Hurdle Model by Site: UBCO

#Load in necessary libraries

```r
#Load in necessary libraries
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.3.3
```

```r
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 4.3.3
```

```
## Loading required package: Matrix
```

```
## Warning: package 'Matrix' was built under R version 4.3.3
```

```r
library(DAAG)
```

```
## Warning: package 'DAAG' was built under R version 4.3.3
```

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```r
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.3.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(MuMIn)
```

```
## Warning: package 'MuMIn' was built under R version 4.3.3
```

```r
library(car)
```

```
## Warning: package 'car' was built under R version 4.3.3
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 4.3.3
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:DAAG':
##
##     vif
```

```r
library(viridis)
```

```
## Warning: package 'viridis' was built under R version 4.3.3
```

```
## Loading required package: viridisLite
```

```r
library(DHARMa)
```

```
## Warning: package 'DHARMa' was built under R version 4.3.3
```

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

```r
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 4.3.3
```

```
##
## Attaching package: 'tidyr'
```

```
## The following objects are masked from 'package:Matrix':
##
##     expand, pack, unpack
```

```r
#Import and Clean Data
data <- read_excel("~/UBCO/Invasion publication/Invasion_publication/Invasion 197198_ID_FINAL.xlsx")

#Remove missing values (SampleID24, 2017)
#Convert NAs to proper data type
data[data == "NA"] <- NA
#Remove nas
data <- drop_na(data)

#Set seed
set.seed(123)

#Review data structure
summary(data)
```

```
##       Year          Plot              SampleID          Plant
##  Min.   :2014   Length:285         Min.   : 1.00   Min.   :1.000
##  1st Qu.:2014   Class :character   1st Qu.: 6.00   1st Qu.:1.000
##  Median :2015   Mode  :character   Median :12.00   Median :3.000
##  Mean   :2015                      Mean   :12.38   Mean   :3.232
##  3rd Qu.:2016                      3rd Qu.:18.00   3rd Qu.:4.000
##  Max.   :2017                      Max.   :24.00   Max.   :5.000
##    Distance          Copies_ul
##  Length:285         Length:285
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
```

```r
head(data)
```

```
## # A tibble: 6 x 6
##    Year Plot      SampleID Plant Distance Copies_ul
##   <dbl> <chr>        <dbl> <dbl> <chr>    <chr>
## 1  2014 Aberdeen         1     1 10       2.1
## 2  2014 Aberdeen         1     1 25       0.08
## 3  2014 Aberdeen         1     1 50       0
## 4  2014 Tutt             2     1 10       0.26
## 5  2014 Tutt             2     1 25       5.5
## 6  2014 Tutt             2     1 50       21.8
```

```r
tail(data)
```

```
## # A tibble: 6 x 6
##    Year Plot  SampleID Plant Distance Copies_ul
##   <dbl> <chr>    <dbl> <dbl> <chr>    <chr>
## 1  2017 KM 2        22     5 10       0
## 2  2017 KM 2        22     5 25       0
## 3  2017 KM 2        22     5 50       0
## 4  2017 KM 4        23     5 10       1
## 5  2017 KM 4        23     5 25       0.25
## 6  2017 KM 4        23     5 50       0.23
```

```r
#Convert copies/uL to numeric type
data$Copies_ul <- as.numeric(data$Copies_ul)

#Pool distances per plant (one observation/plant)
data <- data %>%
  group_by(SampleID, Year, Plant, Plot) %>%
  summarize(Copies_ul = sum(Copies_ul, na.rm = TRUE), .groups = "drop")
summary(data)
```

```
##      SampleID           Year          Plant           Plot
##  Min.   : 1.00   Min.   :2014   Min.   :1.000   Length:95
##  1st Qu.: 6.50   1st Qu.:2014   1st Qu.:2.000   Class :character
##  Median :12.00   Median :2015   Median :3.000   Mode  :character
##  Mean   :12.38   Mean   :2015   Mean   :3.232
##  3rd Qu.:18.00   3rd Qu.:2016   3rd Qu.:4.000
##  Max.   :24.00   Max.   :2017   Max.   :5.000
##    Copies_ul
##  Min.   :  0.000
##  1st Qu.:  0.000
##  Median :  0.280
##  Mean   :  5.641
##  3rd Qu.:  2.740
##  Max.   :142.900
```

```r
#Create Predictor Labels
#Plant 3 and 5 = inoculated (1), plant 1 and 4 = uninoculated (0)
#Plants 1 and 3 = achillea; 4 and 5, = artemesia
data <- data %>%
  mutate(
    hostID = ifelse(Plant > 3, "Artemesia", "Achillea"),
    inoculation = case_when(
      Plant %in% c(1,4) ~ 0L,
      Plant %in% c(3,5) ~ 1L,
      TRUE ~ NA_integer_
    ),
    Site = case_when(
      Plot %in% c("Aberdeen","Quail","Tutt") ~ "UBCO",
      Plot %in% c("KM 2","KM 4","KM 5") ~ "Knox",
      TRUE ~ NA_character_
    )
  )
head(data)
```

```
## # A tibble: 6 x 8
##   SampleID  Year Plant Plot    Copies_ul hostID   inoculation Site
##      <dbl> <dbl> <dbl> <chr>       <dbl> <chr>          <int> <chr>
## 1        1  2014     1 Aberdeen     2.18 Achillea           0 UBCO
## 2        1  2015     1 Aberdeen     2.8  Achillea           0 UBCO
## 3        1  2016     1 Aberdeen     1.08 Achillea           0 UBCO
## 4        1  2017     1 Aberdeen     3.49 Achillea           0 UBCO
## 5        2  2014     1 Tutt        27.6  Achillea           0 UBCO
## 6        2  2015     1 Tutt         4.86 Achillea           0 UBCO
```

