

Assignment 5

Elliot Gavrin

2023-02-06

https://github.com/19eag3/BIOL432_Assignment5

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

```
MData <-read.csv("C:/Users/egavr/OneDrive/Documents/BIOL432/a1/measurements.csv")
head(MData)
```

##	X.1	X	Organisms	Limb_width	Limb_length	Observer	Limb_volume
## 1	1	1	Cyanocitta cristata	2.302585	1.609438	William P.	3.705868
## 2	2	2	Anas acuta	2.484907	2.833213	Cameron N.	7.040271
## 3	3	3	Cyanocitta cristata	2.772589	2.302585	William P.	6.384121
## 4	4	4	Pica serica	2.564949	2.564949	Elliot G.	6.578965
## 5	5	5	Cyanocitta cristata	2.564949	2.484907	Elliot G.	6.373660
## 6	6	6	Columba livia	1.609438	2.302585	William P.	3.705868

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. HINT: You can think of the regular expression commands like sub, grep, grepl and gsub as 'functions' that you can use with dplyr commands just like mean(), sd(), and desc() in the Data Science tutorial.

```
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
```

```
MData$Sp <-sub("(\\w)\\w+", "\\1.", MData$Organisms)
head(MData)
```

##	X.1	X	Organisms	Limb_width	Limb_length	Observer	Limb_volume
## 1	1	1	Cyanocitta cristata	2.302585	1.609438	William P.	3.705868
## 2	2	2	Anas acuta	2.484907	2.833213	Cameron N.	7.040271
## 3	3	3	Cyanocitta cristata	2.772589	2.302585	William P.	6.384121
## 4	4	4	Pica serica	2.564949	2.564949	Elliot G.	6.578965
## 5	5	5	Cyanocitta cristata	2.564949	2.484907	Elliot G.	6.373660
## 6	6	6	Columba livia	1.609438	2.302585	William P.	3.705868
##			Sp				
## 1			C. cristata				
## 2			A. acuta				
## 3			C. cristata				
## 4			P. serica				
## 5			C. cristata				
## 6			C. livia				

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.

```
cols <- names(MData)
cols <- c(cols[1:2], "Sp", cols[3:(length(cols) - 1)])
SPData <- MData[, cols]
SPData <-SPData %>%
  select(everything(), -Organisms)
head(SPData)
```

##	X.1	X	Sp	Limb_width	Limb_length	Observer	Limb_volume
## 1	1	1	C. cristata	2.302585	1.609438	William P.	3.705868
## 2	2	2	A. acuta	2.484907	2.833213	Cameron N.	7.040271
## 3	3	3	C. cristata	2.772589	2.302585	William P.	6.384121
## 4	4	4	P. serica	2.564949	2.564949	Elliot G.	6.578965
## 5	5	5	C. cristata	2.564949	2.484907	Elliot G.	6.373660
## 6	6	6	C. livia	1.609438	2.302585	William P.	3.705868

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

```
write.csv(SPData,"MeasShort.csv")
```