Behavioral Tolerance, Vigor, and resilience

David R. Clark, Jason Walsman, Rachael Kramp, and Jessica Stephenson

2024-03-25

Contents

1 Introduction **1**

1.1 Overall summary 1

1.2 Quick experimental summary 1

1.3 Load in packages needed for analysis 1

1.4 Load in dataframe 1

1.5 Calculate metrics needed for further analyses 2

1.6 Description of data, structure, and type 18

1.7 Visualizing the relationships between variables in our dataset 19

1.8 Visualize some patterns in the raw data 20

1.9 What hypotheses we want to test with these data and what data we can use to test them? . . 21

2 Do infected and uninfected individuals differ in their average and variation in velocity? 21

2.1 Visually inspection of the explanatory variables that will be used in the analyses 21

2.2 Does infection or sex variation impact the average velocity of individauls? 22

2.3 Does infection or sex variation impact the variance in velocity of individauls? 28

3 Do we see differences in change in behavior over time based on infection status and sexual variation? 34

3.1 Visually inspection of the explanatory variables that will be used in the analyses 34

1 Introduction

1.1 Overall summary

1.2 Quick experimental summary

1.3 Load in packages needed for analysis

```r
# Clear the working environment
rm(list = ls())
# Visualization
library(ggplot2)
library(visreg)
source("http://highstat.com/Books/BGS/GAMM/RCodeP2/HighstatLibV6.R")
# (generalized) Linear mixed modeling
library(lme4)
library(glmmTMB)
library(lmodel2)
# Statistcal analysis reporting and model validation
library(performance)
library(car)
library(lmtest)
library(DHARMa)
# Data wrangling
library(dplyr)
library(plyr)
library(tidyverse)
library(tidylog)
```

1.4 Load in dataframe

```r
# Loading in data set
IndBehav <- read_csv("VIEBehavior_20210913_V2.csv", col_types = cols(fishID = col_character(),
    Infection = col_character(), InfDate = col_character(), BehavGroup = col_character(),
    dayofinf = col_character(), CountDay = col_character(), Behavdate = col_character(),
    TOD = col_character(), Duration = col_double(), Velocity = col_double(), Distance = col_double()))

# Filtering Data set so we can perform our analysis
IndBehav1 <- IndBehav %>%
    # Removing columns we dont need
select(-c(Blind.ID, InfLength, LateLength, dayofinf, Countintial, Notes, InfWeight))
```

```
## select: dropped 7 variables (Blind.ID, InfWeight, InfLength, LateLength, dayofinf, ...)
```

1.5 Calculate metrics needed for further analyses

1.5.1 Calculating all body condition metrics

```r
##### This code is just to examine how many fish cleared their infection
##### (13)#### IndBehavRec<- IndBehav1 %>% filter(Recov==1)%>% filter(Fish==1)


##### This code is just to examine how many fish died during their infection
##### (2)##### IndBehavDead<- IndBehav1 %>% filter(Died==1)%>% filter(Fish==1)

FemaleOnly <- IndBehav1 %>%
    filter(Sex == "F")
```

```
## filter: removed 541 rows (49%), 564 rows remaining
```

```
MaleOnly <- IndBehav1 %>%
    filter(Sex == "M")
```

```
## filter: removed 565 rows (51%), 540 rows remaining
```

```
# Calculating SMI metrics Preinfection SMI#####

# PreinfectionSMI for females To calculate the SMI we take the OLS slope
lmodel2(log(as.numeric(PreWeight) + 1) ~ log(as.numeric(PreLength) + 1), data = FemaleOnly)
```

```
## RMA was not requested: it will not be computed.

## No permutation test will be performed

##
## Model II regression
##
## Call: lmodel2(formula = log(as.numeric(PreWeight) + 1) ~
## log(as.numeric(PreLength) + 1), data = FemaleOnly)
##
## n = 564   r = 0.9103118   r-square = 0.8286675
## Parametric P-values:   2-tailed = 1.890218e-217    1-tailed = 9.451089e-218
## Angle between the two OLS regression lines = 4.099607 degrees
##
## Regression results
##   Method Intercept     Slope Angle (degrees) P-perm (1-tailed)
## 1    OLS -1.114762 0.4206982        22.81640               NA
## 2     MA -1.155627 0.4341311        23.46717               NA
## 3    SMA -1.240856 0.4621474        24.80390               NA
##
## Confidence intervals
##   Method 2.5%-Intercept 97.5%-Intercept 2.5%-Slope 97.5%-Slope
## 1    OLS      -1.163000       -1.066525  0.4048487   0.4365478
## 2     MA      -1.205686       -1.106162  0.4178711   0.4505866
## 3    SMA      -1.289899       -1.193466  0.4465696   0.4782687
##
## Eigenvalues: 0.009625885 0.0002518475
##
## H statistic used for computing C.I. of MA: 0.0001893902
```

```
# Take the median length from summary
summary(FemaleOnly$PreLength)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   16.60   18.40   20.20   20.03   21.20   24.20
```

```
# Adding PreSMI to Dataframe
FemaleOnly <- FemaleOnly %>%
    group_by(fishID) %>%
    mutate(PreSMI = PreWeight * ((20.2/PreLength)^0.4206982))
```

```
## group_by: one grouping variable (fishID)

## mutate (grouped): new variable 'PreSMI' (double) with 47 unique values and 0% NA
```

```
# PreinfectionSMI for males
lmodel2(log(as.numeric(PreWeight) + 1) ~ log(as.numeric(PreLength) + 1), data = MaleOnly)
```

```
## RMA was not requested: it will not be computed.

## No permutation test will be performed

##
## Model II regression
##
## Call: lmodel2(formula = log(as.numeric(PreWeight) + 1) ~
## log(as.numeric(PreLength) + 1), data = MaleOnly)
##
## n = 540   r = 0.7784225   r-square = 0.6059415
## Parametric P-values:   2-tailed = 7.086423e-111    1-tailed = 3.543211e-111
## Angle between the two OLS regression lines = 6.76859 degrees
##
## Regression results
##   Method  Intercept     Slope Angle (degrees) P-perm (1-tailed)
## 1    OLS -0.4620565 0.1938185        10.96898                NA
## 2     MA -0.4750867 0.1984879        11.22660                NA
## 3    SMA -0.6160126 0.2489888        13.98170                NA
##
## Confidence intervals
##   Method 2.5%-Intercept 97.5%-Intercept 2.5%-Slope 97.5%-Slope
## 1    OLS     -0.4990055      -0.4251076  0.1805813   0.2070557
## 2     MA     -0.5130172      -0.4373522  0.1849657   0.2120803
## 3    SMA     -0.6539329      -0.5800547  0.2361033   0.2625776
##
## Eigenvalues: 0.00431447 9.773707e-05
##
## H statistic used for computing C.I. of MA: 0.0001701002
```

```
# Take the median length from summary
summary(MaleOnly$PreLength)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.70   14.50   15.30   15.32   16.00   18.50
```

```
# Adding PreSMI to Dataframe
MaleOnly <- MaleOnly %>%
    group_by(fishID) %>%
    mutate(PreSMI = PreWeight * ((15.3/PreLength)^0.1938185))
```

```
## group_by: one grouping variable (fishID)

## mutate (grouped): new variable 'PreSMI' (double) with 44 unique values and 0% NA
```

```r
##### Late Infection SMI####

# Late infection SMI for females Take OLS slope from this model
lmodel2(log(as.numeric(LateWeight) + 1) ~ log(as.numeric(PreLength) + 1), data = FemaleOnly)
```

```
## RMA was not requested: it will not be computed.

## No permutation test will be performed

##
## Model II regression
##
## Call: lmodel2(formula = log(as.numeric(LateWeight) + 1) ~
## log(as.numeric(PreLength) + 1), data = FemaleOnly)
##
## n = 516   r = -0.127178   r-square = 0.01617424
## Parametric P-values:   2-tailed = 0.003807557    1-tailed = 0.001903778
## Angle between the two OLS regression lines = 57.58662 degrees
##
## Regression results
##   Method  Intercept       Slope Angle (degrees) P-perm (1-tailed)
## 1    OLS   2.033718  -0.5976124       -30.86306                NA
## 2     MA 107.560934 -35.3034151       -88.37748                NA
## 3    SMA  14.504560  -4.6990243       -77.98610                NA
##
## Confidence intervals
##   Method 2.5%-Intercept 97.5%-Intercept  2.5%-Slope 97.5%-Slope
## 1    OLS      0.8050897        3.262347   -1.001497  -0.1937279
## 2     MA     64.2522935      331.105245 -108.822693 -21.0600644
## 3    SMA     13.3291795       15.785299   -5.120234  -4.3124649
##
## Eigenvalues: 0.1896132 0.008435402
##
## H statistic used for computing C.I. of MA: 0.0003658854
```

```r
# Take the median length from summary
summary(FemaleOnly$PreLength)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   16.60   18.40   20.20   20.03   21.20   24.20
```

```r
# Adding PreSMI to Dataframe
FemaleOnly <- FemaleOnly %>%
    group_by(fishID) %>%
    mutate(LateSMI = PreWeight * ((20.1/PreLength)^-0.5976124))
```

```
## group_by: one grouping variable (fishID)

## mutate (grouped): new variable 'LateSMI' (double) with 47 unique values and 0% NA
```

```r
# Late Infection SMI for males Take OLS slope from this model
lmodel2(log(as.numeric(LateWeight) + 1) ~ log(as.numeric(PreLength) + 1), data = MaleOnly)
```

```
## RMA was not requested: it will not be computed.

## No permutation test will be performed

##
## Model II regression
##
## Call: lmodel2(formula = log(as.numeric(LateWeight) + 1) ~
## log(as.numeric(PreLength) + 1), data = MaleOnly)
##
## n = 456   r = -0.06095677   r-square = 0.003715728
## Parametric P-values:   2-tailed = 0.1938354    1-tailed = 0.09691771
## Angle between the two OLS regression lines = 64.27059 degrees
##
## Regression results
##   Method  Intercept       Slope Angle (degrees) P-perm (1-tailed)
## 1    OLS   1.492739   -0.472241       -25.27860                NA
## 2     MA 348.385697 -124.982883       -89.54158                NA
## 3    SMA  21.760991   -7.747145       -82.64494                NA
##
## Confidence intervals
##   Method 2.5%-Intercept 97.5%-Intercept 2.5%-Slope 97.5%-Slope
## 1    OLS     -0.4947141        3.480193  -1.185443   0.2409615
## 2     MA    138.8775568     -682.296809 244.961036 -49.7839131
## 3    SMA     19.8652425       23.839279  -8.493107  -7.0667022
##
## Eigenvalues: 0.204658 0.003396832
##
## H statistic used for computing C.I. of MA: 0.0001459962
```

```r
# Take the median length from summary
summary(MaleOnly$PreLength)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.70   14.50   15.30   15.32   16.00   18.50
```

```r
# Adding PreSMI to Dataframe
MaleOnly <- MaleOnly %>%
    group_by(fishID) %>%
    mutate(LateSMI = PreWeight * ((15.3/PreLength)^-0.472241))
```

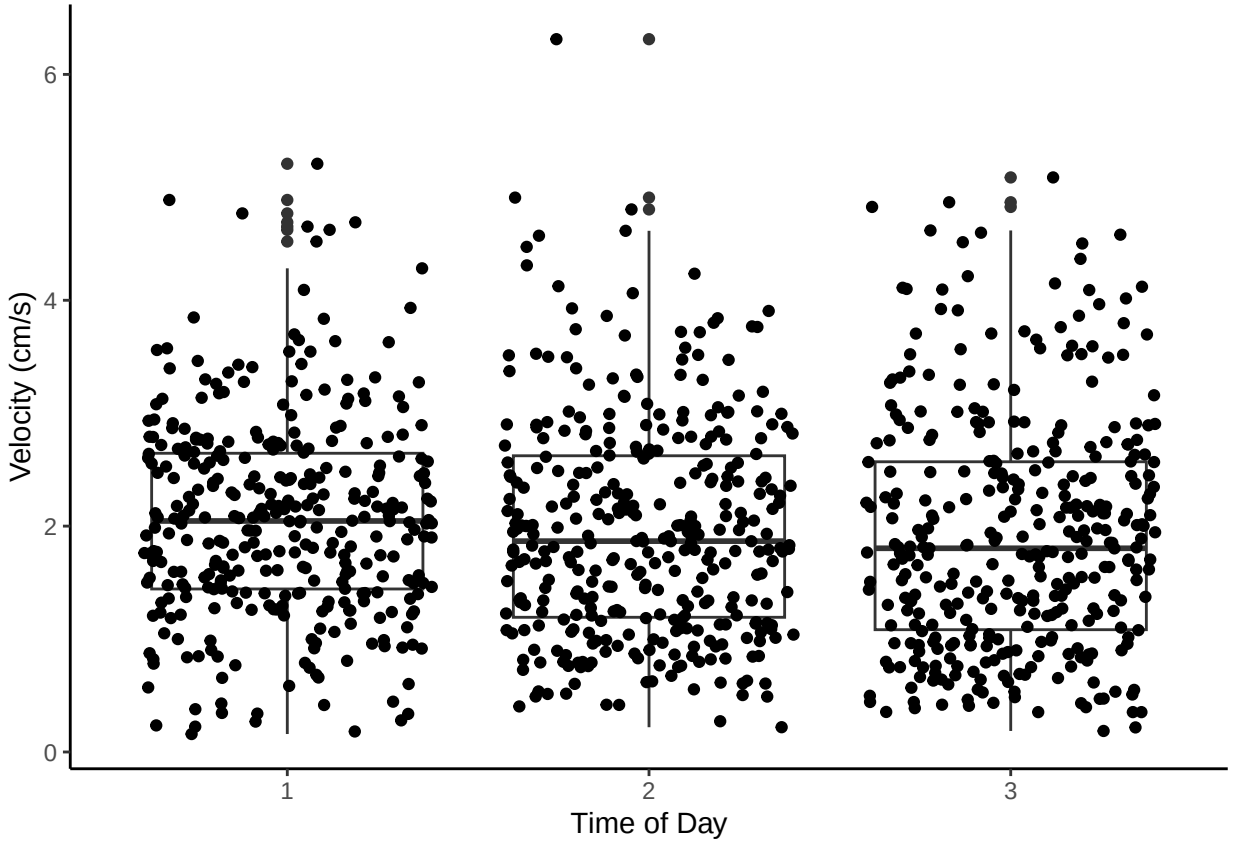
```
## group_by: one grouping variable (fishID)

## mutate (grouped): new variable 'LateSMI' (double) with 44 unique values and 0% NA
```

```
#####

# Bringing the male and female dataframes together to create our overall
# dataframe
IndBehav2 <- rbind(FemaleOnly, MaleOnly)


# View(IndBehav2)
```

