

BIOL432_Assignment1

SiruiZHAO

2023-01-11

my GitHub repository (https://github.com/siruiZHAO/BIOL432_Assignment1)

First, load library

```
library(ggplot2)
library(dplyr)
```

Then, we need to source two R scripts.

1). dataGenerato.R generate 4 variables, organism, limb width, limb length and observer, and combine them into a data frame.

2). volumeEstimato.R add one variable called limb volume into the data frame. The limb volume calculated from the $v = \pi * (\text{limb_width}/2)^2 * \text{limb_length}$

```
source("dataGenerato.R", local = knitr::knit_global())
source("volumeEstimato.R", local = knitr::knit_global())
```

```
## 'data.frame':   100 obs. of  4 variables:
## $ organism    : chr  "\"Bos gaurus\"" "\"Antelope cervicapra\"" "\"Gazella bennett
ii\"" "\"Boselaphus tragocamelus\"" ...
## $ Limb_width  : num  3.01 2.78 3.12 3.47 1.2 ...
## $ Limb_length: num  17.6 55.6 51.7 46.7 57.7 ...
## $ Observer    : chr  "\"Charles Darwin\"" "\"Charles Darwin\"" "\"Charles Darwin\""
" "\"Charles Darwin\"" ...
```

1. Load the revised measurements.csv and library

```
mydata <- read.csv("measurements.csv")
```

2. Use dplyr to sort the data by species, then by observer, then by limb volume

```
mydata <- mydata %>% #use pipe in dplyr
  arrange(organism, Observer, limb_volume) # use arrange to sort column in order
head(mydata)
```

```
##           organism Limb_width Limb_length      Observer limb_volume
## 1 "Antelope cervicapra"   3.062553   12.31966 "Charles Darwin"    90.75192
## 2 "Antelope cervicapra"   2.779865   24.60835 "Charles Darwin"   149.35499
## 3 "Antelope cervicapra"   4.204600   22.03541 "Charles Darwin"   305.95692
## 4 "Antelope cervicapra"   2.784150   55.59689 "Charles Darwin"   338.47434
## 5 "Antelope cervicapra"   3.606322   37.04045 "Charles Darwin"   378.35125
## 6 "Antelope cervicapra"   3.700756   62.28123 "Charles Darwin"   669.92770
```

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

```
avg_volume <- mydata %>%      #select data frame
  group_by(organism) %>%      #group rows by column values in the DataFrame
  summarise(mean_volum = mean(limb_volume)) #calculate mean of limb volume
avg_volume
```

```
## # A tibble: 5 × 2
##   organism          mean_volum
##   <chr>              <dbl>
## 1 "\"Antelope cervicapra\""      386.
## 2 "\"Bos gaurus\""              267.
## 3 "\"Boselaphus tragocamelus\""  226.
## 4 "\"Canis lupus\""             306.
## 5 "\"Gazella bennettii\""       303.
```

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

```
observation <- mydata %>%      #select data frame
  group_by(organism, Observer) %>% #group rows by two column values
  summarise(observations = n(), .group = "drop") %>% #count frequency
  as.data.frame()
```

