

Discovering Associations - Model 1

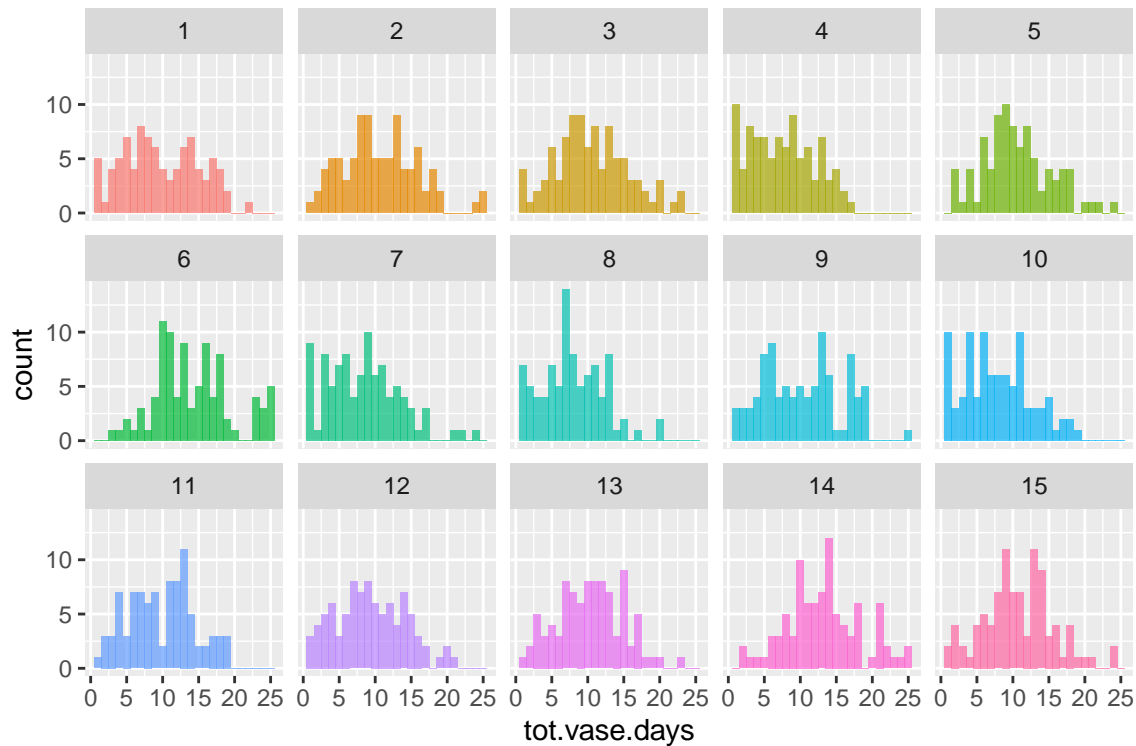
Count data of rose vase life days.

Poisson-mixed effects modelling.

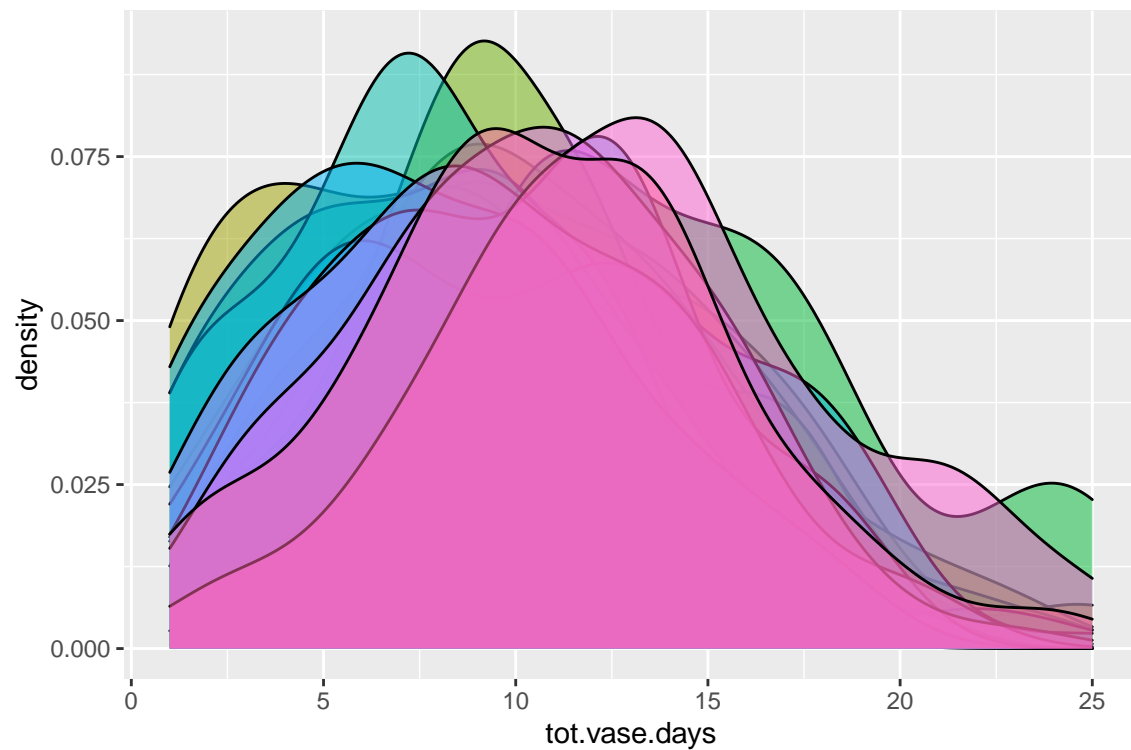
Out of 1440 datapoints, we have 60 missing outcomes (total vase days), which is 4.1% of the data. More data description, can copy some from simulation description.

