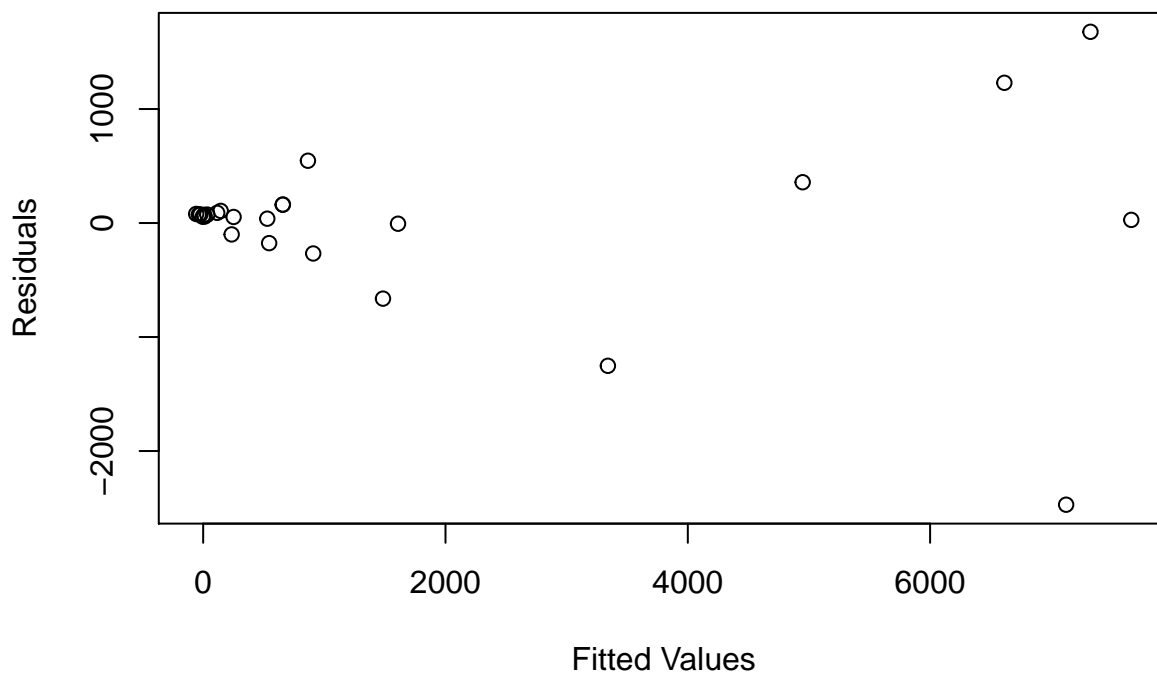


```
## Demo for transformations and interactions

## Florida oranges revisited
dat <- read.csv("csv/florange.csv")
lm.1 <- lm(dat$boxes ~ dat$acres)
summary(lm.1)

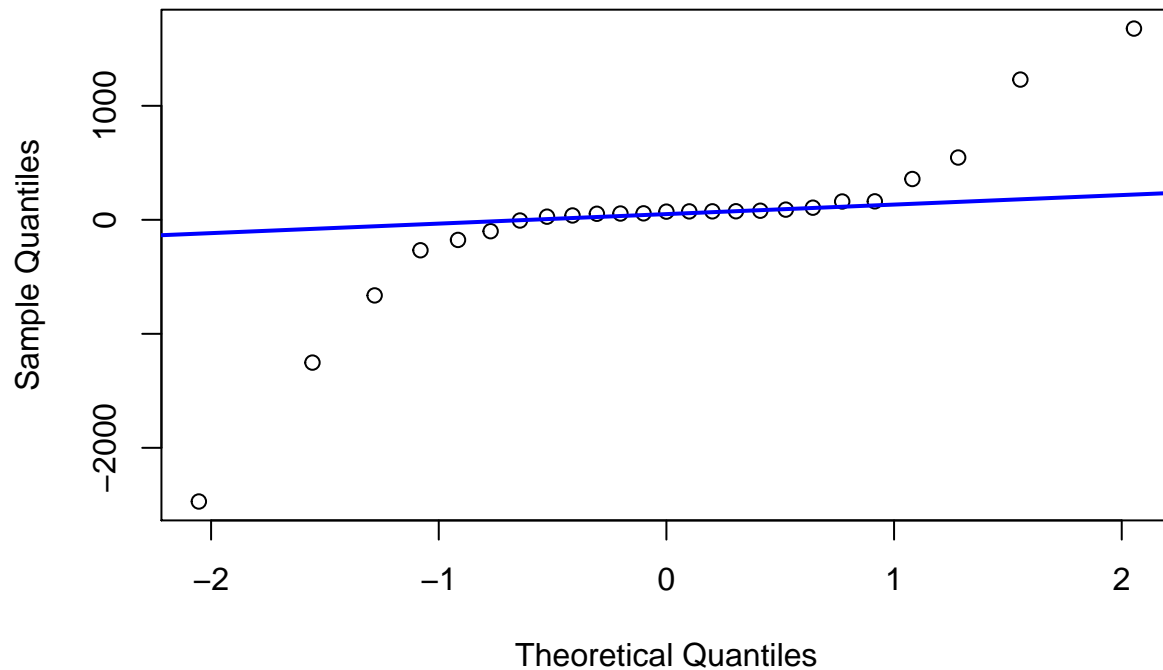
##
## Call:
## lm(formula = dat$boxes ~ dat$acres)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2470.81    -6.17    71.72   106.46  1677.32
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -85.391989  186.178031  -0.459   0.651
## dat$acres     0.116717   0.006761  17.263 1.16e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 754.4 on 23 degrees of freedom
## Multiple R-squared:  0.9284, Adjusted R-squared:  0.9252
## F-statistic: 298 on 1 and 23 DF, p-value: 1.164e-14

# Recall: residuals had non-constant variance
# (variance increases with fitted values)
plot(lm.1$fitted.values,
     lm.1$residuals,
     xlab = "Fitted Values",
     ylab = "Residuals")
```



```
qqnorm(lm.1$residuals)
qqline(lm.1$residuals, col = "blue", lwd = 2)
```

