# HW1

September 14, 2018

```
In [1]: MortalityRate <- read.table("MortalityRate.txt", header = T)
```

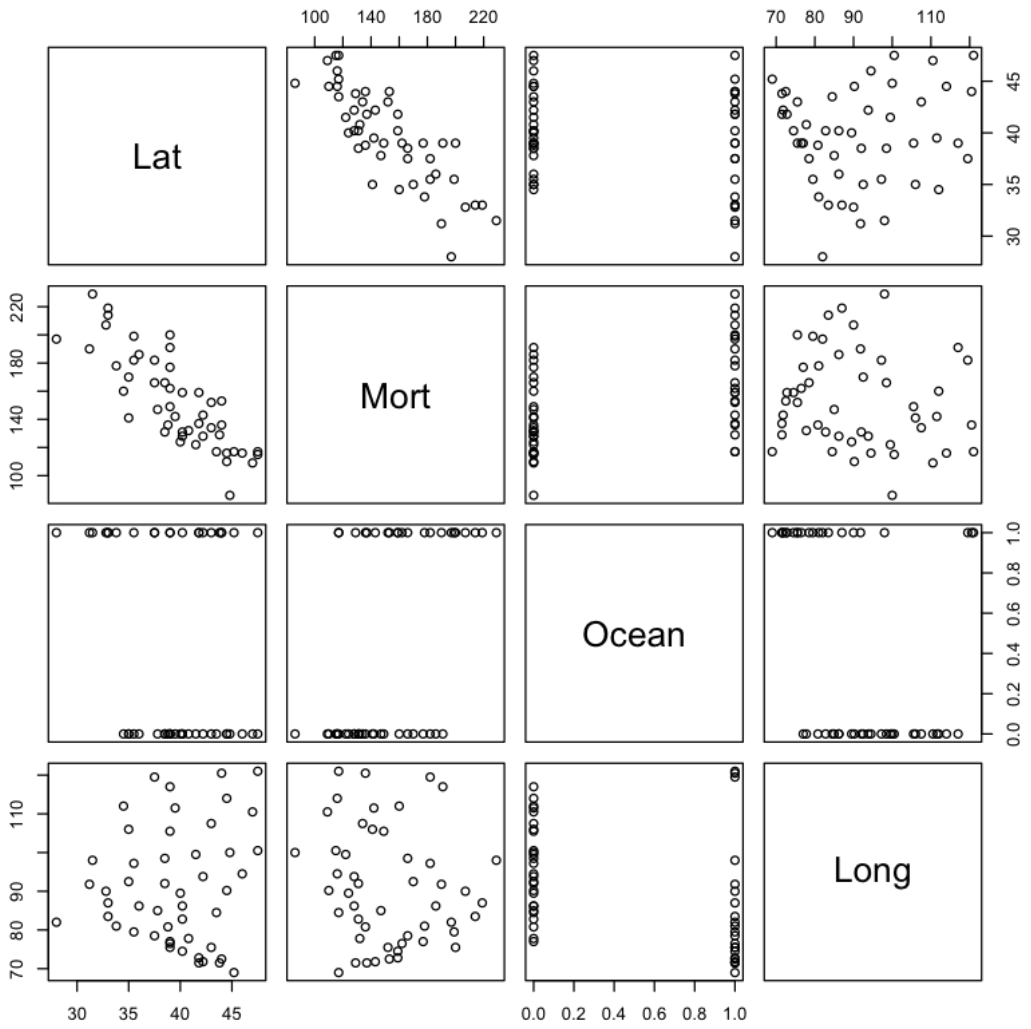
```
In [2]: head(MortalityRate)
```

| State | Lat | Mort | Ocean | Long |
|---|---|---|---|---|
| Alabama | 33.0 | 219 | 1 | 87.0 |
| Arizona | 34.5 | 160 | 0 | 112.0 |
| Arkansas | 35.0 | 170 | 0 | 92.5 |
| California | 37.5 | 182 | 1 | 119.5 |
| Colorado | 39.0 | 149 | 0 | 105.5 |
| Connecticut | 41.8 | 159 | 1 | 72.8 |

```
In [3]: cat("Mean:\n", mean(MortalityRate$Mort), "\n\n")
        cat("Standard Deviation:\n", sd(MortalityRate$Mort), "\n\n")
        cat("Median:\n", median(MortalityRate$Mort), "\n\n")
        cat("Variance:\n", var(MortalityRate$Mort), "\n\n")
```

```
Mean:
 152.8776

Standard Deviation:
 33.42818

Median:
 147

Variance:
 1117.443
```

```
In [4]: pairs(MortalityRate[2:5])
```

