Rock Pool Temperature Model

Andy

12/7/2021

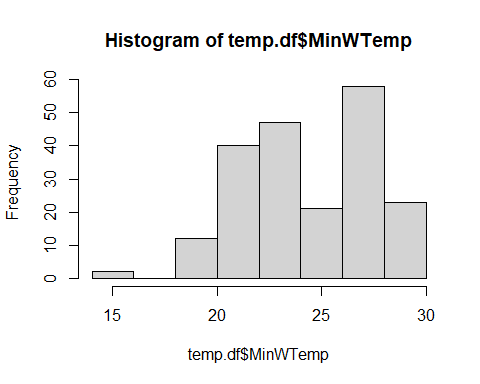
temp.df <- read.csv("temperature\_data.csv")

First, let’s take a look at some of the data and explore it for correlations.

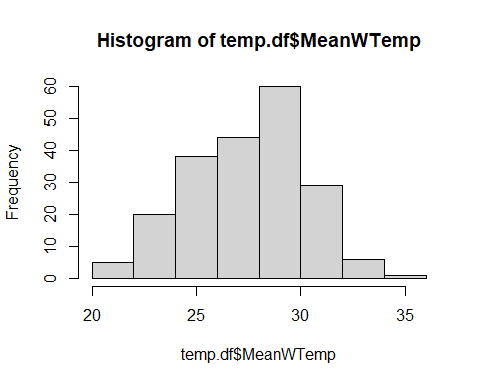
# Histograms

## Temperature Metrics

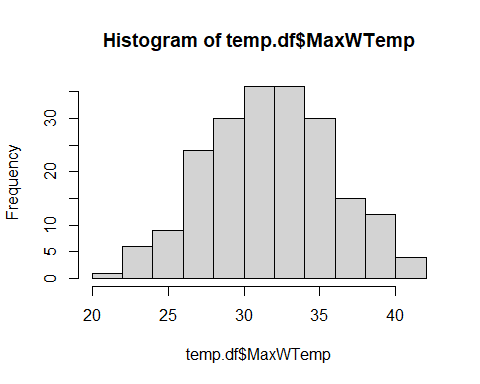
hist(temp.df$MinWTemp)



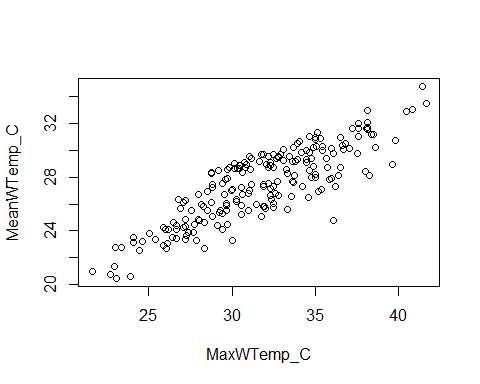
hist(temp.df$MeanWTemp)



hist(temp.df$MaxWTemp)



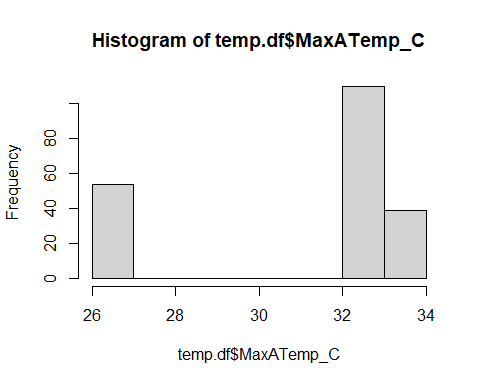
plot(MeanWTemp\_C ~ MaxWTemp\_C, data = temp.df)



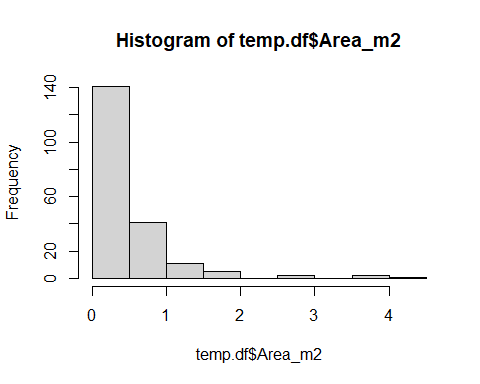
For this analysis, I’m choosing to focus in on maximum water temperature as our dependent variable, because it has a nice and pretty distribution and correlates well with mean water temperature (which is sort of what we’re more interested in here). We could easily use mean temperature as well, but I am hesitant to use it because it does not represent a true daily “mean”, because our data were collected over 8 hour periods and not a full 24 hours.

