

Assignment 9

Jaclyn Narleski

2025-03-20

Setup

```
# Load necessary packages  
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --  
## v dplyr      1.1.4      v readr      2.1.5  
## v forcats    1.0.0      v stringr   1.5.1  
## v ggplot2    3.5.1      v tibble    3.2.1  
## v lubridate  1.9.4      v tidyr     1.3.1  
## v purrr      1.0.2  
## -- Conflicts ----- tidyverse_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag()     masks stats::lag()  
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(here)
```

```
## here() starts at /Users/jaclynrose/Desktop/ENV710/week9/j-narleski
```

```
library(naniar)  
library(DHARMa)
```

```
## This is DHARMa 0.4.7. For overview type '?DHARMa'. For recent changes, type news(package = 'DHARMa')
```

```
library(gtsummary)  
library(broom)
```

```
# Read csvs  
lizards <- read.csv("jrn_lizard.csv")  
lizard_npp <- read.csv("jrn_lizard_npp.csv")
```

Logistic Regression

Do snout-to-vent length, sex, and vegetation zone at time of capture significantly predict if a lizard tail is recorded as whole?

Data Tidying

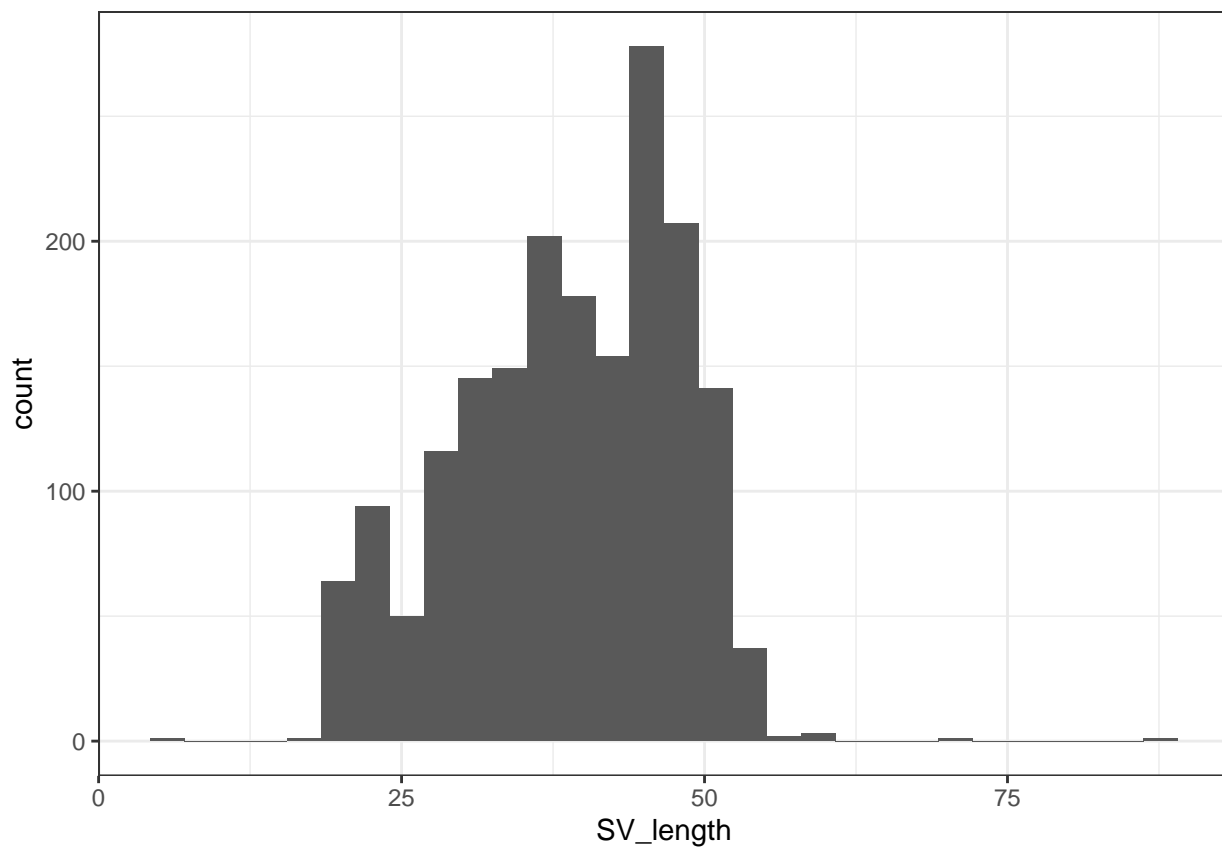
```
# Filter dataset
lizards <- lizards %>%
  filter(spp == "UTST")
```

Examine Data

```
# Create exploratory figures for each variable
(hist1 <- ggplot(data = lizards, aes(x = SV_length)) +
  geom_histogram() +
  theme_bw())
```

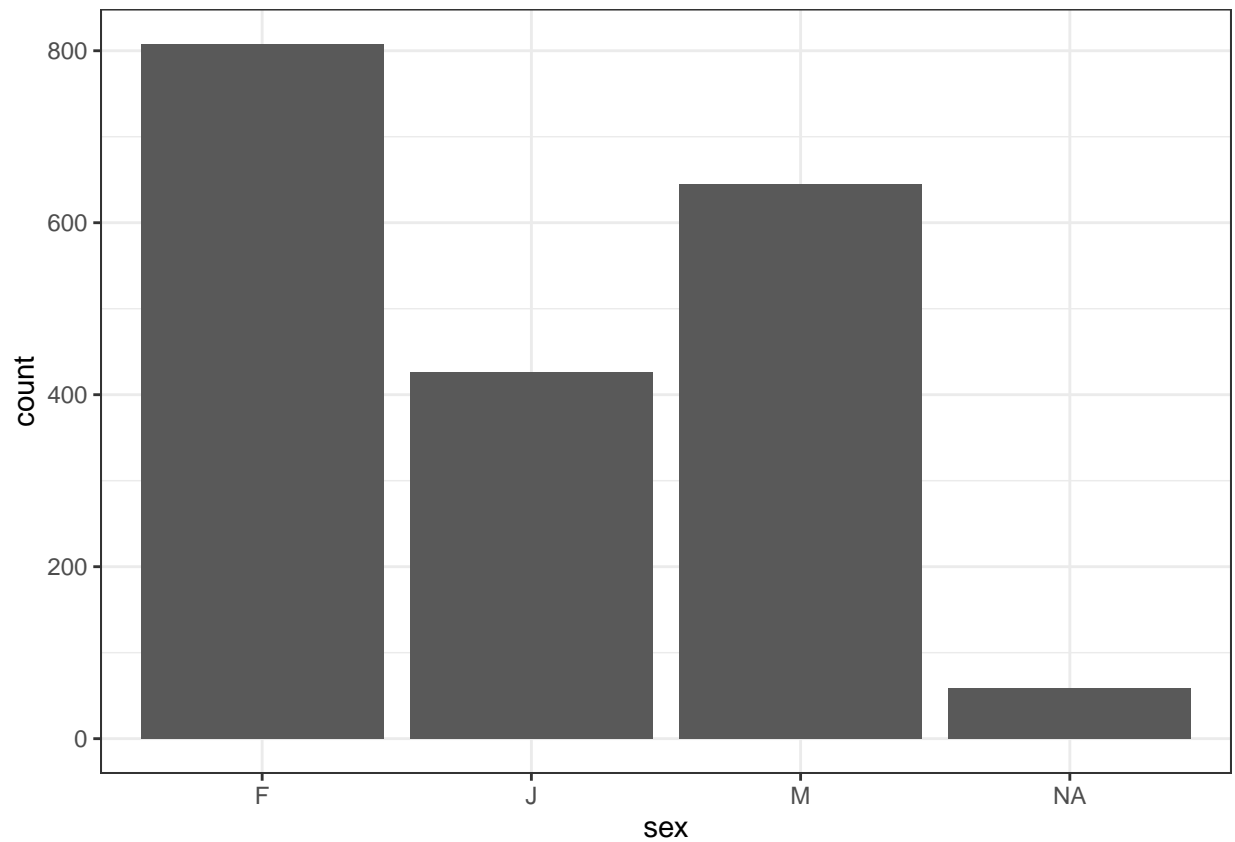
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

```
## Warning: Removed 114 rows containing non-finite outside the scale range
## ('stat_bin()').
```



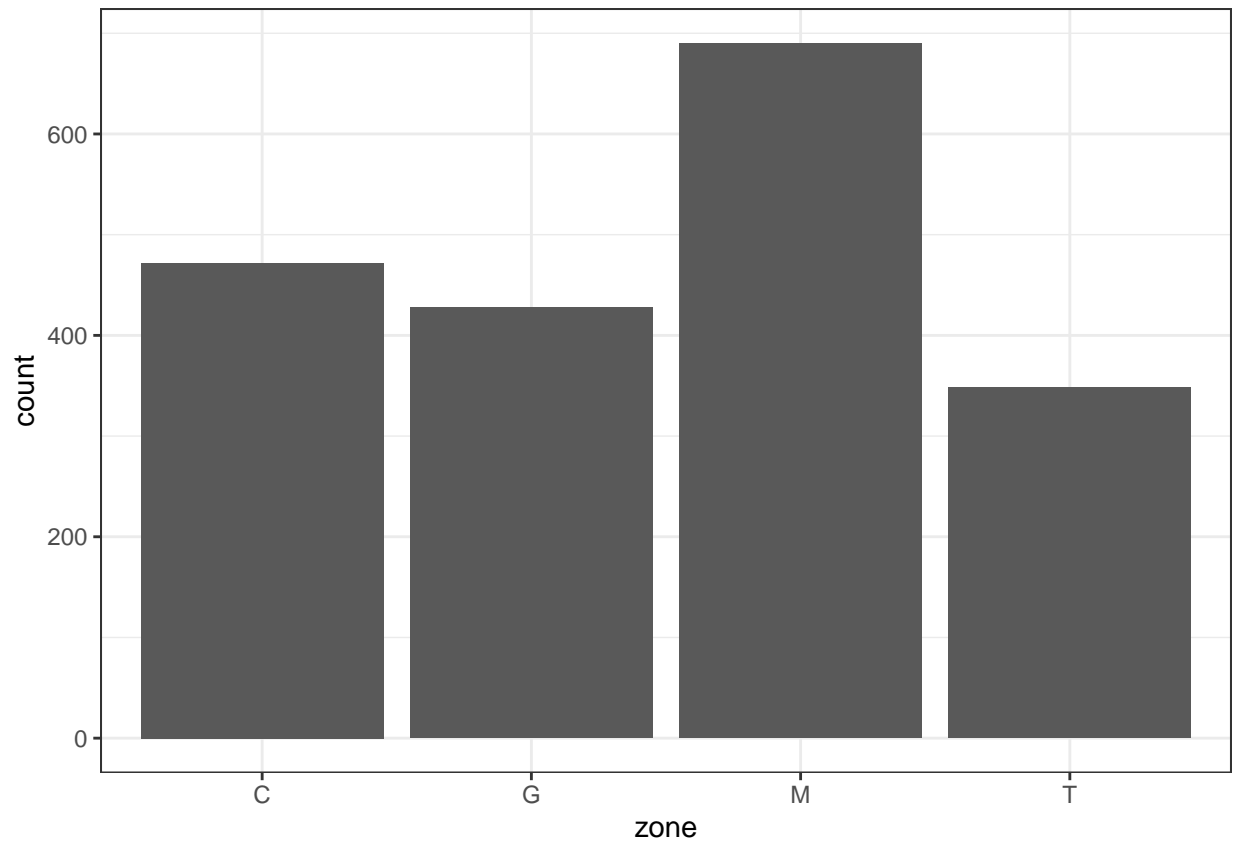
```
(hist2 <- ggplot(data = lizards, aes(x = sex)) +
  geom_histogram(stat = "count", na.rm = TRUE) +
  theme_bw())
```

```
## Warning in geom_histogram(stat = "count", na.rm = TRUE): Ignoring unknown
## parameters: 'binwidth', 'bins', and 'pad'
```



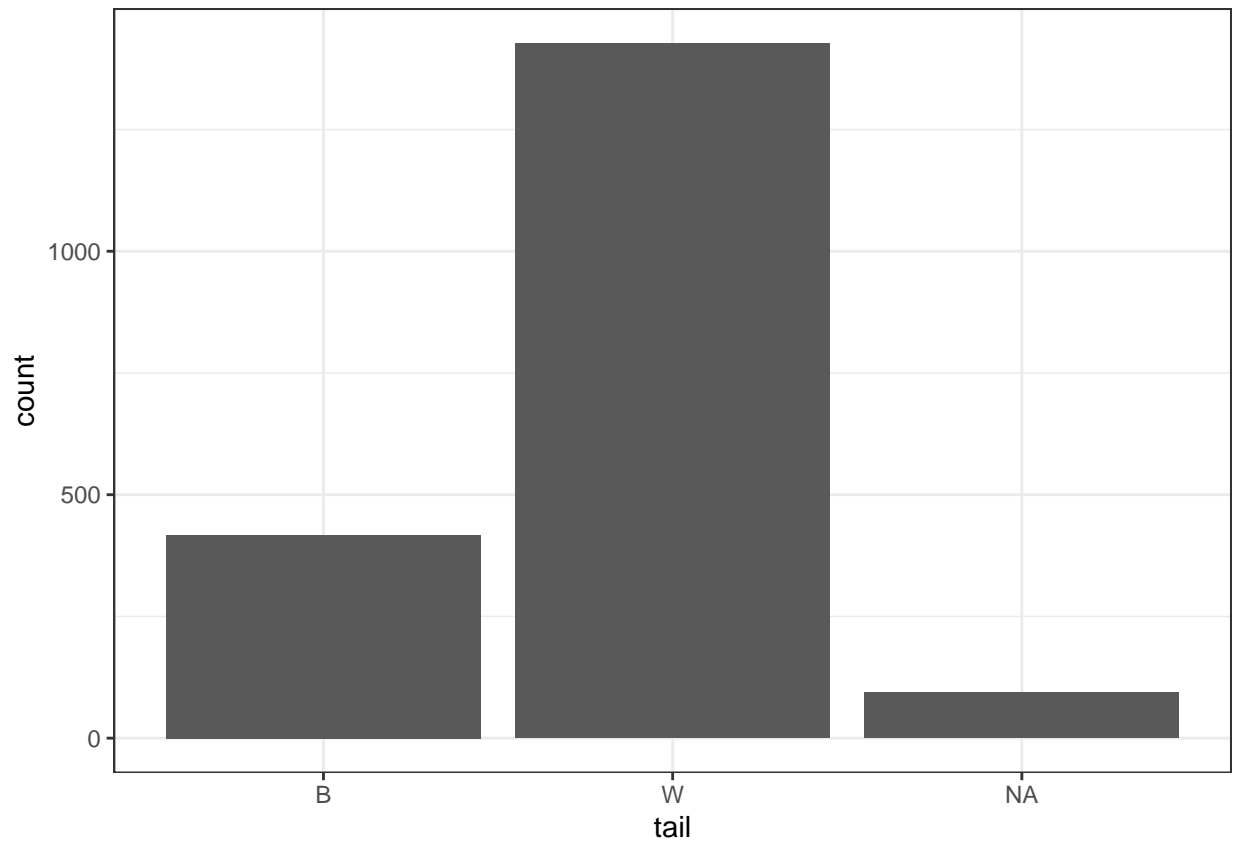
```
(hist3 <- ggplot(data = lizards, aes(x = zone)) +  
  geom_histogram(stat = "count", na.rm = TRUE) +  
  theme_bw())
```

```
## Warning in geom_histogram(stat = "count", na.rm = TRUE): Ignoring unknown
## parameters: 'binwidth', 'bins', and 'pad'
```



