

Assignment 5: Part B**Web Scraping** (New York Times Best Seller List) 10 points

Due Date: May 28, Tuesday, 11:59pm

The objective of this assignment is to scrape New York Times Best Sellers List (Combined Print & E-Book Fiction) data from 2011 to 2018. For example, the top hit currently for 2018 is "The Reckoning" and for 2011, it is "11/22/63". A weblink for the year 2017 is as follows:

<https://www.nytimes.com/books/best-sellers/2017/12/31/combined-print-and-e-book-fiction/?action=click&contentCollection=Books&referrer®ion=Header&module=ArrowNav&version=Right&pgtype=Reference>

Scrape the following information and present it in a dataframe (6 points):

1. Year
2. Title
3. Author
4. Number of weeks on the list
5. Publisher

A. Find the mean of the number of weeks the books have been on the list. and create a Box plot for the number of weeks. (2 points)

B. Create a horizontal bar plot for a sublist of the books that have been the longest on the list: the length of the bar depicts number of weeks. Additional information on the figure should include the name of the book and year. Save the plot. (2 points)

Submission Requirements

1. Create a directory called "Assignment 5B". Place the ipynb file and the plot inside this directory. Call it Assignment 5B.ipynb and Assignment5Bplot (with appropriate extension name). Submit the zipped directory.