

Week 9 Assignment

Becca Cox

2025-03-18

Set Up

```
lizards <- read.csv("jrn_lizard.csv")
```

```
library(here)
```

```
## here() starts at /Users/beccacox/Desktop/MEM 24-25/ENV710_Lab/beccaeco.git
```

```
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(naniar)
```

```
library(DHARMA)
```

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

```
library(gtsummary)
```

```
library(broom)
```

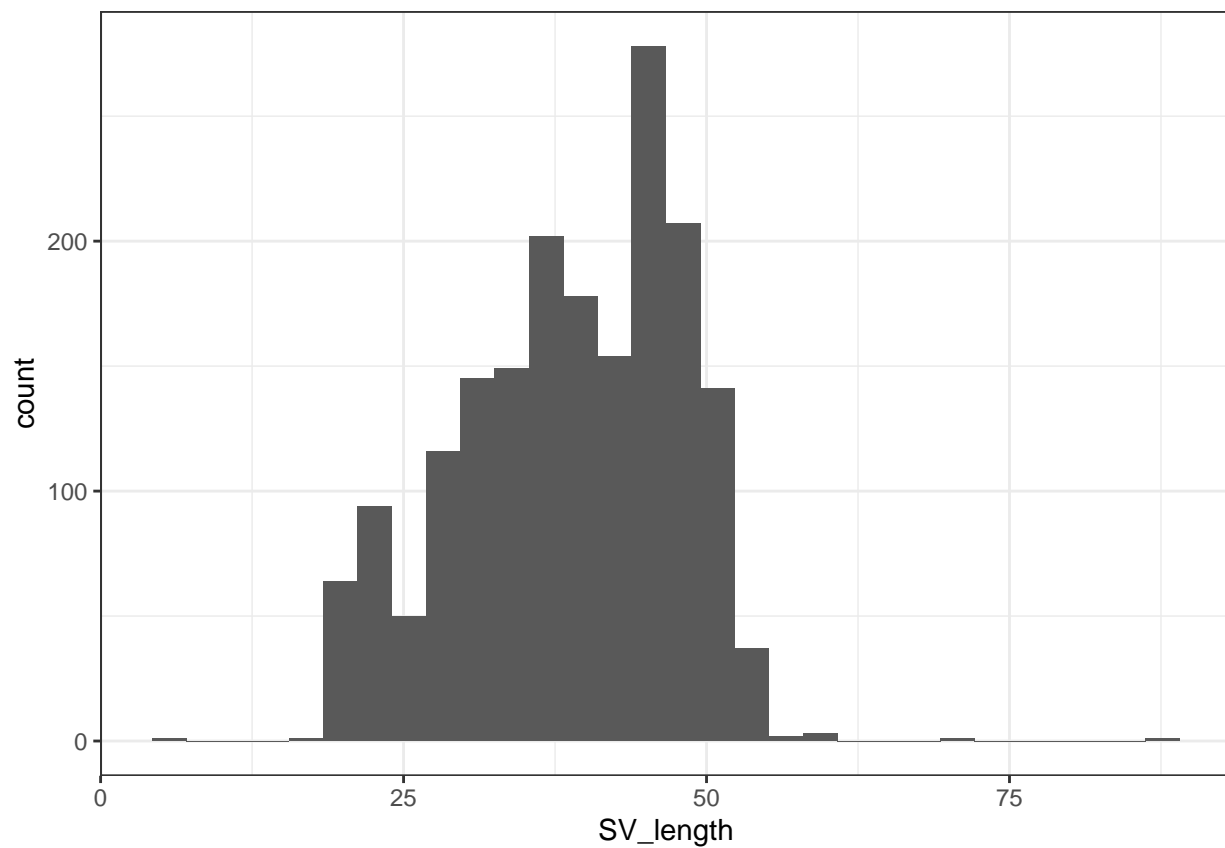
```
lizards_filtered <- lizards %>%
  filter(spp == "UTST")
```

Exploratory Figures

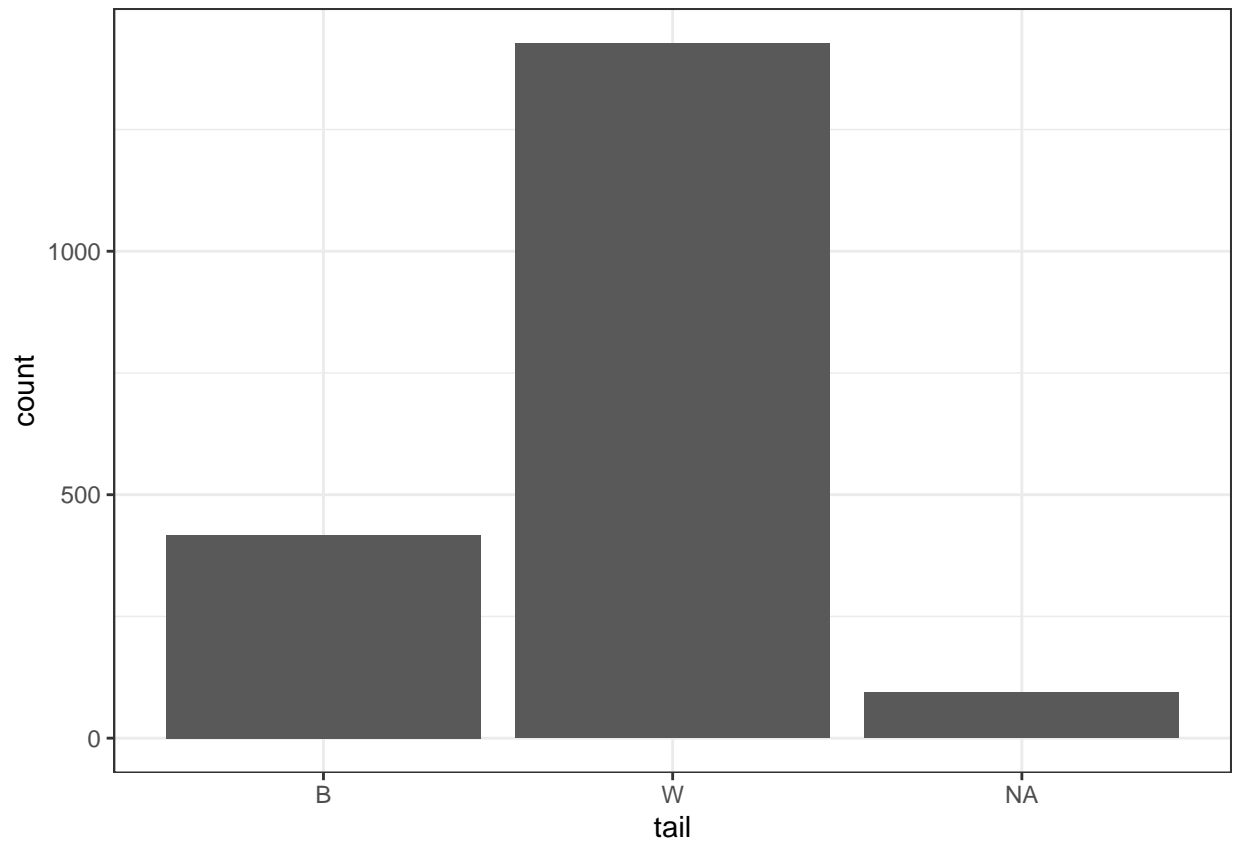
```
hist1 <- ggplot(lizards_filtered, aes(x = SV_length)) +
  geom_histogram() +
  theme_bw()
hist1
```

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

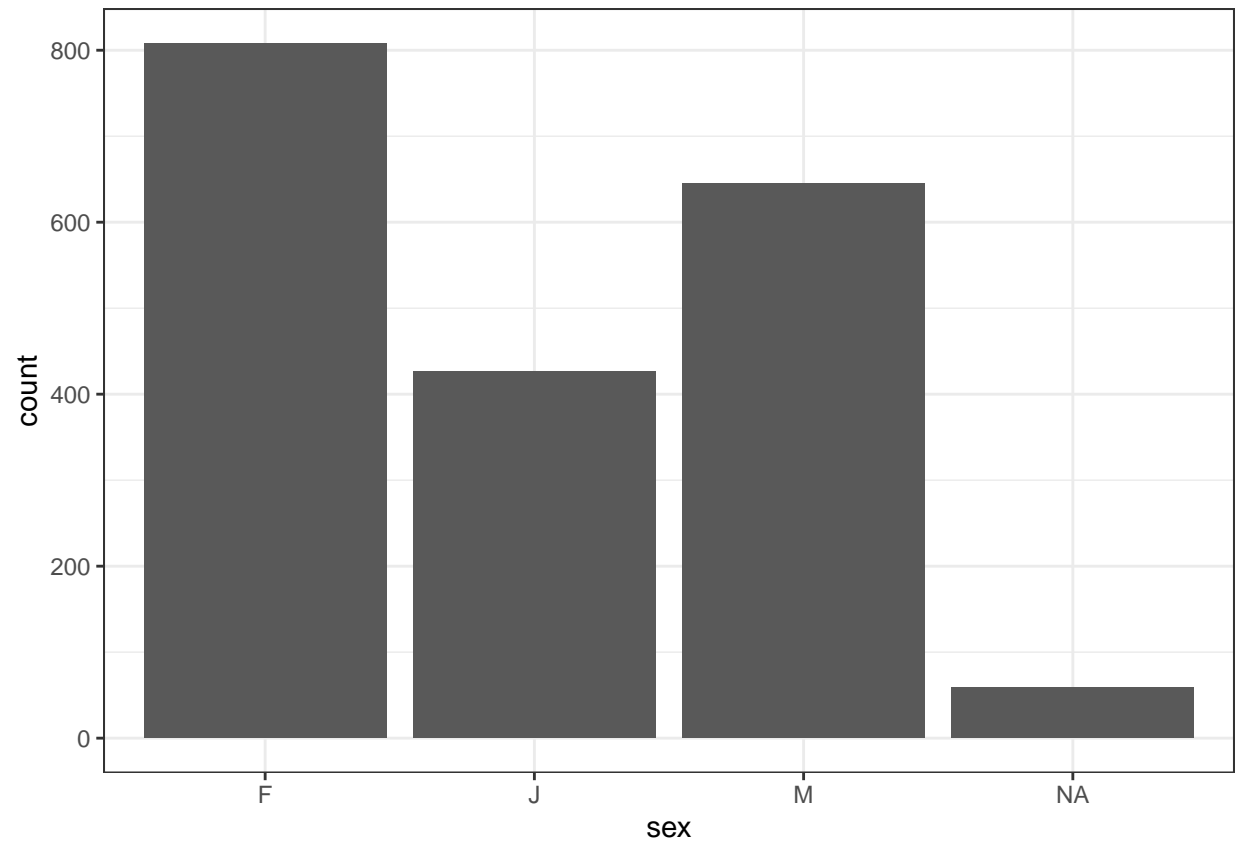
```
## Warning: Removed 114 rows containing non-finite outside the scale range
## ('stat_bin()').
```