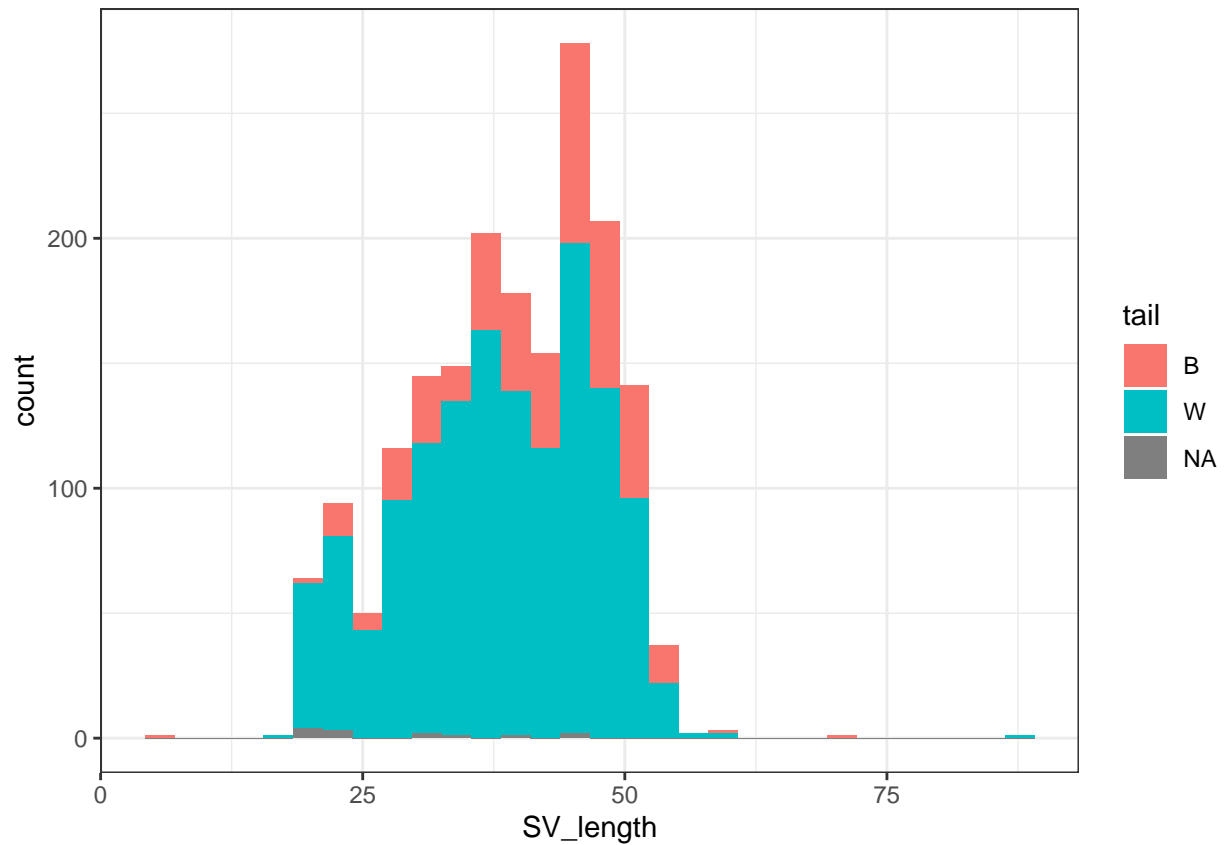
```
(hist4 <- ggplot(data = lizards, aes(x = tail)) +  
  geom_histogram(stat = "count", na.rm = TRUE) +  
  theme_bw())
```

```
## Warning in geom_histogram(stat = "count", na.rm = TRUE): Ignoring unknown  
## parameters: 'binwidth', 'bins', and 'pad'
```



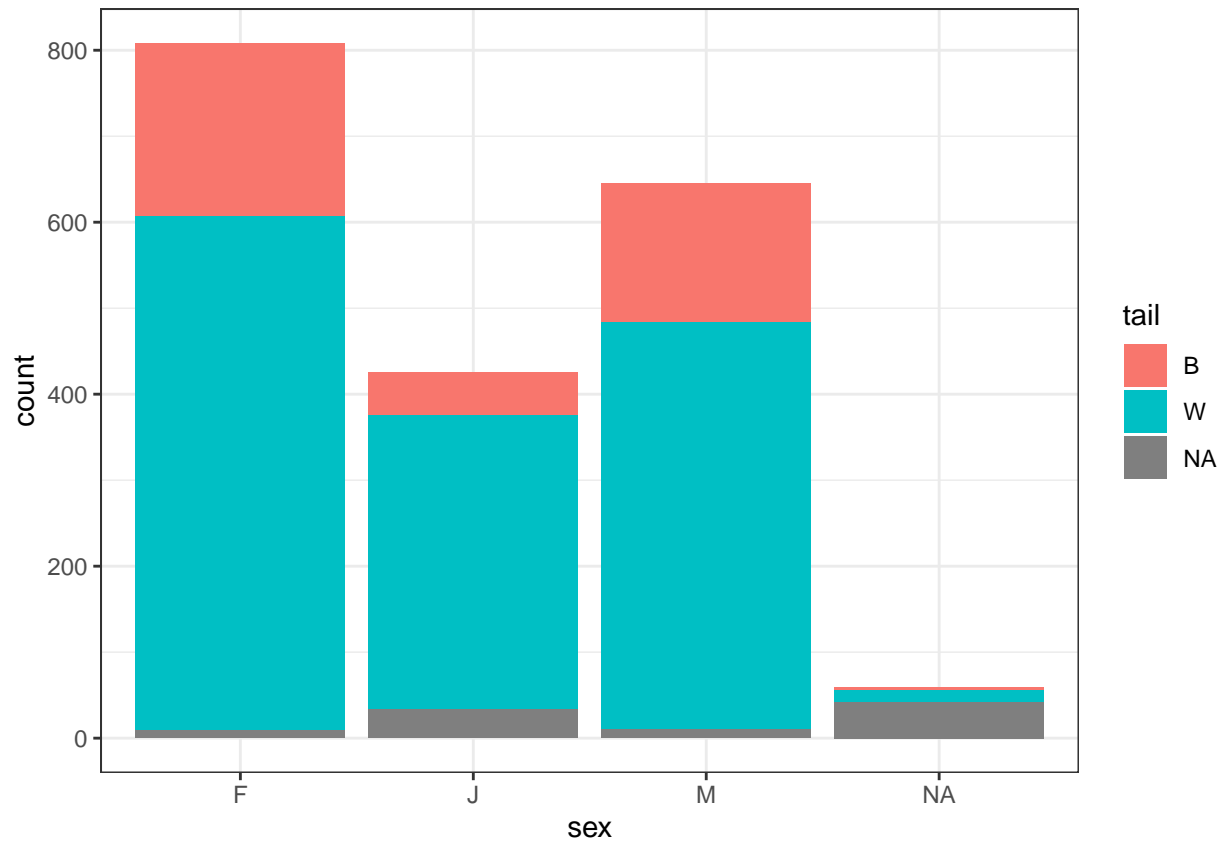
```
# Create exploratory figures between variables
(hist5 <- ggplot(data = lizards, aes(x = SV_length,
                                     fill = tail)) +
  geom_histogram(na.rm = TRUE) +
  theme_bw())
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



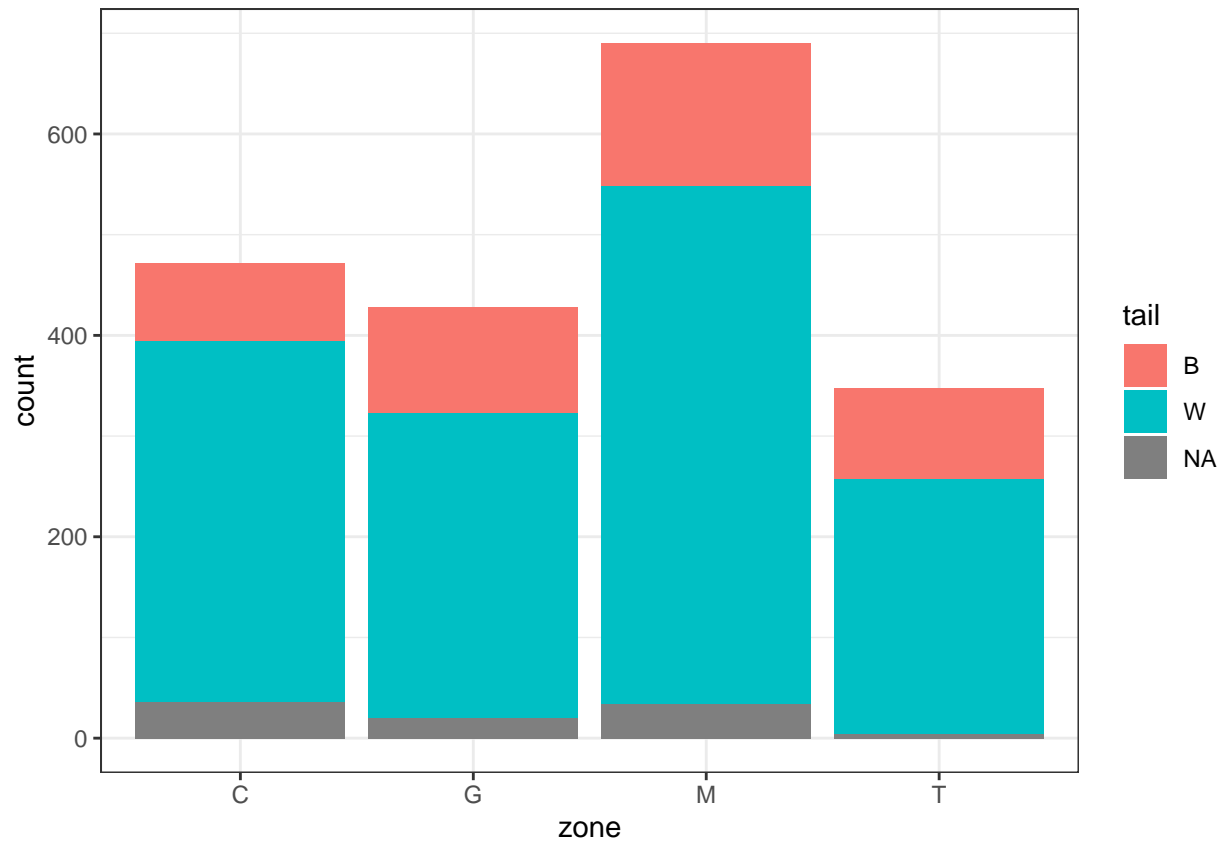
```
(hist6 <- ggplot(data = lizards, aes(x = sex,
                                     fill = tail)) +
  geom_histogram(stat = "count", na.rm = TRUE) +
  theme_bw())
```

```
## Warning in geom_histogram(stat = "count", na.rm = TRUE): Ignoring unknown
## parameters: 'binwidth', 'bins', and 'pad'
```



```
(hist7 <- ggplot(data = lizards, aes(x = zone,
                                     fill = tail)) +
  geom_histogram(stat = "count", na.rm = TRUE) +
  theme_bw())
```

```
## Warning in geom_histogram(stat = "count", na.rm = TRUE): Ignoring unknown
## parameters: 'binwidth', 'bins', and 'pad'
```



Fit Regression Model

```
# Turn character vectors into factors
lizards$sex <- factor(lizards$sex,
                      levels = c("F", "J", "M", "NA"))

lizards$zone <- factor(lizards$zone,
                      levels = c("C", "G", "M", "T"))

lizards$tail <- factor(lizards$tail,
                      levels = c("B", "W", "NA"))

# Fit regression model
tail_mod <- glm(tail ~ SV_length + zone + sex,
                data = lizards,
                family = "binomial")

# Summary of regression model
summary(tail_mod)
```

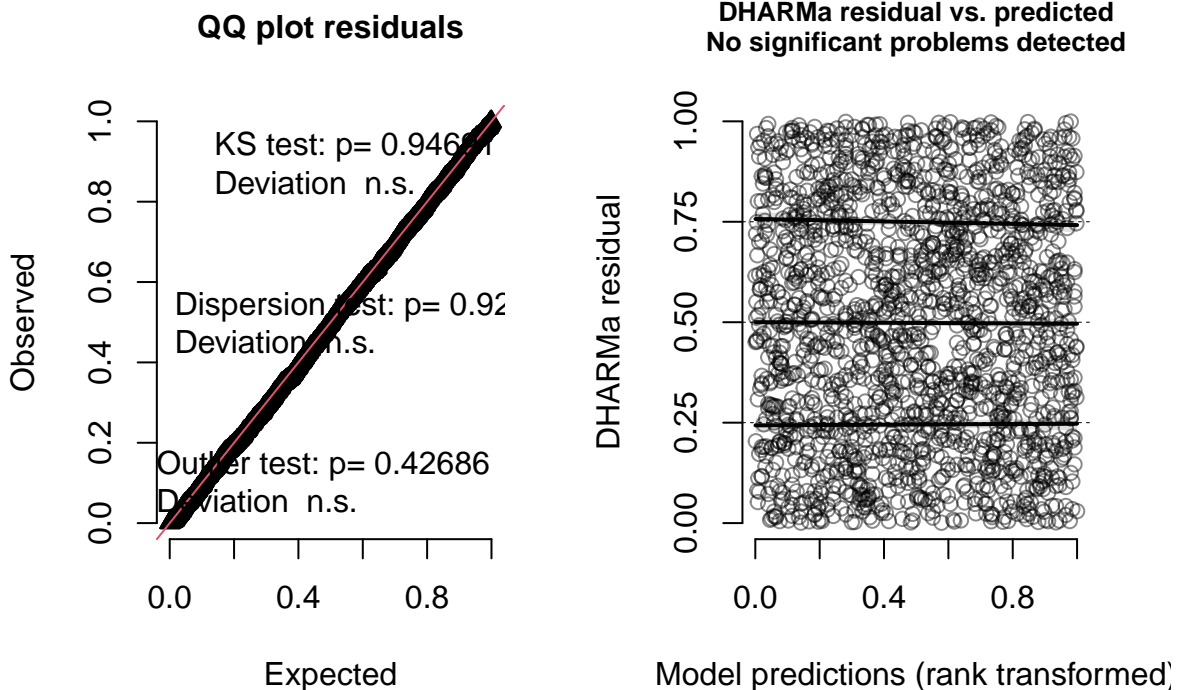
```
##
## Call:
## glm(formula = tail ~ SV_length + zone + sex, family = "binomial",
##      data = lizards)
```