Checking the distribution of the Vase Days of the count data. It looks like there is a larger degree of overdispersion than the poisson model currently accounts for (nope, it's just conditional on the compound: the goodfit should be done per compound).

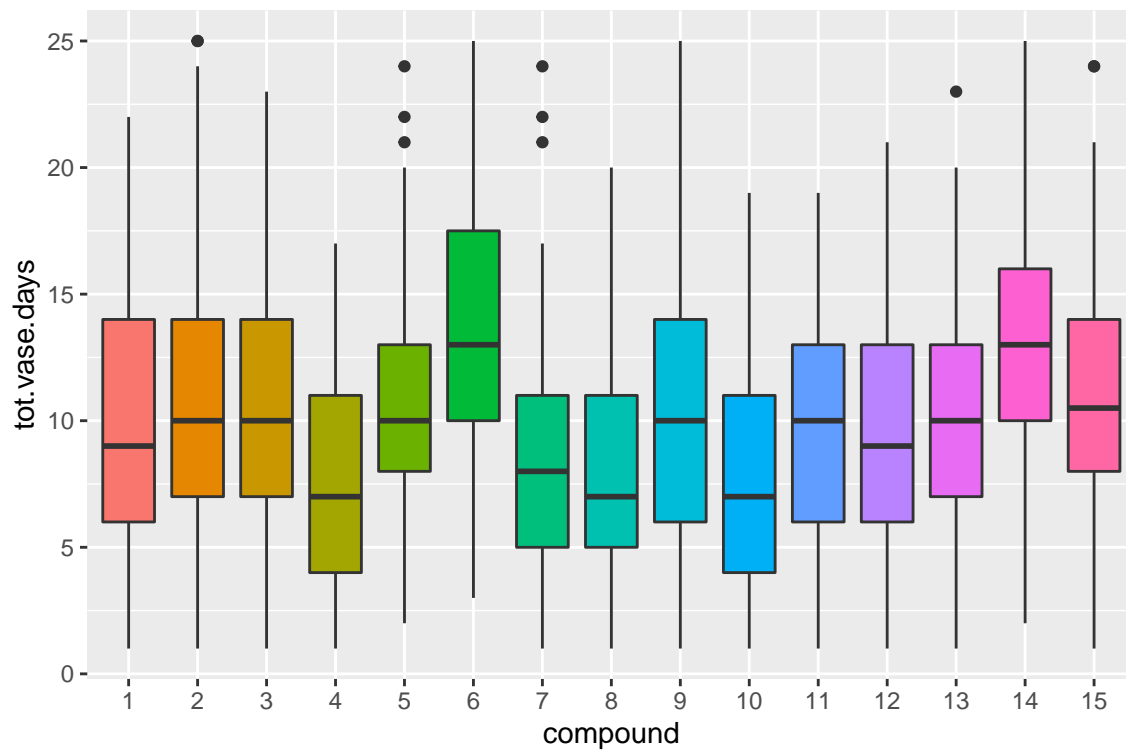
```
ggplot(d, aes(x=tot.vase.days, fill=compound))+geom_histogram(alpha=0.7, bins=25)+facet_wrap(~compound,
```



```
ggplot(d, aes(x=tot.vase.days, fill=compound))+geom_density(alpha=0.5)+theme(legend.position = "none")
```