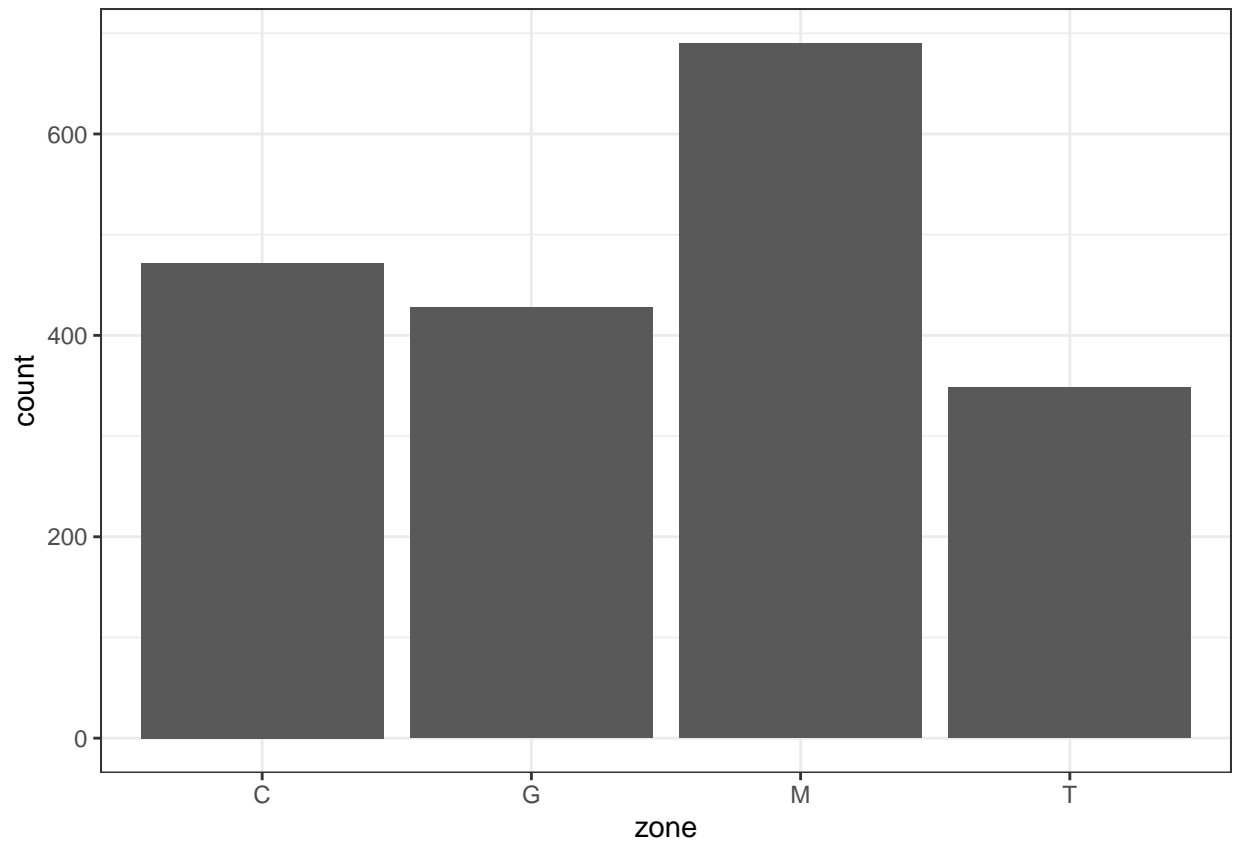
```
bar1 <- ggplot(lizards_filtered, aes(x = tail)) + #cannot do histogram without continuous x
  geom_bar() +
  theme_bw()
bar1
```



```
bar2 <- ggplot(lizards_filtered, aes(x = sex)) + geom_bar() + theme_bw()  
bar2
```



```
bar3 <- ggplot(lizards_filtered, aes(x = zone)) + geom_bar() + theme_bw()  
bar3
```



As Factors and Model

```
unique(lizards_filtered$zone)
```

```
## [1] "C" "G" "M" "T"
```

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

```
unique(lizards_filtered$sex)
```

```
## [1] "J" "F" NA  "M"
```

```
lizards_filtered$sex <- factor(lizards_filtered$sex,  
                                levels = c("J", "F", "M"))
```

```
unique(lizards_filtered$tail)
```

```
## [1] "B" "W" NA
```

```
lizards_filtered$tail <- factor(lizards_filtered$tail,
                                levels = c("B", "W"))
```

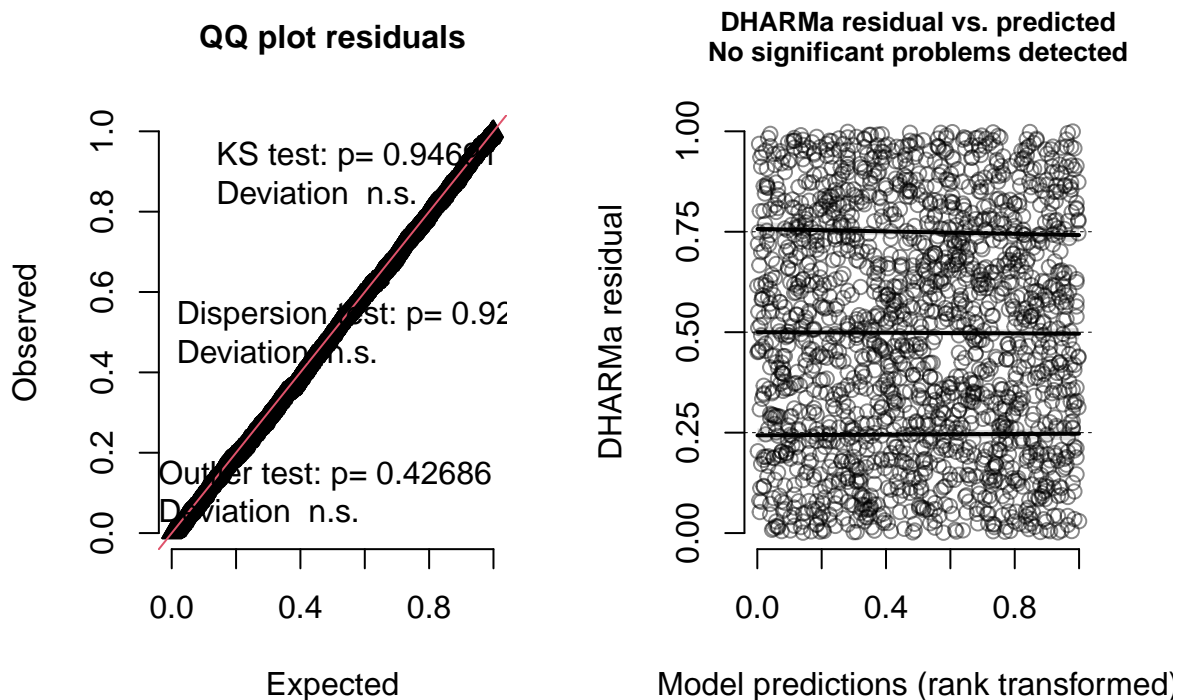
```
liz_mod <- glm(tail~SV_length+sex+zone,
               data = lizards_filtered,
               family = "binomial")
```

```
summary(liz_mod)
```

```
##
## Call:
## glm(formula = tail ~ SV_length + sex + zone, family = "binomial",
##      data = lizards_filtered)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.493940   0.332420  10.511 < 2e-16 ***
## SV_length   -0.045230   0.008877  -5.095 3.49e-07 ***
## sexF        -0.203871   0.206033  -0.990 0.32241
## sexM        -0.138956   0.223664  -0.621 0.53442
## 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 *
## ---
## 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
```

```
simulateResiduals(liz_mod) %>% plot()
```

DHARMA residual



The results of a logistic regression suggest that the log-odds of having a whole tail are greater with lower snout-to-vent length ($B = -0.045$, $p < 0.001$). There was also a significant decrease in log-odds of having a whole tail in vegetation zones with black grama grassland ($B = -0.565$, $p = 0.001$) and tarbush shrubland ($B = -0.404$, $p = 0.025$) relative to creosotebush shrubland. Mesquite duneland vegetation zones ($B = -0.301$, $p = 0.063$) did not have significant log-odds of having a whole tail relative to creosotebush shrubland. Neither females ($B = -0.204$, $p = 0.322$) nor males ($B = -0.139$, $p = 0.534$) had significant log-odds of having a whole tail relative to juveniles.

Predictive Figures

```
# variables are SV length, sex, and zone

range(lizards_filtered$SV_length, na.rm=TRUE)

## [1] 5 87

length_vector <- rep(seq(from = 5, to = 87), 12) #83, 3, and 4 don't have common factors
sex_vector <- rep(c("J", "F", "M"), 332)

zone_vector <- c(rep("C", 249),
                 rep("G", 249),
                 rep("M", 249),
                 rep("T", 249))
```

```

data_pred <- data.frame(sex_vector, length_vector, zone_vector)
colnames(data_pred) <- c("sex", "SV_length", "zone")

prediction <- predict(liz_mod,
                     newdata = data_pred,
                     type = "response",
                     se.fit = TRUE)

data_fig <- data.frame(data_pred,
                      prediction$fit,
                      prediction$se.fit)

colnames(data_fig) <- c("Sex", "Length", "Zone", "probability", "se")

ggplot(data_fig, aes(x = Length,
                    y = probability)) +
  geom_smooth(aes(color = Zone)) +
  geom_ribbon(aes(ymin = probability - se,
                ymax = probability + se,
                fill = Zone), alpha = 0.3) +
  labs(x = "Length", y = "Probability of Whole Tail", color = "Zone", fill = "Zone") +
  theme_bw()

```

```
## 'geom_smooth()' using method = 'loess' and formula = 'y ~ x'
```

```
# jagged because there are so many replicates?
```

```

ggplot(data_fig, aes(x = Length,
                    y = probability)) +
  geom_smooth(aes(color = Sex)) +
  geom_ribbon(aes(ymin = probability - se,
                ymax = probability + se,
                fill = Sex), alpha = 0.3) +
  labs(x = "Length", y = "Probability of Whole Tail", color = "Sex", fill = "Sex") +
  theme_bw()

```

```
## 'geom_smooth()' using method = 'loess' and formula = 'y ~ x'
```

The probability of having a whole tail decreases as snout-to-vent length increases for all sexes and vegetation zones.