```r
tail(data)
```

```
## # A tibble: 6 x 8
##   SampleID  Year Plant Plot  Copies_ul hostID    inoculation Site
##      <dbl> <dbl> <dbl> <chr>     <dbl> <chr>           <int> <chr>
## 1      23  2015     5 KM 4       0     Artemesia           1 Knox
## 2      23  2016     5 KM 4       0     Artemesia           1 Knox
## 3      23  2017     5 KM 4       1.48  Artemesia           1 Knox
## 4      24  2014     5 KM 5      11.8   Artemesia           1 Knox
## 5      24  2015     5 KM 5       3.5   Artemesia           1 Knox
## 6      24  2016     5 KM 5       0     Artemesia           1 Knox
```

```r
#Filter to site (UBCO)
SITE <- "UBCO"
data <- data %>% filter(Site == SITE)

#Ensure all predictors and response are the correct type. Baseline year 2014 for
#intercept estimate
data <- data %>% mutate(
  hostID = factor(hostID),
  Site = factor(Site),
  Plot = factor(Plot),
  inoculation = factor(inoculation),
  Copies_ul = as.numeric(Copies_ul),
  Year = as.integer(Year),
  Year_centered = Year - 2014
)

#Check zero inflation
data <- data %>% mutate(Presence = if_else(Copies_ul > 0, 1, 0))
prop_zero <- mean(data$Presence == 0)
print(prop_zero) #14%
```
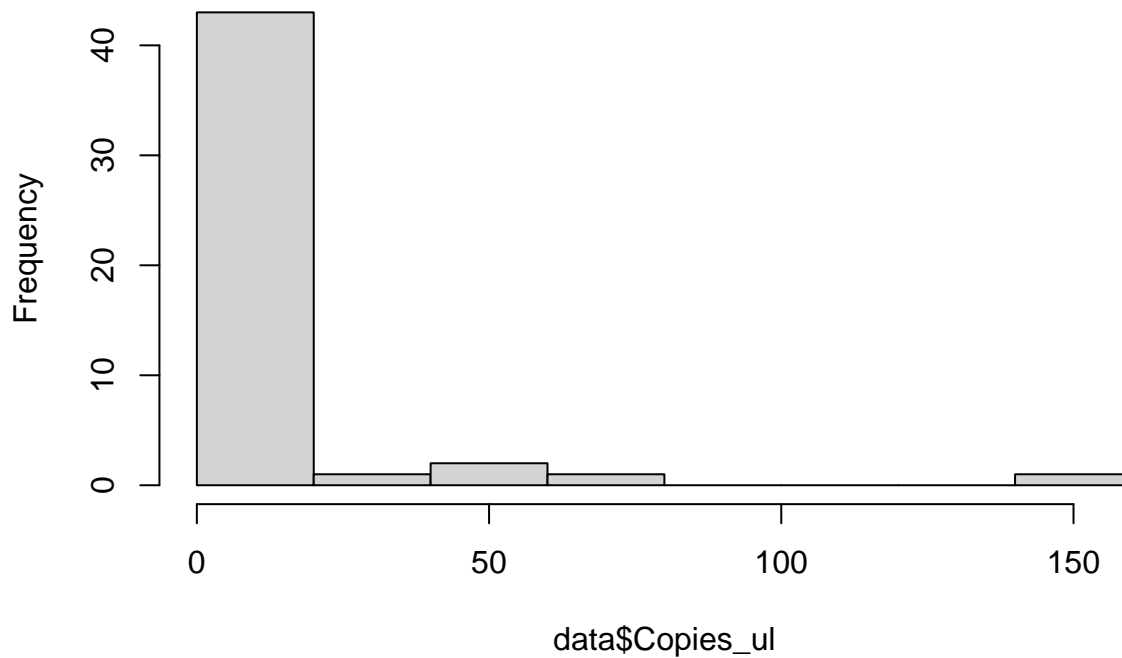
```
## [1] 0.1458333
```

```r
hist(data$Copies_ul) #zero inflation
```

Histogram of data$Copies_ul



```
zero_df <- data %>% filter(Presence == 0)
View(zero_df)
SUBSET <- data %>% filter(Copies_ul > 0) %>% mutate(Year_centered = Year - 2014)

#Simple One-Way Visualizations
#Create the mean presence plots
#Inoculation
barplot(tapply(data$Presence, data$inoculation, mean, na.rm = TRUE),
        ylim = c(0, 1),
        ylab = "Mean presence",
        main = "Mean Presence by Inoculation (UBCO)")
mtext("(a)", side = 3, line = 1, adj = 0, font = 2) #Add label
```

(a) **Mean Presence by Inoculation (UBCO)**
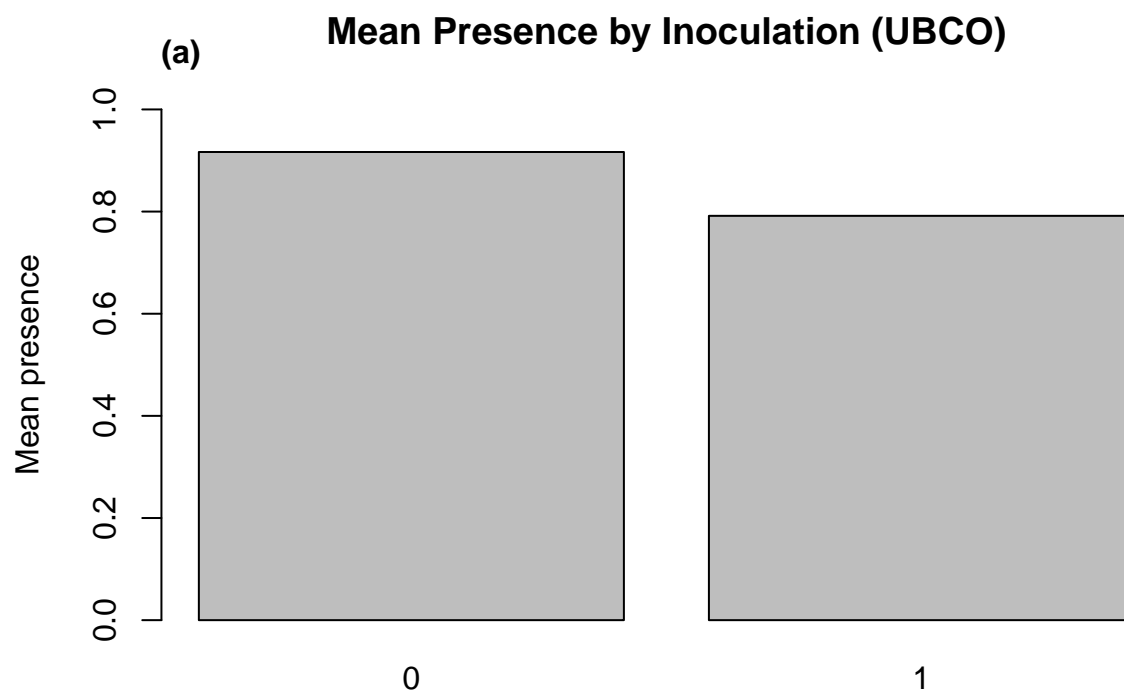

