

# mid project

2024-11-22

```
require(tidyverse)

## Loading required package: tidyverse

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## vforcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyv     1.3.1
## v purrr    1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

require(deSolve)

## Loading required package: deSolve

Big_data <- read_csv("/Users/andrewbatmunkh/Desktop/eeb313/Big.csv")

## Rows: 30496 Columns: 18
## -- Column specification -----
## Delimiter: ","
## chr (8): state, rMapState, rMapCounty, age, sex, repstat, disease_time_step...
## dbl (10): month, day, year, mass, fa, Pd_intro, years_Pd, county_centroid_lo...
##
## 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.

head(Big_data)

## # A tibble: 6 x 18
##   state rMapState rMapCounty month   day   year age   sex   repstat   mass   fa
##   <chr> <chr>     <chr>      <dbl> <dbl> <dbl> <chr> <chr> <chr>   <dbl> <dbl>
## 1 GA    georgia   chattooga     5    29  2013 adult male  testes-~  17    48
## 2 GA    georgia   carroll      5    30  2013 adult male  non-rep-~  15    47
## 3 GA    georgia   gordon       5    20  2013 adult female pregnant  25    49
## 4 GA    georgia   gordon       5    20  2013 adult female pregnant  23.8   46
## 5 GA    georgia   gordon       5    21  2013 adult female pregnant  26.5   48
## 6 GA    georgia   gordon       5    20  2013 adult female pregnant  20.5   44
## # i 7 more variables: Pd_intro <dbl>, years_Pd <dbl>, disease_time_step <chr>,
## #   county_centroid_lon <dbl>, county_centroid_lat <dbl>, site_mask <chr>,
## #   fips <dbl>
```

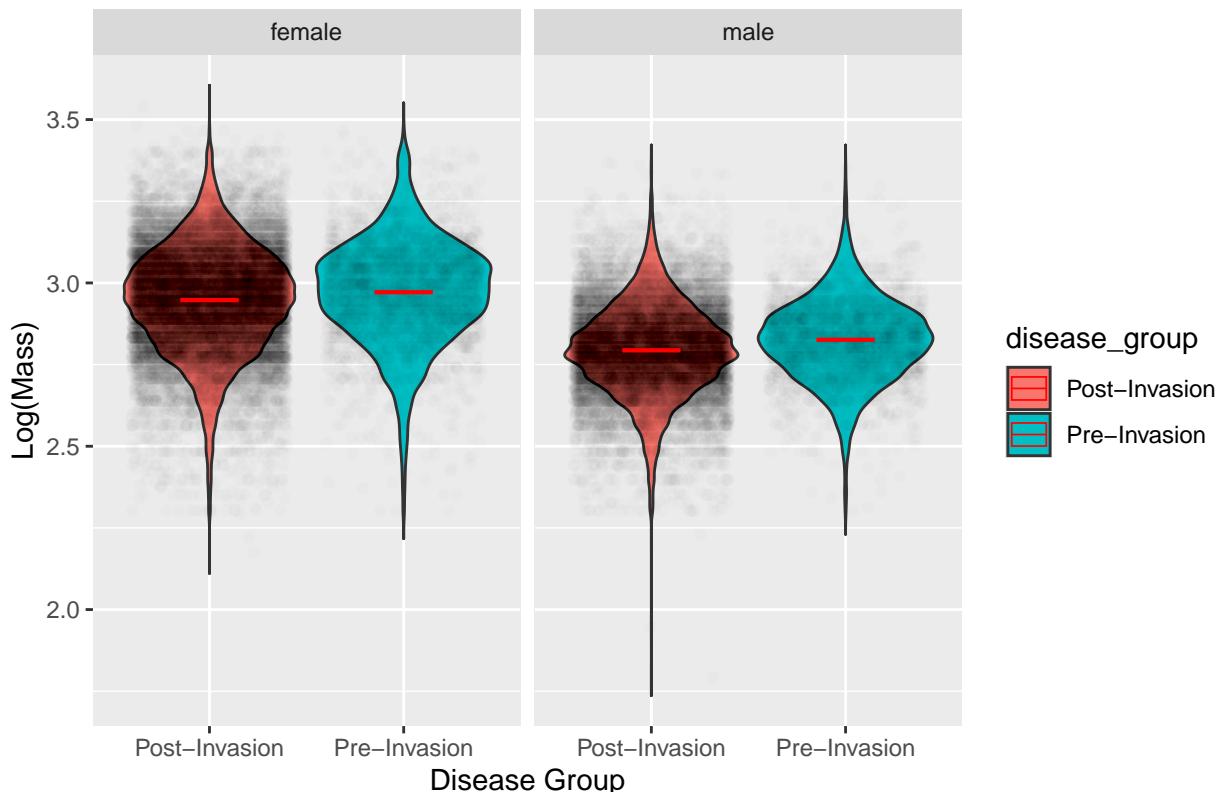
### Violin Plot of Log(Mass) by Disease Group and Sex

```
data_sex <- Big_data %>%
  mutate(disease_group = ifelse(disease_time_step == "pre-invasion",
                                "Pre-Invasion",
                                "Post-Invasion")) %>%
  mutate(log_mass = log(mass))

ggplot(data_sex, aes(x = disease_group, y = log_mass, fill = disease_group)) +
  geom_violin(trim = FALSE) +
  geom_point(alpha = 0.01, position = "jitter") +
  stat_summary(fun = mean, geom = "crossbar",
               width = 0.3, color = "red", size = 0.3) +
  facet_wrap(~sex) +
  labs(
    title = "Violin Plot of Log(Mass) by Disease Group and Sex",
    x = "Disease Group",
    y = "Log(Mass)"
  )

## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

### Violin Plot of Log(Mass) by Disease Group and Sex