## Predictor Variables

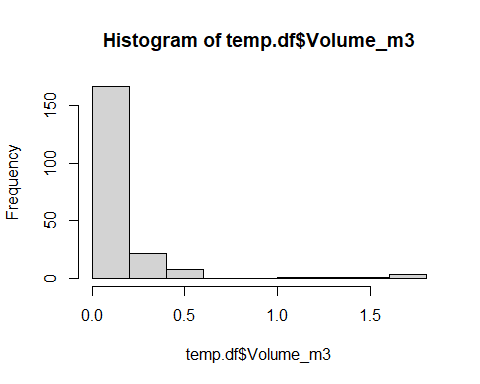
hist(temp.df$MaxATemp\_C)



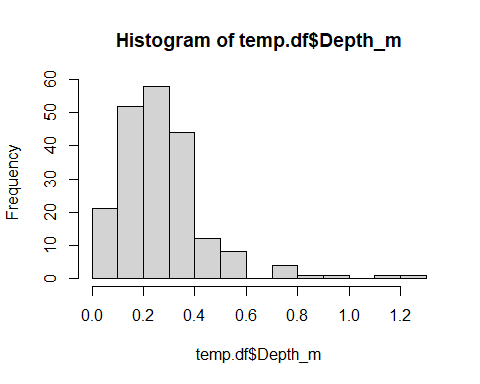
hist(temp.df$Area\_m2)



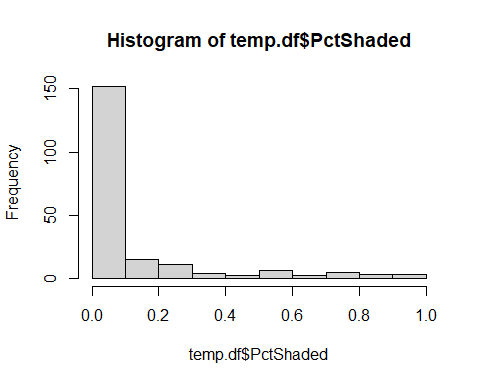
hist(temp.df$Volume\_m3)



hist(temp.df$Depth\_m)



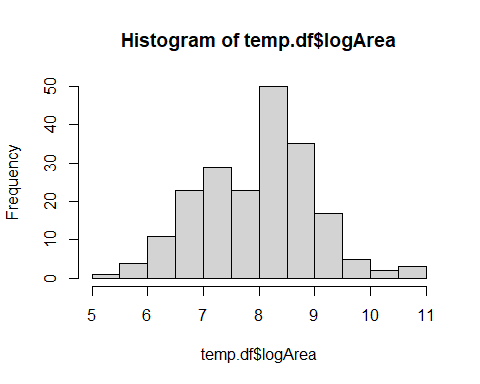
hist(temp.df$PctShaded)



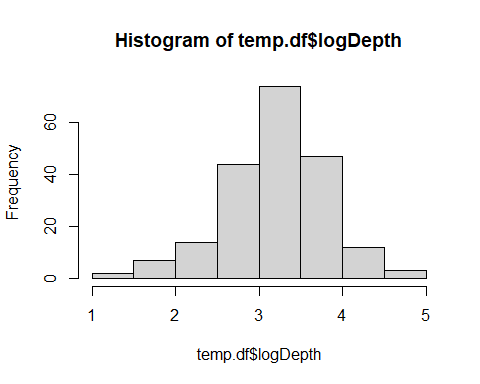
## Applying Transformations

Lots of right-skewing going on, so let’s transform these variables. We’ll use log +1 for Area and Depth, and either logit or arcsine for PctShaded since it’s a proportion.

