String Processing (Section 3.1 of Unit4)

Chih Hui Wang

September 4, 2015

```
library(stringr)
```

1. Only the strings "cat", "at", and "t".

```
check_cat1 <- function(str){
  data.frame(string=str, check=str_detect(CAT, "^cat$|^at$|^t$"))
}
#test
CAT <- c("blackcat", "cat", "at", "atm", "t")
check_cat1(CAT)
```

```
    string check
1 blackcat FALSE
2      cat  TRUE
3       at  TRUE
4      atm FALSE
5        t  TRUE
```

2. The strings "cat", "caat", "caaat", etc.

```
check_cat2 <- function(str){
  data.frame(string=str, check=str_detect(str, "ca+t"))
}
#Test2
CAT2 <- c("bcabct", "cat", "caat", "caaat", "cabat",  "caaata")
check_cat2(CAT2)
```

```
  string check
1 bcabct FALSE
2    cat  TRUE
3   caat  TRUE
4  caaat  TRUE
5  cabat FALSE
6 caaata  TRUE
```

3. "dog", "Dog", "dOg", "doG", "DOg", etc. (the word dog in any combination of lower and upper case).

```
check_dog <- function(str){
  data.frame(string=str, check=str_detect(str, fixed("dog", ignore_case=TRUE)))
}
#Test
DOG <- c("dog", "Dog", "dOg", "doG", "DOg", "DoG", "dOG", "DOG")
check_dog(DOG)
```

```
  string check
1    dog  TRUE
2    Dog  TRUE
3    dOg  TRUE
4    doG  TRUE
5    DOg  TRUE
6    DoG  TRUE
7    dOG  TRUE
8    DOG  TRUE
```

4. Any positive number with or without a decimal point.

```
check_decimal <- function(number){
  str <- as.character(number)
  with <- str_detect(str, fixed("."))
  list <- c()
  list$w_decimal <- number[with]; list$wo_decimal <- number[!with]
  list$boolean <- with
  list
}
#Test
n <- c(1, 1.5, 2, 2.5, 3)
check_decimal(n)
```

```
$w_decimal
[1] 1.5 2.5

$wo_decimal
[1] 1 2 3

$boolean
[1] FALSE  TRUE FALSE  TRUE FALSE
```

5. Any line with exactly two words separated by any amount of whitespace (spaces or tabs). There may or may not be whitespace at the beginning or end of the line.

```
exact_twoword_seperate <- function(str){
  str_detect(str, " *[[:word:]]+ +[[:word:]]+ *|\t*[[:word:]]+\t+[[:word:]]+\t*")
}
#Test
my_str <- c("helloworld", "hello world", "hello  world",
            "hello   world", "hello world ", " hello word", " helloworld ")
exact_twoword_seperate(my_str)
```

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