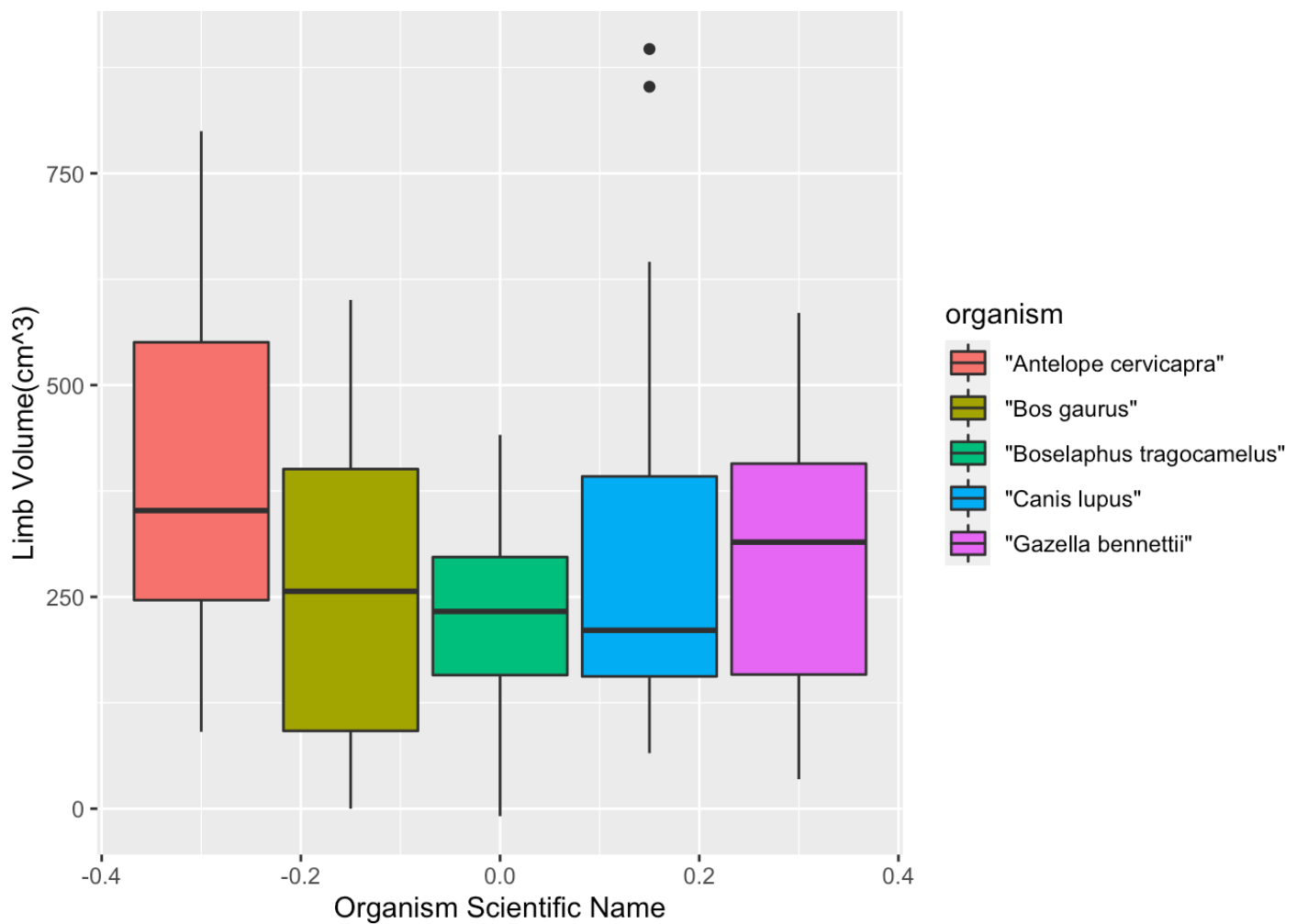
```
## `summarise()` has grouped output by 'organism'. You can override using the
## `.groups` argument.
```

```
observation
```

##	organism	Observer	observations	.group
## 1	"Antelope cervicapra"	"Charles Darwin"	7	drop
## 2	"Antelope cervicapra"	"Gregor Mendel"	6	drop
## 3	"Antelope cervicapra"	"Robert Hooke"	7	drop
## 4	"Bos gaurus"	"Charles Darwin"	7	drop
## 5	"Bos gaurus"	"Gregor Mendel"	7	drop
## 6	"Bos gaurus"	"Robert Hooke"	6	drop
## 7	"Boselaphus tragocamelus"	"Charles Darwin"	6	drop
## 8	"Boselaphus tragocamelus"	"Gregor Mendel"	7	drop
## 9	"Boselaphus tragocamelus"	"Robert Hooke"	7	drop
## 10	"Canis lupus"	"Charles Darwin"	6	drop
## 11	"Canis lupus"	"Gregor Mendel"	7	drop
## 12	"Canis lupus"	"Robert Hooke"	7	drop
## 13	"Gazella bennettii"	"Charles Darwin"	7	drop
## 14	"Gazella bennettii"	"Gregor Mendel"	6	drop
## 15	"Gazella bennettii"	"Robert Hooke"	7	drop

