

STA312H5S Assignment 1

The state of Texas maintains a website with a list of all the people on death row executed since 1984. The website can be found here http://www.tdcj.state.tx.us/death_row/dr_executed_offenders.html. The website consists of a table with links to each inmates last statement, as well as some other information which is much less interesting (like ages, race, etc...). In this assignment you will build a program that scrapes the website and extracts the last statements of the inmates and puts it into an excel table.

You will find that I have graciously provided you with some code to start you off. Try running it! You will find that it outputs a beautiful excel file with all the information from the website except for the last statement. Before you begin this task, there are a few caveats I should warn you about.

- You will have to open a new page to get to the last statement, as the table on the webpage only contains a link to each statement.
- When you press the link you will notice that the last statement is NOT in a table. So if you look at the HTML code for the page (always a good idea), you will find something like this

```
<p> last statement </p>
```

So you should parse for the paragraph HTML tag and not the column or cell HTML tag.

- A few characters seem to not convert from unicode properly. There is a quick fix for this. You will want to replace these with the empty string. This is more of a technical issue, but nonetheless still needs to be addressed to be able to do this assignment properly.

Rememeber if you are confused you can ask the course instructor or me for help. I expect you to print out your code and submit it to me along with the excel file printout. This assignment is due at the beginning of next tutorial.

Good Luck!