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.

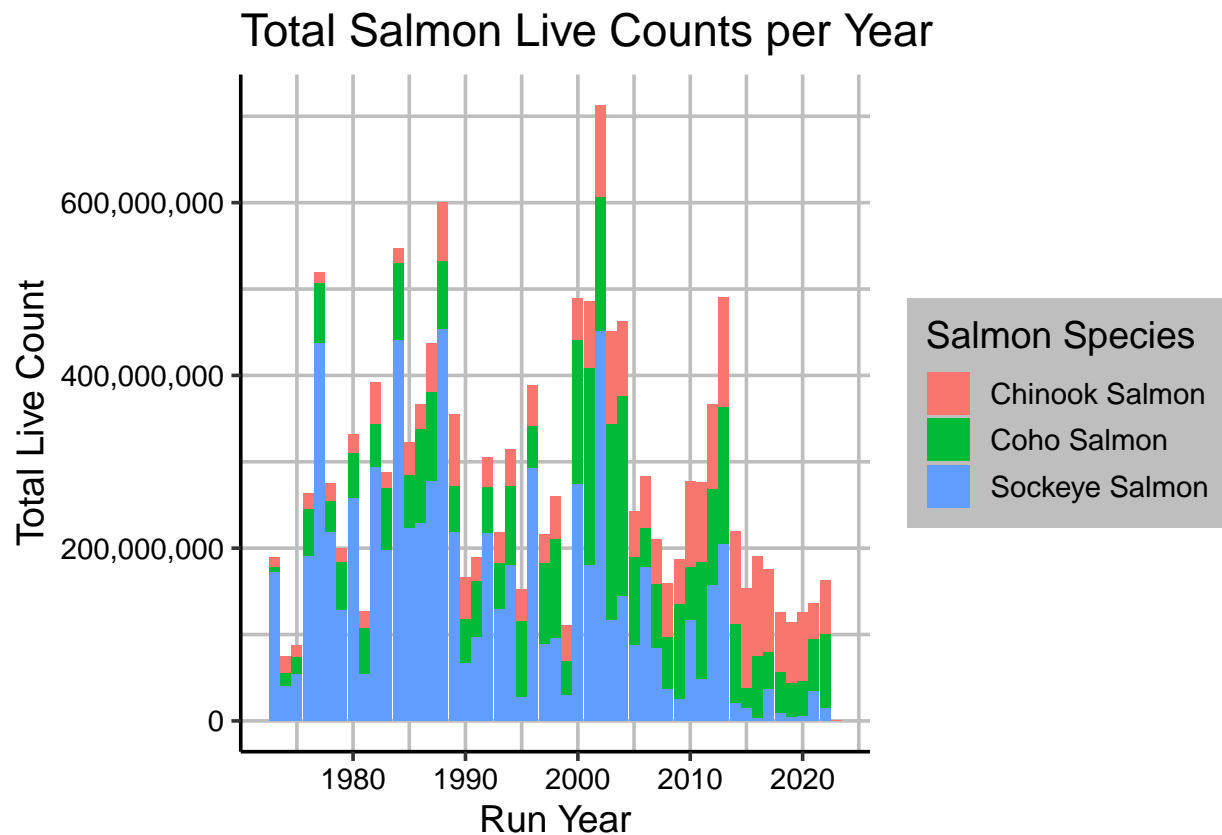## Data Wrangling

Exploratory Analysis

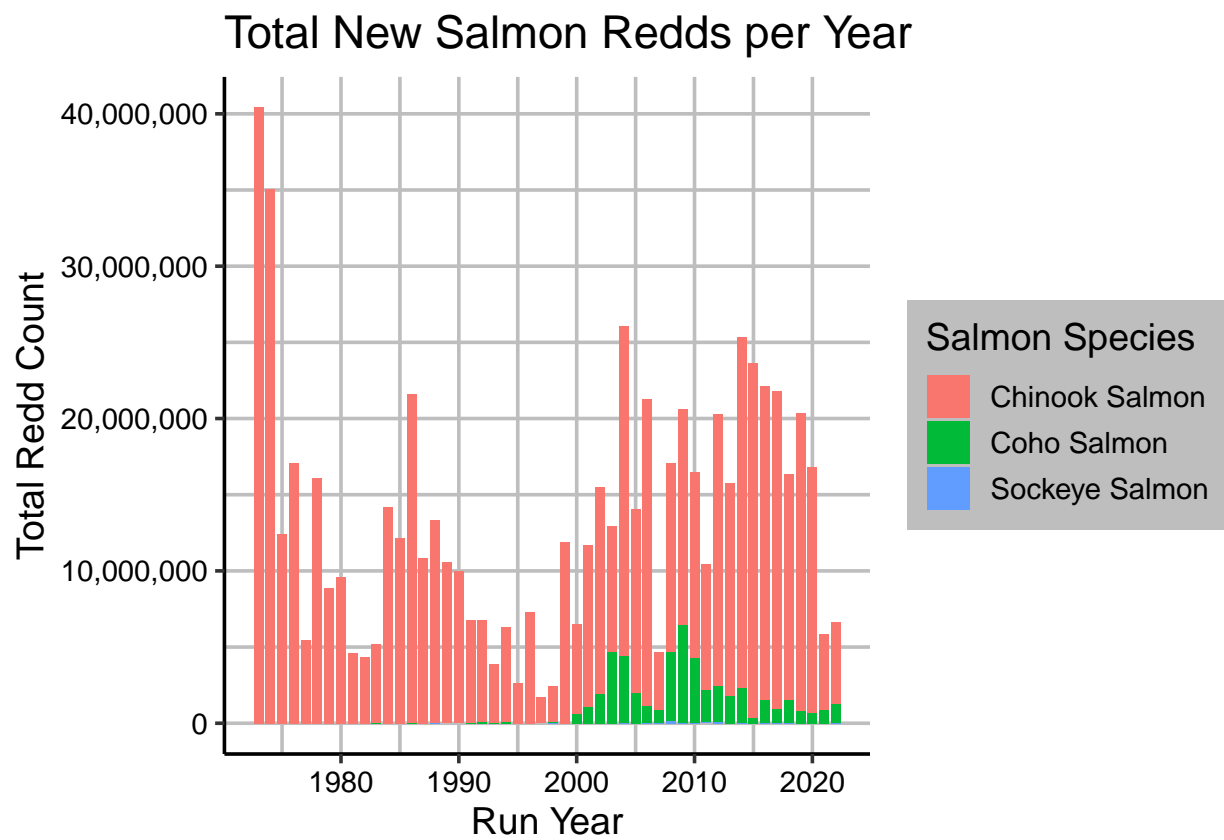Figure 1: Total live salmon observed from 1973 to present day.

