Homework4

Evelin Reyes

10/26/2021

Answer 1

```
library(rvest)
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──
```

```
## ✓ ggplot2 3.3.5     ✓ purrr   0.3.4
## ✓ tibble  3.1.5     ✓ dplyr   1.0.7
## ✓ tidyr   1.1.3     ✓ stringr 1.4.0
## ✓ readr   2.0.1     ✓ forcats 0.5.1
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## x dplyr::filter()         masks stats::filter()
## x readr::guess_encoding() masks rvest::guess_encoding()
## x dplyr::lag()            masks stats::lag()
```

```
library(httr)
library(dplyr)
library(tidyr)

url = "https://introdatasci.dlilab.com/schedule_materials/"
x = url %>%
  httr::GET(config = httr::config(ssl_verifypeer = FALSE)) %>%
  read_html()

df1 = html_table(x)[[1]]
df1
```

```
## # A tibble: 30 × 5
##    Date   Topic                                    Notes HW    Reading
##    <chr>  <chr>                                    <chr> <chr> <chr>
##  1 Aug 24 About the course                         "📙 " "-"   "Leek & Peng 2015"
##  2 Aug 26 Data science project cycle               "📙 " ""    "Mason and Wiggins 2…
##  3 Aug 31 Class cancelled because of Hurrican…     ""    ""    ""
##  4 Sep 2  Class cancelled because of Hurrican…     ""    ""    ""
##  5 Sep 7  Introduction and install tools           "📙 " ""    "Cooper & Hsing 2017"
##  6 Sep 9  Version control with Git                 "📙 " ""    "Blischak et al. 201…
##  7 Sep 14 Introduction to GitHub                   "📙 " ""    ""
##  8 Sep 16 RStudio project and dynamic documen…     "📙 " "01"  "Xie et al, Chapter …
##  9 Sep 21 The file system and basic unix shell     "📙 " ""    "Allesina & Wilmes, …
## 10 Sep 23 R basics: data types, vectors, matr…     "📙 " ""    ""
## # … with 20 more rows
```

Answer 2

```
str_split_fixed(df1$Date," ", 2)
```

```
##       [,1]  [,2]
##  [1,] "Aug" "24"
##  [2,] "Aug" "26"
##  [3,] "Aug" "31"
##  [4,] "Sep" "2"
##  [5,] "Sep" "7"
##  [6,] "Sep" "9"
##  [7,] "Sep" "14"
##  [8,] "Sep" "16"
##  [9,] "Sep" "21"
## [10,] "Sep" "23"
## [11,] "Sep" "28"
## [12,] "Sep" "30"
## [13,] "Oct" "5"
## [14,] "Oct" "7"
## [15,] "Oct" "12"
## [16,] "Oct" "14"
## [17,] "Oct" "19"
## [18,] "Oct" "26"
## [19,] "Oct" "28"
## [20,] "Nov" "2"
## [21,] "Nov" "4"
## [22,] "Nov" "9"
## [23,] "Nov" "11"
## [24,] "Nov" "16"
## [25,] "Nov" "18"
## [26,] "Nov" "23"
## [27,] "Nov" "25"
## [28,] "Nov" "30"
## [29,] "Dec" "2"
## [30,] "Dec" "14"
```

```
df2 <- df1 %>%
  separate(Date, c("Month","Day"), " ", remove = F)
df2
```

```
## # A tibble: 30 × 7
##    Date   Month Day   Topic                          Notes HW    Reading
##    <chr>  <chr> <chr> <chr>                          <chr> <chr> <chr>
##  1 Aug 24 Aug   24    About the course               "📙 " "-"   "Leek & Peng 201…
##  2 Aug 26 Aug   26    Data science project cycle     "📙 " ""    "Mason and Wiggi…
##  3 Aug 31 Aug   31    Class cancelled because of …   ""    ""    ""
##  4 Sep 2  Sep   2     Class cancelled because of …   ""    ""    ""
##  5 Sep 7  Sep   7     Introduction and install to…   "📙 " ""    "Cooper & Hsing …
##  6 Sep 9  Sep   9     Version control with Git       "📙 " ""    "Blischak et al.…
##  7 Sep 14 Sep   14    Introduction to GitHub         "📙 " ""    ""
##  8 Sep 16 Sep   16    RStudio project and dynamic…   "📙 " "01"  "Xie et al, Chap…
##  9 Sep 21 Sep   21    The file system and basic u…   "📙 " ""    "Allesina & Wilm…
## 10 Sep 23 Sep   23    R basics: data types, vecto…   "📙 " ""    ""
## # … with 20 more rows
```

