INSTRUCTOR VERISON OF QUBES LESSON:

Spotting Interactions: *Spotted Salamanders, Beneficial Buffers, and Helpful Hydroperiods*

Prerequisites: Student should have a basic understanding of:

- Regression and ANOVA models
- R and RStudio

NOTE: This lesson contains two separates R (.Rmd) files, one for the instructor and one for the students. There is an optional activity at the end where students will need to fill in R codes based on prompts within the R and interpret the statistical output. Similar codes are given in the lesson but you may refer to 'InstructorR-QUBESBuffAmphs.Rmd' for the working codes to examine the same questions for wood frogs.

Lesson Outline:

This lesson should take 30-45 minutes to complete with extra time needed to read the primary literature, answer questions and to fill in the optional additional assignment of analyzing the data for wood frogs.

PART 1: Experiment and Lesson

- Students and instructor should have read the focal paper before the beginning of the lesson
- Present the student lesson (Lesson-QUBESBuffAmphs.pdf) in class

PART 2: Data analysis in R

- Check the data for normality
- Run the generalized linear model (GLM) without the Poisson distribution
- Look at Fig. 5 or the recreation available in R using a linear model (Note this graph gives up the output without Poisson)
- Next, run the GLM with Poisson distribution
- Make the GLM graph with Poisson
- Check models to see if the Poisson distribution works better with our dataset

PART 3: Questions and additional analysis of Frog Data

- Answer the discussion question either as a group or assign for homework
- Optional: Have students fill in R code for the wood frog data, this is great practice for students who may be newer to R.

By Liz Buikema
Spring 2022
Virginia Commonwealth University
Center for Environmental Studies

Spotting Interactions: *Spotted Salamanders, Beneficial Buffers, and Helpful Hydroperiods*

Lesson Purpose

The purpose of this lesson is to help us better understand how we are able to conserve wetland-dependent wildlife. We will be exploring, manipulating, and analyzing data in R using a Poisson distributed generalize linear model (GLM) with interacting effects. The data consists of two pool-breeding amphibians and how their populations respond to the hydrology and surrounding terrestrial habitats.

Ambystoma maculatum

Prerequisites

This lesson assumes that you have:

(1) Read the article by Powell and Babbitt, 2015.

(2) A basic understanding of linear models (Regression and ANOVA).

For refreshers with these see: Linear Regression and ANOVA.

(3) Basic skills in R with software readily installed.

Pool-Breeding Amphibians

Globally, habitat loss and destruction due to human activities is frequently cited as a major contributor to species decline. According to the IUCN red list of endangered species, 41% of amphibians are threatened with extinction. Amphibians have complex life histories that require the use of aquatic and terrestrial habitats at various life stages.

There are a variety of aquatic habitats used by amphibians, like those who require the use of vernal pools (also referred to as ephemeral pools or ponds) or other lentic (slow or motionless) waters. These pools are unique wetlands that do not harbor any fishes which may prey on egg masses, which is essential for pool-breeding amphibians. In addition to these wetlands, the areas surrounding them are used for the amphibian's terrestrial life phase, many migrate back to their natal pools once they have reached reproductive age.

We will be getting to know the spotted salamander (*Ambystoma maculatum*) and wood frog (*Lithobates sylvaticus*) both of which require the use of vernal pools and their surrounding areas for population maintenance and growth.

Lithobates sylvaticus

Below you will find a summary of the life histories for each of these organisms, notice the number of eggs that females will lay during each breeding event and the amount of time that they require aquatic habitats (hatching time through metamorphosis). The range in time for hatching and metamorphosis depends on the amount of heat, light, and water inundation length. If a pool does not hold water for extended lengths of time, metamorphosis time can increase, but this results in smaller organisms that may have weakened immune systems.

Species	Max. Life	Clutch Size	Hatching Time	Metamorphosis	Typical Breeding Interval
Wood Frog	6 years	1000 - 3000 eggs	9 - 30 days	2 - 4.5 months	Once
Spotted Salamander	32 years	100 - 370 eggs	28 - 53 days	2 - 4 months	Multiple times

Question 1: Based on the information in this table, what inferences can you make about these two species and their fecundity (potential for reproduction)?

- *By looking at this table we can see that the wood frog is capable of laying egg masses upwards of 10x larger than those of the spotted salamander. This indicates that the wood frog's population can rebound faster than the salamander's, thus having higher fecundity.*
- *While the spotted salamander has a longer maximum life and the ability to breed more than once, the number of eggs in each mass as well as the amount of time from hatching to metamorphosis can be less advantageous to quick population recovery if they are subject to catastrophic events, thus having less fecundity than the wood frog.*

Vernal Pools

Vernal pools are small temporary wetlands that come and go depending on local climatic elements. Some important aspects of vernal pools are their lack of fresh groundwater sources and their drying ability, this ensures that there are no fish inhabitants. This wetland can be found all over the world in a variety of biomes.

For this lesson, we will be focusing on vernal pools found in the northeastern United States.

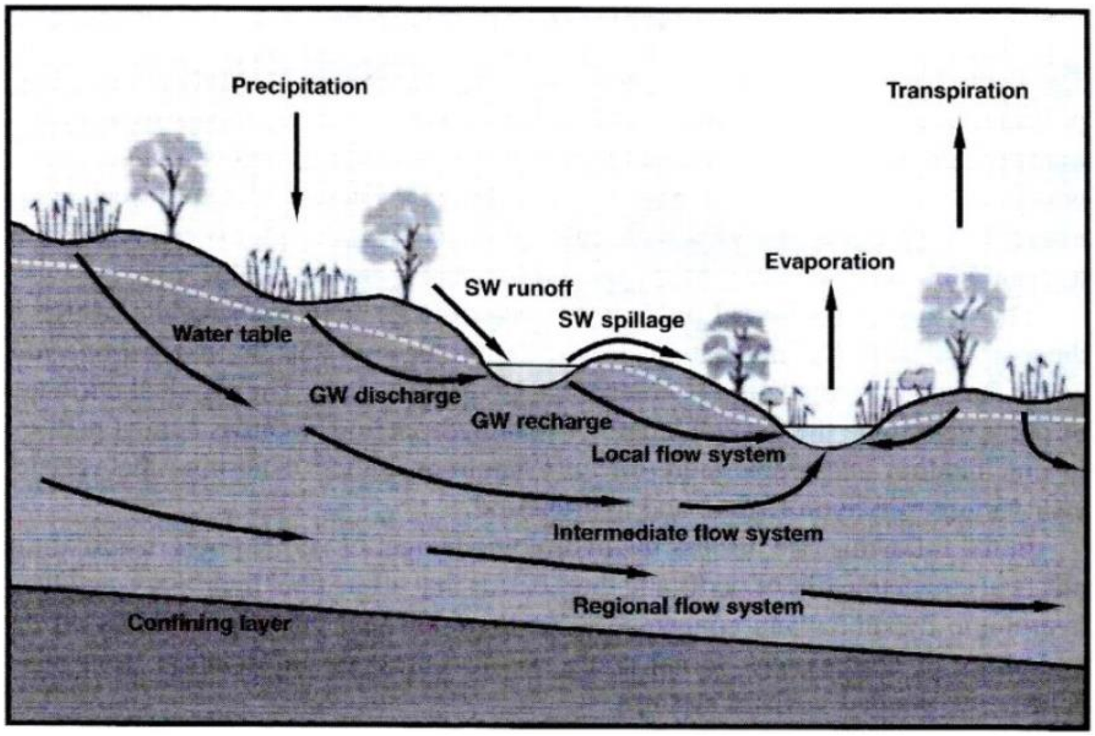
For more information on Vernal pools, check out the Vernal Pool Association's website.

