

Ruminant violins

Walker Mellon

3/18/2025

###Make Violins

Read data in. This dataset has been specifically designed to be used just for this violin script.

```
library(tidyr)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v purrr      1.0.2
## v forcats    1.0.0      v readr      2.1.5
## v ggplot2    3.5.1      v stringr    1.5.1
## v lubridate  1.9.3      v tibble     3.2.1
## -- 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
```

```
library(cowplot)
```

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

```
# lh <- read.csv("LifeHistory.csv")
# neo <- read.csv("min10prev.csv")
# newData<-left_join(lh, neo, by = "taxa_id")
# DataRum <- subset(newData, Family == "Bovidae" | Family == "Cervidae" | Family == "Giraffidae" | Family == "Elephantidae")
# Data<-DataRum
```

```
Data <- read.csv("min10ruminants.csv")
```

```
#
# d2<-read.csv("min20-2022.05.16.csv")
#
#
# # Step 1: Get common columns (columns shared by d2 and Data)
# common_cols <- intersect(colnames(d2), colnames(Data))
#
# # Step 2: Get extra columns in Data (not found in d2)
# extra_cols <- setdiff(colnames(Data), colnames(d2))
#
```

```

# # Step 3: Combine the columns: first common (in d2 order), then extras from Data
# desired_order <- c(common_cols, extra_cols)
#
# # Step 4: Reorder Data according to desired_order
# Data_reordered <- Data[, desired_order]
#
# write.csv(Data_reordered, "min10ruminants.csv")

```

Subset by clade

```

# Subset by specific species

```

```

Data$FemaleNeoplasiaPrevalence<-Data$FemaleNeoplasia/Data$Female
Data$MaleNeoplasiaPrevalence<-Data$MaleNeoplasia/Data$Male

```

```

Data_long <- pivot_longer(Data,
                           cols = c(FemaleNeoplasiaPrevalence, MaleNeoplasiaPrevalence),
                           names_to = "Sex",
                           values_to = "Prevalence")

```

```

# Clean up 'Sex' column values

```

```

Data_long$Sex <- gsub("NeoplasiaPrevalence", "", Data_long$Sex)

```

wilcox test

```

wil_neo<-wilcox.test(Data$FemaleNeoplasiaPrevalence, Data$MaleNeoplasiaPrevalence, paired = TRUE)

```

```

## Warning in wilcox.test.default(Data$FemaleNeoplasiaPrevalence,
## Data$MaleNeoplasiaPrevalence, : cannot compute exact p-value with ties

```

```

## Warning in wilcox.test.default(Data$FemaleNeoplasiaPrevalence,
## Data$MaleNeoplasiaPrevalence, : cannot compute exact p-value with zeroes

```

```

wilcox_p <- signif(wil_neo$p.value, 3)

```

Violin code

```

mam_vio <- ggplot(Data_long, aes(x = Sex, y = 100 * Prevalence, fill = Sex)) +
  geom_violin(adjust = 1) +
  theme(
    legend.title = element_text(size = 21, face = "bold"),
    legend.text = element_text(size = 20),
    plot.title = element_text(size = 20, face = "bold"),
    axis.text = element_text(size = 16),
    axis.title = element_text(size = 16)
  ) +
  labs(
    title = "Neoplasia Prevalence by Sex in Ruminants",
    subtitle = paste0("Wilcoxon Signed-Rank Test: P-Value = ", wilcox_p), # Update this as needed
    fill = 'Sex'
  ) +
  ylab("Neoplasia Prevalence (%)") +
  xlab("Sex")

```

Add some mean bars and clean up the background with cowplot

```

mam_vio + geom_jitter(shape=16, position=position_jitter(0.2))+
stat_summary(fun=mean, geom="crossbar", size=0.7) +