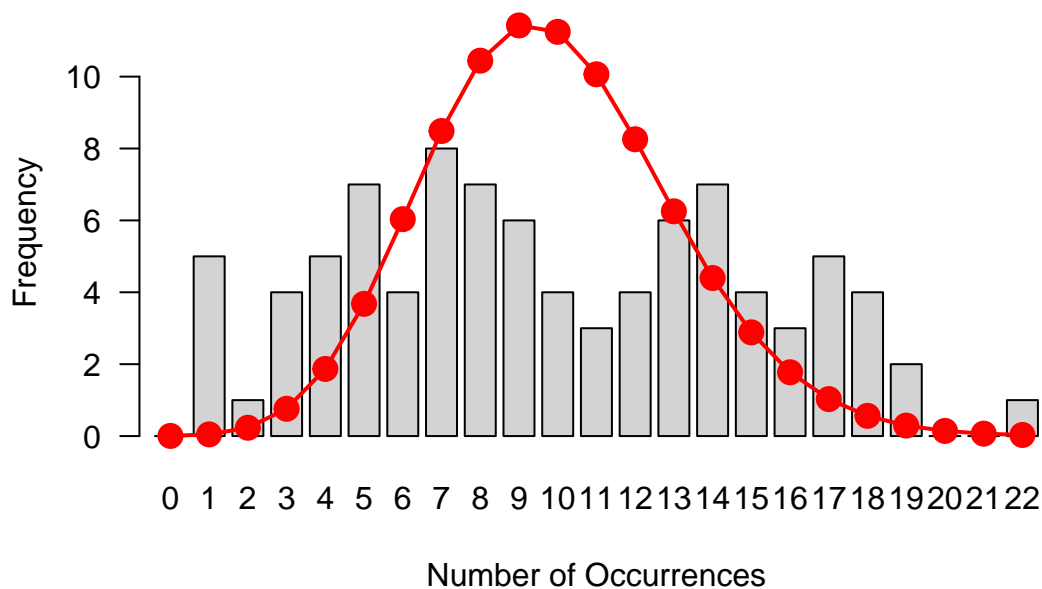
```
#ggplot(d %>% filter(compound %in% c(1, 6, 14)), aes(x=tot.vase.days, fill=as.factor(compound)))+geom_d
# d %>% group_by(compound) %>%
#   dplyr::summarise(mean=mean(tot.vase.days),
#                     var=var(tot.vase.days)) %>%
#   ggplot(aes(x=mean, y=var))+
#   geom_abline(slope=1, intercept=0)+
#   geom_point()
ggplot(d, aes(x=compound,y=tot.vase.days, fill=compound))+geom_boxplot()+theme(legend.position = "none")
```



```
gf <- goodfit(d$tot.vase.days[d$compound==1], "poisson")
summary(gf)
```

```
##
## Goodness-of-fit test for poisson distribution
##
##          X^2 df      P(> X^2)
## Likelihood Ratio 100.8575 18 1.542306e-13
```

```
plot(gf, type="standing", scale="raw") #I guess the fit would be better if we did it per compound
```