```

glm_model <- glm(log_mass ~ disease_group * sex, data = data_sex)

summary(glm_model)

## 
## Call:
## glm(formula = log_mass ~ disease_group * sex, data = data_sex)
## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                2.947483  0.001262 2334.968 < 2e-16 ***
## disease_groupPre-Invasion 0.024042  0.003509   6.852 7.44e-12 ***
## sexmale                   -0.153842  0.001920  -80.135 < 2e-16 ***
## disease_groupPre-Invasion:sexmale 0.008289  0.005156   1.608   0.108  
## ---                        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for gaussian family taken to be 0.02380785)
## 
## Null deviance: 902.67 on 30495 degrees of freedom
## Residual deviance: 725.95 on 30492 degrees of freedom
## AIC: -27436
## 
## Number of Fisher Scoring iterations: 2

```

### Violin Plot of Log(Mass) by Disease Group and Age

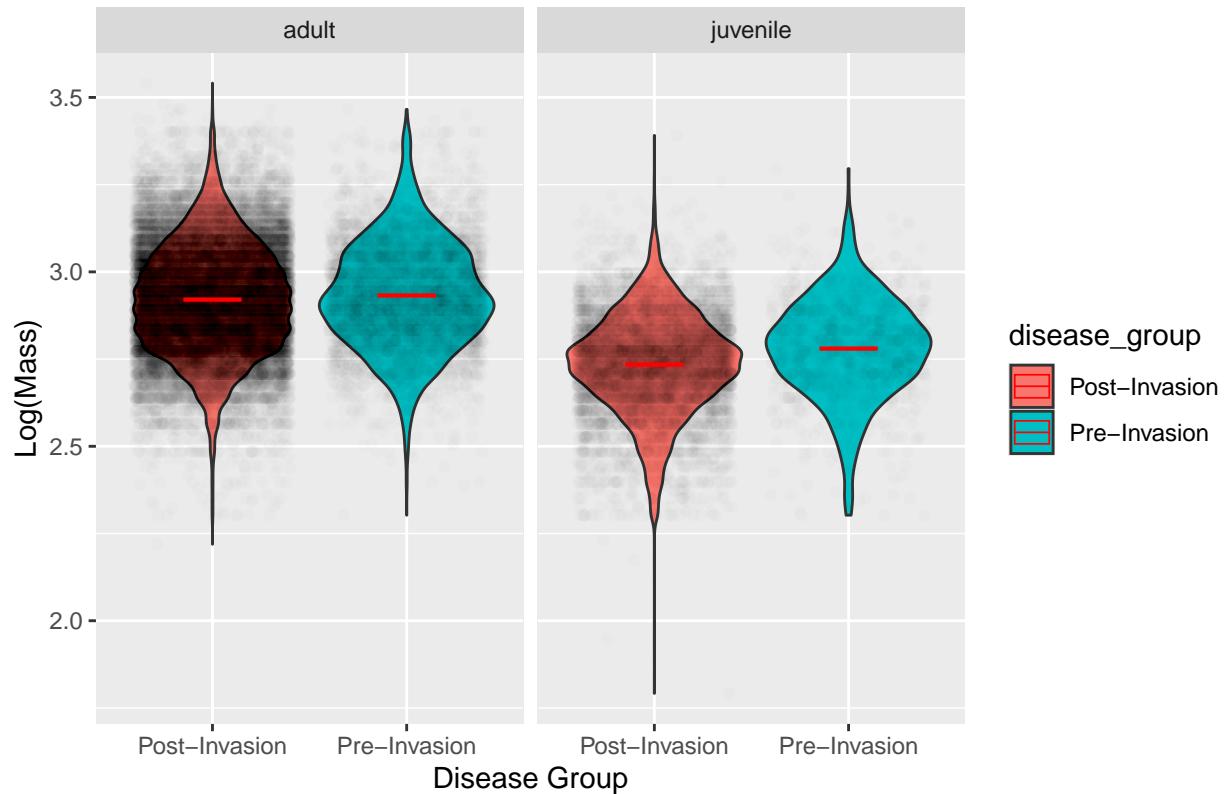
```