```
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.290069   0.407027   8.083 6.31e-16 ***
## SV_length   -0.045230   0.008877  -5.095 3.49e-07 ***
## zoneG       -0.565369   0.173745  -3.254 0.00114 **
## zoneM       -0.301313   0.162032  -1.860 0.06294 .
## zoneT       -0.404349   0.179938  -2.247 0.02463 *
## sexJ         0.203871   0.206033   0.990 0.32241
## sexM         0.064915   0.126708   0.512 0.60842
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 1931.7  on 1799  degrees of freedom
## Residual deviance: 1867.3  on 1793  degrees of freedom
## (138 observations deleted due to missingness)
## AIC: 1881.3
##
## Number of Fisher Scoring iterations: 4
```

```
# Simulate residuals in plot form
simulateResiduals(tail_mod) %>% plot()
```

DHARMa residual



```
## Create Prediction Model
```

```

# Create vector of snout-to-vent lengths
svlength_vector <- rep(seq(from = 0, to = 100), 4)

# Vector of sex categories
sex_vector <- c(rep("F", 101),
                rep("M", 101),
                rep("J", 101),
                rep(NA, 101))

# Vector of zone categories
zone_vector2 <- c("G", "C", "M", "T")
# Randomize
zone_vector_random <- sample(zone_vector2, size = 404, replace = TRUE)

# Generate data frame
data_pred <- data.frame(svlength_vector, sex_vector, zone_vector_random)
# Name columns
colnames(data_pred) <- c("SV_length", "sex", "zone")

# Generate prediction model
prediction <- predict(tail_mod,
                     newdata = data_pred,
                     type = "response",
                     se.fit = TRUE)

# Generate data frame of predictions
data_fig <- data.frame(data_pred,
                      prediction$fit,
                      prediction$se.fit)

# Name columns
colnames(data_fig) <- c("SV_length",
                      "Sex", "Zone", "probability", "se")

# Generate plot for with zone coloring
zoneplot <- ggplot(data_fig, aes(x = SV_length,
                                y = probability)) +
  geom_line(aes(color = Zone)) +
  geom_ribbon(aes(ymin = probability - se,
                ymax = probability + se,
                fill = Zone), alpha = 0.3) +
  labs(x = "Snout-to-Vent Length (cm)",
       y = "Probability of Whole Tail",
       color = "Zone",
       fill = "Zone",
       title = "Prediction Model for Likelihood of Whole Tail Based on Zone",
       caption = "This figure displays the probability of a lizard having a whole tail based on its snout-to-vent length and habitat zone.") +
  scale_color_manual(values = c("chartreuse3", "darkolivegreen3", "yellow3",
                                "palegreen3"),
                    labels = c("Black Grama Grassland",
                                "Creosotebush Shrubland",
                                "Mesquite Duneland", "Tarbush Shrubland")) +
  scale_fill_manual(values = c("greenyellow", "darkolivegreen", "yellow4",
                                "palegreen3"))

```

```

      "palegreen3"),
      labels = c("Black Grama Grassland",
                  "Creosotebush Shrubland",
                  "Mesquite Duneland", "Tarbush Shrubland")) +

theme_bw() +
theme(plot.caption = element_text(hjust = 0.5))

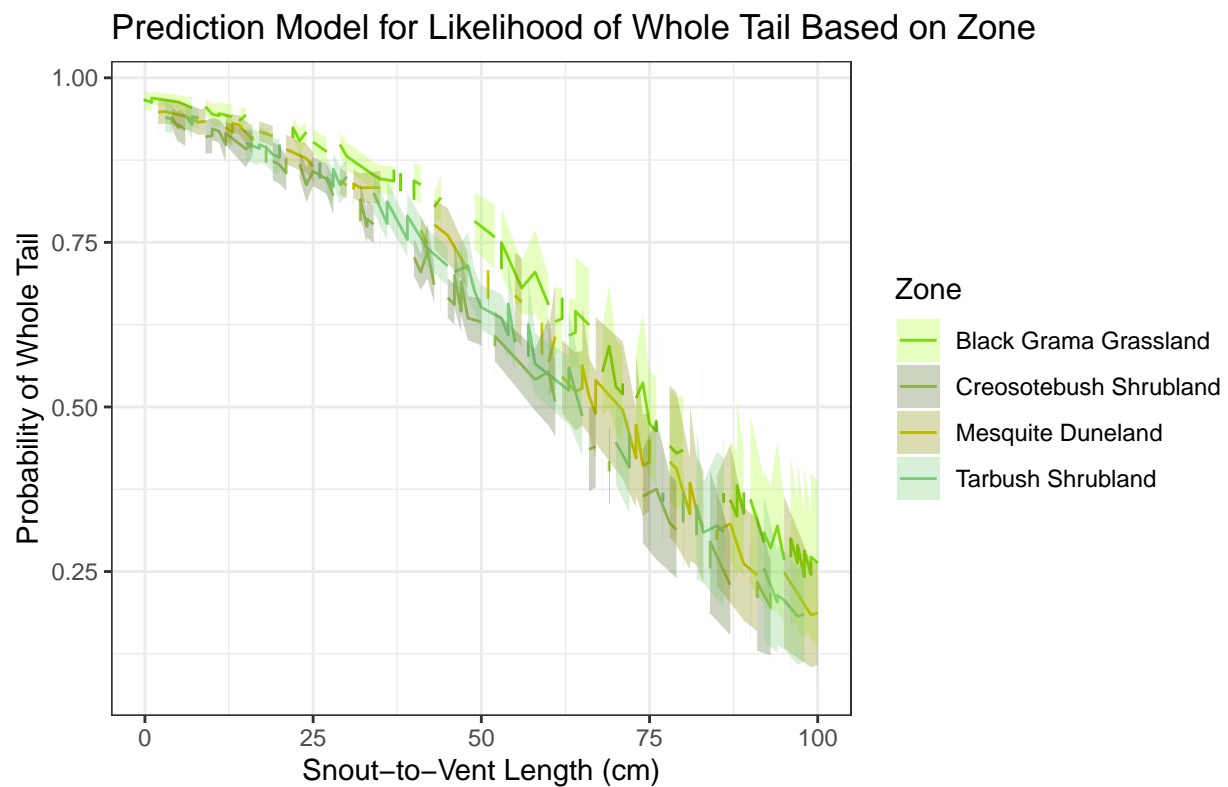
# View plot
zoneplot

```

```

## Warning: Removed 3 rows containing missing values or values outside the scale range
## ('geom_line()').

```



```

# Vector of sex categories
sex_vector2 <- c("F", "M", "J", NA)
# Randomize
sex_vector_random <- sample(sex_vector2, size = 404, replace = TRUE)

# Vector of zone categories
zone_vector <- c(rep("C", 101),
                  rep("G", 101),
                  rep("M", 101),
                  rep("T", 101))

```

```

# Generate data frame
data_pred2 <- data.frame(svlength_vector, sex_vector_random, zone_vector_random)
# Name columns
colnames(data_pred2) <- c("SV_length", "sex", "zone")

# Generate prediction model
prediction <- predict(tail_mod,
                     newdata = data_pred2,
                     type = "response",
                     se.fit = TRUE)

# Generate data frame of predictions
data_fig2 <- data.frame(data_pred2,
                       prediction$fit,
                       prediction$se.fit)

# Name columns
colnames(data_fig2) <- c("SV_length",
                       "Sex", "Zone", "probability", "se")

# Generate plot for sex coloring
sexplot <- ggplot(data_fig2, aes(x = SV_length,
                                y = probability)) +
  geom_line(aes(color = Sex)) +
  geom_ribbon(aes(ymin = probability - se,
                 ymax = probability + se,
                 fill = Sex), alpha = 0.3) +
  labs(x = "Snout-to-Vent Length (cm)",
       y = "Probability of Whole Tail",
       color = "Sex",
       fill = "Sex",
       title = "Prediction Model of Whole Tail Based on Sex",
       caption = "This figure displays the probability of a lizard having a
                 whole tail based on its snout-to-vent length and sex.") +
  scale_color_manual(values = c("violetred3", "purple3", "gold3",
                                "snow3"),
                    labels = c("Female", "Male", "J", "No Data")) +
  scale_fill_manual(values = c("violetred3", "purple3", "gold3",
                                "snow3"),
                   labels = c("Female", "Male", "J", "No Data")) +
  theme_bw() +
  theme(plot.caption = element_text(hjust = 0.5))

# View plot
sexplot

```

```

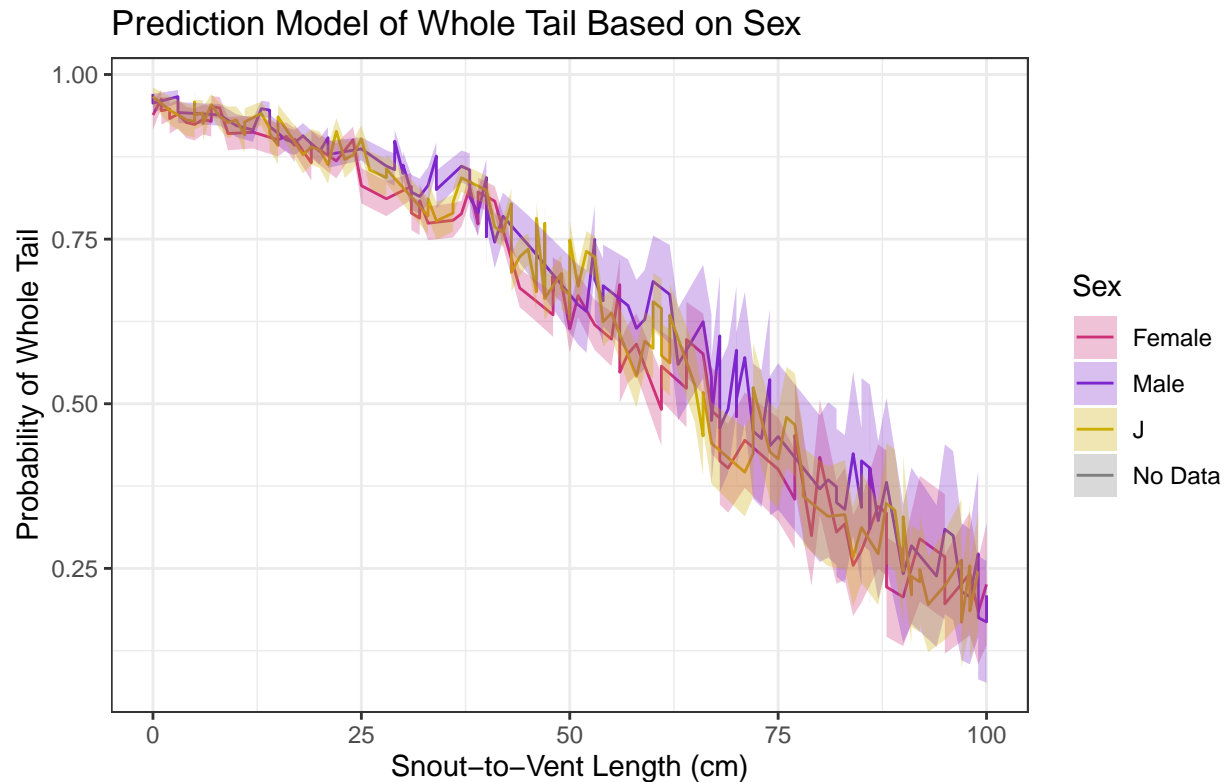
## Warning: Removed 108 rows containing missing values or values outside the scale range
## ('geom_line()').

```