1.5.2 Checking that fish velocity is not different at different times of day.

We recorded the velocity of the fish at 3 time periods over the course of the day. Once in the monring between 0900-1100, once in the afternoon 1200-1400, 1500-1700. Therefore we wanted to confirm that there is no consistent Time of day effect before we collapse the behavioral measurements into means for each day.

```
##### First visually look at the TOD effect for behavior####
ggplot(IndBehav2, aes(TOD, Velocity)) + geom_boxplot() + geom_jitter() + xlab("Time of Day") +
    ylab("Velocity (cm/s)") + theme_classic()
```

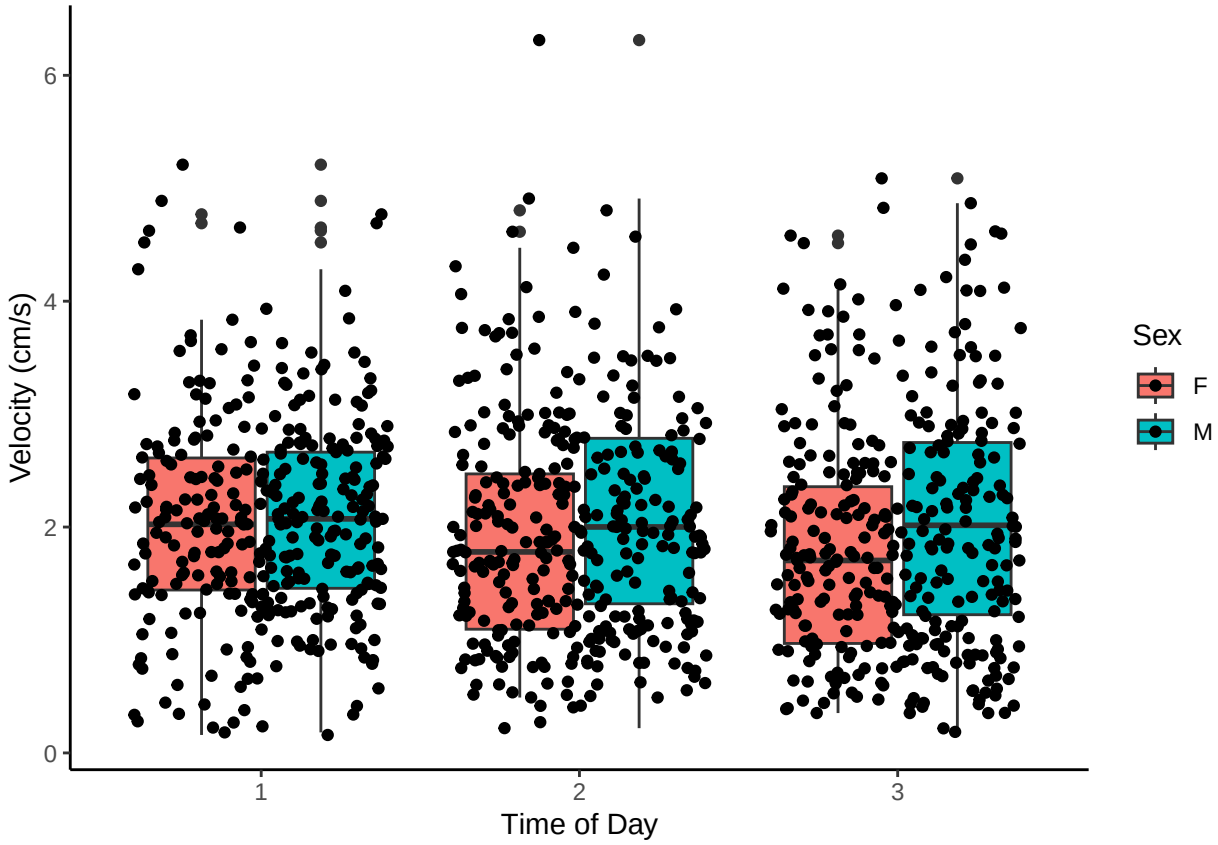



```
#### Verify that this is not statistically significant#####
anova(lm(Velocity ~ TOD, IndBehav2))
```

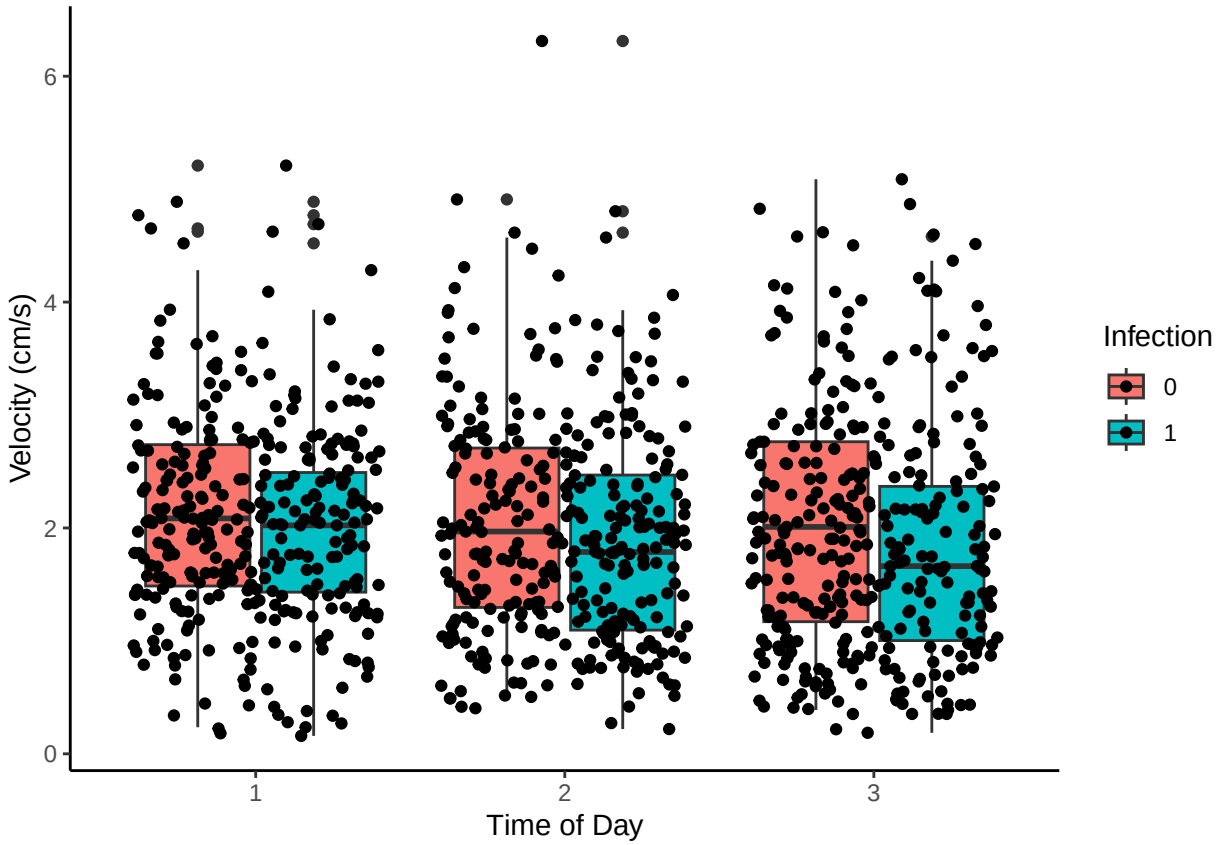
```
## Analysis of Variance Table
##
```

```
## Response: Velocity
##              Df Sum Sq Mean Sq F value Pr(>F)
## TOD           2   3.34 1.67073  1.7383 0.1763
## Residuals 1017 977.45 0.96111
```

```
##### Lets also confirm this does not change by sex####
ggplot(IndBehav2, aes(TOD, Velocity, fill = Sex)) + geom_boxplot() + geom_jitter() +
    xlab("Time of Day") + ylab("Velocity (cm/s)") + theme_classic()
```




```
##### Lets also confirm this does n0t change by Infection Status####
ggplot(IndBehav2, aes(TOD, Velocity, fill = Infection)) + geom_boxplot() + geom_jitter() +
    xlab("Time of Day") + ylab("Velocity (cm/s)") + theme_classic()
```



Visually there does not seem to be a distinguished pattern overall and an anova confirms it. It also does not seem to visually differ by sex or infection status. Therefore we can collapse the three different time points within a day to a mean velocity and variance of velocity across each time point, Before, Early, Late, and Later.