```
#Year
barplot(tapply(data$Presence, data$Year, mean, na.rm = TRUE),
        ylim = c(0, 1),
        ylab = "Mean presence",
        main = "Mean Presence by Year (UBCO)")
mtext("(b)", side = 3, line = 1, adj = 0, font = 2)
```

Mean Presence by Year (UBCO)

(b)
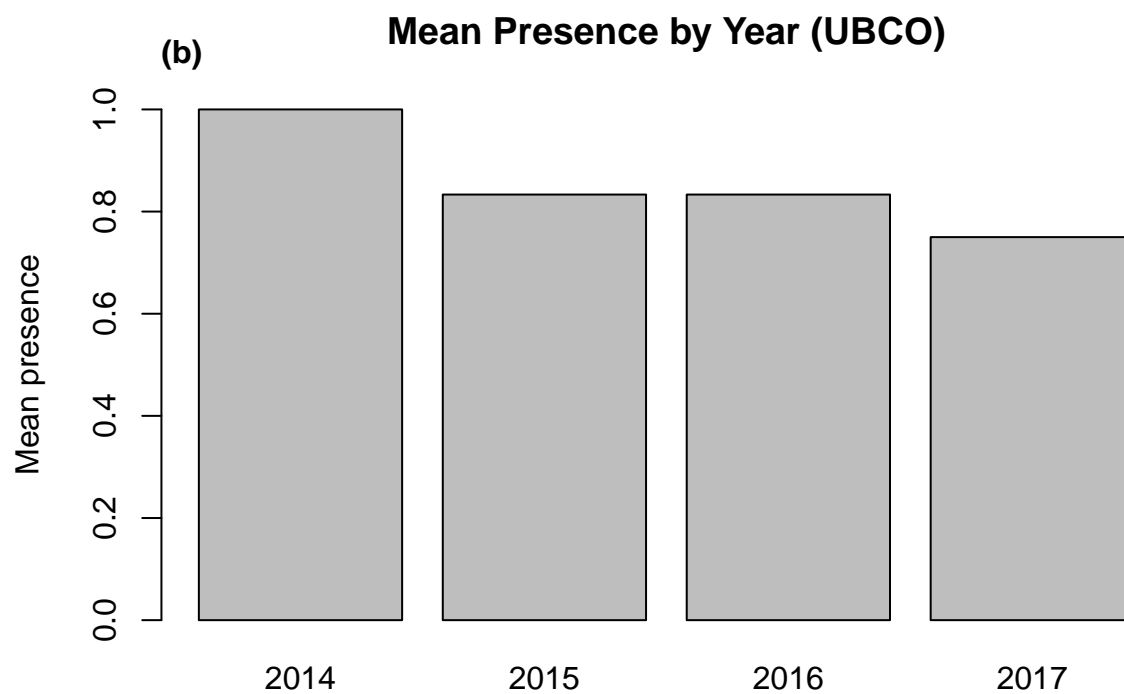