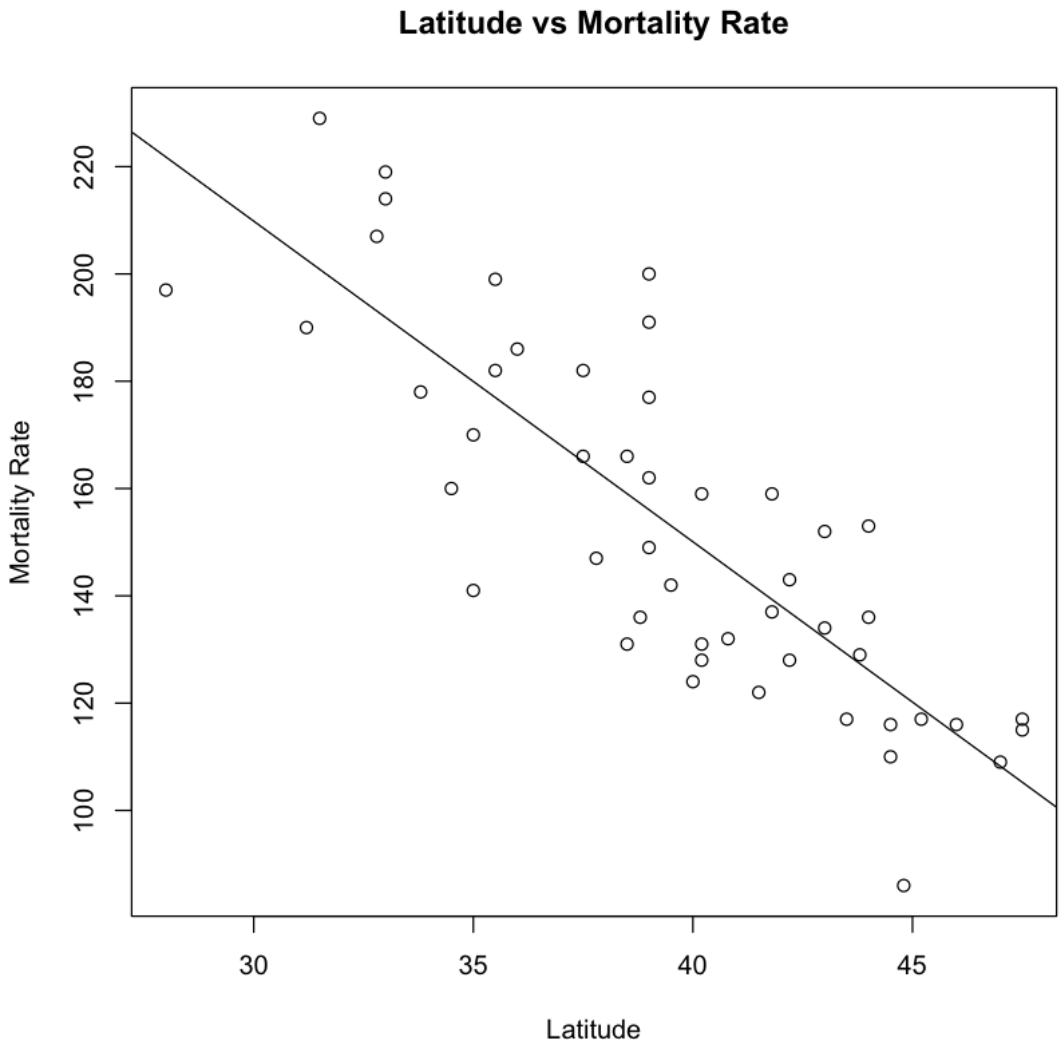



The correlation matrix appears to show a strong negative correlation between latitude and mortality rate.

