Final project

Jared Adam

12/6/2024

# Packages

# Data import and wrangling

set.seed(654321)  
s21 <- read\_csv('data/2021 Sentinel Prey Assessment.csv')  
s22 <- read\_csv("data/PSA\_CE2\_SentinelPrey.csv")  
s23 <- read\_csv('data/PSA\_Sent.prey.2023.csv')  
  
# I need to get total predation into a column as a binary. 1 = predation, 0 = not  
  
# 2021 cleaning   
s21

## # A tibble: 5,281 × 17  
## location date n.weather d.weather growth\_stage plot\_id rep.block treatment  
## <chr> <chr> <dbl> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 PA 6/16/2… 15.3 18.44 V3 101 1 1  
## 2 PA 6/16/2… 15.3 18.44 V3 101 1 1  
## 3 PA 6/16/2… 15.3 18.44 V3 101 1 1  
## 4 PA 6/16/2… 15.3 18.44 V3 101 1 1  
## 5 PA 6/16/2… 15.3 18.44 V3 101 1 1  
## 6 PA 6/16/2… 15.3 18.44 V3 101 1 1  
## 7 PA 6/16/2… 15.3 18.44 V3 102 1 3  
## 8 PA 6/16/2… 15.3 18.44 V3 102 1 3  
## 9 PA 6/16/2… 15.3 18.44 V3 102 1 3  
## 10 PA 6/16/2… 15.3 18.44 V3 102 1 3  
## # ℹ 5,271 more rows  
## # ℹ 9 more variables: row <dbl>, sample <dbl>, n.absent <chr>, n.partial <chr>,  
## # n.predated <chr>, d.absent <chr>, d.partial <chr>, d.predated <chr>,  
## # to.predated <chr>

clean21 <- s21 %>%   
 mutate(year = '2021') %>%   
 dplyr::select(location, year, growth\_stage, plot\_id, rep.block, treatment, to.predated) %>%   
 mutate(to.predated = as.double(to.predated)) %>%   
 dplyr::rename(block = rep.block) %>%   
 group\_by(location, year, growth\_stage, plot\_id, block, treatment) %>%   
 # dplyr::summarise(total = sum(to.predated)) %>%   
 na.omit() %>%   
 mutate(treatment = case\_when(  
 treatment == '33' ~ '3',  
 .default = as.factor(treatment))) %>%   
 dplyr::filter(treatment != '6',  
 treatment != '7',  
 treatment != '8') %>%   
 mutate\_at(vars(1:6), as.factor)   
  
  
# 2022 cleaning   
s22

## # A tibble: 3,246 × 19  
## location date am.weather pm.weather growth\_stage plotid block treatment  
## <chr> <chr> <dbl> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 PA 6/22/2022 22.7 26.1 V3 101 1 1  
## 2 PA 6/22/2022 22.7 26.1 V3 101 1 1  
## 3 PA 6/22/2022 22.7 26.1 V3 101 1 1  
## 4 PA 6/22/2022 22.7 26.1 V3 101 1 1  
## 5 PA 6/22/2022 22.7 26.1 V3 101 1 1  
## 6 PA 6/22/2022 22.7 26.1 V3 101 1 1  
## 7 PA 6/22/2022 22.7 26.1 V3 102 1 3  
## 8 PA 6/22/2022 22.7 26.1 V3 102 1 3  
## 9 PA 6/22/2022 22.7 26.1 V3 102 1 3  
## 10 PA 6/22/2022 22.7 26.1 V3 102 1 3  
## # ℹ 3,236 more rows  
## # ℹ 11 more variables: row <dbl>, sample <dbl>, am.absent <chr>,  
## # am.partial <chr>, am.predators <chr>, pm.absent <chr>, pm.partial <chr>,  
## # pm.predators <chr>, to.predated <dbl>, n.predated <dbl>, d.predated <dbl>

unique(s22$treatment)

## [1] 1 3 2 4

unique(s22$growth\_stage)

## [1] "V3" "V5" "R3" "R2"

clean22 <- s22 %>%   
 mutate(year = '2022') %>%   
 dplyr::select(location, year, growth\_stage, plotid, block, treatment, to.predated) %>%   
 dplyr::rename(plot\_id = plotid) %>%   
 mutate(growth\_stage = case\_when(growth\_stage == 'R2' ~ 'R3',  
 .default = as.character(growth\_stage))) %>%   
 dplyr::group\_by(location, year, growth\_stage, plot\_id, block, treatment) %>%   
 # dplyr::summarise(total = sum(to.predated)) %>%   
 mutate\_at(vars(1:6), as.factor)  
  
  
# 2023 cleaning   
  
clean23 <- s23 %>%  
 mutate(year = '2023') %>%  
 relocate(am.partial, am.absent, pm.partial, pm.absent) %>%   
 mutate\_at(vars(1:4), as.double) %>%   
 mutate(to.predated = if\_else(am.partial | am.absent | pm.partial | pm.absent == 1, 1, 0)) %>%   
 relocate(to.predated)%>%   
 mutate(growth\_stage = case\_when((location == 'NC' & date == '7/20/2023') ~ 'R3',  
 .default = as.character(growth\_stage))) %>%   
 dplyr::select(location, year, growth\_stage, plotid, block, treatmetn, to.predated) %>%   
 dplyr::rename(plot\_id = plotid,   
 treatment = treatmetn) %>%  
 group\_by(location, year, growth\_stage, plot\_id, block, treatment) %>%   
 na.omit() %>%   
 filter(treatment != 5) %>%   
 mutate\_at(vars(1:6),as.factor)  
  
# and in the darkness, bind them   
sent <- rbind(clean21, clean22, clean23)  
as\_tibble(sent)

## # A tibble: 12,432 × 7  
## location year growth\_stage plot\_id block treatment to.predated  
## <fct> <fct> <fct> <fct> <fct> <fct> <dbl>  
## 1 PA 2021 V3 101 1 1 0  
## 2 PA 2021 V3 101 1 1 0  
## 3 PA 2021 V3 101 1 1 0  
## 4 PA 2021 V3 101 1 1 0  
## 5 PA 2021 V3 101 1 1 0  
## 6 PA 2021 V3 101 1 1 0  
## 7 PA 2021 V3 102 1 3 1  
## 8 PA 2021 V3 102 1 3 1  
## 9 PA 2021 V3 102 1 3 0  
## 10 PA 2021 V3 102 1 3 1  
## # ℹ 12,422 more rows

dim(sent)

## [1] 12432 7

unique(sent$location)

## [1] PA FL IL IA KS KY MD NE NY NC OH TX VT VA WI DE  
## Levels: DE FL IA IL KS KY MD NC NE NY OH PA TX VA VT WI

# Experimental design

Data: Sentinel Prey assessment of arthropod-predator activity in corn fields.