Analysis

Total New Salmon Redds per Year

Figure 2: Total salmon redds observed from 1973 to present day
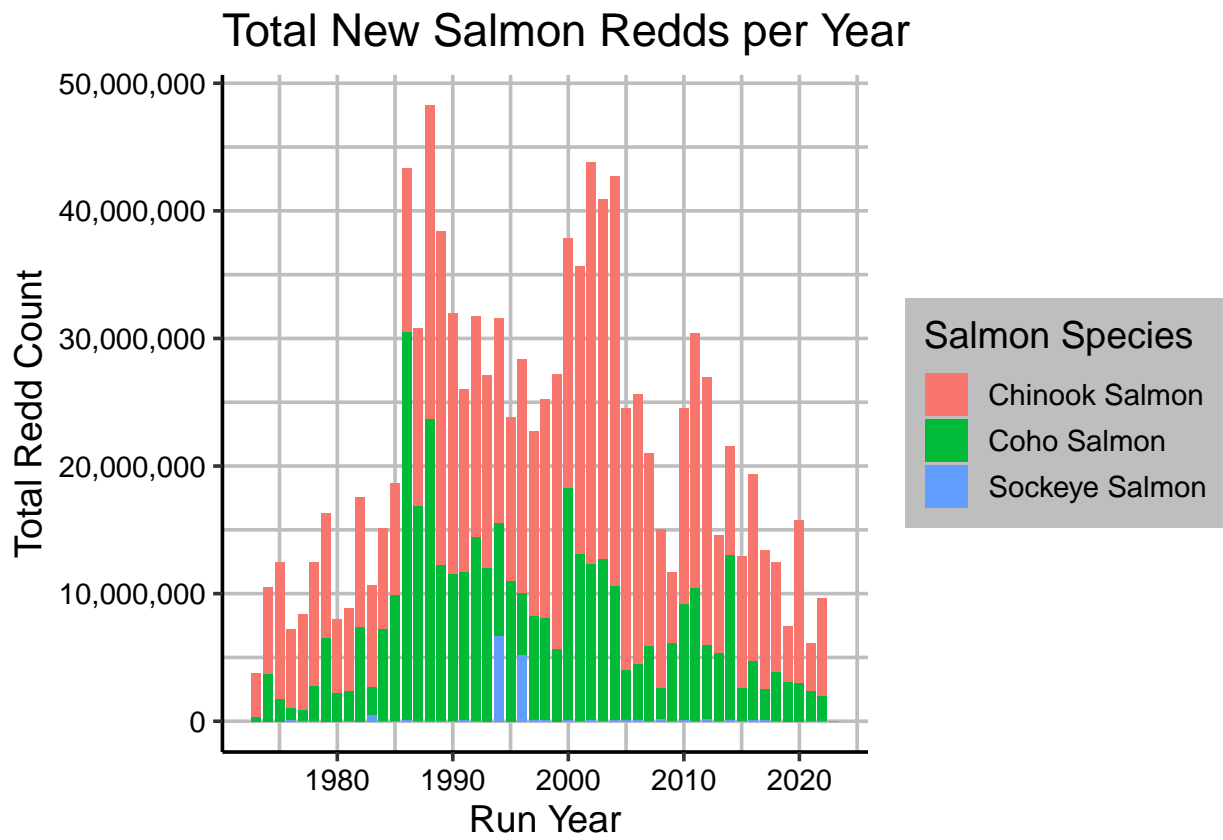
Figure 3: Total new salmon redds observed from 1973 to present day

```r
# 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
```

```r
# 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
## -19.83 -14.83  -6.99  -1.37 587.17
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.193      6.936   0.028  0.97781