Vernal Pool located in Virginia

The hydrology of vernal pools is referred to as its hydroregime and is unique to each pool. The period of time that they are inundated with water, via precipitation, is referred to as the hydroperiod, these wetlands can appear for short periods of time or stay for multiple seasons. The hydroperiod takes into account many landscape features like: surrounding vegetation, tree cover, pool size, and climatic events. Drying typically occurs in the summer months when evaporation from heat and evapotranspiration from plants are at their peak.

Vernal pools are important habitats for many obligate species that require them for survival, like the fairy shrimp and our focal species, the spotted salamander and the wood frog.

However, most states in North America do not offer protections to these wetlands, paving the way for them to be filled in and built upon.

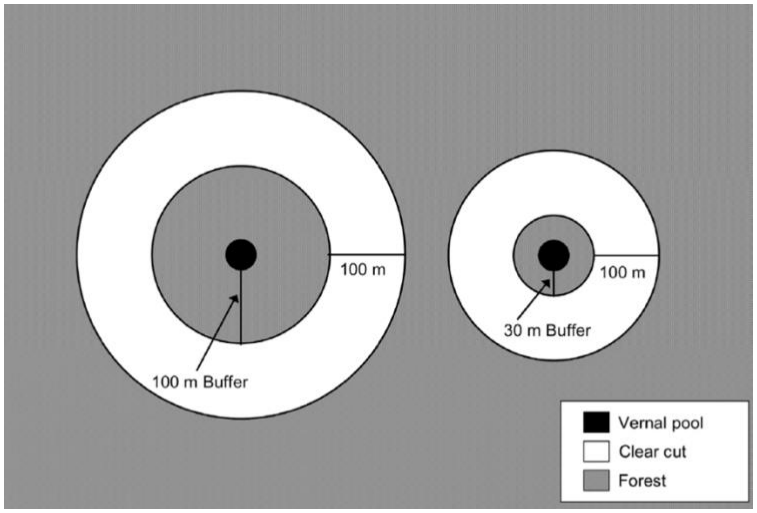


Note. Diagram demonstrating the basics of a vernal pool hydroregime. A. Calhoun, & P. DeMaynadier (Eds.). 2008. *Science and Conservation of Vernal Pools in Northeastern North America*, CRC Press (pp. 34).

Question 2: In regards to the hydroregime, what might you expect to happen to obligate species when there are changes to the climate?

- *Changes to climate can increase and/or decrease the period of water inundation.*
- *An increase in precipitation could result in the pools inability to dry out, possibly leading to fish being established in the pool, which prey upon amphibian egg masses, resulting in a decline.*
- *A decrease in hydroperiod via increased temperatures or less precipitation could cause the amphibians inability to lay eggs, or increase metamorphosis – both of which results in a less robust population.*
- *While not specifically asked in the question, a decrease in buffers around the pool make them more susceptible to contaminants via runoff. Changes in buffer could also cause changes to the vernal pools hydroperiod. More on buffers in the next section.*

Buffers



(Fig 1) Diagram of buffered areas from the paper

Vegetative buffers consist of plants found along the edges of many types of aquatic habitats and wetlands. These buffer zones are wonderful at helping to filter out pollutants from runoff and also help combat erosion in lotic (moving-water) systems like streams and rivers. From a water quality stand point, buffers between 15m-30m are sufficient in maintaining nutrient runoff loads. Bonus, they can also be used as a mitigation technique for wetland wildlife!

Amphibians require the use of both aquatic and terrestrial habitats throughout their lives. After metamorphosis, these amphibians spend much of their lives in the surrounding terrestrial habitats, this is especially true for the spotted salamander who live in burrows surrounding the pool.

Larger buffers can support greater populations of breeding adults. Some amphibians have been known to migrate up to 290m away from their natal pools and use these buffered areas to help maintain their population. Loss of buffers also results in a loss of genes for these species. In the US these regions are protected up to 30 meters through the Clean Water Act. However, this does not apply to vernal pools in many areas.

Why does this matter?

The ability for animals to move through patches of habitat is important to their ability to find mates outside of their local population. Increased mating within a local population, especially a small one, reduces their genetic diversity. A reduction in genetic diversity increases a species likelihood of becoming extinct.

Successful amphibian conservation would increase connectivity (through reduced destruction and alterations) that would include not only the pools they use for breeding and metamorphosis but also for the terrestrial areas surrounding the pools. Since amphibians require both aquatic and terrestrial habitats, this would need to be an overarching effort

These patches of habitat are important, but the hydroperiods are increasingly important as well. While a large buffer can help increase the populations success, longer periods of inundation further increase success. Resources are used up more quickly with more population abundance; this can cause slower metamorphosis due to less resource availability. However, when the hydroperiod is increased, slower metamorphosis can successfully take place minimizing the impacts of large populations.

A spotted salamander over its egg mass surrounded by fairy shrimp

Question 3: Given that we now know amphibians can migrate upwards of 290m from their natal waterbody, what effects do you think having buffers of 15-30m will have on their populations?

Since many amphibians migrate away from their native pools, the increase in habitat fragmentation by these small buffers can cause a decline in populations due to smaller mate selection resulting in more inbreeding. Less genetic diversity would cause the local populations to be less adaptable to change and increase extinction rates.

