Processing Strings in R

7 May 2019

Excercise

The local zoo has made an inventory of its animal stock.[1] However, the zoo keepers did a messy job with writing up totals as you can see below. You are hired to clean up the mess using *R*.

```
library(stringr)
library(quanteda)
```

```
## Package version: 1.3.0
```

```
## Parallel computing: 2 of 4 threads used.
```

```
## See https://quanteda.io for tutorials and examples.
```

```
##
## Attaching package: 'quanteda'
```

```
## The following object is masked from 'package:utils':
##
##     View
```

```
zoo <- c("bear x2", "Ostric7", "platypus x60", "x7 Eliphant", "x16 conDOR")
```

Use the functions in the `stringr` to clean up the string, taking out typos, etc. Generate an *R* dataframe with the following variables: *Species* (character), *Number* (numeric).

```
zoo <- c("bear x2", "Ostric7", "platypus x60", "x7 Eliphant", "x16 conDOR")

zoo <- str_to_lower(zoo)
zoo <- str_replace(zoo, "x", "")
zoo[2] <- str_replace(zoo[2], "7", "h 7")
zoo[4] <- str_replace(zoo[4], "eliphant", "elephant")
animal <- str_extract(zoo, "[a-z]+")
str(animal)
```

```
##  chr [1:5] "bear" "ostrich" "platypus" "elephant" "condor"
```

```
number <- str_extract(zoo, "\\d+")
str(number)
```

```
##  chr [1:5] "2" "7" "60" "7" "16"
```

```
data <- data.frame(animal = as.character(animal), number = as.numeric(number), stringsAsFactors = FALSE)
print(data)
```

```
##     animal number
## 1     bear      2
## 2  ostrich      7
## 3 platypus     60
## 4 elephant      7
## 5   condor     16
```

[1] This exercise is based on an example from Automated Data Collection With R, by Munzert *et al* (2015).

```r
str(data)
```

```
## 'data.frame':    5 obs. of  2 variables:
##  $ animal: chr  "bear" "ostrich" "platypus" "elephant" ...
##  $ number: num  2 7 60 7 16
```

```r
#make a plot
library(ggplot2)

ggplot(data, aes(animal, number)) +
  geom_col() + theme_minimal()
```