## Month1970-01-06  -0.193     52.821  -0.004  0.99709
## Month1970-01-07  -0.193     52.821  -0.004  0.99709
## Month1970-01-09   1.566      8.573   0.183  0.85506
## Month1970-01-10   6.801      7.534   0.903  0.36685
## Month1970-01-11  19.634      7.435   2.641  0.00838 **
## Month1970-01-12  14.639      7.761   1.886  0.05951 .
## Month1970-01-13   1.173      8.309   0.141  0.88771
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 52.36 on 1215 degrees of freedom
##   (10143 observations deleted due to missingness)
## Multiple R-squared:  0.01956,    Adjusted R-squared:  0.01391
## F-statistic: 3.462 on 7 and 1215 DF,  p-value: 0.001119
```
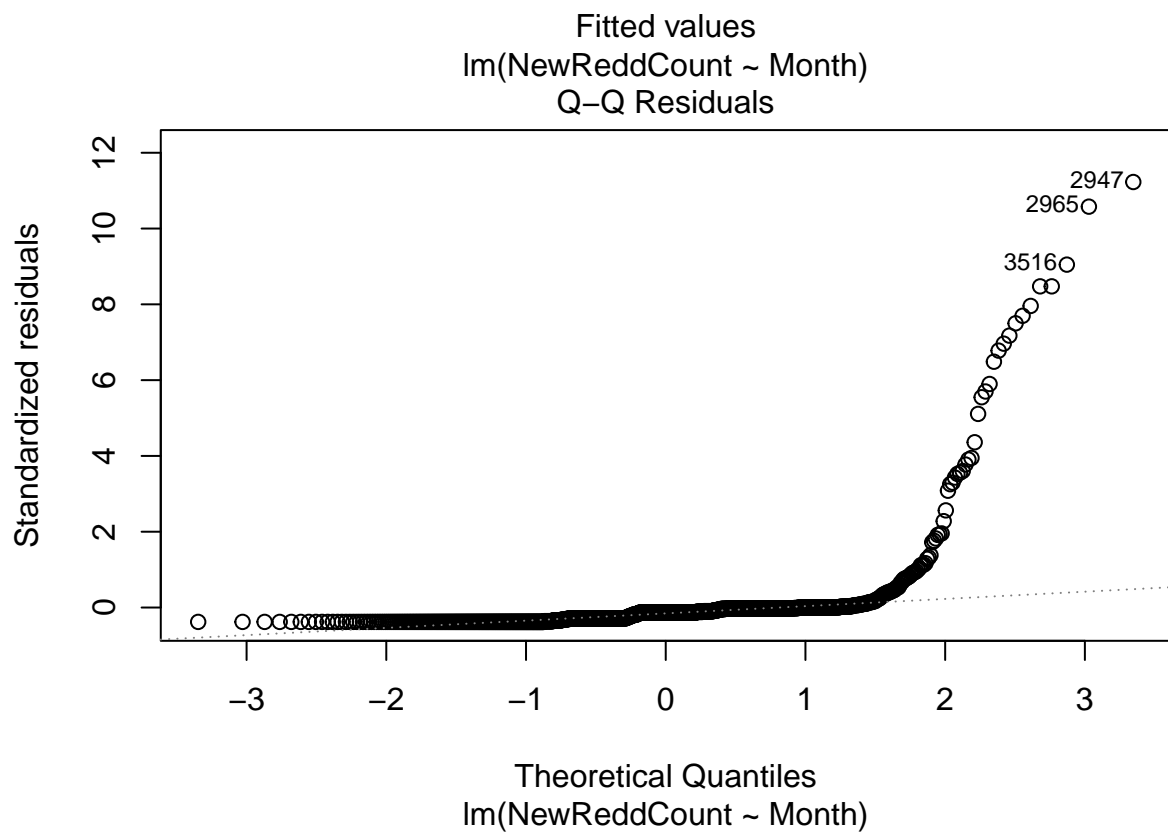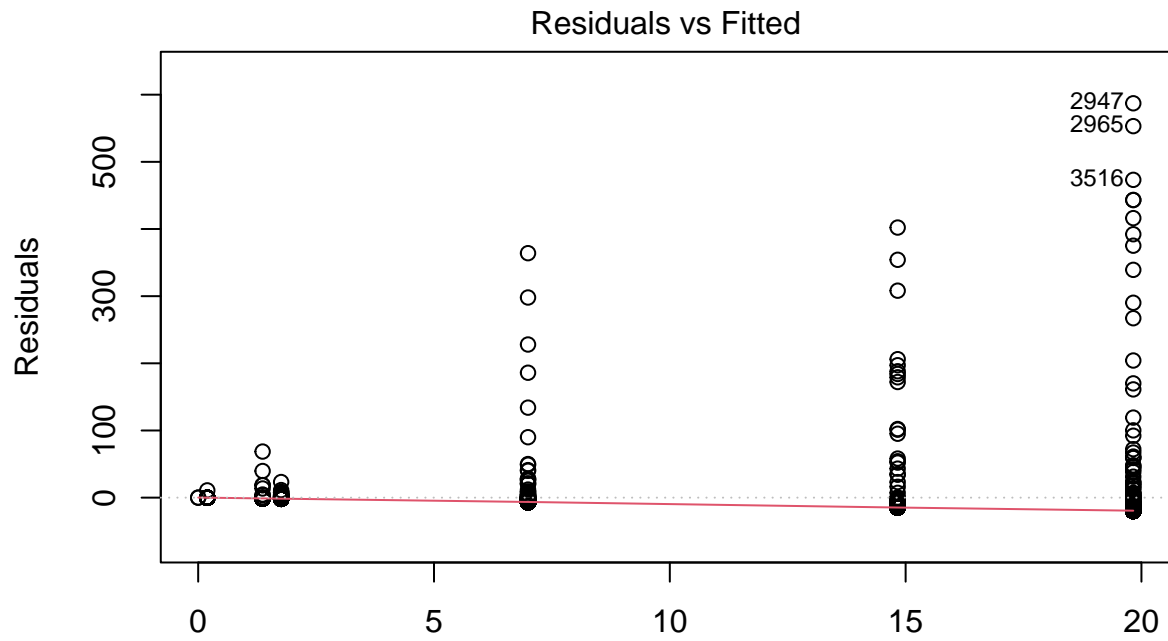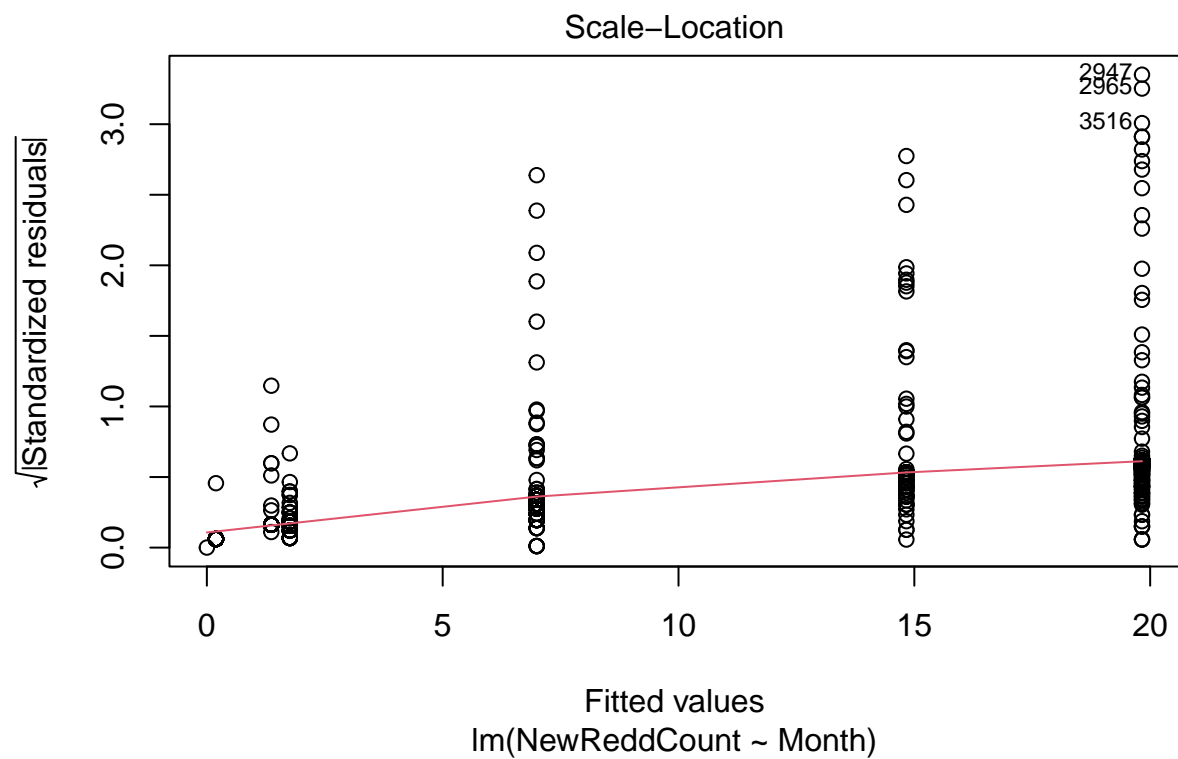
```
plot(NewReddCountssock_month)
```

```
## Warning: not plotting observations with leverage one:
##   744
```

