Initial Milestone EDA

JingjianGao

2022-10-25

EDA of the data

```
library(dplyr)
```

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

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

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

```
library(ggplot2)
library(rstanarm)
```

```
## Loading required package: Rcpp
```

```
## This is rstanarm version 2.21.3
```

```
## - See https://mc-stan.org/rstanarm/articles/priors for changes to default priors!
```

```
## - Default priors may change, so it's safest to specify priors, even if equivalent to the defaults.
```

```
## - For execution on a local, multicore CPU with excess RAM we recommend calling
```

```
##   options(mc.cores = parallel::detectCores())
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
```

```
## v tibble  3.1.8     v purrr   0.3.4
## v tidyr   1.2.1     v stringr 1.4.1
## v readr   2.1.2     v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(devtools)
```

```
## Loading required package: usethis
```

```
library(ggpubr)
quabbin_data_all <- read.csv("/Users/billg/Desktop/MA 675/Initial-Milestone/binomial_quabbin_data_all.cs

#Data Cleaning, data has values where time.delay = "one " instead of "one", get rid of the blank
```

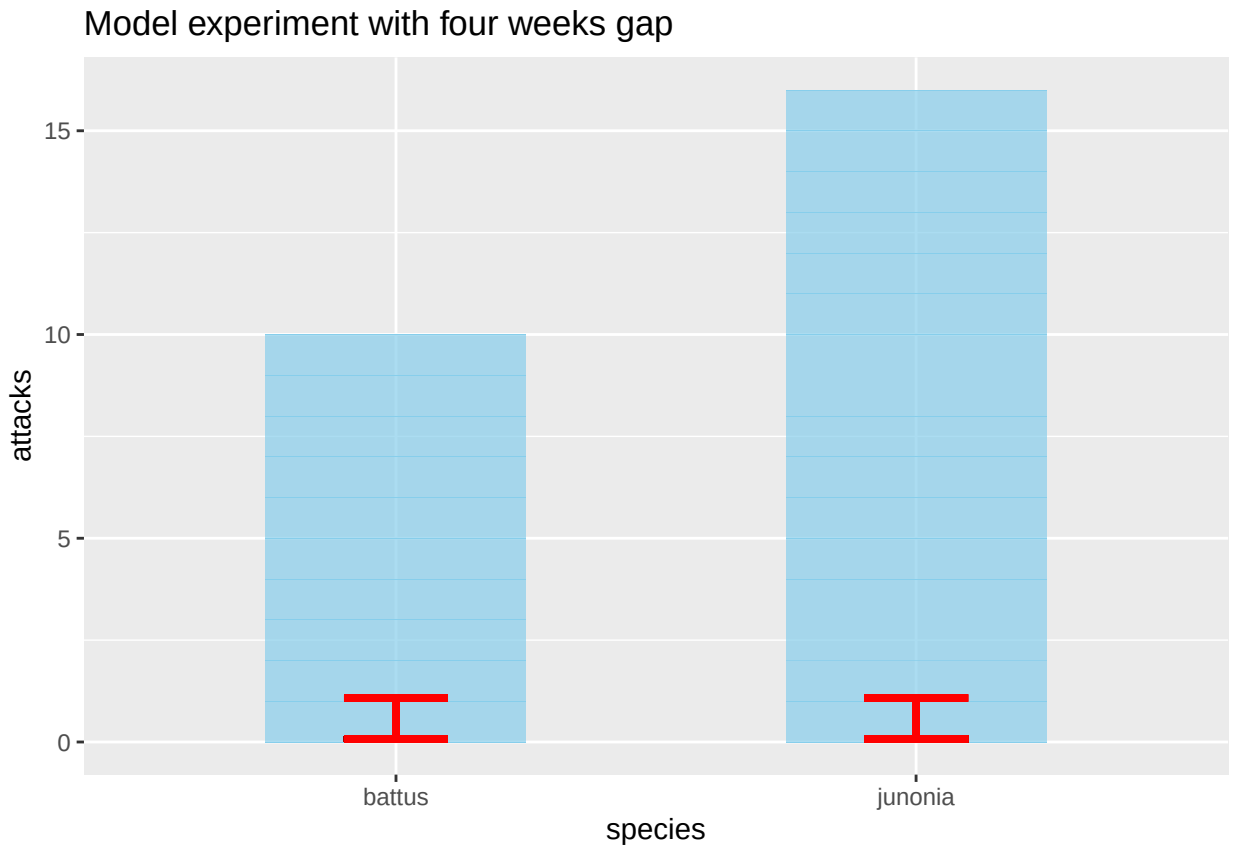
```
one_to_one <- function(string){
  if (string == "one "){
    return(string ="one")
  }
  else {return(string)}
}

quabbin_data_all$time.delay <- sapply(quabbin_data_all$time.delay, one_to_one)

#Separate the data_all file based on time.delay
# Model
quabbin_model <- subset(quabbin_data_all, trial %in% "model")
quabbin_model_four <- subset(quabbin_model, time.delay %in% "four")
quabbin_model_two  <- subset(quabbin_model, time.delay %in% "two")
quabbin_model_one  <- subset(quabbin_model, time.delay %in% "one")

# Mimic
quabbin_mimic <- subset(quabbin_data_all, trial %in% "mimic")
quabbin_mimic_four <- subset(quabbin_mimic, time.delay %in% "four")
quabbin_mimic_two  <- subset(quabbin_mimic, time.delay %in% "two")
quabbin_mimic_one  <- subset(quabbin_mimic, time.delay %in% "one")

# Simultaneous
quabbin_simul <- subset(quabbin_data_all, trial %in% "simultaneously")
```

