Individual project pre analysis

Claire Campion

5/11/2022

1. Overview

Bumble bee populations are increasingly under threat from habitat fragmentation, pesticides, pathogens, and climate change. Climate is likely a prime driver of declines in abundance and distribution, as bumble bees are limited by their thermal tolerance (Kerr et al. 2015). However, the tolerance of whole organisms often exceeds that of their gametes; for example, insects can be sterilized by temperatures below their upper thermal tolerances (Heerwaarden and Sgrò 2021; David et al. 2005). In bumble bees, males are independent from the colony and can withstand extreme temperatures, but whether these temperatures compromise spermatozoa is unclear. Using commercially-reared bumble bee (*Bombus impatiens*) males, we measured how exposure to sublethal temperatures near male critical thermal minimum and maximum impact spermatozoa viability. We measured temperature effects on spermatozoa in intact males to determine if males are potentially protecting spermatozoa. A LIVE/DEAD™ Cell Imaging Kit (Invitrogen™) and cell counter (Nexcelom Cellometer Spectrum Image Cytometry System) were used to estimate sperm count and viability in males exposed to 45°C for 85 minutes and 4°C for 85 min and 48 hours. Cells were stained with LIVE/DEAD kit immediately after temperature exposures and were compared to control males which were held at room temperature (22°C) throughout experiment.The purpose of this study is to determine how exposure time may influence temperature effects on viability. This gap is important to address as bumble bee populations continue to decline and potential sterility to males could be a contributing factor.

My goal for this project is to be able to create box plots and violin graphs to visualize these results and begin to understand the effects of temperature on the reproductive fitness in bumble bee males. I hope to begin to get these figures in a place where that can be published!

2. Dataset

The dataset I will be using for this experiment include cell counts of male spermatozoa. The data consists of live cells, dead cells, total cells, and percent viability(live/total). Other information included in the datasets are treatment, date, colony, species, date, mass, beeid, image number, and sex. I get my data from a automated microscope and I take 3 images of each sample and average them to get a accurate calculation of the cell count. I have 2 datasets, one for cold and one for heat treatment as explained above.

I anticipate there not being many issues with this data but I want to explore some addition aspects I could add to my boxplot such as changing the y axis to log scale or creating a histogram to see if differences are due to mass or other variables.

```
library(tidyverse, message== FALSE)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(readxl, message== FALSE)
library(ggplot2, message== FALSE)
library(patchwork, message== FALSE)
```

```
chillvia <- read_xlsx("chillcoma exp.xlsx")
str(chillvia)
```

```
## tibble [93 x 13] (S3: tbl_df/tbl/data.frame)
##  $ beeid       : num [1:93] 1 1 1 2 2 2 3 3 3 4 ...
##  $ image       : num [1:93] 1 2 3 1 2 3 1 2 3 1 ...
##  $ mass        : num [1:93] 0.118 0.118 0.118 0.0989 0.0989 ...
##  $ treat       : chr [1:93] "Control" "Control" "Control" "Chillcoma" ...
##  $ time        : num [1:93] 0 0 0 48 48 48 48 48 48 48 ...
##  $ live        : num [1:93] 1236 882 624 528 381 ...
##  $ dead        : num [1:93] 138 158 169 130 118 123 131 103 135 239 ...
##  $ total       : num [1:93] 1374 1040 1093 658 499 ...
##  $ pc_viability: num [1:93] 89.7 84.5 84.3 80.1 76.1 77.7 75.2 80.8 81.4 72.1 ...
##  $ date        : POSIXct[1:93], format: "2022-04-20" "2022-04-20" ...
##  $ colony      : chr [1:93] "simon" "simon" "simon" "simon" ...
##  $ species     : chr [1:93] "impatiens" "impatiens" "impatiens" "impatiens" ...
##  $ sex         : chr [1:93] "male" "male" "male" "male" ...
```

```
heatvia <- read_xlsx("heat exp.xlsx")
str(heatvia)
```

```
## tibble [117 x 14] (S3: tbl_df/tbl/data.frame)
##  $ cellometer  : logi [1:117] NA NA NA NA NA NA ...
##  $ beeid       : num [1:117] 1 1 1 2 2 2 3 3 3 4 ...
##  $ image       : num [1:117] 1 2 3 1 2 3 1 2 3 1 ...
##  $ mass        : num [1:117] 0.161 0.161 0.161 0.167 0.167 ...
##  $ treat       : chr [1:117] "control" "control" "control" "hs" ...
##  $ time        : num [1:117] 0 0 0 29 29 29 29 29 29 0 ...
##  $ live        : num [1:117] 127 142 141 540 514 438 438 374 355 272 ...
##  $ dead        : num [1:117] 30 40 33 157 161 172 112 84 76 67 ...
##  $ total       : num [1:117] 157 162 174 697 675 610 550 458 431 339 ...
##  $ pc_viability: num [1:117] 80.9 78 81.1 77.4 76 71.8 79.7 81.8 82.4 80.4 ...
##  $ date        : POSIXct[1:117], format: "2021-11-11" "2021-11-11" ...
##  $ colony      : chr [1:117] "aspen" "aspen" "aspen" "willow" ...
##  $ species     : chr [1:117] "impatiens" "impatiens" "impatiens" "impatiens" ...
##  $ sex         : chr [1:117] "male" "male" "male" "male" ...
```