These data come from the Precision Sustainable Agriculture effort through the USDA. I am the lead on the entomology component of this project and responsible for analyzing this three year data set which spans multiple states. This effort began during my Master’s degree, but I only analyzed Pennsylvania data for my thesis.

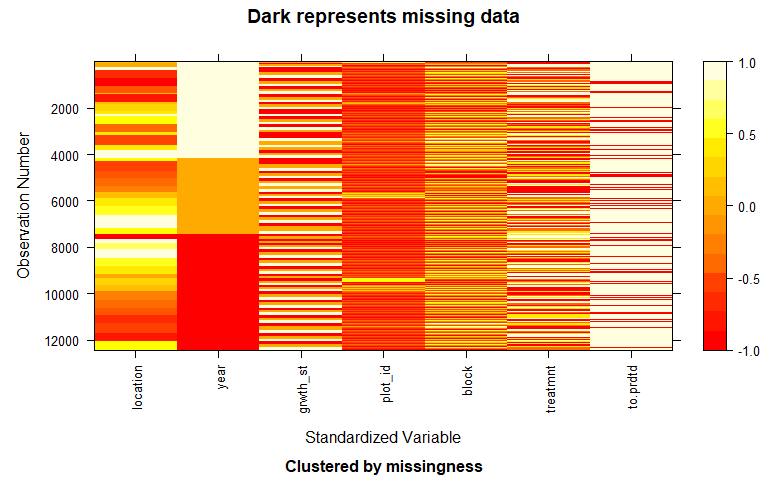
**Study design:**  **Treatments** = 4; No cover crop, early-terminated cover crop, late-terminated cover crop, planting green **Plots** = 20; 5-6 blocks (location dependent) composed of 4 plots each = 20 plots / study site / year **Years** = 3 (2021,2022,2023) **Locations** = This project comprises 16 states. Not all states collected sentinel prey data every year. Each site year was in a different field. **Effort** = Data were collected at three corn growth stages / year (V3,V5,R3). **Sample** = 6 sentinel prey traps were placed in each plot = 120 samples collected / growth stage. Total sample effort per state per season = 360 samples.

**Variables:** **Response** = Total level of predation. This is a binomial of 6 traps/ counts per plot. Pseudoreplication is account for in the random term. **Explanatory** = Crop growth stage (timing, three levels) and treatment (four levels). I am not interested in the fixed effects of location. **Random effects** = Plot in block in location, in year. I want to account for pseduoreplication and all of the site/year combinations.

Plots were randomly assigned to each block. Field sites were as random as they could be at each respective research station. Sentinel prey traps were placed between pre-determined rows and at specific length intervals within each plot to maintain consistency.

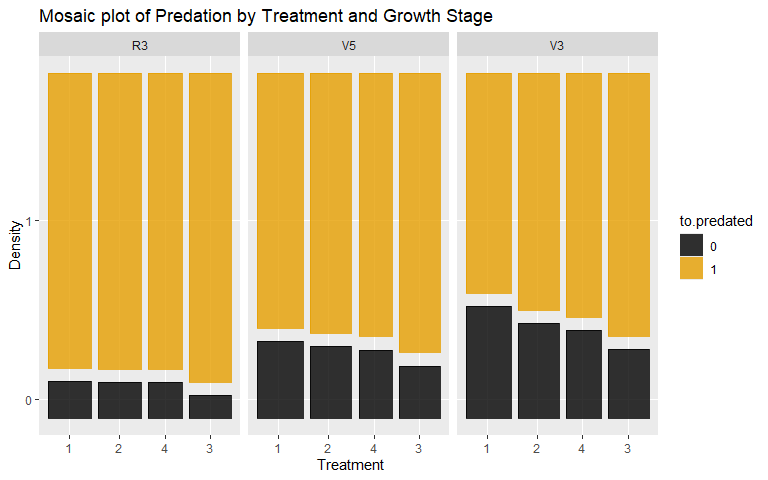
# Locating missing values

library(mi)  
# make an object of the missing df and then present the image  
tdf <- missing\_data.frame(data.frame(sent))  
image(tdf)



# Preliminary data display

sent %>%   
 ggplot() +  
 geom\_mosaic(aes(x = product(treatment), fill = to.predated), offset = 0.04)+  
 facet\_wrap(~growth\_stage)+  
 scale\_fill\_colorblind()+  
 labs(title = 'Mosaic plot of Predation by Treatment and Growth Stage',  
 y = 'Density',  
 x = ' Treatment')



# Modeling

Theoretical model:

${Predation|Treatment,GrowthStage} \sim \beta\_0 + \beta\_1I\_{Treatment = 2}\_i + \beta\_2I\_{Treatment = 4}\_i + \beta\_3I\_{Treatment = 3}\_i + \beta\_4I\_{GrowthStage = V5}\_i + \beta\_5I\_{GrowthStage = V3}\_i + \Year\_i + \Location\_i + \Block\_i + \Plot\_i + \Error\_i$

Add the appropriate stuff below

$\pi\{Predation|Treatment, GrowthStage} \sim \beta\_0 + \beta\_1I\_{Treatment = 2} + \beta\_2I\_{Treatment = 4} + \beta\_3I\_{Treatment = 3} + \beta\_4I\_{GrowthStage = V5} + \beta\_5I\_{GrowthStage = V3}$

