Name:_____ Date:_____

LAB 3E: Scraping web data
Response Sheet

Directions: Record your responses to the lab questions in the spaces provided.

The web as a data source

Our first web scraper

(1) Briefly describe what the data on the website is about.

(2) Write down 3 questions you'd be interested in answering by analyzing this data.

HTML

(3) How is the data table in `HTML` different than the data tables we're used to seeing in `R`, for example, when we use the `View()` function?

(4) What do you think the *tags* `<TABLE>`, `<TR>`, `<TH>`, `<TD>` mean? How does `HTML` use these *tags* to display the table?

Get to scraping!

Find the URL address for the site and assign it the name `data_url` in `R`.

(5) Then fill in the blanks below to have `R` scrape *every* web table available on the site:

```
tables <- readHTMLTable(_____)
```

Name:_____ Date:_____

LAB 3E: Scraping web data
Response Sheet

Find our data

(6) Write and run code using the `length()` function to find out how many tables of data were scraped in our set of `tables`.

Saving tables

(7) Write and run code using `readHTMLTable()` to re-scrape the data from the web but this time use the `which` argument to scrape just the individual table.

Check, save and use!

(8) Fill in the blanks to save the data and give it a file name.

```
save(___________, file = "___________.Rda")
```

(9) What is the mean and standard deviation of `elev_ft`?

(10) Which `state` has the most mountains in our data?