1.5.3 Calculating the behavioral tolerance metrics needed for analyses

We are calculating behavioral tolerance through two metrics. First is a measure of behavioral tolerance using linear mixed models using slope random effects for fishID. This allows us to extract each individuals change in activity with parasite burden and therefore allow us to calculate the tolerance (fishID slope) for each individual. This also allows us to use the random effect intercept for fishID as a measure of behavioral vigor. The second way we could calculate these metrics are a measure of change in behavior between two points of behavior (often referred to as point tolerance, CITATION FOR THIS). This is calculated in the second half of the code as the change in behavior between pre-infection and early, late, and later points of infection.

```
# Calulating the behavioral tolerance metrics were interested in for our
# analysis

# Random slope from random effects model Random effects model with random slope
# for each fish by worm burden prior to measuring its activity. This slope is
# the tolerance an individual has to changing its behavior as infection
# increases. The slope of the behavior is the individuals behavioral vigor, or
# its pre-infection behavior#####
BehavTolLM <- lmer(Velocity ~ (Wormbf | fishID), IndBehav2)
summary(BehavTolLM)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Velocity ~ (Wormbf | fishID)
##    Data: IndBehav2
##
## REML criterion at convergence: 1363
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7511 -0.6468 -0.0308  0.5698  4.3071
##
## Random effects:
##  Groups   Name        Variance  Std.Dev. Corr
##  fishID   (Intercept) 3.172e-01 0.563247
##           Wormbf      9.477e-06 0.003078 -0.17
##  Residual             5.781e-01 0.760314
## Number of obs: 541, groups:  fishID, 48
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   1.8515     0.0876   21.14
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.345298 (tol = 0.002, component 1)
## Model is nearly unidentifiable: very large eigenvalue
##  - Rescale variables?
```

```
##### Extracting the intercept and slope from the linear mixed model####
betaTol <- data.frame(coef(BehavTolLM)$fishID)
##### Renaming the slope and intercept to align with our metrics#####
betaTol <- betaTol %>%
    rename(BehavTol = Wormbf, BehavVig = X.Intercept.)
```

```
## rename: renamed 2 variables (BehavTol, BehavVig)
```

```
##### Creating a fishID column for combination with future dataframe####
betaTol <- cbind(fishID = rownames(betaTol), betaTol)

# Calculating the mean and variance in velocity per day, per fish

# We need to split the dataframes into the differnet time periods to calculate
# averages since mutate isnt splitting by TrialTime.

##### creating a dataframe to calculate the mean and variance of velocity for
##### before measurements.####

InBehavBefore <- IndBehav2 %>%
    drop_na(Velocity) %>%
    filter(TrialTime == "Before") %>%
    select(-c(InfDiff, Fish, Behavdate, Duration, Distance, comp3, Totworm, rate05,
        rate012, rate512, rate1216, rate18end, TOD, day, linf, maxlinf, maxworm,
        peakday, Ddate, Rdate, CountDay)) %>%
    mutate(AvgVel = mean(Velocity), VarVel = var(Velocity)) %>%
    distinct(fishID, .keep_all = TRUE)
```

```
## drop_na (grouped): removed 84 rows (8%), 1,020 rows remaining
```

```
## filter (grouped): removed 750 rows (74%), 270 rows remaining

## select: dropped 21 variables (Fish, CountDay, day, Totworm, comp3, ...)

## mutate (grouped): new variable 'AvgVel' (double) with 91 unique values and 0% NA

##                   new variable 'VarVel' (double) with 91 unique values and 0% NA

## distinct (grouped): removed 179 rows (66%), 91 rows remaining
```

```
##### creating a dataframe to calculate the mean and variance of velocity for
##### early measurements.#####

InBehavEarly <- IndBehav2 %>%
    drop_na(Velocity) %>%
    filter(TrialTime == "Early") %>%
    select(-c(InfDiff, Fish, Behavdate, Duration, Distance, comp3, Totworm, rate05,
        rate012, rate512, rate1216, rate18end, TOD, day, linf, maxlinf, maxworm,
        peakday, Ddate, Rdate, CountDay)) %>%
    mutate(AvgVel = mean(Velocity), VarVel = var(Velocity)) %>%
    distinct(fishID, .keep_all = TRUE)
```

```
## drop_na (grouped): removed 84 rows (8%), 1,020 rows remaining

## filter (grouped): removed 760 rows (75%), 260 rows remaining

## select: dropped 21 variables (Fish, CountDay, day, Totworm, comp3, ...)

## mutate (grouped): new variable 'AvgVel' (double) with 87 unique values and 0% NA

##                   new variable 'VarVel' (double) with 87 unique values and 0% NA

## distinct (grouped): removed 173 rows (67%), 87 rows remaining
```

```
##### creating a dataframe to calculate the mean and variance of velocity for
##### Late measurements. #####

InBehavLate <- IndBehav2 %>%
    drop_na(Velocity) %>%
    filter(TrialTime == "Late") %>%
    select(-c(InfDiff, Fish, Behavdate, Duration, Distance, comp3, Totworm, rate05,
        rate012, rate512, rate1216, rate18end, TOD, day, linf, maxlinf, maxworm,
        peakday, Ddate, Rdate, CountDay)) %>%
    mutate(AvgVel = mean(Velocity), VarVel = var(Velocity)) %>%
    distinct(fishID, .keep_all = TRUE)
```

```
## drop_na (grouped): removed 84 rows (8%), 1,020 rows remaining

## filter (grouped): removed 761 rows (75%), 259 rows remaining
```

```
## select: dropped 21 variables (Fish, CountDay, day, Totworm, comp3, ...)

## mutate (grouped): new variable 'AvgVel' (double) with 87 unique values and 0% NA

##                   new variable 'VarVel' (double) with 87 unique values and 0% NA

## distinct (grouped): removed 172 rows (66%), 87 rows remaining
```

```
##### creating a dataframe to calculate the mean and variance of velocity for
##### Later measurements.####

InBehavLater <- IndBehav2 %>%
    drop_na(Velocity) %>%
    filter(TrialTime == "Later") %>%
    select(-c(InfDiff, Fish, Behavdate, Duration, Distance, comp3, Totworm, rate05,
        rate012, rate512, rate1216, rate18end, TOD, day, linf, maxlinf, maxworm,
        peakday, Ddate, Rdate, CountDay)) %>%
    mutate(AvgVel = mean(Velocity), VarVel = var(Velocity)) %>%
    distinct(fishID, .keep_all = TRUE)
```

```
## drop_na (grouped): removed 84 rows (8%), 1,020 rows remaining

## filter (grouped): removed 789 rows (77%), 231 rows remaining

## select: dropped 21 variables (Fish, CountDay, day, Totworm, comp3, ...)

## mutate (grouped): new variable 'AvgVel' (double) with 79 unique values and 0% NA

##                   new variable 'VarVel' (double) with 79 unique values and 0% NA

## distinct (grouped): removed 152 rows (66%), 79 rows remaining
```

```
#### Adding these dataframes together to create one large one####

IndBehav3 <- rbind(InBehavBefore, InBehavEarly, InBehavLate, InBehavLater)


##### Pivoting wider so I can subtract velocity changes between periods to get
##### change in behavior####
IndBehav4 <- IndBehav3 %>%
    group_by(TrialTime, fishID, add = TRUE) %>%
    select(-Velocity) %>%
    pivot_wider(names_from = TrialTime, values_from = c(Wormbf, Ratebf, AvgVel, VarVel))
```

```
## group_by: 2 grouping variables (fishID, TrialTime)

## select: dropped one variable (Velocity)

## pivot_wider: reorganized (Wormbf, TrialTime, Ratebf, AvgVel, VarVel) into (Wormbf_Before, Wormbf_Earl
```

```
##### subset down to infected individuals only and calculate point tolerance
##### metrics for each timeframe. #####
IndBehav4Inf <- IndBehav4 %>%
    filter(Infection == 1) %>%
    mutate(EChBe = AvgVel_Early - AvgVel_Before, LChBe = AvgVel_Late - AvgVel_Before,
        LtrChBe = AvgVel_Later - AvgVel_Before, ERatebf = (Wormbf_Early - Wormbf_Before)/6,
        LRatebf = (Wormbf_Late - Wormbf_Early)/6, LtrRatebf = (Wormbf_Later - Wormbf_Late)/6) %>%
    pivot_longer(cols = starts_with("Wormbf"), names_to = "TrialTime", names_prefix = "Wormbf_",
        values_to = "Wormbf") %>%
    select(-c(AvgVel_Before, AvgVel_Early, AvgVel_Late, AvgVel_Later, Ratebf_Before,
        Ratebf_Early, Ratebf_Late, Ratebf_Later)) %>%
    filter(fishID != c(44, 118, 139, 94))
```

```
## filter (grouped): removed 39 rows (43%), 52 rows remaining

## mutate (grouped): new variable 'EChBe' (double) with 49 unique values and 8% NA

##                   new variable 'LChBe' (double) with 49 unique values and 8% NA

##                   new variable 'LtrChBe' (double) with 42 unique values and 21% NA

##                   new variable 'ERatebf' (double) with 23 unique values and 10% NA

##                   new variable 'LRatebf' (double) with 42 unique values and 10% NA

##                   new variable 'LtrRatebf' (double) with 37 unique values and 23% NA

## pivot_longer: reorganized (Wormbf_Before, Wormbf_Early, Wormbf_Late, Wormbf_Later) into (TrialTime, W

## select: dropped 8 variables (Ratebf_Before, Ratebf_Early, Ratebf_Late, Ratebf_Later, AvgVel_Before,

## filter (grouped): removed 4 rows (2%), 204 rows remaining
```

```
##### Make two dataframes with pivot longer for wormbf and velocity then merge
##### them together on fishID and TrialTime####
IndBehav4InfV <- IndBehav4 %>%
    filter(Infection == 1) %>%
    pivot_longer(cols = starts_with("AvgVel"), names_to = "TrialTime", names_prefix = "AvgVel_",
        values_to = "AvgVel") %>%
    select(fishID, AvgVel, TrialTime) %>%
    filter(fishID != c(44, 118, 139, 94))
```

```
## filter (grouped): removed 39 rows (43%), 52 rows remaining

## pivot_longer: reorganized (AvgVel_Before, AvgVel_Early, AvgVel_Late, AvgVel_Later) into (TrialTime, A

## select: dropped 24 variables (Sex, PreWeight, PreLength, Treatment, Infection, ...)

## filter (grouped): removed 4 rows (2%), 204 rows remaining
```

```
##### Making the growth rate columns into one column for infected
##### individuals####
IndBehavGRInf <- IndBehav4 %>%
    filter(Infection == 1) %>%
    pivot_longer(cols = starts_with("Ratebf"), names_to = "TrialTime", names_prefix = "Ratebf_",
        values_to = "Ratebf") %>%
    select(fishID, TrialTime, Ratebf)
```

```
## filter (grouped): removed 39 rows (43%), 52 rows remaining

## pivot_longer: reorganized (Ratebf_Before, Ratebf_Early, Ratebf_Late, Ratebf_Later) into (TrialTime, R

## select: dropped 24 variables (Sex, PreWeight, PreLength, Treatment, Infection, ...)
```

```
##### Making the new growth rate columns into one column for infected
##### individuals#####
IndBehavNGRInf <- IndBehav4Inf %>%
    select(fishID, ERatebf, LRatebf, LtrRatebf) %>%
    rename(NRatebf_Early = ERatebf, NRatebf_Late = LRatebf, NRatebf_Later = LtrRatebf) %>%
    add_column(NRatebf_Before = 0) %>%
    pivot_longer(cols = starts_with("NRatebf"), names_to = "TrialTime", names_prefix = "NRatebf_",
        values_to = "NRatebf") %>%
    select(fishID, TrialTime, NRatebf) %>%
    distinct(fishID, TrialTime, .keep_all = TRUE)
```

```
## select: dropped 21 variables (Sex, PreWeight, PreLength, Treatment, Infection, ...)

## rename: renamed 3 variables (NRatebf_Early, NRatebf_Late, NRatebf_Later)

## pivot_longer: reorganized (NRatebf_Early, NRatebf_Late, NRatebf_Later, NRatebf_Before) into (TrialTi

## select: no changes

## distinct (grouped): removed 608 rows (75%), 208 rows remaining
```

```
##### Making the Variance columns into one column for infected individuals#####
IndBehavVarInf <- IndBehav4Inf %>%
    select(-TrialTime) %>%
    pivot_longer(cols = starts_with("VarVel"), names_to = "TrialTime", names_prefix = "VarVel_",
        values_to = "VarVel") %>%
    select(fishID, TrialTime, VarVel) %>%
    distinct(fishID, TrialTime, .keep_all = TRUE)
```

```
## select: dropped one variable (TrialTime)

## pivot_longer: reorganized (VarVel_Before, VarVel_Early, VarVel_Late, VarVel_Later) into (TrialTime, V

## select: dropped 19 variables (Sex, PreWeight, PreLength, Treatment, Infection, ...)

## distinct (grouped): removed 608 rows (75%), 208 rows remaining
```

```
##### Merging the tolerance and vigor metrics ####
IndBehav4Inf1 <- merge(IndBehav4Inf, betaTol, by.x = "fishID", by.y = "fishID", .keep = all)

##### Merge the two dataframes together based on the fishID and TrialTime to
##### get all variables we want into one dataframe#####
IndBehav5Inf <- merge(IndBehav4InfV, IndBehav4Inf1, by.x = c("fishID", "TrialTime"),
    by.y = c("fishID", "TrialTime"), all = TRUE)
IndBehav5Inf2 <- merge(IndBehav5Inf, IndBehavGRInf, by.x = c("fishID", "TrialTime"),
    by.y = c("fishID", "TrialTime"), all = TRUE)
IndBehav5Inf3 <- merge(IndBehav5Inf2, IndBehavNGRInf, by.x = c("fishID", "TrialTime"),
    by.y = c("fishID", "TrialTime"), all = TRUE)
IndBehav5Inf4 <- merge(IndBehav5Inf3, IndBehavVarInf, by.x = c("fishID", "TrialTime"),
    by.y = c("fishID", "TrialTime"), all = TRUE)
# Subset down to uninfected individuals and calculate their point tolerance
# metrics Creating a dataframe for calculating the tolerance metrics and then
# pivot longer by Wormbf####
IndBehav4Unf <- IndBehav4 %>%
    filter(Infection == 0) %>%
    mutate(EChBe = AvgVel_Early - AvgVel_Before, LChBe = AvgVel_Late - AvgVel_Before,
        LtrChBe = AvgVel_Later - AvgVel_Before, NRatebf_Early = (Wormbf_Early - Wormbf_Before)/5,
        NRatebf_Late = (Wormbf_Late - Wormbf_Early)/5, NRatebf_Later = (Wormbf_Later -
            Wormbf_Late)/5, ) %>%
    pivot_longer(cols = starts_with(c("Wormbf")), names_to = "TrialTime", names_prefix = c("Wormbf_"),
        values_to = "Wormbf") %>%
    select(-c(AvgVel_Before, AvgVel_Early, AvgVel_Late, AvgVel_Later, Ratebf_Before,
        Ratebf_Early, Ratebf_Late, Ratebf_Later))
```

```
## filter (grouped): removed 52 rows (57%), 39 rows remaining

## mutate (grouped): new variable 'EChBe' (double) with 39 unique values and 0% NA

##                   new variable 'LChBe' (double) with 39 unique values and 0% NA

##                   new variable 'LtrChBe' (double) with 39 unique values and 3% NA

##                   new variable 'NRatebf_Early' (double) with one unique value and 100% NA

##                   new variable 'NRatebf_Late' (double) with one unique value and 100% NA

##                   new variable 'NRatebf_Later' (double) with one unique value and 100% NA

## pivot_longer: reorganized (Wormbf_Before, Wormbf_Early, Wormbf_Late, Wormbf_Later) into (TrialTime, W

## select: dropped 8 variables (Ratebf_Before, Ratebf_Early, Ratebf_Late, Ratebf_Later, AvgVel_Before,
```

```
##### Creating a second dataframe where we pivot longer for velocity####
IndBehav4UnfV <- IndBehav4 %>%
    filter(Infection == 0) %>%
    pivot_longer(cols = starts_with("AvgVel"), names_to = "TrialTime", names_prefix = "AvgVel_",
        values_to = "AvgVel") %>%
    select(fishID, TrialTime, AvgVel)
```

```
## filter (grouped): removed 52 rows (57%), 39 rows remaining

## pivot_longer: reorganized (AvgVel_Before, AvgVel_Early, AvgVel_Late, AvgVel_Later) into (TrialTime, A

## select: dropped 24 variables (Sex, PreWeight, PreLength, Treatment, Infection, ...)
```

```
##### Making the growth rate columns into one column for uninfected
##### individuals#####
IndBehavGRUnf <- IndBehav4 %>%
    filter(Infection == 0) %>%
    pivot_longer(cols = starts_with("Ratebf"), names_to = "TrialTime", names_prefix = "Ratebf_",
        values_to = "Ratebf") %>%
    select(fishID, TrialTime, Ratebf)
```

```
## filter (grouped): removed 52 rows (57%), 39 rows remaining

## pivot_longer: reorganized (Ratebf_Before, Ratebf_Early, Ratebf_Late, Ratebf_Later) into (TrialTime, R

## select: dropped 24 variables (Sex, PreWeight, PreLength, Treatment, Infection, ...)
```

```
##### Making the new growth rate columns into one column for infected
##### individuals#####
IndBehavNGRUnf <- IndBehav4Unf %>%
    select(-TrialTime) %>%
    add_column(NRatebf_Before = 0) %>%
    pivot_longer(cols = starts_with("NRatebf"), names_to = "TrialTime", names_prefix = "NRatebf_",
        values_to = "NRatebf") %>%
    select(fishID, TrialTime, NRatebf) %>%
    distinct(fishID, TrialTime, .keep_all = TRUE)
```

```
## select: dropped one variable (TrialTime)

## pivot_longer: reorganized (NRatebf_Early, NRatebf_Late, NRatebf_Later, NRatebf_Before) into (TrialTi

## select: dropped 20 variables (Sex, PreWeight, PreLength, Treatment, Infection, ...)

## distinct (grouped): removed 468 rows (75%), 156 rows remaining
```

```
##### Making the Variance columns into one column for infected individuals#####
IndBehavVarUnf <- IndBehav4Unf %>%
    select(-TrialTime) %>%
    pivot_longer(cols = starts_with("VarVel"), names_to = "TrialTime", names_prefix = "VarVel_",
        values_to = "VarVel") %>%
    select(fishID, TrialTime, VarVel) %>%
    distinct(fishID, TrialTime, .keep_all = TRUE)
```

```
## select: dropped one variable (TrialTime)

## pivot_longer: reorganized (VarVel_Before, VarVel_Early, VarVel_Late, VarVel_Later) into (TrialTime, V

## select: dropped 19 variables (Sex, PreWeight, PreLength, Treatment, Infection, ...)

## distinct (grouped): removed 468 rows (75%), 156 rows remaining
```

```r
##### Merge the different datasets together#####
IndBehav5Unf <- merge(IndBehav4UnfV, IndBehav4Unf, by.x = c("fishID", "TrialTime"),
    by.y = c("fishID", "TrialTime"), all = TRUE)
IndBehav5Unf2 <- merge(IndBehav5Unf, IndBehavGRUnf, by.x = c("fishID", "TrialTime"),
    by.y = c("fishID", "TrialTime"), all = TRUE)
IndBehav5Unf3 <- merge(IndBehav5Unf2, IndBehavNGRUnf, by.x = c("fishID", "TrialTime"),
    by.y = c("fishID", "TrialTime"), all = TRUE)
IndBehav5Unf4 <- merge(IndBehav5Unf3, IndBehavVarUnf, by.x = c("fishID", "TrialTime"),
    by.y = c("fishID", "TrialTime"), all = TRUE)

# Creating the BehavTol and BehavVig columns for the uninfected and calculating
# the metric for uninfected individual so we can combine these dataframes
# together
IndBehav5Unf4 <- IndBehav5Unf4 %>%
    group_by(fishID) %>%
    add_column(BehavTol = 0, BehavVig = 0)
```

```
## group_by: one grouping variable (fishID)
```

```r
# Rbinding our infected and uninfected dataframes together into one overall
# dataframe
IndBehav5 <- rbind(IndBehav5Unf4, IndBehav5Inf4)

## Pivot longer for Change in behavior
IndBehav5ChBe <- IndBehav5 %>%
    select(-TrialTime) %>%
    rename(ChBehav_Early = EChBe, ChBehav_Late = LChBe, ChBehav_Later = LtrChBe) %>%
    add_column(ChBehav_Before = 0) %>%
    pivot_longer(cols = starts_with("ChBehav"), names_to = "TrialTime", names_prefix = "ChBehav_",
        values_to = "ChBehav") %>%
    select(fishID, TrialTime, ChBehav) %>%
    distinct(fishID, TrialTime, .keep_all = TRUE)
```

```
## select: dropped one variable (TrialTime)

## rename: renamed 3 variables (ChBehav_Early, ChBehav_Late, ChBehav_Later)

## pivot_longer: reorganized (ChBehav_Early, ChBehav_Late, ChBehav_Later, ChBehav_Before) into (TrialTi

## select: dropped 29 variables (AvgVel, Sex, PreWeight, PreLength, Treatment, ...)

## distinct (grouped): removed 1,092 rows (75%), 364 rows remaining
```

```r
# Merge dataframes together for one large dataframe
IndBehav6 <- merge(IndBehav5, IndBehav5ChBe, by.x = c("fishID", "TrialTime"), by.y = c("fishID",
    "TrialTime"), all = TRUE)

# Removing some of the columns we dont want in the dataframe Create a point
# tolerance metric for each timepoint
IndBehav6 <- IndBehav6 %>%
    select(-c(ERatebf, LRatebf, LtrRatebf, NRatebf_Early, NRatebf_Late, NRatebf_Later,
```

```
        VarVel_Before, VarVel_Early, VarVel_Late, VarVel_Later)) %>%
    mutate(PTol = ChBehav/Wormbf) %>%
    mutate(PTol = replace(PTol, PTol == NaN, 0)) %>%
    mutate(PTol = replace(PTol, PTol == -Inf, 0)) %>%
    mutate(PTol = replace(PTol, PTol == Inf, 0)) %>%
    select(-c(EChBe, LChBe, LtrChBe)) %>%
    distinct(fishID, TrialTime, .keep_all = TRUE) %>%
    drop_na(Sex)
```

```
## select: dropped 10 variables (VarVel_Before, VarVel_Early, VarVel_Late, VarVel_Later, NRatebf_Early,

## mutate: new variable 'PTol' (double) with 123 unique values and 63% NA

## mutate: no changes

## mutate: changed 10 values (3%) of 'PTol' (0 new NA)

## mutate: changed 5 values (1%) of 'PTol' (0 new NA)

## select: dropped 3 variables (EChBe, LChBe, LtrChBe)

## distinct: no rows removed

## drop_na: removed 17 rows (5%), 347 rows remaining
```

```
# Remove dataframes used for data wrangling and managing rm(c(IndBehav,
# IndBehav1,IndBehav2,IndBehav3,IndBehav3Inf1,IndBehav4,IndBehav4Inf,IndBehav4Inf1,IndBehav4InfV,IndBeh

# Setting some characters as factors
IndBehav6$TrialTime <- as.factor(IndBehav6$TrialTime)
IndBehav6$Sex <- as.factor(IndBehav6$Sex)
IndBehav6$Infection <- as.factor(IndBehav6$Infection)
IndBehav6$ContInf <- as.factor(IndBehav6$ContInf)
IndBehav6$Died <- as.factor(IndBehav6$Died)
IndBehav6$Treatment <- as.factor(IndBehav6$Treatment)


# #Exporting dataframe for saving
write.csv(IndBehav6, "IndividualBehaviors_20240416.csv")
```

