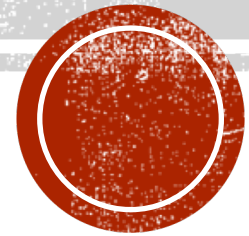


HW6 WEBSCRAPING

Data-X GSI Team



BEFORE WE START

- Beautiful Soup
 - [Data-X Web scraping Lecture Slides](#)
 - [Data-X Web scraping Code](#)
 - [Beautiful Soup Documentation](#)



■ Part I Statistics in Presidential Debates

- Find target content (third debate)
- Find required information

October 19, 2016: The Third Clinton- Trump Presidential Debate

Debate char length	93765
war_count	6
most_common_w	the
most_common_w_count	761

In This Section

→ [Debate Videos](#)

→ [Debate Transcripts](#)

→ [Citizen Resources](#)

→ [State Boards of Election](#)

Debate Transcripts

Unofficial transcripts of most presidential and vice presidential debates are available on this site.

2016 Transcripts

[September 26, 2016: The First Clinton-Trump Presidential Debate](#)

[October 4, 2016: The Kain-Pence Vice Presidential Debate](#)

[October 9, 2016: The Second Clinton-Trump Presidential Debate](#)

[October 19, 2016: The Third Clinton-Trump Presidential Debate](#)

2012 Transcripts

[October 3, 2012: The First Obama-Romney Presidential Debate](#)

[October 11, 2012: The Biden-Ryan Vice Presidential Debate](#)

[October 16, 2012: The Second Obama-Romney Presidential Debate](#)

[October 22, 2012: The Third Obama-Romney Presidential Debate](#)



■ Part II Download and read in specific line from many data sets

Authors	Counts
Helmut Spaeth	
S Chatterjee B Price	
R J Freund and P D Minton	
D G Kleinbaum and L L Kupper	
S C Narula J F Wellington	
K A Brownlee	
S Chatterjee and B Price	

Index of /~jburkardt/datasets/regression

- [Parent Directory](#)
- [regression.html](#)
- [x01.txt](#)
- [x02.txt](#)
- [x03.txt](#)
- [x04.txt](#)
- [x05.txt](#)
- [x06.txt](#)
- [x07.txt](#)

```
# x01.txt
#
# Reference:
#
# Helmut Spaeth,
# Algorithms for Linear Regression,
# Academic Press, 1991, page 304,
# ISBN 0-12-656460-4.
#
# S Weisberg,
# Applied Linear Regression,
# Wiley, 1980, pages 128-129.
#
# Discussion:
#
# The data records the average weight of the brain and body for
# a number of mammal species.
#
# There are 62 rows of data. The 3 data columns include:
#
# I, the index,
# A1, the brain weight;
# B, the body weight.
#
# We seek a model of the form:
#
# B = A1 * X1.
#
# 3 columns
# 62 rows
# Index
# Brain Weight
# Body Weight
#
# 1 3.385 44.500
# 2 0.480 15.500
# 3 1.350 8.100
# 4 465.000 423.000
# 5 36.330 119.500
# 6 27.660 115.000
# 7 14.830 98.200
# 8 1.040 5.500
# 9 4.190 58.000
# 10 0.425 6.400
# 11 0.101 4.000
# 12 0.920 5.700
# ...
```



SUBMISSION

1. You will be able to run otter grader on your local (pip install otter-grader)
2. Run the whole Jupyter Notebook (Kernal -> Restart & Run all)
3. Make sure all the outputs are shown correctly
4. Submit the Jupyter Notebook (.ipynb) to Gradescope