```

## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf

```



Interpret Models

The results of a logistic regression suggest that the odds of a lizard having a whole tail are greater when the snout-to-vent length is lower ($\beta = -0.04523$, $p < 0.01$) and the lizard is less likely to be located in the black grama grassland ($\beta = -0.56537$, $p = 0.001$) or the tarbush shrubland ($\beta = -0.40435$, $p < 0.05$) zones. Lizards with a whole tail were not significantly correlated with the male ($\beta = 0.064915$, $p = 0.61$) or “J” ($\beta = 0.203871$, $p = 0.32$) species or with living in the mesquite duneland zone ($\beta = -0.301313$, $p = 0.063$).

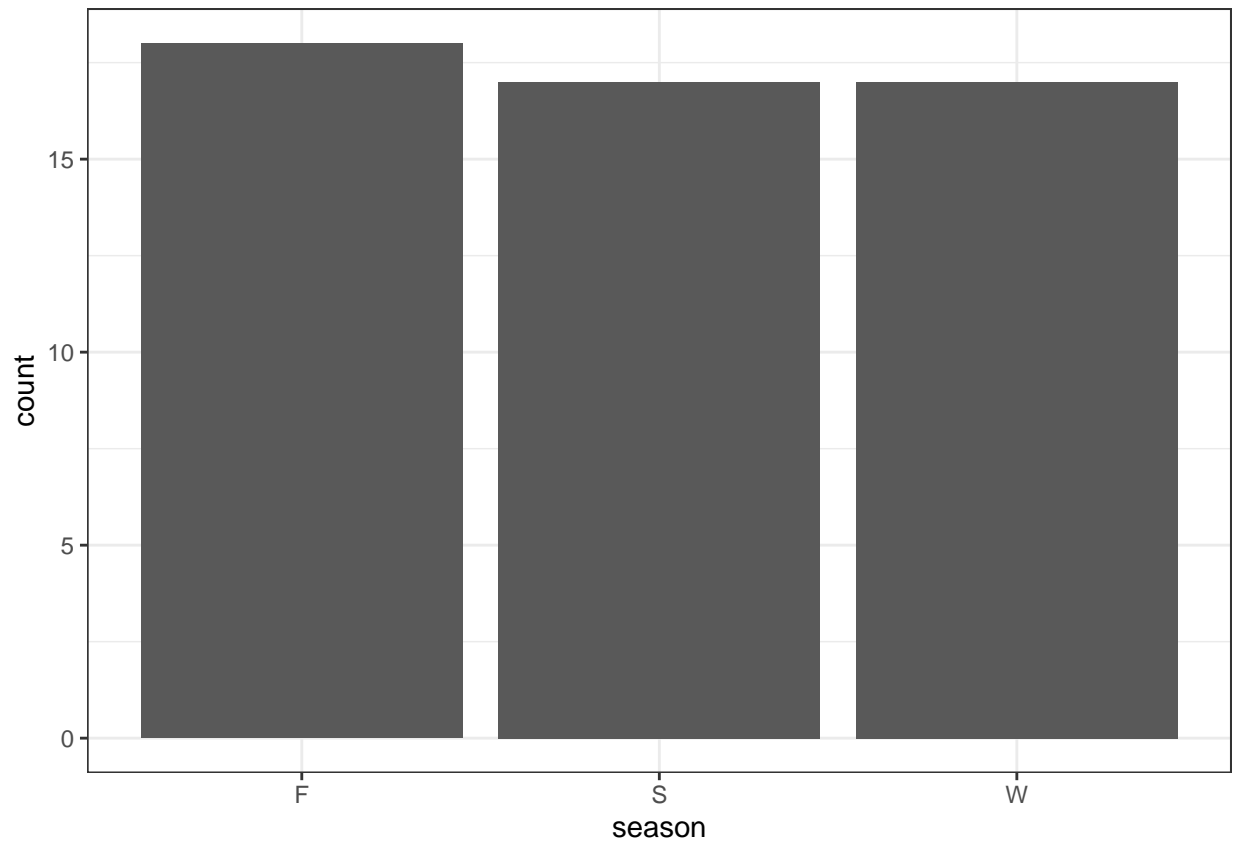
Poisson Regression

Do season and plant species percent cover significantly predict lizard counts?

Examine Data

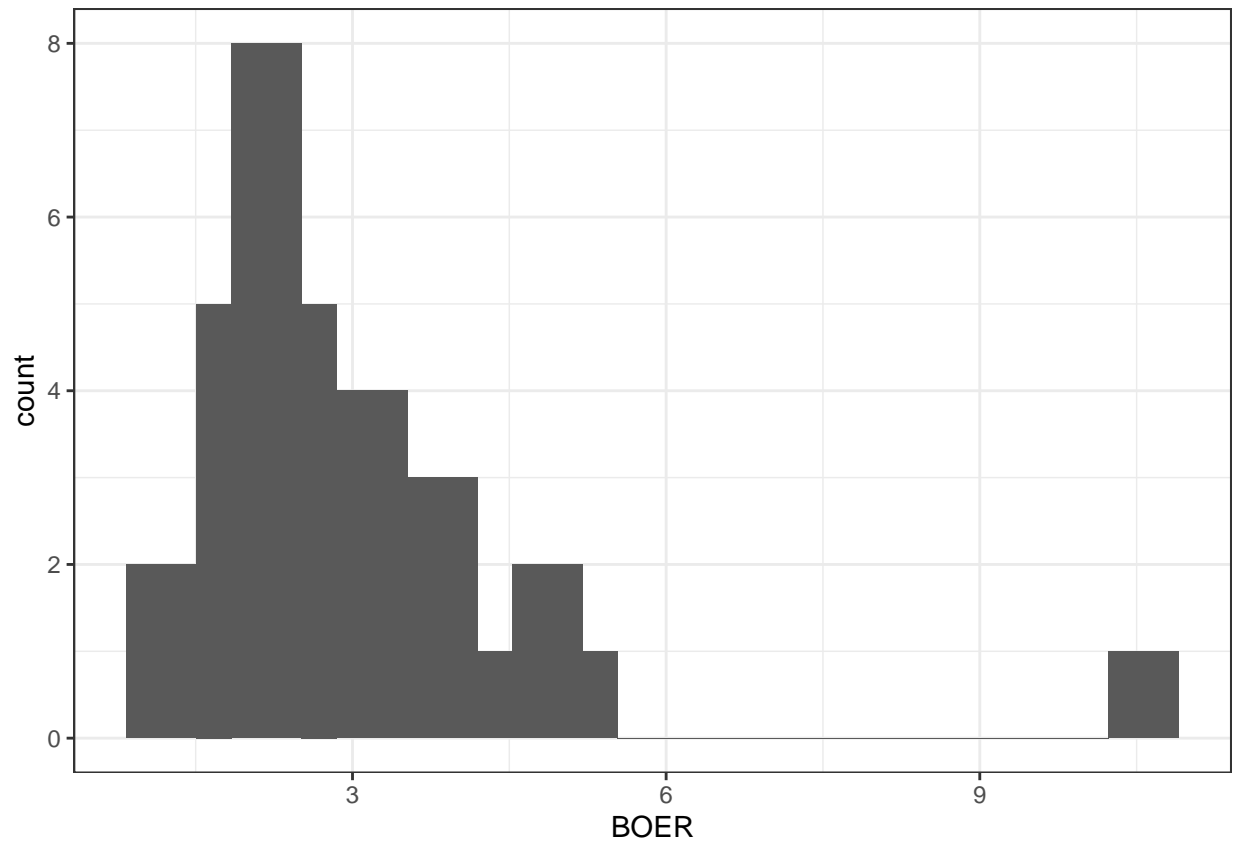
```
# Generate exploratory plots for each visual
(hist8 <- ggplot(data = lizard_npp, aes(x = season)) +
  geom_histogram(stat = "count", na.rm = TRUE) +
  theme_bw())
```

```
## Warning in geom_histogram(stat = "count", na.rm = TRUE): Ignoring unknown
## parameters: 'binwidth', 'bins', and 'pad'
```



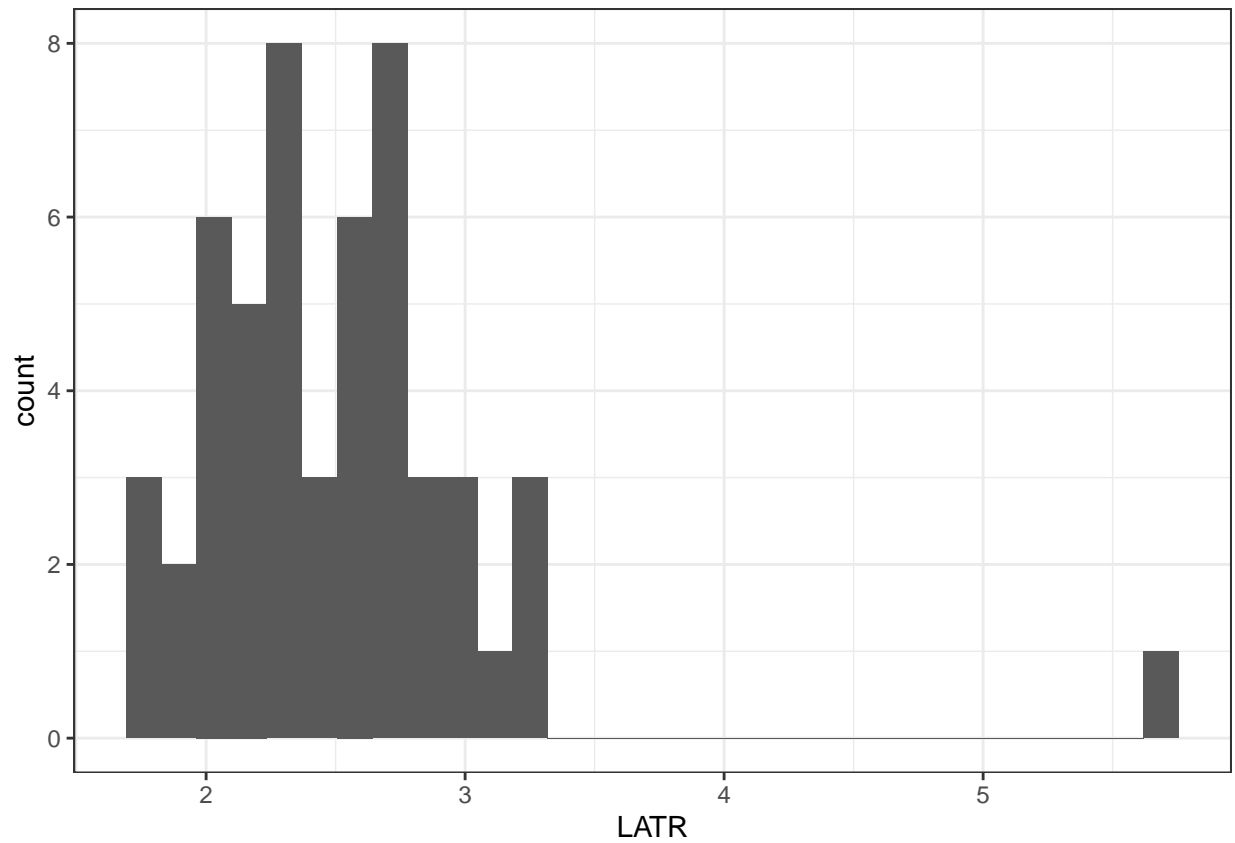
```
(hist9 <- ggplot(data = lizard_npp, aes(x = BOER)) +  
  geom_histogram(na.rm = TRUE) +  
  theme_bw())
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



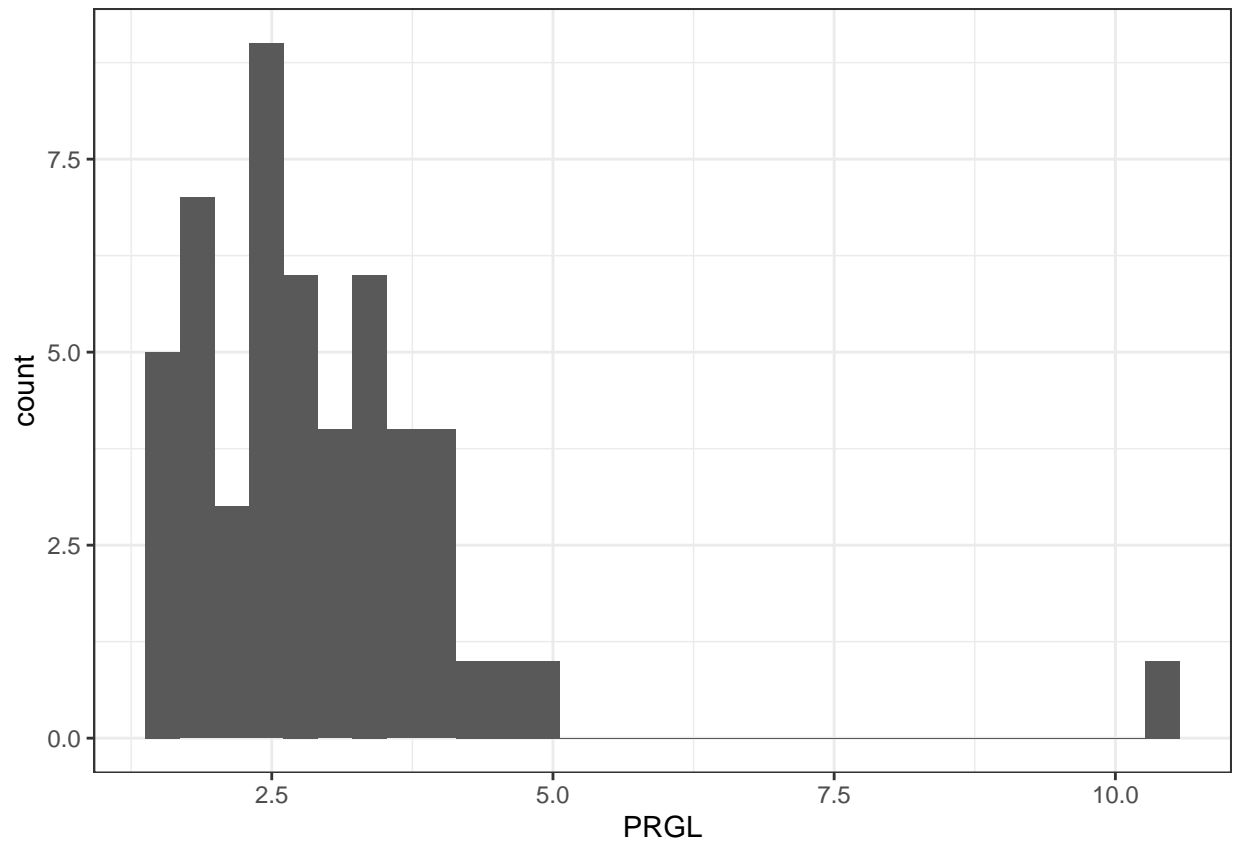
```
(hist10 <- ggplot(data = lizard_npp, aes(x = LATR)) +  
  geom_histogram(na.rm = TRUE) +  
  theme_bw())
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



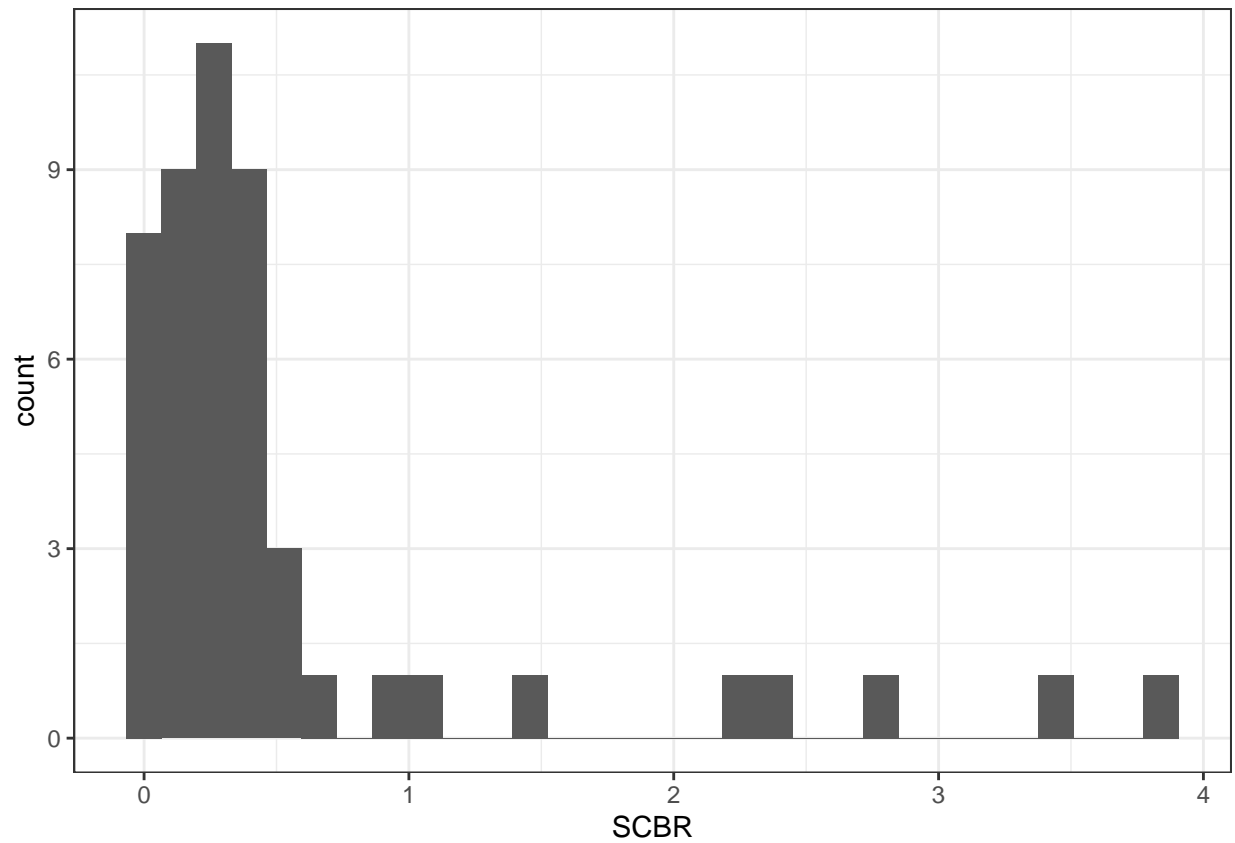
```
(hist11 <- ggplot(data = lizard_npp, aes(x = PRGL)) +  
  geom_histogram(na.rm = TRUE) +  
  theme_bw())
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