```

```
scale_fill_manual(values = c("#FFC627", "#8C1D40", "#F0E442")) +
theme_cowplot(12)+
  theme(legend.position = "none")
```

```
## 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.

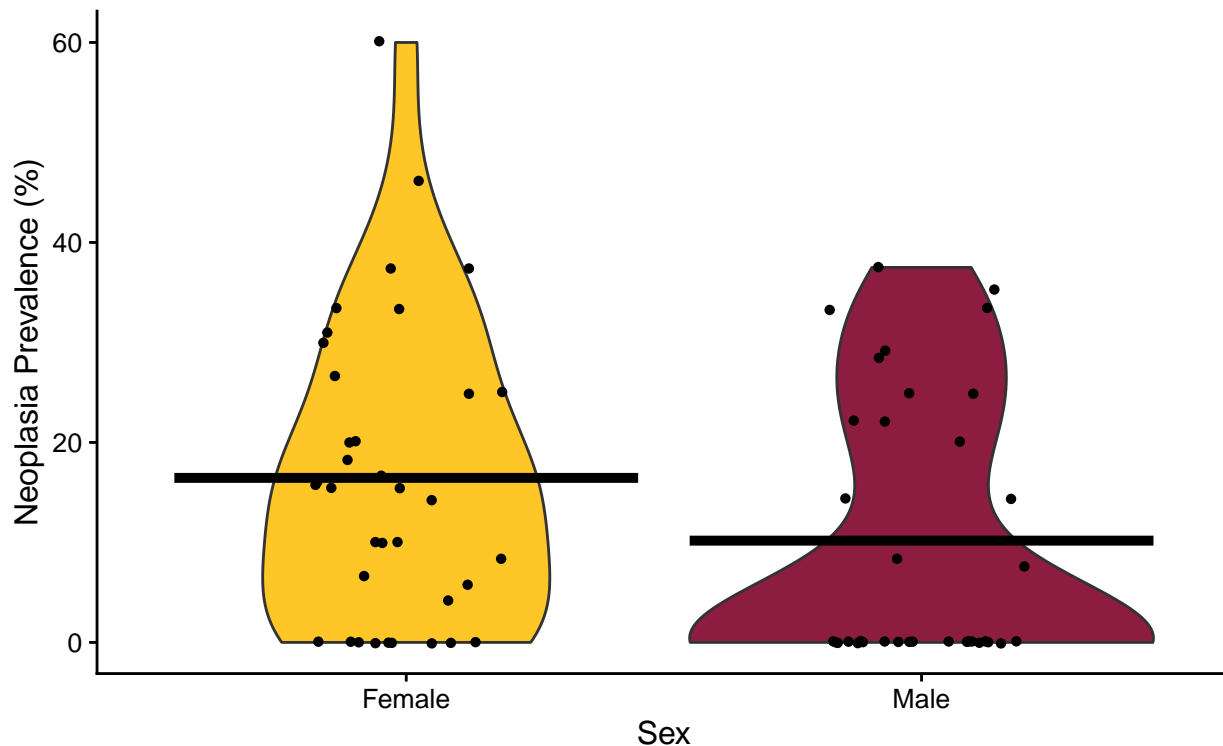
## Warning: Removed 5 rows containing non-finite outside the scale range
## (`stat_ydensity()`).

## Warning: Removed 5 rows containing non-finite outside the scale range
## (`stat_summary()`).

## Warning: Removed 5 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

Neoplasia Prevalence by Sex in Ruminants

Wilcoxon Signed-Rank Test: P-Value = 0.13



```
ggsave("RuminantSexNeoViolins.png", width = 10, height = 6, units = "in")
```

```
## Warning: Removed 5 rows containing non-finite outside the scale range
## (`stat_ydensity()`).

