

Envr_Violin_Plots

2024-10-08

#Download packages

```
library(dplyr) #Used for working with data frames
```

```
##
```

```
## 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(lubridate) #Used for time-date conversions
```

```
##
```

```
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## date, intersect, setdiff, union
```

```
library(readr) #Used to read the CSV file
```

```
library(ggplot2) #plot with ggplot
```

```
library(cowplot) #arrange ggplots
```

```
##
```

```
## Attaching package: 'cowplot'
```

```
## The following object is masked from 'package:lubridate':
```

```
##
```

```
## stamp
```

```
library(ggpmisc)
```

```
## Loading required package: ggpp
```

```
## Registered S3 methods overwritten by 'ggpp':
##   method          from
##   heightDetails.titleGrob ggplot2
##   widthDetails.titleGrob  ggplot2

##
## Attaching package: 'ggpp'

## The following object is masked from 'package:ggplot2':
##
##   annotate
```

Salinity file read-in and cleaning

```
#set working directory to files location
setwd("/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/src/NM")
#Read in salinity files
DEBY_sal <- read.csv("/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/data/envr_raw_data/DEBY_sal.csv")
FL_sal <- read.csv("/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/data/envr_raw_data/FL_sal.csv")
JR_sal <- read.csv("/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/data/envr_raw_data/JR_sal.csv")
#remove uncorrected salinity column and rename corrected_salinity to salinity
JR_sal <- subset(JR_sal, select = -c(salinity))

colnames(JR_sal) <- c("site_name", "datetime", "salinity")

LA_sal <- read.csv("/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/data/envr_raw_data/LA_sal.csv")
LOLA_sal <- read.csv("/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/data/envr_raw_data/LOLA_sal.csv")
ME_sal <- read.csv("/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/data/envr_raw_data/ME_sal.csv")
NH_sal <- read.csv("/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/data/envr_raw_data/NH_sal.csv")
TX_sal <- read.csv("/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/data/envr_raw_data/TX_sal.csv")

#convert all date times to POSIXct
DEBY_sal$datetime <- as.POSIXct(DEBY_sal$datetime, "%Y-%m-%d %H:%M:%S", tz = "")

FL_sal$datetime <- as.POSIXct(FL_sal$datetime, "%m/%d/%y %H:%M", tz = "")

JR_sal$datetime <- as.POSIXct(JR_sal$datetime, "%Y-%m-%d %H:%M:%S", tz = "")

LA_sal$datetime <- as.POSIXct(LA_sal$datetime, "%Y-%m-%d %H:%M:%S", tz = "")

LOLA_sal$datetime <- as.POSIXct(LOLA_sal$datetime, "%Y-%m-%d %H:%M:%S", tz = "")

ME_sal$datetime <- as.POSIXct(ME_sal$datetime, "%m/%d/%y %H:%M", tz = "")

NH_sal$datetime <- as.POSIXct(NH_sal$datetime, "%m/%d/%y %H:%M", tz = "")

TX_sal$datetime <- as.POSIXct(TX_sal$datetime, "%m/%d/%y %H:%M", tz = "")
```

```
#ME only has 2 data points in October 2022 and one in July 2022, so exclude those points
ME_sal <- ME_sal[-c(26312, 26313, 26314), ]
```

```
#make all salinity variables numeric
DEBY_sal$salinity <- as.numeric(DEBY_sal$salinity)
FL_sal$salinity <- as.numeric(FL_sal$salinity)
JR_sal$salinity <- as.numeric(JR_sal$salinity)
LA_sal$salinity <- as.numeric(LA_sal$salinity)
```

```
## Warning: NAs introduced by coercion
```

```
LOLA_sal$salinity <- as.numeric(LOLA_sal$salinity)
ME_sal$salinity <- as.numeric(ME_sal$salinity)
NH_sal$salinity <- as.numeric(NH_sal$salinity)
TX_sal$salinity <- as.numeric(TX_sal$salinity)
