# Untitled

Suliat Alli

10/26/2021

**Homework 4**

```
# install.packages("rvest")
?rvest
library(rvest)
library(tidyr)
library(dplyr)
www <- read_html('https://introdatasci.dlilab.com/schedule_materials/')

content <- www %>%
  html_table(fill = T)
table <- content[[1]]
table
```

**Question 1**

```
## No documentation for 'rvest' in specified packages and libraries:
## you could try '??rvest'
## # A tibble: 30 x 5
##    Date   Topic                            Notes    HW    Reading
##    <chr>  <chr>                            <chr>    <chr> <chr>
##  1 Aug 24 About the course                 "\U0001f~ "-"   "Leek & Peng 2015"
##  2 Aug 26 Data science project cycle       "\U0001f~ ""    "Mason and Wiggins~
##  3 Aug 31 Class cancelled because of Hurric~ ""       ""    ""
##  4 Sep 2  Class cancelled because of Hurric~ ""       ""    ""
##  5 Sep 7  Introduction and install tools   "\U0001f~ ""    "Cooper & Hsing 20~
##  6 Sep 9  Version control with Git         "\U0001f~ ""    "Blischak et al. 2~
##  7 Sep 14 Introduction to GitHub           "\U0001f~ ""    ""
##  8 Sep 16 RStudio project and dynamic docum~ "\U0001f~ "01"  "Xie et al, Chapte~
##  9 Sep 21 The file system and basic unix sh~ "\U0001f~ ""    "Allesina & Wilmes~
## 10 Sep 23 R basics: data types, vectors, ma~ "\U0001f~ ""    ""
## # ... with 20 more rows
```

```
table <- table %>%
  separate(Date, c("Month", "Day"))
table
```

**Question 2**

```
## # A tibble: 30 x 6
##    Month Day   Topic                          Notes      HW    Reading
##    <chr> <chr> <chr>                          <chr>      <chr> <chr>
##  1 Aug   24    About the course               "\U0001f~ "-"   "Leek & Peng 2015"
##  2 Aug   26    Data science project cycle     "\U0001f~ ""    "Mason and Wiggin~
##  3 Aug   31    Class cancelled because of Hu~ ""         ""    ""
##  4 Sep   2     Class cancelled because of Hu~ ""         ""    ""
##  5 Sep   7     Introduction and install tools "\U0001f~ ""    "Cooper & Hsing 2~
##  6 Sep   9     Version control with Git       "\U0001f~ ""    "Blischak et al. ~
##  7 Sep   14    Introduction to GitHub         "\U0001f~ ""    ""
##  8 Sep   16    RStudio project and dynamic d~ "\U0001f~ "01"  "Xie et al, Chapt~
##  9 Sep   21    The file system and basic uni~ "\U0001f~ ""    "Allesina & Wilme~
## 10 Sep   23    R basics: data types, vectors~ "\U0001f~ ""    ""
## # ... with 20 more rows
```

```
table %>%
  group_by(Month) %>%
  summarise(num_lecture = n()) %>%
  arrange(desc(num_lecture))
```

**Question 3**

```
## # A tibble: 5 x 2
##   Month num_lecture
##   <chr>       <int>
## 1 Nov             9
## 2 Sep             9
## 3 Oct             7
## 4 Aug             3
## 5 Dec             2
```

```
library(stringr)

word_list <- strsplit(table$Topic, " ")
words <- unlist(word_list)
words <- table(words)

freq_words <- words %>%
  as.data.frame() %>%
  arrange(desc(Freq))

top_n(freq_words, 5)
```

**Question 4**

```
##      words Freq
## 1        R    9
## 2      and    8
## 3     data    6
## 4     with    6
## 5  basics:    4
## 6     Data    4
## 7  project    4
```