data_age <- Big_data %>%
  mutate(disease_group = ifelse(disease_time_step == "pre-invasion",
                                "Pre-Invasion",
                                "Post-Invasion")) %>%
  mutate(log_mass = log(mass))

ggplot(data_age, aes(x = disease_group, y = log_mass, fill = disease_group)) +
  geom_violin(alpha = 1) +
  geom_point(alpha = 0.01, position = "jitter") +
  stat_summary(fun = mean, geom = "crossbar",
              width = 0.3, color = "red", size = 0.3) +
  facet_wrap(~age) +
  labs(
    title = "Violin Plot of Log(Mass) by Disease Group and Age",
    x = "Disease Group",
    y = "Log(Mass)"
  )

```

## Violin Plot of Log(Mass) by Disease Group and Age



```
glm_age <- glm(log_mass ~ disease_group * age, data = data_age)

summary(glm_age)
```

```
##
## Call:
## glm(formula = log_mass ~ disease_group * age, data = data_age)
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                2.920468   0.001076 2712.948 < 2e-16
## disease_groupPre-Invasion  0.011866   0.002875   4.127 3.69e-05
## agejuvenile              -0.186521   0.002339  -79.735 < 2e-16
## disease_groupPre-Invasion:agejuvenile  0.034202   0.006550    5.222 1.78e-07
##
## (Intercept)                 ***
## disease_groupPre-Invasion ***
## agejuvenile                 ***
## disease_groupPre-Invasion:agejuvenile ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.02404237)
##
## Null deviance: 902.67  on 30495  degrees of freedom
```

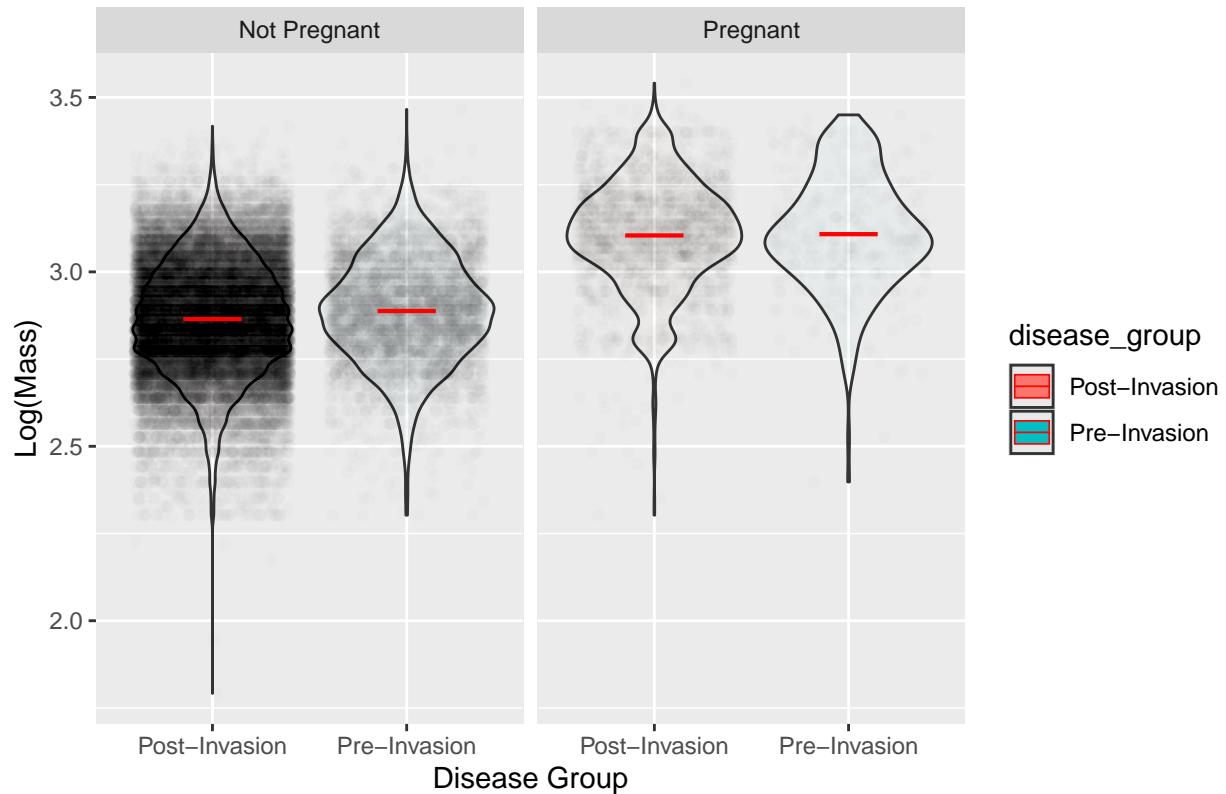
```
## Residual deviance: 733.10  on 30492  degrees of freedom
## AIC: -27137
##
## Number of Fisher Scoring iterations: 2
```

### Violin Plot of Log(Mass) by Disease Group and Pregnancy Status