Bar Plots with error bars

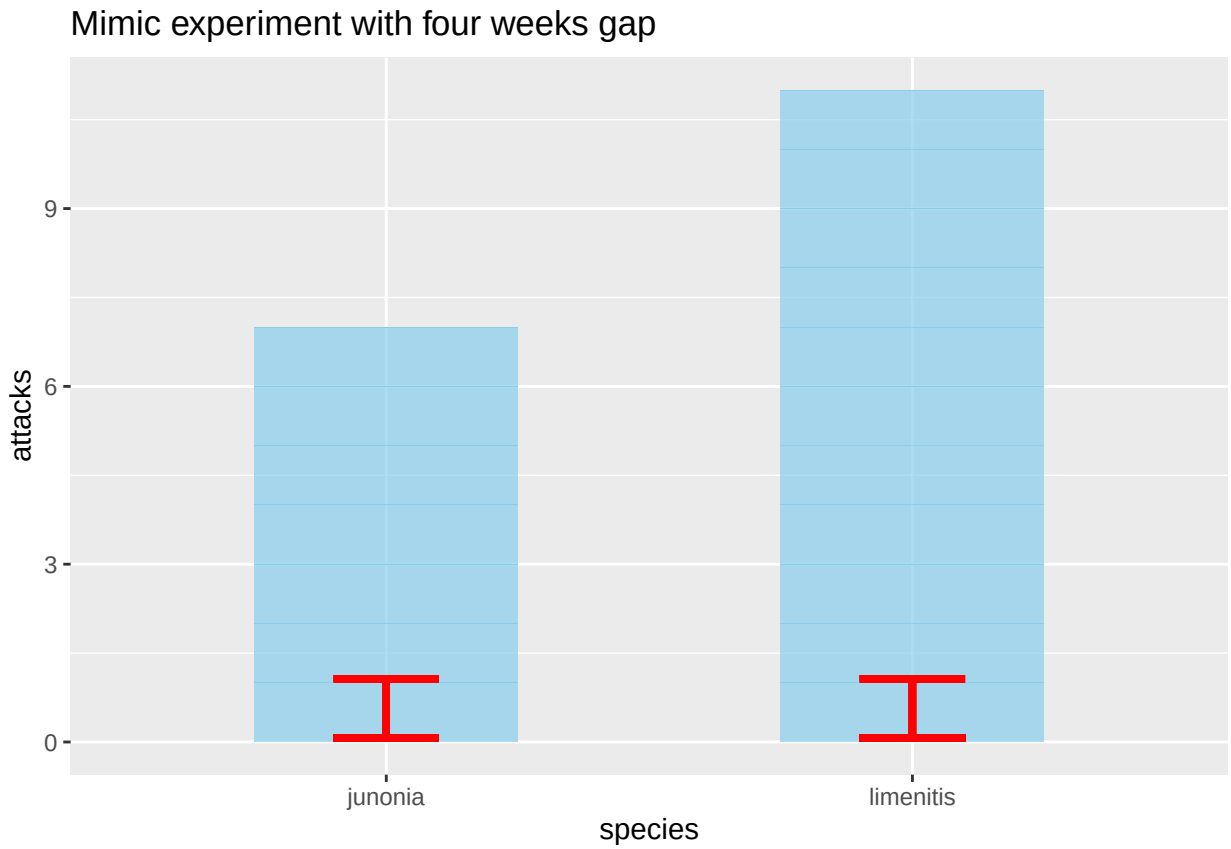
#The length of an Error Bar helps reveal the uncertainty of a data point: a short Error Bar shows that values are concentrated, signalling that the plotted average value is more likely, while a long Error Bar would indicate that the values are more spread out and less reliable.

Species

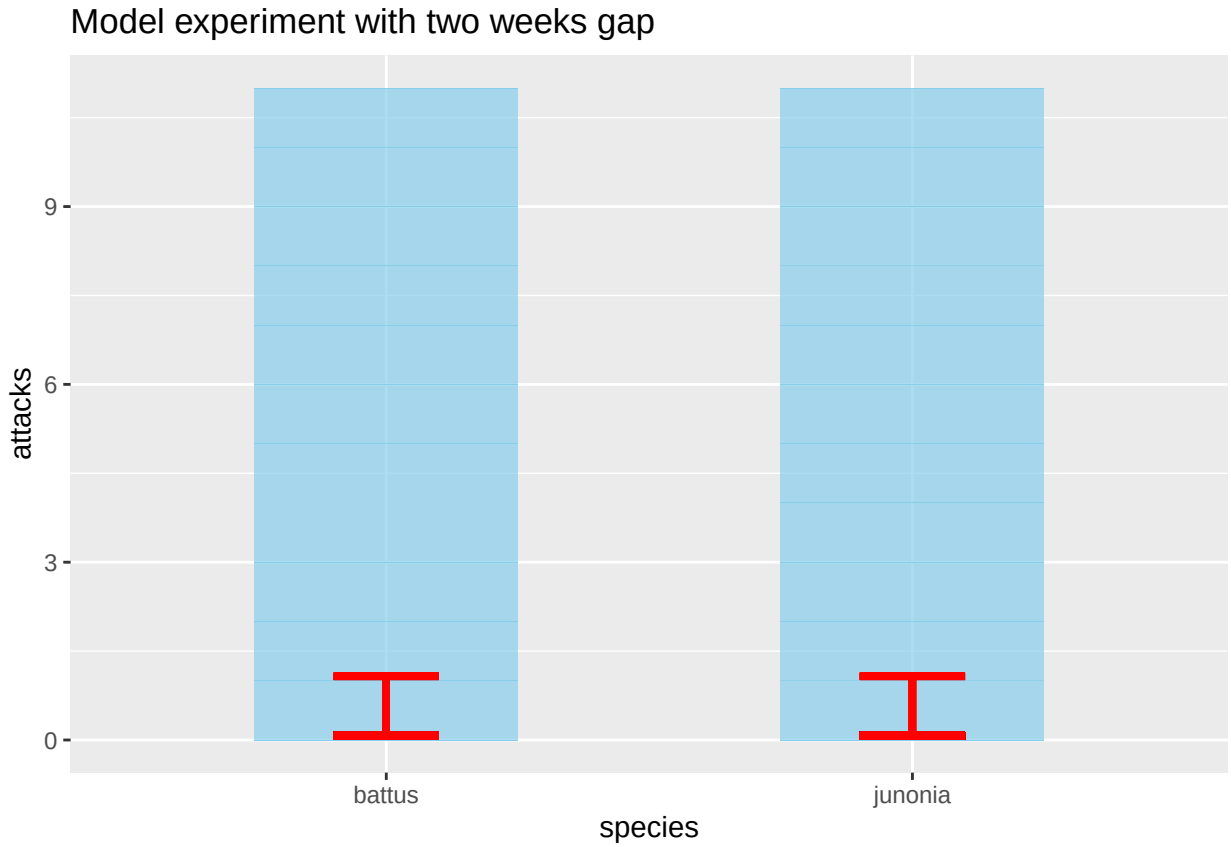
```
# Plot based on species for "Model" with four weeks gap
ggplot_model_species_four <- ggplot(quabbin_model_four) +
  ggtitle("Model experiment with four weeks gap")+
  geom_bar(aes(x=species, y=attacks), stat="identity", fill="skyblue", alpha=0.7,width=0.5) +
  geom_errorbar(aes(x=species, ymin=sd(attacks), ymax=attacks+sd(attacks)), width=0.2,
                colour="red",alpha=0.9, size=1.3)
ggplot_model_species_four
```