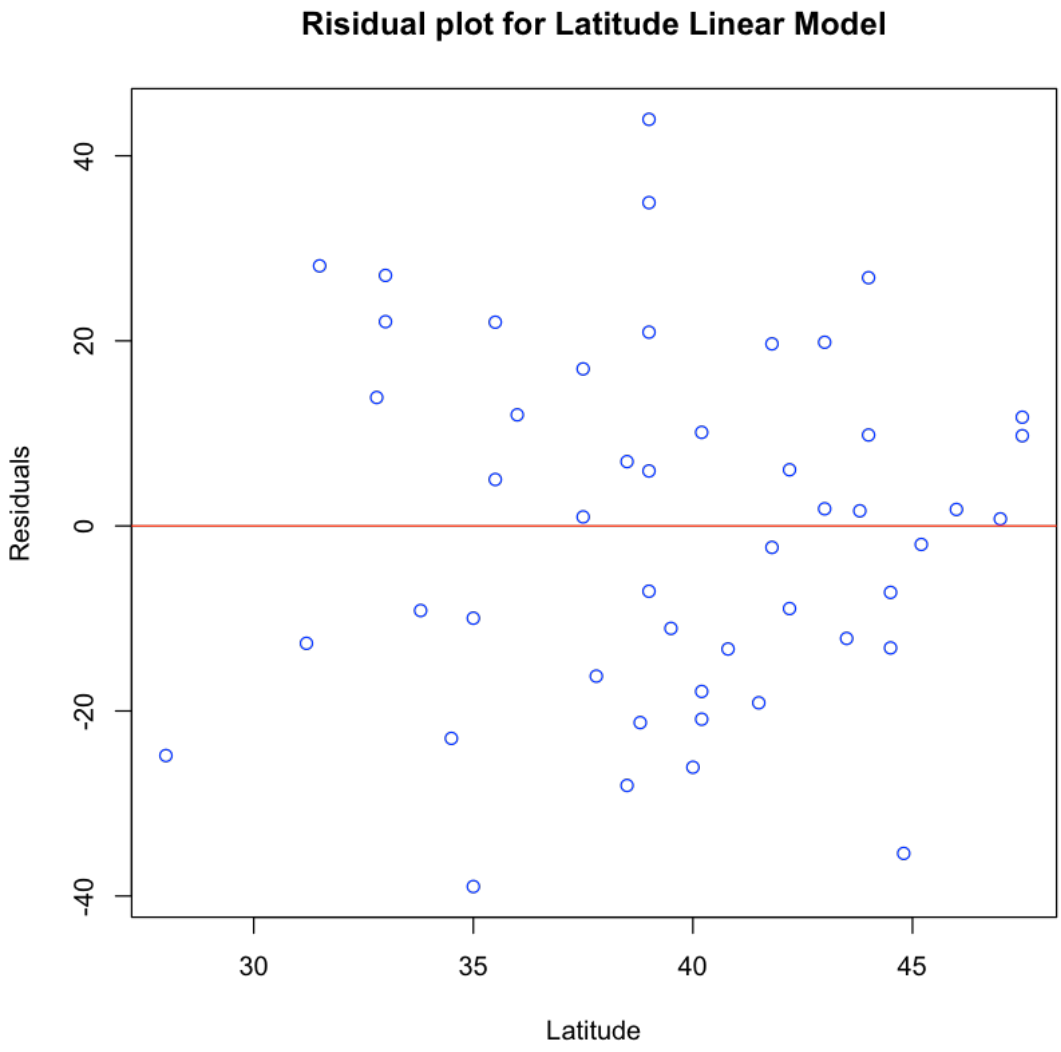
```
In [5]: linear_lat <- lm(formula = MortalityRate$Mort ~ MortalityRate$Lat)

        summary(linear_lat)

        plot(MortalityRate$Lat, MortalityRate$Mort, xlab="Latitude", ylab="Mortality Rate", ma
        abline(linear_lat)
```

```
Call:
lm(formula = MortalityRate$Mort ~ MortalityRate$Lat)
```