```r
#hostID
barplot(tapply(data$Presence, data$hostID, mean, na.rm = TRUE),
        ylim = c(0, 1),
        ylab = "Mean presence",
        main = "Mean Presence by Host (UBCO)")
mtext("(c)", side = 3, line = 1, adj = 0, font = 2)
```

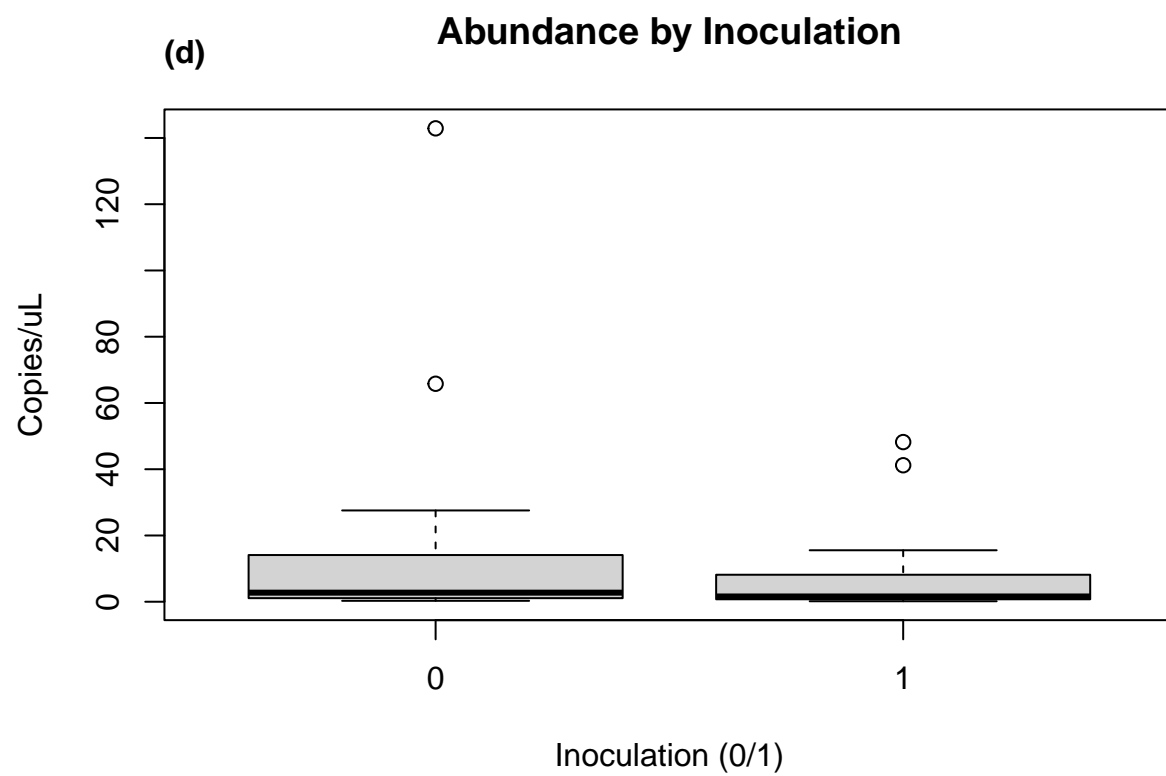## Mean Presence by Host (UBCO)

(c)
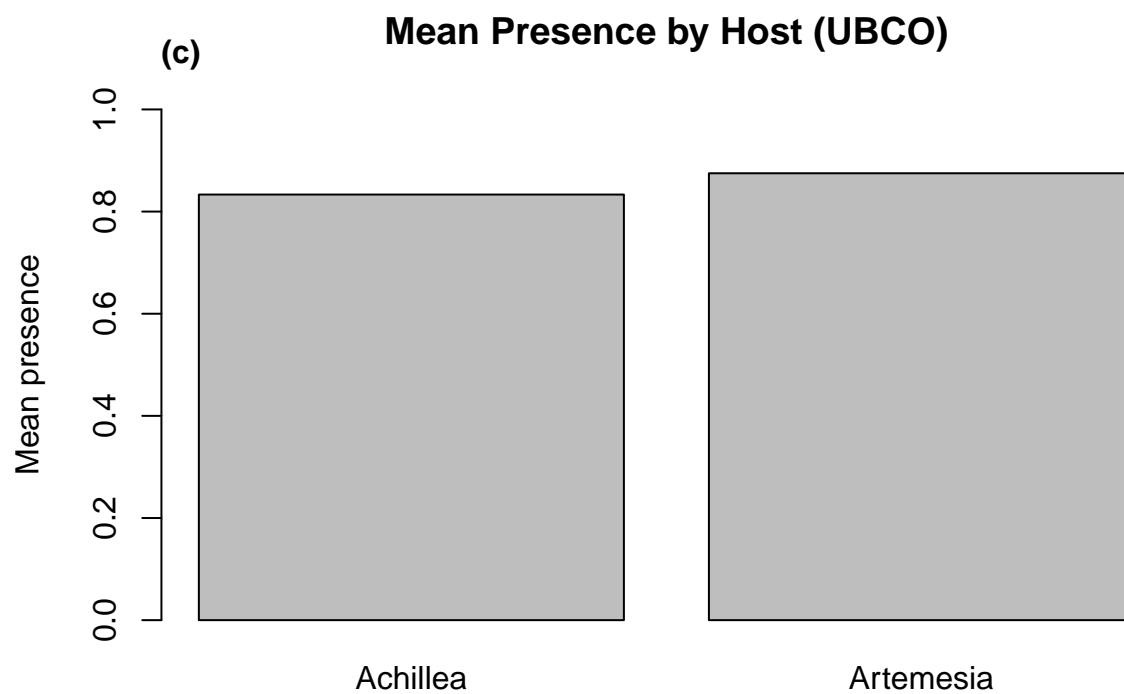

```r
#Median abundance
#Inoculation
boxplot(Copies_ul ~ inoculation, data = SUBSET,
        xlab = "Inoculation (0/1)",
        ylab = "Copies/uL",
        main = "Abundance by Inoculation")
mtext("(d)", side = 3, line = 1, adj = 0, font = 2)
```

(d)

Abundance by Inoculation


```
#Year
boxplot(Copies_ul ~ Year, data = SUBSET,
        xlab = "Year", ylab = "Copies/uL", main = "Abundance by Year")
mtext("(e)", side = 3, line = 1, adj = 0, font = 2)
```

Abundance by Year

(e)
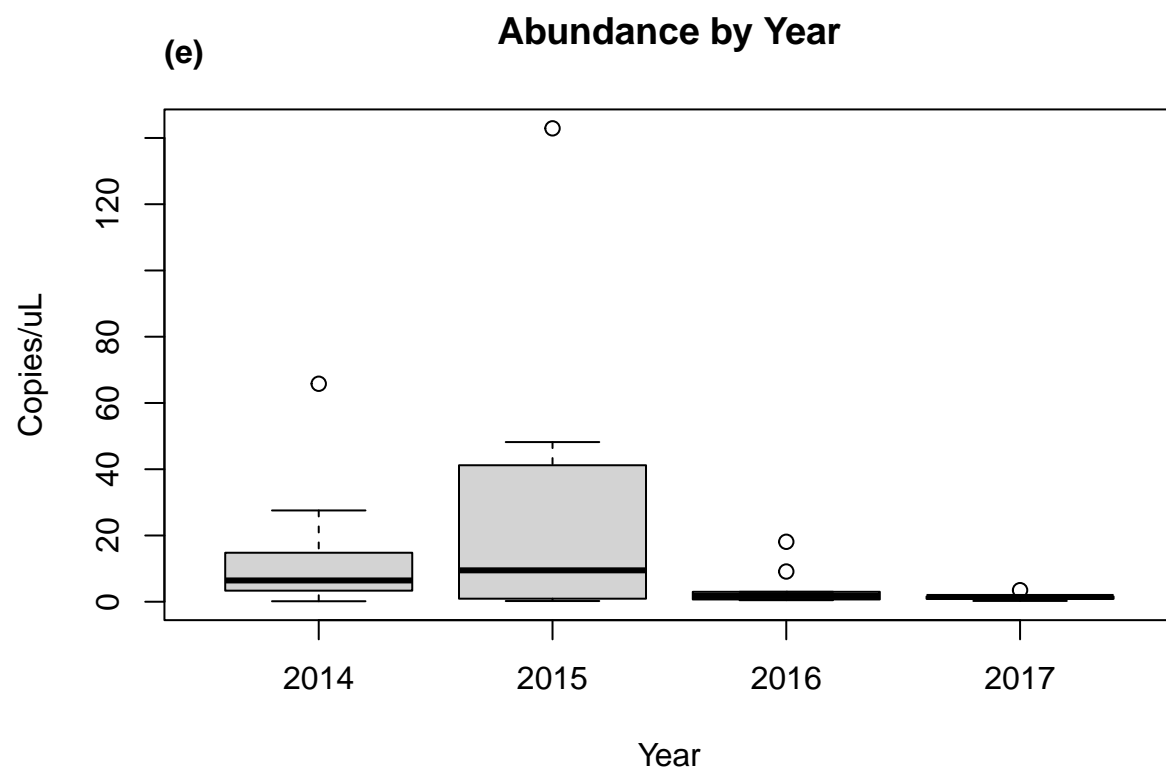