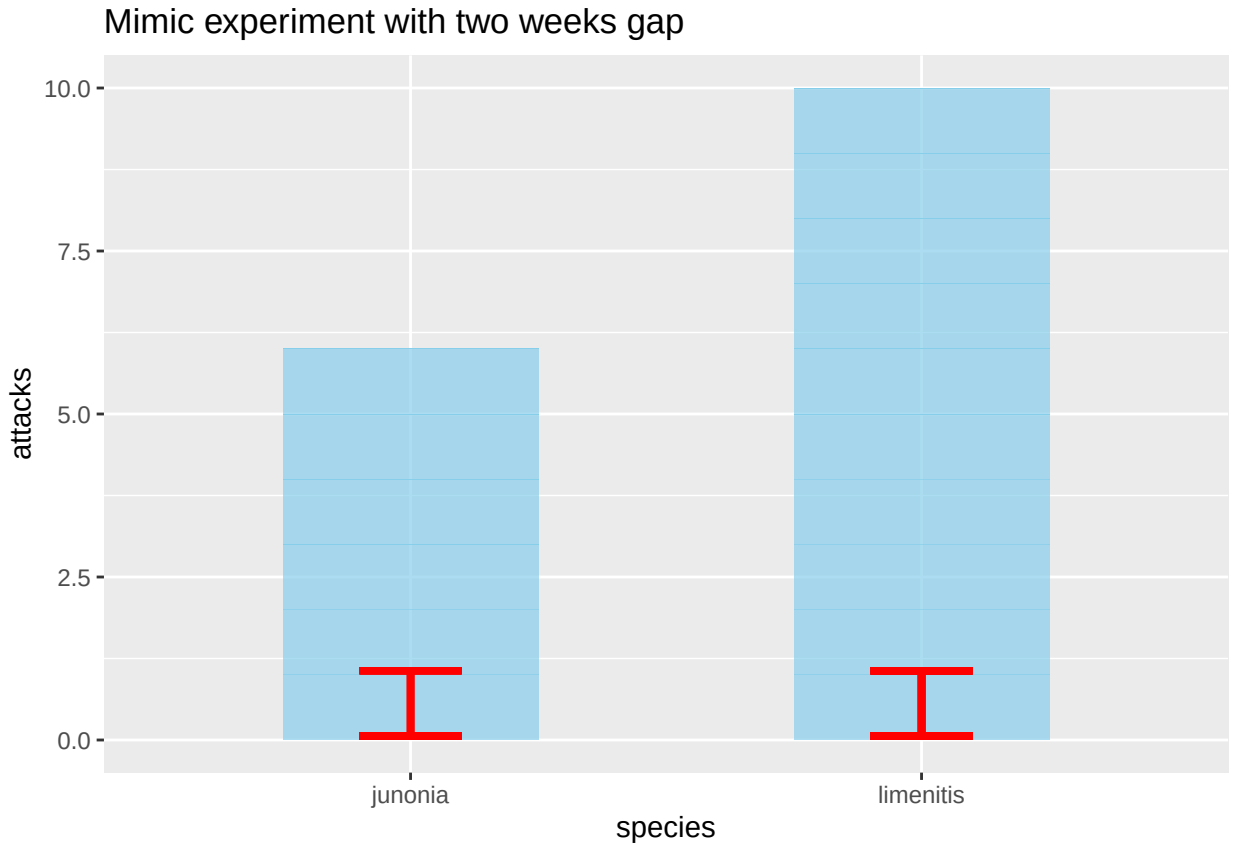

```
#Plot based on species for "Mimic" with four weeks gap
ggplot_mimic_species_four <- ggplot(quabbin_mimic_four) +
  ggtitle("Mimic experiment with four weeks gap")+
  geom_bar(aes(x=species, y=attacks), stat="identity", fill="skyblue", alpha=0.7,width=0.5) +
  geom_errorbar(aes(x=species, ymin=sd(attacks), ymax=attacks+sd(attacks)), width=0.2,
                colour="red",alpha=0.9, size=1.3)
ggplot_mimic_species_four
```



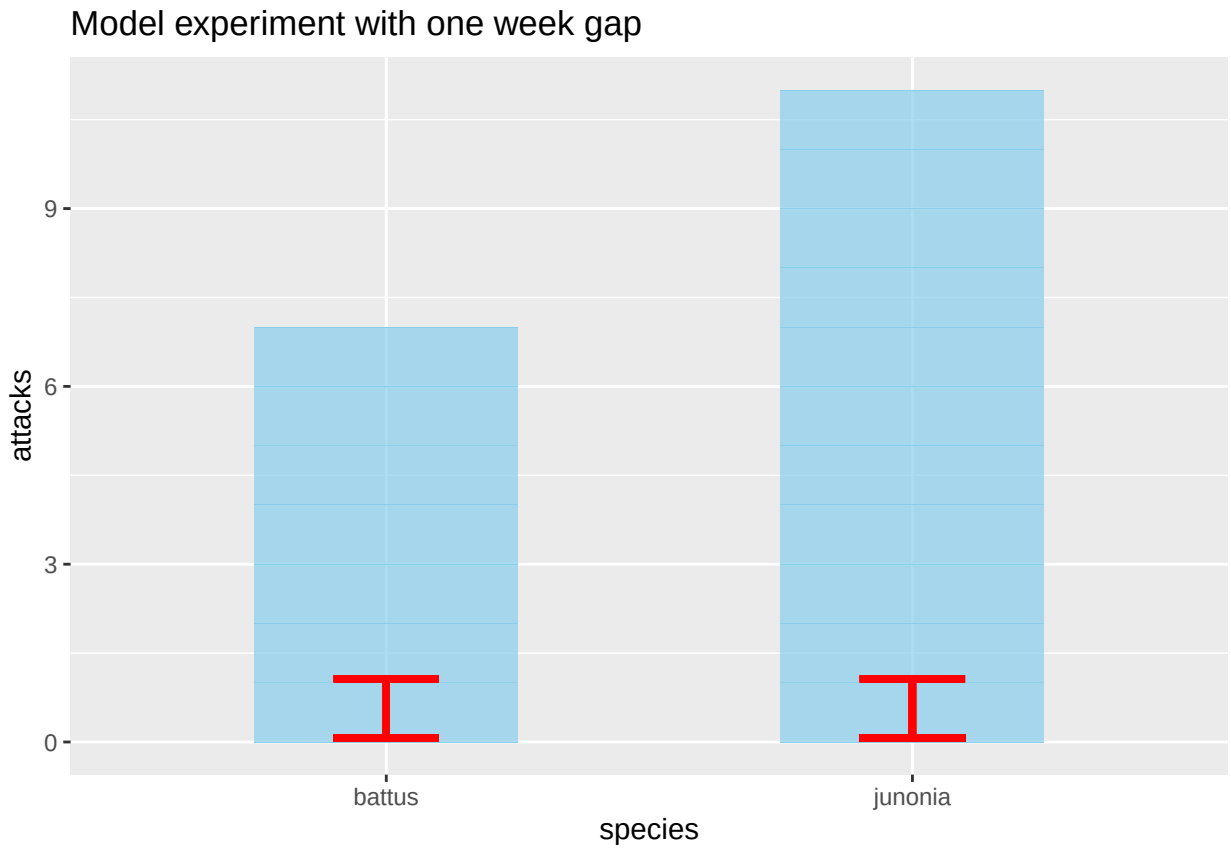

```
# Plot based on species for "Model with two weeks gap
ggplot_model_species_two <- ggplot(quabbin_model_two) +
  ggtitle("Model experiment with two weeks gap")+
  geom_bar(aes(x=species, y=attacks), stat="identity", fill="skyblue", alpha=0.7,width=0.5) +
  geom_errorbar(aes(x=species, ymin=sd(attacks), ymax=attacks+sd(attacks)), width=0.2,
                colour="red",alpha=0.9, size=1.3)
ggplot_model_species_two
```




```
#Plot based on species for "Mimic" with two weeks gap
ggplot_mimic_species_two <- ggplot(quabbin_mimic_two) +
  ggtitle("Mimic experiment with two weeks gap")+
  geom_bar(aes(x=species, y=attacks), stat="identity", fill="skyblue", alpha=0.7,width=0.5) +
  geom_errorbar(aes(x=species, ymin=sd(attacks), ymax=attacks+sd(attacks)), width=0.2,
                colour="red",alpha=0.9, size=1.3)
ggplot_mimic_species_two
```

