

HW_04

LeoFontenot

10/26/2021

```
library(rvest)
url <- read_html("https://introdatasci.dlilab.com/schedule_materials/")

SnM_table <- url %>% html_elements("table") |> html_table()

'SMdf' <- as.data.frame(SnM_table)
```

```
days <- stringi::stri_sub(SMdf[,1], -2)
days <- as.numeric(days)
SMdf_new <- cbind(SMdf, days)

month <- stringi::stri_sub(SMdf[,1], from = 1, to = 3)
month <- as.factor(month)
SMdf_new <- cbind(SMdf, month)
```

```
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
```

```
lec_per_m <- SMdf_new |> group_by(month) |> summarise(month, n = n()) |> distinct() |> arrange(month, desc(n))
```

```
## 'summarise()' has grouped output by 'month'. You can override using the '.groups' argument.
```

```
topic <- stringr::str_split(SMdf_new$Topic, " ")
in_words <- unlist(topic)
top5 <- sort(prop.table(table(in_words)), decreasing = T)
```