Data-607 Week-3 Assignment

Student Name : Sachid Deshmukh

Date : 09/16/2018

- GitHub Location for rmd file
- GitHub Location for SQL script
- GitHub Location for pdf file
- RPubs location of published file

```
library(stringr)
```

```
## Warning: package 'stringr' was built under R version 3.4.3
```

Question 3: Copy the introductory example. The vector name stores the extracted names.

```
raw.data <-"555-1239Moe Szyslak(636) 555-0113Burns, C. Montgomery555-6542Rev. Timothy Lovejoy555 8904Ne

name = unlist(str_extract_all(raw.data, "[[:alpha:] .,]{2,}"))
name
```

```
## [1] "Moe Szyslak"          "Burns, C. Montgomery" "Rev. Timothy Lovejoy"
## [4] "Ned Flanders"         "Simpson, Homer"       "Dr. Julius Hibbert"
```

* (a)

```
name.clean = str_replace_all(name,",", "")
name.fl = unlist(str_replace_all(name.clean, "[:alpha:]*[:punct:] ", ""))
name.fl
```

```
## [1] "Moe Szyslak"      "Burns Montgomery" "Timothy Lovejoy"
## [4] "Ned Flanders"     "Simpson Homer"    "Julius Hibbert"
```

* (b)

```
str_detect(name.clean, unlist(str_extract_all(name.clean, "[:alpha:]{2,}[:punct:]")))
```

```
## [1] FALSE FALSE  TRUE FALSE FALSE  TRUE
```

* (c)

```
str_detect(name.clean, unlist(str_extract_all(name.clean, "[A-Z][:punct:]{1}")))
```

```
## [1] FALSE  TRUE FALSE FALSE FALSE FALSE
```

Question 4: Describe the type of strings that conform to the following regular expression

* (a) [0-9]+\$: This regex match one or more digit follwed by $ sign

Example

```
str = "The value of this product is 100$"
amount = unlist(str_extract_all(str, "[0-9]+\\$"))
amount
```

```
## [1] "100$"
```

* (b) \b[a-z]{1,4}\b : This regex match any word in lower case whihc is 1 to 4 chars long

Example

```
str = "This is my batbinton bat"
bat = unlist(str_extract_all(str, "\\b[a-z]{1,4}\\b"))
bat
```

```
## [1] "is"  "my"  "bat"
```

* (c) .*?\.txt$: This regex match any word endting wiht .txt # Example

```
files = c("Program.R", "Program.cpp", "Program.txt")
txt = unlist(str_extract_all(files, ".*?\\.txt$"))
txt
```

```
## [1] "Program.txt"
```

* (d) \d{2}/\d{2}/\d{4} : This regex match date pattern mm/dd/yyyy

Example

```
str = "Today's date is 09/16/2018"
date = unlist(str_extract_all(str,"\\d{2}/\\d{2}/\\d{4}" ))
date
```

```
## [1] "09/16/2018"
```

* (e) <(.+?)>.+?</\1> : This regex matches any word wrapped inside <> and </> HTML tags

Example

```
str = "To print in bold write <b>Bold</b> in Html"
html = unlist(str_extract_all(str, "<(.+?)>.+?</\\1>"))
html
```

```
## [1] "<b>Bold</b>"
```

Question 9 : Secret Message

```
sm =  "clcopCow1zmstc0d87wnkig7OvdicpNuggvhryn92Gjuwczi8hqrfpRxs5Aj5dwpn0TanwoUwisdij7Lj8kpf03AT5Idr3co
decoded <- unlist(str_extract_all(sm, "[[:upper:].]{1,}"))
decoded <- str_replace_all(paste(decoded, collapse = ''), "[.]", " "); decoded
```

```
## [1] "CONGRATULATIONS YOU ARE A SUPERNERD"
```