Initial Model (Model 0).

```
#https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html#overdispersion
overdisp_fun <- function(model) {
  rdf <- df.residual(model)
  rp <- residuals(model,type="pearson")
  Pearson.chisq <- sum(rp^2)
  pratt <- Pearson.chisq/rdf #pearson residuals over the residual degrees of freedom
  pval <- pchisq(Pearson.chisq, df=rdf, lower.tail=FALSE)
  c(chisq=Pearson.chisq, ratio=pratt, rdf=rdf, p=pval)
}
```

```
#inspired by the first answer here for nested data formulation https://stats.stackexchange.com/question
glmer_out <- glmer(tot.vase.days ~ compound + species + garden + (1|rater) + (1|subplotID/bushID), fami
summary(glmer_out)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: poisson ( log )
## Formula: tot.vase.days ~ compound + species + garden + (1 | rater) + (1 |
## subplotID/bushID)
## Data: d
##
##      AIC      BIC    logLik deviance df.resid
## 7222.6   7327.2  -3591.3   7182.6     1360
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -2.7555 -0.7878 -0.0459  0.6249  3.7431
##
## Random effects:
##   Groups          Name          Variance Std.Dev.
##   bushID:subplotID (Intercept) 0.01129  0.1062
##   subplotID        (Intercept) 0.05483  0.2342
##   rater             (Intercept) 0.12488  0.3534
## Number of obs: 1380, groups:  bushID:subplotID, 96; subplotID, 16; rater, 6
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.135101   0.171288  12.465 < 2e-16 ***
## compound2    0.070574   0.046365   1.522 0.127975
## compound3    0.063235   0.046126   1.371 0.170398
## compound4   -0.272911   0.050428  -5.412 6.24e-08 ***
## compound5    0.062465   0.046940   1.331 0.183279
## compound6    0.346820   0.043616   7.952 1.84e-15 ***
## compound7   -0.167143   0.048947  -3.415 0.000638 ***
## compound8   -0.248028   0.050442  -4.917 8.78e-07 ***
## compound9    0.042653   0.046612   0.915 0.360164
## compound10  -0.223922   0.049918  -4.486 7.26e-06 ***
## compound11   0.003953   0.047305   0.084 0.933410
## compound12  -0.055446   0.047967  -1.156 0.247713
## compound13   0.046951   0.046540   1.009 0.313063
## compound14   0.292411   0.044307   6.600 4.12e-11 ***
## compound15   0.082048   0.045998   1.784 0.074466 .
## species2    -0.033620   0.028069  -1.198 0.231009
## garden2      0.186699   0.120430   1.550 0.121077
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Correlation matrix not shown by default, as p = 17 > 12.
## Use print(x, correlation=TRUE) or
##      vcov(x)          if you need it
```

```
-2*logLik(glmer_out)
```

```
## 'log Lik.' 7182.625 (df=20)
```

```
AIC(glmer_out)
```

```
## [1] 7222.625
```

```
BIC(glmer_out)
```

```
## [1] 7327.222
```

```
overdisp_fun(glmer_out) #quite oki! no significance, yey. i think this might be more relevant than good
```