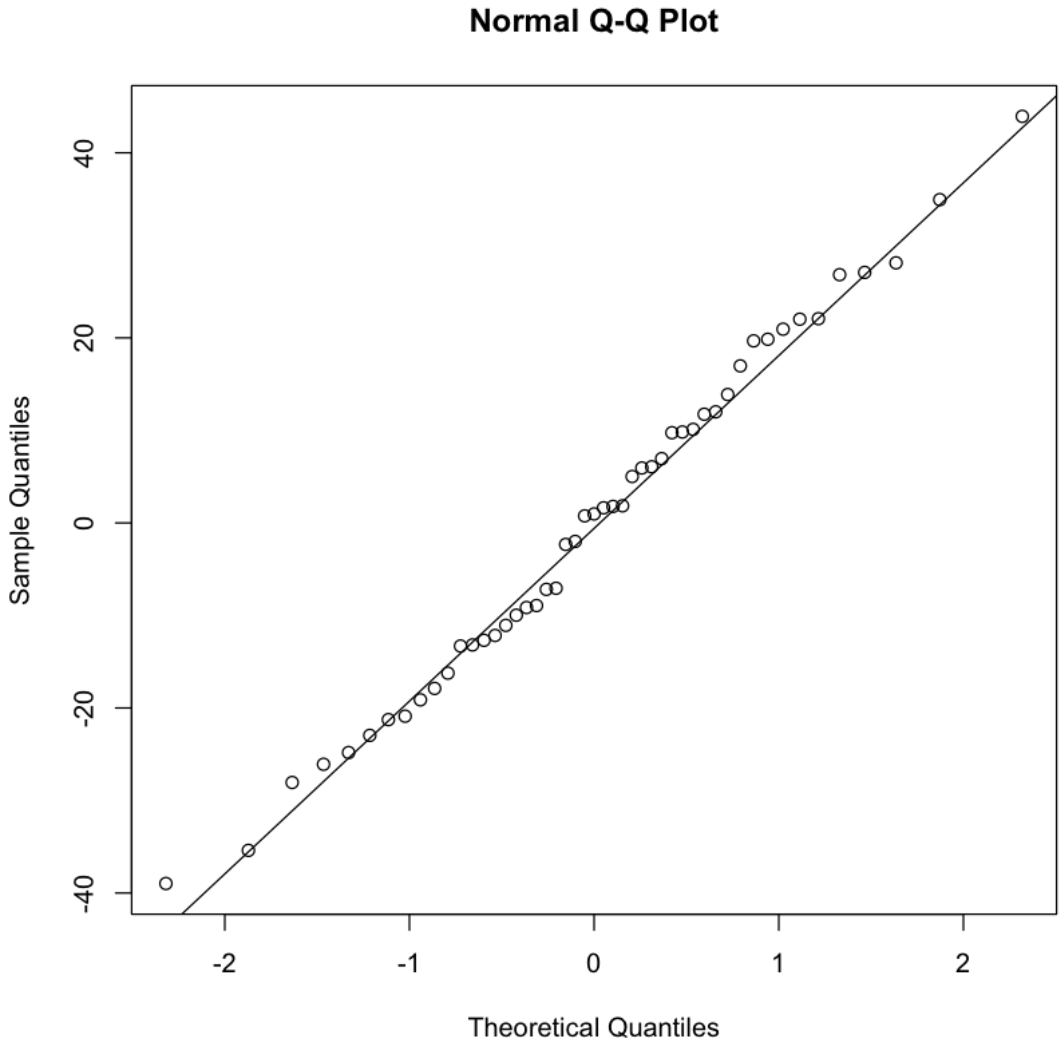
```
Residuals:
    Min      1Q  Median      3Q     Max
-38.972 -13.185   0.972  12.006  43.938

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       389.1894    23.8123   16.34  < 2e-16 ***
MortalityRate$Lat  -5.9776     0.5984   -9.99 3.31e-13 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 19.12 on 47 degrees of freedom
Multiple R-squared:  0.6798,Adjusted R-squared:  0.673
F-statistic:  99.8 on 1 and 47 DF,  p-value: 3.309e-13
```



