Salmon Population Dynamics in Washington State

https://github.com/fekelley4/KelleyLiddle_ENV872_EDA_FinalProject.git

Fiona Kelley & David Liddle

2023-12-12

Contents

Rationale

Salmon species native to Washington waters are critical components in maintaining economic growth, environmental health, and cultural identities. Salmon fisheries annually generate 23,000 jobs and millions for the State economy from harvest, recreation, and tourism. As a keystone species, salmon support a diverse ecosystem, sustaining 138 wildlife species such as orcas. Additionally, the unique migratory patterns of salmon species facilitates the transport and exchange of nutrients between freshwater and saltwater environments (Washington State Recreation & Conservation Office, 2022).

Salmon species are subject to various anthropogenic and natural stressors due to its exposure to different environments. Within Washington, 14 salmon population groups are listed as threatened or endangered under the Endangered Species Act. Increased water temperatures, as a result of climate change, habitat degradation, infrastructure development, overfishing, and pollution are placing immense pressures on salmon populations; these stressors are limiting breeding grounds, disrupting migration, altering genetic diversity, and depleting fish stocks (Washington State Recreation & Conservation Office, 2022). Utilizing a dataset monitoring the status of salmon fisheries from the Washington Department of Fish and Wildlife, salmon abundance and spawning is examined in relation to temperature and seasonal components. The objective of this analysis is to evaluate changes in counts of live salmon and redds (nests) for Chinook, Coho, and Sockeye salmon, species possessing population groups listed as threatened or endangered under the ESA.

Research Questions

Question 1: How has the observation of live salmon and new redd counts changed over the past 50 years? Do certain species exhibit more significant changes in live counts or redd counts compared to others?

Question 2: To what extent do factors such as water temperature and seasonality influence salmon stock counts, and is there variation in these influences between different salmon species?

Dataset Information

The selected dataset is a Spawning Ground Survey (SGS) database obtained from the Washington Department of Fish and Wildlife. The enormous survey, possessing 59 columns, contains observations of salmon status from 1930 to present day across Washington State streams. The data is displayed via several organizational methods, including latitude and longitude coordinates, watershed code, stream name, and mileage upstream. The data was collected irregularly, meaning that not every location was sampled on a regular basis, nor was every single component sampled in each event. Thus, the data had to be wrangled carefully, in order to draw meaningful trends from the data.

Data Wrangling

Exploratory Analysis

Analysis

```
# Create AIC to determine what set of variables is the live count predictor
TPAIC2 <- lm(data = salmon_data, NewReddCount ~ WaterTempFarenheit + Month)
step(TPAIC2)
```

```
## Start:  AIC=42557.87
## NewReddCount ~ WaterTempFarenheit + Month
##
##                       Df Sum of Sq      RSS   AIC
## - Month                1         2 14552871 42556
## <none>                             14552869 42558
## - WaterTempFarenheit   1    391644 14944512 42699
##
## Step:  AIC=42555.87
## NewReddCount ~ WaterTempFarenheit
##
##                       Df Sum of Sq      RSS   AIC
## <none>                             14552871 42556
## - WaterTempFarenheit   1    395009 14947880 42698

##
## Call:
## lm(formula = NewReddCount ~ WaterTempFarenheit, data = salmon_data)
##
## Coefficients:
##        (Intercept)  WaterTempFarenheit
##            -3.4525              0.4531
```