## Initial model

sent

## # A tibble: 12,432 × 7  
## # Groups: location, year, growth\_stage, plot\_id, block, treatment [2,091]  
## location year growth\_stage plot\_id block treatment to.predated  
## <fct> <fct> <fct> <fct> <fct> <fct> <dbl>  
## 1 PA 2021 V3 101 1 1 0  
## 2 PA 2021 V3 101 1 1 0  
## 3 PA 2021 V3 101 1 1 0  
## 4 PA 2021 V3 101 1 1 0  
## 5 PA 2021 V3 101 1 1 0  
## 6 PA 2021 V3 101 1 1 0  
## 7 PA 2021 V3 102 1 3 1  
## 8 PA 2021 V3 102 1 3 1  
## 9 PA 2021 V3 102 1 3 0  
## 10 PA 2021 V3 102 1 3 1  
## # ℹ 12,422 more rows

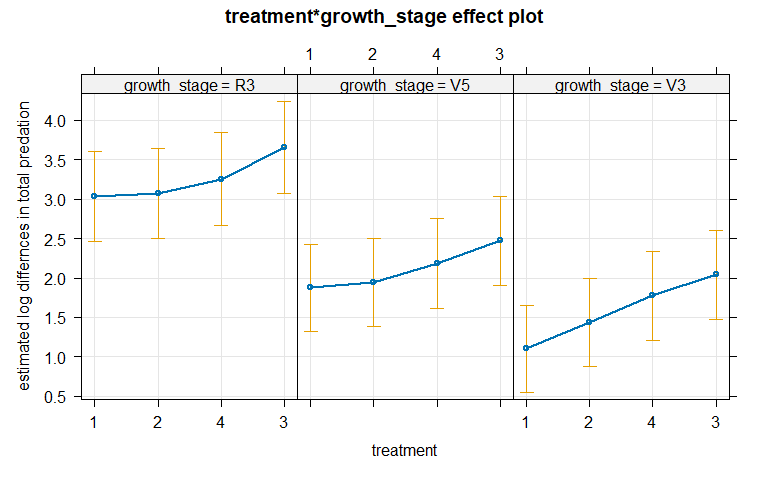
m2 <- glmer(to.predated ~ treatment\*growth\_stage + (1|year/location/block/plot\_id) , family = binomial, data = sent)  
summary(m2)

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula:   
## to.predated ~ treatment \* growth\_stage + (1 | year/location/block/plot\_id)  
## Data: sent  
##   
## AIC BIC logLik deviance df.resid   
## 9135.0 9253.8 -4551.5 9103.0 12416   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -11.4203 0.0801 0.2125 0.4224 2.9757   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## plot\_id:(block:(location:year)) (Intercept) 2.611e-01 0.511004  
## block:(location:year) (Intercept) 1.701e-01 0.412483  
## location:year (Intercept) 2.689e+00 1.639802  
## year (Intercept) 1.545e-05 0.003931  
## Number of obs: 12432, groups:   
## plot\_id:(block:(location:year)), 744; block:(location:year), 191; location:year, 39; year, 3  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 3.03307 0.29061 10.437 < 2e-16  
## treatment2 0.03549 0.16020 0.222 0.824674  
## treatment4 0.22142 0.17252 1.283 0.199352  
## treatment3 0.62055 0.17515 3.543 0.000396  
## growth\_stageV5 -1.15589 0.13170 -8.776 < 2e-16  
## growth\_stageV3 -1.92785 0.13398 -14.389 < 2e-16  
## treatment2:growth\_stageV5 0.03110 0.18796 0.165 0.868581  
## treatment4:growth\_stageV5 0.08776 0.20154 0.435 0.663242  
## treatment3:growth\_stageV5 -0.02574 0.20392 -0.126 0.899541  
## treatment2:growth\_stageV3 0.29511 0.18920 1.560 0.118818  
## treatment4:growth\_stageV3 0.44923 0.20103 2.235 0.025439  
## treatment3:growth\_stageV3 0.31249 0.20397 1.532 0.125508  
##   
## Correlation of Fixed Effects:  
## (Intr) trtmn2 trtmn4 trtmn3 grw\_V5 grw\_V3 t2:\_V5 t4:\_V5 t3:\_V5  
## treatment2 -0.272   
## treatment4 -0.248 0.458   
## treatment3 -0.246 0.453 0.421   
## grwth\_stgV5 -0.293 0.511 0.472 0.466   
## grwth\_stgV3 -0.300 0.511 0.474 0.464 0.647   
## trtmnt2:\_V5 0.199 -0.724 -0.332 -0.327 -0.693 -0.446   
## trtmnt4:\_V5 0.185 -0.334 -0.719 -0.305 -0.646 -0.415 0.453   
## trtmnt3:\_V5 0.184 -0.330 -0.306 -0.744 -0.639 -0.412 0.448 0.417   
## trtmnt2:\_V3 0.203 -0.727 -0.334 -0.330 -0.451 -0.692 0.622 0.294 0.290  
## trtmnt4:\_V3 0.192 -0.339 -0.728 -0.310 -0.424 -0.655 0.297 0.623 0.273  
## trtmnt3:\_V3 0.191 -0.335 -0.309 -0.747 -0.420 -0.645 0.292 0.272 0.645  
## t2:\_V3 t4:\_V3  
## treatment2   
## treatment4   
## treatment3   
## grwth\_stgV5   
## grwth\_stgV3   
## trtmnt2:\_V5   
## trtmnt4:\_V5   
## trtmnt3:\_V5   
## trtmnt2:\_V3   
## trtmnt4:\_V3 0.459   
## trtmnt3:\_V3 0.454 0.427  
## optimizer (Nelder\_Mead) convergence code: 0 (OK)  
## Model failed to converge with max|grad| = 0.0296809 (tol = 0.002, component 1)

