

Lab 12

Frankie Lin

4/25/2019

```
library(XML)
library(xml2)
library(rvest)

##
## Attaching package: 'rvest'

## The following object is masked from 'package:XML':
##
##      xml

library(magrittr)
library(stringr)
```

Your turn*

```
# Assemble url (so it fits on screen)
basket <- "https://www.basketball-reference.com"
gsw <- "/teams/GSW/2017.html"
gsw_url <- paste0(basket, gsw)

# download HTML file to your working directory
download.file(gsw_url, 'gsw-roster-2017.html')

# Read GSW Roster html table
gsw_roster <- readHTMLTable('gsw-roster-2017.html')
lol <- gsw_roster[[1]]

# Extracting html
nba_html <- paste0(basket, "/leagues/NBA_2017.html")

xml_doc <- read_html(nba_html)

xml_text <- xml_doc %>% html_text()

xml_tables <- xml_doc %>%
  html_nodes("table") %>%
  extract(1:2)

# Pulling out the names of the
hrefs <- xml_tables %>%
  html_nodes("a") %>%
  html_attr("href")

hrefs
```

```

teams <- str_extract(hrefs, "[A-Z][A-Z][A-Z]")
files <- paste0(teams, "-roster-2017.csv")
basket <- "https://www.basketball-reference.com"

for (i in 1:length(teams)){
  url <- paste0(basket, hrefs[i])
  download.file(url, paste0(teams[i], '-roster-2017.html'))
  # Read Roster html table
  table <- readHTMLTable(paste0(teams[i], '-roster-2017.html'))
  # Download
  write.csv(table[1], file = files[i])
}

```

Challenge*

```

base = data.frame()
for (i in 1:length(teams)){
  base <- rbind(base, read.csv(files[i]))
}
write.csv(base, 'nba-rosters-2017.csv')

```