hw\_4

ahmed tohamy

10/28/2021

library(rvest)  
library(httr)  
library(stringr)  
library(stringi)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(tidyr)  
  
webpage <- ("https://introdatasci.dlilab.com/schedule\_materials/")  
xpath <- '//\*[@id="main"]/table'  
  
table <- webpage %>%  
read\_html() %>%  
html\_nodes(xpath = xpath) %>%  
html\_table(., fill = TRUE)   
table <- table[[1]]  
print(table, n =30) #Q1

## # 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… "📙" "" ""   
## 11 Sep 28 More R basics: lists, dates, etc. "📙" "" "Hadley, Chapter 4"   
## 12 Sep 30 R programming basics: conditional s… "📙" "02" ""   
## 13 Oct 5 R programming basics: loops, apply "📙" "" ""   
## 14 Oct 7 Strings and Regular expressions "📙" "03" "Peng, Chapter 17"   
## 15 Oct 12 API and data scraping "📙" "" ""   
## 16 Oct 14 Data input and output "📙" "" "Hadley, Chapter 11"   
## 17 Oct 19 Data manipulation with R "📙" "04" "Hadley, Chapter 5"   
## 18 Oct 26 More data manipulation with R "📙" "" "Hadley, Chapter 5"   
## 19 Oct 28 Data visualization with R "📙" "05" "Holmes and Huber, C…  
## 20 Nov 2 Exploratory data analysis "📙" "" "Hadley, Chapter 7"   
## 21 Nov 4 Regression methods "📙" "06" ""   
## 22 Nov 9 More on Regression methods "📙" "" "Navarro, Chapter 15"  
## 23 Nov 11 Write your own functions "📙" "" "Hadley, Chapter 19"   
## 24 Nov 16 Write your own R package "📙" "07" "Hadley, Chapter 2"   
## 25 Nov 18 Open Science and automating things … "" "" ""   
## 26 Nov 23 Ethics in data science (virtual) "" "" ""   
## 27 Nov 25 Thanksgiving, no class "" "" ""   
## 28 Nov 30 Final project presentation "" "" ""   
## 29 Dec 2 Final project presentation and wrap… "" "" ""   
## 30 Dec 14 Final grades due "" "" ""

table$month <- word(table$Date, 1)  
table$day <- stri\_sub(table$Date,-2,-1)  
table$day <-as.numeric(table$day)  
print(table, n=30) #Q2

## # A tibble: 30 × 7  
## Date Topic Notes HW Reading month day  
## <chr> <chr> <chr> <chr> <chr> <chr> <dbl>  
## 1 Aug 24 About the course "📙" "-" "Leek & Peng 201… Aug 24  
## 2 Aug 26 Data science project cycle "📙" "" "Mason and Wiggi… Aug 26  
## 3 Aug 31 Class cancelled because of … "" "" "" Aug 31  
## 4 Sep 2 Class cancelled because of … "" "" "" Sep 2  
## 5 Sep 7 Introduction and install to… "📙" "" "Cooper & Hsing … Sep 7  
## 6 Sep 9 Version control with Git "📙" "" "Blischak et al.… Sep 9  
## 7 Sep 14 Introduction to GitHub "📙" "" "" Sep 14  
## 8 Sep 16 RStudio project and dynamic… "📙" "01" "Xie et al, Chap… Sep 16  
## 9 Sep 21 The file system and basic u… "📙" "" "Allesina & Wilm… Sep 21  
## 10 Sep 23 R basics: data types, vecto… "📙" "" "" Sep 23  
## 11 Sep 28 More R basics: lists, dates… "📙" "" "Hadley, Chapter… Sep 28  
## 12 Sep 30 R programming basics: condi… "📙" "02" "" Sep 30  
## 13 Oct 5 R programming basics: loops… "📙" "" "" Oct 5  
## 14 Oct 7 Strings and Regular express… "📙" "03" "Peng, Chapter 1… Oct 7  
## 15 Oct 12 API and data scraping "📙" "" "" Oct 12  
## 16 Oct 14 Data input and output "📙" "" "Hadley, Chapter… Oct 14  
## 17 Oct 19 Data manipulation with R "📙" "04" "Hadley, Chapter… Oct 19  
## 18 Oct 26 More data manipulation with… "📙" "" "Hadley, Chapter… Oct 26  
## 19 Oct 28 Data visualization with R "📙" "05" "Holmes and Hube… Oct 28  
## 20 Nov 2 Exploratory data analysis "📙" "" "Hadley, Chapter… Nov 2  
## 21 Nov 4 Regression methods "📙" "06" "" Nov 4  
## 22 Nov 9 More on Regression methods "📙" "" "Navarro, Chapte… Nov 9  
## 23 Nov 11 Write your own functions "📙" "" "Hadley, Chapter… Nov 11  
## 24 Nov 16 Write your own R package "📙" "07" "Hadley, Chapter… Nov 16  
## 25 Nov 18 Open Science and automating… "" "" "" Nov 18  
## 26 Nov 23 Ethics in data science (vir… "" "" "" Nov 23  
## 27 Nov 25 Thanksgiving, no class "" "" "" Nov 25  
## 28 Nov 30 Final project presentation "" "" "" Nov 30  
## 29 Dec 2 Final project presentation … "" "" "" Dec 2  
## 30 Dec 14 Final grades due "" "" "" Dec 14

table\_lec <- table %>% group\_by(month) %>% summarise(n())  
table\_lec\_order <- table\_lec[order(-table\_lec$`n()`),]  
print(table\_lec\_order) #Q3

## # A tibble: 5 × 2  
## month `n()`  
## <chr> <int>  
## 1 Nov 9  
## 2 Sep 9  
## 3 Oct 7  
## 4 Aug 3  
## 5 Dec 2

len <- length(ncol(table))  
word\_list <- vector(mode="list", length = len)  
word\_list <- strsplit(table$Topic, split= " ")  
words <- unlist(word\_list)  
words <- sort((table(words)), decreasing = TRUE)  
  
print(words) #Q4 The top 5 words are: R, and, data, with & basics.

## 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