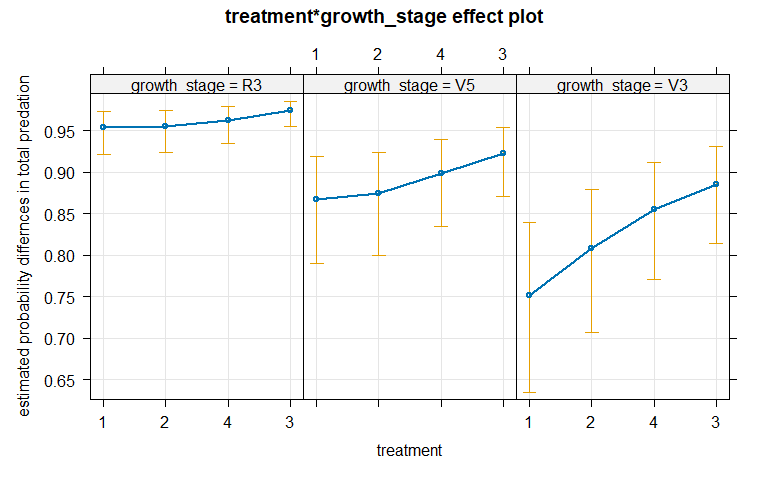
Anova(m2)

## Analysis of Deviance Table (Type II Wald chisquare tests)  
##   
## Response: to.predated  
## Chisq Df Pr(>Chisq)  
## treatment 60.9286 3 3.722e-13  
## growth\_stage 526.0579 2 < 2.2e-16  
## treatment:growth\_stage 8.2544 6 0.22

plot(allEffects(m2), type = 'link',ylab = 'estimated log differnces in total predation', grid = T)



plot(allEffects(m2), type = 'response', xlim = c(0,1),ylab = 'estimated probability differnces in total predation', grid = T)

 \* There is weak evidence against the null of no interaction between treatment and growth stage Chi squared(6) = 8.25, p = 0.22, after controlling for the random effect of plot nested in block nested in location nested in year, and will remove the interaction term from the model.

## Model refinement

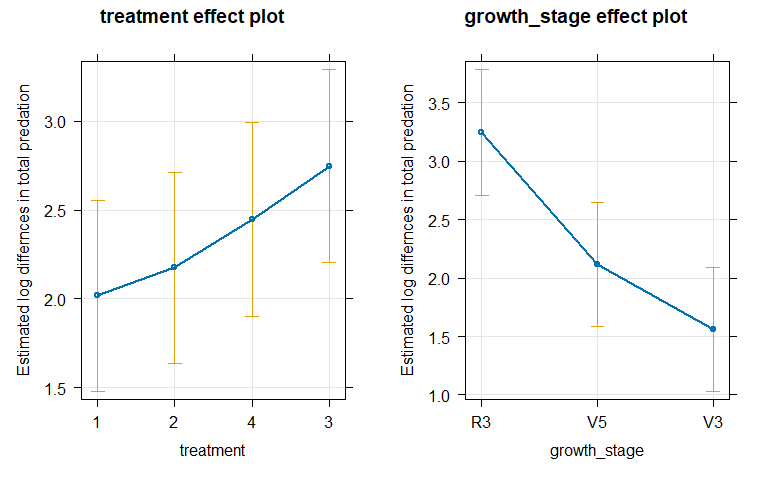
m3 <- glmer(to.predated ~ treatment+growth\_stage + (1|year/location/block/plot\_id) , family = binomial, data = sent)  
summary(m3)

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula:   
## to.predated ~ treatment + growth\_stage + (1 | year/location/block/plot\_id)  
## Data: sent  
##   
## AIC BIC logLik deviance df.resid   
## 9131.1 9205.4 -4555.5 9111.1 12422   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -11.4488 0.0811 0.2124 0.4229 2.9455   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## plot\_id:(block:(location:year)) (Intercept) 2.600e-01 0.50988   
## block:(location:year) (Intercept) 1.700e-01 0.41236   
## location:year (Intercept) 2.672e+00 1.63473   
## year (Intercept) 5.569e-06 0.00236   
## Number of obs: 12432, groups:   
## plot\_id:(block:(location:year)), 744; block:(location:year), 191; location:year, 39; year, 3  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 2.92687 0.27926 10.481 < 2e-16  
## treatment2 0.15828 0.09462 1.673 0.0944  
## treatment4 0.42821 0.10271 4.169 3.06e-05  
## treatment3 0.72957 0.09942 7.338 2.16e-13  
## growth\_stageV5 -1.13027 0.07256 -15.578 < 2e-16  
## growth\_stageV3 -1.68191 0.07334 -22.932 < 2e-16  
##   
## Correlation of Fixed Effects:  
## (Intr) trtmn2 trtmn4 trtmn3 grw\_V5  
## treatment2 -0.161   
## treatment4 -0.138 0.451   
## treatment3 -0.141 0.468 0.436   
## grwth\_stgV5 -0.172 -0.003 -0.015 -0.026   
## grwth\_stgV3 -0.176 -0.007 -0.024 -0.041 0.625  
## optimizer (Nelder\_Mead) convergence code: 0 (OK)  
## Model failed to converge with max|grad| = 0.0104819 (tol = 0.002, component 1)

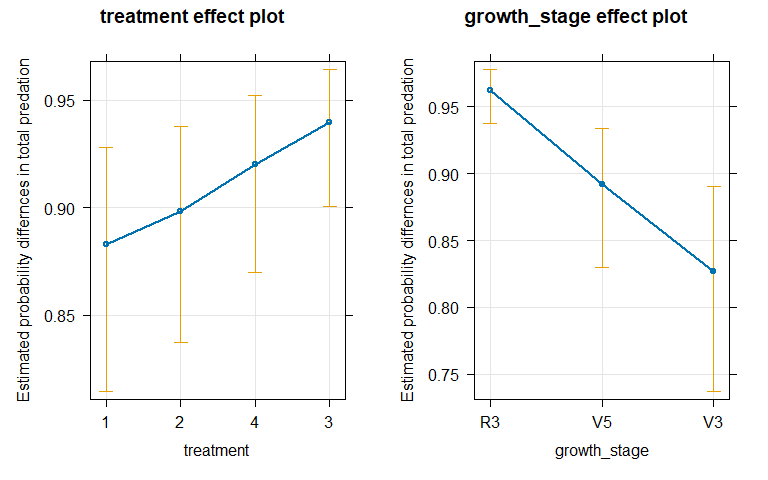
Anova(m3)