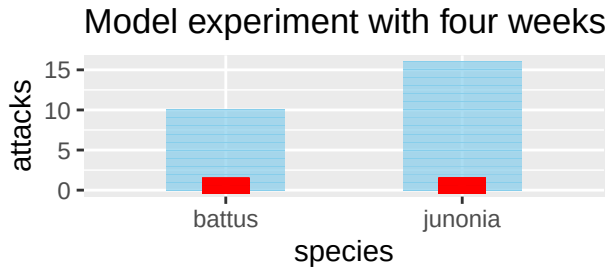


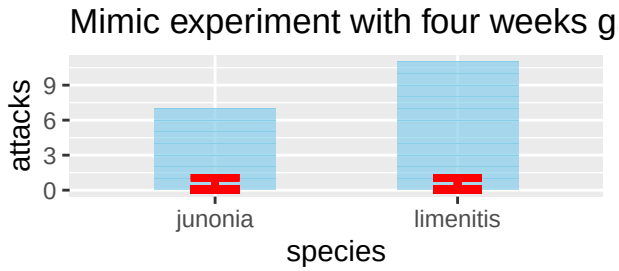

```
# Plot based on species for "Model" with one week gap
ggplot_model_species_one <- ggplot(quabbin_model_one) +
  ggtitle("Model experiment with one week gap")+
  geom_bar(aes(x=species, y=attacks), stat="identity", fill="skyblue", alpha=0.7,width=0.5) +
  geom_errorbar(aes(x=species, ymin=sd(attacks), ymax=attacks+sd(attacks)), width=0.2,
                colour="red",alpha=0.9, size=1.3)
ggplot_model_species_one
```

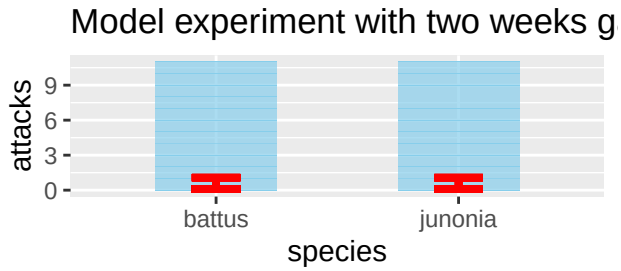


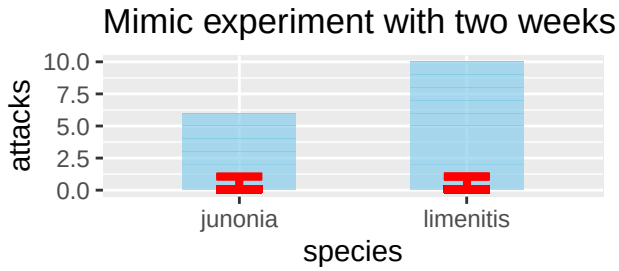

```
#Plot based on species for "Mimic" with one week gap
ggplot_mimic_species_one <- ggplot(quabbin_mimic_one) +
  ggtitle("Mimic experiment with one week gap")+
  geom_bar(aes(x=species, y=attacks), stat="identity", fill="skyblue", alpha=0.7,width=0.5) +
  geom_errorbar(aes(x=species, ymin=sd(attacks), ymax=attacks+sd(attacks)), width=0.2,
                colour="red",alpha=0.9, size=1.3)
ggplot_mimic_species_one
```

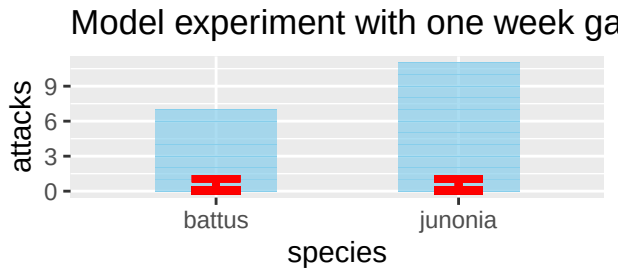

```
ggarrange(ggplot_model_species_four,ggplot_mimic_species_four,ggplot_model_species_two,ggplot_mimic_spe
          ggplot_model_species_one,ggplot_mimic_species_one,
          ncol=2,nrow=3)
```