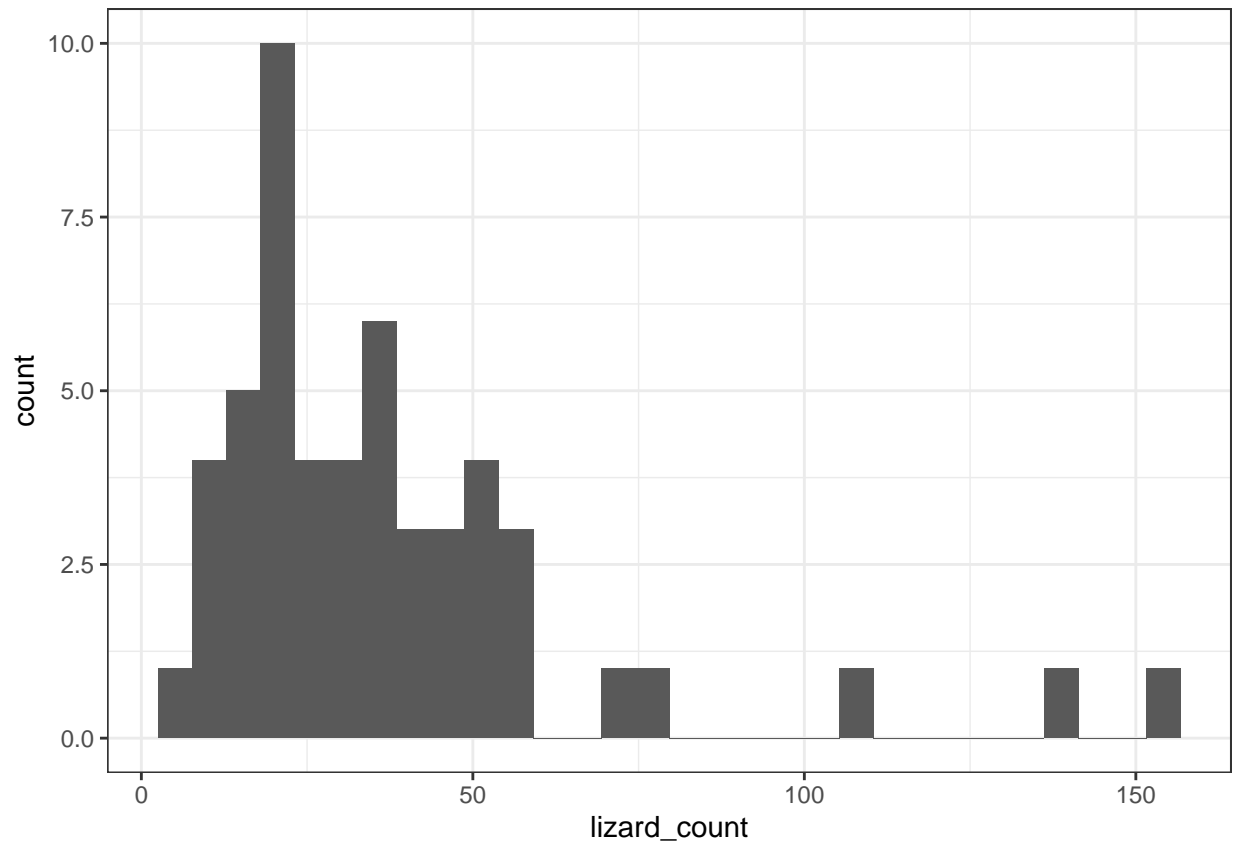
```
(hist12 <- ggplot(data = lizard_npp, aes(x = SCBR)) +  
  geom_histogram(na.rm = TRUE) +  
  theme_bw())
```

'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.



```
(hist13 <- ggplot(data = lizard_npp, aes(x = lizard_count)) +  
  geom_histogram(na.rm = TRUE) +  
  theme_bw())
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



Fit Regression Model

```
# Make season characters into factors
lizard_npp$season <- factor(lizard_npp$season,
                             levels = c("F", "W", "S"))

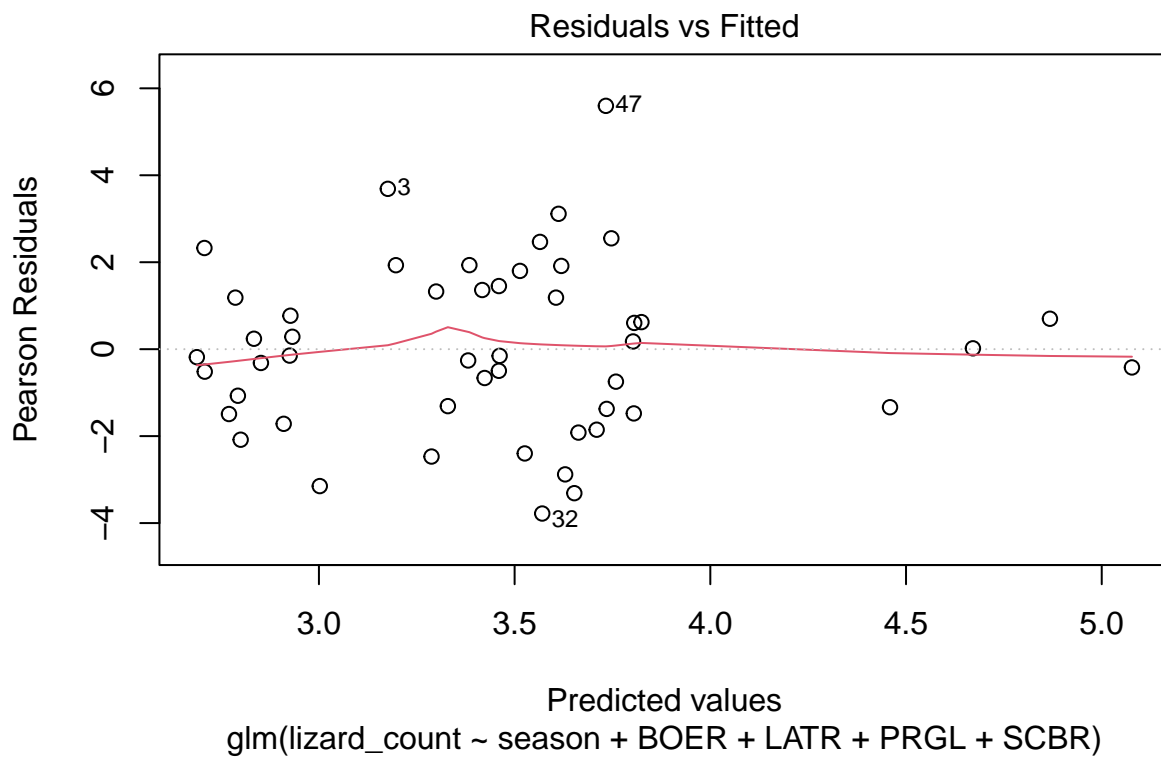
# Fit regression model
lizard_nppmod <- glm(lizard_count ~ season + BOER + LATR + PRGL + SCBR,
                     data = lizard_npp,
                     family = "poisson")

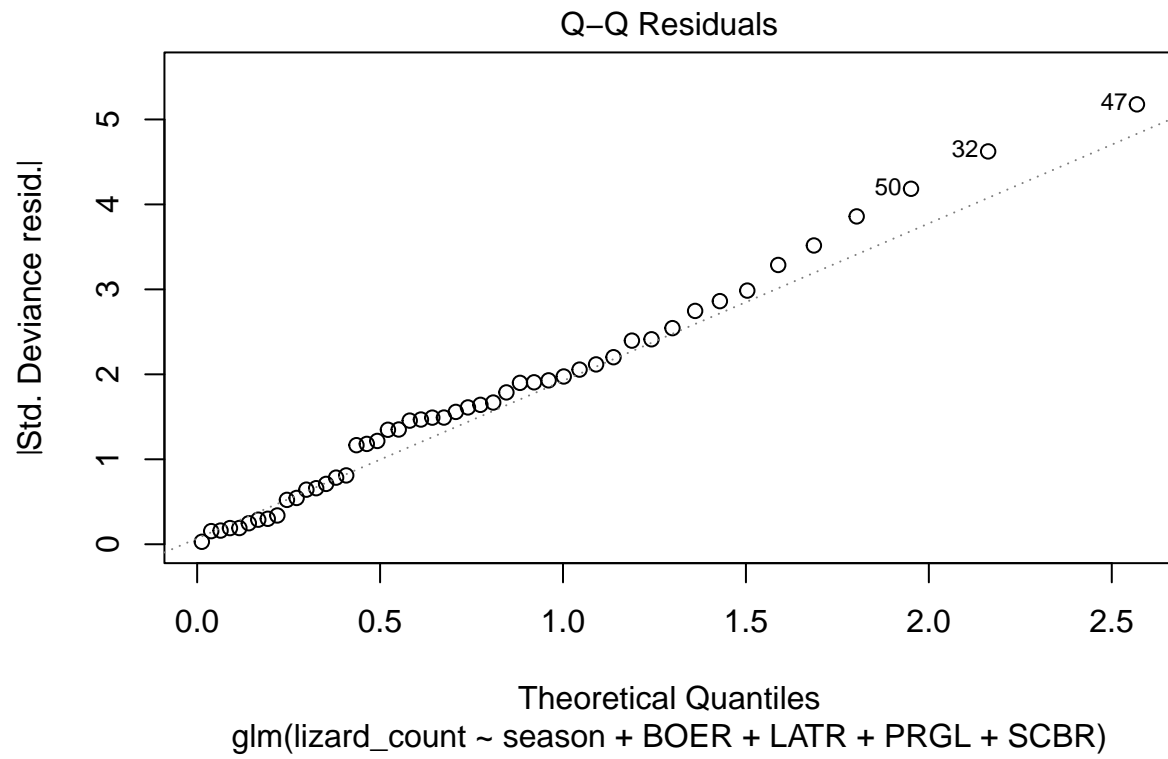
# View summary
summary(lizard_nppmod)
```

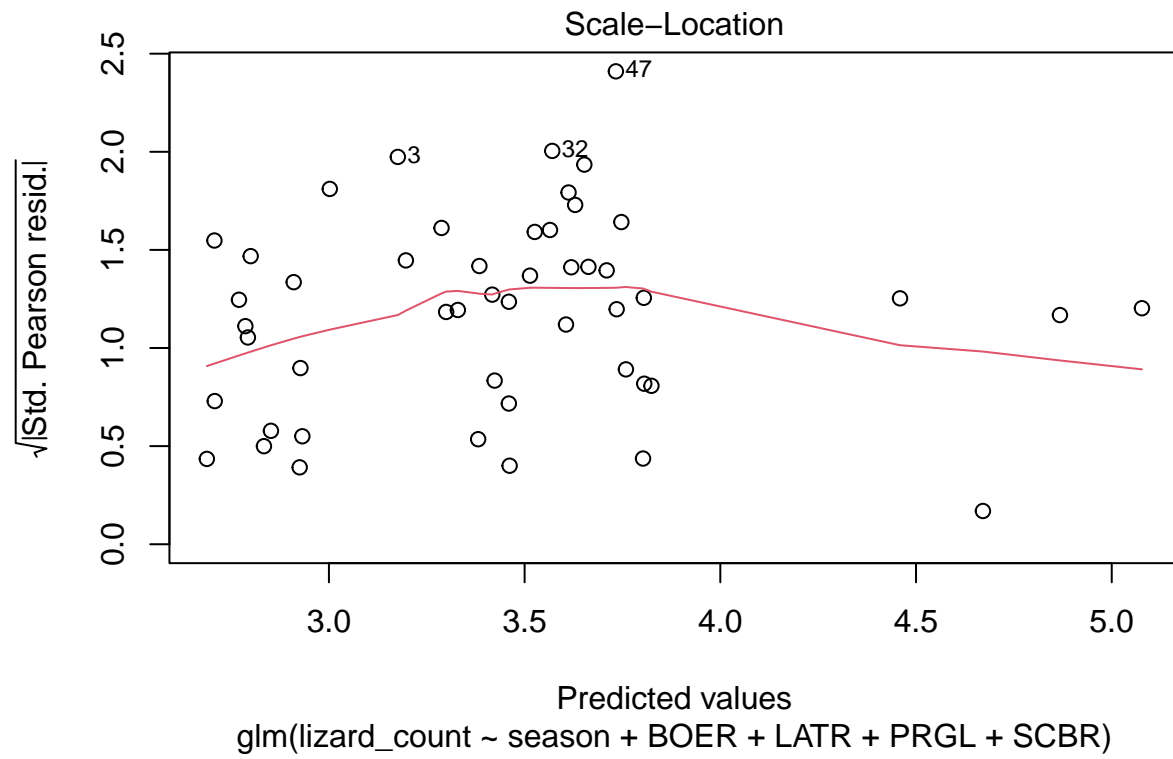
```
##
## Call:
## glm(formula = lizard_count ~ season + BOER + LATR + PRGL + SCBR,
##      family = "poisson", data = lizard_npp)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   3.76785    0.09880  38.136 < 2e-16 ***
## seasonW       0.20278    0.07464   2.717  0.00659 **
## seasonS      -0.75234    0.06970 -10.795 < 2e-16 ***
```

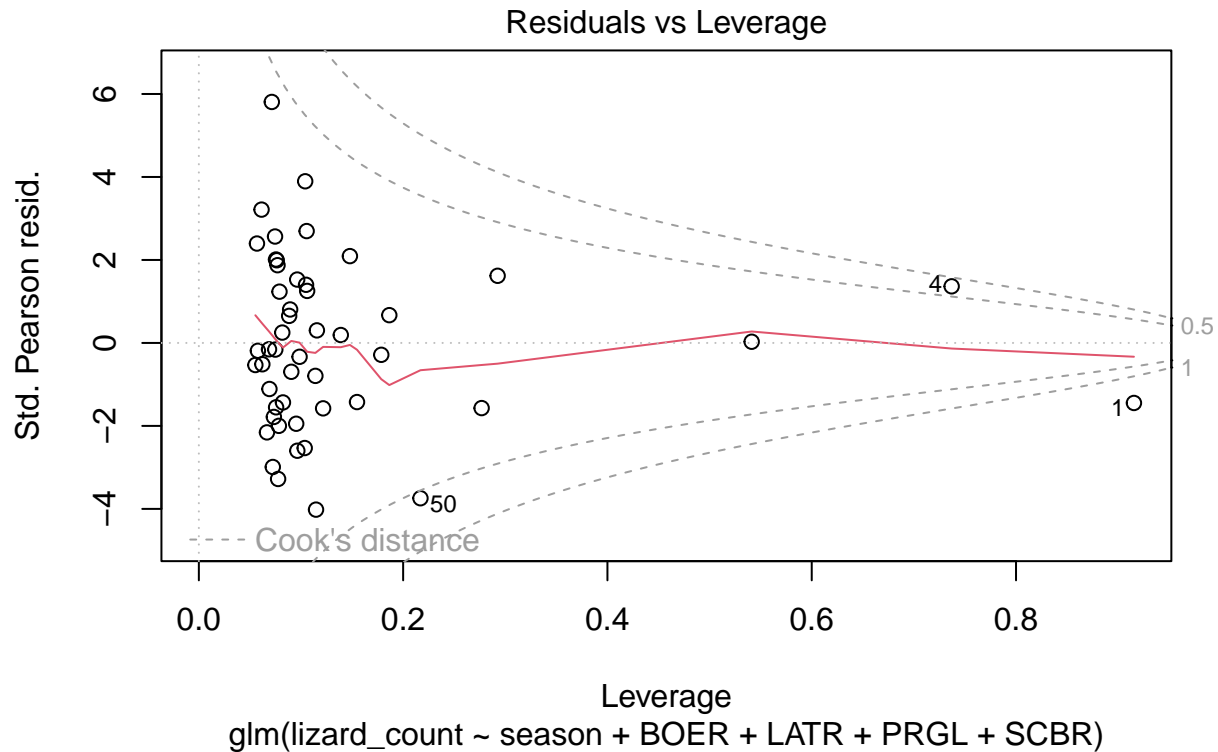
```
## BOER      -0.03260    0.02160   -1.509   0.13120
## LATR      -0.34305    0.07757   -4.423   9.74e-06 ***
## PRGL       0.20656    0.04412    4.681   2.85e-06 ***
## SCBR       0.37960    0.04199    9.040   < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 911.44  on 48  degrees of freedom
## Residual deviance: 184.19  on 42  degrees of freedom
## (3 observations deleted due to missingness)
## AIC: 453.85
##
## Number of Fisher Scoring iterations: 4
```