## Warning: Removed 5 rows containing non-finite outside the scale range
## (`stat_summary()`).

## Warning: Removed 5 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

Subset by clade

```
# Subset by specific species
```

```
Data$FemaleMalignancyPrevalence<-Data$FemaleMalignant/Data$Female  
Data$MaleMalignancyPrevalence<-Data$MaleMalignant/Data$Male
```

```
Data_long_mal <- pivot_longer(Data,  
                               cols = c(FemaleMalignancyPrevalence, MaleMalignancyPrevalence),  
                               names_to = "Sex",  
                               values_to = "Prevalence")
```

```
# Clean up 'Sex' column values
```

```
Data_long_mal$Sex <- gsub("MalignancyPrevalence", "", Data_long$Sex)
```

wilcox test

```
wil_mal<-wilcox.test(Data$FemaleMalignancyPrevalence, Data$MaleMalignancyPrevalence, paired = TRUE)
```

```
## Warning in wilcox.test.default(Data$FemaleMalignancyPrevalence,  
## Data$MaleMalignancyPrevalence, : cannot compute exact p-value with ties
```

```
## Warning in wilcox.test.default(Data$FemaleMalignancyPrevalence,  
## Data$MaleMalignancyPrevalence, : cannot compute exact p-value with zeroes
```

```
wilcox_p <- signif(wil_mal$p.value, 3)
```

Violin code

```
mam_vio_mal <- ggplot(Data_long_mal, aes(x = Sex, y = 100 * Prevalence, fill = Sex)) +  
  geom_violin(adjust = 1) +  
  theme(  
    legend.title = element_text(size = 21, face = "bold"),  
    legend.text = element_text(size = 20),  
    plot.title = element_text(size = 20, face = "bold"),  
    axis.text = element_text(size = 16),  
    axis.title = element_text(size = 16)  
  ) +  
  labs(  
    title = "Malignancy Prevalence by Sex in Ruminants",  
    subtitle = paste0("Wilcoxon Signed-Rank Test: P-Value = ", wilcox_p), # Update this as needed  
    fill = 'Sex'  
  ) +  
  ylab("Malignancy Prevalence (%)") +  
  xlab("Sex")
```

Add some mean bars and clean up the background with cowplot

```
mam_vio_mal + geom_jitter(shape=16, position=position_jitter(0.2))+  
stat_summary(fun=mean, geom="crossbar", size=0.7) +  
scale_fill_manual(values = c("#FFC627", "#8C1D40", "#F0E442")) +  
theme_cowplot(12)+  
  theme(legend.position = "none")
```

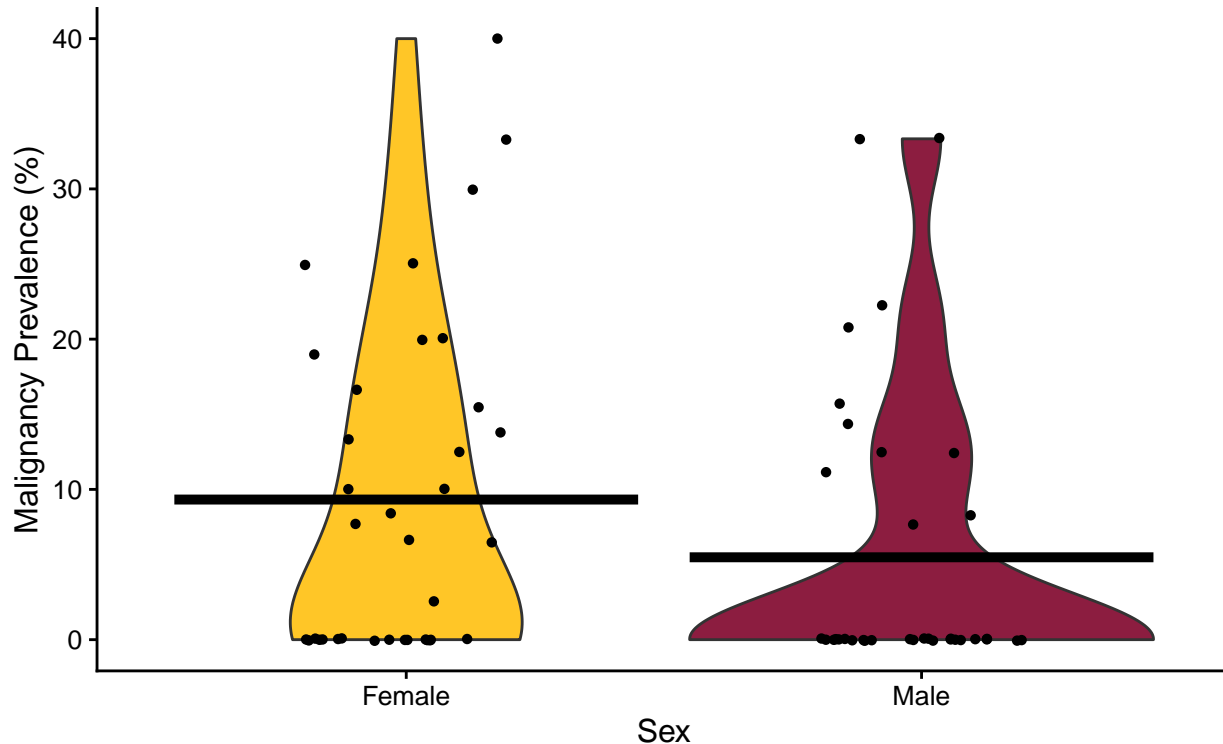
```
## Warning: Removed 5 rows containing non-finite outside the scale range  
## (`stat_ydensity()`).
```

```
## Warning: Removed 5 rows containing non-finite outside the scale range  
## (`stat_summary()`).
```

```
## Warning: Removed 5 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

Malignancy Prevalence by Sex in Ruminants

Wilcoxon Signed-Rank Test: P-Value = 0.11