Poisson Regression: Exploratory Plots

```
lizards2 <- read_csv("jrn_lizard_npp.csv")
```

```
## Rows: 52 Columns: 7
```

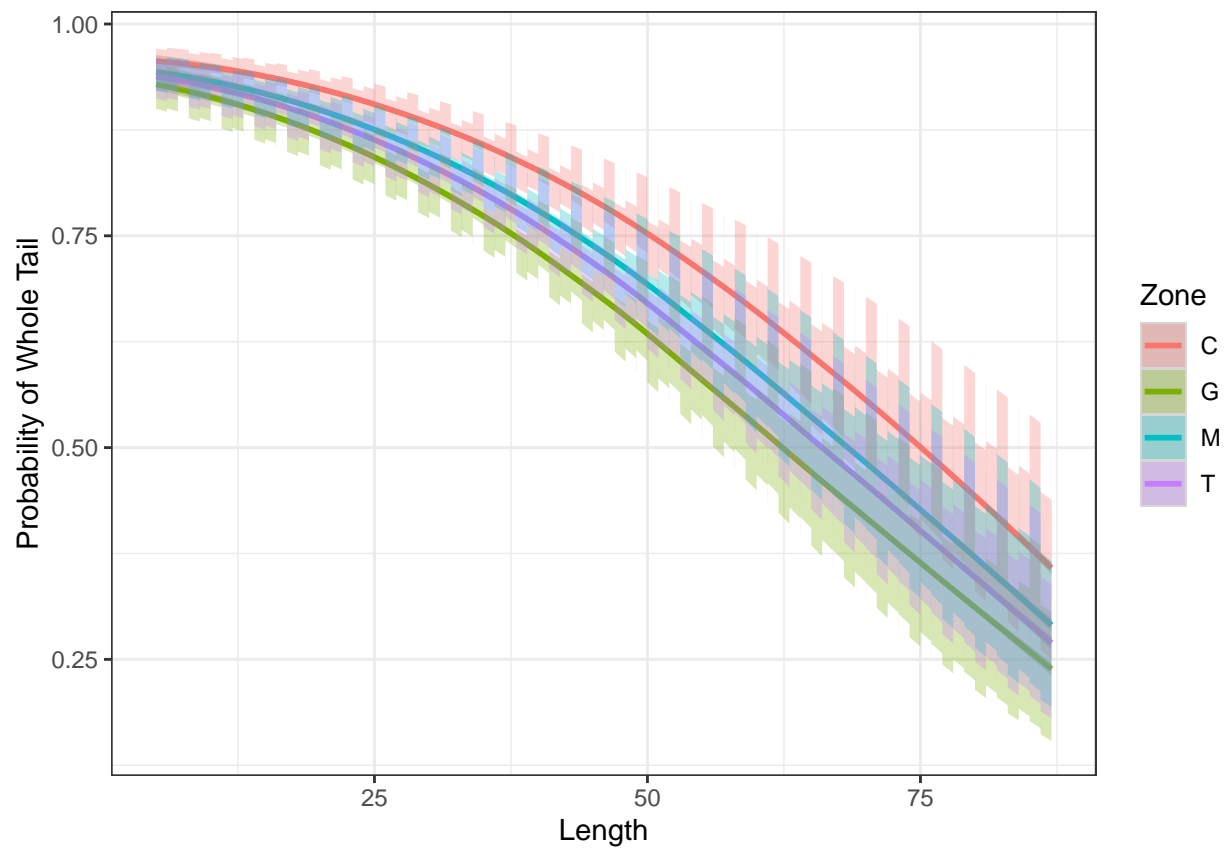



Figure 1: Figure 1. Predictive model of the probability of a lizard tail being whole in relation to snout-to-vent length by vegetation zone.

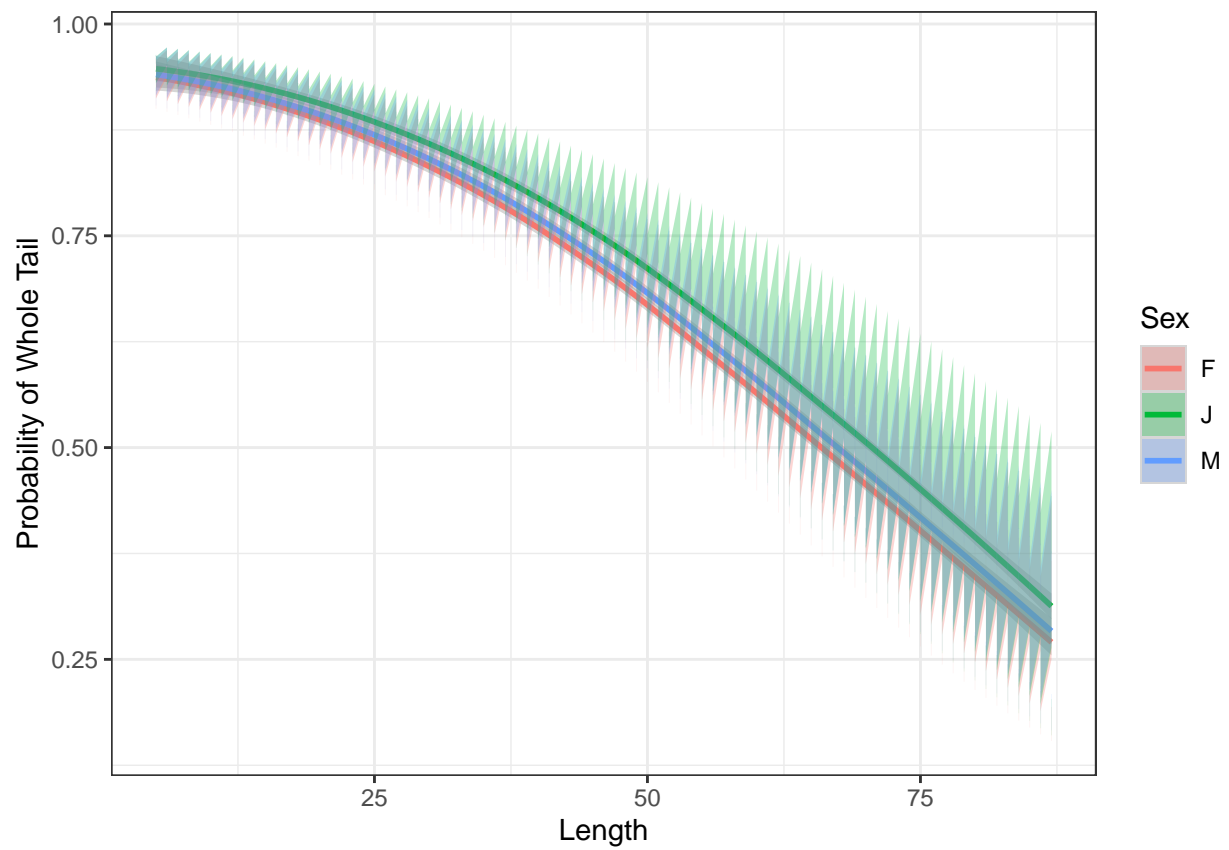
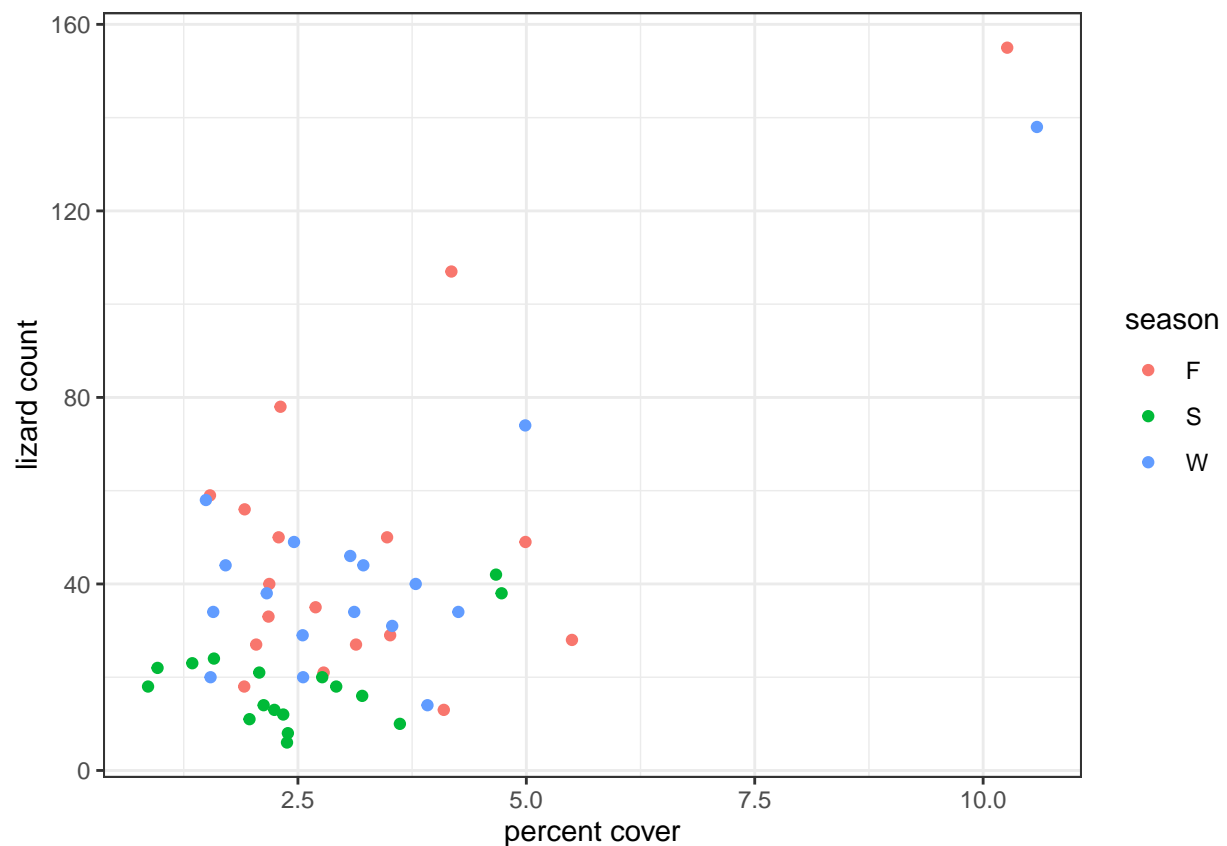