```
# View plots
plot(lizard_nppmod)
```









```
# Remove more outlier rows
lizard_npp2 <- lizard_npp[-c(32, 47, 50, 1),]

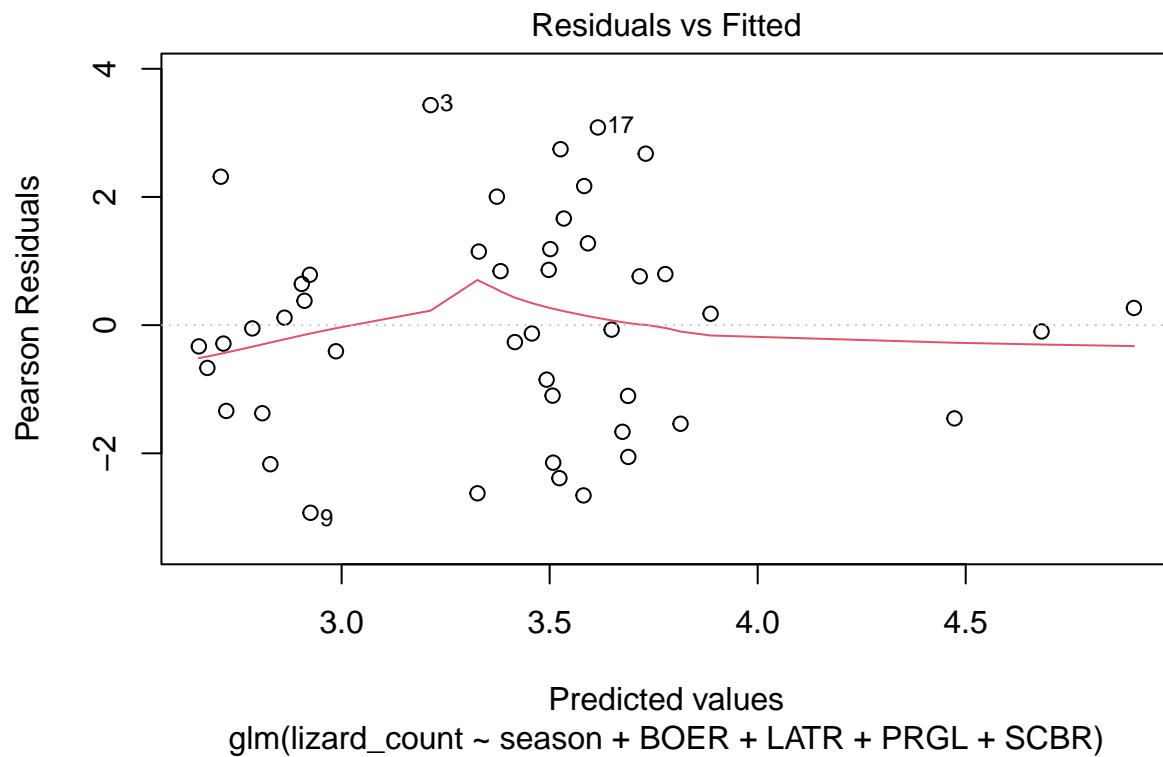
# Fit other regression model(s) and view summary/plots
lizard_nppmod2 <- glm(lizard_count ~ season + BOER + LATR + PRGL + SCBR,
  data = lizard_npp2,
  family = "poisson")

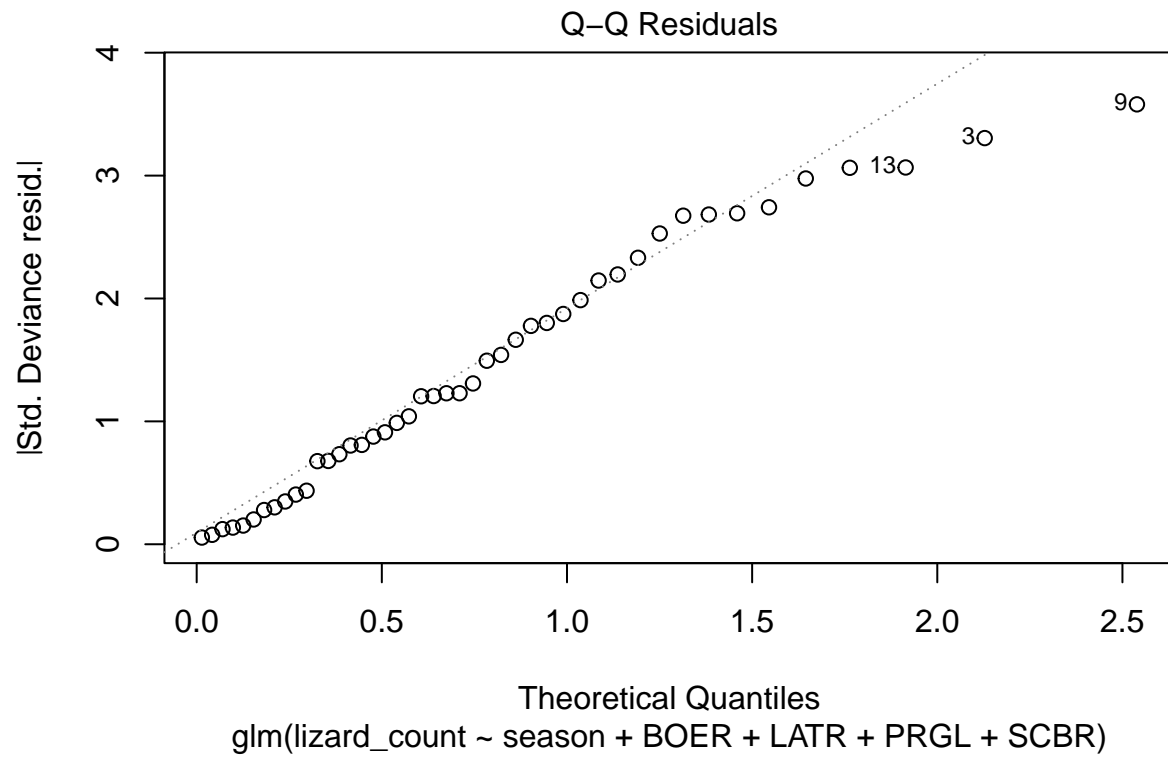
summary(lizard_nppmod2)
```

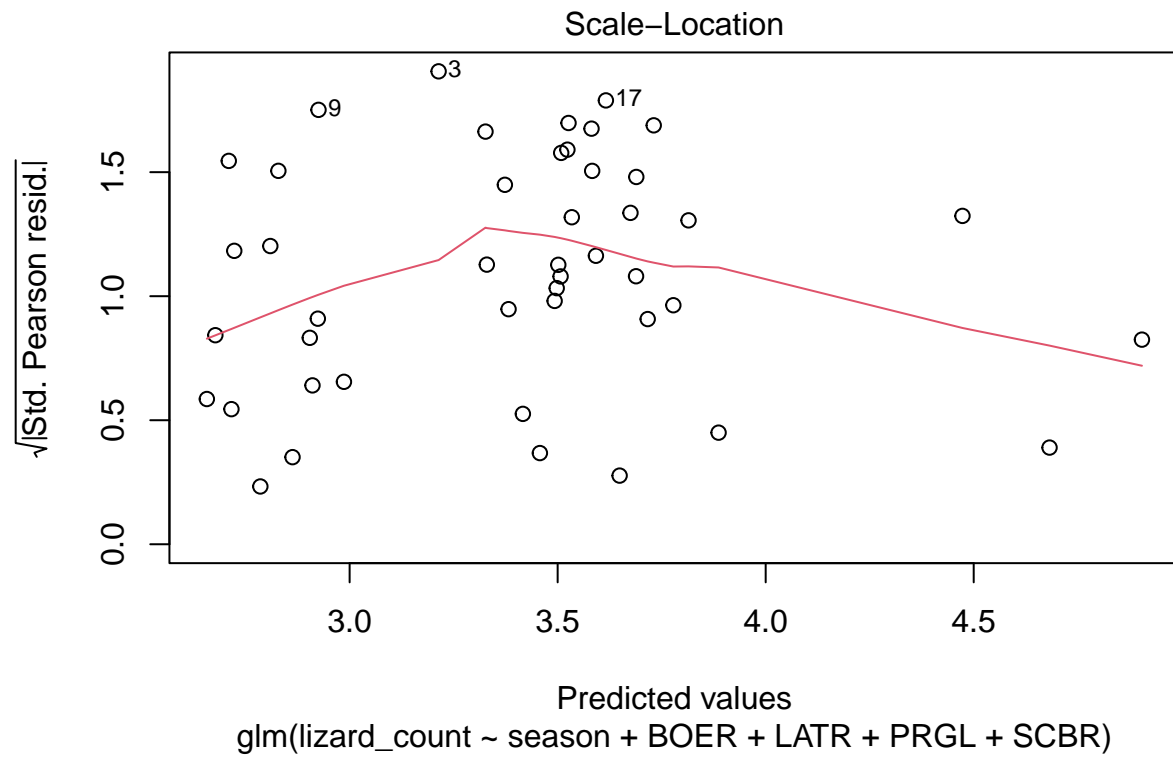
```
##
## Call:
## glm(formula = lizard_count ~ season + BOER + LATR + PRGL + SCBR,
##      family = "poisson", data = lizard_npp2)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.43721    0.17553  19.581 < 2e-16 ***
## seasonW      0.14259    0.07833   1.820 0.068695 .
## seasonS     -0.80036    0.07279 -10.995 < 2e-16 ***
## BOER        -0.03995    0.02402  -1.663 0.096265 .
## LATR        -0.15178    0.09293  -1.633 0.102425
## PRGL         0.17461    0.05157   3.386 0.000709 ***
## SCBR         0.40488    0.04560   8.878 < 2e-16 ***
## ---
```

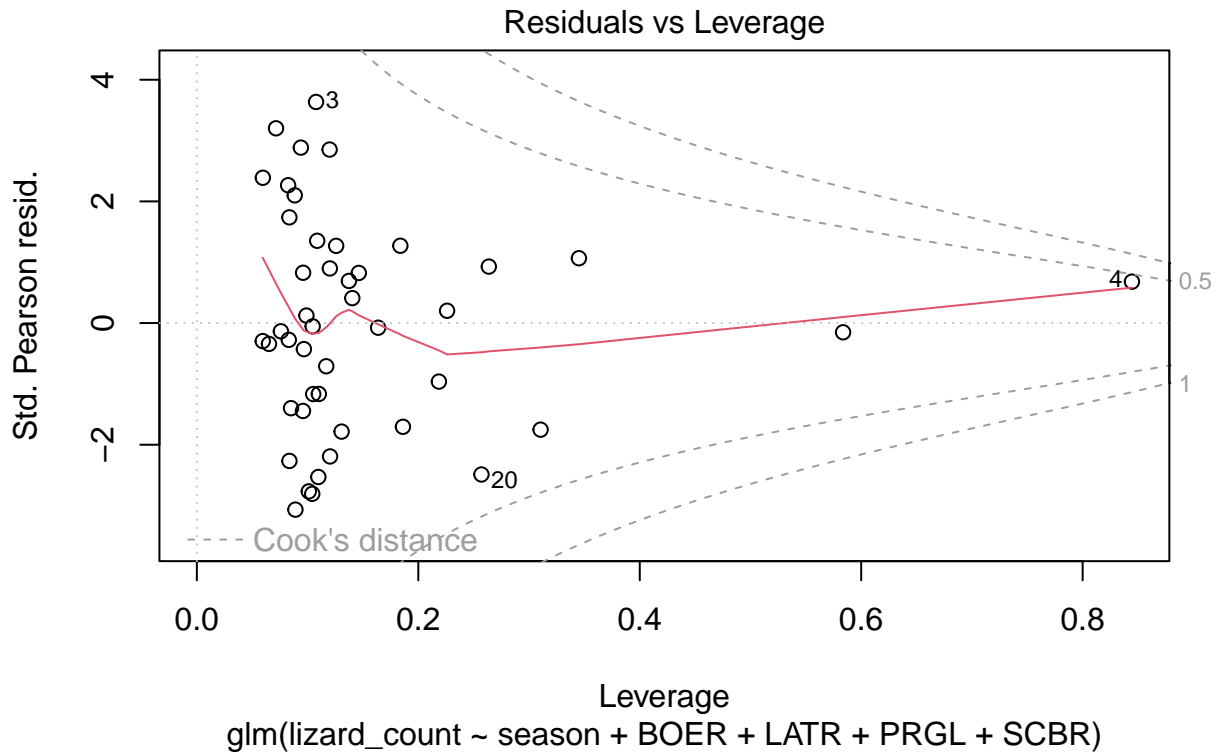
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##    Null deviance: 629.31  on 44  degrees of freedom
## Residual deviance: 121.17  on 38  degrees of freedom
##   (3 observations deleted due to missingness)
## AIC: 368.6
##
## Number of Fisher Scoring iterations: 4
```

```
plot(lizard_nppmod2)
```









