

ECON 370: Lab #10 Questions

Write a script that scrapes Williams web pages to calculate the answers to the questions below. Make sure that the answers display when someone runs your code. If you are in a group of 1 or 2, you can skip Questions 1 and 2 (if you wish). If you are in a group of 3, you can skip Question 1 or Question 2 (but not both). Make sure your script clearly indicates (using comments) where the answer to one question ends and the answer to the next question begins.

1. What is the average height of players on the men's basketball team?
2. Which team has a higher proportion of athletes from New England: women's swimming and diving or men's swimming and diving?
3. Which department has a higher proportion of female faculty: economics or political science? Faculty are defined (in this context) as individuals listed on the department Faculty & Staff or People web page that have "professor" in their title and are not emeritus.
4. What proportion of economics faculty did their undergraduate study at a liberal arts college?
 - (a) Hint 1: faculty are defined as above. You can drop visiting faculty, since their CVs are not always available on the web.
 - (b) Hint 2: you will need to collect the urls for faculty profile pages from the main People page for the department, and then you will need to loop through those urls to collect data on faculty educational histories.
 - (c) Hint 3: you should clean the data on educational histories enough to separate undergraduate institutions from graduate institutions, and to make sure that the undergraduate institution is listed for every faculty member. If a data on a faculty member's undergraduate institution is missing, you should be able to get it from their CV ("by hand").