

# Northeastern University

**Course:** DA5020  
**Assignment:** Web Scraping with R  
**Total Points:** 100  
**Date Due:** Posted on Blackboard

## Learning Objectives

In this assignment, you will learn how to:

- scrape data from HTML through R programming
- specify search parameters through a URL
- parse HTML

## Tasks

The objective of this assignment is to learn how to extract data from web pages through R programming.

- Select a website from which to scrape the data
- (20 points) Declare (in your R code as comments) exactly what data the code scrapes and from where
- (20 points) Make the scraping parameterized, i.e., allow a data scientist to select search parameters -- choose a website that uses GET requests
- (50 points) Write a function to scrape the data and return the data as a data frame
- (10 points) Provide additional code that validates the data retrieved is the expected data set. This could be a simple count of table records, a distribution of known values, or retrieval of specific records from the data frame.