## Analysis of Deviance Table (Type II Wald chisquare tests)  
##   
## Response: to.predated  
## Chisq Df Pr(>Chisq)  
## treatment 60.988 3 3.615e-13  
## growth\_stage 528.427 2 < 2.2e-16

plot(allEffects(m3), type = 'link', ylab = 'Estimated log differnces in total predation', grid = T)



plot(allEffects(m3), type = 'response', xlim = c(0,1),ylab = 'Estimated probability differnces in total predation', grid = T)



cld(emmeans(m3, ~treatment, adjust = 'tukey', type = 'response'), Letters = letters)

## treatment prob SE df asymp.LCL asymp.UCL .group  
## 1 0.880 0.0290 Inf 0.787 0.935 a   
## 2 0.895 0.0257 Inf 0.812 0.944 a   
## 4 0.918 0.0209 Inf 0.849 0.957 b   
## 3 0.938 0.0161 Inf 0.884 0.968 c   
##   
## Results are averaged over the levels of: growth\_stage   
## Confidence level used: 0.95   
## Conf-level adjustment: sidak method for 4 estimates   
## Intervals are back-transformed from the logit scale   
## P value adjustment: tukey method for comparing a family of 4 estimates   
## Tests are performed on the log odds ratio scale   
## significance level used: alpha = 0.05   
## NOTE: If two or more means share the same grouping symbol,  
## then we cannot show them to be different.  
## But we also did not show them to be the same.

cld(emmeans(m3, ~growth\_stage, adjust = 'tukey', type = 'response'), Letters = letters)

## growth\_stage prob SE df asymp.LCL asymp.UCL .group  
## V3 0.828 0.03850 Inf 0.717 0.902 a   
## V5 0.893 0.02584 Inf 0.814 0.941 b   
## R3 0.963 0.00981 Inf 0.931 0.980 c   
##   
## Results are averaged over the levels of: treatment   
## Confidence level used: 0.95   
## Conf-level adjustment: sidak method for 3 estimates   
## Intervals are back-transformed from the logit scale   
## P value adjustment: tukey method for comparing a family of 3 estimates   
## Tests are performed on the log odds ratio scale   
## significance level used: alpha = 0.05   
## NOTE: If two or more means share the same grouping symbol,  
## then we cannot show them to be different.  
## But we also did not show them to be the same.

### R squared and ICC

r.squaredGLMM(m3)

## R2m R2c  
## theoretical 0.08286345 0.527981  
## delta 0.05596765 0.356609

library(performance)  
r2(m3)

## # R2 for Mixed Models  
##   
## Conditional R2: 0.528  
## Marginal R2: 0.083

summary(m3)

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula:   
## to.predated ~ treatment + growth\_stage + (1 | year/location/block/plot\_id)  
## Data: sent  
##   
## AIC BIC logLik deviance df.resid   
## 9131.1 9205.4 -4555.5 9111.1 12422   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -11.4488 0.0811 0.2124 0.4229 2.9455   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## plot\_id:(block:(location:year)) (Intercept) 2.600e-01 0.50988   
## block:(location:year) (Intercept) 1.700e-01 0.41236   
## location:year (Intercept) 2.672e+00 1.63473   
## year (Intercept) 5.569e-06 0.00236   
## Number of obs: 12432, groups:   
## plot\_id:(block:(location:year)), 744; block:(location:year), 191; location:year, 39; year, 3  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 2.92687 0.27926 10.481 < 2e-16  
## treatment2 0.15828 0.09462 1.673 0.0944  
## treatment4 0.42821 0.10271 4.169 3.06e-05  
## treatment3 0.72957 0.09942 7.338 2.16e-13  
## growth\_stageV5 -1.13027 0.07256 -15.578 < 2e-16  
## growth\_stageV3 -1.68191 0.07334 -22.932 < 2e-16  
##   
## Correlation of Fixed Effects:  
## (Intr) trtmn2 trtmn4 trtmn3 grw\_V5  
## treatment2 -0.161   
## treatment4 -0.138 0.451   
## treatment3 -0.141 0.468 0.436   
## grwth\_stgV5 -0.172 -0.003 -0.015 -0.026   
## grwth\_stgV3 -0.176 -0.007 -0.024 -0.041 0.625  
## optimizer (Nelder\_Mead) convergence code: 0 (OK)  
## Model failed to converge with max|grad| = 0.0104819 (tol = 0.002, component 1)

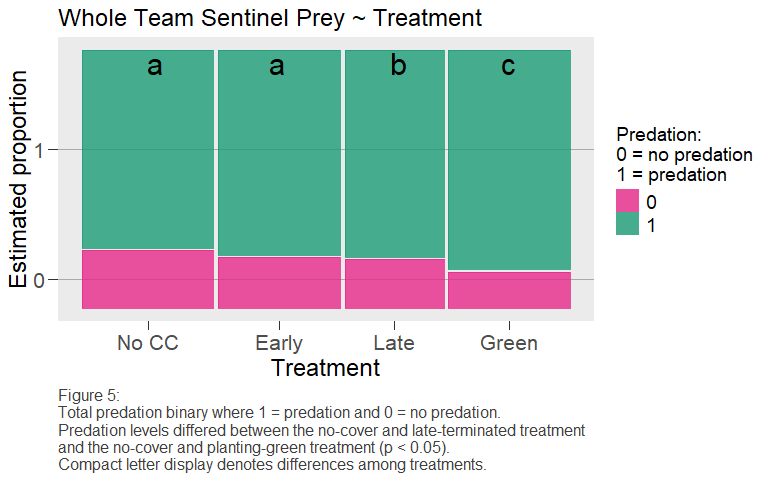
# for plot   
(2.600e-01) / (2.600e-01 + 1.700e-01 + .672e+00 + 5.569e-06)