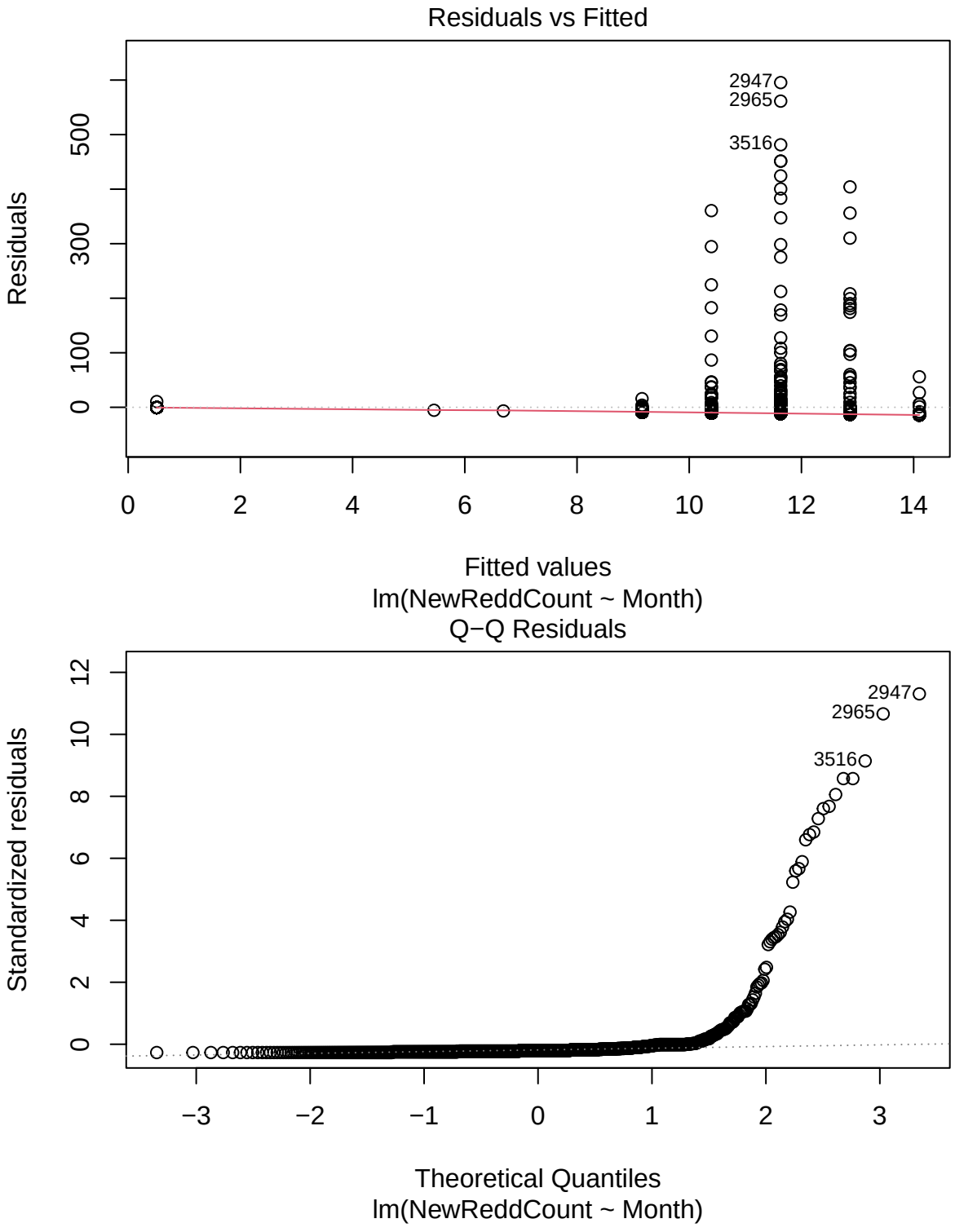
```
# Define the multiple regression model
model2 <- lm(data = salmon_data, NewReddCount ~ WaterTempFarenheit + Month)

# Fit the model
summary(model)
```