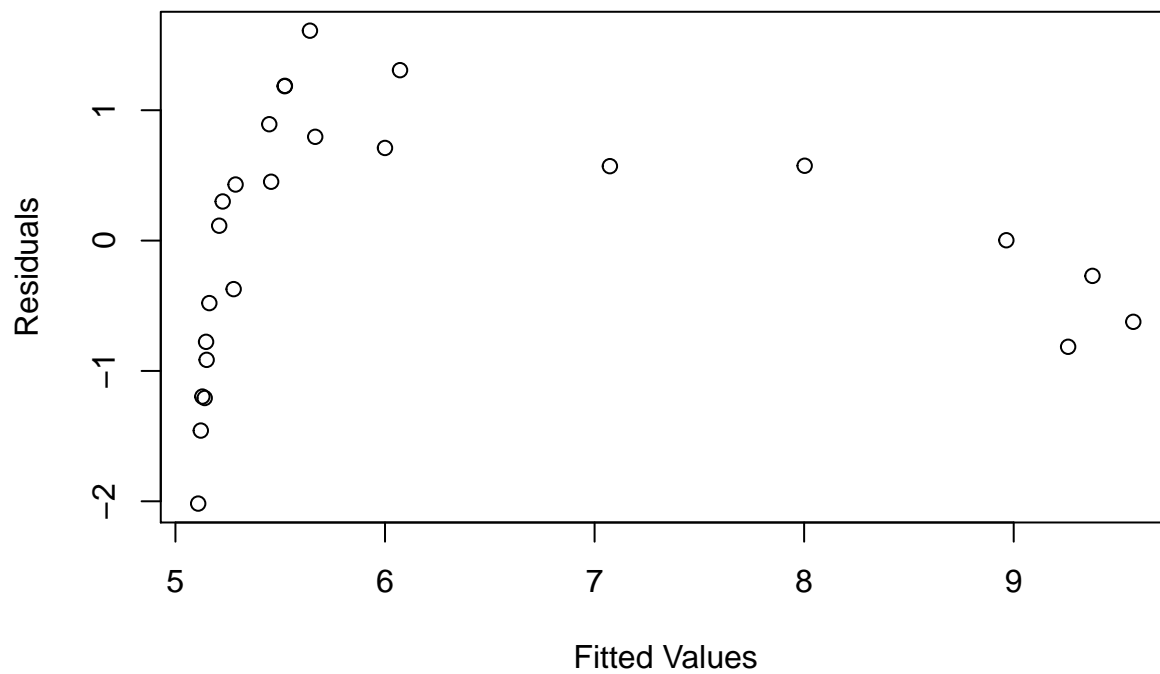
Normal Q-Q Plot



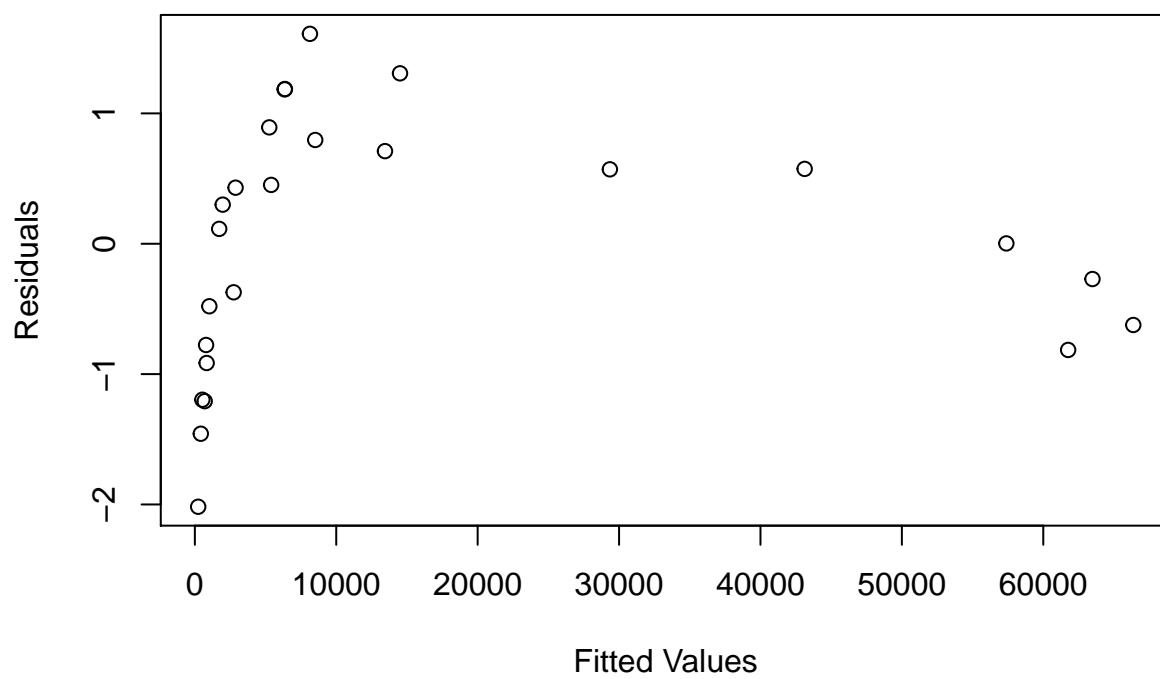
```
# Try log-transforming y
lm.log <- lm(log(dat$boxes) ~ dat$acres)
summary(lm.log)
```

```
##
## Call:
## lm(formula = log(dat$boxes) ~ dat$acres)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0175 -0.7767  0.1142  0.7106  1.6102
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.093e+00  2.425e-01  20.997  < 2e-16 ***
## dat$acres    6.748e-05  8.808e-06   7.661  8.95e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9828 on 23 degrees of freedom
## Multiple R-squared:  0.7184, Adjusted R-squared:  0.7062
## F-statistic: 58.69 on 1 and 23 DF,  p-value: 8.948e-08
```

```
plot(lm.log$fitted.values,
     lm.log$residuals,
     xlab = "Fitted Values",
     ylab = "Residuals")
```

