

A4_WU_CHENYANG

Chenyang Wu

2022/2/2

R-egex Assignment

Project Info

GitHub user name: Wuris

Date: 2022/2/2

GitHub Link: https://github.com/Wuris/Biol432_Assignments.git

Part I: Data Science

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

```
setwd("E:/Undergraduate/4th Fourth year/BIOL 432/Week 3/Biol432_Assignment3/Assignment 4/Input")
MData <- read.csv("measurement.csv")
```

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) #Load the package we will use
```

```
##
##   'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
MData <- MData %>%
  mutate(Sp = sub("(\\w)\\w+", "\\1.", MData$Organism)) #Create a new column called Sp, with the abbrev
```

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.

```
MData_new <- MData %>%  
  select(Sp, Limb.width, Limb.length, Observer) # Create the new data frame
```

Save the dataset as a new file called MeasShort.csv

```
setwd("E:/Undergraduate/4th Fourth year/BIOL 432/Week 3/Biol432_Assignment3/Assignment 4/Output")  
write.csv(MData_new, "MeasShort.csv", row.names = FALSE) # Save the new data frame
```