3. Tidying Data

In this section, my goal is to tidy data, make sure all columns are correct data type, removing any potential NAs, calculating the mean live, dead, and total counts for each unique beeID, and grouping by bee ID.

```
#For the chill dataset, I using the mutate function to assign each column is
#the correct data type.  I want my treatment as a factor and the rest as
#numeric value.
chillvia <- chillvia %>%
  mutate(treat=factor(treat),
         time=as.numeric(time),
         live=as.numeric(live),
         dead=as.numeric(dead),
         total=as.numeric(total),
         pc_viability=as.numeric(pc_viability))

#Next I am grouping the dataset by BeeID and using summarize to calculate the
#means for viability, live, dead, total cell counts, mass, and time. I also am
#taking the unique value of each beeIDs treatment and colony.
cvia1 <- chillvia %>%
  group_by(beeid) %>%
  summarize(via = mean(pc_viability, na.rm=TRUE),
            live = mean(live, na.rm=TRUE),
            dead = mean(dead, na.rm=TRUE),
            total= live+dead,
            treat = unique(treat),
            time = mean(time, na.rm=T),
            mass = mean(mass, na.rm=TRUE),
            colony = unique(colony))
cvia1
```

```
## # A tibble: 31 x 9
##    beeid   via  live  dead total treat      time   mass colony
##    <dbl> <dbl> <dbl> <dbl> <dbl> <fct>     <dbl>  <dbl> <chr>
##  1     1  86.2  914  155   1069  Control       0 0.118  simon
##  2     2  78.0  448. 124.   572  Chillcoma    48 0.0989 simon
##  3     3  79.1  480. 123    603. Chillcoma    48 0.123  fleetwood
##  4     4  77.4  821  432   1253  Chillcoma    48 0.154  fleetwood
##  5     5  82.0  882. 193.  1075  Chillcoma    48 0.181  mac
##  6     6  74.9  321  101    422  Control       0 0.119  mac
##  7     8  94.6 1559.  87   1646. Chillcoma    48 0.0805 fleetwood
##  8     9  86.2 1066  175   1241  Control       0 0.113  fleetwood
##  9    10  86   1624. 267.  1891  Chillcoma    48 0.104  fleetwood
## 10    11  96.2 1773.  69.7 1842. Control       0 0.140  mac
## # ... with 21 more rows
```

```
#For the heat dataset, I using the mutate function to assign each column is the
#correct data type. I want my treatment as a factor and the rest as numeric
#values.
heatvia <- heatvia %>%
  mutate(treat=factor(treat),
         time=as.numeric(time),
         live=as.numeric(live),
```

```
         dead=as.numeric(dead),
         total=as.numeric(total),
         pc_viability=as.numeric(pc_viability))

#Next I am grouping the dataset by BeeID and using summarize to calculate the
#means for viability, live, dead, total cell counts, mass, and time. I also am
#taking the unique value of each beeIDs treatment and colony.
hvia1 <-  heatvia %>%
  group_by(beeid) %>%
  summarize(via = mean(pc_viability, na.rm=TRUE),
            live = mean(live, na.rm=TRUE),
            dead = mean(dead, na.rm=TRUE),
            total= live+dead,
            treat = unique(treat),
            time = mean(time, na.rm=T),
            mass = mean(mass, na.rm=TRUE),
            colony = unique(colony))
hvia1
```

```
## # A tibble: 39 x 9
##    beeid   via   live  dead total treat    time   mass colony
##    <dbl> <dbl>  <dbl> <dbl> <dbl> <fct>   <dbl>  <dbl> <chr>
##  1     1  80   137.    34.3 171   control     0 0.161  aspen
##  2     2  75.1 497.   163.  661.  hs         29 0.166  willow
##  3     3  81.3 389     90.7 480.  heat       29 0.102  aspen
##  4     4  83.2 275.    56   331.  control     0 0.138  aspen
##  5     5  47.6 279    307   586   heat       29 0.197  aspen
##  6     6  69.6 244.   106.  351.  hs         29 0.105  aspen
##  7     7  80.2 732.   184   916.  control     0 0.128  willow
##  8     8  54.4 145.   121   266.  heat       36 0.0988 aspen
##  9     9  24.4   7.33  22.3  29.7 hs         36 0.0655 aspen
## 10    10  86.9 401.    62   463.  control     0 0.209  aspen
## # ... with 29 more rows
```

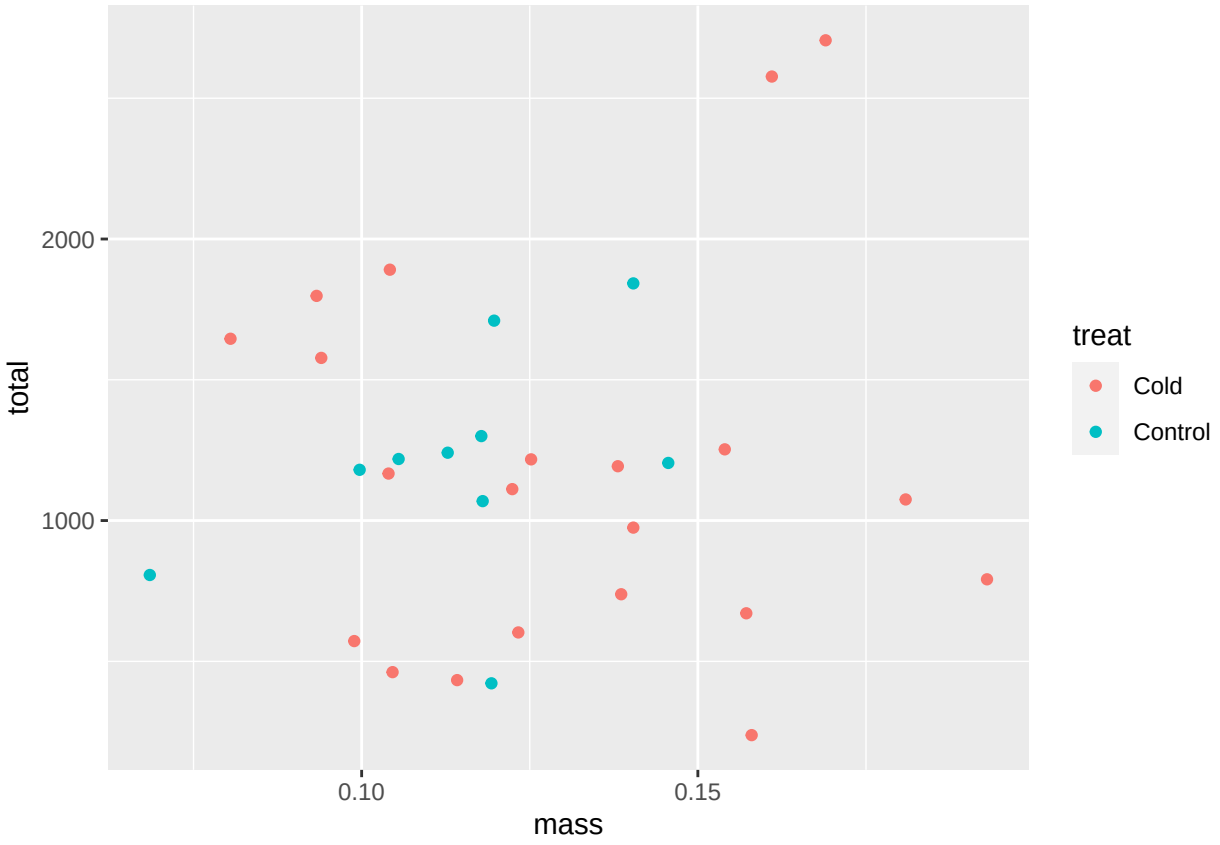
```
#renaming variables to be more decriptive for graphs

cvia1<-cvia1%>%
  mutate(treat=ifelse(treat=="Control", "Control", ifelse(treat=="Chillcoma",
 "Cold", treat)))

hvia1<-hvia1%>%
  mutate(treat=ifelse(treat=="control", "Control", ifelse(treat=="heat",
 "Heat",ifelse(treat=="hs", "Heat Stupor", treat))))
```