The Statistics: An Introduction

Within this study, the authors use generalized linear mixed modeling (GLMM) to run their statistics, which is an extension of generalized linear model (GLM) that includes random mixed effect in addition to fixed effects. The statistics ran by the authors include the interactions among the year, treatment, sex ratios, recapture rates, average hydroperiod, and standard deviation of the hydroperiod.

Our approach for this lesson, will focus on examining if changes in buffer and hydroperiod have an interacting effect on the number of breeding adult salamanders. Since we will be working with count data that lacks a normal distribution, an assumption of many statistical analyses, the Poisson distribution will be used in R to help us accurately portray the data. We will learn about statistical interactions, how to interpret the R output, and how to build a graphical model with the Poisson.

Fairy Shrimp, an obligate species to vernal pools

The Statistics: ANCOVA

Assumptions of ANCOVA:

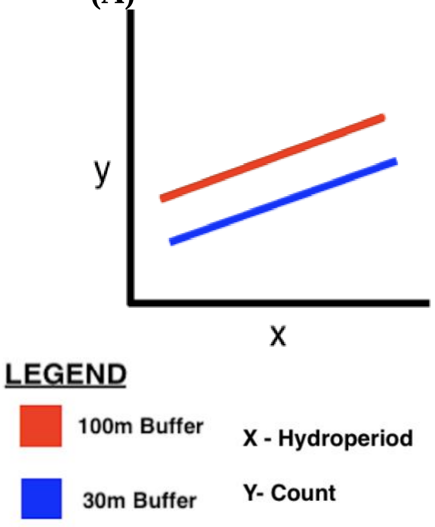
- Independent observations
- Normal distribution of data
- Linear relationship
- Homogeneity of data and regression slopes

We will be using some concepts from analysis of covariance (ANCOVA) and apply it to the GLM with an interaction. An ANCOVA is essentially an extension of both regression analysis and one-way ANOVA. It seeks to tells us if there is a significant difference in the *slopes* of different treatment groups; or how the independent variable acts based on the dependent variables. An ANCOVA can help to remove covariates (variables you do not wish to study) from your statistical model and seeks to tell you if there is an interaction based on your categorical and continuous variables.

The Statistics: Statistical Interaction

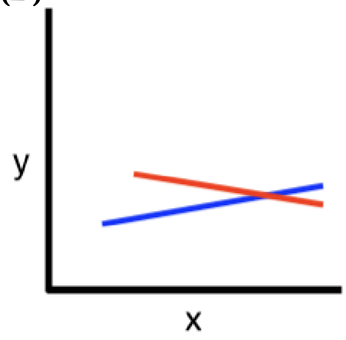
In linear modeling, two or more explanatory variables interact if changes to one or both of these variables also changes the value of the response variable. For this lesson, the explanatory variables are mean hydroperiod (a continuous variable) and buffer size (a categorical variable), the response variable is the number of breeding adults. Below are examples of graphical outputs that show no interaction (A) and an interaction (B) of our response variables.

(A)



Interpretation of graph (A):
Here we can see that as hydroperiod (X) increases, so does the count (Y) for both buffer treatments (blue and red lines). These lines have different intercepts and are parallel; the slopes are the same. Since these slopes are not different from each other, there is no statistical interaction taking place. The treatment does not have an effect, hydroperiod does.

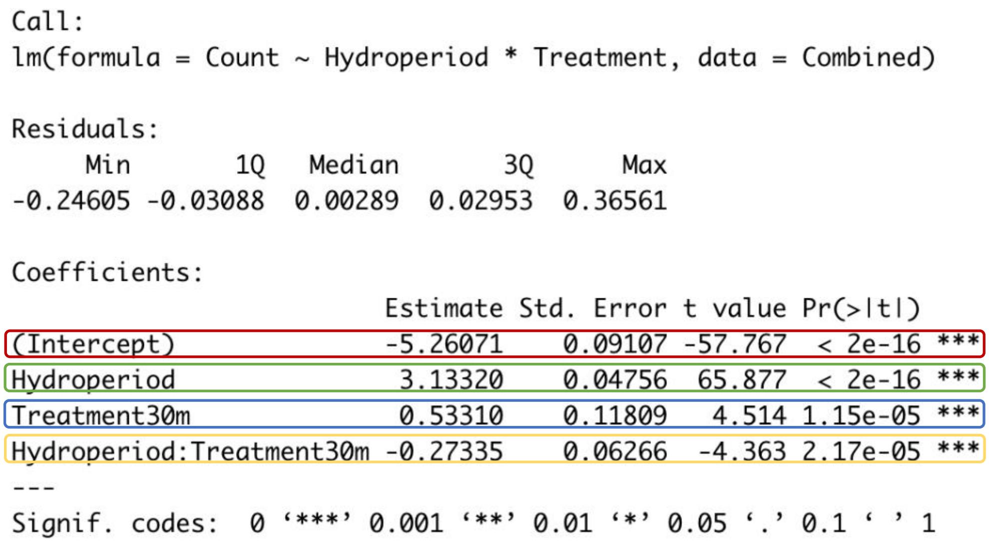
(B)



Interpretation of graph (B):
For the 30m buffer (blue line), we can see that as hydroperiod (X) increases, so does the count (Y). For the 100m buffer (red line), as hydroperiod increases, count decreases. These different slopes indicate that the treatments (30m/100m) and hydroperiod have an interacting effect on count. When looking at both of the treatments, it is obvious that the slopes are not the same. However, we cannot just state that the slopes are different from each other, we must see if these differences are statistically significant.

