A1_Jewitt_20096665

Ally Jewitt

10/01/2023

Load libraries

```
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(tidyr)
```

Set theme

```
source("./ThemePub.R")
```

Run R scripts

This script creates 4 vectors for Organism, Limb_width, Limb_Length, and Observer, and arranges them into a data.frame.

```
MyData<-data.frame(Organism=sample(c("Ailuropoda melanoleuca", "Panthera tigris", "El
ephas maximus", "Gorilla beringei", "Pongo pygmaeus"), 100, replace=T), Limb_width=rn
orm(100, 40, 5), Limb_length=rnorm(100, 70, 5), Observer=sample(c("Dr. Jane Goodall",
"Dian Fossey", "Dr. Biruté Galdikas"), 100, replace=T))
```

This script adds a Volume column to the measurements.csv file using the equation for the volume of a cylinder (πr^2h).

```
MyData<-MyData %>%
  mutate(Volume=pi*((MyData$Limb_width/2)^2)*MyData$Limb_length)
```

Link to GitHub

GitHub repository (https://github.com/allyjewitt/BIOL432_Assignment1)

1. Load the revised measurements.csv

```
head(MyData)
```

```
##                   Organism Limb_width Limb_length            Observer    Volume
## 1        Pongo pygmaeus     34.61060    71.59704 Dr. Biruté Galdikas  67360.16
## 2 Ailuropoda melanoleuca     33.08102    65.58121    Dr. Jane Goodall  56367.28
## 3       Gorilla beringei     38.55947    76.77909         Dian Fossey  89659.23
## 4        Pongo pygmaeus     51.59535    72.80619 Dr. Biruté Galdikas 152222.67
## 5       Gorilla beringei     35.15736    71.32304    Dr. Jane Goodall  69239.25
## 6 Ailuropoda melanoleuca     39.02188    77.12242         Dian Fossey  92233.11
```

2. Use dplyr to sort the data by Organism, then by Observer, then by Volume

```
MyData <- MyData %>%
  arrange(Organism, Observer, Volume)
head(MyData)
```

```
##                   Organism Limb_width Limb_length    Observer    Volume
## 1 Ailuropoda melanoleuca     34.22033    69.02229 Dian Fossey 63481.57
## 2 Ailuropoda melanoleuca     38.78348    66.05577 Dian Fossey 78035.85
## 3 Ailuropoda melanoleuca     41.25293    60.14268 Dian Fossey 80386.33
## 4 Ailuropoda melanoleuca     37.09819    79.36765 Dian Fossey 85790.42
## 5 Ailuropoda melanoleuca     41.41126    68.17360 Dian Fossey 91821.23
## 6 Ailuropoda melanoleuca     39.02188    77.12242 Dian Fossey 92233.11
```

3. Create a table showing the average volume for each species

```
AvgVolume<-aggregate(MyData$Volume, list(MyData$Organism), mean)
AvgVolume %>%
  rename(
    Mean = x,
    Organism = Group.1
  )
```

```
##                   Organism     Mean
## 1 Ailuropoda melanoleuca 90059.97
## 2       Elephas maximus 92748.42
## 3       Gorilla beringei 94861.35
## 4       Panthera tigris 90090.18
## 5       Pongo pygmaeus 90711.73
```

4. Create a table showing the number of observations (i.e. rows) for each combination of species and observer.

```
NumObservations<- MyData %>% group_by(Organism, Observer) %>%
  summarize(NumObservations = n())
```

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

```
NumObservations
```

```
## # A tibble: 15 × 3
## # Groups:   Organism [5]
##    Organism              Observer            NumObservations
##    <chr>                 <chr>                         <int>
##  1 Ailuropoda melanoleuca Dian Fossey                      8
##  2 Ailuropoda melanoleuca Dr. Biruté Galdikas             12
##  3 Ailuropoda melanoleuca Dr. Jane Goodall                 6
##  4 Elephas maximus        Dian Fossey                      5
##  5 Elephas maximus        Dr. Biruté Galdikas             11
##  6 Elephas maximus        Dr. Jane Goodall                 7
##  7 Gorilla beringei       Dian Fossey                     11
##  8 Gorilla beringei       Dr. Biruté Galdikas              5
##  9 Gorilla beringei       Dr. Jane Goodall                 7
## 10 Panthera tigris        Dian Fossey                      1
## 11 Panthera tigris        Dr. Biruté Galdikas              1
## 12 Panthera tigris        Dr. Jane Goodall                 3
## 13 Pongo pygmaeus         Dian Fossey                      5
## 14 Pongo pygmaeus         Dr. Biruté Galdikas             10
## 15 Pongo pygmaeus         Dr. Jane Goodall                 8
```