```
ggsave("RuminantSexMalViolins.png", width = 10, height = 6, units = "in")
```

```
## Warning: Removed 5 rows containing non-finite outside the scale range
## (`stat_ydensity()`).
```

```
## Warning: Removed 5 rows containing non-finite outside the scale range
## (`stat_summary()`).
```

```
## Warning: Removed 5 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

Overall Violins

```
Data_overall_long <- pivot_longer(Data,
                                   cols = c(NeoplasiaPrevalence, MalignancyPrevalence),
                                   names_to = "Type",
                                   values_to = "Prevalence")
```

```
# Clean up Type labels if needed
```

```
Data_overall_long$Type <- gsub("Prevalence", "", Data_overall_long$Type)
```

```
wil_overall <- wilcox.test(Data$NeoplasiaPrevalence,
                           Data$MalignancyPrevalence,
                           paired = TRUE)
```

```
## Warning in wilcox.test.default(Data$NeoplasiaPrevalence,
## Data$MalignancyPrevalence, : cannot compute exact p-value with zeroes
```

```
wil_overall_p <- signif(wil_overall$p.value, 3)

overall_vio <- ggplot(Data_overall_long, aes(x = Type, y = 100 * Prevalence, fill = Type)) +
  geom_violin(adjust = 1, trim = FALSE) +
  geom_jitter(shape = 16, position = position_jitter(0.1), size = 1.5, alpha = 0.7) +
  stat_summary(fun = mean, geom = "crossbar", width = 0.3, size = 0.7, color = "black") +
  scale_fill_manual(values = c("#FFC627", "#F0E442")) + # Customize colors if you like
  theme_cowplot(12) +
  theme(
    legend.position = "none",
    plot.title = element_text(size = 20, face = "bold"),
    axis.text = element_text(size = 14),
    axis.title = element_text(size = 16)
  ) +
  labs(
    title = "Comparison of Neoplasia and Malignancy Prevalence in Ruminants",
    subtitle = paste0("Wilcoxon Signed-Rank Test: P-Value = ", wil_overall_p),
    y = "Prevalence (%)",
    x = ""
  )
)
```