Output Example from R and How to Read it

We will run the GLM in R to help us determine if the slopes are statistically different from each other. Where Count is the Y variable, hydroperiod is the X variable, and treatment interaction is calculated. We need to be careful when interpreting these results, when we have a significant main effect, you also need to consider the interaction.



```
Call:
lm(formula = Count ~ Hydroperiod * Treatment, data = Combined)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.24605 -0.03088  0.00289  0.02953  0.36561 

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)                -5.26071    0.09107 -57.767  < 2e-16 ***
Hydroperiod                 3.13320    0.04756  65.877  < 2e-16 ***
Treatment30m                0.53310    0.11809   4.514 1.15e-05 ***
Hydroperiod:Treatment30m -0.27335    0.06266  -4.363 2.17e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

NOTE: Graphically this data output does not correspond to the above example graph (B), this is an example.

So, what are we looking at?

Red Box *(Intercept)* is the statistical output for *Treatment 100m*, more specifically under *Estimate* this is the intercept of the line (-5.26). This is the baseline or reference. R chooses the reference alphanumerically.

Green Box *Hydroperiod* is the slope of *Treatment 100m* under *Estimate*. *Treatment 100m* (our reference) increases 3.133 for each unit of *Hydroperiod*.

Blue Box *Treatment30m* under *Estimate* is the statistical output for change in the intercept of *Treatment30m* compared to the reference. To get this output you must add -5.26071 + 0.53310 = -4.72761; this is the intercept of *Treatment30m*.

Yellow Box *Hydroperiod:Treatment30m* under *Estimate* is the statistical output for change in the slope of *Treatment30m*. To get this output you must add 3.13320 + (-0.27335) = 2.85985; this is the slope of *Treatment30m*. This line also tells us if the slopes of our treatments are significantly different or not (note the colon [**:**] in the name), under the p-value column (*Pr(>|t|)*.

-Here we see a very small p-value, these slopes are significantly different from each other!

Question 4: How do we know if an interaction is occurring with our data?

We know that there is an interaction occurring when: (1) we have two or more predicting variables, (2) the slopes of the response variable lines are not equivalent, and (3) the difference in these slopes were found to be statistically significant.

The Statistics: Poisson Distribution

To learn more about the Poisson distribution formula and mathematical appliance go here.

The Poisson distribution is used in cases when you have data that is not normally distributed. It is appropriate to use a Poisson probability distribution when:

(1) observations consist of counts. *

(2) counts are obtained from defined units, like time intervals, and can be put into frequency distributions. **

(3) fewer objects being counted* are in the sampling unit** than it could contain, therefore are considered *rare*.

(4) counts are independent of each other, random.

The Statistics: R code and Analysis

Our Question: *Is an interaction occurring between the buffer treatments and the mean hydroperiod of these pools on the number of breeding adult salamanders?*

From above we know that we will be using count data (*TOTAL.adults*) to examine the average hydroperiod's (*MEAN.hydro*) interaction with the experiment's treatment levels (*30m* or *100m*). According to these parameters, the Poisson distribution is appropriate.