```

```
#remove extra columns from NH and TX
```

```
NH_sal <- NH_sal[, c(1,2,3)]
TX_sal <- TX_sal[, c(1,2,3)]
```

```
#merge data frames
```

```
full_sal <- rbind(DEBY_sal, FL_sal, JR_sal, LA_sal, LOLA_sal, ME_sal, NH_sal, TX_sal)
```

```
#remove NAs
```

```
na_full_sal <- is.na(full_sal) # store our NAs in a variable
```

```
summary(na_full_sal) # we have 3696 NAs in datetime and 192381 NAs in salinity that are stored as "TRUE"
```

```
##   site_name      datetime      salinity
## Mode :logical   Mode :logical   Mode :logical
## FALSE:2695809   FALSE:2695255   FALSE:2503428
##                TRUE :554        TRUE :192381
```

```
full_sal <- na.omit(full_sal) #remove NAs using na.omit
```

```
#filter out temps below 0 and above 40
```

```
filtered_sal <- full_sal %>%
  filter(between(salinity, 0, 40))
```

```
#average monthly salinity
```

```
monthly_sal <- filtered_sal %>%
  mutate(year = year(datetime), month = month(datetime)) %>%
  group_by(site_name, year, month) %>%
  summarise(mean_sal = mean(salinity))
```

```
## 'summarise()' has grouped output by 'site_name', 'year'. You can override using
## the '.groups' argument.
```

```
#make sure all 8 sites are still present
```

```
monthly_sal_sites <- list(unique(monthly_sal$site_name))
```

```
monthly_sal_sites
```

```

## [[1]]
## [1] "DEBY" "FL"    "JR"    "LA"    "LOLA" "ME"    "NH"    "TX"

#write to csv for future analyses
write.csv(monthly_sal, "/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/data/envr_raw_

#Temperature

#set working directory to files location
setwd("/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/src/NM")
#Read in temp files
DEBY_temp <- read.csv("/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/data/envr_raw_

FL_temp <- read.csv("/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/data/envr_raw_da

JR_temp <- read.csv("/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/data/envr_raw_da

#remove uncorrected salinity column and rename corrected_salinity to salinity
JR_temp <- subset(JR_temp, select = -c(temp))
colnames(JR_temp) <- c("site_name", "datetime", "temp")

LA_temp <- read.csv("/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/data/envr_raw_da

LOLA_temp <- read.csv("/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/data/envr_raw_

ME_temp <- read.csv("/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/data/envr_raw_da

NH_temp <- read.csv("/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/data/envr_raw_da

TX_temp <- read.csv("/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/data/envr_raw_da

#convert all date times to POSIXct
DEBY_temp$datetime <- as.POSIXct(DEBY_temp$datetime, "%Y-%m-%d %H:%M:%S", tz = "")

FL_temp$datetime <- as.POSIXct(FL_temp$datetime, "%m/%d/%y %H:%M", tz = "")

JR_temp$datetime <- as.POSIXct(JR_temp$datetime, "%Y-%m-%d %H:%M", tz = "")

LA_temp$datetime <- as.POSIXct(LA_temp$datetime, "%Y-%m-%d %H:%M:%S", tz = "")

LOLA_temp$datetime <- as.POSIXct(LOLA_temp$datetime, "%Y-%m-%d %H:%M:%S", tz = "")

ME_temp$datetime <- as.POSIXct(ME_temp$datetime, "%m/%d/%y %H:%M", tz = "")

NH_temp$datetime <- as.POSIXct(NH_temp$datetime, "%m/%d/%y %H:%M", tz = "")

TX_temp$datetime <- as.POSIXct(TX_temp$datetime, "%m/%d/%y %H:%M", tz = "")

#make all temp variables numeric
DEBY_temp$temp <- as.numeric(DEBY_temp$temp)
FL_temp$temp <- as.numeric(FL_temp$temp)
JR_temp$temp <- as.numeric(JR_temp$temp)
LA_temp$temp <- as.numeric(LA_temp$temp)