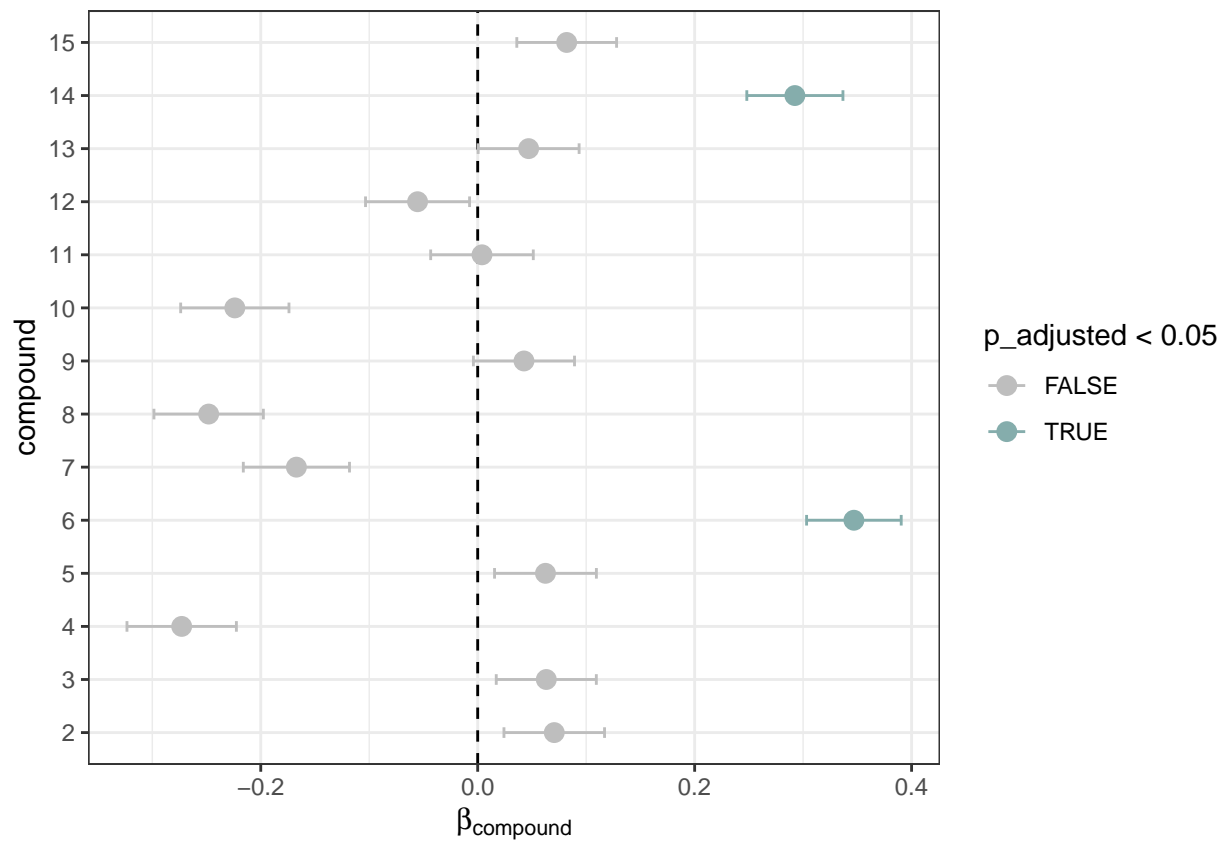
```
##          chisq          ratio          rdf          p
## 1416.6189390    1.0416316 1360.0000000    0.1392604
```

```
#https://fukamilab.github.io/BI0202/04-B-binary-data.html#glm_for_count_data
drop1(glmer_out, test = "Chi")
```

```
## Single term deletions
##
## Model:
## tot.vase.days ~ compound + species + garden + (1 | rater) + (1 |
## subplotID/bushID)
##      npar      AIC      LRT Pr(Chi)
## <none>      7222.6
## compound   14 7615.7 421.07 <2e-16 ***
## species    1 7222.1   1.43  0.2315
## garden     1 7222.9   2.23  0.1356
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
glmer_coefficients<-as.data.frame(summary(glmer_out)$coeff) %>%
  rownames_to_column("predictor") %>%
  filter(grepl("compound",predictor)) %>%
  dplyr::rename(pval=`Pr(>|z|)` ) %>%
  #we want to have p-adjusted (Holm) values for one-sided test H.alt: lambda(compound)>lambda(water)
  dplyr::mutate(one_sided_pval=ifelse(`z value`>0, pval/2, (1-pval/2)),
    p_adjusted=p.adjust(one_sided_pval, method="holm"),
    significant_higher=ifelse(p_adjusted<0.05, T, F))

ggplot(glmer_coefficients %>%
  mutate(compound=factor(gsub("compound","",predictor), levels=2:15)),
  aes(x=compound, y=Estimate, color=p_adjusted<0.05))+
  geom_hline(yintercept=0, linetype="dashed")+
  geom_errorbar(aes(ymin=Estimate - `Std. Error`, ymax=Estimate + `Std. Error`), width=0.2)+geom_point(s
  scale_color_manual(values=c("grey", "#85ADAC"))+coord_flip()+
  ylab(expression(beta["compound"])))
```



Conclusion: compounds 6 and 14 significantly increase rose vase days (mention estimates \pm sd error, also backcalculated in days, alpha, one-sided Wald test, maybe the exact z and p-values, correction Holm).

Models comparison.

Model 1. Dropping Compounds.

```
glmer_out1 <- glmer(tot.vase.days ~ species + garden + (1|rater) + (1|subplotID/bushID), family=poisson,
-2*logLik(glmer_out1))
```

```
## 'log Lik.' 7603.694 (df=6)
```

```
AIC(glmer_out1)
```

```
## [1] 7615.694
```

```
BIC(glmer_out1)
```

```
## [1] 7647.073
```

```
LRT.1=2*(logLik(glmer_out)-logLik(glmer_out1))
c(LRT.1,1-pchisq(LRT.1,1))
```

```
## [1] 421.0685 0.0000
```

Model 2. Dropping species.

```
glmer_out2 <- glmer(tot.vase.days ~ compound + garden + (1|rater) + (1|subplotID/bushID), family=poisson)
-2*logLik(glmer_out2)
```

```
## 'log Lik.' 7184.057 (df=19)
```

```
AIC(glmer_out2)
```

```
## [1] 7222.057
```

```
BIC(glmer_out2)
```

```
## [1] 7321.424
```

```
LRT.2=2*(logLik(glmer_out)-logLik(glmer_out2))
c(LRT.2,1-pchisq(LRT.2,1))
```

```
## [1] 1.431591 0.231505
```