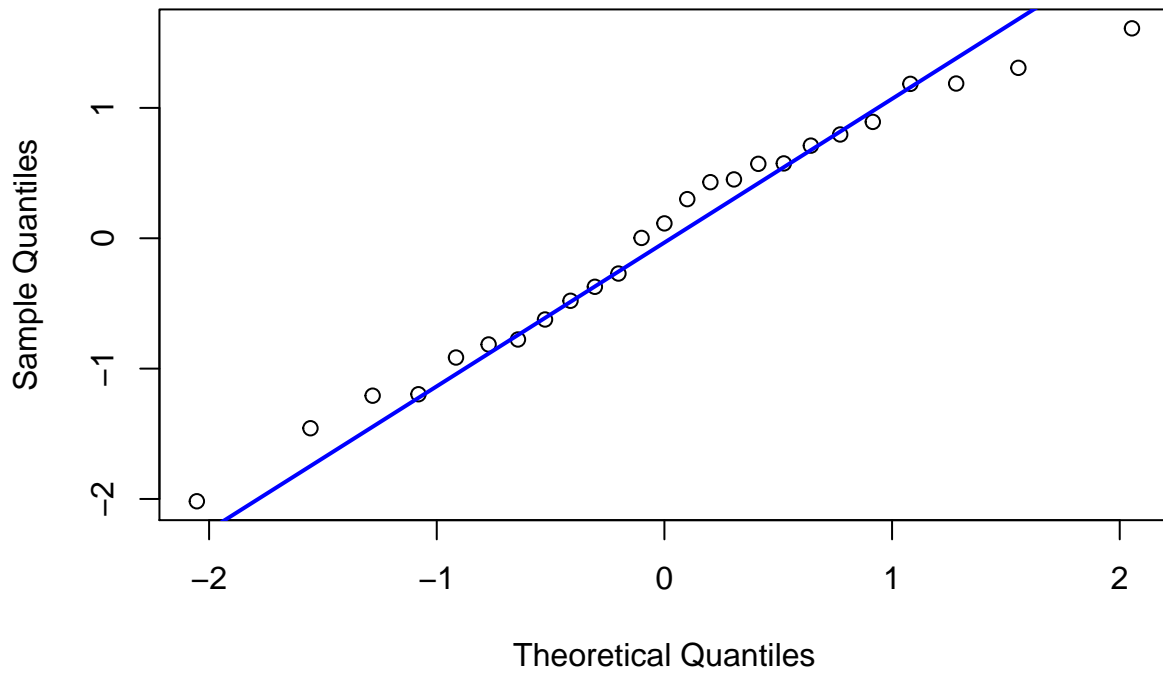


```
plot(dat$acres, lm.log$residuals, xlab = "Fitted Values", ylab = "Residuals")
```

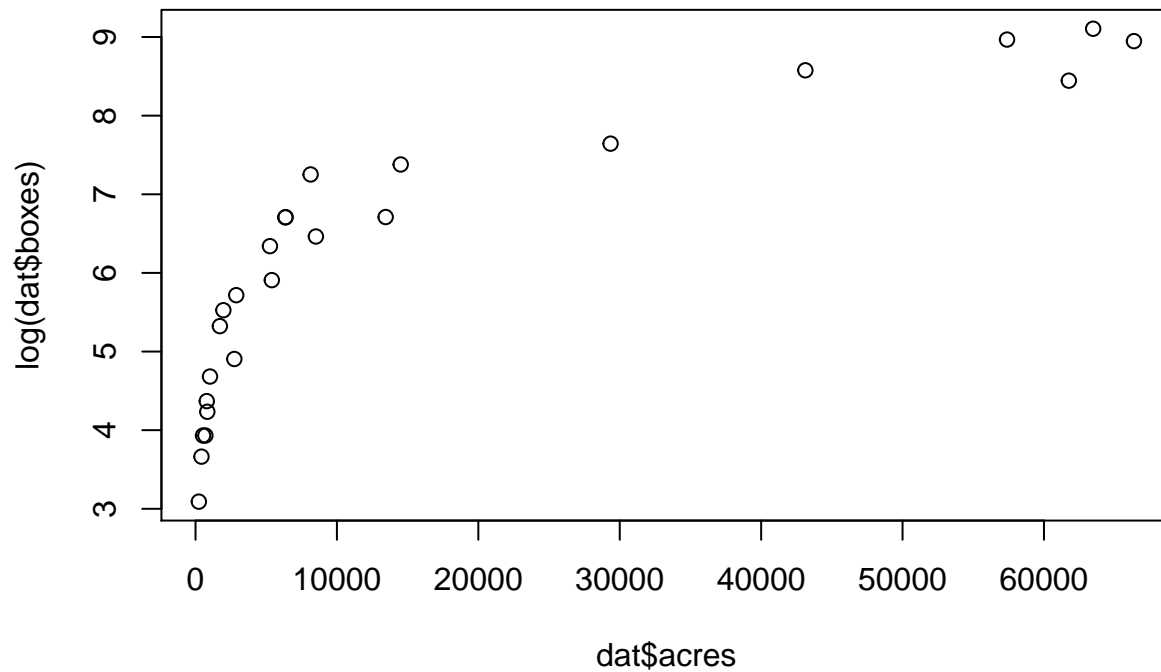


```
qqnorm(lm.log$residuals)
qqline(lm.log$residuals, col = "blue", lwd = 2)
```