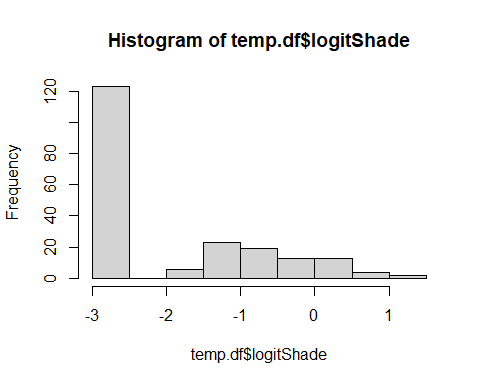
temp.df$Area\_cm2 <- temp.df$Area\_m2 \* 10000  
temp.df$logArea <- log(temp.df$Area\_cm2 + 1)  
temp.df$logDepth <- log(temp.df$Depth\_cm + 1)  
temp.df$logitShade <- log10(temp.df$PctShaded/(1-temp.df$PctShaded))  
temp.df$arcShade <- asin(sqrt(temp.df$PctShaded))  
  
hist(temp.df$logArea)



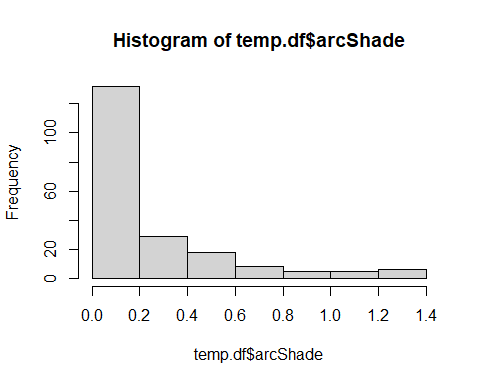
hist(temp.df$logDepth)



hist(temp.df$logitShade)



hist(temp.df$arcShade)



Arcsine seems to do a better job than logit, in this case, but it is still pretty zero inflated.

# Correlation Plots

require(GGally)

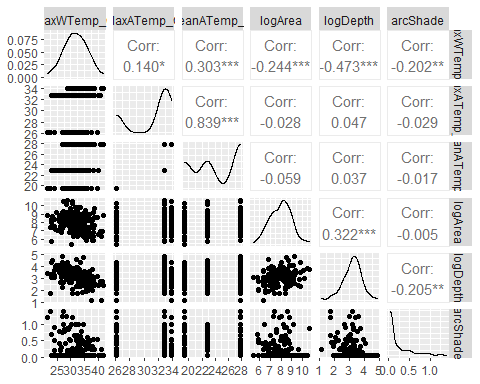
## Loading required package: GGally

## Warning: package 'GGally' was built under R version 4.1.2

## Loading required package: ggplot2

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

temp2.df <- subset(temp.df, select = c("MaxWTemp\_C", "MaxATemp\_C", "MeanATemp\_C", "logArea", "logDepth", "arcShade"))  
ggpairs(data = temp2.df)



Area and depth look like good predictors, as does maximum air temperature and shade. Oddly, there is also some colinearity going on with some of our predictor variables, too (e.g., area x depth, shade x depth).

# Preliminary Modeling

Let’s try making a basic GLMM using these variables, with area and depth as fixed effects and maximum air temperature as a random effect.

require(glmmTMB)

## Loading required package: glmmTMB

## Warning: package 'glmmTMB' was built under R version 4.1.2

## Warning in checkMatrixPackageVersion(): Package version inconsistency detected.  
## TMB was built with Matrix version 1.3.4  
## Current Matrix version is 1.3.3  
## Please re-install 'TMB' from source using install.packages('TMB', type = 'source') or ask CRAN for a binary version of 'TMB' matching CRAN's 'Matrix' package

require(performance)

## Loading required package: performance

## Warning: package 'performance' was built under R version 4.1.2

mod1 <- glmmTMB(MaxWTemp\_C ~ logDepth \* logArea \* arcShade + (1 | MaxATemp\_C), data = temp.df )  
summary(mod1)