4. Scatterplots of mass

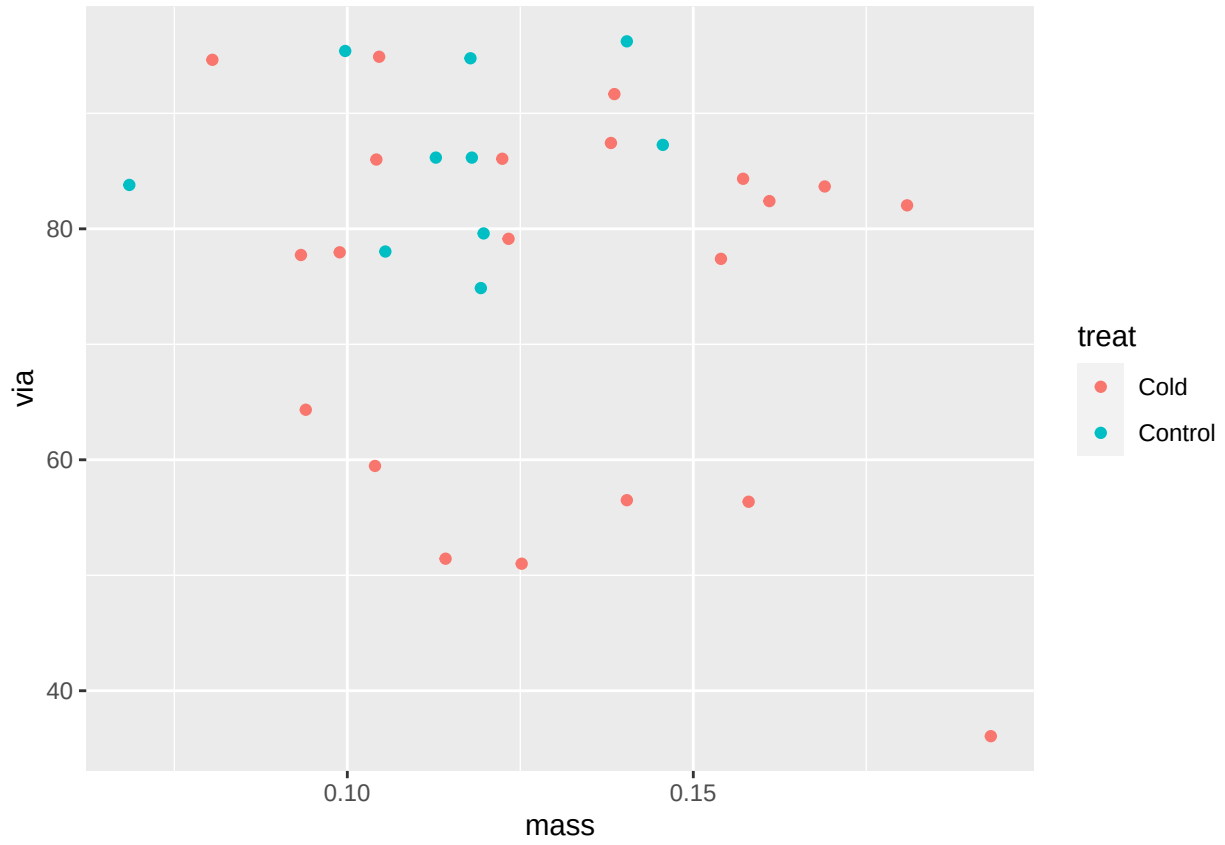
Using scatterplot, I will visualize how mass varies among treatments in both datasets. This is to ensure any results are not due to mass differences.

```
#comparing mass in cold treatment experiments by total cell count and %
#viability
```

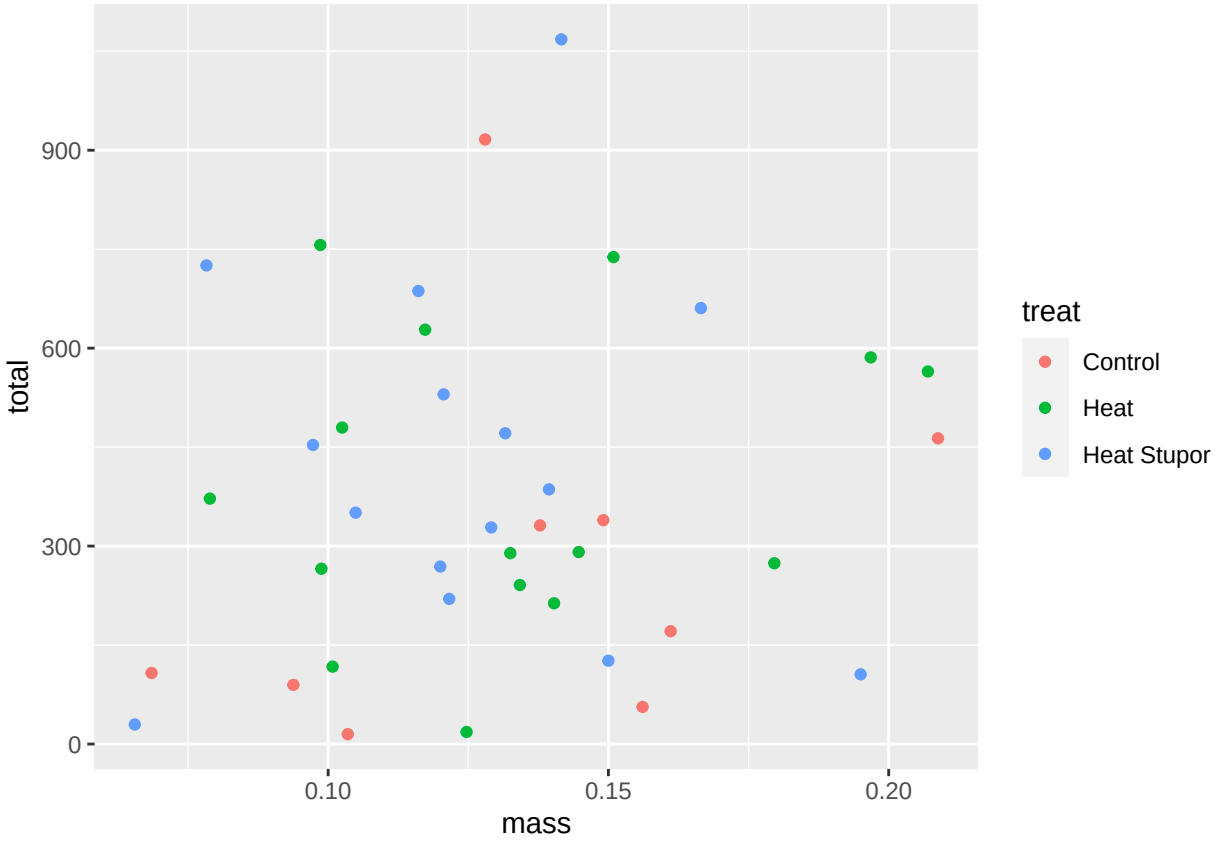
```
cvia1 %>%
  ggplot(aes(x=mass, y=total, col=treat)) +
  geom_point()
```



