Elementary Web Mining Using rvest and bankrates.com

Robert Schnitman

November 26, 2019

```
suppressMessages(lapply(c('tidyverse', 'magrittr', 'knitr', 'kableExtra', 'rvest'),
                        function(x) library(x, character.only = TRUE)))
```

```
links <- data.frame(category = c('Checking Accounts', 'Savings', 'Money Moarket'),
                    url = c('https://www.bankrate.com/banking/checking/
                            best-checking-accounts/',
                            'https://www.bankrate.com/banking/savings/rates/',
                            'https://www.bankrate.com/banking/money-market/rates/'),
                    stringsAsFactors = FALSE)
```

```
read_html_table <- function(url) {


  read_html(url) %>%
    html_nodes('table') %>%
    html_table() %>%
    as.data.frame() %>%
    map_df(~ gsub('\\*', '', .))

} # Output: list

mykable <- function(df, ...) {

  kable(df, ..., booktabs = TRUE) %>%
    kable_styling(full_width = TRUE)

}
```

```
### Test with only the first link

col_names <- c('Bank', 'Monthly Fee', 'APY',
               'Minimum Opening Balance', 'Minimum Balance to Avoid Fees')

read_html_table(links$url[[1]]) %>%
  mykable(caption = links$category[[1]], col.names = col_names) %>%
  collapse_rows(1, valign = 'top')
```

Table 1: Checking Accounts

Bank	Monthly Fee	APY	Minimum Opening Balance	Minimum Balance to Avoid Fees
HSBC Premier Checking	$50	0%	$1	$75,000
Ally Interest Checking	$0	0.10%	$0	$0
Capital One 360 Checking	$0	0.20%	$0	$0
Discover Cashback Debit Checking	$0	0%	$0	$0
Chase Premier Plus Checking	$25	0.01%	Varies by location	Varies by account type
Heritage Bank eCentive Checking	$0	3.33%	$100	$0
Simple Individual Checking	$0	2.02%	$0	$0
TIAA Bank Yield Pledge Checking	$0	1.01%	$100	$0
Radius Bank Rewards Checking	$0	1.20%	$100	$0
NBKC Bank Personal Account	$0	1.01%	$5	$0