5. Create a single box plot (or similar) to compare key features of the distributions of volumes for each species.

```
ylab <- expression(Volume ~ cm^3)
ggplot(aes(x=Organism, y=Volume), data=MyData) + geom_boxplot() + ylab(ylab) + theme_
pub()
```
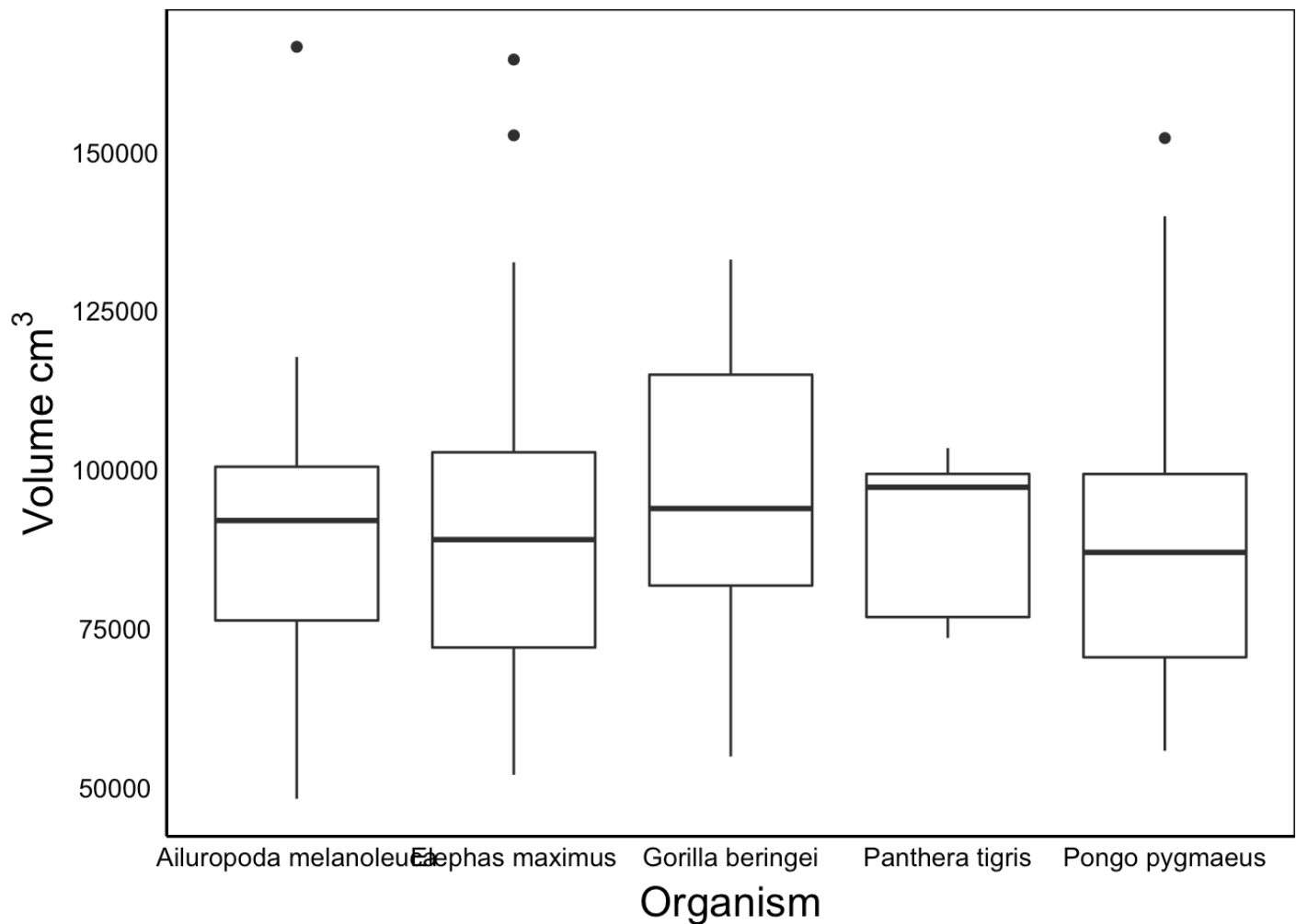
Figure 1. Box plot comparing the key features of the distribution volumes of each species. The boxes represent the interquartile range, and the horizontal line represents the median value. Upper and lower tails represent the maximum and minimum values. The dots represent any outliers. Each species has a similar median value for limb volume.

