

BIOL432 Assignment 5

SiruiZHAO

2023-02-08

Part I: Data Science

my GitHub repository (https://github.com/siruiZHAO/BIOL432_Assignment_5.git)

```
#load library
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
```

1. Import your measurements.csv file from the Week 1 Assignment into an object called MData

```
#Import data
MData <- read.csv("measurements.csv")

head(MData)
```

```
##           organism Limb_width Limb_length Observer limb_volume
## 1      "Bos gaurus"   3.362429   38.03622 "Charles Darwin"    337.7486
## 2 "Antelope cervicapra" 3.381003   32.62935 "Charles Darwin"    292.9473
## 3   "Gazella bennettii" 3.848400   25.32446 "Charles Darwin"    294.5714
## 4 "Boselaphus tragocamelus" 2.725810   42.00863 "Charles Darwin"    245.1431
## 5      "Canis lupus"   4.009360   22.21913 "Charles Darwin"    280.5221
## 6      "Bos gaurus"   3.526292   28.43048 "Charles Darwin"    277.6582
```

```
#check data structure
str(MData)
```

```
## 'data.frame':   100 obs. of  5 variables:
## $ organism     : chr  "\"Bos gaurus\"" "\"Antelope cervicapra\"" "\"Gazella bennett
ii\"" "\"Boselaphus tragocamelus\"" ...
## $ Limb_width   : num  3.36 3.38 3.85 2.73 4.01 ...
## $ Limb_length  : num  38 32.6 25.3 42 22.2 ...
## $ Observer     : chr  "\"Charles Darwin\"" "\"Charles Darwin\"" "\"Charles Darwin\""
" "\"Charles Darwin\"" ...
## $ limb_volume  : num  338 293 295 245 281 ...
```

2. Use a dplyr command AND regular expression(s) to add a new column called Sp that shortens the genus name in the Organism column to only the first letter and period. For example: Homo sapiens in the original Organism column becomes H. sapiens in the Sp column.

```
#use mutate() create new column and gsub() in regular expressions to shortens the let
ters after first capital letters
MData <- MData %>%
  mutate(Sp = gsub("([A-Z]).* (.*)", "\\1. \\2", organism))

#check data
head(MData)
```

```
##           organism Limb_width Limb_length      Observer limb_volume
## 1      "Bos gaurus"   3.362429   38.03622 "Charles Darwin"    337.7486
## 2    "Antelope cervicapra" 3.381003   32.62935 "Charles Darwin"    292.9473
## 3    "Gazella bennettii"  3.848400   25.32446 "Charles Darwin"    294.5714
## 4 "Boselaphus tragocamelus" 2.725810   42.00863 "Charles Darwin"    245.1431
## 5      "Canis lupus"   4.009360   22.21913 "Charles Darwin"    280.5221
## 6      "Bos gaurus"   3.526292   28.43048 "Charles Darwin"    277.6582
##           Sp
## 1      "B. gaurus"
## 2      "A. cervicapra"
## 3      "G. bennettii"
## 4 "B. tragocamelus"
## 5      "C. lupus"
## 6      "B. gaurus"
```

3. Use a dplyr command to create a new dataset BUT with Sp column instead of the Organism column, moved to the beginning in place of the original Organism column.

```
#create new dataset only with Sp
#move Sp column to the beginning
NewData <- MData %>%
  select(Sp, Limb_width, Limb_length, limb_volume, Observer)

#check data
head(NewData)
```

##		Sp	Limb_width	Limb_length	limb_volume	Observer
## 1		"B. gaurus"	3.362429	38.03622	337.7486	"Charles Darwin"
## 2		"A. cervicapra"	3.381003	32.62935	292.9473	"Charles Darwin"
## 3		"G. bennettii"	3.848400	25.32446	294.5714	"Charles Darwin"
## 4		"B. tragocamelus"	2.725810	42.00863	245.1431	"Charles Darwin"
## 5		"C. lupus"	4.009360	22.21913	280.5221	"Charles Darwin"
## 6		"B. gaurus"	3.526292	28.43048	277.6582	"Charles Darwin"

4. Save the dataset as a new file called MeasShort.csv

```
#save as csv file
write.csv(NewData, "MeasShort.csv")
```

Thanks!