```

```
## Warning: NAs introduced by coercion
```

```
LOLA_temp$temp <- as.numeric(LOLA_temp$temp)
ME_temp$temp <- as.numeric(ME_temp$temp)
NH_temp$temp <- as.numeric(NH_temp$temp)
TX_temp$temp <- as.numeric(TX_temp$temp)

#remove extra columns from NH
NH_temp <- NH_temp[, c(1,2,3)]

#merge data frames
full_temp <- rbind(DEBY_temp, FL_temp, JR_temp, LA_temp, LOLA_temp, ME_temp, NH_temp, TX_temp)

#remove NAs
na_full_temp <- is.na(full_temp) # store our NAs in a variable
summary(na_full_temp) # we have 3820 NAs in datetime and 180627 NAs in temp that are stored as "TRUE"
```

```
##   site_name      datetime      temp
##   Mode :logical   Mode :logical   Mode :logical
##   FALSE:2704146   FALSE:2703590   FALSE:2523519
##                   TRUE :556       TRUE :180627
```

```
full_temp <- na.omit(full_temp) #remove NAs using na.omit

#filter out temps below 0 and above 40
filtered_temp <- full_temp %>%
  filter(between(temp, 0, 40))
```

Violin plots with the raw data cause R to crash due to large data size. Therefore, I will make the violin plots using monthly temp/salinity averages rather than raw data.

```
#average monthly temperatures
monthly_temp <- filtered_temp %>%
  mutate(year = year(datetime), month = month(datetime)) %>%
  group_by(site_name, year, month) %>%
  summarise(mean_temp = mean(temp))
```

```
## 'summarise()' has grouped output by 'site_name', 'year'. You can override using
## the '.groups' argument.
```

```
#make sure all 8 sites are still present
monthly_temp_sites <- list(unique(monthly_temp$site_name))

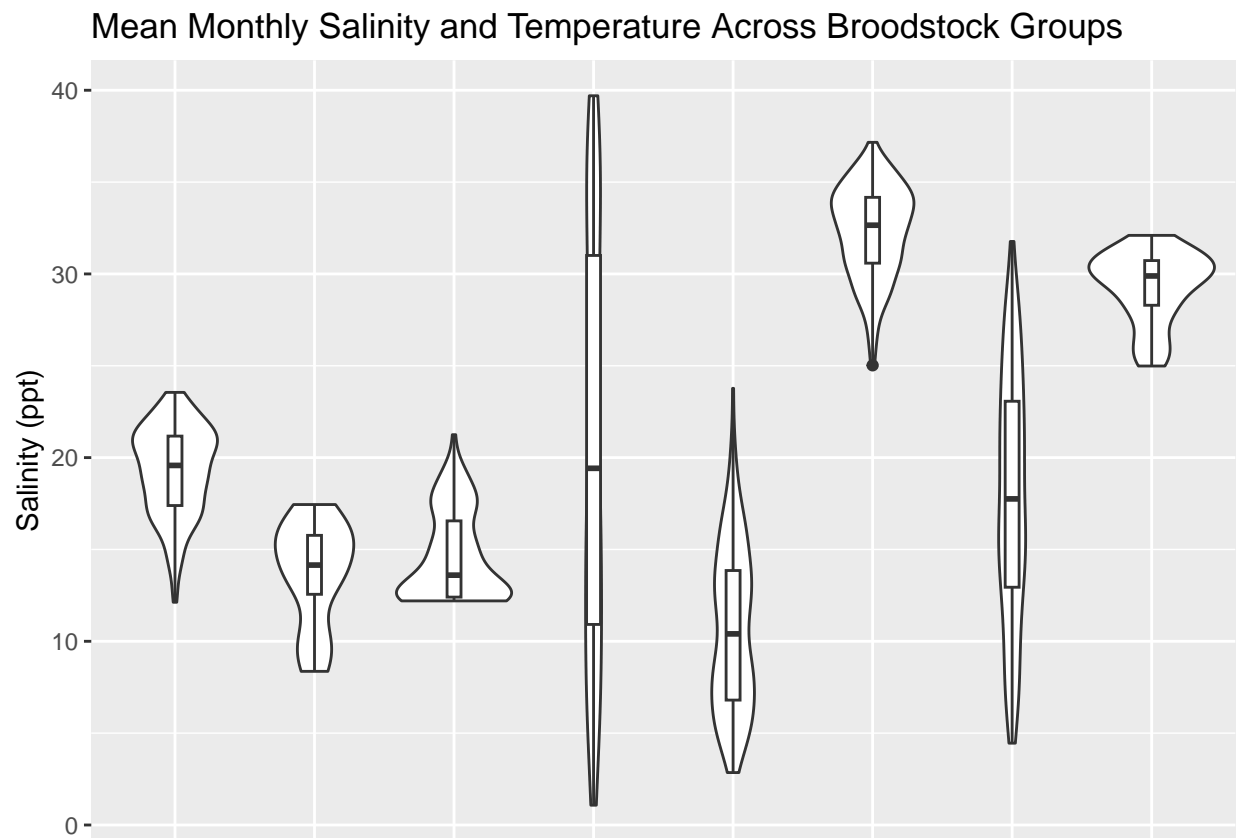
monthly_temp_sites
```

```
## [[1]]
## [1] "DEBY" "FL"   "JR"   "LA"   "LOLA" "ME"   "NH"   "TX"
```

```
#write to csv for future analyses
write.csv(monthly_temp, "/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/data/envr_ra
```

Violin Plots

```
sal_violin <- ggplot(monthly_sal, aes(x = site_name, y = mean_sal)) +  
  geom_violin()+  
  geom_boxplot(width = .1) +  
  ggtitle("Mean Monthly Salinity and Temperature Across Broodstock Groups")+  
  ylab("Salinity (ppt)") +  
  scale_x_discrete(name = "Site Name", limits = c("DEBY", "LOLA", "JR", "TX", "LA", "FL", "NH", "ME")) +  
  theme(axis.title.x = element_blank(), axis.ticks.x = element_blank(), axis.text.x = element_blank())  
  
sal_violin
```