6. Create a multi-panel plot showing frequency histograms for each species.

```
xlab <- expression(Volume ~ cm^3)
ggplot(MyData, aes(Volume)) + geom_histogram(bins=7) + facet_wrap(.~Organism) + xlab(
xlab) + ylab("Count") + theme_pub()
```
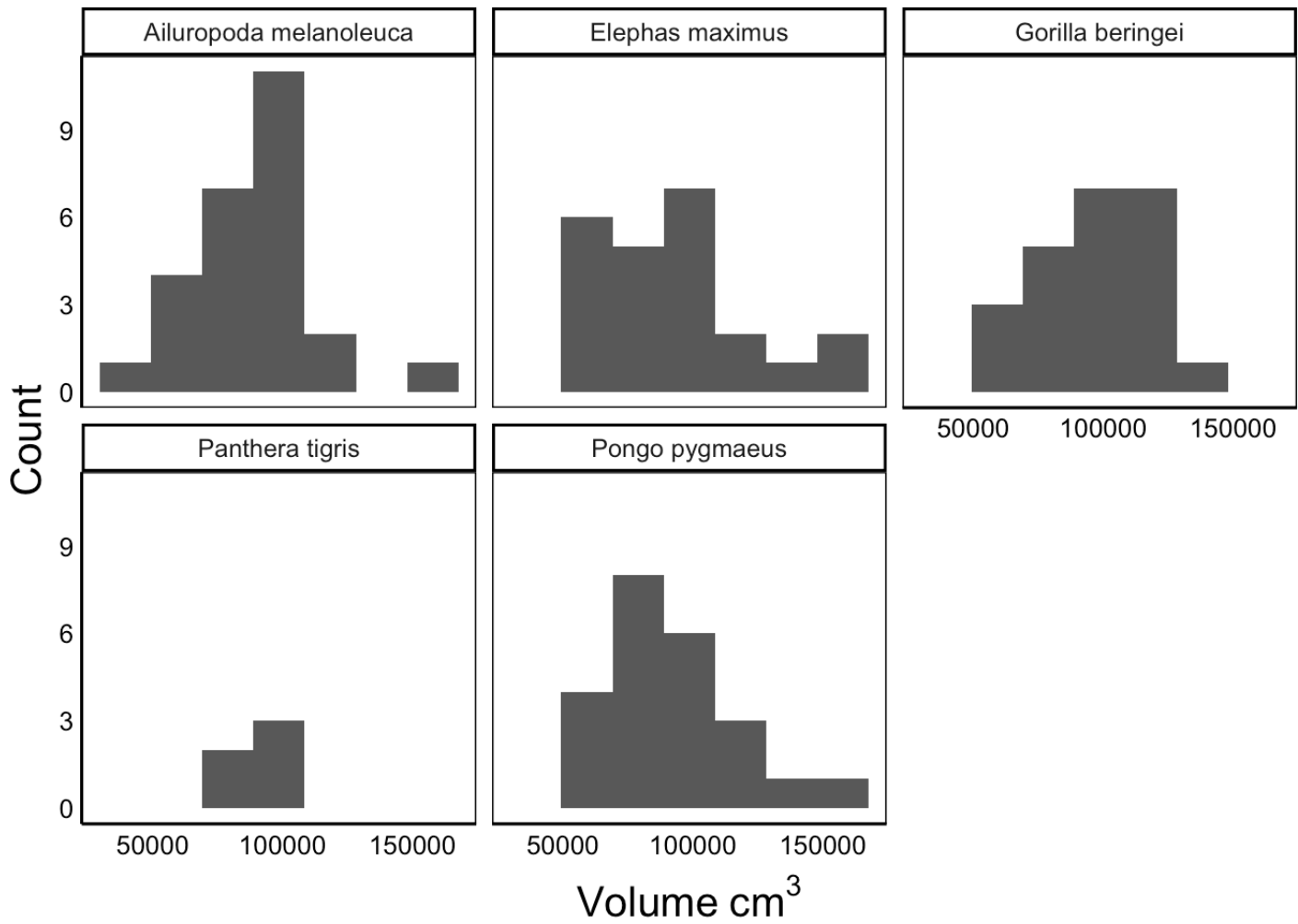
Figure 2. Multi-panel plot showing the frequency histograms for the limb volume for each species, *Ailuropoda melanoleuca*, *Elephas maximus*, *Gorilla beringei*, *Panthera tigris*, and *Pongo pygamaeus*. Each histogram shows a similar distribution for limb volume with the highest frequency at the mean value.