Residuals vs Fitted

Fitted values
lm(NewReddCount ~ Month)

Q–Q Residuals

Theoretical Quantiles
lm(NewReddCount ~ Month)
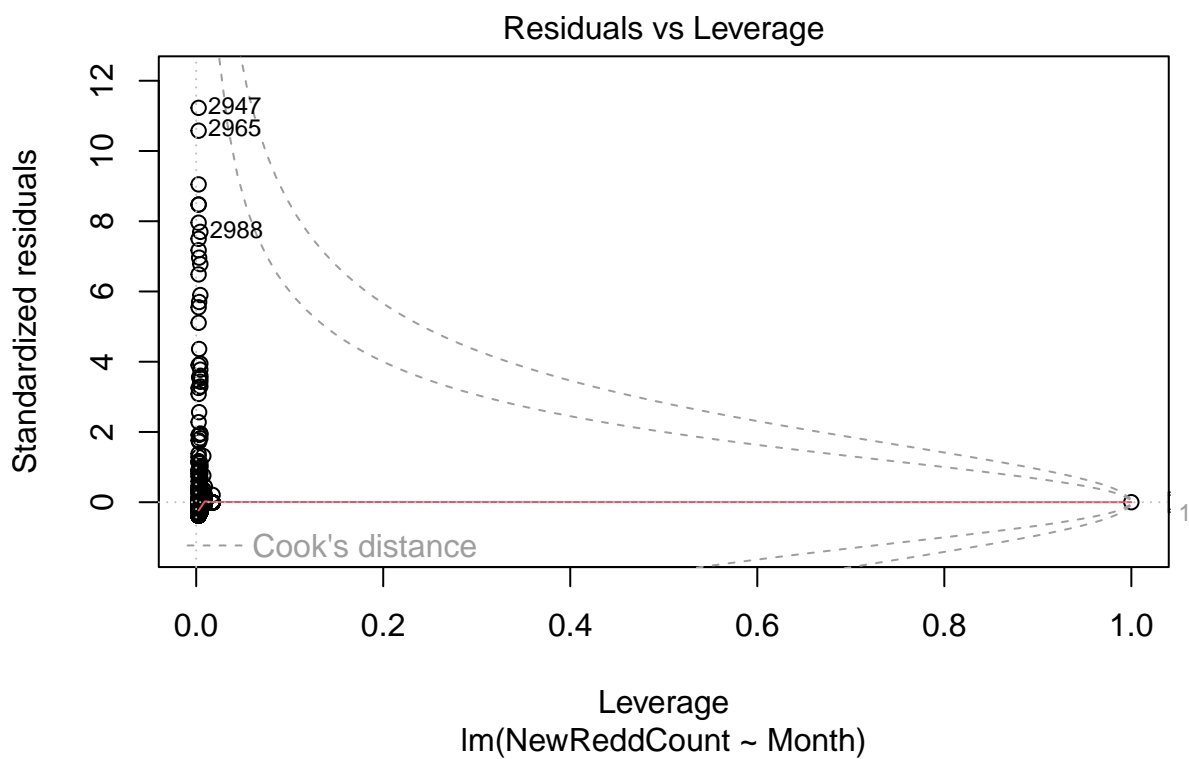
Scale−Location

