Split_Strings_Quiz.Rmd

eCornell

3/15/2021

Use these libraries and this data set to answer the questions below.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(lubridate)
library(utf8)

calls <- read.csv("calls_Q1_Q1.csv", stringsAsFactors = FALSE)
```

CQ1: Just like there was a structure to the desc variable (i.e., components were separated by a semicolon), there is structure to the date/time variable, dt, because the date and the time are separated by @. Fill in the code below to split each dt value at @ and save the results in two new columns within the calls data frame called date and time. Save the data frame as CQ1.

```
# Initialize vectors to store each of the new variables
dates <- c()
times <- c()

for(i in 1:nrow(calls)){ # loop over emergency calls

  # get the i^th date/time
  dtI <- calls[i, "dt"]

  # split the date/time text based on "@" --> gives a vector of substrings
  splitCallDT <- str_split(dtI, "@", simplify = TRUE)

    # store the date and the time
    dates[i] <- splitCallDT[1]
    times[i] <- splitCallDT[2]

}

# Add new variables to the data frame
```

```
calls$date <- dates
calls$time <- times
head(calls)
```

```
##   X                                                                  desc   zip
## 1 1      BRADFIELD RD & SUSQUEHANNA RD; ABINGTON; 2020-01-01 @ 00:04:06 19001
## 2 2 E CITY AVE & PRESIDENTIAL BLVD; LOWER MERION; 2020-01-01 @ 00:02:25 19004
## 3 3            MAPLE AVE AND W 6TH ST;  LANSDALE; 2020-01-01 @ 00:07:21 19446
## 4 4                        DEKALB ST; BRIDGEPORT; 2020-01-01 @ 00:07:53    NA
## 5 5                         BEECH ST;  POTTSTOWN; 2020-01-01 @ 00:20:15 19464
## 6 6           DEKALB ST AND W 5TH ST; BRIDGEPORT; 2020-01-01 @ 00:20:36 19405
##                      title                         address         towns
## 1          Fire: FIRE ALARM  BRADFIELD RD & SUSQUEHANNA RD      ABINGTON
## 2 Traffic: VEHICLE ACCIDENT E CITY AVE & PRESIDENTIAL BLVD  LOWER MERION
## 3          EMS: LACERATIONS           MAPLE AVE & W 6TH ST      LANSDALE
## 4    Fire: WOODS/FIELD FIRE                      DEKALB ST    BRIDGEPORT
## 5             EMS: STABBING                       BEECH ST     POTTSTOWN
## 6 Traffic: VEHICLE ACCIDENT           DEKALB ST & W 5TH ST    BRIDGEPORT
##                        dt         date      time
## 1  2020-01-01 @ 00:04:06  2020-01-01   00:04:06
## 2  2020-01-01 @ 00:02:25  2020-01-01   00:02:25
## 3  2020-01-01 @ 00:07:21  2020-01-01   00:07:21
## 4  2020-01-01 @ 00:07:53  2020-01-01   00:07:53
## 5  2020-01-01 @ 00:20:15  2020-01-01   00:20:15
## 6  2020-01-01 @ 00:20:36  2020-01-01   00:20:36
```

CQ2: There's also structure to the title variable: The type of first responders (EMS vs. Fire vs. Traffic) is separated from the reason for the call (e.g., "Syncopal episode", "Fracture", etc.) by a colon, :. Continue modifying the calls data set by separating the title variable. First, load the modified calls data set using the code in the first line of the code chunk below. Then, fill in the code below to split the title variable into two new variables (representing the responder type and the call reason). Add these variables to the data frame as two new columns called responder and reason.

```
calls <- read.csv("calls_Q2_Q1.csv", stringsAsFactors = FALSE)

# Initialize vectors to store each of the new variables
responders <- c()
reasons <- c()

for(i in 1:nrow(calls)) { # loop over emergency calls

  # get the i^th call title
  titleI <- calls[i, "title"]

  # split the title text
  splitCallTitle <- str_split(titleI, ":", simplify = TRUE)

  # store the responder and reason
  responders[i] <- splitCallTitle[1]
  reasons[i] <- splitCallTitle[2]
}

# Add new variables to the data frame
calls$responder <- responders
calls$reason <- reasons
```

```r
head(calls)
```

```
##                                                                  desc   zip
## 1      BRADFIELD RD & SUSQUEHANNA RD; ABINGTON; 2020-01-01 @ 00:04:06 19001
## 2 E CITY AVE & PRESIDENTIAL BLVD; LOWER MERION; 2020-01-01 @ 00:02:25 19004
## 3            MAPLE AVE AND W 6TH ST;  LANSDALE; 2020-01-01 @ 00:07:21 19446
## 4                       DEKALB ST; BRIDGEPORT; 2020-01-01 @ 00:07:53    NA
## 5                        BEECH ST;  POTTSTOWN; 2020-01-01 @ 00:20:15 19464
## 6           DEKALB ST AND W 5TH ST; BRIDGEPORT; 2020-01-01 @ 00:20:36 19405
##                       title                          address         towns
## 1          Fire: FIRE ALARM  BRADFIELD RD & SUSQUEHANNA RD      ABINGTON
## 2 Traffic: VEHICLE ACCIDENT E CITY AVE & PRESIDENTIAL BLVD  LOWER MERION
## 3          EMS: LACERATIONS           MAPLE AVE & W 6TH ST      LANSDALE
## 4    Fire: WOODS/FIELD FIRE                      DEKALB ST    BRIDGEPORT
## 5             EMS: STABBING                       BEECH ST     POTTSTOWN
## 6 Traffic: VEHICLE ACCIDENT           DEKALB ST & W 5TH ST    BRIDGEPORT
##                        dt         date      time responder            reason
## 1  2020-01-01 @ 00:04:06  2020-01-01   00:04:06      Fire        FIRE ALARM
## 2  2020-01-01 @ 00:02:25  2020-01-01   00:02:25   Traffic  VEHICLE ACCIDENT
## 3  2020-01-01 @ 00:07:21  2020-01-01   00:07:21       EMS       LACERATIONS
## 4  2020-01-01 @ 00:07:53  2020-01-01   00:07:53      Fire  WOODS/FIELD FIRE
## 5  2020-01-01 @ 00:20:15  2020-01-01   00:20:15       EMS          STABBING
## 6  2020-01-01 @ 00:20:36  2020-01-01   00:20:36   Traffic  VEHICLE ACCIDENT
```