We will be looking at the response of number of breeding adults on the predictors of mean hydroperiod, a continuous variable, and treatment, a categorical variable with 2 levels. Within the paper, the authors ran statistics with 3 treatment levels: reference (uncut), 30m buffer, and 100m buffer. Reference has been eliminated from our data for simplicity.

In R we will be running our model with and without the Poisson distribution to see why it is important to use in our data set.

An Experimental Test of Buffer Utility as a Technique for Managing Pool-Breeding Amphibians

The scientists who wrote this article and designed the study wanted to test buffer size effects on breeding adult spotted salamanders and wood frogs, citing that only observational studies have been done in the past but buffer size specifically had not been measured. This study consisted of 11 different vernal pools, all roughly the same size, in Maine over a six-year period in one of three different experimental buffer groups: reference (uncut), 30m, and 100m. They then examined whether there were changes to the abundance, proportion recaptured, and sex ratios according to their experimental buffers.

Meet the Scientist

Jessica Veysey Powell

Jessica Veysey Powell received her Ph.D. from the University of New Hampshire in wetland ecology and policy. Her research revolves around land use and its effects on biological components of the ecosystem. More information about Dr. Veysey Powell's work can be found at ResearchGate.

Kimberly Babbitt

Kimberly J. Babbitt is an associate dean at the University of New Hampshire with the department of Natural Resources and the Environment. She received her Ph.D. from the University of Florida in Wildlife Ecology and Conservation and has conducted extensive research on buffer utility and amphibians. More information about Dr. Babbitt's work can be found here.

The Data

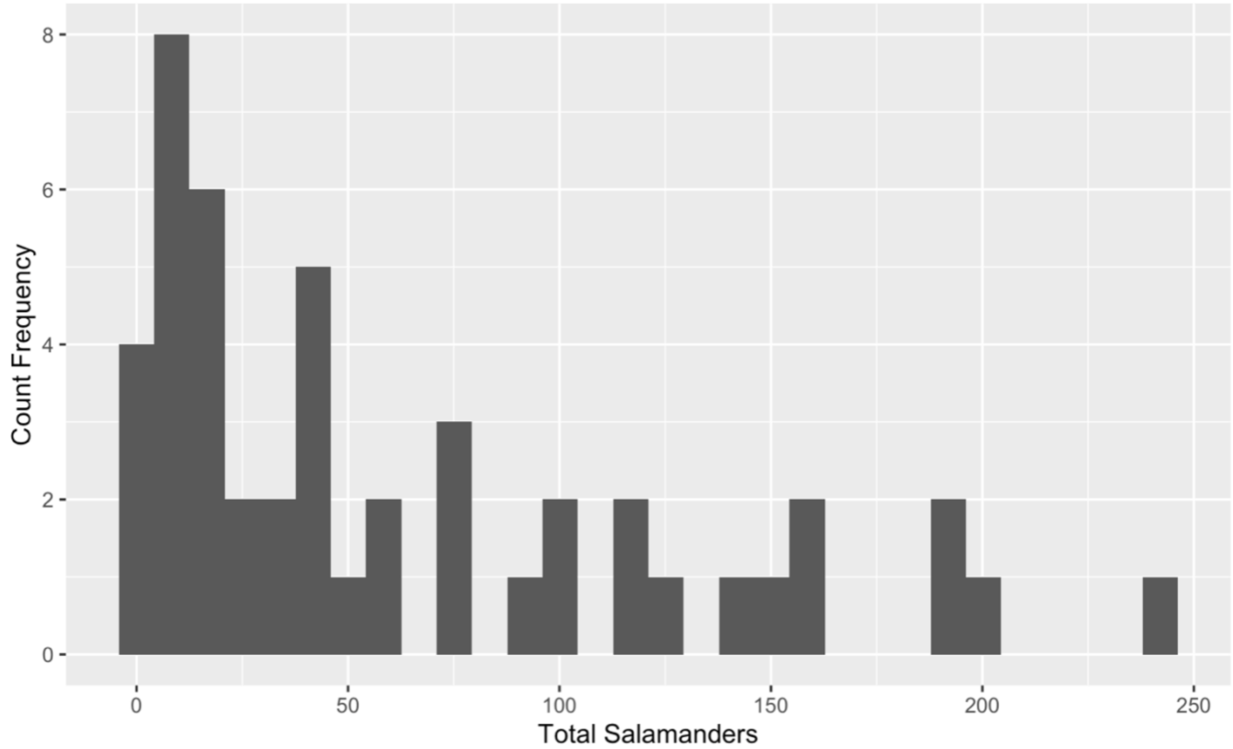
For this QUBES lesson we will be using an edited version of the study data, the original form of this data can be found on Dryad. We will be utilizing the columns of: treatment, species, breeding adults, and mean hydroperiod.

- The treatments have 2 levels, 30m and 100m buffer; reference was removed.
- The species were simplified to say either '*FROG*' for the wood frog or '*SALA'* for the spotted salamander.
- The *TOTAL.adults* column contains the total number of breeding adults captured in a given year, both male and female.
- The mean days that the pool contained water, it's hydroperiod, is depicted in the *MEAN.hydro* column, calculated over the length of the study. According to the 'readme' file provided by the authors on Dryad,

> *"We calculated hydroperiod for each pool in each year as the number of days a pool held water between ice-out (i.e., < 75% of the pool was covered in ice) and the day the pool dried completely. To facilitate statistical analyses, we assigned a hydroperiod end date of October 28th to pools that did not dry in a given year. We chose this end date because we still had evidence of persistent water in these pools on this date, but it was sufficiently late in the year that most amphibians at our study pools were inactive."*

The R Output

First let's visualize our data

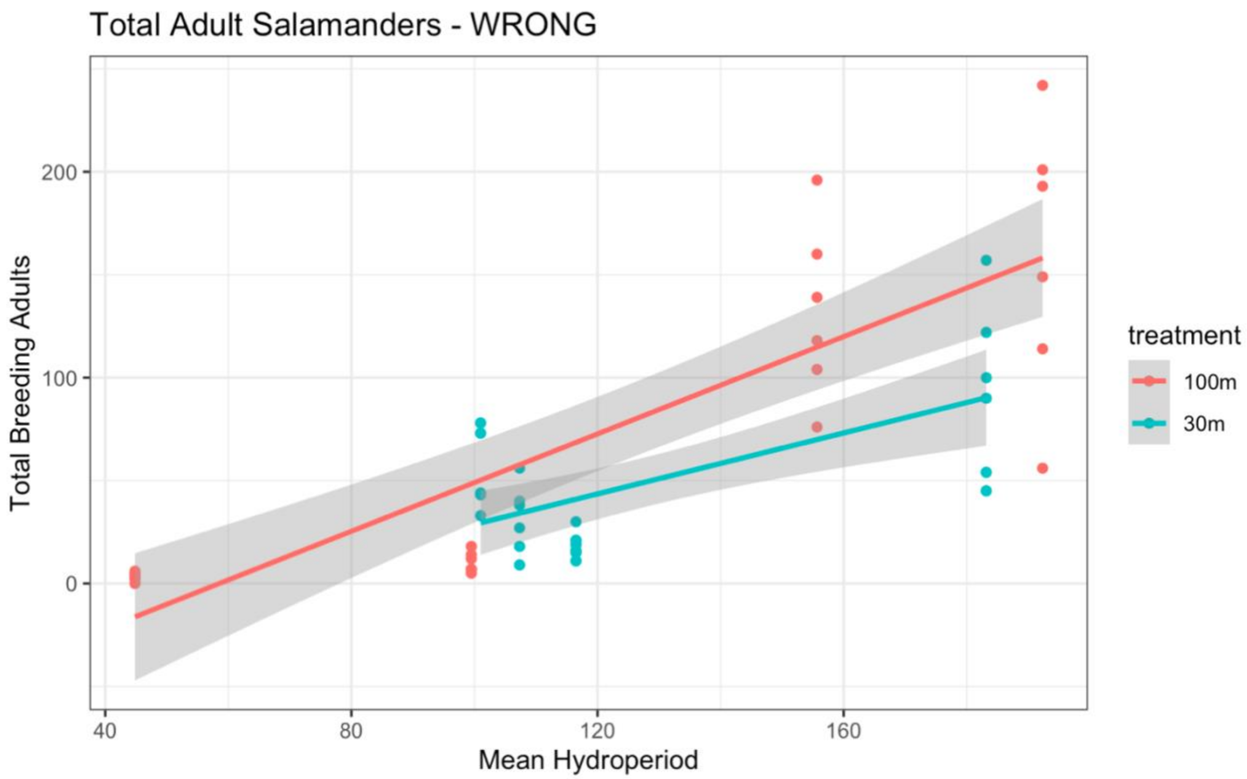


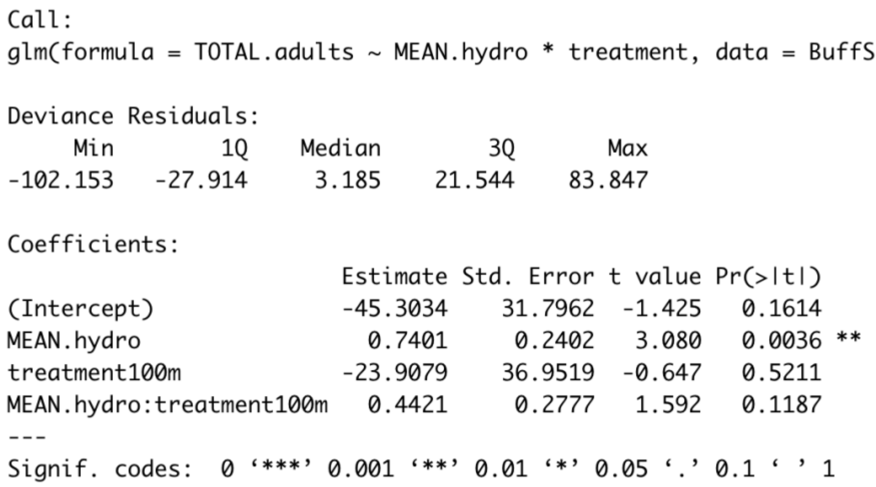
Question 5: Is this data normally distributed? Why or why not?

No, this data does not have a bell-shaped curve which would indicate that we have a normal distribution. Instead, we see that this data is skewed to the left with smaller counts being dominate. This is not a surprise, as we are working with count data (number of salamanders). Since this data is not normally distributed, we will need to use the Poisson distribution.

<u>GLM without Poisson</u>

Let do this the wrong way to see how it can be improved