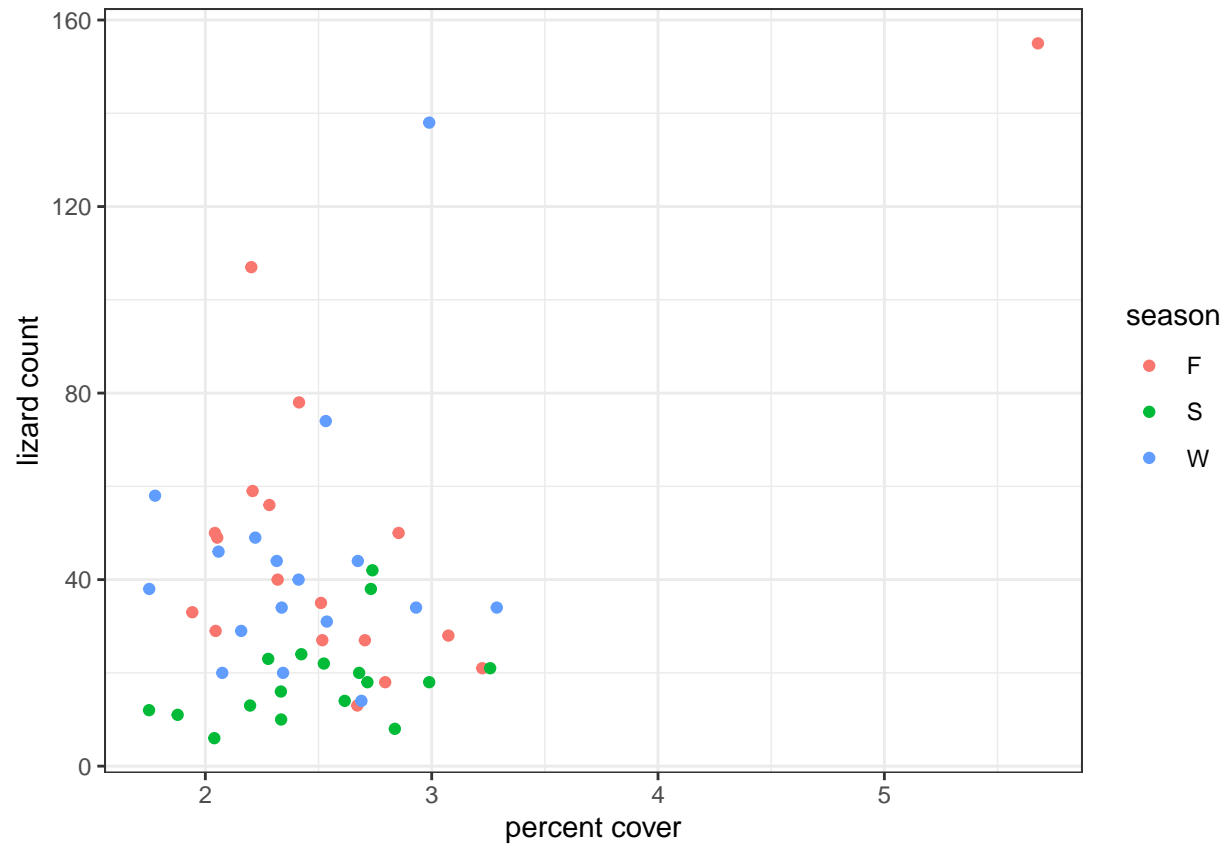
Figure 2: Figure 2. Predictive model of the probability of a lizard tail being whole in relation to snout-to-vent length by sex.

```
## -- Column specification -----
## Delimiter: ","
## chr (1): season
## dbl (6): sample_year, lizard_count, BOER, LATR, PRGL, SCBR
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

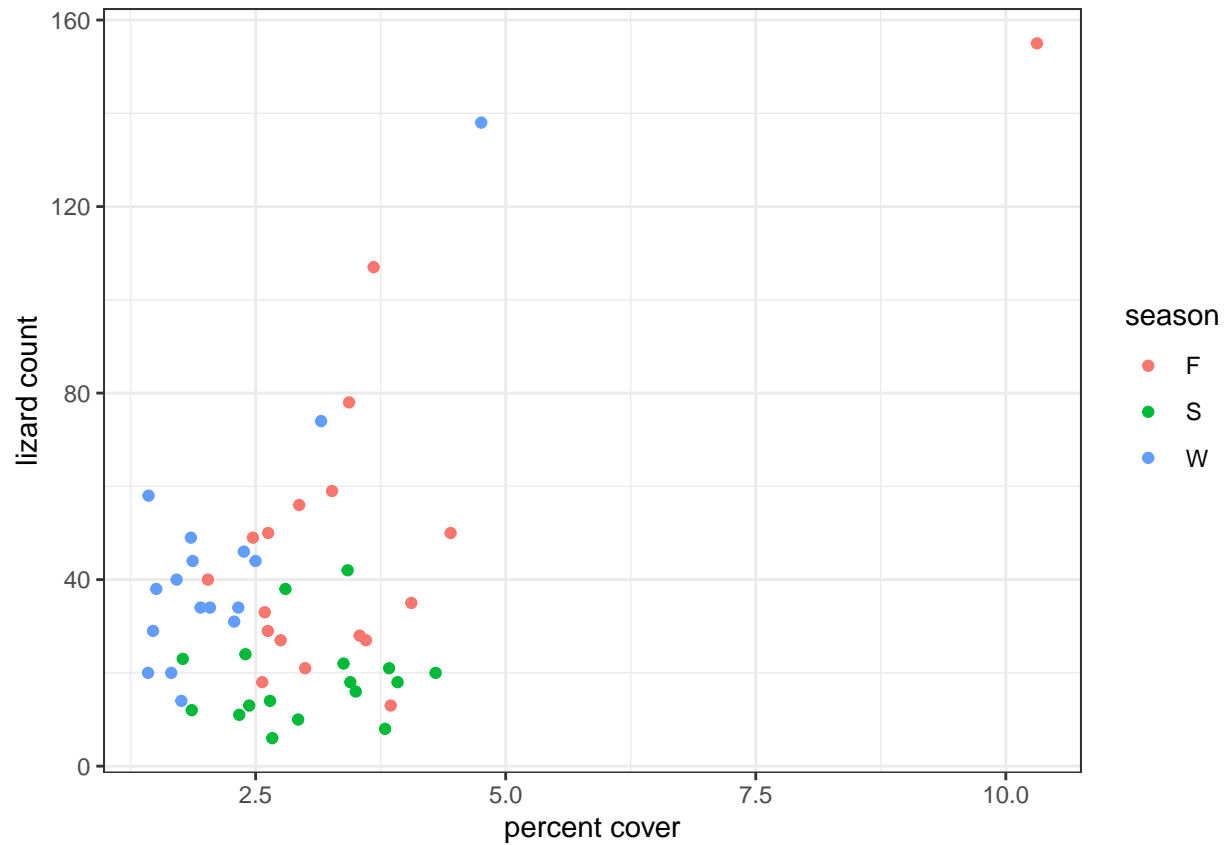
```
ggplot(lizards2, aes(x = BOER,
                     y = lizard_count)) +
  geom_point(aes(color = season)) +
  labs(x = "percent cover", y = "lizard count") +
  theme_bw()
```



```
ggplot(lizards2, aes(x = LATR,
                     y = lizard_count)) +
  geom_point(aes(color = season)) +
  labs(x = "percent cover", y = "lizard count") +
  theme_bw()
```

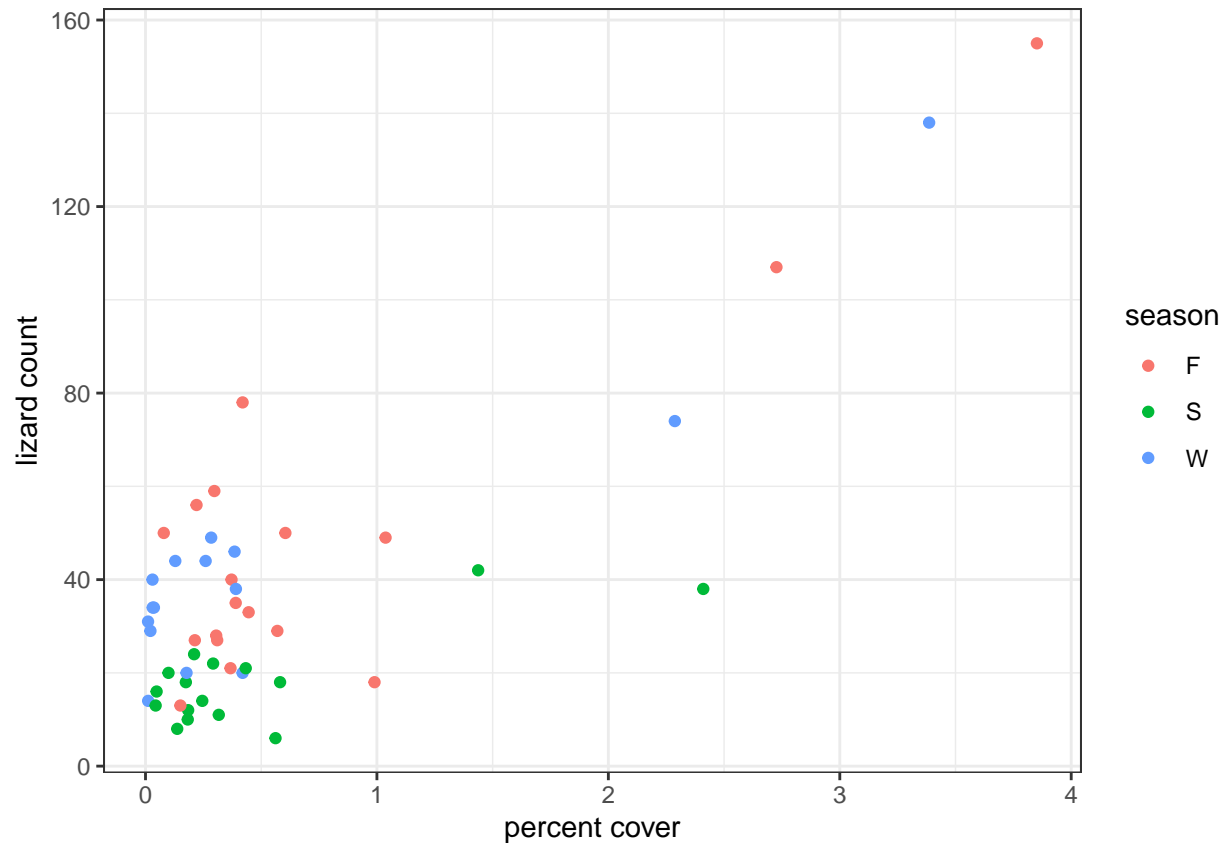