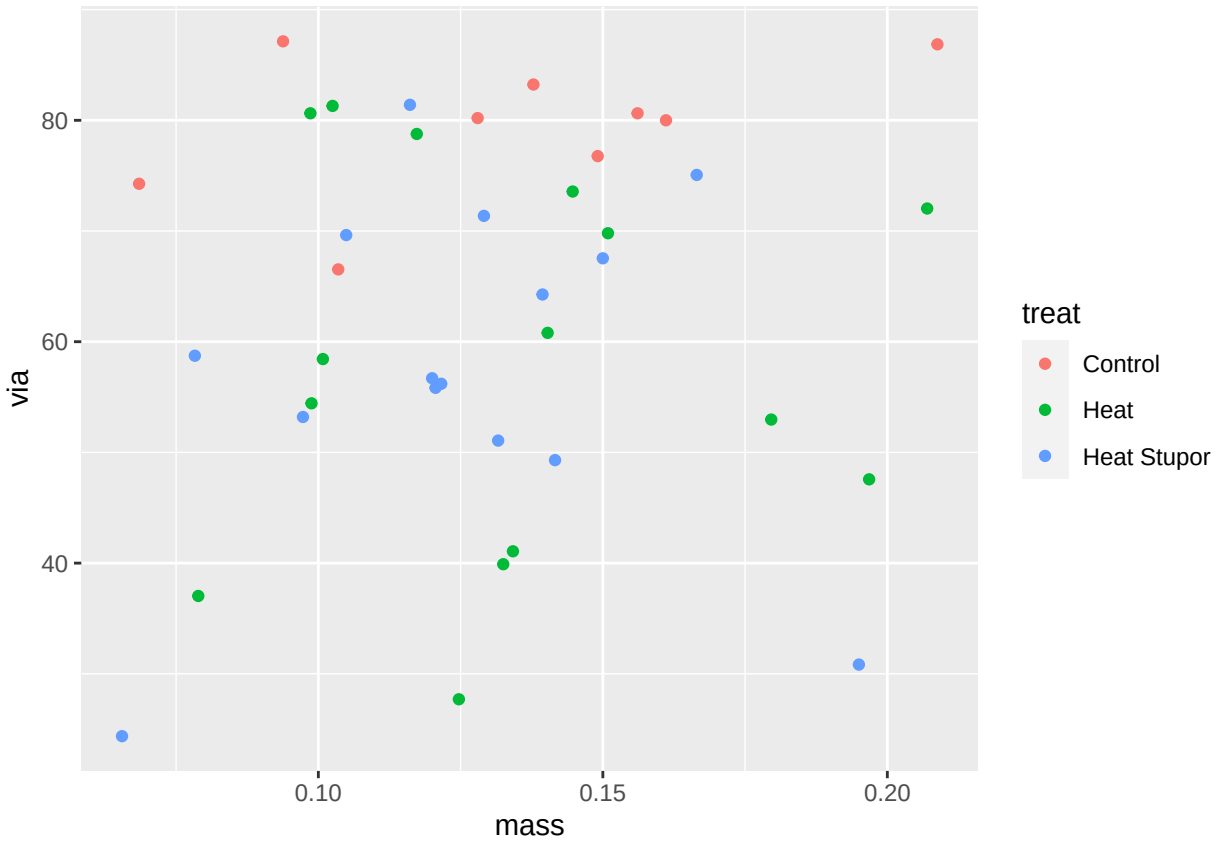

```
cvia1 %>%
  ggplot(aes(x=mass, y=via, col=treat)) +
  geom_point()
```




```
#comparing mass in heat treatment experiments by total cell count and %
#viability

hvia1 %>%
  ggplot(aes(x=mass, y=total, col=treat)) +
  geom_point()
```




```
hvia1 %>%
  ggplot(aes(x=mass, y=via, col=treat)) +
  geom_point()
```

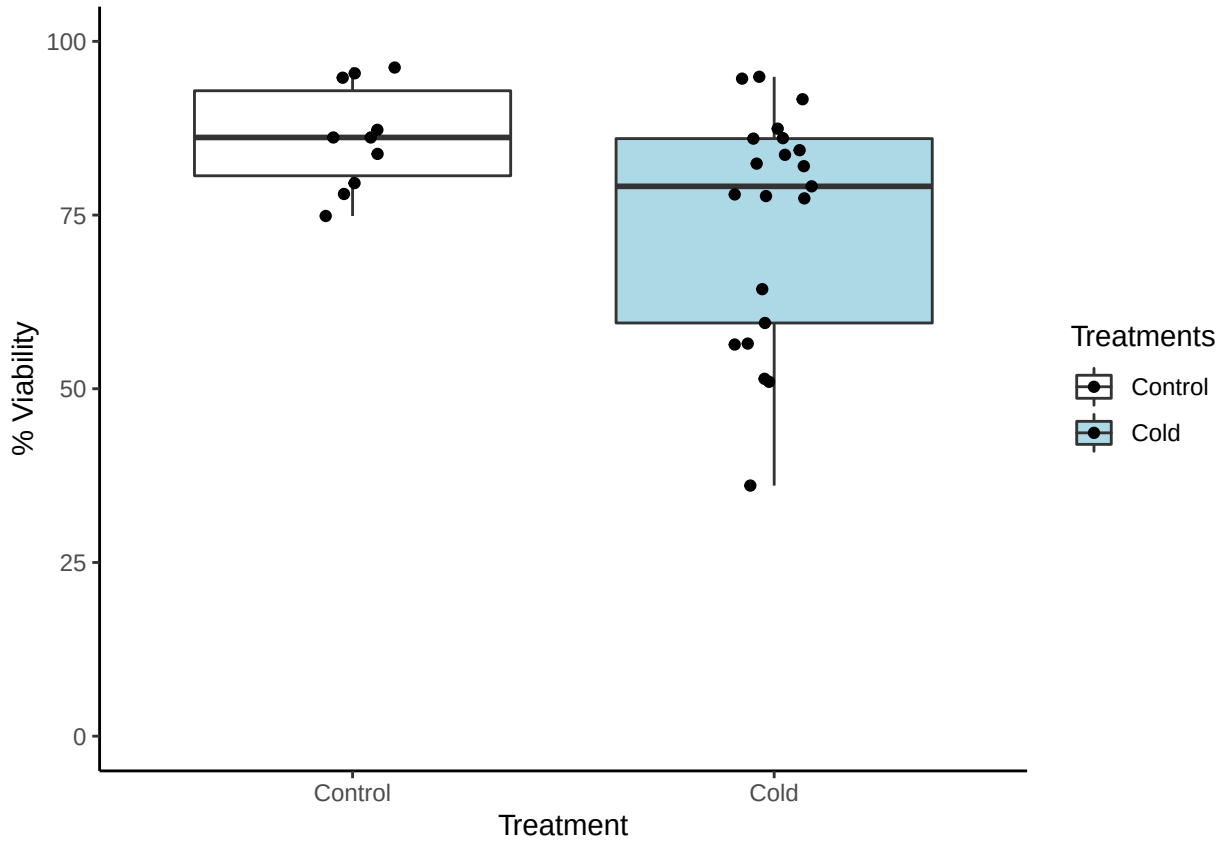


These graphs are showing that mass and viability/ cell count are likely not correlated. This is important to consider before moving forward so we know that a decrease in viability is not due to mass of organism.