Fitted values
lm(NewReddCount ~ Month)

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```

Residuals vs Leverage
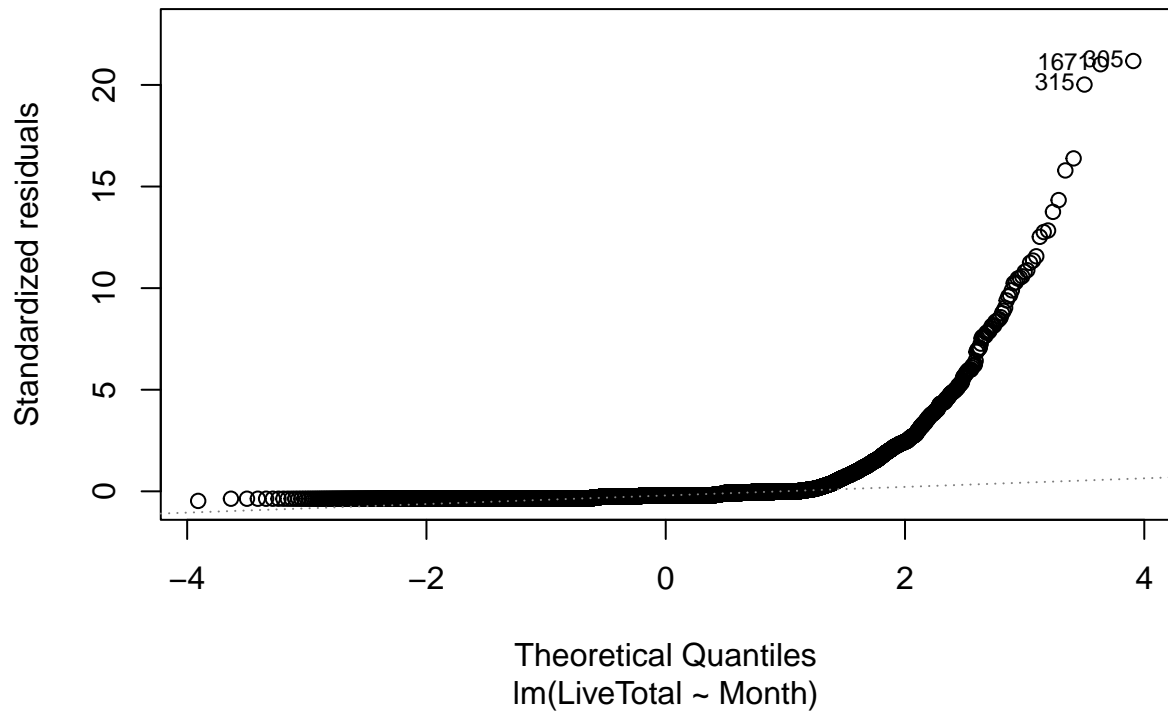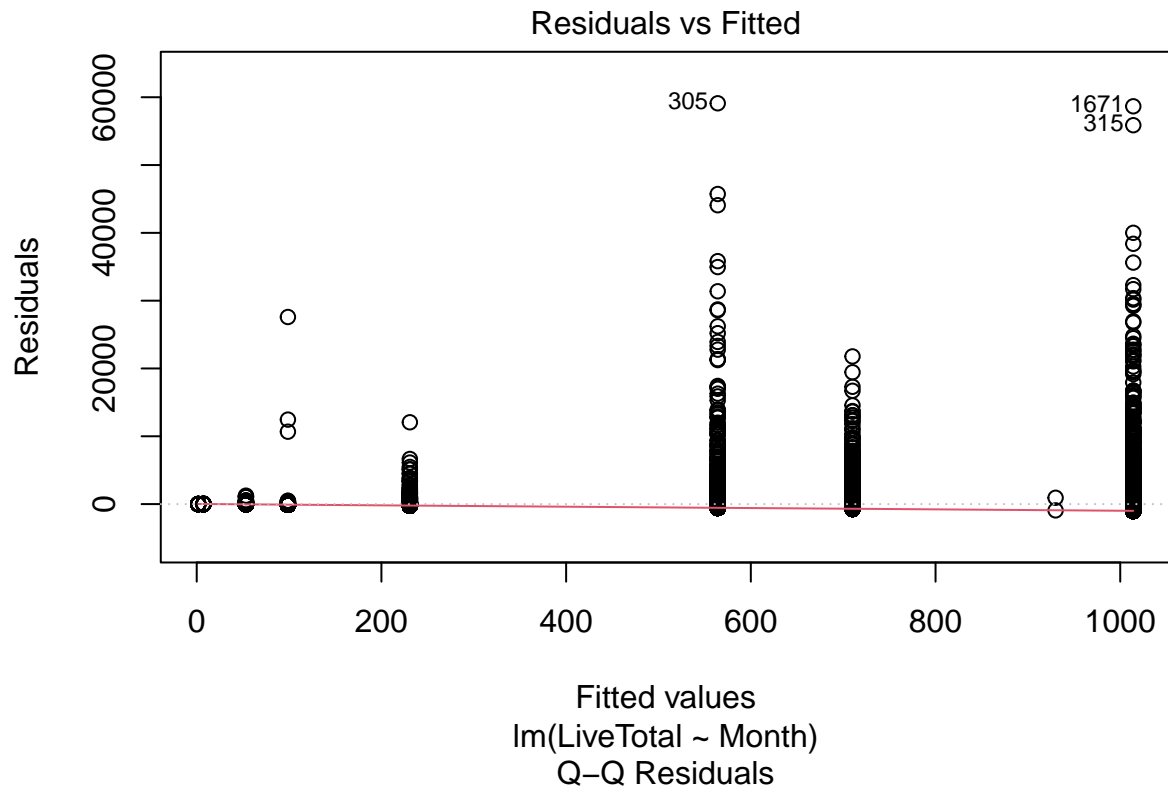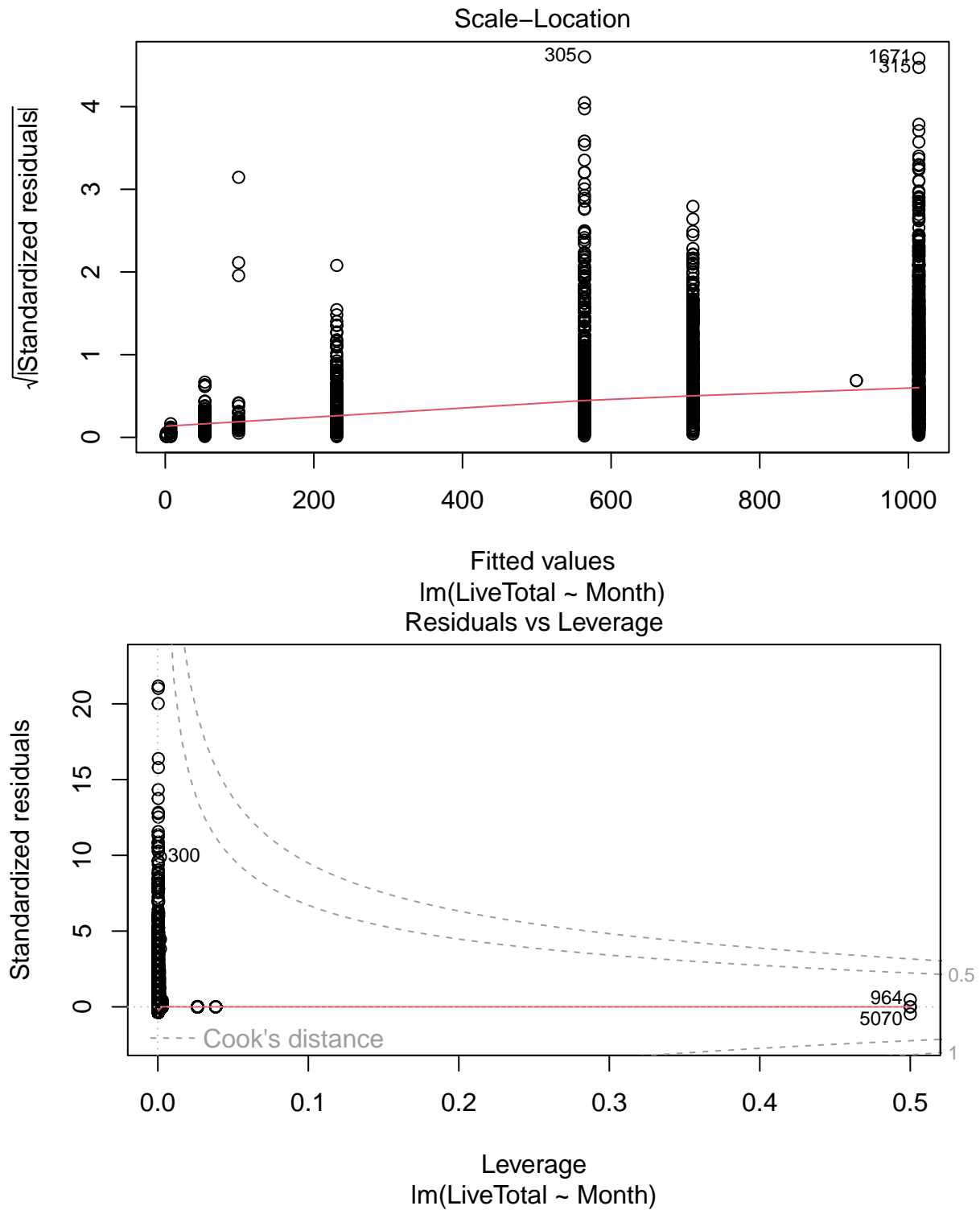