Now we have to decided how best to calculate whether an individual has recovered. In the experiment only had 13 fish recover from infection totally, therefore we are thinking of another metric that indicates a fish is in recovery during infection.

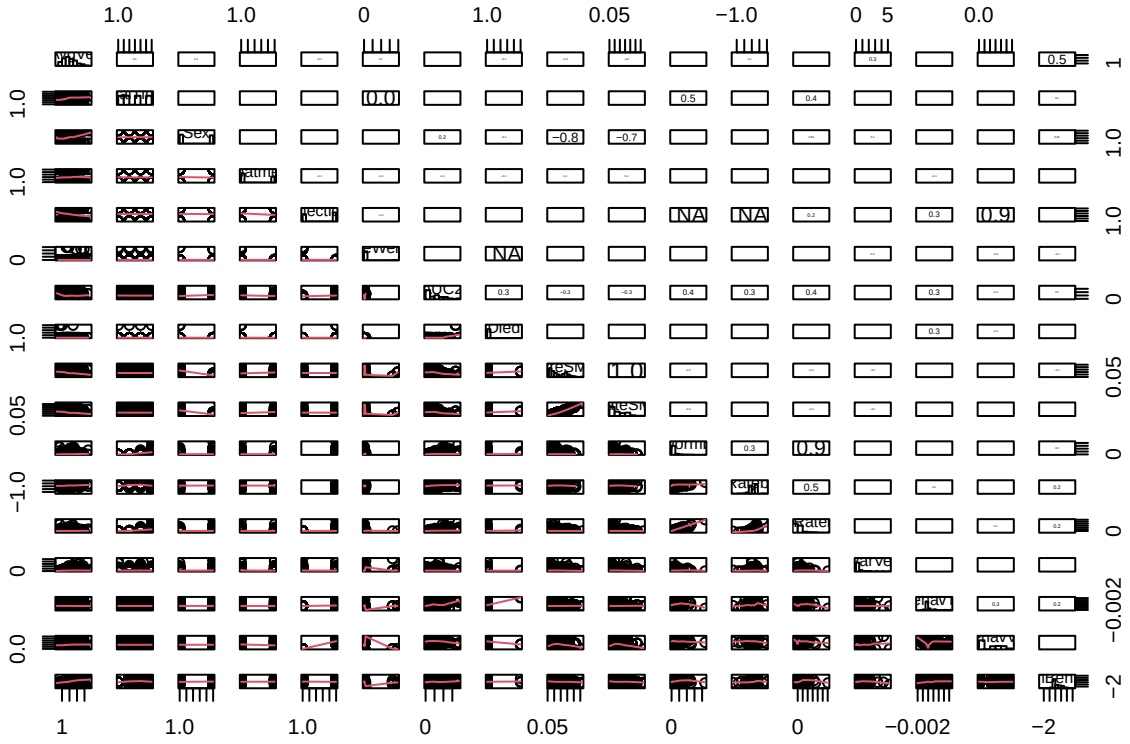
1.6 Description of data, structure, and type

fishID: The individual ID of each fish used in the trial.
TrialTime: The point of infection where each behavior trial happened. Early - early infection (6 days), Late - late in infection (12 days), Later - later in infection (18 days)
AvgVel: The average velocity from three separate behavioral trials of each fish for each of the Trial Times.

(cm/s)
Sex: The sex of the individual. F - female, M - male
PreWeight: The weight of the individual prior to their fist behavior trial and pre-infection. $(grams)$
PreLength: The length of the individual prior to their first behavior trial and pre-infection (mm)
Treatment: What treatment the fish received prior to their first behavior and pre-infection. VIE - visible implant elastomer implant, UNTOUCHED - control individual, received no injection nor implant (mm)
Infection: Whether or not the individual was infected with Gyrodactylus turnbulli. 1 - infected, 0 - uninfected
LateWeight: The weight of the individual after their final behavior trial and after infection. $(grams)$
wormJump: The number of worms that jumped from the donor fish to the trial fish during manual infections.
AUC2: The area under the curve of infection over the total infection trajectory for each individual.
RecovPeriod: The time frame in which an individual started recovering from infection. This is calculated by looking at the growth rate of the worms between each count and see when the worm growth rate was decreasing overall.
ContInf: Whether or not the individual was controling infection. We qualified controling infection by having a negative growth rate post peak infection. 1 - Controlled, 0 - Uncontrolled
Died: Whether or not the individual Died from infection during the experimental trial. 1 - Died, 0 - did not died
PreSMI: The body condition of the individual prior to their first behavioral trial and pre infection. (mm/g)
LateSMI: The body condition of the individual after to their last behavioral trial and after infection. (mm/g)
Wormbf:The number of worms on the fish prior to each trial.
Ratebf:The rate of growth of worms on the fish prior to each trial from the worm count immediately prior to the trial.
NRatebf:The rate of growth of worms on the fish prior to each trial calculated as the change in worms between each time point. (For Example, From before to early).
VarVel:The variance in velocity from the separate behavioral trials during each time point. $((cm/s)^2)$
BehavTol:The behavioral tolerance of each individual calculated as the slope of a random effect model where fishID is the random effect term.
BehavVig:The behavioral Vigor of each individual calculated as the intercept of a random effect model where fishID is the random effect term
ChBehav:The change in behavior between trial times. (cm/s)
PTol:Point tolerance metric calculated by dividing the change in behavior between time points by the worms on the fish before the trial $(cm/s/worms)$

1.7 Visualizing the relationships between variables in our dataset

```
View(IndBehav6)
pairs(~AvgVel + TrialTime + Sex + Treatment + Infection + LateWeight + AUC2 + Died +
    PreSMI + LateSMI + Wormbf + Ratebf + NRatebf + VarVel + BehavTol + BehavVig +
    ChBehav, lower.panel = panel.smooth, diag.panel = panel.hist, upper.panel = panel.cor,
    data = IndBehav6)
```



1.8 Visualize some patterns in the raw data

```
# Load in dataset from above. This bit of code is meant to save time so people
# dont have to rerun the entire data parsing and calculating step above. Code
# will be saved for reproducibility.
IndBehav7 <- read_csv("IndividualBehaviors_20240424.csv")
```

```
## Rows: 347 Columns: 25
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (3): TrialTime, Sex, Treatment
## dbl (21): fishID, AvgVel, PreWeight, PreLength, Infection, LateWeight, wormJ...
## lgl  (1): ContrPeriod
##
## 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.
```

```
# Setting some of the factors back to factors
IndBehav7$fishID <- as.factor(IndBehav7$fishID)
IndBehav7$TrialTime <- as.factor(IndBehav7$TrialTime)
IndBehav7$Sex <- as.factor(IndBehav7$Sex)
IndBehav7$Infection <- as.factor(IndBehav7$Infection)
IndBehav7$ContInf <- as.factor(IndBehav7$ContInf)
IndBehav7$Died <- as.factor(IndBehav7$Died)
IndBehav7$Treatment <- as.factor(IndBehav7$Treatment)


# Calculating tissue tolerance for each individual.
```

```
IndBehav7 <- IndBehav7 %>%
    mutate(ChSMI = LateSMI - PreSMI) %>%
    mutate(TisTol = ChSMI/Totworm)
```

```
## mutate: new variable 'ChSMI' (double) with 85 unique values and 0% NA
## mutate: new variable 'TisTol' (double) with 47 unique values and 46% NA
```

```
# Filtering the data set for individuals who are not controlling for infection.
IndBehavCont <- IndBehav7 %>%
    filter(ContInf == "0" | ContInf == "1")
```

```
## filter: removed 152 rows (44%), 195 rows remaining
```

```
# Filtering to only infected individuals for tolerance metrics
IndBehavI <- IndBehav7 %>%
    filter(Infection == "1")
```

```
## filter: removed 156 rows (45%), 191 rows remaining
```

1.9 What hypotheses we want to test with these data and what data we can use to test them?

For uninfected and infected individuals only:

Do infected and uninfected individuals differ in their average and variation in velocity?

`Is there sexual variation in the average and variation in velocity?`

Do we see differences in the change in behavior and between infected

For infected individuals only:

Is there sexual variation in host behavioral tolerance and behavioral vigor?

Do host behavioral tolerance, tissue tolerance, and behavioral vigor correlate together and does this differ by sex?

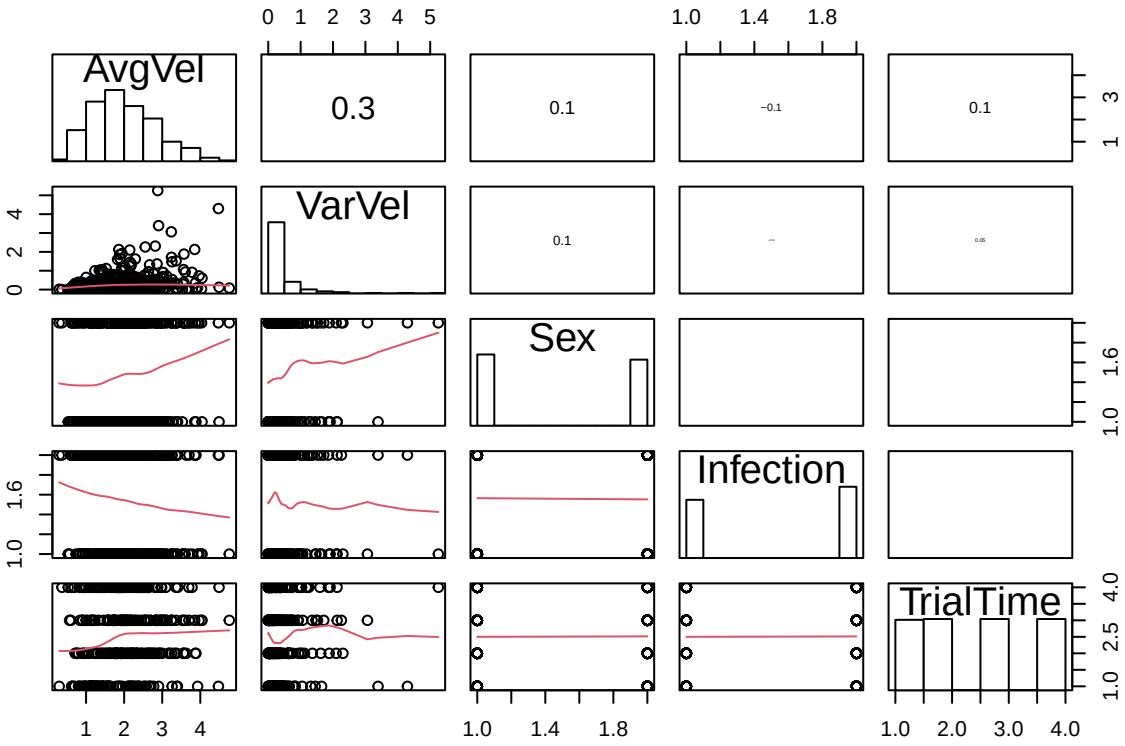
Do hosts with lower behavioral tolerance have higher behavior resilience (i.e. how hosts behave once they have started to clear infection/have cleared infection)

Is there sexual variation in individuals behavioral resilience?

2 Do infected and uninfected individuals differ in their average and variation in velocity?

2.1 Visually inspection of the explanatory variables that will be used in the analyses

```
pairs(~AvgVel + VarVel + Sex + Infection + TrialTime, lower.panel = panel.smooth,
    diag.panel = panel.hist, upper.panel = panel.cor, data = IndBehav7)
```



2.2 Does infection or sex variation impact the average velocity of individauls?

Individuals had 3 behavioral trials per time period of infection (i.e. 3 behavioral trials before infection) and therefore using preliminary analysis we showed that there is no difference due to time of day of these recordings so we averaged and quantified the variance of the velocities for that day to get an average velocity per trial time.

This analysis uses the average velocity for each individual at each trial point.

2.2.1 Description, development, and fitting of linear model for the analysis

We will use a linear mixed model to analyze how average velocity differs by infection status and sexual variation. FishID is included as a random term to allow for non-independence of individuals due to multiple measurements per individual across time.