## [1] 0.2359335

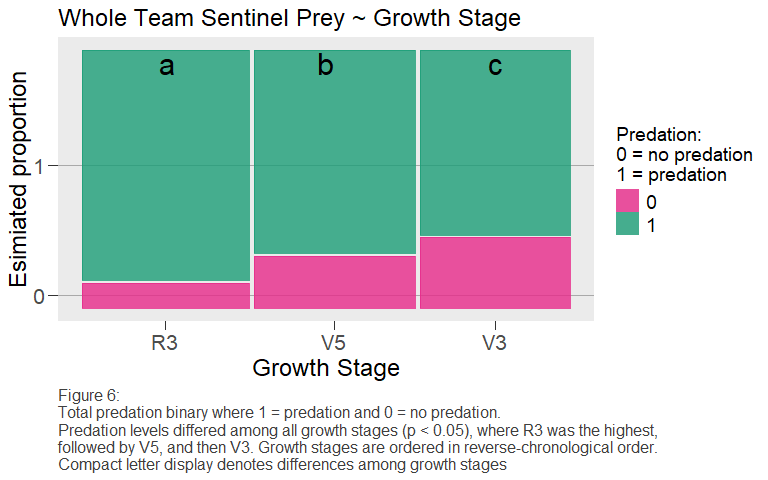
The ICC for plot id is 0.236.

# Final plots

sent %>%   
 ggplot()+  
 geom\_mosaic(aes(x = product(treatment), fill = to.predated))+  
 scale\_fill\_manual(values = c("#E7298A","#1B9E77"))+  
 scale\_x\_productlist(labels=c("No CC", "Early", "Late", "Green"))+  
 labs(title = 'Whole Team Sentinel Prey ~ Treatment',  
 x = 'Treatment',  
 y = 'Estimated proportion',  
 caption = 'Figure 5:\nTotal predation binary where 1 = predation and 0 = no predation.\nPredation levels differed between the no-cover and late-terminated treatment\nand the no-cover and planting-green treatment (p < 0.05).\nCompact letter display denotes differences among treatments.')+  
 guides(fill = guide\_legend(title = 'Predation:\n0 = no predation\n1 = predation'))+  
 theme(legend.title = element\_text(size = 14),  
 legend.text = element\_text(size = 14),  
 axis.text.x = element\_text(size=16),  
 axis.text.y = element\_text(size = 16),  
 axis.title = element\_text(size = 18),  
 plot.title = element\_text(size = 18),  
 plot.subtitle = element\_text(size = 12),   
 panel.grid.major.y = element\_line(color = "darkgrey"),  
 panel.grid.major.x = element\_blank(),  
 panel.grid.minor = element\_blank(),  
 strip.text = element\_text(size = 12),  
 plot.caption = element\_text(hjust = 0, size = 12, color = "grey25"),  
 axis.ticks.length=unit(.25, "cm"))+  
 annotate('text', x = 0.15, y = 0.95, label = 'a', size = 8)+  
 annotate('text', x = .4, y = 0.95, label = 'a', size = 8)+  
 annotate('text', x = .65, y = 0.95, label = 'b', size = 8)+  
 annotate('text', x = .875, y = 0.95, label = 'c', size = 8)



sent %>%   
 ggplot()+  
 geom\_mosaic(aes(x = product(growth\_stage), fill = to.predated))+  
 scale\_fill\_manual(values = c("#E7298A","#1B9E77"))+  
 labs(title = 'Whole Team Sentinel Prey ~ Growth Stage',  
 x = 'Growth Stage',  
 y = 'Esimiated proportion',  
 caption = 'Figure 6:\nTotal predation binary where 1 = predation and 0 = no predation.\nPredation levels differed among all growth stages (p < 0.05), where R3 was the highest,\nfollowed by V5, and then V3. Growth stages are ordered in reverse-chronological order.\nCompact letter display denotes differences among growth stages')+  
 guides(fill = guide\_legend(title = 'Predation:\n0 = no predation\n1 = predation'))+  
 theme(legend.title = element\_text(size = 14),  
 legend.text = element\_text(size = 14),  
 axis.text.x = element\_text(size=16),  
 axis.text.y = element\_text(size = 16),  
 axis.title = element\_text(size = 18),  
 plot.title = element\_text(size = 18),  
 plot.subtitle = element\_text(size = 12),   
 panel.grid.major.y = element\_line(color = "darkgrey"),  
 panel.grid.major.x = element\_blank(),  
 panel.grid.minor = element\_blank(),  
 strip.text = element\_text(size = 12),  
 plot.caption = element\_text(hjust = 0, size = 12, color = "grey25"),  
 axis.ticks.length=unit(.25, "cm"))+  
 annotate('text', x = 0.175, y = 0.95, label = 'a', size = 8)+  
 annotate('text', x = .5, y = 0.95, label = 'b', size = 8)+  
 annotate('text', x = .85, y = 0.95, label = 'c', size = 8)



# Table 1

# table as a proportion   
sent %>%   
 group\_by(location, treatment, growth\_stage) %>%   
 summary()

## location year growth\_stage plot\_id block treatment  
## IL :1296 2021:5004 R3:4334 203 : 608 1:2681 1:3439   
## OH :1173 2022:3246 V5:4196 103 : 607 2:2491 2:3203   
## VT :1080 2023:4182 V3:3902 303 : 607 3:2493 4:2605   
## PA :1074 401 : 603 4:2514 3:3185   
## KY :1054 102 : 591 5:1969   
## KS : 848 202 : 589 6: 284   
## (Other):5907 (Other):8827   
## to.predated   
## Min. :0.0000   
## 1st Qu.:1.0000   
## Median :1.0000   
## Mean :0.8103   
## 3rd Qu.:1.0000   
## Max. :1.0000   
##

sent %>%   
 group\_by(growth\_stage) %>%   
 summary()

## location year growth\_stage plot\_id block treatment  
## IL :1296 2021:5004 R3:4334 203 : 608 1:2681 1:3439   
## OH :1173 2022:3246 V5:4196 103 : 607 2:2491 2:3203   
## VT :1080 2023:4182 V3:3902 303 : 607 3:2493 4:2605   
## PA :1074 401 : 603 4:2514 3:3185   
## KY :1054 102 : 591 5:1969   
## KS : 848 202 : 589 6: 284   
## (Other):5907 (Other):8827   
## to.predated   
## Min. :0.0000   
## 1st Qu.:1.0000   
## Median :1.0000   
## Mean :0.8103   
## 3rd Qu.:1.0000   
## Max. :1.0000   
##

tally(to.predated ~ treatment + growth\_stage, data = sent)