```
data_pregnant <- Big_data %>%
  mutate(pregnancy_status = ifelse(repstat == "pregnant",
                                    "Pregnant",
                                    "Not Pregnant")) %>%
  mutate(disease_group = ifelse(disease_time_step == "pre-invasion",
                                 "Pre-Invasion",
                                 "Post-Invasion")) %>%
  mutate(log_mass = log(mass))

ggplot(data_pregnant, aes(x = disease_group, y = log_mass, fill = disease_group)) +
  geom_violin(alpha = 0.02) +
  geom_point(alpha = 0.01, position = "jitter") +
  stat_summary(fun = mean, geom = "crossbar",
              width = 0.3, color = "red", size = 0.3) +
  facet_wrap(~pregnancy_status) +
  labs(
    title = "Violin Plot of Log(Mass) by Disease Group and Pregnancy Status",
    x = "Disease Group",
    y = "Log(Mass)"
  )
```

## Violin Plot of Log(Mass) by Disease Group and Pregnancy Status



```
glm_pregnancy <- glm(log_mass ~ disease_group * pregnancy_status, data = data_pregnant)
summary(glm_pregnancy)
```

```
##
## Call:
## glm(formula = log_mass ~ disease_group * pregnancy_status, data = data_pregnant)
##
## Coefficients:
##                               Estimate Std. Error
## (Intercept)                2.864696  0.001029
## disease_groupPre-Invasion  0.022977  0.002785
## pregnancy_statusPregnant  0.239677  0.003948
## disease_groupPre-Invasion:pregnancy_statusPregnant -0.019309  0.010460
##                                         t value Pr(>|t|)
## (Intercept)                2784.622 <2e-16 ***
## disease_groupPre-Invasion   8.251  <2e-16 ***
## pregnancy_statusPregnant   60.705 <2e-16 ***
## disease_groupPre-Invasion:pregnancy_statusPregnant -1.846   0.0649 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.02596528)
##
## Null deviance: 902.67 on 30495 degrees of freedom
```

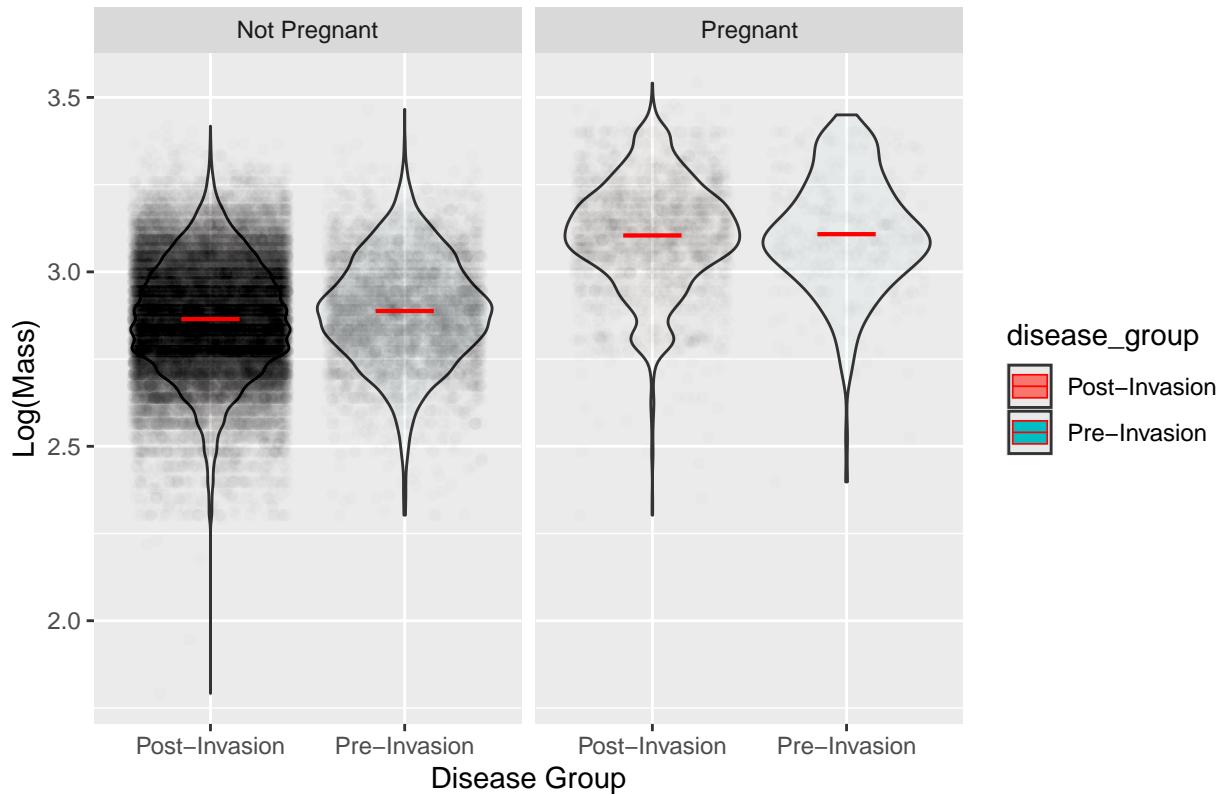
```
## Residual deviance: 791.73  on 30492  degrees of freedom
## AIC: -24791
##
## Number of Fisher Scoring iterations: 2
```