- Deterministic
- $AvgVel_{det}$ = a + b_1TrialTime + b_2Infection + b_3Sex + a_i
- Stochastic
  - AvgVel ~ $N(AvgVel_{det}, \sigma^2)$
  - $a_i \sim N(0, \sigma^2_{fishID})$
- Fixed
  - TrialTime
  - Infection
  - Sex

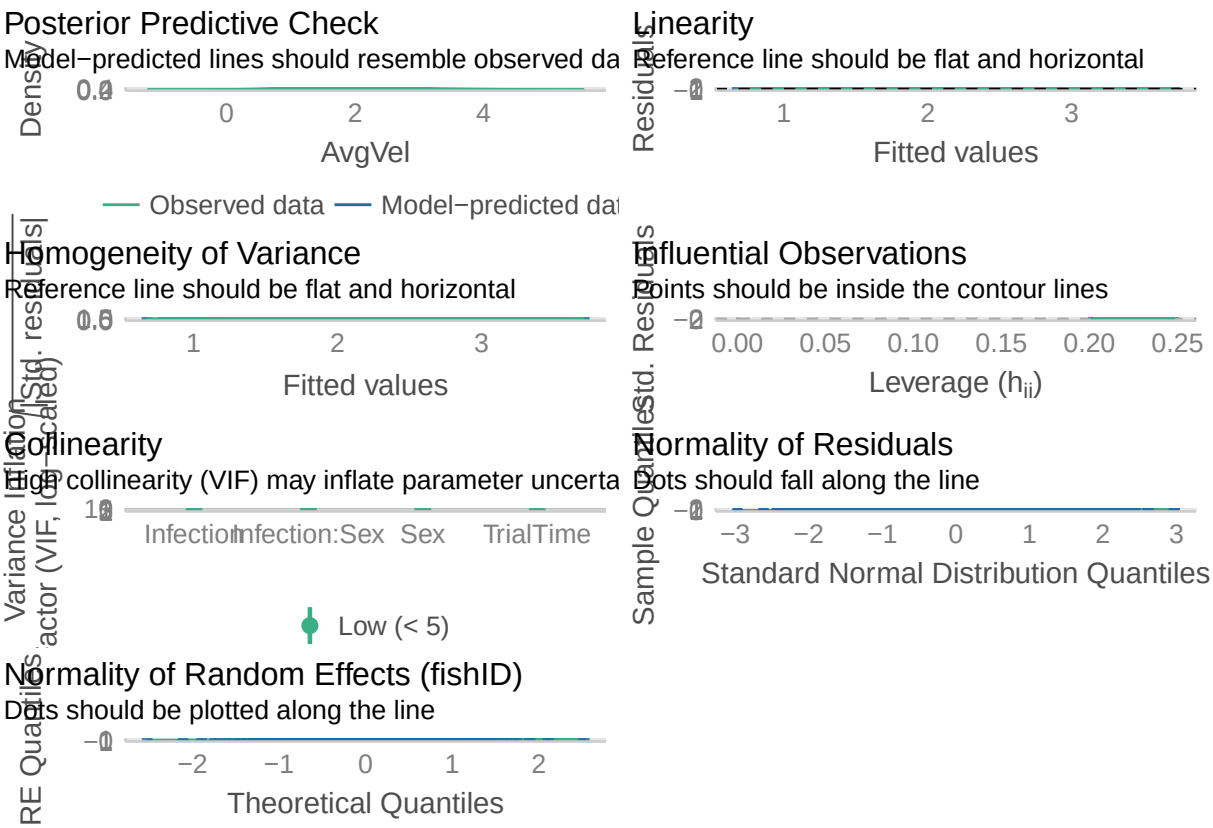
- Random
  - fishID

```
# Fit a linear model for checking what explanatory factors are important for
# Average Velocity Note this is a linear mixed model because we have multiple
# measures per fish and therefore, need to account for non-independence between
# measures.
AvgVelLM <- lmer(AvgVel ~ TrialTime + Infection * Sex + (1 | fishID), IndBehav7)



# Summary to see the relationship of the variables.
summary(AvgVelLM)
```

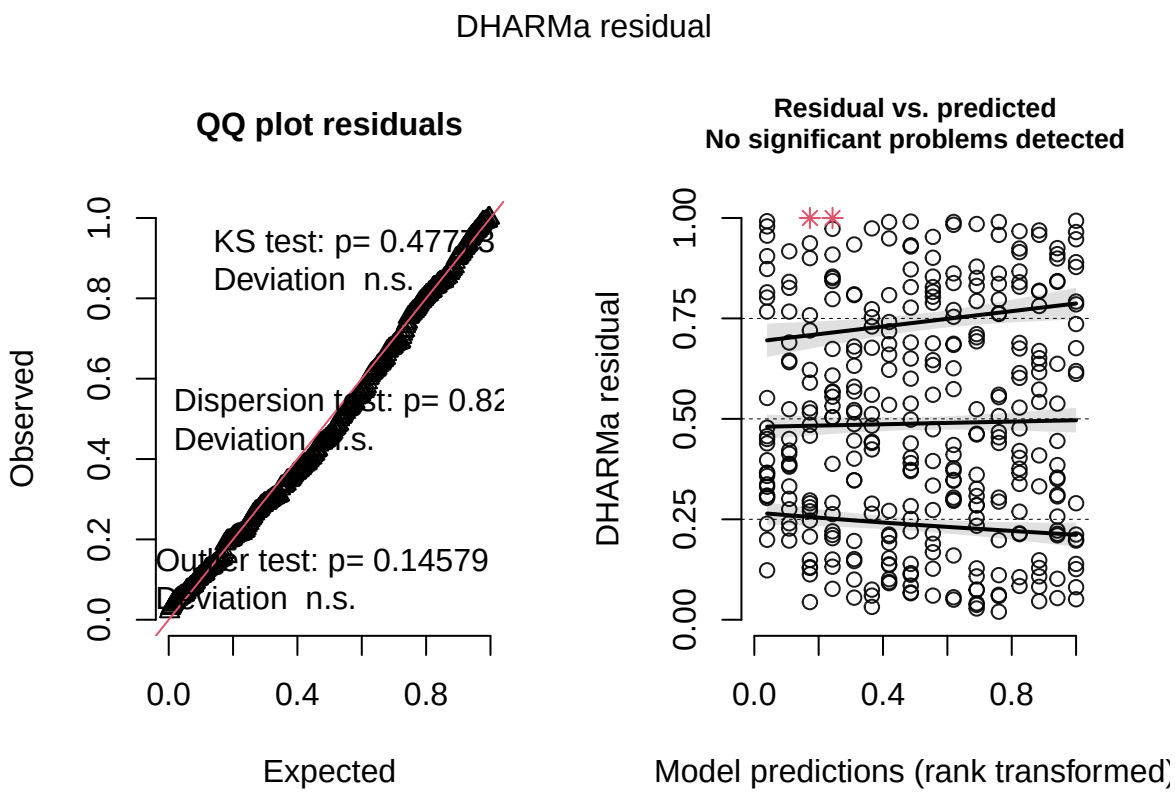
```
## Linear mixed model fit by REML ['lmerMod']
## Formula: AvgVel ~ TrialTime + Infection * Sex + (1 | fishID)
##    Data: IndBehav7
##
## REML criterion at convergence: 737.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.4345 -0.5983 -0.1208  0.6133  3.3557
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  fishID   (Intercept) 0.2838   0.5327
##  Residual             0.3534   0.5945
## Number of obs: 336, groups:  fishID, 86
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)       1.55954    0.14738  10.582
## TrialTimeEarly    0.46775    0.09066   5.160
## TrialTimeLate     0.60388    0.09066   6.661
## TrialTimeLater    0.25170    0.09351   2.692
## Infection1       -0.01296    0.18312  -0.071
## SexM              0.45148    0.19562   2.308
## Infection1:SexM -0.40505    0.26545  -1.526
##
## Correlation of Fixed Effects:
##             (Intr) TrlTmE TrilTmLt TrlTmLtr Infct1 SexM
## TrialTmErly -0.308
## TrialTimeLt -0.308  0.500
## TrialTimLtr -0.308  0.485  0.485
## Infection1  -0.690  0.000  0.000    0.003
## SexM        -0.646  0.000  0.000    0.004    0.519
## Infctn1:SxM  0.474  0.000  0.000    0.010   -0.690 -0.737
```

2.2.2 Validate that the model fits well and there are no problems

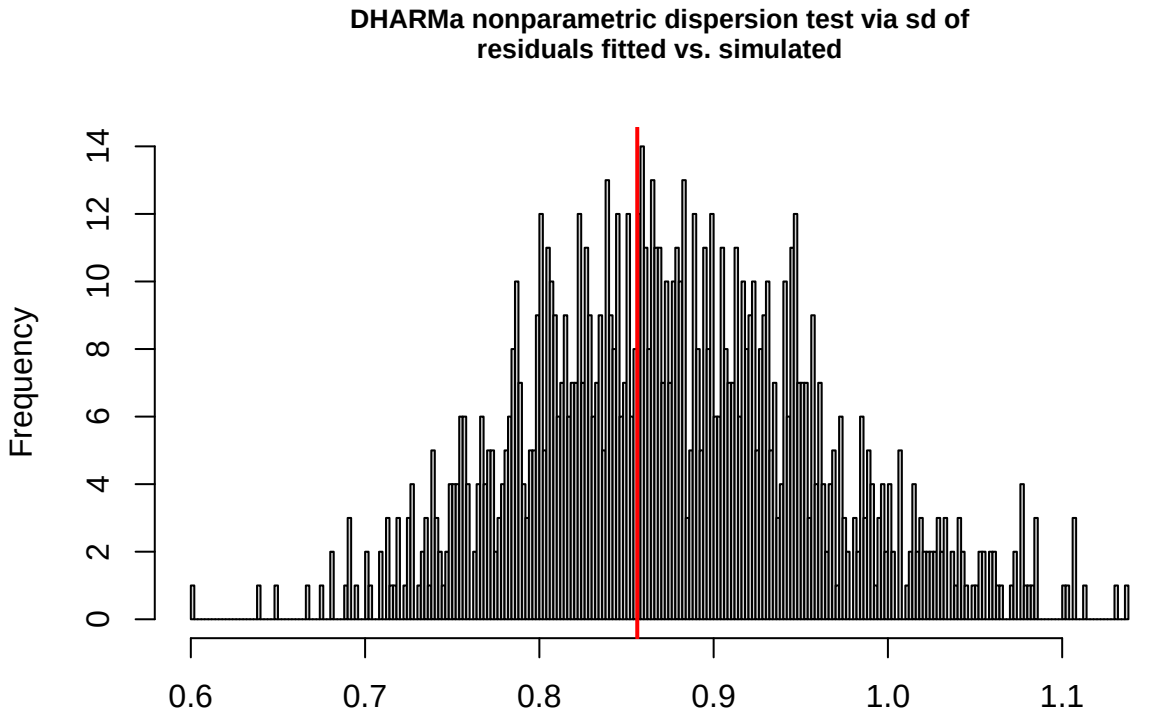
```
# Using the check_model function from the perforamnce package to check the
# model validation

check_model(AvgVelLM)
```




```
# Using the Dharma package to check quantile residuals First simulating the
# quantile residuals
sim_residuals_AvgVelLM <- simulateResiduals(AvgVelLM, 1000)
# Plotting the quantile residuals to test how quantile residuals look
plot(sim_residuals_AvgVelLM)
```




```
# Testing for dispersion
testDispersion(sim_residuals_AvgVelLM)
```



```
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## dispersion = 0.97694, p-value = 0.828
## alternative hypothesis: two.sided
```

```
# All model validation looks good.
```

2.2.3 Testing the significance of factors in our model using a Kenward-Rodgers F test

```
# F test to test for signficance of slope of variables
Anova(AvgVelLM, test = "F", type = 3)
```

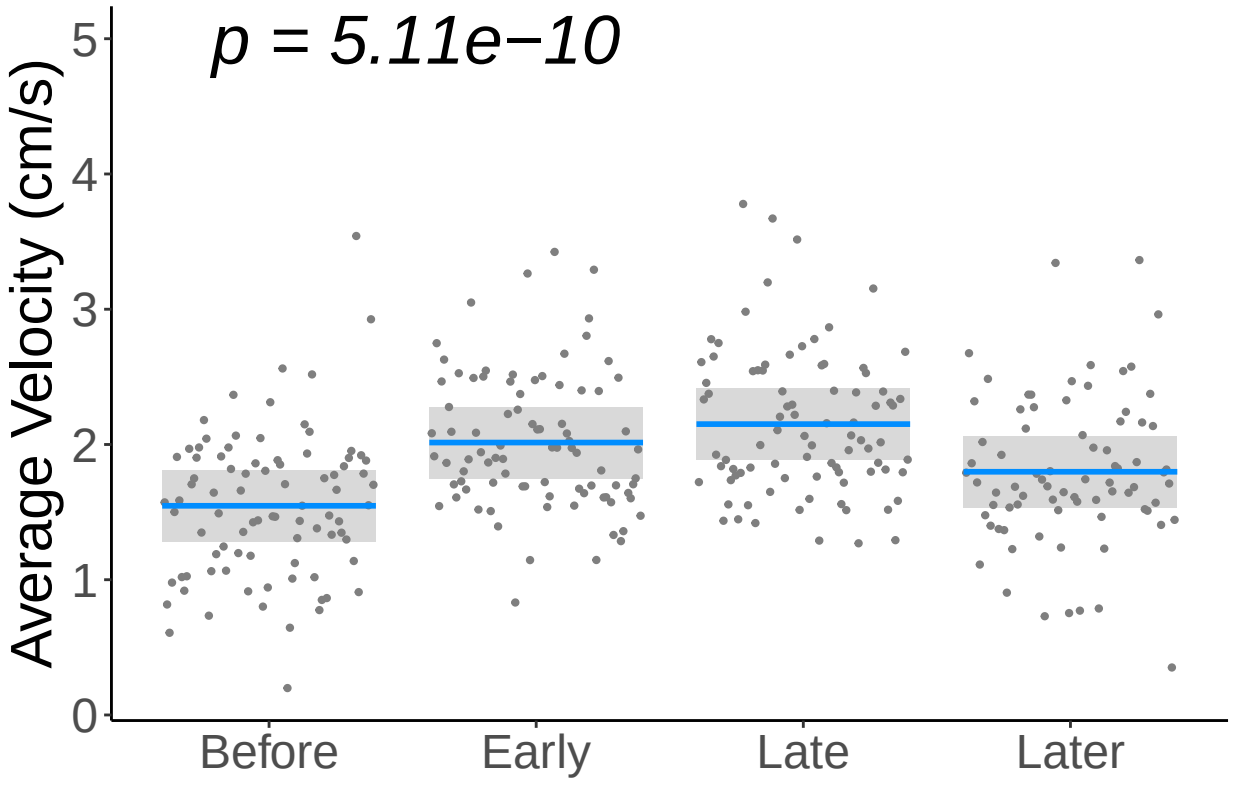
```
## Analysis of Deviance Table (Type III Wald F tests with Kenward-Roger df)
##
## Response: AvgVel
##                       F Df  Df.res    Pr(>F)
## (Intercept)   111.9753  1 109.313 < 2.2e-16 ***
## TrialTime      16.8953  3 247.962 5.118e-10 ***
## Infection       0.0050  1  81.104   0.94376
## Sex             5.3265  1  81.208   0.02355 *
## Infection:Sex   2.3282  1  81.944   0.13090
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.2.4 Visualize the important explanatory factors for average velocity