## , , growth\_stage = R3  
##   
## treatment  
## to.predated 1 2 4 3  
## 0 128 125 96 77  
## 1 1018 1024 802 1064  
##   
## , , growth\_stage = V5  
##   
## treatment  
## to.predated 1 2 4 3  
## 0 277 232 179 167  
## 1 918 836 692 895  
##   
## , , growth\_stage = V3  
##   
## treatment  
## to.predated 1 2 4 3  
## 0 371 281 221 204  
## 1 727 705 615 778

datasummary\_balance( ~ to.predated,data = sent)

|  | | 0 | | 1 | |
| --- | --- | --- | --- | --- | --- |
|  |  | N | Pct. | N | Pct. |
| location | DE | 193 | 8.2 | 405 | 4.0 |
|  | FL | 19 | 0.8 | 692 | 6.9 |
|  | IA | 16 | 0.7 | 705 | 7.0 |
|  | IL | 95 | 4.0 | 1201 | 11.9 |
|  | KS | 345 | 14.6 | 503 | 5.0 |
|  | KY | 373 | 15.8 | 681 | 6.8 |
|  | MD | 192 | 8.1 | 408 | 4.1 |
|  | NC | 19 | 0.8 | 341 | 3.4 |
|  | NE | 65 | 2.8 | 706 | 7.0 |
|  | NY | 94 | 4.0 | 474 | 4.7 |
|  | OH | 531 | 22.5 | 642 | 6.4 |
|  | PA | 240 | 10.2 | 834 | 8.3 |
|  | TX | 22 | 0.9 | 698 | 6.9 |
|  | VA | 49 | 2.1 | 239 | 2.4 |
|  | VT | 38 | 1.6 | 1042 | 10.3 |
|  | WI | 67 | 2.8 | 503 | 5.0 |
| year | 2021 | 883 | 37.4 | 4121 | 40.9 |
|  | 2022 | 676 | 28.7 | 2570 | 25.5 |
|  | 2023 | 799 | 33.9 | 3383 | 33.6 |
| growth\_stage | R3 | 426 | 18.1 | 3908 | 38.8 |
|  | V5 | 855 | 36.3 | 3341 | 33.2 |
|  | V3 | 1077 | 45.7 | 2825 | 28.0 |
| plot\_id | 101 | 161 | 6.8 | 425 | 4.2 |
|  | 102 | 112 | 4.7 | 479 | 4.8 |
|  | 103 | 126 | 5.3 | 481 | 4.8 |
|  | 104 | 119 | 5.0 | 426 | 4.2 |
|  | 201 | 112 | 4.7 | 452 | 4.5 |
|  | 202 | 117 | 5.0 | 472 | 4.7 |
|  | 203 | 113 | 4.8 | 495 | 4.9 |
|  | 204 | 129 | 5.5 | 445 | 4.4 |
|  | 301 | 111 | 4.7 | 478 | 4.7 |
|  | 302 | 112 | 4.7 | 455 | 4.5 |
|  | 303 | 123 | 5.2 | 484 | 4.8 |
|  | 304 | 120 | 5.1 | 450 | 4.5 |
|  | 401 | 130 | 5.5 | 473 | 4.7 |
|  | 402 | 112 | 4.7 | 477 | 4.7 |
|  | 403 | 106 | 4.5 | 461 | 4.6 |
|  | 404 | 101 | 4.3 | 473 | 4.7 |
|  | 501 | 108 | 4.6 | 400 | 4.0 |
|  | 502 | 79 | 3.4 | 422 | 4.2 |
|  | 503 | 82 | 3.5 | 442 | 4.4 |
|  | 504 | 71 | 3.0 | 386 | 3.8 |
|  | 601 | 2 | 0.1 | 111 | 1.1 |
|  | 602 | 1 | 0.0 | 75 | 0.7 |
|  | 603 | 9 | 0.4 | 86 | 0.9 |
|  | 604 | 11 | 0.5 | 96 | 1.0 |
|  | 505 | 5 | 0.2 | 22 | 0.2 |
|  | 506 | 3 | 0.1 | 33 | 0.3 |
|  | 701 | 5 | 0.2 | 49 | 0.5 |
|  | 702 | 2 | 0.1 | 34 | 0.3 |
|  | 706 | 2 | 0.1 | 34 | 0.3 |
|  | 803 | 3 | 0.1 | 33 | 0.3 |
|  | 804 | 7 | 0.3 | 25 | 0.2 |
|  | 805 | 1 | 0.0 | 35 | 0.3 |
|  | 901 | 4 | 0.2 | 32 | 0.3 |
|  | 905 | 6 | 0.3 | 27 | 0.3 |
|  | 906 | 4 | 0.2 | 50 | 0.5 |
|  | 105 | 2 | 0.1 | 16 | 0.2 |
|  | 106 | 11 | 0.5 | 7 | 0.1 |
|  | 205 | 3 | 0.1 | 15 | 0.1 |
|  | 206 | 3 | 0.1 | 15 | 0.1 |
|  | 305 | 4 | 0.2 | 14 | 0.1 |
|  | 306 | 5 | 0.2 | 13 | 0.1 |
|  | 405 | 6 | 0.3 | 12 | 0.1 |
|  | 406 | 4 | 0.2 | 13 | 0.1 |
|  | 605 | 0 | 0.0 | 30 | 0.3 |
|  | 705 | 2 | 0.1 | 33 | 0.3 |
|  | 801 | 3 | 0.1 | 32 | 0.3 |
|  | 902 | 1 | 0.0 | 17 | 0.2 |
|  | 606 | 0 | 0.0 | 9 | 0.1 |
|  | 802 | 4 | 0.2 | 13 | 0.1 |
|  | 904 | 1 | 0.0 | 17 | 0.2 |
| block | 1 | 631 | 26.8 | 2050 | 20.3 |
|  | 2 | 479 | 20.3 | 2012 | 20.0 |
|  | 3 | 468 | 19.8 | 2025 | 20.1 |
|  | 4 | 435 | 18.4 | 2079 | 20.6 |
|  | 5 | 331 | 14.0 | 1638 | 16.3 |
|  | 6 | 14 | 0.6 | 270 | 2.7 |
| treatment | 1 | 776 | 32.9 | 2663 | 26.4 |
|  | 2 | 638 | 27.1 | 2565 | 25.5 |
|  | 4 | 496 | 21.0 | 2109 | 20.9 |
|  | 3 | 448 | 19.0 | 2737 | 27.2 |