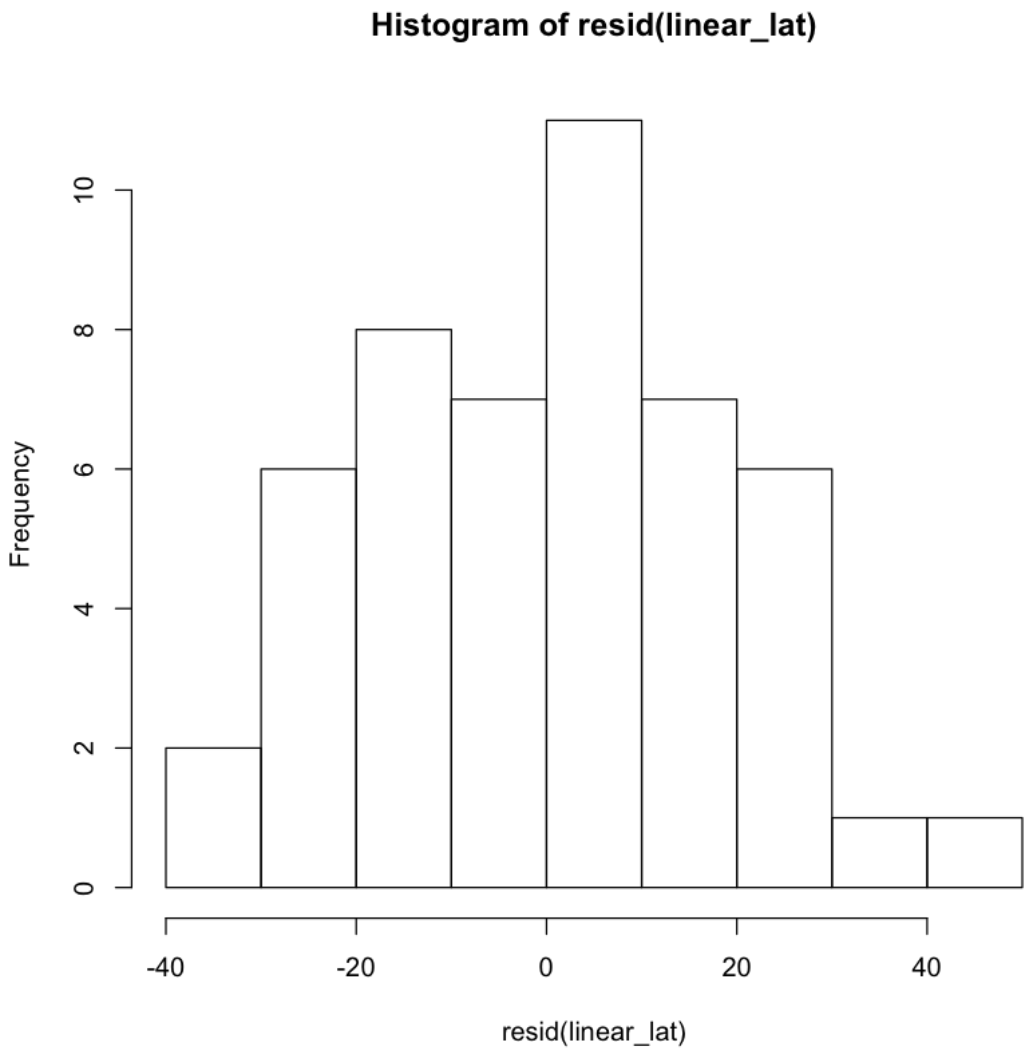
```
In [6]: plot(x = MortalityRate$Lat, y = resid(linear_lat), col = 'blue', xlab="Latitude", ylab=
        abline(h = 0, col = 'red')
```




```
In [25]: qqnorm(resid(linear_lat))
         qqline(resid(linear_lat))
```