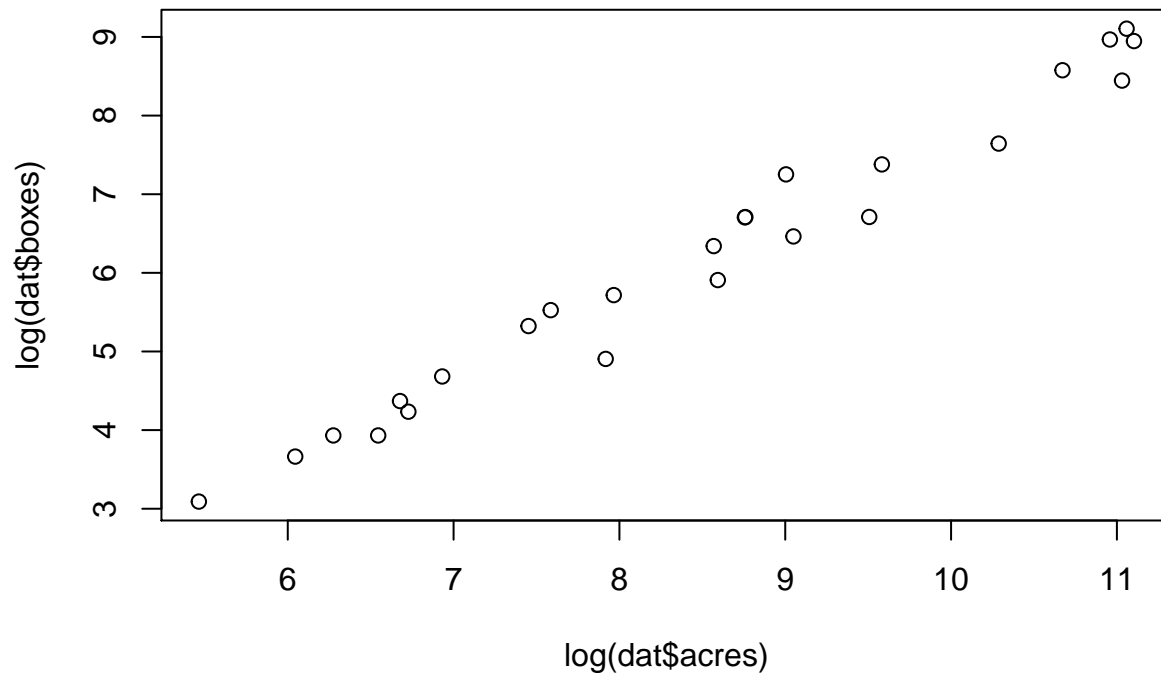
Normal Q-Q Plot



```
# Does the plot of residuals vs x suggest a problem  
# Let's take a closer look  
plot(dat$acres, log(dat$boxes)) # evidently not linear!
```

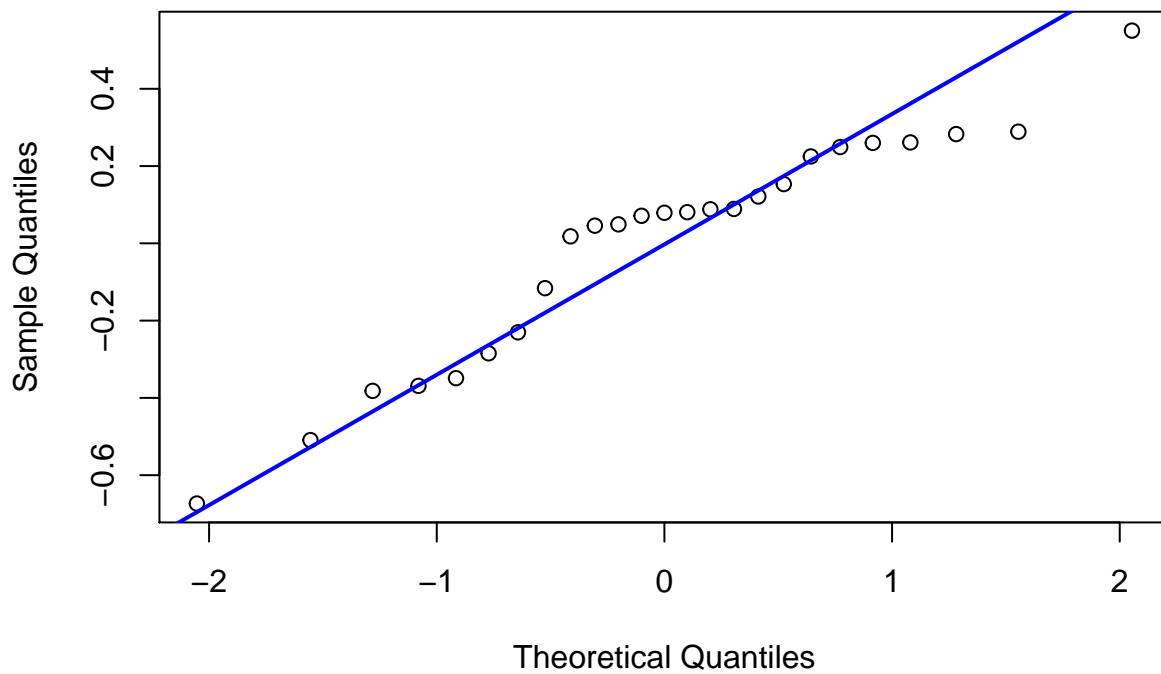


```
# Log-transform x as well  
plot(log(dat$acres), log(dat$boxes)) # looks much more linear!
```

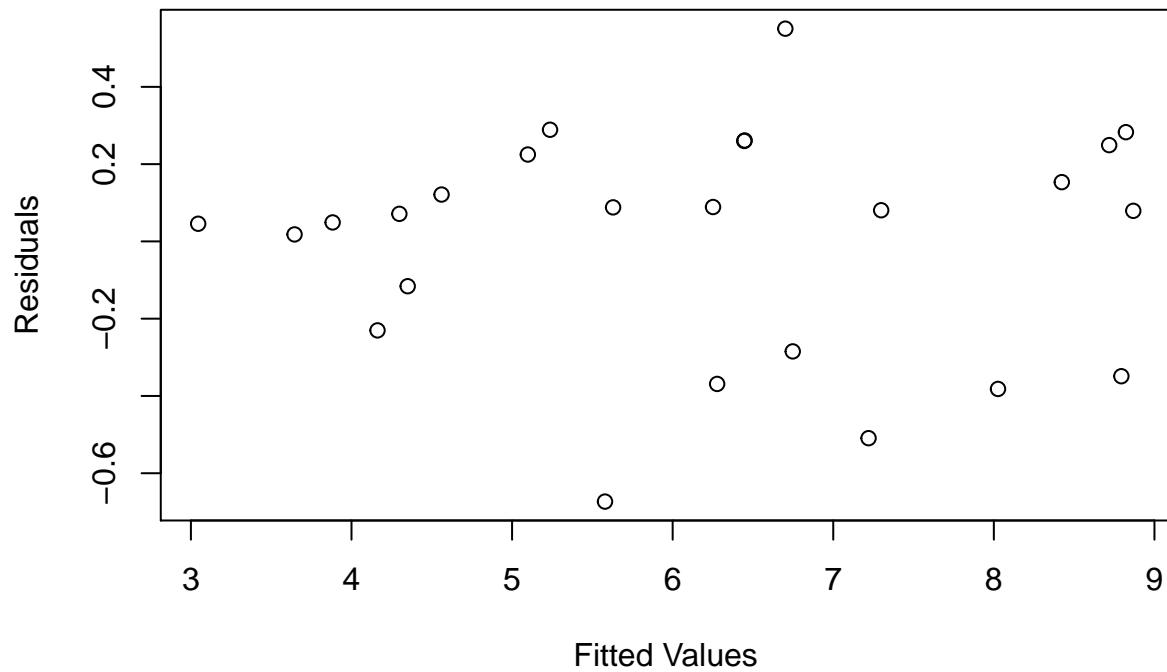


```
lm.loglog <- lm(log(dat$boxes) ~ log(dat$acres))
qqnorm(lm.loglog$residuals)
qqline(lm.loglog$residuals, col = "blue", lwd = 2)
```

