Splitting Strings Into Useable Pieces

Step 1: Load and examine your data.

```
# Load the calls data set:
library(tidyverse)
calls <- read.csv("calls.csv", stringsAsFactors = FALSE)

# Then, examine it to see how it is organized:
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
## 1          Fire: FIRE ALARM
## 2 Traffic: VEHICLE ACCIDENT
## 3          EMS: LACERATIONS
## 4    Fire: WOODS/FIELD FIRE
## 5             EMS: STABBING
## 6 Traffic: VEHICLE ACCIDENT
```

Step 2: Split the description of the first call in the data set.

Extract the description of the first call and assign it to the variable temp:

```
temp <- calls$desc[1]
temp
```

```
## [1] "BRADFIELD RD & SUSQUEHANNA RD; ABINGTON; 2020-01-01 @ 00:04:06"
```

Since the description uses semicolons to split each fragment, that is what you'll use to separate your data when you use the str_split() function:

```
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
## 1          Fire: FIRE ALARM
## 2 Traffic: VEHICLE ACCIDENT
## 3          EMS: LACERATIONS
```

```
## 4    Fire: WOODS/FIELD FIRE
## 5             EMS: STABBING
## 6 Traffic: VEHICLE ACCIDENT
```

Split the first call description into components.

```
str_split(temp, ";", simplify = TRUE)
```

```
##      [,1]                           [,2]        [,3]
## [1,] "BRADFIELD RD & SUSQUEHANNA RD" " ABINGTON" " 2020-01-01 @ 00:04:06"
```

Step 3: Split all call descriptions with a for loop.

To separate the descriptions of each call in the whole data set quickly, you'll want to write a for loop. The comments within the following for loop describe what each step of the loop does.

```
# Initialize vectors to store each of the new variables:

address <- c()
town <- c()
dt <- c()

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

  # get the description of the i^th call
  callI <- calls[i, "desc"]

  # split the description text based on ";" --> gives a matrix of substrings
  splitCallDesc <- str_split(callI, ";", simplify = TRUE)

  # store the street address, town, and date/time
  address[i] <- splitCallDesc[1]
  town[i] <- splitCallDesc[2]
  dt[i] <- splitCallDesc[3]

}
```

Step 4: Store the new vectors in the calls data frame.

Now that you've created the address, town, and date-time (dt) vectors, you can add them to the data frame as new variables:

```
calls$address <- address
calls$towns <- town
calls$dt <- dt
```

Step 5: Examine the new data set.

```
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 AND W 6TH ST      LANSDALE
## 4    Fire: WOODS/FIELD FIRE                      DEKALB ST    BRIDGEPORT
## 5             EMS: STABBING                       BEECH ST     POTTSTOWN
## 6 Traffic: VEHICLE ACCIDENT         DEKALB ST AND W 5TH ST    BRIDGEPORT
##                       dt
## 1  2020-01-01 @ 00:04:06
## 2  2020-01-01 @ 00:02:25
## 3  2020-01-01 @ 00:07:21
## 4  2020-01-01 @ 00:07:53
## 5  2020-01-01 @ 00:20:15
## 6  2020-01-01 @ 00:20:36
```