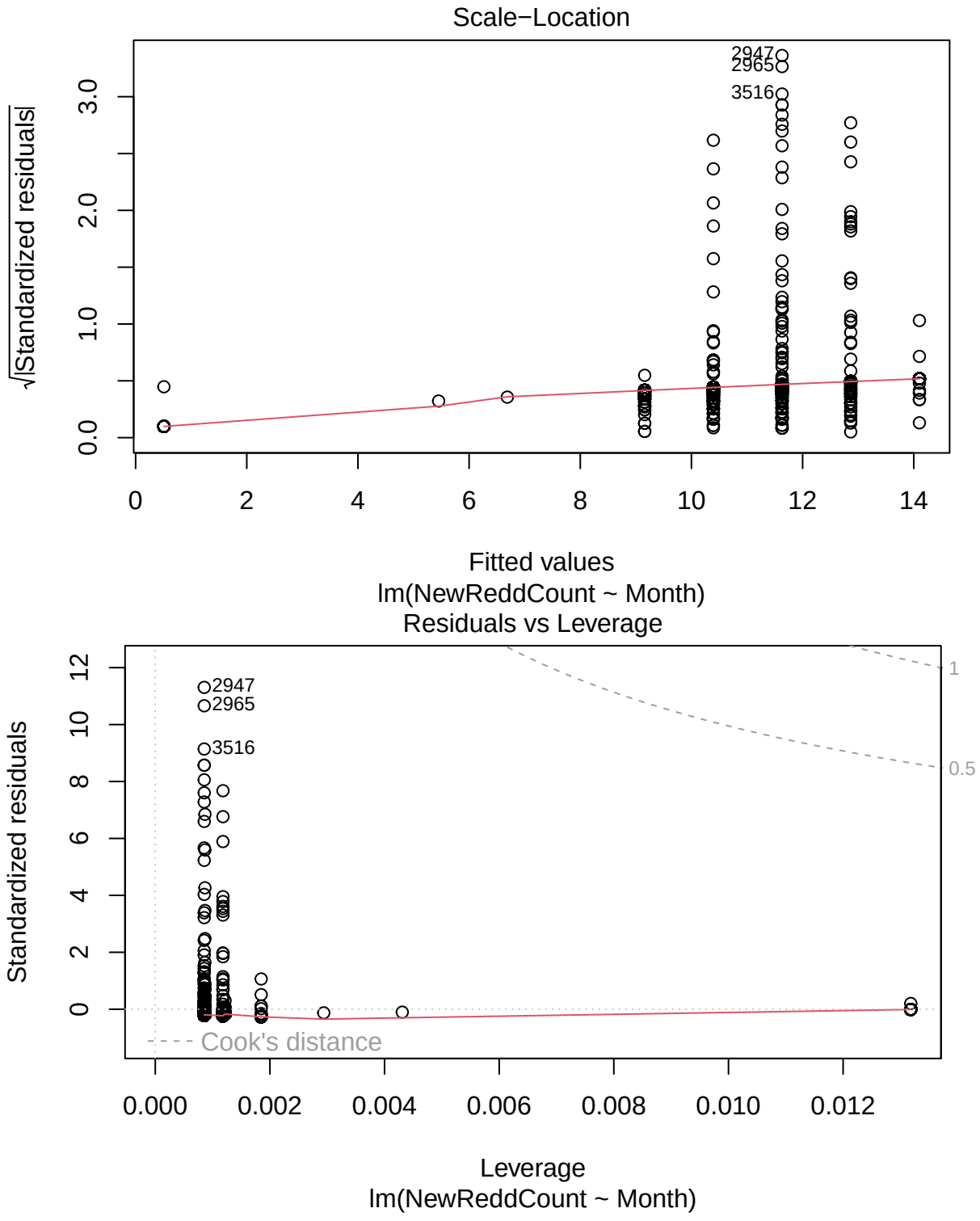
```
##
## Call:
## lm(formula = LiveTotal ~ WaterTempFarenheit + Month, data = salmon_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -72.3  -41.0  -11.3   -5.8 4121.2
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.718831   7.760889   0.093    0.926
## WaterTempFarenheit  0.924890   0.088290  10.476   <2e-16 ***
## Month              -0.002387   0.728266  -0.003    0.997
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 147.8 on 6843 degrees of freedom
##   (606795 observations deleted due to missingness)
## Multiple R-squared:  0.01579,    Adjusted R-squared:  0.0155
## F-statistic: 54.89 on 2 and 6843 DF,  p-value: < 2.2e-16
```

```
#Because Sockeye New Redd was the only factor with not enough evidence to reject the null, this was bec
#Run Sockeye New Redd and Live Count Single Regression
#Redd Count
NewReddCountssock_month<- lm(data = salmon_sock, NewReddCount ~ Month)
summary(NewReddCountssock_month)
```

```
##
## Call:
## lm(formula = NewReddCount ~ Month, data = salmon_sock)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -14.10 -11.63 -10.39  -7.75 595.37
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.7270     6.7152  -0.108    0.914
## Month         1.2357     0.6862   1.801    0.072 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 52.68 on 1221 degrees of freedom
##   (10143 observations deleted due to missingness)
## Multiple R-squared:  0.002649,   Adjusted R-squared:  0.001832
## F-statistic: 3.243 on 1 and 1221 DF,  p-value: 0.07199
```

```
plot(NewReddCountssock_month)
```


