

# R Notebook

```
include <- function(library_name){  
  if( !(library_name %in% installed.packages()) )  
    install.packages(library_name)  
  library(library_name, character.only=TRUE)  
}  
include("rvest")  
include("tidyr")  
include("stringr")
```

```
my_url <- "http://ems.csuchico.edu/APSS/schedule/spr2019/CSCI.shtml"  
html1 <- read_html(my_url)  
my_url <- "http://ems.csuchico.edu/APSS/schedule/spr2020/CSCI.shtml"  
html2 <- read_html(my_url)  
my_url <- "http://ems.csuchico.edu/APSS/schedule/spr2019/MATH.shtml"  
html3 <- read_html(my_url)  
my_url <- "http://ems.csuchico.edu/APSS/schedule/spr2020/MATH.shtml"  
html4 <- read_html(my_url)
```

This chunk does the actual scraping of the html

The function returns a tibble of the data scraped

```
read_class_schedule <- function(url) {  
  
  classes <- url %>% html_nodes(".classrow")  
  
  subject <- classes %>%  
    html_nodes("td.subj") %>%  
    html_text()  
  
  catNum <- classes %>%  
    html_nodes("td.cat_num") %>%  
    html_text() %>%  
    as.integer()  
  
  sect <- classes %>%  
    html_nodes("td.sect") %>%  
    html_text() %>%  
    as.integer()  
  
  type <- classes %>%  
    html_nodes("td.comp") %>%  
    html_text()  
  
  title <- classes %>%  
    html_nodes("td.title") %>%  
    html_text()  
  
  units <- classes %>%
```

```

    html_nodes("td.units") %>%
    html_text() %>%
    as.integer()

instruct <- classes %>%
  html_nodes("td.Instructor") %>%
  html_text()

totenro <- classes %>%
  html_nodes("td.enrtot") %>%
  html_text() %>%
  as.integer()

temp <- url %>%
  html_nodes("#bd > div > div > h1") %>%
  html_text()

pattern1 <- "(Spring|Fall)\\ (\\d*)"
yearSem <- tibble(strings = stringr::str_match_all(temp, pattern1))

returned <- tibble(subject= subject, catNum= catNum, sect= sect, type=type, title=title, units=units,
  return(returned)
}

csci2019 <- read_class_schedule(html1)

```

```
## Warning in function_list[[k]](value): NAs introduced by coercion
```

```
head(csci2019)
```

```
## # A tibble: 6 x 10
##   subject catNum  sect type  title      units instruct  totenro sem  year
##   <chr>    <int> <int> <chr> <chr>    <int> <chr>    <int> <chr> <chr>
## 1 CSCI      101    1 DIS  Intro to~    3 " "      0 Spri~ 2019
## 2 CSCI      102    1 DIS  Living W~    3 Juliano,B~ 26 Spri~ 2019
## 3 CSCI      111    2 ACT  Programm~   NA Gibson,To~ 29 Spri~ 2019
## 4 CSCI      111    4 LEC  Programm~    4 Raigoza,J~ 49 Spri~ 2019
## 5 CSCI      111    6 ACT  Programm~   NA Raigoza,J~ 19 Spri~ 2019
## 6 CSCI      211    2 ACT  Programm~   NA Donnelly,~ 27 Spri~ 2019

```

```
csci2020 <- read_class_schedule(html2)
```

```
## Warning in function_list[[k]](value): NAs introduced by coercion
```

```
head(csci2020)
```

```
## # A tibble: 6 x 10
##   subject catNum  sect type  title      units instruct  totenro sem  year
##   <chr>    <int> <int> <chr> <chr>    <int> <chr>    <int> <chr> <chr>
## 1 CSCI      101    1 DIS  Intro to ~    3 Herring,~    1 Spri~ 2020

```

```
## 2 CSCI      102      1 DIS   Living Wi~      3 Harris,K~      0 Spri~ 2020
## 3 CSCI      111      2 ACT   Programmi~    NA Gibson,T~      1 Spri~ 2020
## 4 CSCI      111      4 LEC   Programmi~      4 Renner,R~      1 Spri~ 2020
## 5 CSCI      111      6 ACT   Programmi~    NA Renner,R~      0 Spri~ 2020
## 6 CSCI      211      2 ACT   Programmi~    NA Herring,~      3 Spri~ 2020
```

```
math2019 <- read_class_schedule(html3)
```

```
## Warning in function_list[[k]](value): NAs introduced by coercion
```

```
head(math2019)
```

```
## # A tibble: 6 x 10
##   subject catNum  sect type  title      units instruct  totenro sem  year
##   <chr>      <int> <int> <chr> <chr>      <int> <chr>      <int> <chr> <chr>
## 1 MATH        31      1 LAB  Foundati~      1 " "          0 Spri~ 2019
## 2 MATH        51      1 LAB  Foundati~      1 " "          0 Spri~ 2019
## 3 MATH        51      3 LAB  Foundati~      1 Huck,Soph~     24 Spri~ 2019
## 4 MATH        51      5 LAB  Foundati~      1 Xiong,Mou~     21 Spri~ 2019
## 5 MATH        51      7 LAB  Foundati~      1 " "          0 Spri~ 2019
## 6 MATH        51      9 LAB  Foundati~      1 Archuleta~    34 Spri~ 2019
```

```
math2020 <- read_class_schedule(html4)
```

```
## Warning in function_list[[k]](value): NAs introduced by coercion
```

```
head(math2020)
```

```
## # A tibble: 6 x 10
##   subject catNum  sect type  title      units instruct  totenro sem  year
##   <chr>      <int> <int> <chr> <chr>      <int> <chr>      <int> <chr> <chr>
## 1 MATH        31      1 LAB  Foundatio~      1 " "          0 Spri~ 2020
## 2 MATH        51      1 LAB  Foundatio~      1 " "          0 Spri~ 2020
## 3 MATH        51      5 LAB  Foundatio~      1 " "          0 Spri~ 2020
## 4 MATH        51      8 LAB  Foundatio~      1 McElwain~      0 Spri~ 2020
## 5 MATH        51     10 LAB  Foundatio~      1 Huff,Ste~      1 Spri~ 2020
## 6 MATH        51     12 LAB  Foundatio~      1 McElwain~      1 Spri~ 2020
```