

CUNY MSDS 607 Lab Week 3

Comments (-)

Share

Hide Toolbars

AJ Bentley

September 14, 2018

Q3. The Simpsons

Load data into R data frame

```
raw.data <- "555-1239Moe Szyslak(636) 555-0113Burns, C. Montgomer
y555-6542Rev. Timothy Lovejoy555 8904Ned Flanders636-555-3226Sim
pson, Homer5553642Dr. Julius Hibbert"
```

```
raw.data
```

```
## [1] "555-1239Moe Szyslak(636) 555-0113Burns, C. Montgomery555
-6542Rev. Timothy Lovejoy555 8904Ned Flanders636-555-3226Simpso
n, Homer5553642Dr. Julius Hibbert"
```

Create name vector (*used code from book*)

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

Translate name list to first_name last_name

Pseudocode

1. for any item in “name” that contains a “,” after a word and before another word, reverse the order of the words before and after.
2. strip out any titles. be careful of Chuck Burns—that C. is an initial and not a title.

```
namef1 <- str_replace_all(name, "([[:alpha:]]+)(, )([[:alpha:]].
+)", "\\3 \\1")
```

```
namef12 <- str_replace_all(namef1, "[[:alpha:]]{2,}\\.", "")
```

```
namef12
```

```
## [1] "Moe Szyslak"          "C. Montgomery Burns" " Timothy Lov
ejoy"
## [4] "Ned Flanders"        "Homer Simpson"      " Julius Hibb
ert"
```