Residuals vs Fitted
Residuals
Fitted values
lm(NewReddCount ~ Month)
2947
2965
3516
Q−Q Residuals
Standardized residuals
Theoretical Quantiles
lm(NewReddCount ~ Month)

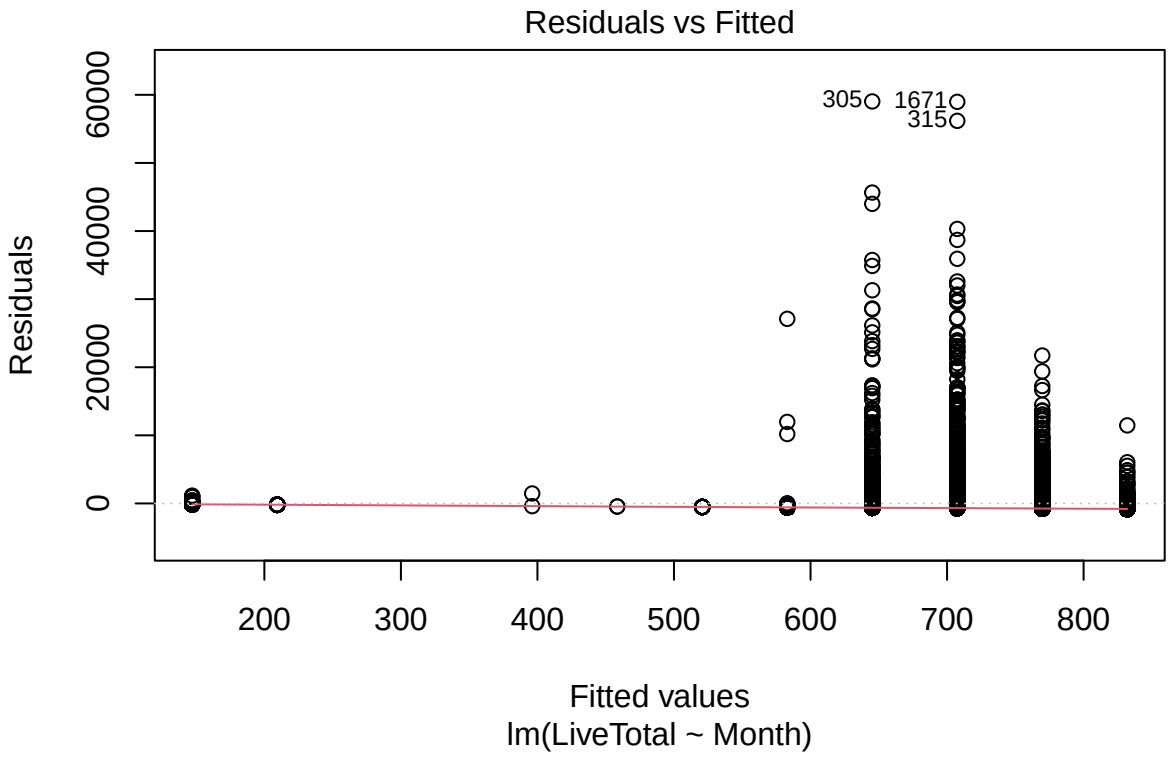


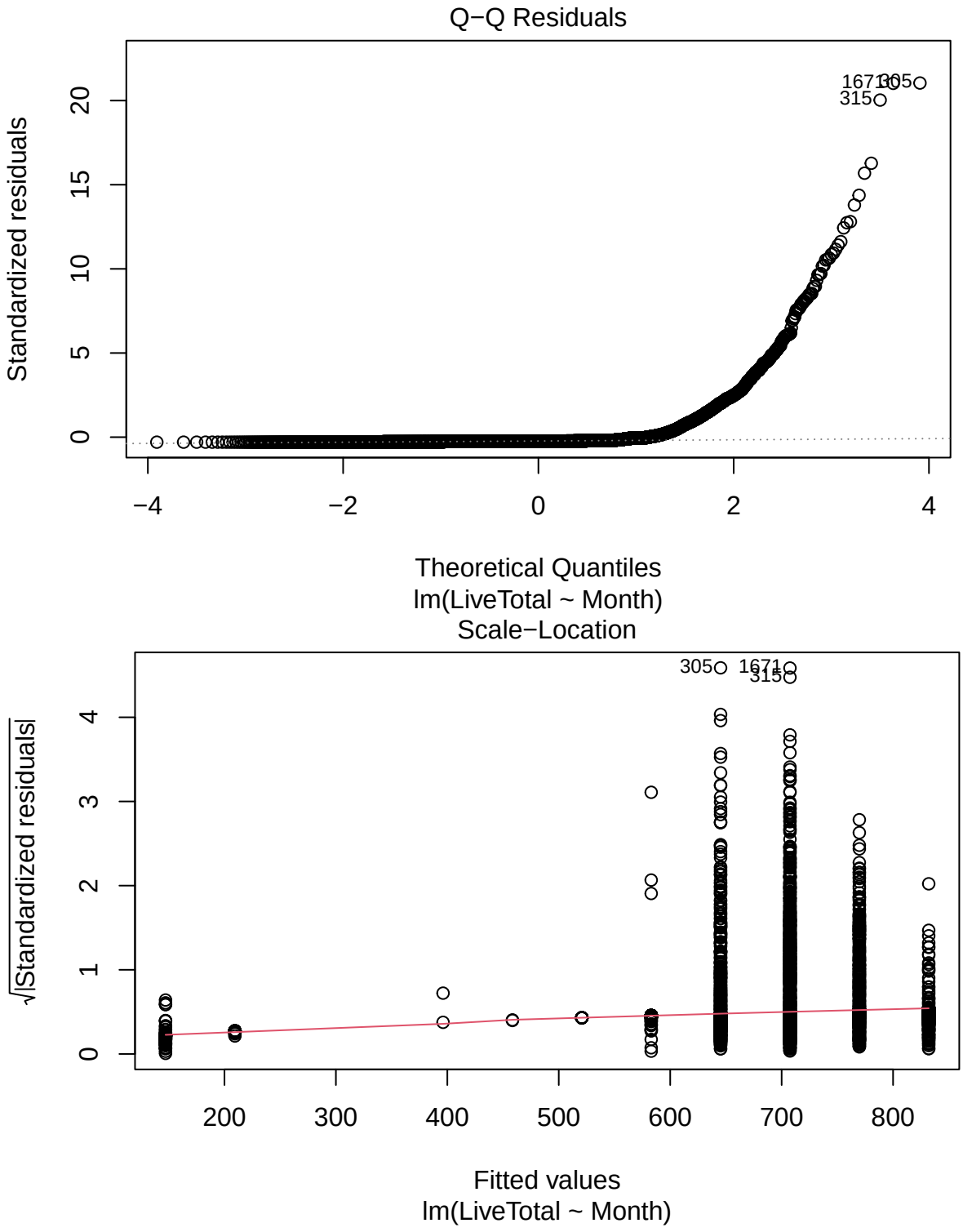

```
#Live Count
LiveCountssock_month<- lm(data = salmon_sock, LiveTotal ~ Month)
summary(LiveCountssock_month)
```