### Violin Plot of Log(Mass) by Disease Group and Pregnancy Status with Mean Lines

```
data_pregnant <- Big_data %>%
  mutate(pregnancy_status = ifelse(repstat == "pregnant",
                                    "Pregnant",
                                    "Not Pregnant")) %>%
  mutate(disease_group = ifelse(disease_time_step == "pre-invasion",
                                "Pre-Invasion",
                                "Post-Invasion")) %>%
  mutate(log_mass = log(mass))

ggplot(data_pregnant, aes(x = disease_group, y = log_mass, fill = disease_group)) +
  geom_violin(alpha = 0.02) +
  geom_point(alpha = 0.01, position = "jitter") +
  stat_summary(fun = mean, geom = "crossbar",
              width = 0.3, color = "red", size = 0.3) +
  facet_wrap(~pregnancy_status) +
  labs(
    title = "Violin Plot of Log(Mass) by Disease Group and Pregnancy Status with Mean Lines",
    x = "Disease Group",
    y = "Log(Mass)"
  )
```

## Violin Plot of Log(Mass) by Disease Group and Pregnancy Status with Mean



```
glm_pregnancy_disease <- glm(log_mass ~ disease_group * pregnancy_status, data = data_pregnant)
summary(glm_pregnancy_disease)
```

```
##
## Call:
## glm(formula = log_mass ~ disease_group * pregnancy_status, data = data_pregnant)
##
## Coefficients:
##                               Estimate Std. Error
## (Intercept)                2.864696  0.001029
## disease_groupPre-Invasion  0.022977  0.002785
## pregnancy_statusPregnant   0.239677  0.003948
## disease_groupPre-Invasion:pregnancy_statusPregnant -0.019309  0.010460
##                                         t value Pr(>|t|)
## (Intercept)                2784.622 <2e-16 ***
## disease_groupPre-Invasion    8.251 <2e-16 ***
## pregnancy_statusPregnant     60.705 <2e-16 ***
## disease_groupPre-Invasion:pregnancy_statusPregnant   -1.846  0.0649 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.02596528)
##
## Null deviance: 902.67 on 30495 degrees of freedom
```

```
## Residual deviance: 791.73  on 30492  degrees of freedom
## AIC: -24791
##
## Number of Fisher Scoring iterations: 2
```