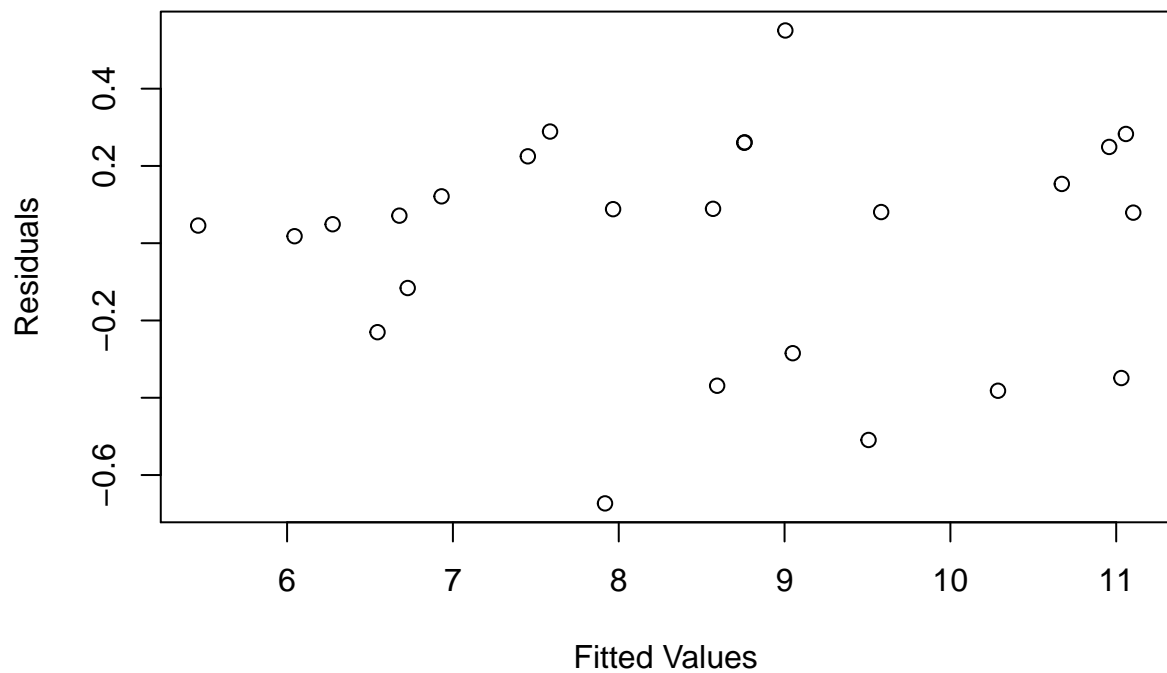
Normal Q-Q Plot



```
plot(lm.loglog$fitted.values,
     lm.loglog$residuals,
     xlab = "Fitted Values",
     ylab = "Residuals")
```

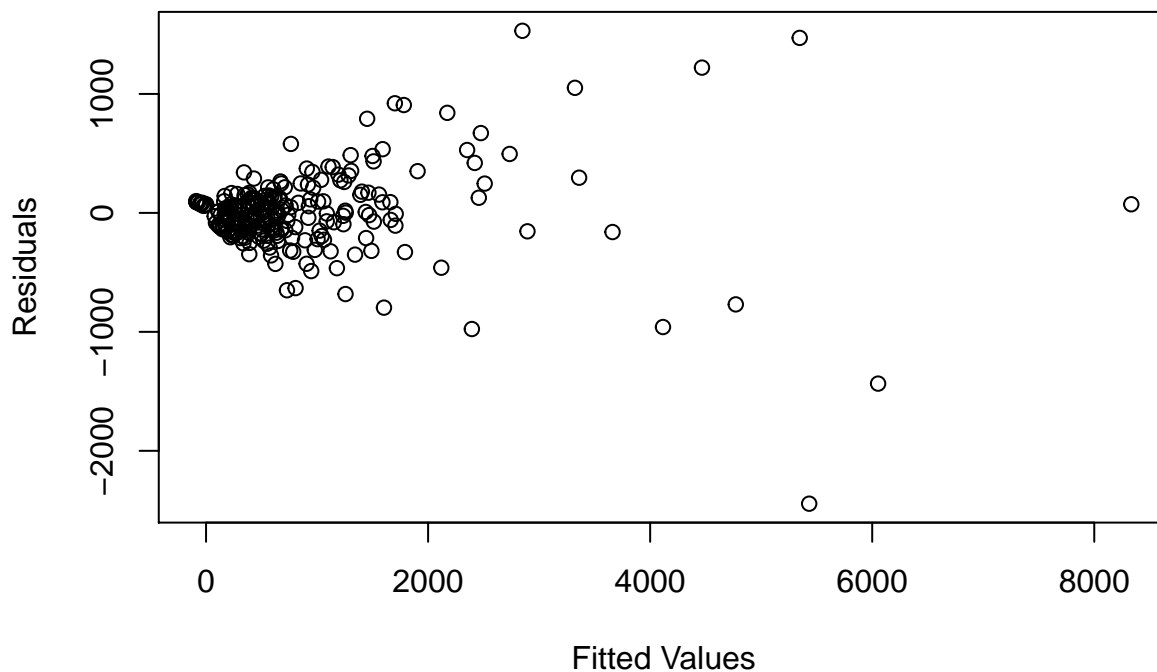


```
plot(log(dat$acres),
     lm.loglog$residuals,
     xlab = "Fitted Values",
     ylab = "Residuals")
```