Model 3. Dropping garden.

```
glmer_out3 <- glmer(tot.vase.days ~ compound + species + (1|rater) + (1|subplotID/bushID), family=poisson)
-2*logLik(glmer_out3)
```

```
## 'log Lik.' 7184.852 (df=19)
```

```
AIC(glmer_out3)
```

```
## [1] 7222.852
```

```
BIC(glmer_out3)
```

```
## [1] 7322.219
```

```
LRT.3=2*(logLik(glmer_out)-logLik(glmer_out3))
c(LRT.3,1-pchisq(LRT.3,1))
```

```
## [1] 2.2272167 0.1355979
```

Model 4.1 - Dropping bushID random effect.

```
glmer_out4.1 <- glmer(tot.vase.days ~ compound + garden + species + (1|rater) + (1|subplotID), family=poisson)
-2*logLik(glmer_out4.1)
```

```
## 'log Lik.' 7218.377 (df=19)
```



```
AIC(glmer_out4.1)
```

```
## [1] 7256.377
```

```
BIC(glmer_out4.1)
```

```
## [1] 7355.744
```

```
LRT.4.1=2*(logLik(glmer_out)-logLik(glmer_out4.1))  
c(LRT.4.1,1-pchisq(LRT.4.1,1))
```

```
## [1] 3.575219e+01 2.240798e-09
```

Model 4.2 - Dropping subplotID random effect

```
glmer_out4.2 <- glmer(tot.vase.days ~ compound + garden + species + (1|rater) + (1|bushID), family=poisson,  
-2*logLik(glmer_out4.2))
```

```
## 'log Lik.' 7261.358 (df=19)
```

```
AIC(glmer_out4.2)
```

```
## [1] 7299.358
```

```
BIC(glmer_out4.2)
```

```
## [1] 7398.725
```

```
LRT.4.2=2*(logLik(glmer_out)-logLik(glmer_out4.2))  
c(LRT.4.1,1-pchisq(LRT.4.1,1))
```

```
## [1] 3.575219e+01 2.240798e-09
```

Dropping rater as random effect.

```
glmer_out5 <- glmer(tot.vase.days ~ compound + garden + species + (1|subplotID/bushID), family=poisson(  
-2*logLik(glmer_out5))
```

```
## 'log Lik.' 7327.074 (df=19)
```

```
AIC(glmer_out5)
```

```
## [1] 7365.074
```

```
BIC(glmer_out5)
```

```
## [1] 7464.441
```

```
LRT.5=2*(logLik(glmer_out)-logLik(glmer_out5))  
c(LRT.5,1-pchisq(LRT.5,1))
```

```
## [1] 144.4489 0.0000
```

Model 6 is generated and compared to models 2 and 3.

```
glmer_out6 <- glmer(tot.vase.days ~ compound + (1|subplotID/bushID) + (1|rater), family=poisson(link =  
-2*logLik(glmer_out6))
```

```
## 'log Lik.' 7186.284 (df=18)
```

```
AIC(glmer_out6)
```

```
## [1] 7222.284
```

```
BIC(glmer_out6)
```

```
## [1] 7316.421
```

```
LRT.6.2=2*(logLik(glmer_out2)-logLik(glmer_out6))  
c(LRT.6.2,1-pchisq(LRT.6.2,1))
```

```
## [1] 2.2269663 0.1356199
```

```
LRT.6.3=2*(logLik(glmer_out3)-logLik(glmer_out6))  
c(LRT.6.3,1-pchisq(LRT.6.3,1))
```

```
## [1] 1.4313404 0.2315458
```

Model 6 is compared with models lacking any of the random effects. Model 7 without subplotID

```
glmer_out7 <- glmer(tot.vase.days ~ compound + (1|bushID) + (1|rater), family=poisson(link = "log"), da  
-2*logLik(glmer_out7))
```

```
## 'log Lik.' 7271.997 (df=17)
```

```
AIC(glmer_out7)
```

```
## [1] 7305.997
```

```
BIC(glmer_out7)
```

```
## [1] 7394.905
```

```
LRT.7=2*(logLik(glmer_out6)-logLik(glmer_out7))  
c(LRT.7,1-pchisq(LRT.7,1))
```

```
## [1] 85.71367 0.00000
```