Answer 3

```
df3 <- df2 %>% group_by(Month)
df3
```

```
## # A tibble: 30 × 7
## # Groups:   Month [5]
##    Date   Month Day   Topic                          Notes HW    Reading
##    <chr>  <chr> <chr> <chr>                          <chr> <chr> <chr>
##  1 Aug 24 Aug   24    About the course               "📙 " "-"   "Leek & Peng 201…
##  2 Aug 26 Aug   26    Data science project cycle     "📙 " ""    "Mason and Wiggi…
##  3 Aug 31 Aug   31    Class cancelled because of …   ""    ""    ""
##  4 Sep 2  Sep   2     Class cancelled because of …   ""    ""    ""
##  5 Sep 7  Sep   7     Introduction and install to…   "📙 " ""    "Cooper & Hsing …
##  6 Sep 9  Sep   9     Version control with Git       "📙 " ""    "Blischak et al.…
##  7 Sep 14 Sep   14    Introduction to GitHub         "📙 " ""    ""
##  8 Sep 16 Sep   16    RStudio project and dynamic…   "📙 " "01"  "Xie et al, Chap…
##  9 Sep 21 Sep   21    The file system and basic u…   "📙 " ""    "Allesina & Wilm…
## 10 Sep 23 Sep   23    R basics: data types, vecto…   "📙 " ""    ""
## # … with 20 more rows
```

```
table = summarise(df3, No.Lectures = n())
table
```

```
## # A tibble: 5 × 2
##   Month No.Lectures
##   <chr>       <int>
## 1 Aug             3
## 2 Dec             2
## 3 Nov             9
## 4 Oct             7
## 5 Sep             9
```

```
df4 <- table[order(-table$No.Lectures),]
df4
```

```
## # A tibble: 5 × 2
##   Month No.Lectures
##   <chr>       <int>
## 1 Nov             9
## 2 Sep             9
## 3 Oct             7
## 4 Aug             3
## 5 Dec             2
```