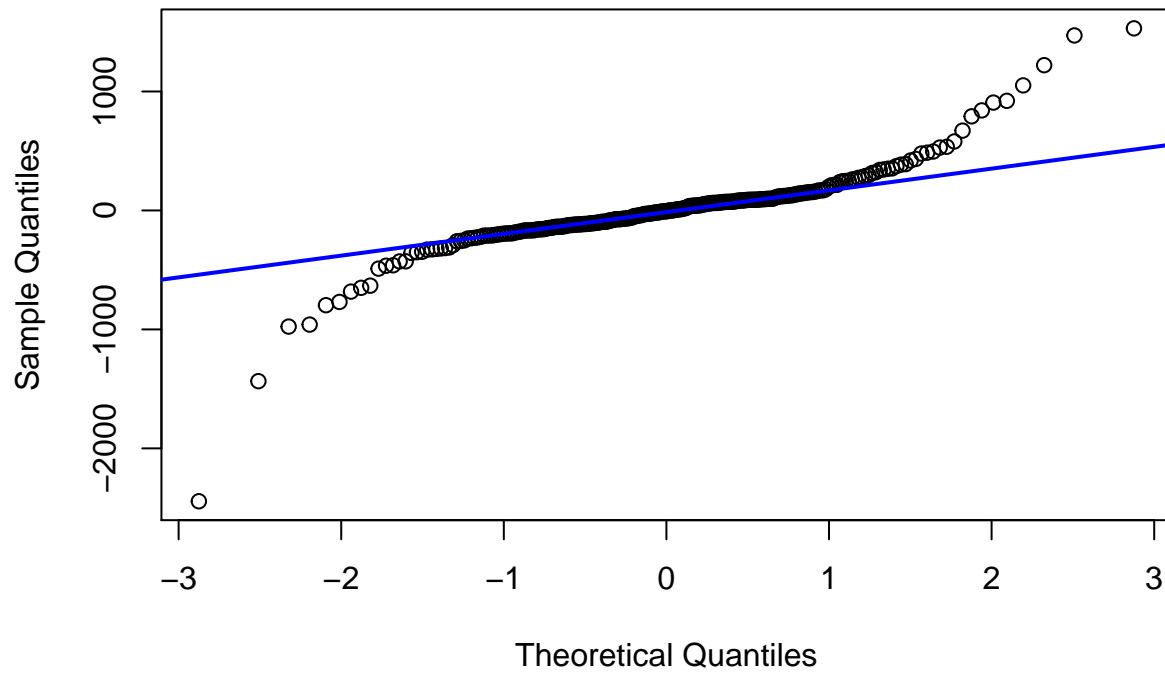
```
## Python data revisited
python <- read.csv("csv/FLpython.csv")
python$male <- ifelse(python$sex == 'M', 1, 0) # 1 = M, 0 = F
mpf2 <- lm(fat ~ male + mass + svl, data = python)
summary(mpf2)
```

```
##
## Call:
## lm(formula = fat ~ male + mass + svl, data = python)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2444.44  -137.38    -6.66   109.22  1530.81
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  204.09840   132.30121    1.543  0.1242
## male        -196.71705    47.16396   -4.171 4.22e-05 ***
## mass          0.11788     0.00524   22.495 < 2e-16 ***
## svl         -1.59841     0.76433   -2.091  0.0375 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 360.2 on 244 degrees of freedom
## Multiple R-squared:  0.897, Adjusted R-squared:  0.8957
## F-statistic: 708.2 on 3 and 244 DF, p-value: < 2.2e-16
# Residual plot: vs fitted values
plot(mpf2$fitted.values,
     mpf2$residuals,
     xlab = "Fitted Values",
     ylab = "Residuals")
```



```
## QQ plot of residuals
qqnorm(mpf2$residuals)
qqline(mpf2$residuals, col = "blue", lwd = 2)
```

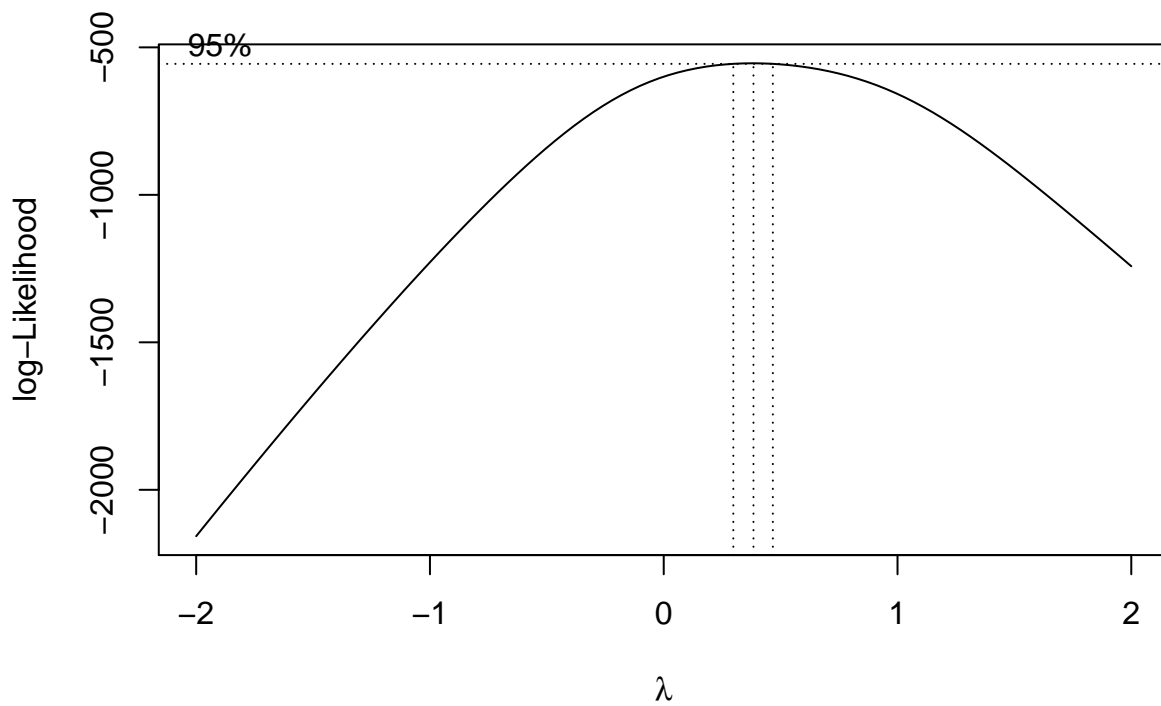
Normal Q-Q Plot



```
library(MASS)
```

```
# Try a Box-Cox transformation
```

```
bc <- boxcox(mpf2)
```



```
lambda <- bc$x[which.max(bc$y)]
```