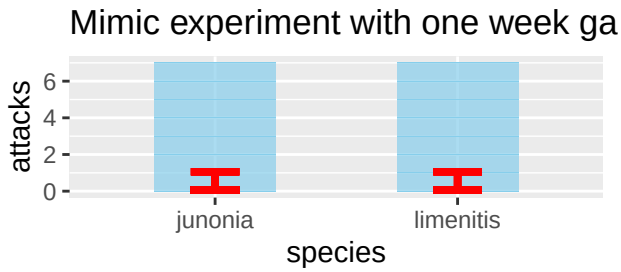






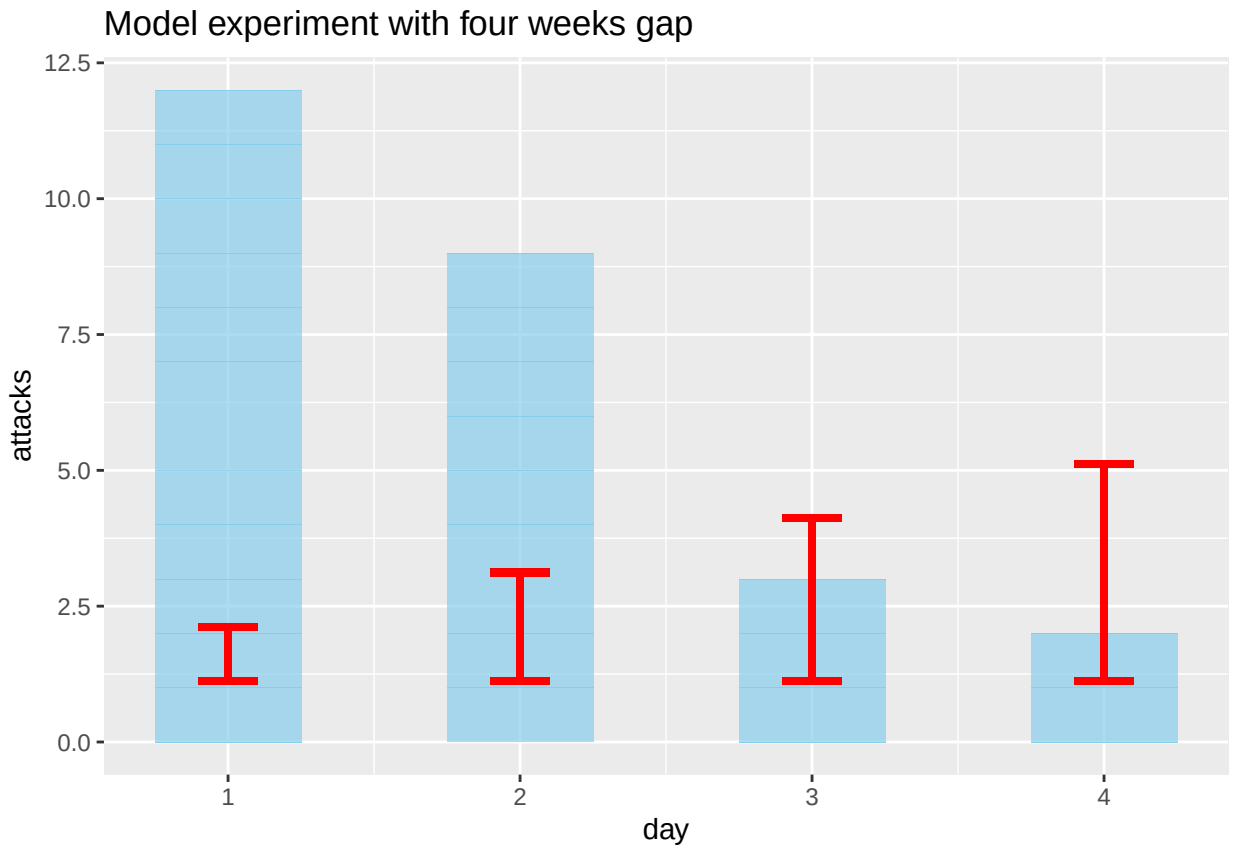




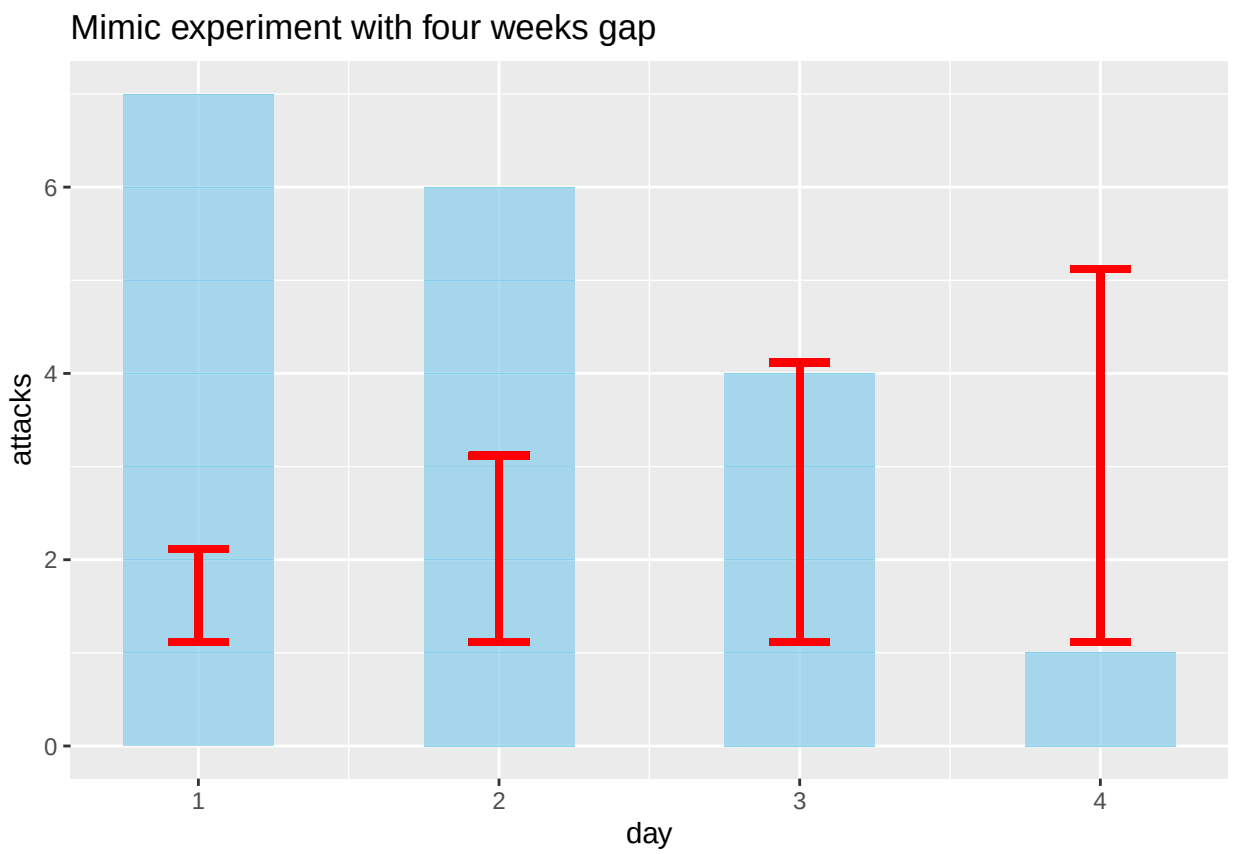


Days

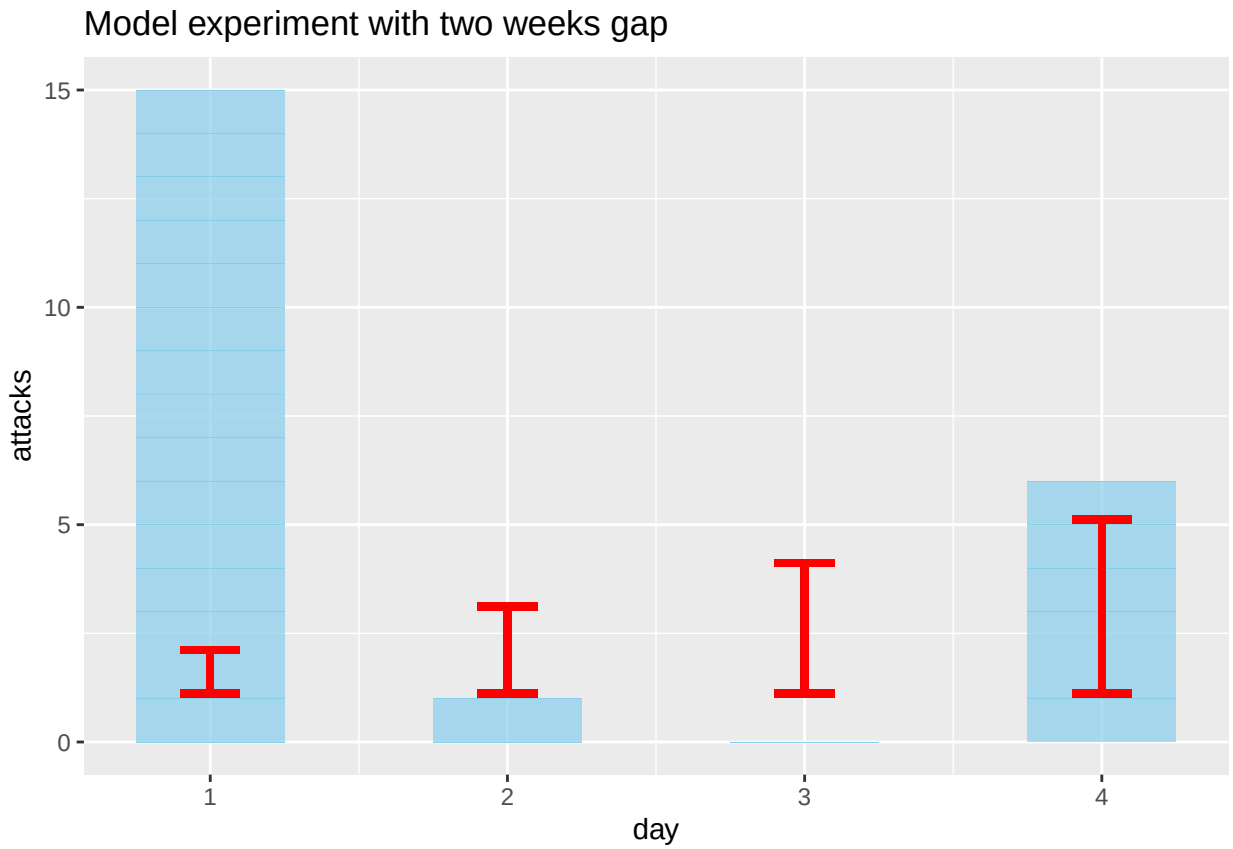
```
# Plot based on days for "Model" with four weeks gap
ggplot_model_day_four <- ggplot(quabbin_model_four) +
  ggtitle("Model experiment with four weeks gap")+
  geom_bar(aes(x=day, y=attacks), stat="identity", fill="skyblue", alpha=0.7,width = 0.5) +
  geom_errorbar(aes(x=day, ymin=sd(day), ymax=day+sd(day)), width=0.2,