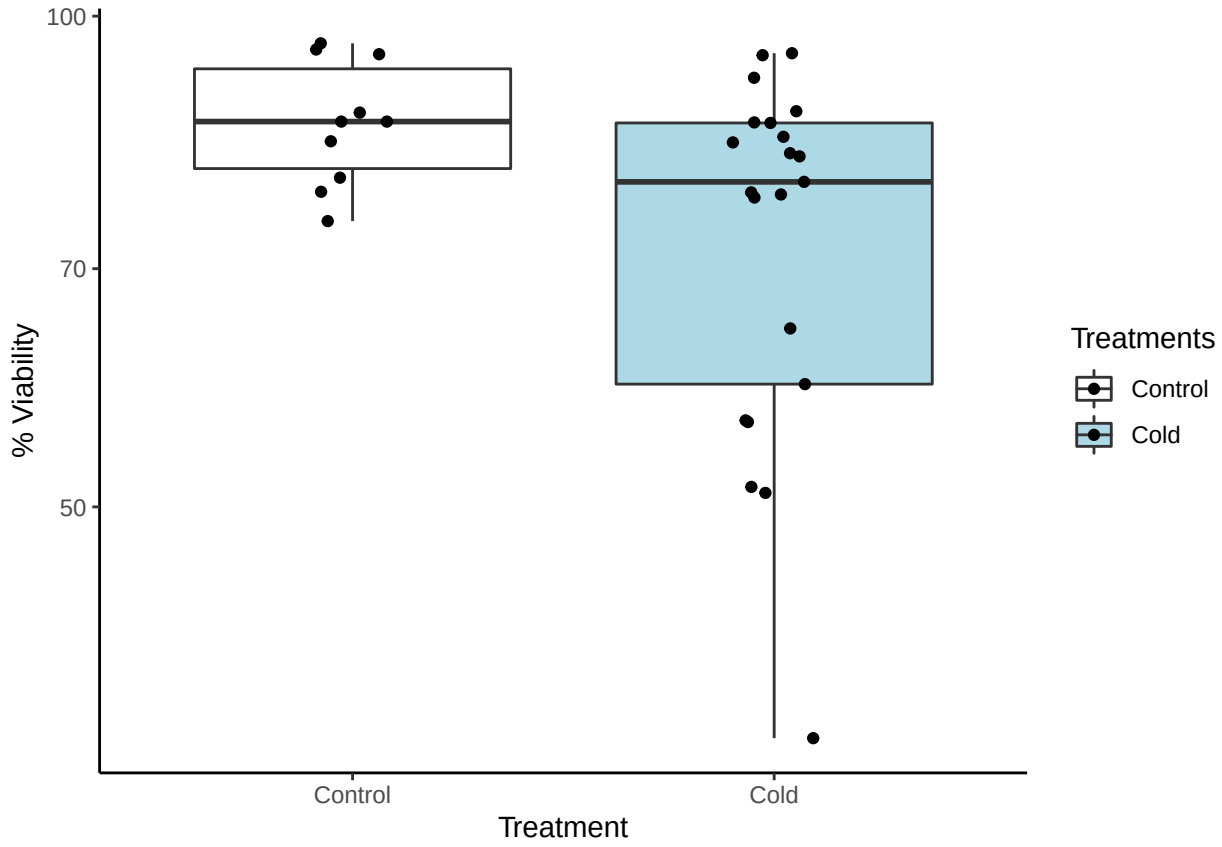
5. Plotting Viability in both datasets with boxplots

Using boxplots, I will visualize how temperature impacts spermatozoa viability in male bumble bees. The first box plot will compare males in control group and males exposed to 4C for 48 hours. This group is called 'coma.' The second plot will compare males who were exposed to 45C and either hit heat stupor (hs) or were just exposed to heat and did not hit heat stupor (heat).

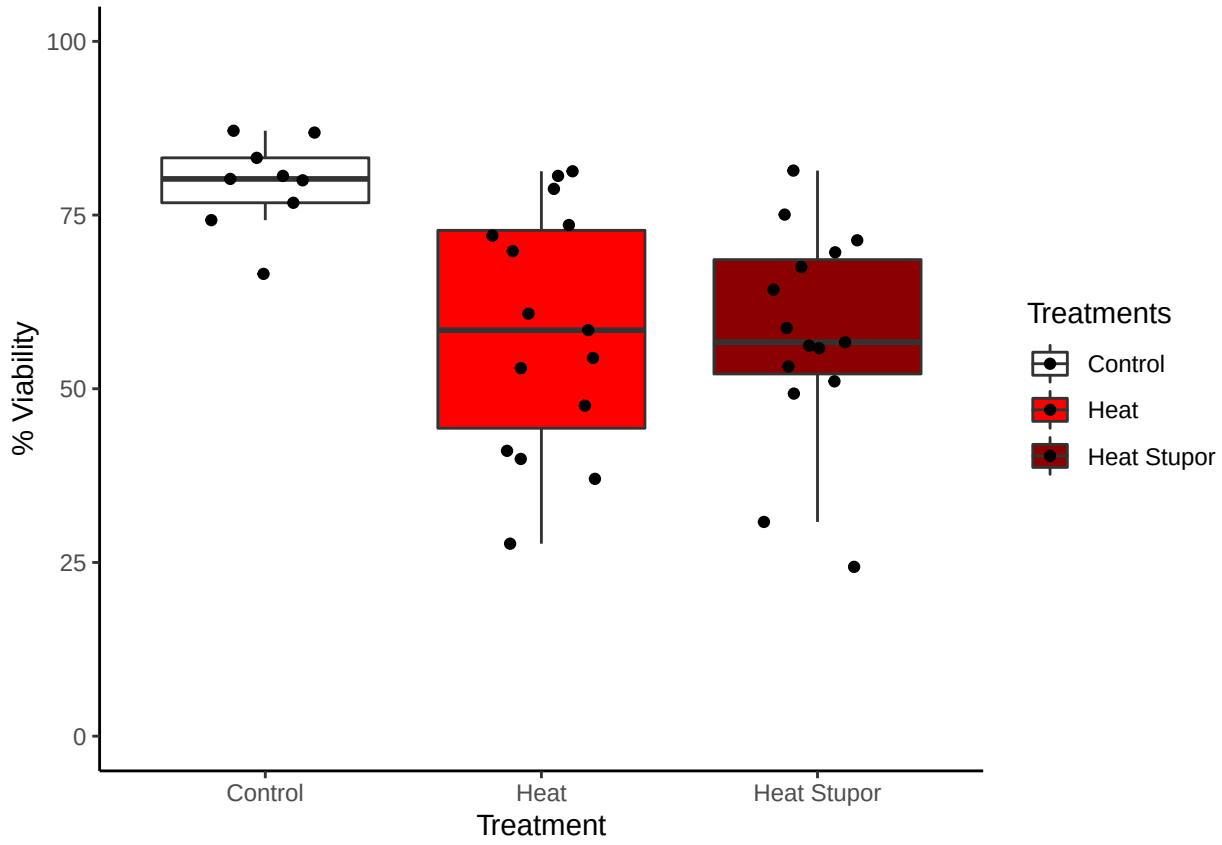
```
#cold treatment boxplot
cvia1$treat<- factor(cvia1$treat, levels= c("Control", "Cold"))
p1 <- cvia1 %>%
  ggplot(aes(x=treat, y=via, fill=treat)) +
  geom_boxplot(outlier.shape = NA) +
  theme_classic() +
  scale_fill_manual(values=c("white","light blue")) +
  labs(x="Treatment", y="% Viability") +
  labs(fill='Treatments') +
  ylim(0,100)+
  geom_jitter(position = position_jitterdodge())

p1
```



