Project 8 Solutions

(Abhimanyu Agarwal)

Collaborators: N/A

TA help:

1) Melissa : Helped me go through Question 4 and 5.

Online resources used: N/A

Question 1

```
#Loads into two distinct dataframes using read.csv()
books <- read.csv("/class/datamine/data/goodreads/csv/goodreads_books.csv")
authors <- read.csv("/class/datamine/data/goodreads/csv/goodreads_book_authors.csv")

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)
head(books$description)
```

```
[1] ""
[2] "Anita Diamants international bestseller The Red Tent brilliantly recreated the ancient world of wo
[3] "Omnibus book club edition containing the Ladies of Madrigyn and the Witches of Wenshar"
[4] "Addie Downs and Valerie Adler were eight when they first met and decided to be best friends foreve
[5] ""
[6] "Relates in vigorous prose the tale of Aeneas the legendary ancestor of Romulus who escaped from th
#The function has one argument. It is called "mycolumn". The name of the function is string_punctation.
```

Question 2

```
#Initialize the test string
test_string <- "This is  a test string  with no punctuation"

#Now we split the string into individual components using strsplit
strsplit(test_string, c(" "))
```

```
[[1]]
 [1] "This"       "is"            ""              "a"           "test"
```

```
 [6] "string"      ""              "with"        "no"          "punctuation"
```
#No, it is not accurate as it takes into account several other punctuations.
#This affects the net result as to how the split turns out to be.

Question 3

```
unlist(strsplit(c(test_string)," "))
```

```
 [1] "This"       "is"          ""            "a"           "test"
 [6] "string"     ""            "with"        "no"          "punctuation"
```
```
sum(unlist(strsplit(test_string, " ") [[1]] != ""))
```

```
[1] 8
```
#It is an alternative way of doing it
#Uses the length instead both account 8
```
length(which(unlist(strsplit(test_string, " ") [[1]] != "")))
```

```
[1] 8
```
#Prints out the number of words

Question 4

```
#Count the number of words in the specified description row & column
count_words <- function(description)
  {
  aux <- gsub('[[:punct:]]+','', description)

  # split the sentence by space and remove extra spaces
  result <- sum(unlist(strsplit(aux, " ")) != "")

  return(result)
  }
count_words(books$description[2])
```

```
[1] 211
```
#Prints out the result

###Question 5
#Function takes in one argument and returns split based on first name, middle name and last. The argume
```
myfunction<- function(names)
{
  return (unlist(strsplit(names, c(" "))))
}
head(authors$name)
```

```
[1] "Ronald J. Fields" "Anita Diamant"    "Barbara Hambly"
[4] "Jennifer Weiner"  "Nigel Pennick"    "Alfred J. Church"
```
```
head(myfunction(authors$name))
```

```
[1] "Ronald"  "J."       "Fields"  "Anita"    "Diamant" "Barbara"
```

```r
tail(myfunction(authors$name))
```

```
[1] "Jolly" "sr@"   "mwrGn" "Barry" "S."    "Brown"
```

Submitting deliverables: project08.RMD, project08.R and project08.pdf

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.