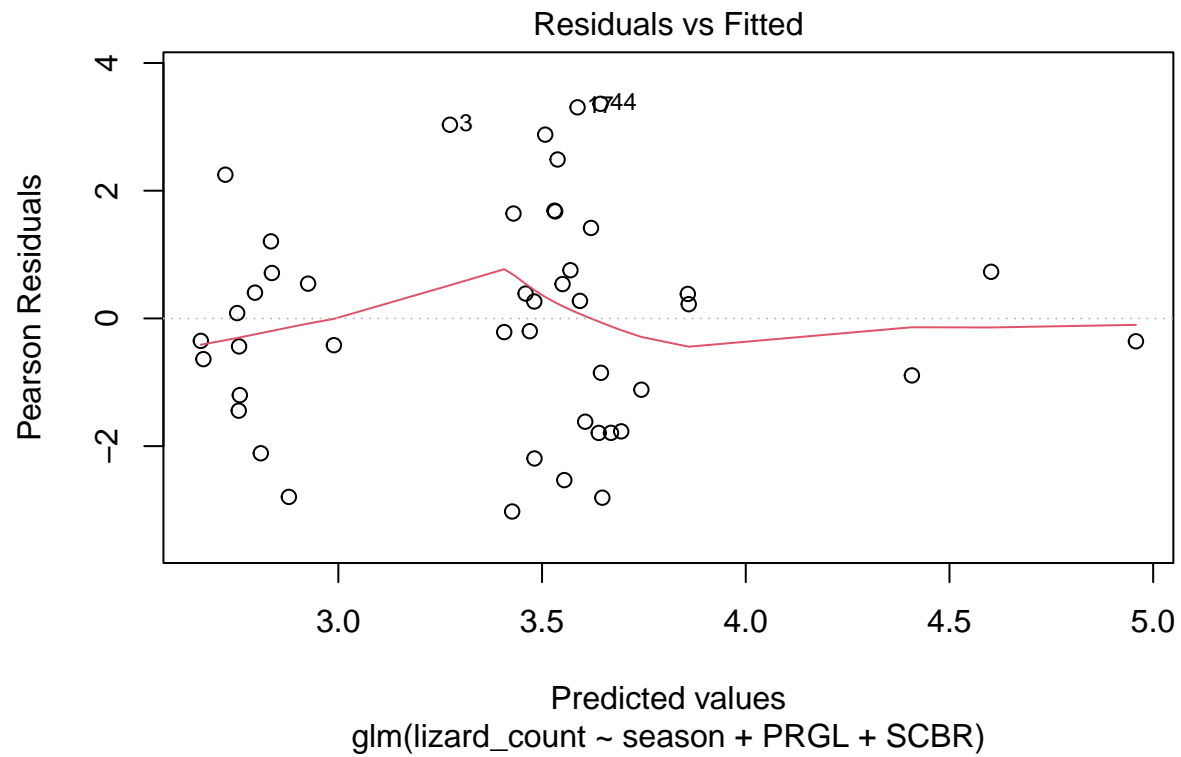
```
lizard_nppmod3 <- glm(lizard_count ~ season + PRGL + SCBR,
  data = lizard_npp2,
  family = "poisson")
```

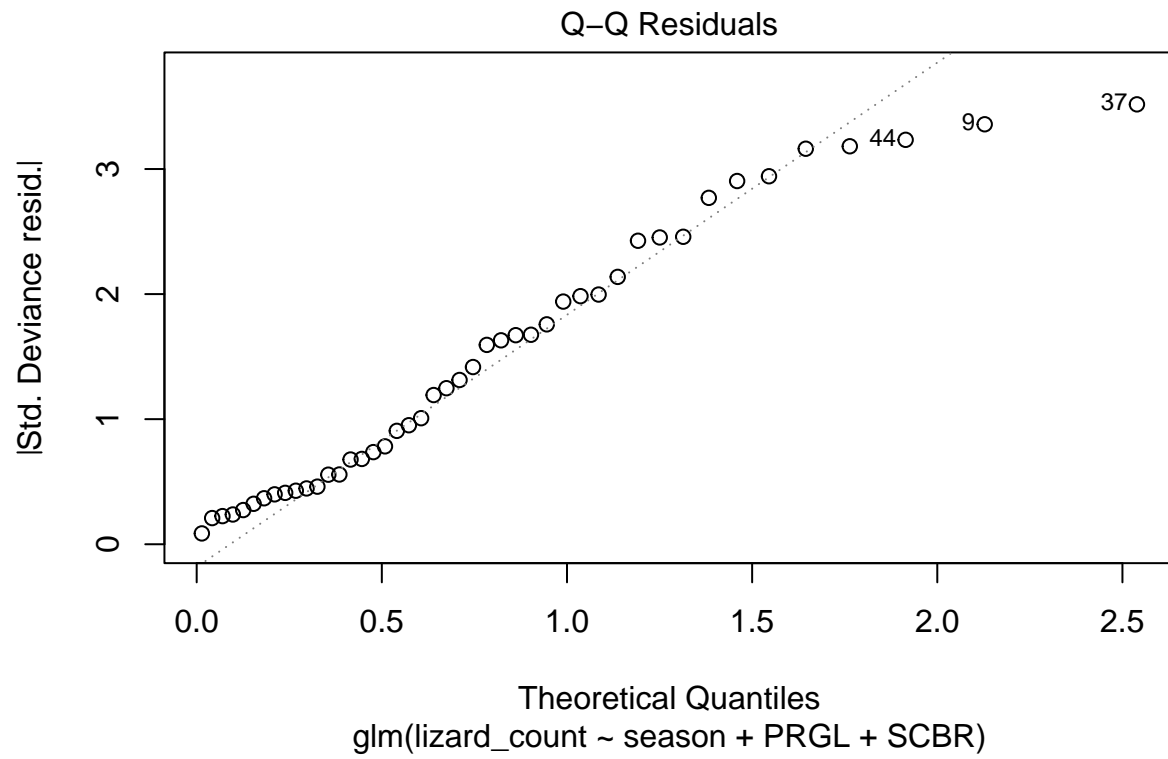
```
summary(lizard_nppmod3)
```

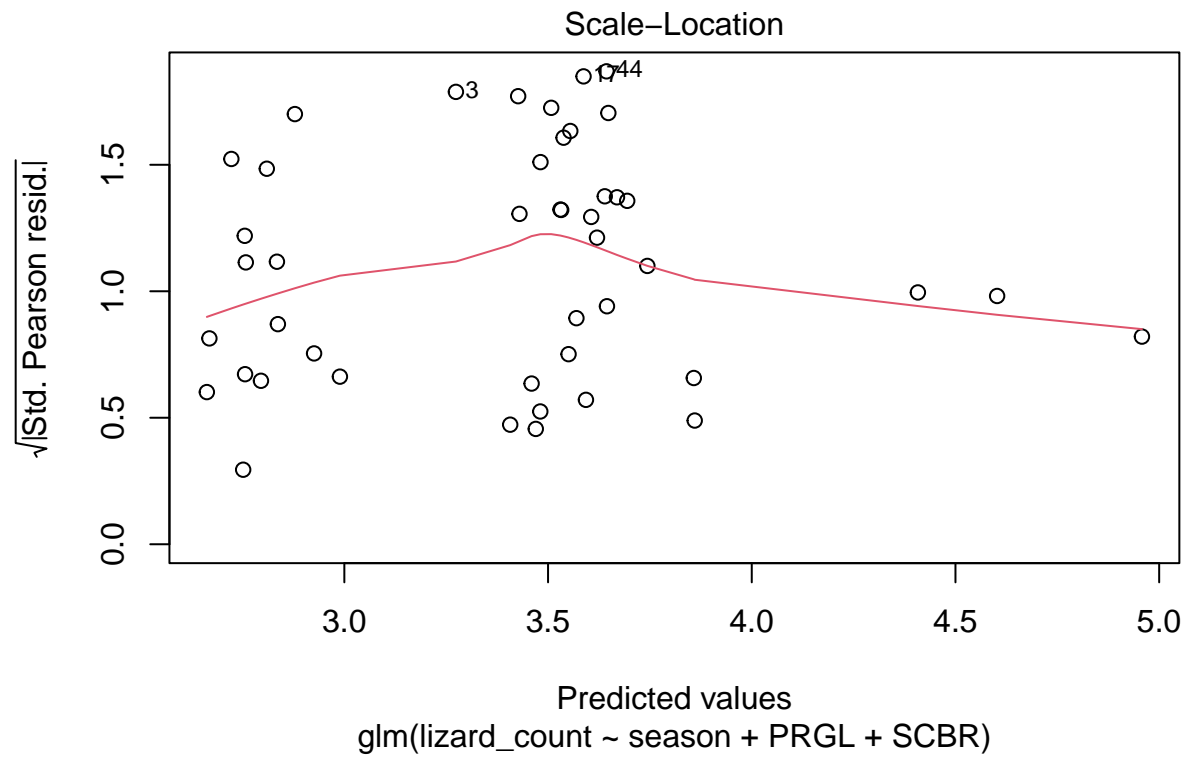
```
##
## Call:
## glm(formula = lizard_count ~ season + PRGL + SCBR, family = "poisson",
##      data = lizard_npp2)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.26177    0.11451  28.485  <2e-16 ***
## seasonW      0.01593    0.06314   0.252  0.8008
## seasonS     -0.81633    0.07117 -11.471  <2e-16 ***
## PRGL         0.08224    0.03870   2.125  0.0336 *
## SCBR         0.38067    0.03247  11.723  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 629.31  on 44  degrees of freedom
## Residual deviance: 129.11  on 40  degrees of freedom
```

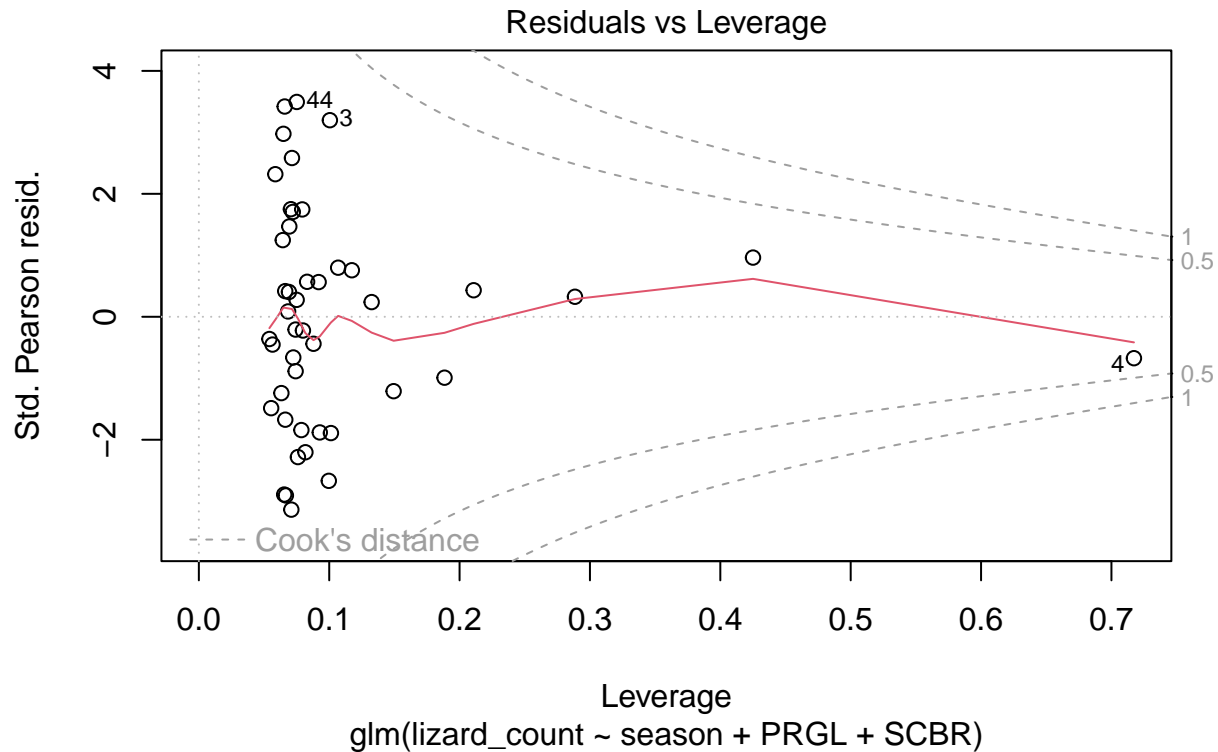
```
## (3 observations deleted due to missingness)
## AIC: 372.54
##
## Number of Fisher Scoring iterations: 4
```

```
plot(lizard_nppmod3)
```