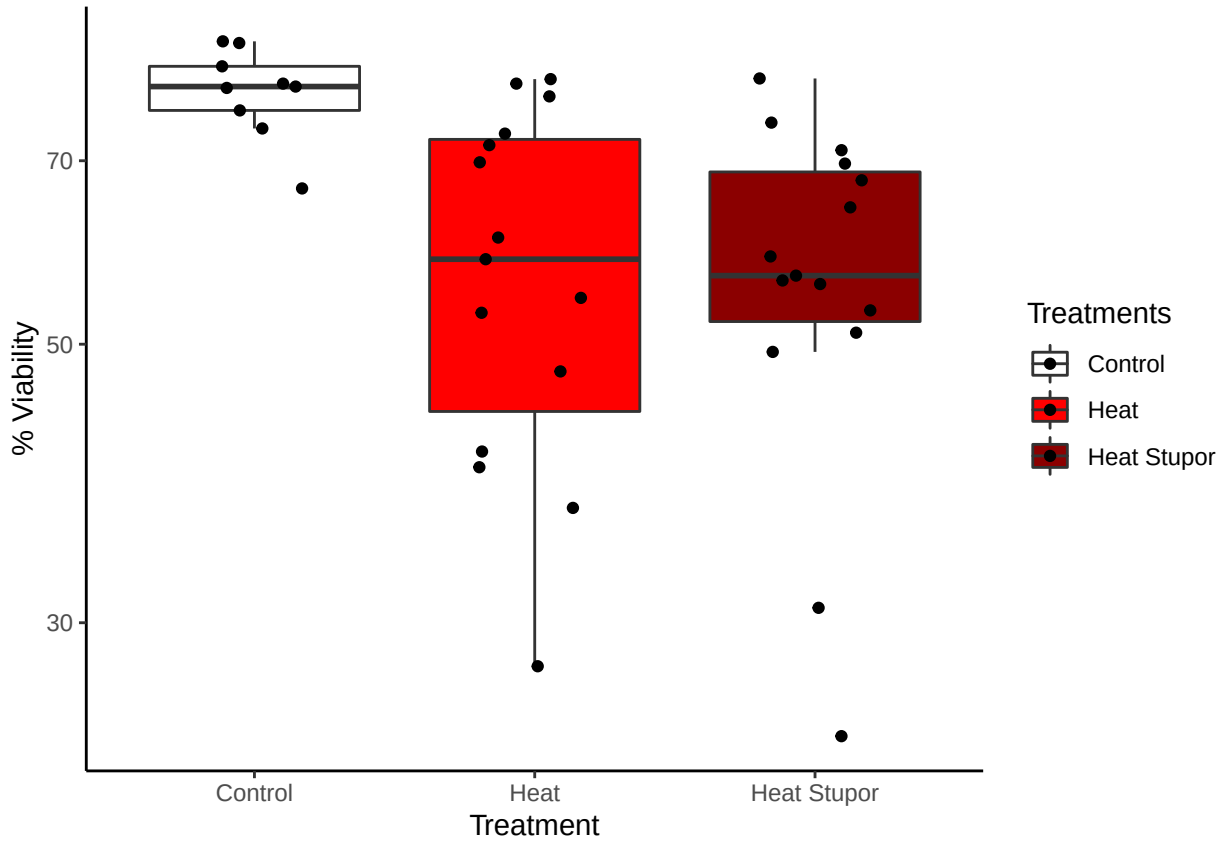

```
#cold treatment boxplot with y axis log scale to see if data is better
#visualized
p2 <- cvia1 %>%
  ggplot(aes(x=treat, y=via, fill=treat)) +
  geom_boxplot(outlier.shape = NA) +
  theme_classic() +
  scale_fill_manual(values=c("white","light blue")) +
  labs(x="Treatment", y="% Viability") +
  labs(fill='Treatments') +
  scale_y_continuous(trans='log10') +
  geom_jitter(position = position_jitterdodge())
p2
```



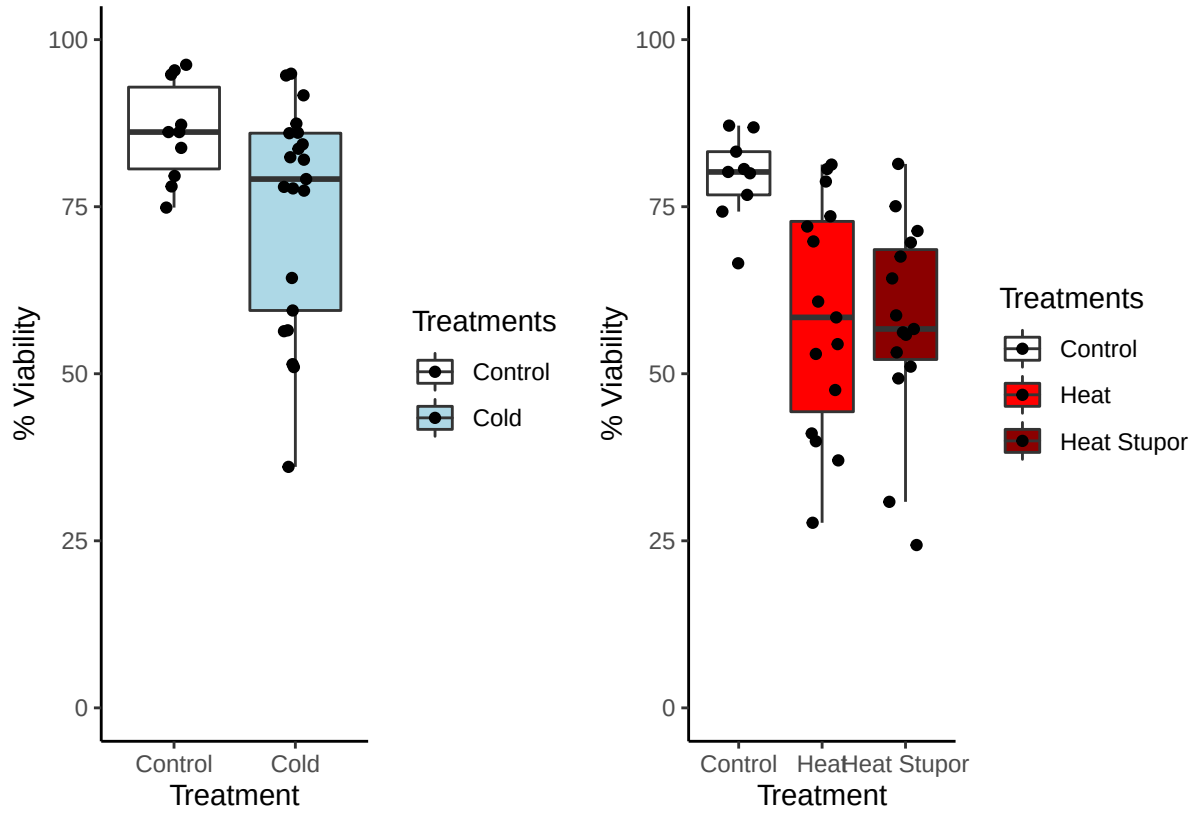

```
#heat treatment boxplot
p3 <- hvia1 %>%
 ggplot(aes(x=treat, y=via, fill=treat)) +
  geom_boxplot(outlier.shape = NA) +
  scale_fill_manual(values=c("white","red", "dark red")) +
  theme_classic() +
  labs(x="Treatment", y="% Viability") +
  labs(fill='Treatments') +
  ylim(0,100)+
  geom_jitter(position = position_jitterdodge())

p3
```




```
#heat treatment boxplot with y axis log scale to see if data is better
#visualized
p4 <- hvia1 %>%
 ggplot(aes(x=treat, y=via, fill=treat)) +
  geom_boxplot(outlier.shape = NA) +
  scale_fill_manual(values=c("white","red", "dark red")) +
  theme_classic() +
  labs(x="Treatment", y="% Viability") +
  labs(fill='Treatments') +
  scale_y_continuous(trans='log10') +
  geom_jitter(position = position_jitterdodge())

p4
```


