

Project 8 Solutions

Kevin Choe

Collaborators: N/A

TA help: Katie Brinkers guided with problems 1-7, explained new concepts

Online resources used: Stat 190 Example Book (All problems)

Question 1

```
books <- read.csv("/class/datamine/data/goodreads/csv/goodreads_books.csv")
authors <- read.csv("/class/datamine/data/goodreads/csv/goodreads_book_authors.csv")

# A function that, given a string (myColumn), returns the string
# without any punctuation.
strip_punctuation <- function(myColumn) {
  # Use regular expressions to identify punctuation.
  # Replace identified punctuation with an empty string ''.
  desc_no_punc <- gsub('[:punct:]+', '', myColumn)

  # Return the result
  return(desc_no_punc)
}
books$description<- strip_punctuation(books$description)
```

The name of the function is strip punctuation. There is one argument which is named myColumn.

Question 2

```
#String split by space
test_string <- "This is a test string with no punctuation"
strsplit("This will be split by space.", " ")
```

```
[[1]]
[1] "This" "will" "be" "" "" "split" "by" "space."
```

It would be not be an accurate counting because there are double spaces (two).

Question 3

```
#Use which to fix issue by unlisting strsplit result
strsplit("This will be split by space.", " ")
```

```
[[1]]
[1] "This" "will" "be" "" "" "split" "by" "space."
```

```
x <- unlist(strsplit(test_string, " "))
length(which(x!=""))
```

```
[1] 8
```

Question 4

```
# word_count is a function that accepts a sentence as an argument,  
# strips punctuation and extra space, and returns the number of  
# words in the sentence.  
word_count <- function(sentence) {  
  # strip punctuation and save into an auxiliary variable  
  aux <- gsub('[:punct:]]+', '', sentence)  
  
  # split the sentence by space and remove extra spaces  
  result <- sum(unlist(strsplit(aux, " ")) != "")  
  return(result)  
}  
second <- books[2, "description"]  
word_count(second)
```

```
[1] 211
```

There are 211 words total.

Question 5

```
#Adds together all words and titles  
title_count <- function(name) {  
  aux <- gsub('[:punct:]]+', '', name$title)  
  result <- sum(unlist(strsplit(aux, " ")) != "")  
  return(result)  
}  
title_count(books)
```

```
[1] 5528803
```

Total number of words and titles combined

Pledge

By submitting this work I hereby pledge that this is my own, personal work. I've acknowledged in the designated place at the top of this file all sources that I used to complete said work, including but not limited to: online resources, books, and electronic communications. I've noted all collaboration with fellow students and/or TA's. I did not copy or plagiarize another's work.

As a Boilermaker pursuing academic excellence, I pledge to be honest and true in all that I do.
Accountable together - We are Purdue.