```
Call:
glm(formula = TOTAL.adults ~ MEAN.hydro * treatment, data = BuffSala)

Deviance Residuals: 
     Min        1Q    Median        3Q       Max  
-102.153   -27.914     3.185    21.544    83.847  

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)   
(Intercept)              -45.3034    31.7962  -1.425   0.1614   
MEAN.hydro                 0.7401     0.2402   3.080   0.0036 **
treatment100m            -23.9079    36.9519  -0.647   0.5211   
MEAN.hydro:treatment100m   0.4421     0.2777   1.592   0.1187   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Question 6: Based on the graph and its corresponding R output what can we say about the data?

- *Visually it appears that our lines are not parallel, but to what degree?*

Let’s do some digging…

- *We will examine our residuals, remember we want our median to be close to 0, with the minimum, 1Q (quadrant), 3Q, and maximum to be roughly even on each side.*

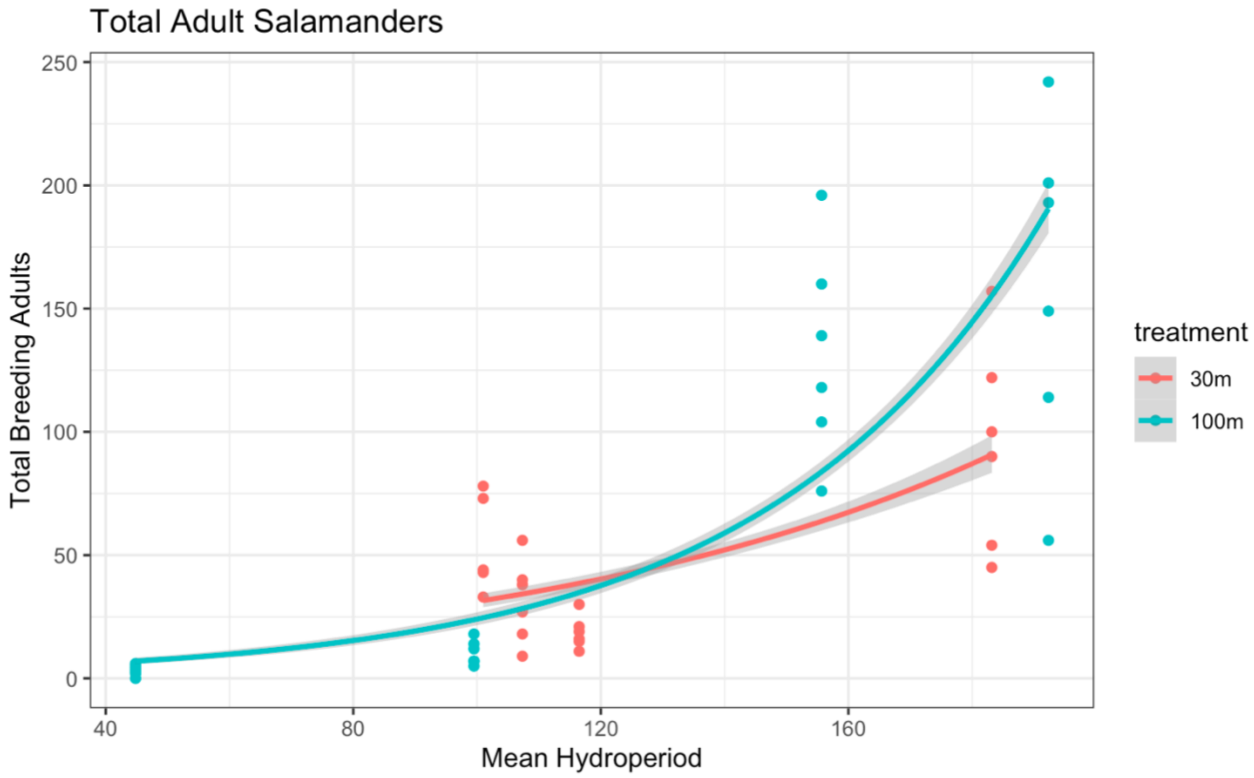
  Here we see:
  Median = 3.185 - not awful but it could surely be better

1Q/3Q = -27.914 / 21.544 – close but could be better
Min / Max = -102.153 / 83.847 - could be much better

- *The standard error (std. error) column on the output and the gray bars on the graph could use a lot of work; we want these numbers to be small, larger numbers represent less certainty.*
- *The slope of 30m treatment is significant, there is a main effect*
- *The 30m treatment (reference) group saw an increase at a slower rate than those in the 100m treatment group.*
- *The 100m treatment group saw a quicker increase in counts, especially at longer hydroperiods.*
- *Our statistical interaction is insignificant. The slopes are not different enough.*