## Family: gaussian ( identity )  
## Formula: MaxWTemp\_C ~ logDepth \* logArea \* arcShade + (1 | MaxATemp\_C)  
## Data: temp.df  
##   
## AIC BIC logLik deviance df.resid   
## 1051.4 1084.5 -515.7 1031.4 193   
##   
## Random effects:  
##   
## Conditional model:  
## Groups Name Variance Std.Dev.  
## MaxATemp\_C (Intercept) 2.477 1.574   
## Residual 9.028 3.005   
## Number of obs: 203, groups: MaxATemp\_C, 3  
##   
## Dispersion estimate for gaussian family (sigma^2): 9.03   
##   
## Conditional model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 49.02214 10.92790 4.486 7.26e-06 \*\*\*  
## logDepth -3.85959 3.39196 -1.138 0.255   
## logArea -0.26667 1.39492 -0.191 0.848   
## arcShade -32.07827 33.66881 -0.953 0.341   
## logDepth:logArea -0.05983 0.42812 -0.140 0.889   
## logDepth:arcShade 7.70537 10.75391 0.717 0.474   
## logArea:arcShade 1.93673 4.27262 0.453 0.650   
## logDepth:logArea:arcShade -0.43604 1.35073 -0.323 0.747   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

r2(mod1)

## # R2 for Mixed Models  
##   
## Conditional R2: 0.479  
## Marginal R2: 0.336

Nothing is significant except the intercept. Given the colinearity between area and depth, let’s try dropping one or both of them.

mod2 <- glmmTMB(MaxWTemp\_C ~ logDepth \* arcShade + (1 | MaxATemp\_C), data = temp.df )  
summary(mod2)

## Family: gaussian ( identity )  
## Formula: MaxWTemp\_C ~ logDepth \* arcShade + (1 | MaxATemp\_C)  
## Data: temp.df  
##   
## AIC BIC logLik deviance df.resid   
## 1046.4 1066.3 -517.2 1034.4 197   
##   
## Random effects:  
##   
## Conditional model:  
## Groups Name Variance Std.Dev.  
## MaxATemp\_C (Intercept) 2.419 1.555   
## Residual 9.168 3.028   
## Number of obs: 203, groups: MaxATemp\_C, 3  
##   
## Dispersion estimate for gaussian family (sigma^2): 9.17   
##   
## Conditional model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 47.568 1.717 27.709 < 2e-16 \*\*\*  
## logDepth -4.558 0.445 -10.243 < 2e-16 \*\*\*  
## arcShade -17.641 3.323 -5.309 1.11e-07 \*\*\*  
## logDepth:arcShade 4.550 1.082 4.206 2.60e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

r2(mod2)

## # R2 for Mixed Models  
##   
## Conditional R2: 0.470  
## Marginal R2: 0.330

mod3 <- glmmTMB(MaxWTemp\_C ~ logArea \* arcShade + (1 | MaxATemp\_C), data = temp.df )  
summary(mod3)

## Family: gaussian ( identity )  
## Formula: MaxWTemp\_C ~ logArea \* arcShade + (1 | MaxATemp\_C)  
## Data: temp.df  
##   
## AIC BIC logLik deviance df.resid   
## 1117.6 1137.5 -552.8 1105.6 197   
##   
## Random effects:  
##   
## Conditional model:  
## Groups Name Variance Std.Dev.  
## MaxATemp\_C (Intercept) 2.871 1.694   
## Residual 13.049 3.612   
## Number of obs: 203, groups: MaxATemp\_C, 3  
##   
## Dispersion estimate for gaussian family (sigma^2): 13   
##   
## Conditional model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 42.8266 2.7253 15.714 < 2e-16 \*\*\*  
## logArea -1.2484 0.3137 -3.979 6.91e-05 \*\*\*  
## arcShade -12.6812 7.2344 -1.753 0.0796 .   
## logArea:arcShade 1.2425 0.8875 1.400 0.1615   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

r2(mod3)

## # R2 for Mixed Models  
##   
## Conditional R2: 0.263  
## Marginal R2: 0.101

Model 2, which excludes area but includes depth, has the lowest AIC. It indicates that depth and shade both negatively impact maximum water temperatures, and there is an interaction - deep, shady pools are probably especially cool. Additionally, maximum air temperature does pull out as a significant random effect.

Model 2 seems to explain less of the variance than the first model, but the parameter estimates are probably more useful.

## Exploring Other Model Types

Does the marginal R2 values being lower than the conditional R2 values suggest that air temperature is not a “useful” random effect? Some studies (e.g., Caissie et al. 2001) suggest that water temperature should be modeled as a logistic function of air temperature, instead.

There’s also the possibility of doing it as a Bayesian hierarchical model, which could be fun, but perhaps not any more (or less) valid.

Ryland also suggested making a composite variable out of area and depth. Mike may know how.