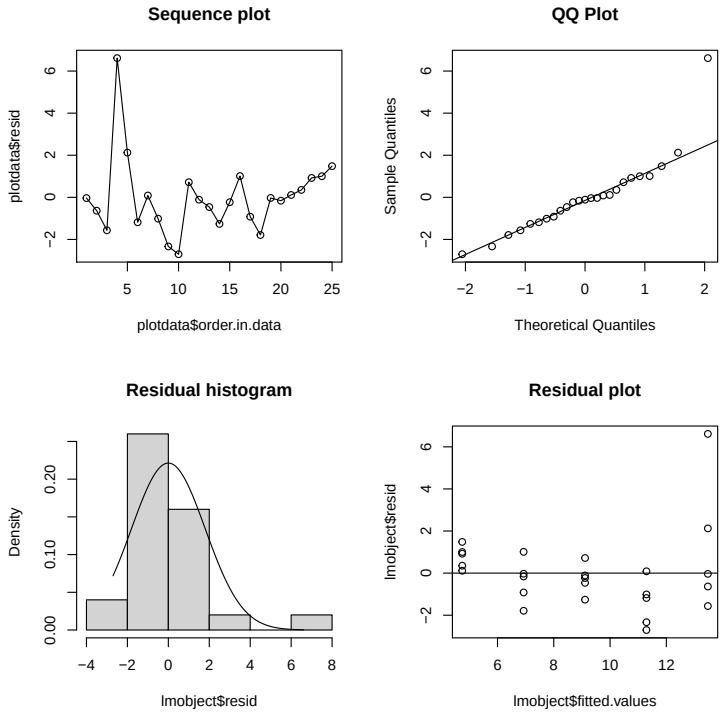
2.2.2: R - Linear Regression Remedial Measures

Stat 5100: Dr. Bean

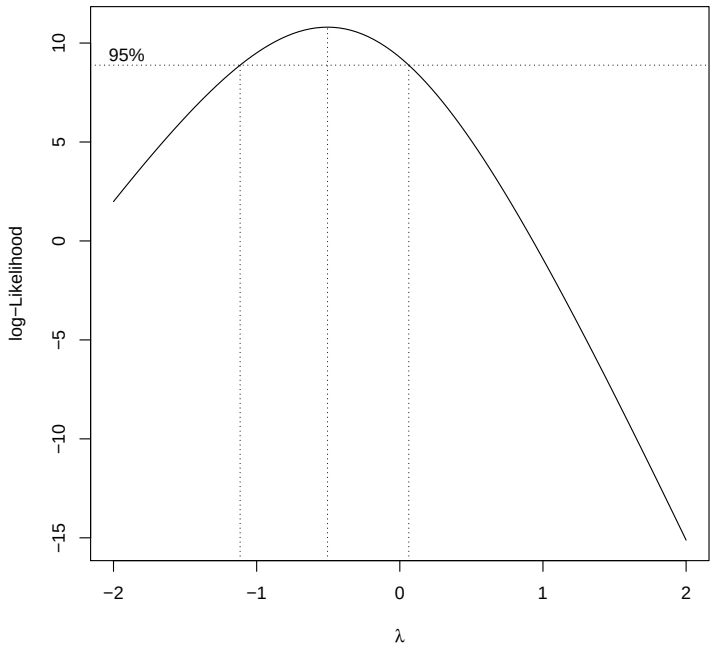
Example: Age and plasma levels for 25 healthy children in a study are reported. Of interest is how plasma level depends on age.

```
# Load data
library(stat5100)
data(plasma)

# Fit regression model and check assumptions
plasma_lm <- lm(level ~ age, data = plasma)

# Check visual assumptions
# (Note: this is a new function you haven't seen before, all this one does
# is it combines the work from seq_plot, qq_plot, residual_hist, and
# residual_plot into one single image)
visual_assumptions(plasma_lm)
```