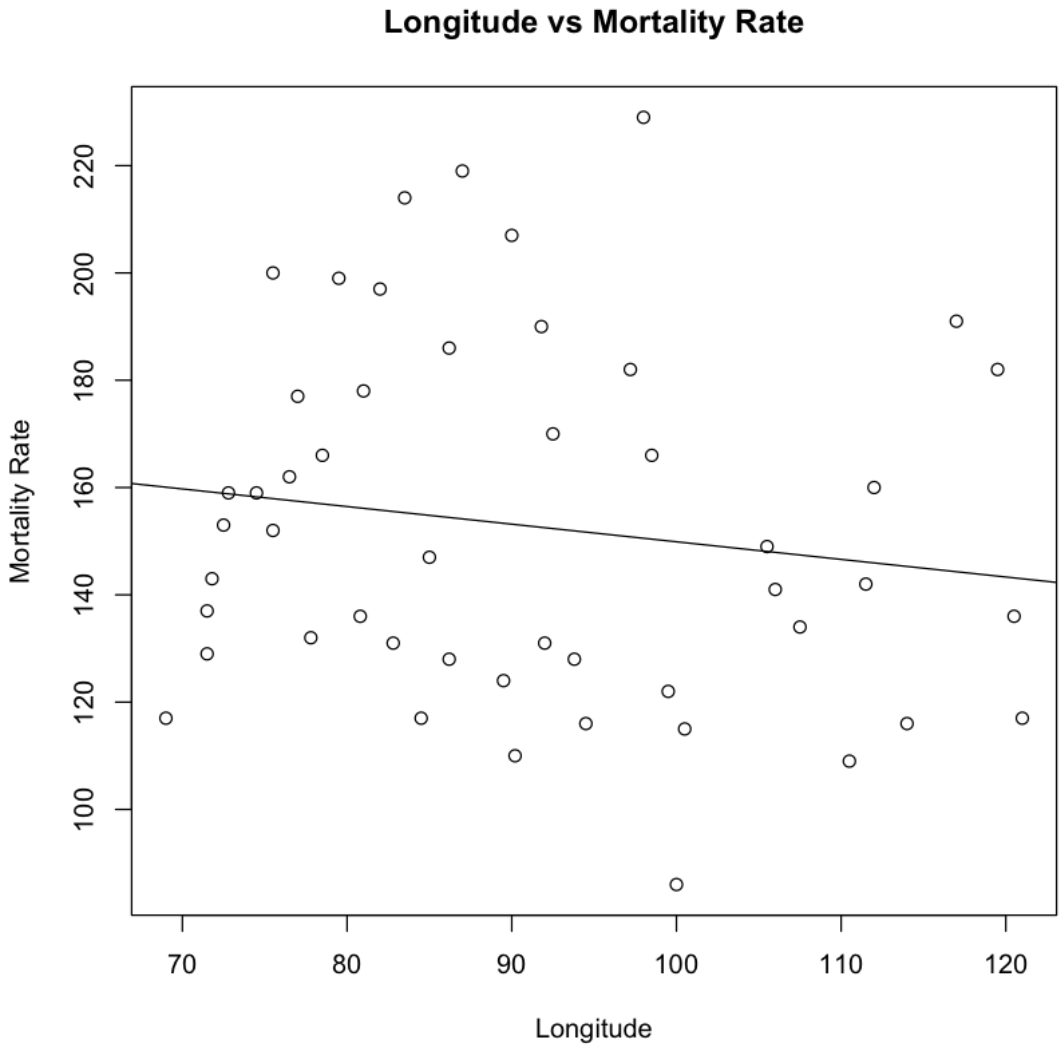



In [26]: `hist(resid(linear_lat))`

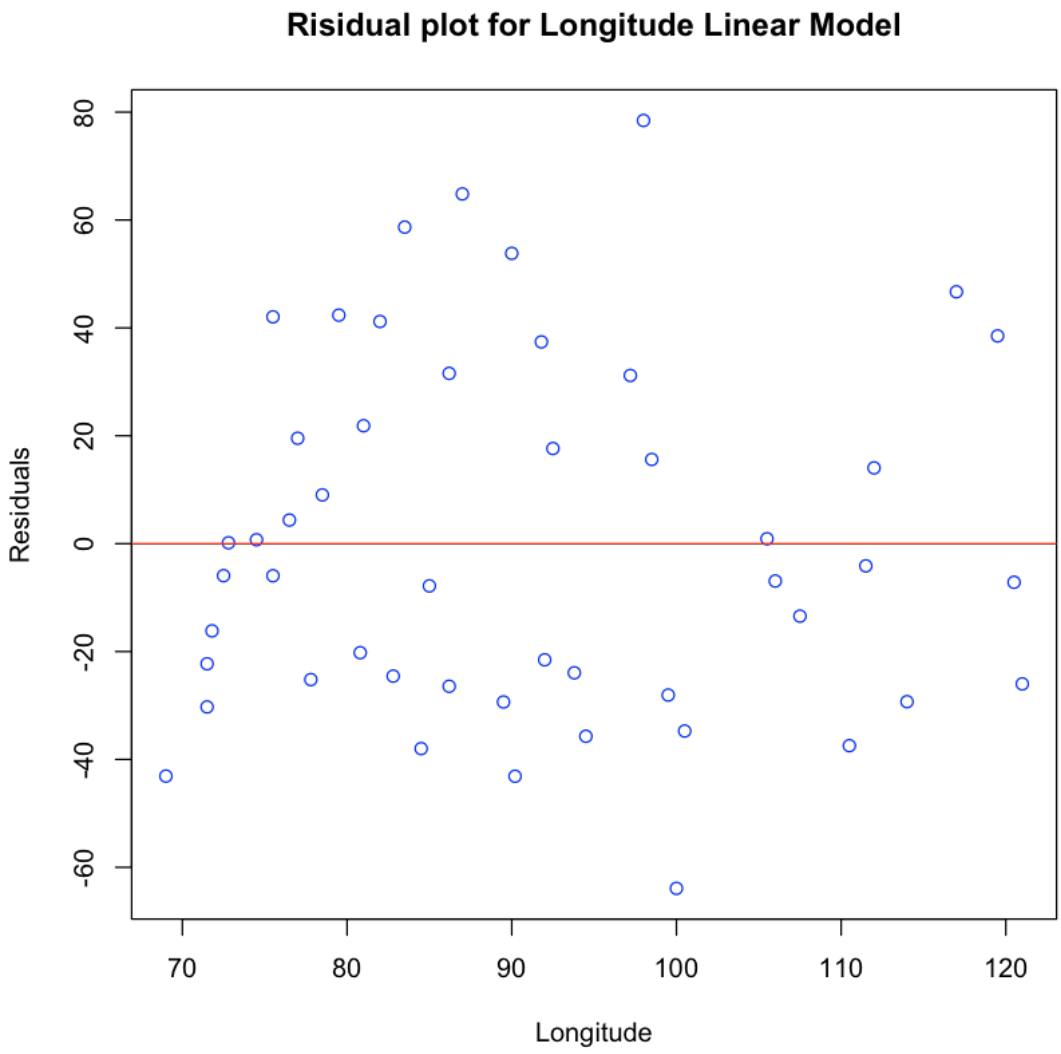



```
In [8]: linear_long <- lm(formula = MortalityRate$Mort ~ MortalityRate$Long)

        summary(linear_long)

        plot(MortalityRate$Long, MortalityRate$Mort, xlab="Longitude", ylab="Mortality Rate", m
        abline(linear_long)


Call:
lm(formula = MortalityRate$Mort ~ MortalityRate$Long)

Residuals:
    Min      1Q  Median      3Q     Max
```