GLM with Poisson

Now let’s check out the GLM with the Poisson distribution



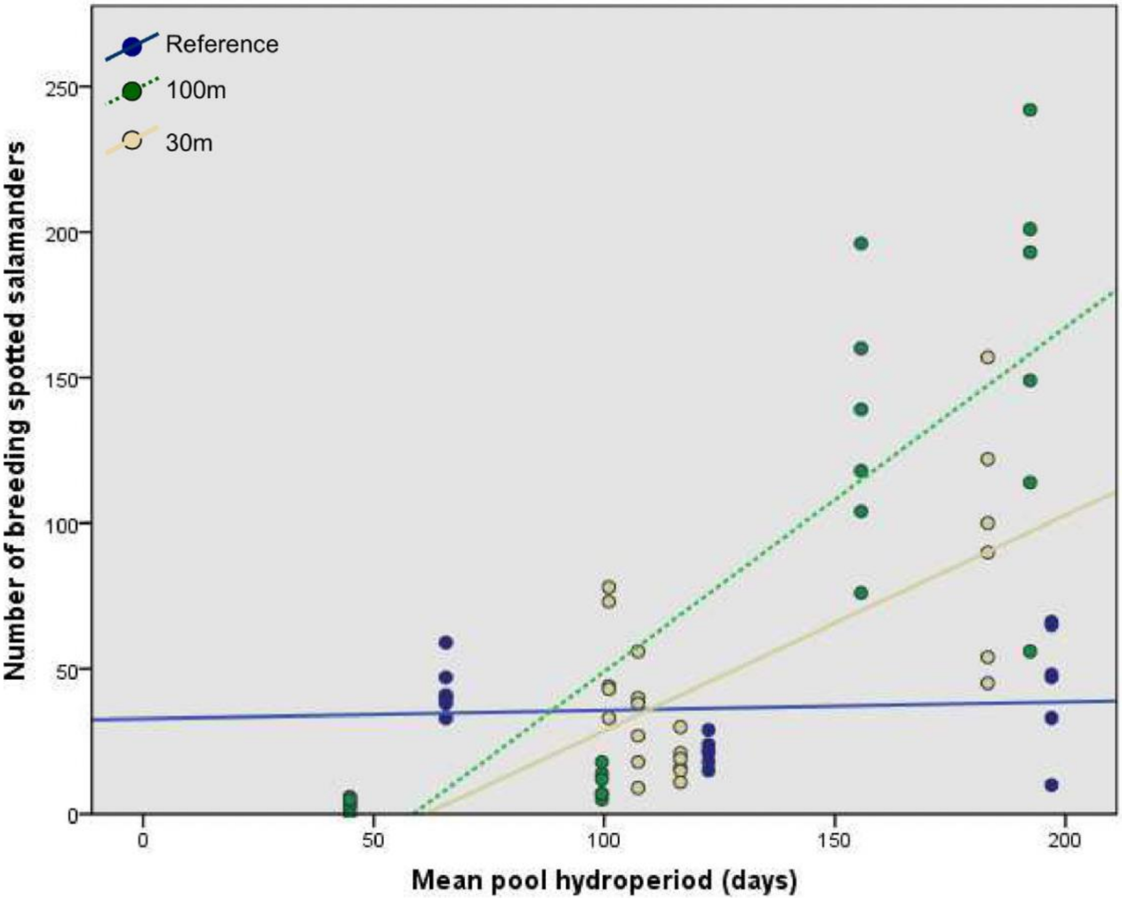
```
Call:
glm(formula = TOTAL.adults ~ MEAN.hydro * treatment, family = poisson(link = "log"),
    data = BuffSala)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-11.489   -3.616   -1.230    2.002   10.425

Coefficients:
                           Estimate Std. Error z value Pr(>|z|)
(Intercept)               2.1568329  0.1188210  18.152  < 2e-16 ***
MEAN.hydro                0.0128330  0.0007961  16.121  < 2e-16 ***
treatment100m            -1.2174970  0.1707136  -7.132 9.91e-13 ***
MEAN.hydro:treatment100m  0.0095804  0.0010607   9.032  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Question 7: Do these results seem like a better fit to the data when compared to the GLM without the Poisson distribution? What does this mean for the salamanders?

- *We see an improvement on our p-value significance levels from the previous GLM without Poisson. All levels are significantly different.*

But of course, let’s do some more digging…

- *Our residuals have improved as well.*
  Here we see:
  Median = -1.230 – closer to 0!
  1Q/3Q = -3.616 / 2.002 - close!
  Min / Max = -11.489 / 10.425 – close!
- *The standard error (std. error) column has improved significantly as well. Smaller numbers represent more certainty with our data.*
- *For the salamanders this shows us that the 100m treatment has an effect up to a point but eventually the hydroperiod allowed the breeding adult population to increase more rapidly especially when comparing it to the 30m treatment. An interaction has occurred!*



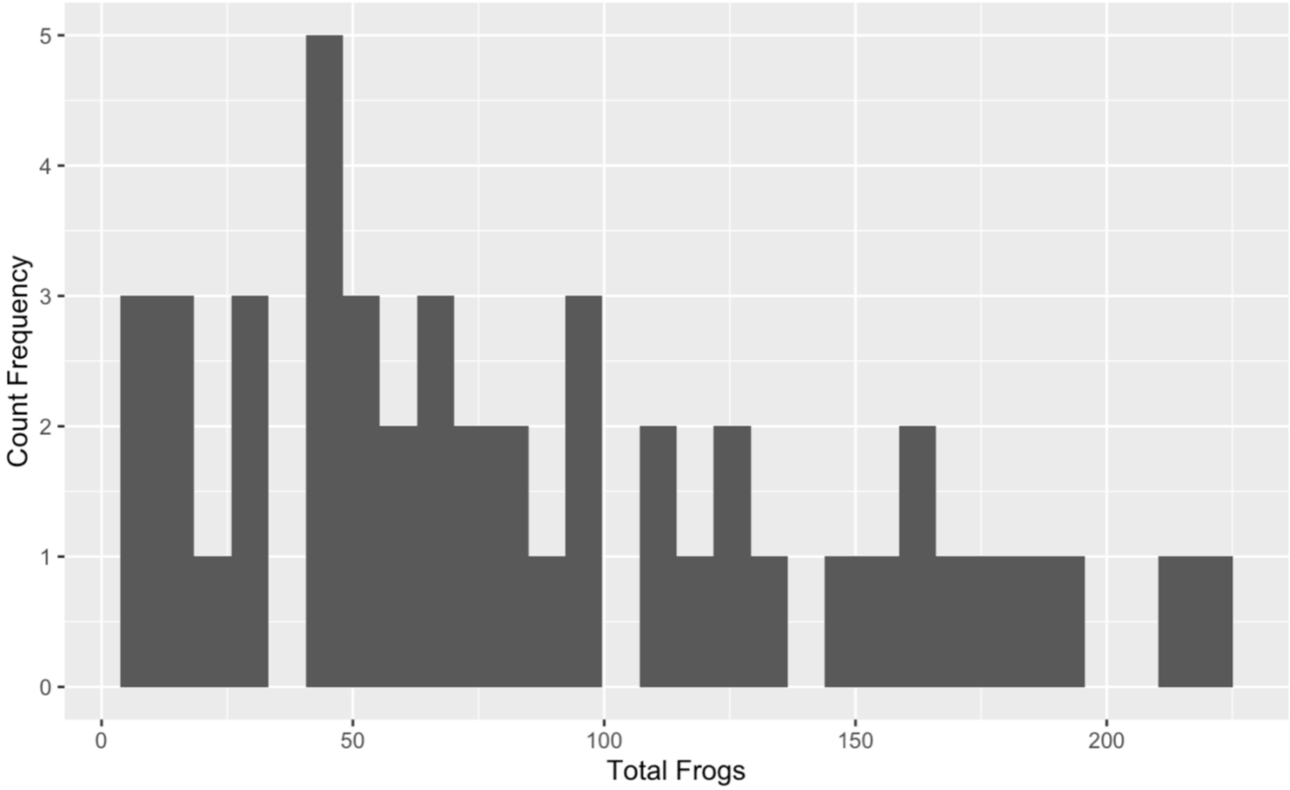
Question 8: The paper included a third treatment level of "reference (uncut)" to show how the 30m and 100m buffer differed. The figure above was taken from the paper (figure 5). How does our paper describe what is happening with the reference pools (blue line)? Is this what you expected to happen?

- *Page 15. They were not able to easily explain why the longer hydroperiods at 100m buffers were more abundant than the reference pools*
- *Page 17. Multiple breeding wetlands could be within a few hundred meters of each other, adults could have switched pools (reduced philopatry); possibly one outside of the study range.*
- *Due to disturbances (clearcutting, possibly toe-clip), the salamanders could have delayed reproduction (temporary emigration).*

Optional Activity Graphs (Not included in student section)

Encourage your students to run these analyses with different names than what was taught for the salamander data, running these commands multiple times with the same names, and expecting different results may confuse R.

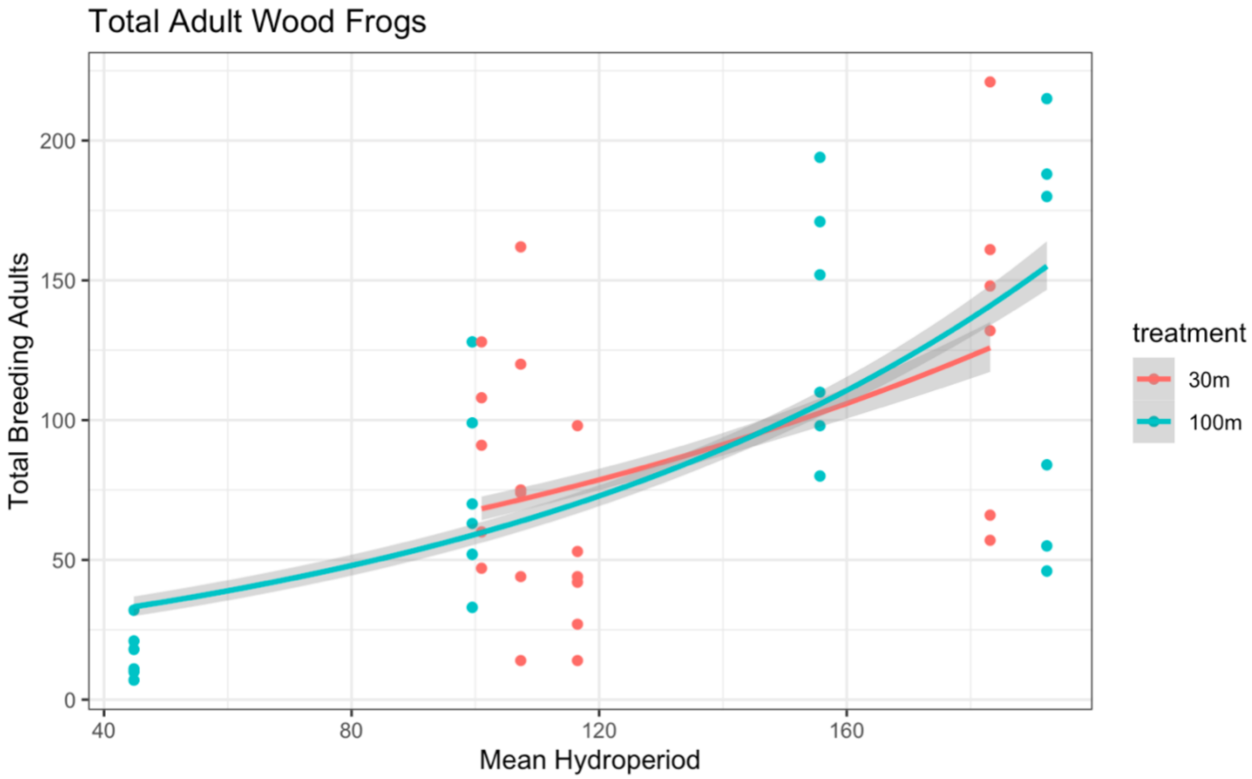
Histogram to see distribution of frog data



We learned our lesson with the salamander data so we will only be running this model with the Poisson distribution. Just to make sure, is this data normally distributed?

No, this data is not normally distributed. See question 5.