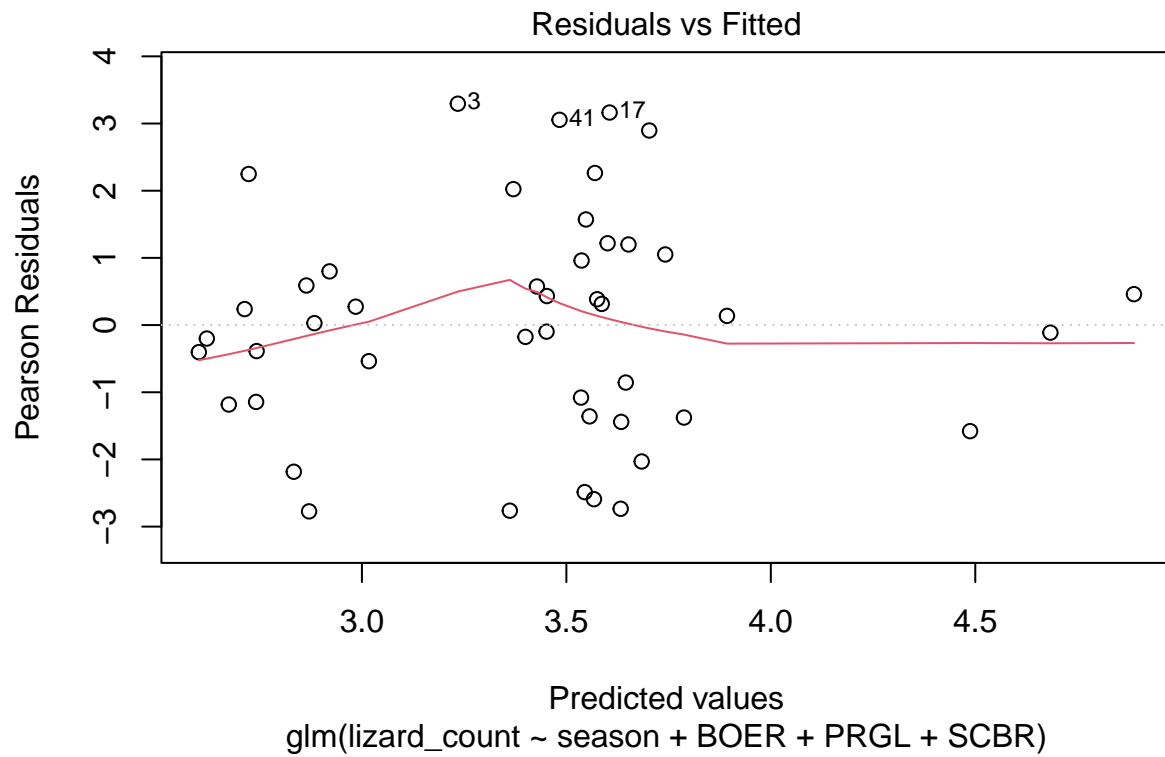
```
lizard_nppmod4 <- glm(lizard_count ~ season + BOER + PRGL + SCBR,
  data = lizard_npp2,
  family = "poisson")
```

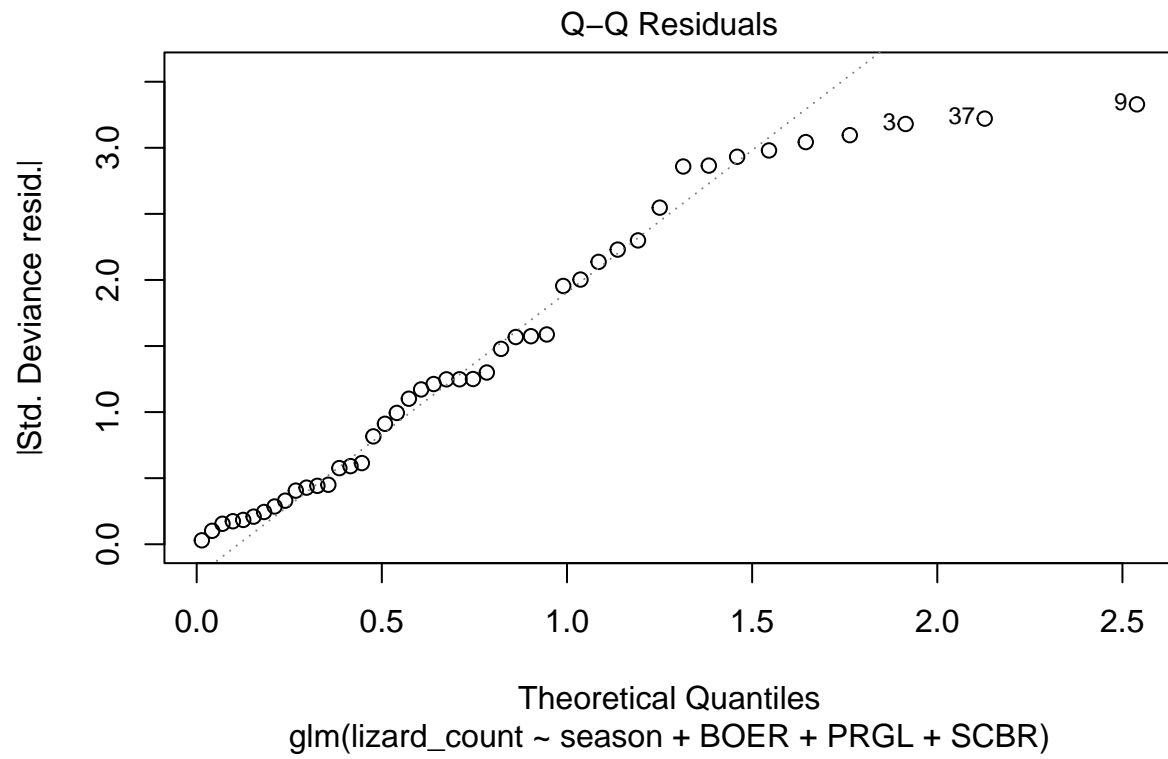
```
summary(lizard_nppmod4)
```

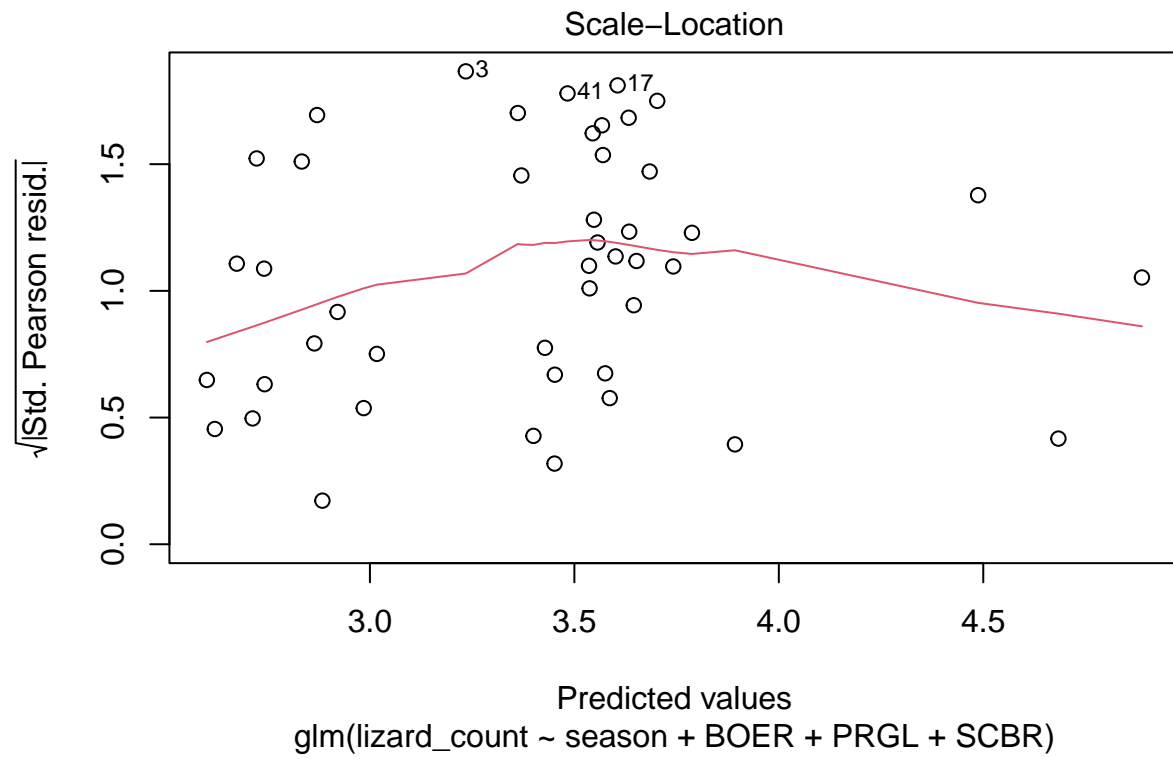
```
##
## Call:
## glm(formula = lizard_count ~ season + BOER + PRGL + SCBR, family = "poisson",
##      data = lizard_npp2)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.22100    0.11689  27.555 < 2e-16 ***
## seasonW      0.10777    0.07525   1.432  0.15208
## seasonS     -0.82462    0.07129 -11.567 < 2e-16 ***
## BOER        -0.05221    0.02286  -2.284  0.02238 *
## PRGL         0.13224    0.04449   2.972  0.00296 **
## SCBR         0.43805    0.04064  10.778 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 629.31  on 44  degrees of freedom
```

```
## Residual deviance: 123.87 on 39 degrees of freedom
## (3 observations deleted due to missingness)
## AIC: 369.29
##
## Number of Fisher Scoring iterations: 4
```

```
plot(lizard_nppmod4)
```

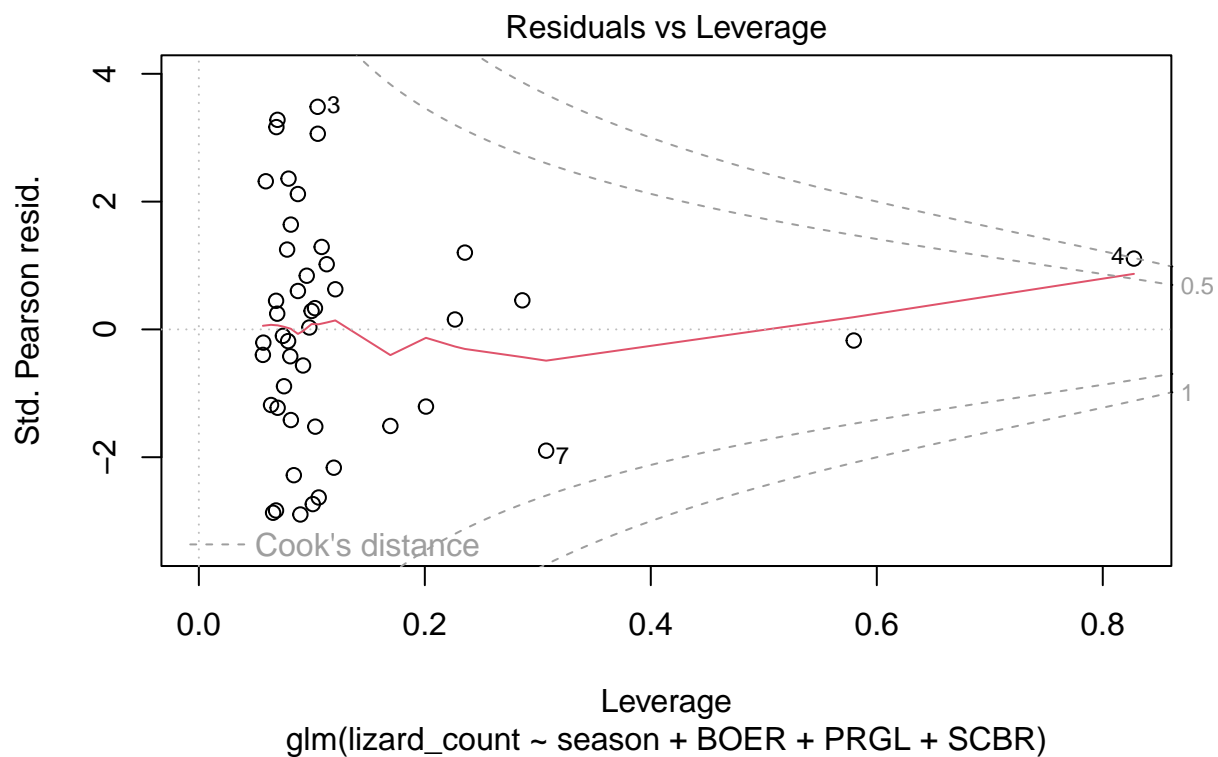






Characteristic	log(IRR)	95% CI	p-value
season			
F	—	—	
W	0.11	-0.04, 0.26	0.2
S	-0.82	-0.97, -0.69	<0.001
BOER	-0.05	-0.10, -0.01	0.022
PRGL	0.13	0.04, 0.22	0.003
SCBR	0.44	0.36, 0.52	<0.001

Abbreviations: CI = Confidence Interval, IRR = Incidence Rate Ratio



```
# This fourth model has the best significance and AIC
# View regression p-values and confidence intervals
tbl_regression(lizard_nppmod4)
```

Interpret Models

The results of a Poisson regression show that lizard counts are lower in the summer (beta = -0.82462, $p < 0.01$), lower in BOER (beta = -0.05221, $p < 0.05$), and higher in PRGL (beta = 0.13224, $p < 0.01$) and in SCBR (beta = 0.43805, $p < 0.01$). There was not a strong correlation between lizard counts and the winter

season ($\beta = 0.10777$, $p = 0.15$). To perform this regression, 4 rows of outliers were removed since they were well out of the range of the expected regression based on the plot visuals of prior Poisson regression models.