```
ggplot(lizards2, aes(x = PRGL,  
                     y = lizard_count)) +  
  geom_point(aes(color = season)) +  
  labs(x = "percent cover", y = "lizard count") +  
  theme_bw()
```



```
ggplot(lizards2, aes(x = SCBR,
                     y = lizard_count)) +
  geom_point(aes(color = season)) +
  labs(x = "percent cover", y = "lizard count") +
  theme_bw()
```

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



Poisson Regression

```
unique(lizards2$season)
```

```
## [1] "F" "S" "W"
```

```
lizards2$season <- factor(lizards2$season,
                          levels = c("F",
                                      "S",
                                      "W"))

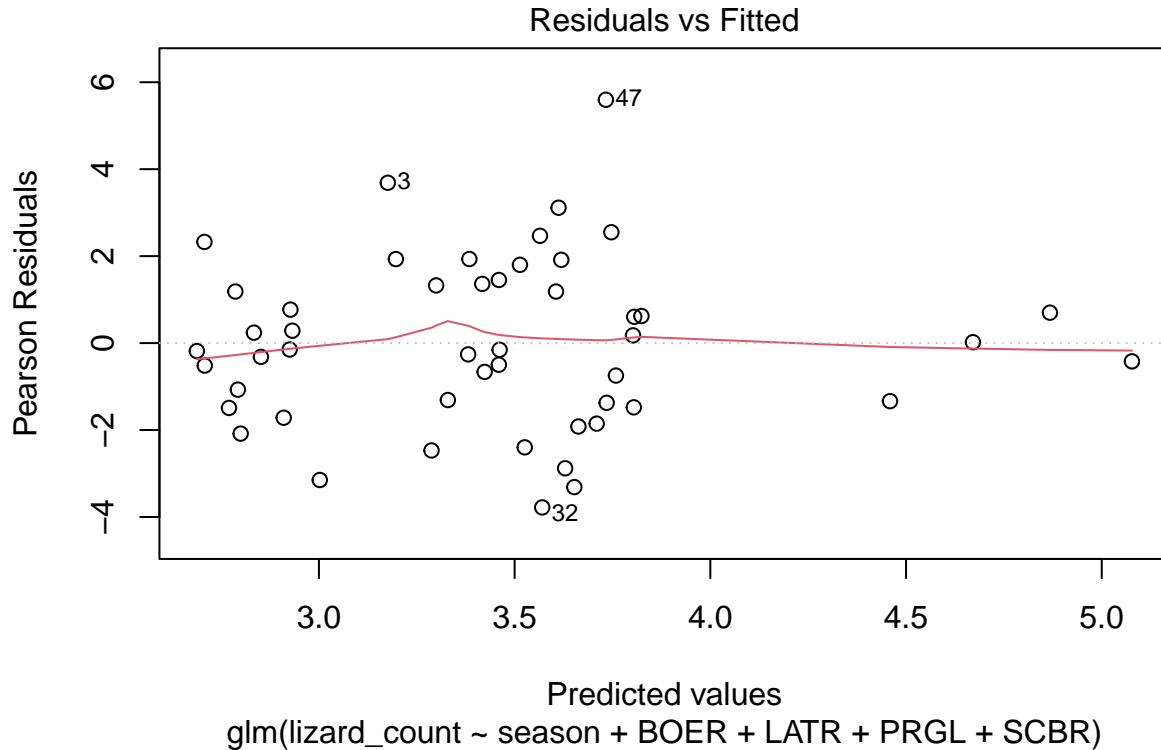
lizard2_mod <- glm(lizard_count~season+BOER+LATR+PRGL+SCBR,
                  data = lizards2,
                  family = "poisson")
```

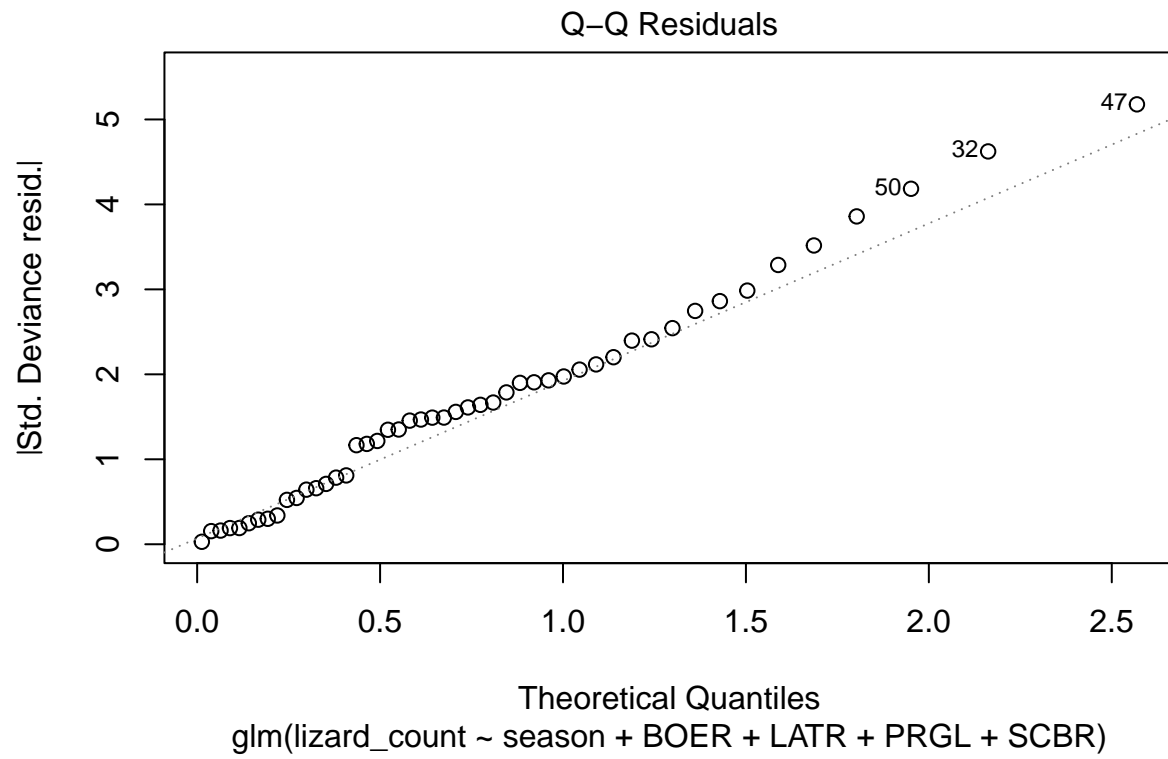
```
summary(lizard2_mod)
```

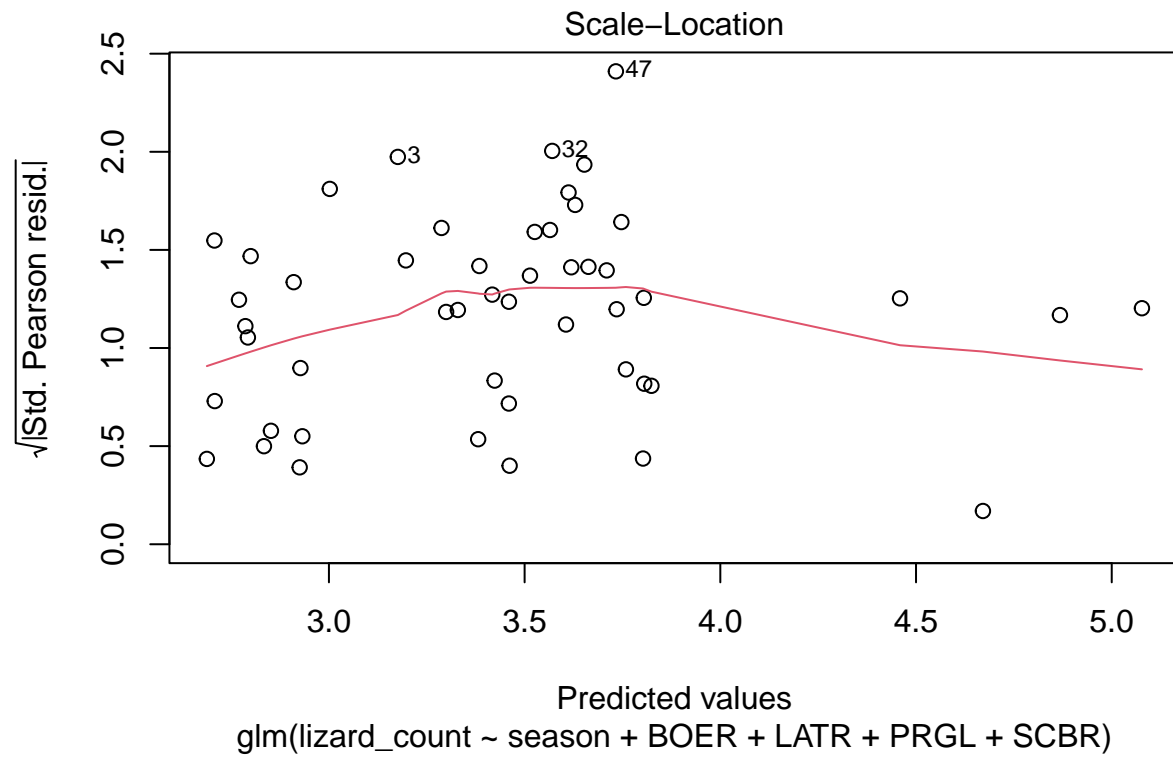
```
##
## Call:
## glm(formula = lizard_count ~ season + BOER + LATR + PRGL + SCBR,
##      family = "poisson", data = lizards2)
##
```

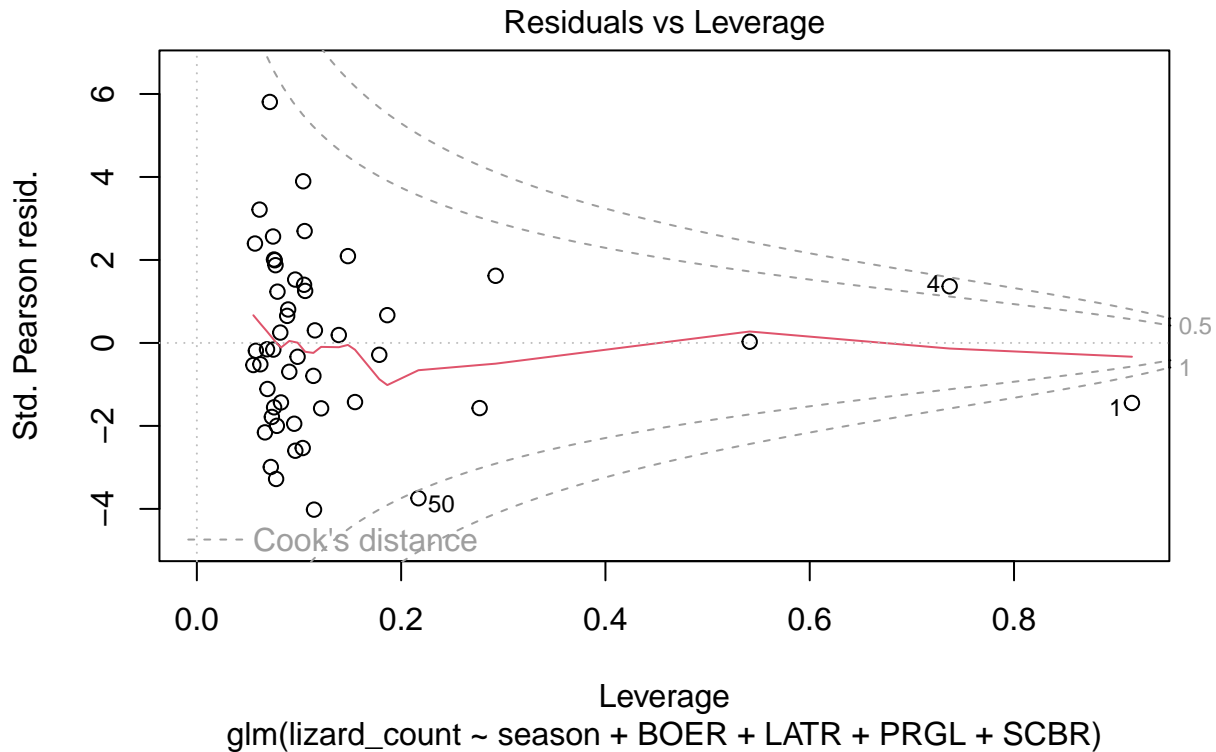
```
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.76785    0.09880  38.136 < 2e-16 ***
## seasonS      -0.75234    0.06970 -10.795 < 2e-16 ***
## seasonW       0.20278    0.07464   2.717 0.00659 **
## BOER         -0.03260    0.02160  -1.509 0.13120
## LATR         -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
```