                colour="red", alpha=0.9, size=1.3)
ggplot_model_day_four
```



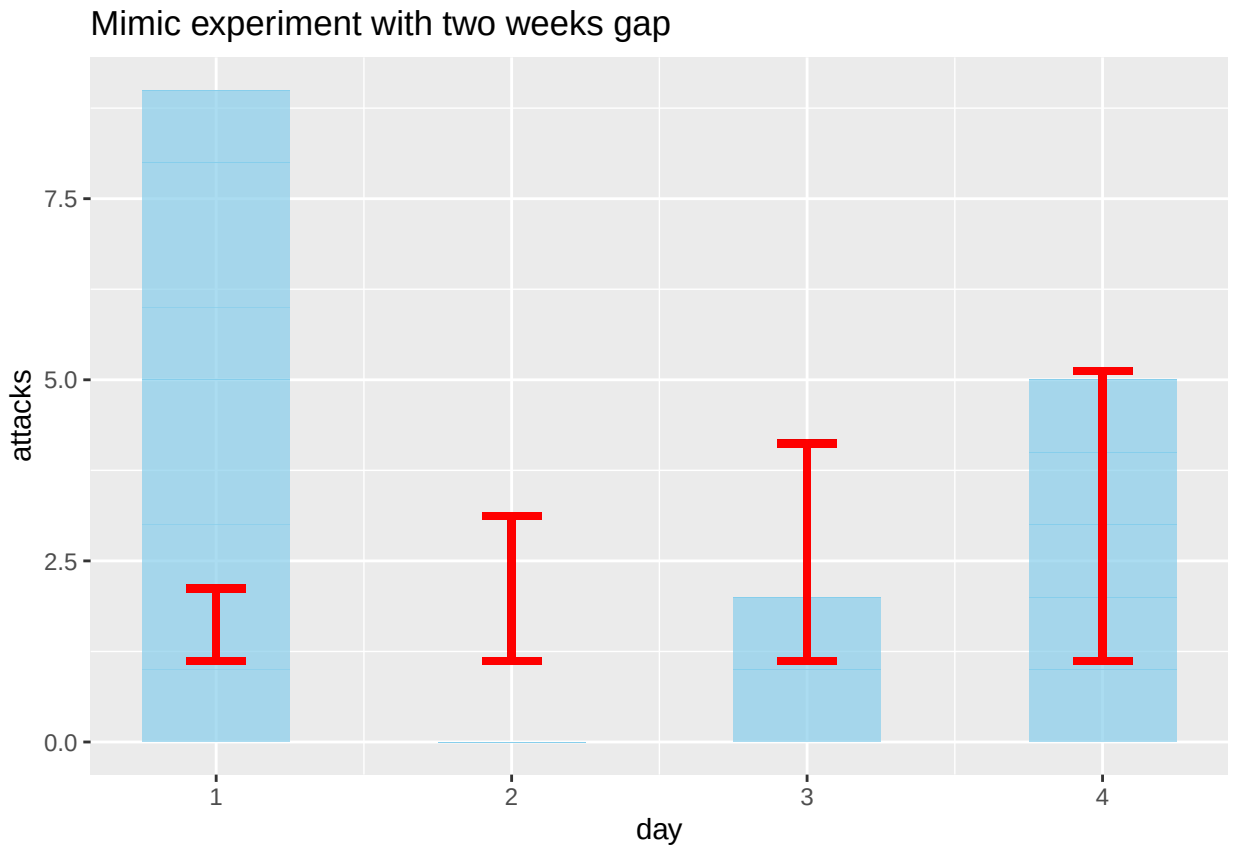

```
# Plot based on days for "Mimic" with four weeks gap
ggplot_mimic_day_four <- ggplot(quabbin_mimic_four) +
  ggtitle("Mimic experiment with four weeks gap")+
  geom_bar(aes(x=day, y=attacks), stat="identity", fill="skyblue", alpha=0.7,width = 0.5) +
  geom_errorbar(aes(x=day, ymin=sd(day), ymax=day+sd(day)), width=0.2,
                colour="red", alpha=0.9, size=1.3)
ggplot_mimic_day_four
```