Model 8 without bushID.

```
glmer_out8 <- glmer(tot.vase.days ~ compound + (1|subplotID) + (1|rater), family=poisson(link = "log"),  
-2*logLik(glmer_out8))
```

```
## 'log Lik.' 7222.654 (df=17)
```

```
AIC(glmer_out8)
```

```
## [1] 7256.654
```

```
BIC(glmer_out8)
```

```
## [1] 7345.561
```

```
LRT.8=2*(logLik(glmer_out6)-logLik(glmer_out8))  
c(LRT.8,1-pchisq(LRT.8,1))
```

```
## [1] 3.636999e+01 1.631966e-09
```

Model 9 without rater.

```
glmer_out9 <- glmer(tot.vase.days ~ compound + (1|subplotID/bushID), family=poisson(link = "log"), data  
-2*logLik(glmer_out9))
```

```
## 'log Lik.' 7329.569 (df=17)
```

```
AIC(glmer_out9)
```

```
## [1] 7363.569
```

```
BIC(glmer_out9)
```

```
## [1] 7452.477
```

```
LRT.9=2*(logLik(glmer_out6)-logLik(glmer_out9))
c(LRT.9,1-pchisq(LRT.9,1))
```

```
## [1] 143.2858 0.0000
```

Results using model 6

```
summary(glmer_out6)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: poisson ( log )
## Formula: tot.vase.days ~ compound + (1 | subplotID/bushID) + (1 | rater)
## Data: d
##
##      AIC      BIC    logLik deviance df.resid
##  7222.3   7316.4  -3593.1   7186.3     1362
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.7683 -0.7958 -0.0394  0.6268  3.6864
##
## Random effects:
## Groups             Name             Variance Std.Dev.
## bushID:subplotID (Intercept) 0.01157  0.1076
## subplotID         (Intercept) 0.06436  0.2537
## rater              (Intercept) 0.12580  0.3547
## Number of obs: 1380, groups:  bushID:subplotID, 96; subplotID, 16; rater, 6
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.21167    0.16203  13.650 < 2e-16 ***
## compound2    0.07052    0.04637   1.521 0.128276
## compound3    0.06326    0.04613   1.371 0.170251
## compound4   -0.27314    0.05043  -5.416 6.08e-08 ***
## compound5    0.06224    0.04694   1.326 0.184858
## compound6    0.34686    0.04362   7.953 1.82e-15 ***
## compound7   -0.16728    0.04895  -3.418 0.000632 ***
## compound8   -0.24821    0.05044  -4.921 8.62e-07 ***
## compound9    0.04249    0.04661   0.912 0.362024
## compound10  -0.22381    0.04992  -4.484 7.34e-06 ***
## compound11   0.00396    0.04730   0.084 0.933284
## compound12  -0.05510    0.04797  -1.149 0.250690
## compound13   0.04699    0.04654   1.010 0.312695
## compound14   0.29228    0.04431   6.597 4.20e-11 ***
## compound15   0.08186    0.04600   1.780 0.075139 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
```

```
## Correlation matrix not shown by default, as p = 15 > 12.
## Use print(x, correlation=TRUE) or
##     vcov(x)         if you need it
```

```
glmer_coefficients<-as.data.frame(summary(glmer_out6)$coeff) %>%
  rownames_to_column("predictor") %>%
  filter(grepl("compound",predictor)) %>%
  dplyr::rename(pval=`Pr(>|z|)` ) %>%
  #we want to have p-adjusted (Holm) values for one-sided test H.alt: lambda(compound)>lambda(water)
  dplyr::mutate(one_sided_pval=ifelse(`z value`>0, pval/2, (1-pval/2)),
    p_adjusted=p.adjust(one_sided_pval, method="holm"),
    significant_higher=ifelse(p_adjusted<0.05, T, F))

ggplot(glmer_coefficients %>%
  mutate(compound=factor(gsub("compound","",predictor), levels=2:15)),
  aes(x=compound, y=Estimate, color=p_adjusted<0.05))+
  geom_hline(yintercept=0, linetype="dashed")+
  geom_errorbar(aes(ymin=Estimate - `Std. Error`, ymax=Estimate + `Std. Error`), width=0.2)+geom_point(s
  scale_color_manual(values=c("grey", "#85ADAC"))+coord_flip()+
  ylab(expression(beta["compound"])))
```