Add some mean bars and clean up the background with cowplot

```
overall_vio + geom_jitter(shape=16, position=position_jitter(0.2))+
stat_summary(fun=mean, geom="crossbar", size=0.7) +
scale_fill_manual(values = c("#FFC627", "#8C1D40", "#F0E442")) +
theme_cowplot(12)+
  theme(legend.position = "none")+ylim(0, NA)
```

```
## Scale for fill is already present.
```

```
## Adding another scale for fill, which will replace the existing scale.
```

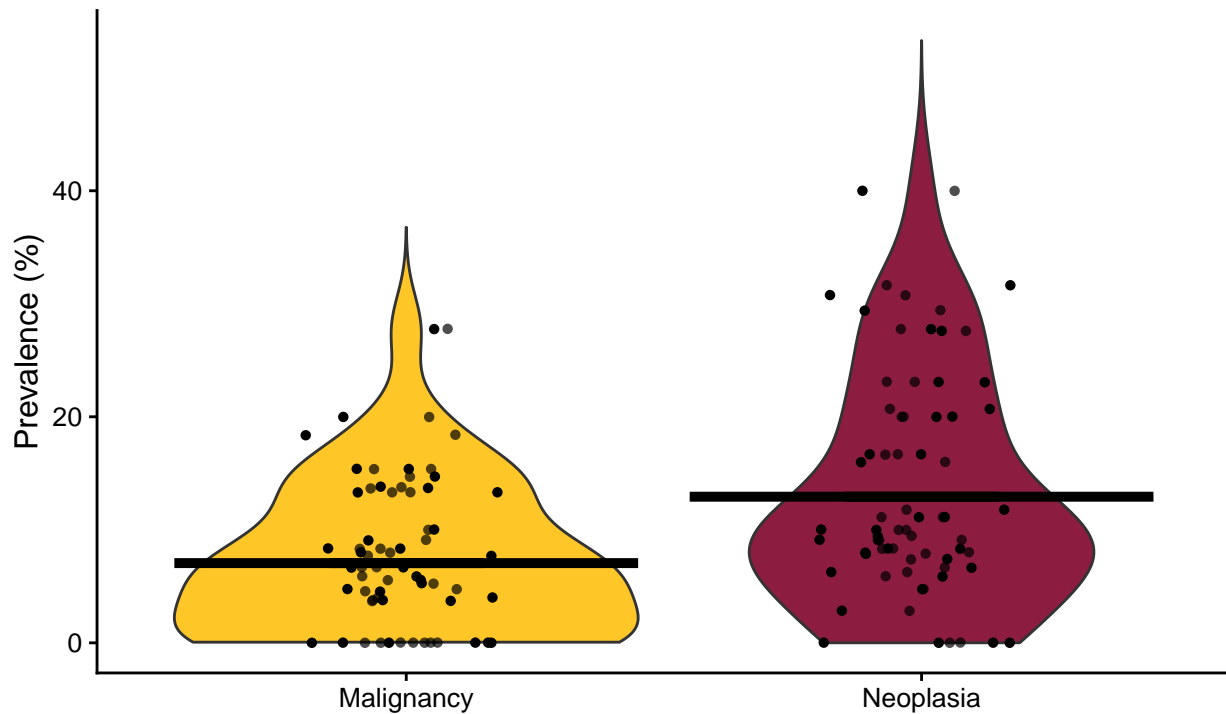
```
## Warning: Removed 203 rows containing missing values or values outside the scale range
## (`geom_violin()`).
```

```
## Warning: Removed 9 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

```
## Warning: Removed 8 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

Comparison of Neoplasia and Malignancy Prevalence in Ruminar

Wilcoxon Signed-Rank Test: P-Value = 1.94e-05



```
ggsave("RuminantPrevViolins.png", width = 10, height = 6, units = "in")
```

```
## Warning: Removed 203 rows containing missing values or values outside the scale range  
## (`geom_violin()`).
```

```
## Warning: Removed 7 rows containing missing values or values outside the scale range  
## (`geom_point()`).
```

```
## Warning: Removed 5 rows containing missing values or values outside the scale range  
## (`geom_point()`).
```