Regex.Rmd

Roshael Chellappah (20103016)

02/02/2022

Project Set-Up

git hub username: RoshaelC

git hub repository: https://github.com/RoshaelC/r-eproducibility.git (https://github.com/RoshaelC/r-eproducibility.git)

```
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
```

```
getwd()
```

```
## [1] "C:/Users/rosch/Documents/BIOL 432/Week 3 - R-eproducibility/r-eproducibility"
```

Part I: Data Science

1. Add new coloumn called "Sp" that shortens the genus name in the "Organism" column to only the first letter and period (i.e., Homo sapiens -> H. sapiens)

```
MData <- mutate(
  read.csv("C:/Users/rosch/Documents/BIOL 432/Week 1 - Introduction/A1_CHELLAPPAH_ROSHAEL.csv"),
  Sp = gsub("[a-z]\\w+\\s", "\\. ", Organism))

# gsub() used to find and replace data
# [a-z] identifies all lower case letters (disregards upper case)
# \\w+ identifies strings of any character that is a letter or digit
# \\s identifies spaces
# \\. replaces the identified string characters with a "."
```

2. Create a new dataset BUT with "Sp" column where "Organism" column is, and remove "Organism" column

```
MData <- transmute(MData, Sp, Limb.width, Limb.length, Observer)
```

3. Save the dataset as a new file called "MeasShort.csv"

```
write.csv(MData, "./MeasShort.csv")
```

Part II: Bash

See .txt file "A4_Chellappah_Roshael.txt"