A1_Yip20168228.PDF

Marco Yip

2023-01-11

#Github link #https://github.com/marcoyip21/BIOL-432_Assignment1

#Run previous R scripts

```
source("./dataGenerato.R")
source("./volumeEstimato.R")
```

```
##
## 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
```

#Q1, Loading library and revised measurements.csv

```
library(dplyr)
library(ggplot2)
InData = read.csv("./measurements.csv")
```

#Q2 Sorting the data by species, then by observer, then by limb volume

```
SpeciesData = InData %>%
  arrange(Species)
ObserverData = InData %>%
  arrange(Observer)
VolumeData = InData %>%
  arrange(volume)
```

#Q3 Table showing average volume for each species

```
NewData = SpeciesData %>%
  group_by(Species)%>%
  summarize(mean(volume))
```

#Q4 Table showing the number of observations for each combination of species and observer

```
DataCount=count(SpeciesData,Species, Observer)
```

#Q5 Box plot to compare key features of the distributions of volumes for each species

```
ggplot(data = SpeciesData, aes(x = Species, y = volume))+
  geom_boxplot()
```
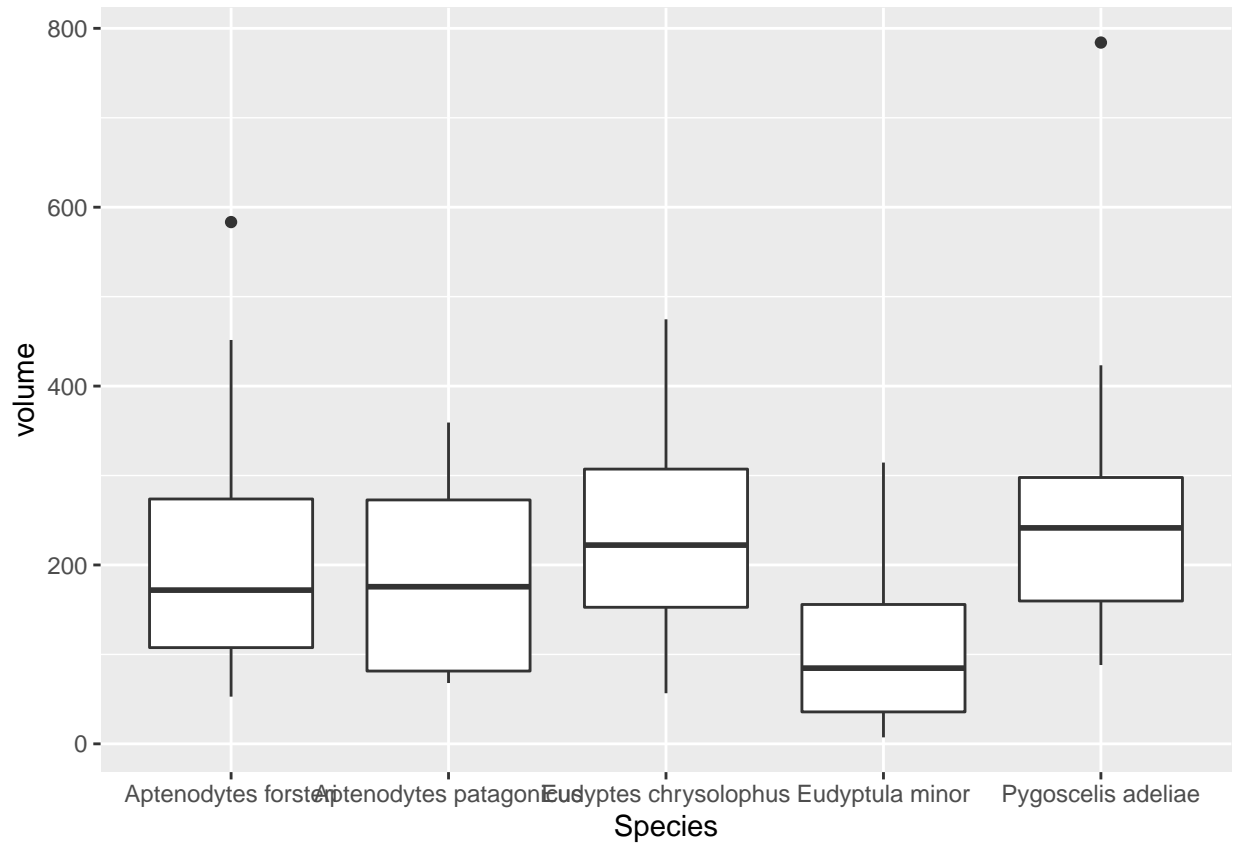

#Q6 Multi-panel plot showing frequency histograms showing the distribution of limb volume for each species.

```
ggplot(SpeciesData, aes(x=volume, fill=Species)) +
  geom_histogram() +
  facet_wrap(vars(Species)) +
  ylab("Number of species")+ xlab("Volume of limb")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```