# install.packages('openxlsx')  
# library(openxlsx)  
#   
# table <- favstats(to.predated ~ treatment + growth\_stage, data = sent)  
# write.xlsx(table, 'test.xlsx')  
# favstats(to.predated ~ growth\_stage, data = sent)  
# favstats(to.predated ~ treatment, data = sent)  
# citation('mosaic')

# Study design figure

tally(~ year, sent)

## year  
## 2021 2022 2023   
## 5004 3246 4182

tally(~location, sent)

## location  
## DE FL IA IL KS KY MD NC NE NY OH PA TX VA VT WI   
## 598 711 721 1296 848 1054 600 360 771 568 1173 1074 720 288 1080 570

tally(~block, sent)

## block  
## 1 2 3 4 5 6   
## 2681 2491 2493 2514 1969 284

tally(~plot\_id, sent)

## plot\_id  
## 101 102 103 104 201 202 203 204 301 302 303 304 401 402 403 404 501 502 503 504   
## 586 591 607 545 564 589 608 574 589 567 607 570 603 589 567 574 508 501 524 457   
## 601 602 603 604 505 506 701 702 706 803 804 805 901 905 906 105 106 205 206 305   
## 113 76 95 107 27 36 54 36 36 36 32 36 36 33 54 18 18 18 18 18   
## 306 405 406 605 705 801 902 606 802 904   
## 18 18 17 30 35 35 18 9 17 18

names1 <- list("Year 2021", "Year 2022", "Year 2023")  
names2 <- list("Loc. 1", "...", "Loc. 16", "...", "Loc. 1", "...", "Loc. 16")  
names3 <- list("Block 1", "...", "Block 6")  
names4 <- list("Plot 1", "...", "Plot 4")  
namesF <- list("Treatment", "Growth Stage")  
namesR <- list("Year", "Location in Year", "Block in Location", "Observation Error")

# Citations

citation('DiagrammeRsvg')

## To cite package 'DiagrammeRsvg' in publications use:  
##   
## Iannone R (2016). \_DiagrammeRsvg: Export DiagrammeR Graphviz Graphs  
## as SVG\_. R package version 0.1,  
## <https://CRAN.R-project.org/package=DiagrammeRsvg>.  
##   
## A BibTeX entry for LaTeX users is  
##   
## @Manual{,  
## title = {DiagrammeRsvg: Export DiagrammeR Graphviz Graphs as SVG},  
## author = {Richard Iannone},  
## year = {2016},  
## note = {R package version 0.1},  
## url = {https://CRAN.R-project.org/package=DiagrammeRsvg},  
## }

citation('car')

## To cite the car package in publications use:  
##   
## Fox J, Weisberg S (2019). \_An R Companion to Applied Regression\_,  
## Third edition. Sage, Thousand Oaks CA.  
## <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.  
##   
## A BibTeX entry for LaTeX users is  
##   
## @Book{,  
## title = {An {R} Companion to Applied Regression},  
## edition = {Third},  
## author = {John Fox and Sanford Weisberg},  
## year = {2019},  
## publisher = {Sage},  
## address = {Thousand Oaks {CA}},  
## url = {https://socialsciences.mcmaster.ca/jfox/Books/Companion/},  
## }

citation('effects')

## To cite effects in publications use:  
##   
## John Fox and Sanford Weisberg (2019). An R Companion to Applied  
## Regression, 3rd Edition. Thousand Oaks, CA  
## <https://socialsciences.mcmaster.ca/jfox/Books/Companion/index.html>  
##   
## For predictor effects or partial residuals also cite:  
##   
## John Fox, Sanford Weisberg (2018). Visualizing Fit and Lack of Fit in  
## Complex Regression Models with Predictor Effect Plots and Partial  
## Residuals. Journal of Statistical Software, 87(9), 1-27. doi  
## 10.18637/jss.v087.i09  
##   
## For generalized linear models also cite:  
##   
## John Fox (2003). Effect Displays in R for Generalised Linear Models.  
## Journal of Statistical Software, 8(15), 1-27. doi  
## 10.18637/jss.v008.i15  
##   
## For usage in multinomial and proportional-odds logit models also cite:  
##   
## John Fox, Jangman Hong (2009). Effect Displays in R for Multinomial  
## and Proportional-Odds Logit Models: Extensions to the effects  
## Package. Journal of Statistical Software, 32(1), 1-24. doi  
## 10.18637/jss.v032.i01  
##   
## To see these entries in BibTeX format, use 'print(<citation>,  
## bibtex=TRUE)', 'toBibtex(.)', or set  
## 'options(citation.bibtex.max=999)'.

citation('emmeans')

## To cite package 'emmeans' in publications use:  
##   
## Lenth R (2024). \_emmeans: Estimated Marginal Means, aka Least-Squares  
## Means\_. R package version 1.10.4, commit  
## 52c445c77326050a97e70999b376424266495419,  
## <https://github.com/rvlenth/emmeans>.  
##   
## A BibTeX entry for LaTeX users is  
##   
## @Manual{,  
## title = {emmeans: Estimated Marginal Means, aka Least-Squares Means},  
## author = {Russell V. Lenth},  
## year = {2024},  
## note = {R package version 1.10.4, commit 52c445c77326050a97e70999b376424266495419},  
## url = {https://github.com/rvlenth/emmeans},  
## }