```
plot(lizard2_mod)
```









```
lizards2_2 <- lizards2[-c(1,4,50),]

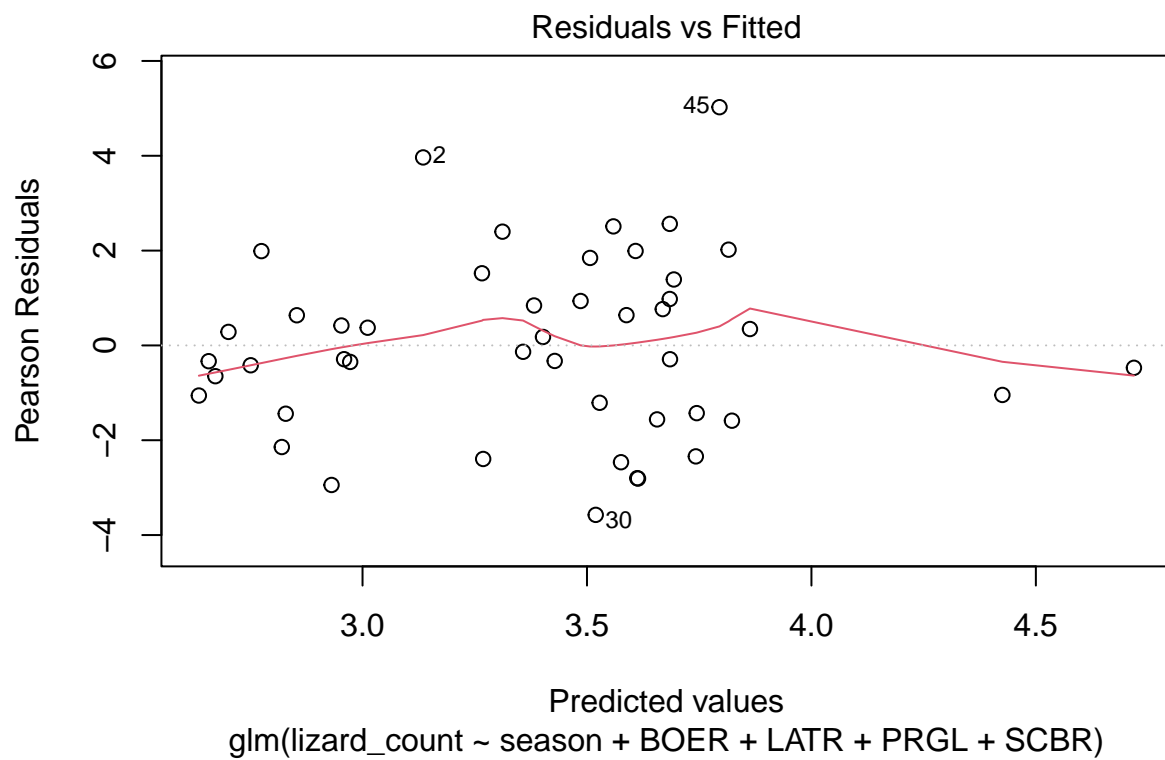
lizard2_mod_2 <- glm(lizard_count~season+BOER+LATR+PRGL+SCBR,
  data = lizards2_2,
  family = "poisson")