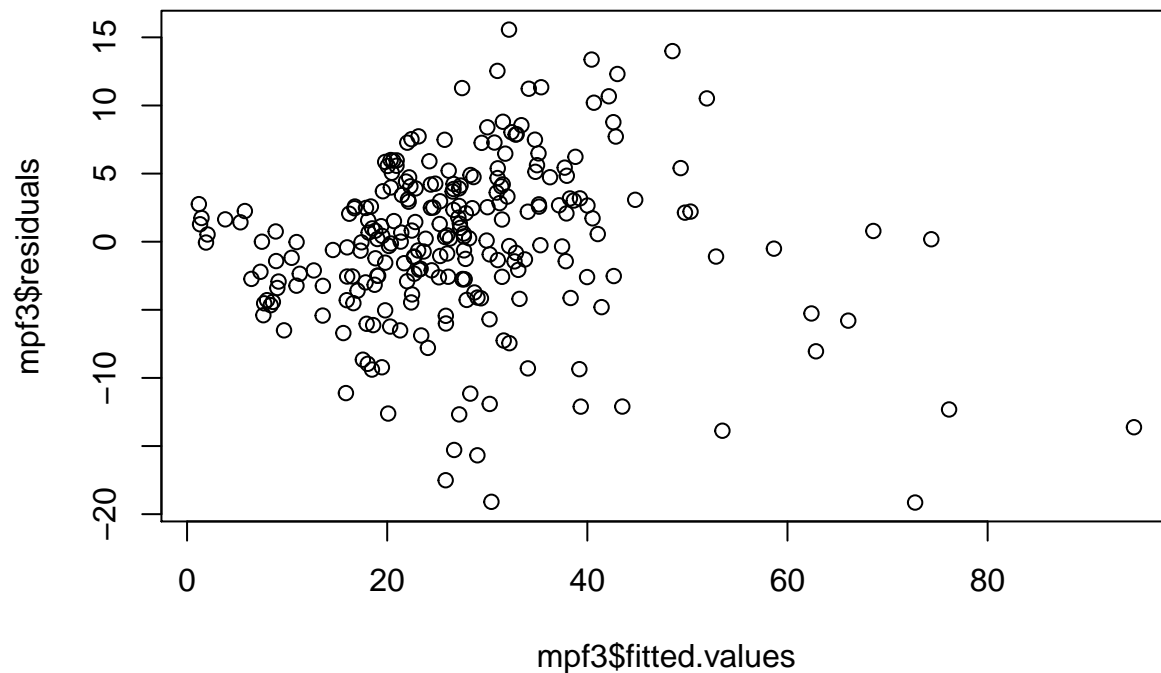
```
mpf3 <- lm((fat ^ lambda - 1) / lambda ~ male + mass + svl, data = python)
```

```
summary(mpf3)
```

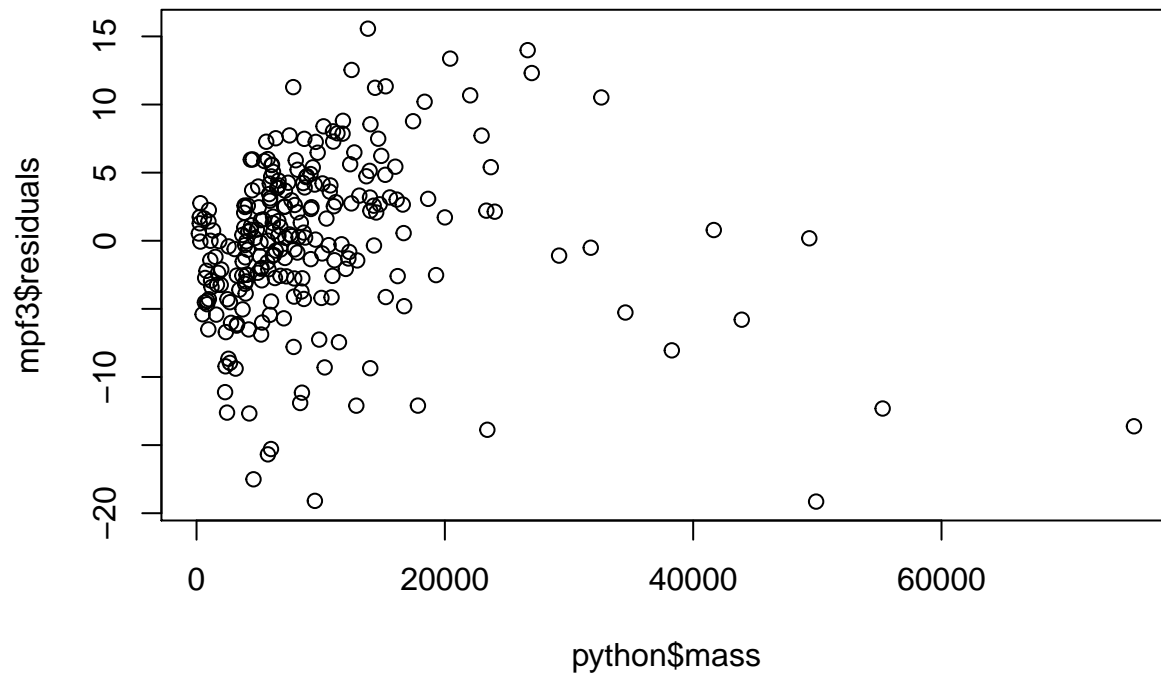


```
##
## Call:
## lm(formula = (fat~lambda - 1)/lambda ~ male + mass + svl, data = python)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.146  -2.910   0.297   3.688  15.568
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.0558134   2.1813183  -3.693 0.000273 ***
## male         -1.7849310   0.7776166  -2.295 0.022560 *
## mass          0.0004461   0.0000864    5.164 5.03e-07 ***
## svl           0.1431492   0.0126019  11.359 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.939 on 244 degrees of freedom
## Multiple R-squared:  0.8356, Adjusted R-squared:  0.8336
## F-statistic: 413.5 on 3 and 244 DF,  p-value: < 2.2e-16
```

