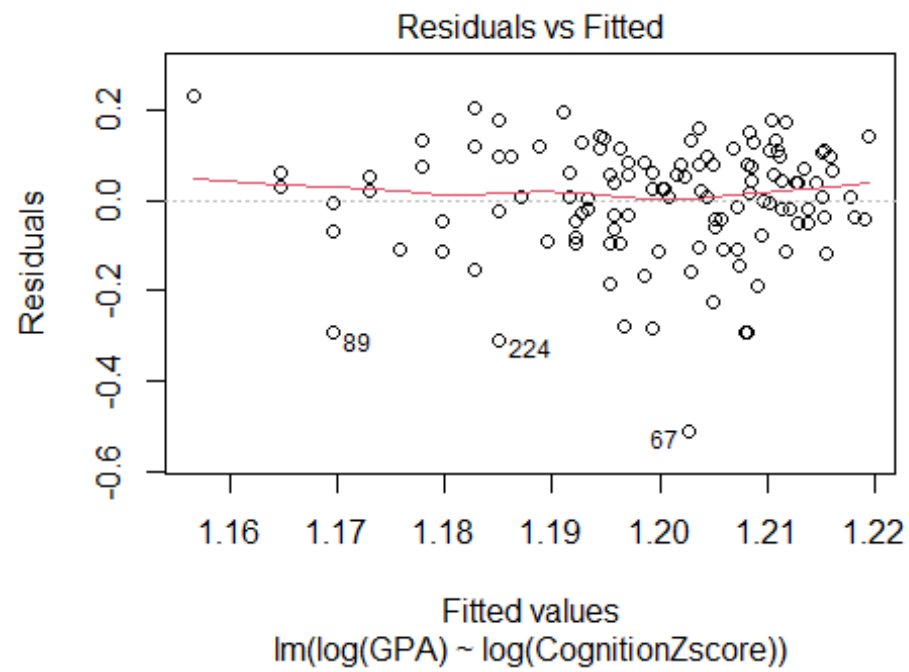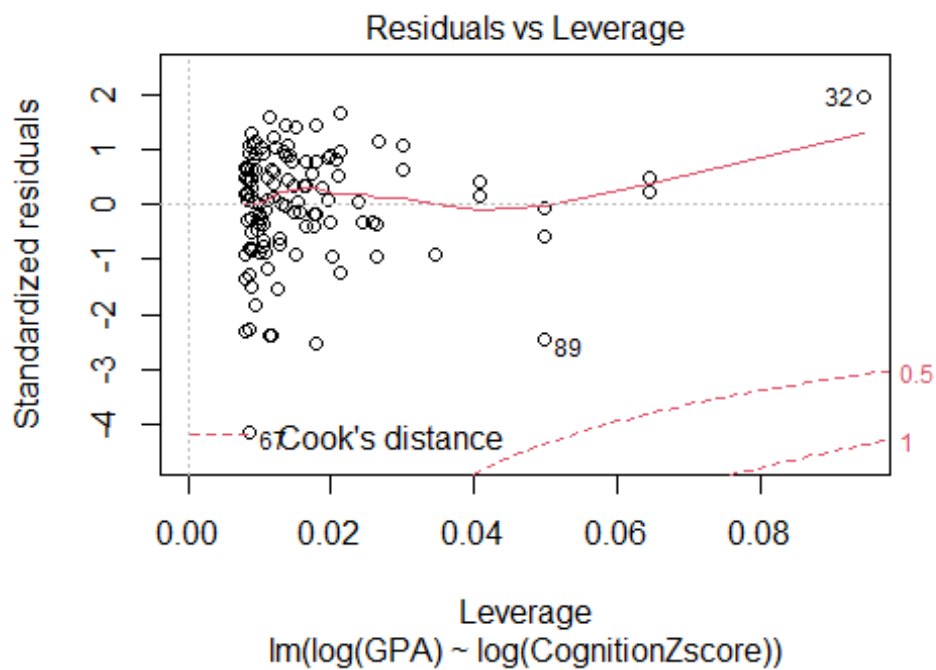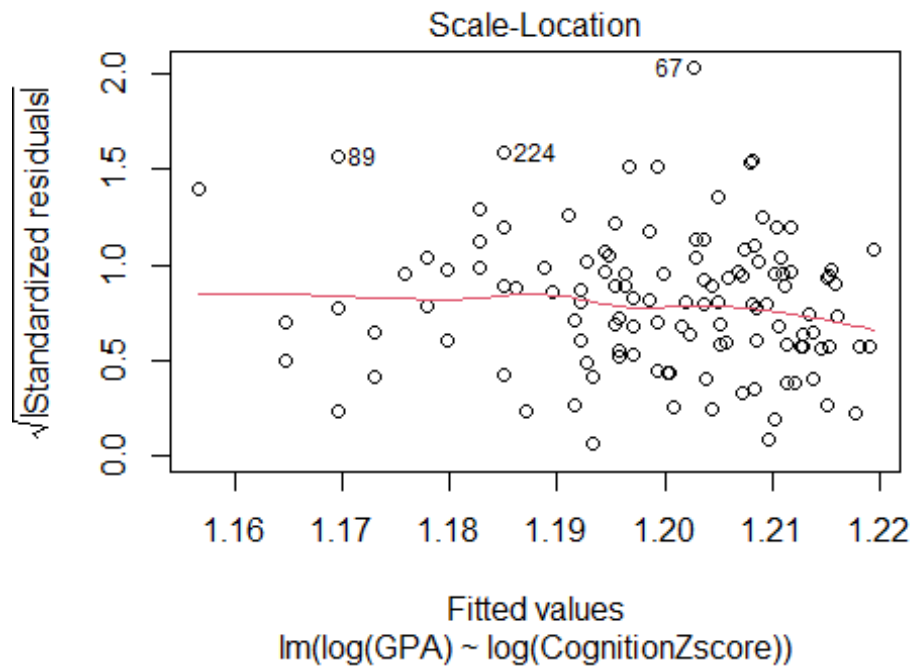
3. Explain why you think that this transformation does or does not improve satisfying the linear model conditions.

This transformation does not improve satisfying the linear model conditions because it does not appear to improve any of the conditions such as linearity, normality, or constant variance.

Note: If you used a different transformation that instead appears to improve the conditions of linearity/normality/constant variance, state this and provide proof based on the plots you create above.