Leverage
lm(NewReddCount ~ Month)

```
#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
##  -1014   -967   -564   -173  59108
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)        53.30     148.78   0.358  0.72019
## Month1970-01-03   -45.98     476.64  -0.096  0.92315
## Month1970-01-06   876.70    1979.43   0.443  0.65784
## Month1970-01-07   -45.80    1979.43  -0.023  0.98154
## Month1970-01-08   -51.72     567.30  -0.091  0.92736
## Month1970-01-09    45.45     188.00   0.242  0.80896
## Month1970-01-10   510.87     157.25   3.249  0.00116 **
## Month1970-01-11   960.84     155.11   6.194 6.06e-10 ***
## Month1970-01-12   656.79     162.97   4.030 5.61e-05 ***
## Month1970-01-13   177.51     174.92   1.015  0.31022
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2791 on 10741 degrees of freedom
##   (615 observations deleted due to missingness)
## Multiple R-squared:  0.01236,    Adjusted R-squared:  0.01154
## F-statistic: 14.94 on 9 and 10741 DF,  p-value: < 2.2e-16
```

```
plot(LiveCountssock_month)
```

Residuals vs Fitted

305

1671
315

Residuals

Fitted values
lm(LiveTotal ~ Month)

Q−Q Residuals

1671 305
315

Standardized residuals

Theoretical Quantiles
lm(LiveTotal ~ Month)

10

Scale–Location

√|Standardized residuals|

Fitted values
lm(LiveTotal ~ Month)

Residuals vs Leverage

Standardized residuals

Cook's distance

Leverage
lm(LiveTotal ~ Month)

Summary & Conclusions

Works Cited