```
#HostID
boxplot(Copies_ul ~ hostID, data = SUBSET,
        xlab = "Host",
        ylab = "Copies/uL",
        main = "Abundance by Host")
mtext("(f)", side = 3, line = 1, adj = 0, font = 2)
```

(f) **Abundance by Host**
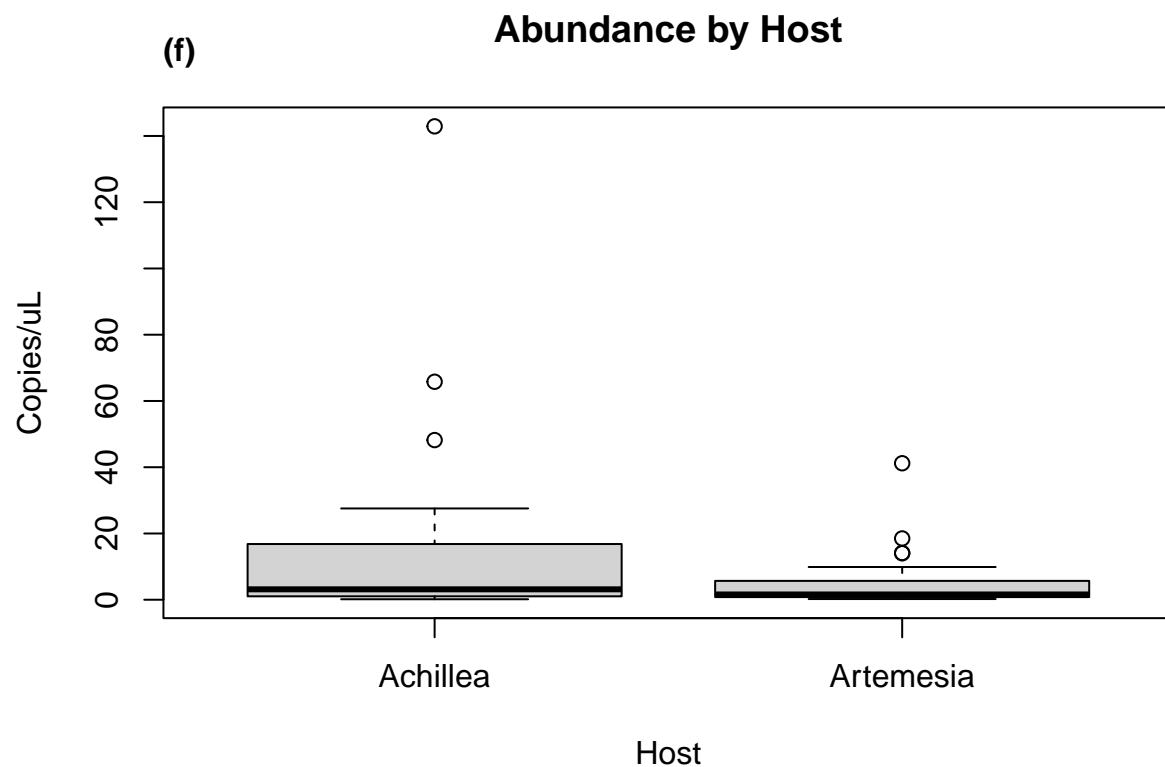

```r
dev.off()
```

```
## null device
##           1
```

```r
#Check Possible Interactions
par(mfrow = c(1,1)) #Revert plot to 1 per row and column

#inoculation x hostID
interaction.plot(data$inoculation, data$hostID, data$Presence,
                 main= "UBCO hostID*inoculation",
                 xlab= "Inoculation",
                 ylab= "Mean presence",
                 legend= TRUE,
                 col= 1:length(unique(data$hostID)),
                 pch= 1:length(unique(data$hostID)),
                 type= "b")
#Likely interaction

#inoculation x Year
interaction.plot(data$Year, data$inoculation, data$Presence,
                 main= "UBCO hostID*inoculation",
                 xlab= "Year",
                 ylab= "Mean presence",
                 legend= TRUE,
```

```
                col= 1:length(unique(data$inoculation)),
                pch= 1:length(unique(data$inoculation)),
                type= "b")
#Possible interaction

#Binomial Model (Presence)
#Note: Plot is often singular or collinear with Site at UBCO; omit.
#Start with the simplest model
#Year as a predictor
binom_FIXED_1 <- glm(Presence ~ Year_centered,
                     family = binomial(link = "logit"),
                     data = data)
summary(binom_FIXED_1)
```

```
##
## Call:
## glm(formula = Presence ~ Year_centered, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)     2.9426     0.9558   3.079  0.00208 **
## Year_centered  -0.6595     0.4197  -1.571  0.11608
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 39.880  on 47  degrees of freedom
## Residual deviance: 37.047  on 46  degrees of freedom
## AIC: 41.047
##
## Number of Fisher Scoring iterations: 5
```

```
#Year+inoculation
binom_FIXED_2 <- glm(Presence ~ Year_centered + inoculation,
                     family = binomial(link = "logit"),
                     data = data)
summary(binom_FIXED_2)
```