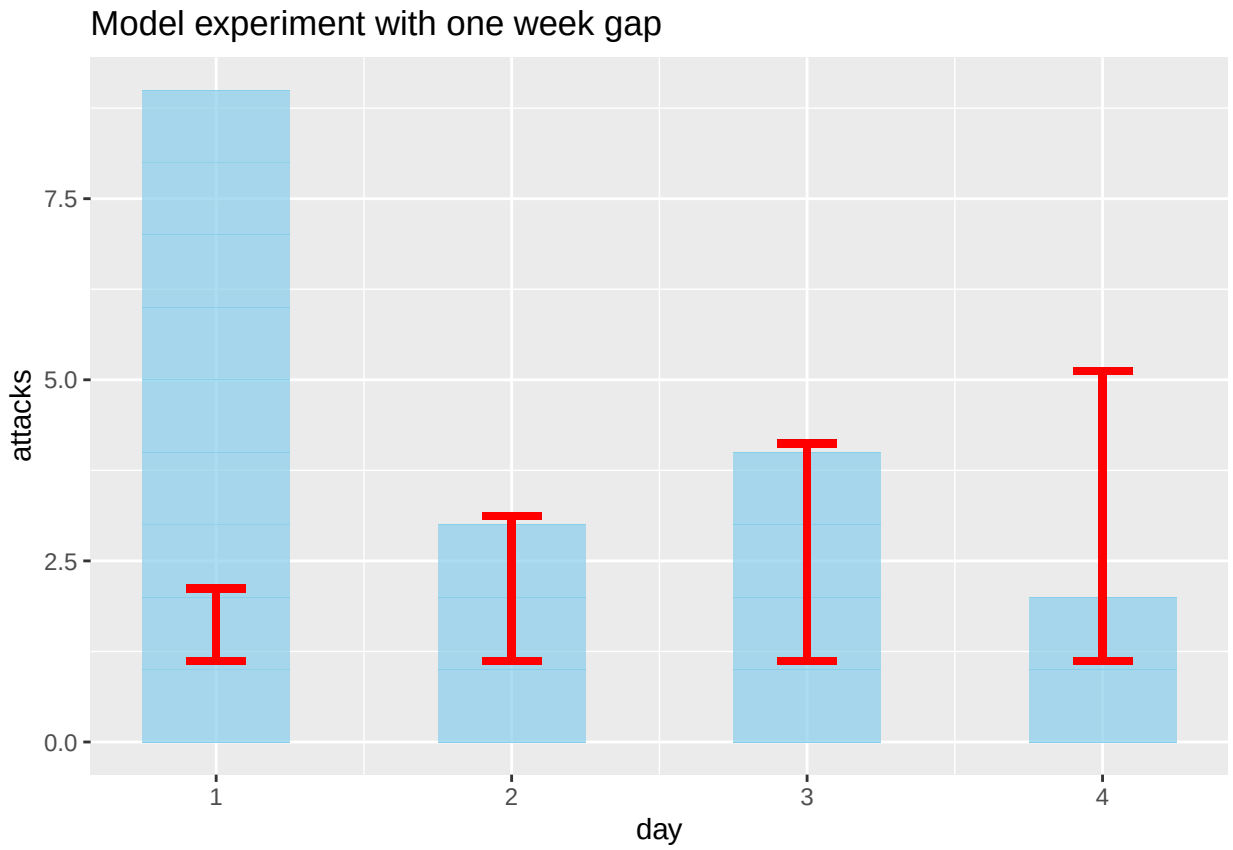

```
# Plot based on days for "Model" with two weeks gap
ggplot_model_day_two <- ggplot(quabbin_model_two) +
  ggtitle("Model experiment with two weeks gap")+
  geom_bar(aes(x=day, y=attacks), stat="identity", fill="skyblue", alpha=0.7,width = 0.5) +
  geom_errorbar(aes(x=day, ymin=sd(day), ymax=day+sd(day)), width=0.2,
                colour="red", alpha=0.9, size=1.3)
ggplot_model_day_two
```



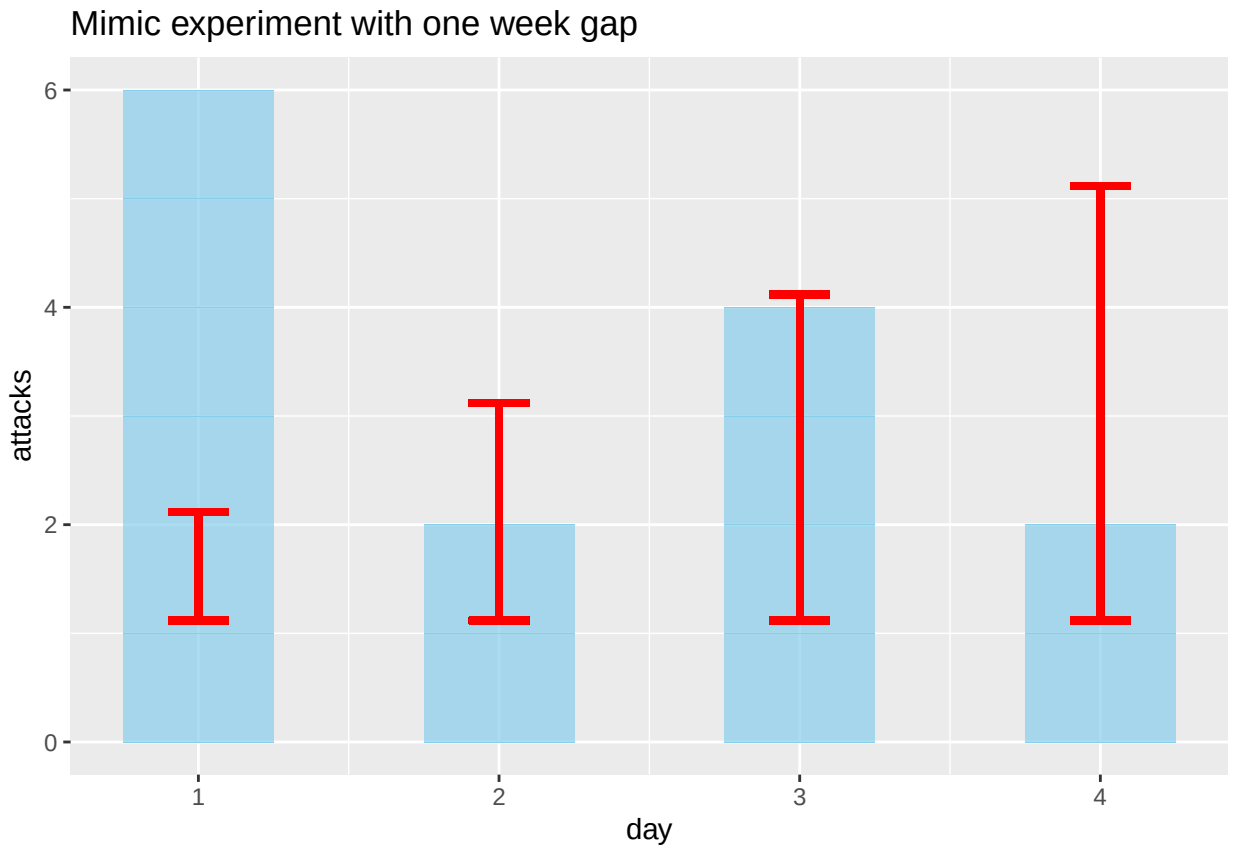

```
# Plot based on days for "Mimic" with two weeks gap
ggplot_mimic_day_two <- ggplot(quabbin_mimic_two) +
  ggtitle("Mimic experiment with two weeks gap")+
  geom_bar(aes(x=day, y=attacks), stat="identity", fill="skyblue", alpha=0.7,width = 0.5) +
  geom_errorbar(aes(x=day, ymin=sd(day), ymax=day+sd(day)), width=0.2,
                colour="red", alpha=0.9, size=1.3)
ggplot_mimic_day_two
```