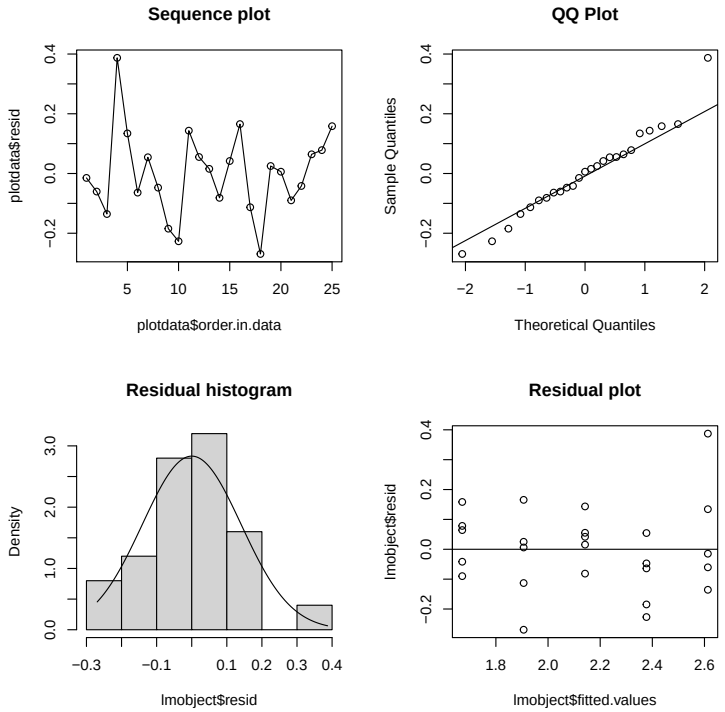

```
# Check assumptions with numerical tests
brown_forsythe_lm(plasma_lm)

## [1] "Brown-forsythe test for constant variance in the residuals:"
## [1] "T-statistic: -1.6903, p-value: 0.1045"
```

```r
cor_normality_lm(plasma_lm)

## Correlation test of normality:
##                   resid expected_norm
## resid         1.0000000     0.9036011
## expected_norm 0.9036011     1.0000000
##
## Total observations: 25
## Make sure to consult with table B.6 for your final result.

ftest_lackfit_lm(plasma_lm)

## Analysis of Variance Table
##
## Model 1: level ~ age
## Model 2: level ~ age
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     23 77.983
## 2     20 55.234  3    22.749 2.7457 0.06994 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Consider transformations
library(MASS)
boxcox(level ~ age, data = plasma)
```

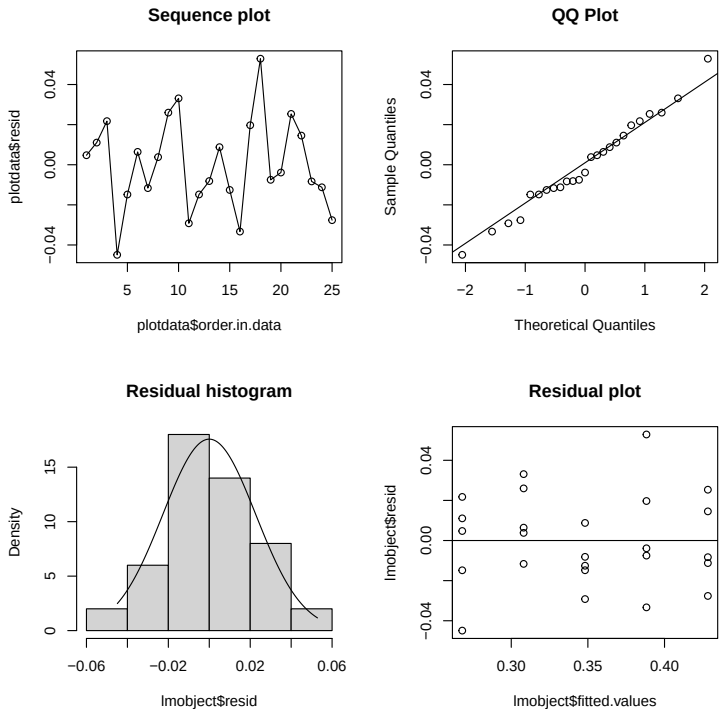



```r
# The above plot tells us we could consider either a log transform or
# possible a 1/(sqrt(response)) type of transform. We will try both and
# show the new visual checks for model assumptions for each.
plasma <- cbind(plasma, log_level = log(plasma$level),
                invsqrt_level = 1 / sqrt(plasma$level))
```

```r
# LOG TRANSFORM
# ---------------

plasma_log_lm <- lm(log_level ~ age, data = plasma)
visual_assumptions(plasma_log_lm)
```




```r
# Numerical checks
brown_forsythe_lm(plasma_log_lm)

## [1] "Brown-forsythe test for constant variance in the residuals:"
## [1] "T-statistic: -0.7958, p-value: 0.4343"

cor_normality_lm(plasma_log_lm)

## Correlation test of normality:
##                    resid expected_norm
## resid          1.0000000     0.9807112
## expected_norm 0.9807112     1.0000000
##
## Total observations: 25
## Make sure to consult with table B.6 for your final result.

# INVERSE SQUARE ROOT TRANSFORM
# -------------------------------

plasma_invsqrt_lm <- lm(invsqrt_level ~ age, data = plasma)
visual_assumptions(plasma_invsqrt_lm)
```




```
# Numerical checks
brown_forsythe_lm(plasma_invsqrt_lm)

## [1] "Brown-forsythe test for constant variance in the residuals:"
## [1] "T-statistic: -0.5031, p-value: 0.6197"

cor_normality_lm(plasma_invsqrt_lm)

## Correlation test of normality:
##                    resid expected_norm
## resid         1.0000000     0.9918794
## expected_norm 0.9918794     1.0000000
##
## Total observations: 25
## Make sure to consult with table B.6 for your final result.
```