```
##
## Call:
## glm(formula = Presence ~ Year_centered + inoculation, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)     3.6554     1.1904   3.071  0.00214 **
## Year_centered  -0.6835     0.4290  -1.593  0.11106
## inoculation1   -1.1299     0.9212  -1.226  0.22001
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 39.880  on 47  degrees of freedom
## Residual deviance: 35.402  on 45  degrees of freedom
## AIC: 41.402
##
## Number of Fisher Scoring iterations: 5
```

```r
#Year+inoculation+hostID
binom_FIXED_3 <- glm(Presence ~ Year_centered + inoculation + hostID,
                     family = binomial(link = "logit"), data = data)
summary(binom_FIXED_3)
```

```
##
## Call:
## glm(formula = Presence ~ Year_centered + inoculation + hostID,
##     family = binomial(link = "logit"), data = data)
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)       3.4890     1.2406   2.812  0.00492 **
## Year_centered    -0.6867     0.4302  -1.596  0.11043
## inoculation1     -1.1353     0.9237  -1.229  0.21900
## hostIDArtemesia   0.3732     0.8693   0.429  0.66772
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 39.880  on 47  degrees of freedom
## Residual deviance: 35.216  on 44  degrees of freedom
## AIC: 43.216
##
## Number of Fisher Scoring iterations: 5
```

```r
#Compare cv binary
cv.binary(binom_FIXED_1)
```

```
##
## Fold:  3 10 2 6 5 4 9 8 7 1
## Internal estimate of accuracy = 0.854
## Cross-validation estimate of accuracy = 0.854
```

```r
cv.binary(binom_FIXED_2)
```

```
##
## Fold:  4 6 8 7 1 2 5 3 9 10
## Internal estimate of accuracy = 0.854
## Cross-validation estimate of accuracy = 0.854
```

```
cv.binary(binom_FIXED_3)
```

```
##
## Fold:  7 10 6 5 8 4 3 9 2 1
## Internal estimate of accuracy = 0.854
## Cross-validation estimate of accuracy = 0.833
```

```
#All comparable, year is likely driving presence

#Test whether adding the interaction improves model fit
binom_FIXED_4 <- glm(Presence ~ Year_centered * inoculation,
                     family = binomial(link = "logit"), data = data)
summary(binom_FIXED_4)
```

```
##
## Call:
## glm(formula = Presence ~ Year_centered * inoculation, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  4.6532     2.4673   1.886   0.0593 .
## Year_centered               -1.1171     0.9623  -1.161   0.2457
## inoculation1                -2.4136     2.6830  -0.900   0.3683
## Year_centered:inoculation1   0.5823     1.0798   0.539   0.5897
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 39.880  on 47  degrees of freedom
## Residual deviance: 35.075  on 44  degrees of freedom
## AIC: 43.075
##
## Number of Fisher Scoring iterations: 6
```

```
#Mixed-Effects Binomial Models
#Add a random effects structure to accout for repeated measure across years
binom_RI <- glmer(Presence ~ Year_centered + inoculation + hostID
                  + (1 | SampleID),
                  family = binomial(link = "logit"),
                  data = data)
summary(binom_RI)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: Presence ~ Year_centered + inoculation + hostID + (1 | SampleID)
##    Data: data
##
##      AIC      BIC   logLik deviance df.resid
##     45.2     54.5    -17.6     35.2       43
```

```
## 
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2641  0.1538  0.2880  0.4398  0.9481
## 
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  SampleID (Intercept) 0.2727   0.5222
## Number of obs: 48, groups:  SampleID, 12
## 
## Fixed effects:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)       3.6559     1.5149   2.413   0.0158 *
## Year_centered    -0.7118     0.4528  -1.572   0.1159
## inoculation1     -1.1556     0.9948  -1.162   0.2454
## hostIDArtemesia   0.3212     0.9672   0.332   0.7398
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Correlation of Fixed Effects:
##             (Intr) Yr_cnt inclt1
## Year_centrd -0.721
## inoculatin1 -0.539  0.114
## hostIDArtms -0.382  0.021  0.049
```

```
#Sample ID does explain some variance

#Remove hostID to decrease model complexity
binom_RI_2 <- glmer(Presence ~ Year_centered + inoculation
                    + (1 | SampleID),
                    family = binomial(link = "logit"),
                    data = data)
summary(binom_RI_2)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: Presence ~ Year_centered + inoculation + (1 | SampleID)
##    Data: data
## 
##      AIC      BIC   logLik deviance df.resid
##     43.3     50.8    -17.6     35.3       44
## 
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9797  0.1634  0.2788  0.3996  0.9336
## 
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  SampleID (Intercept) 0.4132   0.6428
## Number of obs: 48, groups:  SampleID, 12
## 
## Fixed effects:
##                 Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)       3.8848     1.4325   2.712  0.00669 **
## Year_centered  -0.7198     0.4541  -1.585  0.11295
## inoculation1    -1.1856     1.0328  -1.148  0.25098
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) Yr_cnt
## Year_centrd -0.760
## inoculatin1 -0.585  0.124
```

#model with and without hostID are near identical, exclude host ID

#Compare AIC of each model
AIC(binom_FIXED_1, binom_FIXED_2, binom_FIXED_3, binom_RI, binom_RI_2)

```
##              df      AIC
## binom_FIXED_1  2 41.04744
## binom_FIXED_2  3 41.40221
## binom_FIXED_3  4 43.21621
## binom_RI       5 45.16251
## binom_RI_2     4 43.26681
```

#No statistical difference between binom_RI_2 and all other models (delta AIC<2)
#Continue with the mixed effects model that includes inoculation and year as
#the two fixed effect predictors, as these are the two predictors of interest

#Residual Diagnostics (DHARMa)
#Simluate residuals using DHARMa
binom_RI_res_2 <- **simulateResiduals**(binom_RI)
#Plot simulated residuals
plot(binom_RI_res_2)*#No issues detected*
#Test dispersion
print(**testDispersion**(binom_RI_res_2)) *#No significant issues*