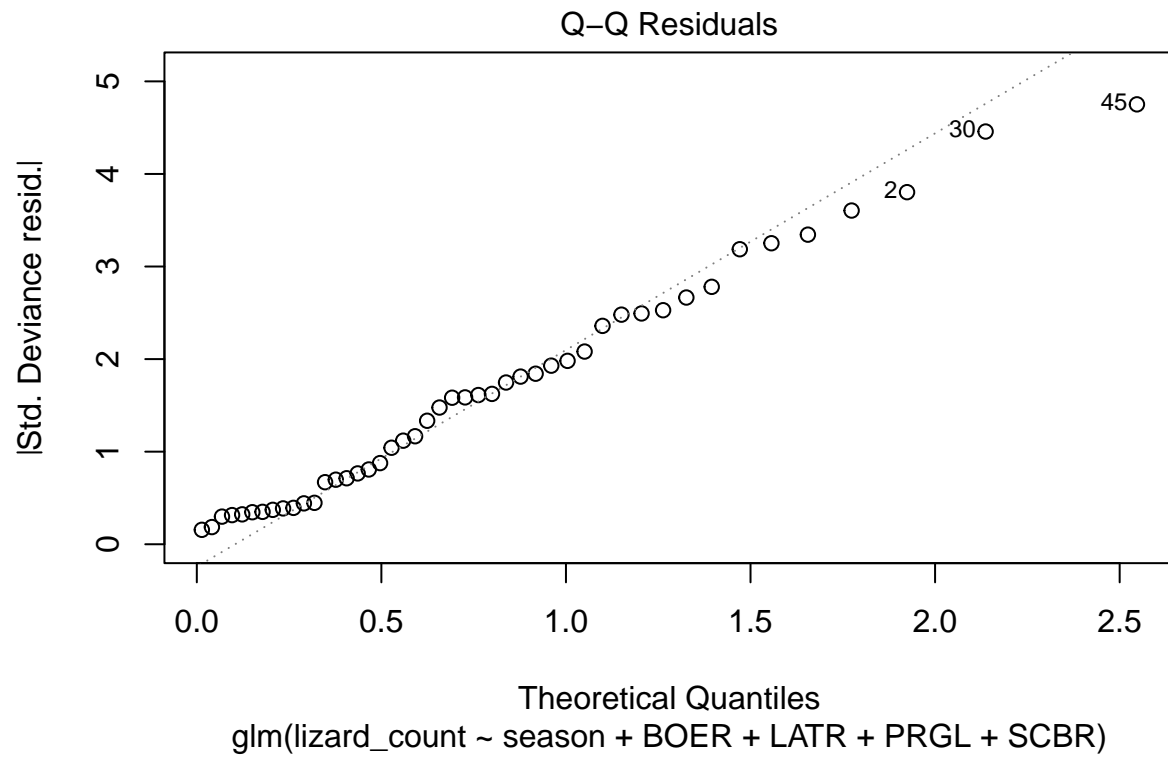
summary(lizard2_mod_2)
```

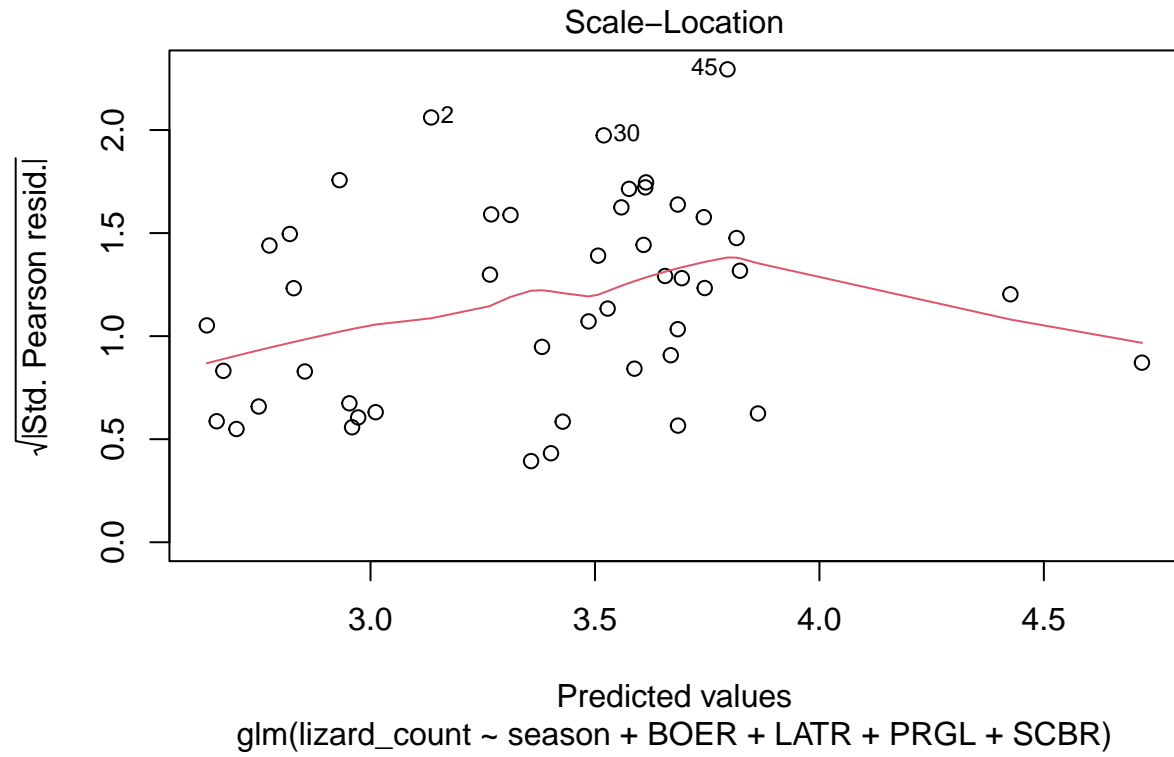
```
##
## Call:
## glm(formula = lizard_count ~ season + BOER + LATR + PRGL + SCBR,
##      family = "poisson", data = lizards2_2)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.58137    0.18748  19.102 < 2e-16 ***
## seasonS      -0.84691    0.07277 -11.638 < 2e-16 ***
## seasonW       0.10419    0.08321   1.252  0.21056
## BOER         -0.10478    0.03249  -3.225  0.00126 **
## LATR         -0.10037    0.09565  -1.049  0.29398
## PRGL          0.14701    0.05450   2.697  0.00699 **
## SCBR          0.46039    0.04565  10.085 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

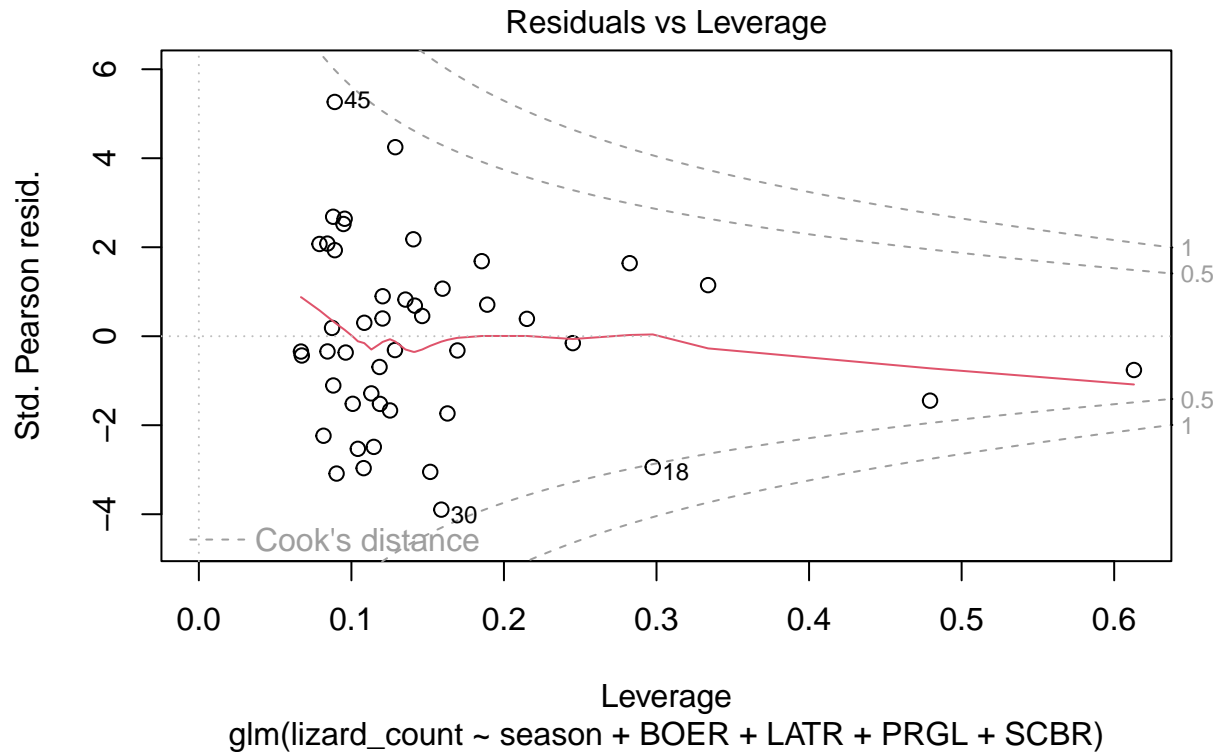
```
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 508.00 on 45 degrees of freedom
## Residual deviance: 160.51 on 39 degrees of freedom
## (3 observations deleted due to missingness)
## AIC: 411.78
##
## Number of Fisher Scoring iterations: 4
```

```
plot(lizard2_mod_2)
```









```
lizards2_3 <- lizards2_2[-c(18),]

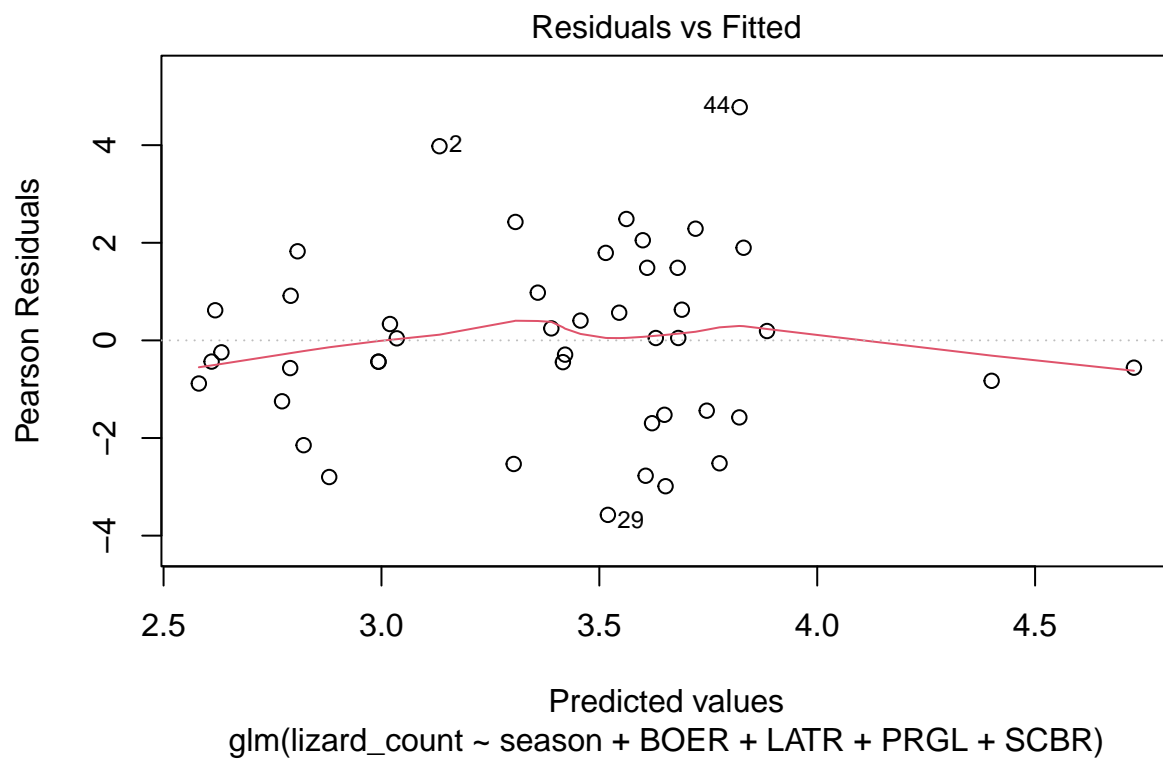
lizard2_mod_3 <- glm(lizard_count~season+BOER+LATR+PRGL+SCBR,
  data=lizards2_3,
  family = "poisson")

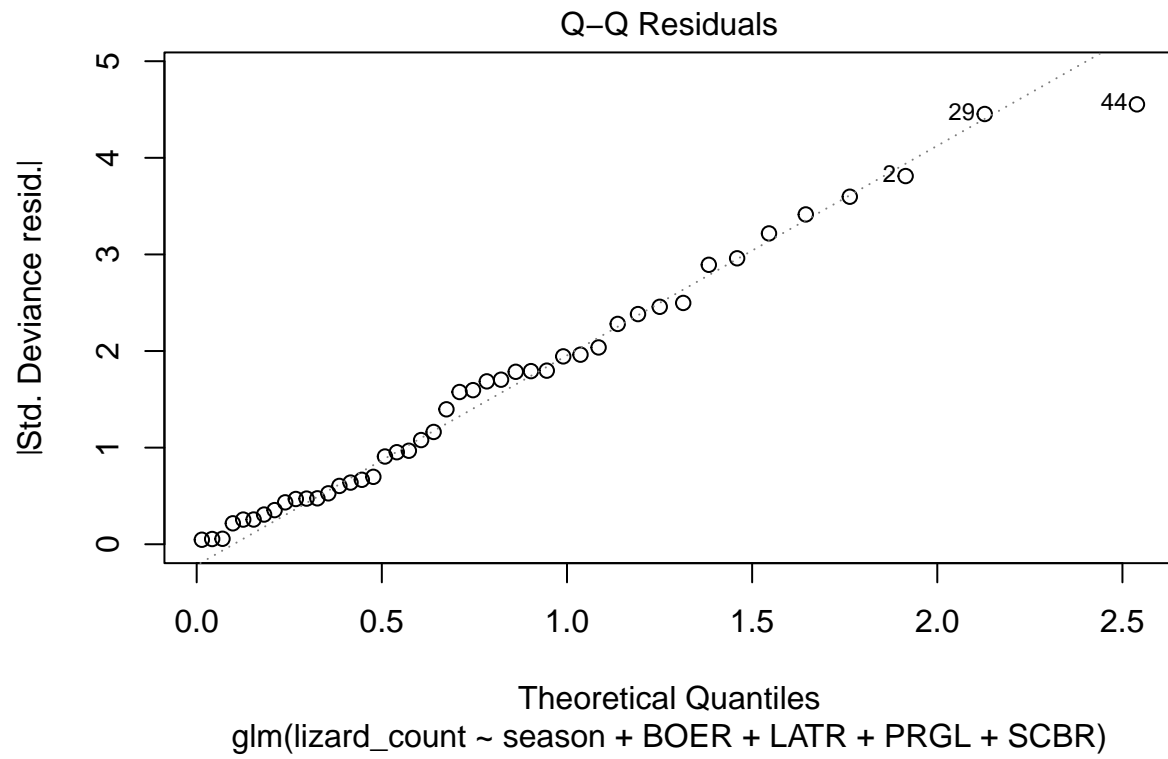
summary(lizard2_mod_3)
```