Answer 4

```
df5 <- str_split(df3$Topic, " ")
df5
```

```
## [[1]]
## [1] "About"  "the"    "course"
##
## [[2]]
## [1] "Data"    "science" "project" "cycle"
##
## [[3]]
## [1] "Class"     "cancelled" "because"   "of"        "Hurricane" "Ida"
##
## [[4]]
## [1] "Class"     "cancelled" "because"   "of"        "Hurricane" "Ida"
##
## [[5]]
## [1] "Introduction" "and"          "install"      "tools"
##
## [[6]]
## [1] "Version" "control" "with"    "Git"
##
## [[7]]
## [1] "Introduction" "to"           "GitHub"
##
## [[8]]
## [1] "RStudio"   "project"   "and"       "dynamic"   "documents" "with"
## [7] "R"         "Markdown"
##
## [[9]]
## [1] "The"    "file"   "system" "and"    "basic"  "unix"   "shell"
##
## [[10]]
## [1] "R"        "basics:"  "data"     "types,"   "vectors," "matrix,"  "data"
## [8] "frame,"   "etc."
##
## [[11]]
## [1] "More"    "R"       "basics:" "lists,"  "dates,"  "etc."
##
## [[12]]
## [1] "R"           "programming" "basics:"     "conditional" "statements"
##
## [[13]]
## [1] "R"           "programming" "basics:"     "loops,"      "apply"
##
## [[14]]
## [1] "Strings"     "and"         "Regular"     "expressions"
##
## [[15]]
## [1] "API"      "and"      "data"     "scraping"
##
## [[16]]
## [1] "Data"   "input"  "and"    "output"
##
## [[17]]
## [1] "Data"         "manipulation" "with"         "R"
##
```

```
## [[18]]
## [1] "More"         "data"         "manipulation" "with"         "R"
##
## [[19]]
## [1] "Data"          "visualization" "with"          "R"
##
## [[20]]
## [1] "Exploratory" "data"        "analysis"
##
## [[21]]
## [1] "Regression" "methods"
##
## [[22]]
## [1] "More"       "on"         "Regression" "methods"
##
## [[23]]
## [1] "Write"     "your"      "own"       "functions"
##
## [[24]]
## [1] "Write"   "your"    "own"     "R"       "package"
##
## [[25]]
## [1] "Open"       "Science"    "and"        "automating" "things"
## [6] "with"       "Makefile"
##
## [[26]]
## [1] "Ethics"    "in"        "data"      "science"   "(virtual)"
##
## [[27]]
## [1] "Thanksgiving," "no"            "class"
##
## [[28]]
## [1] "Final"        "project"      "presentation"
##
## [[29]]
## [1] "Final"        "project"      "presentation" "and"          "wrap"
## [6] "up"
##
## [[30]]
## [1] "Final"  "grades" "due"
```

```
words <- unlist(df5)
words
```

```
##   [1] "About"         "the"           "course"        "Data"
##   [5] "science"       "project"       "cycle"         "Class"
##   [9] "cancelled"     "because"       "of"            "Hurricane"
##  [13] "Ida"           "Class"         "cancelled"     "because"
##  [17] "of"            "Hurricane"     "Ida"           "Introduction"
##  [21] "and"           "install"       "tools"         "Version"
##  [25] "control"       "with"          "Git"           "Introduction"
##  [29] "to"            "GitHub"        "RStudio"       "project"
##  [33] "and"           "dynamic"       "documents"     "with"
##  [37] "R"             "Markdown"      "The"           "file"
##  [41] "system"        "and"           "basic"         "unix"
##  [45] "shell"         "R"             "basics:"       "data"
##  [49] "types,"        "vectors,"      "matrix,"       "data"
##  [53] "frame,"        "etc."          "More"          "R"
##  [57] "basics:"       "lists,"        "dates,"        "etc."
##  [61] "R"             "programming"   "basics:"       "conditional"
##  [65] "statements"    "R"             "programming"   "basics:"
##  [69] "loops,"        "apply"         "Strings"       "and"
##  [73] "Regular"       "expressions"   "API"           "and"
##  [77] "data"          "scraping"      "Data"          "input"
##  [81] "and"           "output"        "Data"          "manipulation"
##  [85] "with"          "R"             "More"          "data"
##  [89] "manipulation"  "with"          "R"             "Data"
##  [93] "visualization" "with"          "R"             "Exploratory"
##  [97] "data"          "analysis"      "Regression"    "methods"
## [101] "More"          "on"            "Regression"    "methods"
## [105] "Write"         "your"          "own"           "functions"
## [109] "Write"         "your"          "own"           "R"
## [113] "package"       "Open"          "Science"       "and"
## [117] "automating"    "things"        "with"          "Makefile"
## [121] "Ethics"        "in"            "data"          "science"
## [125] "(virtual)"     "Thanksgiving," "no"            "class"
## [129] "Final"         "project"       "presentation"  "Final"
## [133] "project"       "presentation"  "and"           "wrap"
## [137] "up"            "Final"         "grades"        "due"
```

```
freq <- table(words)
freq
```

```
## words
##     (virtual)         About      analysis           and           API
##             1             1             1             8             1
##         apply    automating         basic       basics:       because
##             1             1             1             4             2
##     cancelled         class         Class   conditional       control
##             2             1             2             1             1
##        course         cycle          data          Data        dates,
##             1             1             6             4             1
##     documents           due       dynamic          etc.        Ethics
##             1             1             1             2             1
##   Exploratory   expressions          file         Final        frame,
##             1             1             1             3             1
##     functions           Git        GitHub        grades     Hurricane
##             1             1             1             1             2
##           Ida            in         input       install  Introduction
##             2             1             1             1             2
##        lists,        loops,      Makefile  manipulation      Markdown
##             1             1             1             2             1
##       matrix,       methods          More            no            of
##             1             2             3             1             2
##            on          Open        output           own       package
##             1             1             1             2             1
##  presentation   programming       project             R    Regression
##             2             2             4             9             2
##       Regular       RStudio       science       Science      scraping
##             1             1             2             1             1
##         shell    statements       Strings        system Thanksgiving,
##             1             1             1             1             1
##           the           The        things            to         tools
##             1             1             1             1             1
##        types,          unix            up      vectors,       Version
##             1             1             1             1             1
## visualization          with          wrap         Write          your
##             1             6             1             2             2
```

```
most <- sort(freq,decreasing = TRUE)
most
```

```
## words
##             R           and          data          with       basics:
##             9             8             6             6             4
##          Data       project         Final          More       because
##             4             4             3             3             2
##     cancelled         Class          etc.     Hurricane           Ida
##             2             2             2             2             2
##  Introduction  manipulation       methods            of           own
##             2             2             2             2             2
##  presentation   programming    Regression       science         Write
##             2             2             2             2             2
##          your     (virtual)         About      analysis           API
##             2             1             1             1             1
##         apply    automating         basic         class   conditional
##             1             1             1             1             1
##       control        course         cycle        dates,     documents
##             1             1             1             1             1
##           due       dynamic        Ethics   Exploratory   expressions
##             1             1             1             1             1
##          file        frame,     functions           Git        GitHub
##             1             1             1             1             1
##        grades            in         input       install        lists,
##             1             1             1             1             1
##        loops,      Makefile      Markdown       matrix,            no
##             1             1             1             1             1
##            on          Open        output       package       Regular
##             1             1             1             1             1
##       RStudio       Science      scraping         shell    statements
##             1             1             1             1             1
##       Strings        system Thanksgiving,           the           The
##             1             1             1             1             1
##        things            to         tools        types,          unix
##             1             1             1             1             1
##            up      vectors,       Version visualization          wrap
##             1             1             1             1             1
```

```
most[1:5]
```

```
## words
##       R     and    data    with basics:
##       9       8       6       6       4
```