```
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## dispersion = 1.0674, p-value = 0.84
## alternative hypothesis: two.sided
```

#Test for significant outliers
print(**testOutliers**(binom_RI_res_2))*#No significant issues*

```
##
##  DHARMa bootstrapped outlier test
##
## data:  binom_RI_res_2
## outliers at both margin(s) = 0, observations = 48, p-value = 1
## alternative hypothesis: two.sided
```

```
##   percent confidence interval:
##    0 0
## sample estimates:
## outlier frequency (expected: 0.000208333333333333 )
##                                                     0
```

```r
#Test for temporal autocorrelation
testTemporalAutocorrelation(
  DHARMa::recalculateResiduals(binom_RI_res_2, group = data$Year_centered),
  time = sort(unique(data$Year_centered))
)
```

```
##
##   Durbin-Watson test
##
## data:  simulationOutput$scaledResiduals ~ 1
## DW = 2.0746, p-value = 0.8989
## alternative hypothesis: true autocorrelation is not 0
```

```r
#Years are not autocorrelated. Residuals suggest a time pattern not captured
#by the model may be present

#Check residuals for each predictor
plotResiduals(binom_RI_res_2, data$Year_centered)
plotResiduals(binom_RI_res_2, data$inoculation)

#Check for collinearity among predictors
performance::check_collinearity(binom_RI_2) #No collinearity
```

```
## # Check for Multicollinearity
##
## Low Correlation
##
##           Term  VIF       VIF 95% CI adj. VIF Tolerance Tolerance 95% CI
##   Year_centered 1.02 [1.00, 2.61e+05]     1.01      0.98      [0.00, 1.00]
##      inoculation 1.02 [1.00, 2.61e+05]     1.01      0.98      [0.00, 1.00]
```

```r
#Move forward with the mixed effects model (inoculation + year + (1|SampleID)),
#Inoculation and Year are chosen as the final predictors, as these are the
#variables of interest. HostID adds complexity and does not enhance model fit.

#Gamma Model (Abundance)
#Create a histogram of the non-zero observations
hist(SUBSET$Copies_ul,
     breaks=20,
     main="Copies/µL distribution",
     xlab="Copies/µL")

#Start with the simplest model (Year as singular predictor)
gamma_1<- glm(Copies_ul ~ Year_centered,
            family = Gamma(link = "log"),
            data = SUBSET)
summary(gamma_1)
```

```
## 
## Call:
## glm(formula = Copies_ul ~ Year_centered, family = Gamma(link = "log"),
##     data = SUBSET)
## 
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     3.4827     0.4808   7.243 9.96e-09 ***
## Year_centered  -0.9464     0.2690  -3.518  0.00112 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for Gamma family taken to be 3.745109)
## 
##     Null deviance: 110.463  on 40  degrees of freedom
## Residual deviance:  85.714  on 39  degrees of freedom
## AIC: 257.65
## 
## Number of Fisher Scoring iterations: 12
```

```r
#Year+inoculation
gamma_2<- glm(Copies_ul ~ Year_centered + inoculation,
              family = Gamma(link = "log"),
              data = SUBSET)
summary(gamma_2)
```

```
## 
## Call:
## glm(formula = Copies_ul ~ Year_centered + inoculation, family = Gamma(link = "log"),
##     data = SUBSET)
## 
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     3.6879     0.5161   7.146 1.56e-08 ***
## Year_centered  -0.9384     0.2478  -3.787 0.000529 ***
## inoculation1   -0.5440     0.5583  -0.974 0.336090
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for Gamma family taken to be 3.177182)
## 
##     Null deviance: 110.463  on 40  degrees of freedom
## Residual deviance:  82.774  on 38  degrees of freedom
## AIC: 257.83
## 
## Number of Fisher Scoring iterations: 12
```

```r
#Year+inoculation+hostID
gamma_3<- glm(Copies_ul ~ Year_centered + inoculation + hostID,
              family = Gamma(link = "log"),
              data = SUBSET)
summary(gamma_3)
```

```
## 
```

```
## Call:
## glm(formula = Copies_ul ~ Year_centered + inoculation + hostID,
##      family = Gamma(link = "log"), data = SUBSET)
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       3.8874     0.5312   7.319 1.07e-08 ***
## Year_centered    -0.9111     0.2257  -4.036 0.000262 ***
## inoculation1     -0.2202     0.5119  -0.430 0.669541
## hostIDArtemesia  -0.9400     0.5113  -1.838 0.074037 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 2.62999)
##
##     Null deviance: 110.463  on 40  degrees of freedom
## Residual deviance:  75.062  on 37  degrees of freedom
## AIC: 254.77
##
## Number of Fisher Scoring iterations: 11
```

```r
#Add sample ID as a random effects structure to accout for repeated measures
#over years
gamma_4<-glmer(Copies_ul ~ Year_centered + inoculation + hostID
               + (1|SampleID),
               family = Gamma(link = "log"),
               data = SUBSET)
summary(gamma_4)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: Gamma  ( log )
## Formula: Copies_ul ~ Year_centered + inoculation + hostID + (1 | SampleID)
##    Data: SUBSET
##
##      AIC      BIC   logLik deviance df.resid
##    249.6    259.9   -118.8    237.6       35
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.9565 -0.5855 -0.2762  0.3239  2.5643
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  SampleID (Intercept) 1.125    1.061
##  Residual             1.016    1.008
## Number of obs: 41, groups:  SampleID, 12
##
## Fixed effects:
##                 Estimate Std. Error t value Pr(>|z|)
## (Intercept)       3.1176     0.6146   5.073 3.92e-07 ***
## Year_centered    -0.8452     0.1441  -5.866 4.46e-09 ***
## inoculation1     -0.6641     0.6776  -0.980    0.327
## hostIDArtemesia  -0.4460     0.6771  -0.659    0.510
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) Yr_cnt inclt1
## Year_centrd -0.359
## inoculatin1 -0.497  0.018
## hostIDArtms -0.556  0.047 -0.058
```