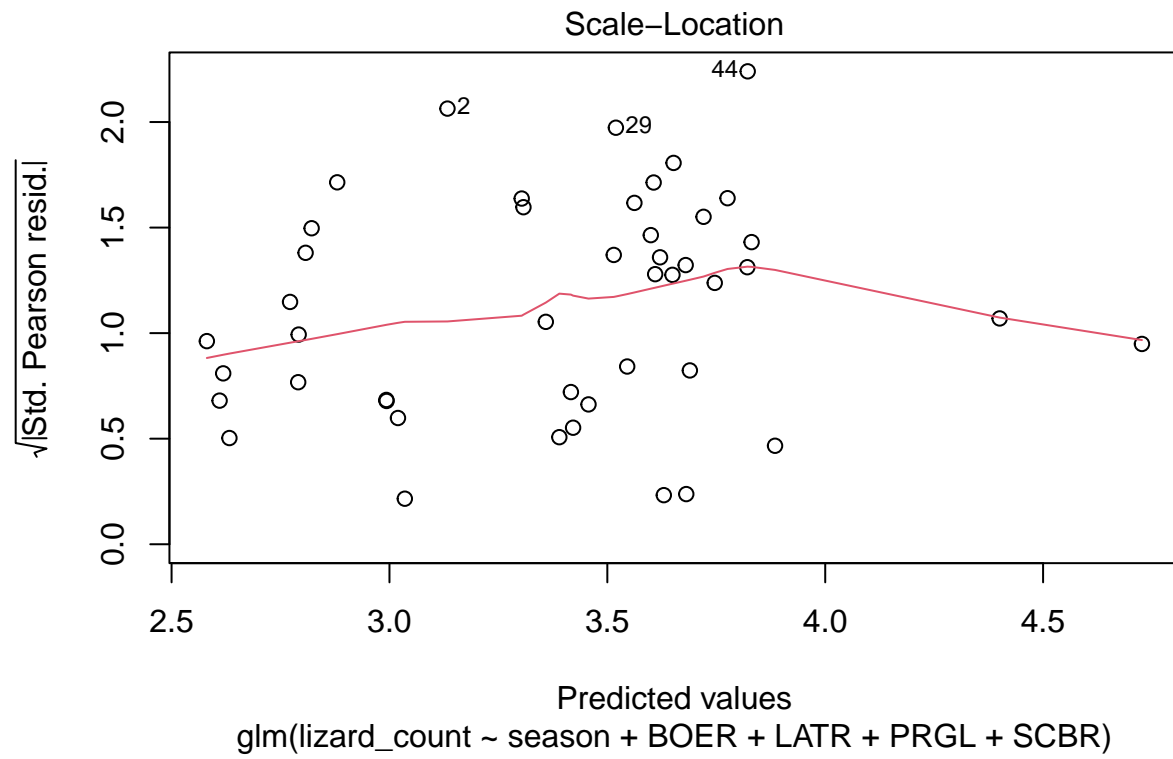
```
##
## Call:
## glm(formula = lizard_count ~ season + BOER + LATR + PRGL + SCBR,
##      family = "poisson", data = lizards2_3)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   3.44840    0.19236  17.927 < 2e-16 ***
## seasonS      -0.91056    0.07600 -11.981 < 2e-16 ***
## seasonW       0.01700    0.08794   0.193 0.846672
## BOER         -0.12535    0.03304  -3.794 0.000148 ***
## LATR          0.06150    0.10972   0.561 0.575104
## PRGL          0.09001    0.05732   1.570 0.116367
## SCBR          0.48990    0.04662  10.509 < 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: 502.98 on 44 degrees of freedom
## Residual deviance: 150.83 on 38 degrees of freedom
## (3 observations deleted due to missingness)
## AIC: 397.22
##
## Number of Fisher Scoring iterations: 4
```

```
plot(lizard2_mod_3)
```

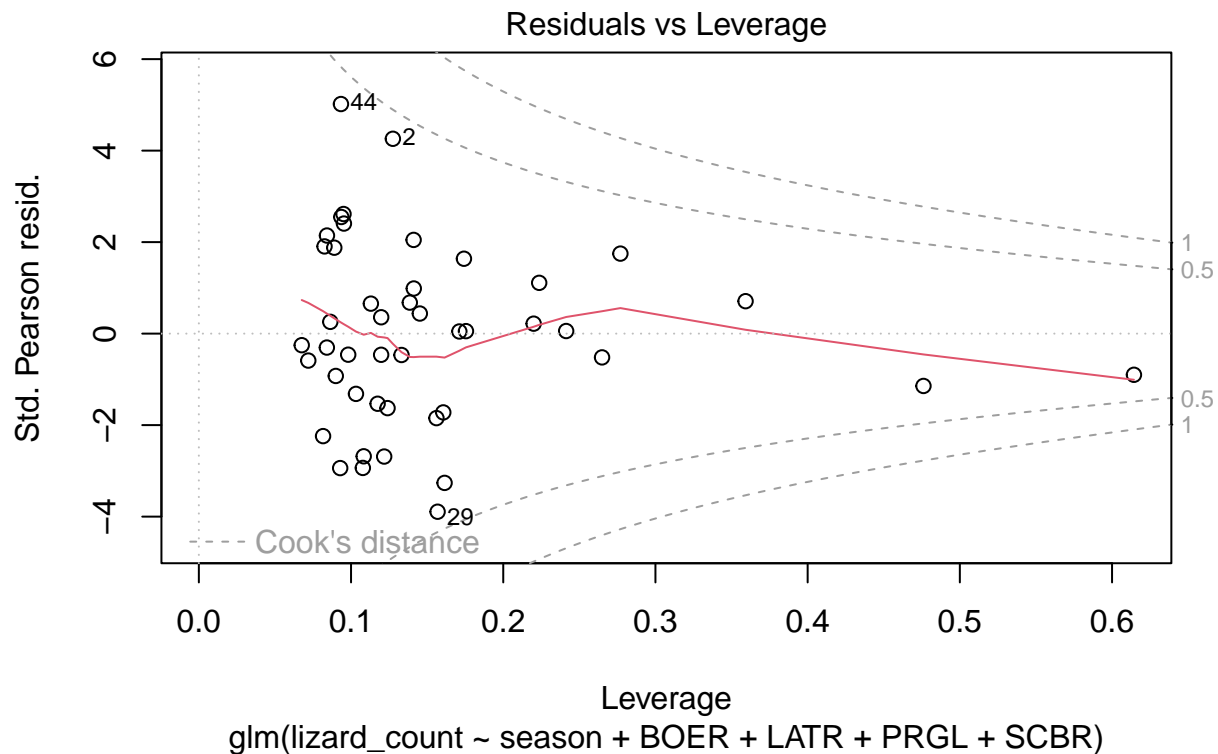






Characteristic	log(IRR)	95% CI	p-value
season			
F	—	—	
S	-0.91	-1.1, -0.76	<0.001
W	0.02	-0.16, 0.19	0.8
BOER	-0.13	-0.19, -0.06	<0.001
LATR	0.06	-0.15, 0.28	0.6
PRGL	0.09	-0.02, 0.20	0.12
SCBR	0.49	0.40, 0.58	<0.001

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



```
tbl_regression(lizard2_mod_3)
```

I removed a total of four outliers that were beyond Cook's distance.

The results of a logistic regression suggest that the log-odds of having higher lizard counts increase with burrograss vegetation ($B = 0.490$, $p < 0.001$). The log-odds of higher lizard counts decrease with increased black grama grass vegetation ($B = -0.125$, $p < 0.001$). There was a significant decrease in the log-odds of having higher lizard counts in the summer ($B = -0.911$, $p < 0.001$) relative to the fall. Neither increased creosote bush ($B = 0.062$, $p = 0.575$) nor honey mesquite ($B = 0.090$, $p = 0.116$) had significant log-odds of having higher lizard counts. Winter ($B = 0.017$, $p = 0.847$) did not have significant log-odds of having higher lizard counts relative to fall.