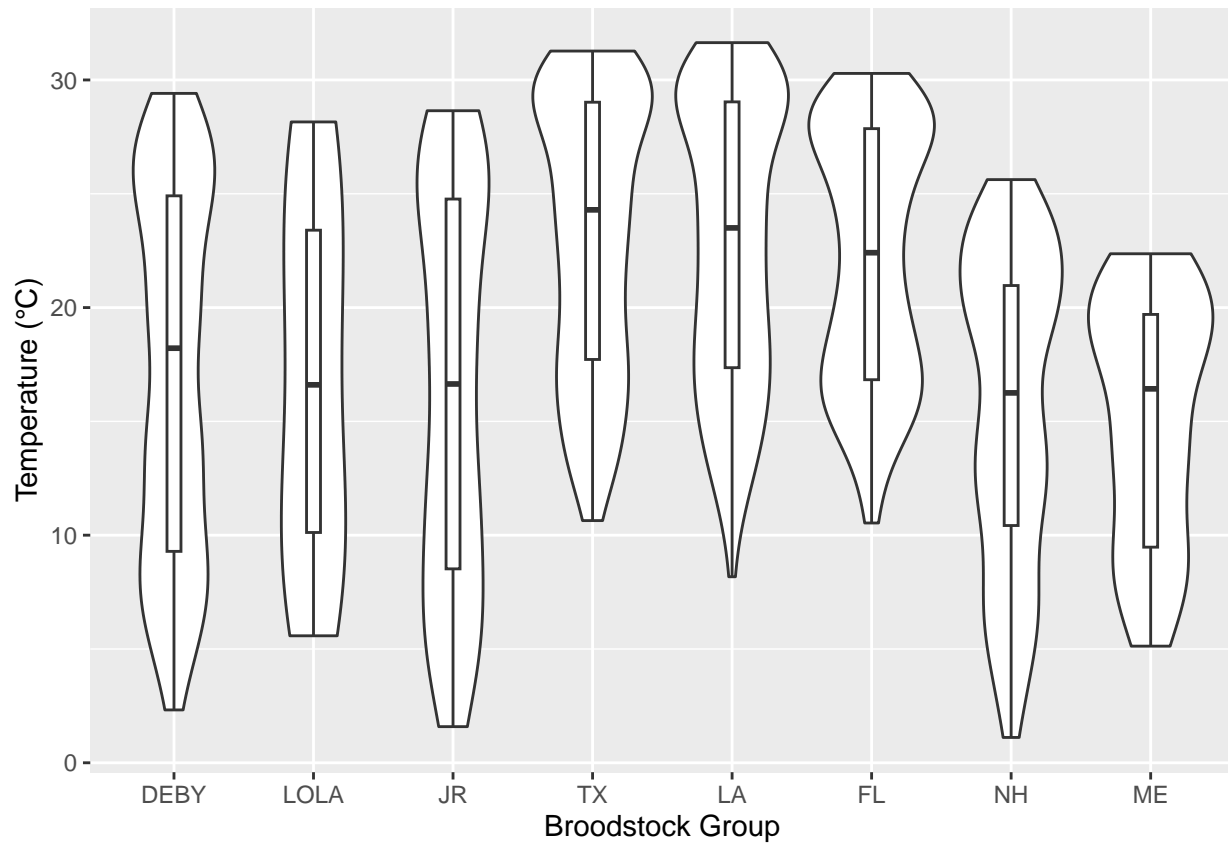
```
ggsave(sal_violin,  
  filename = "salinity_violinplot.png",  
  device = "png",  
  path = "/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/results/NM_results")
```

Saving 6.5 x 4.5 in image

```
temp_violin <- ggplot(monthly_temp, aes(x = site_name, y = mean_temp)) +  
  geom_violin()+  
  geom_boxplot(width = .1) +  
  xlab("Broodstock Group") +
```

```
ylab("Temperature (°C)") +
scale_x_discrete(name = "Broodstock Group", limits = c("DEBY", "LOLA", "JR", "TX", "LA", "FL", "NH", "ME")) +
theme(plot.title = element_blank())
```

```
temp_violin
```



```
ggsave(temp_violin,
        filename = "temperature_violinplot.png",
        device = "png",
        path = "/Users/nicolemongillo/Desktop/GitHub/MVP_Chesapeake_VIMS_hatchery/results/NM_results")
```

```
## Saving 6.5 x 4.5 in image
```

```
##merge plots
plot_grid(sal_violin, temp_violin, ncol = 1, align = "v")
```

Mean Monthly Salinity and Temperature Across Broodstock Groups