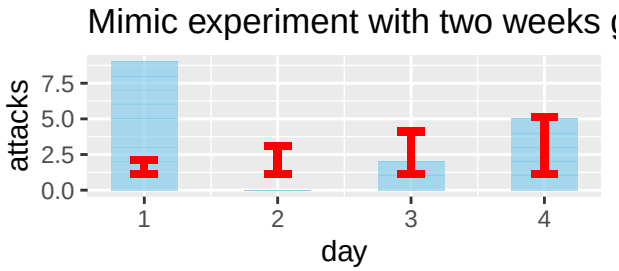

```
# Plot based on days for "Model" with one weeks gap
ggplot_model_day_one <- ggplot(quabbin_model_one) +
  ggtitle("Model experiment with one week gap")+
  geom_bar(aes(x=day, y=attacks), stat="identity", fill="skyblue", alpha=0.7,width = 0.5) +
  geom_errorbar(aes(x=day, ymin=sd(day), ymax=day+sd(day)), width=0.2,
                colour="red", alpha=0.9, size=1.3)
ggplot_model_day_one
```

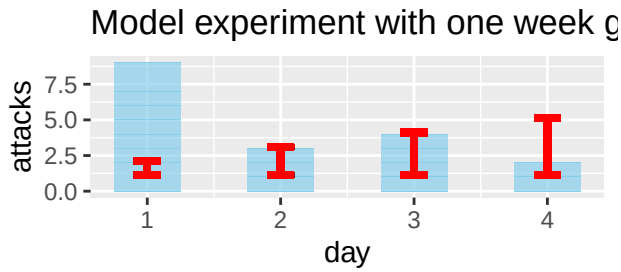



```
# Plot based on days for "Mimic" with one weeks gap
ggplot_mimic_day_one <- ggplot(quabbin_mimic_one) +
  ggtitle("Mimic experiment with one week gap")+
  geom_bar(aes(x=day, y=attacks), stat="identity", fill="skyblue", alpha=0.7,width = 0.5) +
  geom_errorbar(aes(x=day, ymin=sd(day), ymax=day+sd(day)), width=0.2,
                colour="red", alpha=0.9, size=1.3)
ggplot_mimic_day_one
```



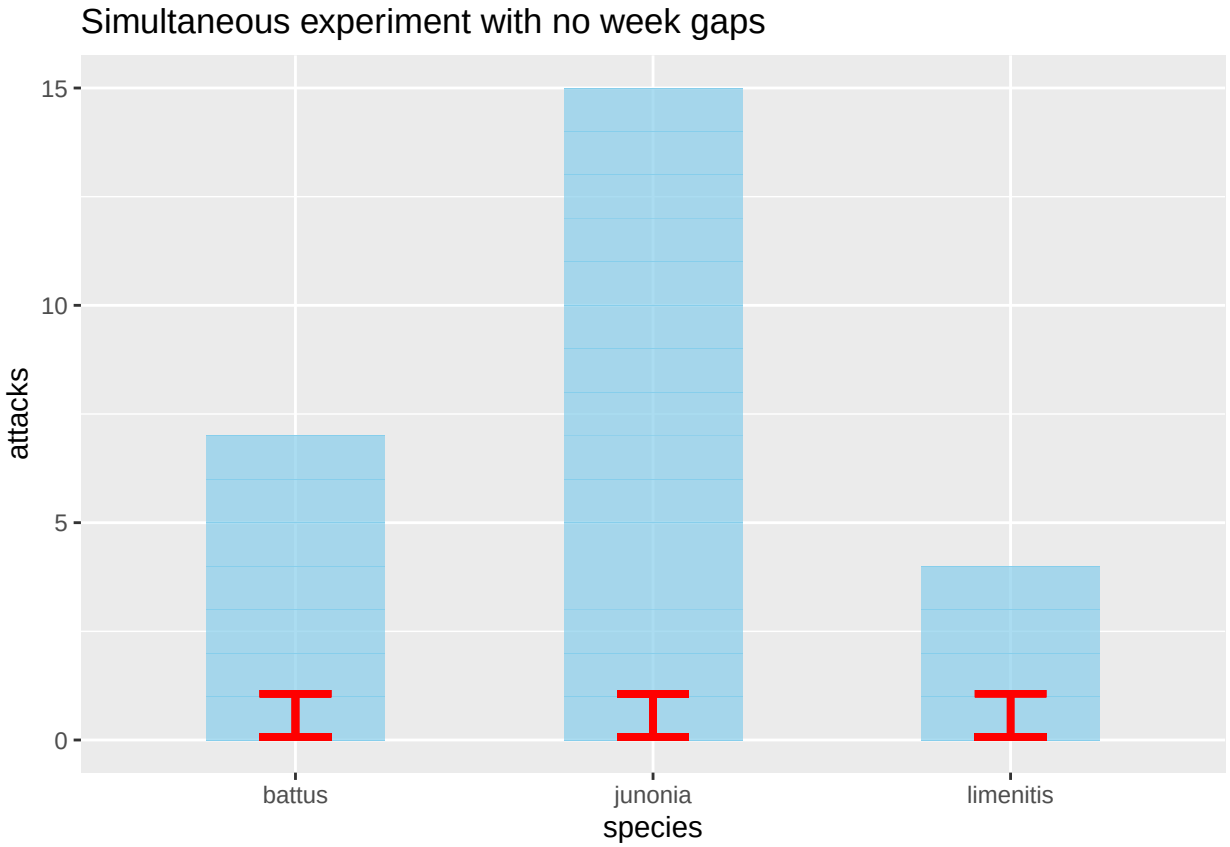

```
ggarrange(ggplot_model_day_four,ggplot_mimic_day_four,ggplot_model_day_two,ggplot_mimic_day_two,
          ggplot_model_day_one,ggplot_mimic_day_one,
          ncol=2,nrow=3)
```





##Simultaneous Graphs

```
ggplot_simul <- ggplot(quabbin_simul) +
  ggtitle("Simultaneous experiment with no week gaps")+
  geom_bar(aes(x=species, y=attacks), stat="identity", fill="skyblue", alpha=0.7,width=0.5) +
  geom_errorbar(aes(x=species, ymin=sd(attacks), ymax=attacks+sd(attacks)), width=0.2,
                colour="red",alpha=0.9, size=1.3)
ggplot_simul
```



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.