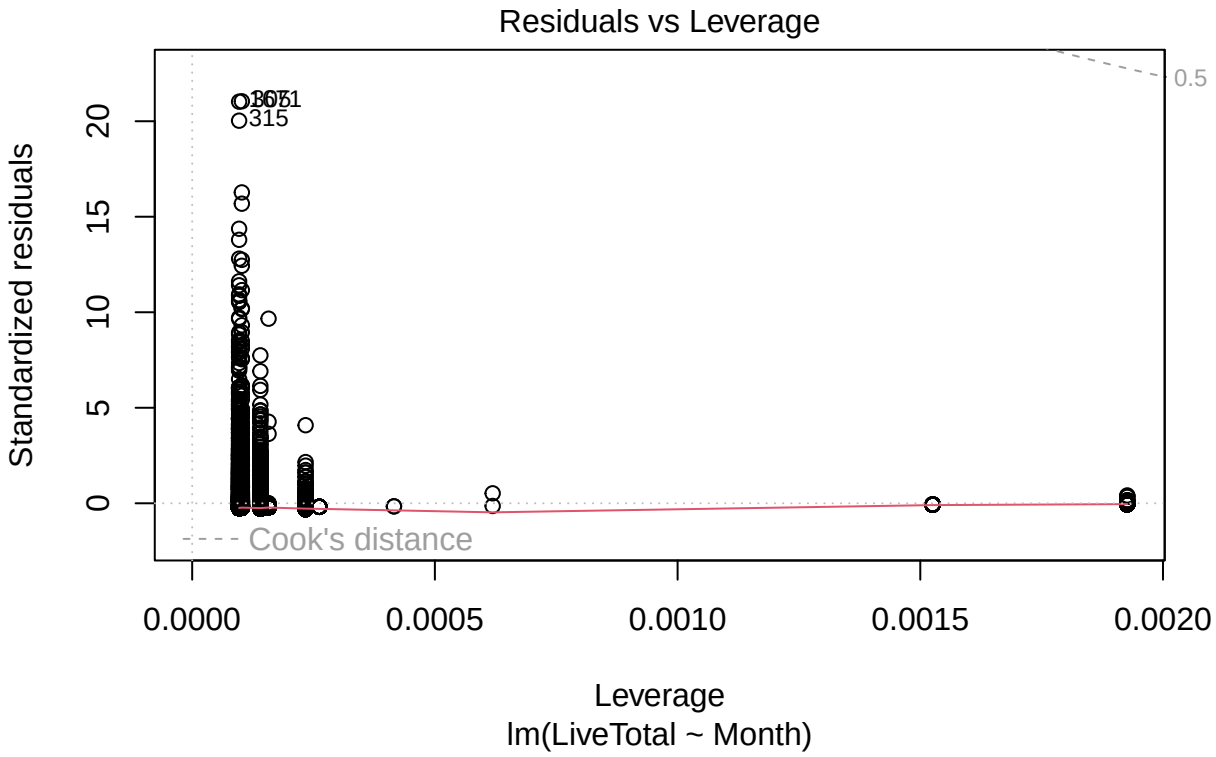
```
##
## Call:
```

```
## lm(formula = LiveTotal ~ Month, data = salmon_sock)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##   -832   -707   -645   -573  59027
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    84.86     136.76   0.620    0.535
## Month          62.26      13.94   4.465 8.09e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2805 on 10749 degrees of freedom
##   (615 observations deleted due to missingness)
## Multiple R-squared:  0.001851,   Adjusted R-squared:  0.001758
## F-statistic: 19.94 on 1 and 10749 DF,  p-value: 8.086e-06
```

```
plot(LiveCountssock_month)
```




Q−Q Residuals
Standardized residuals
Theoretical Quantiles
lm(LiveTotal ~ Month)
Scale−Location
√|Standardized residuals|
Fitted values
lm(LiveTotal ~ Month)



Summary & Conclusions

Works Cited