```
# TrialTimeGraph
AvgVelbyTT = visreg(AvgVelLM, scale = "response", "TrialTime", partial = T, gg = TRUE) +
    theme_classic() + theme(legend.position = "none") + ylab("Average Velocity (cm/s)") +
    xlab(" ") + theme(text = element_text(size = 22)) + annotate("text", x = 0.25,
    y = 5, label = "p = 5.11e-10", size = 9, fontface = "italic")
```

```
## Conditions used in construction of plot
## Infection: 1
## Sex: F
## fishID: 1
```

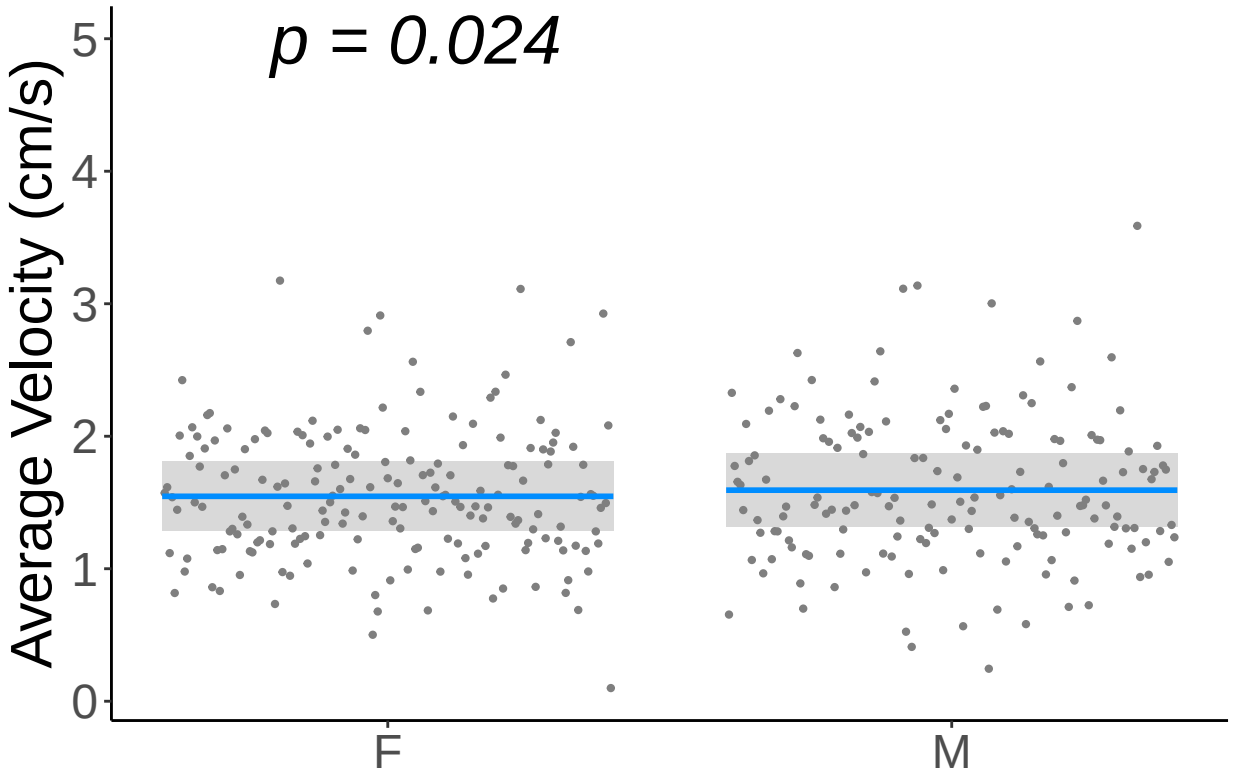
```
# Print the graph
print(AvgVelbyTT)
```




```
# TrialTimeGraph
AvgVelbySex = visreg(AvgVelLM, scale = "response", "Sex", partial = T, gg = TRUE) +
    theme_classic() + theme(legend.position = "none") + ylab("Average Velocity (cm/s)") +
    xlab(" ") + theme(text = element_text(size = 22)) + annotate("text", x = 0.25,
    y = 5, label = "p = 0.024", size = 9, fontface = "italic")
```

```
## Conditions used in construction of plot
## TrialTime: Before
## Infection: 1
## fishID: 1
```

```
# Print the graph
print(AvgVelbySex)
```



2.3 Does infection or sex variation impact the variance in velocity of indivi-dauls?

This analysis uses the variance of velocity for each individual at each trial point.

2.3.1 Description, development, and fitting of linear model for the analysis

We will use a linear mixed model to analyze how variance in velocity differs by infection status and sexual variation. FishID is included as a random term to allow for non-independence of individuals due to multiple measurements per individual across time.

- Deterministic
- $VarVel_{det}$ = a + b_1TrialTime + b_2Infection + b_3Sex + a_i
- Stochastic
  - VarVel ~ $N(VarVel_{det}, \sigma^2)$
  - $a_i \sim N(0, \sigma^2_{fishID})$
- Fixed
  - TrialTime
  - Infection
  - Sex
- Random

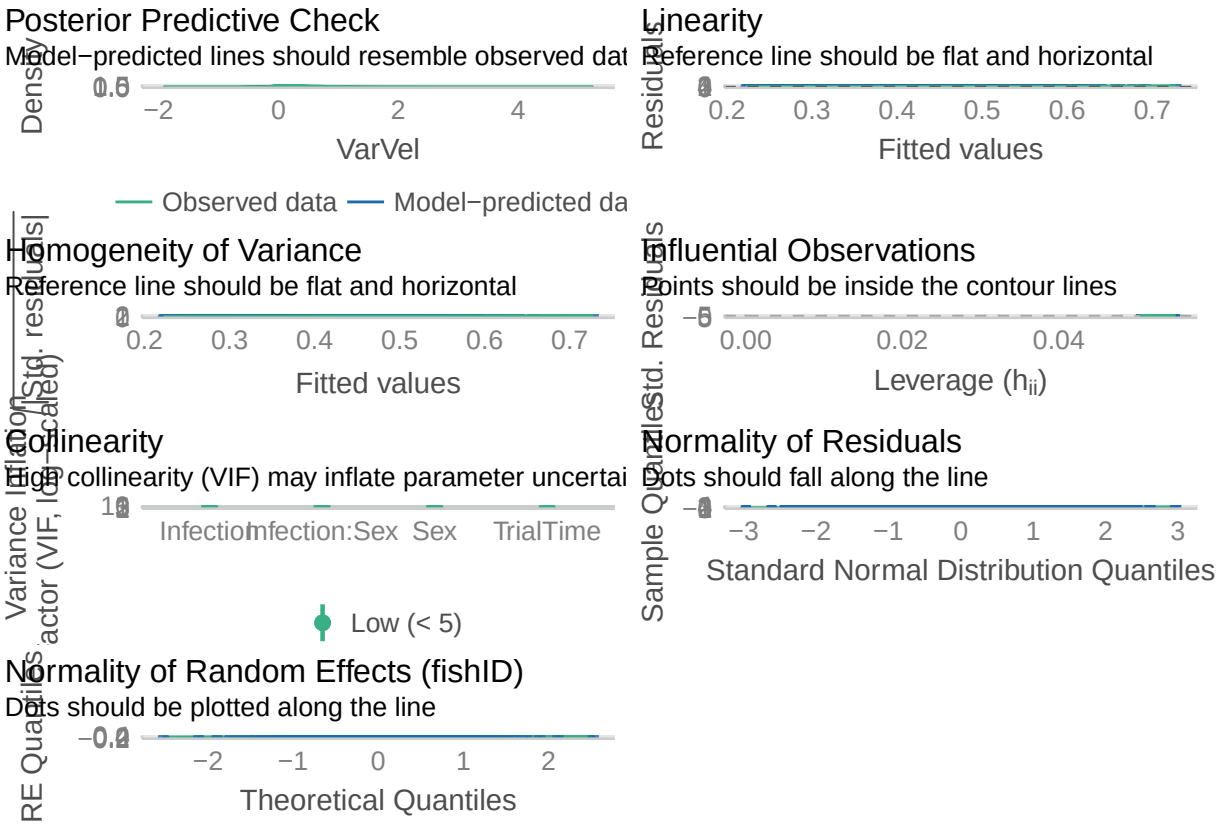
- fishID

```
# Fit a linear model for checking what explanatory factors are important for
# Variance in Velocity Note this is a linear mixed model because we have
# multiple measures per fish and therefore, need to account for
# non-independence between measures.
VarVelLM <- lmer(VarVel ~ TrialTime + Infection * Sex + (1 | fishID), IndBehav7)

# Summary to see the relationship of the variables.
summary(VarVelLM)
```

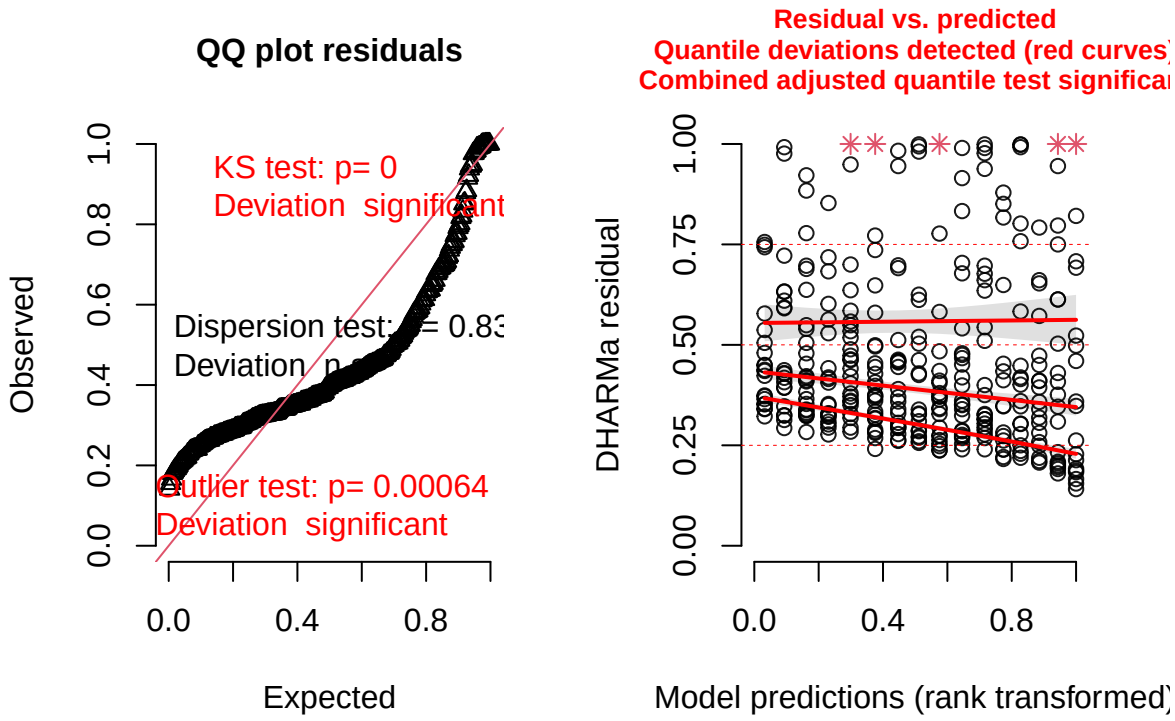
```
## Linear mixed model fit by REML ['lmerMod']
## Formula: VarVel ~ TrialTime + Infection * Sex + (1 | fishID)
##    Data: IndBehav7
##
## REML criterion at convergence: 614.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.1442 -0.4932 -0.2441  0.1172  7.8163
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  fishID   (Intercept) 0.0127   0.1127
##  Residual             0.3350   0.5788
## Number of obs: 336, groups:  fishID, 86
##
## Fixed effects:
##                 Estimate Std. Error t value
## (Intercept)      0.29427    0.08815   3.338
## TrialTimeEarly  -0.03672    0.08826  -0.416
## TrialTimeLate    0.01992    0.08826   0.226
## TrialTimeLater   0.07158    0.09064   0.790
## Infection1       0.04223    0.09335   0.452
## SexM             0.24678    0.09979   2.473
## Infection1:SexM -0.18959    0.13612  -1.393
##
## Correlation of Fixed Effects:
##             (Intr) TrlTmE TrilTmLt TrlTmLtr Infct1 SexM
## TrialTmErly -0.501
## TrialTimeLt -0.501  0.500
## TrialTimLtr -0.501  0.487  0.487
## Infection1  -0.587  0.000  0.000    0.005
## SexM        -0.550  0.000  0.000    0.006    0.518
## Infctn1:SxM  0.397  0.000  0.000    0.017   -0.686 -0.733
```

2.3.2 Validate that the model fits well and there are no problems

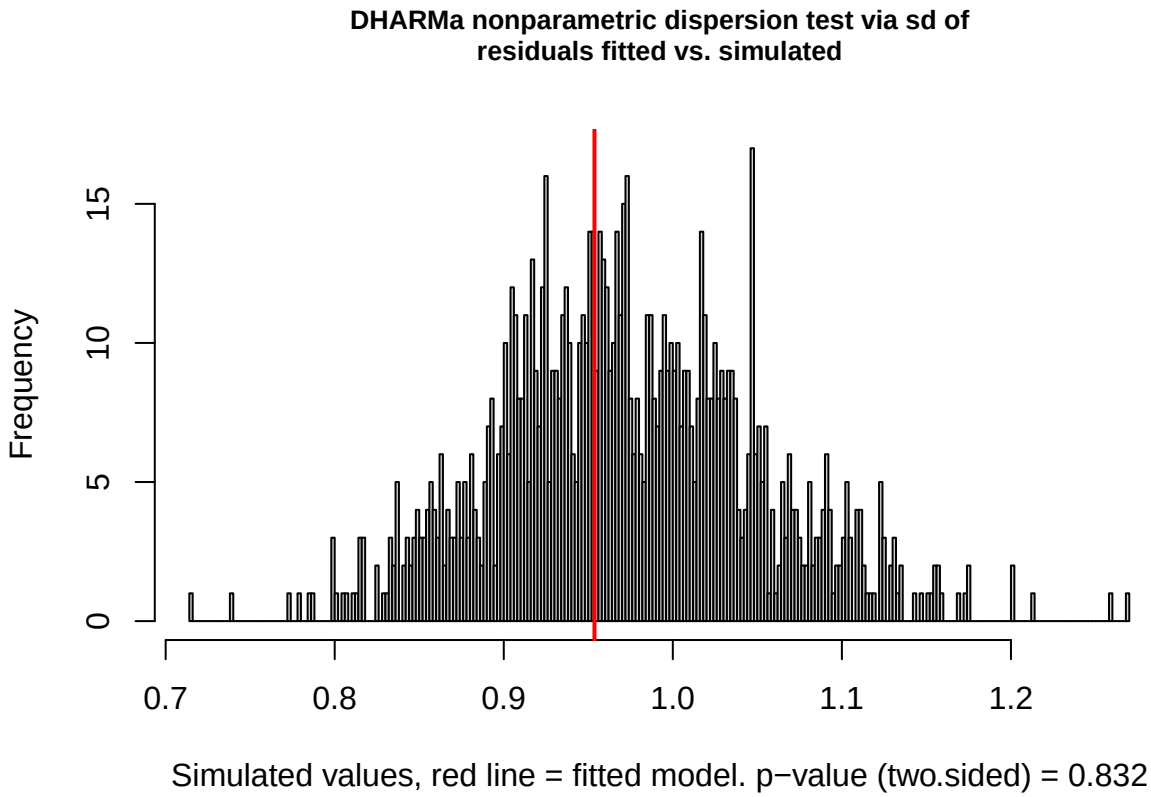
```
# Using the check_model function from the perforamnce package to check the
# model validation

check_model(VarVelLM)
```




```
# Using the Dharma package to check quantile residuals First simulating the
# quantile residuals
sim_residuals_VarVelLM <- simulateResiduals(VarVelLM, 1000)
# Plotting the quantile residuals to test how quantile residuals look
plot(sim_residuals_VarVelLM)
```