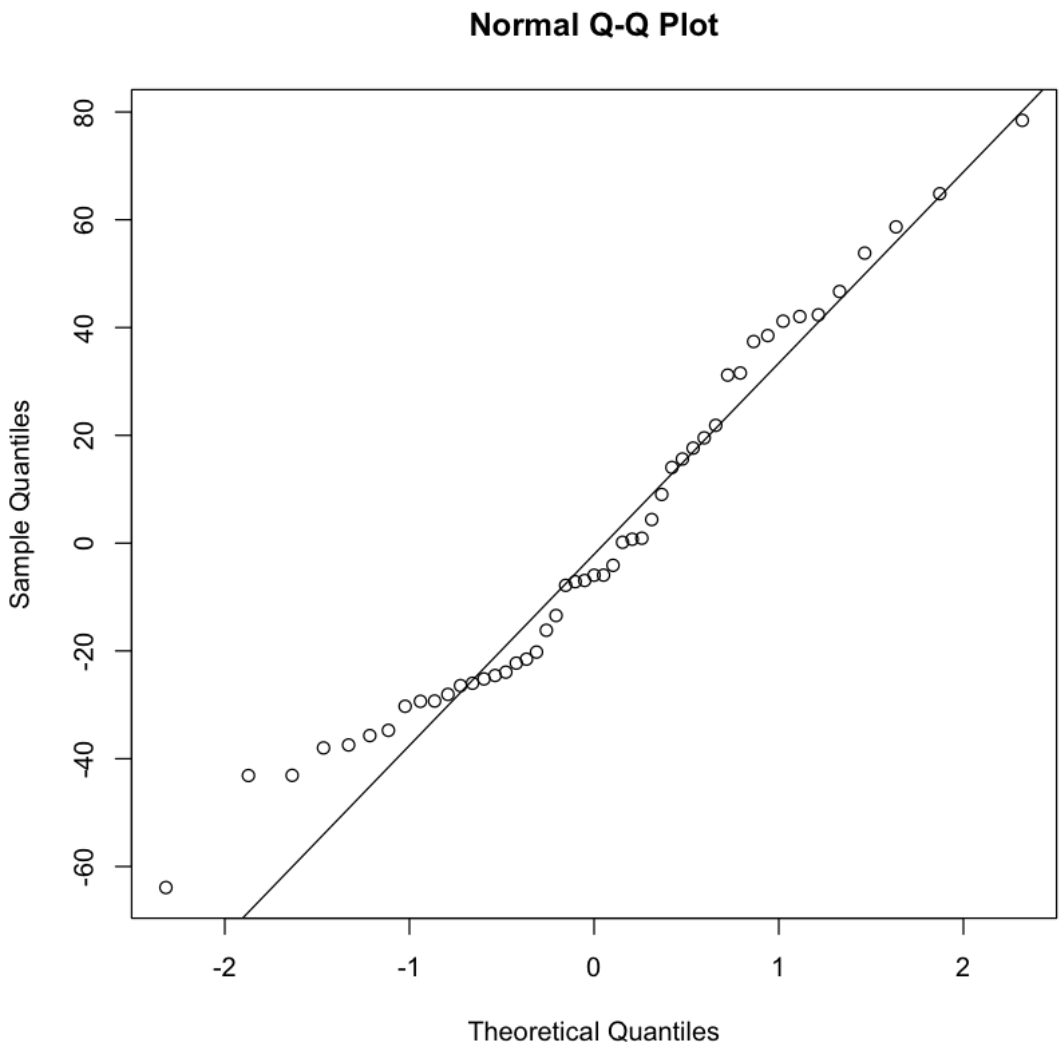
```
-63.898 -25.995  -5.952  21.856  78.444

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        182.7696    29.8893   6.115  1.8e-07 ***
MortalityRate$Long  -0.3287     0.3245  -1.013    0.316
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 33.42 on 47 degrees of freedom
Multiple R-squared:  0.02137,Adjusted R-squared:  0.0005491
F-statistic: 1.026 on 1 and 47 DF,  p-value: 0.3162
```