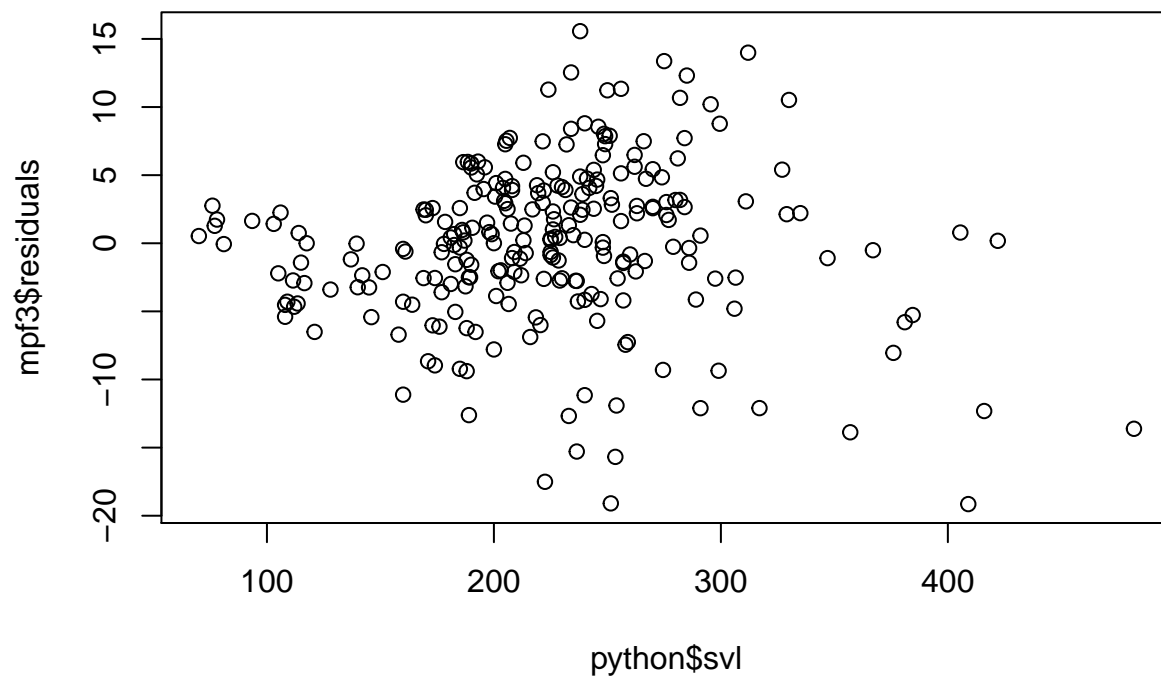
```
plot(mpf3$fitted.values, mpf3$residuals)
```



```
plot(python$mass, mpf3$residuals)
```

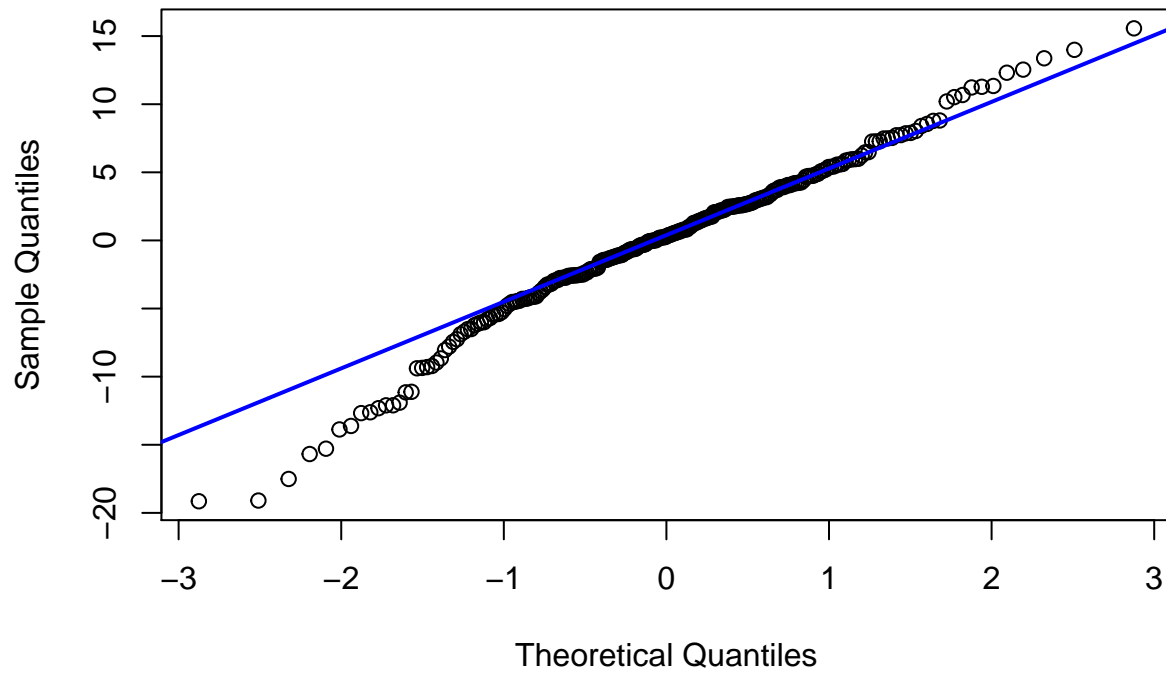


```
plot(python$svl, mpf3$residuals)
```



```
qqnorm(mpf3$residuals)
qqline(mpf3$residuals, col = "blue", lwd = 2)
```

Normal Q-Q Plot



still some skew, but better!