Project 20

Name:

Using RStudio, we start to learn how to extract data from the web.

Use the data from the Billboard Hot 100 for question 1. Please use the data from the week you were born. For instance, if I solve question 1, I would use the data located here: https://www.billboard.com/charts/hot-100/1976-10-13

On the Hot 100 chart, from the day of your birth:

1a. Extract the titles of the songs ranked #2 through #100.

Explain solution:

1b. Extract the artists for those 99 songs.

 # Place code here

Explain solution:

1c. Extract the title of the number 1 song for that day.

 # Place code here

Explain solution:

1d. Extract the artist for the number 1 song for that day.

 # Place code here

Explain solution:

Using the data from the National Park Service: https://www.nps.gov/

2a. Extract the cities where each of the National Park properties are located.

 # Place code here

Explain solution:

2b. Extract the zip codes for the properties.

 # Place code here

Explain solution:

3a. Identify three potential websites that you are interested to try to scrape yourself, during the upcoming seminars. Look for websites with data that is (relatively) easy to scrape, for instance: Systematic URL's that are easy to understand; (relative) consistency in how the data is stored; and make sure that the data is embedded in the page, rather than in csv files that are already prepared for download. (We want to actually scrape some data.)

1.

2.

3.

3b. For each of the three websites that you identified, give a very brief description of the kind of data that you want to scrape.

1.

2.

3.