

Chaudhary_homework_8

Jyoti Chaudhary

November 18, 2016

R Markdown

```

clean_who_number_spaces <- function(x){
  ## Change formatting of numbers from xxx xxx to xxxxxx

  clean_triplet_pair <-
    function(x) str_replace(x, "([0-9]{1,3}) ([0-9]{3})", "\\1\\2")
  number_of_iterations <- max(map_int(x, ~ str_count(., " ")))
  for (i in 1:number_of_iterations){
    x <- clean_triplet_pair(x)
  }
}

fix_who_column <- function(df, col1, col2, col3){

  ## Take col1 as input and output 3 column dataframe

  library(tidyverse)
  library(stringr)
  library(dplyr)

  pos_openbraces <- str_locate(df[[col1]], "\\(")
  pos_closebraces <- str_locate(df[[col1]], "\\)")
  pos_dash <- str_locate(df[[col1]], "-")

  pos_openbraces <- pos_openbraces[,1]
  pos_closebraces <- pos_closebraces[,1]
  pos_dash <- pos_dash[,1]

  df[[col2]] <- ifelse(pos_openbraces != "NA", str_sub(df[[col1]], pos_openbraces + 1, pos_dash
- 1), "NA")

  df[[col3]] <- ifelse(pos_openbraces != "NA", str_sub(df[[col1]], pos_dash + 1, pos_closebraces
- 1), "NA")

  df[[col1]] <- ifelse(pos_openbraces != "NA", str_sub(df[[col1]], end = pos_openbraces - 1), "N
A")

  clean_who_number_spaces(df[[col1]])
  clean_who_number_spaces(df[[col2]])
  clean_who_number_spaces(df[[col3]])

  result <- data.frame("Patient.count" = df[[col1]], "lower.bound" = df[[col2]], "upper.bound" =
df[[col3]])
  result
}

```

```

## Read "Number of People Living with HIV" dataset
HIV_data <- read_csv("data.csv", skip=1)

```

```
## Parsed with column specification:
## cols(
##   `WHO region` = col_character(),
##   `2015` = col_character()
## )
```

```
colnames(HIV_data) <- c("Region", "Patient.count")
```

```
## Call function to read "Patient.count" as input and output a dataframe with 3 columns
```

```
fix_who_column(HIV_data, "Patient.count", "lower.bound", "upper.bound")
```

```
##   Patient.count lower.bound upper.bound
## 1    25 500 000    23 000 000    28 400 000
## 2     3 400 000     3 000 000     3 900 000
## 3     3 500 000     3 000 000     4 100 000
## 4     2 500 000     2 300 000     2 700 000
## 5      330 000      240 000      490 000
## 6     1 400 000     1 100 000     2 000 000
## 7    36 700 000    34 000 000    39 800 000
```