```
# Testing for dispersion
testDispersion(sim_residuals_VarVelLM)
```



```
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## dispersion = 0.98009, p-value = 0.832
## alternative hypothesis: two.sided
```

```
# There are some problems with this model validation. It doesnt look model
# breaking but definitely should look at other model error structures to
# resolve the issues.
```

2.3.3 Testing the significance of factors in our model using a Kenward-Rodgers F test

```
# F test to test for signficance of slope of variables
Anova(VarVelLM, test = "F", type = 3)
```

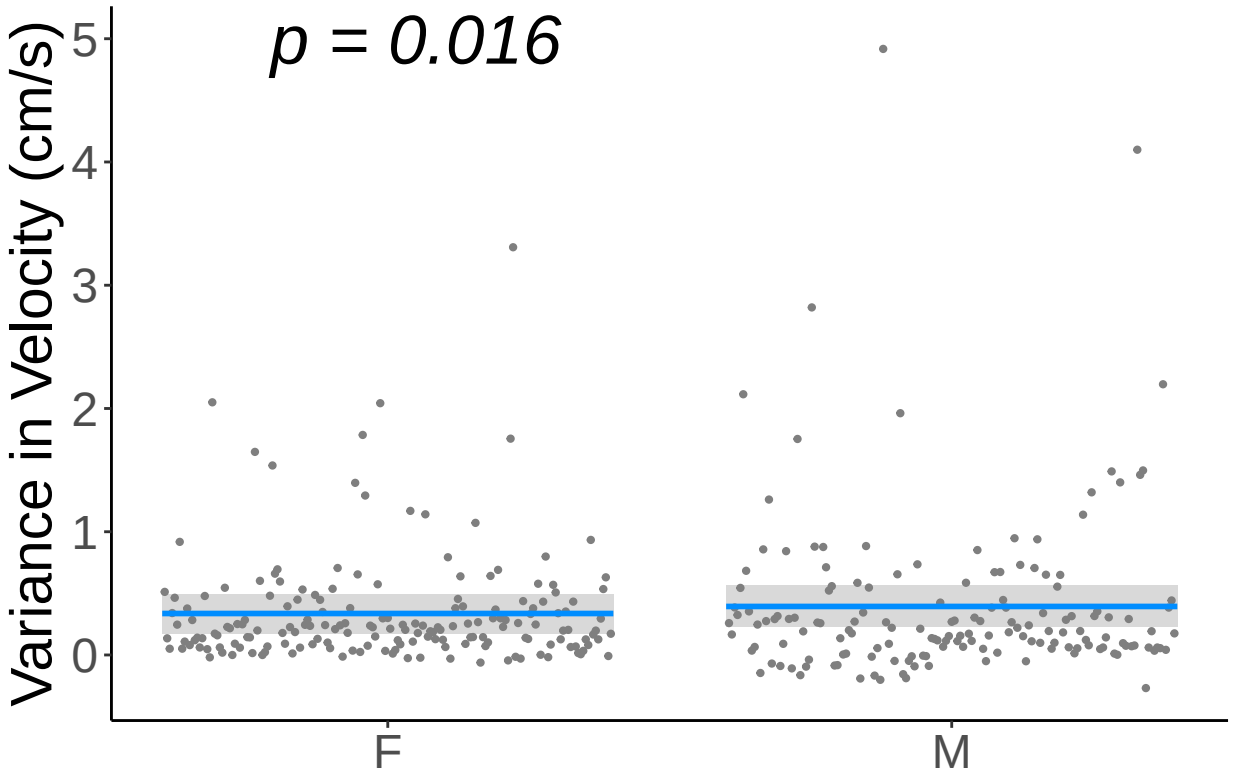
```
## Analysis of Deviance Table (Type III Wald F tests with Kenward-Roger df)
##
## Response: VarVel
##                      F Df  Df.res  Pr(>F)
## (Intercept)   11.1449  1 183.824 0.00102 **
## TrialTime      0.4943  3 249.871 0.68653
## Infection      0.2047  1  79.390 0.65222
## Sex            6.1148  1  79.655 0.01553 *
## Infection:Sex  1.9396  1  81.661 0.16749
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.3.4 Visualize the important explanatory factors for variance of velocity

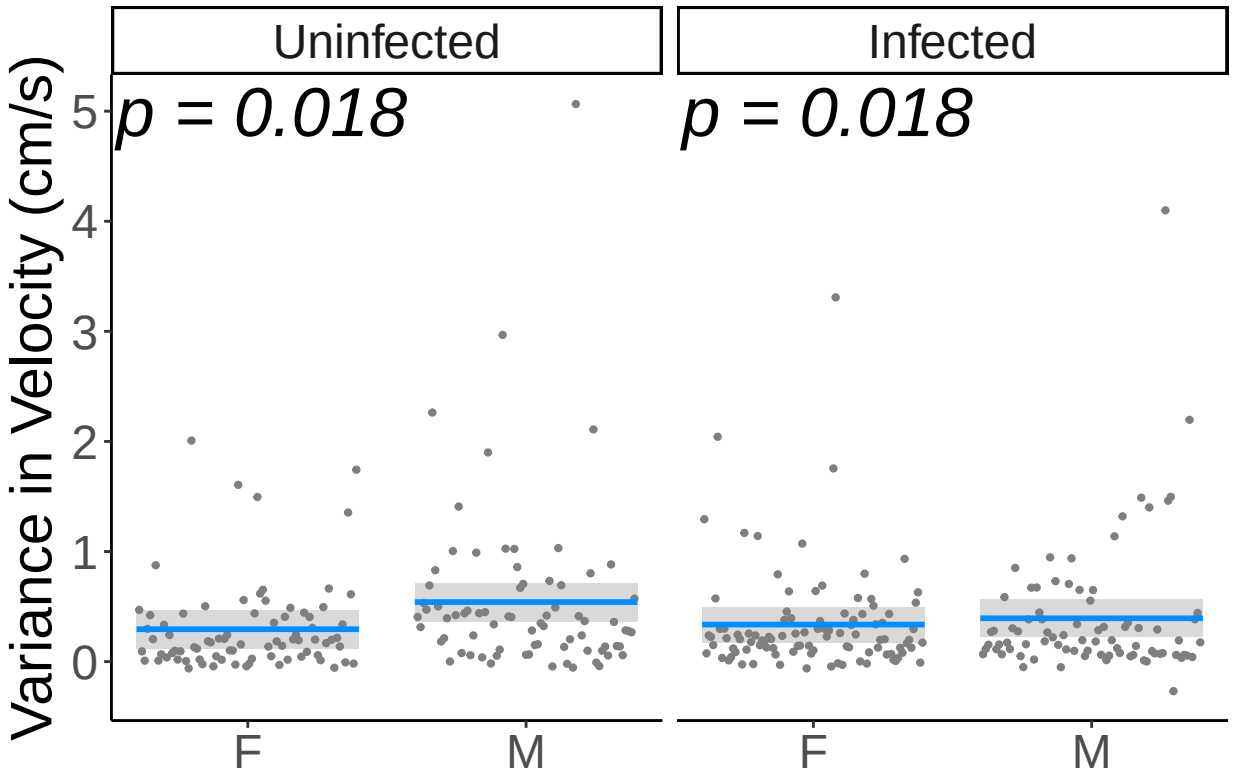
```
# VariancebySex
VarVelbySex <- visreg(VarVelLM, scale = "response", "Sex", partial = T, gg = TRUE) +
    theme_classic() + theme(legend.position = "none") + ylab("Variance in Velocity (cm/s)") +
    xlab(" ") + theme(text = element_text(size = 22)) + annotate("text", x = 0.25,
    y = 5, label = "p = 0.016", size = 9, fontface = "italic")
```

```
## Conditions used in construction of plot
## TrialTime: Before
## Infection: 1
## fishID: 1
```

```
# Print the graph
print(VarVelbySex)
```



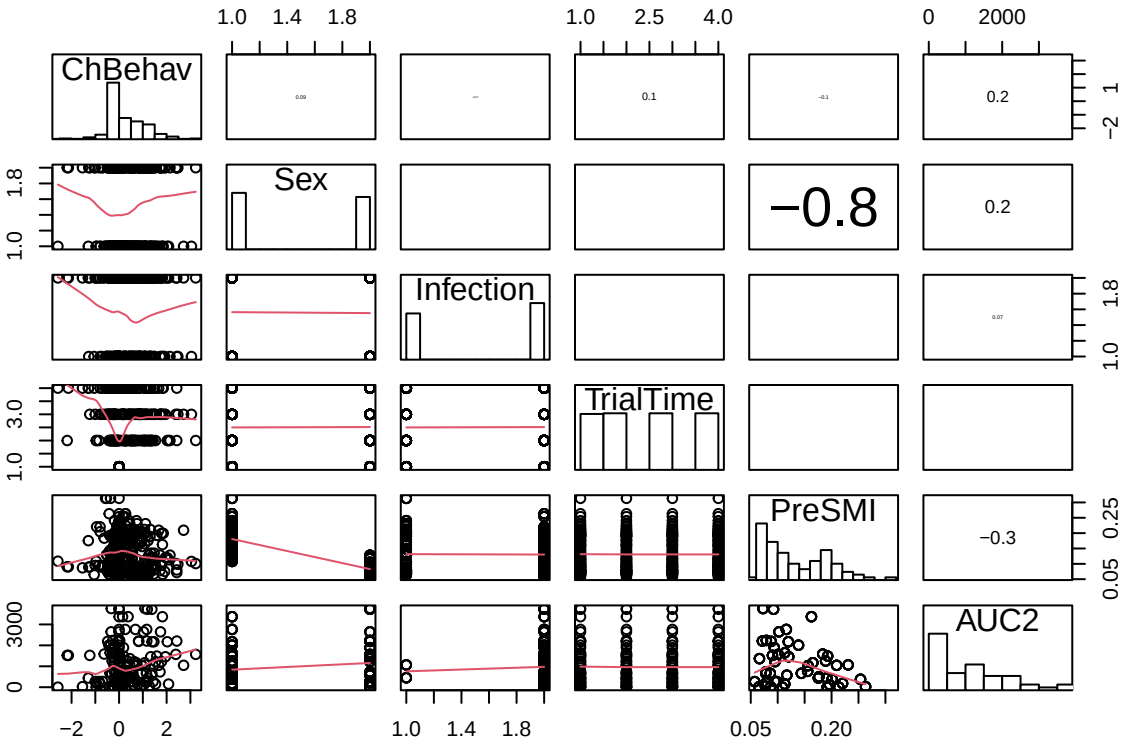

```
# Variance by sex and split by infection
Infnames <- c("Uninfected", "Infected")
names(Infnames) <- c("0", "1")

VarVelbySexnInf <- visreg(VarVelLM, scale = "response", "Sex", "Infection", partial = T,
    gg = TRUE) + theme_classic() + theme(legend.position = "none") + ylab("Variance in Velocity (cm/s)")
    xlab(" ") + theme(text = element_text(size = 22)) + annotate("text", x = 0.25,
    y = 5, label = "p = 0.018", size = 9, fontface = "italic") + facet_wrap(~Infection,
    labeller = labeller(Infection = Infnames))

# Print the graph
print(VarVelbySexnInf)
```



3 Do we see differences in change in behavior over time based on infection status and sexual variation?

3.1 Visually inspection of the explanatory variables that will be used in the analyses

```
pairs(~ChBehav + Sex + Infection + TrialTime + PreSMI + AUC2, lower.panel = panel.smooth,
    diag.panel = panel.hist, upper.panel = panel.cor, data = IndBehav7)
```



3.1.1 Description, development, and fitting of linear model for the analysis

We will use a linear mixed model to analyze how Change in velocity differs by infection status and sexual variation. FishID is included as a random term to allow for non-independence of individuals due to multiple measurements per individual across time.

