A4_Tianyi_Zuo

Tianyi Zuo

2022/2/2

GitHub username: Lydia12138

Date: 2022-02-02
Repo:https://github.com/Lydia12138/Assignment3_TianyiZuo (https://github.com/Lydia12138/Assignment3_TianyiZuo)

```
library(dplyr)#package loading
```

```
## Warning: package 'dplyr' was built under R version 3.6.2
```

```
##
## 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
```

```
MyData<-read.csv("./InputData/measurements.csv")#loading the data
str(MyData)#check the structure of the Data
```

```
## 'data.frame':    49 obs. of  4 variables:
##  $ organism   : Factor w/ 13 levels "Acinonyx jubatus",..: 9 10 3 10 4 11 6 9 8 2
...
##  $ Limb.width : num  3.45 0.04 2.42 0.02 2.63 2.56 0.52 1.24 8.54 9.67 ...
##  $ Limb.length: num  21.2 102 0.6 9 10 34 1.4 12 23 34 ...
##  $ Observer   : Factor w/ 17 levels "Abby","Amy Y",..: 12 15 16 15 2 2 16 12 17 9
...
```

```
head(MyData)
```

```
##              organism Limb.width Limb.length   Observer
## 1     Panthera Tigris       3.45        21.2     Kelvin
## 2   Passer virginanus       0.04       102.0      Sally
## 3    Corvus splendens       2.42         0.6 Victoria X
## 4   Passer virginanus       0.02         9.0      Sally
## 5    Felis domesticus       2.63        10.0      Amy Y
## 6 Psittacula eupatria       2.56        34.0      Amy Y
```

```
class(MyData)
```

```
## [1] "data.frame"
```

```
dim(MyData)
```

```
## [1] 49  4
```

```
tail(MyData)
```

```
##               organism Limb.width Limb.length Observer
## 44       Rattus rattus       8.90         0.6     Ella
## 45       Rattus rattus       1.50        21.0     Abby
## 46   Passer virginanus       3.56        35.0 Oliver R
## 47    Acinonyx jubatus       1.56        23.0 Oliver R
## 48 Camelus dromedaries       0.31        23.0    Cindy
## 49   lxodes scapularis       2.31         5.1   Hans P
```

```
summary(MyData)
```

```
##                   organism    Limb.width     Limb.length
##  Camelus dromedaries  : 5   Min.   :0.010   Min.   :  0.20
##  lxodes scapularis    : 5   1st Qu.:1.230   1st Qu.:  2.50
##  Oryctolagus cuniculus: 5   Median :2.350   Median :  9.00
##  Passer virginanus    : 5   Mean   :2.963   Mean   : 15.74
##  Psittacula eupatria  : 5   3rd Qu.:3.670   3rd Qu.: 23.00
##  Felis domesticus     : 4   Max.   :9.670   Max.   :102.00
##  (Other)              :20
##        Observer
##  Oliver R  : 7
##  Wuris     : 7
##  Cindy     : 5
##  Victoria X: 5
##  Jayce M   : 4
##  Kelvin    : 4
##  (Other)   :17
```

```
# Use a dplyr command and regular expression to add a new column called Sp
# Here I shortens the genus name in the Organism column to only the first letter and
 period
MyData <- mutate(MyData, Sp = sub("(\\w)\\w+ (.*)", "\\1. \\2", MyData$organism))
head(MyData)
```

```
##              organism Limb.width Limb.length   Observer            Sp
## 1     Panthera Tigris       3.45        21.2     Kelvin     P. Tigris
## 2   Passer virginanus       0.04       102.0      Sally P. virginanus
## 3    Corvus splendens       2.42         0.6 Victoria X  C. splendens
## 4   Passer virginanus       0.02         9.0      Sally P. virginanus
## 5    Felis domesticus       2.63        10.0      Amy Y F. domesticus
## 6 Psittacula eupatria       2.56        34.0      Amy Y   P. eupatria
```

```
# creating new data set using dplyr which replace organism with Sp.
NewData <- MyData %>% select(Sp, Limb.width, Limb.length, Observer)
head(NewData)
```

```
##              Sp Limb.width Limb.length   Observer
## 1     P. Tigris       3.45        21.2     Kelvin
## 2 P. virginanus       0.04       102.0      Sally
## 3  C. splendens       2.42         0.6 Victoria X
## 4 P. virginanus       0.02         9.0      Sally
## 5 F. domesticus       2.63        10.0      Amy Y
## 6   P. eupatria       2.56        34.0      Amy Y
```

```
# Output a new file named "MeasShort.csv"
write.csv(NewData, "Output/MeasShort.csv")
```