70
50
30
% Viability
Control
Heat
Heat Stupor
Treatment
Treatments
Control
Heat
Heat Stupor

```
p1+p3
```



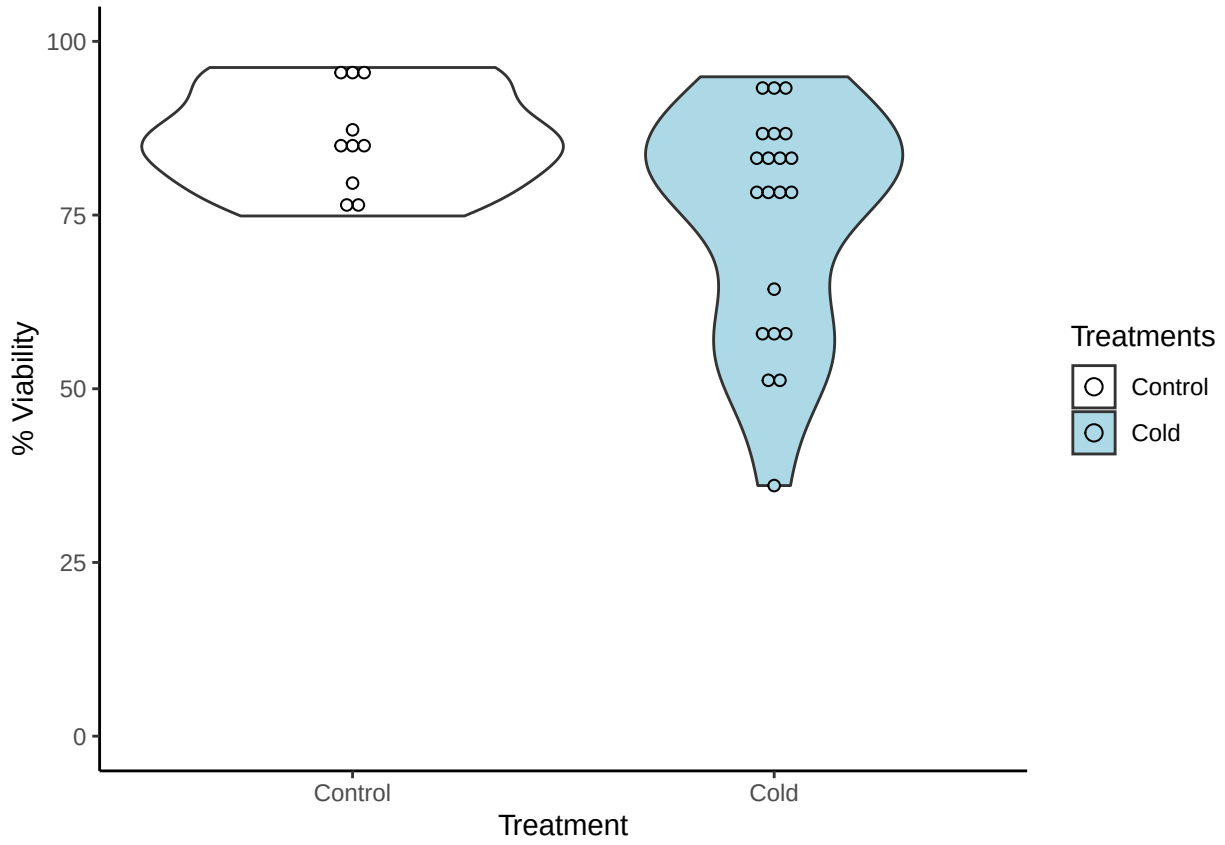
These graphs are showing that sperm count is effected by temperature compared to the control for both heat and cold treatments. More statistical testing is needed to determine if these treatment cause a significant decrease.

6. Plotting Viability in both datasets with violin plots

Another way to visualize this dataset could be using violin plots.

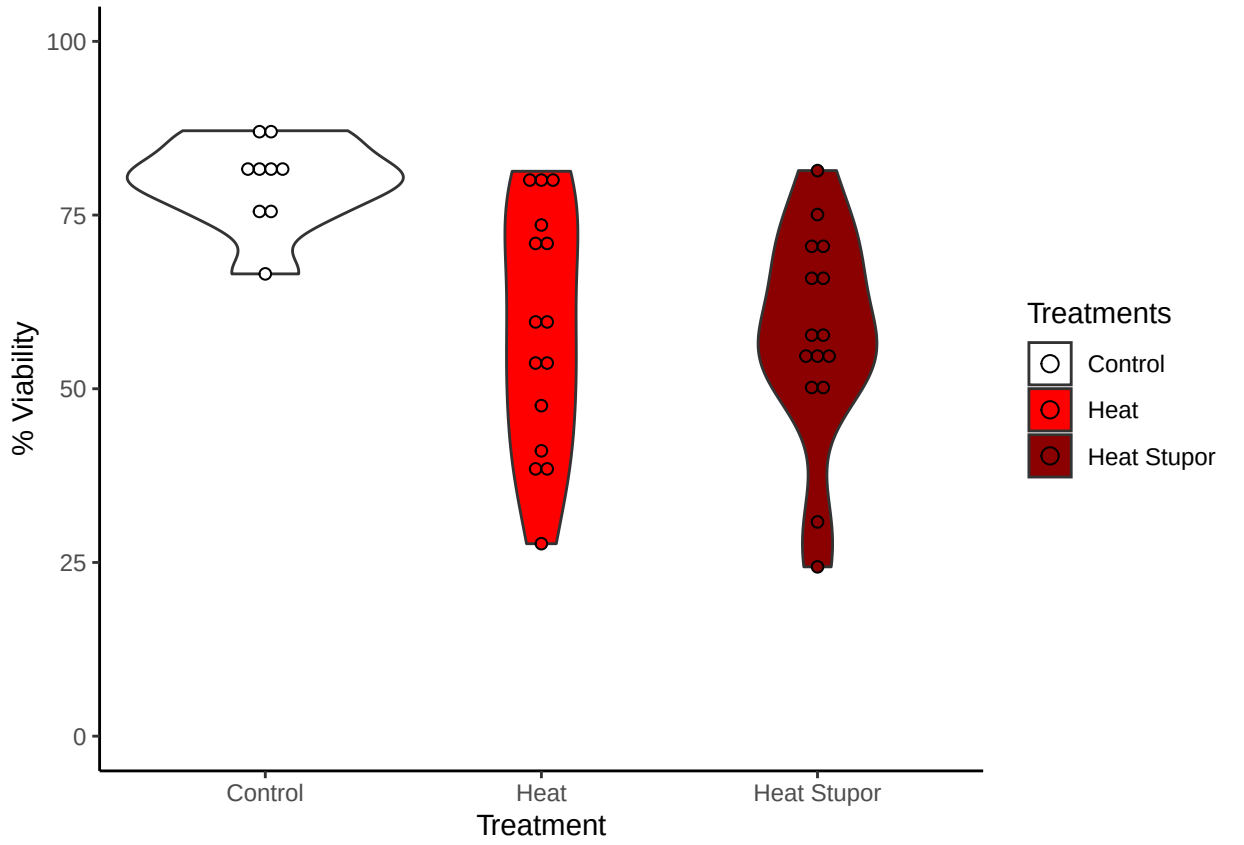
```
#cold treatment violin plot with points
cvia1$treat<- factor(cvia1$treat, levels= c("Control", "Cold"))
h1<- cvia1 %>%
 ggplot(aes(x=treat, y=via, fill=treat)) +
    geom_violin(width=1) +
  geom_dotplot(binaxis = "y",
               stackdir = "center",
               dotsize = 0.5)+
    theme_classic() +
    scale_fill_manual(values=c("white", "light blue")) +
    labs(x="Treatment", y="% Viability") +
    ylim(0,100)+
    labs(fill='Treatments')
h1
```

```
## Bin width defaults to 1/30 of the range of the data. Pick better value with `binwidth`.
```