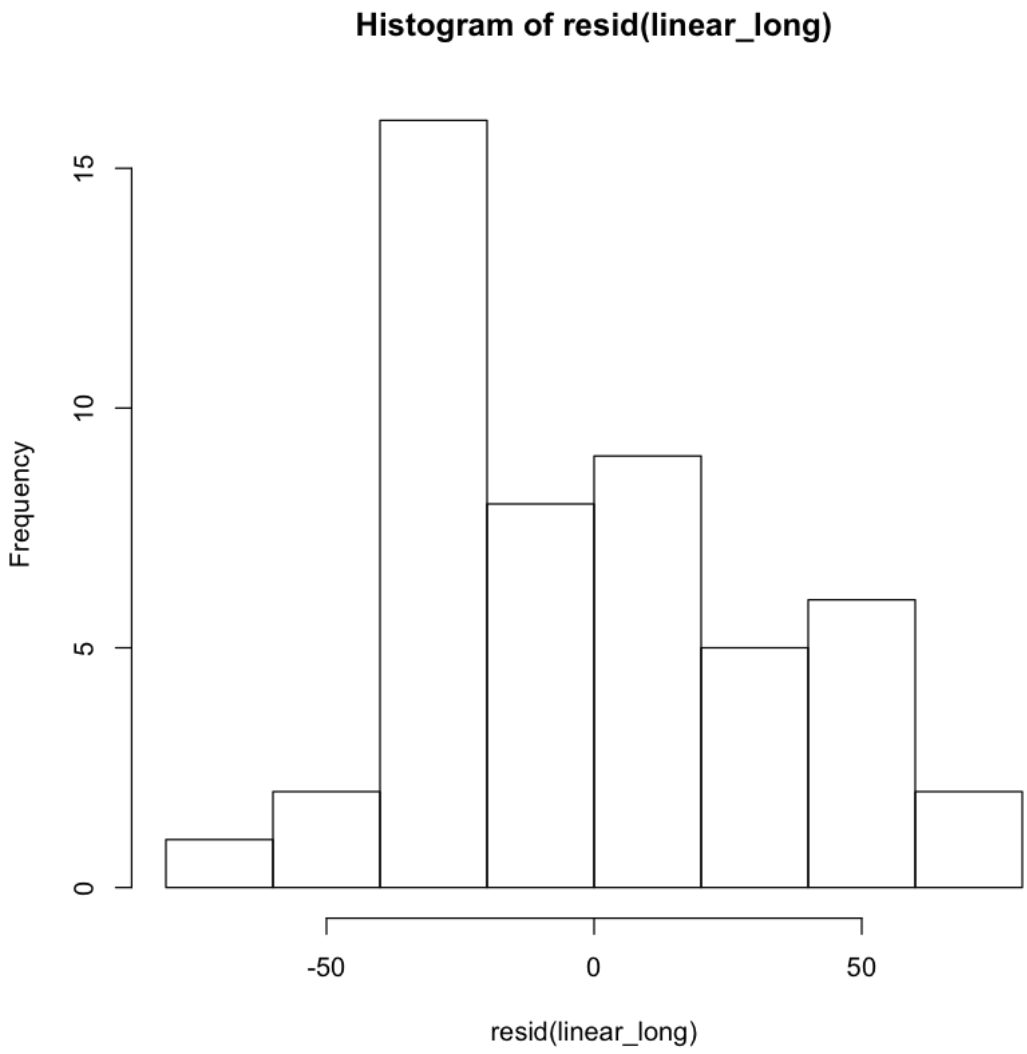
```
In [9]: plot(x = MortalityRate$Long, y = resid(linear_long), col = 'blue', xlab="Longitude", yl
        abline(h = 0, col = 'red')
```




```
In [22]: qqnorm(resid(linear_long))
         qqline(resid(linear_long))
```




In [27]: `hist(resid(linear_long))`




The residual plot for latitude shows even variance about the mean, while the points for longitude show uneven variance, suggesting latitude is a better candidate for a linear model. The qq plot for latitude closely follows a 45 degree upward sloping line suggesting that the error is approximately normal. The qq plot for longitude is skewed suggesting the error does not follow a normal distribution.