```r
#Compare the mixed effects model with and without hostID
gamma_5<-glmer(Copies_ul ~ Year_centered + inoculation
               + (1|SampleID),
               family = Gamma(link = "log"),
               data = SUBSET)
summary(gamma_5)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: Gamma  ( log )
## Formula: Copies_ul ~ Year_centered + inoculation + (1 | SampleID)
##    Data: SUBSET
##
##      AIC      BIC   logLik deviance df.resid
##    248.0    256.6   -119.0    238.0       36
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.9590 -0.5693 -0.2837  0.2808  2.4364
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  SampleID (Intercept) 1.205    1.098
##  Residual             1.015    1.007
## Number of obs: 41, groups:  SampleID, 12
##
## Fixed effects:
##              Estimate Std. Error t value Pr(>|z|)
## (Intercept)    2.8917     0.5237   5.522 3.35e-08 ***
## Year_centered -0.8409     0.1435  -5.858 4.69e-09 ***
## inoculation1  -0.6905     0.6925  -0.997    0.319
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) Yr_cnt
## Year_centrd -0.389
## inoculatin1 -0.646  0.021
```

```r
#Compare AIC values
AIC(gamma_1, gamma_2, gamma_3, gamma_4, gamma_5)
```

```
##         df      AIC
## gamma_1  3 257.6541
```

```
## gamma_2   4 257.8287
## gamma_3   5 254.7656
## gamma_4   6 249.5772
## gamma_5   5 247.9954
```

```
#Gamma 4 and gamma 5 are the best AIC fit, exclude host ID to decrease model
#complexity. This does not affect conclusions and retains the two predictors
#of highest interest.

#DHARMa Diagnostics for the selected Gamma Model
#Simulate residuals
res_gamma_5 <- simulateResiduals(gamma_5)
#Plot residuals
plot(res_gamma_5)
#Test dispersion
print(testDispersion(res_gamma_5))
```

```
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## dispersion = 0.5957, p-value = 0.6
## alternative hypothesis: two.sided
```

```
#Test for significant outliers
print(testOutliers(res_gamma_5))
```

```
##
##  DHARMa outlier test based on exact binomial test with approximate
##  expectations
##
## data:  res_gamma_5
## outliers at both margin(s) = 0, observations = 41, p-value = 1
## alternative hypothesis: true probability of success is not equal to 0.007968127
## 95 percent confidence interval:
##  0.00000000 0.08604384
## sample estimates:
## frequency of outliers (expected: 0.00796812749003984 )
##                                                      0
```

```
#Check for collinearity among predictors
performance::check_collinearity(gamma_5)
```

```
## # Check for Multicollinearity
##
## Low Correlation
##
##          Term  VIF  VIF 95% CI adj. VIF Tolerance Tolerance 95% CI
##  Year_centered 1.00 [1.00, Inf]     1.00      1.00    [0.00, 1.00]
##     inoculation 1.00 [1.00, Inf]     1.00      1.00    [0.00, 1.00]
```

```
#Check residuals for each predictor
plotResiduals(res_gamma_5,SUBSET$Year_centered)
plotResiduals(res_gamma_5, SUBSET$inoculation)

#Outlier Sensitivity Test
#Filter out the outlier
filtered <- SUBSET %>% filter(Copies_ul <= 100)

#Create the model with the filtered dataset
if (nrow(filtered) >= 10) {
  filtered$Year_centered <- filtered$Year - 2014
  gamma_FILTERED <- glmer(Copies_ul ~ Year_centered + inoculation
                          + (1|SampleID),
                          family = Gamma(link = "log"),
                          data = filtered)
  summary(gamma_FILTERED)
}
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: Gamma  ( log )
## Formula: Copies_ul ~ Year_centered + inoculation + (1 | SampleID)
##    Data: filtered
##
##      AIC      BIC   logLik deviance df.resid
##    230.7    239.2   -110.4    220.7       35
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.9904 -0.5804 -0.2884  0.4266  2.6755
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  SampleID (Intercept) 0.8969   0.9470
##  Residual             0.9456   0.9724
## Number of obs: 40, groups:  SampleID, 12
##
## Fixed effects:
##               Estimate Std. Error t value Pr(>|z|)
## (Intercept)     2.7208     0.4813   5.653 1.58e-08 ***
## Year_centered  -0.8140     0.1374  -5.924 3.14e-09 ***
## inoculation1   -0.5340     0.6349  -0.841      0.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Yr_cnt
## Year_centrd -0.411
## inoculatin1 -0.640  0.038
```

```
#Removal of the possible outlier does not affect conclusions, model is robust to
#outlier inclusion
```

```
#Summarize final selected models
summary(binom_RI_2)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: Presence ~ Year_centered + inoculation + (1 | SampleID)
##    Data: data
##
##      AIC      BIC   logLik deviance df.resid
##     43.3     50.8    -17.6     35.3       44
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9797  0.1634  0.2788  0.3996  0.9336
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  SampleID (Intercept) 0.4132   0.6428
## Number of obs: 48, groups:  SampleID, 12
##
## Fixed effects:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)     3.8848     1.4325   2.712  0.00669 **
## Year_centered  -0.7198     0.4541  -1.585  0.11295
## inoculation1   -1.1856     1.0328  -1.148  0.25098
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Yr_cnt
## Year_centrd -0.760
## inoculatin1 -0.585  0.124
```

```
summary(gamma_5)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: Gamma  ( log )
## Formula: Copies_ul ~ Year_centered + inoculation + (1 | SampleID)
##    Data: SUBSET
##
##      AIC      BIC   logLik deviance df.resid
##     248.0    256.6   -119.0    238.0       36
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.9590 -0.5693 -0.2837  0.2808  2.4364
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  SampleID (Intercept) 1.205    1.098
```

```
##  Residual              1.015     1.007
## Number of obs: 41, groups:  SampleID, 12
##
## Fixed effects:
##              Estimate Std. Error t value Pr(>|z|)
## (Intercept)    2.8917     0.5237   5.522 3.35e-08 ***
## Year_centered -0.8409     0.1435  -5.858 4.69e-09 ***
## inoculation1  -0.6905     0.6925  -0.997    0.319
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) Yr_cnt
## Year_centrd -0.389
## inoculatin1 -0.646  0.021
```