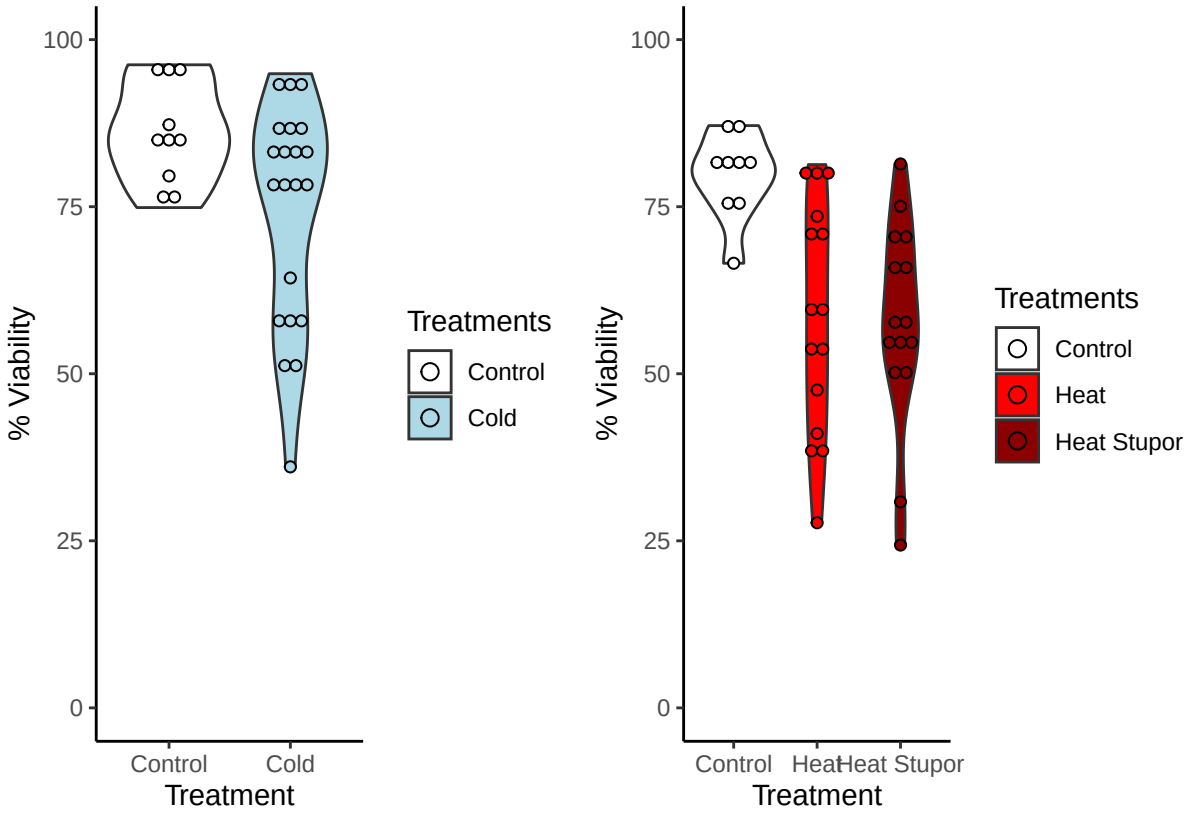

```
#heat treatment violin plot with points
h2 <- hvia1 %>%
 ggplot(aes(x=treat, y=via, fill=treat)) +
    geom_violin(width=1) +
    geom_dotplot(binaxis = "y",
                 stackdir = "center",
                 dotsize = 0.5)+
    theme_classic() +
    scale_fill_manual(values=c("white","red", "dark red")) +
    labs(x="Treatment", y="% Viability") +
    ylim(0,100)+
    labs(fill='Treatments')
h2
```

```
## Bin width defaults to 1/30 of the range of the data. Pick better value with `binwidth`.
```




```
h1+h2
```

```
## Bin width defaults to 1/30 of the range of the data. Pick better value with `binwidth`.
## Bin width defaults to 1/30 of the range of the data. Pick better value with `binwidth`.
```



These graphs are just showing another way to visualize how temperature decreases sperm viability in males.

8. Feedback Implementation

After receiving feedback from my peers I implemented changes to my final project. I made some minor changes such as making sure my words did not run off the page, italicizing scientific names, mute loading packages, and changed the order of cold treatment graphs so control is first. I also added summaries below each plot of findings, added points to my violin plots, used patchwork to combine the best two graphs together, renamed treatments, and adjusted y axis.

Overall, getting feedback from my peers was helpful in making my figures closer to being ready to be published. I also was able to implement their feedback to make this document easier to read and understand.

7. References

David, J. R., L. O. Araripe, M. Chakir, H. Legout, B. Lemos, G. Pétavy, C. Rohmer, D. Joly, and B. Moreteau. 2005. "Male Sterility at Extreme Temperatures: A Significant but Neglected Phenomenon for Understanding Drosophila Climatic Adaptations." *Journal of Evolutionary Biology* 18 (4): 838–46. https://doi.org/10.1111/j.1420-9101.2005.00914.x.

Heerwaarden, Belinda van, and Carla M. Sgrò. 2021. "Male Fertility Thermal Limits Predict Vulnerability to Climate Warming." *Nature Communications* 12 (1): 2214. https://doi.org/10.1038/s41467-021-22546-w.

Kerr, J. T., A. Pindar, P. Galpern, L. Packer, S. G. Potts, S. M. Roberts, P. Rasmont, et al. 2015. "Climate Change Impacts on Bumblebees Converge Across Continents." *Science* 349 (6244): 177–80. https://doi.org/10.1126/science.aaa7031.