```
Call:
glm(formula = TOTAL.adults ~ MEAN.hydro * treatment, family = poisson(link = "log"),
    data = BuffFrog)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-10.3111   -4.1928   -0.2146    2.8068    9.1542

Coefficients:
                           Estimate Std. Error z value Pr(>|z|)
(Intercept)               3.4718571  0.0876285  39.620  < 2e-16 ***
MEAN.hydro                0.0074438  0.0006174  12.057  < 2e-16 ***
treatment100m            -0.4365816  0.1139251  -3.832 0.000127 ***
MEAN.hydro:treatment100m  0.0029992  0.0007682   3.904 9.46e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```



- *Our residuals look pretty good.*
  Here we see:
    Median = -0.215 – not bad
    1Q/3Q = -4.193/ 2.807 – good
    Min / Max = -10.311 / 9.154 -good
- *The standard error (std. error) column looks good, and gray bars on the graph look good too. Smaller numbers represent more certainty with our data.*
- *Here the 100m buffer is increasing faster than the 30m buffer, the difference in these slopes is significant.*

References

Childs DZ, Petchey OL, Beckerman AP. Advancing Your Statistics in R. In: *Getting Started with R*. Oxford University Press; 2017. doi:10.1093/acprof:oso/9780198787839.003.0006

Cushman, S. A. (2006). Effects of habitat loss and fragmentation on amphibians: A review and prospectus. *Biological Conservation*, *128*(2), 231–240. https://doi.org/10.1016/j.biocon.2005.09.031

Fowler, J., Cohen, L., & Jarvis, P. (2009). Probability. In *Practical Statistics for Field Biology* (2nd ed., pp. 54–55). essay, John Wiley & Sons.

Hector A. *The New Statistics with R: An Introduction for Biologists*. Oxford University Press; 2015. doi:10.1093/acprof:oso/9780198729051.001.0001

Powell JSV, Babbitt KJ. An Experimental Test of Buffer Utility as a Technique for Managing Pool-Breeding Amphibians. PloS one. 2015;10(7):e0133642-e0133642. doi:10.1371/journal.pone.0133642

Semlitsch, R. D., & Skelly, D. K. (2008). Ecology and Conservation of Pool-Breeding Amphibians. In *Science and Conservation of Vernal Pools in Northeastern North America* (pp. 128–142). essay, CRC Press.

The IUCN Red List of Threatened Species. IUCN Red List of Threatened Species. (n.d.). Retrieved April 4, 2022, from https://www.iucnredlist.org/resources/summary-statistics

Veysey Powell, Jessica S.; Babbitt, Kimberly J. (2016), Data from: An experimental test of buffer utility as a technique for managing pool-breeding amphibians, Dryad, Dataset, https://doi.org/10.5061/dryad.547rp