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

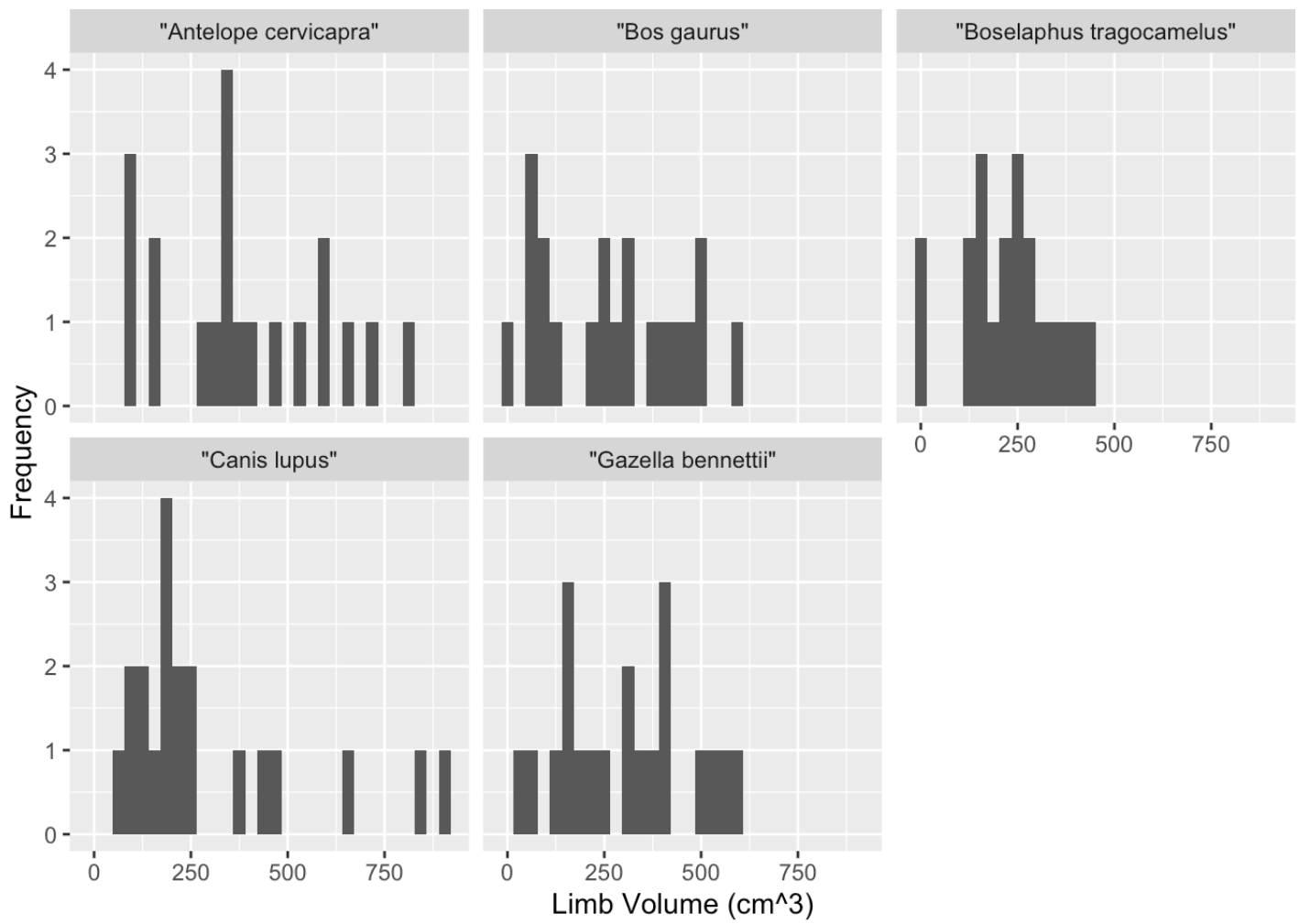
```
#use ggplot2 to create box plot
plot1 <- ggplot(mydata, aes(y = limb_volume, fill = organism)) + geom_boxplot() + lab
s(x = "Organism Scientific Name", y = "Limb Volume(cm^3)")
plot1
```



6. Create a multi-panel plot showing frequency histograms showing the distribution of limb volume for each species. Note that this may require you to add code to your R markdown file to change the data from the wide to long format.

```
#use facet_wrap() in ggplot2 to create multi-panel plot
ggplot(mydata, aes(x = limb_volume)) +
  geom_histogram() +
  facet_wrap(vars(organism)) + labs(x = "Limb Volume (cm^3)", y = "Frequency")
```

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



Thanks!