Harvesting Data From the Web With Rvest: Solutions

Below are the solutions to these exercises on "Harvesting Data From the Web With Rvest."

```r
# load package
library(rvest)

# Exercise 1
webpage <- read_html(x = "https://money.cnn.com/data/us_markets/"

# Exercise 2

## <session> https://money.cnn.com/data/us_markets/
##   Status: 200
##   Type:   text/html; charset=utf-8
##   Size:   82914

# Exercise 3
html_text(html_nodes(webpage, "a[href*='sectors']"))

## [1] "Communications"     "Consumer Durables"
## [3] "Consumer Non-Durables" "Commercial Services"
```
Exercise 4

Exercise 5

Exercise 6
# Exercise 7

html_text(html_nodes(webpage, ".HeadlineList a"))

## [1] "Jamie Dimon on the trade war, infrastructure 'emergency' and Trump"
## [2] "Papa John's is worried that Papa John will try to take over the company"
## [3] "Watch: What keeps Jamie Dimon up at night"
## [4] "This is how Jamie Dimon would fix income inequality"
## [5] "Sergio Marchionne, auto legend, steps down as CEO of Fiat Chrysler"
## [6] "Overstock is making a push into real estate"
## [7] "The Gates Foundation is now one of Berkshire's largest shareholders"
## [8] "AI is hurting people of color and the poor. Experts want to fix that"
## [9] "Tariffs and Brexit could kill this steel company"
## [10] "Global immigration backlash could hurt India's top tech companies"
## [11] "Ryanair's profit has dropped 20% and more trouble is coming"
## [12] "New York Daily News to slash 50% of its newsroom"
# Exercise 8

```r
webpage %>% html_nodes(".wsod_disclaimer > span") %>% html_attrs() %>% [[1]]
```

## stream

```
"time_144221|196723|276301|417335648|218057|143954|170715|263373|38911044|207106"
```

## streamjstime

```
"1532359505000"
```

## streamdateformat

```
"g%3Ai%3Asa%20x"
```

# Exercise 9

```r
webpage %>% html_nodes(".scale div") %>% html_attr("class")
```

```
[1] "bars pct100" "bars pct60" "bars pct60" "bars pct60" "bars pct50"
[6] "bars pct50" "bars pct50" "bars pct40" "bars pct40" "bars pct40"
```

# Exercise 10

```r
webpage %>% html_nodes("img[src$='svg']") %>% html_attr("src")
```