- Deterministic
- $ChVel_{det}$ = a + b_1TrialTime + b_2Infection + b_3Sex + a_i
- Stochastic
  - ChVel ~ $N(ChVel_{det}, \sigma^2)$
  - $a_i \sim N(0, \sigma^2_{fishID})$
- Fixed
  - TrialTime
  - Infection
  - Sex
- Random
  - fishID

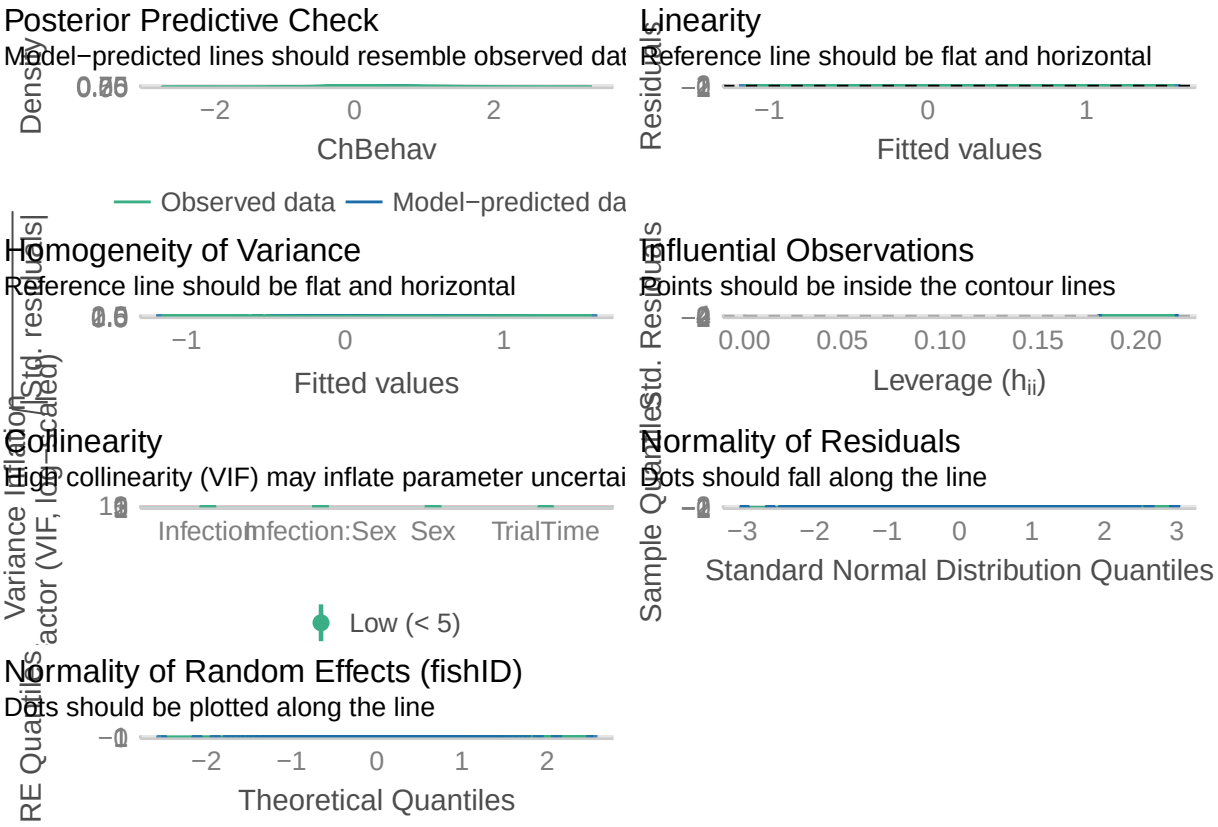
```
# Fit a linear model for checking what explanatory factors are important for
# Variance in Velocity Note this is a linear mixed model because we have
# multiple measures per fish and therefore, need to account for
# non-independence between measures.
ChVelLM <- lmer(ChBehav ~ TrialTime + Infection * Sex + (1 | fishID), IndBehav7)
```

```r
# Summary to see the relationship of the variables.
summary(VarVelLM)
```

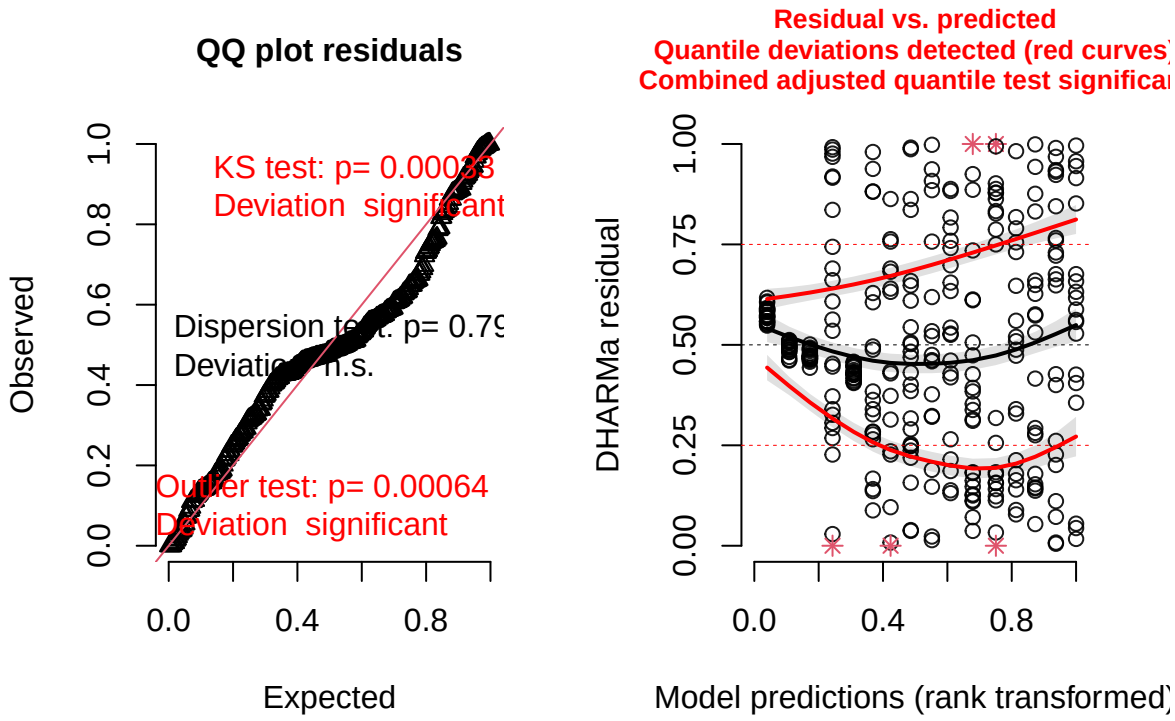
```
## Linear mixed model fit by REML ['lmerMod']
## Formula: VarVel ~ TrialTime + Infection * Sex + (1 | fishID)
##    Data: IndBehav7
##
## REML criterion at convergence: 614.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.1442 -0.4932 -0.2441  0.1172  7.8163
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  fishID   (Intercept) 0.0127   0.1127
##  Residual             0.3350   0.5788
## Number of obs: 336, groups:  fishID, 86
##
## Fixed effects:
##                 Estimate Std. Error t value
## (Intercept)      0.29427    0.08815   3.338
## TrialTimeEarly  -0.03672    0.08826  -0.416
## TrialTimeLate    0.01992    0.08826   0.226
## TrialTimeLater   0.07158    0.09064   0.790
## Infection1       0.04223    0.09335   0.452
## SexM             0.24678    0.09979   2.473
## Infection1:SexM -0.18959    0.13612  -1.393
##
## Correlation of Fixed Effects:
##             (Intr) TrlTmE TrilTmLt TrlTmLtr Infct1 SexM
## TrialTmErly -0.501
## TrialTimeLt -0.501  0.500
## TrialTimLtr -0.501  0.487  0.487
## Infection1  -0.587  0.000  0.000    0.005
## SexM        -0.550  0.000  0.000    0.006    0.518
## Infctn1:SxM  0.397  0.000  0.000    0.017   -0.686 -0.733
```

3.1.2 Validate that the model fits well and there are no problems

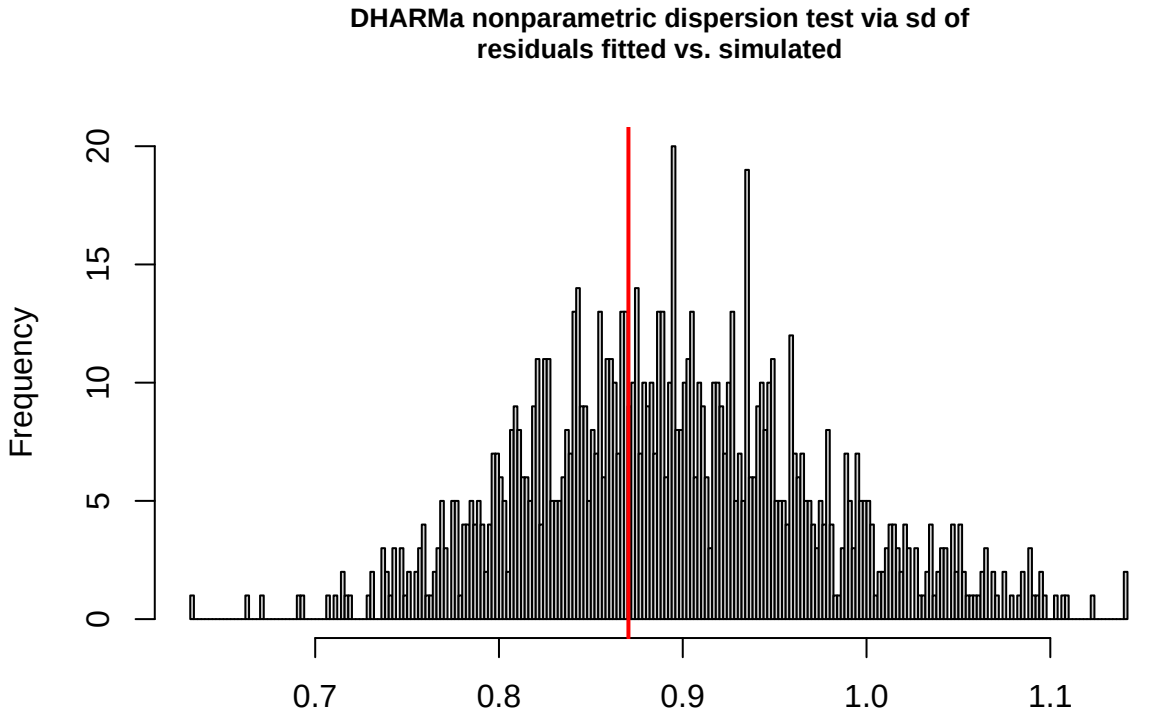
```r
# Using the check_model function from the perforamnce package to check the
# model validation

check_model(ChVelLM)
```




```
# Using the Dharma package to check quantile residuals First simulating the
# quantile residuals
sim_residuals_ChVelLM <- simulateResiduals(ChVelLM, 1000)
# Plotting the quantile residuals to test how quantile residuals look
plot(sim_residuals_ChVelLM)
```




```
# Testing for dispersion
testDispersion(sim_residuals_ChVelLM)
```



```
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## dispersion = 0.97279, p-value = 0.794
## alternative hypothesis: two.sided
```

```
# There are some problems with this model validation. It doesnt look model
# breaking but definitely should look at other model error structures to
# resolve the issues.
```

3.1.3 Testing the significance of factors in our model using a Kenward-Rodgers F test

```
# F test to test for signficance of slope of variables
Anova(ChVelLM, test = "F", type = 3)
```

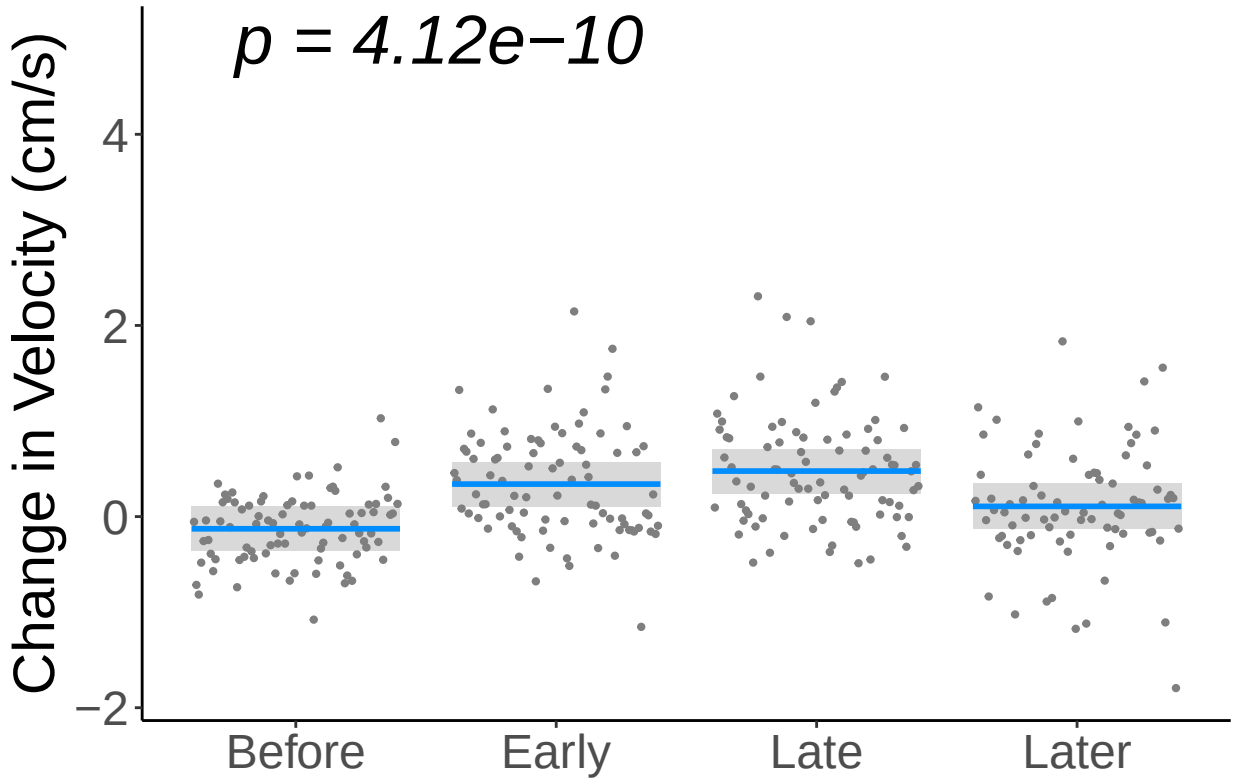
```
## Analysis of Deviance Table (Type III Wald F tests with Kenward-Roger df)
##
## Response: ChBehav
##                      F Df  Df.res    Pr(>F)
## (Intercept)    0.0000  1 118.842    0.9998
## TrialTime     17.0705  3 248.251 4.123e-10 ***
## Infection      0.6487  1  80.840    0.4230
## Sex            0.4838  1  80.972    0.4887
## Infection:Sex  0.0533  1  81.917    0.8180
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3.1.4 Visualize the important explanatory factors for change in velocity

```
# TrialTimeGraph
ChVelbyTT = visreg(ChVelLM, scale = "response", "TrialTime", partial = T, gg = TRUE) +
    theme_classic() + theme(legend.position = "none") + ylab("Change in Velocity (cm/s)") +
    xlab(" ") + theme(text = element_text(size = 22)) + annotate("text", x = 0.25,
    y = 5, label = "p = 4.12e-10", size = 9, fontface = "italic")
```

```
## Conditions used in construction of plot
## Infection: 1
## Sex: F
## fishID: 1
```

```
# Print the graph
print(ChVelbyTT)
```


p = 4.12e−10
Change in Velocity (cm/s)
4
2
0
−2
Before
Early
Late
Later