STAT 527 HW 8

Satoshi Ido (ID: 34788706)

10 April 2023

```r
library("stringr")
# get the data from R file
# source("Purdue_STAT527/hw8.R")
```

```r
## Q1
v1 <- "The sun was shining on the sea. Sally saw the seagulls and smiled."
## Q2
v2 <- c("apple", "banana", "123", "456", "grape", "789", "orange", "melon")
## Q3
v3 <- "The cat sat on the mat. The cat slept well."
## Q4
v4 <- c("jane.doe@example.com", "john_smith@company.co.uk", "info@myblog.com",
        "james.bond@mi6.gov.uk", "support@amazon.com", "not_an_email_address",
        "user@localhost", "example.com", "myemail@com"
        )
## Q5
v5 <- c("1-800-555-1234", "555-123-4567", "(555) 555-7890", "123-456-7890",
        "555.555.1212", "1112223333", "123", "456-789-0123", "555/123-4567"
        )
```

Q1

```r
# extract all the words that start with "s" from the sentence
str_extract_all(v1, "(?<=\\b)s[^\\s]+")
```

```
## [[1]]
## [1] "sun"      "shining"  "sea."      "saw"       "seagulls" "smiled."
```

Q2

```r
# extract all the numbers from the given vector
x <- as.numeric(gsub(".*?([0-9]+).*", "\\1", v2))
x[!is.na(x)]
```

```
## [1] 123 456 789
```

Q3

```
# replace all the occurrences of "cat" with "dog" in the given sentence
gsub("cat", "dog", v3)
```

```
## [1] "The dog sat on the mat. The dog slept well."
```

Q4

```
# extract all the email addresses from the given vector
y <- str_match(v4, "^[[:alnum:]|[._%+].._]+@[[:alnum:]|[-.]-.]+\\.+[[:alpha:]]{2,}+$")
y[!is.na(y)]
```

```
## [1] "jane.doe@example.com"      "john_smith@company.co.uk"
## [3] "info@myblog.com"           "james.bond@mi6.gov.uk"
## [5] "support@amazon.com"
```

Q5

```
# extract all the phone numbers in the given vector
z <- str_match(v5, "\\(?\\d{3}\\)?[.-]? *\\d{3}[.-]? *[.-]?\\d{4}")
z[!is.na(z)]
```

```
## [1] "800-555-1234"    "555-123-4567"    "(555) 555-7890" "123-456-7890"
## [5] "555.555.1212"    "1112223333"      "456-789-0123"
```